(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111511759 B
(45)授权公告日 2024.07.30

(21)申请号 201880078542.7
(22)申请日 2018.10.11
(65)同一申请的已公布的文献号
申请公布号 CN 111511759 A
(43)申请公布日 2020.08.07
(30)优先权数据
62/571,672 2017.10.12 US
62/608,478 2017.12.20 US
62/616,281 2018.01.11 US
62/624,629 2018.01.31 US
(85)PCT国际申请进入国家阶段日
2020.06.04
(86)PCT国际申请的申请数据
PCT/US2018/055412 2018.10.11
(87)PCT国际申请的公布数据
WO2019/075200 EN 2019.04.18

(73)专利权人 杰克逊实验室
地址 美国缅因州
(72)发明人 A·程 N·吉勒特 杜梦涵
(74)专利代理机构 北京市金杜律师事务所 11256
专利代理师 陈文平 徐志明

(51)Int.Cl.
C07K 14/435(2006.01)
C07K 19/00(2006.01)
C12N 9/14(2006.01)
C12N 15/09(2006.01)
C12N 15/62(2006.01)
(56)对比文件
CN 1350582 A,2002.05.22
审查员 张婷

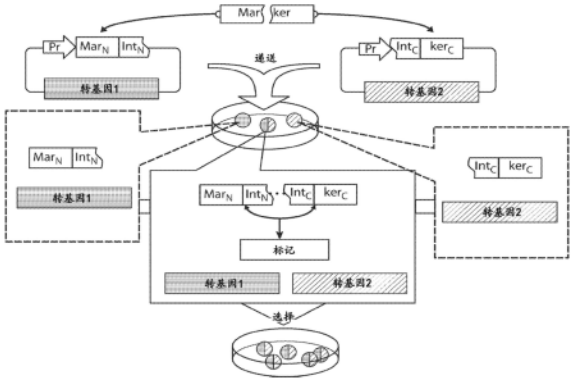
权利要求书2页 说明书53页
序列表（电子公布） 附图29页

(54)发明名称

转基因选择方法和组合物

(57)摘要

本公开内容提供了用于产生和选择转基因细胞的拆分内含肽选择标记系统。



CN 111511759 B

1.一种体外方法,其包括向真核细胞递送:

(a)第一载体,其包含(i)编码选择标记蛋白的N-末端片段的核苷酸序列,其位于编码拆分内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;和

(b)第二载体,其包含(i)编码所述拆分内含肽的C-末端片段的核苷酸序列,其位于所述选择标记蛋白的C-末端片段的上游,和(ii)编码第二目标分子的核苷酸序列,

其中所述拆分内含肽的N-末端片段和C-末端片段催化所述选择标记蛋白的N-末端片段和C-末端片段的接合以产生全长选择标记蛋白,

且其中所述第一目标分子和所述第二目标分子由不同的转基因编码。

2.根据权利要求1所述的体外方法,进一步包括将所述真核细胞维持在允许所述第一和第二载体引入所述真核细胞中以产生转基因真核细胞的条件下。

3.根据权利要求2所述的体外方法,进一步包括选择包含所述全长选择标记蛋白的所述转基因真核细胞。

4.根据权利要求1-3任一项所述的体外方法,其中所述真核细胞是哺乳动物细胞。

5.根据权利要求1-3任一项所述的体外方法,其中所述选择标记蛋白是抗生素抗性蛋白。

6.根据权利要求1-3任一项所述的体外方法,其中所述选择标记蛋白是荧光蛋白。

7.根据权利要求1-3任一项所述的体外方法,其中所述拆分内含肽是DnaE内含肽或DnaB内含肽。

8.根据权利要求7所述的体外方法,其中所述DnaE内含肽选自集胞藻(Synechocystis sp.)DnaE(SspDnaE)内含肽和点形念珠藻(Nostoc punctiforme)(NpuDnaE)内含肽。

9.根据权利要求7所述的体外方法,其中所述DnaB内含肽是SspDnaB S1内含肽。

10.根据权利要求1-3任一项所述的体外方法,其中所述第一和/或第二分子是蛋白质或非编码核糖核酸(RNA)。

11.根据权利要求1-3任一项所述的体外方法,其中所述第一和/或第二载体是质粒载体或病毒载体。

12.一种试剂盒,其包含

(a)第一载体,其包含(i)编码选择标记蛋白的N-末端片段的核苷酸序列,其位于编码拆分内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;和

(b)第二载体,其包含(i)编码所述拆分内含肽的C-末端片段的核苷酸序列,其位于所述选择标记蛋白的C-末端片段的上游,和(ii)编码第二目标分子的核苷酸序列,

其中所述拆分内含肽的N-末端片段和C-末端片段催化所述选择标记蛋白的N-末端片段和C-末端片段的接合以产生全长选择标记蛋白,

且其中所述第一目标分子和所述第二目标分子由不同的转基因编码。

13.一种体外方法,包括向真核细胞递送

(a)第一载体,其包含(i)编码选择标记蛋白的N-末端片段的核苷酸序列,其位于编码第一拆分内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

(b)第二载体,其包含(i)编码所述第一拆分内含肽的C-末端片段的核苷酸序列,其位于编码所述选择标记蛋白的中心片段的核苷酸序列的上游,该编码所述选择标记蛋白的中心片段的核苷酸序列位于编码第二拆分内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

(c)第三载体,其包含(i)编码所述第二拆分内含肽的C-末端片段的核苷酸序列,其位于编码所述选择标记蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,

其中所述第一拆分内含肽的N-末端片段和C-末端片段催化所述选择标记蛋白的N-末端片段与所述选择标记蛋白的中心片段的接合,而所述第二拆分内含肽的N-末端片段和C-末端片段催化所述选择标记蛋白的中心片段与所述选择标记蛋白的C-末端片段的接合,以产生全长选择标记蛋白,

且其中所述第一目标分子、所述第二目标分子和所述第三目标分子由不同的转基因编码。

14.根据权利要求13所述的体外方法,进一步包括将所述真核细胞维持在允许所述第一、第二和第三载体引入所述真核细胞中以产生转基因真核细胞的条件下。

15.根据权利要求14所述的体外方法,进一步包括选择包含所述全长选择标记蛋白的所述转基因真核细胞。

16.一种试剂盒,其包含:

(a)第一载体,其包含(i)编码选择标记蛋白的N-末端片段的核苷酸序列,其位于编码第一拆分内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

(b)第二载体,其包含(i)编码所述第一拆分内含肽的C-末端片段的核苷酸序列,其位于编码所述选择标记蛋白的中心片段的核苷酸序列的上游,该编码所述选择标记蛋白的中心片段的核苷酸序列位于编码第二拆分内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,及

(c)第三载体,其包含编码(i)所述第二拆分内含肽的C-末端片段的核苷酸序列,其位于编码所述选择标记蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,

其中所述第一拆分内含肽的N-末端片段和C-末端片段催化所述选择标记蛋白的N-末端片段与所述选择标记蛋白的中心片段的接合,而所述第二拆分内含肽的N-末端片段和C-末端片段催化所述选择标记蛋白的中心片段与所述选择标记蛋白的C-末端片段的接合,以产生全长选择标记蛋白,

且其中所述第一目标分子、所述第二目标分子和所述第三目标分子由不同的转基因编码。

转基因选择方法和组合物

[0001] 相关申请

[0002] 本申请依据35U.S.C.§119(e)要求2018年1月11日提交的美国临时申请号62/616,281,2017年12月20日提交的美国临时申请号62/608,478,2018年1月31日提交的美国临时申请号62/624,629,2017年10月12日提交的美国临时申请号62/571,672的权益,将其全部内容按引用并入本文中。

[0003] 序列表

[0004] 本申请包括计算机可读形式的序列表(文件名:J022770007WO00-SEQ-HJD;1.50MB-ASCII文本文件;创建于2018年10月3日),将其全部内容按引用并入本文中并形成公开内容的一部分。

背景技术

[0005] 选择标记广泛采用于转基因和基因组编辑中以用于选择具有所需基因型的工程化细胞。抗生素抗性基因(编码抗生素抗性蛋白)提供对特定抗生素的抗性,使得只有表达这些抗性基因的细胞才能存活并繁殖。可用于真核细胞中的抗生素抗性基因/抗生素包括hygB/潮霉素、neo/Geneticin®/G418、pac/嘌呤霉素、Sh ble/腐草霉素D1(Zeocin™)和bsd/杀稻瘟菌素。荧光蛋白,如绿色荧光蛋白(GFP),提供了另一种细胞选择的方式,例如,通过荧光激活细胞分选(FACS)技术或荧光显微术。

发明内容

[0006] 可用于真核(例如,哺乳动物)细胞中的抗生素抗性基因/抗生素数量有限,因此用于鉴定含有多个转基因的细胞的选择方案受到限制。不仅在真核细胞中赋予抗生素抗性的独特基因的数量有限,而且同时使用少至三个不同抗生素抗性基因也可能不利地影响转基因细胞的健康。尽管抗生素选择可以连续地进行,但这个过程很耗时。当需要鉴定其中已经引入多个转基因(例如,以产生转基因生物体,例如动物模型,例如小鼠模型)的细胞时,对用于鉴定转基因细胞的选择方案的这些限制是个问题。

[0007] 本文提供了可用于产生和/或鉴定例如带有两个或更多个转基因(例如,双转基因、三转基因等)的细胞和/或生物体的方法、组合物和试剂盒。例如,组合物和试剂盒可以用于产生和/或鉴定带有两个、三个或四个转基因的细胞和/或生物体。这种技术至少部分地基于由内含肽(intein)自动加工结构域启动的蛋白质剪接机制,该机制促进特别地多转基因细胞中多个(例如两个、三个或四个)分开的选择标记蛋白片段的接合(缀合)(双转基因细胞,三转基因细胞或四转基因细胞)。在多转基因细胞中两个、三个、四个或更多个分开的选择标记蛋白片段的连接产生全长选择标记蛋白,其赋予例如抗生素抗性(抗生素抗性蛋白)或能够在适当的光波长下发出荧光(荧光蛋白)。表达全长抗生素抗性基因的细胞在相应抗生素的存在下存活,并因此作为多转基因(例如,双转基因、三转基因或四转基因)细胞被选择。同样地,表达全长功能性荧光蛋白的细胞在适当的光波长下发出荧光,并因此作为多转基因(例如,双转基因、三转基因或四转基因)细胞被选择。

[0008] 因此,在一些实施方式中,本公开提供了包括将两个或更多个载体递送至包含真核细胞的组合物的方法,其中每个载体包含(i)编码与N-末端内含肽蛋白片段和/或C-末端内含肽蛋白片段连接的选择标记蛋白片段的核苷酸序列和(ii)编码目标分子的核苷酸序列,其中内含肽蛋白片段在框内接合以形成全长功能性蛋白时催化选择标记蛋白片段的接合,以产生全长选择标记蛋白。例如,当两个载体递送至细胞群体时(例如,在转染条件下),一些细胞摄取第一载体(载体被引入细胞中),一些细胞摄取第二载体和一些细胞摄取两个载体。只有摄取两个载体的那些细胞能够表达全长功能性选择标记蛋白,因此只有那些细胞作为双转基因细胞被选择。

[0009] 在一些实施方式中,本文的方法包括将(a)第一载体和(b)第二载体递送至包含真核细胞的组合物,所述第一载体包含(i)在编码N-末端内含肽蛋白片段的核苷酸序列上游的编码第一选择标记蛋白片段(例如,抗生素抗性蛋白片段或荧光蛋白片段)的核苷酸序列和(ii)编码第一分子的核苷酸序列,所述第二载体包含(i)在第二选择标记蛋白片段(例如,抗生素抗性蛋白片段或荧光蛋白片段)上游的编码C-末端内含肽蛋白片段的核苷酸序列和(ii)编码第二分子的核苷酸序列,其中N-末端内含肽蛋白片段和C-末端内含肽蛋白片段催化第一选择标记蛋白片段与第二选择标记蛋白片段的接合以产生全长选择标记蛋白。当将两个载体递送至细胞群体时(例如,在转染条件下),一些细胞摄取第一载体(载体被引入细胞中),一些细胞摄取第二载体和一些细胞摄取两个载体。只有摄取两个载体的那些细胞能够表达全长功能性选择标记蛋白,因此只有那些细胞作为双转基因细胞被选择。

[0010] 在其他实施方式中,方法包括将(a)第一载体、(b)第二载体和(c)第三载体递送至真核细胞,所述第一载体包含(i)编码选择标记蛋白(例如,抗生素抗性蛋白或荧光蛋白)的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,所述第二载体包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的中心片段的核苷酸序列的上游,该编码选择标记蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,且所述第三载体包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,其中第一内含肽的N-末端片段和C-末端片段催化选择标记蛋白的N-末端片段与选择标记蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化选择标记蛋白的中心片段与选择标记蛋白的C-末端片段的接合,以产生全长选择标记蛋白。当三个载体递送至细胞群体时(例如,在转染条件下),一些细胞摄取第一载体(载体被引入细胞中),一些细胞摄取第二载体,一些细胞摄取第三载体,一些细胞摄取两个不同的载体和一些细胞摄取全部三个载体。只有摄取全部三个载体的那些细胞能够表达全长功能性选择标记蛋白,因此只有那些细胞作为三转基因细胞被选择。

[0011] 再在其他实施方式中,方法包括将(a)第一载体、(b)第二载体、(c)第三载体和(d)第四载体递送至真核细胞,所述第一载体包含(i)编码选择标记蛋白(例如,抗生素抗性蛋白或荧光蛋白)的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,所述第二载体包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的第一中心片段的核苷酸序列的上游,该编码选择标记蛋白的第一中心片段的核苷酸序列位于编码第二内含肽的N-末端

片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,所述第三载体包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的第二中心片段的核苷酸序列的上游,该编码选择标记蛋白的第二中心片段的核苷酸序列位于编码第三内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,且所述第四载体包含(i)编码第三内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,其中第一内含肽的N-末端片段和C-末端片段催化选择标记蛋白的N-末端片段与选择标记蛋白的第一中心片段的接合,第二内含肽的N-末端片段和C-末端片段催化选择标记蛋白的第一中心片段与选择标记蛋白的第二中心片段的接合,第三内含肽的N-末端片段和C-末端片段催化选择标记蛋白的第二中心片段与选择标记蛋白的C-末端片段的接合,以产生全长选择标记蛋白。当四个载体递送至细胞群体时(例如,在转染条件下),一些细胞摄取第一载体(载体被引入细胞中),一些细胞摄取第二载体,一些细胞摄取第三载体,一些细胞摄取第四载体,一些细胞摄取两个不同的载体,一些细胞摄取三个不同载体和一些细胞摄取全部四个载体。只有摄取全部四个载体的那些细胞能够表达全长功能性选择标记蛋白,因此只有那些细胞作为四转基因细胞被选择。

[0012] 应当理解本文所述的任一个实施方式,包括只在实施例或说明书的一个部分公开的那些,旨在能够与任一个或多个其他实施方式结合,除非明确声明排除。

附图说明

[0013] 图1A-1B.用于两个单独的转基因载体的抗生素共选择的拆分选择标记。(图1A)将选择标记的编码序列拆分成N-末端片段(MarN)和C-末端片段(kerC)并且在两个各自携带不同转基因的不同载体上单独地克隆到分别拆分内含肽的N-末端片段(IntN)的上游和拆分内含肽的C-末端片段(IntC)的下游。这些载体递送至细胞,从而产生含有任一个载体或两个载体的细胞的亚群。只有在同时具有两个表达两个内含肽拆分选择标记片段(“markertron”)的载体的细胞中,发生蛋白质反式剪接以重建全长选择标记,使得允许双转基因细胞的特异性选择和富集。(图1B)为了筛选与用于抗生素抗性基因的内含肽相容的拆分点,我们根据测试的内含肽类型的连接需求鉴定潜在的拆分点,随后将相应的N-末端和C-末端片段克隆至配备有TagBFP2或mCherry荧光蛋白(其用作我们的测试转基因来评估选择效率)的慢病毒载体上的拆分内含肽支架。这些通过慢病毒转导递送至细胞中。随后将细胞分入重复平板中,一个进行抗生素选择,而另一个维持在非选择培养基中。抗生素选择后,通过流式细胞术来分析重复的培养物。

[0014] 图2A-2F.内含肽拆分抗性(Intres)基因(也称为选择标记基因)的拆分点和质粒的详细内容。(图2A)用于潮霉素抗性蛋白的拆分点(SEQ ID NO:1)。呈现了潮霉素抗性蛋白的氨基酸序列,这个研究中表征的拆分点用浮框(cloud)进行标记。在标记内,上面一行表示对应于表1的质粒编号。下面一行表示N-末端片段中最后一个氨基酸的残基编号、所用的内含肽的种类和C-末端片段中的第一个氨基酸的残基编号。“^C”表示半胱氨酸的插入。(图2B)用于嘌呤霉素抗性蛋白的拆分点(SEQ ID NO:2)。呈现了嘌呤霉素抗性蛋白的氨基酸序列,这个研究中表征的拆分点用浮框进行标记。在标记内,上面一行表示对应于表1的质粒编号。下面一行表示N-末端片段中的最后一个氨基酸的残基编号、所用的内含肽的种类和

C-末端片段中的第一个氨基酸的残基编号。“^C”表示半胱氨酸的插入。(图2C)用于新霉素抗性蛋白的拆分点(SEQ ID NO:3)。呈现了新霉素抗性基因的氨基酸序列,这个研究中表征的拆分点用浮框进行标记。在标记内,上面一行表示对应于表1的质粒编号。下面一行表示N-末端片段中的最后一个氨基酸的残基编号、所用的内含肽的种类和C-末端片段中的第一个氨基酸的残基编号。(图2D)用于杀稻瘟菌素抗性蛋白的拆分点(SEQ ID NO:4)。呈现了杀稻瘟菌素抗性基因的氨基酸序列,这个研究中表征的拆分点用浮框进行标记。在标记内,上面一行表示对应于表1的质粒编号。下面一行表示N-末端片段中的最后一个氨基酸的残基编号、所用的内含肽的种类和C-末端片段中的第一个氨基酸的残基编号。(图2E)用于绿色荧光蛋白的拆分点(SEQ ID NO:5)。(图2F)用于mSarlet荧光蛋白的拆分点(SEQ ID NO:6)。呈现了mSarlet基因的氨基酸序列,这个研究中表征的拆分点用浮框进行标记。在标记内,上面一行表示对应于表1的质粒编号。下面一行表示N-末端片段中的最后一个氨基酸的残基编号、所用的内含肽的种类和C-末端片段中的第一个氨基酸的残基编号。“^C”表示半胱氨酸的插入。

[0015] 图3.2-markertron潮霉素(Hygro)内含肽拆分抗性(Intres)基因。上图显示了针对潮霉素抗性基因测试的拆分点。棒糖形顶部指明N-末端片段的最后一个残基。圆形棒糖形表示使用NpuDnaE内含肽的拆分点,而方形棒糖形表示使用SspDnaB内含肽的那些。划掉和加阴影的棒糖形表示未能给予细胞潮霉素抗性的拆分对。下面的柱状图显示了通过流式细胞术分析的非选择的(白色柱)和选择的(蓝色柱)培养物中双转基因细胞(BFP+mCherry+)的百分比。

[0016] 图4.2-markertron嘌呤霉素(Puro)Intres基因。上图显示了针对嘌呤霉素抗性基因测试的拆分点,而下面的柱状图显示了非选择的(白色柱)和选择的(棕色柱)培养物中双转基因细胞的百分比。

[0017] 图5.2-markertron新霉素(Neo)抗性基因。上图显示了针对新霉素抗性基因测试的拆分点,而下面的柱状图显示了非选择的(白色柱)和选择的(橙色柱)培养物中双转基因细胞的百分比。

[0018] 图6.2-markertron杀稻瘟菌素(Blast)Intres基因。上图显示了针对杀稻瘟菌素抗性基因测试的拆分点,而下面的柱状图显示了非选择的(白色柱)和选择的(青色柱)培养物中双转基因细胞的百分比。

[0019] 图7A-7C.具有2-markertron Intres标志物的Gateway相容慢病毒目的载体。(图7A)用于每个拆分Intres标志物的Gateway相容慢病毒目的载体套装由N-载体和C-载体组成。N-载体含有病毒LTR、CAGGS启动子、gateway目的盒AttL、ccdB基因、允许LR克隆酶介导的携带转基因的Gateway供体载体重组的氯霉素抗性基因,接着是允许N-markertron的多顺反子表达的内部核糖体进入位点(IRES)。相似地,C-载体含有C-markertron并允许另一转基因的重组。(图7B)通过Gateway重组将TagBFP2(作为转基因1)和mCherry(作为转基因2)克隆至2-markertron Intres质粒中并通过慢病毒转导递送至细胞,接着是抗生素选择和流式细胞术分析。柱状图显示了2-markertron潮霉素(Hygro,蓝色柱)、嘌呤霉素(Puro,棕色柱)和新霉素(Neo,橙色柱)实验的选择培养物相对于其相应的非选择柱培养物(白色柱)中BFP+mCherry+双阳性细胞的百分比。(图7C)用GFP荧光标记核的NLS-GFP(作为转基因1)和用mScarlet荧光标记F-肌动蛋白的lifeAct-mScarlet(作为转基因2)重组至表达全长

非拆分潮霉素抗性基因的慢病毒载体或具有2-markertron潮霉素Intres基因的慢病毒载体中,并用于转导U2OS细胞以制备双标记细胞。代表性荧光显微图像显示了潮霉素选择两周后细胞的GFP、mScarlet和合并的通道。

[0020] 图8A-8C.用于两个单独的转基因载体的荧光介导的共选择的拆分mScarlet。(图8A)2-markertron mScarlet蛋白。上图显示了针对mScarlet测试的拆分点。棒糖形的顶部指示N-末端片段的最后一个残基。(图8B)为了筛选用于mScarlet的NpuDnaE内含肽相容的拆分点,我们根据NpuDnaE内含肽的连接需求鉴定了潜在的拆分点,随后将相应的N-末端和C-末端片段克隆至配备有TagBFP2或EGFP荧光蛋白(其用作我们的测试转基因来评估选择效率)的慢病毒载体上的拆分内含肽支架。将这些通过慢病毒转导递送至细胞中。具有两种慢病毒的细胞含有必要的蛋白质剪接机制和mScarlet片段以重构全长mScarlet荧光蛋白,以及表达TagBFP2和EGFP两个转基因。细胞进行FACS分析。加框的图显示了质粒对33+34的FACS分析的实例。将P1群体针对活的单线态细胞的前向散射和侧向散射进行门控。在其中,17.8%的细胞是TagBFP2和EGFP转基因双阳性的。当P1细胞进一步针对mScarlet阳性(mCherry通道)进行门控时,99.4%的细胞是TagBFP2和EGFP转基因双阳性的。(图8C)下面的柱状图显示了每个所示拆分点的mScarlet阳性细胞的百分比。上面的柱状图显示了P1细胞(白色柱)和P1细胞的mScarlet阳性子集(红色柱)中TagBFP2+EGFP+细胞的百分比。

[0021] 图9A-9D.用于三个或更多个转基因载体的共选择的多拆分选择标记。(图9A)将选择标记分成三个片段(M_1、M_2和M_3)。第一标记片段(M_1)在第一拆分内含肽的N-末端片段(I_{N1})的上游融合。第二标记片段(M_2)在第一拆分内含肽的C-末端片段(I_{C1})的下游和第二拆分内含肽的N-末端片段(I_{N2})的上游融合。第三标记片段(M_3)在第二拆分内含肽的C-末端片段(I_{C2})的下游融合。第一拆分内含肽催化M_1与M_2的接合,而第二拆分内含肽催化M_2与M_3的接合,从而有效地重构全长选择标记。(图9B)通过“内含肽链”机制的k-拆分选择标记的设计。与3-拆分情况相似,将选择标记分割成k个片段,并且通过插入拆分内含肽介导的蛋白质反式剪接重构。(图9C)从2-拆分选择标记鉴定的拆分点组合使用以产生3-拆分选择标记。相应的片段克隆至慢病毒载体中以形成3-拆分选择标记结构和每个载体的报告荧光转基因。细胞随后用从这些载体制备的病毒转导,拆分至选择性或非选择性培养基中。在适当的选择期后,通过流式细胞术分析培养物。(图9D)3-markertron潮霉素(Hygro)Intres。上图显示了针对潮霉素抗性基因测试的拆分点,圆形或方形棒糖形上方指示了N-末端片段的最后一个氨基酸的残基编号,分别表示NpuDnaE和SspDnaB内含肽。测试了六个3-markertron潮霉素Intres,每个用编号的线来表示,圆形或方形表示用于每种情况中的两个拆分点。下面的柱状图显示了针对由下面的编号指示的3-markertron潮霉素Intres来自非选择(白色柱)和选择(蓝色柱)培养物的三转基因(BFP+GFP+mCherry+)细胞的百分比。

[0022] 图10A-10C.具有3-markertron潮霉素Intres基因的Gateway相容性慢病毒目的载体。(图10A)Gateway相容性慢病毒目的载体,其具有病毒LTR、CAGGS启动子、gateway目的盒AttL、ccdB基因、允许LR克隆酶介导的携带转基因的Gateway供体载体重组的氯霉素抗性基因,接着是内部核糖体进入位点(IRES),其允许三个3-拆分潮霉素markertrons中每一个的多顺反子表达。(图10B)将TagBFP2(作为转基因1)和EGFP(作为转基因2)和mCherry(作为转基因3)通过Gateway重组克隆至3-拆分Intres质粒中并通过慢病毒转导递送至细胞,接着进行抗生素选择和流式细胞术分析。(图10C)柱状图显示了潮霉素选择的培养物(蓝色柱)

相对于其相应的非选择培养物(白色柱)中BFP+GFP+mCherry+三阳性细胞的百分比。

[0023] 图11.四拆分Hygro inters。(a)由四个不同的质粒表示的4-拆分hygro intres的markertrons。质粒115表示由潮霉素抗性基因的氨基酸1-89[Hygro(1-89)]与NpuDnaE(N)和亮氨酸拉链A基序(LZA)融合形成的markertron。质粒116表示从N-至C-末端由亮氨酸拉链B基序(LZB)-NpuDnaGEP(C)、Hygro(90-200)和SspDnaB(N)融合形成的markertron。质粒117表示从N-至C-末端由SspDnaB(C)、Hygro(201-240)、NpuDnaE(N)-LZA融合形成的markertron。质粒118表示由LZB-NpuDnaGEP(C)与Hygro(241-341)融合形成的markertron。

[0024] 图12A-12E.Intres标记允许来自CRISPR/Cas介导的敲入实验的双等位基因靶向细胞的富集。设计了含有对于AAVS1安全港基因座的同源臂的靶向构建体对以包含全长(FL)非拆分或拆分Intres标记并测试其通过抗生素选择富集双等位基因靶向细胞的能力。(图12A)质粒107和108含有由AAVS1基因座处的内源性PPP1R12C启动子驱动的FL新霉素(Neo)抗性基因、由EF1a启动子驱动的FL潮霉素(Hygro)基因和rtTA Dox-响应性反式激活子以及表达的FL杀稻瘟菌素(Blast)及来自dox诱导型TetO启动子的EGFP(质粒107)和mScarlet(质粒108)。质粒106含有Cas9和靶向AAVS基因座的sgRNA。2A:自切割的2A肽。将质粒106、107和108共同转染至HEK293T细胞中,拆分并在含dox的潮霉素、杀稻瘟菌素或非选择培养基中传代两周,并通过流式细胞术进行分许以测定双等位基因靶向的效率。(图12B)质粒109和110含有与质粒107和108相似的结构,但具有拆分Blast Intres而非FL Blast。(图12C)质粒111和112含有由2A肽分隔的EF1a驱动的FL Blast和TetO驱动的FL Hygro、硝基还原酶(NTR)、荧光蛋白(EGFP或mCherry)。(图12D)质粒113和114与质粒111和112相似,但具有Hygro Intres而非FL Hygro。(图12E)在含dox的非选择培养基(选择:无)、杀稻瘟菌素选择培养基(Blast)和潮霉素选择培养基(Hygro)中培养两周后,用质粒106(Cas9+AAVS-sgRNA)和所示的靶向构建体对转染的细胞的流式细胞术分析。

具体实施方式

[0025] 在一些方面中,本文提供了产生转基因(例如,多转基因,如双转基因或三转基因)生物体的方法,所述生物体中引入了超过一个转基因(或其他遗传元件)。如图1A中所示,本公开的示例性方法包括将(a)编码N-末端内含肽蛋白片段上游的第一选择标记蛋白片段和第一目标转基因的载体及(b)编码第二选择标记蛋白片段上游的C-末端内含肽蛋白片段和第二(例如,不同的)目标转基因的另一个载体递送至细胞群体。所述群体中的一些细胞摄取单个载体(只携带一个内含肽片段、一个选择标记蛋白片段和单个转基因),而所述群体中的其他细胞摄取两个载体(并且因此携带两个内含肽片段、两个选择标记蛋白片段和两个目标转基因)。在摄取两个载体的细胞中,翻译后,内含肽蛋白片段自发地且非共价地装配(协同地折叠)成内含肽结构以催化第一选择标记蛋白片段与第二选择标记蛋白片段的接合,产生全长选择标记蛋白,其能够特异性地选择那些双转基因细胞。例如,如果选择标记蛋白是抗生素抗性蛋白,只有表达全长(功能性)抗生素抗性蛋白的双转基因细胞在特定抗生素存在下的选择中存活。作为另一个实例,如果选择标记蛋白是荧光蛋白,只有表达全长(功能性)荧光蛋白的双转基因细胞发射可检测的信号,使得只选择那些发射信号的细胞。

[0026] 本公开的另一种示例性方法包括将(a)第一载体、(b)第二载体和(c)第三载体递

送至细胞群体,所述第一载体包含(i)编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,所述第二载体包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的中心片段的核苷酸序列的上游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,所述第三载体包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列。所述群体中的一些细胞摄取单个载体(只携带一个内含肽片段、一个选择标记蛋白片段和单个转基因),而所述群体中的其他细胞摄取两个载体或全部三个载体(并且因此携带全部内含肽片段、全部选择标记蛋白片段和全部目标转基因)。在摄取全部三个载体的细胞中,翻译后,内含肽蛋白片段自发地且非共价地装配(协同地折叠)成内含肽结构以催化选择标记蛋白的N-末端片段与中心片段的接合以及中心片段与选择标记蛋白的C-末端片段的接合,从而产生全长选择标记蛋白,其能够特异性地选择那些三转基因细胞。例如,如果选择标记蛋白是抗生素抗性蛋白,只有表达全长(功能性)抗生素抗性蛋白的三转基因细胞在特定抗生素存在下的选择中存活。作为另一个实例,如果选择标记蛋白是荧光蛋白,只有表达全长(功能性)荧光蛋白的三转基因细胞发射可检测的信号,使得只选择那些发射信号的细胞。

[0027] 内含肽

[0028] 内含肽(插入蛋白)进行独特的自动加工事件,称为蛋白质剪接,其中内含肽通过切割两个肽键从较大的前体多肽中将其自身切除,并且在该过程中通过新肽键的形成将侧翼外显肽(外部蛋白)序列连接起来。这种重排在翻译后(或可能是共翻译地)发生,因为发现内含肽基因框内嵌入其他蛋白编码基因中。此外,内含肽介导的蛋白质剪接是自发的;它不需要外部因子或能量源,仅需要内含肽结构域的折叠。实质上,前体蛋白包含三个片段-N-外显肽(蛋白质的N-末端部分),然后是内含肽,接着是C-外显肽(蛋白质的C-末端部分)。剪接后,所得蛋白质包含与C-外显肽连接的N-外显肽。

[0029] 存在两种类型的内含肽:顺式-剪接内含肽是嵌入宿主蛋白质中的单一多肽,而反式-剪接内含肽(称为拆分内含肽)是分开的多肽,其在内含肽片段和其蛋白质载物结合后介导蛋白质剪接(参见,例如,Paulus,H Annu Rev Biochem 69:447-496(2000);和Saleh L,Perler FB Chem Rec 6:183-193(2006))。拆分内含肽催化一系列需要内含肽正确装配和折叠的化学重排。剪接的第一步涉及N-S酰基重排(acyl shift),其中N-外显肽多肽转移至内含肽的第一个残基的侧链。这随后接着是反式-(硫)酯化反应,其中这个酰基单元转移至C-外显肽的第一个残基(其是丝氨酸、苏氨酸或半胱氨酸)以形成分支中间体。在该过程的倒数第二个步骤中,通过涉及内含肽的C-末端天冬酰胺残基的转酰胺反应将这个分支中间体从内含肽切割。则这建立了该过程的最后一个步骤,其涉及S-N酰基转移以在两个外显肽之间形成正常的肽键(Lockless,SW,Muir,TW PNAS106(27):10999-11004(2009))。

[0030] 迄今为止,存在至少70种不同的内含肽等位基因,不仅通过内含肽嵌入其中的宿主基因类型来区分,而且还通过该宿主基因内的整合点来区分(Perler,FB Nucleic Acids Res.30:383-384(2002);Pietrokovski,S Trends Genet.17:465-472(2001))。小部分(低于5%的)鉴定的内含肽基因编码拆分内含肽。与更常见的连续内含肽不同,拆分内含肽作

为两个分开的多肽,N-内含肽和C-内含肽,单独转录和翻译,其各自与一个外显肽融合。翻译时,内含肽片段自发地且非共价地装配(协同地折叠)成正则的内含肽结构以进行反式蛋白质剪接。从蓝藻集胞藻(Synechocystis)种PCC6803(Ssp)和点形念珠藻(Nostoc punctiforme)PCC73102(Npu)鉴定的前两种拆分内含肽是天然发现插入DNA聚合酶III(DnaE)的α亚基中的直系同源物。Npu尤其值得注意,因为它的蛋白质反式剪接速度非常快(30℃下$t_{1/2}$=50s)。这一半衰期明显比Ssp(30℃下$t_{1/2}$=80min)短(Shah,NH等,J.Am.Chem.Soc.135:5839(2013))。

[0031] 在本文中,拆分内含肽用于催化选择标记蛋白(如抗生素抗性蛋白或荧光蛋白)的两个片段(例如,N-末端片段和C-末端片段)的接合以以产生功能性的、全长蛋白(例如,图1A和1B)。

[0032] 拆分内含肽可以是天然的拆分内含肽或工程化的拆分内含肽。天然拆分内含肽天然地存在于多种不同生物体中。在至少20个蓝藻种的DnaE基因内发现了最大的已知拆分内含肽家族(Caspi J等,Mol.Microbiol.50:1569-1577(2003))。因此,在本公开的一些实施方式中,天然拆分内含肽选自DnaE内含肽。DnaE内含肽的非限制性实例包括集胞藻属(Synechocystis sp.)DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0033] 在一些实施方式中,拆分内含肽是工程化拆分内含肽。工程化拆分内含肽可以从连续内含肽产生(其中将连续内含肽人工拆分)或可以是修饰的天然拆分内含肽,例如,这促进有效的蛋白质纯化、连接、修饰和环化(例如,Npu_{GEP}和Cfa_{GEP},如Stevens,AJ PNAS114(32):8538-8543(2017)所述的))。例如,Aranko,AS等,Protein Eng Des Sel.27(8):263-271(2014)描述了用于工程化拆分内含肽的方法,其按引用并入本文中。在一些实施方式中,工程化拆分内含肽从DnaB内含肽工程化(Wu,H等,Biochim Biophys Acta 1387(1-2):422-432(1998))。例如,工程化拆分内含肽可以是SspDnaB S1内含肽。在一些实施方式中,工程化拆分内含肽从GyrB内含肽工程化。例如,工程化拆分内含肽可以是SspGyrB S11内含肽。

[0034] 在一些实施方式中,其中产生三转基因物,例如,第一内含肽可以与第二内含肽相同(例如,都是DnaE内含肽)。在其他实施方式中,可以使用两种不同的内含肽(例如,DnaE内含肽和DnaB内含肽)。在一些实施方式中,第一内含肽是NpuDnaE内含肽,而第二内含肽是NpuDnaE内含肽。

[0035] 选择标记蛋白

[0036] 基于全长选择标记蛋白的表达来选择本公开的转基因(例如,双和/或三转基因)细胞。选择标记蛋白通常赋予适用于人工选择的性状。合适的选择标记蛋白的实例包括抗生性抗性蛋白和荧光蛋白。

[0037] 抗生素抗性基因是编码赋予对特定的抗生素或一类抗生素的抗性的蛋白质的基因。用于真核细胞中的抗生素抗性基因的非限制性实例包括编码赋予潮霉素、G418、嘌呤霉素、腐草霉素D1或杀稻瘟菌素抗性的蛋白质的那些。用于原核细胞中的抗生素抗性基因的非限制性实例包括编码赋予潮霉素、G418、嘌呤霉素、腐草霉素D1、杀稻瘟菌素、卡那霉素、壮观霉素、链霉素、氨苄青霉素、羧苄青霉素、博来霉素、红霉素、多粘菌素D1、四环素和氯霉素抗性的蛋白质的那些。

[0038] 潮霉素B是由细菌吸水链霉菌(Streptomyces hygroscopicus)产生的抗生素。其

是通过抑制蛋白质合成杀灭细菌、真菌和高等真核细胞的氨基糖苷。由最初源自大肠杆菌(Escherichia coli)的hpt基因(也称为hph或aphIV基因)编码的潮霉素磷酸转移酶(HPT)对氨基环醇抗生素潮霉素B解毒。因此,在一些实施方式中,本公开的选择标记基因是hpt基因。

[0039] G418(GENETICIN®)是结构与庆大霉素B1相似的氨基糖苷抗生素。其由红橙小单孢菌(Micromonospora rhodorangea)产生。G418通过抑制原核和真核细胞中的延伸步骤阻断多肽合成。通过来自编码氨基糖苷3'-磷酸转移酶APT 3'II的Tn5的neo基因赋予对G418的抗性。G418是硫酸新霉素的类似物,并且具有与新霉素相似的机制。因此,在一些实施方式中,本公开的选择标记基因是neo基因。

[0040] 嘌呤霉素是源自白黑链霉菌(Streptomyces alboniger)的氨基核苷抗生素,其导致在核糖体中发生翻译过程中的过早链终止。嘌呤霉素对于原核生物或真核生物是选择性的。通过嘌呤霉素N-乙酰基-转移酶(pac)基因的表达赋予对嘌呤霉素的抗性。因此,在一些实施方式中,本公开的选择标记基因是pac基因。

[0041] 腐草霉素D1(例如,ZEOCIN®)是糖肽抗生素并且是来自轮枝链霉菌(Streptomyces verticillus)的一种腐草霉素,属于抗生素的博来霉素家族。其是有效对抗大部分细菌、丝状真菌、酵母、植物和动物细胞的广谱抗生素。其通过插入DNA中引起细胞死亡并诱导DNA的双链断裂。通过首先从印度斯坦链异壁菌(Streptoalloteichus hindustanus)分离的Sh ble基因的产物赋予对腐草霉素D1的抗性。因此,在一些实施方式中,本公开的选择标记基因是Sh ble基因。

[0042] 杀稻瘟菌素S是由灰产色链霉菌(Streptomyces griseochromogenes)产生的抗生素。杀稻瘟菌素通过抑制翻译的终止步骤以及经由核糖体的肽键形成(程度较低)阻止真核和原核细胞的生长。通过至少三个不同基因赋予对杀稻瘟菌素的抗性:来自链轮丝菌属(Streptoverticillum spp.)的bls(酰基转移酶);来自蜡状芽孢杆菌(Bacillus cereus)的bsr(杀稻瘟菌素-S脱氨酶)(其他bsr基因也是已知的);和来自土曲霉(Aspergillus terreus)的bsd(另一种脱氨酶)。因此,在一些实施方式中,本公开的选择标记基因是bls基因、bsr基因或bsd基因。

[0043] 如本文提供的可以使用的荧光蛋白的非限制性实例包括TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0044] 在一些实施方式中,通过在同一细胞中接合两个选择标记基因片段来产生全长选择标记基因。在一些实施方式中,关于任何全长蛋白,一个片段是N-末端片段(N-外显肽),而另一片段是C-末端片段(C-外显肽)。因此,在一些实施方式中,第一抗生素抗性蛋白片段是N-末端抗生素抗性蛋白片段,而第二抗生素抗性蛋白片段是C-末端抗生素抗性蛋白片

段。在其他实施方式中,第一荧光蛋白片段是N-末端荧光蛋白片段,而第二荧光蛋白片段是C-末端荧光蛋白片段。

[0045] 在其他实施方式中,通过在同一细胞中接合三个或更多个选择标记基因片段来产生全长选择标记基因。在一些实施方式中,关于任何全长蛋白,一个片段是N-末端片段,一个或多个(例如,1、2或3个)片段是中心片段,和一个片段是C-末端片段。

[0046] N-末端片段可以是包括全长蛋白的游离胺基(-NH2)的任何蛋白质片段。C-末端片段可以是包括游离羧基(-COOH)的任何蛋白质片段。中心片段可以是位于全长蛋白的N-末端片段和C-末端片段之间的任何蛋白质片段。

[0047] 例如,编码潮霉素(341个氨基酸的蛋白质)的基因的氨基酸1-89可以称为N-末端蛋白质片段,而氨基酸90-341可以称为C-末端片段。相似地,参照图5,编码潮霉素的基因的氨基酸1-200可以称为N-末端蛋白质片段,而氨基酸201-341可以称为C-末端片段。图6显示了另外的实例,其中氨基酸1-53、1-240或1-292被认为是全长潮霉素的N-末端蛋白质片段,所述潮霉素含有氨基酸54-341、241-341或293-341作为相应的C-末端片段。

[0048] 作为另一个实例,编码潮霉素(341个氨基酸的蛋白质)的基因的氨基酸1-52可以称为N-末端蛋白质片段,氨基酸53-89可以称为中心蛋白片段,而氨基酸90-341可以称为C-末端片段。相似地,编码潮霉素的基因的氨基酸1-89可以称为N-末端蛋白质片段,氨基酸90-240可以称为中心片段,而氨基酸241-341可以称为C-末端片段。

[0049] 转基因和其他目标分子

[0050] 在一些实施方式中,本公开的方法和组合物用于产生多转基因(例如,双和/或三转基因)细胞和/或生物体。因此,在一些实施方式中,所述方法使用编码第一分子(第一目标分子)的一个载体和编码第二分子(第二目标分子)的另一个载体。在一些实施方式中,所述方法使用编码第三目标分子的再另一个载体。另外的载体(例如,编码选择标记蛋白的另外的中心片段)可以编码另外的目标分子。目标分子可以是例如多肽(例如,蛋白质和肽)或多核苷酸(例如,核酸,如DNA或RNA)。

[0051] 在一些实施方式中,第一分子(例如,位于第一载体上)是蛋白质。在一些实施方式中,第二分子(例如,位于第二载体上)是蛋白质。在一些实施方式中,第三分子(例如,位于第三载体上)是蛋白质。目标蛋白质的实例包括,但不限于,酶、细胞因子、转录因子、激素、生长因子、血液因子、抗原和抗体。

[0052] 在一些实施方式中,第一分子是肽。在一些实施方式中,第二分子是肽。在一些实施方式中,第三分子是肽。

[0053] 在一些实施方式中,第一分子是信使RNA(mRNA)。在一些实施方式中,第二分子是mRNA。在一些实施方式中,第三分子是mRNA。在一些实施方式中,mRNA编码疫苗或其他抗原性分子。

[0054] 在一些实施方式中,第一分子是非编码RNA(不编码蛋白质的RNA)。在一些实施方式中,第二分子是非编码RNA。在一些实施方式中,第三分子是非编码RNA。非编码RNA的实例包括,但不限于,RNA干扰分子,如微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0055] 载体

[0056] 本公开的方法包括使用至少两个或至少三个不同的载体。载体是可以用作介质将

外源(外来)遗传物质携带至细胞中的任何核酸。在一些实施方式中,载体是包括插入序列(例如,转基因)和用作载体主链的较大序列的DNA序列。载体的非限制性实例包括质粒、病毒/病毒载体、粘粒和人工染色体,其中任何一个可以按照本文所述的使用。在一些实施方式中,载体是病毒载体,如病毒颗粒。在一些实施方式中,载体是基于RNA的载体,如自我复制RNA载体。在一些实施方式中,第一载体是质粒,第二载体是质粒和/或第三载体是质粒。如本文提供的载体包含可操作地连接于编码内含肽的片段和选择标记蛋白的片段的核酸的启动子。在一些实施方式中,载体还包含可操作地连接于编码目标分子的核酸(如转基因)的启动子。

[0057] 在一些实施方式中,一个载体(例如,第一载体)包含位于编码N-末端内含肽蛋白片段的核苷酸序列上游的编码第一选择标记蛋白片段的核苷酸序列,而其他载体(例如,第二载体)包含位于第二抗生素抗性蛋白片段上游的编码C-末端内含肽蛋白片段的核苷酸序列(参见,例如,图1A)。这种构型等同于一个载体(例如,第一载体)包含位于编码第一选择标记蛋白片段的核苷酸序列下游的编码N-末端内含肽蛋白片段的核苷酸序列,而其他载体(例如,第二载体)包含位于编码C-末端内含肽蛋白片段的核苷酸序列下游的第二抗生素抗性蛋白片段。术语“上游”和“下游”是指核酸中的相对位置。每个核酸具有5’端和3’端,这是针对脱氧核糖(或核糖)环上的碳位置来命名的。例如,当考虑双链DNA时,上游朝向编码链的5’端,而下游朝向3’端。

[0058] 在一些实施方式中,(a)第一载体包含编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,(b)第二载体包含编码第一内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的中心片段的核苷酸序列的上游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(c)第三载体包含编码第二内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的C-末端片段的核苷酸序列的上游。这种构型等同于(a)第一载体包含编码第一内含肽的N-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的N-末端片段的核苷酸序列的下游,(b)第二载体包含编码第二内含肽的N-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的中心片段的核苷酸序列的下游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第一内含肽的C-末端片段的核苷酸序列的下游,和(c)第三载体包含抗生素抗性蛋白的C-末端片段,其位于编码第二内含肽的C-末端片段的核苷酸序列的下游。

[0059] 细胞

[0060] 本公开的方法可以用于通过将本文所述的载体(例如,第一和第二载体)引入宿主细胞中产生转基因细胞和生物体。其中引入载体的细胞可以是真核或原核的。在一些实施方式的,细胞是真核的。如本文所述使用的真核细胞的实例包括哺乳动物细胞、植物细胞(例如,作物细胞)、昆虫细胞(例如,果蝇(Drosophila))和真菌细胞(例如,酵母(Saccharomyces))。哺乳动物细胞例如可以是人类细胞(干细胞或来自建立的细胞系的细胞)、灵长类细胞、马类细胞、牛类细胞、猪类细胞、犬类细胞、猫类细胞或啮齿动物细胞(例如,小鼠或大鼠)。如本文所述使用的哺乳动物细胞的实例包括,但不限于,中国仓鼠卵巢(CHO)细胞、人胚肾(HEK)293细胞、HeLa细胞和NSO细胞。在一些实施方式中,细胞是原核的。如本文所述使用的原核细胞的实例包括细菌细胞。细菌细胞可以是例如埃希氏菌属

(Escherichia spp.)(例如,大肠杆菌(Escherichia coli))、链球菌属(Streptococcu spp.)(例如,酿脓链球菌(Streptococcus pyogenes)、绿色链球菌(Streptococcus viridans)、肺炎链球菌(Streptococcus pneumoniae))、奈瑟氏球菌属(Neisseria spp.)(例如,Neisseria gibirrhoea、脑膜炎奈瑟氏球菌(Neisseria meningitidis))、棒杆菌属(Corynebacterium spp.)(例如,白喉棒杆菌(Corynebacterium diphtheriae))、芽孢杆菌属(Bacillis spp.)(例如,炭疽芽孢杆菌(Bacillis anthracis)、枯草芽孢杆菌(Bacillis subtilis))、乳杆菌属(Lactobacillus spp.)、梭菌属(Clostridium spp.)(例如,破伤风梭菌(Clostridium tetani)、产气荚膜梭菌(Clostridium perfringens)、诺氏梭菌(Clostridium novyii))、分枝杆菌属(Mycobacterium spp.)(例如,结核分枝杆菌(Mycobacterium tuberculosis))、志贺氏菌属(Shigella spp.)(例如,福氏志贺氏菌(Shigella flexneri)、痢疾志贺氏菌(Shigella dysenteriae))、沙门氏菌属(Salmonella spp.)(例如,伤寒沙门氏菌(Salmonella typhi)、肠炎沙门氏菌(Salmonella enteritidis))、克雷伯氏菌属(Klebsiella spp.)(例如,肺炎克雷伯氏菌(Klebsiella pneumoniae))、耶尔森氏菌属(Yersinia spp.)(例如,鼠疫耶尔森氏菌(Yersinia pestis))、沙雷氏菌属(Serratia spp.)(例如,粘质沙雷氏菌(Serratia marcescens))、假单胞菌属(Pseudomonas spp.)(例如,铜绿假单胞菌(Pseudomonas aeruginosa)、鼻疽假单胞菌(Pseudomonas mallei))、艾肯氏菌属(Eikenella spp.)(例如,啮蚀艾肯氏菌(Eikenella corrodens))、嗜血菌属(Haemophilus spp.)(例如,流感嗜血菌(Haemophilus influenza)、杜氏嗜血菌(Haemophilus ducreyi)、埃及嗜血菌(Haemophilus aegyptius))、弧菌属(Vibrio spp.)(例如,霍乱弧菌(Vibrio cholera)、需钠弧菌(Vibrio natriegens))、军团菌属(Legionella spp.)(例如,麦氏军团菌(Legionella micdadei)、博氏军团菌(Legionella bozemani))、布鲁氏菌属(Brucella spp.)(例如,流产布鲁氏菌(Brucella abortus))、枝原体属(Mycoplasma spp.)(例如,肺炎枝原体(Mycoplasma pneumoniae))或链霉菌属(Streptomyces spp.)(例如,天蓝色链霉菌(Streptomyces coelicolor)、浅青紫链霉菌(Streptomyces lividans)、白色链霉菌(Streptomyces albus))。

[0061] 递送和选择方法

[0062] 在一些实施方式中,本公开的方法包括将载体递送至包含细胞的组合物,和将该组合物维持在允许核酸(例如,第一、第二和第三载体)引入细胞中并允许核酸在细胞中表达以产生真核细胞的条件下。将核酸(例如,载体)引入细胞中所需的条件是公知的。这些条件包括例如(原核细胞的)转化条件、(真核细胞的)转染条件、(通过病毒/病毒载体的)转导条件和电穿孔条件,其中任何一种可以如本文所述使用。因此,在一些实施方式中,本公开的方法包括转染真核(哺乳动物)细胞,而在其他实施方式中,所述方法包括转化原核(例如,细菌)细胞。

[0063] 转基因细胞,例如,多转基因,如双、三和/或四转基因细胞的选择取决于所用的选择标记的类型。例如,如果选择标记蛋白是抗生素抗性蛋白,选择步骤可以包括将细胞暴露于特定的抗生素并且只选择存活的那些细胞。如果选择标记蛋白是荧光蛋白,选择步骤可以包括在显微镜下简单地观察细胞并选择发出荧光的细胞,或选择步骤可以包括其他荧光选择方法,如荧光激活细胞分选(FACS)。

[0064] 在一些实施方式中,用携带如本文所述的核酸的病毒载体(例如,病毒)转导细胞。在一些实施方式中,在转导(或其他转染方法)之前,将细胞以1×10^4至1×10^6/孔的密度接种于例如孔板(例如,12-孔板)上。在一些实施方式中,将100μL至500μL,例如,100、150、200、250、300、350、400、450或500μL的每种病毒载体加入每个孔中。

[0065] 试剂盒

[0066] 本公开还提供了可以用于例如产生和筛选转基因细胞和/或生物体的试剂盒。所述试剂盒可以包括本文所述的任意两种或更多种组分。例如,试剂盒可以包含(a)第一载体,其包含位于编码N-末端内含肽蛋白片段的核苷酸序列上游的编码第一选择标记蛋白片段的核苷酸序列;和(b)第二载体,其包含位于第二选择标记蛋白片段上游的编码C-末端内含肽蛋白片段的核苷酸序列,其中N-末端内含肽蛋白片段和C-末端内含肽蛋白片段催化第一选择标记蛋白片段与第二选择标记蛋白片段的接合,以产生全长抗生素抗性蛋白。

[0067] 在一些实施方式中,所述试剂盒包括本文所述的任意两种或更多种组分。例如,试剂盒可以包含(a)第一载体,其包含编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,(b)第二载体,其包含编码第一内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的中心片段的核苷酸序列的上游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(c)第三载体,其包含编码第二内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的C-末端片段的核苷酸序列的上游,其中第一内含肽的N-末端片段和C-末端片段催化抗生性抗性蛋白的N-末端片段与抗生素抗性蛋白的中心片段的接合,且第二内含肽的N-末端片段和C-末端片段催化抗生性抗性蛋白的中心片段与抗生素抗性蛋白的C-末端片段的接合,以产生全长抗生素抗性蛋白。

[0068] 在一些实施方式中,试剂盒进一步包含以下组分中的任意一种或多种:缓冲剂、盐、克隆酶(例如,LR克隆酶)、感受态细胞(例如,感受态细菌细胞)、转染试剂、抗生素和/或用于进行本文所述方法的说明。

[0069] 另外的实施方式

[0070] 本公开的另外的实施方式由以下编号的段落涵盖:

[0071] 1.一种方法,包括向真核细胞递送

[0072] (a)第一载体,其包含(i)编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码内含肽N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;和

[0073] (b)第二载体,其包含(i)编码内含肽C-末端片段的核苷酸序列,其位于抗生素抗性蛋白的C-末端片段的上游,和(ii)编码第二目标分子的核苷酸序列,

[0074] 其中内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的N-末端片段和C-末端片段的接合以产生全长抗生素抗性蛋白。

[0075] 2.段落1的方法,进一步包括将真核细胞维持在允许第一和第二载体引入真核细胞中以产生转基因真核细胞的条件下。

[0076] 3.段落2的方法,进一步包括选择包含全长抗生素抗性蛋白的转基因真核细胞。

[0077] 4.段落1-3任一项的方法,其中真核细胞是哺乳动物细胞。

[0078] 5.段落1-4任一项的方法,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤霉素、腐

草霉素D1或杀稻瘟菌素的抗性。

[0079] 6.段落1-5任一项的方法,其中内含肽是拆分内含肽。

[0080] 7.段落6的方法,其中拆分内含肽是天然拆分内含肽。

[0081] 8.段落7的方法,其中天然拆分内含肽选自DnaE内含肽。

[0082] 9.段落8的方法,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0083] 10.段落6的方法,其中拆分内含肽是工程化拆分内含肽。

[0084] 11.段落10的方法,其中工程化拆分内含肽是从DnaB内含肽工程化的。

[0085] 12.段落11的方法,其中工程化拆分内含肽是SspDnaB S1内含肽。

[0086] 13.段落12的方法,其中工程化拆分内含肽是从GyrB内含肽工程化的。

[0087] 14.段落13的方法,其中工程化拆分内含肽是SspGyrB S11内含肽。

[0088] 15.段落1-14任一项的方法,其中第一和/或第二分子是蛋白质。

[0089] 16.段落1-15任一项的方法,其中第一和/或第二分子是非编码核糖核酸(RNA)。

[0090] 17.段落16的方法,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0091] 18.段落1-17任一项的方法,其中第一和/或第二载体是质粒载体或病毒载体。

[0092] 19.一种方法,包括向真核细胞递送

[0093] (a)第一载体,其包含(i)hygB基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0094] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于hygB基因的C-末端片段上游,和(ii)第二目标分子,

[0095] 其中内含肽的N-末端片段和C-末端片段催化由hygB基因的N-末端片段编码的蛋白质片段与由hygB基因的C-末端片段编码的蛋白质片段的接合以产生全长潮霉素B磷酸转移酶。

[0096] 20.段落19的方法,其中由第二hygB基因片段编码的蛋白质片段的第一氨基酸是半胱氨酸。

[0097] 21.段落23的方法,其中

[0098] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-89识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸90-341识别的氨基酸序列;

[0099] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-200识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸201-341识别的氨基酸序列;

[0100] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-53识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸54-341识别的氨基酸序列;

[0101] 由hygB基因的N-末端片段编码的蛋白质片段包含由SEQ ID NO:1的氨基酸1-240识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含由SEQ ID NO:1的氨基酸241-341识别的氨基酸序列;

[0102] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-292识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸293-341识别的氨基酸序列。

[0103] 22.段落23-21任一项的方法,其中

[0104] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0105] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0106] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0107] 23.一种方法,包括向真核细胞递送

[0108] (a)第一载体,其包含(i)bsr基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0109] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于bsr基因的C-末端片段上游,和(ii)第二目标分子,

[0110] 其中内含肽的N-末端片段和C-末端片段催化由bsr基因的N-末端片段编码的蛋白质片段与由bsr基因的C-末端片段编码的蛋白质片段的接合以产生全长杀稻瘟菌素-S脱氨酶。

[0111] 24.段落23的方法,其中由bsr基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:4的氨基酸1-102识别的氨基酸序列,而由bsr基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:4的氨基酸103-140识别的氨基酸序列。

[0112] 25.段落22或23的方法,其中

[0113] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0114] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0115] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0116] 26.一种方法,包括向真核细胞递送

[0117] (a)第一载体,其包含(i)pac基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0118] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于pac基因的C-末端片段的上游,和(ii)第二目标分子,

[0119] 其中内含肽的N-末端片段和C-末端片段催化由pac基因的N-末端片段编码的蛋白质片段与由pac基因的C-末端片段编码的蛋白质片段的接合以产生全长嘌呤霉素N-乙酰基-转移酶。

[0120] 27.段落26的方法,其中

[0121] 由pac基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸1-63识别的氨基酸序列,而由pac基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的

氨基酸64-199识别的氨基酸序列;

[0122] 由pac基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸1-119识别的氨基酸序列,而由pac基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸120-199识别的氨基酸序列;

[0123] 由pac基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸1-100识别的氨基酸序列,而由pac基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸101-199识别的氨基酸序列。

[0124] 28.段落26或27的方法,其中

[0125] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0126] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0127] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0128] 29.一种方法,包括向真核细胞递送

[0129] (a)第一载体,其包含(i)neo基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0130] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于neo基因的C-末端片段的上游,和(ii)第二目标分子,

[0131] 其中内含肽的N-末端片段和C-末端片段催化由neo基因的N-末端片段编码的蛋白质片段与由neo基因的C-末端片段编码的蛋白质片段的接合以产生全长氨基糖苷3'-磷酸转移酶。

[0132] 30.段落29的方法,其中

[0133] 由neo基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸1-133识别的氨基酸序列,而由neo基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸134-267识别的氨基酸序列;或

[0134] 由neo基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸1-194识别的氨基酸序列,而由neo基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸195-267识别的氨基酸序列。

[0135] 31.段落29或30的方法,其中

[0136] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0137] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0138] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0139] 32.一种方法,包括向包含真核细胞的组合物递送

[0140] (a)第一载体,其包含(i)编码荧光蛋白的N-末端片段的核苷酸序列,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;和

[0141] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于荧光蛋白的C-末端片段的上游,和(ii)编码第二目标分子的核苷酸序列,

[0142] 其中内含肽的N-末端片段和C-末端片段催化荧光蛋白的N-末端片段与荧光蛋白的C-末端片段的接合以产生全长荧光蛋白。

[0143] 33.段落51的方法,进一步包括将真核细胞维持在允许第一和第二载体引入真核细胞中以产生转基因真核细胞的条件下。

[0144] 34.段落33的方法,进一步包括选择包含全长荧光蛋白的转基因真核细胞。

[0145] 35.段落32-34任一项的方法,其中真核细胞是哺乳动物细胞。

[0146] 36.段落32-35任一项的方法,其中荧光蛋白选自TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0147] 37.段落32-36任一项的方法,其中内含肽是拆分内含肽。

[0148] 38.段落37的方法,其中拆分内含肽是天然拆分内含肽。

[0149] 39.段落38的方法,其中天然拆分内含肽选自DnaE内含肽。

[0150] 40.段落39的方法,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0151] 41.段落40的方法,其中拆分内含肽是工程化拆分内含肽。

[0152] 42.段落41的方法,其中工程化拆分内含肽是从DnaB内含肽工程化的。

[0153] 43.段落42的方法,其中工程化拆分内含肽是SspDnaB S1内含肽。

[0154] 44.段落42的方法,其中工程化拆分内含肽是从GyrB内含肽工程化的。

[0155] 45.段落44的方法,其中工程化拆分内含肽是SspGyrB S11内含肽。

[0156] 46.段落32-35任一项的方法,其中第一和/或第二分子是蛋白质。

[0157] 47.段落32-46任一项的方法,其中第一和/或第二分子是非编码核糖核酸(RNA)。

[0158] 48.段落47的方法,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0159] 49.段落32-48任一项的方法,其中第一和/或第二载体是质粒载体或病毒载体。

[0160] 50.一种方法,包括向真核细胞递送

[0161] (a)第一载体,其包含(i)egfp基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0162] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于egfp基因的C-末端片段的上游,和(ii)第二目标分子,

[0163] 其中内含肽的N-末端片段和C-末端片段催化由egfp基因的N-末端片段编码的蛋白质片段与由egfp基因的C-末端片段编码的蛋白质片段的接合以产生EGFP蛋白。

[0164] 51.段落50的方法,其中由egfp基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:5的氨基酸1-175识别的氨基酸序列,而由egfp基因的C-末端片段编码的蛋白质片段

包含通过SEQ ID NO:5的氨基酸175-239识别的氨基酸序列。

[0165] 52.段落50或51的方法,其中

[0166] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0167] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0168] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0169] 53.一种方法,包括向真核细胞递送

[0170] (a)第一载体,其包含(i)mScarlet基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0171] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于mScarlet基因的C-末端片段上游,和(ii)第二目标分子,

[0172] 其中内含肽的N-末端片段和C-末端片段催化由mScarlet基因的N-末端片段编码的蛋白质片段与由mScarlet基因的C-末端片段编码的蛋白质片段的接合以产生全长mScarlet蛋白。

[0173] 54.段落53的方法,其中

[0174] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-46识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸47-232识别的氨基酸序列;

[0175] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-48识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸49-232识别的氨基酸序列;

[0176] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-51识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸52-232识别的氨基酸序列;

[0177] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-75识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸76-232识别的氨基酸序列;

[0178] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-122识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸123-232识别的氨基酸序列;

[0179] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-140识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸141-232识别的氨基酸序列;

[0180] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-163识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸164-232识别的氨基酸序列。

[0181] 55.段落53或54的方法,其中

[0182] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0183] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0184] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0185] 56.一种真核细胞,其包含

[0186] (a)第一载体,其包含(i)编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;和

[0187] (b)第二载体,其包含(i)编码内含肽的C-末端片段的核苷酸序列,其位于抗生素抗性蛋白的C-末端片段的上游,和(ii)编码第二目标分子的核苷酸序列,

[0188] 其中内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的N-末端片段和C-末端片段的接合以产生全长抗生素抗性蛋白。

[0189] 57.段落56的细胞,其中真核细胞是哺乳动物细胞。

[0190] 58.段落56或57的细胞,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤霉素、腐草霉素D1或杀稻瘟菌素的抗性。

[0191] 59.段落56-58任一项的细胞,其中内含肽是拆分内含肽。

[0192] 60.段落59的细胞,其中拆分内含肽是天然拆分内含肽。

[0193] 61.段落60的细胞,其中天然拆分内含肽选自DnaE内含肽。

[0194] 62.段落61的细胞,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0195] 63.段落59的细胞,其中拆分内含肽是工程化拆分内含肽。

[0196] 64.段落63的细胞,其中工程化拆分内含肽是从DnaB内含肽工程化的。

[0197] 65.段落64的细胞,其中工程化拆分内含肽是SspDnaB S1内含肽。

[0198] 66.段落65的细胞,其中工程化拆分内含肽是从GyrB内含肽工程化的。

[0199] 67.段落66的细胞,其中工程化拆分内含肽是SspGyrB S11内含肽。

[0200] 68.段落56-67任一项的细胞,其中第一和/或第二分子是蛋白质。

[0201] 69.段落56-68任一项的细胞,其中第一和/或第二分子是非编码核糖核酸(RNA)。

[0202] 70.段落69的细胞,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0203] 71.段落56-70任一项的细胞,其中第一和/或第二载体是质粒载体或病毒载体。

[0204] 72.一种细胞,其包含

[0205] (a)第一载体,其包含(i)hygB基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0206] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于hygB基因的C-末端片段的上游,和(ii)第二目标分子,

[0207] 其中内含肽的N-末端片段和C-末端片段催化由hygB基因的N-末端片段编码的蛋白质片段与由hygB基因的C-末端片段编码的蛋白质片段的接合以产生全长潮霉素B磷酸转

移酶。

[0208] 73.段落72的细胞,其中由第二hygB基因片段编码的蛋白质片段的第一氨基酸是半胱氨酸。

[0209] 74.段落73的细胞,其中

[0210] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-89识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸90-341识别的氨基酸序列;

[0211] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-200识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸201-341识别的氨基酸序列;

[0212] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-53识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸54-341识别的氨基酸序列;

[0213] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-240识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸241-341识别的氨基酸序列;

[0214] 由hygB基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸1-292识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:1的氨基酸293-341识别的氨基酸序列。

[0215] 75.段落72-74任一项的细胞,其中

[0216] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0217] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0218] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0219] 76.一种真核细胞,其包含

[0220] (a)第一载体,其包含(i)bsr基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0221] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于bsr基因的C-末端片段的上游,和(ii)第二目标分子,

[0222] 其中内含肽的N-末端片段和C-末端片段催化由bsr基因的N-末端片段编码的蛋白质片段与由bsr基因的C-末端片段编码的蛋白质片段的接合以产生全长杀稻瘟菌素-S脱氨酶。

[0223] 77.段落76的细胞,其中由bsr基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:4的氨基酸1-102识别的氨基酸序列,而由hygB基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:4的氨基酸103-140识别的氨基酸序列。

[0224] 78.段落76或77的细胞,其中

[0225] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID

NO:17识别;

[0226] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0227] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0228] 79.一种真核细胞,其包含

[0229] (a)第一载体,其包含(i)pac基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0230] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于pac基因的C-末端片段上游,和(ii)第二目标分子,

[0231] 其中内含肽的N-末端片段和C-末端片段催化由pac基因的N-末端片段编码的蛋白质片段与由pac基因的C-末端片段编码的蛋白质片段的接合以产生全长嘌呤霉素N-乙酰基-转移酶。

[0232] 80.段落79的细胞,其中

[0233] 由pac基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸1-63识别的氨基酸序列,而由pac基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸64-199识别的氨基酸序列;

[0234] 由pac基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸1-119识别的氨基酸序列,而由pac基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸120-199识别的氨基酸序列;

[0235] 由pac基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸1-100识别的氨基酸序列,而由pac基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:2的氨基酸101-199识别的氨基酸序列。

[0236] 81.段落79或80的细胞,其中

[0237] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0238] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0239] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0240] 82.一种真核细胞,其包含

[0241] (a)第一载体,其包含(i)neo基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0242] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于neo基因的C-末端片段上游,和(ii)第二目标分子,

[0243] 其中内含肽的N-末端片段和C-末端片段催化由neo基因的N-末端片段编码的蛋白质片段与由neo基因的C-末端片段编码的蛋白质片段的接合以产生全长氨基糖苷3’-磷酸转移酶。

[0244] 83.段落82的细胞,其中

[0245] 由neo基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸1-133识别的氨基酸序列,而由neo基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸134-267识别的氨基酸序列;或

[0246] 由neo基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸1-194识别的氨基酸序列,而由neo基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:3的氨基酸195-267识别的氨基酸序列.

[0247] 84.段落82或83的细胞,其中

[0248] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0249] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0250] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0251] 85.一种真核细胞,其包含

[0252] (a)第一载体,其包含(i)编码荧光蛋白的N-末端片段的核苷酸序列,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;和

[0253] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于荧光蛋白的C-末端片段的上游,和(ii)编码第二目标分子的核苷酸序列,

[0254] 其中内含肽的N-末端片段和C-末端片段催化荧光蛋白的N-末端片段与荧光蛋白的C-末端片段的接合以产生全长荧光蛋白。

[0255] 86.段落85的细胞,进一步包括将真核细胞维持在允许第一和第二载体引入真核细胞中以产生转基因真核细胞的条件下。

[0256] 87.段落86的细胞,进一步包括选择包含全长荧光蛋白的转基因真核细胞。

[0257] 88.段落85-87任一项的细胞,其中真核细胞是哺乳动物细胞。

[0258] 89.段落85-88任一项的细胞,其中荧光蛋白选自TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0259] 90.段落95-89任一项的细胞,其中内含肽是拆分内含肽。

[0260] 91.段落90的细胞,其中拆分内含肽是天然拆分内含肽。

[0261] 92.段落91的细胞,其中天然拆分内含肽选自DnaE内含肽。

[0262] 93.段落92的细胞,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0263] 94.段落93的细胞,其中拆分内含肽是工程化拆分内含肽。

[0264] 95.段落94的细胞,其中工程化拆分内含肽是从DnaB内含肽工程化的。

[0265] 96.段落95的细胞,其中工程化拆分内含肽是SspDnaB S1内含肽。

[0266] 97.段落95的细胞,其中工程化拆分内含肽是从GyrB内含肽工程化的。

[0267] 98.段落97的细胞,其中工程化拆分内含肽是SspGyrB S11内含肽。

[0268] 99.段落85-98任一项的细胞,其中第一和/或第二分子是蛋白质。

[0269] 100.段落85-99任一项的细胞,其中第一和/或第二分子是非编码核糖核酸(RNA)。

[0270] 101.段落100的细胞,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0271] 102.段落85-101任一项的细胞,其中第一和/或第二载体是质粒载体或病毒载体。

[0272] 103.一种真核细胞,其包含

[0273] (a)第一载体,其包含(i)egfp基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0274] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于efgp基因的C-末端片段的上游,和(ii)第二目标分子,

[0275] 其中内含肽的N-末端片段和C-末端片段催化由egfp基因的N-末端片段编码的蛋白质片段与由egfp基因的C-末端片段编码的蛋白质片段的接合以产生EGFP蛋白。

[0276] 104.段落103的细胞,其中由egfp基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:5的氨基酸1-175识别的氨基酸序列,而由egfp基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:5的氨基酸175-239识别的氨基酸序列。

[0277] 105.段落103或104的细胞,其中

[0278] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0279] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0280] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0281] 106.一种真核细胞,其包含

[0282] (a)第一载体,其包含(i)mScarlet基因的N-末端片段,其位于编码内含肽的N-末端片段的核苷酸序列的上游,和(ii)第一目标分子;和

[0283] (b)第二载体,其包含(ii)编码内含肽的C-末端片段的核苷酸序列,其位于mScarlet基因的C-末端片段的上游,和(ii)第二目标分子,

[0284] 其中内含肽的N-末端片段和C-末端片段催化由mScarlet基因的N-末端片段编码的蛋白质片段与由mScarlet基因的C-末端片段编码的蛋白质片段的接合以产生全长mScarlet蛋白。

[0285] 107.段落106的细胞,其中

[0286] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-46识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸47-232识别的氨基酸序列;

[0287] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-48识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸49-232识别的氨基酸序列;

[0288] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-51识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸52-232识别的氨基酸序列;

[0289] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-75识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸76-232识别的氨基酸序列;

[0290] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-122识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸123-232识别的氨基酸序列;

[0291] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-140识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸141-232识别的氨基酸序列;

[0292] 由mScarlet基因的N-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸1-163识别的氨基酸序列,而由mScarlet基因的C-末端片段编码的蛋白质片段包含通过SEQ ID NO:6的氨基酸164-232识别的氨基酸序列。

[0293] 108.段落106或107的细胞,其中

[0294] 内含肽的N-末端片段通过SEQ ID NO:16识别,而内含肽的C-末端片段通过SEQ ID NO:17识别;

[0295] 内含肽的N-末端片段通过SEQ ID NO:7识别,而内含肽的C-末端片段通过SEQ ID NO:8识别;或

[0296] 内含肽的N-末端片段通过SEQ ID NO:18或SEQ ID NO:9识别,而内含肽的C-末端片段通过SEQ ID NO:19或SEQ ID NO:10识别。

[0297] 109.一种组合物,其包含段落85-108任一项的细胞。

[0298] 110.一种试剂盒,其包含

[0299] (a)第一载体,其包含编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码内含肽的N-末端片段的核苷酸序列的上游;和

[0300] (b)第二载体,其包含编码内含肽的C-末端片段的核苷酸序列,其位于抗生素抗性蛋白的C-末端片段的上游,

[0301] 其中内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的N-末端片段和C-末端片段的接合以产生全长抗生素抗性蛋白。

[0302] 111.段落110的试剂盒,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤霉素、腐草霉素D1或杀稻瘟菌素的抗性。

[0303] 112.一种试剂盒,其包含

[0304] (a)第一载体,其包含编码荧光蛋白的N-末端片段的核苷酸序列,其位于编码内含肽的N-末端片段的核苷酸序列的上游;和

[0305] (b)第二载体,其包含编码内含肽的C-末端片段的核苷酸序列,其位于荧光蛋白的C-末端片段的上游,

[0306] 其中内含肽的N-末端片段和C-末端片段催化荧光蛋白的N-末端片段与荧光蛋白的C-末端片段的接合以产生全长荧光蛋白。

[0307] 113.段落112的试剂盒,其中荧光蛋白选自TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0308] 114.段落110-113任一项的试剂盒,其中内含肽是拆分内含肽。

[0309] 115.段落114的试剂盒,其中拆分内含肽是天然拆分内含肽或工程化拆分内含肽。

[0310] 116.段落115的试剂盒,其中天然拆分内含肽选自DnaE内含肽。

[0311] 117.段落116的试剂盒,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0312] 118.段落115的试剂盒,其中工程化拆分内含肽是从DnaB内含肽或GyrB内含肽工程化的。

[0313] 119.段落118的试剂盒,其中工程化拆分内含肽是SspDnaB S1内含肽。

[0314] 120.段落118的试剂盒,其中工程化拆分内含肽是是SspGyrB S11内含肽。

[0315] 121.段落112-120任一项的试剂盒,进一步包含以下组分中的任意一种或多种:缓冲剂、盐、克隆酶、感受态细胞、转染试剂、抗生素和/或用于进行本文所述方法的说明。

[0316] 122.一种方法,包括向真核细胞递送

[0317] (a)第一载体,其包含(i)编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

[0318] (b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的中心片段的核苷酸序列的上游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

[0319] (c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,

[0320] 其中第一内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的N-末端片段与抗生素抗性蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的中心片段与抗生素抗性蛋白的C-末端片段的接合,以产生全长抗生素抗性蛋白。

[0321] 123.段落112的方法,进一步包括将真核细胞维持在允许第一、第二和第三载体引入真核细胞中以产生转基因真核细胞的条件下。

[0322] 124.段落123的方法,进一步包括选择包含全长抗生素抗性蛋白的转基因真核细胞。

[0323] 125.段落112-124任一项的方法,其中真核细胞是哺乳动物细胞。

[0324] 126.段落112-125任一项的方法,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤

霉素、腐草霉素D1或杀稻瘟菌素的抗性。

[0325] 127.段落126的方法,其中抗生素抗性蛋白赋予对潮霉素的抗性。

[0326] 128.段落112-127任一项的方法,其中第一内含肽是拆分内含肽。

[0327] 129.段落112-128任一项的方法,其中第二内含肽是拆分内含肽。

[0328] 130.段落128或129的方法,其中拆分内含肽是天然拆分内含肽。

[0329] 131.段落130的方法,其中天然拆分内含肽选自DnaE内含肽。

[0330] 132.段落131的方法,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0331] 133.段落132的方法,其中第一内含肽是NpuDnaE内含肽,和第二内含肽是NpuDnaE内含肽。

[0332] 134.段落112-133任一项的方法,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是蛋白质。

[0333] 135.段落112-133任一项的方法,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是非编码核糖核酸(RNA)。

[0334] 136.段落135的方法,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0335] 137.段落112-136任一项的方法,其中第一载体、第二载体、第三载体或其任意组合是质粒载体或病毒载体。

[0336] 138.一种方法,包括向真核细胞递送

[0337] (a)第一载体,其包含(i)hgyB基因的N-末端片段,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

[0338] (b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于hgyB基因的中心片段的上游,该hgyB基因的中心片段位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

[0339] (c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于hgyB基因的C-末端片段的上游,和(ii)编码第三目标分子的核苷酸序列,

[0340] 其中第一内含肽的N-末端片段和C-末端片段催化由hgyB基因的N-末端片段编码的蛋白质片段与由hgyB基因的中心片段编码的蛋白质片段的接合,而第二内含肽的N-末端片段和C-末端片段催化由hgyB基因的中心片段编码的蛋白质片段与由hgyB基因的C-末端片段编码的蛋白质片段的接合,以产生全长潮霉素B磷酸转移酶。

[0341] 139.段落138的方法,其中第一载体编码通过SEQ ID NO:29识别的序列,第二载体编码通过SEQ ID NO:61识别的序列和第三载体通过SEQ ID NO:23识别的序列。

[0342] 140.段落138的方法,其中第一载体编码通过SEQ ID NO:21识别的序列,第二载体编码通过SEQ ID NO:61识别的序列和第三载体通过SEQ ID NO:35识别的序列。

[0343] 141.一种真核细胞,其包含

[0344] (a)第一载体,其包含(i)编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

[0345] (b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编

码抗生素抗性蛋白的中心片段的核苷酸序列的上游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

[0346] (c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,

[0347] 其中第一内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的N-末端片段与抗生素抗性蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的中心片段与抗生素抗性蛋白的C-末端片段的接合,以产生全长抗生素抗性蛋白。

[0348] 142.段落112的真核细胞,其中真核细胞是哺乳动物细胞。

[0349] 143.段落141或142的真核细胞,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤霉素、腐草霉素D1或杀稻瘟菌素的抗性。

[0350] 144.段落143的真核细胞,其中抗生素抗性蛋白赋予对潮霉素的抗性。

[0351] 145.段落141-144任一项的真核细胞,其中第一内含肽是拆分内含肽。

[0352] 146.段落142-145任一项的真核细胞,其中第二内含肽是拆分内含肽。

[0353] 147.段落145或146的真核细胞,其中拆分内含肽是天然拆分内含肽。

[0354] 148.段落147的真核细胞,其中天然拆分内含肽选自DnaE内含肽。

[0355] 149.段落148的真核细胞,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0356] 150.段落149的真核细胞,其中第一内含肽是NpuDnaE内含肽,和第二内含肽是NpuDnaE内含肽。

[0357] 151.段落142-150任一项的真核细胞,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是蛋白质。

[0358] 152.段落142-150任一项的真核细胞,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是非编码核糖核酸(RNA)。

[0359] 153.段落152的真核细胞,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0360] 154.段落142-153任一项的真核细胞,其中第一载体、第二载体、第三载体或其任意组合是质粒载体或病毒载体。

[0361] 155.一种组合物,其包含段落142-154任一项的真核细胞。

[0362] 156.一种试剂盒,其包含

[0363] (a)第一载体,其包含编码抗生素抗性蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,

[0364] (b)第二载体,其包含编码第一内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的中心片段的核苷酸序列的上游,该编码抗生素抗性蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和

[0365] (c)第三载体,其包含编码第二内含肽的C-末端片段的核苷酸序列,其位于编码抗生素抗性蛋白的C-末端片段的核苷酸序列的上游,

[0366] 其中第一内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的N-末端片段与抗生素抗性蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化抗生素抗性蛋白的中心片段与抗生素抗性蛋白的C-末端片段的接合,以产生全长抗生素抗性蛋白。

[0367] 157.段落156的试剂盒,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤霉素、腐草霉素D1或杀稻瘟菌素的抗性。

[0368] 158.段落157的试剂盒,其中抗生素抗性蛋白赋予对潮霉素的抗性。

[0369] 159.段落156-158任一项的试剂盒,其中第一内含肽是拆分内含肽。

[0370] 160.段落156-159任一项的试剂盒,其中第二内含肽是拆分内含肽。

[0371] 161.段落159或160的试剂盒,其中拆分内含肽是天然拆分内含肽。

[0372] 162.段落161的试剂盒,其中天然拆分内含肽选自DnaE内含肽。

[0373] 163.段落162的试剂盒,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0374] 164.段落163的试剂盒,其中第一内含肽是NpuDnaE内含肽,和第二内含肽是NpuDnaE内含肽。

[0375] 165.段落156-164任一项的试剂盒,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是蛋白质。

[0376] 166.段落156-164任一项的试剂盒,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是非编码核糖核酸(RNA)。

[0377] 167.段落166的试剂盒,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0378] 168.段落156-167任一项的试剂盒,其中第一载体、第二载体、第三载体或其任意组合是质粒载体或病毒载体。

[0379] 169.一种方法,包括向真核细胞递送

[0380] (a)第一载体,其包含(i)编码荧光蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

[0381] (b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码荧光蛋白的中心片段的核苷酸序列的上游,该编码荧光蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

[0382] (c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码荧光蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,

[0383] 其中第一内含肽的N-末端片段和C-末端片段催化荧光蛋白的N-末端片段与荧光蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化荧光蛋白的中心片段与荧光蛋白的C-末端片段的接合,以产生全长荧光蛋白。

[0384] 170.段落169的方法,进一步包括将真核细胞维持在允许第一、第二和第三载体引入真核细胞中以产生转基因真核细胞的条件下。

[0385] 171.段落170的方法,进一步包括选择包含全长荧光蛋白的转基因真核细胞。

[0386] 172.段落169-171任一项的方法,其中真核细胞是哺乳动物细胞。

[0387] 173.段落169-172任一项的方法,其中荧光蛋白选自TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0388] 174.段落173的方法,其中荧光蛋白是mScarlet。

[0389] 175.段落169-174任一项的方法,其中第一内含肽是拆分内含肽。

[0390] 176.段落169-175任一项的方法,其中第二内含肽是拆分内含肽。

[0391] 177.段落175或176的方法,其中拆分内含肽是天然拆分内含肽。

[0392] 178.段落177的方法,其中天然拆分内含肽选自DnaE内含肽。

[0393] 179.段落178的方法,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0394] 180.段落179的方法,其中第一内含肽是NpuDnaE内含肽,和第二内含肽是NpuDnaE内含肽。

[0395] 181.段落169-170任一项的方法,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是蛋白质。

[0396] 182.段落169-180任一项的方法,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是非编码核糖核酸(RNA)。

[0397] 183.段落182的方法,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0398] 184.段落169-183任一项的方法,其中第一载体、第二载体、第三载体或其任意组合是质粒载体或病毒载体。

[0399] 185.一种方法,包括向真核细胞递送

[0400] (a)第一载体,其包含(i)mScarlet基因的N-末端片段,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

[0401] (b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于mScarlet基因的中心片段的上游,该mScarlet基因的中心片段位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

[0402] (c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于mScarlet基因的C-末端片段的上游,和(ii)编码第三目标分子的核苷酸序列,

[0403] 其中第一内含肽的N-末端片段和C-末端片段催化由mScarlet基因的N-末端片段编码的蛋白质片段与由mScarlet基因的中心片段编码的蛋白质片段的接合,而第二内含肽的N-末端片段和C-末端片段催化由mScarlet基因的中心片段编码的蛋白质片段与由mScarlet基因的C-末端片段编码的蛋白质片段的接合,以产生全长mScarlet蛋白。

[0404] 186.段落185的方法,其中第一载体编码通过SEQ ID NO:121识别的序列,第二载体编码通过SEQ ID NO:123识别的序列和第三载体通过SEQ ID NO:125识别的序列。

[0405] 187.一种真核细胞,其包含:

[0406] (a)第一载体,其包含(i)编码荧光蛋白的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列,

[0407] (b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码荧光蛋白的中心片段的核苷酸序列的上游,该编码荧光蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,和

[0408] (c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码荧光蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,

[0409] 其中第一内含肽的N-末端片段和C-末端片段催化荧光蛋白的N-末端片段与荧光蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化荧光蛋白的中心片段与荧光蛋白的C-末端片段的接合,以产生全长荧光蛋白。

[0410] 188.段落187的真核细胞,其中真核细胞是哺乳动物细胞。

[0411] 189.段落187或188的真核细胞,其中荧光蛋白选自TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0412] 190.段落189的真核细胞,其中荧光蛋白是mScarlet。

[0413] 191.段落187-190任一项的真核细胞,其中第一内含肽是拆分内含肽。

[0414] 192.段落185-191任一项的真核细胞,其中第二内含肽是拆分内含肽。

[0415] 193.段落191或192的真核细胞,其中拆分内含肽是天然拆分内含肽。

[0416] 194.段落193的真核细胞,其中天然拆分内含肽选自DnaE内含肽。

[0417] 195.段落194的真核细胞,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0418] 196.段落195的真核细胞,其中第一内含肽是NpuDnaE内含肽,和第二内含肽是NpuDnaE内含肽。

[0419] 197.段落185-196任一项的真核细胞,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是蛋白质。

[0420] 198.段落185-196任一项的真核细胞,其中第一目标分子、第二目标分子、第三目标分子或其任意组合是非编码核糖核酸(RNA)。

[0421] 199.段落198的真核细胞,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0422] 200.段落185-199任一项的真核细胞,其中第一载体、第二载体、第三载体或其任意组合是质粒载体或病毒载体。

[0423] 201.一种组合物,包含段落185-200任一项的真核细胞。

[0424] 202.一种试剂盒,其包含:

[0425] (a)第一载体,其包含编码荧光蛋白的N-末端片段的核苷酸序列,其位于编码第一

内含肽的N-末端片段的核苷酸序列的上游，

[0426] (b)第二载体，其包含编码第一内含肽的C-末端片段的核苷酸序列，其位于编码荧光蛋白的中心片段的核苷酸序列的上游，该编码荧光蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游，和

[0427] (c)第三载体，其包含编码第二内含肽的C-末端片段的核苷酸序列，其位于编码荧光蛋白的C-末端片段的核苷酸序列的上游，

[0428] 其中第一内含肽的N-末端片段和C-末端片段催化荧光蛋白的N-末端片段与荧光蛋白的中心片段的接合，而第二内含肽的N-末端片段和C-末端片段催化荧光蛋白的中心片段与荧光蛋白的C-末端片段的接合，以产生全长荧光蛋白。

[0429] 203.段落202的试剂盒，其中荧光蛋白选自TagCFP、mTagCFP2、Czurite、ECFP2、mKalama1、Sirius、Sapphire、T-Sapphire、ECFP、Cerulean、SCFP3C、mTurquoise、mTurquoise2、单体Midoriishi-Cyan、TagCFP、mTFP1、EGFP、Emerald、Superfolder GFP、单体Czami Green、TagGFP2、mUKG、mWasabi、Clover、mNeonGreen、EYFP、Citrine、Venus、SYFP2、TagYFP、单体Kusabira-Orange、mKOκ、mKO2、mOrange、mOrange2、mRaspberry、mCherry、mStrawberry、mScarlet、mTangerine、tdTomato、TagRFP、TagRFP-T、mCpple、mRuby、mRuby2、mPlum、HcRed-Tandem、mKate2、mNeptune、NirFP、TagRFP657、IFP1.4和iRFP。

[0430] 204.段落203的试剂盒，其中荧光蛋白是mScarlet。

[0431] 205.段落202-204任一项的试剂盒，其中第一内含肽是拆分内含肽。

[0432] 206.段落202-205任一项的试剂盒，其中第二内含肽是拆分内含肽。

[0433] 207.段落206的试剂盒，其中拆分内含肽是天然拆分内含肽。

[0434] 208.段落207的试剂盒，其中天然拆分内含肽选自DnaE内含肽。

[0435] 209.段落208的试剂盒，其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0436] 210.段落209的试剂盒，其中第一内含肽是NpuDnaE内含肽，和第二内含肽是NpuDnaE内含肽。

[0437] 211.段落202-210任一项的试剂盒，其中第一目标分子、第二目标分子、第三目标分子或其任意组合是蛋白质。

[0438] 212.段落202-210任一项的试剂盒，其中第一目标分子、第二目标分子、第三目标分子或其任意组合是非编码核糖核酸(RNA)。

[0439] 213.段落212的试剂盒，其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)或短发夹RNA(shRNA)。

[0440] 214.段落202-213任一项的试剂盒，其中第一载体、第二载体、第三载体或其任意组合是质粒载体或病毒载体。

[0441] 215.段落202-214任一项的试剂盒，进一步包含以下组分中的任意一种或多种:缓冲剂、盐、克隆酶、感受态细胞、转染试剂、抗生素和/或用于进行本文所述方法的说明。

[0442] 216.一种转基因选择方法，包括向包含真核细胞的组合物递送:(a)第一载体，其包含(i)在编码N-末端内含肽蛋白片段的核苷酸序列上游的编码第一选择标记蛋白片段(例如，抗生素抗性蛋白片段或荧光蛋白片段)的核苷酸序列和(ii)编码第一分子的核苷酸序列，和(b)第二载体，其包含(i)第二选择标记蛋白片段(例如，抗生素抗性蛋白片段或荧

光蛋白片段)上游的编码C-末端内含肽蛋白片段的核苷酸序列,和(ii)编码第二分子的核苷酸序列,其中N-末端内含肽蛋白片段和C-末端内含肽蛋白片段催化第一选择标记蛋白片段与第二选择标记蛋白片段的接合以产生全长选择标记蛋白。

[0443] 217.一种转基因选择方法,包括向真核细胞递送(a)第一载体,其包含(i)编码选择标记蛋白(例如,抗生素抗性蛋白或荧光蛋白)的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;(b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的中心片段的核苷酸序列的上游,该编码选择标记蛋白的中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列;和(c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,其中第一内含肽的N-末端片段和C-末端片段催化选择标记蛋白的N-末端片段与选择标记蛋白的中心片段的接合,而第二内含肽的N-末端片段和C-末端片段催化选择标记蛋白的中心片段与选择标记蛋白的C-末端片段的接合,以产生全长选择标记蛋白。

[0444] 218.一种转基因选择方法,包括向真核细胞递送(a)第一载体,其包含(i)编码选择标记蛋白(例如,抗生素抗性蛋白或荧光蛋白)的N-末端片段的核苷酸序列,其位于编码第一内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第一目标分子的核苷酸序列;(b)第二载体,其包含(i)编码第一内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的第一中心片段的核苷酸序列的上游,该编码选择标记蛋白的第一中心片段的核苷酸序列位于编码第二内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第二目标分子的核苷酸序列,(c)第三载体,其包含(i)编码第二内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的第二中心片段的核苷酸序列的上游,该编码选择标记蛋白的第二中心片段的核苷酸序列位于编码第三内含肽的N-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列;和(d)第四载体,其包含(i)编码第三内含肽的C-末端片段的核苷酸序列,其位于编码选择标记蛋白的C-末端片段的核苷酸序列的上游,和(ii)编码第三目标分子的核苷酸序列,其中第一内含肽的N-末端片段和C-末端片段催化选择标记蛋白的N-末端片段与选择标记蛋白的第一中心片段的接合,第二内含肽的N-末端片段和C-末端片段催化选择标记蛋白的第一中心片段与选择标记蛋白的第二中心片段的接合,第三内含肽的N-末端片段和C-末端片段催化选择标记蛋白的第二中心片段与选择标记蛋白的C-末端片段的接合,以产生全长选择标记蛋白。

[0445] 219.段落216-218任一项的方法,进一步包括将真核细胞维持在允许将载体引入真核细胞中以产生转基因真核细胞的条件下。

[0446] 220.段落219的方法,进一步包括选择包含全长选择标记蛋白的转基因真核细胞。

[0447] 221.段落216-220任一项的方法,其中真核细胞是哺乳动物细胞。

[0448] 222.段落216-221任一项的方法,其中抗生素抗性蛋白赋予对潮霉素、G418、嘌呤霉素、腐草霉素D1或杀稻瘟菌素的抗性。

[0449] 223.段落216-222任一项的方法,其中内含肽是拆分内含肽。

[0450] 224.段落223的方法,其中拆分内含肽是天然拆分内含肽。

[0451] 225.段落224的方法,其中天然拆分内含肽选自DnaE内含肽。

[0452] 226.段落225的方法,其中DnaE内含肽选自集胞藻属DnaE(SspDnaE)内含肽和点形念珠藻(NpuDnaE)内含肽。

[0453] 227.段落223的方法,其中拆分内含肽是工程化拆分内含肽。

[0454] 228.段落2278的方法,其中工程化拆分内含肽是从DnaB内含肽工程化的。

[0455] 229.段落228的方法,其中工程化拆分内含肽是SspDnaB S1内含肽。

[0456] 230.段落229的方法,其中工程化拆分内含肽是从GyrB内含肽工程化的。

[0457] 231.段落230的方法,其中工程化拆分内含肽是SspGyrB S11内含肽。

[0458] 232.段落216-231任一项的方法,其中所述分子选自蛋白质。

[0459] 233.段落216-231任一项的方法,其中所述分子选自非编码核糖核酸(RNA)。

[0460] 234.段落233的方法,其中非编码RNA是微RNA(miRNA)、反义RNA、短干扰RNA(siRNA)和短发夹RNA(shRNA)。

[0461] 235.段落216-234任一项的方法,其中所述载体选自质粒载体和病毒载体。

[0462] 实施例

[0463] 通过以下实施例进一步说明本公开。提供这些实施例来帮助理解所述公开内容,并且不应当解释为是对其的限制。

[0464] 实施例1.抗生素抗性标记

[0465] 选择标记常常用于遗传工程化中以分离具有所需基因型的细胞[1]。然而,充分表征的用于真核细胞中的抗生素抗性基因的数量有限,并且可以通过普通实验室中的设备明确地区分其光谱的荧光蛋白的数量也有限。如果研究人员要将多个转基因整合到细胞中,他们经常会遇到没有足够的选择标记可以选择的问题。另一方面,同一时间使用多种抗生素进行选择常常对细胞有伤害。“选择标记再循环(selectable marker recycling)”可能提供一种暂时的解决方法,然而,需要多轮转基因、选择和选择标记的去除[2]。为了允许同一时间通过一个选择方案选择多个转基因,我们创建了断裂抗生素抗性和荧光蛋白基因,其中编码抗生素抗性或荧光蛋白的基因被拆分成两个或更多个与内含肽融合的片段(“markertron”),其可以通过蛋白质反式剪接重新接合[3](图1A)。将每个markertron插入到携带特定转基因的转基因载体上。含有一组markertron的转基因载体的递送产生了带有markertron的子集或完整集的细胞。只有含有完整markertron集的细胞才通过蛋白质剪接产生完全重建的标记蛋白并因此通过选择,而具有部分markertron集的细胞被消除,从而实现了含有全部预期转基因的细胞的共同选择。

[0466] 我们从用于双重转基因的2-markertron内含肽断裂抗性(Intres)基因的工程化开始。由于侧翼残基和局部蛋白折叠可能影响内含肽介导的反式剪接的效率,因此我们着手鉴定与两个充分表征的由NpuDnaE[4,5]和SspDnaB[6]衍生的拆分内含肽相容的四种常用抗生素抗性基因每一个中的拆分点。为了促进对双重转基因选择的有效性的评估,我们将markertron克隆至表达TagBFP2或mCherry荧光蛋白(作为测试转基因)的慢病毒载体上(图1B)。将病毒制剂转导至U2OS细胞中,然后将其分入含有非选择或选择培养基的重复平板中。经过适当的用于抗生素选择的传代后,通过流式细胞术分析了两种细胞培养物。对于潮霉素(Hygro)抗性基因,测试具有侧翼残基“GS”的一个“天然”SspDnaB拆分点(G200:S201)和具有“YC”残基的一个“天然”NspDnaE拆分点(Y89:C90)。当N-和C-markertron都转导时能够实现的成功选择在选择培养物中产生了>99%的BFP+mCherry+双转基因细胞,与

非选择培养物中<10％的双阳性细胞相比(图3;质粒对3、4和5、6)。用两个markertron中任一个转导的细胞没有在潮霉素选择中存活。相反,具有常规全长非断裂潮霉素载体的双重转基因只允许BFP+mCherry+细胞的～20％富集(质粒对97、98)。我们对于NpuDnaE筛选了三个另外的潜在拆分点(52S:53C)、(240A:241C)和(292R:293C),具有C-外显肽接合处的必需半胱氨酸残基和N-外显肽接合处之前报告7中支持显著反式剪接活性的残基。我们还通过在C-外显肽接合处插入“人工”半胱氨酸并入了六个另外的NpuDnaE拆分点以支持在异位位点的剪接,从而产生另外的拆分点。总之,测试的十一个拆分点中的八个支持潮霉素选择(图3)。相似地,对于嘌呤霉素(Puro)(图4)、新霉素(Neo)(图5)和杀稻瘟菌素(Blast)(图6)抗性基因,我们分别鉴定了四个、两个或一个功能性Intres对。在所有这些情况中,用任一markertron转导的细胞没有在选择中存活,而用两者转导的细胞在选择培养物中产生了>95％的双转基因细胞,与非选择培养物中<50％相比,除了杀稻瘟菌素(102)Intres,其获得了较低但仍然显著的91％双转基因细胞的富集(图3-6)。Intres基因的拆分点和质粒的详细内容呈现于图2A-2D和表1中。

[0467] 表1.质粒

[0468]

质 粒 #*	质粒名称	Markertron

[0469]

3	pLX-Hygro(1-89)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-89)-NpuDnaE(N)
4	pLX-NpuDnaE(C)-Hygro(90-341)-IRES-mCherry	NpuDnaE(C)-Hygro(90-341)
5	pLX-Hygro(1-200)-SspDnaB(N)-IRES-TagBFP2	Hygro(1-200)-SspDnaB(N)
6	pLX-SspDnaB(C)-Hygro(201-341)-IRES-mCherry	SspDnaB(C)-Hygro(201-341)
7	pLX-Hygro(1-52)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-52)-NpuDnaE(N)
8	pLX-NpuDnaE(C)-Hygro(53-341)-IRES-mCherry	NpuDnaE(C)-Hygro(53-341)
9	pLX-Hygro(1-240)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-240)-NpuDnaE(N)
10	pLX-NpuDnaE(C)-Hygro(241-341)-IRES-mCherry	NpuDnaE(C)-Hygro(241-341)
11	pLX-Hygro(1-292)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-292)-NpuDnaE(N)
12	pLX-NpuDnaE(C)-Hygro(293-341)-IRES-mCherry	NpuDnaE(C)-Hygro(293-341)
13	pLX-Blast(1-102)-NpuDnaE(N)-IRES-TagBFP2	Blast(1-102)-NpuDnaE(N)
14	pLX-NpuDnaE(C)-Blast(103-140)-IRES-mCherry	NpuDnaE(C)-Blast(103-140)
17	pLX-Puro(1-119)-NpuDnaE(N)-IRES-TagBFP2	Puro(1-119)-NpuDnaE(N)
18	pLX-NpuDnaE(C)-Puro(insCys;120-199)-IRES-mCherry	NpuDnaE(C)-Puro(insCys;120-199)
19	pLX-Puro(1-100)-SspDnaB(N-S0)-IRES-TagBFP2	Puro(1-100)-SspDnaB(N-S0)
20	pLX-SspDnaB(C-S0)-Puro(101-199)-IRES-mCherry	SspDnaB(C-S0)-Puro(101-199)
21	pLX-Neo(1-133)-NpuDnaE(N)-IRES-TagBFP2	Neo(1-133)-NpuDnaE(N)
22	pLX-NpuDnaE(C)-Neo(134-267)-IRES-mCherry	NpuDnaE(C)-Neo(134-267)
23	pLX-Neo(1-194)-NpuDnaE(N)-IRES-TagBFP2	Neo(1-194)-NpuDnaE(N)
24	pLX-NpuDnaE(C)-Neo(195-267)-IRES-mCherry	NpuDnaE(C)-Neo(195-267)
25	pLX-NpuDnaE(C)_Hygro(53-89)-NpuDnaE(N)-IRES-GFP	NpuDnaE(C)_Hygro(53-89)-NpuDnaE(N)
26	pLX-NpuDnaE(C)_Hygro(53-239)-NpuDnaE(N)-IRES-GFP	NpuDnaE(C)_Hygro(53-239)-NpuDnaE(N)
27	pCR8-BsaI->ccdbCam<-BsaI-NpuDnaE(N)-MD1-68-15	
28	pCR8-NpuDnaE(C)_BsaI->ccdbCam<-BsaI-MD1-68-18	
29	pCR8-BsaI->ccdbCam<-BsaI-SspDnaE(N)-MD1-68-12	
30	pCR8-SspDnaE(C)_BsaI->ccdbCam<-BsaI-MD1-68-13	
31	pCR8-BsaI->ccdbCam<-BsaI-SspDnaB(N-S0)-25-135-18	
32	pCR8-SspDnaB(C-S0)_BsaI->ccdbCam<-BsaI-25-155-41	
33	pLX-mScarlet(1-46)-NpuDnaE(N)_LZA-IRES-TagBFP2	mScarlet(1-46)-NpuDnaE(N)_LZA
34	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;47-232)-IRES-TagBFP2	LZB_NpuDnaE(C)-mScarlet(insCys;47-232)
35	pLX-mScarlet(1-48)-NpuDnaE(N)_LZA-IRES-TagBFP2	mScarlet(1-48)-NpuDnaE(N)_LZA
36	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;49-232)-IRES-GFP	LZB_NpuDnaE(C)-mScarlet(insCys;49-232)
37	pLX-mScarlet(1-51)-NpuDnaE(N)_LZA -IRES-TagBFP2	mScarlet(1-51)-NpuDnaE(N)_LZA

[0470]

38	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;52-232)-IRES-GFP	LZB_NpuDnaE(C)-mScarlet(insCys;52-232)
39	pLX-mScarlet(1-75)-NpuDnaE(N)_LZA-IRES-TagBFP2	mScarlet(1-75)-NpuDnaE(N)_LZA
40	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;76-232)-IRES-GFP	LZB_NpuDnaE(C)-mScarlet(insCys;76-232)
41	pLX-mScarlet(1-122)-NpuDnaE(N)_LZA-IRES-TagBFP2	mScarlet(1-122)-NpuDnaE(N)_LZA
42	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;123-232)-IRES-GFP	LZB_NpuDnaE(C)-mScarlet(insCys;123-232)
43	pLX-mScarlet(1-140)-NpuDnaE(N)_LZA-IRES-TagBFP2	mScarlet(1-140)-NpuDnaE(N)_LZA
44	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;141-232)-IRES-GFP	LZB_NpuDnaE(C)-mScarlet(insCys;141-232)
45	pLX-mScarlet(1-163)-NpuDnaE(N)_LZA-IRES-TagBFP2	mScarlet(1-163)-NpuDnaE(N)_LZA
46	pLX-LZB_NpuDnaE(C)-mScarlet(insCys;164-232)-IRES-GFP	LZB_NpuDnaE(C)-mScarlet(insCys;164-232)
47	pCR8-TagBFP2	TagBFP2
48	pCR8-mCherry	mCherry
49	pLX-DEST-IRES-Hygro(1-89)-NpuDnaE(N)	Hygro(1-89)-NpuDnaE(N)
50	pLX-DEST-IRES-NpuDnaE(C)-Hygro(90-341)	NpuDnaE(C)-Hygro(90-341)
51	pLX-[TagBFP2]-IRES-Hygro(1-89)-NpuDnaE(N)	Hygro(1-89)-NpuDnaE(N)
52	pLX-[mCherry]-IRES-NpuDnaE(C)-Hygro(90-341)	NpuDnaE(C)-Hygro(90-341)
53	pLX-DEST-IRES-Puro(1-119)-NpuDnaE(N)	Puro(1-119)-NpuDnaE(N)
54	pLX-DEST-IRES-NpuDnaE(C)-Puro(120-199)	NpuDnaE(C)-Puro(insCys;120-199)
55	pLX-[TagBFP2]-IRES-Puro(1-119)-NpuDnaE(N)	Puro(1-119)-NpuDnaE(N)
56	pLX-[mCherry]-IRES-NpuDnaE(C)-Puro(120-199)	NpuDnaE(C)-Puro(insCys;120-199)
57	pLX-DEST-IRES-Neo(1-194)-NpuDnaE(N)	Neo(1-194)-NpuDnaE(N)
58	pLX-DEST-IRES-NpuDnaE(C)-Neo(195-267)	NpuDnaE(C)-Neo(195-267)
59	pLX-[TagBFP2]-IRES-Neo(1-194)-NpuDnaE(N)	Neo(1-194)-NpuDnaE(N)
60	pLX-[mCherry]-IRES-NpuDnaE(C)-Neo(195-267)	NpuDnaE(C)-Neo(195-267)
64	pLX-Hygro(1-69)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-69)-NpuDnaE(N)
65	pLX-NpuDnaE(C)-Hygro(^C;70-341)-IRES-mCherry	NpuDnaE(C)-Hygro(^C;70-341)
66	pLX-Hygro(1-131)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-131)-NpuDnaE(N)
67	pLX-NpuDnaE(C)-Hygro(^C;132-341)-IRES-mCherry	NpuDnaE(C)-Hygro(^C;132-341)
68	pLX-Hygro(1-171)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-171)-NpuDnaE(N)
69	pLX-NpuDnaE(C)-Hygro(^C;172-341)-IRES-mCherry	NpuDnaE(C)-Hygro(^C;172-341)
70	pLX-Hygro(1-218)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-218)-NpuDnaE(N)
71	pLX-NpuDnaE(C)-Hygro(^C;219-341)-IRES-mCherry	NpuDnaE(C)-Hygro(^C;219-341)
72	pLX-Hygro(1-259)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-259)-NpuDnaE(N)

[0471]

73	pLX-NpuDnaE(C)-Hygro(^C;260-341)-IRES-mCherry	NpuDnaE(C)-Hygro(^C;260-341)
74	pLX-Hygro(1-277)-NpuDnaE(N)-IRES-TagBFP2	Hygro(1-277)-NpuDnaE(N)
75	pLX-NpuDnaE(C)-Hygro(^C;278-341)-IRES-mCherry	NpuDnaE(C)-Hygro(^C;278-341)
76	pLX-Puro(1-32)-NpuDnaE(N)-IRES-TagBFP2	Puro(1-32)-NpuDnaE(N)
77	pLX-NpuDnaE(C)-Puro(^C;33-199)-IRES-mCherry	NpuDnaE(C)-Puro(^C;33-199)
78	pLX-Puro(1-84)-NpuDnaE(N)-IRES-TagBFP2	Puro(1-84)-NpuDnaE(N)
79	pLX-NpuDnaE(C)-Puro(^C;85-199)-IRES-mCherry	NpuDnaE(C)-Puro(^C;85-199)
80	pLX-Puro(1-137)-NpuDnaE(N)-IRES-TagBFP2	Puro(1-137)-NpuDnaE(N)
81	pLX-NpuDnaE(C)-Puro(^C;138-199)-IRES-mCherry	NpuDnaE(C)-Puro(^C;138-199)
82	pLX-Puro(1-158)-NpuDnaE(N)-IRES-TagBFP2	Puro(1-158)-NpuDnaE(N)
83	pLX-NpuDnaE(C)-Puro(^C;159-199)-IRES-mCherry	NpuDnaE(C)-Puro(^C;159-199)
84	pLX-Puro(1-180)-NpuDnaE(N)-IRES-TagBFP2	Puro(1-180)-NpuDnaE(N)
85	pLX-NpuDnaE(C)-Puro(^C;181-199)-IRES-mCherry	NpuDnaE(C)-Puro(^C;181-199)
86	pLX-Blast(1-58)-NpuDnaE(N)-IRES-TagBFP2	Blast(1-58)-NpuDnaE(N)
87	pLX-NpuDnaE(C)-Blast(59-140)-IRES-mCherry	NpuDnaE(C)-Blast(59-140)
88	pLX-NpuDnaE(C)-HygroBA-SspDnaB(N-S0)-IRES-EGFP	NpuDnaE(C)-Hygro(53-200)-SspDnaB(N-S0)
89	pLX-SspDnaB(C-S0)-Hygro(201-341)-IRES-mCherry	SspDnaB(C-S0)-Hygro(201-341)
90	pLX-NpuDnaE(C)-Hygro(90-200)-SspDnaB(N-S0)-IRES-EGFP	NpuDnaE(C)-Hygro(90-200)-SspDnaB(N-S0)
91	pLX-Hygro(1-200)-SspDnaB(N-S0)-IRES-TagBFP2	Hygro(1-200)-SspDnaB(N-S0)
92	pLX-SspDnaB(C-S0)-Hygro(201-240)-NpuDnaE(N)-IRES-EGFP	SspDnaB(C-S0)-Hygro(201-240)-NpuDnaE(N)
93	pLX-SspDnaB(C-S0)-Hygro(201-292)-NpuDnaE(N)-IRES-EGFP	SspDnaB(C-S0)-Hygro(201-292)-NpuDnaE(N)
94	pLX-DEST-IRES-TagBFP2	
95	pLX-DEST-IRES-EGFP	
96	pLX-DEST-IRES-mCherry	
97	pLX-Hygro-IRES-TagBFP2	非断裂 Hygro
98	pLX-Hygro-IRES-mCherry	非断裂 Hygro
99	pLX-Puro-IRES-TagBFP2	非断裂 Puro
100	pLX-Puro-IRES-mCherry	非断裂 Puro
101	pLX-Hygro-IRES-EGFP	非断裂 Hygro
102	pLX-NLS_GFP-IRES-Hygro	非断裂 Hygro
103	pLX-LifeAct_mCherry-IRES-Hygro	非断裂 Hygro
104	pLX-NLS_GFP-IRES-Hygro(1-89)-NpuDnaE(N)	Hygro(1-89)-NpuDnaE(N)
105	pLX-LifeAct_mScarlet-IRES-NpuDnaE(C)-Hygro(90-341)	NpuDnaE(C)-Hygro(90-341)
106	pX330-AAVS1	

[0472]

107	pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-Blast-P2A-EGFP	非断裂 Blast
108	pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-Blast-P2A-mScarlet	非断裂 Blast
109	pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-Blast(1-102)_NpuDnaE(N)-P2A-EGFP	Blast(1-102)-NpuDnaE(N)
110	pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-NpuDnaE(C)_Blast(103-140)-P2A-mScarlet	NpuDnaE(C)-Blast(103-140)
111	pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-Hygro-P2A-NTR-E2A-EGFP	非断裂 Hygro
112	pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-Hygro-P2A-NTR-E2A-mCherry	非断裂 Hygro
113	pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-Hygro(1-89)-NpuDnaE(N)-P2A-NTR-E2A-EGFP	Hygro(1-89)-NpuDnaE(N)
114	pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-NpuDnaE(C)-Hygro(90-341)-P2A-NTR-E2A-mCherry	NpuDnaE(C)-Hygro(90-341)
115	pLX-Hygro(1-89)_NpuDnaE(N)_LZA-IRES-TagBFP2	Hygro(1-89)-NpuDnaE(N)-LZA
116	pLX-LZB_NpuDnaGEP(C)_Hygro(90-200)_SspDnaB(N-S0)-IRES-GFP	LZB-NpuDnaGEP(C)-Hygro(90-200)-SspDnaB(N-S0)
117	pLX-SspDnaB(C-S0)_Hygro(201-240)_NpuDnaE(N)_LZA-IRES-GFP	SspDnaB(C-S0)-Hygro(201-240)-NpuDnaE(N)-LZA
118	pLX-LZB_NpuDnaGEP(C)_Hygro(241-341)-IRES-mCherry	LZB-NpuDnaGEP(C)-Hygro(241-341)

[0473] 实施例2.Gateway相容的慢病毒载体

[0474] 为了促进Intres标记的采用,我们建立了Gateway相容的慢病毒载体用于方便的转基因的限制-连接非依赖性LR克隆酶重组8(图7A)。我们通过将TagBPF2和mCherry分别与N-和C-Intres载体重组测试了这些载体的功能性,并且发现双转基因细胞的稳定的选择(图7B)。Intres载体的一个潜在用途是在细胞中安置不同的荧光标记以标记不同的细胞区室。为了利用这种用途,我们通过Gateway重组在NLS-GFP和LifeAct-mScarlet 9(其分别标记核和F-肌动蛋白)中进行克隆,以获得常规全长(FL)非断裂潮霉素选择载体或2-markertron潮霉素Intres载体,并用任一组质粒转导细胞,接着进行抗生素选择(图7C)。用非断裂选择质粒转导的样品含有单一和双重标记的细胞两者,而用Intres质粒转导的细胞全部是双重标记的(图7C)。

[0475] 实施例3.荧光标记

[0476] 为了测试拆分荧光标记是否可以用于转基因选择,我们针对mScarlet荧光蛋白筛选NpuDnaE拆分点(图8A)并且鉴定了四个在mScarlet-门控群体中允许双转基因细胞>96%的富集的拆分点和允许>60%双转基因细胞富集的三个另外的拆分点,与非门控群体中<20%的双转基因细胞相比(图8B)。

[0477] 实施例4.高阶拆分标记

[0478] 使用针对2-markertron Intres基因鉴定的拆分点,我们着手对高阶拆分标记进行工程化。我们测试了拆分点的组合以将标记基因分割成三个或更多个markertron,以允许使用一种抗生素共同选择超过两个“未连锁的”转基因(图9A-9B)。为了鉴定允许这样的“Intres链”的拆分点的对,我们将3-拆分markertron克隆至三个慢病毒载体中,每个载体携带三个荧光转基因TagBFP2、EGFP或mCherry中的一个,这允许我们通过流式细胞术评价

选择的有效性(图9C)。由于潮霉素抗性基因是最长的,且提供了最多的用于测试的拆分点,我们聚焦于3-markertron潮霉素Intres的工程化。我们使用两个插入NpuDnaE内含肽测试了两个3-markertron潮霉素Intres,两个使用NpuDnaE用于第一内含肽的和SspDnaB用于第二内含肽,以及两个使用SspDnaB用于第一内含肽和NpuDnaE用于第二内含肽(图9D)。这六个3-markertron潮霉素Intres中的五个能够在潮霉素选择培养物中实现>97%的三重转基因选择,而剩余的一个能够实现80%的三重转基因选择,与非选择培养物中<15%的三转基因细胞相比。使用留一法(leave-one-out)转导的样品在潮霉素选择后没有产生任何活细胞,而用非拆分潮霉素载体转导的细胞在选择后仅产生7%的三转基因细胞。

[0479] 为利于3-markertron Intres的使用,我们建立了具有这些标记的Gateway相容性慢病毒载体(图10A)。通过将TagBFP(作为转基因1)、EGFP(作为转基因2)和mCherry(作为转基因3)重组至N-、M-和C-Intres Gateway目的载体中测试了这些载体的三个组,并用于转导U2OS细胞,其随后分割并在潮霉素选择或非选择培养基中培养(图10B)。在选择后两周,通过流式细胞术分析细胞。所有三组3-markertron潮霉素Intres质粒支持>99%的三转基因细胞选择,与非选择培养物中的<25%相比(图10C)。

[0480] 我们进一步测试了4-markertron潮霉素Intres基因的可行性(图11)。在此,我们使用了与亮氨酸拉链基序11融合的称为NpuDnaGEP 10的NpuDnaE内含肽的增强变体,其与SspDnaB内含肽组合。尽管所有四个含有组成markertron的质粒的转导产生了在潮霉素选择下存活的细胞,但留一法转导没有产生任何存活(表2)。

[0481] 表2.用(“+”)或未用(“-”)从所示质粒制备的慢病毒转导的细胞的存活

[0482]

	质粒115	质粒116	质粒117	质粒118	存活
样品1	+	+	+	+	是
样品2	-	+	+	+	否
样品3	+	-	+	+	否
样品4	+	+	-	+	否
样品5	+	+	+	-	否

[0483] 实施例5.AAAVS1基因座处的双等位基因敲入

[0484] CRISPR/Cas最近成为一种用于基因组工程和编辑的强有力的技术。尽管基于NHEJ介导的插入/缺失(indels)的基因敲除以高频率发生,但基于使用同源修复模板(也称为靶向构建体)的同源定向修复(HDR)的精确编辑和敲入是低效率的。我们测试了拆分选择标记是否可以用于富集在AAVS1基因座具有双等位基因敲入的细胞。我们构建了具有侧邻靶位点的同源臂的靶向构建体,并剪接受体2A肽以将markertron捕捉到宿主基因PPP1R12C的内含子1内。然而,在使用这些靶向构建体的CRISPR/Cas敲入实验和两周抗生素选择后,我们没有获得任何活的细胞(数据未显示)。我们怀疑宿主基因PPP1R12C的内源启动子可能不驱动markertron的充分表达以重构足够的抗生素抗性蛋白来抵抗抗生素的作用。因此,我们测试了通过TetO启动子来表达Intres markertron的替代策略,该启动子的活性可以通过强力霉素(dox)浓度进行滴定。为了允许比较Intres介导的双等位基因选择与全长(FL)非拆分选择标记,我们实施了几种不同的靶向构建体设计。首先,我们在组成型EF1a启动子下驱动全长(FL)抗性基因(例如Hygro)与rtTA一起表达,并在dox诱导型TetO启动子下驱动单独的测试Intres(例如Blast Intres)的表达(图12B,质粒109和110)。这允许比较相同构建

体内的全长和拆分选择标记。为了公平地比较由相同TetO启动子驱动的全长标记与拆分标记,我们构建了两个相似的质粒107和108(参见质粒109和110),其中全长抗生素抗性基因(Blast)位于TetO启动子的下游(图12A)。为实现双等位基因靶向的单细胞定量和证明将两个转基因并入两个AAVS1等位基因中的可行性,我们通过自切割2A肽在测试拆分或非拆分标记的下游附加EGFP和mScarlet荧光基因。类似地,为了测试Hygro Intres,我们交换了EF1a和TetO驱动的标记,从而将FL Hygro或Hygro Intres置于TetO的下游,而将FL Blast置于EF1a的下游(图12C-12D;质粒111-114)。我们共转染了含有Cas9和sgRNA靶向AAVS1的pX330-AAVS1(质粒106),且不同靶向构建体对用于HEK293T细胞,并在随后传代时分入一式三份的不含抗生素、含杀稻瘟菌素或含潮霉素的含强力霉素培养基中。在选择后两周,我们通过GFP和RFP荧光的流式细胞术测量分析了培养物的双等位基因靶向(图12E)。如所预期的,非选择的培养物带有小部分(<1%)的双等位基因敲入的GFP+/RFP+细胞(图12E;选择=无)。在靶向构建体上存在相应的FL抗生素抗性基因的抗生素选择产生了<30%的双等位基因敲入细胞(图12E;Blast:TC a,c,d;Hygro:TC a,b,c)。相反,在靶向构建体上存在相应Intres的抗生素选择产生了75%(图12E;Blast Intres:TC b)和88%(图12E;Hygro Intres:TC d)的双等位基因敲入细胞。

[0485] 在上述实施例中,我们对断裂抗生素抗性和荧光蛋白基因进行工程化,其允许两个或更多个“未连锁”转基因的选择。通过在选择标记处插入非天然残基,我们证明可以利用新的高效率拆分点,从而扩展可用于工程化的位置。我们在CRISPR/Cas9基因组编辑实验中证明了可以将拆分选择标记整合到慢病毒载体或基因靶向构建体中以实现具有双重转基因或双等位基因敲入的细胞的富集。通过组合两个或更多个拆分点,我们显示了可以生成3-和4-拆分标记以允许进行更高阶的转基因选择。甚至更高阶的拆分选择标记的未来开发可能使包含数十个转基因或靶向敲入的细胞“超工程化(hyper-engineering)”成为可能。

[0486] 材料和方法

[0487] 克隆

[0488] 为了对于每个markertron生成测试质粒,我们首先生成了包含其ORF的Gateway供体质粒,然后重组至具有TagBFP2(质粒94:pLX-DEST-IRES-TagBFP2)、EGFP(质粒95:pLX-DEST-IRES-EGFP)或mCherry(质粒96:pLX-DEST-IRES-mCherry)报告基因的慢病毒目的载体中,其通过去除嘌呤霉素抗性基因并将IRES荧光基因插入Gateway盒下游从pLX302(addgene.org/25896/)衍生。通过嵌套融合PCR程序生成markertron-ORF Gateway供体质粒以将内含肽与选择标记片段的编码序列结合,然后通过非序列-和连接-依赖性的克隆(SLIC)插入pCR8-GW-TOPO质粒中(Li,MZ&Elledge,SJ SLIC:a method for sequence-and ligation-independent cloning.Gene Synthesis:Methods and Protocols,51-59(2012)),或者PCR扩增选择标记的相关片段,然后通过SLIC插入含有内含肽序列的“支架”质粒(质粒27~32)中。对编码内含肽的DNA序列针对人进行密码子优化,并如GBlock(IDT)进行合成,具有编码NpuDnaE内含肽的AC1947GB、编码SspDnaB内含肽的AC1949GB。从含有这些标记的质粒扩增选择标记片段。质粒参见表1。

[0489] 细胞培养

[0490] 所有细胞均在含有10%胎牛血清(FBS)(Lonza),4% Glutamax(Gibco),1%丙酮

酸钠(Gibco)和青霉素-链霉素(Gibco)的Dulbecco's改良Eagle's培养基(DMEM)(Sigma)中培养。培养箱条件为37℃和5% CO_2。

[0491] 病毒产生

[0492] 使用Lipofectamine 3000将pLP1、pLP2和VSV-G的病毒包装混合物用每种慢病毒载体共转染到前一天以每孔1.2×10^6个细胞的浓度接种在6孔板中的Lenti-X 293T细胞(ClonTech)中。转染后6小时更换培养基,然后孵育过夜。转染后28小时,使用45uM PES过滤器过滤含有病毒的培养基上清液,然后保存在-80℃下直至使用。

[0493] 转导

[0494] 转导前一天,将靶细胞(HEK293T、MCF7、U2-OS)以每孔1.5×10^5个细胞的密度接种到12孔板中。转导之前,将培养基更换为含有10μg/mL聚凝胺的培养基,每孔1mL。将250μL的每种相应的病毒(对于添加两种病毒的实验样品,共500μL)添加到每个孔中,并孵育过夜。感染后24小时更换培养基。感染后4天,将细胞分入重复的平板中。感染后5天,将具有抗生素(潮霉素)的培养基添加到一个重复平板的每个相应孔中(另一个保持无选择)。抗生素选择持续2周,然后进行FACS分析。

[0495] 荧光激活细胞分选

[0496] 细胞用胰蛋白酶处理,悬浮在培养基中,然后在HP Z230工作站上在使用FACSDiVa软件(版本8)的LSRFortessa X-20(BD Bioscience)流式细胞仪上进行分析。每次运行收集五万个事件。

[0497] 构建体和序列

[0498] NpuDnaE(N)

[0499] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN(SEQ ID NO:7)

[0500] NpuDnaE(C)

[0501] IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN(SEQ ID NO:8)

[0502] SspDnaB(N-S0)

[0503] CISGDSLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRLGRTIKATANHRFLTIDGWKRLDELSLKEHIALPRKLESSSLQL(SEQ ID NO:9)

[0504] SspDnaB(C-S0)

[0505] SPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHN(SEQ ID NO:10)

[0506] NpuDnaE(N)-LZA

[0507] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPNGGGGSGSAQLEKELQALEKKLAQLEWENQALEKELAQ(SEQ ID NO:11)

[0508] LZB-NpuDnaGEP(C)

[0509] AQLKKKLQANKKELAQLKWKLQALKKKLAQGGGGSGSMIKIATRKYLGKQNVYDIGVGEPHNFALKNGFIASN(SEQ ID NO:12)NpuDnaGFP(C)

[0510] IKIATRKYLGKQNVYDIGVGEPHNFALKNGFIASN(SEQ ID NO:13)

[0511] LZA

[0512] AQLEKELQALEKKLAQLEWENQALEKELAQ(SEQ ID NO:14)

[0513] LZB

[0514] AQLKKKLQANKKELAQLKWKLQALKKKLAQ(SEQ ID NO:15)

[0515] SspDnaE(N)

[0516] CLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEVLEYELEDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLDAGTIK(SEQ ID NO:16)

[0517] SspDnaE(C)

[0518] VKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN(SEQ ID NO:17)

[0519] SspDnaB(N)

[0520] CISGDSLISLA(SEQ ID NO:18)

[0521] SspDnaB(C)

[0522] STGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRLGRTIKATANHRFLTIDGWKRLDELSLKEHIALPRKLESSSLQLSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHN(SEQ ID NO:19)

[0523] 质粒3:pLX-Hygro(1-89)-NpuDnaE(N)-IRES-TagBFP2

[0524] 蛋白质＝Hygro(1-89)-NpuDnaE(N)

[0525] 载体序列(SEQ ID NO:20)

[0526] 氨基酸序列(SEQ ID NO:21)

[0527] 质粒4:pLX-NpuDnaE(C)-Hygro(90-341)-IRES-mCherry

[0528] 蛋白质＝NpuDnaE(C)-Hygro(90-341)

[0529] 载体序列(SEQ ID NO:22)

[0530] 氨基酸序列(SEQ ID NO:23)

[0531] 质粒5:pLX-Hygro(1-200)-SspDnaB(N)-IRES-TagBFP2

[0532] 蛋白质＝Hygro(1-200)-SspDnaB(N)

[0533] 载体序列(SEQ ID NO:24)

[0534] 氨基酸序列(SEQ ID NO:25)

[0535] 质粒6:pLX-SspDnaB(C)-Hygro(201-341)-IRES-mCherry

[0536] 蛋白质＝SspDnaB(C)-Hygro(201-341)

[0537] 载体序列(SEQ ID NO:26)

[0538] 氨基酸序列(SEQ ID NO:27)

[0539] 质粒7:pLX-Hygro(1-52)-NpuDnaE(N)-IRES-TagBFP2

[0540] 蛋白质＝Hygro(1-52)-NpuDnaE(N)

[0541] 载体序列(SEQ ID NO:28)

[0542] 氨基酸序列(SEQ ID NO:29)

[0543] 质粒8:pLX-NpuDnaE(C)-Hygro(53-341)-IRES-mCherry

[0544] 蛋白质＝NpuDnaE(C)-Hygro(53-341)

[0545] 载体序列(SEQ ID NO:30)

[0546] 氨基酸序列(SEQ ID NO:31)

[0547] 质粒9:pLX-Hygro(1-240)-NpuDnaE(N)-IRES-TagBFP2

[0548] 蛋白质＝Hygro(1-240)-NpuDnaE(N)

[0549] 载体序列(SEQ ID NO:32)

[0550] 氨基酸序列(SEQ ID NO:33)

[0551] 质粒10:pLX-NpuDnaE(C)-Hygro(241-341)-IRES-mCherry

[0552] 蛋白质＝NpuDnaE(C)-Hygro(241-341)

[0553] 载体序列(SEQ ID NO:34)

[0554] 氨基酸序列(SEQ ID NO:35)

[0555] 质粒11:pLX-Hygro(1-292)-NpuDnaE(N)-IRES-TagBFP2

[0556] 蛋白质＝Hygro(1-292)-NpuDnaE(N)

[0557] 载体序列(SEQ ID NO:36)

[0558] 氨基酸序列(SEQ ID NO:37)

[0559] 质粒12:pLX-NpuDnaE(C)-Hygro(293-341)-IRES-mCherry

[0560] 蛋白质＝NpuDnaE(C)-Hygro(293-341)

[0561] 载体序列(SEQ ID NO:38)

[0562] 氨基酸序列(SEQ ID NO:39)

[0563] 质粒13:pLX-Blast(1-102)-NpuDnaE(N)-IRES-TagBFP2

[0564] 蛋白质＝Blast(1-102)-NpuDnaE(N)

[0565] 载体序列(SEQ ID NO:40)

[0566] 氨基酸序列(SEQ ID NO:41)

[0567] 质粒14:pLX-NpuDnaE(C)-Blast(103-140)-IRES-mCherry

[0568] 蛋白质＝NpuDnaE(C)-Blast(103-140)

[0569] 载体序列(SEQ ID NO:42)

[0570] 氨基酸序列(SEQ ID NO:43)

[0571] 质粒17:pLX-Puro(1-119)-NpuDnaE(N)-IRES-TagBFP2

[0572] 蛋白质＝Puro(1-119)-NpuDnaE(N)

[0573] 载体序列(SEQ ID NO:44)

[0574] 氨基酸序列(SEQ ID NO:45)

[0575] 质粒18:pLX-NpuDnaE(C)-Puro(insCys;120-199)-IRES-mCherry

[0576] 蛋白质＝NpuDnaE(C)-Puro(insCys;120-199)

[0577] 载体序列(SEQ ID NO:46)

[0578] 氨基酸序列(SEQ ID NO:47)

[0579] 质粒19:pLX-Puro(1-100)-SspDnaB(N-S0)-IRES-TagBFP2

[0580] 蛋白质＝Puro(1-100)-SspDnaB(N-S0)

[0581] 载体序列(SEQ ID NO:48)

[0582] 氨基酸序列(SEQ ID NO:49)

[0583] 质粒20:pLX-SspDnaB(C-S0)-Puro(101-199)-IRES-mCherry

[0584] 蛋白质＝SspDnaB(C-S0)-Puro(101-199)

[0585] 载体序列(SEQ ID NO:50)

[0586] 氨基酸序列(SEQ ID NO:51)

[0587] 质粒21:pLX-Neo(1-133)-NpuDnaE(N)-IRES-TagBFP2

[0588] 蛋白质＝Neo(1-133)-NpuDnaE(N)

[0589] 载体序列(SEQ ID NO:52)

[0590] 氨基酸序列(SEQ ID NO:53)

[0591] 质粒22:pLX-NpuDnaE(C)-Neo(134-267)-IRES-mCherry

[0592] 蛋白质＝NpuDnaE(C)-Neo(134-267)

[0593] 载体序列(SEQ ID NO:54)

[0594] 氨基酸序列(SEQ ID NO:55)

[0595] 质粒23:pLX-Neo(1-194)-NpuDnaE(N)-IRES-TagBFP2

[0596] 蛋白质＝Neo(1-194)-NpuDnaE(N)

[0597] 载体序列(SEQ ID NO:56)

[0598] 氨基酸序列(SEQ ID NO:57)

[0599] 质粒24:pLX-NpuDnaE(C)-Neo(195-267)-IRES-mCherry

[0600] 蛋白质＝NpuDnaE(C)-Neo(195-267)

[0601] 载体序列(SEQ ID NO:58)

[0602] 氨基酸序列(SEQ ID NO:59)

[0603] 质粒25:pLX-NpuDnaE(C)_Hygro(53-89)-NpuDnaE(N)-IRES-GFP

[0604] 蛋白质＝NpuDnaE(C)_Hygro(53-89)-NpuDnaE(N)

[0605] 载体序列(SEQ ID NO:60)

[0606] 氨基酸序列(SEQ ID NO:61)

[0607] 质粒26:pLX-NpuDnaE(C)_Hygro(53-239)-NpuDnaE(N)-IRES-GFP

[0608] 蛋白质＝NpuDnaE(C)_Hygro(53-239)-NpuDnaE(N)

[0609] 载体序列(SEQ ID NO:62)

[0610] 氨基酸序列(SEQ ID NO:63)

[0611] 质粒27:pCR8-BsaI->ccdbCam<-BsaI-NpuDnaE(N)-MD1-68-15(SEQ ID NO:64)

[0612] 质粒28:pCR8-NpuDnaE(C)_BsaI->ccdbCam<-BsaI-MD1-68-18(SEQ ID NO:65)

[0613] 质粒29:pCR8-BsaI->ccdbCam<-BsaI-SspDnaE(N)-MD1-68-12(SEQ ID NO:66)

[0614] 质粒30:pCR8-SspDnaE(C)_BsaI->ccdbCam<-BsaI-MD1-68-13(SEQ ID NO:67)

[0615] 质粒31:pCR8-BsaI->ccdbCam<-BsaI-SspDnaB(N-S0)-25-135-18(SEQ ID NO:68)

[0616] 质粒32:pCR8-SspDnaB(C-S0)_BsaI->ccdbCam<-BsaI-25-155-41(SEQ ID NO:69)

[0617] 质粒33:pLX-mScarlet(1-46)-NpuDnaE(N)_LZA-IRES-TagBFP2

[0618] 蛋白质＝mScarlet(1-46)-NpuDnaE(N)_LZA

[0619] 载体序列(SEQ ID NO:70)

[0620] 氨基酸序列(SEQ ID NO:71)

[0621] 质粒34:

[0622] pLX-LZB_NpuDnaE(C)-mScarlet(insCys;47-232)-IRES-TagBFP2

[0623] 蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;47-232)

[0624] 载体序列(SEQ ID NO:72)

[0625] 氨基酸序列(SEQ ID NO:73)

[0626] 质粒35:pLX-mScarlet(1-48)-NpuDnaE(N)_LZA-IRES-TagBFP2

[0627] 蛋白质＝mScarlet(1-48)-NpuDnaE(N)_LZA

[0628] 载体序列(SEQ ID NO:74)

[0629] 氨基酸序列(SEQ ID NO:75)

[0630] 质粒36:pLX-LZB_NpuDnaE(C)-mScarlet(insCys;49-232)-IRES-GFP

[0631] 蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;49-232)

[0632] 载体序列(SEQ ID NO:76)

[0633] 氨基酸序列(SEQ ID NO:77)

[0634] 质粒37:pLX-mScarlet(1-51)-NpuDnaE(N)_LZA-IRES-TagBFP2蛋白质＝mScarlet(1-51)-NpuDnaE(N)_LZA

[0635] 载体序列(SEQ ID NO:78)

[0636] 氨基酸序列(SEQ ID NO:79)

[0637] 质粒38:pLX-LZB_NpuDnaE(C)-mScarlet(insCys;52-232)-IRES-GFP蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;52-232)

[0638] 载体序列(SEQ ID NO:80)

[0639] 氨基酸序列(SEQ ID NO:81)

[0640] 质粒39:pLX-mScarlet(1-75)-NpuDnaE(N)_LZA-IRES-TagBFP2蛋白质＝mScarlet(1-75)-NpuDnaE(N)_LZA

[0641] 载体序列(SEQ ID NO:82)

[0642] 氨基酸序列(SEQ ID NO:83)

[0643] 质粒40:pLX-LZB_NpuDnaE(C)-mScarlet(insCys;76-232)-IRES-GFP蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;76-232)

[0644] 载体序列(SEQ ID NO:84)

[0645] 氨基酸序列(SEQ ID NO:85)

[0646] 质粒41:pLX-mScarlet(1-122)-NpuDnaE(N)_LZA-IRES-TagBFP2蛋白质＝mScarlet(1-122)-NpuDnaE(N)_LZA

[0647] 载体序列(SEQ ID NO:86)

[0648] 氨基酸序列(SEQ ID NO:87)

[0649] 质粒42:pLX-LZB_NpuDnaE(C)-mScarlet(insCys;123-232)-IRES-GFP蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;123-232)

[0650] 载体序列(SEQ ID NO:88)

[0651] 氨基酸序列(SEQ ID NO:89)

[0652] 质粒43:pLX-mScarlet(1-140)-NpuDnaE(N)_LZA-IRES-TagBFP2蛋白质＝mScarlet(1-140)-NpuDnaE(N)_LZA

[0653] 载体序列(SEQ ID NO:90)

[0654] 氨基酸序列(SEQ ID NO:91)

[0655] 质粒44:pLX-LZB_NpuDnaE(C)-mScarlet(insCys;141-232)-IRES-GFP蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;141-232)

[0656] 载体序列(SEQ ID NO:92)

[0657] 氨基酸序列(SEQ ID NO:93)

[0658] 质粒45:pLX-mScarlet(1-163)-NpuDnaE(N)_LZA-IRES-TagBFP2蛋白质＝mScarlet(1-163)-NpuDnaE(N)_LZA

[0659] 载体序列(SEQ ID NO:94)

[0660] 氨基酸序列(SEQ ID NO:95)

[0661] 质粒46:pLX-LZB_NpuDnaE(C)-mScarlet(insCys;164-232)-IRES-GFP蛋白质＝LZB_NpuDnaE(C)-mScarlet(insCys;164-232)

[0662] 载体序列(SEQ ID NO:96)

[0663] 氨基酸序列(SEQ ID NO:97)

[0664] 质粒47:pCR8-TagBFP2

[0665] 蛋白质＝TagBFP2

[0666] 载体序列(SEQ ID NO:98)

[0667] 氨基酸序列(SEQ ID NO:99)

[0668] 质粒48:pCR8-mCherry

[0669] 蛋白质＝mCherry

[0670] 载体序列(SEQ ID NO:100)

[0671] 氨基酸序列(SEQ ID NO:101)

[0672] 质粒49:pLX-DEST-IRES-Hygro(1-89)-NpuDnaE(N)

[0673] 蛋白质＝Hygro(1-89)-NpuDnaE(N)

[0674] 载体序列(SEQ ID NO:102)

[0675] 氨基酸序列(SEQ ID NO:103)

[0676] 质粒50:pLX-DEST-IRES-NpuDnaE(C)-Hygro(90-341)

[0677] 蛋白质＝NpuDnaE(C)-Hygro(90-341)

[0678] 载体序列(SEQ ID NO:104)

[0679] 氨基酸序列(SEQ ID NO:105)

[0680] 质粒51:pLX-[TagBFP2]-IRES-Hygro(1-89)-NpuDnaE(N)

[0681] 载体序列(SEQ ID NO:106)

[0682] 质粒52:pLX-[mCherry]-IRES-NpuDnaE(C)-Hygro(90-341)

[0683] 载体序列(SEQ ID NO:107)

[0684] 质粒53:pLX-DEST-IRES-Puro(1-119)-NpuDnaE(N)

[0685] 蛋白质＝Puro(1-119)-NpuDnaE(N)

[0686] 载体序列(SEQ ID NO:108)

[0687] 氨基酸序列(SEQ ID NO:109)

[0688] 质粒54:pLX-DEST-IRES-NpuDnaE(C)-Puro(120-199)

[0689] 蛋白质＝NpuDnaE(C)-Puro(120-199)

[0690] 载体序列(SEQ ID NO:110)

[0691] 氨基酸序列(SEQ ID NO:111)

[0692] 质粒55:pLX-[TagBFP2]-IRES-Puro(1-119)-NpuDnaE(N)

[0693] 载体序列(SEQ ID NO:112)

[0694] 质粒56:pLX-[mCherry]-IRES-NpuDnaE(C)-Puro(120-199)

[0695] 载体序列(SEQ ID NO:113)

[0696] 质粒57:pLX-DEST-IRES-Neo(1-194)-NpuDnaE(N)

[0697] 蛋白质＝Neo(1-194)-NpuDnaE(N)

[0698] 载体序列(SEQ ID NO:114)

[0699] 氨基酸序列(SEQ ID NO:115)

[0700] 质粒58:pLX-DEST-IRES-NpuDnaE(C)-Neo(195-267)

[0701] 蛋白质＝NpuDnaE(C)-Neo(195-267)

[0702] 载体序列(SEQ ID NO:116)

[0703] 氨基酸序列(SEQ ID NO:117)

[0704] 质粒59:pLX-[TagBFP2]-IRES-Neo(1-194)-NpuDnaE(N)

[0705] 载体序列(SEQ ID NO:118)

[0706] 质粒60:pLX-[mCherry]-IRES-NpuDnaE(C)-Neo(195-267)

[0707] 载体序列(SEQ ID NO:119)

[0708] 质粒61:pLX-mScarlet(1-51)-NpuDnaE(N)-LZA-IRES-TagBFP2

[0709] 蛋白质＝mScarlet(1-51)-NpuDnaE(N)-LZA

[0710] 载体序列(SEQ ID NO:120)

[0711] 氨基酸序列(SEQ ID NO:121)

[0712] 质粒62:

[0713] pLX-LZB-NpuDnaE(C)-mScarlet(^C,52-163)-NpuDnaE(N)_LZA-IRES-EGFP

[0714] 蛋白质＝LZB-NpuDnaE(C)-mScarlet(^C;52-163)-NpuDnaE(N)_LZA

[0715] 载体序列(SEQ ID NO:122)

[0716] 氨基酸序列(SEQ ID NO:123)

[0717] 质粒63:pLX-LZB-NpuDnaE(C)-mScarlet(^C;164-232)-IRES-EGFP

[0718] 蛋白质＝LZB-NpuDnaE(C)-mScarlet(^C;164-232)

[0719] 载体序列(SEQ ID NO:124)

[0720] 氨基酸序列(SEQ ID NO:125)

[0721] 质粒64:pLX-Hygro(1-69)-NpuDnaE(N)-IRES-TagBFP2

[0722] 蛋白质＝Hygro(1-69)-NpuDnaE(N)

[0723] 载体序列(SEQ ID NO:126)

[0724] 氨基酸序列(SEQ ID NO:127)

[0725] 质粒65:pLX-NpuDnaE(C)-Hygro(^C;70-341)-IRES-mCherry

[0726] 蛋白质＝NpuDnaE(C)-Hygro(^C;70-341)

[0727] 载体序列(SEQ ID NO:128)

[0728] 氨基酸序列(SEQ ID NO:129)

[0729] 质粒66:pLX-Hygro(1-131)-NpuDnaE(N)-IRES-TagBFP2

[0730] 蛋白质＝Hygro(1-131)-NpuDnaE(N)

[0731] 载体序列(SEQ ID NO:130)

[0732] 氨基酸序列(SEQ ID NO:131)

[0733] 质粒67:pLX-NpuDnaE(C)-Hygro(^C;132-341)-IRES-mCherry

[0734] 蛋白质＝NpuDnaE(C)-Hygro(^C;132-341)

[0735] 载体序列(SEQ ID NO:132)

[0736] 氨基酸序列(SEQ ID NO:133)

[0737] 质粒68:pLX-Hygro(1-171)-NpuDnaE(N)-IRES-TagBFP2

[0738] 蛋白质＝Hygro(1-171)-NpuDnaE(N)

[0739] 载体序列(SEQ ID NO:134)

[0740] 氨基酸序列(SEQ ID NO:135)

[0741] 质粒69:pLX-NpuDnaE(C)-Hygro(^C;172-341)-IRES-mCherry

[0742] 蛋白质＝NpuDnaE(C)-Hygro(^C;172-341)

[0743] 载体序列(SEQ ID NO:136)

[0744] 氨基酸序列(SEQ ID NO:137)

[0745] 质粒70:pLX-Hygro(1-218)-NpuDnaE(N)-IRES-TagBFP2

[0746] 蛋白质＝Hygro(1-218)-NpuDnaE(N)

[0747] 载体序列(SEQ ID NO:138)

[0748] 氨基酸序列(SEQ ID NO:139)

[0749] 质粒71:pLX-NpuDnaE(C)-Hygro(^C;219-341)-IRES-mCherry

[0750] 蛋白质＝NpuDnaE(C)-Hygro(^C;219-341)

[0751] 载体序列(SEQ ID NO:140)

[0752] 氨基酸序列(SEQ ID NO:141)

[0753] 质粒72:pLX-Hygro(1-259)-NpuDnaE(N)-IRES-TagBFP2

[0754] 蛋白质＝Hygro(1-259)-NpuDnaE(N)

[0755] 载体序列(SEQ ID NO:142)

[0756] 氨基酸序列(SEQ ID NO:143)

[0757] 质粒73:pLX-NpuDnaE(C)-Hygro(^C;260-341)-IRES-mCherry

[0758] 蛋白质＝NpuDnaE(C)-Hygro(^C;260-341)

[0759] 载体序列(SEQ ID NO:144)

[0760] 氨基酸序列(SEQ ID NO:145)

[0761] 质粒74:pLX-Hygro(1-277)-NpuDnaE(N)-IRES-TagBFP2

[0762] 蛋白质＝Hygro(1-277)-NpuDnaE(N)

[0763] 载体序列(SEQ ID NO:146)

[0764] 氨基酸序列(SEQ ID NO:147)

[0765] 质粒75:pLX-NpuDnaE(C)-Hygro(^C;278-341)-IRES-mCherry

[0766] 蛋白质＝NpuDnaE(C)-Hygro(^C;278-341)

[0767] 载体序列(SEQ ID NO:148)

[0768] 氨基酸序列(SEQ ID NO:149)

[0769] 质粒76:pLX-Puro(1-32)-NpuDnaE(N)-IRES-TagBFP2

[0770] 蛋白质＝Puro(1-32)-NpuDnaE(N)

[0771] 载体序列(SEQ ID NO:150)

[0772] 氨基酸序列(SEQ ID NO:151)

[0773] 质粒77:pLX-NpuDnaE(C)-Puro(^C;33-199)-IRES-mCherry

[0774] 蛋白质＝NpuDnaE(C)-Puro(^C;33-199)

[0775] 载体序列(SEQ ID NO:152)

[0776] 氨基酸序列(SEQ ID NO:153)

[0777] 质粒78:pLX-Puro(1-84)-NpuDnaE(N)-IRES-TagBFP2

[0778] 蛋白质＝Puro(1-84)-NpuDnaE(N)

[0779] 载体序列(SEQ ID NO:154)

[0780] 氨基酸序列(SEQ ID NO:155)

[0781] 质粒79:pLX-NpuDnaE(C)-Puro(^C;85-199)-IRES-mCherry

[0782] 蛋白质＝NpuDnaE(C)-Puro(^C;85-199)

[0783] 载体序列(SEQ ID NO:156)

[0784] 氨基酸序列(SEQ ID NO:157)

[0785] 质粒80:pLX-Puro(1-137)-NpuDnaE(N)-IRES-TagBFP2

[0786] 蛋白质＝Puro(1-137)-NpuDnaE(N)

[0787] 文件＝pLX-[PuroKC3(N)-NpuDnaE(N)-25-131-29″]-IRES-TagBFP2-25-133-6

[0788] 载体序列(SEQ ID NO:158)

[0789] 氨基酸序列(SEQ ID NO:159)

[0790] 质粒81:pLX-NpuDnaE(C)-Puro(^C;138-199)-IRES-mCherry

[0791] 蛋白质＝NpuDnaE(C)-Puro(^C;138-199)

[0792] 载体序列(SEQ ID NO:160)

[0793] 氨基酸序列(SEQ ID NO:161)

[0794] 质粒82:pLX-Puro(1-158)-NpuDnaE(N)-IRES-TagBFP2

[0795] 蛋白质＝Puro(1-158)-NpuDnaE(N)

[0796] 载体序列(SEQ ID NO:162)

[0797] 氨基酸序列(SEQ ID NO:163)

[0798] 质粒83:pLX-NpuDnaE(C)-Puro(^C;159-199)-IRES-mCherry

[0799] 蛋白质＝NpuDnaE(C)-Puro(^C;159-199)

[0800] 载体序列(SEQ ID NO:164)

[0801] 氨基酸序列(SEQ ID NO:165)

[0802] 质粒84:pLX-Puro(1-180)-NpuDnaE(N)-IRES-TagBFP2

[0803] 蛋白质＝Puro(1-180)-NpuDnaE(N)

[0804] 载体序列(SEQ ID NO:166)

[0805] 氨基酸序列(SEQ ID NO:167)

[0806] 质粒85:pLX-NpuDnaE(C)-Puro(^C;181-199)-IRES-mCherry

[0807] 蛋白质＝NpuDnaE(C)-Puro(^C;181-199)

[0808] 载体序列(SEQ ID NO:168)

[0809] 氨基酸序列(SEQ ID NO:169)

[0810] 质粒86:pLX-Blast(1-58)-NpuDnaE(N)-IRES-TagBFP2

[0811] 蛋白质＝Blast(1-58)-NpuDnaE(N)

[0812] 载体序列(SEQ ID NO:170)

[0813] 氨基酸序列(SEQ ID NO:171)

[0814] 质粒87:pLX-NpuDnaE(C)-Blast(59-140)-IRES-mCherry

[0815] 蛋白质＝NpuDnaE(C)-Blast(59-140)

[0816] 载体序列(SEQ ID NO:172)

[0817] 氨基酸序列(SEQ ID NO:173)

[0818] 质粒88:pLX-NpuDnaE(C)-HygroBA-SspDnaB(N-S0)-IRES-EGFP

[0819] 蛋白质＝NpuDnaE(C)-Hygro(53-200)-SspDnaB(N-S0)

[0820] 载体序列(SEQ ID NO:174)

[0821] 氨基酸序列(SEQ ID NO:175)

[0822] 质粒89:pLX-SspDnaB(C-S0)-Hygro(201-341)-IRES-mCherry

[0823] 蛋白质＝SspDnaB(C-S0)-Hygro(201-341)

[0824] 载体序列(SEQ ID NO:176)

[0825] 氨基酸序列(SEQ ID NO:177)

[0826] 质粒90:pLX-NpuDnaE(C)-Hygro(90-200)-SspDnaB(N-S0)-IRES-EGFP

[0827] 蛋白质＝NpuDnaE(C)-Hygro(90-200)-SspDnaB(N-S0)

[0828] 载体序列(SEQ ID NO:178)

[0829] 氨基酸序列(SEQ ID NO:179)

[0830] 质粒91:pLX-Hygro(1-200)-SspDnaB(N-S0)-IRES-TagBFP2

[0831] 蛋白质＝Hygro(1-200)-SspDnaB(N-S0)

[0832] 载体序列(SEQ ID NO:180)

[0833] 氨基酸序列(SEQ ID NO:181)

[0834] 质粒92:

[0835] pLX-SspDnaB(C-S0)-Hygro(201-240)-NpuDnaE(N)-IRES-EGFP

[0836] 蛋白质＝SspDnaB(C-S0)-Hygro(201-240)-NpuDnaE(N)

[0837] 载体序列(SEQ ID NO:182)

[0838] 氨基酸序列(SEQ ID NO:183)

[0839] 质粒93:

[0840] pLX-SspDnaB(C-S0)-Hygro(201-292)-NpuDnaE(N)-IRES-EGFP

[0841] 蛋白质＝SspDnaB(C-S0)-Hygro(201-292)-NpuDnaE(N)

[0842] 载体序列(SEQ ID NO:184)

[0843] 氨基酸序列(SEQ ID NO:185)

[0844] 质粒94:pLX-DEST-IRES-TagBFP2(SEQ ID NO:186)

[0845] 质粒95:pLX-DEST-IRES-EGFP(SEQ ID NO:187)

[0846] 质粒96:pLX-DEST-IRES-mCherry(SEQ ID NO:188)

[0847] 质粒97:pLX-Hygro-IRES-TagBFP2

[0848] 载体序列(SEQ ID NO:189)

[0849] 质粒98:pLX-Hygro-IRES-mCherry

[0850] 载体序列(SEQ ID NO:190)

[0851] 质粒99:pLX-Puro-IRES-TagBFP2

[0852] 载体序列(SEQ ID NO:191)

[0853] 质粒100:pLX-Puro-IRES-mCherry

[0854] 载体序列(SEQ ID NO:192)

[0855] 质粒101:pLX-Hygro-IRES-EGFP

[0856] 载体序列(SEQ ID NO:193)

[0857] 质粒102:pLX-NLS_GFP-IRES-Hygro

[0858] 载体序列(SEQ ID NO:194)

[0859] 质粒103:pLX-LifeAct_mCherry-IRES-Hygro

[0860] 载体序列(SEQ ID NO:195)

[0861] 质粒104:pLX-NLS_GFP-IRES-Hygro(1-89)-NpuDnaE(N)

[0862] 载体序列(SEQ ID NO:196)

[0863] 质粒105:pLX-LifeAct_mScarlet-IRES-NpuDnaE(C)-Hygro(90-341)

[0864] 载体序列(SEQ ID NO:197)

[0865] 质粒106:pX330-AAVS1

[0866] sgRNA间隔区序列:gACCCCACAGTGGGGCCACTA(第一g不匹配基因组)(SEQ ID NO:198)

[0867] 载体序列(SEQ ID NO:199)

[0868] 质粒107:pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-Blast-P2A-EGFP

[0869] 载体序列(SEQ ID NO:200)

[0870] 质粒108:

[0871] pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-Blast-P2A-mScarlet

[0872] 载体序列(SEQ ID NO:201)

[0873] 质粒109:

[0874] pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-Blast(1-102)_NpuDnaE(N)-P2A-EGFP

[0875] 载体序列(SEQ ID NO:202)

[0876] 质粒110:

[0877] pAAVS1-Nst-EF1aHygro2ArtTA3(-)_TetO-NpuDnaE(C)_Blast(103-140)-P2A-mScarlet

[0878] 载体序列(SEQ ID NO:203)

[0879] 质粒111:

[0880] pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-Hygro-P2A-NTR-E2A-EGFP

[0881] 载体序列(SEQ ID NO:204)

[0882] 质粒112:

[0883] pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-Hygro-P2A-NTR-E2A-mCherry

[0884] 载体序列(SEQ ID NO:205)

[0885] 质粒113:

[0886] pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-Hygro(1-89)-NpuDnaE(N)-P2A-NTR-E2A-EGFP

[0887] 载体序列(SEQ ID NO:206)

[0888] 质粒114:

[0889] pAAVS1-Nst-EF1aBlast2ArtTA3(-)_TetO-NpuDnaE(C)-Hygro(90-341)-P2A-NTR-E2A-mCherry

[0890] 载体序列(SEQ ID NO:207)

[0891] 质粒115:pLX-Hygro(1-89)_NpuDnaE(N)_LZA-IRES-TagBFP2

[0892] 蛋白质＝Hygro(1-89)-NpuDnaE(N)-LZA

[0893] 载体序列(SEQ ID NO:208)

[0894] 氨基酸序列(SEQ ID NO:209)

[0895] 质粒116:

[0896] pLX-LZB_NpuDnaGEP(C)_Hygro(90-200)_SspDnaB(N-S0)-IRES-GFP

[0897] 蛋白质＝LZB-NpuDnaGEP(C)-Hygro(90-200)-SspDnaB(N-S0)

[0898] 载体序列(SEQ ID NO:210)

[0899] 氨基酸序列(SEQ ID NO:211)

[0900] 质粒117:

[0901] pLX-SspDnaB(C-S0)_Hygro(201-240)_NpuDnaE(N)_LZA-IRES-GFP

[0902] 蛋白质＝SspDnaB(C-S0)-Hygro(201-240)-NpuDnaE(N)-LZA

[0903] 载体序列(SEQ ID NO:212)

[0904] 氨基酸序列(SEQ ID NO:213)

[0905] 质粒118:pLX-LZB_NpuDnaGEP(C)_Hygro(241-341)-IRES-mCherry

[0906] 蛋白质＝LZB-NpuDnaGEP(C)-Hygro(241-341)

[0907] 载体序列(SEQ ID NO:214)

[0908] 氨基酸序列(SEQ ID NO:215)

[0909] AC1947GB(SEQ ID NO:216)

[0910] AC1949GB(SEQ ID NO:217)

[0911] pCR8-ccdbCam(SEQ ID NO:218)

[0912] 参考文献

[0913] 1.Shearer,R.F.&Saunders,D.N.Experimental design for stable genetic manipulation in mammalian cell lines:lentivirus and alternatives.Genes to cells:devoted to molecular&cellular mechanisms 20,1-10(2015).

[0914] 2.Abuin,A.&Bradley,A.Recycling selectable markers in mouse embryonic stem cells.Molecular and cellular biology 16,1851-1856(1996).

[0915] 3.Shah,N.H.&Muir,T.W.Inteins:Nature's Gift to Protein Chemists.Chemical science 5,446-461(2014).

[0916] 4.Zettler,J.,Schütz,V.&Mootz,H.D.The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction.FEBS letters 583,909-914(2009).

[0917] 5.Iwai,H.,Züger,S.,Jin,J.&Tam,P.-H.Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme.FEBS

letters 580,1853-1858(2006).

[0918] 6.Sun,W.,Yang,J.&Liu,X.Q.Synthetic two-piece and three-piece split inteins for protein trans-splicing.The Journal of biological chemistry 279,35281-35286(2004).

[0919] 7.Cheriyan,M.,Pedamallu,C.S.,Tori,K.&Perler,F.Faster protein splicing with the Nostoc punctiforme DnaE intein using non-native extein residues.The Journal of biological chemistry 288,6202-6211(2013).

[0920] 8.Chee,J.&Chin,C.Gateway cloning technology:Advantages and drawbacks.Cloning Transgenes 4,138(2015).

[0921] 9.Bindels,D.S.等,mScarlet:a bright monomeric red fluorescent protein for cellular imaging.nature methods 14,53(2017).

[0922] 10.Stevens,A.J.等,A promiscuous split intein with expanded protein engineering applications.Proceedings of the National Academy of Sciences 114,8538-8543(2017).

[0923] 11.Ghosh,I.,Hamilton,A.D.&Regan,L.Antiparallel leucine zipper-directed protein reassembly:application to the green fluorescent protein.Journal of the American Chemical Society 122,5658-5659(2000).

[0924] 12.Wang,H.,La Russa,M.&Qi,L.S.CRISPR/Cas9 in genome editing and beyond.Annual review of biochemistry 85,227-264(2016).

[0925] 13.Peng,R.,Lin,G.&Li,J.Potential pitfalls of CRISPR/Cas9-mediated genome editing.The FEBS journal 283,1218-1231(2016).

[0926] 14.Oceguera-Yanez,F.等,Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives.Methods 101,43-55(2016).

[0927] 本文所公开的所有参考文献、专利和专利申请相对于其所引用的主题按引用并入本文中,其在某些情况下可以涵盖所述文件的整体。

[0928] 除非明确指出相反的含义,否则本文中在说明书和权利要求书中使用的不定冠词“一个(a)”和“一种(an)”应理解为表示“至少一个”。

[0929] 还应该理解的是,除非有明显的相反指示,否则在本文要求保护的包括超过一个步骤或动作的任何方法中,所述方法的步骤或动作的顺序不一定限于其中叙述所述方法步骤或动作的顺序。

[0930] 在权利要求以及以上的说明书中,所有过渡性短语,如“包含”、“包括”、“携带”、“具有”、“含有”、“涉及”、“持有”、“组成”等,应理解为开放式的,即表示包括但不限于。如美国专利局专利审查程序手册第2111.03节所述,仅过渡性短语“由……组成”和“基本上由……组成”应分别是封闭式或半封闭式的过渡性短语。

[0931] 数值前的术语“约”和“基本上”表示所述数值的±10%。

[0932] 在提供值的范围的情况中,在此具体考虑和描述了范围上端和下端之间的每个值。

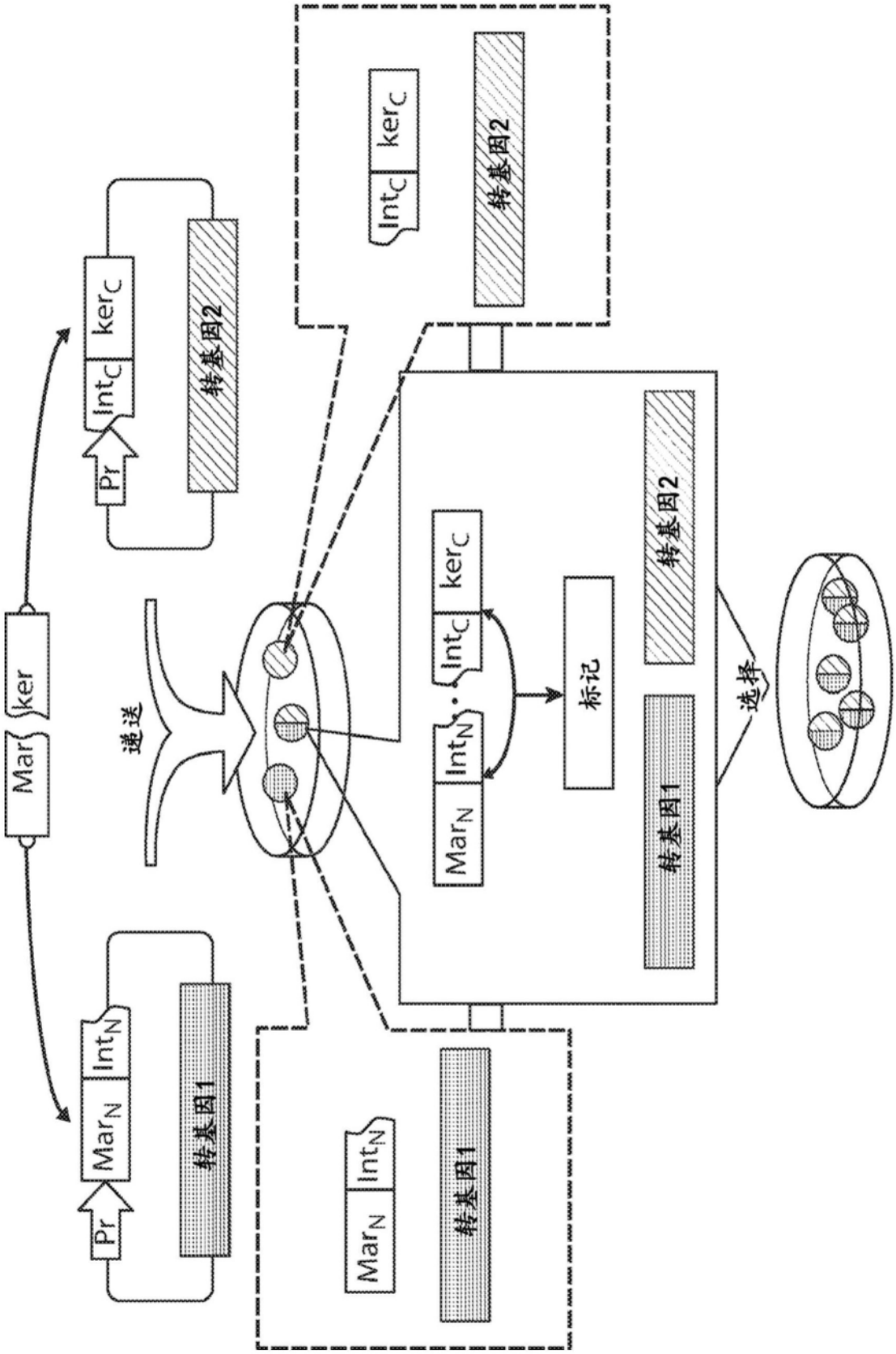

Mar
ker
Pr
Mar$_N$
Int$_N$
转基因1
Pr
Int$_C$
ker$_C$
转基因2
递送
Mar$_N$
Int$_N$
转基因1
Int$_C$
ker$_C$
转基因2
Int$_C$
ker$_C$
Int$_N$
标记
转基因2
转基因1
选择

图1A

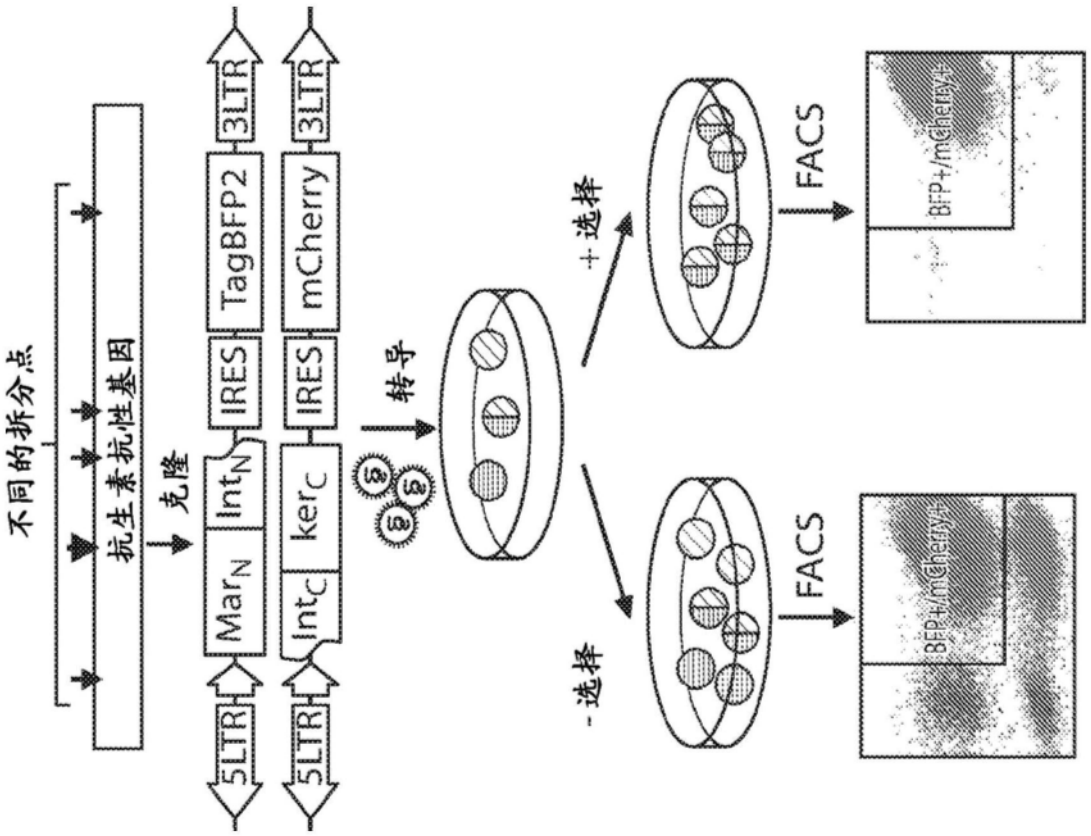

不同的拆分点
抗生素抗性基因
克隆
5LTR
Mar$_N$
Int$_N$
IRES
TagBFP2
3LTR
5LTR
Int$_C$
ker$_C$
IRES
mCherry
3LTR
转导
+选择
-选择
FACS
FACS
BFP+/mCherry+
BFP+/mCherry+

图1B

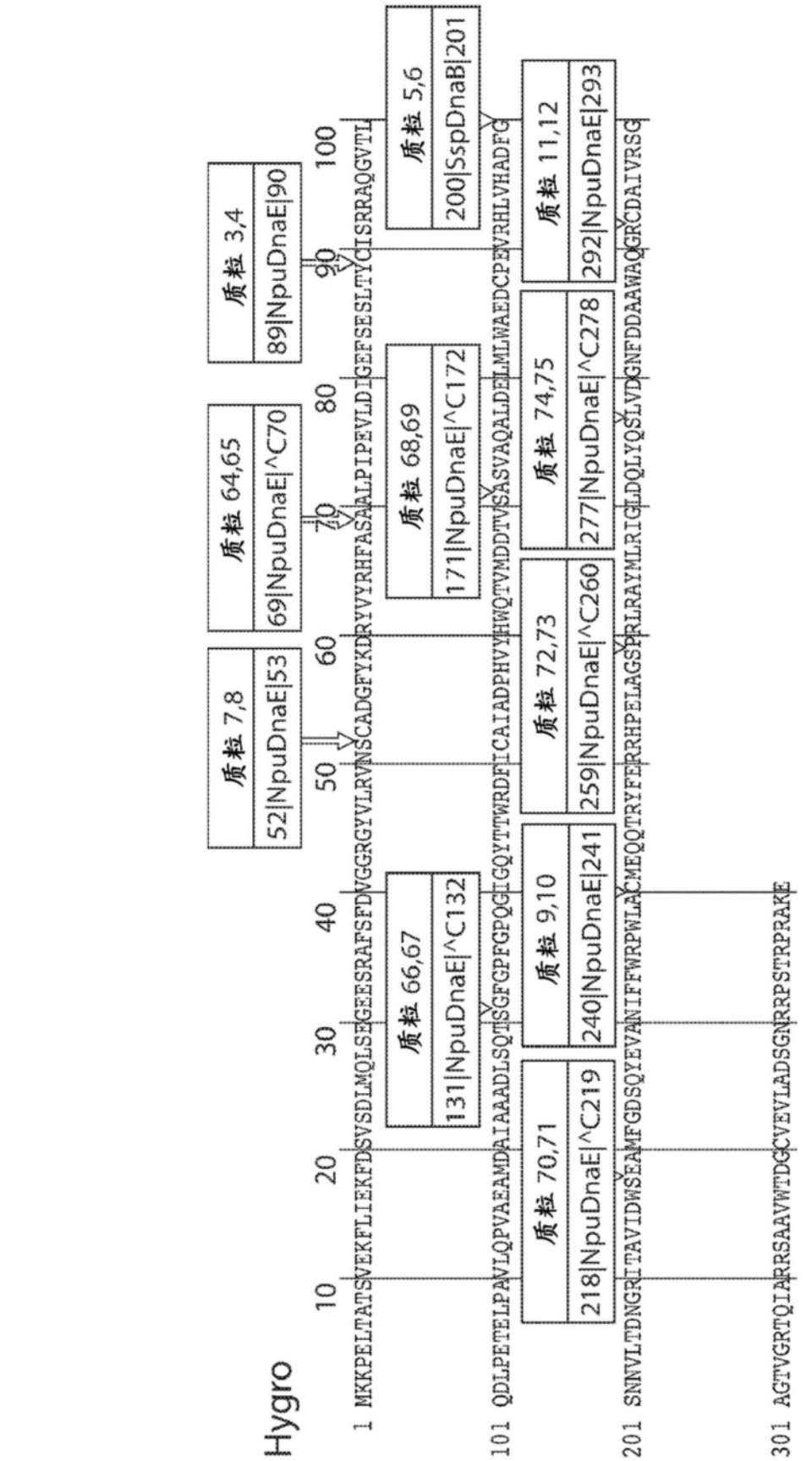

Hygro
质粒 7,8
52|NpuDnaE|53
质粒 64,65
69|NpuDnaE|^C70
质粒 3,4
89|NpuDnaE|90
质粒 66,67
131|NpuDnaE|^C132
质粒 68,69
171|NpuDnaE|^C172
质粒 5,6
200|SspDnaB|201
质粒 70,71
218|NpuDnaE|^C219
质粒 9,10
240|NpuDnaE|241
质粒 72,73
259|NpuDnaE|^C260
质粒 74,75
277|NpuDnaE|^C278
质粒 11,12
292|NpuDnaE|293

图2A

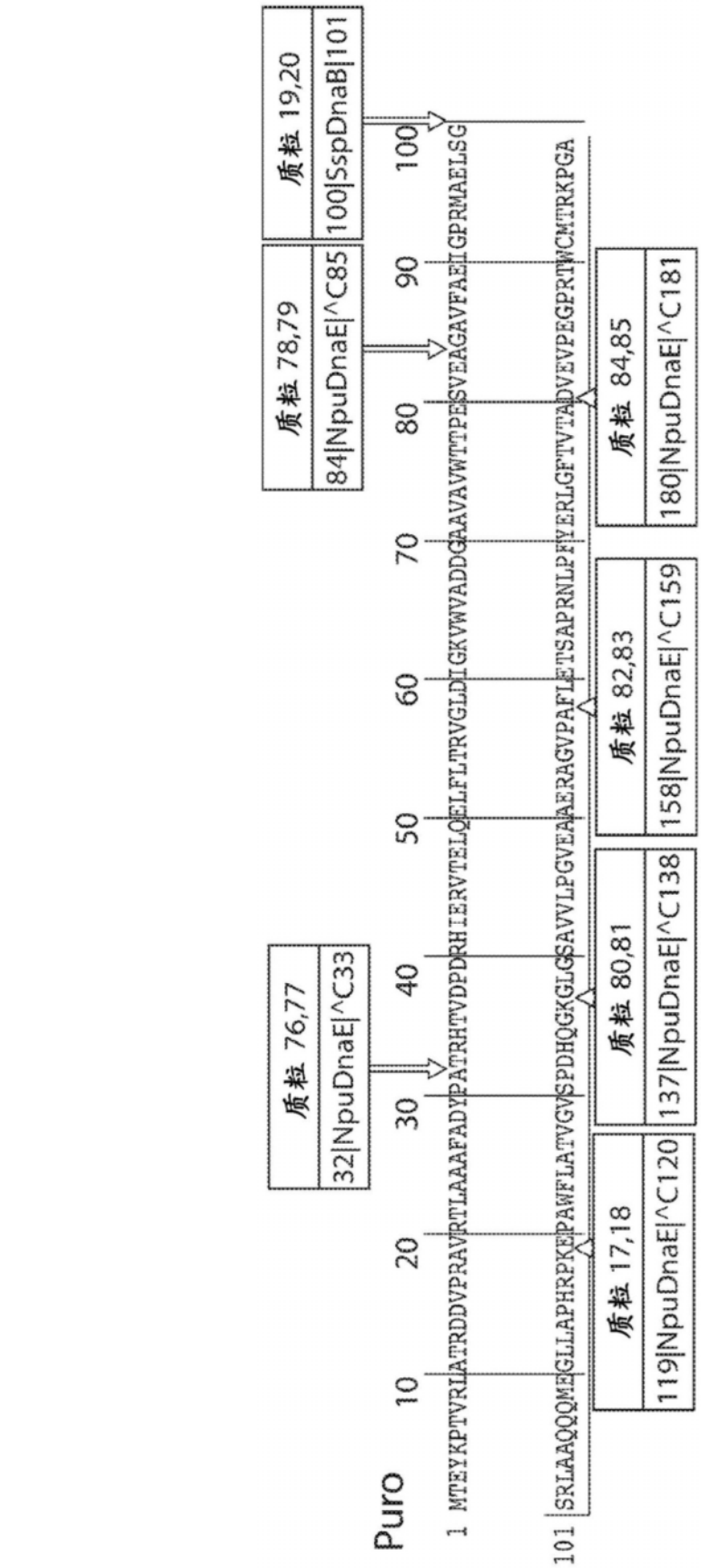

质粒 19,20
100|SspDnaB|101
质粒 78,79
84|NpuDnaE|^C85
质粒 76,77
32|NpuDnaE|^C33
质粒 84,85
180|NpuDnaE|^C181
质粒 82,83
158|NpuDnaE|^C159
质粒 80,81
137|NpuDnaE|^C138
质粒 17,18
119|NpuDnaE|^C120
Puro
1 MTEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHTVDPDRHIERVTELQELFLTRVGLDIGKVWVADDGAAVAVWTTPESVEAGAVFAEIGPRMAELSG
101 SRLAAQQQMEGLLAPHRPKEPAWFLATVGVSPDHQGKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLPFYERLGFTVTADVEVPEGPRTWCMTRKPGA
10
20
30
40
50
60
70
80
90
100

图2B

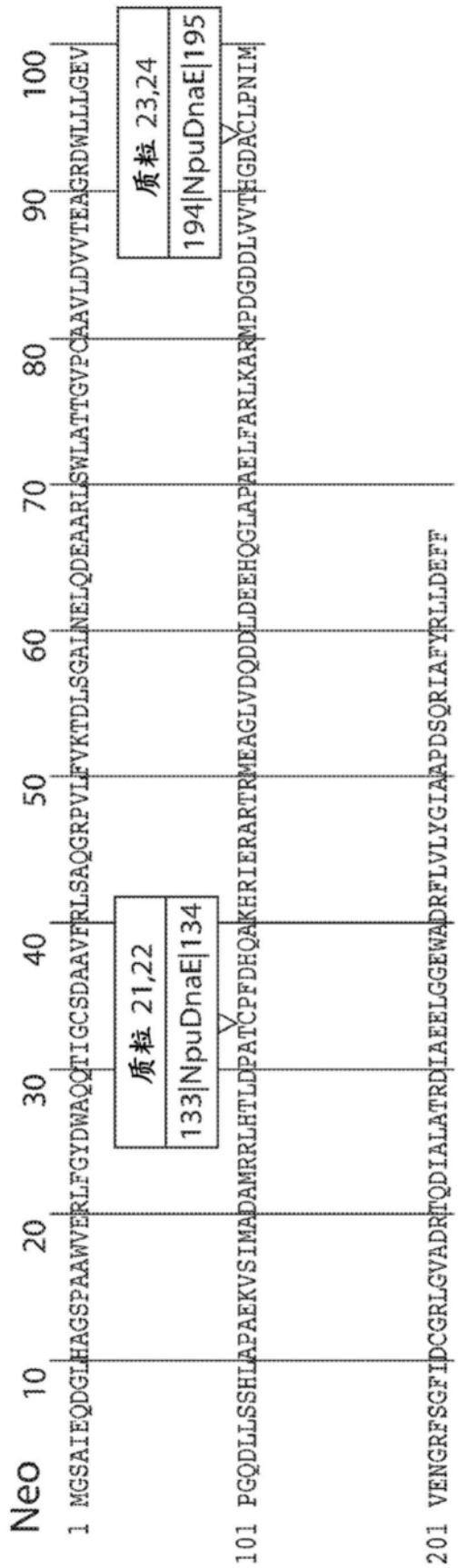


图2C

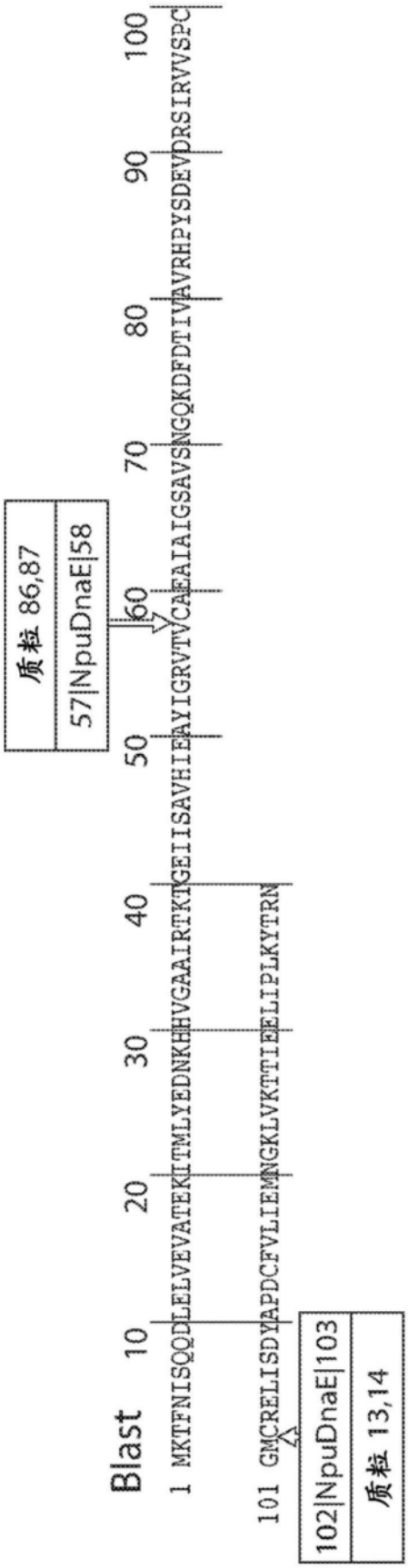

质粒 86,87
57|NpuDnaE|58
Blast
1 MKTFNISQQDLELVEVATEKITMLYEDNKHHVGAAIRTKTGEIISAVHIEAYIGRVTVCAEAIAIGSAVSNGQKDFDTIVAVRHPYSDEVDRSIRVVSPC
101 GMCRELISDYAPDCFVLIEMNGKLVKTTIEELIPLKYTRN
102|NpuDnaE|103
质粒 13,14

图2D

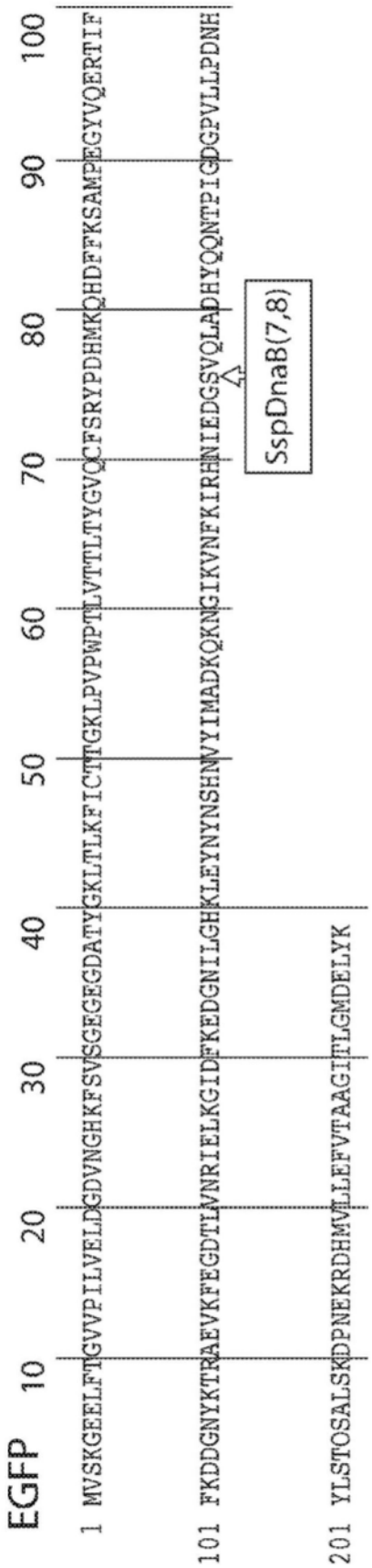

EGFP
10
20
30
40
50
60
70
80
90
100
1 MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIF
101 FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNH
201 YLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
SspDnaB(7,8)

图2E

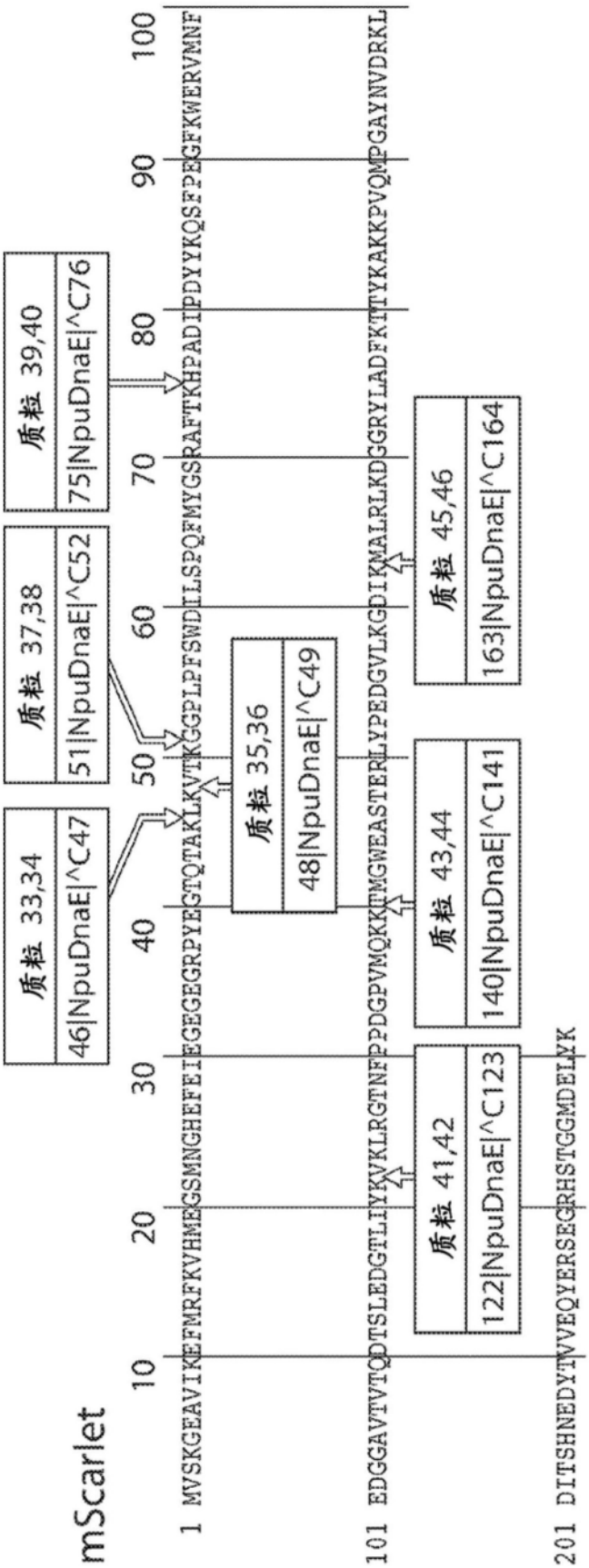

mScarlet
质粒 33,34
46|NpuDnaE|^C47
质粒 37,38
51|NpuDnaE|^C52
质粒 39,40
75|NpuDnaE|^C76
质粒 35,36
48|NpuDnaE|^C49
质粒 41,42
122|NpuDnaE|^C123
质粒 43,44
140|NpuDnaE|^C141
质粒 45,46
163|NpuDnaE|^C164
1 MVSKGEAVIKEFMRFKVHMEGSMNGHEFEIEGEGEGRPYE
101 EDGGAVTVTQDTSLEDGTLIYKVKLRGTNFPPDGPVMQKKTMGWEASTERLYPEDGVLKGDIKMALRLKDGGRYLADFKTTYKAKKPVQMPGAYNVDRKL
201 DITSHNEDYTVVEQYERSEGRHSTGGMDELYK
10
20
30
40
50
60
70
80
90
100

图2F

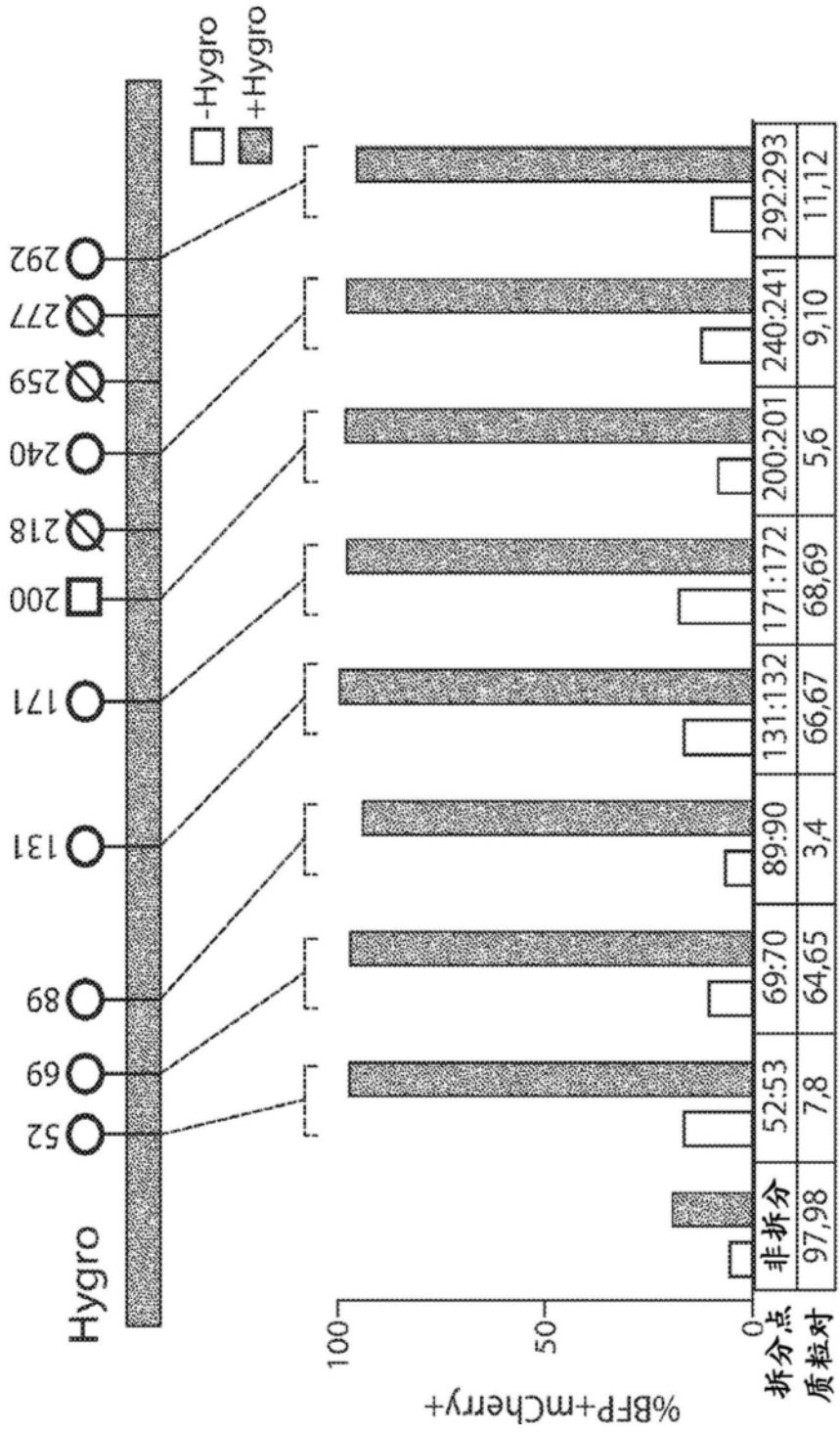

-Hygro
+Hygro
Hygro
292
277
259
240
218
200
171
131
89
69
52
%BFP+mCherry+
100
50
0
拆分点
质粒对
非拆分
97,98
52:53
7,8
69:70
64,65
89:90
3,4
131:132
66,67
171:172
68,69
200:201
5,6
240:241
9,10
292:293
11,12

图3

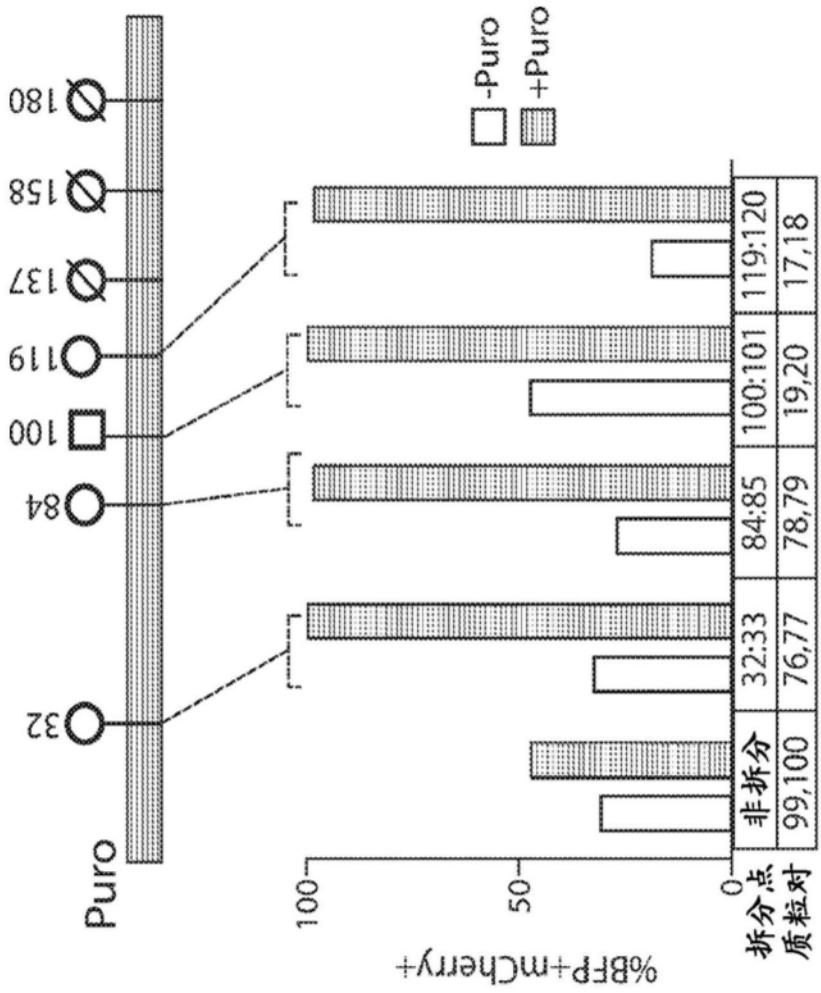

-Puro
+Puro
180
158
137
119
100
84
32
Puro
100
50
0
%BFP+mCherry+
拆分点
质粒对
非拆分
99,100
32:33
76,77
84:85
78,79
100:101
19,20
119:120
17,18

图4

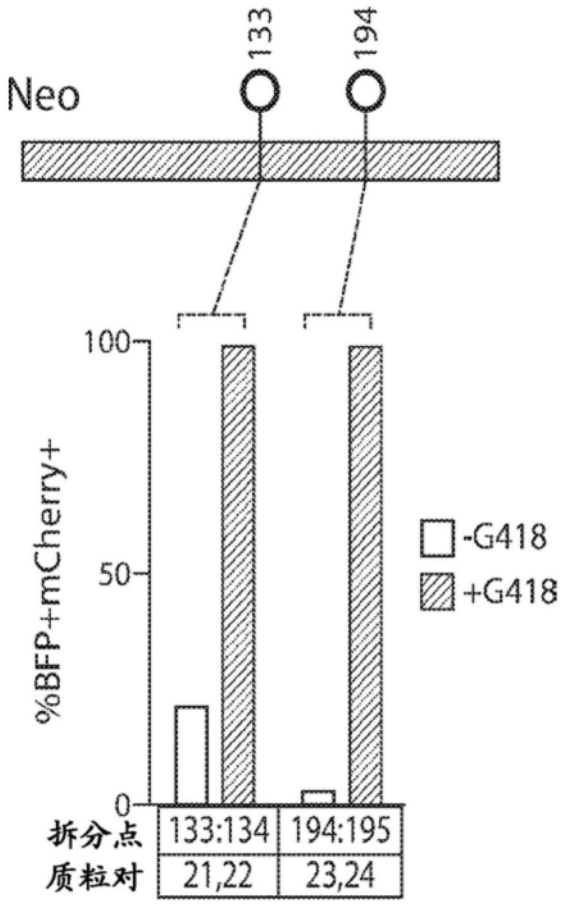

Neo
133
194
100
50
0
%BFP+mCherry+
-G418
+G418
拆分点
质粒对
133:134
194:195
21,22
23,24

图5

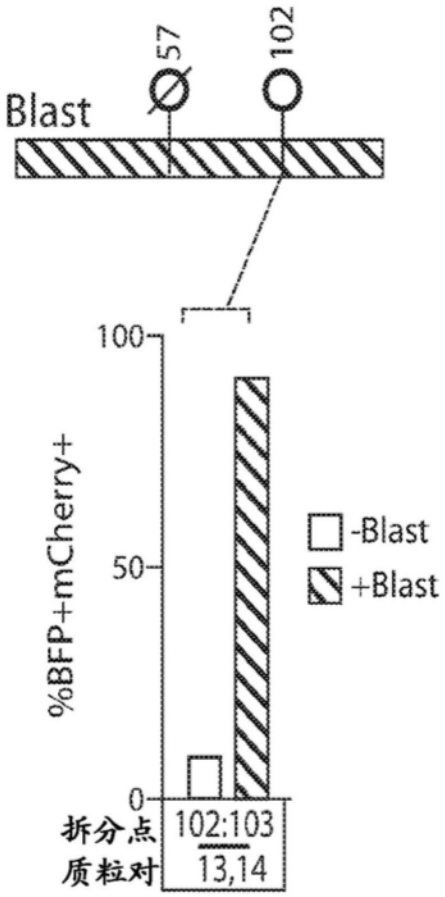

Blast
57
102
100
50
0
%BFP+mCherry+
-Blast
+Blast
拆分点
质粒对
102:103
13,14

图6

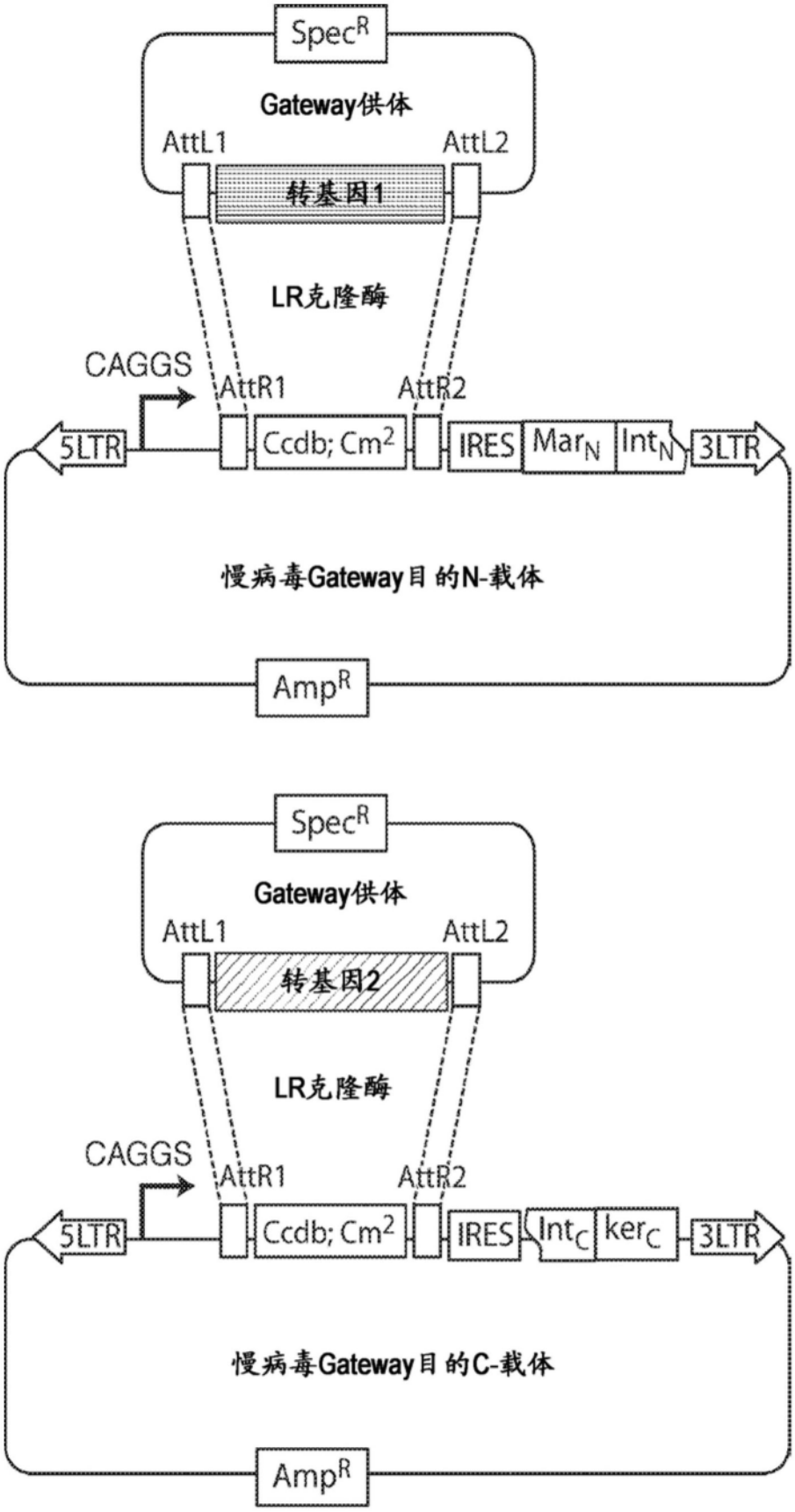
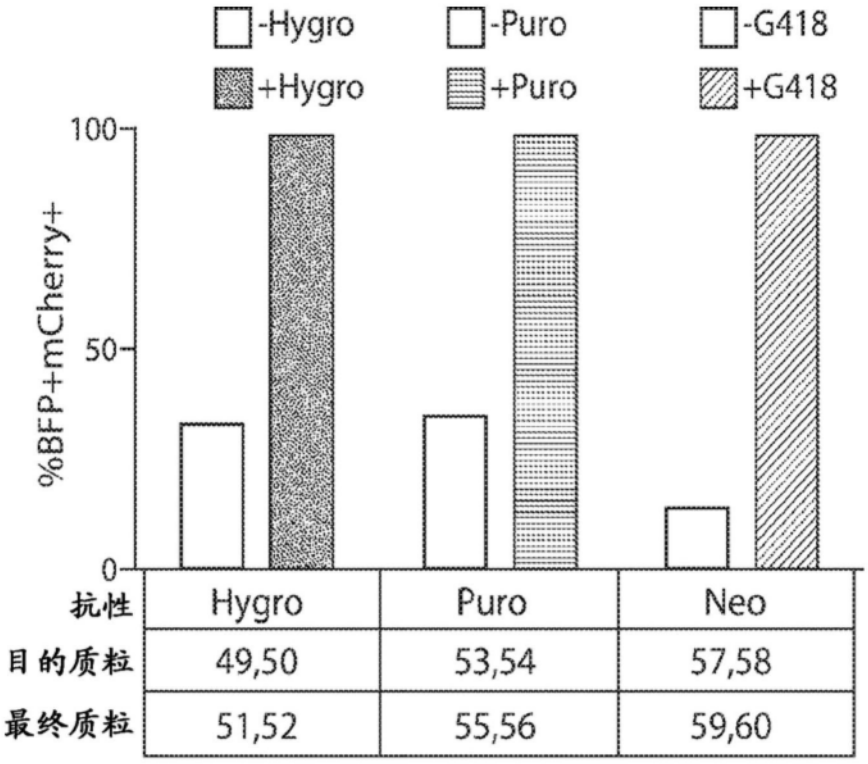

SpecR
Gateway供体
AttL1
AttL2
转基因1
LR克隆酶
CAGGS
AttR1
AttR2
5LTR
Ccdb; Cm2
IRES
MarN
IntN
3LTR
慢病毒Gateway目的N-载体
AmpR
SpecR
Gateway供体
AttL1
AttL2
转基因2
LR克隆酶
CAGGS
AttR1
AttR2
5LTR
Ccdb; Cm2
IRES
IntC
kerC
3LTR
慢病毒Gateway目的C-载体
AmpR

图7A



抗性	Hygro	Puro	Neo
目的质粒	49,50	53,54	57,58
最终质粒	51,52	55,56	59,60

图7B

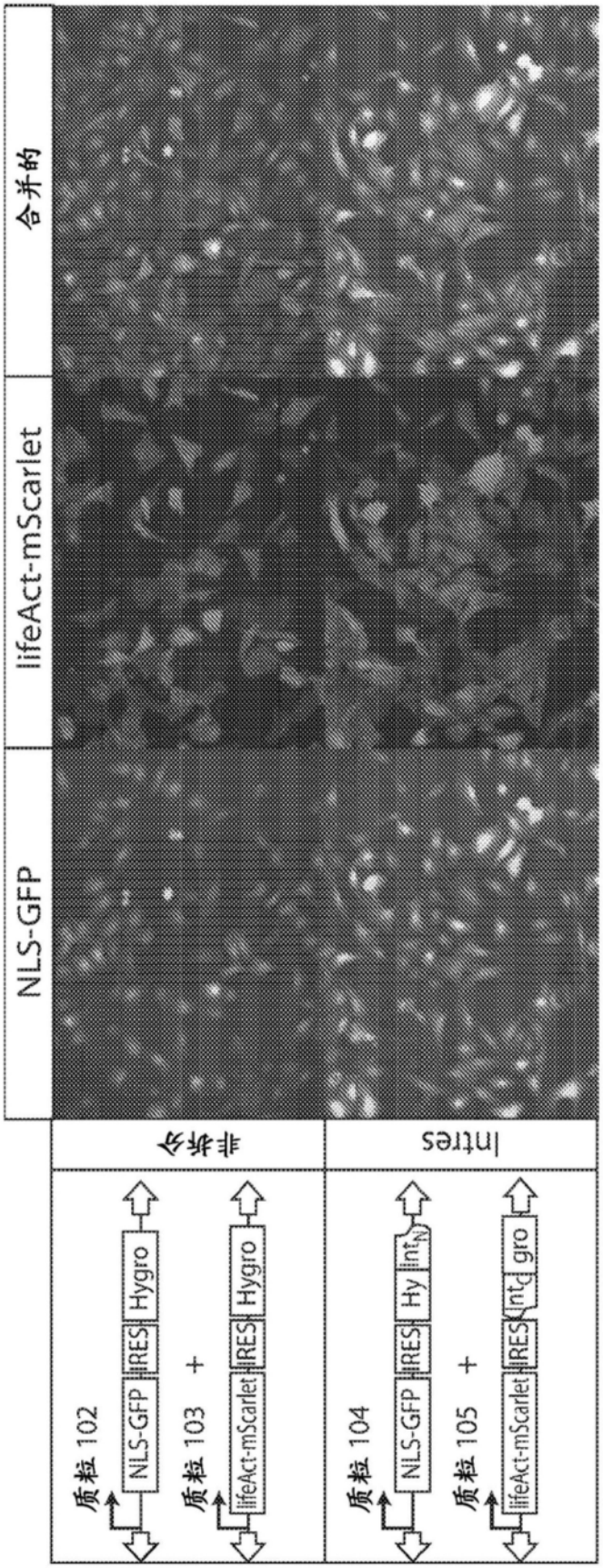

NLS-GFP
lifeAct-mScarlet
合并的
非共焦
Intres
质粒 102
NLS-GFP
IRES
Hygro
质粒 103 +
lifeAct-mScarlet
IRES
Hygro
质粒 104
NLS-GFP
IRES
Hy
int_N
质粒 105 +
lifeAct-mScarlet
IRES
int_C
gro

图7C

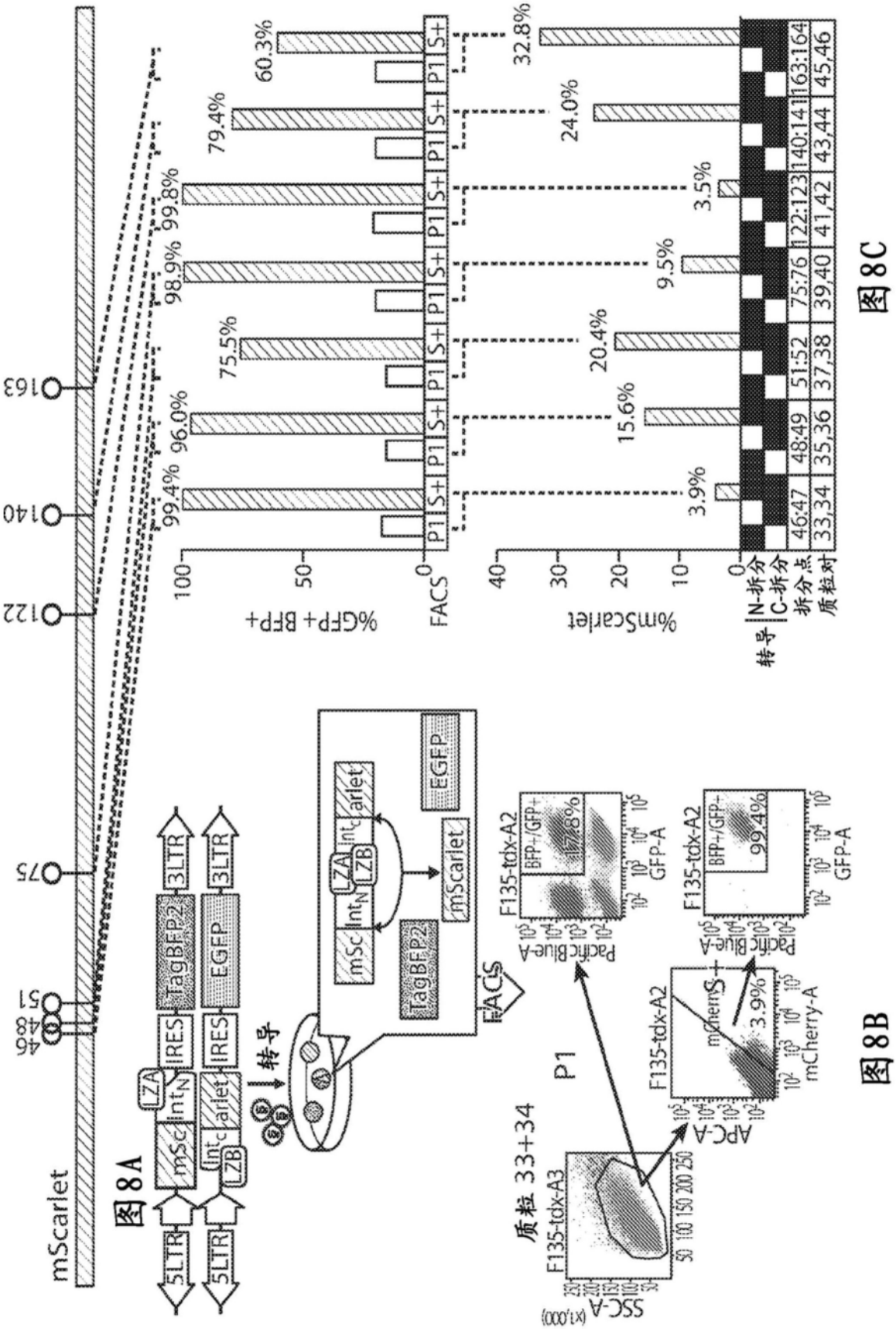

mScarlet
5LTR
LZA
IntN
mSc
IntC
arlet
LZB
IRES
TagBFP2
EGEP
3LTR
转导
EGFP
mScarlet
FACS
质粒 33+34
P1
F135-tdx-A3
SSC-A
F135-tdx-A2
APC-A
mCherry-A
Pacific Blue-A
GFP-A
BFP+/GFP+
67.8%
BFP+/GFP-
99.4%
3.9%
%GFP+ BFP+
%mScarlet
FACS
转导
N-拆分
C-拆分
拆分点
质粒对
46:47
48:49
51:52
75:76
122:123
140:141
163:164
33,34
35,36
37,38
39,40
41,42
43,44
45,46
99.4%
96.0%
75.5%
98.9%
99.8%
79.4%
60.3%
3.9%
15.6%
20.4%
9.5%
3.5%
24.0%
32.8%

图8A 图8B 图8C

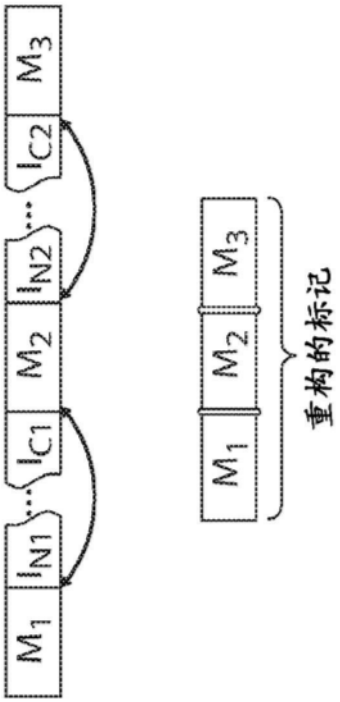


图9A

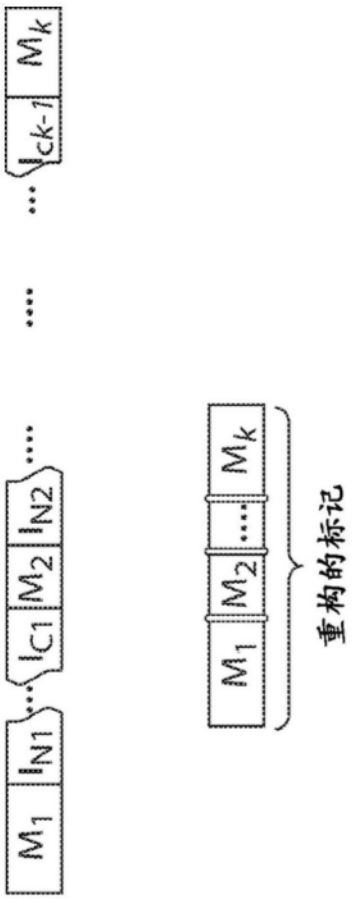


图9B

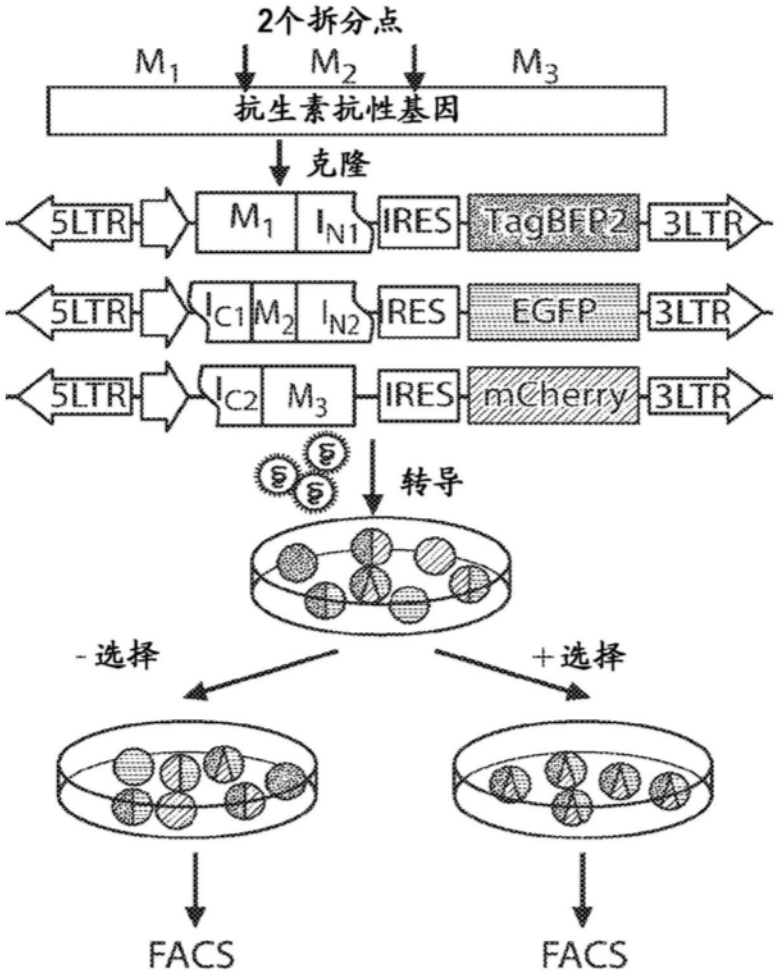

2个拆分点
M1
M2
M3
抗生素抗性基因
克隆
5LTR
M1
IN1
IRES
TagBFP2
3LTR
5LTR
IC1
M2
IN2
RES
EGFP
3LTR
5LTR
IC2
M3
IRES
mCherry
3LTR
转导
-选择
+选择
FACS
FACS

图9C

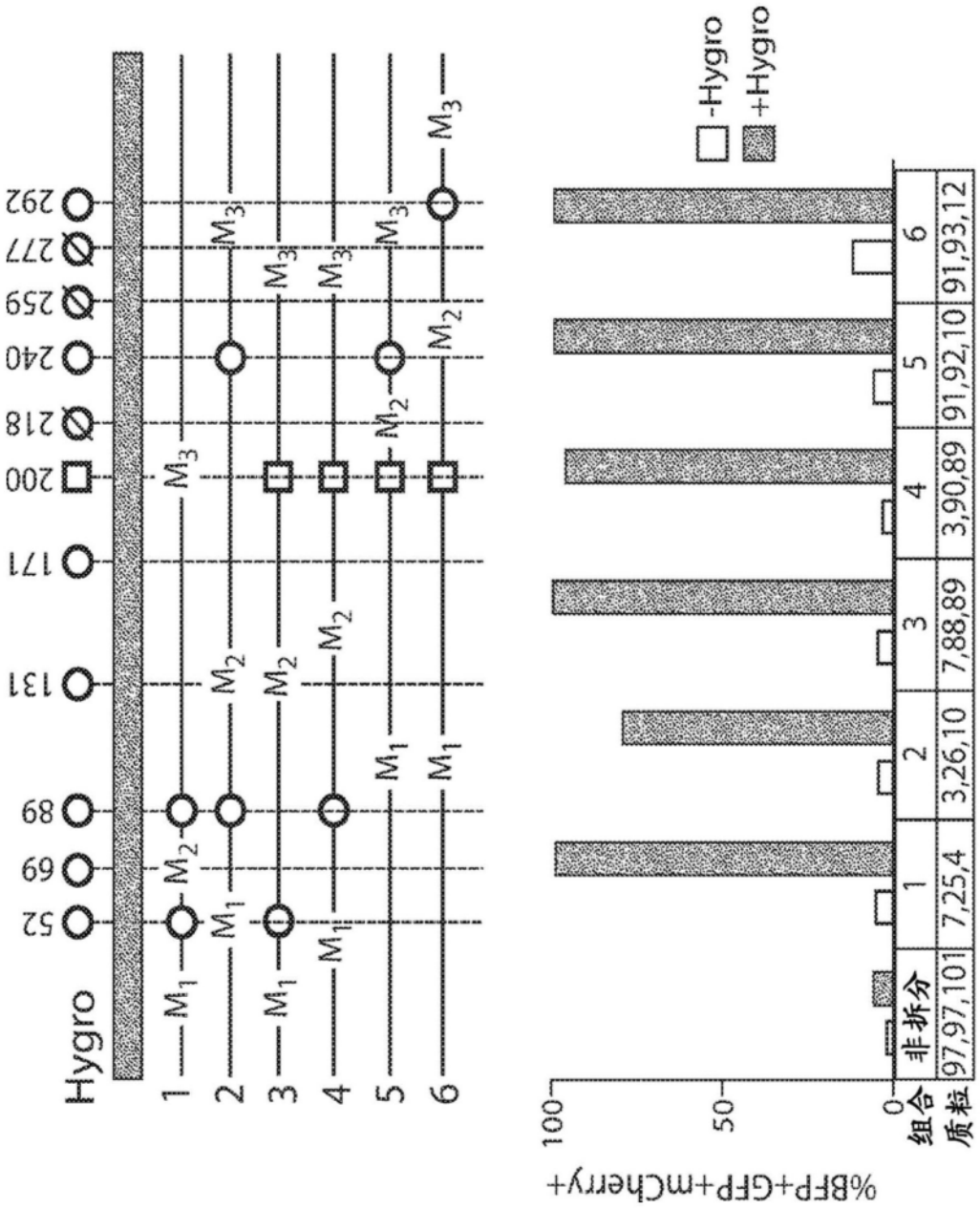

Hygro
52
69
89
131
171
200
218
240
259
277
292
1
2
3
4
5
6
M1
M2
M3
-Hygro
+Hygro
%BFP+GFP+mCherry+
100
50
0
组合
质粒
非拆分
97,97,101
7,25,4
3,26,10
7,88,89
3,90,89
91,92,10
91,93,12

图9D

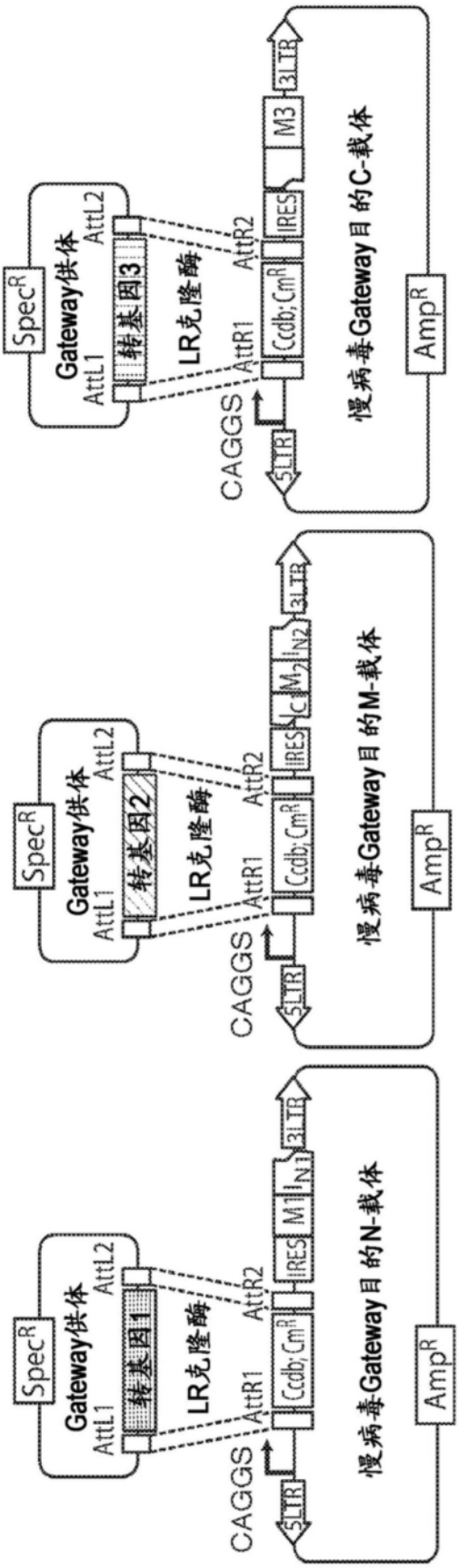

Spec^R
Gateway供体
AttL1
AttL2
转基因1
LR克隆酶
AttR1
AttR2
Ccdb;Cm^R
IRES
M1
IN1
3LTR
CAGGS
5LTR
慢病毒-Gateway目的N-载体
Amp^R
转基因2
IC
M2
IN2
慢病毒-Gateway目的M-载体
转基因3
M3
慢病毒-Gateway目的C-载体

图10A

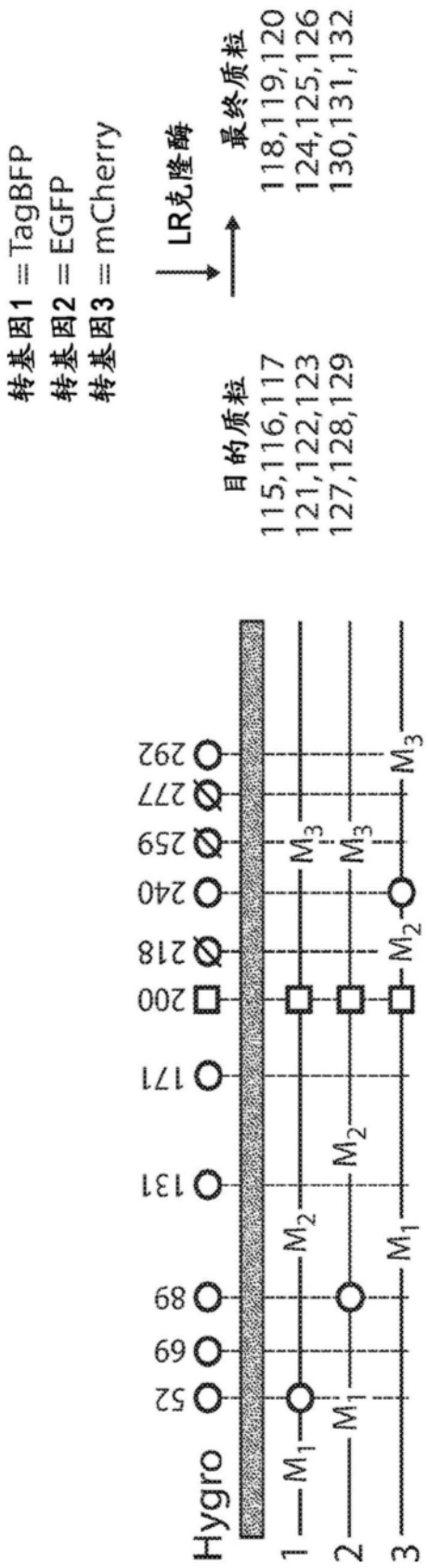

转基因1 = TagBFP
转基因2 = EGFP
转基因3 = mCherry
LR克隆酶
目的质粒
115,116,117
121,122,123
127,128,129
最终质粒
118,119,120
124,125,126
130,131,132
Hygro
52
69
89
131
171
200
218
240
259
277
292
1
2
3
M1
M2
M3

图10B

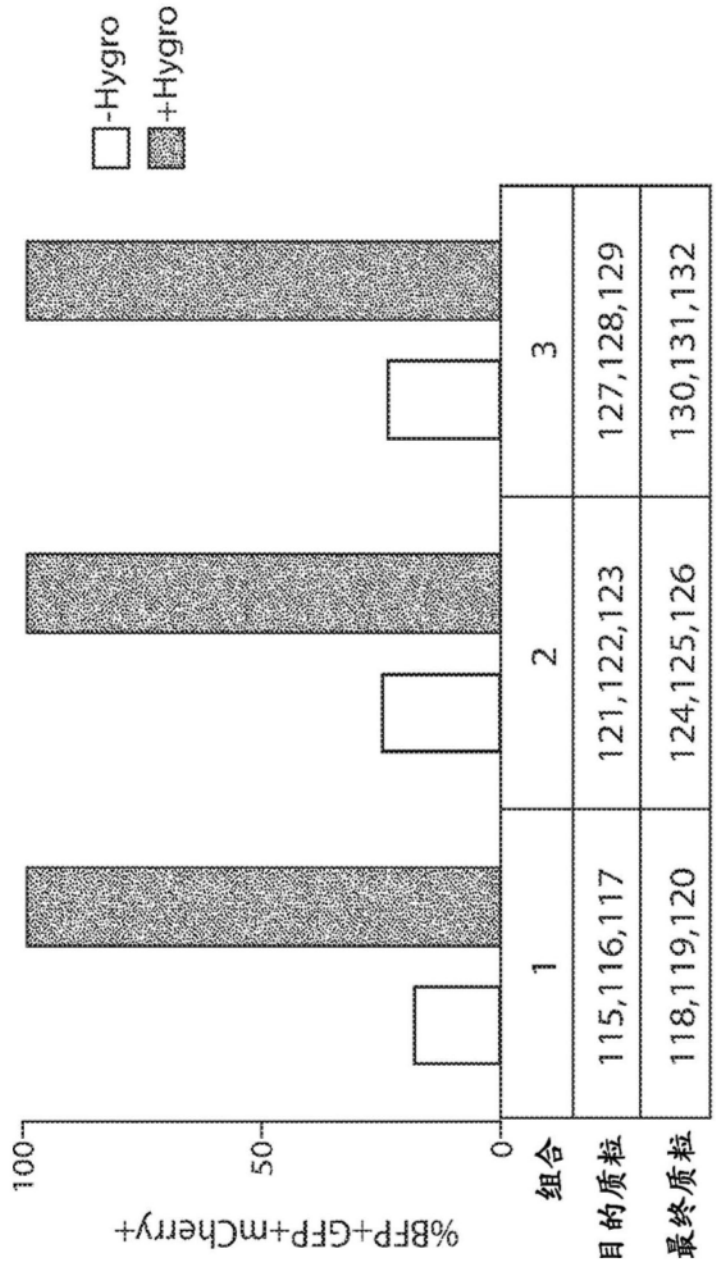

-Hygro
+Hygro
%BFP+GFP+mCherry+
100
50
0
组合
目的质粒
最终质粒
1
2
3
115,116,117
121,122,123
127,128,129
118,119,120
124,125,126
130,131,132

图10C

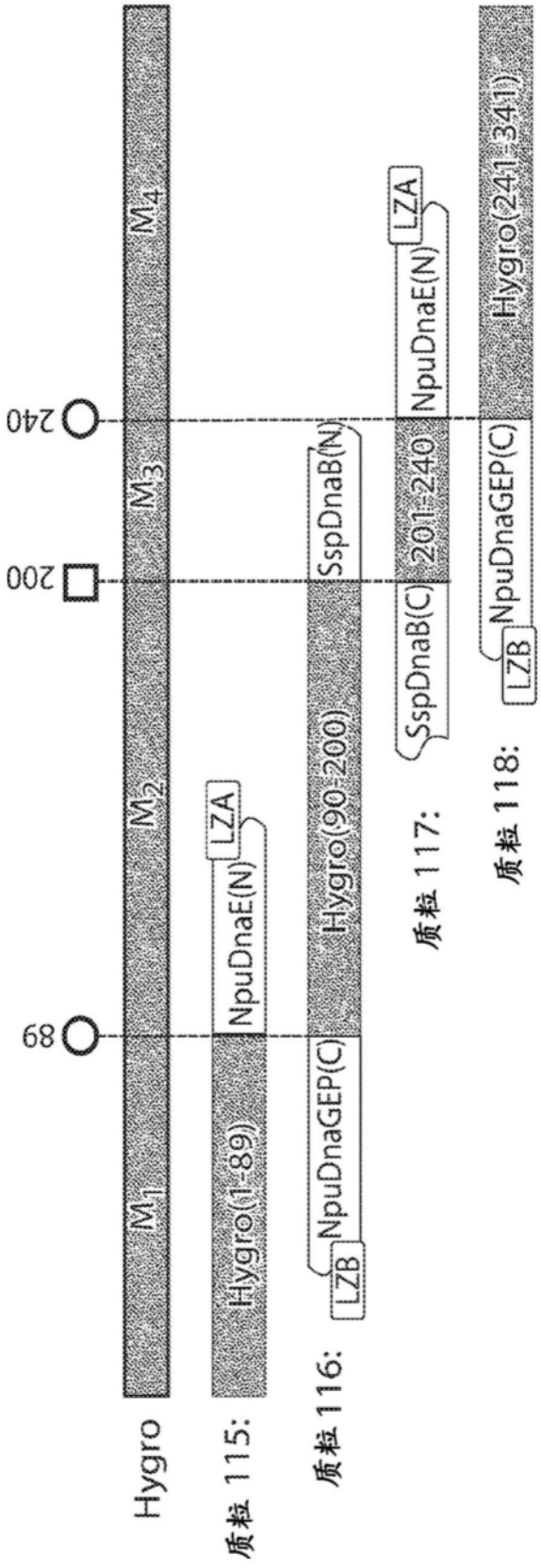

Hygro
质粒 115:
质粒 116:
质粒 117:
质粒 118:
M1
M2
M3
M4
Hygro(1-89)
NpuDnaE(N)
LZA
NpuDnaGEP(C)
LZB
Hygro(90-200)
SspDnaB(N)
SspDnaB(C)
201-240
NpuDnaE(N)
LZA
NpuDnaGEP(C)
LZB
Hygro(241-341)
89
200
240

图11

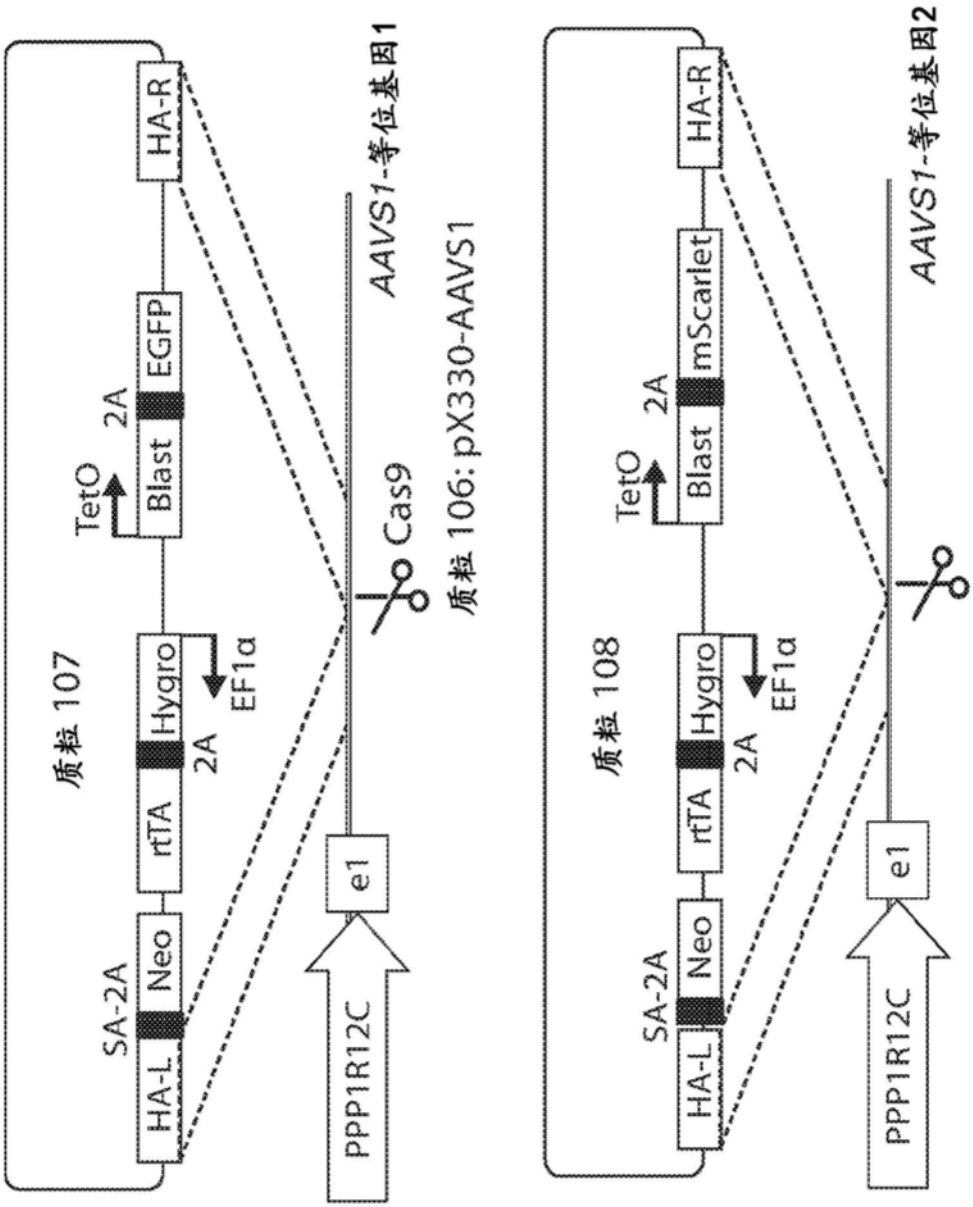

质粒 107
HA-L
SA-2A
Neo
rtTA
2A
Hygro
EF1α
TetO
2A
Blast
EGFP
HA-R
PPP1R12C
e1
Cas9
AAVS1-等位基因1
质粒 106: pX330-AAVS1
质粒 108
HA-L
SA-2A
Neo
rtTA
2A
Hygro
EF1α
TetO
2A
Blast
mScarlet
HA-R
PPP1R12C
e1
AAVS1-等位基因2

图12A

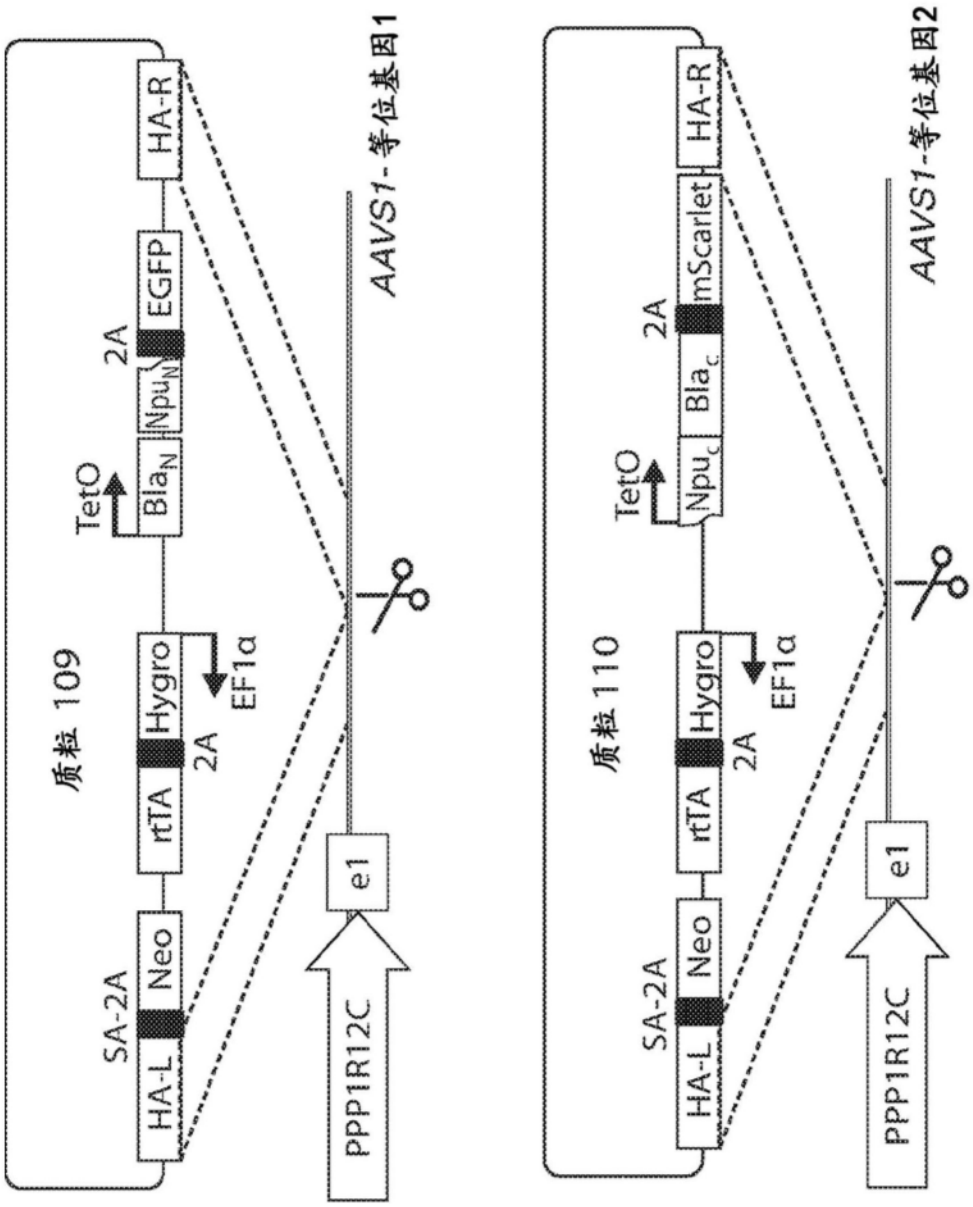

质粒 109
HA-L
SA-2A
Neo
rtTA
2A
Hygro
EF1α
e1
PPP1R12C
AAVS1-等位基因1
HA-R
EGFP
2A
Npu$_N$
Bla$_N$
TetO
质粒 110
HA-L
SA-2A
Neo
rtTA
2A
Hygro
EF1α
e1
PPP1R12C
AAVS1-等位基因2
HA-R
mScarlet
2A
Bla$_C$
Npu$_C$
TetO

图12B

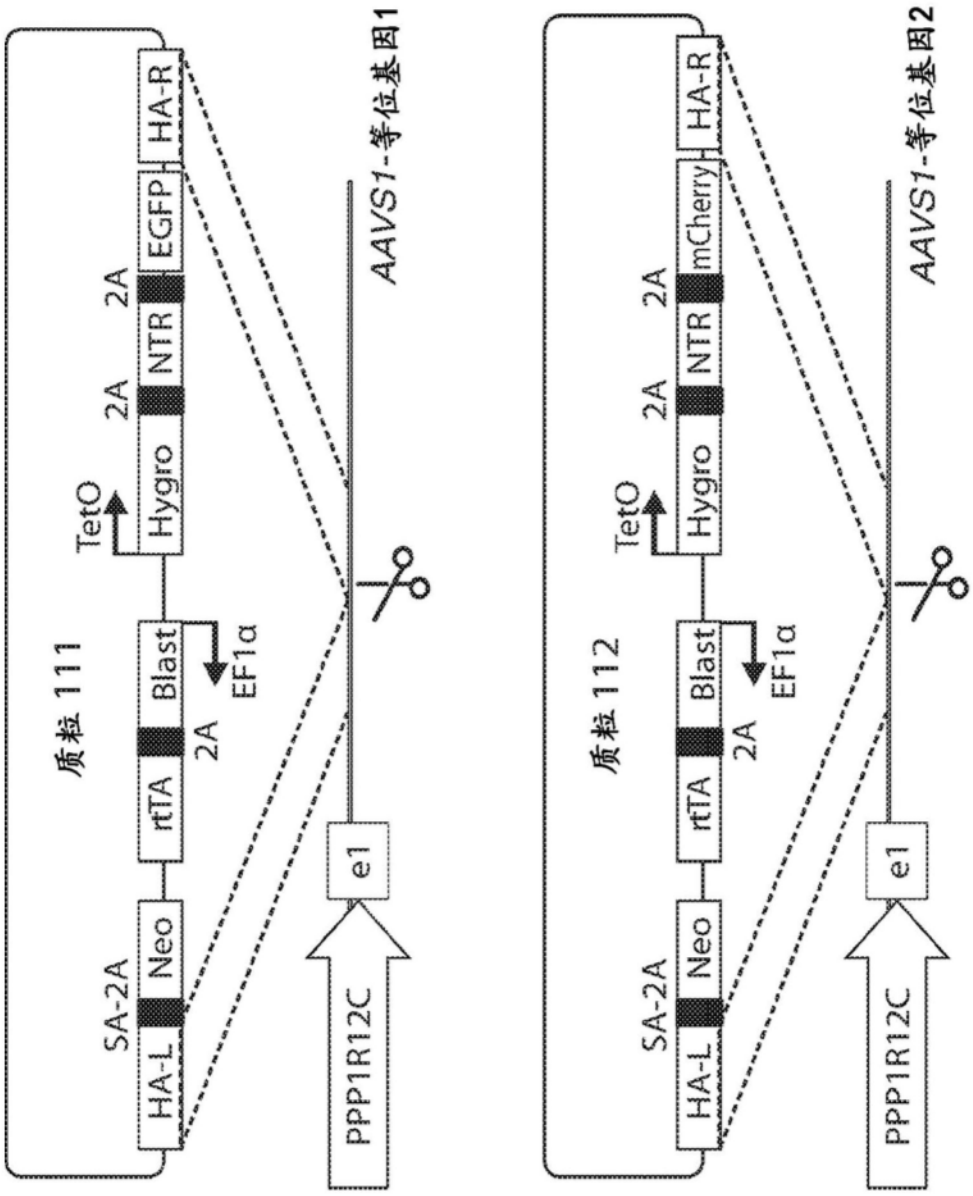

质粒 111
HA-L
SA-2A
Neo
rtTA
2A
Blast
EF1α
TetO
Hygro
2A
NTR
2A
EGFP
HA-R
e1
PPP1R12C
AAVS1-等位基因1
质粒 112
HA-L
SA-2A
Neo
rtTA
2A
Blast
EF1α
TetO
Hygro
2A
NTR
2A
mCherry
HA-R
e1
PPP1R12C
AAVS1-等位基因2

图12C

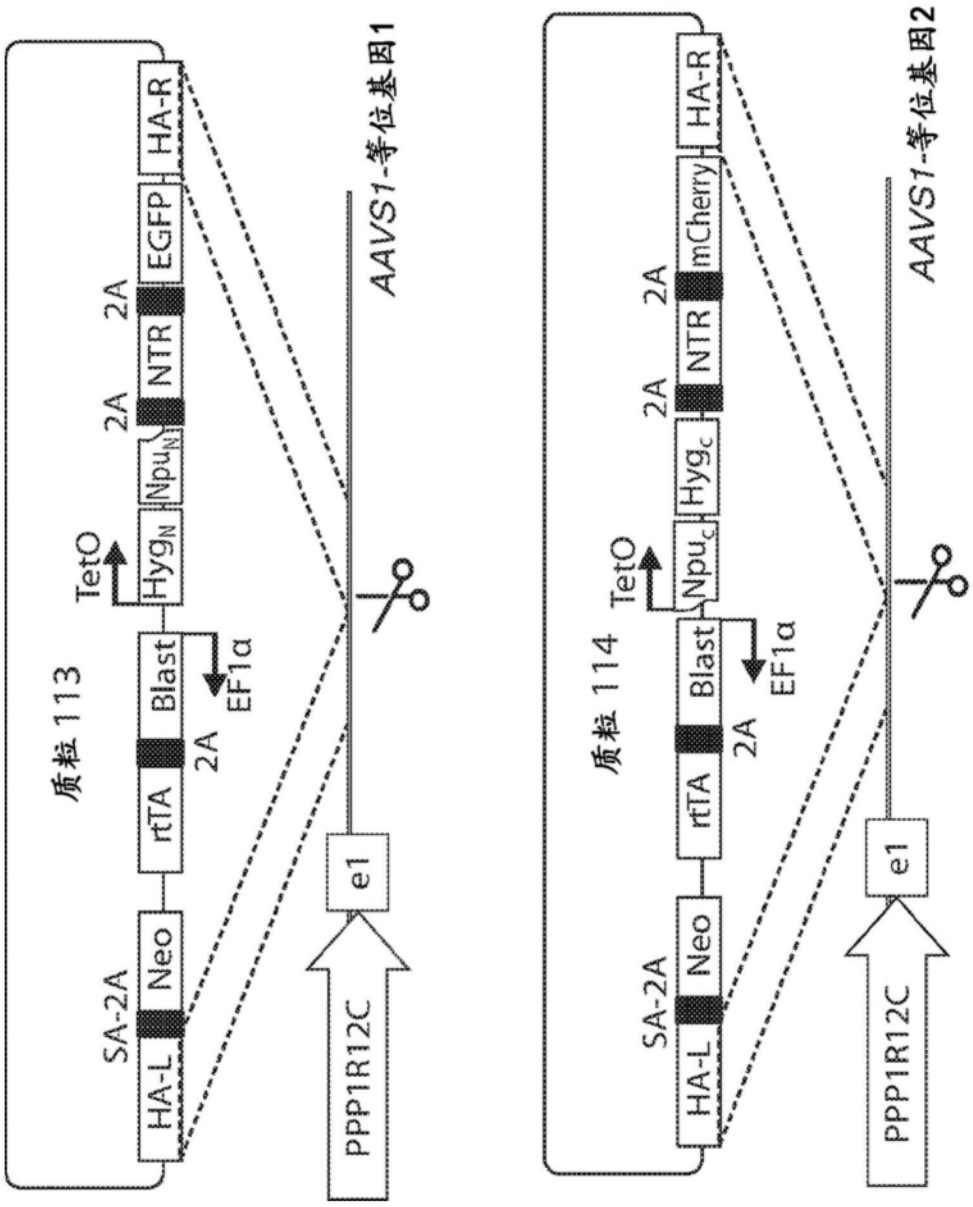

质粒 113
SA-2A
HA-L
Neo
2A
rtTA
Blast
2A
EF1α
TetO
Hyg$_N$
Npu$_N$
2A
NTR
2A
EGFP
HA-R
e1
PPP1R12C
AAVS1-等位基因1
质粒 114
SA-2A
HA-L
Neo
2A
rtTA
Blast
2A
EF1α
TetO
Npu$_C$
Hyg$_C$
2A
NTR
2A
mCherry
HA-R
e1
PPP1R12C
AAVS1-等位基因2

图12D

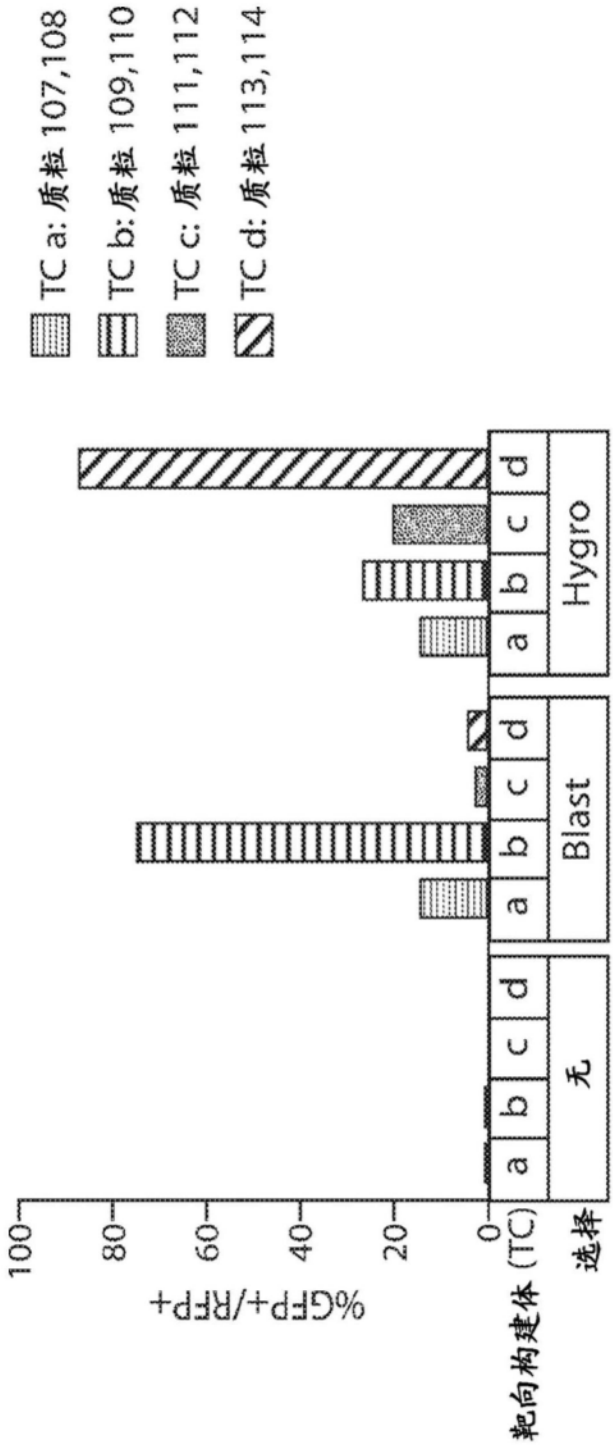

TC a: 质粒 107,108
TC b: 质粒 109,110
TC c: 质粒 111,112
TC d: 质粒 113,114
100
80
60
40
20
0
%GFP+/RFP+
a
b
c
d
无
a
b
c
d
Blast
a
b
c
d
Hygro
靶向构建体 (TC)
选择

图12E