(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119278267 A
(43)申请公布日 2025.01.07

(21)申请号 202380037194.X
(22)申请日 2023.04.28
(30)优先权数据
22170965.2 2022.04.29 EP
(85)PCT国际申请进入国家阶段日
2024.10.29
(86)PCT国际申请的申请数据
PCT/EP2023/061394 2023.04.28
(87)PCT国际申请的公布数据
WO2023/209227 EN 2023.11.02
(71)申请人 庞培法布拉大学
地址 西班牙巴塞罗纳
(72)发明人 M•盖尔卡戈尔
A•桑切斯-梅希亚斯加西亚
D•伊万西奇
(74)专利代理机构 隆天知识产权代理有限公司 72003
专利代理师 付文川
(51)Int.Cl.
C12N 15/86(2006.01)
C12N 15/87(2006.01)
C07K 14/16(2006.01)

权利要求书2页 说明书46页
序列表（电子公布） 附图15页

(54)发明名称

利用慢病毒载体的靶向DNA整合及其用途

(57)摘要

本发明涉及一种慢病毒颗粒组合物,其包含突变的整合酶和RNA-引导的核酸酶,用于所靶向的基因插入。

CN 119278267 A

1.组合物,包含:

(i)编码具有整合活性受损的慢病毒载体的核酸,可选地包含适体结合蛋白,

(ii)编码包含RNA引导的核酸酶或切口酶的多肽或蛋白的核酸,

(iii)编码与适体融合的引导RNA(gRNA)的核酸,

(iv)编码目标转基因的核酸,以及

(v)编码融合蛋白的核酸,所述融合蛋白包含:

-第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,

-第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及

-可选地,第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成。

2.根据权利要求1所述的组合物,其中:

-所述适体结合蛋白是MS2噬菌体衣壳蛋白(MCP),优选与SEQ ID NO:61具有至少75%的同一性,并且

-所述适体是MS2 RNA四环结合序列,优选与SEQ ID NO:63具有至少75%的同一性。

3.根据权利要求1或2所述的组合物,其中所述RNA引导的核酸酶或切口酶不与整合酶融合。

4.根据权利要求1至3中任一项所述的组合物,其中所述RNA引导的核酸酶或切口酶是Cas蛋白。

5.根据权利要求1至4中任一项所述的组合物,其中所述RNA引导的核酸酶或切口酶与突变的高活性PiggyBac转座酶结合。

6.根据权利要求1至5中任一项所述的组合物,其中(i)的整合酶和/或(v)的POL多聚蛋白中包含的整合酶当存在时,在选自以下位置处包含至少一个氨基酸突变:SEQ ID NO:1的HIV-1整合酶的10、11、13、64、94、116、117、119、120、122、124、128、152、164、168、170、185、186、231、264、266,或在另一整合酶中的相应位置处。

7.根据权利要求1至6中任一项所述的组合物,其中(i)和/或(v)的整合酶包含至少一个选自以下的氨基酸突变:SEQ ID NO:1的HIV-1整合酶的D10K、E11K、E13K、D64A、D64E、G94D、G94E、G94R、G94K、D116A、D116E、N117D、N117E、N117R、N117K、S119A、S119P、S119T、S119G、S119D、S119E、S119R、S119K、N120D、N120E、N120R、N120K、T122K、T122I、T122V、T122A、T122R、A124D、A124E、A124R、A124K、A128T、E152A、E152D、D164N、Q168L、Q168A、E170G、F185K、K186E、R231G、R231K、R231D、R231E、R231S、K264R、K266R和K273R,或在另一整合酶的相应位置处。

8.根据权利要求1至7中任一项所述的组合物,其中(i)的慢病毒载体的整合活性通过整合酶序列中的至少一个突变而降低或消除,所述突变选自:D116N取代,编号基于SEQ ID NO:1的HIV-1整合酶;D164N取代,编号基于SEQ ID NO:1的HIV-1整合酶;早熟终止密码子插入;以及整合酶基因的缺失(ΔIN)。

9.慢病毒颗粒的制备方法,包括以下步骤:

(i)用权利要求1至8中任一项所述的组合物转染在合适的培养基中培养的慢病毒生产细胞,和

(ii)收集所述慢病毒生产细胞的培养基中的慢病毒颗粒。

10.慢病毒颗粒群,其可通过权利要求9所述的方法获得。

11.包含慢病毒颗粒的慢病毒颗粒群,所述慢病毒颗粒包含以下或由以下组成:

a.慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,

b.RNA引导的核酸酶或切口酶,优选Cas9蛋白,或编码其的核酸序列,

c.引导RNA,其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,以及

d.融合蛋白,其包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及可选地,(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,或编码所述融合蛋白的核酸序列。

12.将目标转基因位点特异性整合到细胞的基因组中的体外方法,所述方法包括用权利要求10或11所述的慢病毒颗粒群感染所述细胞,其中所述慢病毒颗粒群包含目标转基因,或其中目标转基因在慢病毒颗粒群之前、同时或之后递送到细胞。

13.根据权利要求10或11所述的慢病毒颗粒群,用作药物。

14.根据权利要求10或11所述的慢病毒颗粒群,用于将目标转基因位点特异性整合到细胞的基因组中,其中所述慢病毒颗粒群包含目标转基因,或其中目标转基因要在慢病毒颗粒群之前、同时或之后递送到细胞。

15.根据权利要求10或11所述的慢病毒颗粒群,用于治疗有需要的受试者的疾病,优选基因疾病,其中所述慢病毒颗粒群包含目标转基因,或其中目标转基因要在慢病毒颗粒群之前、同时或之后递送到细胞。

利用慢病毒载体的靶向DNA整合及其用途

发明领域

[0001] 本发明涉及基因编辑和基因治疗领域,特别是靶向基因插入。

[0002] 发明背景

[0003] 病理性基因缺陷可以是几个碱基到大的缺失。哺乳动物基因组中基因大小的片段的靶向DNA整合可以实现治疗信息和新功能的精确添加。已经开发出基于非同源末端连接(NHEJ)的精确基因递送方法,例如不依赖于同源性的靶向整合(HITI)。这种方法已被证明可以插入几千个碱基,但对于非常大的编辑仍然效率低下。虽然HITI可能可以递送外显子,但它的效率可能不足以稳健递送基因例如DMD(~14kb)或ABCA4(~6.8kb)的cDNA。最近,已通过融合DNA结合结构域扩展了HITI以提高DNA效率。在细菌中,已经证明了使用CRISPR可编程转座子进行精确基因递送,但该技术尚未用于哺乳动物细胞。

[0004] 基因组编辑新方法的开发使得人类基因治疗(HGT)应用得以扩展。许多HGT策略都基于DNA载荷的加入,该载荷可以弥补基因缺陷和/或为受体细胞提供合成特性。传统上,治疗性载荷的递送基于病毒和非病毒载体,这些载体会不受控制地整合到基因组中。虽然精确的递送技术正在出现,但病毒载体介导的不受控制的递送已使大量患者受益。已经观察到插入性突变事件,这些事件导致基因治疗试验的首次尝试在白血病病例与基于逆转录病毒的SCID-X1基因治疗相关联后被叫停。最近在CAR-T治疗中再次观察到载体驱动的克隆扩增,并且在与基于逆转录病毒的药物Strimvelis相关联的白血病病例中也观察到了这种现象。rAAV可引起插入性突变并诱导克隆扩增。PiggyBac转座酶(PB)等无病毒技术与致癌诱导有关。慢病毒载体在治疗上用于不受控制地插入转基因,用于离体应用和体内应用。我们观察到通过突变整合酶蛋白并在慢病毒颗粒中整合SpCas9核酸酶,慢病毒载体可以被重编程以精确整合转基因DNA。

[0005] 使用ZFN、TALEN和CRISPR Cas9等工具发现和实施精确的递送策略,有可能避免不希望的基因破坏或激活,并能同时进行基因的失活和进入,这对于治疗性基因组调节是关键特征。此类技术也引发了与安全相关的担忧。据报道,载荷的整合可能不精确,产生不同的意外整合结果,例如供体在脱靶的切割位点处整合以及诸如倒位或易位的基因组重排。

[0006] 因此,对精确且安全的基因递送策略,特别是非常大的序列的递送,存在尚未满足的需求。

发明内容

[0007] 本发明涉及一种组合物,其包含:

[0008] (i)编码具有整合活性受损的慢病毒载体的核酸,可选地包含适体结合蛋白,

[0009] (ii)编码包含RNA引导的核酸酶或切口酶的多肽或蛋白的核酸,

[0010] (iii)编码与适体融合的引导RNA(gRNA)的核酸,

[0011] (iv)编码目标转基因的核酸,以及

[0012] (v)可选地,编码融合蛋白的核酸,所述融合蛋白包含:

[0013] -第一蛋白,包含GAG多聚蛋白或由GAG多聚蛋白组成,

[0014] -第二蛋白,包含适体结合蛋白或由适体结合蛋白组成,以及

[0015] -可选地,第三蛋白,包含POL多聚蛋白或由POL多聚蛋白组成。

[0016] 在一些实施方案中,适体结合蛋白是MS2噬菌体衣壳蛋白(MCP),优选与SEQ ID NO:61具有至少75%同一性;并且适体是MS2 RNA四环结合序列,优选与SEQ ID NO:63具有至少75%同一性。

[0017] 在一些实施方案中,RNA引导的核酸酶或切口酶不与整合酶融合。

[0018] 在一些实施方案中,RNA引导的核酸酶或切口酶是Cas蛋白。

[0019] 在一些实施方案中,RNA引导的核酸酶或切口酶与突变的高活性PiggyBac转座酶结合。

[0020] 在一些实施方案中,(i)的整合酶和/或(v)的POL多聚蛋白中包含的整合酶当存在时包含选自以下的位置处的至少一个氨基酸突变:SEQ ID NO:1的HIV-1整合酶的10、11、13、64、94、116、117、119、120、122、124、128、152、164、168、170、185、186、231、264、266,或在另一整合酶中的相应位置处。

[0021] 在一些实施方案中,(i)和/或(v)的整合酶包含至少一个选自以下的氨基酸突变:SEQ ID NO:1的HIV-1整合酶的D10K、E11K、E13K、D64A、D64E、G94D、G94E、G94R、G94K、D116A、D116E、N117D、N117E、N117R、N117K、S119A、S119P、S119T、S119G、S119D、S119E、S119R、S119K、N120D、N120E、N120R、N120K、T122K、T122I、T122V、T122A、T122R、A124D、A124E、A124R、A124K、A128T、E152A、E152D、D164N、Q168L、Q168A、E170G、F185K、K186E、R231G、R231K、R231D、R231E、R231S、K264R、K266R和K273R,或另一整合酶中的相应位置处。

[0022] 在一些实施方案中,(i)的慢病毒载体的整合活性通过整合酶序列中的至少一个突变而降低或消除,所述突变选自:D116N取代、D164N取代、插入早熟终止密码子以及整合酶基因的缺失(ΔIN),编号基于SEQ ID NO:1的HIV-1整合酶。

[0023] 本发明还涉及一种生产慢病毒颗粒的方法,包括以下步骤:

[0024] (i)用本发明的组合物转染在合适培养基中培养的慢病毒生产细胞,以及

[0025] (ii)收集所述慢病毒生产细胞的培养基中的慢病毒颗粒。

[0026] 本发明还涉及可通过本发明的生产慢病毒颗粒的方法获得的慢病毒颗粒群。

[0027] 本发明还涉及慢病毒颗粒群,其包含慢病毒颗粒,所述慢病毒颗粒包含以下或由以下组成:

[0028] a.慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,

[0029] b.RNA引导的核酸酶或切口酶,优选Cas9蛋白,或编码其的核酸序列,

[0030] c.引导RNA,其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,以及

[0031] d.可选地,融合蛋白,包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及可选地(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,或编码其的核酸序列。

[0032] 本发明还涉及一种将目标转基因位点特异性整合到细胞的基因组中的体外方法,所述方法包括用本发明的慢病毒颗粒群感染所述细胞,其中慢病毒颗粒群包含目标转基因,或其中目标转基因在慢病毒颗粒群之前、同时或之后递送到细胞。

[0033] 本发明还涉及本发明的慢病毒颗粒群,用作药物。

[0034] 本发明还涉及本发明的慢病毒颗粒群,用于将目标转基因位点特异性整合到细胞

的基因组中,其中慢病毒颗粒群包含目标转基因,或其中目标转基因将在慢病毒颗粒群之前、同时或之后递送到细胞。

[0035] 本发明还涉及本发明的慢病毒颗粒群,用于治疗有需要的受试者的疾病,优选基因疾病,其中慢病毒颗粒群包含目标转基因,或其中目标转基因将在慢病毒颗粒群之前、同时或之后递送至细胞。

[0036] 定义

[0037] 在本公开中,任何使用单数形式“a”、“an”或“the”均包括单数指代,也包括复数指代,除非上下文另有明确说明。因此,例如,对“药剂”的指代包括单个药剂和多个这样的药剂。

[0038] “核酸(序列)”和“核苷酸序列”可互换使用用于指由单体核苷酸组成或包含单体核苷酸的任何分子。核酸可以是寡核苷酸或多核苷酸;它可以是DNA、RNA或两者的混合。它可以是化学修饰的或人工的;例如,它包括肽核酸(PNA)、吗啉代(morpholinos)和锁核酸(LNA),以及乙二醇核酸(GNA)和苏糖核酸(TNA)。这些核酸都是通过分子骨架的变化而区别于天然存在的DNA或RNA。此外,可以使用硫代磷酸核苷酸。其他脱氧核苷酸类似物包括但不限于甲基膦酸酯、磷酰胺、二硫代磷酸酯、N3’P5’磷酰胺和寡核糖核苷酸硫代磷酸酯及其2’O-烯丙基类似物和2’O-甲基核糖核苷酸甲基膦酸酯,其可用于本公开的核酸中。

[0039] “多肽”、“肽”、“蛋白”和“氨基酸序列”可互换使用用于指氨基酸残基的聚合物。除非另有说明,氨基酸残基的聚合物可以是任意长度。该术语还适用于其中一个或多个氨基酸是相应天然存在的氨基酸的化学类似物或修饰衍生物的氨基酸聚合物。

[0040] “核酸酶”是指催化核酸序列内核酸水解的酶。核酸酶活性可导致单链或双链核酸分子的断裂,其中核酸分子可以是DNA或RNA。

[0041] “Cas9”或“Cas9核酸酶”指RNA引导的核酸酶,包含Cas9蛋白或其片段(例如,包含Cas9的活性或非活性DNA切割结构域和/或Cas9的gRNA结合结构域的蛋白)。Cas9核酸酶有时也称为casn1核酸酶或CRISPR(成簇规律间隔短回文重复序列)结合核酸酶。CRISPR是一种适应性免疫系统,可产生对抗移动遗传元件(病毒、转座元件和接合质粒)的保护作用。CRISPR簇包含间隔区(与先前的移动元件互补的序列),并且靶向入侵的核酸。CRISPR簇被转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,pre-crRNA的正确加工需要反式编码小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA可作为核糖核酸酶3辅助的pre-crRNA加工的向导。随后,Cas9/crRNA/tracrRNA复合物通过内切核酸酶切割与间隔区互补的线性或环状dsDNA靶标。不与crRNA互补的靶标链首先通过内切核酸酶被切割,然后通过外切核酸酶进行3’-5’修剪。在自然界中,DNA结合和切割通常需要蛋白和两种RNA。然而,可以对单引导RNA(“sgRNA”或简称为“gRNA”)进行工程化,以将crRNA和tracrRNA两者的各个方面整合到单RNA分子中。

[0042] Cas9识别CRISPR重复序列中的短基序(PAM或原型间隔区相邻基序),以帮助区分自我和非我。Cas9核酸酶序列和结构是本领域技术人员所熟知的。Cas9直系同源物已在各种物种中有描述,包括但不限于酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开,本领域技术人员将清楚其他合适的Cas9核酸酶和序列,并且此类Cas9核酸酶和序列包括来自Chylinski等,2013(RNA Biol.10(5):726-37)(其全部内容通过引用并入本文)中公开的生物体和基因座的Cas9序列。

[0043] 在一些实施方案中,Cas9核酸酶具有无活性的(例如,失活的)DNA切割结构域。核酸酶失活的Cas9蛋白可以互换地称为“dCas9”蛋白(指核酸酶“死”的Cas9)。本领域已知用于生成具有无活性DNA切割结构域的Cas9蛋白(或其片段)的方法(参见,例如,Jinek等,2012.Science.337(6096):816-821;Qi等,2013.Cell.152(5):1173-83,其整体内容通过引用并入本文)。

[0044] “外源”是指任何不天然存在于目标细胞或生物体中,但可以通过一种或多种基因、生物化学或其他方法引入其中的分子。分子在细胞或生物体中的天然存在也可以根据其特定发育阶段和环境条件来确定。因此,例如,仅在肌肉胚胎发育期间存在的分子对于成年肌肉细胞而言是外源分子。类似地,由热休克诱导的分子对于未热休克的细胞而言是外源分子。外源分子可以包括例如功能失调的内源分子的有功能版本或正常功能的内源分子的功能失调版本。相比之下,术语“内源”是指在特定环境条件下在特定发育阶段正常存在于细胞或生物体中的任何分子。

[0045] “融合蛋白”是指包含两个或更多个融合在一起的氨基酸序列(即来自两个或更多个不同的蛋白和/或肽)的单链杂合多肽。两个或更多个氨基酸序列可以通过直接肽键融合在一起,也可以通过肽接头间接融合在一起。融合蛋白可以特别地完全由单个核酸序列编码。

[0046] “基因”通常指编码蛋白的DNA区域(即编码区)。该术语还可包括本身不编码蛋白的DNA区域(即非编码区)。后者包括例如转录成功能性非编码RNA分子(例如转移RNA、核糖体RNA、调节性RNA等)的区域。其他非编码区调节编码区的转录和翻译(即调节元件),或用作建构元件(例如支架/基质附着区)、DNA复制起点、着丝粒或端粒等。调节元件包括但不限于启动子序列、终止子、翻译调节序列(例如核糖体结合位点[RBS]和内部核糖体进入位点[IRES])、增强子、沉默子、绝缘子、边界元件、复制起点、基质附着位点和基因座控制区。

[0047] “真核”是指细胞或由此类细胞组成的生物体,这些细胞不是细菌细胞(原核细胞)。真核细胞包括但不限于真菌细胞、植物细胞和动物细胞(例如,哺乳动物细胞和人类细胞)。

[0048] “转染”及其任何词形变化是指通过非病毒方式将一个或几个核酸分子(DNA和/或RNA)引入一个或多个细胞,无论是体外还是体内。转染方法是本领域公知的,包括例如脂质转染和电穿孔。

[0049] “转导”及其任何词形变化是指利用病毒载体,例如病毒、病毒颗粒或病毒载体,包括但不限于逆转录病毒(包括慢病毒)、腺病毒、腺相关病毒(AAV)及其衍生的载体,将一种或多种核酸分子(DNA和/或RNA)引入一个或多个细胞。

[0050] “切割”是指DNA分子共价骨架的断裂。切割可通过多种方法启动,包括但不限于磷酸二酯键的酶促或化学水解。单链切割和双链切割均有可能,双链切割可由两个不同的单链切割事件引起。DNA切割可导致产生平末端或交错末端。

[0051] “特异性”是指选择性结合与选定序列具有一定序列同一性的序列的能力。

[0052] “插入”和“整合”是指将核酸序列加入到第二核酸序列或基因组或其部分中。插入可以是“特异的”、“位点特异性的”、“靶向的”或“中靶的(on-targeted)”:这些形容词定义将核酸插入第二核酸的特定位点或基因组或其部分的特定位点(即,特意选择用于插入的位点)。与之相比,形容词“随机的”、“非靶向的(non-targeted)”或“脱靶的(off-

targeted)”是指将核酸非特异性和/或非预期地插入不希望的位点。术语“总计”或“全部”是指插入的总数。

[0053] “突变”是指将序列(例如核酸或氨基酸序列)内的一个残基用另一个残基取代;和/或是指核酸或氨基酸序列内一个或多个残基的缺失或插入。突变通常在本文中通过表明原始残基、随后是该残基在序列内的位置、然后表明新取代的残基来描述。本文提供的进行氨基酸取代(突变)的各种方法都是本领域公知的,并且由例如Green&Sambrook,2012(Molecular cloning:alaboratory manual(4th Ed.).Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.)提供。

[0054] “整合酶”是指由逆转录病毒(例如HIV)基因组产生的酶,可催化核酸序列在宿主细胞基因组内的整合。与瞬时表达的核酸序列(例如转染的质粒)相反,整合的核酸序列由细胞稳定表达而无论有丝分裂事件如何,并与基因组的其余部分一起复制。

[0055] “转座酶”是指结合到转座子末端并通过剪切粘贴机制或复制转座机制催化其移动到基因组另一部分的酶。

[0056] “突变的”,指代序列(例如氨基酸序列或核酸序列)时是指该序列与参考序列(例如野生型序列)不同。通常,与参考序列(例如相应的野生型序列)相比,突变序列包含以下至少一种:一个或多个残基的取代、添加或缺失。

[0057] “接头”是指连接两个相邻分子或部分的化学基团或分子。

[0058] “同一性(identity)”或“同一(identical)”在用于两个或更多个氨基酸序列或两个或更多个核酸序列之间的序列关系时,是指氨基酸序列或核酸序列之间的序列关联程度,由两个或更多个氨基酸残基或核酸残基的链条之间的匹配数决定。“同一性”测量两个或更多个序列中较小序列之间的相同匹配百分比,其中空位比对(如果有的话)由特定数学模型或计算机程序(即“算法”)解决。相关氨基酸序列或核酸序列的同一性可以通过已知方法容易地计算。此类方法包括但不限于Lesk A.M.(1988).Computational molecular biology:Sources and methods for sequence analysis.New York,NY:Oxford University Press;Smith D.W.(1993).Biocomputing:Informatics and genome projects.San Diego,CA:Academic Press;Griffin A.M.&Griffin H.G.(1994).Computer analysis of sequence data,Part 1.Totowa,NJ:Humana Press;von Heijne G.(1987).Sequence analysis in molecular biology:treasure trove or trivial pursuit.San Diego,CA:Academic press;Gribskov M.R.&Devereux J.(1991).Sequence analysis primer.New York,NY:Stockton Press;Carillo et al.,1988.SIAM J Appl Math.48(5):1073-82中描述的方法。确定同一性的优选方法设计为提供所测序列之间的最大匹配。确定同一性的方法在公开的计算机程序中有描述。确定两个序列之间同一性的优选计算机程序方法包括GCG程序包,包括GAP(Genetics Computer Group,University of Wisconsin,Madison,WI;Devereux etal.,1984.Nucleic Acids Res.12(1Pt 1):387-95)、BLASTP、BLASTN和FASTA(Altschul et al.,1990.J Mol Biol.215(3):403-10)。BLASTX程序可从美国国家生物技术信息中心(NCBI)和其他来源(BLAST手册,Altschul et al.NCB/NLM/NIH Bethesda,Md.20894)公开获得。公知的Smith Waterman算法也可用于确定同一性。

[0059] “受试者”是指哺乳动物,优选人类。受试者可以是“患者”,即温血动物,更优选人类,其正在等待接受医疗护理或正在接受医疗护理或曾经/现在/将成为医疗程序的对象,

或正在被监测疾病的发展。术语“哺乳动物”在此是指任何哺乳动物,包括人类、家养和农场动物以及动物园、运动或宠物动物,例如狗、猫、牛、马、绵羊、猪、山羊、兔子等。优选地,哺乳动物是灵长类动物,更优选人类。

[0060] “治疗”、“缓解”、“治愈”及其任何词形变化均指治疗性治疗,不包括预防性或防止性措施;其目的是减缓、减轻、停止甚至逆转(部分或全部)目标病理状况或病症的发展。需要治疗的者包括已患有该病症的以及疑似患有该病症的。如果在接受治疗后受试者显示与目标病理状况或病症相关的一种或多种症状的可观察到的和/或可测量的减少或消失;在一定程度上的缓解;发病率和/或死亡率降低;和/或生活质量问题得到改善,则该受试者针对所述病理状况或病症被成功“治疗”。用于评估疾病的成功治疗和改善的上述参数可以通过医生熟悉的常规程序容易地测量。

[0061] “预防”及其任何词形变化指预防性或防止性措施,其中目的是减少受试者在给定时间段内发展给定病理状况或病症的机率。这种减少可以反映为例如受试者的病理状况或病症的至少一种症状的延迟发作。

[0062] 发明详述

[0063] 本发明涉及一种组合物,包含:

[0064] (i)编码具有整合活性受损的慢病毒载体的核酸,所述慢病毒载体可选地包含适体结合蛋白,

[0065] (ii)编码包含RNA引导的核酸酶或切口酶的多肽或蛋白的核酸,

[0066] (iii)编码与适体融合的引导RNA(gRNA)的核酸,

[0067] (iv)编码目标转基因的核酸,以及

[0068] (v)可选地,编码融合蛋白的核酸,所述融合蛋白包含:

[0069] -第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,

[0070] -第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成;以及

[0071] -可选地,第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成。

[0072] 在一些实施方案中,组合物中的核酸是脱氧核糖核酸(DNA)、核糖核酸(RNA)或其组合。在一些实施方案中,组合物中的核酸是DNA。在一些实施方案中,组合物中的核酸是RNA。

[0073] 在一些实施方案中,组合物的核酸包含天然核苷酸、非天然核苷酸或其组合,或由以上组成。在一些实施方案中,组合物的核酸包含天然核苷酸或由天然核苷酸组成。在一些实施方案中,天然核苷酸选自包含以下或由以下组成的组:腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶和尿嘧啶。在一些实施方案中,天然核苷酸选自包含以下或由以下组成的组:腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶。在一些实施方案中,天然核苷酸选自包含以下或由以下组成的组:腺嘌呤、鸟嘌呤、胞嘧啶和尿嘧啶。在一些实施方案中,组合物的核酸包含非天然核苷酸(例如,化学修饰的)或由其组成。

[0074] 在一些实施方案中,组合物中的核酸是单链、双链或它们的组合。在一些实施方案中,组合物中的核酸是单链。在一些实施方案中,组合物中的核酸是双链。

[0075] 在一些实施方案中,组合物中的核酸选自包含以下或由以下组成的组:质粒、信使RNA(mRNA)、裸DNA、粘粒、fosmid、原核染色体(例如,细菌人工染色体)、真核染色体(例如,酵母人工染色体或人类人工染色体)或它们的组合。在一些实施方案中,组合物中的核酸是

质粒、mRNA或它们的组合。在一些实施方案中,组合物中的核酸是质粒。在一些实施方案中,组合物中的核酸是mRNA。

[0076] 在一些实施方案中,组合物中的每种核酸均包含在其自己的载体中(例如,质粒、粘粒、fosmid、染色体等)。或者,组合物中的两种核酸可包含在单个载体中。例如,(i)和(ii)、或(i)和(iii)、或(i)和(iv)、或(i)和(v)、或(ii)和(iii)、或(ii)和(iv)、或(ii)和(v)、或(iii)和(iv)、或(iii)和(v)、或(iv)和(v)的核酸可包含在单个载体中。或者,组合物中的三种核酸可包含在单个载体中。例如,(i)和(ii)和(iii)、或(i)和(ii)和(iv)、或(i)和(ii)和(v)、或(i)和(iii)和(iv)、或(i)和(iii)和(v)、或(i)和(iv)和(v)、或(ii)和(iii)和(iv)、或(ii)和(iii)和(v)、或(ii)和(iv)和(v)、或(iii)和(iv)和(v)的核酸可包含在单个载体中。或者,组合物的四种核酸可包含在单个载体中。例如,(i)和(ii)和(iii)和(iv)、或(i)和(ii)和(iii)和(v)、或(i)和(ii)和(iv)和(v)、或(i)和(iii)和(iv)和(v)、或(ii)和(iii)和(iv)和(v)的核酸可包含在单个载体中。或者,组合物的五种核酸可包含在单个载体中。

[0077] 在一些实施方案中,组合物中的核酸包含至少一个编码序列,和可选的至少一个非编码序列,例如调控序列或复制起点序列。

[0078] 在一些实施方案中,组合物中的核酸编码蛋白(包括翻译后切割和/或修饰的多聚蛋白和前多肽)、核酸或其组合。换句话说,组合物中的核酸包含至少一个表达盒。

[0079] 本文进一步描述了由组合物中的核酸编码的蛋白或核酸。

[0080] 根据本发明,所述组合物包含整合活性受损的慢病毒载体。

[0081] 本文使用的表述“整合活性受损”是指与野生型慢病毒载体相比,慢病毒载体的整合酶的整合活性降低或消失。优选地,慢病毒载体与野生型慢病毒载体相比,整合活性被消除。为了实现整合活性的降低或消除,对慢病毒载体的整合酶进行基因修饰,例如,利用一个或几个氨基酸取代改变其整合活性,或者通过部分或完全缺失整合酶编码序列。

[0082] 在一些实施方案中,与野生型(WT)整合酶相比(例如,与具有SEQ ID NO:1的野生型HIV-1整合酶相比),慢病毒载体的整合酶包含至少一个氨基酸突变。如本文所用,术语“至少一个”表示1、2、3、4、5、6、7、8、9、10或更多。如本文所用,术语“氨基酸突变”涵盖取代、添加、缺失等。

[0083] 在下文中,除非另有说明,氨基酸残基编号均基于具有SEQ ID NO:1的HIV-1整合酶的序列。本领域技术人员可以容易地确定其他整合酶的序列中的相应位置,例如通过序列比对。

[0084]

SEQ ID NO: 1 – HIV-1 整合酶

FLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEA
MHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAY
FLLKLAGRWPVKTVHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQ
GVIESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGE
RIVDIIATDIQTKELQKQITKIQNFRVYYRDSRDPVWKGPAKLLWKGEGA
VVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED

[0085] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的位置处的至少一个氨基酸突变:10、11、13、64、94、116、117、119、120、122、124、128、152、164、168、170、185、186、231、264、266和273。

[0086] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的至少一个氨基酸突变:E10K、E11K、E13K、D64V、D64A、D64E、G94D、G94E、G94R、G94K、D116N、D116A、D116E、N117D、N117E、N117R、N117K、S119A、S119P、S119T、S119G、S119D、S119E、S119R、S119K、N120D、N120E、N120R、N120K、T122K、T122I、T122V、T122A、T122R、A124D、A124E、A124R、A124K、A128T、E152V、E152A、E152D、D164N、Q168L、Q168A、E170G、F185K、K186Q、K186E、R231G、R231K、R231D、R231E、R231S、K264R、K266R和K273R。

[0087] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的位置处的至少一个氨基酸突变:10、13、64、116、119、152、164、186和231。在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的至少一个氨基酸突变:E10K、E13K、D64V、D64A、D64E、D116N、D116A、D116E、S119A、S119P、S119T、S119G、S119D、S119E、S119R、S119K、E152V、E152A、E152D、D164N、K186Q、K186E、R231G、R231K、R231D、R231E和R231S。

[0088] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的至少一个氨基酸突变:E10K、E13K、D64V、D116N、S119G、E152V、D164N、K186Q和R231G。

[0089] 在一些实施方案中,慢病毒载体的整合酶包含破坏与p75相互作用的至少一个突变。在一些实施方案中,慢病毒载体的整合酶包含选自包含双突变E10K/E13K和K186Q或由其组成的组的至少一个氨基酸突变。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变E10K和E13K,通常慢病毒载体的整合酶具有SEQ ID NO:2的氨基酸序列。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变K186Q,通常慢病毒载体的整合酶具有SEQ ID NO:8的氨基酸序列。

[0090] 在一些实施方案中,慢病毒载体的整合酶在其催化位点包含至少一个突变。在一些实施方案中,慢病毒载体的整合酶包含选自包含D64V、D116N和E152V或由其组成的组的至少一个氨基酸突变。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变D64V,通常慢病毒载体的整合酶具有SEQ ID NO:3的氨基酸序列。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变D116N,通常慢病毒载体的整合酶具有SEQ ID NO:4的氨基酸序列。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变E152V,通常慢病毒载体的整合酶具有SEQ ID NO:6的氨基酸序列。

[0091] 在一些实施方案中,慢病毒载体的整合酶包含至少一个突变,该突变使病毒整合重新定位而远离基因密集区域。在一些实施方案中,慢病毒载体的整合酶包含选自包含S119G和R231G或由其组成的组的至少一个氨基酸突变。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变S119G,通常慢病毒载体的整合酶具有SEQ ID NO:5的氨基酸序列。在一些实施方案中,慢病毒载体的整合酶包含氨基酸突变R231G,通常慢病毒载体的整合酶具有SEQ ID NO:9的氨基酸序列。

[0092] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的位置处的至少一个氨基酸突变:94、117、119、120、122、124和231。在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的至少一个氨基酸突变:G94D、G94E、

G94R、G94K、N117D、N117E、N117R、N117K、S119A、S119P、S119T、S119G、S119D、S119E、S119R、S119K、N120D、N120E、N120R、N120K、T122K、T122I、T122V、T122A、T122R、A124D、A124E、A124R、A124K、R231G、R231K、R231D、R231E和R231S。

[0093] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的位置处的至少一个氨基酸突变:264、266和273。在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的至少一个氨基酸突变:K264R、K266R和K273R。

[0094] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的位置处的至少一个氨基酸突变:10、13、64、116、128、152、168和170。在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的至少一个氨基酸突变:E10K、E13K、D64A、D64E、D116A、D116E、A128T、E152A、E152D、Q168L、Q168A和E170G。

[0095] 在一些实施方案中,慢病毒载体的整合酶包含至少一个位置168处的氨基酸突变。在一些实施方案中,慢病毒载体的整合酶包含至少一个氨基酸突变Q168L或Q168A。

[0096] 在一些实施方案中,慢病毒载体的整合酶包含选自包含以下或由以下组成的组的至少一个氨基酸突变:D116N取代、D164N取代、在其编码序列中插入早熟终止密码子以及其编码序列的缺失(ΔIN)。

[0097] 在一些实施方案中,慢病毒载体的整合酶具有SEQ ID NO:4的氨基酸序列。在一些实施方案中,慢病毒载体的整合酶具有SEQ ID NO:68的氨基酸序列。

[0098] 在一些实施方案中,编码慢病毒载体整合酶的核酸序列包含早熟终止密码子(即UAA、UAG或UGA)。在一些实施方案中,所得整合酶是截短的且无功能。在一些实施方案中,所得整合酶被降解,例如,通过细胞的质量控制系统降解。

[0099] 在一些实施方案中,编码整合酶的核酸序列从慢病毒载体的pol基因中缺失,称为“ΔIN”。

[0100] 在一些实施方案中,慢病毒载体编码选自POL、GAG和ENV的至少一种慢病毒多聚蛋白。在一些实施方案中,慢病毒载体编码GAG和POL。

[0101] 在一些实施方案中,慢病毒载体包含慢病毒载体基因组,所述慢病毒载体基因组包含选自gag和pol的至少一个基因。在一些实施方案中,慢病毒载体包含慢病毒载体基因组,所述慢病毒载体基因组包含选自gag、pol和vpr的至少一个基因。在一些实施方案中,慢病毒载体包含慢病毒载体基因组,所述慢病毒载体基因组包含选自gag、pol、env和vpr的至少一个基因。在一些实施方案中,慢病毒载体包含慢病毒载体基因组,所述慢病毒载体基因组包含选自gag、pol、env、vpr、vpu、vif、nef、tat和rev的至少一个基因。

[0102] 在一些实施方案中,慢病毒载体来源于人类免疫缺陷病毒1型(HIV-1)或人类免疫缺陷病毒2型(HIV-2)。在一些实施方案中,慢病毒载体来源于HIV-1。

[0103] 在一些实施方案中,慢病毒载体还包含至少一种非慢病毒蛋白,或编码其的核酸,特别是至少一种非慢病毒表面蛋白。在一些实施方案中,非慢病毒蛋白能够结合真核细胞的表面蛋白,优选哺乳动物细胞,更优选人类细胞。

[0104] 根据本发明,整合活性受损的慢病毒载体可选地包含适体结合蛋白。

[0105] 在一些实施方案中,适体结合蛋白与慢病毒载体的GAG多聚蛋白融合。在一些实施方案中,适体结合蛋白的C末端与GAG多聚蛋白的N末端融合,可选地通过接头融合。在一些实施方案中,适体结合蛋白的N末端与GAG多聚蛋白的C末端融合,可选地通过接头融合。

[0106] 在一些实施方案中,适体结合蛋白与慢病毒载体的POL多聚蛋白融合。在一些实施方案中,适体结合蛋白的C末端与POL多聚蛋白的N末端融合,可选地通过接头融合。在一些实施方案中,适体结合蛋白的N末端与POL多聚蛋白的C末端融合,可选地通过接头融合。

[0107] 在一些实施方案中,适体结合蛋白选自包含以下或由以下组成的组:MS2噬菌体衣壳蛋白(MCP)、PP7衣壳蛋白(PCP)、λN22肽和COM。

[0108] 在一些实施方案中,适体结合蛋白是MS2噬菌体衣壳蛋白(MCP)。

[0109] 在一些实施方案中,MCP与SEQ ID NO:61(例如,由具有SEQ ID NO:60的核酸序列编码)具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的同一性。

[0110]

SEQ ID NO: 60 – MS2 噬菌体衣壳蛋白 – DNA 编码序列

atggcttcaaactttactcagttcgtgctcgtggacaatggtgggacaggggatgtgacagtggctccttctaatttcgc
taatggggtggcagagtggatcagctccaactcacggagccaggcctacaaggtgacatgcagcgtcaggcagtct
agtgcccagaagagaaagtataccatcaaggtggaggtccccaaagtggctacccagacagtgggcggagtcgaa
ctgcctgtcgccgcttggaggtcctacctgaacatggagctcactatcccaattttcgctaccaattctgactgtgaact
catcgtgaaggcaatgcaggggctcctcaaagacggtaatcctatcccttccgccatcgccgctaactcaggtatcta
c

SEQ ID NO: 61 – MS2 噬菌体衣壳蛋白

[0111]

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSV
RQSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATN
SDCELIVKAMQGLLKDGNPIPSAIAANSGIY

[0112] 在一些实施方案中,适体结合蛋白是MS2噬菌体衣壳蛋白(MCP),优选与SEQ ID NO:61具有至少75%的同一性,并且适体是MS2 RNA四环结合序列,优选与SEQ ID NO:63具有至少75%的同一性。

[0113] 根据本发明,组合物包含编码包含RNA引导的核酸酶或切口酶的多肽或蛋白的核酸。

[0114] 在一些实施方案中,RNA引导的核酸酶或切口酶是包含活性DNA切割结构域和引导RNA结合结构域的核酸酶蛋白。

[0115] 在一些实施方案中,RNA引导的核酸酶或切口酶是Cas蛋白,包括但不限于酿脓链球菌Cas9(SpCas9)、金黄色葡萄球菌(Staphylococcus aureus)Cas9(SaCas9)或空肠弯曲杆菌(Campylobacter jejuni)Cas9(CjCas9)或其变体,切口酶Cas9(nCas9)、死Cas9(dCas9)、Cas12a蛋白、Cas12b蛋白、Cas12f蛋白、Cpf1蛋白或CasX蛋白,包括其变体和功能片段。

[0116] 在一些实施方案中,RNA引导的核酸酶或切口酶是Cas9蛋白,包括其变体和功能片段。

[0117] 在一些实施方案中,RNA引导的核酸酶或切口酶是Cas9蛋白。

[0118] 在一些实施方案中,Cas9蛋白与选自包含以下或由以下组成的组的蛋白的序列具

有至少80％、90％、95％、99％或至少100％的氨基酸序列同一性:SEQ ID NO:34的酿脓链球菌Cas9(SpCas9)、SEQ ID NO:35的金黄色葡萄球菌Cas9(SaCas9)、SEQ ID NO:36的Cpf1、SEQ ID NO:37的空肠弯曲杆菌Cas9(CjCas9)、SEQ ID NO:38的酿脓链球菌Cas9切口酶(nCas9)以及SEQ ID NO:39的金黄色葡萄球菌Cas9切口酶。

[0119] 在一些实施方案中,Cas9蛋白与选自包含以下或由以下组成的组的蛋白的序列具有至少80％、90％、95％、99％或至少100％的氨基酸序列同一性:SEQ ID NO:34的金黄色葡萄球菌Cas9(SaCas9)和SEQ ID NO:35的酿脓链球菌Cas9(SpCas9)。

[0120] 在一些实施方案中,Cas9蛋白是SEQ ID NO:34的金黄色葡萄球菌Cas9(SaCas9)。在一些实施方案中,Cas9蛋白是SEQ ID NO:35的酿脓链球菌Cas9(SpCas9)。

[0121] 在一些实施方案中,Cas9变体包含具有一个或多个氨基酸取代的Cas9蛋白的氨基酸序列。

[0122] 在一些实施方案中,RNA引导的核酸酶或切口酶是Cas9蛋白的变体或功能片段。

[0123] 在一些实施方案中,Cas9变体是人源化Cas9(hCas9)或其功能片段。本文所用的术语“人源化Cas9”或“hCas9”是指针对人类细胞进行序列优化的Cas9蛋白。在一些实施方案中,hCas9蛋白具有与SEQ ID NO:40具有至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或约100％序列同一性的氨基酸序列。

[0124] 在一些实施方案中,RNA引导的核酸酶或切口酶是CasX蛋白。在一些实施方案中,CasX具有与SEQ ID NO:41具有至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或约100％序列同一性的氨基酸序列。

[0125] 在一些实施方案中,RNA引导的核酸酶或切口酶是死Cas9蛋白。在一些实施方案中,死Cas9具有与SEQ ID NO:42具有至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或约100％序列同一性的氨基酸序列。

[0126] 在一些实施方案中,RNA引导的核酸酶或切口酶是TnpB(来自转座子PsiTn554的转座酶B)蛋白。在一些实施方案中,TnpB具有与SEQ ID NO:43具有至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或约100％序列同一性的氨基酸序列。

[0127] 在一些实施方案中,RNA引导的核酸酶或切口酶是Cas12f蛋白。在一些实施方案中,Cas12f蛋白来自细菌Acidibacillus sulfuroxidans(AsCas12f)。在一些实施方案中,Cas12f具有与SEQ ID NO:78具有至少约70％、至少约75％、至少约80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％或约100％序列同一性的氨基酸序列。

[0128] 在一些实施方案中,RNA引导的核酸酶或切口酶与转座酶结合。

[0129] 在一些实施方案中,RNA引导的核酸酶或切口酶与突变的高活性PiggyBac转座酶(hyPB)结合。在一些实施方案中,RNA引导的核酸酶或切口酶结合至与SEQ ID NO:10具有至少75％的氨基酸序列同一性的高活性PiggyBac转座酶。

[0130]

SEQ ID NO: 10 －高活性 PiggyBac 转座酶

MGSSLDDEHILSALLQSDDELVGEDSDSEVSDHVSEDDVQSDTEEAFIDEV
HEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKP
TRRSRVSALNIVRSQRGPTRMCRNIYDPLLCFKLFFTDEIISEIVKWTNAEIS
LKRRESMTSATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLFDRSLSM
VYVSVMSRDRFDFLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQ
NYTPGAHLTIDEQLLGFRGRCPFRVYIPNKPSKYGIKILMMCDSGTKYMIN
GMPYLGRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNL
LQEPYKLTIVGTVRSNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPK
PAKMVYLLSSCDEDASINESTGKPQMVMYYNQTKGGVDTLDQMCSVMT
CSRKTNRWPMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLY
MGLTSSFMRKRLEAPTLKRYLRDNISNILPKEVPGTSDDSTEEPVMKKRT
YCTYCPSKIRRKASASCKKCKKVICREHNIDMCQSCF

[0131] 如本文所用,术语“突变的hyPB”是指与氨基酸序列SEQ ID NO:10的hyPB相比,包含一个或多个氨基酸取代的转座酶,通常不超过1、2、3、4、5、6、7、8、9或10个氨基酸取代。更具体地说,突变的hyPB包含(i)一个或多个氨基酸取代,以与氨基酸序列SEQ ID NO:10的hyPB相比增加切割活性,和/或(ii)一个或多个氨基酸取代,以与氨基酸序列SEQ ID NO:10的hyPB相比降低DNA结合活性。在一些实施方案中,突变的hyPB包含与SEQ ID NO:10中所示的序列至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。

[0132] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以增加切割活性。

[0133] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以增加切割活性,所述氨基酸突变选自氨基酸位置编号[194-200]、[214-222]、[434-442]或[446-456]定义的区域内的氨基酸突变,例如位置D198、D201、R202、M212和/或S213处的氨基酸取代,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。

[0134] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以增加切割活性,所述氨基酸突变选自位置450、560、564、573、589、592和/或594处的氨基酸突变,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。

[0135] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以增加切割活性,所述氨基酸突变选自位置M194和/或D450处的氨基酸突变,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号,优选地选自M194V和/或D450N的氨基酸取代。

[0136] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以降低DNA结合活性。

[0137] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以降低DNA结合活性,所述氨基酸突变选自位置254、275、277、347、372、375和/或465处的氨基酸突变,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。

[0138] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以降低DNA结合活性,所述氨基酸突变选自R275、N347、R372、K375、R376、E377和E380,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。

[0139] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以降低DNA结合活性,所述氨基酸突变选自R372、K375、R376、E377和E380,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号,优选地选自氨基酸取代R372A、K375A、R376A、E377A和/或E380A。

[0140] 在一些实施方案中,突变的高活性PiggyBac包含一个或多个氨基酸突变以降低DNA结合活性,所述氨基酸突变选自N347、R372和K375,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号,优选地选自氨基酸取代N347S、N347A、R372A、K375A,更优选地选自氨基酸取代N347S、N347A。

[0141] 在一些实施方案中,突变的高活性PiggyBac包含如上文所述的一个或多个氨基酸突变以增加切割活性;以及如上文所述的一个或多个氨基酸突变以降低DNA结合活性。

[0142] 在一些实施方案中,突变的高活性PiggyBac包括位置D450处的至少一个氨基酸取代以增加切割活性,以及位置N347、R372和K375处的至少两个氨基酸取代以降低DNA结合活性,优选地,所述突变的高活性PiggyBac的转座酶包括双突变N347S和D450N或三重突变D450N、R372A和K375A,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。在更优选的实施方案中,突变的高活性PiggyBac的转座酶包括双突变N347S和D450N,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。

[0143] 在一些实施方案中,如先前实施方案所公开的突变的高活性PiggyBac还包含氨基酸位置编号[158-169]所定义的区域中的至少一个突变,例如A166S;和/或位置Y527、R518、K525、N463处的至少一个突变。

[0144] 通常,所述突变的高活性PiggyBac包含与SEQ ID NO:1的突变的高活性PiggyBac具有至少85%、至少90%、至少95%同一性或100%同一性的氨基酸序列。

[0145] 在一些实施方案中,所述突变的高活性PiggyBac是SEQ ID NO:10的高活性PiggyBac的变体,与SEQ ID NO:10相比具有一个或多个氨基酸取代,通常具有不超过1、2、3、4、5、6、7、8、9或10个氨基酸取代。

[0146] 在一些实施方案中,所述突变的高活性PiggyBac还包含以下氨基酸突变中的一个或多个:位置34、43、117、202、230、245、268、275、277、287、290、315、325、341、346、347、350、351、356、357、388、409、411、412、432、447、460、461、465、517、560、564、571、573、576、586、587、589、592和/或594处的氨基酸突变,位置编号对应于高活性PiggyBac序列(SEQ ID NO:10)的氨基酸编号。

[0147] 在一些实施方案中,所述突变的高活性PiggyBac包含以下突变或突变组合:V34M、T43I、Y177H、R202K、S230N、R245A、D268N、K287A、K290A、K287A/K290A、R315A、G325A、R341A、D346N、N347A、N347S、T350A、S351E、S351P、S351A、K356E、N357A、R388A、K409A、A411T、K412A、K432A、D447A、D447N、D450N、R460A、K461A、W465A、S517A、T560A、S564P、S571N、

S573A、K576A、H586A、I587A、M589V、S592G、或F594L、D450N/R372A/K375A、R275A/R277A 、K409A/K412A 、R460A/K461A 、R275A/R277A/N347S/K375A/T560A/S573A/M589V/S592G 和R245A/R275A/R277A/R372A/W465A。

[0148] 在一些实施方案中,所述突变的高活性PiggyBac包含以下氨基酸取代或氨基酸取代的组合:

[0149] -R372A/K375A/D450N,

[0150] -R372A/K375A/R376A/D450N,

[0151] -K375A/R376A/E377A/E380A/D450N,

[0152] -R372A/K375A/R376A/E377A/E380A/D450N,

[0153] -M194V,

[0154] -M194V/R372A/K375A,

[0155] -S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L,

[0156] -R245A/R275A/R277A/R372A/W465A/M589V,

[0157] -R275A/325A/R372A/T560A,

[0158] -N347A/D450N,

[0159] -N347S/D450N/T560A/S573A/F594L,

[0160] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,R275A/N347S/K375A/D450N/S592G,

[0161] -R275A/N347S/R372A/D450N/T560A/F594L,

[0162] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,

[0163] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,

[0164] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,

[0165] -V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P,

[0166] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,

[0167] -S230N/R277A/N347S/K375A/D450N,

[0168] -T43I/R372A/K375A/A411T/D450N,

[0169] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,和

[0170] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,

[0171] 位置编号对应于高活性PiggyBac序列(SEQ ID NO:10)的氨基酸编号。

[0172] 本公开使用的优选的突变高活性PiggyBac转座酶包括包含以下氨基酸取代组合的突变高活性PiggyBac:

[0173] -R372A/K375A/D450N,

[0174] -S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L,

[0175] -R245A/R275A/R277A/R372A/W465A/M589V,

[0176] -N347A/D450N,

[0177] -N347S/D450N/T560A/S573A/F594L,

[0178] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,

[0179] -R275A/N347S/K375A/D450N/S592G,

[0180] -R275A/N347S/R372A/D450N/T560A/F594L,

[0181] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,

[0182] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,

[0183] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,

[0184] -V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P,

[0185] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,

[0186] -S230N/R277A/N347S/K375A/D450N,

[0187] -T43I/R372A/K375A/A411T/D450N,

[0188] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,

[0189] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,和

[0190] -R275A/325A/R372A/T560A,

[0191] 位置编号对应于高活性PiggyBac序列(SEQ ID NO:10)的氨基酸编号。

[0192] 在一些实施方案中,所述突变的高活性PiggyBac包含以下氨基酸取代或氨基酸取代的组合:

[0193] -R245A/R275A/R277A/R372A/W465A/M589V,

[0194] -R275A/325A/R372A/T560A,

[0195] -N347A/D450N,

[0196] -N347S/D450N/T560A/S573A/F594L,

[0197] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,

[0198] -R275A/N347S/K375A/D450N/S592G,

[0199] -R275A/N347S/R372A/D450N/T560A/F594L,

[0200] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,

[0201] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,

[0202] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,

[0203] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,

[0204] -S230N/R277A/N347S/K375A/D450N,

[0205] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,和

[0206] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,

[0207] 位置编号对应于高活性PiggyBac序列(SEQ ID NO:10)的氨基酸编号。

[0208] 在更优选的实施方案中,所述突变的高活性PiggyBac包含以下的氨基酸取代组合:

[0209] -N347A/D450N,

[0210] -N347S/D450N/T560A/S573A/F594L,

[0211] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,

[0212] -R275A/N347S/K375A/D450N/S592G,

[0213] -R275A/N347S/R372A/D450N/T560A/F594L,

[0214] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,

[0215] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,

[0216] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,

[0217] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,

[0218] -S230N/R277A/N347S/K375A/D450N,

[0219] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,

[0220] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,

[0221] 位置编号对应于高活性PiggyBac序列(SEQ ID NO:10)的氨基酸编号。

[0222] 在一些实施方案中,所述突变的转座酶包含以下的氨基酸取代组合:R372A/K375A/D450N,所述位置编号对应于SEQ ID NO:10的未突变的高活性PiggyBac的氨基酸编号。在一些实施方案中,所述突变的转座酶具有SEQ ID NO:11的氨基酸序列。

[0223] 在一些实施方案中,所述突变的转座酶具有选自SEQ ID NO:12至33中任一个的氨基酸序列。

[0224] 在一些实施方案中,所述突变的转座酶具有选自SEQ ID NO:12至SEQ ID NO:20中任一个的氨基酸序列。

[0225] 在一些实施方案中,所述突变的转座酶具有选自SEQ ID NO:21至SEQ ID NO:33中任一个的氨基酸序列。

[0226] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,所述突变涉及保守的催化三联体,例如在对应于SEQ ID NO:10的氨基酸编号的氨基酸268和/或346处(例如,D268N和/或D346N)。

[0227] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,这些突变对于切割来说至关重要,例如在对应于SEQ ID NO:10的氨基酸编号的氨基酸287、287/290和/或460/461处(例如K287A、K287A/K290A和/或R460A/K461A)。

[0228] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,这些突变涉及靶连接(target joining),例如在对应于SEQ ID NO:10的氨基酸编号的氨基酸351、356和/或379处(例如S351E、S351P、S351A和/或K356E)。

[0229] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,这些突变对于整合来说至关重要,例如在对应于SEQ ID NO:10的氨基酸编号的氨基酸560、564、571、573、589、592和/或594处(例如T560A、S564P、S571N、S573A、M589V、S592G和/或F594L)。

[0230] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,这些突变在比对(alignment)中涉及,例如在对应于SEQ ID NO:10的氨基酸编号的氨基酸325、347、350、357和/或465处(例如G325A、N347A、N347S、T350A和/或W465A)。

[0231] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,这些突变是非常保守的,例如在对应于SEQ ID NO:10的氨基酸编号的氨基酸576和/或587处(例如K576A和/或I587A)。

[0232] 在一些实施方案中,突变的转座酶可以相对于hyPB包含一个或多个突变,这些突变涉及$Zn2^{+}$结合,例如对应于SEQ ID NO:10的氨基酸编号的586(例如H586A)。

[0233] 在一些实施方案中,可编程转座酶可以相对于hyPB包含一个或多个突变,这些突变涉及整合,例如对应于SEQ ID NO:10的氨基酸编号的315、341、372和/或375(例如,R315A、R341A、R372A和/或K375A)。

[0234] 在一些实施方案中,突变的高活性PiggyBac包含与SEQ ID NO:10中所示的序列至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一的氨基酸序列。在一些实施方案中,突变的高活性PiggyBac因其与高活性PiggyBac相比具有将DNA整合

到基因组中的高特异性而被选择。在一些实施方案中,突变的高活性PiggyBac包含相对于SEQ ID NO:10具有本文公开的一个或多个修饰的氨基酸序列,并且保留与SEQ ID NO:12至SEQ ID NO:33中所示的序列至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一的序列。

[0235] 在一些实施方案中,突变的高活性PiggyBac转座酶可以包含选自以下氨基酸的一个或多个氨基酸的突变:245、275、277、325、347、351、372、375、388、450、465、560、564、573、589、592、594,对应于SEQ ID NO:10的氨基酸编号。

[0236] 在一些实施方案中,突变的高活性PiggyBac转座酶突变可以包含选自以下的一个或多个氨基酸修饰:R245A、R275A、R277A、R275A/R277A、G325A、N347A、N347S、S351E、S351P、S351A、R372A、K375A、R388A、D450N、W465A、T560A、S564P、S573A、M589V、S592G或F594L,对应于SEQ ID NO:10的氨基酸编号。

[0237] 在一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰D450N,对应于SEQ ID NO:10的氨基酸编号。

[0238] 在一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰R245A和D450,对应于SEQ ID NO:10的氨基酸编号。

[0239] 在一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰R245A、G325A和S573P,对应于SEQ ID NO:10的氨基酸编号。

[0240] 在一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰R245A、G325A、D450和S573P,对应于SEQ ID NO:10的氨基酸编号。

[0241] 在一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰N347S或N347A,对应于SEQ ID NO:10的氨基酸编号。

[0242] 在一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰N347S和D450N,对应于SEQ ID NO:10的氨基酸编号。

[0243] 在另一个实施方案中,突变的高活性PiggyBac转座酶包含氨基酸修饰N347A和D450N,对应于SEQ ID NO:10的氨基酸编号。在一些实施方案中,所述突变的高活性PiggyBac转座酶包含SEQ ID NO:21的氨基酸序列。

[0244] 在一些实施方案中,突变的高活性PiggyBac转座酶包含氨基酸序列SEQ ID NO:10,其中:

[0245] -位置34处的氨基酸残基为V或M,

[0246] -位置43处的氨基酸残基为T或I,

[0247] -位置177处的氨基酸残基为Y或H,

[0248] -位置202处的氨基酸残基为R或K,

[0249] -位置230处的氨基酸残基为S或N,

[0250] -位置245处的氨基酸残基为A,

[0251] -位置268处的氨基酸残基为D或N,

[0252] -位置277处的氨基酸残基为R或A,

[0253] -位置275处的氨基酸残基为R或A,

[0254] -位置277处的氨基酸残基为R或A,

[0255] -位置325处的氨基酸残基为A或G,

[0256] -位置347处的氨基酸残基为S或A,

[0257] -位置351处的氨基酸残基为E、P或A,

[0258] -位置372处的氨基酸残基为R或A,

[0259] -位置375处的氨基酸残基为K或A,

[0260] -位置388处的氨基酸残基为R或A,

[0261] -位置409处的氨基酸残基为K或A,

[0262] -位置411处的氨基酸残基为A或T,

[0263] -位置412处的氨基酸残基为K或A,

[0264] -位置450处的氨基酸残基为D或N,

[0265] -位置460处的氨基酸残基为R或A,

[0266] -位置465处的氨基酸残基为W或A,

[0267] -位置517处的氨基酸残基为S或A,

[0268] -位置560处的氨基酸残基为T或A,

[0269] -位置564处的氨基酸残基为P或S,

[0270] -位置571处的氨基酸残基为S或N,

[0271] -位置573处的氨基酸残基为S或A,

[0272] -位置576处的氨基酸残基为K或A,

[0273] -位置586处的氨基酸残基为H或A,

[0274] -位置587处的氨基酸残基为I或A,

[0275] -位置589处的氨基酸残基为M或V,

[0276] -位置592处的氨基酸残基为G或S,和/或,-位置594处的氨基酸残基为L或F。

[0277] 在一些实施方案中,突变的高活性PiggyBac转座酶包含与选自包含以下或由以下组成的组的氨基酸序列具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、99.9%或100%序列同一性的氨基酸序列或由其组成:SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:20、SEQ ID NO:21、SEQ ID NO:22、SEQ ID NO:23、SEQ ID NO:24、SEQ ID NO:25、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32和SEQ ID NO:33。

[0278] 在一些实施方案中,突变的转座酶不是Himar1C9突变体。

[0279] 在一些实施方案中,RNA引导的核酸酶或切口酶和hyPB以融合蛋白的形式结合在一起。在一些实施方案中,RNA引导的核酸酶或切口酶和hyPB包含在融合蛋白中。在一些实施方案中,融合蛋白还包含接头。

[0280] 在一些实施方案中,RNA引导的核酸酶/切口酶-hyPB融合蛋白包含以下或由以下组成:

[0281] (i)第一蛋白,其包含RNA引导的核酸酶或切口酶或由RNA引导的核酸酶或切口酶组成,如上文所述,

[0282] (ii)可选地,接头,以及

[0283] (iii)第二蛋白,其包含转座酶或由转座酶组成,所述转座酶是与SEQ ID NO:10的hyPB相比包含一个或多个氨基酸突变的突变hyPB,如上文所述。

[0284] 在一些实施方案中,RNA引导的核酸酶/切口酶-hyPB融合蛋白中包含的RNA引导的核酸酶或切口酶是如上文所述的Cas9蛋白。

[0285] 在一些实施方案中,RNA引导的核酸酶/切口酶-hyPB融合蛋白包含以下或由以下组成:

[0286] (i)第一蛋白,其包含Cas9蛋白或其变体或由Cas9蛋白或其变体组成,如上文所述,

[0287] (ii)可选地,接头,以及

[0288] (iii)第二蛋白,其包含转座酶或由转座酶组成,所述转座酶是与SEQ ID NO:10的hyPB相比包含一个或多个氨基酸突变的突变hyPB,如上文所述。

[0289] 在一些实施方案中,第一蛋白和第二蛋白在融合蛋白中可以以任意顺序排列。

[0290] 在一些实施方案中,RNA引导的核酸酶或切口酶的N末端与hyPB的C末端直接融合或通过接头间接融合。在另一个实施方案中,hyPB的N末端与RNA引导的核酸酶或切口酶的C末端直接融合或通过接头间接融合。

[0291] 在一些实施方案中,RNA引导的核酸酶或切口酶(例如,Cas9蛋白)与慢病毒蛋白融合,优选与结构慢病毒蛋白(例如,衣壳或基质蛋白)融合,以有利于其在慢病毒颗粒中的包装。

[0292] 在一些实施方案中,RNA引导的核酸酶或切口酶与病毒蛋白R(VPR)结合。在一些实施方案中,VPR具有与SEQ ID NO:67具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、99.9%或100%序列同一性的氨基酸序列。

[0293]

SEQ ID NO: 67

MEQAPEDQGPQREPYNEWTLELLEELKREAVRHFPRPWLHGLGQHIYET

YGDTWTGVEAIIRILQRLLFVHFRIGCQHSRIGILRQRRARNGASRS

[0294] 在一些实施方案中,VPR-RNA引导的核酸酶/切口酶融合蛋白包含以下或由以下组成:

[0295] (i)第一蛋白,其包含VPR或其变体或由VPR或其变体组成,

[0296] (ii)可选地,接头,以及

[0297] (iii)第二蛋白,其包含RNA引导的核酸酶或切口酶或由RNA引导的核酸酶或切口酶组成,如上文所述。

[0298] 在一些实施方案中,RNA引导的核酸酶或切口酶和VPR以融合蛋白的形式结合在一起。在一些实施方案中,融合蛋白中包含的RNA引导的核酸酶或切口酶是如上文所述的Cas9蛋白。

[0299] 在一些实施方案中,VPR-Cas9融合蛋白包含以下或由以下组成:

[0300] (i)第一蛋白,其包含VPR或其变体或由VPR或其变体组成,

[0301] (ii)可选地,接头,以及

[0302] (iii)第二蛋白,其包含Cas9蛋白或其变体或由Cas9蛋白或其变体组成,如上文所述。

[0303] 在一些实施方案中,第一蛋白和第二蛋白在融合蛋白中可以以任意顺序排列。

[0304] 在一些实施方案中,Cas9的N末端与VPR的C末端直接融合或通过接头间接融合。在

一些实施方案中,VPR-Cas9融合蛋白与SEQ ID NO:66具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、99.9%或100%的序列同一性。

[0305] 在一些实施方案中,Cas9的C末端与VPR的N末端直接融合或通过接头间接融合。

[0306] 在一些实施方案中,VPR-Cas9融合蛋白进一步与RNA引导的核酸酶或切口酶融合,如上文所述。

[0307] 在一些实施方案中,RNA引导的核酸酶或切口酶与晶状体上皮衍生生长因子(Lens Epithelium Derived Growth Factor,LEDGF,与P75可互换使用)结合。

[0308] 在一些实施方案中,P75-RNA引导的核酸酶/切口酶融合蛋白包含以下或由以下组成:

[0309] (i)第一蛋白,其包含P75或其变体或由P75或其变体组成,

[0310] (ii)可选地,接头,以及

[0311] (iii)第二蛋白,其包含RNA引导的核酸酶或切口酶或由RNA引导的核酸酶或切口酶组成,如上文所述。

[0312] 在一些实施方案中,RNA引导的核酸酶或切口酶和P75以融合蛋白的形式结合在一起。在一些实施方案中,融合蛋白中包含的RNA引导的核酸酶或切口酶是如上文所述的Cas9蛋白。

[0313] 在一些实施方案中,P75-Cas9融合蛋白包含以下或由以下组成:

[0314] (i)第一蛋白,其包含P75或其变体或由P75或其变体组成,

[0315] (ii)可选地,接头,以及

[0316] (iii)第二蛋白,其包含Cas9蛋白或其变体或由Cas9蛋白或其变体组成,如上文所述。

[0317] 在一些实施方案中,第一蛋白和第二蛋白在融合蛋白中可以以任意顺序排列。

[0318] 在一些实施方案中,Cas9的N末端与P75的C末端直接融合或通过接头间接融合。在另一个实施方案中,Cas9的C末端与P75的N末端直接融合或通过接头间接融合。

[0319] 在一些实施方案中,P75-Cas9融合蛋白具有与SEQ ID NO:69至73中任一个至少75%、80%、85%、90%、95%、99%或100%同一的氨基酸序列。在一些实施方案中,P75-Cas9融合蛋白具有选自包含SEQ ID NO:69至73或由SEQ ID NO:69至73组成的组的氨基酸序列。在一些实施方案中,P75-Cas9融合蛋白具有SEQ ID NO:69的氨基酸序列。在一些实施方案中,P75-Cas9融合蛋白具有SEQ ID NO:70的氨基酸序列。在一些实施方案中,P75-Cas9融合蛋白具有SEQ ID NO:71的氨基酸序列。在一些实施方案中,P75-Cas9融合蛋白具有SEQ ID NO:72的氨基酸序列。在一些实施方案中,P75-Cas9融合蛋白具有SEQ ID NO:73的氨基酸序列。

[0320] 在一些实施方案中,P75-Cas9融合蛋白进一步与如上文所述的RNA引导的核酸酶或切口酶融合。

[0321] 在一些实施方案中,RNA引导的核酸酶或切口酶不与整合酶融合。在一些实施方案中,RNA引导的核酸酶或切口酶不与(i)的慢病毒载体的整合酶融合。在一些实施方案中,RNA引导的核酸酶或切口酶不与由(v)的核酸编码的融合蛋白的整合酶(如果存在的话)融合。

[0322] 在一些实施方案中,RNA引导的核酸酶或切口酶与如本文定义的整合酶或其突变

体结合。

[0323] 在一些实施方案中,整合酶-RNA引导的核酸酶/切口酶融合蛋白包含以下或由以下组成:

[0324] (i)第一蛋白,其包含本文定义的整合酶或其突变体或由本文定义的整合酶或其突变体组成,

[0325] (ii)可选地,接头,以及

[0326] (iii)第二蛋白,其包含RNA引导的核酸酶或切口酶或由RNA引导的核酸酶或切口酶组成,如上文所述。

[0327] 在一些实施方案中,本文定义的RNA引导的核酸酶或切口酶和整合酶或其突变体以融合蛋白的形式结合在一起。在一些实施方案中,融合蛋白中包含的RNA引导的核酸酶或切口酶是如上文所述的Cas9蛋白。

[0328] 在一些实施方案中,整合酶-Cas9融合蛋白包含以下或由以下组成:

[0329] (i)第一蛋白,其包含本文定义的整合酶或其突变体或由本文定义的整合酶或其突变体组成,或其变体,

[0330] (ii)可选地,接头,以及

[0331] (iii)第二蛋白,其包含Cas9蛋白或其变体或由Cas9蛋白或其变体组成,如上文所述。

[0332] 在一些实施方案中,第一蛋白和第二蛋白在融合蛋白中可以以任意顺序排列。

[0333] 在一些实施方案中,Cas9的N末端与本文定义的整合酶或其突变体的C末端直接融合或通过接头间接融合。在另一个实施方案中,Cas9的C末端与本文定义的整合酶或其突变体的N末端直接融合或通过接头间接融合。

[0334] 在一些实施方案中,整合酶-Cas9融合蛋白进一步与如上文所述的RNA引导的核酸酶或切口酶融合。

[0335] 在一些实施方案中,RNA引导的核酸酶或切口酶与GAG和/或POL结合。

[0336] 在一些实施方案中,GAG和/或POL-RNA引导的核酸酶/切口酶融合蛋白包含以下或由以下组成:

[0337] (i)第一蛋白,其包含GAG和/或POL或其变体或由GAG和/或POL或其变体组成,

[0338] (ii)可选地,接头,以及

[0339] (iii)第二蛋白,其包含RNA引导的核酸酶或切口酶或由RNA引导的核酸酶或切口酶组成,如上文所述。

[0340] 在一些实施方案中,RNA引导的核酸酶或切口酶与GAG和/或POL以融合蛋白的形式结合在一起。在一些实施方案中,融合蛋白中包含的RNA引导的核酸酶或切口酶是如上文所述的Cas9蛋白。

[0341] 在一些实施方案中,GAG和/或POL-Cas9融合蛋白包含以下或由以下组成:

[0342] (i)第一蛋白,其包含GAG和/或POL或其变体或由GAG和/或POL或其变体组成,

[0343] (ii)可选地,接头,以及

[0344] (iii)第二蛋白,其包含Cas9蛋白或其变体或由Cas9蛋白或其变体组成,如上文所述。

[0345] 在一些实施方案中,第一蛋白和第二蛋白在融合蛋白中可以以任意顺序排列。

[0346] Cas9的N末端与GAG和/或POL的C末端直接融合或通过接头间接融合。在另一个实施方案中,Cas9的C末端与GAG和/或POL的N末端直接融合或通过接头间接融合。

[0347] 在一些实施方案中,GAG和/或POL-Cas9融合蛋白进一步与如上文所述的RNA引导的核酸酶或切口酶融合。

[0348] 根据本发明,组合物包含编码与至少一个适体融合的引导RNA(gRNA)的核酸。

[0349] 本文中使用的术语"至少一个"表示1、2、3、4、5、6、7、8、9、10或更多。

[0350] 在一些实施方案中,适体是包含四环的RNA序列。术语"四环"可与术语"茎环"和"发夹环"互换使用。

[0351] 在一些实施方案中,本文所述的适体结合蛋白能够结合至少一个适体。

[0352] 在一些实施方案中,所述至少一个四环是MS2 RNA四环结合序列。在一些实施方案中,所述至少一个MS2 RNA四环结合序列与SEQ ID NO:63(例如,由SEQ ID NO:62的DNA序列编码)具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%或100%的同一性。

[0353]

SEQ ID NO: 62 –MS2 RNA 四环结合序列 – DNA 编码序列

ggggccactagggacaggatgttttagagctaggccaacatgaggatcacccatgtctgcagggcctagcaagtta

aaataaggctagtccgttatcaacttggccaacatgaggatcacccatgtctgcagggccaagtggcaccgagtcgg

tgcttttt

SEQ ID NO: 63 –MS2 RNA 四环结合序列

ggggccacuagggacaggauguuuuagagcuaggccaacaugaggaucacccaugucugcagggccuag

caaguuaaaauaaggcuaguccguuaucaacuuggccaacaugaggaucacccaugucugcagggccaag

uggcaccgagucggugcuuuuuu

[0354] 在一些实施方案中,gRNA能够与上文所述的RNA引导的核酸酶或切口酶例如Cas9蛋白或包含Cas9的融合蛋白形成复合物。

[0355] 在一些实施方案中,gRNA能够将上文所述的RNA引导的核酸酶或切口酶靶向细胞(优选哺乳动物细胞,更优选人类细胞)基因组的特定序列或区域。在一些实施方案中,gRNA所靶向的特定序列与对Cas9蛋白特异的原型间隔区相邻基序(PAM)相邻。

[0356] 在一些实施方案中,gRNA所靶向的特定序列可位于细胞基因组中的安全港基因座(safe harbor locus)内。"安全港基因座"是指细胞基因组的区域,在此整合的物质可以充分表达而不会扰乱内源基因结构或功能。安全港基因座包括但不限于AAVS1(PPP1R12C的内含子1)、HPRT、HI 1、hRosa26、白蛋白和F-A区域。安全港基因座可以是普遍表达的基因和/或具有组织特异性表达(例如肌肉)的基因的外显子或内含子。安全港基因座可以选自:PPP1R12C的外显子1、内含子1或外显子2,HPRT的外显子1、内含子1或外显子2,hRosa26的外显子1、内含子1或外显子2,白蛋白基因的内含子1。安全港基因座还可以包括缺乏内源基因且具有开放染色质的基因组区域,所述开放染色质允许表达插入的转基因而不会扰乱基因组结构或功能。

[0357] 在一些实施方案中,gRNA所靶向的特定序列可位于细胞基因组中的基因内或基因附近。例如,可以涵盖的是,gRNA所靶向的特定序列位于需要基因修复或基因修饰的基因内

或基因附近(通过整合所述基因的外源拷贝,其由(iv)中编码目标转基因的核酸带来)。

[0358] 根据本发明,组合物包含编码目标转基因的核酸。

[0359] 本文所用的术语“目标转基因”是指要插入细胞(优选真核细胞、更优选哺乳动物细胞)的基因组的外源核酸序列,其包含至少一个开放阅读框。通常,目标转基因是编码序列,即,其编码至少一个蛋白或多肽。然而,目标转基因也可以是非编码序列。

[0360] 在一些实施方案中,目标转基因包含至少一个基因和至少一个非编码序列,例如调控元件或启动子。

[0361] 在一些实施方案中,目标转基因包含至少一个基因。在一些实施方案中,所述至少一个基因不存在于细胞的基因组中。在一些实施方案中,所述至少一个基因是细胞的基因组的基因的等位基因,其中该等位基因对于细胞的基因组不是内源的。

[0362] 在一些实施方案中,目标转基因包含至少一个与细胞来自同一物种的基因。在另一个实施方案中,目标转基因包含至少一个与细胞来自不同物种的基因。

[0363] 在一些实施方案中,目标转基因的长度为至少500个碱基对(bp)。如本文所用,至少500意指500、600、700、800、900、1,000、10,000、20,000bp或更多。

[0364] 在一些实施方案中,目标转基因是DNA分子、RNA分子或其组合。在一些实施方案中,目标转基因是DNA分子。在一些实施方案中,目标转基因是RNA分子。

[0365] 在一些实施方案中,目标转基因的侧翼可以是反向末端重复(ITR)序列。这在RNA引导的核酸酶或切口酶与如上文所述的转座酶结合时尤其期望如此。

[0366] 根据本发明,组合物可选地包含编码融合蛋白的核酸,所述融合蛋白包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及可选的(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成。

[0367] 本文所用的术语“GAG”是指逆转录病毒多聚蛋白。GAG多聚蛋白通常在翻译后由病毒蛋白酶(PR)加工成MA(基质)、CA(衣壳)和NC(核衣壳)部分,有时还会加工成更多部分。例如,HIV-1病毒的GAG多聚蛋白在翻译后被加工成基质蛋白p17、衣壳蛋白p24、核衣壳蛋白p7和p6-gag。在一些实施方案中,融合蛋白含有包含HIV-1GAG多聚蛋白的蛋白或含有由HIV-1GAG多聚蛋白组成的蛋白。

[0368] 本文使用的术语“POL”是指逆转录病毒多聚蛋白。POL多聚蛋白通常在翻译后由病毒蛋白酶(PR)加工成PR(蛋白酶)、RT(逆转录酶)、p15(核糖核酸酶H)和INT(整合酶)部分。蛋白酶本身通过自催化裂解释放。在一些实施方案中,融合蛋白含有包含HIV-1POL多聚蛋白的蛋白或含有由HIV-1POL多聚蛋白组成的蛋白。

[0369] 在一些实施方案中,融合蛋白包含POL多聚蛋白,因此组合物可包含编码两个不同整合酶的核酸分子:一个具有整合活性受损的整合酶在(i)的慢病毒载体中;另一种整合酶在(v)的融合蛋白中。

[0370] 在一些实施方案中,适体结合蛋白与GAG多聚蛋白融合。在一些实施方案中,适体结合蛋白的C末端与GAG多聚蛋白的N末端融合,可选地通过接头融合。在一些实施方案中,适体结合蛋白的N末端与GAG多聚蛋白的C末端融合,可选地通过接头融合。

[0371] 在一些实施方案中,适体结合蛋白的C末端与多聚蛋白GAG的N末端融合,可选地通过接头融合,通常地,融合蛋白具有SEQ ID NO:64的氨基酸序列。

[0372]

SEQ ID NO: 64

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSV
RQSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATN
SDCELIVKAMQGLLKDGNPIPSAIAANSGIYSQNYPIVQSGLRSRAQASNS
AVDGTAGPGSMDSGRDFLTLHGLQDDEDLQALLKGSQLLKVKSSSWRRE
RFYKLQEDCKTIWQESRKVMRTPESQLFSIEDIQEVRMGHRTEGLEKFAR
DVPEDRCFSIVFKDQRNTLDLIAPSPADAQHWVLGLHKIIHHSGSMDQRQ
KLQHWIHSCLRKADKNKDNKMSFKELQNFLKELNIQPRAEFTLAARASV
LSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQ
NKSKKKAQQAAADTGHSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWV
KVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETI
NEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNP
PIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRA
EQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVG
GPGHKARVLAEAMSQVTNTATIMMQRGNFRNQRKMVKCFNCGKEGHT
ARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFL
QSRPEPTAPPFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLR
SLFGNDPSSQ

[0373] 在一些实施方案中,适体结合蛋白与POL多聚蛋白(如果存在的话)融合。在一些实施方案中,适体结合蛋白的C末端与POL多聚蛋白(优选GAG-POL多聚蛋白)的N末端融合,可选地通过接头融合。在一些实施方案中,适体结合蛋白的N末端与POL多聚蛋白(优选GAG-POL多聚蛋白)的C末端融合,可选地通过接头融合。

[0374] 在一些实施方案中,适体结合蛋白的C末端与多聚蛋白GAG-POL的N末端融合,可选地通过接头融合,通常地,融合蛋白具有SEQ ID NO:65的氨基酸序列。

[0375]

SEQ ID NO: 65

MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSV
RQSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATN
SDCELIVKAMQGLLKDGNPIPSAIAANSGIYSQNYPIVQSGLRSRAQASNS
AVDGTAGPGSMDSGRDFLTLHGLQDDEDLQALLKGSQLLKVKSSSWRRE
RFYKLQEDCKTIWQESRKVMRTPESQLFSIEDIQEVRMGHRTEGLEKFAR
DVPEDRCFSIVFKDQRNTLDLIAPSPADAQHWVLGLHKIIHHSGSMDQRQ
KLQHWIHSCLRKADKNKDNKMSFKELQNFLKELNIQPRAEFTLAARASV
LSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQ
NKSKKKAQQAAADTGHSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWV
KVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETI
NEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNP
PIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRA
EQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVG
GPGHKARVLAEAMSQVTNTATIMMQRGNFRNQRKMVKCFNCGKEGHT
ARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFL
QSRPEPTAPPFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLR
SLFGNDPSSQRSVASRRPCWTPAPTTPCWRRACPAGGSPRSAASAASSRG
STTRSSRSAATRPSAPCWWDLHLTSSAGTCPRSAAPTSPSAPSRPCPSSPA
WTALRSSGPPRRRSRPWWRSAPRWRRRARSARSAPRTPTTPPCSPSRRRT

[0376]

APSGGSWWTSGSTRGPRTSGRCSWASPTPPARRRRAPCWTWATPTSACP
WTRTSGSIPPSPSPASTTRPPASATSTTCCPRAGRAAPPSSRAAQRSWSPSR
SRTPTSSISTWTTCTWAATWRSASTGPRSRSGSTCGGAPPPTRSTRRSPHS
CGWATSCTPTSGPCSPSCCPRRTAGPTTFRSWWASTGPARSTLASRGSCA
SCGAQRLPRSPPRRPSWSWPRTGRSRSPCTACTTTPARTSPRSRSRARASG
PTRSTRSPSRTRPASTPACAAPTPTTSSPRPCRRSPPRASSGARLLSSSCPSR
RRPGRPGGPSTGRPPGFPSGSSTPLPWSCGISWRRSPSWAPRPSTWTAPPT
GRPSWARPATPTRAARRWCPPTPPTRRPSCRLSTWPCRTQAWRTSPTAST
PWASSRPSPTRARASWTRSSSSSRRRRCTWPGCPPTRASAATSRWTSWAP
ASGRSCSWTASTRPRTSTRSTTATGGLWLATSTCLPWWLRRSWPAATSA
SRARPCTARWTAAPASGSWIAPTWRARSWWPCTWPPATSRPRSPPRPAR
RPPTSCSWPAAGPRPSTPTTAATSPAPPRLPAGGPASSRSSASPTTPSLRAW
WRATRSRRSSARGTRPSTRPPCRWPCSSTTSRGRAASAATAPARGSWTSS
PPTSRPRSCRSRSPRSRTSGCTTGTAGTLCGRAPPSCCGRARAPWSRTTAT
SRWCPGGRPRSSGTTASRWPATTAWPPGRTRT

[0377] 上文已描述了适体结合蛋白,其描述经适当修改后亦适用于此。

[0378] 在一些实施方案中,当融合蛋白含有包含POL多聚蛋白的第三蛋白或含有由POL多聚蛋白组成的第三蛋白时,POL多聚蛋白的整合酶可具有受损的整合活性。

[0379] 上文已描述了损害整合酶整合活性的突变(包括取代和缺失),其描述经适当修改后亦适用于此。

[0380] 在一些实施方案中,POL多聚蛋白的整合酶与(i)的慢病毒载体的整合酶相比可包含相同或不同的突变。

[0381] 在一些实施方案中,融合蛋白包含一个或几个接头。

[0382] 在一些实施方案中,接头是肽接头。在一个实施方案中,肽接头选自包含以下或由以下组成的组:$(GGS)_n$、$(GGGGS)_n$、$(G)_n$、$(EAAAK)_n$、XTEN接头和$(XP)_n$基序,以及这些中的任意组合,其中n独立地为1至50之间的整数。

[0383] 在一些实施方案中,接头的长度为1至24个氨基酸,或由长度为3至72个核苷酸的核酸序列编码。在一个实施方案中,接头的长度为1至12个氨基酸,或由长度为3至36个核苷酸的核酸序列编码。在一个实施方案中,接头的长度为1至6个氨基酸,或由长度为3至18个核苷酸的核酸序列编码。

[0384] 一些实施方案中,接头是XTEN接头或$(GGS)_n$接头。

[0385] 在一些实施方案中,接头选自表1所示的接头。

[0386] 表1:接头

[0387]

接头	核酸序列 (SEQ ID NO)	氨基酸序列 (SEQ ID NO)
GGSx3	ggtggatctggcggtggatctggtggcggt (SEQ ID NO: 44)	GGSGGGSGGG (SEQ ID NO: 45)
GGS4x	ggagggagtggtgggtccggtggtagtggcgg atcc (SEQ ID NO: 46)	GGSGGSGGSGGS (SEQ ID NO: 47)
GGS5x	ggaggctccggtgggtctggtgggagcggtgg tagtggcggatcc (SEQ ID NO: 48)	GGSGGSGGSGGSGGS (SEQ ID NO: 49)
GGS6x	ggaggcagtggtgggagcggtggttccgggg gtagtggtggttccgggggatcc (SEQ ID NO: 50)	GGSGGSGGSGGSGGS GGS (SEQ ID NO: 51)
GGS7x	ggaggttctggaggctccggtgggtccggggg aagtgggggtcaggcggatcaggaggatcc (SEQ ID NO: 52)	GGSGGSGGSGGSGGS GGSGGS (SEQ ID NO: 53)
GGS8x	ggaggtagcggaggttccggagggagcggcg ggagtgggggaagcgggggaagtggaggatc cgggggaggatcc (SEQ ID NO: 54)	GGSGGSGGSGGSGGS GGSGGS (SEQ ID NO: 55)

[0388]

接头 XTEN	tccggtagcgaaacaccggggacttcagaatcg gccaccccggagtct (SEQ ID NO: 56)	SGSETPGTSESATPES (SEQ ID NO: 57)
接头 B	ggaagcgccggtagtgcggctgggtctggcga gttc (SEQ ID NO: 58)	GSAGSAAGSGEF (SEQ ID NO: 59)

[0389] 在一些实施方案中,所述接头包含选自包含以下或由以下组成的组的氨基酸序列:SEQ ID NO:45、SEQ ID NO:47、SEQ ID NO:49、SEQ ID NO:51、SEQ ID NO:53、SEQ ID NO:55、SEQ ID NO:57、SEQ ID NO:59或其任意组合;分别由SEQ ID NO:44、SEQ ID NO:46、SEQ ID NO:48、SEQ ID NO:50、SEQ ID NO:52、SEQ ID NO:54、SEQ ID NO:56、SEQ ID NO:58的示例性核酸序列编码。

[0390] 在一些实施方案中,接头包含SEQ ID NO:45的氨基酸序列或由SEQ ID NO:45的氨基酸序列组成;由SEQ ID NO:44的示例性核酸序列编码。

[0391] 在一些实施方案中,适体结合蛋白以任意顺序与突变整合酶融合。在一些实施方案中,适体结合蛋白通过N末端与突变整合酶融合。在一些实施方案中,适体结合蛋白通过C末端与整合酶融合。在一些实施方案中,适体结合蛋白的N末端与突变整合酶的C末端融合,可选地通过接头融合。在一些实施方案中,适体结合蛋白的C末端与突变整合酶的N末端融合,可选地通过接头融合。

[0392] 在一些实施方案中,融合蛋白还包含病毒蛋白R(VPR)。在一些实施方案中,适体结合蛋白与POL多聚蛋白的整合酶融合,并且POL多聚蛋白的整合酶进一步与VPR融合。在一些实施方案中,VPR具有与SEQ ID NO:67具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、99.9%或100%序列同一性的氨基酸序列。在一些实施方案中,适体结合蛋白与GAG多聚蛋白融合,并且GAG多聚蛋白进一步与VPR融合。在一些实施方案中,适体结合蛋白与GAG-POL多聚蛋白融合,并且GAG-POL多聚蛋白进一步与VPR融合。

[0393] 在一些实施方案中,POL多聚蛋白的整合酶的N末端与VPR的C末端融合,可选地通过接头融合,并且适体结合蛋白的N末端与POL多聚蛋白的整合酶的C末端融合,可选地通过接头融合。在一些实施方案中,POL多聚蛋白的整合酶的N末端与适体结合蛋白的C末端融合,可选地通过接头融合,并且VPR的N末端与-POL多聚蛋白的整合酶的C末端融合,可选地通过接头融合。

[0394] 在一些实施方案中,GAG多聚蛋白的N末端与VPR的C末端融合,可选地通过接头融合,并且适体结合蛋白的N末端与GAG多聚蛋白的C末端融合,可选地通过接头融合。在一些实施方案中,GAG多聚蛋白的N末端与适体结合蛋白的C末端融合,可选地通过接头融合,并且VPR的N末端与GAG多聚蛋白的C末端融合,可选地通过接头融合。

[0395] 在一些实施方案中,GAG-POL多聚蛋白的N末端与VPR的C末端融合,可选地通过接头融合,并且适体结合蛋白的N末端与GAG-POL多聚蛋白的C末端融合,可选地通过接头融合。在一些实施方案中,GAG-POL多聚蛋白的N末端与适体结合蛋白的C末端融合,可选地通过接头融合,并且VPR的N末端与GAG-POL多聚蛋白的C末端融合,可选地通过接头融合。

[0396] 本发明还涉及通过上述生产慢病毒颗粒的方法获得的慢病毒颗粒群。

[0397] 在一些实施方案中,慢病毒颗粒群包含慢病毒颗粒,所述慢病毒颗粒包含以下或由以下组成:

[0398] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,

[0399] (ii)RNA引导的核酸酶或切口酶,优选Cas9蛋白,可选地融合至突变的hyPB和/或VPR和/或P75和/或GAG和/或POL和/或整合酶或其突变体,如上文所述,或编码其的核酸序列,

[0400] (iii)引导RNA(gRNA),其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,如上文所述,以及

[0401] (iv)可选地,融合蛋白,所述融合蛋白包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,和可选地(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,如上文所述,或编码所述融合蛋

白的核酸序列。

[0402] 在一些实施方案中,慢病毒颗粒群中的慢病毒颗粒还包含目标转基因,如上文所述。

[0403] 在一些实施方案中,慢病毒颗粒悬浮并储存在合适的溶剂中,优选生物相容性溶剂(例如生物缓冲液、生理血清等)。

[0404] 在一些实施方案中,慢病毒颗粒储存在适合防止核酸降解的温度下。在一些实施方案中,慢病毒颗粒储存在包含4℃至-200℃的温度下。如本文所用,表述"4℃至-200℃"涵盖4、3、2、1、0、-1、-2、-3、-4、-5、-6、-7、-8、-9、-10、-11、-12、-13、-14、-15、-16、-17、-18、-19、-20、-30、-40、-50、-60、-70、-80、-90、-100、-120、-140、-160、-180、-200°C。在一些实施方案中,慢病毒颗粒储存在包含4℃至-80℃的温度下。在一些实施方案中,慢病毒颗粒储存在包含4℃至-20℃的温度下。在一些实施方案中,慢病毒颗粒储存在包含4℃至0℃的温度下。在一些实施方案中,慢病毒颗粒储存在包含0℃至-200℃的温度下。在一些实施方案中,慢病毒颗粒储存在包含-20℃至-200℃的温度下。在一些实施方案中,慢病毒颗粒储存在包含-80℃至-200℃的温度下。在一些实施方案中,慢病毒颗粒储存在-196℃的温度下。

[0405] 在一些实施方案中,慢病毒颗粒是基本上纯的。

[0406] 在一些实施方案中,慢病毒颗粒还包含至少一种治疗剂。

[0407] 在一些实施方案中,当慢病毒颗粒包含含有GAG多聚蛋白、适体结合蛋白和可选的POL多聚蛋白的融合蛋白时,该融合蛋白(或编码其的核酸)和RNA引导的核酸酶或切口酶(或编码其的核酸)可以包含在不同的慢病毒颗粒中。

[0408] 在一些实施方案中,当慢病毒颗粒包含含有GAG多聚蛋白、适体结合蛋白和可选的POL多聚蛋白的融合蛋白时,该融合蛋白(或编码其的核酸)和RNA引导的核酸酶或切口酶(或编码其的核酸)包含在2个不同的慢病毒颗粒中。

[0409] 在一些实施方案中,慢病毒颗粒群包含:

[0410] 第一慢病毒颗粒,其包含以下或由以下组成:

[0411] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,

[0412] (ii)RNA引导的核酸酶或切口酶,优选Cas9蛋白,可选地融合至突变的hyPB和/或VPR和/或P75和/或GAG和/或POL和/或整合酶或其突变体,如上文所述,或编码其的核酸序列,和

[0413] (iii)引导RNA(gRNA),其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,如上文所述;以及

[0414] 第二慢病毒颗粒,其包含以下或由以下组成:

[0415] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,以及

[0416] (ii)融合蛋白,包含:(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成;(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成;以及可选地,(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,如上文所述,或编码所述融合蛋白的核酸序列。

[0417] 在一些实施方案中,第一慢病毒颗粒、第二慢病毒颗粒或两者包含目标转基因,如上文所述。

[0418] 在一些实施方案中,慢病毒颗粒群包含:

[0419] 第一慢病毒颗粒,其包含以下或由以下组成:

[0420] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,以及

[0421] (ii)RNA引导的核酸酶或切口酶,优选Cas9蛋白,可选地融合至突变的hyPB和/或VPR和/或P75和/或GAG和/或POL和/或整合酶或其突变体,如上文所述,或编码其的核酸序列;和

[0422] 第二慢病毒颗粒,其包含以下或由以下组成:

[0423] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,

[0424] (ii)融合蛋白,包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及可选地(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,如上文所述,或编码所述融合蛋白的核酸序列,以及

[0425] (iii)引导RNA(gRNA),其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,如上文所述。

[0426] 在一些实施方案中,第一慢病毒颗粒、第二慢病毒颗粒或两者包含目标转基因,如上文所述。

[0427] 在一些实施方案中,当慢病毒颗粒包含含有GAG多聚蛋白、适体结合蛋白和可选的POL多聚蛋白的融合蛋白时,该融合蛋白(或编码其的核酸)和RNA引导的核酸酶或切口酶(或编码其的核酸)包含在3个不同的慢病毒颗粒中。

[0428] 在一些实施方案中,慢病毒颗粒群包含:

[0429] 第一慢病毒颗粒,其包含以下或由以下组成:

[0430] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,以及

[0431] (ii)RNA引导的核酸酶或切口酶,优选Cas9蛋白,可选地融合至突变的hyPB和/或VPR和/或P75和/或GAG和/或POL和/或整合酶或其突变体,如上文所述,或编码其的核酸序列;和

[0432] 第二慢病毒颗粒,其包含以下或由以下组成:

[0433] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,和

[0434] (ii)融合蛋白,包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及可选地(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,如上文所述,或编码所述融合蛋白的核酸序列;以及

[0435] 第三慢病毒颗粒,其包含以下或由以下组成:

[0436] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,以及

[0437] (ii)引导RNA(gRNA),其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,如上文所述。

[0438] 在一些实施方案中,第一慢病毒颗粒、第二慢病毒颗粒、第三慢病毒颗粒或它们中

的两个或所有三个包含如上文所述的目标转基因。

[0439] 在一些实施方案中,当慢病毒颗粒包含含有GAG多聚蛋白、适体结合蛋白和可选的POL多聚蛋白的融合蛋白时,该融合蛋白(或编码其的核酸)和RNA引导的核酸酶或切口酶(或编码其的核酸)包含在4个不同的慢病毒颗粒中。

[0440] 在一些实施方案中,慢病毒颗粒群包含:

[0441] 第一慢病毒颗粒,其包含以下或由以下组成:

[0442] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,以及

[0443] (ii)RNA引导的核酸酶或切口酶,优选Cas9蛋白,可选地融合至突变的hyPB和/或VPR和/或P75和/或GAG和/或POL和/或整合酶或其突变体,如上文所述,或编码其的核酸序列;和

[0444] 第二慢病毒颗粒,其包含以下或由以下组成:

[0445] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,和

[0446] (ii)融合蛋白,包含(1)第一蛋白,其包含GAG多聚蛋白或由GAG多聚蛋白组成,(2)第二蛋白,其包含适体结合蛋白或由适体结合蛋白组成,以及可选地(3)第三蛋白,其包含POL多聚蛋白或由POL多聚蛋白组成,如上文所述,或编码所述融合蛋白的核酸序列;以及

[0447] 第三慢病毒颗粒,其包含以下或由以下组成:

[0448] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,和

[0449] (ii)引导RNA(gRNA),其与适体融合,优选与至少一个MS2 RNA四环结合序列融合,如上文所述;以及

[0450] 第四慢病毒颗粒,其包含以下或由以下组成:

[0451] (i)慢病毒蛋白和/或基因或其部分,其中慢病毒整合酶具有受损的整合活性,如上文所述,和

[0452] (ii)如上文所述的目标转基因。

[0453] 本发明还涉及一种生产慢病毒颗粒的方法,包括以下步骤:

[0454] (i)用本发明的组合物转染在合适的培养基中培养的慢病毒生产细胞,以及

[0455] (ii)收集所述慢病毒生产细胞的培养基中的慢病毒颗粒。

[0456] 在一些实施方案中,将慢病毒生产细胞用上述组合物的1、2、3、4或5种核酸转染。例如,可以涵盖的是,将不同批次的慢病毒生产细胞用上述组合物的不同核酸转染,以产生不同的慢病毒颗粒。

[0457] 可以通过本领域中已知的方法转染细胞;非限制性例子包括脂质转染、电穿孔、声孔技术、基因枪、显微注射或基于化学的方法,例如磷酸钙。

[0458] 在一些实施方案中,细胞同时和/或分别用组合物的核酸分子转染。在一些实施方案中,细胞同时用组合物的核酸分子转染。在一些实施方案中,细胞分别用组合物的核酸分子转染。

[0459] 在一些实施方案中,慢病毒颗粒可通过分裂系统获得,例如,反式互补系统(载体/包装系统),通过用含有慢病毒载体基因组组分的质粒体外转染允许细胞(例如,HEK 293T

细胞),其中所述慢病毒载体基因组组分包括编码多聚蛋白GAG、POL和包膜多肽的gag、pol和env序列,或这些多肽中足以形成逆转录病毒颗粒的一部分。

[0460] 适合于生产慢病毒颗粒的细胞包括但不限于真核和原核的细胞和/或细胞系。此类细胞或由此类细胞产生的细胞系的非限制性实例包括例如COS、CHO(例如CHO-S、CHO-K1、CHO-DG44、CHO-DUXB11、CHO-DUKX、CHOK1SV)、VERO、MDCK、WI38、V79、B14AF28-G3、BHK、HaK、NS0、SP2/0-Ag14、HeLa、HEK293(例如HEK293-F、HEK293-H、HEK293-T)和perC6细胞以及昆虫细胞(例如草地贪夜蛾(Spodoptera fugiperda,Sf))或真菌细胞(例如酵母菌(Saccharomyces)、毕赤酵母(Pichia)和裂殖酵母(Schizosaccharomyces))。

[0461] 在一些实施方案中,慢病毒颗粒由慢病毒生产细胞分泌到其培养基中。在一些实施方案中,慢病毒颗粒在慢病毒生产细胞的培养基中收集。

[0462] 在一些实施方案中,慢病毒颗粒在转染后12小时至一周收集,优选在转染后24至72小时收集。

[0463] 在一些实施方案中,慢病毒颗粒从慢病毒生产细胞的培养基中纯化。可以通过任何合适的方法纯化慢病毒颗粒以提高浓度,例如通过密度梯度纯化(例如,氯化铯(CsCl))、通过色谱技术(例如,柱色谱或批处理色谱)或通过超速离心。例如,本发明的慢病毒颗粒可以进行两个或三个CsCl密度梯度纯化步骤。期望使用包括裂解细胞、将裂解物施加到色谱树脂、从色谱树脂洗脱慢病毒颗粒以及收集含有慢病毒颗粒的级分的方法,从感染的细胞中纯化慢病毒颗粒。

[0464] 本发明还涉及一种生产细胞,其包含(即,转染了)本发明的组合物。

[0465] 在一些实施方案中,包含本发明组合物的生产细胞生产慢病毒颗粒。通常,慢病毒颗粒从细胞的质膜出芽。

[0466] 在一些实施方案中,生产细胞来自永生化细胞系。在一些实施方案中,生产细胞来自原代细胞培养物。

[0467] 在一些实施方案中,生产细胞是真核细胞。在一些实施方案中,生产细胞是动物细胞或植物细胞。在一些实施方案中,生产细胞是动物细胞。在一些实施方案中,生产细胞是脊椎动物细胞或昆虫细胞。在一些实施方案中,生产细胞是脊椎动物细胞。在一些实施方案中,生产细胞是哺乳动物细胞。

[0468] 在一些实施方案中,生产细胞是非人哺乳动物细胞(例如,非人灵长类动物、小鼠、大鼠)。在一些实施方案中,生产细胞是人类细胞。本领域公知的培养细胞系(例如,HeLa细胞、HEK293细胞、HAP-1细胞等)。在一些实施方案中,生产细胞来自HEK293细胞系。在一些实施方案中,生产细胞来自HEK293T细胞系。

[0469] 在一些实施方案中,生产细胞与其来源的细胞系相比包含一个或多个突变。在一些实施方案中,生产细胞与其来源的细胞系相比未经过修饰。

[0470] 维持细胞培养的方法在本领域中是公知的。在一些实施方案中,生产细胞维持在合适的培养基中,例如杜氏改良伊格尔培养基(DMEM)。在一些实施方案中,生产细胞通常在37℃、95%空气和5% CO_2下培养。

[0471] 本发明还涉及一种将目标转基因位点特异性整合到靶细胞基因组中的方法,所述方法包括用上述慢病毒颗粒群感染所述细胞。

[0472] 在一些实施方案中,所述方法是体外方法。

[0473] 在一些实施方案中,慢病毒颗粒包含目标转基因。在一些实施方案中,慢病毒颗粒群中的一些慢病毒颗粒包含目标转基因。

[0474] 在一些实施方案中,目标转基因包含在慢病毒颗粒群的一个或几个或所有慢病毒颗粒中。在一些实施方案中,慢病毒颗粒不包含目标转基因,目标转基因通过合适的方法递送至靶细胞,例如通过转染(例如,脂质转染、电穿孔等)。在一些实施方案中,目标转基因在慢病毒颗粒之前递送至靶细胞。在一些实施方案中,目标转基因与慢病毒颗粒同时递送至靶细胞。在一些实施方案中,目标转基因在慢病毒颗粒之后递送至靶细胞。

[0475] 在一些实施方案中,所述方法旨在编辑目标细胞的基因组,特别是进行目标转基因的靶向插入。

[0476] 在一些实施方案中,靶细胞可被本发明的慢病毒颗粒感染。在一些实施方案中,靶细胞在其细胞外表面表达可被慢病毒颗粒识别,特别是被慢病毒颗粒的表面蛋白识别的至少一种蛋白或其部分或蛋白复合物。在一些实施方案中,该至少一种蛋白或蛋白复合物允许慢病毒颗粒结合(或锚定)和/或进入靶细胞内。

[0477] 在一些实施方案中,靶细胞表达CD4、CCR5和/或CXCR4表面蛋白。通常,CD4、CCR5和/或CXCR4与HIV-1表面蛋白Gp120和/或Gp41相互作用。

[0478] 在一些实施方案中,靶细胞是原代细胞或来自适合慢病毒感染的细胞系的细胞,例如,HEK 293T细胞、CHO细胞、HSC细胞、iPS细胞、原代免疫细胞、原代T细胞、原代NK细胞、原代巨噬细胞、原代肝细胞、K-562细胞和Jurkat-T细胞。

[0479] 在一些实施方案中,靶细胞人工表达至少一种允许慢病毒颗粒结合和/或进入的表面蛋白。“人工表达”是指细胞不内源性表达所述表面蛋白,但已被修饰或工程化以瞬时或稳定地表达所述表面蛋白。

[0480] 在一些实施方案中,靶细胞是从生物体、优选哺乳动物生物体、更优选人类生物体分离的细胞。在一些实施方案中,靶细胞是骨髓细胞。在一些实施方案中,靶细胞是淋巴细胞。

[0481] 在一些实施方案中,所述方法用于执行在靶细胞群中进行的靶基因插入,所述靶细胞群在从生物体中提取、在体外培养并重新引入生物体。根据该实施方案,所述方法被称为离体方法。

[0482] 应当理解,慢病毒颗粒的内容物在被递送到靶细胞中后,与包含至少一个适体(例如,MS2 RNA四环结合序列)的gRNA形成复合物的RNA引导的核酸酶或切口酶(例如,Cas9)将与融合至整合酶(来自具有受损整合活性的慢病毒载体和/或来自包含第三蛋白的融合蛋白,所述第三蛋白包含POL多聚蛋白或由POL多聚蛋白组成(当存在时),如上文所述)的适体结合蛋白(例如,MCP)相互作用。具体而言,这种相互作用通过适体/适体结合蛋白相互作用(例如,MS2/MCP)发生。

[0483] 在一些实施方案中,gRNA的位点特异性将RNA引导的核酸酶或切口酶/整合酶复合物活性引导至靶细胞基因组的特定区域或位点(例如,引导至安全港基因座或在靶细胞基因组中的基因内或附近)。

[0484] 在一些实施方案中,将目标转基因插入到gRNA特异性识别的位点。在一些实施方案中,目标转基因插入到距离gRNA特异性识别的序列1至20个核苷酸的位置,其中“1至20”包括1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19和20。

[0485] 在一些实施方案中,突变的整合酶活性导致目标转基因的靶向基因整合,也称为位点特异性整合。

[0486] 在一些实施方案中,目标转基因不插入到靶细胞的基因组的编码序列和/或调控序列中。在一些实施方案中,使用本发明的方法产生的靶向基因插入不会破坏、改变或降低一个或多个基因的表达。

[0487] 在一些实施方案中,将目标转基因插入到目标细胞的基因组的编码序列和/或调控序列中,导致编码序列部分或完全失活。

[0488] 在一些实施方案中,目标转基因是通过非同源末端连接(NHEJ)插入的。在一些实施方案中,目标转基因是通过同源重组(HR)插入的。

[0489] 在一些实施方案中,插入的目标转基因由细胞表达。在一些实施方案中,插入的目标转基因由细胞稳定表达。在一些实施方案中,插入的目标转基因的表达受启动子或调控元件的控制。

[0490] 在一些实施方案中,通过对靶位点(即,gRNA识别的位点)周围的基因组区域进行测序来控制目标转基因的位点特异性整合。

[0491] 在一些实施方案中,慢病毒颗粒数量与细胞数量的比率(感染复数,MOI)为MOI 0.5至MOI 40。

[0492] 本发明还涉及如上文所述的慢病毒颗粒群(加上可选的目标转基因,当所述目标转基因不包含在慢病毒颗粒群的慢病毒颗粒中时),用作药物。

[0493] 本发明还涉及如上文所述的慢病毒颗粒群(加上可选的目标转基因,当所述目标转基因不包含在慢病毒颗粒群的慢病毒颗粒中时),用于将目标转基因位点特异性整合到靶细胞的基因组中。

[0494] 本发明还涉及如上文所述的慢病毒颗粒群(加上可选的目标转基因,当所述目标转基因不包含在慢病毒颗粒群的慢病毒颗粒中时),用于治疗有需要的受试者的疾病。本发明还涉及一种治疗有需要的受试者的疾病的方法,包括向受试者施用治疗有效量的如上文所述的慢病毒颗粒群(加上可选的目标转基因,当所述目标转基因不包含在慢病毒颗粒群的慢病毒颗粒中时)。

[0495] 在一些实施方案中,疾病的治疗是通过将目标转基因位点特异性整合到有需要的受试者的细胞的基因组中来实现的。

[0496] 在一些实施方案中,所述疾病是基因疾病。在一些实施方案中,基因疾病的特征在于至少一个基因部分或完全沉默、失活或功能失调。在一些实施方案中,基因疾病的特征在于有需要的受试者的细胞中的至少一个基因编码至少一种无功能蛋白。在一些实施方案中,基因疾病的特征在于有需要的受试者的细胞中的至少一个基因编码至少一种对细胞具有有害作用的蛋白或多肽。

[0497] 在一些实施方案中,疾病是癌症。在一些实施方案中,癌症是血液癌症(例如,白血病)。

[0498] 通常,靶向基因插入会恢复与疾病有关的所述至少一个基因的正常表达和功能。

[0499] 在一些实施方案中,慢病毒颗粒群将在第二治疗剂之前、同时或之后施用。

[0500] 附图简要说明

[0501] 图1A-D涉及实现可编程插入的几种包装策略。图1A显示同时加载核酸酶和载荷

DNA(共同包装)比单独添加组分(编辑器+DNA)在靶向整合上更有效。图1B显示了具有融合至多聚蛋白GAG或多聚蛋白GAG-POL的适体结合蛋白(MCP)的LV载荷的靶向整合。图1C显示了具有不同包装策略的LV载荷的靶向整合:与NLS(NLS-Cas9)、与病毒蛋白R融合的整合酶(VPRIN-Cas9)、LEDGF(p75-LEDGF-Cas9)和Gag-Pol(GAGPOL-Cas9)共同包装。图1D显示了整合酶突变体慢病毒蛋白如何确定在同时的SpCas9核酸酶RNP递送和LV载体感染后中靶整合水平的效率。

[0502] 图2显示了使用不同RNA引导的核酸酶的靶向DNA整合:SpCas9、SaCas9、Cas12f和TnpB可有效用于靶向递送。

[0503] 图3显示了使用MS2系统在慢病毒颗粒中以DNA、RNA或蛋白形式递送Cas9核酸酶或与可编程转座酶(FiCAT)融合的Cas9后,NHEJ激活细胞的百分比。通过在报告细胞系中进行的NHEJ来测量加载。

[0504] 图4A-C涉及整合接合测序。图4A显示了在HEK293T细胞中使用野生型慢病毒载体(WT-LV)以及共同递送与MCP融合的突变整合酶、Cas9蛋白和含有MS2的gRNA(PILV)所获得的靶向整合滴度(TU/mL)的比较,比较了中靶整合和脱靶整合。图4B显示了在HEK293T细胞的每条染色体上检测的插入事件;19号染色体上的较大条带对应于中靶整合。图4C举例说明了整合接合测序。

[0505] 图5A和5B显示了Cas9引入慢病毒颗粒的加载策略。图5A显示了病毒生产概览和相应的基因组分。图5B说明了将Cas9加载到VLP中的七个(1至7)加载策略。第1行)过表达Cas9;第2行)与VPR(病毒衣壳蛋白)融合;第3行)与野生型病毒整合酶蛋白融合;第4行)与突变病毒整合酶蛋白融合;第5行)Cas9与VPR和突变病毒整合酶两者融合;第6行)与GAP-POL病毒基因融合;或第7行)与人类内源性p75蛋白融合。

[0506] 图6A和6B显示了Cas9在VLP中的加载和活性。图6A为蛋白质印迹结果,显示了上述1至7的七个加载策略中的Cas9相对加载效率;条件0代表没有Cas9存在的非整合对照。图6B显示了使用常规RNA(阴性)和带有四价MS2结构域(用于MCP特异性结合)的gRNA,在依赖于上述七个封装策略的病毒颗粒感染的细胞的下一代illumina靶向测序中,通过NHEJ读段测量的编辑活性。

[0507] 图7A和7B说明了用于Cas9在VLP中加载的MS2-MCP系统和活性。图7A是MS2-MCP系统作为gRNA中的适体用于Cas9在VLP中有效加载的示意图。图7B是蛋白质印迹的结果,显示了上述1至7的七个加载策略中的相对Cas9加载效率;条件0代表没有Cas9存在的非整合对照。在所有情况下,产生的细胞中表达的gRNA均是gRNA-MS2变体。

[0508] 图8A和8B显示了PILV(可编程插入慢病毒载体)在可编程插入报告细胞系中的表现。图8A是用于可编程整合检测的分裂GFP报告系统的示意代表图。报告载荷被克隆到使用LTR而非ITR的慢病毒载体中。图8B说明了在Hershey报告系统中测试的不同包装系统的可编程插入(GFP)和整体插入(RFP)容量。通过流式细胞术分析测量和比较了效率。

[0509] 图9说明了PILV在细胞系模型中可编程插入内源性靶标的表现。在HEK293T细胞中感染PILV包装策略和对照后,绘制了通过接合qPCR测量的靶向整合和通过Illumina测序测量的靶向NHEJ活性,以观察关联性。包装策略如下:“克隆”,具有已知靶向插入事件的克隆细胞系作为阳性对照;“NI”,非整合阴性对照;“PILV”,采用MCP-GAGPOL、MS2-gRNA和Cas9过表达的包装策略;“MCP-GAG”,MCP与GAG蛋白融合,gRNA-MS2用于包装Cas9;“Cas9_

Insynthetic”,策略3)与病毒整合酶蛋白野生型融合;“Cas9_IND64Vsynth”,策略4)与突变病毒整合酶蛋白融合;“Cas9_Ms2_mRNA”,MCP与GAG融合,表达具有MS2标签的Cas9以封装mRNA而非RNP;“VPRIN_Cas9-186+11”,策略5)Cas9与VPR和突变病毒整合酶两者融合;“GAGPOL-NLS-Cas9”,策略6)与GAP-POL病毒基因融合。

[0510] 图10A、10B和10C显示了改进的HDR供体的PILV性能。图10A是用于PILV的HDR设计供体和使用HDR的可编程整合机制的示意代表图。图10B显示了HEK293T细胞中所有三个供体的非整合和PILV载体的GFP信号。图10C显示了使用非整合慢病毒颗粒(NI)或PILV转导HDR载荷后7天的GFP信号。

[0511] 图11A和11B显示了用于T细胞工程化的改进的HDR供体的PILV性能。图11A是用于TRAC基因座的PILV的HDR设计供体的示意代表图。图11B表示在两个可选PBMC供体中原代T细胞中TRAC可编程插入的非整合GFP对照和PILV载体的GFP信号。

实施例

[0512] 通过下面的实施例进一步说明本发明。

[0513] 实施例1:使用慢病毒载体的靶向DNA整合

[0514] 材料和方法

[0515] 质粒构建

[0516] 从Addgene获得以下质粒:

[0517] -hCas9(质粒#41815),pcDNA™3.3-TOPO™质粒,在CMV启动子的控制下表达人类密码子优化的Cas9核酸酶;

[0518] -psPAX2(质粒#12260),空的第二代慢病毒包装质粒,包含gag、pol、tat和rev基因;

[0519] -pCMV-VSV-G(质粒#8454),在CMV启动子的控制下表达水疱性口炎病毒的G糖蛋白的质粒;

[0520] -pMDLg/pRRE(质粒#12251),第三代慢病毒包装质粒,表达HIV-1Gag(编码病毒体主要结构蛋白)、HIV-1Pol(编码HIV-1蛋白酶、HIV-1逆转录酶和HIV-1整合酶)和HIV-1Rev反应元件(RRE,促进RNA从细胞核中输出的Rev蛋白的结合位点);

[0521] -pRSV Rev(质粒#12253),第三代慢病毒包装质粒,表达Rev蛋白。

[0522] 使用QuikChange Lightning多位点定向诱变试剂盒(Agilent#210513)获得整合酶突变。整合酶突变体的序列公开于SEQ ID NO:2至SEQ ID NO:9。

[0523] 所有其余载体均通过Golden Gate组装法使用Esp3I和T4连接酶构建。使用SEQ ID NO:74的质粒pSICO(Addgene参考质粒#41815;包含目标转基因的通用载体)、SEQ ID NO:75的pRRL双重(包含目标转基因的通用载体)、SEQ ID NO:76的psMCP-GAG和SEQ ID NO:77的psMCP-GAGPOL。

[0524] 细胞培养

[0525] HEK293T细胞(ATCC CRL-3216)在补充有高葡萄糖(Gibco,Thermo Fisher)、10%胎牛血清、2mM谷氨酰胺和100U青霉素、0.1mg/mL链霉素的杜氏改良伊格尔培养基(DMEM)中生长。

[0526] K-562细胞(ATCC CRL-3343)和Jurkat-T细胞(克隆E6-1,ATCC IB 152)在补充有

10％ FBS、1％青霉素-链霉素(Gibco)和1％ GlutaMAX(100x)(Gibco)的RPMI 1640培养基(Gibco)中生长。

[0527] 所有细胞均在37℃、95％空气和5％ CO_2下生长。

[0528] 慢病毒生产

[0529] 慢病毒质粒根据Addgene网站https://www.addgene.org/protocols/lentivirus production/提供的方案生产,并重述如下:

[0530] 介绍

[0531] 该方案可以使用聚乙烯亚胺(PEI)转染方案将慢病毒载体转染到Lenti-X293T细胞中,从而产生慢病毒。该程序可修改为使用其他包装细胞系或转染试剂。慢病毒一旦产生,便可用于各种下游应用,例如稳定细胞系的生产。

[0532] 工作流程时间表

[0533] 第0天:接种293T包装细胞

[0534] 第1天(下午):转染包装细胞

[0535] 第2天(上午):转染后18小时。移除培养基,换成新鲜培养基

[0536] 第3-4天(上午):收获病毒

[0537] 试剂制备

[0538] 1.DMEM完全培养基:10％v/v FBS和4mM L-丙氨酰-L-谷氨酰胺

[0539] 向500mL瓶装DMEM高糖培养基中添加55mL热灭活FBS和11mL 200mM L-丙氨酰-L-谷氨酰胺,4℃保存。

[0540] 2.25mM二磷酸氯喹

[0541] 将0.129g二磷酸氯喹盐溶解于10mL无菌水中。

[0542] 通过0.22μm过滤器过滤除菌。

[0543] 分装为50～100μL,-20℃保存。

[0544] 分装试样可以解冻,并在使用前储存于4℃。解冻的分装试样应在1-2个月后丢弃。

[0545] 3.1mg/mL PEI,线性MW 25,000Da

[0546] 将100mg粉末溶解于100mL去离子水中。

[0547] 搅拌的同时,缓慢添加盐酸直至溶液变清澈。

[0548] 检查溶液的PH值。

[0549] 使用盐酸或氢氧化钠将pH值调整至7.0。通常溶液呈碱性,需要先用盐酸进行调整。

[0550] 让溶液混合10分钟,然后重新检查pH值以确保其没有波动。

[0551] 将溶液通过0.22μm膜过滤。

[0552] 将500至1000μL分装到无菌试管中。

[0553] 将试管储存在-80℃下。

[0554] 解冻后,溶液可在4℃下保存长达2个月。2个月后,丢弃试管并解冻新的工作储备液。

[0555] 对于每一批新的1mg/mL PEI和每种细胞系,需要根据经验确定最佳DNA质量:PEI质量比率。

[0556] 程序

[0557] 将293T包装细胞以3.8×10^{6}个细胞/板接种于10cm组织培养板中的DMEM完全培养基中。

[0558] 将细胞在37℃、5% CO2下孵育约20小时。

[0559] 轻柔吸出培养基,添加10mL含有25μM二磷酸氯喹的新鲜DMEM完全培养基,孵育约5小时。

[0560] 对于10mL DMEM完全培养基,添加10μL 25mM二磷酸氯喹。

[0561] 制备3种转染质粒的混合物:

[0562] psPAX2:1.3pmol

[0563] pMD2.G:0.72pmol

[0564] 转移质粒:1.64pmol

[0565] OptiPro SFM至总体积为500μL

[0566] 将上述500μL混合物稀释到500μL PEI-OptiPro SFM中,PEI足量以使得μg DNA:μg PEI比率为1:3(每10cm培养皿总共1000μL)。

[0567] 使用转移质粒pHAGE TRE dCas9-KRAB(质粒DNA总μg为27.8μg),每10cm培养皿中的416.6μL OptiPro SFM中将含有83.4μL 1mg/mL PEI。

[0568] 将稀释的PEI轻柔添加至稀释的DNA中。在逐滴添加稀释的PEI的同时轻柔摇晃稀释的DNA管。将混合物在室温下孵育15至20分钟。

[0569] 小心地将转染混合物转移到Lenti-X 293T包装细胞中。逐滴添加转染混合物,注意不要移动细胞。

[0570] 将细胞孵育18小时,或直到第二天早上。

[0571] 第二天早上,小心地吸出培养基。用15mL DMEM完全培养基更换培养基。

[0572] 孵育细胞。

[0573] 可在转染后48、72和96小时收获病毒,可单独收获或将所有单独收获物合并在一起。如果合并收获物,则将收获的培养基转移到聚丙烯储存管中,并在收获之间储存在4℃下。

[0574] 将病毒上清液以~500g的速度离心5分钟,以使收获过程中收集的任何包装细胞沉淀。

[0575] 将上清液通过0.45μm PES过滤器进行过滤。

[0576] 病毒上清液可以在4℃下保存数小时,但应尽快分装并在液氮中快速冷冻并保存在-80℃下,以避免滴度损失。

[0577] 在转染前一天接种了4.9×10^{6}HEK293T细胞的10cm培养皿中生产细胞,转染使用0.72pmol pCMV-VSV-G包膜质粒、1.64pmol pSICO或pRRL双重载荷(转移质粒)和1.30pmol psPAX2质粒(野生型或pol突变体)。

[0578] 对于核酸酶包装,0.65pmol psPAX2和0.65pmol psMCP-GAG或psMCP-GAGPOL质粒与0.65pmol核酸酶质粒一起使用。

[0579] 对于转染,将质粒在500μl Optimem和100mg聚乙烯亚胺(PEI)中混合。质粒转染后两天,收获上清液并过滤,在4000g和4℃下离心过夜。弃去上清液并重悬慢病毒颗粒以达到100x载体浓度。

[0580] 核酸酶活性的测定

[0581] 对于VPR/IN/LEDGF核酸酶融合活性,使用交通灯报告基因(Certo et al.,2011.Nat Methods.8(8):671-676)HEK293T细胞系,该细胞系含有框外mCherry ORF,其在Cas9靶向后通过NHEJ活性重建。使用BD LSR Fortessa仪器(黄绿561nm激光,带610/20滤光片)通过流式细胞术测量荧光。

[0582] 通过在240 000个细胞/孔的p12细胞中用gRNA转染细胞,来评估SaCas9、AsCas12f和TnpB编辑以及gRNA中的MS2支架在HEK293T中的影响。添加核酸酶后三天,将细胞沉淀并用KAPPA快速提取法来提取DNA。对相应的靶位点进行PCR扩增,并通过合成MiSeq™平台使用MiSeq试剂盒v2在Illumina测序仪上进行测序(300个循环,2×150配置)。

[0583] 中靶整合的测定

[0584] 建立了报告细胞系(Hershey)来评估靶向整合,其中在中靶整合后荧光蛋白ORF的重建会产生可测量的荧光信号(GFP)。

[0585] 将含有1/2Emerald GFP(emGFP)和1/2内含子的病毒载荷质粒包装到具有相应包装系统和/或整合酶突变体的慢病毒载体中。将剪接受体前的无启动子的emGFP的C端半部分(C-t)随机插入HEK293T细胞基因组中,以构建报告细胞系。在C-t emGFP上游添加“靶位点”。

[0586] 为了进行实验,将200 000个HEK293T Hershey报告细胞接种在12个孔/板中,并在次日在转染的同时感染慢病毒颗粒,这导致在添加载荷质粒后产生emGFP重建。

[0587] 慢病毒采用标准方案生产。感染后四天,通过FACS测量emGFP荧光信号(BD LSR Fortessa仪器;蓝色488nm激光,带530/30滤光片,黄绿561nm激光,带610/20滤光片)。

[0588] 整合接合测序

[0589] INSERTseq方法结合了整合DNA的靶向扩增、基于UMI的PCR偏差校正和Oxford Nanopore长读段测序,可对基因组中的DNA整合进行稳健分析。INSERT-seq能够检测发生频率达0.1%的事件。INSERT-seq可完整处理所有插入,而无论重复的大小如何。

[0590] 文库制备和测序

[0591] 使用Nanobind试剂盒(Circulomics,目录号NB-900-001-01)提取DNA,使用g-TUBE™(Covaris,目录号520079)剪切成约2kb的片段。将Nanopore PCR Barcoding Kit(SQK PBK004)中的WGP引物混合物额外添加到第二次PCR中。在Flongle R9.4.1流动池中进行测序,总输出量为约300 000个读段。为了计算检测限(LOD),将来自HEK293T细胞的单克隆样品(带有一个真正的慢病毒插入)用WT基因组材料以1/100、1/1 000和1/10 000的比例稀释。稀释物在Flongle R9.4.1流动池中进行测序,分别获得500 000、500 000和150 000个读段的测序输出。分析是在INSERTseq分析之后进行的。

[0592] 整合位点分析

[0593] 使用Guppy 4.0.11(由ONT通过其网页https://community.nanoporetech.com提供)对Nanopore原始读段进行碱基调用(base-call)。

[0594] 使用源自NanoPack的NanoStats NanoQC和NanoPlot评估读段质量(De Coster et al.,2018.Bioinformatics.34(15):2666-2669)。使用源自NanoPack的NanoFilt按质量(>10)和长度(>200)对读段进行过滤。使用两个先前发布的管线的组合和改编,通过UMI对读段进行聚类(ONT发布的pipeline-umi-amplicon,https://github.com/nanoporetech/pipeline-umi-amplicon)。聚类是通过使用Python脚本提取UMI序列进行的,序列使用

vsearch(Rognes et al.,2016.PeerJ.4:e2584)进行聚类,而聚类的共有序列则通过进行两轮racon修饰(Vaser et al.,2017.Genome Res.27(5):737-746)和两轮medaka(https://github.com/nanoporetech/medaka)获得。

[0595] 为了分析插入,读段通过过滤以强制存在所使用的接合体(adapter),并使用cutadapt(Martin,2011.EMBnet j.17(1):10-12)进行修剪以从读段中移除接合体和插入序列。使用minimap2(Li,2018.Bioinformatics.34(18):3094-3100)“map-ont”默认参数将读段映射到参考基因组上,并使用Python脚本进行过滤,选择映射质量高于30的独特映射读段。使用bedtools(Quinlan&Hall,2010.Bioinformatics.26(6):841-842)返回第一输出,格式为bed,包含所有映射的读段。之后,使用Python脚本进行峰调用步骤,峰按形状过滤。当将峰值覆盖拟合至beta分布的残差平方和(RSS)低于1时,认为峰通过了形状过滤器。

[0596] 优化的INSERTseq方案的实施

[0597] 提取基因组DNA,进行片段化、末端修复和A-尾化,然后连接包含用于读段聚类的UMI的接合体以及用于在计算管线中进行样本解复用的条形码。读段通过UMI聚类,整合序列和接合体经过过滤和修剪,将读段映射至参考基因组,并报告和注释显著峰。

[0598] 结果

[0599] 核酸酶包装策略

[0600] 鉴于慢病毒DNA在双链断裂(DSB)中整合,因此推断将功能性Cas9-gRNA包装到慢病毒颗粒中应该会产生靶向整合。

[0601] 测试了六种不同的包装系统配置,以通过将SpCas9蛋白包装到慢病毒中来实现靶向整合:

[0602] (i)在载体生产时共表达spCas9,

[0603] (ii)共表达与病毒体靶向蛋白Vpr p6-GAG相互作用结构域融合的SpCas9,如先前所述(Indikova&Indik,2020.Nucleic Acids Res.48(14):8178-8187),

[0604] (iii)与整合酶蛋白融合的SpCas9,

[0605] (iv)与VPR整合酶融合,如先前所述(Montagna et al.,2018.Mol Ther Nucleic Acids.12:453-462),

[0606] (v)与LEDGF p75染色质对接因子的整合酶结合结构域融合,如先前所述(Hare et al.,2009.PLoS Pathog.5(1):e1000259),以及

[0607] (vi)与包装质粒中的GAGPOL多聚蛋白直接融合。

[0608] 在所有包装系统中均检测到SpCas9蛋白,但在慢病毒施用至培养细胞后未检测到非同源末端连接(NHEJ)编辑。一种假设是,缺乏编辑可能是由于缺少SpCas9 gRNA,因为U6驱动的表达将gRNA定位在细胞核中,而细胞质gRNA表达已被证明可以挽救gRNA在慢病毒衍生囊泡中的包装。

[0609] 使用MCP-GAGPOL包装融合蛋白和含有SpCas9 gRNA的MS2茎-环,以将gRNA募集到慢病毒颗粒中。使用此配置,挽救了用慢病毒颗粒转导时的编辑活性(图1A)。

[0610] 比较了包含适体结合蛋白MS2衣壳蛋白(MCP)与慢病毒多聚蛋白GAG-POL或单独的多聚蛋白GAG融合的慢病毒颗粒之间的靶向整合效率(图1B)。两种情况导致相似的靶向整合率。

[0611] 在Jurkat-T细胞中在同时的慢病毒感染和Cas9 RNP递送时,观察到靶向整合。使

用四种不同的包装策略生成可编程慢病毒(PILV):Cas9与NLS(NLS-Cas9)、与融合至病毒蛋白R的整合酶(VPRIN-Cas9)、与LEDGF(p75-LEDGF-Cas9)或与Gag-Pol(GAGPOL-Cas9)共同包装。NLS-Cas9和VPRIN-Cas9产生最高的插入率(图1C)。

[0612] 测试了几种整合酶(IN)突变体,所有突变体都表现出比野生型整合酶更好的靶向整合性能(图1D)。

[0613] 最后,使用MS2系统在慢病毒颗粒中使用编辑工具(所述编辑工具为Cas9核酸酶或包含Cas9和可编程转座酶(PT条件)的融合蛋白)递送DNA载荷时,比较了NHEJ激活细胞的比例。此外,编辑工具以DNA、RNA或蛋白(RNP)形式提供。

[0614] 与其他核酸酶的兼容性

[0615] 鉴于自SpCas9以来已描述的可编程核酸酶种类繁多,我们用不同的可编程核酸酶(AsCas12和TnpB)测试了可编程慢病毒包装系统(PILV)。在所有系统中均检测到靶向整合(图2),表明该系统与任何核酸酶兼容。

[0616] 与可编程转座酶的共同递送

[0617] 使用MCP-MS2系统,将Cas9或包含Cas9和高活性PiggyBac转座酶(FiCAT,具有R372A/K375A/D450N取代)的融合蛋白加载到慢病毒颗粒中。在用以mRNA或蛋白(RNP)形式加载基因编辑器的慢病毒颗粒转导的交通灯报告HEK293T细胞中,测量编辑活性(图3)。

[0618] 结果显示,两种情况下都存在基因编辑活性,但当基因编辑器以蛋白形式递送时获得了更好的结果。

[0619] 中靶整合的测量

[0620] 使用INSERTseq方法测量中靶整合,如图4B和图4C所示。

[0621] 如图4A所示,在野生型慢病毒颗粒(WT-LV)和可编程慢病毒(PILV)之间比较了总体整合和中靶整合活性。与WT LV相比,PILV能够实现精确的中靶整合,使其成为治疗性应用的最佳系统。

[0622] 实施例2:将Cas9引入慢病毒颗粒的策略

[0623] 为了将Cas9核酸酶引入病毒样颗粒(VLP)中,可以采用几种在慢病毒生产细胞中表达核酸酶的方法。除了在生产细胞的细胞质上过表达和被动加载(图5A)之外,还探索了其他6种Cas9加载策略。这些策略(图5B)包括:1)Cas9过表达;2)与VPR(病毒衣壳蛋白)融合;3)与野生型病毒整合酶蛋白融合;4)与突变病毒整合酶蛋白融合;5)Cas9与VPR和突变病毒整合酶两者融合;6)与GAP-POL病毒基因融合;或7)与人类内源性p75蛋白融合。

[0624] 材料和方法

[0625] 细胞培养和质粒克隆

[0626] HEK293T细胞(ATCC CRL-3216)在补充有高葡萄糖(Gibco,Thermo Fisher)、10%胎牛血清、2mM谷氨酰胺和100U青霉素、0.1mg/mL链霉素的杜氏改良伊格尔培养基(DMEM)中生长。

[0627] K-562细胞(ATCC CRL-3343)和Jurkat-T细胞(克隆E6-1,ATCC IB 152)在补充有10% FBS、1%青霉素-链霉素(Gibco)和1% GlutaMAX(100x)(Gibco)的RPMI 1640培养基(Gibco)中生长。

[0628] 所有细胞系均在37℃、95%空气和5% CO_2下生长。

[0629] 实施例1中未包括的其他载体通过Golden Gate组装法使用Esp3I和T4连接酶构

建。使用含有密码子优化的IN和Cas9的融合体的质粒pRRL_Cas9_Insynth(SEQ ID NO:79),含有VPR、突变的IN和Cas9的融合体的质粒pRRL_VPR-IN186opt-Cas9(SEQ ID NO:80)。

[0630] 慢病毒生产

[0631] 慢病毒质粒根据Addgene网站https://www.addgene.org/protocols/lentivirus production/提供的方案生产,并重述如下:

[0632] 介绍

[0633] 该方案可以用于通过聚乙烯亚胺(PEI)转染方案将慢病毒载体转染到Lenti-X 293T细胞中,从而产生慢病毒。该程序可修改为使用其他包装细胞系或转染试剂。慢病毒一旦产生,便可用于各种下游应用,例如稳定细胞系的生产。

[0634] 工作流程时间表

[0635] 第0天:接种293T包装细胞

[0636] 第1天(下午):转染包装细胞

[0637] 第2天(上午):转染后18小时。移除培养基,换成新鲜培养基

[0638] 第3-4天(上午):收获病毒

[0639] 试剂制备

[0640] 1.DMEM完全培养基:10%v/v FBS和4mM L-丙氨酰-L-谷氨酰胺

[0641] 向500mL瓶装DMEM高糖培养基中添加55mL热灭活FBS和11mL 200mM L-丙氨酰-L-谷氨酰胺。4℃保存。

[0642] 2.25mM二磷酸氯喹

[0643] 将0.129g二磷酸氯喹盐溶解于10mL无菌水中。

[0644] 通过0.22μm过滤器过滤除菌。

[0645] 分装为50~100μL,-20℃保存。

[0646] 分装试样可以解冻,并在使用前储存于4℃。解冻的分装试样应在1-2个月后丢弃。

[0647] 3.1mg/mL PEI,线性MW 25,000Da

[0648] 将100mg粉末溶解于100mL去离子水中。

[0649] 搅拌的同时,缓慢添加盐酸直至溶液变清澈。

[0650] 检查溶液的PH值。

[0651] 使用盐酸或氢氧化钠将pH值调整至7.0。通常溶液呈碱性,需要先用盐酸进行调整。

[0652] 让溶液混合10分钟,然后重新检查pH值以确保其没有波动。

[0653] 将溶液通过0.22μm膜过滤。

[0654] 将500至1000μL分装到无菌试管中。

[0655] 将试管储存在-80℃下。

[0656] 解冻后,溶液可在4℃下保存长达2个月。2个月后,丢弃试管并解冻新的工作储备液。

[0657] 对于每一批新的1mg/mL PEI和每种细胞系,需要根据经验确定最佳DNA质量:PEI质量比率。

[0658] 程序

[0659] 将293T包装细胞以3.8×10^6个细胞/板接种于10cm组织培养板中的DMEM完全培养

基中。

[0660] 将细胞在37℃、5% CO2下孵育约20小时。

[0661] 轻柔吸出培养基,添加10mL含有25μM二磷酸氯喹的新鲜DMEM完全培养基,孵育约5小时。

[0662] 对于10mL DMEM完全培养基,添加10μL 25mM二磷酸氯喹。

[0663] 制备3种转染质粒的混合物:

[0664] psPAX2:1.3pmol

[0665] pMD2.G:0.72pmol

[0666] 转移质粒:1.64pmol

[0667] OptiPro SFM至总体积为500μL

[0668] 将上述500μL混合物稀释到500μL PEI-OptiPro SFM中,PEI足以使得μgDNA:μg PEI比率为1:3(每10cm培养皿总共1000μL)。

[0669] 使用转移质粒pHAGE TRE dCas9-KRAB(质粒DNA总μg为27.8μg),每10cm培养皿中的416.6μL OptiPro SFM中将含有83.4μL 1mg/mL PEI。

[0670] 将稀释的PEI轻柔添加至稀释的DNA中。在逐滴添加稀释的PEI的同时轻柔摇晃稀释的DNA管。将混合物在室温下孵育15至20分钟。

[0671] 小心地将转染混合物转移到Lenti-X 293T包装细胞中。逐滴添加转染混合物,注意不要移动细胞。

[0672] 将细胞孵育18小时,或直到第二天早上。

[0673] 第二天早上,小心地吸出培养基。用15mL DMEM完全培养基更换培养基。

[0674] 孵育细胞。

[0675] 可在转染后48、72和96小时收获病毒,可单独收获或将所有单独收获物合并在一起。如果合并收获物,则将收获的培养基转移到聚丙烯储存管中,并在收获之间储存在4℃下。

[0676] 将病毒上清液以~500g的速度离心5分钟,以使收获过程中收集的任何包装细胞沉淀。

[0677] 将上清液通过0.45μm PES过滤器进行过滤。

[0678] 病毒上清液可以在4℃下保存数小时,但应尽快分装并在液氮中快速冷冻并保存在-80℃下,以避免滴度损失。

[0679] 在转染前一天接种了4.9×10^{6}HEK293T细胞的10cm培养皿中生产细胞,转染使用0.72pmol pCMV-VSV-G包膜质粒、1.64pmol pSICO或pRRL双重载荷(转移质粒)和1.30pmol psPAX2质粒(野生型或pol突变体)。

[0680] 对于核酸酶包装,0.65pmol psPAX2和0.65pmol psMCP-GAG或psMCP-GAGPOL质粒与0.65pmol核酸酶质粒一起使用。

[0681] 对于转染,将质粒在500μl Optimem和100mg聚乙烯亚胺(PEI)中混合。质粒转染后两天,收获上清液并过滤,在4000g和4℃下离心过夜。弃去上清液并重悬慢病毒颗粒以达到100x载体浓度。

[0682] 核酸酶活性的测定

[0683] 所生产的使用不同策略加载Cas9和gRNA的慢病毒颗粒用于感染HEK293T或其他细

胞系模型。在病毒颗粒中添加核酸酶后三天,将细胞沉淀并用KAPPA快速提取法来提取DNA。对相应的靶位点进行PCR扩增,并通过合成MiSeq™平台使用MiSeq试剂盒v2在Illumina测序仪上进行测序(300个循环,2×150配置)。

[0684] 核酸酶活性的测定

[0685] 将慢病毒颗粒在RIPA缓冲液中裂解以获得蛋白。按照标准方法通过蛋白质印迹分析蛋白,简而言之:将蛋白加载并运行于SDS-PAGE凝胶(Invitrogen)中,然后转移到硝酸纤维素印迹中,使用抗Cas9和抗p24抗体(AbCam)以及HRP偶联二抗进行蛋白检测。使用鲁米诺作为底物,通过化学发光显色检测和定量蛋白。

[0686] 结果

[0687] 尽管通过7种测试策略中的大多数都可以实现Cas9的存在(图6A),使用这些方法无法检测到核酸酶活性(图6B)。

[0688] 接下来,我们利用MS2-MCP RNA蛋白系统来利用gRNA与Cas9的相互作用,测试了几种可选加载策略。在其中一项策略中,我们共表达了gRNA中的四价MS2和与MCP蛋白融合的修饰的GAG-POL病毒基因,以将Cas9主动加载到病毒颗粒中,同时加载gRNA,用于充分靶向和切割感染细胞中的靶DNA(图6A)。

[0689] 该方法表明,当使用MCP-MS2系统时,Cas9的存在增加(图7B),并且感染细胞中的核酸酶活性得到恢复,但令人惊讶的是,当使用MCP-MS2系统时,感染细胞中的核酸酶活性完全恢复(图6B)。

[0690] 实施例3:性能最佳的Cas9加载PILV原型的验证

[0691] 材料和方法

[0692] 使用根据不同加载策略生产(生产方法与实施例1和2相同,不同的Cas9和gRNA加载策略)的慢病毒颗粒来感染HEK293T报告细胞或K562细胞。对感染细胞分析如下。

[0693] 核酸酶活性的测定

[0694] 在病毒颗粒中添加核酸酶后三天,将细胞沉淀并用KAPPA快速提取法来提取DNA。对相应的靶位点进行PCR扩增,并通过合成MiSeq™平台使用MiSeq试剂盒v2在Illumina测序仪上进行测序(300个循环,2×150配置)。

[0695] 中靶整合的测定

[0696] 建立了报告细胞系(Hershey)来评估靶向整合,其中在中靶整合时荧光蛋白ORF的重建会产生可测量的荧光信号(GFP)。

[0697] 利用相应包装系统和/或整合酶突变体将含有1/2Emerald GFP(emGFP)和1/2内含子的病毒载荷质粒包装在慢病毒载体中。将剪接受体前的无启动子的emGFP的C端半部分(C-t)随机插入HEK293T细胞基因组中,以构建报告细胞系。在C-t emGFP上游添加“靶位点”。

[0698] 为了进行实验,将200 000个HEK293T Hershey报告细胞接种在12个孔/板中,并在次日在转染的同时感染慢病毒颗粒,这导致在加入载荷质粒时emGFP的重建。

[0699] 慢病毒采用标准方案生产。感染后四天,通过FACS测量emGFP荧光信号(BD LSR Fortessa仪器;蓝色488nm激光,带530/30滤光片,黄绿561nm激光,带610/20滤光片)。

[0700] 对于K562细胞,在感染后5天提取了基因组DNA,并通过半定量标准方法(SYBR Green qPCR,Applied Biosystems)对靶位点和病毒载荷的连接PCR进行了扩增,并与总病

毒拷贝数和内源性基因拷贝数进行了比较。

[0701] 结果

[0702] 在建立了Cas9和gRNA的加载系统以将有效的核酸酶活性并入感染的供体后,在整合酶缺陷病毒颗粒的背景下测试几种整合酶变体,以实现可编程整合。

[0703] 为了评估gRNA在特定靶位点的可编程整合,我们使用了基于分裂GFP报告基因的报告HEK293T细胞系(图8A)。评估了每种包装系统的效率,结果表明,性能最佳的系统是MCP-GAG与gRNA-MS2的融合以及使用MCP-MS2系统封装mRNA而不使用mRNA。使用MCP-GAGPOL的PILV显示出非常有前景的结果(图8B)。通过提高生产率,使用MCP-GAGPOL的PILV在后续重复中得到了改进。与整合酶或POL直接融合的结果最差。

[0704] 除了用于可编程整合的细胞报告系统外,还测试并验证了PILV在可选细胞系模型K562的内源性靶位点中的可编程插入效率。通过接合靶向qPCR评估了游离体衰变后的可编程整合效率,将相对拷贝数作图并与靶向NHEJ活性进行了比较(图9)。策略1涉及MCP-GAGPOL融合和MS2-MCP相互作用于gRNA和MCP-GAG融合,显示出最有前景的结果,其次是mRNA封装在病毒样颗粒中而非RNP。

[0705] 实施例4:通过同源性依赖重组(HDR)对PILV基因组的可编程整合

[0706] 材料和方法

[0707] 细胞培养和质粒克隆

[0708] HEK293T细胞(ATCC CRL-3216)在补充有高葡萄糖(Gibco,Thermo Fisher)、10%胎牛血清、2mM谷氨酰胺和100U青霉素、0.1mg/mL链霉素的杜氏改良伊格尔培养基(DMEM)中生长。

[0709] K-562细胞(ATCC CRL-3343)和T细胞(分离自人类供体,并激活)在补充有10%FBS、1%青霉素-链霉素(Gibco)和1% GlutaMAX(100x)(Gibco)的RPMI 1640培养基(Gibco)中生长。

[0710] 从多达4名人类供体的白细胞层(buffy coat)分离T细胞;纯化PBMC并使用标准方法激活T细胞以在感染前进行主动增殖。

[0711] 所有细胞系均在37℃、95%空气和5% CO_2下生长。

[0712] 实施例1和2中未包括的其余载体是使用Esp3I和T4连接酶通过Golden Gate组装法构建的。使用质粒pLV RAB11A-HA-GFP-HA,带有GFP标签的慢病毒载荷供体,用于在RAB11A位点进行靶向整合,SEQ ID NO:81;质粒pLV RAB11A-HA-GFPinv-HA,带有反向GFP标签的慢病毒载荷供体,用于在RAB11A位点进行靶向整合,SEQ ID NO:82;质粒pLV FBL-HA-mGFP-HA,带有GFP标签的慢病毒载荷供体,用于在FBL位点进行靶向整合,SEQ ID NO:83;质粒pLV GFP-TRAC-HA,带有GFP标签的慢病毒载荷供体,用于在TRAC位点进行靶向整合,SEQ ID NO:84;质粒pLV eGFP-TRAC-HA,带有eGFP标签的慢病毒载荷供体,用于在TRAC位点进行靶向整合,SEQ ID NO:85。

[0713] 中靶整合的测定

[0714] 原代T细胞或HEK293T、K562细胞系模型用通过与实施例1和2相同的方法产生的病毒颗粒感染,其中在内源基因上带有HDR基因捕获报告子,其中在蛋白上添加了GFP标签,以进行位点特异性插入检测。

[0715] 为了进行实验中,用慢病毒颗粒感染200 000个HEK293T或K562,或500000个活化T

细胞。用FACS(BD LSR Fortessa仪器;蓝色488nm激光,带530/30滤光片,黄绿561nm激光,带610/20滤光片)在不同时间点测量GFP荧光信号,以追踪信号。

[0716] 结果

[0717] 通过生成用于可编程插入的病毒基因组,其具有与目标靶位点同源的同源臂以提高在活跃分裂细胞中的效率,从而改进了载货供体。设计并生成了三种可选的PILV载货(图10A)。在HEK293T细胞中(图10B),RAB11A报告构建体产生了高水平的非特异性信号,但该启动子的反向版本大大降低了噪声信号并增加了PILV条件下的荧光,这与载荷的可编程插入相一致。图10C显示两种细胞系上的PILV信号均略有增加,这与HDR载荷的可编程插入相一致。

[0718] 此外,FBL报告基因也显示出类似的结果,HEK293T细胞中PILV的特异性GFP信号高达10%(图10B)。在细胞模型K562和JurcaT中也观察到类似的HDR载荷结果,存在游离性泄漏表达,并且两种载荷中的PILV信号均增加(图10C)。

[0719] 为了验证慢病毒载体的可编程整合在细胞工程化中的治疗潜力,设计并构建了TRAC特异性基因陷阱,用于生成在TRAC基因座中插入治疗性载货的T细胞(图11A)。在两个不同PBMC供体的原代T细胞中测试了基于HDR的TRAC特异性PILV载荷的可编程插入。与无转导效率的非整合对照相比,可编程插入载体的转导效率超过50%(图11B)。

[0720] 这些结果验证了PILV的转移潜力,作为一种安全、可控的对T细胞和其他原代细胞进行治疗目的的工程改造的方式。

[0721] 总结

[0722] 哺乳动物基因组中基因大小片段的靶向DNA整合可以实现精确添加治疗信息和新性能。慢病毒载体在治疗上用于不受控制地插入转基因以离体应用于先进细胞治疗,以及体内基因治疗。通过突变整合酶蛋白并将SpCas9核酸酶整合到慢病毒颗粒中,我们观察到慢病毒载体可以重新编程以精确整合转基因DNA。

[0723] 精确的DNA整合在永生化细胞模型和原代细胞中是一致的。PILV还可以递送转基因至小鼠肝脏的ROSA26基因座中。我们还在同时递送两个转基因后检测到PILV顺式整合,这为离体模型多重KO-KI提供了一种可靠的方法。通过在慢病毒生产时改变两个包装载体,慢病毒载体可以变得可编程,从而实现靶向性DNA添加,可用于离体细胞制造、体内基因替换和哺乳动物基因组工程化。

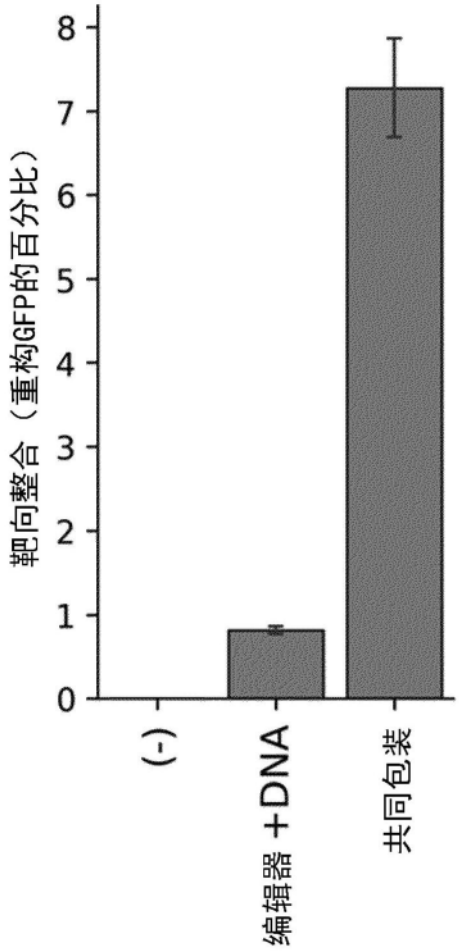


图1A

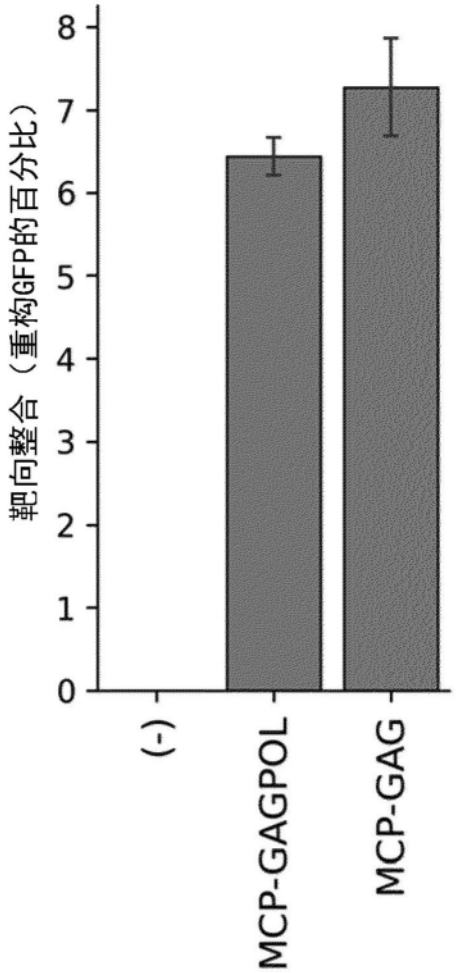


图1B

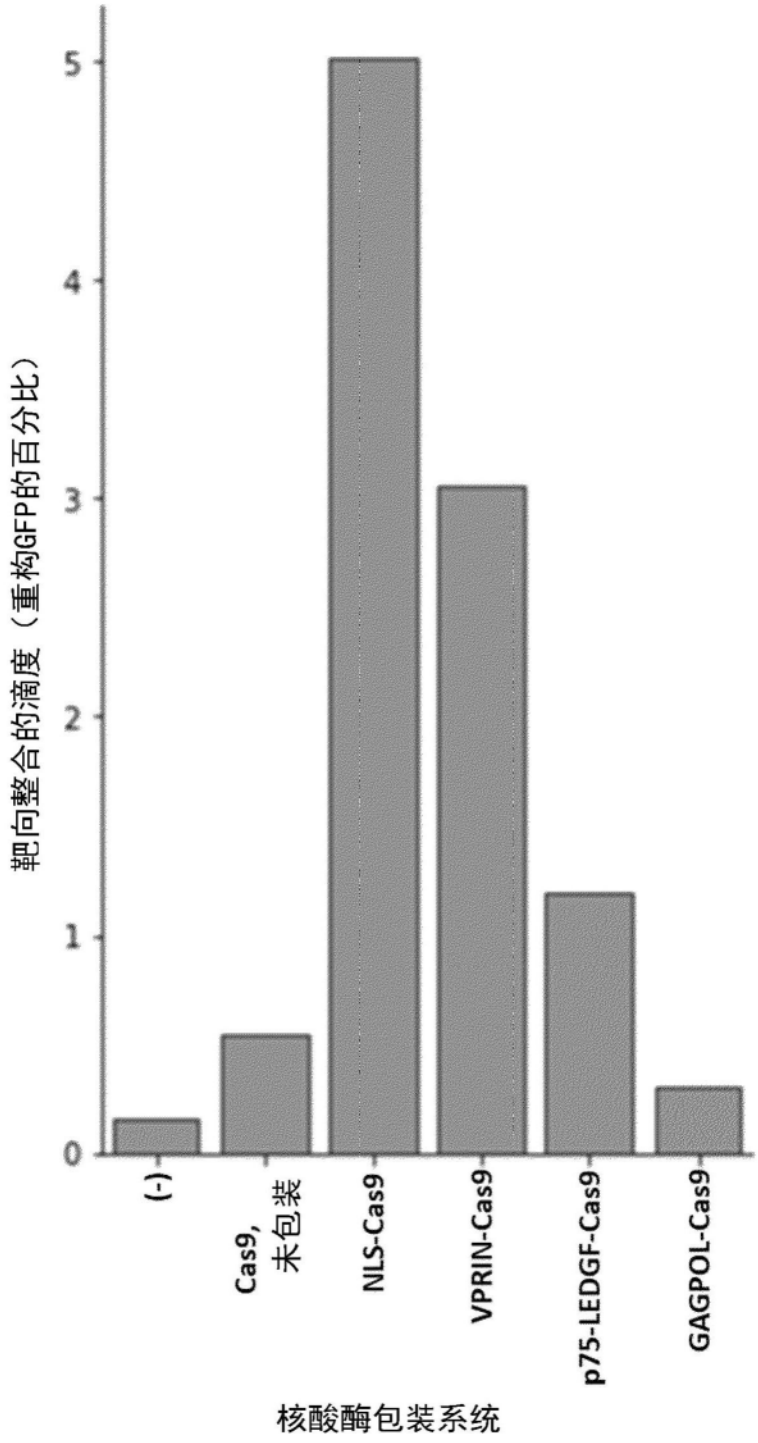

靶向整合的滴度（重构GFP的百分比）
0
1
2
3
4
5
(-)
Cas9,
未包装
NLS-Cas9
VPRIN-Cas9
p75-LEDGF-Cas9
GAGPOL-Cas9
核酸酶包装系统

图1C

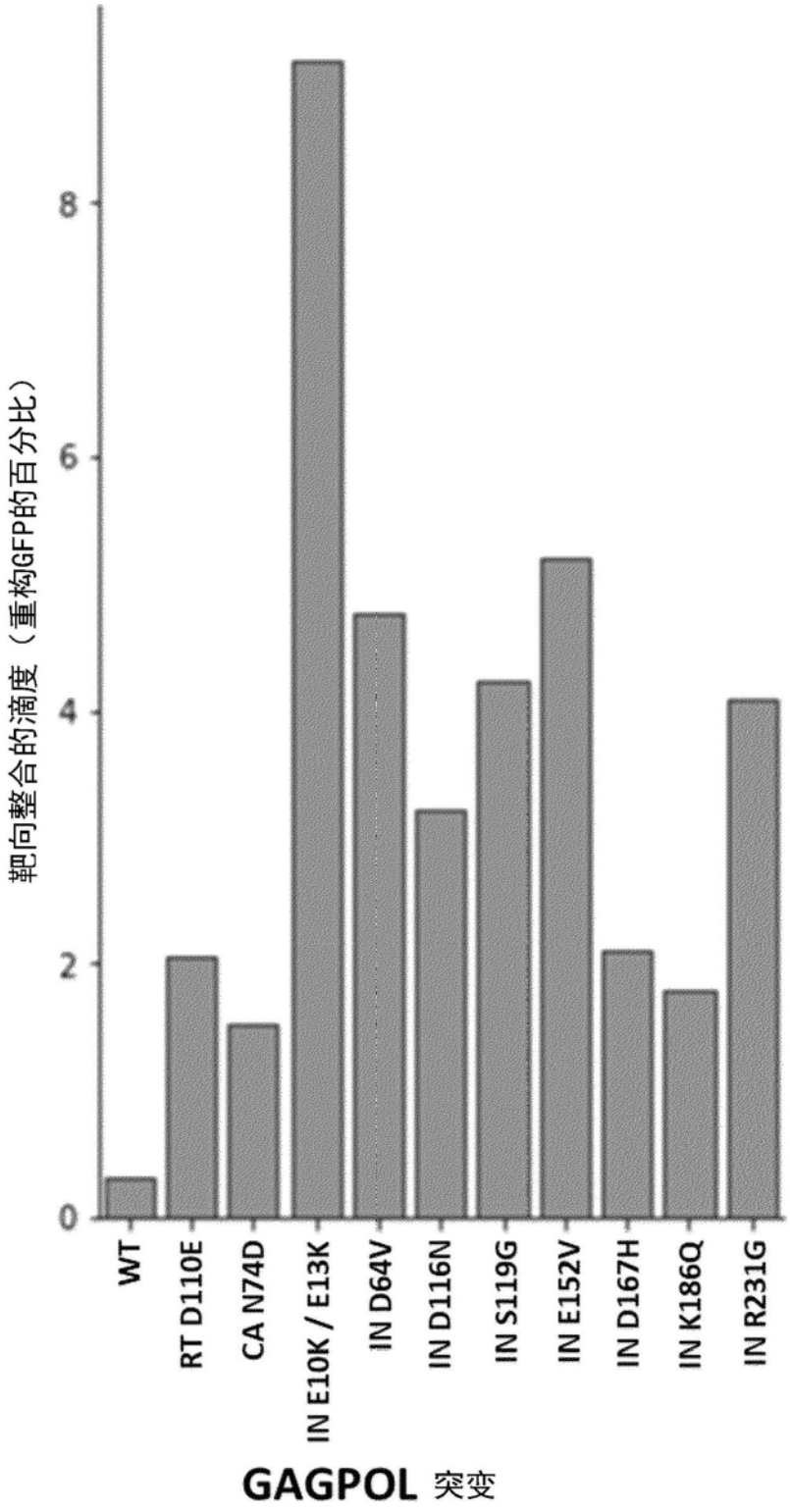

靶向整合的滴度（重构GFP的百分比）
0
2
4
6
8
WT
RT D110E
CA N74D
IN E10K / E13K
IN D64V
IN D116N
IN S119G
IN E152V
IN D167H
IN K186Q
IN R231G
GAGPOL 突变

图1D

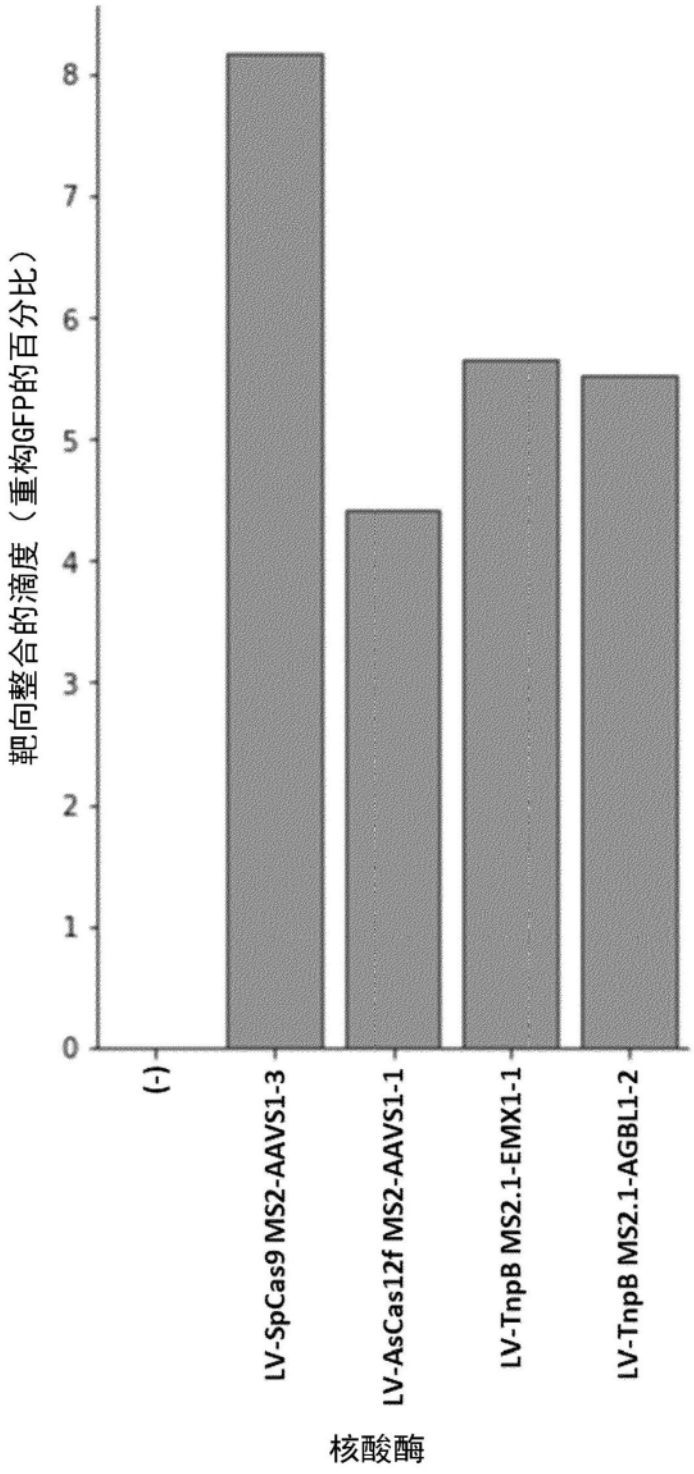

靶向整合的滴度（重构GFP的百分比）
0
1
2
3
4
5
6
7
8
(-)
LV-SpCas9 MS2-AAVS1-3
LV-AsCas12f MS2-AAVS1-1
LV-TnpB MS2.1-EMX1-1
LV-TnpB MS2.1-AGBL1-2
核酸酶

图2

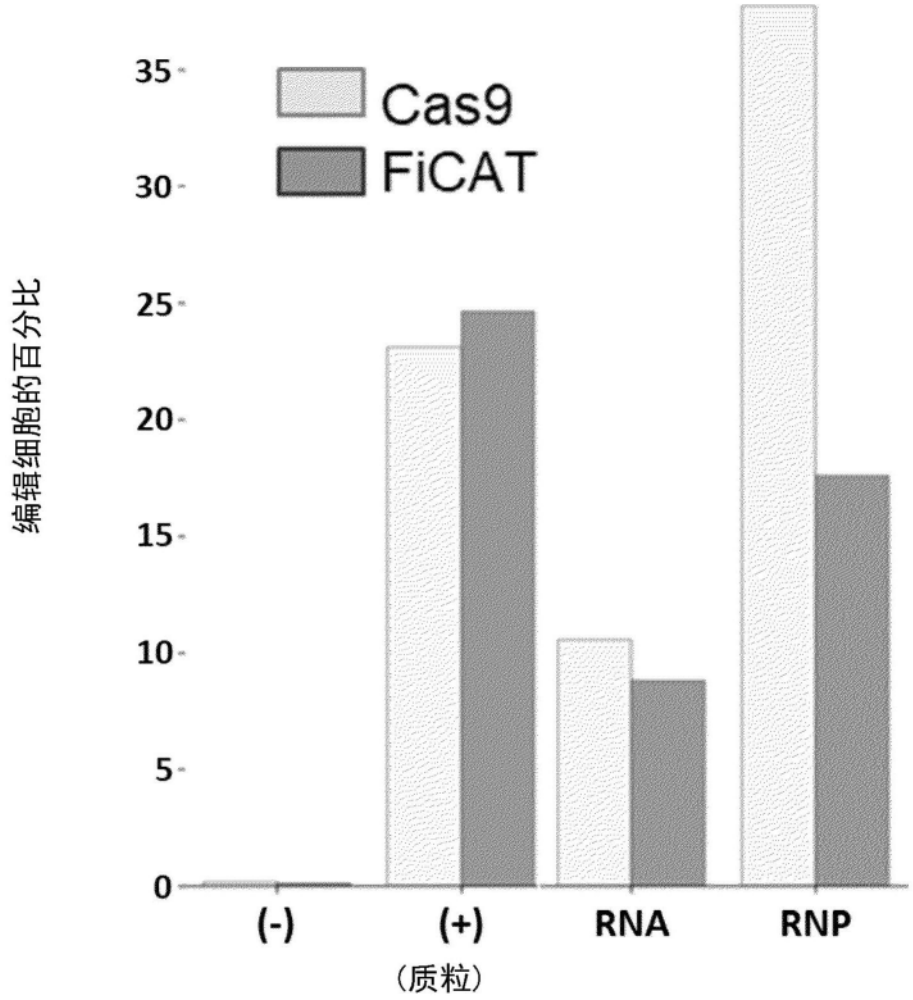


图3

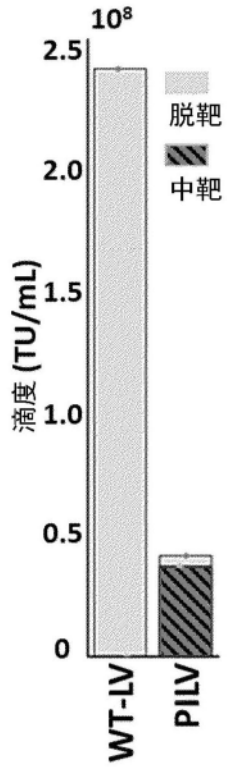

10⁸
2.5
2.0
1.5
1.0
0.5
0
滴度 (TU/mL)
脱靶
中靶
WT-LV
PILV

图4A

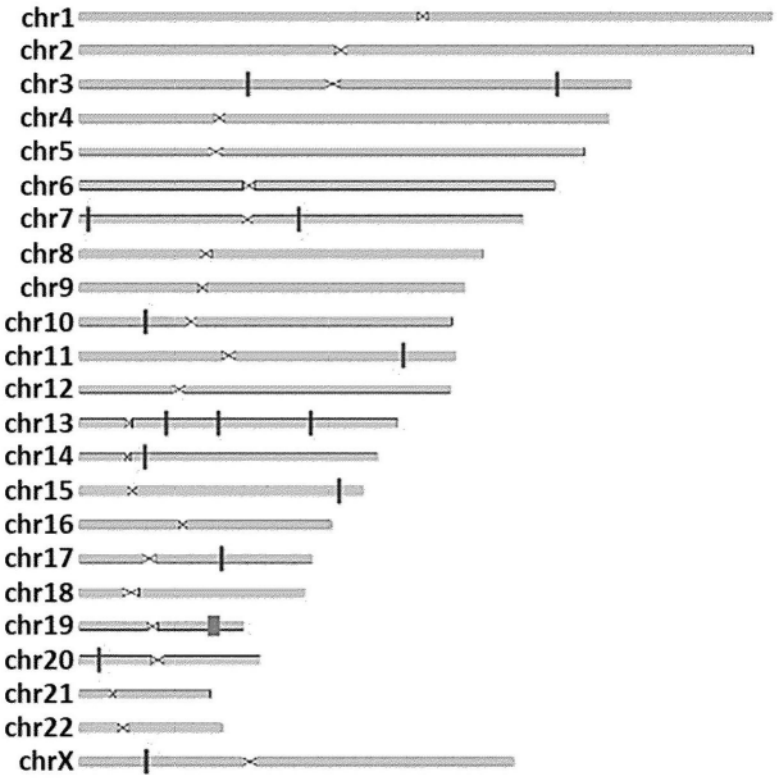

chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX

图4B

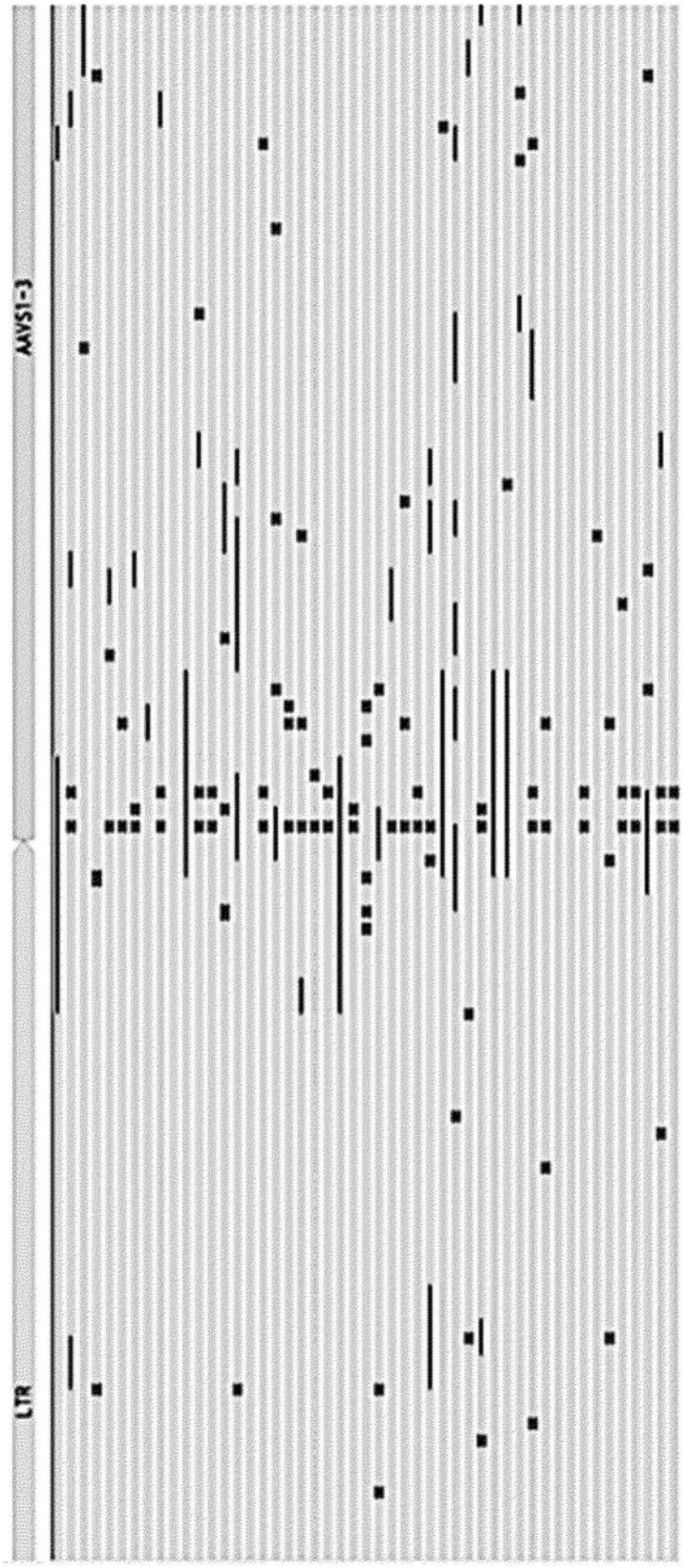


图4C

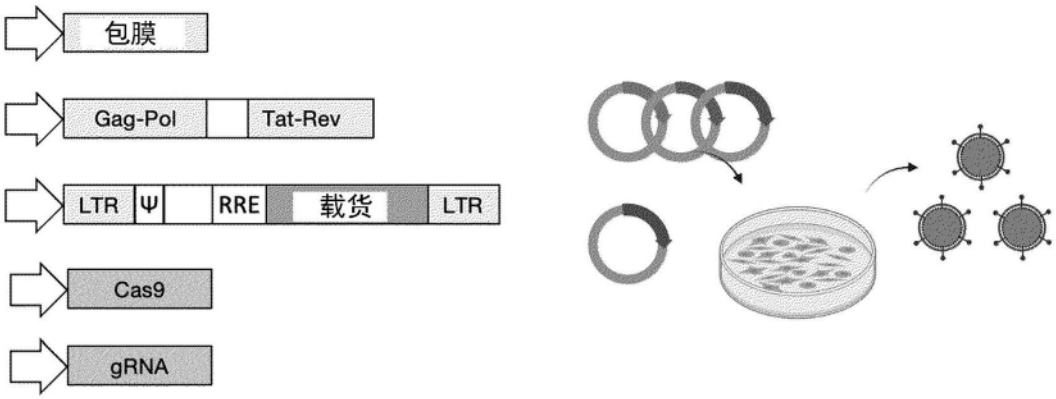

包膜
Gag-Pol
Tat-Rev
LTR
Ψ
RRE
载货
LTR
Cas9
gRNA

图5A

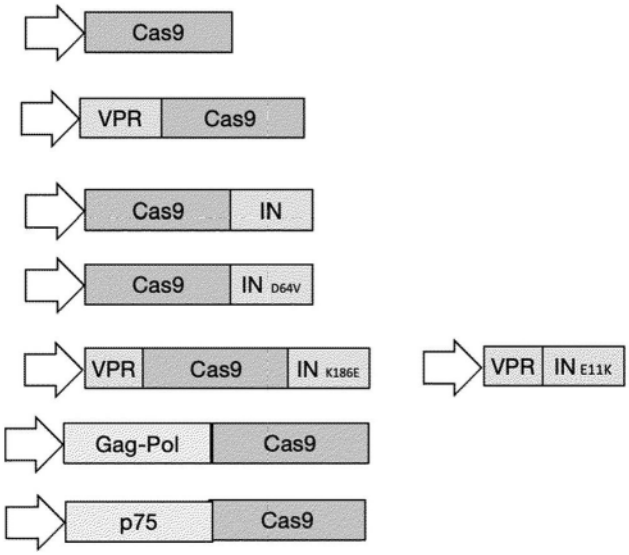

Cas9
VPR
Cas9
Cas9
IN
Cas9
IN D64V
VPR
Cas9
IN K186E
VPR
IN E11K
Gag-Pol
Cas9
p75
Cas9

图5B

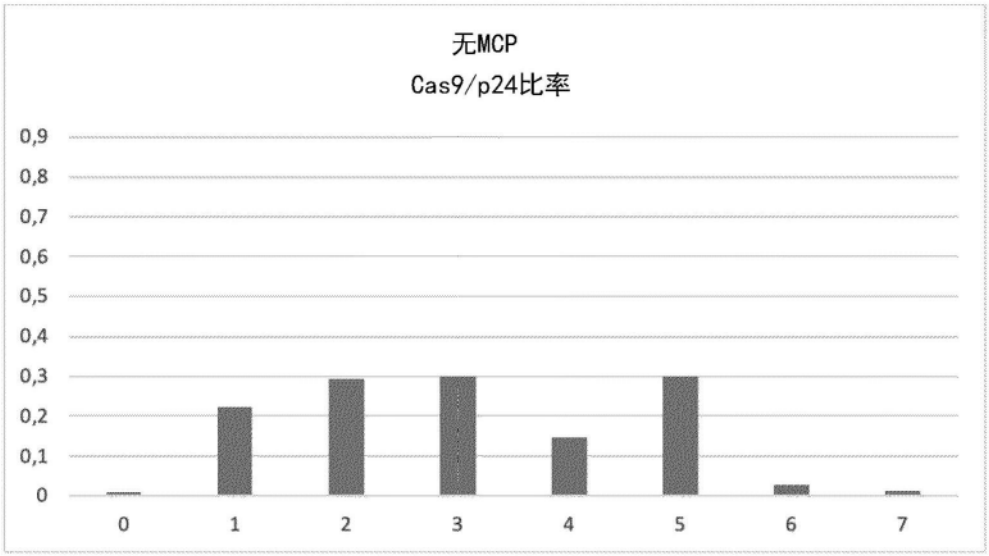

无MCP
Cas9/p24比率
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
1
2
3
4
5
6
7

图6A

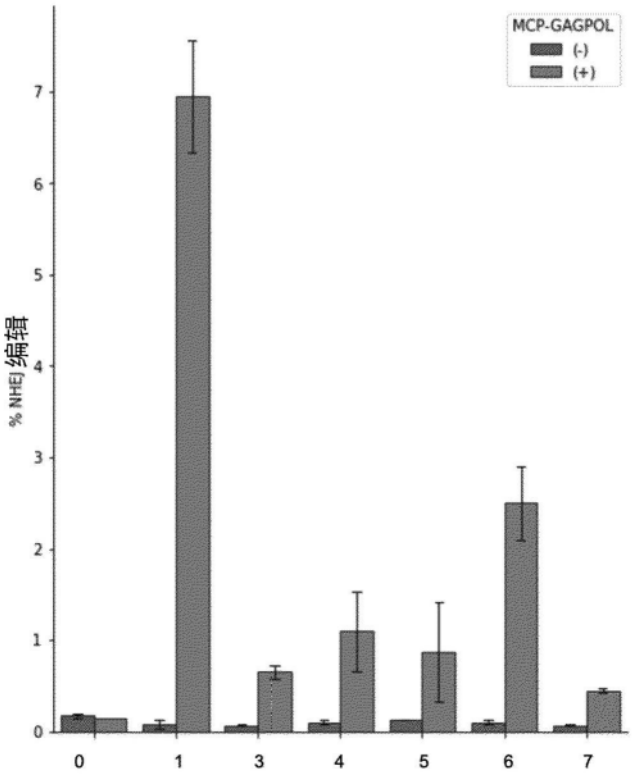

MCP-GAGPOL
(-)
(+)
% NHEJ编辑
7
6
5
4
3
2
1
0
0
1
3
4
5
6
7

图6B

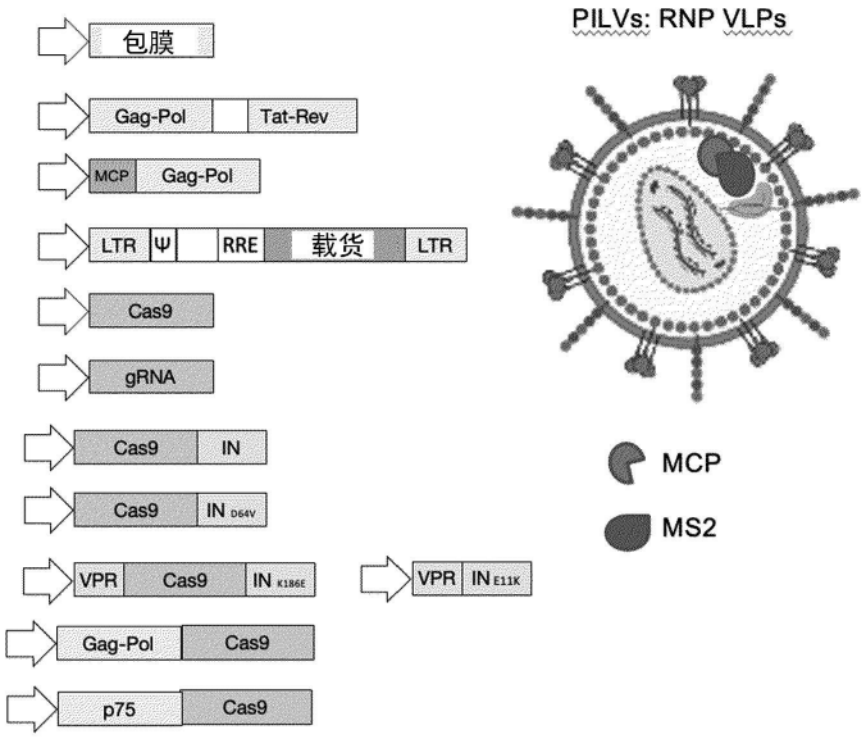

包膜
Gag-Pol
Tat-Rev
MCP
Gag-Pol
LTR
Ψ
RRE
载货
LTR
Cas9
gRNA
Cas9
IN
Cas9
IN D64V
VPR
Cas9
IN K186E
VPR
IN E11K
Gag-Pol
Cas9
p75
Cas9
PILVs: RNP VLPs
MCP
MS2

图7A

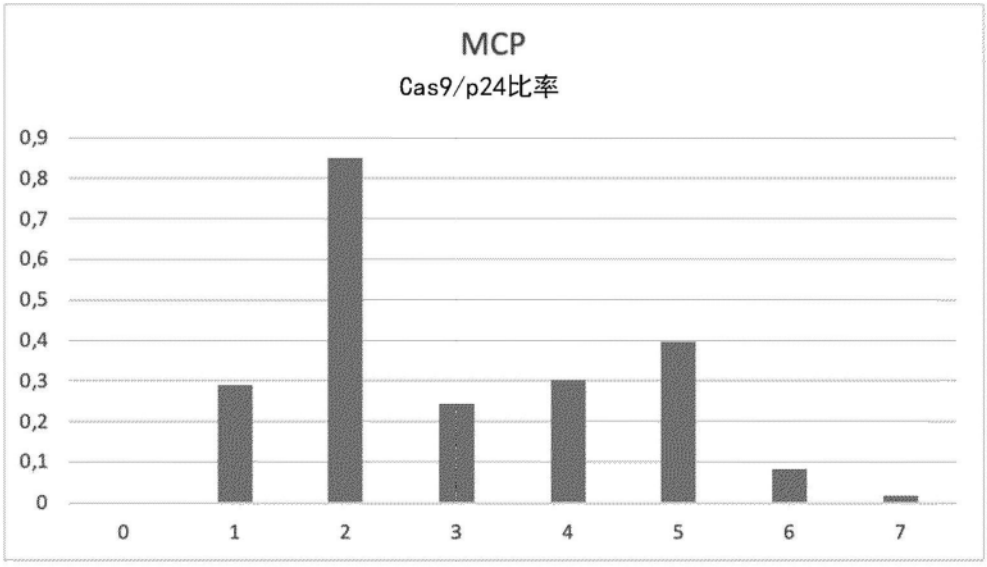

MCP
Cas9/p24比率
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
1
2
3
4
5
6
7

图7B

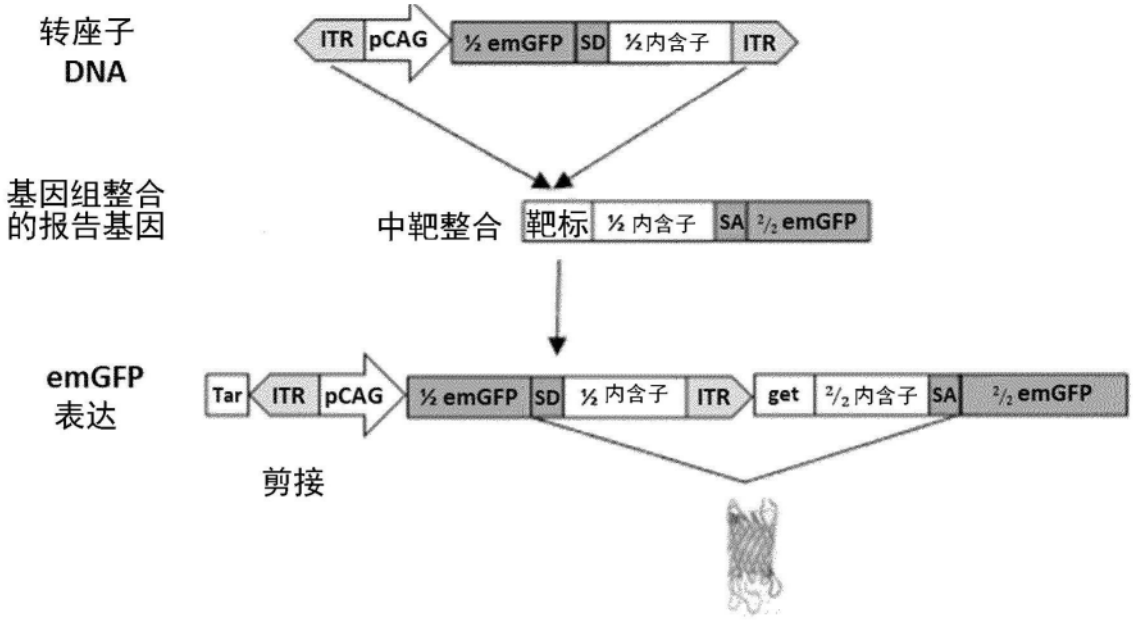

转座子
DNA
ITR
pCAG
½ emGFP
SD
½内含子
ITR
基因组整合
的报告基因
中靶整合
靶标
½ 内含子
SA
²/₂ emGFP
emGFP
表达
Tar
ITR
pCAG
½ emGFP
SD
½ 内含子
ITR
get
²/₂ 内含子
SA
²/₂ emGFP
剪接

图8A

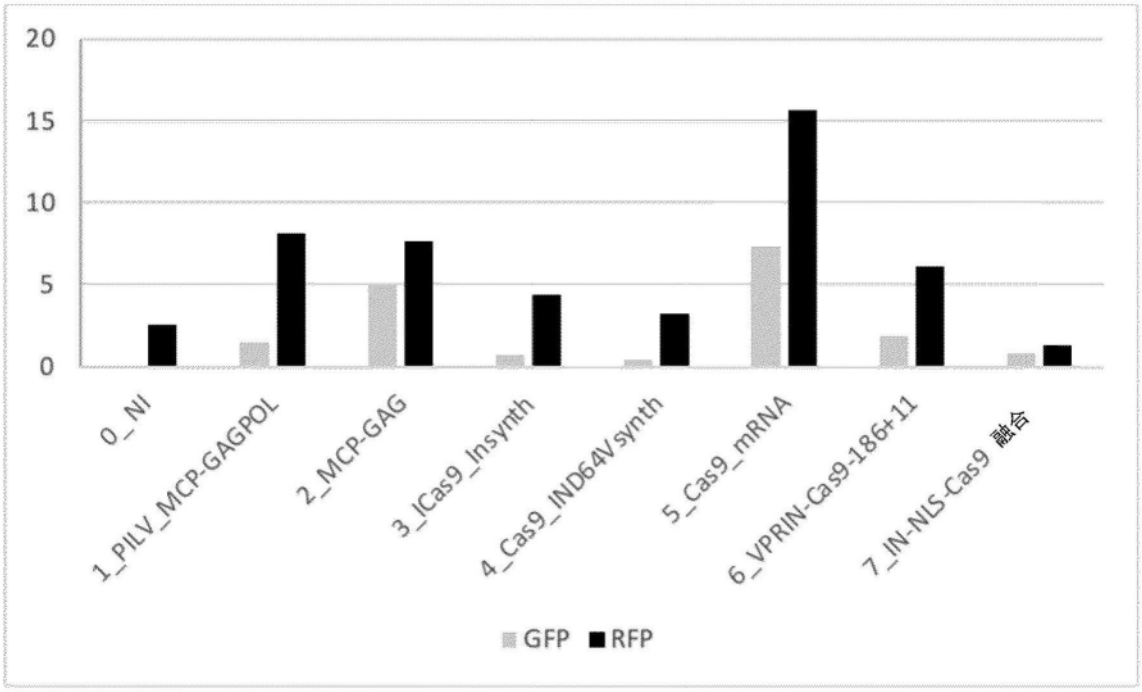

20
15
10
5
0
0_NI
1_PILV_MCP-GAGPOL
2_MCP-GAG
3_ICas9_Insynth
4_Cas9_IND64Vsynth
5_Cas9_mRNA
6_VPRIN-Cas9-186+11
7_IN-NLS-Cas9 融合
GFP
RFP

图8B

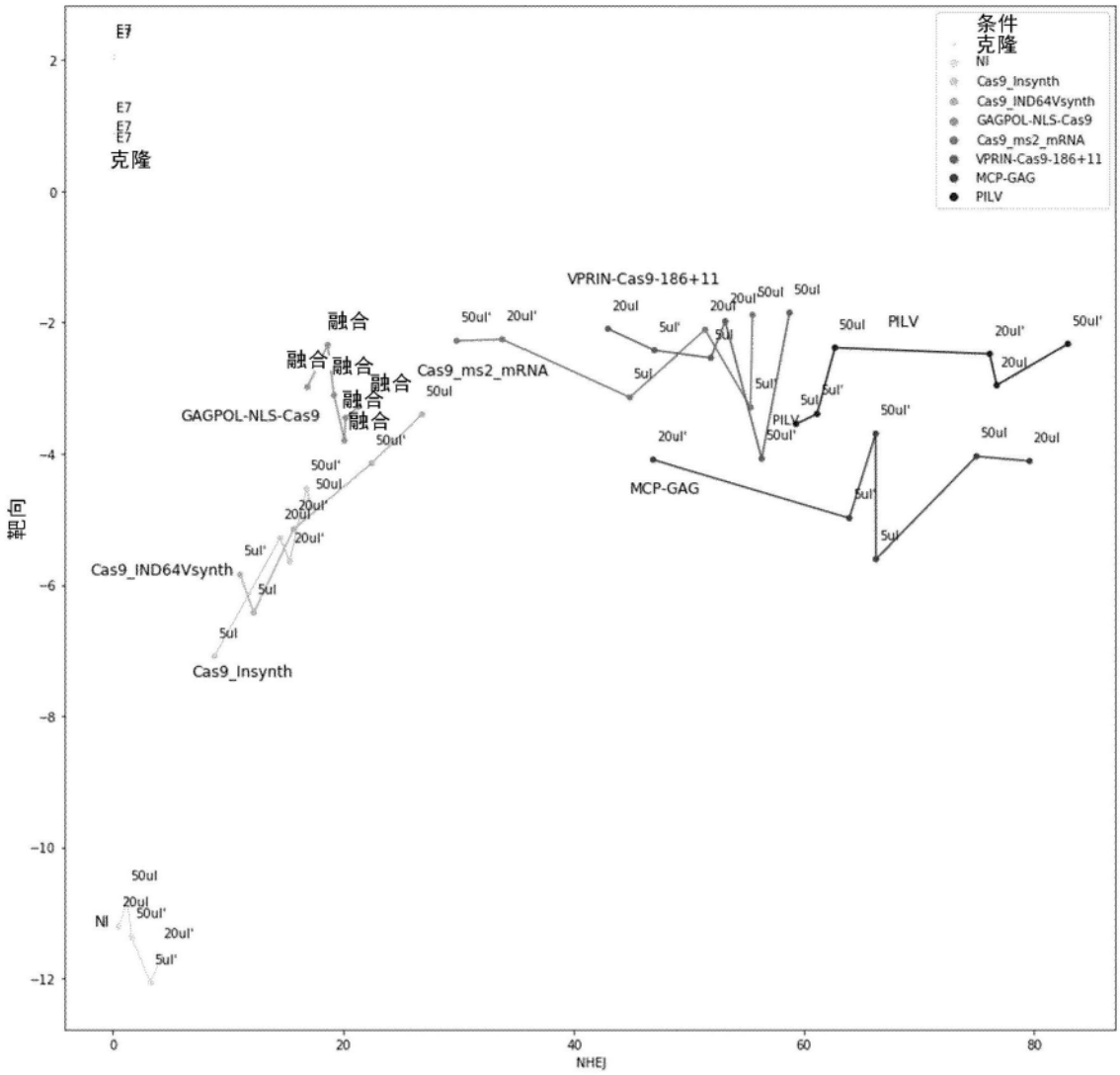

条件
克隆
NI
Cas9_Insynth
Cas9_IND64Vsynth
GAGPOL-NLS-Cas9
Cas9_ms2_mRNA
VPRIN-Cas9-186+11
MCP-GAG
PILV
克隆
融合
Cas9_ms2_mRNA
GAGPOL-NLS-Cas9
VPRIN-Cas9-186+11
MCP-GAG
Cas9_IND64Vsynth
Cas9_Insynth
PILV
NI
向性
NHEJ

图9

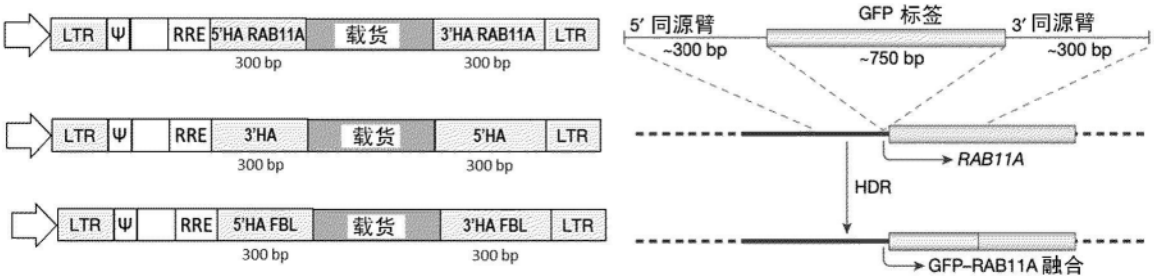

LTR
Ψ
RRE
5'HA RAB11A
载货
3'HA RAB11A
300 bp
3'HA
5'HA
5'HA FBL
3'HA FBL
5′ 同源臂
GFP 标签
3′ 同源臂
~300 bp
~750 bp
RAB11A
HDR
GFP–RAB11A 融合

图10A

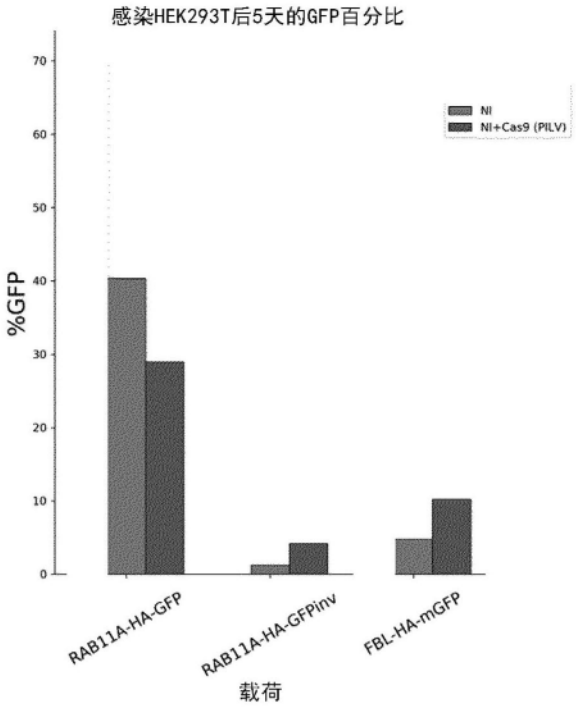

感染HEK293T后5天的GFP百分比
NI
NI+Cas9 (PILV)
%GFP
0
10
20
30
40
50
60
70
RAB11A-HA-GFP
RAB11A-HA-GFPinv
FBL-HA-mGFP
载荷

图10B

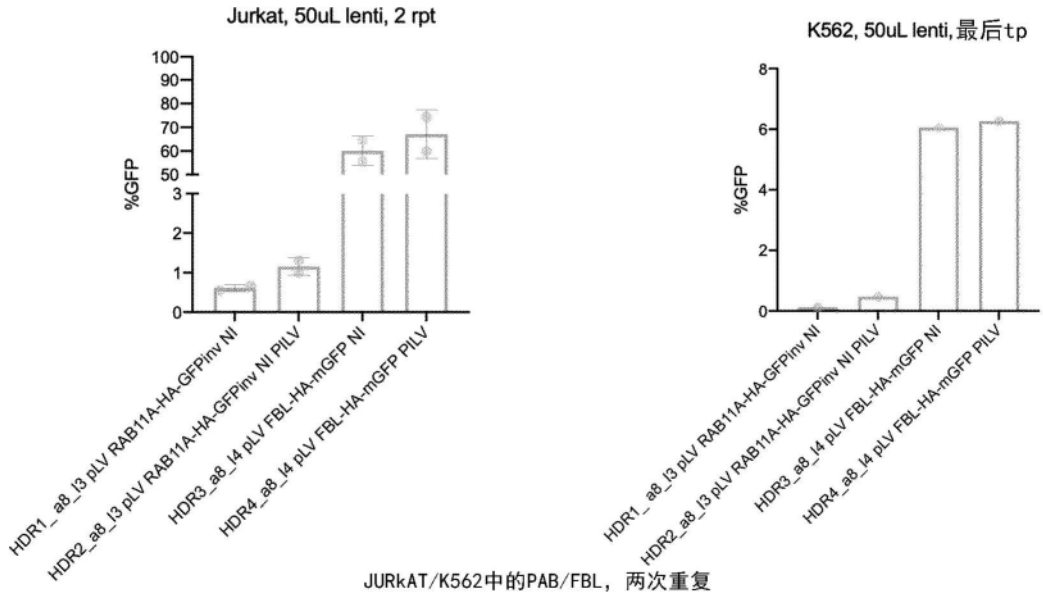

Jurkat, 50uL lenti, 2 rpt
%GFP
HDR1_a8_I3 pLV RAB11A-HA-GFPinv NI
HDR2_a8_I3 pLV RAB11A-HA-GFPinv NI PILV
HDR3_a8_I4 pLV FBL-HA-mGFP NI
HDR4_a8_I4 pLV FBL-HA-mGFP PILV
K562, 50uL lenti,最后tp
%GFP
HDR1_a8_I3 pLV RAB11A-HA-GFPinv NI
HDR2_a8_I3 pLV RAB11A-HA-GFPinv NI PILV
HDR3_a8_I4 pLV FBL-HA-mGFP NI
HDR4_a8_I4 pLV FBL-HA-mGFP PILV
JURkAT/K562中的PAB/FBL，两次重复

图10C

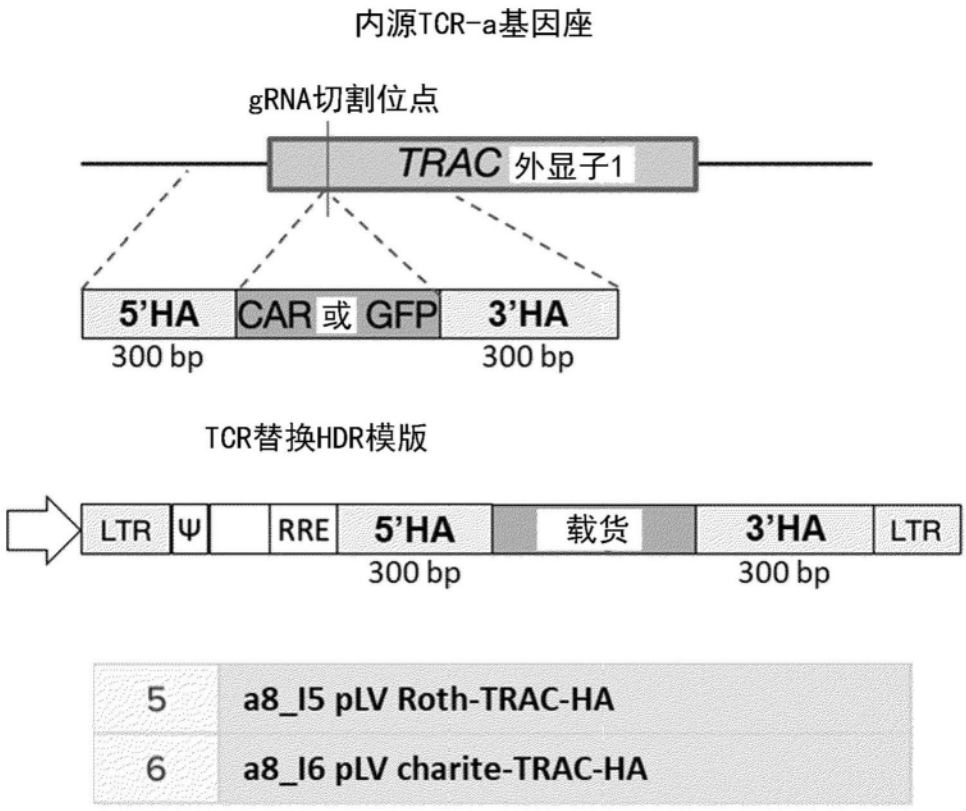

内源TCR-a基因座
gRNA切割位点
TRAC 外显子1
5'HA
CAR 或 GFP
3'HA
300 bp
300 bp
TCR替换HDR模版
LTR
Ψ
RRE
5'HA
载货
3'HA
LTR
300 bp
300 bp
5
a8_I5 pLV Roth-TRAC-HA
6
a8_I6 pLV charite-TRAC-HA

图11A

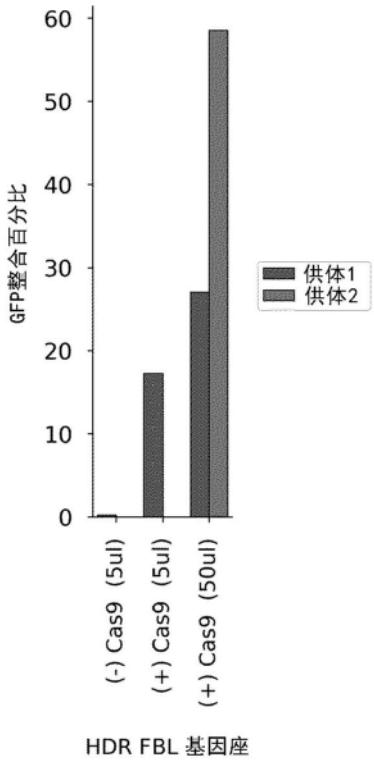

60
50
40
30
20
10
0
GFP整合百分比
供体1
供体2
(-) Cas9 (5ul)
(+) Cas9 (5ul)
(+) Cas9 (50ul)
HDR FBL 基因座

图11B