(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112585268 B
(45)授权公告日 2025.03.04

(21)申请号 201980056488.0
(22)申请日 2019.06.28
(65)同一申请的已公布的文献号
申请公布号 CN 112585268 A
(43)申请公布日 2021.03.30
(30)优先权数据
62/691573 2018.06.28 US
(85)PCT国际申请进入国家阶段日
2021.02.26
(86)PCT国际申请的申请数据
PCT/IB2019/000833 2019.06.28
(87)PCT国际申请的公布数据
WO2020/003006 EN 2020.01.02
(73)专利权人 克里斯珀医疗股份公司
地址 瑞士楚格
(72)发明人 T·D·卡洛 R·L·博戈拉德
(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 黄登高 黄希贵
(51)Int.Cl.
C12N 15/11(2006.01)
C12N 15/90(2006.01)
A61K 31/7088(2006.01)
A61K 48/00(2006.01)
(56)对比文件
US 2011229880 A1,2011.09.22
WO 2017077386 A1,2017.05.11
S. J. ORLANDO ET AL.Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology.NUCLEIC ACIDS RESEARCH ADVANCE ACCESS.2010,第38卷(第15期),摘要,第2页右栏,图4.
审查员 李德利
权利要求书5页 说明书132页
序列表25页 附图10页

(54)发明名称

通过插入供体多核苷酸用于基因组编辑的组合物和方法

(57)摘要

本公开内容提供了校正或诱导gDNA中的突变的供体多核苷酸、基因组编辑系统、方法和试剂盒。

CN 112585268 B

1.一种包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述双链DNA分子包含校正细胞中的基因组DNA(gDNA)分子中的突变的核苷酸序列,所述突变引起糖原贮积病1a,其中所述突变是接近人G6PC基因中的3'剪接位点的蛋白质编码突变,所述供体多核苷酸包含:

(i)第一链,其从5'到3'包含(a)包含校正细胞中的突变的外显子序列的核苷酸序列和(b)包含一种或多种剪接信号以控制由gDNA转录的前体mRNA(前mRNA)的加工的核苷酸序列;和

(ii)第二链,其包含与第一链互补的核苷酸序列,

其中所述供体多核苷酸为长度10-100个核苷酸,并且其中当所述供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正突变,

其中所述一种或多种剪接信号是分支点、多聚嘧啶片、或天然或增强的3'剪接位点,和其中所述分支点包含核苷酸序列TTCAT,所述多聚嘧啶片包含核苷酸序列CTTGTTCTGTTTTTTTT(SEQ ID NO:109),和所述3'剪接位点包含核苷酸序列TAG。

2.权利要求1的供体多核苷酸,其中所述供体多核苷酸插入DSB内导致在包含所述外显子序列的gDNA中的外显子形成。

3.权利要求2的供体多核苷酸,其中所述一种或多种剪接信号指导外显子包含到mRNA中,所述外显子包含校正突变的外显子序列。

4.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸为长度20-80个核苷酸。

5.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸为长度30-70个核苷酸。

6.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸为长度40-60个核苷酸。

7.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸为长度40-70个核苷酸。

8.权利要求7的供体多核苷酸,其中所述供体多核苷酸为长度50-60个核苷酸。

9.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸为长度40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷酸。

10.权利要求9的供体多核苷酸,其中所述供体多核苷酸为长度50个核苷酸。

11.权利要求1-3中任一项的供体多核苷酸,其中所述定点核酸酶构成CRISPR/Cas系统。

12.权利要求11的供体多核苷酸,其中所述CRISPR/Cas系统包含一种或多种引导RNA(gRNA)。

13.权利要求12的供体多核苷酸,其中所述定点核酸酶包含衍生自化脓性链球菌的Cas9多肽(SpCas9)。

14.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸包含天然核苷酸。

15.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸包含一个或多个非

天然和/或修饰的核苷酸。

16.权利要求15的供体多核苷酸,其中所述一个或多个非天然和/或经修饰的核苷酸是2'-O-甲基核苷酸。

17.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸包含一种或多种主链修饰。

18.权利要求17的供体多核苷酸,其中所述一种或多种主链修饰是硫代磷酸酯。

19.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸包含两个平端。

20.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸包含一个平端,并且包括包含突出物的一端。

21.权利要求20的供体多核苷酸,其中所述突出物是5'或3'突出物。

22.权利要求1-3中任一项的供体多核苷酸,其中所述核苷酸序列包含一个或多个核苷酸,其阻止定点核酸酶识别且切割所述供体多核苷酸。

23.权利要求1的供体多核苷酸,其中所述突变位于人染色体17q21上的人G6PC基因中。

24.权利要求1的供体多核苷酸,其中所述突变是G6PC基因中的位置83处的精氨酸(R)的氨基酸取代,并且其中所述外显子序列包含与G6PC基因中的位置83处的密码子相对应的编码精氨酸(R)的密码子。

25.权利要求1的供体多核苷酸,其中所述G6PC基因中的突变导致人G6PC蛋白中的R83C氨基酸取代。

26.权利要求1的供体多核苷酸,其中所述G6PC基因中的突变导致人G6PC蛋白中的R83H氨基酸取代。

27.权利要求1-3中任一项的供体多核苷酸,其中所述外显子序列包含核苷酸序列GATTCTCTTTGGACAGCGCCCTTACT。

28.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸的核苷酸序列在SEQ ID NO:30(CH3454-0)中阐述。

29.权利要求1-3中任一项的供体多核苷酸,其中所述供体多核苷酸的核苷酸序列在SEQ ID NO:20(CH3250-0)中阐述。

30.一种包含非复制性dsDNA分子的供体多核苷酸,所述dsDNA分子包含校正细胞中的gDNA分子中的人葡萄糖苷酶α(GAA)基因突变的核苷酸序列,所述突变引起庞贝氏病1a,其中GAA的剪接信号中的所述突变导致缺乏外显子2和/或一个或多个隐蔽剪接位点的激活的GAA基因的mRNA转录物,所述供体多核苷酸包含:

(i)第一链,其从5'到3'包含校正突变的第一核苷酸序列,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号包含3'剪接位点、多聚嘧啶片和分支点序列的组合;和

(ii)第二链,其从5'到3'包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述一种或多种剪接信号包含3'剪接位点、多聚嘧啶片和分支点序列的组合,其中所述第二链与第一链互补,

其中所述供体多核苷酸为长度10-100个核苷酸,并包含两个平端,其中所述供体多核

苷酸的每条链的最5'核苷酸包含5'磷酸部分,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,NHEJ DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,其中所述供体多核苷酸被配置用于双向插入DSB断裂内,其中以任一方向的插入形成3'剪接位点、多聚嘧啶片和校正突变的分支点序列,其中所述剪接信号指导外显子2包括到mRNA内,从而校正了突变,

其中所述分支点包含核苷酸序列TATTAAC、GTTAATA、TACTGAC或TTAAC,所述多聚嘧啶片包含核苷酸序列CTTCTTCTCTTCTTCC或TTTTTTTCTTTTT,和所述3'剪接位点包含核苷酸序列TAG或CAG。

31.权利要求30的供体多核苷酸,其中所述供体多核苷酸为长度20-80个核苷酸。

32.权利要求30的供体多核苷酸,其中所述供体多核苷酸为长度30-70个核苷酸。

33.权利要求30的供体多核苷酸,其中所述供体多核苷酸为长度40-60个核苷酸。

34.权利要求30的供体多核苷酸,其中所述供体多核苷酸的核苷酸序列在SEQ ID NO:63(GAA_50-0)中阐述。

35.一种校正细胞中的基因组DNA分子(gDNA)中的突变的系统,所述系统包含以下组分:

(a)根据权利要求1-34中任一项的供体多核苷酸;

(b)一种或多种gRNA分子;和

(c)定点核酸酶,

其中当所述系统引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

36.权利要求35的系统,其中所述定点核酸酶由mRNA编码。

37.权利要求35的系统,其中所述定点核酸酶是多肽。

38.权利要求37的系统,其中所述一种或多种gRNA分子和所述定点核酸酶包含核糖核蛋白。

39.权利要求35-38中任一项的系统,其中所述定点核酸酶是Cas核酸酶。

40.权利要求39的系统,其中所述Cas核酸酶是化脓性链球菌Cas9(SpCas9),或者其同源物、衍生物或修饰形式。

41.权利要求35-38中任一项的系统,其中所述一种或多种gRNA分子包含选自以下的修饰:主链修饰、糖修饰、修饰的核苷间键合、或者修饰的或非天然的核碱基。

42.权利要求41的系统,其中所述一种或多种gRNA分子包含主链修饰。

43.权利要求42的系统,其中所述主链修饰是硫代磷酸酯修饰。

44.权利要求35-38中任一项的系统,其中所述供体多核苷酸、gRNA分子和定点核酸酶在脂质体或脂质纳米颗粒中个别地配制或共配制。

45.权利要求44的系统,其中所述供体多核苷酸、gRNA分子和定点核酸酶在脂质体或脂质纳米颗粒中个别地配制。

46.权利要求44的系统,其中所述供体多核苷酸、gRNA分子和定点核酸酶在脂质体或脂质纳米颗粒中共配制。

47.权利要求35-38中任一项的系统,其中所述gRNA的核苷酸序列包含SEQ ID NO:107

(CH32突变体-CTX1 sgRNA)中阐述的序列。

48.权利要求35-38中任一项的系统,其中所述gRNA的核苷酸序列包含SEQ ID NO:108(CH34突变体-CTX1 sgRNA)中阐述的序列。

49.权利要求35-38中任一项的系统,其中所述gRNA的核苷酸序列包含SEQ ID NO:93(突变体GAA sgRNA间隔区)中阐述的序列。

50.一种细胞,其包含权利要求1-34中任一项的供体多核苷酸或权利要求35-38中任一项的系统。

51.权利要求50的细胞,其中所述细胞是分裂细胞或非分裂细胞。

52.权利要求51的细胞,其中所述细胞是分裂细胞。

53.权利要求51的细胞,其中所述细胞是非分裂细胞。

54.权利要求50的细胞,其中所述细胞是患者特异性诱导多能干细胞(iPSC)。

55.权利要求50的细胞,其中所述细胞是原代肝细胞。

56.一种药物组合物,其包含根据权利要求1-34中任一项的供体多核苷酸和药学上可接受的载体。

57.一种药物组合物,其包含根据权利要求35的系统和药学上可接受的载体。

58.一种药物组合物,其包含根据权利要求50的细胞和药学上可接受的载体。

59.权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物,其用于通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者,所述治疗包括:从所述患者中分离细胞,使细胞与所述供体多核苷酸、系统或药物组合物接触,其中当所述供体多核苷酸、系统或药物组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

60.权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物,其用于通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者,所述治疗包括:向所述患者施用有效量的所述供体多核苷酸、系统或药物组合物,其中当施用所述供体多核苷酸、系统或组合物时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

61.权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物,用于制造通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者的药剂的用途。

62.根据权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物在制备药物中的用途,其中制备所述药物以用于校正细胞中的基因组DNA分子(gDNA)中的突变的方法,所述方法包括:使细胞与所述供体多核苷酸、所述系统、或所述药物组合物接触,其中当所述供体多核苷酸、系统或组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

63.根据权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物在制备药物中的用途,其中制备所述药物

以用于通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者的方法，所述方法包括:从所述患者中分离细胞,使细胞与所述供体多核苷酸、所述系统、或所述药物组合物接触,其中当所述供体多核苷酸、系统或组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

64.根据权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物在制备药物中的用途,其中制备所述药物以用于通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者的方法，所述方法包括:向所述患者施用有效量的所述供体多核苷酸、所述系统、或所述药物组合物,其中当施用所述供体多核苷酸、系统或组合物时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

65.权利要求62的用途,其中所述细胞是患者特异性诱导多能干细胞(iPSC)。

66.权利要求62的用途,其中所述细胞是肝细胞。

67.权利要求65的用途,其中所述方法进一步包括使包含校正的突变的iPSC分化成分化细胞;并且将所述分化细胞植入患者内。

68.权利要求62的用途,其中治疗导致由基因组DNA分子(gDNA)转录的mRNA的翻译,所述gDNA包含插入的供体多核苷酸,其中所述翻译导致翻译产物(蛋白质)的形成,所述翻译产物减轻疾病或者并不引起或促进疾病。

69.一种试剂盒,其包括容器和包含使用说明书的包装插页,所述容器包含根据权利要求1-34中任一项的供体多核苷酸、根据权利要求35-38中任一项的系统、或根据权利要求56-58中任一项的药物组合物,其用于校正细胞中的基因组DNA分子(gDNA)中的突变。

通过插入供体多核苷酸用于基因组编辑的组合物和方法

[0001] 相关信息

[0002] 本申请要求于2018年6月28日提交的美国临时申请序列号62/691,573的权益。上述临时专利申请的全部内容引入本文作为参考。

背景技术

[0003] 使用与设计的外源DNA修复模板分子组合的、可编程和/或改造的核酸酶的基因组编辑疗法,已开发用于治疗难治性病,例如病毒感染(Lin等人(2014)Mol Ther Nucleic Acids 3:e186)、酶缺乏症(Yin等人(2016)Nat Biotechnol 34(3):328-333)、以及遗传性肌病(Long等人(2014)Science 345:1184-1188;Long等人(2016)Science 351:400-403;Tabebordbar等人(2016)Science 351:407-411)。治疗性靶向基因组的方法经常依赖于同源性指导的修复(HDR)途径,其使得能够使用外源单链或双链DNA修复模板(例如供体多核苷酸)对引入的双链断裂(DSB)的准确基因组修复,但在非分裂细胞(例如,G_1期细胞)中经常被高度压制(Orthwein等人(2015)Nature 528(7582):422-6)。尽管在这个生物医学研究领域中的快速进展、以及用于临床应用的潜力,但由于目前的方法特别对于组成大多数成体组织的非分裂细胞是低效率的,用于体内治疗目的的转基因或其它多核苷酸的靶向整合仍是挑战性的。

[0004] 可以通过双链断裂修复(DSBR)机制,例如HDR途径和非同源末端连接(NHEJ)DNA修复途径,在细胞中修复通过外部来源(例如改造的核酸酶)在基因组内诱导的DSB。已知规范的HDR途径在分裂细胞(例如,处于S期的细胞)中起作用,因为它需要同源姐妹染色单体来执行,而NHEJ途径可以在分裂细胞和非分裂细胞两者中发挥功能,并且不依赖于细胞周期(Iyama & Wilson(2013)DNA Repair 12(8),620-636)。与HDR途径形成对比,NHEJ修复途径经常用于基因组编辑方法中,所述基因组编辑方法不采用外源DNA修复模板(例如供体多核苷酸),并且针对在DSB位点处的一个或多个核苷酸的随机插入或缺失('插入缺失(indel)')的形成,导致例如启动子或增强子中的编码序列的翻译读码框或反式作用因子的结合位点的破坏。

[0005] 由于遗传异常被广泛公认为众多疾病的主要病因学基础,因此治疗性靶向基因组编辑的无能或低效率对于开发用于广泛范围的遗传病症的治疗造成了技术障碍。因此,需要治疗性地靶向特别是非分裂细胞中的基因组的新方法。

发明内容

[0006] 准确的前信使RNA(前mRNA)剪接对于正确的蛋白质表达是至关重要的。脊椎动物基因体系结构经常由相对较长的内含子和较短的内部外显子组成。不受理论的束缚,与跨越内含子形成对比,外显子限定模型假定剪接位点跨越外显子配对,并且在具有大内含子的前mRNA中,剪接机制在外显子极性中搜索一对紧密间隔的剪接位点(即,上游的3'剪接位点和下游的5'剪接位点)(Berget(1995)J Biol Chem 270:2411-2414)。通过剪接机制识别外显子周围相对较短(50至300个核苷酸)区域中的一对剪接位点,降低了识别内含子内随

机分布的隐蔽内含子剪接位点的频率,因此增加了剪接的准确性。已知众多致病突变位于外显子内(例如,蛋白质编码突变)或内含子内(例如,剪接信号突变),其可以导致蛋白质生产或功能中的异常,最终导致或促成疾病。

[0007] 本公开内容至少部分地基于以下发现:可以通过基因组编辑组合物和方法,特别是通过使用如本文所述的供体多核苷酸,在基因组DNA分子(gDNA)中校正或诱导突变(例如,有害或致病突变),其导致gDNA的核苷酸序列中的所需改变并调节外显子限定,从而导致由编辑的gDNA转录的RNA转录物(例如,前mRNA)中包括所需改变。相应地,本公开内容提供了供体多核苷酸,当用于修复通过定点核酸酶(例如,Cas核酸酶)引入gDNA内的双链断裂(DSB)时,通过在校正或诱导的突变附近掺入一种或多种剪接信号,所述供体多核苷酸校正或诱导突变并调节外显子限定。在一些实施方案中,由本公开内容提供的供体多核苷酸被设计为包含核苷酸序列,所述核苷酸序列包含校正靶核酸(例如,基因组DNA)中的突变的所需改变、以及作用于调节剪接位点识别的一种或多种剪接信号。在一些实施方案中,由本公开内容提供的供体多核苷酸被设计为包含核苷酸序列,所述核苷酸序列包含所需改变和作用于调节外显子限定的一种或多种剪接信号,指导细胞的剪接机制将所需改变掺入转录产物(例如,前mRNA)内。

[0008] 本公开内容还提供了通过将线性、双链DNA供体多核苷酸插入在突变附近生成的CRISPR/Cas9介导的DSB内,用于校正或诱导gDNA中的突变(例如,致病突变)的方法,包括细胞、离体和体内方法。虽然活跃分裂细胞可以使用HDR和NHEJ途径两者来修复DNA损伤,但非分裂细胞占优势地使用NHEJ途径。CRISPR/Cas9的使用允许引入位点特异性DNA断裂,其可以通过NHEJ途径进行修复,使得将外源DNA多核苷酸连接或插入分裂细胞或非分裂细胞的基因组DNA内。在一些实施方案中,本公开内容的供体多核苷酸包含这样的核苷酸序列,其既校正或诱导突变又包含一种或多种剪接信号,从而通过建立关于所需外显子(例如,包含致病突变的外显子)的外显子限定、和/或破坏关于不希望有的外显子(例如,包含致病突变的外显子)的外显子限定,来促进所需RNA加工事件发生。本文还提供的是用于执行此类方法的组合物、系统和试剂盒。还提供的是通过此类方法产生的细胞。

[0009] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含:

[0010] (i)第一链,其从5'到3'包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的核苷酸序列、以及包含一种或多种剪接信号以控制由gDNA转录的前体mRNA(前mRNA)加工的核苷酸序列;和

[0011] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0012] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核

苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0013] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含:

[0014] (i)第一链,其从5'到3'包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的核苷酸序列、以及包含一种或多种剪接信号以控制由gDNA转录的前体mRNA(前mRNA)加工的核苷酸序列;和

[0015] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0016] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0017] 在一些实施方案中,突变是取代、错义、无义、插入、缺失或移码突变。在一些实施方案中,突变是取代突变。在一些实施方案中,突变是错义突变。在一些实施方案中,突变是无义突变。在一些实施方案中,突变是插入突变。在一些实施方案中,突变是缺失突变。在一些实施方案中,突变是移码突变。

[0018] 在一些实施方案中,突变在外显子中。在一些实施方案中,突变是取代、插入或缺失,并且突变在内含子中。在一些实施方案中,突变是位于内含子中的取代突变。在一些实施方案中,突变是位于内含子中的插入突变。在一些实施方案中,突变是位于内含子中的缺失突变。

[0019] 在一些实施方案中,突变接近gDNA中的剪接信号。在一些实施方案中,突变接近gDNA中的3'剪接位点。在一些实施方案中,突变接近gDNA中的5'剪接位点。在一些实施方案中,突变在gDNA中的剪接信号中。在一些实施方案中,突变在gDNA中的3'剪接位点中。在一些实施方案中,突变在gDNA中的5'剪接位点中。在一些实施方案中,突变在多聚嘧啶片中。在一些实施方案中,突变在分支点序列中。在一些实施方案中,突变是蛋白质编码突变。在一些实施方案中,突变与疾病相关或引起疾病。

[0020] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含内含子序列。在一些实施方案中,内含子序列校正了突变。

[0021] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含外显子序列。在一些实施方案中,外显子序列校正了突变。

[0022] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含选自以下的一种或多种剪接信号:

[0023] (a)天然或增强的3'剪接位点;

[0024] (b)天然或增强的5'剪接位点;

[0025] (c)多聚嘧啶片;

[0026] (d)分支点;

[0027] (e)外显子剪接增强子(ESE);

[0028] (f)内含子剪接增强子(ISE);

[0029] (g)外显子剪接沉默子(ESS);

[0030] (h)内含子剪接沉默子(ISS);和

[0031] (i)(a)-(h)中任一种的组合。

[0032] 在一些实施方案中,一种或多种剪接信号是天然或增强的3'剪接位点。在一些实施方案中,一种或多种剪接信号是天然的3'剪接位点。在一些实施方案中,一种或多种剪接信号是增强的3'剪接位点。在一些实施方案中,一种或多种剪接信号是天然或增强的5'剪接位点。在一些实施方案中,一种或多种剪接信号是天然的5'剪接位点。在一些实施方案中,一种或多种剪接信号是增强的3'剪接位点。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施方案中,一种或多种剪接信号是包含分支点的核苷酸序列。在一些实施方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,一种或多种剪接信号是包含天然或增强的3'剪接位点和多聚嘧啶片的组合。在一些实施方案中,一种或多种剪接信号是包含天然或增强的3'剪接位点、多聚嘧啶片和分支点的组合。

[0033] 在一些实施方案中,本公开内容提供了供体多核苷酸,其包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的非复制性双链DNA分子(dsDNA)、以及控制由gDNA转录的前体mRNA(前mRNA)加工的一种或多种剪接信号,其中所述供体多核苷酸包括包含分支点序列的第一剪接信号,其中所述供体多核苷酸包含:

[0034] (i)第一链,其从5'到3'包括包含第一分支点序列和校正gDNA中的突变的第一核苷酸序列的核苷酸序列;和

[0035] (ii)第二链,其从5'到3'包括包含第二分支点序列和校正gDNA中的突变的第二核苷酸序列的核苷酸序列,其中所述第二链与第一链互补,

[0036] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,包含第一剪接信号的核苷酸序列符合任一链上的分支点共有序列,其中第一分支点序列和第二分支点序列的核苷酸序列是互补的。

[0037] 在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一

些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0038] 在一些实施方案中,本公开内容提供了供体多核苷酸,其包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的非复制性双链DNA分子(dsDNA)、以及控制由gDNA转录的前体mRNA(前mRNA)加工的一种或多种剪接信号,其中所述供体多核苷酸包括包含分支点序列的第一剪接信号,其中所述供体多核苷酸包含:

[0039] (i)第一链,其从5'到3'包括包含第一分支点序列和校正gDNA中的突变的第一核苷酸序列的核苷酸序列;和

[0040] (ii)第二链,其从5'到3'包括包含第二分支点序列和校正gDNA中的突变的第二核苷酸序列的核苷酸序列,其中所述第二链与第一链互补,

[0041] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,包含第一剪接信号的核苷酸序列符合任一链上的分支点共有序列,其中第一分支点序列和第二分支点序列的核苷酸序列是互补的。

[0042] 在一些实施方案中,分支点共有序列是YTNAY(SEQ ID NO:49),其中Y是包含胞嘧啶(C)或胸腺嘧啶(T)核碱基的核苷酸,并且其中N是包含选自以下的核碱基的核苷酸:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C)。在一些实施方案中,Y包含胞嘧啶(C)。在一些实施方案中,Y包含胸腺嘧啶(T)。在一些实施方案中,N包含腺嘌呤(A)。在一些实施方案中,N包含鸟嘌呤(G)。在一些实施方案中,N包含胸腺嘧啶(T)。在一些实施方案中,N包含胞嘧啶(C)。

[0043] 在一些实施方案中,第一分支点序列是TATTAAC(SEQ ID NO:50)。

[0044] 在一些实施方案中,第二分支点序列是GTTAATA(SEQ ID NO:51)。

[0045] 在一些实施方案中,第二分支点序列是TACTGAC(SEQ ID NO:52)。

[0046] 在一些实施方案中,供体多核苷酸包括包含多聚嘧啶片的第二剪接信号,其中第一链包含位于第一分支点序列下游的第一多聚嘧啶片;并且第二链包含位于第二分支点序列下游的第二多聚嘧啶片。在一些实施方案中,包含第一多聚嘧啶片和第二多聚嘧啶片的核苷酸序列各自包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),并且其中所述核苷酸序列为约100%、约90%-100%、或约80%-90%的嘧啶核碱基。在一些实施方案中,核苷酸序列是约100%的嘧啶核碱基。在一些实施方案中,核苷酸序列是约90%-100%的嘧啶核碱基。在一些实施方案中,核苷酸序列是约80%-90%的嘧啶核碱基。

[0047] 在一些实施方案中,包含多聚嘧啶片的核苷酸序列是TTTTTTCT(SEQ ID NO:53)。在一些实施方案中,包含多聚嘧啶片的核苷酸序列是TTTTTTCTTTTT(SEQ ID NO:54)。在一些实施方案中,包含多聚嘧啶片的核苷酸序列是CTTCTTCTCTTCTTCC(SEQ ID NO:55)。

[0048] 在一些实施方案中,第一分支点序列和第一多聚嘧啶片彼此相邻。在一些实施方案中,第二分支点序列和第二多聚嘧啶片彼此相邻。

[0049] 在一些实施方案中,供体多核苷酸包含第三剪接信号,其中所述第三剪接信号包含3'剪接位点,其中第一链包含核苷酸序列,其包含位于第一多聚嘧啶片下游的第一3'剪接位点;并且其中第二链包含核苷酸序列,其包含位于第二多聚嘧啶片下游的第二3'剪接位点。在一些实施方案中,第一3'剪接位点和第二3'剪接位点包含核苷酸序列YAG,并且其中Y是包含选自胸腺嘧啶(T)和胞嘧啶(C)的核碱基的核苷酸。在一些实施方案中,Y包含胸腺嘧啶(T)。在一些实施方案中,Y包含胞嘧啶(C)。

[0050] 在一些实施方案中,供体多核苷酸包含编码序列,其中第一链包含第一编码序列,其中第二链包含第二编码序列,其中校正gDNA中的突变的第一核苷酸序列包含第一编码序列,其中校正gDNA中的突变的第二核苷酸序列包含第二编码序列,其中所述第一编码序列位于第一3'剪接位点的下游,并且其中所述第二编码序列位于第二3'剪接位点的下游。在一些实施方案中,包含第一编码序列和第二编码序列的核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C)。在一些实施方案中,包含(i)和(ii)的编码序列是不等同的或不互补的,以降低自退火。

[0051] 在一些实施方案中,供体多核苷酸包含一个或多个定界符序列(delimiter sequence),其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中所述核苷酸序列为长度约1-40、约1-30、约1-20、约1-15、约1-10、约30、约20、约10、约9、约8、约7、约6、约5、约4、约3、约2或1个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约1-40个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约1-30个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约1-20个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约1-15个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约1-10个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约30个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约20个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约10个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约9个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约8个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约7个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约6个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约5个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约4个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约3个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约2个核苷酸。在一些实施方案中,包含一个或多个定界符序列的核苷酸序列为长度约1个核苷酸。

[0052] 在一些实施方案中,一个或多个定界符序列位于第一分支点序列和第二分支点序列之间。在一些实施方案中,一个或多个定界符序列位于第一分支点序列和第一多聚嘧啶片之间。在一些实施方案中,一个或多个定界符序列位于第二分支点和第二多聚嘧啶片之间。

[0053] 在一些实施方案中,本公开内容提供了配置用于双向插入DSB内的供体多核苷酸,其中当供体多核苷酸以任一取向插入DSB内时,第一剪接信号和第二剪接信号、任选地第三

剪接信号和编码序列构成有义链,从而校正了突变并提供一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工。

[0054] 在一些实施方案中,本公开内容提供了供体多核苷酸,其包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的非复制性双链DNA分子(dsDNA)、以及控制由gDNA转录的前体mRNA(前mRNA)加工的一种或多种剪接信号,其中所述供体多核苷酸包含:

[0055] (i)第一链,其从5'到3'包括包含第一5'剪接位点序列和校正gDNA中的突变的第一核苷酸序列的核苷酸序列;和

[0056] (ii)第二链,其从5'到3'包括包含第二5'剪接位点序列和校正gDNA中的突变的第二核苷酸序列的核苷酸序列,其中所述第二链与第一链互补,

[0057] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,第一链包含第一编码序列,其中第二链包含第二编码序列,其中所述第一编码序列位于第一5'剪接位点的上游,并且其中第二编码序列位于第二5'剪接位点的上游,并且其中第一链和第二链中的编码序列是不互补的(或者包含一个、两个、三个、四个或更多个错配),以降低自退火。在一些实施方案中,供体多核苷酸包含位于第一5'剪接位点和第二5'剪接位点之间的定界符序列,其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中所述核苷酸序列为长度约1-40、约1-30、约1-20、约1-15、约1-10、约30、约20、约10、约9、约8、约7、约6、约5、约4、约3、约2或1个核苷酸。在一些实施方案中,供体多核苷酸被配置用于双向插入DSB内,其中当供体多核苷酸以第一取向插入DSB内时,第一5'剪接位点和第一编码序列构成有义链,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,并且其中当供体多核苷酸以第二取向插入DSB内时,第二5'剪接位点和第二编码序列构成有义链,从而校正了突变并提供一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0058] 在一些实施方案中,本公开内容提供了供体多核苷酸,其包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的非复制性双链DNA分子(dsDNA)、以及控制由gDNA转录的前体mRNA(前mRNA)加工的一种或多种剪接信号,其中所述供体多核苷酸包含:

[0059] (i)第一链,其从5'到3'包括包含第一5'剪接位点序列和校正gDNA中的突变的第一核苷酸序列的核苷酸序列;和

[0060] (ii)第二链,其从5'到3'包括包含第二5'剪接位点序列和校正gDNA中的突变的第

二核苷酸序列的核苷酸序列,其中所述第二链与第一链互补,

[0061] 其中所述供体多核苷酸的长度为约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,第一链包含第一编码序列,其中第二链包含第二编码序列,其中所述第一编码序列位于第一5'剪接位点的上游,并且其中第二编码序列位于第二5'剪接位点的上游,并且其中第一链和第二链中的编码序列是不互补的(或者包含一个、两个、三个、四个或更多个错配),以降低自退火。在一些实施方案中,供体多核苷酸包含位于第一5'剪接位点和第二5'剪接位点之间的定界符序列,其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中所述核苷酸序列为长度约1-40、约1-30、约1-20、约1-15、约1-10、约30、约20、约10、约9、约8、约7、约6、约5、约4、约3、约2或1个核苷酸。在一些实施方案中,供体多核苷酸被配置用于双向插入DSB内,其中当供体多核苷酸以第一取向插入DSB内时,第一5'剪接位点和第一编码序列构成有义链,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,并且其中当供体多核苷酸以第二取向插入DSB内时,第二5'剪接位点和第二编码序列构成有义链,从而校正了突变并提供一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工。

[0062] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0063] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0064] (ii)第二链,其包含与第一链互补的核苷酸序列,其中所述一种或多种剪接信号选自:

[0065] (a)天然或增强的3'剪接位点;

[0066] (b)多聚嘧啶片;

[0067] (c)分支点;

[0068] (d)外显子剪接增强子(ESE);

[0069] (e)内含子剪接增强子(ISE);

[0070] (f)外显子剪接沉默子(ESS);

[0071] (g)内含子剪接沉默子(ISS);和

[0072] (h)(a)-(g)中任一种的组合,

[0073] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,一种或多种剪接信号是天然或增强的3'剪接位点。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施

方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0074] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0075] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0076] (ii)第二链,其包含与第一链互补的核苷酸序列,其中所述一种或多种剪接信号选自:

[0077] (a)天然或增强的3'剪接位点;

[0078] (b)多聚嘧啶片;

[0079] (c)分支点;

[0080] (d)外显子剪接增强子(ESE);

[0081] (e)内含子剪接增强子(ISE);

[0082] (f)外显子剪接沉默子(ESS);

[0083] (g)内含子剪接沉默子(ISS);和

[0084] (h)(a)-(g)中任一种的组合,

[0085] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,一种或多种剪接信号是天然或增强的3'剪接位点。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-

60个核苷酸。

[0086] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0087] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA(gDNA)分子中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0088] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0089] 其中所述一种或多种剪接信号选自:

[0090] (a)天然或增强的5'剪接位点;

[0091] (b)外显子剪接增强子(ESE);

[0092] (c)内含子剪接增强子(ISE);

[0093] (d)外显子剪接沉默子(ESS);

[0094] (e)内含子剪接沉默子(ISS);和

[0095] (f)(a)-(e)中任一种的组合,

[0096] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,一种或多种剪接信号是天然或增强的5'剪接位点。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0097] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0098] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA(gDNA)分子中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0099] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0100] 其中所述一种或多种剪接信号选自:

[0101] (a)天然或增强的5'剪接位点;

[0102] (b)外显子剪接增强子(ESE);

[0103] (c)内含子剪接增强子(ISE);

[0104] (d)外显子剪接沉默子(ESS);

[0105] (e)内含子剪接沉默子(ISS);和

[0106] (f)(a)-(e)中任一种的组合,

[0107] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,一种或多种剪接信号是天然或增强的5'剪接位点。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0108] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0109] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中至少一种剪接信号是天然或增强的3'剪接位点;和

[0110] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0111] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,

供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0112] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0113] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中至少一种剪接信号是天然或增强的3'剪接位点;和

[0114] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0115] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0116] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0117] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号是天然或增强的3'剪接位点和多聚嘧啶片的组合;和

[0118] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0119] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0120] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体

多核苷酸,所述dsDNA包含

[0121] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号是天然或增强的3’剪接位点和多聚嘧啶片的组合;和

[0122] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0123] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0124] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0125] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号是天然或增强的3’剪接位点、多聚嘧啶片和分支点的组合;和

[0126] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0127] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0128] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0129] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第一链包

含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号是天然或增强的3'剪接位点、多聚嘧啶片和分支点的组合;和

[0130] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0131] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0132] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0133] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变在3'剪接位点中,其中所述第一链包含内含子序列,任选地外显子序列,其中所述内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0134] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0135] 其中所述一种或多种剪接信号选自:

[0136] (a)多聚嘧啶片;

[0137] (b)分支点;

[0138] (c)外显子剪接增强子(ESE);

[0139] (d)内含子剪接增强子(ISE);

[0140] (e)外显子剪接沉默子(ESS);

[0141] (f)内含子剪接沉默子(ISS);和

[0142] (g)(a)-(f)中任一种的组合,

[0143] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷

酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0144] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0145] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变在3'剪接位点中,其中所述第一链包含内含子序列,任选地外显子序列,其中所述内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0146] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0147] 其中所述一种或多种剪接信号选自:

[0148] (a)多聚嘧啶片;

[0149] (b)分支点;

[0150] (c)外显子剪接增强子(ESE);

[0151] (d)内含子剪接增强子(ISE);

[0152] (e)外显子剪接沉默子(ESS);

[0153] (f)内含子剪接沉默子(ISS);和

[0154] (g)(a)-(f)中任一种的组合,

[0155] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,一种或多种剪接信号是多聚嘧啶片。在一些实施方案中,一种或多种剪接信号是分支点。在一些实施方案中,一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中,一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中,一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中,一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0156] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0157] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述供体多核苷酸包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中至少一种剪接信号是天然或增强的5'剪接位点;和

[0158] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0159] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-

100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0160] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0161] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近5’剪接位点的蛋白质编码突变,其中所述供体多核苷酸包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中至少一种剪接信号是天然或增强的5’剪接位点;和

[0162] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0163] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0164] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0165] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变在5’剪接位点中,其中所述第一链包含内含子序列,任选地外显子序列,其中所述内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0166] (ii)第二链,其包含与第一链互补的核苷酸序列,其中所述一种或多种剪接信号选自:

[0167] (a)外显子剪接增强子(ESE);

[0168] (b)内含子剪接增强子(ISE);

[0169] (c)外显子剪接沉默子(ESS);

[0170] (d)内含子剪接沉默子(ISS);和

[0171] (e)(a)-(d)中任一种的组合，

[0172] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸，并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时，非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内，所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内，从而校正了突变。在一些实施方案中，一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中，一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中，一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中，一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中，供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中，供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中，供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中，供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中，供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中，供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中，供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中，供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中，供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中，供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中，供体多核苷酸为长度约40-60个核苷酸。

[0173] 在一些实施方案中，本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸，所述dsDNA包含

[0174] (i)第一链，其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列，其中所述致病突变在5’剪接位点中，其中所述第一链包含内含子序列，任选地外显子序列，其中所述内含子序列校正了突变，其中所述第一链包含一种或多种剪接信号，以控制由gDNA转录的前体mRNA(前mRNA)的加工；和

[0175] (ii)第二链，其包含与第一链互补的核苷酸序列，其中所述一种或多种剪接信号选自：

[0176] (a)外显子剪接增强子(ESE)；

[0177] (b)内含子剪接增强子(ISE)；

[0178] (c)外显子剪接沉默子(ESS)；

[0179] (d)内含子剪接沉默子(ISS)；和

[0180] (e)(a)-(d)中任一种的组合，

[0181] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸，并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时，非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内，所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内，从而校正了突变。在一些实施方案中，一种或多种剪接信号是外显子剪接增强子(ESE)。在一些实施方案中，一种或多种剪接信号是内含子剪接增强子(ISE)。在一些实施方案中，一种或多种剪接信号是外显子剪接沉默子(ESS)。在一些实施方案中，一种或多种剪接信号是内含子剪接沉默子(ISS)。在一些实施方案中，供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中，供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中，供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中，供体多核苷酸为长度

约40-60个核苷酸。

[0182] 在一些实施方案中,供体多核苷酸插入DSB内导致在包含外显子序列的gDNA中的外显子形成。在一些实施方案中,一种或多种剪接信号指导包含外显子序列的外显子的包括,所述外显子序列校正mRNA内的突变。

[0183] 在一些实施方案中,供体多核苷酸的插入导致包含内含子序列的内含子形成,其中所述内含子序列校正了突变。

[0184] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含第一链和第二链,其中所述第二链与第一链互补,其中所述第一链从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述第一链包含下式:

[0185] 5'-$[B]_a$-$[S1]_b$-$[P]_c$-$[S2]_d$-X-E-3',其中

[0186] (i)B(如果存在的话)是分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a是整数,其值指示B包含的核苷酸数目,其中a=0或5-7;

[0187] (ii)P是多聚嘧啶片,其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c是整数,其值指示P包含的核苷酸数目,其中c=9-20,其中构成P的核苷酸序列为约100%、约90%-100%、约80%-90%的嘧啶核碱基;

[0188] (iii)E是包含校正突变的核苷酸序列的外显子序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C);

[0189] (iv)X是包含3'剪接位点的核苷酸序列;和

[0190] (v)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和d各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和d = 0-20,

[0191] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于定向插入双链DNA断裂(DSB)内,其中当供体多核苷酸插入DSB内时,B(如果存在的话)、P、X(如果存在的话)和E(如果存在的话)构成有义链,其中B(如果存在的话)、P和X(如果存在的话)包含一种或多种剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0192] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在B。在一些实施方案中,不存在B。在一些实施方案中,a = 0。在一些实施方案中,a = 5-7。在一些实施方案中,a = 5。在一些实施方案中,a = 6。在一些实施方案中,a = 7。在一些实施方案中,构成P的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案

中,c = 9。在一些实施方案中,c = 10。在一些实施方案中,c = 11。在一些实施方案中,c = 12。在一些实施方案中,c = 13。在一些实施方案中,c = 14。在一些实施方案中,c = 15。在一些实施方案中,c = 16。在一些实施方案中,c = 17。在一些实施方案中,c = 18。在一些实施方案中,c = 19。在一些实施方案中,c = 20。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,d = 0。在一些实施方案中,d = 1。在一些实施方案中,d = 2。在一些实施方案中,d = 3。在一些实施方案中,d = 4。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,d = 8。在一些实施方案中,d = 9。在一些实施方案中,d = 10。在一些实施方案中,d = 11。在一些实施方案中,d = 12。在一些实施方案中,d = 13。在一些实施方案中,d = 14。在一些实施方案中,d = 15。在一些实施方案中,d = 16。在一些实施方案中,d = 17。在一些实施方案中,d = 18。在一些实施方案中,d = 19。在一些实施方案中,d = 20。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0193] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含第一链和第二链,其中所述第二链与第一链互补,其中所述第一链从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述第一链包含下式:

[0194] 5’-$[B]_a$-$[S1]_b$-$[P]_c$-$[S2]_d$-X-E-3’,其中

[0195] (i)B(如果存在的话)是分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a是整数,其值指示B包含的核苷酸数目,其中a=0或5-7;

[0196] (ii)P是多聚嘧啶片,其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c是整数,其值指示P包含的核苷酸数目,其中c=9-

20,其中构成P的核苷酸序列为约100%、约90%-100%、约80%-90%的嘧啶核碱基;

[0197] (iii)E是包含校正突变的核苷酸序列的外显子序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C);

[0198] (iv)X是包含3'剪接位点的核苷酸序列;和

[0199] (v)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和d各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和d = 0-20,

[0200] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于定向插入双链DNA断裂(DSB)内,其中当供体多核苷酸插入DSB内时,B(如果存在的话)、P、X(如果存在的话)和E(如果存在的话)构成有义链,其中B(如果存在的话)、P和X(如果存在的话)包含一种或多种剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0201] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在B。在一些实施方案中,不存在B。在一些实施方案中,a = 0。在一些实施方案中,a = 5-7。在一些实施方案中,a = 5。在一些实施方案中,a = 6。在一些实施方案中,a = 7。在一些实施方案中,构成P的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,c = 9。在一些实施方案中,c = 10。在一些实施方案中,c = 11。在一些实施方案中,c = 12。在一些实施方案中,c = 13。在一些实施方案中,c = 14。在一些实施方案中,c = 15。在一些实施方案中,c = 16。在一些实施方案中,c = 17。在一些实施方案中,c = 18。在一些实施方案中,c = 19。在一些实施方案中,c = 20。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,d = 0。在一些实施方案中,d = 1。在一些实施方案中,d = 2。在一些实施方案中,d = 3。在一些实施方案中,d = 4。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,d = 8。在一些实施方案中,d = 9。在一些实施方案中,d = 10。在一些实施方案中,d = 11。在一些实施方案中,d = 12。在一些实施方案中,d = 13。在一些实施方案中,d = 14。在一些实施方案中,d = 15。在一些实施方案中,d = 16。在一些实施方案中,d = 17。在一些实施方案中,d = 18。在一些实施方案中,d = 19。在一些实施方案中,d = 20。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80

个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0202] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含第一链和第二链,其中所述第二链与第一链互补,其中所述第一链从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,

[0203] 其中所述第一链包含下式:

[0204] 5'-E-Y-I-3',其中

[0205] (i)E是包含校正突变的核苷酸序列的外显子序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C);

[0206] (ii)Y是包含5'剪接位点的核苷酸序列;和

[0207] (iii)I(如果存在的话)包含内含子序列,其包括包含选自以下的核碱基的核苷酸:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),

[0208] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于定向插入双链DNA断裂(DSB)内,其中当供体多核苷酸插入DSB内时,E(如果存在的话)、Y和I(如果存在的话)构成有义链,其中Y包含一种或多种剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0209] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在E。在一些实施方案中,不存在E。在一些实施方案中,存在I。在一些实施方案中,不存在I。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0210] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含第一链和第二链,其中所述第二链与第一链互补,其中所述第一链从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,

[0211] 其中所述第一链包含下式:

[0212] 5'-E-Y-I-3',其中

[0213] (i)E是包含校正突变的核苷酸序列的外显子序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C);

[0214] (ii)Y是包含5'剪接位点的核苷酸序列;和

[0215] (iii)I(如果存在的话)包含内含子序列,其包括包含选自以下的核碱基的核苷酸:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),

[0216] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于定向插入双链DNA断裂(DSB)内,其中当供体多核苷酸插入DSB内时,E(如果存在的话)、Y和I(如果存在的话)构成有义链,其中Y包含一种或多种剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0217] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在E。在一些实施方案中,不存在E。在一些实施方案中,存在I。在一些实施方案中,不存在I。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0218] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0219] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变接近3'剪接位点,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0220] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变接近3'剪接位点,其中所述第二链包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0221] 5'-$[P1]_a$-$[S1]_b$-$[B]_c$-$[S2]_d$-$[P2]_e$-3',其中

[0222] (a)B包含分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c是整数,其值指示B包含的核苷酸数目,其中c = 5-7;

[0223] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和e是整数,其值分别指示P1和P2包含的核苷酸数目,其中a=9-20和e=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,并且其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;和

[0224] (c)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧

啶(C),其中b和d各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和d = 0-20,

[0225] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B和P2构成有义链,并且B和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B和P1构成有义链,并且B和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0226] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,a = 9。在一些实施方案中,a = 10。在一些实施方案中,a = 11。在一些实施方案中,a = 12。在一些实施方案中,a = 13。在一些实施方案中,a = 14。在一些实施方案中,a = 15。在一些实施方案中,a = 16。在一些实施方案中,a = 17。在一些实施方案中,a = 18。在一些实施方案中,a = 19。在一些实施方案中,a = 20。在一些实施方案中,e = 9。在一些实施方案中,e = 10。在一些实施方案中,e = 11。在一些实施方案中,e = 12。在一些实施方案中,e = 13。在一些实施方案中,e = 14。在一些实施方案中,e = 15。在一些实施方案中,e = 16。在一些实施方案中,e = 17。在一些实施方案中,e = 18。在一些实施方案中,e = 19。在一些实施方案中,e = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,d = 0。在一些实施方案中,d = 1。在一些实施方案中,d = 2。在一些实施方案中,d = 3。在一些实施方案中,d = 4。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,d = 8。在一些实施方案中,d = 9。在一些实施方案中,d = 10。在一些实施方案中,d = 11。在一些实施方案中,d = 12。在一些实施方案中,d = 13。在一些实施方案中,d = 14。在一些实施方案中,d = 15。在一些实施方案中,d = 16。在一些实施方案中,d = 17。在一些实施方案中,d = 18。在一些实施方案中,d = 19。在一些实施方案中,d = 20。在一些实施方案中,供体多核苷酸

为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0227] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0228] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变接近3’剪接位点,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0229] (ii)第二链,其从5’到3’包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变接近3’剪接位点,其中所述第二链包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0230] 5’-$[P1]_a$-$[S1]_b$-$[B]_c$-$[S2]_d$-$[P2]_e$-3’,其中

[0231] (a)B包含分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c是整数,其值指示B包含的核苷酸数目,其中c = 5-7;

[0232] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和e是整数,其值分别指示P1和P2包含的核苷酸数目,其中a=9-20和e=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,并且其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;和

[0233] (c)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和d各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和d = 0-20,

[0234] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B和P2构成有义链,并且B和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B和P1构成有义链,并且B和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0235] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接

(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,a = 9。在一些实施方案中,a = 10。在一些实施方案中,a = 11。在一些实施方案中,a = 12。在一些实施方案中,a = 13。在一些实施方案中,a = 14。在一些实施方案中,a = 15。在一些实施方案中,a = 16。在一些实施方案中,a = 17。在一些实施方案中,a = 18。在一些实施方案中,a = 19。在一些实施方案中,a = 20。在一些实施方案中,e = 9。在一些实施方案中,e = 10。在一些实施方案中,e = 11。在一些实施方案中,e = 12。在一些实施方案中,e = 13。在一些实施方案中,e = 14。在一些实施方案中,e = 15。在一些实施方案中,e = 16。在一些实施方案中,e = 17。在一些实施方案中,e = 18。在一些实施方案中,e = 19。在一些实施方案中,e = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,d = 0。在一些实施方案中,d = 1。在一些实施方案中,d = 2。在一些实施方案中,d = 3。在一些实施方案中,d = 4。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,d = 8。在一些实施方案中,d = 9。在一些实施方案中,d = 10。在一些实施方案中,d = 11。在一些实施方案中,d = 12。在一些实施方案中,d = 13。在一些实施方案中,d = 14。在一些实施方案中,d = 15。在一些实施方案中,d = 16。在一些实施方案中,d = 17。在一些实施方案中,d = 18。在一些实施方案中,d = 19。在一些实施方案中,d = 20。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0236] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0237] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变接近3’剪接位点,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0238] (ii)第二链,其从5’到3’包含校正细胞中的gDNA中的致病突变的第二核苷酸序

列,其中所述致病突变接近3'剪接位点,其中所述第二链包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0239] 5'-$[P1]_a$-$[S1]_b$-$[B1]_c$-$[S2]_d$-$[B2]_e$-$[S3]_f$-$[P2]_g$-3',其中

[0240] (a)B1(如果存在的话)和B2各自为分支点序列,其中B2包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中构成B1(如果存在的话)的分支点序列相对于构成B2的分支点序列处于相反取向,并且在供体多核苷酸的相反链上,其中B1和B2各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c和e是整数,其值分别指示B1和B2包含的核苷酸数目,其中c = 0或5-7,其中e = 5-7;

[0241] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中a和g是整数,其值分别指示P1和P2包含的核苷酸数目,其中a = 9-20和g = 9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;和

[0242] (c)S1、S2和S3(如果任一种存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b、d和f各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和f = 0或1-20,其中d = 0或1-40,

[0243] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B2和P2构成有义链,并且B2和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B1和P1构成有义链,并且B1和P1分别提供了第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0244] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在B1。在一些实施方案中,不存在B1。在一些实施方案中,c = 0。在一些实施方案中,c = 5-7。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,e = 5。在一些实施方案中,e = 6。在一些实施方案中,e = 7。在一些实施方案中,a = 9。在一些实施方案中,a = 10。在一些实施方案中,a = 11。在一些实施方案中,a = 12。在一些实施方案中,a = 13。在一些实施方案中,a = 14。在一些实施方案中,a = 15。在一些实施方案中,a = 16。在一些实施方案中,a = 17。在一些实施方案中,a = 18。在一些实施方案中,a = 19。在一些实施方案中,a = 20。在一些实施方案中,g = 9。在一些实施方案中,g = 10。在一些实施方案中,g = 11。在一些实施方案中,g = 12。在一些实施方案中,g = 13。在一些实施方案中,g = 14。在一些实施方案中,g = 15。在一些实施方案中,g = 16。在一些实施方案中,g = 17。在一些实施方案中,g = 18。在一些实施方案中,g = 19。在一些实施方案中,g = 20。在一

些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,存在S3。在一些实施方案中,不存在S3。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,f = 0。在一些实施方案中,f = 1。在一些实施方案中,f = 2。在一些实施方案中,f = 3。在一些实施方案中,f = 4。在一些实施方案中,f = 5。在一些实施方案中,f = 6。在一些实施方案中,f = 7。在一些实施方案中,f = 8。在一些实施方案中,f = 9。在一些实施方案中,f = 10。在一些实施方案中,f = 11。在一些实施方案中,f = 12。在一些实施方案中,f = 13。在一些实施方案中,f = 14。在一些实施方案中,f = 15。在一些实施方案中,f = 16。在一些实施方案中,f = 17。在一些实施方案中,f = 18。在一些实施方案中,f = 19。在一些实施方案中,f = 20。在一些实施方案中,d = 0。在一些实施方案中,d = 1。在一些实施方案中,d = 2。在一些实施方案中,d = 3。在一些实施方案中,d = 4。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,d = 8。在一些实施方案中,d = 9。在一些实施方案中,d = 10。在一些实施方案中,d = 11。在一些实施方案中,d = 12。在一些实施方案中,d = 13。在一些实施方案中,d = 14。在一些实施方案中,d = 15。在一些实施方案中,d = 16。在一些实施方案中,d = 17。在一些实施方案中,d = 18。在一些实施方案中,d = 19。在一些实施方案中,d = 20。在一些实施方案中,d = 21。在一些实施方案中,d = 22。在一些实施方案中,d = 23。在一些实施方案中,d = 24。在一些实施方案中,d = 25。在一些实施方案中,d = 26。在一些实施方案中,d = 27。在一些实施方案中,d = 28。在一些实施方案中,d = 29。在一些实施方案中,d = 30。在一些实施方案中,d = 31。在一些实施方案中,d = 32。在一些实施方案中,d = 33。在一些实施方案中,d = 34。在一些实施方案中,d = 35。在一些实施方案中,d = 36。在一些实施方案中,d = 37。在一些实施方案中,d = 38。在一些实施方案中,d = 39。在一些实施方案中,d = 40。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷

酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0245] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0246] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变接近3'剪接位点,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0247] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变接近3'剪接位点,其中所述第二链包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0248] 5'-$[P1]_a$-$[S1]_b$-$[B1]_c$-$[S2]_d$-$[B2]_e$-$[S3]_f$-$[P2]_g$-3',其中

[0249] (a)B1(如果存在的话)和B2各自为分支点序列,其中B2包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中构成B1(如果存在的话)的分支点序列相对于构成B2的分支点序列处于相反取向,并且在供体多核苷酸的相反链上,其中B1和B2各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c和e是整数,其值分别指示B1和B2包含的核苷酸数目,其中c = 0或5-7,其中e = 5-7;

[0250] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中a和g是整数,其值分别指示P1和P2包含的核苷酸数目,其中a = 9-20和g = 9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;和

[0251] (c)S1、S2和S3(如果任一种存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b、d和f各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和f = 0或1-20,其中d = 0或1-40,

[0252] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B2和P2构成有义链,并且B2和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B1和P1构成有义链,并且B1和P1分别提供了第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0253] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在B1。在一些实施方案中,不存在B1。在一些实施方案中,c = 0。在一些实施方案中,c = 5-7。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,e = 5。在一些实施

方案中,e = 6。在一些实施方案中,e = 7。在一些实施方案中,a = 9。在一些实施方案中,a = 10。在一些实施方案中,a = 11。在一些实施方案中,a = 12。在一些实施方案中,a = 13。在一些实施方案中,a = 14。在一些实施方案中,a = 15。在一些实施方案中,a = 16。在一些实施方案中,a = 17。在一些实施方案中,a = 18。在一些实施方案中,a = 19。在一些实施方案中,a = 20。在一些实施方案中,g = 9。在一些实施方案中,g = 10。在一些实施方案中,g = 11。在一些实施方案中,g = 12。在一些实施方案中,g = 13。在一些实施方案中,g = 14。在一些实施方案中,g = 15。在一些实施方案中,g = 16。在一些实施方案中,g = 17。在一些实施方案中,g = 18。在一些实施方案中,g = 19。在一些实施方案中,g = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,存在S3。在一些实施方案中,不存在S3。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,f = 0。在一些实施方案中,f = 1。在一些实施方案中,f = 2。在一些实施方案中,f = 3。在一些实施方案中,f = 4。在一些实施方案中,f = 5。在一些实施方案中,f = 6。在一些实施方案中,f = 7。在一些实施方案中,f = 8。在一些实施方案中,f = 9。在一些实施方案中,f = 10。在一些实施方案中,f = 11。在一些实施方案中,f = 12。在一些实施方案中,f = 13。在一些实施方案中,f = 14。在一些实施方案中,f = 15。在一些实施方案中,f = 16。在一些实施方案中,f = 17。在一些实施方案中,f = 18。在一些实施方案中,f = 19。在一些实施方案中,f = 20。在一些实施方案中,d = 0。在一些实施方案中,d = 1。在一些实施方案中,d = 2。在一些实施方案中,d = 3。在一些实施方案中,d = 4。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,d = 8。在一些实施方案中,d = 9。在一些实施方案中,d = 10。在一些实施方案中,d = 11。在一些实施方案中,d = 12。在一些实施方案中,d = 13。在一些实施方案中,d = 14。在一些实施方案中,d = 15。在一些实施方案中,d = 16。在一些实施方案中,d = 17。在一些实施方案中,d = 18。在一些实施方案中,d = 19。在一些实施方案中,d = 20。在一些实施方案中,d = 21。在一些实施方案中,d = 22。在一些实施方案中,d = 23。在一些实施方案中,d = 24。在一些实施方案中,d = 25。在一些实施方案中,d = 26。在一些实施方案中,d = 27。在一些实施方案中,d = 28。在一些实施方案中,d = 29。在一些实施方案中,d = 30。在一些实施方案中,d = 31。在一些实施方案中,d = 32。在一些实施方案中,d = 33。在一些实施方案中,d = 34。在一些实施方案中,d = 35。在一些实施方案中,d = 36。在一些实

施方案中,d = 37。在一些实施方案中,d = 38。在一些实施方案中,d = 39。在一些实施方案中,d = 40。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0254] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0255] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0256] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0257] 5'-$[E1]_a$-X1-$[P1]_b$-$[S1]_c$-$[B]_d$-$[S2]_e$-$[P2]_f$-X2-$[E2]_g$-3',其中

[0258] (a)B是分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中d是整数,其值指示B包含的核苷酸数目,其中d = 5-7;

[0259] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和f是整数,其值分别指示P1和P2包含的核苷酸数目,其中b=9-20和f=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;

[0260] (c)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和g是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在供体多核苷酸的相反链上,其中构成E1和E2的核苷酸序列不互补;

[0261] (d)X1和X2各自为包含3'剪接位点的核苷酸序列,其中构成X1的核苷酸序列相对于构成X2的核苷酸序列处于相反方向,且在相反链上;和

[0262] (e)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中c和e各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中c = 0或1-20和e = 0或1-20,

[0263] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B、P2、E2和X2构成有义链,其中B、P2和X2分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变

并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B、P1、E1和X1构成有义链,其中B、P1和X1分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0264] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,f = 9。在一些实施方案中,f = 10。在一些实施方案中,f = 11。在一些实施方案中,f = 12。在一些实施方案中,f = 13。在一些实施方案中,f = 14。在一些实施方案中,f = 15。在一些实施方案中,f = 16。在一些实施方案中,f = 17。在一些实施方案中,f = 18。在一些实施方案中,f = 19。在一些实施方案中,f = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,c = 0。在一些实施方案中,c = 1。在一些实施方案中,c = 2。在一些实施方案中,c = 3。在一些实施方案中,c = 4。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,c = 8。在一些实施方案中,c = 9。在一些实施方案中,c = 10。在一些实施方案中,c = 11。在一些实施方案中,c = 12。在一些实施方案中,c = 13。在一些实施方案中,c = 14。在一些实施方案中,c = 15。在一些实施方案中,c = 16。在一些实施方案中,c = 17。在一些实施方案中,c = 18。在一些实施方案中,c = 19。在一些实施方案中,c = 20。在一些实施方案中,e = 0。在一些实施方案中,e = 1。在一些实施方案中,e = 2。在一些实施方案中,e = 3。在一些实施方案中,e = 4。在一些实施方案中,e = 5。在一些实施方案中,e = 6。在一些实施方案中,e = 7。在一些实施方案中,e = 8。在一些实施方案中,e = 9。在一些实施方案中,e = 10。在一些实施方案中,e = 11。在一些实施方案中,e = 12。在一些实施方案中,e = 13。在一些实施方案中,e = 14。在一些实施方案中,e = 15。在一些实施方案中,e = 16。在一些实施方案中,e = 17。在一些实施方案中,e = 18。在一些实施方案中,e = 19。在一些实施方案中,e = 20。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施

方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0265] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0266] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0267] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0268] 5'-$[E1]_a$-X1-$[P1]_b$-$[S1]_c$-$[B]_d$-$[S2]_e$-$[P2]_f$-X2-$[E2]_g$-3',其中

[0269] (a)B是分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中d是整数,其值指示B包含的核苷酸数目,其中d = 5-7;

[0270] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和f是整数,其值分别指示P1和P2包含的核苷酸数目,其中b=9-20和f=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;

[0271] (c)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和g是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在供体多核苷酸的相反链上,其中构成E1和E2的核苷酸序列不互补;

[0272] (d)X1和X2各自为包含3'剪接位点的核苷酸序列,其中构成X1的核苷酸序列相对于构成X2的核苷酸序列处于相反方向,且在相反链上;和

[0273] (e)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中c和e各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中c = 0或1-20和e = 0或1-20,

[0274] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B、P2、E2和X2构成有义链,其中B、P2和X2分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B、P1、E1和X1构成有义链,其中B、P1和X1分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供

剪接信号,以控制由gDNA转录的前mRNA的加工,

[0275] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,f = 9。在一些实施方案中,f = 10。在一些实施方案中,f = 11。在一些实施方案中,f = 12。在一些实施方案中,f = 13。在一些实施方案中,f = 14。在一些实施方案中,f = 15。在一些实施方案中,f = 16。在一些实施方案中,f = 17。在一些实施方案中,f = 18。在一些实施方案中,f = 19。在一些实施方案中,f = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,c = 0。在一些实施方案中,c = 1。在一些实施方案中,c = 2。在一些实施方案中,c = 3。在一些实施方案中,c = 4。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,c = 8。在一些实施方案中,c = 9。在一些实施方案中,c = 10。在一些实施方案中,c = 11。在一些实施方案中,c = 12。在一些实施方案中,c = 13。在一些实施方案中,c = 14。在一些实施方案中,c = 15。在一些实施方案中,c = 16。在一些实施方案中,c = 17。在一些实施方案中,c = 18。在一些实施方案中,c = 19。在一些实施方案中,c = 20。在一些实施方案中,e = 0。在一些实施方案中,e = 1。在一些实施方案中,e = 2。在一些实施方案中,e = 3。在一些实施方案中,e = 4。在一些实施方案中,e = 5。在一些实施方案中,e = 6。在一些实施方案中,e = 7。在一些实施方案中,e = 8。在一些实施方案中,e = 9。在一些实施方案中,e = 10。在一些实施方案中,e = 11。在一些实施方案中,e = 12。在一些实施方案中,e = 13。在一些实施方案中,e = 14。在一些实施方案中,e = 15。在一些实施方案中,e = 16。在一些实施方案中,e = 17。在一些实施方案中,e = 18。在一些实施方案中,e = 19。在一些实施方案中,e = 20。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0276] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0277] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所

述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0278] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,和

[0279] 其中所述第一链和第二链各自包含下式:

[0280] 5'-$[E1]_a$-X1-$[P1]_b$-$[S1]_c$-$[B1]_d$-$[S2]_e$-$[B2]_f$-$[S3]_g$-$[P2]_h$-X2-$[E2]_i$-3',其中

[0281] (a)B1(如果存在的话)和B2各自为分支点序列,其中B2包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中构成B1(如果存在的话)的分支点序列相对于构成B2的分支点序列处于相反取向,并且在供体多核苷酸的相反链上,其中B1和B2各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中d和f是整数,其值分别指示B1和B2包含的核苷酸数目,其中d和f = 5-7;

[0282] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中b和h是整数,其值分别指示P1和P2包含的核苷酸数目,其中b=9-20和h=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;

[0283] (c)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和i是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在相反链上,其中构成E1和E2的核苷酸序列不互补;

[0284] (d)X1和X2各自包括包含3'剪接位点的核苷酸序列,其中X1的核苷酸序列相对于X2的核苷酸序列处于相反方向,且在供体多核苷酸的相反链上;和

[0285] (e)S1、S2和S3(如果任一种存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c、e和g各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中c = 0或1-20和g = 0或1-20,其中e = 0或1-40,

[0286] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B2、P2、E2和X2构成有义链,其中B2、P2和X2分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B1、P1、E1和X1构成有义链,其中B1、P1和X1分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0287] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在B1。在一些实施方案中,不存在B1。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一

些实施方案中,f = 5。在一些实施方案中,f = 6。在一些实施方案中,f = 7。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,h = 9。在一些实施方案中,h = 10。在一些实施方案中,h = 11。在一些实施方案中,h = 12。在一些实施方案中,h = 13。在一些实施方案中,h = 14。在一些实施方案中,h = 15。在一些实施方案中,h = 16。在一些实施方案中,h = 17。在一些实施方案中,h = 18。在一些实施方案中,h = 19。在一些实施方案中,h = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,存在S3。在一些实施方案中,不存在S3。在一些实施方案中,c = 0。在一些实施方案中,c = 1。在一些实施方案中,c = 2。在一些实施方案中,c = 3。在一些实施方案中,c = 4。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,c = 8。在一些实施方案中,c = 9。在一些实施方案中,c = 10。在一些实施方案中,c = 11。在一些实施方案中,c = 12。在一些实施方案中,c = 13。在一些实施方案中,c = 14。在一些实施方案中,c = 15。在一些实施方案中,c = 16。在一些实施方案中,c = 17。在一些实施方案中,c = 18。在一些实施方案中,c = 19。在一些实施方案中,c = 20。在一些实施方案中,g = 0。在一些实施方案中,g = 1。在一些实施方案中,g = 2。在一些实施方案中,g = 3。在一些实施方案中,g = 4。在一些实施方案中,g = 5。在一些实施方案中,g = 6。在一些实施方案中,g = 7。在一些实施方案中,g = 8。在一些实施方案中,g = 9。在一些实施方案中,g = 10。在一些实施方案中,g = 11。在一些实施方案中,g = 12。在一些实施方案中,g = 13。在一些实施方案中,g = 14。在一些实施方案中,g = 15。在一些实施方案中,g = 16。在一些实施方案中,g = 17。在一些实施方案中,g = 18。在一些实施方案中,g = 19。在一些实施方案中,g = 20。在一些实施方案中,e = 0。在一些实施方案中,e = 1。在一些实施方案中,e = 2。在一些实施方案中,e = 3。在一些实施方案中,e = 4。在一些实施方案中,e = 5。在一些实施方案中,e = 6。在一些实施方案中,e = 7。在一些实施方案中,e = 8。在一些实施方案中,e = 9。在一些实施方案中,e = 10。在一些实施方案中,e = 11。在一些实施方案中,e = 12。在一些实施方案中,e = 13。在一些实施方案中,e = 14。在一些实施方案中,e = 15。在一些实施方案中,e = 16。在一些实施方案中,e = 17。在一些实施方案中,e = 18。在一些实施方案中,e = 19。在一些实施方案中,e = 20。在一些实施方案中,e = 21。在一些实施方案中,e = 22。在一些实施方案中,e = 23。在一些实施方案中,e = 24。在一些实施方案中,e = 25。在一些实施方案中,e = 26。在一些实施方案中,e = 27。在一些实施方案中,e = 28。在一些实施方案中,e = 29。在一些实施方案中,e = 30。在一些实施方案中,e = 31。在一些实施方案中,e = 32。在一些实施方案中,e = 33。在一些实施方案中,e = 34。在一些实施方案中,e = 35。在一些

实施方案中,e = 36。在一些实施方案中,e = 37。在一些实施方案中,e = 38。在一些实施方案中,e = 39。在一些实施方案中,e = 40。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0288] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0289] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0290] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,和

[0291] 其中所述第一链和第二链各自包含下式:

[0292] 5'-$[E1]_a$-X1-$[P1]_b$-$[S1]_c$-$[B1]_d$-$[S2]_e$-$[B2]_f$-$[S3]_g$-$[P2]_h$-X2-$[E2]_i$-3',其中

[0293] (a)B1(如果存在的话)和B2各自为分支点序列,其中B2包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中构成B1(如果存在的话)的分支点序列相对于构成B2的分支点序列处于相反取向,并且在供体多核苷酸的相反链上,其中B1和B2各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中d和f是整数,其值分别指示B1和B2包含的核苷酸数目,其中d和f = 5-7;

[0294] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中b和h是整数,其值分别指示P1和P2包含的核苷酸数目,其中b=9-20和h=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;

[0295] (c)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和i是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在相反链上,其中构成E1和E2的核苷酸序列不互补;

[0296] (d)X1和X2各自包括包含3'剪接位点的核苷酸序列,其中X1的核苷酸序列相对于X2的核苷酸序列处于相反方向,且在供体多核苷酸的相反链上;和

[0297] (e)S1、S2和S3(如果任一种存在的话)各自为包含一个或多个核苷酸的定界符序

列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c、e和g各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中c = 0或1-20和g = 0或1-20,其中e = 0或1-40,

[0298] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B2、P2、E2和X2构成有义链,其中B2、P2和X2分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B1、P1、E1和X1构成有义链,其中B1、P1和X1分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0299] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在B1。在一些实施方案中,不存在B1。在一些实施方案中,d = 5。在一些实施方案中,d = 6。在一些实施方案中,d = 7。在一些实施方案中,f = 5。在一些实施方案中,f = 6。在一些实施方案中,f = 7。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,h = 9。在一些实施方案中,h = 10。在一些实施方案中,h = 11。在一些实施方案中,h = 12。在一些实施方案中,h = 13。在一些实施方案中,h = 14。在一些实施方案中,h = 15。在一些实施方案中,h = 16。在一些实施方案中,h = 17。在一些实施方案中,h = 18。在一些实施方案中,h = 19。在一些实施方案中,h = 20。在一些实施方案中,构成P1的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P1的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约90%-100%的嘧啶核碱基。在一些实施方案中,构成P2的核苷酸序列为约80%-90%的嘧啶核碱基。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,存在S2。在一些实施方案中,不存在S2。在一些实施方案中,存在S3。在一些实施方案中,不存在S3。在一些实施方案中,c = 0。在一些实施方案中,c = 1。在一些实施方案中,c = 2。在一些实施方案中,c = 3。在一些实施方案中,c = 4。在一些实施方案中,c = 5。在一些实施方案中,c = 6。在一些实施方案中,c = 7。在一些实施方案中,c = 8。在一些实施方案中,c = 9。在一些实施方案中,c = 10。在一些实施方案中,c = 11。在一些实施方案中,c = 12。在一些实施方案中,c = 13。在一些实施方案中,c = 14。在一些实施方案中,c = 15。在一些实施方案中,c = 16。在一些实施方案中,c = 17。在一些实施方案中,c = 18。在一些实施方案中,c = 19。在一些实施方案中,c = 20。在一些实施方案中,g = 0。在一些实施方案中,g = 1。在一些实施方案中,g = 2。在一些实施方案中,g = 3。在一些实施方案中,g = 4。在一些实施方案中,g = 5。在一些实施方案中,g = 6。在一些实施方案中,g = 7。在一些实施方案中,g = 8。在一些实施方案中,g = 9。在一些实施方案中,g = 10。在一些实施方案中,g = 11。在一

些实施方案中,g = 12。在一些实施方案中,g = 13。在一些实施方案中,g = 14。在一些实施方案中,g = 15。在一些实施方案中,g = 16。在一些实施方案中,g = 17。在一些实施方案中,g = 18。在一些实施方案中,g = 19。在一些实施方案中,g = 20。在一些实施方案中,e = 0。在一些实施方案中,e = 1。在一些实施方案中,e = 2。在一些实施方案中,e = 3。在一些实施方案中,e = 4。在一些实施方案中,e = 5。在一些实施方案中,e = 6。在一些实施方案中,e = 7。在一些实施方案中,e = 8。在一些实施方案中,e = 9。在一些实施方案中,e = 10。在一些实施方案中,e = 11。在一些实施方案中,e = 12。在一些实施方案中,e = 13。在一些实施方案中,e = 14。在一些实施方案中,e = 15。在一些实施方案中,e = 16。在一些实施方案中,e = 17。在一些实施方案中,e = 18。在一些实施方案中,e = 19。在一些实施方案中,e = 20。在一些实施方案中,e = 21。在一些实施方案中,e = 22。在一些实施方案中,e = 23。在一些实施方案中,e = 24。在一些实施方案中,e = 25。在一些实施方案中,e = 26。在一些实施方案中,e = 27。在一些实施方案中,e = 28。在一些实施方案中,e = 29。在一些实施方案中,e = 30。在一些实施方案中,e = 31。在一些实施方案中,e = 32。在一些实施方案中,e = 33。在一些实施方案中,e = 34。在一些实施方案中,e = 35。在一些实施方案中,e = 36。在一些实施方案中,e = 37。在一些实施方案中,e = 38。在一些实施方案中,e = 39。在一些实施方案中,e = 40。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0300] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0301] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0302] (ii)第二链,其从5’到3’包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3’剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,和

[0303] 其中所述第一链和第二链各自包含下式:

[0304] 5’-$[E1]_a$-Y1-$[S1]_b$-Y2-$[E2]_c$-3’,其中

[0305] (a)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(T),其中a和c是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在相反链上,其中构成E1和E2的核苷酸序列不互补;

[0306] (b)Y1和Y2各自包括包含5’剪接位点的核苷酸序列,其中Y1的核苷酸序列相对于Y2的核苷酸序列处于相反方向,且在供体多核苷酸的相反链上;和

[0307] (c)S1(如果存在的话)是包含一个或多个核苷酸的定界符序列,所述一个或多个

核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b是整数,其值指示定界符序列包含的核苷酸数目,其中b = 0或1-50,

[0308] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,E2和Y2构成有义链,并且E2包含剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,E1和Y1构成有义链,并且E1包含剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0309] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,b = 21。在一些实施方案中,b = 22。在一些实施方案中,b = 23。在一些实施方案中,b = 24。在一些实施方案中,b = 25。在一些实施方案中,b = 26。在一些实施方案中,b = 27。在一些实施方案中,b = 28。在一些实施方案中,b = 29。在一些实施方案中,b = 30。在一些实施方案中,b = 31。在一些实施方案中,b = 32。在一些实施方案中,b = 33。在一些实施方案中,b = 34。在一些实施方案中,b = 35。在一些实施方案中,b = 36。在一些实施方案中,b = 37。在一些实施方案中,b = 38。在一些实施方案中,b = 39。在一些实施方案中,b = 40。在一些实施方案中,b = 41。在一些实施方案中,b = 42。在一些实施方案中,b = 43。在一些实施方案中,b = 44。在一些实施方案中,b = 45。在一些实施方案中,b = 46。在一些实施方案中,b = 47。在一些实施方案中,b = 48。在一些实施方案中,b = 49。在一些实施方案中,b = 50。在一些实施方案中,供体多核苷酸为长度约10-400、约10-300或约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-500个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-400个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-300个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-200个核苷酸。在一些实施方案中,供体多核苷酸为长度约10-100个核苷酸。在一些实施方案中,供体多核苷酸为长度约20-80个核苷酸。在一些实施方案中,供体多核苷酸为长度约30-70个核苷酸。在一些实施方案中,供体多核苷酸为长度约40-60个核苷酸。

[0310] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0311] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的

第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0312] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,和

[0313] 其中所述第一链和第二链各自包含下式:

[0314] 5'-$[E1]_a$-Y1-$[S1]_b$-Y2-$[E2]_c$-3',其中

[0315] (a)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(T),其中a和c是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在相反链上,其中构成E1和E2的核苷酸序列不互补;

[0316] (b)Y1和Y2各自包括包含5'剪接位点的核苷酸序列,其中Y1的核苷酸序列相对于Y2的核苷酸序列处于相反方向,且在供体多核苷酸的相反链上;和

[0317] (c)S1(如果存在的话)是包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b是整数,其值指示定界符序列包含的核苷酸数目,其中b = 0或1-50,

[0318] 其中所述供体多核苷酸为长度约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,E2和Y2构成有义链,并且E2包含剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,E1和Y1构成有义链,并且E1包含剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0319] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,存在S1。在一些实施方案中,不存在S1。在一些实施方案中,b = 0。在一些实施方案中,b = 1。在一些实施方案中,b = 2。在一些实施方案中,b = 3。在一些实施方案中,b = 4。在一些实施方案中,b = 5。在一些实施方案中,b = 6。在一些实施方案中,b = 7。在一些实施方案中,b = 8。在一些实施方案中,b = 9。在一些实施方案中,b = 10。在一些实施方案中,b = 11。在一些实施方案中,b = 12。在一些实施方案中,b = 13。在一些实施方案中,b = 14。在一些实施方案中,b = 15。在一些实施方案中,b = 16。在一些实施方案中,b = 17。在一些实施方案中,b = 18。在一些实施方案中,b = 19。在一些实施方案中,b = 20。在一些实施方案中,b = 21。在一些实施方案中,b = 22。在一些实施方案中,b = 23。在一些实施方案中,b = 24。在一些实施方案中,b = 25。在一些实施方案中,b = 26。在一些实施方案中,b = 27。在一些实施方案中,b = 28。在一些实施方案中,b = 29。在一些实施方案中,b = 30。在一些实施方案中,b = 31。在一些实施方案中,b = 32。在一些实施方案中,b = 33。在一些实施方案中,b = 34。在一些实

施方案中,b = 35。在一些实施方案中,b = 36。在一些实施方案中,b = 37。在一些实施方案中,b = 38。在一些实施方案中,b = 39。在一些实施方案中,b = 40。在一些实施方案中,b = 41。在一些实施方案中,b = 42。在一些实施方案中,b = 43。在一些实施方案中,b = 44。在一些实施方案中,b = 45。在一些实施方案中,b = 46。在一些实施方案中,b = 47。在一些实施方案中,b = 48。在一些实施方案中,b = 49。在一些实施方案中,b = 50。

[0320] 在一些实施方案中,构成有义链的外显子序列校正了突变。在一些实施方案中,构成有义链的包含3'剪接位点的核苷酸序列校正了突变。在一些实施方案中,构成有义链的包含5'剪接位点的核苷酸序列校正了突变。在一些实施方案中,构成有义链的包含多聚嘧啶片的核苷酸序列校正了突变。在一些实施方案中,构成有义链的包含分支点序列的核苷酸序列校正了突变。

[0321] 在一些实施方案中,供体多核苷酸的最5'核苷酸包含5'磷酸基。在一些实施方案中,供体多核苷酸的长度为约40-70个核苷酸或约50-60个核苷酸。在一些实施方案中,供体多核苷酸的长度为约50-60个核苷酸。在一些实施方案中,供体多核苷酸的长度为40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷酸。在一些实施方案中,供体多核苷酸的长度为40个核苷酸。在一些实施方案中,供体多核苷酸的长度为41个核苷酸。在一些实施方案中,供体多核苷酸的长度为42个核苷酸。在一些实施方案中,供体多核苷酸的长度为43个核苷酸。在一些实施方案中,供体多核苷酸的长度为44个核苷酸。在一些实施方案中,供体多核苷酸的长度为45个核苷酸。在一些实施方案中,供体多核苷酸的长度为46个核苷酸。在一些实施方案中,供体多核苷酸的长度为47个核苷酸。在一些实施方案中,供体多核苷酸的长度为49个核苷酸。在一些实施方案中,供体多核苷酸的长度为42个核苷酸。在一些实施方案中,供体多核苷酸的长度为50个核苷酸。在一些实施方案中,供体多核苷酸的长度为51个核苷酸。在一些实施方案中,供体多核苷酸的长度为52个核苷酸。在一些实施方案中,供体多核苷酸的长度为53个核苷酸。在一些实施方案中,供体多核苷酸的长度为54个核苷酸。在一些实施方案中,供体多核苷酸的长度为55个核苷酸。在一些实施方案中,供体多核苷酸的长度为56个核苷酸。在一些实施方案中,供体多核苷酸的长度为57个核苷酸。在一些实施方案中,供体多核苷酸的长度为58个核苷酸。在一些实施方案中,供体多核苷酸的长度为59个核苷酸。在一些实施方案中,供体多核苷酸的长度为60个核苷酸。

[0322] 在一些实施方案中,供体多核苷酸包含天然核苷酸。在一些实施方案中,供体多核苷酸包含天然存在的核苷酸。在一些实施方案中,供体多核苷酸包含一个或多个非天然和/或修饰的核苷酸。在一些实施方案中,一个或多个非天然和/或修饰的核苷酸是2'-0-甲基核苷酸。在一些实施方案中,供体多核苷酸包含一种或多种主链修饰。在一些实施方案中,一种或多种主链修饰是硫代磷酸酯。在一些实施方案中,供体多核苷酸包含两个平端。在一些实施方案中,供体多核苷酸包含一个平端,并且包括包含突出物(例如5'或3'突出物)的一端。在一些实施方案中,供体多核苷酸的核苷酸序列包含一个或多个核苷酸,其阻止定点核酸酶识别且切割供体多核苷酸。

[0323] 在一些实施方案中,疾病是糖原贮积病1a(GSD1a)。在一些实施方案中,突变位于人染色体17q21上的人G6PC基因中。在一些实施方案中,G6PC基因中的突变导致人G6PC蛋白中的R83C、R83H或E110K氨基酸取代。在一些实施方案中,G6PC基因中的突变导致人G6PC蛋白中的R83C氨基酸取代。在一些实施方案中,G6PC基因中的突变导致人G6PC蛋白中的R83H

氨基酸取代。在一些实施方案中,G6PC基因中的突变导致人G6PC蛋白中的E110K氨基酸取代。

[0324] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含校正细胞中的基因组DNA分子(gDNA)中的突变的核苷酸序列,所述突变引起糖原贮积病1a,其中所述突变位于人染色体17q21上的人G6PC基因中,并且导致氨基酸取代R83C或R83H,所述供体多核苷酸包含:

[0325] (i)第一链,其从5'到3'包括包含校正突变的外显子序列的核苷酸序列、以及包含一种或多种剪接信号以控制由基因组DNA分子转录的前体mRNA(前mRNA)加工的核苷酸序列,其中所述外显子序列包含与G6PC基因中的位置83处的密码子相对应的编码精氨酸(R)的密码子,其中所述一种或多种剪接信号是3'剪接位点、多聚嘧啶片和分支点序列的组合;和

[0326] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0327] 其中所述供体多核苷酸为长度约40-70个核苷酸,并包含两个平端,其中所述供体多核苷酸的每条链的最5'核苷酸包含5'磷酸部分,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,其中所述供体多核苷酸的插入在包含校正突变的外显子序列的gDNA中形成外显子,其中所述剪接信号指导外显子包括到mRNA内,从而校正了突变。在一些实施方案中,分支点序列包含核苷酸序列TTCAT,其中多聚嘧啶片包含核苷酸序列CTTGTTCTGTTTTTTT,其中3'剪接位点包含核苷酸序列TAG,并且其中外显子序列包含核苷酸序列GATTCTCTTTGGACAGCGCCCTTACT。

[0328] 在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:30(CH34 54-0)中阐述。在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:20(CH32 50-0)中阐述。

[0329] 在一些实施方案中,疾病是庞贝氏病。在一些实施方案中,突变位于人染色体17q25.3上的人葡萄糖苷酶α(GAA)基因中。在一些实施方案中,突变在GAA的剪接信号中,其导致缺乏外显子2和/或一个或多个隐蔽剪接位点的激活的GAA基因的mRNA转录物。

[0330] 在一些实施方案中,本公开内容提供了包含非复制性dsDNA分子的供体多核苷酸,所述dsDNA分子包含校正细胞中的gDNA分子中的突变的核苷酸序列,所述突变引起庞贝氏病,其中所述突变在GAA的剪接信号中,其导致缺乏外显子2和/或一个或多个隐蔽剪接位点的激活的GAA基因的mRNA转录物,所述供体多核苷酸包含:

[0331] (i)第一链,其从5'到3'包含校正突变的第一核苷酸序列,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号包含3'剪接位点、多聚嘧啶片和分支点序列的组合;和

[0332] (ii)第二链,其从5'到3'包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述一种或多种剪接信号包含3'剪接位点、多聚嘧啶片和分支点序列的组合,其中所述第二链与第一链互补,

[0333] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-

100、约20-80、约30-70或约40-60个核苷酸,并包含两个平端,其中所述供体多核苷酸的每条链的最5'核苷酸包含5'磷酸部分,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,NHEJ DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,其中所述供体多核苷酸被配置用于双向插入DSB断裂内,其中以任一方向的插入形成3'剪接位点、多聚嘧啶片和校正突变的分支点序列,其中所述剪接信号指导外显子2包括到mRNA内,从而校正了突变。

[0334] 在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:63(GAA_50-0)中阐述。

[0335] 在一些实施方案中,本公开内容提供了校正细胞中的基因组DNA分子(gDNA)中的突变的系统,该系统包含以下组分:

[0336] (a)前述供体多核苷酸中的任何一种;

[0337] (b)一种或多种gRNA分子;和

[0338] (c)定点核酸酶,

[0339] 其中当系统引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,定点核酸酶由mRNA编码。在一些实施方案中,定点核酸酶是多肽。在一些实施方案中,一种或多种gRNA分子和定点核酸酶包含核糖核蛋白。在一些实施方案中,定点核酸酶是Cas核酸酶。在一些实施方案中,Cas核酸酶是化脓性链球菌(*S. pyogenes*)Cas9(SpCas9),或者其同源物、衍生物或修饰形式。在一些实施方案中,Cas核酸酶是金黄色葡萄球菌(*S. aureus*)Cas9(SaCas9),或者其同源物、衍生物或修饰形式。

[0340] 在一些实施方案中,一种或多种gRNA分子包含选自以下的修饰:主链修饰、糖修饰、修饰的核苷间键合、或者修饰的或非天然的核碱基。在一些实施方案中,一种或多种gRNA分子包含主链修饰。在一些实施方案中,主链修饰是硫代磷酸酯修饰。

[0341] 在由本公开内容提供的系统的一些实施方案中,供体多核苷酸、gRNA分子和定点核酸酶在脂质体或脂质纳米颗粒中个别地配制或共配制。在一些实施方案中,供体多核苷酸、gRNA分子和定点核酸酶在脂质体或脂质纳米颗粒中个别地配制。在一些实施方案中,供体多核苷酸、gRNA分子和定点核酸酶在脂质体或脂质纳米颗粒中共配制。在一些实施方案中,gRNA的核苷酸序列包含SEQ ID NO:107(CH32突变体-CTX1 sgRNA间隔区)中阐述的序列。在一些实施方案中,gRNA的核苷酸序列包含SEQ ID NO:108(CH34突变体-CTX1 sgRNA间隔区)中阐述的序列。在一些实施方案中,gRNA的核苷酸序列包含SEQ ID NO:127(突变体GAA sgRNA间隔区)中阐述的序列。

[0342] 在一些实施方案中,本公开内容提供了细胞,其包含前述供体多核苷酸或系统中的任何一种。在一些实施方案中,细胞是分裂细胞或非分裂细胞。在一些实施方案中,细胞是分裂细胞。在一些实施方案中,细胞是非分裂细胞。在一些实施方案中,细胞是患者特异性诱导多能干细胞(iPSC)。在一些实施方案中,细胞是原代肝细胞。

[0343] 在一些方面,本公开内容提供了药物组合物,其包含前述供体多核苷酸中的任何一种和药学上可接受的载体。在一些实施方案中,本公开内容提供了药物组合物,其包含前述系统中的任何一种和药学上可接受的载体。在一些实施方案中,本公开内容提供了药物组合物,其包含前述细胞中的任何一种和药学上可接受的载体。

[0344] 在一些实施方案中,本公开内容提供了前述供体多核苷酸、系统或药物组合物中的任何一种,其用于通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者,所述治疗包括:从患者中分离细胞,使细胞与供体多核苷酸、系统或药物组合物接触,其中当供体多核苷酸、系统或药物组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0345] 在一些实施方案中,本公开内容提供了前述供体多核苷酸、系统或药物组合物中的任何一种,其用于通过校正细胞中的基因组DNA分子(gDNA)中的突变来治疗患有疾病的患者,所述治疗包括:向患者施用有效量的供体多核苷酸、系统或药物组合物,其中当施用供体多核苷酸、系统或组合物时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0346] 在一些实施方案中,本公开内容提供了校正细胞中的基因组DNA分子(gDNA)中的突变的方法,所述方法包括:使细胞与前述供体多核苷酸、系统或药物组合物中的任何一种接触,其中当供体多核苷酸、系统或组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0347] 在一些实施方案中,本公开内容提供了通过校正细胞中的基因组DNA分子(gDNA)中的突变,来治疗患有疾病的患者的方法,所述方法包括:从患者中分离细胞,使细胞与前述供体多核苷酸、系统或药物组合物中的任何一种接触,其中当供体多核苷酸、系统或组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0348] 在一些实施方案中,本公开内容提供了通过校正细胞中的基因组DNA分子(gDNA)中的突变,来治疗患有疾病的患者的方法,所述方法包括:向患者施用有效量的前述供体多核苷酸、系统或药物组合物中的任何一种,其中当施用供体多核苷酸、系统或组合物时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。(体内)。

[0349] 在由本公开内容提供的方法的一些实施方案中,细胞是患者特异性诱导多能干细胞(iPSC)。在一些实施方案中,细胞是肝细胞。在由本公开内容提供的方法的一些实施方案中,该方法进一步包括使包含校正的突变的iPSC分化成分化细胞;并且将分化细胞植入患者内。

[0350] 在由本公开内容提供的方法的一些实施方案中,治疗导致由基因组DNA分子(gDNA)转录的mRNA的翻译,所述gDNA包含插入的供体多核苷酸,其中所述翻译导致翻译产物(蛋白质)的形成,所述翻译产物减轻疾病或者并不引起或促进疾病。

[0351] 在一些实施方案中,本公开内容提供了包括容器和包含使用说明书的包装插页的试剂盒,所述容器包含前述供体多核苷酸、系统或药物组合物中的任何一种,其用于校正细胞中的基因组DNA分子(gDNA)中的突变。

附图说明

[0352] 图1提供了示意图,其显示了靶核酸中的内源基因座、以及在供体多核苷酸插入后的基因座。指示了剪接信号、剪接位点、内含子和外显子的相对位置。

[0353] 图2A-2C提供了描绘了总测序读数的百分比的点图,所述总测序读数对应于由供体多核苷酸编码的特异性核苷酸变化(例如,校正编辑),指示了NHEJ介导的供体多核苷酸插入靶基因(例如,G6PC外显子2)内,如通过下一代测序(NGS)确定的。图2A-2C描绘了由三种不同的gRNA诱导的*G6PC*外显子2基因组DNA(gDNA)基因座中的DSB的校正编辑的百分比。

[0354] 图3提供了显示校正编辑的百分比的线图,所述校正编辑起因于NHEJ介导的供体多核苷酸在CH34 sgRNA靶位点处插入*G6PC*外显子2 gDNA基因座(DNA)内。含有另外的剪接信号的供体多核苷酸的插入和ORF中的核苷酸变化导致*G6PC* DNA中的校正编辑(黑线),并且还导致*G6PC*基因的转录,其产生含有所需预测编辑(灰色线)的mRNA。校正编辑的水平也显示为所使用的供体多核苷酸的量(供体的ng)的函数。

[0355] 图4A提供了显示对应于校正编辑的总测序读数的百分比的点图,所述校正编辑起因于NHEJ介导的供体多核苷酸插入小鼠肝细胞的*G6PC*外显子2 gDNA基因座内,所述肝细胞用包含供体多核苷酸、编码Cas9的mRNA和靶向*G6PC*的gRNA的脂质纳米颗粒(LNP)进行处理。图4B提供了点图,其显示了在包含Cas9 mRNA、供体多核苷酸和靶向*G6PC*的gRNA的LNP的体内施用之后,起因于完全NHEJ介导的供体多核苷酸插入的gDNA和转录的mRNA中的校正编辑的百分比。

[0356] 图5提供了描述双向供体多核苷酸的设计的示意图。供体被设计为当在正向方向或反向方向上插入DNA DSB内时掺入剪接序列。

[0357] 图6A-6B提供了描述用于插入*GAA*基因内的双向供体多核苷酸的设计的示意图。图6A描绘了dsODN供体多核苷酸的有义链和反义链中的剪接信号。图6B描绘了所需剪接信号的插入,而不管dsODN供体是在正向方向上还是在反向方向上插入*GAA*切割位点内。

[0358] 图7A和图7C提供了条形图,其显示了用单向供体多核苷酸处理的细胞中的校正编辑的百分比,当在正向方向上插入*G6PC*(图7A)或*GAA*基因(图7C)中诱导的DSB内时,所述单向供体多核苷酸编码校正编辑。图7B和图7D提供了条形图,其显示了用双向供体多核苷酸处理的细胞中的校正编辑的百分比,当在正向方向或反向方向上插入*G6PC*(图7B)或*GAA*基因(图7D)中诱导的DSB内时,所述双向供体多核苷酸编码校正编辑。

[0359] 图8提供了显示校正编辑的百分比的条形图,所述校正编辑起因于50nt双向供体插入原代人肝细胞中的*GAA*基因中诱导的DSB内。描绘的是起因于供体的完全正向插入、完全反向插入、以及组合的总完全插入的总序列读数的百分比。

[0360] 图9提供了显示校正编辑的百分比的线图,所述校正编辑起因于不同长度的双向供体插入Huh-7细胞中的*GAA*基因中诱导的DSB内。显示的是起因于完全插入、以及在正向方向(例如,完全插入)或反向方向(例如,完全反向插入)上的插入的校正编辑的百分比。

[0361] 图10A-10C提供了显示校正编辑的百分比的条形图,所述校正编辑起因于双向供体插入成纤维细胞中的*GAA*基因中诱导的DSB内,所述成纤维细胞衍生自患有庞贝病的患者。图10A中显示的是如通过NGS确定的,具有由供体多核苷酸编码的校正编辑的gDNA的序列读数的百分比。图10B中显示的是如通过定量PCR(qPCR)确定的,针对缺乏外显子2的mRNA转录物数量标准化的编码外显子2的mRNA转录物的定量。图10C中显示的是如通过NGS确定的,与模拟转染的细胞相比,在用Cas9/gRNA连同或不连同供体多核苷酸一起转染的细胞中,编码外显子2的序列读数的数目中的百分比增加。

具体实施方式

[0362] 用于修复DNA双链断裂(DSB)的主导途径是非同源末端连接(NHEJ)途径、以及同源性指导的修复(HDR)途径(例如,也称为同源重组或HR)。通过使用定点核酸酶(例如,CRISPR/Cas9核酸内切酶)在靶基因中诱导位点特异性DSB,可以使用同源供体多核苷酸作为模板通过HDR修复引入基因编辑。然而,HDR修复是低效率的,特别是在非分裂细胞中。因此,使用NHEJ修复用于诱导由供体多核苷酸提供的所需基因编辑的方法,对于编辑非分裂细胞和分裂细胞两者均为有利的。本公开内容至少部分地基于供体多核苷酸的设计,所述供体多核苷酸使用NHEJ修复途径,插入靶基因中通过定点核酸酶(例如,CRISPR/Cas9核酸内切酶)诱导的DSB处。供体多核苷酸被设计用于NHEJ介导的靶基因DSB中的插入,以校正位于外显子(例如,蛋白质编码突变)和内含子(例如,剪接信号突变)中的致病突变,所述致病突变导致异常的蛋白质生产或功能。因此,本文提供的是包含供体多核苷酸的方法和组合物,所述供体多核苷酸提供了基因组DNA(gDNA)的核苷酸序列中的所需改变和/或调节外显子限定,从而导致所需改变包括在由编辑的gDNA转录的RNA转录物(例如,前mRNA)中。

[0363] 基因组编辑

[0364] 基因组编辑一般指优选以精确、期望和/或预定的方式,编辑或改变基因组的核苷酸序列的过程。本文所述的基因组编辑的组合物、系统和方法的实例使用定点核酸酶,以在基因组中的精确靶位置处切开或切割DNA,从而在DNA中产生双链断裂(DSB)。此类断裂可以通过内源性DNA修复途径进行修复,所述修复途径例如同源性指导的修复(HDR)和/或非同源末端连接(NHEJ)修复(参见例如,Cox等人(2015)Nature Medicine 21(2):121-31)。在非分裂细胞中的有效基因组编辑的主要障碍之一是缺乏同源性指导的修复(HDR)。如果没有HDR,非分裂细胞依靠非同源末端连接(NHEJ),以修复基因组中出现的双链断裂(DSB)。NHEJ介导的DSB的DNA修复的结果可以包括DSB的校正修复、或者一个或多个核苷酸或多核苷酸的缺失或插入。

[0365] 供体多核苷酸

[0366] 在人中,外显子限定由跨越外显子而不是跨越内含子配对的剪接位点决定。其它剪接信号,包括但不限于分支点序列、多聚嘧啶片、以及外显子和内含子剪接增强子和沉默子,也促成外显子适当剪接在一起,以形成成熟mRNA。在前mRNA剪接过程中,剪接机制以外显子极性搜索一对紧密间隔的剪接位点(即上游3'剪接位点和下游5'剪接位点)(Berget(1995)J Biol Chem 270:2411-2414)。

[0367] 相应地,本公开内容提供了供体多核苷酸,其在插入DSB内之后,校正或诱导靶核酸(例如,基因组DNA)中的突变,并且包含剪接信号(例如,一种或多种剪接信号),通过外显子限定的调节,所述剪接信号允许剪接机制部分地将校正的突变(或诱导突变)掺入转录产物(例如前mRNA)内。在一些实施方案中,由本公开内容提供的供体多核苷酸被细胞的NHEJ机制识别,并且用于修复通过定点核酸酶引入靶核酸内的双链断裂(DSB),其中DSB的修复导致供体多核苷酸插入靶核酸内。在一些实施方案中,供体多核苷酸被配置用于单向插入DSB内。在一些实施方案中,供体多核苷酸被配置用于双向插入DSB内。

[0368] *单向供体多核苷酸*

[0369] 在一些实施方案中,由本公开内容提供的供体多核苷酸是线性、非复制性的双链DNA分子(dsDNA),其包含校正或诱导细胞中的基因组DNA(gDNA)分子中的突变的核苷酸序

列、以及包含一种或多种剪接信号以控制由gDNA转录的前体mRNA(前mRNA)加工的核苷酸序列,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。在一些实施方案中,供体多核苷酸包含第一DNA链和第二DNA链,其中所述第二DNA链包含与第一DNA链互补的核苷酸序列。

[0370] 在一些实施方案中,供体多核苷酸包含校正或诱导突变的核苷酸序列,其中校正或诱导突变的核苷酸序列包含单个核苷酸。在一些实施方案中,校正或诱导突变的核苷酸序列包含两个或更多个核苷酸。在一些实施方案中,校正或诱导突变的核苷酸序列包含密码子。在一些实施方案中,校正或诱导突变的核苷酸序列包含一个或多个密码子。在一些实施方案中,校正或诱导突变的核苷酸序列包含外显子序列。

[0371] 在一些实施方案中,供体多核苷酸包含校正或诱导突变的核苷酸序列,其中校正或诱导突变的核苷酸序列包含内含子序列。在一些实施方案中,校正或诱导突变的核苷酸序列包含剪接信号。

[0372] 在一些实施方案中,供体多核苷酸序列与靶核酸的内源序列等同或基本上等同(具有至少一个核苷酸差异)。在一些实施方案中,内源序列包含细胞的基因组序列。在一些实施方案中,内源序列包含染色体或染色体外序列。在一些实施方案中,供体多核苷酸序列包含的序列与细胞中的DSB处或附近的内源序列的一部分基本上等同(包含至少一个核苷酸差异/变化)。在一些实施方案中,用供体多核苷酸修复靶核酸分子导致靶核酸分子的一个或多个核苷酸的插入、缺失或取代。在一些实施方案中,一个或多个核苷酸的插入、缺失或取代导致由包含靶序列的基因表达的蛋白质中的一个或多个氨基酸变化。在一些实施方案中,一个或多个核苷酸的插入、缺失或取代导致由靶基因表达的RNA中的一个或多个核苷酸变化。在一些实施方案中,一个或多个核苷酸的插入、缺失或取代改变靶基因的表达水平。在一些实施方案中,一个或多个核苷酸的插入、缺失或取代导致靶基因的表达增加或减少。在一些实施方案中,一个或多个核苷酸的插入、缺失或取代导致基因敲减。在一些实施方案中,一个或多个核苷酸的插入、缺失或取代导致基因敲除。在一些实施方案中,用供体多核苷酸修复靶核酸分子导致靶基因的外显子序列、内含子序列、转录控制序列、翻译控制序列、包含剪接信号的序列、或非编码序列的替换。

[0373] 供体多核苷酸具有合适的长度,以校正或诱导gDNA中的突变。在一些实施方案中,供体多核苷酸包含长度为10、15、20、25、50、75、100个或更多个核苷酸。在一些实施方案中(例如本文所述的那些,其中供体多核苷酸作为通过非同源末端连接介导的插入而掺入切割的核酸内),供体多核苷酸不具有同源臂。在一些实施方案中,供体多核苷酸的长度为约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,供体多核苷酸的长度为约10-500个核苷酸。在一些实施方案中,供体多核苷酸的长度为约10-400个核苷酸。在一些实施方案中,供体多核苷酸的长度为约10-300个核苷酸。在一些实施方案中,供体多核苷酸的长度为约10-200个核苷酸。在一些实施方案中,供体多核苷酸的长度为约10-100个核苷酸。在一些实施方案中,供体多核苷酸的长度为约20-80个核苷酸。在一些实施方案中,供体多核苷酸的长度为约30-70个核苷酸。在一些实施方案中,供体多核苷酸的长度为约40-60个核苷酸。在一些实施方案中,供体多核苷酸的长度为40、41、42、43、44、45、46、46、48、49、50、51、52、53、54、55、56、57、58、59或60个核苷

酸。在一些实施方案中,供体多核苷酸的长度为40个核苷酸。在一些实施方案中,供体多核苷酸的长度为41个核苷酸。在一些实施方案中,供体多核苷酸的长度为42个核苷酸。在一些实施方案中,供体多核苷酸的长度为43个核苷酸。在一些实施方案中,供体多核苷酸的长度为44个核苷酸。在一些实施方案中,供体多核苷酸的长度为45个核苷酸。在一些实施方案中,供体多核苷酸的长度为46个核苷酸。在一些实施方案中,供体多核苷酸的长度为47个核苷酸。在一些实施方案中,供体多核苷酸的长度为48个核苷酸。在一些实施方案中,供体多核苷酸的长度为49个核苷酸。在一些实施方案中,供体多核苷酸的长度为50个核苷酸。在一些实施方案中,供体多核苷酸的长度为51个核苷酸。在一些实施方案中,供体多核苷酸的长度为52个核苷酸。在一些实施方案中,供体多核苷酸的长度为53个核苷酸。在一些实施方案中,供体多核苷酸的长度为54个核苷酸。在一些实施方案中,供体多核苷酸的长度为55个核苷酸。在一些实施方案中,供体多核苷酸的长度为56个核苷酸。在一些实施方案中,供体多核苷酸的长度为57个核苷酸。在一些实施方案中,供体多核苷酸的长度为58个核苷酸。在一些实施方案中,供体多核苷酸的长度为59个核苷酸。在一些实施方案中,供体多核苷酸的长度为60个核苷酸。

[0374] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含内含子序列。在一些实施方案中,供体多核苷酸包含校正或诱导gDNA中的突变的内含子序列。在一些实施方案中,供体多核苷酸包含外显子序列。在一些实施方案中,供体多核苷酸包含校正或诱导gDNA中的突变的外显子序列。

[0375] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含一种或多种剪接信号,以控制由供体多核苷酸插入其内的gDNA转录的前体mRNA(前mRNA)的加工(例如剪接)。

[0376] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工(例如剪接),在一些实施方案中,其中所述一种或多种剪接信号选自:

[0377] (a)天然或增强的3'剪接位点;

[0378] (b)天然或增强的5'剪接位点;

[0379] (c)多聚嘧啶片;

[0380] (d)分支点;

[0381] (e)外显子剪接增强子(ESE);

[0382] (f)内含子剪接增强子(ISE);

[0383] (g)外显子剪接沉默子(ESS);

[0384] (h)内含子剪接沉默子(ISS);和

[0385] (i)(a)-(h)中任一种的组合。

[0386] 在一些实施方案中,供体多核苷酸包含天然或增强的3'剪接位点。在一些实施方案中,供体多核苷酸包含天然存在的3'剪接位点。在一些实施方案中,供体多核苷酸包含增强的3'剪接位点。天然的3'剪接位点是天然存在的3'剪接位点。在一些实施方案中,天然的3'剪接位点的核苷酸序列与供体多核苷酸插入位点处或附近的3'剪接位点的核苷酸序列等同或基本上等同(具有至少一个核苷酸差异)。增强的3'剪接位点是包含与3'剪接位点的共有序列等同或基本上等同(具有至少一个核苷酸差异)的核苷酸序列的剪接位点。3'剪接位点的共有序列是核苷酸序列YAG,并且其中Y是包含选自胸腺嘧啶(T)和胞嘧啶(C)的核碱

基的核苷酸。参见例如,Reed(1989)Genes Dev 3(12B):2113-2123,其进一步描述了哺乳动物中的3'剪接位点序列的组构。

[0387] 在一些实施方案中,供体多核苷酸包含天然或增强的5'剪接位点。在一些实施方案中,供体多核苷酸包含天然的5'剪接位点。在一些实施方案中,供体多核苷酸包含增强的5'剪接位点。在一些实施方案中,天然的5'剪接位点是天然存在的5'剪接位点。在一些实施方案中,天然的5'剪接位点的核苷酸序列与供体多核苷酸插入位点处或附近的5'剪接位点的核苷酸序列等同或基本上等同(具有至少一个核苷酸差异)。在一些实施方案中,增强的5'剪接位点是包含与5'剪接位点的共有序列等同或基本上等同(具有至少一个核苷酸差异)的核苷酸序列的剪接位点。5'剪接位点的共有序列是核苷酸序列GTRAG,并且其中R是包含选自腺嘌呤(A)和鸟嘌呤(G)的核碱基的核苷酸。

[0388] 在一些实施方案中,供体多核苷酸包含多聚嘧啶片。多聚嘧啶片是顺式作用序列元件,其指导在前mRNA剪接中的内含子去除。已发现多聚嘧啶片的逐渐缺失取消正确的套索(lariate)形成、剪接体组装和剪接。研究已显示,含有连续尿苷的嘧啶束是强嘧啶束(Coolidge等人(1997)Nucleic Acids Res 25(4):888-896)。在一些实施方案中,多聚嘧啶片的长度为5、6、7、8、9、10、11、12、13、14、15、16、17或18个核苷酸。在一些实施方案中,多聚嘧啶片的长度为16个核苷酸。在一些实施方案中,多聚嘧啶片的长度为13个核苷酸。在一些实施方案中,多聚嘧啶片包含核苷酸序列TTTTTTTCTTTTT(SEQ ID NO:54)。在一些实施方案中,多聚嘧啶片的长度为9个核苷酸。在一些实施方案中,多聚嘧啶片包含核苷酸序列TTTTTTTCT(SEQ ID NO:53)。

[0389] 在一些实施方案中,供体多核苷酸包含分支点。在一些实施方案中,供体多核苷酸包括包含分支点的核苷酸序列。在高等真核生物中,前mRNA剪接由简并剪接顺式元件介导,所述元件包含分支点序列(BPS)、多聚嘧啶片(PPT)、5'和3'剪接位点以及外显子/内含子剪接增强子/沉默子,并且前mRNA在由剪接体介导的两个序贯酯交换反应中进行剪接。在一些实施方案中,供体多核苷酸包含分支点,其包含与分支点共有序列等同或基本上等同(具有至少一个核苷酸差异)的核苷酸序列。在一些实施方案中,分支点共有序列是YTNAY,其中Y = C或T,并且其中N = A、G、C或T(SEQ ID NO:49。在一些实施方案中,分支点序列是TATTAAC(SEQ ID NO:50)。在一些实施方案中,分支点序列是GTTAATA(SEQ ID NO:51)。在一些实施方案中,分支点序列是TACTGAC(SEQ ID NO:52)。

[0390] 在一些实施方案中,供体多核苷酸包含外显子剪接增强子(ESE)(参见例如,Blencowe(2000)Trends Biochem Sci 25(3):106-110)。在一些实施方案中,供体多核苷酸包含外显子剪接沉默子(ESS)(参见例如,Wang等人(2006)Mol Cell 23(1):61-70)。在一些实施方案中,供体多核苷酸包含内含子剪接增强子(ISE)(参见例如,Wang等人(2012)Nat Struct Mol Biol 19(10):1044-1052)。在一些实施方案中,供体多核苷酸包含内含子剪接沉默子(ISS)(参见例如,Carstens等人(2000)Mol Cell Biol 20(19):7388-7400)。

[0391] 在一些实施方案中,供体多核苷酸包含天然或增强的3'剪接位点和多聚嘧啶片。在一些实施方案中,供体多核苷酸包含天然或增强的3'剪接位点、多聚嘧啶片和分支点。

[0392] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0393] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的

核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0394] (ii)第二链,其包含与第一链互补的核苷酸序列,其中所述一种或多种剪接信号选自:

[0395] (a)天然或增强的3'剪接位点;

[0396] (b)多聚嘧啶片;

[0397] (c)分支点;

[0398] (d)外显子剪接增强子(ESE);

[0399] (e)内含子剪接增强子(ISE);

[0400] (f)外显子剪接沉默子(ESS);

[0401] (g)内含子剪接沉默子(ISS);和

[0402] (h)(a)-(g)中任一种的组合,

[0403] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0404] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0405] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA(gDNA)分子中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0406] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0407] 其中所述一种或多种剪接信号选自:

[0408] (a)天然或增强的5'剪接位点;

[0409] (b)外显子剪接增强子(ESE);

[0410] (c)内含子剪接增强子(ISE);

[0411] (d)外显子剪接沉默子(ESS);

[0412] (e)内含子剪接沉默子(ISS);和

[0413] (f)(a)-(e)中任一种的组合,

[0414] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0415] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0416] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包

含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中至少一种剪接信号是天然或增强的3'剪接位点;和

[0417] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0418] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0419] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0420] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号是天然或增强的3'剪接位点和多聚嘧啶片的组合;和

[0421] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0422] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0423] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0424] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号是天然或增强的3'剪接位点、多聚嘧啶片和分支点的组合;和

[0425] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0426] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0427] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0428] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变在3'剪接位点中,其中所述第一链包含内含子序列,任选地外显子序列,其中所述内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0429] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0430] 其中所述一种或多种剪接信号选自:

[0431] (a)多聚嘧啶片;

[0432] (b)分支点;

[0433] (c)外显子剪接增强子(ESE);

[0434] (d)内含子剪接增强子(ISE);

[0435] (e)外显子剪接沉默子(ESS);

[0436] (f)内含子剪接沉默子(ISS);和

[0437] (g)(a)-(f)中任一种的组合,

[0438] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0439] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0440] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述供体多核苷酸包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中至少一种剪接信号是天然或增强的5'剪接位点;和

[0441] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0442] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0443] 在一些实施方案中,供体多核苷酸插入DSB内导致在包含外显子序列的gDNA中的外显子形成。在一些实施方案中,一种或多种剪接信号指导包含外显子序列的外显子包括到mRNA内。

[0444] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0445] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变在5'剪接位点中,其中所述第一链包含内含子序列,任选地外显子序列,其中所述内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0446] (ii)第二链,其包含与第一链互补的核苷酸序列,其中所述一种或多种剪接信号选自:

[0447] (a)外显子剪接增强子(ESE);

[0448] (b)内含子剪接增强子(ISE);

[0449] (c)外显子剪接沉默子(ESS);

[0450] (d)内含子剪接沉默子(ISS);和

[0451] (e)(a)-(d)中任一种的组合,

[0452] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0453] 在一些实施方案中,供体多核苷酸的插入导致包含内含子序列的内含子形成,其中所述内含子序列校正了突变。

[0454] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含第一链和第二链,其中所述第二链与第一链互补,其中所述第一链从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述第一链包含下式:

[0455] 5'-[B]a-[S1]b-[P]c-[S2]d-X-E-3',其中

[0456] (i)B(如果存在的话)是分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a是整数,其值指示B包含的核苷酸数目,其中a=0或5-7;

[0457] (ii)P是多聚嘧啶片,其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c是整数,其值指示P包含的核苷酸数目,其中c=9-20,其中构成P的核苷酸序列为约100%、约90%-100%、约80%-90%的嘧啶核碱基;

[0458] (iii)E是包含校正突变的核苷酸序列的外显子序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C);

[0459] (iv)X是包含3'剪接位点的核苷酸序列;和

[0460] (v)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和d各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和d = 0-20,

[0461] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于定向插入双链DNA断裂(DSB)内,其中当供体多核苷酸插入DSB内时,B(如果存在的话)、P、X(如果存在的话)和E(如果存在的话)构成有义链,其中B(如果存在的话)、P和X(如果存在的话)包含一种或多种剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0462] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0463] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含第一链和第二链,其中所述第二链与第一链互补,其中所述第一链从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的核苷酸序列,其中所

述致病突变是接近5'剪接位点的蛋白质编码突变,其中所述第一链包含内含子序列和外显子序列,其中所述外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,

[0464] 其中所述第一链包含下式:

[0465] 5'-E-Y-I-3',其中

[0466] (i)E是包含校正突变的核苷酸序列的外显子序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C);

[0467] (ii)Y是包含5'剪接位点的核苷酸序列;和

[0468] (iii)I(如果存在的话)包含内含子序列,其包括包含选自以下的核碱基的核苷酸:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),

[0469] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于定向插入双链DNA断裂(DSB)内,其中当供体多核苷酸插入DSB内时,E(如果存在的话)、Y和I(如果存在的话)构成有义链,其中Y包含一种或多种剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0470] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0471] *双向供体多核苷酸*

[0472] 在一些实施方案中,由本公开内容提供的供体多核苷酸是包含两个游离DNA末端的线性dsDNA分子。因此,通过细胞的NHEJ机制将供体多核苷酸插入DSB内可以在两个取向之一上发生;正向和反向。相应地,在一些方面,本公开内容提供了供体多核苷酸,其被配置用于双向插入通过定点核酸酶引入gDNA内的DSB内,其中当以任一取向插入DSB内时,所述供体多核苷酸校正或诱导细胞中的基因组DNA(gDNA)分子中的突变,并且提供一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工。

[0473] 在一些实施方案中,供体多核苷酸包括包含分支点序列的第一剪接信号,其中所述供体多核苷酸包含:

[0474] (i)第一链,其从5'到3'包括包含第一分支点序列和校正gDNA中的突变的第一核苷酸序列的核苷酸序列;和

[0475] (ii)第二链,其从5'到3'包括包含第二分支点序列和校正gDNA中的突变的第二核苷酸序列的核苷酸序列,其中所述第二链与第一链互补,

[0476] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,第一分支点序列和第二分支点序列的核苷酸序列符合分支点共有序列,并且第一分支点序列和第二分支点序列的核苷酸序列是互补的。在一些实施方案中,分支点共有序列是YTNAY(SEQ ID NO:49),其中Y是包含胞嘧啶(C)或胸腺嘧啶(T)核碱基的核苷酸,并且其中N是包含选自以下的核碱基的核苷酸:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C)。在一些实施方案中,第一分支点序列是TATTAAC(SEQ ID NO:50)。在一些实施方案中,第二分支点序列是GTTAATA(SEQ ID NO:51)。在一些实施方案中,第二分支点序列是TACTGAC(SEQ ID NO:52)。

[0477] 在一些实施方案中,供体多核苷酸包括包含多聚嘧啶片的第二剪接信号,其中第一链包含位于第一分支点序列下游的第一多聚嘧啶片;并且第二链包含位于第二分支点序列下游的第二多聚嘧啶片。

[0478] 在一些实施方案中,包含第一多聚嘧啶片和第二多聚嘧啶片的核苷酸序列各自包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),并且其中所述核苷酸序列为约100%、约90%-100%、或约80%-90%的嘧啶核碱基。在一些实施方案中,包含多聚嘧啶片的核苷酸序列是TTTTTTTCT(SEQ ID NO:53)。在一些实施方案中,包含多聚嘧啶片的核苷酸序列是TTTTTTTCTTTTT(SEQ ID NO:54)。在一些实施方案中,包含多聚嘧啶片的核苷酸序列是CTTCTTCTCTTCTTCC(SEQ ID NO:55)。

[0479] 在一些实施方案中,第一分支点序列和第一多聚嘧啶片彼此相邻。在一些实施方案中,第二分支点序列和第二多聚嘧啶片彼此相邻。

[0480] 在一些实施方案中,供体多核苷酸包括包含3'剪接位点的第三剪接信号,其中第一链包含核苷酸序列,其包含位于第一多聚嘧啶片下游的第一3'剪接位点,并且第二链包含核苷酸序列,其包含位于第二多聚嘧啶片下游的第二3'剪接位点。在一些实施方案中,第一3'剪接位点和第二3'剪接位点包含核苷酸序列YAG,并且其中Y是包含选自胸腺嘧啶(T)和胞嘧啶(C)的核碱基的核苷酸。

[0481] 在一些实施方案中,供体多核苷酸包含编码序列,其中第一链包含第一编码序列,其中第二链包含第二编码序列,其中校正gDNA中的突变的第一核苷酸序列包含第一编码序列,其中校正gDNA中的突变的第二核苷酸序列包含第二编码序列,其中所述第一编码序列位于第一3'剪接位点的下游,并且其中所述第二编码序列位于第二3'剪接位点的下游。在一些实施方案中,包含第一编码序列和第二编码序列的核苷酸序列包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C)。在一些实施方案中,包含(i)和(ii)的编码序列和/或剪接信号是不等同的或不互补的,以降低自退火。

[0482] 在一些实施方案中,由本公开内容提供的供体多核苷酸包含一个或多个定界符序列,其包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中所述核苷酸序列为长度约1-40、约1-30、约1-20、约1-15、约1-10、约30、约20、约10、约9、约8、约7、约6、约5、约4、约3、约2或1个核苷酸。在一些实施方案中,一个或多个定界符序列位于第一分支点序列和第二分支点序列之间。在一些实施方案中,一个或多个定界符序列位于第一分支点序列和第一多聚嘧啶片之间。在一些实施方案中,一个或多个定界符序列位于第二分支点和第二多聚嘧啶片之间。

[0483] 在一些实施方案中,供体多核苷酸被配置用于双向插入通过定点核酸酶引入gDNA内的DSB内,其中当供体多核苷酸以任一取向插入DSB内时,第一剪接信号和第二剪接信号,任选地第三剪接信号和编码序列构成有义链,从而校正了突变并提供一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工。

[0484] 在一些实施方案中,本公开内容提供了包括包含5'剪接位点的一种或多种剪接信号的供体多核苷酸,其中所述供体多核苷酸包含:

[0485] (i)第一链,其从5'到3'包括包含第一5'剪接位点序列和校正gDNA中的突变的第一核苷酸序列的核苷酸序列;和

[0486] (ii)第二链,其从5'到3'包括包含第二5'剪接位点序列和校正gDNA中的突变的第

二核苷酸序列的核苷酸序列,其中所述第二链与第一链互补,

[0487] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸。在一些实施方案中,第一链包含第一编码序列,并且第二链包含第二编码序列,其中所述第一编码序列位于第一5'剪接位点的上游,并且其中第二编码序列位于第二5'剪接位点的上游,并且其中第一链和第二链中的编码序列是不互补的(或者包含一个、两个、三个、四个或更多个错配),以降低自退火。在一些实施方案中,供体多核苷酸包含在第一5'剪接位点和第二5'剪接位点之间的定界符序列,其中所述定界符序列包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中所述核苷酸序列为长度约1-40、约1-30、约1-20、约1-15、约1-10、约30、约20、约10、约9、约8、约7、约6、约5、约4、约3、约2或1个核苷酸。

[0488] 在一些实施方案中,供体多核苷酸被配置用于双向插入DSB内,其中当供体多核苷酸以第一取向插入DSB内时,第一5'剪接位点和第一编码序列构成有义链,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,并且其中当供体多核苷酸以第二取向插入DSB内时,第二5'剪接位点和第二编码序列构成有义链,从而校正了突变并提供一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工。

[0489] 在一些实施方案中,供体多核苷酸包含非复制性双链DNA分子(dsDNA),其包含

[0490] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变接近3'剪接位点,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0491] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变接近3'剪接位点,其中所述第二链包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0492] 5'-[P1]a-[S1]b-[B]c-[S2]d-[P2]e-3',其中

[0493] (a)B包含分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中c是整数,其值指示B包含的核苷酸数目,其中c = 5-7;

[0494] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中a和e是整数,其值分别指示P1和P2包含的核苷酸数目,其中a=9-20和e=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,并且其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;和

[0495] (c)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)和胞嘧啶(C),其中b和d各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和d = 0-20,

[0496] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-

100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B和P2构成有义链,并且B和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B和P1构成有义链,并且B和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0497] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0498] 在一些实施方案中,供体多核苷酸包含非复制性双链DNA分子(dsDNA),其包含

[0499] (i)第一链,其从5’到3’包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变接近3’剪接位点,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0500] (ii)第二链,其从5’到3’包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变接近3’剪接位点,其中所述第二链包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0501] 5’-[P1]a-[S1]b-[B1]c-[S2]d-[B2]e-[S3]f-[P2]g-3’,其中

[0502] (a)B1(如果存在的话)和B2各自为分支点序列,其中B2包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中构成B1(如果存在的话)的分支点序列相对于构成B2的分支点序列处于相反取向,并且在供体多核苷酸的相反链上,其中B1和B2各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中c和e是整数,其值分别指示B1和B2包含的核苷酸数目,其中c = 0或5-7,其中e = 5-7;

[0503] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中a和g是整数,其值分别指示P1和P2包含的核苷酸数目,其中a=9-20和g=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;和

[0504] (c)S1、S2和S3(如果任一种存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中b、d和f各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中b和f = 1-20,其中d = 1-40,

[0505] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B2和P2构成有义链,并且B2和P2分别包含第一剪接信号和第二剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B1和P1构成有义链,并且B1和P1分别提供了第一剪接信号和第二剪接信号,从而校正了突变并提供剪

接信号,以控制由gDNA转录的前mRNA的加工,

[0506] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0507] 在一些实施方案中,供体多核苷酸包含非复制性双链DNA分子(dsDNA),其包含

[0508] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0509] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,并且其中所述第一链和第二链各自包含下式:

[0510] 5'-[E1]a-X1-[P1]b-[S1]c-[B]d-[S2]e-[P2]f-X2-[E2]g-3',其中

[0511] (a)B是分支点序列,其包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中B包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中d是整数,其值指示B包含的核苷酸数目,其中d = 5-7;

[0512] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中b和f是整数,其值分别指示P1和P2包含的核苷酸数目,其中b=9-20和f=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;

[0513] (c)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中a和g是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在供体多核苷酸的相反链上,其中构成E1和E2的核苷酸序列不互补;

[0514] (d)X1和X2各自为包含3'剪接位点的核苷酸序列,其中构成X1的核苷酸序列相对于构成X2的核苷酸序列处于相反方向,且在相反链上;和

[0515] (e)S1和S2(如果任一存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中c和e各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中c和e = 1-20,

[0516] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B、P2、E2和X2构成有义链,其中B、P2和X2分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B、P1、E1和X1构成有义链,其中B、P1和X1分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0517] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接

(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0518] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0519] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0520] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,和

[0521] 其中所述第一链和第二链各自包含下式:

[0522] 5'-[E1]a-X1-[P1]b-[S1]c-[B1]d-[S2]e-[B2]f-[S3]g-[P2]h-X2-[E2]i-3',其中

[0523] (a)B1(如果存在的话)和B2各自为分支点序列,其中B2包含符合供体多核苷酸的每条链上的分支点共有序列的核苷酸序列,其中构成B1(如果存在的话)的分支点序列相对于构成B2的分支点序列处于相反取向,并且在供体多核苷酸的相反链上,其中B1和B2各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中d和f是整数,其值分别指示B1和B2包含的核苷酸数目,其中d和f = 5-7;

[0524] (b)P1和P2是多聚嘧啶片,其各自包括包含选自以下的核碱基的核苷酸序列:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中b和h是整数,其值分别指示P1和P2包含的核苷酸数目,其中b=9-20和h=9-20,其中构成P1和P2的核苷酸序列各自为约100%、约90%-100%、约80%-90%的嘧啶核碱基,其中P1相对于P2处于相反取向且在供体多核苷酸的相反链上;

[0525] (c)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中a和i是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在相反链上,其中构成E1和E2的核苷酸序列不互补;

[0526] (d)X1和X2各自包括包含3'剪接位点的核苷酸序列,其中X1的核苷酸序列相对于X2的核苷酸序列处于相反方向,且在供体多核苷酸的相反链上;和

[0527] (e)S1、S2和S3(如果任一种存在的话)各自为包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中c、e和g各自为整数,其值分别指示定界符序列包含的核苷酸数目,其中c和g = 1-20,其中e = 1-40,

[0528] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,B2、P2、E2和X2构成有义链,其中B2、P2和X2分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突

变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,B1、P1、E1和X1构成有义链,其中B1、P1和X1分别包含第一剪接信号、第二剪接信号和第三剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0529] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0530] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含

[0531] (i)第一链,其从5'到3'包含校正细胞中的基因组DNA分子(gDNA)中的致病突变的第一核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第一链包含第一内含子序列和第一外显子序列,其中所述第一外显子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工;和

[0532] (ii)第二链,其从5'到3'包含校正细胞中的gDNA中的致病突变的第二核苷酸序列,其中所述致病突变是接近3'剪接位点的蛋白质编码突变,其中所述第二链包含第二内含子序列和第二外显子序列,其中所述第二外显子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述第二链与第一链互补,和

[0533] 其中所述第一链和第二链各自包含下式:

[0534] 5'-[E1]a-Y1-[S1]b-Y2-[E2]c-3',其中

[0535] (a)E1和E2各自为外显子序列,其各自包含校正突变的核苷酸序列,其中所述核苷酸序列包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中a和c是整数,其值分别指示E1和E2包含的核苷酸数目,其中构成E1的外显子序列相对于构成E2的外显子序列处于相反方向,且在相反链上,其中构成E1和E2的核苷酸序列不互补;

[0536] (b)Y1和Y2各自包括包含5'剪接位点的核苷酸序列,其中Y1的核苷酸序列相对于Y2的核苷酸序列处于相反方向,且在供体多核苷酸的相反链上;和

[0537] (c)S1(如果存在的话)是包含一个或多个核苷酸的定界符序列,所述一个或多个核苷酸包含选自以下的核碱基:腺嘌呤、鸟嘌呤、胸腺嘧啶和胞嘧啶,其中b是整数,其值指示定界符序列包含的核苷酸数目,其中b = 1-50,

[0538] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,其中所述供体多核苷酸被配置用于双向插入双链DNA断裂(DSB)内,其中当供体多核苷酸以第一取向插入DSB内时,E2和Y2构成有义链,并且E2包含剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,其中当供体多核苷酸以第二取向插入DSB内时,E1和Y1构成有义链,并且E1包含剪接信号,从而校正了突变并提供剪接信号,以控制由gDNA转录的前mRNA的加工,

[0539] 并且其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,从而校正了突变。

[0540] 在一些实施方案中,构成有义链的外显子序列校正了突变。在一些实施方案中,构

成有义链的包含3'剪接位点的核苷酸序列校正了突变。在一些实施方案中,构成有义链的包含5'剪接位点的核苷酸序列校正了突变。在一些实施方案中,构成有义链的包含多聚嘧啶片的核苷酸序列校正了突变。在一些实施方案中,构成有义链的包含分支点序列的核苷酸序列校正了突变。

[0541] 在一些实施方案中,由本公开内容提供的任何供体多核苷酸的每条链的最5'核苷酸包含5'磷酸基。在一些实施方案中,供体多核苷酸包含两个平端。在一些实施方案中,供体多核苷酸包含一个平端,并且包括包含突出物(例如5'或3'突出物)的一端。

[0542] 在一些实施方案中,供体多核苷酸包括包含一个或多个核苷酸的核苷酸序列,所述一个或多个核苷酸阻止定点核酸酶识别且切割供体多核苷酸。

[0543] 在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含校正细胞中的基因组DNA分子(gDNA)中的突变的核苷酸序列,所述突变引起糖原贮积病1a,其中所述突变位于人染色体17q21上的人G6PC基因中,并且导致氨基酸取代R83C或R83H,所述供体多核苷酸包含:

[0544] (i)第一链,其从5'到3'包括包含校正突变的外显子序列的核苷酸序列、以及包含一种或多种剪接信号以控制由基因组DNA分子转录的前体mRNA(前mRNA)加工的核苷酸序列,其中所述外显子序列包含与G6PC基因中的位置83处的密码子相对应的编码精氨酸(R)的密码子,其中所述一种或多种剪接信号是3'剪接位点、多聚嘧啶片和分支点序列的组合;和

[0545] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0546] 其中所述供体多核苷酸为长度约40-70个核苷酸,并包含两个平端,其中所述供体多核苷酸的每条链的最5'核苷酸包含5'磷酸部分,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,其中所述供体多核苷酸的插入在包含校正突变的外显子序列的gDNA中形成外显子,其中所述剪接信号指导外显子包括到mRNA内,从而校正了突变。在一些实施方案中,分支点序列包含核苷酸序列TTCAT,其中多聚嘧啶片包含核苷酸序列CTTGTTCTGTTTTTTT,其中3'剪接位点包含核苷酸序列TAG,并且其中外显子序列包含核苷酸序列GATTCTCTTTGGACAGCGCCCTTACT。

[0547] 在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:30(CH34 54-0)中阐述。

[0548] 在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:20(CH32 50-0)中阐述。

[0549] 制备和测试供体多核苷酸的方法

[0550] 由本公开内容提供的供体多核苷酸通过本领域已知的合适的DNA合成方法或手段来产生。DNA合成是脱氧核糖核酸(DNA)分子的天然或人工产生。术语DNA合成指DNA复制,DNA生物合成(例如体内DNA扩增),酶促DNA合成(例如聚合酶链反应(PCR);体外DNA扩增)或化学DNA合成。

[0551] 在一些实施方案中,供体多核苷酸的每条链通过寡核苷酸合成产生。寡核苷酸合成是具有确定的化学结构(序列)的相对较短的单链核酸片段或链的化学合成。寡核苷酸合成的方法是本领域已知的(参见例如,Reese(2005)Organic & Biomolecular Chemistry 3

(21):3851)。然后可以将两条链退火在一起或双链体化,以形成供体多核苷酸。

[0552] 在一些方面,通过本领域已知的合适方法确定供体多核苷酸插入DSB内。例如,在插入事件后,使用侧接DSB位点的PCR引物生成的PCR扩增子的核苷酸序列,就包含供体多核苷酸的核苷酸序列的存在进行分析。下一代测序(NGS)技术用于确定供体多核苷酸插入DSB内的程度,其就供体多核苷酸序列的存在或不存在分析PCR扩增子。进一步地,由于每个供体多核苷酸是线性的dsDNA分子,其可以在两个取向的任一上插入,因此NGS分析可以用于确定供体多核苷酸以任一方向的插入程度。

[0553] 在一些方面,通过由供体多核苷酸插入其内的gDNA转录的mRNA的核苷酸序列分析,来确定供体多核苷酸的插入及其校正突变的能力。通过本领域已知的合适方法,来分析由含有插入的供体多核苷酸的gDNA转录的mRNA。例如,从细胞(其用由本公开内容提供的供体多核苷酸或系统进行处理或者与之接触)中提取的mRNA被酶促转化成cDNA,其通过NGS分析进行进一步分析,以确定包含校正的突变的mRNA分子的程度。

[0554] 在其它方面,通过由mRNA(其由供体多核苷酸插入其内的gDNA转录)翻译的多肽的蛋白质序列分析,来确定供体多核苷酸的插入及其校正突变的能力。在一些实施方案中,供体多核苷酸通过将密码子掺入外显子来校正或诱导突变,所述密码子在包含gDNA分子的基因中制备氨基酸变化,其中来自含有插入的供体多核苷酸的基因的mRNA翻译生成包含氨基酸变化的多肽。多肽中的氨基酸变化通过蛋白质序列分析来确定,使用技术包括但不限于桑格测序、质谱法、测量多肽的酶促活性的功能测定、或使用对氨基酸变化反应性的抗体的免疫印迹。

[0555] 供体多核苷酸的使用

[0556] *DNA修复途径*

[0557] 细胞中的DNA断裂(例如DSB)修复主要通过两个DNA修复途径来完成,即非同源末端连接(NHEJ)修复途径和同源性指导的修复(HDR)途径。

[0558] 在NHEJ期间,Ku70/80异二聚体与DNA末端结合,并且募集DNA蛋白激酶(DNA-PK)(Cannan & Pederson(2015)J Cell Physiol 231:3-14)。一旦结合,DNA-PK就激活其自身的催化亚基(DNA-PKcs),并且进一步招募核酸内切酶Artemis(也称为SNM1c)。在DSB的子集处,Artemis去除过量的单链DNA(ssDNA),并且生成通过DNA连接酶IV进行连接的底物。通过NHEJ的DNA修复涉及经由规范DNA-PKcs/Ku70/80复合物,不依赖于序列同源性的平端连接机制。在细胞周期过程中,NHEJ占优势地在G0/G1和G2中发生(Chiruvella等人(2013)Cold Spring Harb Perspect Biol 5:a012757)。目前的研究已显示,NHEJ是在G0和G1中活跃的唯一DSB修复途径,而HDR主要在S和G2期中发挥功能,在复制相关的DSB的修复中起主要作用(Karanam等人(2012)Mol Cell 47:320-329;Li和Xu(2016)Acta Biochim Biophys Sin 48(7):641-646)。与HDR不同,NHEJ在分裂细胞和非分裂细胞两者,而不只是分裂细胞中是活跃的,这使得能够开发基于基因组编辑用于非分裂成体细胞(例如眼、脑、胰腺或心脏的细胞)的疗法。

[0559] 在通过HDR的DNA修复期间,主要通过MRE11-RAD50-NBS1(MRN)复合物,将DSB末端切除,以暴露3’ssDNA尾部(Heyer等人(2010)Annu Rev Genet 44: 113-139)。在生理条件下,相邻的姐妹染色单体将用作修复模板,提供同源序列,并且ssDNA将侵入由重组酶Rad51介导的模板,置换完整链以形成D环。D环延伸随后为分支迁移,以产生双霍利迪连接体,其

重溶完成了修复循环。HDR经常需要易错聚合酶,但通常视为无错的(Li和Xu(2016)Acta Biochim Biophys Sin 48(7):641-646)

[0560] 第三种修复机制是微同源性介导的末端连接(MMEJ),也称为“替代NHEJ”,其遗传结果与NHEJ相似,因为可以在切割位点处发生小缺失和插入。MMEJ利用侧接DNA断裂位点的几个核苷酸的同源序列,以驱动更有利的DNA末端连接修复结果,并且最近的报道已进一步阐明了这一过程的分子机制(Cho和Greenberg,(2015)Nature 518:174-176;Mateos-Gomez等人(2015)Nature 518,254-257;Ceccaldi等人(2015)Nature 528,258-262。

[0561] *使用供体多核苷酸来校正或诱导突变*

[0562] 在一些实施方案中,通过在由定点核酸酶(例如本文所述的那些)诱导的切割位点(例如DSB)处,将供体多核苷酸插入靶核酸(例如gDNA)内,由本公开内容提供的供体多核苷酸用于校正或诱导细胞中的gDNA中的突变。在一些实施方案中,细胞的NHEJ DNA修复机制使用供体多核苷酸来修复DSB,从而将供体多核苷酸插入DSB内,并且将供体多核苷酸的核苷酸序列加入gDNA中。在一些实施方案中,供体多核苷酸包含校正细胞中的gDNA中的致病突变的核苷酸序列,并且包括包含一种或多种剪接信号、以控制由gDNA转录的前体mRNA(前mRNA)加工的核苷酸序列。在一些实施方案中,将供体多核苷酸插入接近突变的位置处,从而校正了突变。在一些实施方案中,突变是取代、错义、无义、插入、缺失或移码突变。在一些实施方案中,突变在外显子中。在一些实施方案中,突变是取代、插入或缺失,并且位于内含子中。在一些实施方案中,突变接近gDNA中的剪接信号。在一些实施方案中,突变接近gDNA中的3’剪接位点。在一些实施方案中,突变接近gDNA中的5’剪接位点。在一些实施方案中,突变在gDNA中的剪接信号或包含剪接信号的序列中。在一些实施方案中,突变在gDNA中的3’剪接位点中。在一些实施方案中,突变在gDNA中的5’剪接位点中。在一些实施方案中,突变在多聚嘧啶片中。在一些实施方案中,突变在分支点序列中。在一些实施方案中,突变是蛋白质编码突变。在一些实施方案中,突变与疾病相关或引起疾病。

[0563] 在一些实施方案中,校正细胞中的gDNA中的致病突变的核苷酸序列包含分支点序列。在一些实施方案中,校正疾病的核苷酸序列包含多聚嘧啶片。在一些实施方案中,通过NHEJ DNA修复将供体多核苷酸插入DSB内。在一些实施方案中,通过NHEJ DNA修复将供体多核苷酸、供体多核苷酸的一部分插入靶核酸切割位点内。在某些方面,经由NHEJ修复将供体多核苷酸插入靶核酸内,可以导致例如内源基因序列的突变、缺失、改变、整合、基因校正、基因替换、加上基因标签、转基因插入、核苷酸缺失、基因破坏、易位和/或基因突变。

[0564] 在一些实施方案中,致病突变引起疾病糖原贮积病1a(GSD1a)。在一些实施方案中,致病突变位于人染色体17q21上的人G6PC基因中。在一些实施方案中,致病突变位于G6PC基因中,并且导致人G6PC蛋白中的R83C、R83H或E110K氨基酸取代。在一些实施方案中,G6PC基因中的致病突变导致人G6PC蛋白中的R83C氨基酸取代。在一些实施方案中,G6PC基因中的致病突变导致人G6PC蛋白中的R83H氨基酸取代。

[0565] 在一些实施方案中,G6PC基因中的致病突变导致人G6PC蛋白中的E110K氨基酸取代。在一些实施方案中,本公开内容提供了供体多核苷酸,其用于修复细胞中通过定点核酸酶(例如,Cas9)引入靶核酸分子(例如,gDNA)内的DSB。在一些实施方案中,供体多核苷酸被细胞的非同源末端连接(NHEJ)修复途径使用,以修复靶核酸分子中的DSB。在一些实施方案中,定点核酸酶是Cas核酸酶。在一些实施方案中,Cas核酸酶是Cas9。本文所述的定点核酸

酶可以将DSB引入细胞中的靶核酸(例如,基因组DNA)中。由定点核酸酶诱导的DSB引入细胞基因组DNA中,将刺激内源性DNA修复途径,例如本文所述的那些。相比之下,通过NHEJ的DSB修复在非分裂细胞中发生,但不需要使用同源多核苷酸序列用于修复。然而,NHEJ途径可以用于在修复期间将多核苷酸(例如供体多核苷酸)插入DSB内。

[0566] 相应地,在一些实施方案中,本公开内容提供了包含非复制性双链DNA分子(dsDNA)的供体多核苷酸,所述dsDNA包含校正细胞中的基因组DNA分子(gDNA)中的突变的核苷酸序列,所述突变引起糖原贮积病1a,其中所述突变位于人染色体17q21上的人G6PC基因中,并且导致氨基酸取代R83C或R83H,所述供体多核苷酸包含:

[0567] (i)第一链,其从5'到3'包括包含校正突变的外显子序列的核苷酸序列、以及包含一种或多种剪接信号以控制由基因组DNA分子转录的前体mRNA(前mRNA)加工的核苷酸序列,其中所述外显子序列包含与G6PC基因中的位置83处的密码子相对应的编码精氨酸(R)的密码子,其中所述一种或多种剪接信号是3'剪接位点、多聚嘧啶片和分支点序列的组合;和

[0568] (ii)第二链,其包含与第一链互补的核苷酸序列,

[0569] 其中所述供体多核苷酸为长度约40-70个核苷酸,并包含两个平端,其中所述供体多核苷酸的每条链的最5'核苷酸包含5'磷酸部分,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入双链DNA断裂(DSB)内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,其中所述供体多核苷酸的插入在包含校正突变的外显子序列的gDNA中形成外显子,其中所述剪接信号指导外显子包括到mRNA内,从而校正了突变。在一些实施方案中,分支点序列包含核苷酸序列TTCAT,其中多聚嘧啶片包含核苷酸序列CTTGTTCTGTTTTTTT,其中3'剪接位点包含核苷酸序列TAG,并且其中外显子序列包含核苷酸序列GATTCTCTTTGGACAGCGCCCTTACT。

[0570] 在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:30(CH34 54-0)中阐述。在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:20(CH32 50-0)中阐述。

[0571] 在一些实施方案中,致病突变引起疾病庞贝氏病。在一些实施方案中,致病突变位于人染色体17q25.3上的人葡萄糖苷酶α(GAA)基因中。在一些实施方案中,突变在GAA的剪接信号中,其导致缺乏外显子2的GAA基因的mRNA转录物。在一些实施方案中,突变在GAA的剪接信号中,其导致一个或多个隐蔽剪接位点的激活。隐蔽剪接位点是这样的mRNA序列,其编码剪接信号并且具有与剪接体相互作用的潜力,但通常不用于mRNA剪接以生成功能性蛋白质产物。基因中的突变可以激活隐蔽剪接信号,其导致异常蛋白质产物的产生。

[0572] 相应地,在一些实施方案中,本公开内容提供了包含非复制性dsDNA分子的供体多核苷酸,所述dsDNA分子包含校正细胞中的gDNA分子中的突变的核苷酸序列,所述突变引起庞贝氏病,其中所述突变在GAA的剪接信号中,其导致缺乏外显子2和/或一个或多个隐蔽剪接位点的激活的GAA基因的mRNA转录物,所述供体多核苷酸包含:

[0573] (i)第一链,其从5'到3'包含校正突变的第一核苷酸序列,其中所述第一链包含第一内含子序列,其中所述第一内含子序列校正了突变,其中所述第一链包含一种或多种剪接信号,以控制由gDNA转录的前体mRNA(前mRNA)的加工,其中所述一种或多种剪接信号包含3'剪接位点、多聚嘧啶片和分支点序列的组合;和

[0574] (ii)第二链,其从5’到3’包含第二内含子序列,其中所述第二内含子序列校正了突变,其中所述第二链包含一种或多种剪接信号,以控制由gDNA转录的前mRNA的加工,其中所述一种或多种剪接信号包含3’剪接位点、多聚嘧啶片和分支点序列的组合,其中所述第二链与第一链互补,

[0575] 其中所述供体多核苷酸为长度约10-500、约10-400、约10-300、约10-200、约10-100、约20-80、约30-70或约40-60个核苷酸,并包含两个平端,其中所述供体多核苷酸的每条链的最5’核苷酸包含5’磷酸部分,其中当供体多核苷酸与定点核酸酶组合引入细胞内时,NHEJ DNA修复途径将供体多核苷酸插入DSB内,所述DSB通过定点核酸酶在接近突变的位置处引入gDNA内,其中所述供体多核苷酸被配置用于双向插入DSB断裂内,其中以任一方向的插入形成3’剪接位点、多聚嘧啶片和校正突变的分支点序列,其中所述剪接信号指导外显子2包括到mRNA内,从而校正了突变。

[0576] 在一些实施方案中,供体多核苷酸的核苷酸序列在SEQ ID NO:63(GAA_50-0)中阐述。

[0577] 相应地,在一些实施方案中,提供了单个供体多核苷酸或相同供体多核苷酸的多重拷贝。在其它实施方案中,提供了两个或更多个供体多核苷酸,使得修复可以在两个或更多个靶位点处发生。例如,提供了不同的供体多核苷酸,以修复细胞中的单个基因或细胞中的两个不同基因。在一些实施方案中,不同的供体多核苷酸以独立的拷贝数提供。

[0578] 在一些实施方案中,供体多核苷酸作为由非同源末端连接(NHEJ)介导的插入掺入靶核酸内。在一些实施方案中,供体多核苷酸序列与切割位点附近的核酸序列没有相似性。在一些实施方案中,提供了单个供体多核苷酸或相同供体多核苷酸的多重拷贝。在其它实施方案中,具有不同序列的两个或更多个供体多核苷酸通过非同源末端连接插入两个或更多个位点处。在一些实施方案中,不同的供体多核苷酸以独立的拷贝数提供。

[0579] CRISPR/Cas核酸酶系统

[0580] 天然存在的CRISPR/Cas系统是遗传防御系统,其在原核生物中提供了某种形式的获得性免疫。CRISPR是成簇规律间隔短回文重复的缩写,是在细菌和古细菌的基因组中发现的DNA序列家族,其含有与先前暴露于细胞的外来DNA具有相似性的DNA(间隔区DNA)片段,例如通过已感染或攻击原核生物的病毒。这些DNA片段被原核生物用于检测且破坏例如在随后的攻击过程中由相似病毒重新引入时的相似外源DNA。CRISPR基因座的转录导致包含间隔区序列的RNA分子的形成,所述间隔区序列结合并靶向能够识别且切割外来的外源性DNA的Cas(CRISPR相关)蛋白。众多类型和种类的CRISPR/Cas系统已得到描述(参见例如,Koonin等人(2017)Curr Opin Microbiol 37:67-78)。

[0581] CRISPR/Cas系统的改造版本已以众多形式开发,以突变或编辑来自其它物种的细胞的基因组DNA。使用CRISPR/Cas系统的一般方法涉及与引导RNA(gRNA)组合的定点核酸酶(例如Cas核酸酶)在细胞内的异源表达或引入,导致在精确的可靶向位置处、在细胞的基因组DNA的主链中的DNA切割事件(例如,形成单链或双链断裂(SSB或DSB))。其中DNA切割事件通过细胞修复的方式提供了通过添加、去除或修饰(取代)DNA核苷酸或序列(例如基因)来编辑基因组的机会。

[0582] *Cas核酸酶*

[0583] 在一些实施方案中,本公开内容提供了包含定点核酸酶的组合物和系统(例如,改

造的CRISPR/Cas系统),其中所述定点核酸酶是Cas核酸酶。Cas核酸酶可以包含与引导RNA(gRNA)相互作用的至少一个结构域。另外,Cas核酸酶通过引导RNA被导向靶序列。引导RNA与Cas核酸酶以及靶序列相互作用,使得一旦被导向靶序列,Cas核酸酶就能够切割靶序列。在一些实施方案中,引导RNA提供了关于靶序列切割的特异性,并且Cas核酸酶是通用的并且与不同的引导RNA配对,以切割不同的靶序列。

[0584] 在一些实施方案中,CRISPR/Cas系统包含衍生自I型、II型或III型系统的组分。CRISPR/Cas基因座的更新分类方案定义了具有I至V或VI型的1类和2类CRISPR/Cas系统(Makarova等人(2015)Nat Rev Microbiol,13(11):722-36;Shmakov等人(2015)Mol Cell,60:385-397)。2类CRISPR/Cas系统具有单一蛋白质效应物。II、V和VI型的Cas蛋白是单蛋白、RNA引导的核酸内切酶,在本文中称为"2类Cas核酸酶"。2类Cas核酸酶包括例如Cas9、Cpf1、C2c1、C2c2和C2c3蛋白。Cpf1核酸酶(Zetsche等人(2015)Cell 163:1-13)与Cas9同源,并且含有RuvC样核酸酶结构域。

[0585] 在一些实施方案中,Cas核酸酶来自II型CRISPR/Cas系统(例如,来自CRISPR/Cas9系统的Cas9蛋白)。在一些实施方案中,Cas核酸酶来自2类CRISPR/Cas系统(单蛋白Cas核酸酶,例如Cas9蛋白或Cpf1蛋白)。Cas9和Cpf1蛋白家族是具有DNA核酸内切酶活性的酶,并且可以通过设计适当的引导RNA来指导它们切割所需的核酸靶,如本文进一步所述。

[0586] II型CRISPR/Cas系统组分来自IIA型、IIB型或IIC型系统。涵盖了Cas9及其直系同源物。Cas9核酸酶或其它组分来自其的非限制性的示例性物种包括化脓性链球菌、嗜热链球菌(*Streptococcus thermophilus*)、链球菌属物种(*Streptococcus sp.*)、金黄色葡萄球菌、无害李斯特菌(*Listeria innocua*)、加氏乳杆菌(*Lactobacillus gasseri*)、新凶手弗朗西丝氏菌(*Francisella novicida*)、产琥珀酸沃林氏菌(*Wolinella succinogenes*)、华德萨特氏菌(*Sutterella wadsworthensis*)、γ-变形菌(*Gamma proteobacterium*)、脑膜炎奈瑟菌(*Neisseria meningitidis*)、空肠弯曲菌(*Campylobacter jejuni*)、多杀性巴氏杆菌(*Pasteurella multocida*)、琥珀酸丝状杆菌(*Fibrobacter succinogene*)、深红红螺菌(*Rhodospirillum rubrum*)、达氏拟诺卡氏菌(*Nocardiopsis dassonvillei*)、始旋链霉菌(*Streptomyces pristinaespiralis*)、绿色产色链霉菌(*Streptomyces viridochromogenes*)、绿色产色链霉菌、玫瑰链孢囊菌(*Streptosporangium roseum*)、玫瑰链孢囊菌、酸热脂环酸芽孢杆菌(*Alicyclobacillus acidocaldarius*)、假蕈状芽孢杆菌(*Bacillus pseudomycoides*)、还原硒酸盐芽孢杆菌(*Bacillus selenitireducens*)、唐松草微小杆菌(*Exiguobacterium sibiricum*)、德氏乳杆菌(*Lactobacillus delbrueckii*)、唾液乳杆菌(*Lactobacillus salivarius*)、布氏乳杆菌(*Lactobacillus buchneri*)、齿垢密螺旋体(*Treponema denticola*)、海洋微颤菌(*Microscilla marina*)、伯克氏菌目细菌(*Burkholderiales bacterium*)、食萘极单胞菌(*Polaromonas naphthalenivorans*)、极地单胞菌属物种(*Polaromonas sp.*)、瓦氏鳄球藻(*Crocosphaera watsonii*)、蓝杆藻属物种(*Cyanothece sp.*)、铜绿微囊藻(*Microcystis aeruginosa*)、聚球藻属物种(*Synechococcus sp*)、阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*)、丹氏制氨菌(*Ammonifex degensii*)、*Caldicelulosiruptor becscii*、*Candidatus Desulforudis*、肉毒杆菌(*Clostridium botulinum*)、艰难梭菌(*Clostridium difficile*)、大芬戈尔德菌(*Finegoldia magna*)、嗜热需碱厌氧菌(*Natranaerobius thermophilus*)、嗜热丙酸降解菌

(*Pelotomaculum thermopropionicum*)、嗜酸喜温硫杆菌(*Acidithiobacillus caldus*)、嗜酸氧化亚铁硫杆菌(*Acidithiobacillus ferrooxidans*)、酒色闪杆菌(*Allochromatium vinosum*)、海杆菌属物种(*Marinobacter sp.*)、嗜盐亚硝化球菌(*Nitrosococcus halophilus*)、瓦氏亚硝化球菌(*Nitrosococcus watsoni*)、游海假交替单胞菌(*Pseudoalteromonas haloplanktis*)、成簇细枝菌(*Ktedonobacter racemifer*)、调查甲烷盐菌(*Methanohalobium evestigatum*)、多变鱼腥藻(*Anabaena variabilis*)、泡沫节球藻(*Nodularia spumigena*)、念珠藻属物种(*Nostoc sp.*)、极大节螺藻(*Arthrospira maxima*)、钝顶节螺藻(*Arthrospira platensis*)、节螺藻属物种(*Arthrospira sp.*)、鞘丝藻属物种(*Lyngbya sp.*)、喜泥微鞘藻(*Microcoleus chthonoplastes*)、颤藻属物种(*Oscillatoria sp.*)、运动石孢菌(*Petrotoga mobilis*)、非洲栖热腔菌(*Thermosipho africanus*)、巴氏链球菌(*Streptococcus pasteurianus*)、灰色奈瑟菌(*Neisseria cinerea*)、海鸥弯曲菌(*Campylobacter lari*)、食淋洗物细小棒菌(*Parvibaculum lavamentivorans*)、白喉棒状杆菌(*Corynebacterium diphtheria*)或*Acaryochloris marina*。在一些实施方案中,Cas9蛋白来自化脓性链球菌(SpCas9)。在一些实施方案中,Cas9蛋白来自嗜热链球菌(StCas9)。在一些实施方案中,Cas9蛋白来自脑膜炎奈瑟菌(NmCas9)。在一些实施方案中,Cas9蛋白来自金黄色葡萄球菌(SaCas9)。在一些实施方案中,Cas9蛋白来自空肠弯曲菌(CjCas9)。

[0587] 在一些实施方案中,Cas核酸酶可以包含多于一个核酸酶结构域。例如,Cas9核酸酶可以包含至少一个RuvC样核酸酶结构域(例如Cpf1)和至少一个HNH样核酸酶结构域(例如Cas9)。在一些实施方案中,Cas9核酸酶在靶序列中引入DSB。在一些实施方案中,Cas9核酸酶被修饰为仅含有一个功能性核酸酶结构域。例如,这样修饰Cas9核酸酶,使得核酸酶结构域之一突变或者全部或部分缺失,以降低其核酸切割活性。在一些实施方案中,Cas9核酸酶被修饰为不含功能性RuvC样核酸酶结构域。在其它实施方案中,Cas9核酸酶被修饰为不含功能性HNH样核酸酶结构域。在其中仅一个核酸酶结构域有功能的一些实施方案中,Cas9核酸酶是能够将单链断裂(“切口”)引入靶序列内的切口酶。在一些实施方案中,Cas9核酸酶结构域内的保守氨基酸被取代,以降低或改变核酸酶活性。在一些实施方案中,Cas核酸酶切口酶包含在RuvC样核酸酶结构域中的氨基酸取代。RuvC样核酸酶结构域中的示例性氨基酸取代包括D10A(基于化脓性链球菌Cas9核酸酶)。在一些实施方案中,切口酶包含在HNH样核酸酶结构域中的氨基酸取代。HNH样核酸酶结构域中的示例性氨基酸取代包括E762A、H840A、N863A、H983A和D986A(基于化脓性链球菌Cas9核酸酶)。在一些实施方案中,本文描述的核酸酶系统包含切口酶、以及分别与靶序列的有义链和反义链互补的一对引导RNA。引导RNA通过在靶序列的相反链上生成切口(即,双重切口),来引导切口酶靶向并引入DSB。使用嵌合Cas9核酸酶,其中蛋白质的一个结构域或区域替换为不同蛋白质的一部分。例如,Cas9核酸酶结构域由来自不同核酸酶(例如Fok1)的结构域替换。Cas9核酸酶是修饰的核酸酶。

[0588] 在替代实施方案中,Cas核酸酶来自I型CRISPR/Cas系统。在一些实施方案中,Cas核酸酶是I型CRISPR/Cas系统的Cascade复合物的组分。例如,Cas核酸酶是Cas3核酸酶。在一些实施方案中,Cas核酸酶衍生自III型CRISPR/Cas系统。在一些实施方案中,Cas核酸酶衍生自IV型CRISPR/Cas系统。在一些实施方案中,Cas核酸酶衍生自V型CRISPR/Cas系统。在

一些实施方案中,Cas核酸酶衍生自VI型CRISPR/Cas系统。

[0589] *引导RNA(gRNA)*

[0590] 改造的CRISPR/Cas系统包含至少两种组分:1)引导RNA(gRNA)分子和2)Cas核酸酶,其相互作用以形成gRNA/Cas核酸酶复合物。gRNA至少包含称为"间隔区"的用户定义的靶向结构域,其包含核苷酸序列和CRISPR重复序列。在改造的CRISPR/Cas系统中,通过生成包含具有核苷酸序列(其能够以互补的方式与特异性靶序列结合)的间隔区的gRNA,将gRNA/Cas核酸酶复合物靶向靶核酸(例如基因组DNA分子)内的特异性目的靶序列。因此,间隔区提供了gRNA/Cas核酸酶复合物的靶向功能。

[0591] 在天然存在的II-CRISPR/Cas型系统中,"gRNA"由两条RNA链组成:1)包含间隔区和CRISPR重复序列的CRISPR RNA(crRNA),以及2)反式激活CRISPR RNA(tracrRNA)。在II型CRISPR/Cas系统中,包含CRISPR重复序列的crRNA的一部分与tracrRNA的一部分杂交,以形成crRNA:tracrRNA双链体,其与Cas核酸酶(例如Cas9)相互作用。如本文使用的,术语"分裂gRNA"或"模块gRNA"指包含两条RNA链的gRNA分子,其中第一RNA链掺入crRNA功能和/或结构,而第二RNA链掺入tracrRNA功能和/或结构,并且其中所述第一RNA链和第二RNA链部分杂交。

[0592] 相应地,在一些实施方案中,由本公开内容提供的gRNA包含两个RNA分子。在一些实施方案中,gRNA包含CRISPR RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)。在一些实施方案中,gRNA是分裂gRNA。在一些实施方案中,gRNA是模块gRNA。在一些实施方案中,分裂gRNA包含第一链,其从5'到3'包含间隔区和第一互补性区域;以及第二链,其从5'到3'包含第二互补性区域;以及任选地尾部结构域。

[0593] 在一些实施方案中,crRNA包含间隔区,其包含的核苷酸序列与这样的序列互补并杂交,所述序列与靶核酸(例如,基因组DNA分子)上的靶序列互补。在一些实施方案中,crRNA包含与tracrRNA的一部分互补并杂交的区域。

[0594] 在一些实施方案中,tracrRNA可以包含来自天然存在的CRISPR/Cas系统的野生型tracrRNA序列的全部或一部分。在一些实施方案中,tracrRNA可以包含野生型tracr RNA的截短或修饰的变体。tracr RNA的长度可能取决于使用的CRISPR/Cas系统。在一些实施方案中,tracrRNA可以包含长度5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、60、70、80、90、100个或多于100个核苷酸。在某些实施方案中,tracrRNA为长度至少26个核苷酸。在另外的实施方案中,tracrRNA为长度至少40个核苷酸。在一些实施方案中,tracrRNA可以包含某些二级结构,例如一种或多种发夹或茎环结构、或者一种或多种凸起结构。

[0595] *单引导RNA(sgRNA)*

[0596] 改造的CRISPR/Cas核酸酶系统经常通过在这些组分之间添加接头,将crRNA和tracrRNA组合成单个RNA分子,在本文中称为"单引导RNA"(sgRNA)。不受理论的束缚,类似于双链体化的crRNA和tracrRNA,sgRNA与Cas核酸酶(例如Cas9)形成复合物,将Cas核酸酶引导至靶序列并激活Cas核酸酶,用于切割靶核酸(例如,基因组DNA)。相应地,在一些实施方案中,gRNA可以包含可操作地连接的crRNA和tracrRNA。在一些实施方案中,sgRNA可以包含与tracrRNA共价连接的crRNA。在一些实施方案中,crRNA和tracrRNA经由接头共价连接。在一些实施方案中,sgRNA可以包含经由crRNA和tracrRNA之间的碱基配对的茎环结构。在

一些实施方案中,sgRNA从5'到3'包含间隔区、第一互补性区域、连接结构域、第二互补性区域以及任选地尾部结构域。

[0597] *间隔区*

[0598] 在一些实施方案中,由本公开内容提供的gRNA包含间隔区序列。间隔区序列是定义靶核酸(例如DNA)的靶位点的序列。靶核酸是双链分子:一条链包含与PAM序列相邻的靶序列,并且被称为"PAM链",而第二链被称为"非PAM链",并且与PAM链和靶序列互补。gRNA间隔区和靶序列两者均与靶核酸的非PAM链互补。gRNA间隔区序列与互补链(例如,靶核酸/靶位点的非PAM链)杂交。在一些实施方案中,间隔区与靶序列的互补链(例如,非PAM链)充分互补,以便将Cas核酸酶靶向靶核酸。在一些实施方案中,间隔区与靶核酸的非PAM链至少80%、85%、90%或95%互补。在一些实施方案中,间隔区与靶核酸的非PAM链100%互补。在一些实施方案中,间隔区包含与靶核酸的非PAM链不互补的1、2、3、4、5、6个或更多个核苷酸。在一些实施方案中,间隔区包含与靶核酸的非PAM链不互补的1个核苷酸。在一些实施方案中,间隔区包含与靶核酸的非PAM链不互补的2个核苷酸。

[0599] 在一些实施方案中,gRNA的最5'核苷酸包含间隔区的最5'核苷酸。在一些实施方案中,间隔区位于crRNA的5'端处。在一些实施方案中,间隔区位于sgRNA的5'端处。在一些实施方案中,间隔区为长度约15-50、约20-45、约25-40或约30-35个核苷酸。在一些实施方案中,间隔区为长度约19-22个核苷酸。在一些实施方案中,间隔区为长度约15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个核苷酸。在一些实施方案中,间隔区为长度19个核苷酸。在一些实施方案中,间隔区为长度20个核苷酸,在一些实施方案中,间隔区为长度21个核苷酸。

[0600] 在一些实施方案中,靶序列和PAM的核苷酸序列包含式5' N_{19-21}-N-R-G-3'(SEQ ID NO:59),其中N是任何核苷酸,并且其中R是包含核碱基腺嘌呤(A)或鸟嘌呤(G)的核苷酸,并且其中三个3'末端核酸N-R-G代表化脓性链球菌PAM。在一些实施方案中,间隔区的核苷酸序列使用计算机程序进行设计或选择。计算机程序可以使用变量,例如预测的解链温度、二级结构形成、预测的退火温度、序列同一性、基因组背景、染色质可及性、% GC、基因组出现频率(例如,等同或相似,但由于错配、插入或缺失而在一个或多个点中改变的序列)、甲基化状态和/或SNP的存在。

[0601] 在一些实施方案中,间隔区包含至少一个或多个修饰的核苷酸,例如本文所述的那些。本公开内容提供了包含间隔区的gRNA分子,所述间隔区可以包含核碱基尿嘧啶(U),而编码gRNA(其包括包含核碱基尿嘧啶(U)的间隔区)的任何DNA,在相应位置中包含核碱基胸腺嘧啶(T)。

[0602] *制备gRNA的方法*

[0603] 本公开内容的gRNA通过本领域中可用的合适手段产生,所述手段包括但不限于体外转录(IVT)、合成和/或化学合成方法或其组合。利用酶促(IVT)、固相、液相、组合合成方法、小区域合成和连接方法。在一个实施方案中,使用IVT酶促合成方法制备gRNA。通过IVT制备多核苷酸的方法是本领域已知的,并且在国际申请PCT/US2013/30062中进行描述。相应地,本公开内容还包括用于体外转录本文所述的gRNA的多核苷酸,例如DNA、构建体和载体。

[0604] 在一些方面,在合成或合成后过程中,将非天然修饰的核碱基引入多核苷酸例如

gRNA内。在某些实施方案中,修饰在核苷间键合、嘌呤或嘧啶碱基、或糖上。在特定实施方案中,修饰在多核苷酸的末端处引入;使用化学合成或聚合酶。修饰的核酸及其合成的实例在PCT申请号PCT/US2012/058519中公开。修饰的多核苷酸的合成也在Verma和Eckstein,Annual Review of Biochemistry,第76卷,99-134(1998)中进行描述。

[0605] 在一些方面,酶促或化学连接方法用于缀合具有不同功能部分的多核苷酸或其区域,例如靶向试剂或递送试剂、荧光标记、液体、纳米颗粒等。多核苷酸和修饰的多核苷酸的缀合物在Goodchild,Bioconjugate Chemistry,第1(3)卷,165-187(1990)中进行综述。

[0606] 本发明的某些实施方案还提供了编码本文所述的gRNA的核酸,例如载体。在一些实施方案中,核酸是DNA分子。在其它实施方案中,核酸是RNA分子。在一些实施方案中,核酸包含编码crRNA的核苷酸序列。在一些实施方案中,编码crRNA的核苷酸序列包含间隔区,其侧翼为来自天然存在的CRISPR/Cas系统的重复序列的全部或一部分。在一些实施方案中,核酸包含编码tracrRNA的核苷酸序列。在一些实施方案中,crRNA和tracrRNA由两种分开的核酸编码。在其它实施方案中,crRNA和tracrRNA由单一核酸编码。在一些实施方案中,crRNA和tracrRNA由单一核酸的相反链编码。在其它实施方案中,crRNA和tracrRNA由单一核酸的同一链编码。

[0607] 在一些实施方案中,通过本领域中描述的任何手段化学合成由本公开内容提供的gRNA(参见例如,WO/2005/01248)。随着化学合成程序不断扩展,通过程序例如高效液相层析(HPLC,其避免使用凝胶如PAGE)纯化此类RNA趋于变得更具挑战性,因为多核苷酸长度显著增加超过了一百多个核苷酸。用于生成更大长度的RNA的一种方法是产生连接在一起的两种或更多种分子。

[0608] 在一些实施方案中,由本公开内容提供的gRNA通过酶促方法(例如,体外转录、IVT)合成。

[0609] 如本领域所述,可以在RNA的化学合成和/或酶促生成期间或之后引入各种类型的RNA修饰,例如,增强稳定性、降低先天性免疫应答的可能性或程度、和/或增强其它属性的修饰。

[0610] 在某些实施方案中,多于一种引导RNA可以与CRISPR/Cas核酸酶系统一起使用。每种引导RNA可能含有不同的靶向序列,使得CRISPR/Cas系统裂解多于一种靶核酸。在一些实施方案中,一种或多种引导RNA可以在Cas9 RNP复合物中具有相同或不同特性,例如活性或稳定性。当使用多于一种引导RNA时,每种引导RNA可以在相同或不同的载体上编码。用于驱动多于一种引导RNA表达的启动子是相同或不同的。

[0611] 引导RNA可以经由crRNA的靶向序列(例如间隔区序列)靶向任何目的序列。在一些实施方案中,引导RNA的靶向序列和靶核酸分子上的靶序列之间的互补性程度为约60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%。在一些实施方案中,引导RNA的靶向序列和靶核酸分子上的靶序列是100%互补的。在其它实施方案中,引导RNA的靶向序列和靶核酸分子上的靶序列可以含有至少一个错配。例如,引导RNA的靶向序列和靶核酸分子上的靶序列可以含有1、2、3、4、5、6、7、8、9或10个错配。在一些实施方案中,引导RNA的靶向序列和靶核酸分子上的靶序列可以含有1-6个错配。在一些实施方案中,引导RNA的靶向序列和靶核酸分子上的靶序列可以含有5或6个错配。

[0612] 靶向序列的长度可能取决于CRISPR/Cas9系统和使用的组分。例如,来自不同细菌

物种的不同Cas9蛋白具有不同的最佳靶向序列长度。相应地,靶向序列可以包含长度5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50个或多于50个核苷酸。在一些实施方案中,靶向序列可以包含长度18-24个核苷酸。在一些实施方案中,靶向序列可以包含长度19-21个核苷酸。在一些实施方案中,靶向序列可以包含长度20个核苷酸。

[0613] 在本公开内容的一些实施方案中,CRISPR/Cas核酸酶系统包括至少一种引导RNA。在一些实施方案中,引导RNA和Cas蛋白可以形成核糖核蛋白(RNP),例如CRISPR/Cas复合物。引导RNA可以将Cas蛋白引导至靶核酸分子(例如,基因组DNA分子)上的靶序列,在其中Cas蛋白切割靶核酸。在一些实施方案中,CRISPR/Cas复合物是Cpf1/引导RNA复合物。在一些实施方案中,CRISPR复合物是II型CRISPR/Cas9复合物。在一些实施方案中,Cas蛋白是Cas9蛋白。在一些实施方案中,CRISPR/Cas9复合物是Cas9/引导RNA复合物。

[0614] *改造的核酸酶*

[0615] 在另外的实施方案中,由本公开内容提供的供体多核苷酸与定点核酸酶组合使用,其中所述定点核酸酶是改造的核酸酶。示例性的改造的核酸酶是大范围核酸酶(例如,归巢核酸内切酶)、ZFN、TALEN和megaTAL。

[0616] 天然存在的大范围核酸酶可以识别且切割约12至40个碱基对的双链DNA序列,并且通常分组成五个家族。在一些实施方案中,大范围核酸酶选自LAGLIDADG家族、GIY-YIG家族、HNH家族、His-Cys盒家族和PD-(D/E)XK家族。在一些实施方案中,大范围核酸酶的DNA结合结构域被改造为识别并结合除其同源靶序列外的序列。在一些实施方案中,大范围核酸酶的DNA结合结构域与异源核酸酶结构域融合。在一些实施方案中,大范围核酸酶,例如归巢核酸内切酶,与TAL模块融合以产生杂合蛋白,例如“megaTAL”蛋白。megaTAL蛋白通过识别大范围核酸酶的DNA结合结构域和TAL模块两者的靶序列,具有改善的DNA靶向特异性。

[0617] ZFN是包含锌指DNA结合结构域(“锌指”或“ZF”)和核酸酶结构域的融合蛋白。每种天然存在的ZF可以结合三个连续的碱基对(DNA三联体),并且ZF重复被组合以识别DNA靶序列并提供足够的亲和力。因此,改造的ZF重复被组合以识别更长的DNA序列,例如9-、12-、15-或18-bp等。在一些实施方案中,ZFN包含与来自限制性核酸内切酶的核酸酶结构域融合的ZF。例如,限制性核酸内切酶是FokI。在一些实施方案中,核酸酶结构域包含二聚化结构域,例如当核酸酶二聚化以具有活性时,并且包含ZF重复和核酸酶结构域的一对ZFN被设计用于靶向靶序列,所述靶序列包含由DNA分子的相反链上的每个ZF重复识别的两个半靶序列,在其之间具有互连序列(其在文献中有时称为间隔区)。例如,互连序列为长度5至7 bp。当该对的两个ZFN均结合时,核酸酶结构域可能二聚化并在互连序列内引入DSB。在一些实施方案中,核酸酶结构域的二聚化结构域包含旋钮入孔(knob-into-hole)基序,以促进二聚化。例如,ZFN包含在FokI的二聚化结构域中的旋钮入孔基序。

[0618] TALEN的DNA结合结构域通常包含可变数目的34或35个氨基酸重复(“模块”或“TAL模块”),其中每个模块与单个DNA碱基对A、T、G或C结合。在每个模块的位置12和13处的相邻残基(“重复可变双残基”或RVD)指定了该模块与之结合的单个DNA碱基对。尽管用于识别G的模块也可能对于A具有亲和力,但TALEN获益于简单的识别代码—对于4种碱基各一个模块—这极大地简化了识别特异性靶序列的DNA结合结构域的定制。在一些实施方案中,TALEN可以包含来自限制性核酸内切酶的核酸酶结构域。例如,限制性核酸内切酶是FokI。

在一些实施方案中,核酸酶结构域可以二聚化以具有活性,并且一对TALENS被设计用于靶向靶序列,所述靶序列包含由DNA分子的相反链上的每个DNA结合结构域识别的两个半靶序列,在其之间具有互连序列。例如,每个半靶序列在10至20 bp的范围内,并且互连序列为长度12至19 bp。当该对的两个TALEN均结合时,核酸酶结构域可能二聚化并在互连序列内引入DSB。在一些实施方案中,核酸酶结构域的二聚化结构域可能包含旋钮入孔基序,以促进二聚化。例如,TALEN可能包含在FokI的二聚化结构域中的旋钮入孔基序。

[0619] *修饰的核酸酶*

[0620] 在某些实施方案中,核酸酶任选地由其野生型配对物进行修饰。在一些实施方案中,核酸酶与至少一个异源蛋白质结构域融合。至少一个蛋白质结构域位于核酸酶的N末端、C末端处或内部位置中。在一些实施方案中,两个或多个异源蛋白质结构域在核酸酶上的一个或多个位置处。

[0621] 在一些实施方案中,蛋白质结构域可以促进核酸酶转运到细胞的核内。例如,蛋白质结构域是核定位信号(NLS)。在一些实施方案中,核酸酶与1-10个NLS融合。在一些实施方案中,核酸酶与1-5个NLS融合。在一些实施方案中,核酸酶与一个NLS融合。在其它实施方案中,核酸酶与多于一个NLS融合。在一些实施方案中,核酸酶与2、3、4或5个NLS融合。在一些实施方案中,核酸酶与2个NLS融合。在一些实施方案中,核酸酶与3个NLS融合。在一些实施方案中,核酸酶不与NLS融合。在一些实施方案中,NLS可以是单分型序列(monopartite sequence),例如SV40 NLS、PKKKRKV(SEQ ID NO:56)或PKKKRRV(SEQ ID NO:57)。在一些实施方案中,NLS是二分型序列(bipartite sequence),例如核纤溶酶的NLS,KRPAATKKAGQAKKKK(SEQ ID NO:58)。在一些实施方案中,NLS由其野生型配对物进行遗传修饰。

[0622] 在一些实施方案中,蛋白质结构域能够修饰核酸酶的细胞内半衰期。在一些实施方案中,核酸酶的半衰期可以是增加的。在一些实施方案中,核酸酶的半衰期是降低的。在一些实施方案中,该实体能够增加核酸酶的稳定性。在一些实施方案中,该实体能够降低核酸酶的稳定性。在一些实施方案中,蛋白质结构域充当蛋白质降解的信号肽。在一些实施方案中,蛋白质降解由蛋白水解酶介导,所述蛋白水解酶例如蛋白酶体、溶酶体蛋白酶或钙蛋白酶。在一些实施方案中,蛋白质结构域包含PEST序列。在一些实施方案中,通过添加泛素或多聚泛素链来修饰核酸酶。在一些实施方案中,泛素是泛素样蛋白(UBL)。泛素样蛋白的非限制性实例包括小泛素样修饰物(SUMO)、泛素交叉反应蛋白(UCRP,也称为干扰素刺激基因15(ISG15))、泛素相关修饰物1(URM1)、神经元前体细胞表达发育下调蛋白8(NEDD8,在酿酒酵母(S. cerevisiae)中也称为Rub 1)、人白细胞抗原F相关的(FAT10)、自噬8(ATG8)和-12(ATG12)、Fau泛素样蛋白(FUB1)、膜锚定UBL(MUB)、泛素折叠修饰物1(UFM1)和泛素样蛋白5(UBLS)。

[0623] 在一些实施方案中,蛋白质结构域是标记物结构域。标记物结构域的非限制性实例包括荧光蛋白、纯化标签、表位标签和报道基因序列。在一些实施方案中,标记物结构域是荧光蛋白。合适的荧光蛋白的非限制性实例包括绿色荧光蛋白(例如,GFP、GFP-2、tagGFP、turboGFP、sfGFP、EGFP、Emerald、Azami Green、Monomeric Azami Green、CopGFP、AceGFP、ZsGreenl)、黄色荧光蛋白(例如,YFP、EYFP、Citrine、Venus、YPet、PhiYFP、ZsYellowl)、蓝色荧光蛋白(例如,EBFP、EBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-

sapphire)、青色荧光蛋白(例如ECFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan)、红色荧光蛋白(例如,mKate、mKate2、mPlum、DsRed monomer、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-Monomer、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRasberry、mStrawberry、Jred)、以及橙色荧光蛋白(mOrange、mKO、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine、tdTomato)、或任何其它合适的荧光蛋白。在其它实施方案中,标记物结构域是纯化标签和/或表位标签。非限制性的示例性标签包括谷胱甘肽-S-转移酶(GST)、几丁质结合蛋白(CBP)、麦芽糖结合蛋白(MBP)、硫氧还蛋白(TRX)、聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、HA、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、6×His(HHHHHH;SEQ ID NO:60)、生物素羧基载体蛋白(BCCP)和钙调蛋白。非限制性的示例性报道基因包括谷胱甘肽-S-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰基转移酶(CAT)、β-半乳糖苷酶、β-葡糖醛酸酶、萤光素酶或荧光蛋白。

[0624] 在另外的实施方案中,蛋白质结构域可以将核酸酶靶向特异性细胞器、细胞类型、组织或器官。

[0625] 在进一步的实施方案中,蛋白质结构域是效应结构域。当核酸酶被导向其靶核酸时,例如,当Cas9蛋白通过引导RNA导向靶核酸时,效应结构域可以修饰或影响靶核酸。在一些实施方案中,效应结构域选自核酸结合结构域、核酸酶结构域、表观遗传修饰结构域、转录激活结构域或转录阻遏物结构域。

[0626] 本发明的某些实施方案还提供了在载体上提供的编码本文所述的核酸酶(例如,Cas9蛋白)的核酸。在一些实施方案中,核酸是DNA分子。在其它实施方案中,核酸是RNA分子。在一些实施方案中,编码核酸酶的核酸是mRNA分子。在某些实施方案中,核酸是编码Cas9蛋白的mRNA。

[0627] 在一些实施方案中,编码核酸酶的核酸进行密码子优化,用于在一种或多种真核细胞类型中的有效表达。在一些实施方案中,编码核酸酶的核酸进行密码子优化,用于在一种或多种哺乳动物细胞中的有效表达。在一些实施方案中,编码核酸酶的核酸进行密码子优化,用于在人细胞中的有效表达。密码子优化方法包括密码子使用表和密码子优化算法是本领域可获得的。

[0628] *靶位点*

[0629] 在一些实施方案中,本文所述的定点核酸酶被导向并切割(例如引入DSB)靶核酸分子。在一些实施方案中,Cas核酸酶通过引导RNA导向靶核酸分子(gDNA)的靶位点,在其中所述引导RNA与靶序列的互补链杂交,并且Cas核酸酶在靶位点处切割靶核酸。在一些实施方案中,靶序列的互补链与引导RNA的靶向序列(例如间隔区序列)互补。在一些实施方案中,引导RNA的靶向序列与其靶序列的相应互补链之间的互补程度为约50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%。在一些实施方案中,靶序列的互补链和引导RNA的靶向序列是100%互补的。在其它实施方案中,靶序列的互补链和引导RNA的靶向序列包含至少一个错配。例如,靶序列的互补链和引导RNA的靶向序列包含1、2、3、4、5、6、7、8、9或10个错配。在一些实施方案中,靶序列的互补链和引导RNA的靶向序列含有1-6个错配。在一些实施方案中,靶序列的互补链和引导RNA的靶向序列含有5或6个错配。

[0630] 靶序列的长度可能取决于使用的核酸酶系统。例如,CRISPR/Cas系统的靶序列包

含长度5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50个或多于50个核苷酸。在一些实施方案中,靶序列包含长度18-24个核苷酸。在一些实施方案中,靶序列包含长度19-21个核苷酸。在一些实施方案中,靶序列包含长度20个核苷酸。当使用切口酶时,靶序列包含被DNA分子的相反链上的一对切口酶识别的一对靶序列。

[0631] 在一些实施方案中,大范围核酸酶的靶序列包含长度为12-40个或更多个的核苷酸。当使用ZFN时,靶序列包含由DNA分子的相反链上的一对ZFN识别的两个半靶序列,在其之间具有互连序列。在一些实施方案中,ZFN的每个半靶序列独立地包含长度9、12、15、18个或更多个核苷酸。在一些实施方案中,ZFN的互连序列包含长度4-20个核苷酸。在一些实施方案中,ZFN的互连序列包含长度5-7个核苷酸。

[0632] 当使用TALEN时,靶序列可以类似地包含由DNA分子的相反链上的一对TALEN识别的两个半靶序列,在其之间具有互连序列。在一些实施方案中,TALEN的每个半靶序列可以独立地包含长度10-20个或更多个核苷酸。在一些实施方案中,TALEN的互连序列可以包含长度4-20个核苷酸。在一些实施方案中,TALEN的互连序列可以包含长度12-19个核苷酸。

[0633] 靶核酸分子是细胞内源或外源的任何DNA分子。如本文使用的,术语“内源序列”指细胞天然的序列。在一些实施方案中,靶核酸分子是来自细胞或在细胞中的基因组DNA(gDNA)分子或染色体。在一些实施方案中,靶核酸分子的靶序列是来自细胞或在细胞中的基因组序列。在其它实施方案中,细胞是真核细胞。在一些实施方案中,真核细胞是哺乳动物细胞。在一些实施方案中,真核细胞可以是啮齿类动物细胞。在一些实施方案中,真核细胞可以是人细胞。在进一步的实施方案中,靶序列可以是病毒序列。在再其它实施方案中,靶序列可以是合成序列。在一些实施方案中,靶序列可以在真核染色体如人染色体上。

[0634] 在一些实施方案中,靶序列可以位于基因的编码序列、基因的内含子序列、基因的转录控制序列、基因的翻译控制序列、或基因之间的非编码序列中。在一些实施方案中,基因可以是蛋白质编码基因。在其它实施方案中,基因可以是非编码RNA基因。在一些实施方案中,靶序列可以包含疾病相关基因的全部或一部分。

[0635] 在一些实施方案中,靶序列可以位于基因组中控制染色质组构方面的非基因功能位点中,例如支架位点或基因座控制区。在一些实施方案中,靶序列可以是遗传安全港位点,即,促进安全遗传修饰的基因座。

[0636] 在一些实施方案中,靶序列可以与前间隔序列邻近基序(PAM)相邻,所述PAM是由CRISPR/Cas9复合物识别的短序列。在一些实施方案中,PAM可以与靶序列的3’端的1、2、3或4个核苷酸相邻或在其内。PAM的长度和序列可能取决于使用的Cas9蛋白。例如,PAM可以选自特异性Cas9核酸酶或Cas9直系同源物的共有区或特定PAM序列,包括Ran等人(2015)Nature,520:186-191(2015)的图1中公开的那些,所述参考文献引入本文作为参考。在一些实施方案中,PAM可以包含长度2、3、4、5、6、7、8、9或10个核苷酸。非限制性的示例性PAM序列包括NGG(SpCas9 WT、SpCas9切口酶、二聚体dCas9-Fok1、SpCas9-HF1、SpCas9 K855A、eSpCas9(1.0)、eSpCas9(1.1))、NGAN或NGNG(SpCas9 VQR变体)、NGAG(SpCas9 EQR变体)、NGCG(SpCas9 VRER变体)、NAAG(SpCas9 QQR1变体)、NNGRRT或NNGRRN(SaCas9)、NNNRRT(KKH SaCas9)、NNNNRYAC(CjCas9)、NNAGAAW(St1Cas9)、NAAAAC(TdCas9)、NGGNG(St3Cas9)、NG(FnCas9)、NAAAAN(TdCas9)、NNAAAAW(StCas9)、NNNNACA(CjCas9)、GNNNCNNA(PmCas9)和

NNNNGATT(NmCas9)(参见例如,Cong等人(2013)Science 339:819-823;Kleinstiver等人(2015)Nat Biotechnol 33:1293-1298;Kleinstiver等人(2015)Nature 523:481-485;Kleinstiver等人(2016)Nature 529:490-495; Tsai等人(2014)Nat Biotechnol 32:569-576;Slaymaker等人(2016)Science 351:84-88;Anders等人(2016)Mol Cell 61:895-902;Kim等人(2017)Nat Comm 8:14500;Fonfara等人(2013)Nucleic Acids Res 42:2577-2590;Garneau等人(2010)Nature 468:67-71;Magadan等人(2012)PLoS ONE 7:e40913;Esvelt等人(2013)Nat Methods 10(11):1116-1121(其中N定义为任何核苷酸,W定义为A或T,R定义为嘌呤(A)或(G),且Y定义为嘧啶(C)或(T))。在一些实施方案中,PAM序列是NGG。在一些实施方案中,PAM序列是NGAN。在一些实施方案中,PAM序列是NGNG。在一些实施方案中,PAM是NNGRRT。在一些实施方案中,PAM序列是NGGNG。在一些实施方案中,PAM序列可以是NNAAAAW。

[0637] 用于基因组编辑的系统

[0638] 在一些方面,本公开内容提供了用于校正基因组DNA分子中的突变的系统。在一些实施方案中,该系统包含定点核酸酶,任选地gRNA,以及供体多核苷酸,例如本文所述的那些。在本公开内容的一些实施方案中,该系统包含改造的核酸酶。在一些实施方案中,该系统包含定点核酸酶。在一些实施方案中,定点核酸酶包含CRISPR/Cas核酸酶系统。在一些实施方案中,Cas核酸酶是Cas9。在一些实施方案中,包含CRISPR/Cas系统的引导RNA是sgRNA。

[0639] 修饰的供体多核苷酸

[0640] 在一些实施方案中,提供了具有适合于在细胞内的递送和稳定性的化学的供体多核苷酸。此外,在一些实施方案中,提供了可用于控制本文所述的供体多核苷酸的药代动力学、生物分布、生物利用度和/或功效的化学。相应地,在一些实施方案中,本文所述的供体多核苷酸可以是修饰的,例如包含修饰的糖部分、修饰的核苷间键合、修饰的核苷、修饰的核苷酸和/或其组合。另外,修饰的供体多核苷酸可以显示出下述中特性的一种或多种:并非免疫刺激性的;是核酸酶抗性的;与未修饰的供体多核苷酸相比,具有改善的细胞摄取;和/或对细胞或哺乳动物是无毒的。

[0641] 核苷酸和核苷修饰已显示使得它们掺入其内的多核苷酸(例如供体多核苷酸)比天然多核苷酸对核酸酶消化更具抗性,并且这些修饰的多核苷酸已显示比未修饰的多核苷酸完整地存活更长时间。修饰的寡核苷酸的具体实例包括包含修饰的主链(即修饰的核苷间键合)的那些寡核苷酸,所述修饰的主链例如硫代磷酸酯、磷酸三酯、甲基膦酸酯、短链烷基或环烷基糖间键合、或者短链杂原子或杂环糖间键合。在一些实施方案中,寡核苷酸可以具有硫代磷酸酯主链;杂原子主链,例如亚甲基(甲基亚氨基)或MMI主链;酰胺主链(参见例如,De Mesmaeker等人Ace. Chem. Res. 1995,28:366- 374);吗啉代主链(参见Summerton和Weller,美国专利号5,034,506);或肽核酸(PNA)主链(其中多核苷酸的磷酸二酯主链由聚酰胺主链替换,核苷酸直接或间接结合聚酰胺主链的氮杂氮原子,参见Nielsen等人Science 1991,254,1497)。含磷的修饰键合包括但不限于具有正常的3'-5'键合的硫代磷酸酯、手性硫代磷酸酯、二硫代磷酸酯、磷酸三酯、氨基烷基磷酸三酯、包含3'亚烷基膦酸酯和手性膦酸酯的甲基和其它烷基膦酸酯、次膦酸酯、包含3'-氨基氨基磷酸酯和氨基烷基氨基磷酸酯的氨基磷酸酯、硫羰氨基磷酸酯、硫羰烷基膦酸酯、硫羰烷基磷酸三酯和硼烷磷酸酯,这些的2'-5'连接的类似物以及具有反转极性的那些,其中相邻的核苷单元对将3'-5'

连接至5'-3'或将2'-5'连接至5'-2';参见美国专利号3,687,808;4,469,863;4,476,301 ;5,023,243;5,177,196;5,188,897;5,264,423;5,276,019;5,278,302;5,286,717;5,321,131 ;5,399,676;5,405,939;5,453,496;5,455,233;5,466,677;5031272.1 5,476,925;5,519,126;5,536,821 ;5,541,306;5,550,111 ;5,563,253;5,571,799;5,587,361 ;以及5,625,050。

[0642] 基于吗啉代的寡聚化合物在以下中进行描述:Dwaine A. Braasch和David R. Corey,Biochemistry,2002,41(14),4503-4510);Genesis,第30卷,第3期,2001 ;Heasman,J.,Dev. Biol.,2002,243,209-214;Nasevicius等人Nat. Genet.,2000,26,216- 220;Lacerra等人Proc. Natl. Acad. Sci.,2000,97,9591-9596;以及于1991年7月23日授权的美国专利号5,034,506。在一些实施方案中,基于吗啉代的寡聚化合物是二氨基磷酸酯吗啉代寡聚物(PMO)(例如,如Iverson,Curr. Opin. Mol. Ther.,3:235-238,2001 ;以及Wang等人J. Gene Med.,12:354-364,2010中所述)。

[0643] 环已烯基核酸寡核苷酸模拟物在Wang等人J. Am. Chem. Soc,2000,122,8595-8602中进行描述。

[0644] 其中不包括磷原子的修饰的寡核苷酸主链具有由短链烷基或环烷基核苷间键合、混合的杂原子和烷基或环烷基核苷间键合、或者一个或多个短链杂原子或杂环核苷间键合形成的主链。这些包括具有以下的那些:吗啉代键合(部分由核苷的糖部分形成);硅氧烷主链;硫化物、亚砜和砜主链;甲酰基和硫代甲酰基主链;亚甲基甲酰基和硫代甲酰基主链;含烯烃的主链;氨基磺酸酯主链;亚甲基亚氨基和亚甲基肼基主链;磺酸酯和磺酰胺主链;酰胺主链;以及具有混合的N、O、S和CH2组分部分的其它;参见美国专利号5,034,506;5,166,315;5,185,444;5,214,134;5,216,141 ;5,235,033;5,264,562;5,264,564;5,405,938;5,434,257;5,466,677;5,470,967;5,489,677;5,541,307;5,561,225;5,596,086;5,602,240;5,610,289;5,602,240;5,608,046;5,610,289;5,618,704;5,623,070;5,663,312;5,633,360;5,677,437;以及5,677,439,所述专利各自引入本文作为参考。

[0645] 在一些实施方案中,例如通过掺入修饰(例如核苷酸修饰),使本公开内容的供体多核苷酸针对溶核降解得到稳定。在一些实施方案中,本公开内容的供体多核苷酸包括在核苷酸序列的5'和/或3'端处的至少第一、第二和/或第三核苷酸间键合处的硫代磷酸酯。在一些实施方案中,本公开内容的供体多核苷酸包括一个或多个2'-修饰的核苷酸,例如2'-脱氧-2'-氟、2'-O-甲基、2'-O-甲氧基乙基(2'-O-MOE)、2'-O-氨基丙基(2'-O-AP)、2'-O-二甲基氨基乙基(2'-O-DMAOE)、2'-O-二甲基氨基丙基(2'-O-DMAP)、2'-O-二甲基氨基乙氧基乙基(2'-O-DMAEOE)或2'-ON-甲基乙酰氨基(2'-O-NMA)。在一些实施方案中,本公开内容的供体多核苷酸包括如本文所述的硫代磷酸酯和2'-修饰的核苷酸。

[0646] 本文所述的任何修饰的化学可以彼此组合,并且在同一分子内可以包括一种、两种、三种、四种、五种或更多种不同类型的修饰。在一些实施方案中,供体多核苷酸包含1、2、3、4、5、6、7、8、9、10种或修饰。

[0647] mRNA组分

[0648] 在一些实施方案中,由本公开内容提供的系统包含由mRNA编码的改造的核酸酶。在一些实施方案中,由本公开内容提供的组合物包含核酸酶系统,其中构成核酸酶系统的核酸酶由mRNA编码。在一些实施方案中,mRNA可以是天然存在或非天然存在的mRNA。在一些

实施方案中,mRNA可以包括一个或多个修饰的核碱基、核苷或核苷酸,如下所述,在这种情况下,它可以被称为“修饰的mRNA”。在一些实施方案中,mRNA可以包括5’非翻译区(5’-UTR)、3’非翻译区(3’-UTR)、和/或编码区(例如开放读码框)。mRNA可以包括任何合适数目的碱基对,包括数十个(例如10、20、30、40、50、60、70、80、90或100)、数百个(例如200、300、400、500、600、700、800或900)、或数千个(例如1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000)碱基对。任何数目(例如,全部、一些或无一)的核碱基、核苷或核苷酸可以是规范种类的类似物、取代的、修饰的或以其它方式非天然存在的。在某些实施方案中,所有特定的核碱基类型都可以是修饰的。在一些实施方案中,如本文所述的mRNA可以包括5’帽结构、链终止核苷酸、任选地Kozak或Kozak样序列(也称为Kozak共有序列)、茎环、多聚A序列和/或多聚腺苷酸化信号。

[0649] 5’帽结构或帽种类是包括通过接头连接的两个核苷部分的化合物,并且可以选自天然存在的帽、非天然存在的帽或帽类似物、或抗反向帽类似物(ARCA)。帽种类可以包括一个或多个修饰的核苷和/或接头部分。例如,天然mRNA帽可以包括鸟嘌呤核苷酸、以及通过在其5’位置处的三磷酸键合连接的在7位置处甲基化的鸟嘌呤(G)核苷酸,例如m^7G(5’)ppp(5’)G,通常写作m^7GpppG。帽种类也可以是抗反向帽类似物。可能的帽种类的非限制性列表包括m^7GpppG、m^7Gpppm^{7}G、$m^7$3’dGpppG、$m_2^{7,03'}$GpppG、$m_2^{7,03'}$GppppG、$m_2^{7,02'}$GppppG、m^7Gpppm^{7}G、$m^7$3’dGpppG、$m_2^{7,03'}$GpppG、$m_2^{7,03'}$GppppG和$m_2^{7,02'}$GppppG。

[0650] mRNA可以代替或另外地包括链终止核苷。例如,链终止核苷可以包括在其糖基的2’和/或3’位置处脱氧的那些核苷。这些种类可以包括3’-脱氧腺苷(虫草素)、3’-脱氧尿苷、3’-脱氧胞嘧啶、3’-脱氧鸟苷、3’-脱氧胸腺嘧啶,以及2’,3’-双脱氧核苷例如2’,3’-双脱氧腺苷、2’,3’-双脱氧尿苷、2’,3’-双脱氧胞嘧啶、2’,3’-双脱氧鸟苷和2’,3’-双脱氧胸腺嘧啶。在一些实施方案中,如例如国际专利公开号WO 2013/103659中所述,例如在3’末端处将链终止核苷酸掺入mRNA内可以导致mRNA的稳定。

[0651] mRNA可以代替或另外地包括茎环,例如组蛋白茎环。茎环可以包含2、3、4、5、6、7、8个或更多个核苷酸碱基对。例如,茎环可以包括4、5、6、7或8个核苷酸碱基对。茎环可以位于mRNA的任何区域中。例如,茎环可以位于非翻译区(5’非翻译区或3’非翻译区)、编码区、或者多聚A序列或尾部之中、之前或之后。在一些实施方案中,茎环可以影响mRNA的一种或多种功能,例如翻译的起始、翻译效率和/或转录终止。

[0652] mRNA可以代替或另外地包括多聚A序列和/或多聚腺苷酸化信号。多聚A序列可以完全或主要由腺嘌呤核苷酸或其类似物或衍生物组成。多聚A序列可以是定位与mRNA的3’非翻译区相邻的尾部。在一些实施方案中,多聚A序列可以影响mRNA的核输出、翻译和/或稳定性。

[0653] 修饰的RNA

[0654] 在一些实施方案中,本公开内容的RNA(例如gRNA或mRNA)包含一个或多个修饰的核碱基、核苷、核苷酸或核苷间键合。在一些实施方案中,修饰的mRNA和/或gRNA可以具有有用的特性,包括与参考的未修饰的mRNA和/或gRNA相比,增强的稳定性、细胞内保留、增强的翻译、和/或mRNA和/或gRNA引入其内的细胞的先天性免疫应答的显著诱导的缺乏。因此,修饰的mRNA和/或gRNA的使用可以增强蛋白质生产的效率、核酸的细胞内保留以及具有降低的免疫原性。

[0655] 在一些实施方案中,mRNA和/或gRNA包括一个或多个(例如1、2、3或4个)不同的修饰的核碱基、核苷、核苷酸或核苷间键合。在一些实施方案中,mRNA和/或gRNA包括一个或多个(例如1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100个或更多个)不同的修饰的核碱基、核苷或核苷酸。在一些实施方案中,相对于相应的未修饰的gRNA,修饰的gRNA在gRNA引入其内的细胞中具有降低的降解。在一些实施方案中,相对于相应的未修饰的mRNA,修饰的mRNA在mRNA引入其内的细胞中具有降低的降解。

[0656] 在一些实施方案中,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(s^2U)、4-硫代尿苷(s^4U)、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-碘尿苷或5-溴尿苷)、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^5s^2U)、5-氨基甲基-2-硫代尿苷(nm^5s^2U)、5-甲基氨基甲基尿苷(mnm^5U)、5-甲基氨基甲基-2-硫代尿苷(mnm^5s^2U)、5-甲基氨基甲基-2-硒代尿苷(mnm^5se^2U)、5-氨基甲酰基甲基尿苷(ncm^5U)、5-羧甲基氨基甲基尿苷($cmnm^5U$)、5-羧甲基氨基甲基-2-硫代尿苷($cmnm^5s^2U$)、5-丙炔基尿苷、1-丙炔基假尿苷、5-牛磺酰甲基尿苷(τm^5U)、1-牛磺酰甲基假尿苷、5-牛磺酰甲基-2-硫代尿苷(τm^5s^2U)、1-牛磺酸甲基-4-硫代假尿苷、5-甲基尿苷(m^5U,即具有核碱基脱氧胸腺嘧啶)、1-甲基假尿苷($m^1\psi$)、5-甲基-2-硫代尿苷(m^5s^2U)、1-甲基-4-硫代假尿苷($m^1s^4\psi$)、4-硫代-1-甲基假尿苷、3-甲基假尿苷($m^3\psi$)、2-硫代-1-甲基假尿苷、1-甲基-1-脱氮杂假尿苷、2-硫代-1-甲基-1-脱氮杂假尿苷、二氢尿苷(D)、二氢假尿苷、5,6-二氢尿苷、5-甲基-二氢尿苷(m^5D)、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^3U)、1-甲基-3-(3-氨基-3-羧丙基)假尿苷($acp^3\psi$)、5-(异戊烯基氨基甲基)尿苷(inm^5U)、5-(异戊烯基氨基甲基)-2-硫代尿苷(inm^5s^2U)、α-硫代尿苷、2'-O-甲基尿苷(Um)、5,2'-O-二甲基尿苷(m^5Um)、2'-O-甲基假尿苷(ψm)、2-硫代-2'-O-甲基尿苷(s^2Um)、5-甲氧羰基甲基-2'-O-甲基尿苷(mcm^5Um)、5-氨基甲酰基甲基-2'-O-甲基尿苷(ncm^5Um)、5-羧甲基氨基甲基-2'-O-甲基尿苷($cmnm^5Um$)、3,2'-O-二甲基尿苷(m^3Um)和5-(异戊烯基氨基甲基)-2'-O-甲基尿苷(inm^5Um)、1-硫代尿苷、脱氧胸苷、2'-F-ara-尿苷、2'-F-尿苷、2'-OH-ara-尿苷、5-(2-羰基甲氧基乙烯基)尿苷和5-[3-(1-E-丙烯基氨基)]尿苷。

[0657] 在一些实施方案中,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5-氮杂胞苷、6-氮杂胞苷、假异胞苷、3-甲基胞苷(m^3C)、N4-乙酰基胞苷(ac^4C)、5-甲酰基胞苷(f^5C)、N4-甲基胞苷(m^4C)、5-甲基胞苷(m^5C)、5-卤代胞苷(例如5-碘胞苷)、5-羟甲基胞苷(hm^5C)、1-甲基-假异胞苷、吡咯并胞苷、吡咯并假异胞苷、2-硫代胞苷(s^2C)、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-硫基-1-甲基-1-脱氮杂假异胞苷、1-甲基-1-脱氮杂假异胞苷、泽布拉林、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、赖胞苷(k_2C)、α-硫代胞苷、2'-O-甲基胞苷(Cm)、5,2'-O-二甲基胞苷(m^5Cm)、N4-乙酰基-2'-O-甲基胞苷(ac^4Cm)、N4,2'-O-二甲基胞苷

(m^4Cm)、5-甲酰基-2'-O-甲基胞苷(f^5Cm)、N4,N4,2'-O-三甲基胞苷(m^4_2Cm)、1-硫代胞苷、2'-F-ara-胞苷、2'-F-胞苷和2'-OH-ara-胞苷。

[0658] 在一些实施方案中,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α-硫代腺苷、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-氯嘌呤)、6-卤代嘌呤(例如6-氯嘌呤)、2-氨基-6-甲基嘌呤、8-叠氮基腺苷、7-脱氮杂腺嘌呤、7-脱氮杂-8-氮杂腺嘌呤、7-脱氮杂-2-氨基嘌呤、7-脱氮杂-8-氮杂-2-氨基嘌呤、7-脱氮杂-2,6-二氨基嘌呤、7-脱氮杂-8-氮杂-2,6-二氨基嘌呤、1-甲基腺苷(m^1A)、2-甲基腺嘌呤(m^2A)、N6-甲基腺苷(m^6A)、2-甲硫基-N6-甲基腺苷(ms^2m^6A)、N6-异戊烯基-腺苷(i^6A)、2-甲硫基-N6-异戊烯基腺苷(ms^2i^6A)、N6-(顺式-羟基异戊烯基)腺苷(io^6A)、2-甲硫基-N6-(顺式-羟基异戊烯基)腺苷(ms^2io^6A)、N6-甘氨酰氨基甲酰基腺苷(g^6A)、N6-苏氨酰氨基甲酰基腺苷(t^6A)、N6-甲基-N6-苏氨酰氨基甲酰基腺苷(m^6t^6A)、2-甲硫基-N6-苏氨酰氨基甲酰腺苷(ms^2g^6A)、N6,N6-二甲基腺苷(m^6_2A)、N6-羟基正缬氨酰氨基甲酰基腺苷(hn^6A)、2-甲硫基-N6-羟基正缬氨酰氨基甲酰基腺苷(ms^2hn^6A)、N6-乙酰基腺苷(ac^6A)、7-甲基腺嘌呤、2-甲硫基腺嘌呤、2-甲氧基-腺嘌呤、α-硫代腺苷、2'-O-甲基腺苷(Am)、N6,2'-O-二甲基腺苷(m^6Am)、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-ara-腺苷、2'-F-腺苷、2'-OH-ara-腺苷和N6-(19-氨基-五氧杂壬癸基)腺苷。

[0659] 在一些实施方案中,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α-硫代鸟苷、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)、改性不足的羟基怀丁苷(OhyW*)、7-脱氮杂鸟苷、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-脱氮杂鸟苷($preQ_0$)、7-氨基甲基-7-脱氮杂鸟苷($preQ_1$)、古嘌苷(G^+)、7-脱氮杂-8-氮杂鸟苷、6-硫代鸟苷、6-硫代-7-脱氮杂鸟苷、6-硫代-7-脱氮杂-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^1G)、N2-甲基鸟苷(m^2G)、N2,N2-二甲基鸟苷(m^2_2G)、N2,7-二甲基鸟苷($m^{2,7}G$)、N2,N2,7-二甲基鸟苷($m^{2,2,7}G$)、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^1Gm)、N2,7-二甲基-2'-O-甲基鸟苷($m^{2,7}Gm$)、2'-O-甲基-肌苷(Im)、1,2'-O-二甲基肌苷(m^1Im)、2'-O-核糖基鸟苷(磷酸盐)(Gr(p))、1-硫代鸟苷、O6-甲基鸟苷、2'-F-ara-鸟苷和2'-F-鸟苷。

[0660] 在一些实施方案中,本公开内容的mRNA和/或gRNA包括前述修饰的核碱基中的一种或多种的组合(例如,前述修饰的核碱基中的2、3或4种的组合)。

[0661] 在一些实施方案中,修饰的核碱基是假尿苷(ψ)、N1-甲基假尿苷($m^1ψ$)、2-硫代尿苷、4'-硫代尿苷、5-甲基胞嘧啶、2-硫代-1-甲基-1-脱氮杂假尿苷、2-硫代-1-甲基假尿苷、2-硫代-5-氮杂尿苷、2-硫代-二氢假尿苷、2-硫代-二氢尿苷、2-硫代假尿苷、4-甲氧基-2-硫代假尿苷、4-甲氧基假尿苷、4-硫代-1-甲基假尿苷、4-硫代假尿苷、5-氮杂尿苷、二氢假尿苷、5-甲氧基尿苷或2'-O-甲基尿苷。在一些实施方案中,本公开内容的mRNA包括前述修饰的核碱基中的一种或多种的组合(例如,前述修饰的核碱基中的2、3或4种的组合)。在一

些实施方案中,修饰的核碱基是N1-甲基假尿苷($m^1\psi$),并且本公开内容的mRNA由N1-甲基假尿苷($m^1\psi$)完全修饰。在一些实施方案中,N1-甲基假尿苷($m^1\psi$)代表mRNA中75-100%的尿嘧啶。在一些实施方案中,N1-甲基假尿苷($m^1\psi$)代表mRNA中100%的尿嘧啶。

[0662] 在一些实施方案中,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N4-乙酰基胞苷(ac^4C)、5-甲基胞苷(m^5C)、5-卤代胞苷(例如5-碘胞苷)、5-羟甲基胞苷(hm^5C)、1-甲基-假异胞苷、2-硫代胞苷(s^2C)、2-硫代-5-甲基胞苷。在一些实施方案中,本公开内容的mRNA包括前述修饰的核碱基中的一种或多种的组合(例如,前述修饰的核碱基中的2、3或4种的组合)。

[0663] 在一些实施方案中,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7-脱氮杂腺嘌呤、1-甲基腺苷(m^1A)、2-甲基腺嘌呤(m^2A)、N6-甲基腺苷(m^6A)。在一些实施方案中,本公开内容的mRNA包括前述修饰的核碱基中的一种或多种的组合(例如,前述修饰的核碱基中的2、3或4种的组合)。

[0664] 在一些实施方案中,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肌苷(I)、1-甲基肌苷(m^1I)、怀俄苷(imG)、甲基怀俄苷(mimG)、7-脱氮杂鸟苷、7-氰基-7-脱氮杂鸟苷($preQ_0$)、7-氨基甲基-7-脱氮杂鸟苷($preQ_1$)、7-甲基鸟苷(m^7G)、1-甲基鸟苷(m^1G)、8-氧代鸟苷、7-甲基-8-氧代鸟苷。在一些实施方案中,本公开内容的mRNA包括前述修饰的核碱基中的一种或多种的组合(例如,前述修饰的核碱基中的2、3或4种的组合)。

[0665] 在一些实施方案中,修饰的核碱基是1-甲基假尿苷($m^1\psi$)、5-甲氧基尿苷(mo^5U)、5-甲基胞苷(m^5C)、假尿苷(ψ)、α-硫代鸟苷或α-硫代腺苷。在一些实施方案中,本公开内容的mRNA包括前述修饰的核碱基中的一种或多种的组合(例如,前述修饰的核碱基中的2、3或4种的组合)。

[0666] 在某些实施方案中,本公开内容的mRNA和/或gRNA就特定修饰进行均匀修饰(即,完全修饰、贯穿整个序列修饰)。例如,mRNA可以由N1-甲基假尿苷($m^1\psi$)或5-甲基胞苷(m^5C)均匀修饰,这意味着mRNA序列中的所有尿苷或所有胞嘧啶核苷都由N1-甲基假尿苷($m^1\psi$)或5-甲基胞苷(m^5C)替换。类似地,本公开内容的mRNA可以通过由修饰的残基(例如上述那些)替换,就序列中存在的任何类型的核苷残基进行均匀修饰。

[0667] 在一些实施方案中,本公开内容的mRNA可以在编码区(例如,编码多肽的开放读码框)中进行修饰。在其它实施方案中,mRNA可以在除编码区之外的区域中进行修饰。例如,在一些实施方案中,提供了5'-UTR和/或3'-UTR,其中任一或两者可以独立地含有一种或多种不同的核苷修饰。在此类实施方案中,核苷修饰也可以存在于编码区中。

[0668] 核糖核蛋白

[0669] 在某些方面,定点多肽(例如Cas核酸酶)和靶向基因组的核酸(例如gRNA或sgRNA)可以各自分开地施用于细胞或受试者。在某些方面,定点多肽可以与一种或多种引导RNA、或者一种或多种sgRNA预复合。此类预复合材料称为核糖核蛋白颗粒(RNP)。在一些实施方案中,核酸酶系统包含核糖核蛋白(RNP)。在一些实施方案中,核酸酶系统包含Cas9 RNP,其包含与gRNA复合的纯化的Cas9蛋白。Cas9蛋白可以通过本领域已知的任何手段进行表达且纯化。核糖核蛋白在体外组装,并且可以使用本领域已知的标准电穿孔或转染技术直接递送至细胞。

[0670] 载体

[0671] 在一些实施方案中,定点核酸酶(例如,Cas核酸酶)和供体多核苷酸可以由一种或多种载体提供。在一些实施方案中,载体可以是DNA载体。在一些实施方案中,载体可以是环状的。在其它实施方案中,载体可以是线性的。非限制性的示例性载体包括质粒、噬菌粒、粘粒、人工染色体、微型染色体、转座子、病毒载体和表达载体。

[0672] 在一些实施方案中,载体可以是病毒载体。在一些实施方案中,病毒载体可以由其野生型配对物进行遗传修饰。例如,病毒载体可以包含一个或多个核苷酸的插入、缺失或取代,以促进克隆或使得载体的一种或多种特性改变。此类特性可以包括包装容量、转导效率、免疫原性、基因组整合、复制、转录和翻译。在一些实施方案中,病毒基因组的一部分可以被缺失,使得病毒能够包装具有更大尺寸的外源序列。在一些实施方案中,病毒载体可以具有增强的转导效率。在一些实施方案中,可以降低在宿主中通过病毒诱导的免疫应答。在一些实施方案中,促进病毒序列整合到宿主基因组内的病毒基因(例如整合酶)可以被突变,使得病毒变得非整合。在一些实施方案中,病毒载体可以是复制缺陷型的。在一些实施方案中,病毒载体可以包含外源转录或翻译控制序列,以驱动载体上的编码序列表达。在一些实施方案中,病毒可以是辅助病毒依赖型的。例如,病毒可能需要一种或多种辅助病毒,以供应将载体扩增且包装到病毒颗粒内所需的病毒组分(例如病毒蛋白)。在这种情况下,一种或多种辅助组分,包括编码病毒组分的一种或多种载体,可以连同本文所述的载体系统一起引入宿主细胞内。在其它实施方案中,病毒可以是无辅助病毒的。例如,病毒可能能够无需任何辅助病毒而扩增且包装载体。在一些实施方案中,本文所述的载体系统还可以编码病毒扩增和包装所需的病毒组分。

[0673] 非限制性的示例性病毒载体包括腺伴随病毒(AAV)载体、慢病毒载体、腺病毒载体、单纯疱疹病毒(HSV-1)载体、细菌噬菌体T4、杆状病毒载体和逆转录病毒载体。在一些实施方案中,病毒载体可以是AAV载体。在其它实施方案中,病毒载体可以是慢病毒载体。在一些实施方案中,慢病毒可以是非整合的。在一些实施方案中,病毒载体可以是腺病毒载体。在一些实施方案中,腺病毒可以是高克隆容量或“无肠”腺病毒,其中除去5’和3’反向末端重复(ITR)和包装信号(Ψ)的所有编码病毒区域都从病毒中缺失,以增加其包装容量。在再其它实施方案中,病毒载体可以是HSV-1载体。在一些实施方案中,基于HSV-1的载体是辅助病毒依赖型的,而在其它实施方案中,它是辅助病毒非依赖型的。例如,仅保留包装序列的扩增子载体需要具有用于包装的结构成分的辅助病毒,而30 kb缺失型HSV-1载体(其去除非必需病毒功能)不需要辅助病毒。在另外的实施方案中,病毒载体可以是细菌噬菌体T4。在一些实施方案中,当病毒的头部被掏空时,细菌噬菌体T4可能能够包装任何线性或环状DNA或RNA分子。在进一步的实施方案中,病毒载体可以是杆状病毒载体。在再进一步的实施方案中,病毒载体可以是逆转录病毒载体。在使用具有较小克隆容量的AAV或慢病毒载体的实施方案中,可能有必要使用多于一种载体,以递送如本文公开的载体系统的所有组分。例如,一种AAV载体可以含有编码Cas9蛋白的序列,而第二AAV载体可以含有一个或多个引导序列和供体多核苷酸的一个或多个拷贝。

[0674] 在某些实施方案中,可以将病毒载体修饰为靶向特定的组织或细胞类型。例如,可以改变病毒表面蛋白,以减少或消除病毒蛋白与其天然细胞表面受体的结合。还可以将表面蛋白改造为与所需细胞类型特异性的受体相互作用。病毒载体可能具有改变的宿主嗜

性,包括有限或重定向的嗜性。某些改造的病毒载体例如在WO2011130749 [HSV]、WO2015009952 [HSV]、美国专利号5,817,491 [逆转录病毒]、WO2014135998 [T4]和WO2011125054 [T4]中进行描述。在一些实施方案中,载体可能能够驱动一种或多种编码序列在细胞中的表达。在一些实施方案中,细胞可以是真核细胞,例如,酵母、植物、昆虫或哺乳动物细胞。在一些实施方案中,真核细胞可以是哺乳动物细胞。在一些实施方案中,真核细胞可以是啮齿类动物细胞。在一些实施方案中,真核细胞可以是人细胞。在不同类型的细胞中驱动表达的合适启动子是本领域已知的。在一些实施方案中,启动子可以是野生型的。在其它实施方案中,可以修饰启动子用于更有效或高效的表达。在再其它实施方案中,启动子可以是截短的,但仍保留其功能。例如,启动子可以具有适合于将载体适当包装到病毒内的正常大小或减小的大小。

[0675] 在一些实施方案中,载体可以包含编码本文描述的核酸酶的核苷酸序列。在一些实施方案中,载体系统可以包含编码核酸酶的核苷酸序列的一个拷贝。在其它实施方案中,载体系统可以包含编码核酸酶的核苷酸序列的多于一个拷贝。在一些实施方案中,编码核酸酶的核苷酸序列可以与至少一种转录或翻译控制序列可操作地连接。在一些实施方案中,编码核酸酶的核苷酸序列可以与至少一个启动子可操作地连接。在一些实施方案中,编码核酸酶的核苷酸序列可以与至少一种转录或翻译控制序列可操作地连接。

[0676] 在一些实施方案中,启动子可以是组成型、诱导型或组织特异性的。在一些实施方案中,启动子可以是组成型启动子。非限制性的示例性组成型启动子包括巨细胞病毒立即早期启动子(CMV)、猿猴病毒(SV40)启动子、腺病毒主要晚期(MLP)启动子、劳斯肉瘤病毒(RSV)启动子、小鼠乳腺肿瘤病毒(MMTV)启动子、磷酸甘油酸激酶(PGK)启动子、延伸因子-α(EF1α)启动子、泛素启动子、肌动蛋白启动子、微管蛋白启动子、免疫球蛋白启动子、其功能片段或前述任一种的组合。在一些实施方案中,启动子可以是CMV启动子。在一些实施方案中,启动子可以是截短的CMV启动子。在其它实施方案中,启动子可以是EF1α启动子。在一些实施方案中,启动子可以是诱导型启动子。非限制性的示例性诱导型启动子包括可通过热休克、光、化学制品、肽、金属、类固醇、抗生素或乙醇诱导的那些。在一些实施方案中,诱导型启动子可以是具有低基础(非诱导)表达水平的启动子,例如Tet-On® 启动子(Clontech)。在一些实施方案中,启动子可以是组织特异性启动子。在一些实施方案中,组织特异性启动子在肝组织中专有地或占优势地表达。非限制性的示例性组织特异性启动子包括B29启动子、CD14启动子、CD43启动子、CD45启动子、CD68启动子、结蛋白启动子、弹性蛋白酶-1启动子、内皮糖蛋白启动子、纤连蛋白启动子、Flt-1启动子、GFAP启动子、GPIIb启动子、ICAM-2启动子、INF-β启动子、Mb启动子、Nphs1启动子、OG-2启动子、SP-B启动子、SYN1启动子和WASP启动子。

[0677] 在一些实施方案中,由载体编码的核酸酶可以是Cas蛋白,例如Cas9蛋白或Cpf1蛋白。载体系统可以进一步包含载体,其包含编码本文所述的引导RNA的核苷酸序列。在一些实施方案中,载体系统可以包含引导RNA的一个拷贝。在其它实施方案中,载体系统可以包含引导RNA的多于一个拷贝。在具有多于一种引导RNA的实施方案中,引导RNA可以是不等同的,使得它们靶向不同的靶序列,或具有其它不同的特性,例如在Cas9 RNP复合物内的活性或稳定性。在一些实施方案中,编码引导RNA的核苷酸序列可以与至少一种转录或翻译控制序列可操作地连接。在一些实施方案中,编码引导RNA的核苷酸序列可以与至少一个启动子

可操作地连接。在一些实施方案中,启动子可以由RNA聚合酶III(Pol III)识别。Pol III启动子的非限制性实例包括U6、H1和tRNA启动子。在一些实施方案中,编码引导RNA的核苷酸序列可以与小鼠或人U6启动子可操作地连接。在其它实施方案中,编码引导RNA的核苷酸序列可以与小鼠或人H1启动子可操作地连接。在一些实施方案中,编码引导RNA的核苷酸序列可以与小鼠或人tRNA启动子可操作地连接。在具有多于一种引导RNA的实施方案中,用于驱动表达的启动子可以是相同或不同的。在一些实施方案中,编码引导RNA的crRNA的核苷酸和编码引导RNA的tracr RNA的核苷酸可以在同一载体上提供。在一些实施方案中,编码crRNA的核苷酸和编码tracr RNA的核苷酸可以由同一启动子驱动。在一些实施方案中,crRNA和tracr RNA可以被转录成单个转录物。例如,crRNA和tracr RNA可以由单个转录物进行加工,以形成双分子引导RNA。可替代地,crRNA和tracr RNA可以被转录成单分子引导RNA。在其它实施方案中,crRNA和tracr RNA可以由其在同一载体上的相应启动子驱动。在再其它实施方案中,crRNA和tracr RNA可以由不同的载体编码。

[0678] 在一些实施方案中,编码引导RNA的核苷酸序列可以位于包含编码Cas9蛋白的核苷酸序列的相同载体上。在一些实施方案中,引导RNA和Cas9蛋白的表达可以由不同的启动子驱动。在一些实施方案中,引导RNA的表达可以由驱动Cas9蛋白表达的相同启动子驱动。在一些实施方案中,引导RNA和Cas9蛋白转录物可以包含在单个转录物内。例如,引导RNA可以在Cas9蛋白转录物的非翻译区(UTR)内。在一些实施方案中,引导RNA可以在Cas9蛋白转录物的5’UTR内。在其它实施方案中,引导RNA可以在Cas9蛋白转录物的3’UTR内。在一些实施方案中,可以通过在其3’UTR内含有引导RNA,且从而缩短其3’UTR的长度,来降低Cas9蛋白转录物的细胞内半衰期。在另外的实施方案中,引导RNA可以在Cas9蛋白转录物的内含子内。在一些实施方案中,可以在引导RNA位于其内的内含子处添加合适的剪接位点,使得引导RNA从转录物中适当地剪接出来。在一些实施方案中,在同一载体上紧密接近的Cas9蛋白和引导RNA的表达可以促进CRISPR复合物的更有效形成。

[0679] 在一些实施方案中,载体系统可以进一步包括包含本文所述的供体多核苷酸的载体。在一些实施方案中,载体系统可以包含供体多核苷酸的一个拷贝。在其它实施方案中,载体系统可以包含供体多核苷酸的多于一个拷贝。在一些实施方案中,载体系统可以包含供体多核苷酸的2、3、4、5、6、7、8、9、10个或更多个拷贝。供体多核苷酸的多重拷贝可以位于相同或不同的载体上。供体多核苷酸的多重拷贝也可以是彼此相邻的,或者通过其它核苷酸序列或载体元件分开。

[0680] 载体系统可以包含1-3个载体。在一些实施方案中,载体系统可以包含一个单一载体。在其它实施方案中,载体系统可以包含两个载体。在另外的实施方案中,载体系统可以包含三个载体。当不同的引导RNA或供体多核苷酸用于多重化时,或当使用引导RNA或供体多核苷酸的多重拷贝时,载体系统可以包含多于三个载体。

[0681] 在一些实施方案中,编码Cas9蛋白的核苷酸序列、编码引导RNA的核苷酸序列和供体多核苷酸可以位于同一或分开的载体上。在一些实施方案中,所有序列可以位于同一载体上。在一些实施方案中,两个或更多个序列可以位于同一载体上。序列可以在载体上以相同或不同方向和以任何次序定向。在一些实施方案中,编码Cas9蛋白的核苷酸序列和编码引导RNA的核苷酸序列可以位于同一载体上。在一些实施方案中,编码Cas9蛋白和供体多核苷酸的核苷酸序列可以位于同一载体上。在一些实施方案中,编码引导RNA和供体多核苷酸

的核苷酸序列可以位于同一载体上。在一些实施方案中,载体系统可以包括包含编码Cas9蛋白的核苷酸序列的第一载体、以及包含编码引导RNA和供体多核苷酸的核苷酸序列的第二载体。

[0682] 纳米颗粒组合物

[0683] 在一些方面,本公开内容提供了包含本文所述的供体多核苷酸、系统或系统的组分的纳米颗粒组合物(例如,脂质纳米颗粒,LNP)。本公开内容的供体多核苷酸、系统或系统的组分可以在纳米颗粒或其它递送媒介物(例如,聚合纳米颗粒)中个别地或组合在一起配制,以促进细胞摄取和/或保护其在递送至受试者(例如,具有突变的患者)时不被降解。说明性的纳米颗粒在Panyam & Labhasetwar(2003)Adv Drug Deliv Rev 55:329-347和Peer等人(2007)Nature Nanotech. 2:751-760中进行描述。

[0684] 纳米颗粒是大小范围通常为1和100至500纳米(nm)的超细颗粒,具有周围的界面层,并且经常显示出大小有关或大小依赖性特性。纳米颗粒组合物是众多的,并且涵盖脂质纳米颗粒(LNP)、脂质体(例如脂质囊泡)和脂质复合物(lipoplexes)。例如,纳米颗粒组合物可以是具有直径500 nm或更小的脂质双层的脂质体。在一些实施方案中,纳米颗粒组合物是包括一个或多个脂质双层的囊泡。在某些实施方案中,纳米颗粒组合物包括由水性区室分开的两个或更多个同心双层。脂质双层可以是功能化和/或彼此交联的。脂质双层可以包括一种或多种配体、蛋白质或通道。

[0685] 纳米颗粒组合物可以通过各种方法进行表征。例如,显微镜检查(例如,透射电子显微镜检查或扫描电子显微镜检查),可以用于检查纳米颗粒组合物的形态和尺寸分布。动态光散射或电位测定法(例如电位滴定法)可以用于测量ζ电位。动态光散射也可以用于确定粒度。仪器例如Zetasizer Nano ZS(Malvern Instruments Ltd,Malvern,Worcestershire,UK)也可以用于测量纳米颗粒组合物的多重特征,例如粒度、多分散性指数和ζ电位。

[0686] 纳米颗粒的大小可以帮助对抗生物反应,例如但不限于炎症,或者可以增加多核苷酸的生物效应。纳米颗粒的大小也可能改变生物分布、免疫应答和细胞摄取。

[0687] 如本文使用的,在纳米颗粒组合物的上下文中,“大小”或“平均大小”指纳米颗粒组合物的平均直径。

[0688] 在一个实施方案中,编码目的多肽的多核苷酸在具有约10至约100 nm直径的脂质纳米颗粒中进行配制,所述直径例如但不限于约10至约20 nm、约10至约30 nm、约10至约40 nm、约10至约50 nm、约10至约60 nm、约10至约70 nm、约10至约80 nm、约10至约90 nm、约20至约30 nm、约20至约40 nm、约20至约50 nm、约20至约60 nm、约20至约70 nm、约20至约80 nm、约20至约90 nm、约20至约100 nm、约30至约40 nm、约30至约50 nm、约30至约60 nm、约30至约70 nm、约30至约80 nm、约30至约90 nm、约30至约100 nm、约40至约50 nm、约40至约60 nm、约40至约70 nm、约40至约80 nm、约40至约90 nm、约40至约100 nm、约50至约60 nm、约50至约70 nm、约50至约80 nm、约50至约90 nm、约50至约100 nm、约60至约70 nm、约60至约80 nm、约60至约90 nm、约60至约100 nm、约70至约80 nm、约70至约90 nm、约70至约100 nm、约80至约90 nm、约80至约100 nm和/或约90至约100 nm。

[0689] 在一个实施方案中,纳米颗粒具有约10至500 nm的直径。在一个实施方案中,纳米颗粒具有大于100 nm、大于150 nm、大于200 nm、大于250 nm、大于300 nm、大于350 nm、大

于400 nm、大于450 nm、大于500 nm、大于550 nm、大于600 nm、大于650 nm、大于700 nm、大于750 nm、大于800 nm、大于850 nm、大于900 nm、大于950 nm或大于1000 nm的直径。

[0690] 在一些实施方案中,纳米颗粒组合物的最大尺寸为1 μm或更短(例如1 μm、900 nm、800 nm、700 nm、600 nm、500 nm、400 nm、300 nm、200 nm、175 nm、150 nm、125 nm、100 nm、75 nm、50 nm或更短)。

[0691] 纳米颗粒组合物可以是相对均匀的。多分散性指数可以用于指示纳米颗粒组合物的均匀性,例如,纳米颗粒组合物的粒度分布。小(例如,小于0.3)多分散性指数一般指示窄粒度分布。纳米颗粒组合物可以具有约0至约0.25的多分散性指数,例如0.01、0.02、0.03、0.04、0.05、0.06、0.07、0.08、0.09、0.10、0.11、0.12、0.13、0.14、0.15、0.16、0.17、0.18、0.19、0.20、0.21、0.22、0.23、0.24或0.25。在一些实施方案中,本文公开的纳米颗粒组合物的多分散性指数可以是约0.10至约0.20。

[0692] 纳米颗粒组合物的ζ电位可以用于指示该组合物的电动电位。例如,ζ电位可以描述纳米颗粒组合物的表面电荷。一般期望具有相对较低的正电荷或负电荷的纳米颗粒组合物,因为带较高电荷的种类可以与体内的细胞、组织和其它要素不期望地相互作用。在一些实施方案中,本文公开的纳米颗粒组合物的ζ电位可以是约-10 mV至约+20 mV、约-10 mV至约+15 mV、约 10 mV至约+10 mV、约-10 mV至约+5 mV、约-10 mV至约0 mV、约-10 mV至约-5 mV、约-5 mV至约+20 mV、约-5 mV至约+15 mV、约-5 mV至约+10 mV、约-5 mV至约+5 mV、约-5 mV至约0 mV、约0 mV至约+20 mV、约0 mV至约+15 mV、约0 mV至约+10 mV、约0 mV至约+5 mV、约+5 mV至约+20 mV、约+5 mV至约+15 mV、或约+5 mV至约+10 mV。

[0693] 在一些实施方案中,脂质纳米颗粒的ζ电位可以是约0 mV至约100 mV、约0 mV至约90 mV、约0 mV至约80 mV、约0 mV至约70 mV、约0 mV至约60 mV、约0 mV至约50 mV、约0 mV至约40 mV、约0 mV至约30 mV、约0 mV至约20 mV、约0 mV至约10 mV、约10 mV至约100 mV、约10 mV至约90 mV、约10 mV至约80 mV、约10 mV至约70 mV、约10 mV至约60 mV、约10 mV至约50 mV、约10 mV至约40 mV、约10 mV至约30 mV、约10 mV至约20 mV、约20 mV至约100 mV、约20 mV至约90 mV、约20 mV至约80 mV、约20 mV至约70 mV、约20 mV至约60 mV、约20 mV至约50 mV、约20 mV至约40 mV、约20 mV至约30 mV、约30 mV至约100 mV、约30 mV至约90 mV、约30 mV至约80 mV、约30 mV至约70 mV、约30 mV至约60 mV、约30 mV至约50 mV、约30 mV至约40 mV、约40 mV至约100 mV、约40 mV至约90 mV、约40 mV至约80 mV、约40 mV至约70 mV、约40 mV至约60 mV、以及约40 mV至约50 mV。在一些实施方案中,脂质纳米颗粒的ζ电位可以是约10 mV至约50 mV、约15 mV至约45 mV、约20 mV至约40 mV、以及约25 mV至约35 mV。在一些实施方案中,脂质纳米颗粒的ζ电位可以是约10 mV、约20 mV、约30 mV、约40 mV、约50 mV、约60 mV、约70 mV、约80 mV、约90 mV和约100 mV。

[0694] 在一些实施方案中,纳米颗粒的pKa为约5-8。在一些实施方案中,纳米颗粒的pKa为约5。在一些实施方案中,纳米颗粒的pKa为约6。在一些实施方案中,纳米颗粒的pKa为约7。在一些实施方案中,纳米颗粒的pKa为约8。

[0695] 术语多核苷酸的“封装效率”描述相对于所提供的初始量,在制备后由纳米颗粒组合物封装或以其它方式结合的多核苷酸的量。如本文使用的,“封装”可以指完全、基本或部分封闭、约束、围绕或包装。

[0696] 封装效率期望很高(例如,接近于100%)。可以例如通过比较在用一种或多种有机

溶剂或去污剂破碎纳米颗粒组合物之前和之后,在含有纳米颗粒组合物的溶液中的多核苷酸量来测量封装效率。

[0697] 荧光可用于测量溶液中的游离多核苷酸量。对于本文所述的纳米颗粒组合物,多核苷酸的封装效率可以是至少50%,例如50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%。在一些实施方案中,封装效率可以是至少80%。在某些实施方案中,封装效率可以是至少90%。

[0698] 在某些实施方案中,本公开内容的供体多核苷酸或系统被封装在纳米颗粒内。在一些实施方案中,包含单个核苷酸序列的一种或多种供体多核苷酸被封装在纳米颗粒内。在一些实施方案中,包含不同核苷酸序列的一种或多种供体多核苷酸被封装在纳米颗粒内。

[0699] 在一些实施方案中,本公开内容的系统被封装在纳米颗粒内。在一些实施方案中,本公开内容的系统的一种或多种组分被个别地封装在纳米颗粒内。在一些实施方案中,本公开内容的系统的每种组分被个别地封装在纳米颗粒内。在一些实施方案中,gRNA被封装在纳米颗粒内。在一些实施方案中,编码核酸酶(例如,Cas9)的mRNA被封装在纳米颗粒内。在一些实施方案中,gRNA和编码核酸酶(例如,Cas9)的mRNA被封装在纳米颗粒内。在一些实施方案中,gRNA、编码核酸酶(例如,Cas9)的mRNA和供体多核苷酸被封装在纳米颗粒内。

[0700] *脂质纳米颗粒*

[0701] 在特定实施方案中,纳米颗粒包括脂质。脂质纳米颗粒包括但不限于脂质体和胶束。可以存在许多脂质中的任一种,包括阳离子和/或可电离的脂质、阴离子脂质、中性脂质、两亲性脂质、缀合的脂质(例如,PEG化脂质)和/或结构脂质。此类脂质可以单独或组合使用。

[0702] 在一些方面,供体多核苷酸、系统或者系统的一种或多种组分,例如本文所述的那些,包含脂质纳米颗粒(LNP)。本文描述的LNP各自可以用作本文所述的供体多核苷酸、系统或者系统的任何一种或多种组分中的任一种的制剂。

[0703] *阳离子/可电离的脂质*

[0704] 在一些实施方案中,脂质纳米颗粒可以包含阳离子和/或可电离的脂质。如本文使用的,术语"可电离的脂质"具有其在本领域中的普通含义,并且可以指包含一个或多个荷电部分的脂质。在一些实施方案中,可电离的脂质可以是带正电或带负电的。可电离的脂质可以是带正电荷的,在这种情况下,它可以被称为"阳离子脂质"。在某些实施方案中,可电离的脂质分子可以包含胺基,并且可以被称为可电离的氨基脂质。如本文使用的,"荷电部分"是携带形式电子电荷的化学部分,例如单价(+1或-1)、二价(+2或-2)、三价(+3或-3)等。荷电部分可以是阴离子的(即,带负电的)或阳离子的(即,带正电的)。带正电荷的部分的实例包括胺基(例如伯胺、仲胺和/或叔胺)、铵基、吡啶鎓基、胍基和咪唑鎓基。在一个特定实施方案中,荷电部分包含胺基。带负电荷的基团或其前体的实例包括羧酸基、磺酸基、硫酸基、膦酸基、磷酸基、羟基等等。在一些情况下,荷电部分的电荷可以随环境条件而变,例如,pH中的变化可以改变该部分的电荷,和/或促使该部分变得荷电或不荷电。一般而言,可以根据需要选择分子的电荷密度。

[0705] 应该理解术语"荷电"或"荷电部分"并非指分子上的"部分负电荷"或"部分正电荷"。术语"部分负电荷"和"部分正电荷"给予其在本领域中的普通含义。当官能团包含这样

的键时,可能导致“部分负电荷”,所述键变得极化,使得电子密度被拉向该键的一个原子,在原子上产生部分负电荷。一般而言,本领域技术人员识别可以以这种方式变得极化的键。

[0706] 在一些实施方案中,可电离的脂质是可电离的氨基脂质,在本领域中有时被称为“可电离的阳离子脂质”。在一个实施方案中,可电离的氨基脂质可以具有经由接头结构连接的带正电荷的亲水头部和疏水尾部。除这些之外,可电离的脂质也可以是包括环状胺基的脂质。

[0707] 在一个实施方案中,可电离的脂质可以选自(但不限于)国际公开号WO2013086354和WO2013116126中描述的可电离的脂质。

[0708] 在再一个实施方案中,可电离的脂质可以选自(但不限于)美国专利号7,404,969的式CLI-CLXXXXII。

[0709] 在一些实施方案中,脂质纳米颗粒可以包括一种或多种(例如1、2、3、4、5、6、7或8种)阳离子和/或可电离的脂质。此类阳离子和/或可电离的脂质包括但不限于3-(双十二烷基氨基)-N1,N1,4-三(十二烷基)-1-哌嗪乙胺(KL10)、N1-[2-(双十二烷基氨基)乙基]-N1,N4,N4-三(十二烷基)-1,4-哌嗪二乙胺(KL22)、14,25-双十三烷基-15,18,21,24-四氮杂-三十八烷(KL25)、1,2-二亚油酰基氧基-N,N-二甲基氨基丙烷(DLin-DMA)、2,2-二亚油酰基-4-二甲基氨基甲基-[1,3]-二氧戊环(DLin-K-DMA)、三十七碳-6,9,28,31-四烯-19-基4-(二甲基氨基)丁酸酯(DLin-MC3-DMA)、2,2-二亚油酰基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLin-KC2-DMA)、2-({8-[(3β)-胆甾-5-烯-3-基氧基]辛基}氧基)-N,N-二甲基-3-[(9Z,12Z)-十八碳-9,12-二烯-1-基氧基]丙-1-胺(Octyl-CLinDMA)、(2R)-2-({8-[(3β)-胆甾-5-烯-3-基氧基]辛基}氧基)-N,N-二甲基-3-[(9Z,12Z)-十八碳-9,12-二烯-1-基氧基]丙-1-胺(Octyl-CLinDMA(2R))、(2S)-2-({8-[(3β)-胆甾-5-烯-3-基氧基]辛基}氧基)-N,N-二甲基-3-[(9Z,12Z)-十八碳-9,12-二烯-1-基氧基]丙-1-胺(Octyl-CLinDMA(2S))、N,N-二油酰基-N,N-二甲基氯化铵(“DODAC”);N-(2,3-二油氧基)丙基-N,N--N-三乙基氯化铵(“DOTMA”);N,N-二硬脂基-N,N-二甲基溴化铵(“DDAB”);N-(2,3-二油酰氧基)丙基)-N,N,N-三甲基氯化铵(“DOTAP”);1,2-二油氧基-3-三甲基氨基丙烷氯化物盐(“DOTAP.Cl”);3-β-(N--(N',N'-二甲基氨基乙烷)-氨基甲酰基)胆固醇(“DC-Chol”)、N-(1-(2,3-二油氧基)丙基)-N-2-(精胺甲酰胺基)乙基)-N,N-二甲基三氟乙酸铵(“DOSPA”)、双十八烷基酰胺基甘氨酰羧基精胺(“DOGS”)、1,2-二油酰基-3-二甲基铵丙烷(“DODAP”)、N,N-二甲基-2,3-二油氧基)丙胺(“DODMA”)、以及N-(1,2-二肉豆蔻酰基氧基丙-3-基)-N,N-二甲基-N-羟乙基溴化铵(“DMRIE”)。

[0710] 另外,可以使用阳离子和/或可电离的脂质的许多商业制剂,例如,LIPOFECTIN®(包括可从GIBCO/BRL获得的DOTMA和DOPE)、以及LIPOFECTAMINE®(包括可从GIBCO/BRL获得的DOSPA和DOPE)。KL10、KL22和KL25例如在美国专利号8,691,750中进行描述。

[0711] *阴离子脂质*

[0712] 适用于本公开内容的脂质纳米颗粒中的阴离子脂质包括但不限于磷脂酰甘油、心磷脂、二酰基磷脂酰丝氨酸、二酰基磷脂酸、N-十二烷酰基磷脂酰乙醇胺、N-琥珀酰基磷脂酰乙醇胺、N-戊二酰基磷脂酰乙醇胺、赖氨酰磷脂酰甘油以及与中性脂质连接的其它阴离子修饰基团。

[0713] *中性脂质*

[0714] 适用于本公开内容的脂质纳米颗粒中的中性脂质(包括非荷电和两性离子脂质两者)包括但不限于二酰基磷脂酰胆碱、二酰基磷脂酰乙醇胺、神经酰胺、鞘磷脂、二氢鞘磷脂、脑磷脂、固醇(例如胆固醇)和脑苷脂。在一些实施方案中,脂质纳米颗粒包含胆固醇。具有不同链长和饱和度的各种酰基链基团的脂质是可获得的,或者可以通过众所周知的技术分离或合成。另外,可以使用具有饱和与不饱和脂肪酸链和环状区域的混合物的脂质。在一些实施方案中,本公开内容中使用的中性脂质是DOPE、DSPC、DPPC、POPC或任何相关的磷脂酰胆碱。在一些实施方案中,中性脂质可以包含鞘磷脂、二氢鞘磷脂、或具有其它头基如丝氨酸和肌醇的磷脂。

[0715] *两亲性脂质*

[0716] 在一些实施方案中,两亲性脂质包括在本公开内容的纳米颗粒中。适用于本公开内容的纳米颗粒中的示例性两亲性脂质包括但不限于鞘脂、磷脂、脂肪酸和氨基脂质。

[0717] 本文公开的药物组合物的脂质组合物可以包含一种或多种磷脂,例如,一种或多种饱和或(多)不饱和磷脂或其组合。一般而言,磷脂包含磷脂部分和一个或多个脂肪酸部分。

[0718] 磷脂部分可以非限制性地选自例如磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰甘油、磷脂酰丝氨酸、磷脂酸、2-溶血磷脂酰胆碱和鞘磷脂。

[0719] 脂肪酸部分可以非限制性地选自例如月桂酸、肉豆蔻酸、肉豆蔻油酸、棕榈酸、棕榈油酸、硬脂酸、油酸、亚油酸、α-亚麻酸、芥酸、植烷酸、花生酸、花生四烯酸、二十碳五烯酸、山嵛酸、二十二碳五烯酸和二十二碳六烯酸。

[0720] 特定的两亲性脂质可以促进与膜的融合。例如,阳离子磷脂可以与膜(例如细胞膜或细胞内膜)的一种或多种带负电荷的磷脂相互作用。磷脂与膜的融合可以允许含脂质的组合物(例如,LNP)的一种或多种成分(例如,治疗剂)穿过膜,允许例如将一种或多种成分递送至靶组织。

[0721] 还考虑了非天然的两亲性脂质种类,包括具有修饰和取代包括分支、氧化、环化和炔烃的天然种类。例如,磷脂可以由一种或多种炔烃(例如其中一个或多个双键由三键替换的烯基)官能化或与之交联。在适当的反应条件下,炔基在暴露于叠氮化物时可以经历铜催化的环加成反应。此类反应可以用于官能化纳米颗粒组合物的脂质双层,以促进膜渗透或细胞识别,或者可用于将纳米颗粒组合物缀合至有用的组分,例如靶向部分或成像部分(例如染料)。

[0722] 磷脂包括但不限于甘油磷脂,例如磷脂酰胆碱、磷脂酰乙醇胺、磷脂酰丝氨酸、磷脂酰肌醇、磷脂酰甘油和磷脂酸。磷脂还包括鞘氨醇磷脂,例如鞘磷脂。

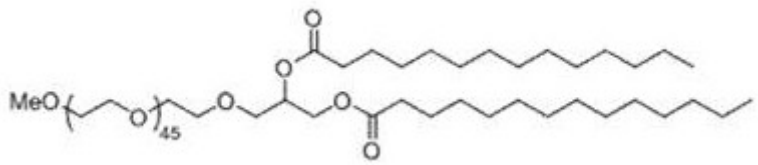
[0723] 还可以使用其它缺乏磷的化合物,例如鞘脂、鞘糖脂家族、二酰基甘油和β-酰氧基酸。另外,此类两亲性脂质可以容易地与其它脂质例如甘油三酯和固醇混合。

[0724] *PEG化脂质*

[0725] 脂质纳米颗粒组合物的脂质组分可以包括一种或多种分子,其包含聚乙二醇例如PEG或PEG修饰的脂质。此类种类可以可替代地被称为PEG化脂质。PEG化脂质(也称为PEG脂质或PEG修饰的脂质)是用聚乙二醇修饰的脂质。PEG化脂质可以非限制性地选自PEG-修饰的磷脂酰乙醇胺、PEG-修饰的磷脂酸、PEG-修饰的神经酰胺、PEG-修饰的二烷基胺、PEG-修饰的二酰基甘油和PEG-修饰的二烷基甘油。例如,聚乙二醇化的脂质可以是PEG-c-DOMG、

PEG-DMG、PEG-DLPE、PEG-DMPE、PEG-DPPC或PEG-DSPE脂质。

[0726] 在一些实施方案中,PEG-修饰的脂质是PEG DMG的修饰形式。PEG-DMG具有下述结构:

[0727] 

[0728] 在一个实施方案中,可用于本发明中的PEG脂质可以是国际公开号WO2012099755中描述的PEG化脂质。本文所述的这些示例性PEG脂质中的任一种可以进行修饰,以包含在PEG链上的羟基。在某些实施方案中,PEG脂质是PEG-OH脂质。如本文一般定义的,"PEG-OH脂质"(在本文中也称为"羟基-PEG化脂质")是在脂质上具有一个或多个羟基(-OH)的PEG化脂质。在某些实施方案中,PEG-OH脂质包括在PEG链上的一个或多个羟基。在某些实施方案中,PEG-OH或羟基-PEG化脂质包含在PEG链的末端处的-OH基团。每种可能性代表本发明的一个分开的实施方案。在一些实施方案中,PEG链的长度包含约250、约500、约1000、约2000、约3000、约5000、约10000个环氧乙烷单元。

[0729] *结构脂质*

[0730] 本文公开的药物组合物的脂质组合物可以包含一种或多种结构脂质。如本文使用的,术语"结构脂质"指固醇以及含有固醇部分的脂质。

[0731] 在脂质纳米颗粒中掺入结构脂质可以帮助减轻颗粒中的其它脂质的聚集。结构脂质可以选自(但不限于)胆固醇、粪甾醇、谷甾醇、麦角固醇、菜油甾醇、豆甾醇、菜籽甾醇、番茄碱、番茄素(tomatine)、熊果酸、α-生育酚、藿烷类化合物、植物固醇、类固醇及其混合物。在一些实施方案中,结构脂质是固醇。如本文定义的,"固醇"是由类固醇类组成的类固醇亚组。在某些实施方案中,结构脂质是类固醇。在某些实施方案中,结构脂质是胆固醇。在某些实施方案中,结构脂质是胆固醇的类似物。

[0732] *靶向部分*

[0733] 在某些实施方案中,期望使用对细胞类型和/或组织类型特异性的靶向部分,来靶向本公开内容的纳米颗粒,例如脂质纳米颗粒。在一些实施方案中,可以使用靶向部分将纳米颗粒靶向特定的细胞、组织和/或器官。在特定实施方案中,纳米颗粒包含靶向部分。示例性的非限制性靶向部分包括配体、细胞表面受体、糖蛋白、维生素(例如核黄素)和抗体(例如全长抗体、抗体片段(例如Fv片段、单链Fv(scFv)片段、Fab'片段或F(ab')2片段)、单结构域抗体、骆驼科动物抗体及其片段、人抗体及其片段、单克隆抗体和多特异性抗体(例如双特异性抗体))。在一些实施方案中,靶向部分可以是多肽。靶向部分可以包括整个多肽(例如肽或蛋白质)或其片段。靶向部分通常以这样的方式定位在纳米颗粒的外表面上,使得靶向部分可用于与靶例如细胞表面受体相互作用。各种不同的靶向部分和方法是本领域已知且可获得的,包括例如Sapra等人Prog. Lipid Res. 42(5):439-62,2003以及Abra等人J. Liposome Res. 12:1-3,2002中描述的那些。

[0734] 在一些实施方案中,脂质纳米颗粒(例如脂质体)可以包括亲水聚合物链如聚乙二醇(PEG)链的表面涂层(参见例如,Allen等人Biochimica et Biophysica Acta 1237: 99-108,1995;DeFrees等人Journal of the American Chemistry Society 118: 6101-6104,

1996;Blume等人Biochimica et Biophysica Acta 1149: 180-184,1993;Klibanov等人Journal of Liposome Research 2: 321-334,1992;美国专利号5,013,556;Zalipsky,Bioconjugate Chemistry 4: 296-299,1993;Zalipsky,FEBS Letters 353: 71-74,1994;Zalipsky,in Stealth Liposomes第9章(Lasic和Martin,编辑)CRC Press,Boca Raton Fla.,1995)。在一种方法中,用于靶向脂质纳米颗粒的靶向部分连接至形成纳米颗粒的脂质的极性头基。在另一种方法中,靶向部分附着至形成亲水聚合物涂层的PEG链的远端(参见例如,Klibanov等人Journal of Liposome Research 2: 321-334,1992;Kirpotin等人FEBS Letters 388: 115-118,1996)。

[0735] 可以使用用于偶联一个或多个靶向部分的标准方法。例如,可以使用磷脂酰乙醇胺(其可以被激活用于附着靶向部分)、或衍生的亲脂性化合物例如脂质衍生的博来霉素。靶向抗体的脂质体可以使用例如掺入蛋白A的脂质体来构建(参见例如,Renneisen等人J. Bio. Chem.,265:16337-16342,1990和Leonetti等人Proc. Natl. Acad. Sci.(USA),87:2448-2451,1990)。抗体缀合的其它实例公开于美国专利号6,027,726中。靶向部分的实例还可以包括对细胞组分特异性的其它多肽,所述细胞组分包括与赘生物或肿瘤相关的抗原。用作靶向部分的多肽可以经由共价键附着至脂质体(参见例如,Heath,Covalent Attachment of Proteins to Liposomes,149 Methods in Enzymology 111-119(Academic Press,Inc. 1987))。其它靶向方法包括生物素-抗生物素蛋白系统。

[0736] 在一些实施方案中,本公开内容的脂质纳米颗粒包括将脂质纳米颗粒靶向细胞的靶向部分,所述细胞包括但不限于肝细胞、结肠细胞、上皮细胞、造血细胞、上皮细胞、内皮细胞、肺细胞、骨细胞、干细胞、间充质细胞、神经细胞、心脏细胞、脂肪细胞、血管平滑肌细胞、心肌细胞、骨骼肌细胞、β细胞、垂体细胞、滑膜衬里细胞、卵巢细胞、睾丸细胞、成纤维细胞、B细胞、T细胞、网织红细胞、白细胞、粒细胞和肿瘤细胞(包括原发性肿瘤细胞和转移性肿瘤细胞)。在特定实施方案中,靶向部分将脂质纳米颗粒靶向肝细胞。

[0737] *类脂质*

[0738] 本文所述的脂质纳米颗粒可以是基于类脂质的。类脂质的合成已得到广泛描述,并且含有这些化合物的制剂特别适合多核苷酸的递送(参见Mahon等人Bioconjug Chem. 2010 21:1448-1454;Schroeder等人J Intern Med. 2010 267:9-21;Akinc等人Nat. Biotechnol. 2008 26:561-569;Love等人Proc Natl Acad Sci USA. 2010 107:1864-1869;Siegwart等人Proc Natl Acad Sci USA. 2011 108:12996-3001)

[0739] 根据本发明,可以制备含有这些类脂质的复合物、胶束、脂质体或颗粒(例如纳米颗粒),并且因此在经由局限性和全身施用途径注射之后,导致供体多核苷酸或系统的有效递送,如通过例如供体多核苷酸插入gDNA内所确定的。包含类脂质复合物的药物组合物可以通过本文公开的各种手段进行施用。

[0740] 取决于靶细胞类型和制剂扩散通过细胞外基质进入血流内的能力,用于肌内或皮下途径的优化的类脂质制剂的特征可能显著不同。尽管由于内皮窗孔(fenestrae)的大小,可能需要小于150 nm的粒度用于有效的肝细胞递送(参见例如,Akinc等人Mol Ther. 2009 17:872-879),但使用类脂质寡核苷酸将制剂递送至其它细胞类型(包括但不限于内皮细胞、髓样细胞和肌肉细胞)可能没有类似的大小限制。

[0741] 在一个方面,对髓样细胞如单核细胞的有效递送,类脂质制剂可以具有相似的组

分摩尔比。类脂质和其它成分的不同比率可以用于优化供体多核苷酸或系统的制剂对不同细胞类型的递送,所述其它成分包括但不限于中性脂质(例如,二酰基磷脂酰胆碱)、胆固醇、PEG化脂质(例如,PEG-DMPE)和脂肪酸(例如,ω-3脂肪酸),所述细胞类型包括但不限于肝细胞、髓样细胞、肌肉细胞等。示例性类脂质包括但不限于DLin-DMA、DLin-K-DMA、DLin-KC2-DMA、98N12-5、C12-200(包括变体和衍生物)、DLin-MC3-DMA及其类似物。使用类脂质制剂经由皮下或肌内递送将核酸局限性递送至细胞(例如但不限于脂肪细胞和肌肉细胞),也可能不需要对于全身递送可能需要的所有制剂组分,并且因此可以包含类脂质和供体多核苷酸或系统。

[0742] 在一个进一步的实施方案中,不同类脂质的组合可以用于改善由本公开内容提供的供体多核苷酸或系统的功效。

[0743] 根据本公开内容,由本公开内容提供的供体多核苷酸或系统可以通过在加入细胞之前,将供体多核苷酸或系统、或系统的个别组分以设定比率与类脂质混合进行配制。体内制剂可能需要添加额外的成分,以促进遍及全身的循环。在颗粒形成后,添加由本公开内容提供的供体多核苷酸、系统或系统的个别组分,并且允许与复合物整合。使用标准的染料排斥测定来确定封装效率。

[0744] 本公开内容的供体多核苷酸和/或系统的体内递送可能受到许多参数的影响,所述参数包括但不限于制剂组成、颗粒PEG化的性质、装载程度、寡核苷酸与脂质的比率、以及生物物理学参数例如粒度(Akinc等人Mol Ther. 2009 17:872-879;整体引入本文作为参考)。例如,聚(乙二醇)(PEG)脂质的锚定链长度中的小变化可能导致对体内功效的显著作用。具有不同类脂质的制剂可以就体内活性进行测试,所述类脂质包括但不限于五[3-(1-月桂基氨基丙酰基)]-三亚乙基四胺盐酸盐(TETA-5LAP;又名98N12-5,参见Murugaiah等人Analytical Biochemistry,401:61(2010))、C12-200(包括衍生物和变体)、MD1、DLin-DMA、DLin-K-DMA、DLin-KC2-DMA和DLin-MC3-DMA。在本文中称为"98N12-5"的类脂质通过Akinc等人Mol Ther. 2009 17:872-879)得到公开。在本文中称为"C12-200"的类脂质通过Love等人Proc Natl Acad Sci USA. 2010 107:1864-1869以及Liu和Huang,Molecular Therapy. 2010 669-670得到公开。

[0745] 类脂质配制的供体多核苷酸或系统在体外或体内改变gDNA中的核苷酸序列(例如,校正或诱导突变)的能力,可以通过本领域已知的或本文所述的任何技术(例如,下一代DNA测序)来确定。

[0746] *其它组分*

[0747] 本文公开的纳米颗粒还可以包括除上述那些之外的一种或多种组分。例如,脂质组合物可以包括一种或多种渗透性增强剂分子、碳水化合物、聚合物、表面改变剂(例如表面活性剂)或其它组分。例如,渗透性增强剂分子可以是通过美国专利申请公开号2005/0222064中描述的分子。碳水化合物可以包括单糖(例如葡萄糖)和多糖(例如糖原及其衍生物和类似物)。

[0748] 药物组合物

[0749] 本公开内容包括药物组合物,其包含与一种或多种药学上可接受的赋形剂、载体或稀释剂组合的供体多核苷酸、gRNA和Cas9蛋白。在特定实施方案中,供体多核苷酸被封装在纳米颗粒例如脂质纳米颗粒中。在一些实施方案中,gRNA被封装在纳米颗粒中。在一些实

施方案中,Cas核酸酶(例如SpCas9)被封装在纳米颗粒中。在特定实施方案中,编码Cas核酸酶的mRNA或封装Cas核酸酶的纳米颗粒存在于药物组合物中。在各种实施方案中,存在于药物组合物中的一种或多种mRNA被封装在纳米颗粒例如脂质纳米颗粒中。在特定实施方案中,第一mRNA与第二mRNA的摩尔比为约1:50、约1:25、约1:10、约1:5、约1:4、约1:3、约1:2、约1:1、约2:1、约3:1、约4:1、或约5:1、约10:1、约25:1或约50:1。在特定实施方案中,第一mRNA与第二mRNA的摩尔比大于

[0750] 在一些实施方案中,脂质组合物和供体多核苷酸之间的比率可以是约5:1、6:1、7:1、8:1、9:1、10:1、11:1、12:1、13:1、14:1、15:1、16:1、17:1、18:1、19:1、20:1、21:1、22:1、23:1、24:1、25:1、26:1、27:1、28:1、29:1、30:1、31:1、32:1、33:1、34:1、35:1、36:1、37:1、38:1、39:1、40:1、41:1、42:1、43:1、44:1、45:1、46:1、47:1、48:1、49:1、50:1、51:1、52:1、53:1、54:1、55:1、56:1、57:1、58:1、59:1或60:1(wt/wt)。在一些实施方案中,脂质组合物与多核苷酸的wt/wt比为约20:1或约15:1。

[0751] 在一个实施方案中,本文所述的脂质纳米颗粒可以包含多核苷酸(例如供体多核苷酸),其脂质:多核苷酸重量比为5:1、10:1、15:1、20:1、25:1、30:1、35:1、40:1、45:1、50:1、55:1、60:1或70:1,或者这些比率的范围或任一种,例如但不限于5:1至约10:1、约5:1至约15:1、约5:1至约20:1、约5:1至约25:1、约5:1至约30:1、约5:1至约35:1、约5:1至约40:1、约5:1至约45:1、约5:1至约50:1、约5:1至约55:1、约5:1至约60:1、约5:1至约70:1、约10:1至约15:1、约10:1至约20:1、约10:1至约25:1、约10:1至约30:1、约10:1至约35:1、约10:1至约40:1、约10:1至约45:1、约10:1至约50:1、约10:1至约55:1、约10:1至约60:1、约10:1至约70:1、约15:1至约20:1、约15:1至约25:1、约15:1至约30:1、约15:1至约35:1、约15:1至约40:1、约15:1至约45:1、约15:1至约50:1、约15:1至约55:1、约15:1至约60:1或约15:1至约70:1。

[0752] 在一个实施方案中,本文所述的脂质纳米颗粒可以包含多核苷酸,其浓度为大约0.1 mg/ml至2 mg/ml,例如但不限于0.1 mg/ml、0.2 mg/ml、0.3 mg/ml、0.4 mg/ml、0.5 mg/ml、0.6 mg/ml、0.7 mg/ml、0.8 mg/ml、0.9 mg/ml、1.0 mg/ml、1.1 mg/ml、1.2 mg/ml、1.3 mg/ml、1.4 mg/ml、1.5 mg/ml、1.6 mg/ml、1.7 mg/ml、1.8 mg/ml、1.9 mg/ml、2.0 mg/ml或大于2.0 mg/ml。

[0753] 治疗方法

[0754] 本文提供的是通过校正基因组DNA分子中的突变来治疗患有疾病的患者的方法。在一些实施方案中,该方法可以包括将本文所述的供体多核苷酸、系统、载体或药物组合物引入细胞内。在一些实施方案中,该方法可以包括将供体多核苷酸、系统、载体或药物组合物施用于有此需要的受试者(例如,患有由突变引起的疾病的患者)。

[0755] 本公开内容的实施方案涵盖用于编辑细胞中的靶核酸分子(基因组DNA)的方法。在一些实施方案中,该方法包括将本文所述的供体多核苷酸引入细胞内。在一些实施方案中,该方法包括使细胞与本文所述的药物组合物接触。在一些实施方案中,该方法包括生成包含靶向的编辑的核酸分子的稳定细胞系。在一些实施方案中,细胞是真核细胞。真核细胞的非限制性实例包括酵母细胞、植物细胞、昆虫细胞、来自无脊椎动物的细胞、来自脊椎动物的细胞、哺乳动物细胞、啮齿类动物细胞、小鼠细胞、大鼠细胞和人细胞。在一些实施方案中,真核细胞可以是哺乳动物细胞。在一些实施方案中,真核细胞可以是啮齿类动物细胞。

在一些实施方案中,真核细胞可以是人细胞。类似地,靶序列可以来自任何此类细胞或在任何此类细胞中。

[0756] 本文所述的供体多核苷酸、系统、载体或药物组合物可以经由本领域已知的任何方法引入细胞内,所述方法例如病毒或细菌噬菌体感染、转染、缀合、原生质体融合、脂转染、电穿孔、磷酸钙沉淀、聚乙烯亚胺(PEI)介导的转染、DEAE-葡聚糖介导的转染、脂质体介导的转染、粒子枪技术、磷酸钙沉淀、剪切驱动的细胞渗透、与细胞穿透肽的融合随后为细胞接触、显微注射和纳米颗粒介导的传递。在一些实施方案中,可以经由病毒感染将载体系统引入细胞内。在一些实施方案中,可以经由细菌噬菌体感染将载体系统引入细胞内。

[0757] 本发明的实施方案还涵盖了用本文所述的供体多核苷酸、系统、载体或药物组合物来治疗患者。在一些实施方案中,该方法可以包括向患者施用本文所述的供体多核苷酸、系统、载体或药物组合物。该方法可以用作单一疗法或与本领域可获得的其它疗法组合使用。在一些实施方案中,患者可以具有在疾病相关基因中的突变(例如,插入、缺失、取代、染色体易位)。在一些实施方案中,供体多核苷酸、系统、载体或药物组合物的施用,可以导致患者中包含疾病相关基因的一个或多个核苷酸的插入、缺失或取代的突变。某些实施方案可以包括修复患者在疾病相关基因中的突变的方法。在一些实施方案中,突变可以导致由疾病相关基因表达的蛋白质中的一个或多个氨基酸改变。在一些实施方案中,突变可以导致由疾病相关基因表达的RNA中的一个或多个核苷酸改变。在一些实施方案中,突变可以改变疾病相关基因的表达水平。在一些实施方案中,突变可以导致基因的表达增加或减少。在一些实施方案中,突变可以导致患者中的基因敲减。在一些实施方案中,供体多核苷酸、系统、载体或药物组合物的施用可以导致患者的疾病相关基因中的突变的校正。在一些实施方案中,供体多核苷酸、系统、载体或药物组合物的施用可以导致患者中的基因敲除。在一些实施方案中,供体多核苷酸、系统、载体或药物组合物系统的施用,可以导致疾病相关基因的外显子序列、内含子序列、转录控制序列、翻译控制序列或非编码序列的替换。

[0758] 在一些实施方案中,供体多核苷酸、系统、载体或药物组合物的施用,可以导致外源序列(例如,供体多核苷酸序列)整合到患者的基因组DNA内。在一些实施方案中,外源序列可以包含可操作地连接至外源启动子序列的蛋白质或RNA编码序列,使得在外源序列整合到患者的基因组DNA内之后,患者能够表达由整合序列编码的蛋白质或RNA。外源序列可以提供补充或替换的蛋白质编码或非编码序列。例如,供体多核苷酸、系统、载体或药物组合物的施用,可以导致患者中的疾病相关基因的突变部分的替换。在一些实施方案中,突变部分可以包括疾病相关基因的外显子。在其它实施方案中,外源序列的整合可以导致来自患者的基因组DNA上存在的内源启动子序列的整合序列的表达。例如,供体多核苷酸、系统、载体或药物组合物的施用,可以导致疾病相关基因的功能基因产物的供应,以纠正患者的突变。在再其它实施方案中,供体多核苷酸、系统、载体或药物组合物的施用,可以导致外显子序列、内含子序列、转录控制序列、翻译控制序列或非编码序列整合到患者的基因组DNA内。

[0759] 本发明的另外实施方案还涵盖了以组织特异性方式治疗患者的方法。在一些实施方案中,该方法可以包括将如本文所述的包含组织特异性启动子的供体多核苷酸、系统、载体或药物组合物施用于患者。用于通过该方法治疗的合适组织的非限制性实例包括免疫系统、神经元、肌肉、胰腺、血液、肾脏、骨骼、肺、皮肤、肝脏和乳腺组织。

[0760] 在一些实施方案中,本公开内容提供了校正细胞中的基因组DNA分子(gDNA)中的突变的方法,所述方法包括使细胞与本文所述的供体多核苷酸,根据本公开内容的包含供体多核苷酸、gRNA和定点核酸酶的系统,或本文所述的药物组合物接触,其中当供体多核苷酸、系统或组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0761] 在一些实施方案中,本公开内容提供了通过校正细胞中的基因组DNA分子(gDNA)中的突变,来治疗患有疾病的患者的方法,所述方法包括从患者中分离细胞,使细胞与本文所述的供体多核苷酸,根据本公开内容的包含供体多核苷酸、gRNA和定点核酸酶的系统,或本文所述的药物组合物接触,其中当供体多核苷酸、系统或组合物接触细胞时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0762] 在一些实施方案中,本公开内容提供了通过校正细胞中的基因组DNA分子(gDNA)中的突变,来治疗患有疾病的患者的方法,所述方法包括向患者施用有效量的本文所述的供体多核苷酸,根据本公开内容的包含供体多核苷酸、gRNA和定点核酸酶的系统,或本文所述的药物组合物,其中当施用供体多核苷酸、系统或组合物时,非同源末端连接(NHEJ)DNA修复途径将供体多核苷酸插入在接近突变的位置处引入gDNA内的双链DNA断裂内,从而校正了突变。

[0763] 在一些实施方案中,细胞是患者特异性诱导多能干细胞(iPSC)。在一些实施方案中,细胞是肝细胞。在一些实施方案中,该方法进一步包括使包含校正的突变的iPSC分化成分化细胞,并且将分化细胞植入患者内。在一些实施方案中,治疗导致由基因组DNA分子(gDNA)转录的mRNA的翻译,所述gDNA包含插入的供体多核苷酸,其中所述翻译导致翻译产物(蛋白质)的形成,所述翻译产物减轻疾病或者并不引起或促进疾病。

[0764] 试剂盒

[0765] 本公开内容提供了试剂盒,其包含如本文公开的供体多核苷酸,包含供体多核苷酸、一种或多种gRNA分子和定点核酸酶的系统,或者包含供体多核苷酸、系统或由系统编辑的细胞的药物组合物,以及使用说明。试剂盒可以包含在合适容器中的供体多核苷酸、系统或药物组合物,一种或多种对照,以及本领域众所周知的各种缓冲液、试剂、酶和其它标准成分。某些实施方案包括具有本文公开的供体多核苷酸的试剂盒,伴随在CRISPR/Cas9系统中用于治疗或延迟受试者中的疾病或病症的进展的说明书。在一些方面,供体多核苷酸和CRISPR/Cas9系统的剩余组分在分开的小瓶中提供。

[0766] 在一些方面,本公开内容提供了包括本文公开的供体多核苷酸、系统或药物组合物可以置于其内的容器的试剂盒,所述容器包括至少一个小瓶、孔、试管、烧瓶、瓶、注射器或其它容器装置。当提供了另外的组分时,该试剂盒可以含有这种组分可以置于其内的另外容器。容器和/或试剂盒可以包括具有使用说明书和/或警告的标签。

[0767] 在一些方面,本公开内容提供了试剂盒,其包含本文公开的供体多核苷酸、系统或药物组合物以及包装插页,所述包装插页包含用于与CRISPR/Cas9系统的组分组合施用的说明书,用于校正细胞中的基因组DNA分子(gDNA)中的突变,其中突变的校正用于治疗或延迟受试者中起因于突变的疾病或病症的进展。

[0768] 在一些实施方案中,本公开内容提供了包括容器和包含使用说明书的包装插页的

试剂盒,所述容器包含本文公开的供体多核苷酸、系统或药物,其用于校正细胞中的基因组DNA分子(gDNA)中的突变。

[0769] 定义

[0770] 除非另有说明,否则权利要求和说明书中使用的术语如下文阐述的进行定义。在与母案临时专利申请中使用的术语直接冲突的情况下,以本申请中使用的术语为准。

[0771] 必须注意的是,如说明书和所附权利要求中使用的,单数形式"一个"、"一种"和"该/所述"包括复数指示物,除非上下文另有明确规定。进一步地,除非上下文另有要求,否则单数术语应该包括复数,并且复数术语应该包括单数。

[0772] *约*:如本文使用的,术语"约"(可替代地"大约")由本领域技术人员理解,并且根据它在其中使用的上下文,在一定程度上变化。如果鉴于它在其中使用的上下文,存在对普通技术人员并不明确的术语使用,则"约"意指特定值的至多加上或减去10%。

[0773] *氨基酸*:如本文使用的,术语"氨基酸"指天然存在的和合成的氨基酸,以及以类似于天然存在的氨基酸的方式发挥功能的氨基酸类似物和氨基酸模拟物。天然存在的氨基酸是由遗传密码编码的氨基酸,以及以后进行修饰的那些氨基酸,例如羟脯氨酸、γ-羧基谷氨酸和O-磷酸丝氨酸。氨基酸类似物指与天然存在的氨基酸具有相同的基本化学结构的化合物,所述基本化学结构即与氢、羧基、氨基和R基团结合的碳,所述化合物例如高丝氨酸、正亮氨酸、甲硫氨酸亚砜、甲硫氨酸甲基锍。此类类似物具有修饰的R基团(例如正亮氨酸)或修饰的肽主链,但保留与天然存在的氨基酸相同的基本化学结构。氨基酸模拟物指这样的化学化合物,其具有与氨基酸的一般化学结构不同的结构,但以与天然存在的氨基酸相似的方式发挥功能。

[0774] 氨基酸可以在本文中通过其通常已知的三字母符号、或由IUPAC-IUB生化命名委员会(Biochemical Nomenclature Commission)推荐的单字母符号来提及。同样地,核苷酸可以通过其普遍公认的单字母代码来提及。

[0775] *氨基酸取代*:如本文使用的,"氨基酸取代"指用第二种不同的"替换"氨基酸残基替换预定的氨基酸序列(起始多肽的氨基酸序列)中的至少一个现有氨基酸残基。"氨基酸插入"指将至少一个另外的氨基酸掺入预定的氨基酸序列内。虽然插入通常由一个或两个氨基酸残基的插入组成,但也可以制备较大的"肽插入",例如约三个至约五个或甚至多达约十个、十五个或二十个氨基酸残基的插入。如上文公开的,插入的残基可以是天然存在的或非天然存在的。"氨基酸缺失"指从预定的氨基酸序列中去除至少一个氨基酸残基。

[0776] *碱基组成*:如本文使用的,术语"碱基组成"指由鸟嘌呤+胞嘧啶或胸腺嘧啶(或尿嘧啶)+腺嘌呤核碱基组成的核酸总碱基的比例。

[0777] *碱基对*:如本文使用的,术语"碱基对"指在相反的互补多核苷酸链或同一链的区域上的两个核碱基,其经由形成特异性氢键而相互作用。如本文使用的,术语"沃森-克里克碱基配对"与"互补碱基配对"可互换使用,指一组碱基配对规则,其中嘌呤总是与嘧啶结合,使得核碱基腺嘌呤(A)与DNA分子中的胸腺嘧啶(T)形成互补碱基对,并且鸟嘌呤(G)与胞嘧啶(C)形成互补碱基对。在RNA分子中,胸腺嘧啶替换为尿嘧啶(U),其类似于胸腺嘧啶(T),与腺嘌呤(A)形成互补碱基对。互补碱基对通过氢键结合在一起,并且氢键的数目在碱基对之间不同。如本领域中已知的,鸟嘌呤(G)-胞嘧啶(C)碱基对通过三(3)个氢键结合,而腺嘌呤(A)-胸腺嘧啶(T)或尿嘧啶(U)碱基对通过两(2)个氢键结合。

[0778] 不遵循这些规则的碱基配对相互作用可以在天然、非天然和合成核酸中发生,并且在本文中被称为“非沃森-克里克碱基配对”或可替代地“非规范碱基配对”。“摆动碱基对”是RNA分子中的两个核碱基之间的配对,其不遵循沃森-克里克碱基对规则。例如,肌苷是在结构上类似于鸟苷,但缺少2-氨基的核苷。肌苷能够与四种天然核碱基各自形成两个氢键(Oda等人(1991)Nucleic Acids Res 19:5263-5267),并且它经常由研究人员用作“通用”碱基,这意味着它可以与所有天然存在的或规范碱基进行碱基配对。四个主要摆动碱基对是鸟嘌呤-尿嘧啶(G-U)碱基对、次黄嘌呤-尿嘧啶(I-U)碱基对、次黄嘌呤-腺嘌呤(I-A)碱基对和次黄嘌呤-胞嘧啶(I-C)碱基对。为了维持核酸命名法的一致性,将“I”用于次黄嘌呤,因为次黄嘌呤是肌苷的核碱基;命名法在其它方面沿用了核碱基及其相应核苷的名称(例如,“G”用于鸟嘌呤和鸟苷两者 - 以及用于脱氧鸟苷)。摆动碱基对的热力学稳定性与沃森-克里克碱基对的热力学稳定性可比较。摆动碱基对在RNA分子的二级结构形成中起作用。

[0779] *密码子*:如本文使用的,术语“密码子”指三个核苷酸的序列,其共同形成DNA或RNA分子中的遗传密码的单位。密码子由起始核苷酸可操作地定义,翻译从所述起始核苷酸开始并且对于一段相继的核苷酸三联体设置框架,其称为“开放读码框”(ORF)。例如,字符串GGGAAACCC如果从第一个位置读取,则含有密码子GGG、AAA和CCC;如果从第二个位置读取,则它含有密码子GGA和AAC;并且如果从第三个位置读取,则含有GAA和ACC。因此,在其5’ → 3’方向上读取的每一个核酸序列包含三个读码框,各自产生可能不同的氨基酸序列(在给定的实例中,分别为Gly-Lys-Pro、Gly-Asn或Glu-Thr)。DNA是双链的,定义了六个可能的读码框,其中三个在一条链上为正向取向,而三个在相反链上为反向。编码多肽的开放读码框通常由起始密码子定义,通常为序列中的第一个AUG密码子。

[0780] *校正或诱导突变*:如本文使用的,术语“校正或诱导突变”指在供体多核苷酸插入gDNA分子中诱导的双链断裂(DSB)内之后,供体多核苷酸(例如本文所述的那些)将所需的改变掺入包含基因组DNA(gDNA)分子的核苷酸序列内,从而改变gDNA的核苷酸序列的功能。

[0781] 术语“校正突变”指通过供体多核苷酸掺入所需改变,其导致包含突变(例如,有害或致病突变)的gDNA中的一个或多个核苷酸改变,使得突变以所需方式被恢复或转换(transmuted)。可以通过比较已知或怀疑包含突变的gDNA的核苷酸序列与野生型gDNA的核苷酸序列,来确定待校正突变的身份。

[0782] 术语“诱导突变”指通过供体多核苷酸掺入所需改变,其导致gDNA中的一个或多个核苷酸改变,使得gDNA以所需方式突变。通过供体多核苷酸诱导的突变可以是本领域已知的任何类型的突变。在一些实施方案中,突变的诱导是出于治疗目的或导致疗效。

[0783] *共价连接的*:如本文使用的,术语“共价连接的”(可替代地“缀合的”、“连接的”、“附着的”、“融合的”或“栓系的”),当就两个或更多个部分而言使用时,意指各部分通过包括化学缀合、重组技术或酶促活性的任何手段,直接或经由充当连接剂的一个或多个另外部分,在物理上彼此结合或连接,以形成足够稳定的结构,使得在其中使用该结构的条件例如生理条件下,所述部分保持物理上结合。

[0784] *互补的*:如本文使用的,术语“互补的”或“互补性”指包含两条多核苷酸链的核苷酸序列或同一多核苷酸链的区域之间的关系,以及包含该链或区域的双链体形成,其中两条链或两个区域之间的连续碱基配对的程度足以生成双链体结构。已知腺嘌呤(A)与胸腺

嘧啶(T)或尿嘧啶(U)形成特异性氢键或“碱基配对”。类似地,已知胞嘧啶(C)与鸟嘌呤(G)碱基配对。还已知非规范核碱基(例如肌苷)可以与天然碱基氢键键合。如果当第一链和第二链以反平行方式排列时,两条链之间的碱基配对程度在其中使用双链体结构的条件(例如细胞中的生理条件)下维持双链体结构,则包含多核苷酸的第一链或者其区域、一部分或片段的核苷酸序列,被说成与包含相同或不同核酸的第二链或者其区域、一部分或片段的核苷酸序列“充分互补”。应当理解,多核苷酸的互补链或区域可以包括一些非互补的碱基对。互补性可以是“部分的”,其中多核苷酸包含的仅一些核碱基根据碱基配对规则匹配。或者,核酸之间可能存在“完全”或“总体”互补性。尽管多核苷酸链或区域之间的互补程度对链或区域之间的杂交效率和强度具有显著作用,但并不要求两个互补多核苷酸在每一个核苷酸位置处碱基配对。在一些实施方案中,第一多核苷酸与第二多核苷酸100%或“完全”互补,且因此在每一个核苷酸位置形成碱基对。在一些实施方案中,第一多核苷酸不是100%互补的(例如是90%或80%或70%互补的),并且在一个或多个核苷酸位置处含有错配的核苷酸。尽管经常需要完全的互补性,但一些实施方案可以包括一个或多个,但优选6、5、4、3、2或1个错配。

[0785] *接触*:如本文使用的,术语“接触”意指在两个或更多个实体之间建立物理连接。例如,使细胞与试剂(例如,本公开内容的RNA、脂质纳米颗粒组合物或其它药物组合物)接触,意指使细胞和试剂共享物理连接。使细胞在体内、体外和离体与外部实体接触的方法是生物学领域众所周知的。在本公开内容的示例性实施方案中,使哺乳动物细胞与组合物(例如,本公开内容的分离的RNA、纳米颗粒或药物组合物)接触的步骤在体内执行。例如,可以通过任何合适的施用途径(例如,对生物的肠胃外施用,包括静脉内、肌内、皮内和皮下施用),来执行脂质纳米颗粒组合物与细胞(例如哺乳动物细胞)的接触,所述细胞可以位于生物体(例如哺乳动物)内。对于在体外存在的细胞,可以例如通过将组合物加入细胞的培养基中,使组合物(例如脂质纳米颗粒或分离的RNA)与细胞接触,并且可以涉及或导致转染。此外,多于一种细胞可以由试剂接触。

[0786] *变性*:如本文使用的,术语“变性”指通过其破坏核酸中的碱基配对的核苷酸之间的氢键合,导致二级和/或三级核酸结构的丧失(例如,先前退火的链的分开)的过程。变性可以通过将外部物质、能量或生物化学过程应用于核酸而发生。

[0787] *双链断裂*:如本文使用的,术语“双链断裂”(DSB)指当DNA分子的两条互补链断裂或切割时生成的DNA损伤,导致两个游离DNA端或末端。DSB可能经由暴露于环境损害(例如,辐射、化学试剂或UV光)而发生、或故意生成(例如,经由改造的核酸酶),并且出于确定的生物学目的(例如,插入供体多核苷酸以校正突变)。

[0788] *平端*:如本文使用的,术语“平端”、“平端的”指双链体或双链核酸(例如DNA)的一端的结构,其中构成双链体的两条互补链至少在一端处在一个碱基对中终止。因此,构成双链体的两条链无一比另一条从端部延伸得更远。

[0789] *双链体*:如本文使用的,术语“双链体”指由双链多核苷酸的互补链、或单链多核苷酸在其自身上回折的互补区域形成的结构。由于互补核苷酸序列通过碱基配对相互作用结合在一起或杂交,出现核酸的双链体结构。

[0790] EC_{50}:如本文使用的,术语“EC_{50}”指这样的组合物浓度,其在体外或体内测定中,诱导其为最大应答的50%的应答,即最大应答和基线之间的一半。

[0791] *有效剂量*:如本文使用的,术语“有效剂量(effective dose)”或“有效剂量(effective dosage)”定义为足以达到或至少部分达到所需效应的量。

[0792] *基因组编辑*:如本文使用的,术语基因组编辑一般指优选以精确或预定的方式,编辑或改变基因组的核苷酸序列的过程。本文所述的基因组编辑方法的实例包括以下方法:使用定点核酸酶在基因组内的精确靶位置或序列处切割基因组DNA,从而在靶序列内产生DNA断裂(例如,DSB),并且这样修复DNA断裂,使得修复的基因组的核苷酸序列已在DNA断裂的位点处或附近改变。

[0793] 双链DNA断裂(DSB)可以并定期通过天然的内源性细胞过程进行修复,所述细胞过程例如同源性指导的修复(HDR)和非同源末端连接(NHEJ)(参见例如,Cox等人(2015)Nature Medicine 21(2):121-131)。

[0794] 通过HDR的DNA修复利用了具有核苷酸序列的多核苷酸(经常被称为“修复模板”或“供体模板”),所述核苷酸序列与侧接DSB的序列同源。通过HDR机制的DNA修复涉及修复模板与切割的基因组DNA分子之间的同源重组。可以这样设计修复模板,使得它们插入基因组DNA分子中的核苷酸或使基因组DNA分子中的核苷酸缺失、或者改变基因组DNA分子的核苷酸序列。

[0795] NHEJ机制可以通过将起因于DSB的DNA端直接连接或联接在一起来修复DSB。通过NHEJ的DSB修复可以涉及一个或多个核苷酸的随机插入或缺失(即插入缺失)。通过NHEJ的DNA修复的这一方面经常在基因组编辑方法中用于破坏基因表达。NHEJ还可以通过以同源性不依赖性方式,将外源多核苷酸插入切割位点内来修复DSB。

[0796] 第三种修复机制是微同源性介导的末端连接(MMEJ),也称为“替代NHEJ”,其中遗传结果与NHEJ相似,因为可以在切割位点处发生小缺失和插入。MMEJ利用侧接DNA断裂位点的几个碱基对的同源序列,以驱动更有利的DNA末端连接修复结果(参见例如,Cho和Greenberg,(2015)Nature 518,174-176);Mateos-Gomez等人Nature 518,254-57(2015);Ceccaldi等人Nature 528,258-62(2015)。在一些情况下,可能能够基于在DNA断裂的位点处的潜在微同源性分析,来预测可能的修复结果。前述DNA修复机制各自可以用于基因组编辑方法中,以产生所需的基因组改变。基因组编辑过程中的第一步通常是在靶序列中产生尽可能靠近预期的突变或改变位点的一个或两个DNA断裂。如本文描述且示出的,这可以经由使用定点核酸酶来实现。

[0797] *定点核酸酶*:如本文使用的,术语“定点核酸酶”指几种不同类别的核酸酶之一,其可以被编程或改造为识别DNA分子(例如,基因组DNA分子)中的特异性靶位点(即,靶核苷酸序列),并且在DNA分子内的特异性位点处、附近或其内生成DNA断裂(例如,DSB)。定点核酸酶可用于基因组编辑方法,例如本文所述的那些。定点核酸酶包括但不限于锌指核酸酶(ZFN)、转录激活因子样效应物(TALE)核酸酶、CRISPR/Cas核酸酶(例如Cas9)、归巢核酸内切酶(也称为大范围核酸酶)和其它核酸酶(参见例如,Hafez和Hausner,Genome 55,553-69(2012);Carroll,Ann. Rev. Biochem. 83,409-39(2014);Gupta和Musunuru,J. Clin. Invest. 124,4154-61(2014);以及Cox等人,同上。它们的主要区别在于其结合DNA并产生靶向的位点特异性DNA断裂的方式。本领域已知的定点核酸酶可能产生单链断裂(SSB)或DSB。为了本发明的目的,本公开内容提及“定点核酸酶”指产生DSB的那些核酸酶。在DSB产生后,与NHEJ或HDR基本上相同的天然细胞DNA修复机制被指定(co-opted)以实现所需的遗

传修饰,因此,考虑使用定点核酸酶的基因组编辑技术或系统可以用于实现本文所述的遗传和治疗结果。

[0798] *需要*:如本文使用的,“需要预防”、“需要治疗”或“有此需要”的受试者指这样的受试者,其通过适当的医学从业者(例如,在人的情况下,医生、护士或执业护士;在非人哺乳动物的情况下,兽医)的判断,将合理地获益于给定治疗。

[0799] *插入*:如本文使用的,“插入”或“添加”指氨基酸或核苷酸序列的变化,导致与参考序列例如在天然存在的分子中发现的序列相比,分别向分子添加一个或多个氨基酸残基或核苷酸。

[0800] 内含子:如本文使用的,术语“内含子”指基因内的任何核苷酸序列,其在最终RNA产物(例如,mRNA)的成熟期间通过RNA剪接机制去除。内含子指基因内的DNA序列和RNA转录物(例如,前mRNA)中的相应序列两者。在RNA剪接后的最终成熟RNA中连接在一起的序列是“外显子”。如本文使用的,术语“内含子序列”指包含内含子或内含子的一部分的核苷酸序列。内含子在大多数真核生物的基因中发现,并且可以位于广泛范围的基因中,包括生成蛋白质、核糖体RNA(rRNA)和转移RNA(tRNA)的基因。当蛋白质由含有内含子的基因生成时,RNA剪接作为RNA加工途径的部分发生,所述RNA加工途径在转录之后并在翻译之前。

[0801] *脂质*:如本文使用的,术语“脂质”指具有疏水或两亲特性的小分子。脂质可以是天然存在的或合成的。脂质类别的实例包括但不限于脂肪、蜡、含固醇的代谢产物、维生素、脂肪酸、甘油脂、甘油磷脂、鞘脂、糖脂和聚酮化合物,以及孕烯醇酮脂质。在一些情况下,一些脂质的两亲特性致使其在水性介质中形成脂质体、囊泡或膜。

[0802] *局部施用*:如本文使用的,“局部施用”或“局部递送”指不依赖组合物或试剂经由血管系统转运至其预期的靶组织或部位的递送。例如,可以通过组合物或试剂的注射或植入、或者通过含有该组合物或试剂的装置的注射或植入,来递送该组合物。在靶组织或部位附近的局部施用之后,组合物或试剂、或者其一种或多种组分可以扩散至预期的靶组织或部位。

[0803] *修饰的*:如本文使用的,“修饰的”或“修饰”指结构中起因于多核苷酸例如DNA的修饰的改变状态或改变。可以以包括化学、结构和/或功能上的各种方式来修饰多核苷酸。例如,本公开内容的DNA分子可以通过掺入提供生物活性的化学修饰的碱基进行修饰。在一个实施方案中,例如,当它涉及天然核碱基腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和胸腺嘧啶(T)时,通过引入非天然或化学修饰的碱基、核苷和/或核苷酸来修饰DNA。

[0804] *mRNA*:如本文使用的,“ mRNA”指信使核糖核酸。mRNA可以是天然存在的或非天然存在的或合成的。例如,mRNA可以包括修饰的和/或非天然存在的组分,例如一个或多个核碱基、核苷、核苷酸或接头。mRNA可以包括帽结构、5’转录前导区、5’非翻译区、起始密码子、开放读码框、终止密码子、链终止核苷、茎环、发夹、多聚A序列、多聚腺苷酸化信号和/或一种或多种顺式调控元件。mRNA可以具有编码多肽的核苷酸序列。mRNA的翻译,例如mRNA在哺乳动物细胞内部的体内翻译,可以产生多肽。传统上,天然mRNA分子的基本组分至少包括编码区、5’非翻译区(5’-UTR)、3’UTR、5’帽和多聚A序列。

[0805] *天然存在的*:如本文使用的,术语“天然存在的”当应用于物体时,指物体可以在自然界中发现的事实。例如,存在于生物(包括病毒)中的多肽或多核苷酸序列(例如剪接位点)、或其组分例如氨基酸或核苷酸是天然存在的,其可以从自然界中的来源中分离,并且

未在实验室中人为地进行有意修饰。

[0806] *非同源末端连接*:如本文使用的,术语“非同源末端连接”指修复DNA中的双链断裂(DSB)的途径。NHEJ被称为“非同源的”,因为与需要同源序列来引导修复的同源性指导的修复(HDR)形成对比,DNA末端无需同源模板而直接连接。

[0807] *非复制性*:如本文使用的,术语“非复制性”指DNA分子不能在细胞或生物内复制的特征。某些DNA分子(例如质粒、病毒基因组)含有序列元件(例如复制起点),其对DNA分子赋予由细胞或生物拷贝或复制的能力。术语“非复制性”意味着不含有此类序列元件的那些DNA分子。

[0808] *核酸*:如本文使用的,术语“核酸”指以单链或双链形式的脱氧核糖核苷酸或核糖核苷酸及其聚合物或寡聚物。除非特别限制,否则该术语涵盖含有天然核苷酸的已知类似物的核酸,所述类似物具有与参考核酸相似的结合特性,并且以与天然存在的核苷酸相似的方式进行代谢。核苷酸的聚合物被称为“多核苷酸”。本公开内容的示例性核酸或多核苷酸包括但不限于核糖核酸(RNA)、脱氧核糖核酸(DNA)、DNA-RNA杂合体、RNAi诱导剂、RNAi剂、siRNA、shRNA、miRNA、反义RNA、核酶、催化DNA、诱导三螺旋形成的RNA、苏糖核酸(TNA)、二醇核酸(GNA)、肽核酸(PNA)、锁核酸(LNA,包括具有β-D-核糖构型的LNA、具有α-L-核糖构型的α-LNA(LNA的非对映异构体)、具有2’-氨基官能化的2’-氨基-LNA、以及具有2’-氨基官能化的2’-氨基-α-LNA)或其杂合体。

[0809] 本文使用的多核苷酸可以由任何多核糖核苷酸或多脱氧核糖核苷酸构成,其可以是未修饰的RNA或DNA或者修饰的RNA或DNA。例如,多核苷酸可以由以下组成:单链和双链DNA,其为单链区和双链区的混合物的DNA,单链和双链RNA,以及其为单链区和双链区的混合物的RNA,包含DNA和RNA的杂交分子,其可以是单链的或更通常地双链的,或者是单链区和双链区的混合物。另外,多核苷酸可以由包含RNA或DNA或RNA和DNA两者的三链区构成。多核苷酸还可以含有出于稳定性或其它原因进行修饰的一个或多个修饰的碱基或者DNA或RNA主链。“修饰的”碱基包括例如三苯甲基化的碱基。“修饰的核苷”包括例如肌苷和胸腺嘧啶,当后者在RNA中发现或构成RNA时。可以对DNA和RNA进行各种修饰;因此,“多核苷酸”包括化学、酶促或代谢修饰的形式。

[0810] *核酸结构*:如本文使用的,术语“核酸结构”指构成核酸(例如RNA)的原子、化学组成成分、元件、基序和/或核碱基序列的排列或组构,和/或可以指核酸的二维或三维状态。相应地,术语“RNA结构”指构成RNA分子(例如mRNA)的原子、化学组成成分、元件、基序和/或核碱基序列的排列或组构,和/或可以指RNA分子的二维和/或三维状态。基于渐增的组构复杂性,核酸结构可以进一步划分成四个组构范畴,在本文中称为“分子结构”、“一级结构”、“二级结构”和“三级结构”。

[0811] *核碱基*:如本文使用的,术语“核碱基”(可替代地“核苷酸碱基”或“含氮碱基”)指在核酸中发现的嘌呤或嘧啶杂环化合物,包括天然存在的嘌呤和嘧啶的任何衍生物或类似物,其对核酸或其一部分或区段赋予改善特性(例如结合亲和力、核酸酶抗性、化学稳定性)。腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和尿嘧啶是天然核酸中占优势地发现的主要或规范的核碱基。其它天然的、非天然的、非规范的和/或合成的核碱基可以掺入核酸内,例如本文公开的那些。

[0812] *核苷/核苷酸*:如本文使用的,术语“核苷”指与核碱基(例如,嘌呤或嘧啶)或其衍

生物或类似物共价连接的、含有糖分子(例如,RNA中的核糖或DNA中的脱氧核糖)的化合物或其衍生物或类似物。如本文使用的,术语“核苷酸”指与磷酸基共价连接的核苷。如本文使用的,术语“核糖核苷”指包含核糖和核碱基的核苷(例如,腺苷(A)、胞苷(C)、鸟苷(G)、5-甲基尿苷(m^5U)、尿苷(U)、或肌苷(I))。

[0813] *可操作地连接*的:如本文使用的,当核酸、或其片段或一部分(例如多核苷酸或寡核苷酸)置于与另一种核酸序列、或其片段或一部分的功能关系内时,它是“可操作地连接的”。

[0814] *多核苷酸/寡核苷酸*:如本文使用的,术语“多核苷酸”和“寡核苷酸”可互换使用,并且指由天然存在的碱基、糖和糖间(主链)键合组成的核苷酸或核苷单体的单链或双链聚合物或寡聚物。术语“多核苷酸”和“寡核苷酸”也包括包含非天然存在的碱基、糖和糖间(主链)键合的聚合物和寡聚物或其一部分,其类似地发挥功能。多核苷酸并不限于任何特定长度的核苷酸序列,因为术语“多核苷酸”涵盖任何长度的核苷酸的聚合形式。短多核苷酸在本领域中通常称为“寡核苷酸”。在本公开内容的上下文中,此类修饰或取代的多核苷酸和寡核苷酸经常优先于天然形式,因为修饰增加了一种或多种期望的或有益的生物学特性或活性,包括但不限于增强的细胞摄取和/或在核酸酶的存在下增强的稳定性。在一些实施方案中,本公开内容的激动剂包含含有修饰的核苷酸的至少一个区域的多核苷酸和寡核苷,其赋予一种或多种有益特性或增加生物活性(例如,增加的核酸酶抗性、增加的细胞内摄取、增加的双链体稳定性、增加的与靶多肽的结合亲和力)。

[0815] *回文序列*:如本文使用的,术语“回文序列”(可替代地“回文”)指其为自互补的核苷酸序列;其中以5’至3’方向的核苷酸序列与构成互补链的核苷酸序列当以5’至3’读取时相同。例如,序列5’-ACCTAGGT-3’是回文序列,因为当以5’至3’方向读取时,其互补序列3’-TGGATCCA-5’与原始序列相同。相比之下,序列5’-AGTGGCTG-3’不是回文序列,因为当以5’至3’方向读取时,其互补序列3’-TCACCGAC-5’与原始序列不同。

[0816] *肠胃外施用*:如本文使用的,“肠胃外施用”、“肠胃外施用的”和其它语法上等价的短语,指除肠内和局部施用外的施用模式,通常通过注射,且包括但不限于静脉内、鼻内、眼内、肌内、动脉内、鞘内、囊内、眶内、心内、皮内、腹膜内、经气管、皮下、表皮下、关节内、囊下、蛛网膜下腔、脊柱内、硬膜外、脑内、颅内、颈动脉内和胸骨内注射和输注。

[0817] *同一性百分比*:如本文使用的,在两个或更多个核酸或多肽序列的上下文中,术语“同一性百分比”指当就最大对应性进行比较并比对时,两个或更多个序列或子序列具有指定百分比的相同核苷酸或氨基酸残基,如使用下文所述的序列比较算法(例如BLASTP和BLASTN或本领域技术人员可获得的其它算法)之一或通过目视检查测量的。取决于应用,“同一性百分比”可以存在于待比较的序列的区域上,例如功能结构域上,或可替代地存在于待比较的两个序列的全长上。对于序列比较,通常一个序列充当测试序列与之比较的参考序列。当使用序列比较算法时,将测试序列和参考序列输入计算机内,需要时指定子序列坐标,并且指定序列算法程序参数。然后,序列比较算法基于指定的程序参数,计算测试序列相对于参考序列的序列同一性百分比。两个序列之间的同一性百分比是由序列共享的等同位置数目的函数(即,%同源性=等同位置#/位置总# x 100),将空位的数目和每个空位的长度考虑进去,所述空位需要引入用于两个序列的最佳比对。序列的比较和两个序列之间的同一性百分比的确定可以使用数学算法来完成,如下文非限制性实例中所述。

[0818] 用于比较的序列的最佳比对可以例如通过以下进行:Smith & Waterman,Adv. Appl. Math. 2:482(1981)的局部同源性算法,Needleman & Wunsch,J. Mol. Biol. 48:443(1970)的同源性比对算法,Pearson & Lipman,Proc. Nat'l. Acad. Sci. USA 85:2444(1988)的相似性搜索方法,这些算法的计算机化实现(Wisconsin Genetics Software Package,Genetics Computer Group,575 Science Dr.,Madison,Wis.中的GAP、BESTFIT、FASTA和TFASTA),或目测检查(一般参见Ausubel等人,下文)。

[0819] 适合于确定序列同一性和序列相似性百分比的算法的一个实例是BLAST算法,其在Altschul等人J. Mol. Biol. 215:403-410(1990)中进行描述。用于执行BLAST分析的软件可通过美国国家生物技术信息中心(National Center for Biotechnology Information)网站公开获得。两个核苷酸序列之间的同一性百分比可以使用GCG软件包(可在http://www.gcg.com处获得)中的GAP程序进行确定,使用NWSgapdna.CMP矩阵,以及40、50、60、70或80的空位权重和 1、2、3、4、5或6的长度权重。两个核苷酸或氨基酸序列之间的同一性百分比也可以使用E. Meyers和W. Miller(CABIOS,4:11-17(1989))的算法进行确定,所述算法已掺入ALIGN程序(版本2.0)内,使用PAM120权重残基表、空位长度罚分12和空位罚分4。另外,两个氨基酸序列之间的同一性百分比可以使用Needleman和Wunsch(J. Mol. Biol.(48):444-453(1970))算法进行确定,所述算法已掺入GCG软件包(可在http://www.gcg.com处获得)中的GAP程序内,使用Blossum 62矩阵或PAM250矩阵,以及16、14、12、10、8、6或4的空位权重和 1、2、3、4、5或6的长度权重。

[0820] 本公开内容的核酸和蛋白质序列可以进一步用作"查询序列",以针对公共数据库执行搜索,以例如鉴定有关序列。可以使用Altschul等人(1990)J. Mol. Biol. 215:403-10的NBLAST和XBLAST程序(版本2.0)来执行此类搜索。可以用NBLAST程序、得分 = 100,字长 = 12来执行BLAST核苷酸搜索,以获得与本发明的核酸分子同源的核苷酸序列。可以用XBLAST程序、得分 = 50、字长= 3来执行BLAST蛋白质搜索,以获得与本发明的蛋白质分子同源的氨基酸序列。为了获得用于比较目的的空位比对,可以如Altschul等人(1997)Nucleic Acids Res. 25(17):3389-3402中所述利用Gapped BLAST。当利用BLAST和Gapped BLAST程序时,可以使用各个程序(例如XBLAST和NBLAST)的缺省参数。参见http://www.ncbi.nlm.nih.gov。

[0821] *药学上可接受的*:如本文使用的,术语"药学上可接受的"指这样的化合物、材料、组合物和/或剂型,其在合理医学判断的范围内,适合与人类和动物的组织、器官和体液接触使用,而无过度的毒性、刺激、变应性应答、或其它问题或并发症,与合理的利益/风险比相称。

[0822] *药学上可接受的载体*:如本文使用的,术语"药学上可接受的载体"指并且包括生理上相容的任何和所有溶剂、分散介质、涂层、抗菌剂和抗真菌剂、等渗剂和吸收延迟剂等等。组合物可以包括药学上可接受的盐,例如酸加成盐或碱加成盐(参见例如,Berge等人(1977)J Pharm Sci 66:1-19)。

[0823] *磷酸盐/酯*:如本文使用的,术语"磷酸盐/酯"意指磷酸的盐或酯。聚磷酸盐/酯是由通过共享氧原子连接在一起的四面体PO_4(磷酸盐/酯)结构单元形成的聚合氧阴离子的盐或酯。如本文使用的,术语"二磷酸盐"指包含两个磷酸盐结构单元的聚磷酸盐。如本文使用的,术语"三磷酸盐"指包含三个磷酸盐结构单元的聚磷酸盐。

[0824] 磷酸盐生物电子等排体:如本文使用的,术语“磷酸盐生物电子等排体”(可替代地“磷酸盐模拟物”)指这样的化学取代基或基团,其具有与磷酸盐类似的物理或化学特性,并且产生与磷酸盐大体上相似的生物特性,包括二磷酸盐和三磷酸部分。在药物设计中,将一种生物电子等排体交换为另一种生物电子等排体的目的是增强化合物的所需生物或物理特性,而不产生化学结构中的显著改变。生物电子等排体的使用在药物开发中是普遍的,并且用于例如降低母体或前导化合物的毒性,改变其生物利用度,或者修饰其活性或代谢(参见例如,Rye和Baell(2005)Curr Med Chem 12(26):3127-3141;Elliot等人(2012)MedChemCom 3(7):735-751)。

[0825] *多肽*:如本文使用的,术语“多肽”、“肽”和“蛋白质”可互换使用,以指氨基酸残基的聚合物。该术语适用于其中一个或多个氨基酸残基是相应的天然存在的氨基酸的人工化学模拟物的氨基酸聚合物,以及天然存在的氨基酸聚合物和非天然存在的氨基酸聚合物。

[0826] *预防*:如本文使用的,术语“预防(preventing)”或“预防(prevent)”当与病况有关使用时,指这样的组合物施用,相对于未接受组合物的受试者,所述组合物施用降低受试者中的医学病况症状的频率、或延迟医学病况症状的发作。

[0827] *纯化的*:如本文使用的,当应用于本文所述的任何蛋白质(抗体或片段)时,术语“纯化的”或“分离的”指已与组分(例如蛋白质或者其它天然存在的生物或有机分子)分开或纯化的多肽,所述组分天然伴随其,例如表达蛋白质的原核生物中的其它蛋白质、脂质和核酸。通常,当多肽构成样品中的总蛋白按重量计的至少60(例如,至少65、70、75、80、85、90、92、95、97或99)%时,它是纯化的。

[0828] *有义链*:如本文使用的,术语“有义链”或“编码链”指双链DNA(例如基因组DNA)内的区段,其具有5’至3’方向性,并且具有与由区段转录的mRNA相同的核苷酸序列。转录产物是前mRNA转录物,其含有与有义链等同的核苷酸序列,除了尿嘧啶在其中胸腺嘧啶位于DNA中的那些位置处掺入mRNA内之外。有义链与DNA的反义链或模板链(其从3’延伸到5’)互补。

[0829] *RNA剪接*:如本文使用的,术语“剪接”(可替代地“RNA剪接”)指将细胞中的新生前体信使RNA(前mRNA)转录物加工或编辑为成熟的信使RNA(mRNA)。新生RNA转录物例如前mRNA由内含子和外显子两者构成。在剪接反应过程中,将内含子从新生RNA转录物中去除,并且将外显子连接在一起。对于核编码的基因,剪接在RNA(例如,前mRNA)转录期间或紧在其之后在核内发生。剪接在通过剪接体(snRNP的复合物)催化的一系列反应中进行,所述剪接体识别新生RNA转录物内含有的顺式作用剪接信号。由剪接体识别的规范剪接信号包括5’剪接位点、分支点、多聚嘧啶片和3’剪接位点。人分支点共有序列是YTNAY(RNA中的YUNAY),其中带下划线的腺嘌呤(A)是分支点,其中“Y”指示嘧啶核碱基(胞嘧啶(C)或胸腺嘧啶(T))的存在,并且其中N指示包含任何核碱基的核苷酸的存在(Gao等人(2008)Nucleic Acids Res 36(7):2257-2267)。剪接通过其发生的生物化学机制在本领域中得到充分描述。术语“剪接信号”、“RNA剪接信号”或“剪接基序”指被细胞的剪接机制识别的初级转录物(例如,前mRNA)内的核苷酸序列。通过在称为“剪接位点”的剪接信号处的识别和切割,从初级转录物(例如前mRNA)中去除内含子。这些位点位于内含子的5’和3’端处,并且在本文中被称为“5’剪接位点”和“3’剪接位点”。最常见地,被去除的内含子RNA序列在其5’端处以二核苷酸GU开始,并且在其3’端处以AG结束。这些共有序列已知是关键的,因为5’或3’剪接位点内的保守核苷酸之一的改变或突变导致剪接的抑制。关于DNA中的3’剪接位点的共有序

列是YAG,其中Y是C或T。关于DNA中的5'剪接位点的共有序列是GTRAG,其中R是A或G。

[0830] 称为"分支点"的剪接信号位于距离内含子3'端上游10至100个核苷酸的任何地方。分支点始终含有腺嘌呤,但它在其它方面是松散保守的。分支点共有序列是YTNAY,其中Y指示嘧啶,N指示任何核苷酸,且T指示包含胸腺嘧啶的核苷酸,且A指示包含腺嘌呤的核苷酸。

[0831] 称为"多聚嘧啶片"的剪接信号包含在前mRNA内的区域,其促进剪接机制或剪接体的组装,所述剪接体是专门用于在转录后修饰过程中进行RNA剪接的蛋白质复合物。多聚嘧啶片富含嘧啶核苷酸,尤其是尿嘧啶,且通常长15-20个碱基对,位于待剪接的内含子的3'端之前约1-40个碱基对。

[0832] 与剪接体通信的其它示例性剪接信号包括但不限于外显子剪接增强子(ESE)、外显子剪接沉默子(ESS)、内含子剪接增强子(ISE)和内含子剪接沉默子(ISS)。已知剪接部分地受嵌入外显子(ESE和ESS)和内含子(ISE和ISS)的核苷酸序列内的ESE、ESS、ISE和ISS促进或影响(参见例如,Mersch等人(2008)BMC Bioinformatics 9:369;Wang等人(2004)Cell 119:831-845;Wang等人(2012)Nat Struct Mol Biol 19(10):1044-1052;Havens等人(2013)RNA 4:247-266)。

[0833] *受试者*:如本文使用的,术语"受试者"包括任何人或非人动物。例如,本发明的方法和组合物可以用于治疗患有病症(例如遗传病症)的受试者。术语"非人动物"包括所有脊椎动物,例如哺乳动物和非哺乳动物,例如非人灵长类动物、绵羊、犬、牛、鸡、两栖动物、爬行动物等。

[0834] *治疗剂*:如本文使用的,术语"治疗剂"指这样的任何试剂,当施用于受试者时,所述试剂具有治疗、诊断和/或预防效应和/或引发所需生物和/或药理效应。

[0835] *治疗有效量*:如本文使用的,术语"治疗有效量"或"治疗有效剂量"或本文使用的类似术语,预期意指引发所需生物学或医学应答的试剂的量,例如在已经患有该疾病的患者中至少部分地阻止该状况或疾病及其并发症(例如,癌症的一种或多种症状中的改善)。对于这种用途有效的量取决于待治疗病症的严重性、以及患者的自身免疫系统的一般状态。

[0836] *治疗*:如本文使用的,术语"治疗(treat)"、"治疗(treating)"和"治疗(treatment)"指本文所述的治疗措施。"治疗"的方法采用将本公开内容的组合物施用于需要这种治疗的受试者,以便治愈、延迟病症或复发性病症、减轻其严重性或改善其一种或多种症状,或者以便延长受试者的存活超过在不存在此类治疗的情况下预计的存活。

[0837] 除非另有定义,否则本文使用的所有技术和科学术语都具有与本公开内容所属领域的普通技术人员通常理解相同的含义。尽管下文描述了优选的方法和材料,但与本文描述的那些相似或等价的方法和材料也可以用于实践或测试本文公开的方法和组合物。

[0838] 等价方案和范围

[0839] 仅使用例行实验,本领域技术人员将认识到或能够确定本文所述的具体实施方案的许多等价方案。本公开内容的范围并不预期限于上文说明书,而是如所附权利要求中阐述的。

[0840] 在权利要求中,冠词如"一个"、"一种"和"该/所述"可以意指一个/种或多于一个/种,除非上下文相反地指出或另外显而易见的。如果在给定的产品或方法中存在、采用或以

其它方式关联一个、多于一个、或所有组成员,则视为满足在组的一个或多个成员之间包括“或”的要求或描述,除非上下文相反地指出或另外显而易见的。本公开内容包括其中在给定的产品或方法中存在、采用或以其它方式关联该组的恰好一个成员的实施方案。本公开内容包括其中在给定的产品或方法中存在、采用或以其它方式关联多于一个或所有组成员的实施方案。此外,应理解,本公开内容涵盖其中将来自一个或多个所列权利要求的一个或多个限制、要素、条款、描述性术语等引入另一个权利要求内的所有变型、组合和变换。例如,可以将从属于另一个权利要求的任何权利要求修改为包括在从属于同一基础权利要求的任何其它权利要求中发现的一个或多个限制。此外,当权利要求叙述组合物时,应理解包括出于本文公开的任何目的使用该组合物的方法,以及包括根据本文公开的任何制备方法或本领域已知的其它方法制备组合物的方法。除非另有说明或除非对本领域的普通技术人员显而易见的是将出现矛盾或不一致。

[0841] 当要素作为例如以Markush组格式的列表呈现时,应理解,还公开了要素的每个子组,并且可以从该组中去除任何要素。应当理解,一般而言,当本发明或本发明的方面称为包含特定要素、特征等时,本发明的某些实施方案或本发明的方面由此类要素、特征等组成或基本上由其组成。为了简单起见,这些实施方案没有在本文中具体地书面阐述。

[0842] 还应注意的是,术语“包含”预期是开放的,并且允许但不要求包括另外的要素或步骤。当术语“包含”在本文中使用时,术语“由……组成”因此也被涵盖且公开。

[0843] 当给出范围时,包括了端点。此外,应理解,除非另外说明或从上下文和本领域普通技术人员的理解显而易见的,否则在本发明的不同实施方案中,表示为范围的值在本发明的不同实施方案中可以采取所述范围内的任何具体值或子范围,至范围下限的单位的十分之一,除非上下文另有明确规定。

[0844] 另外,应理解,落入现有技术内的本发明的任何特定实施方案可以从任何一个或多个权利要求中明确地排除。由于此类实施方案被视为本领域普通技术人员已知的,因此即使在本文中没有明确地阐述排除,也可以将它们排除。本发明的组合物的任何特定实施方案(例如,由此编码的任何核酸或蛋白质;任何生产方法;任何使用方法;等)都可以出于任何原因从任何一个或多个权利要求中排除,无论是否与现有技术的存在有关。

[0845] 所有引用的来源,例如,本文引用的参考文献、出版物、数据库、数据库条目和技术,都引入本申请作为参考,即使在引用中没有明确地陈述。在引用来源的陈述与本申请冲突的情况下,以本申请中的陈述为准。

实施例

[0846] 本公开内容通过参考下述实施例将得到更充分地理解。然而,它们不应被解释为限制本公开内容的范围。应理解,本文描述的实施例和实施方案仅用于说明性目的,并且鉴于其的各种修改或改变将被建议给本领域技术人员,并且被包括在本申请的精神和范围以及所附权利要求的范围内。

[0847] 实施例1:设计供体多核苷酸以调节外显子限定并校正gDNA分子中的突变

[0848] RNA剪接是所有真核生物中共享的过程,其允许通过内含子的去除和外显子的连接,将转录的RNA加工成mRNA。在其中外显子较小而内含子较大的高等真核生物,如脊椎动物中,剪接机制将外显子用作识别单位。识别且配对跨越外显子的剪接位点的这一过程称

为外显子限定,解释了脊椎动物内部外显子中可见的窄尺寸分布(平均大小为134nt)。3'和5'剪接位点之间的距离对于外显子的准确识别和包括是关键的。先前的研究已显示了,剪接体形成和外显子限定受内含子和外显子大小的影响。使用cDNA序列将组成型内部外显子的3'和5'剪接位点之间的距离扩展到>300nt已显示,由于剪接体限定外显子的能力的低效率,导致弱外显子识别(Sterner等人(1996)Proc Natl Acad Sci USA 93(26):15081-15085)。

[0849] 供体多核苷酸设计为通过改变一个或多个核苷酸,确定供体多核苷酸调节外显子限定且改变gDNA分子中的核苷酸序列(例如,校正或诱导外显子或内含子中的突变)的有效性。至少基于下述标准来设计供体多核苷酸:选择改变(例如,校正或诱导突变)的核苷酸序列;并且选择一种或多种剪接信号,以调节剪接位点识别或外显子限定。剪接信号包括分支点、多聚嘧啶片、3'剪接位点、5'剪接位点、外显子或内含子剪接增强子或沉默子、或其组合中的至少一种或多种。

[0850] 在插入gDNA内时,供体多核苷酸的第一部分或区段包括包含一个或多个核苷酸的核苷酸序列,所述一个或多个核苷酸以所需方式改变gDNA核苷酸序列(例如,校正或诱导突变),并且供体多核苷酸的第二部分或区段包含所需的一种或多种剪接信号,用于将构成所需改变序列的一个或多个核苷酸准确地包括到最终的mRNA转录物内。例如,通过供体多核苷酸改变gDNA核苷酸序列,以校正或诱导表达的基因中的特异性突变,并且同时提供所需的一种或多种剪接信号,以确保将校正或诱导的突变转录成mRNA。将供体多核苷酸引入细胞内,并且确定细胞的NHEJ DNA修复机制将供体多核苷酸插入DSB内的效率,所述DSB通过定点核酸酶(例如Cas核酸酶)引入gDNA内,以及确定改变的核苷酸序列转录成mRNA的效率。

[0851] 例如,如描绘供体多核苷酸的有义链的图1中的示意图中所示,供体多核苷酸设计为通过插入双链断裂(DSB)内来校正外显子中的有害或致病突变(由X指示),所述DSB使用定点核酸酶(例如,Cas9)在致病突变附近引入。通过NHEJ介导的供体多核苷酸(以黑色显示)插入来修复DSB,所述供体多核苷酸包含校正突变的一个或多个核苷酸(由*指示),并且包含一种或多种剪接信号(例如3'剪接位点、多聚嘧啶片和/或分支点)。将供体多核苷酸插入DSB内生成具有用于外显子的两个潜在3'剪接位点的基因。

[0852] 不受理论的束缚,供体多核苷酸的插入有效地抑制或破坏上游(内源)3'剪接位点参与外显子限定的能力,部分地通过增加上游(内源)3'剪接位点和下一个可用的5'剪接位点之间的核苷酸(由//指示)数目。在供体多核苷酸插入后,剪接机制无法将内源3'和5'剪接位点识别为限定外显子的一对。相反,剪接机制将包含供体多核苷酸(以黑色显示)的新的3'剪接位点和下一个可用的5'剪接位点识别为限定外显子的一对剪接位点,因为这些剪接位点这样放置,使得它们与RNA剪接的外显子限定理论一致。此外,在校正突变(由*指示)的一个或多个核苷酸附近和上游存在新的3'剪接位点(以及任选地另外的剪接信号),导致编码校正突变的外显子的限定及其包括到mRNA转录物内。

[0853] 实施例2:通过NHEJ插入DSB内的供体多核苷酸改变了gDNA核苷酸序列,并且在外显子中引入密码子变化

[0854] 先前的工作已显示,通过使用供体多核苷酸来通过HDR修复DSB的基因组DNA(gDNA)中的突变校正,在分裂细胞中以比未分裂细胞中更高的速率发生(数据未显示)。尽管HDR修复在非分裂细胞(例如处于细胞周期的G0或G1期中的细胞)中经常是低效率的,但

NHEJ修复途径在分裂细胞和非分裂细胞两者中仍是活跃的DSB修复途径。因此,利用NHEJ修复途径用于将供体多核苷酸插入DSB内,对于在分裂细胞和非分裂细胞两者中引入突变的位点特异性校正是理想的。与使用供体多核苷酸通过HDR修复DSB(其中所述供体多核苷酸包含将供体多核苷酸定向在单个方向上用于DSB的修复的同源臂)不同,使用供体多核苷酸通过NHEJ修复机制的DSB修复,可以导致供体多核苷酸在两个取向之一上插入DSB内;正向或反向。为了评估细胞使用供体多核苷酸经由NHEJ DNA修复机制来修复gDNA中CRISPR/Cas9介导的DSB,并且以所需方式改变gDNA的核苷酸序列的能力,确定NHEJ介导的供体多核苷酸在导致所需改变(例如,校正编辑)的取向上插入(与通过同源重组的供体多核苷酸整合相反)分裂的肝细胞衍生细胞的gDNA内。

[0855] NHEJ介导的供体多核苷酸在产生掺入所需核苷酸序列改变的取向上插入DSB内,在本文中被称为"校正编辑"或"校正插入"。"校正插入"包括"完全插入",其中实现插入而不对校正DSB周围的gDNA序列或插入供体多核苷酸的末端引入突变(例如,将供体多核苷酸无缝插入gDNA DSB内)。"校正插入"还包括NHEJ介导的插入,其中在校正DSB周围的gDNA序列内、在插入供体多核苷酸的末端处、或在插入供体多核苷酸内的内部序列处出现一个或多个小突变,但仍得到编码所需核苷酸序列改变(例如,由供体多核苷酸编码的基因校正和/或剪接信号)的gDNA序列。

[0856] 基于实施例1中阐述的标准来设计供体多核苷酸,以在编码外显子中已知引起疾病的位置处引入核苷酸改变。通过引入密码子改变来改变核苷酸序列可用于校正编码外显子中的致病突变。例如,葡萄糖-6-磷酸酶催化亚基(*G6PC*)基因的外显子2中的R83C突变,已知在人中引起糖原贮积病1a(GSD1a)。校正导致GSD1a的R83C突变的治疗方法是在外显子2中引入密码子变化,其导致G6PC多肽中的C83R恢复,并且提供一种或多种剪接信号,以指导回复密码子包括到编码G6PC多肽的mRNA内。作为用于验证校正R83C突变的这一治疗策略的原理证明,供体多核苷酸设计为改变*G6PC*基因中的外显子2的核苷酸序列,其中将R83C突变引入野生型*G6PC*基因的外显子2内。这通过设计且测试供体多核苷酸来完成,在所述供体多核苷酸中,对应于人G6PC多肽中的R83的野生型*G6PC*基因中的精氨酸(R)密码子(CGT)突变为致病性半胱氨酸(C)密码子(TGC)。

[0857] 简言之,选择单个gRNA(sgRNA),其包含对于葡萄糖-6-磷酸酶催化亚基(*G6PC*)基因的外显子2特异性的间隔区(sgRNA CH32,SEQ ID NO:6;sgRNA CH32间隔区,SEQ ID NO:81;G6PC靶基因序列 + PAM,SEQ ID NO:82)。具有硫代磷酸酯键和2'-O-甲基核苷酸残基的修饰的sgRNA可以用于生成所需的靶基因DSB。制备了其为长度25至125个核苷酸的双链寡核苷酸(dsODN)的供体多核苷酸,其包含2种序列元件:(1)外显子序列,其在对应于G6PC多肽的位置83的外显子2的开放读码框中掺入单个密码子变化;以及(2)一种或多种剪接信号,其选自单独、或与多聚嘧啶片组合、或进一步与分支点组合的3'剪接位点(如表1中所示)。

[0858] CH32 sgRNA的靶基因序列显示于表4中,CH32 sgRNA间隔区序列显示于表3中,而CH32 sgRNA序列显示于表2中。Cas9/sgRNA在距离PAM序列起点上游的三个核苷酸处形成DSB。由dsODN供体多核苷酸编码的外显子序列在距离PAM序列起点下游的三个核苷酸处与靶基因序列重叠,使得dsODN供体多核苷酸插入切割位点内,伴随所需的剪接信号以及恢复具有编码的TGC突变的外显子序列。

[0859] 将sgRNA和每种供体多核苷酸转染到组成型表达SpCas9的HuH-7细胞(肝细胞衍生的细胞癌细胞系)内。使用包含阳离子脂质体的转染试剂(Lipofectamine®MessengerMax™;ThermoFisher Scientific,Waltham,MA),将sgRNA和供体多核苷酸同时转染到HuH-7细胞内。用含有34ng sgRNA、200ng dsDNA供体多核苷酸、以及由制造商建议的转染试剂量的50μl溶液,转染96孔板中的每个孔的细胞。

[0860] 如表1中所示,测试了渐增长度(25-125)并包含一种或多种剪接信号序列的供体多核苷酸。在转染后48小时后,根据制造商的说明,使用商购可得的DNA提取试剂盒(PrepGem;Zygem,Charlottesville,VA),从转染的HuH-7细胞中分离基因组DNA(gDNA)。为了确定供体多核苷酸是否通过NHEJ插入通过CH32 sgRNA:spCas9复合物(CH32靶位点,SEQ ID NO:82)生成的DSB内,从分离的gDNA生成PCR扩增子并通过下一代测序(NGS)进行分析。在Illumina NextSeq机器(Illumina,Inc.,San Diego,CA)上,使用高达300 nt的配对端运行来执行NGS测序。通过标准双重条形码来鉴定样品。有限循环的第一轮PCR扩增了对应于CH32靶位点周围~800bp区域的扩增子。第二轮使用先前生成的扩增子,并且生成待测序的核心扩增子,侧接CH32靶位点的~200bp。在最后一轮(第三轮)PCR中添加了NGS所需的衔接子。

[0861] 在核心扩增子的配对端测序后,分析所得到的FASTQ文件,以确定插入率。简言之,使用PANDAseq程序(Masella等人(2012)BMC Bioinformatics 13:31)连接读数的端部,并且使用FASTX-Toolkit(http://hannonlab.cshl.edu/fastx_toolkit/index.html)中提供的FASTQ Quality Filter程序,鉴定且去除质量差的读数。将等同的读数制成表格,合并,然后搜索起因于整个供体多核苷酸插入CH32切割位点内的预测序列。展示指示供体多核苷酸的预测5'和3'连接两者、以及外显子2中的单个密码子变化的测序读数被视为对于校正插入呈阳性,将这些与总序列读数进行比较,并且结果在表1中显示为%校正插入。

[0862] 表1. 用于在*G6PC*基因中引入密码子变化的单向dsODN供体多核苷酸

[0863]

供体多核苷酸(大小)	%校正插入	核苷酸序列(5'-3'，有义链)	SEQ ID NO
CH32_25-0 (25 nt)	0.012%	*tag*GATTCTCTTTGGACAGTGCCCT	16
CH32_50-0 (50 nt)	8.088%	cccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	18
CH32_75-0 (75 nt)	0.019%	tgggcaaaagcattcattcagtaaccccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	10
CH32_100-0 (100 nt)	0.015%	acactcttcttgaaggtgtaggctttgggcaaaagcattcattcagtaaccccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	12
CH32_125-0 (125 nt)	0.000%	caacatgtgaaatccttctcaggctacactcttcttgaaggtgtaggctttgggcaaaagcattcattcagtaaccccagaaacttgttctgtttttccca*tag*GATTCTCTTTGGACAGTGCCCT	14

供体多核苷酸符号表	
注释	供体多核苷酸设计元件
小写字母	内含子序列
斜体小写字母	3'剪接位点
实线下划线小写字母	多聚嘧啶片
虚线下划线小写字母	分支点序列
大写字母	外显子序列
实线下划线大写字母	半胱氨酸密码子
粗体大写字母	所需的改变核苷酸

[0864] 表2．用于指导*G6PC*基因中的位点特异性DSB的sgRNA序列

[0865]

名称	sgRNA 核苷酸序列(5'-3')	SEQ ID NO
CH32 sgRNA	UCUUUGGACAGCGUCCAUACGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	6
CH34 sgRNA	UGGACAGCGUCCAUACUGGUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	7
CH36 sgRNA	GUAUCCAAAACCCACCAGUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	8
CH42 sgRNA	GUCAGUCUCACAGGUUACAGGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	70

[0866] 表3. gRNA间隔区序列

[0867]

名称	sgRNA间隔区序列	SEQ ID NO
CH32 sgRNA间隔区	UCUUUGGACAGCGUCCAUAC	81
CH34 sgRNA间隔区	UGGACAGCGUCCAUACUGGU	83
CH36 sgRNA间隔区	GUAUCCAAAACCCACCAGUA	85
CH42 sgRNA间隔区	GUCAGUCUCACAGGUUACAG	88

[0868] 表4. *G6PC*基因中的sgRNA靶基因序列

[0869]

名称	靶基因(PAM)序列	SEQ ID NO
CH32靶基因	TCTTTGGACAGCGTCCATAC(TGG)	82
CH34靶基因	TGGACAGCGTCCATACTGGT(GGG)	84
CH36靶基因	GTATCCAAAACCCACCAGTA(TGG)	86
CH42靶基因	GTCAGTCTCACAGGTTACAG(GGG)	87

[0870] 结果证实了,供体多核苷酸可以通过分裂细胞中的NHEJ DNA修复机制精确地插入Cas9介导的DSB的位点内,允许在gDNA中的特定位置处以所需方式改变核苷酸序列。另外,数据提示了,在这些条件下用于插入的供体多核苷酸长度的下限和上限分别为长度约25 nt和长度约75 nt,并且根据测试的那些,用于校正编辑插入的供体多核苷酸的最佳长度为长度约50 nt。

[0871] 为了进一步评估供体多核苷酸长度和gRNA靶位点的位置对细胞经由NHEJ DNA修复机制,以导致gDNA中的所需核苷酸序列改变的方式,将供体多核苷酸插入gDNA中CRISPR/

Cas9介导的DSB内的能力的作用,如表2中所示选择对于G6PC外显子2特异性的三种不同sgRNA(CH32 sgRNA(SEQ ID NO:6)、CH34 sgRNA(SEQ ID NO:7)和CH36 sgRNA(SEQ ID NO:8);每种gRNA靶向G6PC外显子2内的不同位点,其靶基因序列显示于表4中)。

[0872] 将sgRNA与各种长度的供体多核苷酸(如表5中阐述的)组合,并且确定了将供体多核苷酸插入分裂的肝细胞衍生细胞的基因组DNA内的不同位置内的能力。简言之,如上文,用对于G6PC外显子2特异性的三种不同的sgRNA(表5中的CH32 sgRNA、CH34 sgRNA和CH36 sgRNA),转染组成型表达SpCas9的HuH-7细胞。sgRNA与相应的供体多核苷酸共转染,每种供体多核苷酸至少包含3'剪接位点和外显子序列,所述外显子序列设计为制备G6PC基因的外显子2中的R至C的单个氨基酸变化。在如上所述转染之后,评估了供体多核苷酸插入通过与SpCas9复合的每种相应sgRNA生成的DSB内。sgRNA和供体多核苷酸的核苷酸序列分别显示于表2和表5中。在转染后48小时后,处理细胞以获得gDNA,并且如上所述,通过NGS分析确定掺入由供体多核苷酸编码的校正编辑的总序列读数的百分比。

[0873] 表5. 用于在*G6PC*基因中引入密码子变化的dsODN供体多核苷酸的序列

[0874]

供体多核苷酸(大小)	供体核苷酸序列(5'-3', 有义链)	SEQ ID NO
En CH32 50-0(50 nt)	ttcataaacttgttctgttttttta*tag*GATTCTCTTTGGACAGTGCCCT	20
En CH32 55-0(55 nt)	ttcatcccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	22
En CH32 60-0(60 nt)	attcattaaccccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	24
CH32 65-0(65 nt)	cattcattcagtaaccccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	26
CH32 70-0(70 nt)	aaaagcattcattcagtaaccccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCT	28
En CH34 54-0(54 nt)	ttcataaacttgttctgttttttta*tag*GATTCTCTTTGGACAGTGCCCTTACT	30
En CH34 59-0(59 nt)	ttcatcccagaaacttgttctgtttttcca*tag*GATTCTCTTTGGACAGTGCCCTTACT	32

[0875]

En CH34 64-0(64 nt)	attcattaaccccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTACT	34
CH34 69-0(69 nt)	cattcattcagtaaccccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTACT	36
CH34 74-0(74 nt)	aaaagcattcattcagtaaccccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTACT	38
En CH36 53-0(53 nt)	ttcataaacttgttctgttttttatagGATTCTCTTTGGACAGTGCCCTTAC	40
En CH36 58-0(58 nt)	ttcatcccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTAC	42
En CH36 63-0(63 nt)	attcattaaccccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTAC	44
CH36 68-0(68 nt)	cattcattcagtaaccccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTAC	46
CH36 73-0(73 nt)	aaaagcattcattcagtaaccccagaaacttgttctgtttttccatagGATTCTCTTTGGACAGTGCCCTTAC	48

供体多核苷酸符号表

注释	供体多核苷酸设计元件
小写字母	内含子序列
斜体小写字母	3'剪接位点
实线下划线小写字母	多聚嘧啶片
粗体小写字母	增强嘧啶束的核苷酸取代
虚线下划线小写字母	分支点序列
大写字母	外显子序列
实线下划线大写字母	半胱氨酸密码子
粗体大写字母	所需的改变核苷酸

[0876] 图2A-2C显示了如上所述确定的校正插入百分比,其中显示了在NHEJ介导的供体多核苷酸dsODN插入DSB内之后的校正编辑百分比,所述DSB通过Cas9以及由SEQ ID NO:6(图2A)、SEQ ID NO:7(图2B)或SEQ ID NO:8(图2C)描绘的sgRNA在*G6PC*基因中诱导。图2A-2C中所示的数据证实了,将测试的供体多核苷酸插入gDNA内并不限于单个sgRNA,并且不依赖于靶位点而发生。另外,结果显示了,长度70个核苷酸或更长的供体多核苷酸并未以可观的频率插入DSB内,进一步提示了在这些条件下用于插入DSB内的供体多核苷酸的长度上限为长度约70个核苷酸,与表1中所示的结果一致。然而,在如下文详述的后续研究中,具有更长长度的dsODN供体多核苷酸允许以与50nt dsODN可比较的效率插入DSB内。如实施例中进一步描述的,关于更长长度的dsODN供体多核苷酸的插入效率可能取决于sgRNA、待靶向的基因基因座以及dsODN试剂的制备。

[0877] 实施例3:供体多核苷酸插入非分裂的原代人肝细胞的*G6PC*基因内调节G6PC前mRNA剪接

[0878] 为了评估供体多核苷酸插入*G6PC*基因座内对表达*G6PC*的非分裂的原代人肝细胞(PHH)中的*G6PC*前mRNA剪接的作用,选择并测试了CH34_54-0供体多核苷酸(CH34_54-0;SEQ ID NO:30),其在分裂的HuH-7细胞中提供了最高水平的插入。CH34_54-0供体多核苷酸包含编码与外显子序列紧靠的3'剪接位点的内含子序列,所述外显子序列对应于包含密码子的*G6PC*外显子2的5'端,所述密码子在插入时在外显子2编码区中产生氨基酸改变(R83C)。

[0879] 简言之,在24孔板中与小鼠成纤维细胞共培养的PHH用转染溶液中的各种浓度的CH34_54-0进行转染,所述转染溶液包括编码SpCas9的mRNA(600 ng)和靶向G6PC外显子2的sgRNA(CH34;SEQ ID NO :7;200ng)。测试了dsODN一系列浓度的供体多核苷酸(从0.56 ng到10 ng)。Lipofectamine RNAiMAX Transfection Reagent如由制造商(ThermoFisher,USA)推荐的使用,以促进细胞的转染。在转染后48小时后,按照制造商的说明(Qiagen,USA),使用AllPrep DNA/RNA Mini Kit 来同时分离总RNA和gDNA。为了测量mRNA中的基因变化的掺入,对*G6PC*外显子1和外显子3特异性的PCR引物用于产生用于NGS分析的cDNA扩增子,以确定包含G6PC外显子2的cDNA扩增子的百分比(%校正编辑),所述G6PC外显子2编码在编码区中的氨基酸变化(R83C)。还处理细胞以获得gDNA,并且通过NGS分析进行分析,以确定供体多核苷酸的校正插入%。将对于校正编辑呈阳性的序列读数(即:具有适当的剪接信号和所需的核苷酸变化)与总序列读数进行比较,以确定校正编辑的百分比。

[0880] 图3显示了起因于NHEJ介导的供体多核苷酸在CH34 sgRNA靶位点(SEQ ID NO:84)处插入*G6PC*外显子2 gDNA基因座(DNA)内的校正编辑的百分比(如上所述确定)。含有另外的剪接信号和ORF中的核苷酸变化的供体多核苷酸插入,导致*G6PC* DNA中的校正编辑(黑线),并且还导致表达含有预测基因编辑的mRNA的*G6PC*基因(灰线)。还显示了校正编辑的水平是存在的供体多核苷酸量(供体的ng)的函数。在DNA水平上,数据证实了供体多核苷酸的插入中的剂量依赖性增加,并且在最高测试浓度下,大约10%的*G6PC*等位基因具有正确插入非分裂的原代人肝细胞中的Cas9介导的DSB内的供体多核苷酸。与使用同源供体多核苷酸和基于HDR的修复方法的基因校正相比,这表示细胞中的校正编辑事件的百分比中的超过15倍增加。先前,HDR方法在非分裂细胞中仅实现~0.5%的校正编辑(数据未显示)。另外,用最高浓度的供体转染的细胞也显示了在成熟的G6PC RNA中相似水平的校正编辑(>15%;图3)。这些数据证实了,将CH34_54-0供体多核苷酸插入*G6PC*基因中的特异性Cas9介导的DSB内,导致改变的多核苷酸序列(即,所需的密码子改变)以剂量依赖性方式包括到成熟的*G6PC* mRNA内。

[0881] 实施例4:用供体多核苷酸治疗的小鼠在体内调节G6PC前mRNA剪接

[0882] 为了在体内评估NHEJ介导的供体多核苷酸在*G6PC*基因座内的插入和校正编辑的效应,用编码SpCas9的mRNA、靶向鼠*G6PC*外显子2的sgRNA(小鼠CH34;SEQ ID NO:7)、以及实施例2中所述的CH34_54-0供体多核苷酸(SEQ ID NO:30)来治疗小鼠。

[0883] 简言之,个别地配制基于类脂质的纳米颗粒(LNP),以含有CH34 sgRNA的鼠同源物(鼠CH34;SEQ ID NO:89)、编码SpCas9的mRNA或供体多核苷酸CH34_54-0(SEQ ID NO:30)。分开地配制LNP,以含有sgRNA、mRNA和供体多核苷酸,各自为20:1的核酸:脂质比率,然后在通过注射施用之前混合LNP。小鼠(C57BL/6,6-8周龄,5只动物/组)用下述三种治疗之一在

尾静脉内进行注射:PBS,含有以0.5 mg/kg的鼠CH34 gRNA LNP和以0.5 mg/kg的SpCas9 mRNA LNP的LNP混合物,或含有以0.5 mg/kg的鼠CH34 gRNA LNP、以0.5 mg/kg的SpCas9 mRNA LNP和以0.5mg/kg的CH34_54-0供体多核苷酸LNP的LNP混合物。在注射后96小时后,收获小鼠肝脏,提取gDNA,执行NGS分析并确定校正编辑,如实施例3中所述。

[0884] 图4A显示了如指示的,起因于NHEJ介导的供体多核苷酸插入的校正编辑百分比(如上所述确定)。另外,在分开的实验中,如实施例2中所述,在gDNA和mRNA转录物两者中,确定了起因于供体多核苷酸的完全插入的校正编辑百分比。如图4B中所示,在gDNA和mRNA转录物的水平上,可见增加水平的校正编辑。值得注意的是,校正编辑的掺入在mRNA转录物的水平上是增加的。已知一定比例的肝细胞具有多重核/细胞,其似乎具有不同的表达水平和模式(Kreutz等人(2017)*Front Physiol.* 8:862)。不受理论的束缚,认为在更活跃转录的G6PC等位基因中,gDNA的校正可以得到增强。如果仅对活跃转录的gDNA进行基因编辑,则编码靶基因并掺入校正编辑的gDNA的百分比将低于编码校正编辑的相应mRNA的百分比。图4A-4B中所示的数据证实了LNP可以递送所有三种组分,并且NHEJ介导的供体多核苷酸插入在体内小鼠肝细胞中发生,导致所需的基因编辑(即,校正编辑)。

[0885] 实施例5:双向供体多核苷酸改善了NHEJ介导的肝细胞衍生细胞中的G6PC或GAA插入和校正

[0886] 当供体多核苷酸在单个(正向)取向上插入时,实施例1-4中描述的供体多核苷酸设计为导致校正插入。然而,考虑到通过NHEJ-修复途径插入DSB内的供体多核苷酸(例如,供体dsODN)可以在正向方向或反向方向上插入,理想的是,不管供体多核苷酸插入的方向如何,都掺入所需的基因编辑和剪接信号。因此,设计了双向供体多核苷酸,当在正向取向或反向取向上插入时,所述双向供体多核苷酸将编码校正编辑和所需的剪接信号,如图5中所示。

[0887] 为了确定双向多核苷酸在肝细胞衍生细胞中的有效性,设计了一组供体多核苷酸,用于双向插入*G6PC*基因或*GAA*基因中,所述*GAA*基因当突变时与庞贝病相关。*GAA*基因中的突变在内含子中发生并影响剪接,因此设计具有嘧啶束和分支点的供体多核苷酸。分支点专门设计为当供体在正向取向或反向取向上插入时发挥功能。因为*G6PC*供体多核苷酸需要外显子序列来校正突变,所以CH42_25-0供体的长度不允许使用双向供体,因此该供体仅以单向方式插入。

[0888] 选择了靶向含有与成人发作型庞贝病相关的最常见突变(IVS 1(-13T>G))的区域的sgRNA,GAA-5 sgRNA(SEQ ID NO:61),并且在表6中进行描绘。IVS 1(-13T>G)突变破坏了GAA基因的外显子2的嘧啶束。破坏导致外显子2跳跃和隐蔽剪接位点的激活,其负面影响GAA基因表达和基因产物的酶促活性。表6中描绘的GAA-5 sgRNA靶向野生型GAA基因中发生该突变的区域(SEQ ID NO:61)。

[0889] 表6. 用于指导*GAA*基因中的位点特异性DSB的gRNA

[0890]

sgRNA	核苷酸序列(5'-3')	SEQ ID NO
WT GAA 靶基因(PAM)	AGCCCGCTTTCTTCTCCCGC(AGG)	90
WT GAA5 sgRNA 间隔区	AGCCCGCUUUCUUCUCCCGC	91
WT GAA5 sgRNA	AGCCCGCUUUCUUCUCCCGCGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUUUU	61
突变体(Mt) GAA 靶基因 (PAM)	AGCCCGCTTGCTTCTCCCGC(AGG)	92
Mt GAA sgRNA 间隔区	AGCCCGCUUGCUUCUCCCGC	93
Mt GAA sgRNA	AGCCCGCUUGCUUCUCCCGCGUUUUAGAGCUAGAAAU AGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUUUU	94

[0891] 设计用于插入*GAA*基因内的双向供体dsODN显示于表7中。在5'至3'方向上,供体dsODN序列包含多聚嘧啶片CTTCTTCTCTTCTTCC(SEQ ID NO:55)的反向和互补体、任选地分支点序列TACTGAC(SEQ ID NO:52)反向和互补体、分支点序列TATTAAC和多聚嘧啶片TTTTTTCTTTTT(SEQ ID NO:54)。由YAG/G(其中Y是胸腺嘧啶或胞嘧啶,而/指示外显子和内含子之间的边界)编码的3'剪接位点位于gDNA中。包含分支点序列和多聚嘧啶片的双向供体的插入允许在gDNA中的3'剪接位点处的正确剪接。例如,由SEQ ID NO:63描绘的50nt dsODN的有义序列和反义序列显示于图6A中。不管50nt dsODN是在正向方向还是反向方向上插入,所得到的基因都编码所需的校正编辑(例如,掺入分支点序列和校正的多聚嘧啶片),如图6B中所示。

[0892] 值得注意的是,在dsODN供体的设计中,在正向方向上编码的剪接信号并不编码与在反向方向上编码的等价剪接信号相同的序列。例如,在正向方向上编码的多聚嘧啶片TTTTTTCTTTTT(SEQ ID NO:54)不同于在反向方向上的多聚嘧啶片CTTCTTCTCTTCTTCC(SEQ ID NO:55),尽管两者均符合多聚嘧啶片共有序列。不受理论的束缚,这种设计元件的目的是避免可能阻碍插入的dsODN的自退火、或避免由具有插入dsODN的基因转录的mRNA的自退火。mRNA的此类自退火可以导致弱mRNA表达水平。因此,对于在正向方向上的给定剪接信号,在反向方向上的等价剪接信号具有不同的序列,但两者均符合实现转录mRNA的剪接所需的给定剪接信号的共有序列。

[0893] 表7. 用于将剪接信号插入*GAA*基因内的双向dsODN供体多核苷酸

[0894]

供体多核苷酸(大小)	核苷酸序列(5'-3')	SEQ ID NO
GAA 25-0_F(25 nt)	gagaagaagTATTAACttttttct	62
GAA_50-0_F(50 nt)	tggaagaagagaagaagctgggTATTAACgcatttttttcttttttaattc	63
GAA_75-0_F(75 nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgTATTAACgcaaatgaattttttttcttttttaattc	64
GAA_100-0_F(100 nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgagccaccgcgccagccatgcaaatcTATTAACgcaaatgaattttttttcttttttaattc	65

供体多核苷酸符号表	
注释	供体多核苷酸设计元件
小写字母	内含子序列
虚线下划线小写字母	定界符序列
粗体小写字母	3'剪接位点和多聚嘧啶片
双下划线小写字母	多聚嘧啶片和3'剪接位点的反向和互补体
实线下划线大写字母	在任一方向上符合分支点共有区的序列
双下划线大写字母	分支点序列 TACTGAC 的反向和互补体

[0895] *CH42 25-0被设计为单向供体多核苷酸。

[0896] 表2和表8中分别阐述了用于NHEJ介导的插入*G6PC*基因内的sgRNA和双向供体多核苷酸(除了其为单向对照的ch42 25-0之外)的组合。双向供体dsODN设计为在插入*G6PC*基因的外显子2的3'端附近诱导的切割位点内时,掺入*G6PC*基因的上游5'剪接位点和野生型外显子编码序列。在5'至3'方向上,双向供体dsODN设计为掺入外显子编码序列、5'剪接位点、5'剪接位点GTGAGT(SEQ ID NO:XX)的反向和互补体、以及外显子编码序列的反向和互补体。不管供体dsODN是在正向方向还是反向方向上插入,所得到的基因都编码所需的校正编辑(例如5'剪接位点上游的外显子编码序列)。

[0897] 表8. 用于在*G6PC*基因中引入密码子变化的双向dsODN供体多核苷酸

[0898]

供体多核苷酸(大小)	核苷酸序列(5'-3')	SEQ ID NO
CH42 0-25*(25 nt)	AACCTGTGAGACTGGACCAG*gtaag*	71
CH42 0-50(50 nt)	AACCTGTGAGACTGGACCAG*gtaag*actcac**ACTTGCGAAACCGGCCCAG**	73
CH42 0-75(75 nt)	AACCTGTGAGACTGGACCAG*gtaagc*gacgcgcatttctcacacggcaggactcac**ACTTGCGAAACCGGCCCAG**	75
CH42 0-100(100 nt)	AACCTGTGAGACTGGACCAG*gtaagc*gacgcgcatttctcacacggcagggagggccacacgcgtttgtttctcaactcac**ACTTGCGAAACCGGCCCAG**	77
CH42 0-125(125 nt)	AACCTGTGAGACTGGACCAG*gtaagc*gacgcgcatttctcacacggcagggagggccacacgcgtttgtttctcacacgatgggcagggcgacacatgttactcac**ACTTGCGAAACCGGCCCAG**	79

供体多核苷酸符号表

注释	供体多核苷酸设计元件
小写字母	内含子序列
斜体小写字母	5'剪接位点
实线下划线小写字母	5'剪接位点的反向和互补体
虚线下划线小写字母	定界符序列
大写字母	外显子序列
粗体大写字母	外显子序列的反向和互补体

[0899] 确定了在*GAA*或*G6PC*基因中的任一方向(正向或反向)上的插入百分比。简言之,如实施例2中所述,用20ng供体多核苷酸、sgRNA和编码SpCas9的mRNA独立地转染HuH-7细胞。Huh-7细胞表达SpCas9,但用编码SpCas9的mRNA进行进一步转染,以改善编辑效率。如上所述,确定了NGS分析和校正插入百分比的确定。

[0900] 图7A-7D描绘了靶向Huh-7细胞中的*G6PC*或*GAA*的双向供体多核苷酸的校正编辑百分比。结果显示了,供体多核苷酸的双向性急剧增加了校正插入的百分比。

[0901] 图7A和图7C显示了当供体多核苷酸在正向方向上(类似于单向供体多核苷酸)插入*G6PC*基因(图7A)和GAA基因(图7C)中时的校正编辑百分比。此外,图7B和图7D证实了,使用双向插入(在正向方向或反向方向上的插入),校正插入的百分比显著增加。双向多核苷酸显示了在25至75个核苷酸之间的长度下的显著插入,其中50个核苷酸显示了最高的插入率。双向50-0供体多核苷酸总体上显示出最高的校正插入百分比。与实施例1-3中观察到的校正插入一致,在用长度在25-75nt之间且至多100nt的双向供体多核苷酸转染的细胞中,观察到显著水平的校正编辑,所述双向供体多核苷酸由不同于上文显示的靶引导的sgRNA(例如,CH42sgRNA;图7A-7B)靶向*G6PC*基因、或人*GAA*基因(图7C-7D)。

[0902] 进一步评估了50nt双向供体插入*GAA*基因内的水平,以确定起因于在正向方向上的插入相对于反向方向的校正编辑的百分比。用对野生型GAA基因基因座中的靶序列特异

性的sgRNA(SEQ ID NO:61)、编码spCas9的mRNA、以及由SEQ ID NO:63显示的50nt双向供体(例如dsODN)转染PHH细胞。在转染后48小时,如实施例2中所述,通过NGS确定GAA基因的序列。确定起因于在正向方向或反向方向上的完全插入的总序列读数的百分比。如图8中所示,发现在正向方向或反向方向上的插入是可比较的,其中在正向方向上的插入诱导略微更高水平的校正编辑。

[0903] 已经观察到插入的效率取决于供体多核苷酸的长度,执行了进一步的研究,以评估通过由表6中的SEQ ID NO:61所示的sgRNA,用于插入*GAA*基因中诱导的位点特异性DSB内的这种依赖性。制备了供体dsODN,其在长度中相差5nt,范围从25nt的长度到100nt的长度。这些包括表7中描述的序列以及具有可变长度的另外序列。设计且评估的供体dsODN的序列显示于表9中。如上所述,在HuH-7细胞中评估了完全的NHEJ介导的在*GAA*基因基因座内的插入。有趣的是,与如图7D中所示的在*GAA*基因中的双向供体插入所观察到的效应形成对比,观察到较长的供体dsODN(例如,>50nt)具有与较短的dsODN(例如,50nt)相似的插入功效。事实上,如图9中所示,100nt供体dsODN诱导与50nt供体dsODN相似水平的总插入(例如,在正向方向或反向方向上的完全插入)。关键区别是用于生成dsODN试剂的供应商(图7D中所示的用于生成插入的dsODN得自TriLink Biotechnologies,San Diego,CA;图9中所示的用于生成插入的dsODN得自BioSpring,Frankfurt,德国)。因此,对于较长长度的供体dsODN(例如,长度大于50nt)实现了高效率插入。然而,不受理论的束缚,较长长度的供体dsODN(例如,长度大于50nt)的插入效率可以取决于供体多核苷酸试剂的质量、待编辑的特定基因基因座和/或基于gRNA选择的特定切割位点。

[0904] 表9. 用于NHEJ介导的插入的双向供体多核苷酸

[0905]

供体多核苷酸(大小)	核苷酸序列(5'-3')	SEQ ID NO
GAA_25(25nt)	gagaagaagTATTAACttttttct	62
GAA_30(30nt)	aagagaagaagTATTAACtttttttctttt	95
GAA_35(35nt)	aagaagagaagaagTATTAACtttttttcttttta	96
GAA_40(40nt)	ggaagaagagaagaagTATTAACtttttttcttttaatt	97
GAA_45(45nt)	tggaagaagagaagaagctTATTAACtttttttcttttaattc	98
GAA_50(50nt)	tggaagaagagaagaagctgggTATTAACgcatttttttcttttaattc	63
GAA_55(55nt)	tggaagaagagaagaagctgggTATTAACgcaaatttttttcttttaaattc	99
GAA_60(60nt)	tggaagaagagaagaagctgggaatgtTATTAACgcaaatgatttttttcttttaattc	100
GAA_65(65nt)	tggaagaagagaagaagctgggaatgGTCAGTATATTAACgcaaatgtttttttcttttaattc	101
GAA_70(70nt)	tggaagaagagaagaagctgggaatgtGTCAGTAagTATTAACgcaaatgaatttttttcttttaattc	102
GAA_75(75nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgTATTAACgcaaatgaatttttttcttttaattc	64
GAA_80(80nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgagccaTATTAACgcaaatgaatttttttcttttaattc	103
GAA_85(85nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgagccaccgcgTATTAACgcaaatgaatttttttcttttaattc	104
GAA_90(90nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgagccaccgcgccagcTATTAACgcaaatgaatttttttcttttaattc	105
GAA_95(95nt)	tgggaagaagagaagaagctgggaatgtGTCAGTAaggcatgagccaccgcgccagccatgcTATTAACgcaaatgaatttttttcttttaattc	106
GAA_100(100nt)	tggaagaagagaagaagctgggaatgtGTCAGTAaggcatgagccaccgcgccagccatgcaaatcTATTAACgcaaatgaatttttttcttttaattc	65
供体多核苷酸符号表		
注释	供体多核苷酸设计元件	

[0906]

小写字母	内含子序列
虚线下划线小写字母	定界符序列
粗体小写字母	3'剪接位点和多聚嘧啶片
双下划线小写字母	多聚嘧啶片和3'剪接位点的反向和互补体
实线下划线大写字母	在任一方向上符合分支点共有区的序列
双下划线大写字母	分支点序列TACTGAC的反向和互补体

[0907] 实施例6:双向供体多核苷酸改善了衍生自庞贝病患者的成纤维细胞中NHEJ介导的GAA基因中的插入

[0908] 在衍生自庞贝病患者的成纤维细胞中,进一步评估了NHEJ介导的双向供体多核苷酸(例如dsODN)插入由Cas9/gRNA在*GAA*基因座中诱导的DSB内的效率。细胞在细胞培养基中进行铺平板,并且用靶向突变体*GAA*等位基因的gRNA(SEQ ID NO:94)、编码spCas9的mRNA和50nt双向供体dsODN(SEQ ID NO:63)进行瞬时转染。在转染后48小时,如实施例2所述,从细胞中分离gDNA,并且经受通过NGS的序列分析。确定了包含dsODN在正向方向或反向方向上的完全插入的总序列读数的百分比。起因于dsODN完全总体插入GAA基因座内的校正编辑的百分比显示于图10A中。与单独的Cas9/gRNA对照相比,用Cas9/gRNA和dsODN两者的转染导致高水平(超过15%的校正编辑)。

[0909] 另外,从转染的细胞中分离总RNA,并且使用qPCR和NGS就校正编辑进行评估。对于通过qPCR的校正编辑定量,使用了两组qPCR探针。一组识别包括外显子2的GAA基因的mRNA转录物,而第二组识别缺乏外显子2的GAA基因的mRNA转录物。通过qPCR确定相对于缺乏外显子2的mRNA水平,包括外显子2的mRNA水平。关于外显子包括的ΔΔCt显示于图10B中。用dsODN转染的细胞在由突变体*GAA*基因转录的mRNA中具有升高的外显子2包括,指示了*GAA*基因在转录水平上的增加校正。通过使用NGS对由GAA基因转录的mRNA进行测序,进一步确证了这一点。如图10C中所示,与模拟转染的细胞(例如,未处理的细胞)的转录物相比,掺入外显子2的转录物中的增加百分比对于用供体dsODN转染的细胞更高。总之,这些结果验证了双向供体DNA用于校正人细胞中的GAA基因中的剪接突变的用途。

[0910] 序列表

[0911]

SEQ ID NO	描述	序列
1	人葡萄糖-6-磷酸酶催化亚基(G6PC)mRNA核酸序列(NCBI参考序列NM_000151.3)	ATAGCAGAGCAATCACCACCAAGCCTGGAATAACTG CAAGGGCTCTGCTGACATCTTCCTGAGGTGCCAAGGA AATGAGGATGGAGGAAGGAATGAATGTTCTCCATGA CTTTGGGATCCAGTCAACACATTACCTCCAGGTGAAT TACCAAGACTCCCAGGACTGGTTCATCTTGGTGTCCG TGATCGCAGACCTCAGGAATGCCTTCTACGTCCTCTT CCCCATCTGGTTCCATCTTCAGGAAGCTGTGGGCATT AAACTCCTTTGGGTAGCTGTGATTGGAGACTGGCTCA ACCTCGTCTTTAAGTGGATTCTCTTTGGACAGCGTCC ATACTGGTGGGTTTTGGATACTGACTACTACAGCAAC ACTTCCGTGCCCCTGATAAAGCAGTTCCCTGTAACCT GTGAGACTGGACCAGGGAGCCCCTCTGGCCATGCCA TGGGCACAGCAGGTGTATACTACGTGATGGTCACATC TACTCTTTCCATCTTTCAGGGAAAGATAAAGCCGACC TACAGATTTCGGTGCTTGAATGTCATTTTGTGGTTGG GATTCTGGGCTGTGCAGCTGAATGTCTGTCTGTCACG AATCTACCTTGCTGCTCATTTTCCTCATCAAGTTGTTG CTGGAGTCCTGTCAGGCATTGCTGTTGCAGAAACTTT CAGCCACATCCACAGCATCTATAATGCCAGCCTCAAG AAATATTTTCTCATTACCTTCTTCCTGTTCAGCTTCGC CATCGGATTTTATCTGCTGCTCAAGGGACTGGGTGTA GACCTCCTGTGGACTCTGGAGAAAGCCCAGAGGTGG TGCGAGCAGCCAGAATGGGTCCACATTGACACCACA CCCTTTGCCAGCCTCCTCAAGAACCTGGGCACGCTCT TTGGCCTGGGGCTGGCTCTCAACTCCAGCATGTACAG GGAGAGCTGCAAGGGGAAACTCAGCAAGTGGCTCCC ATTCCGCCTCAGCTCTATTGTAGCCTCCCTCGTCCTCC TGCACGTCTTTGACTCCTTGAAACCCCCATCCCAAGT CGAGCTGGTCTTCTACGTCTTGTCCTTCTGCAAGAGT GCGGTAGTGCCCCTGGCATCCGTCAGTGTCATCCCCT

[0912]

		ACTGCCTCGCCCAGGTCCTGGGCCAGCCGCACAAGA
		AGTCGTTGTAAGAGATGTGGAGTCTTCGGTGTTTAAA
		GTCAACAACCATGCCAGGGATTGAGGAGGACTACTA
		TTTGAAGCAATGGGCACTGGTATTTGGAGCAAGTGAC
		ATGCCATCCATTCTGCCGTCGTGGAATTAAATCACGG
		ATGGCAGATTGGAGGGTCGCCTGGCTTATTCCCATGT
		GTGACTCCAGCCTGCCCTCAGCACAGACTCTTTCAGA
		TGGAGGTGCCATATCACGTACACCATATGCAAGTTTC
		CCGCCAGGAGGTCCTCCTCTCTCTACTTGAATACTCT
		CACAAGTAGGGAGCTCACTCCCACTGGAACAGCCCA
		TTTTATCTTTGAATGGTCTTCTGCCAGCCCATTTTGAG
		GCCAGAGGTGCTGTCAGCTCAGGTGGTCCTCTTTTAC
		AATCCTAATCATATTGGGTAATGTTTTTGAAAAGCTA
		ATGAAGCTATTGAGAAAGACCTGTTGCTAGAAGTTG
		GGTTGTTCTGGATTTTCCCCTGAAGACTTACTTATTCT
		TCCGTCACATATACAAAAGCAAGACTTCCAGGTAGG
		GCCAGCTCACAAGCCCAGGCTGGAGATCCTAACTGA
		GAATTTTCTACCTGTGTTCATTCTTACCGAGAAAAGG
		AGAAAGGAGCTCTGAATCTGATAGGAAAAGAAGGCT
		GCCTAAGGAGGAGTTTTTAGTATGTGGCGTATCATGC
		AAGTGCTATGCCAAGCCATGTCTAAATGGCTTTAATT
		ATATAGTAATGCACTCTCAGTAATGGGGGACCAGCTT
		AAGTATAATTAATAGATGGTTAGTGGGGTAATTCTGC
		TTCTAGTATTTTTTTTACTGTGCATACATGTTCATCGT
		ATTTCCTTGGATTTCTGAATGGCTGCAGTGACCCAGA
		TATTGCACTAGGTCAAAACATTCAGGTATAGCTGACA
		TCTCCTCTATCACATTACATCATCCTCCTTATAAGCCC
		AGCTCTGCTTTTTCCAGATTCTTCCACTGGCTCCACAT
		CCACCCCACTGGATCTTCAGAAGGCTAGAGGGCGAC
		TCTGGTGGTGCTTTTGTATGTTTCAATTAGGCTCTGAA
		ATCTTGGGCAAAATGACAAGGGGAGGGCCAGGATTC
		CTCTCTCAGGTCACTCCAGTGTTACTTTTAATTCCTAG

[0913]

		AGGGTAAATATGACTCCTTTCTCTATCCCAAGCCAAC
		CAAGAGCACATTCTTAAAGGAAAAGTCAACATCTTCT
		CTCTTTTTTTTTTTTTTTGAGACAGGGTCTCACTATGT
		TGCCCAGGCTGCTCTTGAATTCCTGGGCTCAAGCAGT
		CCTCCCACCCTACCACAGCGTCCCGCGTAGCTGGGAC
		TACAGGTGCAAGCCACTATGTCCAGCTAGCCAACTCC
		TCCTTGCCTGCTTTTCTTTTTTTTTCTTTTTTTGAGACG
		GCGCACCTATCACCCAGGCTGGAGTGGAGTGGCACG
		ATCTTGGCTCACTGCAACCTCTTCCTCCTGGTTCAAG
		CGATTCTCATGTCTCAGCCTCCTCAGTAGCTAGGACT
		ACCGGCGTGCACCACCATGCCAGGCTAATTTTTATAT
		TTTTAGAATTTTAGAAGAGATGGGATTTCATCATGTT
		GGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGATC
		CACCTGCCTTGGCCTCCCAAGGTGCTAGGATTACAGG
		CATGAGCCACCGCACCGGGCCCTCCTTGCCTGTTTTT
		CAATCTCATCTGATATGCAGAGTATTTCTGCCCCACC
		CACCTACCCCCCAAAAAAAGCTGAAGCCTATTTATTT
		GAAAGTCCTTGTTTTTGCTACTAATTATATAGTATAC
		CATACATTATCATTCAAAACAACCATCCTGCTCATAA
		CATCTTTGAAAAGAAAAATATATATGTGCAGTATTTT
		ATTAAAGCAACATTTTATTTAAGAATAAAGTCTTGTT
		AATTACTATATTTTAGATGCAATGTGATCTGAAGTTT
		CTAATTCTGGCCCAACTAAATTTCTAGCTCTGTTTCCC
		TAAACAAATAATTTGGTTTCTCTGTGCCTGCATTTTCC
		CTTTGGAGAAGAAAAGTGCTCTCTCTTGAGTTGACCG
		AGAGTCCCATTAGGGATAGGGAGACTTAAATGCATC
		CACAGGGGCACAGGCAGAGTTGAGCACATAAACGGA
		GGCCCAAAATCAGCATAGAACCAGAAAGATTCAGAG
		TTGGCCAAGAATGAACATTGGCTACCAGACCACAAG
		TCAGCATGAGTTGCTCTATGGCATCAAATTGCAACTT
		GAGAGTAGATGGGCAGGGTCACTATCAAATTAAGCA
		ATCAGGGCACACAAGTTGCAGTAACACAACAAGACT

[0914]

		AGGCCAGCTCTGGAATCCAGTAACTCAGTGTCAGCA AGGTTTTGGGTTATAGTTCAAGAAAGTCTAAACAGAG CCAGTCACAGCACCAAGGAATGCTCAAGGGAGCTAT TGCAGGTTTCTCTGCTAAGAGATTTATTTCATCCTGG GTGCAGGGTTCGACCTCCAAAGGCCTCAAATCATCAC CGTATCAATGGATTTCCTGAGGGTAAGCTCCGCTATT TCACACCTGAACTCCGGAGTCTGTATATTCAGGGAAG ATTGCATTCTCCTACTGGATTTGGGCTCTCAGAGGGC GTTGTGGGAACCAGGCCCCTCACAGAATCAAATGGT CCCAACCAGGGAGAAAGAAAATAGTCTTTTTTTTTTT TTTAATAGAGATGGGGGTCTCACTATGCTGCCCAGGC TGGTCTTGAACTCCTGGGTTCAAGTGATCCTCCTGCC TCAGCCTCCCAAAGTGCTGGGATTACAGTGTGAGCCA CTGCGCTTGGCCAGAAATGGTTTTGATCTGTCTGAAC TGAACCCTACTGCTTAGGCATAGCCCCATCCTTGATA ATCTATTTGCTCCCAAGGACCAAGTCCAAGATCCTTA CAAGAAAGGTCTGCCAGAAAGTAAATACTGCCCCCA CTCCCTGAAGTTTATGAGGTTGATAAGAAAACATAAC AGATAAAGTTTATTGAGTGCTAACTTTAAAAAAAAA
2	人葡萄糖-6-磷酸酶催化亚基(G6PC)氨基酸序列(NCBI 参考序列 NP_000142.2)	MEEGMNVLHDFGIQSTHYLQVNYQDSQDWFILVSVIAD LRNAFYVLFPIWFHLQEAVGIKLLWVAVIGDWLNLVFK WILFGQRPYWWVLDTDYYSNTSVPLIKQFPVTCETGPG SPSGHAMGTAGVYYYVMVTSTLSIFQGKIKPTYRFRCLN VILWLGFWAVQLNVCLSRIYLAAHFPHQVVAGVLSGIA VAETFSHIHSIYNASLKKYFLITFFLFSFAIGFYLLLKGLG VDLLWTLEKAQRWCEQPEWVHIDTTPFASLLKNLGTLF GLGLALNSSMYRESCKGKLSKWLPFRLSSIVASLVLLH VFDSLKPPSQVELVFYVLSFCKSAVVPLASVSVIPYCLA QVLGQPHKKSL
3	CH32-CTX1 sgRNA	USCSUSUUGGACAGCGUCCAUACGUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUSUSUSU

[0915]

4	CH34-CTX1 sgRNA	USGSGSACAGCGUCCAUACUGGUGUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUSUSUSU
5	CH36-CTX1 sgRNA	GSUSASUCCAAAACCCACCAGUAGUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUSUSUSU
6	CH32 sgRNA	UCUUUGGACAGCGUCCAUACGUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU
7	CH34 sgRNA	UGGACAGCGUCCAUACUGGUGUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU
8	CH36 sgRNA	GUAUCCAAAACCCACCAGUAGUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU
9	CH32_75-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAG AACAAGTTTCTGGGGTTACTGAATGAATGCTTTTGCC CA
10	CH32_75-0 有义	TGGGCAAAAGCATTCATTCAGTAACCCCAGAAACTT GTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCC T
11	CH32_100-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAG AACAAGTTTCTGGGGTTACTGAATGAATGCTTTTGCC CAAAGCCTACACCTTCAAGAAGAGTGT
12	CH32_100-0 有义	ACACTCTTCTTGAAGGTGTAGGCTTTGGGCAAAAGCA TTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCA TAGGATTCTCTTTGGACAGTGCCCT
13	CH32_125-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAG AACAAGTTTCTGGGGTTACTGAATGAATGCTTTTGCC CAAAGCCTACACCTTCAAGAAGAGTGTAGCCTGAGA AGGATTTCACATGTTG
14	CH32_125-0 有义	CAACATGTGAAATCCTTCTCAGGCTACACTCTTCTTG

[0916]

		AAGGTGTAGGCTTTGGGCAAAAGCATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCT
15	CH32_25-0 反义	AGGGCACTGTCCAAAGAGAATCCTA
16	CH32_25-0 有义	TAGGATTCTCTTTGGACAGTGCCCT
17	CH32_50-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGG
18	CH32_50-0 有义	CCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCT
19	增强的 CH32_50-0 反义	AGGGCACTGTCCAAAGAGAATCCTATAAAAAAACAGAACAAGTTTATGAA
20	增强的 CH32_50-0 有义	TTCATAAACTTGTTCTGTTTTTTTATAGGATTCTCTTTGGACAGTGCCCT
21	增强的 CH32_55-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGATGAA
22	增强的 CH32_55-0 有义	TTCATCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCT
23	增强的 CH32_60-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTAATGAAT
24	增强的 CH32_60-0 有义	ATTCATTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCT
25	CH32_65-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTACTGAATGAATG
26	CH32_65-0 有义	CATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCT
27	CH32_70-0 反义	AGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTACTGAATGAATGCTTTT
28	CH32_70-0 有义	AAAAGCATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCT
29	增强的	AGTAAGGGCACTGTCCAAAGAGAATCCTATAAAAAA

[0917]

	CH34_54-0(C) 反义	ACAGAACAAGTTTATGAA
30	增强的 CH34_54-0(C) 有义	TTCATAAACTTGTTCTGTTTTTTTATAGGATTCTCTTTGGACAGTGCCCTTACT
31	增强的 CH34_59-0 反义	AGTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGATGAA
32	增强的 CH34_59-0 有义	TTCATCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTACT
33	增强的 CH34_64-0 反义	AGTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTAATGAAT
34	增强的 CH34_64-0 有义	ATTCATTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTACT
35	CH34_69-0 反义	AGTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTACTGAATGAATG
36	CH34_69-0 有义	CATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTACT
37	CH34_74-0 反义	AGTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTACTGAATGAATGCTTTT
38	CH34_74-0 有义	AAAAGCATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTACT
39	增强的 CH36_53-0 反义	GTAAGGGCACTGTCCAAAGAGAATCCTATAAAAAAACAGAACAAGTTTATGAA
40	增强的 CH36_53-0 有义	TTCATAAACTTGTTCTGTTTTTTTATAGGATTCTCTTTGGACAGTGCCCTTAC
41	增强的 CH36_58-0 反义	GTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGATGAA
42	增强的	TTCATCCCAGAAACTTGTTCTGTTTTTCCATAGGATTC

[0918]

	CH36_58-0 有义	TCTTTGGACAGTGCCCTTAC
43	增强的CH36_63-0反义	GTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTAATGAAT
44	增强的CH36_63-0 有义	ATTCATTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTAC
45	CH36_68-0反义	GTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTACTGAATGAATG
46	CH36_68-0 有义	CATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTAC
47	CH36_73-0反义	GTAAGGGCACTGTCCAAAGAGAATCCTATGGAAAAACAGAACAAGTTTCTGGGGTTACTGAATGAATGCTTTT
48	CH36_73-0 有义	AAAAGCATTCATTCAGTAACCCCAGAAACTTGTTCTGTTTTTCCATAGGATTCTCTTTGGACAGTGCCCTTAC
49	分支点共有序列	YTNAY，其中 Y = C 或 T；其中 N = A、G、C 或 T
50	分支点序列	TATTAAC
51	分支点序列	GTTAATA
52	分支点序列	TACTGAC
53	多聚嘧啶片	TTTTTTTCT
54	多聚嘧啶片	TTTTTTTCTTTTT
55	多聚嘧啶片	CTTCTTCTCTTCTTCC
56	SV40 核定位信号(NLS)	PKKKRKV
57	SV40 核定位信号(NLS)	PKKKRRV
58	核质蛋白核定位信号(NLS)	KRPAATKKAGQAKKKK
59	化脓性链球菌靶序列和 PAM	$N_{19\text{-}21}$NRG，其中 N = A、G、C 或 T；并且 R = A 或 G
60	6xHis	HHHHHH

[0919]

61	GAA5 sgRNA	AGCCCGCUUUCUUCUCCCGCGUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU
62	GAA_25-0_F	GAGAAGAAGTATTAACTTTTTTTCT
63	GAA_50-0_F	TGGAAGAAGAGAAGAAGCTGGGTATTAACGCATTTT TTTCTTTTTAATTC
64	GAA_75-0_F	TGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAG GCATGTATTAACGCAAATGAATTTTTTTCTTTTTAATT C
65	GAA_100-0_F	TGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAG GCATGAGCCACCGCGCCAGCCATGCAAATCTATTAA CGCAAATGAATTTTTTTCTTTTTAATTC
66	GAA_25-0_R	AGAAAAAAAGTTAATACTTCTTCTC
67	GAA_50-0_R	GAATTAAAAAGAAAAAAATGCGTTAATACCCAGCTT CTTCTCTTCTTCCA
68	GAA_75-0_R	GAATTAAAAAGAAAAAAATTCATTTGCGTTAATACAT GCCTTACTGACACATTCCCAGCTTCTTCTCTTCTTCCA
69	GAA_100-0_R	GAATTAAAAAGAAAAAAATTCATTTGCGTTAATAGA TTTGCATGGCTGGCGCGGTGGCTCATGCCTTACTGAC ACATTCCCAGCTTCTTCTCTTCTTCCA
70	CH42 sgRNA	GUCAGUCUCACAGGUUACAGGUUUUAGAGCUAGAA AUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU
71	CH42_25-0_F	AACCTGTGAGACTGGACCAGGTAAG
72	CH42_25-0_R	CTTACCTGGTCCAGTCTCACAGGTT
73	CH42_50-0_F	AACCTGTGAGACTGGACCAGGTAAGACTCACACTTG CGAAACCGGCCCAG
74	CH42_50-0_R	CTGGGCCGGTTTCGCAAGTGTGAGTCTTACCTGGTCC AGTCTCACAGGTT
75	CH42_75-0_F	AACCTGTGAGACTGGACCAGGTAAGCGACGCGCATT TCTCACACGGCAGGACTCACACTTGCGAAACCGGCC CAG

[0920]

76	CH42_75-0_R	CTGGGCCGGTTTCGCAAGTGTGAGTCCTGCCGTGTGAGAAATGCGCGTCGCTTACCTGGTCCAGTCTCACAGGTT
77	CH42_100-0_F	AACCTGTGAGACTGGACCAGGTAAGCGACGCGCATTTCTCACACGGCAGGGAGGGCCACACGCGTTTGTTTCTCAACTCACACTTGCGAAACCGGCCCAG
78	CH42_100-0_R	CTGGGCCGGTTTCGCAAGTGTGAGTTGAGAAACAAACGCGTGTGGCCCTCCCTGCCGTGTGAGAAATGCGCGTCGCTTACCTGGTCCAGTCTCACAGGTT
79	CH42_125-0_F	AACCTGTGAGACTGGACCAGGTAAGCGACGCGCATTTCTCACACGGCAGGGAGGGCCACACGCGTTTGTTTCTCACACGATGGGCAGGGCGACACATGTTACTCACACTTGCGAAACCGGCCCAG
80	CH42_125-0_R	CTGGGCCGGTTTCGCAAGTGTGAGTAACATGTGTCGCCCTGCCCATCGTGTGAGAAACAAACGCGTGTGGCCCTCCCTGCCGTGTGAGAAATGCGCGTCGCTTACCTGGTCCAGTCTCACAGGTT
81	CH32 sgRNA 间隔区	UCUUUGGACAGCGUCCAUAC
82	CH32 靶基因 (PAM)	TCTTTGGACAGCGTCCATAC(TGG)
83	CH34 sgRNA 间隔区	UGGACAGCGUCCAUACUGGU
84	CH34 靶基因 (PAM)	TGGACAGCGTCCATACTGGT(GGG)
85	CH36 sgRNA 间隔区	GUAUCCAAAACCCACCAGUA
86	CH36 靶基因 (PAM)	GTATCCAAAACCCACCAGTA(TGG)
87	CH42 靶基因	GTCAGTCTCACAGGTTACAG(GGG)

[0921]

	(PAM)	
88	CH42 sgRNA 间隔区	GUCAGUCUCACAGGUUACAG
89	鼠 CH34-CTX1 sgRNA	USGSGSACAACGCCCGUAUUGGUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUSUSUSU
90	WT GAA 靶基因(PAM)	AGCCCGCTTTCTTCTCCCGC(AGG)
91	WT GAA5 sgRNA 间隔区	AGCCCGCUUUCUUCUCCCGC
92	突变体(Mt)GAA 靶基因(PAM)	AGCCCGCTTGCTTCTCCCGC(AGG)
93	Mt GAA sgRNA 间隔区	AGCCCGCUUGCUUCUCCCGC
94	Mt GAA sgRNA	AGCCCGCUUGCUUCUCCCGCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU
95	GAA_30 F	AAGAGAAGAAGTATTAACTTTTTTTCTTTT
96	GAA_35 F	AAGAAGAGAAGAAGTATTAACTTTTTTTCTTTTTA
97	GAA_40 F	GGAAGAAGAGAAGAAGTATTAACTTTTTTTCTTTTTAATT
98	GAA_45 F	TGGAAGAAGAGAAGAAGCTTATTAACTTTTTTTCTTTTAATTC
99	GAA_55 F	TGGAAGAAGAGAAGAAGCTGGGTATTAACGCAAATTTTTTTTCTTTTTAATTC
100	GAA_60 F	TGGAAGAAGAGAAGAAGCTGGGAATGTTATTAACGCAAATGATTTTTTTCTTTTTAATTC
101	GAA_65 F	TGGAAGAAGAGAAGAAGCTGGGAATGGTCAGTATATTAACGCAAATGTTTTTTTCTTTTTAATTC
102	GAA_70 F	TGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAGTATTAACGCAAATGAATTTTTTTCTTTTTAATTC

103	GAA_80 F	TGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAGGCATGAGCCATATTAACGCAAATGAATTTTTTTCTTTTTAATTC
104	GAA_85 F	TGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAGGCATGAGCCACCGCGTATTAACGCAAATGAATTTTTTTCTTTTTAATTC
105	GAA_90 F	TGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAGGCATGAGCCACCGCGCCAGCTATTAACGCAAATGAATTTTTTCTTTTTAATTC
106	GAA_95 F	TGGGAAGAAGAGAAGAAGCTGGGAATGTGTCAGTAAGGCATGAGCCACCGCGCCAGCCATGCTATTAACGCAAATGAATTTTTTTCTTTTTAATTC
107	CH32 突变体 CTX1 sgRNA 间隔区	UCUUUGGACAGCGUCCAUAC
108	CH34 突变体 CTX1 sgRNA 间隔区	UGGACAGCGUCCAUACUGGU

[0922] 。

序列表

<110> 克里斯珀医疗股份公司

<120> 通过插入供体多核苷酸用于基因组编辑的组合物和方法

<130> AC2461 PCT S3

<140> PCT/IB2019/00833

<141> 2019-06-28

<150> US 62/691,573

<151> 2018-06-28

<160> 110

<170> PatentIn version 3.5

<210> 1

<211> 4169

<212> DNA

<213> 智人

<220>

<221> misc_feature

<223> 人葡萄糖6磷酸酶催化亚基（G6PC）mRNA核酸序列

<400> 1

[0001]

atagcagagc aatcaccacc aagcctggaa taactgcaag ggctctgctg acatcttcct 60
gaggtgccaa ggaaatgagg atggaggaag gaatgaatgt tctccatgac tttgggatcc 120
agtcaacaca ttacctccag gtgaattacc aagactccca ggactggttc atcttggtgt 180
ccgtgatcgc agacctcagg aatgccttct acgtcctctt ccccatctgg ttccatcttc 240
aggaagctgt gggcattaaa ctcctttggg tagctgtgat tggagactgg ctcaacctcg 300
tctttaagtg gattctcttt ggacagcgtc catactggtg ggttttggat actgactact 360
acagcaacac ttccgtgccc ctgataaagc agttccctgt aacctgtgag actggaccag 420
ggagcccctc tggccatgcc atgggcacag caggtgtata ctacgtgatg gtcacatcta 480
ctctttccat ctttcaggga aagataaagc cgacctacag atttcggtgc ttgaatgtca 540
ttttgtggtt gggattctgg gctgtgcagc tgaatgtctg tctgtcacga atctaccttg 600
ctgctcattt tcctcatcaa gttgttgctg gagtcctgtc aggcattgct gttgcagaaa 660
ctttcagcca catccacagc atctataatg ccagcctcaa gaaatatttt ctcattacct 720
tcttcctgtt cagcttcgcc atcggatttt atctgctgct caagggactg ggtgtagacc 780
tcctgtggac tctggagaaa gcccagaggt ggtgcgagca gccagaatgg gtccacattg 840
acaccacacc ctttgccagc ctcctcaaga acctgggcac gctctttggc ctggggctgg 900
ctctcaactc cagcatgtac agggagagct gcaaggggaa actcagcaag tggctcccat 960
tccgcctcag ctctattgta gcctccctcg tcctcctgca cgtctttgac tccttgaaac 1020
ccccatccca agtcgagctg gtcttctacg tcttgtcctt ctgcaagagt gcggtagtgc 1080
ccctggcatc cgtcagtgtc atcccctact gcctcgccca ggtcctgggc cagccgcaca 1140
agaagtcgtt gtaagagatg tggagtcttc ggtgtttaaa gtcaacaacc atgccaggga 1200
ttgaggagga ctactatttg aagcaatggg cactggtatt tggagcaagt gacatgccat 1260
ccattctgcc gtcgtggaat taaatcacgg atggcagatt ggagggtcgc ctggcttatt 1320
cccatgtgtg actccagcct gccctcagca cagactcttt cagatggagg tgccatatca 1380
cgtacaccat atgcaagttt cccgccagga ggtcctcctc tctctacttg aatactctca 1440
caagtaggga gctcactccc actggaacag cccattttat ctttgaatgg tcttctgcca 1500
gcccattttg aggccagagg tgctgtcagc tcaggtggtc ctcttttaca atcctaatca 1560

[0002]

tattgggtaa tgtttttgaa aagctaatga agctattgag aaagacctgt tgctagaagt 1620
tgggttgttc tggattttcc cctgaagact tacttattct tccgtcacat atacaaaagc 1680
aagacttcca ggtagggcca gctcacaagc ccaggctgga gatcctaact gagaattttc 1740
tacctgtgtt cattcttacc gagaaaagga gaaaggagct ctgaatctga taggaaaaga 1800
aggctgccta aggaggagtt tttagtatgt ggcgtatcat gcaagtgcta tgccaagcca 1860
tgtctaaatg gctttaatta tatagtaatg cactctcagt aatgggggac cagcttaagt 1920
ataattaata gatggttagt ggggtaattc tgcttctagt attttttta ctgtgcatac 1980
atgttcatcg tatttccttg gatttctgaa tggctgcagt gacccagata ttgcactagg 2040
tcaaaacatt caggtatagc tgacatctcc tctatcacat tacatcatcc tccttataag 2100
cccagctctg ctttttccag attcttccac tggctccaca tccaccccac tggatcttca 2160
gaaggctaga gggcgactct ggtggtgctt ttgtatgttt caattaggct ctgaaatctt 2220
gggcaaaatg acaaggggag ggccaggatt cctctctcag gtcactccag tgttacttt 2280
aattcctaga gggtaaatat gactcctttc tctatcccaa gccaaccaag agcacattct 2340
taaaggaaaa gtcaacatct tctctctttt ttttttttt tgagacaggg tctcactatg 2400
ttgcccaggc tgctcttgaa ttcctgggct caagcagtcc tcccacccta ccacagcgtc 2460
ccgcgtagct gggactacag gtgcaagcca ctatgtccag ctagccaact cctccttgcc 2520
tgcttttctt ttttttctt tttttgagac ggcgcaccta tcacccaggc tggagtggag 2580
tggcacgatc ttggctcact gcaacctctt cctcctggtt caagcgattc tcatgtctca 2640
gcctcctcag tagctaggac taccggcgtg caccaccatg ccaggctaat ttttatattt 2700
ttagaatttt agaagagatg ggatttcatc atgttggcca ggctggtctc gaactcctga 2760
cctcaagtga tccacctgcc ttggcctccc aaggtgctag gattacaggc atgagccacc 2820
gcaccgggcc ctccttgcct gtttttcaat ctcatctgat atgcagagta tttctgcccc 2880
acccacctac cccccaaaaa aagctgaagc ctatttattt gaaagtcctt gtttttgcta 2940
ctaattatat agtataccat acattatcat tcaaaacaac catcctgctc ataacatctt 3000
tgaaaagaaa aatatatatg tgcagtattt tattaaagca acattttatt taagaataaa 3060
gtcttgttaa ttactatatt ttagatgcaa tgtgatctga agtttctaat tctggcccaa 3120
ctaaatttct agctctgttt ccctaaacaa ataatttggt ttctctgtgc ctgcattttc 3180
cctttggaga agaaaagtgc tctctcttga gttgaccgag agtcccatta gggataggga 3240
gacttaaatg catccacagg ggcacaggca gagttgagca cataaacgga ggcccaaaat 3300
cagcatagaa ccagaaagat tcagagttgg ccaagaatga acattggcta ccagaccaca 3360
agtcagcatg agttgctcta tggcatcaaa ttgcaacttg agagtagatg ggcagggtca 3420
ctatcaaatt aagcaatcag ggcacacaag ttgcagtaac acaacaagac taggccagct 3480
ctggaatcca gtaactcagt gtcagcaagg ttttgggtta tagttcaaga aagtctaaac 3540
agagccagtc acagcaccaa ggaatgctca agggagctat tgcaggtttc tctgctaaga 3600
gatttatttc atcctgggtg cagggttcga cctccaaagg cctcaaatca tcaccgtatc 3660
aatggatttc ctgagggtaa gctccgctat ttcacacctg aactccggag tctgtatatt 3720
cagggaagat tgcattctcc tactggattt gggctctcag agggcgttgt gggaaccagg 3780
cccctcacag aatcaaatgg tcccaaccag ggagaaagaa aatagtcttt ttttttttt 3840
taatagagat gggggtctca ctatgctgcc caggctggtc ttgaactcct gggttcaagt 3900
gatcctcctg cctcagcctc ccaaagtgct gggattacag tgtgagccac tgcgcttggc 3960
cagaaatggt tttgatctgt ctgaactgaa ccctactgct taggcatagc cccatccttg 4020
ataatctatt tgctcccaag gaccaagtcc aagatcctta caagaaaggt ctgccagaaa 4080
gtaaatactg cccccactcc ctgaagttta tgaggttgat aagaaaacat aacagataaa 4140
gtttattgag tgctaacttt aaaaaaaaa 4169

<210> 2
<211> 357
<212> PRT
<213> 智人
<220>
<221> misc_feature
<223> 人葡萄糖6磷酸酶催化亚基（G6PC）氨基酸序列
<400> 2

[0003]

Met Glu Glu Gly Met Asn Val Leu His Asp Phe Gly Ile Gln Ser Thr
1 5 10 15
His Tyr Leu Gln Val Asn Tyr Gln Asp Ser Gln Asp Trp Phe Ile Leu
20 25 30
Val Ser Val Ile Ala Asp Leu Arg Asn Ala Phe Tyr Val Leu Phe Pro
35 40 45
Ile Trp Phe His Leu Gln Glu Ala Val Gly Ile Lys Leu Leu Trp Val
50 55 60
Ala Val Ile Gly Asp Trp Leu Asn Leu Val Phe Lys Trp Ile Leu Phe
65 70 75 80
Gly Gln Arg Pro Tyr Trp Trp Val Leu Asp Thr Asp Tyr Tyr Ser Asn
85 90 95
Thr Ser Val Pro Leu Ile Lys Gln Phe Pro Val Thr Cys Glu Thr Gly
100 105 110
Pro Gly Ser Pro Ser Gly His Ala Met Gly Thr Ala Gly Val Tyr Tyr
115 120 125
Val Met Val Thr Ser Thr Leu Ser Ile Phe Gln Gly Lys Ile Lys Pro
130 135 140
Thr Tyr Arg Phe Arg Cys Leu Asn Val Ile Leu Trp Leu Gly Phe Trp
145 150 155 160
Ala Val Gln Leu Asn Val Cys Leu Ser Arg Ile Tyr Leu Ala Ala His
165 170 175
Phe Pro His Gln Val Val Ala Gly Val Leu Ser Gly Ile Ala Val Ala
180 185 190
Glu Thr Phe Ser His Ile His Ser Ile Tyr Asn Ala Ser Leu Lys Lys
195 200 205
Tyr Phe Leu Ile Thr Phe Phe Leu Phe Ser Phe Ala Ile Gly Phe Tyr
210 215 220
Leu Leu Leu Lys Gly Leu Gly Val Asp Leu Leu Trp Thr Leu Glu Lys
225 230 235 240
Ala Gln Arg Trp Cys Glu Gln Pro Glu Trp Val His Ile Asp Thr Thr
245 250 255
Pro Phe Ala Ser Leu Leu Lys Asn Leu Gly Thr Leu Phe Gly Leu Gly
260 265 270
Leu Ala Leu Asn Ser Ser Met Tyr Arg Glu Ser Cys Lys Gly Lys Leu
275 280 285

Ser Lys Trp Leu Pro Phe Arg Leu Ser Ser Ile Val Ala Ser Leu Val
290 295 300
Leu Leu His Val Phe Asp Ser Leu Lys Pro Pro Ser Gln Val Glu Leu
305 310 315 320
Val Phe Tyr Val Leu Ser Phe Cys Lys Ser Ala Val Val Pro Leu Ala
325 330 335
Ser Val Ser Val Ile Pro Tyr Cys Leu Ala Gln Val Leu Gly Gln Pro
340 345 350
His Lys Lys Ser Leu
355

<210> 3
<211> 106
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH32-CTX1 sgRNA
<400> 3
uscsusuugg acagcgucca uacguuuuag agcuagaaau agcaaguuaa aauaaggcua 60
guccguuauc aacuugaaaa aguggcaccg agucggugcu sususu 106

[0004]

<210> 4
<211> 106
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH34-CTX1 sgRNA
<400> 4
usgsgsacag cguccauacu gguguuuuag agcuagaaau agcaaguuaa aauaaggcua 60
guccguuauc aacuugaaaa aguggcaccg agucggugcu sususu 106
<210> 5
<211> 106
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH36-CTX1 sgRNA
<400> 5
gsusasucca aaacccacca guaguuuuag agcuagaaau agcaaguuaa aauaaggcua 60
guccguuauc aacuugaaaa aguggcaccg agucggugcu sususu 106
<210> 6
<211> 100
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH32 sgRNA
<400> 6

ucuuuggaca gcguccauac guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 7
<211> 100
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH34 sgRNA
<400> 7
uggacagcgu ccauacuggu guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 8
<211> 100
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH36 sgRNA
<400> 8
guauccaaaa cccaccagua guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 9
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_75-0 Antisense
<400> 9
agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg gttactgaat 60
gaatgctttt gccca 75

<210> 10
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_75-0 有义
<400> 10
tgggcaaaag cattcattca gtaaccccag aaacttgttc tgtttttcca taggattctc 60
tttggacagt gccct 75

<210> 11
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_100-0 反义

[0005]

<400> 11
agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg gttactgaat 60
gaatgctttt gcccaaagcc tacaccttca agaagagtgt 100
<210> 12
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_100-0 有义
<400> 12
acactcttct tgaaggtgta ggctttgggc aaaagcattc attcagtaac cccagaaact 60
tgttctgttt ttccatagga ttctctttgg acagtgccct 100
<210> 13
<211> 125
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_125-0 反义
<400> 13
agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg gttactgaat 60
gaatgctttt gcccaaagcc tacaccttca agaagagtgt agcctgagaa ggatttcaca 120
tgttg 125

[0006]

<210> 14
<211> 125
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_125-0 有义
<400> 14
caacatgtga aatccttctc aggctacact cttcttgaag gtgtaggctt tgggcaaaag 60
cattcattca gtaaccccag aaacttgttc tgtttttcca taggattctc tttggacagt 120
gccct 125
<210> 15
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_25-0 反义
<400> 15
agggcactgt ccaaagagaa tccta 25
<210> 16
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 合成的: CH32_25-0 有义
<400> 16
taggattctc tttggacagt gccct 25
<210> 17
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_50-0 反义
<400> 17
agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg 50

<210> 18
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_50-0 有义
<400> 18
cccagaaact tgttctgttt ttccatagga ttctctttgg acagtgccct 50

[0007]

<210> 19
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH32_50-0 反义
<400> 19
agggcactgt ccaaagagaa tcctataaaa aaacagaaca agtttatgaa 50
<210> 20
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH32_50-0 有义
<400> 20
ttcataaact tgttctgttt ttttatagga ttctctttgg acagtgccct 50
<210> 21
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH32_55-0 反义
<400> 21

agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg atgaa 55

<210> 22

<211> 55

<212> DNA

<213> 人工序列

<220>

<223> 合成的: 增强的 CH32_55-0 有义

<400> 22

ttcatcccag aaacttgttc tgtttttcca taggattctc tttggacagt gccct 55

<210> 23

<211> 60

<212> DNA

<213> 人工序列

<220>

<223> 合成的: 增强的 CH32_60-0 反义

<400> 23

agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg gttaatgaat 60

<210> 24

<211> 60

<212> DNA

<213> 人工序列

<220>

[0008]

<223> 合成的: 增强的 CH32_60-0 有义

<400> 24

attcattaac cccagaaact tgttctgttt ttccatagga ttctctttgg acagtgccct 60

<210> 25

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 合成的: CH32_65-0 反义

<400> 25

agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg gttactgaat 60

gaatg 65

<210> 26

<211> 65

<212> DNA

<213> 人工序列

<220>

<223> 合成的: CH32_65-0 有义

<400> 26

cattcattca gtaaccccag aaacttgttc tgtttttcca taggattctc tttggacagt 60

gccct 65

<210> 27

<210> 70
<211> DNA
<212> 人工序列
<213>
<220>
<223> 合成的: CH32_70-0 反义
<400> 27
agggcactgt ccaaagagaa tcctatggaa aaacagaaca agtttctggg gttactgaat 60
gaatgctttt 70
<210> 28
<211> 70
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32_70-0 有义
<400> 28
aaaagcattc attcagtaac cccagaaact tgttctgttt ttccatagga ttctctttgg 60
acagtgccct 70
<210> 29
<211> 54
<212> DNA
<213> 人工序列
<220>
[0009]
<223> 合成的: 增强的 CH34_54-0 (C)反义
<400> 29
agtaagggca ctgtccaaag agaatcctat aaaaaaacag aacaagttta tgaa 54
<210> 30
<211> 54
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH34_54-0 (C)有义
<400> 30
ttcataaact tgttctgttt ttttatagga ttctctttgg acagtgccct tact 54
<210> 31
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH34_59-0 反义
<400> 31
agtaagggca ctgtccaaag agaatcctat ggaaaaacag aacaagtttc tgggatgaa 59
<210> 32
<211> 59
<212> DNA

<213> 人工序列
<220>
<223> 合成的：增强的 CH34_59-0 有义
<400> 32
ttcatcccag aaacttgttc tgtttttcca taggattctc tttggacagt gcccttact 59
<210> 33
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 合成的：增强的 CH34_64-0 反义
<400> 33
agtaagggca ctgtccaaag agaatcctat ggaaaaacag aacaagtttc tggggttaat 60
gaat 64
<210> 34
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 合成的：增强的 CH34_64-0 有义
<400> 34
[0010]
attcattaac cccagaaact tgttctgttt ttccatagga ttctctttgg acagtgccct 60
tact 64
<210> 35
<211> 69
<212> DNA
<213> 人工序列
<220>
<223> 合成的：CH34_69-0 反义
<400> 35
agtaagggca ctgtccaaag agaatcctat ggaaaaacag aacaagtttc tggggttact 60
gaatgaatg 69
<210> 36
<211> 69
<212> DNA
<213> 人工序列
<220>
<223> 合成的：CH34_69-0 有义
<400> 36
cattcattca gtaaccccag aaacttgttc tgtttttcca taggattctc tttggacagt 60
gcccttact 69
<210> 37
<211> 74
<212> DNA

<213> 人工序列
<220>
<223> 合成的: CH34_74-0 反义
<400> 37
agtaagggca ctgtccaaag agaatcctat ggaaaaacag aacaagtttc tggggttact 60
gaatgaatgc tttt 74
<210> 38
<211> 74
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH34_74-0 有义
<400> 38
aaaagcattc attcagtaac cccagaaact tgttctgttt ttccatagga ttctctttgg 60
acagtgccct tact 74
<210> 39
<211> 53
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH36_53-0 反义
<400> 39
[0011]
gtaagggcac tgtccaaaga gaatcctata aaaaaacaga acaagtttat gaa 53
<210> 40
<211> 53
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH36_53-0 有义
<400> 40
ttcataaact tgttctgttt ttttatagga ttctctttgg acagtgccct tac 53
<210> 41
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 增强的 CH36_58-0 反义
<400> 41
gtaagggcac tgtccaaaga gaatcctatg gaaaaacaga acaagtttct gggatgaa 58
<210> 42
<211> 58
<212> DNA
<213> 人工序列
<220>

<223> 合成的：增强的 CH36_58-0 有义
<400> 42
ttcatcccag aaacttgttc tgtttttcca taggattctc tttggacagt gcccttac 58
<210> 43
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成的：增强的 CH36_63-0 反义
<400> 43
gtaagggcac tgtccaaaga gaatcctatg gaaaaacaga acaagtttct ggggttaatg 60
aat 63
<210> 44
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成的：增强的 CH36_63-0 有义
<400> 44
attcattaac cccagaaact tgttctgttt ttccatagga ttctctttgg acagtgccct 60
tac 63

[0012]

<210> 45
<211> 68
<212> DNA
<213> 人工序列
<220>
<223> 合成的：CH36_68-0 反义
<400> 45
gtaagggcac tgtccaaaga gaatcctatg gaaaaacaga acaagtttct ggggttactg 60
aatgaatg 68
<210> 46
<211> 68
<212> DNA
<213> 人工序列
<220>
<223> 合成的：CH36_68-0 有义
<400> 46
cattcattca gtaaccccag aaacttgttc tgtttttcca taggattctc tttggacagt 60
gcccttac 68
<210> 47
<211> 73
<212> DNA
<213> 人工序列
<220>

[0013]

<223> 合成的: CH36_73-0 反义
<400> 47
gtaagggcac tgtccaaaga gaatcctatg gaaaaacaga acaagtttct ggggttactg 60
aatgaatgct ttt 73
<210> 48
<211> 73
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH36_73-0 有义
<400> 48
aaaagcattc attcagtaac cccagaaact tgttctgttt ttccatagga ttctctttgg 60
acagtgccct tac 73
<210> 49
<211> 5
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 分支点共有序列
<220>
<221> misc_feature
<222> (3)..(3)
<223> n 是 a, c, g 或 t
<400> 49
ytnay 5
<210> 50
<211> 7
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 分支点序列
<400> 50
tattaac 7
<210> 51
<211> 7
<212> DNA
<213> 人工序列
<220>
<223> 合成的: 分支点序列
<400> 51
gttaata 7
<210> 52
<211> 7
<212> DNA

[0014]

<213> 人工序列
<220>
<223> 合成的：分支点序列
<400> 52
tactgac 7
<210> 53
<211> 9
<212> DNA
<213> 人工序列
<220>
<223> 合成的：多聚嘧啶片
<400> 53
tttttttct 9
<210> 54
<211> 13
<212> DNA
<213> 人工序列
<220>
<223> 合成的：多聚嘧啶片
<400> 54
ttttttctt ttt 13
<210> 55
<211> 16
<212> DNA
<213> 人工序列
<220>
<223> 合成的：多聚嘧啶片
<400> 55
cttcttctct tcttcc 16
<210> 56
<211> 7
<212> PRT
<213> 人工序列
<220>
<223> 合成的：SV40 核定位信号(NLS)
<400> 56
Pro Lys Lys Lys Arg Lys Val
1 5
<210> 57
<211> 7
<212> PRT
<213> 人工序列
<220>
<223> 合成的：SV40 核定位信号(NLS)

<400> 57
Pro Lys Lys Lys Arg Arg Val
1 5
<210> 58
<211> 16
<212> PRT
<213> 人工序列
<220>
<223> 合成的: 核质蛋白核定位信号(NLS)
<400> 58
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15
<210> 59
<211> 24
<212> DNA
<213> 化脓性链球菌
<220>
<221> misc_feature
<223> 化脓性链球菌靶序列和 PAM
<220>
<221> misc_feature
[0015] <222> (1)..(22)
<223> n 是 a, c, g 或 t
<220>
<221> misc_feature
<222> (20)..(21)
<223> n 可存在或可不存在
<400> 59
nnnnnnnnnn nnnnnnnnnn nnrg 24
<210> 60
<211> 6
<212> PRT
<213> 人工序列
<220>
<223> 合成的: 6xHis
<400> 60
His His His His His His
1 5
<210> 61
<211> 100
<212> RNA
<213> 人工序列
<220>
<223> 合成的: GAA5 sgRNA

<400> 61
agcccgcuuu cuucucccgc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100
<210> 62
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_25-0_F
<400> 62
gagaagaagt attaactttt tttct 25
<210> 63
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_50-0_F
<400> 63
tggaagaaga gaagaagctg ggtattaacg catttttttc tttttaattc 50
<210> 64
<211> 75
<212> DNA
[0016] <213> 人工序列
<220>
<223> 合成的: GAA_75-0_F
<400> 64
tggaagaaga gaagaagctg ggaatgtgtc agtaaggcat gtattaacgc aaatgaattt 60
ttttcttttt aattc 75
<210> 65
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_100-0_F
<400> 65
tggaagaaga gaagaagctg ggaatgtgtc agtaaggcat gagccaccgc gccagccatg 60
caaatctatt aacgcaaatg aatttttttc tttttaattc 100
<210> 66
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_25-0_R
<400> 66

agaaaaaaag ttaatacttc ttctc 25

<210> 67
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_50-0_R
<400> 67
gaattaaaaa gaaaaaaatg cgttaatacc cagcttcttc tcttcttcca 50

<210> 68
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_75-0_R
<400> 68
gaattaaaaa gaaaaaaatt catttgcgtt aatacatgcc ttactgacac attcccagct 60
tcttctcttc ttcca 75

<210> 69
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_100-0_R
<400> 69
gaattaaaaa gaaaaaaatt catttgcgtt aatagatttg catggctggc gcggtggctc 60
atgccttact gacacattcc cagcttcttc tcttcttcca 100

[0017]

<210> 70
<211> 100
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH42 sgRNA
<400> 70
gucagucuca cagguuacag guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

<210> 71
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_25-0_F
<400> 71
aacctgtgag actggaccag gtaag 25

[0018]

<210> 72
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_25-0_R
<400> 72
cttacctggt ccagtctcac aggtt 25
<210> 73
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_50-0_F
<400> 73
aacctgtgag actggaccag gtaagactca cacttgcgaa accggcccag 50
<210> 74
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_50-0_R
<400> 74
ctgggccggt ttcgcaagtg tgagtcttac ctggtccagt ctcacaggtt 50
<210> 75
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_75-0_F
<400> 75
aacctgtgag actggaccag gtaagcgacg cgcatttctc acacggcagg actcacactt 60
gcgaaaccgg cccag 75
<210> 76
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_75-0_R
<400> 76
ctgggccggt ttcgcaagtg tgagtcctgc cgtgtgagaa atgcgcgtcg cttacctggt 60
ccagtctcac aggtt 75
<210> 77
<211> 100

[0019]

<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_100-0_F
<400> 77
aacctgtgag actggaccag gtaagcgacg cgcatttctc acacggcagg gagggccaca 60
cgcgtttgtt tctcaactca cacttgcgaa accggcccag 100
<210> 78
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_100-0_R
<400> 78
ctgggccggt ttcgcaagtg tgagttgaga aacaaacgcg tgtggccctc cctgccgtgt 60
gagaaatgcg cgtcgcttac ctggtccagt ctcacaggtt 100
<210> 79
<211> 125
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_125-0_F
<400> 79
aacctgtgag actggaccag gtaagcgacg cgcatttctc acacggcagg gagggccaca 60
cgcgtttgtt tctcacacga tgggcagggc gacacatgtt actcacactt gcgaaaccgg 120
cccag 125
<210> 80
<211> 125
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH42_125-0_R
<400> 80
ctgggccggt ttcgcaagtg tgagtaacat gtgtcgccct gcccatcgtg tgagaaacaa 60
acgcgtgtgg ccctccctgc cgtgtgagaa atgcgcgtcg cttacctggt ccagtctcac 120
aggtt 125
<210> 81
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH32 sgRNA 间隔区
<400> 81
ucuuuggaca gcguccauac 20

[0020]

<210> 82
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH32 靶基因(PAM)
<400> 82
tctttggaca gcgtccatac tgg 23
<210> 83
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH34 sgRNA 间隔区
<400> 83
uggacagcgu ccauacuggu 20
<210> 84
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH34 靶基因(PAM)
<400> 84
tggacagcgt ccatactggt ggg 23
<210> 85
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH36 sgRNA 间隔区
<400> 85
guauccaaaa cccaccagua 20
<210> 86
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成的: CH36 靶基因 (PAM)
<400> 86
gtatccaaaa cccaccagta tgg 23
<210> 87
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 合成的: CH42 靶基因 (PAM)
<400> 87
gtcagtctca caggttacag ggg 23
<210> 88
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH42 sgRNA 间隔区
<400> 88
gucagucuca cagguuacag 20
<210> 89
<211> 106
<212> RNA
<213> 人工序列
<220>
<223> 合成的: 鼠 CH34-CTX1 sgRNA
<400> 89
usgsgsacaa cgcccguauu gguguuuuag agcuagaaau agcaaguuaa aauaaggcua 60
guccguuauc aacuugaaaa aguggcaccg agucggugcu sususu 106

[0021]

<210> 90
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成的: WT GAA 靶基因(PAM)
<400> 90
agcccgcttt cttctcccgc agg 23
<210> 91
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: WT GAA5 sgRNA 间隔区
<400> 91
agcccgcuuu cuucucccgc 20
<210> 92
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 合成的:突变体 (Mt) GAA 靶基因(PAM)
<400> 92

agcccgcttg cttctcccgc agg 23

<210> 93
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: Mt GAA sgRNA 间隔区
<400> 93
agcccgcuug cuucucccgc 20

<210> 94
<211> 100
<212> RNA
<213> 人工序列
<220>
<223> 合成的: Mt GAA sgRNA
<400> 94
agcccgcuug cuucucccgc guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60
cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 100

[0022]

<210> 95
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_30 F
<400> 95
aagagaagaa gtattaactt ttttttcttt 30

<210> 96
<211> 35
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_35 F
<400> 96
aagaagagaa gaagtattaa ctttttttct tttta 35

<210> 97
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_40 F
<400> 97
ggaagaagag aagaagtatt aacttttttt ctttttaatt 40

<210> 98
<211> 43

<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_45 F
<400> 98
tggaagaaga gaagaagctt attaactttt tttcttttaa ttc 43
<210> 99
<211> 53
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_55 F
<400> 99
tggaagaaga gaagaagctg ggtattaacg caaatttttt ttcttttaa ttc 53
<210> 100
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_60 F
<400> 100
[0023] tggaagaaga gaagaagctg ggaatgttat taacgcaaat gatttttttc tttttaattc 60
<210> 101
<211> 65
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_65 F
<400> 101
tggaagaaga gaagaagctg ggaatggtca gtatattaac gcaaatgttt ttttcttttt 60
aattc 65
<210> 102
<211> 70
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_70 F
<400> 102
tggaagaaga gaagaagctg ggaatgtgtc agtaagtatt aacgcaaatg aatttttttc 60
ttttttaattc 70
<210> 103
<211> 80
<212> DNA
<213> 人工序列

<220>
<223> 合成的: GAA_80 F

<400> 103
tggaagaaga gaagaagctg ggaatgtgtc agtaaggcat gagccatatt aacgcaaatg 60
aatttttttc tttttaattc 80
<210> 104
<211> 85
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_85 F
<400> 104
tggaagaaga gaagaagctg ggaatgtgtc agtaaggcat gagccaccgc gtattaacgc 60
aaatgaattt ttttcttttt aattc 85
<210> 105
<211> 90
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_90 F
[0024] <400> 105
tggaagaaga gaagaagctg ggaatgtgtc agtaaggcat gagccaccgc gccagctatt 60
aacgcaaatg aatttttttc tttttaattc 90
<210> 106
<211> 96
<212> DNA
<213> 人工序列
<220>
<223> 合成的: GAA_95 F
<400> 106
tgggaagaag agaagaagct gggaatgtgt cagtaaggca tgagccaccg cgccagccat 60
gctattaacg caaatgaatt tttttctttt taattc 96
<210> 107
<211> 20
<212> RNA
<213> 人工序列
<220>
<223> 合成的: CH32 突变体-CTX1 sgRNA 间隔区
<400> 107
ucuuuggaca gcguccauac 20
<210> 108
<211> 20
<212> RNA

<213> 人工序列
<220>
<223> 合成的：CH34 突变体-CTX1 sgRNA 间隔区
<400> 108
uggacagcgu ccauacuggu 20
<210> 109
<211> 16
<212> DNA
<213> 人工序列
<220>
[0025] <223> 合成的：多聚嘧啶片
<400> 109
cttgttctgt tttttt 16
<210> 110
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 合成的：外显子序列
<400> 110
gattctcttt ggacagcgcc cttact 26

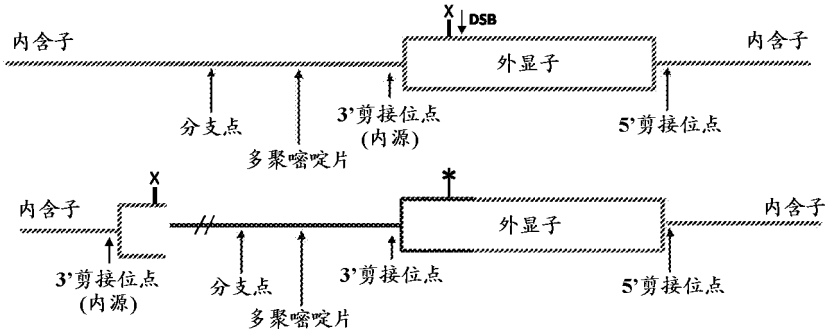

DSB
内含子
外显子
内含子
分支点
3'剪接位点
(内源)
5'剪接位点
多聚嘧啶片
内含子
外显子
内含子
3'剪接位点
(内源)
分支点
3'剪接位点
5'剪接位点
多聚嘧啶片

图 1

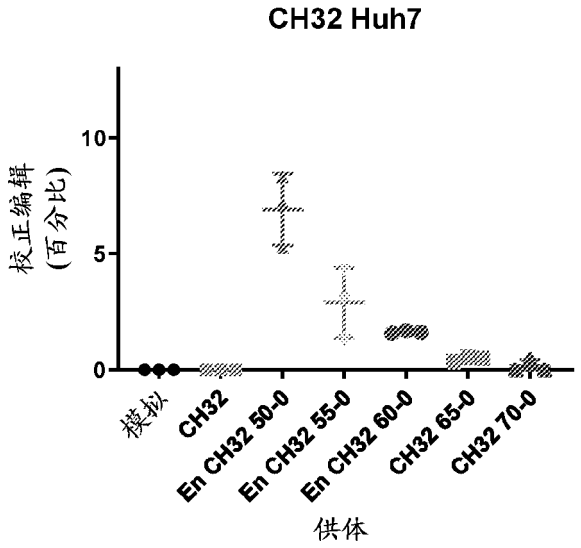

CH32 Huh7
校正编辑
(百分比)
10
5
0
模拟
CH32
En CH32 50-0
En CH32 55-0
En CH32 60-0
CH32 65-0
CH32 70-0
供体

图 2A

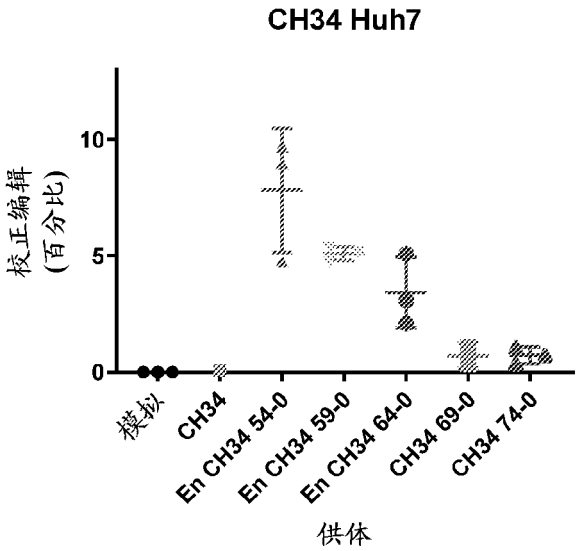

CH34 Huh7
校正编辑
(百分比)
10
5
0
模拟
CH34
En CH34 54-0
En CH34 59-0
En CH34 64-0
CH34 69-0
CH34 74-0
供体

图 2B

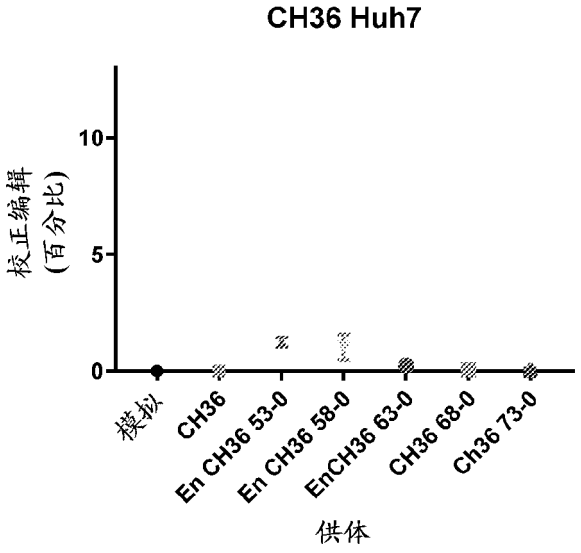

CH36 Huh7
校正编辑
(百分比)
10
5
0
模拟
CH36
En CH36 53-0
En CH36 58-0
EnCH36 63-0
CH36 68-0
Ch36 73-0
供体

图 2C

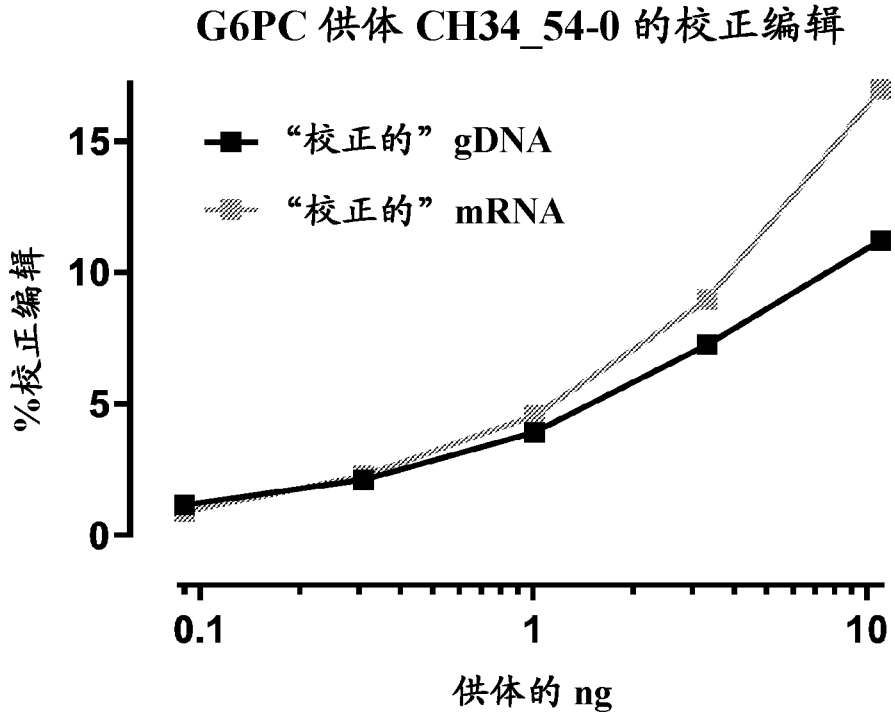

G6PC 供体 CH34_54-0 的校正编辑
“校正的”gDNA
“校正的”mRNA
%校正编辑
15
10
5
0
0.1
1
10
供体的 ng

图 3

体内插入

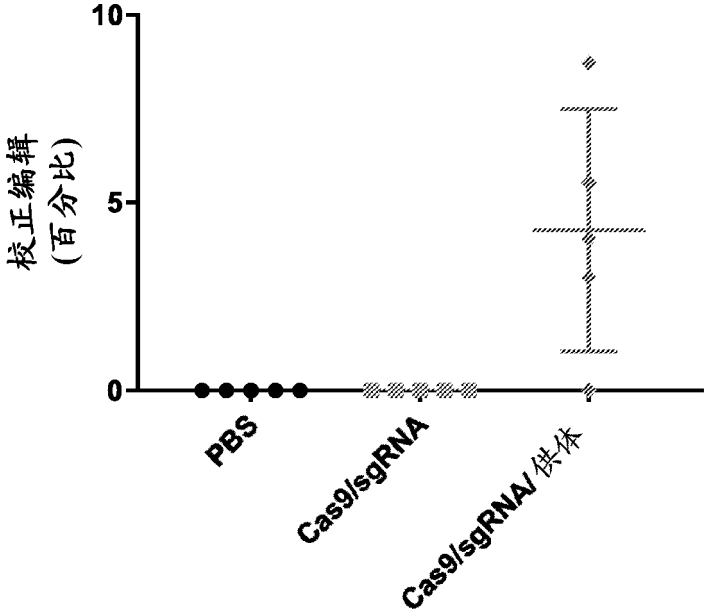

校正编辑
(百分比)
10
5
0
PBS
Cas9/sgRNA
Cas9/sgRNA/供体

图 4A

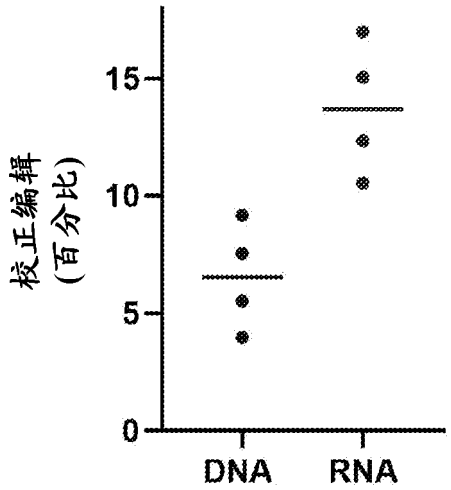

校正编辑
(百分比)
15
10
5
0
DNA
RNA

图 4B

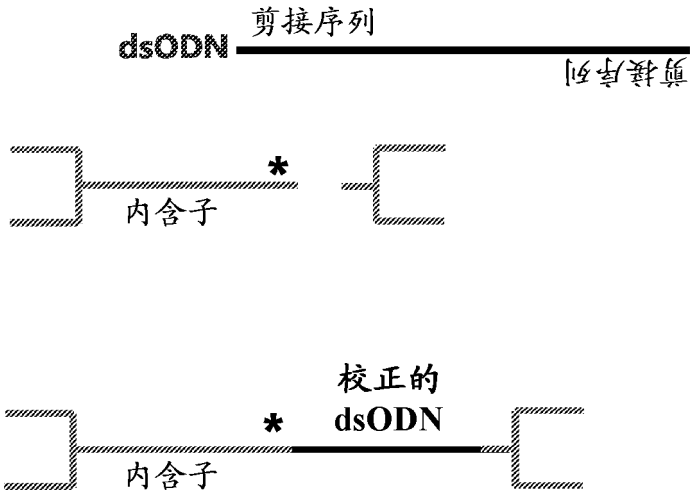


图 5

50nt 供体 dsODN:
有义链:
反义链:
分支点
多聚
嘧啶片
5′ - tggaagaagagaagaagctgggTATTAACgcatttttttctttttaattc - 3′
3′ - accttcttctcttcttcgacccATAATTGcgtaaaaaaaagaaaaattaag - 5′

图 6A

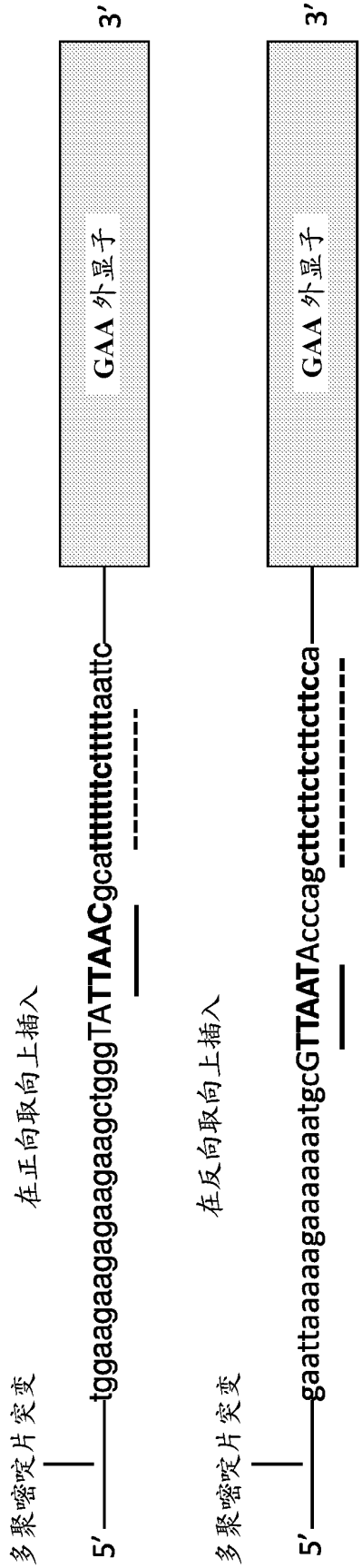

3′
GAA 外显子
在正向取向上插入
多聚嘧啶片突变
5′
tggaagagaagaagctggGTATTAACgcatttttttctttttaattc
3′
GAA 外显子
在反向取向上插入
多聚嘧啶片突变
5′
gaattaaaaagaaaaaaatgcGTTAATAcccagcttcttctcttcttcca

图 6B

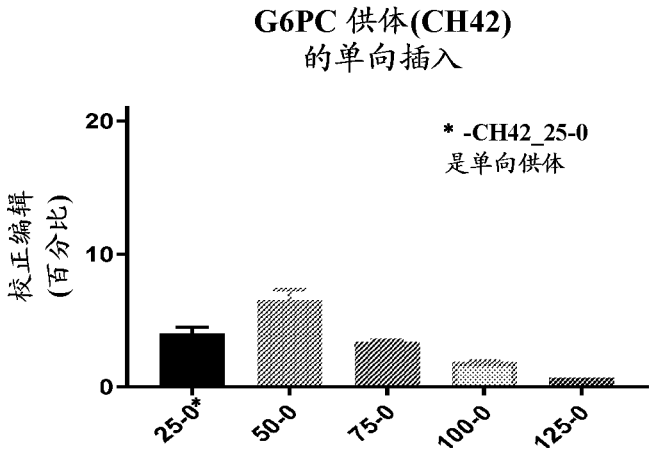

G6PC 供体(CH42)
的单向插入
* -CH42_25-0
是单向供体
校正编辑
(百分比)
20
10
0
25-0*
50-0
75-0
100-0
125-0

图 7A

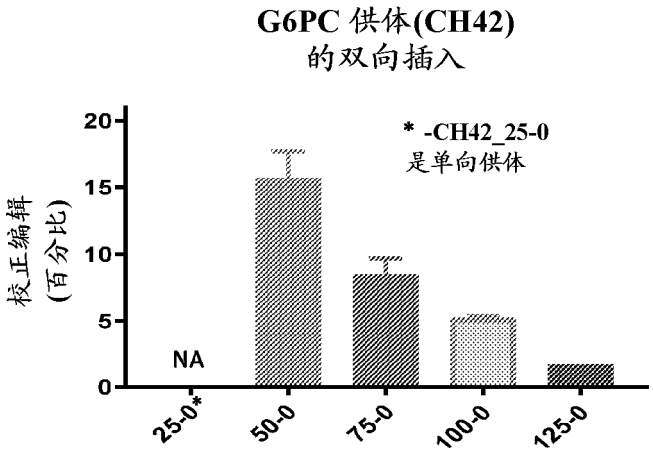

G6PC 供体(CH42)
的双向插入
* -CH42_25-0
是单向供体
校正编辑
(百分比)
20
15
10
5
0
NA
25-0*
50-0
75-0
100-0
125-0

图 7B

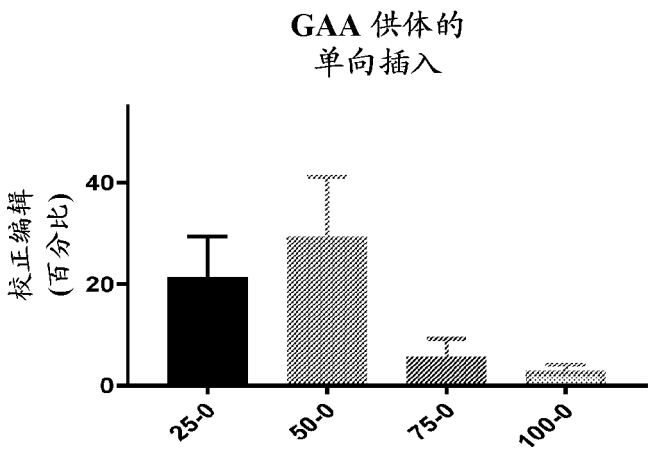

GAA 供体的
单向插入
校正编辑
(百分比)
40
20
0
25-0
50-0
75-0
100-0

图 7C

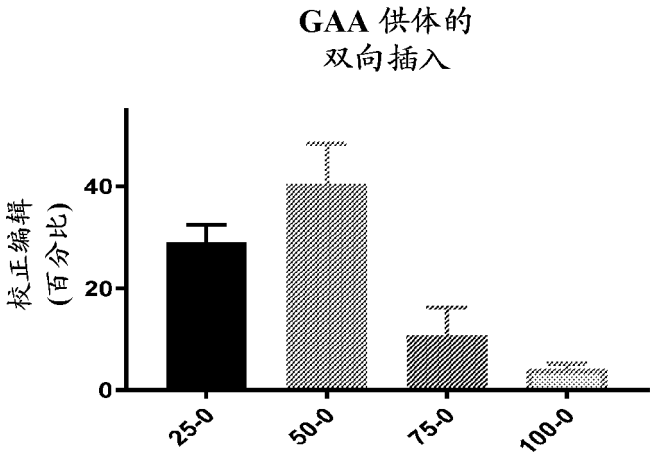

GAA 供体的
双向插入
校正编辑
(百分比)
40
20
0
25-0
50-0
75-0
100-0

图 7D

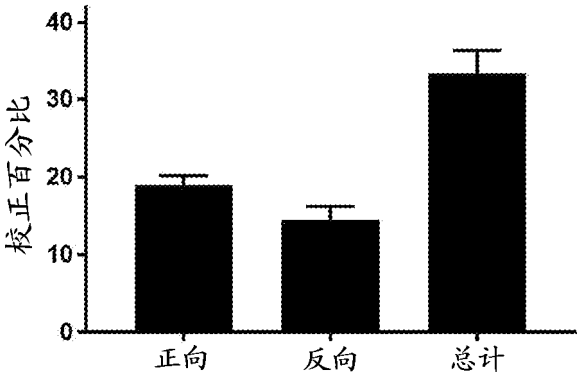

PHH 中的插入校正
校正百分比
40
30
20
10
0
正向
反向
总计

图 8

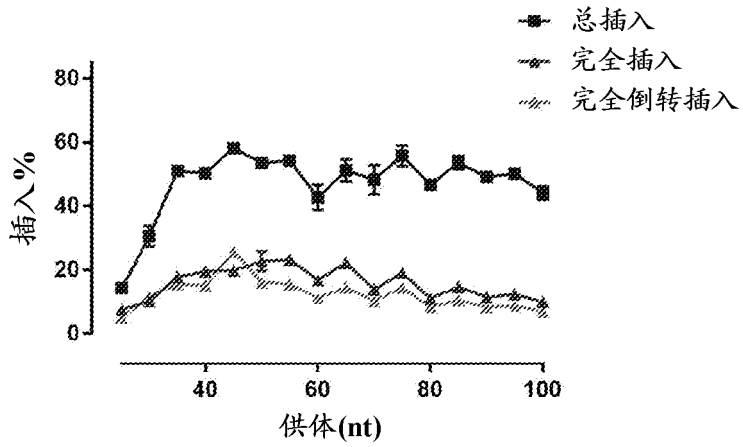

Huh7 中的插入
总插入
完全插入
完全倒转插入
插入%
80
60
40
20
0
40
60
80
100
供体(nt)

图 9

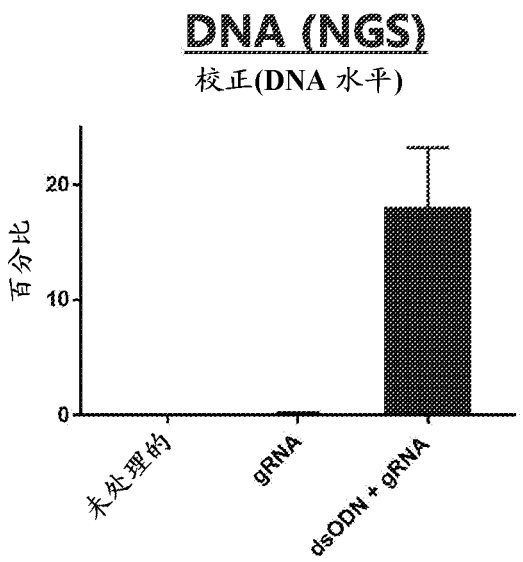

DNA (NGS)
校正(DNA 水平)
百分比
20
10
0
未处理的
gRNA
dsODN + gRNA

图 10A

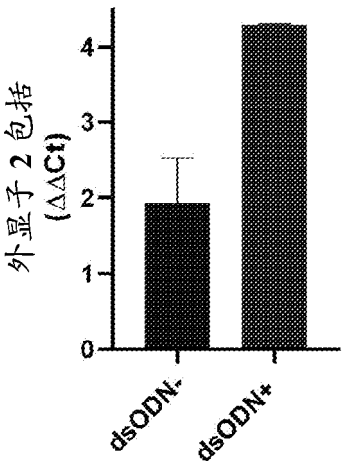

mRNA 外显子 2 包括(qPCR)
外显子 2 包括
(ΔΔCt)
4
3
2
1
0
dsODN-
dsODN+

图 10B

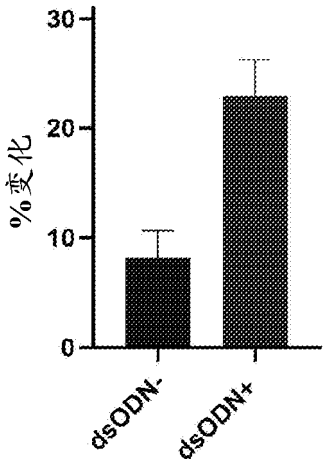


图 10C