



(12) 发明专利

(10) 授权公告号 CN 111032881 B

(45) 授权公告日 2023. 12. 26

(21) 申请号 201880055271.3

汉娜·希夫

(22) 申请日 2018.08.23

(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

(65) 同一申请的已公布的文献号
申请公布号 CN 111032881 A

专利代理师 张艳 马军芳

(43) 申请公布日 2020.04.17

(51) Int.Cl.

(30) 优先权数据

C12Q 1/6806 (2006.01)

17188047.9 2017.08.25 EP

C12Q 1/6855 (2006.01)

(85) PCT国际申请进入国家阶段日
2020.02.25

C12Q 1/6865 (2006.01)

C12Q 1/6869 (2006.01)

(86) PCT国际申请的申请数据
PCT/EP2018/072754 2018.08.23

(56) 对比文件

US 2005214825 A1, 2005.09.29

CN 101395280 A, 2009.03.25

(87) PCT国际申请的公布数据
W02019/038372 EN 2019.02.28

Michael S Akhras等.PathogenMip assay:
a multiplex pathogen detection assay.PloS
One.2007,第2卷(第2期),e233.

(73) 专利权人 瑞士联邦水科学与技术研究所
地址 瑞士杜本多夫

审查员 马璐

(72) 发明人 马努·塔米宁 蒂莫西·朱利安
珍妮·斯帕克 莉亚·卡杜夫

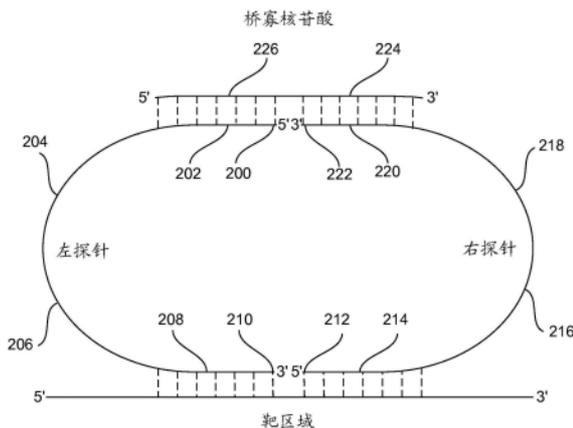
权利要求书3页 说明书14页
序列表6页 附图7页

(54) 发明名称

核酸的精确和大规模平行定量

(57) 摘要

本发明公开内容涉及下一代DNA测序方法,以及用于一种或更多种核酸靶标的精确和大规模平行定量的用途。更具体地,本发明涉及包含探针的方法和试剂盒,所述探针用于检测和定量复杂DNA库中的基因靶标,所述复杂DNA库主要用于所述人类和动物群体以及环境样品中的基因靶标和变体检测。此外,本发明在检测从人体(包括但不限于活组织检查样品、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)等)获得的样品中致病基因改变的领域中找到了特别的应用。本发明包括每个基因靶标的一种或更多种靶标特异性核酸探针(左探针和右探针)和桥寡核苷酸。



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、第二探针和桥寡核苷酸,

其中,所述第一探针从分子的5'末端开始包括第一桥寡核苷酸特异性序列和在所述第一探针的3'末端的第一靶标特异性部分;

并且其中,所述第二探针从分子的5'末端开始包括第二靶标特异性部分和在所述第二探针的3'末端的第二桥寡核苷酸特异性序列;

并且其中,所述桥寡核苷酸含有分别与所述第一探针和所述第二探针中的所述第一桥寡核苷酸特异性序列和所述第二桥寡核苷酸特异性序列互补的序列;

其中,存在第一序列条形码、第二序列条形码和第三条形码中的至少一个;在所述第一序列条形码存在的情况下,所述第一序列条形码位于所述第一探针中的所述第一桥寡核苷酸特异性序列与所述第一靶标特异性部分之间;在所述第二序列条形码存在的情况下,所述第二序列条形码位于所述第二探针中的所述第二靶标特异性部分与第二桥寡核苷酸特异性序列之间;在所述第三条形码存在的情况下,所述第三条形码位于所述桥寡核苷酸中;并且

其中,T7 RNA聚合酶的启动子序列存在于所述第一探针、所述第二探针或所述桥寡核苷酸中,

(ii) 对于所述一种或多种靶标核苷酸序列的每一种,将所述第一探针和第二探针与所述桥寡核苷酸接触并使得能够自退火成多个连接复合体;

(iii) 将存在于待测试所述靶标核苷酸序列的所述多个样品中的核酸与所述多个连接复合体接触;

(iv) 使得所述第一探针和第二探针各自的所述第一靶标特异性部分和第二靶标特异性部分能够与所述靶标核苷酸序列上基本上相邻的区段杂交,从而形成杂交复合体;

(v) 将所述杂交复合体中的所述第一探针和所述第二探针连接,以提供经连接的连接复合体;

(vi) 汇集来自所述多个样品的所述经连接的连接复合体;

(viii) 使用T7 RNA聚合酶从一个或多个经连接的连接复合体扩增RNA,所述T7 RNA聚合酶从嵌在所述经连接的连接复合体中的T7 RNA聚合酶启动子开始RNA合成;

(xii) 对RNA分子进行高通量测序,以确定识别序列;以及

(xiii) 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/或所述第一序列条形码和/或第二序列条形码的至少一部分,和/或所述第三条形码的至少一部分来识别所述多个样品中所述靶标核苷酸序列的存在和/或数量,

其中,所述方法涉及非诊断目的。

2. 根据权利要求1所述的方法,其中,进行步骤(ii)时,每个样品在单独的管中。

3. 根据权利要求1所述的方法,其中,所述第一探针、所述第二探针,或所述桥寡核苷酸包括脱氧尿苷部分,所述脱氧尿苷部分允许通过使用尿嘧啶特异性切除试剂的切割来线性化。

4. 根据权利要求1所述的方法,其中,所述测序通过下一代RNA测序的方式来进行。

5. 根据权利要求1所述的方法,其中,所述第一探针的3'末端或所述第二探针的5'末端,或者两者都被修饰,以允许所述第一探针与所述第二探针的化学连接。

6. 根据权利要求1所述的方法,其中,所述第一探针或所述第二探针,或者两者的桥接部分包括经化学修饰的碱基,以允许改善与所述桥寡核苷酸的结合。

7. 根据权利要求1所述的方法,其中,所述第一靶标特异性部分、所述第二靶标特异性部分、所述第一桥寡核苷酸特异性序列,和/或所述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。

8. 根据权利要求1所述的方法,其中,通过计算每种靶标和每个样品的分子条形码的数量来允许基因靶标计数。

9. 根据权利要求1所述的方法,其中,所述第一探针进一步包括第一通用序列,在所述第一序列条形码存在的情况下,所述第一通用序列位于所述第一桥寡核苷酸特异性序列与所述第一序列条形码之间;在所述第一序列条形码不存在的情况下,所述第一通用序列位于所述第一桥寡核苷酸特异性序列与所述第一靶标特异性部分之间,

其中,所述第二探针进一步包括第二通用序列,在所述第二序列条形码存在的情况下,所述第二通用序列位于所述第二序列条形码与所述第二桥寡核苷酸特异性序列之间;在所述第二序列条形码不存在的情况下,所述第二通用序列位于所述第二靶标特异性部分与所述第二桥寡核苷酸特异性序列之间,并且

其中,用通用的第一引物和通用的第二引物扩增所述RNA分子,以提供扩增产物,所述通用的第一引物和通用的第二引物与所述经连接的连接复合体中存在的所述第一通用序列和第二通用序列互补。

10. 根据权利要求1所述的方法,其中,对于两个以上样品或者两个以上基因座/等位基因组合,条形码序列用于对所述样品进行一种或多种序列和/或多态性的基因型分析。

11. 根据权利要求10所述的方法,所述多态性是SNP和/或插入缺失。

12. 含多个容器的成套的试剂盒,其中,至少一个容器包括一组或多组的第一探针和第二探针,并且至少一个容器包括一种或多种桥寡核苷酸,

其中,所述第一探针从分子的5'末端开始包括第一桥寡核苷酸特异性序列和在所述第一探针的3'末端的第一靶标特异性部分;

其中,所述第二探针从分子的5'末端开始包括第二靶标特异性部分和在第二探针的3'末端的第二桥寡核苷酸特异性序列;

其中,所述桥寡核苷酸包括分别与所述第一探针和所述第二探针中的所述第一桥寡核苷酸特异性序列和所述第二桥寡核苷酸特异性序列互补的序列;

并且其中,存在第一序列条形码、第二序列条形码和第三条形码中的至少一个;在所述第一序列条形码存在的情况下,所述第一序列条形码位于所述第一探针中的所述第一桥寡核苷酸特异性序列与所述第一靶标特异性部分之间;在所述第二序列条形码存在的情况下,所述第二序列条形码位于所述第二探针中的所述第二靶标特异性部分与第二桥寡核苷酸特异性序列之间;在所述第三条形码存在的情况下,所述第三条形码位于所述桥寡核苷酸中;

并且其中,T7 RNA聚合酶的启动子序列存在于所述第一探针、所述第二探针或所述桥寡核苷酸中。

13. 根据权利要求12所述的成套的试剂盒,其中,所述第一探针的3'末端或所述第二探针的5'末端,或者两者都被修饰,以允许所述第一探针与所述第二探针的化学连接。

14. 根据权利要求13所述的成套的试剂盒,其中,所述桥寡核苷酸在与所述第一探针的序列互补的序列中或在与所述第二探针的序列互补的序列中或在两者中包括一个或多个经化学修饰的核苷酸。

15. 根据权利要求13或14所述的成套的试剂盒,其中,所述第一靶标特异性部分、所述第二靶标特异性部分、所述第一桥寡核苷酸特异性序列,和/或所述第二桥寡核苷酸特异性序列彼此独立地含有一个或多个经化学修饰的核苷酸。

16. 第一探针、第二探针和桥寡核苷酸在多个样品中确定至少一种靶标序列是否存在和/或对至少一种靶标序列进行定量的高通量测序中的用途,

其中,所述第一探针从分子的5'末端开始包括第一桥寡核苷酸特异性序列和在所述第一探针的3'末端的第一靶标特异性部分;

并且其中,所述第二探针从分子的5'末端开始包括第二靶标特异性部分和在所述第二探针的3'末端的第二桥寡核苷酸特异性序列;

并且其中,所述桥寡核苷酸包括分别与所述第一探针和所述第二探针中的所述第一桥寡核苷酸特异性序列和所述第二桥寡核苷酸特异性序列互补的序列;

其中,存在第一序列条形码、第二序列条形码和第三条形码中至少一个,所述第一序列条形码位于所述第一探针中的所述第一桥寡核苷酸特异性序列与所述第一靶标特异性部分之间;在所述第二序列条形码存在的情况下,所述第二序列条形码位于所述第二探针中的所述第二靶标特异性部分与第二桥寡核苷酸特异性序列之间;在所述第三条形码存在的情况下,所述第三条形码位于所述桥寡核苷酸中;并且

其中,T7 RNA聚合酶的启动子序列存在于所述第一探针、所述第二探针或所述桥寡核苷酸中;并且

其中,当所述第一探针和第二探针各自的靶标特异性区段与所述靶标序列上基本相邻的区段杂交时,所述第一探针与所述第二探针连接,以提供经连接的探针;汇集并使用T7 RNA聚合酶扩增来自所述多个样品的所述经连接的探针;并且对所述扩增的产物进行高通量测序以确定识别序列;并且

其中,所述用途涉及非诊断目的。

核酸的精确和大规模平行定量

技术领域

[0001] 本发明公开内容涉及下一代DNA测序方法,以及用于一种或更多种核酸靶标的精确和大规模平行定量的用途。更具体地,本发明涉及含探针的方法和试剂盒,所述探针用于检测和定量复杂DNA库中的基因靶标,所述复杂DNA库主要用于人类和动物群体以及环境样品中的基因靶标和变体检测。此外,本发明在检测从人体获得的样品(包括但不限于活组织检查样品、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)等)中引起疾病的基因变化的领域中发现了特别的应用。本发明包括每种基因靶标的一种或更多种靶标特异性核酸探针(左探针和右探针)和桥寡核苷酸。

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同样的变异并不麻烦。然而,需要在技术上取得进展,以便对基因改变(诸如突变和其他基因变异)的核苷酸序列进行非常有价值的研究,从而为促进预防和治疗措施提供必要的信息。核苷酸测序是确定感兴趣的基因靶标中核苷酸顺序的过程。感兴趣的基因靶标的核苷酸序列中突变的检测是使用核苷酸测序方法进行的。最常见的测序方法,诸如但不限于用于核苷酸序列的检测的化学测序法和一代测序法。核酸测序技术在诸如分子生物学、进化生物学、宏基因组学、医学和法医学等众多领域发挥着至关重要的作用。理想情况下,检测和定量基因靶标,例如,定量母体血液中的DNA是无创产前三染色体性筛查的基础;定量循环DNA中的罕见突变是癌症液体活组织检查的基础;定量抗生素抗性基因是抗生素抗性基因传播风险评估的基础。

[0003] 通常,癌症诊断涉及对循环DNA中与肿瘤相关的突变进行检测和定量。类似地,胎儿异常涉及对母体血液中少量存在的胎儿DNA进行检测和定量。尽管降低了测序成本,但是检测和精确定量这样的微弱信号目前仍然是费力和昂贵的。可以更准确地表述各种问题,诸如在一致背景下检测基因信号的特异性、为了检测弱基因信号的灵敏度、为了经检测信号准确定量的准确性、每次试验的经靶向基因靶标的通量数量、每次试验的成本、缩放以在当平行检测多个样品时确定试验成本规模以及确定从采样到结果的时间长度的周转(turn-over)。

[0004] 目前,液体活组织检查和概念上类似的检测(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:

[0005] 定量PCR:定量PCR(qPCR),是一种基于分子生物学的技术,其包括在PCR过程中(即实时)扩增经靶向DNA分子。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)。

[0006] 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,考虑到建立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,因此随着

样品数量的增加,实际操作时间会急剧增加。估计每个样品的每种定量靶标至少有40个反应,排除人工成本将单价设定为大约80\$,为多少反应可以定量设定了明确的限制。

[0007] 阵列PCR:PCR阵列是分析相关、途径或疾病集中基因组表达的最可靠工具。每个96孔板、384孔板或100孔板PCR阵列都包括SYBR绿色经优化引物试验,用于对一组集中基因进行彻底研究。qPCR技术的一个新的迭代是阵列qPCR,它使单个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,使用上述设置分析数千个样品仍然非常昂贵。

[0008] 数字PCR:数字聚合酶链式反应(Digital PCR, DigitalPCR, dPCR, dePCR)是一种通过微滴流体和荧光检测提供靶标绝对定量的方法。这种方法相对经济有效(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不足以覆盖数千个样品。

[0009] 多重连接依赖探针扩增(MLPA)

[0010] MLPA提供了一种方法来简化单个样品中多个基因靶标的检测。然而,MLPA只提供靶标的相对定量,并要求对每个样品进行单独的检测实验。最近,MLPA的一个变体引入了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。

[0011] 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NIPT筛查。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,位基因座的非靶向取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。

[0012] 因此,根据前述讨论,需要通过精确和大规模平行定量核酸靶标来克服前述缺点,诸如但不限于特异性、灵敏度、准确性、通量、成本、规模和周转。

发明内容

[0013] 本发明寻求提供一种用于确定多重连接试验(MLA)的方法和组合物。多重连接试验是一种基于分子连接的试验,旨在将条形码化的MLPA的简单性和平行性潜力与qPCR的准确性和灵敏度相结合。

[0014] 本发明更具体地涉及一种通过利用连接依赖性试验在大量样品中高通量检测靶标核苷酸序列的方法。在一个实施方案中,靶标核苷酸序列可以是DNA或RNA,优选DNA。在一个实施方案中,所述方法和组合物包括在多个样品中确定至少一种靶标序列是否存在和/或对至少一种靶标序列进行定量的高通量测序分析靶标的步骤。

[0015] 在一个方面,本发明的实施方案提供了该方法,该方法进一步包括针对每个样品中的每种靶标核苷酸序列的一组靶标特异性探针。靶标特异性探针组件的设计包括每种基

因靶标两个靶标特异性核酸探针(左探针和右探针),或者可替代地用三个靶标特异性核酸探针设计靶标特异性探针组件,其中两个对基因靶标特异(左探针和右探针),一个通用(桥寡核苷酸)。通用(桥寡核苷酸)用于连接左探针和右探针。

[0016] 在一个实施方案中,第一探针从分子的5'末端开始包括可选的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,和在第一探针3'末端的第一靶标特异性部分。第二探针从分子的5'末端开始包括可选的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,和在第二探针3'末端的第二桥寡核苷酸特异性序列。优选地,第一序列条形码或第二序列条形码中的至少一个分别存在于第一探针或第二探针中,其中,第一序列条形码或第二序列条形码,或两者可以是随机序列,或者可以含有靶标核苷酸序列识别序列、样品识别序列和/或用于靶标计数的分子条形码。

[0017] 优选地,通用桥寡核苷酸探针含有T7 RNA聚合酶的启动子序列、分别与第一探针和第二探针中的第一桥寡核苷酸特异性序列和第二桥寡核苷酸特异性序列互补的序列,以及可选的第三条形码,其可以是随机序列或可以含有样品识别序列或靶标识别序列。优选地,T7RNA聚合酶的启动子序列存在于桥寡核苷酸中,然而,启动子也能存在于第一探针或第二探针中,而不是启动子存在于桥寡核苷酸中。然而,在这种情况下,探针和寡核苷酸的设计必须使得T7 RNA聚合酶能够转录识别样品和靶标以及计数靶标序列所需的所有序列。

[0018] 第一探针和第二探针包括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。第一探针和第二探针包含彼此独立的序列。第一探针和第二探针优选用一个或更多个经化学修饰的核苷酸序列进行化学修饰。

[0019] 第一探针和第二探针与桥寡核苷酸接触(优选对于每个样品在单独的管中)。多个第一探针和第二探针与多个桥寡核苷酸自退火成多个连接复合体。优选地,每个连接复合体对于第一靶标特异性序列、第二靶标特异性序列和一个或更多个条形码序列的组合是唯一的。这使得能够在扩增和结果分析之后计数靶标序列。

[0020] 随后,使每个样品中的一种或更多种靶标核苷酸序列与连接复合体接触,使得能够自退火成多个杂交的连接复合体。相应的第一探针和第二探针的第一靶标特异性部分和第二靶标特异性部分与靶标序列上基本上相邻的区段杂交,从而形成杂交复合体。形成的杂交复合体中的第一探针和第二探针的连接是酶促地或化学地进行的,以提供经连接的连接复合体。然后将一个或更多个样品中的经连接的连接复合体汇集。

[0021] 可选地,如果第一探针和第二探针之间存在间隙,通过引入聚合酶和一个或更多个核苷酸来填充。聚合酶添加(a)互补于通用桥寡核苷酸序列的核苷酸和(b)互补于靶标序列的核苷酸,从而填充第一探针和第二探针之间的两个间隙,使得左探针和右探针连接。桥寡核苷酸从与经连接探针互补的5'位点或3'位点延伸,使得靶标序列识别序列整合到桥寡核苷酸中,从而形成一个或更多个经连接的连接复合体。

[0022] 为了防止聚合酶在第一探针和第二探针在靶标序列部分处连接之前将桥寡核苷酸延伸到第一探针和第二探针上的相应靶标序列中,优选使用合适的聚合酶,更优选不延伸到双链DNA序列中的聚合酶。这种聚合酶的典型例子是Taq聚合酶。

[0023] 接下来,T7 RNA聚合酶结合在经连接的连接复合体的桥寡核苷酸序列上的双链T7 RNA聚合酶启动子位点上,并转录下游的RNA。从一个或更多个经连接的连接复合体中扩增

经合成的RNA是通过例如使用乳液逆转录酶聚合酶(RT)-PCR或T7 RNA聚合酶进行的。游离样品和探针核酸的去除是通过可选地在T7 RNA聚合酶合成RNA之后但在合成cDNA或RT-PCR之前将DNA特异性核酸外切酶和核酸内切酶添加到混合扩增反应中来进行的。

[0024] 优选地,通过使用纯化柱或珠纯化去除连接反应混合物的成分来进行经连接的片段的纯化过程。可选地,使用与通用位点2反向互补的DNA寡核苷酸分子从RNA分子制备cDNA。在制备cDNA后,可以使用常规的或优选的乳液PCR来增加信号。或者,可以使用RT-PCR或乳液RT-PCR。

[0025] 通过对经扩增的RNA或cDNA分子进行高通量测序技术以确定识别序列,从而确定靶标核苷酸序列是否存在。

[0026] 另一方面,本发明的实施方案提供了一种用于本发明方法和组合物的试剂盒。该试剂盒包括多个容器,其中,至少一个容器包括一组第一探针和第二探针,并且其中,至少一个容器包括桥寡核苷酸。在一个特定的实施方案中,含一组第一探针和第二探针的至少一个容器和含桥寡聚物的至少一个容器是同一个容器。在这种情况下,三个探针可以被预退火并已形成经连接的复合体。

[0027] 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),检测和扩增感兴趣的靶标序列。探针被设计成具有经改进的结合性质,并引起更高的分析特异性、灵敏度和准确性。本发明可应用于分子生物学、进化生物学、宏基因组学、基因分型领域,并且更具体地,但不限于癌症诊断和胎儿染色体疾病。

[0028] 从附图和结合所附权利要求解释的说明性实施方案的详细描述中,本发明的其他方面、优点、特征和目的将变得显而易见。

[0029] 应当理解,在不脱离由所附权利要求限定的本发明的范围的情况下,本发明的特征易于以各种组合进行结合。

附图说明

[0030] 当结合附图阅读时,可以更好地理解以上概述以及以下说明性实施方案的详细描述。为了说明本发明,在附图中示出了本发明的示例性结构。然而,本发明不限于这里公开的具体方法和手段。此外,本领域技术人员将理解附图不是按比例绘制的。只要有可能,相同的元素用相同的数字表示。

[0031] 现在将参考以下附图仅以示例的方式描述本发明的实施方案,其中:

[0032] 图1示出了根据本文实施方案的多重连接试验(MLA)的流程图;

[0033] 图2A示出了根据本文实施方案的具有多个探针实体的探针三联体的原理结构;

[0034] 图2B示出了根据本文实施方案的第一探针和第二探针之间的间隙填充;

[0035] 图3显示了增加探针结合效率和提高连接效率的可能的化学修饰;

[0036] 图4提供了根据本文实施方案的经复制染色体计数数据分析的概况;

[0037] 图5提供了根据本文实施方案的用于定量地识别多个基因突变变体比例的概况;

和

[0038] 图6示出了根据本文实施方案的多重连接依赖性探针扩增的流程图。

具体实施方式

[0039] 以下详细描述示出了本发明的实施方案及其实现方式。尽管已经公开了实现本发明的一些模式,但是本领域技术人员将认识到,用于实现或实践本发明的其他实施方案也是可能的。

[0040] 在第一实施方案中,本发明提供了一种高通量检测多个样品中一种或更多种靶标核苷酸序列的方法,该方法包括以下步骤:

[0041] (i) 为每个样品中的每种靶标核苷酸序列提供:第一探针、第二探针和桥寡核苷酸,其中

[0042] 第一探针从分子的5'末端开始包括可选的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,和在第一探针3'末端的第一靶标特异性部分;并且其中

[0043] 第二探针从分子的5'末端开始包括可选的5'磷酸、第二靶标特异性序列、可选的第二序列条形码、可选的第二通用序列,和在第二探针3'末端的第二桥寡核苷酸特异性序列;并且其中

[0044] 桥寡核苷酸包括分别与第一探针和第二探针中的第一桥寡核苷酸特异性序列和第二桥寡核苷酸特异性序列互补的序列,以及可选的第三条形码,第三条形码可以是随机序列或可以含有样品识别序列或靶标识别序列;并且其中

[0045] 第一序列条形码或第二序列条形码或第三条形码中的至少一个分别存在于第一探针或第二探针或桥寡核苷酸中,其中,第一序列条形码或第二序列条形码,或两者可以是随机序列,或者可以含有靶标核苷酸序列识别序列、样品识别序列和/用于靶标计数的或分子条形码;并且其中

[0046] T7 RNA聚合酶的启动子序列存在于第一探针、第二探针或桥寡核苷酸中,并且其中,优选地,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,和/或第二桥寡核苷酸特异性序列彼此独立地包括一个或更多个经化学修饰的核苷酸;

[0047] (ii) 对于一种或更多种靶标核苷酸序列中的每一个,将第一探针与第二探针与桥寡核苷酸接触(优选每个样品在单独的管中),并使得能够自退火成多个连接复合体;

[0048] (iii) 将存在于待测试靶标核苷酸序列的多个样品中的核酸与多个连接复合体接触;

[0049] (iv) 使第一探针和第二探针各自的第一靶标特异性部分和第二靶标特异性部分能够与靶标序列上基本上相邻的区段杂交,从而形成杂交复合体;

[0050] (v) 将杂交复合体中的第一探针和第二探针连接,以提供经连接的连接复合体;

[0051] (vi) 汇集来自多个样品的经连接的连接复合体;

[0052] (vii) 可选地,向杂交复合体提供聚合酶和一种或更多个核苷酸,使得桥寡核苷酸上的第一探针和第二探针之间存在的间隙被填充,使得经连接的探针从5'位点或3'位点延伸,使得桥寡核苷酸中存在的条形码序列整合到经连接的探针中,或者使得桥寡核苷酸从5'位点或3'位点延伸,使得存在于第一探针或第二探针中的条形码序列整合到桥寡核苷酸中,或者两者皆有,从而形成一个或更多个经连接的连接复合体;

[0053] (viii) 使用T7 RNA聚合酶从一个或更多个经连接的连接复合体扩增的RNA,所述T7RNA聚合酶从嵌在经连接的复合体中的T7 RNA聚合酶启动子开始RNA合成;

[0054] (ix) 可选地,向经汇集的扩增反应中加入DNA特异性核酸外切酶和核酸内切酶,以去除样品和探针DNA;

[0055] (x) 可选地,使用珠子或柱净化来纯化剩余的RNA;

[0056] (xi) 可选地,使用与通用位点2反向互补的DNA寡核苷酸分子从RNA分子制备cDNA;

[0057] (xii) 对经扩增的RNA或cDNA分子进行高通量测序技术,以确定识别序列;以及

[0058] (xiii) 通过确定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、第一条形码序列和/或第二条形码序列的至少一部分,和/或第三条形码序列的至少一部分来识别多个样品中靶标核苷酸序列的存在和/或数量。

[0059] 定义:

[0060] 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,给出的术语指的是连续核苷酸序列以及具有互补序列的核酸分子。靶标序列优选是代表多态性或与多态性相关的核苷酸序列。

[0061] 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位点。多态位点可以小到一个碱基对。

[0062] 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。优选制备根据本发明方法提供的样品,以便至少提取靶标核酸,并使这些核酸可用于本发明中的探针。特别地,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(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))获得的两个或更多个样品,从环境(包括水、废水、土壤等)获得的两个或更多个样品。

[0063] 探针:术语探针是可变长度(通常50-1000个碱基长,优选50-200个碱基长)的DNA或RNA片段,探针可用于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,提供一对左探针和右探针,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,本发明提供了用于接合左探针和右探针的桥寡核苷酸。

[0064] 通用(Universal):当用于描述扩增过程时,术语通用指的是能够使用单个引物或引物组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路复用,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”。

[0065] 杂交:术语杂交(或杂化)描述脱氧核糖核酸(DNA)或核糖核酸(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。

[0066] 连接: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(的末端)之间形成磷酸二酯键的酶。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末端都被修饰以使得能够化学连接。

[0067] 扩增:本文使用的术语扩增指的是使用聚合酶链式反应(PCR)来增加特定核苷酸序列在核苷酸序列混合物中的浓度。“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通过加热样品而变性。术语引物是RNA或DNA的短链(通常约18-22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。T7 RNA聚合酶能够将一个单一的DNA分子转录并扩增成多个RNA拷贝,这些拷贝可以转化回cDNA。

[0068] 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。本文所用的特异性聚合酶,T7 RNA聚合酶,是来自T7噬菌体的RNA聚合酶,它在5'→3'方向催化RNA的形成。T7 RNA聚合酶需要部分双链的DNA模板和Mg²⁺离子作为合成RNA的辅因子。T7 RNA聚合酶能够将单个DNA分子转录并扩增成多个RNA拷贝。

[0069] 高通量:术语高通量表示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。

[0070] 尿嘧啶特异性切除试剂(USER):尿嘧啶特异性切除试剂也称为USER,允许通过切割存在脱氧尿苷核苷酸的地方来线性化环状DNA分子。

[0071] 本发明公开了一种使用下一代测序允许的技术确定复杂核酸库中基因靶标序列的方法。本发明提供了一种方法,通过利用连接依赖性试验,在多个样品中,优选在非常大量的样品中,对多种基因靶标进行分析。本发明提供了一种多重连接依赖性探针扩增的方法,该方法能够查询多个样品中的不同靶标核酸。本发明内容为本领域技术人员提供了一种用于靶标核酸测序的方法和组合物,该方法和组合物对样品中存在的感兴趣的核酸靶标进行精确和大规模平行定量,用于检测基因变体是否存在。此外,在处理测序数据时,独特的序列识别用于基因靶标的识别和样品库中单个样品的绝对定量。

[0072] 在一个优选实施方案中,本发明的方法和组合物涉及使用高通量测序技术识别多个核酸样品。本发明的方法和组合物使得能够对多个样品中的一种或更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。本发明的方法和组合物利用三种核酸探针,其中两种靶标特异性核酸探针(左探针和右探针)对基因靶标特异,而一种核酸探针是通用的(桥寡核苷酸)。左探针和右探针与桥探针杂交,形成连接复合体。在环境条件下,使得样品DNA/RNA上具有靶标识别位点的连接复合体(含有一个或更多条形码序列)能够与待查询样品的互补靶标序列杂交。杂交后,左探针和右探针以化学的或通过DNA连接酶酶促的方式连接,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,将形成多个这种经连接的连接复合体。

[0073] T7 RNA聚合酶利用下一代测序平台(包括但不限于Illumina MiSeq、HiSeq或NextSeq)扩增用于测序的经接合的探针。样品从序列数据中分离(解卷积)出来,序列靶标在DNA测序后用计算机定量。在一个优选实施方案中,提供了根据本发明的方法,其中通过下一代DNA或RNA测序来进行测序。

[0074] 在一个实施方案中,“多个样品”可以指但不限于从人体(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))获得的两个或更多样品,从环境(包括水、废水、土壤等)获得的两个或更多样品。所用的样品可以是使用任何现有提取技术(诸如乙醇沉淀、苯酚-氯仿提取、硅胶柱等)提取的DNA或RNA。

[0075] 在一个实施方案中,该方法和组合物包括在检测和/或定量多个样品中的至少一种靶标序列时使用高通量测序分析靶标的步骤。靶标序列包括需要检测的任何感兴趣的核苷酸序列。本发明的靶标核苷酸序列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是从凋亡/坏死的癌细胞中获得的,或者母体血液中的一部分DNA是从胎儿获得的。进一步地,分析结果用于,例如,评估个体患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。

[0076] 在一个实施方案中,提供了一种用于高通量检测一种或更多种靶标核苷酸序列的方法,所述靶标核苷酸序列可以源自一个或更多个样品。在某些实施方案中,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。如本文所用,术语探针组包括第一探针、第二探针和桥寡核苷酸。

[0077] 在某些实施方案中,第一探针从分子的5'末端开始包括可选的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'末端的第一靶标特异性部分。在某些实施方案中,第二探针从分子的5'末端开始包括可选的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'末端的第二桥寡核苷酸特异性序列。

[0078] 在优选实施方案中,要么第一探针要么第二探针含有第一序列条形码或第二序列条形码中的至少一个。第一序列条形码或第二序列条形码,或两者都可以是随机序列,或者可以含有靶标核苷酸序列识别序列、样品识别序列和/或用于靶标计数的分子条形码。

[0079] 在优选实施方案中,桥寡核苷酸包含T7 RNA聚合酶的启动子序列、分别与第一探针和第二探针中的第一桥寡核苷酸特异性序列和第二桥寡核苷酸特异性序列互补的序列、通用序列,和/或可以包含第三条形码,第三条形码可以是随机序列或可以含有样品识别序列或序列识别序列。在这方面,第三条形码不一定意味着已经存在第一条形码和第二条形码。如前所述,在经连接的连接复合体中应该存在至少一个条形码,这使得能够在所有经测试样品的所有连接复合体中唯一地定义该复合体。

[0080] 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优选彼此独立地含有至少一个经化学修饰的核苷酸。增强探针结合的化学修饰包括但不限于核糖核酸、肽核酸和锁核酸。在某些实施方案中,化学修饰允许相邻探针的化学连接。前述探针与完全相邻的基因座,或者相隔至多50个碱基对,优选相隔至多40个碱基对,更优选相隔至多30个碱基对,更优选相隔至多20个碱基对,更优选相隔至多10个碱基对,最优选相隔至多5个碱基对结合。

[0081] 在探针与含靶标序列的样品接触之前,使第一探针和第二探针与桥寡核苷酸接触(优选每个样品在单独的管中),并使得能够自退火成连接复合体。此后,使多个样品中的一种或更多种靶标核苷酸序列与多个连接复合体接触。相应的第一探针和第二探针的第一靶标特异性部分和第二靶标特异性部分与靶标序列上基本上相邻的区段杂交,从而形成杂交复合体。形成的杂交复合体中的第一探针和第二探针的连接是酶促地或化学地进行的,以提供经连接的连接复合体。然后将经连接的连接复合体从一种或更多种靶标样品中汇集。

[0082] 第一探针和第二探针之间的间隙(如果存在)可以通过引入聚合酶和一个或更多个核苷酸来填充。聚合酶添加(a)互补于通用桥寡核苷酸序列的核苷酸和/或(b)互补于条形码序列的核苷酸,并从而填充第一探针和第二探针之间的两个间隙,使得左探针和右探针连接,并将通用序列和/或第三条形码序列包含到桥互补链中。桥寡核苷酸从与经连接探针互补的5'位点或3'位点延伸,使得第一探针或第二探针中存在的靶标序列识别序列整合

到桥寡核苷酸中。优选地,使用不断裂双链DNA的聚合酶(例如Taq聚合酶),以便当第一探针和第二探针都与靶标序列退火时,不干扰第一探针和第二探针的连接。

[0083] 优选地,探针或桥寡核苷酸中的任一个含有脱氧尿苷部分,该部分允许通过使用尿嘧啶特异性切除试剂切割来线性化。这使得能够使用T7 RNA聚合酶从线性化的经连接的连接复合体转录RNA,该聚合酶从嵌在桥寡核苷酸或第一或第二探针中的T7 RNA聚合酶启动子启动RNA合成。扩增后,游离样品和探针核酸的去除是通过可选地向经汇集的扩增反应中加入DNA特异性外核酸酶和内切核酸酶来进行的。如前所述,T7 RNA聚合酶启动子序列可以位于桥寡核苷酸、第一探针或第二探针中。脱氧尿苷部分也是如此。然而,对于本领域技术人员来说明确的是,T7 RNA聚合酶启动子序列和脱氧尿苷部分应该定位成使得T7 RNA聚合酶能够转录用于计数不同样品中不同靶标所需的所有信息。在T7 RNA聚合酶和脱氧尿苷部分之间应该出现以下序列:至少一个能够识别靶标的序列,至少一个能够识别样品的序列,以及至少一个能够确定样品中靶标序列拷贝数的唯一识别序列。

[0084] 经连接片段的纯化过程可以通过使用纯化柱或珠纯化除去连接反应混合物的成分来进行。可选地,使用与通用位点2反向互补的DNA-寡核苷酸分子,从RNA分子制备cDNA。RNA分子可选地被转化成cDNA,并可选地通过PCR或乳液PCR,使用与探针通用部分结合的引物进行扩增。

[0085] 通过高通量测序技术确定第一和/或第二靶标特异性部分的至少一部分、第一和/或第二条形码的至少一部分和/或第三条形码的至少一部分来识别多个样品中靶标核苷酸序列的存在和/或数量。优选地,通过计数每种靶标和每个样品的分子条形码的数量来允许基因靶标计数。

[0086] 通过对经扩增的RNA或cDNA分子进行高通量测序技术以确定识别序列,从而确定靶标核苷酸序列是否存在。在优选实施方案中,提供了根据本发明的方法,其中,用第一引物和第二引物扩增RNA分子或cDNA分子,以提供扩增产物。优选使用通用的第一引物和通用的第二引物,其与经连接的复合体中存在的第一通用序列和第二通用序列互补。

[0087] 与现有技术中用于核苷酸测序和扩增的方法相比,本发明的方法和组合物涉及各种各样的有利方面。本发明公开的方法和组合物旨在提供优于任何现有靶标,而不限于定量PCR(qPCR)、阵列PCR、数字PCR、多重连接依赖探针扩增等的分析。

[0088] 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分析。本发明的另一个方面是,本发明的方法和组合物使得能够对包括人类和动物群体在内的多个样品中的多个核酸靶标进行精确和大规模平行定量。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修饰核苷酸,可提高退火和结合效率。结合特性的改善引起更高的分析特异性、灵敏度和准确性。本发明的方法和组合物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更多种序列和/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型分析。在一个优选的实施方案中,提供了根据本发明的方法,其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,使用条形码序列对样品进行一个或更多个序列和/或多态性(诸如SNP和/或插入缺失)的基因型分析。

[0089] 在另一方面,本发明提供了包括多个容器的成套的试剂盒,其中,至少一个容器包

括一组或更多组第一探针和第二探针,并且至少一个容器包括一种或更多种桥寡核苷酸,

[0090] 其中,第一探针从分子的5'末端开始包括可选的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,和第一探针3'末端的第一靶标特异性部分;

[0091] 其中,第二探针从分子的5'末端开始包括可选的5'磷酸、可选的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,和第二探针3'末端的第二桥寡核苷酸特异性序列;

[0092] 其中,桥寡核苷酸包含分别与第一探针和第二探针中的第一桥寡核苷酸特异性序列和第二桥寡核苷酸特异性序列互补的序列,以及可选的第三条形码,其可以是随机序列或可以含有靶标核苷酸识别序列或样品识别序列;并且

[0093] 其中,第一序列条形码或第二序列条形码或第三条形码中的至少一个分别存在于第一探针或第二探针或桥寡核苷酸中,其中,第一序列条形码或第二序列条形码或两者可以是随机序列,或者可以含有靶标核苷酸序列识别序列、样品识别序列和/或用于靶标计数的分子条形码;并且其中,T7 RNA聚合酶的启动子序列存在于第一探针、第二探针或桥寡核苷酸中,

[0094] 其中,优选地,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,和/或第二寡核苷酸特异性序列可以彼此独立地包括一个或更多个经化学修饰的核苷酸。

[0095] 优选地,第一探针的3'末端或第二探针的5'末端,或两者都被修饰,以允许第一探针与第二探针的化学连接。

[0096] 优选地,桥寡核苷酸在与第一探针序列互补的序列中或在与第二探针序列互补的序列中或在两者中包含一个或更多个经化学修饰的核苷酸。

[0097] 优选地,第一探针、第二探针或桥寡核苷酸包含脱氧尿苷部分,脱氧尿苷部分允许通过使用尿嘧啶特异性切除试剂切割来线性化。

[0098] 优选地,第一探针的3'末端或第二探针的5'末端,或两者都被修饰,以允许第一探针与第二探针的化学连接。

[0099] 优选地,第一探针或第二探针的桥接部分,或两者包括经化学修饰的碱基,以允许与桥寡核苷酸的结合得到改善。

[0100] 在另一方面,本发明提供了第一探针、第二探针和桥寡核苷酸在高通量测序中用于确定多个样品中一种或更多种靶标序列是否存在和/或对一种或更多种靶标进行定量的用途。

[0101] 第一探针从分子的5'末端开始包括可选的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,和第一探针3'末端的第一靶标特异性部分。

[0102] 第二探针从分子的5'末端开始包括可选的5'磷酸、可选的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,和第二探针3'末端的第二桥寡核苷酸特异性序列。

[0103] 桥寡核苷酸包括分别与第一探针和第二探针中的第一桥寡核苷酸特异性序列和第二桥寡核苷酸特异性序列互补的序列,以及可选的第三条形码,其可以是随机序列或者可以包含靶标核苷酸识别序列或样品识别序列。

[0104] 第一序列条形码或第二序列条形码或第三条形码中的至少一个分别存在于第一探针或第二探针或桥寡核苷酸中,其中,第一序列条形码或第二序列条形码或两者可以是随机序列,或者可以含有靶标核苷酸序列识别序列、样品识别序列和/或用于靶标计数的分子条形码;并且其中,T7 RNA聚合酶的启动子序列存在于第一探针、第二探针或桥寡核苷酸中,

[0105] 优选地,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和/或第二寡核苷酸特异性序列可以彼此独立地含有一个或多个经化学修饰的核苷酸。

[0106] 当探针的相应靶标特异性区段与靶标序列上基本相邻的区段杂交时,第一探针与第二探针连接,以提供经连接的探针。优选在汇集多个样品之后,T7 RNA聚合酶用于从多个样品中扩增经连接的探针。扩增产物进一步经历高通量测序技术以确定识别序列。

[0107] 优选地,第一探针的3'末端或第二探针的5'末端,或两者都被修饰,以允许第一探针与第二探针的化学连接。

[0108] 优选地,桥寡核苷酸在与第一探针序列互补的序列中或在与第二探针序列互补的序列中或在两者中包括一个或多个经化学修饰的核苷酸。

[0109] 优选地,第一探针、第二探针或桥寡核苷酸包含脱氧尿苷部分,脱氧尿苷部分允许通过使用尿嘧啶特异性切除试剂切割来线性化。

[0110] 优选地,第一探针的3'末端或第二探针的5'末端或两者都被修饰,以允许第一探针与第二探针的化学连接。

[0111] 优选地,第一探针或第二探针的桥接部分或两者包括经化学修饰的碱基,以允许与桥寡核苷酸的结合得到改善。

[0112] 本发明的另一方面涉及用于本发明方法的试剂盒。如本文所述,试剂盒包括第一探针、第二探针和桥寡核苷酸。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),检测和扩增感兴趣的靶标序列。探针的设计具有改进的结合特性,这带来了更高的分析特异性、灵敏度和准确性。本发明可应用于癌症诊断或胎儿染色体疾病,但不限于对一种或更多种序列和/或多态性(诸如SNP和/或插入缺失)的样品进行基因型分析。

[0113] 在一个特别优选的实施方案中,桥寡核苷酸包括识别样品的信息,并包括唯一的识别子。在这种情况下,第一探针和第二探针普遍适用于所有样品(并且仅包括识别靶标的信息)。因此,在一个优选的实施方案中,提供了根据本发明的方法、试剂盒或用途,其中,桥寡核苷酸包括条形码,该条形码包括能够计数每个样品的靶标序列的唯一序列。

[0114] 实验部分

[0115] 下面的实验提供了本发明的各种实施方案的范例,然而并不局限于这些实验。

[0116] 病例1:一名患有晚期肺癌的老年人预先接受化疗作为第一治疗路线。由于化疗和患者的状况,通过组织活组织检查分析肿瘤是不适用的。因此,不可能从患者的肿瘤中获得组织DNA。对于二线治疗,进行Guardant360液体活组织检查,并根据结果,对患者进行癌症靶向治疗。对于随访,应遵循靶向突变和相关的抗性突变。治疗费用为每月7000€,并且每周的样品对于检测所用药物的可能耐药性是最佳的。如果在产生耐药性的时候,能够非侵入性地进行个体化的靶向突变分析(具有快速的TA时间),则可以考虑新的大范围液体活组织检查,并且可以考虑新的靶向药物。

[0117] 在1-3个月的治疗中,忽略不再有效的耐药性进展的药物省去了€7-€21000。进

一步地,下一代治疗方案的益处实现得更快,并可能改善疾病的结果。

[0118] 病例2:检测母体血液中胎儿染色体三染色体性的多重连接试验。

[0119] 每一个相关的基因都被多个能够检测典型突变的探针靶向。多个探针的使用允许定量识别多种基因突变变体的比例。通常,10个不同的连接复合体靶向每个染色体,用于从母体血液中检测胎儿染色体疾病。10个不同探针的使用使得每条染色体有10个独立的重复。这组多重连接复合体检测显著的突变。使用统计工具分析经复制的染色体计数数据,以检测与染色体靶标的标准偏差模式的显著偏差。与染色体靶标的标准偏差模式的偏差允许识别缺失(诸如威廉姆斯综合征)和三染色体性(如帕托综合征、爱德华兹综合征或唐氏综合征)。

[0120] 病例3:检测癌症患者循环DNA中肿瘤相关突变的多重连接试验。

[0121] 为了检测癌症患者循环DNA中的肿瘤相关基因突变,每个相关基因被多重连接复合体靶向,该复合体可以检测癌症类型的典型突变。多重连接试验允许定量识别多基因突变变体的比例。

[0122] 案例4:用于检测和定量粪便或废水样品中的多重抗生素抗性基因的多重连接试验。

[0123] 为了检测从粪便或废水材料中提取的DNA中的抗生素抗性基因,每个抗生素抗性基因被多个特异性连接复合体靶向。每种靶标基因使用几个不同的连接复合体使得每个基因独立的技术复制。使用统计工具分析该复制的计数数据,以定量多个样品中多个抗生素抗性基因的拷贝数。

[0124] 附图的详细说明

[0125] 图1示出了所述发明的一个实施方案的工作流程。在步骤1中,将从样品中提取的核酸(DNA或RNA)(102)与一组连接复合体接触(104)。连接复合体在靶标核酸上退火(106)。在步骤2中,连接经退火的连接复合体,产生经连接的连接复合体(108)。在步骤3中,将来自多个样品的经连接的连接复合体(110)汇集在一起(112)。在步骤4中,使用T7 RNA聚合酶从经连接的连接复合体合成RNA(114)。可选地使用核酸内切酶和核酸外切酶的混合物去除探针和样品DNA。在步骤5中,将所述RNA转化成cDNA(116),并可选地使用PCR或乳液PCR进行扩增。在步骤6中,使用下一代DNA测序法对经扩增的DNA进行测序。在步骤7中,使用生物学流程(pipeline)将DNA测序结果转换成靶标计数。

[0126] 图2A示出了根据本文实施方案的具有多个探针实体的探针三联体的原理结构。多个探针实体包括在样品退火之前组装的左探针、右探针和桥寡核苷酸。左探针的第一个碱基可选地包括用于酶连接的磷酸部分或允许化学连接到相邻探针的5'末端的修饰,称为修饰1(200)。左探针的15-25个碱基包括桥结合序列1(202),其进一步包括用于有效结合桥寡核苷酸的经化学修饰碱基,称为桥位点1。前述片段还可选包括T7 RNA聚合酶启动子的反向互补序列。如前所述,该启动子序列可以存在于第一探针或第二探针中,而不存在于桥寡核苷酸中,条件是寡核苷酸和探针被设计成使得T7 RNA聚合酶能够转录用于在不同样品中计数靶标序列所需的所有信息。左探针的接下来15-30个碱基可选地包括用于PCR引物的通用结合位点,在此称为通用位点1(204)。左探针还可选地包括从5'末端开始的接下来10-20个碱基,其包括形成分子特异性条形码或样品特异性条形码的随机核苷酸片段,称为条形码1(206)。左探针进一步包括从5'末端开始的与基因靶标结合的接下来15-30个碱基基因

(208)。202或208的一些或全部核苷酸可包括增加探针对靶标或桥寡核苷酸(226)的亲力和化学修饰。左探针的最后一个碱基可选地包括用于酶连接的磷酸部分或修饰允许化学连接到相邻探针的5'末端的修饰,称为修饰1(210)。

[0127] 右探针的第一个碱基可选地包括用于酶连接的磷酸部分或修饰允许化学连接到相邻探针的5'末端的修饰,称为修饰2(212)。右探针从5'末端的15-30个碱基包括与基因靶标结合的右探针的一部分(214)。右探针从5'末端开始的接下来10-20个碱基可选地包括随机核苷酸片段,其形成分子特异性条形码或样品特异性条形码,称为条形码2(216)。右探针从5'末端的接下来15-30个碱基可选地包括通用的PCR引物结合位点,称为通用位点2(218)。右探针的最后15-25个碱基包括用于有效结合桥寡核苷酸的锚定序列,称为桥序列2(220)。前述片段还可选地包括T7 RNA聚合酶启动子的反向互补序列。如前所述,该启动子序列可以存在于第一探针或第二探针中,而不存在于桥寡核苷酸中,条件是寡核苷酸和探针被设计成使得T7 RNA聚合酶能够转录在不同样品中计数靶标序列所需的所有信息。214或220的一些或全部核苷酸可包括增加探针对靶标或桥寡核苷酸(224)的亲力和化学修饰。右探针的最后一个碱基可选地包括用于酶连接磷酸部分或允许化学连接到相邻探针的5'末端的修饰,称为修饰2(222)。

[0128] 桥寡核苷酸5'末端的前15-25个碱基与右探针的桥序列1(202)反向互补,称为桥序列3(226),并可选地包括经化学修饰的核苷酸以增强结合。桥的接下来15-25个碱基与左探针的桥序列2序列(220)反向互补,称为桥序列4(224),并可选地包括经化学修饰的核苷酸以增强结合。桥序列3(226)或桥序列4(224),可选地包括T7 RNA聚合酶启动子的序列。

[0129] 图2B示出了根据本文实施方案的第一探针和第二探针之间的间隙填充。这里,桥寡核苷酸包含桥序列3(226)和桥序列4(224)之间的间隙序列(228)。间隙序列(228)可以可选地包括T7 RNA聚合酶启动子的序列。通过引入聚合酶和一个或多个核苷酸来填充左探针和右探针之间的间隙。在这个过程中,可以使用Stoffel片段、Taq聚合酶或Phusion聚合酶。聚合酶添加(a)互补于通用桥寡核苷酸序列的核苷酸和(b)互补于靶标序列的核苷酸,从而填充第一探针和第二探针之间的两个间隙(即间隙1和间隙2),引起与桥寡核苷酸互补的左探针和右探针的连接。这样,如果位置228的桥寡核苷酸中存在条形码,则条形码被整合到互补序列中。可选地,桥寡核苷酸探针224从与所连接的探针互补的5'位点或3'位点延伸,使得条形码(如果存在于探针1或探针2中)和/或靶标序列208和214整合到桥寡核苷酸中,从而形成一个或多个经连接的连接复合体。然而,应注意聚合酶的作用不会干扰第一探针和第二探针在靶标序列位点的连接。例如,可以使用当到达双链DNA部分(例如存在于第一探针和第二探针的与靶标序列杂交的部分)时停止作用的聚合酶。T7 RNA聚合酶启动子序列可以嵌入连接复合体的桥寡核苷酸的位置226、228或224。脱氧尿苷部分可以嵌入至左探针的位置204和206之间,或位置206内。

[0130] 图3是改进与靶标结合或允许化学连接的探针修饰实例的示意图。改善靶标结合的化学修饰包括但不限于核糖核酸、肽核酸和/或锁核酸。允许化学连接的化学修饰包括,但不限于,5'-叠氮化物3'-炔基修饰,其可以在铜催化的环加成中接合(joinded)。

[0131] 图4提供了使用统计工具分析复制的染色体计数数据,以便由染色体靶标的标准偏差模式检测显著偏差的概况。使用多种靶标特异性连接复合体混合物(402)对多个样品中的多种基因靶标(404)进行剖析。进一步地,多种基因靶标404包括从母体血液中提取的

核酸。多探针靶向每一个可以检测典型突变的相关基因。典型地,至少10种,优选至少100种,更优选至少500种,最优选至少1000种不同的探针靶向每个染色体,用于从母体血液中检测胎儿染色体疾病。使用至少10种不同的探针会引起每个染色体至少有10个独立的重复。这组多探针检测特定的突变。使用统计工具分析复制的染色体计数数据,以检测与染色体靶标的标准偏差模式的显著偏差。与染色体靶标的标准偏差模式的偏差允许识别缺失(诸如威廉姆斯综合征)和三染色体性(如帕托综合征、爱德华兹综合征或唐氏综合征)。

[0132] 图5提供了通过将一组靶标特异性连接复合体混合物(502)与多种基因靶标(504)混合来定量地识别多基因突变变体比例的概况。多种基因靶标(504)包括从患者血浆中提取的核酸(DNA)。使用多种靶标特异性连接复合体混合物(502)对非常大量的样品中的多基因靶标(504)进行剖析。每一个相关基因都被多种不同的探针靶向,这些探针可以检测典型的突变。为了对多种基因靶标(504)进行分析,20种不同的探针靶向每个基因,以检测患有癌症的患者中循环DNA的肿瘤相关突变。

[0133] 图6示出了根据本文实施方案的多重连接依赖性探针扩增的流程图。在步骤604中,从两个或更多个样品中提取(靶标)核酸DNA/RNA。这种样品可以预先从人体(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))获得,从环境(包括水、废水、土壤等)获得的两个或更多个样品。在步骤606中,为每个样品提供连接复合体。连接复合体包括左探针、右探针和桥寡核苷酸探针。在步骤608中,将靶标样品与连接复合体混合,以使得能够自退火。在可选的步骤610中,可选地,在第一探针和第二探针之间填充间隙,并且进一步使用聚合酶进行桥寡核苷酸探针延伸,使得形成环状连接的连接复合体。例如,当探针三联体被组装时,或者当结合的靶特异性部分被延伸和连接时,就会发生这种情况。在步骤612中,使用连接酶来对连接复合体进行连接,从而形成环状的经连接的连接复合体。在步骤614中,从多个样品组中汇集多个这种经连接的连接复合体。在可选的步骤616中,加入可选的尿嘧啶特异性切除试剂,通过切割环状的经连接的连接体复合体来促进线性化。在步骤618中,加入T7 RNA聚合酶,其结合经连接的连接复合体的桥寡核苷酸序列(或第一或第二探针)上的T7 RNA聚合酶启动子位点,以转录RNA。在可选的步骤620中,向反应混合物中加入DNA特异性核酸外切酶和核酸内切酶,以除去游离样品和探针。在步骤622中,通过反转录过程从RNA合成cDNA。在可选的步骤624中,通过PCR或乳液PCR来扩增cDNA。在步骤626中,通过高通量测序进行测序。

[0134] 在不脱离由所附权利要求限定的本发明的范围的情况下,对前述本发明的实施方案的修改是可能的。用于描述和要求保护本发明的诸如“包括”、“包含”、“并入”、“具有”、“是”的表达意在以非排他性的方式解释,即允许没有明确描述的项目、组件或元件也存在。对单数的引用也应被解释为涉及复数。

- [0001] 序列表
- [0002] <110> 瑞士联邦水科学与技术研究所
- [0003] <120> 核酸的精确和大规模平行定量
- [0004] <130> Genomil1001
- [0005] <140> EP17188047.9
- [0006] <141> 2017-08-25
- [0007] <160> 28
- [0008] <170> PatentIn version 3.5
- [0009] <210> 1
- [0010] <211> 31
- [0011] <212> DNA
- [0012] <213> 人工序列
- [0013] <220>
- [0014] <223> 图1的示例性序列1
- [0015] <400> 1
- [0016] ggggcccgcc gtcgatcgga gccgtagga t 31
- [0017] <210> 2
- [0018] <211> 31
- [0019] <212> DNA
- [0020] <213> 人工序列
- [0021] <220>
- [0022] <223> 图1的示例性序列2
- [0023] <400> 2
- [0024] ttaaggtgcc gtcgatcgga gccgacgtac g 31
- [0025] <210> 3
- [0026] <211> 31
- [0027] <212> DNA
- [0028] <213> 人工序列
- [0029] <220>
- [0030] <223> 图1的示例性序列3
- [0031] <400> 3
- [0032] ttaaggtgcc gtcgatcgga gccgacgtac g 31
- [0033] <210> 4
- [0034] <211> 31
- [0035] <212> DNA
- [0036] <213> 人工序列
- [0037] <220>
- [0038] <223> 图1的示例性序列4

- [0039] <400> 4
[0040] tataatagag gtcgtgcagt cacgacccgg t 31
[0041] <210> 5
[0042] <211> 31
[0043] <212> DNA
[0044] <213> 人工序列
[0045] <220>
[0046] <223> 图1的示例性序列5
[0047] <400> 5
[0048] accaggtgcc gtcgatcgga gccgacccgg t 31
[0049] <210> 6
[0050] <211> 31
[0051] <212> DNA
[0052] <213> 人工序列
[0053] <220>
[0054] <223> 图1的示例性序列6
[0055] <400> 6
[0056] gggccgggag gtcgtgcact cacgttagga t 31
[0057] <210> 7
[0058] <211> 31
[0059] <212> DNA
[0060] <213> 人工序列
[0061] <220>
[0062] <223> 图1的示例性序列7
[0063] <400> 7
[0064] tccaggtgag tcgatccgtc acgtacgtac g 31
[0065] <210> 8
[0066] <211> 31
[0067] <212> DNA
[0068] <213> 人工序列
[0069] <220>
[0070] <223> 图1的示例性序列8
[0071] <400> 8
[0072] aaaattttag cgtacgtcgt acgttttagga t 31
[0073] <210> 9
[0074] <211> 31
[0075] <212> DNA
[0076] <213> 人工序列
[0077] <220>

- [0078] <223> 图1的示例性序列9
[0079] <400> 9
[0080] tataatagag gtcgtgcagt cacgaccgg t 31
[0081] <210> 10
[0082] <211> 31
[0083] <212> DNA
[0084] <213> 人工序列
[0085] <220>
[0086] <223> 图1的示例性序列10
[0087] <400> 10
[0088] aggaccttga gtcgatccgc acgtaccgg t 31
[0089] <210> 11
[0090] <211> 31
[0091] <212> DNA
[0092] <213> 人工序列
[0093] <220>
[0094] <223> 图1的示例性序列11
[0095] <400> 11
[0096] agcgaccgag gtcgtgcagt cacgacgtac g 31
[0097] <210> 12
[0098] <211> 31
[0099] <212> DNA
[0100] <213> 人工序列
[0101] <220>
[0102] <223> 图1的示例性序列12
[0103] <400> 12
[0104] tataatagag gtcgtgcagt cacgaccgg t 31
[0105] <210> 13
[0106] <211> 31
[0107] <212> DNA
[0108] <213> 人工序列
[0109] <220>
[0110] <223> 图1的示例性序列13
[0111] <400> 13
[0112] ggaaaaagcc gtcgatcgga gccgttagga t 31
[0113] <210> 14
[0114] <211> 31
[0115] <212> DNA
[0116] <213> 人工序列

- [0117] <220>
[0118] <223> 图1的示例性序列14
[0119] <400> 14
[0120] atatacagag gtcgtgcagt caccttagga t 31
[0121] <210> 15
[0122] <211> 30
[0123] <212> DNA
[0124] <213> 人工序列
[0125] <220>
[0126] <223> 图1的示例性序列15
[0127] <400> 15
[0128] gagaccgac ctcgtccagt cagaccgt 30
[0129] <210> 16
[0130] <211> 31
[0131] <212> DNA
[0132] <213> 人工序列
[0133] <220>
[0134] <223> 图1的示例性序列16
[0135] <400> 16
[0136] cgcacgcgag gtcgtgcagt cacgacgtac g 31
[0137] <210> 17
[0138] <211> 29
[0139] <212> DNA
[0140] <213> 人工序列
[0141] <220>
[0142] <223> 图1的示例性序列17
[0143] <400> 17
[0144] tacaagccgt cgatcggagc cgaccgggt 29
[0145] <210> 18
[0146] <211> 31
[0147] <212> DNA
[0148] <213> 人工序列
[0149] <220>
[0150] <223> 图1的示例性序列18
[0151] <400> 18
[0152] tataatagag gtcgtgcagt cagaccgg t 31
[0153] <210> 19
[0154] <211> 31
[0155] <212> DNA

- [0156] <213> 人工序列
[0157] <220>
[0158] <223> 图1的示例性序列19
[0159] <400> 19
[0160] gggcaattag cgtacgtcgt acgtacgtac g 31
[0161] <210> 20
[0162] <211> 31
[0163] <212> DNA
[0164] <213> 人工序列
[0165] <220>
[0166] <223> 图1的示例性序列20
[0167] <400> 20
[0168] tatgcgagcc gtcgatcgga gccgacgtac g 31
[0169] <210> 21
[0170] <211> 31
[0171] <212> DNA
[0172] <213> 人工序列
[0173] <220>
[0174] <223> 图1的示例性序列21
[0175] <400> 21
[0176] aggaccttga gtcgatccgc acgtacccgg t 31
[0177] <210> 22
[0178] <211> 31
[0179] <212> DNA
[0180] <213> 人工序列
[0181] <220>
[0182] <223> 图1的示例性序列22
[0183] <400> 22
[0184] attacaagcc gtcgatcgga gccgacccgg t 31
[0185] <210> 23
[0186] <211> 31
[0187] <212> DNA
[0188] <213> 人工序列
[0189] <220>
[0190] <223> 图1的示例性序列23
[0191] <400> 23
[0192] gaagaattag cgtacgtcgt acgttttagga t 31
[0193] <210> 24
[0194] <211> 31

- [0195] <212> DNA
[0196] <213> 人工序列
[0197] <220>
[0198] <223> 图1的示例性序列24
[0199] <400> 24
[0200] gggcaattag cgtacgtcgt acgtacgtac g 31
[0201] <210> 25
[0202] <211> 31
[0203] <212> DNA
[0204] <213> 人工序列
[0205] <220>
[0206] <223> 图1的示例性序列25
[0207] <400> 25
[0208] gggcaattag cgtacgtcgt acgtacgtac g 31
[0209] <210> 26
[0210] <211> 31
[0211] <212> DNA
[0212] <213> 人工序列
[0213] <220>
[0214] <223> 图1的示例性序列26
[0215] <400> 26
[0216] gagcacttag cgtacgtcgt acgtacccgg t 31
[0217] <210> 27
[0218] <211> 31
[0219] <212> DNA
[0220] <213> 人工序列
[0221] <220>
[0222] <223> 图1的示例性序列27
[0223] <400> 27
[0224] aaaggggcga gtcgatccgc acgttttagga t 31
[0225] <210> 28
[0226] <211> 31
[0227] <212> DNA
[0228] <213> 人工序列
[0229] <220>
[0230] <223> 图1的示例性序列28
[0231] <400> 28
[0232] agcgcgcgcc gtcgatcgga gccgtttagga t 31

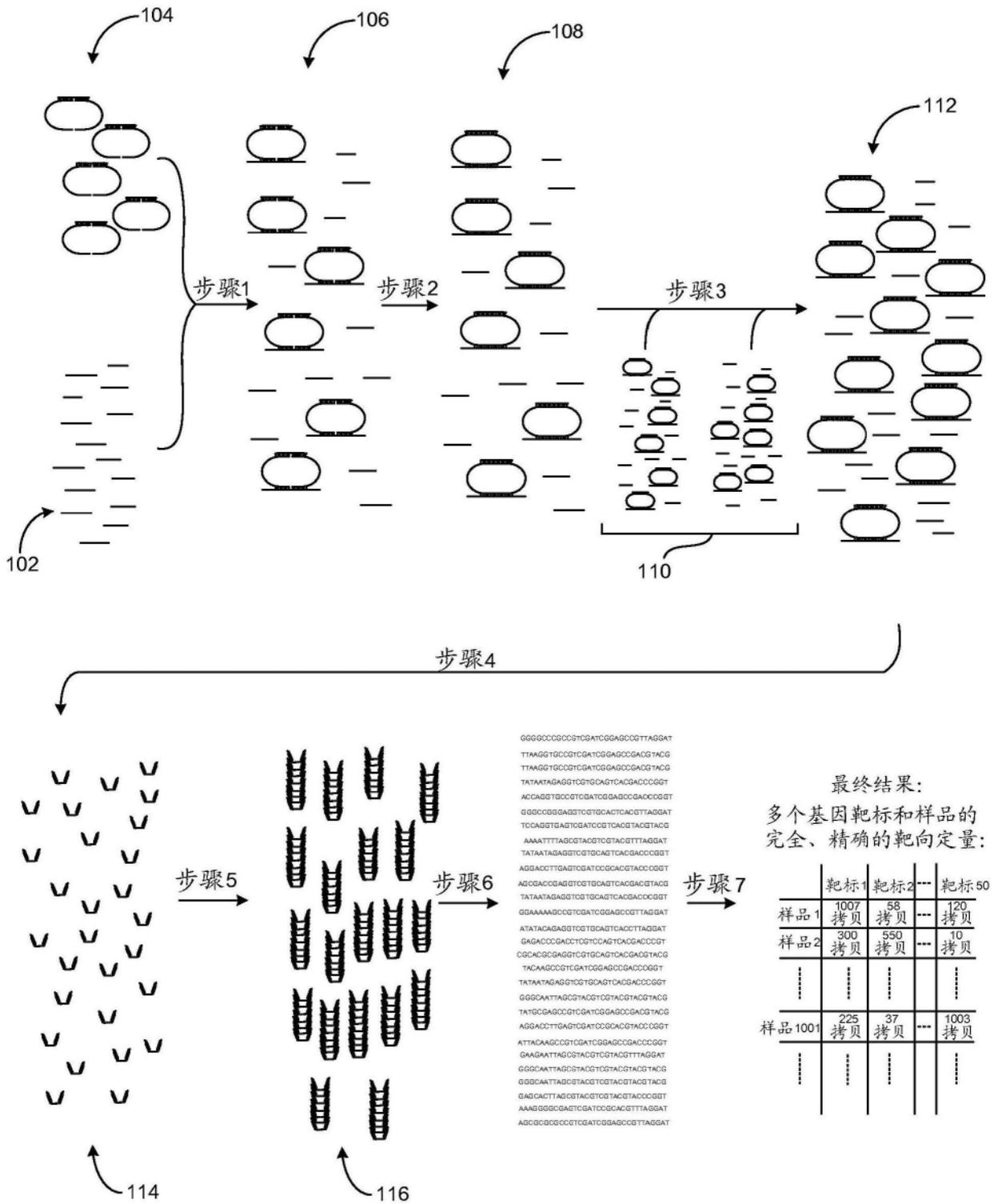


图1

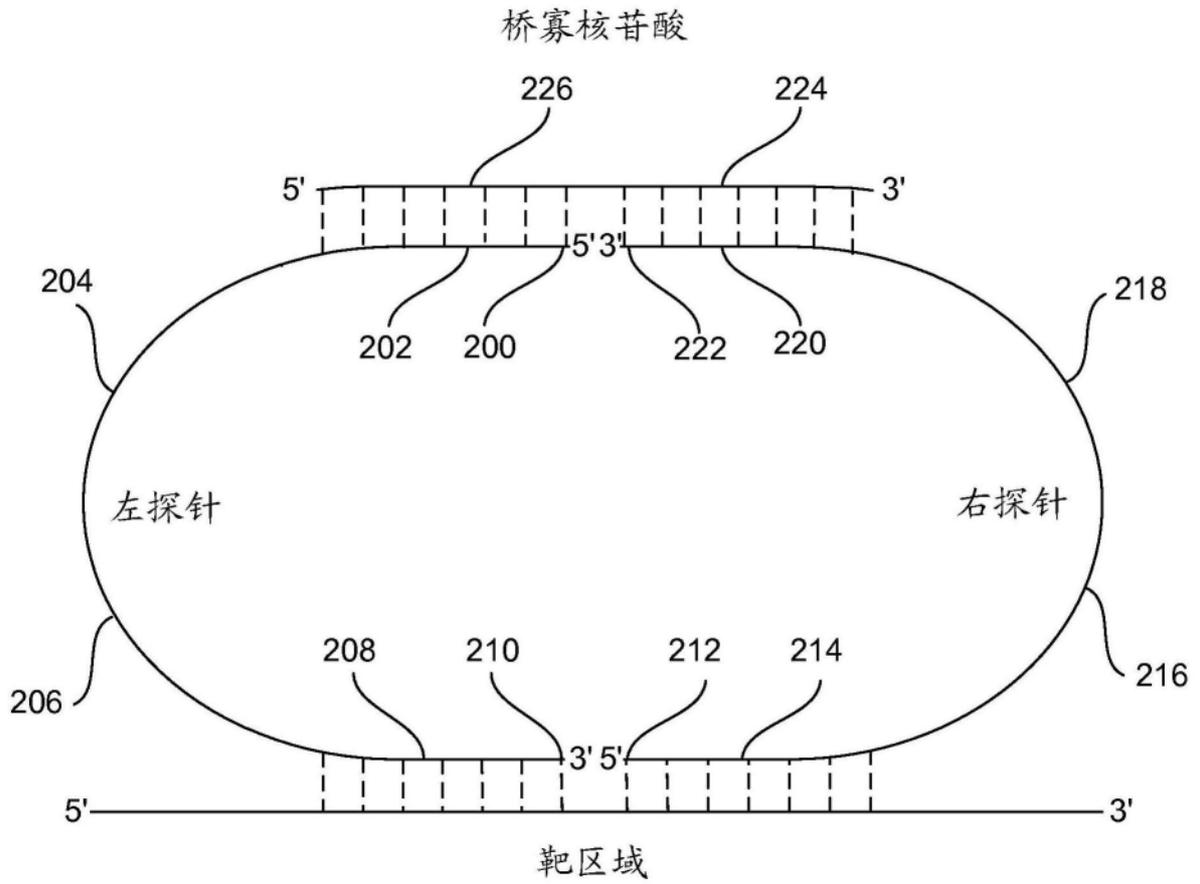


图2A

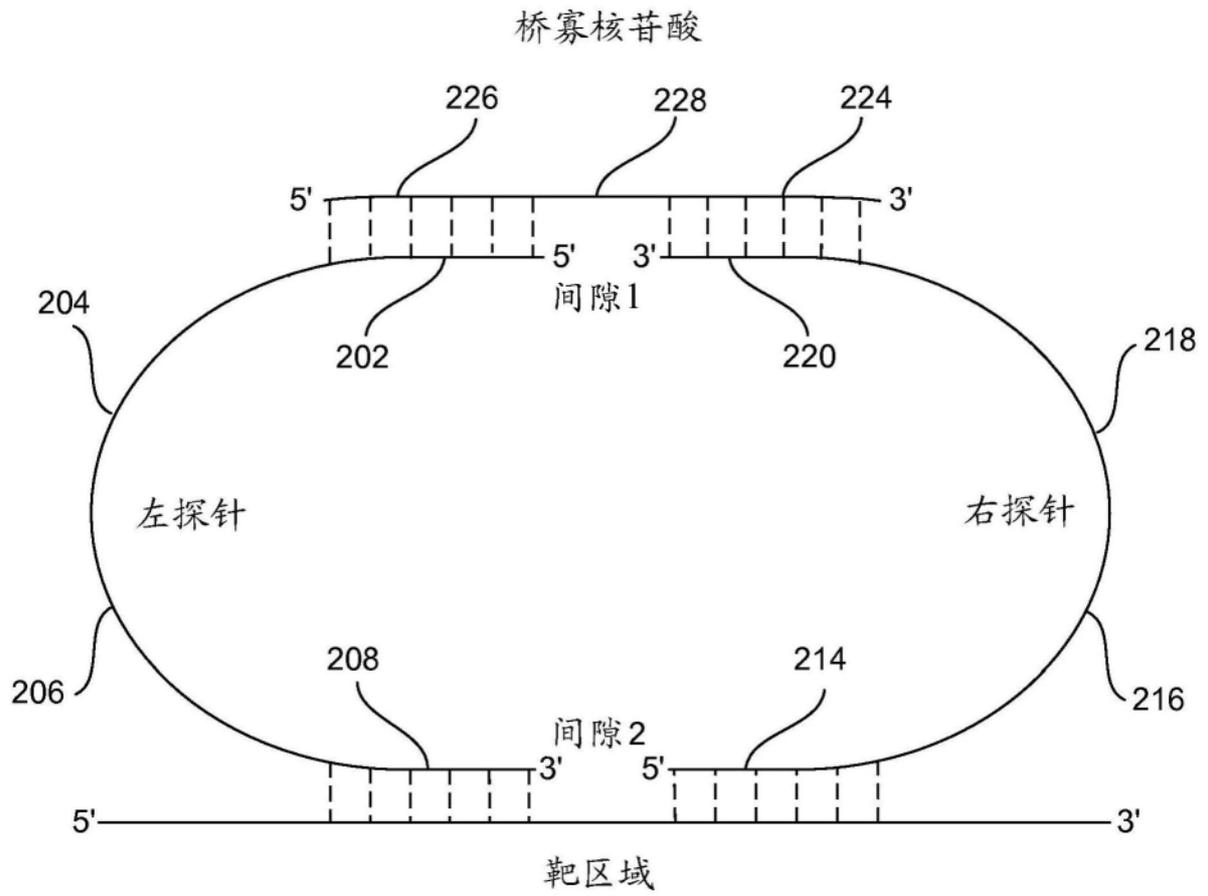


图2B

在区域 202、208、214、224 和 226 中的经化学修饰的核苷酸的例子。

在化学连接所需的位置 200、210、212 和 222 进行化学修饰的例子

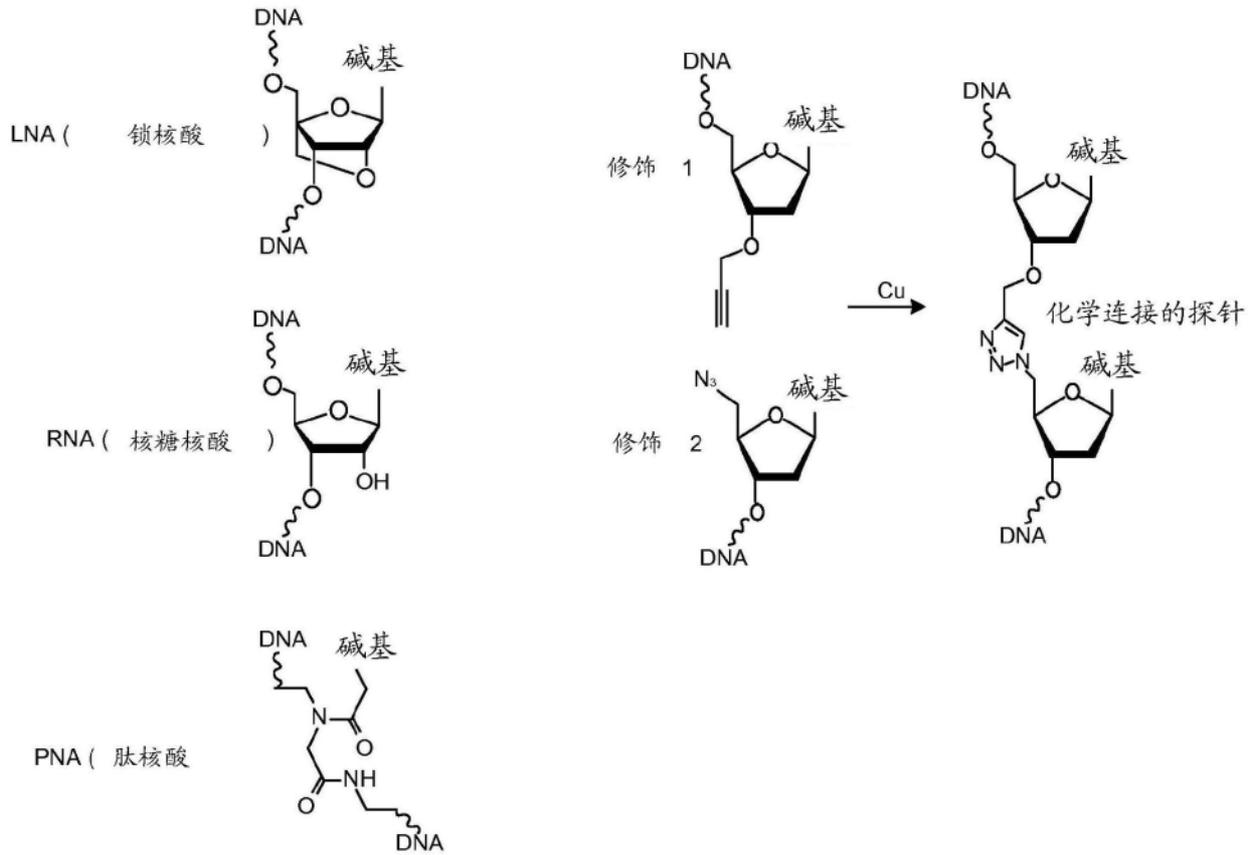
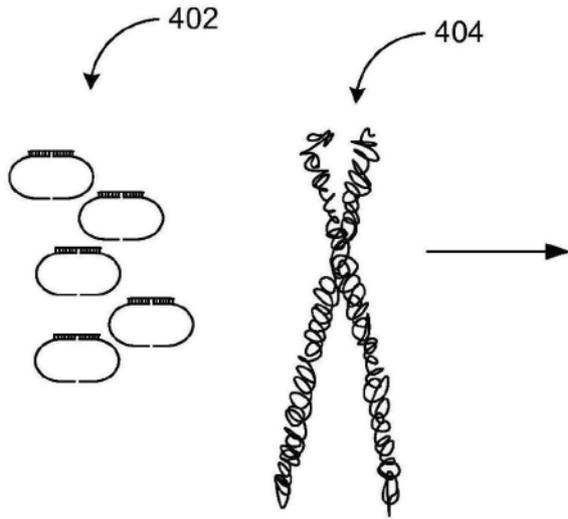


图3



完全、精确的染色体拷贝数定量

	靶标 1	靶标 2	-----	靶标 10
染色体 1	1007 拷贝	995 拷贝	-----	1120 拷贝
染色体 2	890 拷贝	970 拷贝	-----	1150 拷贝
⋮	⋮	⋮		
染色体 23	923 拷贝	1002 拷贝	-----	1107 拷贝

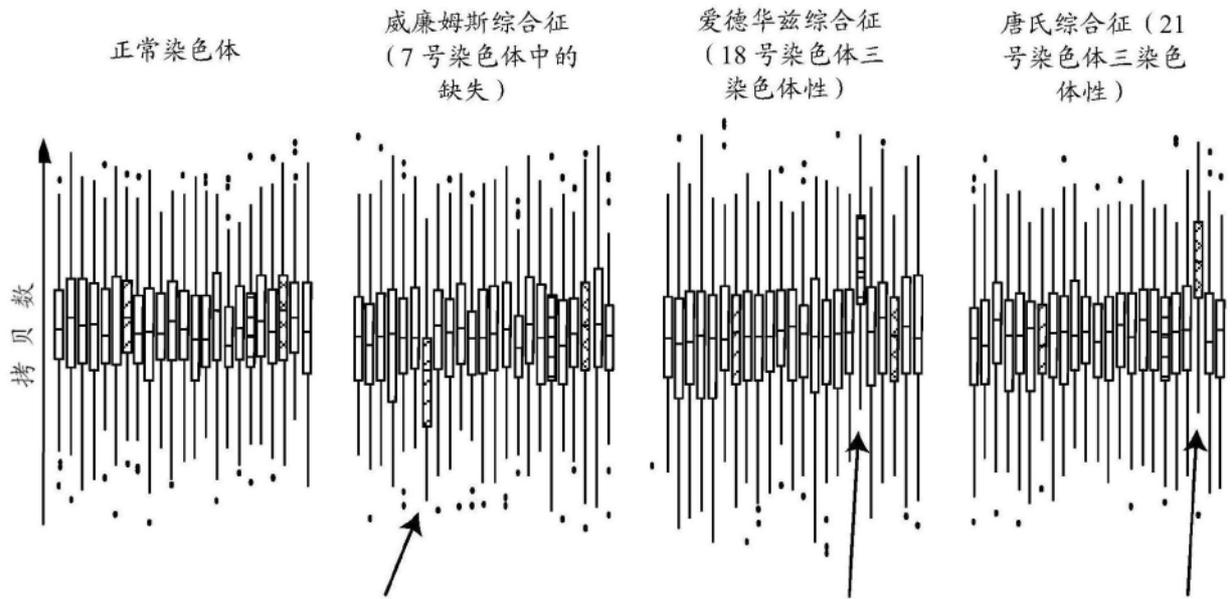


图4

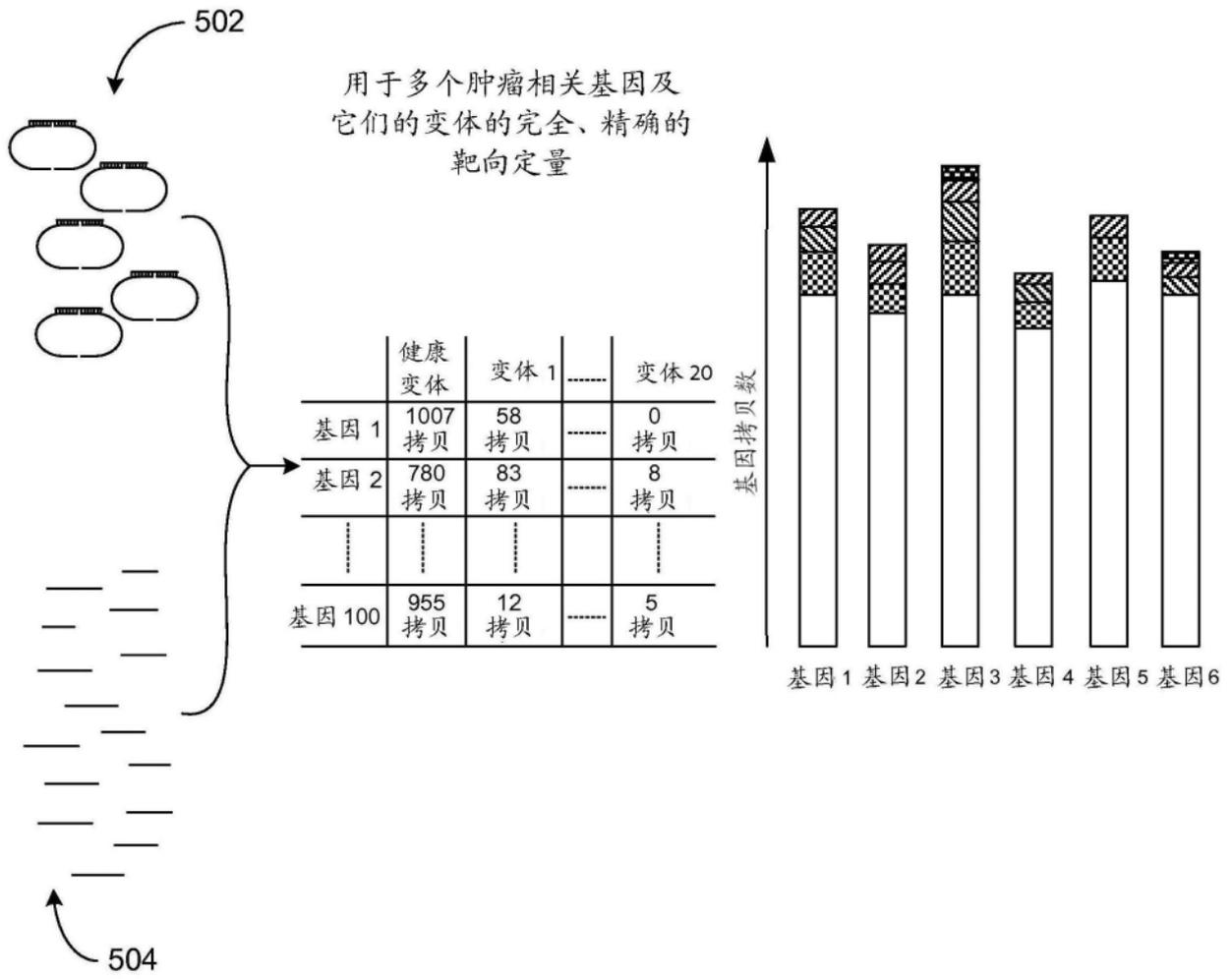


图5

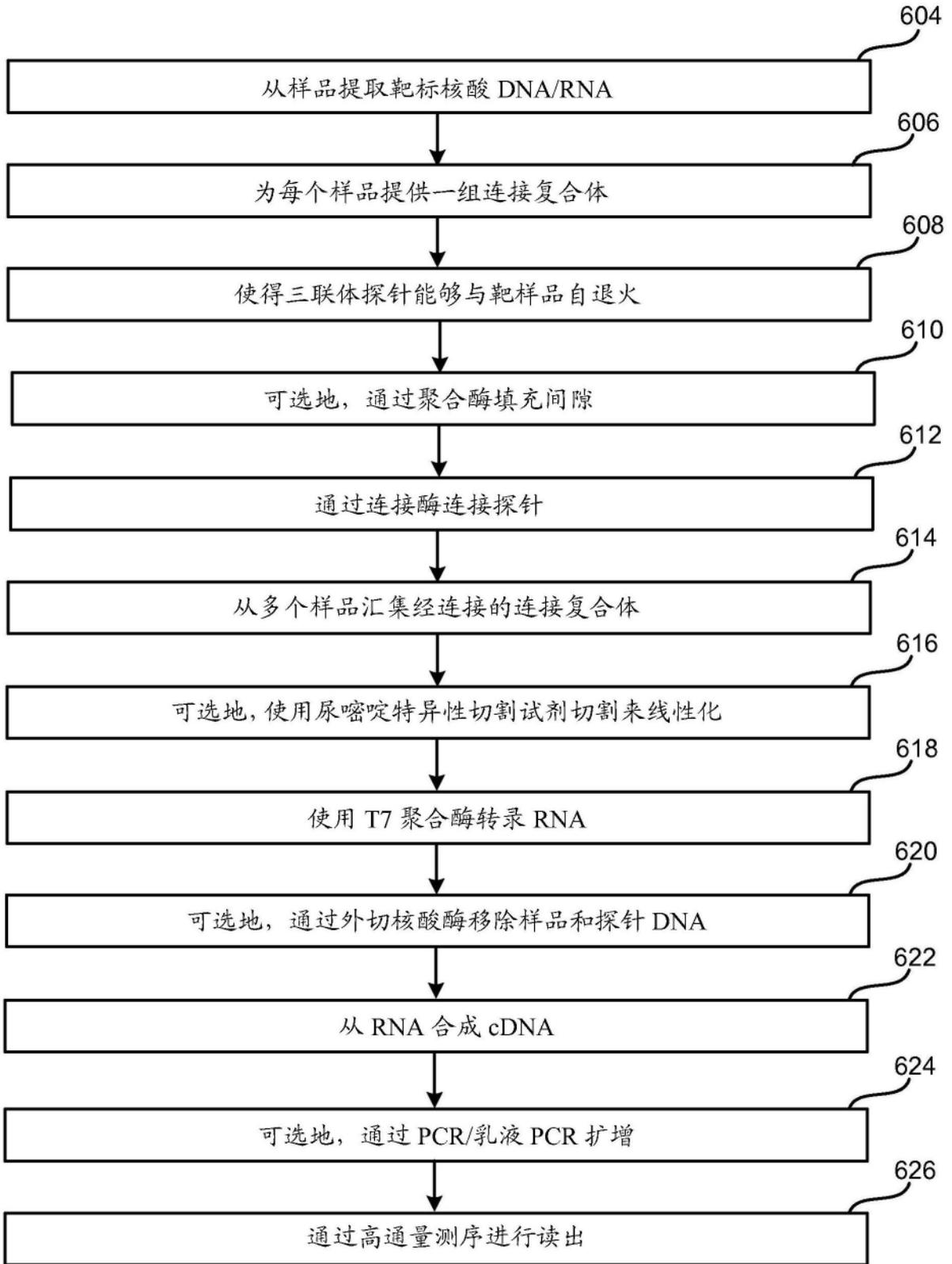


图6