



(12) 发明专利申请

(10) 申请公布号 CN 114277114 A

(43) 申请公布日 2022.04.05

(21) 申请号 202111646690.4

C12Q 1/6858 (2018.01)

(22) 申请日 2021.12.30

C12Q 1/6886 (2018.01)

(71) 申请人 深圳海普洛斯医学检验实验室

地址 518000 广东省深圳市南山区西丽街
道松坪山路1号源兴科技大厦南座
1106

(72) 发明人 许明炎 张晓妮 周书雄

(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 李小焦 彭家恩

(51) Int. Cl.

C12Q 1/6869 (2018.01)

C12N 15/11 (2006.01)

G40B 50/06 (2006.01)

C12Q 1/6806 (2018.01)

权利要求书3页 说明书14页
序列表3页

(54) 发明名称

一种扩增子测序添加唯一性标识符的方法
及应用

(57) 摘要

本申请公开了一种扩增子测序添加唯一性标识符的方法及应用。本申请方法包括,用第一引物对模板核酸进行一次延伸;第一引物包括平台上游引物结合区、UMI和靶标上游引物序列,且平台上游引物结合区和靶标上游引物序列的T替换为dU;然后,向反应体系加入第二引物进行一次延伸;第二引物包括平台下游引物结合区和靶标下游引物序列;然后,向反应体系加入UDG/UNG酶消化dU;消化后,向反应体系加入第三引物进行PCR;第三引物为第一引物平台上游引物结合区序列。本申请方法,通过对引物进行特殊设计,依序向反应体系中加入三种引物,实现使每个原始模板核酸只对应一个UMI,校正每个靶点扩增性偏差,实现靶标基因拷贝数定量检测。

1. 一种扩增子测序添加唯一性标识符的方法,其特征在于:包括以下步骤,

配制反应体系,采用第一引物对模板核酸进行一次延伸,得到互补链;所述第一引物由5'端到3'端依序包括测序平台上游引物结合区、唯一性标识符和靶标特异性上游引物序列;并且,所述第一引物中,测序平台上游引物结合区和靶标特异性上游引物序列中的碱基T替换为脱氧尿嘧啶,所述测序平台上游引物结合区对应测序平台的上游测序引物的3'末端;

第一引物延伸结束后,向反应体系中加入第二引物,利用第二引物对第一引物延伸的互补链进行一次延伸,获得由测序平台上游引物结合区、唯一性标识符、靶标序列和测序平台下游引物结合区组成的产物;所述第二引物由5'端到3'端依序包括测序平台下游引物结合区和靶标特异性下游引物序列,所述测序平台下游引物结合区对应测序平台的下游测序引物的3'末端;

第二引物延伸结束后,向反应体系中加入UDG/UNG酶消化脱氧尿嘧啶,从而消化第一引物以及第一引物的延伸链;

UDG/UNG酶消化完成后,向反应体系中加入第三引物,利用第三引物和第二引物对第二引物延伸的产物进行PCR扩增富集,获得所述模板核酸的所有扩增子都添加相同唯一性标识符的产物;所述第三引物为第一引物的测序平台上游引物结合区自5'端起的全部或部分序列,并且,第三引物中的碱基T不被替换为脱氧尿嘧啶。

2. 根据权利要求1所述的方法,其特征在于:所述第一引物中,唯一性标识符的序列中插入有至少一个脱氧尿嘧啶,通过插入脱氧尿嘧啶的分隔,使得唯一性标识符的连续碱基数小于5;

优选的,所述PCR扩增富集的扩增循环数大于或等于5。

3. 一种基于权利要求1或2所述方法的测序文库构建方法,其特征在于:包括以下步骤,对所述PCR扩增富集的产物进行纯化,获得纯化产物;

采用第四引物和第五引物,对纯化产物进行建库扩增,获得测序文库;所述第四引物为带有测序接头和Barcode的测序平台上游测序引物,所述第五引物为带有测序接头和Barcode的测序平台下游测序引物。

4. 根据权利要求3所述的测序文库构建方法,其特征在于:所述纯化为磁珠法纯化、过柱法纯化和凝胶法纯化中的至少一种。

5. 一种扩增子测序添加唯一性标识符的试剂盒,其特征在于:包括第一引物、第二引物、第三引物和UDG/UNG酶;

所述第一引物由5'端到3'端依序包括测序平台上游引物结合区、唯一性标识符和靶标特异性上游引物序列;并且,所述第一引物中,测序平台上游引物结合区和靶标特异性上游引物序列中的碱基T替换为脱氧尿嘧啶,所述测序平台上游引物结合区对应测序平台的上游测序引物的3'末端;

所述第二引物由5'端到3'端依序包括测序平台下游引物结合区和靶标特异性下游引物序列,所述测序平台下游引物结合区对应测序平台的下游测序引物的3'末端;

所述第三引物为第一引物的测序平台上游引物结合区自5'端起的全部或部分序列,并且,第三引物中的碱基T不被替换为脱氧尿嘧啶。

6. 根据权利要求5所述的试剂盒,其特征在于:还包括PCR扩增试剂。

7. 一种基于权利要求1或2所述方法进行测序文库构建的试剂盒,其特征在于:包括权利要求5或6所述的试剂盒,以及第四引物和第五引物;

所述第四引物为带有测序接头和Barcode的测序平台上游测序引物,所述第五引物为带有测序接头和Barcode的测序平台下游测序引物。

8. 根据权利要求1或2所述的方法或权利要求3或4所述的测序文库构建方法在制备肺癌EGFR L858R和19De1突变率检测试剂或PIK3CA基因突变率检测试剂中的应用。

9. 一种用于肺癌EGFR L858R和19De1突变率检测的试剂盒,其特征在于:包括第一引物、第二引物、第三引物、第四引物和第五引物;

所述第一引物由Seq ID No.1所示序列的EGFR L858R检测上游引物和Seq ID No.2所示序列的EGFR 19De1检测上游引物组成,

Seq ID No.1:

5' -CACGACGCUCUCCGAUCUNNNUNNNNCGUACUGGUGAAAACACCGCA-3' ,

Seq ID No.2:

5' -CACGACGCUCUCCGAUCUNNNUNNNNACUCUGGAUCCCAGAAGGUG-3' ;

所述第二引物由Seq ID No.3所示序列的EGFR L858R检测下游引物和Seq ID No.4所示序列的EGFR 19De1检测下游引物组成,

Seq ID No.3:

5' -AGACGTGTGCTCTCCGATCTGAACTCACATCGAGGATT-3' ,

Seq ID No.4:

5' -AGACGTGTGCTCTCCGATCTGAACTCACATCGAGGATT-3' ;

所述第三引物为Seq ID No.5所示序列,

Seq ID No.5:5' -CACGACGCTCTCCGATCT-3' ;

所述第四引物为Seq ID No.6所示序列,

Seq ID No.6:

5' -AATGATACGGCGACCACCGAGATCTACACNNNNNACACTCTTCCCTACACGACGCTCTCCGATCT-3' ;

所述第五引物为Seq ID No.7所示序列,

Seq ID No.7:

5' -CAAGCAGAAGACGGCATAACGAGATNNNNNNGTACTGGAGTTCAGACGTGTGCTCTCCGATCT-3' 。

10. 一种用于PIK3CA基因突变率检测的试剂盒,其特征在于:包括第一引物、第二引物、第三引物、第四引物和第五引物;

所述第一引物为Seq ID No.8所示序列,

Seq ID No.8:

5' -CACGACGCUCUCCGAUCUNNNUNNNNAGCAAUUCUACACGAGAUCUCUCU-3' ;

所述第二引物为Seq ID No.9所示序列,

Seq ID No.9:

5' -AGACGTGTGCTCTCCGATCTCTGGGCTACTTCATCTCTTGAAT-3' ;

所述第三引物为Seq ID No.5所示序列,

Seq ID No.5:5' -CACGACGCTCTCCGATCT-3' ;

所述第四引物为Seq ID No.6所示序列，

Seq ID No.6:

5' -AATGATACGGCGACCACCGAGATCTACACNNNNNACACTCTTCCCTACACGACGCTCTCCGATCT-
3' ;

所述第五引物为Seq ID No.7所示序列，

Seq ID No.7:

5' -CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-3' 。

一种扩增子测序添加唯一性标识符的方法及应用

技术领域

[0001] 本申请涉及测序技术领域,特别是涉及一种扩增子测序添加唯一性标识符的方法及应用。

背景技术

[0002] 高通量测序技术具有低测序成本、高测序通量的优势,最高可同时并行处理几百万条DNA分子,彻底改变了传统基因组研究方法,在基因检测方面有巨大的应用价值。随着二代测序代表公司美国Illumina开发的MiSeq测序平台通过美国FDA技术认证,高通量测序开始从科研领域走向临床领域。近年来,基于二代测序平台的肿瘤基因检测试剂盒陆续获批,华大智造开发的部分测序仪通过医疗器械备案,标志着高通量测序技术正在向临床诊断应用迈进。同时,在科研上,高通量测序技术助力对各个物种进行基因组学研究、转录组学研究、表观遗传学等研究,为科研学术研究提供了有利工具和方法。

[0003] 高通量测序临床应用主要技术有扩增子测序技术和液相杂交测序技术,两种方法各有优劣。其中,扩增子测序是一种高靶向性方法,可用于分析特定基因组区域中的基因变异。扩增子的超深度测序可以有效地识别变异并对其进行特征分析。可用于目标区域扩增测序、微生物测序,如分析细菌、真菌、古菌分类方面物种之间的差异等。此外,扩增子测序在海洋、土壤、肠道等微生物研究中也具有广泛应用。

[0004] 一般来说,扩增子测序技术基于多重PCR扩增,结合测序技术一次性可检测多个靶点,不仅可检测已知的点突变、插入缺失、部分拷贝数变异、结构变异等信息,还可检测扩增片段中的未知突变序列。相比于使用探针捕获靶标的液相杂交测序技术,扩增子测序技术优势在于:1.起始模板量要求低,DNA可以低至1ng;2.步骤简单,操作简单,实验耗时短,最快只需要1-2h即可完成建库;3.特异性强,基本可达95%以上;4.基于PCR检测,灵敏度高,目前市售PCR检测产品灵敏度最高可达0.1%;5.靶向针对性强,数据量需求少,有效降低检测成本。

[0005] 虽然扩增子测序技术有上述优点,但在某些问题上也存在缺点,例如该技术在PCR扩增环节可能出现扩增偏好性,在建库和测序过程中可能出现碱基错误配对,实验过程中可能会引入人源误差。这些缺点都是扩增子测序建库无法忽视的技术缺点,会导致测序结果只能定性无法定量、超低频阳性突变是否假阳性等问题。

[0006] 为了克服上述问题,有必要在靶向扩增之前将唯一性标识符(缩写UMI)引入至起始DNA模板中,让每一个原始的DNA模板都带有一个唯一性的分子标签。分析测序数据时,相同UMI的DNA序列为同一个模板来源,从而保留起始DNA分子的唯一性,解决PCR扩增偏差的问题。归类所有UMI,可准确分析原始模板中某条序列对应的真实数量,达到定量的目的。对于靶向扩增、建库及测序中引入的测序错误,或者人为因素引入的测序错误,通过添加UMI可起到矫正作用,相同的UMI连接的DNA中都有同一的多样性位点,若仅有其中的一个或者几个UMI含有多多样性位点,此多样性可判断为扩增错误或者人为引入的假阳性。加入UMI的扩增子测序方法具有良好的准确性、特异性、灵敏性和通用性。

[0007] 主流应用中,以RNA为起始模板的扩增子测序在反转录的时候利用反转录引物添加UMI,得到唯一性标签的cDNA模板,再根据cDNA进行多重PCR,技术上容易实现。但在实际情况中,一些检测在DNA层面上更加有利且易实现,例如一些肿瘤基因突变检测等。其次, RNA样本在采集、保存、运输、核酸提取过程,相对DNA要求更加严格,操作更加繁琐,存在易降解的问题,在多样性研究中, RNA丰度与DNA相比也可能存在差距。因此,以基因组DNA为模板的扩增子测序还有开发的空间。目前,基于DNA模板的多重PCR的扩增子测序中,通常是给每条多重PCR特异性引物加上一条随机的UMI,即每条引物都有一个自己的UMI,导致在新一轮PCR的扩增中,子链会重新引入一个新的UMI,最终结果就是无法分辨相同靶序列是否来自同一条模板母链。例如,一般UMI是由随机N碱基构成的,每一条相同的特异性引物的UMI是“唯一”的,在反应过程中,母链会通过相同的特异性扩增引物,从一个母链扩增为两个、四个、八个、N个子链;但是,每次扩增的特异性引物带入的UMI都不一致,导致同一母链扩增的所有子链UMI均不一致,无法区分相同的插入序列是否来自于同一个母链模板,从而无法体现UMI的作用。

[0008] 因此,扩增子测序方法无法校正最重要的特异性扩增步骤,只能校正后续建库扩增中引入的错误,无法实现每一个原始模板对应一个UMI,通俗来讲,该类方法无法真正意义上实现绝对定量,只能有限地校正扩增偏差、有限地消除扩增错误。如何实现在扩增子测序中使每一个原始模板只对应一个UMI标签,是亟待解决的问题。

发明内容

[0009] 本申请的目的是提供一种新的扩增子测序添加唯一性标识符的方法及应用。

[0010] 为了实现上述目的,本申请采用了以下技术方案:

[0011] 本申请的一方面公开了一种扩增子测序添加唯一性标识符的方法,包括以下步骤:

[0012] 配制反应体系,采用第一引物对模板核酸进行一次延伸,得到互补链;第一引物由5'端到3'端依序包括测序平台上游引物结合区、唯一性标识符和靶标特异性上游引物序列;并且,第一引物中,测序平台上游引物结合区和靶标特异性上游引物序列中的碱基T替换为脱氧尿嘧啶(即dU,在序列中以U表示),测序平台上游引物结合区对应测序平台的上游测序引物的3'末端;

[0013] 第一引物延伸结束后,向反应体系中加入第二引物,利用第二引物对第一引物延伸的互补链进行一次延伸,获得由测序平台上游引物结合区、唯一性标识符、靶标序列和测序平台下游引物结合区组成的产物;第二引物由5'端到3'端依序包括测序平台下游引物结合区和靶标特异性下游引物序列,测序平台下游引物结合区对应测序平台的下游测序引物的3'末端;

[0014] 第二引物延伸结束后,向反应体系中加入UDG/UNG酶消化脱氧尿嘧啶,从而消化第一引物以及第一引物的延伸链;

[0015] UDG/UNG酶消化完成后,向反应体系中加入第三引物,利用第三引物和第二引物对第二引物延伸的产物进行PCR扩增富集,获得模板核酸的所有扩增子都添加相同唯一性标识符的产物;第三引物为第一引物的测序平台上游引物结合区自5'端起的全部或部分序列,并且,第三引物中的碱基T不被替换为脱氧尿嘧啶。

[0016] 本申请的方法中，“进行一次延伸”是指利用引物退火杂交到靶标序列上后，只进行这条引物的延伸即可，不进行再次的变性、退火。这样能够确保一个模板核酸母链被一个唯一的UMI标记。当然，在第二引物加入后，虽然设计的是第二引物进行退火杂交、延伸；但是，这个时候，第一引物也会退火杂交、延伸；不过，由于第二引物的退火杂交、延伸也是只“进行一次延伸”；这种情况下，由测序平台上游引物结合区、唯一性标识符、靶标序列和测序平台下游引物结合区组成的产物同样只有最开始第一引物延伸标记UMI的母链。最后，在第三引物和第二引物的指数倍PCR扩增富集下，也只有最开始第一引物延伸标记的UMI母链的扩增子能够被指数倍富集。并且，在第三引物和第二引物进行PCR扩增富集之前，利用UDG/UNG酶消化去除了第一引物，也避免了第一引物在新的一轮PCR扩增中再次引入新的UMI。本申请的模板核酸，可以是DNA或cDNA。

[0017] 需要说明的是，本申请的方法，通过对第一引物、第二引物和第三引物进行特殊设计，依序向反应体系中添加三种引物，可以实现对某一个模板核酸母链的所有扩增子链都添加相同的UMI，这对于突变检测尤其重要。例如，可以直接通过UMI确定相同的特异性引物扩增获得的扩增子链中哪些是来源于突变或非突变，从而对突变进行定量检测，获得准确的突变率。

[0018] 还需要说明的是，本申请的关键在于，引物结构的设计以及各引物的添加顺序的设计，使得最终扩增富集的扩增子中都带有相同的UMI；至于具体的引物序列，可以根据具体针对的靶标序列和具体的测序平台而定。例如，采用常规的引物设计软件，针对具体的靶标序列设计第一引物的靶标特异性上游引物序列，针对预计采用的测序平台设计第一引物的测序平台上游引物结合区，由此组成第一引物。

[0019] 本申请的一种实现方式中，第一引物中，唯一性标识符的序列中插入有至少一个脱氧尿嘧啶，通过插入脱氧尿嘧啶的分隔，使得唯一性标识符的连续碱基数量小于5。

[0020] 需要说明的是，本申请在引物中插入脱氧尿嘧啶或者利用脱氧尿嘧啶替换T，目的是为了在不需要的时候，利用UDG/UNG酶消化引物。在唯一性标识符中插入脱氧尿嘧啶，同样是为了尽量避免随机的UMI可能在后续扩增中出现非特异性扩增，唯一性标识符序列中可以选择使用一个或几个固定的脱氧尿嘧啶插入至其碱基序列中间，脱氧尿嘧啶左右两边的连续性N碱基数量小于5nt，这样能够更有效的避免非特异性扩增；当然，如果不考虑非特异性扩增的可能性，唯一性标识符中也可以不插入脱氧尿嘧啶。

[0021] 本申请的一种实现方式中，PCR扩增富集的扩增循环数大于或等于5。

[0022] 需要说明的是，第三引物和第二引物的PCR扩增富集，主要是为了使最开始第一引物延伸标记UMI的母链的扩增子能够被指数倍扩增富集，获得更多的来源于同一模板核酸的、且UMI都相同的扩增子链，以便于后续建库和测序。

[0023] 本申请的另一面公开了一种基于本申请的扩增子测序添加唯一性标识符的方法的测序文库构建方法，包括以下步骤：

[0024] 对PCR扩增富集的产物进行纯化，获得纯化产物；

[0025] 采用第四引物和第五引物，对纯化产物进行建库扩增，获得测序文库；第四引物为带有测序接头和Barcode的测序平台上游测序引物，第五引物为带有测序接头和Barcode的测序平台下游测序引物。

[0026] 需要说明的是，本申请的测序文库构建方法，实际上就是对第三引物和第二引物

的PCR扩增富集产物进行扩增建库。

[0027] 本申请的一种实现方式中,纯化为磁珠法纯化、过柱法纯化和凝胶法纯化中的至少一种。

[0028] 本申请的再一面公开了一种扩增子测序添加唯一性标识符的试剂盒,其包括第一引物、第二引物、第三引物和UDG/UNG酶;第一引物由5'端到3'端依序包括测序平台上游引物结合区、唯一性标识符和靶标特异性上游引物序列;并且,第一引物中,测序平台上游引物结合区和靶标特异性上游引物序列中的碱基T替换为脱氧尿嘧啶,测序平台上游引物结合区对应测序平台的上游测序引物的3'末端;第二引物由5'端到3'端依序包括测序平台下游引物结合区和靶标特异性下游引物序列,测序平台下游引物结合区对应测序平台的下游测序引物的3'末端;第三引物为第一引物的测序平台上游引物结合区自5'端起的全部或部分序列,并且,第三引物中的碱基T不被替换为脱氧尿嘧啶。

[0029] 需要说明的是,本申请扩增子测序添加唯一性标识符的试剂盒,实际上就是将本申请扩增子测序添加唯一性标识符的方法中使用的第一引物、第二引物和第三引物组装成试剂盒,以便于实现本申请的扩增子测序添加唯一性标识符的方法。

[0030] 本申请的一种实现方式中,本申请的扩增子测序添加唯一性标识符的试剂盒还包括PCR扩增试剂。

[0031] 可以理解,PCR扩增试剂可以根据需求组装到本申请的试剂盒中,也可以另行购买常规使用的PCR扩增试剂,例如PCR反应缓冲液、酶等。

[0032] 本申请的再一面公开了一种基于本申请的扩增子测序添加唯一性标识符的方法进行测序文库构建的试剂盒,其包括本申请的扩增子测序添加唯一性标识符的试剂盒,以及第四引物和第五引物;第四引物为带有测序接头和Barcode的测序平台上游测序引物,第五引物为带有测序接头和Barcode的测序平台下游测序引物。可以理解,本申请的文库构建试剂盒,实际上就是相应的增加了文库扩增引物。

[0033] 需要说明的是,本申请的关键之一在于引物结构的设计,至于具体的引物序列,可以根据具体针对的靶标序列和具体的测序平台而定。例如,采用常规的引物设计软件,针对具体的靶标序列设计第一引物的靶标特异性上游引物序列,针对预计采用的测序平台设计第一引物的测序平台上游引物结合区,由此组成第一引物。

[0034] 本申请的再一面公开了本申请的扩增子测序添加唯一性标识符的方法,或者本申请的基于本申请的扩增子测序添加唯一性标识符的方法的测序文库构建方法,在制备肺癌EGFR L858R和19De1突变率检测试剂或PIK3CA基因突变率检测试剂中的应用。

[0035] 需要说明的是,本申请的应用,主要是指根据本申请的扩增子测序添加唯一性标识符的方法,针对肺癌EGFR L858R和19De1突变率检测或PIK3CA基因突变率检测,设计相应的第一引物、第二引物、第三引物、第四引物和第五引物。

[0036] 本申请的再一面公开了一种用于肺癌EGFR L858R和19De1突变率检测的试剂盒,包括第一引物、第二引物、第三引物、第四引物和第五引物;

[0037] 第一引物由Seq ID No.1所示序列的EGFR L858R检测上游引物和Seq ID No.2所示序列的EGFR 19De1检测上游引物组成,

[0038] Seq ID No.1:

[0039] 5' -CACGACGCUCUCCGAUCUNNNUNNNCGUACUGGUGAAAACACCGCA -3' ,

[0040] Seq ID No.2:

[0041] 5'-CACGACGCUCUCCGAUCUNNNNUNNNNACUCUGGAUCCCAGAAGGUG-3' ;

[0042] 第二引物由Seq ID No.3所示序列的EGFR L858R检测下游引物和Seq ID No.4所示序列的EGFR 19De1检测下游引物组成,

[0043] Seq ID No.3:

[0044] 5'-AGACGTGTGCTCTTCCGATCTGAAACTCACATCGAGGATT-3' ,

[0045] Seq ID No.4:

[0046] 5'-AGACGTGTGCTCTTCCGATCTGAAACTCACATCGAGGATT-3' ;

[0047] 第三引物为Seq ID No.5所示序列,

[0048] Seq ID No.5:5'-CACGACGCTCTTCCGATCT-3' ;

[0049] 第四引物为Seq ID No.6所示序列,

[0050] Seq ID No.6:

[0051] 5'-AATGATACGGCGACCACCGAGATCTACACNNNNNNACACTCTTCCCTACACGACGCTCTTCCGATCT-3' ;

[0052] 第五引物为Seq ID No.7所示序列,

[0053] Seq ID No.7:

[0054] 5'-CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' 。

[0055] 其中,Seq ID No.6所示序列的第四引物和Seq ID No.7所示序列的第五引物中,“NNNNN”是指长度为6-10nt的Index,即Barcode。本申请用于肺癌EGFR L858R和19De1突变率检测的试剂盒中,Seq ID No.6所示序列的第四引物和Seq ID No.7所示序列的第五引物的“NNNNN”是不同的;例如,本申请的一种实现方式中,Seq ID No.6所示序列的第四引物的“NNNNN”具体为“ATCGGTTA”,Seq ID No.7所示序列的第五引物的“NNNNN”具体为“TCTAATGG”。

[0056] 本申请的再一面公开了一种用于PIK3CA基因突变率检测的试剂盒,包括第一引物、第二引物、第三引物、第四引物和第五引物;

[0057] 第一引物为Seq ID No.8所示序列,

[0058] Seq ID No.8:

[0059] 5'-CACGACGCUCUCCGAUCUNNNNUNNNNAGCAAUUUCUACACGAGAUCUCUCU-3' ;

[0060] 第二引物为Seq ID No.9所示序列,

[0061] Seq ID No.9:

[0062] 5'-AGACGTGTGCTCTTCCGATCTCTGGGCTACTTCATCTCTTGAAT-3' ;

[0063] 第三引物为Seq ID No.5所示序列,

[0064] Seq ID No.5:5'-CACGACGCTCTTCCGATCT-3' ;

[0065] 第四引物为Seq ID No.6所示序列,

[0066] Seq ID No.6:

[0067] 5'-AATGATACGGCGACCACCGAGATCTACACNNNNNNACACTCTTCCCTACACGACGCTCTTCCGATCT-3' ;

[0068] 第五引物为Seq ID No.7所示序列,

[0069] Seq ID No.7:

[0070] 5' -CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-3'。

[0071] 需要说明的是,本申请用于PIK3CA基因突变率检测的试剂盒中,其第三引物、第四引物和第五引物与本申请用于肺癌EGFR L858R和19De1突变率检测的试剂盒几乎相同,这是因为本申请采用相同的测序平台进行检测。当然,本申请用于PIK3CA基因突变率检测的试剂盒中,Seq ID No.6所示序列的第四引物和Seq ID No.7所示序列的第五引物的“NNNNN”也是不同的;例如,本申请的一种实现方式中,Seq ID No.6所示序列的第四引物的“NNNNN”具体为“CTATCACA”,Seq ID No.7所示序列的第五引物的“NNNNN”具体为“TTAGTAGC”。

[0072] 可以理解,本申请用于肺癌EGFR L858R和19De1突变率检测的试剂盒和用于PIK3CA基因突变率检测的试剂盒,只是本申请的扩增子测序添加唯一性标识符的方法、试剂盒及其文库构建试剂盒等的具体应用,在本申请相同的发明构思下,还可以设计出更多的针对其它基因突变率检测的试剂盒,在此不作具体限定。

[0073] 由于采用以上技术方案,本申请的有益效果在于:

[0074] 本申请扩增子测序添加唯一性标识符的方法,通过对第一引物、第二引物和第三引物进行特殊设计,依序向反应体系中加入三种引物,可以实现使每一个原始模板核酸只对应一个UMI标签,从而校正每个靶点扩增性偏差,校正PCR扩增错误,校正建库过程中人为引入的扩增错误;本申请的方法能够对每个原始模板核酸进行标记,从而实现靶标基因的拷贝数定量检测。

具体实施方式

[0075] 扩增子测序中,UMI是通过引物延伸添加到基因组DNA中的,但是,为了区别不同的原始模板DNA,对于同一种特异性引物而言,每条引物带的UMI都是不同的,这就使得基于同一个原始模板DNA获得的扩增子链,带有不同的UMI。

[0076] 为了解决上述问题,本申请研发了一种在扩增子测序中以基因组DNA为起始模板的加入唯一性标识符(UMI)的方法。本申请方法大致原理为:设计特异性UMI序列,可以延伸DNA模板链的互补链并加入UMI,后利用UDG/UNG酶消化体系中未被使用的UMI序列,实现每条模板分子的延伸链UMI唯一性。再利用此互补链扩增富集靶向区域,设计相应的测序引物,给富集产物加入Barcode/Index及测序接头,完成建库。

[0077] 本申请方法通用性强,可操作强,不仅可以用于常规扩增子测序,还可以兼容几乎所有常规的单重/多重PCR扩增,起始模板可以是DNA或cDNA,同时具有PCR实验的高特异性,高灵敏度等优点,文库具有UMI唯一性的特性,保证相同UMI的分子链都来自于同一条模板,可用于DNA分子检测相关的定量、定性实验,可以矫正因PCR错误或人为因素导致的假阳性问题,保证数据真实性。

[0078] 基于以上研究和认识,本申请提出了一种扩增子测序添加唯一性标识符的方法,包括以下步骤:

[0079] 配制反应体系,采用第一引物对模板核酸进行一次延伸,得到互补链;第一引物由5'端到3'端依序包括测序平台上游引物结合区、唯一性标识符和靶标特异性上游引物序

列;并且,第一引物中,测序平台上游引物结合区和靶标特异性上游引物序列中的碱基T替换为脱氧尿嘧啶,测序平台上游引物结合区对应测序平台的上游测序引物的3'末端;

[0080] 第一引物延伸结束后,向反应体系中加入第二引物,利用第二引物对第一引物延伸的互补链进行一次延伸,获得由测序平台上游引物结合区、唯一性标识符、靶标序列和测序平台下游引物结合区组成的产物;第二引物由5'端到3'端依序包括测序平台下游引物结合区和靶标特异性下游引物序列,测序平台下游引物结合区对应测序平台的下游测序引物的3'末端;

[0081] 第二引物延伸结束后,向反应体系中加入UDG/UNG酶消化脱氧尿嘧啶,从而消化第一引物以及第一引物的延伸链;

[0082] UDG/UNG酶消化完成后,向反应体系中加入第三引物,利用第三引物和第二引物对第二引物延伸的产物进行PCR扩增富集,获得模板核酸的所有扩增子都添加相同唯一性标识符的产物;第三引物为第一引物的测序平台上游引物结合区自5'端起的全部或部分序列,并且,第三引物中的碱基T不被替换为脱氧尿嘧啶。

[0083] 进一步的,测序文库构建方法,在以上步骤的基础上进行,包括对PCR扩增富集的产物进行纯化,获得纯化产物;然后,采用第四引物和第五引物,对纯化产物进行建库扩增,获得测序文库;第四引物为带有测序接头和Barcode的测序平台上游测序引物,第五引物为带有测序接头和Barcode的测序平台下游测序引物。

[0084] 本申请方法中,第一引物、第二引物、第三引物、第四引物和第五引物的设计思路如下:

[0085] 第一引物为UMI序列,包括三个部分:第一部分为15-25nt的固定序列,对应测序平台上游测序引物3'末端,第二部分为6-8位随机性的N碱基序列,即UMI,第三部分为靶标特异性上游引物序列。序列连接顺序为:5'-第一部分-第二部分-第三部分-3'。

[0086] 其中,在第一、三部分序列中,使用dU(脱氧尿嘧啶)碱基替换T(胸腺嘧啶)碱基。第一部分15-25nt固定序列可参考不同测序平台的上游接头完整3'末端序列,例如参考Illumina NovaSeq6000测序平台的P5端3'末尾19nt序列设计为Seq ID No.10所示序列,

[0087] Seq ID No.10:5'-CACGACGCUCUCCGAUCU-3'。

[0088] 进一步地,为了尽量避免随机的UMI可能在后续扩增中出现非特异性扩增,第二部分序列中可以选择使用一个或几个固定的脱氧尿嘧啶插入至6-8位随机性的N碱基序列中间,脱氧尿嘧啶左右两边的连续性N碱基数量小于5nt,当然,如果不考虑非特异性扩增的可能性,第二部分序列可以不插入脱氧尿嘧啶。例如,第二部分序列可以是5'-NNNNUNNN-3', 5'-NNNUNNNUNN-3'或5'-NNNNNNNN-3'等。

[0089] 进一步地,第三部分序列为靶标特异性上游引物序列,可根据NCBI等权威数据库查找靶标基因,根据碱基互补配对及引物设计原则,结合所需位点,自主设计上游引物。可根据多靶点设计多个特异性强的PCR引物,混合使用。

[0090] 第二引物,包括两个部分:第一部分为下游特异性引物序列,第二部分为15-25nt的固定序列,对应测序平台下游测序引物3'末端互补序列。序列连接顺序为:5'-第二部分-第一部分-3'。

[0091] 进一步地,第二引物的第一部分序列为靶标特异性下游引物序列,可根据NCBI等权威数据库查找靶标基因,根据碱基互补配对及引物设计原则,结合所需位点,自主设计下

游引物。可根据多靶点设计多个特异性强的PCR引物,混合使用。第二引物的第二部分15-25nt固定序列可参考不同测序平台的下游接头完整3'末端序列,例如参考Illumina NovaSeq6000测序平台的P7端3'末尾21nt序列的互补序列设计为Seq ID No.11所示序列,

[0092] Seq ID No.11:5' -AGACGTGTGCTCTTCCGATCT-3'。

[0093] 第三引物为与第一引物序列中15-25nt的固定序列相同的序列,需要注意得的是,此段序列T碱基不可被脱氧尿嘧啶替换。例如参考Illumina NovaSeq6000测序平台的P5端3'末尾19nt序列设计为Seq ID No.5所示序列,

[0094] Seq ID No.5:5' -CACGACGCTCTTCCGATCT-3'。

[0095] 第四引物的序列为:带有Barcode的完整上游测序接头序列,Index长度可以为6-10nt。例如参考Illumina NovaSeq6000测序平台的P5序列设计为Seq ID No.6所示序列,

[0096] Seq ID No.6:

[0097] 5' -AATGATACGGCGACCACCGAGATCTACACNNNNNNACACTCTTCCCTACACGACGCTCTTCCGATCT-3'。

[0098] Seq ID No.6所示序列的第四引物中“NNNNN”即Index序列。

[0099] 第五引物的序列为:带有Barcode的下游测序接头序列的互补序列,Index长度可以为6-10nt。例如参考Illumina NovaSeq6000测序平台的P7序列设计为Seq ID No.7所示序列,

[0100] Seq ID No.7:

[0101] 5' -CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'。

[0102] Seq ID No.7所示序列的第五引物中“NNNNN”即Index序列。

[0103] 基于本申请的第一引物至第五引物的快速加入UMI的方法及完整的文库制备技术,流程包括:1.使用第一引物延伸模板核酸,得到互补链,引入UMI标签。2.加入第二引物完全延伸第1步中的互补链。3.使用UDG/UNG酶消化脱氧尿嘧啶,消化第一引物。4.加入第三引物对第2步产物进行扩增富集;并纯化扩增产物,去除体系、多余引物及基因组污染。5.使用第四引物及第五引物进行建库扩增,加上Barcode及测序接头,完成建库。

[0104] 具体地,技术流程详细说明如下:

[0105] 第一步:引入UMI标签,模板互补链延伸:

[0106] 取含靶区域的模板核酸,例如DNA/cDNA,总量1-100ng,优选地基因组DNA为模板,优选地核酸量为100ng。

[0107] 配制延伸体系:可以使用市售PCR扩增试剂盒或者自研PCR扩增试剂盒,其主要组分可以包括但不限于:DNA聚合酶、Mg离子、dNTPs、缓冲体系。

[0108] 若实验设计为多重PCR反应,则优选选择市售或自研多重PCR扩增试剂盒。

[0109] 在配制好的延伸体系中,加入第一引物,其工作浓度为可以为50-500mM;优选地,工作浓度设置为200mM。

[0110] 在配制好的延伸体系中,加入准备好的模板DNA/cDNA,充分混合后,进行模板互补链延伸反应。反应程序参数应参照PCR扩增试剂盒说明书进行设置。应注意的是,需将反应的延伸时间调整为大于“目标区域长度/延伸速度”,即是目标区域能够被充分、完整的延伸;不设置PCR循环数或者循环数设置为1个,即只进行一次延伸,不重复进行变形、退火、延

伸。

[0111] 完成此反应后,得到的产物链即为已经加入UMI的模板互补链。

[0112] 第二步:模板互补链互补延伸

[0113] 取第一步反应产物,加入第二引物,其工作浓度为可以为50-500mM;优选地,工作浓度为200mM。

[0114] 混合均匀后,放入PCR程序,PCR程序与第一步保持完全一致。

[0115] 完成此反应后,得到的产物链即为已经加入UMI的文库片段,且序列与模板上的目的片段一致。

[0116] 第三步:使用UDG/UNG酶消化脱氧尿嘧啶

[0117] 准备热敏性UDG/UNG酶,可以为市售或自制酶。

[0118] 取出第二步反应产物,加入准备好的热敏性UDG/UNG酶,酶加入量根据酶活力及消化效率调整,一般来说,酶活力大于1U/ μ L时,加入1 μ L即可。充分混匀后,根据酶最适反应温度及条件,将体系中所有含脱氧尿嘧啶的序列消化。

[0119] 此步的目的是为了消化体系中富余的第一引物以及参与第一步反应的第一引物延伸序列。最终,得到的产物中只含有初始加入的DNA模板以及第二步产出的带有唯一性UMI标签的延伸产物链。

[0120] 第四步:特异性扩增富集

[0121] 第三引物加入至混合的第三步产物中,其工作浓度为可以为50-500mM;优选地,工作浓度为200mM。

[0122] 充分混合后,进行模板互补链延伸反应。反应程序参数应参照PCR扩增试剂盒说明书进行设置。PCR循环数可以根据项目需求及试剂盒性能个性化设置,优选5个循环以上。

[0123] 扩增完成后,将产物全部取出,进行核酸纯化。得到纯度高无杂质的纯化产物,进行下一步建库扩增。纯化方法可以是但不限于磁珠法、过柱法或者凝胶法。

[0124] 第五步:建库扩增

[0125] 将第四步得到的纯化产物,加入扩增体系,及第四引物、第五引物,充分混合。第四引物、第五引物工作浓度可以为200-2000mM,优选1500mM。

[0126] 反应程序参数应参照PCR扩增试剂盒说明书进行设置。PCR循环数可以根据项目需求及试剂盒性能个性化设置。

[0127] 反应完成后,产物即为带有完整接头信息的可上机文库。通过核酸纯化,得到高纯度的文库。质检定量后,文库即可用于上机测序。

[0128] 下面通过具体实施例对本申请作进一步详细说明。以下实施例仅对本申请进行进一步说明,不应理解为对本申请的限制。

[0129] 实施例一

[0130] 按照以上方法和思路本例设计了检测肺癌患者的EGFR L858R及19De1突变率的第一引物至第五引物,并进行相应的试验,具体如下:

[0131] 选取一例肺癌突变患者样本,其EGFR L858R及19De1突变率使用深圳市海普洛斯生物科技有限公司的产品“人EGFR/ALK基因突变联合检测试剂盒(可逆末端终止测序法)”(国械注准20213400832)进行测定。测定结果作为本例的对照试验。所用样本由深圳市海普洛斯生物科技有限公司提供和保存。

[0132] 按照以上思路设计第一引物至第五引物：

[0133] 第一引物，本例的第一引物为两种突变检测的混合引物，混合比例为1:1，即第一引物由Seq ID No.1所示序列的EGFR L858R检测上游引物和Seq ID No.2所示序列的EGFR 19De1检测上游引物组成，

[0134] Seq ID No.1：

[0135] 5' -CACGACGCUCUCCGAUCUNNNUNNNNCGUACUGGUGAAAAACACCGCA-3'，

[0136] Seq ID No.2：

[0137] 5' -CACGACGCUCUCCGAUCUNNNUNNNNACUCUGGAUCCCAGAAGGUG-3'。

[0138] 第二引物，也是两种突变检测的混合引物，混合比例为1:1，即第二引物由Seq ID No.3所示序列的EGFR L858R检测下游引物和Seq ID No.4所示序列的EGFR 19De1检测下游引物组成，

[0139] Seq ID No.3：

[0140] 5' -AGACGTGTGCTCTTCCGATCTGAAACTCACATCGAGGATT-3'，

[0141] Seq ID No.4：

[0142] 5' -AGACGTGTGCTCTTCCGATCTGAAACTCACATCGAGGATT-3'。

[0143] 第三引物为Seq ID No.5所示序列，

[0144] Seq ID No.5:5' -CACGACGCTCTTCCGATCT-3'。

[0145] 第四引物为Seq ID No.6所示序列，

[0146] Seq ID No.6：

[0147] 5' -AATGATACGGCGACCACCGAGATCTACACNNNNNNACACTCTTCCCTACACGACGCTCTCCGATCT-3'；

[0148] Seq ID No.6所示序列的第四引物中，“NNNNN”具体为“ATCGGTTA”。

[0149] 第五引物为Seq ID No.7所示序列，

[0150] Seq ID No.7：

[0151] 5' -CAAGCAGAAGACGGCATAACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'；

[0152] Seq ID No.7所示序列的第五引物中，“NNNNN”具体为“TCTAATGG”。

[0153] 设计合成以上引物后，用TE buffer将引物稀释，第一引物、第二引物、第三引物稀释至浓度5 μ M，第四引物、第五引物稀释至浓度30 μ M。

[0154] 取100ng上述该例样本基因组DNA，使用QIAGEN MultiplexPCR试剂盒进行扩增实验。

[0155] 配制反应体系，依次加入到一个新的0.2mL管中，反应体系为：第一引物4.5 μ L、基因组DNA 100ng、PCR Master mix 25 μ L、Q-solution 5 μ L，补充NF水至45 μ L。

[0156] 轻轻混匀混有样本及试剂的0.2mL样本管后，放入BIORAD T100PCR仪上进行PCR反应，反应程序：95 $^{\circ}$ C变性15min，然后94 $^{\circ}$ C 30s、60 $^{\circ}$ C 90s、72 $^{\circ}$ C 90s，最后72 $^{\circ}$ C延伸5min，4 $^{\circ}$ C待机。

[0157] 反应完成后，取出反应产物，加入5 μ L第二引物，混匀后放入BIORAD T100PCR仪上进行PCR反应，反应程序：95 $^{\circ}$ C变性15min，然后94 $^{\circ}$ C 30s、60 $^{\circ}$ C 90s、72 $^{\circ}$ C 90s，最后72 $^{\circ}$ C延伸5min，4 $^{\circ}$ C待机。

[0158] 反应完成后,取出反应产物,加入1 μ L热敏性UDG酶(Heat-labile UDG, Vazyme),混匀后放入以下程序中反应:25 $^{\circ}$ C消化10min,55 $^{\circ}$ C灭活5min,95 $^{\circ}$ C 5min,4 $^{\circ}$ C待机。

[0159] 反应完成后,取出反应产物,加入5 μ L第三引物,混匀后放入BIORAD T100PCR仪上进行PCR反应,反应程序:95 $^{\circ}$ C变性15min,然后进入20个循环:94 $^{\circ}$ C 30s、60 $^{\circ}$ C 90s、72 $^{\circ}$ C 90s,循环结束后72 $^{\circ}$ C延伸5min,4 $^{\circ}$ C待机。

[0160] 反应完成后,取出反应产物,使用磁珠纯化,详细步骤如下:

[0161] 1.使用1.2 \times AMpure XP beads磁珠进行多重PCR产物纯化:取新1.5mL样本管,将多重PCR产物50 μ L与60 μ L混合均匀的AMpure XP beads磁珠加入新1.5mL样本管中,涡旋混匀,置于室温10min,使DNA与磁珠充分结合。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心移除上清液。

[0162] 2.再加入500 μ L的80%乙醇,180度旋转样本管使磁珠穿过溶液吸至另一侧管壁,旋转2-3次,静置15s后弃上清液。

[0163] 3.重复步骤2一次;

[0164] 4.自然静置1.5mL样本管,待酒精全部挥发后,向1.5mL样本管中加入20 μ L无核酸酶水,充分混匀。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心吸出上清液放入新0.2 μ L样本管中,得到纯化产物。

[0165] 测序文库构建,将第四引物、第五引物、扩增试剂与纯化产物混合,进行建库扩增。使用KAPAHiFi Hotstart Ready Mix试剂进行扩增,混合比例如下:2 \times KAPAHiFi Hotstart Ready Mix 25 μ L、第四引物2.5 μ L、第五引物2.5 μ L、纯化产物20 μ L,补充NF水至50 μ L。

[0166] 混合均匀后,放入以下程序中反应:98 $^{\circ}$ C变性45s,然后进入5个循环:98 $^{\circ}$ C 15s、60 $^{\circ}$ C 30s、72 $^{\circ}$ C 30s,循环结束后72 $^{\circ}$ C延伸1min,4 $^{\circ}$ C待机。

[0167] 程序结束后,得到建库扩增PCR产物50 μ L,使用1 \times AMpure XP磁珠进行多重PCR产物纯化:

[0168] 取新1.5mL样本管,将多重PCR产物50 μ L与50 μ L混合均匀的AMpure XP磁珠加入新1.5mL样本管中,涡旋混匀,置于室温10min,使DNA与磁珠充分结合。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心移除上清液。

[0169] 加入500 μ L的80%乙醇,180度旋转样本管使磁珠穿过溶液吸至另一侧管壁,旋转2-3次,静置15s后弃上清液;重复此步骤一次。

[0170] 自然静置1.5mL样本管,待酒精全部挥发后,向1.5mL样本管中加入20 μ L Nuclease-Free Water,充分混匀。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心吸出上清液,标记保存,即完成扩增子文库建库。

[0171] 高通量测序:使用Illumina NovaSeq 6000高通量测序系统进行上机测序,上机模式为PE151+8+8+151。

[0172] 结果:

[0173] 对测序获得的下机数据进行常规的质控过滤和测序深度阈值过滤后,分析序列UMI。相同UMI的序列,如果Reads序列一致,可认为来源于同一母链模板。比较来自同一模板的突变型Reads与野生型Reads数量,可以直观测定样本突变率。对比本例方法检测获得的L858R及19Del的突变率,以及“人EGFR/ALK基因突变联合检测试剂盒(可逆末端终止测序法)”检测获得的突变率,结果如表1所示。

[0174] 表1 L858R和19De1的突变率检测结果

试验方法	突变类型	突变率
本例方法	L858R	5.1%
对照方法	L858R	4.9%
本例方法	19De1	12.2%
对照方法	19De1	12.5%

[0176] 表1的结果显示,本例方法与试剂盒的对照方法在突变率结果上具有良好的一致性,可以用于准确定量实验。

[0177] 此外,本例方法在测序错误中也有良好的矫正作用,可以矫正扩增偏差带来的不准确定量、矫正因扩增错误或实验引入的错误导致的阳性率虚高/假阳性。例如,采用本例的方法,如果不进行UMI矫正,L858R的突变率为12.2%,远高于试剂盒的检测结果,说明存在大量的假阳性,而校正后L858R的突变率为5.1%,与试剂盒检测结果一致。同样的,如果不进行UMI矫正,19De1的突变率为20.3%,远高于试剂盒的检测结果,说明存在大量的假阳性,而校正后19De1的突变率为12.2%,与试剂盒检测结果具有较好的一致性。

[0178] 实施例二

[0179] 按照以上方法和思路本例设计了检测一例人PIK3CA基因突变率的第一引物至第五引物,并进行相应的试验,具体如下:

[0180] 选用一个经过液态探针杂交捕获技术测序的人类样本,结果得知,该样本的PIK3CA基因,发生了碱基突变c.1624G>A、c.1633G>A,突变率分别为1.7%和2.2%,其中c.1635G>T未发生突变,突变率0%。此NGS结果作为对照数据。取该样本的原始基因组DNA,用本例方法进行扩增子建库实验,对比本例方法与对照数据突变率的一致性。所用样本由深圳市海普洛斯生物科技有限公司提供和保存。

[0181] 按照以上思路设计第一引物至第五引物:

[0182] 第一引物为G1624A、G1633A两种突变检测通用上游引物,序列为Seq ID No.8所示序列,

[0183] Seq ID No.8:

[0184] 5' -CACGACGCUCUCCGAUCUNNNNUNNNNAGCAAUUUCUACACGAGAUCUCUCU-3'。

[0185] 第二引物为两种突变的通用下游引物,序列为Seq ID No.9所示序列,

[0186] Seq ID No.9:

[0187] 5' -AGACGTGTGCTCTTCCGATCTCTGGGCTACTTCATCTCTTGAAT-3'。

[0188] 第三引物为Seq ID No.5所示序列,

[0189] Seq ID No.5:5' -CACGACGCTCTTCCGATCT-3'。

[0190] 第四引物为Seq ID No.6所示序列,

[0191] Seq ID No.6:

[0192] 5' -AATGATACGGCGACCACCGAGATCTACACNNNNNNACACTCTTCCCTACACGACGCTCTCCGATCT-3' ;

[0193] Seq ID No.6所示序列的第四引物中,“NNNNN”具体为“CTATCACA”。

[0194] 第五引物为Seq ID No.7所示序列,

[0195] Seq ID No.7:

[0196] 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-3' ;

[0197] Seq ID No.7所示序列的第五引物中,“NNNNN”具体为“TTAGTAGC”。

[0198] 设计合成以上引物后,用TE buffer将引物稀释,第一引物、第二引物、第三引物稀释至浓度5 μ M,第四引物、第五引物稀释至浓度30 μ M。

[0199] 取100ng上述该例样本基因组DNA,使用QIAgen MultiplexPCR试剂盒进行扩增实验。

[0200] 配制反应体系,依次加入到一个新的0.2mL管中,反应体系为:第一引物4.5 μ L、基因组DNA100ng、PCR Master mix 25 μ L、Q-solution 5 μ L,补充NF水至45 μ L。

[0201] 轻轻混匀混有样本及试剂的0.2mL样本管后,放入BIORAD T100PCR仪上进行PCR反应,反应程序:95 $^{\circ}$ C变性15min,然后94 $^{\circ}$ C 30s、55 $^{\circ}$ C 90s、72 $^{\circ}$ C 90s,最后72 $^{\circ}$ C延伸5min,4 $^{\circ}$ C待机。

[0202] 反应完成后,取出反应产物,加入5 μ L第二引物,混匀后放入BIORAD T100PCR仪上进行PCR反应,反应程序:95 $^{\circ}$ C变性15min,然后94 $^{\circ}$ C 30s、55 $^{\circ}$ C 90s、72 $^{\circ}$ C 90s,最后72 $^{\circ}$ C延伸5min,4 $^{\circ}$ C待机。

[0203] 反应完成后,取出反应产物,加入1 μ L热敏感性UDG酶(Heat-labile UDG,Vazyme),混匀后放入以下程序中反应:25 $^{\circ}$ C消化10min,55 $^{\circ}$ C灭活5min,95 $^{\circ}$ C 5min,4 $^{\circ}$ C待机。

[0204] 反应完成后,取出反应产物,加入5 μ L第三引物,混匀后放入BIORAD T100PCR仪上进行PCR反应,反应程序:95 $^{\circ}$ C变性15min,然后进入20个循环:94 $^{\circ}$ C 30s、55 $^{\circ}$ C 90s、72 $^{\circ}$ C 90s,循环结束后72 $^{\circ}$ C延伸5min,4 $^{\circ}$ C待机。

[0205] 反应完成后,取出反应产物,使用磁珠纯化,详细步骤如下:

[0206] 1.使用1.2 \times AMPure XP beads磁珠进行多重PCR产物纯化:取新1.5mL样本管,将多重PCR产物50 μ L与60 μ L混合均匀的AMPure XP beads磁珠加入新1.5mL样本管中,涡旋混匀,置于室温10min,使DNA与磁珠充分结合。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心移除上清液。

[0207] 2.再加入500 μ L的80%乙醇,180度旋转样本管使磁珠穿过溶液吸至另一侧管壁,旋转2-3次,静置15s后弃上清液。

[0208] 3.重复步骤2一次。

[0209] 4.自然静置1.5mL样本管,待酒精全部挥发后,向1.5mL样本管中加入20 μ L无核酸酶水,充分混匀。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心吸出上清液放入新0.2 μ L样本管中,得到纯化产物。

[0210] 测序文库构建,将第四引物、第五引物、扩增试剂与纯化产物混合,进行建库扩增。使用KAPA HiFi Hotstart Ready Mix试剂进行扩增,混合比例如下:2 \times KAPA HiFi Hotstart Ready Mix 25 μ L、第四引物2.5 μ L、第五引物2.5 μ L、纯化产物20 μ L,补充NF水至50 μ L。

[0211] 混合均匀后,放入以下程序中反应:98 $^{\circ}$ C变性45s,然后进入5个循环:98 $^{\circ}$ C 15s、60 $^{\circ}$ C 30s、72 $^{\circ}$ C 30s,循环结束后72 $^{\circ}$ C延伸1min,4 $^{\circ}$ C待机。

[0212] 程序结束后,得到建库扩增PCR产物50 μ L,使用1 \times AMPure XP磁珠进行多重PCR产物纯化:

[0213] 取新1.5mL样本管,将多重PCR产物50 μ L与50 μ L混合均匀的Ampure XP磁珠加入新1.5mL样本管中,涡旋混匀,置于室温10min,使DNA与磁珠充分结合。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心移除上清液。

[0214] 加入500 μ L的80%乙醇,180度旋转样本管使磁珠穿过溶液吸至另一侧管壁,旋转2-3次,静置15s后弃上清液;重复此步骤一次。

[0215] 自然静置1.5mL样本管,待酒精全部挥发后,向1.5mL样本管中加入20 μ L Nuclease-FreeWater,充分混匀。将1.5mL样本管置于磁力架上,进行磁珠吸附,直至溶液澄清,小心吸出上清液,标记保存,即完成扩增子文库建库。

[0216] 高通量测序:使用Illumina NovaSeq 6000高通量测序系统进行上机测序,上机模式为PE151+8+8+151。

[0217] 结果:

[0218] 对测序获得的下机数据进行常规的质控过滤和测序深度阈值过滤后,分析序列UMI。相同UMI的序列,如果Reads序列一致,可认为来源于同一母链模板。比较来自同一模板的突变型Reads与野生型Reads数量,可以直观测定样本突变率。对比本例方法检测获得的突变率和NGS结果,如表2所示。

[0219] 表2 PIK3CA基因突变率检测结果

[0220]

方法	碱基变化	突变率
本例方法	G1624A	1.7%
NGS结果	G1624A	1.7%
本例方法	G1633A	2.3%
NGS结果	G1633A	2.2%
本例方法	G1635T	0%
NGS结果	G1635T	0%

[0221] 表2的结果显示,本例方法与NGS结果具有良好的一致性,可以用于准确的定量实验。

[0222] 以上内容是结合具体的实施方式对本申请所作的进一步详细说明,不能认定本申请的具体实施只局限于这些说明。对于本申请所属技术领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干简单推演或替换。

SEQUENCE LISTING

<110> 深圳海普洛斯医学检验实验室

<120> 一种扩增子测序添加唯一性标识符的方法及应用

<130> 21I33106

<160> 11

<170> PatentIn version 3.3

<210> 1

<211> 49

<212> RNA

<213> 人工序列

<220>

<221> misc_feature

<222> (20) .. (23)

<223> n is a, c, g, or u

<220>

<221> misc_feature

<222> (25) .. (28)

<223> n is a, c, g, or u

<400> 1

cacgacgcuc uuccgaucun nnnunnnncg uacuggugaa aacaccgca 49

<210> 2

<211> 48

<212> RNA

<213> 人工序列

<220>

<221> misc_feature

<222> (20) .. (23)

<223> n is a, c, g, or u

<220>

<221> misc_feature

<222> (25) .. (28)

<223> n is a, c, g, or u

<400> 2

cacgacgcuc uuccgaucun nnnunnnnac ucuggauc cc agaaggug 48

<210> 3

<211> 40

<212> DNA

<213> 人工序列

<400> 3
agacgtgtgc tcttccgata tgaaactcac atcgaggatt 40
<210> 4
<211> 40
<212> DNA
<213> 人工序列
<400> 4
agacgtgtgc tcttccgata tgaaactcac atcgaggatt 40
<210> 5
<211> 19
<212> DNA
<213> 人工序列
<400> 5
cacgacgctc ttccgatct 19
<210> 6
<211> 68
<212> DNA
<213> 人工序列
<220>
<221> misc_feature
<222> (30)..(35)
<223> n is a, c, g, or t
<400> 6
aatgatacgg cgaccaccga gatctacacn nnnnacact ctttccctac acgacgctct 60
tccgatct 68
<210> 7
<211> 64
<212> DNA
<213> 人工序列
<220>
<221> misc_feature
<222> (25)..(30)
<223> n is a, c, g, or t
<400> 7
caagcagaag acggcatacg agatnnnnnn gtgactggag ttcagacgtg tgctcttccg 60
atct 64
<210> 8
<211> 54
<212> RNA

<213> 人工序列
<220>
<221> misc_feature
<222> (20)..(23)
<223> n is a, c, g, or u
<220>
<221> misc_feature
<222> (25)..(28)
<223> n is a, c, g, or u
<400> 8
cacgacgcuc uuccgaucun nnnunnnnag caauuucuaac acgagauccu cucu 54
<210> 9
<211> 44
<212> DNA
<213> 人工序列
<400> 9
agacgtgtgc tcttccgatc tctgggctac ttcattcttt gaat 44
<210> 10
<211> 19
<212> RNA
<213> 人工序列
<400> 10
cacgacgcuc uuccgaucu 19
<210> 11
<211> 21
<212> DNA
<213> 人工序列
<400> 11
agacgtgtgc tcttccgatc t 21