



(12) 发明专利

(10) 授权公告号 CN 111534518 B

(45) 授权公告日 2021.07.23

(21) 申请号 202010421923.X

C40B 50/06 (2006.01)

(22) 申请日 2020.05.18

C12Q 1/6806 (2018.01)

(65) 同一申请的已公布的文献号

(56) 对比文件

申请公布号 CN 111534518 A

CN 108456713 A, 2018.08.28

CN 108676846 A, 2018.10.19

(43) 申请公布日 2020.08.14

CN 108949941 A, 2018.12.07

(73) 专利权人 纳昂达(南京)生物科技有限公司

WO 2018183808 A1, 2018.10.04

地址 210031 江苏省南京市江北新区华康路142号南京生物医药谷加速器三期A01栋南侧3-4层

审查员 常子月

(72) 发明人 胡玉刚 汪彪 郑文莉 吴强

(74) 专利代理机构 北京康信知识产权代理有限公司 11240

代理人 路秀丽

(51) Int. Cl.

权利要求书2页 说明书13页

C12N 15/113 (2010.01)

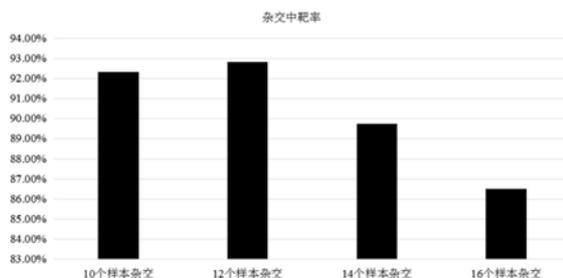
序列表7页 附图4页

(54) 发明名称

通用封闭序列及其应用

(57) 摘要

本发明提供了一种通用封闭序列及其应用。该通用封闭序列按照从5'到3'的方向包括依次连接的左侧非标签区封闭序列、中间标签区封闭序列及右侧非标签区封闭序列，左侧非标签区封闭序列包括5~7个LAN或BNA修饰的碱基，中间标签区封闭序列为通用封闭碱基序列，右侧非标签区封闭序列包括7~10个LAN或BNA修饰的碱基，且右侧非标签区封闭序列的3'端带有封闭修饰。通过在左侧和右侧非标签区分别设计5~7个和7~10个LNA或BNA修饰的碱基序列能够显著增强与待封闭序列的结合能力，从而增加封闭效果；3'端进行封闭修饰能够减少或避免文库的非特异性捕获，提高目的文库捕获的中靶率。



1. 一种通用封闭序列组合,其特征在于,所述通用封闭序列组合为MGI测序平台的带有第一标签序列的P1接头的封闭序列和带有第二标签序列的P2接头的封闭序列的组合,其中,所述组合选自:

组合1):所述P1接头的封闭序列为:

CTCTCA+GTACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCACAGA+ACGA+CATGG+CTACGATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer, /3SpC3/为3'端的C3间隔臂修饰;

所述P2接头的封闭序列为:

GCA+TGGC+GA+CCTT+ATCA+GXXXXXXXXXXTTGTCTT+CCTA+AGA+CCGC+TTG+GCC+TCCGA+CTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer, /3SpC3/为3'端的C3间隔臂修饰;

或者

组合2):所述P1接头的封闭序列为:

CTC+TCA+GT+ACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer, /3SpC3/为3'端的C3间隔臂修饰;

所述P2接头的封闭序列:

GCA+TG+GC+GA+CC+TT+ATCA+GXXXXXXXXXXTTG+TCTT+CCTA+AGA+CC+GC+TTG+GCC+TCC+GA+CTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer, /3SpC3/为3'端的C3间隔臂修饰;

或者

组合3)P1接头的封闭序列为:

CTC+T+CA+GT+ACG+TCA+GCA+GT+TXXXXXXXXXXCAACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGA+CTT/3SpC3/,其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer, /3SpC3/为3'端的C3间隔臂修饰;

P2接头的封闭序列为:

G+CA+TG+GC+GA+CC+TT+ATCAGXXXXXXXXXXTTGTCTT+CCTA+AGA+CC+GC+TTG+GCC+TC+C+GA+CTT/3SpC3/,其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer, /3SpC3/为3'端的C3间隔臂修饰。

2. 一种捕获试剂盒,所述捕获试剂盒包括通用封闭序列组合,其特征在于,所述通用封闭序列组合为权利要求1所述的通用封闭序列组合。

3. 根据权利要求2所述的试剂盒,其特征在于,所述试剂盒中通用捕获探针的工作浓度为0.4~0.8 $\mu$ g所述通用封闭序列组合中的任一条/1 $\mu$ g待捕获文库。

4. 一种文库杂交捕获方法,所述方法包括采用捕获试剂盒对待捕获文库进行捕获,其特征在于,所述捕获试剂盒采用权利要求2或3所述的捕获试剂盒,所述通用封闭序列组合中的任一条封闭序列与待捕获文库按照摩尔比为20:1的比例进行封闭。

5. 根据权利要求4所述的方法,其特征在于,

当通用封闭序列组合为所述组合1)时,任一条封闭序列加入浓度为200 $\mu$ mol/L;

当通用封闭序列组合为所述组合2)时,任一条封闭序列加入浓度为100 $\mu$ mol/L;

当通用封闭序列组合为所述组合3)时,任一条封闭序列加入浓度为100 $\mu\text{mol/L}$ 。

6. 一种多文库捕获方法,其特征在于,所述方法包括:

构建片段化文库;

采用权利要求2或3所述的试剂盒,或者采用权利要求4或5所述的方法对所述片段化文库进行杂交捕获,得到捕获文库;

对所述捕获文库进行PCR扩增,得到测序文库;

其中,对所述片段化文库进行杂交捕获的步骤中,单次不超过12个所述片段化文库进行混合杂交捕获,且所述通用封闭序列的工作浓度为0.4~0.8 $\mu\text{g}$ 所述通用封闭序列组合中的任一条/1 $\mu\text{g}$ 待捕获文库。

## 通用封闭序列及其应用

### 技术领域

[0001] 本发明涉及高通量测序文库构建领域,具体而言,涉及一种通用封闭序列及其应用。

### 背景技术

[0002] 随着高通量测序在临床应用的辅助诊断重要性的提升,如何降低测序成本是一个很关键的问题,降低测序成本在不同的层面有相应的表现:华大智造(MGI)不断推出更高测序通量的测序仪,测序成本不断降低,相继推出MGI-200、MGI-2000和T7测序仪,其中T7测序仪是目前市场上测序通量最高和测序成本最低的测序仪。此外靶向捕获测序也是在检测目标序列的同时实现大规模降低检测成本的有效途径。

[0003] 在测序的过程中,不同样本用不同的Index序列来区分,进而将多个样本混合测序,也是高通量测序能够降低单个样本成本的方式。但是,如果采用单端Index,由于Index接头或引物在合成、文库构建实验操作以及测序的各个环节,都不可避免地会导致污染和/或相互串扰。因此,需要一种方式解决样本之间的低频度的相互串扰,目前解决这个问题的方法是采用双端Index来区分不同的样本,用双端Index的方法可以有效去除样本之间的相互串扰。

[0004] 通过杂交捕获可以有效地降低检测目标的测序成本,同时如果在杂交封闭的过程中能够提升捕获目标区域的占比,也能节省测序成本。然而,如何对双端带有index的样本文库进行高效捕获,目前尚无有效的解决方案。

### 发明内容

[0005] 本发明的主要目的在于提供一种通用封闭序列及其应用,以解决现有技术中双端index文库杂交捕获效率较低的问题。

[0006] 为了实现上述目的,根据本发明的一个方面,提供了一种通用封闭序列,该通用封闭序列按照从5'到3'的方向包括依次连接的左侧非标签区封闭序列、中间标签区封闭序列及右侧非标签区封闭序列,其中,左侧非标签区封闭序列包括5~7个LAN或BNA修饰的碱基,中间标签区封闭序列为通用封闭碱基序列,右侧非标签区封闭序列包括7~10个LAN或BNA修饰的碱基,且右侧非标签区封闭序列的3'端带有封闭修饰。

[0007] 进一步地,3'端的封闭修饰为MGB修饰、C3间隔臂修饰,磷酸化修饰,地高辛修饰或生物素修饰,或者3'端碱基为双脱氧碱基。

[0008] 进一步地,通用封闭碱基为次黄嘌呤或C3间隔臂。

[0009] 进一步地,通用封闭序列为MGI测序平台的带有第一标签序列的P1接头的封闭序列或带有第二标签序列的P2接头的封闭序列,其中,P1接头的封闭序列为SEQ ID NO:3:

[0010] CTCTCA+GTACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCACAGA+ACGA+CATGG+CTACGATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的C3间隔臂修饰;

[0011] P2接头的封闭序列为SEQ ID NO:4:

[0012] GCA+TGGC+GA+CCTT+ATCA+GXXXXXXXXXXTTGTCTT+CCTA+AGA+CCGC+TTG+GCC+TCCGA+CTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的C3间隔臂修饰。

[0013] 进一步地,通用封闭序列为MGI测序平台的带有第一标签序列的P1接头的封闭序列或带有第二标签序列的P2接头的封闭序列;

[0014] P1接头的封闭序列为SEQ ID NO:5:

[0015] CTC+TCA+GT+ACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的C3间隔臂修饰;

[0016] P2接头的封闭序列SEQ ID NO:6:

[0017] GCA+TG+GC+GA+CC+TT+ATCA+GXXXXXXXXXXTTG+TCTT+CCTA+AGA+CC+GC+TTG+GCC+TCC+GA+CTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的C3间隔臂修饰。

[0018] 为了实现上述目的,根据本发明的第二个方面,提供了一种捕获试剂盒,该捕获试剂盒包括通用封闭序列,封闭序列为上述任一种通用封闭序列。

[0019] 进一步地,试剂盒中通用捕获探针的工作浓度以单个文库计,为0.4~0.8μg通用封闭序列/1μg待捕获文库。

[0020] 根据本发明的第三个方面,提供了一种文库杂交捕获方法,该方法包括采用捕获试剂盒对待捕获文库进行捕获,捕获试剂盒采用上述任一种捕获试剂盒。

[0021] 进一步地,采用捕获试剂盒对待捕获文库进行捕获的步骤包括,将封闭序列与待捕获文库按照摩尔比为10:1~20:1的比例进行封闭。

[0022] 根据本发明的第四个方面,提供了一种建库方法,该建库方法包括:构建片段化文库;对片段化文库进行杂交捕获,得到捕获文库;对捕获文库进行PCR扩增,得到测序文库;采用上述任一种捕获试剂盒进行杂交捕获,或者采用上述任一种方法进行杂交捕获。

[0023] 应用本发明的技术方案,通过在左侧非标签区封闭序列和右侧非标签区封闭序列上分别对5~7个和7~10个碱基进行LNA或BNA修饰,能够增强对待封闭序列的结合能力,从而提升封闭效果;在右侧非标签区封闭序列的3'端进行封闭修饰,使得本申请的封闭序列在文库捕获时能够减少或避免对非目标文库的捕获,提升目标文库的捕获率(或者叫目标文库的中靶率)。

## 附图说明

[0024] 构成本申请的一部分的说明书附图用来提供对本发明的进一步理解,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。在附图中:

[0025] 图1示出了目前MGI平台双端Index建库流程的示意图;

[0026] 图2示出了现有技术中采用通用封闭序列进行捕获时能够提升目标序列捕获的中靶率的原理示意图;

[0027] 图3示出了现有技术中MGI平台双端Index文库采用12个双端Index所构建的文库之间相互串扰的统计结果图;

[0028] 图4A和图4B示出了改进后的通用封闭序列对杂交文库的正常封闭和异常封闭的原理示意图；

[0029] 图5示出了采用不同修饰数量和不同浓度的通用封闭序列对目标序列捕获的中靶率的影响示意图；

[0030] 图6示出了过少和过多的碱基修饰的通用封闭序列对封闭效果的影响；

[0031] 图7示出了修饰碱基数量相同但修饰位置不同的通用封闭序列对封闭效果的影响；

[0032] 图8示出了不同文库投入量对封闭效果影响。

### 具体实施方式

[0033] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。下面将结合实施例来详细说明本发明。

[0034] 术语解释:

[0035] 双端Index接头:高通量测序时需要每个片段末端连接通用的测序接头,接头的非互补区域各有一个可变序列区域序列是Index序列,是用来测序时拆分数据用。

[0036] 接头封闭序列:在文库捕获时,每个文库都具有相同和相近的接头序列,杂交时目标片段和非目标片段接头部分会相互结合,降低中靶率,接头封闭序列是用来特异结合接头部分序列,起到提升中靶率的作用。通用封闭序列,就是针对文库中带有不同Index的接头均能实现封闭的序列。

[0037] C3间臂:C3 Splicer主要用于模仿核糖的3'和5'羟基间的三碳间隔,或“替代”一个序列中未知的碱基,在核酸序列中间主要是起连接作用,并不能和互补的碱基配对而起到稳定作用,仅对前后碱基起到连接作用。

[0038] 需要说明的是,本申请中所提及的MGI测序平台用的index长度是以10bp为例进行说明的。本申请中的针对index的标签封闭段的长度可以通过调整次黄嘌呤I和C3间隔臂的长度来适应不同长度的index接头的封闭。

[0039] 如背景技术所提到的,现在的高通量测序仪需要降低测序成本,单次测序通量越来越高,现有MGI测序仪为了节约测序成本,也推出了目前市场上通量最高,测序成本最低的T7测序仪,测序通量提升了就要求一次必须混合很多测序样本测序,高通量测序就是通过不同的样本连接不同的Index接头来实现多个样本混合在一起测序。MGI最先推出的是单端Index接头的建库方案,该方案有一个显著的问题就是会有样本的相互串扰,为了解决串扰问题MGI也推出了双端Index的建库方案,这样就能解决由于接头合成、实验操作和测序过程相互影响导致的低频串扰问题,通过两个Index数据拆分把串扰的数据过滤掉。

[0040] 但针对其双端index文库进行捕获文库测序时,还存在对含目标片段的文库的捕获效率低的缺陷,为了进一步提高该双端Index文库的捕获效率,本申请利用纳昂达针对MGI测序仪推出的双端Index建库方案(具体的双端Index建库流程如图1),提出了相应的杂交捕获改进方案。

[0041] 另一个降低测序成本的有效途径是靶向捕获测序,人的基因组大小是3Gb,编码基因的区域占比不到2%的区域,(以IDT的全外显子V2版本来说)涵盖了人的大概2万个基因,约34Mb大小,所以一个全基因组测序成本可以测10个全外显子测序,而肿瘤靶向用药的检

测区域与全基因组测序相比,成本差异更大,更能体现出靶向测序的重要性。而且肿瘤的突变都是低频突变,需要考虑两个主要问题,一个是低频突变不能有样本之间的串扰,或者说如果串扰不能避免必须能有方法剔除串扰的数据,双端Index就是当串扰不能避免时的一种可剔除串扰的必选方案,所以这也是双端Index存在的最大意义。能够检测低频突变的另一个条件是测序深度要有保证,一般需要几千乘到几万乘的测序深度。

[0042] 在靶向测序的过程中,有两方面的关键因素决定靶向捕获的中靶率,一方面是设计区域的特异性,设计的探针不能落到高度重复区域,另一方面是捕获时的接头封闭效果。探针序列是由待捕获区域的序列决定的,一般在设计探针时都会避开高度重复的区域,所以我们能够改进的是靶向捕获时的封闭效果。如果捕获时不加入封闭序列,理论上以及实际测试均表明,(目标片段文库)的中靶率不会超过50%。如图2所示,在不加入封闭接头序列时,接头部分会结合非靶向文库从而减低中靶率。在捕获探针区域的特异性比较好的前提下,加入不同等级的封闭序列可以使中靶率在45-90%之间波动。因此,本申请就以MGI平台的双端Index接头文库为例,解释说明改进设计的通用封闭序列及其所达到的封闭效果,能够大幅提升中靶率进而降低测序成本。

[0043] MGI平台的双端Index的文库两端分别有一个接头序列,不同编号的Index接头所携带的index序列都是一个10bp的可变区,该可变区是用来对混合捕获和混合测序中的不同样本进行区分的。上文提到双端Index的目的是去除相互串扰的,如图3所示,每万条序列通过双端Index可以过滤掉3-7条序列,如果不用双端Index对千分之一及以下的突变检测是不可信的数据,通过双端Index可以提升检测的准确性。在杂交捕获的过程中为了提升中靶率,本发明开发了针对MGI平台双端Index的通用封闭序列。这个通用序列的特征之一是10bp的Index区域是选择通用的碱基(次黄嘌呤或C3 spacer)起到封闭/占位作用,为了提升封闭效果两端的固定序列区域进行了提升杂交温度的碱基修饰替代,固定序列部分碱基进行LNA或BNA修饰。双端Index的通用探针的具体特征和要求如下:(1) 标签序列区域为通用封闭碱基,比如次黄嘌呤(I)、C3间隔臂等间隔序列或者其组合;(2) 为了增强该通用型封闭序列的结合效率,在封闭序列的上下游的非标签序列区域的部分碱基进行LNA或BNA修饰,并且修饰碱基的数量分别在5~7和7~10,或者20%-40%之间。(3) 通用封闭序列最适使用浓度和修饰碱基数量成反比,修饰的碱基多,最适浓度相对浓度低,太高浓度会产生起反作用;相反修饰的碱基少,要达到封闭效果需要更高的封闭序列浓度。

[0044] 本发明根据MGI双端Index的接头序列特征设计了通用封闭修饰序列,MGI两条通用封闭的原始序列如下:

[0045] SEQ ID NO:1:

[0046] CTCTCAGTACGTCAGCAGTTNNNNNNNNNCAACTCCTTGGCTCACAGAACGACATGGCTACGATCCGACTT。

[0047] SEQ ID NO:2:

[0048] GCATGGCGACCTTATCAGNNNNNNNNNTTGTCTTCTAAGACCGCTTGGCCTCCGACTT。

[0049] N的部分是Index序列的封闭序列,10bp长度的Index相比6bp和8bp时的Index,其好处是可以增加在设计不同Index时的选择,缺点是对通用封闭序列设计而言,增加了难度和不稳定性。在设计成通用封闭序列时,Index区域的碱基设计成简并碱基N、C3间隔臂和次黄嘌呤,所获得的通用封闭序列的不稳定性更高,这就要求Index两端通用封闭序列修饰

的碱基更多,因此,本申请将提升封闭效果的修饰碱基的数量由3-6个提升到5-10个。

[0050] 在本申请一种优选的实施例中,P1(5+7修饰)端封闭序列为SEQ ID NO:3:

[0051] CTCTCA+GTACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCACAGA+ACGA+CATGG+CTACGATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。在一种优选的实施例中,P2端封闭序列为SEQ ID NO:4:

[0052] GCA+TGGC+GA+CCTT+ATCA+GXXXXXXXXXXTTGTCTT+CCTA+AGA+CCGC+TTG+GCC+TCCGA+CTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。

[0053] 在本申请另一种优选的实施例中,P1端封闭序列为SEQ ID NO:5:

[0054] CTC+TCA+GT+ACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。在一种优选的实施例中,P2端封闭序列为SEQ ID NO:6:

[0055] GCA+TG+GC+GA+CC+TT+ATCA+GXXXXXXXXXXTTG+TCTT+CCTA+AGA+CC+GC+TTG+GCC+TCC+GA+CTT/3SpC3/;其中+表示LAN或BNA修饰,X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。

[0056] 进一步研究发现,修饰碱基的数量比较多,同时加入的通用封闭序列的浓度比较高时反而封闭效果不好,如图4A和图4B所示,当封闭序列中修饰碱基数量多了,在局部区域形成强的结合能力,大部分会像图4A所示,封闭序列可以正确封闭文库序列,但同时也会形成十字星型封闭,捕获下来一个非靶向序列,从而降低文库中靶率,如图4B所示。

[0057] 进一步地发现在杂交的过程中,加入的文库量也有最大量的限制,以250bp的长度插入片段文库计算,在单次杂交不能超过6.5 $\mu$ g(25pmol/L,这里指单端序列),在用全外显子测试发现每个文库在杂交时投入500ng,12个文库杂交效果要比14-16个文库杂交效果好,在杂交时由于文库投入量过多,文库空间距离变短也会有部分文库两个文库和通用封闭形成图4B所示的双文库和双封闭序列的十字星型结构,从而导致捕获效率降低。

[0058] 基于上述研究和发现,申请人提出了本申请的技术方案。在一种典型的实施方式中,提供了一种通用封闭序列,该通用封闭序列按照从5'到3'的方向包括:包括依次连接的左侧非标签区封闭序列、中间标签区封闭序列及右侧非标签区封闭序列,其中,左侧非标签区封闭序列包括5~7个LAN(Locked nucleic acid,锁核酸)或BNA(桥连核酸Bridged nucleic acid 2',4'-BNA<sup>NC</sup>,即2'-0,4'-aminoethylene bridged nucleic acid是含有具有NO键的六元桥连结构的化合物)修饰的碱基(主要对是C碱基进行LNA或BNA修饰),中间标签区封闭序列为通用封闭碱基序列,右侧非标签区封闭序列包括7~10个LAN或BNA修饰的碱基,且右侧非标签区封闭序列的3'端带有封闭修饰。

[0059] 通过在左侧非标签区封闭序列上5~7个碱基,并在右侧非标签区封闭序列上的7~10个碱基进行LNA或BNA修饰,能够显著增强与待封闭序列的结合能力,从而增加封闭效果;而在右侧非标签区封闭序列的3'端进行封闭修饰,使得文库中多余的接头不能被作为引物对其他文库的接头进行扩增,进而能够减少或避免文库的非特异性捕获,提高目的文库捕获的中靶率。

[0060] 上述右侧非标签区封闭序列的3'端的封闭修饰,可以采用MGB修饰、C3间隔臂修饰,3'磷酸化修饰,3'地高辛修饰,3'生物素修饰或3'端的碱基为双脱氧碱基。在本申请中,

优选采用C3间隔臂修饰。

[0061] 上述标签区封闭序列采用通用封闭碱基可以采用与A、T、C和G四种碱基均存在弱结合能力的碱基序列。在本申请一种优选的实施例中，该通用封闭碱基为次鸟嘌呤I和/或C3间隔臂。具体标签区封闭序列的碱基数目并非局限于10bp，可以根据待封闭的文库中的样本标签的碱基数目合理设定。比如，还可以是6bp、7bp、8bp、9bp、11bp或12bp等。

[0062] 当采用C3间隔臂作为标签区封闭序列时，以前述MGI平台的P1和P2接头的封闭序列为例，其是10个C3间隔臂，或10个次黄嘌呤(I)。次黄嘌呤的优点是能与所有的碱基都有微弱的配对能力，而C3间隔臂只是占位一个碱基，与配对碱基没有结合能力，不能起到稳定的作用。

[0063] 上述通用封闭序列中，左侧非标签区封闭序列和右侧非标签区封闭序列中LNA或BNA修饰的碱基的数目，通常认为修饰碱基数据量与左侧非标签区封闭序列或右侧非标签区封闭序列的序列长度呈负相关，序列长则需要修饰的碱基数量少，序列短则需要修饰的碱基数量多。然而在本申请中，发明人发现，对于特定长度的非标签区域的封闭序列来看，LNA或BNA修饰的碱基在左侧非标签区封闭序列或右侧非标签区封闭序列中的数目为5~10个碱基时，封闭序列与目的接头的结合能力最强。而当少于5个碱基时，与目的接头的结合不稳定，从而使得捕获效率较低。

[0064] 此外，在利用上述通用封闭序列进行文库捕获时，待捕获文库的总量与加入的通用封闭序列的量最好也要匹配，若添加的文库量过多，容易导致封闭序列之间杂交形成星状结构，从而导致非特异性捕获，减低捕获效率。比如，当通用封闭序列的添加量为2.4μg时，以每个文库500ng计算，共杂交12个文库，即总量为6μg时，杂交捕获效果比杂交14~16个文库的捕获效果要好。当然，每个文库小于500ng，比如400ng时，2.4μg的通用封闭序列，同时杂交15个文库的捕获效率最高。

[0065] 针对MGI平台的双端index接头，本申请也提供了能够抑制MGI测序平台的双端Index测序文库捕获的通用封闭序列。在本申请一种优选的实施例中，P1(5+7修饰)端封闭序列为SEQ ID NO:3:

[0066] CTCTCA+GTACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCACAGA+ACGA+CATGG+CTACGATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰，X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。在一种优选的实施例中，P2端封闭序列为SEQ ID NO:4:

[0067] GCA+TGGC+GA+CCTT+ATCA+GXXXXXXXXXXTTGTCTT+CCTA+AGA+CCGC+TTG+GCC+TCCGA+CTT/3SpC3/;其中+表示LAN或BNA修饰，X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。

[0068] 在本申请另一种优选的实施例中，P1端封闭序列为SEQ ID NO:5:

[0069] CTC+TCA+GT+ACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGACTT/3SpC3/;其中+表示LAN或BNA修饰，X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。在一种优选的实施例中，P2端封闭序列为SEQ ID NO:6:

[0070] GCA+TG+GC+GA+CC+TT+ATCA+GXXXXXXXXXXTTG+TCTT+CCTA+AGA+CC+GC+TTG+GCC+TCC+GA+CTT/3SpC3/;其中+表示LAN或BNA修饰，X表示次黄嘌呤或C3 spacer,/3SpC3/为3'端的封闭修饰。

[0071] 上述两种优选的实施例所提供的通用封闭序列，不仅在数量上增加了修饰碱基从

而提高了与目的接头的结合能力,而且上述通用封闭序列上的修饰碱基的具体位置也相比修饰其他位置的碱基时,对目的接头的结合能力强。也就是说,上述优选的通用封闭序列对目的接头的封闭效果最佳,杂交捕获时,目的文库的捕获效率最高。

[0072] 在上述各种改进的通用封闭序列的基础上,在本申请第二种典型的实施方式中,提供了一种捕获试剂盒,该捕获试剂盒包括上述任一种通用封闭序列。该捕获试剂盒中的通用封闭序列与目的接头的结合能力强,在用于捕获文库构建时,能够实现目的文库的高效捕获。

[0073] 为了进一步提高目标文库的捕获效率(即中靶率),在一种优选实施例中,上述试剂盒中通用捕获探针的工作浓度为0.4~0.8 $\mu$ g通用封闭序列/1 $\mu$ g待捕获文库。按照上述用量进行捕获,能够进一步避免文库量过多而形成十字星型封闭,从而降低文库中靶率。上述工作浓度可以根据具体封闭方案的不同有所不同,比如,当采用本申请的SEQ ID NO:5和SEQ ID NO:6的封闭方案时,该工作浓度按照0.4 $\mu$ g通用封闭序列/1 $\mu$ g待捕获文库进行捕获,目标文库的中靶率较高。当采用本申请的SEQ ID NO:3和SEQ ID NO:4的封闭方案时,该工作浓度按照0.8 $\mu$ g通用封闭序列/1 $\mu$ g待捕获文库进行捕获目标文库的中靶率较高。

[0074] 在本申请第三种典型的实施方式中,还提供了一种文库杂交捕获方法,该方法包括采用捕获试剂盒对待捕获文库进行捕获,捕获试剂盒采用上述捕获试剂盒。该捕获试剂盒中的封闭序列在捕获文库构建时,能够实现目的文库的高效捕获。

[0075] 发明人还发现通用封闭序列与待捕获文库按照摩尔比为10:1~20:1时封闭效果更好。因而在本申请一种优选的实施例中,采用捕获试剂盒对待捕获文库进行捕获的步骤中,将封闭序列与待捕获文库按照摩尔比为10:1~20:1的比例进行封闭。

[0076] 在本申请第四种典型的实施方式中,提供了一种建库方法,该建库方法包括:构建片段化文库;对片段化文库进行杂交捕获,得到捕获文库;对捕获文库进行PCR扩增,得到测序文库;采用上述捕获试剂盒进行杂交捕获,或者采用上述任一种方法进行杂交捕获。采用本申请的建库方法所构建的文库中目的文库的占比较高,文库产出的有效数据占比高。

[0077] 下面将结合具体的实施例来进一步说明本申请的有益效果。

[0078] 需要说明的是,以下实施例采用NadPrep<sup>TM</sup> DNA文库构建试剂盒(for MGI)(201909Version2.0)(纳昂达(南京)生物科技有限公司)所提供的文库构建流程进行。还需要说明的是,以下实施例仅是示例性说明,并不限定本申请的方法仅能采用如下方法。具体流程简述如下:

[0079] DNA样本片段化---末端修复和加A---接头连接---片段筛选---PCR扩增---文库纯化、定量和质检---使用MGI平台测序或靶向捕获后测序。

[0080] 实施例1修饰多少和加入浓度差异的通用封闭方案

[0081] 步骤:文库构建参考NadPrep<sup>TM</sup> DNA文库构建试剂盒(for MGI)(201909Version2.0)说明书进行。其中杂交捕获的步骤按以下进行,真空浓缩后进行多文库混合杂交捕获时,具体的杂交文库混合步骤如下表:

[0082] 表1:

组分	总文库量	数量
总文库	6 $\mu$ g	1~12
Human Cot DNA	5 $\mu$ l	/

通用封闭序列(下面列出序列)	2 $\mu$ l	/
----------------	-----------	---

[0084] 1) 本申请改进的MGI接头双端Index通用型封闭序列

[0085] 1.1 SEQ ID NO:3所示的P1端封闭序列和SEQ ID NO:4所示的P2端封闭序列。

[0086] CTCTCA+GTACG+TCA+GCA+GT+T10XXXXXXXXXXCA+ACTCCT+TGGC+TCACAGA+ACGA+CATGG+CTACGATC+CGACTT/3SpC3/;

[0087] GCA+TGGC+GA+CCTT+ATCA+GXXXXXXXXXXTTGTCTT+CCTA+AGA+CCGC+TTG+GCC+TCCGA+CCTT/3SpC3/

[0088] 1.2 SEQ ID NO:5:所示的P1端封闭序列和SEQ ID NO:6所示的P2端封闭序列CTC+TCA+GT+ACG+TCA+GCA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGACTT/3SpC3/;

[0089] GCA+TG+GC+GA+CC+TT+ATCA+GXXXXXXXXXXTTG+TCTT+CCTA+AGA+CC+GC+TTG+GCC+TCC+GA+CCTT/3SpC3/

[0090] 上述四条序列中,X表示C3 Spacer/次黄嘌呤,+N表示LNA或BNA修饰碱基(两种修饰效果相当,以LNA修饰为例),/3SpC3/表示3'的C3间隔臂封闭。

[0091] 在杂交过程中放入的通用封闭序列和浓度表如下:

[0092] 表2:

封闭组合	5+7 修饰封闭(封闭方案1)			7+10 修饰组合(封闭方案2)		
浓度( $\mu$ mol/L)	50	100	200	50	100	200

[0094] 2) 具体杂交捕获步骤如下:

[0095] 1、按照上表将各组分混合于一个0.2/1.5ml的低吸附离心管中,涡旋混匀,瞬时离心。

[0096] 2、将离心管放入提前预热至60 $^{\circ}$ C的真空浓缩仪中干燥。

[0097] 3、待全部液体蒸发并完全干燥后,将离心管密封后备用。

[0098] 4、取出xGen<sup>®</sup>Exome Research Panel v2.0在冰上自然融解,使用后按需小量分装。

[0099] 5、根据下表3配制杂交反应液,使用移液器混合均匀后加入到已经真空浓缩干燥的离心管底部,使用移液器轻柔吹吸混匀15~20次,瞬时离心,25 $^{\circ}$ C孵育5~10min。

[0100] 表3:

[0101]	杂交缓冲液 (2X Hybridization Buffer)	8.5 $\mu$ l
	杂交缓冲液增强子 (Hybridization Buffer Enhancer)	2.7 $\mu$ l
	无 RNA 酶的水 (Nuclease Free Water)	1.8 $\mu$ l
	xGen® Exome Research Panel v2.0	4 $\mu$ l
	总计	17 $\mu$ l

[0102] 6、涡旋混匀杂交反应混合液,瞬时离心后,将离心管中的全部17 $\mu$ l杂交反应混合液转移至一个新的0.2ml PCR管中,瞬时离心,放进PCR仪中,启动如下杂交程序:

[0103] 表4:

[0104]	杂交程序 (热盖 100°C)	
	95°C	30 s
	65°C	4 ~ 16 h
	65°C	Hold
	洗脱程序 (热盖 70°C)	
	65°C	Hold

[0105] 7、杂交文库洗脱

[0106] (1) 准备工作

[0107] 1、取出xGen®Hybridization and Wash Kit中的其他试剂室温自然融解,涡旋混合均匀(注意:Wash Buffer I如果无法融解,可于65°C水浴孵育至完全融解)。

[0108] 2、Dynabeads™ M-270 Streptavidin Beads涡旋混合均匀,室温平衡30min后方可进行链霉亲和素磁珠的清洗和捕获步骤。

[0109] (2) 试剂配制

[0110] 1、洗脱缓冲液的配制

[0111] 根据下表体系配制洗脱缓冲液的1X工作液:

[0112] 表5:

[0113]	组分名称	Rnase-free水	缓冲液	总计
	2X磁珠洗脱缓冲液	160 $\mu$ l	160 $\mu$ l	320 $\mu$ l

10X洗脱缓冲液I	252 $\mu$ l	28 $\mu$ l	280 $\mu$ l
10X洗脱缓冲液II	144 $\mu$ l	16 $\mu$ l	160 $\mu$ l
10X洗脱缓冲液III	144 $\mu$ l	16 $\mu$ l	160 $\mu$ l
10X Stringent洗脱缓冲液	288 $\mu$ l	32 $\mu$ l	320 $\mu$ l

[0114] 2、磁珠悬浮液配制,如表6。

[0115] 表6:

杂交缓冲液 (2X Hybridization Buffer)	8.5 $\mu$ l
杂交缓冲液增强子 (Hybridization Buffer Enhancer)	2.7 $\mu$ l
无 RNA 酶的水 (Nuclease Free Water)	58 $\mu$ l
总计	17 $\mu$ l

[0117] (3) 亲和素磁珠清洗

[0118] 1、将Dynabeads™ M-270 Streptavidin Beads涡旋混匀15s,确保完全混匀。吸取50 $\mu$ l M270磁珠至1个1.5ml低吸附离心管中。

[0119] 2、向离心管中加入100 $\mu$ l 1X Bead Wash Buffer,轻柔吹吸混匀10次,瞬时离心,置于磁力架上数分钟,待液体完全澄清,使用移液器移弃上清。将离心管从磁力架上移出。

[0120] 3、重复步骤2两次。

[0121] 4、向离心管中加入17 $\mu$ l磁珠悬浮液,轻柔吹吸混匀,将全部磁珠悬浮液转移至1个新的0.2ml低吸附PCR管中。

[0122] (4) 链霉亲和素磁珠捕获

[0123] 1、16h杂交反应后,调节PCR仪进入到洗脱程序。

[0124] 2、将重悬的链霉亲和素磁珠加入到杂交体系中,并使用移液器轻柔吹吸混匀或涡旋混匀。

[0125] 3、65 $^{\circ}$ C孵育45min,每10~12min轻柔涡旋一次,确保磁珠完全重悬。

[0126] (5) 热洗脱(注意:热洗脱过程操作要迅速;吹吸混匀过程中尽量避免产生气泡)

[0127] 1、孵育结束后从PCR仪上取下PCR管,并向其中加入100 $\mu$ l 65 $^{\circ}$ C 1X Wash Buffer I,吹吸混匀含有磁珠的杂交体系。

[0128] 2、将PCR管置于磁力架上1min,待液体完全澄清后,使用移液器吸取移弃上清。

[0129] 3、将PCR管从磁力架上移出,加入150 $\mu$ l 65 $^{\circ}$ C 1X Stringent Wash Buffer,轻柔吹吸10次混合均匀,放进PCR仪中65 $^{\circ}$ C孵育5min。

[0130] 4、重复步骤2和3一次。

[0131] (6) 室温洗脱

[0132] 1、将PCR管瞬时离心后置于磁力架上1min,待液体完全澄清后吸取移弃上清,加入150 $\mu$ l室温1X Wash Buffer I,涡旋混匀,室温孵育2min,期间涡旋混匀30s后静置30s,交替

进行,确保充分混匀。

[0133] 2、将PCR管瞬时离心后置于磁力架上1min,待液体完全澄清后吸取移弃上清,加入150 $\mu$ l室温1X Wash Buffer II,涡旋混匀,室温孵育2min,期间涡旋混匀30s后静置30s,交替进行,确保充分混匀。

[0134] 3、将PCR管瞬时离心后置于磁力架上1min,待液体完全澄清后吸取移弃上清,加入150 $\mu$ l室温1X Wash Buffer III,涡旋混匀,室温孵育2min,期间涡旋混匀30s后静置30s,交替进行,确保充分混匀。

[0135] 4、将PCR管瞬时离心后置于磁力架上1min,待液体完全澄清后吸取移弃上清,之后换用10 $\mu$ l吸头移去少量残余Buffer。

[0136] 5、将PCR管从磁力架上移出,加入22.5 $\mu$ l Nuclease Free Water,使用移液器轻柔吹吸10次,确保混合均匀,转移全部液体至一个新的0.2ml PCR管中。

[0137] 后续的PCR扩增及文库纯化和定量步骤按照NadPrep<sup>TM</sup> DNA文库构建试剂盒 (for MGI) (201909Version2.0)的说明书进行操作即可。

[0138] 本实施例测试发现在Index区域是用通用碱基时,两端固定的区域碱基修饰的多少和通用封闭序列的使用浓度对最终的封闭效果影响加大,偏少的修饰碱基(5+7)时达到最佳封闭效果需要加入200 $\mu$ mol/L(封闭序列和文库比例是20:1);偏多的修饰碱基(7+10)是100 $\mu$ mol/L就能达到最佳效果,达到200 $\mu$ mol/L时已经产生了抑制,如图5所示,可能是局部区域产生强的结合能力,当加入的通用封闭序列过多时,在反应体系中碰撞机会增加,从而形成异常的十字星结构(如图4B所示)。

[0139] 实施例2

[0140] 实施例2与实施例1的步骤相同,唯一不同的在于,所采用的通用封闭序列的修饰碱基数量不同。本实施例中,通用封闭序列的修饰碱基数目如下表所示:

[0141] 表8:

封闭组合	4+6修饰封闭(封闭方案3)	8+11修饰组合(封闭方案4)
浓度( $\mu$ mol/L)	200	100

[0143] 在封闭方案3中,P1序列不变(SEQ ID NO:3所示的P1端封闭序列,

[0144] CTCTCA+GTACG+TCA+GCA+GT+T10XXXXXXXXXXCA+ACTCCT+TGGC+TCACAGA+ACGA+CATGG+CTACGATC+CGACTT/3SpC3/),P2序列为:SEQ ID NO:7

[0145] GCA+TGGC+GA+CCTT+ATCAGXXXXXXXXXXTTGTCTT+CCTA+AGA+CCGC+TTG+GCCTCCGA+CTT/3SpC3/。

[0146] 在封闭方案4中,P1封闭序列为SEQ ID NO:8

[0147] CTC+TCA+GT+ACG+TCA+G+CA+GT+TXXXXXXXXXXCA+ACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGA+CTT/3SpC3/;P2封闭序列为SEQ ID NO:9

[0148] GCA+TG+GC+GA+CC+TT+AT+CA+GXXXXXXXXXXTTG+TCTT+CCTA+AGA+CC+GC+TTG+GCC+TC+C+GA+CTT/3SpC3/。

[0149] 方案3是在方案1的基础上保留P1的修饰不变,把P2的标签两端的封闭修饰碱基数减少一个,修饰额的数量是4+6个;方案4是在方案是在方案3的基础上在标签两端的通用封闭各加一个封闭修饰碱基。当标签序列左侧的封闭修饰碱基是4个,右侧是6个时封闭效果明显不如方案1的5+7组合方案。同时方案4的8+11修饰方式也比方案2的7+10方案封闭效

果差,结果见图6。所以左侧非标签区封闭序列包括5~7个修饰的碱基,和右侧非标签区封闭序列包括7~10个修饰的碱基是效果更好的方案。

[0150] 实施例3

[0151] 实施例3与实施例1的步骤相同,通用封闭序列的修饰碱基数目也与实施例1的方案2相同,唯一不同之处在于,通用封闭序列的修饰碱基位置不同,具体序列如下:

[0152] 封闭方案5:P1封闭序列为SEQ ID NO:10

[0153] CTC+T+CA+GT+ACG+TCA+GCA+GTTXXXXXXXXXXCAACTCCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATC+CGA+CTT/3SpC3/,P2封闭序列为SEQ ID NO:11

[0154] G+CA+TG+GC+GA+CC+TT+ATCAGXXXXXXXXXXTTGTCTT+CCTA+AGA+CC+GC+TTG+GCC+TC+C+GA+CTT/3SpC3/。

[0155] 封闭方案6:P1封闭序列为SEQ ID NO:12

[0156] CTCTCA+GT+ACG+TCA+G+CA+GT+TXXXXXXXXXXCA+ACT+CCT+TGGC+TCAC+AGA+ACGA+CAT+GG+CTAC+GATCCGACTT/3SpC3/;P2封闭序列为SEQ ID NO:13

[0157] GCATG+GC+GA+CC+TT+AT+CA+GXXXXXXXXXXTTG+TCTT+C+CTA+AGA+CC+GC+TTG+GCC+TCC+GACTT/3SpC3/。

[0158] 封闭方案5是与方案2修饰数量相同的基础上改变了修饰的位置,方案5的是减少了两端的各一个修饰,增加了离中间标签比较近的位置的修饰;封闭方案6是正好相反,增加了两端的封闭修饰碱基,减少了离中间标签比较近的修饰。通用封闭方案5,方案6与方案2的封闭序列使用浓度都是100 $\mu\text{mol/L}$ ,用相同的投入量文库杂交。结果发现方案5和方案2的效果接近,方案6的效果变差,见图7所示,说明不仅是修饰的碱基数量影响封闭效果,修饰的位置了对封闭有影响,本发明发现均衡修饰和离中间标签序列比较近的区域增加修饰会明显好于两个末端增加修饰数量的效果好。

[0159] 实施例4多个文库杂交测试

[0160] 建库和杂交的实验步骤和实施例1一样,通用封闭选择7+10的碱基修饰,浓度用100 $\mu\text{mol/L}$ ,区别是多个文库混合杂交,具体文库投入数量及文库投入总量如下:

[0161] 表7.

[0162] 投入文库数量 (500ng/文库)	10个	12个	14个	16个
文库投入量总量	5 $\mu\text{g}$	6 $\mu\text{g}$	7 $\mu\text{g}$	8 $\mu\text{g}$

[0163] 在杂交捕获时投入量是500ng/文库时测序的各项指标表现比较好,如果单次杂交能够允许投入更多的量,也就是单次杂交更多的样本会降低每个样本的杂交捕获成本。本申请的通用封闭序列的测试发现单次不超过12个样本一起杂交时,杂交的各项指标表现较好,而杂交的文库数量达到14个和16个样本时,中靶率会有一定程度的下降。如图8所示,14个文库和16个文库的投入总量分别是7 $\mu\text{g}$ 和8 $\mu\text{g}$ ,在固定的捕获体系中,随着文库数量的增加,部分文库有机会出现两个文库与通用封闭形成十字星结构封闭,从而降低中靶率。

[0164] 综上所述,本发明针对MGI平台的双端Index文库杂交设计了通用的封闭序列,通过在Index的可变区域用通用碱基替代,并在index两侧的固定序列中增加提升退火温度的碱基修饰,能够提高双端index文库的捕获效率。进一步地,本申请还发现通用封闭序列的封闭的效果与提升退火温度的修饰碱基的数量控制在左侧为5~7个,右侧为7~10个时,封闭序列的封闭效果更好。此外还发现如果进一步优化修饰碱基的具体位置为本申请实施例

的优选位置时,封闭效果最好。相应地,通用封闭序列的使用浓度也对捕获目的文库的中靶率有影响。当通用封闭序列的用量为0.4~0.8 $\mu$ g时,本申请可支持同时进行12个样本一起杂交,这大大降低了单个样本的杂交捕获成本。

[0165] 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。

- [0001] 序列表
- [0002] <110> 纳昂达(南京)生物科技有限公司
- [0003] <120> 通用封闭序列及其应用
- [0004] <130> PN132184NAGD
- [0005] <160> 13
- [0006] <170> SIPOSequenceListing 1.0
- [0007] <210> 1
- [0008] <211> 72
- [0009] <212> DNA
- [0010] <213> 人工序列(Artificial Sequence)
- [0011] <220>
- [0012] <221> misc\_feature
- [0013] <222> (21) .. (30)
- [0014] <223> 标签封闭序列,n表示A,T,C,或G;
- [0015] <400> 1
- [0016] ctctcagtac gtcagcagtt nnnnnnnnnn caactccttg gctcacagaa cgacatggct 60
- [0017] acgatccgac tt 72
- [0018] <210> 2
- [0019] <211> 60
- [0020] <212> DNA
- [0021] <213> 人工序列(Artificial Sequence)
- [0022] <220>
- [0023] <221> misc\_feature
- [0024] <222> (19) .. (28)
- [0025] <223> 标签封闭序列,n表示A,T,C或G
- [0026] <400> 2
- [0027] gcatggcgac cttatcagnn nnnnnnnntt gtcttcctaa gaccgcttgg cctccgactt 60
- [0028] <210> 3
- [0029] <211> 72
- [0030] <212> DNA
- [0031] <213> 人工序列(Artificial Sequence)
- [0032] <220>
- [0033] <221> misc\_feature
- [0034] <222> (21) .. (30)
- [0035] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0036] <220>
- [0037] <221> misc\_feature
- [0038] <222> (1) .. (20)

- [0039] <223> 左侧非标签区封闭序列,第6、12、15、18和20位的碱基为LAN或BNA修饰;
- [0040] <220>
- [0041] <221> misc\_feature
- [0042] <222> (31) .. (72)
- [0043] <223> 右侧非标签区封闭序列,第33、39、43、50、54、59和67位的碱基为LAN或BNA修饰,第72位末端为C3间隔臂修饰;
- [0044] <400> 3
- [0045] ctctcagtac gtcagcagtt nnnnnnnnnn caactccttg gctcacagaa cgacatggct 60
- [0046] acgatccgac tt 72
- [0047] <210> 4
- [0048] <211> 60
- [0049] <212> DNA
- [0050] <213> 人工序列(Artificial Sequence)
- [0051] <220>
- [0052] <221> misc\_feature
- [0053] <222> (19) .. (28)
- [0054] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0055] <220>
- [0056] <221> misc\_feature
- [0057] <222> (1) .. (18)
- [0058] <223> 左侧非标签区封闭序列,第4、8、10、14和18的碱基为LAN或BNA修饰;
- [0059] <220>
- [0060] <221> misc\_feature
- [0061] <222> (1) .. (18)
- [0062] <223> 右侧非标签区封闭序列,第36、40、43、47、50、53和58位的碱基为LAN或BNA修饰;
- [0063] <400> 4
- [0064] gcatggcgac cttatcagnn nnnnnnnntt gtcttctctaa gaccgcttgg cctccgactt 60
- [0065] <210> 5
- [0066] <211> 72
- [0067] <212> DNA
- [0068] <213> 人工序列(Artificial Sequence)
- [0069] <220>
- [0070] <221> misc\_feature
- [0071] <222> (21) .. (30)
- [0072] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0073] <220>
- [0074] <221> misc\_feature
- [0075] <222> (1) .. (20)

- [0076] <223> 左侧非标签区封闭序列,第4、7、9、12、15、18和20位的碱基为LAN或BNA修饰;
- [0077] <220>
- [0078] <221> misc\_feature
- [0079] <222> (31) .. (72)
- [0080] <223> 右侧非标签区封闭序列,第33、39、43、47、50、54、57、59、63和67位的碱基为LAN或BNA修饰,第72位末端为C3间隔臂修饰;
- [0081] <400> 5
- [0082] ctctcagtac gtcagcagtt nnnnnnnnnn caactccttg gctcacagaa cgacatggct 60
- [0083] acgatccgac tt 72
- [0084] <210> 6
- [0085] <211> 60
- [0086] <212> DNA
- [0087] <213> 人工序列(Artificial Sequence)
- [0088] <220>
- [0089] <221> misc\_feature
- [0090] <222> (19) .. (28)
- [0091] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0092] <220>
- [0093] <221> misc\_feature
- [0094] <222> (1) .. (18)
- [0095] <223> 左侧非标签区封闭序列,第4、6、8、10、12、14和18的碱基为LAN或BNA修饰;
- [0096] <220>
- [0097] <221> misc\_feature
- [0098] <222> (29) .. (60)
- [0099] <223> 右侧非标签区封闭序列,第32、36、40、43、45、47、50、53、56和58位的碱基为LAN或BNA修饰,第60位末端为C3间隔臂修饰;
- [0100] <400> 6
- [0101] gcatggcgac cttatcagnn nnnnnnnntt gtcttctctaa gaccgcttgg cctccgactt 60
- [0102] <210> 8
- [0103] <211> 60
- [0104] <212> DNA
- [0105] <213> 人工序列(Artificial Sequence)
- [0106] <220>
- [0107] <221> misc\_feature
- [0108] <222> (19) .. (28)
- [0109] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0110] <220>
- [0111] <221> misc\_feature

- [0112] <222> (1) .. (18)
- [0113] <223> 左侧非标签区封闭序列,第4、8、10、14位碱基为LAN或BNA修饰;
- [0114] <220>
- [0115] <221> misc\_feature
- [0116] <222> (29) .. (60)
- [0117] <223> 右侧非标签区封闭序列,第36、40、43、47、50和58位的碱基为LAN或BNA修饰,第60位末端为C3间隔臂修饰;
- [0118] <400> 8
- [0119] gcatggcgac cttatcagnn nnnnnnnntt gtcttcctaa gaccgcttgg cctccgactt 60
- [0120] <210> 8
- [0121] <211> 72
- [0122] <212> DNA
- [0123] <213> 人工序列(Artificial Sequence)
- [0124] <220>
- [0125] <221> misc\_feature
- [0126] <222> (21) .. (30)
- [0127] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0128] <220>
- [0129] <221> misc\_feature
- [0130] <222> (1) .. (20)
- [0131] <223> 左侧非标签区封闭序列,第4、7、9、12、15、16、18和20位的碱基为LAN或BNA修饰;
- [0132] <220>
- [0133] <221> misc\_feature
- [0134] <222> (31) .. (72)
- [0135] <223> 右侧非标签区封闭序列,第33、39、43、47、50、54、57、59、63、67、70位碱基为LNA或BAN修饰,第72位碱基末端进行C3间隔臂封闭修饰
- [0136] <400> 8
- [0137] ctctcagtac gtcagcagttnnnnnnnnnn caactccttg gctcacagaa cgacatggct 60
- [0138] acgatccgac tt 72
- [0139] <210> 9
- [0140] <211> 60
- [0141] <212> DNA
- [0142] <213> 人工序列(Artificial Sequence)
- [0143] <220>
- [0144] <221> misc\_feature
- [0145] <222> (19) .. (28)
- [0146] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0147] <220>

- [0148] <221> misc\_feature
- [0149] <222> (1) .. (18)
- [0150] <223> 左侧非标签区封闭序列,第4、6、8、10、12、14、16和18的碱基为LAN或BNA修饰;
- [0151] <220>
- [0152] <221> misc\_feature
- [0153] <222> (29) .. (60)
- [0154] <223> 右侧非标签区封闭序列,第32、36、40、43、45、47、50、53、55、56和58位的碱基为LAN或BNA修饰,第60位末端为C3间隔臂修饰;
- [0155] <400> 9
- [0156] gcatggcgac cttatcagmn nnnnnnnntt gtcttcctaa gaccgcttgg cctccgactt 60
- [0157] <210> 10
- [0158] <211> 72
- [0159] <212> DNA
- [0160] <213> 人工序列(Artificial Sequence)
- [0161] <220>
- [0162] <221> misc\_feature
- [0163] <222> (21) .. (30)
- [0164] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0165] <220>
- [0166] <221> misc\_feature
- [0167] <222> (1) .. (20)
- [0168] <223> 签区封闭序列,第4、5、7、9、12、15和18位的碱基为LAN或BNA修饰;
- [0169] <220>
- [0170] <221> misc\_feature
- [0171] <222> (31) .. (72)
- [0172] <223> 右侧非标签区封闭序列,第39、43、47、50、54、57、59、63、67、70位碱基为LNA或BAN修饰,第72位碱基末端进行C3间隔臂封闭修饰
- [0173] <400> 10
- [0174] ctctcagtac gtcagcagtt nnnnnnnnnn caactccttg gctcacagaa cgacatggct 60
- [0175] acgatccgac tt 72
- [0176] <210> 11
- [0177] <211> 60
- [0178] <212> DNA
- [0179] <213> 人工序列(Artificial Sequence)
- [0180] <220>
- [0181] <221> misc\_feature
- [0182] <222> (19) .. (28)
- [0183] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;

- [0184] <220>
- [0185] <221> misc\_feature
- [0186] <222> (1) .. (18)
- [0187] <223> 左侧非标签区封闭序列,第2、4、6、8、10、12和14位碱基为LAN或BNA修饰;
- [0188] <220>
- [0189] <221> misc\_feature
- [0190] <222> (29) .. (60)
- [0191] <223> 右侧非标签区封闭序列,第36、40、43、45、47、50、53、55、56和58位的碱基为LAN或BNA修饰,第60位末端为C3间隔臂修饰;
- [0192] <400> 11
- [0193] gcatggcgac cttatcagmn nnnnnnnntt gtcttcctaa gaccgcttgg cctccgactt 60
- [0194] <210> 12
- [0195] <211> 72
- [0196] <212> DNA
- [0197] <213> 人工序列(Artificial Sequence)
- [0198] <220>
- [0199] <221> misc\_feature
- [0200] <222> (21) .. (30)
- [0201] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;
- [0202] <220>
- [0203] <221> misc\_feature
- [0204] <222> (1) .. (20)
- [0205] <223> 签区封闭序列,第7、9、12、15、16、18和20位的碱基为LAN或BNA修饰;
- [0206] <220>
- [0207] <221> misc\_feature
- [0208] <222> (31) .. (72)
- [0209] <223> 右侧非标签区封闭序列,第33、36、39、43、47、50、54、57、59和63位碱基为LNA或BAN修饰,第72位碱基末端进行C3间隔臂封闭修饰
- [0210] <400> 12
- [0211] ctctcagtac gtcagcagtt nnnnnnnnnn caactccttg gctcacagaa cgacatggct 60
- [0212] acgatccgac tt 72
- [0213] <210> 13
- [0214] <211> 60
- [0215] <212> DNA
- [0216] <213> 人工序列(Artificial Sequence)
- [0217] <220>
- [0218] <221> misc\_feature
- [0219] <222> (19) .. (28)
- [0220] <223> 标签区封闭序列,n表示次黄嘌呤或C3间隔臂;

- [0221] <220>
- [0222] <221> misc\_feature
- [0223] <222> (1) .. (18)
- [0224] <223> 左侧非标签区封闭序列,第6、8、10、12、14、16和18的碱基为LAN或BNA修饰;
- [0225] <220>
- [0226] <221> misc\_feature
- [0227] <222> (29) .. (60)
- [0228] <223> 右侧非标签区封闭序列,第32、36、37、40、43、45、47、50、53和56位的碱基为LAN或BNA修饰,第60位末端为C3间隔臂修饰;
- [0229] <400> 13
- [0230] gcatggcgac cttatcagmn nnnnnnnntt gtcttcctaa gaccgcttgg cctccgactt 60

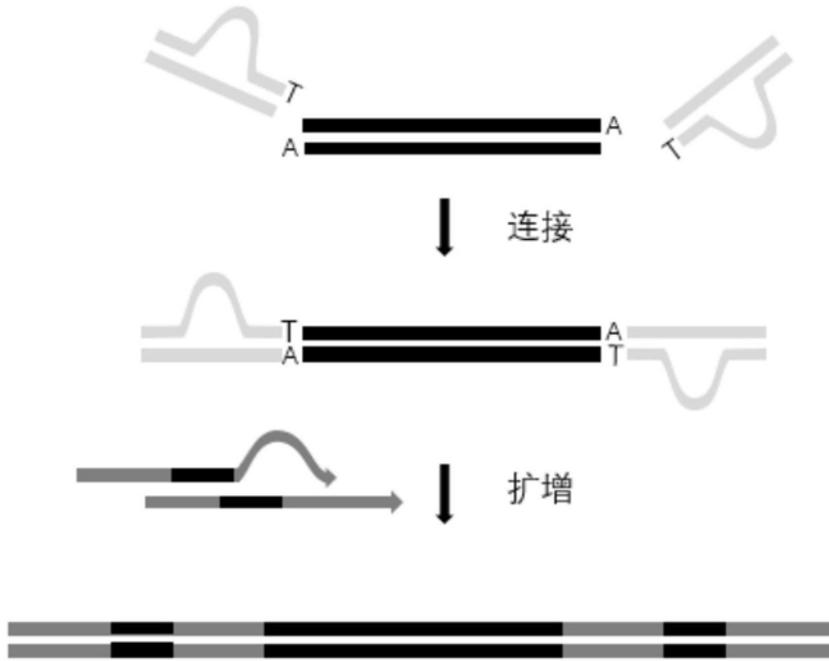


图1

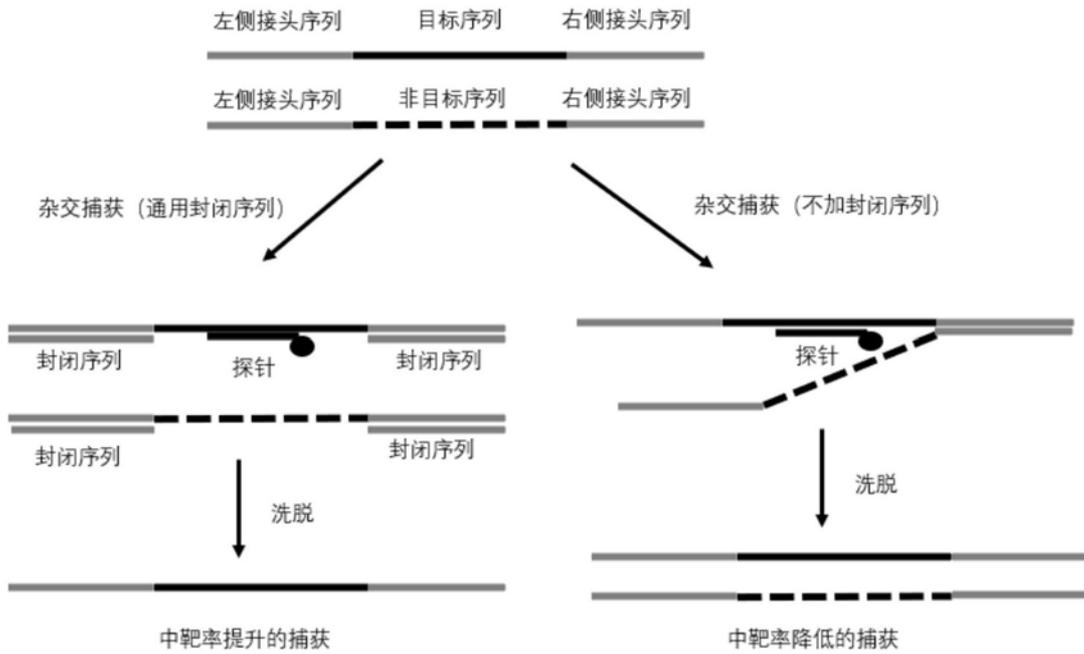


图2

	Index2-1	Index2-2	Index2-3	Index2-4	Index2-5	Index2-6	Index2-7	Index2-8	Index2-9	Index2-10	Index2-11	Index2-12	干扰数据
Index1-1	10000	0.25	0.26	0.55	0.34	0.23	0.3	0.73	0.38	0.43	0.53	0.45	4.45
Index1-2	0.32	10000	3.65	0.51	0.42	0.24	0.48	0.24	0.28	0.36	0.58	0.52	7.6
Index1-3	0.26	0.54	10000	2.25	0.92	0.28	0.44	0.3	0.55	0.41	0.54	0.39	6.88
Index1-4	0.51	0.21	0.46	10000	0.74	0.29	0.26	0.99	0.33	0.41	0.52	0.4	5.12
Index1-5	0.35	0.27	0.37	0.84	10000	0.25	0.42	0.26	0.5	0.37	0.47	0.36	4.46
Index1-6	0.24	0.33	0.36	0.27	0.22	10000	0.43	0.26	0.25	0.27	0.29	0.32	3.24
Index1-7	0.34	0.33	0.41	0.38	0.21	0.23	10000	0.29	0.19	0.27	0.42	0.36	3.43
Index1-8	0.54	0.21	0.22	0.81	0.26	0.22	0.28	10000	1.97	0.93	0.32	0.33	6.09
Index1-9	0.41	0.29	0.45	0.44	0.36	0.27	0.32	0.62	10000	1.68	0.43	0.45	5.72
Index1-10	0.54	0.23	0.35	0.39	0.39	0.21	0.31	0.49	0.42	100000	0.49	0.44	4.26
Index1-11	0.35	0.33	0.24	0.27	0.28	0.2	0.35	0.16	0.23	0.31	10000	0.31	3.03
Index1-12	0.27	0.25	0.17	0.29	0.28	0.29	0.36	0.22	0.28	0.33	0.37	10000	3.11
干扰数据	4.13	3.24	6.94	7	4.42	2.71	3.95	4.56	5.38	5.77	4.96	4.33	

图3

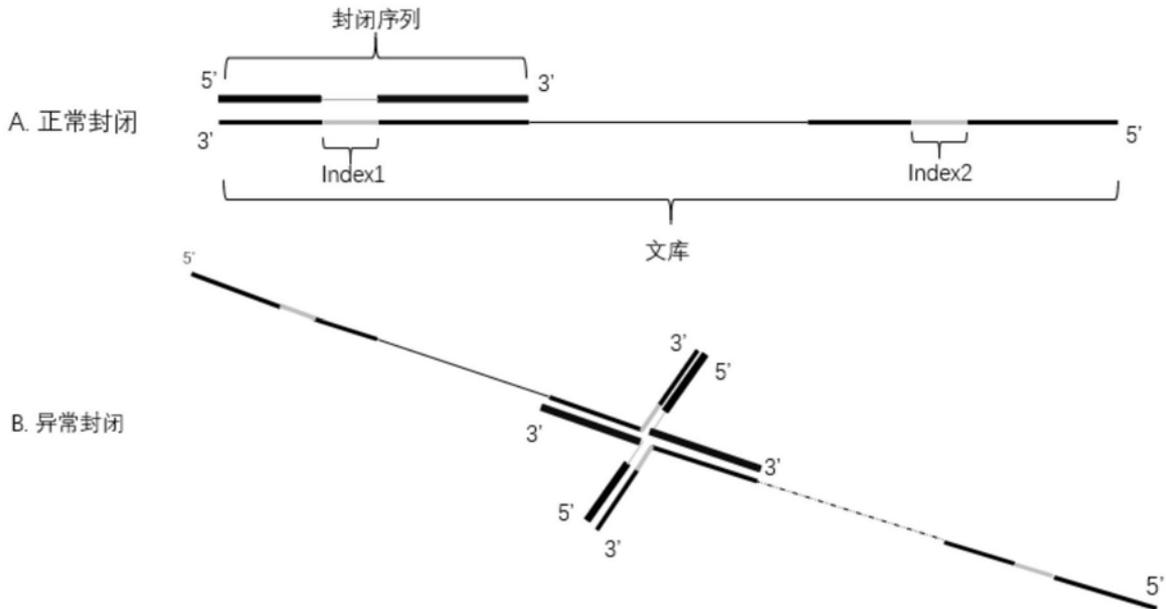


图4A和图4B

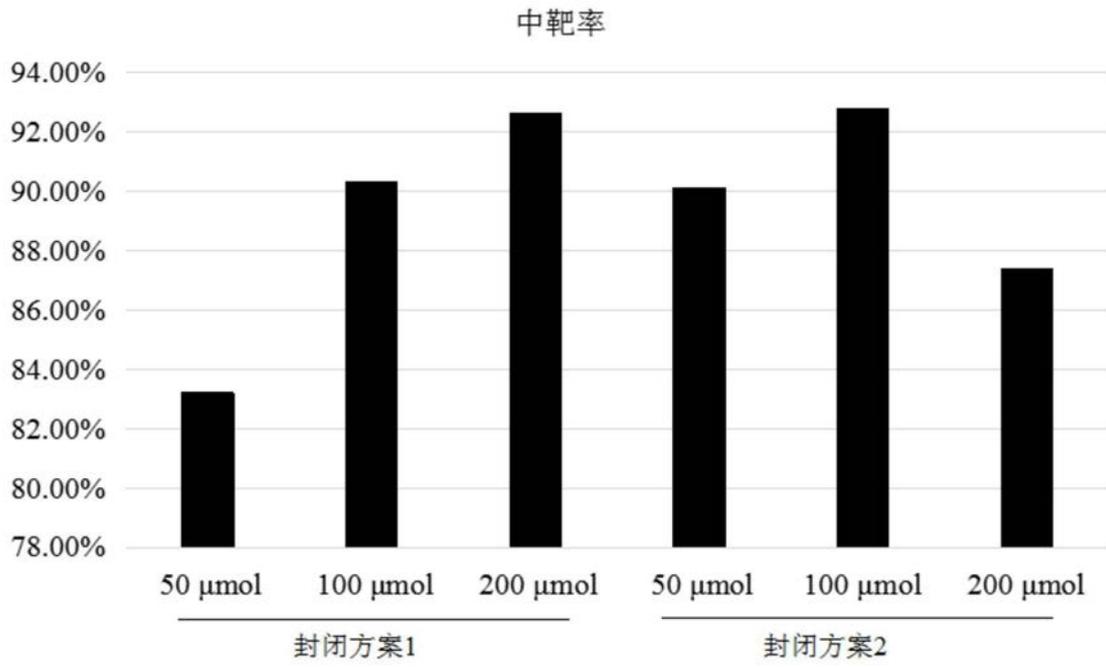


图5

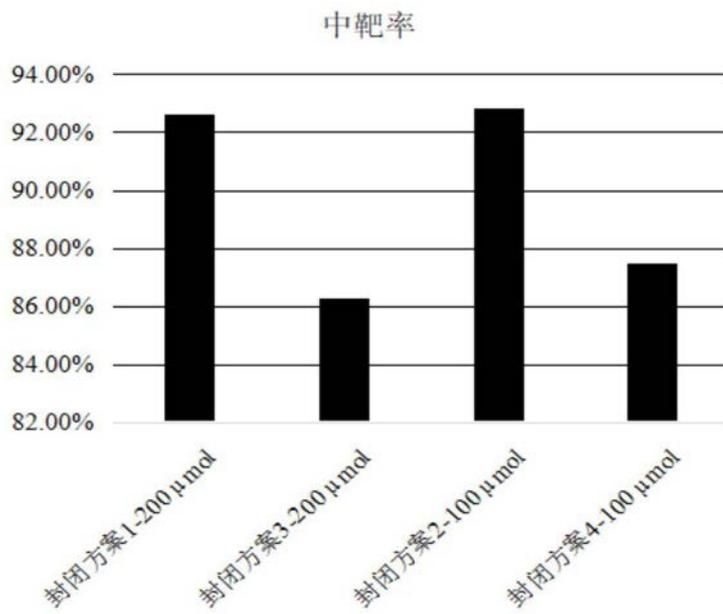


图6

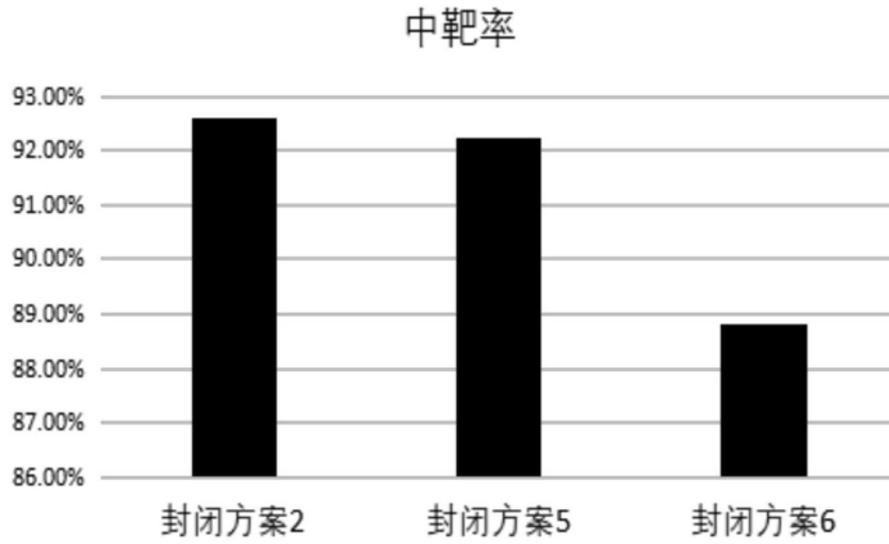


图7

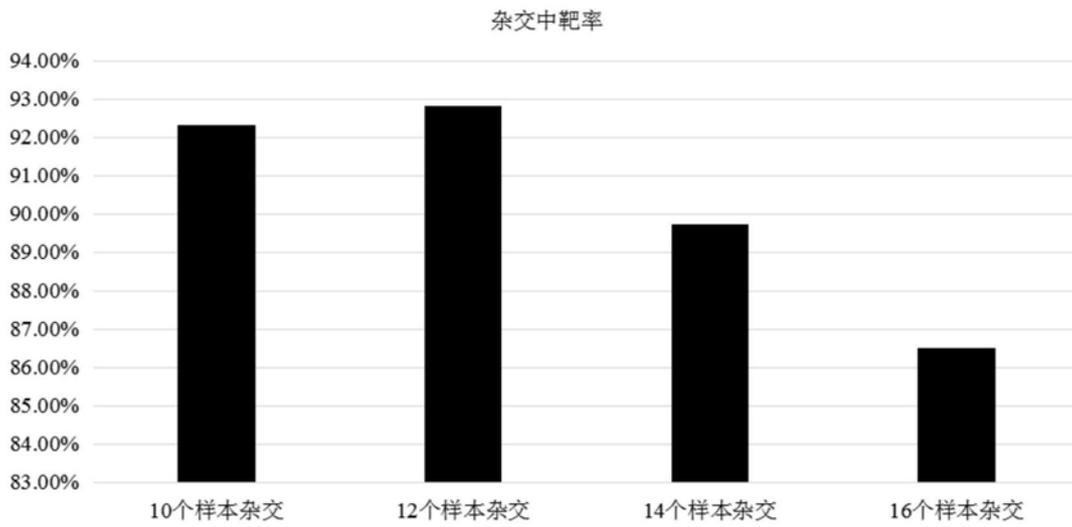


图8