



(12)发明专利申请

(10)申请公布号 CN 111304309 A

(43)申请公布日 2020.06.19

(21)申请号 202010152944.6

(22)申请日 2020.03.06

(71)申请人 上海韦翰斯生物医药科技有限公司

地址 201318 上海市浦东新区康新公路
3399弄25号534室

(72)发明人 林健 杨敬敏 覃振东 唐嘉婕
朱学萍

(74)专利代理机构 上海光华专利事务所(普通
合伙) 31219

代理人 朱凌娇 许亦琳

(51)Int.Cl.

C12Q 1/6869(2018.01)

G16B 30/10(2019.01)

权利要求书1页 说明书9页
序列表8页 附图1页

(54)发明名称

一种测序平台标签序列污染的检测方法

(57)摘要

本发明提供一种测序平台标签序列污染的检测方法,至少包括以下步骤:(1)将待测标签序列与已知序列连接,获得标签序列-已知序列,测序平台的标签序列种类固定且已知;(2)对步骤(1)获得的序列进行测序,获得测序结果,所述测序结果包括序列的碱基排列和序列的数量;(3)根据标签序列的种类的不同,对测序结果进行拆分,若一已知序列 r_m 除了在其对应的标签序列 T_m 的分类结果中出现外,还在其他标签序列 T_n 的分类结果中出现,则所述标签序列 T_m 被所述其他标签序列 T_n 污染;其中, m, n 为自然数,且 $m \neq n$ 。本发明通过验证标签序列的单一性,保证后续的下机数据的可信性。

1. 一种测序平台标签序列污染的检测方法,至少包括以下步骤:

(1) 将待测标签序列与已知序列连接,获得标签序列-已知序列,测序平台的标签序列种类固定且已知;

(2) 对步骤(1)获得的序列进行测序,获得测序结果,所述测序结果包括序列的碱基排列和序列的数量;

(3) 根据标签序列的种类的不同,对测序结果进行拆分,若一已知序列 r_m 除了在其对应的标签序列 T_m 的分类结果中出现外,还在其他标签序列 T_n 的分类结果中出现,则所述标签序列 T_m 被所述其他标签序列 T_n 污染;其中, m, n 为自然数,且 $m \neq n$ 。

2. 如权利要求1所述的测序平台标签序列污染的检测方法,其特征在于,步骤(1)中,所述已知序列源于任意物种基因组。

3. 如权利要求2所述的测序平台标签序列污染的检测方法,其特征在于,步骤(1)中,所述已知序列源于古菌或者噬菌体基因组。

4. 如权利要求3所述的测序平台标签序列污染的检测方法,其特征在于,所述噬菌体选自 λ 噬菌体。

5. 如权利要求1所述的测序平台标签序列污染的检测方法,其特征在于,还包括以下特征中的一项或多项:

1) 所述标签序列选自单端barcode、双端barcode;

2) 步骤(1)中,当待测标签序列的种类大于1种时,不同标签序列连接的已知序列不同。

6. 如权利要求1所述的测序平台标签序列污染的检测方法,其特征在于,所述检测方法还包括以下步骤:计算标签序列 T_m 被标签序列 T_n 污染的比例,所述比例采用以下方法计算:

已知序列 r_m 在标签序列 T_n 的分类结果中的序列数量/已知序列 r_m 在标签序列 T_m 的分类结果中的序列数量*100%。

7. 如权利要求1所述的测序平台标签序列污染的检测方法,其特征在于,所述已知序列的长度为100~250bp。

8. 如权利要求1所述的测序平台标签序列污染的检测方法,其特征在于,所述待测标签序列与已知序列通过PCR的方式或连接酶进行连接。

9. 如权利要求1所述的测序平台标签序列污染的检测方法,其特征在于,所述测序平台为Illumina二代测序平台。

10. 如权利要求1-9任一所述的测序平台标签序列污染的检测方法,在基因测序中的用途。

一种测序平台标签序列污染的检测方法

技术领域

[0001] 本发明涉及生物信息学和生物技术领域,特别是涉及一种测序平台标签序列污染的检测方法。

背景技术

[0002] 近些年,随着技术的发展和测序成本的下降,高通量测序已经由科研向着人们的日常生活渗透。目前,该技术主要有几方面的应用:基因组测序、RNA测序、DNA甲基化等,在肿瘤、遗传病、宏基因组等应用更加普遍。高通量测序的数据由标签序列将对应的样本分选,但是index序列是否单一、或者在操作过程中引入了其他污染导致标签序列不再纯粹,会导致后续的样本分选、生信分析出现误差。

发明内容

[0003] 鉴于以上所述现有技术的缺点,本发明的目的在于提供一种测序平台标签序列污染的检测方法。所述至少包括以下步骤:

[0004] (1) 将待测标签序列与已知序列连接,获得标签序列-已知序列,测序平台的标签序列种类固定且已知;

[0005] (2) 对步骤(1)获得的序列进行测序,获得测序结果,所述测序结果包括序列的碱基排列和序列的数量;

[0006] (3) 根据标签序列的种类的不同,对测序结果进行拆分,若一已知序列 r_m 除了在其对应的标签序列 T_m 的分类结果中出现外,还在其他标签序列 T_n 的分类结果中出现,则所述标签序列 T_m 被所述其他标签序列 T_n 污染;其中, m, n 为自然数,且 $m \neq n$ 。

[0007] 本发明还提供前述测序平台标签序列污染的检测方法,在基因测序中的用途。

[0008] 如上所述,本发明的检测测序平台标签序列污染的方法,具有以下有益效果:

[0009] 本发明利用单一的DNA序列,进行建库,上机测序,通过对下机数据中的已知序列对标签序列进行测试,保证标签序列无其他标签序列污染。通过验证标签序列的单一性,保证后续的下机数据的可信性。

附图说明

[0010] 图1已知序列PCR扩增的电泳结果。

具体实施方式

[0011] 以下通过特定的具体实例说明本发明的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本发明的精神下进行各种修饰或改变。

[0012] 在进一步描述本发明具体实施方式之前,应理解,本发明的保护范围不局限于下

述特定的具体实施方案;还应当理解,本发明实施例中使用的术语是为了描述特定的具体实施方案,而不是为了限制本发明的保护范围;在本发明说明书和权利要求书中,除非文中另外明确指出,单数形式“一个”、“一”和“这个”包括复数形式。

[0013] 当实施例给出数值范围时,应理解,除非本发明另有说明,每个数值范围的两个端点以及两个端点之间任何一个数值均可选用。除非另外定义,本发明中使用的所有技术和科学术语与本技术领域技术人员通常理解的意义相同。除实施例中使用的具体方法、设备、材料外,如本技术领域的技术人员对现有技术的掌握及本发明的记载,还可以使用与本发明实施例中所述的方法、设备、材料相似或等同的现有技术的任何方法、设备和材料来实现本发明。

[0014] 除非另外说明,本发明中所公开的实验方法、检测方法、制备方法均采用本技术领域常规的分子生物学、生物化学、染色质结构和分析、分析化学、细胞培养、重组DNA技术及相关领域的常规技术。

[0015] 本发明一实施例的测序平台标签序列污染的检测方法,至少包括以下步骤:

[0016] (1) 将待测标签序列与已知序列连接,获得标签序列-已知序列,测序平台的标签序列种类固定且已知;

[0017] (2) 对步骤(1)获得的序列进行测序,获得测序结果,所述测序结果包括序列的碱基排列和序列的数量;

[0018] (3) 根据标签序列的种类的不同,对测序结果进行拆分,若一已知序列 r_m 除了在其对应的标签序列 T_m 的分类结果中出现外,还在其他标签序列 T_n 的分类结果中出现,则所述标签序列 T_m 被所述其他标签序列 T_n 污染;其中, m, n 为自然数,且 $m \neq n$ 。

[0019] 所述标签序列是指一段核酸片段,用于在混合测序时,区分不同样本。

[0020] 所述已知序列是指碱基排列顺序已知的核酸片段。

[0021] 所述序列的数量是指核酸片段的条数。

[0022] 进一步的,检测测序平台标签序列是否被污染时,可以采用步骤(1)-(3),直接对标签序列进行检测。

[0023] 所述其他已知序列 r_n 可以是一种或多种。当所述其他已知序列 r_n 是多种时,表明所述标签序列 T_m 将多种标签序列污染。例如,标签序列 T_1 下除了含有所述标签序列 T_1 对应的已知序列 r_1 外,还含有其他已知序列 r_2 和 r_5 ,则表明,所述已知序列 r_1 对应的标签序列 T_1 将所述其他已知序列 r_2 对应的标签序列 T_2 和其他已知序列 r_5 对应的标签序列 T_5 污染。

[0024] 进一步的,步骤(1)中,所述已知序列源于任意物种基因组。

[0025] 在更进一步的实施方式中,步骤(1)中,所述已知序列源于古菌或者噬菌体基因组。古菌或者噬菌体的基因组为已知序列的来源。便于使用,防止存在多拷贝序列。

[0026] 优选的,在一次检测中,所述已知序列均来自同一种生物或同一个生物个体。能够最大限度的保证每个已知序列均不同。

[0027] 在一种实施方式中,所述噬菌体选自 λ 噬菌体。

[0028] 在一种实施方式中,所述古菌选自初古菌门(Korarchaeota)、纳古菌门(Nanoarchaeota)、奇古菌门(Thaumarchaeota)或广古菌门(Euryarchaeota)。例如热球菌属(Thermococcus)或热网菌属(Pyrodictium)或嗜盐杆菌(Halobacterium salinarum)。

[0029] 所述标签序列可以为单端barcode或双端barcode。

[0030] 本发明可以检测对一种标签序列或同时检测多种标签序列是否被污染。

[0031] 步骤(1)中,当待测标签序列的种类大于1种时,不同标签序列连接的已知序列不同。即在步骤1中,标签序列与已知序列需要一一对应连接,即一种待测标签序列只对应连接一种已知序列。

[0032] 进一步的,所述检测方法还包括以下步骤:计算标签序列 T_m 被标签序列 T_n 污染的比例,所述比例采用以下方法计算:

[0033] 已知序列 r_m 在标签序列 T_n 的分类结果中的序列数量/已知序列 r_m 在标签序列 T_m 的分类结果中的序列数量*100%。

[0034] 优选的,所述已知序列的长度为100~250bp。

[0035] 所述待测标签序列与已知序列可通过PCR的方式或连接酶进行连接。所述标签序列与已知序列可以直接相连或间接相连。间接相连时,可以通过引物序列进行连接。所述引物序列是指测序中的常用接头,如短Y接头或长Y接头。

[0036] 在一种实施方式中,多重建库的测序模式中,可通过PCR方法将标签序列与已知序列相连。

[0037] 在一种实施方式中,全基因组建库的测序模式中,如果采用短Y接头,先将标签序列与短Y接头连接,再通过PCR方法将已知序列连接到短Y接头上,获得标签序列-短Y接头-已知序列。

[0038] 在一种实施方式中,全基因组建库的测序模式中,如果采用长Y接头,所述长Y接头上已经含有标签序列,可以通过连接酶将长Y接头与已知序列连接,实现已知序列与标签序列的间接连接。

[0039] 可选的,所述测序平台为Illumina二代测序平台。

[0040] 前述的测序平台标签序列污染的检测方法可以用于基因测序。

[0041] 实施例1

[0042] 1.1用到的试剂:2×Taq Plus Master Mix (Dye Plus)、Ligation Module (翊圣)、DNA纯化磁珠 (翊圣)、2×Kapa Enzyme Mix、NextSeq 500/550Mid Output Reagent Cartridge V2、NextSeq Accessory Box V2、NextSeq 500/550Mid Output Flow Cell Cartridge V2、NextSeq 500/550Buffer Cartridge V2

[0043] 1.2片段扩增:用2×Taq Plus Master Mix (Dye Plus),以 λ 噬菌体DNA为模板,表1所列序列为引物,表2反应体系,表3反应条件分别进行DNA扩增。

[0044] 表1引物列表

[0045]

	F	SEQ ID	R	SEQ ID
r1	GCTGACATTTTCGGT	NO: 1	TGGCCTGCCGCAGTT	NO: 2
r2	CAGCCAGGAACTATT	NO: 3	GTTTTCCAGTCCGGA	NO: 4
r3	ATCCGTGAGGTGAAT	NO: 5	CAGCGACGGAATATC	NO: 6
r4	GATATTGAACAGGAA	NO: 7	TAAGATACTGCTCCT	NO: 8
r5	GTCATCCGCCAGCAG	NO: 9	AGTCTTTGACAATCT	NO: 10
r6	TATCGACTCCCAGCT	NO: 11	CATTCTGCACCATT	NO: 12
r7	TCCGTCTACGGAAAG	NO: 13	TCGGGAAGTGAACGG	NO: 14
r8	GACGCAATGAGGCAC	NO: 15	TCATCCTCTCCGGAT	NO: 16
r9	ATGACCTGATGACAG	NO: 17	ATACATAAAATCCTG	NO: 18
r10	GAATATGCCGGTTATC	NO: 19	CCTGATGCAGCTGGAT	NO: 20
r11	GAAGCGGCATGGAAAG	NO: 21	CTGACCATCCGGAAct	NO: 22
r12	TATTACGTCAGCGAG	NO: 23	TGCCCCGTCCTCCACGG	NO: 24
r13	CAGCGTGATGGAGCA	NO: 25	CCAATCCAGCCGGTCA	NO: 26
r14	TGCAGACGGCTCAGGA	NO: 27	AAAGTACGCCACGAC	NO: 28

[0046]

	F	SEQ ID	R	SEQ ID
r15	GAAAGAAGTTCAGGA	NO: 29	GATTCAAATGCTGCA	NO: 30

[0047] 表2反应体系

[0048]

PCR组分	体积(μL)
λDNA, 50ng/μL	1
正向引物, 10μM	1
反向引物, 10μM	1
2×Taq Plus Master Mix	25
水	22
Total	50

[0049] 表3反应条件

[0050]	循环数	温度	时间
	1cycle	96°C	2min
	30cycles	96°C	30s
		55°C	30s
		72°C	30s
	1cycle	72°C	5min
	1cycle	4°C	hold

[0051] PCR产物经琼脂糖凝胶电泳检测,由图1可见,检测扩增片段大小与预期目的片段大小一致。

[0052] 1.3 PCR产物的纯化:PCR产物使用翊圣DNA纯化磁珠1.4×样本体积纯化,80%乙醇洗涤两次,50μL TE洗脱,检测浓度。分别获得所述已知序列r1-r15。

[0053] 1.4 连接:以表5所示配制连接反应体系,表6所示反应条件,通过引物序列将待测标签序列与已知序列相连,所述引物序列的核苷酸序列为:

[0054] CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCCTTGGCACCCGAGAATTC, (SEQ ID NO:46),其中NNNNN代表标签序列,具体序列如表4所述。将引物序列与已知序列分别连接,即为将已知序列与待测标签序列T1-T15分别进行连接,获得标签序列-已知序列。具体的,为T1-r1,T2-r2,T3-r3,T4-r4,T5-r5,T6-r6,T7-r7,T8-r8,T9-r9,T10-r10,T11-r11,T12-r12,T13-r13,T14-r14,T15-r15。待测标签序列为barcode序列

[0055] 表4标签序列

标签序号	标签序列	SEQ ID
T1	TTATAT	NO:31
T2	ACCAAC	NO:32
T3	CTATGC	NO:33
T4	ATTCCT	NO:34
T5	CAACTC	NO:35
T6	TTAGGC	NO:36
T7	AGGATC	NO:37
T8	CAGCAA	NO:38
T9	AAGTAG	NO:39
T10	ACAGTG	NO:40
T11	GGTCCA	NO:41
T12	GATCAG	NO:42
T13	ATTATG	NO:43
T14	GGCTAC	NO:44
T15	GAACCT	NO:45

[0057] 表5连接体系

[0058]	试剂名称	体积/ μL
	翊圣连接buffer	10
	连接酶	2.5
	DNA (10ng/ μL)	3
	待测标签序列 (barcode序列)	2
	水	32.5
	total	50

[0059] 表6反应条件

[0060]	反应温度	反应时间
	4 $^{\circ}\text{C}$	Hold
	22 $^{\circ}\text{C}$	60min
	4 $^{\circ}\text{C}$	Hold

[0061] 连接结束后,连接产物使用翊圣DNA纯化磁珠1.0 \times 样本体积纯化,80%乙醇洗涤两次,

[0062] 12 μL TE洗脱;

[0063] 1.5文库扩增:将所有的标签序列-已知序列混合,以表7所示配制连接反应体系,表8所示反应条件,进行接头连接(P5/P7)。

[0064] 表7连接体系

[0065]	试剂名称	体积/ μL
	KAPA HiFi Mix	12.5
	P5/P7接头	2
	DNA (标签序列-已知序列)	10.5
	total	25

[0066] 表8反应条件

[0067]	循环数	温度	时间
		98 $^{\circ}\text{C}$	45s
	9cycles	98 $^{\circ}\text{C}$	15s
		60 $^{\circ}\text{C}$	30s
		72 $^{\circ}\text{C}$	30s
		72 $^{\circ}\text{C}$	1min
		4 $^{\circ}\text{C}$	hold

[0068] 扩增结束后,连接产物使用翊圣DNA纯化磁珠1.2 \times 样本体积纯化,80%乙醇洗涤两次,30 μL TE洗脱,得到需要检测的DNA文库;

[0069] 1.6质检

[0070] 用Qubit染料以及1.5%琼脂糖凝胶电泳测定扩增后的测序文库中的DNA浓度及片

段大小；

[0071] 1.7文库稀释:将构建好的文库稀释至10nM,以每个文库产生1M数据量进行混合；

[0072] 1.8上机测序:使用Illumina CN500二代测序平台对样本进行测序,测序完成后通过标签序列对barcode对其进行拆分,结果如表9、表10、表11所示；

[0073] 表9生信结果-1

标签序列	SEQ ID	上游引物	数量	已知序列名称
[0074] T2	NO: 1	GCTGACATTTTCGGT	0	r1
	NO: 3	CAGCCAGGAACTATT	5603	r2
	NO: 5	ATCCGTGAGGTGAAT	0	r3
	NO: 7	GATATTGAACAGGAA	0	r4
	NO: 9	GTCATCCGCCAGCAG	0	r5
	NO: 11	TATCGACTCCCAGCT	0	r6
	NO: 13	TCCGTCTACGGAAAG	0	r7

标签序列	SEQ ID	上游引物	数量	已知序列名称
[0075]	NO: 15	GACGCAATGAGGCAC	0	r8
	NO: 17	ATGACCTGATGACAG	0	r9
	NO: 19	GAATATGCCGGTTATC	0	r10
	NO: 21	GAAGCGGCATGGAAAG	0	r11
	NO: 23	TATTACGTCAGCGAG	0	r12
	NO: 25	CAGCGTGATGGAGCA	0	r13
	NO: 27	TGCAGACGGCTCAGGA	0	r14
	NO: 29	GAAAGAAGTTCAGGA	0	r15

[0076] 表10生信结果-2

标签序列	SEQ ID	上游引物	数量	已知序列名称	
[0077]	T14	NO: 1	GCTGACATTTTCGGT	0	r1
		NO: 3	CAGCCAGGAACTATT	0	r2
		NO: 5	ATCCGTGAGGTGAAT	0	r3
		NO: 7	GATATTGAACAGGAA	0	r4
		NO: 9	GTCATCCGCCAGCAG	0	r5
		NO: 11	TATCGACTCCCAGCT	0	r6
		NO: 13	TCCGTCTACGGAAAG	0	r7
		NO: 15	GACGCAATGAGGCAC	0	r8
		NO: 17	ATGACCTGATGACAG	0	r9
		NO: 19	GAATATGCCGGTTATC	0	r10
		NO: 21	GAAGCGGCATGGAAAG	4	r11
		NO: 23	TATTACGTCAGCGAG	0	r12
		NO: 25	CAGCGTGATGGAGCA	0	r13
		NO: 27	TGCAGACGGCTCAGGA	7638	r14
		NO: 29	GAAAGAAGTTCAGGA	0	r15

[0078] 表11生信结果-3

标签序列	SEQ ID	上游引物	数量	已知序列名称	
[0079]	T11	NO: 1	GGCAATGAGCTTCTT	0	r1

[0080]	NO: 3	TCTCACAGTGTGTCC	0	r2
	NO: 5	CCAGCTGAATTCCAA	0	r3
	NO: 7	CCTCCTGCTTCTCCT	0	r4
	NO: 9	GCTGGCGATAGATTT	0	r5
	NO: 11	CTCCTCTGGTTTCAT	0	r6
	NO: 13	GTTGATGAGGCTTAT	0	r7
	NO: 15	ATCAGTGGGACCTGC	0	r8
	NO: 17	AGATACATGTTATTT	0	r9
	NO: 19	TGCAGTATCTGCGAA	0	r10
	NO: 21	ACTATTTTCAGCAGCT	5612	r11
	NO: 23	GCCATGGAGGCGGCT	3	r12
	NO: 25	ACTGGAAATGGAGGT	0	r13
	NO: 27	CACCTGGGATTTAAT	0	r14
	NO: 29	CAATTCCAGCATCTT	0	r15

[0081] 在下机数据中进行分析:表9所示,通过r2对应的标签序列T2拆分,结果发现,标签序列T2下除了r2外,没有其他已知序列,表明r2对应的barcode未污染其余14个barcode;表10所示,通过r14对应的标签序列T14进行拆分,结果发现,标签序列T14下除了r14外,还有r11,则r14对应的barcode污染了r11对应的barcode,根据表11,通过r11对应的标签序列T11进行拆分后,可知r11的数量为5612,则污染比例为 $4 \div 5612 = 0.07\%$ 。

[0082] 以上的实施例是为了说明本发明公开的实施方案,并不能理解为对本发明的限制。此外,本文所列出的各种修改以及发明中方法、组合物的变化,在不脱离本发明的范围和精神的前提下对本领域内的技术人员来说是显而易见的。虽然已结合本发明的多种具体优选实施例对本发明进行了具体的描述,但应当理解,本发明不应仅限于这些具体实施例。事实上,各种如上所述的对本领域内的技术人员来说显而易见的修改来获取发明都应包括在本发明的范围内。

序列表

- <110> 上海韦翰斯生物医药科技有限公司
- <120> 一种测序平台标签序列污染的检测方法
- <160> 46
- <170> SIPOSequenceListing 1.0
- <210> 1
- <211> 15
- <212> DNA
- <213> 人工序列 (Artificial Sequence)
- <400> 1
- gctgacattt tcggt 15
- <210> 2
- <211> 15
- <212> DNA
- <213> 人工序列 (Artificial Sequence)
- <400> 2
- tggcctgccg cagtt 15
- <210> 3
- <211> 15
- <212> DNA
- <213> 人工序列 (Artificial Sequence)
- <400> 3
- cagccaggaa ctatt 15
- <210> 4
- <211> 16
- <212> DNA
- <213> 人工序列 (Artificial Sequence)
- <400> 4
- gttttccagt tccgga 16
- <210> 5
- <211> 15
- <212> DNA
- <213> 人工序列 (Artificial Sequence)
- <400> 5
- atccgtgagg tgaat 15
- <210> 6
- <211> 15
- <212> DNA

<213> 人工序列 (Artificial Sequence)
<400> 6
cagcgacgga atatc 15
<210> 7
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 7
gatattgaac aggaa 15
<210> 8
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 8
taagatactg ctctc 15
<210> 9
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 9
gtcatccgcc agcag 15
<210> 10
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 10
agtctttgac aatct 15
<210> 11
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 11
tatcgactcc cagct 15
<210> 12
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 12
catttctgca ccatt 15

<210> 13
<211> 15
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 13
tccgtctacg gaaag 15
<210> 14
<211> 15
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 14
tcgggaagtg aacgg 15
<210> 15
<211> 15
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 15
gacgcaatga ggcac 15
<210> 16
<211> 15
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 16
tcatcctctc cggat 15
<210> 17
<211> 15
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 17
atgacctgat gacag 15
<210> 18
<211> 15
<212> DNA
<213> 人工序列(Artificial Sequence)
<400> 18
atacataaaa tcctg 15
<210> 19
<211> 16
<212> DNA

<213> 人工序列 (Artificial Sequence)
<400> 19
gaatatgccg gttatc 16
<210> 20
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 20
cctgatgcag ctggat 16
<210> 21
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 21
gaagcggcat ggaaag 16
<210> 22
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 22
ctgaccatcc ggaact 16
<210> 23
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 23
tattacgtca gcgag 15
<210> 24
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 24
tgcccgtcct ccacgg 16
<210> 25
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 25
cagcgtgatg gagca 15

<210> 26
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 26
ccaatccagc cgtca 16
<210> 27
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 27
tgcagacggc tcagga 16
<210> 28
<211> 16
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 28
aaagtacgcc cacgac 16
<210> 29
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 29
gaaagaagtt cagga 15
<210> 30
<211> 15
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 30
gattcaaagc ctgca 15
<210> 31
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 31
ttatat 6
<210> 32
<211> 6
<212> DNA

<213> 人工序列 (Artificial Sequence)
<400> 32
accaac 6
<210> 33
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 33
ctatgc 6
<210> 34
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 34
attcct 6
<210> 35
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 35
caactc 6
<210> 36
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 36
ttaggc 6
<210> 37
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 37
aggatc 6
<210> 38
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 38
cagcaa 6

<210> 39
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 39
aagtag 6
<210> 40
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 40
acagtg 6
<210> 41
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 41
ggtcca 6
<210> 42
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 42
gatcag 6
<210> 43
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 43
attatg 6
<210> 44
<211> 6
<212> DNA
<213> 人工序列 (Artificial Sequence)
<400> 44
ggctac 6
<210> 45
<211> 6
<212> DNA

<213> 人工序列 (Artificial Sequence)

<400> 45

gaacct 6

<210> 46

<211> 61

<212> DNA

<213> 人工序列 (Artificial Sequence)

<400> 46

caagcagaag acggcatacg agatnnnnnn gtgactggag ttccttgca cccgagaatt 60

c 61

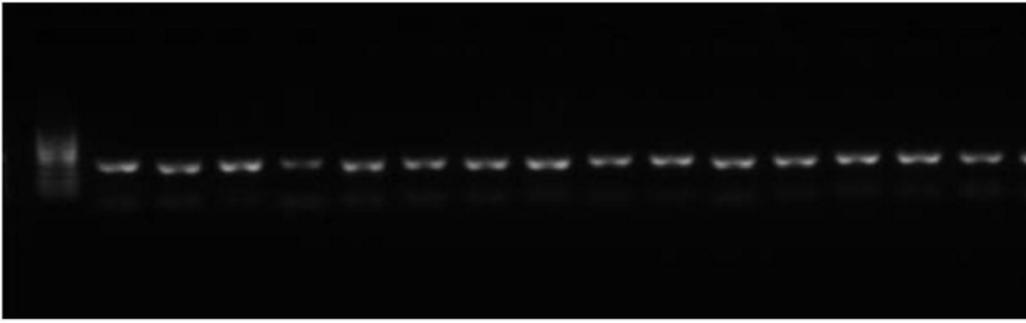


图1