



(12) 发明专利申请



(10) 申请公布号 CN 119487184 A

(43) 申请公布日 2025.02.18

(21) 申请号 202380039707.0

(22) 申请日 2023.05.12

(30) 优先权数据

22191966.5 2022.08.24 EP

63/341,619 2022.05.13 US

(85) PCT国际申请进入国家阶段日

2024.11.11

(86) PCT国际申请的申请数据

PCT/EP2023/062736 2023.05.12

(87) PCT国际申请的公布数据

W02023/218021 EN 2023.11.16

(71) 申请人 英特格拉治疗公司

地址 西班牙巴塞罗那

申请人 庞培法布拉大学

(72) 发明人 A·桑切斯-梅希亚斯加西亚

M·盖尔卡戈尔

M·帕拉斯马西米亚

(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.

G12N 9/22 (2006.01)

G12N 15/62 (2006.01)

G12N 15/85 (2006.01)

权利要求书1页 说明书44页

序列表(电子公布) 附图9页

(54) 发明名称

转座酶用于改善转基因表达和核定位的用途

(57) 摘要

本发明涉及利用转座酶增加细胞群中目的转基因的表达和/或增加细胞群中目的转基因的核定位的方法。

1. 一种在细胞群中增加编码至少一种目的转基因的核酸分子的核定位的体外方法,包括使细胞群与以下物质接触:

- 包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,以及
- 编码所述至少一种目的转基因的核酸分子。

2. 根据权利要求1所述的体外方法,其中编码所述至少一种目的转基因的核酸分子还包含至少一个末端反向重复(ITR)序列。

3. 根据权利要求1或2所述的体外方法,其中包含转座酶或其片段的蛋白或多肽是包含所述转座酶或其片段和至少一种额外的多肽或蛋白的融合蛋白。

4. 根据权利要求3所述的体外方法,其中所述至少一种额外的多肽或蛋白是核酸酶。

5. 根据权利要求3或4所述的体外方法,其中所述融合蛋白与SEQ ID NO:2具有至少75%的氨基酸序列同一性。

6. 根据权利要求1至5中任一项所述的体外方法,进一步包括使细胞群与指导RNA(gRNA)接触。

7. 根据权利要求3至6中任一项所述的体外方法,其中所述至少一种额外的多肽或蛋白是适体结合蛋白。

8. 根据权利要求7所述的体外方法,其中所述适体结合蛋白是优选具有SEQ ID NO:39的MS2噬菌体包被蛋白(MCP)。

9. 根据权利要求8所述的体外方法,其中所述融合蛋白通过MCP与gRNA分子共价或非共价相互作用,或能够通过MCP与gRNA共价或非共价相互作用,其中所述gRNA包含至少一个MS2适体,优选地,所述MS2适体具有SEQ ID NO:40的核糖核酸序列。

10. 根据权利要求1至9中任一项所述的方法,其中所述转座酶选自由高活性PiggyBac转座酶、PiggyBac转座酶、睡美人转座酶、SB11转座酶、Tol2转座酶、Mos1转座酶和青蛙王子转座酶组成的组。

11. 根据权利要求1至10中任一项所述的方法,其中所述转座酶为修饰的高活性PiggyBac转座酶,其与具有SEQ ID NO:1的高活性PiggyBac转座酶的氨基酸序列相比包含至少一个氨基酸突变。

12. 根据权利要求1至11中任一项所述的方法,其中所述转座酶或其片段是催化失活的。

13. 根据权利要求12所述的方法,其中所述催化失活的转座酶与SEQ ID NO:3具有至少75%的氨基酸序列同一性。

14. 一种包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,其用于治疗有需要的受试者的遗传疾病,

其中向有需要的受试者进一步施用编码至少一种目的转基因的核酸分子,

其中目的转基因在有需要的受试者的至少一个细胞中的表达补偿了导致所述遗传疾病的基因缺陷。

15. 根据权利要求14所述用途的包含转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸,所述用途还包括根据权利要求2至13中任一项或几项的特征。

转座酶用于改善转基因表达和核定位的用途

发明领域

[0001] 本发明涉及改善细胞群中转基因表达的方法。

[0002] 发明背景

[0003] 具有将核酸或蛋白引入细胞和组织的有效方法是基因组工程的关键。为此目的，已经开发了不同的方法，例如病毒感染、电穿孔、聚合物和纳米颗粒转染。虽然这些方法可以极大提高转染效率，但将目的DNA引入细胞核仍然是一个巨大的挑战。

[0004] 确实，外来DNA进入细胞后，在到达细胞核之前将面临各种障碍。首先，哺乳动物细胞具有旨在感知和调控外来DNA的防御机制，例如干扰素系统。Semenova等人 (Nucleic acids research 2019, vol. 47, 19) 报道称，细胞溶质DNA传感器在转染后的数小时内与转染的DNA结合，从而导致促炎反应。其次，外来DNA分子必须穿过核膜的屏障。Vandenbroucke等人 (Nucleic acids research 2007, vol. 35, 12) 提出利用两亲性分子来调节核孔复合物 (NPC) 的活性，以增加核质转运；然而，两亲性分子与一定程度的毒性有关，并且它们似乎无法增加某些DNA载体的核摄取。

[0005] 外来DNA易位到细胞核的其他额外的障碍可能包括运输、细胞类型特异性和细胞分裂 (详情请参阅Bai等人. "Cytoplasmic transport and nuclear import of plasmid DNA." Bioscience reports, 2017, vol. 37, 6)。

[0006] 因此，外来DNA难以到达真核细胞 (体外和体内) 的细胞核是许多转染方法试图克服的问题。本发明提供一种解决方案，包括用ITR包围目的基因，并向转染混合物中添加转座酶，从而帮助将外来DNA定位到细胞核中。

[0007] 转座酶是自然进化的酶，可以将DNA片段“剪切粘贴”或“复制粘贴”到原核生物 (例如Tn3、Tn5和Tn7) 或真核生物 (例如睡美人和PiggyBac) 的基因组DNA中。在某些物种的基因组中发现这些酶具有较低的转座效率，以避免基因组毒性。转座酶与含有特定末端反向重复 (ITR) (称为转座子) 的有效载荷DNA结合，复制或剪切其序列并将其粘贴到随机或半随机的基因组位点中。与大多数DNA结合蛋白一样，转座酶由于核定位信号而自然定位到细胞核中 (例如参见Keith等人. "Analysis of the piggyBac transposase reveals a functional nuclear targeting signal in the 94c-terminal residues." BMC Molecular Biology 2008, vol. 9, 72)。

[0008] 最先进的技术包括不同的转座酶，其被改造以提高其效率并创建用于哺乳动物基因组编辑的工具，例如SB100或高活性PiggyBac。转座酶也已与DNA结合结构域融合，例如Cas9 (参见例如W02022129438或W02020243085)、死亡-Cas9 (参见例如Hew等人. "RNA-guided piggyBac transposition in human cells." Synthetic Biology 2019, vol. 4, 1), 以及ZNF, 以便特异性地靶向和编辑某个基因组DNA位置。

[0009] 然而，虽然转座酶已被描述为基因组编辑的工具 (例如参见Zhao等人. "PiggyBac transposon vectors: the tools of the human gene encoding." Translational Lung Cancer Research 2016, vol. 5, 1), 它们的应用依赖于它们的插入活性，要么将基因插入宿主细胞的DNA中，要么破坏基因，这也带来了插入诱变的问题。

[0010] 此外,转染通常在活跃分裂(增殖)细胞上进行,因为它们比非分裂静止细胞更容易内化核酸(Bai等人.2017),并且转染汇合细胞或某些在培养中不分裂的细胞类型尤其困难。因此,需要改善非分裂细胞的转染效率。

[0011] 因此,本发明提供了证据,表明使用转座酶和促进核酸细胞质转导的载体,可以增加目的DNA的核定位或核摄取,然后它可以在那里被细胞表达(即,瞬时表达、转染)或插入基因组DNA中。本发明的方法还能够有效转染静止细胞。最后,本发明的方法还能够在使用催化无活性的转座酶时实现无插入活性的有效转染。

[0012] 发明概述

[0013] 本发明的一个目的是增加细胞群中编码至少一种目的转基因的核酸分子的表达的方法,包括使细胞群与下列物质接触:

[0014] -包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,以及

[0015] -编码至少一种目的转基因的核酸分子。

[0016] 本发明的另一个目的是一种在细胞群中增加编码至少一种目的转基因的核酸分子的核定位的方法,包括使细胞群与下列物质接触:

[0017] -包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,以及

[0018] -编码至少一种目的转基因的核酸分子。

[0019] 本发明的另一个目的是一种编辑细胞群基因组的方法,包括使细胞群与下列物质接触:

[0020] -包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,以及

[0021] -编码至少一种目的转基因的核酸分子。

[0022] 本发明的另一个目的是一种治疗有需要的受试者的遗传病的方法,包括向所述受试者施用治疗有效量的:

[0023] -包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,以及

[0024] -编码至少一种目的转基因的核酸分子,

[0025] 其中目的转基因在有需要的受试者的至少一种细胞中的表达补偿了导致所述遗传病的基因缺陷。

[0026] 在一些实施方案中,编码至少一种目的转基因的核酸分子是脱氧核糖核酸(DNA)分子。在一些实施方案中,所述DNA分子是互补DNA(cDNA)分子。在一些实施方案中,所述DNA分子是基因组DNA(gDNA)分子。

[0027] 在一些实施方案中,编码至少一种目的转基因的核酸分子还包含至少一个末端反向重复(ITR)序列。在一些实施方案中,编码至少一种目的转基因的核酸分子还包含两个ITR序列。在一些实施方案中,至少一个ITR序列与目的转基因相邻。在一些实施方案中,编码至少一种目的转基因的核酸分子还包含两个位于目的转基因侧翼的ITR序列。在一些实施方案中,至少一个ITR序列与转座酶相互作用或非共价结合。

[0028] 在一些实施方案中,DNA分子被插入到所述细胞群的基因组中。在一些实施方案中,编码至少一种目的转基因的核酸分子被细胞群稳定表达。

[0029] 在一些实施方案中,目的转基因是外源基因。在一些实施方案中,目的转基因是内源基因。

[0030] 在一些实施方案中,编码至少一种目的转基因的核酸分子包含在质粒、F黏粒、粘

粒、人工染色体或病毒载体中。在一些实施方案中,编码至少一种目的转基因的核酸分子包含在质粒中。在一些实施方案中,编码至少一种目的转基因的核酸分子包含在病毒载体中。在一些实施方案中,编码至少一种目的转基因的核酸分子包含在基于DNA病毒的载体中,所述载体选自由双链DNA病毒域(Duplodnaviria)、单链DNA病毒域(Monodnaviria)和多链DNA病毒域(Varidnaviria)的病毒组成的组。

[0031] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽增加编码至少一种目的转基因的核酸分子的核定位。在一些实施方案中,包含转座酶或其片段的蛋白或多肽将编码至少一种目的转基因的核酸分子易位至细胞核,或促进该核酸分子易位至细胞核。

[0032] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽是包含转座酶或其片段和至少一种额外的多肽或蛋白的融合蛋白。在一些实施方案中,至少一种额外的多肽或蛋白是核酸酶。在一些实施方案中,至少一种额外的多肽或蛋白是RNA引导的核酸酶。在一些实施方案中,至少一种额外的多肽或蛋白是Cas核酸酶。在一些实施方案中,至少一种额外的多肽或蛋白是Cas9核酸酶。

[0033] 在一些实施方案中,融合蛋白还包含接头。

[0034] 在一些实施方案中,融合蛋白与SEQ ID NO:2具有至少75%的氨基酸序列同一性。在一些实施方案中,融合蛋白包含SEQ ID NO:2的氨基酸序列或由SEQ ID NO:2的氨基酸序列组成。

[0035] 在一些实施方案中,至少一种额外的多肽或蛋白是适体结合蛋白。在一些实施方案中,至少一种额外的多肽或蛋白是MS2噬菌体包被蛋白(MCP)。在一些实施方案中,融合蛋白通过MCP与包含至少一个MS2适体的gRNA分子共价或非共价相互作用,或能够共价或非共价相互作用。在一些实施方案中,gRNA分子通过至少一个MS2适体与RNA引导的核酸酶共价或非共价相互作用,或能够共价或非共价相互作用。

[0036] 在一些实施方案中,转座酶选自由高活性PiggyBac转座酶、PiggyBac转座酶、睡美人转座酶、SB11转座酶、Tol2转座酶、Mos1转座酶和青蛙王子转座酶组成的组。在一些实施方案中,转座酶选自由高活性PiggyBac转座酶和睡美人转座酶组成的组。在一些实施方案中,转座酶是高活性的PiggyBac转座酶。在一些实施方案中,转座酶是具有SEQ ID NO:1的氨基酸序列的高活性PiggyBac转座酶。在一些实施方案中,转座酶是修饰的高活性PiggyBac转座酶,其与SEQ ID NO:1的高活性PiggyBac转座酶的氨基酸序列相比包含至少一个氨基酸突变。在一些实施方案中,转座酶是睡美人转座酶。在一些实施方案中,转座酶或其片段具有降低的催化活性。在一些实施方案中,转座酶或其片段是催化失活的。在一些实施方案中,催化失活的转座酶与SEQ ID NO:3具有至少75%的氨基酸序列同一性。在一些实施方案中,催化失活的转座酶包含SEQ ID NO:3的氨基酸序列或由SEQ ID NO:3的氨基酸序列组成。

[0037] 在一些实施方案中,蛋白或多肽包含转座酶片段。在一些实施方案中,转座酶片段包含至少一个转座酶功能结构域或由至少一个转座酶功能结构域组成。在一些实施方案中,转座酶片段包含ITR结合结构域或由ITR结合结构域组成。在一些实施方案中,转座酶片段是高活性PiggyBac转座酶或睡美人转座酶的片段。在一些实施方案中,转座酶片段是高活性PiggyBac转座酶的片段,优选地包含ITR结合结构域。在一些实施方案中,转座酶片段是睡美人转座酶的片段,优选地选自由睡美人转座酶的SB100结构域和N57结构域组成的

组。

[0038] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是核糖核酸(RNA)分子或DNA分子。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是RNA分子。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是信使RNA(mRNA)分子。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是cDNA分子或质粒。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸通过转染或转化递送至细胞群。在一些实施方案中,转染选自脂质转染、电穿孔、超声处理、纳米颗粒、显微注射和病毒载体感染,包括非整合型和整合型病毒载体感染。在一些实施方案中,转化包括使用整合型病毒载体或改良的整合型病毒。

[0039] 在一些实施方案中,将编码包含转座酶或其片段的蛋白或多肽的核酸与编码至少一种目的转基因的核酸分子一起递送至细胞群。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前递送至细胞群。在一些实施方案中,将编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前1小时至72小时递送至细胞群。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前约4小时递送至细胞群。

[0040] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸包含在药物组合物中,所述药物组合物还包含至少一种可接受的赋形剂。

[0041] 在一些实施方案中,细胞群是真核细胞群或原核细胞群。在一些实施方案中,细胞群是真核细胞群。在一些实施方案中,细胞群是原核细胞群。在一些实施方案中,该方法是在体外进行的。在一些实施方案中,细胞群是体外培养的细胞群。在一些实施方案中,该方法是离体进行的。在一些实施方案中,该方法是在体内进行的。

[0042] 在一些实施方案中,细胞群包含在活生物体的组织或器官中。在一些实施方案中,生物体是动物。在一些实施方案中,生物体是哺乳动物。在一些实施方案中,生物体是人。

[0043] 在一些实施方案中,该方法包括以下步骤:

[0044] -使细胞群与包含编码包含转座酶或其片段的蛋白或多肽的核酸的载体接触,

[0045] -培养细胞群1小时至72小时的时间段,以及

[0046] -使细胞群与包含编码至少一种目的转基因的核酸分子的载体接触。

[0047] 在一些实施方案中,该方法包括以下步骤:

[0048] -使细胞群与包含含有转座酶或其片段的蛋白或多肽的载体接触,以及

[0049] -使细胞群与包含编码至少一种目的转基因的核酸分子的载体接触。

[0050] 定义

[0051] 在本发明中,下列术语具有如下含义:

[0052] “Cas9”或“Cas9核酸酶”是指包含Cas9蛋白或其片段的RNA引导的核酸酶(例如,包含Cas9的活性或非活性DNA切割结合结构域和/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(成簇的规律间隔的短回文重复序列)相关核酸酶。CRISPR是一种适应性免疫系统,可提供针对移动遗传元件(病毒、转座元件和接合质粒)的保护。CRISPR簇包含间隔区(与先前的移动元件互补的序列),并靶向入侵的核酸。CRISPR簇被转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,正确处理pre-crRNA需要反式

编码的小RNA (tracrRNA)、内源性核糖核酸酶3 (rnc) 和Cas9蛋白。tracrRNA可用作用于核糖核酸酶3辅助加工前crRNA的指导序列。随后,Cas9/crRNA/tracrRNA复合物以核酸内切法切割与间隔区互补的线性或环状dsDNA靶标。首先通过核酸内切法切割不与crRNA互补的靶链,然后通过核酸外切法修剪3'-5'。在自然界中,DNA结合和裂解通常需要蛋白和两种RNA。然而,单指导RNA (“sgRNA”或简称“gRNA”) 可以进行工程化,从而将crRNA和tracrRNA这两者的各方面整合到单个RNA类型中。Cas9识别CRISPR重复序列中的短基序 (PAM或原间隔区相邻基序) 以帮助区分自身与非自身。Cas9核酸酶序列和结构是本领域技术人员所熟知的。Cas9直系同源物已在多种物种中得到描述,包括但不限于化脓性链球菌 (*S. pyogenes*) 和嗜热链球菌 (*S. thermophilus*)。

[0053] “外源”是指在目的细胞或生物体中不是天然存在但可以通过一种或多种遗传、生物化学或其他方法引入其中的任何分子。分子在细胞或生物体中的天然存在也可以根据其特定的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,对于非热休克细胞而言,热休克诱导的分子是外源分子。外源分子可以包括例如功能失调的内源分子的功能性版本或功能正常的内源分子的功能失调版本。相比之下,术语“内源”指的是在特定环境条件下的特定发育阶段,细胞或生物体中通常存在的任何分子。

[0054] “融合蛋白”是指包含两个或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) 的单链杂交多肽。两个或更多个氨基酸序列可以经由直接肽键或间接通过肽接头融合在一起。融合蛋白尤其可以完全由单个核酸序列编码。

[0055] “基因”通常是指编码蛋白的DNA区域 (即编码区)。该术语还可以包括本身不编码蛋白的DNA区域 (即非编码区域)。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RNA等) 的区域。其他非编码区调节编码区域的转录和翻译 (即,调控元件),或用作结构元件 (例如,支架/基质附着区)、DNA复制的起点、着丝粒或端粒等。调控元件包括但不限于启动子序列、终止子、翻译调控序列 (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。

[0056] “同一性”或“同一”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关系时,是指氨基酸序列或核酸序列之间的序列相关程度,如通过两个或更多个氨基酸残基或核酸残基串之间的匹配数目确定的。“同一性”衡量两个或更多个序列中较小序列之间相同匹配的百分比,这些序列具有通过特定数学模型或计算机程序 (即“算法”) 解决的空位比对 (如果有的话)。相关氨基酸序列或核酸序列的同一性可以很容易通过已知方法进行计算。这样的方法包括但不限于以下文献中描述的那些方法: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等人,

1988.SIAM J Appl Math.48(5):1073-82。确定同一性的优选方法被设计为使测试序列之间给出最大的匹配。确定同一性的方法描述在公众可得的计算机程序中。用于确定两个序列之间的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 (遗传学计算机课题组), University of Wisconsin, Madison, WI; Devereux等人, 1984.Nucleic Acids Res.12(1Pt 1):387-95)、BLASTP、BLASTN和FASTA (Altschul等人, 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等人.NCB/NLM/NIH Bethesda, Md.20894) 公众获得。也可以使用著名的Smith Waterman算法来确定同一性。

[0057] “插入”和“整合”是指将一个核酸序列添加到第二核酸序列或基因组或其部分中。插入可能是“特异性的”、“位点特异性的”、“靶向的”或“在靶的 (on-targeted)”：这些形容词定义了将核酸插入到第二核酸的特定位置或基因组或其部分的特定位置 (即, 专门选择插入的位点)。相反, 形容词“随机的”、“非靶向的”或“脱靶的 (off-targeted)”是指核酸非特异性和/或非预期地插入到不想要的位点。术语“总计”或“整体”是指插入的总数。

[0058] “接头”是指连接两个相邻分子或部分的化学基团或分子。

[0059] “修饰的”是指与相应未修饰的蛋白或核酸序列不同的蛋白或核酸序列。

[0060] 与序列 (例如, 氨基酸序列或核酸序列) 相关的“突变的 (Mutated)”, 是指与参考序列 (例如野生型序列) 不同的序列。通常, 与参考序列 (例如相应的野生型序列) 相比, 突变序列包括一个或几个残基的取代、添加或缺失中的至少一个。

[0061] “突变 (Mutation)”是指序列 (例如核酸或氨基酸序列) 内的残基被另一个残基取代; 和/或在核酸或氨基酸序列内一个或多个残基的缺失或插入。本文通常通过识别原始残基、随后识别残基在序列中的位置、然后识别新取代的残基来描述突变。本文提供的用于进行氨基酸取代 (突变) 的各种方法是本领域中众所周知的, 并且由例如Green&Sambrook, 2012 (Molecular cloning: a laboratory manual (4th Ed.)). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) 提供。

[0062] “核酸酶”是指催化核酸序列内核酸水解的酶。核酸酶活性可以导致单链或者双链核酸分子的断裂, 其中核酸分子可以是DNA或RNA。

[0063] “核酸 (分子/序列)”和“核苷酸序列”可以互换使用, 是指由单体核苷酸组成或者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或它们的混合物。它可以是化学修饰的或人工的; 例如, 它包括肽核酸 (PNA)、吗啉核酸 (morpholinos) 和锁核酸 (LNA), 以及甘油核酸 (GNA) 和苏糖核酸 (TNA)。这些核酸中的每一种通过分子主链的变化与天然存在的DNA或RNA进行区分。此外, 也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磷酸酯、磷酸酰胺、二硫代磷酸酯、N3'P5'磷酸酰胺和寡核糖核苷酸硫代磷酸酯以及它们的2'0-烯丙基类似物和2'0-甲基核糖核苷酸甲基磷酸酯, 其可用于本公开的核酸中。

[0064] “多肽”、“肽”、“蛋白”和“氨基酸序列”可互换使用, 是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。该术语还适用于氨基酸聚合物, 其中一个或多个氨基酸是相应的天然存在的氨基酸的化学类似物或修饰衍生物。

[0065] “预防”及其任何词尾变化是指预防性和防止性措施, 其中目的是减少受试者在给定的时间段内发展出给定病理状况或病症的机会。例如, 这种减少可以反映在受试者的病

理状况或病症的至少一种症状的延迟发作中。

[0066] “特异性”是指选择性地结合与选定序列具有一定序列同一性的序列的能力。

[0067] “受试者”是指哺乳动物,优选是人。受试者可以是“患者”,即温血动物,更优选地是人,其正在等待接受医疗护理或正在接受医疗护理或曾经/是/将成为医疗程序的对象,或被监测疾病的发展。术语“哺乳动物”在此是指任何哺乳动物,包括人、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、羊、猪、山羊、兔子等。优选地,哺乳动物是灵长类动物,更优选地是人。

[0068] “转导”及其任何词尾变化是指利用病毒载体运载体(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。

[0069] “转染”及其任何词尾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,无论是在体外还是在体内。换句话说,“转染”是指本领域已知的任何促进或增加核酸分子或货物(cargo)的细胞质转导的方法、技术或媒介物。转染方法是本领域所熟知的,包括例如脂质转染、PEI和电穿孔。

[0070] “转基因”是指外源核酸序列,特别是编码基因产物的外源DNA或cDNA。基因产物可以是RNA、肽或蛋白。除了基因产物的编码区(CDS)外,转基因还可以包括一个或多个操作序列或者与一个或多个操作序列相关,以促进或增强表达,例如启动子、增强子、反应元件、报告元件、绝缘子元件、多聚腺苷酸化信号和/或其他功能元件。除非另有说明,本公开的实施方案可以利用任何已知合适的启动子、增强子、反应元件、报告元件、绝缘子元件、多聚腺苷酸化信号和/或其他功能元件。合适的元件和序列对于本领域技术人员来说是众所周知的。

[0071] “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。在本发明的范围内,转座酶可以是转座酶的片段,例如,ITR结合结构域或功能结构域,优选ITR结合结构域。

[0072] “治疗”、“缓解”、“治愈”及其任何词尾变化是指治疗性处理,不包括预防性或防止性措施;其中目的是减缓、减轻、停止或甚至逆转(部分或全部)目标病理状况或病症的进化。需要治疗的那些包括已经患有该病症的那些以及疑似患有该病症的那些。如果受试者接受治疗后,显示出与病理状况或病症相关的一种或多种症状明显和/或可测量的减少或消失;在一定程度上得到缓解;发病率和/或死亡率降低;和/或生活质量问题得到改善,则认为受试者的目标病理状况或病症得到成功“治疗”。用于评估成功治疗和疾病改善的上述参数可以通过医师熟悉的常规程序轻松测量。

[0073] 如本文所用,“载体”是指可以携带例如第二目的多核苷酸并且例如可以将基因序列转移到靶细胞的任何多核苷酸。因此,该术语包括克隆和表达载体以及整合载体。

[0074] 发明详述

[0075] 当尝试在细胞群中表达目的转基因时,可能出现的主要障碍之一是表达水平较差,这可能是由于携带目的转基因的DNA分子难以到达细胞核造成的。本发明解决了这个问题。

[0076] 因此,本发明涉及一种在细胞群中增加编码至少一种目的转基因的核酸分子的表达的方法。

[0077] 本发明还涉及一种在细胞群中增加编码至少一种目的转基因的核酸分子的核定

位的方法。

[0078] 本发明还涉及一种编辑细胞群基因组的方法,特别是通过在细胞群基因组中插入编码至少一种目的转基因的核酸分子。

[0079] 本发明还涉及在细胞群(优选非分裂细胞群)中转染编码至少一种目的转基因的核酸分子的方法。

[0080] 在一些实施方案中,编码至少一种目的转基因的核酸分子通过至少一种技术转染到细胞群中,所述至少一种技术选自包含下列或由下列组成的组:脂质转染、电穿孔、超声处理、纳米颗粒、显微注射、PEI和病毒载体感染,包括非整合型和整合型病毒载体感染。

[0081] 本发明是基于发明人的观察:细胞质转座酶能够将核酸递送至细胞核,从而增加其核定位。具体来说,细胞质转座酶能够将核酸易位或促进易位至细胞(包括静止细胞)的细胞核。因此,核定位的增加会诱导核酸表达水平的增加。有趣的是,发明人观察到转座酶增加的核定位并不依赖于其酶活性。

[0082] 应当理解,编码至少一种目的转基因的DNA分子在被递送到细胞的细胞质中时,将会被细胞的免疫系统感知,从而引发针对存在的外来DNA的反应(例如,炎症反应)。此外,编码至少一种目的转基因的DNA分子也必须穿过核膜才能到达细胞核。当编码至少一种目的转基因的DNA分子被细胞质转座酶易位时,这些问题就得到了克服。

[0083] 根据本发明,这些方法包括使细胞群与下列物质接触:

[0084] -编码至少一种目的转基因的核酸分子,以及

[0085] -包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸。

[0086] 在一些实施方案中,编码至少一种目的转基因的核酸分子通过转染递送至细胞群。在一些实施方案中,编码至少一种目的转基因的核酸分子通过至少一种技术转染到细胞群中,所述至少一种技术选自包含下列或由下列组成的组:脂质转染、电穿孔、超声处理、纳米颗粒、显微注射、PEI和病毒载体感染,包括非整合型和整合型病毒载体感染。

[0087] 在一些实施方案中,可以以任何顺序使细胞群与编码至少一种目的转基因的核酸分子、以及包含转座酶或其片段的蛋白或多肽或编码该蛋白或多肽的核酸接触。

[0088] 在一些实施方案中,使细胞群与编码至少一种目的转基因的核酸分子接触,并与包含转座酶或其片段的蛋白或多肽或编码该蛋白或多肽的核酸同时接触。

[0089] 在一个实施方案中,细胞群在与编码至少一种目的转基因的核酸分子接触之前,先与包含转座酶或其片段的蛋白或多肽或编码该蛋白或多肽的核酸接触。在另一个实施方案中,细胞群在与包含转座酶或其片段的蛋白或多肽或编码该蛋白或多肽的核酸接触之前,先与编码至少一种目的转基因的核酸分子接触。

[0090] 术语“编码至少一种目的转基因的核酸分子”是指待插入细胞(优选真核细胞,更优选哺乳动物细胞(包括人细胞))的基因组中的核酸序列,其编码至少一种目的产物。目的产物可以是蛋白或其片段;在这种情况下,目的转基因被称为编码核酸序列。然而,该术语还涵盖非编码核酸序列,例如不编码蛋白或其片段而是表达“RNA基因”(或“非编码RNA”) (例如转移RNA、核糖体RNA、小RNA、长非编码RNA等)的核酸序列。

[0091] 在一些实施方案中,目的转基因是编码肽或蛋白(例如但不限于酶、转录因子、生长因子、营养因子、激素、细胞因子、抗体、抗原、受体、免疫调节剂、分化因子、自杀蛋白、细胞周期修饰蛋白、抗增殖蛋白、血管生成因子、抗血管生成因子、基因组编辑器、核酸酶、重

组酶、转座酶、神经递质和报告因子,包括其任何前体,以及融合蛋白)的核酸序列。在这种情况下,目的序列通常可以是(或来源于)mRNA、cDNA、gDNA、合成核酸,或它们的任何组合。

[0092] 在一些实施方案中,目的转基因可选地是非编码RNA的核酸序列。

[0093] 非编码RNA的实例包括但不限于转移RNA(tRNA)、核糖体RNA(rRNA)、小核RNA(snRNA)、小核仁RNA(snoRNA)、SmY RNA、小卡哈尔体特异性RNA(scaRNA)、指导RNA(gRNA)、Y RNA、端粒酶RNA组分(TEC)、剪接前导RNA(SL RNA)、催化RNA(即核酶;例如核糖核酸酶P、核糖核酸酶MRP等)、反义RNA(aRNA)、顺式天然反义转录本(cis-NAT)、CRISPR RNA(crRNA)、长非编码RNA(lncRNA)、微小RNA(miRNA)、piwi相互作用RNA(piRNA)、小干扰RNA(siRNA)、短发夹RNA(shRNA)、反式作用siRNA(tasiRNA)、重复相关siRNA(rasiRNA)、7SK RNA(7SK)、增强子RNA(eRNA)和RNA适体。

[0094] 目的转基因的实例包括编码具有治疗利益的分子的任何核酸序列,例如编码肽或蛋白的任何核酸序列,或非编码RNA,其在受试者中是缺失的、不足的和/或无功能的。

[0095] 在一些实施方案中,编码至少一种目的转基因的核酸分子包含至少一个转基因和至少一个调控元件。调控元件的实例包括但不限于启动子、增强子、沉默子、绝缘子等。优选地,至少一个调控元件位于至少一种目的转基因的上游,即5'处。

[0096] 编码至少一种目的转基因的核酸分子可以是双链的或单链的。

[0097] 编码至少一种目的转基因的核酸分子可以是脱氧核糖核酸(DNA)分子、核糖核酸(RNA)分子,或它们的混合物;优选地,编码至少一种目的转基因的核酸分子是DNA分子。

[0098] 编码至少一种目的转基因的核酸分子通常包含天然核苷酸。然而它也可能包含非天然核苷酸。本文中使用的“天然核苷酸”是指腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)、胸腺嘧啶(T)和尿嘧啶(U)。如本文所用,术语“非天然核苷酸”是指化学修饰的A、T、U、C或G核苷酸。

[0099] 在一些实施方案中,编码至少一种目的转基因的核酸分子的长度为至少10个碱基对(bp),例如至少10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000bp或更多。

[0100] 在一些实施方案中,编码至少一种目的转基因的核酸分子是DNA分子,该DNA分子选自包含互补DNA(cDNA)和基因组DNA(gDNA)或由互补DNA(cDNA)和基因组DNA(gDNA)组成的组。在一些实施方案中,DNA分子是cDNA。在一些实施方案中,DNA分子是gDNA。

[0101] 在一些实施方案中,编码至少一种目的转基因的核酸分子还包含至少一个末端反向重复(ITR)序列。

[0102] 如本文所用,“ITR序列”是指天然存在于真核转座元件(或转座子)末端的核酸序列。它们被称为“5' ITR”和“3' ITR”。通常,5' ITR和3' ITR是互补的并且能够形成发夹结构。ITR序列可用于转座酶识别转座子。在一些实施方案中,ITR序列是回文的。

[0103] ITR序列的一些非限制性实例包括SEQ ID NO:55至60,以及其互补序列。

[0104] SEQ ID NO:55 ccctagaaagatagtgctgcgtaaaattgacgcatg

[0105] SEQ ID NO:56

[0106] ccctagaaagatagtgctgcgtaaaattgacgcatgagataatcaatattgtgacgtacgttaa

[0107] SEQ ID NO:57

[0108] ccctagaaagatagtgctgcgtaaaattgacgcatgagataatcaatattgtgacgtacgttaaagata

atcat gcgtaaaattgacgcatg

[0109] SEQ ID NO:58gattatctttctaggg

[0110] SEQ ID NO:59cacaatatgattatctttctaggg

[0111] SEQ ID NO:60

[0112] catgcgtcaattttacgcatgattatctttaacgtacgtacgtcacaatatgattatctttctaggg

[0113] 在一些实施方案中,至少一个ITR序列与包含转座酶或其片段的蛋白或多肽相互作用或非共价结合。在一些实施方案中,包含转座酶或其片段的蛋白或多肽特异性识别并结合至少一个ITR序列。

[0114] 在一些实施方案中,至少一个ITR序列与目的转基因相邻。“相邻”意指没有核苷酸将目的转基因与该至少一个ITR序列分开。或者,至少一个ITR序列可与目的转基因间隔至少一个核苷酸,例如1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100个或更多个核苷酸。

[0115] 在一些实施方案中,编码至少一种目的转基因的核酸分子包含两个ITR序列。优选地,这两个ITR序列位于目的转基因的两侧。在一些实施方案中,这两个ITR序列直接位于目标转基因的两侧,即没有核苷酸将目标转基因与这两个侧翼ITR序列分隔开。或者,目的转基因可与这两个ITR序列中的至少一个或两个ITR序列相隔至少一个核苷酸,例如1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100个或更多个核苷酸。

[0116] 在一些实施方案中,编码至少一种目的转基因的核酸分子由细胞群稳定或瞬时表达,优选稳定表达。

[0117] 在一些实施方案中,编码至少一种目的转基因的核酸分子是瞬时表达的。在优选的实施方案中,编码至少一种目的转基因的核酸分子被转染。在一些实施方案中,转染包括使用至少一种载体和/或方法来促进或增加编码至少一种目的转基因的核酸分子的细胞质转导。转染技术的实例包括但不限于脂质转染、电穿孔、超声处理、纳米颗粒、显微注射、PEI和病毒载体感染,包括非整合型和整合型病毒载体感染。

[0118] 在一些实施方案中,编码至少一种目的转基因的核酸分子是插入到细胞群的细胞的基因组中的DNA分子。应该理解,术语“插入基因组”意味着DNA分子在有丝分裂期间与细胞基因组一起复制。

[0119] 目的转基因可以是外源基因或内源基因。

[0120] 在一些实施方案中,目的转基因是外源基因。如本文所用,“外源基因”是指细胞中不是天然存在的,但可以通过一种或多种遗传、生物化学或其他方法引入细胞中的基因。可以根据细胞的特定发育阶段和环境条件确定在细胞中的天然存在。例如,在胚胎发育期间仅存在于细胞中的分子对于成年阶段的细胞而言是外源的。类似地,通过热休克细胞诱导的分子对于非热休克细胞而言是外源的。

[0121] 在一些实施方案中,目的转基因是内源基因。如本文所用,“内源基因”是指在细胞群的基因组中天然存在至少一个拷贝的基因。在一些实施方案中,至少一个拷贝可以是等位基因,即包含至少一个突变的基因版本。在某些实施方案中,转基因内源基因的表达增加该基因的总表达水平(即转录本)。

[0122] 在一些实施方案中,编码至少一种目的转基因的核酸分子包含在载体中,例如但不限于质粒、F黏粒、粘粒、人工染色体或病毒载体。

[0123] 在一些实施方案中,编码至少一种目的转基因的核酸分子包含在质粒中。质粒可以是环状的或线状的,优选为环状的。在一些实施方案中,编码至少一种目的转基因的核酸分子包含在F黏粒中。在一些实施方案中,编码至少一种目的转基因的核酸分子包含在粘粒中。在一些实施方案中,编码至少一种目的转基因的核酸分子包含在人工染色体(例如,人类人工染色体)中。

[0124] 在一些实施方案中,编码至少一种目的转基因的核酸分子包含在病毒载体中。

[0125] 病毒载体可以是基于DNA病毒的载体。此类基于DNA病毒的载体的一些非限制性实例包括源自双链DNA病毒域、单链DNA病毒域或多链DNA病毒域的病毒的载体。

[0126] 在一些实施方案中,基于DNA病毒的载体源自双链DNA病毒域的病毒。双链DNA病毒域的病毒包括疱疹病毒目(Herpesvirales)病毒。因此,在一些实施方案中,基于DNA病毒的载体是源自疱疹病毒科(Herpesviridae)的载体,例如源自单纯疱疹病毒。

[0127] 在一些实施方案中,基于DNA病毒的载体源自单链DNA病毒域的病毒。单链DNA病毒域的病毒包括乳头瘤病毒科(Papillomaviridae)和多瘤病毒科(Polyomaviridae)的病毒。因此,在一些实施方案中,基于DNA病毒的载体是源自乳头瘤病毒科或多瘤病毒科的载体。

[0128] 在一些实施方案中,基于DNA病毒的载体源自多链DNA病毒域的病毒。多链DNA病毒域的病毒包括腺病毒科(Adenoviridae)和痘病毒科(Poxviridae)的病毒。因此,在一些实施方案中,基于DNA病毒的载体是源自腺病毒科的载体,例如源自腺病毒;或源自痘病毒科的载体,例如源自痘苗病毒。

[0129] 在一些实施方案中,可以从载体中切除编码至少一种目的转基因的核酸分子。在一些实施方案中,编码至少一种目的转基因的核酸分子可以通过至少一个ITR序列(通常通过转座酶或其变体或片段)从载体中切除。

[0130] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽增加编码至少一种目的转基因的核酸分子的核定位。

[0131] 如本文所用,“增加核定位”是指与不存在包含转座酶或其片段的蛋白或多肽的情况相比,在存在包含转座酶或其片段的蛋白或多肽的情况下,编码至少一种目的转基因的核酸分子的核定位增加至少1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍、100倍或更多倍。

[0132] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽将编码至少一种目的转基因的核酸分子易位或促进易位到细胞核。在一些实施方案中,转座酶或其片段(i)与编码至少一种目的转基因的核酸分子相互作用和/或非共价结合,和(ii)易位到细胞核中。

[0133] 在一些实施方案中,转座酶或其片段不会将编码至少一种目的转基因的核酸分子插入和/或易位到基因组中。

[0134] 在一些实施方案中,转座酶或其片段使编码至少一种目的转基因的核酸分子穿过核膜增加至少1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、20倍、30倍、40倍、50倍、60倍、70倍、80倍、90倍、100倍或更多倍。

[0135] 在一些实施方案中,转座酶或其片段保护编码至少一种目的转基因的核酸分子免受细胞的传感机制和/或防御机制的影响。在一些实施方案中,转座酶或其片段不会诱导促炎和/或炎症反应。

[0136] 在一些实施方案中,转座酶选自由高活性PiggyBac转座酶、PiggyBac转座酶、睡美人转座酶、SB11转座酶、Tol2转座酶、Mos1转座酶和青蛙王子转座酶组成的组。

[0137] 在一些实施方案中,转座酶选自由高活性PiggyBac转座酶和睡美人转座酶组成的组。

[0138] 在一些实施方案中,转座酶是高活性PiggyBac转座酶。

[0139] 在一些实施方案中,转座酶是具有SEQ ID NO:1的氨基酸序列的高活性PiggyBac转座酶。

SEQ ID NO: 1

[0140] MGSSLDDEHILSALLQSDDELVGEDSDSEVSDHVSEDDVQSDTEEAFIDEV
HEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKPT
RRSRVSALNIVRSQRGPTRMCRNIYDPLLCKLFFTDIIEIVKWTNAEISLK
RRESMTSATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLDFDRSLSMVYV
SVMSRDRDFDLIRCLRMDDKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTP
GAHLTIDEQLLGFGRGCPFRVYIPNKPSKYGIKILMMCDGSKYMINGMPYL
GRGTQTNGVPLGEYYVKELSKPVHGSCRNITCDNWFTSIPLAKNLLQEPYK
LTIVGTVRSNKREIPEVLKNSRSRPVGTSMFCFDGPLTLVSYKPKPAKMVYL
LSSCDEDASINESTGKPMVMYNYNQTGGVDTLDQMCSVMTCRKTNRW
PMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMGLTSSFMRK

[0141] RLEAPTLKRYLRDNISNILPKEVPGTSDDSTEETVMKKRTYCTYCPSKIRRK
ASASCKKCKKVICREHNIDMCQSCF

[0142] 在一些实施方案中,转座酶是修饰的高活性PiggyBac转座酶。“修饰的高活性PiggyBac转座酶”是指与具有SEQ ID NO:1的氨基酸序列的高活性PiggyBac转座酶相比,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,修饰的高活性PiggyBac转座酶可以包含(i)一个或多个氨基酸取代,以与具有SEQ ID NO:1的氨基酸序列的高活性PiggyBac转座酶相比提高切除活性,和/或(ii)一个或多个氨基酸取代,以与具有SEQ ID NO:1的氨基酸序列的高活性PiggyBac转座酶相比降低DNA结合活性。

[0143] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含与SEQ ID NO:1所示的序列至少80%、85%、90%、95%、96%、97%、98%或99%同一的氨基酸序列。

[0144] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以增加切除活性。

[0145] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以增加切除活性,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具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。

[0146] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以

增加切除活性,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具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。

[0147] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以增加切除活性,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具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号;优选地,氨基酸取代是M194V和/或D450N。

[0148] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以降低DNA结合活性。

[0149] 在一些实施方案中,修饰的高活性PiggyBac转座酶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具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。

[0150] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以降低DNA结合活性,所述突变选自R275、N347、R372、K375、R376、E377和E380,所述位置编号对应于SEQ ID NO:1的高活性PiggyBac的氨基酸编号。

[0151] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以降低DNA结合活性,所述突变选自R372、K375、R376、E377和E380,所述位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号;优选地,氨基酸取代是R372A、K375A、R376A、E377A和/或E380A。

[0152] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含一个或多个氨基酸突变以降低DNA结合活性,所述突变选自N347、R372和K375,所述位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号;优选地,氨基酸取代是N347S、N347A、R372A和/或K375A;更优选地,氨基酸取代是N347S或N347A。

[0153] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含如上所定义的一个或多个氨基酸突变以增加切除活性;以及如上所定义的一个或多个氨基酸突变以降低DNA结合活性。

[0154] 在一些实施方案中,修饰的高活性PiggyBac转座酶包括至少一个氨基酸取代以增加位置D450的切除活性,以及至少两个氨基酸取代以降低位置N347、R372和K375的DNA结合活性;优选地,所述修饰的高活性PiggyBac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具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。

[0155] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含双突变N347S和D450N,所述位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。

[0156] 在一些实施方案中,如先前实施方案所公开的修饰的高活性PiggyBac转座酶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。

[0157] 在一些实施方案中,修饰的高活性PiggyBac转座酶进一步包括在位置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处的一个或多个氨基酸取代,位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸

编号。

[0158] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含以下氨基酸取代之一或氨基酸取代的组合: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、F594L、D450N/R372A/K375A、R275A/R277A、K409A/K412A、R460A/K461A、R275A/R277A/N347S/K375A/T560A/S573A/M589V/S592G 或R245A/R275A/R277A/R372A/W465A。

[0159] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含以下氨基酸取代之一或氨基酸取代的组合:

- [0160] -R372A/K375A/D450N,
- [0161] -R372A/K375A/R376A/D450N,
- [0162] -K375A/R376A/E377A/E380A/D450N,
- [0163] -R372A/K375A/R376A/E377A/E380A/D450N,
- [0164] -M194V,
- [0165] -M194V/R372A/K375A,
- [0166] -S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L,
- [0167] -R245A/R275A/R277A/R372A/W465A/M589V,
- [0168] -R275A/325A/R372A/T560A,
- [0169] -N347A/D450N,
- [0170] -N347S/D450N/T560A/S573A/F594L,
- [0171] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,
- [0172] -R275A/N347S/K375A/D450N/S592G,
- [0173] -R275A/N347S/R372A/D450N/T560A/F594L,
- [0174] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,
- [0175] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,
- [0176] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,
- [0177] -V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P,
- [0178] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,
- [0179] -S230N/R277A/N347S/K375A/D450N,
- [0180] -T43I/R372A/K375A/A411T/D450N,
- [0181] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,或
- [0182] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,
- [0183] 位置编号对应具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。
- [0184] 一些优选的修饰的高活性PiggyBac转座酶包括修饰的高活性PiggyBac,其包含下列氨基酸取代的组合之一:

- [0185] -R372A/K375A/D450N,
- [0186] -S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L,
- [0187] -R245A/R275A/R277A/R372A/W465A/M589V,

- [0188] -N347A/D450N,
- [0189] -N347S/D450N/T560A/S573A/F594L,
- [0190] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,
- [0191] -R275A/N347S/K375A/D450N/S592G,
- [0192] -R275A/N347S/R372A/D450N/T560A/F594L,
- [0193] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,
- [0194] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,
- [0195] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,
- [0196] -V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P,
- [0197] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,
- [0198] -S230N/R277A/N347S/K375A/D450N,
- [0199] -T43I/R372A/K375A/A411T/D450N,
- [0200] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,
- [0201] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,或
- [0202] -R275A/325A/R372A/T560A,
- [0203] 位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。
- [0204] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含下列氨基酸取代的组合之一:
- [0205] -R245A/R275A/R277A/R372A/W465A/M589V,
- [0206] -R275A/325A/R372A/T560A,
- [0207] -N347A/D450N,
- [0208] -N347S/D450N/T560A/S573A/F594L,
- [0209] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,
- [0210] -R275A/N347S/K375A/D450N/S592G,
- [0211] -R275A/N347S/R372A/D450N/T560A/F594L,
- [0212] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,
- [0213] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,
- [0214] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,
- [0215] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,
- [0216] -S230N/R277A/N347S/K375A/D450N,
- [0217] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,或
- [0218] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,
- [0219] 位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。
- [0220] 在一些优选的实施方案中,修饰的高活性PiggyBac转座酶包含下列氨基酸取代的组合之一:
- [0221] -N347A/D450N,
- [0222] -N347S/D450N/T560A/S573A/F594L,
- [0223] -R202K/R275A/N347S/R372A/D450N/T560A/F594L,
- [0224] -R275A/N347S/K375A/D450N/S592G,

- [0225] -R275A/N347S/R372A/D450N/T560A/F594L,
- [0226] -R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L,
- [0227] -R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G,
- [0228] -R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L,
- [0229] -G325A/N347S/K375A/D450N/S573A/M589V/S592G,
- [0230] -S230N/R277A/N347S/K375A/D450N,
- [0231] -G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G,或
- [0232] -Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G,
- [0233] 位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。
- [0234] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含R372A/K375A/D450N取代,所述位置编号对应于具有SEQ ID NO:1的高活性PiggyBac的氨基酸编号。所述修饰的转座酶具有SEQ ID NO:4的氨基酸序列。
- [0235] 在一些实施方案中,修饰的高活性PiggyBac转座酶具有选自包含SEQ ID NO:5-26或由SEQ ID NO:5-26组成的组的氨基酸序列。
- [0236] 在一些实施方案中,修饰的高活性PiggyBac转座酶具有选自包含SEQ ID NO:5-13或由SEQ ID NO:5-13组成的组的氨基酸序列。
- [0237] 在一些实施方案中,修饰的高活性PiggyBac转座酶具有选自包含SEQ ID NO:14-26或由SEQ ID NO:14-26组成的组的氨基酸序列。
- [0238] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变参与保守的催化三联体,例如在对应于SEQ ID NO:1的氨基酸编号的氨基酸268和/或346处(例如,D268N和/或D346N)。
- [0239] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变对于切除至关重要,例如在对应于SEQ ID NO:1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)。
- [0240] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,所述突变参与靶连接,例如在对应于SEQ ID NO:1的氨基酸编号的氨基酸351、356和/或379处(例如,S351E、S351P、S351A和/或K356E)。
- [0241] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变对于整合至关重要,例如在对应于SEQ ID NO:1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)。
- [0242] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变参与比对,例如在对应于SEQ ID NO:1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)。
- [0243] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变是高度保守的,例如在对应于SEQ ID NO:1的氨基酸编号的氨基酸576和/或587处(例如,K576A和/或I587A)。

[0244] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变参与Zn²⁺结合,例如在对应于SEQ ID NO:1的氨基酸编号的586处(例如,H586A)。

[0245] 在一些实施方案中,修饰的高活性PiggyBac转座酶可以包含相对于高活性PiggyBac转座酶的一个或多个突变,这些突变参与整合,例如,在对应于SEQ ID NO:1的氨基酸编号的315、341、372和/或375处(例如,R315A、R341A、R372A和/或K375A)。

[0246] 在一些实施方案中,选择修饰的高活性PiggyBac是因为与高活性PiggyBac相比,它对DNA整合到基因组中具有更高的特异性。

[0247] 在一些实施方案中,修饰的高活性PiggyBac包含具有本文所公开的相对于SEQ ID NO:1的一个或多个修饰的氨基酸序列,并且与SEQ ID NO:5-26中的任一个保持至少80%、85%、90%、95%、96%、97%、98%、99%或更高的序列同一性。

[0248] 在一些实施方案中,修饰的高活性PiggyBac转座酶可包含选自对应于SEQ ID NO:1的氨基酸编号的氨基酸245、275、277、325、347、351、372、375、388、450、465、560、564、573、589、592和594处的一个或多个氨基酸的突变。

[0249] 在一些实施方案中,修饰的高活性PiggyBac转座酶突变可以包含选自以下的对应于SEQ ID NO:1的氨基酸编号的一个或多个氨基酸取代:R245A、R275A、R277A、R275A/R277A、G325A、N347A、N347S、S351E、S351P、S351A、R372A、K375A、R388A、D450N、W465A、T560A、S564P、S573A、M589V、S592G和F594L。

[0250] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代D450N。

[0251] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代R245A和D450。

[0252] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代R245A、G325A和S573P。

[0253] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代R245A、G325A、D450和S573P。

[0254] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代N347S或N347A。

[0255] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代N347S和D450N。

[0256] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含对应于SEQ ID NO:1的氨基酸编号的氨基酸取代N347A和D450N。该修饰的高活性PiggyBac转座酶包含SEQ ID NO:14的氨基酸序列。

[0257] 在一些实施方案中,修饰的高活性PiggyBac转座酶包含具有SEQ ID NO:1的氨基酸序列,其中:

[0258] -位置34的氨基酸残基为V或M,

[0259] -位置43的氨基酸残基为T或I,

[0260] -位置177的氨基酸残基为Y或H,

[0261] -位置202的氨基酸残基为R或K,

- [0262] -位置230的氨基酸残基为S或N,
- [0263] -位置245的氨基酸残基为A,
- [0264] -位置268的氨基酸残基为D或N,
- [0265] -位置277的氨基酸残基为R或A,
- [0266] -位置275的氨基酸残基为R或A,
- [0267] -位置277的氨基酸残基为R或A,
- [0268] -位置325的氨基酸残基为A或G,
- [0269] -位置347的氨基酸残基是S或A,
- [0270] -位置351的氨基酸残基是E、P或A, -位置372的氨基酸残基是R或A,
- [0271] -位置375的氨基酸残基是K或A,
- [0272] -位置388的氨基酸残基是R或A,
- [0273] -位置409的氨基酸残基是K或A,
- [0274] -位置411的氨基酸残基是A或T,
- [0275] -位置412的氨基酸残基是K或A,
- [0276] -位置450的氨基酸残基是D或N,
- [0277] -位置460的氨基酸残基是R或A,
- [0278] -位置465的氨基酸残基是W或A,
- [0279] -位置517的氨基酸残基是S或A,
- [0280] -位置560的氨基酸残基是T或A,
- [0281] -位置564的氨基酸残基是P或S,
- [0282] -位置571的氨基酸残基是S或N,
- [0283] -位置573的氨基酸残基是S或A,
- [0284] -位置576的氨基酸残基是K或A,
- [0285] -位置586位置的氨基酸残基是H或A, -位置587的氨基酸残基是I或A,
- [0286] -位置589的氨基酸残基是M或V,
- [0287] -位置592的氨基酸残基是G或S, 和/或,
- [0288] -位置594的氨基酸残基是L或F。

[0289] 在一些实施方案中,修饰的高活性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%或更高的序列同一性的氨基酸序列或其组成:SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、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和SEQ ID NO:26。

[0290] 在一些实施方案中,修饰的高活性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%或更高的序列同一性的氨基酸序列或其组成:SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、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和SEQ ID NO:26。

[0291] 在一些实施方案中,修饰的转座酶不是Himar1C9突变体。

[0292] 在一些实施方案中,转座酶是睡美人转座酶。

[0293] 在一些实施方案中,转座酶是具有SEQ ID NO:61的氨基酸序列的高活性睡美人SB100转座酶。

SEQ ID NO: 61

[0294] MGKSKEISQDLRKRIVDLHKSGSSLGAISKRLAVPRSSVQTIVRKYKHHGTT
QPSYRSGRRRVLSPRDERTLVRKVQINPRTTAKDLVKMLEETGTKVSISTVK
RVLYRHNLKGHSARKKPLLQNRHKARLRFATAHGDKDRTFWRNVLWSD
ETKIELFGHNDHRYVWRKKGEACKPKNTIPTVKHGGGSIMLWGCFAAGGT
GALHKIDGIMDAVQYVDILKQHLKTSVRKLLKGRKWVVFQHDNDPKHTSK

[0295] VVAKWLKDNKVKVLEWPSQSPDLNPIENLWAEKRRVRRARRPTNLTQLHQ
LCQEEWAKIHPNYCGKLVEGYPKRLTQVKQFKGNATKY

[0296] 在一些实施方案中,转座酶是修饰的高活性睡美人SB100转座酶,与未修饰的睡美人转座酶的氨基酸序列相比包含至少一个氨基酸取代,例如1、2、3、4、5、6、7、8、9、10或更多个取代。

[0297] 在一些实施方案中,修饰的高活性睡美人SB100转座酶包含在选自对应于SEQ ID NO:61的氨基酸编号的C176、H187、I212、P247和K248的位置处的一个或多个氨基酸取代。

[0298] 在一些实施方案中,修饰的高活性睡美人SB100转座酶包含选自对应于SEQ ID NO:61的氨基酸编号的C176S、H187V/P、I212S、P247R/S和K248A/C/I/L/M/N/R/S/T/V的一个或多个氨基酸取代。

[0299] 在一些实施方案中,转座酶或其片段具有降低的催化活性。如本文所用,“降低的催化活性”意指与野生型转座酶和/或高活性转座酶相比降低1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2倍、3倍、4倍、5倍、10倍、20倍或更多。

[0300] 在一些实施方案中,转座酶或其片段是催化失活的。在一些实施方案中,转座酶或其片段不发挥任何催化活性。在一些实施方案中,催化失活的转座酶保留其与其他蛋白、多肽和/或核酸分子结合的能力。

[0301] 在一些实施方案中,催化失活的转座酶与具有SEQ ID NO:3的催化失活的高活性PiggyBac转座酶(失活hyPB)具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%、99.9%或更高的序列同一性。在一些实施方案中,失活hyPB包含具有SEQ ID NO:3的氨基酸序列或由具有SEQ ID NO:3的氨基酸序列组成。

SEQ ID NO: 3

[0302] MGSSLDDEHILSALLQSDDELVGEDSDSEVSDHVSEDDVQSDTEEFIDEV
HEVQPTSSGSEILDEQNVIEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKPT
RRSRVSALNIVRSQRGPTRMCRNIYDPLLCKLFFTDDEIISEIVKWTNAEISLK
RRESMTSATFRDTNEDEIYAFFGILVMTAVRKDNHMSTDDLDFDRSLSMVYV
SVMSRDRDFDLIRCLRMDDKSIRPTLRENDVFTVVRKIWDLFIHQCIQNYTP

[0303] GAHLTIAEQLLGFRGRCPFRVYIPNKPSKYGIKILMMCDGSKYMINGMPYL
GRGTQTNGVPLGEYYVKELSKPVHGSCRNITCANWFTSIPLAKNLLQEPYK
LTIVGTVASNAREIPEVLKNSRSPVGTSMFCFDGPLTLVSYKPKPAKMVYL
LSSCDEDASINESTGKPMVMYYNQTKGGVDTLNQMCVMTCSRKTNRW
PMALLYGMINIACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMGLTSSFMRK
RLEAPTLKRYLRDNISNILPKEVPGTSDDSTEPPVMKKRTYCTYCPSKIRRK
ASASCKKCKKVICREHNIDMCQSCF

[0304] 在一些实施方案中,蛋白或多肽包含转座酶片段。

[0305] 在一些实施方案中,转座酶片段包含至少一个转座酶功能结构域或由至少一个转座酶功能结构域组成。

[0306] 在一些实施方案中,转座酶片段包含ITR结合结构域或由ITR结合结构域组成。

[0307] ITR结合结构域的实例包括但不限于SEQ ID NO:62至64,其中SEQ ID NO:62和63是高活性PiggyBac转座酶的ITR结合结构域,而SEQ ID NO:64是睡美人转座酶的ITR结合结构域(也称为“N57靶向结构域”)。

[0308] SEQ ID NO:62

[0309] ILPKEVPGTSDDSTEPPVMKKRTYCTYCPSKIRRKASASCKKCKKVI CREHNIDMCQSCF

[0310] SEQ ID NO:63

[0311] MKKRTYCTYCPSKIRRKASASCKKCKKVICREHNIDMCQSCF

[0312] SEQ ID NO:64

[0313] MGKSKEISQDLRKRIVDLHKSGSSLGAISKRLAVPRSSVQTIVRKYK HHGTTQPSYR

[0314] 在一些实施方案中,转座酶片段是高活性PiggyBac转座酶或睡美人转座酶的片段。

[0315] 在一些实施方案中,转座酶片段是高活性PiggyBac转座酶的片段,优选地包含ITR结合结构域。

[0316] 在一些实施方案中,转座酶片段是睡美人转座酶的片段,优选地选自睡美人转座酶的SB100结构域和N57结构域组成的组。

[0317] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽是包含转座酶或其片段和至少一种额外的多肽或蛋白的融合蛋白。

[0318] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽是包含转座酶或其片段以及至少两种另外的多肽或蛋白的融合蛋白。

[0319] 在一些实施方案中,融合蛋白包含单个连续的多肽链。

[0320] 在一些实施方案中,融合蛋白还包含融合蛋白的两个多肽或蛋白之间的至少一个接头,特别是至少一个肽接头。

[0321] 示例性接头包括但不限于 $(G)_n$ 、 $(GS)_n$ 、 $(GGS)_n$ 、 $(GGGS)_n$ (具有SEQ ID NO:49)、 $(GGGGS)_n$ (具有SEQ ID NO:50)、 $(EAAAK)_n$ (具有SEQ ID NO:51)、XTEN接头和 $(XP)_n$ 接头,以及它们的任意组合,其中n是1至50之间的整数。

[0322] 在一些实施方案中,肽接头是富含甘氨酸/丝氨酸的接头。在一些实施方案中,肽接头具有选自包含以下或由以下组成的组的氨基酸序列:SEQ ID NO:41、SEQ ID NO:42、SEQ ID NO:43、SEQ ID NO:44、SEQ ID NO:45和SEQ ID NO:46。

[0323] SEQ ID NO:41:GGSGGSGGG

[0324] SEQ ID NO:42:GGSGGSGGSGGS

[0325] SEQ ID NO:43:GGSGGSGGSGGSGGS

[0326] SEQ ID NO:44:GGSGGSGGSGGSGGSG

[0327] SEQ ID NO:45:GGSGGSGGSGGSGGSGGSGGS

[0328] SEQ ID NO:46:GGSGGSGGSGGSGGSGGSGGSGGS

[0329] 在一些实施方案中,肽接头是包含SEQ ID NO:47或由SEQ ID NO:47组成的XTEN接头。

[0330] SEQ ID NO:47:SGSETPGTSESATPES

[0331] 在一些实施方案中,肽接头包含SEQ ID NO:48或由SEQ ID NO:48组成。

[0332] SEQ ID NO:48:GSAGSAAGSGEF

[0333] 在一些实施方案中,肽接头包含SEQ ID NO:55或由SEQ ID NO:55组成。

[0334] SEQ ID NO:55:GIHGVPAA

[0335] 可以使用本领域技术人员熟知的方法来构建含有融合蛋白的编码序列以及适当的转录/翻译控制信号的表达载体。这些方法包括体外重组DNA技术、合成技术和体内重组/基因重组。例如,请参阅Sambrook等人,2012 (Molecular cloning:A laboratory manual (第4版).Cold Spring Harbor Laboratory Press)中描述的技术。表达载体可以是质粒、病毒的一部分,或者可以是核酸片段。表达载体包括表达盒,编码融合蛋白的多核苷酸(即编码区)以可操作的方式与启动子和/或其他转录或翻译控制元件连接克隆到该表达盒中。如本文所用,“编码区”是由翻译成氨基酸的密码子组成的核酸部分。尽管“终止密码子”(TAG、TGA或TAA)不会翻译成氨基酸,但如果存在,它可以被视为编码区的一部分,但是任何侧翼序列,例如启动子、核糖体结合位点、转录终止子、内含子、5'-和3'-非翻译区等,不是编码区的一部分。两个或更多个编码区可存在于单个多核苷酸构建体中(例如,在单个载体上),或存在于单独的多核苷酸构建体中(例如,在单独的(不同的)载体上)。此外,任何载体可以含有单个编码区,或者可以包含两个或更多个编码区,例如,本发明的载体可以编码一个或多个多肽,这些多肽在翻译后或共翻译过程中经由蛋白水解裂解分离成最终的蛋白。此外,本发明的载体、多核苷酸或核酸可以编码异源编码区,这些编码区可以融合或不融合至编码本发明的融合蛋白(片段)或其变体或衍生物的多核苷酸。异源编码区包括但不限于专门的元件或基序,例如分泌信号肽或异源功能结构域。可操作连接是指基因产物(例如多肽)的编码区与一个或多个调控序列连接,从而使基因产物的表达受到调控序列的影响或

控制。如果启动子功能的诱导导致编码所需基因产物的mRNA转录,并且两个DNA片段之间的连接性质不干扰表达调控序列指导基因产物表达的能力或干扰DNA模板被转录的能力,则两个DNA片段(例如多肽编码区和与其相关的启动子)是“可操作连接的”。因此,如果启动子能够实现该核酸的转录,则启动子区将与编码多肽的核酸可操作地连接。启动子可以是细胞特异性启动子,其仅在预定的细胞中指导DNA的大量转录。除启动子之外的其他转录控制元件,例如增强子、操纵子、阻遏子和转录终止信号,也可以可操作地与多核苷酸连接以指导细胞特异性转录。本文公开了合适的启动子和其他转录控制区。本领域技术人员已知多种转录控制区。这些包括但不限于在脊椎动物细胞中起作用的转录控制区,例如但不限于来自巨细胞病毒的启动子和增强子片段(例如,与内含子-A结合的立即早期启动子)、猿猴病毒40(例如,早期启动子)和逆转录病毒(例如,Rous肉瘤病毒)。其他转录控制区包括源自脊椎动物基因的转录控制区,例如肌动蛋白、热休克蛋白、牛生长激素和兔 α -珠蛋白,以及其他能够控制真核细胞中基因表达的序列。另外的合适的转录控制区包括组织特异性启动子和增强子以及诱导型启动子(例如,四环素诱导型启动子)。类似地,多种翻译控制元件是本领域普通技术人员已知的。这些包括但不限于核糖体结合位点、翻译起始和终止密码子、以及源自病毒系统的元件(特别是内部核糖体进入位点或IRES,也称为CITE序列)。表达盒还可以包括其他特征,例如复制起点和/或染色体整合元件,例如逆转录病毒长末端重复序列(LTR)或腺相关病毒(AAV)末端反向重复(ITR)。

[0336] 可通过本领域已知技术纯化如本文所述制备的融合蛋白,例如高效液相色谱法、离子交换色谱法、凝胶电泳、亲和色谱法、尺寸排阻色谱法等。用于纯化特定蛋白的实际条件将部分取决于诸如净电荷、疏水性、亲水性等因素,并且对于本领域技术人员来说是显而易见的。对于亲和色谱法纯化,可以使用融合蛋白结合的抗体、配体、受体或抗原。融合蛋白的纯度可以通过多种熟知的分析方法中的任一种来确定,包括凝胶电泳、高压液相色谱法等。

[0337] 在一些实施方案中,编码融合蛋白的核酸可以表达为编码整个融合蛋白的单个核酸分子或表达为共表达的多个(例如,两个或更多个)核酸分子。共表达的核酸分子编码的多肽可以通过例如二硫键或其他方式缔合,形成功能性融合蛋白。

[0338] 在一些实施方案中,融合蛋白包含直接或经由接头间接地在至少一种蛋白或多肽的C末端融合的转座酶或其片段或由直接或经由接头间接地在至少一种蛋白或多肽的C末端融合的转座酶或其片段组成。在其他实施方案中,融合蛋白包含直接或经由接头间接地在至少一种蛋白或多肽的N末端融合的转座酶或其片段或由直接或经由接头间接地在至少一种蛋白或多肽的N末端融合的转座酶或其片段组成。

[0339] 在一些实施方案中,融合蛋白是三重融合蛋白,即融合蛋白包含转座酶或其片段,以及至少两种额外的多肽或蛋白。

[0340] 在一些实施方案中,融合蛋白还包含核定位序列(NLS)。一般来说,一个或多个NLS应该具有足够的强度来驱动融合蛋白在细胞的细胞核中积累。一般来说,核定位活性的强度取决于NLS的数量和位置,以及融合蛋白中使用的一个或多个特定NLS。

[0341] 在一些实施方案中,NLS可以位于融合蛋白的N端和/或C端。在一些实施方案中,融合蛋白包含约1、2、3、4、5、6、7、8、9、10或更多个NLS。在一些实施方案中,融合蛋白在N端或N端附近包含约1、2、3、4、5、6、7、8、9、10个或更多个NLS。在一些实施方案中,融合蛋白在C端

或C端附近包含约1、2、3、4、5、6、7、8、9、10个或更多个NLS。在一些实施方案中,融合蛋白包含这些的组合,例如位于N端的一个或多个NLS和位于C端的一个或多个NLS。

[0342] 当存在多于一个NLS时,每个NLS均可以独立于其他NLS进行选择。在一些实施方案中,融合蛋白包含两个NLS,例如,两个NLS分别位于N端和C端。

[0343] 一般来说,NLS由暴露在蛋白表面的一个或多个带正电荷的赖氨酸或精氨酸的短序列组成,但本领域也知道其他类型的NLS。NLS的非限制性实例包括(M)KKRKV(具有SEQ ID NO:52)、(M)PKKKRKV(具有SEQ ID NO:53)或(M)SGGSPKKRKV(具有SEQ ID NO:54),其中(M)表示启动子蛋氨酸,其可以在NLS位于N端时存在或者在翻译后被去除,而在NLS位于C端时不存在。

[0344] 此外,融合蛋白还可以包括其他定位序列,例如细胞质定位序列、叶绿体定位序列、线粒体定位序列等,这取决于融合蛋白在细胞中所需的定位。

[0345] 在一些实施方案中,至少一种额外的多肽或蛋白是核酸酶。在一些实施方案中,核酸酶是内切核酸酶或外切核酸酶。在一些实施方案中,核酸酶是脱氧核糖核酸酶或核糖核酸酶。

[0346] 在一些实施方案中,至少一种额外的多肽或蛋白是RNA引导的核酸酶。

[0347] 在一些实施方案中,至少一种额外的多肽或蛋白是Cas核酸酶。

[0348] 在一些实施方案中,Cas核酸酶选自包含以下或由以下组成的组:Cas9、Cas12a(Cpf1)、Cas12b、Cas12f和CasX。应当理解也包括其变体及功能片段,如切口酶Cas(nCas)或失活Cas(dCas)变体。

[0349] Cas9核酸酶的实例包括但不限于化脓性链球菌(*Streptococcus pyogenes*) Cas9(SpCas9)、溶血葡萄球菌(*Staphylococcus haemolyticus*) Cas9(ShCas9)和空肠弯曲菌(*Campylobacter jejuni*) Cas9(CjCas9)。

[0350] 在一些实施方案中,Cas核酸酶与具有SEQ ID NO:27的SpCas9、具有SEQ ID NO:28的ShCas9、具有SEQ ID NO:29的Cpf1、具有SEQ ID NO:30的CjCas9、具有SEQ ID NO:31的nCas9和/或具有SEQ ID NO:32的nCas9的序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性。

SEQ ID NO: 27

MDKKYSIGLDIGTNSVGVAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
LALAHMIKFRGHFLIEGDLNPDNSVDKLFIQLVQTYNQLFEENPINASGVD
AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAE
DAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI
DGGASQEEFYKFIKPILEKMDGTEELLVVLNREDLLRKQRTFDNGSIPHQIH
LGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTR
KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
TLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQK
NSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRKGSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREIN

[0351]

NYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI
 GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
 VRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
 GFDSPTVAYSVLV VAKVEK GKSKKLSVKELLGITIMERS SFEKNPIDFLEA
 KGYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEK LKGS PEDNEQKQLFVEQH KHYLDEIIEQISEFSKR VILADANL
 DKVLSAYNKHRDKPIREQAENIHLFTLTNLGAPAAF KYFDTTIDRKRYTST
 KEVL DATLIHQ SITGLYETRIDL SQLGGD

SEQ ID NO: 28

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
 ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEE
 FSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQL
 ERLKKDGEVRGSINRFKTS DYVKEAKQLLK VQKAYHQLDQSFIDTYIDLLE
 TRRTY YEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLY
 NALNDLNNLVITRDENEKLEYEYEFQIENVFKQKKKPTLKQIAKEILVNEE
 DIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIENAELLDQIAKILTIYQSSE
 DIQEELTNL NSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQI
 AIFNRLKLV PPKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYG
 LPNDIII ELAREKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIE
 KIKLHDMQEGKCLYSLEAIPLEDLLNPNFNYEVDHIIPRSVSFDNSFNKVL
 VKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLA KGKGRISKTKKEYL
 LEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSING
 GFTSFLRRKWKFKKERNKGYKHH AEDALI ANADFIFKEWKKLDKAKKVM
 ENQMFEEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
 RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDK LKKLINKSPEKLLMYH
 HDPQTYQKLKLIMEQYGDEKNPLYKY YEETGNYLTKYSKKNPVIKKIK
 YYGNKLN AHLDITDDYPNSRNKVVKLSLKP YRFDVYLDNGVYKFVTVKN
 LDVIKENYYEVNSKCYEEAKK LKKISNQA EFIASFYNNDLIKINGELYRVI
 GVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL
 GNLYEVKSKKHPQIIKKG

[0352]

SEQ ID NO: 29

SNAMSKLEKFTNCYLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYK
 GVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEIN
 LRKEIAKAFKGNEGYKSLFKKDIETILPEFLDDKDEIALVNSFNGFTTAFTGF
 FDNRENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVD AIFDKHEVQEIKE
 KILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYIN
 LYNQKTKQKLPKFKPLYKQVLS DRESLSFYGEGYTSDEEVLEVFRNTLNKN
 SEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAE
 YDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKE
 IIIQKVDEIYKVYGSSEKLFDA DVFLEKSLKKNDAVVAIMKDLLDSVKS FEN
 YIKAFFGEGKETNRDES FYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKD
 KFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKI
 DKDDVNGNYEKINYKLLPGPNKMLPKVFFS KKWMAYYNPSEDIQKIYKNG
 TFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFY
 REVEEQGYKVSFESASKKEVDKLV EEGKLYMFQIYNKDFSDKSHGTPNLHT
 MYFKLLFDENNHGQIRLSGGAE LFMRRASLKKEELVVHPANSPIANKNPDN
 PKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDD
 NPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINN FNIRIKTDYHSLLD
 KKEKERFEARQNWTSIENIKELKAGYISQV VHKICELVEKYDAVIALEDLNS
 GFKNSRVKVEKQVYQKFEKMLIDK LNYMVDKKS NPCATGGALKGYQITN
 KFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISFD
 RIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVF
 DWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLM
 LQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAY
 NIARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQT SVKH

[0353]

SEQ ID NO: 30

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARS
 ARKRLARRKARLNHLKHLIANEFKLN YEDYQSFDESLAKAYKGS LISPYEL
 RFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKL
 ANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLI

FKKQREFGFSFSKKFEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNS
 PLAFMFVALTRIINLLNNLNKNTGILYTKDDLNALLNEVLKNGTLTYKQTKK
 LLGLSDDYEFKGEKGYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIK
 DEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPMLLEGKKY
 DEACNELNLKVAINEDKKDFLPAFNETYYKDEVTPVVLRAIKEYRKVLNA
 LLKKYGVHVKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKL
 GLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRFD
 DSYMKNVLVFTKQNQEKLNQTPEAFGNDSAKWQKIEVLAKNLPTKKQK
 RILDKNYKDKEQKNFKDRNLNDTRYIARLVLNNTKDYLDLPLSDDENTKL
 NDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYAN
 NSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKI
 DEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVN
 GKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEI
 KDWILMDENYEFCSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKH
 DNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEF
 RQREDFKK

[0354]

SEQ ID NO: 31

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
 LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEE
 SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
 LALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD
 AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAE
 DAKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
 TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI
 DGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIH
 LGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAMTR
 KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
 VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
 TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK

QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
 NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
 NSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
 ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRSGKSDNVPSEEVVK
 KMKNYWRQLLNAKLITQRKFDNLTKAERGGELSELDKAGFIKRQLVETRQIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREIN
 NYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI
 GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
 VRKVLSPQVNVIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
 GFDSPTVAYSVLVVAKVEKGGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEA
 KGYKEVKKDLIKLPKYSLEFLENKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANL
 DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVLDTLIHQISITGLYETRIDLSQLGGD

[0355]

SEQ ID NO: 32

MKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRG
 ARRLKRRRRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEE
 FSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQL
 ERLKKDGEVVRGSINRFKTSYVKEAKQLLVQKAYHQLDQSFIDTYIDLLE
 TRRTYYEGPGEKSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLY
 NALNDLNNLVITRDENEKLEYEYEFQIENVFKQKKKPTLKQIAKEILVNEE
 DIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIEAELLDQIAKILTIYQSSE
 DIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQI
 AIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYG
 LPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIIRTGKENAKYLIE
 KIKLHDMQEGKCLYSLEAIPLEDLLNPNFYEVVDHIIPRSVSFDNSFNKVL
 VKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAGKGRISKTKKEYL
 LEERDINRFSVQKDFINRNLVDTRYATRGLMNLRSYFRVNNLDVKVKSING

[0356]

GFTSFLRRKWKFKKERNKGYKHHAEDALIINANADFIFKEWKKLDKAKKVM
 ENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
 RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYH
 HDPQTYQKLKLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNNGPVIKKIK
 YYGNKLNAHLDDYDYPNSRNKVVKLSLKPFRFDVYLDNGVYKFVTVKN
 LDVIKKENYYEVNSKCYEEAKKLLKISNQAEFIASFYNNDLIKINGELYRVI
 GVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDIL
 GNLYEVKSKKHPQIIKKG

[0357] 在一些实施方案中,Cas核酸酶与具有SEQ ID NO:28的ShCas9或具有SEQ ID NO:27的SpCas9的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性。

[0358] 在一些实施方案中,Cas核酸酶是具有SEQ ID NO:28的ShCas9。在一些实施方案中,Cas核酸酶是具有SEQ ID NO:27的SpCas9。

[0359] 在一些实施方案中,至少一种额外的多肽或蛋白是Cas9核酸酶的变体或功能片段。与所述Cas9的野生型氨基酸序列相比,Cas9变体通常包含一个或多个氨基酸取代。在一些实施方案中,Cas9变体是人源化Cas9 (hCas9) 或其功能片段。如本文所用,术语“人源化Cas9”或“hCas9”是指用于在人类细胞中表达的优化的Cas9蛋白序列。在一些实施方案中,hCas9具有与SEQ ID NO:33的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的氨基酸序列。在一些实施方案中,hCas9具有由SEQ ID NO:33组成的氨基酸序列。

[0360]

SEQ ID NO: 33

MDKKYSIGLDIGTNSVGVAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
 LFDSETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEE
 SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
 LALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD
 AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAE
 DAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
 TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI

[0361]

DGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIH
 LGELHAILRRQEDFYPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTR
 KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
 VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL
 TLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
 QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
 NLAGSPAIKKGILQTVKVVDELVKVMGRHHPENIVIEMARENQTTQKGQK
 NSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
 ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
 KMKNYWRQLLNAKLITQRKFDNLTKAERGGELSELDKAGFIKRQLVETRQIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIN
 NYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI
 GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
 VRKVLSPQVNVKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
 GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEA
 KGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANL
 DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVLDTLIHQISITGLYETRIDLSQLGGD

[0362] 在一些实施方案中,所述至少一种额外的多肽或蛋白是CasX。在一些实施方案中,CasX具有与SEQ ID NO:34的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的氨基酸序列。在一些实施方案中,CasX具有由SEQ ID NO:34组成的氨基酸序列。

[0363]

SEQ ID NO: 34
 MAPKKKRKVSMEIKRINKIRRRRLVKDSNTKKAGKTGPMKTLVLRVMTDP
 LRERLENLRKKPENIPQISNTSRANLNKLLTDYTEMKKAILHVYWEFQK
 DPVGLMSRVAQPAPKNIDQRKLIQVVDGNERLTSSGFACSCCQPLYVYKLE
 QVNDKGGKPHNTNYFGRCNVSEHERLILLSPHKPEANDELVTYSLGKFGQRAL

[0364]

DFYSIHVTRESNHPVKPLEQIGGNASCASGPVKGKALSDACMGAVASFLTQYQ
 DIILEHQKVIKKNEKRLANLKDASANGLAFPKITLPPQPHTKEGIEAYNNVV
 AQIVIWVNLNLWQKLKIGRDEAKPLQRLKGFPSFPLVERQANEVDWWDW
 VCNVKKLINEKKEDGKVFVWQNLAGYKRQEALLPYLSSEEDRKKGKFFAR
 YQFGDLLLHLEKKHGEDWGKVYDEAWERIDKKVEGLSKHIKLEEERSED
 AQSKAALTDWLRKASVIEGLKEADKDEFRCCELKLQKWYGDLRGKPF
 IEAENSILDISGFSKQYNCAFIWQKDGVKLNLYLIINYFKGGKLRFKKIKPE
 AFEANRFYTVINKKSGEIVPMEVNFNFDDPNLILPLAFGKRQGREFIWDL
 LSLETGSLKLANGRVIEKTLYNRRTRQDEPALFVALTFERREVLDSSNIKPMN
 LIGIDRGENIPAVIALTDPEGCPLSRFKDSLGNPTHILRIGESYKEKQRTIQAA
 KEVEQRRAGGYSRKYASKAKNLADDMVRNTARDLLYYAVTQDAMLIFEN
 LSRGFGRQGKRTFMAERQYTRMEDWLTAKLAYEGLPSKTYLSKTLAQYTS
 KTCSNCGFTITSADYDRVLEKLLKKTATGWMTTINGKELKVEGQITYYNRYK
 RQNVVKDLSVELDRLSEESVNNDISSWTKGRSGEALSLLKKRFSHRPVQEK
 FVCLNCGFETHADEQAALNIARSWLFLRSQEYKQYQTNKTTGNTDKRAV
 ETWQSFYRKKLKEVWKPAVTS

[0365] 在一些实施方案中,至少一种额外的多肽或蛋白是失活Cas9蛋白(dCas9)。在一些实施方案中,dCas9具有与SEQ ID NO:35的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的氨基酸序列。在一些实施方案中,dCas9具有由SEQ ID NO:35组成的氨基酸序列。

[0366]

SEQ ID NO: 35

MDKKYSIGLAIGTNSVGVAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
 LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL
 SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
 LALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVD
 AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAE
 DAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
 TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI

[0367]

DGGASQEEFYKFIKPILEKMDGTEELLVVLNREDLLRKQRTFDNGSIPHQIH
 LGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTR
 KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
 VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
 KKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTL
 TLFEDREMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
 QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
 NLAGSPAIKKGILQTVKVVDELVKVMGRHHPENIVIAMARENQTTQKGQK
 NSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
 ELDINRLSDYDVAIVPQSFLKDDSIDNKVLRTR

[0368] 在一些实施方案中,至少一种额外的多肽或蛋白是TnpB(来自转座子PsiTn554的转座酶B)。在一些实施方案中,TnpB具有与SEQ ID NO:36的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的氨基酸序列。在一些实施方案中,TnpB具有由SEQ ID NO:36组成的氨基酸序列。

[0369]

SEQ ID NO: 36

MIRNKAFVVRLYPNAAQTELINRTLGSARFVYNHFLARRIAAYKESGKGLT
 YGQTSSSELTLKQAEETSWLSEVDKFALQNSLKNLETAYKNFFRTVKQSGK
 KVGFPFRFRKKRTGESYRTQFTNNNIQIGEGRLKLPKLGWVVKTKGQQDIQGG
 ILNVTVRRRIHEGHYEASVLCEVEIPYLPAPKFAAGVDVGIKDFAIVTDGVR
 FKHEQNPKYRSTLKRLRKAQQTLSRRKKGSARYGKAKTKLARIHKRIVN
 KRQDFLHKLTTSLVREYEIIGTEHLKPDNMRKNRRLALSISDAGWGEFIRQL
 EYKAAWYGRLVSKVSPYFPSSQLCHDCGFKNPEVKNLAVRTWTCPCGET
 HDRDENAALNIRREALVAAGISDTLNAHGGYVRPASAGNGLRSENHATLVV

[0370] 在一些实施方案中,至少一种额外的多肽或蛋白是Cas12f。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:37的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的氨基酸序列。

[0371] 在一些实施方案中,Cas12f具有由SEQ ID NO:37组成的氨基酸序列。

SEQ ID NO: 37

[0372]

MIKVYRYEIVKPLDLDWKEFGTILRQLQQETRFALNKATQLAWEWMGFSS
 DYKDNHGEYPKSKDILGYTNVHGYAYHTIKTKAYRLNSGNLSQTIKRATDR
 FKAYQKEILRGDMSIPSYKRDIPDLIKENISVNRNMNHGDYIASLSLLSNPAK
 QEMNVKRKISVIIIVRGAGKTIMDRILSGEYQVSASQIIHDDRKNKWYLNIS
 YDFEPQTRVLDLNKIMGIDLGVAVAVYMAFQHTPARYKLEGGEIENFRRQV
 ESRRISMLRQGGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSTRYI
 VDMAIKEGCGTIQMEDLTNIRDIGSRFLQNWTTYDLQKIIYKAEEAGIKVI
 KIDPQYTSQRCSECGNIDSGNRIGQAIKFCRACGYEANADYNAARNIAIPNI
 DKIIAESIK

[0373] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽是包含转座酶或其片段和如上所述的核酸酶的融合蛋白。

[0374] 在一些实施方案中,融合蛋白与SEQ ID NO:2具有至少75%、80%、85%、90%、95%、96%、97%、98%、99%或更高的氨基酸序列同一性。在一些实施方案中,融合蛋白包含SEQ ID NO:2的氨基酸序列或由SEQ ID NO:2的氨基酸序列组成。

SEQ ID NO: 2

[0375]

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
 LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLLEE
 SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
 LALAHMIKFRGHFLIEGDLNPDNSVDKLFIQLVQTYNQLFEENPINASGVD
 AKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAE
 DAKLQLSKDXYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI
 TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYI
 DGGASQEEFYKFIKPILEKMDGTEELLVKNREDLLRKQRTFDNGSIPHQIH
 LGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTR
 KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT
 VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF

[0376]

KKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTL
 TLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
 QSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA
 NLAGSPAIKKGIHQTVKVVDELVKVMGRHHPENIVIAMARENQTTQKGQK
 NSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQ
 ELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
 KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
 KHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVIN
 NYHHAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEI
 GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT
 VRKVLSPQVNVIVKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYG
 GFDSPTVAYSVLVVAKVEKGGKSKKLKSVKELLGITIMERSSSFENPIDFLEA
 KGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSKYVNF
 LYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANL
 DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVLDTLIHQSTGLYETRIDLSQLGGSGSETPGTSESATPESGSSLDDEHIL
 SALLQSDDELVGEDSDSEVSDHVSEDDVQSDTEEAFIDEVHEVQPTSSGSEI
 LDEQNVIEQPGSSLASNRILTLPQRTIRGKNKHCWSTSKPTRRSRVSALNIVR
 SQRGPTRMCRNIYDPLLCKLFFTDEIIEIVKWTNAEISLKRRESMTSATFR
 DTNEDEIYAFFGILVMTAVRKDNHMSTDDLDFDRSLSMVYVSVMSRDRDFDL
 IRCLRMDKKSIRPTLRENDVFTPVRKIWDLFIHQCIQNYTPGAHLTIDEQLLG
 FRGRCPFRVYIPNKPSKYGIKILMMCDSGTKYMINGMPYLGRGTQTNGVPL
 GEYYVKELSKPVHGSRNITCDNWFTSIPLAKNLLQEPYKLTIVGTVASNAR
 EIPEVLKNSRSPVGTSMFCFDGPLTLVSYKPKPAKMVYLLSSCEDEDASINE
 STGKPMVMYYNQTKGGVDTLNQMCVMTCSRKTNRWPMALLYGMINI
 ACINSFIIYSHNVSSKGEKVQSRKKFMRNLYMGLTSSFMRKRLEAPTLKRYL
 RDNISNILPKEVPGTSDDSTEPPVMKKRTYCTYCPKIRRKASASCKKCKKV
 ICREHNIDMCQSCFSRADPKKKRKRKVR

[0377] 在一些实施方案中,至少一种额外的多肽或蛋白是适体结合蛋白。

[0378] 在一些实施方案中,至少一种额外的多肽或蛋白是选自包含以下或由以下组成的组的适体结合蛋白:MS2噬菌体包被蛋白(MCP)、PP7包被蛋白(PCP)、λN22肽和COM。

[0379] 在一些实施方案中,至少一种额外的多肽或蛋白是MS2噬菌体包被蛋白(MCP)。在一些实施方案中,MCP具有与SEQ ID NO:39的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的氨基酸序列。

SEQ ID NO: 39

[0380]

MASNFTQFVLVDNNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVR
 QSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDC
 ELIVKAMQGLLKDGNPIPSAIAANSIY

[0381] MCP能够结合MS2 RNA四环结合序列(或“MS2适体”)。在一些实施方案中,MS2适体具有与SEQ ID NO:40的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的核酸序列。

SEQ ID NO: 40

[0382]

ggggccacuagggacaggauguuuuagagcuagggcaacaugaggauccaccaugucugcagggccuagca
 aguuuuuuuagggcuaguccguuuaucaacuuggccaacaugaggauccaccaugucugcagggccaagugg
 caccgagucggugcuuuuuu

[0383] 在一些实施方案中,融合蛋白通过MCP与包含至少一个MS2适体的指导RNA (gRNA) 分子共价或非共价相互作用或能够共价或非共价相互作用。

[0384] 因此,本发明的方法还可以包括使细胞群与融合至至少一种适体的指导RNA (gRNA) 接触。

[0385] 在一些实施方案中,适体是包含四环的RNA序列。术语“四环”是指四碱基发夹环基序。该术语在本文中术语“茎环”或“发夹环”可互换使用。

[0386] 在一些实施方案中,适体是MS2 RNA四环序列(或MS2适体)。在一些实施方案中,MS2适体具有与SEQ ID NO:40的氨基酸序列具有至少80%、85%、90%、95%、99%或更高的氨基酸序列同一性的核酸序列。在一些实施方案中,适体包含具有SEQ ID NO:40的核酸序列或由具有SEQ ID NO:40的核酸序列组成。

[0387] 在一些实施方案中,gRNA能够与如上所述的RNA指导的核酸酶形成复合物,例如Cas蛋白或包含Cas蛋白的融合蛋白。在一些实施方案中,gRNA分子通过至少一个MS2适体与RNA指导的核酸酶共价或非共价相互作用,或能够共价或非共价相互作用。

[0388] 在一些实施方案中,gRNA能够将如上所述的RNA指导的核酸酶靶向至细胞(优选哺乳动物细胞,更优选人类细胞)的基因组的特定序列或区域。在一些实施方案中,gRNA靶向的特定序列与特异于Cas蛋白的原间隔区相邻基序(PAM)相邻。

[0389] 在一些实施方案中,gRNA靶向的特异性序列可能位于细胞基因组中的安全港基因座内。“安全港基因座”是指细胞基因组的某个区域,其中整合的物质可以充分表达,而不会扰乱内源基因的结构或功能。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由以下组成的组: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;或白蛋白基因的内含子1。安全港基因座也可以包括缺乏内源基因并具有开放的染色质的基因组区域,其允许插入的转基因表达,而不会扰乱基因组的结构或功能。

[0390] 在一些实施方案中,gRNA包含20、25、30、35、40、45、50个或更多个核苷酸。

[0391] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽以蛋白的形式递送至细胞。

[0392] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽以编码所述转座酶或其片段的核酸的形式递送至细胞。核酸可以是核糖核酸(RNA)或脱氧核糖核酸(DNA)。

[0393] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是RNA分子。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是信使RNA(mRNA)分子。

[0394] 或者,编码包含转座酶或其片段的蛋白或多肽的核酸可以是DNA分子。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是互补DNA(cDNA)。在一些实施方案中,DNA分子可以包含在载体内,例如质粒。

[0395] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是线性的或环状的。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸是双链或单链的。

[0396] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸通过转染或转化递送至细胞群。

[0397] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸通过转染递送至细胞群。转染技术的实例包括但不限于脂质转染、电穿孔、超声处理、纳米颗粒、显微注射和病毒载体感染,包括非整合型和整合型病毒载体感染。

[0398] 在一些实施方案中,转染是通过脂质转染进行的。进行脂质转染的方法在本部分中是已知的并且包括使用脂质转染试剂,例如lipofectamine™。在一些实施方案中,转染是通过电穿孔进行的。在一些实施方案中,转染是通过超声处理进行的。在一些实施方案中,转染是通过纳米颗粒(例如,聚合物纳米颗粒,如JetPEI)进行的。在一些实施方案中,转染是通过显微注射进行的。在一些实施方案中,转染是通过病毒载体感染进行的。在一些实施方案中,病毒载体是整合型或非整合型,优选非整合型。非整合型病毒载体的非限制性实例包括腺病毒载体、腺相关病毒(AAV)载体、痘病毒载体、单纯疱疹病毒载体等。

[0399] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸通过转化递送至细胞群。转化技术的实例包括但不限于使用选自包含以下或由以下组成的组的整合型载体:整合型病毒载体、整合型质粒、酶或基因组编辑方法。在一些实施方案中,转化包括使用整合型病毒载体或修饰的整合型病毒。

[0400] 在一些实施方案中,整合型病毒载体或修饰的整合型病毒是整合型病毒或其变体或其突变体,其选自包含以下或由以下组成的组:逆转录病毒科(Retroviridae)、腺病毒科(Adenoviridae)、黄病毒科(Flaviviridae)、疱疹病毒科(Herpesviridae)、嗜肝DNA病毒科(Hepadnaviridae)、乳头瘤病毒科(Papillomaviridae)、多瘤病毒科(Polyomaviridae)、细小病毒科(Parvoviridae)、沙粒病毒科(Arenaviridae)、博尔纳病毒科(Bornaviridae)、布尼亚病毒科(Bunyaviridae)、丝状病毒科(Filoviridae)和副黏病毒科(Paramyxoviridae)的病毒,优选地选自包含逆转录病毒科、腺病毒科和黄病毒科的组的病毒或由逆转录病毒科、腺病毒科和黄病毒科的病毒组成的组的病毒。在一些实施方案中,整合型病毒载体属于逆转录病毒科。在一些实施方案中,整合型病毒载体是慢病毒或修饰的慢病毒。

[0401] 在一些实施方案中,将编码包含转座酶或其片段的蛋白或多肽的核酸与编码至少一种目的转基因的核酸分子一起递送至细胞群。在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸和编码至少一种目的转基因的核酸分子在同一载体中递送。或

者,编码包含转座酶或其片段的蛋白或多肽的核酸以及编码至少一种目的转基因的核酸分子在不同的载体中同时递送。

[0402] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前递送至细胞群,例如在编码至少一种目的转基因的核酸分子之前1分钟、2分钟、3分钟、4分钟、5分钟、10分钟、30分钟、60分钟、2小时、3小时、4小时、5小时、6小时、12小时、24小时、36小时、48小时、60小时、72小时、4天、5天、6天、7天或更长时间。

[0403] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前1小时到72小时递送至细胞群。

[0404] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前2小时至72小时、3小时至72小时、4小时至72小时、5小时至72小时、6小时至72小时、7小时至72小时、8小时至72小时、9小时至72小时、10小时至72小时、11小时至72小时、12小时至72小时、24小时至72小时、36小时至72小时、48小时至72小时递送至细胞群。

[0405] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前1小时至60小时、1小时至48小时、1小时至36小时、1小时至24小时、1小时至12小时、1小时至11小时、1小时至10小时、1小时至9小时、1小时至8小时、1小时至7小时、1小时至6小时、1小时至5小时、1小时至4小时、1小时至3小时、1小时至2小时递送至细胞群。

[0406] 在一些实施方案中,将编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前1小时至12小时、2小时至8小时、3小时至6小时递送至细胞群。

[0407] 在一些实施方案中,编码包含转座酶或其片段的蛋白或多肽的核酸在编码至少一种目的转基因的核酸分子之前约4小时递送至细胞群。

[0408] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸包含在药物组合物中,所述药物组合物还包含至少一种药学上可接受的赋形剂。

[0409] 如本文所用,术语“药学上可接受的赋形剂”是指当施用于动物(优选人类)或细胞群时不会产生不良、过敏或其他不想要反应的赋形剂。它包括任何和所有溶剂、分散介质、包衣、抗菌剂和抗真菌剂、等渗剂和吸收延缓剂等。可接受的赋形剂是指任何类型的无毒固体、半固体或液体填充剂、稀释剂、包封材料或制剂辅料。对于人类施用,制剂应符合EMA或FDA生物制品标准办公室的无菌性、热原性、一般安全性和纯度标准。

[0410] 在一些实施方案中,可接受的赋形剂选自包含以下或由以下组成的组:溶剂、稀释剂、运载体、分散介质、包衣、抗菌剂、抗真菌剂、等渗剂、吸收延缓剂,及其任意组合。赋形剂在与该方法的蛋白、多肽或核酸分子相容并且在施用于个体或细胞群后不是有害的意义上必须是“可接受的”。通常,赋形剂在被施用于个体(优选人类个体)或细胞群时,不会产生不良、过敏或其他不想要的反应。

[0411] 用于治疗用途的可接受的赋形剂是本领域中众所周知的,并且例如在Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.R.Gennaro ed.1985)中进行了描述。可以根据预期的给药途径和标准实践选择合适的赋形剂。

[0412] 在一些实施方案中,细胞群是真核细胞群或原核细胞群。在一些实施方案中,细胞

群是真核细胞群。在一些实施方案中,细胞群是原核细胞群。

[0413] 在一些实施方案中,细胞群分离自供体。如本文所用,“供体”是指动物,优选哺乳动物,更优选人类。当细胞群被分离时,供体可以是活的或死的,优选是活的。在一些实施方案中,细胞群是从供体的至少一个器官或组织中分离的,任选地是从供体的多于一个器官或组织中分离的。在一些实施方案中,细胞群是同质的(即,它含有单一类型的细胞)或异质的(即,它含有多于一种类型的细胞),优选是同质的。

[0414] 在一些实施方案中,细胞群在体外培养。在一些实施方案中,细胞群包含在原代细胞培养物中。

[0415] 在一些实施方案中,细胞群来自永生化细胞系。

[0416] 在一些实施方案中,该方法是在体外进行的。在一些实施方案中,细胞群是体外培养的细胞群。

[0417] 体外培养细胞的方法是本领域中众所周知的并且是常规实践。通常,细胞在受控的气氛(例如37°C、5% CO₂)中维持,并在合适的培养基(例如杜氏改良的Eagle培养基)中维持。

[0418] 在一些实施方案中,将细胞群培养2小时至72小时、3小时至72小时、4小时至72小时、5小时至72小时、6小时至72小时、7小时至72小时、8小时至72小时、9小时至72小时、10小时至72小时、11小时至72小时、12小时至72小时、24小时至72小时、36小时至72小时、48小时至72小时。在一些实施方案中,将细胞群培养1小时至60小时、1小时至48小时、1小时至36小时、1小时至24小时、1小时至12小时、1小时至11小时、1小时至10小时、1小时至9小时、1小时至8小时、1小时至7小时、1小时至6小时、1小时至5小时、1小时至4小时、1小时至3小时、1小时至2小时。

[0419] 在一个实施方案中,细胞群体是非分裂、非复制、终末分化或静止细胞群。术语“非分裂细胞群”、“非复制细胞群”、“终末分化”和“静止细胞群”在本文中可互换使用。

[0420] 如本文所用,“非分裂”细胞、“非复制”细胞或“静止”细胞意指细胞仍处于细胞周期之外,但保留分裂能力,换句话说,细胞处于可逆性生长停滞的状态。静止可以由非模仿性的接触抑制、化学或药理学药剂、信号传导蛋白、激素、抑制剂等诱导,或者由缺乏信号(例如生长因子)或与一种或多种细胞类型的接触诱导。在一些实施方案中,在培养基中存在一种或多种生长抑制剂的情况下,细胞群是静止的。在一些实施方案中,一种或多种生长抑制剂选自包含以下或由以下组成的组:化学药剂、药理学药剂、信号传导蛋白、激素、生长因子和抑制剂。在一些实施方案中,在培养基中不存在一种或多种生长刺激剂时,细胞群是静止的。在一些实施方案中,一种或多种生长刺激剂选自包含以下或由以下组成的组:化学药剂、药理学药剂、信号传导蛋白、激素、生长因子、氨基酸、糖、脂质和脂肪酸。如本文所用,“终末分化”细胞是指细胞脱离细胞周期,失去分裂能力,换句话说,细胞处于不可逆的生长停滞状态。

[0421] 在一些实施方案中,非分裂或静止细胞群是汇合细胞培养物(即,静止是由接触抑制诱导的)。

[0422] 在一些实施方案中,非分裂或静止细胞群由至少一种在培养中不复制的细胞类型组成。

[0423] 在一些实施方案中,非分裂或静止细胞群是衰老的。

[0424] 在一些实施方案中,非分裂或终末分化细胞群由至少一种已完全丧失进行细胞分裂能力的细胞类型组成。

[0425] 在一些实施方案中,细胞群是非主动分裂的。

[0426] 在另一个实施方案中,细胞群是主动分裂的、复制的或增殖的。在一些实施方案中,分裂细胞群处于细胞培养的指数期。

[0427] 在一些实施方案中,该方法是在离体进行的。

[0428] 在一些实施方案中,该方法是在体内进行的。在一些实施方案中,该方法在动物受试者(优选哺乳动物受试者)体内进行。在一些实施方案中,该方法是在人类受试者体内进行的。在一些实施方案中,细胞群包含在活生物体的组织或器官中。在一些实施方案中,生物体是动物。在一些实施方案中,生物体是哺乳动物。在一些实施方案中,生物体是人类。

[0429] 本发明还涉及包含至少一个拷贝的含有转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸、和/或至少一个拷贝的至少一种目的转基因的细胞或细胞群。在一些实施方案中,细胞或细胞群包含至少一个拷贝的含有转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸,以及至少一个拷贝的至少一种目的转基因。

[0430] 在一些实施方案中,包含转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸、至少一种目的转基因以及细胞或细胞群还包含如上文所公开的特征。

[0431] 本发明还涉及一种治疗有需要的受试者的遗传疾病的方法,包括向所述受试者施用:

[0432] -如上文所述的编码至少一种目的转基因的核酸分子,以及

[0433] -如上文所述的包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸。

[0434] 本发明还涉及如上文所述的编码至少一种目的转基因的核酸分子,以及如上文所述的包含转座酶或其片段的蛋白或多肽,或编码该蛋白或多肽的核酸,其用于治疗有需要的受试者的遗传疾病。

[0435] 本发明还涉及如上文所述的编码至少一种目的转基因的核酸分子、以及如上文所述的包含转座酶或其片段的蛋白或多肽、或编码该蛋白或多肽的核酸在制造用于治疗有需要的受试者的遗传疾病的药物中的用途。

[0436] 在一些实施方案中,目的转基因补偿了检测出导致受试者遗传疾病的基因。

[0437] 在一些实施方案中,该方法还包括向受试者施用如上所述的与至少一种适体融合的指导RNA(gRNA)。

[0438] 在一些实施方案中,该方法可以包括向有需要的受试者施用至少一种额外的治疗剂。在一些实施方案中,至少一种治疗剂用于治疗遗传疾病。

[0439] 在一些实施方案中,遗传疾病的特征在于,在受试者的基因组中至少一个基因发生突变,从而该至少一个基因编码的蛋白的功能受损或无功能,或被细胞蛋白质量控制(例如蛋白酶体)降解,或不产生;换句话说,至少一个基因的功能至少部分(如果不是完全)丧失。

[0440] 遗传疾病的非限制性实例包括镰状细胞性贫血(sickle cell anemia)、囊性纤维化(cystic fibrosis)、亨廷顿病(Huntington disease)、先天性肌营养不良症(congenital muscular dystrophy)、杜氏肌营养不良症(Duchenne muscular dystrophy)、法布里病(Fabry disease)、马凡氏综合征(Marfan syndrome)、地中海贫血

(thalassemia)、胱氨酸病(cystinosis)、家族性高胆固醇血症(familial hypercholesterolemia)、血色素沉着症(hemochromatosis)等。

[0441] 在一些实施方案中,目的转基因部分或完全(优选完全)恢复突变的至少一个基因的功能。在一些实施方案中,目的转基因与突变的至少一个基因编码相同的蛋白。在一些实施方案中,目的转基因编码的蛋白与突变的至少一个基因编码的蛋白不同,但具有相似或相同的生物学功能。

[0442] 附图简要说明

[0443] 图1是一组照片,举例说明了表达GFP的转座子和转座酶在用转座酶(单独的hyPB或与Cas9核酸酶融合的hyPB)电穿孔的胚胎干细胞H9中的活力和转染效率。BF:明视野;GFP:绿色荧光蛋白。

[0444] 图2是显示表达或不表达hyPB的HepG2细胞中红色荧光蛋白(RFP)报告基因的转录效率的直方图。

[0445] 图3是显示表达或不表达hyPB的Huh7细胞中RFP报告基因的转录效率的直方图。

[0446] 图4和图5是显示表达或不表达hyPB的Huh7细胞中GFP报告基因随时间(图4中为3小时、6小时、9小时和12小时;图5中为8小时、10小时、12小时、14小时和16小时)的转录效率的图。

[0447] 图6是显示表达或不表达催化失活hyPB的Huh7细胞中GFP报告基因的转录效率的直方图。

[0448] 图7是显示在表达包含Cas9核酸酶和SB100转座酶或N57转座酶的融合蛋白的Huh7细胞中GFP报告基因的转录效率的直方图。

[0449] 图8A-8B是比较表达或不表达hyPB的Huh7细胞中GFP报告基因的转录效率的一组直方图,其中GFP DNA转座子包含一个末端反向重复(ITR)(图8A)或两个ITR(图8B)。

[0450] 图9是显示用hyPB mRNA或催化失活的hyPB mRNA注射后一天或四周,小鼠体内荧光素酶报告基因的转录效率的一组照片。

[0451] 图10是显示使用聚合物纳米颗粒载体(jetPEI)在用hyPB DNA转染或未转染的小鼠中注射后一天或一周体内荧光素酶报告基因的转录效率的一组照片。

[0452] 图11是显示在用或未用hyPB转染的小鼠中注射一天后体内荧光素酶报告基因的转录效率的一组照片,其中DNA有效载荷(荧光素酶报告基因)和hyPB mRNA使用相同载体(脂质纳米颗粒)共转染。

[0453] 图12是直方图,显示用hyPB转染的非分裂NIH-3T3细胞与不表达hyPB的细胞相比,具有更高的DNA转座子摄取和表达能力。结果表示为平均值 \pm s.e.m, $n=4$ 。

[0454] 图13A-13B是一组显示转染后20小时间隔内测量的复制(图13A)和非分裂(图13B)NIH/3T3细胞中核DNA易位的图表。结果表示为平均值 \pm s.e.m, $n=2$ 。统计学显著性表示为***(p 值 <0.001)和**(p 值 <0.005)。

[0455] 图14是显示用hyPB转染的培养物中非分裂原代小鼠肝细胞与不表达hyPB的细胞相比具有更高的DNA荧光素酶编码转座子摄取和表达能力的直方图。数据表示为平均值 \pm SEM, $n=3$ 。

实施例

[0456] 通过下列实施例进一步说明本发明。

[0457] 实施例1:转座酶驱动的核心定位

[0458] 材料和方法

[0459] H9干细胞的电穿孔

[0460] 单独使用1.8 μ g微环GFP转座子或与1 μ g hyPB mRNA或1.5 μ g FiCAT mRNA和1.75 μ g 靶向AAVS1基因座的sgRNA一起使用1.8 μ g微环GFP转座子对胚胎干细胞H9进行电穿孔(CRG细胞和组织设施)。使用荧光显微镜并使用明视野对细胞进行成像,估算活力和GFP信号的百分比。

[0461] HepG2细胞的脂质转染

[0462] 第0天:将400k个细胞/孔接种于p6孔板中。

[0463] 第1天:根据制造商的方案,用lipofectamine 3000转染1.25 μ g hyPB质粒或puc19模拟质粒。将细胞与Optimem+转染混合物一起保存2小时,然后添加完全培养基(杜氏改良的Eagle培养基10%胎牛血清、2mM谷氨酰胺和100U青霉素/0.1mg/ml链霉素)

[0464] 第2天:将转染的细胞取出,以150k个细胞/孔的密度接种到p12孔板中

[0465] 第3天:按照制造商的方案用0.75 μ g RFP DNA转座子和0.25 μ g puc19填充DNA和lipofectamine 3000转染细胞。将细胞与Optimem+转染混合物一起保存2小时,然后添加完全培养基。

[0466] 第4天:第4天转染后24小时测量荧光

[0467] 或者,在第1天用Cas9_SB100质粒和Cas9_N57质粒代替hyPB转染细胞。在第3天,用含有SB100 ITR的GFP转座子和TCR1 gRNA质粒转染细胞。

[0468] Huh7细胞的脂质转染

[0469] 转染前24小时将400,000个细胞接种于p6孔中。根据制造商的方案,使用1.8 μ g转座子DNA和1.95 μ g hyPB或大小相似的模拟mRNA用lipofectamine 3000进行转染。将细胞与Optimem+转染混合物一起保存2小时,然后添加完全培养基。转染后每3小时或每14小时提取细胞一次,并用流式细胞术测量荧光。

[0470] 或者,在第1天用Cas9_SB100质粒和Cas9_N57质粒代替hyPB转染细胞。在第3天,用含有SB100 ITR的GFP转座子以及用TCR1 gRNA质粒转染细胞。

[0471] 体内流体动力学注射

[0472] 给6-7周龄小鼠注射了总计10至10.2 μ g的核酸(3.2 μ g MC-荧光素酶转座子,1.42 μ g hyPB/hyPB_dead mRNA)。用PBS稀释核酸,并经由眼眶后全身注射在不到7秒的时间内注射动物体重的7% (按ml计)。

[0473] JetPEI体内转染

[0474] 采用DNA有效载荷和hyPB并采用不同的递送方法:根据制造商的说明,使用体内JetPEI将35.6 μ g MC-荧光素酶DNA转座子以N/P比率7静脉内注射到5周龄小鼠。静脉内注射后4小时注射载有2.5 μ g hyPB mRNA的LNP。采用Benchtop Nanossembler (Precision Nanosystems) 进行LNP制造。

[0475] 采用DNA有效载荷和hyPB并采用相同的递送方法:将15 μ g载有MC-荧光素酶DNA转座子的LNP静脉内、与或不与2.5 μ g载有hyPB的LNP注射到5周龄小鼠。使用Benchtop

Nanossemblr (Precision Nanosystems) 进行LNP制造。

[0476] 脂质比例为50% Dlin-MC3、10% DSPC辅助脂质、38.5%胆固醇、1.5%PEG-2000和N/P比为6。纳米颗粒的尺寸为100-130nm,包封率高于90%。

[0477] 全身成像

[0478] 使用IVIS光谱成像系统(Caliper Life Sciences)对FiCAT-gRNA转座子或转座子对照给药后不同时间点的荧光素酶表达进行全身成像。根据制造商的说明,在腹膜内注射D-荧光素钾盐(Gold Biotechnology)后5分钟拍摄图像。FiCAT:Cas9-hyPB融合。

[0479] 结果

[0480] 用转座酶(hyPB或FiCAT,FiCAT为Cas9-hyPB融合蛋白)电穿孔的胚胎干细胞H9显示GFP信号(图1),这意味着GFP转座子DNA已进入细胞核并已被翻译和转录。虽然不使用hyPB或FiCAT的GFP转座子的附加体信号较低(效率为20%),但hyPB产生更强的GFP信号(转染效率为70%)。

[0481] 先前表达hyPB的HepG2肝细胞系比不表达hyPB的细胞表现出更好的RFP DNA转座子的内化(图2)。利用Huh7肝细胞系也获得了这一结果(图3)。此外,与用模拟mRNA转染的细胞相比,用hyPB mRNA转染的Huh7细胞表现出更高的摄取和表达DNA转座子的能力(图4和5)。

[0482] 与用模拟mRNA转染的细胞相比,用hyPB催化失活的mRNA转染的Huh7肝细胞系表现出更高的核摄取和表达DNA转座子的能力(图6)。

[0483] 有趣的是,显示与不使用hyPB的细胞相比hyPB转座酶对转座子的核摄取能力更高的上述结果可以推广到其他转座酶,如SB100和N57(图7)。

[0484] 然后测试了DNA有效载荷的ITR要求。先前用hyPB表达的Huh7肝细胞系比不表达hyPB的细胞(先前用Puc19质粒DNA转染)表现出更好的GFP DNA转座子内化,这两者都是带有1个ITR或2个ITR的GFP转座子(图8A-8B)。这些数据表明,即使仅存在1个ITR(因此hyPB分子可以结合到其有效载荷中),hyPB也能促进有效载荷DNA的核内化。

[0485] 在体内条件下进一步测试了hyPB改善转染的能力(图9)。hyPB转座酶在DNA有效载荷内化和表达方面表现出优势,即使其催化活性受到损害(失活-hyPB)。在单独有效载荷的流体动力学注射后24小时,3只小鼠中有1只显示出荧光素酶表达,而当与hyPB或催化失活的hyPB共同注射时,这增加到2/3或3/3只小鼠。

[0486] 当使用不同的递送方法施用有效载荷DNA和hyPB mRNA时,hyPB转座酶在DNA有效载荷内化和表达方面表现出优势(图10)。注射DNA时,50%的小鼠在24小时后显示荧光素酶信号,而使用LNP共同施用hyPB时100%的小鼠显示出信号。

[0487] 当使用相同的递送方法施用有效载荷DNA和hyPB mRNA时,hyPB转座酶在DNA有效载荷内化和表达方面表现出优势(图11)。注射DNA solo时,小鼠不会显示荧光素酶信号,而在这两种情况下使用LNP共同施用hyPB时信号都会恢复。这一结果表明hyPB有助于将DNA内化到细胞核中,因此它能够被表达。

[0488] 实施例2:非分裂细胞中的核定位

[0489] 材料和方法

[0490] NIH/3T3培养

[0491] NIH/3T3(ATCC)在含有10%或1%热灭活胎牛血清、100U/ml青霉素和0.1mg/ml链

霉素的DMEM中在37℃加湿CO₂ (5%) 培养箱中培养。

[0492] 对静止NIH-3T3细胞的剂量依赖性效应测定

[0493] 接种60000个NIH-3T3细胞,并将它们在达到汇合后在补充有1% FBS的DMEM中培养48小时以诱导衰老。然后,转染细胞。根据制造商的方案,使用单独的0.2ug转座子DNA (附加体) 或与指定量的hyPB-wt或催化失活hyPB (失活-hyPB) 组合用lipofectamine 3000一起进行转染。将细胞与在Optimem中的转染混合物一起孵育过夜,然后添加完全培养基。DMEM添加后1小时,对细胞进行胰蛋白酶消化,并通过流式细胞术评估GFP表达。(表示为平均值±SEM, n=4)

[0494] 在分裂和非分裂NIH-3T3细胞中转染的比较

[0495] 在复制细胞中,转染前当天接种了60000个细胞。对于非分裂的3T3,一旦它们达到汇合,就将其在补充有1% FBS的DMEM中培养48小时以诱导衰老。然后,转染细胞。根据制造商的方案,使用单独的0.25ug转座子DNA (附加体) 或与0.25ug hyPB-wt或催化失活hyPB (失活-hyPB) 组合用lipofectamine 3000一起进行转染,并以3小时的间隔进行胰蛋白酶消化,以采用流式细胞术进行GFP评估。

[0496] 使用双向方差分析 (ANOVA) 检验来分析组间差异的统计学意义。

[0497] 原代肝细胞分离

[0498] 使用肝灌注缓冲液 (HBSS KCl 0.4g.L⁻¹, 葡萄糖1g.L⁻¹, NaHCO₃ 2.1g.L⁻¹, EDTA 0.2g.L⁻¹) 对肝脏进行灌注,然后使用肝消化缓冲液 (DMEM-GlutaMAX 1g.L⁻¹葡萄糖, HEPES 15mM pH 7.4, 青霉素/链霉素1%、每只小鼠5mg胶原酶IV (C5138 Sigma)) 进行消化。切除后,将肝脏放在冰上的铺板培养液中 (M199, 胎牛血清10%、青霉素/链霉素1%、丙酮酸钠1%、L-谷氨酰胺1%、1nm胰岛素、1mM地塞米松、2mg.mL⁻¹牛血清白蛋白 (BSA))。使用镊子将组织均质化,然后在铺板培养基中过滤。然后将细胞在铺板培养基中洗涤三次,然后将其铺到铺板培养基中的胶原蛋白包被板 (Thermo Fisher Scientific) 上。第二天早上更换培养基并在转染前将细胞孵育48小时。

[0499] 原代肝细胞的培养和转染

[0500] 从8周龄野生型小鼠中分离原代肝细胞,每孔接种75,000个细胞。培养48小时后,转染细胞。根据制造商的方案,使用单独的0.2ug转座子DNA (附加体) 或与0.25ug hyPB-wt或催化失活hyPB (失活-hyPB) 组合用lipofectamine 3000一起进行转染。将细胞与在Optimem中的转染混合物一起孵育5小时,然后添加完全培养基。转染后24小时,用PBS洗涤细胞一次,并用适当的缓冲液裂解。转染后24小时,使用荧光素酶检测系统 (Promega, Madison, WI, USA) 对荧光素酶活性进行定量,一式两份地进行。数据显示为相对于未转染肝细胞荧光素酶值的倍数变化,并以任意单位 (AU) 表示。

[0501] 采用单向或双向方差分析 (ANOVA) 检验来分析组间差异的统计学意义。在适用的情况下,使用双侧学生t检验来分析成对比较。结果表示为平均值±s.e.m。

[0502] 结果

[0503] 为了进一步支持PiggyBac转座酶等DNA转座酶在DNA转座子核输入中的作用,在非分裂细胞中进行了实验。

[0504] 选择了两种不同的模型:在图12和图13A-13B中,NIH/3T3成纤维细胞系是衰老诱导的,在图14中,从小鼠分离的原代肝细胞 (在培养时不会复制) 用于测试非分裂 (静止) 细

胞中的PiggyBac (hyPB) DNA核易位活性。

[0505] 测试了增加的量的PiggyBac转座酶 (hyPB),以探究其剂量依赖性效应;此外,添加了催化失活版本的转座酶 (失活-hyPB) 作为单独DNA易位活性而没有转座酶的插入活性的对照。图12显示,即使使用最低剂量,在表达hyPB或失活-hyPB的细胞中DNA转座子 (GFP) 的摄取和表达也会增加。

[0506] 此外,还对递送到主动分裂和非分裂的NIH/3T3细胞的细胞质中的DNA有效载荷的转导和表达效率进行了并行比较,以便支持转座酶蛋白在这些有效载荷的核定位中的作用。(图13A-13B)。图13B显示,与没有转座酶的情况相比,在表达hyPB或失活-hyPB的非分裂NIH-3T3细胞中,GFP有效载荷的表达效率显著增加。

[0507] 原代肝细胞的结果也支持该结论,并表明PiggyBac转座酶 (无论是催化活性 (hyPB) 还是无活性 (失活-hyPB)) 的存在均增加了转座子的表达 (图14),这是其核运输改善的结果。

[0508] 总体而言,这些结果证明了:DNA转座酶增加了DNA转座子的核输入,与其插入活性无关,特别是在非分裂细胞中。

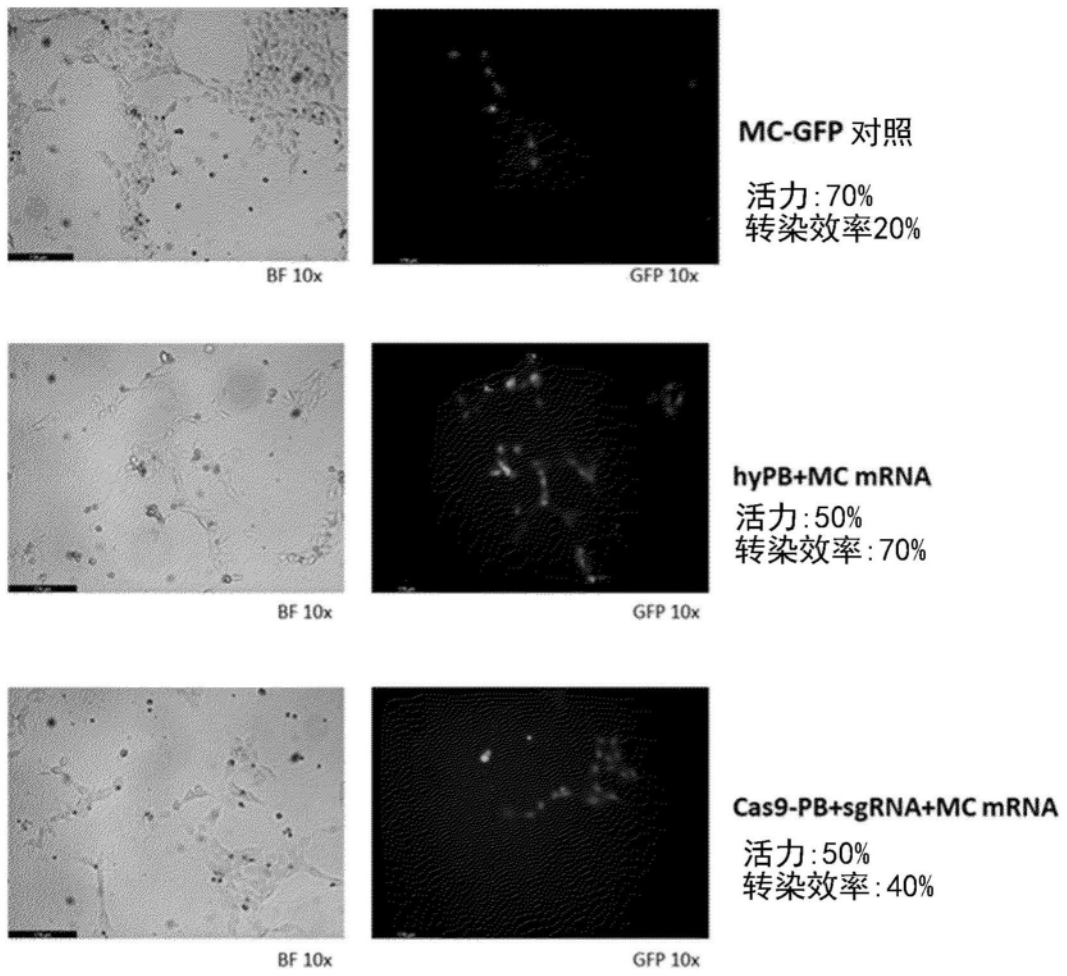


图1

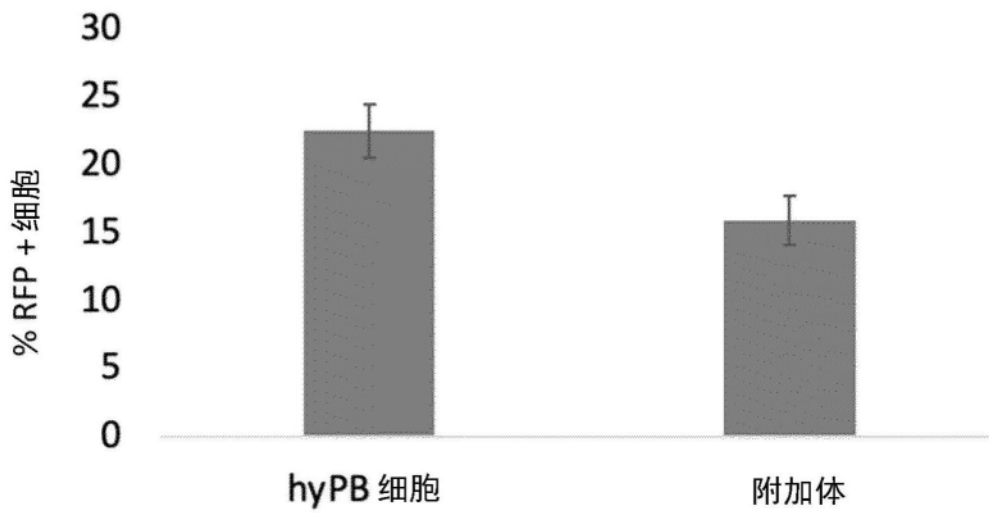


图2

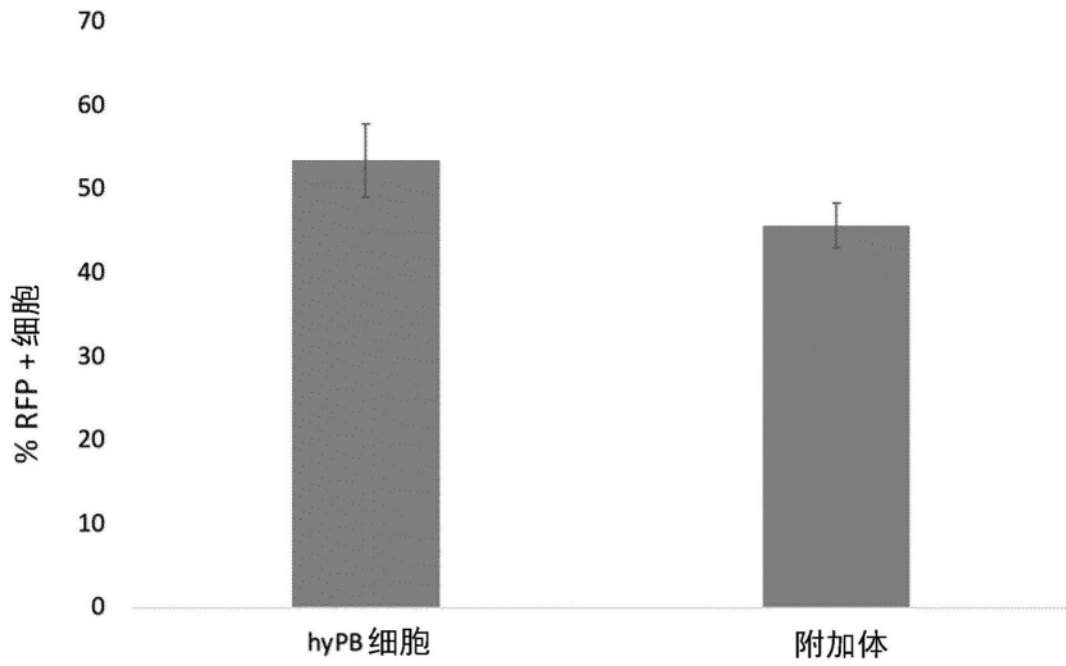


图3

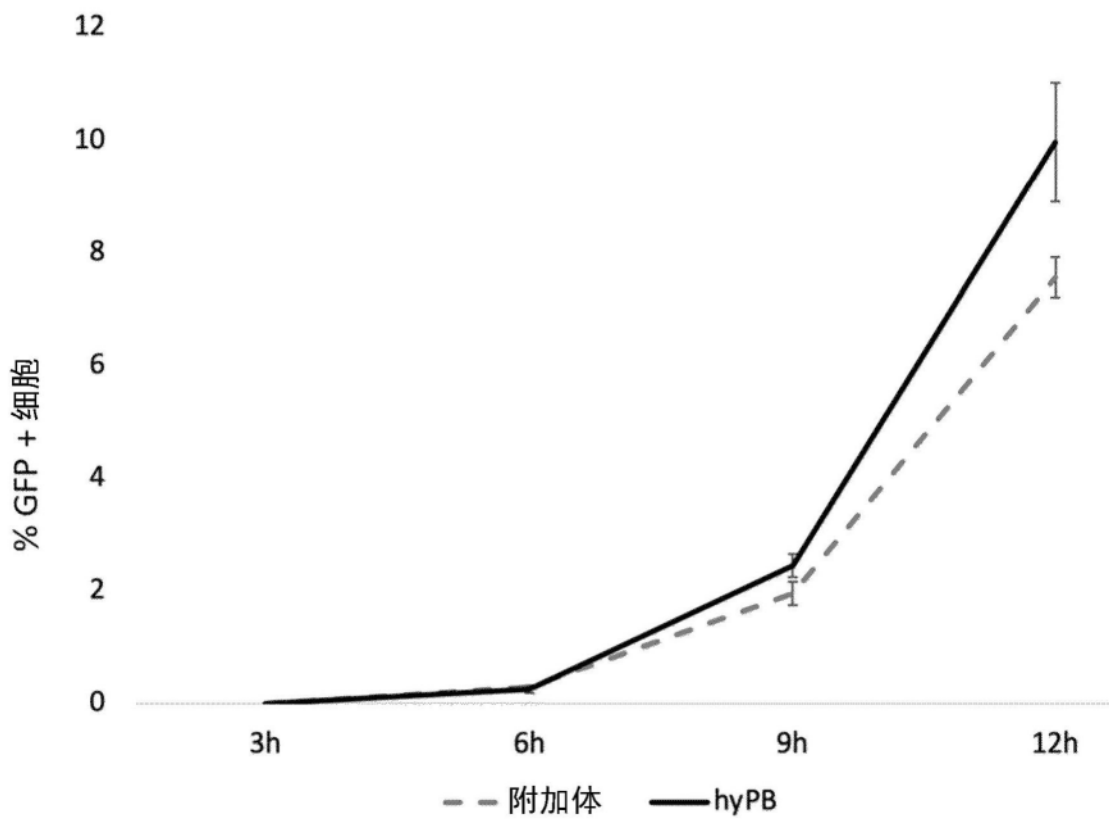


图4

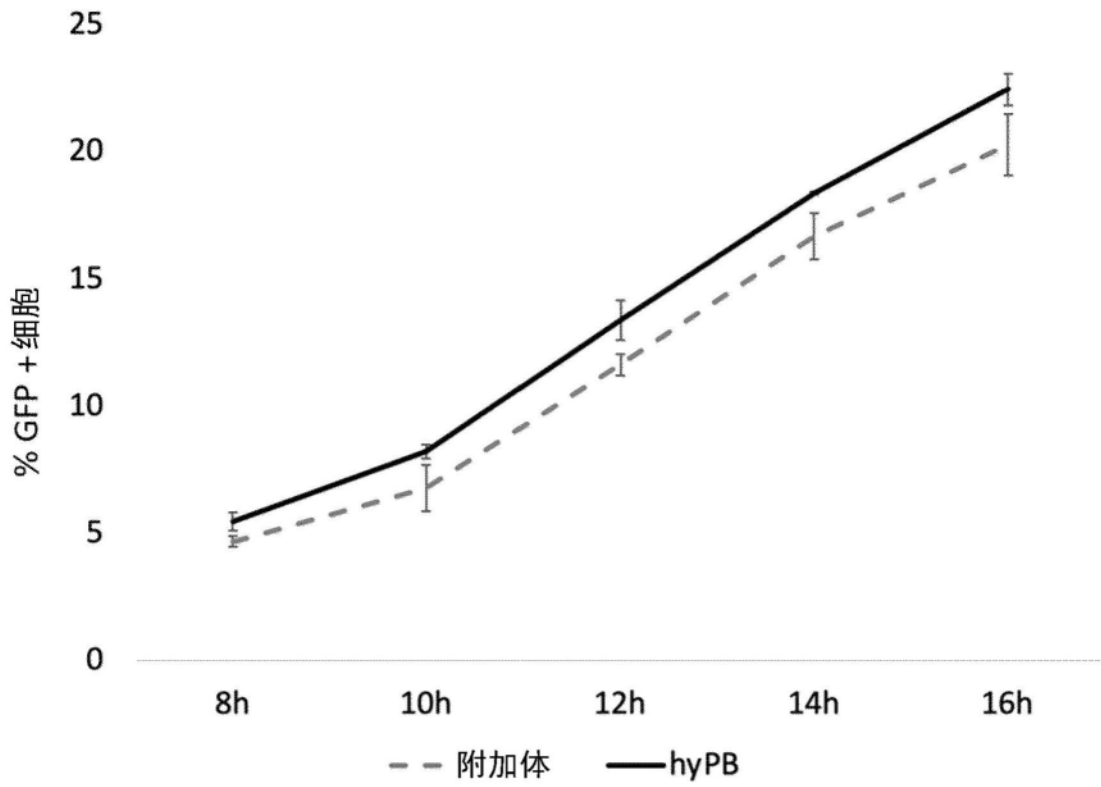


图5

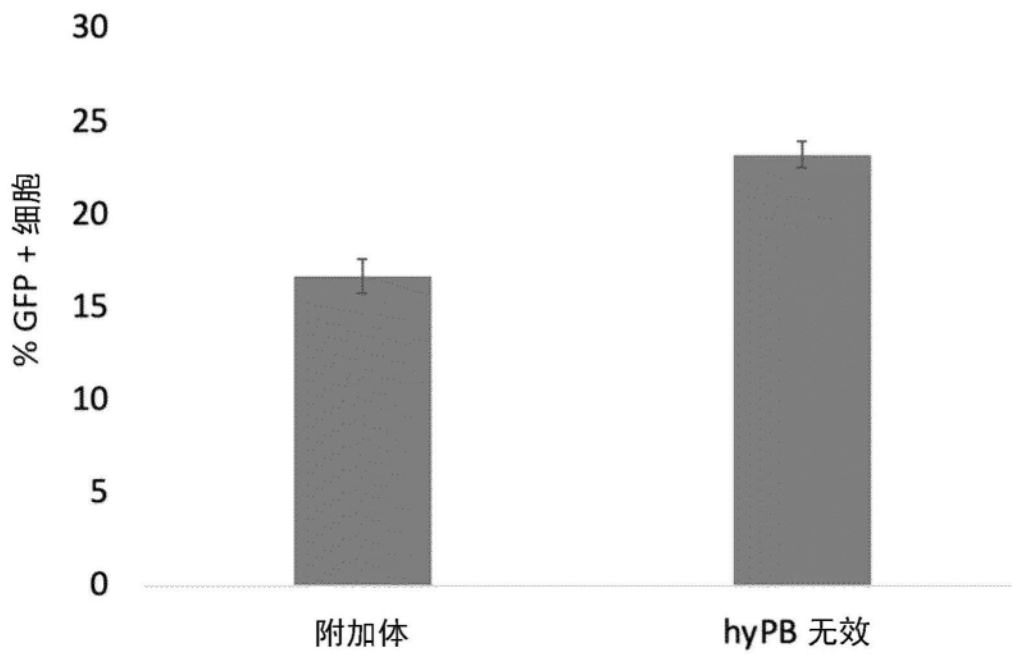


图6

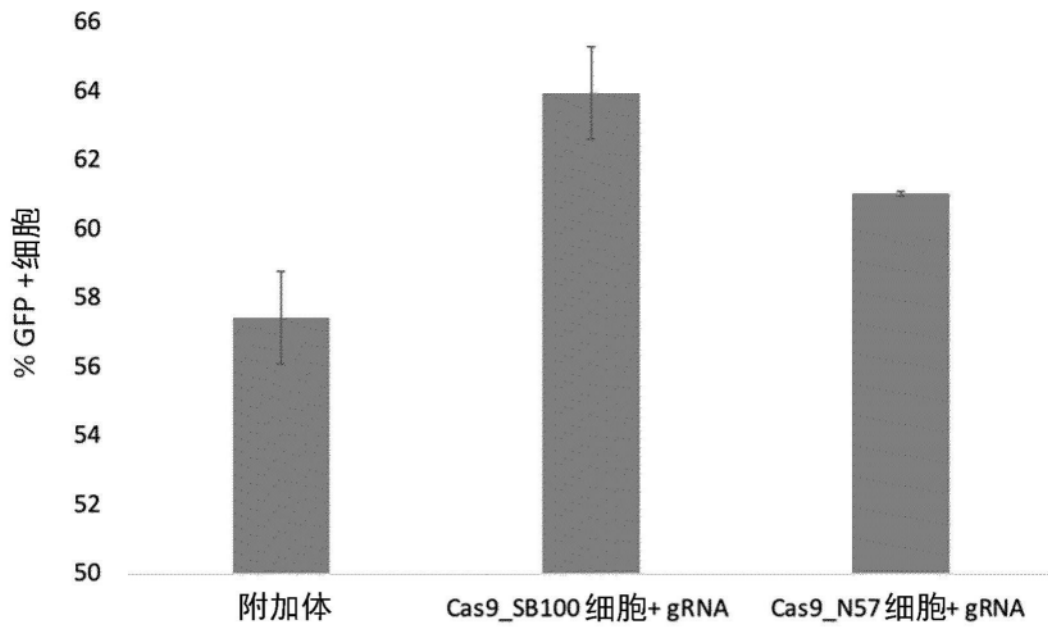


图7

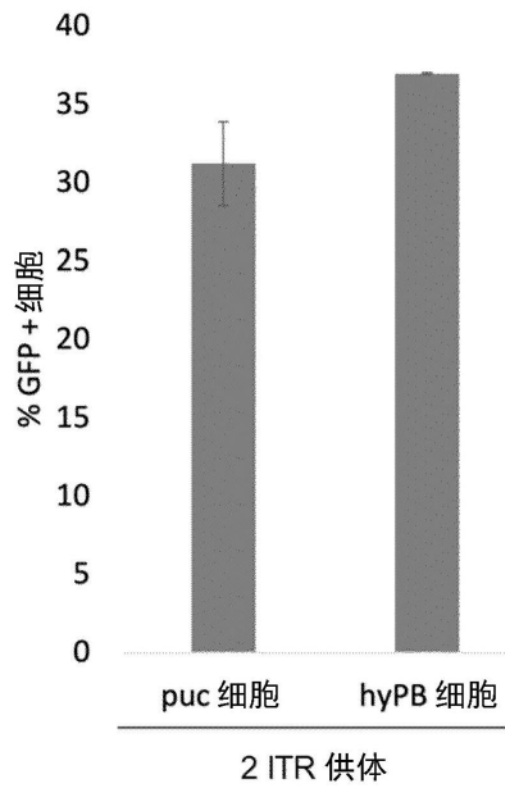


图8A

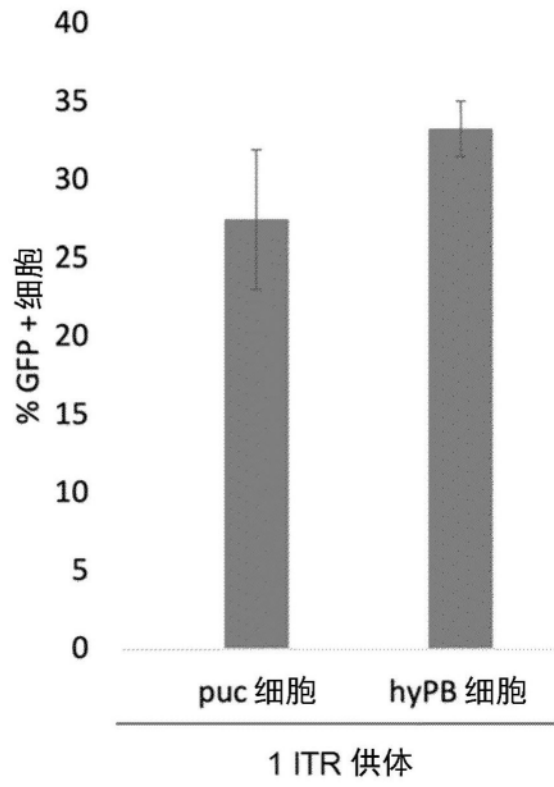


图8B

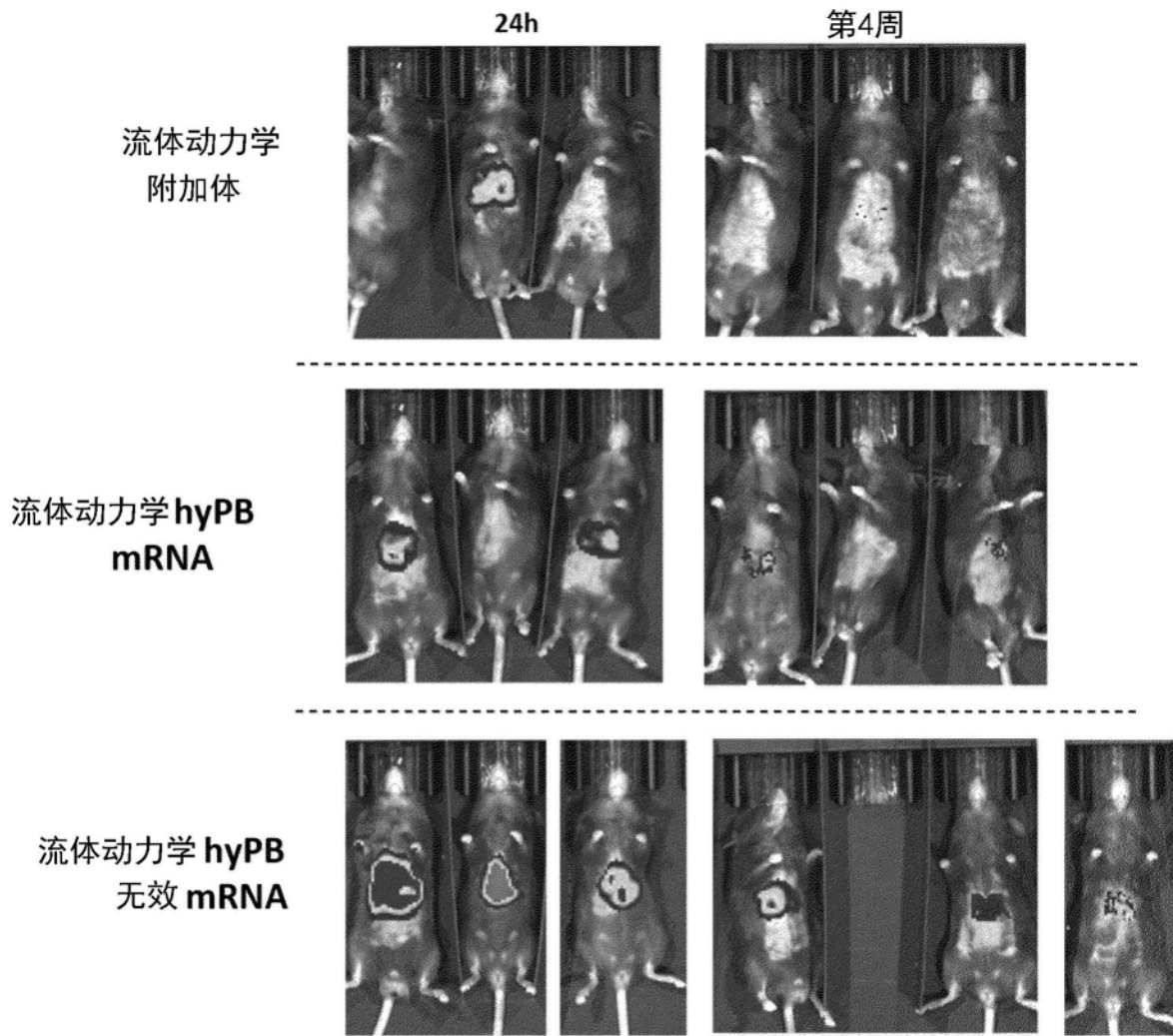


图9

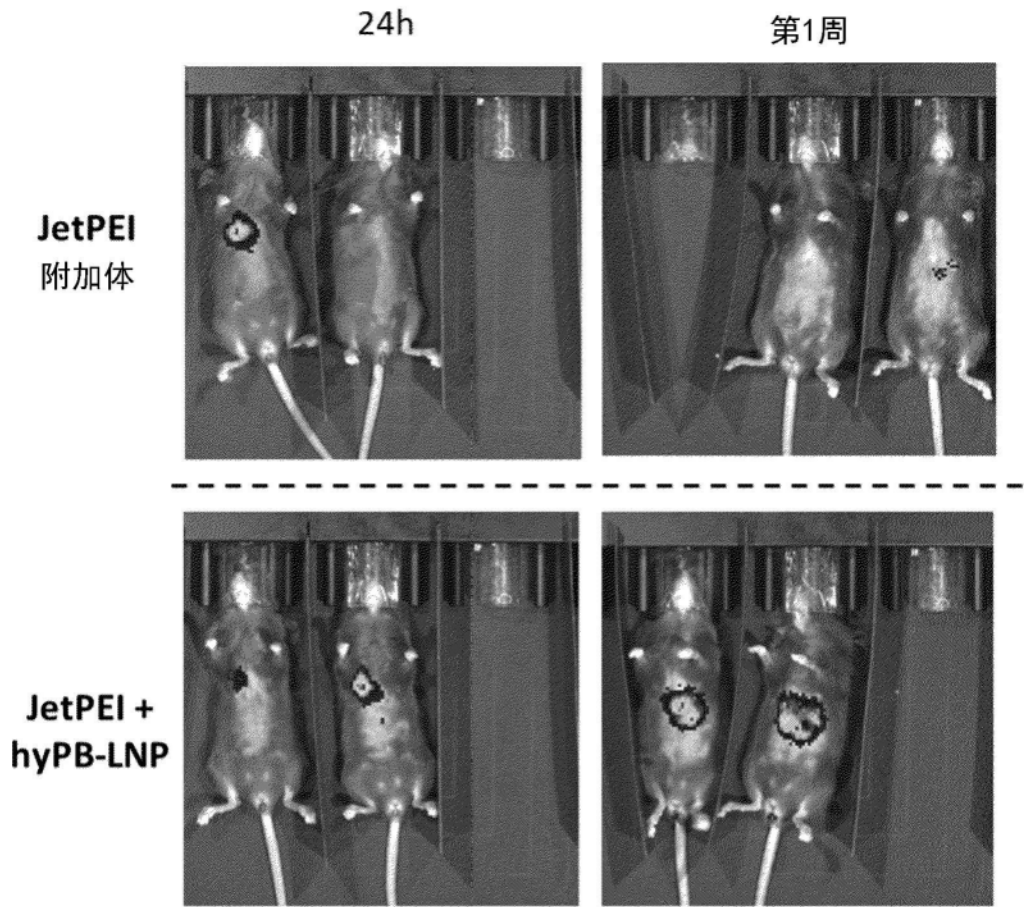


图10

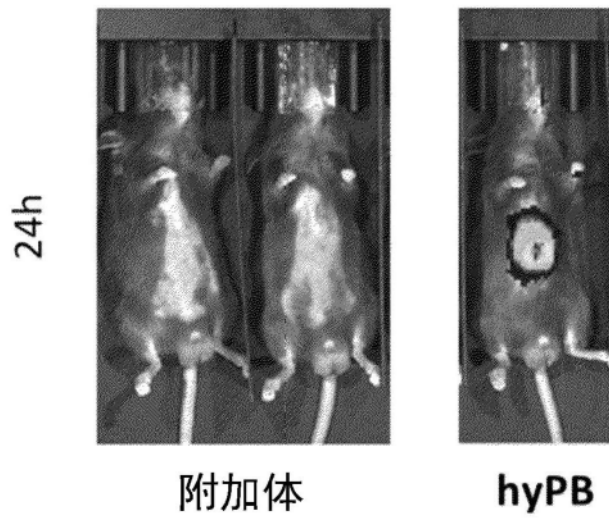


图11

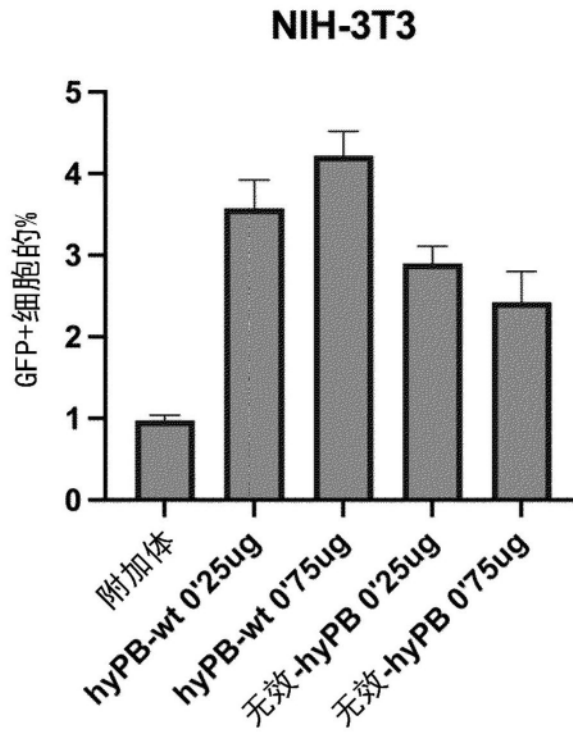


图12

NIH/3T3 细胞系 分裂

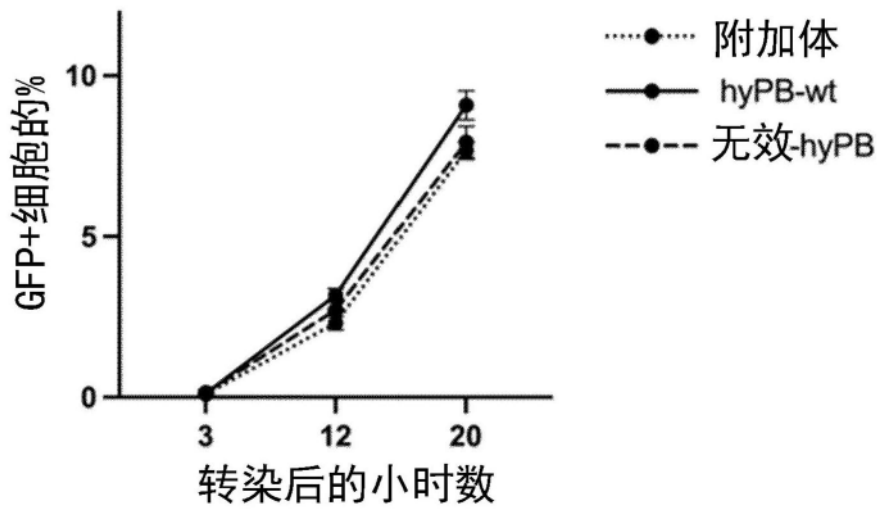


图13A

NIH/3T3 细胞系 非分裂

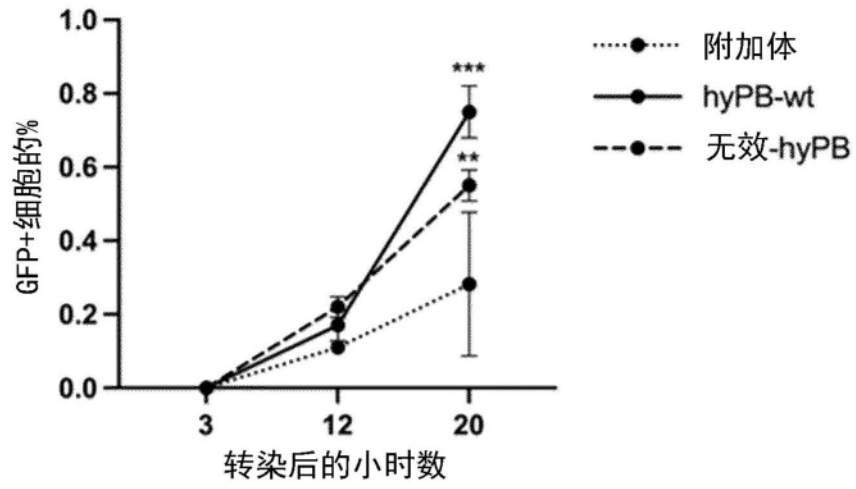


图13B

非分裂细胞中的DNA易位 原代肝细胞

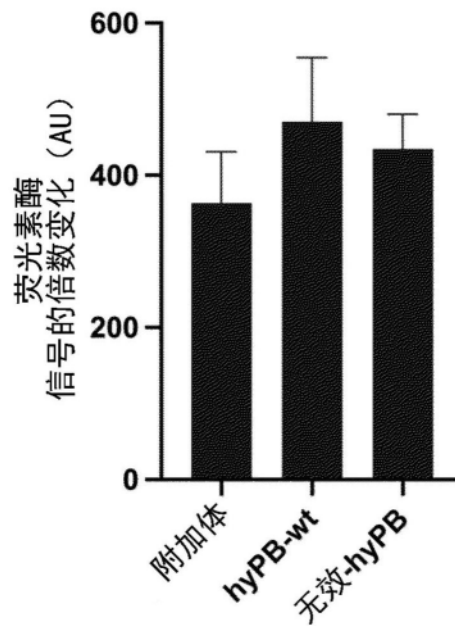


图14