



(12) 发明专利

(10) 授权公告号 CN 105518137 B

(45) 授权公告日 2021.04.30

(21) 申请号 201580000469.8

(22) 申请日 2015.06.11

(65) 同一申请的已公布的文献号
申请公布号 CN 105518137 A

(43) 申请公布日 2016.04.20

(85) PCT国际申请进入国家阶段日
2015.09.29

(86) PCT国际申请的申请数据
PCT/CN2015/081227 2015.06.11

(87) PCT国际申请的公布数据
W02016/197355 ZH 2016.12.15

(73) 专利权人 深圳市第二人民医院
地址 518037 广东省深圳市福田区笋岗西
路3002号

(72) 发明人 蔡志明 牟丽莎 刘璐 谢崇伟
陈鹏飞 张军方 高汉超 陆赢

(74) 专利代理机构 深圳鼎合诚知识产权代理有
限公司 44281
代理人 孙银行 彭愿洁

(51) Int.Cl.
C12N 15/113 (2010.01)
C12N 15/867 (2006.01)

(56) 对比文件
CN 101679950 A, 2010.03.24
CN 104651399 A, 2015.05.27
CN 103820454 A, 2014.05.28

审查员 严金波

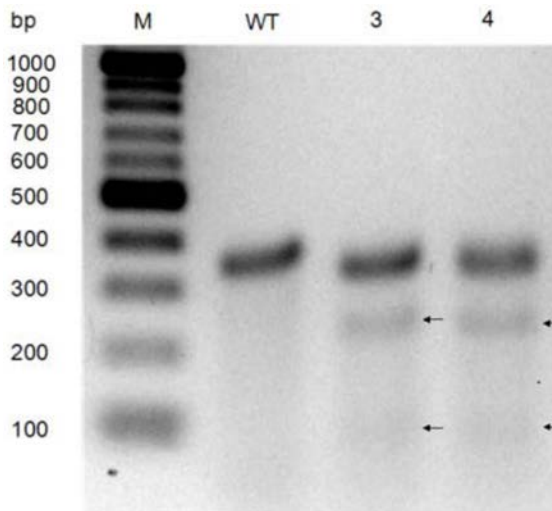
权利要求书2页 说明书11页
序列表15页 附图3页

(54) 发明名称

CRISPR-Cas9特异性敲除猪SALL1基因的方法及用于特异性靶向SALL1基因的sgRNA

(57) 摘要

本发明公开了一种运用CRISPR-Cas9特异性敲除猪SALL1基因的方法及用于特异性靶向SALL1基因的sgRNA。本发明的特异性靶向SALL1基因的sgRNA在SALL1基因上的靶序列符合5'-N(20)NGG-3'的序列排列规则,其中N(20)表示20个连续的碱基,其中每个N表示A或T或C或G;在SALL1基因上的靶序列位于SALL1基因的N端的3个外显子编码区或与相邻内含子的交界处;在SALL1基因上的靶序列是唯一的。本发明的sgRNA用于CRISPR-Cas9特异性敲除猪SALL1基因的方法中,能够快速、精确、高效、特异性地敲除猪SALL1基因,有效地解决构建SALL1基因敲除猪周期长和成本高的问题。



1. 在运用CRISPR-Cas9特异性敲除猪SALL1基因中用于特异性靶向SALL1基因的sgRNA, 其特征在于:

(1) 所述sgRNA在SALL1基因上的靶序列符合5' -N(20) NGG-3' 的序列排列规则, 其中N(20)表示20个连续的碱基, 其中每个N表示A或T或C或G, 符合所述规则的靶序列位于正义链或反义链;

(2) 所述sgRNA在SALL1基因上的靶序列位于SALL1基因的N端的3个外显子编码区, 或靶序列的一部分位于SALL1基因的N端的3个外显子, 其余部分跨越与相邻内含子的交界, 位于相邻内含子;

(3) 所述sgRNA在SALL1基因上的靶序列是唯一的;

所述靶序列为序列表中SEQ ID NO:3或4所示的序列。

2. sgRNA或其在SALL1基因上的靶序列在制备运用CRISPR-Cas9特异性敲除猪SALL1基因的试剂盒中的用途, 其特征在于, 所述用途包括如下步骤:

(1) 在权利要求1所述的sgRNA的靶序列的5' -端加上用于形成粘性末端的序列, 合成得到正向寡核苷酸序列; 在权利要求1所述的sgRNA的靶序列对应的互补序列的两端加上合适的用于形成粘性末端的序列, 合成得到反向寡核苷酸序列; 将合成的所述正向寡核苷酸序列与反向寡核苷酸序列退火、复性, 形成具有粘性末端的双链寡聚核苷酸;

(2) 将所述双链寡聚核苷酸连入线性化的携带Cas9基因的表达载体, 得到携带含相应靶序列的sgRNA寡聚核苷酸和Cas9基因的表达载体, 转化感受态细菌, 筛选鉴定出正确的阳性克隆, 并对所述阳性克隆摇菌、提取质粒;

(3) 用所述携带有sgRNA寡聚核苷酸和Cas9基因的表达载体、包装质粒和包装细胞系包装出同时携带靶向SALL1基因的sgRNA和Cas9的假型慢病毒;

(4) 使用所述假型慢病毒感染目的细胞, 并进一步培养; 然后收集被感染的目的细胞, 以其基因组DNA为模板扩增包含所述靶序列的基因片段, 经过变性、复性及酶切, 确定SALL1基因的敲除情况。

3. 根据权利要求2所述的用途, 其特征在于, 所述表达载体为序列表中SEQ ID NO:50所示序列的载体。

4. 根据权利要求2或3所述的用途, 其特征在于, 所述用途包括如下步骤:

(1) 在权利要求1所述的sgRNA的靶序列的5' -端加上CACCG序列, 合成得到正向寡核苷酸序列; 在权利要求1所述的sgRNA的靶序列对应的互补序列的5' -端加上AAAC序列、3' -端加上C, 合成得到反向寡核苷酸序列; 将合成的所述正向寡核苷酸序列与反向寡核苷酸序列退火、复性, 形成具有粘性末端的双链寡聚核苷酸;

(2) 将所述双链寡聚核苷酸连入如序列表中SEQ ID NO:50所示序列的表达载体lentiCRISPR v2经BsmBI限制性内切酶酶切得到的线性化载体, 得到携带sgRNA寡聚核苷酸的重组表达载体lentiCRISPR v2-SALL1, 转化感受态细菌, 筛选鉴定出正确的阳性克隆, 并对所述阳性克隆摇菌、提取质粒;

(3) 用所述表达载体lentiCRISPR v2-SALL1、包装质粒和包装细胞系包装出同时携带靶向SALL1基因的sgRNA和Cas9的假型慢病毒;

(4) 使用所述假型慢病毒感染目的细胞, 并进一步培养; 然后收集被感染的目的细胞, 以其基因组DNA为模板扩增包含所述靶序列的基因片段, 经过变性、复性及酶切, 确定SALL1

基因的敲除情况。

5. 根据权利要求4所述的用途,其特征在于,所述包装质粒为质粒pLP1、质粒pLP2和质粒pLP/VSVG;所述包装细胞系为HEK293T细胞。

6. 根据权利要求4所述的用途,其特征在于,所述目的细胞为猪PIEC细胞;

所述以其基因组DNA为模板扩增包含所述靶序列的基因片段,经过变性、复性及酶切,确定SALL1基因的敲除情况,具体为:

(a) 以感染病毒的目的细胞的基因组DNA为模板,用SALL1基因的上下游引物扩增包含所述sgRNA的靶序列的SALL1基因片段,同时用相同引物扩增未感染病毒的野生型细胞的基因组DNA;

(b) 纯化上述扩增到的SALL1基因片段,然后将来自感染病毒的目的细胞的SALL1基因片段和来自野生型细胞的SALL1基因片段分别加热变性、复性,形成杂交DNA分子;

(c) 用Cruiser酶切割复性后的杂交DNA分子;

(d) 电泳检测酶切产物,检测靶序列介导的SALL1基因敲除效果。

7. 在CRISPR-Cas9特异性敲除猪SALL1基因的方法中用到的重组表达载体lentiCRISPR v2-SALL1,其特征在于,所述重组表达载体的骨架载体的序列如序列表中SEQ ID NO:50所示;所携带的靶序列选自序列表中SEQ ID NO:3或4所示的靶序列。

8. 如权利要求7所述的重组表达载体lentiCRISPR v2-SALL1在制备运用CRISPR-Cas9特异性敲除猪SALL1基因的试剂盒中的用途。

CRISPR-Cas9特异性敲除猪SALL1基因的方法及用于特异性靶向SALL1基因的sgRNA

技术领域

[0001] 本发明涉及基因工程技术领域,尤其涉及基因敲除技术领域,具体涉及CRISPR-Cas9特异性敲除猪SALL1基因的方法及用于特异性靶向SALL1基因的sgRNA。

背景技术

[0002] 器官移植是治疗器官衰竭疾病最有效的治疗手段。迄今为止,全球已有近百万的患者通过器官移植而延续生命。随着人口老龄化及医疗技术的进步,需要进行器官移植手术的病人越来越多,但供体器官的短缺严重制约了器官移植手术的开展。以肾脏移植为例,我国每年需要进行肾移植的患者多达30万,而可用于移植的捐献肾脏不超过1万例,大部分患者死于肾衰竭。依靠死后器官捐献已不能满足器官移植的需要。通过基因工程改造其他物种,以提供合适于人体移植的器官,成为解决人类供体器官短缺问题的主要途径。

[0003] 目前,根据生物安全性、生理功能指标、经济性及稀有物种保护等多方面评价,猪成为了最为理想的异种器官来源。但猪和人之间存在巨大的差异,直接将猪的器官移植到会产生强烈的免疫排斥反应。因此,通过基因工程对猪进行改造,以产生适合于人体移植的器官,成为异种移植的终极目标。

[0004] 传统的技术路线通过减少猪与人的免疫差异以获得可用于移植的猪的品系。近年来,利用器官发育缺陷型猪作为培养环境产生由人类细胞构成的器官成为新的思路。通过基因工程有效干扰控制猪自身器官发育的基因,使猪在发育过程中某个器官缺失,从而为人源细胞器官的发育提供了关键的培养环境。

[0005] 目前已知SALL1基因是肾脏发育中的必需基因。SALL1在果蝇中存在同源基因,在进化上高度保守。SALL1基因仅在肾发育阶段表达,组织学上分布于后肾间叶细胞以及肾基质,对于输尿管芽进入间质组织是必需的。缺失SALL1基因会导致新生鼠肾发育不全或缺失,表明SALL1调控肾发育过程中的关键步骤。敲除SALL1基因能够使猪在发育过程中不产生肾脏,为人类细胞来源肾提供了良好的发育环境。而准确高效的敲除猪的SALL1基因,是首要步骤。

[0006] 目前,常见的基因敲除技术包括同源重组(Homologous Recombination,HR)技术、类转录激活效应子核酸酶(Transcription Activator-Like Effector Nuclease,TALEN)技术、锌指核酸酶(Zinc-Finger Nuclease,ZFN)技术以及最近发展的规律成簇间隔短回文重复(Clustered Regularly Interspaced Short Palindromic Repeat,CRISPR)技术。HR技术由于重组效率低下(效率大约只有 10^{-6}),对突变体的筛选工作非常耗时和低效,已逐渐被取代。TALEN技术和ZFN技术的切割效率一般能达到20%,但都需要构建可以识别特定序列的蛋白质模块,前期工作繁琐费时。ZFN技术的模块设计较为复杂且有较高的脱靶率,其应用有限。

[0007] CRISPR是一种源于原核生物的后天免疫系统,该系统执行干扰功能的复合物由蛋白质Cas和CRISPR-RNA(crRNA)组成。目前该系统已发现有三种类型,其中第二类Cas9系统

组成简单,已被积极应用于基因工程领域。Cas9靶向切割DNA是通过两种小RNA——crRNA (CRISPR RNA) 和tracrRNA (trans-activating crRNA) 与靶序列互补识别的原理实现的。现在已经将两种小RNA融合成一条RNA链,简称sgRNA (single guide RNA),能够识别特定的基因序列,引导Cas9蛋白进行切割。在真核生物中,DNA被切断后发生非同源重组末端连接,造成移码突变,最终导致基因功能性敲除。

[0008] 相比于上述3种技术,CRISPR操作简单、筛选效率高,能够实现精确的靶向切割。因此,通过CRISPR技术敲除SALL1基因能够极大地提高SALL1缺失细胞及肾发育缺失基因工程猪的筛选效率。但是该路径的关键技术难题是设计并制备精确靶向的sgRNA,因为基因的靶向精确度高度依赖于sgRNA靶序列,能否成功设计出精确靶向的sgRNA成为敲除目的基因的关键技术问题,本发明意在解决该技术问题从而为敲除SALL1基因提供坚实的基础。

发明内容

[0009] 本发明的目的在于提供CRISPR-Cas9特异性敲除猪SALL1基因的方法及用于特异性靶向SALL1基因的sgRNA。

[0010] 根据本发明的第一方面,本发明提供在CRISPR-Cas9特异性敲除猪SALL1基因中用于特异性靶向SALL1基因的sgRNA,该sgRNA具有以下特点:

[0011] (1) 该sgRNA在SALL1基因上的靶序列符合5' -N (20) NGG-3' 的序列排列规则,其中N (20) 表示20个连续的碱基,其中每个N表示A或T或C或G,符合上述规则的靶序列位于正义链或反义链;

[0012] (2) 该sgRNA在SALL1基因上的靶序列位于SALL1基因的N端的3个外显子编码区,或靶序列的一部分位于SALL1基因的N端的3个外显子,其余部分跨越与相邻内含子的交界,位于相邻内含子;

[0013] (3) 该sgRNA在SALL1基因上的靶序列是唯一的。

[0014] 作为本发明的优选方案,上述靶序列为序列表中SEQ ID NO:1~49中任一条序列所示的序列。

[0015] 作为本发明的优选方案,上述靶序列为序列表中SEQ ID NO:3或4所示的序列。

[0016] 根据本发明的第二方面,本发明提供运用CRISPR-Cas9特异性敲除猪SALL1基因的方法,该方法包括如下步骤:

[0017] (1) 在第一方面所述的sgRNA的靶序列的5' -端加上用于形成粘性末端的序列,合成得到正向寡核苷酸序列;在第一方面所述的sgRNA的靶序列对应的互补序列的两端加上合适的用于形成粘性末端的序列,合成得到反向寡核苷酸序列;将合成的正向寡核苷酸序列与反向寡核苷酸序列退火、复性,形成具有粘性末端的双链寡聚核苷酸;

[0018] (2) 将上述双链寡聚核苷酸连入线性化的携带Cas9基因的表达载体,得到携带含相应靶序列的sgRNA寡聚核苷酸和Cas9基因的表达载体,转化感受态细菌,筛选鉴定出正确的阳性克隆,并对阳性克隆摇菌、提取质粒;

[0019] (3) 用上述携带有sgRNA寡聚核苷酸和Cas9基因的表达载体、包装质粒和包装细胞系包装出同时携带靶向SALL1基因的sgRNA和Cas9的假型慢病毒;

[0020] (4) 使用上述假型慢病毒感染目的细胞,并进一步培养;然后收集被感染的目的细胞,以其基因组DNA为模板扩增包含上述靶序列的基因片段,经过变性、复性及酶切,确定

SALL1基因的敲除情况。

[0021] 作为本发明的优选方案,上述表达载体为序列表中SEQ ID NO:50所示序列的载体。

[0022] 作为本发明的优选方案,上述方法包括如下步骤:

[0023] (1)在第一方面所述的sgRNA的靶序列的5'-端加上CACCG序列,合成得到正向寡核苷酸序列;在第一方面所述的sgRNA的靶序列对应的互补序列的5'-端加上AAAC序列、3'-端加上C,合成得到反向寡核苷酸序列;将合成的正向寡核苷酸序列与反向寡核苷酸序列退火、复性,形成具有粘性末端的双链寡聚核苷酸;

[0024] (2)将上述双链寡聚核苷酸连入如序列表中SEQ ID NO:50所示序列的表达载体lentiCRISPR v2经BsmBI限制性内切酶酶切得到的线性化载体,得到携带sgRNA寡聚核苷酸的重组表达载体lentiCRISPR v2-SALL1,转化感受态细菌,筛选鉴定出正确的阳性克隆,并对阳性克隆摇菌、提取质粒;

[0025] (3)用上述表达载体lentiCRISPR v2-SALL1、包装质粒和包装细胞系包装出同时携带靶向SALL1基因的sgRNA和Cas9的假型慢病毒;

[0026] (4)使用上述CRISPR假型慢病毒感染目的细胞,并进一步培养;然后收集被感染的目的细胞,以其基因组DNA为模板扩增包含上述靶序列的基因片段,经过变性、复性及酶切,确定SALL1基因的敲除情况。

[0027] 作为本发明的优选方案,上述包装质粒为质粒pLP1、质粒pLP2和质粒pLP/VSVG;上述包装细胞系为HEK293T细胞。

[0028] 作为本发明的优选方案,上述目的细胞为猪PIEC细胞。

[0029] 作为本发明的优选方案,上述以其基因组DNA为模板扩增包含上述靶序列的基因片段,经过变性、复性及酶切,确定SALL1基因的敲除情况,具体为:

[0030] (a)以感染病毒的目的细胞的基因组DNA为模板,用SALL1基因的上下游引物扩增包含上述sgRNA的靶序列的SALL1基因片段,同时用相同引物扩增未感染病毒的野生型细胞的基因组DNA;

[0031] (b)纯化上述扩增到的SALL1基因片段,然后将来自感染病毒的目的细胞的SALL1基因片段和来自野生型细胞的SALL1基因片段分别加热变性、复性,形成杂交DNA分子;

[0032] (c)用Cruiser酶切割复性后的杂交DNA分子;

[0033] (d)电泳检测酶切产物,检测靶序列介导的SALL1基因敲除效果。

[0034] 根据本发明的第三方面,本发明提供在CRISPR-Cas9特异性敲除猪SALL1基因的方法中用到的重组表达载体lentiCRISPR v2-SALL1,该重组表达载体的骨架载体的序列如序列表中SEQ ID NO:50所示;所携带的靶序列如第一方面的sgRNA的靶序列,优选序列表中SEQ ID NO:3或4所示的靶序列。

[0035] 根据本发明的第四方面,本发明提供如第一方面所述的sgRNA或第三方面所述的重组表达载体lentiCRISPR v2-SALL1在CRISPR-Cas9特异性敲除猪SALL1基因的方法中的用途。

[0036] 本发明的针对CRISPR-Cas9特异性敲除猪SALL1基因,成功地找到特异性靶向SALL1基因的sgRNA,将本发明的sgRNA用于CRISPR-Cas9特异性敲除猪SALL1基因的方法中,能够快速、精确、高效、特异性地敲除猪SALL1基因,有效地解决构建SALL1基因敲除猪周期

长和成本高的技术问题。

附图说明

- [0037] 图1为本发明实施例中使用的载体质粒lentiCRISPR v2的质粒图谱；
- [0038] 图2为本发明实施例中使用的包装质粒pLP1的质粒图谱；
- [0039] 图3为本发明实施例中使用的包装质粒pLP2的质粒图谱；
- [0040] 图4为本发明实施例中使用的包装质粒pLP/VSVG的质粒图谱；
- [0041] 图5为本发明实施例中酶切验证靶序列的基因敲除效果的电泳检测结果图，其中M表示DNA Marker，WT表示未经过病毒感染和Cas9切割的野生型细胞的PCR产物Cruiser酶切检测结果，3和4分别表示表1中第3号和第4号靶序列对SALL1基因的靶向切割效果，箭头处表示经Cruiser酶切割得到的小片段。

具体实施方式

[0042] 下面结合附图和具体实施例对本发明的技术方案做进一步说明。这些附图和具体实施例不用来限制本发明的范围。若未特别指明，实施例中所用的技术手段为本领域技术人员所熟知的常规手段，所用原料均为市售商品。

[0043] 以下实施例中涉及的试验材料和试剂：lentiCRISPR v2质粒购自Addgene公司，包装质粒pLP1、pLP2和pLP/VSVG购自Invitrogen公司，包装细胞系HEK293T细胞购自美国模式培养物集存库(ATCC)，PIEC细胞购自中国科学院细胞库，DMEM培养基、Opti-MEM培养基和胎牛血清FBS购自Gibco公司，Lipofectamine2000购自Invitrogen公司。

[0044] 以下实施例中未作具体说明的分子生物学实验方法，均参照《分子克隆实验指南》(第三版)J.萨姆布鲁克一书中描述的具体方法进行，或者按照试剂盒和产品说明书进行。

[0045] 本发明的概括性的技术方案包括以下五个部分：

[0046] 一、Sus scrofa (猪) SALL1基因sgRNA靶序列的选择和设计

[0047] 1. SALL1基因的sgRNA靶序列选择：

[0048] 在SALL1基因外显子区寻找合适的20bp寡核苷酸序列作为靶序列。

[0049] 2. SALL1基因的sgRNA靶序列设计：

[0050] 将上述靶序列及互补序列分别添加接头，形成正向寡核苷酸序列和反向寡核苷酸序列。

[0051] 二、构建SALL1基因的CRISPR载体

[0052] 1. 合成上述正向寡核苷酸序列和反向寡核苷酸序列，复性形成具有粘性末端的双链DNA片段(即双链靶序列寡聚核苷酸，也可以称为双链寡聚核苷酸)。

[0053] 2. 构建CRISPR-sgRNA表达载体：

[0054] 将上述双链DNA片段构建至目标载体(如lentiCRISPR v2，其质粒图谱如图1所示)，形成如lentiCRISPR v2-SALL1的慢病毒CRISPR载体。

[0055] 三、获得表达SALL1sgRNA的假型慢病毒

[0056] 利用包装质粒、包装细胞系与慢病毒CRISPR载体生产表达SALL1sgRNA的CRISPR假型慢病毒。

[0057] 四、感染目的细胞并检测SALL1基因敲除效果

[0058] 1.慢病毒感染目的细胞:

[0059] 将如lentiCRISPR v2-SALL1的假型慢病毒加入目的细胞培养基进行感染并进一步培养。

[0060] 2.检测SALL1基因敲除效果:

[0061] 收集目的细胞,以基因组DNA为模板扩增包含靶序列的基因片段,经过变性、复性及酶切,确定SALL1基因的敲除情况。

[0062] 五、SALL1基因敲除单克隆的挑选和鉴定

[0063] 1.对于有确定敲除效果的目的细胞群,通过稀释和单克隆培养,分离出若干单细胞来源的细胞株。

[0064] 2.鉴定单克隆的SALL1敲除情况。

[0065] 以下通过实施例详细说明本发明的技术方案及其有益效果。

[0066] 实施例一、Sus scrofa (猪) SALL1基因sgRNA靶序列的选择和设计

[0067] 靶序列决定了sgRNA的靶向特异性和诱导Cas9切割目的基因的效率。因此,高效特异的靶序列选择和设计是构建sgRNA表达载体的前提。

[0068] 1.SALL1基因的sgRNA靶序列选择

[0069] 针对SALL1基因,在靶序列选择上应该遵循下列原则:

[0070] (1)在SALL1基因外显子编码区寻找符合5'-N(20)NGG-3'规则的靶序列,其中N(20)表示20个连续的碱基,其中每个N表示A或T或C或G,符合上述规则的靶序列位于正义链或反义链;

[0071] (2)选择靠近N端的3个外显子编码区序列,靶序列可以位于SALL1基因的N端的3个外显子编码区,或靶序列的一部分位于SALL1基因的N端的3个外显子,其余部分跨越与相邻内含子的交界,位于相邻内含子;这样的编码区序列的切割会造成SALL1基因的功能敲除,残留截短的序列不会形成有功能的蛋白;

[0072] (3)如果存在多种剪切体,则在共有外显子编码区进行选择,针对SALL1基因选择靠近N端的3个外显子编码区序列即可满足该条件;

[0073] (4)利用在线序列分析工具(<http://crispr.mit.edu/>)分析以上靶序列在猪基因组中的同源情况,舍弃存在显著同源序列的靶序列,根据评分进一步挑选,所挑选的靶序列在SALL1基因上是唯一的。

[0074] 基于以上原则,选择出表1所示的靶序列集合。

[0075] 表1靶序列集合

[0076]

编号	序列
1	TTGCTTCCTCCGCGACATGC
2	AGGCCACTTCGGGGTCGGAT
3	TTTCCAATCCGACCCCGAAG
4	GAGCGAGGCCACTTCGGGGT
5	GGGTCGGATTGGAATGTTG
6	GCCTCGCTCCCCGGCGAGA
7	CGGGGGAGCGAGGCCACTTC
8	CCGAAGTGGCCTCGCTCCCC

[0077]

9	GGGGGAGCGAGGCCACTTCG
10	CCGGGGGAGCGAGGCCACTT
11	ACAGCACCGGCCGACAGCGT
12	CATTTGTTTCATCGGGATGAT
13	TTGCTCTTAGTGGTGCGGTT
14	CATTTACGATTAAAAACGAGT
15	CTTGCTCTTAGTGGTGCGGT
16	TCATTTGTTTCATCGGGATGA
17	CAACCGCACCCTAAGAGCA
18	ATTTGTTTCATCGGGATGATG

19	CACAGCACCGGCCGACAGCG
20	AGGATGCCCACGTCTGCGGC
21	GTGGGCATCCTTGCTCTTAG
22	CAAAGAACTCGGCACAGCAC
23	AGCAAGGATGCCCACGTCTG
24	CATCCTTGCTCTTAGTGGTG
25	TTTTCGTGCAGCTCTTCTTG
26	TTCATCGGGATGATGGGGAG
27	CGTGTCAATTCATTTGTTTCAT
28	ACGGTTAACAAAAACAGAGCA
29	GGACTGGGGGAGAAGGGTTC
30	GCCGCCTTTGTTAGCCACCG
31	GAGGCCCGGTGGCTAACAA
32	AGAACACCACGGGCCGGACA
33	TGCCGCCTTTGTTAGCCACC
34	CAGAACACCACGGGCCGGAC
35	GCCCCGGTGGCTAACAAAGG
36	CTGTCCGGCCCGTGGTGTTTC
37	GTGGCTAACAAAGGCGGCAG
38	GCGACCTTCCAGAACACCAC
39	CCGGCCCGTGGTGTTCTGGA
40	ACTCTTCCCTGTCCGGCCCG
41	CCTTCCAGAACACCACGGGC
42	CTGCCGCCTTTGTTAGCCAC
43	TCCATGGACTCTTCCCTGTC

[0078]

44	AGCGACCTTCCAGAACACCA
45	AAAGGCGGCAGTGGCCCCCT
46	GTGGCAGCAGCGATGGCAGT
47	GGACAGGGAAGAGTCCATGG

48	GCCGGACAGGGAAGAGTCCA
49	CAAAGCGGCAGTGGCCCCC

[0079] 2. SALL1基因的sgRNA靶序列设计:

[0080] (1) 以lentiCRISPR v2质粒作为表达载体,根据lentiCRISPR v2质粒的特点,在上述N(20)靶序列的5'-端添加CACCG序列,形成正向寡核苷酸序列:

[0081] 5'-CACCGNNNNNNNNNNNNNNNNNNNN-3';

[0082] (2) 在上述N(20)靶序列的反向互补序列的两端添加序列,形成反向寡核苷酸序列:

[0083] 5'-AAACNNNNNNNNNNNNNNNNNNNNNC-3';

[0084] 正向寡核苷酸序列和反向寡核苷酸序列可以互补形成具有粘性末端的双链DNA片段:

[0085] 5'-CACCGNNNNNNNNNNNNNNNNNNNN-3'

[0086] 3'-CNNNNNNNNNNNNNNNNNNNNCAA-5'。

[0087] 实施例二、构建SALL1基因的sgRNA表达载体

[0088] 1. 合成DNA插入片段

[0089] (1) 合成上述设计的正向和反向寡核苷酸序列

[0090] 寡核苷酸序列可以由商业化的公司(如Invitrogen公司)根据提供的序列具体合成。本实施例及以下实施例研究了表1中所列的第3号和第4号序列所示的靶序列对SALL1基因的敲除效果。

[0091] 第3号靶序列对应的正向寡核苷酸序列和反向寡核苷酸序列如下:

[0092] CACCGTTTCCAATCCGACCCCGAAG (SEQ ID NO:51);

[0093] AAACCTTCGGGGTCGGATTGGAAC (SEQ ID NO:52)。

[0094] 第4号靶序列对应的正向寡核苷酸序列和反向寡核苷酸序列如下:

[0095] CACCGGAGCGAGGCCACTTCGGGGT (SEQ ID NO:53);

[0096] AAACACCCCGAAGTGGCCTCGCTCC (SEQ ID NO:54)。

[0097] 将对应的正向和反向寡核苷酸序列退火、复性,形成具有粘性末端的双链DNA片段。

[0098] 反应体系(20 μ L)如下所示:

[0099] 正向寡核苷酸(10 μ M):1 μ L

[0100] 反向寡核苷酸(10 μ M):1 μ L

[0101] 10 \times PCR buffer:2 μ L

[0102] ddH₂O:16 μ L

[0103] 将上述反应体系放入PCR仪,并按以下程序进行反应。

[0104] 反应程序:

[0105] 95 $^{\circ}$ C,5min;

[0106] 80 $^{\circ}$ C,5min;

[0107] 70 $^{\circ}$ C,5min;

[0108] 60 $^{\circ}$ C,5min;

[0109] 50 $^{\circ}$ C,5min;

- [0110] 自然降至室温。
- [0111] 2. 构建sgRNA表达载体
- [0112] (1) 利用BsmB I限制性内切酶酶切目标载体lentiCRISPR v2质粒(其序列如序列表中SEQ ID NO:50所示)。
- [0113] 按照以下反应体系进行配制:
- [0114] LentiCRISPR v2质粒:1 μ g
- [0115] 10 \times 酶切buffer:2 μ L
- [0116] BsmB I限制性内切酶:2 μ L
- [0117] 补充ddH₂O至总体积20 μ L
- [0118] 将酶切反应体系置于37 $^{\circ}$ C反应4h。
- [0119] (2) 电泳分离并纯化载体片段
- [0120] 酶切结束后,将酶切混合物通过琼脂糖凝胶电泳进行分离,选择载体片段(约12kb)进行切割,并通过DNA凝胶回收柱进行回收。
- [0121] (3) 将合成的双链DNA片段与载体主片段进行连接并转化大肠杆菌
- [0122] 将复性得到的双链DNA片段与回收得到的载体片段进行连接反应,按照以下反应体系进行配制:
- [0123] LentiCRISPR v2载体片段:100ng
- [0124] 双链DNA片段:200ng
- [0125] T4连接酶:1 μ L
- [0126] T4连接反应buffer:1 μ L
- [0127] 补充ddH₂O至总体积10 μ L
- [0128] 将连接混合物置于25 $^{\circ}$ C反应2h。
- [0129] 反应结束后将连接混合物转化大肠杆菌DH5 α 菌株:向连接混合物中加入100 μ L大肠杆菌DH5 α 感受态细胞,冰上孵育30min;将混合物放入42 $^{\circ}$ C水浴,热激90s后放入冰上冷却;向混合物加入100 μ L LB培养基,37 $^{\circ}$ C摇床培养20min;将混合物涂Amp LB平板,37 $^{\circ}$ C培养14h。
- [0130] (4) 鉴定正确的转化克隆
- [0131] 从Amp LB平板上挑选若干菌落进行扩大培养,提取质粒进行酶切鉴定。挑选可能正确的克隆进行测序,验证插入序列是否正确。对于正确的lentiCRISPR v2-SALL1载体克隆进行保种。
- [0132] 实施例三、获得表达SALL1sgRNA的假型慢病毒
- [0133] 1. 材料准备
- [0134] 扩增并抽提包装质粒pLP1、pLP2和pLP/VSVG(购自Invitrogen,其图谱分别如图2、图3和图4所示);扩增并抽提载体质粒lentiCRISPR v2-SALL1;培养包装细胞系HEK293T细胞(购自ATCC);DMEM培养基、Opti-MEM培养基和胎牛血清FBS(购自Gibco);Lipofectamine2000(购自Invitrogen);HEK293T细胞培养于含5%CO₂的37 $^{\circ}$ C培养环境中,培养基为含10%FBS的DMEM培养基。
- [0135] 2. 转染和病毒包装
- [0136] 第一天:将包装细胞系HEK293T传代至10cm dish,约30%融合度;

- [0137] 第二天:在HEK293T达到80%融合度时按照下列配方进行转染:
- [0138] 配制混合物1,包含:
- [0139] lentiCRISPR v2-SALL1:6 μ g
- [0140] pLP1:6 μ g
- [0141] pLP2:6 μ g
- [0142] pLP/VSVG:3 μ g
- [0143] Opti-MEM:500 μ L。
- [0144] 配制混合物2,包含:
- [0145] Lipofectamine 2000:30 μ L
- [0146] Opti-MEM:500 μ L。
- [0147] 静置5min后,将混合物1和混合物2混匀成转染混合物,静置20min。
- [0148] 将HEK293T培养基换为无血清DMEM培养基,加入转染混合物,37 $^{\circ}$ C培养8h后换为20%FBS的DMEM培养基,继续培养。
- [0149] 3. 病毒收集与保存
- [0150] 第三天:转染48h后收集含病毒的HEK293T培养基上清,用0.45 μ m滤头过滤后,分装,放置-80 $^{\circ}$ C保存。
- [0151] 实施例四、感染目的细胞并检测靶序列的敲除效果
- [0152] 1. 材料准备
- [0153] 培养目的细胞系猪髂动脉血管内皮细胞PIEC(购自中国科学院细胞库);DMEM培养基和胎牛血清FBS(购自Gibco);不同靶序列(序列3和序列4)的lentiCRISPR v2-SALL1假型慢病毒;PIEC细胞培养于含5%CO₂的37 $^{\circ}$ C培养环境中,培养基为含10%FBS的DMEM培养基。
- [0154] 2. 慢病毒感染目的细胞
- [0155] 第一天:将目的细胞传代至6孔板,约20%融合密度。每一种病毒需要一个6孔,同时需要效率对照一个6孔。
- [0156] 第二天:待目的细胞约40%融合密度时加入1mL lentiCRISPR v2-SALL1假型慢病毒上清及1mL DMEM培养基。效率对照不需要添加慢病毒。
- [0157] 第三天:感染24h后去除含病毒培养基,换成正常培养基,加入嘌呤霉素至终浓度2 μ g/mL,没有感染病毒的效率对照样品也同时加入嘌呤霉素作为对照,培养48h。
- [0158] 3. 细胞感染效率检测和培养
- [0159] 第五天:未感染的效率对照细胞在嘌呤霉素的作用下应该全部凋亡(>95%)。根据感染慢病毒细胞的凋亡情况判断细胞的感染效率,通常可以达到90%以上的感染效率(凋亡率<10%)。必要时可以将病毒上清进行浓缩或梯度稀释后进行感染以达到合适的感染效率。
- [0160] 经过嘌呤霉素筛选后,未感染的细胞发生凋亡。将目的细胞重新传代并换为普通培养基培养48h。
- [0161] 4. 检测SALL1基因敲除效果
- [0162] (1) 设计上下游引物以扩增SALL1基因片段,其中上下游引物序列如下所示:
- [0163] GAGCCCCTCTATGATTAATCGCAATGCA (SEQ ID NO:55)
- [0164] GTGGGTCCAAGTGTGCGTGAGTG (SEQ ID NO:56)。

[0165] 目的扩增片段包含sgRNA靶序列,大小为371bp。靶序列至片段两端的位置不少于100bp。

[0166] (2)收集部分目的细胞,使用promega基因组DNA试剂盒抽提基因组DNA。同时抽提野生型目的细胞的基因组DNA。

[0167] (3)以基因组DNA为模板扩增包含靶序列的SALL1基因片段(包括感染的突变样品和野生型样品)。

[0168] 扩增反应体系(20 μ L)如下:

[0169] 上游引物(10 μ M):1 μ L

[0170] 下游引物(10 μ M):1 μ L

[0171] 2 \times PCR Mix:10 μ L

[0172] 基因组DNA:100ng

[0173] 以上述反应体系进行配制,放入PCR仪,并按下列程序进行反应。

[0174] 反应程序:

[0175] 95 $^{\circ}$ C,3min

[0176] 95 $^{\circ}$ C,30s

[0177] 58 $^{\circ}$ C,20s

[0178] 72 $^{\circ}$ C,20s

[0179] 72 $^{\circ}$ C,3min;

[0180] 其中第二步至第四步重复35个循环。

[0181] (4)电泳检测PCR产物并回收纯化

[0182] (5)将纯化后的DNA片段分别加热变性、复性,形成杂交DNA分子(包括突变样品和野生型样品)。

[0183] 反应体系如下所示:

[0184] 基因组PCR片段:200ng

[0185] 5 \times 反应buffer:2 μ L

[0186] 反应体系共9 μ L

[0187] 以上述反应体系进行配制,放入PCR仪,并按下列程序进行反应。

[0188] 反应程序:

[0189] 95 $^{\circ}$ C,5min;

[0190] 80 $^{\circ}$ C,5min;

[0191] 70 $^{\circ}$ C,5min;

[0192] 60 $^{\circ}$ C,5min;

[0193] 50 $^{\circ}$ C,5min;

[0194] 自然降至室温。

[0195] (6)用Cruiser酶切割复性后的杂交DNA(包括突变样品和野生型样品)向经过变性、复性的反应混合物加入1 μ L Cruiser酶,45 $^{\circ}$ C孵育20min。

[0196] (7)电泳检测酶切产物,检测靶序列介导的SALL1基因敲除效果。

[0197] 将经过酶切的DNA片段用2%的琼脂糖凝胶进行电泳分析,100V,25min。确定目的片段的切割情况,判断靶序列的基因敲除效果。

[0198] 对突变DNA的切割识别基于以下原理:经过感染的细胞会表达sgRNA和Cas9。基因组DNA如果被sgRNA介导的Cas9蛋白靶向切割,经过修复后会在切割位点附近引入突变(野生型变为突变型)。由于野生型和突变型序列在该位置不匹配,以此为模板扩增出的野生型DNA与突变型DNA经过变复性形成的杂交分子会就产生局部的环形(loop)结构。而后者可以被Cruiser酶识别并切断,导致杂交DNA分子被切割成小片段。突变样品中由于含有部分野生型DNA成分,因而变复性后可以形成含有局部环形结构的杂交分子。

[0199] 结果如图5所示,未经过病毒感染的野生型细胞未产生切割,因此未检测到小片段;而序列3和序列4能够有效靶向SALL1基因产生切割,因此检测到小片段的存在,表明序列3和序列4能够作为CRISPR-Cas9特异性敲除猪SALL1基因的靶序列。

[0200] 实施例五、SALL1基因敲除单克隆的挑选和鉴定

[0201] 1. 单克隆的挑选(基于序列3和序列4的靶序列)

[0202] (1) 将部分感染的目的细胞群进行传代,取100个单细胞转移至10cm dish培养。

[0203] (2) 培养约10天后,有相当数量的单克隆生长到肉眼可见的水平。

[0204] (3) 用移液器头刮取独立的克隆,将细胞转移至24孔板中培养,每个孔对应一个克隆。

[0205] (4) 再经过约一周的培养后,有部分克隆长至足够的数量,准备做进一步的鉴定。

[0206] 2. 鉴定单克隆的SALL1敲除情况

[0207] (1) 收集待检的单克隆及野生型细胞,分别抽提基因组DNA。

[0208] (2) 按照前述方法,分别扩增单克隆及野生型细胞的SALL1基因片段,所扩增的基因片段包含sgRNA靶序列。

[0209] (3) 将等量的单克隆PCR片段与野生型PCR片段混合,加热变性、复性,形成杂交DNA分子。

[0210] (4) 用Cruiser酶切割退火后的杂交DNA,45℃孵育20min。

[0211] (5) 电泳检测酶切产物,根据是否有切割片段确定单克隆是否发生有效突变。

[0212] 结果显示,基于序列3所示的靶序列的lentiCRISPR v2-SALL1假型慢病毒感染目的细胞,从100个单细胞中随机挑选的20个单克隆经Cruiser酶酶切电泳检测,其中有18个单克隆能检测到切割小片段,表明基因敲除发生,基因敲除效率能够达到90%以上,说明序列3所示的靶序列具有很高的靶向敲除SALL1基因的作用。基于序列4所示的靶序列的lentiCRISPR v2-SALL1假型慢病毒感染目的细胞,从100个单细胞中随机挑选的20个单克隆经Cruiser酶酶切电泳检测,其中有19个单克隆能检测到切割小片段,表明基因敲除发生,基因敲除效率能够达到95%以上,说明序列4所示的靶序列具有很高的靶向敲除SALL1基因的作用。

[0213] 以上内容是结合具体的实施方式对本发明所作的进一步详细说明,不能认定本发明的具体实施只局限于这些说明。对于本发明所属技术领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干简单推演或替换。

SEQUENCE LISTING

<110>	深圳市第二人民医院	
<120>	CRISPR-Cas9 特异性敲除猪 SALL1 基因的方法及用于特异性靶向 SALL1 基因的 sgRNA	
<130>	15P21486	
<160>	56	
<170>	PatentIn version 3.3	
<210>	1	
<211>	20	
<212>	DNA	
<213>	人工序列	
<400>	1	
	ttgcttctc cgcgacatgc	20
<210>	2	
<211>	20	
<212>	DNA	
<213>	人工序列	
<400>	2	
	aggccacttc ggggtcggat	20
[0001]		
<210>	3	
<211>	20	
<212>	DNA	
<213>	人工序列	
<400>	3	
	tttccaatcc gaccccgaag	20
<210>	4	
<211>	20	
<212>	DNA	
<213>	人工序列	
<400>	4	
	gagcgaggcc acttcggggt	20
<210>	5	
<211>	20	
<212>	DNA	
<213>	人工序列	
<400>	5	
	gggtcggatt ggaaatgttg	20
<210>	6	

	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 6	
	gcctcgctcc cccggcgaga	20
	<210> 7	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 7	
	cgggggagcg aggccaattc	20
	<210> 8	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 8	
	ccgaagtggc ctcgctcccc	20
[0002]	<210> 9	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 9	
	ccgaagtggc ctcgctcccc	20
	<210> 10	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 10	
	ccgggggagc gaggccactt	20
	<210> 11	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 11	
	acagcaccgg ccgagacgt	20
	<210> 12	
	<211> 20	
	<212> DNA	

	<213> 人工序列	
	<400> 12	
	catttgttca tcgggatgat	20
	<210> 13	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 13	
	ttgctcttag tgggtcgggt	20
	<210> 14	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 14	
	catttacgat taaaacgagt	20
	<210> 15	
	<211> 20	
	<212> DNA	
[0003]	<213> 人工序列	
	<400> 15	
	cttgcctta gtggtcgggt	20
	<210> 16	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 16	
	tcatttggtc atcgggatga	20
	<210> 17	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 17	
	caaccgcacc actaagagca	20
	<210> 18	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 18	

	atttgttcat cgggatgatg	20
	<210> 19	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 19	
	cacagcaccg gccgcagacg	20
	<210> 20	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 20	
	aggatgccca cgtctgcggc	20
	<210> 21	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 21	
[0004]	gtgggcatcc ttgctcttag	20
	<210> 22	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 22	
	caaagaactc ggcacagcac	20
	<210> 23	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 23	
	agcaaggatg cccacgtctg	20
	<210> 24	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 24	
	catccttgct cttagtgggtg	20

	<210> 25	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 25	
	ttttcgtgca gctcttcttg	20
	<210> 26	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 26	
	ttcatcggga tgatggggag	20
	<210> 27	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 27	
	cgtgtcattc atttgttcat	20
[0005]	<210> 28	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 28	
	acggttaaca aaacagagca	20
	<210> 29	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 29	
	ggactggggg agaagggttc	20
	<210> 30	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 30	
	gccgcctttg ttagccaccg	20
	<210> 31	
	<211> 20	

	<212> DNA	
	<213> 人工序列	
	<400> 31	
	gaggccccgg tggctaacaa	20
	<210> 32	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 32	
	agaacaccac gggccggaca	20
	<210> 33	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 33	
	tgccgccttt gttagccacc	20
	<210> 34	
	<211> 20	
	<212> DNA	
[0006]	<213> 人工序列	
	<400> 34	
	cagaacacca cgggccggac	20
	<210> 35	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 35	
	gccccggtgg ctaacaaagg	20
	<210> 36	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 36	
	ctgtccggcc cgtggtgttc	20
	<210> 37	
	<211> 20	
	<212> DNA	
	<213> 人工序列	

	<400> 37	
	gtggctaaca aaggcggcag	20
	<210> 38	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 38	
	gcgaccttcc agaacaccac	20
	<210> 39	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 39	
	ccggcccgtg gtgttctgga	20
	<210> 40	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
[0007]	<400> 40	
	actettccct gtccggcccc	20
	<210> 41	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 41	
	ccttccagaa caccacgggc	20
	<210> 42	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 42	
	ctgccgcctt tgttagccac	20
	<210> 43	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 43	
	tccatggact cttccetgtc	20

	<210> 44	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 44	
	agcgaccttc cagaacacca	20
	<210> 45	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 45	
	aaaggcggca gtggccccct	20
	<210> 46	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 46	
	gtggcagcag cgatggcagt	20
[0008]	<210> 47	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 47	
	ggacagggaa gagtccatgg	20
	<210> 48	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 48	
	gccggacagg gaagagtcca	20
	<210> 49	
	<211> 20	
	<212> DNA	
	<213> 人工序列	
	<400> 49	
	caaaggcggc agtggcccc	20
	<210> 50	

	<211>	14873	
	<212>	DNA	
	<213>	表达载体 lentiCRISPR v2	
	<400>	50	
	gtcgacggat	cgggagatct	cccgatcccc
	tatggtgcac	tctcagtaca	atctgctctg
	60		
	atgccgcata	gttaagccag	tatctgctcc
	ctgcttgtgt	gttggaggtc	gctgagtagt
	120		
	gcgcgagcaa	aatttaagct	acaacaaggc
	aaggcttgac	cgacaattgc	atgaagaatc
	180		
	tgcttaggg	taggcgtttt	gcgctgcttc
	gcgatgtacg	ggccagatat	acgcgttgac
	240		
	attgattatt	gactagttaa	taatagtaat
	caattacggg	gtcattagtt	catagcccat
	300		
	atatggagtt	ccgcgttaca	taacttacgg
	taaatggccc	gccttgctga	ccgcccacag
	360		
	acccccgccc	attgacgtca	ataatgacgt
	atgttcccat	agtaacgcca	atagggactt
	420		
	tccattgacg	tcaatgggtg	gagtatttac
	ggtaaactgc	ccacttgcca	gtacatcaag
	480		
	tgtatcatat	gccaagtacg	ccccctattg
	acgtcaatga	cggtaaatgg	cccgcctggc
	540		
	attatgccc	gtacatgacc	ttatgggact
	ttcctacttg	gcagtacatc	tacgtattag
	600		
	tcatcgctat	taccatgggtg	atgcggtttt
	ggcagtacat	caatgggctg	ggatagcgg
	660		
	ttgactcacg	gggatttcca	agtctccacc
	ccattgacgt	caatgggagt	ttgttttggc
	720		
	acccaaatca	acgggacttt	ccaaaatgic
	gtaacaactc	cgccccattg	acgcaaatgg
	780		
	gcggtaggcg	tgtacgggtg	gaggtctata
	taagcagcgc	gttttgctg	tactgggtct
	840		
	ctctggttag	accagatctg	agcctgggag
	ctctctggct	aactagggaa	cccactgctt
	900		
	aagcctcaat	aaagcttgc	ttgagtgctt
	caagtagtgt	gtgccctct	gttgtgtgac
	960		
	tctgtaact	agagatccct	cagacccttt
	tagtcagtgt	ggaaaatctc	tagcagtggc
	1020		
	gcccgaacag	ggacttgaaa	gcaaaagga
	aaccagagga	gctctctcga	cgcaggactc
	1080		
	ggcttgctga	agcgcgcacg	gcaagaggcg
	agggcgccg	actggtgagt	acgccccaaa
	1140		
	ttttgactag	cggaggctag	aaggagagag
	atgggtgcga	gagcgtcagt	attaagcggg
	1200		
	ggagaattag	atcgcgatgg	gaaaaaatc
	ggttaaggcc	agggggaaag	aaaaaatata
	1260		
	aattaaaaca	tatagtatgg	gcaagcaggg
	agctagaacg	attcgcagtt	aatcctggcc
	1320		
	tgttagaaac	atcagaaggc	tgtagacaaa
	tactgggaca	gctacaacca	tcccttcaga
	1380		
	caggatcaga	agaacttaga	tcattatata
	atacagtagc	aacctctat	tgtgtgcatc
	1440		
	aaaggataga	gataaaagac	accaaggaag
	ctttagacaa	gatagaggaa	gagcaaaaaca
	1500		
	aaagtaagac	caccgcacag	caagcggccg
	ctgatcttca	gacctggagg	aggagatatg
	1560		
	agggacaatt	ggagaagtga	attatataaa
	tataaagtag	taaaaattga	accattagga
	1620		
	gtagcaccca	ccaaggcaaa	gagaagagtg
	gtcagagag	aaaaaagagc	agtgggaata
	1680		
	ggagctttgt	tccttgggtt	cttgggagca
	gcaggaagca	ctatgggcgc	agcgtcaatg
	1740		
	acgctgacgg	tacagggcag	acaattattg
	tctggtatag	tgcagcagca	gaacaatttg
	1800		
	ctgagggcta	ttgagggcga	acagcatctg
	ttgcaactca	cagtctgggg	catcaagcag
	1860		
	ctccaggcaa	gaatcctggc	tgtggaaaga
	tacctaaagg	atcaacagct	cctggggatt
	1920		
	tggggttgct	ctggaactct	catttgcacc
	actgctgtgc	cttggaatgc	tagttggagt
	1980		
	aataaatctc	tggaacagat	ttggaatcac
	acgacctgga	tggagtggga	cagagaaatt
	2040		
	aacaattaca	caagcttaat	acactcctta
	attgaagaat	cgcaaaacca	gcaagaaaag
	2100		
	aatgaacaag	aattattgga	attagataaa
	tgggcaagtt	tgtggaattg	gtttaacata
	2160		
	acaaattggc	tgtggtatat	aaaattatc
	ataatgatag	taggaggctt	ggtaggttta
	2220		
	agaatagttt	ttgctgtact	ttctatagtg
	aatagagtta	ggcagggata	ttcaccatta
	2280		
	tcgtttcaga	cccacctccc	aaccccgagg
	ggacctgaca	ggcccgaagg	aatagaagaa
	2340		
	gaaggtggag	agagagacag	agacagatcc
	attcgattag	tgaacggatc	ggcactgcgt
	2400		

[0009]

	gcgccaattc	tgacagacaaa	tggcagatatt	catccacaat	tttaaaagaa	aaggggggat	2460
	tggggggtac	agtgcagggg	aaagaatagt	agacataata	gcaacagaca	tacaaactaa	2520
	agaattacaa	aaacaaatta	caaaaattca	aaattttcgg	gtttattaca	gggacagcag	2580
	agatccagtt	tggttaatta	aggtagcag	ggcctatttc	ccatgattcc	ttcatatttg	2640
	catatacgat	acaaggctgt	tagagagata	attagaatta	atgtgactgt	aaacacaaag	2700
	atattagtac	aaaatacgtg	acgtagaaaag	taataatttc	ttgggtagtt	tgacagttta	2760
	aaattatgtt	ttaaaatgga	ctatcatatg	cttaccgtaa	cttgaagta	tttcgatttc	2820
	ttgcttttat	atatcttggtg	gaaaggacga	aacaccggag	acggttgtaa	atgagcacac	2880
	aaaatacaca	tgctaaaata	ttatattcta	tgacctttat	aaaatcaacc	aaaatcttct	2940
	ttttaataac	tttagtatca	ataattagaa	ttttatggt	cctttttgca	aacttttaat	3000
	aaaaatgagc	aaaataaaaa	aacgctagtt	ttagtaactc	gcgttgtttt	cttcaccttt	3060
	aataatagct	actccaccac	ttgttcttaa	gcggtcagct	cctgcttcaa	tcattttttg	3120
	agcatcttca	aatgttctaa	ctccaccagc	tgctttaact	aaagcattgt	ctttaacaac	3180
	tgacttcatt	agtttaacat	cttcaaatgt	tgacctgat	tttgaaaatc	ctgttgatgt	3240
	tttaacaaat	tctaattccag	cttcaacagc	tatttcaaa	gctttcatga	tttcttcttt	3300
	tgtaataaaa	caattttcca	taatacattt	aacaacatgt	gatccagctg	ctttttttac	3360
	agctttcatg	tcttctaaaa	ctaattcata	atttttgtct	tttaatgcac	caatatttaa	3420
	taccatatca	atctctgttg	caccatcttt	aattgcttca	gaaacttcca	atgcttttgt	3480
	agctgttggtg	catgcacctc	gaggaaaacc	tacaacattt	gttatttcta	cattttgtcc	3540
	ttttaataat	tctttacaat	agcttgttca	atatgaatta	acacaaactg	ttgcaaaatc	3600
	aaattcaatt	gcttcatcac	ataattgttt	aatttcagct	ttcgtagcat	cttgttttta	3660
[0010]	taatgtgtga	tctatatatt	tgtttagttt	cattttttct	cctatatatt	catttttaat	3720
	tttaattctt	taataatttc	gtctacttta	acttttagct	tttgaacaga	ttcaccaca	3780
	cctataaaat	aaatttttag	tttaggttca	gttccacttg	ggcgaacagc	aaatcatgac	3840
	ttatcttcta	aataaaaattt	tagtaagtct	tgctctggca	tattatacat	tccatcgatg	3900
	tagtcttcaa	cattaacaac	tttaagtcca	gcaatttgag	ttaagggtgt	tgctctcaat	3960
	gatttcatta	atggttcaat	ttttaatttc	tttcttctg	gtttaaaatt	caagtttaaa	4020
	gtgaaagtgt	aatatgcacc	catttcttta	aataaatctt	ctaaatagtc	tactaatggt	4080
	ttattttggt	ttttataaaa	tcaagcagcc	tctgctatta	atatagaagc	ttgtattcca	4140
	tctttatctc	tagctgagtc	atcaattaca	tatccataac	tttcttcata	agcaaaaaca	4200
	aaatttaate	cgttatcttc	ttcttttagca	atctctctac	ccattcattt	aaatccagtt	4260
	aaagttttta	caatattaac	tccatatttt	tcatgagcga	ttctatcacc	caaatcactt	4320
	gttacaaaac	ttgaatatag	agccggattt	tttggaatgc	tatttaagcg	ttttagattt	4380
	gataattttc	aatcaattaa	aattggctct	gtttgatttc	catctaatct	tacaaaatga	4440
	ccatcatggt	ttattgcat	tccaaatctg	tcagcatctg	ggtcattcat	aataataata	4500
	tctgcatcat	gttttaatacc	atattcaagc	ggtatttttc	atgcaggatc	aaattctgga	4560
	tttggattta	caacattttt	aaatgtttca	tcttcaaatg	catgctcttc	aacctcaata	4620
	acgttatatc	ctgattcacg	taatatTTTT	ggggtaaatt	tagttctctg	tccattaact	4680
	gcgctaaaaa	taatttttaa	atctttttta	gcttcttctg	cttttttgta	cgctctctgt	4740
	ttagagctag	aaatagcaag	ttaaaataag	gctagctcgt	tatcaacttg	aaaaagtggc	4800
	accgagtcgg	tgcttttttg	aattcgtctg	ctaggtcttg	aaaggagtgg	gaattggctc	4860
	cggtgcccg	cagtgggcag	agcgcacatc	gccacagtc	cccagagaag	tggggggagg	4920
	ggtcggcaat	tgatccggtg	cctagagaag	gtggcgcggg	gtaaactggg	aaagtgatgt	4980
	cgtgtactgg	ctccgccttt	ttcccagagg	tgggggagaa	ccgtatataa	gtgcagtagt	5040

	cgccgtgaac gttctttttc gcaacggggt tgccgccaga acacaggacc ggttctagag	5100
	cgctgccacc atggacaaga agtacagcat cggcctggac atcggcacca actctgtggg	5160
	ctgggccgtg atcaccgacg agtacaaggt gccagcaag aaattcaagg tgctgggcaa	5220
	caccgaccgg cacagcatca agaagaacct gatcggagcc ctgctgttcg acagcggcga	5280
	aacagccgag gccacccggc tgaagagaac gccagaaga agatacacca gacggaagaa	5340
	ccggatctgc tatctgcaag agatcttcag caacgagatg gccaaagggtg acgacagctt	5400
	cttccacaga ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcacc	5460
	catcttcggc aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca	5520
	cctgagaaag aaactggtgg acagcaccga caaggccgac ctgcggtga tctatctggc	5580
	cctggccccc atgatcaagt tccggggcca ctctctgac gagggcgacc tgaaccccga	5640
	caacagcgac gtggacaagc tgttcatcca gctgggtgcag acctacaacc agctgttcga	5700
	ggaaaacccc atcaacgcca gcggcgtgga cgcaaggcc atcctgtctg ccagactgag	5760
	caagagcaga cggctggaaa atctgatcgc ccagctgccc ggcgagaaga agaatggcct	5820
	gttcggaaac ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttca	5880
	cctggccgag gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa	5940
	cctgtctggc cagatcggcg accagtacgc cgacctgtt ctggccgcca agaactgtc	6000
	cgacccatc ctgctgagc acatcctgag agtgaacacc gagatcacca aggccccct	6060
	gagcgcctct atgatcaaga gatacagca gcaccaccag gacctgacc tgctgaaagc	6120
	tctctgctgg cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa	6180
	cggctacgcc ggctacattg acggcggagc cagccaggaa gatttctaca agttcatcaa	6240
	gcccatcctg gaaaagatgg acggcaccga ggaactgctc gtgaaactga acagagagga	6300
[0011]	cctgtctcgg aagcagcggc ccttcgacaa cggcagcatc ccccaccaga tccacctggg	6360
	agagctgcac gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg	6420
	ggaaaagatc gagaagatcc tgaccttcg catcccctac tacgtgggccc ctctggccag	6480
	gggaaacagc agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa	6540
	cttcgaggaa gtggtggaca agggcgcttc cgcccagagc ttcacgagc ggatgaccaa	6600
	cttcgataag aacctgcccc acgagaaggt gctgcccagg cacagcctgc tgtacgagta	6660
	cttcaccgtg tataacgagc tgaccaaagt gaaatacgtg accgagggaa tgagaaagcc	6720
	cgcttctctg agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg	6780
	gaaagtgacc gtgaagcagc tgaagagga ctacttcaag aaaatcgagt gcttcgactc	6840
	cgtggaaatc tccggcgtgg aagatcgggt caacgcctcc ctgggcacat accacgatct	6900
	gctgaaaatt atcaaggaca aggacttctt ggacaatgag gaaaacgagg acattctgga	6960
	agatatcgtg ctgacctga cactgtttga ggacagagag atgatcgagg aacggctgaa	7020
	aacctatgcc cacctgttcg acgacaaagt gatgaagcag ctgaaagcggc ggagatacac	7080
	cggctggggc aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccgcaa	7140
	gacaatcctg gatttctga agtccgacgg cttcgccaac agaaacttca tgcagctgat	7200
	ccacgacgac agcctgacct ttaaagagga catccagaaa gccaggtgt cggccaggg	7260
	cgatagcctg cacgagcaca ttgccaatct ggccggcagc cccgccatta agaaggcat	7320
	cctgcagaca gtgaaggtgg tggacgagct cgtgaaagt atgggcccgc acaagcccga	7380
	gaacatcgtg atcgaaatgg ccagagagaa ccagaccacc cagaagggac agaagaacag	7440
	ccgcgagaga atgaagcggc tcgaagaggg catcaaagag ctgggcagcc agatcctgaa	7500
	agaacacccc gtggaaaaca cccagctgca gaacgagaag ctgtacctgt actacctgca	7560
	gaatgggccc gatatgtacg tggaccagga actggacatc aaccggctgt ccgactacga	7620
	tgtggacat atcgtgcctc agagctttct gaaggacgac tccatcgaca acaaggtgct	7680

	gaccagaagc	gacaagaacc	ggggcaagag	cgacaacgtg	ccctccgaag	aggctcgtgaa	7740
	gaagatgaag	aactactggc	ggcagctgct	gaacgccaaag	ctgattaccc	agagaaaagt	7800
	cgacaatctg	accaaggccg	agagaggcgg	cctgagcгаа	ctggataagg	ccggcttcat	7860
	caagagacag	ctggtggaaa	cccggcagat	cacaaagcac	gtggcacaga	tcctggactc	7920
	ccggatgaac	actaagtacg	acgagaatga	caagctgatc	cgggaaagtga	aagtgatcac	7980
	cctgaagtcc	aagctggtgt	ccgatttccg	gaaggatttc	cagttttaca	aagtgcgcga	8040
	gatcaacaac	taccaccacg	cccacgacgc	ctacctgaac	gccgtcgtgg	gaaccgcct	8100
	gatcaaaaag	taccctaagc	tggaaagcga	gttcgtgtac	ggcgactaca	agggtgtacga	8160
	cgtgcggaag	atgatcgcca	agagcgagca	ggaaatcggc	aaggctaccg	ccaagtactt	8220
	cttctacagc	aacatcatga	actttttcaa	gaccgagatt	accctggcca	acggcgagat	8280
	ccggaagcgg	cctctgatcg	agacaaacgg	cgaaacgggg	gagatcgtgt	gggataaggg	8340
	ccgggatttt	gccaccgtgc	gaaaagtgct	gagcatgccc	caagtgaata	tcgtgaaaaa	8400
	gaccgaggtg	cagacaggcg	gcttcagcaa	agagtctatc	ctgcccaaga	ggaacagcga	8460
	taagctgata	gccagaaaga	aggactggga	ccctaagaag	tacggcggct	tcgacagccc	8520
	caccgtggcc	tattctgtgc	tgggtggtgc	caaagtggaa	aaggccaagt	ccaagaaact	8580
	gaagagtgtg	aaagagctgc	tggggatcac	catcatggaa	agaagcagct	tcgagaagaa	8640
	tcccatcgac	tttctggaag	ccaagggcta	caaagaagtg	aaaaaggacc	tgatcatcaa	8700
	gctgcctaag	tactccctgt	tcgagctgga	aaacggccgg	aagagaatgc	tggcctctgc	8760
	cggcgaactg	cagaagggaa	acgaactggc	cctgccctcc	aaatatgtga	acttctctga	8820
	cctggccagc	cactatgaga	agctgaaggg	ctccccgag	gataatgagc	agaaacagct	8880
	gtttgtgga	cagcacaagc	actacctgga	cgagatcadc	gagcagatca	gcgagtcttc	8940
[0012]	caagagagtg	atcctggccg	acgctaactc	ggacaaagtg	ctgtccgct	acaacaagca	9000
	ccgggataag	cccacagag	agcaggccga	gaatatcadc	cacctgttta	ccctgaccaa	9060
	tctgggagcc	cctgccgct	tcaagtactt	tgacaccacc	atcgaccgga	agaggtacac	9120
	cagcaccaaa	gaggtgctgg	acgccaccct	gatccaccag	agcatcaccg	gcctgtacga	9180
	gacacggata	gacctgtctc	agctgggagg	cgacaagcga	cctgccgcca	caaagaaggc	9240
	tggacaggct	aagaagaaga	aagattacaa	agacgatgac	gataagggat	ccggcgcaac	9300
	aaacttctct	ctgctgaaac	aagccggaga	tgctgaagag	aatcctggac	cgaccgagta	9360
	caagcccacg	gtgcgcctcg	ccaccgcgca	cgacgtcccc	agggccgtac	gcaccctcgc	9420
	cgcccgcttc	gccgactacc	ccgccacgcg	ccacaccgtc	gatccggacc	gccacatcga	9480
	gcgggtcacc	gagctgcaag	aactcttctc	cacgcgcgtc	gggctcgaca	tcggcaaggt	9540
	gtgggtcgcg	gacgacggcg	ccgcggtggc	ggtctggacc	acgccggaga	gcgtcgaagc	9600
	ggggcggtg	ttcgccgaga	tcggcccgcg	catggccgag	ttgagcggtt	cccggctggc	9660
	cgcgacgcaa	cagatggaag	gcctcctggc	gccgcaccgg	cccaaggagc	ccgcgtggtt	9720
	cctggccacc	gtcggagtct	cgcccaccca	ccagggaag	ggtctgggca	gcgccgtcgt	9780
	gctccccgga	gtggaggcgg	ccgagcgcgc	cggggtgccc	gccttctctg	agacctccgc	9840
	gccccgcaac	ctccccttct	acgagcggct	cggttcacc	gtcaccgccg	acgtcgaggt	9900
	gccccgaagga	ccgcgcacct	ggtgcatgac	ccgcaagccc	ggtgcctgaa	cgcgtaaagt	9960
	cgacaatcaa	cctctggatt	acaaaatttg	tgaagattg	actggtatc	ttaactatgt	10020
	tgctcctttt	acgetatgtg	gatacgtgct	tttaatgcct	ttgtatcatg	ctattgettc	10080
	ccgatggct	ttcattttct	cctccttgta	taaatcctgg	ttgctgtctc	tttatgagga	10140
	gttgtggccc	gttgtcaggc	aacgtggcgt	ggtgtgact	gtgtttgctg	acgcaacccc	10200
	cactggttgg	ggcattgcca	ccacctgtca	gctcctttcc	gggactttcg	ctttcccct	10260
	ccctattgcc	acggcggaac	tcatcgccgc	ctgccttggc	cgctgctgga	caggggctcg	10320

gctgttgggc actgacaatt ccgtggtgtt gtcggggaaa tcatcgtcct ttccttggct 10380
 gctcgcctgt gttgccacct ggattctgcg cgggacgtcc ttctgctacg tcccttcggc 10440
 cctcaatcca gcggaccttc ctccccggc cctgctgccg gctctgcggc ctcttcggcg 10500
 tcttcgcctt cgcctcaga cgagtcggat ctcccttgg gccgcctccc cgcgtcgact 10560
 ttaagaccaa tgacttaciaa ggcagctgta gatcttagcc actttttaa agaaaagggg 10620
 ggactggaag ggctaattca ctcccaacga agacaagatc tgccttttgc ttgtactggg 10680
 tctctctggt tagaccagat ctgagcctgg gagctctctg gctaactagg gaaccactg 10740
 cttaaagctc aataaagctt gccttgagt cttcaagtag tgtgtgcccc tctgttgtgt 10800
 gactctggta actagagatc cctcagacc ttttagtcag tgtggaaaat ctctagcagg 10860
 gcccgttaa acccgctgat cagcctcgac tgtgccttct agttgccagc catctgttgt 10920
 ttgcccctcc cccgtgcctt ccttgaccct ggaaggtgcc actcccactg tcccttcta 10980
 ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc tggggggtgg 11040
 ggtggggcag gacagcaagg gggaggattg ggaagacaat agcaggcatg ctggggatgc 11100
 ggtgggctct atggcttctg aggcggaaag aaccagctgg ggctctaggg ggtatcccca 11160
 cgcgccctgt agcggcgcac taagcgcggc ggggtgtggtg gttacgcgca gcgtgaccgc 11220
 tacacttgc agcgcctag cgcctcctc tttcgttct tcccttct tctcgcac 11280
 gttcggcgcc tttcccgtc aagctctaaa tcgggggctc ccttaggt tccgatttag 11340
 tgctttacgg cacctcgacc caaaaaact tgattagggt gatggtcac gtagtgggcc 11400
 atcgcctga tagacggtt ttcgccctt gacgttgag tccacgttct ttaatagtg 11460
 actctgttc caaactgaa caaactcaa cctatctc gtctattct ttgattata 11520
 agggattttg ccgatttcg cctattggt aaaaaatgag ctgatttaac aaaaattta 11580
 cgcaattaa ttctgtgaa tgtgtgctc ttaggtgtg gaaagtccc aggctccca 11640
 gcaggcagaa gtatgcaaa catgcatctc aattagtcag caaccagtg tggaaagtc 11700
 ccaggctccc cagcaggcag aagtatgcaa agcatgcat tcaattagtc agcaaccata 11760
 gtcccgcctc taactccgc catcccgc ctaactccgc ccagtccgc ccattctccg 11820
 ccccatggct gactaattt ttttattat gcagaggccg aggccgcctc tgcctctgag 11880
 ctattccaga agtagtgagg aggcctttt ggaggcctag gcttttgca aaagctcccg 11940
 ggagcttcta tatccatctt cggatctgat cagcacgtgt tgacaattaa tcatcgcat 12000
 agtatatcgg catagtataa tacgacaagg tgaggaacta aaccatggcc aagttgacca 12060
 gtgccgttcc ggtgctcacc gcgcgcgacg tcgccggagc ggtcagttc tggaccgacc 12120
 ggctcgggtt ctcccgggac ttcgtggagg acgacttcgc cgggtgtggtc cgggacgacg 12180
 tgacctgtt catcagcgcg gtccaggacc aggtggtgcc ggacaacacc ctggcctggg 12240
 tgtgggtgcg cggcctggac gagctgtacg ccgagtgtc ggaggtcgtg tccacgaact 12300
 tccgggacgc ctccgggccc gccatgaccg agatcggcga gcagccgtgg gggcgggagt 12360
 tcgccctgcg cgaccggccc ggcaactgcg tgcacttct ggccaggag caggactgac 12420
 acgtgctacg agatttcgat tccaccgccc cttctatga aaggttgggc ttcggaatcg 12480
 tttccggga cgcggctgg atgatctcc agcgcgggga tctcatctg gattctctc 12540
 cccaccccaa cttgtttatt gcagcttata atggttacia ataaagcaat agcatcacia 12600
 atttcaciaa taaagcattt ttttactgc attctagttg tggttgtcc aaactcatca 12660
 atgtatctta tcatgtctgt ataccgtcga cctctagcta gagcttggcg taatcatggt 12720
 catagctgtt tctgtgtga aattgttatc cgtcaciaa tccacacaac atacgagccg 12780
 gaagcataaa gtgtaaagcc tgggtgcct aatgagtgag ctaactcaca ttaattgcgt 12840
 tgcgtcact gcccgttctc cagtcgggaa acctgctgt ccagctgat taatgaatcg 12900
 gccaacgcgc ggggagaggc gtttgcgta ttgggcgctc ttcgcttcc tcgctcactg 12960

[0013]

[0014]

```

actcgtcgcg ctcggtcggt cggctgcggc gagcggatc agctcactca aaggcggtaa 13020
tacggttata cacagaatca ggggataacg caggaaagaa catgtgagca aaaggccagc 13080
aaaaggccag gaaccgtaaa aaggccgcgt tgctggcggt tttccatagg ctccgcccc 13140
ctgacgagca tcacaaaaat cgacgctcaa gtcagagggt gcgaaacccg acaggactat 13200
aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctggt ccgaccctgc 13260
cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt tctcatagct 13320
cacgctgtag gtatctcagt tcgggtgtagg tcgttcgctc caagctgggc tgtgtgcacg 13380
aaccccccg tccagcccagc cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc 13440
cggtaagaca cgacttatcg ccaactggcag cagccactgg taacaggatt agcagagcga 13500
ggtatgtagg cgggtctaca gaggttctga agtgggtggc taactacggc taccatagaa 13560
gaacagtatt tggatctcgc gctctgctga agccagttac cttcgaaaa agagttggtg 13620
gctcttgatc cggcaaaaa accaccgctg gtagcgggtg ttttttggt tgcaagcagc 13680
agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatctttct acggggctcg 13740
acgctcagtg gaacgaaac tcacgttaag ggattttggt catgagatta tcaaaaagga 13800
tcttcaccta gatcctttta aattaaat gaagtttta atcaatctaa agtatatatg 13860
agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc tcagcgatct 13920
gtctatttcg ttcattcata gttgcctgac tccccgctg gtagataact acgatacggg 13980
agggttacc atctggcccc agtgctgcaa tgataccgcg agaccacgc tcaccggctc 14040
cagatttata agcaataaac cagccagccg gaaggccga gcgcagaagt ggtcctgcaa 14100
ctttatccgc ctccatccag tctattaatt gttccggga agctagagta agtagttcgc 14160
cagttaatag tttgcgcaac gttggtgcca ttgctacagg catcgtggtg tcacgctcgt 14220
cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt acatgatccc 14280
ccatgttggt caaaaaagcg gttagctcct tcggctctcc gatcgttgtc agaagtaagt 14340
tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt actgtcatgc 14400
catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc tgagaatagt 14460
gtatcggcgc accgagttgc tcttgcccgg cgtaataac ggataatacc gcgccacata 14520
gcagaacttt aaaagtgctc atcattggaa aacgttcttc ggggcgaaaa ctctcaagga 14580
tcttaccgct gttgagatcc agttcagatg aaccactcgc tgcacccaac tgatcttcag 14640
catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa 14700
aaaagggaat aaggcgacga cggaaatgtt gaatactcat actcttcctt tttcaatatt 14760
attgaagcat ttatcagggt tattgtctca tgagcggata catattgaa tgtatttaga 14820
aaaataaaca aatagggtt cgcgcacat ttccccgaaa agtgccacct gac 14873

```

```

<210> 51
<211> 25
<212> DNA
<213> 人工序列
<400> 51
caccgtttcc aatccgacct cgaag 25

```

```

<210> 52
<211> 25
<212> DNA
<213> 人工序列

```

	<400> 52	
	aaaccttcgg gtcggattg gaaac	25
	<210> 53	
	<211> 25	
	<212> DNA	
	<213> 人工序列	
	<400> 53	
	caccggagcg aggccattc ggggt	25
	<210> 54	
	<211> 25	
	<212> DNA	
	<213> 人工序列	
[0015]	<400> 54	
	aaacaccccg aagtgccctc gctcc	25
	<210> 55	
	<211> 28	
	<212> DNA	
	<213> 人工序列	
	<400> 55	
	gagccctct atgattaatc gcaatgca	28
	<210> 56	
	<211> 23	
	<212> DNA	
	<213> 人工序列	
	<400> 56	
	gtgggtccaa gtgtcgtga gtg	23

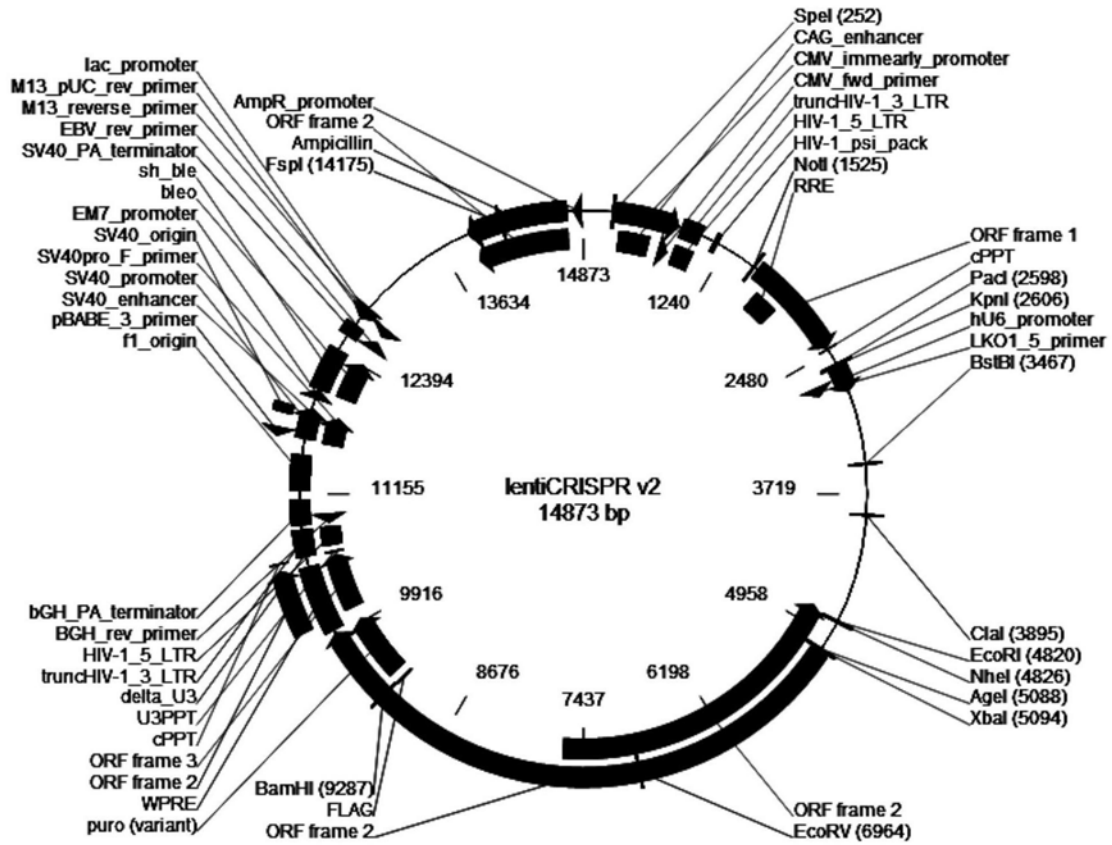


图1

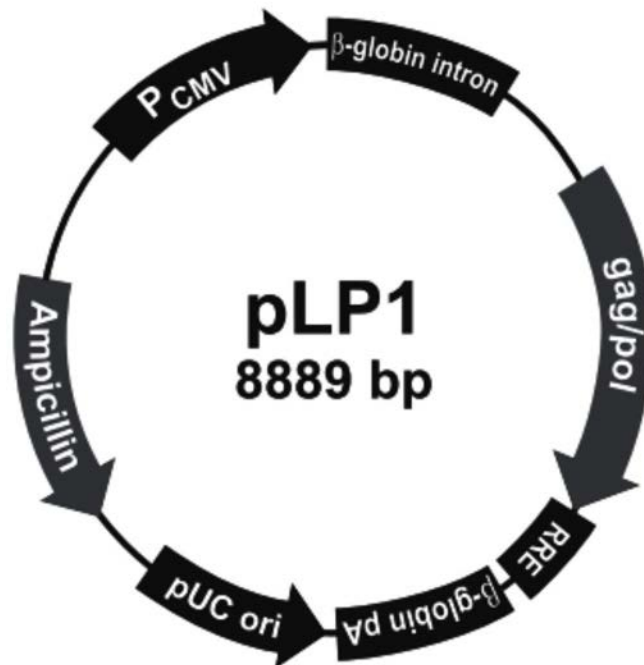


图2

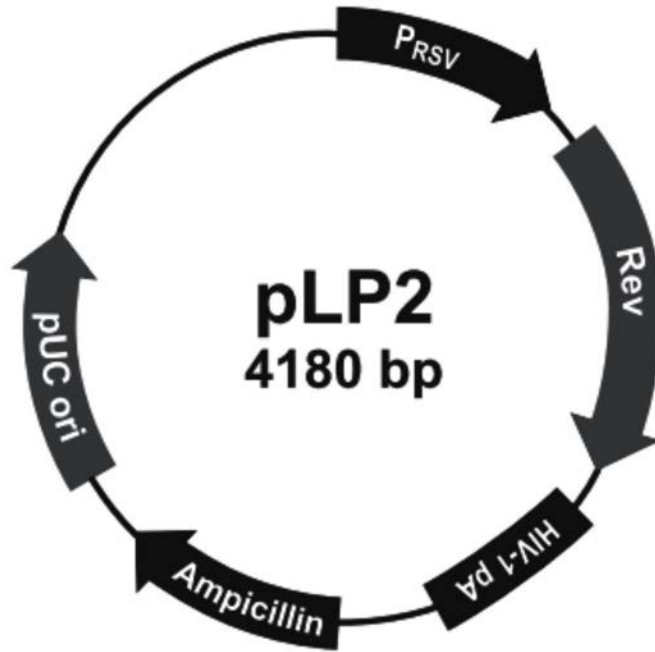


图3

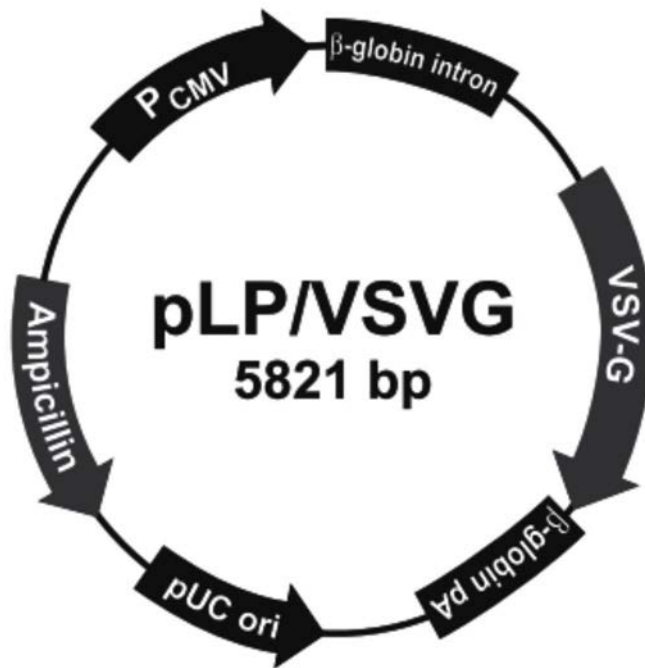


图4

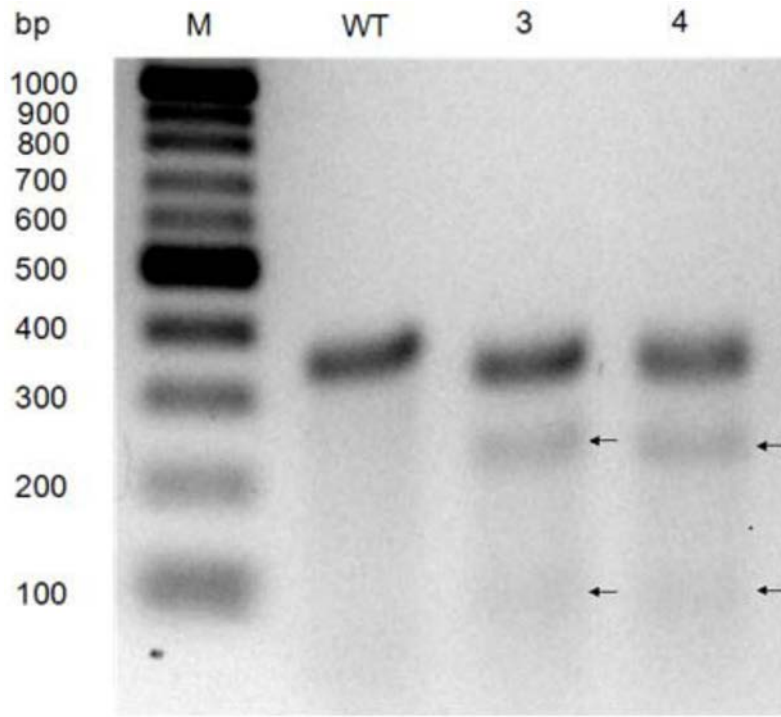


图5