



(12)发明专利申请

(10)申请公布号 CN 107922970 A

(43)申请公布日 2018.04.17

(21)申请号 201680045912.8

(74)专利代理机构 中国专利代理(香港)有限公司 72001

(22)申请日 2016.08.03

代理人 梁谋 黄希贵

(30)优先权数据

62/201727 2015.08.06 US

(51)Int.Cl.

C12Q 1/6855(2018.01)

(85)PCT国际申请进入国家阶段日

2018.02.05

(86)PCT国际申请的申请数据

PCT/EP2016/068546 2016.08.03

(87)PCT国际申请的公布数据

W02017/021449 EN 2017.02.09

(71)申请人 豪夫迈·罗氏有限公司

地址 瑞士巴塞尔

(72)发明人 B.戈德文

权利要求书1页 说明书9页 附图2页

(54)发明名称

通过单探针引物延伸的靶标富集

(57)摘要

本发明包括用单引物延伸和低偏差限制性扩增富集靶核酸的方法和组合物。

1. 一种扩增靶序列的方法,其包括以下步骤:
 - a) 使所述靶核酸与引物和聚合酶接触,其中所述引物包含靶结合位点和独特的分子鉴定标签 (UID);
 - b) 进行聚合酶延伸反应和终止以产生单链引物延伸产物;
 - c) 将衔接子连接至单链引物延伸产物的每一端以产生连接产物,其中衔接子包含至少一个通用引发位点;
 - d) 在扩增反应中扩增所述连接产物以产生扩增的靶序列,其利用结合至至少一个通用引发位点的至少一个引物。
2. 权利要求1的方法,其中所述引物和至少一个衔接子包含相互相容的通用连接位点。
3. 权利要求1-2的方法,其中所述靶标结合位点是预先设计的靶标特异性序列。
4. 权利要求1-2的方法,其中所述靶标结合位点是随机序列。
5. 权利要求1-4的方法,其中所述终止通过选自以下的方法实现:温度转换,添加特异性酶抑制剂,添加螯合剂,掺入含尿苷的碱基、随后用尿嘧啶-N-DNA糖基化酶处理。
6. 权利要求1-5的方法,其中至少一个衔接子包含条形码。
7. 权利要求6的方法,其中所述条形码是多重样品ID (MID)。
8. 权利要求1-7的方法,其中所述扩增是线性扩增。
9. 权利要求1-7的方法,其中所述扩增是指数扩增。
10. 权利要求1-9的方法,其进一步包括在步骤b) 和c) 中的至少一个之后的纯化步骤。
11. 一种用于扩增靶序列的试剂盒,其包含:
 - a) 引物,其包含靶结合位点、独特分子鉴定标签 (UID) 和通用连接位点;
 - b) 至少一个衔接子,其包含至少一个通用引发位点、多重样品ID (MID) 和通用连接位点。
12. 权利要求11的试剂盒,其包含具有不同通用引发位点的两个衔接子,但仅一个衔接子包含通用连接位点和MID。
13. 权利要求11-12的试剂盒,其进一步包含以下中的一种或多种:核酸聚合酶、连接酶、热稳定DNA聚合酶和通用引物。

通过单探针引物延伸的靶标富集

发明领域

[0001] 本发明通常涉及富集样品中的核酸靶标,并且更具体地涉及富集用于核酸测序(包括高通量测序)的靶标。

[0002] 发明背景

本发明属于一类允许使用者聚焦于待测序核酸内的目标区域的技术。这降低了与测序反应和随后的数据分析相关的成本。目前有三种一般类型的选择性捕获样品中存在的核酸内的目标区域的技术。第一种技术是杂交捕获,其中通过可以选择性结合至捕获表面的探针的杂交来捕获目标区域。该捕获允许除去非靶标核酸,随后释放和收集捕获的靶分子。这种类型的技术具有优点,包括捕获外显子组大小的区域和包含未知结构变异的区域的能力。缺点包括长而复杂的方案,其倾向于花费超过8小时来完成。复杂性主要由在杂交前制备随机片段化的鸟枪文库的需求所引起。单独的杂交步骤可能需要三天来完成。这种类型的技术的实例包括SeqCap EZ (NimbleGen, Madison, Wisc.)和SureSelect靶标富集系统(Agilent, Santa Clara, Cal.)。

[0003] 靶标富集的另一方法是基于双重靶标引物的扩增。在该方法中,在靶标的边界上使用两个探针来富集目标区域。所述方法倾向于花费少于8小时来完成,并且比杂交捕获方法更简单。然而,基于双重引物的技术不能富集具有未知结构变异的序列。最成熟的双重引物方法是多重PCR。这是一个非常简单的单一过程,但在每个反应管只能扩增数十种靶标。目前可用其他更新的技术,包括TruSeq Amplicon (Illumina, San Diego, Cal.)和Ion Torrent Ampliseq (Life Technologies, Grand Island, NY)产品,其能够在单个反应管中扩增数百至数千种靶标,并且仅需要几个处理步骤。

[0004] 第三种技术是基于单靶标引物的扩增。在该方法中,通过扩增由单个靶引物和末端连接的通用引物限定的区域来富集靶标。类似于基于杂交的方法;这些技术需要在靶寡核苷酸的选择性杂交之前产生随机片段化的鸟枪文库。然而,不是使用该寡核苷酸来捕获靶标并洗掉非靶标分子,而是采用扩增步骤,其选择性扩增随机产生的末端和靶标特异性寡核苷酸之间的区域。该技术的优点在于,不同于双重引物技术,其允许检测具有未知结构变异的序列。它比基于杂交的技术更快且更简单。然而,这种类型的技术仍比基于双重引物的方法更慢且更复杂。这种类型的技术的实例是Archer的锚定多重PCR (Archer Dx, Boulder, Colo.)和Ovation®靶标富集系统(NuGen, San Carlos, Cal.)。

[0005] 对于也将容纳靶标序列中未知的结构变异的快速和简单的靶标富集的方法,仍然存在未满足的需求。

[0006] 发明简述

在一个实施方案中,本发明是扩增靶序列的方法,其包括以下步骤:使所述靶核酸与引物和聚合酶接触,其中所述引物包含靶结合位点和独特的分子鉴定标签(UID);进行聚合酶延伸反应和终止以产生单链引物延伸产物;将衔接子连接至单链引物延伸产物的每一端以产生连接产物,其中衔接子包含至少一个通用引发位点;在扩增反应中扩增所述连接产物以产生扩增的靶序列,其利用结合至至少一个通用引发位点的至少一个引物。在一些实施

方案中,所述引物和至少一个衔接子包含相互相容的通用连接位点。在一些实施方案中,所述靶标结合位点是预先设计的靶标特异性序列。在一些实施方案中,所述靶标结合位点是随机序列。在一些实施方案中,所述终止通过选自以下的方法实现:温度转换,添加特异性酶抑制剂,添加螯合剂,掺入含尿苷的碱基、随后用尿嘧啶-N-DNA糖基化酶处理。在一些实施方案中,至少一个衔接子包含条形码。所述条形码可以是多重样品ID (MID)。所述扩增可以是线性扩增或指数扩增。在一些实施方案中,所述方法在引物延伸和连接中的至少一种之后进一步包括纯化步骤。

[0007] 在其他实施方案中,本发明是用于扩增靶序列的试剂盒,其包含:引物,其包含靶结合位点、独特分子鉴定标签 (UID) 和通用连接位点;至少一个衔接子,其包含至少一个通用引发位点、多重样品ID (MID) 和通用连接位点。在一些实施方案中,所述试剂盒包含具有不同通用引发位点的两个衔接子,但仅一个衔接子包含通用连接位点和MID。在一些实施方案中,所述试剂盒进一步包含以下中的一种或多种:核酸聚合酶、连接酶、热稳定DNA聚合酶和通用引物。

[0008] 附图简述

图1是所述方法的步骤的示意图。

[0009] 发明详述

定义

如本文所用,“探针”意指能够选择性结合至特异性预期的靶生物分子(例如,待被探针结合、捕获或杂交的目标核酸序列)的任何分子。

[0010] 如本文所用,“衔接子”意指可以被添加至另一序列、以便将额外的特性输入该序列的核苷酸序列。衔接子可以是单链或双链的,或者可以同时具有单链部分和双链部分。

[0011] 如本文所用,“条形码”意指为分子赋予身份的核苷酸序列。条形码可以为单个分子(及其拷贝)赋予独特身份。此条形码是独特ID (UID)。条形码可以为来自相同来源(例如,患者)的整个分子群体(及其拷贝)赋予身份。此条形码是多重ID (MID)。

[0012] 如本文所用,“连接位点”是可以便于连接的核酸分子的一部分(不同于双链分子的平端)。存在于两个分子上的“相容连接位点”使得两个分子能够与彼此优先结合。

[0013] 如本文所用,“单链连接”是以至少一个单链底物开始并且通常涉及一个或多个双链或部分双链衔接子的连接程序。

[0014] 如本文所用,“通用引物”和“通用引发位点”是指不是天然存在于靶序列中的引物和引发位点。通常,所述通用引发位点存在于衔接子或靶标特异性引物中。所述通用引物可以结合至通用引发位点并指导从通用引发位点的引物延伸。

[0015] 本发明的方法可以用作测序方案(包括高通量单分子测序方案)的一部分。本发明的方法产生待测序的靶核酸的文库。文库中的靶核酸可以并入用于分子鉴定和样品鉴定的条形码。

[0016] 靶标特异性引物延伸

本发明包括用于靶标特异性引物的线性引物延伸步骤。所述线性延伸步骤与本领域中实施的指数扩增相比具有几个优点。每种靶核酸的特征在于独特的合成速率,其取决于靶标特异性引物的退火速率和聚合酶可以通过特定靶序列读取的速率。延伸速率和合成速率的差异造成了偏差,其可导致单轮合成的轻微差异。然而,在PCR期间,所述轻微差异变得指

数扩增。所得差距被称为PCR偏差。所述偏差可掩盖样品中每个序列的初始数量的任何差异,并排除任何定量分析。

[0017] 本发明将靶标特异性引物(包括基因特异性引物和偶然对基因组内的结合位点特异性的简并引物)的延伸限定至单个步骤。任何指数扩增都用不受模板依赖性偏差影响或者比靶标特异性引物受偏差影响更小的通用引物进行。

[0018] 参考图1,所述方法包括引物延伸。所述方法包括反应建立步骤(步骤1,引物杂交),随后为聚合酶添加步骤(步骤2,引物延伸)。任选地,所述引物杂交和延伸步骤同时进行,即在相同反应条件下作为单一步骤进行。在其他实施方案中,所述步骤作为具有不同反应条件的两步法分开进行。

[0019] 所述引物杂交步骤由引物的靶标特异性区域介导。在一些实施方案中,所述靶标特异性区域能够与位于基因的外显子、内含子或非翻译部分或基因的未转录部分(例如启动子或增强子)中的基因的区域杂交。在一些实施方案中,所述基因是蛋白编码基因,但在其他实施方案中,所述基因不是蛋白编码基因,诸如RNA编码基因或假基因。在还有其他实施方案中,所述靶标特异性区域位于基因间区域中。对于RNA靶标,所述引物可以包含寡聚-dT序列。

[0020] 替代预先设计的靶标特异性区域,所述引物可以含有简并序列,即随机引入核苷酸串。此引物也可以在基因组内找到结合位点,并且充当该结合位点的靶标特异性引物。

[0021] 除了靶标特异性区域之外,所述引物可以包含额外的序列。在一些实施方案中,这些序列位于所述靶标特异性区域的5'-末端。在其他实施方案中,可以在引物内别处包括这些序列,只要靶标特异性区域能够与靶标杂交并驱动引物延伸反应,如下所述。所述引物内的额外序列可以包括一个或多个条形码序列,诸如独特的分子鉴定序列(UID)或多重样品鉴定序列(MID)。所述条形码序列可以作为单个序列或两个或更多个序列存在。

[0022] 在一些实施方案中,所述额外序列包括便于连接至引物的5'-末端的序列。所述引物可以含有通用连接序列,其使得能够连接如以下部分中所述的衔接子。

[0023] 在一些实施方案中,所述额外序列包括一个或多个通用扩增引物的一个或多个结合位点。

[0024] 所述引物延伸步骤通过核酸聚合酶进行。取决于所分析的核酸的类型,所述聚合酶可以是DNA依赖性DNA聚合酶(“DNA聚合酶”)或RNA依赖性DNA聚合酶(“逆转录酶”)。

[0025] 在一些实施方案中,期望控制在引物延伸反应中合成的核酸链的长度。(图1,延伸停止)。如下所述,该链的长度决定经受该方法的后续步骤和任何下游应用的核酸的长度。延伸反应可以通过本领域已知的任何方法来终止。所述反应可以物理(例如通过温度转换或添加聚合酶抑制剂)停止。在一些实施方案中,所述反应通过将反应物置于冰上来停止。在其他实施方案中,所述反应通过提高温度以使非热稳定聚合酶失活来终止。在还有其他实施方案中,通过添加能够螯合酶的关键辅因子的螯合剂(诸如EDTA)或能够可逆地或不可逆地使酶失活的另一种化学或生物物质化合物来终止反应。

[0026] 控制引物延伸产物的长度的另一种方法是通过限制关键组分(例如,dNTP)以直接限制延伸长度或限制Mg²⁺以减慢延伸速率并改善控制延伸停止点的能力来限制(starving)延伸反应。本领域技术人员能够在实验上或理论上确定允许有限引物延伸以主要产生所需长度产物的关键组分的适当量。

[0027] 控制引物延伸产物的长度的另一种方法是添加终止子核苷酸,包括可逆的终止子核苷酸。本领域技术人员能够在实验上或理论上确定允许有限引物延伸以主要产生所需长度产物的终止子和非终止子核苷酸的适当比率。终止子核苷酸的实例包括如US8163487中所述的双脱氧核苷酸、2'-磷酸核苷酸、3'-O-封闭的可逆终止子和3'未封闭的可逆终止子,如例如US20140242579和Guo, J.,等人, *Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides*, P.N.A.S. 2008 105 (27) 9145-9150中所述。控制引物延伸产物的长度的又另一种方法是向引物延伸反应中添加有限量的尿嘧啶(dUTP)。含尿嘧啶的DNA然后可以用尿嘧啶-N-DNA糖基化酶处理以产生无碱基位点。具有无碱基位点的DNA可通过以任选添加碱进行热处理而降解,以改善降解效率,如US8669061中所述。本领域技术人员能够在实验上或理论上确定扩增反应中dUTP与dTTP的适当比例,其允许在核酸内切酶处理后有限地包括dUTP以主要产生期望长度产物。

[0028] 在一些实施方案中,延伸产物的长度固有地受输入核酸的长度限制。例如,存在于母体血浆中的无细胞DNA长度低于200bp,其中大部分长度为166bp。Yu, S.C.Y., 等人, *Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing*, PNAS USA 2014; 111 (23):8583-8。在健康个体和癌症患者的血浆中发现的无细胞DNA的中值长度为约185-200bp。Giacona, M.B., 等人, *Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls*, Pancreas 1998; 17 (1):89-97。保存不佳或化学处理的样品可含有化学或物理降解的核酸。例如,福尔马林固定的石蜡包埋组织(FFPET)通常产生平均长度为150bp的核酸。

[0029] 纯化

在一些实施方案中,本发明的方法在通过DNA聚合酶或逆转录酶进行引物延伸之后包括一个或多个纯化步骤。所述纯化将除去未使用的引物分子和用于产生引物延伸产物的模板分子。在一些实施方案中,通过外切核酸酶消化来除去模板核酸和除了延伸的引物以外的所有核酸片段。在该实施方案中,用于引物延伸的引物可以具有使得引物 and 任何延伸产物对外切核酸酶消化具有抗性的5'-末端修饰。此修饰的实例包括硫代磷酸酯键。在其他实施方案中,可以通过节约DNA的酶促处理,例如RNase消化,包括RNA酶H消化来除去RNA模板。在还有其他实施方案中,通过尺寸排阻方法,例如凝胶电泳、色谱或等速电泳,将引物和大尺寸模板DNA与延伸产物分开。

[0030] 在一些实施方案中,纯化是通过亲和结合。在该实施方案的变型中,亲和力针对特定靶序列(序列捕获)。在其他实施方案中,所述引物包含亲和和标签。可以使用本领域已知的任何亲和和标签,例如生物素或抗体或特异性抗体存在的抗原。亲和和标签的亲和配偶体可以存在于溶液中,例如在悬浮的颗粒或珠粒上,或者结合至固体支持物。在亲和纯化的过程中,反应混合物的未结合组分被洗掉。在一些实施方案中,采取额外步骤来除去未使用的引物。

[0031] 连接通用引发序列

在一些实施方案中,本发明包括连接步骤。例如,可以将均聚物尾部添加至核酸的3'末端。在该实施方案中,所述均聚物可以充当反向互补均聚物的结合位点(类似于具有用于

mRNA的聚-T引物的聚-A尾部)。所述连接将一个或多个衔接子序列添加至前一步骤中产生的引物延伸产物。所述衔接子序列提供一个或多个通用引发位点(用于扩增或测序)以及任选一个或多个条形码。连接衔接子的确切模式是不重要的,只要所述衔接子变得与引物延伸产物缔合并能够实现下面描述的后续步骤。

[0032] 在上面描述的一些实施方案中,所述方法涉及包括通用引发序列(“引发位点”)的靶标特异性引物并产生具有单一引发位点的引物延伸产物。在此类实施方案中,仅提供一个额外的引发序列(“引发位点”)以实现指数扩增。在其他实施方案中,所述靶标特异性引物不包括通用引发位点。在此类实施方案中,需要提供两个引发位点以实现指数扩增。具有通用引发位点的衔接子可以通过本领域可用的任何单链连接方法添加。

[0033] 在延伸引物包含通用连接位点的实施方案中,可以使用单链连接方法的一个实例。在此类实施方案中,具有与引物中的通用连接位点互补的双链区和单链突出端的衔接子可以退火并连接,如图1,步骤4所示。衔接子的单链3'-突出端与引物的5'-末端的通用连接位点的退火产生在含有引物延伸产物的链中具有缺口的双链区域。两条链可以通过DNA连接酶或另一种酶或可以催化引物延伸产物的5'-磷酸酯和衔接子的3'-OH之间的反应的非酶促试剂在切口处连接。通过连接衔接子,所述连接在引物延伸产物的一个末端提供通用引发位点。

[0034] 单链连接方法的另一个实例可用于将通用引发位点添加至引物延伸产物的相对末端(或者在延伸引物不包含通用连接位点的实施方案中,添加至延伸产物的两侧)。对于该实施方案,待连接的引物延伸产物的一个或两个末端不具有通用连接位点。此外,在一些实施方案中,待连接的引物延伸产物的至少一个末端具有未知序列(例如,由于随机终止事件或未知序列变异)。在此类实施方案中,采用序列非依赖性的单链连接方法。一种示例性方法描述于美国申请公开号20140193860中。本质上,该方法使用其中单链3'-末端突出端不具有通用连接位点的衔接子群体具有随机序列,例如随机六聚体序列。在该方法的一些实施方案中,所述衔接子也具有发夹结构。另一个实例是由Accel-NGS™ 1S DNA文库试剂盒(Swift Biosciences, Ann Arbor, Mich.)实现的方法。

[0035] 该方法的连接步骤利用连接酶或具有相似活性的另一种酶或非酶试剂。所述连接酶可以是例如病毒或细菌来源的DNA或RNA连接酶,诸如T4或大肠杆菌连接酶,或热稳定连接酶*Afu*、*Taq*、*Tfl*或*Tth*。在一些实施方案中,可以使用替代酶,例如拓扑异构酶。此外,非酶促试剂可用于在引物延伸产物的5'-磷酸酯和衔接子的3'-OH之间形成磷酸二酯键,如US20140193860中所述和参考。

[0036] 任选的引物延伸和平端连接

在该方法的一些实施方案中,所述衔接子的第一连接随后为任选的引物延伸。连接的衔接子具有游离的3'-末端,其可以延伸以产生双链核酸。与衔接子相对的末端则将变得适合于另一衔接子的平端连接。为了避免需要单链连接程序,分子的该双链末端可通过任何连接酶或另一种酶促或非酶促方式连接至双链衔接子。所述双链衔接子序列提供一个或多个通用引发位点(用于扩增或测序)以及任选一个或多个条形码。

[0037] 纯化

在一些实施方案中,本发明的方法包括在连接步骤之后的一个或多个纯化步骤。所述纯化将除去未使用的衔接子分子。通过尺寸排阻方法,例如凝胶电泳、色谱或等速电泳,将

衔接子和大尺寸连接产物与延伸产物分开。

[0038] 在一些实施方案中,纯化是通过亲和结合。在该实施方案的变型中,亲和力针对特定靶序列(序列捕获)。在其他实施方案中,所述衔接子包含亲和和标签。可以使用本领域已知的任何亲和和标签,例如生物素或抗体或其特异性抗体存在的抗原。亲和和标签的亲和配偶体可以存在于溶液中,例如在悬浮的颗粒或珠粒上,或者结合至固体支持物。在亲和纯化的过程中,反应混合物的未结合组分被洗掉。在一些实施方案中,采取额外步骤来除去未使用的衔接子。

[0039] 扩增

在一些实施方案中,本发明包括扩增步骤。该步骤可以涉及线性或指数扩增,例如PCR。用于扩增的引物可以包括存在于所扩增的核酸内的任何序列,并且可以支持一条或两条链的合成。扩增可能是等温的或涉及热循环。

[0040] 在一些实施方案中,所述扩增是指数的并涉及PCR。期望减少PCR扩增偏差。如果使用一种或多种基因特异性引物以降低偏差,则该方法涉及有限数量的扩增循环,例如约10个或更少的循环。在这些实施方案中的其他变型中,使用通用引物来合成两条链。所述通用引物序列可以是一个或两个连接的衔接子的原始延伸引物的一部分。可以使用一个或两个通用引物。上述延伸引物和一个或两个衔接子可以被工程改造成具有相同的引物结合位点。在该实施方案中,可以使用单个通用引物来合成两条链。在其他实施方案中,待扩增的分子的一侧的延伸引物(或衔接子)和另一侧的衔接子含有不同的通用引物结合位点。通用引物可以与另一个通用引物(具有相同或不同的序列)配对。在其他实施方案中,所述通用引物可以与基因特异性引物配对。因为使用通用引物的PCR具有降低的序列偏差,所以扩增循环的数量不需要在与使用基因特异性引物的PCR相同的程度上被限制。使用通用引物的扩增循环数可以是低的,但也可以高达约20、30或更多个循环。

[0041] 条形码

本发明包括使用分子条形码。所述条形码通常由4至36个核苷酸组成。在一些实施方案中,条形码被设计成具有在彼此的10°C或更少内的解链温度。条形码可以设计成形成最低限度的交叉杂交集合,即在期望的反应条件下尽可能少地与彼此形成稳定的杂交体的序列的组合。用于序列鉴定和计数的条形码的设计、放置和使用是本领域已知的。参见例如美国专利号7,393,665、8,168,385、8,481,292、8,685,678和8,722,368。

[0042] 条形码可用于鉴定样品中的每个核酸分子及其子代(即,使用原始核酸分子产生的核酸分子的集合)。此类条形码是“独特的ID”(UID)。

[0043] 条形码也可用于鉴定所分析的核酸分子来源的样品。此类条形码是“多重样品ID”(“MID”)。源自相同样品的所有分子都共有相同的MID。

[0044] 条形码包含每个条形码特征性的核苷酸的独特序列。在一些实施方案中,条形码的序列是预先设计的。在其他实施方案中,所述条形码序列是随机的。条形码内的全部或一些核苷酸可以是随机的。已知序列内的随机序列和随机核苷酸碱基分别被称为“简并序列”和“简并碱基”。在一些实施方案中,分子包含两个或更多个条形码:一个用于分子鉴定(UID),一个用于样品鉴定(MID)。有时,UID或MID每个都包含几个条形码,其当一起使用时可以鉴定分子或样品。

[0045] 在一些实施方案中,反应中UID的数量可以超过待标记的分子的数量。在一些实施

方案中,使用一个或多个条形码来将序列分组或归仓(bin)。例如,在一些实施方案中,使用一个或多个UID来将序列分组或归仓(bin),其中每个仓中的序列含有相同的UID,即是衍生自单个靶分子的扩增子。在一些实施方案中,使用UID来比对序列。在其他实施方案中,所述靶标特异性区域用于比对序列。在本发明的一些实施方案中,在初始引物延伸事件中引入UID,而在连接的衔接子中引入样品条形码(MID)。

[0046] 测序

已进行连接之后,即在步骤4或任选的步骤5(图1)之后,核酸产物可以被测序。可以通过本领域已知的任何方法进行测序。特别有利的是高通量单分子测序。此类技术的实例包括454 Life Sciences GS FLX平台(454 Life Sciences, Branford, Conn.) Illumina HiSeq平台(Illumina, San Diego, Cal.), Ion Torrent平台(Life Technologies, Grand Island, NY),利用SMRT的Pacific BioSciences平台(Pacific Biosciences, Menlo Park, Cal.)以及任何其他目前存在的或未来的单分子测序技术,其涉及或不涉及边测序边合成。在这些实施方案的变型中,所述测序利用存在于一个或两个衔接子序列中或一个或两个引物序列中的通用引物位点。在这些实施方案中的还有的其他变型中,基因特异性引物被用于测序。然而,应注意的是,与基因特异性引物相比,通用引物与降低的测序偏差有关。

[0047] 在一些实施方案中,所述测序步骤涉及序列比对。在一些实施方案中,使用比对来确定来自多个序列(例如具有相同的独特分子ID (UID)的多个序列)的共有序列。在一些实施方案中,使用比对来鉴定序列变异,例如单核苷酸变异(SNV)。在一些实施方案中,从全部具有相同UID的多个序列确定共有序列。在其他实施方案中,使用UID来消除人为结果,即存在于单个分子(特征在于特定UID)的子代中的变异。可以使用UID来消除由PCR误差或测序误差导致的这些人为结果。

[0048] 在一些实施方案中,可以通过定量具有相同多重样品ID (MID)的群体中的具有每个UID的序列的相对数量来定量样品中的每个序列的数量。每个UID代表原始样品中的单个分子,并且计数与每个序列变体相关的不同UID可以确定原始样品中每个序列变体的分数,其中所有分子都共有相同的MID。本领域技术人员将能够确定测定共有序列所必需的序列读数的数目。在一些实施方案中,相关数目是精确定量结果所必需的每个UID的读数(“序列深度”)。在一些实施方案中,期望深度是每个UID 5-50个读数。

[0049] 样品

在本发明的方法中使用的样品包含含有核酸的任何个体(例如人,患者)或环境样品。所述多核苷酸可以从样品中提取,或者样品可以直接进行本发明的方法。起始样品也可以是提取或分离的核酸、DNA或RNA。所述样品可以构成从生物体获得的任何组织或流体。例如,所述样品可以是肿瘤活检样品或血液或血浆样品。在一些实施方案中,所述样品是福尔马林固定、石蜡包埋(FFPE)样品。所述样品可以包含来自一种或多种来源(例如一个或多个患者)的核酸。在一些实施方案中,所述组织可以被病原体感染,并因此含有宿主和病原体的核酸。

[0050] DNA提取的方法是本领域众所周知的。参见J. Sambrook等人,“Molecular Cloning: A Laboratory Manual,”1989,第2版,Cold Spring Harbor Laboratory Press: New York, N.Y.)。各种试剂盒可商售用于从生物样品提取核酸(DNA或RNA)(例如,

BD Biosciences Clontech (Palo Alto, Cal.), Epicentre Technologies (Madison, Wisc.); Genra Systems, Inc. (Minneapolis, Minn.); 和 Qiagen, Inc. (Valencia, Cal.), Ambion, Inc. (Austin, Tex.); BioRad Laboratories (Hercules, Cal.); 等等。

[0051] 在一些实施方案中,本发明的方法中使用的起始样品是文库,例如包含多种多核苷酸的基因组文库或表达文库。在其他实施方案中,通过本发明的方法产生文库。在起始材料是生物样品的情况下,该方法产生扩增文库或代表各种或序列的扩增子集合。文库可以被储存和使用多次,用于进一步扩增或测序文库中的核酸。

实施例

[0052] 实施例1 (预测性)用基因特异性引物的靶标富集和线性扩增

使用DNeasy血液和组织试剂盒(Qiagen, Valencia, Cal.)从人血浆样品中分离核酸。添加基因特异性引物。引物被设计为具有与人EGFR基因的外显子19杂交的基因特异性部分。引物还具有6聚体独特的鉴定序列(UID)和通用连接序列。所述引物在5'-末端被修饰以防止外切核酸酶消化。使所述引物在等温扩增缓冲液(New England Biolabs, Ipswich, Mass., "NEB")中在60°C下杂交20分钟,并添加Bst聚合酶2.0(NEB),一种非热稳定的DNA聚合酶并在65°C下孵育20秒。通过在95°C下加热杀死聚合酶3分钟来终止反应。核酸的模板链用5' ssDNA特异性外切核酸酶RecJF(NEB)和5' dsDNA特异性λ外切核酸酶的组合消化。使用Ampure珠粒纯化(Beckman Coulter, Brea, Cal.)除去未延伸的引物。

[0053] 将由引物延伸产生的单链纯化并添加至连接反应中。添加两种连接衔接子。5'-衔接子被设计为含有通用连接位点,用于扩增的通用引物位点和用于测序的通用引物位点。5'-衔接子还含有多重样品ID(MID)。3'-衔接子被设计为含有用于扩增的通用引物位点和用于测序的通用引物位点。使用来自Accel-NGS™ 1S DNA文库试剂盒(Swift Biosciences, Ann Arbor, Mich.)的试剂进行单链连接。

[0054] 如上所述,通过Ampure纯化将未连接的衔接子与连接产物分开。

[0055] 对于线性扩增,使连接产物与包含对应于3'-末端衔接子中的引物结合位点的单一通用引物的反应混合物接触。在扩增之后,将反应混合物的样品转移至包含通用测序引物的测序反应中。

[0056] 实施例2 (预测性)用简并引物的靶标富集和指数扩增

使用DNeasy血液和组织试剂盒(Qiagen, Valencia, Cal.)从人血浆样品中分离核酸。添加含有简并序列的引物。所述引物被设计为具有6个核苷酸的随机序列,6聚体独特鉴定序列(UID)和通用连接序列。所述引物在5'-末端被修饰以防止外切核酸酶消化。使所述引物在等温扩增缓冲液(NEB)中在60°C下杂交20分钟,并添加Bst聚合酶2.0(NEB),一种非热稳定的DNA聚合酶并在65°C下孵育20秒。通过在95°C下加热杀死聚合酶3分钟来终止反应。核酸的模板链用5' ssDNA特异性外切核酸酶RecJF(NEB)和5' dsDNA特异性λ外切核酸酶的组合消化。使用Ampure珠粒纯化来除去未延伸的引物,如实施例1所述。

[0057] 将由引物延伸产生的单链纯化并添加至连接反应中。添加两种连接衔接子。5'-衔接子被设计为含有通用连接位点,用于扩增的通用引物位点和用于测序的通用引物位点。5'-衔接子还含有多重样品ID(MID)。3'-衔接子被设计为含有用于扩增的通用引物位点和用于测序的通用引物位点。基本上如公开US20140193860中所述进行单链连接。

[0058] 使用Ampure珠粒纯化来除去未连接的衔接子,如实施例1所述。

[0059] 对于指数扩增,将连接产物与包含一对通用扩增引物的PCR反应混合物接触。在扩增之后,将反应混合物的样品转移至包含通用测序引物的测序反应中。

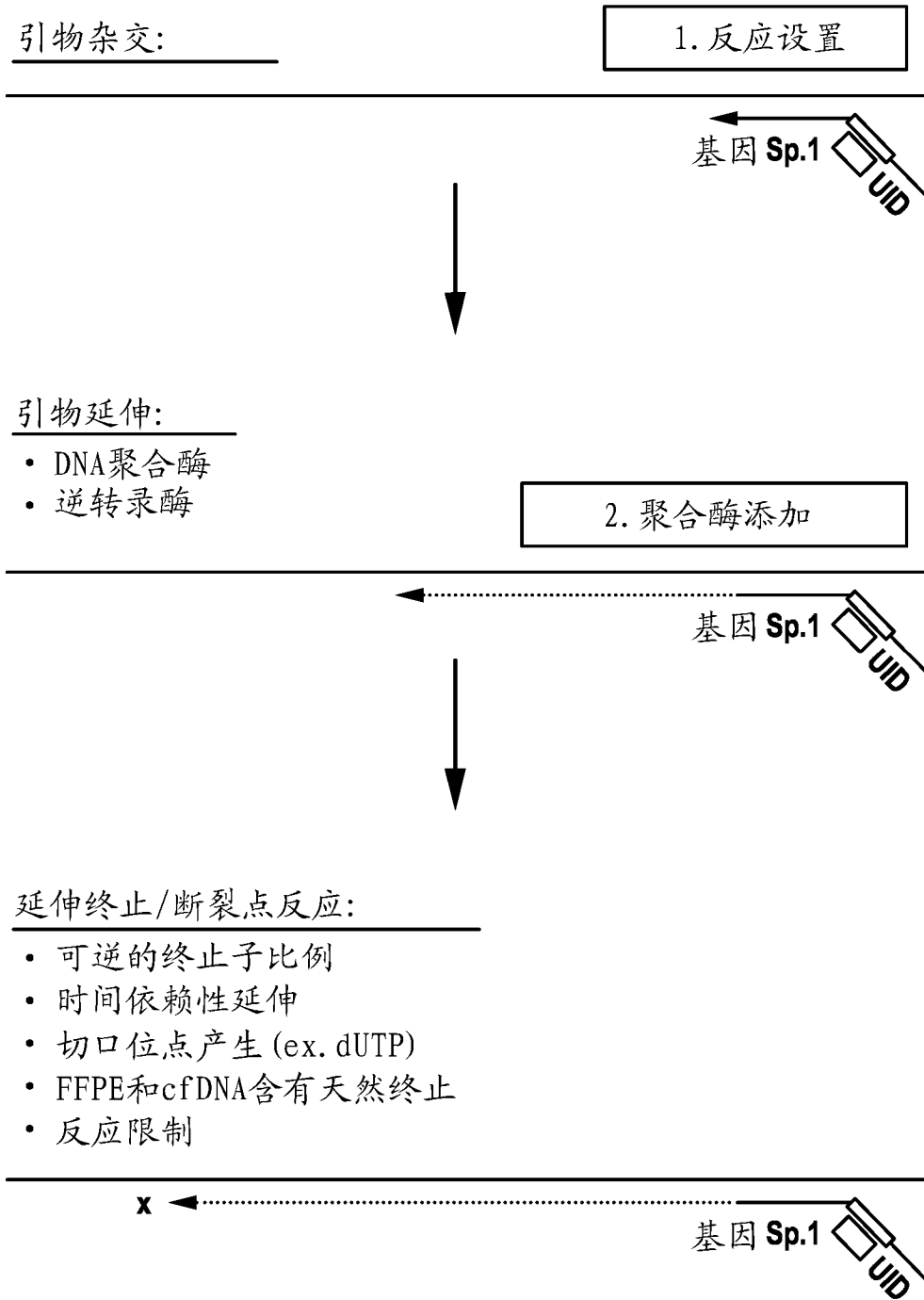
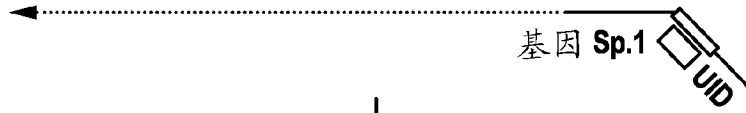


图 1

未延伸的DNA的去除:

- 如果原始引物具有5' 封闭物, 则5' Exo
- 如果引物受保护, 则ssDNA endo ssDNA将保留 (未显示)

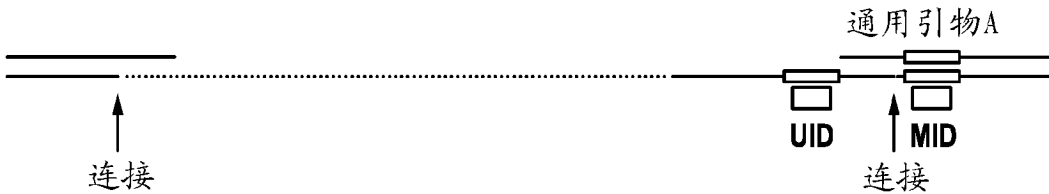
3. 外切核酸酶添加



3' ssDNA 连接:

- 转换Bio 1S试剂盒
- Penn State发夹

4. 连接酶和衔接子添加



LM PCR:

5. PCR试剂盒热循环



图 1续