



(12) 发明专利申请

(10) 申请公布号 CN 117625764 A

(43) 申请公布日 2024. 03. 01

(21) 申请号 202311034878.2

C12N 15/11 (2006.01)

(22) 申请日 2023.08.16

(30) 优先权数据

17/899,851 2022.08.31 US

(71) 申请人 基诺米尔健康公司

地址 芬兰图尔库

(72) 发明人 朱哈-佩卡·珀斯黑默

塔图·赫沃宁

安托尼·科尔基亚科斯基

马努·塔米宁

(74) 专利代理机构 北京新知远方知识产权代理

事务所(普通合伙) 11397

专利代理师 张艳 马军芳

(51) Int. Cl.

C12Q 1/6869 (2018.01)

权利要求书3页 说明书16页

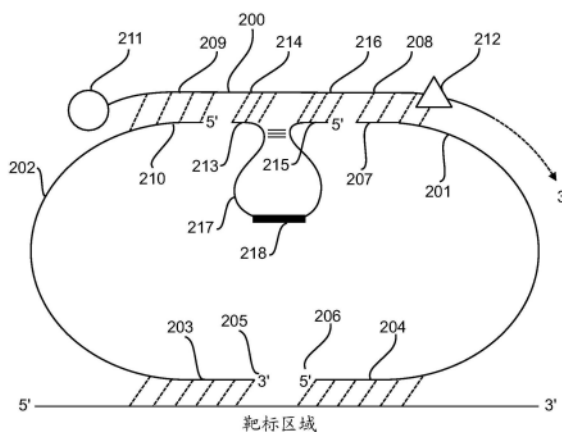
序列表(电子公布) 附图7页

(54) 发明名称

准确地平行检测和定量核酸的方法

(57) 摘要

本发明公开的内容涉及下一代DNA测序方法和用途,用于对一种或多种核酸靶标(例如在大量的未纯化的样品材料中)进行准确且大规模平行定量。更具体地,本发明涉及包括用于检测和定量复杂样品中的基因靶标的探针的方法和试剂盒。本发明包括两个靶标特异性核酸探针/每种基因靶标、条形码环状寡核苷酸和桥寡核苷酸或桥寡核苷酸复合体。



1. 一种高通量检测多个样品中一种或多种靶标核苷酸序列的方法,所述方法包括以下步骤:

(i) 为每个样品中的每种靶标核苷酸序列提供:第一探针、第二探针、条形码环状寡核苷酸,和桥寡核苷酸或能够与所述条形码环状寡核苷酸退火以形成桥寡核苷酸复合体的多个桥寡核苷酸,

其中,所述第一探针包括位于所述第一探针的5'端的第一桥寡核苷酸特异性序列和位于所述第一探针的3'端的第一靶标特异性部分;

其中,所述第二探针包括位于所述第二探针的5'端的第二靶标特异性部分和位于所述第二探针的3'端的第二桥寡核苷酸特异性序列;

其中,所述条形码环状寡核苷酸自分子的5'端起包括第三桥寡核苷酸特异性序列、条形码环状序列和第四桥寡核苷酸特异性序列;

其中,所述桥寡核苷酸或多个桥寡核苷酸包括与所述第一探针中的第一桥寡核苷酸特异性序列和所述第二探针中的第二桥寡核苷酸特异性序列互补的序列,和与所述条形码环状寡核苷酸中的第三桥寡核苷酸特异性序列和第四桥寡核苷酸特异性序列互补的序列;

并且其中,可选地,所述第一探针、所述第二探针、所述条形码环状寡核苷酸、所述桥寡核苷酸或者所述多个桥寡核苷酸的寡核苷酸中的至少一者包括核酸内切酶的识别序列;

(ii) 对于所述一种或多种靶标核苷酸序列中的每一种,优选对于在单独管中的每个样品,使所述第一探针和第二探针与所述条形码环状寡核苷酸和所述桥寡核苷酸或多个桥寡核苷酸接触并允许自退火为连接复合体;

(iii) 使存在于待测试所述靶标核苷酸序列的每个样品中的核酸与所述连接复合体接触;

(iv) 允许所述第一探针的第一靶标特异性部分和所述第二探针的第二靶标特异性部分与靶标序列上基本相邻的区段杂交,从而形成杂交复合体;

(v) 可选地,将来自所述多个样品的杂交复合体汇集起来;

(vi) 将所述杂交复合体中的探针连接起来,以提供经连接的连接复合体;

(vii) 使用滚环扩增法,用链置换聚合酶从一个或多个经连接的连接复合体中扩增核酸,从而获得单链的多联体序列;

(viii) 可选地,在存在步骤(i)中指定的识别序列的条件下,进行通过以下方式获得核酸片段的步骤:

(a) 切割在步骤(vii)中获得的单链的多联体序列,或

(b) 使在步骤(vii)中获得的单链的多联体序列与包含核酸内切酶的识别序列的特异性寡核苷酸退火,其中所述寡核苷酸与步骤(i)中指定的所述识别序列退火,从而获得所述核酸内切酶的识别位点,以及用所述核酸内切酶切割经退火的复合体;

(ix) 对步骤(vii)中获得的所述多联体序列或步骤(viii)中获得的核酸片段进行高通量测序技术以确定所述条形码序列;以及

(x) 通过确定所述第一靶标特异性部分和/或第二靶标特异性部分的至少一部分,和/或与所述条形码环状寡核苷酸中的条形码对应的条形码序列的至少一部分,来鉴定所述多个样品中靶标核苷酸序列的存在和/或数量,

其中,步骤(v)和步骤(vi)可按任何顺序进行。

2. 根据权利要求1所述的方法,其中,所述条形码环状寡核苷酸含有核酸内切酶,诸如缺口核酸内切酶的一个或多个识别序列。

3. 根据权利要求2所述的方法,其中,所述条形码环状寡核苷酸含有能够相互退火的两个识别序列,从而获得双链核酸内切酶识别位点。

4. 根据权利要求3所述的方法,其中,所述方法不包括权利要求1中指定的步骤(viii),但在步骤(vii)与步骤(ix)之间包括以下步骤:允许所述条形码环状寡核苷酸中的两个识别序列退火,并用对所述识别位点具有特异性的核酸内切酶,诸如缺口核酸内切酶来裂解所得双链核酸内切酶识别位点。

5. 根据前述权利要求中任一项所述的方法,其中,所述桥寡核苷酸,或者所述多个桥寡核苷酸的一个或多个寡核苷酸包括,在与所述第一探针、第二探针或条形码环状寡核苷酸不互补的区域的多个通用碱基类似物,以允许引入适合用作靶标列举的分子条形码的随机序列,并且其中,作为步骤(vi)的一部分,使用聚合酶和核苷酸进行间隙填充步骤以生成这种随机序列。

6. 根据权利要求5的方法,其中,所述多个通用碱基类似物是多个5-硝基吡啶。

7. 根据前述权利要求中任一项所述的方法,其中,所述第一探针、第二探针或桥寡核苷酸或者多个桥寡核苷酸的寡核苷酸进一步包括序列条形码。

8. 根据前述权利要求中任一项所述的方法,其中

在步骤(v)之后,但在步骤(vii)之前,进行步骤(a)和步骤(b),其中步骤(a)包括允许经连接的连接复合体与所述靶标核苷酸序列解离,并且步骤(b)包括加入靶标特异性探针,所述靶标特异性探针包括与所述靶标核苷酸序列对应的序列,其中所述靶标特异性探针能够与所述经连接的连接复合体退火,并允许所述靶标特异性探针与所述经连接的连接复合体退火,从而形成扩增模板,并且

其中,在步骤(vii)中,用链置换聚合酶通过滚环扩增对所述扩增模板进行扩增。

9. 根据前述权利要求中任一项所述的方法,其中,所述第一探针、所述第二探针、所述条形码环状寡核苷酸、所述桥寡核苷酸或者所述多个桥寡核苷酸的寡核苷酸中的至少一者包括第一捕获部,并且其中,在步骤(iv)和步骤(v)之间进行中间步骤(iv)(a),所述中间步骤(iv)(a)包括使杂交复合体与含第二捕获部的固体支持物接触,允许所述第一捕获部与所述第二捕获部相互作用,从而使得所述杂交复合体与所述固体支持物相连,以及使与固体支持物相连的杂交复合体与样品中未与所述固体支持物相连的组分分离。

10. 根据前述权利要求中任一项所述的方法,其中,所述多个样品包括血液样品、唾液样品、尿样品或粪便样品。

11. 根据前述权利要求中任一项所述的方法,其中,所述桥寡核苷酸或多个桥寡核苷酸的寡核苷酸包括:

(i) 一至五个3'突出的碱基,和/或

(ii) 3'磷酸,和/或

(iii) 从3'端起的一个或多个硫代磷酸酯修饰。

12. 根据前述权利要求中任一项所述的方法,其中,对所述第一探针的3'端或所述第二探针的5'端,或者所述第一探针的3'端和所述第二探针的5'端进行修饰以允许所述第一探针与所述第二探针的化学连接。

13. 根据前述权利要求中任一项所述的方法,其中,所述第一探针或所述第二探针的桥接部分,或者所述第一探针的桥接部分和所述第二探针的桥接部分,或者所述条形码环状寡核苷酸、所述桥寡核苷酸或者所述多个桥寡核苷酸中的寡核苷酸包括经化学修饰的碱基以改进与所述桥寡核苷酸或桥寡核苷酸复合体的结合。

14. 根据前述权利要求中任一项所述的方法,其中,所述第一靶标特异性部分、第二靶标特异性部分、第一桥寡核苷酸特异性序列和/或第二桥寡核苷酸特异性序列,彼此独立地包含一个或多个经化学修饰的核苷酸。

15. 根据前述权利要求中任一项所述的方法,其中,所述桥寡核苷酸或者所述多个桥寡核苷酸的寡核苷酸,包括一个或多个经化学修饰的核苷酸。

16. 根据前述权利要求中任一项所述的方法,其中,使用phi29聚合酶或Bst聚合酶进行步骤(vii)。

17. 根据前述权利要求中任一项所述的方法,其中,紧邻步骤(ix)且在步骤(ix)之前,使用与所述第一探针和所述第二探针的通用部分结合的引物进行PCR扩增,其中所述引物可选地包括用于步骤(ix)中的后续测序的接头。

18. 根据前述权利要求中任一项所述的方法,其中,使用纳米孔测序进行步骤(ix)中的测序,其中可选地使用转座复合体对步骤(vii)中获得的多联体序列进行片段化。

19. 根据前述权利要求中任一项所述的方法,其中,通过对每种靶标和每个样品的分子条形码的数量进行计数来允许对基因靶标进行列举。

20. 根据前述权利要求中任一项所述的方法,其中,对于两个或更多个的样品或者对于两个或更多个基因座/等位组合,使用条形码序列来对样品的一种或多种序列和/或多态性,诸如SNP和/或插入缺失,进行基因分型。

21. 包括多个容器的成套的试剂盒,其中,至少一个容器容纳一组或多组的第一探针和第二探针,至少一个容器容纳条形码环状寡核苷酸,并且至少一个容器容纳一种或多种桥寡核苷酸或者能够与所述条形码环状寡核苷酸形成桥寡核苷酸复合体的多个桥寡核苷酸;

其中,所述第一探针包括位于所述第一探针的5'端的第一桥寡核苷酸特异性序列和位于所述第一探针的3'端的第一靶标特异性部分;

其中,所述第二探针包括位于所述第二探针的5'端的第二靶标特异性部分和位于所述第二探针的3'端的第二桥寡核苷酸特异性序列;

其中,所述条形码环状寡核苷酸自分子的5'端起包括第三桥寡核苷酸特异性序列、条形码环状序列和第四桥寡核苷酸特异性序列,

其中,所述桥寡核苷酸或多个桥寡核苷酸包含与所述第一探针中的第一桥寡核苷酸特异性序列和所述第二探针中的第二桥寡核苷酸特异性序列互补的序列,和与所述条形码环状寡核苷酸中的第三桥寡核苷酸特异性序列和第四桥寡核苷酸特异性序列互补的序列;

并且其中,可选地,所述第一探针、所述第二探针、所述条形码环状寡核苷酸、所述桥寡核苷酸和所述多个桥寡核苷酸的寡核苷酸中的至少一者包括核酸内切酶的识别序列;

并且其中,可选地,所述成套的试剂盒进一步包括能够与所述识别序列退火的寡核苷酸,从而获得所述核酸内切酶的识别位点。

准确地平行检测和定量核酸的方法

技术领域

[0001] 本发明公开的内容涉及改进的下一代DNA测序方法,用于对一种或多种核酸靶标进行准确且大规模平行定量。更特别的是,本公开涉及包括用于检测和定量复杂DNA池中基因靶标的探针的方法和试剂盒,主要用于基因靶标和变体检测。本发明使用一种或多种靶标特异性核酸探针/每种基因靶标、条形码环状寡核苷酸和一种或多种桥寡核苷酸。

背景技术

[0002] 随着研究基因变异技术的进步,在植物和动物中检测基因变异并不麻烦。然而,尽管测序成本下降,但检测和准确定量基因变异(诸如突变),特别是在信号较弱的样品中,目前仍然是麻烦、费力和昂贵的。各种问题可更准确地表述如下:诸如特异性以在共有背景下检测基因信号、灵敏度以检测弱基因信号、准确性以准确定量经检测的信号、每次测定的经靶向基因靶标的通量数量、每次测定的成本、当平行测定多个样品时确定测定成本规模的规模化以及确定从采样到结果的时间长度的周转(turn-over)。

[0003] 目前,液体活组织检查和概念上类似的测定(诸如抗生素抗性基因检测)的典型定量方法包括定量PCR(qPCR)、阵列qPCR、数字PCR、多重连接依赖式探针扩增(MLPA)或来自下一代DNA测序数据的定量。虽然定量方法是稳健和成熟的方法,但每种方法都与下面更详细讨论的具体问题有关系:

[0004] 定量PCR:定量PCR(qPCR)是一种包括在PCR过程中(即实时)扩增经靶向的DNA分子的技术。实时PCR可以定量使用(定量实时PCR),以及也可以半定量使用,即高于/低于一定量的DNA分子(半定量实时PCR)。定量PCR(qPCR)是基因靶标定量的金标准。目前,qPCR反应的实验室成本约为2\$。然而,考虑到设置反应所需的大量实际操作时间(人工成本)、标准曲线的需要以及每个经定量靶标的重复,实际成本要高得多。由于每种基因靶标都需要单独的定量实验,因此实际操作时间会随着样品数量的增加而急剧增加。

[0005] 阵列PCR:PCR阵列是分析相关通路-或聚焦疾病的一组基因表达的最可靠工具。每个96孔板、384孔板或100孔盘PCR阵列都包括SYBR Green优化的引物测定,用于对一组聚焦基因进行彻底研究。qPCR技术的一个新的迭代是阵列qPCR,其将单个qPCR反应小型化。阵列PCR降低了单个qPCR反应的成本,并提高了该方法对多个靶标和样品的可扩展性。然而,该方法目前局限于以每芯片数千美元的成本加上读出基础设施的巨大资本成本来剖析来自12个样品的384种靶标(或者相反地,来自384个样品的12种靶标)。因此,使用上述设置剖析数千个样品仍然非常昂贵。

[0006] 数字PCR:数字聚合酶链式反应(Digital PCR, DigitalPCR, dPCR, dePCR)是一种通过微滴微流体和荧光检测来提供靶标绝对定量的方法。这种方法相对成本有效(每个样品一个靶标成本约为3\$),但每个样品中每种靶标的准备、设置和运行单独实验的实际时间使得难以扩展至数千个样品。

[0007] 多重连接依赖式探针扩增(MLPA)提供了一种简化单个样品中多个基因靶标检测的方法。然而,MLPA只提供靶标的相对定量,并需要对每个样品进行单独的检测实验。最近,

MLPA的一个变型引入了DNA条形码的概念。与传统的MLPA工作流程相比,该概念允许更好的定量分辨率和样品复用。

[0008] 基于下一代测序的方法:下一代测序(NGS),也称为高通量测序,使得基于序列的基因表达分析成为模拟技术的“数字”替代物。随着DNA测序成本的不断降低,从下一代DNA测序数据中进行靶标计数正变得越来越有吸引力,并且目前正用于例如无创产前测试。然而,目前的方法存在测序文库制备成本高和测序工作浪费在非相关基因靶标测序上的问题。例如,在癌症相关液体活组织检查中,非靶向方法导致肿瘤学的非相关基因座测序工作的浪费。在胎儿诊断学中,基因座的非靶向采样极大地限制了解释数据的统计选择。Guardant Health Inc.提供了更有靶向性的测序方法,其中一系列RNA捕获探针丰富了下一代DNA测序的靶标。

[0009] Akhras et al. (2007) PLoS ONE 2(2):e223公开了一种涉及条形码化靶标特异性探针、靶标环化和测序的多重病原体检测测定。还公开了使用桥寡核苷酸来连接靶标特异性探针。

[0010] W02018109206描述了使用挂锁探针和滚环扩增来检测样品中分析物的方法。没有描述使用桥寡核苷酸。

[0011] W02019038372描述了一种下一代测序方法,其中通过体外转录从含有T7聚合酶启动子的连接复合体中选择性地扩增感兴趣的靶序列,然后进行cDNA合成和测序。虽然这种方法可以对样品中的许多靶标序列进行准确和平行的检测和定量,但对于更复杂、大体积、稀释和/或不纯的样品仍具有挑战性。

[0012] 因此,根据前述讨论,需要通过对核酸靶标的准确和大规模平行定量来克服前述缺点,诸如但不限于特异性、灵敏度、准确性、通量、成本、规模化(scaling)和周转。

发明内容

[0013] 本发明提供了一种使用下一代测序技术在大体积样品(高达数十毫升)和/或稀释和/或未纯化的样品材料中进行高灵敏度、可规模化且准确的靶定量的方法。而且,避免了诸如W02019038372中描述的RNA扩增步骤,使该方法更加简单。此外,本发明的方法包括使用单独的条形码环状寡核苷酸进行样品鉴定。条形码环状寡核苷酸的使用允许同时进行靶查询和样品索引,允许有效的样品汇集,从而使定量测定更具成本效益和灵活性,并为引入唯一分子标识符(UMI)序列提供一种替代方法。由此产生的方案省略了对单独样品索引的需求,因此允许更短的读取长度,并为短读取和纳米孔测序应用提供好处。

[0014] 在第一个主要方面,本发明涉及一种高通量检测多个样品中一种或多种靶标核苷酸序列的方法,所述方法包括以下步骤:

[0015] (i) 为每个样品中的每种靶标核苷酸序列提供:第一探针、第二探针、条形码环状寡核苷酸,和桥寡核苷酸或能够与所述条形码环状寡核苷酸退火以形成桥寡核苷酸复合体的多个桥寡核苷酸,

[0016] 其中,所述第一探针包括位于所述第一探针的5'端的第一桥寡核苷酸特异性序列和位于所述第一探针的3'端的第一靶标特异性部分;

[0017] 其中,所述第二探针包括位于所述第二探针的5'端的第二靶标特异性部分和位于所述第二探针的3'端的第二桥寡核苷酸特异性序列;

[0018] 其中,所述条形码环状寡核苷酸自该分子的5'端起包括第三桥寡核苷酸特异性序列,条形码环状序列和第四桥寡核苷酸特异性序列;

[0019] 其中,所述桥寡核苷酸或多个桥寡核苷酸包括与所述第一探针中的第一桥寡核苷酸特异性序列和所述第二探针中的第二桥寡核苷酸特异性序列互补的序列,和与所述条形码环状寡核苷酸中的第三桥寡核苷酸特异性序列和第四桥寡核苷酸特异性序列互补的序列;

[0020] 并且其中,可选地,所述第一探针、所述第二探针、所述条形码环状寡核苷酸、所述桥寡核苷酸或者所述多个桥寡核苷酸的寡核苷酸中的至少一者包括核酸内切酶的识别序列;

[0021] (ii)对于所述一种或多种靶核苷酸序列中的每一种,优选对于在单独管中的每个样品,使所述第一探针和第二探针与所述条形码环状寡核苷酸和所述桥寡核苷酸或多个桥寡核苷酸接触并允许自退火为连接复合体;

[0022] (iii)使存在于待测试所述靶核苷酸序列的每个样品中的核酸与所述连接复合体接触;

[0023] (iv)允许所述第一探针的第一靶标特异性部分和所述第二探针的第二靶标特异性部分与靶标序列上基本相邻的区段杂交,从而形成杂交复合体;

[0024] (v)可选地,将来自所述多个样品的杂交复合体汇集起来;

[0025] (vi)将所述杂交复合体中的探针连接起来,以提供经连接的连接复合体;

[0026] (vii)使用滚环扩增法,用链置换聚合酶从一个或多个经连接的连接复合体来扩增核酸,从而获得单链的多联体(concatemeric)序列;

[0027] (viii)可选地,在存在步骤(i)中指定的识别序列的条件下,进行通过以下方式获得核酸片段的步骤:

[0028] (a)切割在步骤(vii)中获得的单链的多联体序列,或

[0029] (b)使在步骤(vii)中获得的单链的多联体序列与包含核酸内切酶的识别序列的特异性寡核苷酸退火,其中所述寡核苷酸与步骤(i)中指定的所述识别序列退火,从而获得所述核酸内切酶的识别位点,以及用所述核酸内切酶切割经退火的复合体;

[0030] (ix)使步骤(vii)中获得的所述多联体序列或步骤(viii)中获得的核酸片段经受高通量测序技术以确定所述条形码序列;以及

[0031] (x)通过确定所述第一靶标特异性部分和/或第二靶标特异性部分的至少一部分,和/或与所述条形码环状寡核苷酸中的条形码对应的条形码序列的至少一部分,来鉴定所述多个样品中靶核苷酸序列的存在和/或数量,

[0032] 其中,步骤(v)和步骤(vi)可按任何顺序进行。

附图说明

[0033] 图1示出了根据本发明实施方式的多重连接测定(Multiplexed Ligation Assay, MLA)的流程图。

[0034] 图2A、图2B、图2C和图2D示出了根据本发明实施方式的探针的原理组合。

[0035] 图3示出了被限制性内切酶消化之前(泳道2)和之后(泳道1)的工作流程中的RCA产物。

[0036] 图4示出了通过列举分子条形码从下一代DNA测序数据中推断出来,在四个重复反应中,实验工作流程对基因靶标数量的对数递减的线性响应。每行代表靶标序列的三个浓度。在三个数量级上的响应是线性的。

具体实施方式

[0037] 定义:

[0038] 靶标核苷酸序列:术语“靶标核苷酸序列”可以是需要检测的任何感兴趣的核苷酸序列。应当理解,给出的术语是指连续核苷酸序列以及具有互补序列的核酸分子。在一些实施方式,靶标序列是代表多态性或与多态性相关的核苷酸序列。

[0039] 多态性:术语“多态性”是指在群体中出现两种或更多种基因决定的替代序列或等位基因。多态性标志物或位点是发生序列差异的基因座。多态性基因座 (polymorphic locus) 可以小到一个碱基对。

[0040] 样品:术语“样品”在本文中用于含有两种或更多种靶标序列的两个或更多个样品。根据本发明方法提供的样品可已制备好,以便至少提取靶标核酸,并使那些核酸可被本发明中使用的探针接近。特别地,在一些实施方式中,每个样品包含至少两种不同的靶标序列,优选至少100种,更优选至少250种,更优选至少500种,最优选至少2000种或更多。术语“样品”可以指但不限于从人体/动物体获得的两个或更多个样品,包括尿液、活组织检查样品、唾液和其他分泌物、呼出的水分提取物、组织、血浆(液体活组织检查样品);或从环境获得的两个或更多个样品,包括水、废水、土壤、植物;或含病毒或细菌等的两个或更多个样品。在一个实施方式中,多个样品包括血液样品、唾液样品、尿液样品或粪便样品、另一体液的样品或身体材料的提取物,例如头发或皮屑。

[0041] 探针:术语“探针”是可变长度(通常50至1000个碱基长,优选50至200个碱基长)的DNA或RNA片段,其可用于DNA或RNA样品以检测与探针中序列互补的核苷酸序列(DNA或RNA靶标)的存在。寡核苷酸探针与靶标序列互补的区段被设计成使得对于样品中的每种靶标序列,提供一对第一探针和第二探针,由此每个探针在其末端含有与靶标序列的一部分互补的区段。此外,本公开还描述了用于接合第一探针和第二探针的桥寡核苷酸或桥寡核苷酸复合体。此外,本公开描述了条形码环状寡核苷酸,其包括侧翼为两个区段的环状区段,这个两个区段可与一个或多个桥寡核苷酸杂交。环状区段不与一个或多个桥寡核苷酸杂交,并包括条形码。

[0042] 通用(Universal):当用于描述扩增过程时,术语“通用”是指能够使用单个引物或一组引物进行多个扩增反应的序列。这种引物的使用极大地简化了多重化,因为只需要两个引物来扩增多个选定的核酸序列。当用于描述引发位点时,术语“通用”是通用引物将杂交的位点。还应该注意的,可以使用通用引发序列/引物的“组”。

[0043] 杂交:术语“杂交(或杂化)”描述DNA或RNA分子退火成互补DNA或RNA的过程。DNA或RNA复制以及DNA转录成RNA都依赖于核苷酸杂交。

[0044] 连接(Ligation):术语“连接”是通过酶的作用接合两个核酸片段。DNA连接酶是能够催化在互补链上相邻位点结合的两条多核苷酸链(的端部)之间形成磷酸二酯键的酶。在一个实施方式中,连接也可以化学方式进行,特别地如果多核苷酸的两个相邻的端部都被修饰以使得能够化学连接。

[0045] 扩增:本文使用的术语“扩增”指的是使用DNA聚合酶来增加核苷酸序列混合物中具体核苷酸序列的浓度。“PCR”或“聚合酶链式反应”是体外酶促扩增特定DNA/RNA片段的快速程序。待扩增的DNA/RNA可通过加热样品而变性。术语“引物”是RNA或DNA的链(通常约18至22个碱基),其作为DNA合成的起点。这是复制DNA所必需的,因为催化这一过程的酶,DNA聚合酶,只能在现有的DNA链上添加新的核苷酸。

[0046] 聚合酶:聚合酶是一种合成长链核酸或核酸聚合物的酶。DNA聚合酶和RNA聚合酶分别用于通过碱基配对相互作用复制DNA或RNA模板链来组装DNA和RNA分子。

[0047] 高通量:术语“高通量”指的是同时处理和筛选大量DNA样品的能力;以及在单个DNA样品中同时筛选大量不同的基因座的能力。高通量测序或筛选,通常缩写为HTS,是一种特别适用于同时有效筛选大量样品的科学实验方法。

[0048] 核酸内切酶:核酸内切酶是一种在随机或特定位置使DNA双链或单链裂解或有缺口的酶。

[0049] 条形码:本发明中使用的探针和寡核苷酸可包括由核苷酸序列组成的一种或多种条形码。条形码序列可包括用于靶标列举(target enumeration)的靶标核苷酸序列标识符序列、样品标识符序列和/或分子条形码(也称为独特分子标识符)。条形码序列可包括随机序列。

[0050] 如上文所述,本公开涉及一种利用连接依赖性测定法在非常多的样品中高通量检测靶标核苷酸序列的方法。本公开提供了一种使用下一代测序允许的技术确定复杂核酸池中基因靶标序列的方法。本公开还提供了一种通过利用连接依赖性测定法在多个样品中(优选在非常多的样品中),对多种基因靶标进行剖析的方法。本公开提供了一种多重连接依赖性探针扩增的方法,其使得能够查询多个样品中的不同靶标核酸。本发明的方法使得能够对多个样品中的一种或多种靶标核苷酸序列进行测序,为不同的靶标核酸提供多个不同的探针组。在处理测序数据时,独特序列标识符用于鉴定基因靶标和对样品池中的单个样品进行绝对定量。

[0051] 在第一个主要方面,本发明涉及一种高通量检测多个样品中一种或多种靶标核苷酸序列的方法,所述方法包括以下步骤:

[0052] (i) 为每个样品中的每种靶标核苷酸序列提供:第一探针、第二探针、条形码环状寡核苷酸,和桥寡核苷酸或能够与所述条形码环状寡核苷酸退火以形成桥寡核苷酸复合体的多个桥寡核苷酸,

[0053] 其中,所述第一探针包括位于所述第一探针的5'端的第一桥寡核苷酸特异性序列和位于所述第一探针的3'端的第一靶标特异性部分;

[0054] 其中,所述第二探针包括位于所述第二探针的5'端的第二靶标特异性部分和位于所述第二探针的3'端的第二桥寡核苷酸特异性序列;

[0055] 其中,所述条形码环状寡核苷酸自该分子的5'端起包括第三桥寡核苷酸特异性序列、条形码环状序列和第四桥寡核苷酸特异性序列;

[0056] 其中,所述桥寡核苷酸或多个桥寡核苷酸包括与所述第一探针中的第一桥寡核苷酸特异性序列和所述第二探针中的第二桥寡核苷酸特异性序列互补的序列,和与所述条形码环状寡核苷酸中的第三桥寡核苷酸特异性序列和第四桥寡核苷酸特异性序列互补的序列;

[0057] 并且其中,可选地,所述第一探针、所述第二探针、所述条形码环状寡核苷酸、所述桥寡核苷酸或者所述多个桥寡核苷酸的寡核苷酸中的至少一者包括核酸内切酶的识别序列;

[0058] (ii)对于所述一种或多种靶标核苷酸序列中的每一种,优选对于在单独管中的每个样品,使所述第一探针和第二探针与所述条形码环状寡核苷酸和所述桥寡核苷酸或多个桥寡核苷酸接触并允许自退火为连接复合体;

[0059] (iii)使存在于待测试所述靶标核苷酸序列的每个样品中的核酸与所述连接复合体接触;

[0060] (iv)允许所述第一探针的第一靶标特异性部分和所述第二探针的第二靶标特异性部分与靶标序列上基本相邻的区段杂交,从而形成多个杂交复合体(或可选的杂交复合体);

[0061] (v)可选地,将来自所述多个样品的杂交复合体汇集起来;

[0062] (vi)将所述杂交复合体中的探针连接起来,以提供经连接的连接复合体;

[0063] (vii)使用滚环扩增法,用链置换聚合酶从一个或多个经连接的连接复合体扩增核酸,从而获得单链的多联体序列;

[0064] (viii)可选地,在存在步骤(i)中指定的识别序列的条件下,进行通过以下方式获得核酸片段的步骤:

[0065] (a)切割在步骤(vii)中获得的单链的多联体序列,或

[0066] (b)使在步骤(vii)中获得的单链的多联体序列与包含核酸内切酶的识别序列的特异性寡核苷酸退火,其中所述寡核苷酸与步骤(i)中指定的所述识别序列退火,从而获得所述核酸内切酶的识别位点,以及用所述核酸内切酶切割经退火的复合体;

[0067] (ix)使步骤(vii)中获得的所述多联体序列或步骤(viii)中获得的核酸片段经受高通量测序技术以确定所述条形码序列;以及

[0068] (x)通过确定所述第一靶标特异性部分和/或第二靶标特异性部分的至少一部分,和/或与所述条形码环状寡核苷酸中的条形码对应的条形码序列的至少一部分,来鉴定所述多个样品中靶标核苷酸序列的存在和/或数量,

[0069] 其中,步骤(v)和步骤(vi)可按任何顺序进行。

[0070] 图1提供了本发明方法实施方式的非限制性说明。

[0071] 本发明的方法利用四种或更多种的核酸分子,其中两种靶标特异性的核酸探针(第一探针和第二探针)对基因靶标具有特异性,另外两种或更多种的其他核酸探针通常是通用的(桥寡核苷酸或桥寡核苷酸复合体和条形码环状寡核苷酸)。第一探针、第二探针和条形码环状寡核苷酸与一个或多个桥探针杂交形成连接复合体。使样品DNA或RNA上具有靶标鉴定位点的连接复合体(含有一种或多种条形码序列)与查询样品的互补靶标序列进行杂交。杂交后,第一探针和第二探针被化学连接或通过DNA连接酶酶法连接,形成经连接的连接复合体。在本发明中,在待分析的多个样品的样品分析过程中,将会形成多个这样的经连接的连接复合体。

[0072] 术语“多个样品”可以指但不限于从人体或动物体获得的两个或更多个样品,包括活组织检查样品、唾液和其他分泌物、呼出的水分提取物、组织、血浆(液体活组织检查样品);从环境获得的两个或更多个样品,包括水、废水、土壤、植物;或者含病毒或细菌的两个

或更多个样品等。在一个实施方式中,使用的样品不需要事先进行核酸的纯化或浓缩。在另一个实施方式中,可以对样品进行预处理,例如裂解细胞以暴露核酸。

[0073] 靶标序列可包括需要检测的任何感兴趣的核苷酸序列。本公开的靶标核苷酸序列可从患者血液中的一部分DNA或母体血液中的一部分DNA获得,但不限于此。患者血液中的一部分DNA可例如从凋亡/坏死的癌细胞中获得,或者母体血液中的一部分DNA可以是胎儿或和/或母体来源的。进一步地,分析结果用于,例如评估个体患上给定类型癌症的风险,确定给定治疗对给定癌症的疗效,肿瘤中耐药性相关突变的发展,或胎儿携带基因疾病(诸如常见的三染色体性唐氏综合征、帕图综合征和爱德华兹综合征)的风险。在某些实施方式中,该方法包括为每种靶标核苷酸序列提供多个不同的探针组。

[0074] 如本文所使用的,术语“探针组”包括第一探针、第二探针、条形码环状寡核苷酸,以及一个或多个桥寡核苷酸。

[0075] 在某些实施方式中,第一探针自分子的5'端起包括可选的5'磷酸、第一桥寡核苷酸特异性序列、可选的第一通用序列、可选的第一序列条形码,和在其3'端的第一靶标特异性部分。

[0076] 在某些实施方式中,第二探针自分子的5'端起包括可选的5'磷酸、第二靶标特异性部分、可选的第二序列条形码、可选的第二通用序列,和在其3'端的第二桥寡核苷酸特异性序列。

[0077] 桥寡核苷酸或多个桥寡核苷酸包括与所述第一探针中的第一桥寡核苷酸特异性序列和所述第二探针中的第二桥寡核苷酸特异性序列互补的序列,可选的通用序列,和与所述条形码环状寡核苷酸中的第三桥寡核苷酸特异性序列和第四桥寡核苷酸特异性序列互补的序列。

[0078] 条形码环状寡核苷酸自分子的5'端起包括第三桥寡核苷酸特异性序列、条形码环状序列和第四桥寡核苷酸特异性序列。条形码可用于在所有测试的样品中的所有连接复合体中唯一地限定复合体。条形码环状寡核苷酸可以有任何合适的长度,例如,长度可在30和100bp之间,诸如40和60bp之间。

[0079] 可选地,第一探针、第二探针、条形码环状寡核苷酸,或者一个或多个桥寡核苷酸中的至少一者包括核酸内切酶的识别序列。核酸内切酶识别序列能够切割多联体序列。在一个实施方式中,识别序列是限制性核酸内切酶(诸如EcoRI)的识别序列。在另一个实施方式中,识别序列是归巢核酸内切酶(诸如I-CeuI)的识别序列。在另一个实施方式中,识别序列是向导DNAaseI或CRISPR-Cas类切割体系的识别序列。在另一个实施方式中,识别序列是缺口核酸内切酶的识别序列。

[0080] 在一个实施方式中,所述条形码环状寡核苷酸含有核酸内切酶(诸如缺口核酸内切酶)的一个或多个识别序列。在又一个实施方式中,所述条形码环状寡核苷酸含有能够相互退火的两个识别序列,从而获得双链核酸内切酶识别位点。在本文又一个实施方式中,所述方法不包括上文中指定的步骤(viii),但(相反)在步骤(vii)与步骤(ix)之间包括以下步骤:使所述条形码环状寡核苷酸中的两个识别序列退火,并用对所述识别位点具有特异性的核酸内切酶(诸如缺口核酸内切酶)来切割所得双链核酸内切酶识别位点。

[0081] 可选地,第一探针、第二探针、条形码环状寡核苷酸,或一个或多个桥寡核苷酸中的至少一者包括第一捕获部。在本文在使用时,“第一捕获部”是指使得探针、连接复合体或

杂交复合体被连接至固体支持物的第二捕获部所捕获(即与之结合)的部分,诸如化学基团。本领域内已知的任何合适的捕获部都可用于此目的。众所周知的合适例子是使用包被有链霉亲和素的磁珠捕获生物素化的分子。因此,在一个实施方式中,第一捕获部是生物素部,它可以与连接到固体支持物(诸如磁珠)上的链霉亲和素或亲和素部(第二捕获部)相互作用。其他选择包括生物素衍生物,诸如双生物素、脱硫生物素或可光解的生物素,它们可用于与链霉亲和素/亲和素缀合。进一步的选择包括使用硫醇和丙烯酰胺(acrydite)基团进行丙烯酰胺(acrydite)/丙烯酰胺(acrylamide)缀合,使用炔基和叠氮基进行点击化学,和使用地高辛用于与抗地高辛抗体缀合。缀合配偶体(conjugation partner)可以设置在任何固体表面上,诸如珠子(磁性或其他)或固体支持物。因此,在本发明方法的一个实施方式中,所述第一探针、第二探针、条形码环状寡核苷酸,或者一个或多个桥寡核苷酸中的至少一者包括第一捕获部,并且,在步骤(iv)和步骤(v)之间进行包括以下的中间步骤(iv)(a):使杂交复合体与含第二捕获部的固体支持物接触,允许所述第一捕获部与第二捕获部相互作用,从而使得所述杂交复合体与所述固体支持物相连;以及使与固体支持物相连的杂交复合体与样品中未与所述固体支持物相连的组分分离。

[0082] 可选地,第一靶标特异性部分、第二靶标特异性部分、第一桥寡核苷酸特异性序列、第二桥寡核苷酸特异性序列、第三桥寡核苷酸特异性序列和/或第四桥寡核苷酸特异性序列彼此独立地包含至少一个经化学修饰的核苷酸以增加探针结合。增加探针结合的化学修饰包括但不限于核糖核酸、肽核酸和锁定核酸(例如,如W02019038372的图3所示,该文件通过引用并入本文)。在一个实施方式中,第一探针、第二探针、条形码环状寡核苷酸或所有这三者的桥接部分包括经化学修饰的碱基,以允许改善与一个或多个桥寡核苷酸的结合。在另一实施方式中,第一靶标特异性部分、第二靶标特异性部分、第一桥寡核苷酸特异性序列、第二桥寡核苷酸特异性序列、第三桥寡核苷酸特异性序列和/或第四桥寡核苷酸特异性序列彼此独立地包含一个或多个经化学修饰的核苷酸。在某些实施方式中,化学修饰允许相邻探针的化学连接。

[0083] 在一些实施方式中,上文描述的探针与完全相邻的基因基因座,即靶标核苷酸序列的相邻区段,或者与相距至多500个碱基对,例如相距至多200个碱基对,诸如相距至多50个碱基对,相距至多40个碱基对,相距至多30个碱基对,相距至多20个碱基对,相距至多10个碱基对或相距至多5个碱基对结合。

[0084] 在一些实施方式中,第一探针、第二探针、条形码环状寡核苷酸,或者一个或多个桥寡核苷酸可包括用于DNA测序平台诸如(但不限于)Illumina的接头序列。这些接头序列允许得到的测序文库与测序设备的检测部件(诸如Illumina流动池)结合。

[0085] 此外,在一些实施方式中,桥寡核苷酸或者多个桥寡核苷酸中的寡核苷酸包括:

[0086] (i)一至五个3'突出的碱基(即,不与第二探针形成双螺旋的额外碱基),和/或

[0087] (ii)3'磷酸,和/或

[0088] (iii)在从3'端起的三个位置内的一个或多个硫代磷酸酯修饰。

[0089] 优选对于在单独的管中的每个样品,在探针与含靶标序列的样品接触之前,使第一探针、第二探针和条形码环状寡核苷酸与桥寡核苷酸(或能够形成桥寡核苷酸复合体的多个寡核苷酸)接触,并允许退火成连接复合体(步骤(ii))。在与第一探针和第二探针退火之前,可对桥寡核苷酸与条形码环进行预退火,或者所有的退火步骤可以一次性完成。在其

中桥不是一个寡核苷酸而是桥是能够与条形码环寡核苷酸退火形成桥寡核苷酸复合体的多个寡核苷酸(诸如两个寡核苷酸)的实施方式中(本文中在2B示出),该多个寡核苷酸可以在与第一和第二探针退火前与条形码环状寡核苷酸预退火,或者所有的退火步骤可以一次完成,或者条形码环状寡核苷酸可以在靶标捕获期间与经预退火的探针复合体退火。

[0090] 优选地,每个连接复合体对第一靶标特异性序列、第二靶标特异性序列和一个或多个条形码序列的组合是独特的。这使得能够在扩增后对靶标序列进行列举,并对结果进行分析。

[0091] 此后,使多个样品中的一个或多个靶标核苷酸序列与多个连接复合体接触(步骤(iii))。第一探针的第一靶标特异性部分和第二探针的第二靶标特异性部分与靶标序列上基本相邻的区段杂交,从而形成杂交复合体(步骤(iv))。如上文所提及的,靶标序列上基本相邻的区段可紧密相邻,也可以存在多达500个碱基对的间隙。

[0092] 在一些实施方式中,样品的体积超过100微升,例如超过1ml。在又一实施方式中,样品的核酸浓度低于5pmol,诸如低于1pmol,例如低于200fmol。在一个实施方式中,多个样品包括一个或多个血液样品、一个或多个唾液样品、一个或多个尿液样品,或者一个或多个粪便样品。

[0093] 随后,在一些实施方式中,如果第一探针、第二探针、条形码环状寡核苷酸或一个或多个桥寡核苷酸中的至少一者包含第一捕获部,则使杂交复合体与含第二捕获部的固体支持物接触,并允许第一捕获部与第二捕获部相互作用,使杂交复合体与固体支持物相连(可选步骤(iv)(a))。此后,使与固体支持物相连的杂交复合体与样品中未与固体支持物相连的组分分离。如果固体支持物是磁珠,则可以用磁铁将磁珠固定住,并可去除剩余的液体样品。可选地,在继续进行之前进行洗涤步骤。

[0094] 步骤(iv)(a)带来核酸的纯化和富集,使得特别是对于高度不纯的样品具有改善的结果。在一个实施方式中,本发明的方法在步骤(iv)(a)之前不包括富集核酸的步骤。因此,在一个实施方式中,本方法在步骤(iv)之前不包含将原始样品中的核酸浓缩大于2倍、大于10倍或大于100倍的步骤。在另一个实施方式中,本发明的方法在步骤(vi)的连接之后不包括纯化步骤。

[0095] 随后,以酶法或化学法进行所形成的杂交复合体中探针的连接,以提供经连接的连接复合体(步骤(vi))。可选地,作为步骤(vi)的一部分,第一探针与第二探针之间、第一探针与条形码环状寡核苷酸之间和/或第二探针与条形码环状寡核苷酸之间的间隙(如果存在)可以通过引入聚合酶和一个或多个核苷酸来填补。聚合酶添加与桥寡核苷酸序列互补的核苷酸(a),从而填补第一探针、第二探针和条形码环状寡核苷酸之间的间隙,使探针连接起来并将条形码环状寡核苷酸纳入桥互补链中。桥寡核苷酸或桥寡核苷酸复合体从5'位点或3'位点延伸与连接的探针互补,从而使存在于第一探针或第二探针中的靶标序列识别序列整合到桥寡核苷酸或桥寡核苷酸复合体中。优选地,使用不破坏双链DNA的聚合酶,例如Taq聚合酶,以便在第一探针和第二探针与靶标序列退火时均不干扰第一探针与第二探针的连接。在一个实施方式中,所述桥寡核苷酸,或所述多个桥寡核苷酸的一个或多个寡核苷酸,在与所述第一探针、第二探针或条形码环状寡核苷酸不互补的区域中,包括多个通用碱基类似物,以允许引入适合用作靶标列举的分子条形码的随机序列。在这样的实施方式中,作为步骤(vi)的一部分,使用聚合酶和核苷酸进行间隙填补步骤,以生成这样的随机

序列。在实施方式中,多个通用碱基类似物是多个5-硝基吡啶。

[0096] 然后从一个或多个靶标样品中可选地汇集连接复合体(步骤(vi))。步骤(v)和步骤(vi)可按指定顺序进行或可替代地以相反顺序进行。

[0097] 接下来,从一个或多个经连接的连接复合体中扩增核酸,从而获得单链多联体序列(步骤(vii))。使用链置换聚合酶,诸如phi29聚合酶(UniProtKB-P03680;DPOL_BPPH2)或Bst聚合酶(P52026;DP01_GEOSE)通过滚环扩增来进行扩增。

[0098] 在一个实施方式中,在步骤(v)之后,但在步骤(vii)之前,进行步骤(a)和步骤(b),其中步骤(a)包括允许经连接的连接复合体与靶标核苷酸序列解离,并且步骤(b)包括加入靶标特异性探针,所述靶标特异性探针包括与所述靶标核苷酸序列对应的序列,其中所述靶标特异性探针能够与经连接的连接复合体退火,并允许所述靶标特异性探针与经连接的连接复合体退火,从而形成扩增模板。在此种实施方式中,在步骤(vii)中,用链置换聚合酶通过滚环扩增对所述扩增模板进行扩增。

[0099] 可选地,在存在步骤(i)中指定的识别序列的情况下,步骤(viii)通过以下来进行以获得核酸片段:

[0100] (a) 切割在步骤(vii)中获得的单链的多联体序列,或

[0101] (b) 使在步骤(vii)中获得的单链的多联体序列与包含核酸内切酶的识别序列的特异性寡核苷酸退火,其中所述寡核苷酸与步骤(i)中指定的所述识别序列退火,从而获得所述核酸内切酶的识别位点;以及,用所述核酸内切酶切割经退火的复合体。

[0102] 随后,在步骤(ix)中,使步骤(vii)中获得的所述多联体序列或者,如果进行了步骤(viii),则使步骤(viii)中获得的核酸片段经受高通量测序技术以确定所述条形码序列。

[0103] 可选地,扩增后,去除固体支持物(如果存在),并且将上清液用于后续处理。例如,如果固体支持物是磁性颗粒,可使用磁铁将它们去除。在本发明方法的其他一些实施方式中,在步骤(v)之后、步骤(vi)之后或步骤(vii)之后,立即破坏第一捕获部与第二捕获部之间的相互作用。例如,如果第一捕获部是生物素且第二捕获部是链霉亲和素,则可通过加入过量的可溶性生物素来破坏相互作用。如果链霉亲和素与磁性颗粒结合,则可随后用磁铁将其去除。

[0104] 此外,可选地,紧邻步骤(ix)且在步骤(ix)之前,使用与所述第一探针和第二探针的通用部分结合的引物进行PCR扩增,其中所述引物可选地包括用于后续步骤(ix)中的测序的接头序列。

[0105] 在另一个实施方式中,步骤(ix)中的测序是使用纳米孔测序进行的,其中可选地使用转座复合体对步骤(vii)中获得的多联体序列进行片段化。Wang等人(2021Nat Biotechnol39(11):1348)对用于纳米孔测序的合适技术进行了综述。

[0106] 可以通过高通量测序技术(步骤(ix)和步骤(x))确定第一和/或第二靶标特异性部分的至少一部分和/或条形码的至少一部分来鉴定多个样品中靶标核苷酸序列的存在和/或数量,例如使用下一代测序平台,包括但不限于Illumina iSeq、MiSeq、HiSeq、NextSeq或NovaSeq。优选地,通过对每种靶标和每个样品的分子条形码的数量进行计数来允许对基因靶标进行列举(enumeration)。样品从序列数据中分离出来(去卷积),并且在DNA测序后在silico中对序列靶标进行定量。

[0107] 与传统的核酸测序技术相比,本发明的优点包括但不限于低成本、高简单性、高特异性、高灵敏度、高准确性、高通量、高可扩展性和高周转性的定量测定。本发明的另一个方面是,由于在工作流程的初始阶段进行了样品索引,本发明的方法可将多个样品汇集在一起,从而在测定成本和速度方面提供了改善。本发明的另一个方面是,本发明的方法使得能够对包括人类和动物群体在内并且包括大体积的未纯化样品材料在内的多个样品中的多种核酸靶标进行精确和大规模平行定量。如所提及的,在优选的实施方式中,使用的样品,诸如尿液样品,不需要事先进行核酸的纯化或浓缩。在另一个实施方式中,可以对样品进行预处理,例如裂解细胞以暴露核酸。本发明的一个特别的优点是能够使用独特的探针设计(即探针三联体),来检测和扩增感兴趣的靶标序列。探针设计有特殊定位的经修饰的核苷酸,可提高退火和结合效率。结合特性的改善引起更高的测定特异性、灵敏度和准确性。本发明的方法同样适用于研究基因变异体并在诊断和预后中找到应用,包括但不限于对一种或多种序列和/或多态性(诸如SNP和/或插入缺失)、癌症诊断或来自母体血液的胎儿染色体疾病的样品进行基因分型。在优选的实施方式中,对于两个或更多个样品或者对于两个或更多个基因座/等位基因组合,使用条形码序列来对样品中的一个或多个序列和/或多态性(诸如SNP和/或插入缺失)进行基因分型。

[0108] 在另一方面,本发明提供了含多个容器的成套的试剂盒(a kit of parts),其中,至少一个容器容纳一组或多组的第一探针和第二探针,至少一个容器容纳条形码环状寡核苷酸,并且至少一个容器容纳一种或多种桥寡核苷酸或能够与所述条形码环状寡核苷酸形成桥寡核苷酸复合体的多个桥寡核苷酸。其中,所述第一探针包括位于所述第一探针的5'端的第一桥寡核苷酸特异性序列和位于所述第一探针的3'端的第一靶标特异性部分;

[0109] 其中,所述第二探针包括位于所述第二探针的5'端的第二靶标特异性部分和位于所述第二探针的3'端的第二桥寡核苷酸特异性序列;

[0110] 其中,所述条形码环状寡核苷酸自分子的5'端起包括第三桥寡核苷酸特异性序列、条形码环状序列和第四桥寡核苷酸特异性序列;

[0111] 其中,所述桥寡核苷酸或多个桥寡核苷酸包含与所述第一探针中的第一桥寡核苷酸特异性序列和所述第二探针中的第二桥寡核苷酸特异性序列互补的序列,和与所述条形码环状寡核苷酸中的第三桥寡核苷酸特异性序列和第四桥寡核苷酸特异性序列互补的序列;

[0112] 其中,可选地,所述第一探针、所述第二探针、所述条形码环状寡核苷酸、所述桥寡核苷酸和所述多个桥寡核苷酸的寡核苷酸中的至少一者包括核酸内切酶的识别序列;

[0113] 并且其中,可选地,所述成套的试剂盒进一步包括能够与所述识别序列退火的寡核苷酸,从而获得所述核酸内切酶的识别位点。

[0114] 优选地,对所述多个第一探针的3'端或所述多个第二探针的5'端,或这两者进行修饰以允许所述多个第一探针与所述多个第二探针的化学连接。

[0115] 优选地,桥寡核苷酸或多个桥寡核苷酸中的寡核苷酸在与第一探针序列互补的序列,或与第二探针序列互补的序列,或这两者中包括一个或多个经化学修饰的核苷酸。

[0116] 优选地,对所述第一探针的3'端或所述第二探针的5'端,或这两者进行修饰以允许所述第一探针与所述第二探针的化学连接。

[0117] 优选地,所述第一探针或第二探针的桥接部分,或这两者,或所述条形码环状寡核

昔酸、所述桥寡核苷酸或所述多个桥寡核苷酸中的寡核苷酸包括经化学修饰的碱基,以改进与所述桥寡核苷酸或桥寡核苷酸复合体的结合。

[0118] 在一个具体实施方式中,容纳第一和第二探针组的至少一个容器、容纳条形码环状寡核苷酸的至少一个容器,和容纳桥寡核苷酸或能够彼此退火以形成桥寡核苷酸复合体的多个寡核苷酸的至少一个容器是同一容器。在这种情况下,四种或更多种的探针可以预先退火并形成连接的复合体。

[0119] 本发明的一个特别的优点是能够使用独特的探针设计,来检测和扩增感兴趣的靶标序列。探针被设计为具有引起更高的测定特异性、灵敏度和准确性的改善的结合特性。本发明应用于分子生物学、进化生物学、元基因组学、基因分型领域,更具体地说,但不限于癌症诊断或胎儿染色体疾病,包括但不限于对样品的一个或多个序列和/或多态性(诸如SNP和/或插入缺失)进行基因分型。

[0120] 在一个尤其优选的实施方式中,条形码环状寡核苷酸包括用于鉴定样品的信息,并包括独特的条形码。在这种情况下,第一和第二探针普遍适用于所有样品(并且仅包含用于鉴定靶标的信息)。因此,在一个优选的实施方式中,提供了根据本发明的方法或试剂盒,其中条形码环状寡核苷酸包括含独特序列的条形码,该独特序列使得能够列举每个样品的靶标序列。

[0121] 实施例

[0122] 方法

[0123] 1、探针复合体的形成

[0124] 探针复合体包含基因组靶向、样品索引和构建Illumina测序文库所需的序列。

[0125] 能够形成包括以下的四部分探针复合体(如图2所示):

[0126] (a) 第一探针,从分子的5'端起具有第一桥寡核苷酸特异性序列,和位于第一探针的3'端的第一靶标特异性部分;

[0127] (b) 第二探针,从分子的5'端起具有第二靶标特异性部分、第二序列条形码和位于第二探针的3'端的第二桥寡核苷酸特异性序列;

[0128] (c) 桥寡核苷酸,其具有分别与第一探针中的第一桥寡核苷酸特异性序列和第二探针中的第二桥寡核苷酸特异性序列互补的序列;

[0129] 和(d) 环状寡核苷酸,其具有与桥寡核苷酸互补的序列。

[0130] 通过在退火反应中使所有三部分(桥、右臂和左臂)以等摩尔量结合来构建探针复合体。反应在热循环仪中进行(退火程序见表1)。

[0131] 表1.

步骤	温度	时间
1	+95°C	5 min
2	+95°C	1 min
-4°C/循环, 到 2 10x		
3	+55°C	10 min
4	+55°C	1 min
-5°C/循环, 到 5 5x		
5	+4°C	保持

[0132] 2、靶标捕获

[0134] 靶向含有感兴趣突变的特定基因组区域。经纯化的DNA(例如来自组织、血浆、尿液或唾液)可以用作样品,或者样品可以是未纯化的,只是经过预处理,例如煮沸和/或离心。

[0135] 使探针复合体通过碱基序列互补作用与靶标区域杂交。为了启动靶标捕获,反应探针和靶标DNA混合并在热循环仪中孵育(靶标捕获和间隙填补程序见表2)。将预先退火的条形码环(带有每个样品的特定索引序列)加入到靶向反应中。

[0136] 表2.

步骤	温度	时间	过程
1	+85°C	4 min	变性
2	+75°C	2.5 min	
3	+65°C	2.5 min	
4	+55°C	120 min	靶标捕获
5	+50°C	10 min	间隙填补
6	+45°C	45 min	
7	+4°C	保持	

[0137] 3、间隙填补(GapFill)反应

[0139] 靶标捕获后,汇集来自不同靶向反应的探针复合体,随后加入Phusion DNA聚合酶、核苷酸和Ampligase DNA连接酶,并在+45°C下孵育45分钟来延伸和连接。

[0140] 4、核酸外切酶处理

[0141] 间隙填补后,通过加入1 μ l的不耐热的核酸外切酶1(NEB,#M0568L)和1 μ l的RecJF核酸外切酶(NEB,#M0264L),并在+37°C下孵育30分钟,来去除线性分子。核酸外切酶通过在+92°C下孵育12分钟来失活。

[0142] 5、滚环扩增

[0143] 延伸和连接后,对环形探针分子进行滚环扩增(RCA)。在RCA反应中,将靶标捕获反应与含EquipPhi29(Thermo Scientific)聚合酶的RCA反应混合物混合。反应在+42°C下孵育30min至2小时。RCA反应后,通过用Qubit荧光仪测量单链DNA(ssDNA)浓度来分析反应效率。

[0144] 6、酶消化

[0145] RCA反应产生长的多联体ssDNA分子,其具有多个靶标文库拷贝。每个完整的靶标文库由EcoRI限制性酶识别序列分隔。该序列通过与含EcoRI限制性酶识别序列的特异性寡核苷酸退火,能够对长的多联体进行序列特异性切割,并释放出准备好的靶标文库。RCA产物在+37°C下用EcoRI消化1小时。

[0146] 7、文库PCR

[0147] 经消化的RCA产物在PCR反应中延伸到测序文库中,其中存在于右探针中的截短测序接头延伸到流式细胞兼容的全长测序接头中。

[0148] 8、文库纯化

[0149] 文库PCR后,通过电泳后从琼脂糖凝胶中提取文库分子或使用粒径选择珠(诸如Macherey Nagel NucleoMag)纯化文库分子。

[0150] 9、测序

[0151] 可使用最先进的测序仪器对经纯化的MiSeq或iSeq100兼容文库进行序列分析。重要的是,这些文库可以通过简单的寡核苷酸修饰转换成适合任何现有测序平台的文库。结合使用Unix命令行工具以及Python和R编程语言来处理测序数据。简而言之,序列处理的原理是鉴定每个读数中的探针序列,对它们之间的基因组区域进行测序,并对与每个基因靶标相关的分子条形码的数量进行计数。

[0152] 实验1

[0153] 在第一个实验中,探针混合物是五个独立样品中的4个环状索引探针复合体的集合,其中前四个样品具有单一环状索引,第五个样品为汇集在一起的1至4样品。它们靶向四个基因融合体。靶标寡核苷酸具有唯一识别序列,允许鉴定每个靶标。

[0154] 作为样品,三种合成的靶标寡核苷酸以对数递增的浓度混合。如上所述进行靶标捕获、延伸和连接反应、滚环扩增、随后的EcoRI消化和文库制备PCR。图3是得到的测序文库的例子。

[0155] 用MiSeq和iSeq100仪器对准备好的文库进行测序,通过匹配每个读数中的探针序列、鉴定探针序列之间的基因组序列区域以及对分子条形码进行计数来检测序列数据中的靶标区域。计数数据准确地反映了汇集的环状寡核苷酸分子的比例(图4)。

[0156] 图1和图2的详细说明

[0157] 图1示出了所描述发明的一个实施方式的工作流程。在步骤1中,使样品(102)中的核酸(DNA或RNA)与一组连接复合体(104)接触。连接复合体在靶标核酸(106)上退火。在步骤2中,从样品材料可选地捕获结合靶标的连接复合体,留下样品杂质(103)。在步骤3中,将来自多个样品(110)的经连接的连接复合体汇集在一起(112)。在步骤4中,将经退火的、汇集的连接复合体进行连接,得到经连接的连接复合体。在步骤5中,使用phi29聚合酶或其他链置换聚合酶通过滚环扩增来扩增探针序列,得到探针的长的多联体拷贝(116)。在步骤6中,使用限制性内切酶(诸如EcoRI)或归巢核酸酶(诸如I-CeuI)将多联体探针拷贝可选地切割为单体单元,并可选地使用PCR或乳液PCR进一步扩增(117)。在步骤7中,使用下一代DNA测序对经扩增的DNA进行测序。在步骤8中,使用生物信息学管线将DNA测序结果转换为靶标计数。

[0158] 图2A示出了根据本文实施方式的探针四件套的原理结构。多个探针实体包括第一探针(202)、第二探针(201)、桥寡核苷酸(200)和条形码环状寡核苷酸(217)。此处,探针复

合体包含第一探针与条形码环状寡核苷酸(210和213)之间、第二探针与条形码环状寡核苷酸(207和215)之间以及第一探针与第二探针(203和204)之间的间隙或缺口。通过引入聚合酶和一种或多种核苷酸来填补这些间隙。在此过程中,可使用Stoffel片段、Taq聚合酶或Phusion聚合酶和DNA连接酶(诸如Ampligase)的混合物。聚合酶填补这些间隙,随后DNA连接酶的作用使探针、桥和条形码环状寡核苷酸连接成环形复合体。

[0159] 第一探针的15-25个碱基包括桥结合序列(210),该序列可选地包括经化学修饰的碱基,用于有效桥接寡核苷酸结合。第一探针自5'端进一步包括与基因靶标结合的15-30个碱基(203)。210的一些或全部核苷酸可包括化学修饰,以增加探针对靶标或桥(209)的亲合力。第一探针的最后一个碱基可选地包括用于酶连接的磷酸部或允许化学连接到相邻探针的5'端的修饰(205)。

[0160] 第二探针的第一个碱基可选地包括用于酶连接的磷酸部或允许化学连接到相邻探针的5'端的修饰(206)。第二探针的5'端的15-30个碱基包括第二探针的与基因靶标结合的部分(204)。第二探针的最后15-25个碱基(207)与桥寡核苷酸(208)反向互补。203、204、207或210的一些或全部核苷酸可包括化学修饰,以增加探针对靶标或桥寡核苷酸的亲合力。

[0161] 桥寡核苷酸的5'端的前15-25个碱基(209)与第一探针的桥寡核苷酸特异性序列(210)反向互补,并可选地包括经化学修饰的核苷酸以增加结合。桥寡核苷酸的最后15-25个碱基(208)与第二探针的序列(207)反向互补,并可选地包括经化学修饰的核苷酸以增加结合。桥寡核苷酸的5'端可选地包括用于捕获连接复合体的捕获部(211)。此外,桥寡核苷酸包括序列214和216,与条形码环状寡核苷酸的序列213和215互补。桥寡核苷酸的3'端可选地包括磷酸(或其他可裂解的)部(212),以防止在间隙填补过程中延伸。

[0162] 条形码环状寡核苷酸的5'端的前15-25个碱基(215)与桥寡核苷酸序列216反向互补。条形码环状寡核苷酸包括含条形码(218)的环状区域。条形码环状寡核苷酸的最后15-25个碱基(213)与桥寡核苷酸序列214反向互补。

[0163] 图2B示出了根据本发明实施方式的具有多个探针实体的探针组的原理结构。探针与图2A中的那些探针相对应,除了本实施方式包括两个桥寡核苷酸(200和220),其中桥寡核苷酸200包括209和214,而桥寡核苷酸220包括208和216。

[0164] 图2C示出了根据本发明实施方式的具有多个探针实体的探针组的原理结构。探针与图2A中的那些探针相对应,除了桥寡核苷酸200包括含多个通用碱基类似物的序列(219),以允许引入适合用作靶标列举的分子条形码的随机序列。可存在一个或两个所标示的219序列。

[0165] 图2D示出了根据本发明实施方式的探针四件套的原理结构。多个探针实体包括第一探针(202)、第二探针(201)、桥寡核苷酸(200)和条形码环状寡核苷酸(217)。此处,探针复合体包含第一探针与条形码环状寡核苷酸(210和213)之间、第二探针与条形码环状寡核苷酸(207和215)之间以及第一探针与第二探针(203和204)之间的间隙。通过引入聚合酶和一种或多种核苷酸来填补这些间隙。在此过程中,可使用Stoffel片段、Taq聚合酶或Phusion聚合酶和DNA连接酶(诸如Ampligase)的混合物。聚合酶填补这些间隙,随后DNA连接酶的作用使探针、桥和条形码环状寡核苷酸连接成环形复合体。

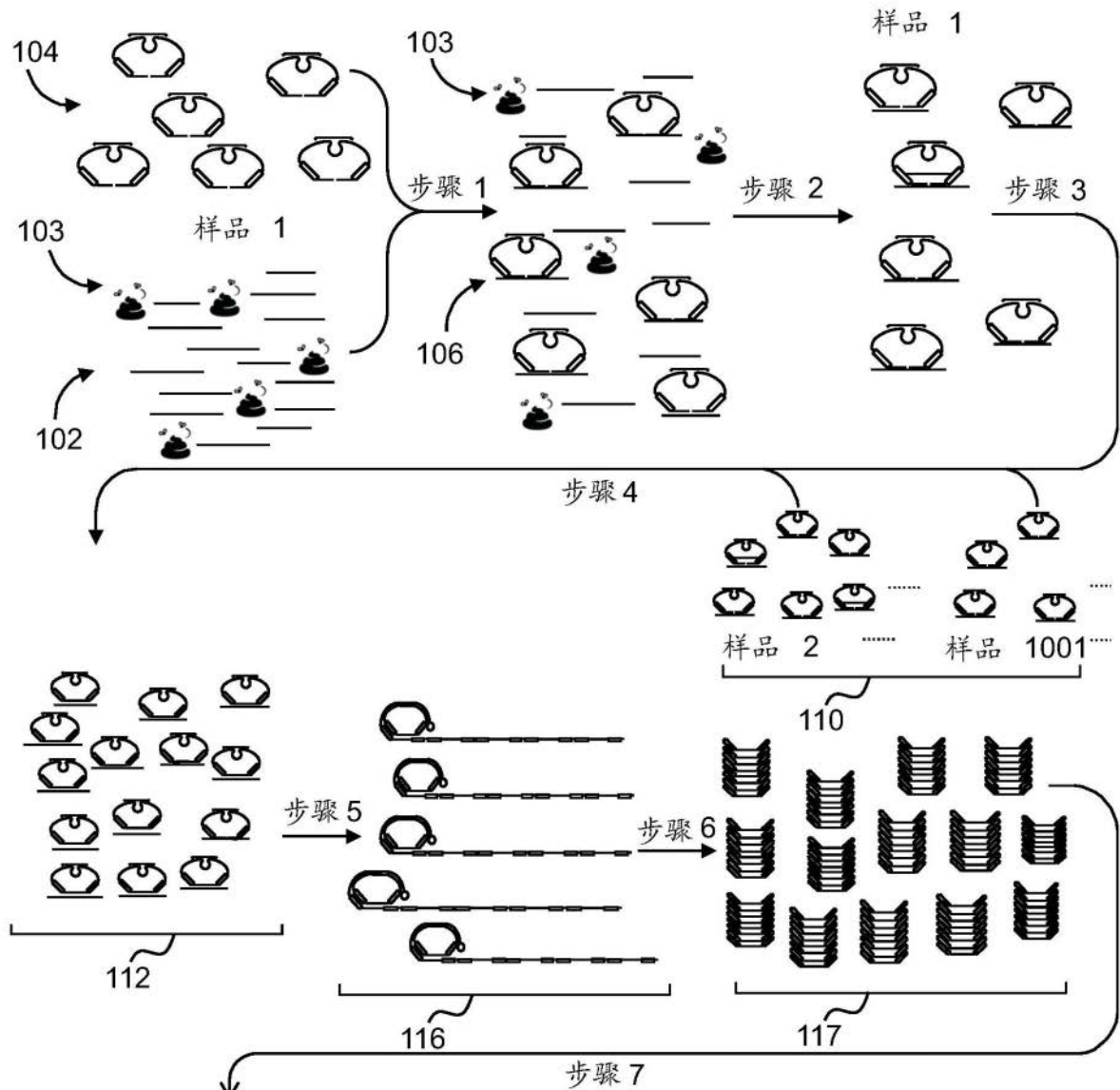
[0166] 第一探针的15-25个碱基包括桥结合序列(210),该序列可选地包括经化学修饰的

碱基,用于有效桥接寡核苷酸结合。第一探针进一步包括用于扩增引物的结合位点(221)和条形码序列(222),以及自5'端起15-30个碱基与遗传靶标结合的序列(203)。210的一些或全部核苷酸可包括化学修饰,以增加探针对靶标或桥(209)的亲合力。第一探针的最后一个碱基可选地包括用于酶连接的磷酸部或允许化学连接到相邻探针的5'端的修饰(205)。

[0167] 第二探针的第一个碱基可选地包括用于酶连接的磷酸部或允许化学连接到相邻探针的5'端的修饰(206)。第二探针的5'端的15-30个碱基包括第二探针的与基因靶标结合的部分(204)。第二探针进一步包括扩增引物的结合位点(223)、测序接头序列(224)、限制性核酸内切酶(诸如EcoRI)或归巢核酸内切酶的识别位点(225)和另一个测序接头序列(207)。第二探针的最后15-25个碱基(207)与桥寡核苷酸(208)反向互补。203、204、207或210的一些或全部核苷酸可包括化学修饰,以增加探针对靶标或桥寡核苷酸的亲合力。

[0168] 桥寡核苷酸的5'端的前15-25个碱基(209)与第一探针的桥寡核苷酸特异性序列(210)反向互补,并可选地包括经化学修饰的核苷酸以增加结合。桥寡核苷酸的最后15-25个碱基(208)与第二探针的序列(207)反向互补,并可选地包括经化学修饰的核苷酸以增加结合。桥寡核苷酸与条形码环状寡核苷酸(220)两端不反向互补的部分可选地包含限制性核酸内切酶的识别位点,以允许对不含环的结构酶降解。桥寡核苷酸的5'端可选地包括用于捕获连接复合体的捕获部(211)。此外,桥寡核苷酸包括与条形码环状寡核苷酸的序列213和215互补的序列214和216。桥寡核苷酸的3'端可选地包括磷酸(或其他可裂解的)部(212),以防止在间隙填补过程中延伸。

[0169] 条形码环状寡核苷酸的5'端的前15-25个碱基(215)与桥寡核苷酸序列216反向互补。条形码环状寡核苷酸包括含条形码(218)的环状区域。条形码环状寡核苷酸的最后15-25个碱基(213)与桥寡核苷酸序列214反向互补。



```

GGGGCCCGCCGTCGATCGGAGCCGTTAGGAT
TTAAGGTGCCGTCGATCGGAGCCGACGTACG
TTAAGGTGCCGTCGATCGGAGCCGACGTACG
TATAATAGAGGTCGTGCAGTCACGACCCGGT
ACGAGCTTGAGTCGATCCGACGTCACCCGGT
GGGCGGGAGGTCGTGCAGTCACGTTAGGAT
TCCAGGTGAGTCGATCCGTCACGTACGTACG
AAAAATTTAGCGTACGTCGTACGTTTAGGAT
TATAATAGAGGTCGTGCAGTCACGACCCGGT
AGGACCTTGAGTCGATCCGACGTCACCCGGT
AGCGACCGAGGTCGTGCAGTCACGACGTACG
TATAATAGAGGTCGTGCAGTCACGACCCGGT
GGAAAAAGCCGTCGATCGGAGCCGTTAGGAT
ATATACAGAGGTCGTGCAGTCACGTTAGGAT
GAGAGCCGAGGTCGTGCAGTCACGACCCGGT
CGCACCGAGGTCGTGCAGTCACGACGTACG
ATTACAAGCCGTCGATCGGAGCCGACCCGGT
TATAATAGAGGTCGTGCAGTCACGACCCGGT
GGGCAATTAGCGTACGTCGTACGTACGTACG
TATGCGAGCCGTCGATCGGAGCCGACGTACG
AGGACCTTGAGTCGATCCGACGTCACCCGGT
ATTACAAGCCGTCGATCGGAGCCGACCCGGT
GAAGAATTAGCGTACGTCGTACGTTTAGGAT
GGGCAATTAGCGTACGTCGTACGTACGTACG
GGGCAATTAGCGTACGTCGTACGTACGTACG
GAGCACTTAGCGTACGTCGTACGTACCCGGT
AAAGGGGCGAGTCGATCCGACGTTTAGGAT
AGCGCGCCGTCGATCGGAGCCGTTAGGAT

```

最终结果:

对多个基因靶标和样品进行绝对、准确的靶标定量

	靶标 1	靶标 2	...	靶标 50
样品 1	1007 拷贝	58 拷贝	...	120 拷贝
样品 2	300 拷贝	550 拷贝	...	10 拷贝
...
样品 1001	225 拷贝	37 拷贝	...	1003 拷贝
...

图1

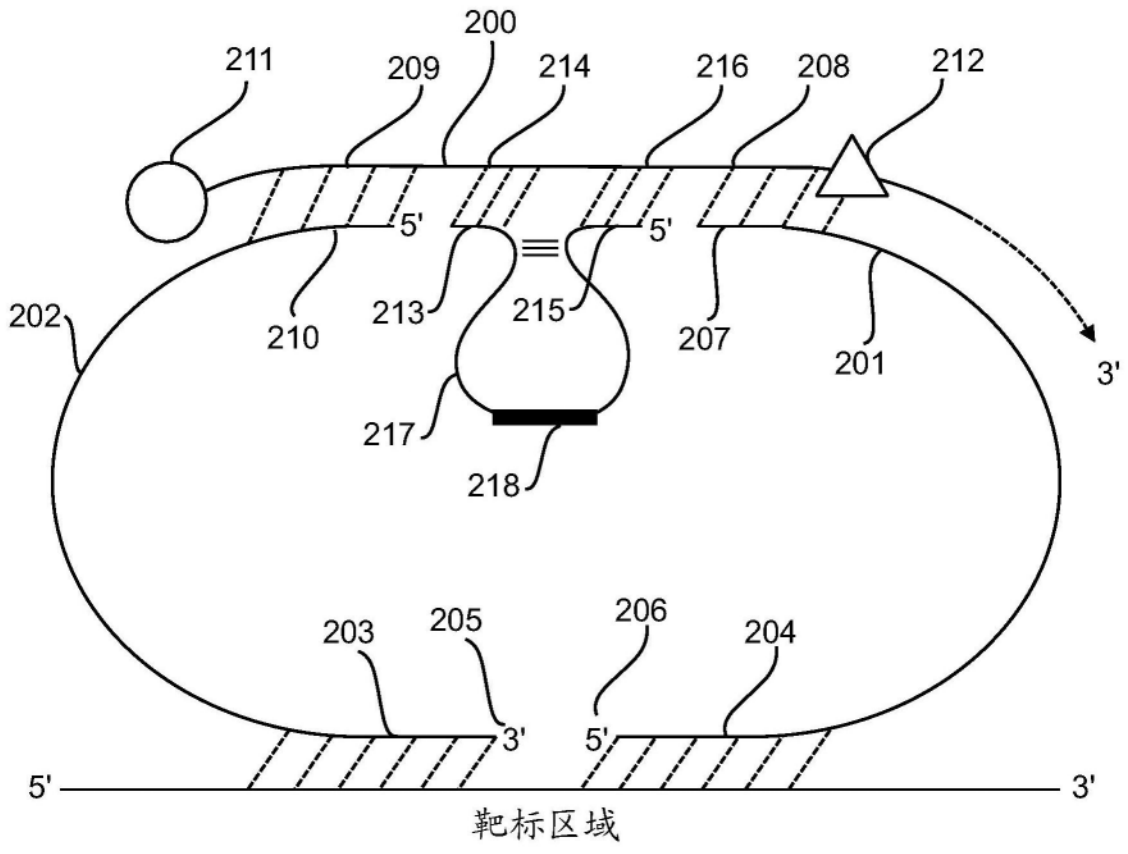


图2A

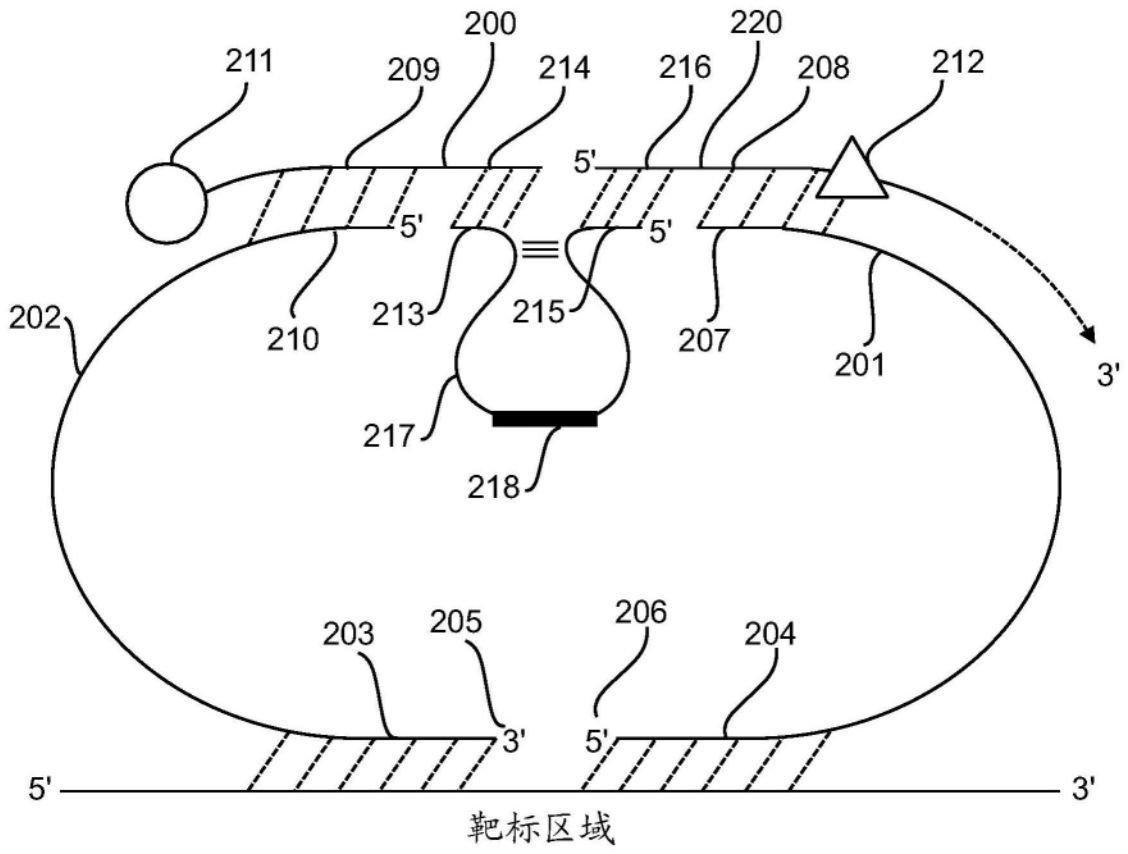


图2B

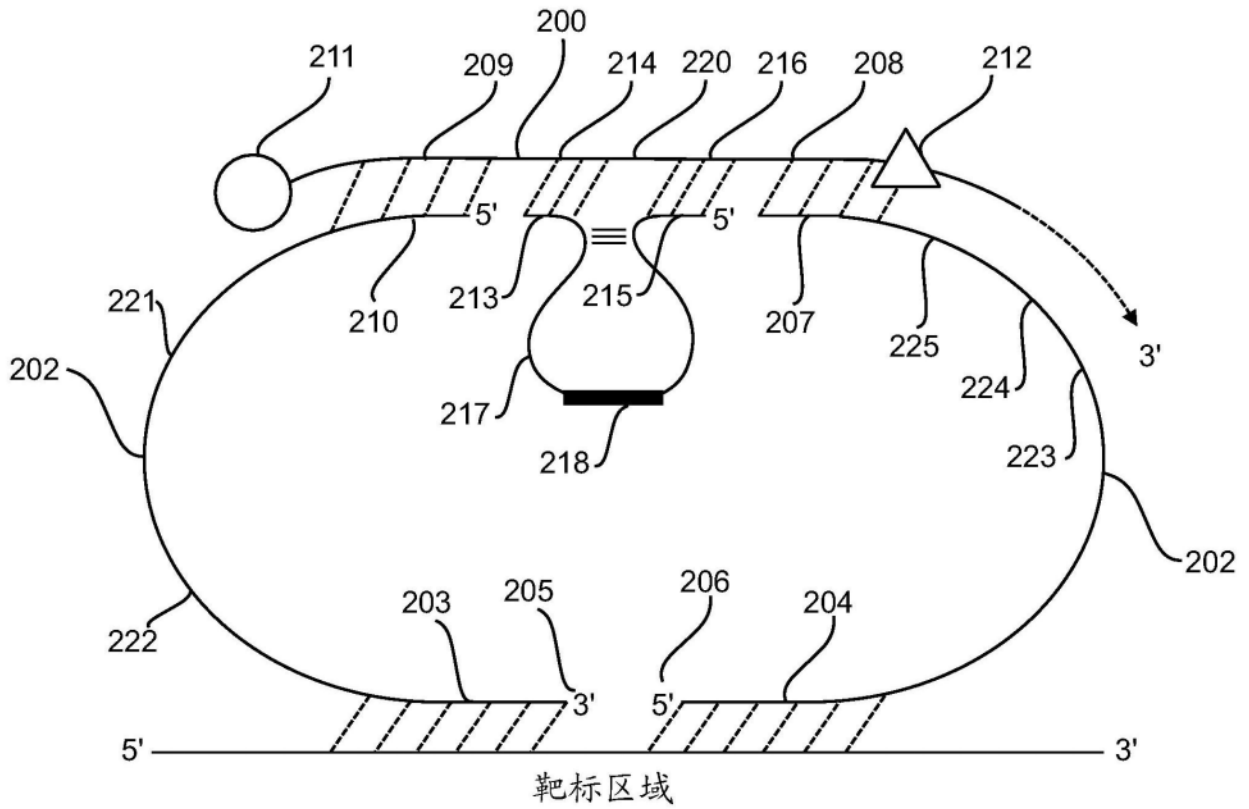


图2D

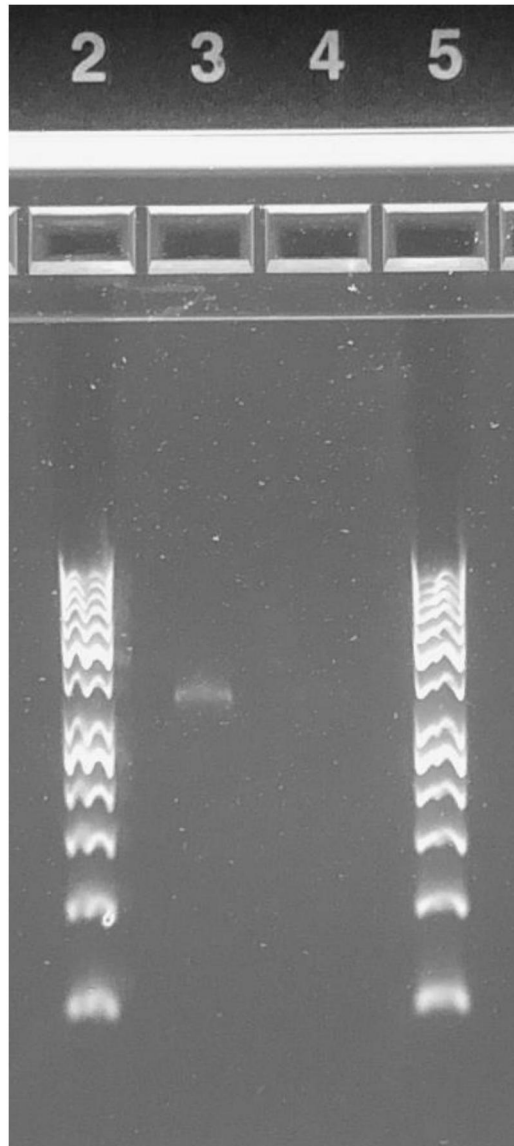


图3

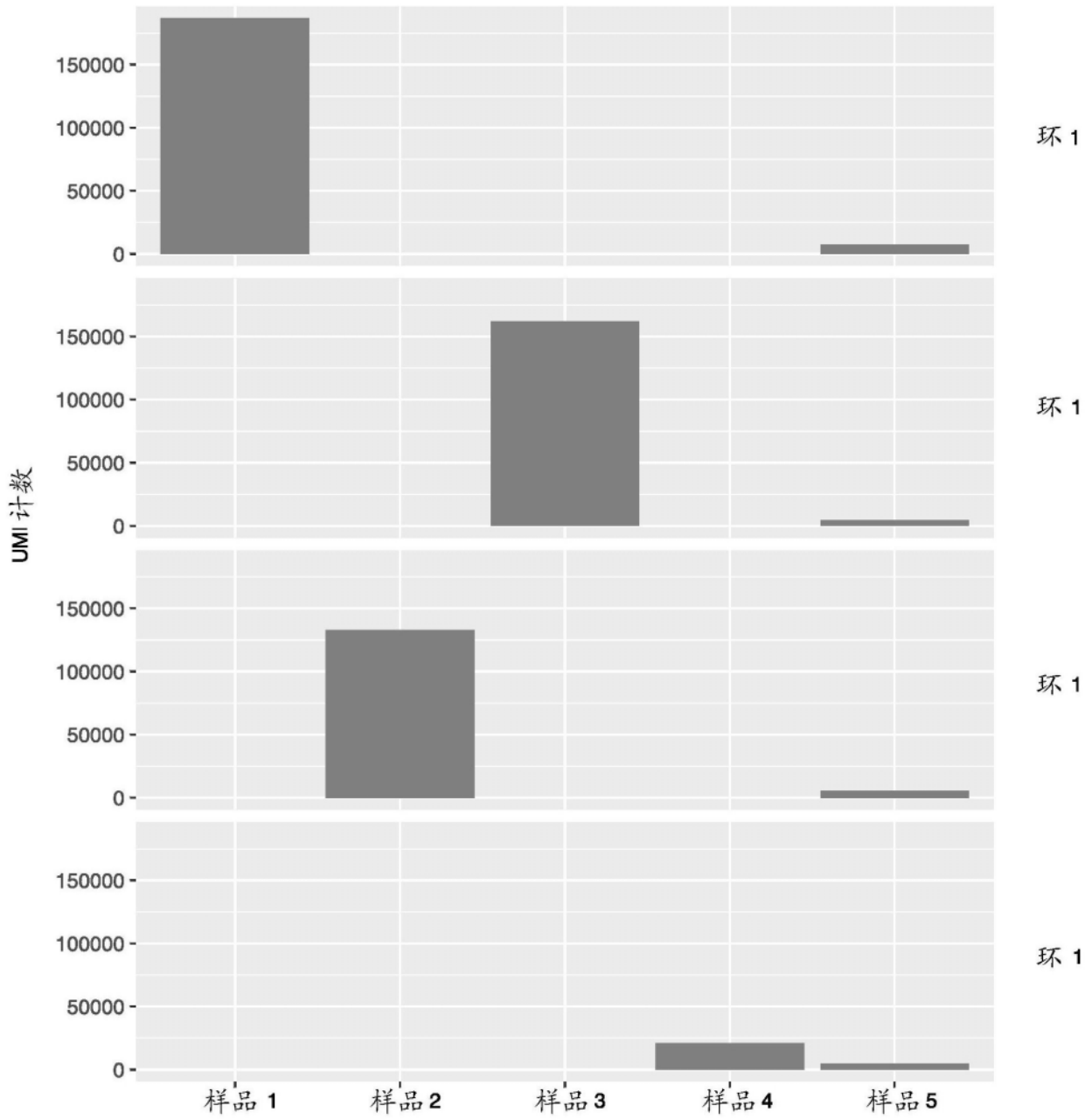


图4