



(12) 发明专利申请

(10) 申请公布号 CN 112074604 A

(43) 申请公布日 2020.12.11

(21) 申请号 201980030296.2

(22) 申请日 2019.05.03

(30) 优先权数据

62/667,194 2018.05.04 US

(85) PCT国际申请进入国家阶段日

2020.11.04

(86) PCT国际申请的申请数据

PCT/US2019/030607 2019.05.03

(87) PCT国际申请的公布数据

W02019/213527 EN 2019.11.07

(71) 申请人 西格马-奥尔德里奇有限责任公司

地址 美国密苏里州

(72) 发明人 J·马斯卡伦哈斯

T·博尔格舒尔特 K·凯泽

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

代理人 翟建伟 黄希贵

(51) Int.Cl.

*C12N 15/63* (2006.01)

*C12N 15/90* (2006.01)

*C12P 21/00* (2006.01)

权利要求书2页 说明书22页

序列表13页 附图7页

(54) 发明名称

具有修饰的宿主细胞蛋白概况的工程改造的细胞

(57) 摘要

经基因工程改造以减少或消除特定宿主细胞蛋白的表达的哺乳动物细胞系,以及使用该工程改造的哺乳动物细胞系生产具有低残余宿主细胞蛋白污染水平的重组蛋白的方法。

1. 用于生产具有降低的宿主细胞蛋白污染水平的重组蛋白产物的方法,所述方法包括  
(a) 在经工程改造以减少或消除至少一种宿主细胞蛋白的表达的哺乳动物细胞系中表达重组蛋白;和

(b) 纯化所述重组蛋白以形成所述重组蛋白产物,其中所述重组蛋白产物的残余宿主细胞蛋白水平低于由非工程改造的亲本哺乳动物细胞系产生的蛋白产物中的残余宿主细胞蛋白水平。

2. 权利要求1的方法,其中所述至少一种宿主细胞蛋白选自羧肽酶B1、羧肽酶D、羧肽酶E、羧肽酶M、组织蛋白酶B、组织蛋白酶D、组织蛋白酶L1、组织蛋白酶Z、硫酸软骨素蛋白聚糖4、簇蛋白、二肽基肽酶3、豆荚蛋白酶 (legumain)、亮氨酸氨基肽酶3、脂蛋白脂肪酶、赖氨酰氧化酶、金属蛋白酶抑制剂1、中性 $\alpha$ -葡萄糖苷酶、巢蛋白1、过氧蛋白酶 (peroxidase)、磷脂酶B-样2、脯氨酰内肽酶、蛋白精氨酸N-甲基转移酶5、蛋白磷酸酶1G、丝氨酸蛋白酶、唾液酸酶1、硫氧还蛋白或硫氧还蛋白还原酶。

3. 权利要求1的方法,其中所述哺乳动物细胞系具有减少或消除的羧肽酶D、组织蛋白酶B、组织蛋白酶D、簇蛋白、脂蛋白脂肪酶、金属蛋白酶抑制剂1、巢蛋白1、过氧蛋白酶 (peroxidase)、丝氨酸蛋白酶、硫氧还蛋白、硫氧还蛋白还原酶或其组合的表达。

4. 权利要求1至3中任一项的方法,其中所述至少一种宿主细胞蛋白的表达经由编码所述至少一种宿主细胞蛋白的染色体序列的至少一个等位基因的失活而减少或消除。

5. 权利要求4的方法,其中编码所述至少一种宿主细胞蛋白的染色体序列的两个等位基因均被失活。

6. 权利要求4或5的方法,其中使用靶向核酸内切酶介导的基因组修饰技术来使所述染色体序列失活。

7. 权利要求6的方法,其中所述靶向核酸内切酶是CRISPR核糖核蛋白复合物或一对锌指核酸酶。

8. 权利要求1至7中任一项的方法,其中所述细胞系是人细胞系。

9. 权利要求8的方法,其中所述人细胞系是人胚胎肾细胞293 (HEK293) 细胞系、HT-1080人结缔组织系或PER.C6人胚胎视网膜细胞系。

10. 权利要求1至7中任一项的方法,其中所述细胞系是非人细胞系。

11. 权利要求10的方法,其中所述非人细胞系是中国仓鼠卵巢 (CHO) 细胞系、幼仓鼠肾 (BHK) 细胞系、NS0小鼠骨髓瘤细胞系、Sp2/0小鼠骨髓瘤细胞系、C127小鼠乳腺细胞系或Vero非洲绿猴肾细胞系。

12. 权利要求1至7中任一项的方法,其中所述细胞系是CHO细胞系。

13. 权利要求1至12中任一项的方法,其中在步骤 (b) 的纯化包括澄清步骤和一个或多个色谱步骤。

14. 权利要求1至13中任一项的方法,其中所述重组蛋白产物中的残余宿主细胞蛋白的水平小于100 ppm。

15. 权利要求1至14中任一项的方法,其中所述重组蛋白产物选自抗体、抗体片段、疫苗、生长因子、细胞因子、激素或凝结因子。

16. 用于生物生产系统中的哺乳动物细胞系,其中所述哺乳动物细胞系被工程改造以减少或消除一种或多种宿主细胞蛋白的表达,所述宿主细胞蛋白选自羧肽酶B1、羧肽酶D、

羧肽酶E、羧肽酶M、组织蛋白酶B、组织蛋白酶D、组织蛋白酶L1、组织蛋白酶Z、硫酸软骨素蛋白聚糖4、簇蛋白、二肽基肽酶3、豆类蛋白酶 (legumain)、亮氨酸氨基肽酶3、脂蛋白脂肪酶、赖氨酰氧化酶、金属蛋白酶抑制剂1、中性 $\alpha$ -葡萄糖苷酶、巢蛋白1、过氧蛋白酶 (peroxidase)、磷脂酶B-样2、脯氨酰内肽酶、蛋白精氨酸N-甲基转移酶5、蛋白磷酸酶1G、丝氨酸蛋白酶、唾液酸酶1、硫氧还蛋白或硫氧还蛋白还原酶。

17. 权利要求16的哺乳动物细胞系,其中所述一种或多种宿主蛋白选自羧肽酶D、组织蛋白酶B、组织蛋白酶D、簇蛋白、脂蛋白脂肪酶、金属蛋白酶抑制剂1、巢蛋白1、过氧蛋白酶 (peroxidase)、丝氨酸蛋白酶、硫氧还蛋白或硫氧还蛋白还原酶。

18. 权利要求16或17的哺乳动物细胞系,其中所述至少一种宿主细胞蛋白的表达经由编码所述至少一种宿主细胞蛋白的染色体序列的至少一个等位基因的失活而减少或消除。

19. 权利要求18的哺乳动物细胞系,其中编码所述至少一种宿主细胞蛋白的染色体序列的两个等位基因均被失活。

20. 权利要求18或19的哺乳动物细胞系,其中使用靶向核酸内切酶介导的基因组修饰技术来使所述染色体序列失活。

21. 权利要求20的哺乳动物细胞系,其中所述靶向核酸内切酶是核糖核蛋白复合物或一对锌指核酸酶。

22. 权利要求16至21中任一项的哺乳动物细胞系,其中所述细胞系是人细胞系。

23. 权利要求22的哺乳动物细胞系,其中所述人细胞系是人胚胎肾细胞293 (HEK293)细胞系、HT-1080人结缔组织系或PER.C6人胚胎视网膜细胞系。

24. 权利要求16至21中任一项的哺乳动物细胞系,其中所述细胞系是非人细胞系。

25. 权利要求24的哺乳动物细胞系,其中所述非人细胞系是中国仓鼠卵巢 (CHO) 细胞系、幼仓鼠肾 (BHK) 细胞系、NS0小鼠骨髓瘤细胞系、Sp2/0小鼠骨髓瘤细胞系、C127小鼠乳腺细胞系或Vero非洲绿猴肾细胞系。

26. 权利要求16至21中任一项的哺乳动物细胞系,其中所述细胞系是CHO细胞系。

27. 权利要求16至26中任一项的哺乳动物细胞系,其中细胞活力、活细胞密度、滴度、生长速率、增殖应答、细胞形态和/或总体细胞健康与非工程改造的亲本哺乳动物细胞系的那些相当。

28. 权利要求16至27中任一项的哺乳动物细胞系,其进一步包含至少一种编码重组蛋白的核酸,所述重组蛋白选自抗体、抗体片段、疫苗、生长因子、细胞因子、激素或凝集因子。

## 具有修饰的宿主细胞蛋白概况的工程改造的细胞

### [0001] 领域

本公开涉及用于生物生产系统中的哺乳动物细胞系,其中所述哺乳动物细胞系被工程改造以减少或消除污染常规生产的重组治疗性蛋白的宿主细胞蛋白的表达。

### [0002] 背景

在重组蛋白生产期间,宿主细胞共同产生与正常细胞功能(诸如细胞生长、增殖、存活、基因转录、蛋白合成等)相关的内源蛋白。由于细胞死亡/凋亡/裂解,内源宿主细胞蛋白也可以被释放至细胞培养基中。在重组蛋白生产期间共表达的所有内源蛋白都称为宿主细胞蛋白(HCP)。HCP构成重组治疗性蛋白(诸如单克隆抗体)中存在的过程相关的杂质的主要部分。这些HCP杂质可以显著影响治疗性蛋白的效力和稳定性,以及引起免疫原性。此外,与治疗性蛋白共纯化的HCP可能难以除去,导致大量下游处理和生产成本增加。例如,已经估计,单克隆抗体生产成本的约80%是由于下游纯化过程。此外,为了满足法规要求,制造商必须表明最终产品中的宿主细胞蛋白的清除率达到1至100 ppm的范围内的水平。

[0003] 因此,需要在治疗性蛋白生产期间减少或消除特定HCP的方式。例如,需要经工程改造以减少或消除HCP的表达的宿主细胞系,所述HCP是丰富的,在下游处理期间难以除去和/或影响产品质量。此类细胞系将简化并降低生物治疗剂生产的成本。

### [0004] 概述

在本公开的各个方面中,提供了用于生物生产系统中的哺乳动物细胞系,其中所述哺乳动物细胞系被工程改造以减少或消除一种或多种宿主细胞蛋白的表达,所述宿主细胞蛋白选自:羧肽酶B1、羧肽酶D、羧肽酶E、羧肽酶M、组织蛋白酶B、组织蛋白酶D、组织蛋白酶L1、组织蛋白酶Z、硫酸软骨素蛋白聚糖4、簇蛋白、二肽基肽酶3、豆荚蛋白酶(legumain)、亮氨酸氨基肽酶3、脂蛋白脂肪酶、赖氨酰氧化酶、金属蛋白酶抑制剂1、中性 $\alpha$ -葡糖苷酶、巢蛋白1、过氧蛋白酶(oxidase)、磷脂酶B-样2、脯氨酰内肽酶、蛋白精氨酸N-甲基转移酶5、蛋白磷酸酶1G、丝氨酸蛋白酶、唾液酸酶1、硫氧还蛋白或硫氧还蛋白还原酶。通常,所述一种或多种蛋白的表达经由使编码所述蛋白的染色体序列的至少一个等位基因的失活而降低。可以使用靶向核酸内切酶介导的基因组修饰(例如,CRISPR核糖核蛋白(RNP)复合物或锌指核酸酶)使染色体序列失活。

[0005] 本公开的另一个方面涵盖用于生产具有降低的宿主细胞蛋白污染水平的重组蛋白产物的方法。所述方法包括在本文公开的任何哺乳动物细胞系中表达重组蛋白,和纯化所述重组蛋白以形成所述重组蛋白产物,其中所述重组蛋白产物的残余宿主细胞蛋白污染水平低于由非工程改造的亲本哺乳动物细胞系产生的蛋白产物中的残余宿主细胞蛋白污染水平。

[0006] 下面更详细地描述了本公开的其他方面和迭代。

### [0007] 附图的简要说明

图1显示模拟物转染或用数对靶向脂蛋白脂肪酶(LPL)或磷脂酶B-样2(PLBL2)的ZFN转染的细胞中的核苷酸错配测定(Ce11测定)的结果。

[0008] 图2呈现在第7天或第15天在模拟物转染或用靶向组织蛋白酶B或组织蛋白酶D的

Cas9 RNP转染的细胞中的核苷酸错配测定 (Cell1测定) 的结果。

[0009] 图3A显示在第10天分批进料的样品中的组织蛋白酶B敲除克隆的生产力和生长概况。

[0010] 图3B呈现在第10天分批进料的样品中的组织蛋白酶D敲除克隆和野生型细胞的生产力和生长概况。

[0011] 图4显示在模拟物转染 (泳道2-4) 或用靶向簇蛋白的Cas9 RNP转染 (泳道5-7) 的细胞中的核苷酸错配测定 (Cell1测定) 的结果。

[0012] 图5A呈现野生型克隆的生产力和生长概况。

[0013] 图5B呈现簇蛋白敲除克隆的生产力和生长概况。

[0014] 图6显示在模拟物转染 (泳道1和6)、用靶向硫氧还蛋白的Cas9 RNPs转染 (泳道2-5) 或用靶向硫氧还蛋白还原酶的Cas9 RNP转染 (泳道7-10) 的细胞中的核苷酸错配测定 (Cell1测定) 的结果。

[0015] 详述

本公开提供了哺乳动物细胞系,其经工程改造以减少或消除特定宿主细胞蛋白的表达,使得由所述细胞系产生的重组蛋白具有非常低水平的污染宿主细胞蛋白。提供了用于产生所述工程改造的细胞系的方法,以及使用所述工程改造的细胞系产生具有低残余宿主细胞蛋白水平的重组蛋白的方法。

[0016] (I) *工程改造的细胞系*

本公开的一个方面涵盖经工程改造以减少或消除一种或多种宿主细胞蛋白 (HCP) 的表达的哺乳动物细胞系。因此,与由非工程改造的亲本细胞 (即,其所述HCP的表达未改变的亲本细胞) 产生的重组蛋白相比,由本文公开的工程改造的细胞系产生的重组蛋白具有降低的一种或多种HCP的水平。

[0017] (a) *靶标HCP*

本文公开的工程改造的细胞系具有减少或消除的一种或多种HCP的表达。如以下实施例1中详述,已经在几种宿主细胞系中鉴定了HCP的子集。这些HCP是高度丰富的,难以在下游纯化过程期间除去,和/或影响产品质量 (例如,残余的蛋白酶可能降解生物治疗产品,由此降低其效力)。具有这些特征的HCP被称为“有问题的” HCP。

[0018] 表A列出靶标HCP,其表达可以在工程改造的细胞系中减少或消除。通常,所述靶标HCP是对于细胞存活和/或细胞功能不是必需的蛋白。

表 A. 靶标宿主细胞蛋白		
蛋白	基因 ID	登录号
羧肽酶 B1	Cpb1	NW_003614612.1
羧肽酶 D	Cpd	NW_003614210.1
羧肽酶 E	Cpe	NW_003613758.1
羧肽酶 M	Cpm	NW_003613642.1
组织蛋白酶 B	Ctsb	NW_003613673.1
组织蛋白酶 D	Ctsd	NW_003614851.1
组织蛋白酶 L1	Ctsl1	NW_003618581.1
组织蛋白酶 Z	Ctsz	NW_003613633.1
硫酸软骨素蛋白聚糖 4	Cspg4	NW_003613669.1
簇蛋白	Clu	NW_003614124.1
二肽基肽酶 3	Dpp3	NW_003614712.1
豆荚蛋白酶(legumain)	Lgmn	NW_003614655.1
亮氨酸氨基肽酶 3	Lap3	NW_003614204.1
脂蛋白脂肪酶	Lpl	NW_003613760.1
赖氨酰氧化酶	Lox	NW_003616945.1
金属蛋白酶抑制剂 1	Timp1	NW_003615404.1
中性 $\alpha$ -葡萄糖苷酶	Ganc	NW_003613677.1
巢蛋白 1	Nid1	NW_003614408.1
过氧蛋白酶(oxidase)	Pxdn	NW_003613848.1
磷脂酶 B-样 2	Plbd2	NW_003614971.1
脯氨酰内肽酶	Prep	NW_003614750.1
蛋白精氨酸 N-甲基转移酶 5	Prmt5	NW_003614213.1
蛋白磷酸酶 1G	Ppm1g	NW_003613640.1
丝氨酸蛋白酶	Htra1	NW_003615400.1
唾液酸酶 1	Neu1	NM_001246800.1
硫氧还蛋白	Trx	NW_003614309.1
硫氧还蛋白还原酶	TxnRd1	NW_003614195.1

[0019] 在一些实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的一种蛋白的表达。在其他实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的两种蛋白的表达。在进一步实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的三种蛋白的表达。在还有其他实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的四种蛋白的表达。在额外实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的五种蛋白的表达。在进一步实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的六种蛋白的表达。在还有其他实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的七种蛋白的表达。在进一步实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的八种蛋白的表达。在额外实施方案中,所述工程改造的细胞系具有减少或消除的表A中所列的八种或更多种蛋白的表达。

[0020] 在一个实施方案中,所述工程改造的细胞系具有减少或消除的组织蛋白酶B、组织蛋白酶D、组织蛋白酶L1和/或组织蛋白酶Z的表达。在另一个实施方案中,所述工程改造的细胞系具有减少或消除的磷脂酶B-样2和/或脂蛋白脂肪酶的表达。在一个进一步实施方案

中,所述工程改造的细胞系具有减少或消除的组织蛋白酶B、组织蛋白酶D、组织蛋白酶L1、组织蛋白酶Z、羧肽酶D、羧肽酶M、羧肽酶B1、羧肽酶E、磷脂酶B-样2、脂蛋白脂肪酶、过氧蛋白酶(peroxidase)、丝氨酸蛋白酶、中性 $\alpha$ -葡糖苷酶、赖氨酰氧化酶和/或二肽基肽酶3的表达。

[0021] 在其他实施方案中,所述工程改造的细胞系具有减少或消除的以下中的一种或多种的表达:羧肽酶D、组织蛋白酶D、簇蛋白、脂蛋白脂肪酶、金属蛋白酶抑制剂1、巢蛋白、过氧蛋白酶(peroxidase)、磷脂酶B-样2、丝氨酸蛋白酶、硫氧还蛋白和/或硫氧还蛋白还原酶。

[0022] 将本文公开的具有减少或消除的一种或多种目标HCP的表达的细胞系进行基因工程改造以修饰编码目标HCP的染色体序列。目标染色体序列可以使用靶向的核酸内切酶介导的基因组编辑技术进行修饰,这在下面的部分(III)中详述。例如,可以修饰染色体序列以包含至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的取代或其组合,使得使阅读框移位并且没有产生蛋白产物(即染色体序列被失活)。编码目标HCP的染色体序列的一个等位基因的失活导致目标HCP的表达降低(即,敲低)。编码目标HCP的染色体序列的两个等位基因的失活导致目标HCP不表达(即,敲除)。

[0023] 在一些实施方案中,目标HCP的水平可以降低至少约5%,至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约60%,至少约70%,至少约80%,至少约90%,至少约95%,至少约99%,或超过约99%。在其他实施方案中,可以将目标HCP的水平降低至使用本领域中标准的技术(例如,Western免疫印迹测定,ELISA酶测定,SDS聚丙烯酰胺凝胶电泳等)检测不到的水平。

[0024] 通常,本文公开的工程改造的细胞系的细胞活力、活细胞密度、滴度、生长速率、增殖应答、细胞形态、凋亡和自噬水平和/或总体细胞健康与其非工程改造的亲本细胞的那些类似。

#### [0025] (b) 细胞类型

本文公开的工程改造的细胞系是哺乳动物细胞系。在一些实施方案中,所述工程改造的细胞系可以源自人细胞系。合适的人细胞系的非限制性实例包括人胚胎肾细胞(HEK293, HEK293T);人结缔组织细胞(HT-1080);人宫颈癌细胞(HELA);人胚胎视网膜细胞(PER.C6);人肾细胞(HKB-11);人肝细胞(Huh-7);人肺细胞(W138);人肝细胞(Hep G2);人U2-OS骨肉瘤细胞,人A549肺细胞,人A-431表皮细胞或人K562骨髓细胞。在其他实施方案中,所述工程改造的细胞系可以源自非人细胞系。合适的细胞系不加限制地包括中国仓鼠卵巢(CHO)细胞;幼仓鼠肾(BHK)细胞;小鼠骨髓瘤NS0细胞;小鼠骨髓瘤Sp2/0细胞;小鼠乳腺C127细胞;小鼠胚胎成纤维细胞3T3细胞(NIH3T3);小鼠B淋巴瘤A20细胞;小鼠黑色素瘤B16细胞;小鼠成肌细胞C2C12细胞;小鼠胚胎间质C3H-10T1/2细胞;小鼠癌CT26细胞、小鼠前列腺DuCuP细胞;小鼠乳房EMT6细胞;小鼠肝癌Hepa1c1c7细胞;小鼠骨髓瘤J5582细胞;小鼠上皮MTD-1A细胞;小鼠心肌MyEnd细胞;小鼠肾RenCa细胞;小鼠胰腺RIN-5F细胞;小鼠黑色素瘤X64细胞;小鼠淋巴瘤YAC-1细胞;大鼠成胶质细胞瘤9L细胞;大鼠B淋巴瘤RBL细胞;大鼠成神经细胞瘤B35细胞;大鼠肝癌细胞(HTC);水牛大鼠肝BRL 3A细胞;犬肾细胞(MDCK);犬乳腺(CMT)细胞;大鼠骨肉瘤D17细胞;大鼠单核细胞/巨噬细胞DH82细胞;猴肾SV-40转化的成纤维细胞(COS7)细胞;猴肾CVI-76细胞;或非非洲绿猴肾(VERO, VERO-76)细胞。哺乳动物细胞系的详

尽列表可见于美国典型培养物保藏中心(American Type Culture Collection)目录(ATCC, Mamassas, VA)中。在一些实施方案中,本文公开的细胞系不同于小鼠细胞系。在某些实施方案中,所述工程改造的细胞系是CHO细胞系。合适的CHO细胞系包括但不限于CHO-K1、CHO-K1SV、CHO GS<sup>-/-</sup>、CHO S、DG44、DuxxB11及其衍生物。

[0026] 在各个实施方案中,所述亲本细胞系可以是谷氨酰胺合成酶(GS)、二氢叶酸还原酶(DHFR)、次黄嘌呤-鸟嘌呤磷酸核糖转移酶(HPRT)或其组合缺陷的。例如,可以使编码GS、DHFR和/或HPRT的染色体序列失活。在具体实施方案中,编码GS、DHFR和/或HPRT的所有染色体序列在亲本细胞系中都是失活的。

[0027] (c) 任选的编码重组蛋白的核酸

在一些实施方案中,本文公开的工程改造的细胞系可以进一步包含至少一种编码重组蛋白的核酸。通常,所述重组蛋白是异源的,这意指所述蛋白对于所述细胞不是天然的。所述重组蛋白可以不加限制地是选自以下的治疗性蛋白:抗体、抗体的片段、单克隆抗体、人源化抗体、人源化单克隆抗体、嵌合抗体、IgG分子、IgG重链、IgG轻链、IgA分子、IgD分子、IgE分子、IgM分子、疫苗、生长因子、细胞因子、干扰素、白介素、激素、凝结(或凝血)因子、血液组分、酶、治疗性蛋白、营养食品蛋白、前述任一种的功能片段或功能变体、或包含前述蛋白和/或其功能片段或变体的任一种的融合蛋白。

[0028] 在一些实施方案中,编码所述重组蛋白的核酸可以连接至编码次黄嘌呤-鸟嘌呤磷酸核糖转移酶(HPRT)、二氢叶酸还原酶(DHFR)和/或谷氨酰胺合成酶(GS)的序列,使得HPRT、DHFR和/或GS可以用作可扩增的可选择标记。编码所述重组蛋白的核酸也可以与编码至少一种抗生素抗性基因的序列和/或编码标记蛋白、诸如荧光蛋白的序列连接。在一些实施方案中,编码所述重组蛋白的核酸可以是表达构建体的一部分。所述表达构建体或载体可以包含额外表达控制序列(例如增强子序列、Kozak序列、聚腺苷酸化序列、转录终止序列等)、可选择标记序列、复制起点等。额外信息可见于“Current Protocols in Molecular Biology” Ausubel等人, John Wiley & Sons, New York, 2003或“Molecular Cloning: A Laboratory Manual” Sambrook和Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 第3版, 2001中。

[0029] 在一些实施方案中,编码所述重组蛋白的核酸可以位于染色体外。也就是说,编码所述重组蛋白的核酸可以从质粒、粘粒、人工染色体、微型染色体或另一染色体外构建体瞬时表达。在其他实施方案中,编码所述重组蛋白的核酸可以染色体性地整合入细胞的基因组中。所述整合可以是随机或靶向的。因此,可以稳定表达所述重组蛋白。在该实施方案的一些迭代中,编码所述重组蛋白的核酸序列可以可操作地连接至适当的异源表达控制序列(即,启动子)。在其他迭代中,编码所述重组蛋白的核酸序列可以被置于内源表达控制序列的控制下。可以使用同源重组、靶向核酸内切酶介导的基因组编辑、病毒载体、转座子、质粒和其他众所周知的方式将编码所述重组蛋白的核酸序列整合入细胞系的基因组中。额外指导可见于Ausubel等人2003(同上)以及Sambrook和Russell, 2001(同上)中。

[0030] (II) 试剂盒

本公开的一个进一步方面提供了用于产生重组蛋白的试剂盒,其中试剂盒包含上面在部分(I)中详述的任何工程改造的细胞系。试剂盒可以进一步包含细胞生长培养基、转染试剂、选择培养基、重组蛋白纯化装置、缓冲液等。本文提供的试剂盒通常包括用于使细胞系

生长并使用它们产生重组蛋白的说明书。所述试剂盒中包括的说明书可以贴至包装材料，或者可以作为包装插页包括。尽管说明书通常是书面或印刷材料，但它们不限于此。本公开考虑能够存储此类说明书并将其传达给最终用户的任何介质。此类介质包括但不限于电子存储介质（例如，磁盘、磁带、盒带、芯片）、光学介质（例如，CD ROM）等。如本文所用，术语“说明书”可以包括提供说明书的互联网站点的地址。

[0031] (III) 用于制备工程改造的细胞系的方法

本公开的又一个方面提供了用于制备或工程改造具有减少或消除的一种或多种HCP的表达的细胞系的方法，其在上文描述于部分(I)中。编码目标HCP的染色体序列可以使用各种技术进行敲低或敲除。通常，使用靶向核酸内切酶介导的基因组修饰方法来制备所述工程改造的细胞系。本领域技术人员理解，也可以使用位点特异性重组系统、随机诱变或本领域中已知的其他方法来制备所述工程改造的细胞系。

[0032] 通常，通过以下方法制备工程改造的细胞系，所述方法包括将至少一种靶向核酸内切酶或编码所述靶向核酸内切酶的核酸引入目标亲本细胞系中，其中所述靶向核酸内切酶被靶向至编码目标HCP的染色体序列。所述靶向核酸内切酶识别并结合特定染色体序列且引入双链断裂。在一些实施方案中，所述双链断裂通过非同源末端接合(NHEJ)修复过程来修复。因为NHEJ是易错的，所以可能发生至少一个核苷酸的缺失、插入和/或取代，由此破坏染色体序列的阅读框，使得不产生蛋白产物。在其他实施方案中，所述靶向核酸内切酶也可用于通过共同引入与靶向的染色体序列的一部分具有实质性序列同一性的多核苷酸，经由同源重组反应来改变染色体序列。在此类情况下，由所述靶向核酸内切酶引入的双链断裂通过同源性引导的修复过程来修复，使得导致染色体序列被变化或改变（例如，通过外源序列的整合）的方式使染色体序列与多核苷酸交换。

[0033] (a) 靶向核酸内切酶

各种靶向核酸内切酶可用于修饰编码目标HCP的染色体序列。所述靶向核酸内切酶可以是天然存在的蛋白或工程改造的蛋白。合适的靶向核酸内切酶不加限制地包括锌指核酸酶(ZFN)、CRISPR核酸酶、转录激活因子样效应物(TALE)核酸酶(TALEN)、大范围核酸酶、嵌合核酸酶、位点特异性核酸内切酶和人工靶向的DNA双链断裂诱导剂。

[0034] (i) 锌指核酸酶

在具体实施方案中，所述靶向核酸内切酶可以是一对锌指核酸酶(ZFN)。ZFN结合特定的靶向的序列并将双链断裂引入靶向的切割位点。通常，ZFN包含DNA结合结构域（即，锌指）和切割结构域（即，核酸酶），其各自在下面描述。

[0035] DNA结合结构域。DNA结合结构域或锌指可以被工程改造以识别和结合任何所选核酸序列。参见，例如，Beerli等人(2002) Nat. Biotechnol. 20:135-141；Pabo等人(2001) Ann. Rev. Biochem. 70:313-340；Isalan等人(2001) Nat. Biotechnol. 19:656-660；Segal等人(2001) Curr. Opin. Biotechnol. 12:632-637；Choo等人(2000) Curr. Opin. Struct. Biol. 10:411-416；Zhang等人(2000) J. Biol. Chem. 275(43):33850-33860；Doyon等人(2008) Nat. Biotechnol. 26:702-708；和Santiago等人(2008) Proc. Natl. Acad. Sci. USA 105:5809-5814。与天然存在的锌指蛋白相比，工程改造的锌指结合结构域可以具有新型结合特异性。工程改造方法包括但不限于合理设计和各种类型的选择。合理设计包括例如使用包含双联体、三联体和/或四联体核苷酸序列

和个别锌指氨基酸序列的数据库,其中各双联体、三联体或四联体核苷酸序列与结合特定三联体或四联体序列的锌指的一个或多个氨基酸序列相关。参见,例如,美国专利号6,453,242和6,534,261,其公开内容通过引用以其整体并入本文。作为一个实例,美国专利6,453,242中描述的算法可用于设计锌指结合结构域以靶向预先选择的序列。替代方法,诸如使用非简并识别代码表的合理设计也可用于设计锌指结合结构域以靶向特定序列(Sera等人(2002) *Biochemistry* 41:7074-7081)。用于鉴定DNA序列中的潜在靶标位点以及设计锌指结合结构域的公开可用的基于网络的工具是本领域中已知的。例如,用于鉴定DNA序列中的潜在靶标位点的工具可见于zincfingertools.org。用于设计锌指结合结构域的工具可见于zifit.partners.org/ZiFiT。(也参见,Mandell等人(2006) *Nuc. Acid Res.* 34:W516-W523; Sander等人(2007) *Nuc. Acid Res.* 35:W599-W605.)。

[0036] 可以设计锌指结合结构域以识别并结合范围为约3个核苷酸至约21个核苷酸长度的DNA序列。在一个实施方案中,可以设计锌指结合结构域以识别并结合范围为约9至约18个核苷酸长度的DNA序列。通常,本文使用的锌指核酸酶的锌指结合结构域包含至少三个锌指识别区域或锌指,其中各锌指结合3个核苷酸。在一个实施方案中,所述锌指结合结构域包含四个锌指识别区域。在另一个实施方案中,所述锌指结合结构域包含五个锌指识别区域。在又一个实施方案中,所述锌指结合结构域包含六个锌指识别区域。可以设计锌指结合结构域以结合任何合适的靶标DNA序列。参见例如,美国专利号6,607,882;6,534,261和6,453,242,其公开内容通过引用以其整体并入本文。

[0037] 选择锌指识别区域的示例性方法包括噬菌体展示和双杂交系统,其描述于美国专利号5,789,538;5,925,523;6,007,988;6,013,453;6,410,248;6,140,466;6,200,759;和6,242,568;以及WO 98/37186;WO 98/53057;WO 00/27878;WO 01/88197和GB 2,338,237,其各自通过引用以其整体并入本文。此外,锌指结合结构域的结合特异性的增强已经描述于例如WO 02/077227,其整个公开内容通过引用并入本文。

[0038] 锌指结合结构域和用于设计和构建融合蛋白(和编码其的多核苷酸)的方法是本领域技术人员已知的且详细描述于例如美国专利号7,888,121,其通过引用以其整体并入本文。可以使用合适的接头序列(包括例如长度为五个或更多个氨基酸的接头)将锌指识别区域和/或多指锌指蛋白连接在一起。对于长度为六个或更多个氨基酸的接头序列的非限制性实例,参见美国专利号6,479,626;6,903,185;和7,153,949,其公开内容通过引用以其整体并入本文。本文所述的锌指结合结构域可以在蛋白的个别锌指之间包括合适的接头的组合。

[0039] 切割结构域。锌指核酸酶也包括切割结构域。所述锌指核酸酶的切割结构域部分可以获得自任何核酸内切酶或核酸外切酶。可以衍生出切割结构域的核酸内切酶的非限制性实例包括但不限于限制核酸内切酶和归巢核酸内切酶。参见,例如,New England Biolabs目录或Belfort等人(1997) *Nucleic Acids Res.* 25:3379-3388。切割DNA的额外酶是已知的(例如,S1核酸酶;绿豆核酸酶;胰腺DNA酶I;微球菌核酸酶;酵母HO核酸内切酶)。还参见Linn等人(编) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993。这些酶(或其功能片段)中的一种或多种可用作切割结构域的来源。

[0040] 切割结构域也可以源自对于切割活性需要二聚化的如上所述的酶或其部分。两个锌指核酸酶可以是切割所需的,因为各核酸酶包含活性酶二聚体的单体。或者,单一锌指核

酸酶可以包含两个单体以产生活性酶二聚体。如本文所用,“活性酶二聚体”是能够切割核酸分子的酶二聚体。两个切割单体可以源自相同的核酸内切酶(或其功能片段),或各单体可以源自不同的核酸内切酶(或其功能片段)。

[0041] 当两个切割单体用于形成活性酶二聚体时,两个锌指的识别位点优选被安置,使得两个锌指与它们的相应识别位点的结合使切割单体彼此处于允许切割单体例如通过二聚化来形成活性酶二聚体的空间取向。作为结果,识别位点的近侧边缘可由约5至约18个核苷酸分开。例如,近侧边缘可由约5、6、7、8、9、10、11、12、13、14、15、16、17或18个核苷酸分开。然而,应理解任何整数个核苷酸或核苷酸对都可插入两个识别位点之间(例如约2至约50个核苷酸对或更多个核苷酸对)。锌指核酸酶的识别位点的近侧边缘(诸如例如本文详细描述的那些)可由6个核苷酸分开。通常,切割位点位于所述识别位点之间。

[0042] 限制核酸内切酶(限制酶)存在于许多物种中且能够序列特异性结合DNA(在识别位点处),且在结合位点处或附近切割DNA。某些限制酶(例如IIS型)在远离识别位点的位点处切割DNA且具有可分开的结合结构域和切割结构域。例如,IIS型酶FokI催化在距它的处于一条链上的识别位点9个核苷酸处且在距它的处于另一条链上的识别位点13个核苷酸处的DNA的双链切割。参见,例如,美国专利号5,356,802;5,436,150和5,487,994;以及Li等人(1992) Proc. Natl. Acad. Sci. USA 89:4275-4279;Li等人(1993) Proc. Natl. Acad. Sci. USA 90:2764-2768;Kim等人(1994a) Proc. Natl. Acad. Sci. USA 91:883-887;Kim等人(1994b) J. Biol. Chem. 269:31978-31982。因此,锌指核酸酶可以包含来自至少一种IIS型限制酶的切割结构域和一个或多个锌指结合结构域,其可以工程改造或可以不工程改造。示例性IIS型限制酶例如描述于国际公开WO 07/014,275,其公开内容通过引用以其整体并入本文。额外限制酶也含有可分开的结合结构域和切割结构域,且这些酶也由本公开涵盖。参见,例如,Roberts等人(2003) Nucleic Acids Res. 31:418-420。

[0043] 其切割结构域可与结合结构域分开的示例性IIS型限制酶是FokI。该特定酶作为二聚体而有活性(Bitinaite等人(1998) Proc. Natl. Acad. Sci. USA 95: 10, 570-10, 575)。因此,出于本公开的目的,锌指核酸酶中使用的FokI酶的部分被认为是切割单体。因此,对于使用FokI切割结构域的靶向的双链切割,可以使用两种锌指核酸酶(其各自包含FokI切割单体)来重构活性酶二聚体。或者,也可以使用含有锌指结合结构域和两个FokI切割单体的单个多肽分子。

[0044] 在某些实施方案中,所述切割结构域包含一个或多个使同二聚化最少或防止同二聚化的工程改造的切割单体。通过非限制实例的方式,在FokI的位置446、447、479、483、484、486、487、490、491、496、498、499、500、531、534、537和538处的氨基酸残基都是影响FokI切割半结构域的二聚化的靶标。形成专性异二聚体的FokI的示例性工程改造的切割单体包括一对,其中第一切割单体在FokI的氨基酸残基位置490和538处包括突变且第二切割单体在氨基酸残基位置486和499处包括突变。

[0045] 因此,在工程改造的切割单体的一个实施方案中,在氨基酸位置490处的突变将Glu(E)替换为Lys(K);在氨基酸残基538处的突变将Iso(I)替换为Lys(K);在氨基酸残基486处的突变将Gln(Q)替换为Glu(E);且在位置499处的突变将Iso(I)替换为Lys(K)。具体地,工程改造的切割单体可通过如下来制备:在一个切割单体中将位置490从E突变为K且将

位置538从I突变为K来产生称为“E490K:I538K”的工程改造的切割单体,且在另一切割单体中将位置486从Q突变为E且将位置499从I突变为K来产生称为“Q486E:I499K”的工程改造的切割单体。上述工程改造的切割单体是使异常切割最少或废除的专性异二聚体突变体。工程改造的切割单体可使用合适的方法,例如,如美国专利号7,888,121(其以其整体并入本文)中所述,通过野生型切割单体(FokI)的定点诱变来制备。

[0046] 额外结构域。在一些实施方案中,所述锌指核酸酶进一步包含至少一个核定位序列(NLS)。NLS是有助于将锌指核酸酶蛋白靶向至核中以在染色体中的靶标序列处引入双链断裂的氨基酸序列。核定位信号是本领域中已知的(参见,例如,Lange等人, J. Biol. Chem., 2007, 282:5101-5105)。核定位信号的非限制性实例包括PKKKRKY (SEQ ID NO:1)、PKKKRRV (SEQ ID NO:2)、KRPAATKKAGQAKKKK (SEQ ID NO:3)、YGRKKRRQRRR (SEQ ID NO:4)、RKKRRQRRR (SEQ ID NO:5)、PAAKRVKLD (SEQ ID NO:6)、RQRRNELKRSP (SEQ ID NO:7)、VSRKRPRP (SEQ ID NO:8)、PPKKARED (SEQ ID NO:9)、PQPKKKPL (SEQ ID NO:10)、SALIKKKKMAP (SEQ ID NO:11)、PKQKKRK (SEQ ID NO:12)、RKLKKKIKKL (SEQ ID NO:13)、REKKKFLKRR (SEQ ID NO:14)、KRKGDEV DGVDEVAKKKSKK (SEQ ID NO:15)、RKCLQAGMNLEARKTKK (SEQ ID NO:16)、NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:17)和RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:18)。所述NLS可以位于所述锌指核酸酶的N-末端、C-末端或内部位置。

[0047] 在额外实施方案中,所述锌指核酸酶也可以包含至少一个细胞穿透结构域。合适的细胞穿透结构域的实例包括但不限于,GRKKRRQRRRPPQPKKKRKY (SEQ ID NO:19)、PLSSIFSRIGDPPKKKRKY (SEQ ID NO:20)、GALFLGWLGAAGSTMGAPKKKRKY (SEQ ID NO:21)、GALFLGLGAAGSTMGAWSQP KKKRKY (SEQ ID NO:22)、KETWWETWWTEWSQP KKKRKY (SEQ ID NO:23)、YARAAARQARA (SEQ ID NO:24)、THRLPRRRRRR (SEQ ID NO:25)、GGRRARRRRR (SEQ ID NO:26)、RRQRRTSKLMKR (SEQ ID NO:27)、GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:28)、KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:29)和RQIKIWFQNRMMKWK (SEQ ID NO:30)。所述细胞穿透结构域可以位于所述锌指核酸酶的N-末端、C-末端或内部位置。

[0048] 在还有其他实施方案中,所述锌指核酸酶可以进一步包含至少一个标记结构域。标记结构域的非限制性实例包括荧光蛋白、纯化标签和表位标签。在一个实施方案中,所述标记结构域可以是荧光蛋白。合适的荧光蛋白的非限制性实例包括绿色荧光蛋白(例如,GFP、GFP-2、tagGFP、turboGFP、EGFP、Emerald、Azami Green、单体Azami Green、CopGFP、AceGFP、ZsGreen1),黄色荧光蛋白(例如YFP、EYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1),蓝色荧光蛋白(例如EBFP、EBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire),青色荧光蛋白(例如ECFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan),红色荧光蛋白(mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-单体、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred)和橙色荧光蛋白(mOrange、mKO、Kusabira-Orange、单体Kusabira-Orange、mTangerine、tdTomato)或任何其他合适的荧光蛋白。在另一个实施方案中,所述标记结构域可以是纯化标签和/或表位标签。合适的标签包括但不限于多(His)标签、FLAG(或DDK)标签、Halo标签、AcV5标签、AU1标签、AU5标签、生物素羧基载体蛋白(BCCP)、钙调蛋白结合蛋白(CBP)、几丁质结合结构域(CBD)、E标签、E2标签、ECS标签、eXact标签、Glu-Glu标签、谷胱甘肽-S-转移

酶(GST)、HA标签、HSV标签、KT3标签、麦芽糖结合蛋白(MBP)、MAP标签、Myc标签、NE标签、NusA标签、PDZ标签、S标签、S1标签、SBP标签、Softag 1标签、Softag 3标签、Spot标签、Strep标签、SUMO标签、T7标签、串联亲和纯化(TAP)标签、硫氧还蛋白(TRX)、V5标签、VSV-G标签和Xa标签。所述标记结构域可以位于所述锌指核酸酶的N-末端、C-末端或内部位置。

[0049] 所述至少一个核定位信号、至少一个细胞穿透结构域和/或至少一个标记结构域可以经由一个或多个化学键(例如,共价键)直接连接至所述锌指核酸酶。或者,所述至少一个核定位信号、至少一个细胞穿透结构域和/或至少一个标记结构域可以经由一个或多个接头间接连接至所述锌指核酸酶。合适的接头包括氨基酸,肽,核苷酸,核酸,有机接头分子(例如,马来酰亚胺衍生物,N-乙氧基苄基咪唑,联苯-3,4',5-三羧酸,对氨基苄氧基羰基等),二硫化物接头,和聚合物接头(例如,PEG)。所述接头可以包括一个或多个间隔基团,包括但不限于亚烷基、亚烯基、亚炔基、烷基、烯基、炔基、烷氧基、芳基、杂芳基、芳烷基、芳烯基、芳炔基等。所述接头可以是中性的,或携带正电荷或负电荷。另外,所述接头可以是可切割的,使得连接所述接头与另一个化学基团的接头的共价键可以在某些条件(包括pH、温度、盐浓度、光、催化剂或酶)下被断裂或切割。在一些实施方案中,所述接头可以是肽接头。所述肽接头可以是柔性氨基酸接头或刚性氨基酸接头。合适接头的额外实例是本领域中众所周知的,并且设计接头的程序是容易可得的(Crasto等人, Protein Eng., 2000, 13(5):309-312)。

#### [0050] (ii) CRISPR核糖核蛋白(RNP)

在其他实施方案中,所述靶向核酸内切酶可以是聚集的规则散布的短回文重复(CRISPR)核酸酶。CRISPR核酸酶是源自细菌或古细菌CRISPR/CRISPR-相关(Cas)系统的RNA-指导的核酸酶。CRISPR RNP系统包含CRISPR核酸酶和指导RNA。

[0051] 核酸酶。CRISPR核酸酶可以源自I型(即IA、IB、IC、ID、IE或IF)、II型(即IIA、IIB或IIC)、III型(即IIIA或IIIB)、V型或VI型CRISPR系统,其存在于各种细菌和古细菌中。例如,CRISPR核酸酶可以来自链球菌属物种(*Streptococcus sp.*) (例如酿脓链球菌(*S. pyogenes*),嗜热链球菌(*S. thermophilus*),巴氏链球菌(*S. pasteurianus*))、弯曲杆菌属物种(*Campylobacter sp.*) (例如空肠弯曲杆菌(*Campylobacter jejuni*))、弗朗西斯菌属物种(*Francisella sp.*) (例如新凶手弗朗西斯菌(*Francisella novicida*))、蓝藻菌属物种(*Acaryochloris sp.*)、醋盐杆菌属物种(*Acetohalobium sp.*)、氨基酸球菌属物种(*Acidaminococcus sp.*)、嗜酸硫杆菌属物种(*Acidithiobacillus sp.*)、脂环酸芽孢杆菌属物种(*Alicyclobacillus sp.*)、闪杆菌属物种(*Allochromatium sp.*)、制氮菌属物种(*Ammonifex sp.*)、鱼腥藻属物种(*Anabaena sp.*)、节旋藻属物种(*Arthrospira sp.*)、芽孢杆菌属物种(*Bacillus sp.*)、伯克霍尔德氏菌属物种(*Burkholderiales sp.*)、*Caldicelulosiruptor*属物种、*Candidatus* 属物种、梭菌属物种(*Clostridium sp.*)、鳄球藻属物种(*Crocospaera sp.*)、蓝丝菌属物种(*Cyanothece sp.*)、微小杆菌属物种(*Exiguobacterium sp.*)、芬戈尔德菌属物种(*Fingoldia sp.*)、纤线杆菌属物种(*Ktedonobacter sp.*)、毛螺菌科物种(*Lachnospiraceae sp.*)、乳杆菌属物种(*Lactobacillus sp.*)、鞘丝藻属物种(*Lyngbya sp.*)、海杆菌属物种(*Marinobacter sp.*)、甲烷盐菌属物种(*Methanohalobium sp.*)、微颤菌属物种(*Microscilla sp.*)、微鞘藻属物种(*Microcoleus sp.*)、微囊藻属物种(*Microcystis sp.*)、盐碱厌氧菌属物种

(*Natronaerobius sp.*)、奈瑟氏菌属物种(*Neisseria sp.*)、亚硝化球菌属物种(*Nitrosococcus sp.*)、拟诺卡氏菌属物种(*Nocardiopsis sp.*)、节球藻属物种(*Nodularia sp.*)、念珠藻属物种(*Nostoc sp.*)、颤藻属物种(*Oscillatoria sp.*)、极单胞菌属物种(*Polaromonas sp.*)、暗色厌氧香肠状菌属物种(*Pelotomaculum sp.*)、假交替单胞菌属物种(*Pseudoalteromonas sp.*)、石袍菌属物种(*Petrotoga sp.*)、普雷沃菌属物种(*Prevotella sp.*)、葡萄球菌属物种(*Staphylococcus sp.*)、链霉菌属物种(*Streptomyces sp.*)、链孢囊菌属物种(*Streptosporangium sp.*)、聚球藻属物种(*Synechococcus sp.*)、栖热腔菌属物种(*Thermosiphon sp.*)或疣微菌门物种(*Verrucomicrobia sp.*)。在其他实施方案中,所述CRISPR核酸酶可以源自古细菌CRISPR系统、CRISPR/CasX系统或CRISPR/CasY系统(Burstein等人, *Nature*, 2017, 542(7640):237-241)。

[0052] 在一些实施方案中,所述CRISPR核酸酶可以源自II型CRISPR核酸酶。例如,所述II型CRISPR核酸酶可以是Cas9蛋白。合适的Cas9核酸酶包括酿脓链球菌Cas9 (SpCas9),新凶手弗朗西斯菌Cas9 (FnCas9),金黄色葡萄球菌(SaCas9),嗜热链球菌Cas9 (StCas9),巴氏链球菌(SpaCas9),空肠弯曲杆菌Cas9 (CjCas9),脑膜炎奈瑟氏菌Cas9 (NmCas9)或灰色奈瑟氏菌Cas9 (NcCas9)。在其他实施方案中,所述CRISPR核酸酶可以源自V型CRISPR核酸酶,诸如Cpf1核酸酶。合适的Cpf1核酸酶包括新凶手弗朗西斯菌Cpf1 (FnCpf1),氨基酸球菌属物种Cpf1 (AsCpf1)或毛螺菌科细菌ND2006 Cpf1 (LbCpf1)。在又一个实施方案中,所述CRISPR核酸酶可以源自VI型CRISPR核酸酶,例如瓦氏纤毛菌(*Leptotrichia wadei*) Cas13a (LwaCas13a)或沙希氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)。

[0053] 所述CRISPR核酸酶可以是野生型CRISPR核酸酶、修饰的CRISPR核酸酶或野生型或修饰的CRISPR核酸酶的片段。可以修饰所述CRISPR核酸酶以增加核酸结合亲和力和/或特异性,改变酶促活性和/或改变蛋白的另一种特性。例如,可以修饰、缺失或失活所述CRISPR核酸酶的核酸酶(即,DNA酶、RNA酶)结构域。所述CRISPR核酸酶可以被截短以除去对于核酸酶的功能不是必需的结构域。

[0054] CRISPR核酸酶包含两个核酸酶结构域。例如,Cas9核酸酶包含切割指导RNA互补链的HNH结构域和切割非互补链的RuvC结构域;Cpf1核酸酶包含RuvC结构域和NUC结构域;且Cas13a核酸酶包含两个HNEPN结构域。当两个核酸酶结构域都有功能时,CRISPR核酸酶引入双链断裂。任一核酸酶结构域可以通过一个或多个突变和/或缺失而失活,由此产生在双链序列的一条链中引入单链断裂的变体。例如,Cas9核酸酶的RuvC结构域中的一个或多个突变(例如,D10A、D8A、E762A和/或D986A)产生HNH切口酶,其使指导RNA互补链产生切口;且Cas9核酸酶的HNH结构域中的一个或多个突变(例如,H840A、H559A、N854A、N856A和/或N863A)产生RuvC切口酶,其使指导RNA非互补链产生切口。相当的突变可以将Cpf1和Cas13a核酸酶转化为切口酶。可以组合使用(经由一对偏移指导RNA)靶向至染色体序列的相对链的两种CRISPR切口酶来在染色体序列中产生双链断裂。双重CRISPR切口酶RNP可以增加靶标特异性并减少脱靶效应。

[0055] 额外结构域。所述CRISPR核酸酶可以进一步包含至少一个核定位序列(NLS)。NLS是一种氨基酸序列,其有助于将所述锌指核酸酶蛋白靶向至核中以在染色体中的靶标序列处引入双链断裂。核定位信号是本领域中已知的(参见,例如,Lange等人, *J. Biol. Chem.*, 2007, 282:5101-5105)。核定位信号的非限制性实例包括PKKKRKV (SEQ ID NO:

1)、PKKKRRV (SEQ ID NO:2)、KRPAATKKAGQAKKKK (SEQ ID NO:3)、YGRKKRRQRRR (SEQ ID NO:4)、RKKRRQRRR (SEQ ID NO:5)、PAAKRVKLD (SEQ ID NO:6)、RQRRNELKRSP (SEQ ID NO:7)、VSRKRPRP (SEQ ID NO:8)、PPKKARED (SEQ ID NO:9)、PQPKKKPL (SEQ ID NO:10)、SALIKKKKMAP (SEQ ID NO:11)、PKQKKRK (SEQ ID NO:12)、RKLKKKIKKL (SEQ ID NO:13)、REKKKFLKRR (SEQ ID NO:14)、KRKGDEV DGVDEVAKKSKK (SEQ ID NO:15)、RKCLQAGMNLEARKTKK (SEQ ID NO:16)、NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:17)和RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:18)。所述NLS可以位于所述CRISPR核酸酶的N-末端、C-末端或内部位置。

[0056] 在额外实施方案中,所述CRISPR核酸酶也可以包含至少一个细胞穿透结构域。合适的细胞穿透结构域的实例包括但不限于,GRKKRRQRRRPPQPKKKRKV (SEQ ID NO:19)、PLSSIFSRIGDPPKKKKRKV (SEQ ID NO:20)、GALFLGWLGAAGSTMGAPKKKKRKV (SEQ ID NO:21)、GALFLGFLGAAGSTMGAWSQPKKKKRKV (SEQ ID NO:22)、KETWWETWWTEWSQPKKKKRKV (SEQ ID NO:23)、YARAAARQARA (SEQ ID NO:24)、THRLPRRRRRR (SEQ ID NO:25)、GGRRARRRRRR (SEQ ID NO:26)、RRQRRTSKLMKR (SEQ ID NO:27)、GWTLSAGYLLGKINLKALAALAKKIL (SEQ ID NO:28)、KALAEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:29)和RQIKIWFQNRMMKWK (SEQ ID NO:30)。所述细胞穿透结构域可以位于所述CRISPR蛋白的N-末端、C-末端或内部位置。

[0057] 在还有其他实施方案中,所述CRISPR核酸酶可以进一步包含至少一个标记结构域。标记结构域的非限制性实例包括荧光蛋白、纯化标签和表位标签。在一个实施方案中,所述标记结构域可以是荧光蛋白。合适的荧光蛋白的非限制性实例包括绿色荧光蛋白(例如,GFP、GFP-2、tagGFP、turboGFP、EGFP、Emerald、Azami Green、单体Azami Green、CopGFP、AceGFP、ZsGreen1),黄色荧光蛋白(例如YFP、EYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1),蓝色荧光蛋白(例如EBFP、EBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire),青色荧光蛋白(例如ECFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan),红色荧光蛋白(mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-单体、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred)和橙色荧光蛋白(mOrange、mKO、Kusabira-Orange、单体Kusabira-Orange、mTangerine、tdTomato)或任何其他合适的荧光蛋白。在另一个实施方案中,所述标记结构域可以是纯化标签和/或表位标签。合适的标签包括但不限于多(His)标签、FLAG(或DDK)标签、Halo标签、AcV5标签、AU1标签、AU5标签、生物素羧基载体蛋白(BCCP)、钙调蛋白结合蛋白(CBP)、几丁质结合结构域(CBD)、E标签、E2标签、ECS标签、eXact标签、Glu-Glu标签、谷胱甘肽-S-转移酶(GST)、HA标签、HSV标签、KT3标签、麦芽糖结合蛋白(MBP)、MAP标签、Myc标签、NE标签、NusA标签、PDZ标签、S标签、S1标签、SBP标签、Softag 1标签、Softag 3标签、Spot标签、Strep标签、SUMO标签、T7标签、串联亲和纯化(TAP)标签、硫氧还蛋白(TRX)、V5标签、VSV-G标签和Xa标签。所述标记结构域可以位于所述CRISPR核酸酶的N-末端、C-末端或内部位置。

[0058] 所述至少一个核定位信号、至少一个细胞穿透结构域和/或至少一个标记结构域可以经由一个或多个化学键(例如,共价键)直接连接至所述CRISPR核酸酶。或者,所述至少一个核定位信号、至少一个细胞穿透结构域和/或至少一个标记结构域可以经由一个或多个接头间接连接至所述CRISPR核酸酶。合适的接头包括氨基酸,肽,核苷酸,核酸,有机接头分子(例如,马来酰亚胺衍生物,N-乙氧基苄基咪唑,联苯-3,4',5-三羧酸,对氨基苄氧基羰

基等),二硫化物接头,和聚合物接头(例如,PEG)。所述接头可以包括一个或多个间隔基团,包括但不限于亚烷基、亚烯基、亚炔基、烷基、烯基、炔基、烷氧基、芳基、杂芳基、芳烷基、芳烯基、芳炔基等。所述接头可以是中性的,或携带正电荷或负电荷。另外,所述接头可以是可切割的,使得连接所述接头与另一个化学基团的接头的共价键可以在某些条件(包括pH、温度、盐浓度、光、催化剂或酶)下被断裂或切割。在一些实施方案中,所述接头可以是肽接头。所述肽接头可以是柔性氨基酸接头或刚性氨基酸接头。合适接头的额外实例是本领域中众所周知的,并且设计接头的程序是本领域中容易的。

[0059] 指导RNA。CRISPR核酸酶通过指导RNA被引导至其靶标位点。所述指导RNA与所述靶标位点杂交并且与所述CRISPR核酸酶相互作用以将所述CRISPR核酸酶引导至所述染色体序列中的靶标位点。所述靶标位点没有序列限制,除了该序列以原间隔区相邻基序(PAM)为边界。来自不同细菌物种的CRISPR蛋白识别不同的PAM序列。例如,PAM序列包括5'-NGG (SpCas9, FnCas9)、5'-NGRRT (SaCas9)、5'-NNAGAAW (StCas9)、5'-NNNGATT (NmCas9)、5'-NNNNRYAC (CjCas9)和5'-TTTV (Cpf1),其中N被定义为任何核苷酸,R被定义为G或A,W被定义为A或T,Y被定义为C或T,且V被定义为A、C或G。Cas9 PAM位于靶标位点的3',且cpf1 PAM位于靶标位点的5'。

[0060] 指导RNA包含三个区域:在5'末端与靶标位点处的序列互补的第一区域,形成茎环结构的第二内部区域,和基本上保持单链的第三3'区域。每个指导RNA的第一区域是不同的,使得每个指导RNA将CRISPR核酸酶引导至特定靶标位点。每个指导RNA的第二和第三区域(也称为支架区域)在所有指导RNA中可以是相同的。

[0061] 所述指导RNA的第一区域与靶标位点处的序列(即,原间隔区序列)互补,使得所述指导RNA的第一区域可以与靶标位点处的序列碱基配对。所述指导RNA的第一区域(即,crRNA)和所述靶标序列之间的互补性可以是至少80%、至少85%、至少90%、至少95%或更高。通常,所述指导RNA的第一区域的序列和所述靶标位点处的序列之间没有错配(即,互补性是完全的)。在各个实施方案中,所述指导RNA的第一区域可以包含约10个核苷酸至多约25个核苷酸。例如,所述指导RNA的第一区域和所述染色体序列中的靶标位点之间的碱基配对的区域可以长度为约10、11、12、13、14、15、16、17、18、19、20、22、23、24、25或多于25个核苷酸。在示例性实施方案中,所述指导RNA的第一区域长度为约19、20或21个核苷酸。

[0062] 所述指导RNA还包含形成二级结构的第二区域。在一些实施方案中,所述二级结构包括茎(或发夹)和环。所述环和所述茎的长度可以变化。例如,所述环可以范围为约3至约10个核苷酸长度,且所述茎可以范围为约6至约20个碱基对长度。所述茎可以包含一个或多个1至约10个核苷酸的凸起。因此,所述第二区域的总长度可以范围为约16至约60个核苷酸长度。在一个示例性实施方案中,所述环为约4个核苷酸长度,且所述茎包含约12个碱基对。

[0063] 所述指导RNA还包含在3'末端基本上保持单链的第三区域。因此,所述第三区域与目标细胞中的任何染色体序列不具有互补性,并且与指导RNA的其余部分没有互补性。所述第三区域的长度可以变化。通常,所述第三区域大于约4个核苷酸长度。例如,所述第三区域的长度可以范围为约5至约60个核苷酸长度。

[0064] 所述指导RNA的第二和第三区域(或支架)的组合长度可以范围为约30至约120个核苷酸长度。在一个方面,所述指导RNA的第二和第三区域的组合长度范围为约70至约100个核苷酸长度。

[0065] 在一些实施方案中,所述指导RNA包含一个包含所有三个区域的分子。在其他实施方案中,所述指导RNA可以包含两个分开的分子。所述第一RNA分子可以包含所述指导RNA的第一(5')区域和所述指导RNA的第二区域的“茎”的一半。所述第二RNA分子可以包含所述指导RNA的第二区域的“茎”的另一半和所述指导RNA的第三区域。因此,在该实施方案中,所述第一和第二RNA分子各自含有彼此互补的核苷酸的序列。例如,在一个实施方案中,所述第一和第二RNA分子各自包含与另一序列碱基配对以形成功能性指导RNA的序列(约6至约20个核苷酸)。

[0066] (iii) 其他靶向核酸内切酶

在进一步实施方案中,所述靶向核酸内切酶可以是大范围核酸酶。大范围核酸酶是内切脱氧核糖核酸酶,其特征在于长识别序列,即所述识别序列通常范围为约12个碱基对至约40个碱基对。作为该要求的结果,所述识别序列通常仅在任何给定基因组中出现一次。在大范围核酸酶中,命名为LAGLIDADG的归巢核酸内切酶的家族已成为基因组和基因组工程改造的研究的一种有价值工具(参见,例如,Arnould等人, 2011, Protein Eng Des Sel, 24(1-2):27-31)。其他合适的大范围核酸酶包括I-CreI和I-Dm01。可以使用本领域技术人员众所周知的技术,通过修饰大范围核酸酶的识别序列来将所述大范围核酸酶靶向至特定染色体序列。

[0067] 在额外实施方案中,所述靶向核酸内切酶可以是转录激活因子样效应物(TALE)核酸酶。TALE是来自植物病原体黄单胞菌属的转录因子,其可以容易被工程改造以结合新的DNA靶标。TALE或其截短版本可以连接至核酸内切酶、诸如FokI的催化结构域以产生称为TALE核酸酶或TALEN的靶向核酸内切酶(Sanjana等人, 2012, Nat Protoc, 7(1):171-192)和Arnould等人, 2011, Protein Engineering, Design & Selection, 24(1-2):27-31)。

[0068] 在替代实施方案中,所述靶向核酸内切酶可以是嵌合核酸酶。嵌合核酸酶的非限制性实例包括ZF-大范围核酸酶、TAL-大范围核酸酶、Cas9-FokI融合体、ZF-Cas9融合体、TAL-Cas9融合体等。本领域技术人员熟悉用于生成此类嵌合核酸酶融合体的方式。

[0069] 在还有其他实施方案中,所述靶向核酸内切酶可以是位点特异性核酸内切酶。具体而言,所述位点特异性核酸内切酶可以是其识别序列很少出现在基因组中的“稀有切点(rare-cutter)”核酸内切酶。或者,所述位点特异性核酸内切酶可以被工程改造以切割目标位点(Friedhoff等人, 2007, Methods Mol Biol 352:1110123)。通常,所述位点特异性核酸内切酶的识别序列仅在基因组中出现一次。在替代的进一步实施方案中,所述靶向核酸内切酶可以是人工的靶向的DNA双链断裂诱导剂。

[0070] (b) 将靶向核酸内切酶递送至细胞

所述方法包括将所述靶向核酸内切酶引入目标亲本细胞系。可以将所述靶向核酸内切酶作为纯化的分离蛋白或作为编码所述靶向核酸内切酶的核酸引入细胞。所述核酸可以是DNA或RNA。在其中编码核酸是mRNA的实施方案中,mRNA可以是5'加帽的和/或3'聚腺苷酸化的。在其中编码核酸是DNA的实施方案中,DNA可以是线性或环状的。所述核酸可以是质粒或病毒载体的一部分,其中编码DNA可以可操作地连接至合适的启动子。本领域技术人员熟悉适当的载体、启动子、其他控制元件以及将载体引入目标细胞的方式。在其中靶向核酸内切酶是CRISPR核酸酶的实施方案中,可以将CRISPR核酸酶系统作为gRNA-蛋白复合物引入细

胞。

[0071] 可以通过各种方式将所述靶向核酸内切酶分子引入细胞中。合适的递送方式包括显微注射、电穿孔、声致穿孔、基因枪法、磷酸钙介导的转染、阳离子转染、脂质体转染、树状聚体转染、热休克转染、核转染、磁转染、脂质体转染、刺穿转染 (impalefection)、光转染、专有试剂增强的核酸摄取以及经由脂质体、免疫脂质体、病毒体或人工病毒粒子递送。在一个具体实施方案中,通过核转染将所述靶向核酸内切酶分子引入细胞。

[0072] 任选的供体多核苷酸。用于靶向的基因组修饰或工程改造的方法可以进一步包括将至少一种供体多核苷酸引入细胞,所述供体多核苷酸包含相对于靶标染色体序列具有至少一个核苷酸变化的序列。所述供体多核苷酸与所述染色体序列中的靶向位点处或附近的序列具有实质性序列同一性,使得可以通过同源性引导的修复过程来修复由所述靶向核酸内切酶引入的双链断裂,并且所述供体多核苷酸的序列可以被插入所述染色体序列或与所述染色体序列交换,由此修饰所述染色体序列。例如,所述供体多核苷酸可以包含与靶标位点的一侧上的序列具有实质性序列同一性的第一序列和与靶标位点的另一侧上的序列具有实质性序列同一性的第二序列。所述供体多核苷酸可以进一步包含用于整合至靶向的染色体序列中的供体序列。例如,所述供体序列可以是外源序列(例如,标记序列),使得所述外源序列的整合破坏阅读框并使靶向的染色体序列失活。

[0073] 所述供体多核苷酸中与所述染色体序列中的靶标位点处或附近的序列具有实质性序列同一性的第一和第二序列的长度可以并且将变化。通常,所述供体多核苷酸中的第一和第二序列各自为至少约10个核苷酸长度。在各个实施方案中,与染色体序列具有实质性序列同一性的供体多核苷酸序列可以是约15个核苷酸、约20个核苷酸、约25个核苷酸、约30个核苷酸、约40个核苷酸、约50个核苷酸、约100个核苷酸或超过100个核苷酸长度。

[0074] 短语“实质性序列同一性”意指多核苷酸中的序列与目标染色体序列具有至少约75%序列同一性。在一些实施方案中,所述多核苷酸中的序列与目标染色体序列具有约75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性。

[0075] 所述供体多核苷酸的长度可以且将变化。例如,所述供体多核苷酸可以范围为约20个核苷酸长度直至约200,000个核苷酸长度。在各个实施方案中,所述供体多核苷酸可以范围为约20个核苷酸至约100个核苷酸长度、约100个核苷酸至约1000个核苷酸长度、约1000个核苷酸至约10,000个核苷酸长度、约10,000个核苷酸至约100,000个核苷酸长度、或约100,000个核苷酸至约200,000个核苷酸长度。

[0076] 通常,所述供体多核苷酸可以是DNA。DNA可以是单链或双链的。DNA可以是线性或环状的。在一些实施方案中,所述供体多核苷酸可以是包含少于约200个核苷酸的单链、线性寡核苷酸。在其他实施方案中,所述供体多核苷酸可以是载体的一部分。合适的载体包括DNA质粒、病毒载体、细菌人工染色体(BAC)和酵母人工染色体(YAC)。在还有其他实施方案中,所述供体多核苷酸可以是与递送媒介物、诸如脂质体或泊洛沙姆复合的PCR片段或核酸。

[0077] 所述供体多核苷酸可以与所述靶向核酸内切酶分子同时引入细胞。或者,可以将所述供体多核苷酸和所述靶向核酸内切酶分子依次引入细胞。所述靶向核酸内切酶分子与所述供体多核苷酸的比率可以并且将变化。通常,靶向核酸内切酶分子与供体多核苷酸的

比率范围为约1:10至约10:1。在各个实施方案中,所述靶向核酸内切酶分子与多核苷酸的比率可以为约1:10、1:9、1:8、1:7、1:6、1:5、1:4、1:3、1:2、1:1、2:1、3:1、4:1、5:1、6:1、7:1、8:1、9:1或10:1。在一个实施方案中,该比率为约1:1。

[0078] (c) 培养细胞

所述方法进一步包括将所述细胞维持在适当的条件下,使得由所述靶向核酸内切酶引入的双链断裂可以通过(i)非同源的末端接合修复过程来修复,使得通过至少一个核苷酸的缺失、插入和/或取代来修饰所述染色体序列,或任选地,通过(ii)同源性引导的修复过程来修复,使得所述染色体序列与所述多核苷酸的序列交换,使得所述染色体序列被修饰。在其中将编码所述靶向核酸内切酶的核酸引入所述细胞的实施方案中,所述方法包括将细胞维持在适当条件下,使得所述细胞表达所述靶向核酸内切酶。

[0079] 通常,将所述细胞维持在适合于细胞生长和/或维持的条件下。合适的细胞培养条件是本领域中众所周知的,并且描述于例如Santiago等人,(2008) PNAS 105:5809-5814; Moehle等人,(2007) PNAS 104:3055-3060; Urnov等人,(2005) Nature 435:646-651; 和 Lombardo等人,(2007) Nat. Biotechnology 25:1298-1306。本领域技术人员理解用于培养细胞的方法是本领域中已知的,并且可以并且将根据细胞类型而变化。在所有情况下,可以使用常规优化来确定用于特定细胞类型的最佳技术。

[0080] 在所述方法的该步骤期间,所述靶向核酸内切酶在所述染色体序列中的靶向的切割位点处识别、结合并产生双链断裂,并且在所述双链断裂的修复期间,将至少一个核苷酸的缺失、插入和/或取代引入所述靶向的染色体序列。在具体实施方案中,所述靶向的染色体序列被失活。

[0081] 在证实目标染色体序列已被修饰后,可以将单细胞克隆分离并基因分型(经由DNA测序和/或蛋白分析)。包含一种修饰的染色体序列的细胞可以经历额外一轮或多轮的靶向的基因组修饰,以修饰额外染色体序列,由此产生双重敲除、三重敲除等。

[0082] (IV) 生产具有低残余HCP水平的重组蛋白

本公开的另一个方面涵盖用于生产具有降低的残余HCP水平的重组蛋白或降低在生物生产系统中生产的重组蛋白中的HCP污染水平的方法。合适的重组蛋白描述于部分(I)(c)中。所述方法包括在上面部分(I)中描述的任何工程改造的细胞系中表达目标重组蛋白,和纯化表达的重组蛋白。用于生产或制造重组蛋白的方式是本领域中众所周知的(参见,例如,“Biopharmaceutical Production Technology”, Subramanian (编), 2012, Wiley-VCH; ISBN: 978-3-527-33029-4)。

[0083] 所述重组蛋白可以经由以下方法进行纯化,所述方法包括澄清(例如,过滤)的步骤和一个或多个色谱(例如,亲和色谱、蛋白A(或G)色谱、离子交换(即,阳离子和/或阴离子)色谱)的步骤。本领域技术人员将理解,可以使用额外的纯化方法,包括但不限于大小排阻色谱,吸附色谱,疏水相互作用色谱,反相色谱,免疫亲和色谱,离心,超速离心,沉淀,免疫沉淀,提取,相分离等。通常,由于污染宿主细胞蛋白的水平较低,由本文公开的哺乳动物细胞系表达的重组蛋白的纯化可以涉及较少的纯化步骤。因此,与常规表达系统相比,可以减少纯化时间和成本。

[0084] 与由非工程改造的亲本细胞系产生的重组蛋白相比,由本文公开的工程改造的细胞系产生的重组蛋白具有降低的HCP水平。通常,由本文公开的细胞系产生的重组蛋白中的

残余HCP水平为小于100 ppm、小于30 ppm、小于10 ppm、小于3 ppm、小于1 ppm、小于0.3 ppm、小于0.1 ppm、小于0.03 ppm、小于0.01 ppm、小于0.003 ppm或小于0.001 ppm,如根据国际协调会议(ICG)指南使用经验证的方法所测量。合适的方法包括Western免疫印迹测定、ELISA酶测定、一维或二维SDS聚丙烯酰胺凝胶电泳(SDS-PAGE)、2D-差异凝胶内电泳(DIGE)、毛细管区带电泳-电喷雾电离-串联质谱(CZE-ESI-MS/MS)、液相色谱-串联质谱(LC-MS/MS)、二维液相色谱-串联质谱(2D-LC-MS/MS)等。

[0085] 定义

除非另外定义,否则本文使用的所有技术和科学术语都具有由本发明所属领域的技术人员通常理解的含义。以下参考文献为技术人员提供本发明中使用的许多术语的一般性定义:Singleton等人, Dictionary of Microbiology and Molecular Biology (第2版1994);The Cambridge Dictionary of Science and Technology (Walker编, 1988);The Glossary of Genetics,第5版, R. Rieger等人(编), Springer Verlag (1991);以及Hale和Marham, The Harper Collins Dictionary of Biology (1991)。如本文所用,以下术语具有归于它们的含义,除非另外指定。

[0086] 当介绍本公开或其优选实施方案的要素时,冠词“一个/种(a)”、“一个/种(an)”、“该”和“所述”意指存在所述要素中的一个/种或多个/种。术语“包含”、“包括”和“具有”意为包括性的且意指可能存在除所列要素以外的额外要素。

[0087] 如本文所用,术语“内源序列”是指对于细胞是天然的染色体序列。

[0088] 术语“外源序列”是指对于细胞不是天然的染色体序列,或被移至不同染色体位置的染色体序列。

[0089] “工程改造的”或“基因修饰的”细胞是指其中基因组已被修饰或工程改造的细胞,即,所述细胞至少含有已被工程改造以含有至少一个核苷酸的插入、至少一个核苷酸的缺失和/或至少一个核苷酸的取代的染色体序列。

[0090] 术语“基因组修饰”和“基因组编辑”是指由此改变特定染色体序列、使得所述染色体序列被修饰的过程。所述染色体序列可以被修饰为包含至少一个核苷酸的插入、至少一个核苷酸的缺失和/或至少一个核苷酸的取代。修饰的染色体序列被失活,使得没有制成产物。或者,可以修饰所述染色体序列,使得制成改变的产物。

[0091] 如本文所用的“基因”是指编码基因产物的DNA区域(包括外显子和内含子)以及调节基因产物的产生的所有DNA区域,无论此类调节序列是否邻近于编码序列和/或转录的序列。因此,基因包括但不必限于启动子序列、终止子、翻译调节序列(诸如核糖体结合位点和内部核糖体进入位点)、增强子、沉默子、绝缘子、边界元件、复制起点、基质衔接位点和基因座控制区域。

[0092] 术语“异源”是指实体对于目标细胞或物种不是天然的。

[0093] 术语“核酸”和“多核苷酸”是指呈线性或环状构型的脱氧核糖核苷酸或核糖核苷酸聚合物。对于本公开的目的,这些术语不应被解释为关于聚合物的长度进行限制。该术语可以涵盖天然核苷酸的已知类似物,以及在碱基、糖和/或磷酸酯部分中进行修饰的核苷酸。通常,特定核苷酸的类似物具有相同的碱基配对特异性;即A的类似物将与T碱基-配对。核酸或多核苷酸的核苷酸可以通过磷酸二酯、硫代磷酸酯、亚磷酰胺、二氨基磷酸酯键或其组合连接。

[0094] 术语“核苷酸”是指脱氧核糖核苷酸或核糖核苷酸。所述核苷酸可以是标准核苷酸(即,腺苷、鸟苷、胞苷、胸苷和尿苷)或核苷酸类似物。核苷酸类似物是指具有修饰的嘌呤或嘧啶碱基或修饰的核糖部分的核苷酸。核苷酸类似物可以是天然存在的核苷酸(例如,肌苷)或非天然存在的核苷酸。核苷酸的糖或碱基部分上的修饰的非限制性实例包括添加(或删除)乙酰基、氨基、羧基、羧甲基、羟基、甲基、磷酰基和硫醇基团,以及用其他原子取代碱基的碳和氮原子(例如7-脱氮嘌呤)。核苷酸类似物也包括双脱氧核苷酸、2'-O-甲基核苷酸、锁定核酸(LNA)、肽核酸(PNA)和吗啉代物。

[0095] 术语“多肽”和“蛋白”可互换使用,是指氨基酸残基的聚合物。

[0096] 术语“有问题的宿主细胞蛋白”是指(i)高度丰富、(ii)在下游处理期间难以除去和/或(iii)影响产物质量的宿主细胞蛋白。

[0097] 如本文所用,术语“靶标位点”或“靶标序列”是指这样的核酸序列,其限定染色体序列的待修饰或编辑的部分,且靶向核酸内切酶被工程改造来对其进行识别和结合(条件是存在足够的结合条件)。

[0098] 术语“上游”和“下游”是指核酸序列中相对于固定位置的定位。上游是指在所述位置的5'(即靠近链的5'末端)的区域,且下游是指在所述位置的3'(即靠近链的3'末端)的区域。

[0099] 用于测定核酸和氨基酸序列同一性的技术是本领域中已知的。通常,此类技术包括测定基因的mRNA的核苷酸序列和/或测定由其编码的氨基酸序列,以及将这些序列与第二核苷酸或氨基酸序列进行比较。也可以该方式测定和比较基因组序列。通常,同一性是指两个多核苷酸或多肽序列的分别核苷酸对核苷酸或氨基酸对氨基酸的精确对应关系。两个或更多个序列(多核苷酸或氨基酸)可以通过确定它们的百分比同一性来进行比较。两个序列(无论是核酸还是氨基酸序列)的百分比同一性是两个比对序列之间的精确匹配的数目除以较短序列的长度且乘以100。核酸序列的近似比对由Smith和Waterman, *Advances in Applied Mathematics* 2:482-489 (1981)的局部同源性算法提供。通过使用由Dayhoff, *Atlas of Protein Sequences and Structure*, M. O. Dayhoff编, 5增刊. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA开发,且由Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986)标准化的评分矩阵,该算法可应用于氨基酸序列。用以确定序列的百分比同一性的该算法的示例性执行程序由Genetics Computer Group (Madison, Wis.)在“BestFit”效用应用中提供。用于计算序列之间的百分比同一性或相似性的其他合适的程序通常是本领域中已知的,例如另一比对程序是以默认参数使用的BLAST。例如,BLASTN和BLASTP可以使用以下默认参数来进行使用:遗传密码=标准;过滤器=无;链=两条;截止值=60;预期值=10;矩阵=BLOSUM62;描述=50个序列;排序依据=高评分;数据库=非冗余,GenBank+EMBL+DDBJ+PDB+GenBank CDS翻译+Swiss蛋白+Spupdate+PIR。这些程序的细节可见于GenBank网站上。关于本文所述的序列,序列同一性的期望程度范围是近似80%至100%以及其之间的任何整数值。通常,序列之间的百分比同一性是至少70-75%、优选80-82%、更优选85-90%、甚至更优选92%、仍更优选95%且最优选98%序列同一性。

[0100] 因为可以在不脱离本发明的范围的情况下在上述细胞和方法中做出各种改变,所以意欲在以上描述中和在以下给出的实施例中含有的所有事项都应解释为说明性的而非

在限制性意义上进行解释。

## 实施例

[0101] 以下实施例举例说明本发明的某些方面。

### [0102] 实施例1:污染宿主细胞蛋白的鉴定

质谱法用于鉴定由几种CHO亲本细胞系产生的HCP。从不同的亲本细胞系收集分批进料的上清液,并通过LC-MS/MS进行分析。类似地,分析蛋白A捕获步骤后的样品洗脱液,以鉴定已经与柱缔合的蛋白。在第二种方法中,通过下游纯化步骤追踪来自重组表达克隆的HCP概况。“有问题的HCP”被表征为宿主细胞蛋白,其(i)高度丰富,(ii)在下游处理期间难以除去,和/或(iii)影响产品质量。考虑到每份样品中鉴定的大量蛋白,进行主组分分析(PCA)以强调变异和挖掘数据模式。表1列出鉴定的一些鉴定的“有问题的”HCP及其特征。

宿主细胞蛋白	丰度	难以纯化	影响产品质量
组织蛋白酶 D	+	+	+
羧肽酶 D	+		+
脂蛋白脂肪酶	+	+	+
磷脂酶 B-样 2	+	+	+
丝氨酸蛋白酶	+	+	+
巢蛋白 1		+	
簇蛋白	+	+	
金属蛋白酶抑制剂 1			+
硫氧还蛋白和硫氧还蛋白还原酶			+
过氧蛋白酶(Peroxidasin)		+	+

[0103] 几种蛋白酶被鉴定为基因编辑的候选物。源自宿主细胞的蛋白酶在细胞培养基中有活性,并且可以影响产品质量。它们的蛋白水解活性可以降解重组表达的多肽,也称为“剪切”,由此产生潜在免疫原性和改变的、例如非功能性或功能性较低的治疗性蛋白。鉴定的HCP被进一步归类为对于宿主细胞生长和生产力必要或非必要的。

### [0104] 实施例2:使用锌指核酸酶的脂蛋白脂肪酶和磷脂酶B-样2基因敲除

用编码靶向至脂蛋白脂肪酶(LPL)或磷脂酶B-样2(PLBL2)基因的一对锌指核酸酶(ZFN)的核酸转染CHO细胞。下面呈现各自的靶标位点(ZFN结合位点以大写显示,且切割位点以小写显示):

#### 脂蛋白脂肪酶

CCTGACTCCAACGTCATTgtggtGGACTGGCTGTATCGGGC (对13, 14) (SEQ ID NO:31)

GGCTGTATCGGGCCCAGCaacactATCCAGTGTCCGGCTGGCT (对15, 16) (SEQ ID NO:32)

#### 磷脂酶B-样2

GGCCTATGCAGCTGGtggtGGAGGCTTCTGTGTCTGAG (SEQ ID NO:33)。

[0105] 在转染后期望的孵育时段之后,收获细胞并分离基因组DNA。ZFN-诱导的切割使用Ce1-1核酸酶测定验证,所述Ce1-1核酸酶测定检测靶向的基因座的等位基因,其由于ZFN-诱导的DNA双链断裂的非同源末端接合(NHEJ)-介导的不完全修复而不同于野生型。ZFN的LLP 13/14对和PLBL2对生成在模拟物处理的细胞中不存在的切割片段(图1),表明将插入/缺失(indel)引入靶向的基因中。

[0106] *实施例3:使用Cas9 RNP的组织蛋白酶B和组织蛋白酶D基因敲除*

用Cas9构建体转染CHO细胞,所述Cas9构建体包含设计为靶向组织蛋白酶B或组织蛋白酶D的基因特异性gRNA。下面呈现gRNA的原间隔区序列。

[0107] 组织蛋白酶B

CCCCTGTTCGGATGACCTGATT (SEQ ID NO:34)

CCACTGTTCGGATGACCTGATTA (SEQ ID NO:35)

CAACAAACGGAATACGACATGG (SEQ ID NO:36)

组织蛋白酶D

CCCCCTGCGCAAGTTCACGTCT (SEQ ID NO:37)

CAAGTTCACGTCTATCCGTCGG (SEQ ID NO:38)

CCTTAAGGGTCCCATAACCACG (SEQ ID NO:39)

[0108] 在转染后第7天和第15天,收获细胞,分离基因组DNA,并进行Ce1-1核酸酶测定。在Cas9 RNP处理的细胞中检测到切割片段(图2)。

[0109] 分离单细胞敲除克隆。比较组织蛋白酶B敲除亚克隆(图3A)、组织蛋白酶D敲除亚克隆和野生型细胞(图3B)间的生产力和生长概况。尽管敲除亚克隆表现出一定的变异性,但敲除克隆和野生型细胞间的滴度和活细胞密度是相似的。

[0110] *实施例4:使用Cas9 RNP的簇蛋白基因敲除*

用Cas9构建体转染细胞,所述Cas9构建体包含设计为靶向簇蛋白的基因特异性gRNA。下面呈现gRNA的原间隔区序列。

GTCTCCGACAATGAGCTCAAGG (SEQ ID NO:40)

CCACTCAAGGGAGTAGGTACAT (SEQ ID NO:41)

GGGAGTAGGTACATTAATAAGG (SEQ ID NO:42)

[0111] 收获细胞,分离基因组DNA,并进行Ce1-1核酸酶测定。在Cas9 RNP处理的细胞中检测到切割片段(图4,泳道5-7)。分离野生型和簇蛋白敲除克隆。在野生型亚克隆(图5A)和簇蛋白敲除亚克隆(图5B)间比较生产力和生长概况。尽管在野生型和敲除亚克隆间存在变异性,但滴度和细胞密度在野生型和敲除细胞间是相似的。

[0112] 在野生型和簇蛋白敲除克隆间比较产品质量。在野生型和簇蛋白敲除克隆中表达模型融合蛋白。使用UPLC SEC分析蛋白产物的大小异质性,并表征为具有非常高的分子量(例如,融合蛋白的二聚体或聚集体,其可以导致额外的下游纯化步骤)、融合蛋白单体和低分子量物质。结果呈现于表1中。通常,簇蛋白敲除克隆具有与野生型克隆相似的概况。

表1. 簇蛋白敲除克隆相比于野生型克隆的产物质量			
亚克隆	聚集(%)	单体(%)	低分子量(%)
WT A01A	11.93	51.58	36.50
WT A01B	14.35	46.26	39.39
WT A05A	22.16	61.99	15.85
WT A05B	22.47	62.22	15.32
WT A06A	12.77	73.27	13.96
WT A06B	12.32	71.85	15.83
WT A09A	8.22	74.44	17.34
WT A09B	9.39	76.13	14.48
WT A10A	17.14	54.39	28.46
WT A10B	17.40	53.81	28.79
WT B10A	16.61	72.52	10.88
WT B10B	16.46	75.54	8.00
WT C03A	10.53	49.89	39.58
WT C03B	11.14	56.77	32.09
WT D07A	18.66	57.76	23.59
WT D07B	20.14	56.71	23.16
WT E01A	11.59	62.03	26.39
WT E01B	13.04	65.60	21.37
CL B01A	13.60	79.19	7.20
CL B01B	13.18	79.61	70.21
CL B07A	22.35	69.61	8.02

CL B07B	22.64	70.69	6.68
CL C02A	18.92	74.43	6.66
CL C02B	19.67	73.89	6.44
CL C07A	8.87	78.54	11.59
CL C07B	11.07	77.40	11.52
CL C09A	6.20	78.44	15.35
CL C09B	5.53	79.15	15.32
CL C10A	5.49	81.99	12.51
CL C10B	5.09	80.20	14.71
CL C11A	15.09	80.20	4.71
CL C11B	13.95	80.60	5.45
CL C12A	9.97	17.14	72.88
CL C12B	8.06	15.26	76.68
CL E02A	11.88	71.16	16.96
CL E02B	10.64	73.06	16.31

[0113] 实施例5:使用Cas9 RNP的硫氧还蛋白和硫氧还蛋白还原酶基因敲除

用Cas9构建体转染细胞,所述Cas9构建体包含设计为靶向硫氧还蛋白或硫氧还蛋白还原酶的基因特异性gRNA。下面呈现gRNA的原间隔区序列。

[0114] 硫氧还蛋白

GAAGCTTTTCAGGAGGCCCTGG (SEQ ID NO:43)

GCTTTTCAGGAGGCCCTGGCGG (SEQ ID NO:44)

CCCTGGCGGCTGCGGGAGACAA (SEQ ID NO:45)

CCACATGGTGTGGACCATGCAA (SEQ ID NO:46)

硫氧还蛋白还原酶

AACTCCTCTCGGAAGTAGATGG (SEQ ID NO:47)

GGAGGAACGTGTGTGAACGTGG (SEQ ID NO:48)

CCAGGCGGCTTTGTTAGGACAA (SEQ ID NO:49)

CCTTGAAAGACTCTCGAAACTA (SEQ ID NO:50)

[0115] 在合适的孵育时段之后,收获细胞,分离基因组DNA,并进行Ce1-1核酸酶测定。在Cas9 RNP处理的细胞中检测到切割片段(图6,泳道2-5和7-10)。

## 序列表

- <110> Sigma-Aldrich Co. LLC  
 Mascarenhas, Joaquina  
 Borgschulte, Trissa  
 Kayser, Kevin
- <120> 具有修饰的宿主细胞蛋白概况的工程改造的细胞
- <130> 047497-623224
- <150> US 62/667,194  
 <151> 2018-05-04
- <160> 50
- <170> PatentIn version 3.5
- <210> 1  
 <211> 7  
 <212> PRT  
 <213> 人工序列
- <220>  
 <223> 合成的
- <400> 1
- [0001] Pro Lys Lys Lys Arg Lys Val  
 1 5
- <210> 2  
 <211> 7  
 <212> PRT  
 <213> 人工序列
- <220>  
 <223> 合成的
- <400> 2
- Pro Lys Lys Lys Arg Arg Val  
 1 5
- <210> 3  
 <211> 16  
 <212> PRT  
 <213> 人工序列
- <220>  
 <223> 合成的
- <400> 3



Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro  
1 5 10

<210> 8  
<211> 8  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 8

Val Ser Arg Lys Arg Pro Arg Pro  
1 5

<210> 9  
<211> 8  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 9

[0003]

Pro Pro Lys Lys Ala Arg Glu Asp  
1 5

<210> 10  
<211> 8  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 10

Pro Gln Pro Lys Lys Lys Pro Leu  
1 5

<210> 11  
<211> 12  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 11

Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro  
1 5 10

<210> 12  
<211> 7  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 12

Pro Lys Gln Lys Lys Arg Lys  
1 5

<210> 13  
<211> 10  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 13

[0004]

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu  
1 5 10

<210> 14  
<211> 10  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 14

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg  
1 5 10

<210> 15  
<211> 20  
<212> PRT  
<213> 人工序列

<220>  
<223> 合成的

<400> 15



<400> 18

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu  
1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys  
20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val  
35 40

<210> 19

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 19

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys  
1 5 10 15

[0006]

Lys Arg Lys Val  
20

<210> 20

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 20

Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys  
1 5 10 15

Arg Lys Val

<210> 21

<211> 24

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 21

Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly  
1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val  
20

<210> 22

<211> 27

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 22

Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly  
1 5 10 15

[0007]

Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val  
20 25

<210> 23

<211> 21

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 23

Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys  
1 5 10 15

Lys Lys Arg Lys Val  
20

<210> 24

<211> 11

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 24

Tyr Ala Arg Ala Ala Ala Arg Gln Ala Arg Ala  
1 5 10

<210> 25

<211> 11

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 25

Thr His Arg Leu Pro Arg Arg Arg Arg Arg Arg  
1 5 10

<210> 26

<211> 11

<212> PRT

<213> 人工序列

<220>

<223> 合成的

[0008]

<400> 26

Gly Gly Arg Arg Ala Arg Arg Arg Arg Arg Arg  
1 5 10

<210> 27

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 27

Arg Arg Gln Arg Arg Thr Ser Lys Leu Met Lys Arg  
1 5 10

<210> 28

<211> 27

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 28

Gly Trp Thr Leu Asn Ser Ala Gly Tyr Leu Leu Gly Lys Ile Asn Leu  
1 5 10 15

Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu  
20 25

<210> 29

<211> 33

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 29

Lys Ala Leu Ala Trp Glu Ala Lys Leu Ala Lys Ala Leu Ala Lys Ala  
1 5 10 15

Leu Ala Lys His Leu Ala Lys Ala Leu Ala Lys Ala Leu Lys Cys Glu  
20 25 30

[0009]

Ala

<210> 30

<211> 16

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 30

Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys  
1 5 10 15

<210> 31

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 31

cctgactcca acgtcattgt ggtggactgg ctgtatcggg c

41

	<210> 32	
	<211> 42	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 32	
	ggctgtatcg ggcccagcaa cactatccag tgtcggctgg ct	42
	<210> 33	
	<211> 40	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 33	
	ggcctatgca gctggtgtgg tggaggcttc tgtgtctgag	40
[0010]	<210> 34	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 34	
	cccactgtcg gatgacctga tt	22
	<210> 35	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 35	
	ccactgtcgg atgacctgat ta	22
	<210> 36	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	

	<223> 合成的	
	<400> 36	
	caacaaacgg aatacgacat gg	22
	<210> 37	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 37	
	ccccctgcgc aagttcacgt ct	22
	<210> 38	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 38	
[0011]	caagttcacg tctatccgtc gg	22
	<210> 39	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 39	
	ccttaagggt cccataacca cg	22
	<210> 40	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 40	
	gtctccgaca atgagctcaa gg	22
	<210> 41	
	<211> 22	

	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的		
	<400> 41		
	ccactcaagg gagtaggtac at		22
	<210> 42		
	<211> 22		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的		
	<400> 42		
	gggagtaggt acattaataa gg		22
	<210> 43		
	<211> 22		
	<212> DNA		
	<213> 人工序列		
[0012]	<220>		
	<223> 合成的		
	<400> 43		
	gaagcttttc aggaggccct gg		22
	<210> 44		
	<211> 22		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的		
	<400> 44		
	gcttttcagg aggccctggc gg		22
	<210> 45		
	<211> 22		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 合成的		
	<400> 45		
	ccctggcggc tgcgggagac aa		22

	<210> 46	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 46	
	ccacatggtg tggacatgc aa	22
	<210> 47	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 47	
	aactcctctc ggaactagat gg	22
[0013]	<210> 48	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 48	
	ggaggaacgt gtgtgaacgt gg	22
	<210> 49	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
	<400> 49	
	ccaggcgget ttgttaggac aa	22
	<210> 50	
	<211> 22	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成的	
[0014]	<400> 50	
	ccttgaaaga ctctcgaaac ta	22

第7天 Cell1

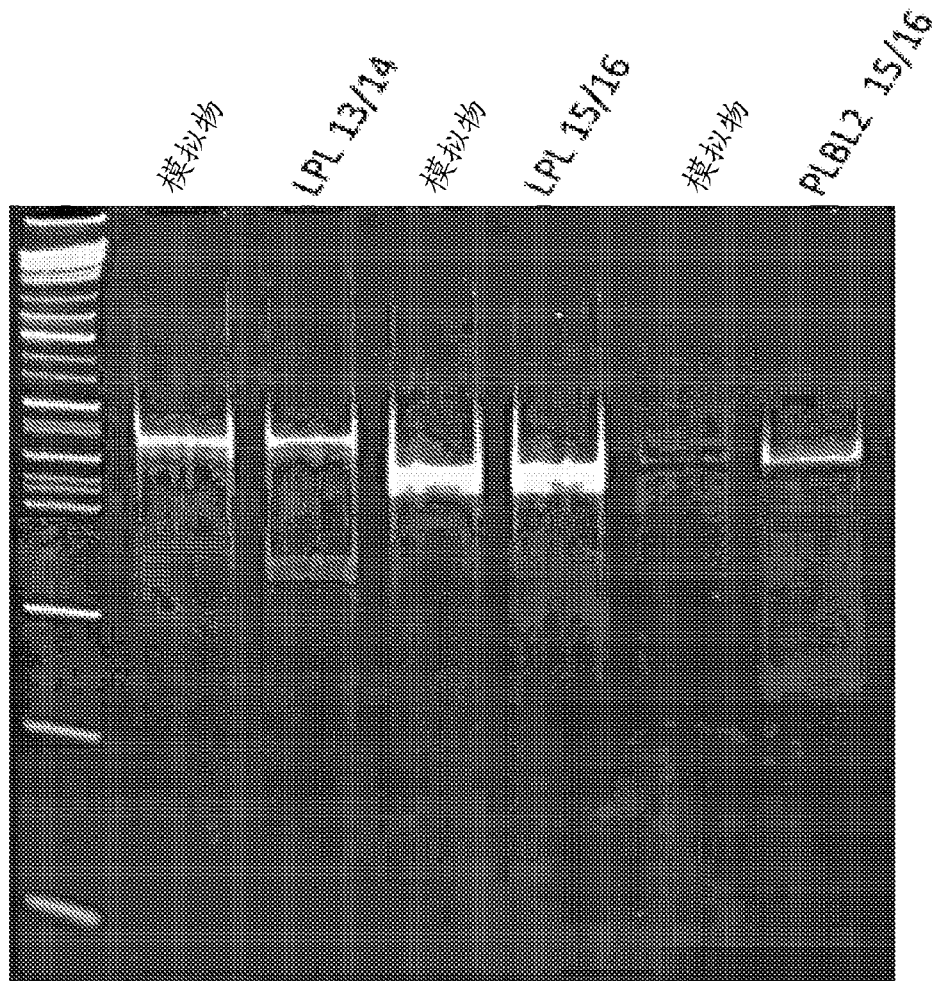


图 1

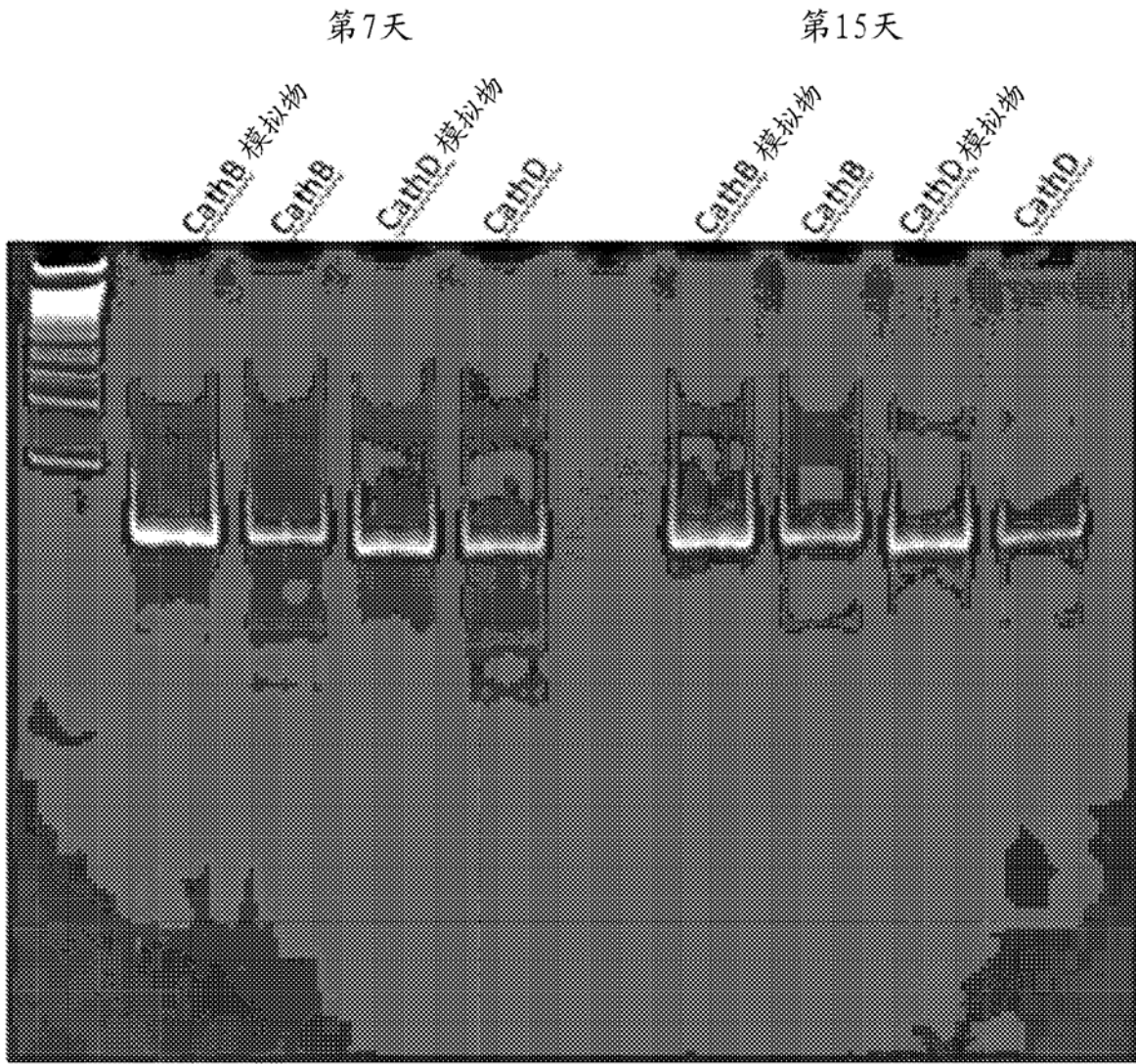


图 2

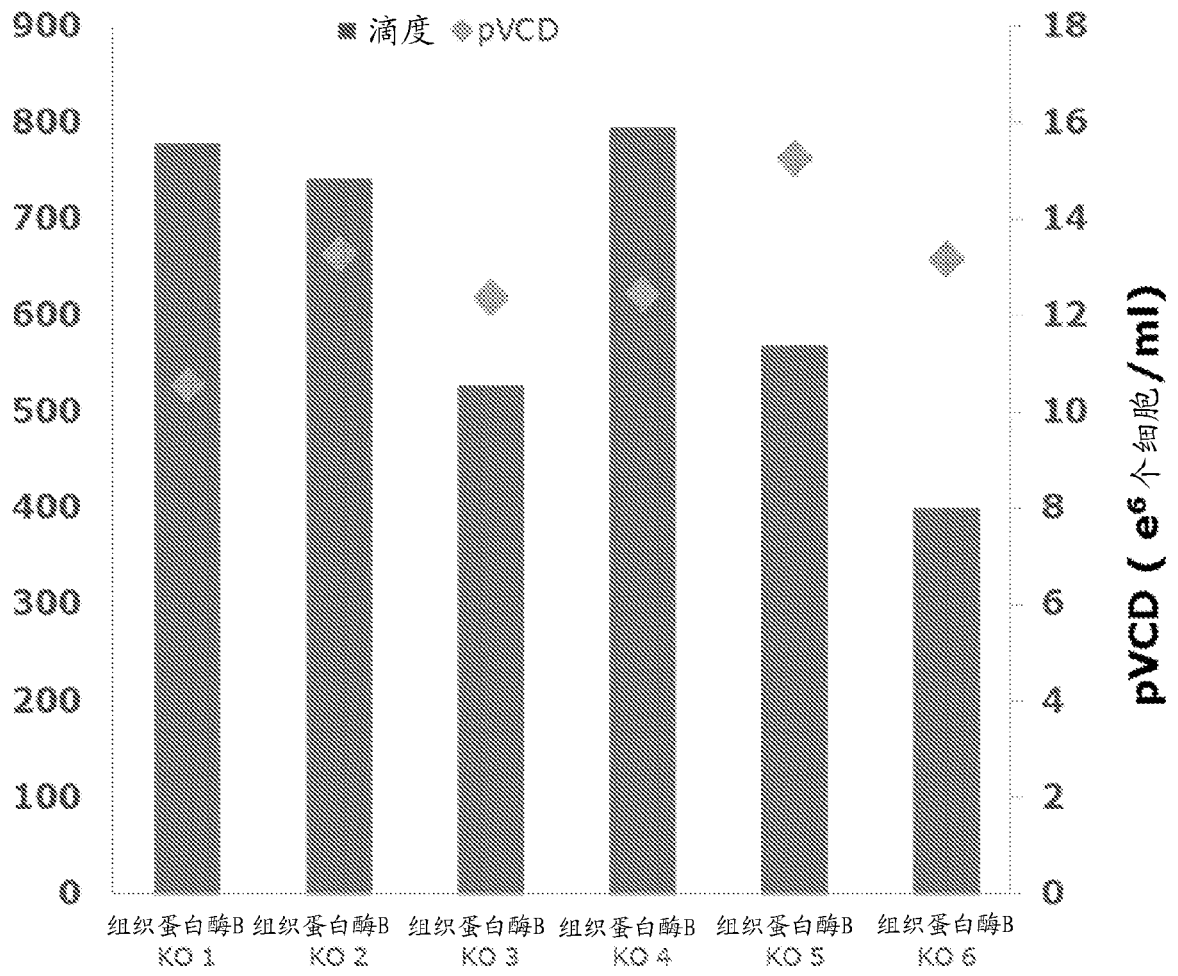


图 3A

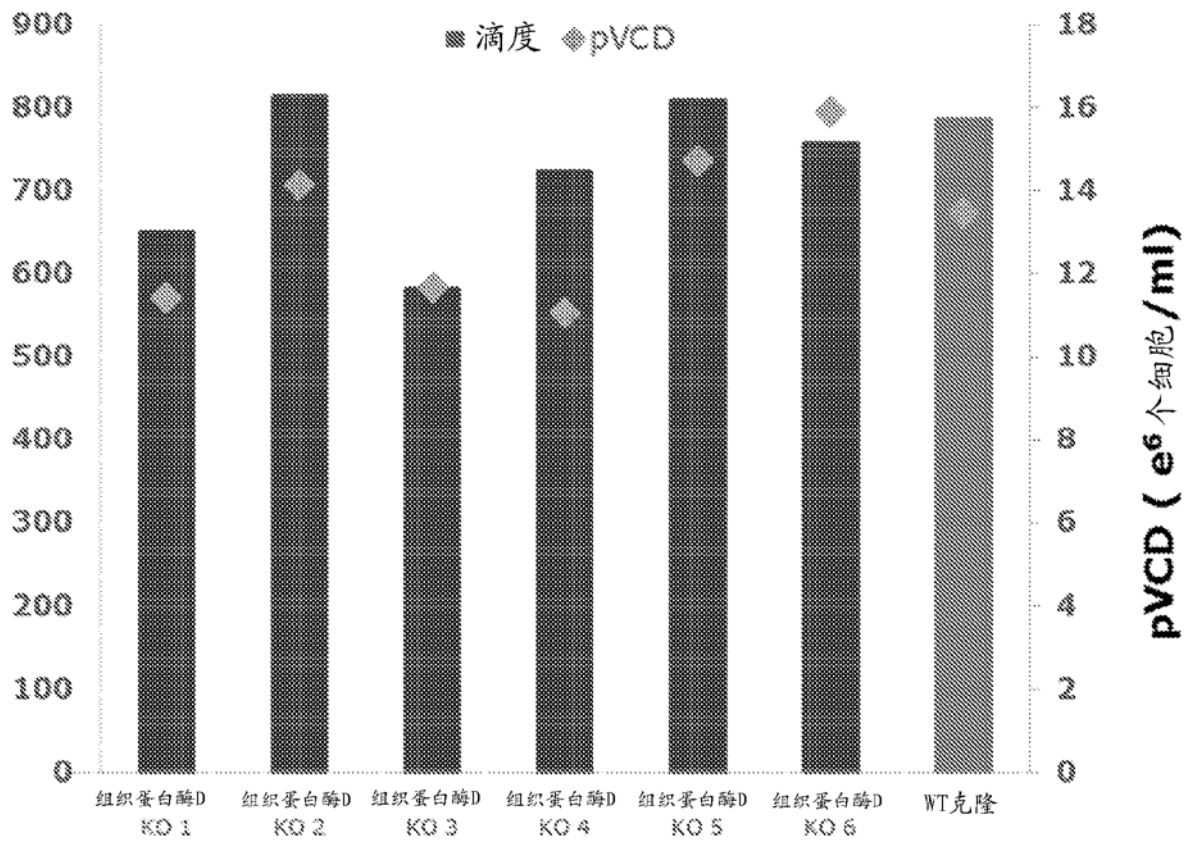


图 3B

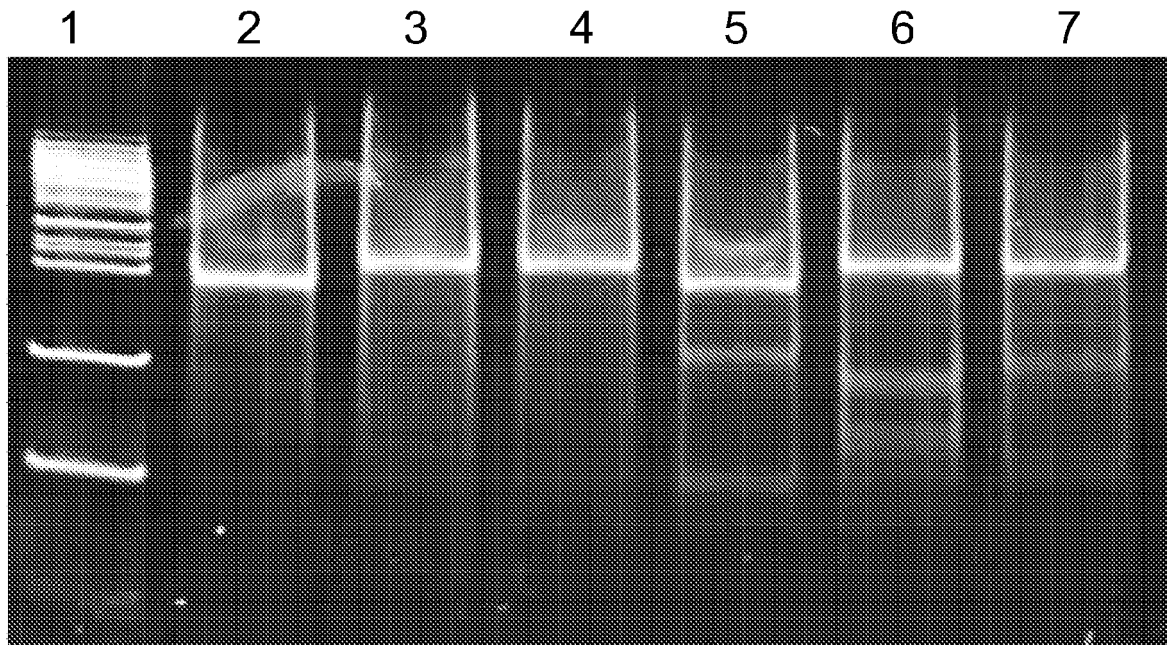


图 4

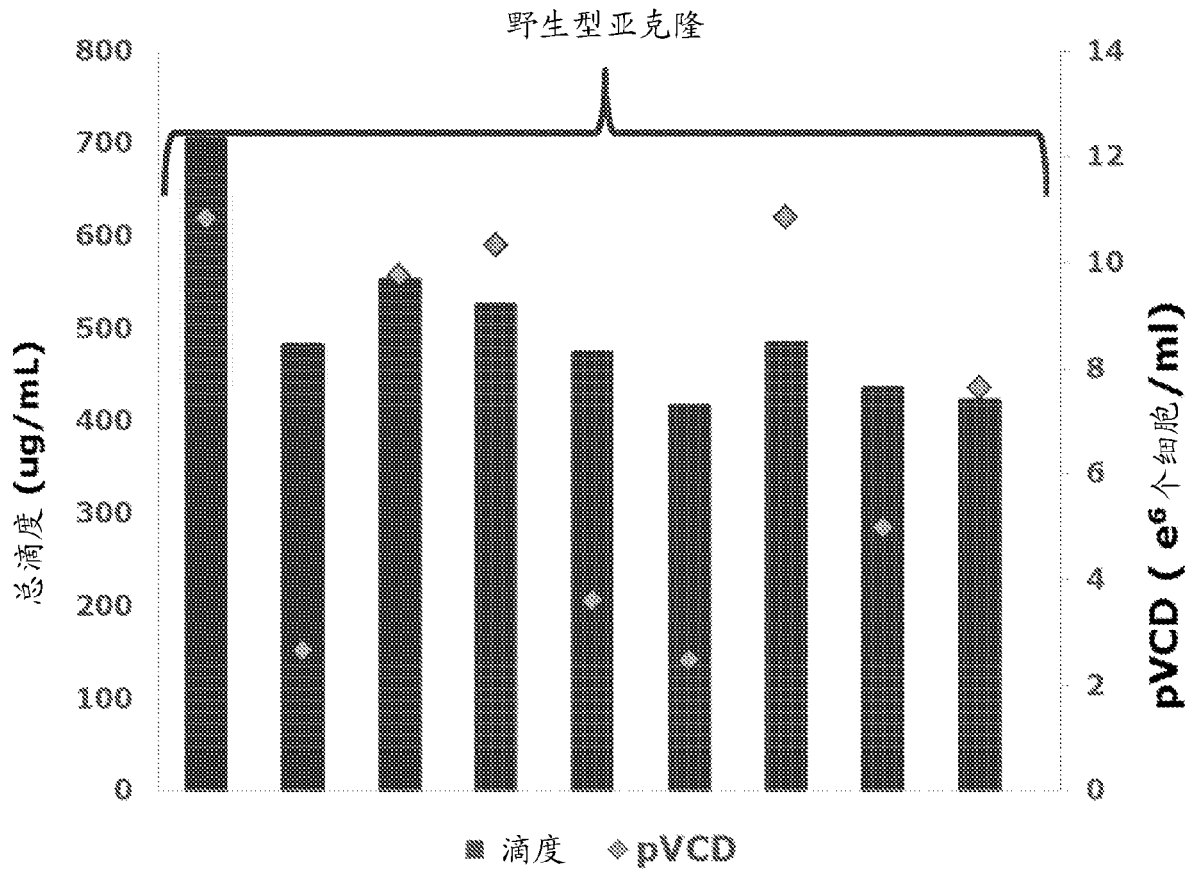


图 5A

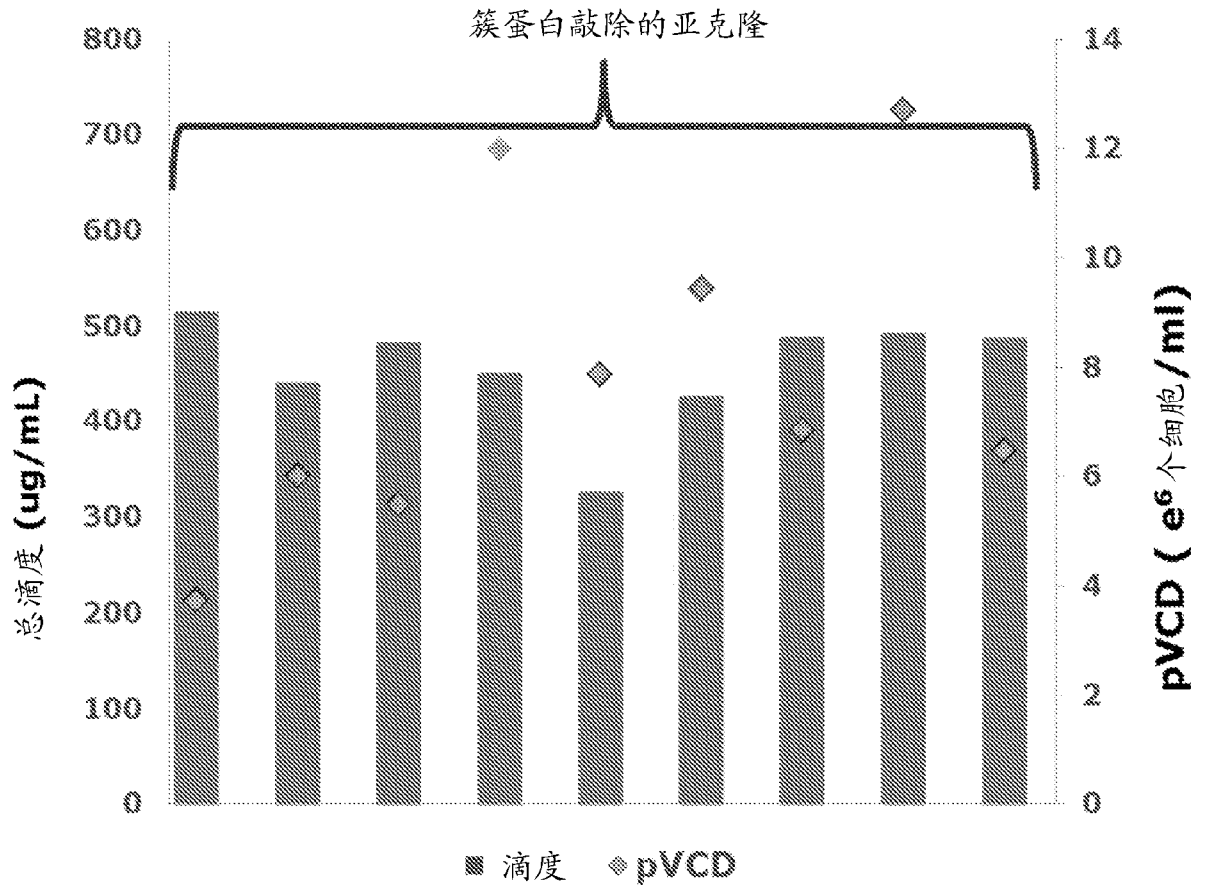


图 5B

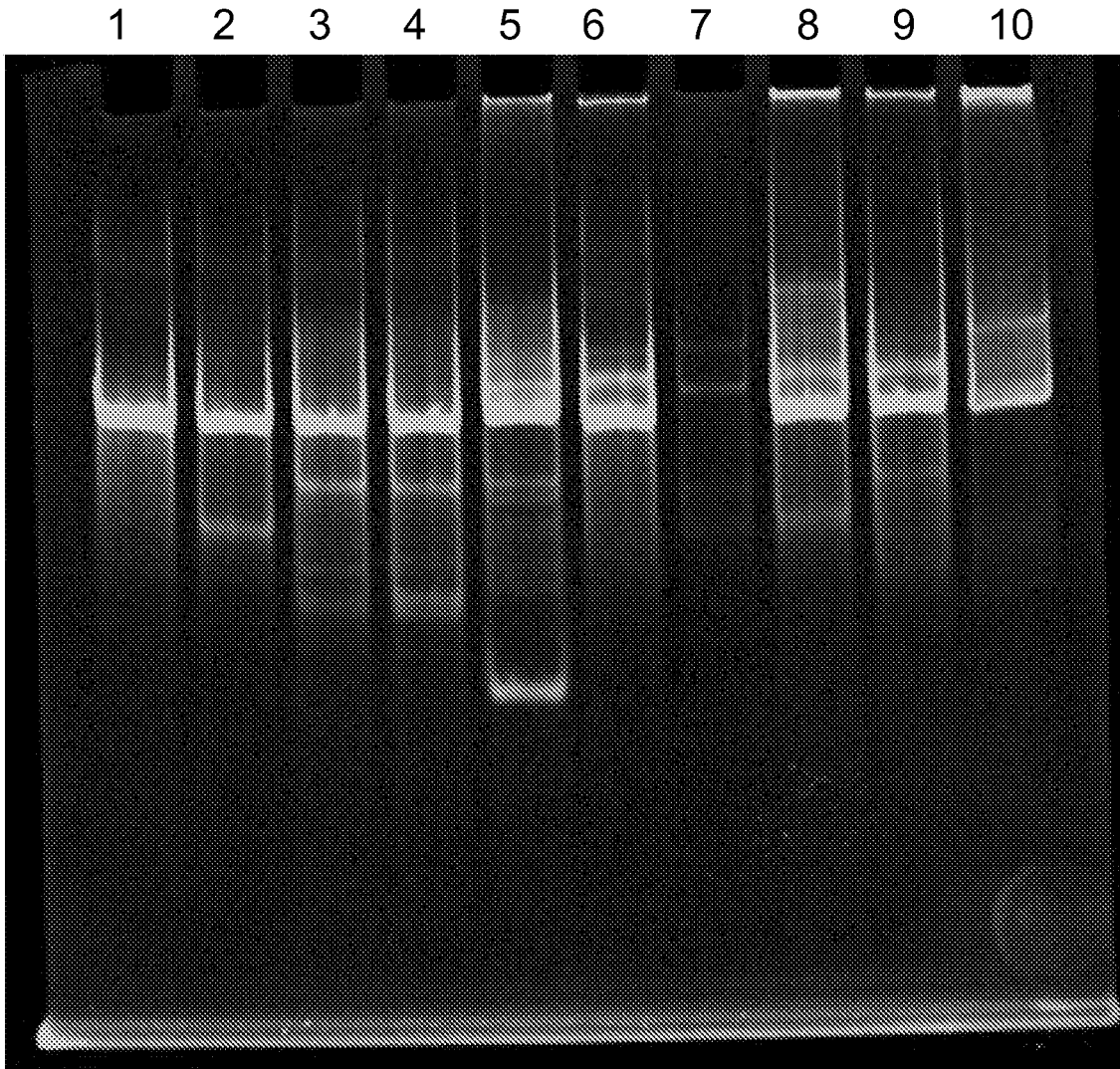


图 6