



(12) 发明专利

(10) 授权公告号 CN 113403373 B

(45) 授权公告日 2025.04.11

(21) 申请号 202110629754.3

(74) 专利代理机构 中国专利代理(香港)有限公司

(22) 申请日 2015.11.19

72001

(65) 同一申请的已公布的文献号

专利代理人 甘霖 黄希贵

申请公布号 CN 113403373 A

(51) Int.CI.

(43) 申请公布日 2021.09.17

C12Q 1/6874 (2018.01)

(30) 优先权数据

C12N 15/11 (2006.01)

62/082883 2014.11.21 US

(56) 对比文件

(62) 分案原申请数据

US 7745129 B1, 2010.06.29

201580073880.8 2015.11.19

WO 2013055995 A2, 2013.04.18

(73) 专利权人 布鲁克空间生物学公司

审查员 马忠强

地址 美国马萨诸塞州

(72) 发明人 J.M.贝彻姆 R.哈菲佐夫

权利要求书2页 说明书24页

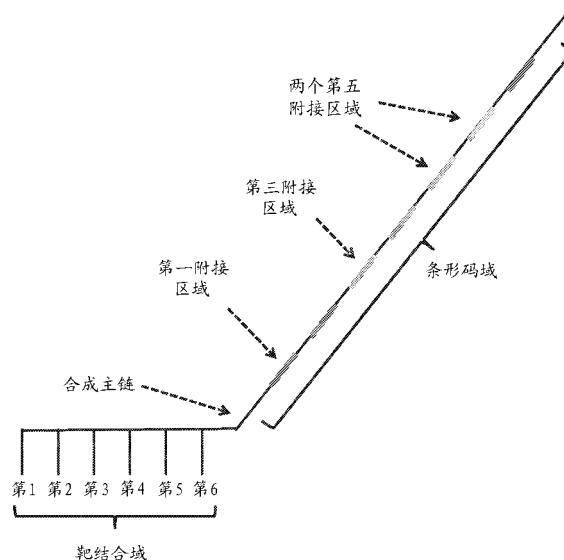
序列表22页 附图43页

(54) 发明名称

无酶且无扩增的测序

(57) 摘要

本发明涉及测序探针、方法、试剂盒和装置，其提供具有长读取长度和低误差率的无酶、无扩增和无文库的核酸测序。使用的序列探针包含靶结合域和条形码域；其中所述靶结合域包含至少四个核苷酸，且能够结合靶核酸；其中所述条形码域包含合成主链，所述条形码域包含至少第一附接区域，所述第一附接区域包含能够被第一互补核酸分子结合的核酸序列，且其中所述第一附接区域的所述核酸序列确定由所述靶结合域的第一个核苷酸结合的所述靶核酸的第一核苷酸的位置和身份。



1. 包含靶结合域和条形码域的测序探针；

其中所述靶结合域包含至少12个核苷酸，并且能够结合靶核酸；

其中所述条形码域包含合成主链，其中所述合成主链包含单链DNA，所述条形码域包含至少2个附接位置，每个附接位置包含至少一个附接区域，所述附接区域包含至少一个能够通过互补核酸分子结合的核酸序列，

其中所述至少2个附接位置对应于所述靶结合域的核苷酸序列，并且其中所述至少2个附接位置中的至少一个附接位置具有不同的核酸序列，且

其中所述至少2个附接位置的每个位置的核酸序列确定了通过所述靶结合域结合的所述靶核酸中的对应核苷酸的位置和身份。

2. 权利要求1的探针，其中所述条形码域包含：

i) 至少3个附接位置；或

ii) 至少4个附接位置。

3. 权利要求1的测序探针，其中所述靶结合域中的核苷酸的数目比所述条形码域中的附接区域的数目多至少两个。

4. 权利要求1的测序探针，其中条形码域中的每个位置具有：(a) 相同数目的附接区域；(b) 一个附接区域；或 (c) 多于一个附接区域。

5. 权利要求1-4中任一项的测序探针，其进一步包含与条形码域的第一附接位置的第一附接区域杂交的互补核酸分子。

6. 权利要求5的测序探针，其中所述互补核酸分子包含可检测标记物。

7. 权利要求5的测序探针，其中所述互补核酸分子是包含可检测标记物的报告复合物的互补核酸分子。

8. 权利要求7的测序探针，其中所述互补核酸分子与一级核酸分子直接或间接连接。

9. 权利要求5的测序探针，其中所述互补核酸包含8个核苷酸至20个核苷酸。

10. 权利要求9的测序探针，其中所述互补核酸包含12个核苷酸。

11. 权利要求8的测序探针，其中所述一级核酸分子与至少一个、两个、三个、四个或五个二级核酸分子杂交。

12. 权利要求11的测序探针，其中所述一个或多个二级核酸分子包含至少一个可检测标记物。

13. 权利要求11的测序探针，其中每个二级核酸分子与至少一个、两个、三个、四个、五个、六个或七个包含至少一个可检测标记物的三级核酸分子杂交。

14. 包含靶结合域和条形码域的测序探针；

其中所述靶结合域包含至少12个核苷酸，并且能够结合靶核酸；

其中所述条形码域包含合成主链，其中所述合成主链包含单链DNA，所述条形码域包含至少4个附接位置，每个附接位置包含至少一个附接区域，所述附接区域包含至少一个能够通过互补核酸分子结合的核酸序列，

其中所述至少4个附接位置对应于所述靶结合域的核苷酸序列，并且其中所述至少2个附接位置各自具有不同的核酸序列，且

其中所述至少4个附接位置的每个位置的核酸序列确定了通过所述靶结合域结合的所述靶核酸中的对应核苷酸的位置和身份。

15. 测序探针群体,其包含多个根据权利要求1-4中任一项的测序探针。

16. 一种方法,其包括以下步骤:

(1) 使包含多个根据权利要求1-4中任一项的测序探针的至少一个测序探针的群体与固定至基质的靶核酸杂交;

(2) 使包含可检测标记物的第一互补核酸分子或包含可检测标记物的第一报告复合物的第一互补核酸分子与所述至少2个附接位置的第一附接位置结合;

(3) 检测结合的第一互补核酸分子的可检测标记物或结合的第一报告复合物的可检测标记物;

(4) 使第一互补核酸分子的可检测标记物或第一报告复合物不结合第一附接位置;

(5) 将包含可检测标记物的第二互补核酸分子或包含可检测标记物的第二报告复合物的第二互补核酸分子与所述至少2个附接位置中的第二附接位置结合;

(6) 检测所述结合的第二互补核酸分子的所述可检测标记物或所述结合的第二报告复合物;

(7) 重复步骤(4)至(6),直到所述至少2个附接位置中的每个附接位置已被包含可检测标记物的互补核酸分子或包含可检测标记物的报告复合物的互补核酸分子结合,并且已检测到所述结合的互补核酸分子的所述可检测标记物或所述结合的报告复合物的所述可检测标记物,从而鉴定被测序探针的靶结合域杂交的靶核酸。

17. 权利要求16的方法,其中步骤(4)和(5)依次或同时发生。

18. 权利要求16的方法,其中所述包含可检测标记物的第一报告复合物的第一互补核酸分子与一级核酸分子直接或间接连接。

19. 权利要求18的方法,其中所述一级核酸分子与至少一个、两个、三个、四个或五个二级核酸分子杂交。

20. 权利要求19的方法,其中每个二级核酸分子包含至少一个可检测标记物。

21. 权利要求19的方法,其中每个二级核酸分子与至少一个、两个、三个、四个、五个、六个或七个包含至少一个可检测标记物的三级核酸分子杂交。

22. 试剂盒,其包含多个根据权利要求1-4中任一项的测序探针,和

至少2个包含可检测标记物的互补核酸分子,其中包含可检测标记物的互补核酸包含与测序探针的至少2个附接位置中的一个杂交的核酸序列,和可检测标记物。

## 无酶且无扩增的测序

[0001] 本申请是以下申请的分案申请:申请日:2015年11月19日;申请号: 201580073880.8 (PCT/US2015/061615);发明名称:同上。

[0002] 相关申请的交叉引用

[0003] 本申请要求2014年11月 21日提交的美国临时申请号62/082,883的权益。上述专利申请的内容通过引用以其整体并入本文。

[0004] 序列表

[0005] 本申请含有已经由EFS-Web以ASCII格式提交且在此通过引用以其整体并入的序列表。所述ASCII拷贝,在2015年11月19日生成,被命名为NATE-025\_ST25.txt,且大小为20,860字节。

### 发明背景

[0006] 目前存在各种核酸测序的方法,即确定核酸分子内核苷酸的精确顺序的方法。目前方法需要酶促地(例如PCR和/或通过克隆)扩增核酸。需要进一步酶促聚合来通过光检测装置产生可检测信号。这种扩增和聚合步骤是昂贵和/或耗时的。因此,本领域需要无扩增和无酶的核酸测序方法。本发明解决了这些需求。

### 发明概述

[0007] 本发明提供测序探针、方法、试剂盒和装置,其提供具有长读取长度和低误差率的无酶、无扩增和无文库的核酸测序。此外,所述方法、试剂盒和装置具有快速的样品至答案(sample-to-answer)能力。这些特征对于在临床环境中的测序特别有用。

[0008] 本文提供了包含靶结合域和条形码域的测序探针。所述靶结合域和条形码域可以可操作地连接,例如共价连接。测序探针任选地包含靶结合域和条形码域之间的间隔物。所述间隔物可以是具有适当机械性能的任何聚合物,例如(1至100个核苷酸,例如2至50个核苷酸的)单链或双链DNA间隔物。双链DNA间隔物的非限制性实例包括SEQ ID NO: 25至SEQ ID NO: 29覆盖的序列。

[0009] 所述靶结合域包含至少四个核苷酸(例如,4、5、6、7、8、9、10、11、12或更多个),并且能够结合靶核酸(例如DNA、RNA和PNA)。所述条形码域包含合成的主链,所述条形码域至少具有包含一个或多个附接区域的第一位置。所述条形码域可以具有1、2、3、4、5、6、7、8、9、10、11、12或更多个位置;每个位置具有一个或多个(例如,一个至五十个)附接区域;每个附接区域包含至少一个(即一个至五十个,例如十个至三十个拷贝的核酸序列)能够可逆结合至互补核酸分子(RNA或DNA)。条形码域中的某些位置可具有比其它位置更多的附接区域;可选地,条形码域中的每个位置具有相同数量的附接区域。第一附接区域的核酸序列确定由靶结合域的第一个核苷酸结合的靶核酸中的第一个核苷酸的位置和身份,而第二附接区域的核酸序列确定由靶结合域的第二个核苷酸结合的靶核酸中的第二个核苷酸的位置和身份。同样,第六附接区域的核酸序列确定由靶结合域的第六个核苷酸结合的靶核酸中的第六个核苷酸的位置和身份。在实施方案中,合成的主链包含多糖、多核苷酸(例如单链或

双链DNA或RNA)、肽、肽核酸或多肽。靶结合域中的核苷酸数目等于或大于条形码域中的位置数(例如,1、2、3、4或更多)。条形码域的特定位置中的每个附接区域可以包括相同核酸序列的一个拷贝和/或相同核酸序列的多个拷贝。然而,附接区域将包括与条形码域的不同位置的附接区域不同的核酸序列,甚至当两个附接区域都标识相同类型的核苷酸(例如腺嘌呤、胸腺嘧啶、胞嘧啶、鸟嘌呤、尿嘧啶和其类似物)时。附接区域可以在合成主链中连接至修饰的单体,例如修饰的核苷酸,从而相对于主链产生分支。附接区域可以是合成主链的多核苷酸序列的一部分。一个或多个附接区域可以与至少一个侧接单链多核苷酸相邻,也就是说,附接区域可以可操作地连接至5'侧接单链多核苷酸和/或3'侧接单链多核苷酸。具有或不具有一个或两个侧接单链多核苷酸的附接区域可以与缺乏可检测标记物的杂交核酸分子杂交。缺乏可检测标记物的杂交核酸分子可以长度为约4至约20个核苷酸之间,例如12个核苷酸或更长。

[0010] 附接区域可以由包含可检测标记物的互补核酸结合。每个互补核酸可以包含可检测标记物。

[0011] 可选地,附接区域可以由作为报告复合物的一部分(包含可检测标记物)的互补核酸结合。互补核酸(包含可检测标记物或报告复合物)可以长度为约4至约20个核苷酸之间,例如约8、10、12和14个核苷酸或更多。在报告复合物中,互补核酸与一级核酸分子(直接或间接)连接。互补核酸可以经由单链或双链核酸接头(例如,包含1至100个核苷酸的多核苷酸)间接连接至一级核酸分子。一级核酸与一个或多个(例如,1、2、3、4、5、6、7、8、9、10或更多个)二级核酸杂交。每个二级核酸与一个或多个(例如,1、2、3、4、5、6、7、8、9、10或更多个)三级核酸杂交;三级核酸包含一个或多个可检测标记物。一个或每个二级核酸可以包含不与一级核酸分子杂交且不与三级核酸分子杂交的区域(“额外手柄”);该区域可以长度为四个或更多个(例如,约6至约40个,例如约8、10、12和14个)核苷酸。不与一级核酸分子杂交并且不与三级核酸分子杂交的区域可以包含与一级核酸分子连接的互补核酸分子的核苷酸序列。该区域可以位于其末端远端的与一级核酸杂交的第二核酸的末端。通过具有包含互补核酸的核苷酸序列的“额外手柄”,报告复合物结合测序探针的可能性和速度大大增加。在本发明的任何实施方案或方面中,当报告复合物包含“额外手柄”时,报告复合物可以经由报告复合物的互补核酸或经由“额外手柄”与测序探针杂交。因此,例如,短语“结合第一附接区域 … 第一报告复合物的第一互补核酸分子”将根据其明确含义理解,并且也被理解为意指“结合第一附接区域 … 第一报告复合物的“额外手柄”。

[0012] 在实施方案中,术语“条形码域”和“合成主链”是同义的。

[0013] 本文提供了用于使用本发明的测序探针对核酸测序的方法。所述方法包括以下步骤:(1)使本发明的至少一个测序探针与(例如,在1、2、3、4、5、6、7、8、9、10或更多个位置)固定至基质的靶核酸杂交;(2)使具有可检测标记物(例如,荧光标记物)的第一互补核酸分子(RNA或DNA)或包含可检测标记物(例如,荧光标记物)的第一报告复合物的第一互补核酸分子与第一附接区域结合;(3)检测可检测标记物,和(4)鉴定固定的靶核酸中的第一核苷酸的位置和身份。任选地,固定的靶核酸在被探针结合之前是延长的。所述方法还包括以下步骤:(5)使第一附接区域(具有或不具有一个或两个侧接单链多核苷酸)与缺乏可检测标记物的第一杂交核酸分子接触,从而未结合具有可检测标记物的第一互补核酸分子或包含可检测标记物的第一报告复合物的第一互补核酸分子且使缺乏可检测标记物的第一杂交

核酸与至少第一附接区域结合；(6)使具有可检测标记物的第二互补核酸分子或包含可检测标记物的第二报告复合物的互补核酸分子与第二附接区域结合；(7)检测可检测标记物；和(8)鉴定固定的靶核酸中的第二核苷酸的位置和身份。重复步骤(5)至(8)，直到已鉴定固定的靶核酸中对应于靶结合域的每个核苷酸。步骤(5)和(6)可以同时或依次发生。每个(例如,第一、第二、第三、第四、第五、第六、第七、第八、第九、第十或更高)互补核酸分子(具有可检测标记物或报告复合物的一部分)与其相应的(即,第一、第二、第三、第四、第五、第六、第七、第八、第九、第十或更高)缺乏可检测标记物的杂交核酸分子具有相同的核酸序列。通过用第一和/或第二捕获探针结合靶核酸的第一位置和/或第二位置而将靶核酸固定至基质；每个捕获探针包含选择性结合基质的亲和标签。第一和/或第二位置可以在靶核酸的末端处或附近。所述基质可以是本领域已知的任何固体支持物，例如包被的载玻片和微流体装置(例如用链霉抗生物素蛋白包被)。位于靶核酸的末端的远端的其它位置可以与基质选择性结合。可以通过施加足以延伸靶核酸的力(例如重力、流体动力、电磁力、流动拉伸、后退弯月面技术(receding meniscus technique)及其组合)来伸长核酸。

[0014] 本文提供了用于使用本发明的一种探针群体或本发明的多种探针群体对核酸测序的方法。所述方法包括以下步骤：(1)使(本发明的)测序探针的第一群体与固定至基质的靶核酸杂交(其中在约相同条件(例如离液剂的水平、温度、盐浓度、pH和流体动力)下，第一群体中的每个测序探针与固定的靶核酸去杂交)；(2)使各自具有可检测标记物的多个第一互补核酸分子或各复合物包含可检测标记物的多个第一报告复合物的多个第一互补核酸分子与第一群体中的各测序探针中的第一附接区域结合；(3)检测可检测标记物；(4)鉴定通过第一群体中的测序探针杂交的固定的靶核酸中的多个第一核苷酸的位置和身份；(5)使第一群体的每个测序探针的每个第一附接区域与多个缺乏可检测标记物的第一杂交核酸分子接触，从而未结合具有可检测标记物或报告复合物的第一互补核酸分子，且使缺乏可检测标记物的第一杂交核酸分子与每个第一附接区域结合；(6)使各自具有可检测标记物的多个第二互补核酸分子或各复合物包含可检测标记物的多个第二报告复合物的多个第二互补核酸分子与第一群体中的各测序探针中的第二附接区域结合；(7)检测可检测标记物；和(8)鉴定通过第一群体中的测序探针杂交的固定的靶核酸中的多个第二核苷酸的位置和身份。在步骤(9)中，重复步骤(5)至(8)，直到已鉴定固定的靶核酸中对应于第一群体中的各测序探针的靶结合域的每个核苷酸。步骤(5)和(6)可以同时或依次发生。从而，对于测序探针的第一群体中的测序探针的靶结合域杂交的固定的靶核酸的区域，鉴定核苷酸的线性顺序。

[0015] 在实施方案中，当使用多个群体(即，多于一个群体)的探针时，所述方法还包括以下步骤：(10)将第一群体的每个测序探针从核酸去杂交；(11)去除第一群体的每个去杂交的测序探针；(12)杂交至少本发明的测序探针的第二群体，其中第二群体中的每个测序探针在约相同条件下与固定的靶核酸去杂交，并且在与第一群体中的测序探针不同的条件与来自固定的靶核酸去杂交；(13)使各自具有可检测标记物的多个第一互补核酸分子或各复合物包含可检测标记物的多个第一报告复合物的多个第一互补核酸分子与第二群体中的各测序探针中的第一附接区域结合；(14)检测可检测标记物；(15)鉴定通过第二群体中的测序探针杂交的固定的靶核酸中的多个第一核苷酸的位置和身份；(16)使第二群体的每个测序探针的每个第一附接区域与多个缺乏可检测标记物的第一杂交核酸分子接触，从而未

结合(具有可检测标记物或来自报告复合物的)第一互补核酸分子,且使缺乏可检测标记物的第一杂交核酸分子与每个第一附接区域结合;(17)使各自具有可检测标记物的多个第二互补核酸分子或各复合物包含可检测标记物的多个第二报告复合物的多个第二互补核酸分子与第二群体中的各测序探针中的第二附接区域结合;(18)检测可检测标记物;(19)鉴定通过第二群体中的测序探针杂交的固定的靶核酸中的多个第二核苷酸的位置和身份;和(20)重复步骤(16)至(19),直到对于通过测序探针的第二群体中的测序探针的靶结合域杂交的固定的靶核酸的区域,已鉴定核苷酸的线性顺序。步骤(16)和(17)可以同时或依次发生。

[0016] 第二群体中的每个测序探针可以在与第一群体中的测序探针从靶核酸去杂交的平均条件不同的条件(例如,更高的温度、更高的离液剂的水平、更高的盐浓度、更高的流速和不同的pH)下从固定的靶核酸去杂交。

[0017] 然而,当使用多于两个探针群体时,则两个连续群体中的探针可以在不同条件下杂交,并且非连续群体中的探针可以在相似的条件下去杂交。作为实例,第一群体和第三群体中的探针可以在相似的条件下去杂交。在实施方案中,连续的探针群体在越来越更严格的条件(例如更高的离液剂水平、盐浓度和温度)下去杂交。对于微流体装置,使用温度作为实例,第一探针群体可以在第一温度下保持杂交,但在高于第一温度的第二温度下去杂交。第二探针群体可以在第二温度下保持杂交,但在高于第二温度的第三温度下去杂交。在该实例中,流过用于初始探针群体的靶核酸的溶液(包含本方法所需的试剂)的温度低于流过用于后续探针群体的靶核酸的溶液。

[0018] 在一些实施方案中,在已使用探针群体之后,将探针群体从靶核酸去杂交,并且使用相同的探针群体的新等分试样。例如,在已将第一探针群体杂交、检测和去杂交之后,将第一探针群体的后续等分试样杂交。可选地,作为实例,可以将第一探针群体去杂交并用第二探针群体替代;一旦已检测和去杂交第二群体,将第一探针群体的后续等分试样与靶核酸杂交。因此,后续群体中的探针可以与已先前测序的靶核酸的区域杂交(从而获得重复和/或验证的序列信息)或后续群体中的探针可以与未先前测序的靶核酸的区域杂交(从而获得新的序列信息)。因此,当先前读取值不令人满意(由于任何原因)和/或改进由测序读取导致的比对的准确性时,可以将探针群体重新等分。

[0019] 在类似条件下杂交和去杂交的探针可以具有相似的其靶结合域的长度、GC含量或重复碱基的频率及其组合。寡核苷酸的T<sub>m</sub>和长度之间的关系教导于例如Sugimoto等人, *Biochemistry*, 34, 11211-6。

[0020] 当使用多于两个探针群体时,如对于测序探针的第一和第二群体所述的步骤用附加的探针群体(例如,10至100至1000个群体)重复。使用的探针群体的数量将取决于多种因素,包括但不限于靶核酸的大小,每个群体中唯一探针的数量,所需的测序探针间的重叠程度以及探针至目标区域的富集。

[0021] 探针群体可以含有针对靶核酸中的特定感兴趣的区域(例如含有突变(例如,点突变)或SNP等位基因的区域)的额外测序探针。探针群体可含有较少的针对靶核酸中的较不感兴趣的特定区域的测序探针。

[0022] 可以将测序探针群体区室化成测序探针的离散的较小集合。区室化可以基于测序探针中的靶结合域的预测解链温度和/或测序探针中的靶结合域的序列基序。区室化可以

基于经验推导的规则。使用不同的反应条件,例如基于温度、盐浓度和/或缓冲液含量,不同的测序探针集合可以与靶核酸反应。可以进行区室化以覆盖具有均匀覆盖度的靶核酸。可以进行区室化以覆盖具有已知覆盖度概况的靶核酸。

[0023] 可以缩短测序探针群体中的靶结合域的长度以增加靶核酸的特定区域中的探针的覆盖度。可以增加测序探针群体中的靶结合域的长度以减少靶核酸的特定区域中的探针的覆盖度,例如高于测序装置的分辨极限。

[0024] 可选地或另外地,可以增加群体中的测序探针的浓度以增加靶核酸的特定区域中的探针的覆盖度。可以降低测序探针的浓度以减少靶核酸的特定区域中的探针的覆盖度,例如高于测序装置的分辨极限。

[0025] 用于对核酸测序的方法还包括组装固定的靶核酸的每个区域的核苷酸的每个鉴定的线性顺序、从而鉴定固定的靶核酸的序列的步骤。组装的步骤使用具有存储在其上的可执行程序的非暂时计算机可读存储介质,其指示微处理器以排列核苷酸的每个鉴定的线性顺序,从而获得核酸的序列。组装可以“实时”发生,即在从测序探针收集数据的同时,而不是在已收集所有数据之后。

[0026] 靶核酸,即测序的靶核酸,可以长度在约4至1,000,000个核苷酸之间。靶标可以包括整个、完整的染色体或其长度,其长度大于1,000,000个核苷酸。

[0027] 本文提供了用于进行本发明的方法的装置。

[0028] 本文提供了包括本发明的测序探针和用于进行本发明的方法的试剂盒。在实施方案中,所述试剂盒包括能够经由捕获探针固定核酸的基质,本发明的多种测序探针,至少一种捕获探针,至少一种具有可检测标记物的互补核酸分子,至少一种缺乏可检测标记物的互补核酸分子,以及使用说明书。在实施方案中,所述试剂盒包含约或至少4096种独特的测序探针。4096是包括每个可能的六聚体组合所必需的独特探针(即,对于条形码域中具有六个附接区域的探针)的最小数量。此处,实现“4096”,是因为对于六个位置有四个核苷酸选项: $4^6$ 。对于在条形码域中具有四个附接区域的探针集合,将仅需要256种(即 $4^4$ 种)独特探针。对于在其靶结合域中具有8个核苷酸的探针集合,将需要 $4^8$ 种(即65,536种)独特探针。对于在其靶结合域中具有10个核苷酸的探针集合,将需要 $4^{10}$ 种(即1,048,576种)独特探针。

[0029] 在实施方案中,所述试剂盒包含约或至少二十四种不同的具有可检测标记物的互补核酸分子,和约或至少二十四种不同的缺乏可检测标记物的杂交核酸分子。作为非限制性实例,互补核酸可以结合具有SEQ ID NO: 1至24之一的序列的附接区域。可以包括在条形码域中的额外示例性序列列于SEQ ID NO: 42至SEQ ID NO: 81中。实际上,核苷酸序列不受限制;优选地,其缺乏与已知核苷酸序列的实质的同源性(例如,50%至99.9%);这有助于避免互补核酸和靶核酸的不期望的杂交。

[0030] 任何上述方面和实施方案可以与任何其它方面或实施方案组合。

[0031] 除非另有定义,本文所用的所有技术和科学术语和本发明所属领域普通技术人员通常理解的具有相同含义。在说明书中,单数形式也包括复数形式,除非上下文另有明确规定;作为实例,术语“一(a)”、“一(an)”和“该(the)”被理解为是单数或复数,且术语“或”被理解为包括性的。通过实例的方式,“一个要素”意指一个或多个元素。在整个说明书中,词语“包含(comprising)”或变型诸如“包含(comprises)”或“包含(comprising)”将被理解为暗示包括所述要素、整数或步骤或要素、整数或步骤的组群,但不排除任何其它要素、整数

或步骤或要素、整数或步骤的组群。约可以被理解为在所示值的10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.5%、0.1%、0.05%或0.01%内。除非从上下文另外清楚，否则本文提供的所有数值都通过术语“约”进行修饰。

[0032] 尽管与本文所述的方法和材料类似或等效的方法和材料也可以用于实施或测试本发明，但下文描述合适的方法和材料。本文提到的所有出版物、专利申请、专利和其它参考文献全文通过引用并入。本文引用的参考文献不被承认是所请求保护的发明的现有技术。在冲突的情况下，将以本说明书，包括定义，为准。此外，材料、方法和实例仅是说明性的，并不旨在进行限制。本发明的其它特征和优点从以下详述和权利要求将是显而易见的。

## 附图简述

[0033] 本专利或申请文件含有至少一个以颜色绘制的图。本专利或专利申请公开的具有彩图的拷贝在请求和支付必要费用后由专利和商标局提供。

[0034] 当结合附图进行时，上述和进一步的特征将从详述更清楚地理解。

[0035] 图1至图5显示本发明的示例性测序探针的示意图。

[0036] 图6A至图6D是显示本发明的测序探针的变体的示意图。

[0037] 图7显示本发明的测序探针的靶结合域的示意图；所述域包括零个、两个或四个具有通用碱基的核苷酸。

[0038] 图8A至图8E说明本发明的测序方法的步骤。

[0039] 图9A显示本发明的测序方法的初始步骤。

[0040] 图9B显示包含可检测标记物的报告复合物的示意图。

[0041] 图9C显示各自包含可检测标记物的多个报告复合物。

[0042] 图9D至9G显示图9A中开始的测序方法的进一步步骤。

[0043] 图10显示图9D和图9E中显示的步骤的替代说明，以及由此获得的示例性数据。所示测序探针的片段具有SEQ ID NO: 82的序列。

[0044] 图11说明图10中显示的方法的变型。所示测序探针的片段同样具有SEQ ID NO: 82的序列。

[0045] 图12说明本发明的方法。

[0046] 图13将本发明的测序方法中所需的步骤与其它测序方法所需的步骤进行比较。

[0047] 图14和图15例举可通过本发明获得的性能测量。

[0048] 图16比较了本发明和各种其它测序方法/装置的测序速率、读数和临床效用。

[0049] 图17表明本发明的测序方法的低原始误差率。所示模板序列具有SEQ ID NO: 83的序列。

[0050] 图18将可从本发明获得的测序数据与其它测序方法进行比较。

[0051] 图19表明本发明的测序方法的单碱基特异性。所示的模板和探针序列(从上至下)具有SEQ ID NO: 84至SEQ ID NO: 88的序列。

[0052] 图20A显示本发明的报告复合物的各种设计。

[0053] 图20B显示从图20A中显示的报告复合物获得的荧光计数。

[0054] 图20C显示用于构建本发明的报告复合物的示例性配方。

[0055] 图21A显示包含“额外手柄”的报告复合物的设计。

- [0056] 图21B显示从具有“额外手柄”的报告复合物获得的荧光计数。
- [0057] 图22A和图22B显示本发明的报告复合物的两种示例性设计的杂交动力学。
- [0058] 图23显示与图8至图12中显示的方法不同的方法中使用的本发明的测序探针的示意图。
- [0059] 图24显示可用于本发明中的可消耗测序卡的示意图。
- [0060] 图25显示如实施例3中所述的10聚体的错配检测。所示核苷酸(从上至下)具有SEQ ID NO: 89至SEQ ID NO: 99的序列。
- [0061] 图26显示根据实施例3中所述的靶结合域的大小的杂交能力。由于非常高的报告物浓度,背景很高,并且没有预先纯化。所示核苷酸(从上至下)具有SEQ ID NO: 100至SEQ ID NO: 104的序列。
- [0062] 图27显示单个斑点相比于全长报告物之间的比较。单个斑点的结果显示杂交速度是全长条形码的1000倍(条件100nM靶标,30分钟杂交)。

## 发明详述

[0063] 本发明提供测序探针、方法、试剂盒和装置,其提供具有长读取长度和低误差率的无酶、无扩增和无文库的核酸测序。

[0064] 测序探针

[0065] 本发明涉及包含靶结合域和条形码域的测序探针。本发明的测序探针的非限制性实例显示于图1至6中。

[0066] 图1显示本发明的测序探针的示意图。该示例性测序探针具有六个核苷酸的靶结合域,其各自对应于条形码域中的位置(其包含一个或多个附接区域)。注意第一附接区域;其对应于靶结合域中由第一核苷酸结合的靶核酸的核苷酸。注意条形码域上的第三个位置。注意包含两个附接区域的第五个位置。条形码域上的每个位置都可以具有多个附接区域。例如,位置可以具有1至50个附接区域。条形码域中的某些位置可具有比其它位置更多的附接区域(如此处位置5相对于位置1至4和6所示);可选地,条形码域中的每个位置具有相同数量的附接区域(参见例如图2、3、5和6)。尽管未显示,但每个附接区域包含能够可逆结合至互补核酸分子(RNA或DNA)的至少一个(即一个至五十个,例如十个至三十个)拷贝的核酸序列。在图1中,附接区域对于组成条形码域的线性多核苷酸分子是必不可少的。

[0067] 图2显示本发明的测序探针的示意图。该示例性测序探针具有六个核苷酸的靶结合域,其各自对应于条形码域中的附接区域。注意第一附接区域;其对应于靶结合域中由第一核苷酸结合的靶核酸的核苷酸。圈出条形码域的第四位置,其包含条形码域的一部分和两个第四附接区域。注意两个第六附接区域。在此,每个位置具有两个附接区域;然而,条形码域上的每个位置可以具有一个附接区域或多个附接区域,例如2至50个附接区域。尽管未显示,但每个附接区域包含能够可逆结合至互补核酸分子(RNA或DNA)的至少一个(即一个至五十个,例如十个至三十个)拷贝的核酸序列。在图2中,条形码域是连接附接区域的线性多核苷酸分子;附接区域对于多核苷酸分子不是必不可少的。

[0068] 图3显示本发明的测序探针的另一个示意图。该示例性测序探针具有四个核苷酸的靶结合域,其中这四个核苷酸对应于条形码域中的四个位置。每个位置显示具有三个连接的附接区域。

[0069] 图4显示本发明的测序探针的另一个示意图。该示例性测序探针具有十个核苷酸的靶结合域。然而,只有前六个核苷酸对应于条形码域中的六个位置。添加第七至第十个核苷酸(由“n<sub>1</sub>至n<sub>4</sub>”指示)以增加靶结合域的长度,从而影响探针将杂交并保持与靶核酸杂交的可能性。在实施方案中,“n”核苷酸可以在对应于条形码域中的位置的核苷酸之前。在实施方案中,“n”核苷酸可以在对应于条形码域中的位置的核苷酸之后。在图4中,显示了四个“n”核苷酸;然而,靶结合域可以包括多于四个“n”核苷酸。“n”核苷酸可以具有通用碱基(例如,肌苷、2'-脱氧肌苷(次黄嘌呤脱氧核苷酸)衍生物、硝基吲哚、硝基吲唑类似物和疏水性芳族非氢键键合基团),其可以与四个典型碱基中任一个碱基配对。

[0070] 本发明的另一个测序探针显示于图5中。此处,“n”核苷酸在对应于条形码域中的位置的核苷酸之前和之后。显示的示例性测序探针具有十个核苷酸的靶结合域。然而,靶结合域中只有三至八个核苷酸对应于条形码域中的六个位置(第一至第六)。添加第一、第二、第九和第十核苷酸(由“n<sub>1</sub>至n<sub>4</sub>”指示)以增加靶结合域的长度。在图5中,显示了四个“n”核苷酸;然而,靶结合域可以包括多于或少于四个“n”核苷酸。

[0071] 图6A至图6D显示图1的测序探针的变体。在图6A中,条形码域中靶结合域中的核苷酸的线性顺序和附接区域的线性顺序从左至右进行(相对于图示)。在图6B中,条形码域中靶结合域中的核苷酸的线性顺序和附接区域的线性顺序从右至左进行(相对于图示)。在图6C中,靶结合域中的核苷酸的线性顺序相对于条形码域中的附接区域的线性顺序反转。在本发明的任何探针中,只要设计探针、使得靶结合域中的每个核苷酸对应于靶结合域中的一个或多个附接域,则靶结合域中的核苷酸和条形码域中的附接区域缺乏严格顺序;图6D中显示缺乏严格顺序。本发明的任何探针(例如,图1至5中例举的那些)可具有如图6中所示的核苷酸和附接区域的顺序。

[0072] 所述靶结合域具有至少四个核苷酸,例如至少4、5、6、7、8、9、10、11、12或更多个核苷酸。所述靶结合域优选为多核苷酸。所述靶结合域能够结合靶核酸。

[0073] 探针可以包括可操作地连接至合成主链的多个拷贝的靶结合域。

[0074] 可以设计探针以控制杂交和/或去杂交的可能性以及这些发生的速率。通常,探针的T<sub>m</sub>越低,探针将与/从靶核酸去杂交越快且可能性越高。因此,使用较低T<sub>m</sub>探针将减少与靶核酸结合的探针的数量。

[0075] 靶结合域的长度部分地影响探针杂交并保持与靶核酸杂交的可能性。通常,靶结合域越长(更多数目的核苷酸),靶核苷酸中将存在互补序列的可能性越小。相反,靶结合域越短,靶核苷酸中将存在互补序列的可能性越大。例如,四聚体序列将位于靶核酸中的概率为1/256,相比之下,六聚体序列将位于靶核酸中的概率为1/4096。因此,当与更长探针的集合相比时,对于给定的核酸区段,较短探针的集合将可能在更多位置结合。

[0076] 图7显示10聚体靶结合域。在一些实施方案中,所述靶结合域包括与四个典型核苷酸(A、G、C和T)中的任一个碱基配对的四个通用碱基(鉴定为“U<sub>b</sub>”)。在实施方案中,所述靶结合域包括一个至六个(例如2个和4个)通用碱基。靶结合域可以不包括通用核苷酸。图7指出,在靶结合域中具有6个特定核苷酸的“完整”探针群体将需要4096个独特探针,并且具有10个特定核苷酸的“完整”探针群体将需要约~100万个独特探针。

[0077] 在一些情况下,优选具有较短靶结合域的探针以增加给定核酸区段中的读数,由此富集靶核酸或靶核酸的一部分、特别是特别感兴趣的部分的覆盖度,例如当检测突变或

SNP等位基因时。

[0078] 然而,可优选具有更少数量的与靶核酸结合的探针,因为存在区域中太多探针可引起其可检测标记物重叠、从而防止分辨两个临近探针。这解释如下。考虑到一个核苷酸长度为0.34nm,并且考虑到测序装置的横向(x-y)空间分辨率为约200nm,测序装置的分辨极限为约588个碱基对(即1个核苷酸/0.34nm x 200nm)。也就是说,当两个探针在彼此的约588个碱基对之内时,上述测序装置将不能分辨来自与靶核酸杂交的两个探针的信号。因此,取决于测序装置的分辨率,两个探针将需要在其可检测标记物可以被分辨为不同的“斑点”之前间隔分开约600bp。所以,在最佳间距下,每600bp的靶核酸应该有单一探针。可以使用各种软件方法(例如,利用荧光强度值和波长依赖比率)来监测、限制和潜在地去卷积在靶核酸的可分辨区域内杂交的探针的数量且相应地设计探针群体。此外,可以选择提供更离散信号的可检测标记物(例如,荧光标记物)。此外,文献中的方法(例如,Small和Parthasarathy:“Superresolution localization methods.” *Annu. Rev. Phys. Chem.*, 2014; 65:107-25)描述了结构照明和各种超分辨率方法,其减少高达10's的纳米的测序显微镜的分辨极限。使用更高分辨率测序装置允许使用具有较短靶结合域的探针。

[0079] 如上所提及,设计探针的T<sub>m</sub>可以影响与靶核酸杂交的探针的数量。可选地或另外地,可以增加群体中的测序探针的浓度以增加靶核酸的特定区域中的探针的覆盖度。可以降低测序探针的浓度以减少靶核酸的特定区域中的探针的覆盖度,例如高于测序装置的分辨极限。

[0080] 术语“靶核酸”应当意指其序列将通过本发明的探针、方法和装置测定的核酸分子(DNA、RNA或PNA)。通常,术语“靶核酸”、“核酸分子”、“核酸序列”、“核酸”、“核酸片段”、“寡核苷酸”和“多核苷酸”可互换使用,并且意在包括,但不限于,可具有各种长度的核苷酸(脱氧核糖核苷酸或核糖核苷酸或其类似物)的聚合形式。核酸的非限制性实例包括基因、基因片段、外显子、内含子、基因间DNA(包括但不限于异染色体DNA)、信使RNA(mRNA)、转移RNA、核糖体RNA、核酶、小干扰RNA(siRNA)、非编码RNA(ncRNA)、cDNA、重组多核苷酸、支链多核苷酸、质粒、载体、序列的分离的DNA、序列的分离的RNA、核酸探针和引物。

[0081] 本方法直接对获得自样品(例如来自生物体的样品)的核酸分子测序,优选不进行转化(或扩增)步骤。作为实例,对于基于RNA的测序,本方法不需要在可获得序列之前将RNA分子转化为DNA分子(即经由cDNA的合成)。由于不需要扩增或转化,当核酸在样品中或当其获得自样品时,本发明中测序的核酸将保留核酸中存在的任何独特碱基和/或表观遗传标记物。在本领域已知的测序方法中丢失了这种独特碱基和/或表观遗传标记物。

[0082] 靶核酸可以获得自核酸的任何样品或来源,例如任何细胞、组织或生物体,在体外,化学合成仪等等。靶核酸可以通过任何本领域公认的方法获得。在实施方案中,核酸获得自临床受试者的血液样品。可以使用本领域熟知的方法和试剂从来源或样品提取、分离或纯化核酸。

[0083] 包含靶核酸的核酸分子可以通过本领域已知的任何方式片段化。优选地,片段化通过酶促或机械方式进行。机械方式可以是超声处理或物理剪切。酶促方式可以通过用核酸酶(例如脱氧核糖核酸酶I(DNase I))或一种或多种限制性内切核酸酶消化来进行。

[0084] 当包含靶核酸的核酸分子是完整的染色体时,应当采取步骤以避免将染色体片段化。

[0085] 靶核酸可以包括如本领域熟知的天然或非天然核苷酸,包括修饰的核苷酸。

[0086] 本发明的探针可以具有约20纳米至约50纳米的总长度(包括靶结合域、条形码域和任何任选域)。探针的主链可以是包含约120个核苷酸的多核苷酸分子。

[0087] 所述条形码域包含合成主链。合成主链和靶结合域可操作地连接,例如共价附接或经由接头附接。合成主链可以包括任何材料,例如多糖、多核苷酸、聚合物、塑料、纤维、肽、肽核酸或多肽。优选地,合成主链是刚性的。在实施方案中,所述主链包含六个DNA双螺旋的“DNA折纸(DNA origami)”(参见例如Lin等人,“Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA.” *Nature Chemistry*; 2012 Oct; 4(10): 832-9)。条形码可以由DNA折纸拼贴制成(Jungmann等人,“Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT”, *Nature Methods*, Vol. 11, No. 3, 2014)。

[0088] 所述条形码域包含多个位置,例如一个、二个、三个、四个、五个、六个、七个、八个、九个、十个或更多个位置。位置数可以小于、等于或大于靶结合域中核苷酸的数量。优选与主链域中的位置数相比在靶结合域中包括额外的核苷酸,例如一个、二个、三个、四个、五个、六个、七个、八个、九个、十个或更多个核苷酸。条形码域的长度不受限,只要对于如上所述的至少四个位置存在足够的空间。

[0089] 条形码域中的每个位置对应于靶结合域中的核苷酸,因此对应于靶核酸中的核苷酸。作为实例,条形码域中的第一位置对应于靶结合域中的第一个核苷酸,并且条形码域中的第六位置对应于靶结合域中的第六个核苷酸。

[0090] 条形码域中的每个位置包含至少一个附接区域,例如一个至五十个或更多个附接区域。条形码域中的某些位置可以具有比其它位置更多的附接区域(例如,第一位置可以具有三个附接区域,而第二位置可以具有两个附接位置);可选地,条形码域中的每个位置具有相同数量的附接区域。每个附接区域包含能够被互补核酸分子(例如,DNA或RNA)可逆结合的至少一个(即一个至五十个,例如十个至三十个)拷贝的核酸序列。在实例中,第一附接区域中的核酸序列确定由靶结合域的第一个核苷酸结合的靶核酸中的第一个核苷酸的位置和身份。每个附接区域可以在合成主链中连接至修饰的单体(例如修饰的核苷酸),使得附接区域从合成主链分支。在实施方案中,附接区域对于多核苷酸主链是必不可少的;也就是说,主链是单个多核苷酸,并且附接区域是单个多核苷酸序列的部分。在实施方案中,术语“条形码域”和“合成主链”是同义的。

[0091] 附接区域中的核酸序列确定由测序探针的靶结合域的核苷酸结合的靶核酸中的核苷酸的位置和身份。在探针中,每个附接区域将具有独特的总体序列。实际上,条形码域上的每个位置可以具有包含核酸序列的附接区域,所述核酸序列编码四种核苷酸之一,即对腺嘌呤、胸腺嘧啶/尿嘧啶、胞嘧啶和鸟嘌呤之一具有特异性。此外,第一位置的附接区域(并且例如编码胞嘧啶)将包括不同于第二位置的附接区域(并且例如编码胸腺嘧啶)的核酸序列。因此,对于编码胸腺嘧啶的第一位置的附接区域中的核酸序列,将不存在鉴定对应于靶结合域的第一个核苷酸的靶核酸中的腺嘌呤的互补核酸分子的结合。此外,对于第二位置的附接区域,将不存在鉴定对应于靶结合域的第一个核苷酸的靶核酸中的胸腺嘧啶的互补核酸分子的结合。

[0092] 条形码域上的每个位置可以包括一个或多个(最多五十个,优选十个至三十个)附

接区域；因此，每个附接区域可以结合一个或多个(最多五十个，优选十个至三十个)互补核酸分子。作为实例，图1中的探针具有包含两个附接区域的第五位置，并且图2中的探针具有具有六个附接区域的第二位置。在实施方案中，位置的附接区域的核酸序列是相同的；因此，结合那些附接区域的互补核酸分子是相同的。在替代实施方案中，位置的附接区域的核酸序列是不同的；因此，结合那些附接区域的互补核酸分子是不同的，例如各自包含不同的核酸序列和/或可检测的标记物。因此，在替代实施方案中，附接至附接区域的不同核酸分子(例如其可检测标记物)的组合提供了用于鉴定靶核酸中的核苷酸的代码。

[0093] 仅为了说明的目的，表1提供了用于测序探针的附件区域的示例性序列，所述测序探针在其条形码域中具有最多六个位置和在与其结合的互补核酸上的可检测标记物。

[0094] 表 1：

[0095] 靶结合域中的核苷酸/条形码域中的位置	核苷酸	附接区域中的核酸序列(5'至3')	互补核酸的可检测标记物	SEQ ID NO
1	A	ATACATCTAG	GFP	1
1	G	GATCTACATA	RFP	2
1	C	TTAGGTAAAG	CFP	3
1	U/T	TCTTCATTAC	YFP	4
2	A	ATGAATCTAC	GFP	5
2	G	TCAATGTATG	RFP	6
2	C	AATTGAGTAC	CFP	7
2	U/T	ATGTTAACGG	YFP	8
3	A	AATTAGGATG	GFP	9
3	G	ATAATGGATC	RFP	10
3	C	TAATAAGGTG	CFP	11
3	U/T	TAGTTAGAGC	YFP	12
4	A	ATAGAGAAGG	GFP	13
4	G	TTGATGATAC	RFP	14
4	C	ATAGTGATTC	CFP	15
4	U/T	TATAACGATG	YFP	16
5	A	TTAACGTTAG	GFP	17
5	G	ATACGTTATG	RFP	18
5	C	TGTACTATAG	CFP	19
5	U/T	TTAACAAAGTG	YFP	20
6	A	AACTATGTAC	GFP	21
6	G	TAACATATGAC	RFP	22
6	C	ACTAATGTTC	CFP	23
6	U/T	TCATTGAATG	YFP	24

[0096] 如表1中所见，第一附接区域的核酸序列可以是SEQ ID NO: 1至SEQ ID NO: 4之一，且第二附接的核酸序列可以是SEQ ID NO: 5至SEQ ID NO: 8之一。当靶核酸中的第一个核苷酸是腺嘌呤时，第一附接区域的核酸序列将具有SEQ ID NO: 1的序列，并且当靶核酸中的第二个核苷酸是腺嘌呤时，第二附接区域的核酸序列将具有SEQ ID NO: 5的序列。

[0097] 在实施方案中，互补核酸分子可以被可检测标记物结合。在替代实施方案中，互补核酸与包含可检测标记物的报告复合物相结合。

[0098] 互补核酸的核苷酸序列不受限制；优选地，其缺乏与已知核苷酸序列的实质的同源性(例如，50%至99.9%)；这有助于避免互补核酸和靶核酸的不期望的杂交。

[0099] 可用于本发明的报告复合物的实例显示于图9B中。在该实例中，互补核酸连接至一级核酸分子，其进而与多个二级核酸分子杂交，所述二级核酸分子各自进而与其上附接

一个或多个可检测标记物的多个三级核酸分子杂交。

[0100] 在实施方案中,一级核酸分子可以包含约90个核苷酸。二级核酸分子可以包含约87个核苷酸。三级核酸分子可以包含约15个核苷酸。

[0101] 图9C显示示例性报告复合物的群体。包括在图9C的左上小图中的是与探针的附接区域1杂交的四个复合物。对于可存在于探针的靶结合域的核苷酸位置1的每个可能的核苷酸,存在一种类型的报告复合物。在此,在进行本发明的序列方法的同时,如果探针的报告域的位置1被具有“蓝色”可检测标记物的报告复合物结合,则靶结合域中的第一个核苷酸被鉴定为腺嘌呤。可选地,如果位置1由具有“绿色”可检测标记物的报告复合物结合,则靶结合域中的第一个核苷酸被鉴定为胸腺嘧啶。

[0102] 报告复合物可以具有各种设计。例如,一级核酸分子可以与至少一个(例如,1、2、3、4、5、6、7、8、9、10或更多个)二级核酸分子杂交。每个二级核酸分子可以与至少一个(例如,1、2、3、4、5、6、7、8、9、10或更多个)三级核酸分子杂交。示例性报告复合物显示于图20A中。在此,“4x3”报告复合物具有与四个二级核酸分子杂交的一个一级核酸分子(其与互补核酸分子连接),所述二级核酸分子各自与三个三级核酸分子(各自包含可检测标记物)杂交。在该图中,复合物的每个互补核酸长度为12个核苷酸(“12个碱基”);然而,互补核酸的长度不受限制,并且可以小于12个或多于12个核苷酸。右下方复合物包括其互补核酸和其一级核酸分子之间的间隔物。所述间隔物被鉴定为20至40个核苷酸长;然而,间隔物的长度是非限制性的,并且其可以短于20个核苷酸或长于40个核苷酸。

[0103] 图20B显示从图20A中所示的四个示例性报告复合物获得的可变平均(荧光)计数。在图20B中,将10pM生物素化的靶模板附接至链霉抗生物素蛋白包被的流动室表面上,将10nM的报告复合物流至流动室上;孵育1分钟后,洗涤流动室,将流动室成像,并计数荧光特征。

[0104] 在实施方案中,所述报告复合物是“预构建的”。也就是说,在将复合物与探针接触之前,将复合物中的每个多核苷酸杂交。用于预构建五个示例性报告复合物的示例性配方显示于图20C中。

[0105] 图21A显示替代的报告复合物,其中二级核酸分子具有与三级核酸分子不杂交并且在一级核酸分子远端的“额外手柄”。在该图中,每个“额外手柄”为12个核苷酸长(“12聚体”);然而,它们的长度不受限制,并且可以小于12个或多于12个核苷酸。在实施方案中,“额外手柄”各自包含互补核酸的核苷酸序列;因此,当报告复合物包含“额外手柄”时,报告复合物可以经由报告复合物的互补核酸或经由“额外手柄”与测序探针杂交。因此,报告复合物结合测序探针的可能性增加。“额外手柄”设计还可以改善杂交动力学。不受理论束缚,“额外手柄”基本上增加报告复合物的互补核酸的有效浓度。

[0106] 图21B显示使用针对图20B描述的程序从具有“额外手柄”的五个示例性报告复合物获得的可变平均(荧光)计数。

[0107] 图22A和22B显示两个示例性报告复合物的杂交动力学和荧光强度。到约5分钟,总计数开始至平台期,表明添加的大多数报告复合物已经发现可用的靶标。

[0108] 可检测部分、标记物或报告物可以以各种方式(包括直接或间接附接可检测部分、诸如荧光部分、比色部分等)结合至互补核酸或三级核酸分子。本领域技术人员可以参考针对标记核酸的参考文献。荧光部分的实例包括但不限于黄色荧光蛋白(YFP)、绿色荧光蛋白

(GFP)、青色荧光蛋白(CFP)、红色荧光蛋白(RFP)、伞形酮、荧光素、异硫氰酸荧光素、罗丹明、二氯三嗪胺荧光素、花菁、丹酰氯、藻青蛋白、藻红蛋白等。荧光标记物及其与核苷酸和/或寡核苷酸的附接描述于许多综述,包括Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, 第九版 (Molecular Probes, Inc., Eugene, 2002); Keller and Manak, *DNA Probes*, 第2版 (Stockton Press, New York, 1993); Eckstein, 编辑 *Oligonucleotides and Analogues: A Practical Approach* (IRL Press, Oxford, 1991); 和Wetmur, *Critical Reviews in Biochemistry and Molecular Biology*, 26: 227-259 (1991)。适用于本发明的具体方法公开于以下参考文献样本中:美国专利号4,757,141;5,151,507;和5,091,519。在一个方面,一种或多种荧光染料用作标记的靶序列的标记物,例如如由美国专利号5,188,934(4,7-二氯荧光素染料);5,366,860(光谱可分辨的罗丹明染料);5,847,162(4,7-二氯罗丹明染料);4,318,846(醚取代的荧光素染料);5,800,996(能量转移染料);Lee等人5,066,580(黄嘌呤染料);5,688,648(能量转移染料);等所公开。标记也可以用量子点进行,如以下专利和专利公开中所公开:美国专利号6,322,901;6,576,291;6,423,551;6,251,303;6,319,426;6,426,513;6,444,143;5,990,479;6,207,392;2002/0045045;和2003/0017264。如本文所用,术语“荧光标记物”包含通过一个或多个分子的荧光吸收和/或发射性能传递信息的信号传导部分。这种荧光性能包括荧光强度、荧光寿命、发射光谱特征、能量转移等。

[0109] 容易掺入核苷酸和/或寡核苷酸序列中的市售荧光核苷酸类似物包括但不限于 Cy3-dCTP、Cy3-dUTP、Cy5-dCTP、Cy5-dUTP (Amersham Biosciences, Piscataway, NJ)、荧光素-12-dUTP、四甲基罗丹明-6-dUTP、TEXAS RED™-5-dUTP、CASCADE BLUE™-7-dUTP、BODIPY TMFL-14-dUTP、BODIPY TMR-14-dUTP、BODIPY TMTR-14-dUTP、RHODAMINE GREEN™-5-dUTP、OREGON GREEN™ 488-5-dUTP、TEXAS RED™-12-dUTP、BODIPY TM 630/650-14-dUTP、BODIPY TM 650/665-14-dUTP、ALEXA FLUOR™ 488-5-dUTP、ALEXA FLUOR™ 532-5-dUTP、ALEXA FLUOR™ 568-5-dUTP、ALEXA FLUOR™ 594-5-dUTP、ALEXA FLUOR™ 546-14-dUTP、荧光素-12-UTP、四甲基罗丹明-6-UTP、TEXAS RED™-5-UTP、mCherry、CASCADE BLUE™-7-UTP、BODIPY TM FL-14-UTP、BODIPY TMR-14-UTP、BODIPY TM TR-14-UTP、RHODAMINE GREEN™-5-UTP、ALEXA FLUOR™ 488-5-UTP、LEXA FLUOR™ 546-14-UTP (Molecular Probes, Inc. Eugene, OR) 等。可选地,可以使用例如亚磷酰胺盐(phosphoroamidite)或NHS化学在寡核苷酸合成期间加入上述荧光团和本文提及的那些。本领域已知用于定制合成具有其它荧光团的核苷酸的方案(参见, Henegariu等人 (2000) *Nature Biotechnol.* 18:345)。2-氨基嘌呤是可以在其合成期间直接掺入寡核苷酸序列的荧光碱基。核酸也可以先用插入染料诸如DAPI、YOYO-1、溴化乙锭、花青染料(例如SYBR Green)等进行染色。

[0110] 可用于合成后附接的其它荧光团包括但不限于ALEXA FLUOR™ 350、ALEXA FLUOR™ 405、ALEXA FLUOR™ 430、ALEXA FLUOR™ 532、ALEXA FLUOR™ 546、ALEXA FLUOR™ 568、ALEXA FLUOR™ 594、ALEXA FLUOR™ 647、BODIPY 493/503、BODIPY FL、BODIPY R6G、BODIPY 530/550、BODIPY TMR、BODIPY 558/568、BODIPY 558/568、BODIPY 564/570、BODIPY 576/589、BODIPY 581/591、BODIPY TR、BODIPY 630/650、BODIPY 650/665、Cascade Blue、Cascade Yellow、Dansyl、丽丝胺罗丹明B、Marina Blue、Oregon Green 488、Oregon Green 514、Pacific Blue、Pacific Orange、罗丹明6G、罗丹明绿、罗丹明红、四甲基罗丹明、得克

萨斯红(可得自Molecular Probes, Inc., Eugene, OR)、Cy2、Cy3、Cy3.5、Cy5、Cy5.5、Cy7(Amersham Biosciences, Piscataway, NJ)等。还可以使用FRET串联荧光团,包括但不限于PerCP-Cy5.5、PE-Cy5、PE-Cy5.5、PE-Cy7、PE-Texas Red、APC-Cy7、PE-Alexa染料(610、647、680)、APC-Alexa染料等。

[0111] 金属银或金颗粒可用于增强来自荧光标记的核苷酸和/或寡核苷酸序列的信号(Lakowicz等人(2003) BioTechniques 34:62)。

[0112] 寡核苷酸序列的其它合适的标记物可以包括荧光素(FAM、FITC)、洋地黄毒昔、二硝基苯酚(DNP)、丹酰基、生物素、溴脱氧尿昔(BrdU)、六组氨酸(6xHis)、磷酸-氨基酸(例如,P-tyr、P-ser、P-thr)等。在一个实施方案中,以下半抗原/抗体对用于检测,其中各抗体用可检测标记物衍生化:生物素/α-生物素、洋地黄毒昔/α-洋地黄毒昔、二硝基苯酚(DNP)/α-DNP、5-羧基荧光素(FAM)/α-FAM。

[0113] 本文描述的可检测标记物是光谱可分辨的。参考多个荧光标记物的“光谱可分辨的”是指标记物的荧光发射带足够不同,即足够不重叠,可以基于相应标记物通过标准光检测系统生成的荧光信号来区分附接各自标记物的分子标签,所述标准光检测系统例如采用带通滤光器和光电倍增管等的系统,如美国专利号4,230,558;4,811,218;等,或Wheless等人,第21-76页,Flow Cytometry: Instrumentation and Data Analysis (Academic Press, New York, 1985)中所述的系统所例举的。在一个方面,光谱可分辨的有机染料,诸如荧光素、罗丹明等,意味着波长发射最大值间隔分开至少20nm,另一方面间隔分开至少40nm。在另一个方面,光谱可分辨的螯合的镧系元素化合物、量子点等意味着波长发射最大值间隔分开至少10nm,在另一个方面,分开至少15nm。

[0114] 测序方法

[0115] 本发明涉及用于使用本发明的测序探针对核酸测序的方法。所述方法的实例显示于图8至12中。

[0116] 所述方法包括使本发明的至少一个测序探针与(例如,在1、2、3、4、5、6、7、8、9、10或更多个位置)固定至基质的靶核酸可逆地杂交。

[0117] 所述基质可以是本领域已知的任何固体支持物,例如包被的载玻片和微流体装置,其能够固定靶核酸。在某些实施方案中,所述基质是表面、膜、珠、多孔材料、电极或阵列。所述靶核酸可以固定至本领域技术人员显而易见的任何基质上。

[0118] 在实施方案中,所述靶核酸由捕获探针结合,所述捕获探针包含与靶核酸的一部分互补的域。所述部分可以是靶核酸的末端,或可以不是朝向末端。

[0119] 示例性有用的基质包括包含选自配体、抗原、碳水化合物、核酸、受体、凝集素和抗体的结合部分的基质。所述捕获探针包含能够与基质的结合部分结合的结合部分。包含反应性部分的示例性的有用的基质包括但不限于包含环氧基、醛、金、酰肼、巯基、NHS-酯、胺、硫醇、羧酸酯、马来酰亚胺、羟甲基膦、亚氨酯、异氰酸酯、羟基、五氟苯基酯、补骨脂素、吡啶基二硫化物或乙烯基砜、聚乙二醇(PEG)、水凝胶或其混合物的表面。这样的表面可以从商业来源获得或根据标准技术制备。包含反应性部分的示例性有用的基质包括但不限于OptArray-DNA NHS基团(Accler8)、Nexterion Slide AL (Schott)和Nexterion Slide E (Schott)。

[0120] 在实施方案中,捕获探针的结合部分是生物素,且基质包含抗生物素蛋白(例如链

霉抗生素蛋白)。包含抗生素蛋白的有用的基质是市售的,包括TB0200 (Accelr8)、SAD6、SAD20、SAD100、SAD500、SAD2000 (Xantec)、SuperAvidin (Array-It)、链霉抗生素蛋白载玻片(目录号#MPC 000, Xenopore)和STREPTAVIDINslide (目录号#439003, Greiner Bio-one)。

[0121] 在实施方案中,捕获探针的结合部分是抗生素蛋白(例如,链霉抗生素蛋白),且基质包含生物素。市售的包含生物素的有用的基质包括但不限于Optiarray-生物素(Acceler8)、BD6、BD20、BD100、BD500和BD2000 (Xantec)。

[0122] 在实施方案中,捕获探针的结合部分可以包含能够通过光活化与基质结合的反应性部分。所述基质可以包含光反应性部分,或者纳米报告物的第一部分可以包含光反应部分。光反应性部分的一些实例包括芳基叠氮化物,例如N((2-吡啶基二硫基)乙基)-4-叠氮基水杨酰胺;氟化芳基叠氮化物,诸如4-叠氮基-2,3,5,6-四氟苯甲酸;基于二苯甲酮的试剂,诸如4-苯甲酰基苯甲酸的琥珀酰亚胺酯;和5-溴-脱氧尿苷。

[0123] 在实施方案中,捕获探针的结合部分可以经由本领域技术人员显而易见的其它结合对固定至基质。

[0124] 在结合至基质之后,可以通过施加足以延伸靶核酸的力(例如重力、流体动力、电磁力“电拉伸”、流动拉伸、后退弯月面技术及其组合)来伸长核酸。

[0125] 所述靶核酸可以由第二捕获探针结合,所述第二捕获探针包含与靶核酸的第二部分互补的域。所述部分可以是靶核酸的末端,或可以不是朝向末端。第二捕获探针的结合可以在靶核酸的延伸之后或期间发生,或者第二捕获探针可以与未延伸的靶核酸结合。第二捕获探针可具有如上所述的结合。

[0126] 捕获探针可以包含可检测标记物(即基准点)或与可检测标记物(即基准点)结合。

[0127] 捕获探针能够从样品分离靶核酸。在此,将捕获探针加入包含靶核酸的样品。捕获探针经由捕获探针的区域结合靶核酸,所述捕获探针的区域与靶核酸的区域互补。当靶核酸接触包含结合捕获探针的结合部分的基质时,核酸被固定至基质上。

[0128] 为了确保使用者从高片段化的样品尽可能多地“捕获”靶核酸分子,包括多个捕获探针、每个捕获探针与靶核酸的不同区域互补是有帮助的。例如,可以存在三个捕获探针集合,其中第一集合与靶核酸的靠近其5'末端的区域互补,第二集合与靶核酸中间的区域互补,且第三集合靠近其3'末端。这可以概括到每个靶核酸的“n个感兴趣区域”。在该实例中,片段化的靶核酸的每个单独集合与包含或结合生物素标签的捕获探针结合。输入样品的1/n(其中n=靶核酸中不同区域的数目)被分离用于每个集合室。捕获探针结合感兴趣的靶核酸。然后将靶核酸经由捕获探针的生物素固定至附接至基质的抗生素蛋白分子。任选地,将靶核酸例如经由流动或静电力拉伸。所有n个集合可以同时拉伸和结合,或者为了最大化完全拉伸分子的数量,可以首先拉伸和结合集合1(其捕获最多5'区域);然后可以拉伸和结合集合2(其捕获靶区域的中间);最后,可以拉伸和结合集合3。

[0129] 所需的不同捕获探针的数量与靶核酸片段的大小成反比。换句话说,对于高度片段化的靶核酸,将需要更多的捕获探针。对于具有高度片段化和降解的靶核酸的样品类型(例如,福尔马林固定石蜡包埋的组织),包括多个捕获探针集合可以是有用的。另一方面,对于具有长靶核酸片段的样品,例如体外获得的分离的核酸,在5'末端的单个捕获探针可以是足够的。

[0130] 靶核酸在两个捕获探针之间或在一个捕获探针之后且在靶核酸的末端之前的区域在本文中称为“缺口”。该缺口是靶核酸可用于由本发明的测序探针结合的一部分。最小缺口是靶结合域长度(例如,4至10个核苷酸),并且最大缺口是整个染色体的大部分。

[0131] 固定的靶核酸显示于图12中。在此,两个捕获探针被鉴定为“5'捕获探针”和“3'捕获探针”。

[0132] 图8A显示与靶核酸结合的测序探针的示意图。在此,靶核酸具有胸苷(T)。在顶部显示包含可检测标记物或报告复合物的第一互补核酸集合,集合的每个成员具有不同的可检测标记物(例如,胸苷通过绿色信号鉴定)和不同的核苷酸序列。靶结合域中的第一个核苷酸结合靶核酸中的T。探针的第一附接区域包括一个或多个核苷酸序列,其指示探针的靶结合域中的第一个核苷酸结合胸苷。因此,仅胸苷的互补核酸结合条形码域的第一位置。如所示,包含可检测标记物的编码胸腺嘧啶的第一互补核酸或包含可检测标记物的报告复合物与探针的条形码域的第一位置的附接区域结合。

[0133] 互补核酸或报告复合物的集合的数量与条形码域中的位置数相同。因此,对于具有六个位置的条形码域,将在探针上循环六个集合。

[0134] 可选地,在使靶核酸与探针接触之前,探针可以在其第一位置与包含可检测标记物的互补核酸或报告复合物杂交。因此,当与其靶核酸接触时,探针能够从其第一位置发射可检测信号,并且不必提供针对条形码域上的第一位置的互补核酸或报告复合物的第一集合。

[0135] 图8B继续图8A中所示的方法。在此,与条形码域的第一位置的附接区域结合的胸苷的第一个互补核酸(或报告复合物)已被用于胸苷且缺乏可检测标记物的第一杂交核酸替代。用于胸苷且缺乏可检测标记物的第一杂交核酸替代包含可检测标记物的先前结合的互补核酸或先前结合的报告复合物。因此,条形码域的位置1不再发射可检测信号。

[0136] 在实施方案中,包含可检测标记物的互补核酸或报告复合物可以从附接区域除去,但未用缺乏可检测标记物的杂交核酸替代。这可以例如通过添加离液剂、增加温度、改变盐浓度、调节pH和/或应用流体动力来进行。在这些实施方案中,需要较少的试剂(即,缺乏可检测标记物的杂交核酸)。

[0137] 图8C继续请求保护的本发明的方法。在此,所述靶核酸具有胞苷(C)和其之前的胸苷(T)。互补核酸或报告复合物的第二集合显示在顶部,集合的每个成员具有不同的可检测标记物和不同的核苷酸序列。此外,第一集合的互补核酸或报告复合物的互补核酸的核苷酸序列不同于第二集合的那些的核苷酸序列。然而,碱基特异性可检测标记物对于互补核酸的集合是共同的,例如胸苷通过绿色信号鉴定。在此,靶结合域中的第二个核苷酸结合靶核酸中的C。探针的第二附接区域具有核苷酸序列,其指示探针的靶结合域中的第二个核苷酸结合胞苷。因此,仅来自第二集合和用于胞苷的包含可检测标记物的互补核酸或报告复合物结合条形码域的第二位置。如所示,编码胞苷的第二互补核酸或报告复合物在探针的条形码域的第二位置结合。

[0138] 在实施方案中,图8C中所示的步骤在图8B中所示的步骤之后。在此,一旦(图8A的)互补核酸或报告复合物的第一集合已经被(图8B中)缺乏可检测标记物的第一杂交核酸替代,则提供互补核酸或报告复合物的第二集合(如图8C中所示)。可选地,图8C中所示的步骤与图8B中所示的步骤同时进行。在此,与互补核酸或报告复合物的第二集合(如图8C中所

示)同时提供缺乏可检测标记物的第一杂交核酸(如图8B中所示)。

[0139] 图8D继续图8C中所示的方法。在此,条形码域上的第一至第五位置由包含可检测标记物的互补核酸或报告复合物结合,并且已被替换为缺乏可检测标记物的杂交核酸。条形码域的第六位置目前由包含可检测标记物的互补核酸或报告复合物结合,其将靶结合域中的第六位置鉴定为与鸟嘌呤(G)结合。

[0140] 如上所提及,包含可检测标记物的互补核酸或报告复合物可以从附接区域除去,但未用缺乏可检测标记物的杂交核酸替代。

[0141] 如果需要,可以通过掺入加速可检测标记物的交换速率的小单链寡核苷酸来加速可检测标记物交换速率(例如,“Toe-Hold” Probes;参见,例如,Seeling等人,“Catalyzed Relaxation of a Metastable DNA Fuel”; *J. Am. Chem. Soc.* 2006, 128 (37), pp12211-12220)。

[0142] 可在条形码域的最终位置(图8D中的第六位置)上替换互补核酸或报告子复合物;然而,当用测序探针被替换为另一个测序探针时,这可能是不必要的。实际上,图8D的测序探针现在可以从靶核酸去杂交并除去,并且被替换为还没有被任何互补核酸结合的第二(重叠的或非重叠的)测序探针,如图8E中所示。第二探针群体中可以包括图8E中的探针。

[0143] 像图8A至8E一样,图9A和9D至9G显示本发明的方法步骤;然而,图9A和9D至9G清楚地显示报告复合物(包含可检测标记物)与测序探针的附接区域结合。图9D和9E显示与报告复合物杂交的探针发射的荧光信号。图9D和9E显示靶核酸具有“T-A”的序列。

[0144] 图10概述图9D和9E中所示的步骤。在图的顶部显示示例性探针的核苷酸序列并鉴定探针的重要域。该探针在其靶结合域和其条形码域之间包含任选的双链DNA间隔物。条形码域依次包含“侧接1”部分、“AR-1”部分、“AR-1/侧接2”部分、“AR-2”部分和“AR-2/侧接3”部分。在步骤1中,“AR-1检测”与探针的“AR-1”和“AR-1/侧接2”部分杂交。“AR-1检测”对应于编码第一位置胸苷的报告复合物或包含可检测标记物的互补核酸。因此,步骤1对应于图9D。在步骤2中,“缺乏1”与探针的“侧接1”和“AR-1”部分杂交。“缺乏1”对应于对探针的第一附接区域特异性的缺乏可检测标记物的杂交核酸(如图9E中显示为覆盖第一附接区域的黑色条)。通过在报告复合物或互补核酸5’的“侧接1”位置杂交,杂交核酸更有效地将报告复合物/互补核酸从探针置换。“侧接”部分也称为“Toe-Holds”。在步骤3中,“AR-2检测”与探针的“AR-2”和“AR-2/侧接3”部分杂交。“AR-2检测”对应于报告复合物或包含可检测标记物的编码第二位置鸟嘌呤的互补核酸。因此,步骤3对应于图9E。在该实施方案中,依次提供缺乏可检测标记物的杂交核酸和包含可检测标记物的互补核酸/报告复合物。

[0145] 可选地,同时提供缺乏可检测标记物的杂交核酸和包含可检测标记物的互补核酸/报告复合物。该替代实施方案显示于图11中。在步骤2中,提供“缺乏1”(缺乏可检测标记物的杂交核酸)连同“AR-2检测”(编码第二位置鸟嘌呤的报告复合物)。该替代实施方案可以比图10中所示的实施方案更加有效,因为它将两个步骤组合成一个。

[0146] 图12说明本发明的方法。在此,将靶核酸捕获并固定在两个位置,从而产生探针能够结合的“缺口”。将第一探针群体杂交至靶核酸上,并检测可检测的标记物。用第二探针群体、第三探针群体至超过100个探针群体重复初始步骤。约100个探针群体的使用提供靶核酸中每个核苷酸的约5倍的覆盖度。图12基于检测来自一个视场(FOV)的信号所需的时间提供了读取时间的估计速率。

[0147] 探针沿靶核酸长度的分布对于可检测信号的分辨是至关重要的。如上所论述,两个可检测标记物的分辨极限为约600个核苷酸。优选地,探针群体中的每个测序探针将彼此结合不近于600个核苷酸。如上所论述,600个核苷酸是典型测序装置的分辨极限。在这种情况下,测序探针将提供单次读取值;这在图12中显示在最左边的分辨限制点。

[0148] 随机地,但部分取决于靶结合域的长度、探针的T<sub>m</sub>和所用探针的浓度,群体中两个不同的测序探针可在彼此的600个核苷酸内结合。在这种情况下,无序多次读取值将从单个分辨限制点发出;这在图12中显示在第二个分辨限制点。

[0149] 替代地或另外地,可以降低群体中的测序探针的浓度以减少靶核酸的特定区域中的探针的覆盖度,例如高于测序装置的分辨极限,从而从分辨限制点产生单个读数。

[0150] 图23显示不同于图8至12中使用的测序探针的示意图。在此,条形码域上的每个位置由包含可检测标记物的互补核酸或报告复合物结合。因此,在该实例中,可以读取六个核苷酸序列而不需要依次替换互补核酸。此测序探针的使用将减少获得序列信息的时间,因为省略所述方法的许多步骤。然而,该探针将受益于不重叠的可检测标记物,例如,荧光团被非重叠波长的光激发或荧光团发射不重叠波长的光。

[0151] 所述方法还包括组装固定的靶核酸的每个区域的核苷酸的每个鉴定的线性顺序、从而鉴定固定的靶核酸的序列的步骤。组装的步骤使用具有存储在其上的可执行程序的非暂时计算机可读存储介质。该程序指示微处理器以排列靶核酸的每个区域的核苷酸的每个鉴定的线性顺序,从而获得核酸的序列。组装可以“实时”发生,即在从测序探针收集数据的同时,而不是在已收集所有数据之后。

[0152] 上述方面和实施方案中的任一个可以与此处在概述和/或详述部分中公开的任何其它方面或实施方案组合。

[0153] 定义:

[0154] 在某些示例性实施方案中,如本文所用的术语“退火”和“杂交”可互换使用以表示形成稳定的双链体。在一个方面,稳定的双链体是指双链体结构在诸如以下的条件下不被严格洗涤破坏:双链体的链的T<sub>m</sub>以下约5°C或以上约5°C的温度和低单价盐浓度,例如小于0.2M,或小于0.1M或本领域技术人员已知的盐浓度。术语“完美匹配的”当参考双链体使用时是指构成双链体的多核苷酸和/或寡核苷酸链与彼此形成双链结构,使得每条链中的每个核苷酸经历与另一条链中的核苷酸的Watson-Crick碱基配对。术语“双链体”包括但不限于可以采用的核苷类似物诸如脱氧肌苷、具有2-氨基嘌呤碱基的核苷、PNA等的配对。两个寡核苷酸之间的双链体中的“错配”意味着双链体中的一对核苷酸不能进行Watson-Crick键合。

[0155] 如本文所用,术语“杂交条件”将通常包括小于约1M、更通常小于约500mM且甚至更通常小于约200mM的盐浓度。杂交温度可以低至5°C,但通常大于22°C,更通常大于约30°C,并且经常超过约37°C。杂交通常在严格条件,例如探针将与其靶序列特异性杂交的条件下进行。严格条件是序列依赖性的,并且在不同情况下是不同的。较长的片段可需要较高的杂交温度用于特异性杂交。由于其它因素可影响杂交的严格性,包括互补链的碱基组成和长度,有机溶剂的存在和碱基错配的程度,参数的组合比单独任一个的绝对测量更重要。

[0156] 通常,选择严格条件以便在定义的离子强度和pH下比特定序列的T<sub>m</sub>低约5°C。示例性严格条件包括在pH 7.0至8.3和至少25°C的温度下至少0.01M至不超过1M Na离子浓度

(或其它盐)的盐浓度。例如,5X SSPE (750 mM NaCl, 50 mM磷酸钠,5mM EDTA,pH7.4) 和25-30℃的温度的条件适用于等位基因特异性探针杂交。对于严格条件,参见例如Sambrook, Fritsche和Maniatis, “Molecular Cloning A Laboratory Manual, 2nd Ed.” Cold Spring Harbor Press (1989) 和Anderson Nucleic Acid Hybridization, 第1版, BIOS Scientific Publishers Limited (1999)。如本文所用,术语“与…特异性杂交”或“与…特异性杂交”或类似术语是指分子在严格条件下实质上与一个或多个特定核苷酸序列结合、双链化或杂交。

[0157] 与探针的特定位置结合的可检测标记物可以“读取”(例如其检测到的荧光)一次或多次;“读取”可以与术语“碱基呼叫(basecall)”同义。多次读取改善准确性。当检测到从单个初始靶分子得到的序列信息的连续序列段时,靶核酸序列被“读取”;通常,这经由多通道共有物产生(如下所定义)。如本文所用,术语“覆盖度”或“覆盖深度”是指靶标区域已被测序(经由离散读取值)并与参考序列比对的次数。读取覆盖度是映射至特定参考目标序列的总读数;碱基覆盖度是在特定基因组位置进行的碱基呼叫的总数。

[0158] 如本文所用,“hybe和seq循环”是指检测特定探针或探针群体上的每个附接区域所需的所有步骤。例如,对于能够检测靶核酸上的六个位置的探针,一个“hybe和seq循环”将至少包括使探针与靶核酸杂交,使互补核酸/报告复合物与探针的条形码域上的六个位置中每一个的附接区域杂交,并且检测与六个位置中的每一个结合的可检测标记物。

[0159] 术语“k-聚体探针”与本发明的探针同义。

[0160] 当比对来自离散读取的两个或更多个序列时,可以组合重叠部分以产生单个共有序列。在重叠部分具有相同碱基的位置(比对的单一列),这些碱基成为共有物。可以使用各种规则来产生在重叠序列之间存在不一致的位置的共有物。简单多数规则将列中最常见碱基用作共有物。“多通道共有物”是来自单个靶分子的所有离散探针读数的比对。根据应用的探针群体/投票的总循环次数,单个靶分子内的每个碱基位置可以用不同的冗余或重叠水平进行查询;通常,冗余增加碱基呼叫的置信水平。

[0161] “原始精确度”是系统正确地鉴定碱基的固有能力的量度。原始精确度取决于测序技术。“共有精确度”是系统使用额外读数和统计能力来正确地鉴定碱基的能力。“特异性”是指每次运行总读数映射至预期靶标的读数的百分比。“均匀性”是指靶区域的序列覆盖度的变异性;高均匀性与低变异性相关。该特征通常被报告为所有靶向区域的平均覆盖深度的 $\geq 20\%$ 所覆盖的靶向区域的分数。随机误差(即固有测序化学误差)可以用相同靶核酸的“多次通过”测序容易地校正;考虑到足够数量的通过,可以实现实质上“完美共有”或“无误差”测序。

[0162] 本文描述的方法可以使用能够实现方法和/或记录结果的任何设备来实现和/或记录结果。可以使用的设备的实例包括但不限于电子计算设备,包括所有类型的计算机。当本文描述的方法在计算机中实现和/或记录时,能够含有计算机程序的任何计算机可读介质中可以含有可用于配置计算机以实施所述方法的步骤的计算机程序。可以使用的计算机可读介质的实例包括但不限于磁盘、CD-ROM、DVD、ROM、RAM、非暂时性计算机可读介质以及其它存储器和计算机存储设备。可用于配置计算机以实施方法的步骤、组装序列信息和/或记录结果的计算机程序也可以通过电子网络,例如通过因特网、内联网或其它网络提供。

[0163] “可消耗的测序卡”(图24)可并入本领域已知的荧光成像设备中。具有许多变化特

征的任何荧光显微镜都能够执行该测序读取。例如:宽场灯、激光、LED、多光子、共聚焦或全内反射照明可用于激发和/或检测。在荧光显微镜的发射检测通道上,具有基于滤光片或基于光栅的光谱分辨(一个或多个光谱分辨发射波长)的相机(单个或多个)和/或光电倍增管(单个或多个)是可能的。标准计算机可以控制可消耗的测序卡,流经卡的试剂和通过荧光显微镜的检测。

[0164] 测序数据可以通过任何数量的标准下一代测序组装器进行分析(参见例如,Wajid 和 Serpedin, “Review of general algorithmic features for genome assemblers for next generation sequencers” *Genomics, proteomics & bioinformatics*, 10 (2), 58-73, 2012)。将在显微镜的单个衍射限制区域内获得的测序数据“局部组装”,以从衍射斑点内的多个读数产生共有序列。然后将多个衍射点组装的读数一起映射,以产生代表整个靶向基因集合的连续序列,或全基因组的从头组装。

[0165] 与本发明相关的额外教导描述于以下的一者或二者中:U.S. 8,148,512、U.S. 7,473,767、U.S. 7,919,237、U.S. 7,941,279、U.S. 8,415,102、U.S. 8,492,094、U.S. 8,519,115、U.S. 2009/0220978、U.S. 2009/0299640、U.S. 2010/0015607、U.S. 2010/0261026、U.S. 2011/0086774、U.S. 2011/0145176、U.S. 2011/0201515、U.S. 2011/0229888、U.S. 2013/0004482、U.S. 2013/0017971、U.S. 2013/0178372、U.S. 2013/0230851、U.S. 2013/0337444、U.S. 2013/0345161、U.S. 2014/0005067、U.S. 2014/0017688、U.S. 2014/0037620、U.S. 2014/0087959、U.S. 2014/0154681和U.S. 2014/0162251,其各自通过引用以其整体并入本文。

## 实施例

[0166] 实施例1:本发明对靶核酸测序的方法是快速的

[0167] 下面描述了本发明的方法中且如图8至12中所示的步骤的时序。

[0168] 本发明需要最少的样品制备。例如,如图13中所示,样品中的核酸可以在2小时或更短或准备时间后开始读取;这是比Ion Torrent (AmpliSeq™)或Illumina (TruSight)测序所需显著更少的时间,其分别需要约12或9小时准备时间。

[0169] 图14中显示示例性运行的计算,并且图15中显示循环时间的计算。

[0170] 探针群体与固定的靶核酸的结合需要约60秒。可以通过在合成主链上利用多个拷贝的靶结合域来加速该反应。用微流体控制的流体交换装置,洗去未结合的探针需要约半秒。

[0171] 添加互补核酸的第一个集合(包含可检测的标记物)并将其结合至条形码域的第一位置中的附接区域需要约15秒。

[0172] 针对四种不同的颜色对每个视场(FOV)进行成像,每种颜色表示单一碱基。放置在5'捕获探针或3'捕获探针(或两者)上的基准点可有助于在两个位置之间仅读取排成一行的(与缺口靶核酸的存在一致的)那些光学条形码。也可以将基准点添加至每个视场,以便在测序过程中的连续步骤后产生图像的相等比对。所有四个图像都可以在单个FOV获得,然后光学读取装置可以移动至新FOV,或者以一种颜色采用所有FOV,然后以第二颜色重新成像。单个FOV可以在约半秒内读取。移动至下一FOV花费约半秒。因此,读取“n”个FOV的时间等于“n” × 1秒)。

[0173] 具有可检测标记物的互补核酸通过加热或用过量缺乏可检测标记物的互补核酸洗涤而从条形码域的第一位置去除。如果需要,可以通过掺入加速可检测标记物的交换速率的小单链寡核苷酸来加速可检测标记物交换速率(例如,“Toe-Hold”探针;参见例如,Seeling等人,“Catalyzed Relaxation of a Metastable DNA Fuel”; *J. Am. Chem. Soc.* 2006, 128(37), pp 12211-12220)。可以对FOV重新成像,以证实在继续移动之前除去具有可检测标记物的所有互补核酸。这花费约十五秒。可以重复该步骤,直到达到背景信号水平。

[0174] 重复上述步骤或探针的条形码域中的剩余位置。

[0175] 读取的总时间等于 $m$ (碱基读取)  $\times$  (15秒+ $n$  FOV  $\times$  1秒+ 15秒)。例如,当条形码域中的位置数为6和20个FOV时,读取的时间等于 $6 \times (30 + 20 + 15)$ 或390秒。

[0176] 将第一群体的探针去杂交。这花费约六十秒。

[0177] 对于第二和随后的探针群体重复上述步骤。如果通过解链温度( $T_m$ )组织测序探针的群体,则每个探针群体将需要多次杂交以确保每个碱基被覆盖至所需深度(这由误差率驱动)。此外,通过分析运行期间的杂交读取值,可在实际测定整个序列之前识别很好地测序的每个单个基因。因此,可以重复循环,直到满足特定期望的误差频率(或覆盖度)。

[0178] 使用上述时序时,连同一些缺口核酸结合密度估计值,可以估计本发明的Nanostring (NSTG) -下一代测序仪的通量。

[0179] 测序仪的净通量通过如下给出:

[0180] 分数碱基占据率  $\times$  <缺口长度>  $\times$  每个FOV的缺口数  $\times$  每个光学条形码的碱基数 / [60秒(探针与靶核酸杂交) + 0.5秒(洗涤) +  $m$ :条形码域中的位置数  $\times$  (15秒(结合互补核酸) +  $n$ FOVs $\times$ 1 + 15秒(未结合互补核酸)) + 60秒(探针与靶核酸去杂交)]。

[0181] 因此,在一个实例中,单个缺口核酸的总“循环”(从图10中所示的方法加在一起):

[0182] 60秒(探针与靶核酸杂交) + 0.5秒(洗涤) +  $m$ -碱基  $\times$  (15秒(结合互补核酸) +  $n$ FOV  $\times$  1 + 15秒(未结合互补核酸)) + 60秒(探针与靶核酸去杂交)。使用 $m = 6$ ,  $n$ FOV = 20, 产生时间 =  $60 + 0.5 + 390 + 60 = 510.5$ 秒。

[0183] 假设:缺口的核酸区域的1%占据率,每个缺口4000个碱基,和每个FOV的5000个缺口的核酸片段,以及6的 $m$ 和20的 $n$ FOV(如上所述)产生以下净通量:

[0184]  $0.01 \times 4000 \times 5000 \times 20 =$  每510.5秒4,000,000个6-碱基读取值 = 47,012.73个碱基/秒。

[0185] 因此,在该实例中,每24小时连续测量的净通量= 4.062千兆碱基(Gb)/天。替代估计每天最多12 Gb。参见图12。

[0186] 如图14中所示,对100个不同靶核酸(“100-plex”)测序所需的运行时间为约4.6小时;对1000个不同靶核酸(“1000-plex”)测序所需的运行时间为约16小时。

[0187] 图16比较了本发明和各种其它测序方法/装置的测序速率、读数和临床效用。

[0188] 实施例2:本发明的方法具有低误差率

[0189] 图17显示,当省略末端位置时,本发明具有约2.1%的原始误差率。

[0190] 对于请求保护的发明,与测序相关的误差率与完全匹配的( $m+n$ )-聚体和单碱基错配的( $m-1+n$ )-聚体之间的自由能差异相关。 $m+n$ 之和是靶结合域中的核苷酸数,且 $m$ 代表条形码域中的位置数。可以使用方程进行杂交选择性的估计(参见,0wczarzy, R. (2005),

Biophys. Chem., 117:207-215 and Integrated DNA Technologies website: at the World Wide Web (www) idtdna.com/analyser/Applications/Instructions/Default.aspxAnalyzerDefinitions=true#MismatchMeltTemp) :

$$[0191] \quad \theta = 1 - \left( \frac{K_a([链2] - [链1]) - 1 + \sqrt{K_a^2([链1] - [链2])^2 + 2K_a([链1] + [链2]) + 1}}{2K_a[链2]} \right)$$

[0192] 其中  $K_a$  是从预测的热力学参数获得的结合平衡常数,

$$[0193] \quad K_a = \exp \left( \frac{-(\Delta H^\circ - T\Delta S^\circ)}{RT} \right)$$

[0194] Theta 代表精确互补序列和单碱基错配序列 (预期其在指定杂交温度下退火至靶标) 的结合百分比。T 是 Kelvins 中的杂交温度,  $\Delta H^\circ$  (焓) 和  $\Delta S^\circ$  (熵) 是从序列和公开的最近邻近热力学参数计算的解链参数, R 是理想气体常数 ( $1.987 \text{ cal.K}^{-1}\text{mole}^{-1}$ ), [链1/2] 是寡核苷酸的摩尔浓度, 且 -273.15 的常数将温度从开尔文转化为摄氏度。从以下出版物获得 DNA/DNA 碱基对 (参见, Allawi, H., SantaLucia, J. *Biochemistry*, 36, 10581)、RNA/DNA 碱基对 (参见, Sugimoto 等人, *Biochemistry*, 34, 11211-6)、RNA/RNA 碱基对 (参见, Xia, T. 等人, *Biochemistry*, 37, 14719) 的最准确、最近邻近参数。

[0195] 作为从 NSTG 测序仪预期的近似误差率的估计的实例如下。对于  $(m + n)$  等于 8' 聚体。考虑以下 8'-聚体条形码及其单碱基错配。

[0196] 5'ATCGTACG3'

[0197] (序列的区域)

[0198] 3'TAGCATGC5'

[0199] (对具有完美匹配的光学条形码测序)

[0200] 3'TAGTATGC5'

[0201] (对具有单碱基错配 (G-T) 配对的光学条形码测序)。

[0202] 使用基于上述方程的 IDT 计算器得到:

[0203] 在  $17.4^\circ\text{C}$  (完美匹配情况的  $T_m$ ) , (50% / 0.3%) 将是在  $T_m$  下与该序列杂交的正确光学条形码与不正确条形码的比率, 得到该序列的估计误差率为 0.6%。

[0204] 非常高的 GC 含量测序计算得到:

[0205] 5'CGCCGGCC3'

[0206] (序列的区域)

[0207] 3'GCGGCCGG5'

[0208] (对具有完美匹配的光学条形码测序)

[0209] 3'GCGGACGG5'

[0210] (对具有单碱基错配 (G-A) 错配对的光学条形码测序)。

[0211] 在  $41.9^\circ\text{C}$  (完美匹配情况的  $T_m$ ) , (50% / 0.4%) 将是在  $T_m$  下与该序列杂交的正确光学条形码与不正确条形码的比率, 得到该序列的估计误差率为 0.8%。

[0212] 8 聚体对的数目的检查得到 0.2% 至 1% 的范围内的误差率分布。尽管上述计算与

使用的条件不同,但是当与误差率可以显著(>> 10%)的其它单分子测序技术(诸如Pacific Biosciences and Oxford Nanopore技术)相比,这些计算提供了本发明的方法将具有相对低的固有误差率的指示。

[0213] 图18表明本发明的原始精确度高于其它测序方法。因此,本发明在比其它测序方法所需的更少的通过之后提供来自单一靶标的共有序列。此外,在30次或更多次通过之后,本发明可以获得“完全一致/无误差”测序(即,99.9999%/Q60),而PacBio测序方法(例如)在70次通过之后不能达到这样的一致。

[0214] 实施例3:本发明具有单碱基对分辨能力

[0215] 图19显示本发明具有单碱基分辨率并且具有低误差率(取决于特定核苷酸取代,在0%至1.5%的范围内)。

[0216] 使用与条形码杂交和使用正常的NanoString基因表达结合技术固定至盒的表面的靶RNA进行另外的实验(参见例如, Geiss等人, “Direct multiplexed measurement of gene expression with color-coded probe pairs”; *Nature Biotechnology*, 26, 317 - 325 (2008)). 测量了具有不同的靶结合域长度和具有完美匹配的条形码(连接至完美10聚体匹配序列的YGBYGR-2um光学条形码)与RNA靶标杂交的能力(图26)。靶结合域的更长长度得到更高的计数。其还显示,10-聚体靶结合域足以注册高于背景的序列。单独单碱基改变的匹配中的每一个都用替代光学条形码合成。计数正确与不正确的光学条形码的比率(图24和25)。

[0217] 10聚体检测SNP的能力,真实序列高于背景>15000个计数,而不正确的序列高于背景最多> 400个计数。在正确探针存在的情况下,预期误差率为实际序列的3%。注意,该数据(本质上)是更糟糕的情况。只有10碱基对杂交序列连接至6.6千碱基光学条形码报告分子(Gen2型)。没有进行特定条件优化。然而,该数据确实表明,NanoString下一代测序方法能够分辨序列的单碱基对。

[0218] 上述研究中利用的详细材料和方法如下:

[0219] 杂交方案探针B加代码集

[0220] · 取25ul要素(194代码集)

[0221] · 将5ul探针B + 互补序列添加至靶标(100uM)

[0222] · 添加15ul Hyb缓冲液(14.56 X SSPE 0.18% Tween 20)

[0223] SSPE (150mM NaCl, NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O 10mM, Na<sub>2</sub>EDTA 10mM)

[0224] · 在冰上孵育10分钟

[0225] · 添加150ul G珠子(40ul 10mg/ml的G珠子加110 ul 5x SSPE 0.1% Tween 20)

[0226] · 在室温孵育10分钟

[0227] · 用0.1SSPE 0.1% Tween 20使用磁性收集器洗涤三次

[0228] · 在45℃下在100ul 0.1x SSPE中洗脱10分钟。

[0229] 靶标杂交方案(750mM NaCl)

[0230] · 取20ul上述洗脱样品

[0231] · 添加10ul hyb缓冲液

[0232] · 添加1ul靶标(100nM生物素化RNA)

[0233] · 在冰上孵育30分钟

[0234] 取15ul并结合至链霉抗生物素蛋白载玻片20分钟,用G钩拉伸流体,用nCounter计数。

[0235] 材料

[0236] 要素194代码集

[0237] 寡核苷酸购自IDT

[0238] SSPE (150mM NaCl, NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O 10mM, Na<sub>2</sub>EDTA 10mM)

[0239] Hyb缓冲液(14.56 X SSPE 0.18% Tween 20)。

[0240] 表 2:12、11、...、8聚体的探针B序列。(SEQ ID NO: 30至SEQ ID NO: 34)

[0241]	GBRYBG	5	GAAGTGTACCCACGCGATGACGTTCTGTCAAGAGTCGCATAATCT	3
	YRBYRG	5	AGAGCTGTACCACAAGAACATCCCTGCTAGCTGAAGGAGGGTCAAAC	3
	YGBYGR	5	GAGAGCTGTACCCCTACGTATATATCCAAGTGGTTATGTCCGACGGC	3
	GBRYGB	5	TGAGAGCTGTACCACCCCTCCAAACGCATTCTTATTGGCAAATGGAA	3
	RYGBRG	5	CTGAGAGCTGTACCCGGGAATCGGCATTTCGCAATTCTTAGGATCTAAA	3

[0242] 表 3:靶序列(以粗体;SEQ ID NO: 35)

[0243]	RNA	5	CAATGTGAGTCTCTTGGTACAGTCTCAGTTAGTCACTCCCTAAG\	3
			Bio TEG\	

[0244] 表 4:10聚体错配的探针B序列(以粗体;SEQ ID NO: 36至SEQ ID NO: 41)

[0245]	10mermis2A	GAGACAGTACCCCTGGTCTAGGTATCTAATTGGTGGGTGGTACT
	10mermis2C	GAGACCGTACCGCTCATTGGAACATACGATTGCGATTACGGAAA
	10mermis2G	GAGACGGTACCTAAAGCTATCCACGAATGTCAAAATGTGGTT
	10mermis1G	GAGAGTGTACCCCAATGCTGCAGTATGTATCCTGATCGTGC
	10mermis1A	GAGAATGTACCCCTACATACCAATGTAAAGTATAGTTAACGCCGT
	10mermis1T	GAGATTGTACCCCTACATATATAGGAAAAGGGAAGGTAGAAGAGCT

[0001]	序列表	
[0002]	<110> NanoString Technologies, Inc.	
[0003]	BEECHEM, Joseph	
[0004]	KHAFIZOV, Rustem	
[0005]	<120> 无酶且无扩增的测序	
[0006]	<130> NATE-025/001	
[0007]	<150> US 62/082,883	
[0008]	<151> 2014-11-21	
[0009]	<160> 104	
[0010]	<170> PatentIn version 3.5	
[0011]	<210> 1	
[0012]	<211> 10	
[0013]	<212> DNA	
[0014]	<213> 人工序列	
[0015]	<220>	
[0016]	<223> 合成的多核苷酸	
[0017]	<400> 1	
[0018]	atacatcttag	10
[0019]	<210> 2	
[0020]	<211> 10	
[0021]	<212> DNA	
[0022]	<213> 人工序列	
[0023]	<220>	
[0024]	<223> 合成的多核苷酸	
[0025]	<400> 2	
[0026]	gatctacata	10
[0027]	<210> 3	
[0028]	<211> 10	
[0029]	<212> DNA	
[0030]	<213> 人工序列	
[0031]	<220>	
[0032]	<223> 合成的多核苷酸	
[0033]	<400> 3	
[0034]	ttaggtaaag	10
[0035]	<210> 4	
[0036]	<211> 10	
[0037]	<212> DNA	
[0038]	<213> 人工序列	

[0039]	<220>	
[0040]	<223>	合成的多核苷酸
[0041]	<400>	4
[0042]	tcttcattac	10
[0043]	<210>	5
[0044]	<211>	10
[0045]	<212>	DNA
[0046]	<213>	人工序列
[0047]	<220>	
[0048]	<223>	合成的多核苷酸
[0049]	<400>	5
[0050]	atgaatctac	10
[0051]	<210>	6
[0052]	<211>	10
[0053]	<212>	DNA
[0054]	<213>	人工序列
[0055]	<220>	
[0056]	<223>	合成的多核苷酸
[0057]	<400>	6
[0058]	tcaatgtatg	10
[0059]	<210>	7
[0060]	<211>	10
[0061]	<212>	DNA
[0062]	<213>	人工序列
[0063]	<220>	
[0064]	<223>	合成的多核苷酸
[0065]	<400>	7
[0066]	aattgagtagc	10
[0067]	<210>	8
[0068]	<211>	10
[0069]	<212>	DNA
[0070]	<213>	人工序列
[0071]	<220>	
[0072]	<223>	合成的多核苷酸
[0073]	<400>	8
[0074]	atgttaatgg	10
[0075]	<210>	9
[0076]	<211>	10
[0077]	<212>	DNA

[0078]	<213>	人工序列	
[0079]	<220>		
[0080]	<223>	合成的多核苷酸	
[0081]	<400>	9	
[0082]	aatttaggatg		10
[0083]	<210>	10	
[0084]	<211>	10	
[0085]	<212>	DNA	
[0086]	<213>	人工序列	
[0087]	<220>		
[0088]	<223>	合成的多核苷酸	
[0089]	<400>	10	
[0090]	ataatggatc		10
[0091]	<210>	11	
[0092]	<211>	10	
[0093]	<212>	DNA	
[0094]	<213>	人工序列	
[0095]	<220>		
[0096]	<223>	合成的多核苷酸	
[0097]	<400>	11	
[0098]	taataagggtg		10
[0099]	<210>	12	
[0100]	<211>	10	
[0101]	<212>	DNA	
[0102]	<213>	人工序列	
[0103]	<220>		
[0104]	<223>	合成的多核苷酸	
[0105]	<400>	12	
[0106]	tagtttagagc		10
[0107]	<210>	13	
[0108]	<211>	10	
[0109]	<212>	DNA	
[0110]	<213>	人工序列	
[0111]	<220>		
[0112]	<223>	合成的多核苷酸	
[0113]	<400>	13	
[0114]	atagagaagg		10
[0115]	<210>	14	
[0116]	<211>	10	

[0117]	<212>	DNA	
[0118]	<213>	人工序列	
[0119]	<220>		
[0120]	<223>	合成的多核苷酸	
[0121]	<400>	14	
[0122]	ttgatgatac		10
[0123]	<210>	15	
[0124]	<211>	10	
[0125]	<212>	DNA	
[0126]	<213>	人工序列	
[0127]	<220>		
[0128]	<223>	合成的多核苷酸	
[0129]	<400>	15	
[0130]	atagtgattc		10
[0131]	<210>	16	
[0132]	<211>	10	
[0133]	<212>	DNA	
[0134]	<213>	人工序列	
[0135]	<220>		
[0136]	<223>	合成的多核苷酸	
[0137]	<400>	16	
[0138]	tataacgatg		10
[0139]	<210>	17	
[0140]	<211>	10	
[0141]	<212>	DNA	
[0142]	<213>	人工序列	
[0143]	<220>		
[0144]	<223>	合成的多核苷酸	
[0145]	<400>	17	
[0146]	ttaagtttag		10
[0147]	<210>	18	
[0148]	<211>	10	
[0149]	<212>	DNA	
[0150]	<213>	人工序列	
[0151]	<220>		
[0152]	<223>	合成的多核苷酸	
[0153]	<400>	18	
[0154]	atacgttatg		10
[0155]	<210>	19	

[0156]	<211>	10	
[0157]	<212>	DNA	
[0158]	<213>	人工序列	
[0159]	<220>		
[0160]	<223>	合成的多核苷酸	
[0161]	<400>	19	
[0162]	tgtactata	g	10
[0163]	<210>	20	
[0164]	<211>	10	
[0165]	<212>	DNA	
[0166]	<213>	人工序列	
[0167]	<220>		
[0168]	<223>	合成的多核苷酸	
[0169]	<400>	20	
[0170]	ttaacaagg	t	10
[0171]	<210>	21	
[0172]	<211>	10	
[0173]	<212>	DNA	
[0174]	<213>	人工序列	
[0175]	<220>		
[0176]	<223>	合成的多核苷酸	
[0177]	<400>	21	
[0178]	aactatgtac		10
[0179]	<210>	22	
[0180]	<211>	10	
[0181]	<212>	DNA	
[0182]	<213>	人工序列	
[0183]	<220>		
[0184]	<223>	合成的多核苷酸	
[0185]	<400>	22	
[0186]	taactatgac		10
[0187]	<210>	23	
[0188]	<211>	10	
[0189]	<212>	DNA	
[0190]	<213>	人工序列	
[0191]	<220>		
[0192]	<223>	合成的多核苷酸	
[0193]	<400>	23	
[0194]	actaatgttc		10

[0195]	<210>	24	
[0196]	<211>	10	
[0197]	<212>	DNA	
[0198]	<213>	人工序列	
[0199]	<220>		
[0200]	<223>	合成的多核苷酸	
[0201]	<400>	24	
[0202]	tcattgaatg		10
[0203]	<210>	25	
[0204]	<211>	14	
[0205]	<212>	DNA	
[0206]	<213>	人工序列	
[0207]	<220>		
[0208]	<223>	合成的多核苷酸	
[0209]	<400>	25	
[0210]	ctgtctcatc tctt		14
[0211]	<210>	26	
[0212]	<211>	26	
[0213]	<212>	DNA	
[0214]	<213>	人工序列	
[0215]	<220>		
[0216]	<223>	合成的多核苷酸	
[0217]	<400>	26	
[0218]	ctgtctcatc tcttgctgca tcctgt		26
[0219]	<210>	27	
[0220]	<211>	38	
[0221]	<212>	DNA	
[0222]	<213>	人工序列	
[0223]	<220>		
[0224]	<223>	合成的多核苷酸	
[0225]	<400>	27	
[0226]	ctgtctcatc tcttgctgca tcctgtcggt tcacgttg		38
[0227]	<210>	28	
[0228]	<211>	36	
[0229]	<212>	DNA	
[0230]	<213>	人工序列	
[0231]	<220>		
[0232]	<223>	合成的多核苷酸	
[0233]	<400>	28	

[0234]	ctgtctcatc ttgctgcatt ctgtcggttc acgttg	36
[0235]	<210> 29	
[0236]	<211> 36	
[0237]	<212> DNA	
[0238]	<213> 人工序列	
[0239]	<220>	
[0240]	<223> 合成的多核苷酸	
[0241]	<400> 29	
[0242]	ctgtctcatt ttgctgcatt ctgtccgttc acgttg	36
[0243]	<210> 30	
[0244]	<211> 43	
[0245]	<212> DNA	
[0246]	<213> 人工序列	
[0247]	<220>	
[0248]	<223> 合成的多核苷酸	
[0249]	<400> 30	
[0250]	gactgtaccc acgcgatgac gttcgtcaag agtcgcataa tct	43
[0251]	<210> 31	
[0252]	<211> 44	
[0253]	<212> DNA	
[0254]	<213> 人工序列	
[0255]	<220>	
[0256]	<223> 合成的多核苷酸	
[0257]	<400> 31	
[0258]	agactgtacc acaagaatcc ctgcttagctg aaggagggtc aaac	44
[0259]	<210> 32	
[0260]	<211> 45	
[0261]	<212> DNA	
[0262]	<213> 人工序列	
[0263]	<220>	
[0264]	<223> 合成的多核苷酸	
[0265]	<400> 32	
[0266]	gagactgtac cctacgtata tatccaagtg gttatgtccg acggc	45
[0267]	<210> 33	
[0268]	<211> 46	
[0269]	<212> DNA	
[0270]	<213> 人工序列	
[0271]	<220>	
[0272]	<223> 合成的多核苷酸	

[0273]	<400>	33	
[0274]	ttagactgta ccaccctcc aaacgcattc ttattggcaa atggaa		46
[0275]	<210>	34	
[0276]	<211>	47	
[0277]	<212>	DNA	
[0278]	<213>	人工序列	
[0279]	<220>		
[0280]	<223>	合成的多核苷酸	
[0281]	<400>	34	
[0282]	ctgagactgt acccggaat cggcattcg cattcttagg atctaaa		47
[0283]	<210>	35	
[0284]	<211>	44	
[0285]	<212>	DNA	
[0286]	<213>	人工序列	
[0287]	<220>		
[0288]	<223>	合成的多核苷酸	
[0289]	<400>	35	
[0290]	caatgtgagt ctcttggtagt agtctcagtt agtcaactccc taag		44
[0291]	<210>	36	
[0292]	<211>	45	
[0293]	<212>	DNA	
[0294]	<213>	人工序列	
[0295]	<220>		
[0296]	<223>	合成的多核苷酸	
[0297]	<400>	36	
[0298]	gagacagtagc cctggcttag gtatctaatt cgtgggtcgg gtact		45
[0299]	<210>	37	
[0300]	<211>	45	
[0301]	<212>	DNA	
[0302]	<213>	人工序列	
[0303]	<220>		
[0304]	<223>	合成的多核苷酸	
[0305]	<400>	37	
[0306]	gagaccgtac cgctcatttt gaacatacga ttgcgattac ggaaa		45
[0307]	<210>	38	
[0308]	<211>	45	
[0309]	<212>	DNA	
[0310]	<213>	人工序列	
[0311]	<220>		

[0312]	<223>	合成的多核苷酸	
[0313]	<400>	38	
[0314]	gagacggta cttaaagcta tccacgaatg tcaaaaatgt ggttt		45
[0315]	<210>	39	
[0316]	<211>	45	
[0317]	<212>	DNA	
[0318]	<213>	人工序列	
[0319]	<220>		
[0320]	<223>	合成的多核苷酸	
[0321]	<400>	39	
[0322]	gagagtgtac ccaatgctt cagttatgtat cctgatcgtg cgtgc		45
[0323]	<210>	40	
[0324]	<211>	45	
[0325]	<212>	DNA	
[0326]	<213>	人工序列	
[0327]	<220>		
[0328]	<223>	合成的多核苷酸	
[0329]	<400>	40	
[0330]	gagaatgtac ctcatacca atgtaaagta tagttaacgc cctgt		45
[0331]	<210>	41	
[0332]	<211>	45	
[0333]	<212>	DNA	
[0334]	<213>	人工序列	
[0335]	<220>		
[0336]	<223>	合成的多核苷酸	
[0337]	<400>	41	
[0338]	gagattgtac cttacatata tagaaaaagg gaaggtagaa gagct		45
[0339]	<210>	42	
[0340]	<211>	12	
[0341]	<212>	DNA	
[0342]	<213>	人工序列	
[0343]	<220>		
[0344]	<223>	合成的多核苷酸	
[0345]	<400>	42	
[0346]	cacgaacgtc ag		12
[0347]	<210>	43	
[0348]	<211>	12	
[0349]	<212>	DNA	
[0350]	<213>	人工序列	

[0351]	<220>	
[0352]	<223>	合成的多核苷酸
[0353]	<400>	43
[0354]	catcgcatgc	ct
[0355]	<210>	44
[0356]	<211>	12
[0357]	<212>	DNA
[0358]	<213>	人工序列
[0359]	<220>	
[0360]	<223>	合成的多核苷酸
[0361]	<400>	44
[0362]	gtcatctcct	ac
[0363]	<210>	45
[0364]	<211>	12
[0365]	<212>	DNA
[0366]	<213>	人工序列
[0367]	<220>	
[0368]	<223>	合成的多核苷酸
[0369]	<400>	45
[0370]	gtcatccgct	ac
[0371]	<210>	46
[0372]	<211>	12
[0373]	<212>	DNA
[0374]	<213>	人工序列
[0375]	<220>	
[0376]	<223>	合成的多核苷酸
[0377]	<400>	46
[0378]	gtcatcgact	ac
[0379]	<210>	47
[0380]	<211>	12
[0381]	<212>	DNA
[0382]	<213>	人工序列
[0383]	<220>	
[0384]	<223>	合成的多核苷酸
[0385]	<400>	47
[0386]	gtcatcttct	ac
[0387]	<210>	48
[0388]	<211>	12
[0389]	<212>	DNA

[0390]	<213>	人工序列	
[0391]	<220>		
[0392]	<223>	合成的多核苷酸	
[0393]	<400>	48	
[0394]	gtcatcacct	ac	12
[0395]	<210>	49	
[0396]	<211>	12	
[0397]	<212>	DNA	
[0398]	<213>	人工序列	
[0399]	<220>		
[0400]	<223>	合成的多核苷酸	
[0401]	<400>	49	
[0402]	gtcatcaactc	ac	12
[0403]	<210>	50	
[0404]	<211>	12	
[0405]	<212>	DNA	
[0406]	<213>	人工序列	
[0407]	<220>		
[0408]	<223>	合成的多核苷酸	
[0409]	<400>	50	
[0410]	gtcatcttcg	ac	12
[0411]	<210>	51	
[0412]	<211>	12	
[0413]	<212>	DNA	
[0414]	<213>	人工序列	
[0415]	<220>		
[0416]	<223>	合成的多核苷酸	
[0417]	<400>	51	
[0418]	gtcatcaact	ac	12
[0419]	<210>	52	
[0420]	<211>	12	
[0421]	<212>	DNA	
[0422]	<213>	人工序列	
[0423]	<220>		
[0424]	<223>	合成的多核苷酸	
[0425]	<400>	52	
[0426]	gtcatccgta	ac	12
[0427]	<210>	53	
[0428]	<211>	12	

[0429]	<212>	DNA	
[0430]	<213>	人工序列	
[0431]	<220>		
[0432]	<223>	合成的多核苷酸	
[0433]	<400>	53	
[0434]	gtcatccgaa	ac	12
[0435]	<210>	54	
[0436]	<211>	12	
[0437]	<212>	DNA	
[0438]	<213>	人工序列	
[0439]	<220>		
[0440]	<223>	合成的多核苷酸	
[0441]	<400>	54	
[0442]	gtcatcacaa	ac	12
[0443]	<210>	55	
[0444]	<211>	12	
[0445]	<212>	DNA	
[0446]	<213>	人工序列	
[0447]	<220>		
[0448]	<223>	合成的多核苷酸	
[0449]	<400>	55	
[0450]	gtcatttgc	ac	12
[0451]	<210>	56	
[0452]	<211>	12	
[0453]	<212>	DNA	
[0454]	<213>	人工序列	
[0455]	<220>		
[0456]	<223>	合成的多核苷酸	
[0457]	<400>	56	
[0458]	gtcatttgc	ct	12
[0459]	<210>	57	
[0460]	<211>	12	
[0461]	<212>	DNA	
[0462]	<213>	人工序列	
[0463]	<220>		
[0464]	<223>	合成的多核苷酸	
[0465]	<400>	57	
[0466]	gtcatccgtc	ct	12
[0467]	<210>	58	

[0468]	<211>	12	
[0469]	<212>	DNA	
[0470]	<213>	人工序列	
[0471]	<220>		
[0472]	<223>	合成的多核苷酸	
[0473]	<400>	58	
[0474]	ctttcacct	ct	12
[0475]	<210>	59	
[0476]	<211>	12	
[0477]	<212>	DNA	
[0478]	<213>	人工序列	
[0479]	<220>		
[0480]	<223>	合成的多核苷酸	
[0481]	<400>	59	
[0482]	ctttcctct	ct	12
[0483]	<210>	60	
[0484]	<211>	12	
[0485]	<212>	DNA	
[0486]	<213>	人工序列	
[0487]	<220>		
[0488]	<223>	合成的多核苷酸	
[0489]	<400>	60	
[0490]	ctttcgact	ct	12
[0491]	<210>	61	
[0492]	<211>	12	
[0493]	<212>	DNA	
[0494]	<213>	人工序列	
[0495]	<220>		
[0496]	<223>	合成的多核苷酸	
[0497]	<400>	61	
[0498]	ctttctgct	ct	12
[0499]	<210>	62	
[0500]	<211>	12	
[0501]	<212>	DNA	
[0502]	<213>	人工序列	
[0503]	<220>		
[0504]	<223>	合成的多核苷酸	
[0505]	<400>	62	
[0506]	ctttctgta	ct	12

[0507]	<210>	63	
[0508]	<211>	12	
[0509]	<212>	DNA	
[0510]	<213>	人工序列	
[0511]	<220>		
[0512]	<223>	合成的多核苷酸	
[0513]	<400>	63	
[0514]	cttttctgtg ct		12
[0515]	<210>	64	
[0516]	<211>	12	
[0517]	<212>	DNA	
[0518]	<213>	人工序列	
[0519]	<220>		
[0520]	<223>	合成的多核苷酸	
[0521]	<400>	64	
[0522]	cttttctgtc ct		12
[0523]	<210>	65	
[0524]	<211>	12	
[0525]	<212>	DNA	
[0526]	<213>	人工序列	
[0527]	<220>		
[0528]	<223>	合成的多核苷酸	
[0529]	<400>	65	
[0530]	cttttcactc ct		12
[0531]	<210>	66	
[0532]	<211>	12	
[0533]	<212>	DNA	
[0534]	<213>	人工序列	
[0535]	<220>		
[0536]	<223>	合成的多核苷酸	
[0537]	<400>	66	
[0538]	cttttcgttc ct		12
[0539]	<210>	67	
[0540]	<211>	12	
[0541]	<212>	DNA	
[0542]	<213>	人工序列	
[0543]	<220>		
[0544]	<223>	合成的多核苷酸	
[0545]	<400>	67	

[0546]	ctttcgtac	ct	12
[0547]	<210>	68	
[0548]	<211>	12	
[0549]	<212>	DNA	
[0550]	<213>	人工序列	
[0551]	<220>		
[0552]	<223>	合成的多核苷酸	
[0553]	<400>	68	
[0554]	ctttccgtc	ct	12
[0555]	<210>	69	
[0556]	<211>	12	
[0557]	<212>	DNA	
[0558]	<213>	人工序列	
[0559]	<220>		
[0560]	<223>	合成的多核苷酸	
[0561]	<400>	69	
[0562]	ctttctgac	ct	12
[0563]	<210>	70	
[0564]	<211>	12	
[0565]	<212>	DNA	
[0566]	<213>	人工序列	
[0567]	<220>		
[0568]	<223>	合成的多核苷酸	
[0569]	<400>	70	
[0570]	aggcatgcga	tg	12
[0571]	<210>	71	
[0572]	<211>	12	
[0573]	<212>	DNA	
[0574]	<213>	人工序列	
[0575]	<220>		
[0576]	<223>	合成的多核苷酸	
[0577]	<400>	71	
[0578]	aggcattgtg	ct	12
[0579]	<210>	72	
[0580]	<211>	12	
[0581]	<212>	DNA	
[0582]	<213>	人工序列	
[0583]	<220>		
[0584]	<223>	合成的多核苷酸	

---

[0585]	<400>	72	
[0586]	aggcattgct ct		12
[0587]	<210>	73	
[0588]	<211>	12	
[0589]	<212>	DNA	
[0590]	<213>	人工序列	
[0591]	<220>		
[0592]	<223>	合成的多核苷酸	
[0593]	<400>	73	
[0594]	aggcatttct ac		12
[0595]	<210>	74	
[0596]	<211>	12	
[0597]	<212>	DNA	
[0598]	<213>	人工序列	
[0599]	<220>		
[0600]	<223>	合成的多核苷酸	
[0601]	<400>	74	
[0602]	aggcataacct ac		12
[0603]	<210>	75	
[0604]	<211>	12	
[0605]	<212>	DNA	
[0606]	<213>	人工序列	
[0607]	<220>		
[0608]	<223>	合成的多核苷酸	
[0609]	<400>	75	
[0610]	aggcatttgc ac		12
[0611]	<210>	76	
[0612]	<211>	12	
[0613]	<212>	DNA	
[0614]	<213>	人工序列	
[0615]	<220>		
[0616]	<223>	合成的多核苷酸	
[0617]	<400>	76	
[0618]	aggcatcgtc ct		12
[0619]	<210>	77	
[0620]	<211>	12	
[0621]	<212>	DNA	
[0622]	<213>	人工序列	
[0623]	<220>		

[0624]	<223>	合成的多核苷酸	
[0625]	<400>	77	
[0626]	tcctgtcggt	tc	12
[0627]	<210>	78	
[0628]	<211>	12	
[0629]	<212>	DNA	
[0630]	<213>	人工序列	
[0631]	<220>		
[0632]	<223>	合成的多核苷酸	
[0633]	<400>	78	
[0634]	gttcaatgct	ct	12
[0635]	<210>	79	
[0636]	<211>	12	
[0637]	<212>	DNA	
[0638]	<213>	人工序列	
[0639]	<220>		
[0640]	<223>	合成的多核苷酸	
[0641]	<400>	79	
[0642]	attcggtgct	ct	12
[0643]	<210>	80	
[0644]	<211>	12	
[0645]	<212>	DNA	
[0646]	<213>	人工序列	
[0647]	<220>		
[0648]	<223>	合成的多核苷酸	
[0649]	<400>	80	
[0650]	gatgcctgct	ct	12
[0651]	<210>	81	
[0652]	<211>	12	
[0653]	<212>	DNA	
[0654]	<213>	人工序列	
[0655]	<220>		
[0656]	<223>	合成的多核苷酸	
[0657]	<400>	81	
[0658]	tttgcttgct	ct	12
[0659]	<210>	82	
[0660]	<211>	100	
[0661]	<212>	DNA	
[0662]	<213>	人工序列	

[0663]	<220>	
[0664]	<223> 合成的多核苷酸	
[0665]	<400> 82	
[0666]	ttcactgttag ctgtctcatt ttgctgcatc ctgtccgttc acgttggagc ttgtcatccg	60
[0667]	tcctcttttc actccttaggc atttgcctat tcggcgtcct	100
[0668]	<210> 83	
[0669]	<211> 10	
[0670]	<212> DNA	
[0671]	<213> 人工序列	
[0672]	<220>	
[0673]	<223> 合成的多核苷酸	
[0674]	<400> 83	
[0675]	cgatctgggtt	10
[0676]	<210> 84	
[0677]	<211> 10	
[0678]	<212> DNA	
[0679]	<213> 人工序列	
[0680]	<220>	
[0681]	<223> 合成的多核苷酸	
[0682]	<400> 84	
[0683]	cgatctgggtt	10
[0684]	<210> 85	
[0685]	<211> 10	
[0686]	<212> DNA	
[0687]	<213> 人工序列	
[0688]	<220>	
[0689]	<223> 合成的多核苷酸	
[0690]	<400> 85	
[0691]	gctagaccaa	10
[0692]	<210> 86	
[0693]	<211> 10	
[0694]	<212> DNA	
[0695]	<213> 人工序列	
[0696]	<220>	
[0697]	<223> 合成的多核苷酸	
[0698]	<400> 86	
[0699]	gctggaccaa	10
[0700]	<210> 87	
[0701]	<211> 10	

[0702]	<212>	DNA	
[0703]	<213>	人工序列	
[0704]	<220>		
[0705]	<223>	合成的多核苷酸	
[0706]	<400>	87	
[0707]	gctcgaccaa		10
[0708]	<210>	88	
[0709]	<211>	10	
[0710]	<212>	DNA	
[0711]	<213>	人工序列	
[0712]	<220>		
[0713]	<223>	合成的多核苷酸	
[0714]	<400>	88	
[0715]	gcttgaccaa		10
[0716]	<210>	89	
[0717]	<211>	10	
[0718]	<212>	DNA	
[0719]	<213>	人工序列	
[0720]	<220>		
[0721]	<223>	合成的多核苷酸	
[0722]	<400>	89	
[0723]	gagactgtac		10
[0724]	<210>	90	
[0725]	<211>	10	
[0726]	<212>	DNA	
[0727]	<213>	人工序列	
[0728]	<220>		
[0729]	<223>	合成的多核苷酸	
[0730]	<400>	90	
[0731]	gagacagtac		10
[0732]	<210>	91	
[0733]	<211>	10	
[0734]	<212>	DNA	
[0735]	<213>	人工序列	
[0736]	<220>		
[0737]	<223>	合成的多核苷酸	
[0738]	<400>	91	
[0739]	gagaccgtac		10
[0740]	<210>	92	

[0741]	<211>	10	
[0742]	<212>	DNA	
[0743]	<213>	人工序列	
[0744]	<220>		
[0745]	<223>	合成的多核苷酸	
[0746]	<400>	92	
[0747]	gagacggtag		10
[0748]	<210>	93	
[0749]	<211>	10	
[0750]	<212>	DNA	
[0751]	<213>	人工序列	
[0752]	<220>		
[0753]	<223>	合成的多核苷酸	
[0754]	<400>	93	
[0755]	gagagtgtac		10
[0756]	<210>	94	
[0757]	<211>	10	
[0758]	<212>	DNA	
[0759]	<213>	人工序列	
[0760]	<220>		
[0761]	<223>	合成的多核苷酸	
[0762]	<400>	94	
[0763]	gagaatgtac		10
[0764]	<210>	95	
[0765]	<211>	10	
[0766]	<212>	DNA	
[0767]	<213>	人工序列	
[0768]	<220>		
[0769]	<223>	合成的多核苷酸	
[0770]	<400>	95	
[0771]	gagattgtac		10
[0772]	<210>	96	
[0773]	<211>	10	
[0774]	<212>	DNA	
[0775]	<213>	人工序列	
[0776]	<220>		
[0777]	<223>	合成的多核苷酸	
[0778]	<400>	96	
[0779]	gagactgtac		10

[0780]	<210>	97	
[0781]	<211>	10	
[0782]	<212>	DNA	
[0783]	<213>	人工序列	
[0784]	<220>		
[0785]	<223>	合成的多核苷酸	
[0786]	<400>	97	
[0787]	gagattgtac		10
[0788]	<210>	98	
[0789]	<211>	10	
[0790]	<212>	DNA	
[0791]	<213>	人工序列	
[0792]	<220>		
[0793]	<223>	合成的多核苷酸	
[0794]	<400>	98	
[0795]	gagaccgtac		10
[0796]	<210>	99	
[0797]	<211>	10	
[0798]	<212>	DNA	
[0799]	<213>	人工序列	
[0800]	<220>		
[0801]	<223>	合成的多核苷酸	
[0802]	<400>	99	
[0803]	gagagtgtac		10
[0804]	<210>	100	
[0805]	<211>	12	
[0806]	<212>	DNA	
[0807]	<213>	人工序列	
[0808]	<220>		
[0809]	<223>	合成的多核苷酸	
[0810]	<400>	100	
[0811]	ctgagactgt ac		12
[0812]	<210>	101	
[0813]	<211>	11	
[0814]	<212>	DNA	
[0815]	<213>	人工序列	
[0816]	<220>		
[0817]	<223>	合成的多核苷酸	
[0818]	<400>	101	

---

[0819]	tgagactgtac	c	11
[0820]	<210>	102	
[0821]	<211>	10	
[0822]	<212>	DNA	
[0823]	<213>	人工序列	
[0824]	<220>		
[0825]	<223>	合成的多核苷酸	
[0826]	<400>	102	
[0827]	gagactgtac		10
[0828]	<210>	103	
[0829]	<211>	12	
[0830]	<212>	DNA	
[0831]	<213>	人工序列	
[0832]	<220>		
[0833]	<223>	合成的多核苷酸	
[0834]	<400>	103	
[0835]	catgtcagag	tc	12
[0836]	<210>	104	
[0837]	<211>	11	
[0838]	<212>	DNA	
[0839]	<213>	人工序列	
[0840]	<220>		
[0841]	<223>	合成的多核苷酸	
[0842]	<400>	104	
[0843]	catgtcagag	t	11

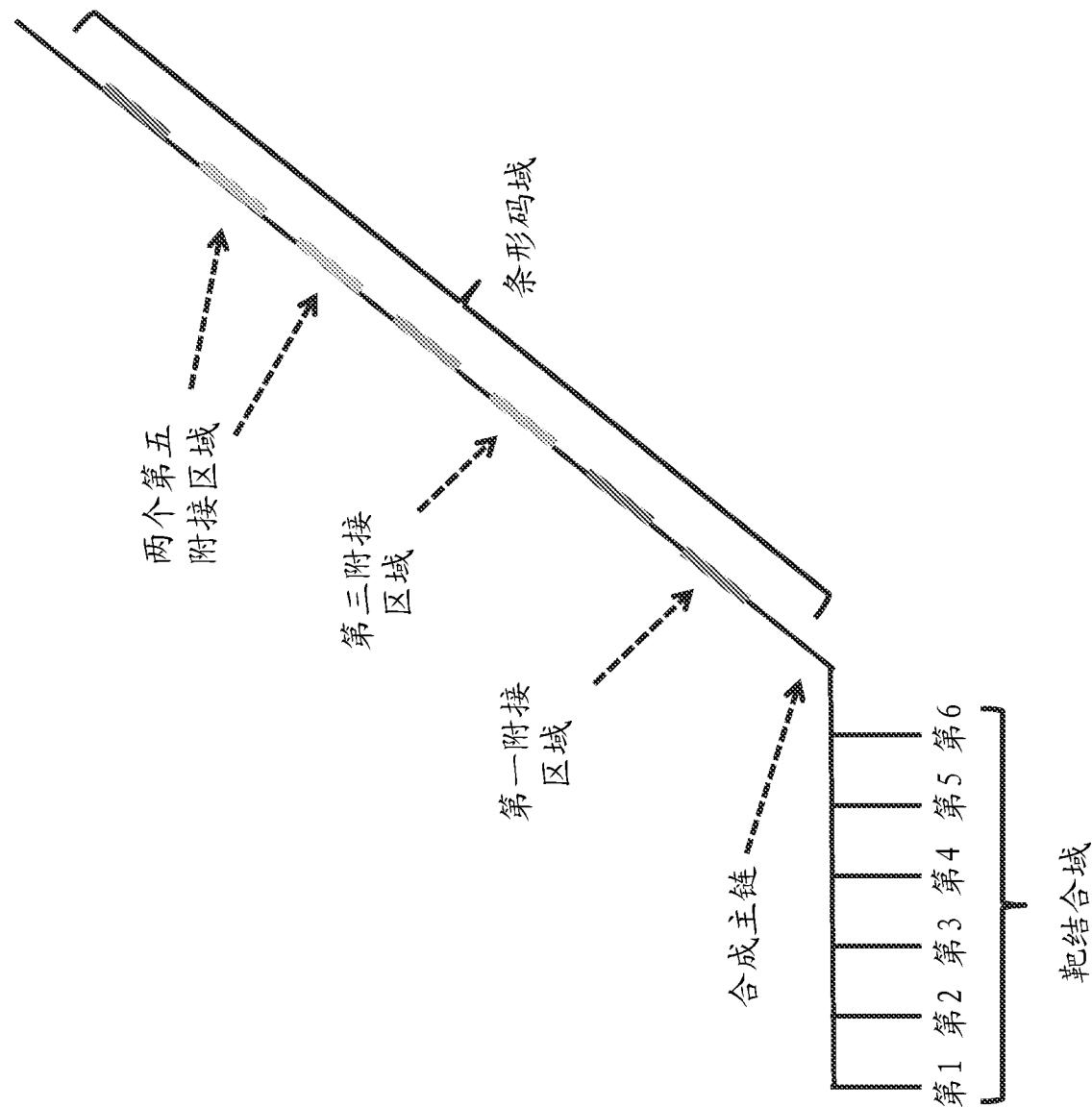


图 1

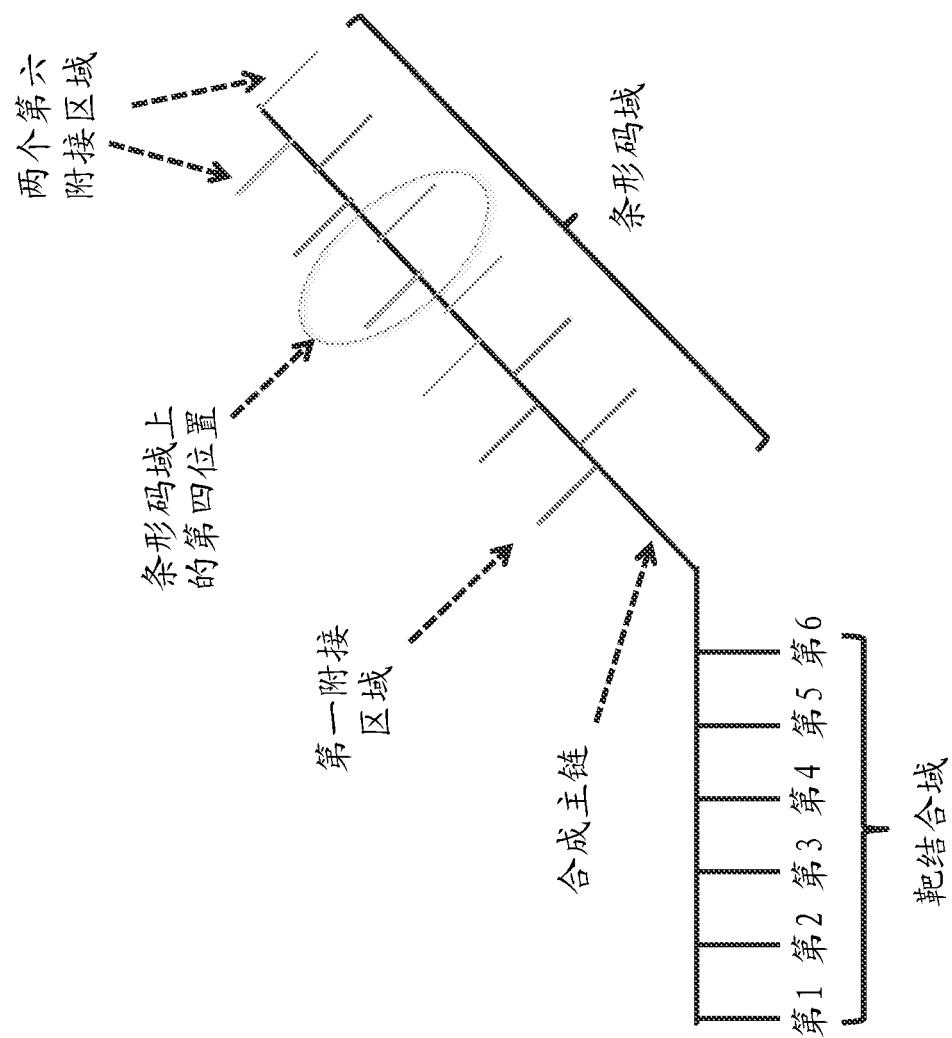


图 2

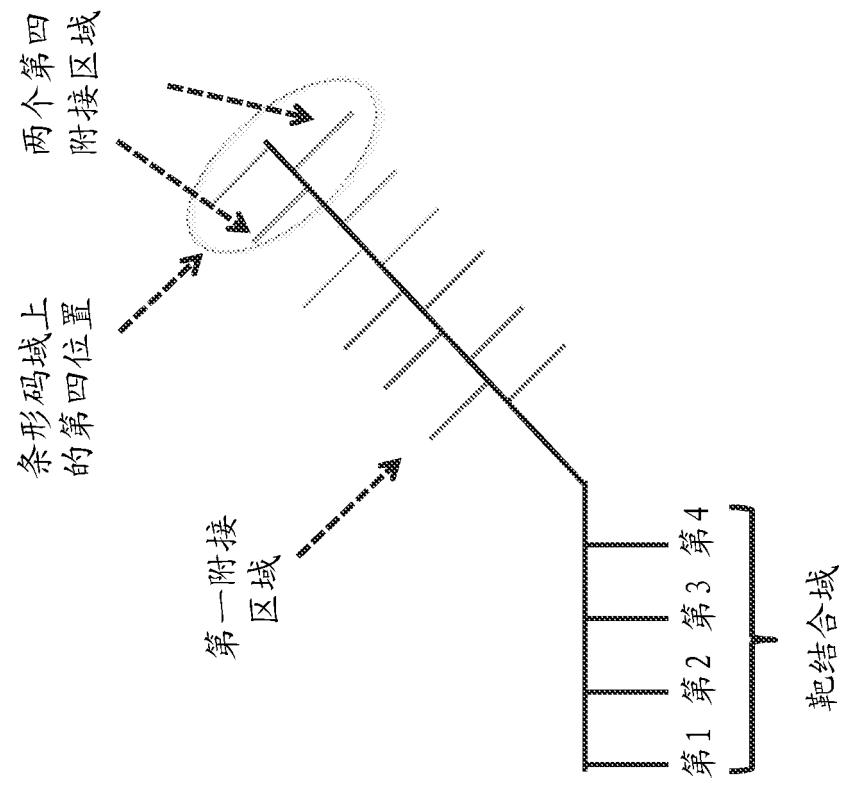


图 3

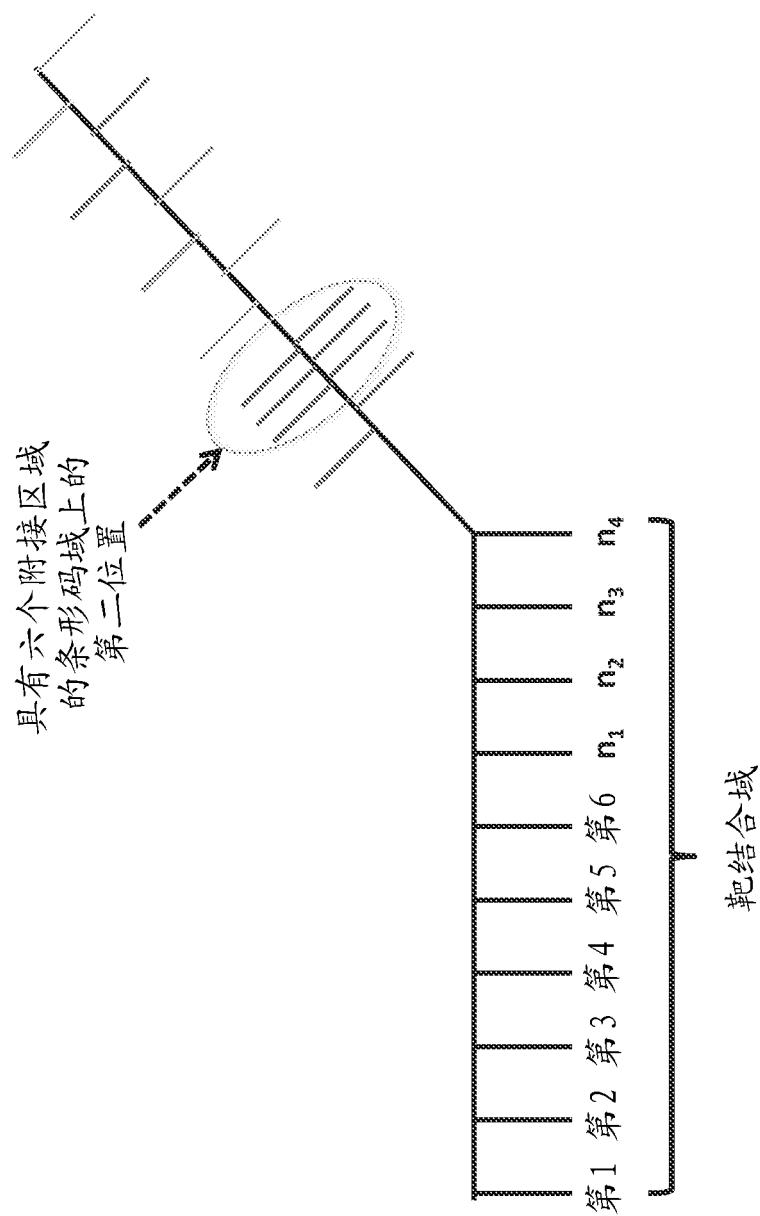


图 4

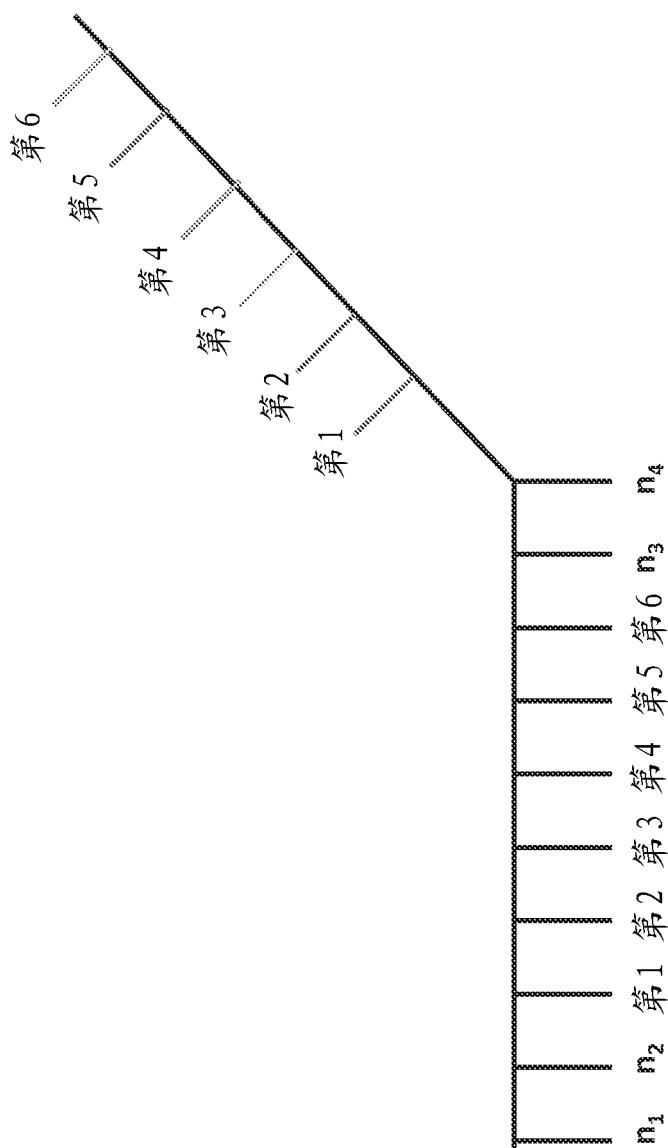


图 5

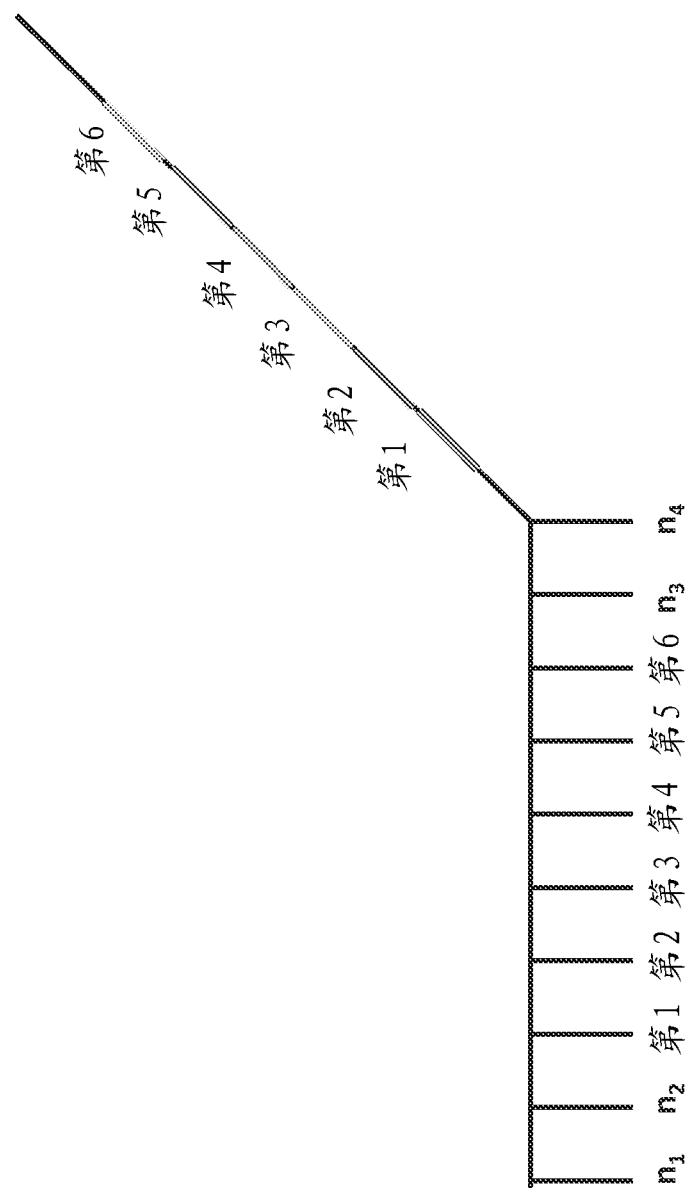


图 6A

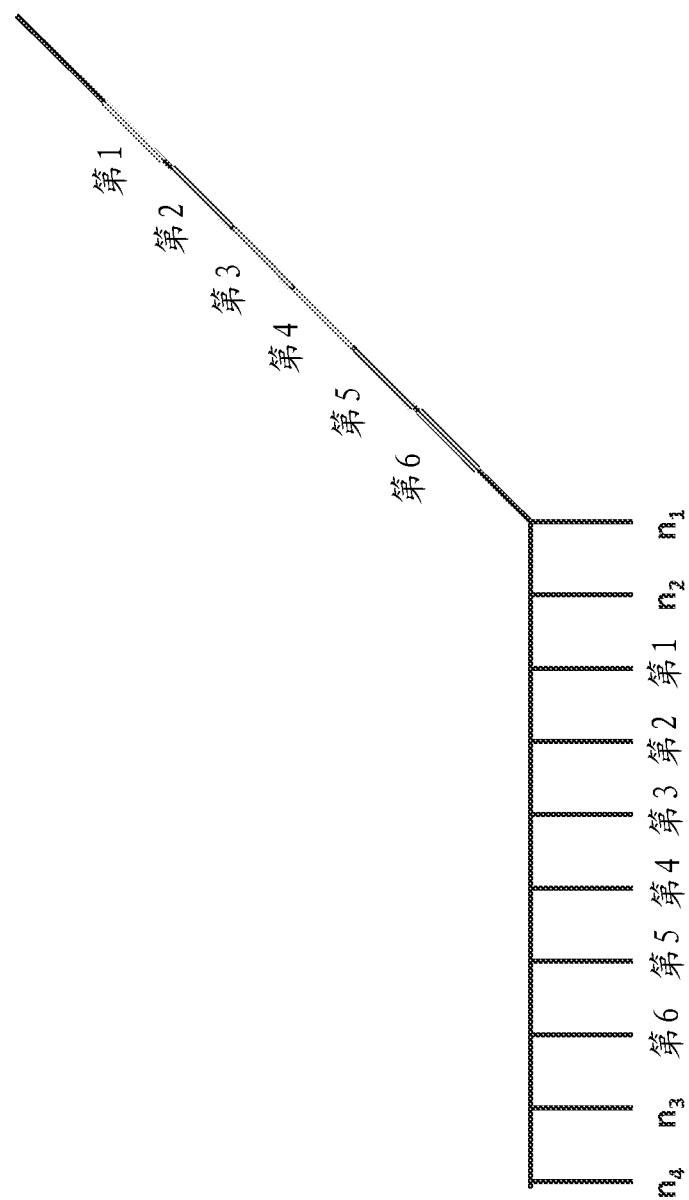


图 6B

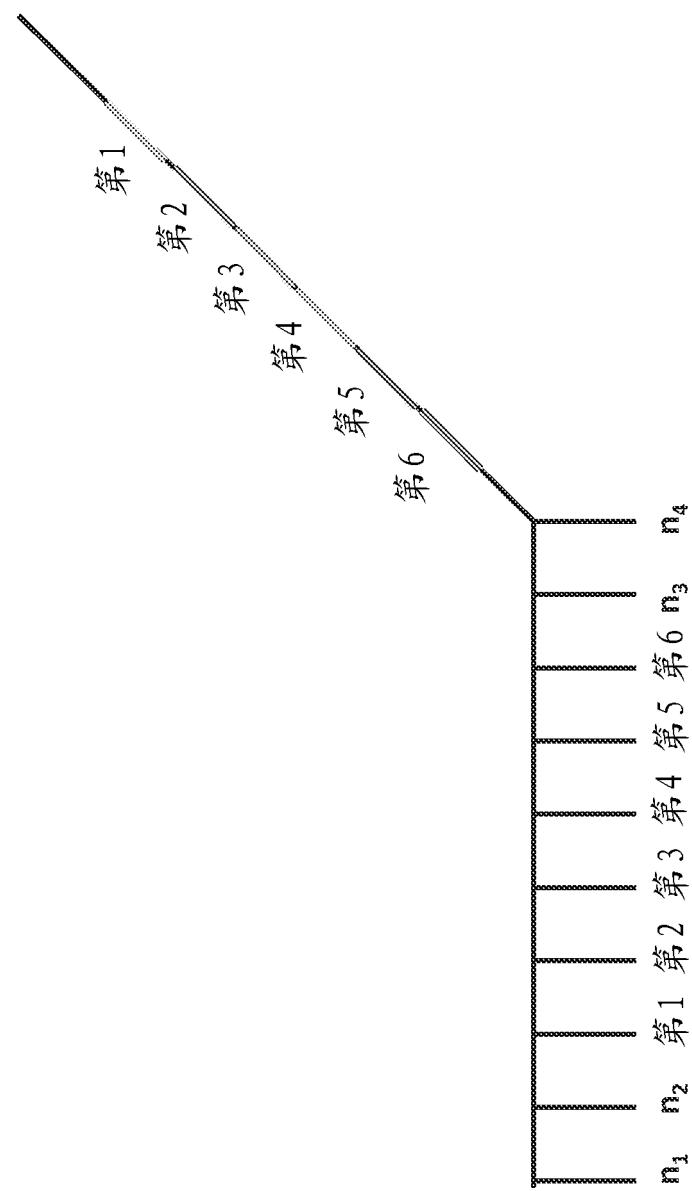


图 6C

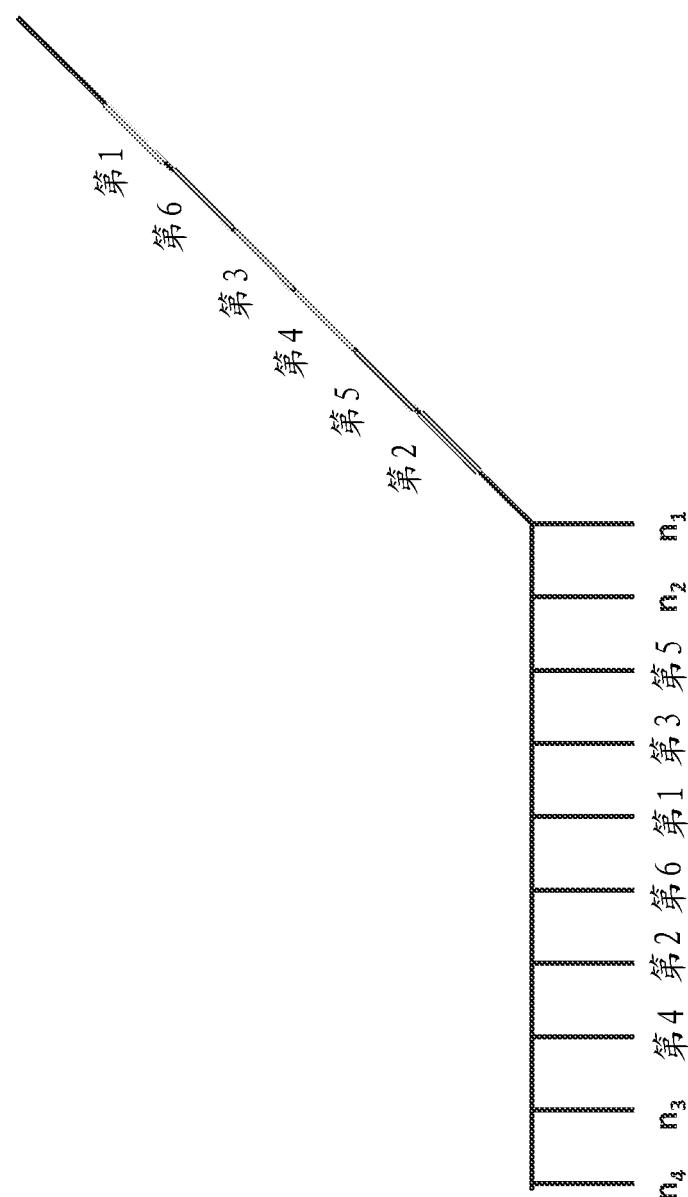


图 6D

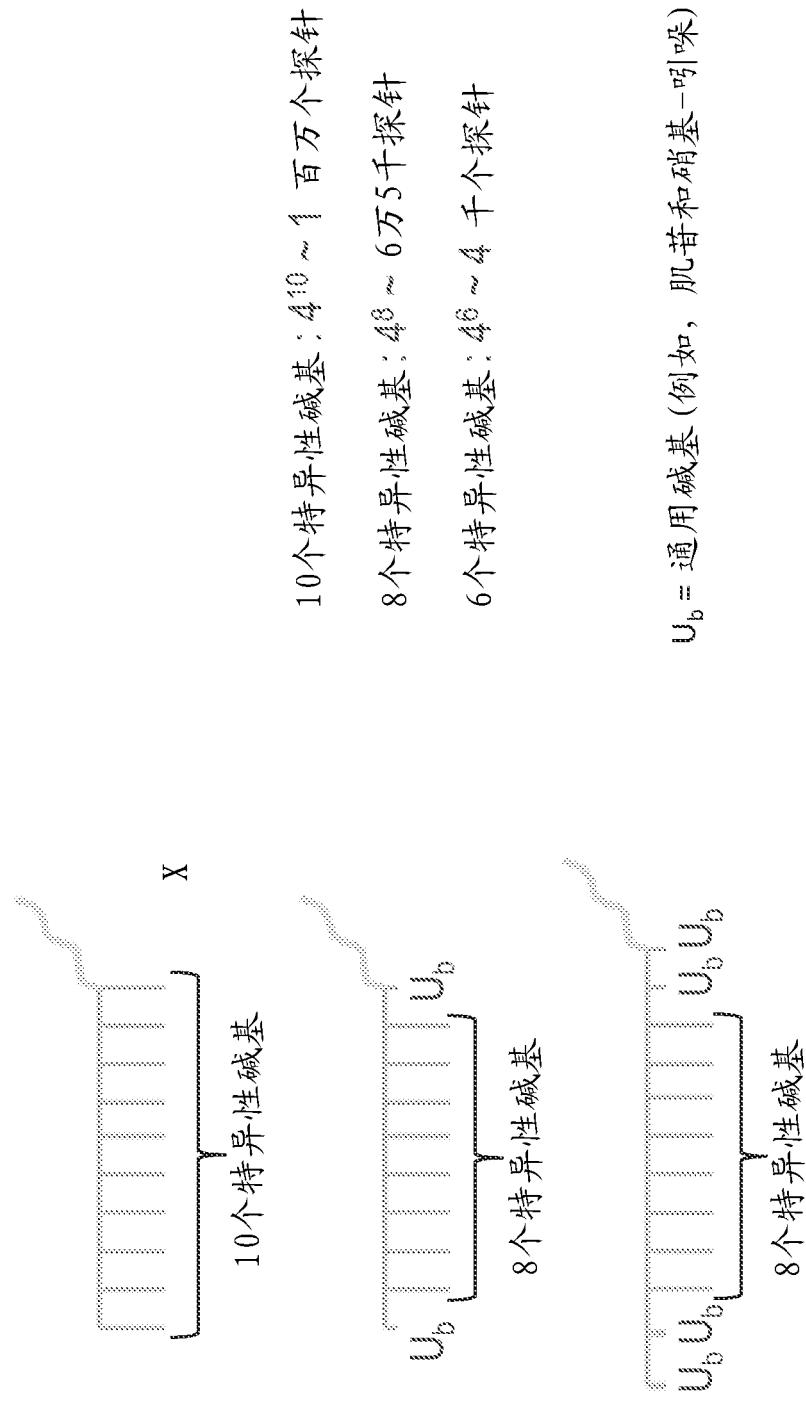


图 7

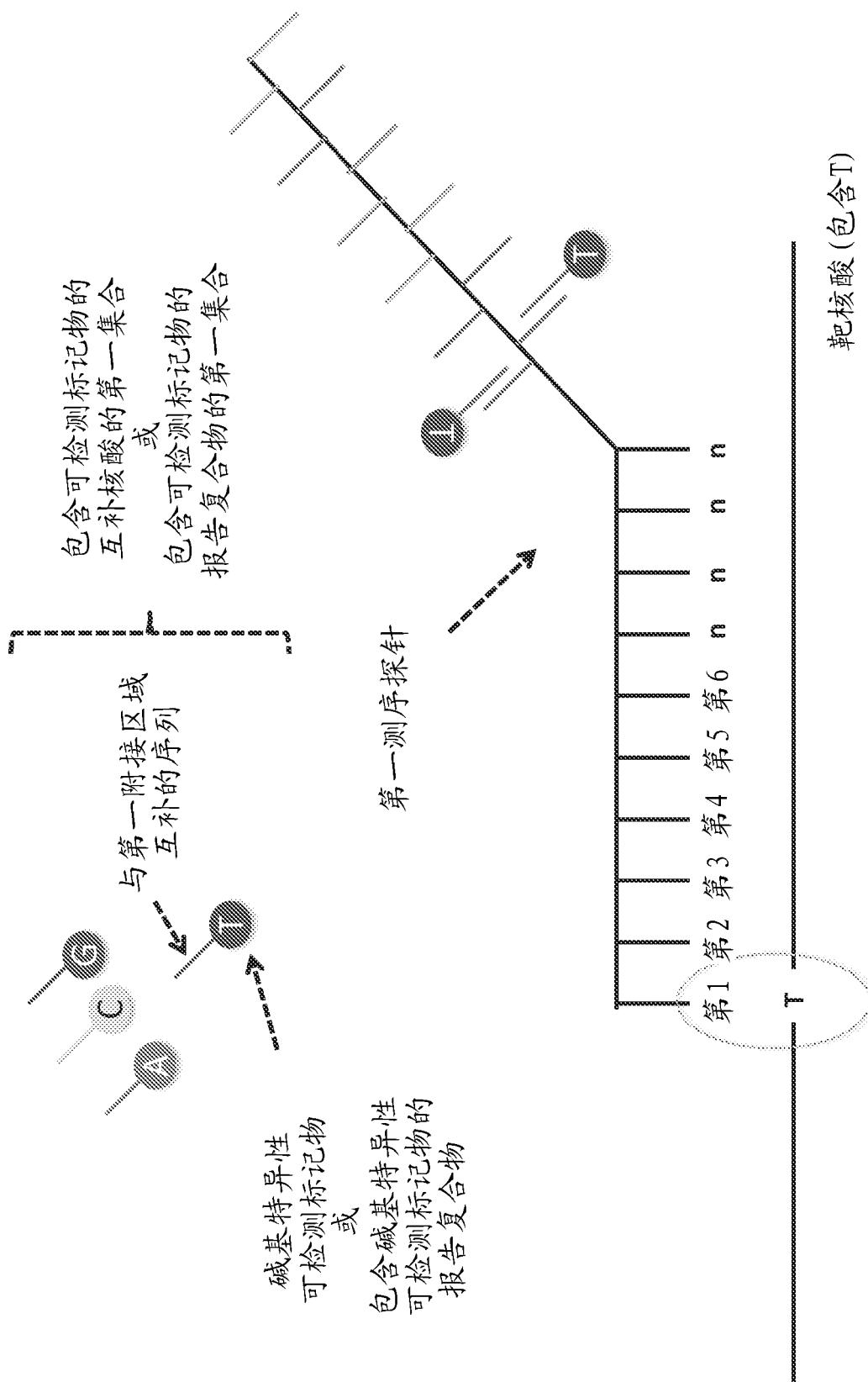


图 8A

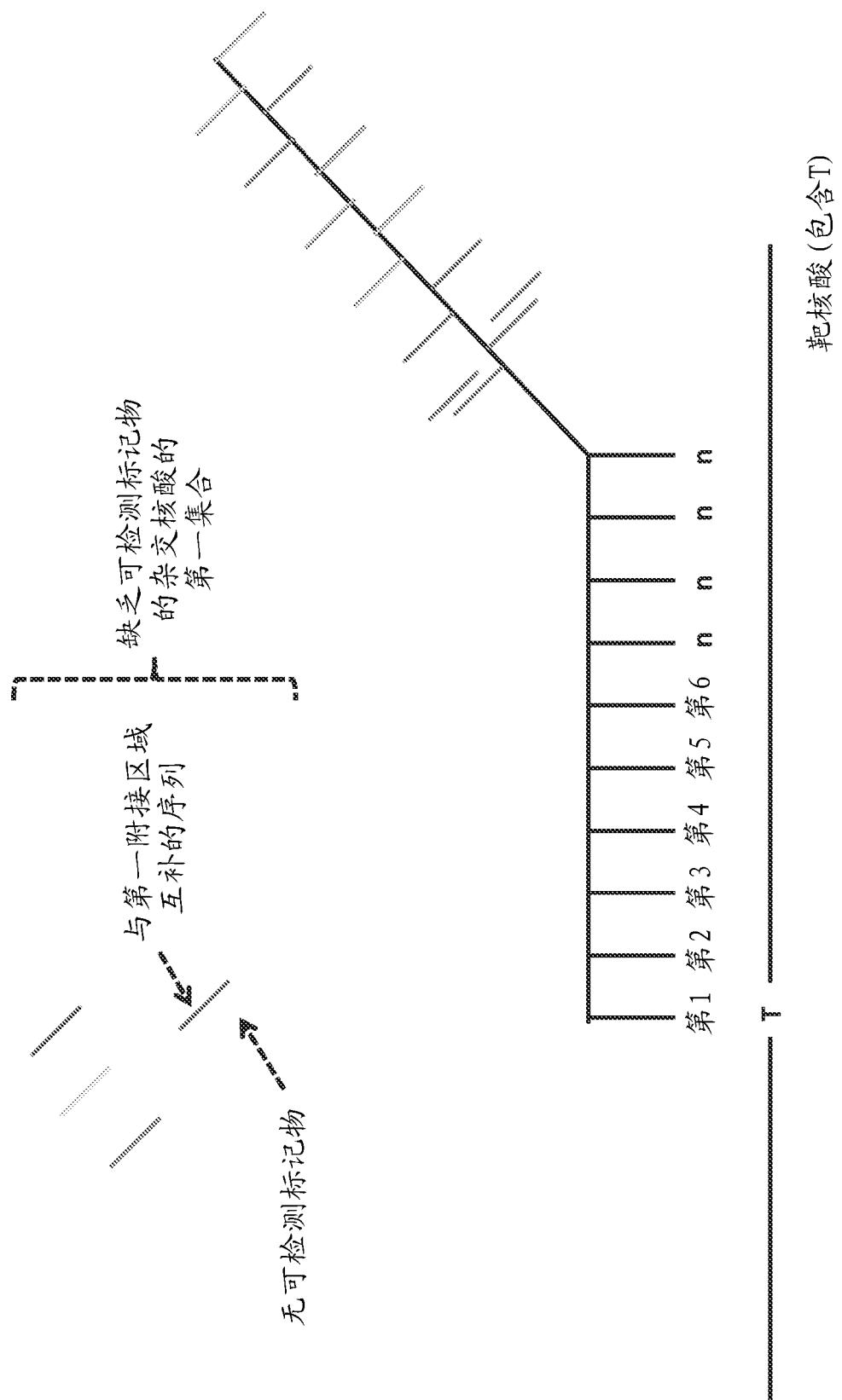
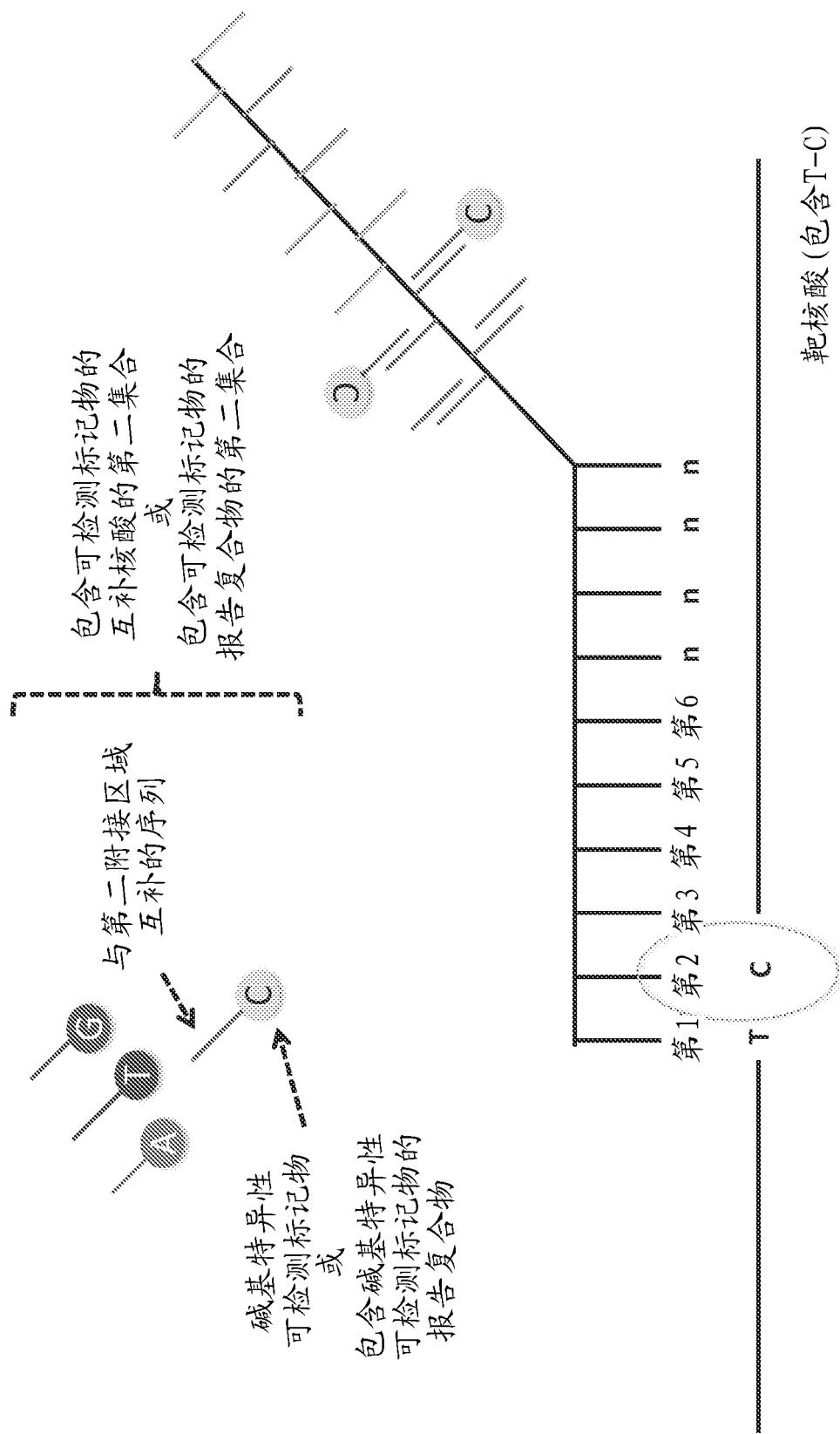


图 8B



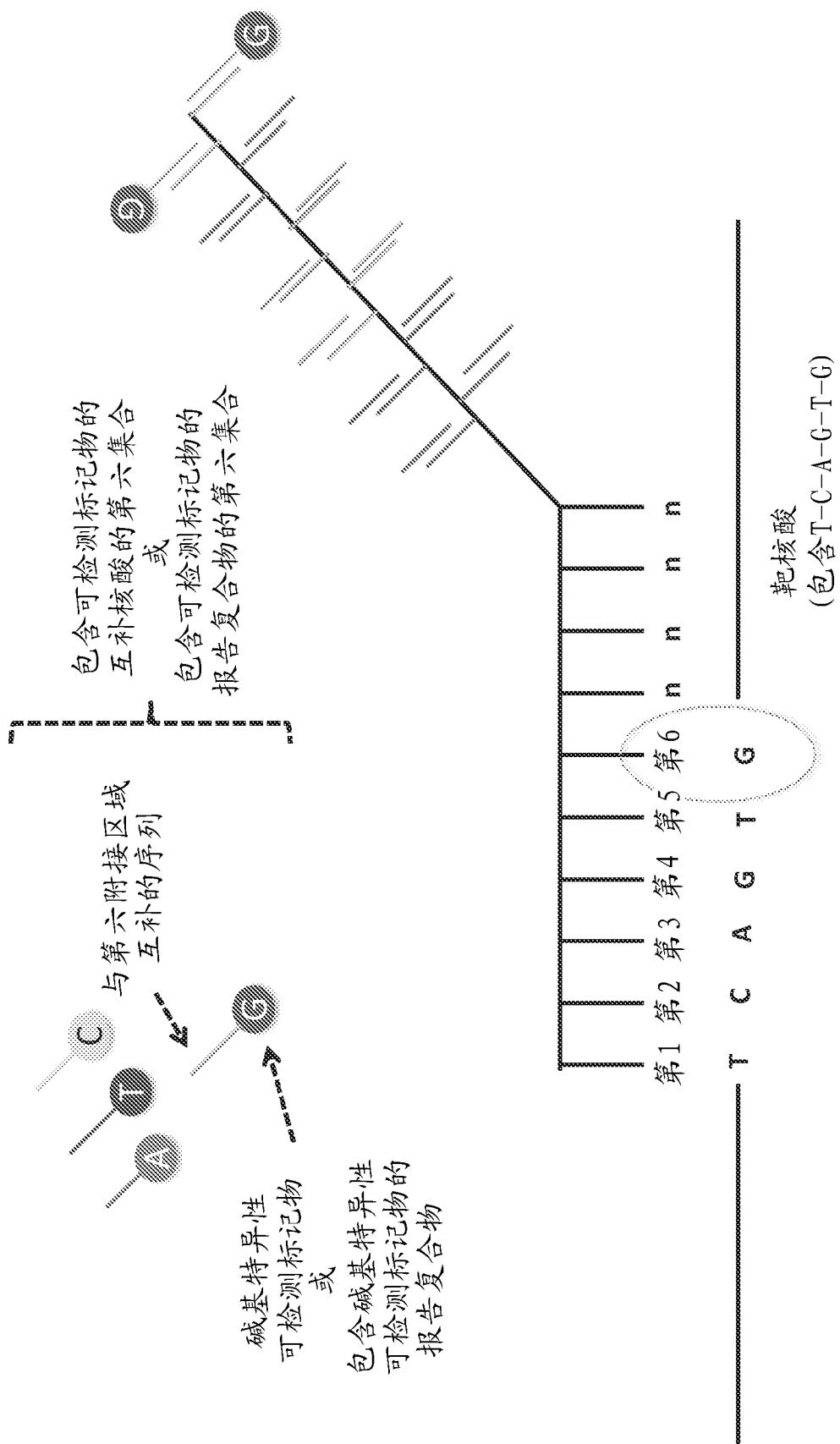


图 8D

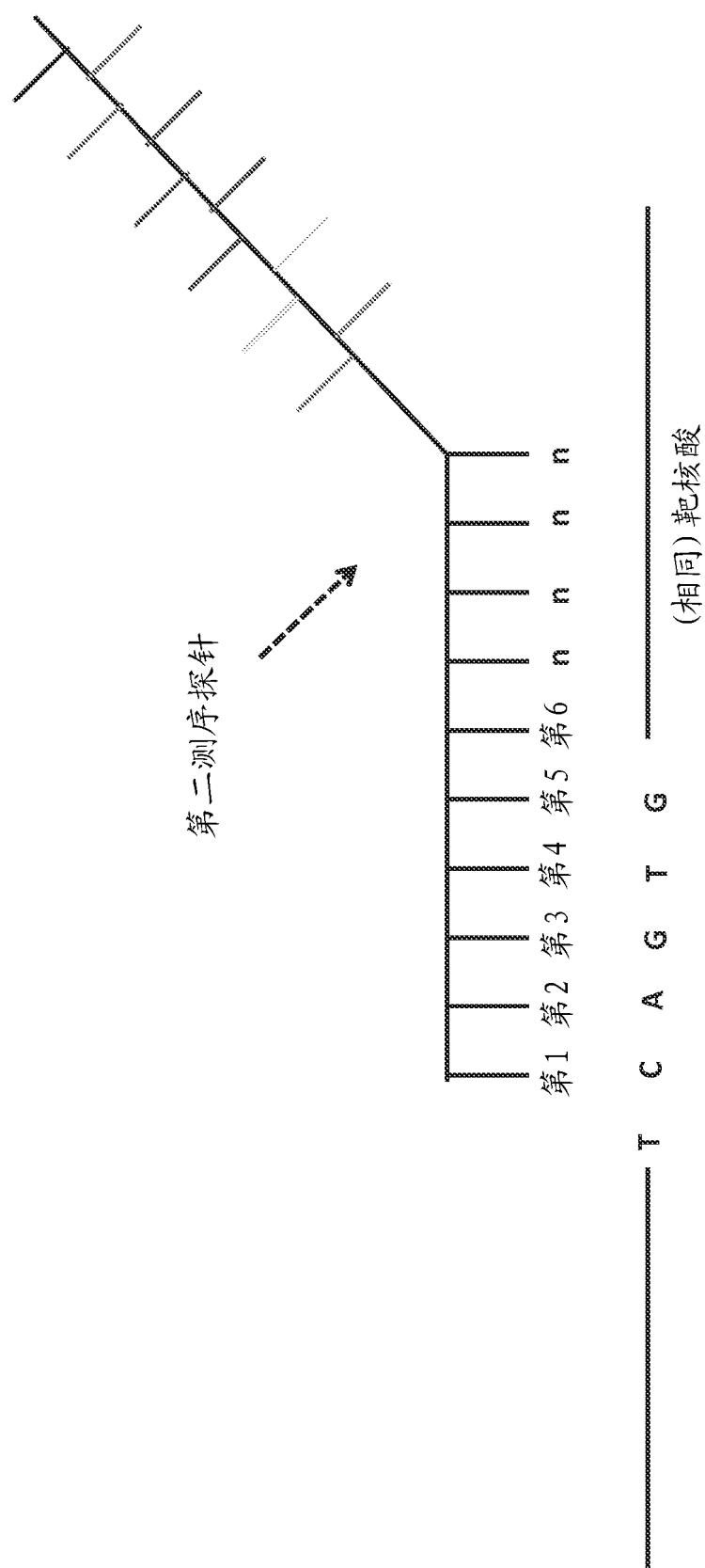


图 8E

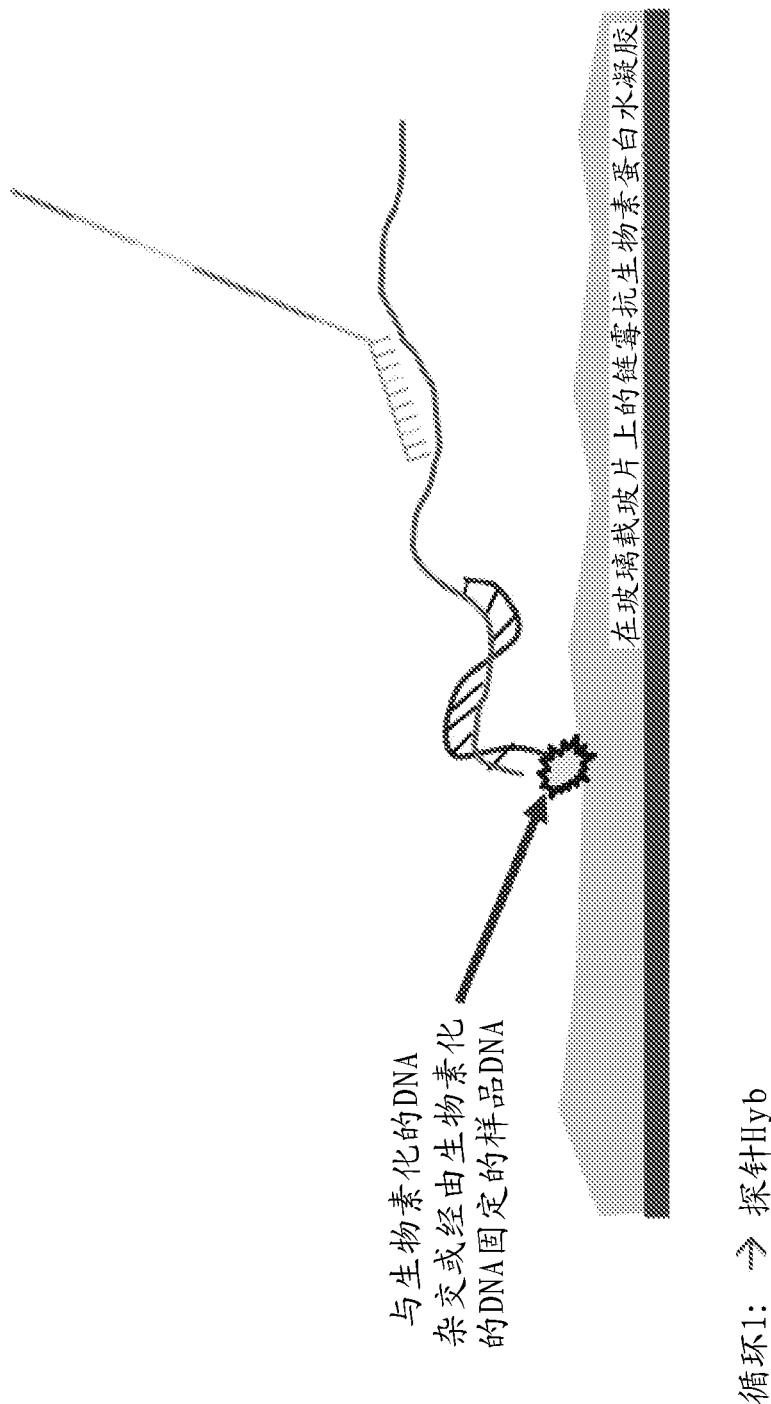


图 9A

包含可检测标记物的报告复合物

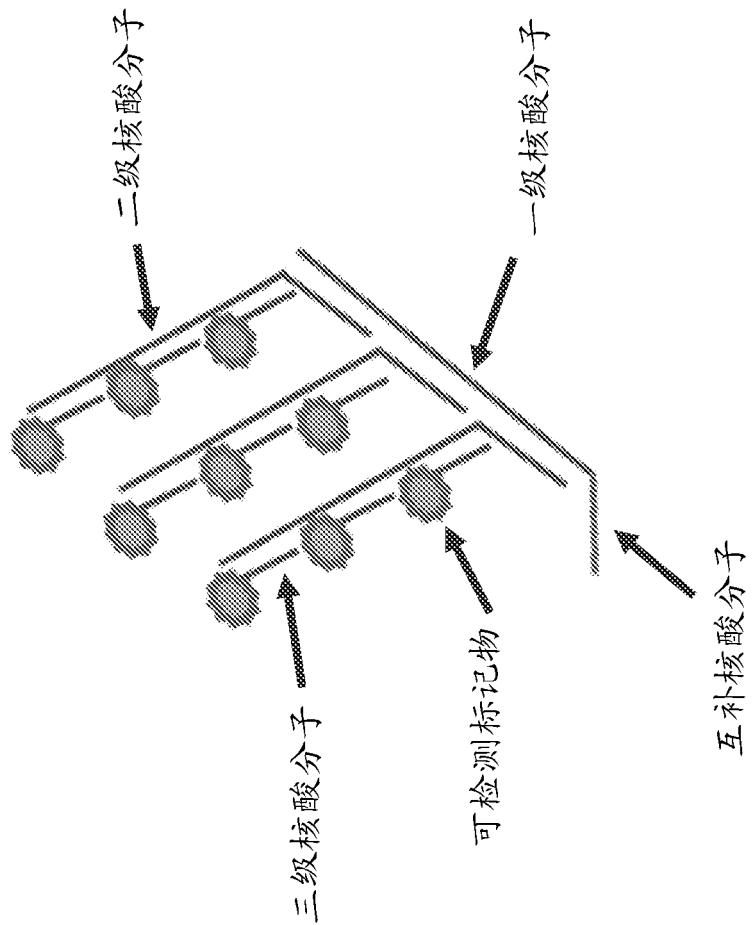


图 9B

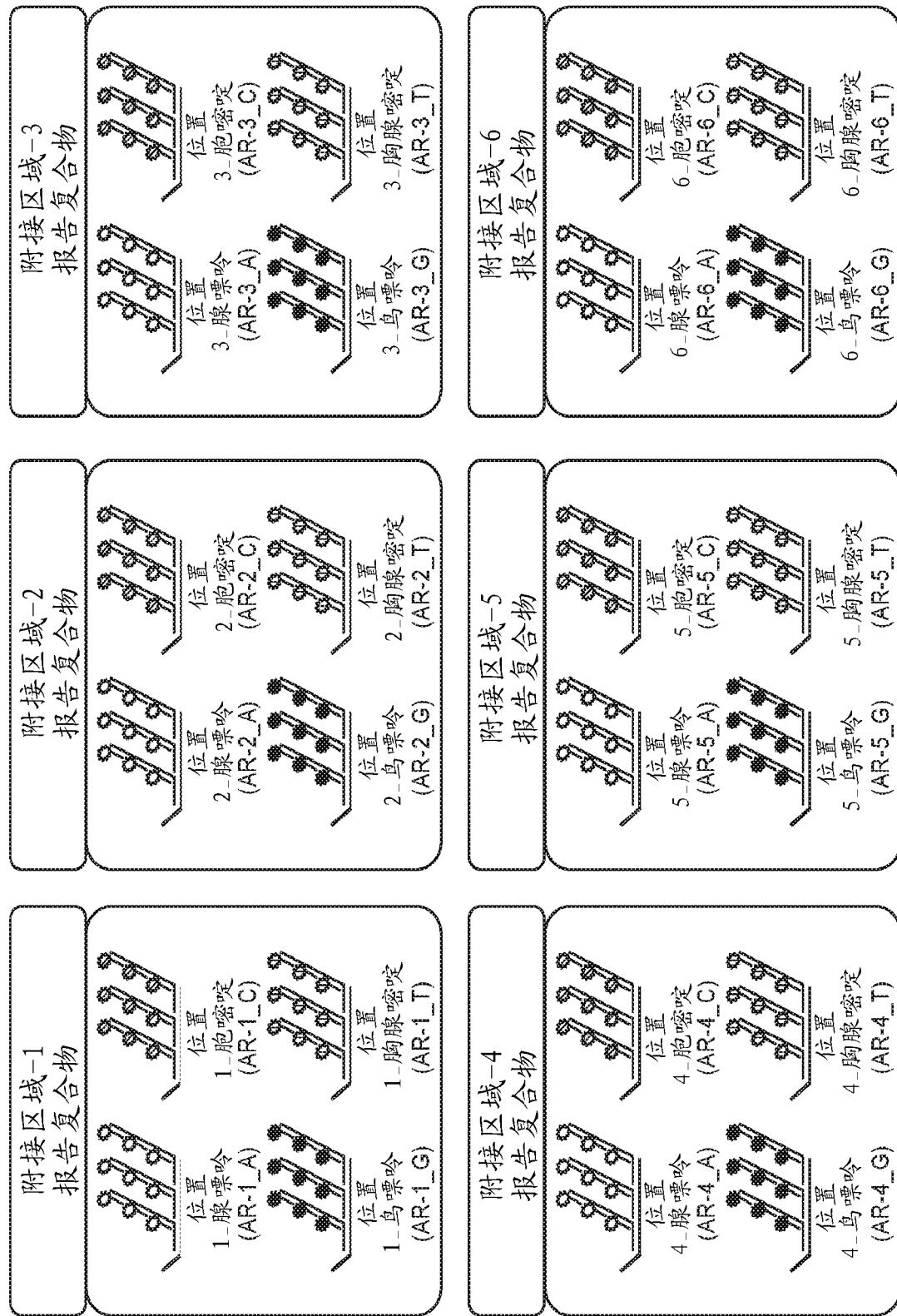


图 9C

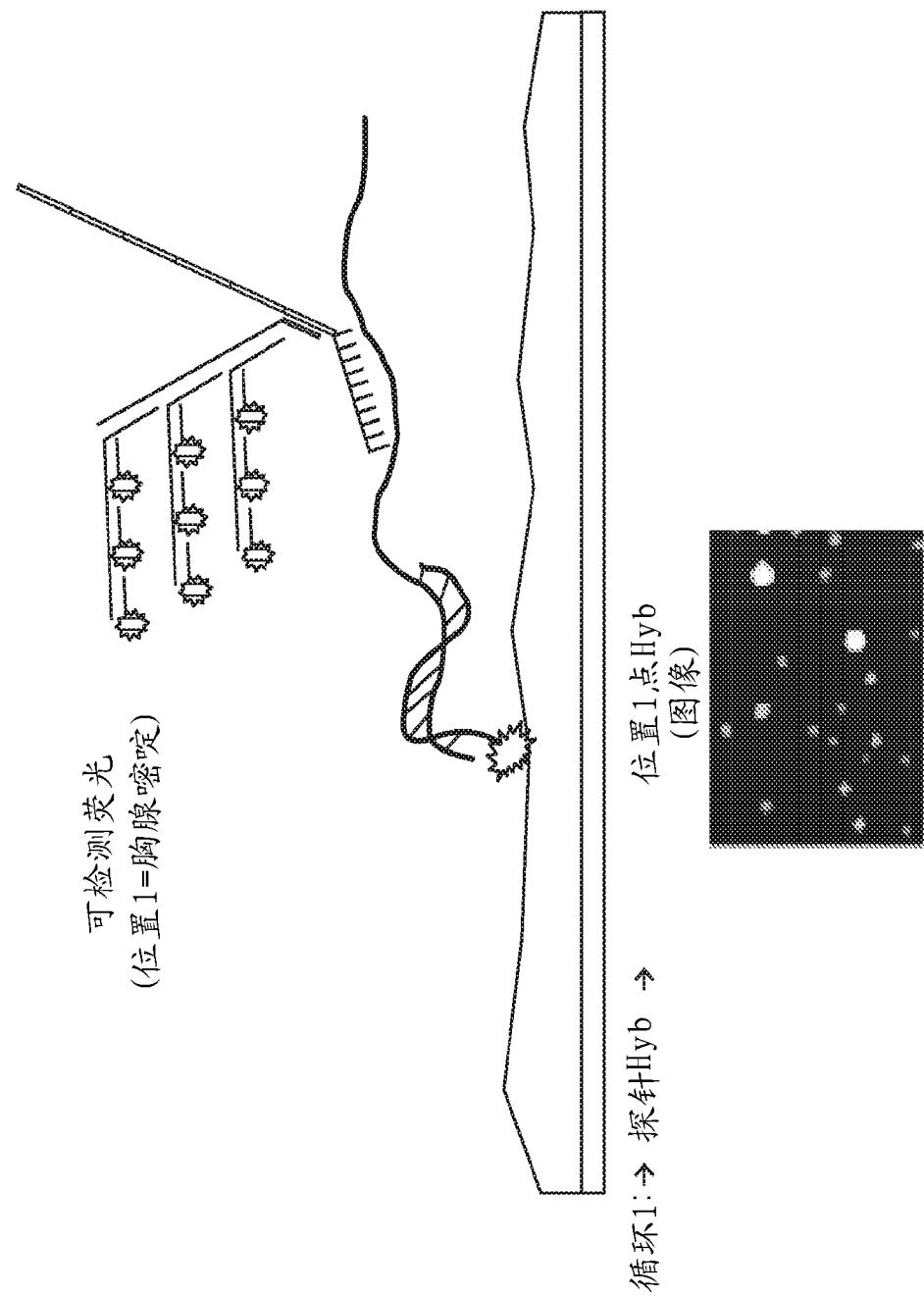


图 9D

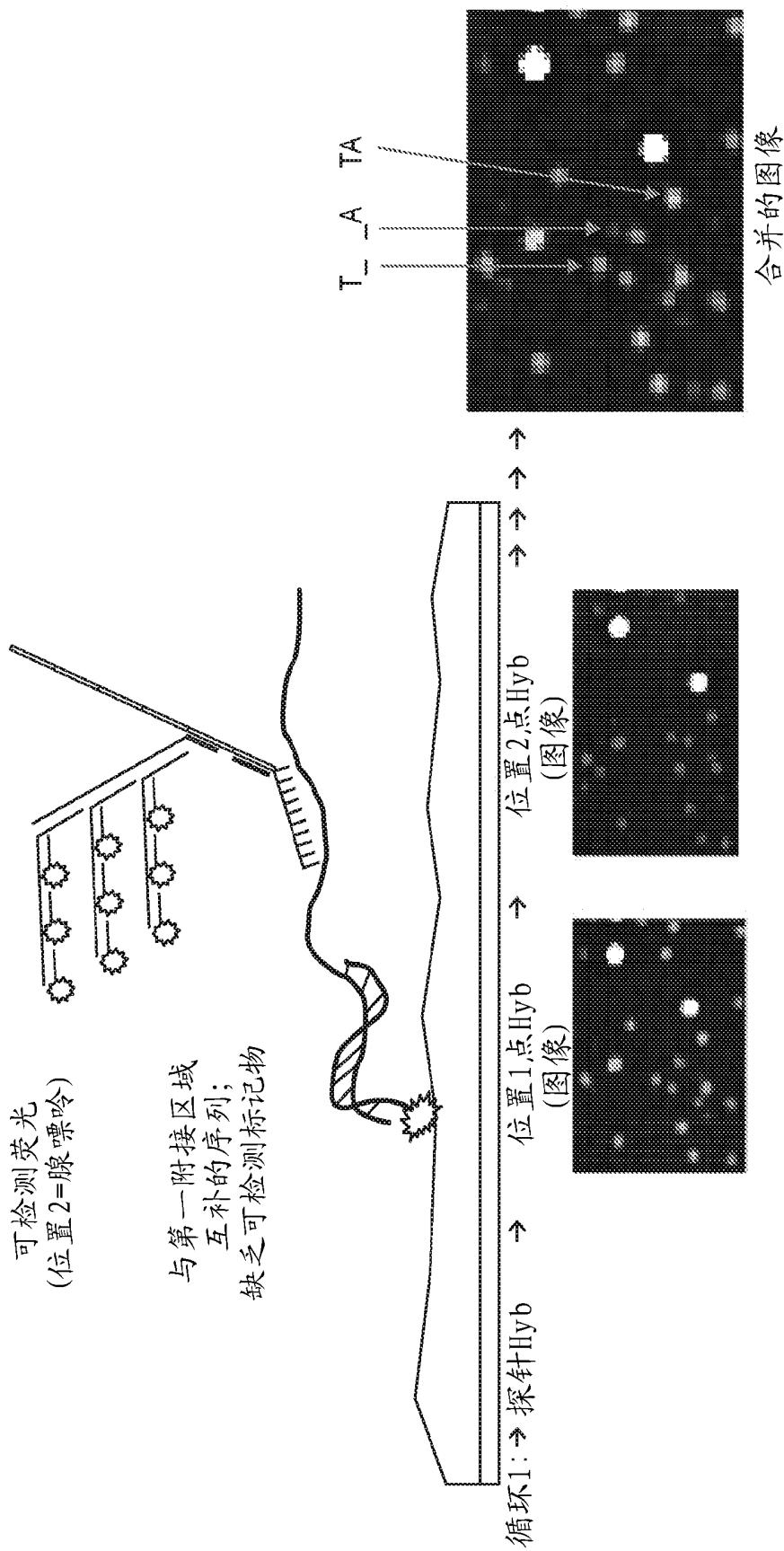


图 9E

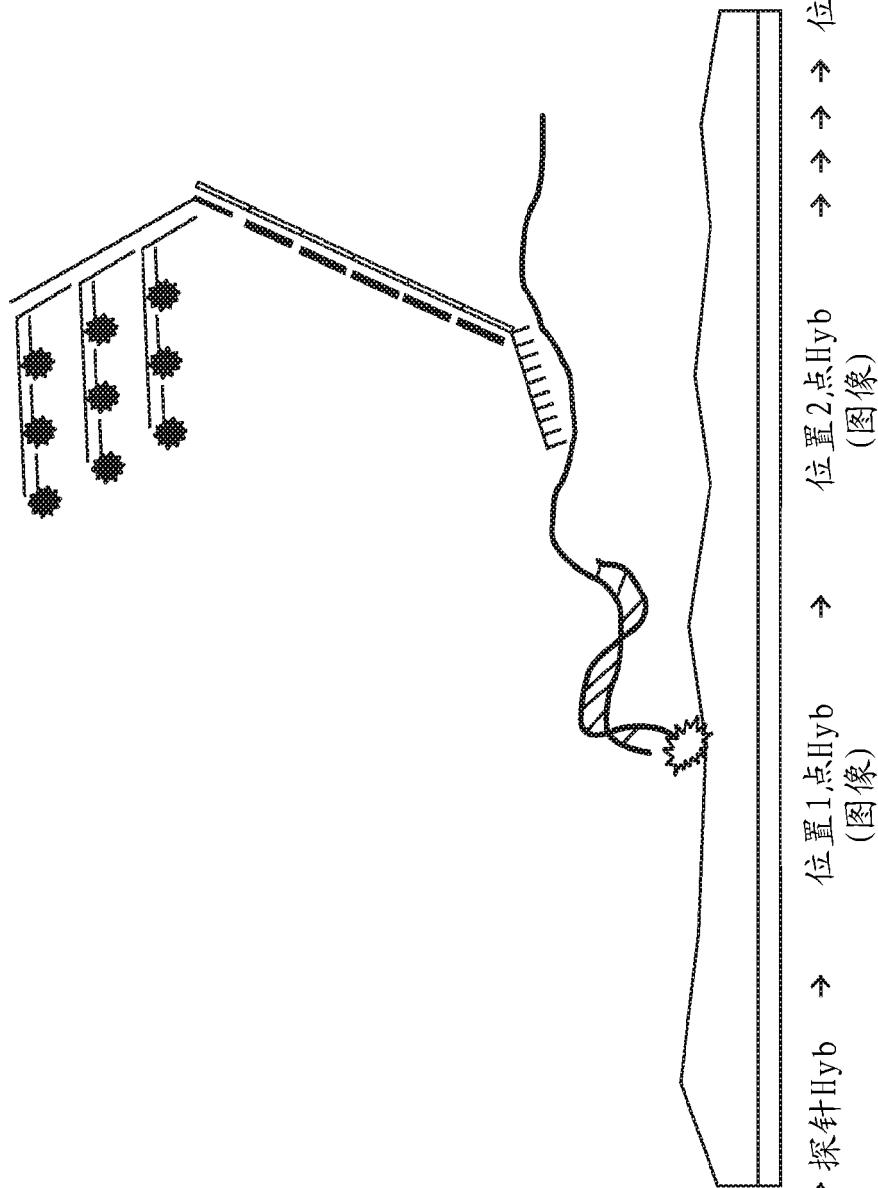
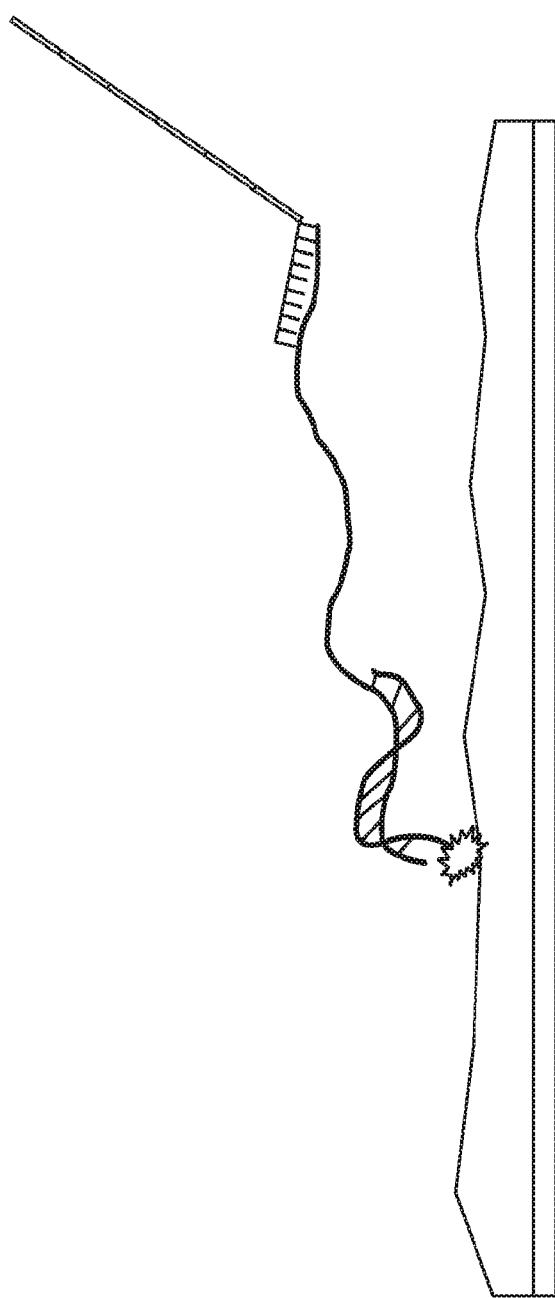
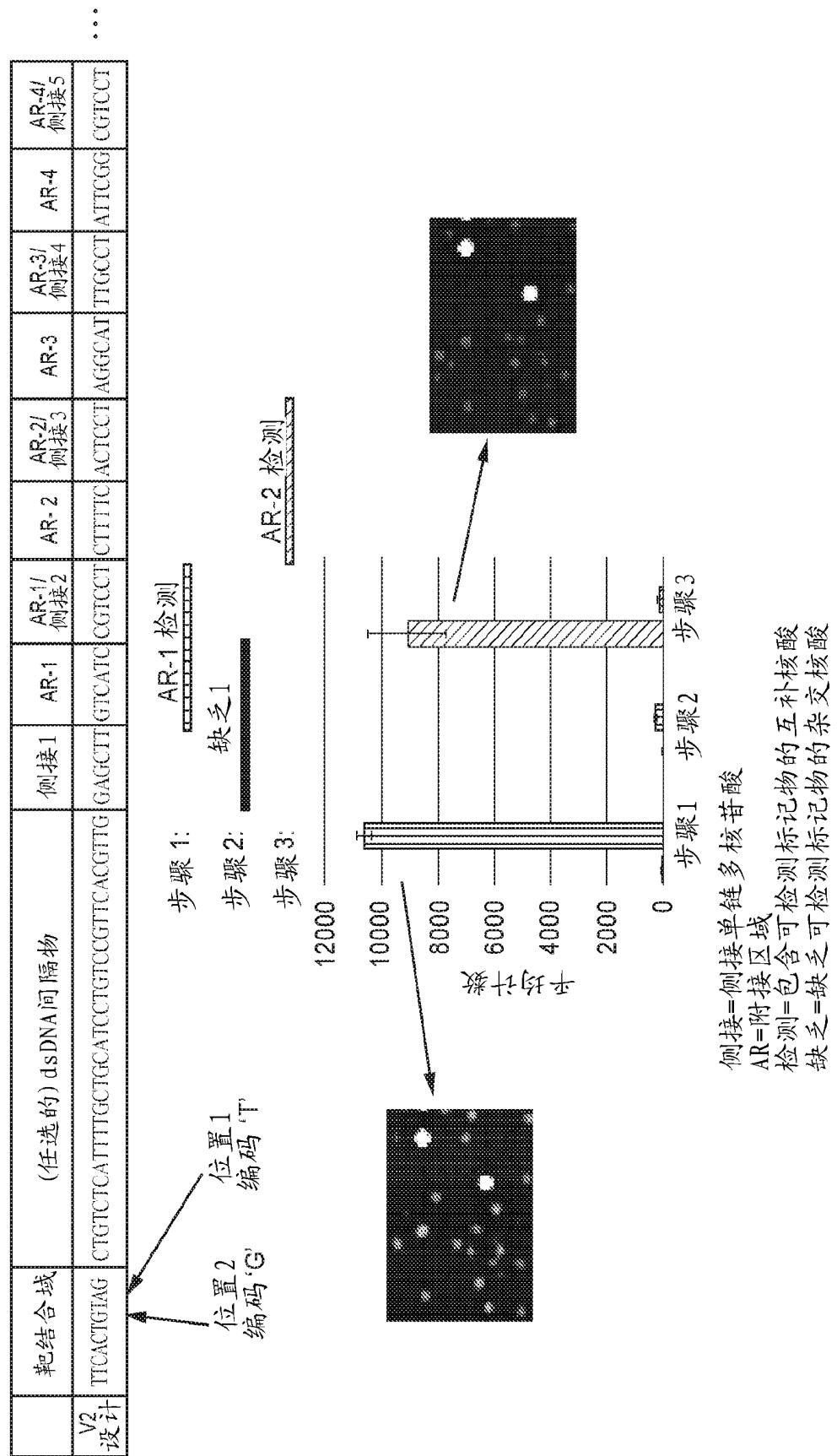


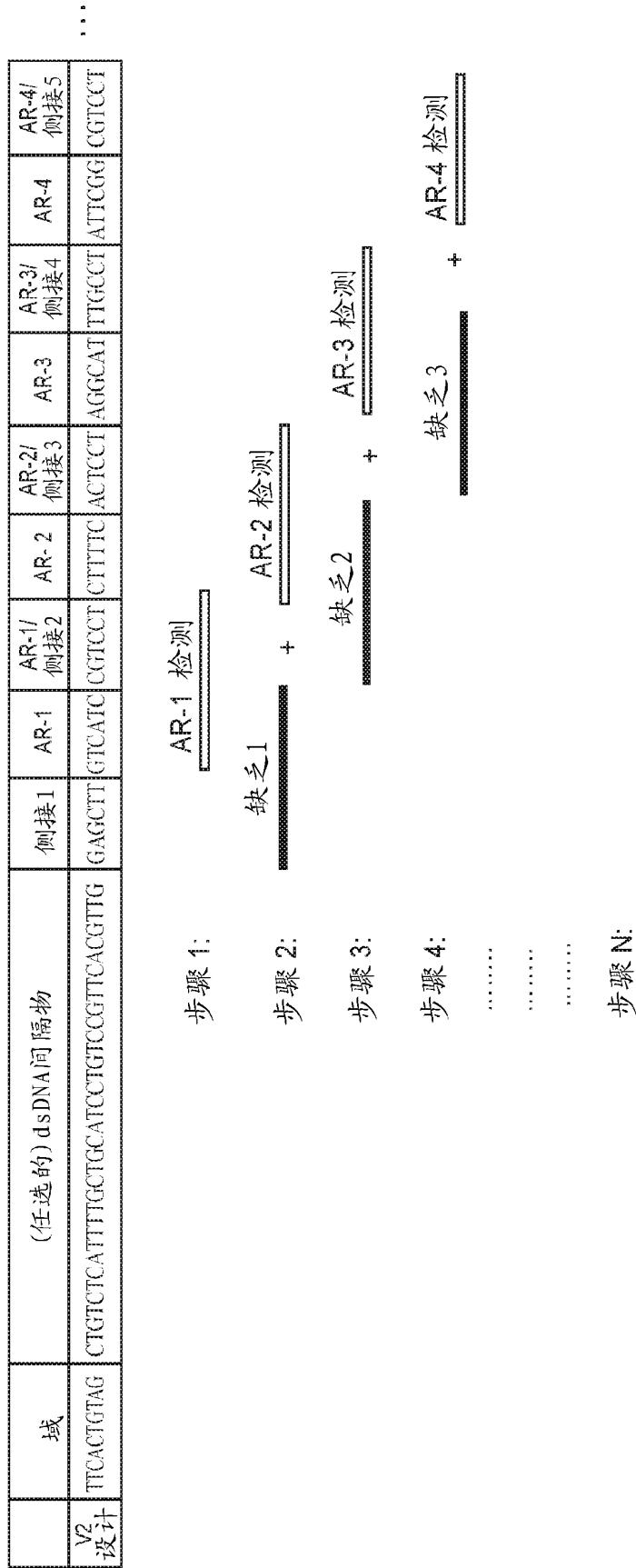
图 9F



循环2: → 第二探针 Hyb

图 9G





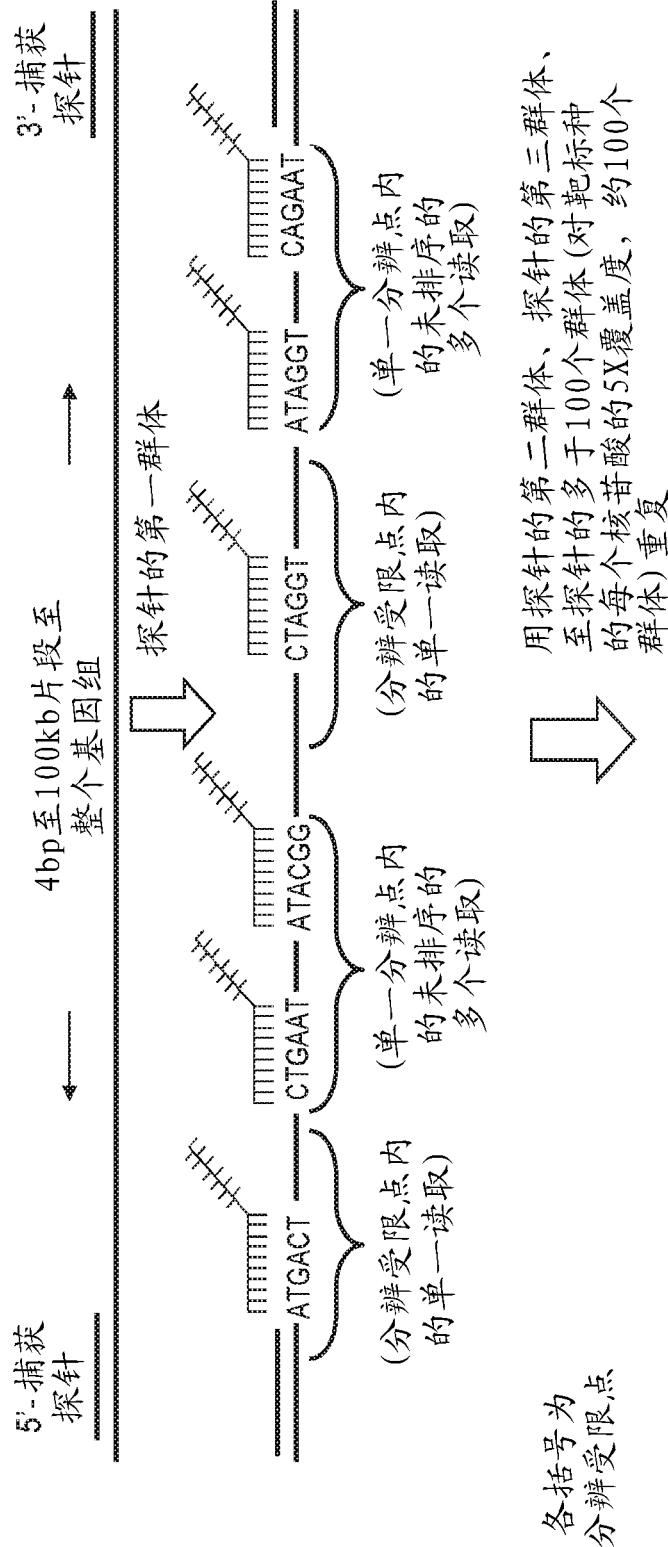


图 12

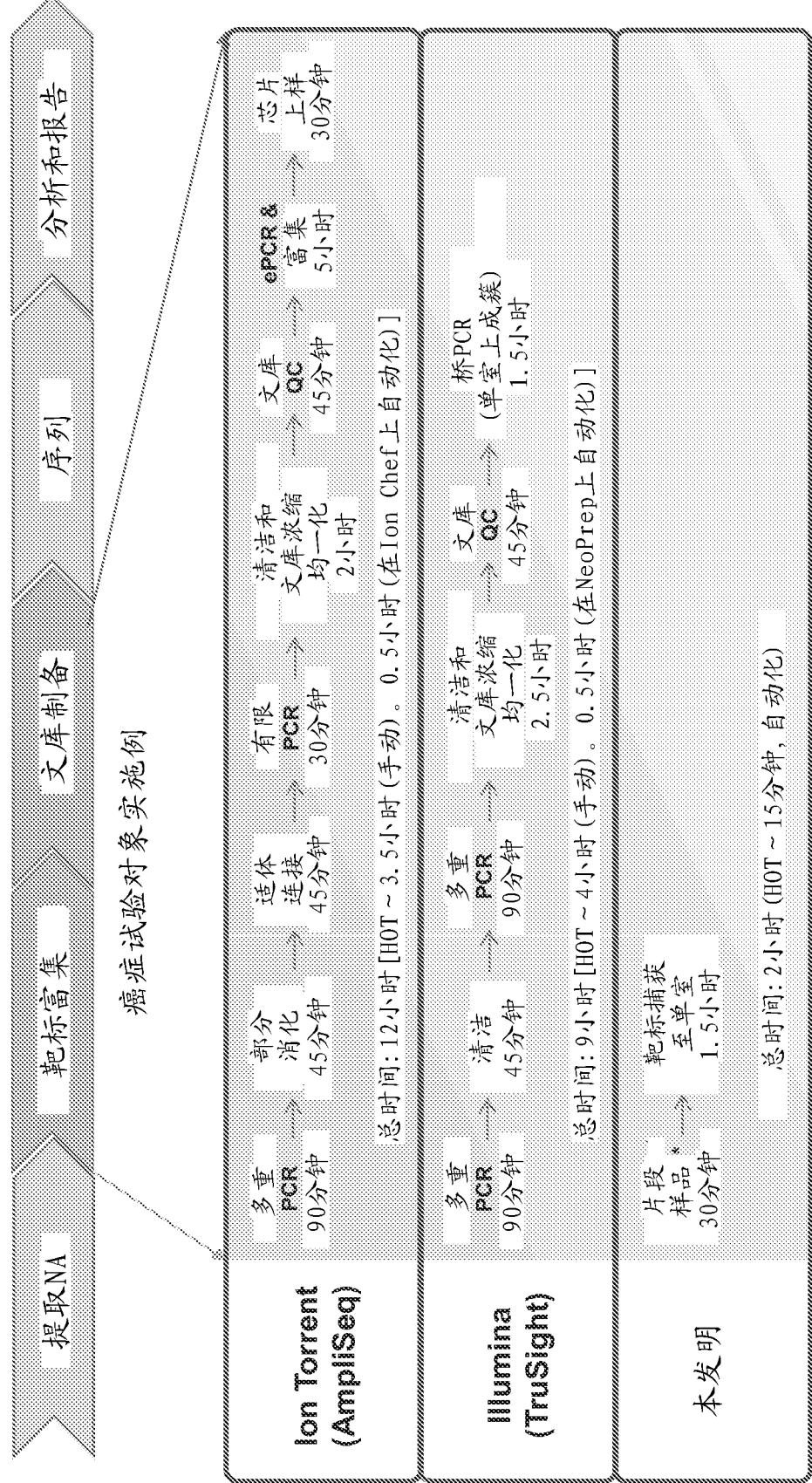
~ 20 FOV (对于基因的100X覆盖度)  $\times$  8分钟 / FOV = 2.7 小时运行 = 3 Gb/24 小时  
(~ 100% MiSeq)

~ 20 FOV  $\times$  5000 个基因 / FOV  $\approx$  100,000 个基因读取 (产生1000个基因的100X)

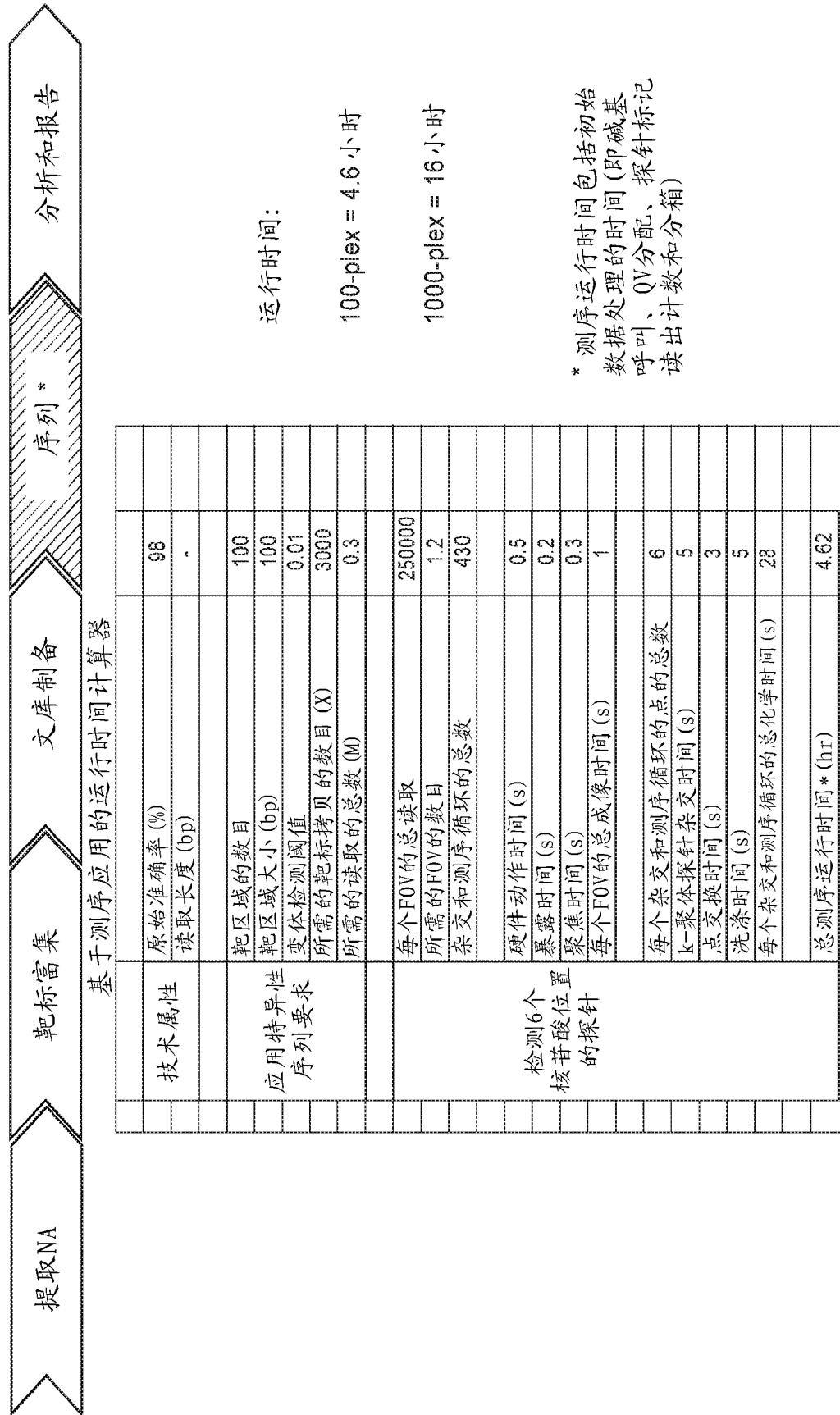
~ 20 FOV (对于基因的100X覆盖度)  $\times$  8分钟 / FOV = 2.7 小时运行 = 12 Gb/24 小时 \*

\* (排序的阵列NSTG通量估计)

## 完全测序流程图 (样品至答案)



完全测序流程图 (样品至答案)



## 基于探针占据的循环计算器

单室上的单分子的总数	30000	
“缺口” 靶标的平均长度 (bp)	100	
每个测序循环的条形码占据率 (%)	0.75	
至少一次“击中” 单室上的靶标的 >95%所需的条形码循环的数目	3	
每个位置的染料占据率 (%)	0.8	
染料位置的总数	6	
每个测序循环的1 bp编码的读取的总数	6	35
每个测序循环的2 bp编码的读取的总数	23	346
每个测序循环的3 bp编码的读取的总数	92	1843
每个测序循环的4 bp编码的读取的总数	369	5530
每个测序循环的5 bp编码的读取的总数	1475	8847
每个测序循环的6 bp编码的读取的总数	5898	5898
每个测序循环的“可用” 读取的总数		22499
以约3X覆盖度覆盖给定单分子靶标 所需的条形码循环的数目	143	
以约5X覆盖度覆盖给定单分子靶标 所需的条形码循环的数目	238	
以约3X覆盖度覆盖所有单分子靶标 所需的条形码循环的数目	430	
以约5X覆盖度覆盖所有单分子靶标 所需的条形码循环的数目	713	

图 15

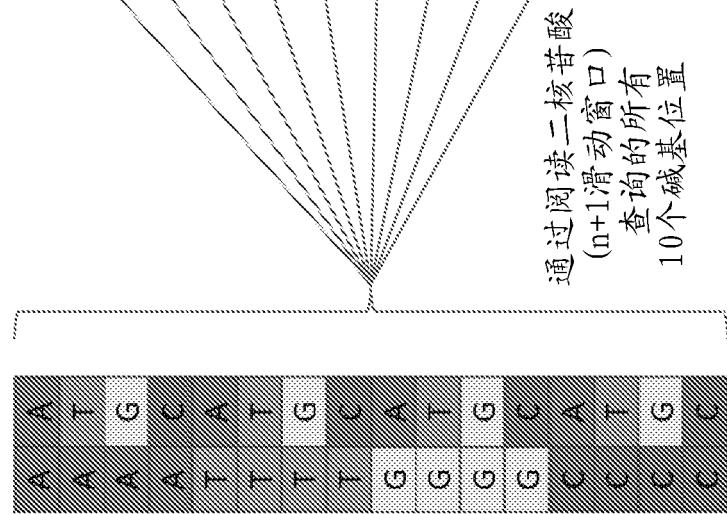
千兆碱基	时间/小时	阅读取	临床效用
Hiseq 2500	300Gb	11天	基因组序列癌症序列
Next Seq	120Gb	27小时	基因组序列
Prism Pi	10Gb	2.4	外显子组转录组
本发明	3至12Gb	3	靶向的 (Dx, cDx) 加多分析物加样品 至答案，长读取
MISeq	3.5Gb	24小时	转录组 靶向的 (Dx)
PGM 318	达到2Gb	7.3/4.4	靶向的
PacBio RS II	2.2Gb/天	3小时	生态位基因组，长读取， 高误差率
PGM 314	0.1Gb	3.7/2.3	靶向的

图 16

所有16中可能的  
二核苷酸组合：

模板： C G A T C T G G T T

测量的16种  
二核苷酸中的9种  
原始误差



	G	T	A	C	G	A	C	C	A	A
G	1.5%	2.3%								
G	0.8%	2.0%								
G	1.7%	3.1%								
G	2.1%	2.0%								
G	1.0%	3.8%								
G	2.0%	1.8%								
G	2.2%	2.8%								
G	0.6%	3.7%								
G	3.0%	32%								

原始误差率  
(平均值: 2.1%,  
无末端位置)

图 17

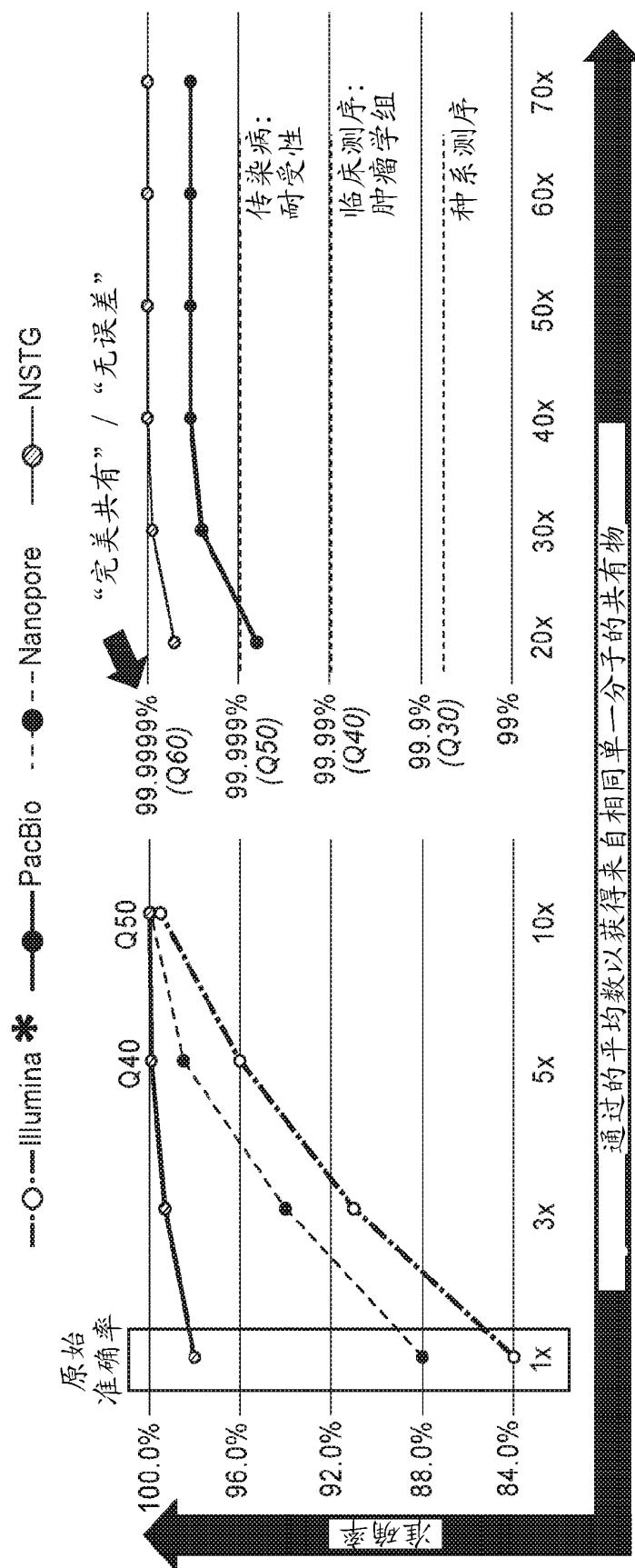
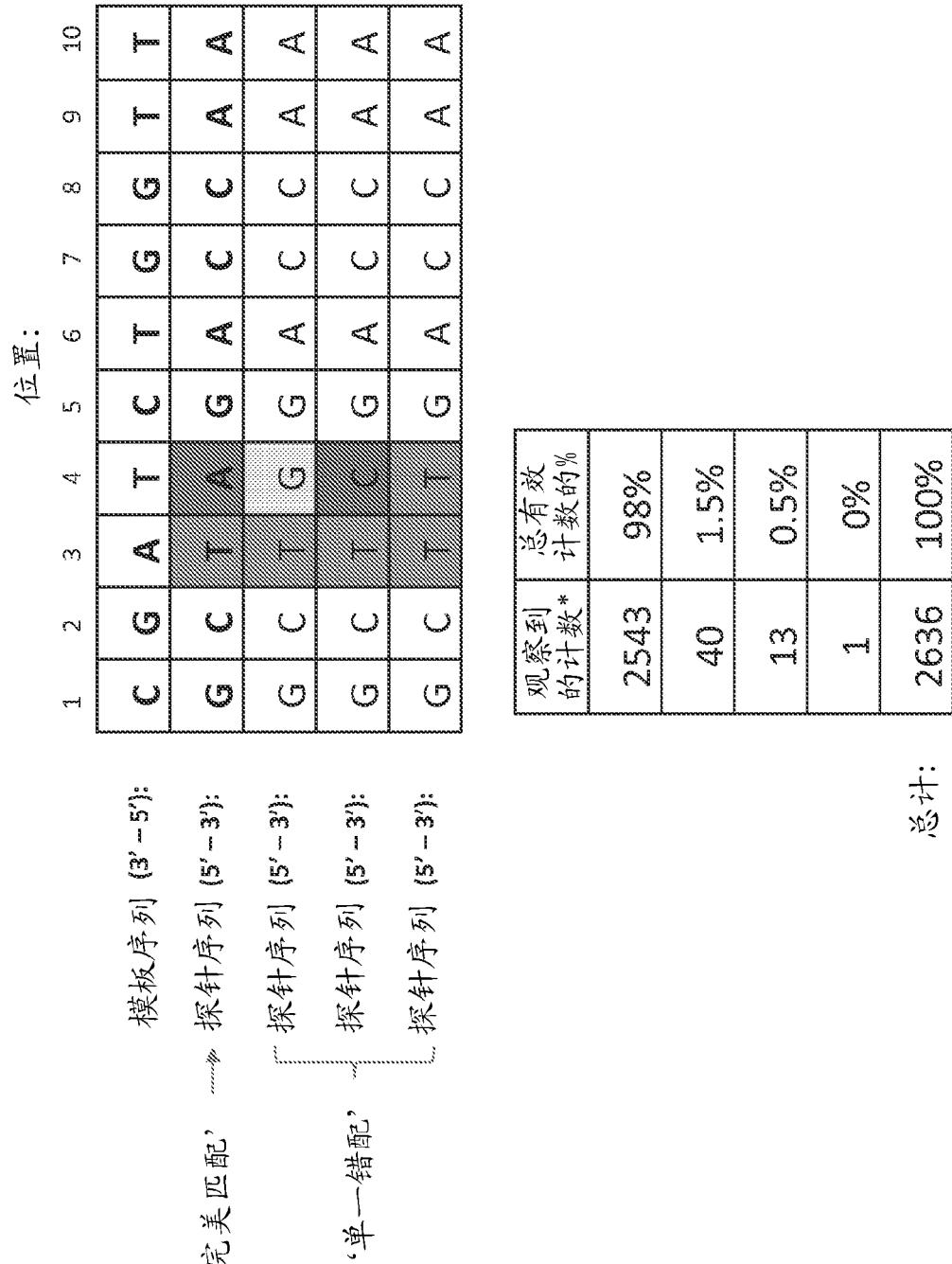


图 18

## 完美匹配

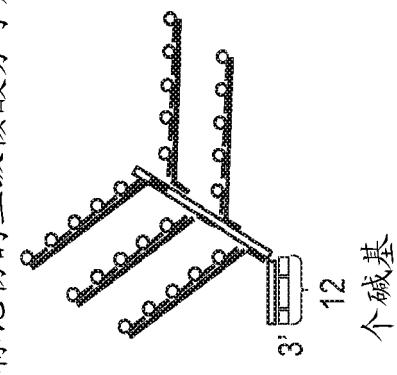


\* 观察到的计数值代表减去背景计数后的计数  
\*\* 所有杂交都在以下缓冲液条件下进行：1xSSPE/0.1% Tween20

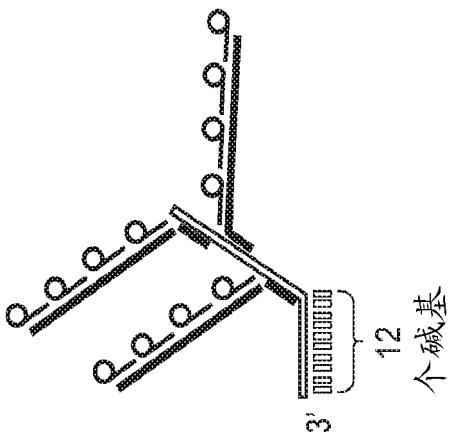
5x5

(1个一级核酸分子, 5个二级  
核酸分子, 5个具有可检测  
标记物的三级核酸分子)

4x3



3x4



间隔物 3x4

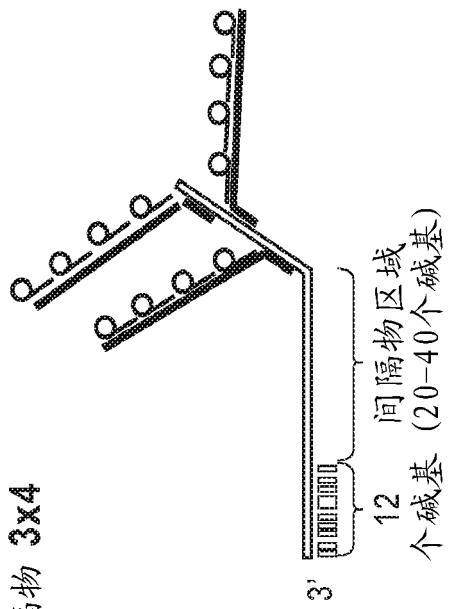


图 20A

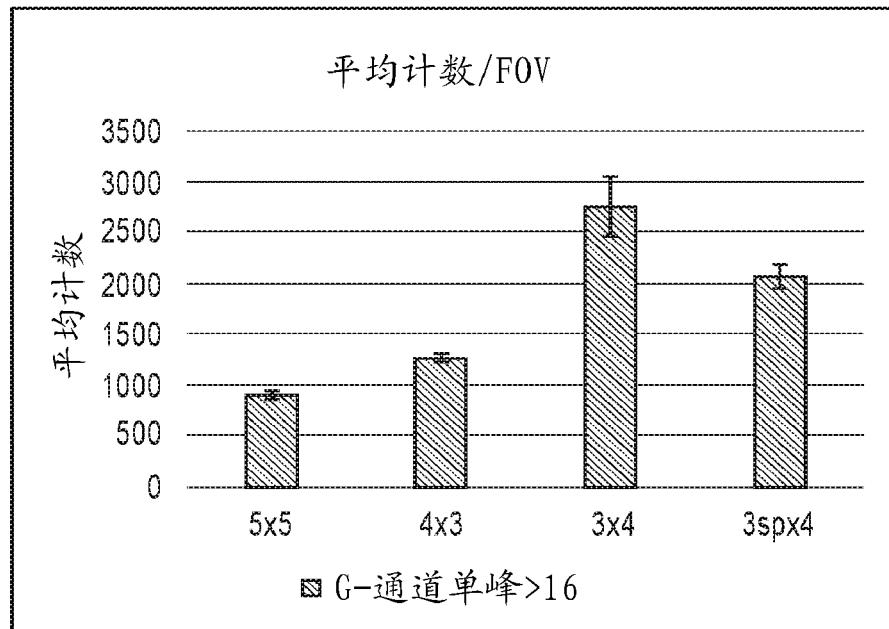


图 20B

用于构建包含可检测标记物的  
报告复合物的示例性配方

体积 ( $\mu$ L)	一级n. a. 分子 (10 $\mu$ M)	二级n. a. 分子 (10 $\mu$ M)	三级n. a. 分子 (100 $\mu$ M)	
5x4	1	4.5	2.25	92.25
5x3	1	4.5	1.8	92.7
4x4	1.28	4.5	2.25	91.97
4x3	1.28	4.5	1.8	92.42
3x4	1.8	4.5	2.25	91.45

表中的颜色对应于在左侧显示的  
复合物的图表中的颜色

图 20C

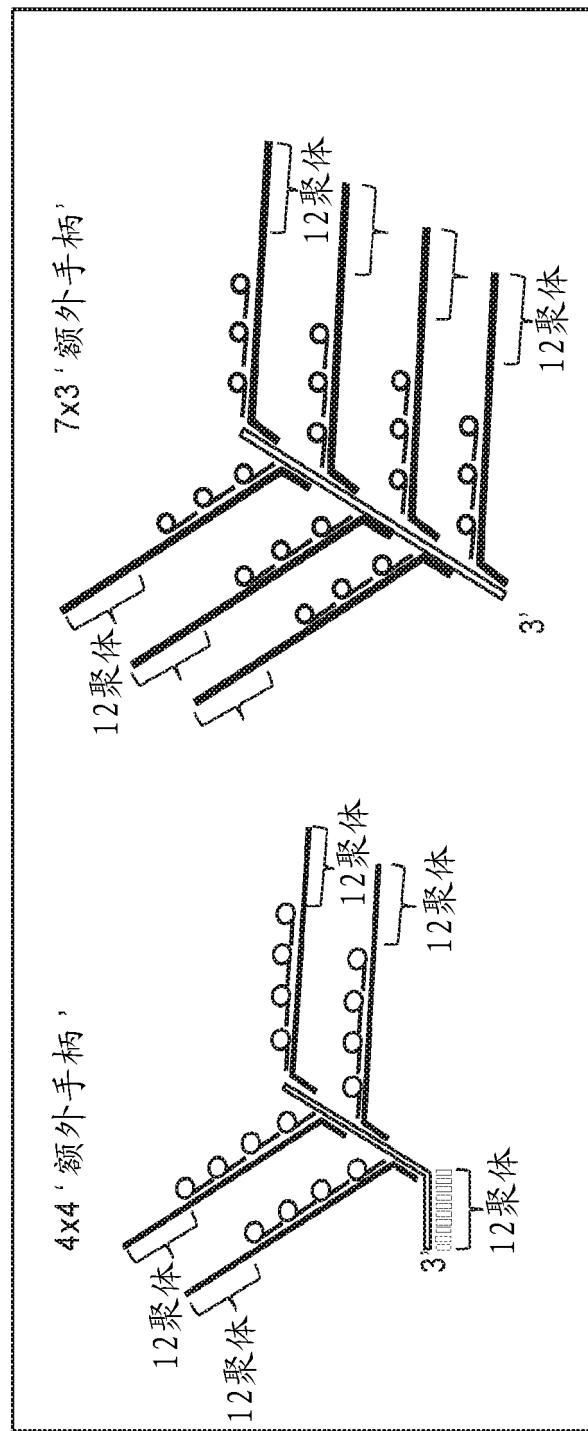


图 21A

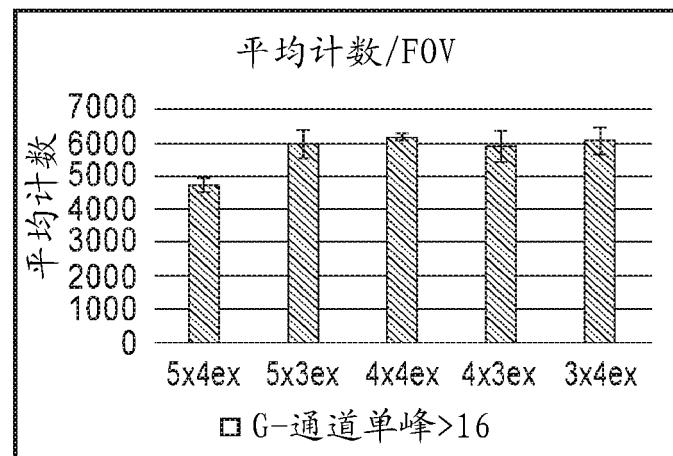


图 21B

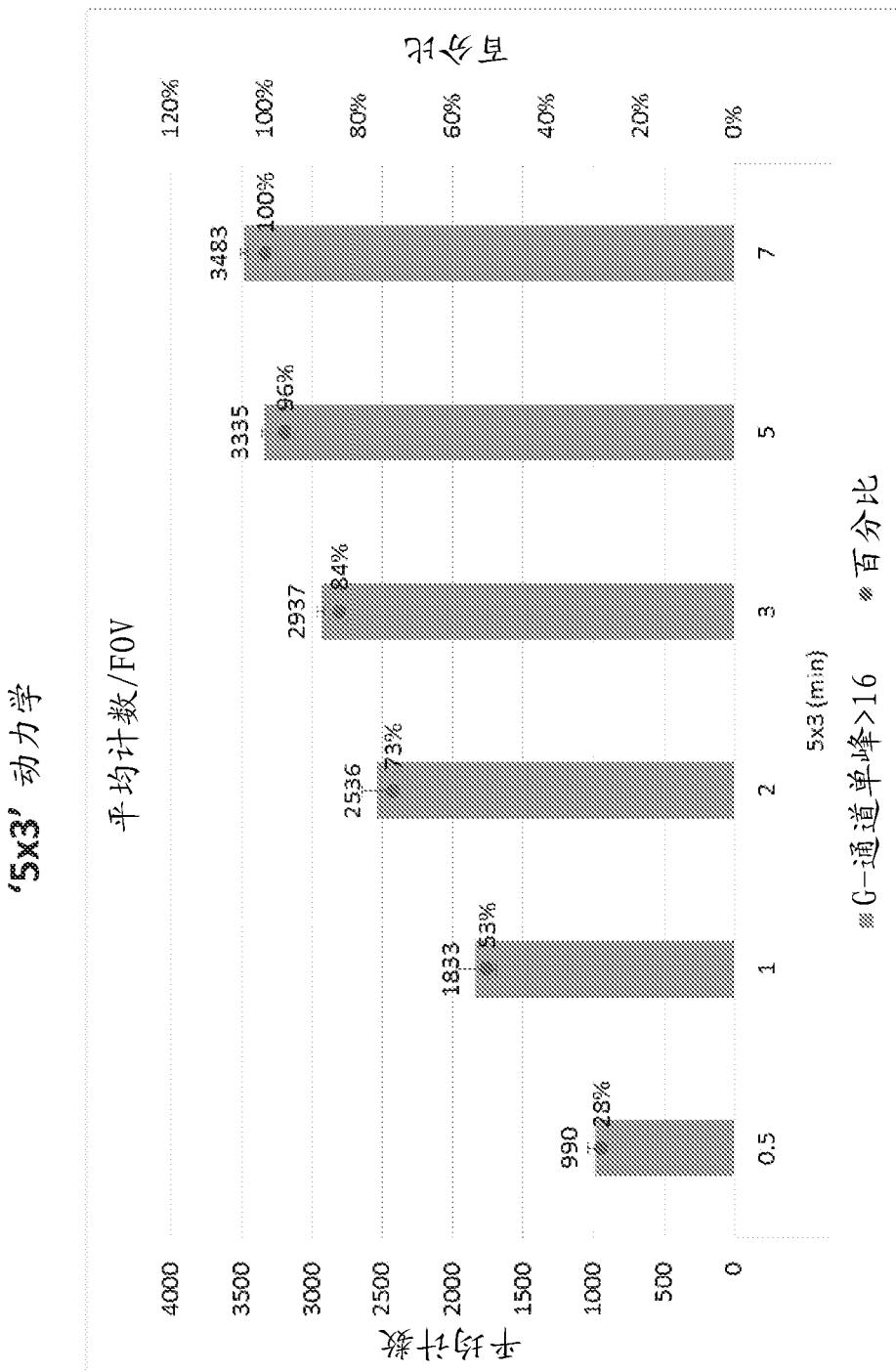


图 22A

“7x3 额外手柄”动力学

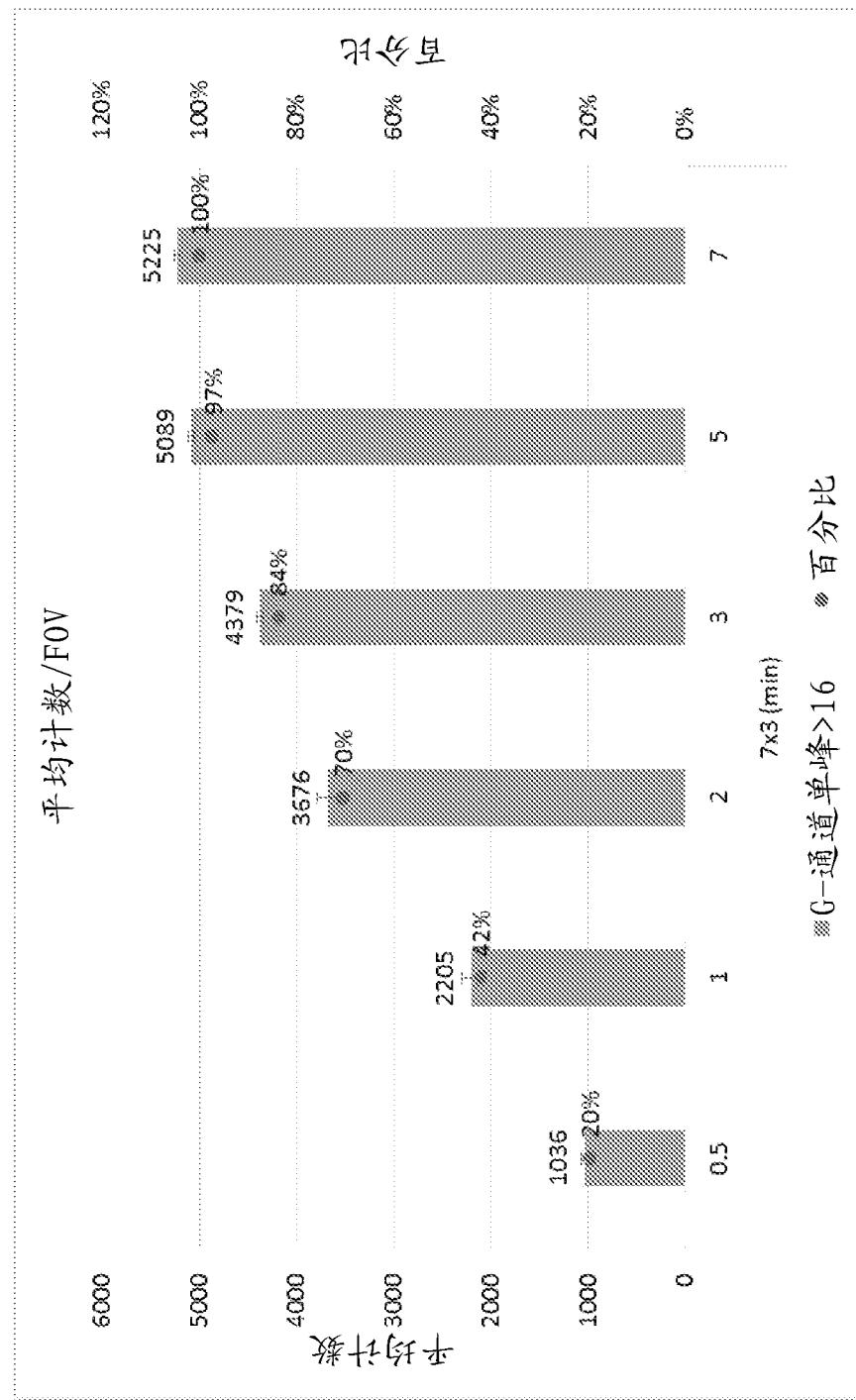


图 22B

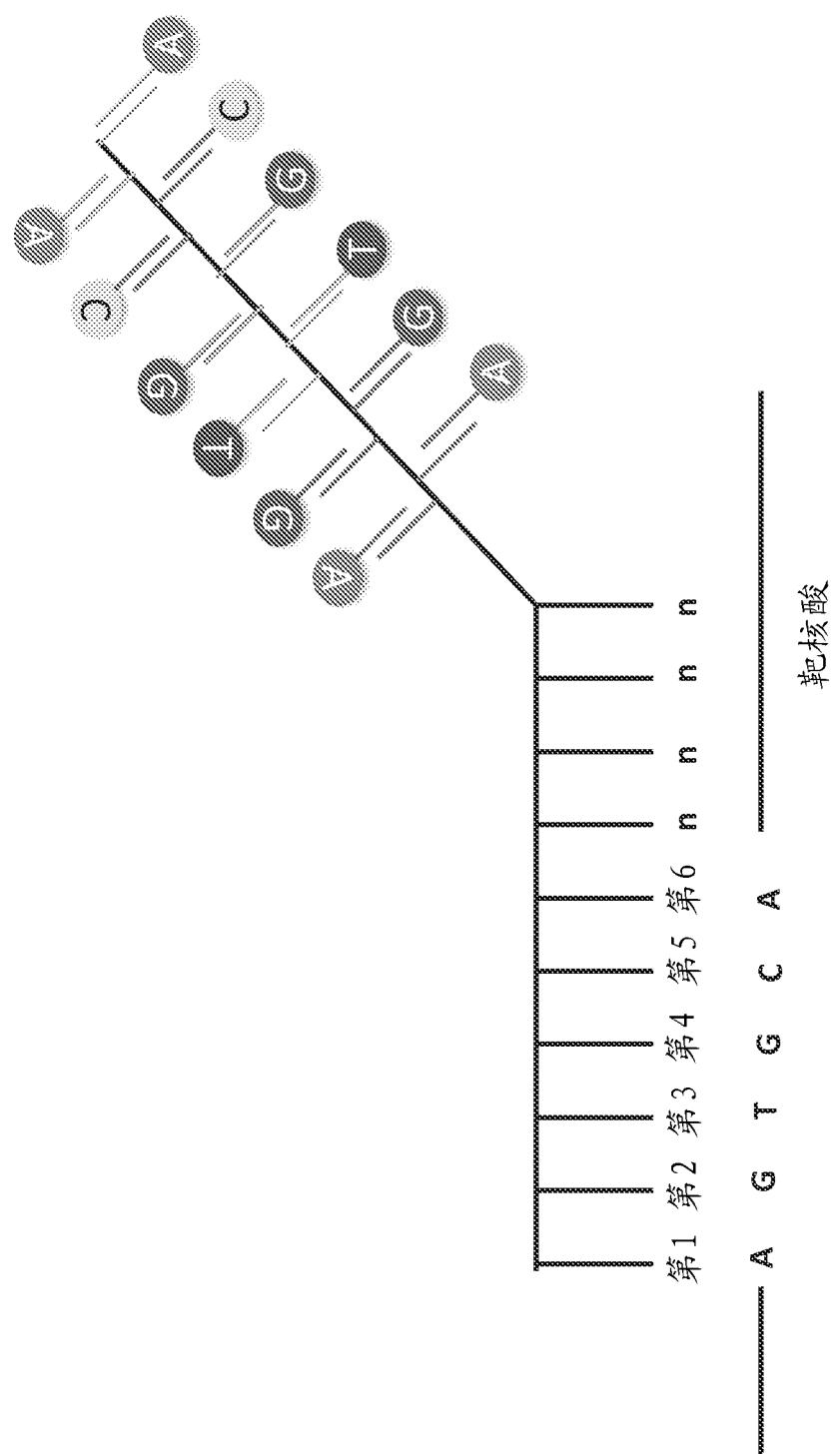


图 23

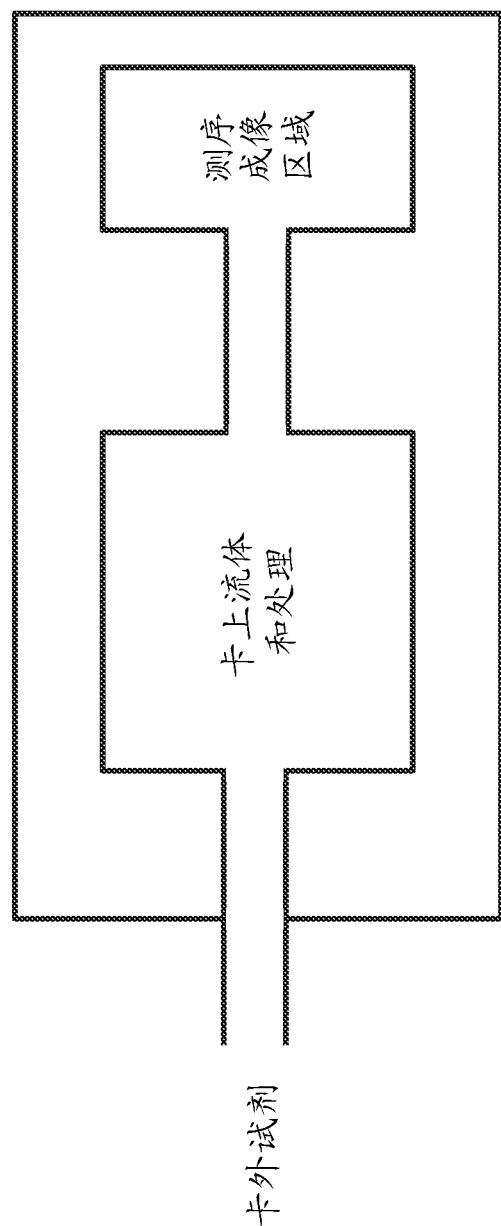


图 24

代码	序列
YGBYGR	GAGACTGTAC - 原始
BGYRGB	GAGACAGTAC
YBRYGB	GAGACCGTAC
RGBYGR	GAGACGGTAC
YRBGYR	GAGAGTGTAC
GRYGRB	GAGAATGTAC
RGYBRG	GAGATTGTAC

代码	计数	序列
YGBYGR	16074	GAGACTGTAC
RGYBRG	1132	GAGATTGTAC
YBRYGB	998	GAGACCGTAC
YRBGYR	784	GAGAGTGTAC
YBRYBG	712	
YBGRBY	699	
YGRYBG	692	
YBRGBR	674	
背景阈值	717	

图 25

序号	代码	计数
12 聚体	CTGAGACTGTAC	RYGBRG
11 聚体	TGAGACTGTAC	GBRYGB
10 聚体	GAGACTGTAC	YGBYGR
9 聚体	AGACTGTAC	YRBYRG
“12聚体”的镜像	CATGTCAGAGTC	GRBGYR
“11聚体”的镜像	CATGTCAAGAGT	BCYRBG
8 聚体	GACTGTAC	GBRYBG
		背景阈值
		622

图 26

全长报告物  
**20 pM** 全长 (6个点) 33个计数/fov

单点报告物

单点 [fM]	计数/FOV
20,000	饱和的
2,000	12000
200	308
<b>20</b>	<b>40</b>
2	11

**20pM NS 报告物  $\simeq$  20fM 单点**

图 27