(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 109715801 B
(45)授权公告日 2022.11.01

(21)申请号 201680080398.1
(22)申请日 2016.12.01
(65)同一申请的已公布的文献号
申请公布号 CN 109715801 A
(43)申请公布日 2019.05.03
(30)优先权数据
62/261661 2015.12.01 US
62/324056 2016.04.18 US
(85)PCT国际申请进入国家阶段日
2018.07.27
(86)PCT国际申请的申请数据
PCT/IB2016/001845 2016.12.01
(87)PCT国际申请的公布数据
WO2017/093804 EN 2017.06.08
(73)专利权人 克里斯普治疗股份公司
地址 瑞士巴塞尔

(72)发明人 C.A.科万 R.L.博戈拉 J.李
A.S.伦贝里 M.J.约翰
J.W.斯特宾斯 T.T.阮

(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 杜艳玲 周齐宏
(51)Int.Cl.
C12N 15/10(2006.01)
C12N 9/22(2006.01)
C12N 15/113(2006.01)

(56)对比文件
谢胜松等.CRISPR/Cas9系统中sgRNA设计与脱靶效应评估.《遗传》.2015,(第11期),

审查员 王小玉

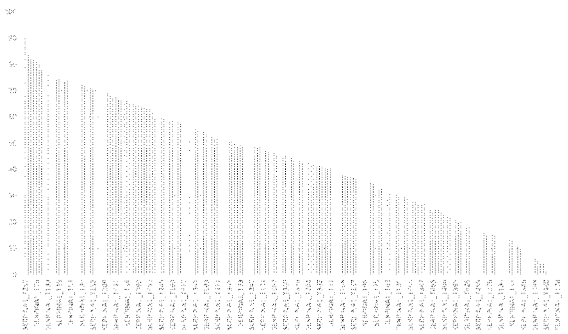
权利要求书1页 说明书82页 附图4页

(54)发明名称

用于治疗α1抗胰蛋白酶缺乏的材料和方法

(57)摘要

本申请提供了用于离体和在体内治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的材料和方法。另外，本申请提供了用于通过基因组编辑来编辑细胞中的SERPINA1基因的材料和方法。



CN 109715801 B

1.包含间隔物序列的引导核糖核酸(gRNA)在制备用于通过基因组编辑来编辑来自具有α1抗胰蛋白酶缺乏(AATD)的患者的细胞中的SERPINA1基因的药物中的用途,其中所述间隔物序列选自SEQ ID NOs: 55,709、60,592、55,559、55,574、55,575、55,577、55,578、55,584、55,586、55,587、55,600、55,602、55,605、55,607、55,616、55,625、55,627、55,628、55,633、55,635、55,639、55,641、55,642、55,646、55,669、55,671、55,672、55,686、55,688、55,694、55,696、55,701、55,704、55,705、55,708、55,716、55,718、55,721、55,722、55,724、55,726、55,727、55,730、55,732、55,743、55,744、55,748、55,749、55,751、55,754、55,762、55,764、55,771、55,778、55,781-55,783、60,476、60,477、60,480-60,482、60,485、60,487、60,490、60,494、60,496、60,497、60,504、60,506、60,515、60,517、60,519、60,527、60,529、60,539-60,541、60,555、60,557、60,558、60,564、60,569-60,572、60,575、60,578、60,581、60,588、60,591、60,599、60,601、60,605、60,607、60,618-60,620、60,622、60,630、60,632、60,633、60,637、60,639、60,640、60,642、60,645、60,648、60,657和60,660中的核酸序列。

2.根据权利要求1所述的用途,其中所述间隔物序列是SEQ ID NO: 55,709或SEQ ID NO: 60,592中的核酸序列。

3.根据权利要求1或2所述的用途,其中所述gRNA是单分子引导RNA (sgRNA)。

4.根据权利要求1-3任一项所述的用途,其中所述gRNA或sgRNA是经修饰的gRNA或经修饰的sgRNA。

5.根据权利要求1-4任一项所述的用途,其中所述细胞是肝细胞。

6.组合物,其包含引导核糖核酸(gRNA),所述引导核糖核酸(gRNA)包含间隔物序列,所述间隔物序列选自SEQ ID NOs: 55,709、60,592、55,559、55,574、55,575、55,577、55,578、55,584、55,586、55,587、55,600、55,602、55,605、55,607、55,616、55,625、55,627、55,628、55,633、55,635、55,639、55,641、55,642、55,646、55,669、55,671、55,672、55,686、55,688、55,694、55,696、55,701、55,704、55,705、55,708、55,716、55,718、55,721、55,722、55,724、55,726、55,727、55,730、55,732、55,743、55,744、55,748、55,749、55,751、55,754、55,762、55,764、55,771、55,778、55,781-55,783、60,476、60,477、60,480-60,482、60,485、60,487、60,490、60,494、60,496、60,497、60,504、60,506、60,515、60,517、60,519、60,527、60,529、60,539-60,541、60,555、60,557、60,558、60,564、60,569-60,572、60,575、60,578、60,581、60,588、60,591、60,599、60,601、60,605、60,607、60,618-60,620、60,622、60,630、60,632、60,633、60,637、60,639、60,640、60,642、60,645、60,648、60,657和60,660中的核酸序列。

7.根据权利要求6所述的组合物,其中所述gRNA是单分子引导RNA (sgRNA)。

8.根据权利要求6所述的组合物,其中所述gRNA是经修饰的gRNA。

9.根据权利要求7所述的组合物,其中所述sgRNA是经修饰的sgRNA。

用于治疗α1抗胰蛋白酶缺乏的材料和方法

[0001] 相关申请

[0002] 本申请要求2015年12月1日提交的美国临时申请号62/261,661和2016年4月18日提交的美国临时申请号62/324,056的权益，它们中的每一篇的内容通过引用以其整体并入本文。

发明领域

[0003] 本申请提供了离体和在体内(*ex vivo* and *in vivo*)治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的材料和方法。另外，本申请提供了通过基因组编辑来调节细胞中的SERPINA1基因和/或α1抗胰蛋白酶(AAT)蛋白的表达、功能和/或活性的材料和方法。

[0004] 通过引用并入序列表

[0005] 本申请含有计算机可读形式的序列表(文件名:CRIS010001WO_ST25.txt;13,706,379字节-ASCII文本文件;2016年11月30日创建)，其通过引用以其整体并入本文并形成本公开的一部分。

[0006] 发明背景

[0007] α1抗胰蛋白酶缺乏(AATD)是常染色体隐性遗传病，其特征在于，对下列增加的风险:慢性阻塞性肺疾病，包括肺气肿、气道阻塞、和/或慢性支气管炎，在至40-50岁的成年人中具有明显表现;肝脏疾病，在小部分受影响患者中，具有如阻塞性黄疸的表现;和在生命早期增加的转氨酶水平。在缺乏儿童期间肝脏并发症史的情况下，成年人中的肝脏疾病甚至可以表现为纤维化和硬化。具有AATD的患者也具有较高的肝细胞癌发病率。(参见例如，Stoller JK, Aboussouan LS,“A review of alpha1-antitrypsin deficiency.”Am J Respir Crit Care Med 2012, 185(3):246-259; Teckman JH,“Liver disease in alpha-1 antitrypsin deficiency:current understanding and future therapy.”COPD 2013, 10 Suppl 1:35-43; 和Stoller JK, et al,“Alpha-1 Antitrypsin Deficiency.”In:GeneReviews(R). 由Pagon RA,等人编辑,Seattle (WA); 1993)。

[0008] AATD的诊断是通过证明血清中低浓度的α1抗胰蛋白酶(AAT)、随后检测功能上缺乏的AAT蛋白或检测编码α1抗胰蛋白酶的基因SERPINA1中的双等位基因致病变体而确定的。(参见例如,Aboussouan LS, Stoller JK;“Detection of alpha-1 antitrypsin deficiency:a review.”Respir Med 2009, 103(3):335-341)。AATD的流行率是大约40:100K (瑞典血统 – 50-60:100K)。描述了AAT基因(SERPINA1)的多于100个变体，但大多数患者(至少具有肝脏表型)是对PI Z纯合的(PI, GLU342LYS ON M1A [dbSNP:rs28929474])。

[0009] 目前对AATD的疗法包括:对慢性阻塞性肺疾病的标准疗法，其包括支气管扩张剂、吸入的皮质类固醇、肺康复、补充供氧和疫苗接种(例如流感和肺炎球菌)。也应用对AATD的特别疗法，包括定期静脉内输注混合的人血清α抗胰蛋白酶(增补疗法(augmentation therapy))。(参见例如,Mohanka M, Khemasuwan D, Stoller JK,“A review of augmentation therapy for alpha-1 antitrypsin deficiency.”Expert Opin Biol

Ther 2012，12(6):685-700)。具有终末期肺疾病的个体常常不得不经历肺移植。对具有肝脏疾病的患者的主要治疗是肝移植，因为其它途径诸如常规疗法(化学分子伴侣、刺激自噬、抗纤维化疗法)常常具有最低限度的效果。

[0010] 近来测试了一些更新的途径。(参见例如，Wewers MD，等人，“Replacement therapy for alpha 1-antitrypsin deficiency associated with emphysema.”N Engl J Med 1987，316(17):1055-1062;Wewers MD，Crystal RG，“Alpha-1 antitrypsin augmentation therapy.”COPD 2013，10 Suppl 1:64-67;和Teckman JH，“Lack of effect of oral 4-phenylbutyrate on serum alpha-1-antitrypsin in patients with alpha-1-antitrypsin deficiency:a preliminary study.”J Pediatr Gastroenterol Nutr 2004，39(1):34-37)。ARC-AAT是单一剂量1期临床研究和ALNY-AAT是1期临床研究。这两个研究都利用RNA干扰途径减少肝细胞中沉积的突变体AAT的量。但是，这些途径不针对循环中正常AAT的缺乏，并且依赖于与增补疗法联合。此外，rAAV1-CB-hAAT是2 期临床研究，其联合内源性AAT与微RNA的下调和外源性AAT的刺激性上调。(参见例如，Flotte TR，等人，“Phase 2 clinical trial of a recombinant adeno-associated viral vector expressing alpha1-antitrypsin:interim results.”Hum Gene Ther 2011，22(10):1239-1247;Mueller C，等人，“Sustained miRNA-mediated knockdown of mutant AAT with simultaneous augmentation of wild-type AAT has minimal effect on global liver miRNA profiles.”Mol Ther 2012，20(3):590-600)。

[0011] 基因组工程指用于活生物体的遗传信息(基因组)的靶向特异性修饰的策略和技术。基因组工程是非常活跃的研究领域，这是因为宽范围的可能应用，特别是在人健康领域中;例如携带有害突变的基因的校正，或为了探究基因的功能。为将基因插入活细胞中而开发的早期技术，诸如转基因作用，经常受限于新序列向基因组中的插入的随机性质。新基因经常盲目地定位，且可能已经被失活或妨碍其它基因的功能、或甚至造成严重的不希望的效应。此外，这些技术通常不提供再现性程度，因为没有确保新序列将在两个不同细胞中插入在相同位置。更近的基因组工程策略诸如ZFNs、TALENs、HEs和MegaTALs使得DNA的特定区域能够被修饰，由此与早期技术相比增加校正或插入的精确度，并提供某种再现性程度。尽管如此，这样的新近基因组工程策略具有限制。

[0012] 尽管一直在尝试解决AATD的世界上的研究人员和医学专业人员做出了努力，并且尽管基因组工程途径有希望，仍然存在对开发用于AATD的安全且有效的治疗的关键性需要。

[0013] 解决AATD的先前途径具有严重限制，诸如有限的成功和足够可利用的疗法的缺乏。本发明如下解决了这些问题:使用基因组工程工具来建立对基因组的永久变化，其可以用单一治疗恢复AAT蛋白活性。因而，本发明校正了造成该疾病的潜在遗传缺陷。

[0014] 发明概述

[0015] 本文提供了如下建立对基因组的永久变化的细胞、离体和体内方法:通过基因组编辑，在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近插入、缺失、校正或调节一个或多个突变或外显子的表达或功能，或将SERPINA1 cDNA或小基因敲入安全港基因座，并恢复α1抗胰蛋白酶(AAT)蛋白活性，所述方法可以用于治疗α1抗胰蛋白酶缺乏(AATD)。本文还提供了用于执行这样的方法的组分、试剂盒和组合物。还提供了通过它

们生产的细胞。

[0016] 本文提供了通过基因组编辑来编辑人细胞中的SERPINA1基因的方法，所述方法包括下述步骤：在所述人细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现一个或多个单链断裂(SSB)或一个或多个双链断裂(DSB)，所述一个或多个单链断裂(SSB)或一个或多个双链断裂(DSB)：在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近，其导致一个或多个突变或外显子的表达或功能的永久插入、缺失、校正或调节中的至少一种，所述一个或多个突变或外显子在SERPINA1基因内或附近、或影响SERPINA1基因的表达或功能；或在安全港基因座内或附近，其导致SERPINA1基因或小基因的永久插入，和导致α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0017] 本文还提供了通过基因组编辑将SERPINA1基因插入人细胞中的方法，所述方法包括下述步骤：在所述人细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现在安全港基因座内或附近的一个或多个单链断裂(SSB)或双链断裂(DSB)，其导致SERPINA1基因或小基因的永久插入，和导致α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0018] 在一个方面，本文提供了通过基因组编辑来编辑人细胞中的SERPINA1基因的方法，所述方法包括在所述细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现一个或多个单链断裂(SSB)或双链断裂(DSB)，所述一个或多个单链断裂(SSB)或双链断裂(DSB)：在所述细胞的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近，其导致一个或多个突变或外显子在SERPINA1基因内或附近的永久缺失、插入或校正；或在安全港基因座内或附近，其导致SERPINA1基因或小基因的永久插入，和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0019] 在另一个方面，本文提供了治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法，所述方法包括下述步骤：建立患者特异性的诱导性多能干细胞(iPSC)；在iPSC的SERPINA1基因或编码iPSC的SERPINA1基因的调节元件的其它DNA序列内或附近编辑，或在iPSC的安全港基因座内或附近编辑；使编辑过基因组的iPSC分化成肝细胞；和将所述肝细胞移植进所述患者中。

[0020] 在某些实施方案中，所述建立患者特异性的诱导性多能干细胞(iPSC)的步骤包括：从所述患者分离体细胞；和将一组多能性相关基因引入所述体细胞中以诱导所述细胞变成多能干细胞。在某些实施方案中，所述体细胞是成纤维细胞。在某些实施方案中，所述组的多能性相关基因是一种或多种选自以下的基因：OCT4、SOX2、KLF4、Lin28、NANOG和cMYC。

[0021] 在iPSC的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近编辑、或在iPSC的SERPINA1基因的安全港基因座内或附近编辑、或在iPSC的SERPINA1基因的第一个外显子的基因座内或附近编辑的步骤可以包括，在所述iPSC中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现一个或多个单链断裂(SSB)或双链断裂(DSB)，所述一个或多个单链断裂(SSB)或双链断裂(DSB)：在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近，其导致一个或多个突变或外显子的表达或功能的永久插入、校正、缺失或调节，所述一个或多个突变或外显子在SERPINA1基因内或附近、或影响SERPINA1基因的表达或功能；或在安全港基因座内或附近，其导致SERPINA1 基因的永久插入，和导致ATT蛋白活性的恢复。所述安全港基因座可以选自AAVS1 (PPP1R12C)、ALB、Angptl3、 ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。所述安全港基因座可以选自AAVS1(PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2和TTR的外显子1-2。

[0022] 在某些实施方案中，所述编辑iPSC的SERPINA1基因的步骤包括在所述iPSC中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现在SERPINA1基因内或附近的一个或多个双链断裂(DSB)，其导致SERPINA1基因内或附近的一个或多个突变的永久缺失、插入或校正和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0023] 在某些实施方案中，使所述编辑过基因组的iPSC分化成肝祖细胞或肝细胞的步骤包括以下的一项或多项：使所述编辑过基因组的iPSC与激活素、B27补体、FGF4、HGF、BMP2、BMP4、制癌蛋白M或Dexametason中的一种或多种接触。

[0024] 在某些实施方案中，将所述肝细胞移植进所述患者中的步骤包括通过移植、局部注射或全身输注或它们的组合将所述肝细胞移植进所述患者中。

[0025] 在另一个方面，本文提供了治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法，所述方法包括下述步骤：执行所述患者的肝的活组织检查；分离肝特异性的祖细胞或原代肝细胞；在所述祖细胞或原代肝细胞的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近编辑，或在所述祖细胞或原代肝细胞的安全港基因座内或附近编辑；和将所述编辑过基因组的祖细胞或原代肝细胞移植进所述患者中。

[0026] 在某些实施方案中，从所述患者分离肝特异性的祖细胞或原代肝细胞的步骤包括：用消化酶灌注新鲜肝组织、细胞差速离心、和细胞培养或它们的组合。

[0027] 在某些实施方案中，在所述祖细胞或原代肝细胞的SERPINA1基因或编码所述祖细胞或原代肝细胞的SERPINA1基因的调节元件的其它DNA序列内或附近编辑、或在所述祖细胞或原代肝细胞的白细胞的安全港基因座内或附近编辑的步骤包括，在所述祖细胞或原代肝细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现一个或多个单链断裂(SSB)或双链断裂(DSB)，所述一个或多个单链断裂(SSB)或双链断裂(DSB)在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近，其导致一个或多个突变或外显子的表达或功能的永久插入、校正、缺失或调节，所述一个或多个突变或外显子在SERPINA1基因内或附近、或影响SERPINA1基因的表达或功能，或在安全港基因座内或附近编辑，其导致一个或多个突变在SERPINA1基因内或附近的永久缺失、插入或校正和α1抗胰蛋白酶(AAT)蛋白活性的恢复。所述安全港基因座可以选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。所述安全港基因座可以选自AAVS1(PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2和TTR的外显子1-2。

[0028] 在某些实施方案中，将祖细胞或原代肝细胞移植进患者中的步骤包括，通过移植、局部注射或全身输注或它们的组合将祖细胞或原代肝细胞移植进患者中。

[0029] 在一个方面，本文提供了治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法，

所述方法包括下述步骤:i)从所述患者分离间充质干细胞; ii)在所述间充质干细胞的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近编辑,或在所述间充质细胞的SERPINA1基因的安全港基因座内或附近编辑;iii)使所述编辑过基因组的间充质干细胞分化成肝细胞;和iv)将所述肝细胞移植进所述患者中。

[0030] 在一个方面,本文提供了治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法,所述方法包括下述步骤:执行患者骨髓的活组织检查;分离间充质干细胞;编辑所述干细胞的SERPINA1 基因;使所述干细胞分化成肝细胞;和将所述肝细胞移植进所述患者中。

[0031] 通过执行患者骨髓的活组织检查,可以从患者骨髓分离间充质干细胞,或可以从外周血分离间充质干细胞。

[0032] 在某些实施方案中,所述分离间充质干细胞的步骤包括:抽吸骨髓并通过使用Percoll™密度离心分离间充质细胞。

[0033] 在某些实施方案中,在干细胞的SERPINA1 基因或编码所述间充质干细胞的SERPINA1 基因的调节元件的其它DNA序列内或附近编辑的步骤包括,向所述间充质干细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现一个或多个单链断裂(SSB)或双链断裂(DSB),所述一个或多个单链断裂(SSB)或双链断裂(DSB):在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近,其导致一个或多个突变或外显子的表达或功能的永久缺失、插入、校正或调节,所述一个或多个突变或外显子在SERPINA1基因内或附近、或影响SERPINA1基因的表达或功能;或在安全港基因座内或附近,其导致在SERPINA1基因内或附近的一个或多个突变,其导致SERPINA1基因或小基因的永久插入,和α1抗胰蛋白酶(AAT)蛋白活性的恢复。所述安全港基因座可以选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。所述安全港基因座可以选自AAVS1(PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2和TTR的外显子1-2。

[0034] 在某些实施方案中,使所述编辑过基因组的间充质干细胞分化成肝细胞的步骤包括以下的一项或多项以使编辑过基因组的间充质干细胞分化成肝细胞:使编辑过基因组的间充质干细胞与胰岛素、转铁蛋白、FGF4、HGF或胆汁酸中的一种或多种接触。

[0035] 在某些实施方案中,将所述肝细胞移植进所述患者中的步骤包括,通过移植、局部注射或全身输注或它们的组合将所述肝细胞移植进所述患者中。

[0036] 在另一个方面,本文提供了治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的体内方法,所述方法包括下述步骤:编辑所述患者的细胞中的SERPINA1基因、或编码SERPINA1基因的调节元件的其它DNA序列,或在所述患者的细胞中的安全港基因座内或附近编辑。

[0037] 在某些实施方案中,编辑所述患者的细胞中的SERPINA1基因的步骤包括,在所述细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个单链断裂(SSB)或双链断裂(DSB),其导致在SERPINA1基因内一个或多个突变或外显子的永久缺失、插入、或校正或调节,其导致SERPINA1基因或小基因的永久插入,和AAT蛋白活性的恢复。

[0038] 在某些实施方案中,编辑所述患者的细胞中的SERPINA1基因的步骤包括,在所述

细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现在SERPINA1基因或编码SERPINA1基因的调节元件的DNA序列内或附近的一个或多个双链断裂(DSB),或在SERPINA1基因的安全港基因座内或附近编辑,其导致一个或多个外显子的表达或功能的永久缺失、插入、校正、或校正或调节,所述一个或多个外显子在SERPINA1基因内或附近、或影响SERPINA1基因的表达或功能,和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0039] 在某些实施方案中,所述一种或多种DNA内切核酸酶是Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (也被称作Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4或Cpf1内切核酸酶;其同系物,天然存在的分子的重组,其经过密码子优化的或经修饰的形式,和前述任一种的组合。

[0040] 在某些实施方案中,所述方法包括在所述细胞中引入编码一种或多种DNA内切核酸酶的一种或多种多核苷酸。在某些实施方案中,所述方法包括在所述细胞中引入编码一种或多种DNA内切核酸酶的一种或多种核糖核酸(RNA)。在某些实施方案中,所述一种或多种多核苷酸或一种或多种RNA是一种或多种经修饰的多核苷酸或一种或多种经修饰的RNA。在某些实施方案中,所述方法包括在所述细胞中引入一种或多种DNA内切核酸酶,其中所述内切核酸酶是蛋白或多肽。

[0041] 在某些实施方案中,所述方法还包括在所述细胞中引入一种或多种引导核糖核酸(gRNA)。在某些实施方案中,所述一种或多种gRNA是单分子引导RNA (sgRNA)。在某些实施方案中,所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA、一种或多种经修饰的sgRNA或它们的组合。

[0042] 在某些实施方案中,所述一种或多种DNA内切核酸酶与一种或多种gRNA、或一种或多种sgRNA或它们的组合预先形成复合物(pre-complexed)。

[0043] 在某些实施方案中,所述方法还包括在所述细胞中引入多核苷酸供体模板,其包含野生型SERPINA1基因的一部分、或小基因(包含天然的或合成的增强子和启动子、一个或多个外显子、和天然的或合成的内含子、和天然的或合成的3’UTR和多腺苷酸化信号)、编码SERPINA1基因的野生型调节元件的DNA序列和/或cDNA。在某些实施方案中,所述野生型SERPINA1基因的一部分或cDNA可以是外显子1、外显子2、外显子3、外显子4、外显子5、内含子区域、它们的片段或组合、或整个SERPINA1基因或cDNA。在某些实施方案中,所述供体模板是单链或双链多核苷酸。在某些实施方案中,所述供体模板具有与14q32.13区域同源的臂。

[0044] 在某些实施方案中,所述方法还包括在所述细胞中引入一种引导核糖核酸(gRNA)和包含野生型SERPINA1基因的至少一部分的多核苷酸供体模板。在某些实施方案中,所述方法还包括在所述细胞中引入一种引导核糖核酸(gRNA)和多核苷酸供体模板,所述多核苷酸供体模板包含经密码子优化的或经修饰的SERPINA1基因的至少一部分。所述一种或多种DNA内切核酸酶是一种或多种Cas9内切核酸酶,其实现在SERPINA1基因(或经密码子优化的或经修饰的SERPINA1基因)或编码SERPINA1基因的调节元件的其它DNA序列内或附近或在安全港基因座内或附近的DSB基因座处的一个双链断裂(DSB),所述断裂促进来自所述多核苷酸供体模板的新序列插入在所述基因座或安全港基因座处的染色体DNA中,所述插入导

致SERPINA1基因的染色体DNA或在所述基因座或安全港基因座近侧的编码SERPINA1基因的调节元件的其它DNA序列的一部分的永久插入或校正和AAT蛋白活性的恢复。在某些实施方案中，所述gRNA包含与所述基因座或安全港基因座的区段互补的间隔物序列。在某些实施方案中，近侧是指在所述基因座或安全港基因座的上游和下游的核苷酸。

[0045] 在某些实施方案中，所述方法还包括在所述细胞中引入两种引导核糖核酸(gRNA)和包含野生型SERPINA1基因的至少一部分的多核苷酸供体模板，且其中所述一种或多种DNA内切核酸酶是两种或更多种Cas9内切核酸酶，其实现在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在安全港基因座内或附近的一对双链断裂(DSB)，第一个是在5’DSB基因座且第二个是在3’DSB基因座，所述断裂促进来自所述多核苷酸供体模板的新序列在所述5’DSB基因座和所述3’DSB基因座之间插入所述染色体DNA中，所述插入导致在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在安全港基因座内或附近在所述5’DSB基因座和所述3’DSB基因座之间的染色体DNA的永久插入或校正，和AAT蛋白活性的恢复，且其中所述第一个引导RNA包含与所述5’ DSB基因座的区段互补的间隔物序列，且所述第二个引导RNA包含与所述3’ DSB基因座的区段互补的间隔物序列。

[0046] 在某些实施方案中，所述一种或多种gRNA是一种或多种单分子引导RNA (sgRNA)。在某些实施方案中，所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA或一种或多种经修饰的sgRNA。

[0047] 在某些实施方案中，所述一种或多种DNA内切核酸酶与一种或多种gRNA或一种或多种sgRNA预先形成复合物。

[0048] 在某些实施方案中，所述野生型SERPINA1基因或cDNA的部分是整个SERPINA1基因或cDNA。

[0049] 在某些实施方案中，所述供体模板是单链或双链的。在某些实施方案中，所述供体模板具有与14q32.13区域同源的臂。

[0050] 在某些实施方案中，所述DSB、或5’DSB和3’DSB是在所述SERPINA1基因的第一个外显子内、第一个内含子内、或第一个外显子和第一个内含子二者内。

[0051] 在某些实施方案中，所述gRNA或sgRNA指向下述病理学变体中的一种或多种：rs764325655、rs121912713、rs28929474、rs17580、rs121912714、rs764220898、rs199422211、rs751235320、rs199422210、rs267606950、rs55819880、rs28931570。

[0052] 在某些实施方案中，所述插入或校正是通过同源性指导的修复(HDR)。

[0053] 在某些实施方案中，所述方法还包括在所述细胞中引入两种引导核糖核酸(gRNA)，且其中所述一种或多种DNA内切核酸酶是两种或更多种Cas9内切核酸酶，其实现在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近或在安全港基因座内或附近的一对双链断裂(DSB)，第一个是在5’DSB基因座且第二个是在3’DSB基因座，所述断裂引起在所述5’DSB基因座和所述3’DSB基因座之间的染色体DNA的缺失，其导致在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在安全港基因座内或附近在所述5’DSB基因座和所述3’DSB基因座之间的染色体DNA的永久缺失，和AAT蛋白活性的恢复，且其中所述第一个引导RNA包含与所述5’ DSB基因座的区段互补的间隔物序列，且所述第二个引导RNA包含与所述3’ DSB基因座的区段互补的间隔物序列。

[0054] 在某些实施方案中，所述两种gRNA是两种单分子引导RNA(sgRNA)。在某些实施方案中，所述两种gRNA或两种sgRNA是两种经修饰的gRNA或两种经修饰的sgRNA。

[0055] 在某些实施方案中，所述一种或多种DNA内切核酸酶与一种或两种gRNA或一种或两种sgRNA预先形成复合物。

[0056] 在某些实施方案中，所述5' DSB和3' DSB都在所述SERPINA1基因的第一个外显子、第二个外显子、第三个外显子、第四个外显子或第五个外显子或内含子内或附近。

[0057] 在某些实施方案中，所述校正是通过同源性指导的修复(HDR)。

[0058] 在某些实施方案中，所述缺失是1kb或更少的缺失。

[0059] 在某些实施方案中，将所述Cas9或Cpf1 mRNA、gRNA和供体模板各自分别配制进脂质纳米颗粒中或都共配制进脂质纳米颗粒中。

[0060] 在某些实施方案中，将所述Cas9或Cpf1 mRNA、gRNA和供体模板配制进单独的外来体中或共配制进外来体中。

[0061] 在某些实施方案中，将所述Cas9或Cpf1 mRNA配制进脂质纳米颗粒中，并将所述gRNA和供体模板都通过病毒载体递送给所述细胞。在某些实施方案中，所述病毒载体是腺伴随病毒(AAV)载体。在某些实施方案中，所述AAV载体是AAV6载体。

[0062] 可以将所述Cas9或Cpf1 mRNA配制进脂质纳米颗粒中，且可以将所述gRNA通过电穿孔递送给所述细胞，且可以将供体模板通过病毒载体递送给所述细胞。在某些实施方案中，所述病毒载体是腺伴随病毒(AAV)载体。在某些实施方案中，所述AAV载体是AAV6载体。

[0063] 在某些实施方案中，所述SERPINA1基因位于染色体14:1,493,319 – 1,507,264上(Genome Reference Consortium – GRCh38/hg38)。

[0064] 可以将AAT蛋白活性的恢复与野生型或正常AAT蛋白活性进行对比。

[0065] 在另一个方面，本文提供了一种或多种引导核糖核酸(gRNA)，其包含选自实施例1-6中和SEQ ID No 54,860-68,297中的用于编辑来自具有α1抗胰蛋白酶缺乏的患者的细胞中的SERPINA1基因的核酸序列的间隔物序列。在某些实施方案中，所述一种或多种gRNA是一种或多种单分子引导RNA(sgRNA)。在某些实施方案中，所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA或一种或多种经修饰的sgRNA。

[0066] 在另一个方面，本文提供了一种或多种引导核糖核酸(gRNA)，其包含选自SEQ ID No:1-54,859中的用于编辑来自具有α1抗胰蛋白酶缺乏的患者的细胞中的安全港基因座的核酸序列的间隔物序列，其中所述安全港基因座选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。在某些实施方案中，所述一种或多种gRNA是一种或多种单分子引导RNA(sgRNA)。在某些实施方案中，所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA或一种或多种经修饰的sgRNA。

[0067] 在另一个方面，本文提供了细胞，其已经通过上述方法修饰成永久地校正SERPINA1基因内的一个或多个突变和恢复AAT蛋白活性。本文还提供了通过给AATD患者施用已经通过上述方法修饰的细胞来改善α1抗胰蛋白酶缺乏(AATD)的方法。

[0068] 在某些实施方案中，本公开的方法和组合物包含一种或多种经修饰的引导核糖核酸(gRNA)。修饰的非限制性例子可以包含一个或多个在糖的2'位置处被修饰的核苷酸，在某些实施方案中，2'-O-烷基、2'-O-烷基-O-烷基或2'-氟-修饰的核苷酸。在某些实施方案

中,RNA修饰包括在嘧啶、脱碱基残基、脱氧核苷酸的核糖上的2'-氟、2'-氨基或2' O-甲基修饰,或在RNA的3'末端处的倒置碱基。

[0069] 在某些实施方案中,所述一种或多种经修饰的引导核糖核酸(gRNA)包含使所述经修饰的gRNA比天然寡核苷酸更耐受核酸酶消化的修饰。这样的修饰的非限制性例子包括包含经修饰的主链的那些,例如,硫代磷酸酯、phosphorothyos、磷酸三酯、膦酸甲酯、短链烷基或环烷基糖间键或短链杂原子或杂环糖间键。

[0070] 在本文中描述和举例说明了各种其它方面。

[0071] 附图简述

[0072] 通过参考附图可以更好地理解在本说明书中公开和描述的用于治疗α1抗胰蛋白酶缺乏(AATD)的材料和方法的各种方面,在附图中:

[0073] 图1A是包含对于化脓性链球菌(S. pyogenes)Cas9内切核酸酶的经密码子优化的基因的质粒(在本文中被称作"CTx-1")的图解。所述CTx-1质粒还包含gRNA支架序列,其包括来自在实施例1和8中列出的序列的20碱基对间隔物序列。

[0074] 图1B是包含对于化脓性链球菌Cas9内切核酸酶的不同的经密码子优化的基因的质粒(在本文中被称作"CTx-2")的图解。所述CTx-2质粒还包含gRNA支架序列,其包括来自在实施例1和8中列出的序列的20碱基对间隔物序列。

[0075] 图1C是包含对于化脓性链球菌Cas9内切核酸酶的另一种不同的经密码子优化的基因的质粒(在本文中被称作"CTx-3")的图解。所述CTx-3质粒还包含gRNA支架序列,其包括来自在实施例1和8中列出的序列的20碱基对间隔物序列。

[0076] 图2A是描绘II型CRISPR/Cas系统的图解。

[0077] 图2B是描绘II型CRISPR/Cas系统的另一个图解。

[0078] 图3是描绘在体外转录且应用messenger Max转染到组成性表达Cas9的HEK293T细胞中的gRNA的切割效率的图,如应用TIDE分析评价的。

[0079] 序列表简述

[0080] SEQ ID NO:1-2,032是关于用化脓性链球菌Cas9内切核酸酶靶向AAVS1 (PPP1R12C)基因的20碱基对间隔物序列。

[0081] SEQ ID NO:2,033-2,203是关于用金黄色葡萄球菌(*S. aureus*)Cas9内切核酸酶靶向AAVS1 (PPP1R12C)基因的20碱基对间隔物序列。

[0082] SEQ ID NO:2,204-2,221是关于用嗜热链球菌(*S. thermophilus*)Cas9内切核酸酶靶向AAVS1 (PPP1R12C)基因的20碱基对间隔物序列。

[0083] SEQ ID NO:2,222-2,230是关于用齿垢密螺旋体(*T. denticola*)Cas9内切核酸酶靶向AAVS1 (PPP1R12C)基因的20碱基对间隔物序列。

[0084] SEQ ID NO:2,231-2,305是关于用脑膜炎奈瑟氏菌(*N. meningitides*)Cas9内切核酸酶靶向AAVS1 (PPP1R12C)基因的20碱基对间隔物序列。

[0085] SEQ ID NO:2,306-3,481是关于用*Acidominococcus*、毛螺菌科(*Lachnospiraceae*)和新凶手弗朗西丝氏菌(*Francisella novicida*)Cpf1内切核酸酶靶向AAVS1 (PPP1R12C)基因的22碱基对间隔物序列。

[0086] SEQ ID NO:3,482-3,649是关于用化脓性链球菌Cas9内切核酸酶靶向Alb基因的20碱基对间隔物序列。

[0087] SEQ ID NO:3,650-3,677是关于用金黄色葡萄球菌Cas9内切核酸酶靶向Alb基因的20碱基对间隔物序列。

[0088] SEQ ID NO:3,678-3,695是关于用嗜热链球菌Cas9内切核酸酶靶向Alb基因的20碱基对间隔物序列。

[0089] SEQ ID NO:3,696-3,700是关于用齿垢密螺旋体Cas9内切核酸酶靶向Alb基因的20碱基对间隔物序列。

[0090] SEQ ID NO:3,701-3,724是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向Alb基因的20碱基对间隔物序列。

[0091] SEQ ID NO:3,725-4,103是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向Alb基因的22碱基对间隔物序列。

[0092] SEQ ID NO:4,104-4,448是关于用化脓性链球菌Cas9内切核酸酶靶向Angptl3基因的20碱基对间隔物序列。

[0093] SEQ ID NO:4,449-4,484是关于用金黄色葡萄球菌Cas9内切核酸酶靶向Angptl3基因的20碱基对间隔物序列。

[0094] SEQ ID NO:4,485-4,507是关于用嗜热链球菌Cas9内切核酸酶靶向Angptl3基因的20碱基对间隔物序列。

[0095] SEQ ID NO:4,508-4,520是关于用齿垢密螺旋体Cas9内切核酸酶靶向Angptl3基因的20碱基对间隔物序列。

[0096] SEQ ID NO:4,521-4,583是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向Angptl3基因的20碱基对间隔物序列。

[0097] SEQ ID NO:4,584-5,431是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向Angptl3基因的22碱基对间隔物序列。

[0098] SEQ ID NO:5,432-5,834是关于用化脓性链球菌Cas9内切核酸酶靶向ApoC3基因的20碱基对间隔物序列。

[0099] SEQ ID NO:5,835-5,859是关于用金黄色葡萄球菌Cas9内切核酸酶靶向ApoC3基因的20碱基对间隔物序列。

[0100] SEQ ID NO:5,860-5,862是关于用嗜热链球菌Cas9内切核酸酶靶向ApoC3基因的20碱基对间隔物序列。

[0101] SEQ ID NO:5,863-5,864是关于用齿垢密螺旋体Cas9内切核酸酶靶向ApoC3基因的20碱基对间隔物序列。

[0102] SEQ ID NO:5,865-5,876是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向ApoC3基因的20碱基对间隔物序列。

[0103] SEQ ID NO:5,877-6,108是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向ApoC3基因的22碱基对间隔物序列。

[0104] SEQ ID NO:6,109-7,876是关于用化脓性链球菌Cas9内切核酸酶靶向ASGR2基因的20碱基对间隔物序列。

[0105] SEQ ID NO:7,877-8,082是关于用金黄色葡萄球菌Cas9内切核酸酶靶向ASGR2基因的20碱基对间隔物序列。

[0106] SEQ ID NO:8,083-8,106是关于用嗜热链球菌Cas9内切核酸酶靶向ASGR2基因的

20碱基对间隔物序列。

[0107] SEQ ID NO:8,107-8,118是关于用齿垢密螺旋体Cas9内切核酸酶靶向ASGR2基因的20碱基对间隔物序列。

[0108] SEQ ID NO:8,119-8,201是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向ASGR2基因的20碱基对间隔物序列。

[0109] SEQ ID NO:8,202-9,641是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向ASGR2基因的22碱基对间隔物序列。

[0110] SEQ ID NO:9,642-9,844是关于用化脓性链球菌Cas9内切核酸酶靶向CCR5基因的20碱基对间隔物序列。

[0111] SEQ ID NO:9,845-9,876是关于用金黄色葡萄球菌Cas9内切核酸酶靶向CCR5基因的20碱基对间隔物序列。

[0112] SEQ ID NO:9,877-9,890是关于用嗜热链球菌Cas9内切核酸酶靶向CCR5基因的20碱基对间隔物序列。

[0113] SEQ ID NO:9,891-9,892是关于用齿垢密螺旋体Cas9内切核酸酶靶向CCR5基因的20碱基对间隔物序列。

[0114] SEQ ID NO:9,893-9,920是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向CCR5基因的20碱基对间隔物序列。

[0115] SEQ ID NO:9,921-10,220是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向CCR5基因的22碱基对间隔物序列。

[0116] SEQ ID NO:10,221-11,686是关于用化脓性链球菌Cas9内切核酸酶靶向F9基因的20碱基对间隔物序列。

[0117] SEQ ID NO:11,687-11,849是关于用金黄色葡萄球菌Cas9内切核酸酶靶向F9基因的20碱基对间隔物序列。

[0118] SEQ ID NO:11,850-11,910是关于用嗜热链球菌Cas9内切核酸酶靶向F9基因的20碱基对间隔物序列。

[0119] SEQ ID NO:11,911-11,935是关于用齿垢密螺旋体Cas9内切核酸酶靶向F9基因的20碱基对间隔物序列。

[0120] SEQ ID NO:11,936-12,088是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向F9基因的20碱基对间隔物序列。

[0121] SEQ ID NO:12,089-14,229是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向F9基因的22碱基对间隔物序列。

[0122] SEQ ID NO:14,230-15,245是关于用化脓性链球菌Cas9内切核酸酶靶向G6PC基因的20碱基对间隔物序列。

[0123] SEQ ID NO:15,246-15,362是关于用金黄色葡萄球菌Cas9内切核酸酶靶向G6PC基因的20碱基对间隔物序列。

[0124] SEQ ID NO:15,363-15,386是关于用嗜热链球菌Cas9内切核酸酶靶向G6PC基因的20碱基对间隔物序列。

[0125] SEQ ID NO:15,387-15,395是关于用齿垢密螺旋体Cas9内切核酸酶靶向G6PC基因的20碱基对间隔物序列。

[0126] SEQ ID NO:15,396-15,485是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向G6PC基因的20碱基对间隔物序列。

[0127] SEQ ID NO:15,486-16,580是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向G6PC基因的22碱基对间隔物序列。

[0128] SEQ ID NO:16,581-22,073是关于用化脓性链球菌Cas9内切核酸酶靶向Gys2基因的20碱基对间隔物序列。

[0129] SEQ ID NO:22,074-22,749是关于用金黄色葡萄球菌Cas9内切核酸酶靶向Gys2基因的20碱基对间隔物序列。

[0130] SEQ ID NO:22,750-20,327是关于用嗜热链球菌Cas9内切核酸酶靶向Gys2基因的20碱基对间隔物序列。

[0131] SEQ ID NO:23,028-23,141是关于用齿垢密螺旋体Cas9内切核酸酶靶向Gys2基因的20碱基对间隔物序列。

[0132] SEQ ID NO:23,142-23,821是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向Gys2基因的20碱基对间隔物序列。

[0133] SEQ ID NO:23,822-32,253是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向Gys2基因的22碱基对间隔物序列。

[0134] SEQ ID NO:32,254-33,946是关于用化脓性链球菌Cas9内切核酸酶靶向HGD基因的20碱基对间隔物序列。

[0135] SEQ ID NO:33,947-34,160是关于用金黄色葡萄球菌Cas9内切核酸酶靶向HGD基因的20碱基对间隔物序列。

[0136] SEQ ID NO:34,161-34,243是关于用嗜热链球菌Cas9内切核酸酶靶向HGD基因的20碱基对间隔物序列。

[0137] SEQ ID NO:34,244-34,262是关于用齿垢密螺旋体Cas9内切核酸酶靶向HGD基因的20碱基对间隔物序列。

[0138] SEQ ID NO:34,263-34,463是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向HGD基因的20碱基对间隔物序列。

[0139] SEQ ID NO:34,464-36,788是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向HGD基因的22碱基对间隔物序列。

[0140] SEQ ID NO:36,789-40,583是关于用化脓性链球菌Cas9内切核酸酶靶向Lp(a)基因的20碱基对间隔物序列。

[0141] SEQ ID NO:40,584-40,993是关于用金黄色葡萄球菌Cas9内切核酸酶靶向Lp(a)基因的20碱基对间隔物序列。

[0142] SEQ ID NO:40,994-41,129是关于用嗜热链球菌Cas9内切核酸酶靶向Lp(a)基因的20碱基对间隔物序列。

[0143] SEQ ID NO:41,130-41,164是关于用齿垢密螺旋体Cas9内切核酸酶靶向Lp(a)基因的20碱基对间隔物序列。

[0144] SEQ ID NO:41,165-41,532是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向Lp(a)基因的20碱基对间隔物序列。

[0145] SEQ ID NO:41,533-46,153是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西

丝氏菌Cpf1内切核酸酶靶向Lp(a)基因的22碱基对间隔物序列。

[0146] SEQ ID NO:46,154-48,173是关于用化脓性链球菌Cas9内切核酸酶靶向PCSK9基因的20碱基对间隔物序列。

[0147] SEQ ID NO:48,174-48,360是关于用金黄色葡萄球菌Cas9内切核酸酶靶向PCSK9基因的20碱基对间隔物序列。

[0148] SEQ ID NO:48,361-48,396是关于用嗜热链球菌Cas9内切核酸酶靶向PCSK9基因的20碱基对间隔物序列。

[0149] SEQ ID NO:48,397-48,410是关于用齿垢密螺旋体Cas9内切核酸酶靶向PCSK9基因的20碱基对间隔物序列。

[0150] SEQ ID NO:48,411-48,550是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向PCSK9基因的20碱基对间隔物序列。

[0151] SEQ ID NO:48,551-50,344 是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向PCSK9基因的22碱基对间隔物序列。

[0152] SEQ ID NO:50,345-51,482是关于用化脓性链球菌Cas9内切核酸酶靶向作为安全港基因座的Serpina1基因的20碱基对间隔物序列。

[0153] SEQ ID NO:51,483-51,575是关于用金黄色葡萄球菌Cas9内切核酸酶靶向作为安全港基因座的Serpina1基因的20碱基对间隔物序列。

[0154] SEQ ID NO:51,576-51,587是关于用嗜热链球菌Cas9内切核酸酶靶向作为安全港基因座的Serpina1基因的20碱基对间隔物序列。

[0155] SEQ ID NO:51,588-51,590是关于用齿垢密螺旋体Cas9内切核酸酶靶向作为安全港基因座的Serpina1基因的20碱基对间隔物序列。

[0156] SEQ ID NO:51,591-51,641是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向作为安全港基因座的Serpina1基因的20碱基对间隔物序列。

[0157] SEQ ID NO:51,642-52,445是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向作为安全港基因座的Serpina1基因的22碱基对间隔物序列。

[0158] SEQ ID NO:52,446-53,277是关于用化脓性链球菌Cas9内切核酸酶靶向TF基因的20碱基对间隔物序列。

[0159] SEQ ID NO:53,278-53,363是关于用金黄色葡萄球菌Cas9内切核酸酶靶向TF基因的20碱基对间隔物序列。

[0160] SEQ ID NO:53,364-53,375是关于用嗜热链球菌Cas9内切核酸酶靶向TF基因的20碱基对间隔物序列。

[0161] SEQ ID NO:53,376-53,382是关于用齿垢密螺旋体Cas9内切核酸酶靶向TF基因的20碱基对间隔物序列。

[0162] SEQ ID NO:53,383-53,426是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向TF基因的20碱基对间隔物序列。

[0163] SEQ ID NO:53,427-54,062是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向TF基因的22碱基对间隔物序列。

[0164] SEQ ID NO:54,063-54,362是关于用化脓性链球菌Cas9内切核酸酶靶向TTR 基因的20碱基对间隔物序列。

[0165] SEQ ID NO:54,363-54,403是关于用金黄色葡萄球菌Cas9内切核酸酶靶向TTR 基因的20碱基对间隔物序列。

[0166] SEQ ID NO:54,404-54,420是关于用嗜热链球菌Cas9内切核酸酶靶向TTR 基因的20碱基对间隔物序列。

[0167] SEQ ID NO:54,421-54,422是关于用齿垢密螺旋体Cas9内切核酸酶靶向TTR基因的20碱基对间隔物序列。

[0168] SEQ ID NO:54,423-54,457是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向TTR基因的20碱基对间隔物序列。

[0169] SEQ ID NO:54,458-54,859是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向TTR基因的22碱基对间隔物序列。

[0170] SEQ ID NO:54,860-61,324是关于用化脓性链球菌Cas9内切核酸酶靶向SERPINA1基因(不同于作为安全港基因座)的20碱基对间隔物序列。

[0171] SEQ ID NO:61,325-61,936是关于用金黄色葡萄球菌Cas9内切核酸酶靶向SERPINA1基因(不同于作为安全港基因座)的20碱基对间隔物序列。

[0172] SEQ ID NO:61,937-62,069是关于用嗜热链球菌Cas9内切核酸酶靶向SERPINA1基因(不同于作为安全港基因座)的20碱基对间隔物序列。

[0173] SEQ ID NO:62,070-62,120是关于用齿垢密螺旋体Cas9内切核酸酶靶向SERPINA1基因(不同于作为安全港基因座)的20碱基对间隔物序列。

[0174] SEQ ID NO:62,121-62,563是关于用脑膜炎奈瑟氏菌Cas9内切核酸酶靶向SERPINA1基因(不同于作为安全港基因座)的20碱基对间隔物序列。

[0175] SEQ ID NO:62,564-68,297是关于用*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌Cpf1内切核酸酶靶向SERPINA1基因(不同于作为安全港基因座)的22碱基对间隔物序列。

[0176] 发明详述

[0177] α1抗胰蛋白酶缺乏(AATD)

[0178] AATD由对SERPINA1基因的突变、或更少见地由对SERPINA1基因的缺失引起。SERPINA1基因位于14q32.13,具有在chr14:1,493,319-1,507,264的基因组坐标(GRCh38)。SERPINA1包括5个外显子并具有12.2 kb的总长度。该基因具有2个启动子,一个启动子控制在巨噬细胞中的表达。已报道了多个转录物变体;但是,它们都编码相同的AAT蛋白。约95%的AATD由病理性等位基因PI Z或PI S或它们的联合引起。

[0179] SERPINA1基因编码α1抗胰蛋白酶(AAT),α1抗胰蛋白酶也称为蛋白酶抑制剂(PI),是一种主要的血浆丝氨酸蛋白酶抑制剂。AAT主要与弹性蛋白酶形成复合物,但也与其它丝氨酸蛋白酶形成复合物。AAT的一个重要的抑制作用是针对嗜中性粒细胞弹性蛋白酶,嗜中性粒细胞弹性蛋白酶是降解肺泡壁的弹性蛋白以及各种组织中的其它结构蛋白的蛋白酶。

[0180] 治疗途径

[0181] 由于AATD的已知形式是具有隐性遗传的单基因病症,校正每个细胞的突变体等位基因之一可能足以校正和恢复或部分恢复AATD功能。一个等位基因的校正可以与一个保留原始突变的拷贝同时,或与被切割且通过非同源末端连接(NHEJ)修复且因此没有适当地校正的拷贝同时。双等位基因校正也可以发生。在下文讨论了可以对特定突变采用的各种编

辑策略。

[0182] 突变体等位基因中的一个或可能两个的校正提供超过现有或潜在疗法诸如通过慢病毒递送和整合引入SERPINA1表达盒的重要改善。基因编辑具有精确基因组修饰和较低不利作用的优点，例如，通过由NHEJ修复途径引起的插入或缺失可以校正突变。如果患者的SERPINA1基因具有插入的或缺失的碱基，靶向切割可以导致恢复框架的NHEJ介导的插入或缺失。使用一种或多种引导RNA，通过NHEJ介导的校正，也可以校正错义突变。基于局部序列和微同源性，可以设计或评价切割(一个或多个)以校正突变的能力或可能性。NHEJ还可以用于缺失基因的区段，无论是直接地还是经由靶向几个位置的一种gRNA、或几种gRNA的切割通过改变剪接供体或受体位点。如果氨基酸、结构域或外显子含有突变且可以被除去或倒置，或如果所述缺失以其它方式恢复对于所述蛋白的功能，这可为有用的。引导链对已经被用于缺失和校正倒位。

[0183] 可选地，用于通过HDR校正的供体含有经校正的序列，其具有小或大侧接同源性臂以允许退火。HDR基本上是在DSB修复过程中使用供给的同源DNA序列作为模板的无误差机制。HDR的比率是突变和切割位点之间的距离的函数，所以选择重叠或附近靶位是重要的。模板可以包括被同源区域侧接的额外序列，或可以含有不同于基因组序列的序列，从而允许序列编辑。

[0184] 除了通过NHEJ或HDR校正突变以外，一系列其它选择是可能的。如果存在小或大缺失或多个突变，可以敲入含有受影响的外显子的cDNA。可以将全长cDNA敲入具有或没有合适调节序列的任何“安全港”——即，不是SERPINA1基因本身的非有害插入点。如果将该构建体敲入在SERPINA1调节元件附近，它应当具有与正常基因类似的生理学控制。两种或更多种(例如，一对)核酸酶可以用于缺失突变的基因区域，尽管将经常必须提供供体来恢复功能。在该情况下，将供给两种gRNA和一种供体序列。

[0185] 本文提供了通过用Cas9和sgRNA诱导双链断裂或者使用两种适当sgRNA在突变周围诱导一对双链断裂来校正基因中的特定突变的方法，和提供供体DNA模板以诱导同源性指导的修复(HDR)的方法。在某些实施方案中，所述供体DNA模板可以是短的单链寡核苷酸、短的双链寡核苷酸、长单链或双链DNA分子。这些方法对于SERPINA1的每种变体使用gRNA和供体DNA分子。

[0186] 本文提供了将SERPINA1 cDNA或小基因(包含一个或多个外显子和内含子或天然的或合成的内含子)敲入对应基因的基因座中的方法。这些方法使用一对sgRNA，其靶向SERPINA1基因的第一个外显子和/或第一个内含子。在某些实施方案中，所述供体DNA是具有与17q21区域同源的臂的单链或双链DNA。

[0187] 本文提供了将SERPINA1 cDNA或小基因(其包含一个或多个外显子和内含子或天然的或合成的内含子)敲入热点例如ALB基因的基因座中的方法。这些方法使用一对sgRNA，其靶向位于肝热点中的基因的第一个外显子和/或第一个内含子。在某些实施方案中，所述供体DNA是具有与对应区域同源的臂的单链或双链DNA。

[0188] 本文提供了如下使用基因组工程工具建立对基因组的永久变化的细胞、离体和体内方法：1)通过由不精确的NHEJ途径引起的插入或缺失，校正在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个突变，2)通过HDR，校正在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个突

变,或3)缺失突变体区域和/或将SERPINA1 cDNA或小基因(其包含天然的或合成的增强子和启动子、一个或多个外显子、和天然的或合成的内含子、和天然的或合成的3'UTR和多腺苷酸化信号)敲入SERPINA1基因的基因座或安全港基因座,并恢复AAT蛋白活性。这样的方法使用内切核酸酶,诸如CRISPR-相关的(CRISPR/Cas9、Cpf1等)核酸酶,以在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的基因组基因座中永久地缺失、插入、编辑、校正或替换一个或多个外显子或其部分(即,在编码序列和/或剪接序列内或附近的突变),或插入SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的基因组基因座中。以此方式,在本公开中阐述的实例可以帮助用SERPINA1基因的单一处理(而不是在患者寿命中递送潜在疗法)恢复所述基因的读码框或野生型序列,或以其它方式校正所述基因。

[0189] 本文提供了治疗具有AATD的患者的方法。这样的方法的一个实施方案是基于离体细胞的疗法。例如,建立患者特异性的诱导性多能干细胞(iPSC)。然后,使用本文描述的材料和方法,编辑这些iPS细胞的染色体DNA。接着,使编辑过基因组的iPSC分化成肝细胞。最后,将所述肝细胞移植进所述患者中。

[0190] 这样的方法的另一个实施方案是基于离体细胞的疗法。例如,执行患者的肝的活组织检查。然后,从所述患者分离肝特异性的祖细胞或原代肝细胞,例如,通过活组织检查。接着,使用本文描述的材料和方法编辑这些祖细胞或原代肝细胞的染色体DNA。最后,将所述编辑过基因组的祖细胞或原代肝细胞移植进所述患者中。

[0191] 这样的方法的另一个实施方案是基于离体细胞的疗法。例如,执行患者骨髓的活组织检查。然后,从所述患者分离间充质干细胞,其可以分离自患者骨髓,例如,通过活组织检查,或分离自外周血。接着,使用本文描述的材料和方法编辑这些间充质干细胞的染色体DNA。接着,使编辑过基因组的间充质干细胞分化成肝细胞。最后,将这些肝细胞移植进所述患者中。

[0192] 离体细胞疗法途径的一个优点是在施用之前进行治疗的综合分析的能力。所有基于核酸酶的治疗剂具有某种水平的靶外(off-target)效应。离体执行基因校正允许人们在植入之前充分地表征经校正的细胞群体。本发明的方面包括对经校正的细胞的整个基因组测序以确保靶外切口(如果有的话)是在与对患者的最小风险有关的基因组位置。此外,可以在植入之前分离特定细胞的群体,包括克隆群体。

[0193] 离体细胞疗法的另一个优点涉及与其它原代细胞来源相比在iPSC中的基因校正。iPSC是多产的,使得容易得到大量细胞,所述细胞将是基于细胞的疗法所需要的。此外,iPSC是用于执行克隆分离的理想细胞类型。这允许筛选正确的基因组校正,没有冒生存力降低的风险。与之相比,其它潜在细胞类型诸如原代肝细胞仅存活几代且难以克隆繁殖。因而,AATD iPSC的操作将廉价得多,并且将缩短产生期望的基因校正所需的时间量。

[0194] 这样的方法的另一个实施方案是基于体内的疗法。在该方法中,使用本文描述的材料和方法校正患者中的细胞的染色体DNA。

[0195] 体内基因疗法的一个优点是治疗性生产和施用的容易性。相同的治疗途径和疗法将具有用于治疗超过一位患者、例如许多具有相同或类似基因型或等位基因的患者的潜力。与之相比,离体细胞疗法通常要求使用患者自身的细胞,其被分离、操作和返回至同一患者。

[0196] 本文还提供了通过基因组编辑来编辑细胞中的SERPINA1基因的细胞方法。例如,

从患者或动物分离细胞。然后,使用本文描述的材料和方法编辑细胞的染色体DNA。

[0197] 不论细胞或离体或体内方法,本发明的方法涉及以下的一项或组合:1)通过由不精确的NHEJ途径引起的插入或缺失,校正在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个突变,2)通过HDR,校正在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个突变,或3)缺失突变体区域和/或将SERPINA1 cDNA或小基因(其包含一个或多个外显子或内含子或天然的或合成的内含子)敲入或将外源SERPINA1 DNA或cDNA序列或其片段引入基因的基因座中或在基因组中的异源位置(诸如安全港基因座,诸如,例如,靶向AAVS1 (PPP1R12C)、ALB基因、Angptl3基因、ApoC3基因、ASGR2基因、CCR5基因、FIX (F9)基因、G6PC基因、Gys2基因、HGD基因、Lp(a)基因、Pcsk9基因、Serpina1基因、TF基因和TTR基因)。HDR介导的cDNA向第一个外显子中的敲入的效率的评估可以利用cDNA向"安全港"位点中的敲入,诸如:具有与下述区域之一同源的臂的单链或双链DNA,例如:ApoC3 (chr11:116829908-116833071)、Angptl3 (chr1:62,597,487-62,606,305)、Serpina1 (chr14:94376747-94390692)、Lp(a) (chr6:160531483-160664259)、Pcsk9 (chr1:55,039,475-55,064,852)、FIX (chrX:139,530,736-139,563,458)、ALB (chr4:73,404,254-73,421,411)、TTR (chr18:31,591,766-31,599,023)、TF (chr3:133,661,997-133,779,005)、G6PC (chr17:42,900,796-42,914,432)、Gys2 (chr12:21,536,188-21,604,857)、AAVS1(PPP1R12C) (chr19:55,090,912-55,117,599)、HGD (chr3:120,628,167-120,682,570)、CCR5 (chr3:46,370,854-46,376,206)、ASGR2 (chr17:7,101,322-7,114,310)。校正和敲入策略都在同源性指导的修复(HDR)中利用供体DNA模板。通过使用一种或多种内切核酸酶在基因组中的特定位点处产生一个或多个双链断裂(DSB),可完成在任一种策略中的HDR。

[0198] 例如,NHEJ校正策略可以牵涉如下恢复SERPINA1基因中的读码框:用一种或多种CRISPR内切核酸酶和gRNA (例如,crRNA + tracrRNA,或sgRNA)在目标基因中诱导一个单链断裂或双链断裂,或用两种或更多种CRISPR内切核酸酶和两种或更多种sgRNA在目标基因中诱导两个或更多个单链断裂或双链断裂。该途径可以需要针对SERPINA1基因开发和优化sgRNA。

[0199] 例如,HDR校正策略牵涉如下恢复SERPINA1基因中的读码框:在外源地引入以指导对同源性指导的修复的细胞DSB应答的供体DNA模板(所述供体DNA模板可以是短的单链寡核苷酸、短的双链寡核苷酸、长的单链或双链DNA分子)存在的情况下,用一种或多种CRISPR内切核酸酶和gRNA (例如,crRNA + tracrRNA,或sgRNA)在目标基因中诱导一个双链断裂,或用一种或多种CRISPR内切核酸酶和两种或更多种适当的sgRNA在目标基因中诱导两个或更多个双链断裂。该途径要求针对SERPINA1基因的主要变体(PI Z)开发和优化gRNA和供体DNA分子。

[0200] 例如,敲入策略牵涉,使用gRNA (例如,crRNA + tracrRNA,或sgRNA)或一对sgRNA,将SERPINA1 cDNA或小基因(其包含天然的或合成的增强子和启动子、一个或多个外显子、和天然的或合成的内含子、和天然的或合成的3'UTR和多腺苷酸化信号)敲入所述基因的基因座中,所述sgRNA靶向SERPINA1基因的第一个或其它外显子和/或内含子的上游或内部,或安全港位点(例如,AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和/或TTR)中。所述供体DNA将是具有与

14q32.13区域同源的臂的单链或双链DNA。

[0201] 例如，缺失侧链牵涉，用一种或多种内切核酸酶和两种或更多种gRNA或sgRNA，在SERPINA1基因的5个外显子的一个或多个中缺失一个或多个突变。

[0202] 上述策略(校正和敲入和缺失)的优点是类似的，基本上包括短期和长期有益的临床和实验室效果。另外，可能仅需要低百分比的AAT活性以提供治疗益处。所有策略的另一个优点是，大多数患者具有低水平基因和蛋白活性，因此提示另外的蛋白表达(例如在基因校正以后)不一定导致针对靶基因产物的免疫应答。敲入途径确实提供超过校正或缺失途径的一个优点——相对于仅患者子集治疗所有患者的能力。尽管在该基因中存在共同突变，也存在许多其它可能的突变，并且使用敲入方法可以治疗它们全部。以此方式使用基因编辑的其它问题是对于HDR而言需要DNA供体。

[0203] 除了以上基因组编辑策略以外，另一种策略牵涉通过在调节序列中编辑来调节SERPINA1的表达、功能或活性。

[0204] 除了上面列出的编辑选项以外，Cas9或类似的蛋白可以用于将效应物结构域靶向至可以为编辑而鉴别出的相同靶位点，或在效应物结构域范围内的另外靶位点。一系列染色质修饰酶、甲基酶或脱甲基酶(demethlyase)可以用于改变靶基因的表达。如果突变导致更低的活性，一种可能性是增加AAT蛋白的表达。这些类型的表观遗传调节具有一些优点，特别是因为它们在可能的靶外效应中受到限制。

[0205] 除了在编码和剪接序列中的突变以外，存在许多类型的基因组靶位点。

[0206] 转录和翻译的调节涉及许多不同种类的与细胞蛋白或核苷酸相互作用的位点。经常可以靶向转录因子或其它蛋白的DNA结合位点进行突变或缺失以研究所述位点的作用，尽管也可以靶向它们来改变基因表达。通过非同源末端连接(NHEJ)或直接基因组编辑(通过同源性指导的修复(HDR))可以添加位点。基因组测序、RNA表达和转录因子结合的全基因组研究的增加的应用已经增加了鉴别所述位点如何导致发育或时间基因调节的能力。这些控制系统可为直接的，或可牵涉广泛协作调节，其可以要求来自多个增强子的活性的整合。转录因子通常结合6-12个碱基长的简并DNA序列。由各个位点提供的低特异性水平提示，在结合和功能结果中牵涉复杂的相互作用和规则。具有更低简并性的结合位点可能提供更简单的调节方式。可以设计人工转录因子来指定更长的序列，其具有在基因组中更低相似性的序列且具有关于靶外切割的更低潜力。可以突变、缺失或甚至建立这些类型的结合位点中的任一种以使得能够在基因调节或表达中变化(Canver, M.C. 等人,*Nature* (2015))。

[0207] 具有这些特征的基因调节区的另一种类型是微RNA (miRNA)结合位点。miRNA是在转录后基因调节中起关键作用的非编码RNA。miRNA可调节所有编码哺乳动物蛋白的基因中的30%的表达。发现了通过双链RNA (RNAi)的特异性的且有效的基因沉默，以及另外的小非编码RNA(Canver, M.C. 等人,*Nature* (2015))。对于基因沉默重要的非编码RNA的最大类型是miRNA。在哺乳动物中，miRNA首先被转录为长的RNA转录物，其可以是单独的转录单元、蛋白内含子的部分或其它转录物。长转录物被称作初级miRNA (初级-miRNA)，其包括不完美地碱基配对的发夹结构。这些初级-miRNA被微处理器切割成一种或多种更短的前体miRNA (前-miRNA)，所述微处理器是在细胞核中的蛋白复合物，牵涉Drosha。

[0208] 前-miRNA是具有2-核苷酸3’-突出端的约70个核苷酸长度的短茎环，其输出成为成熟的19-25核苷酸miRNA:miRNA*双链体。具有较低碱基配对稳定性的miRNA链(引导链)被

加载到RNA诱导的沉默复合物(RISC)上。过客引导链(用*标记)可为功能性的,但是经常被降解。成熟的miRNA将RISC连接至主要见于3′非翻译区(UTR)内的靶mRNA中的部分互补序列基序并诱导转录后基因沉默(Bartel, D.P. *Cell* 136, 215-233 (2009); Saj, A. & Lai, E.C. *Curr Opin Genet Dev* 21, 504-510 (2011))。

[0209] miRNA在哺乳动物和其它多细胞生物中的发育、分化、细胞周期和生长控制中以及在实际上所有的生物途径中是重要的。miRNA也牵涉细胞周期控制、凋亡和干细胞分化、血细胞生成、低氧、肌肉发育、神经发生、胰岛素分泌、胆固醇代谢、老化、病毒复制和免疫应答。

[0210] 单一miRNA可以靶向数百种不同的mRNA转录物,尽管许多不同的miRNA可以靶向个别转录物。在miRBase的最新版本(v.21)中已经注解了超过28645种微RNA。一些miRNA由多个基因座编码,其中的一些从串联地共转录的簇表达。所述特征允许具有多个途径和反馈控制的复杂调节网络。miRNA是这些反馈和调节回路的基本部分(integral parts),并且可以通过将蛋白生产保持在限度内来帮助调节基因表达(Herranz, H. & Cohen, S.M. *Genes Dev* 24, 1339-1344 (2010); Posadas, D.M. & Carthew, R.W. *Curr Opin Genet Dev* 27, 1-6 (2014))。

[0211] miRNA在与异常miRNA表达有关的大量人疾病中也是重要的。该关联强调了miRNA调节途径的重要性。最近的miRNA缺失研究已经将miRNA与免疫应答的调节相联系(Stern-Ginossar, N. 等人,*Science* 317, 376-381 (2007))。

[0212] miRNA还具有与癌症的强联系,并且可在不同类型的癌症中起作用。已经发现miRNA在许多肿瘤中下调。miRNA在关键的癌症相关途径诸如细胞周期控制和DNA损伤应答的调节中是重要的,且因此被用在诊断中且在临床上被靶向。微RNA微妙地调节血管生成的平衡,使得消耗所有微RNA的实验抑制肿瘤血管生成(Chen, S. 等人,*Genes Dev* 28, 1054-1067 (2014))。

[0213] 如关于蛋白编码基因已经显示的,miRNA基因也接受随癌症发生的表观遗传变化。许多miRNA基因座与CpG岛有关,增加它们通过DNA甲基化而调节的机会(Weber, B., Stresemann, C., Brueckner, B. & Lyko, F. *Cell Cycle* 6, 1001-1005 (2007))。大多数研究已经使用用染色质重塑药物的治疗来揭示表观遗传上沉默的miRNA。

[0214] 除了它们在RNA沉默中的作用以外,miRNA还可以活化翻译(Posadas, D.M. & Carthew, R.W. *Curr Opin Genet Dev* 27, 1-6 (2014))。敲除这些位点可导致减少的靶向基因的表达,而引入这些位点可增加表达。

[0215] 通过突变对于结合特异性而言重要的种子序列(微RNA的碱基2-8),可以最有效地敲除个别miRNA。在该区域中切割、随后通过NHEJ修复错误可以通过阻断与靶位点的结合而有效地消除miRNA功能。通过邻近回文序列的特定环区域的特异性靶向,也可以抑制miRNA。在催化上无活性的Cas9也可以用于抑制shRNA表达(Zhao, Y. 等人,*Sci Rep* 4, 3943 (2014))。除了靶向miRNA以外,还可以靶向和突变结合位点以防止通过miRNA的沉默。

[0216] 人细胞

[0217] 为了改善AATD,如本文中所述和解释的,用于基因编辑的主要靶标是人细胞。例如,在离体方法中,所述人细胞是体细胞,其在使用所述的技术修饰以后,可以产生肝细胞或祖细胞。例如,在体内方法中,所述人细胞是肝细胞、肾细胞或来自其它受影响的器官的

细胞。

[0218] 通过在源自需要的患者且因此已经与所述患者完全匹配的自体细胞中执行基因编辑,可能产生这样的细胞:其可以安全地重新引入所述患者中,且有效地产生将有效地改善与患者的疾病有关的一种或多种临床状况的细胞群体。

[0219] 祖细胞(在本文中也称为干细胞)能够繁殖并产生更多的祖细胞,它们进而具有产生大量母细胞的能力,所述母细胞进而可以产生分化的或可分化的子细胞。所述子细胞本身可以被诱导以繁殖和产生后代,所述后代随后分化成一种或多种成熟的细胞类型,同时也保留一种或多种具有亲本发育潜力的细胞。术语“干细胞”则指这样的细胞:其在特定情况下具有分化成更专门的或分化的表型的能力或潜力,且其在某些情况下保留基本上不分化地繁殖的能力。在一个实施方案中,术语祖细胞或干细胞指一般化的母细胞,其后代(子代)通过分化而专门化(经常在不同的方向),例如,通过完全获得个体特征,如在胚胎细胞和组织的进行性多样化中发生的。细胞的分化是通常通过许多细胞分裂而发生的复杂过程。分化的细胞可源自多潜能细胞,所述多潜能细胞本身源自多潜能细胞,以此类推。尽管这些多潜能细胞中的每一种可以被视作干细胞,但是每种可以产生的细胞类型的范围可相当大地变化。一些分化的细胞也具有产生更大发育潜力的细胞的能力。这样的能力可为天然的,或可在用各种因素处理后人工地诱导。在许多生物学情况下,干细胞也是“多潜能的”,因为它们可以产生超过一种独特细胞类型的后代,但这不是“干性(stem-ness)”所要求的。

[0220] 自我更新是干细胞的另一个重要方面。在理论上,自我更新可以通过两种主要机制中的任一种发生。干细胞可不对称地分裂,一个子代保留干状态,且另一个子代表达某种独特的其它特定功能和表型。可选地,群体中的某些干细胞可以对称地分裂成两个干,从而维持所述群体中的某些干细胞作为整体,而所述群体中的其它细胞仅产生分化的后代。通常,“祖细胞”具有更原始(即,沿着发育途径或进展处于比完全分化的细胞更早的步骤)的细胞表型。祖细胞也经常具有显著的或非常高的增殖潜力。祖细胞可以产生多种不同的分化的细胞类型或单一分化的细胞类型,这取决于发育途径和所述细胞在其中发育和分化的环境。

[0221] 在细胞个体发育的背景下,形容词“分化的”或“分化中的”是相对术语。“分化的细胞”是这样的细胞:其相对于与其进行对比的细胞已经沿着发育途径向下进一步发展。因而,干细胞可以分化成谱系限制的前体细胞(诸如肌细胞祖细胞),所述前体细胞进而可以沿着所述途径向下进一步分化成其它类型的前体细胞(诸如肌细胞前体),然后分化成末期分化的细胞,诸如肌细胞,其在某种组织类型中起独特作用,且可保留或可不保留进一步增殖的能力。

[0222] 诱导性多能干细胞

[0223] 在某些实施方案中,本文描述的遗传工程改造的人细胞是诱导性多能干细胞(iPSC)。使用iPSC的一个优点是,所述细胞可以源自要给其施用祖细胞的相同对象。也就是说,体细胞可以得自对象,重新程序化成诱导性多能干细胞,然后重新分化成要施用给对象的祖细胞(例如,自体细胞)。因为祖先基本上源自自体来源,与使用来自另一对象或对象组的细胞相比,植入排斥或变应性应答的风险减小。另外,iPSC的应用消除对得自胚胎来源的细胞的需要。因而,在一个实施方案中,在公开的方法中使用的干细胞不是胚胎干细胞。

[0224] 尽管分化在生理学背景下通常是不可逆的，最近已经开发了几种方法将体细胞重新程序化为iPSC。示例性的方法是本领域技术人员已知的并在下文中简要描述。

[0225] 术语“重新程序化”指改变或逆转分化的细胞(例如，体细胞)的分化状态的过程。换而言之，重新程序化指驱动细胞向后分化成更未分化的或更原始的细胞类型的过程。应当指出，将许多原代细胞置于培养中可以导致完全分化的特征的某种丧失。因而，在术语分化的细胞中包括的简单地培养这样的细胞不使这些细胞成为未分化细胞(例如，未分化的细胞)或多能细胞。分化的细胞向多能性的转变要求重新程序化刺激，其超过导致分化特征在培养中的部分丧失的刺激。相对于原代细胞亲本(其通常在培养中具有仅有限分裂次数的能力)，重新程序化的细胞也具有在不丧失生长潜力的情况下延长的传代的能力的特征。

[0226] 可以使要重新程序化的细胞在重新程序化之前部分地或最终地分化。在某些实施方案中，重新程序化包括分化的细胞(例如，体细胞)向多能状态或多潜能状态的分化状态的完全逆转。在某些实施方案中，重新程序化包括分化的细胞(例如，体细胞)向未分化的细胞(例如，胚胎样细胞)的分化状态的完全或部分逆转。重新程序化可以导致通过所述细胞的特定基因的表达，其表达进一步促成重新程序化。在本文所述的某些实施方案中，分化的细胞(例如，体细胞)的重新程序化造成分化的细胞呈现未分化的状态(例如，是未分化的细胞)。得到的细胞被称作“重新程序化的细胞”或“诱导性多能干细胞(iPSC或iPS细胞)”。

[0227] 重新程序化可以牵涉在细胞分化期间发生的核酸修饰(例如，甲基化)、染色质凝聚、表观遗传的变化、基因组印迹等中的至少一些可遗传模式的改变，例如逆转。重新程序化不同于简单地维持已经是多能的细胞的既存未分化状态或维持已经是多潜能细胞(例如，肌源性干细胞)的细胞的既存的小于完全分化的状态。重新程序化也不同于促进已经是多能或多潜能的细胞的自我更新或繁殖，尽管本文描述的组合物和方法在某些实施方案中也可以用于这样的目的。

[0228] 本领域已知许多可以用于从体细胞产生多能干细胞的方法。将体细胞重新程序化为多能表型的任何这样的方法将适合用在本文所述的方法中。

[0229] 已经描述了使用转录因子的确定组合产生多能细胞的重新程序化方法学。通过Oct4、Sox2、Klf4和c-Myc的直接转导，可以将小鼠体细胞转化成具有扩大的发育潜力的ES细胞样细胞；参见，例如，Takahashi和Yamanaka, Cell 126(4):663-76 (2006)。iPSC类似于ES细胞，因为它们恢复多能性相关的转录线路和许多表观遗传的景象。另外，小鼠iPSC满足多能性的所有标准测定：具体地，在体外分化成三个胚层的细胞类型，畸胎瘤形成，促成嵌合体，种系传递[参见，例如，Maherali和Hochedlinger, Cell Stem Cell. 3(6):595-605 (2008)]，和四倍体互补。

[0230] 使用类似的转导方法可以得到人iPSC，并且已经将转录因子三重小组OCT4、SOX2和NANOG确立为控制多能性的转录因子的核心集合；参见，例如，Budniatzky和Gepstein, Stem Cells Transl Med. 3(4):448-57 (2014); Barrett等人, Stem Cells Trans Med 3:1-6 sctm.2014-0121 (2014); Focosi等人, Blood Cancer Journal 4:e211 (2014); 和其中引用的参考文献。通过在历史上使用病毒载体将编码干细胞相关基因的核酸序列引入成年体细胞中，可以实现iPSC的生产。

[0231] 可以从终末分化的体细胞、以及从成年干细胞或成体干细胞产生或衍生iPSC。也就是说，通过重新程序化可以使非多能祖细胞成为多能的或多潜能的。在这样的情况下，可

能不必包括象重新程序化终末分化的细胞所需要的那么多的重新程序化因子。此外，通过重新程序化因子的非病毒引入，例如，通过引入所述蛋白本身，或通过引入编码重新程序化因子的核酸，或通过引入在翻译后产生重新程序化因子的信使RNA，可以诱导重新程序化(参见例如，Warren等人，Cell Stem Cell，7(5):618-30 (2010)。通过引入编码干细胞相关基因的核酸的组合可以实现重新程序化，所述干细胞相关基因包括，例如Oct-4 (也被称作Oct-3/4或Pouf51)、Sox1、Sox2、Sox3、Sox 15、Sox 18、NANOG、Klf1、Klf2、Klf4、Klf5、NR5A2、c-Myc、1-Myc、n-Myc、Rem2、Tert和LIN28。在一个实施方案中，使用本文描述的方法和组合物的重新程序化可以进一步包含向体细胞引入以下的一种或多种:Oct-3/4，Sox家族的成员，Klf家族的成员，和Myc家族的成员。在一个实施方案中，本文描述的方法和组合物还包含引入用于重新程序化的Oct-4、Sox2、Nanog、c-MYC和Klf4中的每一种的一种或多种。如上面所指出的，用于重新程序化的确切方法对于本文描述的方法和组合物而言不一定是关键性的。但是，在从重新程序化的细胞分化的细胞要用于例如人疗法中的情况下，在一个实施方案中，所述重新程序化不是通过改变基因组的方法实现。因而，在这样的实施方案中，例如，在不使用病毒或质粒载体的情况下，实现重新程序化。

[0232] 通过各种试剂例如小分子的添加，可以增强从开始细胞群体衍生出的重新程序化的效率(即，重新程序化的细胞的数目)，如以下所示:Shi等人，Cell-Stem Cell 2:525-528 (2008); Huangfu等人，Nature Biotechnology 26(7):795-797 (2008)和Marson等人，Cell-Stem Cell 3:132-135 (2008)。因而，增强诱导性多能干细胞生产的效率或速率的试剂或试剂组合可以用于生产患者特异性的或疾病特异性的iPSC。增强重新程序化效率的试剂的一些非限制性例子包括可溶性的Wnt、Wnt条件培养基、BIX-01294 (G9a组蛋白甲基转移酶)、PD0325901 (MEK抑制剂)、DNA甲基转移酶抑制剂、组蛋白脱乙酰基酶(HDAC)抑制剂、丙戊酸、5'-氮胞苷、地塞米松、suberoylanilide、氧肟酸(SAHA)、维生素C和曲古抑菌素(TSA)，以及其它。

[0233] 重新程序化增强剂的其它非限制性例子包括:辛二酰苯胺异羟肟酸(SAHA (例如，MK0683，伏林司他)和其它氧肟酸)，BML-210，Depudecin (例如，(-)-Depudecin)，HC毒素，Nullscript (4-(1,3-二氧代-1H,3H-苯并[de]异喹啉-2-基)-N-羟基丁酰胺)，苯基丁酸盐(例如，苯基丁酸钠)和丙戊酸((VP A)和其它短链脂肪酸)，Scriptaid，舒拉明钠，曲古抑菌素A (TSA)，APHA化合物8，Apicidin，丁酸钠，丁酸新戊酰基氧基甲酯(Pivanex，AN-9)，Trapoxin B，Chlamydocin，缩肽(也被称作FR901228或FK228)，苯甲酰胺(例如，CI-994 (例如，N-乙酰基地那林)和MS-27-275)，MGCD0103，NVP-LAQ-824，CBHA (间-羧基cinnaminic acid双氧肟酸)，JNJ16241199，Tubacin，A-161906，proxamide，oxamflatin，3-Cl-UCHA (例如，6-(3-氯苯基脲基)己酸异羟肟酸(6-(3-chlorophenylureido)caproic hydroxamic acid))，AOE (2-氨基-8-氧代-9,10-环氧癸酸)，CHAP31和CHAP 50。其它重新程序化增强剂包括，例如，HDAC的显性失活形式(例如，催化上无活性的形式)，HDAC的siRNA抑制剂，和特异性地结合HDAC的抗体。这样的抑制剂可得自，例如，BIOMOL International，Fukasawa，Merck Biosciences，Novartis，Gloucester Pharmaceuticals，Titan Pharmaceuticals，MethylGene，和Sigma Aldrich。

[0234] 为了证实用于随本文所述的方法使用的多能干细胞的诱导，可以针对干细胞标志

物的表达测试分离的克隆。在从体细胞衍生出的细胞中的这种表达将所述细胞鉴别为诱导性多能干细胞。干细胞标志物选自包括SSEA3、SSEA4、CD9、Nanog、Fbxl5、Ecatl、Esgl、Eras、Gdf3、Fgf4、Cripto、Daxl、Zpf296、Slc2a3、Rexl、Utfl和Natl的非限制性集合。在一个实施方案中，将表达Oct4或Nanog的细胞鉴别为多能的。用于检测这样的标志物的表达的方法可以包括，例如，RT-PCR和检测编码的多肽的存在的免疫学方法，诸如蛋白质印迹或流式细胞计数分析。在某些实施方案中，检测不仅牵涉RT-PCR，而且包括蛋白标志物的检测。通过RT-PCR或蛋白检测方法诸如免疫细胞化学，可最佳地鉴别细胞内标志物，而通过例如免疫细胞化学容易地鉴别细胞表面标志物。

[0235] 通过评价iPSC的分化成三个胚层中的每一个的细胞的能力的测试，可以证实分离的细胞的多能干细胞特性。作为一个例子，在裸鼠中的畸胎瘤形成可以用于评价分离的克隆的多能特性。将所述细胞引入裸鼠中，并在从所述细胞产生的肿瘤上执行组织学和/或免疫组织化学。例如，包含来自所有三个胚层的细胞的肿瘤的生长进一步指示，所述细胞是多能干细胞。

[0236] 肝细胞

[0237] 在某些实施方案中，本文描述的遗传工程改造的人细胞是肝细胞。肝细胞是肝的主要实质组织的细胞。肝细胞占肝的质量的70-85%。这些细胞牵涉：蛋白合成；蛋白贮存；碳水化合物的转化；胆固醇、胆汁盐和磷脂的合成；外源性和内源性物质的解毒、修饰和排泄；以及胆汁的形成的起始和分泌。

[0238] SERPINA1主要在肝细胞(实质肝细胞)中表达，其是循环蛋白的主要来源，和在单核细胞及嗜中性粒细胞中次要表达。因此，SERPINA1的校正将主要靶向肝细胞和肝。

[0239] 建立患者特异性的iPSC

[0240] 本发明的离体方法的一个步骤牵涉，建立一个患者特异性的iPS细胞、多个患者特异性的iPS细胞或患者特异性的iPS细胞系。在本领域中存在许多确定的用于建立患者特异性的iPS细胞的方法，如在以下中所述：Takahashi和Yamanaka 2006; Takahashi, Tanabe等人. 2007。例如，所述建立步骤包括：a)从所述患者分离体细胞，诸如皮肤细胞或成纤维细胞；和b)将一组多能性相关基因引入所述体细胞中以诱导所述细胞变成多能干细胞。在某些实施方案中，所述组的多能性相关基因是一种或多种选自以下的基因：OCT4、SOX2、KLF4、Lin28、NANOG和cMYC。

[0241] 执行患者的肝或骨髓的活组织检查或抽吸

[0242] 活组织检查或抽吸是从身体取出的组织或流体的样品。存在许多不同种类的活组织检查或抽吸。它们几乎全部牵涉使用尖锐的工具取出少量组织。如果活组织检查将是在皮肤或其它敏感区域上，首先应用麻醉机器。根据本领域的已知方法中的任一种，可执行活组织检查或抽吸。例如，在肝活组织检查中，将针穿过腹部皮肤注射进肝中，捕获肝组织。例如，在骨髓抽吸中，使用大针进入骨盆骨中以收集骨髓。

[0243] 分离肝特异性的祖细胞或原代肝细胞

[0244] 根据本领域已知的任意方法可分离肝特异性的祖细胞和原代肝细胞。例如，从新鲜外科手术样本(例如，自体样品)分离人肝细胞。使用健康的肝组织通过胶原酶消化来分离肝细胞。将得到的细胞悬浮液穿过100-mm尼龙网过滤，并通过在50g离心5分钟进行沉淀，再悬浮，并在冷的洗涤介质中洗涤2-3次。通过在严谨条件下培养从新鲜肝制品得到的肝细

胞，得到人肝干细胞。将接种在胶原包被的平板上的肝细胞培养2周。2周以后，将存活的细胞取出，并针对干细胞标志物的表达进行表征(Herrera等人, STEM CELLS 2006;24:2840-2850)。

[0245] 分离间充质干细胞

[0246] 根据本领域已知的任意方法可分离间充质干细胞，诸如从患者骨髓或外周血。例如，将骨髓抽吸物收集进具有肝素的注射器中。将细胞洗涤并在Percoll™上离心。使用Percoll™分离细胞，诸如血细胞、肝细胞、间质细胞、巨噬细胞、肥大细胞和胸腺细胞。在含有10%胎牛血清(FBS)的Dulbecco氏改良的伊格尔培养基(DMEM)(低葡萄糖)中培养细胞(Pittinger MF, Mackay AM, Beck SC等人, Science 1999; 284:143-147)。

[0247] 基因组编辑

[0248] 基因组编辑通常指改变基因组的核苷酸序列的过程，优选地以精确的或预定的方式。本文描述的基因组编辑的方法的例子包括使用定位核酸酶在基因组中的精确靶位置处切割脱氧核糖核酸(DNA)、由此在基因组内的特定位置处建立双链或单链DNA断裂的方法。这样的断裂可以且经常通过天然的内源性细胞过程诸如同源性指导的修复(HDR)和非同源末端连接(NHEJ)来修复，如最近在Cox等人, Nature Medicine 21(2), 121-31 (2015)中综述的。这两种主要的DNA修复过程由旁路途径家族组成。NHEJ直接连接从双链断裂产生的DNA末端，有时具有核苷酸序列的损失或添加，其可破坏或增强基因表达。HDR利用同源序列或供体序列作为在断裂点处插入确定的DNA序列的模板。所述同源序列可在内源性基因组诸如姐妹染色单体中。可选地，所述供体可为外源核酸，诸如质粒、单链寡核苷酸、双链寡核苷酸、双链体寡核苷酸或病毒，其具有与核酸酶切割的基因座高同源性的区域，但是其还可含有另外的序列或序列变化，包括可以掺入经切割的靶基因座中的缺失。第三种修复机制是微同源性介导的末端连接(MMEJ)，也被称作“可选NHEJ”，其中基因结果类似于NHEJ，因为小缺失和插入可以发生在切割位点处。MMEJ利用侧接DNA断裂位点的几个碱基对的同源序列来驱动更有利的DNA末端连接修复结果，且最近的报道已经进一步阐明了该过程的分子机制；参见，例如，Cho和Greenberg, Nature 518, 174-76 (2015)；Kent等人, Nature Structural and Molecular Biology, Adv. Online doi:10.1038/nsmb.2961(2015)；Mateos-Gomez等人, Nature 518, 254-57 (2015); Ceccaldi等人, Nature 528, 258-62 (2015)。在某些情况下，基于在DNA断裂位点处的潜在微同源性分析可能预测可能的修复结果。

[0249] 这些基因组编辑机制中的每一种可以用于建立期望的基因组改变。基因组编辑过程中的一个步骤是在尽可能接近预期突变位点的靶基因座中建立一个或两个DNA断裂，后者作为双链断裂或作为两个单链断裂。这可以通过使用定位多肽来实现，如本文中所述和解释的。

[0250] 定位多肽诸如DNA内切核酸酶可以在核酸例如基因组DNA中引入双链断裂或单链断裂。所述双链断裂可以刺激细胞的内源性DNA-修复途径(例如，同源性依赖性的修复或非同源末端连接或可选非同源末端连接(A-NHEJ)或微同源性介导的末端连接)。NHEJ可以修复经切割的靶核酸，无需同源模板。这有时可以导致在靶核酸中在切割位点处的小缺失或插入(插入/缺失(indels))，并且可以导致基因表达的破坏或改变。当同源修复模板或供体可得到时，HDR可以发生。同源供体模板包含与侧接靶核酸切割位点的序列同源的序列。姐

妹染色单体通常被细胞用作修复模板。但是，为了基因组编辑的目的，所述修复模板经常作为外源核酸诸如质粒、双链体寡核苷酸、单链寡核苷酸、双链寡核苷酸或病毒核酸来供给。应用外源供体模板，经常在同源性的侧接区域之间引入另外的核酸序列(诸如转基因)或修饰(诸如单个或多个碱基变化或缺失)，使得另外的或改变的核酸序列也变成掺入靶基因座中。MMEJ导致与NHEJ类似的基因结果，因为小缺失和插入可以发生在切割位点处。MMEJ利用侧接切割位点的几个碱基对的同源序列来驱动更有利的末端连接DNA修复结果。在某些情况下，基于在核酸酶靶区域中的潜在微同源性分析可能预测可能的修复结果。

[0251] 因而，在某些情况下，使用同源重组将外源多核苷酸序列插入靶核酸切割位点中。外源多核苷酸序列在本文中被称作供体多核苷酸(或供体或供体序列或多核苷酸供体模板)。在某些实施方案中，将所述供体多核苷酸、所述供体多核苷酸的一部分、所述供体多核苷酸的一个拷贝或所述供体多核苷酸的一个拷贝的部分插入靶核酸切割位点中。在某些实施方案中，所述供体多核苷酸是外源多核苷酸序列，即，在靶核酸切割位点处不天然地发生的序列。

[0252] 由NHEJ和/或HDR引起的靶DNA的修饰可以导致，例如，突变、缺失、改变、整合、基因校正、基因替换、基因加标签、转基因插入、核苷酸缺失、基因破坏、易位和/或基因突变。缺失基因组DNA和将非天然核酸掺入基因组DNA中的过程是基因组编辑的例子。

[0253] CRISPR内切核酸酶系统

[0254] CRISPR(簇集的规则地间隔的短回文重复序列(Clustered Regularly Interspaced Short Palindromic Repeats))基因组基因座可以见于许多原核生物(例如细菌和古细菌)的基因组中。在原核生物中，CRISPR基因座编码这样的产物：其作为一类免疫系统起作用以帮助保护所述原核生物免于外来侵入物，诸如病毒和噬菌体。存在CRISPR基因座功能的三个阶段：新序列整合进CRISPR基因座中，CRISPR RNA (crRNA)的表达，和外来侵入核酸的沉默。已经鉴别了五类CRISPR系统(例如，类型I、类型II、类型III、类型U和类型V)。

[0255] CRISPR基因座包括许多短的重复序列，被称作“重复序列”。当表达时，所述重复序列可以形成二级发夹结构(例如，发夹)和/或包含未结构化的单链序列。所述重复序列经常发生在簇中且频繁地在物种之间差异化。所述重复序列被独特的间插序列(被称作“间隔物”)规则地隔开，导致重复序列-间隔物-重复序列基因座体系结构。所述间隔物与已知的外来侵入序列相同或具有高同源性。间隔物-重复序列单元编码crisprRNA (crRNA)，其被加工成间隔物-重复序列单元的成熟形式。crRNA包含牵涉靶向靶核酸的“种子”或间隔物序列(在原核生物中以天然存在的形式，所述间隔物序列靶向外来侵入核酸)。间隔物序列位于crRNA的5’或3’末端处。

[0256] CRISPR基因座也包含编码CRISPR相关(Cas)基因的多核苷酸序列。Cas基因编码牵涉原核生物中的crRNA功能的生源说和干扰阶段的内切核酸酶。一些Cas基因包含同源的二级和/或三级结构。

[0257] II型CRISPR系统

[0258] 在自然界中在II型CRISPR系统中的crRNA生源说要求反式活化的CRISPR RNA (tracrRNA)。所述tracrRNA被内源性RNA酶III修饰，然后与前-crRNA阵列中的crRNA重复序列杂交。内源性RNA酶III被募集以切割前-crRNA。使经切割的crRNA接受核糖核酸外切酶修

剪以产生成熟的crRNA形式(例如,5'修剪)。所述tracrRNA保持与crRNA杂交,并且tracrRNA和crRNA与定位多肽(例如,Cas9)结合。crRNA-tracrRNA-Cas9复合物的crRNA将所述复合物引导至所述crRNA可以与其杂交的靶核酸。所述crRNA与靶核酸的杂交活化Cas9用于靶向核酸切割。在II型CRISPR系统中的靶核酸被称作原间隔物邻近基序(PAM)。在自然界中,所述PAM是促进定位多肽(例如,Cas9)与靶核酸的结合所必需的。II型系统(也被称作Nmeni或CASS4)被进一步细分成类型II-A (CASS4)和II-B (CASS4a)。Jinek等人,*Science, 337* (6096):816-821 (2012)显示,CRISPR/Cas9系统可用于RNA-可编程的基因组编辑,并且国际专利申请公开号WO2013/176772提供了CRISPR/Cas内切核酸酶系统用于位点特异性的基因编辑的众多例子和应用。

[0259] V型CRISPR系统

[0260] V型CRISPR系统具有与II型系统的几个重要差异。例如,Cpf1是单个RNA引导的内切核酸酶,其与II型系统相比缺乏tracrRNA。实际上,Cpf1-相关的CRISPR阵列被加工成成熟的crRNAS,不需要另外的反式活化的tracrRNA。V型CRISPR阵列被加工成42-44个核苷酸长度的短的成熟的crRNA,和每种成熟的crRNA开始于直接重复序列的19个核苷酸,继之以间隔物序列的23-25个核苷酸。与之相比,在II型系统中的成熟的crRNA开始于间隔物序列的20-24个核苷酸,继之以直接重复序列的约22个核苷酸。并且,Cpf1利用富含T的原间隔物-邻近基序,使得Cpf1-crRNA复合物有效地切割在短的富含T的PAM以后的靶DNA,这与对于II型系统而言在靶DNA以后的富含G的PAM形成对比。因而,V型系统在远离PAM的点处切割,而II型系统在邻近PAM的点处切割。另外,与II型系统相比,Cpf1通过具有4或5个核苷酸5'突出端的交错DNA双链断裂而切割DNA。II型系统通过平端双链断裂而切割。类似于II型系统,Cpf1含有预测的RuvC-样内切核酸酶结构域,但是缺乏第二个HNH内切核酸酶结构域,这与II型系统形成对比。

[0261] Cas基因/多肽和原间隔物邻近基序

[0262] 示例性的CRISPR/Cas多肽包括Fonfara等人,*Nucleic Acids Research, 42*:2577-2590 (2014)的图1中的Cas9多肽。因为发现了Cas基因,CRISPR/Cas基因命名系统已经经历全面的重写。Fonfara(出处同上)的图5提供了来自各种物种的Cas9多肽的PAM序列。

[0263] 定位多肽(Site-Directed Polypeptides)

[0264] 定位多肽是在基因组编辑中用于切割DNA的核酸酶。所述定位可作为一种或多种多肽或一种或多种编码所述多肽的mRNA施用给细胞或患者。

[0265] 在CRISPR/Cas或CRISPR/Cpf1系统的背景下,定位多肽可以结合引导RNA,所述引导RNA进而指定所述多肽所针对的靶DNA中的位点。在本文的CRISPR/Cas或CRISPR/Cpf1系统的实施方案中,所述定位多肽是内切核酸酶,诸如DNA内切核酸酶。

[0266] 在某些实施方案中,定位多肽包含多个核酸切割(即,核酸酶)结构域。两个或更多个核酸切割结构域可以经由接头连接在一起。在某些实施方案中,所述接头包含柔性接头。接头可包含1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40个或更多个氨基酸的长度。

[0267] 天然存在的野生型Cas9酶包含两个核酸酶结构域:HNH核酸酶结构域和RuvC结构域。在本文中,"Cas9"指天然存在的和重组的Cas9。在本文中考虑的Cas9酶包含HNH或HNH-样核酸酶结构域、和/或RuvC或RuvC-样核酸酶结构域。

[0268] HNH或HNH-样结构域包含McrA-样折叠。HNH或HNH-样结构域包含两个反平行的β-链和一个α-螺旋。HNH或HNH-样结构域包含金属结合位点(例如,二价阳离子结合位点)。HNH或HNH-样结构域可以切割靶核酸的一条链(例如,crRNA靶向链的互补链)。

[0269] RuvC或RuvC-样结构域包含RNA酶H或RNA酶H-样折叠。RuvC/RNA酶H结构域牵涉多种多样组的基于核酸的功能,包括作用于RNA和DNA。RNA酶H结构域包含被多个α-螺旋包围的5个β-链。RuvC/RNA酶H或RuvC/RNA酶H-样结构域包含金属结合位点(例如,二价阳离子结合位点)。RuvC/RNA酶H或RuvC/RNA酶H-样结构域可以切割靶核酸的一条链(例如,双链靶DNA的非互补链)。

[0270] 定位多肽可以在核酸例如基因组DNA中引入双链断裂或单链断裂。所述双链断裂可以刺激细胞的内源性DNA-修复途径(例如,同源性依赖性的修复(HDR)或非同源末端连接(NHEJ)或可选非同源末端连接(A-NHEJ)或微同源性介导的末端连接(MMEJ))。NHEJ可以修复经切割的靶核酸,无需同源模板。这有时可以导致在靶核酸中在切割位点处的小缺失或插入(插入/缺失),并且可以导致基因表达的破坏或改变。当同源修复模板或供体可得到时,HDR可以发生。同源供体模板包含与侧接靶核酸切割位点的序列同源的序列。姐妹染色单体通常被细胞用作修复模板。但是,为了基因组编辑的目的,所述修复模板经常作为外源核酸诸如质粒、双链体寡核苷酸、单链寡核苷酸或病毒核酸来供给。应用外源供体模板,经常在同源性的侧接区域之间引入另外的核酸序列(诸如转基因)或修饰(诸如单个或多个碱基变化或缺失),使得另外的或改变的核酸序列也变成掺入靶基因座中。MMEJ导致与NHEJ类似的基因结果,因为小缺失和插入可以发生在切割位点处。MMEJ利用侧接切割位点的几个碱基对的同源序列来驱动有利的末端连接DNA修复结果。在某些情况下,基于在核酸酶靶区域中的潜在微同源性分析可能预测可能的修复结果。

[0271] 因而,在某些情况下,使用同源重组将外源多核苷酸序列插入靶核酸切割位点中。外源多核苷酸序列在本文中被称作供体多核苷酸(或供体或供体序列)。在某些实施方案中,将所述供体多核苷酸、所述供体多核苷酸的一部分、所述供体多核苷酸的一个拷贝或所述供体多核苷酸的一个拷贝的部分插入靶核酸切割位点中。在某些实施方案中,所述供体多核苷酸是外源多核苷酸序列,即,在靶核酸切割位点处不天然地发生的序列。

[0272] 由NHEJ和/或HDR引起的靶DNA的修饰可以导致,例如,突变、缺失、改变、整合、基因校正、基因替换、基因加标记、转基因插入、核苷酸缺失、基因破坏、易位和/或基因突变。缺失基因组DNA和将非天然核酸掺入基因组DNA中的过程是基因组编辑的例子。

[0273] 在某些实施方案中,所述定位多肽包含与野生型示例性定位多肽[例如,来自化脓性链球菌的Cas9,US2014/0068797序列ID No. 8或Sapranauskas等人, *Nucleic Acids Res, 39*(21):9275-9282 (2011)]和各种其它定位多肽)具有至少10%、至少15%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少99%或100%氨基酸序列同一性的氨基酸序列。

[0274] 在某些实施方案中,所述定位多肽包含与野生型示例性定位多肽(例如,来自化脓性链球菌的Cas9, 出处同上)的核酸酶结构域具有至少10%、至少15%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少99%或100%氨基酸序列同一性的氨基酸序列。

[0275] 在某些实施方案中,定位多肽包含与野生型定位多肽(例如,来自化脓性链球菌的

Cas9，出处同上)在10个连续氨基酸上的至少70、75、80、85、90、95、97、99或100%同一性。在某些实施方案中，定位多肽包含与野生型定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)在10个连续氨基酸上的至多70、75、80、85、90、95、97、99或100%同一性。在某些实施方案中，定位多肽包含与野生型定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)在所述定位多肽的HNH核酸酶结构域中的10个连续氨基酸上的至少70、75、80、85、90、95、97、99或100%同一性。在某些实施方案中，定位多肽包含与野生型定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)在所述定位多肽的HNH核酸酶结构域中的10个连续氨基酸上的至多70、75、80、85、90、95、97、99或100%同一性。在某些实施方案中，定位多肽包含与野生型定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)在所述定位多肽的RuvC核酸酶结构域的10个连续氨基酸上的至少70、75、80、85、90、95、97、99或100%同一性。在某些实施方案中，定位多肽包含与野生型定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)在所述定位多肽的RuvC核酸酶结构域的10个连续氨基酸上的至多70、75、80、85、90、95、97、99或100%同一性。

[0276] 在某些实施方案中，所述定位多肽包含野生型示例性定位多肽的经修饰形式。所述野生型示例性定位多肽的经修饰形式包含减小所述定位多肽的核酸切割活性的突变。在某些实施方案中，所述野生型示例性定位多肽的经修饰形式具有所述野生型示例性定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)的核酸切割活性的小于90%、小于80%、小于70%、小于60%、小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1%。所述定位多肽的经修饰形式可以不具有实质的核酸切割活性。当定位多肽是不具有实质的核酸切割活性的经修饰形式时，它在本文中被称作“无酶活性的”。

[0277] 在某些实施方案中，所述定位多肽的经修饰形式包含突变，使得它可以诱导靶核酸上的单链断裂(SSB) (例如，通过仅切割双链靶核酸的糖-磷酸主链之一)。在某些实施方案中，所述突变导致所述野生型定位多肽(例如，来自化脓性链球菌的Cas9，出处同上)的多个核酸切割结构域的一个或多个中的小于90%、小于80%、小于70%、小于60%、小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或小于1%的核酸切割活性。在某些实施方案中，所述突变导致多个核酸切割结构域中的一个或多个保留切割靶核酸的互补链的能力、但是减小它的切割靶核酸的非互补链的能力。在某些实施方案中，所述突变导致多个核酸切割结构域中的一个或多个保留切割靶核酸的非互补链的能力、但是减小它的切割靶核酸的互补链的能力。例如，野生型示例性化脓性链球菌Cas9多肽中的残基诸如Asp10、His840、Asn854和Asn856被突变以失活多个核酸切割结构域(例如，核酸酶结构域)中的一个或多个。在某些实施方案中，要突变的残基对应于野生型示例性化脓性链球菌Cas9多肽中的残基Asp10、His840、Asn854和Asn856 (例如，如通过序列和/或结构比对所确定的)。突变的非限制性例子包括D10A、H840A、N854A或N856A。本领域技术人员会认识到，除了丙氨酸置换以外的突变是合适的。

[0278] 在某些实施方案中，将D10A突变与H840A、N854A或N856A突变中的一种或多种组合以产生基本上缺乏DNA切割活性的定位多肽。在某些实施方案中，将H840A突变与D10A、N854A或N856A突变中的一种或多种组合以产生基本上缺乏DNA切割活性的定位多肽。在某些实施方案中，将N854A突变与H840A、D10A或N856A突变中的一种或多种组合以产生基本上缺乏DNA切割活性的定位多肽。在某些实施方案中，将N856A突变与H840A、N854A或D10A突变

中的一种或多种组合以产生基本上缺乏DNA切割活性的定位多肽。包含一个基本上无活性的核酸酶结构域的定位多肽被称作“切口酶”。

[0279] RNA引导的内切核酸酶例如Cas9的切口酶变体可以用于增加CRISPR介导的基因组编辑的特异性。野生型Cas9通常由单个引导RNA来引导，所述引导RNA被设计成与靶序列中的指定的约20个核苷酸序列(诸如内源性基因组基因座)杂交。但是，可以在引导RNA和靶基因座之间容忍几个错配，将所述靶位点中需要的同源性的长度有效地减小至例如小至13个核苷酸的同源性，且由此导致CRISPR/Cas9复合物在靶基因组别处的结合和双链核酸切割——也被称作靶外切割的升高的可能性。因为Cas9的切口酶变体为了建立双链断裂各自仅切割一条链，必须使一对切口酶紧密靠近和在靶核酸的相对链上结合，由此建立一对切口，其为双链断裂的等同物。这要求两种单独的引导RNA——对于每一切口酶为一种——必须紧密靠近和在靶核酸的相对链上结合。该要求使双链断裂发生所需的同源性的最小长度基本上倍增，由此减小以下可能性：双链切割事件将发生在基因组的别处，在该处所述两个引导RNA位点——如果它们存在的话——不可能彼此充分靠近以使双链断裂形成。如在本领域中描述的，还可以使用切口酶来相对于NHEJ促进HDR。通过使用有效地介导期望的变化的特定供体序列，可以使用HDR将选择的变化引入基因组中的靶位点。用于基因编辑中的各种CRISPR/Cas系统的描述可以见于，例如，国际专利申请公开号WO2013/176772，和*Nature Biotechnology 32*, 347-355 (2014)，和其中引用的参考文献。

[0280] 考虑的突变包括置换、添加和缺失，或它们的任意组合。在某些实施方案中，所述突变将突变的氨基酸转化成丙氨酸。在某些实施方案中，所述突变将突变的氨基酸转化成另一种氨基酸(例如，甘氨酸、丝氨酸、苏氨酸、半胱氨酸、缬氨酸、亮氨酸、异亮氨酸、甲硫氨酸、脯氨酸、苯丙氨酸、酪氨酸、色氨酸、天冬氨酸、谷氨酸、天冬酰胺、谷氨酰胺、组氨酸、赖氨酸或精氨酸)。在某些实施方案中，所述突变将突变的氨基酸转化成非天然的氨基酸(例如，硒代甲硫氨酸)。在某些实施方案中，所述突变将突变的氨基酸转化成氨基酸模仿物(例如，磷酸模仿物)。在某些实施方案中，所述突变是保守突变。例如，所述突变可以将突变的氨基酸转化成类似于突变的氨基酸的大小、形状、电荷、极性、构象和/或旋转异构体的氨基酸(例如，半胱氨酸/丝氨酸突变、赖氨酸/天冬酰胺突变、组氨酸/苯丙氨酸突变)。在某些实施方案中，所述突变造成读码框的转移和/或未成熟终止密码子的建立。在某些实施方案中，突变造成影响一种或多种基因的表达的基因或基因座的调节区的变化。

[0281] 在某些实施方案中，所述定位多肽(例如，变体的、突变的、无酶活性的和/或有条件地无酶活性的定位多肽)靶向核酸。在某些实施方案中，所述定位多肽(例如，变体的、突变、无酶活性的和/或有条件地无酶活性的核糖核酸内切酶)靶向DNA。在某些实施方案中，所述定位多肽(例如，变体的、突变的、无酶活性的和/或有条件地无酶活性的核糖核酸内切酶)靶向RNA。

[0282] 在某些实施方案中，所述定位多肽包含一种或多种非天然序列(例如，所述定位多肽是融合蛋白)。

[0283] 在某些实施方案中，所述定位多肽包含含有与来自细菌(例如，化脓性链球菌)的Cas9至少15%的氨基酸同一性的氨基酸序列、核酸结合结构域和两个核酸切割结构域(即，HNH结构域和RuvC结构域)。

[0284] 在某些实施方案中，所述定位多肽包含含有与来自细菌(例如，化脓性链球菌)的

Cas9至少15%的氨基酸同一性的氨基酸序列和两个核酸切割结构域(即,HNH结构域和RuvC结构域)。

[0285] 在某些实施方案中,所述定位多肽包含含有与来自细菌(例如,化脓性链球菌)的Cas9至少15%的氨基酸同一性的氨基酸序列和两个核酸切割结构域,其中所述核酸切割结构域中的一个或两个含有与来自细菌(例如,化脓性链球菌)的Cas9的核酸酶结构域至少50%的氨基酸同一性。

[0286] 在某些实施方案中,所述定位多肽包含含有与来自细菌(例如,化脓性链球菌)的Cas9至少15%的氨基酸同一性的氨基酸序列、两个核酸切割结构域(即,HNH结构域和RuvC结构域)和非天然序列(例如,核定位信号)或将所述定位多肽连接至非天然序列的接头。

[0287] 在某些实施方案中,所述定位多肽包含含有与来自细菌(例如,化脓性链球菌)的Cas9至少15%的氨基酸同一性的氨基酸序列、两个核酸切割结构域(即,HNH结构域和RuvC结构域),其中所述定位多肽包含在所述核酸切割结构域中的一个或两个中的突变,所述突变使所述核酸酶结构域的切割活性减小至少50%。

[0288] 在某些实施方案中,所述定位多肽包含含有与来自细菌(例如,化脓性链球菌)的Cas9至少15%的氨基酸同一性的氨基酸序列和两个核酸切割结构域(即,HNH结构域和RuvC结构域),其中所述核酸酶结构域之一包含天冬氨酸10的突变,和/或其中所述核酸酶结构域之一包含组氨酸840的突变,且其中所述突变使所述核酸酶结构域(一个或多个)的切割活性减小至少50%。

[0289] 在本发明的某些实施方案中,一种或多种定位多肽例如DNA内切核酸酶包括两种切口酶,它们一起实现在基因组中特定基因座处的一个双链断裂,或四种切口酶,它们一起实现在基因组中特定基因座处的两个双链断裂。可选地,一种定位多肽例如DNA内切核酸酶实现在基因组中特定基因座处的一个双链断裂。

[0290] 靶向基因组的核酸

[0291] 本公开提供了靶向基因组的核酸,其可以指导有关多肽(例如,定位多肽)对靶核酸内的特定靶序列的活性。在某些实施方案中,所述靶向基因组的核酸是RNA。靶向基因组的RNA在本文中被称作“引导RNA”或“gRNA”。引导RNA至少包含与目标靶核酸序列杂交的间隔物序列和CRISPR重复序列。在II型系统中,所述gRNA也包含被称作tracrRNA序列的第二种RNA。在II型引导RNA (gRNA)中,CRISPR重复序列和tracrRNA序列彼此杂交以形成双链体。在V型引导RNA (gRNA)中,crRNA形成双链体。在两种系统中,所述双链体结合定位多肽,使得引导RNA和定位多肽形成复合物。所述靶向基因组的核酸通过它与定位多肽的结合为所述复合物提供靶标特异性。所述靶向基因组的核酸因而指导所述定位多肽的活性。

[0292] 示例性的引导RNA包括在SEQ ID No:1-68,297,例如在SEQ ID No:54,860-68,297中的间隔物序列,其与它们的靶序列和有关的Cas9或Cpf1切割位点的基因组位置一起显示,其中所述基因组位置是基于GRCh38/hg38人基因组组件。如本领域普通技术人员理解的,将每种引导RNA设计成包括与它的基因组靶序列互补的间隔物序列。例如,在SEQ ID No:1-68,297,例如在SEQ ID No:54,860-68,297中的每个间隔物序列可放入单个RNA嵌合体或crRNA中(与对应的tracrRNA一起)。参见Jinek等人, Science, 337, 816-821 (2012)和Deltcheva等人, Nature, 471, 602-607 (2011)。

[0293] 在某些实施方案中,所述靶向基因组的核酸是双分子引导RNA。在某些实施方案

中,所述靶向基因组的核酸是单分子引导RNA。

[0294] 双分子引导RNA包含两条RNA链。第一链在5’至3’方向包含任选的间隔物延伸序列、间隔物序列和最小CRISPR重复序列。第二链包含最小tracrRNA序列(与最小CRISPR重复序列互补)、3’tracrRNA序列和任选的tracrRNA延伸序列。

[0295] 在II型系统中的单分子引导RNA (sgRNA)在5’至3’方向包含任选的间隔物延伸序列、间隔物序列、最小CRISPR重复序列、单分子引导接头、最小tracrRNA序列、3’tracrRNA序列和任选的tracrRNA延伸序列。所述任选的tracrRNA延伸可包含为引导RNA贡献另外的功能性(例如,稳定性)的元件。所述单分子引导接头连接最小CRISPR重复序列和最小tracrRNA序列以形成发夹结构。所述任选的tracrRNA延伸包含一个或多个发夹。

[0296] 在V型系统中的单分子引导RNA (sgRNA)在5’至3’方向包含最小CRISPR重复序列和间隔物序列。

[0297] 作为举例说明,在CRISPR/Cas系统中使用的引导RNA或其它更小的RNA可以通过化学方式容易地合成,如在下面举例说明的和在本领域中描述的。尽管化学合成程序在继续增加,但是通过诸如高效液相色谱法(HPLC,其避免使用凝胶诸如PAGE)的程序的这样的RNA的纯化倾向于变得更有挑战性,因为多核苷酸长度显著地增加超过100个左右核苷酸。用于产生较大长度的RNA的一个途径是生产连接在一起的两个或更多个分子。通过酶法更容易地产生长得多的RNA,诸如编码Cas9或Cpf1内切核酸酶的那些。在RNA的化学合成和/或酶法产生期间或以后,可以引入各种类型的RNA修饰,例如,如本领域中描述的增加稳定性、减小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。

[0298] 间隔物延伸序列

[0299] 在靶向基因组的核酸的某些实施方案中,间隔物延伸序列可以提供稳定性和/或提供用于修饰靶向基因组的核酸的位置。间隔物延伸序列可以改变在靶或靶外(on- or off-target)活性或特异性。在某些实施方案中,提供了间隔物延伸序列。间隔物延伸序列可具有超过1、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、120、140、160、180、200、220、240、260、280、300、320、340、360、380、400、1000、2000、3000、4000、5000、6000或7000个或更多个核苷酸的长度。间隔物延伸序列可具有小于1、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、120、140、160、180、200、220、240、260、280、300、320、340、360、380、400、1000、2000、3000、4000、5000、6000、7000个或更多个核苷酸的长度。在某些实施方案中,间隔物延伸序列的长度小于10个核苷酸。在某些实施方案中,间隔物延伸序列的长度在10-30个核苷酸之间。在某些实施方案中,间隔物延伸序列的长度在30-70个核苷酸之间。

[0300] 在某些实施方案中,所述间隔物延伸序列包含另一个部分(例如,稳定性控制序列、核糖核酸内切酶结合序列、核酶)。在某些实施方案中,所述部分减小或增加靶向核酸的核酸的稳定性。在某些实施方案中,所述部分是转录终止子区段(即,转录终止序列)。在某些实施方案中,所述部分在真核细胞中起作用。在某些实施方案中,所述部分在原核细胞中起作用。在某些实施方案中,所述部分在真核和原核细胞二者中起作用。合适的部分的非限制性例子包括:5’帽(例如,7-甲基鸟苷酸酯帽(m7 G))、核糖开关序列(例如,以允许通过蛋白和蛋白复合物的经调节的稳定性和/或经调节的可达性)、形成dsRNA双链体(即,发夹)的序列、将RNA靶向至亚细胞位置(例如,细胞核、线粒体、叶绿体等)的序列、提供跟踪的修饰或序列(例如,直接缀合至荧光分子,缀合至促进荧光检测的部分,允许荧光检测的序列

等)和/或提供蛋白(例如,在DNA上起作用的蛋白,包括转录活化剂、转录阻遏物、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰基转移酶、组蛋白脱乙酰基酶等)的结合位点的修饰或序列。

[0301] 间隔物序列

[0302] 所述间隔物序列与目标靶核酸中的序列杂交。靶向基因组的核酸的间隔物通过杂交(即,碱基配对)以序列特异性的方式与靶核酸相互作用。所述间隔物的核苷酸序列因而取决于目标靶核酸的序列而变化。

[0303] 在本文的CRISPR/Cas系统中,将间隔物序列设计成与靶核酸杂交,所述靶核酸位于在所述系统中使用的Cas9酶的PAM的5'侧。所述间隔物可与靶序列完美匹配或可具有错配。每种Cas9酶具有它在靶DNA中识别的特定PAM序列。例如,化脓性链球菌在靶核酸中识别包含序列5'-NRG-3'的PAM,其中R包含A或G,其中N是任意核苷酸,且N在被间隔物序列靶向的靶核酸序列的3'紧邻处。

[0304] 在某些实施方案中,所述靶核酸序列包含20个核苷酸。在某些实施方案中,所述靶核酸包含小于20个核苷酸。在某些实施方案中,所述靶核酸包含超过20个核苷酸。在某些实施方案中,所述靶核酸包含至少5、10、15、16、17、18、19、20、21、22、23、24、25、30个或更多个核苷酸。在某些实施方案中,所述靶核酸包含至多5、10、15、16、17、18、19、20、21、22、23、24、25、30个或更多个核苷酸。在某些实施方案中,所述靶核酸序列包含在PAM的第一个核苷酸的5'侧紧邻的20个碱基。例如,在包含5'-NNNNNNNNNNNNNNNNNNNNNRG-3' (SEQ ID NO:68,298)的序列中,靶核酸包含与N对应的序列,其中N是任意核苷酸,且标下划线的NRG序列是化脓性链球菌PAM。

[0305] 在某些实施方案中,与靶核酸杂交的间隔物序列具有至少约6个核苷酸(nt)的长度。所述间隔物序列可以是至少约6个核苷酸、至少约10个核苷酸、至少约15个核苷酸、至少约18个核苷酸、至少约19个核苷酸、至少约20个核苷酸、至少约25个核苷酸、至少约30个核苷酸、至少约35个核苷酸或至少约40个核苷酸、约6个核苷酸至约80个核苷酸、约6个核苷酸至约50个核苷酸、约6个核苷酸至约45个核苷酸、约6个核苷酸至约40个核苷酸、约6个核苷酸至约35个核苷酸、约6个核苷酸至约30个核苷酸、约6个核苷酸至约25个核苷酸、约6个核苷酸至约20个核苷酸、约6个核苷酸至约19个核苷酸、约10个核苷酸至约50个核苷酸、约10个核苷酸至约45个核苷酸、约10个核苷酸至约40个核苷酸、约10个核苷酸至约35个核苷酸、约10个核苷酸至约30个核苷酸、约10个核苷酸至约25个核苷酸、约10个核苷酸至约20个核苷酸、约10个核苷酸至约19个核苷酸、约19个核苷酸至约25个核苷酸、约19个核苷酸至约30个核苷酸、约19个核苷酸至约35个核苷酸、约19个核苷酸至约40个核苷酸、约19个核苷酸至约45个核苷酸、约19个核苷酸至约50个核苷酸、约19个核苷酸至约60个核苷酸、约20个核苷酸至约25个核苷酸、约20个核苷酸至约30个核苷酸、约20个核苷酸至约35个核苷酸、约20个核苷酸至约40个核苷酸、约20个核苷酸至约45个核苷酸、约20个核苷酸至约50个核苷酸、或约20个核苷酸至约60个核苷酸。在某些实施方案中,所述间隔物序列包含20个核苷酸。在某些实施方案中,所述间隔物包含19个核苷酸。

[0306] 在某些实施方案中,在所述间隔物序列和所述靶核酸之间的互补性百分比是至少约30%、至少约40%、至少约50%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约97%、至少约98%、至少约99%或100%。在某些实施

方案中，在所述间隔物序列和所述靶核酸之间的互补性百分比是至多约30%、至多约40%、至多约50%、至多约60%、至多约65%、至多约70%、至多约75%、至多约80%、至多约85%、至多约90%、至多约95%、至多约97%、至多约98%、至多约99%或100%。在某些实施方案中，在所述靶核酸的互补链的靶序列的最5'侧6个连续核苷酸上，在所述间隔物序列和所述靶核酸之间的互补性百分比是100%。在某些实施方案中，在约20个连续核苷酸上，在所述间隔物序列和所述靶核酸之间的互补性百分比是至少60%。所述间隔物序列和所述靶核酸的长度可以相差1-6个核苷酸，其可以被视作一个凸起或多个凸起。

[0307] 在某些实施方案中，使用计算机程序设计或选择间隔物序列。所述计算机程序可以使用变量，诸如预测的解链温度、二级结构形成、预测的退火温度、序列同一性、基因组背景、染色质可达性、%GC、基因组发生频率(例如，相同或类似的序列，但是由于错配、插入或缺失而在一个或多个点中变化)、甲基化状态、SNP的存在等。

[0308] 最小CRISPR重复序列

[0309] 在某些实施方案中，最小CRISPR重复序列是与参比CRISPR重复序列(例如，来自化脓性链球菌的crRNA)具有至少约30%、约40%、约50%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%或100%序列同一性的序列。

[0310] 最小CRISPR重复序列包含可以与细胞中的最小tracrRNA序列杂交的核苷酸。所述最小CRISPR重复序列和最小tracrRNA序列形成双链体，即碱基配对的双链结构。所述最小CRISPR重复序列和所述最小tracrRNA序列一起结合定位多肽。所述最小CRISPR重复序列的至少一部分与最小tracrRNA序列杂交。在某些实施方案中，所述最小CRISPR重复序列的至少一部分包含与最小tracrRNA序列的至少约30%、约40%、约50%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%或100%互补性。在某些实施方案中，所述最小CRISPR重复序列的至少一部分包含与最小tracrRNA序列的至多约30%、约40%、约50%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%或100%互补性。

[0311] 所述最小CRISPR重复序列可以具有约7个核苷酸至约100个核苷酸的长度。例如，所述最小CRISPR重复序列的长度是约7个核苷酸(nt)至约50个核苷酸、约7个核苷酸至约40个核苷酸、约7个核苷酸至约30个核苷酸、约7个核苷酸至约25个核苷酸、约7个核苷酸至约20个核苷酸、约7个核苷酸至约15个核苷酸、约8个核苷酸至约40个核苷酸、约8个核苷酸至约30个核苷酸、约8个核苷酸至约25个核苷酸、约8个核苷酸至约20个核苷酸、约8个核苷酸至约15个核苷酸、约15个核苷酸至约100个核苷酸、约15个核苷酸至约80个核苷酸、约15个核苷酸至约50个核苷酸、约15个核苷酸至约40个核苷酸、约15个核苷酸至约30个核苷酸、或约15个核苷酸至约25个核苷酸。在某些实施方案中，所述最小CRISPR重复序列的长度是大约9个核苷酸。在某些实施方案中，所述最小CRISPR重复序列的长度是大约12个核苷酸。

[0312] 在某些实施方案中，所述最小CRISPR重复序列与参比最小CRISPR重复序列(例如，来自化脓性链球菌的野生型crRNA)在至少6、7或8个连续核苷酸的区段上至少约60%相同。例如，所述最小CRISPR重复序列与参比最小CRISPR重复序列在至少6、7或8个连续核苷酸的区段上至少约65%相同、至少约70%相同、至少约75%相同、至少约80%相同、至少约85%相同、至少约90%相同、至少约95%相同、至少约98%相同、至少约99%相同或100%相同。

[0313] 最小tracrRNA序列

[0314] 在某些实施方案中，最小tracrRNA序列是与参比tracrRNA序列(例如，来自化脓性

链球菌的野生型tracrRNA)具有至少约30%、约40%、约50%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%或100%序列同一性的序列。

[0315] 最小tracrRNA序列包含与细胞中的最小CRISPR重复序列杂交的核苷酸。最小tracrRNA序列和最小CRISPR重复序列形成双链体,即碱基配对的双链结构。所述最小tracrRNA序列和所述最小CRISPR重复序列一起结合定位多肽。所述最小tracrRNA序列的至少一部分可以与所述最小CRISPR重复序列杂交。在某些实施方案中,所述最小tracrRNA序列与所述最小CRISPR重复序列至少约30%、约40%、约50%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%或100%互补。

[0316] 所述最小tracrRNA序列可以具有约7个核苷酸至约100个核苷酸的长度。例如,所述最小tracrRNA序列可以是约7个核苷酸(nt)至约50个核苷酸、约7个核苷酸至约40个核苷酸、约7个核苷酸至约30个核苷酸、约7个核苷酸至约25个核苷酸、约7个核苷酸至约20个核苷酸、约7个核苷酸至约15个核苷酸、约8个核苷酸至约40个核苷酸、约8个核苷酸至约30个核苷酸、约8个核苷酸至约25个核苷酸、约8个核苷酸至约20个核苷酸、约8个核苷酸至约15个核苷酸、约15个核苷酸至约100个核苷酸、约15个核苷酸至约80个核苷酸、约15个核苷酸至约50个核苷酸、约15个核苷酸至约40个核苷酸、约15个核苷酸至约30个核苷酸或约15个核苷酸至约25个核苷酸长。在某些实施方案中,所述最小tracrRNA序列的长度是大约9个核苷酸。在某些实施方案中,所述最小tracrRNA序列是大约12个核苷酸。在某些实施方案中,所述最小tracrRNA由在Jinek等人(出处同上)中描述的tracrRNA核苷酸23-48组成。

[0317] 在某些实施方案中,所述最小tracrRNA序列与参比最小tracrRNA (例如,来自化脓性链球菌的野生型tracrRNA)序列在至少6、7或8个连续核苷酸的区段上至少约60%相同。例如,所述最小tracrRNA序列与参比最小tracrRNA序列在至少6、7或8个连续核苷酸的区段上至少约65%相同、约70%相同、约75%相同、约80%相同、约85%相同、约90%相同、约95%相同、约98%相同、约99%相同或100%相同。

[0318] 在某些实施方案中,在最小CRISPR RNA和最小tracrRNA之间的双链体包含双螺旋。在某些实施方案中,在最小CRISPR RNA和最小tracrRNA之间的双链体包含至少约1、2、3、4、5、6、7、8、9或10个或更多个核苷酸。在某些实施方案中,在最小CRISPR RNA和最小tracrRNA之间的双链体包含至多约1、2、3、4、5、6、7、8、9或10个或更多个核苷酸。

[0319] 在某些实施方案中,所述双链体包含错配(即,所述双链体的两条链不是100%互补)。在某些实施方案中,所述双链体包含至少约1、2、3、4或5个或更多个错配。在某些实施方案中,所述双链体包含至多约1、2、3、4或5个或更多个错配。在某些实施方案中,所述双链体包含不超过2个错配。

[0320] 凸起

[0321] 在某些实施方案中,在所述最小CRISPR RNA和所述最小tracrRNA之间的双链体中存在"凸起"。所述凸起是所述双链体内的核苷酸的未配对区域。在某些实施方案中,所述凸起促进所述双链体与定位多肽的结合。凸起包含在双链体的一侧上的未配对的5'-XXXY-3'(其中X是任意嘌呤,且Y包含可以与相对链上的核苷酸形成摆动对的核苷酸)和在双链体的另一侧上的未配对的核苷酸区域。在双链体的两侧上的未配对的核苷酸的数目可以是不同的。

[0322] 在一个实例中,所述凸起包含在所述凸起的最小CRISPR重复序列链上的未配对的

嘌呤(例如,腺嘌呤)。在某些实施方案中,凸起包含所述凸起的最小tracrRNA序列链的未配对的5’-AAGY-3’,其中Y包含可以与最小CRISPR重复序列链上的核苷酸形成摆动配对的核苷酸。

[0323] 在某些实施方案中,在所述双链体的最小CRISPR重复序列侧上的凸起包含至少1、2、3、4或5个或更多个未配对的核苷酸。在某些实施方案中,在所述双链体的最小CRISPR重复序列侧上的凸起包含至多1、2、3、4或5个或更多个未配对的核苷酸。在某些实施方案中,在所述双链体的最小CRISPR重复序列侧上的凸起包含1个未配对的核苷酸。

[0324] 在某些实施方案中,在所述双链体的最小tracrRNA序列侧上的凸起包含至少1、2、3、4、5、6、7、8、9或10个或更多个未配对的核苷酸。在某些实施方案中,在所述双链体的最小tracrRNA序列侧上的凸起包含至多1、2、3、4、5、6、7、8、9或10个或更多个未配对的核苷酸。在某些实施方案中,在所述双链体的第二侧(例如,所述双链体的最小tracrRNA序列侧)上的凸起包含4个未配对的核苷酸。

[0325] 在某些实施方案中,凸起包含至少一个摆动配对。在某些实施方案中,凸起包含至多一个摆动配对。在某些实施方案中,凸起包含至少一个嘌呤核苷酸。在某些实施方案中,凸起包含至少3个嘌呤核苷酸。在某些实施方案中,凸起序列包含至少5个嘌呤核苷酸。在某些实施方案中,凸起序列包含至少一个鸟嘌呤核苷酸。在某些实施方案中,凸起序列包含至少一个腺嘌呤核苷酸。

[0326] 发夹

[0327] 在各种实施方案中,一个或多个发夹位于3’ tracrRNA序列中的最小tracrRNA的3’侧。

[0328] 在某些实施方案中,所述发夹开始于在最小CRISPR重复序列和最小tracrRNA序列双链体中的最后一个配对核苷酸的3’侧的至少约1、2、3、4、5、6、7、8、9、10、15或20个或更多个核苷酸。在某些实施方案中,所述发夹可以开始于最小CRISPR重复序列和最小tracrRNA序列双链体中的最后一个配对核苷酸的3’ 侧的至多约1、2、3、4、5、6、7、8、9或10个或更多个核苷酸。

[0329] 在某些实施方案中,发夹包含至少约1、2、3、4、5、6、7、8、9、10、15或20个或更多个连续核苷酸。在某些实施方案中,发夹包含至多约1、2、3、4、5、6、7、8、9、10、15个或更多个连续核苷酸。

[0330] 在某些实施方案中,发夹包含CC二核苷酸(即,两个连续胞嘧啶核苷酸)。

[0331] 在某些实施方案中,发夹包含形成双链体的核苷酸(例如,杂交在一起的在发夹中的核苷酸)。例如,发夹包含与3’ tracrRNA序列的发夹双链体中的GG二核苷酸杂交的CC二核苷酸。

[0332] 一个或多个发夹可以与定位多肽的引导RNA-相互作用区域相互作用。

[0333] 在某些实施方案中,存在两个或更多个发夹,且在某些实施方案中,存在三个或更多个发夹。

[0334] 3’ tracrRNA序列

[0335] 在某些实施方案中,3’ tracrRNA序列包含与参比tracrRNA序列(例如,来自化脓性链球菌的tracrRNA)具有至少约30%、约40%、约50%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%或100%序列同一性的序列。

[0336] 在某些实施方案中，所述3’tracrRNA序列具有约6个核苷酸至约100个核苷酸的长度。例如，所述3’tracrRNA序列可以具有约6个核苷酸(nt)至约50个核苷酸、约6个核苷酸至约40个核苷酸、约6个核苷酸至约30个核苷酸、约6个核苷酸至约25个核苷酸、约6个核苷酸至约20个核苷酸、约6个核苷酸至约15个核苷酸、约8个核苷酸至约40个核苷酸、约8个核苷酸至约30个核苷酸、约8个核苷酸至约25个核苷酸、约8个核苷酸至约20个核苷酸、约8个核苷酸至约15个核苷酸、约15个核苷酸至约100个核苷酸、约15个核苷酸至约80个核苷酸、约15个核苷酸至约50个核苷酸、约15个核苷酸至约40个核苷酸、约15个核苷酸至约30个核苷酸、或约15个核苷酸至约25个核苷酸的长度。在某些实施方案中，所述3’tracrRNA序列具有大约14个核苷酸的长度。

[0337] 在某些实施方案中，所述3’tracrRNA序列与参比3’tracrRNA序列(例如，来自化脓性链球菌的野生型3’tracrRNA序列)在至少6、7或8个连续核苷酸的区段上至少约60%相同。例如，所述3’tracrRNA序列与参比3’tracrRNA序列(例如，来自化脓性链球菌的野生型3’tracrRNA序列)在至少6、7或8个连续核苷酸的区段上至少约60%相同、约65%相同、约70%相同、约75%相同、约80%相同、约85%相同、约90%相同、约95%相同、约98%相同、约99%相同或100%相同。

[0338] 在某些实施方案中，3’tracrRNA序列包含超过一个形成双链体的区域(例如，发夹，杂交的区域)。在某些实施方案中，3’tracrRNA序列包含两个形成双链体的区域。

[0339] 在某些实施方案中，所述3’tracrRNA序列包含茎环结构。在某些实施方案中，所述3’tracrRNA中的茎环结构包含至少1、2、3、4、5、6、7、8、9、10、15或20个或更多个核苷酸。在某些实施方案中，所述3’tracrRNA中的茎环结构包含至多1、2、3、4、5、6、7、8、9或10个或更多个核苷酸。在某些实施方案中，所述茎环结构包含功能性部分。例如，所述茎环结构可包含适体、核酶、蛋白-相互作用发夹、CRISPR阵列、内含子或外显子。在某些实施方案中，所述茎环结构包含至少约1、2、3、4或5个或更多个功能性部分。在某些实施方案中，所述茎环结构包含至多约1、2、3、4或5个或更多个功能性部分。

[0340] 在某些实施方案中，所述3’tracrRNA序列中的发夹包含P-结构域。在某些实施方案中，所述P-结构域包含在发夹中的双链区域。

[0341] tracrRNA延伸序列

[0342] 可提供tracrRNA延伸序列，无论所述tracrRNA是在单分子引导还是双分子引导的背景下。在某些实施方案中，tracrRNA延伸序列具有约1个核苷酸至约400个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列具有超过1、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、120、140、160、180、200、220、240、260、280、300、320、340、360、380或400个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列具有约20至约5000个或更多个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列具有超过1000个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列具有小于1、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、120、140、160、180、200、220、240、260、280、300、320、340、360、380、400个或更多个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列可以具有小于1000个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列包含小于10个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列是10-30个核苷酸的长度。在某些实施方案中，tracrRNA延伸序列是30-70个核苷酸的长度。

[0343] 在某些实施方案中，所述tracrRNA延伸序列包含功能性部分(例如，稳定性控制序列、核酶、核糖核酸内切酶结合序列)。在某些实施方案中，所述功能性部分包含转录终止子区段(即，转录终止序列)。在某些实施方案中，所述功能性部分具有约10个核苷酸(nt)至约100个核苷酸、约10个核苷酸至约20个核苷酸、约20个核苷酸至约30个核苷酸、约30个核苷酸至约40个核苷酸、约40个核苷酸至约50个核苷酸、约50个核苷酸至约60个核苷酸、约60个核苷酸至约70个核苷酸、约70个核苷酸至约80个核苷酸、约80个核苷酸至约90个核苷酸或约90个核苷酸至约100个核苷酸、约15个核苷酸至约80个核苷酸、约15个核苷酸至约50个核苷酸、约15个核苷酸至约40个核苷酸、约15个核苷酸至约30个核苷酸或约15个核苷酸至约25个核苷酸的总长度。在某些实施方案中，所述功能性部分在真核细胞中起作用。在某些实施方案中，所述功能性部分在原核细胞中起作用。在某些实施方案中，所述功能性部分在真核和原核细胞二者中起作用。

[0344] 合适的tracrRNA延伸功能性部分的非限制性例子包括3’聚-腺苷酸化的尾巴、核糖开关序列(例如，以允许通过蛋白和蛋白复合物的经调节的稳定性和/或经调节的可达性)、形成dsRNA双链体(即，发夹)的序列、将RNA靶向至亚细胞位置(例如，细胞核、线粒体、叶绿体等)的序列、提供跟踪的修饰或序列(例如，直接缀合至荧光分子，缀合至促进荧光检测的部分，允许荧光检测的序列等)和/或提供蛋白(例如，在DNA上起作用的蛋白，包括转录活化剂、转录阻遏物、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰基转移酶、组蛋白脱乙酰基酶等)的结合位点的修饰或序列。在某些实施方案中，tracrRNA延伸序列包含引物结合位点或分子指标(例如，条形码序列)。在某些实施方案中，所述tracrRNA延伸序列包含一个或多个亲和标签。

[0345] 单分子引导接头序列

[0346] 在某些实施方案中，单分子引导核酸的接头序列具有约3个核苷酸至约100个核苷酸的长度。在Jinek等人(出处同上)中，例如，使用简单的4个核苷酸“四环”(-GAAA-)，Science, 337(6096):816-821 (2012)。示例性接头具有约3个核苷酸(nt)至约90个核苷酸、约3个核苷酸至约80个核苷酸、约3个核苷酸至约70个核苷酸、约3个核苷酸至约60个核苷酸、约3个核苷酸至约50个核苷酸、约3个核苷酸至约40个核苷酸、约3个核苷酸至约30个核苷酸、约3个核苷酸至约20个核苷酸、约3个核苷酸至约10个核苷酸的长度。例如，所述接头可以具有约3个核苷酸至约5个核苷酸、约5个核苷酸至约10个核苷酸、约10个核苷酸至约15个核苷酸、约15个核苷酸至约20个核苷酸、约20个核苷酸至约25个核苷酸、约25个核苷酸至约30个核苷酸、约30个核苷酸至约35个核苷酸、约35个核苷酸至约40个核苷酸、约40个核苷酸至约50个核苷酸、约50个核苷酸至约60个核苷酸、约60个核苷酸至约70个核苷酸、约70个核苷酸至约80个核苷酸、约80个核苷酸至约90个核苷酸或约90个核苷酸至约100个核苷酸的长度。在某些实施方案中，单分子引导核酸的接头是在4-40个核苷酸之间。在某些实施方案中，接头是至少约100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500或7000个或更多个核苷酸。在某些实施方案中，接头是至多约100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500或7000个或更多个核苷酸。

[0347] 接头可以包含许多种序列中的任一种，尽管优选地，所述接头将不包含与引导RNA的其它部分具有广泛同源性区域的序列，所述广泛同源性区域可能造成可干扰所述引导物的其它功能区域的分子内结合。在Jinek等人(出处同上)中，使用简单的4个核苷酸序列-

GAAA-, Science, 337(6096):816-821 (2012),但是同样可以使用众多其它序列,包括更长的序列。

[0348] 在某些实施方案中,所述接头序列包含功能性部分。例如,所述接头序列可包含一个或多个特征,包括适体、核酶、蛋白-相互作用发夹、蛋白结合位点、CRISPR阵列、内含子或外显子。在某些实施方案中,所述接头序列包含至少约1、2、3、4或5个或更多个功能性部分。在某些实施方案中,所述接头序列包含至多约1、2、3、4或5个或更多个功能性部分。

[0349] 通过插入、校正、缺失或替换所述基因内部或附近的一个或多个突变或外显子或通过将SERPINA1 cDNA或小基因敲入对应基因或安全港位点的基因座中以校正细胞的基因组工程策略

[0350] 本公开的方法可以牵涉校正所述突变体等位基因中的一个或两个。校正所述突变的基因编辑具有恢复正确表达水平和时序控制的优点。对患者的SERPINA1等位基因测序允许设计基因编辑策略以最佳地校正鉴别出的突变(一个或多个)。

[0351] 本发明的离体方法的步骤牵涉使用基因组工程编辑/校正患者特异性的iPS细胞。可选地,本发明的离体方法的步骤牵涉编辑/校正祖细胞、原代肝细胞、间充质干细胞、或肝祖细胞。同样地,本发明的体内方法的步骤牵涉使用基因组工程编辑/校正AATD患者中的细胞。类似地,本发明的细胞方法中的步骤牵涉通过基因组工程编辑/校正人细胞中的SERPINA1基因。

[0352] AATD患者表现出SERPINA1基因中的一个或多个突变。因此,不同的患者将通常需要不同的校正策略。任何CRISPR内切核酸酶可以用在本发明的方法中,每种CRISPR内切核酸酶具有它自身的有关的PAM,其可为或可不为疾病特异性的。例如,已经在SEQ ID No:54,860-61,324中鉴别出关于用来自化脓性链球菌的CRISPR/Cas9内切核酸酶靶向SERPINA1基因的gRNA间隔物序列。已经在SEQ ID No:61,325-61,936中鉴别出关于用来自金黄色葡萄球菌的CRISPR/Cas9内切核酸酶靶向SERPINA1基因的gRNA间隔物序列。已经在SEQ ID No:61,9347-62,069中鉴别出关于用来自嗜热链球菌的CRISPR/Cas9内切核酸酶靶向SERPINA1基因的gRNA间隔物序列。已经在SEQ ID No:62,070-62,120中鉴别出关于用来自齿垢密螺旋体的CRISPR/Cas9内切核酸酶靶向SERPINA1基因的gRNA间隔物序列。已经在SEQ ID No:62,121-62,563中鉴别出关于用来自脑膜炎奈瑟氏菌的CRISPR/Cas9内切核酸酶靶向SERPINA1基因的gRNA间隔物序列。已经在SEQ ID No:62,564-68,297中鉴别出关于用来自*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌的CRISPR/Cpf1内切核酸酶靶向SERPINA1基因的gRNA间隔物序列。已经在实施例7中鉴别出关于用来自化脓性链球菌的CRISPR/Cas9内切核酸酶靶向AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR(例如,靶向它们的外显子1-2)的gRNA间隔物序列。已经在实施例8中鉴别出关于用来自金黄色葡萄球菌的CRISPR/Cas9内切核酸酶靶向AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR(例如,靶向它们的外显子1-2)的gRNA间隔物序列。已经在实施例9中鉴别出关于用来自嗜热链球菌的CRISPR/Cas9内切核酸酶靶向AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR(例如,靶向它们的外显子1-2)的gRNA间隔物序列。已经在实施例10中鉴别出关于用来自齿垢密螺旋体的CRISPR/Cas9内切核酸酶靶向AAVS1 (PPP1R12C)、ALB、 Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR(例如,靶向它们的外显子1-2)的gRNA间隔物序列。已经在实施例11中鉴别出关于用来自脑膜炎奈瑟氏菌的CRISPR/Cas9内切核酸酶靶向AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR(例如,靶向它们的外显子1-2)的gRNA间隔物序列。已经在实施例12中鉴别出关于用来自*Acidominococcus*、毛螺菌科和新凶手弗朗西丝氏菌的CRISPR/Cas9内切核酸酶靶向AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR(例如,靶向它们的外显子1-2)的gRNA间隔物序列。

[0353] 例如,通过由不精确的NHEJ修复途径引起的插入或缺失,可以校正突变。如果患者的SERPINA1基因具有插入的或缺失的碱基,靶向切割可以导致恢复框架的NHEJ介导的插入或缺失。使用一种或多种引导RNA通过NHEJ介导的校正也可以校正错义突变。基于局部序列和微-同源性,可以设计或评价切口(一个或多个)的校正突变的能力或可能性。还可以使用NHEJ缺失基因的区段,无论是直接地还是通过改变剪接供体或受体位点(通过一种靶向几个位置的gRNA、或几种gRNA的切割)。如果氨基酸、结构域或外显子含有突变且可以被除去或倒置,或如果所述缺失以其它方式对所述蛋白恢复功能,这可能是有用的。引导链对已经用于缺失和校正倒位。

[0354] 可选地,用于通过HDR校正的供体含有经校正的序列,其具有小或大侧接同源性臂以允许退火。HDR基本上是在DSB修复过程中使用供给的同源DNA序列作为模板的无误差机制。同源性指导的修复(HDR)的比率是突变和切割位点之间的距离的函数,所以选择重叠或最近的靶位是重要的。模板可以包括被同源区域侧接的额外序列,或可以含有不同于基因组序列的序列,从而允许序列编辑。

[0355] 除了通过NHEJ或HDR校正突变以外,一系列其它选择是可能的。如果存在小或大缺失或多个突变,可以敲入含有受影响的外显子的cDNA。可以将全长cDNA敲入任何“安全港”,但是必须使用供给的或其它的启动子。如果将该构建体敲入正确位置,它将具有与正常基因类似的生理学控制。核酸酶对可以用于缺失突变的基因区域,尽管经常必须提供供体来恢复功能。在该情况下,将供给两种gRNA和一种供体序列。

[0356] 一些基因组工程策略牵涉校正SERPINA1基因中或附近的一个或多个突变,或缺失突变体SERPINA1 DNA和/或将SERPINA1 cDNA或小基因(其包含一个或多个外显子和内含子或天然的或合成的内含子)通过同源性指导的修复(HDR)(其也被称作同源重组(HR))敲入对应基因或安全港基因座的基因座中。同源性指导的修复是一种用于治疗患者的策略,所述患者具有在SERPINA1基因中或附近的失活突变。这些策略将恢复SERPINA1基因,并完全地逆转、治疗和/或减轻患病的状态。该策略将需要基于患者的失活突变(一个或多个)的位置的更加定制的方案。用于校正突变的供体核苷酸是小的(< 300碱基对)。这是有利的,因为HDR效率可能与供体分子的大小负相关。还预期供体模板可以配合进大小受限的病毒载体分子、例如腺伴随病毒(AAV)分子中,这已经显示为供体模板递送的有效方式。还预期供体模板可以配合进其它大小受限的分子(作为非限制性例子,包括血小板和/或外来体或其它微囊泡)中。

[0357] 同源性指导的修复是用于修复双链断裂(DSB)的细胞机制。最常见形式是同源重组。存在另外的用于HDR的途径,包括单链退火和可选HDR。基因组工程工具允许研究人员操

作细胞的同源重组途径以建立对基因组的位点特异性修饰。已经发现,细胞可以使用以反式提供的合成的供体分子修复双链断裂。因此,通过在特定突变附近引入双链断裂和提供合适的供体,可以在基因组中做出靶向变化。与$1/10^6$的接受单独同源供体的细胞的比率相比,特异性切割使HDR比率增加超过1,000倍。在特定核苷酸处的同源性指导的修复(HDR)的比率是到切割位点的距离的函数,所以选择重叠的或最近的靶位点是重要的。基因编辑提供超过基因添加的优点,因为原位校正使基因组的剩余部分不受扰乱。

[0358] 为通过HDR编辑供给的供体显著地变化,但是通常含有预期的序列,其具有小或大侧接同源性臂以允许与基因组DNA退火。侧接引入的基因变化的同源区域可以是30个碱基对或更小,或大至可以含有启动子、cDNA等的几千碱基盒。已经使用单链和双链寡核苷酸供体。这些寡核苷酸的大小范围为从小于100个核苷酸至超过许多kb,尽管还可以产生和使用更长的ssDNA。经常使用双链供体,包括PCR扩增子、质粒和小型环(mini-circles)。一般而言,已经发现,AAV载体是递送供体模板的非常有效的方式,尽管个别供体的包装限制是<5kb。供体的活性转录使HDR增加3倍,指示启动子的包含可增加转化。相反,供体的CpG甲基化减少基因表达和HDR。

[0359] 除了野生型内切核酸酶诸如Cas9以外,存在切口酶变体,其具有一个或其它被失活的核酸酶结构域,导致仅一条DNA链的切割。可以从个别Cas切口酶或使用侧接靶区域的切口酶对指导HDR。供体可以是单链的、切口的或dsDNA。

[0360] 可以以核酸酶或独立地通过多种不同的方法例如通过转染、纳米颗粒、显微注射或病毒转导供给供体DNA。已经提出一系列连接选择来增加供体对于HDR的可用性。例子包括将供体连接至核酸酶,连接至结合在附近的DNA结合蛋白,或连接至牵涉DNA末端结合或修复的蛋白。

[0361] 修复途径选择可以通过许多培养条件诸如影响细胞周期的那些或通过DNA修复和有关蛋白的靶向来引导。例如,为了增加HDR,可以抑制关键的NHEJ分子,诸如KU70、KU80或DNA连接酶IV。

[0362] 在没有供体存在时,使用几种非同源修复途径可以连接来自DNA断裂的末端或来自不同断裂的末端,其中在连接处几乎没有或没有碱基配对的情况下连接DNA末端。除了规范NHEJ以外,存在类似的修复机制,诸如alt-NHEJ。如果存在两个断裂,可以缺失或倒置间插区段。NHEJ修复途径可以导致在接头处的插入、缺失或突变。

[0363] 在核酸酶切割以后,使用NHEJ将15-kb可诱导的基因表达盒插入人细胞系中的确定基因座中。Maresca, M., Lin, V.G., Guo, N. & Yang, Y., *Genome Res* 23, 539-546 (2013)。

[0364] 除了通过NHEJ或HDR的基因组编辑以外,已经进行了使用NHEJ途径和HR的位点特异性的基因插入。组合方案在某些场合可能是适用的,可能包括内含子/外显子边框。NHEJ可证实对于内含子中的连接是有效的,而无误差HDR可更好地适于编码区。

[0365] 如以前所述,SERPINA1基因含有5个外显子。为了校正突变和恢复无活性的AAT蛋白,可修复5个外显子或附近的内含子中的任意一个或多个。可选地,存在各种与AATD有关的突变,它们是错义、无义、移码和其它突变的组合,具有失活AAT的共同作用。为了恢复无活性的AAT,可修复所述突变中的任意一个或多个。例如,可校正下述病理学变体中的一个或多个:rs764325655、rs121912713、rs28929474、rs17580、rs121912714、rs764220898、 rs199422211、rs751235320、rs199422210、rs267606950、rs55819880、rs28931570(参见表1)。这些变体包括缺失、插入和单核苷酸多态性。作为另一个替代方案，可将SERPINA1 cDNA或小基因(包含天然的或合成的增强子和启动子、一个或多个外显子、和天然的或合成的内含子、和天然的或合成的3'UTR和多腺苷酸化信号)敲入对应基因的基因座，或敲入安全港位点，诸如AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和/或TTR。所述安全港基因座可以选自AAVS1(PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2和TTR的外显子1-2。在某些实施方案中，所述方法提供了一个gRNA或一对gRNA，其可以用于促进来自多核苷酸供体模板的新序列的掺入，以校正一个或多个突变或敲入部分或整个SERPINA1基因或cDNA。

[0366] 表1

[0367]

变体	位置	变体类型
rs764325655	94,378,547 – 94,378,548	插入
rs121912713	94,378,561	单核苷酸变异
rs28929474	94,378,610	单核苷酸变异
rs17580	94,380,925	单核苷酸变异
rs121912714	94,380,949	单核苷酸变异
rs764220898	94,381,043	单核苷酸变异
rs199422211	94,381,067	单核苷酸变异
rs751235320	94,382,591	单核苷酸变异
rs199422210	94,382,686	单核苷酸变异
rs267606950	94,382,686	插入
rs55819880	94,383,008	单核苷酸变异
rs28931570	94,383,051	单核苷酸变异

[0368] 所述方法的一些实施方案提供了通过将基因切割2次而产生缺失的gRNA对，一个gRNA在一个或多个突变的5'末端处切割，且另一个gRNA在一个或多个突变的3'末端处切割，所述切割促进来自多核苷酸供体模板的新序列的插入以替换所述一个或多个突变，或缺失可不包括突变体氨基酸或邻近它的氨基酸(例如，未成熟终止密码子)，并导致功能性蛋白的表达或恢复可读框。所述切割可由各自在基因组中产生DSB的一对DNA内切核酸酶或一起在基因组中产生DSB的多个切口酶完成。

[0369] 可选地，所述方法的一些实施方案提供了一种gRNA以在一个或多个突变周围产生一个双链切口，其促进来自多核苷酸供体模板的新序列的插入以替换所述一个或多个突变。所述双链切口可由单一DNA内切核酸酶或一起在基因组中产生DSB的多个切口酶产生，或单一gRNA可导致缺失(MMEJ)，其可不包括突变体氨基酸(例如，未成熟终止密码子)，并导致功能性蛋白的表达或恢复可读框。

[0370] 在SERPINA1基因内的示例性修饰包括在上面提及的突变内或附近(近侧)的替换，诸如在特定突变的上游或下游小于3 kb、小于2kb、小于1 kb、小于0.5 kb的区域内。鉴于在

SERPINA1基因内的突变的相对宽泛的变异，应当理解，上面提及的替换的众多变化(包括但不限于更大以及更小的缺失)会预期导致AAT蛋白活性的恢复。

[0371] 这样的变体包括在5’和/或3’方向比特定目标突变更大、或在任一个方向更小的替换。因此，就特定替换而言的“附近”或“近侧”意指，与期望的替换边界(在本文中也被称作端点)有关的SSB或DSB基因座可在离指出的参照基因座小于约3 kb的区域内。在某些实施方案中，所述DSB基因座是在更近侧且在2 kb内，在1 kb内，在0.5 kb内，或在0.1 kb内。在小替换的情况下，期望的端点是在参照基因座处或“邻近”参照基因座，由此意指所述端点是在离参照基因座100 bp内，在50 bp内，在25 bp内，或小于约10 bp至5 bp。

[0372] 预期包含更大或更小替换的实施方案提供相同的益处，只要恢复AAT蛋白活性。因而预期，在本文中描述和举例说明的替换的许多变化对改善AATD将是有效的。

[0373] 另一种基因组工程策略牵涉外显子缺失。特定外显子的靶向缺失是用单一治疗混合物治疗大患者子集的有吸引力的策略。缺失可以是单外显子缺失或多外显子缺失。尽管多外显子缺失可以到达较大数目的患者，对于较大的缺失，缺失的效率随着增加的大小而极大地下降。因此，缺失范围可以是40-10,000碱基对(bp)大小。例如，缺失可为40-100、100-300、300-500、500-1,000、1,000-2,000、2,000-3,000、3,000-5,000或5,000-10,000碱基对大小的范围。

[0374] 缺失可以发生在增强子、启动子、第1个内含子和/或导致基因表达的上调的3’UTR，和/或通过调节元件的缺失。

[0375] 如前述，AAT基因包含5个外显子。所述5个外显子中的任一个或多个、或异常内含子剪接受体或供体位点可被缺失，以便恢复AAT读码框。在某些实施方案中，所述方法提供gRNA对，其可以用于缺失外显子1、2、3、4或5或它们的任何组合。

[0376] 为了确保前-mRNA在外显子缺失以后被适当地加工，可以缺失周围的剪接信号。剪接供体和受体通常是在邻近内含子的100碱基对内。因此，在某些实例中，方法可以提供就每个目标外显子/内含子连接部而言在大约±100-3100碱基对处切割的所有gRNA。

[0377] 对于任何基因组编辑策略，可以通过测序或PCR分析来证实基因编辑。

[0378] 靶序列选择

[0379] 相对于特定参比基因座在5’边界和/或3’边界的位置中的转移被用于促进或增强基因编辑的特定应用，这部分地取决于为所述编辑选择的内切核酸酶系统，如在本文中进一步描述和举例说明的。

[0380] 在这样的靶序列选择的第一个非限制性例子中，许多内切核酸酶系统具有指导初步选择潜在靶位点用于切割的规则或标准，诸如在CRISPR II型或V型内切核酸酶的情况下在邻近DNA切割位点的特定位置对PAM序列基序的要求。

[0381] 在靶序列选择或优化的另一个非限制性例子中，相对于在靶活性的频率评估关于靶序列和基因编辑内切核酸酶的特定组合的“靶外”活性频率(即在选择的靶序列以外的位点处发生的DSB的频率)。在某些情况下，已经在期望的基因座处正确地编辑的细胞可具有相对于其它细胞的选择性优点。选择性优点的示例性的但是非限制性的例子包括属性的获得，诸如增强的复制速率、持久性、对某些条件的抗性、增强的成功移植物植入的比率或引入患者中以后的体内持久性、和与维持有关的其它属性、或增加的这样的细胞的数目或生存力。在其它情况下，通过一种或多种用于鉴别、分类或以其它方式选择已经被正确地编辑

的细胞的筛选方法，可确定地选择已经在期望的基因座处正确地编辑的细胞。选择性优点和定向选择方法可利用与校正有关的表型。在某些实施方案中，可将细胞编辑两次或更多次以便建立第二个修饰，其建立用于选择或纯化预期的细胞群体的新表型。通过添加针对可选择的或可筛选的标志物的第二种gRNA，可以建立这样的第二个修饰。在某些情况下，使用含有cDNA以及也含有选择标记的DNA片段，可以在期望的基因座处正确地编辑细胞。

[0382] 不论任何选择性优点是否适用或任何定向选择是否应用于特定情况，通过靶外频率的考虑也指导靶序列选择，以便增强应用的有效性和/或减小在期望的靶标以外的位点处的不希望的改变的潜力。如在本文中和在本领域中进一步描述和举例说明的，靶外活性的发生受许多因素影响，包括靶位点和各种靶外位点之间的相似性和不同性、以及使用的特定内切核酸酶。生物信息学工具是可得到的，其辅助靶外活性的预测，且这样的工具经常还可以用于鉴别靶外活性的最可能位点，然后可以将其在实验场合中评估以评价靶外活性与在靶活性的相对频率，由此允许选择具有较高相对在靶活性的序列。这样的技术的示例性例子提供在本文中，且其它是本领域已知的。

[0383] 靶序列选择的另一个方面涉及同源重组事件。共享同源性区域的序列可以充当导致间插序列缺失的同源重组事件的焦点。这样的重组事件发生在染色体和其它DNA序列的正常复制过程中，并且也发生在当合成DNA序列时的其它时间，诸如在修复双链断裂(DSB)的情况下，所述双链断裂(DSB)在正常细胞复制周期中定期发生，但是也可通过各种事件(诸如紫外线和DNA断裂的其它诱导物)的发生或某些试剂(诸如各种化学诱导物)的存在而增强。许多这样的诱导物造成DSB在基因组中无差别地发生，并且DSB经常在正常细胞中被诱导和修复。在修复过程中，可以完全保真度重构原始序列，但是，在某些情况下，在DSB位点处引入小插入或缺失(被称作“插入/缺失”)。

[0384] 也可在特定位置特异性地诱导DSB，如在本文描述的内切核酸酶系统的情况下，其可以用于在选择的染色体位置处造成定向或优先基因修饰事件。在许多情况下可以利用同源序列在DNA修复(以及复制)的背景下接受重组的趋势，并且它是基因编辑系统诸如CRISPR的一种应用的基础，其中使用同源性指导的修复将通过使用“供体”多核苷酸提供的目标序列插入期望的染色体位置。

[0385] 还可以使用在特定序列之间的同源性区域(其可以是小的“微同源性”的区域，其可包含少至10个碱基对或更少)来实现期望的缺失。例如，将单一DSB引入在表现出与附近序列的微同源性的位点处。在这样的DSB的正常修复过程中，以高频率发生的结果是间插序列的缺失，其原因是DSB和伴随的细胞修复过程促进了重组。

[0386] 但是，在某些情况下，选择在同源性区域内的靶序列也可以引起大得多的缺失，包括基因融合(当缺失是在编码区中时)，其鉴于特定情况可为期望的或可不为期望的。

[0387] 本文提供的实例进一步举例说明了选择各种靶区域用于建立DSB，所述DSB被设计成诱导导致AAT蛋白活性恢复的替换，以及选择这样的区域内的特定靶序列，其被设计成相对于在靶事件使靶外事件最小化。

[0388] 核酸修饰

[0389] 在某些实施方案中，引入细胞中的多核苷酸包含一种或多种修饰，所述修饰可以个别地或组合地使用，例如，以增强活性、稳定性或特异性，改变递送，减少在宿主细胞中的先天性免疫应答，或用于其它增强，如在本文中进一步描述的和本领域已知的。

[0390] 在某些实施方案中，将经修饰的核苷酸用在CRISPR/Cas9/Cpf1系统中，在该情况下，可以修饰引导RNA (单分子引导或双分子引导)和/或编码引入细胞中的Cas或Cpf1内切核酸酶的DNA或RNA，如下面描述和举例说明的。这样的经修饰的核苷酸可以用在CRISPR/Cas9/Cpf1系统中以编辑任意一个或多个基因组基因座。

[0391] 为了这样的应用的非限制性举例说明的目的使用CRISPR/Cas9/Cpf1系统，可以使用引导RNA的修饰来增强包含引导RNA(其可为单分子引导或双分子)和Cas或Cpf1内切核酸酶的CRISPR/Cas9/Cpf1基因组编辑复合物的形成或稳定性。引导RNA的修饰也可以或可选地用于增强基因组编辑复合物与基因组中的靶序列之间的相互作用的起始、稳定性或动力学，这可以用于例如增强在靶活性。引导RNA的修饰也可以或可选地用于增强特异性，例如，在在靶位点处的基因组编辑相对于在其它(靶外)位点处的效应的相对比率。

[0392] 修饰也可以或可选地用于增加引导RNA的稳定性，例如，通过增加它对存在于细胞中的核糖核酸酶(RNA酶)的降解的抗性，由此造成它在所述细胞中的半衰期增加。在其中通过RNA(其需要翻译以便产生内切核酸酶)将Cas或Cpf1内切核酸酶引入要编辑的细胞中的实施方案中，增强引导RNA半衰期的修饰可以是特别有用的，因为增加与编码内切核酸酶的RNA同时引入的引导RNA的半衰期可以用于增加所述引导RNA和所述编码的Cas或Cpf1内切核酸酶在细胞中共存的时间。

[0393] 修饰也可以或可选地用于减小引入细胞中的RNA引起先天性免疫应答的可能性或程度。这样的应答(其已经在RNA干扰(RNAi)的背景下确定地表征，包括小干扰RNA(siRNA)，如在下面和在本领域中描述的)倾向于与减小的RNA半衰期和/或细胞因子或其它因子(其与免疫应答有关)的引起相关。

[0394] 还可以对引入细胞中的编码内切核酸酶的RNA做出一种或多种类型的修饰，包括但不限于，增强RNA的稳定性的修饰(诸如通过增加存在于细胞中的RNA酶对它的降解)，增强得到的产物(即内切核酸酶)的翻译的修饰，和/或减小引入细胞中的RNA引起先天性免疫应答的可能性或程度的修饰。

[0395] 同样可以使用修饰的组合，诸如前述和其它。在CRISPR/Cas9/Cpf1的情况下，例如，可以对引导RNA (包括上面举例说明的那些)做出一种或多种类型的修饰，和/或可以对编码Cas内切核酸酶的RNA (包括上面举例说明的那些)做出一种或多种类型的修饰。

[0396] 作为举例说明，在CRISPR/Cas9/Cpf1系统中使用的引导RNA或其它更小的RNA可以通过化学方式容易地合成，使许多修饰能容易地掺入，如在下面举例说明的和在本领域中描述的。尽管化学合成程序在继续增加，但是通过诸如高效液相色谱法(HPLC，其避免使用凝胶诸如PAGE)的程序对这样的RNA的纯化倾向于变得更有挑战性，因为多核苷酸长度显著地增加超过了100个左右核苷酸。用于产生较大长度的化学修饰的RNA的一个方案是生产连接在一起的两个或更多个分子。通过酶法更容易地产生长得多的RNA，诸如编码Cas9内切核酸酶的那些。尽管更少类型的修饰通常可用于用在酶促地生产的RNA中，仍然存在可以用于例如增强稳定性、减小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，如在下面和本领域中进一步描述的;并且新类型的修饰正在定期开发中。

[0397] 作为各种类型的修饰的例证，特别是经常与较小的化学合成的RNA使用的那些，修饰可以包含一个或多个在糖的2’位置处修饰的核苷酸，在某些实施方案中，为2’-O-烷基、2’-O-烷基-O-烷基或2’-氟-修饰的核苷酸。在某些实施方案中，RNA修饰包括在嘧啶核糖、

脱碱基残基、或位于RNA的3’末端处的倒置碱基上的2’-氟、2’-氨基或2’ O-甲基修饰。这样的修饰常规地掺入寡核苷酸中，且这些寡核苷酸已经显示具有比2’-脱氧寡核苷酸更高的对给定靶标的Tm (即，更高的靶标结合亲和力)。

[0398] 已经显示许多核苷酸和核苷修饰使得它们掺入其中的寡核苷酸比天然寡核苷酸更耐受核酸酶消化；这些经修饰的寡核苷酸比未修饰的寡核苷酸完整地存活更长的时间。经修饰的寡核苷酸的具体例子包括包含经修饰的主链(例如，硫代磷酸酯、磷酸三酯、膦酸甲酯、短链烷基或环烷基糖间键或短链杂原子或杂环糖间键)的那些。一些寡核苷酸是具有硫代磷酸酯主链的寡核苷酸和具有杂原子主链的那些，特别是CH_2-NH-O-CH_2、CH、~N(CH_3)~O~CH_2 (被称作亚甲基(甲基亚氨基)或MMI主链)、CH_2 --O--N (CH_3)-CH_2、CH_2 -N (CH_3)-N (CH_3)-CH_2和O-N(CH_3)-CH_2-CH_2主链，其中天然磷酸二酯主链被指为O- P-- O- CH)；酰胺主链[参见De Mesmaeker等人，Ace. Chem. Res.，28:366-374 (1995)]；吗啉代主链结构(参见Summerton和Weller，美国专利号5,034,506)；肽核酸(PNA)主链(其中寡核苷酸的磷酸二酯主链被聚酰胺主链替代，所述核苷酸直接地或间接地结合至聚酰胺主链的氮杂氮原子，参见Nielsen等人，Science 1991，254，1497)。含磷键包括但不限于硫代磷酸酯、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、甲基和其它烷基膦酸酯(包含3’亚烷基膦酸酯和手性膦酸酯)、次膦酸酯、氨基磷酸酯(包含3’-氨基氨基磷酸酯和氨基烷基氨基磷酸酯)、硫羰氨基磷酸酯、硫羰烷基膦酸酯、硫羰烷基磷酸三酯和具有正常3’-5’键的硼代磷酸酯、这些的2’-5’连接的类似物和具有倒置极性的那些，其中邻近的核苷单元对以3’-5’至5’-3’或2’-5’至5’-2’连接；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；5,476,925；5,519,126；5,536,821；5,541,306；5,550,111；5,563,253；5,571,799；5,587,361；和5,625,050。

[0399] 基于吗啉代的寡聚体化合物描述在Braasch和David Corey，Biochemistry，41 (14):4503-4510 (2002)；Genesis，第30卷，第3期，(2001)；Heasman，Dev. Biol.，243:209-214 (2002)；Nasevicius等人，Nat. Genet.，26:216-220 (2000)；Lacerra等人，Proc. Natl. Acad. Sci.，97:9591-9596 (2000)；和1991年7月23日授权的美国专利号5,034,506中。

[0400] 环已烯基核酸寡核苷酸模拟物描述在Wang等人，J. Am. Chem. Soc.，122：8595-8602 (2000)中。

[0401] 在其中不包括磷原子的经修饰的寡核苷酸主链具有通过短链烷基或环烷基核苷间键、混合的杂原子和烷基或环烷基核苷间键或一个或多个短链杂原子或杂环核苷间键形成的主链。这些包括具有吗啉代键(部分地从核苷的糖部分形成)的那些；硅氧烷主链；硫化物，亚砜和砜主链；甲乙酰基和硫代甲乙酰基主链；亚甲基甲乙酰基和硫代甲乙酰基主链；含有烯烃的主链；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；和具有混合的N、O、S和CH_2组成部分的其它主链；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;和5,677,439,它们中的每一篇通过引用并入本文。

[0402] 还可以包括一个或多个被取代的糖部分,例如,在2’位置处的以下之一:OH、SH、SCH_3、F、OCN、OCH_3 OCH_3、OCH_3 $O(CH_2)_n$ CH_3、$O(CH_2)_n$ NH_2或$O(CH_2)_n$ CH_3,其中n是从1至约10;C_1至C_{10}低级烷基、烷氧基烷氧基、被取代的低级烷基、烷芳基或芳烷基;Cl;Br;CN;CF_3;OCF_3;O-、S-或N-烷基;O-、S-或N-烯基;$SOCH_3$;SO_2 CH_3;ONO_2;NO_2;N_3;NH_2;杂环烷基;杂环烷芳基;氨基烷基氨基;聚烷基氨基;被取代的甲硅烷基;RNA切割基团;报告基团;嵌入剂;用于改善寡核苷酸的药代动力学性质的基团;或用于改善寡核苷酸的药效动力学性质的基团,和具有类似性质的其它取代基。在某些实施方案中,修饰包括2’-甲氧基乙氧基(2’-O-$CH_2CH_2OCH_3$,也被称作2’-O-(2-甲氧基乙基))(Martin等人, HeIv. Chim. Acta, 1995, 78, 486)。其它修饰包括2’-甲氧基(2’-O-CH_3)、2’-丙氧基(2’-OCH_2 CH_2CH_3)和2’-氟(2’-F)。还可以在寡核苷酸的其它位置、特别是3’末端核苷酸上的糖的3’位置和5’末端核苷酸的5’位置处做出类似的修饰。寡核苷酸也可具有糖模拟物,诸如替代呋喃型戊糖基的环丁基。

[0403] 在某些实施方案中,用新颖的基团替代核苷酸单元的糖和核苷间键,即主链。为了与适当的核酸目标化合物杂交维持碱基单位。一种这样的寡聚体化合物,即已显示具有优秀杂交性能的寡核苷酸模拟物,被称作肽核酸(PNA)。在PNA化合物中,用含有酰胺的主链例如氨基乙基甘氨酸主链替代寡核苷酸的糖-主链。保留核苷碱基,且将其直接地或间接地结合至主链的酰胺部分的氮杂氮原子。教导PNA化合物的制备的代表性美国专利包含但不限于美国专利号5,539,082、5,714,331和5,719,262。PNA化合物的其它教导可以参见Nielsen等人, Science, 254:1497-1500 (1991)。

[0404] 引导RNA还可以额外地或可选地包括核苷碱基(经常在本领域中简单地称作“碱基”)修饰或置换。本文中使用的“未修饰的”或“天然的”核苷碱基包括腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、胞嘧啶(C)和尿嘧啶(U)。经修饰的核苷碱基包括在天然核酸中仅稀少地或短暂地发现的核苷碱基,例如,次黄嘌呤,6-甲基腺嘌呤,5-Me嘧啶,特别是5-甲基胞嘧啶(也被称作5-甲基-2’ 脱氧胞嘧啶,且经常在本领域中称作5-Me-C),5-羟基甲基胞嘧啶(HMC),糖基HMC和龙胆二糖基HMC,以及合成的核苷碱基,例如,2-氨基腺嘌呤、2-(甲基氨基)腺嘌呤、2-(咪唑基烷基)腺嘌呤、2-(氨基烷基氨基)腺嘌呤或其它杂取代的烷基腺嘌呤、2-硫尿嘧啶、2-硫胸腺嘧啶、5-溴尿嘧啶、5-羟基甲基尿嘧啶、8-氮杂鸟嘌呤、7-脱氮鸟嘌呤、N6 (6-氨基己基)腺嘌呤和2,6-二氨基嘌呤。Kornberg, A., DNA Replication, W. H. Freeman & Co., San Francisco, 第75-77页(1980); Gebeyehu等人, Nucl. Acids Res. 15:4513 (1997)。还可以包括本领域已知的“通用”碱基,例如肌苷。已经显示5-Me-C置换使核酸双链体稳定性增加0.6-1.2℃(Sanghvi, Y. S., 见Crooke, S. T. 和Lebleu, B., 编, Antisense Research and Applications, CRC Press, Boca Raton, 1993, 第276-278页),并且是碱基置换的实施方案。

[0405] 经修饰的核苷碱基包含其它的合成的和天然的核苷碱基,诸如5-甲基胞嘧啶(5-me-C)、5-羟基甲基胞嘧啶、黄嘌呤、次黄嘌呤、2-氨基腺嘌呤、腺嘌呤和鸟嘌呤的6-甲基和其它烷基衍生物、腺嘌呤和鸟嘌呤的2-丙基和其它烷基衍生物、2-硫尿嘧啶、2-硫胸腺嘧啶和2-硫胞嘧啶、5-卤代尿嘧啶和胞嘧啶、5-丙炔基尿嘧啶和胞嘧啶、6-偶氮尿嘧啶、胞嘧啶

和胸腺嘧啶、5-尿嘧啶(假-尿嘧啶)、4-硫尿嘧啶、8-卤代、8-氨基、8-巯基、8-硫基烷基、8-羟基和其它8-取代的腺嘌呤和鸟嘌呤、5-卤代、特别是5-溴、5-三氟甲基和其它5-取代的尿嘧啶和胞嘧啶、7-甲基鸟嘌呤(quanine)和7-甲基腺嘌呤、8-氮杂鸟嘌呤和8-氮杂腺嘌呤、7-脱氮鸟嘌呤和7-脱氮腺嘌呤、和3-脱氮鸟嘌呤和3-脱氮腺嘌呤。

[0406] 核苷碱基进一步包含在美国专利号3,687,808中公开的那些,在'The Concise Encyclopedia of Polymer Science And Engineering',第858-859页,Kroschwitz,J.I.,编John Wiley & Sons,1990中公开的那些,由Englisch等人,Angewandle Chemie,International Edition',1991,30,第613页公开的那些,和由Sanghvi,Y.S.,第15章,Antisense Research and Applications',第289- 302页,Crooke,S.T.和Lebleu,B. ea.,CRC Press,1993公开的那些。这些核苷碱基中的某些对于增加本发明的寡聚体化合物的结合亲和力是特别有用的。这些包括5-取代的嘧啶、6-氮杂嘧啶和N-2、N-6和0-6取代的嘌呤,包含2-氨基丙基腺嘌呤、5-丙炔基尿嘧啶和5-丙炔基胞嘧啶。已经显示5-甲基胞嘧啶置换使核酸双链体稳定性增加0.6-1.2°C (Sanghvi,Y.S.,Crooke,S.T. 和Lebleu,B.,编,'Antisense Research and Applications',CRC Press,Boca Raton,1993,第276-278页),并且是碱基置换的实施方案,甚至更具体地当与2'-0-甲氧基乙基糖修饰组合时。经修饰的核苷碱基描述在美国专利号3,687,808,以及4,845,205;5,130,302;5,134,066;5,175,273;5,367,066;5,432,272;5,457,187;5,459,255;5,484,908;5,502,177;5,525,711;5,552,540;5,587,469;5,596,091;5,614,617;5,681,941;5,750,692;5,763,588;5,830,653;6,005,096;和美国专利申请公开2003/0158403中。

[0407] 因而,术语“经修饰的”指非天然的糖、磷酸酯或碱基,其掺入引导RNA、内切核酸酶、或引导RNA和内切核酸酶二者中。不一定将给定的寡核苷酸中的所有位置均匀地修饰,并且实际上,前述修饰中的超过一种可掺入单个寡核苷酸中,或甚至掺入寡核苷酸内的单个核苷中。

[0408] 在某些实施方案中,将引导RNA和/或编码内切核酸酶的mRNA (或DNA)化学连接至一个或多个增强寡核苷酸的活性、细胞分布或细胞摄取的部分或缀合物。这样的部分包含但不限于:脂质部分诸如胆固醇部分[Letsinger等人,Proc. Natl. Acad. Sci. USA,86:6553-6556 (1989)];胆酸[Manoharan等人,Bioorg. Med. Chem. Let.,4:1053-1060 (1994)];硫醚,例如,已基-S-三苯甲基硫醇[Manoharan等人,Ann. N. Y. Acad. Sci.,660:306-309 (1992)和Manoharan等人,Bioorg. Med. Chem. Let.,3:2765-2770 (1993)];硫代胆固醇[Oberhauser等人,Nucl. Acids Res.,20:533-538 (1992)];脂族链,例如,十二烷二醇或十一烷基残基[Kabanov等人,FEBS Lett.,259:327-330 (1990)和Svinarchuk等人,Biochimie,75:49- 54 (1993)];磷脂,例如,二-十六烷基-消旋-甘油或三乙基铵1,2-二-0-十六烷基-消旋-甘油-3-H-膦酸酯[Manoharan等人,Tetrahedron Lett.,36:3651-3654 (1995)和Shea等人,Nucl. Acids Res.,18:3777-3783 (1990)];多胺或聚乙二醇链[Mancharan等人,Nucleosides & Nucleotides,14:969-973 (1995)];金刚烷乙酸[Manoharan等人,Tetrahedron Lett.,36:3651-3654 (1995)];棕榈基部分[(Mishra等人,Biochim. Biophys. Acta,1264:229-237 (1995)];或十八烷基胺或已基氨基-羰基-t羟胆固醇部分[Crooke等人,J. Pharmacol. Exp. Ther.,277:

923-937 (1996)]。也参见美国专利号4,828,979；4,948,882；5,218,105；5,525,465；5,541,313；5,545,730；5,552,538；5,578,717，5,580,731；5,580,731；5,591,584；5,109,124；5,118,802；5,138,045；5,414,077；5,486,603；5,512,439；5,578,718；5,608,046；4,587,044；4,605,735；4,667,025；4,762,779；4,789,737；4,824,941；4,835,263；4,876,335；4,904,582；4,958,013；5,082,830；5,112,963；5,214,136；5,082,830；5,112,963；5,214,136；5,245,022；5,254,469；5,258,506；5,262,536；5,272,250；5,292,873；5,317,098；5,371,241，5,391,723；5,416,203，5,451,463；5,510,475；5,512,667；5,514,785；5,565,552；5,567,810；5,574,142；5,585,481；5,587,371；5,595,726；5,597,696；5,599,923；5,599，928和5,688,941。

[0409] 糖和其它部分可以用于将包含核苷酸的蛋白和复合物诸如阳离子多核糖体和脂质体靶向至特定位点。例如，肝细胞指导的转移可以由无唾液酸糖蛋白受体(ASGPR)介导；参见，例如，Hu，等人，Protein Pept Lett. 21(10):1025-30 (2014)。本领域已知的和不断开发的其它系统可以用于将在本情况下使用的生物分子和/或其复合物靶向至特定目标靶细胞。

[0410] 这些靶向部分或缀合物可以包括与官能团诸如伯或仲羟基共价地结合的缀合物基团。本发明的缀合物基团包括嵌入剂、报告分子、多胺、聚酰胺、聚乙二醇、聚醚、增强寡聚体的药效动力学性质的基团、和增强寡聚体的药代动力学性质的基团。典型的缀合物基团包括胆固醇、脂质、磷脂、生物素、吩嗪、叶酸盐、菲啶、蒽醌、吖啶、荧光素、罗丹明、香豆素和染料。在本发明的上下文中，增强药效动力学性质的基团包括改善摄取、增强对降解的抗性和/或强化与靶核酸的序列特异性杂交的基团。在本发明的上下文中，增强药代动力学性质的基团包括改善本发明的化合物的摄取、分布、代谢或排泄的基团。代表性的缀合物基团公开在1992年10月23日提交的国际专利申请号PCT/US92/09196和美国专利号6,287,860中，它们通过引用并入本文。缀合物部分包括但不限于：脂质部分诸如胆固醇部分、胆酸、硫醚例如己基-5-三苯甲基硫醇、硫代胆固醇、脂族链例如十二烷二醇或十一烷基残基、磷脂例如二-十六烷基-消旋-甘油或三乙基铵1,2-二-O-十六烷基-消旋-甘油-3-H-膦酸盐、多胺或聚乙二醇链、或金刚烷乙酸、棕榈基部分或十八烷基胺或己基氨基-羰基-氧基胆固醇部分。参见，例如，美国专利号4,828,979；4,948,882；5,218,105；5,525,465；5,541,313；5,545,730；5,552,538；5,578,717，5,580,731；5,580,731；5,591,584；5,109,124；5,118,802；5,138,045；5,414,077；5,486,603；5,512,439；5,578,718；5,608,046；4,587,044；4,605,735；4,667,025；4,762,779；4,789,737；4,824,941；4,835,263；4,876,335；4,904,582；4,958,013；5,082,830；5,112,963；5,214,136；5,082,830；5,112,963；5,214,136；5,245,022；5,254,469；5,258,506；5,262,536；5,272,250；5,292,873；5,317,098；5,371,241，5,391,723；5,416,203，5,451,463；5,510,475；5,512,667；5,514,785；5,565,552；5,567,810；5,574,142；5,585,481；5,587,371；5,595,726；5,597,696；5,599,923；5,599,928和5,688,941。

[0411] 还可以通过各种方式修饰不太容易化学合成且通常通过酶促合成来生产的较长多核苷酸。这样的修饰可以包括，例如，某些核苷酸类似物的引入，特定序列或其它部分在分子的5'或3'末端处的掺入，和其它修饰。作为举例说明，编码Cas9的mRNA是大约4 kb长度，且可以通过体外转录来合成。对mRNA的修饰可以用于，例如，增加它的翻译或稳定性(诸

如通过增加它对细胞降解的抗性)，或减小RNA引起先天性免疫应答的趋势，所述先天性免疫应答经常在引入外源RNA、特别是较长RNA(诸如编码Cas9的那种)以后在细胞中观察到。

[0412] 在本领域中已经描述了众多这样的修饰，诸如聚腺苷酸尾巴、5'帽类似物(例如，抗反向帽类似物(Anti Reverse Cap Analog,ARCA)或m7G(5')ppp(5')G (mCAP))、经修饰的5'或3' 非翻译区(UTR)、经修饰的碱基(诸如假-UTP、2-硫-UTP、5-甲基胞苷-5'-三磷酸(5-甲基-CTP)或N6-甲基-ATP)的应用、或除去5'末端磷酸酯的磷酸酶处理。这些和其它修饰是本领域已知的，且RNA的新修饰正在不断开发中。

[0413] 经修饰的RNA存在众多商业供应商，包括例如，TriLink Biotech、AxoLabs、Bio-Synthesis Inc.、Dharmacon和许多其它。如TriLink所述的，例如，5-甲基-CTP可以用于赋予合乎需要的特征，诸如增加的核酸酶稳定性、增加的翻译、或减少的先天性免疫受体与体外转录的RNA的相互作用。还已经显示5-甲基胞苷-5'-三磷酸(5-甲基-CTP)、N6-甲基-ATP、以及假-UTP和2-硫-UTP在培养物中和在体内减少先天性免疫刺激，同时增强翻译，如在下面提及的Kormann等人和Warren等人的出版物中所举例说明的。

[0414] 已经显示，在体内递送的化学修饰的mRNA可以用于实现改善的治疗效果；参见，*例如*，Kormann等人, Nature Biotechnology 29, 154-157 (2011)。这样的修饰可以用于，例如，增加RNA分子的稳定性和/或减小它的免疫原性。使用化学修饰诸如假-U、N6-甲基-A、2-硫-U和5-甲基-C，发现分别用2-硫-U和5-甲基-C取代仅1/4的尿苷和胞苷残基导致小鼠中toll-样受体(TLR)介导的mRNA识别的显著减少。通过减少先天性免疫系统的活化，这些修饰可以用于有效地增加体内mRNA的稳定性和寿命；参见，例如，Kormann等人，出处同上。

[0415] 还已经显示，掺入被设计成绕过先天性抗病毒应答的修饰的合成的信使RNA的重复施用可以将分化的人细胞重新程序化至多能性的。参见，例如，Warren, 等人, Cell Stem Cell, 7(5):618-30 (2010)。作为主要重新程序化蛋白起作用的这样的经修饰的mRNA可以是重新程序化多种人细胞类型的有效工具。这样的细胞被称作诱导的多能性干细胞(iPSC)，并且发现酶促地合成的掺入5-甲基-CTP、假-UTP和抗反向帽类似物(ARCA)的RNA可以用于有效地避免细胞的抗病毒应答；参见，例如，Warren等人，出处同上。

[0416] 在本领域中描述的多核苷酸的其它修饰包括，例如，聚腺苷酸尾巴的应用，5'帽类似物(诸如m7G(5')ppp(5')G (mCAP))的添加，5'或3' 非翻译区(UTR)的修饰，或除去5'末端磷酸酯的磷酸酶处理，且新方案正在不断开发中。

[0417] 结合RNA干扰(RNAi)(包括小干扰RNA (siRNA))的修饰已经开发了许多适用于产生用于本文中的用途的经修饰的RNA的组合物和技术。siRNA提出了特定体内挑战，因为它们通过mRNA干扰对基因沉默的影响通常是短暂的，这可以需要重复施用。另外，siRNA是双链RNA (dsRNA)，且哺乳动物细胞具有已经进化以检测和中和dsRNA(其经常是病毒感染的副产物)的免疫应答。因而，存在哺乳动物酶诸如PKR (dsRNA-响应性的激酶)和潜在地视黄酸-可诱导的基因I (RIG-I)(其可以介导对dsRNA的细胞应答)、以及Toll-样受体(诸如TLR3、TLR7和TLR8)(其可以响应于这样的分子触发细胞因子的诱导)；参见，例如，以下文献的综述：Angart等人, Pharmaceuticals (Basel) 6(4):440-468 (2013)；Kanasty等人, Molecular Therapy 20(3):513-524 (2012)；Burnett等人, Biotechnol J. 6(9):1130-46 (2011)；Judge和MacLachlan, Hum Gene Ther 19(2):111-24 (2008)；和其中引用的参考文献。

[0418] 已经开发了多种修饰,并用于增强RNA稳定性、减少先天性免疫应答和/或实现其它益处,所述其它益处关于多核苷酸向人细胞中的引入可以是有用的,如本文中所述的;参见,例如,以下文献的综述:Whitehead KA等人, Annual Review of Chemical and Biomolecular Engineering, 2:77-96 (2011); Gaglione和Messere, Mini Rev Med Chem, 10(7):578-95 (2010); Chernolovskaya等人, Curr Opin Mol Ther., 12(2):158-67 (2010); Deleavey等人, Curr Protoc Nucleic Acid Chem Chapter 16:Unit 16.3 (2009); Behlke, Oligonucleotides 18(4):305-19 (2008); Fucini等人, Nucleic Acid Ther 22(3):205-210 (2012); Bremsen等人, Front Genet 3:154 (2012)。

[0419] 如上面所指出的,经修饰的RNA存在许多商业供应商,其中的许多已经擅长于被设计成改善siRNA的有效性的修饰。基于在文献中报道的各种发现,提供了多种途径。例如,Dharmacon注意到,用硫对非桥连氧的替换(硫代磷酸酯, PS)已经被广泛地用于改善siRNA的核酸酶抗性,如Kole, Nature Reviews Drug Discovery 11:125-140 (2012)所报道的。已经报道核糖的2'-位置的修饰改善核苷酸间磷酸酯键的核酸酶抗性,同时增加双链体稳定性(Tm),这也已经显示提供不受免疫活化的保护。适度PS主链修饰与小的良好耐受的2'-取代(2'-O-甲基、2'-氟、2'-氢)的组合已经与用于体内应用的高度稳定的siRNA关联,如Soutschek等人. Nature 432:173-178 (2004)所报道的;并且已经报道2'-O-甲基修饰有效地改善稳定性,如Volkov, Oligonucleotides 19:191-202 (2009)所报道的。关于减少先天性免疫应答的诱导,已经报道用2'-O-甲基、2'-氟、2'-氢修饰特定序列减少TLR7/TLR8相互作用,同时通常保留沉默活性;参见,例如,Judge等人, Mol. Ther. 13:494-505 (2006);和Cekaite等人, J. Mol. Biol. 365:90-108 (2007)。还已经显示另外的修饰诸如2-硫尿嘧啶、假尿嘧啶、5-甲基胞嘧啶、5-甲基尿嘧啶和N6-甲基腺苷使由TLR3、TLR7和TLR8介导的免疫效应最小化;参见,例如,Kariko, K. 等人, Immunity 23:165-175 (2005)。

[0420] 也如本领域已知的和商购可得的,许多缀合物可以应用于本文所用的多核苷酸诸如RNA,其可以增强它们通过细胞的递送和/或摄取,包括例如,胆固醇、生育酚和叶酸、脂质、肽、聚合物、接头和适体;参见,例如,Winkler, Ther. Deliv. 4:791-809 (2013)的综述,和其中引用的参考文献。

[0421] 密码子优化

[0422] 在某些实施方案中,根据本领域的用于在含有目标靶DNA的细胞中表达的标准方法,将编码定位多肽的多核苷酸进行密码子优化。例如,如果预期的靶核酸是在人细胞中,考虑将人密码子优化的编码Cas9的多核苷酸用于生产Cas9多肽。

[0423] 靶向基因组的核酸和定位多肽的复合物

[0424] 靶向基因组的核酸与定位多肽(例如,核酸引导的核酸酶诸如Cas9)相互作用,由此形成复合物。靶向基因组的核酸将所述定位多肽引导至靶核酸。

[0425] RNP

[0426] 如以前所述,可将定位多肽和靶向基因组的核酸各自分别施用给细胞或患者。另一方面,可使定位多肽与一种或多种引导RNA、或一种或多种crRNA连同tracrRNA预先形成复合物。然后可将预先形成复合物的物质施用给细胞或患者。这样的预先形成复合物的物

质被称作核糖核蛋白颗粒(RNP)。

[0427] 编码系统组分的核酸

[0428] 在另一个方面,本公开提供了包含编码本公开的靶向基因组的核酸的核苷酸序列的核酸、本公开的定位多肽、和/或进行本公开的方法的实施方案所必需的任何核酸或蛋白性分子。

[0429] 在某些实施方案中,编码本公开的靶向基因组的核酸的核酸、本公开的定位多肽和/或进行本公开的方法的实施方案所必需的任何核酸或蛋白性分子包含载体(例如,重组表达载体)。

[0430] 术语"载体"指能够运输它已经连接的另一种核酸的核酸分子。一类载体是"质粒",其指可以在其中连接另外的核酸区段的圆形双链DNA环。另一类载体是病毒载体,其中可以将另外的核酸区段连接进病毒基因组中。某些载体能够在它们所引入的宿主细胞中自主复制(例如,具有细菌复制起点的细菌载体和附加型哺乳动物载体)。其它载体(例如,非-附加型哺乳动物载体)在引入宿主细胞中后整合进宿主细胞的基因组中,且由此与宿主基因组一起复制。

[0431] 在某些实施方案中,载体能够指导它们可操作地连接的核酸的表达。这样的载体在本文中被称作"重组表达载体",或更简单地称作"表达载体",其起等同功能。

[0432] 术语"可操作地连接的"是指,以允许表达核苷酸序列的方式,将目标核苷酸序列连接至调节序列(一个或多个)。术语"调节序列"意图包括,例如,启动子、增强子和其它表达控制元件(例如,多腺苷酸化信号)。这样的调节序列是本领域众所周知的,且描述在,例如,Goeddel; Gene Expression Technology:Methods in Enzymology 185, Academic Press, San Diego, CA (1990)中。调节序列包括在许多类型的宿主细胞中指导核苷酸序列的组成型表达的那些,和仅在某些宿主细胞中指导核苷酸序列的表达的那些(例如,组织特异性的调节序列)。本领域技术人员将理解,表达载体的设计可以取决于诸如靶细胞的选择、期望的表达水平等因素。

[0433] 考虑的表达载体包括但不限于基于痘苗病毒、脊髓灰质炎病毒、腺病毒、腺伴随病毒、SV40、单纯疱疹病毒、人免疫缺陷病毒、逆转录病毒(例如,鼠白血病病毒、脾坏死病毒和从逆转录病毒诸如劳斯肉瘤病毒、哈维肉瘤病毒、禽造白细胞组织增生病毒、慢病毒、人免疫缺陷病毒、骨髓增生性肉瘤病毒和乳腺肿瘤病毒衍生出的载体)的病毒载体和其它重组载体。为真核靶细胞考虑的其它载体包括但不限于载体pXT1、pSG5、pSVK3、pBPV、pMSG和pSVLSV40 (Pharmacia)。为真核靶细胞考虑的另外的载体包括但不限于载体pCTx-1、pCTx-2和pCTx-3,它们描述在图1A至1C中。可以使用其它载体,只要它们与宿主细胞相容。

[0434] 在某些实施方案中,载体包含一个或多个转录和/或翻译控制元件。取决于利用的宿主/载体系统,在表达载体中可使用许多合适的转录和翻译控制元件中的任一种,包括组成型和诱导型启动子、转录增强子元件、转录终止子等。在某些实施方案中,所述载体是失活病毒序列或CRISPR机制的组分或其它元件的自失活载体。

[0435] 合适的真核启动子(即,在真核细胞中有功能的启动子)的非限制性例子包括来自巨细胞病毒(CMV)立即早期、单纯疱疹病毒(HSV)胸苷激酶、早期和晚期SV40的那些,来自逆转录病毒的长末端重复序列(LTR),人延伸因子-1启动子(EF1),包含与鸡β-肌动蛋白启动子(CAG)融合的巨细胞病毒(CMV)增强子的杂合构建体,鼠干细胞病毒启动子(MSCV),磷酸

甘油酸激酶-1基因座启动子(PGK),和小鼠金属硫蛋白-I。

[0436] 为了表达小RNA,包括与Cas内切核酸酶关联使用的引导RNA,各种启动子诸如RNA聚合酶III启动子(包括例如U6和H1)可以是有利的。关于增强这样的启动子的应用的描述和参数是本领域已知的,且另外的信息和途径正在不断描述中;参见,例如,Ma, H.等人, *Molecular Therapy - Nucleic Acids 3*, e161 (2014) doi:10.1038/mtna.2014.12。

[0437] 所述表达载体还可含有用于翻译起始的核糖体结合位点和转录终止子。所述表达载体还可包括用于放大表达的适当序列。所述表达载体还可包括编码与定位多肽融合的非天然标签(例如,组氨酸标签、血凝素标签、绿色荧光蛋白等)的核苷酸序列,从而产生融合蛋白。

[0438] 在某些实施方案中,启动子是诱导型启动子(例如,热激启动子、四环素-调节型启动子、类固醇-调节型启动子、金属-调节型启动子、雌激素受体-调节型启动子等)。在某些实施方案中,启动子是组成型启动子(例如,CMV启动子、UBC启动子)。在某些实施方案中,所述启动子是空间上受限的和/或时间上受限的启动子(例如,组织特异性的启动子、细胞类型特异性的启动子等)。

[0439] 在某些实施方案中,将编码本公开的靶向基因组的核酸和/或定位多肽的核酸包装到递送媒介物内部或递送媒介物的表面上用于递送给细胞。考虑的递送媒介物包括但不限于纳米球、脂质体、量子点、纳米颗粒、聚乙二醇颗粒、水凝胶和胶束。如在本领域中描述的,多种靶向部分可以用于增强这样的媒介物与期望的细胞类型或位置的优先相互作用。

[0440] 本公开的复合物、多肽和核酸向细胞中的引入可以通过病毒或噬菌体感染、转染、缀合、原生质体融合、脂转染、电穿孔、核转染、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、颗粒枪技术、磷酸钙沉淀、直接显微注射、纳米颗粒介导的核酸递送等而发生。

[0441] 递送

[0442] 通过本领域已知的病毒或非病毒递送媒介物,可以递送引导RNA多核苷酸(RNA或DNA)和/或内切核酸酶多核苷酸(一种或多种)(RNA或DNA)。可选地,通过本领域已知的非病毒递送媒介物,诸如电穿孔或脂质纳米颗粒,可递送内切核酸酶多肽(一种或多种)。在某些实施方案中,所述DNA内切核酸酶可作为一种或多种多肽递送,无论是单独地还是与一种或多种引导RNA、或一种或多种crRNA连同tracrRNA预先形成复合物。

[0443] 可通过非病毒递送媒介物来递送多核苷酸,所述非病毒递送媒介物包括但不限于,纳米颗粒、脂质体、核糖核蛋白、带正电荷的肽、小分子RNA-缀合物、适体-RNA嵌合体和RNA-融合蛋白复合物。一些示例性的非病毒递送媒介物描述在Peer和Lieberman, Gene Therapy, 18:1127-1133 (2011)中(其聚焦于siRNA的非病毒递送媒介物,其也可用于递送其它多核苷酸)。

[0444] 通过脂质纳米颗粒(LNP),可将多核苷酸诸如引导RNA、sgRNA和编码内切核酸酶的mRNA递送给细胞或患者。

[0445] LNP指具有小于1000 nm、500 nm、250 nm、200 nm、150 nm、100 nm、75 nm、50 nm或25 nm的直径的任何颗粒。可选地,纳米颗粒可在1-1000 nm、1-500 nm、1-250 nm、25-200 nm、25-100 nm、35-75 nm或25-60 nm的大小范围内。

[0446] LNP可由阳离子脂质、阴离子脂质或中性脂质制成。可在LNP中包括中性脂质诸如

融原(fusogenic)磷脂DOPE或膜组分胆固醇作为'辅助脂质'来增强转染活性和纳米颗粒稳定性。阳离子脂质的限制包括由差稳定性和快速清除引起的低效力、以及炎症性应答或抗炎应答的产生。

[0447] LNP也可包含疏水脂质、亲水脂质、或疏水脂质和亲水脂质。

[0448] 本领域已知的任何脂质或脂质的组合可用于生产LNP。用于生产LNP的脂质的例子是:DOTMA、DOSPA、DOTAP、DMRIE、DC-胆固醇、DOTAP-胆固醇、GAP-DMORIE-DPyPE和GL67A-DOPE-DMPE-聚乙二醇(PEG)。阳离子脂质的例子是:98N12-5、C12-200、DLin-KC2-DMA(KC2)、DLin-MC3-DMA (MC3)、XTC、MD1和7C1。中性脂质的例子是:DPSC、DPPC、POPC、DOPE和SM。PEG修饰的脂质的例子是:PEG-DMG、PEG-CerC14和PEG-CerC20。

[0449] 所述脂质可以任意数目的摩尔比组合以生产LNP。另外,所述多核苷酸(一种或多种)可与脂质(一种或多种)以宽范围的摩尔比组合以生产LNP。

[0450] 如以前所述,可将所述定位多肽和靶向基因组的核酸各自分别施用给细胞或患者。另一方面,可将所述定位多肽与一种或多种引导RNA、或一种或多种crRNA连同tracrRNA预先形成复合物。然后可将所述预先形成复合物的物质施用给细胞或患者。这样的预先形成复合物的物质被称作核糖核蛋白颗粒(RNP)。

[0451] RNA能够与RNA或DNA形成特异性相互作用。尽管该性质被用在许多生物学过程中,它也伴有在富含核酸的细胞环境中杂乱相互作用的风险。该问题的一个解决方案是形成核糖核蛋白颗粒(RNP),其中将所述RNA与内切核酸酶预先形成复合物。所述RNP的另一个益处是保护所述RNA免于降解。

[0452] 所述RNP中的内切核酸酶可为经修饰的或未修饰的。同样地,所述gRNA、crRNA、tracrRNA或sgRNA可为经修饰的或未修饰的。众多修饰是本领域已知的且可使用。

[0453] 所述内切核酸酶和sgRNA通常以1:1摩尔比组合。可选地,所述内切核酸酶、crRNA和tracrRNA通常以1:1:1摩尔比组合。但是,宽范围的摩尔比可用于生产RNP。

[0454] 重组腺伴随病毒(AAV)载体可用于递送。生产rAAV颗粒的技术在本领域中是标准的,其中向细胞提供:要包装的AAV基因组,所述AAV基因组包括要递送的多核苷酸,rep和cap基因,和辅助病毒功能。rAAV的生产要求下述组分存在于单个细胞(在本文中称作包装细胞)内:rAAV基因组,与rAAV基因组分离(即,不在其里面)的AAV rep和cap基因,和辅助病毒功能。所述AAV rep和cap基因可来自重组病毒可以衍生出的任何AAV血清型,且可来自不同于rAAV基因组ITR的AAV血清型,包括但不限于,AAV血清型AAV-1、AAV-2、AAV-3、AAV-4、AAV-5、AAV-6、AAV-7、AAV-8、AAV-9、AAV-10、AAV-11、AAV-12、AAV-13和AAV rh.74。假型化的rAAV的生产公开在例如国际专利申请公开号WO 01/83692中。参见表2。

[0455] 表2

[0456]

AAV血清型	Genbank登记号
AAV-1	NC_002077.1
AAV-2	NC_001401.2
AAV-3	NC_001729.1
AAV-3B	AF028705.1
AAV-4	NC_001829.1
AAV-5	NC_006152.1

AAV-6	AF028704.1
AAV-7	NC_006260.1
AAV-8	NC_006261.1
AAV-9	AX753250.1
AAV-10	AY631965.1
AAV-11	AY631966.1
AAV-12	DQ813647.1
AAV-13	EU285562.1

[0457] 产生包装细胞的方法牵涉,建立稳定地表达用于AAV颗粒生产的所有必要组分的细胞系。例如,将一个质粒(或多个质粒)整合进细胞的基因组中,所述质粒包含缺乏AAV rep和cap基因的rAAV基因组、与rAAV基因组分离的AAV rep和cap基因、和选择标记诸如新霉素抗性基因。通过程序诸如GC加尾(Samulski等人, 1982, Proc. Natl. Acad. S6. USA, 79:2077-2081)、含有限制性内切核酸酶切割位点的合成接头的添加(Laughlin等人, 1983, Gene, 23:65-73)、或通过直接平端连接(Senapathy & Carter, 1984, J. Biol. Chem., 259:4661-4666),已经将AAV基因组引入细菌质粒中。然后用辅助病毒诸如腺病毒感染包装细胞系。该方法的优点是,所述细胞是可选择的,并且适合用于大规模生产rAAV。合适方法的其它例子采用腺病毒或杆状病毒而不是质粒将rAAV基因组和/或rep和cap基因引入包装细胞中。

[0458] rAAV生产的一般原理综述在例如Carter, 1992, Current Opinions in Biotechnology, 1533-539;和Muzyczka, 1992, Curr. Topics in Microbial. and Immunol., 158:97-129)中。各种途径描述在Ratschin等人, Mol. Cell. Biol. 4:2072 (1984); Hermonat等人, Proc. Natl. Acad. Sci. USA, 81:6466 (1984); Tratschin等人, Mol. Cell. Biol. 5:3251 (1985); McLaughlin等人, J. Virol., 62:1963 (1988);和Lebkowski等人, 1988 Mol. Cell. Biol., 7:349 (1988). Samulski等人(1989, J. Virol., 63:3822-3828);美国专利号5,173,414; WO 95/13365和对应的美国专利号5,658.776; WO 95/13392; WO 96/17947; PCT/US98/18600; WO 97/09441 (PCT/US96/14423);WO 97/08298 (PCT/US96/13872);WO 97/21825 (PCT/US96/20777);WO 97/06243 (PCT/FR96/01064);WO 99/11764; Perrin等人(1995) Vaccine 13:1244-1250; Paul等人(1993) Human Gene Therapy 4:609-615; Clark等人(1996) Gene Therapy 3:1124-1132;美国专利号5,786,211;美国专利号5,871,982;和美国专利号6,258,595中。

[0459] AAV载体血清型可以与靶细胞类型匹配。例如,可用尤其是指定的AAV血清型转导下述示例性的细胞类型。参见表3。

[0460] 表3

[0461]

组织/细胞类型	血清型
肝	AAV3, AAV5, AAV8, AAV9
骨骼肌	AAV1, AAV7, AAV6, AAV8, AAV9
中枢神经系统	AAV5, AAV1, AAV4
RPE	AAV5, AAV4
感光细胞	AAV5

肺	AAV9
心脏	AAV8
胰腺	AAV8
肾	AAV2，AAV8

[0462] 除了腺伴随病毒载体以外，可以使用其它病毒载体。这样的病毒载体包括但不限于慢病毒、甲病毒、肠道病毒、瘟病毒、杆状病毒、疱疹病毒、爱泼斯坦-巴尔病毒、乳多空病毒、痘病毒、痘苗病毒和单纯疱疹病毒。

[0463] 在某些实施方案中，将Cas9 mRNA、靶向SERPINA1基因中的一个或两个基因座的sgRNA和供体DNA各自分别配制进脂质纳米颗粒中，或者都共配制进一个脂质纳米颗粒中，或共配制进两个或多个脂质纳米颗粒中。

[0464] 在某些实施方案中，将Cas9 mRNA配制进脂质纳米颗粒中，而将sgRNA和供体DNA在AAV载体中递送。在某些实施方案中，将Cas9 mRNA和sgRNA共配制进脂质纳米颗粒中，而将供体DNA在AAV载体中递送。

[0465] 可得到将Cas9核酸酶作为DNA质粒、作为mRNA或作为蛋白来递送的选项。所述引导RNA可以从相同DNA表达，或也可以作为RNA递送。可以将所述RNA化学修饰以改变或提高它的半衰期，或减小免疫应答的可能性或程度。所述内切核酸酶蛋白可以在递送之前与gRNA形成复合物。病毒载体允许有效的递送；可以将Cas9的裂解形式和Cas9的更小直系同源物包装进AAV中，对于HDR的供体照样可以。还存在一系列可以递送这些组分中的每一种的非病毒递送方法，或者可以串联采用非病毒和病毒方法。例如，纳米颗粒可以用于递送蛋白和引导RNA，而AAV可以用于递送供体DNA。

[0466] 外来体

[0467] 外来体是一类以磷脂双层为边界的微囊泡，其可以用于将核酸递送至特定组织。体内许多不同类型的细胞天然地分泌外来体。当核内体内陷并形成多囊泡-核内体(MVE)时，外来体在细胞质内形成。当MVE与细胞膜融合时，所述外来体分泌进细胞外隙中。直径在30-120nm之间变化，外来体可以将各种分子以细胞-至-细胞通信的形式从一个细胞穿梭运输至另一个细胞。可以在遗传上改变天然地生产外来体的细胞诸如肥大细胞以生产具有靶向特定组织的表面蛋白的外来体，可选地可以将外来体从血流分离。可以用电穿孔将特定核酸放在经工程改造的外来体内。当全身性地引入时，所述外来体可以将核酸递送至特定靶组织。

[0468] 遗传修饰的细胞

[0469] 术语“遗传修饰的细胞”指包含至少一个通过基因组编辑(例如，使用CRISPR/Cas系统)引入的遗传修饰的细胞。在本文的某些离体实施方案中，所述遗传修饰的细胞是遗传修饰的祖细胞。在本文的某些体内实施方案中，所述遗传修饰的细胞是遗传修饰的肝细胞。在本文中考虑包含外源的靶向基因组的核酸和/或外源的编码靶向基因组的核酸的核酸的遗传修饰的细胞。

[0470] 术语“对照处理群体”描述了除了加入基因组编辑组分以外，已经用相同的介质、病毒诱导、核酸序列、温度、汇合、烧瓶大小、pH等处理过的细胞群体。本领域已知的任意方法可以用于测量SERPINA1基因或蛋白表达或活性的恢复，例如，AAT蛋白的蛋白质印迹分析或定量SERPINA1 mRNA。

[0471] 术语“分离的细胞”指已经从最初发现它的生物取出的细胞，或这样的细胞的后代。任选地，所述细胞已经在体外培养，例如，在确定的条件下，或在有其它细胞存在下。任选地，在以后将所述细胞引入第二种生物或重新引入从其中分离出它(或它来自其中的细胞)的生物。

[0472] 关于分离的细胞群体的术语“分离的群体”指已经从混合的或异质的细胞群体取出和分离的细胞群体。在某些实施方案中，与从其中分离或富集细胞的异质群体相比，分离的群体是基本上纯的细胞群体。在某些实施方案中，所述分离的群体是分离的人祖细胞群体，例如，与包含人祖细胞和人祖细胞来自其中的细胞的细胞的异质群体相比，基本上纯的人祖细胞群体。

[0473] 就特定细胞群体而言，术语“实质上增强的”指这样的细胞群体，其中特定类型的细胞的发生相对于预先存在的水平或参比水平增加了至少2倍、至少3-、至少4-、至少5-、至少6-、至少7-、至少8-、至少9、至少10-、至少20-、至少50-、至少100-、至少400-、至少1000-、至少5000-、至少20000-、至少100000-倍或更多倍，这取决于例如用于改善AATD的这样的细胞的期望水平。

[0474] 就特定细胞群体而言，术语“实质上富集的”指相对于构成总细胞群体的细胞为至少约10%、约20%、约30%、约40%、约50%、约60%、约70%或更多的细胞群体。

[0475] 就特定细胞群体而言，术语“实质上富集的”或“基本上纯的”指相对于构成总细胞群体的细胞至少约75%、至少约85%、至少约90%或至少约95%纯的细胞群体。也就是说，就祖细胞群体而言，术语“基本上纯的”或“基本上纯化的”指含有少于约20%、约15%、约10%、约9%、约8%、约7%、约6%、约5%、约4%、约3%、约2%、约1%或小于1%的不是由本文中的术语定义的祖细胞的细胞的细胞群体。

[0476] 编辑过基因组的iPSC分化成肝细胞

[0477] 本发明的离体方法的另一个步骤牵涉使编辑过基因组的iPSC分化成肝细胞。所述分化步骤可根据本领域已知的任意方法执行。例如，使用各种处理，包括激活素和B27补体(Life Technology)，使hiPSC分化成确定内胚层。使确定内胚层进一步分化成肝细胞，所述处理包括:FGF4、HGF、BMP2、BMP4、制癌蛋白M、Dexametason等(Duan等人, STEM CELLS; 2010;28:674-686, Ma等人, STEM CELLS TRANSLATIONAL MEDICINE *2013;2:409-419*)。

[0478] 编辑过基因组的间充质干细胞分化成肝细胞

[0479] 本发明的离体方法的另一个步骤牵涉使编辑过基因组的间充质干细胞分化成肝细胞。所述分化步骤可根据本领域已知的任意方法执行。例如，用包括胰岛素、转铁蛋白、FGF4、HGF、胆汁酸的各种因子和激素处理hMSC (Sawitza I等人, Sci Rep. 2015; 5:13320)。

[0480] 将细胞移植进患者中

[0481] 本发明的离体方法的另一个步骤牵涉将所述肝细胞移植进患者中。该移植步骤可使用本领域已知的任何植入方法完成。例如，可以将遗传修饰的细胞直接注射进患者的肝中或以其它方式施用给患者。

[0482] 本发明的离体方法的另一个步骤牵涉将所述祖细胞或原代肝细胞移植进患者中。该移植步骤可使用本领域已知的任何植入方法完成。例如，可将遗传修饰的细胞直接注射进患者的肝中或以其它方式施用给患者。使用选择的标志物，可离体纯化遗传修饰的细胞。

[0483] 药学上可接受的载体

[0484] 在本文中考虑的将祖细胞施用给对象的离体方法牵涉使用包含祖细胞的治疗组合物。

[0485] 治疗组合物含有生理学上可耐受的载体连同细胞组合物、和任选地至少一种作为活性成分溶解或分散在其中的如本文中所述的另外的生物活性剂。在某些实施方案中，当为了治疗目的而施用给哺乳动物或人患者时，所述治疗组合物不是实质上免疫原性的，除非期望如此。

[0486] 一般而言，将本文描述的祖细胞作为具有药学上可接受的载体的悬浮液施用。本领域技术人员会认识到，要用在细胞组合物中的药学上可接受的载体不会包括实质上干扰要递送给对象的细胞的生存力的量的缓冲剂、化合物、冷冻保存剂、防腐剂或其它试剂。包含细胞的制剂可以包括例如，允许维持细胞膜完整性的渗透缓冲剂，和任选地，在施用后维持细胞生存力或增强移植物植入的营养物。这样的制剂和悬浮液是本领域技术人员已知的，和/或可以适合使用例行实验与如本文中所述的祖细胞一起使用。

[0487] 还可以将细胞组合物乳化或呈现为脂质体组合物，前提条件是，所述乳化程序没有不利地影响细胞生存力。所述细胞和任意其它活性成分可以以适合用在本文描述的治疗方法中的量与药学上可接受的且与活性成分相容的赋形剂混合。

[0488] 在细胞组合物中包括的另外的试剂可以包括在其中的组分的药学上可接受的盐。药学上可接受的盐包括与以下酸形成的酸加成盐(与多肽的游离氨基形成)：无机酸诸如，例如，盐酸或磷酸，或这样的有机酸如乙酸、酒石酸、扁桃酸等。与游离羧基形成的盐还可以衍生自无机碱诸如，例如，钠、钾、铵、钙或铁的氢氧化物，和有机碱诸如异丙胺、三甲胺、2-乙基氨基乙醇、组氨酸、普鲁卡因等。

[0489] 生理学上可耐受的载体是本领域众所周知的。示例性的液体载体是不含有除了活性成分和水以外的物质的无菌水溶液，或含有缓冲剂诸如在生理pH值的磷酸钠、生理盐水或二者，诸如磷酸盐缓冲盐水。更进一步，水性载体可以含有超过一种缓冲盐，以及盐诸如氯化钠和氯化钾、右旋糖、聚乙二醇和其它溶质。液体组合物还可以含有除了水以外和排除水的液相。这样的另外的液相的示例是甘油、植物油诸如棉籽油、和水-油乳剂。在细胞组合物中使用的有效地治疗特定病症或病症的活性化合物的量将取决于病症或病症的性质，且可以通过标准的临床技术来确定。

[0490] 施用和效力

[0491] 在通过方法或途径将细胞例如祖细胞放入对象中的背景下，术语“施用”、“引入”和“移植”可互换使用，所述方法或途径导致引入的细胞在期望的部位、诸如损伤或修复的部位处的至少部分定位，从而产生期望的效应(一种或多种)。通过任意适当的途径可以施用细胞例如祖细胞或它们的分化的后代，所述途径导致递送至对象中的期望位置，在该处植入的细胞的至少一部分或所述细胞的组分保持存活。施用给对象后细胞的生存期间可以短至几小时，例如，24小时，至几天，至长达几年，或甚至患者的生存期，即，长期移植物植入。例如，在本文所述的某些实施方案中，经由全身性施用途径，诸如腹膜内或静脉内途径，施用有效量的肌源性祖细胞。

[0492] 术语“个体”、“对象”、“宿主”和“患者”在本文中可互换地使用，且指对其期望诊断、治疗或疗法的任何对象。在某些实施方案中，所述对象是哺乳动物。在某些实施方案中，

所述对象是人类。

[0493] 当预防性地提供时，在AATD的任何症状之前，例如，在肺气肿、气道阻塞、慢性支气管炎或肝疾病的发展之前，可以将本文描述的祖细胞施用给对象。因此，肝祖细胞群体的预防性施用用于预防AATD。

[0494] 当治疗性地提供时，在AATD的症状或指征的开始时(或之后)，例如，在肺或肝疾病的开始后，提供肝祖细胞。

[0495] 在本文所述的某些实施方案中，根据本文所述的方法施用的肝祖细胞群体包含得自一个或多个供体的同种异体肝祖细胞。“同种异体的”指得自相同物种的一个或多个不同供体(其中在一个或多个基因座处的基因不是相同的)的肝祖细胞或包含肝祖细胞的生物样品或生物样品。例如，施用给对象的肝祖细胞群体可以源自一个更无关的供体对象，或源自一个或多个不相同的同胞。在某些实施方案中，可以使用同基因的肝祖细胞群体，诸如得自遗传上相同的动物或得自单卵孪生的那些。在其它实施方案中，所述肝祖细胞是自体细胞；也就是说，所述肝祖细胞得自或分离自对象并施用给相同对象，即，供体和接受者是相同的。

[0496] 在一个实施方案中，术语“有效量”指预防或减轻AATD的至少一个或多个体征或症状所需的祖细胞或它们的后代的群体的量，且涉及提供期望的作用、例如治疗具有AATD的对象的组合物的足够量。术语“治疗有效量”因此指，当施用给典型对象、诸如具有AATD或处于AATD的风险中的对象时足以促进特定效应的祖细胞或包含祖细胞的组合物的量。有效量也将包括足以预防或延迟疾病的症状的发展、改变疾病的症状的进程(例如但不限于，减慢疾病的症状的进展)、或逆转疾病的症状的量。应当理解，对于任何给定的病例，适当的“有效量”可以由本领域普通技术人员使用例行实验确定。

[0497] 为了用在本文描述的各个实施方案中，祖细胞的有效量包含至少10^2个祖细胞、至少5×10^2个祖细胞、至少10^3个祖细胞、至少5×10^3个祖细胞、至少10^4个祖细胞、至少5×10^4个祖细胞、至少10^5个祖细胞、至少2×10^5个祖细胞、至少3×10^5个祖细胞、至少4×10^5个祖细胞、至少5×10^5个祖细胞、至少6×10^5个祖细胞、至少7×10^5个祖细胞、至少8×10^5个祖细胞、至少9×10^5个祖细胞、至少1×10^6个祖细胞、至少2×10^6个祖细胞、至少3×10^6个祖细胞、至少4×10^6个祖细胞、至少5×10^6个祖细胞、至少6×10^6个祖细胞、至少7×10^6个祖细胞、至少8×10^6个祖细胞、至少9×10^6个祖细胞或它们的倍数。所述祖细胞源自一个或多个供体，或得自自体来源。在本文所述的某些实施方案中，在施用给有此需要的对象之前，在培养物中扩增所述祖细胞。

[0498] 对于改善疾病的一种或多种症状、对于增加长期存活和/或对于减小与其它治疗有关的副作用，在具有AATD的患者的细胞中表达的功能性AAT的水平的适度和增量增加可以是有益的。在将这样的细胞施用给人患者后，生产增加的功能性AAT水平的肝祖先的存在是有益的。在某些实施方案中，对象的有效治疗导致相对于治疗的对象中的总AAT而言至少约3%、5%或7%的功能性AAT。在某些实施方案中，功能性AAT将是总AAT的至少约10%。在某些实施方案中，功能性AAT将是总AAT的至少约20%至30%。类似地，甚至相对有限的具有显著升高的功能性AAT水平的细胞亚群的引入可以在各种患者中是有益的，因为在某些情况下，正常化的细胞将具有相对于患病细胞的选择性优点。但是，对于改善患者中的AATD的一个或多个方面，甚至适中水平的具有升高的功能性AAT水平的肝祖先可以是有益的。在某些实施

方案中，在给其施用这样的细胞的患者中约10%、约20%、约30%、约40%、约50%、约60%、约70%、约80%、约90%或更多的肝祖先正在生产增加的水平的功能性AAT。

[0499] “施用”指通过方法或途径将祖细胞组合物递送进对象中，所述方法或途径导致所述细胞组合物在期望的部位处的至少部分定位。可以通过导致在对象中的有效治疗的任意适当途径施用细胞组合物，即施用导致向对象中的期望位置的递送，其中递送的组合物的至少一部分、即至少1×10^4个细胞被递送至期望部位一定的时间段。施用模式包括注射、输注、滴注或摄取。“注射”包括但不限于，静脉内、肌肉内、动脉内、鞘内、心室内、囊内、眶内、心内、真皮内、腹膜内、经气管、皮下、表皮下、关节内、囊下、蛛网膜下、椎管内、脑脊髓内和胸骨内注射和输注。在某些实施方案中，所述途径是静脉内的。对于细胞的递送，可以进行通过注射或输注的施用。

[0500] 在一个实施方案中，全身性地施用所述细胞。短语“全身施用”、“全身性地施用”、“周围施用”和“在周围施用”指除了直接进入靶位置、组织或器官以外的祖细胞群体的施用，使得它替代性地进入对象的循环系统，并因而接受代谢和其它类似的过程。

[0501] 熟练的临床医师可以确定包含用于治疗AATD的组合物的治疗的效力。但是，如果以有益的方式(例如，增加至少10%)改变功能性AAT的任意一种或所有体征或症状(但是作为一个例子，其水平)，或改进或改善疾病的其它临床上接受的症状或标志物，认为治疗是“有效的治疗”。通过个体的恶化失败也可以测量效力，如通过住院治疗或对医学干预的需要所评估的(例如，慢性阻塞性肺疾病、或疾病的进展被暂停或至少减慢)。本领域技术人员已知和/或本文描述了测量这些指标的方法。治疗包括个体或动物(一些非限制性例子包括人或哺乳动物)中的疾病的任何治疗，且包括：(1)抑制疾病，例如，阻止或减慢症状的进展；或(2)缓解疾病，例如，造成症状的消退；和(3)预防或减小症状发展的可能性。

[0502] 根据本发明的治疗通过增加个体中的功能性AAT的量来改善一种或多种与AATD有关的症状。通常与AATD有关的早期体征包括例如，肺气肿、气道阻塞、慢性支气管炎或肝脏疾病(阻塞性黄疸和在生命早期增加的转氨酶水平，和在缺乏儿童期间肝脏并发症史的情况下的纤维化和硬化)。

[0503] 试剂盒

[0504] 本公开提供了用于执行本发明的方法的试剂盒。试剂盒可以包括以下一种或多种：靶向基因组的核酸，编码靶向基因组的核酸的多核苷酸，定位多肽，编码定位多肽的多核苷酸，和/或执行本发明的方法的实施方案所需的任何核酸或蛋白性分子，或它们的任意组合。

[0505] 在某些实施方案中，试剂盒包含：(1)载体，其包含编码靶向基因组的核酸的核苷酸序列，和(2)定位多肽或载体，所述载体包含编码定位多肽的核苷酸序列，和(3)用于重构和/或稀释所述载体(一个或多个)和/或多肽的试剂。

[0506] 在某些实施方案中，试剂盒包含：(1)载体，其包含(i)编码靶向基因组的核酸的核苷酸序列，和(ii)编码定位多肽的核苷酸序列，和(2)用于重构和/或稀释所述载体的试剂。

[0507] 在以上试剂盒中的任一种的某些实施方案中，所述试剂盒包含靶向基因组的单分子引导核酸。在以上试剂盒中的任一种的某些实施方案中，所述试剂盒包含靶向基因组的双分子核酸。在以上试剂盒中的任一种的某些实施方案中，所述试剂盒包含两种或更多种双分子引导或单分子引导。在某些实施方案中，所述试剂盒包含编码靶向核酸的核酸的载

体。

[0508] 在以上试剂盒中的任一种的某些实施方案中，所述试剂盒可以进一步包含要插入以实现期望的遗传修饰的多核苷酸。

[0509] 试剂盒的组分可在分开的容器中，或被组合在单个容器中。

[0510] 在某些实施方案中，上述的试剂盒进一步包含一种或多种另外的试剂，其中这样的另外的试剂选自缓冲液、用于将多肽或多核苷酸引入细胞中的缓冲液、洗涤缓冲液、对照试剂、对照载体、对照RNA多核苷酸、用于从DNA体外生产多肽的试剂、用于测序的衔接子等。缓冲液可以是稳定缓冲液、重构缓冲液、稀释缓冲液等。在某些实施方案中，试剂盒还可以包括一种或多种组分，其可用于促进或增强内切核酸酶对DNA的在靶结合或切割，或改善靶向的特异性。

[0511] 除了上述组分以外，试剂盒还可以包括关于使用所述试剂盒的组分来实践所述方法的说明书。关于实践所述方法的说明书通常记录在合适的记录介质上。例如，所述说明书可印刷在基底诸如纸或塑料等上。所述说明书可作为包装插页呈现在试剂盒中，在所述试剂盒或其组分的容器的标签中(即，伴随所述包装或分包装)等。所述说明书可以作为电子存储数据文件呈现，所述电子存储数据文件存在于合适的计算机可读的存储介质例如CD-ROM、软盘、闪存等上。在某些情况下，实际的说明书没有存在于试剂盒中，但是可以提供关于从遥远来源得到说明书的方式(例如经由因特网)。该实施方案的一个例子是包括网址的试剂盒，在所述网址可以浏览说明书和/或从所述网址可以下载说明书。与说明书一样，用于得到说明书的该方式可以记录在合适的基底上。

[0512] 引导RNA制剂

[0513] 取决于特定施用模式和剂型，用药学上可接受的赋形剂诸如载体、溶剂、稳定剂、佐剂、稀释剂等配制本发明的引导RNA。通常配制引导RNA组合物以实现生理上相容的pH，和从约3的pH至约11的pH、约pH 3至约pH 7的范围，这取决于制剂和施用途径。在某些实施方案中，将pH调至约pH 5.0至约pH 8的范围。在某些实施方案中，所述组合物包含治疗有效量的至少一种如本文中所述的化合物、连同一种或多种药学上可接受的赋形剂。任选地，所述组合物包含本文所述化合物的组合，或可包括用于治疗或预防细菌生长的第二种活性成分(例如但不限于，抗细菌剂或抗微生物剂)，或可包括本发明的试剂的组合。

[0514] 合适的赋形剂包括，例如，载体分子，其包括大的、缓慢代谢的大分子，诸如蛋白、多糖、聚乳酸、聚乙醇酸、聚合的氨基酸、氨基酸共聚物和无活性的病毒颗粒。其它示例性的赋形剂包括抗氧化剂(例如但不限于，抗坏血酸)、螯合剂(例如但不限于，EDTA)、碳水化合物(例如但不限于，糊精、羟基烷基纤维素和羟基烷基甲基纤维素)、硬脂酸、液体(例如但不限于，油、水、盐水、甘油和乙醇)、润湿剂或乳化剂、pH缓冲物质等。

[0515] 其它可能的治疗途径

[0516] 使用经工程改造成靶向特定序列的核酸酶可以进行基因编辑。迄今为止，存在4大类核酸酶：大范围核酸酶和它们的衍生物、锌指核酸酶(ZFN)、转录活化剂如效应物核酸酶(TALEN)、和CRISPR-Cas9核酸酶系统。核酸酶平台在设计难度、靶向密度和作用模式上变化，特别是因为ZFN和TALEN的特异性是通过蛋白-DNA相互作用，而RNA-DNA相互作用主要引导Cas9。Cas9切割也需要邻近基序，即PAM，其在不同的CRISPR系统之间是不同的。来自化脓性链球菌(*Streptococcus pyogenes*)的Cas9使用NRG PAM切割，来自脑膜炎奈瑟氏菌

(Neisseria meningitidis)的CRISPR可以在具有PAM(包括NNNNGATT、NNNNNGTTT和NNNNGCTT)的位点处切割。许多其它的Cas9直系同源物靶向与可选PAM邻近的原间隔物。

[0517] CRISPR内切核酸酶诸如Cas9可以用在本发明的方法中。但是,本文描述的教导,诸如治疗性靶位点,可以应用于内切核酸酶的其它形式,诸如ZFN、TALEN、HE或MegaTAL,或使用核酸酶的组合。但是,为了将本发明的教导应用于这样的内切核酸酶,人们将尤其需要工程改造针对特定靶位点的蛋白。

[0518] 可以将另外的结合结构域与Cas9蛋白融合以增加特异性。这些构建体的靶位点将映射至(map to)鉴别出的gRNA指定的位点,但是将需要另外的结合基序,诸如针对锌指结构域。在Mega-TAL的情况下,可以将大范围核酸酶与TALE DNA-结合结构域融合。大范围核酸酶结构域可以增加特异性和提供切割。类似地,可以将失活的或死的Cas9 (dCas9)与切割结构域融合,并且需要sgRNA/Cas9靶位点和对于融合的DNA-结合结构域的邻近结合位点。这可能将需要除了催化失活以外的dCas9的某种蛋白质工程,以减少在没有另外的结合位点时的结合。

[0519] 锌指核酸酶

[0520] 锌指核酸酶(ZFN)是包含与II型内切核酸酶FokI的催化结构域连接的、经工程改造的锌指DNA结合结构域的模块蛋白。因为FokI仅作为二聚体起作用,必须将一对ZFN工程改造成结合相对DNA链上的同源靶"半位点"序列且在它们之间具有精确间距以使催化活性的FokI二聚体能够形成。在FokI结构域(其本身不具有序列特异性)的二聚化后,在ZFN半位点之间产生DNA双链断裂作为基因组编辑中的开始步骤。

[0521] 每一ZFN的DNA结合结构域通常包含丰富的Cys2-His2体系结构的3-6个锌指,每个指主要识别靶DNA序列的一条链上的核苷酸的三联体,尽管与第四个核苷酸的交叉链相互作用也可以是重要的。在与DNA发生关键接触的位置中的指的氨基酸的改变改变给定指的序列特异性。因而,四指锌指蛋白将选择性地识别12碱基对靶序列,其中所述靶序列是每个指促成的三联体偏好的复合物,尽管三联体偏好可以在不同程度上受邻近指影响。ZFN的一个重要方面是,简单地通过修饰各个指可以将它们容易地重新靶向至几乎任意基因组地址,尽管将其做好需要大量专家经验。在ZFN的大多数应用中,使用4-6个指的蛋白,分别识别12-18个碱基对。因此,一对ZFN将通常识别24-36个碱基对的组合靶序列,不包括半位点之间的5-7个碱基对间隔物。所述结合位点可以被包括15-17个碱基对的更大间隔物进一步隔开。该长度的靶序列可能是在人基因组中独有的,假定在设计过程中不包括重复序列或基因同系物。尽管如此,ZFN蛋白-DNA相互作用在它们的特异性方面不是绝对的,所以靶外结合和切割事件确实发生,无论是作为两个ZFN之间的异源二聚体,还是作为ZFN中的一个或另一个的同源二聚体。已经如下有效地消除了后一种可能性:工程改造FokI结构域的二聚化界面以建立"正"和"负"变体,也被称作专性异源二聚体变体,其可以仅彼此二聚化,且不与自身二聚化。强制专性异源二聚体阻止同源二聚体的形成。这已经极大地增强了ZFN以及采用这些FokI变体的任意其它核酸酶的特异性。

[0522] 在本领域中已经描述了多种基于ZFN的系统,不断报道了其修改,且众多参考文献描述了用于指导ZFN的设计的规则和参数;参见,例如,Segal等人,*Proc Natl Acad Sci USA 96*(6):2758-63 (1999);Dreier B等人,*J Mol Biol. 303*(4):489-502 (2000);Liu Q等人,*J Biol Chem. 277*(6):3850-6 (2002);Dreier等人,*J Biol Chem 280*(42):35588-

97 (2005);和Dreier等人,*J Biol Chem. 276*(31):29466-78 (2001)。

[0523] 转录活化剂-样效应物核酸酶(TALEN)

[0524] TALEN代表模块核酸酶的另一种形式,由此与ZFN一样,经工程改造的DNA结合结构域连接至FokI核酸酶结构域,且一对TALEN以串联方式运行以实现靶向的DNA裂解。与ZFN的主要差别是DNA结合结构域的性质和有关的靶DNA序列识别性能。TALEN DNA结合结构域源自TALE蛋白,其最初描述在植物细菌性病原体黄单胞菌属物种(*Xanthomonas sp.*)中。TALE包含33-35个氨基酸重复序列的串联阵列,每个重复序列识别通常多达20个碱基对长度的靶DNA序列中的单个碱基对,产生多达40个碱基对的总靶序列长度。通过仅仅包括在位置12和13处的2个氨基酸的重复可变二残基(RVD)确定每个重复序列的核苷酸特异性。碱基鸟嘌呤、腺嘌呤、胞嘧啶和胸腺嘧啶主要被四种RVD分别识别:Asn-Asn、Asn-Ile、His-Asp和Asn-Gly。这构成比对于锌指简单得多的识别代码,且因而代表就核酸酶设计而言超过后者的一个优点。尽管如此,与ZFN一样,TALEN的蛋白-DNA相互作用在它们的特异性方面不是绝对的,且TALEN还已经受益于FokI结构域的专性异源二聚体变体减小靶外活性的用途。

[0525] 已经建立了FokI结构域的另外变体,其在它们的催化功能方面被失活。如果TALEN或ZFN对的一半含有无活性的FokI结构域,那么仅单链DNA裂解(产生切口)将发生在靶位点,而不是DSB。所述结果与CRISPR/Cas9"切口酶"突变体的应用相当,在后者中所述Cas9切割结构域之一已经被失活。DNA切口可以用于驱动通过HDR的基因组编辑,但是以比使用DSB更低的效率。主要益处是靶外切口被快速地和准确地修复,这不同于易于NHEJ介导的错误修复的DSB。

[0526] 在本领域中已经描述了多种基于TALEN的系统,且不断报道了其修改;参见,例如,Boch, *Science 326*(5959):1509-12 (2009);Mak等人,*Science 335*(6069):716-9 (2012);和Moscou等人,*Science 326*(5959):1501 (2009)。多个小组已经描述了基于"Golden Gate"平台的TALEN或克隆方案的应用;参见,例如,Cermak等人,*Nucleic Acids Res. 39*(12):e82 (2011);Li等人,*Nucleic Acids Res. 39*(14):6315-25(2011);Weber等人,*PLoS One. 6*(2):e16765 (2011);Wang等人,*J Genet Genomics 41*(6):339-47, 2014年5月17日电子公开(2014);和Cermak T等人,*Methods Mol Biol. 1239*:133-59 (2015)。

[0527] 归巢内切核酸酶

[0528] 归巢内切核酸酶(HE)是具有长识别序列(14-44个碱基对)并以高特异性(经常在基因组所特有的位点)切割DNA的序列特异性的内切核酸酶。存在至少6个通过它们的结构分类的已知HE家族,包括LAGLIDADG (SEQ ID NO:68,299)、GIY-YIG、His-Cis box、H-N-H、PD-(D/E)xK和Vsr-样,它们源自宽范围的宿主,包括真核生物、原生生物、细菌、古细菌、蓝细菌和噬菌体。与ZFN和TALEN一样,HE可以用于在靶基因座处建立DSB作为基因组编辑中的初始步骤。另外,一些天然的和经工程改造的HE仅切割DNA的单链,由此作为位点特异性的切口酶起作用。HE的大靶序列和它们提供的特异性已经使它们成为有吸引力的建立位点特异性的DSB的候选物。

[0529] 在本领域中已经描述了多种基于HE的系统,且不断报道了其修改;参见,例如,Steentoft等人,*Glycobiology 24*(8):663-80 (2014); Belfort和Bonocora, *Methods Mol Biol. 1123*:1-26 (2014); Hafez和Hausner, *Genome 55*(8):553-69 (2012)的综述;和其中引用的参考文献。

[0530] MegaTAL/Tev-mTALEN/MegaTev

[0531] 作为杂合核酸酶的其它例子，MegaTAL平台和Tev-mTALEN平台使用TALE DNA结合结构域和催化活性的HE的融合体，利用TALE的可调节的DNA结合和特异性，以及HE的切割序列特异性；参见，例如，Boissel等人，*NAR 42*:2591-2601 (2014)；Kleinstiver等人，*G3 4*:1155-65 (2014)；和Boissel和Scharenberg，*Methods Mol. Biol. 1239*:171-96 (2015)。

[0532] 在另一种变异中，MegaTev体系结构是大范围核酸酶(Mega)与源自GIY-YIG归巢内切核酸酶I-TevI(Tev)的核酸酶结构域的融合体。所述两个活性部位在DNA底物上的位置间隔约30个碱基对，并产生具有不相容的粘性末端的两个DSB；参见，例如，Wolfs等人，*NAR 42*，8816-29 (2014)。预期现有的基于核酸酶的途径的其它组合将演化，并且可用于实现本文描述的靶向基因组修饰。

[0533] dCas9-FokI和其它核酸酶

[0534] 组合上述的核酸酶平台的结构和功能性质提供另一个基因组编辑途径，其可以潜在地克服一些固有缺陷。作为一个例子，CRISPR基因组编辑系统通常使用单一Cas9内切核酸酶来建立DSB。靶向的特异性由引导RNA中的20或22个核苷酸序列驱动，所述序列经历与靶DNA (在来自化脓性链球菌的Cas9的情况下，加上相邻NAG或NGG PAM序列中的另外2个碱基)的沃森-克里克碱基配对。这样的序列足够长以致在人基因组中为独特的，但是，RNA/DNA相互作用的特异性不是绝对的，有时耐受显著的混乱，特别是在靶序列的5’一半中，有效地减小驱动特异性的碱基的数目。其一个解决方案曾是将Cas9催化功能完全地失活——仅保留RNA引导的DNA结合功能——和替代性地将FokI结构域与失活的Cas9融合；参见，例如，Tsai等人，*Nature Biotech 32:*569-76 (2014)；和Guilinger等人，*Nature Biotech. 32:*577-82 (2014)。因为FokI必须二聚化才能变成催化活性的，需要两个引导RNA连接紧密靠近的两个Cas9-FokI融合体以形成二聚体和切割DNA。这基本上使组合的靶位点中的碱基数目翻倍，由此增加基于CRISPR的系统的靶向严谨性。

[0535] 作为其它实例，TALE DNA结合结构域与催化活性的HE诸如I-TevI的融合利用了TALE的可调节的DNA结合和特异性，以及I-TevI的切割序列特异性，例外是，可进一步减少靶外切割。

[0536] 本发明的方法和组合物

[0537] 因此，本公开具体地涉及以下非限制性的发明：在第一种方法(方法1)中，本公开提供了通过基因组编辑来编辑人细胞中的SERPINA1基因的方法，所述方法包括下述步骤：在所述人细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个单链断裂(SSB)或双链断裂(DSB)，其导致一个或多个突变或外显子的表达或功能的永久缺失、插入、或校正或调节，所述一个或多个突变或外显子在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或影响SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的表达或功能，和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0538] 在另一种方法(方法2)中，本公开提供了通过基因组编辑将SERPINA1基因插入人细胞中的方法，所述方法包括：在所述人细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶，以实现在安全港基因座内或附近的一个或多个单链断裂(SSB)或双链断裂(DSB)，其导致SERPINA1基因或小基因的永久插入，和导致AAT活性的恢复。

[0539] 在另一种方法(方法3)中,本公开提供了用于治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法,所述方法包括下述步骤:i)建立患者特异性的诱导性多能干细胞(iPSC);ii)在所述iPSC的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在所述iPSC的安全港基因座内或附近编辑;iii)使编辑过基因组的iPSC分化成肝细胞;和iv)将所述肝细胞移植进所述患者中。

[0540] 在另一种方法(方法4)中,本公开提供了方法3的方法,其中所述建立步骤包括:a)从所述患者分离体细胞;和b)将一组多能性相关基因引入所述体细胞中以诱导所述体细胞变成多能干细胞。

[0541] 在另一种方法(方法5)中,本公开提供了方法4的方法,其中所述体细胞是成纤维细胞。

[0542] 在另一种方法(方法6)中,本公开提供了方法4的方法,其中所述组的多能性相关基因是一种或多种选自以下的基因:OCT4、SOX2、KLF4、Lin28、NANOG和cMYC。

[0543] 在另一种方法(方法7)中,本公开提供了方法3-6中的任一种的方法,其中所述编辑步骤包括:在所述iPSC中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现一个或多个单链断裂(SSB)或双链断裂(DSB),所述一个或多个单链断裂(SSB)或双链断裂(DSB):在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近,其导致一个或多个突变或外显子的表达或功能的永久缺失、插入、或校正或调节,所述一个或多个突变或外显子在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或影响SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的表达或功能;或在安全港基因座内或附近,其导致SERPINA1或小基因的永久插入,和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0544] 在另一种方法(方法8)中,本公开提供了方法7的方法,其中所述安全港基因座选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。

[0545] 在另一种方法(方法9)中,本公开提供了方法3-8中的任一种的方法,其中所述分化步骤包括以下一个或多个以使编辑过基因组的iPSC分化成肝细胞:使所述编辑过基因组的iPSC与激活素、B27补体、FGF4、HGF、BMP2、BMP4、制癌蛋白M、Dexametason中的一种或多种接触。

[0546] 在另一种方法(方法10)中,本公开提供了方法3-9中的任一种的方法,其中所述移植步骤包括,通过移植、局部注射或全身输注或它们的组合将所述肝细胞移植进所述患者中。

[0547] 在另一种方法(方法11)中,本公开提供了用于治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法,所述方法包括下述步骤:i)执行所述患者的肝的活组织检查; ii)分离肝特异性的祖细胞或原代肝细胞; iii)在所述祖细胞或原代肝细胞的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近编辑,或在所述祖细胞或原代肝细胞的安全港基因座内或附近编辑;和iv)将所述编辑过基因组的祖细胞或原代肝细胞移植进所述患者中。

[0548] 在另一种方法(方法12)中,本公开提供了方法11的方法,其中所述分离步骤包括:用消化酶灌注新鲜肝组织、细胞差速离心、细胞培养和它们的组合。

[0549] 在另一种方法(方法13)中,本公开提供了方法11-12中的任一种的方法,其中所述编辑步骤包括:在所述祖细胞或原代肝细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现一个或多个单链断裂(SSB)或双链断裂(DSB),所述一个或多个单链断裂(SSB)或双链断裂(DSB):在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近,其导致一个或多个突变或外显子的表达或功能的永久缺失、插入、校正或调节,所述一个或多个突变或外显子在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或影响SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的表达或功能;或在安全港基因座内或附近,其导致SERPINA1基因或小基因的永久插入,和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0550] 在另一种方法(方法14)中,本公开提供了方法13的方法,其中所述安全港基因座选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。

[0551] 在另一种方法(方法15)中,本公开提供了方法11-14中的任一种的方法,其中所述移植步骤包括通过移植、局部注射或全身输注或它们的组合将编辑过基因组的祖细胞或原代肝细胞移植进所述患者中。

[0552] 在另一种方法(方法16)中,本公开提供了用于治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的离体方法,所述方法包括下述步骤:i)执行患者骨髓的活组织检查; ii)从所述患者分离间充质干细胞; iii)在所述间充质干细胞的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近编辑,或在所述间充质干细胞的安全港基因座内或附近编辑; iv)使所述编辑过基因组的间充质干细胞分化成肝细胞;和v)将所述肝细胞移植进所述患者中。

[0553] 在另一种方法(方法17)中,本公开提供了方法16的方法,其中所述间充质干细胞分离自患者骨髓或外周血。

[0554] 在另一种方法(方法18)中,本公开提供了方法16的方法,其中所述分离步骤包括:抽吸骨髓并使用Percoll™通过密度离心分离间充质细胞。

[0555] 在另一种方法(方法19)中,本公开提供了方法16-18中的任一种的方法,其中所述编辑步骤包括:向所述间充质干细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现一个或多个单链断裂(SSB)或双链断裂(DSB),所述一个或多个单链断裂(SSB)或双链断裂(DSB):在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近,其导致一个或多个突变或外显子的表达或功能的永久缺失、插入、校正或调节,所述一个或多个突变或外显子在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或影响SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的表达或功能;或在安全港基因座内或附近,其导致SERPINA1基因或小基因的永久插入,和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0556] 在另一种方法(方法20)中,本公开提供了方法19的方法,其中所述安全港基因座选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。

[0557] 在另一种方法(方法21)中,本公开提供了方法16-19中的任一种的方法,其中所述分化步骤包括以下一个或多个以使编辑过基因组的干细胞分化成肝细胞:使所述编辑过基

因组的间充质干细胞与胰岛素、转铁蛋白、FGF4、HGF或胆汁酸中的一种或多种接触。

[0558] 在另一种方法(方法22)中,本公开提供了方法16-21中的任一种的方法,其中所述移植步骤包括,通过移植、局部注射或全身输注或它们的组合将所述肝细胞移植进所述患者中。

[0559] 在另一种方法(方法23)中,本公开提供了用于治疗具有α1抗胰蛋白酶缺乏(AATD)的患者的体内方法,所述方法包括下述步骤:编辑所述患者的细胞中的SERPINA1基因,或编码SERPINA1基因的调节元件的其它DNA序列,或在所述患者的细胞的安全港基因座内或附近编辑。

[0560] 在另一种方法(方法24)中,本公开提供了方法23的方法,其中所述编辑步骤包括:在所述细胞中引入一种或多种脱氧核糖核酸(DNA)内切核酸酶,以实现一个或多个单链断裂(SSB)或双链断裂(DSB),所述一个或多个单链断裂(SSB)或双链断裂(DSB):在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近,其导致在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近的一个或多个突变或外显子的永久缺失、插入、或校正或调节;或在安全港基因座内或附近,其导致SERPINA1基因或小基因的永久插入,和α1抗胰蛋白酶(AAT)蛋白活性的恢复。

[0561] 在另一种方法(方法25)中,本公开提供了方法1、2、7、13、19或24中的任一种的方法,其中所述一种或多种DNA内切核酸酶是Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8、Cas9 (也被称作Csn1和Csx12)、Cas100、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4或Cpf1内切核酸酶;或其同系物,天然存在的分子的重组,其经过密码子优化的或经修饰的形式,和它们的组合。

[0562] 在另一种方法(方法26)中,本公开提供了方法25的方法,其中所述方法包括在所述细胞中引入编码一种或多种DNA内切核酸酶的一种或多种多核苷酸。

[0563] 在另一种方法(方法27)中,本公开提供了方法25的方法,其中所述方法包括在所述细胞中引入编码一种或多种DNA内切核酸酶的一种或多种核糖核酸(RNA)。

[0564] 在另一种方法(方法28)中,本公开提供了方法26或27中的任一种的方法,其中所述一种或多种多核苷酸或一种或多种RNA是一种或多种经修饰的多核苷酸或一种或多种经修饰的RNA,本公开提供了方法。

[0565] 在另一种方法(方法29)中,本公开提供了方法26的方法,其中所述DNA内切核酸酶是蛋白或多肽。

[0566] 在另一种方法(方法30)中,本公开提供了上述方法中的任一种的方法,其中所述方法还包括在所述细胞中引入一种或多种引导核糖核酸(gRNA)。

[0567] 在另一种方法(方法31)中,本公开提供了方法30的方法,其中所述一种或多种gRNA是单分子引导RNA (sgRNA)。

[0568] 在另一种方法(方法32)中,本公开提供了方法30-31中的任一种的方法,其中所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA或一种或多种经修饰的sgRNA。

[0569] 在另一种方法(方法33)中,本公开提供了方法30-32中的任一种的方法,其中所述

一种或多种DNA内切核酸酶与一种或多种gRNA或一种或多种sgRNA预先形成复合物。

[0570] 在另一种方法(方法34)中，本公开提供了上述方法中的任一种的方法，其中所述方法还包括在所述细胞中引入包含野生型SERPINA1基因的至少一部分或小基因或cDNA的多核苷酸供体模板。

[0571] 在另一种方法(方法35)中，本公开提供了方法34的方法，其中所述野生型SERPINA1基因的至少一部分或小基因或cDNA是外显子1、外显子2、外显子3、外显子4、外显子5、内含子区域、它们的片段或组合、整个SERPINA1基因、编码SERPINA1基因的野生型调节元件的DNA序列、小基因或cDNA。

[0572] 在另一种方法(方法36)中，本公开提供了方法34-35中的任一种的方法，其中所述供体模板是单链或双链多核苷酸。

[0573] 在另一种方法(方法37)中，本公开提供了方法34-36中的任一种的方法，其中所述供体模板具有与14q32.13区域同源的臂。

[0574] 在另一种方法(方法38)中，本公开提供了方法1、2、7、13、19或24中的任一种的方法，其中所述方法还包括在所述细胞中引入一种引导核糖核酸(gRNA)和包含野生型SERPINA1基因的至少一部分的多核苷酸供体模板，且其中所述一种或多种DNA内切核酸酶是一种或多种Cas9内切核酸酶，其实现在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在安全港基因座内或附近在DSB基因座处的一个双链断裂(DSB)，所述断裂促进来自所述多核苷酸供体模板的新序列插入在所述基因座或安全港基因座处的染色体DNA中，所述插入导致在所述基因座或安全港基因座的近侧的SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列的染色体DNA的一部分的永久插入或校正，和AAT蛋白活性的恢复，且其中所述gRNA包含与所述基因座或安全港基因座的区段互补的间隔物序列。

[0575] 在另一种方法(方法39)中，本公开提供了方法38的方法，其中近侧是指在所述基因座或安全港基因座的上游和下游的核苷酸。

[0576] 在另一种方法(方法40)中，本公开提供了方法1、2、7、13、19或24中的任一种的方法，其中所述方法还包括在所述细胞中引入两种引导核糖核酸(gRNA)和包含野生型SERPINA1基因的至少一部分的多核苷酸供体模板，且其中所述一种或多种DNA内切核酸酶是两种或更多种Cas9内切核酸酶，其实现在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在安全港基因座内或附近的一对双链断裂(DSB)，第一个是在5' DSB基因座且第二个是在3' DSB基因座，所述断裂促进来自所述多核苷酸供体模板的新序列在5' DSB基因座和3' DSB基因座之间插入染色体DNA中，所述插入导致在SERPINA1基因或编码SERPINA1基因的调节元件的其它DNA序列内或附近、或在安全港基因座内或附近在5' DSB基因座和3' DSB基因座之间的染色体DNA的永久插入或校正，和AAT蛋白活性的恢复，且其中所述第一引导RNA包含与所述5' DSB基因座的区段互补的间隔物序列，且所述第二引导RNA包含与所述3' DSB基因座的区段互补的间隔物序列。

[0577] 在另一种方法(方法41)中，本公开提供了方法38-40中的任一种的方法，其中所述一种或两种gRNA是一种或两种单分子引导RNA (sgRNA)。

[0578] 在另一种方法(方法42)中，本公开提供了方法38-41中的任一种的方法，其中所述一种或两种gRNA或一种或两种sgRNA是一种或两种经修饰的gRNA或一种或两种经修饰的

sgRNA。

[0579] 在另一种方法(方法43)中，本公开提供了方法38-42中的任一种的方法，其中所述一种或多种DNA内切核酸酶与一种或两种gRNA或一种或两种sgRNA预先形成复合物。

[0580] 在另一种方法(方法44)中，本公开提供了方法38-43中的任一种的方法，其中所述野生型SERPINA1基因的至少一部分或小基因或cDNA是外显子1、外显子2、外显子3、外显子4、外显子5、内含子区域、它们的片段或组合、整个SERPINA1基因、编码SERPINA1基因的野生型调节元件的DNA序列、小基因或cDNA。

[0581] 在另一种方法(方法45)中，本公开提供了方法38-44中的任一种的方法，其中所述供体模板是单链或双链多核苷酸。

[0582] 在另一种方法(方法46)中，本公开提供了方法38-45中的任一种的方法，其中所述供体模板具有与14q32.13区域同源的臂。

[0583] 在另一种方法(方法47)中，本公开提供了方法44的方法，其中所述DSB、或5' DSB和3' DSB是在所述SERPINA1基因的第一个、第二个、第三个、第四个、第五个外显子或内含子中。

[0584] 在另一种方法(方法48)中，本公开提供了方法1、2、7、13、19、24、30-33或38-41中的任一种的方法，其中所述gRNA或sgRNA指向下述病理学变体中的一种或多种：rs764325655、rs121912713、rs28929474、rs17580、rs121912714、rs764220898、rs199422211、rs751235320、rs199422210、rs267606950、rs55819880、rs28931570。

[0585] 在另一种方法(方法49)中，本公开提供了方法1、2、7、13、19或24-48中的任一种的方法，其中所述插入或校正是通过同源性指导的修复(HDR)。

[0586] 在另一种方法(方法50)中，本公开提供了方法1、2、7、13、19或24中的任一种的方法，其中所述方法还包括在所述细胞中引入两种引导核糖核酸(gRNA)，和其中所述一种或多种DNA内切核酸酶是两种或更多种Cas9内切核酸酶，其实现在SERPINA1基因内或附近的一对双链断裂(DSB)，第一个是在5' DSB基因座且第二个是在3' DSB基因座，所述断裂引起5' DSB基因座和3' DSB基因座之间的染色体DNA的缺失，其导致SERPINA1基因内或附近的5' DSB基因座和3' DSB基因座之间的染色体DNA的永久缺失和AAT蛋白活性的恢复，并且其中第一引导RNA包含与5' DSB基因座的区段互补的间隔物序列且第二引导RNA包含与3' DSB基因座的区段互补的间隔物序列。

[0587] 在另一种方法(方法51)中，本公开提供了方法50的方法，其中所述两种gRNA是两种单分子引导RNA(sgRNA)。

[0588] 在另一种方法(方法52)中，本公开提供了方法50-51的任一种的方法，其中所述两种gRNA或两种sgRNA是两种修饰的gRNA或两种修饰的sgRNA。

[0589] 在另一种方法(方法53)中，本公开提供了方法50-52的任一种的方法，其中所述一种或多种DNA内切核酸酶与一种或两种gRNA或一种或两种sgRNA预先形成复合物。

[0590] 在另一种方法(方法54)中，本公开提供了方法50-53的任一种的方法，其中所述5' DSB和3' DSB都在SERPINA1基因的第一个外显子、第二个外显子、第三个外显子、第四个外显子或第五个外显子中或附近。

[0591] 在另一种方法(方法55)中，本公开提供了方法50-54的任一种的方法，其中所述缺失是1kb或更少的缺失。

[0592] 在另一种方法(方法56)中,本公开提供了方法1、2、7、13或19-55的任一种的方法,其中所述Cas9或Cpf1 mRNA、gRNA和供体模板各自配制进单独的脂质纳米颗粒中或都共配制进脂质纳米颗粒中。

[0593] 在另一种方法(方法57)中,本公开提供了方法1、6、11、15或19-55中的任一种的方法,其中将所述Cas9或Cpf1 mRNA配制进脂质纳米颗粒中,并将所述gRNA和供体模板通过病毒载体递送。

[0594] 在另一种方法(方法58)中,本公开提供了方法57的方法,其中所述病毒载体是腺伴随病毒(AAV)。

[0595] 在另一种方法(方法59)中,本公开提供了方法58的方法,其中所述AAV载体是AAV6载体。

[0596] 在另一种方法(方法60)中,本公开提供了方法1、2、7、13、19或24-55中的任一种的方法,其中将所述Cas9或Cpf1 mRNA、gRNA和供体模板各自配制进单独的外来体中或都共配制进外来体中。

[0597] 在另一种方法(方法61)中,本公开提供了方法1、2、7、13、19或24-55中的任一种的方法,其中将所述Cas9或Cpf1 mRNA配制进脂质纳米颗粒中,并将所述gRNA通过电穿孔递送给所述细胞,和将供体模板通过病毒载体递送给所述细胞,本公开提供了方法。

[0598] 在另一种方法(方法62)中,本公开提供了方法61的方法,其中所述病毒载体是腺伴随病毒(AAV)载体。

[0599] 在另一种方法(方法63)中,本公开提供了方法62的方法,其中所述AAV载体是AAV6载体。

[0600] 在另一种方法(方法64)中,本公开提供了上述方法中的任一种的方法,其中所述SERPINA1基因位于染色体14:1,493,319 – 1,507,264上(Genome Reference Consortium – GRCh38/hg38)。

[0601] 在另一种方法(方法65)中,本公开提供了方法1、2、7、13、19或24-55中的任一种的方法,其中所述AAT蛋白活性的恢复是相比于野生型或正常AAT蛋白活性。

[0602] 在另一种方法(方法66)中,本公开提供了方法1的方法,其中所述人细胞是肝细胞。

[0603] 在另一种方法(方法67)中,本公开提供了方法23的方法,其中所述细胞是肝细胞。

[0604] 在另一种方法(方法68)中,本公开提供了方法1、2、7、13、19或24-55中的任一种的方法,其中将所述SERPINA1基因可操作地连接至驱动SERPINA1基因的表达的外源启动子。

[0605] 在另一种方法(方法69)中,本公开提供了方法1、2、7、13、19或24-55中的任一种的方法,其中所述一个或多个DSB发生在内源启动子基因座的3’侧紧邻位置处。

[0606] 本公开也提供了一种或多种引导核糖核酸(gRNA)的组合物(组合物1),所述一种或多种引导核糖核酸(gRNA)用于在来自具有α1抗胰蛋白酶缺乏(AATD)的患者的细胞中编辑SERPINA1基因,所述一种或多种gRNA包含选自SEQ ID No:54,860-68,297中的核酸序列的间隔物序列,所述核酸序列用于编辑来自具有α1抗胰蛋白酶缺乏的患者的细胞中的SERPINA1基因。

[0607] 在另一种组合物(组合物2)中,本公开提供了组合物1的组合物,其中所述一种或多种gRNA是一种或多种单分子引导RNA (sgRNA)。

[0608] 在另一种组合物(组合物3)中,本公开提供了组合物1或组合物2的组合物,其中所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA或一种或多种经修饰的sgRNA。

[0609] 在另一种组合物(组合物4)中,本公开提供了组合物1或组合物2或组合物3的组合物,其中所述细胞是肝细胞。

[0610] 本公开也提供了一种或多种引导核糖核酸(gRNA)的组合物(组合物5),所述一种或多种引导核糖核酸(gRNA)用于在来自具有α1抗胰蛋白酶缺乏(AATD)的患者的细胞中编辑安全港基因座,所述一种或多种gRNA包含选自SEQ ID No:1-54,859中的核酸序列的间隔物序列,所述核酸序列用于编辑来自具有α1抗胰蛋白酶缺乏的患者的细胞中的安全港基因座,其中所述安全港基因座选自AAVS1 (PPP1R12C)、ALB、Angptl3、ApoC3、ASGR2、CCR5、FIX(F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、Serpina1、TF和TTR。

[0611] 在另一种组合物(组合物6)中,本公开提供了组合物5的组合物,其中所述一种或多种gRNA是一种或多种单分子引导RNA (sgRNA)。

[0612] 在另一种组合物(组合物7)中,本公开提供了组合物5或组合物6的组合物,其中所述一种或多种gRNA或一种或多种sgRNA是一种或多种经修饰的gRNA或一种或多种经修饰的sgRNA。

[0613] 在另一种组合物(组合物8)中,本公开提供了组合物5或组合物6或组合物7的组合物,其中所述细胞是肝细胞。

[0614] 定义

[0615] 术语“包含(comprising)”或“包含(comprises)”参考本发明必要的组合物、方法及其各自的组分(一种或多种)使用,并且无论是否必要都仍然对未指定要素的包含保持开放。

[0616] 术语“基本上由……组成”指给定的实施方案要求的那些要素。该术语允许存在不实质上影响本发明该实施方案的基础的和新颖的或功能性的特征(一种或多种)的额外要素。

[0617] 术语”由……组成”指本文所述的组合物、方法及其各自的组分,其排除在实施方案的该描述中没有列出的任何要素。

[0618] 单数形式“一(a)”、“一(an)”和“所述(the)”包括复数指代物,除非上下文另外清楚地指明。

[0619] 本文中呈现的某些数值的前面是术语“约”。术语“约”用于为术语“约”后面的数值以及大约为所述数值的数值提供字面支持,也就是说,近似的未列举的数值可为这样的数值:其在呈现它的上下文中为明确地列举的数值的实质等同物。术语“约”是指在列举的数值的±10%内的数值。

[0620] 当本文中呈现数值的范围时,考虑在本公开内包括在所述范围的下限和上限之间的每个居间值、为所述范围的上限和下限的值、和具有所述范围的所有所述值。本公开也考虑在所述范围的下限和上限内的所有可能的子范围。

实施例

[0621] 通过参考下述实施例将更充分地理解本发明,下述实施例提供了本发明的示例性

的非限制性实施方案。

[0622] 所述实施例描述了CRISPR系统作为示例性基因组编辑技术在SERPINA1基因中建立确定的治疗性基因组替换(在本文中称作“基因组修饰”)的用途,所述治疗性基因组替换导致基因组基因座中的突变的永久校正或在异源基因座处的表达,其恢复AAT蛋白活性。确定的治疗性修饰的引入代表AATD的潜在改善的新颖治疗策略,如本文中所述和解释的。

[0623] 实施例1 -用于SERPINA1基因(不同于作为安全港基因座)的CRISPR/SpCas9靶位点

[0624] 针对靶位点扫描SERPINA1基因的区域。针对具有序列NRG的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在SEQ ID No:54,860-61,324中所示。

[0625] 实施例2 – 用于SERPINA1基因(不同于作为安全港基因座)的CRISPR/SaCas9靶位点

[0626] 针对靶位点扫描SERPINA1基因的区域。针对具有序列NNGRRT的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在SEQ ID No:61,324-61,936中所示。

[0627] 实施例3 – 用于SERPINA1基因(不同于作为安全港基因座)的CRISPR/StCas9靶位点

[0628] 针对靶位点扫描SERPINA1基因的区域。针对具有序列NNAGAAW的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在SEQ ID No:61,937-62,069中所示。

[0629] 实施例4 – 用于SERPINA1基因(不同于作为安全港基因座)的CRISPR/TdCas9靶位点

[0630] 针对靶位点扫描SERPINA1基因的区域。针对具有序列NAAAAC的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在SEQ ID No:69,070-62,120中所示。

[0631] 实施例5 – 用于SERPINA1基因(不同于作为安全港基因座)的CRISPR/NmCas9靶位点

[0632] 针对靶位点扫描SERPINA1基因的区域。针对具有序列NNNNGHTT的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在SEQ ID No:62,121-62,563中所示。

[0633] 实施例6 – 用于SERPINA1基因(不同于作为安全港基因座)的CRISPR/Cpf1靶位点

[0634] 针对靶位点扫描SERPINA1基因的区域。针对具有序列YTN的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 22碱基对间隔物序列,如在SEQ ID No:62,564-68,297中所示。

[0635] 实施例7 – 用于安全港基因座的CRISPR/SpCas9靶位点

[0636] 针对靶位点扫描下列安全港基因座:AAVS1 (PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2、和TTR的外显子1-2。

针对具有序列NRG的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在下列序列中显示:AAVS1 (PPP1R12C):SEQ ID No. 1-2,032; ALB:SEQ ID No. 3,482-3,649; Angptl3:SEQ ID No. 4,104-4,448; ApoC3:SEQ ID No. 5,432-5,834; ASGR2:SEQ ID No. 6,109-7,876; CCR5:SEQ ID No. 9,642-9,844; FIX (F9):SEQ ID No. 10,221-11,686; G6PC:SEQ ID No. 14,230-15,245; Gys2:SEQ ID No. 16,581-22,073; HGD:SEQ ID No. 32,254-33,946; Lp(a):SEQ ID No. 36,789-40,583; Pcsk9:SEQ ID No. 46,154-48,173; Serpina1:SEQ ID No. 50,345-51,482; TF:SEQ ID No. 52,446-53,277; 和TTR:SEQ ID No. 54,063-54,362。

[0637] 实施例8 - 用于安全港基因座的CRISPR/SaCas9靶位点

[0638] 针对靶位点扫描下列安全港基因座:AAVS1 (PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2、和TTR的外显子1-2。针对具有序列NNGRRT的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在下列序列中显示:AAVS1 (PPP1R12C):SEQ ID No. 2,033-2,203; ALB:SEQ ID No. 3,650-3,677; Angptl3:SEQ ID No. 4,449-4,484; ApoC3:SEQ ID No. 5,835-5,859; ASGR2:SEQ ID No. 7,877-8,082; CCR5:SEQ ID No. 9,845-9,876; FIX (F9):SEQ ID No. 11,687-11,849; G6PC:SEQ ID No. 15,246-15,362; Gys2:SEQ ID No. 22,074-22,749; HGD:SEQ ID No. 33,947-34,160; Lp(a):SEQ ID No. 40,584-40,993; Pcsk9:SEQ ID No. 48,174-48,360; Serpina1:SEQ ID No. 51,483-51,575; TF:SEQ ID No. 53,278-53,363; 和TTR:SEQ ID No. 54,363-54,403。

[0639] 实施例9 - 用于安全港基因座的CRISPR/StCas9靶位点

[0640] 针对靶位点扫描下列安全港基因座:AAVS1 (PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2、和TTR的外显子1-2。针对具有序列NNAGAAW的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在下列序列中显示:AAVS1 (PPP1R12C):SEQ ID No. 2,204-2,221; ALB:SEQ ID No. 3,678-3,695; Angptl3:SEQ ID No. 4,485-4,507; ApoC3:SEQ ID No. 5,860-5,862; ASGR2:SEQ ID No. 8,083-8,106; CCR5:SEQ ID No. 9,877-9,890; FIX (F9):SEQ ID No. 11,850-11,910; G6PC:SEQ ID No. 15,363-15,386; Gys2:SEQ ID No. 22,750-20,327; HGD:SEQ ID No. 34,161-34,243; Lp(a):SEQ ID No. 40,994-41,129; Pcsk9:SEQ ID No. 48,361-48,396; Serpina1:SEQ ID No. 51,576-51,587; TF:SEQ ID No. 53,364-53,375; 和TTR:SEQ ID No. 54,404-54,420。

[0641] 实施例10 - 用于安全港基因座的CRISPR/TdCas9靶位点

[0642] 针对靶位点扫描下列安全港基因座:AAVS1 (PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2、和TTR的外显子1-2。

针对具有序列NAAAAC的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在下列序列中显示:AAVS1 (PPP1R12C):SEQ ID No. 2,222-2,230; ALB:SEQ ID No. 3,696-3,700; Angptl3:SEQ ID No. 4,508-4,520; ApoC3:SEQ ID No. 5,863-5,864; ASGR2:SEQ ID No. 8,107-8,118; CCR5:SEQ ID No. 9,891-9,892; FIX (F9):SEQ ID No. 11,911-11,935; G6PC:SEQ ID No. 15,387-15,395; Gys2:SEQ ID No. 23,028-23,141; HGD:SEQ ID No. 34,244-34,262; Lp(a):SEQ ID No. 41,130-41,164; Pcsk9:SEQ ID No. 48,397-48,410; Serpina1:SEQ ID No. 51,588-51,590; TF:SEQ ID No. 53,376-53,382; 和TTR:SEQ ID No. 54,421-54,422。

[0643] 实施例11 - 用于安全港基因座的CRISPR/NmCas9靶位点

[0644] 针对靶位点扫描下列安全港基因座:AAVS1 (PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2、和TTR的外显子1-2。针对具有序列NNNNGHTT的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 20碱基对间隔物序列,如在下列序列中显示:AAVS1 (PPP1R12C):SEQ ID No. 2,231-2,305; ALB:SEQ ID No. 3,701-3,724; Angptl3:SEQ ID No. 4,521-4,583; ApoC3:SEQ ID No. 5,865-5,876; ASGR2:SEQ ID No. 8,119-8,201; CCR5:SEQ ID No. 9,893-9,920; FIX (F9):SEQ ID No. 11,936-12,088; G6PC:SEQ ID No. 15,396-15,485; Gys2:SEQ ID No. 23,142-23,821; HGD:SEQ ID No. 34,263-34,463; Lp(a):SEQ ID No. 41,165-41,532; Pcsk9:SEQ ID No. 48,411-48,550; Serpina1:SEQ ID No. 51,591-51,641; TF:SEQ ID No. 53,383-53,426; 和TTR:SEQ ID No. 54,423-54,457。

[0645] 实施例12 - 用于安全港基因座的CRISPR/Cpf1靶位点

[0646] 针对靶位点扫描AAVS1 (PPP1R12C)基因的外显子1-2。针对靶位点扫描下列安全港基因座:AAVS1 (PPP1R12C)的外显子1-2、ALB的外显子1-2、Angptl3的外显子1-2、ApoC3的外显子1-2、ASGR2的外显子1-2、CCR5的外显子1-2、FIX (F9)的外显子1-2、G6PC的外显子1-2、Gys2的外显子1-2、HGD的外显子1-2、Lp(a)的外显子1-2、Pcsk9的外显子1-2、Serpina1的外显子1-2、TF的外显子1-2、和TTR的外显子1-2。针对具有序列YTN的原间隔物邻近基序(PAM)扫描每个区域。鉴别出与PAM对应的gRNA 22碱基对间隔物序列,如在下列序列中显示:AAVS1 (PPP1R12C):SEQ ID No. 2,306-3,481; ALB:SEQ ID No. 3,725-4,103; Angptl3:SEQ ID No. 4,584-5,431; ApoC3:SEQ ID No. 5,877-6,108; ASGR2:SEQ ID No. 8,202-9,641; CCR5:SEQ ID No. 9,921-10,220; FIX (F9):SEQ ID No. 12,089-14,229; G6PC:SEQ ID No. 15,486-16,580; Gys2:SEQ ID No. 23,822-32,253; HGD:SEQ ID No. 34,464-36,788; Lp(a):SEQ ID No. 41,533-46,153; Pcsk9:SEQ ID No. 48,551-50,344; Serpina1:SEQ ID No. 51,642-52,445; TF:SEQ ID No. 53,427-54,062; 和TTR:SEQ ID No. 54,458-54,859。

[0647] 实施例13 -引导链的生物信息学分析

[0648] 将在多步骤过程中筛选和选择候选引导,所述过程牵涉理论结合和实验评估的活性。作为举例说明,为了评估在意图的那些以外的染色体位置处的效应的可能性,使用如在下面更详细地描述和举例说明的可用于评估靶外结合的多种生物信息学工具中的一种或

多种,可以针对它们在具有类似序列的靶外位点处切割的潜力评估这样的候选引导:其具有使特定在靶位点诸如在SERPINA1基因内的位点与邻近PAM匹配的序列。然后可以实验评估预测具有相对较低的靶外活性潜力的候选物以测量它们的在靶活性,然后测量在各种位点处的靶外活性。优选的引导具有足够高的在靶活性以在选定的基因座处达到期望的基因编辑水平,并具有相对较低的靶外活性以减小在其它染色体基因座处的改变的可能性。在靶活性与靶外活性的比率经常被称作引导的"特异性"。

[0649] 对于预测的靶外活性的初期筛选,存在许多已知的且和公众可得到的生物信息学工具,其可以用于预测最可能的靶外位点;且由于与CRISPR/Cas9/Cpf1核酸酶系统中的靶位点的结合由互补序列之间的沃森-克里克碱基配对驱动,不同性的程度(和因此减少的靶外结合潜力)基本上与以下基本序列差异有关:错配和凸起,即改变成非互补碱基的碱基,和相对于靶位点在潜在靶外位点中的碱基的插入或缺失。一种被称作COSMID (具有错配、插入和缺失的CRISPR靶外位点) (可在网络上在crispr.bme.gatech.edu得到)的示例性生物信息学工具编译这样的相似性。其它生物信息学工具包括但不限于GUIDO、autoCOSMID、和CCtop。

[0650] 使用生物信息学使靶外切割最小化以便减小突变和染色体重排的有害作用。在CRISPR/Cas9系统上的研究提示,由引导链与具有碱基对错配和/或凸起的DNA序列(特别是在远离PAM区域的位置处)的非特异性杂交引起的高靶外活性的可能性。因此,重要的是具有这样的生物信息学工具:其可以鉴别除了碱基对错配以外还具有在RNA引导链和基因组序列之间的插入和/或缺失的潜在靶外位点。基于生物信息学的工具COSMID (具有错配、插入和缺失的CRISPR靶外位点)因此被用于针对潜在CRISPR靶外位点来检索基因组(可在网络上在crispr.bme.gatech.edu得到)。COSMID输出基于错配的数目和位置的潜在靶外位点的排序列表,允许靶位点的更知情选择,并避免具有更可能的靶外切割的位点的应用。

[0651] 采用了另外的生物信息学渠道,其权衡区域中的gRNA靶向位点的估计在靶和/或靶外活性。可用于预测活性的其它特征包括关于目标细胞类型、DNA可达性、染色质状态、转录因子结合位点、转录因子结合数据和其它CHIP-seq数据的信息。权衡另外的因素,其预测编辑效率诸如gRNA对的相对位置和方向、局部序列特征和微-同源性。

[0652] <u>实施例14-针对在靶活性在细胞中测试优选引导</u>

[0653] 然后将针对在模型细胞系诸如Huh-7细胞中的在靶活性测试预测具有最低靶外活性的gRNA,并使用TIDE或下一代测序针对插入/缺失频率进行评价。TIDE是快速地评估CRISPR-Cas9对由引导RNA (gRNA或sgRNA)确定的靶基因座的基因组编辑的网络工具。基于来自两个标准毛细管测序反应的定量序列跟踪数据,TIDE软件量化编辑效力和鉴别在靶向细胞集合的DNA中的插入和缺失(插入/缺失)的主要类型。关于详细的解释和实施例,参见Brinkman等人, Nucl. Acids Res. (2014)。下一代测序(NGS)也被称作高通量测序,是用于描述许多不同的现代测序技术的概括(catch-all)术语,所述现代测序技术包括:Illumina (Solexa)测序、Roche 454测序、Ion torrent:Proton/PGM测序和SOLiD测序。这些最近的技术允许人们比以前使用的Sanger测序更快速和便宜得多地对DNA和RNA测序,且因此已经彻底改变了基因组学和分子生物学的研究。

[0654] 组织培养细胞的转染允许筛选不同的构建体和测试活性和特异性的稳健方式。将组织培养细胞系诸如Huh-7细胞容易地转染并导致高活性。将评价这些或其它细胞系以确

定提供最佳替代物的细胞系。然后将这些细胞用于许多早期阶段测试。例如，使用适合于在人细胞中表达的质粒诸如，例如，在图1A-1C中描述的CTx-1、CTx-2或CTx-3，将化脓性链球菌Cas9的个别gRNA转染进所述细胞中。几天以后，收获基因组DNA，并通过PCR扩增靶位点。通过游离DNA末端的NHEJ修复所引入的插入、缺失和突变的比率，可以测量切割活性。尽管该方法不可将正确地修复的序列与未切割的DNA区分开，通过错误修复的量可以计量切割的水平。通过扩增鉴别的假定靶外位点和使用类似的方法检测切割，可以观察靶外活性。还可以使用侧接切割位点的引物测定易位，以确定特异性的切割和易位是否发生。已经开发了未引导的测定，其允许包括引导序列(guide-seq)的靶外切割的互补测试。然后可以在培养的细胞中跟踪具有显著活性的gRNA或gRNA对以测量SERPINA1突变的校正。可以再次跟踪靶外事件。这些实验允许优化核酸酶和供体设计和递送。

[0655] 实施例15-针对靶外活性在细胞中测试优选引导

[0656] 然后将使用全基因组测序针对靶外活性测试在以上实施例中来自TIDE和下一代测序研究的具有最佳在靶活性的gRNA。

[0657] 实施例16-针对HDR基因编辑测试不同的途径

[0658] 在针对在靶活性和靶外活性测试gRNA以后，将针对HDR基因编辑测试突变校正和敲入策略。

[0659] 对于突变校正途径，供体DNA模板将提供为短的单链寡核苷酸、短的双链寡核苷酸(完整的PAM序列/突变的PAM序列)、长的单链DNA分子(完整的PAM序列/突变的PAM序列)或长的双链DNA分子(完整的PAM序列/突变的PAM序列)。另外，将通过AAV递送供体DNA模板。

[0660] 对于cDNA敲入途径，具有与14q32.13区域同源的臂的单链或双链DNA可包括SERPINA1基因的第一个外显子(第一个编码外显子)的超过40个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后内含子的至少40个核苷酸。具有与14q32.13区域同源的臂的单链或双链DNA包括SERPINA1基因的第一个外显子的超过80个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后内含子的至少80个核苷酸。具有与14q32.13区域同源的臂的单链或双链DNA可包括SERPINA1基因的第一个外显子的超过100个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后内含子的至少100个核苷酸。具有与14q32.13区域同源的臂的单链或双链DNA可包括SERPINA1基因的第一个外显子的超过150个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后内含子的至少150个核苷酸。具有与14q32.13区域同源的臂的单链或双链DNA可包括SERPINA1基因的第一个外显子的超过300个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后内含子的至少300个核苷酸。具有与14q32.13区域同源的臂的单链或双链DNA可包括SERPINA1基因的第一个外显子的超过400个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后第一个内含子的至少400个核苷酸。可选地，将通过AAV递送DNA模板。

[0661] 对于cDNA 或小基因敲入途径，具有与14q32.13区域同源的臂的单链或双链DNA包括SERPINA1基因的第二个外显子(第一个编码外显子)的超过80个核苷酸、SERPINA1基因的完整CDS和SERPINA1基因的3' UTR、和以后内含子的至少80个核苷酸。可选地，DNA模板将通过AAV递送。

[0662] 接着，将评估HDR介导的SERPINA1的PI Z等位基因的常见突变的校正和cDNA 小基因(包含天然的或合成的增强子和启动子、一个或多个外显子、和天然的或合成的内含子、

和天然的或合成的3' UTR和多腺苷酸化信号)向第2个外显子中的敲入的效率。

[0663] 实施例17-先导CRISPR-Cas9/DNA供体组合的重新评估

[0664] 测试用于HDR基因编辑的不同策略以后,将针对缺失的效率、重组和靶外特异性在原代人肝细胞中重新评估先导CRISPR-Cas9/DNA供体组合。Cas9 mRNA或RNP将被配制进脂质纳米颗粒中用于递送,sgRNA将被配制进纳米颗粒中或作为AAV递送,并且供体DNA将被配制进纳米颗粒中或作为AAV递送。

[0665] 实施例18-在有关小鼠模型中的体内测试

[0666] 在已经重新评估CRISPR-Cas9/DNA供体组合以后,将在动物模型中在体内测试先导制剂。作为非限制性例子,合适的动物模型包括具有重新植入人肝细胞或iPSC衍生出的肝细胞(正常的或AAT缺陷的)的肝的FGR小鼠模型。

[0667] 在人细胞中的培养允许在如上所述的人靶标和背景人基因组上直接测试。

[0668] 通过经修饰的小鼠或人肝细胞在FGR小鼠中的移植物植入,可以观察临床前效力和安全性评价。可以在移植物植入以后几个月中观察经修饰的细胞。

[0669] 实施例19 - gRNA的筛选

[0670] 为了鉴别大范围的(a large spectrum of)能够编辑SERPINA1 DNA靶区域的gRNA对,进行体外转录的(IVT)gRNA筛选。提交SERPINA1基因组序列应用gRNA设计软件进行分析。得到的gRNA列表缩窄至约200个gRNA的列表,这基于序列的独特性——仅筛选在基因组其它处缺乏完美匹配的gRNA——和最低限度预测的靶外。在体外转录该gRNA组和应用messenger Max将其转染到组成性表达Cas9的HEK293T细胞中。转染后48小时收获细胞,分离基因组DNA,并用TIDE分析评估切割效率。gRNA的切割效率(%插入/缺失)在表4中列出。

[0671] 发现大约14%的被测试gRNA引起超过70%的切割效率。(图3)。

[0672] 表4. gRNA序列和在HEK293T-Cas9细胞中的切割效率

[0673]

引导名称	引导序列	% 插入/缺失
SERPINA1_T61	CCCAATGTCTAGAAGGTCTT (SEQ ID NO: 55,783)	70.3
SERPINA1_T37	GCCCAATGTCTAGAAGGTCT (SEQ ID NO: 55,782)	50.6
SERPINA1_T185	ATCAAGGCCCAATGTCTAGA (SEQ ID NO: 55,781)	58.1
SERPINA1_T70	GCCCAAGACCTTCTAGACAT (SEQ ID NO: 60,476)	56.1
SERPINA1_T88	CCCAAGACCTTCTAGACATT (SEQ ID NO: 60,477)	60
SERPINA1_T162	GTGTCAGTATGGATAAATCA (SEQ ID NO: 55,778)	59.2
SERPINA1_T170	CAAACCTTTCTGTGTCAGTA (SEQ ID NO: 55,776)	7.1
SERPINA1_T390	TTATCCATACTGACACAGAA (SEQ ID NO: 60,480)	64.7
SERPINA1_T388	CATACTGACACAGAAAGGTT (SEQ ID NO: 60,481)	63.6
SERPINA1_T283	ATACTGACACAGAAAGGTTT (SEQ ID NO: 60,482)	70.5
SERPINA1_T282	TTTGGGCTAAGTTGTTTCAA (SEQ ID NO: 60,485)	40.5
SERPINA1_T72	CACCAAATCTCACAGATCGA (SEQ ID NO: 55,771)	81.6
SERPINA1_T73	CTCCTTCGATCTGTGAGATT (SEQ ID NO: 60,487)	47.3
SERPINA1_T259	GCTAAAGATGACACTTATTT (SEQ ID NO: 60,490)	73.9
SERPINA1_T275	CACTTATTTTGGAAAACTAA (SEQ ID NO: 60,492)	0.6
SERPINA1_T182	ATGGAACTGCAGTTGTTCAT (SEQ ID NO: 55,764)	52.8
SERPINA1_T323	AACAACTGCAGTTCCATGAA (SEQ ID NO: 60,494)	66.2
SERPINA1_T383	CAAGATAATGCAGCCATTCA (SEQ ID NO: 55,762)	50.7
SERPINA1_T177	CATGAATGGCTGCATTATCT (SEQ ID NO: 60,495)	22
SERPINA1_T280	ATGAATGGCTGCATTATCTT (SEQ ID NO: 60,496)	57.8
SERPINA1_T244	TGAATGGCTGCATTATCTTG (SEQ ID NO: 60,497)	51.8
SERPINA1_T66	GGCTGCATTATCTTGGGGTC (SEQ ID NO: 60,498)	18
SERPINA1_T128	GCTGCATTATCTTGGGGTCT (SEQ ID NO: 60,499)	34.8
SERPINA1_T385	CACTGTGAAGGTCACTGCCA (SEQ ID NO: 60,504)	50.8
SERPINA1_T194	CTCCTTGAGGACACGGACCC (SEQ ID NO: 55,756)	15.2
SERPINA1_T418	TGCCAGGGTCCGTGTCCTCA (SEQ ID NO: 60,506)	49.6
SERPINA1_T59	ACACGGCTTGAAGCTCCTTG (SEQ ID NO: 55,754)	49.3
SERPINA1_T257	GCTCTCTCCTTTCTAGTACA (SEQ ID NO: 55,751)	90.1
SERPINA1_T13	CTTCAAGCCGTGTACTAGAA (SEQ ID NO: 60,511)	15.5
SERPINA1_T291	TACTAGAAAGGAGAGAGCCC (SEQ ID NO: 60,515)	45.8
SERPINA1_T431	TAGAAAGGAGAGAGCCCTGG (SEQ ID NO: 60,517)	60.1

[0674]

SERPINA1_T341	TCACTCCACGTCTGCCTCCA	(SEQ ID NO: 55,749)	83.9
SERPINA1_T471	GTCACTCCACGTCTGCCTCC	(SEQ ID NO: 55,748)	51.7
SERPINA1_T294	GAGAGCCCTGGAGGCAGACG	(SEQ ID NO: 60,519)	44.8
SERPINA1_T187	TGCACCCACAACTCAGAACA	(SEQ ID NO: 55,744)	77.1
SERPINA1_T92	GTGCACCCACAACTCAGAAC	(SEQ ID NO: 55,743)	46.5
SERPINA1_T314	CTCTTCCCTGTTCTGAGTTG	(SEQ ID NO: 60,522)	27.8
SERPINA1_T120	AGTTGTGGGTGCACCTGAGC	(SEQ ID NO: 60,526)	10.6
SERPINA1_T224	GTTGTGGGTGCACCTGAGCA	(SEQ ID NO: 60,527)	72.5
SERPINA1_T274	TTGTGGGTGCACCTGAGCAG	(SEQ ID NO: 60,528)	39.8
SERPINA1_T167	AAGCGCCTCTCCCCCTGCTC	(SEQ ID NO: 55,740)	36.8
SERPINA1_T437	TGTGGGTGCACCTGAGCAGG	(SEQ ID NO: 60,529)	42.7
SERPINA1_T81	CAGGGGGAGAGGCGCTTGTC	(SEQ ID NO: 60,534)	15.8
SERPINA1_T168	AGAGGCGCTTGTCAGGAAGA	(SEQ ID NO: 60,536)	38
SERPINA1_T336	CTTGTCAGGAAGATGGACAG	(SEQ ID NO: 60,539)	46
SERPINA1_T360	TTGTCAGGAAGATGGACAGA	(SEQ ID NO: 60,540)	45.1
SERPINA1_T405	TGTCAGGAAGATGGACAGAG	(SEQ ID NO: 60,541)	80.2
SERPINA1_T379	AAGGCTTTGGCTGATGGGGC	(SEQ ID NO: 55,737)	17.5
SERPINA1_T180	CCTCAAGGCTTTGGCTGATG	(SEQ ID NO: 55,736)	14
SERPINA1_T135	TCCTCAAGGCTTTGGCTGAT	(SEQ ID NO: 55,735)	25.9
SERPINA1_T197	CTCCTCAAGGCTTTGGCTGA	(SEQ ID NO: 55,734)	20.3
SERPINA1_T345	CTTGCTCCTCCTCAAGGCTT	(SEQ ID NO: 55,733)	23.8
SERPINA1_T150	CCCCATCAGCCAAAGCCTTG	(SEQ ID NO: 60,547)	34.9
SERPINA1_T328	CATCAGCCAAAGCCTTGAGG	(SEQ ID NO: 60,549)	35.7
SERPINA1_T138	ATAGGCCTTGCTCCTCCTCA	(SEQ ID NO: 55,732)	76.2
SERPINA1_T348	CAAAGCCTTGAGGAGGAGCA	(SEQ ID NO: 60,552)	24.4
SERPINA1_T176	AGGAGCAAGGCCTATGTGAC	(SEQ ID NO: 60,554)	32.7
SERPINA1_T376	CCTCTCCCTCCCTGTCACAT	(SEQ ID NO: 55,730)	70.1
SERPINA1_T308	GGAGCAAGGCCTATGTGACA	(SEQ ID NO: 60,555)	58.8
SERPINA1_T181	GCAAGGCCTATGTGACAGGG	(SEQ ID NO: 60,557)	71
SERPINA1_T265	CAAGGCCTATGTGACAGGGA	(SEQ ID NO: 60,558)	44.6
SERPINA1_T465	CCTATGTGACAGGGAGGGAG	(SEQ ID NO: 60,561)	24.5
SERPINA1_T392	CAGGGAGGGAGAGGATGTGC	(SEQ ID NO: 60,563)	26.9
SERPINA1_T372	AGGGAGGGAGAGGATGTGCA	(SEQ ID NO: 60,564)	74.3
SERPINA1_T459	GGGAGAGGATGTGCAGGGCC	(SEQ ID NO: 60,566)	16.3
SERPINA1_T444	GGAGAGGATGTGCAGGGCCA	(SEQ ID NO: 60,567)	32.7
SERPINA1_T113	TCACTCCCCCTGGACGGCCC	(SEQ ID NO: 55,728)	9.7
SERPINA1_T253	GTGCAGGGCCAGGGCCGTCC	(SEQ ID NO: 60,569)	67.9

[0675]

SERPINA1_T312	TGCAGGGCCAGGGCCGTCCA	(SEQ ID NO: 60,570)	63.7
SERPINA1_T445	GCAGGGCCAGGGCCGTCCAG	(SEQ ID NO: 60,571)	59.7
SERPINA1_T441	CAGGGCCAGGGCCGTCCAGG	(SEQ ID NO: 60,572)	55.8
SERPINA1_T51	AAGCGCTCACTCCCCCTGGA	(SEQ ID NO: 55,727)	72.9
SERPINA1_T114	CAGGAAGCGCTCACTCCCCC	(SEQ ID NO: 55,726)	69.2
SERPINA1_T164	CAGGGGGAGTGAGCGCTTCC	(SEQ ID NO: 60,575)	56.3
SERPINA1_T155	AGGGGGAGTGAGCGCTTCCT	(SEQ ID NO: 60,576)	37.3
SERPINA1_T240	GGGAGTGAGCGCTTCCTGGG	(SEQ ID NO: 60,578)	71.3
SERPINA1_T46	GGCTCACGTGGACACCTCCC	(SEQ ID NO: 55,724)	65.4
SERPINA1_T30	GGCCTCGAGCAAGGCTCACG	(SEQ ID NO: 55,722)	72.2
SERPINA1_T7	GTCCACGTGAGCCTTGCTCG	(SEQ ID NO: 60,581)	58.6
SERPINA1_T209	CTGATCCCAGGCCTCGAGCA	(SEQ ID NO: 55,721)	69.3
SERPINA1_T476	CGTGAGCCTTGCTCGAGGCC	(SEQ ID NO: 60,582)	13.4
SERPINA1_T67	GTGAGCCTTGCTCGAGGCCT	(SEQ ID NO: 60,583)	0.2
SERPINA1_T34	CACGTTGTAAGGCTGATCCC	(SEQ ID NO: 55,718)	66
SERPINA1_T139	AGAAGCAGAGACACGTTGTA	(SEQ ID NO: 55,716)	76.4
SERPINA1_T49	GCCTTATGCACGGCCTGGAG	(SEQ ID NO: 55,709)	66.2
SERPINA1_T19	AGCCTTATGCACGGCCTGGA	(SEQ ID NO: 55,708)	74.3
SERPINA1_T52	CAGCCTTATGCACGGCCTGG	(SEQ ID NO: 55,707)	25.1
SERPINA1_T32	GCACAGCCTTATGCACGGCC	(SEQ ID NO: 55,705)	65.9
SERPINA1_T54	TCCCCTCCAGGCCGTGCATA	(SEQ ID NO: 60,588)	73.8
SERPINA1_T107	GGTCAGCACAGCCTTATGCA	(SEQ ID NO: 55,704)	82.2
SERPINA1_T9	TTCAGTCCCTTTCTCGTCGA	(SEQ ID NO: 55,701)	42
SERPINA1_T5	GTGCTGACCATCGACGAGAA	(SEQ ID NO: 60,591)	53.9
SERPINA1_T14	TGCTGACCATCGACGAGAAA	(SEQ ID NO: 60,592)	40.5
SERPINA1_T355	GAGAAAGGGACTGAAGCTGC	(SEQ ID NO: 60,594)	15.3
SERPINA1_T75	GGGTATGGCCTCTAAAAACA	(SEQ ID NO: 55,697)	33.5
SERPINA1_T175	TGCTGGGGCCATGTTTTTAG	(SEQ ID NO: 60,599)	45.3
SERPINA1_T213	GGGGGGGATAGACATGGGTA	(SEQ ID NO: 55,696)	55
SERPINA1_T11	ACCTCGGGGGGGATAGACAT	(SEQ ID NO: 55,695)	39.3
SERPINA1_T33	GACCTCGGGGGGGATAGACA	(SEQ ID NO: 55,694)	49.4
SERPINA1_T6	ACCCATGTCTATCCCCCCCG	(SEQ ID NO: 60,601)	66.5
SERPINA1_T31	TGTTGAACTTGACCTCGGGG	(SEQ ID NO: 55,692)	13.3
SERPINA1_T29	TTGTTGAACTTGACCTCGGG	(SEQ ID NO: 55,691)	4.6
SERPINA1_T23	TTTGTTGAACTTGACCTCGG	(SEQ ID NO: 55,690)	36.4
SERPINA1_T18	GTTTGTTGAACTTGACCTCG	(SEQ ID NO: 55,689)	20
SERPINA1_T74	GGTTTGTTGAACTTGACCTC	(SEQ ID NO: 55,688)	56.9

SERPINA1_T63	GGGTTTGTTGAACTTGACCT	(SEQ ID NO: 55,687)	32.3
SERPINA1_T317	TCAATCATTAAGAAGACAAA	(SEQ ID NO: 55,686)	63.2
SERPINA1_T276	TTCAATCATTAAGAAGACAA	(SEQ ID NO: 55,685)	15.5
SERPINA1_T365	TCCCATGAAGAGGGGAGACT	(SEQ ID NO: 55,681)	28.2
SERPINA1_T188	TACCAAGTCTCCCCTCTTCA	(SEQ ID NO: 60,604)	27.4
SERPINA1_T243	ACCAAGTCTCCCCTCTTCAT	(SEQ ID NO: 60,605)	46.3
SERPINA1_T289	CACCACTTTTCCCATGAAGA	(SEQ ID NO: 55,678)	20.6
SERPINA1_T293	TCACCACTTTTCCCATGAAG	(SEQ ID NO: 55,677)	24.5
SERPINA1_T334	TCCCCTCTTCATGGGAAAAG	(SEQ ID NO: 60,607)	48.9
SERPINA1_T252	AGAGGCAGTTATTTTTGGGT	(SEQ ID NO: 55,674)	14.1
SERPINA1_T346	GAGAGGCAGTTATTTTTGGG	(SEQ ID NO: 55,673)	9.5
SERPINA1_T134	AGCGAGAGGCAGTTATTTTT	(SEQ ID NO: 55,672)	47.8
SERPINA1_T331	GAGCGAGAGGCAGTTATTTT	(SEQ ID NO: 55,671)	54.3
SERPINA1_T399	GGGAGGGGTTGAGGAGCGAG	(SEQ ID NO: 55,669)	54.1
SERPINA1_T464	GGGAGGGGGCCAGGGATGGA	(SEQ ID NO: 55,658)	13.9
SERPINA1_T460	AGGGAGGGGGCCAGGGATGG	(SEQ ID NO: 55,657)	1.5
SERPINA1_T462	CAACCCCTCCCCTCCATCCC	(SEQ ID NO: 60,608)	21.9
SERPINA1_T256	CCCTTCTTTAATGTCATCCA	(SEQ ID NO: 55,646)	55
[0676] SERPINA1_T96	ACCCTTCTTTAATGTCATCC	(SEQ ID NO: 55,645)	35.6
SERPINA1_T229	ACATTAAAGAAGGGTTGAGC	(SEQ ID NO: 60,615)	30
SERPINA1_T361	TTTACAGTCACATGCAGGCA	(SEQ ID NO: 55,642)	50.7
SERPINA1_T201	ATTTACAGTCACATGCAGGC	(SEQ ID NO: 55,641)	49.4
SERPINA1_T132	AGGGATTTACAGTCACATGC	(SEQ ID NO: 55,639)	70.7
SERPINA1_T316	AGACTCAGAGAAAACATGGG	(SEQ ID NO: 55,635)	43.6
SERPINA1_T302	GGGAGACTCAGAGAAAACAT	(SEQ ID NO: 55,633)	41.3
SERPINA1_T271	ATACAGCCTCAGCAGGCAAA	(SEQ ID NO: 55,628)	67.7
SERPINA1_T193	CATACAGCCTCAGCAGGCAA	(SEQ ID NO: 55,627)	40.5
SERPINA1_T340	GAGTCTCCCTTTGCCTGCTG	(SEQ ID NO: 60,618)	72.6
SERPINA1_T169	AGCCCACATACAGCCTCAGC	(SEQ ID NO: 55,625)	49.8
SERPINA1_T129	TTGCCTGCTGAGGCTGTATG	(SEQ ID NO: 60,619)	74.3
SERPINA1_T145	TGCCTGCTGAGGCTGTATGT	(SEQ ID NO: 60,620)	41.2
SERPINA1_T204	TGAGGCTGTATGTGGGCTCC	(SEQ ID NO: 60,622)	42.6
SERPINA1_T35	CCGAAGACAGCACTGTTACC	(SEQ ID NO: 55,620)	29
SERPINA1_T184	CCAGGTAACAGTGCTGTCTT	(SEQ ID NO: 60,624)	0
SERPINA1_T178	CAGGTAACAGTGCTGTCTTC	(SEQ ID NO: 60,625)	0
SERPINA1_T157	CTCCATGAACACAGTTCAGG	(SEQ ID NO: 55,617)	0
SERPINA1_T140	GCTCCATGAACACAGTTCAG	(SEQ ID NO: 55,616)	41.6

[0677]

SERPINA1_T147	AACTGTGTTCATGGAGCATC	(SEQ ID NO: 60,628)	18.2
SERPINA1_T273	TGTTCATGGAGCATCTGGCT	(SEQ ID NO: 60,630)	61.9
SERPINA1_T247	CATGGAGCATCTGGCTGGGT	(SEQ ID NO: 60,632)	49.1
SERPINA1_T380	CTGGCTGGGTAGGCACATGC	(SEQ ID NO: 60,633)	77.7
SERPINA1_T456	CACATGCTGGGCTTGAATCC	(SEQ ID NO: 60,636)	0
SERPINA1_T352	ACATGCTGGGCTTGAATCCA	(SEQ ID NO: 60,637)	63.2
SERPINA1_T443	ATGCTGGGCTTGAATCCAGG	(SEQ ID NO: 60,639)	71.8
SERPINA1_T377	TGCTGGGCTTGAATCCAGGG	(SEQ ID NO: 60,640)	57.3
SERPINA1_T76	GCTGAGGATTCAGTCCCCCC	(SEQ ID NO: 55,607)	80.9
SERPINA1_T44	GGGCCCAGGTCCGTAAGCTG	(SEQ ID NO: 55,605)	42.7
SERPINA1_T160	GGGACTGAATCCTCAGCTTA	(SEQ ID NO: 60,642)	58.7
SERPINA1_T4	GAATCCTCAGCTTACGGACC	(SEQ ID NO: 60,643)	13.7
SERPINA1_T20	AATCCTCAGCTTACGGACCT	(SEQ ID NO: 60,644)	6.1
SERPINA1_T207	CTCCAGAAACAGATGGGCCC	(SEQ ID NO: 55,602)	43
SERPINA1_T260	GGAGCCCTCCAGAAACAGAT	(SEQ ID NO: 55,600)	74.5
SERPINA1_T403	TGGAGCCCTCCAGAAACAGA	(SEQ ID NO: 55,599)	37.8
SERPINA1_T87	GACCTGGGCCCATCTGTTTC	(SEQ ID NO: 60,645)	49.4
SERPINA1_T284	CTGGGCCCATCTGTTTCTGG	(SEQ ID NO: 60,647)	32.7
SERPINA1_T228	TGGGCCCATCTGTTTCTGGA	(SEQ ID NO: 60,648)	72.4
SERPINA1_T406	CAAGACAGGACAAGGAAGAC	(SEQ ID NO: 55,595)	23.2
SERPINA1_T304	CAGTCTTCCTTGTCCTGTCT	(SEQ ID NO: 60,650)	37.4
SERPINA1_T126	CTTCTTGGGGACTCCAAGAC	(SEQ ID NO: 55,591)	30.1
SERPINA1_T137	CTGTCTTGGAGTCCCCAAGA	(SEQ ID NO: 60,654)	0
SERPINA1_T324	CTCCCCTGTGATTCCTTCTT	(SEQ ID NO: 55,587)	41.3
SERPINA1_T367	CCTCCCCTGTGATTCCTTCT	(SEQ ID NO: 55,586)	46.3
SERPINA1_T248	GTCCCCAAGAAGGAATCACA	(SEQ ID NO: 60,657)	51
SERPINA1_T310	TCCCCAAGAAGGAATCACAG	(SEQ ID NO: 60,658)	27.8
SERPINA1_T342	CCAAGAAGGAATCACAGGGG	(SEQ ID NO: 60,660)	49
SERPINA1_T332	GTGGAGCCTGGGGTCATGGC	(SEQ ID NO: 55,584)	48.7
SERPINA1_T358	CTTGGTGGAGCCTGGGGTCA	(SEQ ID NO: 55,583)	30
SERPINA1_T395	CAGATACCAGCCATGACCCC	(SEQ ID NO: 60,664)	20.8
SERPINA1_T362	AAGATGCTTGGTGGAGCCTG	(SEQ ID NO: 55,582)	21.1
SERPINA1_T235	GAAGATGCTTGGTGGAGCCT	(SEQ ID NO: 55,581)	4.1
SERPINA1_T238	TGAAGATGCTTGGTGGAGCC	(SEQ ID NO: 55,580)	5.8
SERPINA1_T255	GGGGACATGAAGATGCTTGG	(SEQ ID NO: 55,578)	59.5
SERPINA1_T296	CAGGGGGACATGAAGATGCT	(SEQ ID NO: 55,577)	64.9
SERPINA1_T455	GGGAGTGGGGGATGAGCAGG	(SEQ ID NO: 55,575)	43.7

[0678]

SERPINA1_T347	GGGGAGTGGGGGATGAGCAG	(SEQ ID NO: 55,574)	67.5
SERPINA1_T387	GGGGGAGTGGGGGATGAGCA	(SEQ ID NO: 55,573)	27
SERPINA1_T416	GGGGGGAGTGGGGGATGAGC	(SEQ ID NO: 55,572)	15.8
SERPINA1_T440	CTCTGGGTGGGGGGGAGTGG	(SEQ ID NO: 55,569)	31.4
SERPINA1_T430	ACTCTGGGTGGGGGGGAGTG	(SEQ ID NO: 55,568)	11.8
SERPINA1_T389	AACTCTGGGTGGGGGGGAGT	(SEQ ID NO: 55,567)	6.7
SERPINA1_T398	CAACTCTGGGTGGGGGGGAG	(SEQ ID NO: 55,566)	6.1
SERPINA1_T415	ATGAGCAACTCTGGGTGGGG	(SEQ ID NO: 55,564)	35.7
SERPINA1_T412	GATGAGCAACTCTGGGTGGG	(SEQ ID NO: 55,563)	30.3
SERPINA1_T396	GGATGAGCAACTCTGGGTGG	(SEQ ID NO: 55,562)	30.4
SERPINA1_T426	AGGATGAGCAACTCTGGGTG	(SEQ ID NO: 55,561)	18.5
SERPINA1_T246	CAGGATGAGCAACTCTGGGT	(SEQ ID NO: 55,560)	35.1
SERPINA1_T373	GCAGGATGAGCAACTCTGGG	(SEQ ID NO: 55,559)	40.8
SERPINA1_T205	CTGGCAGGATGAGCAACTCT	(SEQ ID NO: 55,558)	#N/A
SERPINA1_T174	CCTGGCAGGATGAGCAACTC	(SEQ ID NO: 55,557)	#N/A
SERPINA1_T219	CCAGAGTTGCTCATCCTGCC	(SEQ ID NO: 60,669)	#N/A
SERPINA1_T171	CAGAGTTGCTCATCCTGCCA	(SEQ ID NO: 60,670)	#N/A

[0679] 关于示例性实施方案的注解

[0680] 尽管为了举例说明本发明的各种方面和/或它的潜在应用的目的，本公开提供了各种具体实施方案的描述，但是应当理解，本领域技术人员会做出变化和修改。因此，本文描述的一项或多项发明应当理解为至少如它们被请求保护那样宽，并且不象本文提供的特定示例性实施方案那样更狭窄地限定。

[0681] 除非另有说明，否则在本文中指出的任何专利、出版物或其它公开材料通过引用以其整体并入本说明书，但是仅仅就并入的材料不与在本说明书中明确地阐述的现有描述、定义、陈述或其它公开材料冲突而言。因此，且就必要而言，在本说明书中阐述的明确公开内容替代通过引用并入的任何冲突材料。被说成通过引用并入本说明书、但是与在本文中阐述的现有定义、陈述或其它公开材料冲突的任何材料或其部分仅仅就在并入的材料和现有公开材料之间不产生冲突而言并入。申请人保留修改本说明书以明确地列举通过引用并入本文的任何主题或其部分的权利。

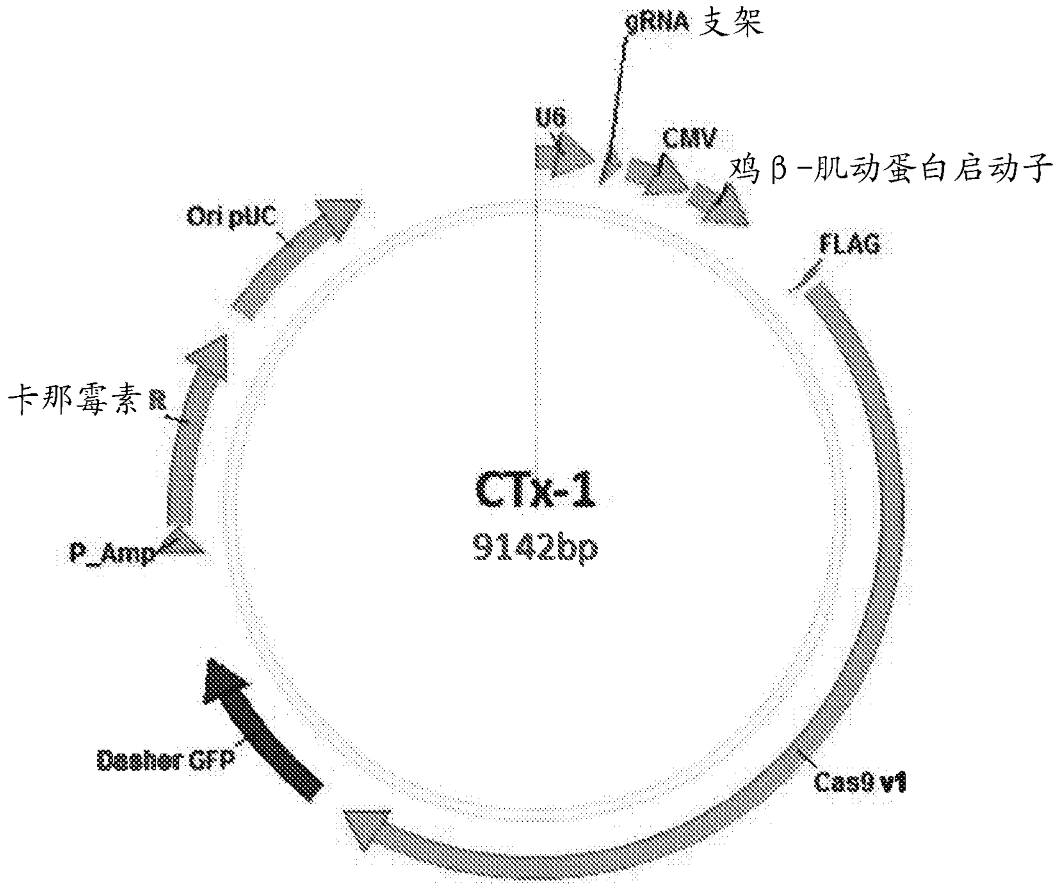

gRNA 支架
U6
CMV
鸡β-肌动蛋白启动子
Ori pUC
FLAG
卡那霉素R
CTx-1
9142bp
P_Amp
Dasher GFP
Cas9 v1

图 1A

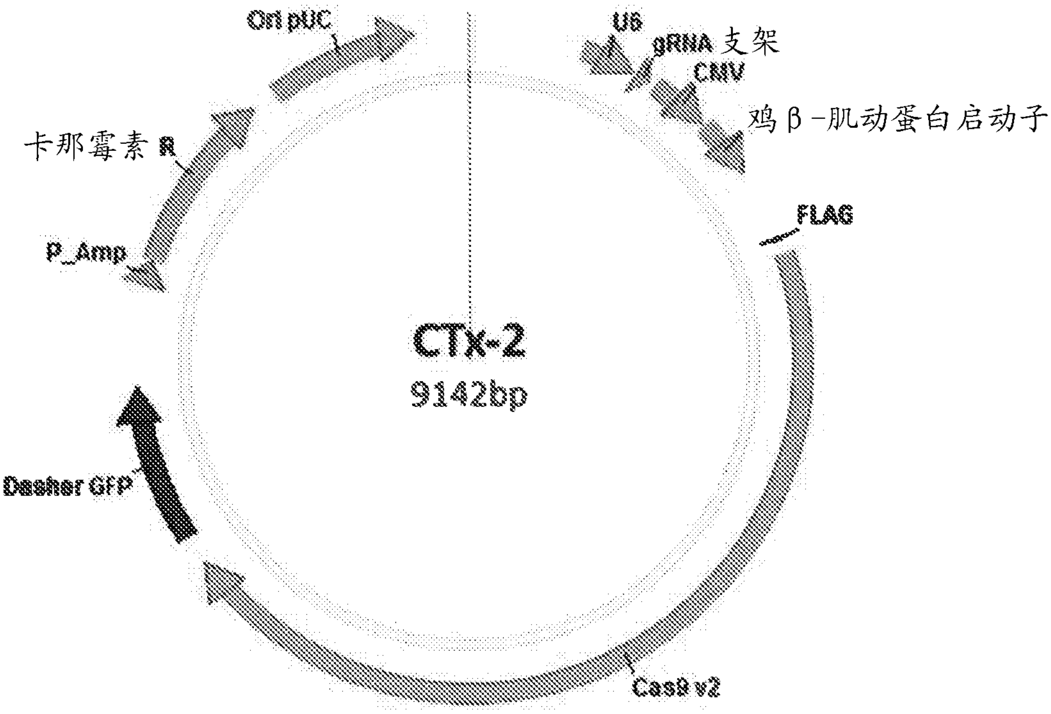

Ori pUC
U6
gRNA 支架
CMV
鸡β-肌动蛋白启动子
卡那霉素 R
P_Amp
FLAG
CTx-2
9142bp
Dasher GFP
Cas9 v2

图 1B

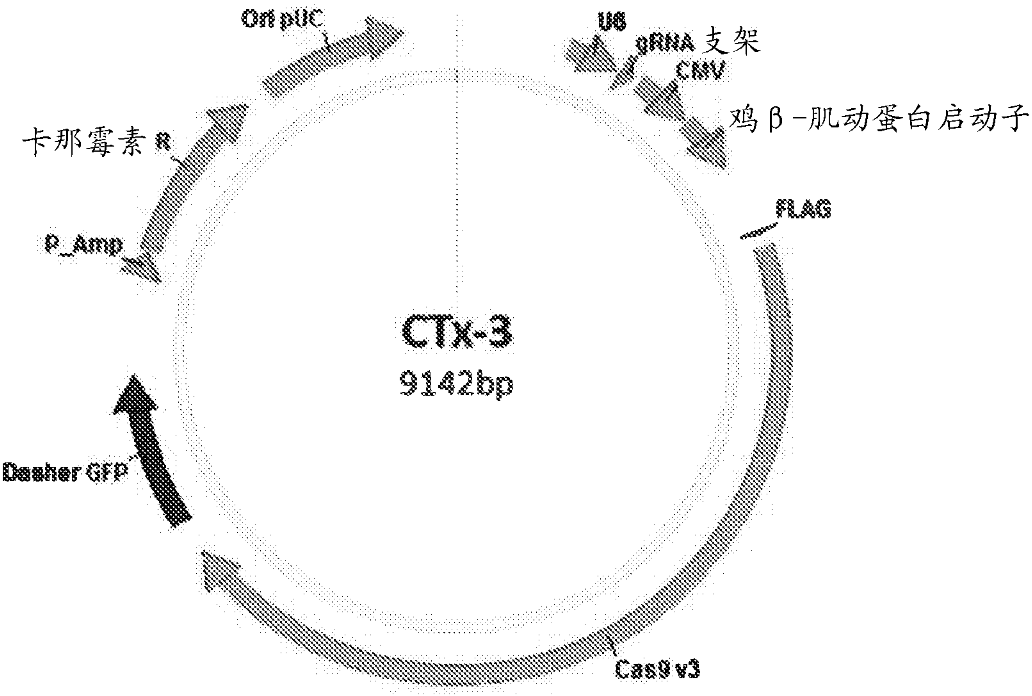

Ori pUC
U6
gRNA 支架
CMV
鸡β-肌动蛋白启动子
卡那霉素 R
P_Amp
FLAG
CTx-3
9142bp
Dasher GFP
Cas9 v3

图 1C

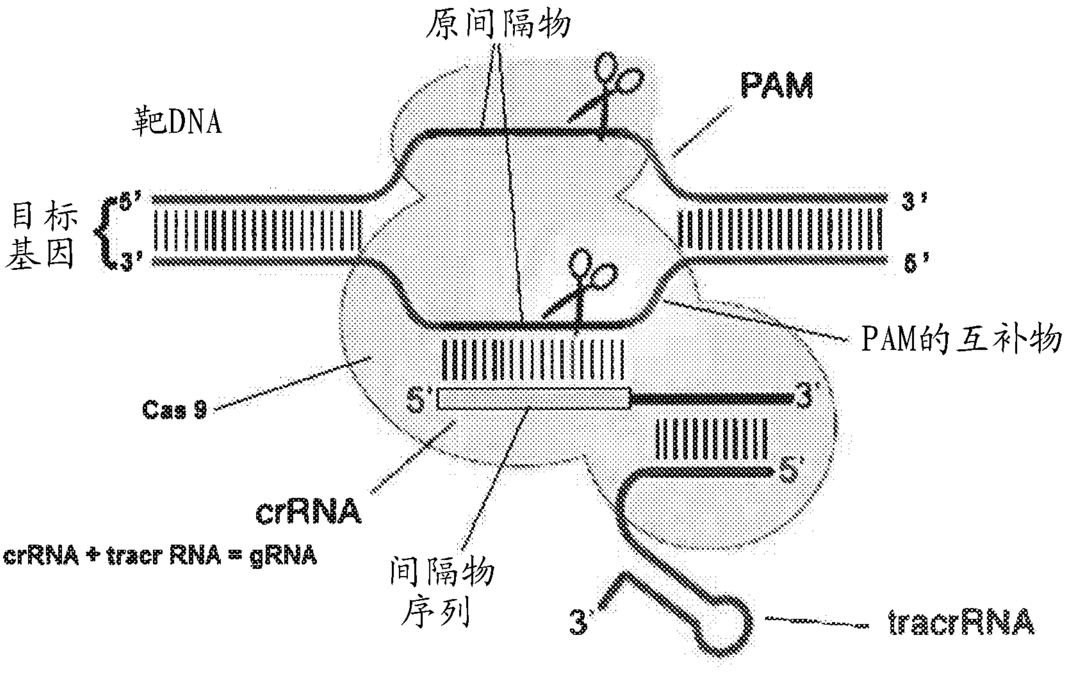

原间隔物
PAM
靶DNA
目标基因
5'
3'
3'
5'
PAM的互补物
Cas 9
5'
3'
5'
crRNA
crRNA + tracr RNA = gRNA
间隔物序列
3'
tracrRNA

图 2A

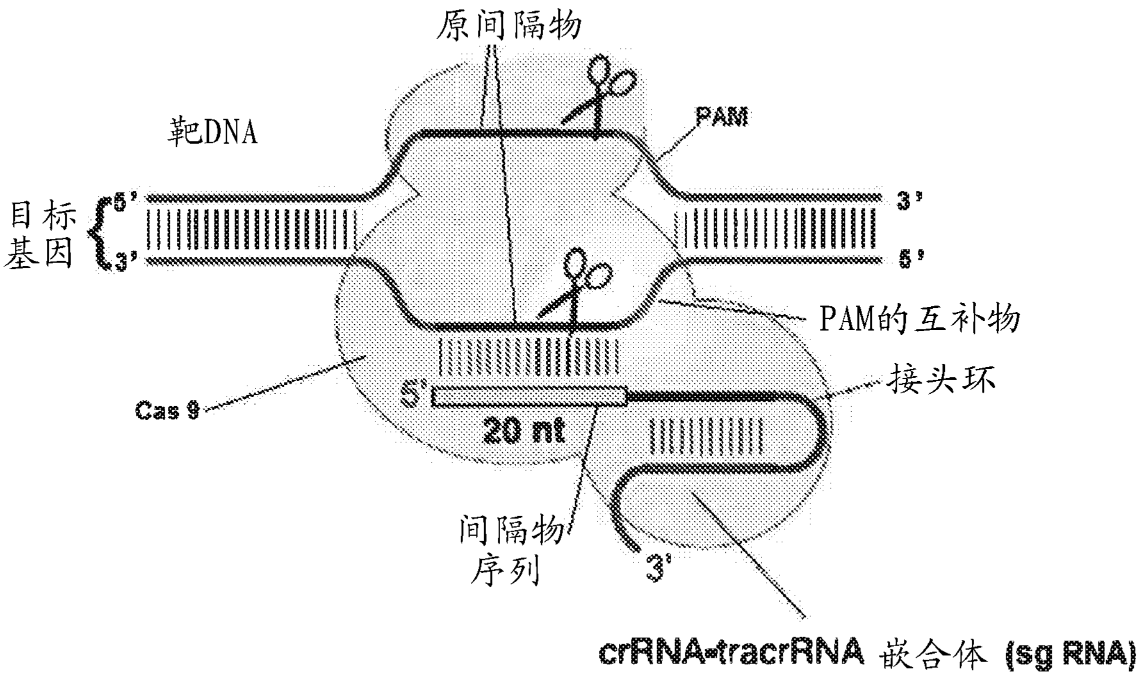

原间隔物
靶DNA
PAM
目标基因
5'
3'
3'
5'
PAM的互补物
接头环
Cas 9
5'
20 nt
间隔物序列
3'
crRNA-tracrRNA 嵌合体 (sg RNA)

图 2B

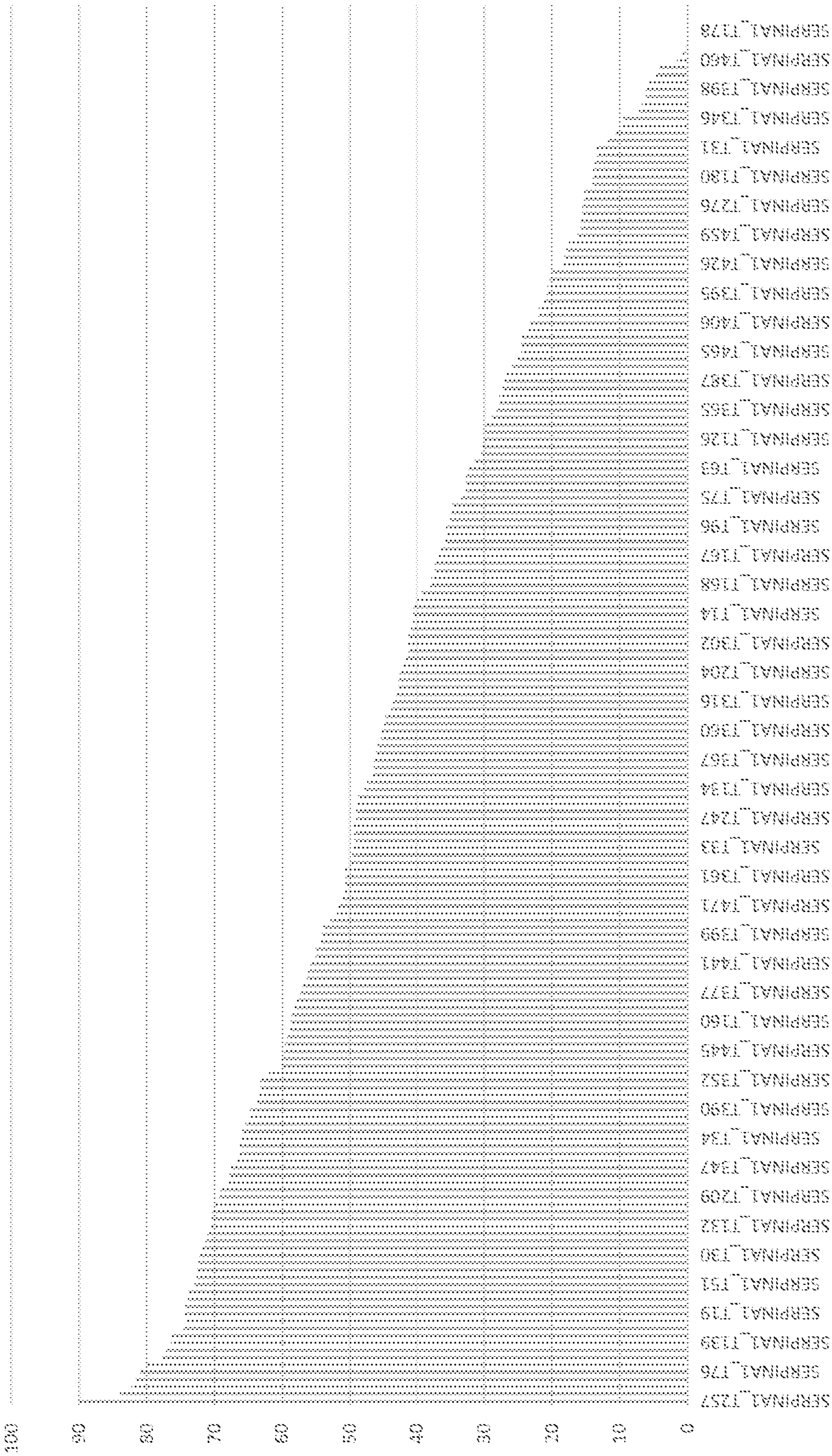

SERPINA1_T178
SERPINA1_T460
SERPINA1_T398
SERPINA1_T346
SERPINA1_T31
SERPINA1_T180
SERPINA1_T276
SERPINA1_T429
SERPINA1_T426
SERPINA1_T395
SERPINA1_T406
SERPINA1_T465
SERPINA1_T387
SERPINA1_T365
SERPINA1_T126
SERPINA1_T63
SERPINA1_T75
SERPINA1_T96
SERPINA1_T167
SERPINA1_T168
SERPINA1_T14
SERPINA1_T302
SERPINA1_T204
SERPINA1_T316
SERPINA1_T360
SERPINA1_T367
SERPINA1_T134
SERPINA1_T247
SERPINA1_T33
SERPINA1_T361
SERPINA1_T471
SERPINA1_T399
SERPINA1_T441
SERPINA1_T377
SERPINA1_T160
SERPINA1_T445
SERPINA1_T352
SERPINA1_T390
SERPINA1_T34
SERPINA1_T347
SERPINA1_T209
SERPINA1_T132
SERPINA1_T30
SERPINA1_T51
SERPINA1_T19
SERPINA1_T139
SERPINA1_T76
SERPINA1_T257
0
10
20
30
40
50
60
70
80
90
100

图 3