(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107208150 A
(43)申请公布日 2017.09.26

(21)申请号 201580074738.5

(22)申请日 2015.12.22

(30)优先权数据
1422982.7 2014.12.22 GB

(85)PCT国际申请进入国家阶段日
2017.07.27

(86)PCT国际申请的申请数据
PCT/GB2015/054125 2015.12.22

(87)PCT国际申请的公布数据
WO2016/102956 EN 2016.06.30

(71)申请人 DNAe集团控股有限公司
地址 英国伦敦

(72)发明人 布莱恩·詹姆斯·麦考恩
卡塔尔·约瑟夫·麦克尔冈

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
代理人 王思琪 郑霞

(51)Int.Cl.
C12Q 1/68(2006.01)

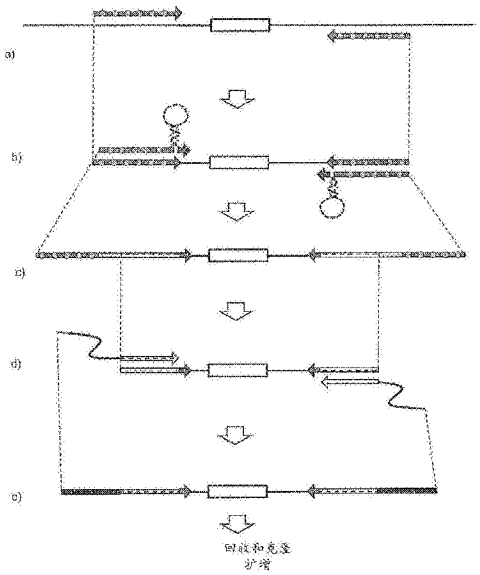
权利要求书3页 说明书8页 附图2页

(54)发明名称

泡状引物

(57)摘要

描述了用于产生核苷酸序列的序列即用片段的方法，该方法利用“泡状引物”，所述泡状引物包括与靶杂交的第一部分和第三部分以及形成未杂交环的部分地自身互补的第二部分。该环包含允许使用测序引物的通用序列。第一部分可以是可降解的，以便产生具有两侧是第三部分和第二部分的通用序列的感兴趣序列的扩增子。在优选的实施方案中，第二部分或第二部分和第三部分之间的区域还包括核苷酸A、C、G、T的四联体(tetrad)，允许校准测序反应。



CN 107208150 A

1.一种用于从起始模板多核苷酸生成多核苷酸片段的方法,所述方法包括:

a)使用第一引物对从所述起始模板扩增感兴趣的区域以形成掺入所述感兴趣的区域的扩增子,

b)用使用第二引物的核酸扩增反应从步骤a)中生成的第一扩增子扩增所述感兴趣的区域,以形成掺入所述第二引物的扩增子,

其中所述第二引物包括具有第一部分、第二部分和第三部分的核酸序列,所述第一部分与所述起始模板的第一部分互补,所述第二部分与所述起始模板不互补,且所述第三部分与所述起始模板的第二部分互补;

其中所述起始模板的所述第一和第二部分相邻或彼此紧密靠近;

其中所述第二引物的所述第一、第二和第三部分以从5'到3'的顺序布置,使得在与所述起始模板杂交时,所述引物的所述第二部分保持不杂交并且在所述第一和第三部分之间形成环;

从而生成包括两侧为所述第二引物的序列的感兴趣的区域的扩增产物。

2.如权利要求1所述的方法,其中步骤b)的扩增反应用第二引物对进行,所述第二引物对的每一个引物是所述第二引物的形式。

3.如权利要求1或权利要求2所述的方法,其中所述第二引物的所述第二部分包括通用序列。

4.如权利要求3所述的方法,其中所述通用序列包括测序引物序列。

5.如权利要求3或权利要求4所述的方法,其中所述通用序列与所述第二引物的所述第三部分相邻。

6.如权利要求5所述的方法,其中所述通用序列通过特定的碱基序列与所述第三部分隔开。

7.如权利要求6所述的方法,其中所述通用序列通过以任何特定的顺序包括四种核苷酸碱基A、T、G和C中的每一个的序列与所述第三部分隔开。

8.如任一前述权利要求所述的方法,其中所述第二引物或每个第二引物的所述第一部分的至少一部分是降解易感的,而所述引物的至少所述第三部分和所述第二部分的至少一部分不是降解易感的;并且所述方法还包括步骤

c)使所述引物或每个引物的易感部分从所述扩增子降解下来。

9.如任一前述权利要求所述的方法,还包括步骤:

d)用第三引物对扩增b)的产物和/或c)的产物,每个引物包括与所述第二引物或每个第二引物的所述第二部分的至少一部分基本上相同的核酸序列。

10.如权利要求9所述的方法,当从属于权利要求3至7中任一项时,其中b)的产物和/或c)的产物被扩增,并且所述第三引物的核酸序列的至少一部分与所述第二引物或每个第二引物的所述第二部分的所述通用序列基本上相同。

11.如任一前述权利要求所述的方法,其中所述模板是基因组的片段。

12.如权利要求11所述的方法,其中所述模板是基因组的基因座。

13.如权利要求2所述的方法,其中所述对中的每个第二引物的所述第二部分是不同的。

14.如任一前述权利要求所述的方法,其中所述模板的所述第一和第二部分被0-20个

核苷酸隔开，优选地被1-10个核苷酸、更优选地被1-6个核苷酸且最优选地被1、2、3、4、5或6个核苷酸隔开。

15.如任一前述权利要求所述的方法，其中所述第二引物的所述第一部分的长度为最多15、20、25、30、35、50个核苷酸，优选20-35个核苷酸，更优选25个核苷酸。

16.如任一前述权利要求所述的方法，其中所述第二引物的所述第二部分包括自身互补区域，使得在杂交时形成的环采取茎环结构，其中所述自身互补区域形成茎。

17.如权利要求8所述的方法，其中所述第二引物的所述第二部分包括第一可降解部分和第二耐受部分。

18.如任一前述权利要求所述的方法，其中所述第二引物的所述第三部分的长度不超过2、4、5、6、7、8、9或10个核苷酸，优选4至6个核苷酸，更优选6个核苷酸。

19.如任一前述权利要求所述的方法，其中所述第二引物的所述第二部分或所述第二和第三部分一起被选择为包括包含全部四种核苷酸碱基(A、C、G、T)的核苷酸四联体。

20.如权利要求9所述的方法，其中步骤d)中的所述第三引物对在5’末端还包括另外的非模板序列。

21.如任一前述权利要求所述的方法，其中步骤b)的扩增是巢式PCR。

22.如权利要求3至21中任一项所述的方法，其中测序引物与所述第二引物的所述第二部分的所述通用序列的互补物杂交。

23.如任一前述权利要求所述的方法，还包括测序所产生的扩增产物的步骤。

24.如任一前述权利要求所述的方法，其中步骤a)和/或步骤b)的扩增是多重扩增。

25.用于核酸扩增的引物，所述引物包括具有第一部分、第二部分和第三部分的核酸序列，所述第一部分与用于扩增的靶序列的第一部分互补，所述第二部分与所述靶序列不互补并且包含通用序列，且所述第三部分与所述靶序列的第二部分互补；

其中所述靶序列的所述第一和第二部分相邻或彼此紧密靠近；

其中所述引物的所述第一、第二和第三部分以从5’到3’的顺序布置，使得在与靶序列杂交时，所述引物的所述第二部分保持不杂交并且在所述第一和第三部分之间形成环。

26.如权利要求25所述的引物，其中所述通用序列的互补物能够与测序引物杂交。

27.如权利要求25或权利要求26所述的引物，其中所述通用序列与所述第三部分相邻。

28.如权利要求25至27中任一项所述的引物，其中所述引物的所述第一部分的长度为最多15、20、25、30、35、50个核苷酸，优选20-35个核苷酸，更优选25个核苷酸。

29.如权利要求25至28中任一项所述的引物，其中所述引物的所述第二部分包括自身互补区域，使得在杂交时形成的环采取茎环结构，其中所述自身互补区域形成茎。

30.如权利要求25至29中任一项所述的引物，其中所述引物的所述第三部分的长度不超过2、4、5、6、7、8、9或10个核苷酸，优选4至6个核苷酸，最优选6个核苷酸。

31.如权利要求25至30中任一项所述的引物，其中所述引物的所述第二部分或所述第二和第三部分一起被选择为包括包含四种核苷酸碱基(A、C、G、T)中的每一个的核苷酸序列。

32.一对根据权利要求25至31中任一项所述的引物，其中所述对中的每个引物的所述第二部分是不同的。

33.如权利要求32所述的引物对，与第二引物对结合，所述第二引物对的每个成员包括

与所述第一引物对的对应成员的至少一部分互补的核酸序列。

34.包括如权利要求33所述的多个引物对的引物对文库，每个对具有包含各自的第一和第二第二部分的第一和第二引物，其中每个第一第二部分相同，并且每个第二第二部分相同。

泡状引物

发明领域

[0001] 本发明涉及从起始模板生成可用于DNA测序分析的多核苷酸片段的方法。这些片段在下一代测序方法中有用。本发明的多个方面涉及在该方法中使用的核酸引物。

[0002] 发明背景

[0003] 自从完成人类基因组序列草案以来，DNA测序分析的生物化学和仪器已经进展到这样的点：对于原始基因组消耗的相同财务支出，现在(2014)对大都市芝加哥(人口270万)中每个男人、女人和儿童以每设备每29小时一个完整基因组的速度产生基因组序列是可能的，每个基因组的每个和每一个区域被30倍覆盖。这种惊人的能力增加是由以通用的方式处理所有被测序的DNA片段的能力造成的，基因组的每个部分被同时暴露于相同的生物化学。“大规模平行”DNA测序通过基因组DNA的随机片段化，然后将人工序列“连接子(adapter)”酶促连接到片段化的DNA小片的每一端来实现。

[0004] 从基因组DNA生成测序“库”是耗时的，而且由于两个不同“风格”的连接子的连接的随机性质，产生其中约一半的模板事实上不可用于分析的片段：许多产物将具有相同的“A型”或“B型”连接子连接在片段的末端，而所需要的是偶然发生的不对称连接，一个风格的连接子(A型)在一个末端而另一个风格的连接子(B型)在另一个末端。一旦开始测序反应，这些不对称产物能够从基因组模板被克隆扩增，并理想地用于生成有价值的序列信息。

[0005] 目前用于NGS的测序文库的制备技术包括以下步骤：

[0006] • 模板DNA的随机片段化

[0007] • 具有所需长度的那些片段的尺寸选择

[0008] • 对片段末端的酶促“末端修复”以允许A型和B型连接子的平端连接

[0009] • 连接子的连接，产生一部分“A/A”和“B/B”冗余产物以及“A/B”所需产物群

[0010] • 连接子修饰的文库片段的克隆扩增。

[0011] 随着基因组重新测序成本降低和测序的速度增加，NGS的应用越来越多地转向临床。然而，很少存在这样的情况，其中读取基因组的全部320万个碱基是有重大关系的；可能的是更有针对性的方法是有实用性的，通过研究与具体状况相关(或可能确认具体状况)的有限数量的遗传位置指导治疗。如果不必读取基因组的全部碱基，那么伴随的是，应用仅为了实现读取基因组的全部碱基的技术和方法学可能不是最佳的。

[0012] 具体区域的靶向测序最有效地要求从可能极度多样化和复杂的主体模板中分离那些序列。有效地，这可以通过将靶区域扩增至它们的数量超过其它未扩增区域的数量的水平来实现。此类扩增产物将可用于NGS末端连接子的那种连接(如以上所述的)，但这些将再次成为混合群体，在该混合群体中很大比例的序列将是不适合支持NGS测序的克隆扩增的“A/A”和“B/B”形式。关键的是，甚至是那些A/B型的构建体也会具有存在于插入在连接子序列和目标是用于测序的基因组DNA的区域之间的大的“引物残留物”区域。如果连接子序列提供测序引物的结合位点，那么这些引物残留物将是产生的第一数据：不必要且不能提供信息的。已知序列的大量区域必须在到达感兴趣的未知序列之前通过NGS法处理。这不仅使用不必要的资源，而且通常测序方法在接近读段起始处最准确，导致之后的未知序列以

较低的保真性被读取。能够产生具有较短的已知区域的序列片段，以降低所进行的不必要的测序的量并提高生成的感兴趣的未知序列的保真性将是期望的。

[0013] 基于连接的连接子策略的另一个缺点是需要获得连接子的不对称掺入(即序列片段的每个末端上的连接子不同)。然而，连接反应通常是非定向的，使得只有一部分片段将包括必需的不对称连接子；其他的将在任一端包括相同的连接子。提供一种允许固有地不对称掺入连接子的方法将是期望的。

[0014] 发明概述

[0015] 根据本发明的第一方面，提供了用于从起始模板多核苷酸生成多核苷酸片段的方法，所述方法包括：

[0016] a)使用第一引物对从所述起始模板扩增感兴趣的区域以形成掺入所述感兴趣的区域的扩增子，

[0017] b)用使用第二引物的核酸扩增反应从步骤a)中生成的第一扩增子扩增感兴趣的区域，以形成掺入第二引物的扩增子，

[0018] 其中所述第二引物包括具有第一部分、第二部分和第三部分的核酸序列，所述第一部分与所述起始模板的第一部分互补，所述第二部分与所述起始模板不互补，所述第三部分与所述起始模板的第二部分互补；

[0019] 其中所述起始模板的所述第一和第二部分相邻或彼此紧密靠近；

[0020] 其中所述第二引物的所述第一、第二和第三部分以从5’到3’的顺序布置，使得在与所述起始模板杂交时，所述引物的所述第二部分保持不杂交并且在所述第一和第三部分之间形成环；

[0021] 从而生成包括两侧为所述第二引物的序列的感兴趣的区域的扩增产物。

[0022] 由此生成的扩增产物包括第二引物的部分，并且所述方法因此可用于生成掺入已知序列的扩增子，所述已知序列可用于测序反应(即产物是“测序即用的”)。

[0023] 优选地，步骤b)的扩增反应用第二引物对进行，所述第二引物对的每一个是所述第二引物的形式。以此方式，扩增产物在每个末端包括引物序列。

[0024] 第二引物的第二部分可以包括通用序列，例如，与测序引物序列至少基本上相同的序列。此通用序列可以是许多或所有可能的第二引物共有的，从而允许扩增子被用于测序反应。

[0025] 通用序列还可以与包括四种核苷酸碱基A、C、G和T的每一个的序列相邻。这可以是四种核苷酸的简单四联体(tetrad)(例如ATCG)或可以包括两个、三个或更多个每个碱基(例如，AACCTTGG)。核苷酸可以是任何顺序。该序列可以将通用序列与引物的第三部分隔开。

[0026] 优选地，所述第二引物或每个第二引物的第一和第二部分的至少一部分是降解易感的，而该引物的至少第三部分和第二部分的至少一部分是降解不易感的；并且所述方法还包括步骤：

[0027] c)将所述引物或每个引物的易感部分从所述扩增子降解下来。

[0028] 这去除了扩增子中的已知序列的区域并且防止残留在扩增子中的第二引物的第二部分再形成茎(当存在茎环结构时)。该茎的再形成否则可以阻止接近另外的引物或引物对的3’末端处的该引物或引物对预期的引物结合位点。

[0029] 所述方法还可包括以下步骤：

[0030] d)用第三引物或引物对扩增b)的产物和/或c)的产物，该第三引物或每个第三引物包括与所述第二引物或每个第二引物的至少一部分基本上相同(并且优选地相同)的3’核酸序列。

[0031] 优选地，第三引物的基本上相同的核酸序列与所述第二引物或每个第二引物的未降解的非易感部分基本上相同，从而生成包括感兴趣的区域和所述第二引物或每个第二引物的未降解部分的扩增产物。该基本上相同的部分可以与包含于第二引物的第二部分内的通用序列基本上相同。

[0032] 此方法解决了现有技术方法固有的困难。特别地，第二引物或引物对(因为在杂交时形成环或泡而被称为“泡状引物”)可以掺入两个具有已知但变化的序列的区域(与靶互补，并因此根据靶变化)和不与靶互补的固定序列的区域。固定序列可以被用于向扩增区域引入测序连接子或其他有用的序列而不需要连接反应。固定序列可以是人工序列，或衍生自另一生物体的不与模板互补的序列。在优选的实施方案中，固定序列可以包括通用序列，例如测序引物序列。

[0033] 当一个部分或序列被描述为与靶或另一序列“不基本上相同”，优选地该部分或序列与该靶或其他序列不相似，使得该部分或序列在扩增反应中使用的条件下不与和靶互补的序列杂交。相似地，当一个序列与靶“不互补”时，其足够地不相似到使得其在扩增反应中使用的条件下不与靶杂交。

[0034] “对降解易感”的部分在本文中还可以被称为“可降解的部分”，而对所述降解不易感的部分在本文中可以被称为“耐受部分”。这些术语被互换使用。

[0035] 模板可以是基因组多核苷酸。模板可以是真核的、原核的或古细菌的。可以提供一个或更多个模板。模板可以代表基因组的片段；例如，单个染色体或单个基因组的基因座(例如，用于等位基因多态性的快速测序)。

[0036] 当存在第二引物对时(即由引物A和B组成)，引物A的第一和第三部分将与引物B的那些部分不同，而第二部分可以不同或可以相同，但优选地不同。在不同的第二引物对中(即引物A和B；和引物A’和B’)，对应引物的第二部分(A和A’；B和B’)将相同，但尽管如此也可以在每个对内不同。第一和第三部分赋予靶特异性并且允许引物的不对称掺入。

[0037] 另外，第二引物(或引物对)的第一和第三部分的用途允许生成泡状部分，使得第一和第三部分紧密接近，并且引物保持对靶的高度特异性以降低非特异性杂交和扩增的几率。

[0038] 优选地，模板的第一和第二部分被0-20个核苷酸隔开，优选地被1-10个核苷酸、更优选地被1-6个核苷酸且最优选地被1、2、3、4、5或6个核苷酸隔开。

[0039] 第二引物(或引物对)的第一部分的长度可以是最多15、20、25、30、35、50个核苷酸，优选20-35个核苷酸，更优选25个核苷酸。

[0040] 第二引物(或引物对)的第二部分可以包括第一可降解部分和第二耐受部分。第一可降解部分优选地与引物的第一部分相邻，并且第二耐受部分与引物的第三部分相邻。

[0041] 第二引物(或引物对)的第二部分优选地包括自身互补区域，使得在杂交时形成的环采取茎环结构，其中自身互补区域形成茎。茎的形成使引物的第一和第三部分靠拢，迫使第三部分与其互补序列亲密接触，如果作为模板DNA的第二部分存在的话。环的长度可以最

小(通常需要约4个核苷酸以形成环),但优选地第二区域还包括形成较大环的非自身互补区域。当第二部分包括可降解部分和耐受部分时,可降解部分优选地形成茎的一半,耐受部分形成茎的另一半加环。

[0042] 第二引物(或引物对)的第三部分的长度优选地不超过2、4、5、6、7、8、9或10个核苷酸。优选的尺寸不超过6个核苷酸、并且优选地4至6个核苷酸、最优选地6个核苷酸。认为此长度(与第一部分一起)为引物提供了足够的特异性,同时降低必须在随后的测序反应中测序的不提供信息的核苷酸的总长度。

[0043] 优选地,第二引物(或引物对)的第二部分或第二和第三部分一起被选择为包括包含全部四种核苷酸碱基(A、C、G、T)的核苷酸四联体。核苷酸的顺序不重要。这允许通过提供在待被测序的区域的开始处具有全部四种核苷酸的已知序列校正测序反应。四联体可以隔开第二和第三部分。当第二部分包括可降解部分和耐受部分时,则四联体可以存在于耐受部分中,或存在于耐受部分连同第三部分中。核苷酸的四联体优选地紧邻测序引物序列的3'末端定位。可以包括超过四个核苷酸,条件是每种核苷酸以已知的数目存在(例如序列可以是AAGGCCTT)。

[0044] 第二引物对的可降解部分可以包括RNA,而耐受部分包括DNA。可选择地,可降解部分可以包括DNA,其中胸腺嘧啶被尿嘧啶代替。可降解部分可以因此被RNA酶H或碱性热解(对于RNA),或被尿嘧啶-N-糖基化酶(对于包含U的DNA)降解,正常DNA耐受其每一个。

[0045] 步骤d)中的第三引物对还可以在5'末端包括另外的非模板序列;这允许在扩增子中掺入另外的功能序列。例如,另外的序列可以包括可选择标志物、用于纯化或检测的标签、用于物理捕获扩增子的部分、克隆扩增序列等。

[0046] 第一引物对可被选择成使得每个引物的3'末端是对应于第二引物或引物对的每个相应引物的第三部分的区域的5'朝向部分。即,第二引物(也被称为"泡状引物")是巢式的,且扩增是巢式PCR。

[0047] 本发明的另外的方面提供了引物,该引物包括具有第一部分、第二部分和第三部分的核酸序列,所述第一部分与用于扩增的起始模板的第一部分互补,所述第二部分与所述起始模板不互补,且所述第三部分与所述起始模板的第二部分互补;

[0048] 其中所述起始模板的所述第一和第二部分彼此相邻或紧密靠近;

[0049] 其中所述引物的所述第一、第二和第三部分以从5'到3'的顺序布置,使得在与所述起始模板杂交时,所述引物的所述第二部分保持不杂交并且在第一和第三部分之间形成环。

[0050] 此处起始模板指将通过此引物扩增的序列,并且该序列在上文定义的方法中将最初被常规引物对扩增以生成扩增子。

[0051] 优选地,所述每个引物的第一和第二部分的至少一部分也是对降解易感的,而该引物的至少第三部分和第二部分的至少一部分不是易感的。

[0052] 还提供了包含上文所述的一对引物的引物对。

[0053] 尽管PCR可能是最广泛使用的扩增方法并且在此作为举例使用,但是可以设想其他非热循环扩增方法。

[0054] 本发明的仍另外的方面提供了如本文所述的引物对的文库,所述文库包含多个引物对,每对具有第一和第二引物,包含相应的第一个和第二第二部分,其中每个第一第二部

分相同，并且每个第二第二部分相同。第一和第三部分可以在引物对之间不同(并且将在引物对的每个引物中不同)。

[0055] 附图简述

[0056] 图1示出了用于在本文所述的方法中使用的引物的示意图。

[0057] 图2展示了用于产生测序即用的多核苷酸片段的方法。

[0058] 发明详述

[0059] 本文公开的方法能够产生NGS(下一代测序)“测序即用”的DNA片段，所述DNA片段是存在于原始模板DNA样品中的总DNA的目标子集。只有那些感兴趣的基因座被例如聚合酶链式反应扩增，使得所产生的扩增子具有两侧是已知序列的末端的感兴趣的模板DNA。这些已知序列在所产生的所有扩增子上相同或基本上相同，并且是有意地和可控地不对称的，具有两个不同序列应用于扩增片段的两个末端的每一个。因此，产生的扩增子功能上等同于常规NGS法中产生的连接子连接的片段，但提供在制备的容易性、时间和成本方面以及随后产生的测序数据的质量方面的明显优势。在随后的NGS操作诸如克隆扩增和DNA测序期间，这些扩增子的末端可适用于通用的“一体适用的”生物化学程序。

[0060] 此外，该方法的实施方案能够使位点特异性引物的3’末端(发明内容中的“第三部分”)相对短以紧密靠近地杂交同一引物上大得多的、稳定杂交的5’元件(发明内容中的“第一部分”)，这两个靶互补区域由非模板序列(发明内容中的第二部分)隔开，所述非模板部分经成功的引物延伸将成为子代扩增子的一部分。非模板序列将掺入在下一代测序中使用的序列，使得测序反应可以从那个点开始。这使得从直接的“连接子连接”策略将不可避免地浪费地生成的已知DNA序列数据的量最小化，避免测序通过大量的“不感兴趣的区域”扩增-引物残留物。

[0061] 此外，该方法的实施方案能够使用NGS用于靶向分析来自复杂DNA模板源内的特定遗传基因座。有效的靶向目标基因测序(targeted-panel sequencing)是可能的(例如来自一个或更多个特定遗传基因座)，而不是当前的大规模平行“全基因组鸟枪测序”。

[0062] 用于在本方法中使用的引物的说明在图1中示出。引物10包括第一部分12、第二部分14和第三部分16。第一部分12被设计为与待被扩增的靶基因组序列的一部分互补，而第三部分16也被设计为与该靶序列的相邻部分互补。第一部分为约25个核苷酸的长度，第三部分为约6个核苷酸。在靶上在与第一部分互补的序列和与第三部分互补的那些序列之间可以存在0-4个核苷酸的缺口。此缺口是为了当第一部分12和第三部分16与靶链杂交时，接纳引物的不互补的第二部分14(茎环结构)。

[0063] 第二部分14不与靶互补，并且包括自身互补区域，使得该序列形成茎环发夹结构。第二部分的环部分和茎部分包括与所选的测序反应中使用的测序引物基本上相同的序列。注意，使用的特定测序化学很大程度上不重要；本文描述的方法是普遍适用的，并且被预期能够在扩增子中掺入相关的测序引物。在某些实施方案中，第二部分可以另外包括或相邻于以任意顺序包含四种核苷酸A、C、G、T中的每一个的序列。优选地，该序列是四联体(例如ACGT)，尽管该序列可以包括每个核苷酸的多个拷贝，但通常(但不必须)以相等的数目(例如AAGGCCTT)。

[0064] 引物10可以包括两个类型的核酸。在引物的5’末端的第一区域可以是对通过选定技术的降解敏感，而在引物的3’末端的第二区域对该技术的降解不敏感。例如，引物的5’末

端可以从RNA形成，而3’末端从DNA形成；RNA部分可以被RNA酶H或碱性热解降解，而DNA部分是耐受的。可选择地，引物的5’末端可以由掺入尿嘧啶代替胸腺嘧啶的DNA形成；这将可被尿嘧啶-N-糖基化酶降解。在优选的实施方案中，可降解部分可被酶降解。

[0065] 在此实例中，可降解部分包括引物的第一部分12的全部和第二部分14的第一段(在第二部分上以双虚线示出)。引物的剩余部分是不可降解的。第二部分的可降解段包括形成茎环结构的茎的一半的那个区域；不可降解部分(在以单虚线示出)形成环和与第三部分相邻的茎的另一半。不可降解部分包括与测序引物的序列至少基本上相同的序列。测序引物与此序列的互补物杂交，所述互补物在经DNA聚合(通常是克隆扩增)生成另一条链时产生。

[0066] 引物10可以成对使用，由正向引物和反向引物构成。正向引物和反向引物包括不同的第一部分和第三部分(因为这些被选择为与待被扩增的模板的区域的端点互补)和不同的第二部分(导致不同的正向和反向测序引物被使用)，因为目标是允许将第二部分不对称地掺入到扩增子中。然而，当提供多个引物对时，每对的第二部分可以是相同的，以允许使用公用测序引物测序所有扩增子。

[0067] 使用引物生成扩增子的方法在图2中示出。此图详述了所执行的顺序性步骤，以生成用于测序反应的通用模板，在该测序反应中最小量的残留引物序列将被询问。

[0068] 该方法允许将多个单独的模板靶快速转化为可用于通用测序流程的产物，并且(最终)具有高的选择性和特异性。顺序性扩增步骤可以在单独的扩增室中分开进行，扩增室物理上分离引物的种类，但本领域技术人员将理解，通过选择引物结合温度、仔细地控制引物浓度(使得某些引物种类被消耗至耗尽)和通过应用时间上从参与整个过程的参与阶段中区分单独的阶段的特异性热循环方案，可以在更小数目的室(理想地只有一个)中进行这些反应。

[0069] 在第一步中[图2a]，使用常规寡核苷酸引物进行标准PCR反应，用扩增的靶富集模板群体。此反应可以有利地多重进行，不同的引物对实现若干靶的相对低的特异性多重扩增，以保证稀有种类被有效地扩增。此初始阶段中的低特异性引物还可以接受靶引物结合位点内的一定程度的非互补碱基配对，如可能在来自例如癌症相关的基因的靶DNA中及其周围遇到的。此初始扩增阶段2a可以为了增加的灵敏度而牺牲特异性；此初始扩增阶段2a是容忍性的，因为任何不当地扩增的物质，包括引物二聚体人工产物将在随后的阶段期间被从进一步的扩增中清除。此步骤生成两侧是引物序列的第一扩增子。应注意，这些引物本身可以是可降解的(例如，由RNA形成或由掺入U代替T的DNA)。这些引物可以被这样设计，使得它们将在变得无效前通过以下的一个或组合产生有限量的扩增子：

[0070] • 高Tm，之后的循环在较低退火温度进行；

[0071] • 低初始浓度的该引物。

[0072] 随后，在步骤b)中，使用上文所述的“泡状引物”或环状引物扩增来自步骤a)的扩增子。在图2b中，新的泡状引物利用在第一步骤2a期间生成的模板的富集池，并高效地仅增殖来自2a的、从正确的靶生成的扩增子，矫正初始扩增可能具有相对低的特异性。因此，此扩增生成利用初始低特异性扩增的高灵敏性的扩增子池(图2a)，但因为泡状引物的3’末端将仅由正确的扩增子接受，所以在此第二阶段重新建立高特异性(图2b)。只有包含泡状引物的“泡序列”的扩增子由现在具有高灵敏度(2a)和高特异性(2b)(组合地)的反应生成。生

成的任何其他脱靶扩增子或人工产物将无法被采用而前行通过该反应方案，因为它们将缺乏在泡状引物的非模板(人工)泡中界定的必需通用序列。

[0073] 泡状引物的序列被这样选择，使得扩增对于步骤a)中的扩增是巢式的；即，泡状引物的第一部分与步骤a)的引物基本上相同，而第三部分是步骤a)的引物的3'末端的3'朝向部分。这意味着，第三部分包含未被呈现在步骤a)的引物中的序列，并且仅允许在初始扩增期间正确生成的那些扩增子的选择性“巢式”PCR，因此该“巢式”PCR可以接受一定程度的降低的特异性。第二和/或第三部分的序列也被理想地选择为使得它包含包括四种核苷酸(A、C、G、T)的每一个的四联体。核苷酸的四联体优选地紧邻与测序引物的区域的3'末端至少基本上相同的区域定位。引物还可以在茎环结构的茎中包括“索引密码子(Index Code)”；例如以鉴定和标记产物。作为一个实例，索引密码子可以被用于鉴定来自特定的单个模板的特定产物。可选择地或另外地，如果被测序，泡状引物的第三部分的六个碱基将正常地足以鉴定在合理大小的多重反应中正被测序的特定靶。

[0074] 步骤c)示出了在步骤b)中生成的扩增子。扩增产物具有紧密靠近靶DNA序列地呈现的非模板序列(即泡状引物的第二部分的序列)。此产物可以具有源自初始靶特异性PCR结合位点的可降解序列(例如RNA)和非模板环的含RNA的残留物。

[0075] 步骤d)中，步骤c)的产物可以被降解(例如通过使用RNA酶H和/或RNA酶A)，以从扩增子去除可降解序列，如果存在的话。此降解还去除了未掺入扩增子的任何过多的可降解引物，功能上去除这些引物的任何另外的活性。因此，剩余的扩增子仅包括扩增的靶序列，该扩增的靶序列掺入了来自引物的第二部分和第三部分的不可降解的非靶序列。任选地，在此阶段，通用PCR扩增还可以用靶向泡状引物的非靶序列的引物(在上文的“发明内容”部分被称作第三引物对)进行。这些另外的引物可以另外携带用作序列捕获标签的非模板人工5'延伸物，序列捕获标签是用于克隆扩增或用于产物的前置扩增后的区域。

[0076] 无论生成的扩增子的5'易感末端是否被消化掉，扩增方案的下一阶段依赖使用与泡状引物的非模板(人工)序列至少基本上相同的引物扩增靶扩增子。在多重反应中生成的所有扩增子适于使用此引物以通用方式扩增，该引物与提供在泡状引物的非模板区域中的人工序列至少基本上相同。此通用引物充当扩增引物，而具有相同或基本上相同的序列的引物可以在测序反应期间被用作最终“测序引物”，该测序引物的3'末端(一般地)靠近待被询问的靶扩增子的区域定位，仅被少数几个靶特异性碱基(根据此模板限定区域的GC含量，理想地在4和10个碱基之间的数量，6个碱基、7个碱基或8个碱基是最理想的)隔开。通用测序引物的3'末端和靶特异性碱基之间的区域被设计或选择为包括核苷酸A、T、G和C的四联体，充当从这些单个碱基掺入事件的每一个生成的信号水平的引物。此核苷酸四联体可以被提供为由每个核苷酸种类的多核苷酸表示(例如AA、TT、GG和CC，或者AAA、TTT、GGG或CCC)。四联体引物内碱基呈现的顺序不重要，且每个碱基的呈现数目可以变化(例如AA、TTT、GG、CCC)。

[0077] 步骤e)示出了最终产物。这包括任选地两侧是可用于捕获/克隆扩增的序列(在步骤d)的扩增中引入)的靶序列；可用于杂交通用测序引物的区域(源自泡状引物的第二部分内)和包含A、T、G和C充当测序期间掺入的每个碱基所生成的信号强度的参考的区域(源自泡状引物的第三部分内，或在第二和第三部分之间)。然后终产物可以被回收并在测序反应中使用。

[0078] 使用与泡状引物的非模板序列至少基本上相同的引物通用扩增靶序列能够获益于包括通用5'标签尾部延伸物，5'标签尾部延伸物可以被用于捕获多重扩增子池中的单个分子并促进这些单个分子(再次)以通用方式克隆扩增。本领域普通技术人员将认识到，对基于人工序列的扩增的依赖给出了靶特异性或整体优化这些扩增的巨大空间，并且整体方案将产生适于NGS技术不可知论的测序的扩增子群。

[0079] 与连接子连接策略实现的～50％随机对称产物相比，本文描述的方法提供了在末端连接了一致的(可靠的不对称的)连接子序列的“末端修饰片段”池：对称产物不适于支持用于NGS测序的克隆扩增，并且因此本发明有效地消除了在NGS中有用的可用模板的减少。

[0080] 本方法能够快速生成DNA短片段的池，其中片段的内部是待由NGS确定的感兴趣的DNA序列，而片段的末端是基本上通用的，允许在生成信号增强所需的克隆群体期间的平行处理。

[0081] 在一种实施方案中，本方法使用利用尿嘧啶碱基代替胸腺嘧啶碱基的引物设计，能够功能上去除这些序列以有利于高效产生期望的产物。在另一个实施方案中，本发明使用在引物的5'末端是RNA而在引物的3'末端是DNA的杂合体的引物设计，使得当与DNA杂交时能够消化RNA组分并功能上去除该组分。

[0082] 泡状引物的3'末端，第三部分，包括有限数量的足以接纳DNA聚合酶的附着和延伸的模板特异性碱基，但限制将在用于NGS反应的终产物中“浪费地”出现并被测序的碱基的数量。

[0083] 本文描述的方法和引物相对于现有技术具有大量优势。在一些实施方案中，DNA序列连接到DNA的特定区域的末端，使得这些不同区域在多重反应中被分析，相同的被应用的生物化学在平行处理中实现NGS测序。本方法和引物在靶DNA区域末端提供通用区域，这些通用区域可用于支持在多种固相和/或水相上捕获和克隆扩增多种靶区域。此外，本方法和引物避免了将DNA连接子连接到由DNA扩增生成的DNA片段的末端的需要，并提供适于高效测序的模板。

[0084] 本方法和引物相对于随后生成克隆扩增产物池的操作(适于在表面、珠或溶液中生成克隆群体)是不可知论的。本技术也是随后用于生成NGS数据的技术不可知论的，并且可以与(例如)Illumina SBS技术、Ion Torrent或Roche 454“一次一个碱基”技术或其他NGS技术诸如纳米孔测序一起使用。通常，如果希望在特定扩增产物的一个或更多个末端引入已定义的序列，则本文所述的方法可能是有利的。

[0085] 本方法和引物主要用于在分析一组选自大得多的可用DNA序列池的DNA靶。

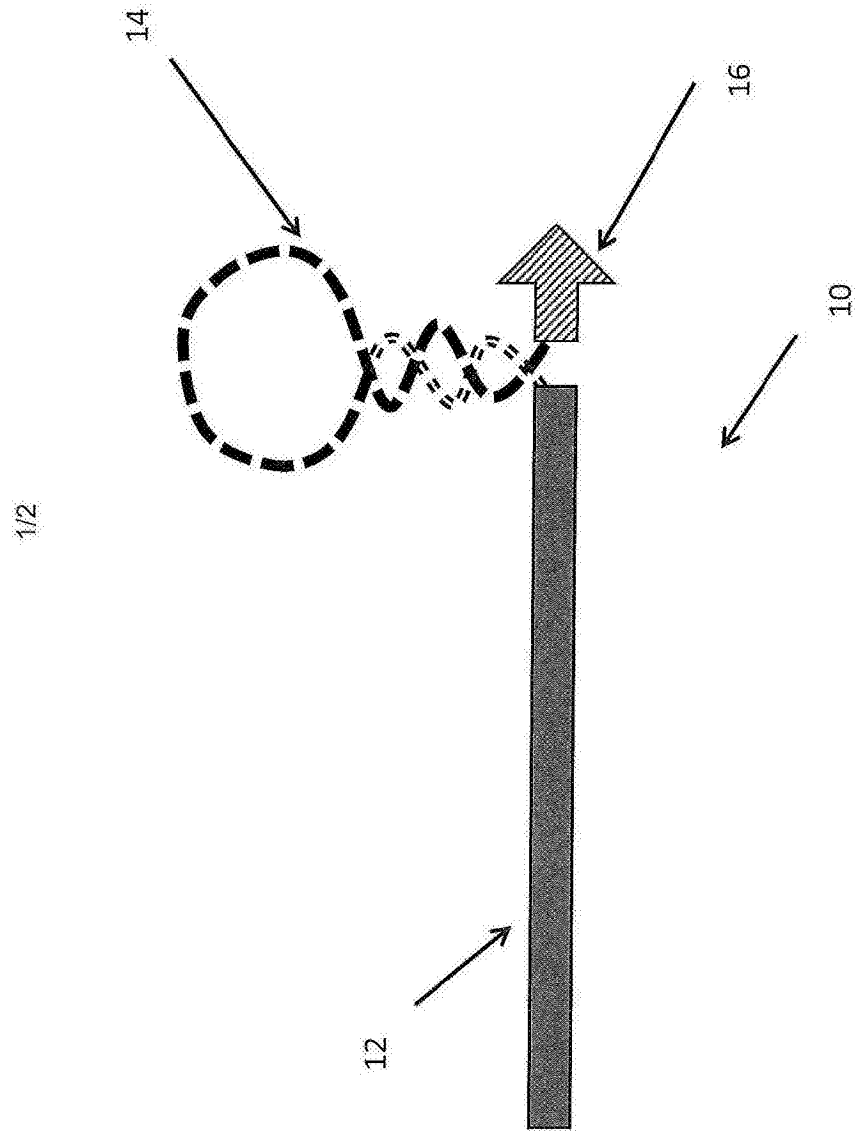


图1

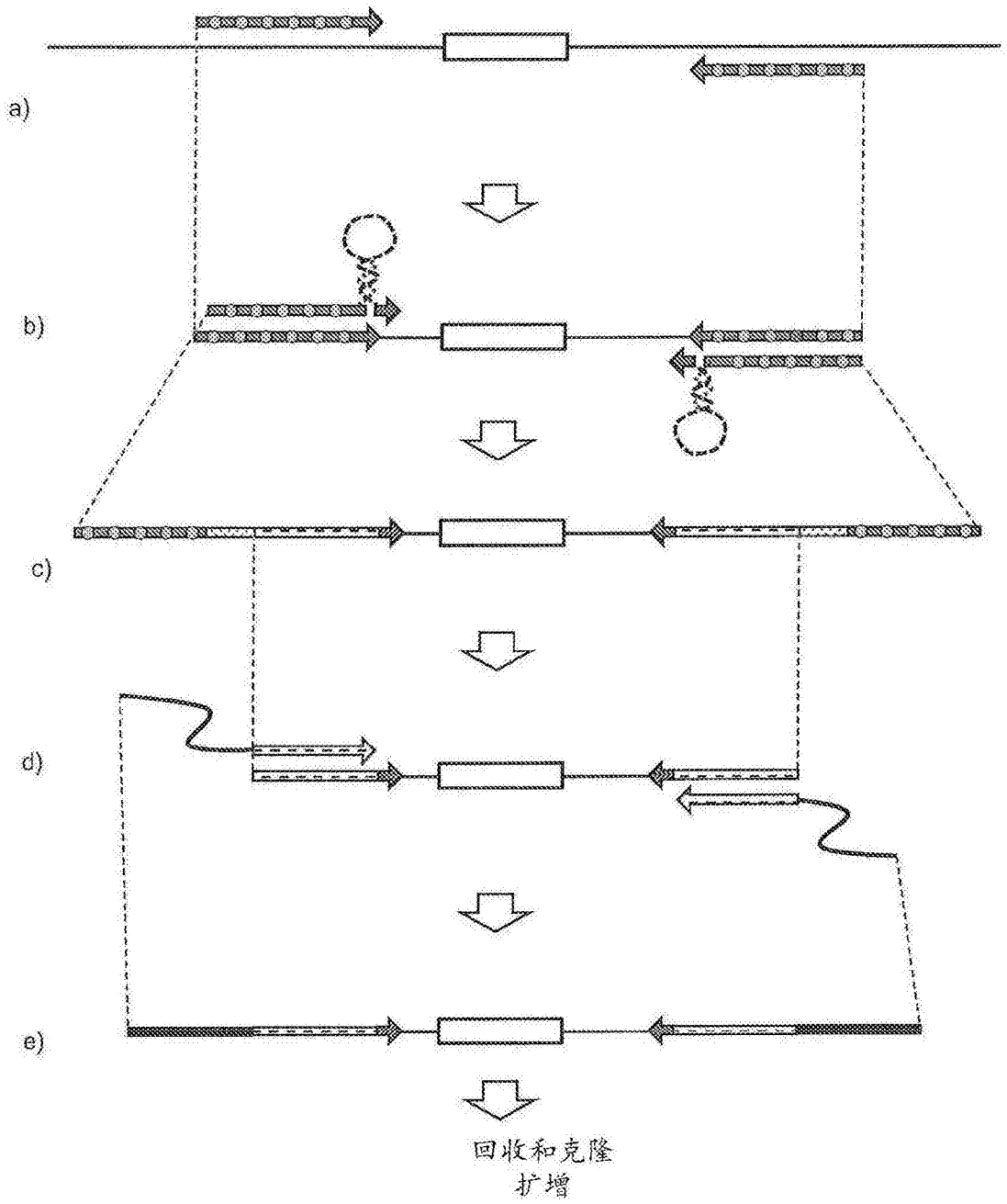


图2