



(12)发明专利申请

(10)申请公布号 CN 109517841 A

(43)申请公布日 2019.03.26

(21)申请号 201811478185.1

(22)申请日 2018.12.05

(71)申请人 华东师范大学

地址 200333 上海市普陀区中山北路3663

号华东师范大学

申请人 上海邦耀生物科技有限公司

(72)发明人 李大力 张晓辉 陈亮 刘明耀

(74)专利代理机构 北京超凡志成知识产权代理

事务所(普通合伙) 11371

代理人 周文波

(51)Int.Cl.

C12N 15/85(2006.01)

C12N 15/10(2006.01)

权利要求书1页 说明书6页

序列表12页 附图3页

(54)发明名称

一种用于核苷酸序列修饰的组合物、方法与应用

(57)摘要

本发明公开了一种用于核苷酸序列修饰的组合物、方法与应用,涉及基因编辑技术领域。该组合物包括第一载体和第二载体;其中,第一载体上具有如下表达元件:胞嘧啶脱氨酶表达元件、腺嘌呤脱氨酶表达元件以及突变型Cas酶表达元件;第二载体上具有如下表达元件:gRNA表达元件和尿嘧啶糖苷酶抑制剂表达元件。采用该组合物,可以对目标核苷酸序列上的位于PAM上游的第1-16位的C碱基进行修饰,使其发生C/G到T/A的转变。相较于现有的基因编辑技术,该组合物和方法具有更宽广的工作窗口,同时具备不引入DSB、indels及脱靶效应极低、更加安全等优点。



1. 一种用于核苷酸序列修饰的组合物,其特征在于,其包括:第一载体和第二载体;
其中,所述第一载体上具有如下表达元件:
胞嘧啶脱氨酶表达元件、腺嘌呤脱氨酶表达元件以及突变型Cas酶表达元件;
所述第二载体上具有如下表达元件:
gRNA表达元件和尿嘧啶糖苷酶抑制剂表达元件。
2. 根据权利要求1所述的组合物,其特征在于,腺嘌呤脱氨酶为野生腺嘌呤脱氨酶、突变型腺嘌呤脱氨酶或其二者的组合。
3. 根据权利要求2所述的组合物,其特征在于,突变型Cas酶为SpCas9n、VQR-Cas9n、Cpf1、SaCas9或其突变体。
4. 根据权利要求1-3任一项所述的组合物,其特征在于,所述gRNA的靶序列选自SEQ ID NO.1-5。
5. 根据权利要求1-3任一项所述的组合物,其特征在于,野生型腺嘌呤脱氨酶的氨基酸序列如SEQ ID NO.6的第199-364位所示。
6. 根据权利要求1-3任一项所述的组合物,其特征在于,突变型腺嘌呤脱氨酶的氨基酸序列如SEQ ID NO.6的第397-562位所示。
7. 根据权利要求1-3任一项所述的组合物,其特征在于,胞嘧啶脱氨酶的氨基酸序列如SEQ ID NO.6的第1-182位所示。
8. 根据权利要求1-3任一项所述的组合物,其特征在于,所述突变型Cas酶为SpCas9n,其氨基酸序列如SEQ ID NO.6的第595-1961位所示。
9. 一种修饰核苷酸序列的方法,其特征在于,其包括:
给予对象施加权利要求1-8任一项所述的组合物。
10. 权利要求1-8中任一项所述的用于核苷酸序列修饰的组合物在基因突变、基因修复、构建有基因突变导致的疾病动物模型、基因治疗、基因功能筛选、药物筛选或疾病诊断中应用。

一种用于核苷酸序列修饰的组合物、方法与应用

技术领域

[0001] 本发明涉及基因编辑技术领域,具体而言,涉及一种用于核苷酸序列修饰的组合物、方法与应用。

背景技术

[0002] 自2013年以来,以CRISPR/Cas9为代表的新一代基因编辑技术进入生物学领域的各个实验,正改变着传统的基因操作手段。

[0003] 2016年4月,David R Liu实验室首次报道了基于大鼠胞嘧啶脱氨酶(Apobec1) CRISPR/Cas9融合而成的单碱基基因编辑技术(cytosine base editor,CBE)(包括BE1, BE2, BE3)在基因组上实现单个碱基C/G到T/A转变的基因组定点编辑。其中,BE3因其高效性而被广泛用于基因组的基因突变或修复,疾病动物模型制作,基因治疗,基因功能筛选等。经典的CBE系统(指BE3)是基于spCas9改造而来,除了识别PAM(例如NGG)外,还存在一个“工作窗口”,即它距离PAM远端数起4-7位,且在5-7位比较高效。

[0004] 已报道的文献,表明它的工作窗口受胞嘧啶脱氨酶的影响,他们也通过胞嘧啶脱氨酶功能域进行氨基酸突变进而筛选到编辑窗口能精确到1-2碱基的单碱基工具,其中以YE1(W90Y+R126E)为最优,保持与BE3相似编辑活性的同时,将编辑窗口精确到1-2个碱基,即可以靶向5-6位的胞嘧啶。后来,David R Liu也通过改变胞嘧啶脱氨酶与Cas9n(D10A)之间的不同的linker—GGG, (GGG)3, XTEN, (GGG)7,实验发现通过改变linker长度并未改变其工作窗口。其后David R Liu通过优化胞嘧啶脱氨酶与Cas9n(D10A)之间的linker长度为32AA,同时额外融合2个尿嘧啶糖苷酶抑制剂,提高了平均约1.5倍的C/G到T/A的编辑效率,同时提高了平均约2.3倍的产物纯度(降低非C到T的突变的效率),然而,其工作窗口依然为3-8,并未发生改变。其后,再无试图通过胞嘧啶脱氨酶与Cas9n(D10A)之间的linker改变其窗口的报道。而后改造的AID-BE3(见图1),即利用人源的胞嘧啶脱氨酶也同样可以实现3-8位C/G到T/A的突变,且5-7位效率最高,略高于BE3。因此,AID介导的Base editor要比Apobec1介导的Base editor具有相对较广且高效的工作窗口。然而,对于靶点20bp范围内,3-8以外的碱基,还不能通过spCas9介导的CBE靶向。而近期报道的BE-Plus是通过10×GCN4募集多个与Scfv融合表达的胞嘧啶脱氨酶,也能靶向4-16位的胞嘧啶,其高效工作窗口为7-13位C,实现了其高效工作窗口的平移,但其方法较为繁琐。这极大地限制了单碱基基因编辑系统的应用范围。

[0005] 鉴于此,特提出本发明。

发明内容

[0006] 本发明的目的在于提供一种用于核苷酸序列修饰的组合物,采用该组合物,可以对目标核苷酸序列上的位于PAM上游的第1-16位的C碱基进行修饰,使其发生C/G到T/A的转变。

[0007] 本发明的另一目的在于提供一种修饰核苷酸序列的方法,采用该方法,可以对目

标核苷酸序列上的位于PAM上游的第1-16位的C碱基进行修饰,使其发生C/G到T/A的转变,相较于现有的基因编辑技术,该方法具有更宽广的工作窗口。

[0008] 本发明的另一目的在于提供上述的组合物在基因修饰中的应用。

[0009] 本发明是这样实现的:

[0010] 一方面,本发明提供了一种用于核苷酸序列修饰的组合物,其包括:第一载体和第二载体;

[0011] 其中,所述第一载体上具有如下表达元件:

[0012] 胞嘧啶脱氨酶表达元件、腺嘌呤脱氨酶表达元件以及突变型Cas酶表达元件;

[0013] 所述第二载体上具有如下表达元件:

[0014] gRNA表达元件和尿嘧啶糖苷酶抑制剂表达元件。

[0015] 本发明通过在胞嘧啶脱氨酶和突变型Cas酶之间插入腺嘌呤脱氨酶,使得在组合物可以对目标核苷酸序列上的位于PAM上游的第1-16位的C碱基进行修饰,使其发生C/G到T/A的转变。相较于现有的基因编辑系统,本发明提供的组合物具有更加宽广的工作窗口。可以对PAM上游的更大范围内的C碱基进行修饰,使其发生C/G到T/A的转变。同时具备不引入DSB、插入、缺失及脱靶效应极低、更加安全的优点。该组合物可以用于基因突变、基因修复、构建有基因突变导致的疾病动物模型、基因治疗、基因功能筛选、药物筛选和疾病诊断等需要进行C/G到T/A的转变的核苷酸修饰的领域。

[0016] 胞嘧啶脱氨酶可以对目标核苷酸序列上位于PAM上游的胞嘧啶(C)脱氨,形成尿嘧啶(U),在尿嘧啶糖苷酶抑制剂存在的条件下,随着基因组的复制,原位于目标核苷酸序列上的胞嘧啶(C)的位点突变为胸腺嘧啶(T),进而实现在该位点的C/G到T/A的突变修饰,实现基因编辑的效果。

[0017] 进一步地,在本发明的一些实施方案中,腺嘌呤脱氨酶为野生腺嘌呤脱氨酶、突变型腺嘌呤脱氨酶或其二者的组合。

[0018] 腺嘌呤脱氨酶表达元件可以是仅表达野生腺嘌呤脱氨酶的编码序列或者是仅表达突变型腺嘌呤脱氨酶的编码序列,可以是同时表达野生腺嘌呤脱氨酶和突变型腺嘌呤脱氨酶的二聚体编码序列。

[0019] 当然,无论是何种类型的腺嘌呤脱氨酶表达元件,其数量可以是一个或者是多个。

[0020] 当然,腺嘌呤脱氨酶可以是人源的,或者是其他非人的动物例如小鼠、大鼠、马、兔、猴、猿等来源的。

[0021] 进一步地,在本发明的一些实施方案中,突变型Cas酶为SpCas9n、VQR-Cas9n、SaCas9或其突变体。

[0022] 突变型Cas酶的作用在于对双链核苷酸中的一条单链形成切口,不具有对双链切割形成DSB的活性,利用其形成单链切口的特性,可以使得胞嘧啶脱氨酶发挥碱基修饰作用。

[0023] 突变型Cas酶的来源可以是来自酿酒酵母的SpCas9n,或者是识别PAM为NGAN的VQR-Cas9n,也可以是金黄色葡萄球菌的中saCas9或其突变体,其识别PAM为NNGRRT或NNNRRT,也可以是Cpf1或其它类型的突变体。只要其具有单链切口形成的活性不具有对双链切割形成DSB的活性即可。

[0024] 进一步地,在本发明的一些实施方案中,所述gRNA的靶序列选自SEQ ID NO.1-5。

[0025] SEQ ID NO.1-4所示的靶序列人的内源基因PD-1的4个靶序列。SEQ ID NO.5所示的靶序列人的内源基因KCNS1的靶序列。采用该组合物,可以对PD-1基因上SEQ ID NO.1-4所示的靶序列和SEQ ID NO.5所示的靶序列位于PAM上游的第1-16位范围内的C碱基进行修饰。

[0026] 应当理解到,除了上述的SEQ ID NO.1-5所示的靶序列外,本来领域技术人员可以采用本发明的组合物,通过设计不同的gRNA针对任何感兴趣的靶序列进行修饰,均属于本发明的保护范围。

[0027] 进一步地,在本发明的一些实施方案中,野生型腺嘌呤脱氨酶的氨基酸序列如SEQ ID NO.6的第199-364位所示。

[0028] 进一步地,在本发明的一些实施方案中,突变型腺嘌呤脱氨酶的氨基酸序列如SEQ ID NO.6的第397-562位所示。

[0029] 进一步地,在本发明的一些实施方案中,胞嘧啶脱氨酶的氨基酸序列如SEQ ID NO.6的第1-182位所示。

[0030] 进一步地,在本发明的一些实施方案中,所述突变型Cas酶为SpCas9n,其氨基酸序列如SEQ ID NO.6的第595-1961位所示。

[0031] 另一方面,本发明提供了一种修饰核苷酸序列的方法,其包括:

[0032] 给予对象施加上述的组合物。

[0033] 另一方面,本发明提供了上述用于核苷酸序列修饰的组合物在基因突变、基因修复、构建有基因突变导致的疾病动物模型、基因治疗、基因功能筛选、药物筛选或疾病诊断中应用。

附图说明

[0034] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本发明的某些实施例,因此不应被看作是对范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

[0035] 图1为实施例1中的第一载体的部分表达元件的结构示意图。

[0036] 图2为实施例1中的第二载体的部分表达元件的结构示意图。

[0037] 图3为实施例2中的针对PD-1-sg6靶点的PCR产物测序峰图。

[0038] 图4为实施例2中的针对PD-1-sg7靶点的PCR产物测序峰图。

[0039] 图5为实施例2中的针对PD-1-sg8靶点的PCR产物测序峰图。

[0040] 图6为实施例2中的针对PD-1-sg10靶点的PCR产物测序峰图。

[0041] 图7为实施例2中的针对KCNS1-sg1靶点的PCR产物测序峰图。

具体实施方式

[0042] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将对本发明实施例中的技术方案进行清楚、完整地描述。实施例中未注明具体条件者,按照常规条件或制造商建议的条件进行。所用试剂或仪器未注明生产厂商者,均为可以通过市售购买获得的常规产品。

[0043] 以下结合实施例对本发明的特征和性能作进一步的详细描述。

[0044] 实施例1

[0045] 载体构建

[0046] (1) 第一载体的构建

[0047] 第一载体上的第一表达元件组如图1所示(图1中PW-CBE-AID),其包括有:胞嘧啶脱氨酶(AID)表达元件、野生型腺嘌呤脱氨酶(TadA)表达元件、突变型腺嘌呤脱氨酶(TadAE59A,其相较于TadA,其第59位的氨基酸残基由E突变为A)表达元件以及突变型Cas酶(SpCas9n)表达元件;各表达元件通过Linker连接。

[0048] 其中,AID的氨基酸序列如SEQ ID NO.6的第1-182位所示,Linker1的氨基酸序列如SEQ ID NO.6的第183-198位所示,TadA的氨基酸序列如SEQ ID NO.6的第199-364位所示,Linker2的氨基酸序列如SEQ ID NO.6的第365-396位所示,TadAE59A的氨基酸序列如SEQ ID NO.6的第397-562位所示,Linker3的氨基酸序列如SEQ ID NO.6的第563-594位所示,SpCas9n的氨基酸序列如SEQ ID NO.6的第595-1961位所示,NLS的氨基酸序列如SEQ ID NO.6的第1966-1972位所示。

[0049] 第一表达元件组的核酸序列如SEQ ID NO.7。

[0050] (2) 第二载体的构建

[0051] 第二载体上的第二表达元件组的结构参考图2,其包括有:gRNA表达元件、尿嘧啶糖苷酶抑制剂(UGI)表达元件以及通过2A即自剪接多肽连接的GFP荧光蛋白编码序列。其中,gRNA表达元件由U6启动子驱动。

[0052] 第二表达元件组的核酸序列如SEQ ID NO.8所示。其中,第1-241位为U6启动子,第268-343位为SgRNA scaffold序列(用于稳定sgRNA结构),第344-350位为U6终止子,第385-892位为CMV启动子序列,第930-950位为Sv40NLS序列,第1035-1268位为UGI序列,第1317-1370为T2A序列。该载体用BbsI酶切完后以后就可以将对应靶点的sgRNA序列连入。

[0053] 在其他的实施例中,gRNA的表达序列可以根据不同的靶序列进行设计。设计出的具有不同gRNA表达元件的第二载体分别与第一载体组合,即可形成用于对目标核苷酸序列修饰的组合物。

[0054] 上述第一表达元件组和第二表达元件组可以基因合成方法合成,并分别克隆至骨架载体Pcmv-BE3(购自addgene,#73021)和pCDNA3.1(购自addgene,#52535),或骨架载体上改造使其具有上述的表达元件组,以被驱动表达,进而对靶序列进行碱基修饰。

[0055] 实施例2

[0056] 验证实施例1的组合物的基因修饰的工作窗口

[0057] (1) 从NCBI下载人的基因PD-1,KCNS1,其中PD-1设计了4个靶点,KCNS1设计了1个靶点(如表-1,表中下划线为PAM),类似于CRISPR/Cas9靶点oligo设计策略,sgRNA以U6为启动子,需要G作为转录起始位点,在针对各靶点的正向oligo的5'端添加CACC,反向oligo是靶点的互补链,在5'端添加AAAC(见表2)。

[0058] 表1人的基因PD-1上的靶核苷酸序列

[0059]	靶点名称	序列 (5'-3')
	PD-1-sg6	<i>TCCAGGCATGCAGATCCCACAGG</i> (SEQ ID NO.1)
	PD-1-sg7	<i>TGCAGATCCCACAGGCGCCCTGG</i> (SEQ ID NO.2)
[0060]	PD-1-sg8	<i>ACGACTGGCCAGGGCGCCTGTGG</i> (SEQ ID NO.3)
	PD-1-sg10	<i>GGGCGGTGCTACAACCTGGGCTGG</i> (SEQ ID NO.4)
	KCNS1-sg1	<i>CACTGTGCCCCACCACCAGCAGG</i> (SEQ ID NO.5)

[0061] 表2不同靶点的正向和反向oligo的序列

[0062]

靶点名称	序列 (5'-3')
PD-1-sg6-up	CACCGTCCAGGCATGCAGATCCCAC
PD-1-sg6-dn	AAACGTGGGATCTGCATGCCTGGAC
PD-1-sg7-up	CACCGTGCAGATCCCACAGGCGCCC
PD-1-sg7-dn	AAACGGGCGCCTGTGGGATCTGCAC
PD-1-sg8-up	CACCGACGACTGGCCAGGGCGCCTG
PD-1-sg8-dn	AAACCAGGCGCCCTGGCCAGTCGTC
PD-1-sg10-up	CACCGGGCGGTGCTACAACCTGGGC
PD-1-sg10-dn	AAACGCCAGTTGTAGCACCGCCC
KCNS1-sg1-up	CACCGCACTGTGCCCCACCACCAGC
KCNS1-sg1-dn	AAACGCTGGTGGTGGGGCACAGTGC

[0063] 表3靶点的PCR鉴定引物

[0064]	靶点名称	序列 (5'-3')
	PD-1-sg6	F: GGAGTGAGTACGGTGTGCACCTGACCTGG GACAGTTTCCCTTC R: GAGTTGGATGCTGGATGGCCACCTACCTA AGAACCATCCTGGC
	PD-1-sg7	
	PD-1-sg8	
	PD-1-sg10	
	KCNS1-sg1	F: GGAGTGAGTACGGTGTGCAAAGGAGGAGG ACGTGGGCTTTAAC R: GAGTTGGATGCTGGATGGTGAAGATGATG GTGATGGGGAGTGC

[0065] (2) 将表2中的各靶点的正向和反向oligo退化后分别连接经BbsI酶切后实施例1

中的第二载体上。

[0066] (3) 将步骤(2)中得到的带有gRNA表达序列的第二载体分别与第一载体按250ng:500ng比例行混合,形成用于修饰基因组载体系统,命名为:PW-CBE-AID,共转293T细胞,其中以已经报道的基因编辑系统AID-BE3作为对照组,120h后,分选收集GFP细胞,提细胞基因组DNA,PCR(引物见表-3)扩增出含靶点约200bp,测序,若存在套峰(图3-图7),再按照HiTOM试剂盒的要求准备样品送深度测序,分析其突变效率。

[0067] 由图3可以看出,PW-CBE-AID可以使靶点PD-1-sg6上位于PAM上游的第2到第11位的C发生C到T的突变,且从峰图上看第7到11位C突变效率较高。相对于AID-BE3也可以实现第1到第11位发生C到T的突变,且从峰图上看第2到7位C突变效率较高。

[0068] 由图4可以看出,PW-CBE-AID可以使靶点PD-1-sg7上位于PAM上游的第3到第12位的C发生C到T的突变,且从峰图上看第8到12位C突变效率较高。相对于AID-BE3也可以实现第3到第9位发生C到T的突变,且从峰图上看第3到9位C突变效率较高。

[0069] 由图5可以看出,PW-CBE-AID可以使靶点PD-1-sg8上位于PAM上游的第2到第10位的C发生C到T的突变,且从峰图上看第5到10位C突变效率较高。相对于AID-BE3也可以实现第2到第9位发生C到T的突变,且从峰图上看第2到5位C突变效率较高。

[0070] 由图6可以看出,PW-CBE-AID可以使靶点PD-1-sg10上位于PAM上游的第4到第12位的C发生C到T的突变,且从峰图上看第9到12位C突变效率较高。相对于AID-BE3也可以实现第4到第12位发生C到T的突变,且从峰图上看第4位C突变效率较高。

[0071] 由图7可以看出,PW-CBE-AID可以使靶点KCNS1-sg1上位于PAM上游的第3到第14位的C发生C到T的突变,且从峰图上看第8到14位C突变效率较高。相对于AID-BE3也可以实现第3到第8位发生C到T的突变,且从峰图上看第3位到第8位C突变效率较高。

[0072] 因此,综上PW-AID-BE3的工作窗口为PAM上游的第1-14位中的C可随机产生突变,且9,10位效率最高,同时也保持着3-8位C可编辑到。相对于,AID-BE3的工作窗口为3-8位,且高效工作窗口为5-7,PW-AID-BE3的工作窗口更加宽广,同时其高活性工作窗口向靠近PAM处平移了4位。这样,PW-AID-BE3则极大可能编辑原来BE3或AID-BE3无法编辑到的位点。例如产生更多的终止密码子实现基因敲除或实现错义突变。

[0073] 以上所述仅为本发明的优选实施例而已,并不用于限制本发明,对于本领域的技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

SEQUENCE LISTING

<110> 华东师范大学、上海邦耀生物科技有限公司

<120> 一种用于核苷酸序列修饰的组合物、方法与应用

<160> 8

<170> PatentIn version 3.5

<210> 1

<211> 23

<212> DNA

<213> 人工序列

<400> 1

tccaggcatg cagatccac agg 23

<210> 2

<211> 23

<212> DNA

<213> 人工序列

<400> 2

tgcagatccc acaggcgccc tgg 23

<210> 3

<211> 23

<212> DNA

<213> 人工序列

<400> 3

acgactggcc agggcgccctg tgg 23

<210> 4

<211> 23

<212> DNA

<213> 人工序列

<400> 4

gggcggtgct acaactgggc tgg 23

<210> 5

<211> 23

<212> DNA

<213> 人工序列

<400> 5

cactgtgccc caccaccagc agg 23

<210> 6

<211> 1972

<212> PRT

<213> 人工序列

<400> 6

Met	Asp	Ser	Leu	Leu	Met	Asn	Arg	Arg	Lys	Phe	Leu	Tyr	Gln	Phe	Lys
1				5					10					15	
Asn	Val	Arg	Trp	Ala	Lys	Gly	Arg	Arg	Glu	Thr	Tyr	Leu	Cys	Tyr	Val
			20					25					30		
Val	Lys	Arg	Arg	Asp	Ser	Ala	Thr	Ser	Phe	Ser	Leu	Asp	Phe	Gly	Tyr
			35				40					45			
Leu	Arg	Asn	Lys	Asn	Gly	Cys	His	Val	Glu	Leu	Leu	Phe	Leu	Arg	Tyr
			50			55					60				
Ile	Ser	Asp	Trp	Asp	Leu	Asp	Pro	Gly	Arg	Cys	Tyr	Arg	Val	Thr	Trp
65					70				75						80
Phe	Thr	Ser	Trp	Ser	Pro	Cys	Tyr	Asp	Cys	Ala	Arg	His	Val	Ala	Asp
				85					90					95	
Phe	Leu	Arg	Gly	Asn	Pro	Asn	Leu	Ser	Leu	Arg	Ile	Phe	Thr	Ala	Arg
			100					105					110		
Leu	Tyr	Phe	Cys	Glu	Asp	Arg	Lys	Ala	Glu	Pro	Glu	Gly	Leu	Arg	Arg
			115				120					125			
Leu	His	Arg	Ala	Gly	Val	Gln	Ile	Ala	Ile	Met	Thr	Phe	Lys	Asp	Tyr
			130				135					140			
Phe	Tyr	Cys	Trp	Asn	Thr	Phe	Val	Glu	Asn	His	Glu	Arg	Thr	Phe	Lys
145					150				155						160
Ala	Trp	Glu	Gly	Leu	His	Glu	Asn	Ser	Val	Arg	Leu	Ser	Arg	Gln	Leu
				165					170					175	
Arg	Arg	Ile	Leu	Leu	Pro	Ser	Gly	Gly	Ser	Pro	Lys	Lys	Lys	Arg	Lys
			180					185					190		
Val	Gly	Ser	Ser	Gly	Ser	Ser	Glu	Val	Glu	Phe	Ser	His	Glu	Tyr	Trp
			195				200					205			
Met	Arg	His	Ala	Leu	Thr	Leu	Ala	Lys	Arg	Ala	Trp	Asp	Glu	Arg	Glu
			210				215				220				
Val	Pro	Val	Gly	Ala	Val	Leu	Val	His	Asn	Asn	Arg	Val	Ile	Gly	Glu
225					230				235					240	
Gly	Trp	Asn	Arg	Pro	Ile	Gly	Arg	His	Asp	Pro	Thr	Ala	His	Ala	Glu
				245					250					255	
Ile	Met	Ala	Leu	Arg	Gln	Gly	Gly	Leu	Val	Met	Gln	Asn	Tyr	Arg	Leu
			260					265					270		
Ile	Asp	Ala	Thr	Leu	Tyr	Val	Thr	Leu	Glu	Pro	Cys	Val	Met	Cys	Ala
			275					280					285		
Gly	Ala	Met	Ile	His	Ser	Arg	Ile	Gly	Arg	Val	Val	Phe	Gly	Ala	Arg

290	295	300
Asp Ala Lys Thr Gly	Ala Ala Gly Ser Leu Met	Asp Val Leu His His
305	310	315
Pro Gly Met Asn His	Arg Val Glu Ile Thr	Glu Gly Ile Leu Ala Asp
325	330	335
Glu Cys Ala Ala Leu	Leu Ser Asp Phe Phe	Arg Met Arg Arg Gln Glu
340	345	350
Ile Lys Ala Gln Lys	Lys Ala Gln Ser Ser	Thr Asp Ser Gly Gly Ser
355	360	365
Ser Gly Gly Ser Ser	Gly Ser Glu Thr Pro	Gly Thr Ser Glu Ser Ala
370	375	380
Thr Pro Glu Ser Ser	Gly Gly Ser Ser Gly	Gly Ser Ser Glu Val Glu
385	390	395
Phe Ser His Glu Tyr	Trp Met Arg His Ala	Leu Thr Leu Ala Lys Arg
405	410	415
Ala Arg Asp Glu Arg	Glu Val Pro Val Gly	Ala Val Leu Val Leu Asn
420	425	430
Asn Arg Val Ile Gly	Glu Gly Trp Asn Arg	Ala Ile Gly Leu His Asp
435	440	445
Pro Thr Ala His Ala	Ala Ile Met Ala Leu	Arg Gln Gly Gly Leu Val
450	455	460
Met Gln Asn Tyr Arg	Leu Ile Asp Ala Thr	Leu Tyr Val Thr Phe Glu
465	470	475
Pro Cys Val Met Cys	Ala Gly Ala Met Ile	His Ser Arg Ile Gly Arg
485	490	495
Val Val Phe Gly Val	Arg Asn Ala Lys Thr	Gly Ala Ala Gly Ser Leu
500	505	510
Met Asp Val Leu His	Tyr Pro Gly Met Asn	His Arg Val Glu Ile Thr
515	520	525
Glu Gly Ile Leu Ala	Asp Glu Cys Ala Ala	Leu Leu Cys Tyr Phe Phe
530	535	540
Arg Met Pro Arg Gln	Val Phe Asn Ala Gln	Lys Lys Ala Gln Ser Ser
545	550	555
Thr Asp Ser Gly Gly	Ser Ser Gly Gly Ser	Ser Gly Ser Glu Thr Pro
565	570	575
Gly Thr Ser Glu Ser	Ala Thr Pro Glu Ser	Ser Ser Gly Gly Ser Ser Gly
580	585	590
Gly Ser Asp Lys Lys	Tyr Ser Ile Gly Leu	Ala Ile Gly Thr Asn Ser
595	600	605

Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys			
610	615	620	
Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu			
625	630	635	640
Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg			
	645	650	655
Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile			
	660	665	670
Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp			
	675	680	685
Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys			
	690	695	700
Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala			
705	710	715	720
Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val			
	725	730	735
Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala			
	740	745	750
His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn			
	755	760	765
Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr			
	770	775	780
Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp			
785	790	795	800
Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu			
	805	810	815
Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly			
	820	825	830
Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn			
	835	840	845
Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr			
	850	855	860
Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala			
865	870	875	880
Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser			
	885	890	895
Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala			
	900	905	910
Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu			

915	920	925
Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe		
930	935	940
Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala		
945	950	955
Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met		
965	970	975
Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu		
980	985	990
Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His		
995	1000	1005
Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr		
1010	1015	1020
Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr		
1025	1030	1035
Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser		
1040	1045	1050
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro		
1055	1060	1065
Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser		
1070	1075	1080
Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu		
1085	1090	1095
Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val		
1100	1105	1110
Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg		
1115	1120	1125
Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp		
1130	1135	1140
Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys		
1145	1150	1155
Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile		
1160	1165	1170
Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His		
1175	1180	1185
Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu		
1190	1195	1200
Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu		
1205	1210	1215

Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala		
1220	1225	1230
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg		
1235	1240	1245
Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile		
1250	1255	1260
Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser		
1265	1270	1275
Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp		
1280	1285	1290
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly		
1295	1300	1305
Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser		
1310	1315	1320
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp		
1325	1330	1335
Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val		
1340	1345	1350
Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys		
1355	1360	1365
Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu		
1370	1375	1380
Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln		
1385	1390	1395
Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg		
1400	1405	1410
Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp		
1415	1420	1425
Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp		
1430	1435	1440
Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly		
1445	1450	1455
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys		
1460	1465	1470
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg		
1475	1480	1485
Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu		
1490	1495	1500
Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg		

1505	1510	1515
Gln Ile Thr Lys His Val	Ala Gln Ile Leu Asp	Ser Arg Met Asn
1520	1525	1530
Thr Lys Tyr Asp Glu Asn	Asp Lys Leu Ile Arg	Glu Val Lys Val
1535	1540	1545
Ile Thr Leu Lys Ser Lys	Leu Val Ser Asp Phe	Arg Lys Asp Phe
1550	1555	1560
Gln Phe Tyr Lys Val Arg	Glu Ile Asn Asn Tyr	His His Ala His
1565	1570	1575
Asp Ala Tyr Leu Asn Ala	Val Val Gly Thr Ala	Leu Ile Lys Lys
1580	1585	1590
Tyr Pro Lys Leu Glu Ser	Glu Phe Val Tyr Gly	Asp Tyr Lys Val
1595	1600	1605
Tyr Asp Val Arg Lys Met	Ile Ala Lys Ser Glu	Gln Glu Ile Gly
1610	1615	1620
Lys Ala Thr Ala Lys Tyr	Phe Phe Tyr Ser Asn	Ile Met Asn Phe
1625	1630	1635
Phe Lys Thr Glu Ile Thr	Leu Ala Asn Gly Glu	Ile Arg Lys Arg
1640	1645	1650
Pro Leu Ile Glu Thr Asn	Gly Glu Thr Gly Glu	Ile Val Trp Asp
1655	1660	1665
Lys Gly Arg Asp Phe Ala	Thr Val Arg Lys Val	Leu Ser Met Pro
1670	1675	1680
Gln Val Asn Ile Val Lys	Lys Thr Glu Val Gln	Thr Gly Gly Phe
1685	1690	1695
Ser Lys Glu Ser Ile Leu	Pro Lys Arg Asn Ser	Asp Lys Leu Ile
1700	1705	1710
Ala Arg Lys Lys Asp Trp	Asp Pro Lys Lys Tyr	Gly Gly Phe Asp
1715	1720	1725
Ser Pro Thr Val Ala Tyr	Ser Val Leu Val Val	Ala Lys Val Glu
1730	1735	1740
Lys Gly Lys Ser Lys Lys	Leu Lys Ser Val Lys	Glu Leu Leu Gly
1745	1750	1755
Ile Thr Ile Met Glu Arg	Ser Ser Phe Glu Lys	Asn Pro Ile Asp
1760	1765	1770
Phe Leu Glu Ala Lys Gly	Tyr Lys Glu Val Lys	Lys Asp Leu Ile
1775	1780	1785
Ile Lys Leu Pro Lys Tyr	Ser Leu Phe Glu Leu	Glu Asn Gly Arg
1790	1795	1800

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu		
1805	1810	1815
Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser		
1820	1825	1830
His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys		
1835	1840	1845
Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile		
1850	1855	1860
Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala		
1865	1870	1875
Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys		
1880	1885	1890
Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu		
1895	1900	1905
Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr		
1910	1915	1920
Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala		
1925	1930	1935
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile		
1940	1945	1950
Asp Leu Ser Gln Leu Gly Gly Asp Ser Gly Gly Ser Pro Lys Lys		
1955	1960	1965
Lys Arg Lys Val		
1970		

<210> 7

<211> 5919

<212> DNA

<213> 人工序列

<400> 7

```

atggacagcc tcttgatgaa ccggaggaag tttctttacc aattcaaaaa tgtccgctgg 60
gctaagggtc ggcgtgagac ctacctgtgc tacgtagtga agaggcgtga cagtgtctaca 120
tcctttttcac tggactttgg ttatcttcgc aataagaacg gctgccacgt ggaattgctc 180
ttcctccgct acatctcgga ctgggacctg gacctggcc gctgctaccg cgtcacctgg 240
ttcacctcct ggagccccctg ctacgactgt gcccgacatg tggccgactt tctgcgaggg 300
aaccccaacc tcagtctgag gatcttcacc gcgcgcctct acttctgtga ggaccgcaag 360
gctgagccccg aggggctgcg gcggtgcac cgcgccgggg tgcaaatagc catcatgacc 420
ttcaaagatt atttttactg ctggaatact tttgtagaaa accacgaaag aactttcaaa 480
gcctgggaag ggctgcatga aaattcagtt cgtctctcca gacagcttcg gcgcacatcctt 540
ttgccctctg gtggttctcc caagaagaag aggaaagtcg gtagttccgg atctagcgag 600

```


gtggagttca gccacgagta ctggatgaga cagccctga ccctggctaa gagagcttgg 660
 gatgagagag aggtgcccgt gggagctgtt ctggttcata acaacagggt gatcggcgag 720
 ggatggaaca gacctatcgg gagacacgac ccaaccgctc atgctgaaat catggccctg 780
 agacaaggag ggctggatgat gcaaaattac agactgatcg acgcaaccct gtacgtgacc 840
 ctggagcctt gtgtgatgtg cgcaggagca atgatccact ccagaatcgg cagagtgggtg 900
 ttcggagcta gagatgccaa aaccggagcc gctggaagcc tgatggacgt tctgcatcac 960
 cccggaatga atcacagagt ggagataacc gagggcattc tggccgacga gtgtgctgct 1020
 ctgctgtctg atttcttcag aatgagaagg caggaaatca aggcccagaa aaaggcccaa 1080
 agcagcaccg acagcggagg atctagcggg gatcaagcg gaagcgagac tcctggaacc 1140
 agcgaaaagc caaccccaga aagcagcggg ggaagtagcg gaggaagctc agaagtcgag 1200
 ttcagccatg agtattggat gagacatgct ctgaccctgg caaagagagc aagagacgag 1260
 agagaggtcc cagtgggagc agttctggtg ctgaacaaca gagtgatcgg ggaggggtgg 1320
 aacagagcaa tcggactgca cgaccctaca gcacacgcag ccataatggc actgagacaa 1380
 ggggggctcg tgatgcaaaa ctacaggtg atcgacgcca ccctgtacgt cacatttgag 1440
 ccctgtgtga tgtgtgccgg agccatgatt cacagtagaa tcggccgggt ggtgttcggt 1500
 gtgagaaacg ctaaaacagg cgccgccgga agcctgatgg atgttctgca ttaccccggc 1560
 atgaatcacc ggggtggagat cacagagggc atcctggctg acgaatgtgc cgctctgctg 1620
 tgttacttct tcagaatgcc ccgacaagtg ttcaacgccc agaagaaagc ccagtcaagc 1680
 accgactctg gcggatctag cgggtgatct agcggctctg agacccttg aacatccgaa 1740
 tccgccactc cagagagcag cggaggctct tctggaggat cagataaaaa gtattctatt 1800
 ggttttagcca tcggcactaa ttccgttgga tgggctgtca taaccgatga atacaaagta 1860
 ccttcaaaga aatttaaggt gttggggaac acagaccgtc attcgattaa aaagaatctt 1920
 atcggtgccc tcctattcga tagtggcgaa acggcagagg cgactcgctt gaaacgaacc 1980
 gctcggagaa ggtatacacg tcgcaagaac cgaatatgtt acttacaaga aatttttagc 2040
 aatgagatgg ccaaagttga cgattctttc tttcacctgt tggaagagtc cttccttgct 2100
 gaagaggaca agaaacatga acggcacccc atctttggaa acatagtaga tgaggtggca 2160
 tatcatgaaa agtacccaac gatttatcac ctcaaaaa agctagttga ctcaactgat 2220
 aaagcggacc tgaggttaat ctacttggct cttgcccata tgataaagtt ccgtgggcac 2280
 tttctcattg agggatgatc aaatccggac aactcggatg tcgacaaact gttcatccag 2340
 ttagtataaaa cctataatca gttgtttgaa gagaacccta taaatgcaag tggcgtggat 2400
 gcgaaggcta ttcttagcgc ccgcctctct aaatcccacg ggctagaaaa cctgatcgca 2460
 caattacccg gagagaagaa aaatgggttg ttcggtaacc ttatagcgtc ctactaggc 2520
 ctgacaccaa attttaagtc gaacttcgac ttagctgaag atgccaaatt gcagcttagt 2580
 aaggacacgt acgatgacga tctcgacaat ctactggcac aaattggaga tcagtatgcg 2640
 gacttatttt tggctgccaa aaaccttagc gatgcaatcc tcctatctga catactgaga 2700
 gtttaacttg agattaccaa ggccgcttta tccgcttcaa tgatcaaaag gtacgatgaa 2760
 catcaccaag acttgacact tctcaaggcc ctagtccgtc agcaactgcc tgagaaatat 2820
 aaggaaatat tctttgatca gtcgaaaaac gggtagcgag gttatattga cggcggagcg 2880
 agtcaagagg aattctacaa gtttatcaaa cccatattag agaagatgga tgggacggaa 2940

gagttgcttg taaaactcaa tcgcaagat ctactgcga agcagcggac tttcgacaac 3000
 ggtagcattc cacatcaa at ccacttaggc gaattgcatg ctatacttag aaggcaggag 3060
 gattttttatc cgttcctcaa agacaatcgt gaaaagattg agaaaatcct aacctttcgc 3120
 ataccttact atgtgggacc cctggcccga gggaactctc ggttcgcatg gatgacaaga 3180
 aagtccgaag aaacgattac tccatggaat tttgaggaag ttgtcgataa aggtgcgtca 3240
 gctcaatcgt tcatcgagag gatgaccaac tttgacaaga atttaccgaa cgaaaaagta 3300
 ttgcctaagc acagtttact ttacgagtat ttcacagtgt acaatgaact cacgaaagtt 3360
 aagtatgtca ctgagggcac gcgtaaaccc gcctttctaa gcggagaaca gaagaaagca 3420
 atagtagatc tgttattcaa gaccaaccgc aaagtgcag ttaagcaatt gaaagaggac 3480
 tactttaaga aaattgaatg cttcgattct gtgcgagatc cgggggtaga agatcgattt 3540
 aatgcgtcac ttggtacgta tcatgacctc ctaaagataa ttaaagataa ggacttcctg 3600
 gataacgaag agaataga tatcttagaa gatatagtgt tgactcttac cctctttgaa 3660
 gatcgggaaa tgattgagga aagactaaaa acatacgtc acctgttcga cgataagggt 3720
 atgaaacagt taaagaggcg tcgctatacg ggctggggac gattgtcgcg gaaacttata 3780
 aacgggataa gagacaagca aagtggtaaa actattctcg attttctaaa gagcgacggc 3840
 ttcgccaata ggaactttat gcagctgac catgatgact ctttaacctt caaagaggat 3900
 atacaaaagg cacaggtttc cggacaaggg gactcattgc acgaacatat tgcgaatctt 3960
 gctggttcgc cagccatcaa aaagggcata ctccagacag tcaaagtagt ggatgagcta 4020
 gttaaggta tgggacgtca caaacgggaa aacattgtaa tcgagatggc acgcgaaaat 4080
 caaacgactc agaaggggca aaaaaacagt cgagagcgga tgaagagaat agaagagggt 4140
 attaaagaac tgggcagcca gatcttaaag gagcatcctg tggaaaatac ccaattgcag 4200
 aacgagaaac ttacctcta ttacctaca aatggaaggg acatgtatgt tgatcaggaa 4260
 ctggacataa accgtttatc tgattacgac gtcgatcaca ttgtaccca atcctttttg 4320
 aaggacgatt caatcgacaa taaagtgtt acacgctcgg ataagaaccg agggaaggt 4380
 gacaatgttc caagcgagga agtcgtaaag aaaatgaaga actattggcg gcagctccta 4440
 aatgcgaaac tgataacgca aagaaagttc gataacttaa ctaaagctga gaggggtggc 4500
 ttgtctgaac ttgacaaggc cggatttatt aaacgtcagc tcgtggaaac ccgccaatc 4560
 acaaagcatg ttgcacagat actagattcc cgaatgaata cgaaatacga cgagaacgat 4620
 aagctgattc gggaagtcaa agtaatcact ttaaagtcaa aattggtgtc ggacttcaga 4680
 aaggattttc aattctataa agttagggag ataaataact accaccatgc gcacgacgct 4740
 tatcttaatg ccgtcgtagg gaccgcactc attaagaaat acccgaagct agaaagttag 4800
 tttgtgtatg gtgattacaa agtttatgac gtccgtaaga tgatcgcgaa aagcgaacag 4860
 gagataggca aggctacagc caaatacttc ttttattcta acattatgaa tttctttaag 4920
 acggaaatca ctctggcaaa cggagagata cgcaaacgac ctttaattga aaccaatggg 4980
 gagacaggtg aaatcgtatg ggataagggc cgggacttcg cgacggtgag aaaagttttg 5040
 tccatgcccc aagtcaacat agtaaagaaa actgaggtgc agaccggagg gttttcaag 5100
 gaatcgattc ttccaaaaag gaatagtgat aagctcatcg ctcgtaaaaa ggactgggac 5160
 ccgaaaaagt acgttggtt cgatagccct acagttgcct attctgtcct agtagtgga 5220
 aaagttgaga agggaaaaatc caagaaactg aagtcagtca aagaattatt ggggataacg 5280

attatggagc gctcgtcttt tgaaaagaac cccatcgact tccttgaggc gaaaggttac 5340
 aaggaagtaa aaaaggatct cataattaaa ctaccaaagt atagtctgtt tgagttagaa 5400
 aatggccgaa aacggatggt ggctagcgcc ggagagcttc aaaaggggaa cgaactcgca 5460
 ctaccgtcta aatacgtgaa tttcctgtat ttagcgtccc attacgagaa gttgaaaggt 5520
 tcacctgaag ataacgaaca gaagcaactt tttgttgagc agcacaacaa ttatctcgac 5580
 gaaatcatag agcaaatctt ggaattcagt aagagagtca tcctagctga tgccaatctg 5640
 gacaaagtat taagcgcata caacaagcac agggataaac ccatacgtga gcaggcggaa 5700
 aatattatcc atttggttac tcttaccac ctcggcgctc cagccgcatt caagtatttt 5760
 gacacaacga tagatcgcaa acgatacact tctaccaagg aggtgctaga cgcgacactg 5820
 attcaccaat ccatcacggg attatatgaa actcggatag atttgtcaca gcttgggggt 5880
 gactctggtg gttctcccaa gaagaagagg aaagtctaa 5919

<210> 8

<211> 2093

<212> DNA

<213> 人工序列

<400> 8

gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgttagagag 60
 ataattagaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga 120
 aagtaataat ttcttgggta gtttgacgtt ttaaaattat gttttaaaat ggactatcat 180
 atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga 240
 cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag ttaaaataag 300
 gctagtcctg tatcaacttg aaaaagtggc accgagtcgg tgcttttttt aggctgaat 360
 tctgcagata tccatcacac tggccgttac ataacttac gtaaattggc cgcctggctg 420
 accgccaac gacccccg ccttgacgtc aataatgacg tatgttccca tagtaacgcc 480
 aatagggact ttccattgac gtcaatgggt ggagtattta cggtaaactg ccacttggc 540
 agtacatcaa gtgtatcata tgccaagtac gccccctatt gacgtcaatg acggtaaatg 600
 gcccgcctgg cattatgccc agtacatgac cttatgggac tttcctactt ggcagtacat 660
 ctacgtatta gtcatcgcta ttaccatggt gatgcgggtt tggcagtaca tcaatgggag 720
 tggatagcgg tttgactcac ggggatttcc aagtctccac ccattgacg tcaatgggag 780
 tttgttttgg caccaaaaac aacgggactt tccaaaatgt cgtaacaact ccgccccatt 840
 gacgcaaatg ggcggtaggc gtgtacggtg ggaggtctat ataagcagag ctggtttagt 900
 gaaccgtcag atccgctagc gccaccatgc ccaagaagaa gaggaaagtc tcgagcgact 960
 acaaagacca tgacggtgat tataaagatc atgacatcga ttacaaggat gacgatgaca 1020
 agtctggtgg ttctactaat ctgtcagata ttattgaaaa ggagaccggt aagcaactgg 1080
 ttatccagga atccatcctc atgctcccag aggaggtgga agaagtcatt gggaacaagc 1140
 cggaaagcga tatactcgtg cacaccgcct acgacgagag caccgacgag aatgtcatgc 1200
 ttctgactag cgacgccct gaatacaagc cttgggctct ggtcatacag gatagcaacg 1260
 gtgagaacaa gattaagatg ctctctggtg gttctcccaa gaagaagagg aaagtcgagg 1320
 gcagaggaag tctgctaaca tgcggtgacg tcgaggagaa tcctggccca gtgagcaagg 1380

gcgaggagct gttcaccggg gtggtgcca tcctggtcga gctggacggc gacgtaaacg 1440
gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc cacctacggc aagctgaccc 1500
tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctc gtgaccaccc 1560
tgacctacgg cgtgcagtgc ttcagccgct accccgacca catgaagcag cagcacttct 1620
tcaagtccgc catgcccga ggctacgtcc aggagcgcac catcttcttc aaggacgacg 1680
gcaactacaa gacccgcgcc gaggtgaagt tcgagggcga caccctggtg aaccgcatcg 1740
agctgaaggg catcgacttc aaggaggacg gcaacatcct ggggcacaag ctggagtaca 1800
actacaacag ccacaacgtc tatatcatgg ccgacaagca gaagaacggc atcaaggtga 1860
acttcaagat ccgccacaac atcgaggacg gcagcgtgca gctcgccgac cactaccagc 1920
agaacacccc catcggcgac ggccccgtgc tgctgcccga caaccactac ctgagcaccc 1980
agtccgccct gagcaaagac cccaacgaga agcgcgatca catggtcctg ctggagttcg 2040
tgaccgccgc cgggatcact ctcggcatgg acgagctgta caaggaattc taa 2093

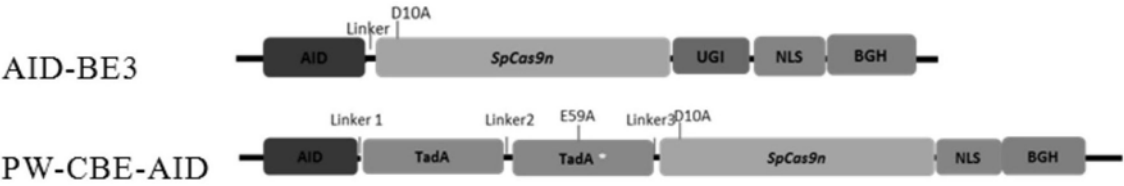


图1



图2

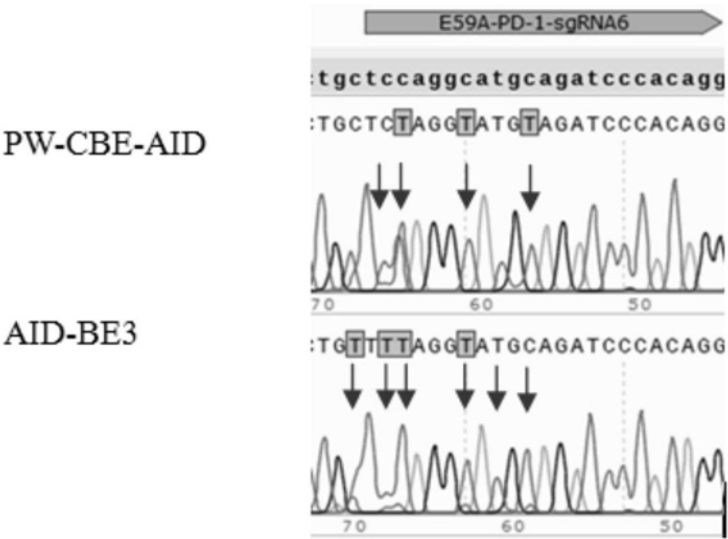


图3

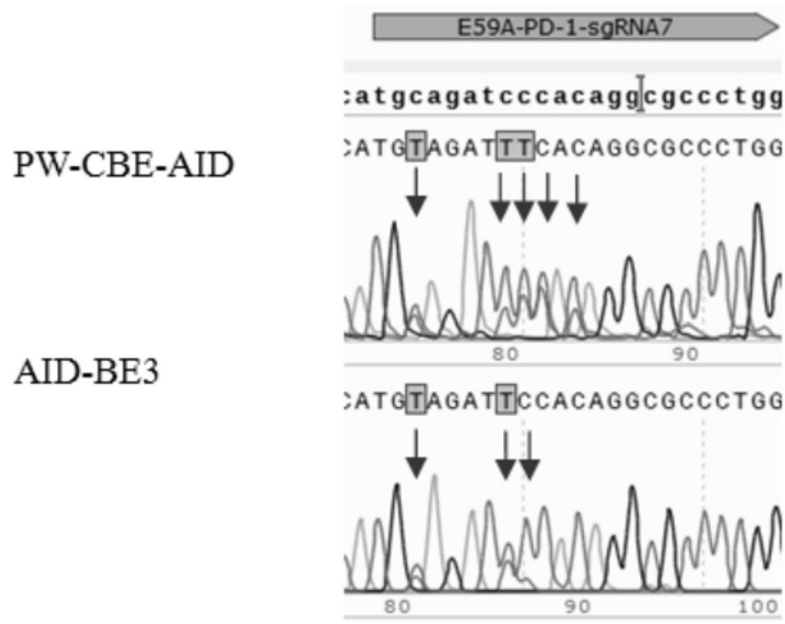


图4

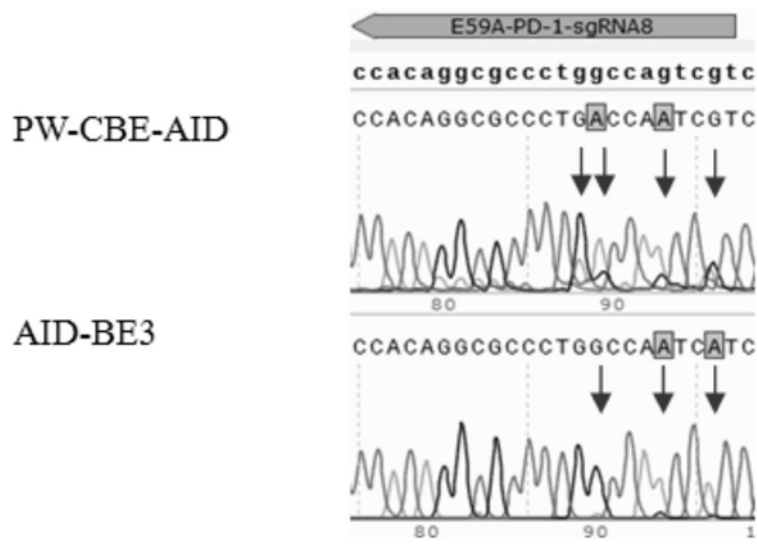


图5

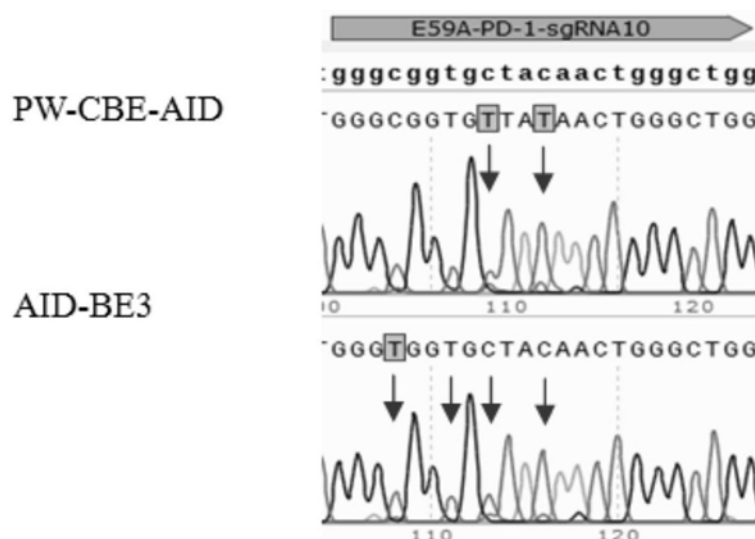


图6

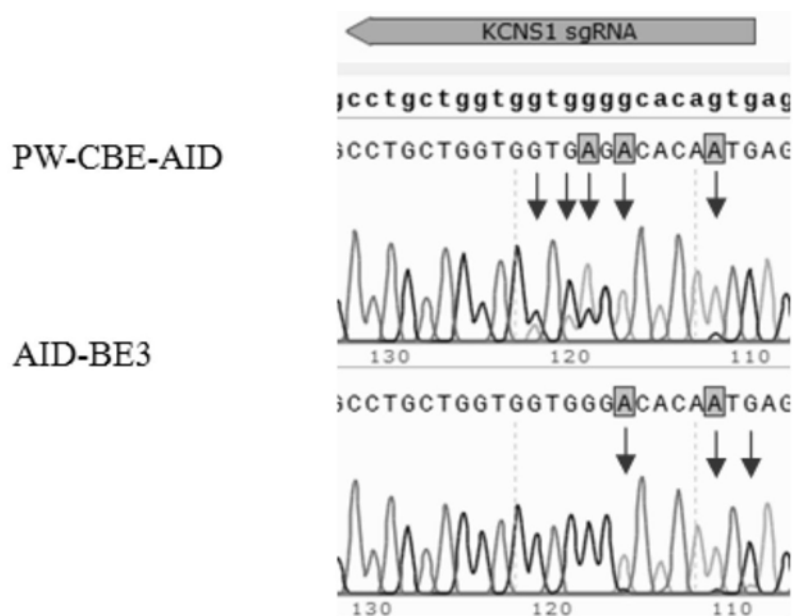


图7