



(12) 发明专利

(10) 授权公告号 CN 112703255 B

(45) 授权公告日 2025. 04. 29

(21) 申请号 201980047067.1
 (22) 申请日 2019.05.14
 (65) 同一申请的已公布的文献号
 申请公布号 CN 112703255 A
 (43) 申请公布日 2021.04.23
 (30) 优先权数据
 62/671091 2018.05.14 US
 62/836327 2019.04.19 US
 (85) PCT国际申请进入国家阶段日
 2021.01.13
 (86) PCT国际申请的申请数据
 PCT/US2019/032168 2019.05.14
 (87) PCT国际申请的公布数据
 W02019/222178 EN 2019.11.21
 (73) 专利权人 布鲁克空间生物学公司
 地址 美国马萨诸塞州

(72) 发明人 D·杜纳威 E·曼罗 J·比彻姆
 R·哈菲佐夫 S·科鲁孔达
 Y·邓 D·金 M·格雷戈里
 M·黄 M·沃尔什 G·梅雷迪斯
 M·麦克尔文 P·斯基尼
 C·伯克

(74) 专利代理机构 中国专利代理(香港)有限公司 72001
 专利代理师 黄登高 黄希贵

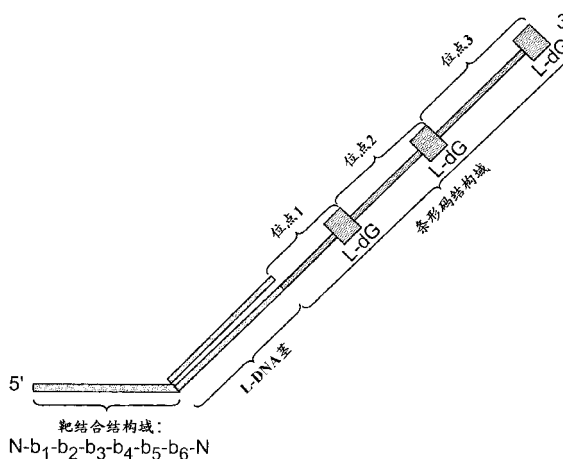
(51) Int. Cl.
 C12Q 1/6869 (2006.01)
 C12Q 1/6874 (2006.01)

(56) 对比文件
 CN 107208144 A, 2017.09.26
 审查员 封丽娜

权利要求书3页 说明书94页
 序列表24页 附图74页

(54) 发明名称
 化学组合物及其使用方法

(57) 摘要
 本公开内容涉及化学组合物、试剂盒和装置,以及用于在各种测定中使用这些组合物、试剂盒和装置的方法。



1. 一种用于鉴定靶核酸中的预定核苷酸序列的存在的体外方法,所述方法包括
 - (1) 使第一探针的靶结合结构域与靶核酸的第一区域杂交,其中所述探针包含靶结合结构域和条形码结构域;
其中所述靶结合结构域是长度至少6个核苷酸并与靶核酸中至少6个互补核苷酸杂交;
其中所述条形码结构域包含合成主链,所述条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列与互补核酸分子杂交,并且其中所述条形码结构域由L-DNA组成,
其中所述至少三个附着位置各自具有不同的核酸序列;
 - (2) 使包含至少一个第一可检测标记和至少一个第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置杂交;
 - (3) 鉴定与第一附着位置杂交的第一互补核酸分子的至少一个第一可检测标记和至少一个第二可检测标记;
 - (4) 去除与第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记;
 - (5) 使包含至少一个第三可检测标记和至少一个第四可检测标记的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置杂交;
 - (6) 鉴定与第二附着位置杂交的第二互补核酸分子的至少一个第三可检测标记和至少一个第四可检测标记;
 - (7) 去除与第二附着位置杂交的至少一个第三可检测标记和至少一个第四可检测标记;
 - (8) 使包含至少一个第五可检测标记和至少一个第六可检测标记的第三互补核酸分子,与条形码结构域的至少三个附着位置的第三附着位置杂交;
 - (9) 鉴定与第三附着位置杂交的第三互补核酸分子的至少一个第五可检测标记和至少一个第六可检测标记,从而基于至少一个第一可检测标记、至少一个第二可检测标记、至少一个第三可检测标记、至少一个第四可检测标记、至少一个第五可检测标记和至少一个第六可检测标记的身份,来确定预定核苷酸序列的存在。
2. 权利要求1的方法,其中所述探针包含在靶结合结构域和条形码结构域之间的单链DNA合成主链和双链DNA间隔区。
3. 权利要求1的方法,其中所述靶结合结构域是长度至少8个核苷酸。
4. 权利要求1的方法,其中所述靶结合结构域是长度至少10个核苷酸。
5. 权利要求1的方法,其中所述靶结合结构域是长度至少12个核苷酸。
6. 权利要求1的方法,其中所述条形码结构域长度为10个核苷酸至100个核苷酸。
7. 权利要求1的方法,其中所述靶结合结构域中的核苷酸数目大于所述条形码结构域中的附着位置数目。
8. 权利要求7的方法,其中所述条形码结构域包含至少四个附着位置。
9. 权利要求1的方法,其中所述条形码结构域中的每个附着位置包含一个附着区域。
10. 权利要求1的方法,其中所述条形码结构域中的每个附着位置的至少一个核酸序列为长度6个至20个核苷酸。
11. 权利要求1的方法,其中所述条形码结构域中的每个附着位置的至少一个核酸序列

为长度9个核苷酸。

12. 权利要求1的方法,其中所述条形码结构域中的每个附着位置的至少一个核酸序列为长度12个核苷酸。

13. 权利要求1的方法,其中所述条形码结构域中的每个附着位置的至少一个核酸序列为长度14个核苷酸。

14. 权利要求1的方法,其中所述条形码结构域中的每个附着位置的至少一个核酸序列为长度16个核苷酸。

15. 权利要求1的方法,其中每个附着位置的至少一个核酸序列包含3'末端鸟苷核苷酸。

16. 权利要求1的方法,其中每个附着位置的至少一个核酸序列包含至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个胞嘧啶核苷酸或其任何组合、以及3'末端鸟苷核苷酸。

17. 权利要求1的方法,其中所述靶结合结构域的至少12个核苷酸的每个核苷酸是D-DNA。

18. 权利要求1的方法,其中步骤(4)和(5)序贯或同时发生。

19. 权利要求1的方法,其中步骤(7)和(8)序贯或同时发生。

20. 权利要求1的方法,其中所述第一可检测标记和第二可检测标记具有相同的发射光谱。

21. 权利要求1的方法,其中所述第一可检测标记和第二可检测标记具有不同的发射光谱。

22. 权利要求1的方法,其中所述第三可检测标记和第四可检测标记具有相同的发射光谱。

23. 权利要求1的方法,其中所述第三可检测标记和第四可检测标记具有不同的发射光谱。

24. 权利要求1的方法,其中所述第五可检测标记和第六可检测标记具有相同的发射光谱。

25. 权利要求1的方法,其中所述第五可检测标记和第六可检测标记具有不同的发射光谱。

26. 权利要求1的方法,其中所述第一互补核酸分子、所述第二互补核酸分子和所述第三互补核酸分子各自包含可切割接头。

27. 权利要求1的方法,其中所述第一互补核酸分子包含报告探针,所述报告探针包含一级核酸,

其中所述一级核酸分子包含至少两个结构域,

与条形码结构域的第一附着位置杂交的第一结构域,和

与六个二级核酸分子杂交的第二结构域,

其中所述二级核酸分子各自与五个三级核酸分子杂交,其中所述三级核酸分子各自包含可检测标记。

28. 权利要求27的方法,其中所述一级核酸分子包含位于第一结构域和第二结构域之间的可切割接头。

29. 权利要求28的方法,其中所述可切割接头是光可切割接头。

30. 权利要求27的方法,其中每个二级核酸分子包含至少两个结构域,与一级核酸分子的第二结构域杂交的第一结构域;和与五个三级核酸分子杂交的第二结构域。

31. 权利要求30的方法,其中每个二级核酸分子包含位于第一结构域和第二结构域之间的可切割接头。

32. 权利要求31的方法,其中所述可切割接头是光可切割接头。

33. 权利要求28、29、31和32中任一项的方法,其中去除与所述第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记包括切割一级核酸的第一结构域和第二结构域之间的可切割接头、切割每个二级核酸的第一结构域和第二结构域之间的可切割接头、或其任何组合。

化学组合物及其使用方法

[0001] 相关申请的交叉引用

[0002] 本申请要求于2018年5月14日提交的美国临时申请号62/671,091、以及于2019年4月19日提交的美国临时申请号62/836,327的优先权和权益。上述专利申请各自的内容整体引入本文作为参考。

[0003] 序列表

[0004] 本申请包含序列表,所述序列表已经由EFS-Web以ASCII格式提交,并且在此整体引入作为参考。于2019年5月13日创建的所述ASCII拷贝命名为“NATE-039_001W0_SeqList.txt”,并且大小为25,129字节。

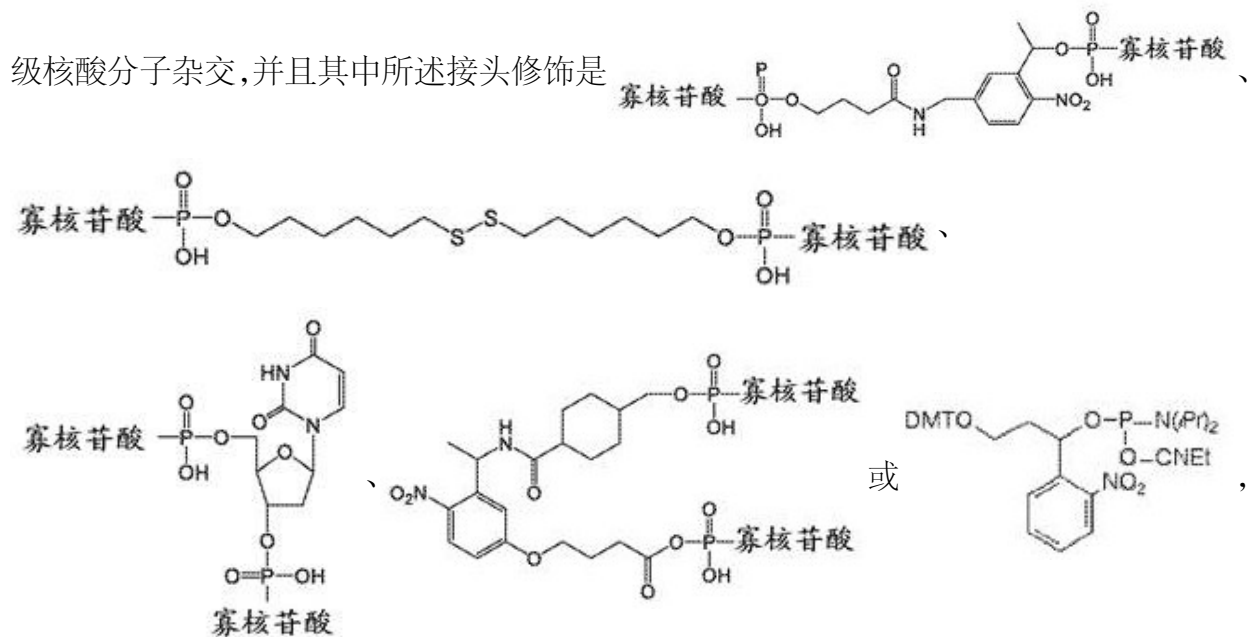
背景技术

[0005] 目前存在用于核酸测序的各种方法,即确定核酸分子内核苷酸的精确次序的过程。目前的方法需要例如PCR和/或通过克隆来酶促扩增核酸。需要进一步的酶促聚合,以产生通过光检测手段可检测的信号。此类扩增和聚合步骤是昂贵和/或耗时的。因此,本领域需要快速且无扩增和无酶的核酸测序方法。本公开内容解决了这些需求。

发明内容

[0006] 本公开内容提供了测序探针、方法、试剂盒和装置,其提供了具有长读取长度和低错误率的快速无酶、无扩增和无文库的核酸测序。本文所述的测序探针包括条形码结构域,其中所述条形码结构域中的每个位置对应于靶结合结构域中的至少两个核苷酸。此外,所述方法、试剂盒和装置具有快速的采样-应答能力。这些特征对于临床环境中的测序特别有用。本公开内容是公开于专利公开号U.S. 2016/0194701中的公开内容的改进,所述专利公开的内容整体引入本文作为参考。

[0007] 本公开内容提供了包含靶结合结构域和条形码结构域的探针;其中所述靶结合结构域包含至少八个核苷酸并且与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸鉴定靶核酸分子中的相应核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包含合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着位置,所述至少一个核酸序列与互补核酸分子杂交,并且其中所述合成主链包含L-DNA,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中所述至少三个附着位置的每个位置的核酸序列,确定由靶结合结构域结合的靶核酸中至少六个核苷酸的相应两个核苷酸的位置和身份;以及与至少三个附着位置的第一附着位置杂交的第一互补一级核酸分子,其中所述第一一级互补核酸分子包含至少两个结构域和可切割接头,其中所述第一结构域与条形码结构域的第一附着位置杂交,并且所述第二结构域能够与至少一个互补的二



并且其中所述接头修饰位于第一结构域和第二结构域之间。

[0008] 探针可以包含约60个核苷酸。探针可以包含在靶结合结构域和条形码结构域之间的单链DNA合成主链和双链DNA间隔区。单链DNA合成主链可以包含L-DNA。单链DNA合成主链可以包含约27个核苷酸。双链DNA间隔区可以包含L-DNA。双链DNA间隔区可以包含长度约25个核苷酸。

[0009] 探针的靶结合结构域中的核苷酸数目可以大于探针的条形码结构域中的附着位置数目。靶结合结构域可以包含八个核苷酸,而条形码结构域可以包含三个附着位置。靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少一个核苷酸,可以在靶结合结构域中的至少六个核苷酸之前,并且其中靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少一个核苷酸,可以在靶结合结构域中的至少六个核苷酸之后。

[0010] 条形码结构域中的附着位置可以包含一个附着区域。条形码结构域中的每个附着位置的至少一个核酸序列可以包含约9个核苷酸。附着位置的至少一个核酸序列可以包含3'末端鸟苷核苷酸。每个附着位置的至少一个核酸序列可以包含至少一个腺嘌呤核苷酸,至少一个胸腺嘧啶核苷酸,至少一个胞嘧啶核苷酸或其任何组合,以及3'末端鸟苷核苷酸。附着位置的至少一个核酸序列的每个核苷酸可以是L-DNA。靶结合结构域的至少八个核苷酸的每个核苷酸可以是D-DNA。

[0011] 互补核酸分子可以是一级核酸分子,其中所述一级核酸分子可以直接结合条形码结构域的至少一个附着位置内的至少一个附着区域。一级核酸分子可以包含至少两个结构域:能够结合条形码结构域的至少一个附着位置内的至少一个附着区域的第一结构域,以及能够结合至少一个互补的二级核酸分子的第二结构域。一级核酸分子的第一结构域可以包含L-DNA。一级核酸分子的第二结构域可以包含D-DNA。一级核酸分子的第一结构域可以包含5'末端胞嘧啶核苷酸。一级核酸分子的第一结构域可以包含至少一个腺嘌呤核苷酸,至少一个胸腺嘧啶核苷酸,至少一个鸟嘌呤核苷酸或其任何组合,以及5'末端胞嘧啶核苷酸。可切割接头可以位于一级核酸分子的第一结构域和一级核酸分子的第二结构域之间。可切割接头可以包含至少一个可切割部分。可切割部分可以是光可切割部分。

[0012] 一级核酸分子可以与条形码结构域的至少一个附着位置内的至少一个附着区域杂交,并且可以与至少一个二级核酸分子杂交。一级核酸分子可以与四个二级核酸分子杂交。

[0013] 二级核酸分子可以包含至少两个结构域:能够与至少一个一级核酸分子中的互补序列结合的第一结构域;以及能够与以下结合的第二结构域:(a) 第一可检测标记和至少第二可检测标记,(b) 至少一个互补的三级核酸分子,或(c) 其组合。二级核酸分子可以包含可切割接头。可切割接头可以位于第一结构域和第二结构域之间。可切割接头可以是光可切割的。二级核酸分子可以与至少一个一级核酸分子杂交,并且与至少一个三级核酸分子杂交。二级核酸分子可以与以下杂交:(a) 至少一个一级核酸分子,(b) 至少一个三级核酸分子,以及(c) 第一可检测标记和至少第二可检测标记。每个二级核酸分子可以与一个三级核酸分子杂交。第一可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。

[0014] 三级核酸分子可以包含至少两个结构域:能够结合二级核酸分子中的互补序列的第一结构域;以及能够结合第一可检测标记和至少第二可检测标记的第二结构域。三级核酸分子包含可切割接头。可切割接头可以位于第一结构域和第二结构域之间。可切割接头可以是光可切割的。三级核酸分子可以与至少一个二级核酸分子杂交,并且可以包含第一可检测标记和至少第二可检测标记。第一可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。

[0015] 位于二级核酸分子上的至少第一可检测标记和第二可检测标记可以具有相同的发射光谱,并且位于三级核酸分子上的至少第一可检测标记和第二可检测标记可以具有相同的发射光谱,并且其中二级核酸分子上的可检测标记的发射光谱可以不同于三级核酸分子上的可检测标记的发射光谱。

[0016] 一级核酸分子可以与四个二级核酸分子杂交,其中所述四个二级核酸分子各自包含四个第一可检测标记,并且其中所述四个二级核酸分子各自与一个三级核酸分子杂交,其中所述三级核酸分子包含五种可检测标记。二级核酸分子的第一可检测标记的发射光谱可以不同于三级核酸分子上的第二可检测标记的发射光谱。

[0017] 本公开内容提供了用于确定核酸的核苷酸序列的方法,其包括(1)使权利要求1的至少一个第一探针的靶结合结构域与靶核酸的第一区域杂交,所述靶核酸任选地在一个或多个位置处固定至基底;(2)使包含至少一个第一可检测标记和至少一个第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置杂交;(3)鉴定与第一附着位置杂交的第一互补核酸分子的至少一个第一可检测标记和至少一个第二可检测标记;(4)去除与第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记;(5)使包含至少一个第三可检测标记和至少一个第四可检测标记的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置杂交;(6)鉴定与第二附着位置杂交的第二互补核酸分子的至少一个第三可检测标记和至少一个第四可检测标记;(7)去除与第二附着位置杂交的至少一个第三可检测标记和至少一个第四可检测标记;(8)使包含至少一个第五可检测标记和至少一个第六可检测标记的第三互补核酸分子,与条形码结构域的至少三个附着位置的第三附着位置杂交;(9)鉴定与第三附着位置杂交的第三互补核酸分子的至少一个第五可检测标记和至少一个第六可检测标记;并且(10)基于至少一个

第一可检测标记、至少一个第二可检测标记、至少一个第三可检测标记、至少一个第四可检测标记、至少一个第五可检测标记和至少一个第六可检测标记的身份,来确定与至少一个第一探针的靶结合结构域的至少六个核苷酸杂交的、任选固定的靶核酸的至少六个核苷酸的核苷酸序列。

[0018] 前述方法可以进一步包括(11)从任选固定的靶核酸的第一区域去除至少一个第一探针;(12)使权利要求1的至少一个第二探针的靶结合结构域与任选固定的靶核酸的第二区域杂交,并且其中第一探针和至少第二探针的靶结合结构域是不同的;(13)使包含至少一个第七可检测标记和至少一个第八可检测标记的第四互补核酸分子,与至少一个第二探针的条形码结构域的至少三个附着位置的第一附着位置杂交;(14)鉴定与第一附着位置杂交的第四互补核酸分子的至少一个第七可检测标记和至少一个第八可检测标记;(15)去除与第一附着位置杂交的至少一个第七可检测标记和至少一个第八可检测标记;(16)使包含至少一个第九可检测标记和至少一个第十可检测标记的第五互补核酸分子,与至少第二探针的条形码结构域的至少三个附着位置的第二附着位置杂交;(17)鉴定与第二附着位置杂交的第五互补核酸分子的至少一个第九可检测标记和至少一个第十可检测标记;(18)去除与第二附着位置杂交的至少一个第九可检测标记和至少一个第十可检测标记;(19)使包含至少一个第十一可检测标记和至少一个第十二可检测标记的第六互补核酸分子,与至少第二探针的条形码结构域的至少三个附着位置的第三附着位置杂交;(20)鉴定与第三附着位置杂交的第六互补核酸分子的至少一个第十一可检测标记和至少一个第十二可检测标记;并且(21)基于至少一个第七可检测标记、至少一个第八可检测标记、至少一个第九可检测标记、至少一个第十可检测标记、至少一个第十一可检测标记和至少一个第十二可检测标记的身份,来确定与至少一个第二探针的靶结合结构域的至少六个核苷酸杂交的、任选固定的靶核酸的至少六个核苷酸的核苷酸序列。

[0019] 前述方法可以进一步包括组装任选固定的靶核酸的至少第一区域和至少第二区域中的每个鉴定的核苷酸线性次序,从而鉴定任选固定的靶核酸的序列。

[0020] 步骤(4)和(5)可以序贯或同时发生。步骤(7)和(8)可以序贯或同时发生。

[0021] 第一可检测标记和第二可检测标记可以具有相同的发射光谱,或具有不同的发射光谱。第三可检测标记和第四可检测标记可以具有相同的发射光谱,或具有不同的发射光谱。第五可检测标记和第六可检测标记可以具有相同的发射光谱,或具有不同的发射光谱。

[0022] 第一互补核酸分子、第二互补核酸分子和第三互补核酸分子可以包含可切割接头。可切割接头可以是光可切割的。

[0023] 第一互补核酸分子可以包含一级核酸、四个二级核酸分子和四个三级核酸分子,其中所述一级核酸与四个二级核酸分子杂交,其中所述四个二级核酸分子各自包含四个第一可检测标记,并且其中所述四个二级核酸分子各自与一个三级核酸分子杂交,其中所述四个三级核酸分子各自包含五个第二可检测标记。

[0024] 一级核酸分子可以包含至少两个结构域:与条形码结构域的第一附着位置杂交的第一结构域,以及与四个二级核酸分子杂交的第二结构域。一级核酸分子可以包含位于第一结构域和第二结构域之间的可切割接头。

[0025] 二级核酸分子可以包含至少两个结构域:与一级核酸分子的第二结构域杂交的第一结构域;以及包含四个第一可检测标记,并且与一个三级核酸分子杂交的第二结构域。二

级核酸分子可以包含位于第一结构域和第二结构域之间的可切割接头。

[0026] 去除与第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记,可以包括切割一级核酸的第一结构域和第二结构域之间的可切割接头、切割每个二级核酸的第一结构域和第二结构域之间的可切割接头、或其任何组合。

[0027] 本公开内容提供了包含至少一种分子复合物的组合物,其中所述至少一种分子复合物包含:(A)从生物样品获得的靶核酸分子,以及(B)至少两种核酸分子复合物,其中第一复合物包含第一部分双链核酸分子,其中所述第一部分双链核酸分子的一条链包含:与靶核酸分子的第一部分杂交的靶特异性结构域、对第一部分双链核酸分子的另一条链退火的双链体结构域、以及至少一个第一亲和力部分,其中所述第一部分双链核酸分子的另一条链包含:对第一部分双链核酸分子的另一条链退火的双链体结构域、与附着至基底的互补核酸杂交的基底特异性结构域、以及至少一个第二亲和力部分,其中所述第二复合物包含第二部分双链核酸分子,其中所述第二部分双链核酸分子的一条链包含:与靶核酸的第二部分杂交的靶特异性结构域,其中所述第一部分和第二部分并不重叠,以及对第二部分双链核酸分子的另一条链退火的双链体结构域,其中所述第二部分双链核酸分子的另一条链包含:对第二部分双链核酸分子的另一条链退火的双链体结构域、鉴定靶核酸由其获得的生物样品的样品特异性结构域、第一单链纯化序列、位于双链体结构域和样品特异性结构域之间的第一可切割部分、以及位于样品特异性结构域和第一单链纯化序列之间的第二可切割部分。

[0028] 本公开内容提供了包含至少一种分子复合物的组合物,其中所述至少一种分子复合物包含:(A)从生物样品获得的靶核酸分子,以及(B)至少两种核酸分子复合物,其中第一复合物包含第一部分双链核酸分子,其中所述第一部分双链核酸分子的一条链包含:与靶核酸分子的第一部分杂交的靶特异性结构域、对第一部分双链核酸分子的另一条链退火的双链体结构域、以及至少一个第一亲和力部分,其中所述第一部分双链核酸分子的另一条链包含:对第一部分双链核酸分子的另一条链退火,并且与靶核酸分子的3'末端可操作连接的双链体结构域,与附着至基底的互补核酸杂交的基底特异性结构域,以及至少一个第二亲和力部分,其中所述第二复合物包含第二部分双链核酸分子,其中所述第二部分双链核酸分子的一条链包含:与靶核酸的第二部分杂交的靶特异性结构域,其中所述第一部分和第二部分并不重叠,以及对第二部分双链核酸分子的另一条链退火的双链体结构域,其中所述第二部分双链核酸分子的另一条链包含:对第二部分双链核酸分子的另一条链退火,并且与靶核酸分子的5'末端可操作连接的双链体结构域,鉴定靶核酸由其获得的生物样品的样品特异性结构域,以及位于双链体结构域和样品特异性结构域之间的第一可切割部分。

[0029] 本公开内容提供了包含至少一种分子复合物的组合物,其中所述至少一种分子复合物包含:(A)从生物样品获得的靶核酸分子,以及(B)至少两种核酸分子复合物,其中第一复合物包含第一部分双链核酸分子,其中所述第一部分双链核酸分子的一条链包含:与靶核酸分子的第一部分杂交的靶特异性结构域、对第一部分双链核酸分子的另一条链退火的双链体结构域、以及至少一个第一亲和力部分,其中所述第一部分双链核酸分子的另一条链包含:对第一部分双链核酸分子的另一条链退火,并且与靶核酸分子的3'末端可操作连接的双链体结构域,与附着至基底的互补核酸杂交的基底特异性结构域,以及至少一个第

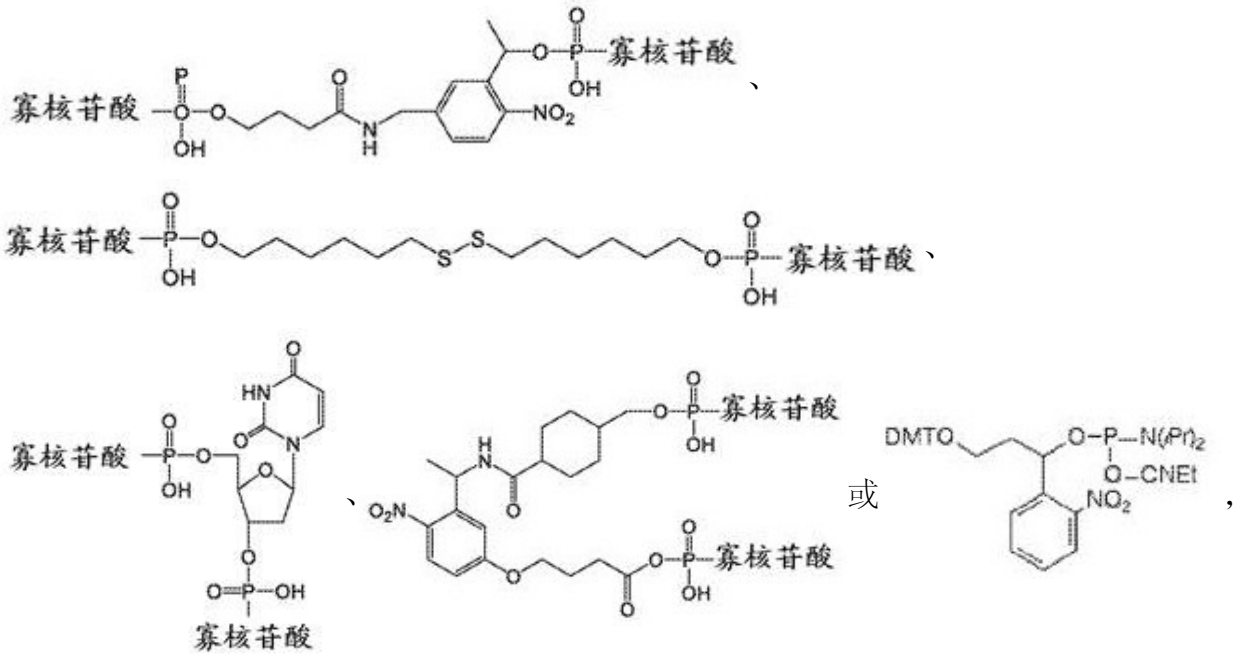
二亲和力部分,其中所述第二复合物包含第二部分双链核酸分子,其中所述第二部分双链核酸分子的一条链包含:与靶核酸的第二部分杂交的靶特异性结构域,其中所述第一部分和第二部分并不重叠,以及对第二部分双链核酸分子的另一条链退火的双链体结构域,其中所述第二部分双链核酸分子的另一条链包含:对第二部分双链核酸分子的另一条链退火,并且与靶核酸分子的5'末端可操作连接的双链体结构域。

[0030] 本公开内容还提供了包含以下的组合物:平面固体支撑基底;在平面固体支撑基底上的第一层;在第一层上的第二层;其中所述第二层包含多个纳米孔,其中每个纳米孔提供了对第一层的暴露部分的接近,其中每个纳米孔包含共价附着到第一层的暴露部分的多个第一寡核苷酸。

[0031] 本公开内容提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含至少八个核苷酸并且与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸鉴定靶核酸分子中的相应核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包含合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列与互补核酸分子杂交,其中所述至少三个附着位置的核酸序列,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的位置和身份,并且其中所述至少三个附着位置各自具有不同的核酸序列。

[0032] 本公开内容还提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含至少八个核苷酸并且与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸鉴定靶核酸分子中的相应核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包含合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着位置,所述至少一个核酸序列与互补核酸分子杂交,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中所述至少三个附着位置的每个位置的核酸序列,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0033] 本公开内容提供了包含以下的复合物:a) 包含靶结合结构域和条形码结构域的组合物;其中所述靶结合结构域包含至少八个核苷酸并且与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸鉴定靶核酸分子中的相应核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包含合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列与互补核酸分子杂交,其中所述至少三个附着位置的核酸序列,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的位置和身份,并且其中所述至少三个附着位置各自具有不同的核酸序列;以及与至少三个附着位置的第一附着位置杂交的第一互补一级核酸分子,其中所述第一一级互补核酸分子包含至少两个结构域和可切割接头,其中所述第一结构域与条形码结构域的第一附着位置杂交,并且其中所述第二结构域能够与至少一个互补的二级核酸分子杂交,并且其中所述可切割接头是



并且其中所述可切割接头位于第一结构域和第二结构域之间。

[0034] 本公开内容提供了用于确定核酸的核苷酸序列的方法,其包括(1)使本公开内容的第一测序探针的靶结合结构域与靶核酸的第一区域杂交,所述靶核酸任选地在一个或多个位置处固定至基底;(2)使包含至少一个第一可检测标记和至少一个第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置杂交;(3)鉴定与第一附着位置杂交的第一互补核酸分子的至少一个第一可检测标记和至少一个第二可检测标记;(4)去除与第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记;(5)使包含至少一个第三可检测标记和至少一个第四可检测标记的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置杂交;(6)鉴定与第二附着位置杂交的第二互补核酸分子的至少一个第三可检测标记和至少一个第四可检测标记;(7)去除与第二附着位置杂交的至少一个第三可检测标记和至少一个第四可检测标记;(8)使包含至少一个第五可检测标记和至少一个第六可检测标记的第三互补核酸分子,与条形码结构域的至少三个附着位置的第三附着位置杂交;(9)鉴定与第三附着位置杂交的第三互补核酸分子的至少一个第五可检测标记和至少一个第六可检测标记;并且(10)基于至少一个第一可检测标记、至少一个第二可检测标记、至少一个第三可检测标记、至少一个第四可检测标记、至少一个第五可检测标记和至少一个第六可检测标记的身份,来确定与第一测序探针的靶结合结构域的至少六个核苷酸杂交的、任选固定的靶核酸的至少六个核苷酸的核苷酸序列。

[0035] 本公开内容提供了用于确定核酸的核苷酸序列的方法,其包括(1)使权利要求113或114的第一测序探针的靶结合结构域,与任选地在一个或多个位置处固定至基底的靶核酸杂交;(2)使包含至少一个第一可检测标记和至少一个第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置杂交;(3)鉴定与第一附着位置杂交的第一互补核酸分子的至少一个第一可检测标记和至少一个第二可检测标记;(4)基于至少一个第一可检测标记和至少一个第二可检测标记的身份,鉴定与靶结合结构域的至少六个核苷酸中的两个杂交的、任选固定的靶核酸中的第一核苷酸和第二核苷酸的位置和

身份；(5) 去除与第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记；(6) 使包含至少一个第三可检测标记和至少一个第四可检测标记的第二互补核酸分子，与条形码结构域的至少三个附着位置的第二附着位置杂交；(7) 鉴定与第二附着位置杂交的第二互补核酸分子的至少一个第三可检测标记和至少一个第四可检测标记；(8) 基于至少一个第三可检测标记和至少一个第四可检测标记的身份，鉴定与靶结合结构域的至少六个核苷酸中的两个杂交的、任选固定的靶核酸中的第三核苷酸和第四核苷酸的位置和身份；(9) 去除与第二附着位置杂交的至少一个第三可检测标记和至少一个第四可检测标记；(10) 使包含至少一个第五可检测标记和至少一个第六可检测标记的第三互补核酸分子，与条形码结构域的至少三个附着位置的第三附着位置杂交；(11) 鉴定与第三附着位置杂交的第三互补核酸分子的至少一个第五可检测标记和至少一个第六可检测标记；并且(12) 基于至少一个第五可检测标记和至少一个第六可检测标记的身份，鉴定与靶结合结构域的至少六个核苷酸中的两个杂交的、任选固定的靶核酸中的第五核苷酸和第六核苷酸的位置和身份；从而确定与第一测序探针的靶结合结构域的至少六个核苷酸杂交的、任选固定的靶核酸的至少六个核苷酸的核苷酸序列。

[0036] 本公开内容还提供了用于鉴定靶核酸中的预定核苷酸序列的存在的方法，其包括(1) 使本公开内容的第一测序探针的靶结合结构域与靶核酸的第一区域杂交，所述靶核酸任选地在一个或多个位置处固定至基底；(2) 使包含至少一个第一可检测标记和至少一个第二可检测标记的第一互补核酸分子，与条形码结构域的至少三个附着位置的第一附着位置杂交；(3) 鉴定与第一附着位置杂交的第一互补核酸分子的至少一个第一可检测标记和至少一个第二可检测标记；(4) 去除与第一附着位置杂交的至少一个第一可检测标记和至少一个第二可检测标记；(5) 使包含至少一个第三可检测标记和至少一个第四可检测标记的第二互补核酸分子，与条形码结构域的至少三个附着位置的第二附着位置杂交；(6) 鉴定与第二附着位置杂交的第二互补核酸分子的至少一个第三可检测标记和至少一个第四可检测标记；(7) 去除与第二附着位置杂交的至少一个第三可检测标记和至少一个第四可检测标记；(8) 使包含至少一个第五可检测标记和至少一个第六可检测标记的第三互补核酸分子，与条形码结构域的至少三个附着位置的第三附着位置杂交；(9) 鉴定与第三附着位置杂交的第三互补核酸分子的至少一个第五可检测标记和至少一个第六可检测标记，从而基于至少一个第一可检测标记、至少一个第二可检测标记、至少一个第三可检测标记、至少一个第四可检测标记、至少一个第五可检测标记和至少一个第六可检测标记的身份，来确定预定核苷酸序列的存在。

[0037] 本公开内容提供了包含以下的试剂盒：(A) 包含第一部分双链核酸分子的第一核酸分子复合物，其中所述第一部分双链核酸分子的一条链包含：与靶核酸分子的第一部分杂交的靶特异性结构域、对第一部分双链核酸分子的另一条链退火的双链体结构域、至少一个第一亲和力部分，其中所述第一部分双链核酸分子的另一条链包含：对第一部分双链核酸分子的另一条链退火的双链体结构域、与附着至基底的互补核酸杂交的基底特异性结构域、以及至少一个第二亲和力部分，以及(B) 包含第二部分双链核酸分子的第二核酸分子复合物，其中所述第二部分双链核酸分子的一条链包含：与靶核酸的第二部分杂交的靶特异性结构域，其中所述第一部分和第二部分并不重叠，以及对第二部分双链核酸分子的另一条链退火的双链体结构域，并且其中所述第二部分双链核酸分子的另一条链包含：对第

二部分双链核酸分子的另一条链退火的双链体结构域、鉴定靶核酸由其获得的生物样品的样品特异性结构域、与附着至基底的互补核酸杂交的基底特异性结构域、第一单链纯化序列、位于双链体结构域和样品特异性结构域之间的第一可切割部分、以及位于样品特异性结构域和第一单链纯化序列之间的第二可切割部分。

[0038] 本公开内容还提供了包含以下的试剂盒:第一单链核酸分子,其包含:与靶核酸分子的第一部分杂交的靶特异性结构域、对第二单链核酸分子的双链体结构域退火的双链体结构域、以及至少一个第一亲和力部分,(B)第二单链核酸分子,其包含:对第一单链核酸分子的双链体结构域退火的双链体结构域、与附着至基底的互补核酸杂交的基底特异性结构域、以及至少一个第二亲和力部分,(C)第三单链核酸分子,其包含:与靶核酸的第二部分杂交的靶特异性结构域,其中所述第一部分和第二部分并不重叠,以及对第四单链核酸分子的双链体结构域退火的双链体结构域,(D)第四单链核酸分子,其包含:对第三单链核酸分子的双链体结构域退火的双链体结构域、鉴定靶核酸由其获得的生物样品的样品特异性结构域、第一单链纯化序列、位于双链体结构域和样品特异性结构域之间的第一可切割部分、以及位于样品特异性结构域和第一单链纯化序列之间的第二可切割部分。

[0039] 以上任何方面都可以与任何其它方面组合。

[0040] 除非另有定义,否则本文中使用的所有技术和科学术语都具有与本公开内容所属领域的普通技术人员通常理解相同的含义。在本说明书中,除非上下文另有明确规定,否则单数形式也包括复数;作为实例,术语“一个”、“一种”和“该/所述”应理解为单数或复数,并且术语“或”应理解为包括性的。举例来说,“要素”意指一个或多个要素。在整个说明书中,词语“包含(comprising)”或者变化如“包含(comprises)”或“包含(comprising)”,应理解为暗示包括陈述的要素、整数或步骤,或者要素、整数或步骤组,但不排除任何其他要素、整数或步骤,或者要素、整数或步骤组。约可以理解为在所述值的10%、9%、8%、7%、6%、5%、4%、3%、2%、1%、0.5%、0.1%、0.05%或0.01%内。除非从上下文另外明确的,否则本文提供的所有数值都由术语“约”修饰。

[0041] 尽管与本文描述的那些类似或等价的方法和材料都可以用于本公开内容的实践或测试中,但下文描述了合适的方法和材料。本文提到的所有出版物、专利申请、专利和其它参考文献都整体引入作为参考。本文引用的参考文献并不承认是本发明的现有技术。在冲突的情况下,以本说明书包括定义为准。另外,材料、方法和实施例仅是说明性的,并不预期是限制性的。根据下述详述和权利要求,本公开内容的其它特征和优点将是显而易见的。

附图说明

[0042] 专利或申请文件含有至少一张用彩色绘制的附图。具有彩色附图的本专利或专利申请公开的副本将在请求和支付必要费用后由专利局提供。

[0043] 当与附图结合时,根据下述详述,将更明确地了解上述以及进一步的特征。

[0044] 图1是本公开内容的一种示例性测序探针的图示。

[0045] 图2显示了本公开内容的标准、三部分测序和单部分接头探针的设计。

[0046] 图3是与本公开内容的示例性测序探针杂交的本公开内容的示例性报告复合物的图示。

[0047] 图4显示了本公开内容的示例性报告探针的示意图。

- [0048] 图5是包含三级核酸的不同排列的本公开内容的几种示例性报告探针的示意图。
- [0049] 图6是包含分支三级核酸的本公开内容的几种示例性报告探针的示意图。
- [0050] 图7显示了本公开内容的示例性报告探针内的可切割接头修饰的可能位置。
- [0051] 图8是使用本公开内容的双捕获探针系统捕获靶核酸的示意图。
- [0052] 图9显示了使用FFPE样品,使用本发明的方法捕获且检测由100个靶组成的多路复用癌症组的实验结果。
- [0053] 图10是本公开内容的测序方法的单个循环的示意图。
- [0054] 图11是本公开内容的测序方法的一个循环的示意图,以及在该循环期间收集的相应成像数据。
- [0055] 图12示出了本公开内容的示例性测序探针池配置,其中八种颜色组合用于设计八个不同的测序探针池。
- [0056] 图13将U.S. 2016/019470中公开的条形码结构域设计与本公开内容的条形码结构域设计进行比较。
- [0057] 图14是本公开内容的测序循环的示意图,其中可切割接头修饰用于使条形码位置变暗。
- [0058] 图15是本公开内容的示例性测序循环的说明性实例,其中条形码结构域内的位置通过一级核酸的置换而变暗。
- [0059] 图16是本公开内容的测序方法如何允许用不同的测序探针对靶核酸的相同碱基进行测序的示意图。
- [0060] 图17显示了可以如何组合从一种或多种测序探针记录的、靶核酸上的特异性核苷酸位置的多重碱基识别 (call),以产生一致序列,从而增加最终碱基识别的准确率。
- [0061] 图18显示了来自使用本公开内容的测序方法获得,并且使用组装算法进行分析的测序实验的结果。对于左图上的图表,从左上方图表开始顺时针方向前进,显示的序列对应于SEQ ID NO:3、4、6、8、7和5。对于右侧上的表格,从顶部开始向下移动,序列对应于SEQ ID NO:3、4、7、8、6和5。
- [0062] 图19显示了来自FFPE样品的癌基因靶的多路复用捕获和测序的实验设计的示意图。
- [0063] 图20显示了直接RNA测序的说明性示意图,以及来自测试RNA分子与本公开内容的测序方法的相容性的实验的结果。
- [0064] 图21显示了使用本公开内容的测序方法,对具有相同核苷酸序列的RNA分子和DNA分子进行测序。
- [0065] 图22显示了本公开内容的标准和三部分测序探针的性能比较。
- [0066] 图23显示了使用个别探针,在本公开内容的示例性的靶结合结构域内的LNA取代的效应。
- [0067] 图24显示了使用九种探针的池,在本公开内容的示例性的靶结合结构域内的LNA取代的效应。
- [0068] 图25显示了在本公开内容的示例性的靶结合结构域中,修饰的核苷酸和核酸类似物取代的效应。
- [0069] 图26显示了来自量化本公开内容的测序方法的原始准确率的实验的结果。

- [0070] 图27显示了当通过多于一种测序探针对靶核酸中的核苷酸进行测序时,来自确定本公开内容的测序方法的准确率的实验的结果。
- [0071] 图28是包含口袋寡核苷酸(pocket oligo)的本公开内容的测序探针的示意图。
- [0072] 图29是本公开内容的测序探针的示意图,其包含在每个附着位置之间的PEG接头区域。
- [0073] 图30是本公开内容的测序探针的示意图,其包含在每个附着位置之间的无碱基区域(abasic region)。
- [0074] 图31是经由连接子寡核苷酸,与本公开内容的示例性测序探针间接杂交的本公开内容的示例性报告复合物的图示。
- [0075] 图32是在本公开内容的方法中使用的奇偶性方案(parity scheme)的图示。
- [0076] 图33是本发明的捕获探针、衔接子寡核苷酸和草坪寡核苷酸复合物(lawn oligonucleotide complex)的示意图。
- [0077] 图34是与靶核酸杂交的本公开内容的c5探针复合物和c3探针复合物的示意图。
- [0078] 图35是在用FEN1消化之后,本公开内容的靶核酸-c3探针-c5探针复合物的示意图。
- [0079] 图36是在连接之后,本公开内容的靶核酸-c3探针-c5探针复合物的示意图。
- [0080] 图37是USER介导的本公开内容的靶核酸-c3探针-c5探针复合物的切割的示意图。
- [0081] 图38是在USER介导的切割之后,本公开内容的靶核酸-c3探针-c5探针复合物的示意图。
- [0082] 图39是UV介导的本公开内容的靶核酸-c3探针-c5探针复合物的切割的示意图。
- [0083] 图40是经由互补核酸附着至基底,在UV介导的切割之后,本公开内容的靶核酸-c3探针-c5探针复合物的示意图。
- [0084] 图41是与靶核酸杂交的本公开内容的c3.2探针复合物和c5.2探针复合物的示意图。
- [0085] 图42是在c3.2和c5.2探针复合物连接之后,本公开内容的靶核酸复合物的示意图。
- [0086] 图43是在本公开内容的靶核酸复合物中,单链纯化序列的切割和释放的示意图。
- [0087] 图44是固定在本公开内容的基底上的本公开内容的靶核酸复合物的示意图。
- [0088] 图45是在本公开内容的靶核酸复合物固定至本公开内容的基底之后,基底特异性结构域的切割和释放的示意图。
- [0089] 图46是固定在本公开内容的基底上的基底特异性结构域释放之后,本公开内容的靶核酸复合物的示意图。
- [0090] 图47是本发明的示例性阵列的示意性横截面。
- [0091] 图48是本发明的示例性阵列的示意性横截面,所述阵列包括具有金字塔形状的纳米孔。
- [0092] 图49是本公开内容的示例性阵列的示意图解,所述阵列包括以随机模式排列的多个圆柱形纳米孔。
- [0093] 图50是本公开内容的示例性阵列的示意图解,所述阵列包括以恒定间距排列在有序网格中的圆柱形纳米孔。

[0094] 图51是本发明的示例性阵列的示意性横截面,其中单个靶核酸复合物固定在每个纳米孔中。

[0095] 图52是本发明的示例性阵列的示意性横截面,其中单个靶核酸复合物固定在每个纳米孔中,从而防止了其它靶核酸复合物的固定。

[0096] 图53是本公开内容的测序探针的示意图,所述测序探针完全由L-DNA组成,并且包含具有3'末端L-dG核苷酸的附着区域。

[0097] 图54是本公开内容的测序探针的示意图,所述测序探针完全由D-DNA组成,并且包含位于附着区域1(位点(spot)1)和附着区域2(位点2)之间、以及附着区域2(位点2)和附着区域3(位点3)之间的口袋寡核苷酸。

[0098] 图55是使用与蛋白质锁组合的捕获探针和草坪寡核苷酸,固定在固体基底上的合成靶核酸的示意图。

[0099] 图56是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。x轴指示了待测序的靶核酸的特异性核苷酸。顶部图表显示了关于LG-间隔和D-口袋测序探针的理论测序多样性,观察到的测序多样性和观察到的测序覆盖率。红色框指示了预测的有问题的测序区域。

[0100] 图57是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。x轴指示了待测序的靶核酸的特异性核苷酸。顶部图表显示了关于LG-间隔和D-口袋测序探针的理论测序多样性,观察到的测序多样性和观察到的测序覆盖率。红色框指示了预测的有问题的测序区域。

[0101] 图58是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。x轴指示了待测序的靶核酸的特异性核苷酸。顶部图表显示了关于LG-间隔和D-口袋测序探针的理论测序多样性,观察到的测序多样性和观察到的测序覆盖率。红色框指示了预测的有问题的测序区域。

[0102] 图59是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。x轴指示了待测序的靶核酸的特异性核苷酸。顶部图表显示了关于LG-间隔和D-口袋测序探针,观察到的测序多样性和观察到的测序覆盖率。

[0103] 图60是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。x轴指示了待测序的靶核酸的特异性核苷酸。顶部图表显示了关于LG-间隔和D-口袋测序探针,观察到的测序多样性和观察到的测序覆盖率。

[0104] 图61是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。x轴指示了待测序的靶核酸的特异性核苷酸。顶部图表显示了关于LG-间隔和D-口袋测序探针,观察到的测序多样性和观察到的测序覆盖率。

[0105] 图62是一系列直方图,其显示了在使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验中,条形码事件的总数和有效的3位点读出的数目。

[0106] 图63是一系列图表,其显示了在使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验中,中靶事件、无效事件、脱靶事件、在 b_1 - b_6 事件时的1次错误、在 b_1 - b_6 事件时的2次错误、在 b_1 - b_6 事件时的3次错误、在 b_1 - b_6 事件时的4次错误、在 b_1 - b_6 事件时的5次错误和在 b_1 - b_6 事件时的6次错误的总数。

[0107] 图64是一系列图表,其显示了在使用本公开内容的LG-间隔的测序探针和D-口袋

测序探针的测序实验中,中靶事件、无效事件、脱靶事件、在 b_1 - b_6 事件时的1次错误、在 b_1 - b_6 事件时的2次错误、在 b_1 - b_6 事件时的3次错误、在 b_1 - b_6 事件时的4次错误、在 b_1 - b_6 事件时的5次错误和在 b_1 - b_6 事件时的6次错误的总数。

[0108] 图65是显示了在使用本公开内容的D-口袋测序探针(循环1-50)和LG-间隔的测序探针(循环51-100)的测序实验的每个循环中,1个观察者(spotter)(可能的三个报告探针中仅一个被成功地记录)、2个观察者(可能的三个报告探针中仅两个被成功地记录)、以及3个观察者(所有三个可能的报告探针都被成功地记录)事件数目的图表。

[0109] 图66是显示了使用本公开内容的LG-间隔的测序探针和D-口袋测序探针的测序实验结果的一系列图表。最左侧的图显示了在测序实验的每个循环中记录的中靶、新六聚体、冗余六聚体、脱靶和无效事件的数目。循环1-50使用D-口袋测序探针执行,并且循环51-100使用本公开内容的LG-间隔的测序探针执行。

[0110] 图67是使用本公开内容的方法和组合物,将靶核酸固定至固体基底的示意图。使用位于捕获探针和草帽寡核苷酸上的生物素部分与中性抗生物素蛋白(neutravidin)部分之间的蛋白质锁,将靶核酸固定。

[0111] 发明详述

[0112] 本公开内容提供了测序探针、报告探针、方法、试剂盒和装置,其提供了具有长读取长度和低错误率的快速无酶、无扩增和无文库的核酸测序。

[0113] 本公开内容的组合物

[0114] 本公开内容提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含表1中所述的任何构建体。示例性的靶结合结构域包含至少八个核苷酸并且能够与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述靶结合结构域中的至少六个核苷酸中的任何可以是修饰的核苷酸或核苷酸类似物,并且其中所述靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何碱基、或者通用碱基或简并碱基。示例性的条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列能够由互补核酸分子结合,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中所述至少三个附着位置的每个位置的核酸序列,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0115] 在其它方面,示例性的靶结合结构域可以包含能够与靶核酸杂交的至少六个核苷酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸;其中所述靶结合结构域中的至少六个核苷酸中的任何核苷酸可以是修饰的核苷酸或核苷酸类似物。

[0116] 本公开内容还提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含至少十个核苷酸并且能够结合靶核酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至

少四个核苷酸并不鉴定靶核酸分子中的相应核苷酸；其中所述条形码结构域包括合成主链，该条形码结构域包含至少三个附着位置，每个附着位置包括包含至少一个核酸序列的至少一个附着区域，所述至少一个核酸序列能够由互补核酸分子结合，其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸，并且所述至少三个附着位置各自具有不同的核酸序列，并且其中所述至少三个附着位置的每个位置的核酸序列，确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0117] 本公开内容还提供了测序探针群体，其包含多种本文公开的任何测序探针。

[0118] 下文更详细地描述了所公开的测序探针的靶结合结构域、条形码结构域和主链、以及互补核酸分子（例如，报告分子或报告复合物）。

[0119] 本公开内容的测序探针包含靶结合结构域和条形码结构域。图1是本公开内容的示例性测序探针的示意图。图1显示了靶结合结构域能够结合靶核酸。靶核酸可以是本公开内容的测序探针可以与其杂交的任何核酸。靶核酸可以是DNA或RNA。靶核酸可以从来自受试者的生物样品获得。术语“靶结合结构域”和“测序结构域”在本文可互换使用。

[0120] 靶结合结构域可以包含一系列核苷酸（例如是多核苷酸）。靶结合结构域可以包含DNA、RNA或其组合。在靶结合结构域是多核苷酸的情况下，靶结合结构域通过与靶核酸的一部分杂交而与靶核酸结合，所述靶核酸的一部分与测序探针的靶结合结构域互补，如图1中所示。

[0121] 可以设计测序探针的靶结合结构域，以控制测序探针杂交和/或去杂交的可能性、以及这些发生的速率。一般地，探针的 T_m 越低，探针与/从靶核酸去杂交越快且越可能。因此，使用较低的 T_m 探针将减少与靶核酸结合的探针的数目。

[0122] 靶结合结构域的长度部分地影响探针与靶核酸杂交并保持杂交的可能性。一般地，靶结合结构域越长（核苷酸的数目越多），互补序列在靶核苷酸中存在的可能性越小。相反，靶结合结构域越短，互补序列在靶核苷酸中存在的可能性越大。例如，四聚体序列位于靶核酸中的概率为 $1/256$ ，而六聚体序列位于靶核酸中的概率为 $1/4096$ 。因此，当与较长探针的集合相比较时，较短探针的集合对于给定的核酸段可能在更多位置中结合。

[0123] 在多种情况下，优选具有具有较短靶结合结构域的探针，以增加在给定核酸段中的读取数目，从而富集靶核酸或靶核酸的一部分，尤其是特别关注的一部分的覆盖，例如在检测突变或SNP等位基因时。

[0124] 靶结合结构域可以是任何量或数目的核苷酸的长度。靶结合结构域可以是长度至少12个核苷酸、长度至少10个核苷酸、长度至少8个核苷酸、长度至少6个核苷酸或长度至少3个核苷酸。

[0125] 靶结合结构域中的每个核苷酸可以鉴定（或编码）靶分子的互补核苷酸。可替代地，靶结合结构域中的一些核苷酸鉴定（或编码）靶分子的互补核苷酸，而靶结合结构域中的一些核苷酸并不鉴定（或编码）靶分子的互补核苷酸。

[0126] 靶结合结构域可以包含至少一个天然碱基。靶结合结构域可以不包含天然碱基。靶结合结构域可以包含至少一个修饰的核苷酸或核酸类似物。靶结合结构域可以不包含修饰的核苷酸或核酸类似物。靶结合结构域可以包含至少一个通用碱基。靶结合结构域可以不包含通用碱基。靶结合结构域可以包含至少一个简并碱基。靶结合结构域可以不包含简

并碱基。

[0127] 靶结构域可以包含任何组合的天然碱基(例如0、1、2、3、4、5、6、7、8、9、10个或更多个天然碱基),修饰的核苷酸或核酸类似物(例如0、1、2、3、4、5、6、7、8、9、10个或更多个修饰的核苷酸或核酸类似物),通用碱基(例如0、1、2、3、4、5、6、7、8、9、10个或更多个通用碱基),或简并碱基(例如0、1、2、3、4、5、6、7、8、9、10个或更多个简并碱基)。当以组合存在时,特定靶结合结构域的天然碱基、修饰的核苷酸或核酸类似物、通用碱基和简并碱基可以以任何次序排列。

[0128] 术语“修饰的核苷酸”或“核酸类似物”包括但不限于锁核酸(LNA),桥接核酸(BNA),丙炔修饰的核酸,拉链核酸(ZNA[®]),异鸟嘌呤,异胞嘧啶,6-氨基-1-(4-羟基-5-羟基甲基-四氢呋喃-2-基)-1,5-二氢-吡啶并[3,4-d]嘧啶-4-酮(PPG),以及2'-修饰的核酸例如2'-O-甲基核酸。靶结合结构域可以包括零至六个(例如0、1、2、3、4、5或6个)修饰的核苷酸或核酸类似物。优选地,修饰的核苷酸或核酸类似物是锁核酸(LNA)。

[0129] 如本文使用的,术语“锁核酸(LNA)”包括但不限于修饰的RNA核苷酸,其中核糖部分包含连接2'氧和4'碳的亚甲基桥。这种亚甲基桥将核糖以3-内构象(也称为北构象,其在A型RNA双链体中发现)锁定。术语难接近的RNA可以与LNA互换使用。如本文使用的,术语“桥接核酸(BNA)”包括但不限于修饰的RNA分子,其包含具有固定的3'-内构象(也称为北构象)的五元或六元桥接结构。桥接结构将核糖的2'氧与核糖的4'碳连接。含有碳、氮和氢原子的各种不同的桥结构是可能的。如本文使用的,术语“丙炔修饰的核酸”包括但不限于在核酸碱基的C5位置处包含丙炔修饰的嘧啶,即胞嘧啶和胸腺嘧啶/尿嘧啶。如本文使用的,术语“拉链核酸(ZNA[®])”包括但不限于与阳离子精胺部分缀合的寡核苷酸。

[0130] 如本文使用的,术语“通用碱基”包括但不限于这样的核苷酸碱基,其并不遵循沃森-克里克碱基对规则,而是可以结合位于靶核酸上的四种规范碱基(A、T/U、C、G)中的任何碱基。如本文使用的,术语“简并碱基”包括但不限于这样的核苷酸碱基,其并不遵循沃森-克里克碱基对规则,而是可以结合四种规范碱基(A、T/U、C、G)中的至少两种,但不是全部四种。简并碱基也可以称为摇摆碱基;这些术语在本文中可互换使用。

[0131] 图1中描绘的示例性的测序探针示出了靶结合结构域,其包含与待测序靶核酸的互补核苷酸1-6特异性杂交的六核苷酸长度(6聚体)序列(b_1 - b_2 - b_3 - b_4 - b_5 - b_6)。靶结合结构域的这个6聚体部分(b_1 - b_2 - b_3 - b_4 - b_5 - b_6)鉴定(或编码)靶序列中的互补核苷酸(1-2-3-4-5-6)。这个6聚体序列在任一侧上侧翼为碱基(N)。由(N)指示的碱基可以独立地是通用碱基或简并碱基。通常,由(N)指示的碱基独立地是规范碱基之一。由(N)指示的碱基并不鉴定(或编码)其在靶序列中结合的互补核苷酸,并且独立于(6聚体)序列(b_1 - b_2 - b_3 - b_4 - b_5 - b_6)的核酸序列。

[0132] 图1中描绘的测序探针可以与本公开内容的测序方法结合使用,以仅使用杂交反应对靶核酸进行测序,而不需要共价化学、酶或扩增。为了对靶核酸分子中的所有可能的6聚体序列进行测序,需要总共4096种测序探针($4^6=4096$)。

[0133] 图1是本公开内容的序列探针的靶结合结构域的一种配置的示例。表1提供了本公开内容的靶结合结构域的几种其它配置。一种优选的靶结合结构域,称为“6 LNA”靶结合结构域,在靶结合结构域的位置b1至b6处包含6个LNA。这6个LNA在任一侧上侧翼为碱基(N)。如本文使用的,(N)碱基可以是独立于(6聚体)序列(b_1 - b_2 - b_3 - b_4 - b_5 - b_6)的核酸序列的

通用/简并碱基或规范碱基。换言之,尽管碱基 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 可能对任何给定的靶序列是特异性的,但(N)碱基可以是通用/简并碱基,或者由对通过碱基 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 决定的靶并非特异性的四种规范碱基中的任何碱基组成。例如,如果待查询的靶序列是CAGGCATA,则靶结合结构域的碱基 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 将是TCCGTA,而靶结合结构域的每个(N)碱基可以独立地是A、C、T或G,使得所得到的靶结合结构域可以具有序列ATCCGTAG、TTCCGTAC、GTCCGTAG或其它16种可能的迭代中的任何种类。可替代地,两个(N)碱基可以在6个LNA之前。还可替代地,两个(N)碱基可以在6个LNA之后。

[0134] 表1

[0135]

靶结合结构域碱基			B1	B2	B3	B4	B5	B6		
“6聚体”			b	b	b	b	b	b		
“8聚体”		b	b	b	b	b	b	b	b	
“10聚体”	b	b	b	b	b	b	b	b	b	b
“天然 I”	N	N	b	b	b	b	b	b	N	N
“天然 II”		N	b	b	b	b	b	b	N	
“2 LNA”		N	b	b	+	+	b	b	N	
		N	b	+	b	b	+	b	N	
		N	+	b	b	b	b	+	N	
“4 LNA”		N	+	+	b	b	+	+	N	
		N	+	b	+	+	b	+	N	
		N	b	+	+	+	+	b	N	
“6 LNA”			+	+	+	+	+	+		
		N	+	+	+	+	+	+	N	
“具有 LNA 的 8 聚体”		N	b/+	b/+	b/+	b/+	b/+	b/+	N	
“MGB”		Q	b	b	b	b	b	b		
	Q	b	b	b	b	b	b	b	b	

[0136] b = 天然碱基;+ = 修饰的核苷酸或核苷酸类似物(例如LNA、2-0'-甲基修饰的碱基、6-氨基-1-(4-羟基-5-羟基甲基-四氢呋喃-2-基)-1,5-二氢-吡啶并[3,4-d]嘧啶-4-酮(PPG));N = 天然、通用或简并碱基;Q是小沟结合剂(例如扭转嵌入核酸(Twisted Intercalating Nucleic Acid)、MGB-BP3、Brostallicin)

[0137] 表1还描述了包含10个天然的、靶特异性碱基的“10聚体”靶结合结构域。表1还描述了包含8个天然的、靶特异性碱基的“8聚体”靶结合结构域。

[0138] 表1进一步描述了在位置b1至b6处包含6个天然碱基的“天然I”靶结合结构域。这6个天然碱基在任一侧上侧翼为2个(N)碱基。可替代地,所有四个(N)碱基都可以在6个天然碱基之前。还可替代地,所有四个(N)碱基都可以在6个天然碱基之后。四个(N)碱基中的任何数目(即1、2、3或4个)都可以在6个天然碱基之前,而剩余的(N)碱基在6个天然碱基之后。

[0139] 表1进一步描述了在位置b1至b6处包含6个天然碱基的“天然II”靶结合结构域。这6个天然碱基在任一侧上侧翼为(N)碱基。可替代地,两个(N)碱基均可以在6个天然碱基之

前。还可替代地,两(N)碱基均可以在6个天然碱基之后。通常,天然II结合结构域的(N)碱基是简并碱基。

[0140] 表1还描述了“2 LNA”靶结合结构域,其在靶结合结构域的位置b1至b6处包含2个LNA和4个天然碱基的组合。2个LNA和4个天然碱基可以以任何次序存在。例如,位置b3和b4可以是LNA,而位置b1、b2、b5和b6是天然碱基。碱基b1至b6在任一侧上侧翼为(N)碱基。可替代地,碱基b1至b6可以之前为两个(N)碱基。还可替代地,碱基b1至b6可以随后为两个(N)碱基。

[0141] 表1进一步描述了“4 LNA”靶结合结构域,其在靶结合结构域的位置b1至b6处包含4个LNA和2个天然碱基的组合。4个LNA和2个天然碱基可以以任何次序存在。例如,位置b2至b5可以是LNA,而位置b1和b6是天然碱基。碱基b1至b6在任一侧上侧翼为(N)碱基。可替代地,碱基b1至b6可以之前为两个(N)碱基。还可替代地,碱基b1至b6可以随后为两个(N)碱基。

[0142] 表1进一步描述了“6 LNA”靶结合结构域,其在靶结合结构域的位置b1至b6处包含6个LNA。碱基b1至b6可以在任一侧上侧翼为(N)碱基。

[0143] 表1进一步描述了“具有LNA的8聚体”靶结合结构域,其在靶结合结构域的位置b1至b6的任何处个别地包含天然碱基或LNA。碱基b1至b6可以在任一侧上侧翼为(N)碱基。

[0144] 靶结合结构域也可以包含小沟结合剂部分。小沟结合剂部分是寡核苷酸的化学修饰,其添加了可以结合寡核苷酸与其杂交的靶核苷酸的小沟的化学部分。不受理论的束缚,小沟结合剂部分的包括增加了靶结合结构域对于靶核酸的亲合力,增加了靶结合结构域-靶核酸双链体的解链温度。较高的结合亲合力可以允许使用较小的靶结合结构域。

[0145] 靶结合结构域也可以包含一种或多种扭转嵌入核酸(TINA)。TINA是稳定来自双链寡核苷酸和三链体形成寡核苷酸的Hoogsteen三链体DNA形成的核酸分子。TINA可以用于稳定双链寡核苷酸,从而改善寡核苷酸探针对靶核酸的特异性和敏感性。

[0146] 靶结合结构域还可以包括包含2'-O-甲基修饰的碱基的核酸分子。2'-O-甲基修饰的碱基是RNA的核苷修饰,其中将甲基加入核糖的2'羟基,以产生2'甲氧基。2'-O-甲基修饰的碱基提供了针对通过核酸酶的碱基水解和消化的极佳保护。不受理论的束缚,2'-O-甲基修饰的碱基的添加也增加了核酸双链体的解链温度。

[0147] 靶结合结构域也可以包含共价连接的芪修饰。芪修饰可以增加核酸双链体的稳定性。

[0148] 本公开内容的测序探针包含合成主链。靶结合结构域,在本文中也描述为测序结构域,与条形码结构域是可操作地连接的。靶结合结构域和条形码结构域可以作为一条合成主链的部分共价附着。靶结合结构域和条形码结构域可以经由接头(例如核酸接头、化学接头)附着。合成主链可以包含任何材料,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。合成主链可以包含单链DNA分子。主链可以包含六个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折纸砖(origami tile)制备(Jungmann等人,“Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT”,*Nature Methods*,第11卷,第3期,2014)。

[0149] 本公开内容的测序探针可以包含部分双链的合成主链。测序探针可以包含在靶结合结构域和条形码结构域之间的单链DNA合成主链和双链DNA间隔区。双链DNA间隔区可以包含至少一个修饰的核苷酸或核酸类似物。可用于双链DNA间隔区中的通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含双链DNA间隔区的每种核酸可以独立地是L-DNA。在一些方面,双链DNA间隔区可以包含L-DNA。双链DNA间隔区可以由L-DNA组成。双链DNA间隔区可以基本上由L-DNA组成。

[0150] 双链DNA间隔区可以包含长度约1个核苷酸至约100个核苷酸。双链DNA间隔区可以包含长度约25个核苷酸。

[0151] 合成主链可以包含L-DNA。合成主链可以由L-DNA组成。合成主链可以基本上由L-DNA组成。单链DNA合成主链可以包含长度约10个核苷酸至约100个核苷酸。单链DNA合成主链可以包含长度约52个核苷酸。单链DNA合成主链可以包含长度约27个核苷酸。

[0152] 条形码结构域可以包含L-DNA。条形码结构域可以由L-DNA组成。条形码结构域可以基本上由L-DNA组成。条形码结构域可以包含约27个核苷酸、或约52个核苷酸、或约99个核苷酸、或约74个核苷酸。条形码结构域可以是长度约27个核苷酸、或约52个核苷酸、或约99个核苷酸、或约74个核苷酸。

[0153] 测序探针可以包含在靶结合结构域和条形码结构域之间的单链DNA合成主链和基于聚合物的间隔区,其具有与双链DNA相似的机械特性。通常的基于聚合物的间隔区包括聚乙二醇(PEG)型聚合物。

[0154] 双链DNA间隔区可以是长度约1个核苷酸至约100个核苷酸;长度约2个核苷酸至约50个核苷酸;长度约20个核苷酸至约40个核苷酸。优选地,双链DNA间隔区是长度约36个核苷酸。

[0155] 在图2的左图中示出了本公开内容的一种测序探针,称为“标准探针”。图2的标准探针包含共价附着至靶结合结构域的条形码结构域,使得靶结合结构域和条形码结构域存在于同一单链寡核苷酸内。在图2,左图中,单链寡核苷酸与茎寡核苷酸结合,以产生称为茎的长36个核苷酸的双链间隔区域。使用这种结构,探针池中的每种测序探针都可以与相同的茎序列杂交。

[0156] 在替代方面,包含条形码结构域以及与标准探针的茎寡核苷酸结合的区域中的每种核酸,可以是规范碱基或修饰的核苷酸或核酸类似物。可用于条形码结构域以及与标准探针的茎寡核苷酸结合的区域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含条形码结构域以及与标准探针的茎寡核苷酸结合的区域中的每种核酸,可以独立地是L-DNA。例如,条形码结构域以及与标准探针的茎寡核苷酸结合的区域可以完全由L-DNA组成。在其它实例中,条形码结构域以及与标准探针的茎寡核苷酸结合的区域可以由L-DNA区段组成,所述L-DNA区段被无碱基的单链核酸区段、或下文进一步描述的与双链DNA具有相似机械特性的聚合物(例如PEG)区段分开。

[0157] 在图2的中图中示出了本公开内容的另一种测序探针,称为“3部分探针”。图2的3部分探针包含条形码结构域,其经由接头附着至靶结合结构域。在该实例中,接头是单链茎寡核苷酸,其与含有靶结合结构域的单链寡核苷酸和含有条形码结构域的单链寡核苷酸杂交,产生了桥接条形码结构域(18个核苷酸)和靶结合结构域(18个核苷酸)的长36个核苷酸的双链间隔区域。使用该实例性探针配置,为了防止条形码结构域的交流,可以这样设计每

个条形码,使得它与独特的茎序列杂交。此外,在将不同的测序探针合并在一起之前,每个条形码结构域也可以与其相应的茎寡核苷酸杂交。

[0158] 在替代方面,包含单链茎寡核苷酸的每种核酸可以是规范碱基或修饰的核苷酸或核酸类似物。可用于单链茎寡核苷酸中的通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含单链茎寡核苷酸的每种核酸可以独立地是L-DNA。

[0159] 在替代方面,包含在单链茎寡核苷酸与之杂交的条形码结构域上的区域的每种核酸,可以是规范碱基或修饰的核苷酸或核酸类似物。可用于单链茎寡核苷酸中的通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含在单链茎寡核苷酸与之杂交的条形码结构域上的区域的每种核酸,可以独立地是L-DNA。

[0160] 在替代方面,包含在单链寡核苷酸上含有单链茎寡核苷酸与之杂交的靶结合结构域的区域上的每种核酸,可以是规范碱基或修饰的核苷酸或核酸类似物。可用于单链茎寡核苷酸中的通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含在单链寡核苷酸上含有单链茎寡核苷酸与之杂交的靶结合结构域的区域上的每种核酸,可以独立地是L-DNA。

[0161] 在图2的右图中示出了本公开内容的另一种测序探针,称为“1-部分接头探针”。图2的1-部分接头探针包含条形码结构域,其经由接头附着至靶结合结构域。在该实例中,接头是PEG分子。可替代地,接头可以是反式芪。还可替代地,接头可以是与双链DNA具有相似机械特性的任何聚合物。通常的基于聚合物的间隔区包括聚乙二醇(PEG)型聚合物。

[0162] 本公开内容的测序探针可以包含约60个核苷酸。本公开内容的测序探针可以包含约107个核苷酸。本公开内容的测序探针可以是长度约60个核苷酸,或长度约107个核苷酸。构成测序探针的核苷酸可以各自个别地是规范碱基、修饰的核苷酸或核酸类似物,包括L-DNA和D-DNA。

[0163] 条形码结构域包括多个附着位置,例如一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个附着位置。附着位置的数目可以少于、等于或多于靶结合结构域中的核苷酸数目。靶结合结构域可以包含比主链结构域中的附着位置数目更多的核苷酸,例如一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个核苷酸。靶结合结构域可以包含八个核苷酸,并且条形码结构域包含三个附着位置。靶结合结构域可以包含十个核苷酸,并且条形码结构域包含三个附着位置。

[0164] 条形码结构域的长度不受限制,只要对于至少三个附着位置存在足够的空间,如下所述。术语“附着位置”、“位置”和“位点”在本文中可互换使用。术语“条形码结构域”和“报告结构域”在本文中可互换使用。

[0165] 条形码结构域中的每个附着位置对应于靶结合结构域中的两个核苷酸(二核苷酸),并且因此对应于靶核酸中与靶结合结构域中的二核苷酸杂交的互补二核苷酸。作为非限制性实例,条形码结构域中的第一附着位置对应于靶结合结构域中的第一核苷酸和第二核苷酸(例如图1,其中R1是条形码结构域中的第一附着位置,并且R1对应于靶结合结构域中的二核苷酸b1和b2 - 其依次又鉴定了靶核酸的二核苷酸1和2);条形码结构域中的第二附着位置对应于靶结合结构域中的第三核苷酸和第四核苷酸(例如图1,其中R2是条形码结构域中的第二附着位置,并且R2对应于靶结合结构域中的二核苷酸b3和b4 - 其依次又鉴定了靶核酸的二核苷酸3和4);并且条形码结构域中的第三附着位置对应于靶结合结构域

中的第五核苷酸和第六核苷酸(例如图1,其中R3是条形码结构域中的第三附着位置,并且R3对应于靶结合结构域中的二核苷酸b5和b6 - 其依次又鉴定了靶核酸的二核苷酸5和6)。在一个进一步的非限制性实例中,条形码结构域中的第一附着位置、条形码结构域中的第二附着位置和条形码结构域中的第三附着位置,共同对应于靶结合结构域中的第一核苷酸至第六核苷酸(例如图1,其中靶结合结构域中的核苷酸b1至b6 - 其依次又鉴定了靶核酸的六个核苷酸)。

[0166] 条形码结构域中的每个附着位置包含至少一个附着区域,例如一个至50个或更多个附着区域。条形码结构域中的某些位置可以比其它位置具有更多的附着区域(例如,第一附着位置可以具有三个附着区域,而第二附着位置可以具有两个附着位置);可替代地,条形码结构域中的每个位置具有相同数目的附着区域。条形码结构域中的每个附着位置可以包含一个附着区域。条形码结构域中的每个附着位置可以包含多于一个附着区域。条形码结构域中的至少三个附着位置中的至少一个,可以包含与条形码结构域中的其它两个附着位置不同数目的附着区域。在一些方面,条形码结构域中的每个附着位置可以包含一个附着区域。

[0167] 每个附着区域包含能够由互补核酸分子(例如,DNA或RNA)可逆地结合的核酸序列的至少一个(即,一至五十,例如十至三十)拷贝。在单个附着位置处的附着区域的核酸序列可以是等同的;因此,结合那些附着区域的互补核酸分子是等同的。可替代地,在一个位置处的附着区域的核酸序列是不等同的;因此,结合那些附着区域的互补核酸分子是不等同的。

[0168] 包含条形码结构域中的每个附着区域的核酸序列可以是长度约6个核苷酸至约20个核苷酸。包含条形码结构域中的每个附着区域的核酸序列可以是长度约12个核苷酸。包含条形码结构域中的每个附着区域的核酸序列可以是长度约16个核苷酸。包含条形码结构域中的每个附着区域的核酸序列可以是长度约14个核苷酸。包含条形码结构域中的每个附着区域的核酸序列可以是长度约8个核苷酸。包含条形码结构域中的每个附着区域的核酸序列可以是长度约9个核苷酸。

[0169] 附着位置、附着区域或附着区域的至少一个核酸序列,可以包含至少一个超T碱基(super T base)(5-羟基丁炔-2'-脱氧尿苷)。附着位置、附着区域或附着区域的至少一个核酸序列,可以包含至少一个3'末端超T碱基(5-羟基丁炔-2'-脱氧尿苷)。附着位置、附着区域或附着区域的至少一个核酸序列,可以包含至少一个5'末端超T碱基(5-羟基丁炔-2'-脱氧尿苷)。

[0170] 包含条形码结构域中的每个附着区域的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。条形码结构域的附着区域中的至少一个、至少两个、至少三个、至少四个、至少五个或至少六个核苷酸可以是修饰的核苷酸或核苷酸类似物。条形码结构域中的修饰的核苷酸或核苷酸类似物与规范碱基的通常比率为1:2至1:8。可用于条形码结构域的附着区域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。例如,修饰的核苷酸或核苷酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善报告物与条形码结构域中的适当附着区域的结合效率和准确率,同时使其它地方(包括与靶)的结合降到最低。

[0171] 条形码结构域内的一个或多个附着区域可以包含L-DNA。L-DNA是天然存在的右旋(right-turning) D-DNA的左旋和镜像版本。L-DNA更稳定并且对酶促消化有抵抗力。由于

L-DNA无法与D-DNA杂交,因此L-DNA可以改善报告物与条形码结构域中的适当附着区域的结合效率和结合准确率,并且防止报告物与测序探针上的其它地方的结合。在一些方面,附着位置的至少一个核酸序列的每个核苷酸可以是L-DNA。

[0172] 包含条形码结构域中的每个附着区域的每种核酸,可以独立地包含腺嘌呤、胞嘧啶、鸟嘌呤或胸腺嘧啶碱基。可替代地,包含条形码结构域中的每个附着区域的每种核酸,可以独立地包含腺嘌呤、鸟嘌呤或胸腺嘧啶碱基。

[0173] 包含条形码结构域中的每个附着区域的每个核酸序列,可以包含至少一个腺嘌呤核苷酸,至少一个胸腺嘧啶核苷酸,至少一个胞嘧啶核苷酸或其任何组合和3'末端鸟苷核苷酸。包含条形码结构域中的每个附着区域的每个核酸序列,可以由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个胞嘧啶核苷酸或其任何组合和3'末端鸟苷核苷酸组成。包含条形码结构域中的每个附着区域的每个核酸序列,可以基本上由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个胞嘧啶核苷酸或其任何组合和3'末端鸟苷核苷酸组成。

[0174] 包含条形码结构域中的每个附着区域的每个核酸序列,可以包含至少一个腺嘌呤核苷酸,至少一个胸腺嘧啶核苷酸,至少一个胞嘧啶核苷酸或其任何组合和5'末端鸟苷核苷酸。包含条形码结构域中的每个附着区域的每个核酸序列,可以由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个胞嘧啶核苷酸或其任何组合和5'末端鸟苷核苷酸组成。包含条形码结构域中的每个附着区域的每个核酸序列,可以基本上由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个胞嘧啶核苷酸或其任何组合和5'末端鸟苷核苷酸组成。

[0175] 在一些方面,条形码结构域的至少一个附着位置中的至少一个附着区域,可以包含3'末端鸟苷核苷酸。在一些方面,条形码结构域的至少两个附着位置中的至少一个附着区域,可以包含3'末端鸟苷核苷酸。在一些方面,条形码结构域的至少三个附着位置中的至少一个附着区域,可以包含3'末端鸟苷核苷酸。3'末端鸟苷核苷酸可以是L-DNA。

[0176] 在一些方面,条形码结构域的至少一个附着位置中的至少一个附着区域,可以包含3'末端鸟苷核苷酸。在一些方面,条形码结构域的至少两个附着位置中的至少一个附着区域,可以包含3'末端鸟苷核苷酸。在一些方面,条形码结构域的至少三个附着位置中的至少一个附着区域,可以包含5'末端鸟苷核苷酸。3'末端鸟苷核苷酸可以是L-DNA,例如L-脱氧鸟苷(L-dG)。末端L-dG核苷酸减轻了附着区域和/或附着位置之间的交叉连接杂交,并且通过提供碱基堆积相互作用来维持稳定性。

[0177] 一个或多个附着区域可以与多核苷酸主链整合;即,主链是单个多核苷酸,并且附着区域是单个多核苷酸序列的部分。一个或多个附着区域可以连接至合成主链中的修饰的单体(例如,修饰的核苷酸),使得附着区域从合成主链分支出来。附着位置可以包含多于一个附着区域,其中一些附着区域从合成主链分支出来,并且一些附着区域对于合成主链是整合的。至少一个附着位置中的至少一个附着区域可以与合成主链整合。至少三个附着位置各自中的每个附着区域可以与合成主链整合。至少一个附着位置中的至少一个附着区域可以从合成主链分支出来。至少三个附着位置各自中的每个附着区域可以从合成主链分支出来。

[0178] 条形码结构域内的每个附着位置对应于十六种二核苷酸之一,即,腺嘌呤-腺嘌呤

呤、腺嘌呤-胸腺嘧啶/尿嘧啶、腺嘌呤-胞嘧啶、腺嘌呤-鸟嘌呤、胸腺嘧啶/尿嘧啶-腺嘌呤、胸腺嘧啶/尿嘧啶-胸腺嘧啶/尿嘧啶、胸腺嘧啶/尿嘧啶-胞嘧啶、胸腺嘧啶/尿嘧啶-鸟嘌呤、胞嘧啶-腺嘌呤、胞嘧啶-胸腺嘧啶/尿嘧啶、胞嘧啶-胞嘧啶、胞嘧啶-鸟嘌呤、鸟嘌呤-腺嘌呤、鸟嘌呤-胸腺嘧啶/尿嘧啶、鸟嘌呤-胞嘧啶或鸟嘌呤-鸟嘌呤。因此,位于条形码结构域的单个附着位置中的一个或多个附着区域对应于十六种二核苷酸之一,并且包含对该附着区域对应于其的二核苷酸特异性的核酸序列。即使在条形码结构域内的这些位置对应于相同的二核苷酸,位于条形码结构域的不同附着位置中的附着区域也含有独特的核酸序列。例如,鉴于本公开内容的测序探针含有具有编码序列A-G-A-G-A-C的六聚体的靶结合结构域,该测序探针的条形码结构域将含有三个位置,其中第一附着位置对应于腺嘌呤-鸟嘌呤二核苷酸,第二附着位置对应于腺嘌呤-鸟嘌呤二核苷酸,而第三附着位置对应于腺嘌呤-胞嘧啶二核苷酸。即使附着位置1和附着位置2两者均对应于二核苷酸腺嘌呤-鸟嘌呤,位于该实例探针的位置一中的附着区域也包含相对于位于位置二中的附着区域的核酸序列独特的核酸序列。这样设计且测试特异性附着位置的序列,使得特定附着位置的互补核酸不与不同的附着位置相互作用。另外,互补核酸的核苷酸序列不受限制;优选地,它与已知的核苷酸序列缺乏基本同源性(例如50%至99.9%);这限制了互补核酸和靶核酸的不期望的杂交。

[0179] 图1显示了包含示例性条形码结构域的本公开内容的一种示例性测序探针的图示。图1中描绘的示例性条形码结构域包含三个附着位置 R_1 、 R_2 和 R_3 。每个附着位置对应于靶结合结构域的6聚体序列(b_1 到 b_6)内存在的特异性二核苷酸。在该实例中, R_1 对应于位置 b_1 和 b_2 , R_2 对应于位置 b_3 和 b_4 ,并且 R_3 对应于位置 b_5 和 b_6 。因此,每个位置解码靶结合结构域的6聚体序列中存在的特定二核苷酸,允许鉴定存在于每个特定二核苷酸中的特定两个碱基(A、C、G或T)。

[0180] 在图1中描绘的示例性条形码结构域中,每个附着位置包含与合成主链整合的单个附着区域。三个附着位置的每个附着区域含有特异性核苷酸序列,其对应于由每个附着位置编码的特定二核苷酸。例如,附着位置 R_1 包含具有对应于二核苷酸 b_1 - b_2 的身份的特异性序列的附着区域。

[0181] 条形码结构域可以进一步包含一个或多个结合区域。条形码结构域可以包含与至少一个附着位置相邻或侧接的至少一个单链核酸序列。条形码结构域可以包含与至少两个附着位置相邻或侧接的至少两个单链核酸序列。条形码结构域可以包含与至少三个附着位置相邻或侧接的至少三个单链核酸序列。这些侧翼部分称为“立足点(Toe-Hold)”,其可以通过对于单链寡核苷酸提供另外的结合位点,来加速与立足点相邻杂交的寡核苷酸的交换速率(例如“立足点”探针;参见例如,Seeling等人,“Catalyzed Relaxation of a Metastable DNA Fuel”;*J. Am. Chem. Soc.* 2006,128(37),第12211-12220页)。

[0182] 条形码结构域内的至少一个附着区域可以在至少一侧上侧翼为双链核酸序列。条形码结构域内的至少两个附着区域可以在至少一侧上侧翼为双链核酸序列。条形码结构域内的至少三个附着区域可以在至少一侧上侧翼为双链核酸序列。

[0183] 条形码结构域内的任何附着区域可以通过称为“口袋寡核苷酸”的双链核酸序列与任何相邻的附着位置分开。图28显示了具有包含三个附着位置的条形码结构域的测序探针的实例。附着位置1通过口袋寡核苷酸与相邻的附着位置2分开。附着位置2进一步通过另

一个口袋寡核苷酸与相邻的附着位置3分开。

[0184] 包含口袋寡核苷酸的每种核酸可以是规范碱基或修饰的核苷酸或核酸类似物。可用于口袋寡核苷酸中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含口袋寡核苷酸的每种核酸可以独立地是L-DNA。口袋寡聚物可以包含至少一个超T碱基(5-羟基丁炔-2'-脱氧尿苷)。口袋寡核苷酸可以是长度约25个核苷酸。

[0185] 在一些方面,条形码结构域中的至少一个、至少两个或至少三个附着位置可以与至少一种侧翼双链多核苷酸相邻。至少一种侧翼双链多核苷酸可以包含至少一种修饰的核苷酸或核酸类似物。至少一种侧翼双链多核苷酸可以包含L-DNA。至少一种侧翼双链多核苷酸可以包含至少一个超T碱基(5-羟基丁炔-2'-脱氧尿苷)。至少一种侧翼双链多核苷酸可以是长度约25个核苷酸。

[0186] 条形码结构域内的至少一个附着区域,可以在至少一侧上侧翼为与双链DNA具有相似机械特性的任何聚合物。通常的基于聚合物的间隔区包括聚乙二醇(PEG)型聚合物。条形码结构域内的至少两个附着区域,可以在至少一侧上侧翼为与双链DNA具有相似机械特性的任何聚合物。条形码结构域内的至少三个附着区域,可以在至少一侧上侧翼为与双链DNA具有相似机械特性的任何聚合物。

[0187] 条形码结构域内的任何附着区域,可以通过与双链DNA具有相似机械特性的任何聚合物与任何相邻的附着位置分开。通常的基于聚合物的间隔区包括聚乙二醇(PEG)型聚合物。图29显示了具有包含三个附着位置的条形码结构域的测序探针的实例。附着位置1通过PEG-接头与相邻的附着位置2分开。附着位置2进一步通过另一个PEG-接头与相邻的附着位置3分开。

[0188] 条形码结构域内的至少一个附着区域可以在至少一侧上侧翼为无碱基的单链核酸分子。无碱基的核酸分子是既没有嘌呤碱基也没有嘧啶碱基的核酸分子。条形码结构域内的至少两个附着区域可以在至少一侧上侧翼为无碱基的单链核酸分子。条形码结构域内的至少三个附着区域可以在至少一侧上侧翼为无碱基的单链核酸分子。

[0189] 条形码结构域内的任何附着区域,可以通过无碱基的单链核酸分子与任何相邻的附着位置分开。图30显示了具有包含三个附着位置的条形码结构域的测序探针的实例。附着位置1通过无碱基的单链核酸分子与相邻的附着位置2分开。附着位置2进一步通过另一个无碱基的单链核酸分子与相邻的附着位置3分开。

[0190] 条形码结构域内的任何附着区域,可以通过3'末端鸟苷核苷酸与任何相邻的附着位置分开。在一些方面,条形码结构域的至少两个附着位置中的至少一个附着区域,可以包含3'末端鸟苷核苷酸。图53显示了具有包含三个附着位置的条形码结构域的测序探针的实例,每一个附着位置通过末端L-G核苷酸分开。附着位置1通过L-G核苷酸与相邻的附着位置2分开。附着位置2进一步通过另一个L-G核苷酸与相邻的附着位置3分开。附着位置3在3'末端上用L-G核苷酸终止。

[0191] 本公开内容的测序探针可以具有约20纳米至约50纳米的总体长度(包括靶结合结构域、条形码结构域和任何任选结构域)。测序探针的主链可以是包含约120个核苷酸、约60个核苷酸、约52个核苷酸或约27个核苷酸的多核苷酸分子。

[0192] 测序探针可以包含可切割接头修饰。可切割接头修饰可以包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、至少十个或

任何数目的可切割部分。可以利用本领域技术人员已知的任何可切割接头修饰或可切割部分。可切割接头修饰和可切割部分的非限制性实例包括但不限于UV光可切割接头、还原剂可切割接头和酶促可切割接头。酶促可切割接头的实例是插入脱氧尿嘧啶,用于通过USER™酶的切割。可切割接头修饰可以位于沿着测序探针的长度的任何位置,包括但不限于靶结合结构域和条形码结构域之间的区域。图7的右图描绘了示例性的可切割接头修饰,其可以掺入本公开内容的探针内。

[0193] 报告探针

[0194] 与本公开内容的测序探针的条形码结构域的至少一个附着位置内的至少一个附着区域内的互补核酸序列结合(例如杂交),并且包含(直接或间接)可检测标记的核酸分子,在本文中称为“报告探针”或“报告探针复合物”,这些术语在本文中可互换使用。报告探针可以是DNA、RNA或PNA。优选地,报告探针是DNA。

[0195] 报告探针可以包含至少两个结构域:能够结合至少一个第一互补核酸分子的第一结构域,以及能够结合第一可检测标记和至少第二可检测标记的第二结构域。图3显示了本公开内容的示例性报告探针的示意图,所述报告探针结合至示例性测序探针的条形码结构域的第一附着位置。在图3中,报告探针的第一结构域(以带阴影的栗色显示)结合条形码结构域的附着位置 R_1 内的互补核酸序列,而报告探针的第二结构域(以灰色显示)与两种可检测标记(一种绿色标记、一种红色标记)结合。

[0196] 可替代地,报告探针可以包含至少两个结构域:能够结合至少一个第一互补核酸分子的第一结构域,以及能够结合至少一个第二互补核酸分子的第二结构域。至少一个第一互补核酸分子和至少一个第二互补核酸分子可以是不同的(具有不同的核酸序列)。

[0197] “一级核酸分子”是包含至少两个结构域的报告探针:能够与测序探针的条形码结构域的至少一个附着位置内的至少一个附着区域内的互补核酸序列结合(例如杂交)的第一结构域,以及与至少一种另外的互补核酸结合(例如杂交)的第二结构域。一级核酸分子可以直接结合测序探针的条形码结构域的至少一个附着位置内的至少一个附着区域内的互补核酸序列。一级核酸分子可以经由核酸接头间接结合测序探针的条形码结构域的至少一个附着位置内的至少一个附着区域内的互补核酸序列。这种核酸接头被称为“连接子寡核苷酸”。

[0198] 连接子寡核苷酸可以包含至少两个结构域:能够与条形码结构域的至少一个附着位置内的至少一个附着区域内的至少一个第一互补核酸序列结合(例如杂交)的第一结构域,以及能够与一级核酸分子的第一结构域结合(例如杂交)的第二结构域。图31显示了经由连接子寡核苷酸与报告探针结合的测序探针。

[0199] 包含连接子寡核苷酸的第一结构域或第二结构域的每种核酸,可以是规范碱基或修饰的核苷酸或核酸类似物。可用于连接子寡核苷酸的第一结构域或第二结构域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。例如,修饰的核苷酸或核苷酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善连接子寡核苷酸的第一结构域与测序探针的条形码结构域的至少一个附着位置内的至少一个附着区域内的适当互补核酸序列的结合效率和准确率,同时使其它地方(包括与靶)的结合降到最低。例如,修饰的核苷酸或核苷酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善连接子寡核苷酸的第二结构域与报告探针的适当第一结构域的结合效率和准确率,同时使其它地方(包括与靶)的结合降到最低。可替

代地,包含连接子寡核苷酸的第一结构域或第二结构域的每种核酸,可以独立地是L-DNA。在连接子寡核苷酸的一个实例中,第一结构域包含D-DNA,并且第二结构域包含L-DNA。在连接子寡核苷酸的另一个实例中,第一结构域包含D-DNA,并且第二结构域包含异鸟嘌呤和/或异胞嘧啶。

[0200] 连接子寡核苷酸的第一结构域可以是长度约8至约16个核苷酸。优选地,连接子寡核苷酸的第一结构域是长度14个核苷酸。连接子寡核苷酸的第二结构域可以是长度约4-12个核苷酸。优选地,连接子寡核苷酸的第二结构域可以是长度约8个核苷酸。

[0201] 在包含连接子寡核苷酸的方面,附着区域可以被称为部分双链的。部分双链的附着区域可以包含双链区域和单链。部分双链的附着区域的单链区域,可以包含与至少一个互补核酸序列结合(例如杂交)的至少一个核酸序列。与部分双链的附着区域的单链区域结合(例如杂交)的至少一个互补核酸序列,可以是一级核酸分子。

[0202] 包含部分双链的附着区域的双链区域的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。部分双链的附着区域的双链区域中的至少一个、两个、至少三个、至少四个、至少五个、至少六个、至少七个或至少八个核苷酸可以是修饰的核苷酸或核酸类似物。条形码结构域中的修饰的核苷酸或核酸类似物与规范碱基的通常比率为1:2至1:8。可用于一级核酸分子的第一结构域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。可替代地,包含部分双链的附着区域的双链区域的每种核酸,可以独立地是L-DNA。

[0203] 包含部分双链的附着区域的单链区域的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。部分双链的附着区域的单链区域中的至少一个、两个、至少三个、至少四个、至少五个、至少六个、至少七个或至少八个核苷酸可以是修饰的核苷酸或核酸类似物。条形码结构域中的修饰的核苷酸或核酸类似物与规范碱基的通常比率为1:2至1:8。可用于部分双链的附着区域的单链区域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。例如,修饰的核苷酸或核酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善部分双链的附着区域的单链区域与一级核酸分子的适当互补核酸序列的结合效率和准确率,同时使其它地方(包括与靶)的结合降到最低。可替代地,包含一级核酸分子的第一结构域的每种核酸,可以独立地是L-DNA。

[0204] 一级核酸分子可以包含可切割接头。可切割接头可以位于第一结构域和第二结构域之间。优选地,可切割接头是光可切割的。可切割接头可以包含至少一个或至少两个可切割部分。该至少一个或至少两个可切割部分可以是光可切割的。

[0205] 一级核酸分子的第一结构域可以是长度约6至16个核苷酸。优选地,一级核酸分子的第一结构域是长度约8个核苷酸。

[0206] 包含一级核酸分子的第一结构域的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。一级核酸分子的第一结构域中的至少一个、两个、至少三个、至少四个、至少五个、至少六个、至少七个或至少八个核苷酸可以是修饰的核苷酸或核酸类似物。条形码结构域中的修饰的核苷酸或核酸类似物与规范碱基的通常比率为1:2至1:8。可用于一级核酸分子的第一结构域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。例如,修饰的核苷酸或核酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善一级核酸分子的第一结构域与测序探针的条形码结构域的至少一个附着位置内的至少一个附着区

域内的适当互补核酸序列的结合效率和准确率,同时使其它地方(包括与靶)的结合降到最低。可替代地,包含一级核酸分子的第一结构域的每种核酸,可以独立地是L-DNA。

[0207] 在一些方面,一级核酸分子的第一结构域可以完全由L-DNA组成,而一级核酸分子的第二结构域可以完全由D-DNA组成。

[0208] 在一些方面,一级核酸分子的第一结构域可以包含3'末端胞嘧啶核苷酸。在一些方面,一级核酸分子的第一结构域可以包含3'末端胞嘧啶核苷酸,其中所述3'末端胞嘧啶核苷酸是L-DNA。

[0209] 在一些方面,一级核酸分子的第一结构域可以包含5'末端胞嘧啶核苷酸。在一些方面,一级核酸分子的第一结构域可以包含5'末端胞嘧啶核苷酸,其中所述5'末端胞嘧啶核苷酸是L-DNA。

[0210] 在一些方面,一级核酸分子的第一结构域可以包含至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个鸟嘌呤核苷酸或其任何组合和3'末端胞嘧啶核苷酸。在一些方面,一级核酸分子的第一结构域可以由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个鸟嘌呤核苷酸或其任何组合和3'末端胞嘧啶核苷酸组成。在一些方面,一级核酸分子的第一结构域可以基本上由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个鸟嘌呤核苷酸或其任何组合和3'末端胞嘧啶核苷酸组成。

[0211] 在一些方面,一级核酸分子的第一结构域可以包含至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个鸟嘌呤核苷酸或其任何组合和5'末端胞嘧啶核苷酸。在一些方面,一级核酸分子的第一结构域可以由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个鸟嘌呤核苷酸或其任何组合和5'末端胞嘧啶核苷酸组成。在一些方面,一级核酸分子的第一结构域可以基本上由至少一个腺嘌呤核苷酸、至少一个胸腺嘧啶核苷酸、至少一个鸟嘌呤核苷酸或其任何组合和5'末端胞嘧啶核苷酸组成。

[0212] 结合一级核酸分子的至少一种另外的互补核酸在本文中被称为“二级核酸分子”。一级核酸分子可以与至少一个、至少两个、至少三个、至少四个、至少五个或更多个二级核酸分子结合(例如杂交)。优选地,一级核酸分子与四个二级核酸分子结合(例如杂交)。

[0213] 二级核酸分子可以包含至少两个结构域:能够与至少一个一级核酸分子中的至少一个互补序列结合(例如杂交)的第一结构域;以及能够与以下结合(例如杂交)的第二结构域:(a)第一可检测标记和至少第二可检测标记,(b)至少一种另外的互补核酸,或(c)其组合。在一些方面,二级核酸分子的第一结构域可以完全由L-DNA组成,而二级核酸分子的第二结构域可以完全由D-DNA组成。在一些方面,二级核酸分子的第一结构域和第二结构域两者均可以完全由D-DNA组成。

[0214] 二级核酸分子可以包含可切割接头。可切割接头可以位于第一结构域和第二结构域之间。优选地,可切割接头是光可切割的。

[0215] 包含二级核酸分子的第一结构域的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。二级核酸分子的第一结构域中的至少一个、两个、至少三个、至少四个、至少五个或至少六个核苷酸可以是修饰的核苷酸或核苷酸类似物。条形码结构域中的修饰的核苷酸或核苷酸类似物与规范碱基的通常比率为1:2至1:8。可用于二级核酸分子的第一结构域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。例如,修饰的核苷酸或核苷酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善二级核酸分子的第一结构域与一

级核酸分子的第二结构域内的适当互补核酸序列的结合效率和准确率,同时使其它地方的结合降到最低。

[0216] 结合二级核酸分子的至少一种另外的互补核酸在本文中被称为“三级核酸分子”。二级核酸分子可以与至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个或更多个三级核酸分子结合(例如杂交)。优选地,至少一个二级核酸分子与一个三级核酸分子结合(例如杂交)。

[0217] 三级核酸分子包含至少两个结构域:能够与至少一个二级核酸分子中的至少一个互补序列结合(例如杂交)的第一结构域,以及能够与第一可检测标记和至少第二可检测标记结合(例如杂交)的第二结构域。可替代地,第二结构域可以在使用例如亚磷酰胺或NHS化学的寡核苷酸合成期间,经由标记的直接或间接附着而包括第一可检测标记和至少第二可检测标记。在一些方面,三级核酸分子的第一结构域可以完全由L-DNA组成,而三级核酸分子的第二结构域可以完全由D-DNA组成。在一些方面,三级核酸分子的第一结构域和第二结构域两者均可以完全由D-DNA组成。三级核酸分子可以包含可切割接头。可切割接头可以位于第一结构域和第二结构域之间。优选地,可切割接头是光可切割的。

[0218] 包含三级核酸分子的第一结构域的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。三级核酸的第一结构域中的至少一个、两个、至少三个、至少四个、至少五个或至少六个核苷酸可以是修饰的核苷酸或核苷酸类似物。三级核酸分子的第一结构域中的修饰的核苷酸或核苷酸类似物与规范碱基的通常比率为1:2至1:8。可用于三级核酸分子的第一结构域中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。例如,修饰的核苷酸或核苷酸类似物如异鸟嘌呤和异胞嘧啶的使用,可以改善三级核酸分子的第一结构域与第二核酸分子的第二结构域内的适当互补核酸序列的结合效率和准确率,同时使其它地方的结合降到最低。

[0219] 报告探针与第一可检测标记和至少第二可检测标记结合,以产生双色组合。荧光染料的这种双重组合可以包括单一颜色的双重性,例如蓝色-蓝色。如本文使用的,术语“标记”包括能够产生可检测信号的单个部分、或者能够产生相同或基本上相同的可检测信号的多重部分。例如,标记包括单一黄色荧光染料例如ALEXA FLUOR™、或双重黄色荧光染料例如ALEXA FLUOR™ 532。

[0220] 报告探针可以与第一可检测标记和至少第二可检测标记结合,其中每种可检测标记是以下四种荧光染料之一:蓝色(B);绿色(G);黄色(Y);和红色(R)。这四种染料的使用产生了10种可能的双色组合:BB;BG;BR;BY;GG;GR;GY;RR;RY;或YY。在一些方面,本公开内容的报告探针用以下8种可能的颜色组合之一标记:BB;BG;BR;BY;GG;GR;GY;或YY,如图3中描绘的。可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。

[0221] 在包含测序探针和一级核酸分子的方面,本公开内容提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含表1中所述的任何构建体。示例性的靶结合结构域包含至少八个核苷酸并且能够与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述靶结合结构域中的至少六个核苷酸中的任何可以是修饰的核苷酸或核苷酸类似物,并且其中所述靶结合结构域中

并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何、或者通用碱基或简并碱基。示例性的条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列由至少一个互补的一级核酸分子结合,其中所述互补的一级核酸分子包含第一可检测标记和至少第二可检测标记,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中与至少三个附着位置的每个位置结合的每个互补一级核酸分子的至少第一可检测标记和至少第二可检测标记,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何碱基、或者通用碱基或简并碱基。

[0222] 在一些方面,靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少一个核苷酸,可以在靶结合结构域中鉴定靶核酸分子中的相应核苷酸的核苷酸之前。在一些方面,靶结合结构域中并不鉴定靶核酸中的相应核苷酸的至少一个核苷酸,可以在靶结合结构域中鉴定靶核酸分子中的相应核苷酸的核苷酸之后。

[0223] 在其它方面,示例性的靶结合结构域可以包含能够与靶核酸杂交的至少六个核苷酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸;其中所述靶结合结构域中的至少六个核苷酸无一或至少六个核苷酸中的任一个,可以是修饰的核苷酸或核苷酸类似物。

[0224] 在包含测序探针和一级核酸分子的方面,本公开内容还提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含至少十个核苷酸并且能够结合靶核酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少四个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列由至少一个互补的一级核酸分子结合,其中所述互补的一级核酸分子包含第一可检测标记和至少第二可检测标记,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,其中与至少三个附着位置的每个位置结合的每个互补一级核酸分子的至少第一可检测标记和至少第二可检测标记,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0225] 在包含测序探针、一级核酸分子和二级核酸分子的方面,本公开内容提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含表1中所述的任何构建体。示例性的靶结合结构域包含至少八个核苷酸并且能够与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述靶结合结构域中的至少六个核苷酸中的任何可以是修饰的核苷酸或核苷酸类似物,并且其中所述靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶

结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何、或者通用碱基或简并碱基。示例性的条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列由至少一个互补的一级核酸分子结合,其中所述互补的一级核酸分子进一步由至少一个互补的二级核酸分子结合,所述互补的二级核酸分子包含第一可检测标记和至少第二可检测标记,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中与至少三个附着位置的每个位置结合的每个互补二级核酸分子的至少第一可检测标记和至少第二可检测标记,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0226] 在其它方面,示例性的靶结合结构域可以包含能够与靶核酸杂交的至少六个核苷酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸;其中所述靶结合结构域中的至少六个核苷酸无一或至少六个核苷酸中的任一个,可以是修饰的核苷酸或核苷酸类似物。

[0227] 在包含测序探针、一级核酸分子和二级核酸分子的方面,本公开内容还提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含至少十个核苷酸并且能够结合靶核酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少四个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列由至少一个互补的一级核酸分子结合,其中所述互补的一级核酸分子进一步由至少一个互补的二级核酸分子结合,所述互补的二级核酸分子包含第一可检测标记和至少第二可检测标记,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,其中与至少三个附着位置的每个位置结合的每个互补二级核酸分子的至少第一可检测标记和至少第二可检测标记,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0228] 在包含测序探针、一级核酸分子、二级核酸分子和三级核酸分子的方面,本公开内容提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含表1中所述的任何构建体。示例性的靶结合结构域包含至少八个核苷酸并且能够与靶核酸杂交,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述靶结合结构域中的至少六个核苷酸中的任何可以是修饰的核苷酸或核苷酸类似物,并且其中所述靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何碱基、或者通用碱基或简并碱基。示例性的条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列由至少一个互补的一级核酸分子结合,其中所述互补的一级核酸分子进一步由至少一个互补的二级核酸分子结合,并且其中所述至少一个互补的二级核酸分

子进一步由至少一个互补的三级核酸分子结合,所述互补的三级核酸分子包含第一可检测标记和至少第二可检测标记,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中与至少三个附着位置的每个位置结合的每个互补三级核酸分子的至少第一可检测标记和至少第二可检测标记,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0229] 在其它方面,示例性的靶结合结构域可以包含能够与靶核酸杂交的至少六个核苷酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸;其中所述靶结合结构域中的至少六个核苷酸无一或至少六个核苷酸中的任一个,可以是修饰的核苷酸或核苷酸类似物。

[0230] 在包含测序探针、一级核酸分子、二级核酸分子和三级核酸分子的方面,本公开内容还提供了包含靶结合结构域和条形码结构域的测序探针;其中所述靶结合结构域包含至少十个核苷酸并且能够结合靶核酸,其中所述靶结合结构域中的至少六个核苷酸能够鉴定靶核酸分子中的相应(互补)核苷酸,并且其中所述靶结合结构域中的至少四个核苷酸并不鉴定靶核酸分子中的相应核苷酸;其中所述条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列由至少一个互补的一级核酸分子结合,其中所述互补的一级核酸分子进一步由至少一个互补的二级核酸分子结合,并且其中所述至少一个互补的二级核酸分子进一步由至少一个互补的三级核酸分子结合,所述互补的三级核酸分子包含第一可检测标记和至少第二可检测标记,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,其中与至少三个附着位置的每个位置结合的每个互补三级核酸分子的至少第一可检测标记和至少第二可检测标记,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0231] 本公开内容还提供了在二级核酸分子和三级核酸分子两者上均具有可检测标记的测序探针和报告探针。例如,二级核酸分子可以结合一级核酸分子,并且二级核酸分子可以包含第一可检测标记和至少第二可检测标记两者,并且还与包含第一可检测标记和至少第二可检测标记的至少一个三级分子结合。位于二级核酸分子上的第一可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。位于三级核酸分子上的第一可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。二级核酸分子上的可检测标记的发射光谱可以与三级核酸分子上的可检测标记的发射光谱相同或不同。

[0232] 图4是本公开内容的示例性报告探针的说明性示意图,所述报告探针包含示例性的一级核酸分子、二级核酸分子和三级核酸分子。在3'末端处,一级核酸包含第一结构域,其中所述第一结构域包含与测序探针条形码结构域的附着位置内的互补附着区域杂交的十二核苷酸序列。在5'末端处的是与六个二级核酸分子杂交的第二结构域。所描绘的示例性的二级核酸分子依次又包含在5'末端中与一级核酸分子杂交的第一结构域,以及在3'部分中与五个三级核酸分子杂交的结构域。

[0233] 三级核酸分子包含至少两个结构域。第一结构域能够与二级核酸分子结合。三级

核酸的第二结构域能够与第一可检测标记和至少第二可检测标记结合。通过一种或多种荧光标记的核苷酸单体直接掺入三级核酸的第二结构域的序列内,可以使三级核酸的第二结构域与第一可检测标记和至少第二可检测标记结合。通过使标记的短多核苷酸与二级核酸的第二结构域杂交,二级核酸分子的第二结构域可以由第一可检测标记和至少第二可检测标记结合。称为“标记的寡核苷酸”的这些短多核苷酸,可以通过直接掺入荧光标记的核苷酸单体、或通过本领域技术人员已知的标记核酸的其它方法进行标记。图4中描绘的示例性的三级核酸分子(其可以被视为“标记的寡核苷酸”),包含与二级核酸分子杂交的第一结构域和第二结构域,所述第二结构域在使用例如NHS化学的寡核苷酸合成期间,通过标记的间接附着进行荧光标记,或在三级核酸分子的合成期间,通过一种或多种荧光标记的核苷酸单体的掺入进行荧光标记。标记的寡核苷酸可以是DNA、RNA或PNA。

[0234] 标记的寡核苷酸可以包含在荧光部分和多核苷酸分子之间的可切割接头。优选地,可切割接头是光可切割的。可切割接头也可以是化学或酶促可切割的。

[0235] 在替代方面,二级核酸的第二结构域能够与第一可检测标记和至少第二可检测标记结合。通过一种或多种荧光标记的核苷酸单体直接掺入二级核酸的第二结构域的序列内,可以使二级核酸的第二结构域与第一可检测标记和至少第二可检测标记结合。通过使标记的短多核苷酸与二级核酸的第二结构域杂交,二级核酸分子的第二结构域可以由第一可检测标记和至少第二可检测标记结合。称为标记的寡核苷酸的这些短多核苷酸,可以通过直接掺入荧光标记的核苷酸单体、或通过本领域技术人员已知的标记核酸的其它方法进行标记。

[0236] 一级核酸分子可以包含约100、约95、约90、约85、约80或约75个核苷酸。一级核酸分子可以包含约100至约80个核苷酸。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90、约85、约80、约75或约70个核苷酸。二级核酸分子可以包含约90至约80个核苷酸。二级核酸分子可以包含约87个核苷酸。二级核酸分子可以包含约25、约20、约15或约10个核苷酸。三级核酸分子可以包含约20至约10个核苷酸。三级核酸分子可以包含约15个核苷酸。

[0237] 本公开内容的报告探针可以具有各种设计。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个或更多个)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个或更多个)三级核酸分子杂交。为了产生用特定双色组合标记的报告探针,这样设计报告探针,使得该探针包含二级核酸分子、三级核酸分子、标记的寡核苷酸,或者二级核酸分子、三级核酸分子和标记的寡核苷酸的任何组合,其用特定双色组合的每种颜色进行标记。例如,图4描绘了本公开内容的报告探针,其包含总共30种染料,其中15种染料用于颜色1,且15种染料用于颜色2。为了防止不同荧光染料之间的颜色交换或交叉杂交,与特异性标记或荧光染料结合的每个三级核酸或标记的寡核苷酸包含独特的核苷酸序列。

[0238] 在一些方面,本公开内容提供了5x5报告探针。5x5报告探针包含一级核酸,其中所述一级核酸包含12个核苷酸的第一结构域。一级核酸还包含第二结构域,其中所述第二结构域包含可以与5个二级核酸分子杂交的核苷酸序列。每个二级核酸包含这样的核苷酸序列,使得由可检测标记结合的5个三级核酸可以与每个二级核酸杂交。

[0239] 在一些方面,本公开内容提供了4x3报告探针。4x3报告探针包含一级核酸,其中所

述一级核酸包含12个核苷酸的第一结构域。一级核酸还包含第二结构域,其中所述第二结构域包含可以与4个二级核酸分子杂交的核苷酸序列。每个二级核酸包含这样的核苷酸序列,使得与可检测标记结合的3个三级核酸可以与每个二级核酸杂交。

[0240] 在一些方面,本公开内容提供了3x4报告探针。3x4报告探针包含一级核酸,其中所述一级核酸包含12个核苷酸的第一结构域。一级核酸还包含第二结构域,其中所述第二结构域包含可以与3个二级核酸分子杂交的核苷酸序列。每个二级核酸包含这样的核苷酸序列,使得与可检测标记结合的4个三级核酸可以与每个二级核酸杂交。

[0241] 在一些方面,本公开内容提供了间隔区3x4报告探针。间隔区3x4报告探针包含一级核酸,其中所述一级核酸包含12个核苷酸的第一结构域。位于一级核酸的第一结构域和第二结构域之间的是由20至40个核苷酸组成的间隔区域。间隔区被鉴定为20至40个核苷酸长;然而,间隔区的长度是非限制性的,并且它可以短于20个核苷酸或长于40个核苷酸。一级核酸的第二结构域包含可以与3个二级核酸分子杂交的核苷酸序列。每个二级核酸包含这样的核苷酸序列,使得与可检测标记结合的4个三级核酸可以与每个二级核酸杂交。

[0242] 在一些方面,一级核酸可以包含12个核苷酸长的第一结构域。然而,一级核酸的第一结构域的长度是不受限制的,并且可以少于12个或多于12个核苷酸。在一个实例中,一级核酸的第一结构域是14个核苷酸。在另一个实例中,一级核酸的第一结构域是9个核苷酸。在一个进一步实例中,一级核酸的第一结构域是8个核苷酸。报告探针的一级核酸的9个核苷酸第一结构域的示例性序列包括表15中的那些。

[0243] 表15

[0244]

报告物位置	9聚体序列	颜色	报告物位置	9聚体序列	颜色
1	CATTGGGTT	BB	2	CGGGGTTTA	GR
1	CTGGTATGT	BG	2	CAAATTGGT	GY
1	CAGTGAGTG	BR	2	CGAAGTGGT	RR
1	CAGGAAGGT	BY	2	CTGTTAGGG	YR
1	CGATGGATG	GG	2	CGTGTGTG	YY
1	CGGTGGAAT	GR	3	CTTTGGTTT	BB
1	CAAAAGAGG	GY	3	CGAGTGGGA	BG
1	CAGGAGAAA	RR	3	CTAGTAGGG	BR
1	CAAGGGTAG	YR	3	CTTTGTGTT	BY
1	CGAGATGAG	YY	3	CATGGGGTG	GG
2	CTTGTGATG	BB	3	CGAAGTTGA	GR
2	CGGGTTAGA	BG	3	CGGTGATTT	GY
2	CGTATGGTT	BR	3	CTATTGTGG	RR
2	CGATTGGTA	BY	3	CTTAGGGAG	YR
2	CATGGTGTA	GG	3	CGGTGGAGG	YY

[0245] 本公开内容的特异性报告探针设计的任何特征,可以与本公开内容的另一种报告探针设计的任何特征组合。例如,可以修饰5x5报告探针,以含有在互补核酸和一级核酸之间的大约20至40个核苷酸的间隔区域。在另一个实例中,可以修饰4x3报告探针,使得4个二级核酸包含核苷酸序列,其允许与可检测标记结合的5个三级核酸与每个二级核酸杂交,从

而产生4x5报告探针。

[0246] 不希望受理论的束缚,5x5报告物(25)含有比4x3报告物(12)更多的荧光标记,并且因此5x5报告物的荧光强度将更大。在任何给定视野FOV中检测到的荧光是各种变量的函数,所述变量包括给定报告探针的荧光强度、以及该FOV内任选结合的靶分子的数目。任选结合的靶分子的数目/视野(FOV)可以是1至250万个靶/FOV。结合的靶分子的通常数目/FOV是20,000至40,000、220,000至440,000或1百万至2百万个靶分子。通常的FOV为.05 mm²至1mm²。通常FOV的进一步实例是.05 mm²至.65mm²。

[0247] 在一些方面,本公开内容提供了报告探针设计,其中二级核酸分子包含“额外手柄”,其不与三级核酸分子杂交并且在一级核酸分子的远端。在一些方面,“额外手柄”可以是12个核苷酸长(“12聚体”);然而,它们的长度是不受限制的,并且可以少于12或多于12个核苷酸。“额外手柄”可以各自包含二级核酸分子与之杂交的一级核酸分子的第一结构域的核苷酸序列。因此,当报告探针包含“额外手柄”时,报告探针可以经由一级核酸分子的第一结构域、或经由“额外手柄”与测序探针杂交。相应地,增加了报告探针与测序探针结合的可能性。“额外手柄”设计还可以改善杂交动力学。不受任何理论的束缚,“额外手柄”可以增加报告探针的互补核酸的有效浓度。预计5x4“额外手柄”报告探针产生大约4750个荧光计数/标准FOV。预计5x3“额外手柄”报告探针、4x4“额外手柄”报告探针、4x3“额外手柄”报告探针和3x4“额外手柄”报告探针全部产生大约6000个荧光计数/标准FOV。本公开内容的任何报告探针设计都可以被修改为包括“额外手柄”。

[0248] 报告探针的各个二级核酸分子可以与三级核酸分子杂交,所述三级核酸分子全部用相同的可检测标记进行标记。例如,图5的左图描绘了“5x6”报告探针。5x6报告探针包括包含第二结构域的一个一级核酸,其中所述第二结构域包含与6个二级核酸分子杂交的核苷酸序列。每个二级核酸包含这样的核苷酸序列,使得与可检测标记结合的5个三级核酸分子与每个二级核酸杂交。与特定二级核酸分子结合的5个三级核酸分子各自用相同的可检测标记进行标记。例如,三个二级核酸分子与用黄色荧光染料标记的三级核酸分子结合,而其它三个二级核酸与用红色荧光染料标记的三级核酸分子结合。

[0249] 报告探针的各个二级核酸分子可以与三级核酸分子杂交,所述三级核酸分子用不同的可检测标记进行标记。例如,图5的中图描绘了“3x2x6”报告探针设计。“3x2x6”报告探针包括包含第二结构域的一个一级核酸,其中所述第二结构域包含与6个二级核酸分子杂交的核苷酸序列。每个二级核酸包含这样的核苷酸序列,使得与可检测标记结合的5个三级核酸与每个二级核酸杂交。每个二级核酸与用黄色荧光染料标记的三级核酸分子、以及用红色荧光染料标记的三级核酸分子两者结合。在该具体实例中,三个二级核酸分子结合两个红色和三个黄色三级核酸分子,而其它三个二级核酸分子结合两个红色和三个黄色三级核酸分子。每个二级核酸分子可以与任何数目的三级核酸分子结合,所述三级核酸分子由不同的可检测标记结合。在图5的中图中,这样排列与各个二级核酸分子结合的三级核酸分子,使得标记的颜色交替(即,红色-黄色-红色-黄色-红色或黄色-红色-黄色-红色-黄色)。

[0250] 在任何描述的报告探针设计中,用不同的可检测标记进行标记的三级核酸可以以任何次序沿着二级核酸排列。例如,图5的右图描绘了“FRET抗性3x2x6”报告探针,其类似于3x2x6报告探针设计,除了沿着每个二级核酸分子的红色和黄色三级核酸分子的排列(例如线性次序或分组)之外。

[0251] 图6描绘了本公开内容的更多示例性的报告探针设计,其包括与不同的三级核酸分子结合的各个二级核酸分子。左图描绘了包含一个一级核酸分子的“6x1x4.5”报告探针,其中所述一级核酸分子包含第二结构域,其中所述第二结构域包含与六个二级核酸分子杂交的核苷酸序列。每个二级核酸分子与五个三级核酸分子杂交。与每个二级核酸分子杂交的五个三级核酸分子中的四个直接用相同颜色的可检测标记进行标记。指示为分支三级核酸的第五三级核酸与双色组合的另一种颜色的5个标记的寡核苷酸结合。在这六个二级核酸中,其中三个与用双色组合的一种颜色(在该实例中为红色)标记的分支三级核酸结合,而其它三个二级核酸与用双色组合的另一种颜色(在该实例中为黄色)标记的分支三级核酸结合。总的来说,6x1x4.5报告探针用总共54种染料进行标记,每种颜色27种染料。图6的中图描绘了“4x1x4.5”报告探针,其与6x1x4.5报告探针共享相同的总体结构,除了4x1x4.5报告探针的一级核酸仅结合4个二级核酸,使得存在总共36种染料,每种颜色18种之外。

[0252] 对于双色组合的每种颜色,报告探针可以包含相同数目的染料。对于双色组合的每种颜色,报告探针可以包含不同数目的染料。可以基于两种染料吸收的光的能级,来作出关于哪种颜色在报告探针内具有更多染料的选择。例如,图6的右图描绘了“5x5能量优化的”报告探针设计。这种报告探针设计包含15种黄色染料(其能量较高)和10种红色染料(其能量较低)。在该实例中,15种黄色染料可以构成第一标记,而10种红色染料可以构成第二标记。

[0253] 可检测部分、标记或报告物可以各种方式与二级核酸分子、三级核酸分子或标记的寡核苷酸结合,包括直接或间接附着可检测部分,例如荧光部分、比色部分等等。本领域技术人员可以查阅涉及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、花青、丹磺酰氯、藻蓝蛋白、藻红蛋白等等。

[0254] 许多综述中描述了荧光标记及其与核苷酸和/或寡核苷酸的附着,包括Haugland, Handbook of Fluorescent Probes and Research Chemicals, Ninth Edition (Molecular Probes, Inc., Eugene, 2002); Keller和Manak, DNA Probes, 第2版 (Stockton Press, New York, 1993); Eckstein, 编辑, Oligonucleotides and Analogues: A Practical Approach (IRL Press, Oxford, 1991); 以及Wetmur, Critical Reviews in Biochemistry and Molecular Biology, 26: 227-259 (1991)。适用于本公开内容的特定方法学公开于下述参考文献实例中:美国专利号4,757,141; 5,151,507; 和5,091,519。一种或多种荧光染料可以用作标记的靶序列的标记,例如,如通过美国专利号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); 等等所公开的。标记也可以用量子点进行,如下述专利和专利公开中所公开的: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。如本文使用的,术语“荧光标记”包含信号传导部分,其通过一种或多种分子的荧光吸收和/或发射特性来传递信息。此类荧光特性包括荧光强度、荧光寿命、发射光谱特征、能量转移等等。

[0255] 易于掺入核苷酸和/或寡核苷酸序列内的商购可得的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 GREENR™ 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、ALEXA FLUOR™ 546-14-UTP (Molecular Probes, Inc. Eugene, OR) 等等。可替代地,可以在使用例如亚磷酰胺或NHS化学的寡核苷酸合成期间,添加上述荧光团和本文提到的那些。用于定制合成具有其它荧光团的核苷酸的方案是本领域已知的(参见, Henegariu等人(2000) Nature Biotechnol. 18:345)。2-氨基嘌呤是荧光碱基,其可以在寡核苷酸序列合成期间直接掺入其中。核酸也可以用嵌入染料例如DAPI、YOYO-1、溴化乙锭、花青染料(例如SYBR Green)等等事先进行染色。

[0256] 可用于合成后附着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、瀑布蓝、瀑布黄、丹磺酰、丽丝胺罗丹明B、海洋蓝、俄勒冈绿488、俄勒冈绿514、太平洋蓝、太平洋橙、罗丹明6G、罗丹明绿、罗丹明红、四甲基罗丹明、德克萨斯红(可得自Molecular Probes, Inc., Eugene, OR)、Cy2、Cy3、Cy3.5、Cy5、Cy5.5、Cy7(Amersham Biosciences, Piscataway, NJ) 等等。也可以使用FRET串联荧光团,包括但不限于PerCP-Cy5.5、PE-Cy5、PE-Cy5.5、PE-Cy7、PE-德克萨斯红、APC-Cy7、PE-Alexa染料(610、647和680)、APC-Alexa染料等等。

[0257] 金属银或金颗粒可以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)。

[0258] 用于寡核苷酸序列的其它合适标记可以包括荧光素(FAM、FITC)、地高辛配基、二硝基苯酚(DNP)、丹磺酰、生物素、溴脱氧尿苷(BrdU)、六组氨酸(6xHis)、磷-氨基酸(例如P-tyr、P-ser、P-thr)等等。下述半抗原/抗体对可以用于检测,其中每种抗体用可检测标记进行衍生化:生物素/ α -生物素、地高辛配基/ α -地高辛配基、二硝基苯酚(DNP)/ α -DNP、5-羧基荧光素(FAM)/ α -FAM。

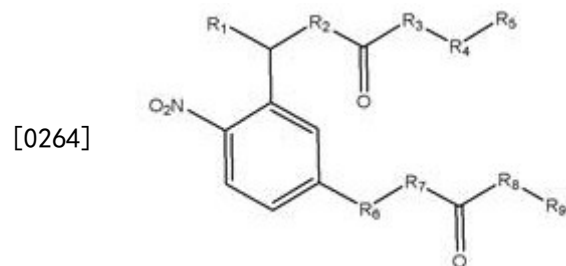
[0259] 本文所述的可检测标记是光谱可分辨的。提及多个荧光标记的“光谱可分辨的”意指标记的荧光发射谱带是足够不同的,即足够不重叠,使得分别的标记与之附着的分子标签可以通过标准的光检测系统,基于通过分别的标记生成的荧光信号来区分,所述光检测系统例如采用带通滤波器和光电倍增管等等的系统,如通过美国专利号4,230,558;4,811,218;等等,或Wheless等人,Flow Cytometry:Instrumentation and Data Analysis (Academic Press, New York, 1985)中的第21-76页中所述的系统例示的。光谱可分辨的有机染料,例如荧光素、罗丹明等等,意指波长发射最大值间隔开至少20 nm,并且在另一个方面,间隔开至少40 nm。对于螯合的镧系元素化合物、量子点等等,光谱可分辨的意指波长发射最大值间隔开至少10 nm或至少15 nm。

[0260] 在条形码结构域中存在三个附着位置,每个位置具有至多10种潜在的双色组合,允许存在至多1000种颜色组合。如果报告探针在少于1000种探针/池中合并,则可以利用使用奇偶校验来克服错误的的能力。可以存在允许奇偶校验的许多潜在的奇偶性方案,图32中显示了一个示例性方案。在该实例中,存在的实际颜色不用作奇偶校验,而是使用在条形码结构域中的每个附着位置处存在的单个(S)颜色报告探针(例如红色)和多色(M)报告探针(例如红色/黄色)。如奇偶性设计中可见的,任何两个报告物位置的状态(S或M)的了解允许预测第三个位置。在所示的实例中,在任何两个位置中观察到S要求未观察到的位置是M,在任何两个位置中观察到S和M意味着另一个位置必须是S,而观察到两个M报告探针要求另一个位置是M。这意味着为了获取具有不正确检测的报告物颜色的三个报告探针的代码,必须进行两次不正确的识别。图32显示了在5%的报告探针错误下的模拟结果,其显示了当应用奇偶校验时,错误过滤中的增加。存在可以应用的多重奇偶性系统,这只是一个实例。

[0261] 另一种错误校正程序是对于每个报告探针池交换彩色调色板。彩色调色板是实际用于测量池的报告探针集合。任何一个池中都不使用多重报告探针,如果池中存在500种报告探针,则仅需要 $\frac{1}{2}$ 可能的颜色组合。实现这点的的最简单方式是具有两个调色板,含有500种报告探针的调色板A和含有其它500种报告探针的调色板B。因此,如果测序池1、3、5、7具有调色板A,而池2、4、6、8具有调色板B,则按次序1、2、3、4、5、6、7、8运行池意味着每个相继测序池具有分开的调色板。因此,来自池2的条形码不存在于之前和之后的池(例如池1和3)中。这允许简单的自动故障排除和检测错误的限制。

[0262] 报告探针可以包含一种或多种可切割接头修饰。一种或多种可切割接头修饰可以置于报告探针中的任何位置。可切割接头修饰可以位于报告探针的一级核酸分子的第一结构域和第二结构域之间。可切割接头修饰可以存在于报告探针的二级核酸分子的第一结构域和第二结构域之间。可切割接头修饰可以存在于报告探针的一级核酸分子和二级核酸分子的第一结构域和第二结构域之间。图7的左图描绘了本公开内容的示例性报告探针,其包含在一级核酸的第一结构域和第二结构域之间、以及在二级核酸的第一结构域和第二结构域之间的可切割接头修饰。在如图7的左图中例示的这种情况下,可切割接头修饰可以包括一个或多个可切割部分,例如图7的左图中例示的那些。

[0263] 可切割接头修饰物可以是式(I)的化合物或其立体异构体或盐:



[0265] (I)

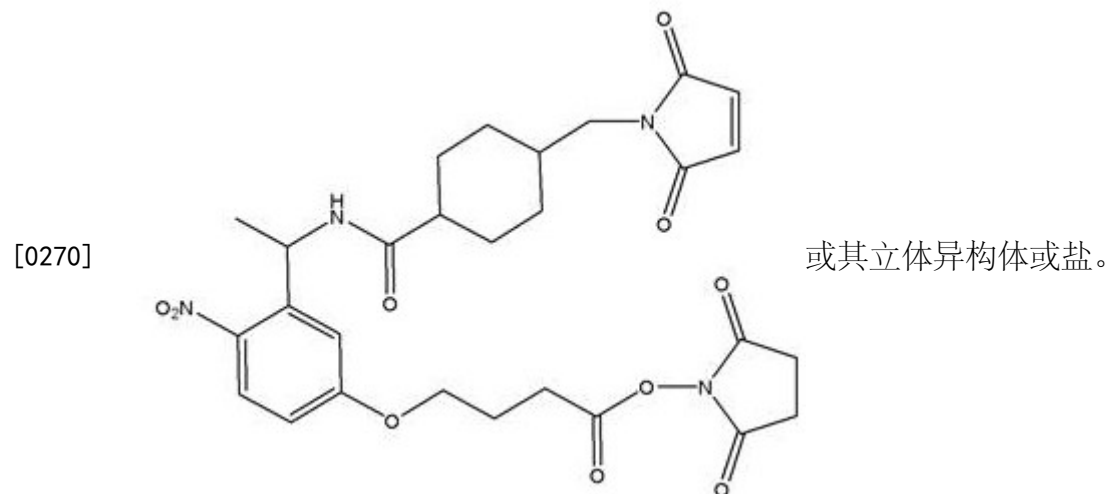
[0266] 其中:R₁为氢、卤素、C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基,其中所述C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基各自独立地任选地由至少一个取代基R₁₀取代;R₂为O、NH或N(C₁₋₆烷基);R₃为环烷基、杂环烷基、芳基或杂芳基,各自任选地由至少一个取代基R₁₀取代;每个R₄和R₇独立地为C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基,其中所述C₁₋₆烷基、C₂₋₆烯基、C₂₋₆炔基各自独立地任选地由至少一个取代基R₁₀取代;R₅和R₉各自独立地为环烷基、杂环烷基、芳基或杂芳基,各自任选地由至少

一个取代基 R_{10} 取代; R_6 为O、NH或N(C_{1-6} 烷基); R_8 为O、NH或N(C_{1-6} 烷基);每个 R_{10} 独立地为氢、卤素、 $-C_{1-6}$ 烷基、 $-C_{2-6}$ 烯基、 $-C_{2-6}$ 炔基、卤代 C_{1-6} 烷基、卤代 C_{2-6} 烯基、卤代 C_{2-6} 炔基、环烷基、杂环基、芳基、杂芳基、 $-CN$ 、 $-NO_2$ 、氧代、 $-OR_{11}$ 、 $-SO_2R_{11}$ 、 $-SO_3^-$ 、 $-COR_{11}$ 、 $-CO_2R_{11}$ 、 $-CONR_{11}R_{12}$ 、 $-C(=NR_{11})NR_{12}R_{13}$ 、 $-NR_{11}R_{12}$ 、 $-NR_{12}COR_{12}$ 、 $-NR_{11}CONR_{12}R_{13}$ 、 $-NR_{11}CO_2R_{12}$ 、 $-NR_{11}SONR_{12}R_{13}$ 、 $-NR_{11}SO_2NR_{12}R_{13}$ 或 $-NR_{11}SO_2R_{12}$;并且可以是相同或不同的 R_{11} 、 R_{12} 和 R_{13} 各自独立地为氢、 $-C_{1-6}$ 烷基、 $-C_{2-6}$ 烯基、 $-C_{2-6}$ 炔基、卤代 C_{1-6} 烷基、卤代 C_{2-6} 烯基、卤代 C_{2-6} 炔基、 C_{1-6} 烷氧基 C_{1-6} 烷基、环烷基、杂环基、芳基或杂芳基。

[0267] 在一个方面, R_1 为 C_{1-6} 烷基,优选 C_{1-3} 烷基,例如甲基、乙基、丙基或异丙基; R_2 为NH或N(C_{1-6} 烷基); R_3 为5-至6-元环烷基,优选环己基; R_4 为 C_{1-6} 烷基,优选 C_{1-3} 亚烷基,例如亚甲基、亚乙基、亚丙基或亚异丙基; R_5 为5-至6-元杂环基,其包含1个氮原子和0或1个另外的选自N、O和S的杂原子,其中所述杂环基任选地由1或2个 R_{10} 取代; R_6 为O; R_7 为 C_{1-6} 烷基,优选 C_{1-3} 亚烷基,例如亚甲基、亚乙基、亚丙基或亚异丙基; R_8 为O; R_9 为5-至6-元杂环基,其包含1个氮原子和0或1个另外的选自N、O和S的杂原子,其中所述杂环基任选地由1或2个 R_{10} 取代;并且每个 R_{10} 独立地为卤素、 C_{1-6} 烷基、卤代 C_{1-6} 烷基、氧代、 $-SO_2H$ 或 $-SO_3^-$ 。

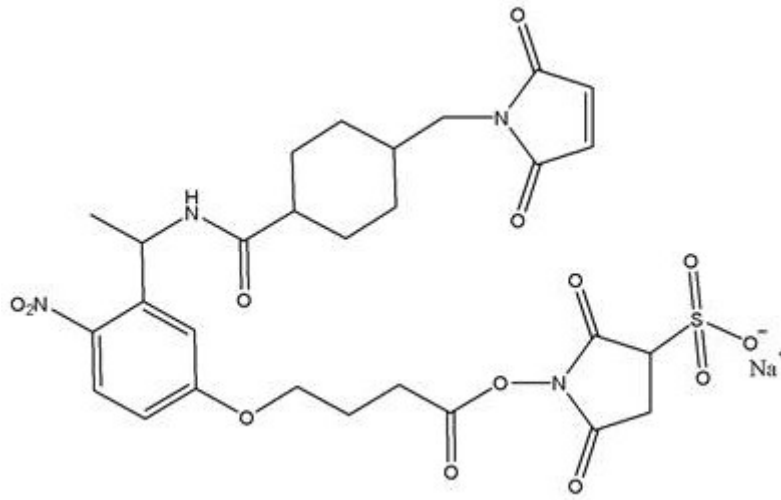
[0268] 在一个方面, R_3 为环己基, R_4 为亚甲基, R_5 为1H-吡咯-2,5-二酮,且 R_9 为吡咯烷-2,5-二酮,任选地由 SO_3^- 取代。

[0269] 接头化合物可以是



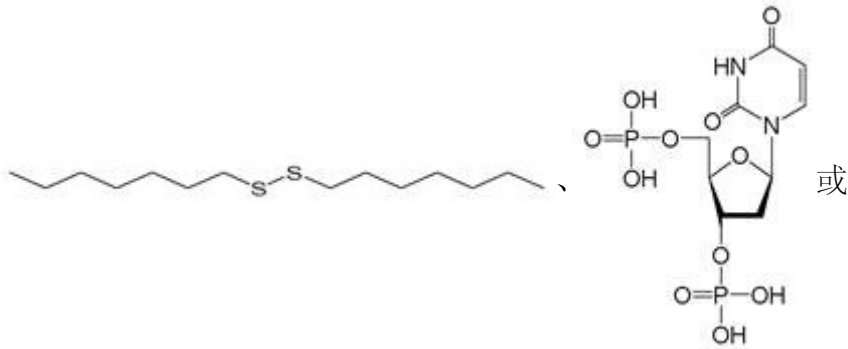
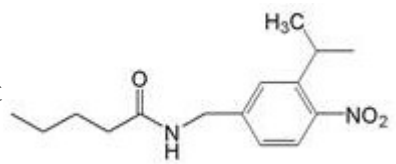
[0271] 接头化合物可以是

[0272]

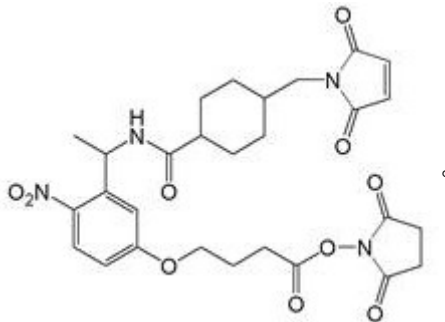


或其立体异构体或盐。

[0273] 接头化合物或接头修饰可以是

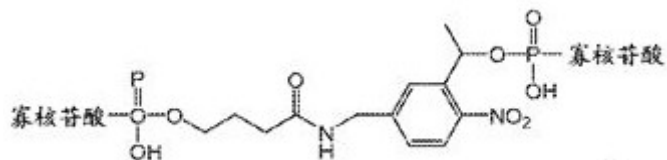


或



[0274] 接头化合物或接头修饰可以是

[0275]



地无需转换(或扩增)步骤。例如,对于直接基于RNA的测序,本方法不需要在获得序列之前,将RNA分子转换为DNA分子(即,经由cDNA的合成)。由于不需要扩增或转化,本公开内容中测序的核酸将保留当核酸处于样品中或当其从样品获得时,核酸中存在的任何独特碱基和/或表观遗传标记物。此类独特碱基和/或表观遗传标记物在本领域已知的测序方法中丢失。

[0283] 本方法可以用于以单分子分辨率进行测序。换言之,本方法允许用户基于从单个靶核酸分子收集的数据来生成最终序列,而不是必须组合来自不同靶核酸分子的数据,保存了该特定靶的任何独特特征。

[0284] 靶核酸可以得自核酸的任何样品或来源,例如任何细胞、组织或生物、体外、化学合成仪等。靶核酸可以通过任何领域公认的方法获得。核酸可以从临床受试者的血液样品获得。可以使用本领域众所周知的方法和试剂盒,从来源或样品中提取、分离或纯化核酸。

[0285] 靶核酸可以通过本领域已知的任何手段进行片段化。优选地,通过酶促或机械手段执行片段化。机械手段可以是超声处理或物理剪切。可以通过用核酸酶(例如,脱氧核糖核酸酶I(DNase I))、或者一种或多种限制性核酸内切酶消化来执行酶促手段。

[0286] 当包含靶核酸的核酸分子是完整染色体时,应采取措施以避免使染色体片段化。

[0287] 靶核酸可以包括天然或非天然核苷酸,包含修饰的核苷酸或核酸类似物,如本领域众所周知的。

[0288] 靶核酸分子可以包括长度高达数百的千碱基(例如1、2、3、4、5、10、20、30、40、50、100、200、500或更多的千碱基)的DNA、RNA和PNA分子。靶核酸分子可以包含约50至约400个核苷酸,或约90至约350个核苷酸。

[0289] 捕获探针

[0290] 靶核酸可以被固定(例如,在一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个位置处)至基底。

[0291] 示例性的有用基底包括包含选自配体、抗原、碳水化合物、核酸、受体、凝集素和抗体的结合部分的那些基底。捕获探针包含能够与基底的结合部分结合的基底结合部分。包含反应性部分的示例性的有用基底包括但不限于包含以下的表面:环氧基、醛、金、酰肼、巯基、NHS-酯、胺、炔、叠氮化物、硫醇、羧酸酯、马来酰亚胺、羟甲基磷、亚氨酸酯、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砷、聚乙二醇(PEG)、水凝胶或其混合物。此类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的有用基底包括但不限于OptArray-DNA NHS基团(Accler8)、Nexterion Slide AL(Schott)和Nexterion Slide E(Schott)。

[0292] 基底可以是本领域已知的任何固体支撑物,例如涂布的载玻片和微流体装置,其能够固定靶核酸。基底可以是表面、膜、珠、多孔材料、电极或阵列。例如,基底可以是聚合物材料、金属、硅、玻璃或石英。靶核酸可以固定到对于本领域技术人员显而易见的任何基底上。

[0293] 当基底是阵列时,基底可以包含孔,其大小和间隔根据待附着的靶核酸分子而变。在一个实例中,这样构建基底,使得附着超密集的有序的靶核酸阵列。基底上的靶核酸阵列密度的实例包括500,000至10,000,000个靶核酸分子/ mm^2 、1,000,000至4,000,000个靶核酸分子/ mm^2 、或850,000至3,500,000个靶核酸分子/ mm^2 。

[0294] 基底中的孔是用于靶核酸分子附着的位置。孔的表面可以用上述反应性部分官能

化,以吸引且结合存在于靶核酸分子上的特异性化学基团,或捕获与靶核酸分子结合的探针,以吸引、固定且结合靶核酸分子。众所周知,这些官能团能够通过各种缀合化学特异性地吸引且结合生物分子。

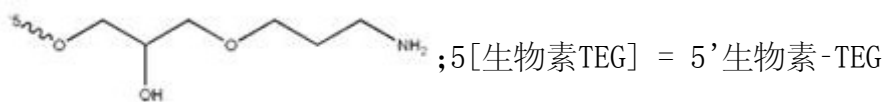
[0295] 对于在基底例如阵列上的单个核酸分子测序,将通用捕获探针或与捕获探针的基底结合部分互补的通用序列附着至每个孔。然后使单个靶核酸分子与通用捕获探针或与捕获探针结合的捕获探针的基底结合部分互补的通用序列结合,并且可以开始测序。

[0296] 对于在基底例如阵列上的单个核酸分子测序,单个靶核酸分子可以与捕获探针结合。捕获探针的基底结合部分然后可以与衔接子寡核苷酸结合。衔接子核苷酸然后与附着至每个孔的草坪寡核苷酸结合,并且可以开始测序。表8中显示了草坪寡核苷酸的示例性序列。

[0297] 表8

示例性的草坪寡核苷酸序列	SEQ ID NO.
5AmMC6/TGGTGAGGTTGTTGGTAGTAGTGAGTTTGTAGG GT	100
5AmMC6/TGGTGAGGTTGTTGGTAGTAGTGAG	101
5AmMC6/TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	102
5AmMC6/CATCTCAAACACCTTCTACAATATGACCTAACAC AC	103
5AmMC6/GTGATGGTTATAAGAGGTGTTGATATATTTATAG TA	104
5AmMC6/TATTGATATTGAGAAAGCGTTTGATGATGTATTG AT	105
5AmMC6/TAGTTATGTAGTAGTTTGCGAAAGAGTTATAGTT AT	106
/5AmMC6/ACTACCCTACTCTACCCTTCTAAGATATACATAT AC	107
5AmMC6/TG/isodG/TGA/isodG/GTT/isodG/TTG/isodG/TA/ isodG/TA/isodG/TGA/isodG/TTT/isodG/TAG/isodG/GT	108
5[生物素 TEG]/L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L- dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L- dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L- dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//L-dT//3AmMO/	114

[0299] 5amMC6 = 具有6碳接头的5'胺;isodG = 异鸟嘌呤;3AmMO = 3'



[0300] 包含草坪寡核苷酸或衔接子寡核苷酸的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。可用于草坪寡核苷酸或衔接子寡核苷酸中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。还可替代地,包含草坪寡核苷酸的每种核酸可以独立地是L-DNA。在一些方面,草坪寡核苷酸可以包含L-DNA。草坪寡核苷酸可以由L-DNA组成。草坪寡核苷酸可以基本上由L-DNA组成。例如,修饰的核苷酸或核苷酸类似物如异鸟嘌呤和异胞嘧啶或L-DNA的使用,可以改善衔接子寡核苷酸与草坪寡核苷酸内的适当互补核酸序列的结合效率和准确率,同时使其它地方的结合降到最低。

[0301] 草坪寡核苷酸可以进一步包含在本文中称为5AmMC6,具有6碳接头的5'胺。5AmMC6可以用于将草坪寡核苷酸附着至基底。

[0302] 捕获探针、衔接子寡核苷酸和草坪寡核苷酸复合物的实例显示于图33中。在该图中,杂交的示例性的衔接子序列和示例性的捕获探针序列为绿色,其为示例性的草坪寡核苷酸的反向互补体的序列为蓝色,并且捕获探针上以红色的示例性序列与靶基因杂交,所述靶基因在该实例中为基因TP53.1。示例性的捕获探针的序列是3' - CCGTCAACCGTTTTGT AGAACAACTCCCGTCCCCTCACTCACTAGCCTCCAGTACCGAAAGC-5' (SEQ ID No:111)。示例性的衔接子序列的序列是5' -GAGTGATCGGAGGTCATGGCTTTCGAC/iMe-isodC/CTA/iMe-isodC/AAA/iMe-isodC/TCA/iMe-isodC/TA/iMe-isodC/TA/iMe-isodC/CAA/iMe-isodC/AAC/iMe-isodC/TCA/iMe-isodC/CA-3' (SEQ ID No:110)。示例性草坪寡核苷酸的序列是TG/iisodG/GAT/iisodG/TTT/iisodG/AGT/iisodG/AT/iisodG/AT/iisodG/GTT/iisodG/TTG/iisodG/AGT/iisodG/GT/5AmMC6 (SEQ ID NO:108)。

[0303] 在一些方面,草坪寡核苷酸可以包含至少一个亲和力部分、至少两个亲和力部分、至少三个亲和力部分、至少四个亲和力部分、至少五个亲和力部分、至少六个亲和力部分、至少七个亲和力部分、至少八个亲和力部分、至少九个亲和力部分或至少十个亲和力部分。亲和力部分可以是生物素。因此,草坪寡核苷酸可以包含至少一个生物素部分、至少两个生物素部分、至少三个生物素部分、至少四个生物素部分、至少五个生物素部分、至少六个生物素部分、至少七个生物素部分、至少八个生物素部分、至少九个生物素部分或至少十个生物素部分。

[0304] 在一些方面,与靶核酸杂交的本公开内容的捕获探针可以包含至少一个第一亲和力部分,例如但不限于生物素部分。然后,与靶核酸杂交的捕获探针可以与基底上的至少一种草坪寡核苷酸直接或间接杂交,其中所述至少一种草坪寡核苷酸包含至少一个第一亲和力部分,例如但不限于生物素部分。在捕获探针与草坪寡核苷酸杂交后,可以使所得的捕获探针-靶核酸-草坪寡核苷酸复合物与第二亲和力部分一起温育,其中所述第二亲和力部分能够结合位于捕获探针上的第一亲和力部分、以及位于草坪寡核苷酸上的第一亲和力部分。在非限制性实例中,如果位于捕获探针上的第一亲和力部分、以及位于草坪寡核苷酸上的第一亲和力部分均为生物素,则中性抗生物素蛋白可以用作第二亲和力部分。第二亲和力部分将结合位于捕获探针上的第一亲和力部分、以及位于草坪寡核苷酸上的第一亲和力部分,产生在本文中被称作“蛋白质锁”的蛋白质桥。蛋白质锁可以用于将靶核酸更稳定地固定到基底上。图67显示了使用生物素化的捕获探针和草坪寡核苷酸和中性抗生物素蛋白的蛋白质锁的示意图。

[0305] 靶核酸可以由一种或多种捕获探针(即,两种、三种、四种、五种、六种、七种、八种、

九种、十种或更多种捕获探针)结合。捕获探针包含与靶核酸的一部分互补的结构域,以及包含基底结合部分的结构域。捕获探针与之互补的靶核酸的一部分可以是靶核酸的末端或不朝向末端。捕获探针可以包含在与靶核酸的一部分互补的结构域、以及包含基底结合部分的结构域之间的可切割部分。

[0306] 可替代地,捕获探针可以包含与靶核酸的一部分互补的第一结构域、包含基底结合部分的第二结构域、以及包含不同的基底结合部分的第三结构域。捕获探针可以包含在任何结构域之间的可切割部分。

[0307] 捕获探针可以在5'末端处进行磷酸化。可替代地,捕获探针可以包含至少一个硫代磷酸酯键。捕获探针可以包含至少两个硫代磷酸酯键。优选地,至少一个或至少两个硫代磷酸酯键位于捕获探针的5'末端处。

[0308] 捕获探针的基底结合部分可以是生物素,并且基底可以是抗生物素蛋白(例如,链霉抗生物素蛋白)。包含抗生物素蛋白的有用基底是商购可得的,包括TB0200 (Accelr8)、SAD6、SAD20、SAD100、SAD500、SAD2000 (Xantec)、SuperAvidin(Array-It)、链霉抗生物素蛋白载玻片(目录#MPC 000, Xenopore)和STREPTAVIDINn载玻片(目录#439003, Greiner Bio-one)。捕获探针的基底结合部分可以是抗生物素蛋白(例如,链霉抗生物素蛋白),并且基底可以是生物素。商购可得的包含生物素的有用基底包括但不限于Optiarray-生物素(Accler8)、BD6、BD20、BD100、BD500和BD2000 (Xantec)。

[0309] 捕获探针的基底结合部分可以是能够通过光活化与基底结合的反应性部分。基底可以包含光反应性部分,或纳米报告物的第一部分可以包含光反应性部分。光反应性部分的一些实例包括芳基叠氮化物,例如N-((2-吡啶基二硫代)乙基)-4-叠氮基水杨酰胺;氟化芳基叠氮化物,例如4-叠氮基-2,3,5,6-四氟苯甲酸;基于二苯甲酮的试剂,例如4-苯甲酰基苯甲酸的琥珀酰亚胺酯;和5-溴-脱氧尿苷。

[0310] 捕获探针的基底结合部分可以是核酸,所述核酸可以与互补的基底的结合部分杂交。包含捕获探针的基底结合部分的每种核酸,可以独立地是规范碱基或修饰的核苷酸或核酸类似物。捕获探针的基底结合部分中的至少一个、至少两个、至少三个、至少四个、至少五个或至少六个核苷酸可以是修饰的核苷酸或核苷酸类似物。捕获探针的基底结合部分中修饰的核苷酸或核苷酸类似物与规范碱基的通常比率为1:2至1:8。可用于捕获探针的基底结合部分中、通常的修饰的核苷酸或核酸类似物是异鸟嘌呤和异胞嘧啶。

[0311] 捕获探针的基底结合部分可以经由对于本领域技术人员显而易见的其它结合对固定至基底。在与基底结合后,可以通过施加足以延伸靶核酸的力(例如重力、水动力、电磁力“电拉伸”、流动拉伸、后退弯月面技术及其组合)来伸长靶核酸。捕获探针可以包含可检测标记或与可检测标记相关联,即基准位点。

[0312] 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结构域。由第二捕获探针结合的靶核酸的第二部分不同于由第一捕获探针结合的靶核酸的第一部分。该部分可以是靶核酸的末端或不朝向末端。第二捕获探针的结合可以在靶核酸伸长之后或期间发生,或者对尚未伸长的靶核酸发生。第二捕获探针可以具有如上所述的结合。

[0313] 靶核酸可以由第三、第四、第五、第六、第七、第八、第九或第十捕获探针结合,所述捕获探针包含与靶核酸的第三、第四、第五、第六、第七、第八、第九或第十部分互补的结构

域。该部分可以是靶核酸的末端或不朝向末端。第三、第四、第五、第六、第七、第八、第九或第十捕获探针的结合可以在靶核酸伸长之后或期间发生,或者对尚未伸长的靶核酸发生。第三、第四、第五、第六、第七、第八、第九或第十捕获探针可以具有如上所述的结合。

[0314] 捕获探针能够从样品中分离靶核酸。在此处,将捕获探针加入包含靶核酸的样品中。捕获探针经由其与靶核酸的区域互补的捕获探针的区域结合靶核酸。当靶核酸接触包含结合捕获探针的基底结合部分的部分的基底时,核酸变成固定到基底上。

[0315] 图8显示了使用本公开内容的双捕获探针系统的靶核酸捕获。基因组DNA在95°C下变性,并且与捕获试剂池杂交。该捕获试剂池包含寡核苷酸探针A、探针B和反义阻断探针。探针A包含在探针的3'末端处的生物素部分,以及与靶核酸的5'末端互补的序列。探针B包含可以通过在探针5'的末端处的顺磁珠结合的纯化结合序列,以及与靶核酸的3'末端互补的核苷酸序列。反义阻断探针包含与待测序的靶核酸一部分的反义链互补的核苷酸序列。在与捕获试剂杂交后,在杂交的探针A和探针B之间的靶核酸上产生测序窗口。靶核酸使用与探针B的5'序列结合的顺磁珠进行纯化。洗掉任何过量的捕获试剂或互补的反义DNA链,导致预期的靶核酸的纯化。然后使纯化的靶核酸流过流动腔室,所述流动腔室包括可以与杂交的探针A上的生物素部分结合的表面,例如链霉抗生物素蛋白。这导致靶核酸的一个末端栓系至流动室的表面。为了捕获另一个末端,对靶核酸进行流动拉伸,并且添加与探针B的纯化结合序列互补的生物素化探针。在与探针B的纯化结合序列杂交后,生物素化的探针可以与流动室的表面结合,导致捕获的靶核酸分子伸长并在两个末端处与流动室表面结合。

[0316] 为了确保用户从高度片段化的样品中“捕获”尽可能多的靶核酸分子,包括多种捕获探针是有帮助的,每种捕获探针与靶核酸的不同区域互补。例如,可以存在3个捕获探针池,其中第一池与靶核酸接近其5'末端的区域互补,第二池与靶核酸的中间区域互补,而第三池与接近3'末端的区域互补。这可以类推至每种靶核酸的“n个目的区域”。在该实例中,片段化靶核酸的每个个别池与包含或结合生物素标签的捕获探针结合。对于每个池腔室,分割输入样品的1/n(其中n = 靶核酸中的不同区域的数目)。捕获探针结合目的靶核酸。然后靶核酸经由捕获探针的生物素固定到粘附至基底的抗生物素蛋白分子。任选地,例如经由流动或静电力拉伸靶核酸。所有n个池可以同时拉伸且结合,或者,为了使完全拉伸的分子数目达到最大,可以首先拉伸且结合池1(其捕获最5'区域);然后可以拉伸且结合池2(其捕获靶的中间区域);最后可以拉伸且结合池3。

[0317] 可以使用本公开内容的“基于两珠的步骤纯化”系统来捕获靶核酸。存在四种捕获探针:探针A、探针B、探针C和探针D。探针A包含OA序列,与靶核酸的5'末端互补的核酸序列,以及附着至生物素部分的核酸序列。OA序列可以包含核苷酸序列CGAAAGCCATGACCTCCGATCACTC(SEQ ID NO:109),并且可以与草帽寡核苷酸结合。附着至生物素部分的核酸序列经由可切割接头连接至与靶核酸的5'末端互补的核酸序列。探针B和探针C包含与靶核酸互补的核酸序列,以及附着至生物素部分的核酸序列。附着至生物素部分的核酸序列经由可切割接头连接至与靶核酸互补的核酸序列。探针D包含与靶核酸的3'末端互补的核酸序列,称为G序列的纯化结合序列和生物素部分。生物素部分经由可切割接头连接至G序列。首先使四种捕获探针与靶核酸杂交。所有探针都沿着靶核酸在不重叠的位置处杂交,其中探针B和探针C在探针A和探针D之间杂交。然后使用链霉抗生物素蛋白顺磁

珠纯化靶核酸,所述链霉抗生物素蛋白顺磁珠与捕获探针上的生物素部分结合。从珠中洗掉过量的非靶基因组DNA。然后通过切割每种捕获探针内的可切割接头,从链霉抗生物素蛋白顺磁珠中释放靶核酸-捕获探针复合物。使用与探针D上的纯化G序列结合的顺磁珠,进一步纯化靶核酸-捕获探针复合物。洗掉过量的捕获探针,并且从顺磁珠中洗脱靶核酸-捕获探针复合物。

[0318] 可以使用本公开内容的“用 λ 核酸外切酶的基于单珠的步骤纯化”系统来捕获靶核酸。存在四种捕获探针:探针A、探针B、探针C和探针D。探针A包含与靶核酸序列的5'末端互补的序列。探针A的5'末端包含两个硫代磷酸酯键。探针B、探针C和探针D包含在探针的3'末端处附着至生物素部分的核酸序列,以及在探针的5'末端处与靶核酸互补的核酸序列。探针B、探针C和探针D的5'末端是磷酸化的。探针A、探针B、探针C和探针D沿着靶核酸在不重叠的位置处杂交。在探针与靶核酸杂交后,使用链霉抗生物素蛋白顺磁珠纯化靶核酸。洗掉过量的gDNA和捕获探针。从珠中洗脱靶核酸-捕获探针复合物。然后使用 λ 核酸外切酶消化探针B、探针C和探针D,所述 λ 核酸外切酶优先降解在5'末端处磷酸化的双链DNA。

[0319] 可以使用本公开内容的“用FEN1的基于单珠的步骤纯化”系统来捕获靶核酸。存在四种捕获探针:探针A、探针B、探针C和探针D。探针A包含不与靶核酸杂交的3'核酸序列,与核酸的5'末端互补的核酸序列,以及不与靶核酸杂交且包含生物素部分的5'核酸序列。探针B和探针C包含不与靶核酸杂交的3'核酸序列,与靶核酸互补的核酸序列,以及不与靶核酸杂交且包含生物素部分的5'核酸序列。探针D包含不与靶核酸杂交的3'序列以及与靶核酸互补的5'序列。探针A、探针B、探针C和探针D与靶核酸杂交,使得探针A与探针B相邻,使得探针A上不与靶核酸序列杂交且包含生物素部分的5'核酸序列、以及探针B上不与靶核酸杂交的3'核酸序列,形成具有5' DNA瓣(flap)的分支双链DNA底物,并且探针B与探针C相邻,使得探针B上不与靶核酸序列杂交且包含生物素部分的5'核酸序列、以及探针C上不与靶核酸杂交的3'核酸序列,形成具有5' DNA瓣的分支双链DNA底物,并且探针C与探针D相邻,使得探针C上不与靶核酸序列杂交且包含生物素部分的5'核酸序列、以及探针D上不与靶核酸杂交的3'核酸序列,形成具有5' DNA瓣的分支双链DNA底物。在探针与靶核酸序列杂交后,使用链霉抗生物素蛋白顺磁珠纯化靶核酸。从珠中洗掉过量的基因组DNA和过量的探针。通过与热稳定的瓣状核酸内切酶1(FEN1)一起温育,从珠中洗脱靶核酸。FEN1切割5' DNA瓣,从而使生物素部分与杂交的捕获探针分开,释放靶核酸-捕获探针复合物。

[0320] 本公开内容还允许用户捕获且同时测序多种靶核酸,多种捕获探针可以与靶核酸的混合样品杂交。多种靶核酸可以包括其中每种核酸含有相同序列的一组多于一种的核酸,或者其中每种核酸不一定含有相同序列的一组多于一种的核酸。同样地,多种捕获探针可以包括在序列中等同的一组多于一种的捕获探针,或者在序列中不一定等同的一组多于一种的捕获探针。例如,使用全部含有相同序列的多种捕获探针,可以允许用户捕获全部含有相同序列的多种靶核酸。通过对含有相同序列的这种多种靶核酸进行测序,由于数据冗余,可以实现更高水平的测序准确率。在另一个实例中,可以使用包括与每种目的基因互补的捕获探针的一组捕获探针,同时捕获且测序两种或更多种特异性目的基因。这允许用户执行特异性基因的多路复用测序。图9显示了来自使用FFPE样品,使用本方法捕获且检测由100种靶组成的多路复用癌症组的实验的结果。

[0321] 捕获探针还可以包含与“多路复用寡核苷酸”结合(例如杂交)的结构域。多路复用

寡核苷酸可以包含至少三个结构域。第一结构域可以包含与捕获探针杂交的核酸序列。第二结构域可以包含鉴定样品的独特核酸序列。第三结构域可以包含基底结合部分。多种多路复用寡核苷酸可以与本公开内容的捕获探针组合使用,以同时测序来自至少两个样品的多种靶核酸。多路复用寡核苷酸可以用于同时测序来自至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个、至少十个、至少100个或至少1000个样品的多种靶核酸。

[0322] 使用多路复用寡核苷酸同时测序来自三个样品的三种靶核酸分子的实例如下:来自三个样品(样品1、样品2和样品3)各自的靶核酸与两种捕获探针:探针A和探针B杂交。探针A包含两个结构域。第一结构域包含基底结合部分。第二结构域包含与靶核酸的5'末端互补的序列。探针B包含两个结构域。第一结构域包含与靶核酸的3'末端互补的序列。第二结构域包含与多路复用寡核苷酸互补的序列。在两种捕获探针与靶核酸杂交后,探针B的第二结构域与多路复用寡核苷酸杂交。多路复用寡核苷酸包含三个结构域。第一结构域包含与探针B的第二结构域互补的序列。第二结构域包含鉴定样品的独特核酸序列。第三结构域包含基底结合部分。在多路复用寡核苷酸杂交后,执行核酸内切酶切割步骤,以去除靶核酸上的任何突出DNA,使得探针A与靶核酸的5'末端杂交,而探针B与靶核酸的3'末端杂交。在核酸内切酶处理后,将多路复用寡核苷酸连接至靶核酸的3'末端,然后去除探针B。将靶核酸-探针A复合物进一步纯化并随后进行测序。由于来自每个样品的每种靶核酸连接至多路复用寡核苷酸,可以通过对多路复用寡核苷酸进行测序,来鉴定靶核酸由其衍生的样品。

[0323] 当需要完全的测序覆盖时,所需的不同的捕获探针的数目与靶核酸片段的大小成反比。换言之,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(例如福尔马林固定的石蜡包埋的组织),包括多重捕获探针池可以是有用的。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核酸,在5'末端处的单个捕获探针可以是足够的。

[0324] 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靶核酸区域,在本文中称为“测序窗口”。在图8中标记了当两种捕获探针用于捕获靶核酸时产生的测序窗口。测序窗口是靶核酸的一部分,其可用于由测序探针结合。最小测序窗口是靶结合结构域长度(例如4至10个核苷酸),而最大测序窗口是整条染色体的大部分。

[0325] 当使用本方法对大的靶核酸分子进行测序时,“阻断物寡核苷酸”或多种阻断物寡核苷酸可以沿着靶核酸的长度进行杂交,以控制测序窗口的大小。阻断物寡核苷酸在特异性位置处与靶核酸杂交,从而阻止测序探针在那些位置处的结合,产生了较小的目的测序窗口。通过产生较小的测序窗口,测序反应局限于靶DNA分子上的特异性目的区域,增加了测序的速度和准确率。当对靶核酸内的已知位置处的特定突变进行测序时,阻断物寡核苷酸的使用是特别有用的,因为无需对整个靶核酸进行测序。在非限制性实例中,本公开内容的方法可以用于两个杂合位点的靶向测序,以区分两种不同的单倍型。

[0326] 捕获探针可以包含核酸分子复合物。核酸分子复合物可以包含部分双链的核酸分子。在一些方面,部分双链的核酸分子可以包含靶特异性结构域、双链体结构域、单链纯化序列、可切割部分、单链突出端结构域、样品特异性结构域、基底特异性结构域或其任何组合。

[0327] 在一些方面,部分双链的核酸分子的任何一条链可以包含约40至约150个核苷酸、

或约60至约135个核苷酸、或约10至约90个核苷酸、或约25至约75个核苷酸、约60个核苷酸、或约50至约100个核苷酸。

[0328] 在一些方面,部分双链的核酸分子的任何一条链可以包含至少一个、或至少两个、或至少三个、或至少四个、或至少五个、或至少六个、或至少七个、或至少八个、或至少九个、或至少十个亲和力部分。

[0329] 在一些方面,部分双链的核酸分子的任何一条链可以包含至少一个交联部分。交联部分可以是化学交联部分或光反应性交联部分。

[0330] 捕获探针可以包含单链核酸分子。在各方面,单链核酸分子可以包含靶特异性结构域、双链体结构域、单链纯化序列、可切割部分、单链突出端结构域、样品特异性结构域、基底特异性结构域或其任何组合。

[0331] 靶特异性结构域、双链体结构域、单链纯化序列、可切割部分、单链突出端结构域、样品特异性结构域或基底特异性结构域,可以包含至少一个天然碱基或不包含天然碱基。在一些方面,靶特异性结构域、双链体结构域、单链纯化序列、可切割部分、单链突出端结构域、样品特异性结构域或基底特异性结构域,可以包含至少一个修饰的核苷酸或核酸类似物或者不包含修饰的核苷酸。

[0332] 靶特异性结构域、双链体结构域、单链纯化序列、可切割部分、单链突出端结构域、样品特异性结构域或基底特异性结构域,可以包含天然碱基(例如,0、1、2、3、4、5、6、7、8、9、10个或更多个天然碱基)和修饰的核苷酸或核酸类似物(例如,0、1、2、3、4、5、6、7、8、9、10个或更多个修饰的)的任何组合。当以组合存在时,天然碱基和修饰的核苷酸或核酸类似物可以以任何次序排列。

[0333] 靶特异性结构域可以包含这样的核酸序列,其与靶核酸分子的一部分互补并且与靶核酸分子杂交。在一些方面,靶特异性结构域可以包含约10至约150个核苷酸、或约25至约100个核苷酸、或约35至约100个核苷酸、或约25至约125个核苷酸、或约15至约100个核苷酸。

[0334] 在一些方面,靶特异性结构域可以在靶核酸分子的3'末端的至少约100个碱基对内杂交。在一些方面,靶特异性结构域可以在靶核酸分子的5'末端的至少约100个内杂交。

[0335] 双链体结构域可以包含这样的核酸序列,其能够对另一条核酸链退火,以形成部分或完全双链的核酸分子。在一些方面,双链体结构域可以包含约14至约45个核苷酸、或约25至约35个核苷酸、或约30个核苷酸、或约10至约60个核苷酸、或约30至约50个核苷酸。

[0336] 单链纯化序列可以包含适合在纯化中使用的核酸序列。单链纯化序列可以包含F标签。单链纯化可以包含F样标签。单链纯化序列可以包含核苷酸序列AACATCACACAGACC (SEQ ID NO:112)。单链纯化序列可以包含核苷酸序列GTCTATCATCACAGC (SEQ ID NO:113)。

[0337] 单链纯化序列可以包含至少一个亲和力部分、或至少两个亲和力部分、或至少三个亲和力部分、或至少四个亲和力、或至少五个亲和力部分、或至少六个亲和力部分、或至少七个亲和力部分、或至少八个亲和力部分、或至少九个亲和力部分或至少十个亲和力部分。亲和力部分可以是生物素。因此,在一些方面,单链纯化序列可以包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个、至少七个、至少八个、至少九个或至少十个生物素部分。

[0338] 单链纯化序列可以包含至少50个核苷酸、或约15至约50个核苷酸。

[0339] 可切割部分可以包含酶促可切割部分。酶促可切割的可以包含用于通过USER酶切割的USER序列。可替代地,可切割部分可以包含光可切割部分。

[0340] 单链突出端结构域可以包含这样的单链核酸序列,其能够连同靶核酸分子一起形成5'突出的瓣结构。

[0341] 样品特异性结构域可以包含鉴定靶核酸分子由其获得的生物样品的核酸序列。样品特异性结构域可以包含L-DNA。样品特异性结构域可以包含D-DNA。样品特异性结构域可以包含L-DNA和D-DNA的组合。样品特异性结构域可以与本公开内容的任何探针杂交。样品特异性结构域可以包含约28个核苷酸。

[0342] 在一些方面,样品特异性结构域可以包含至少一个附着位置或至少两个附着位置。在其中样品特异性结构域包含至少一个附着位置或至少两个附着位置的方面,附着位置可以包含约14个核苷酸、或约10个核苷酸、或约8个核苷酸。

[0343] 基底特异性结构域可以包含与附着至基底的互补核酸分子杂交的核酸序列。基底可以是阵列。基底特异性结构域可以包含与草帽寡核苷酸杂交的核酸序列。

[0344] 基底特异性结构域可以包含多聚A序列。基底特异性结构域可以包含多聚T序列。基底特异性结构域可以包含L-多聚A序列,其中所述多聚 A序列的核苷酸是L-DNA。基底特异性结构域可以包含L-多聚T序列,其中所述多聚 T序列的核苷酸是L-DNA。基底特异性结构域可以包含L-DNA。基底特异性结构域可以包含约30个核苷酸。

[0345] 图34显示了示例性捕获探针的示意图,所述捕获探针包含与靶核酸结合的称为“c5探针复合物”的核酸分子复合物。c5探针复合物包含部分双链的核酸分子。部分双链的核酸分子的一条链包含与靶核酸杂交的靶特异性结构域,对部分双链的核酸分子的另一条链退火的双链体结构域,第一单链纯化序列,以及位于靶特异性结构域和双链体结构域之间的可切割部分。在该非限制性实例中,单链纯化序列包含F样标签,并且可切割部分包含酶促可切割的USER序列。部分双链的核酸分子的另一条链包含对部分双链的核酸分子的另一条链退火的双链体结构域,以及单链突出端结构域。在该非限制性实例中,单链突出端结构域和靶核酸分子形成5'突出的瓣结构。

[0346] 图34还显示了示例性捕获探针的示意图,所述捕获探针包含与靶核酸结合的称为“c3探针复合物”的核酸分子复合物。c3探针复合物包含部分双链的核酸分子。部分双链的核酸分子的一条链包含与靶核酸杂交的靶特异性结构域,对部分双链的核酸分子的另一条链退火的双链体结构域,以及位于靶特异性结构域和双链体结构域之间的可切割部分。在该非限制性实例中,可切割部分包含酶促可切割的USER序列。部分双链的核酸分子的另一条链包含对部分双链的核酸分子的另一条链退火的双链体结构域,样品特异性结构域,基底特异性结构域,单链纯化序列,以及位于单链纯化序列和基底特异性结构域之间的可切割部分。在该非限制性实例中,样品特异性结构域包含L-DNA,基底特异性结构域包含L-DNA,单链纯化序列包含F标签,并且可切割部分是光可切割部分。

[0347] 图41显示了示例性捕获探针的示意图,所述捕获探针包含与靶核酸分子结合的称为“c3.2探针复合物”的核酸分子复合物。c3.2探针复合物包含部分双链的核酸分子。部分双链的核酸分子的一条链包含与靶核酸杂交的靶特异性结构域,以及对部分双链的核酸分子的另一条链退火的双链体结构域。在一些方面,这条链可以任选地包含至少一个第一亲和力和部分。在一些方面,这条链可以任选地包含位于靶特异性结构域和双链体结构域之间

的可切割部分。部分双链的核酸分子的另一条链包含对部分双链的核酸分子的另一条链退火的双链体结构域,以及基底特异性结构域。在一些方面,这条链可以任选地包含至少一个、或至少两个、或至少三个第二亲和力部分。

[0348] 图41还显示了示例性捕获探针的示意图,所述捕获探针包含与靶核酸分子结合的称为“c5.2探针复合物”的核酸分子复合物。c5.2探针复合物包含部分双链的核酸分子。部分双链的核酸分子的一条链包含与靶核酸杂交的靶特异性结构域,对部分双链的核酸分子的另一条链退火的双链体结构域。在一些方面,这条链可以任选地包含位于靶特异性结构域和双链体结构域之间的可切割部分。部分双链的核酸分子的另一条链包含对部分双链的核酸分子的另一条链退火的双链体结构域,样品特异性结构域和第一单链纯化序列,位于双链体结构域和样品特异性结构域之间的第一可切割部分,以及位于样品特异性结构域和第一单链纯化序列之间的第二可切割部分。在一些方面,第一单链纯化序列可以包含至少一个亲和力部分,例如至少一个生物素部分。在一些方面,第一单链纯化序列可以替换为至少一个生物素部分,使得部分双链的核酸分子的另一条链包含对部分双链的核酸分子的另一条链退火的双链体结构域、样品特异性结构域、至少一个生物素部分,位于双链体结构域和样品特异性结构域之间的第一可切割部分、以及位于样品特异性结构域和至少一个生物素部分之间的第二可切割部分。

[0349] 本公开内容的样品制备方法

[0350] 本公开内容提供了样品制备的方法,其包括将靶核酸分子固定至基底。

[0351] 本发明的样品制备方法可以包括基于CRISPR的片段化步骤(参见例如,Baker和Mueller,“CRISPR-mediated isolation of specific megabase segments of genomic DNA”,*Nucleic Acids Research* 2017,45(19),e165;Tsai等人,“Amplification-free, CRISPR-Cas9 targeted enrichment and SMRT sequencing of repeat-expansion disease causative genomic regions”,*bioRxiv* 203919;doi:<https://doi.org/10.1101/203919>;Nachmanson等人,“Targeted genome fragmentation with CRISPR/Cas9 improves hybridization capture, reduces PCR bias, and enables efficient high-accuracy sequencing of small targets”,*bioRxiv* 207027;doi:<https://doi.org/10.1101/207027>)。CRISPR片段化可以包括通过对位于gDNA内的前间隔序列邻近基序(PAM)位点近端切割,从生物样品获得的体外片段化基因组DNA(gDNA)。PAM位点可以包含核苷酸序列NGG,其中N是任何核碱基。可替代地,PAM位点可以包含核苷酸序列NGA,其中N是任何核碱基。通过基于CRISPR的片段化产生的片段,可以使用生物素化的CRISPR复合物或用抗CAS9抗体进行纯化。

[0352] 用于捕获靶核酸的方法可以包括:(1)使用基于CRISPR的片段化步骤,使gDNA片段化;(2)使片段化的gDNA与至少两种捕获探针接触,其中所述至少两种捕获探针中的至少一种是如上所述的c5探针复合物,并且所述至少两种捕获探针中的至少一种是如上所述的c3探针复合物,使得c3探针复合物和c5探针复合物与靶核酸杂交,以形成图34中所示的复合物;(3)通过使组合物与FEN1接触而去除5'突出的瓣结构;(4)将靶核酸的3'末端与包含基底特异性结构域的c3探针复合物的链的5'末端连接;(5)使c5探针复合物的单链纯化序列与第一基底结合;(6)切割位于c3和c5探针复合物各自的双链体结构域和靶特异性结构域之间的可切割部分;(7)使c3探针复合物的单链纯化序列与第二基底结合;(8)切割位于连

接的c3探针复合物的单链纯化序列和基底特异性结构域之间的可切割部分；并且(9)使基底特异性结构域与附着至第三基底的互补核酸分子杂交。

[0353] 在前述方法的一些方面,步骤(9)可以在步骤(8)之前执行。

[0354] 在前述方法的一些方面,步骤(3)和(4)可以同时执行。在前述方法的一些方面,步骤(3)和(4)可以同时执行。

[0355] 在一些方面,前述方法可以任选地包括在步骤(6)和(7)之间的步骤,其中将衍生自不同生物样品的靶核酸-捕获探针复合物合并在一起。在这个方面,衍生自不同样品的靶核酸-捕获探针复合物将包括包含独特的样品特异性结构域的c3探针复合物,使得靶特异性结构域鉴定每种靶核酸由其获得的生物样品。

[0356] 图34-40中显示了本公开内容的样品制备方法的实例。在该非限制性实例中,首先使用基于CRISPR的片段化,使从生物样品获得的gDNA片段化。在片段化后,使靶核酸与两种捕获探针杂交,如图34中所示。在该非限制性实例中,两种捕获探针是如上所述的c3探针复合物和c5探针复合物。c3探针复合物和c5探针复合物沿着靶核酸在不重叠的位置处与靶核酸杂交。c3探针复合物在靶核酸3'末端的不多于8个核苷酸内,经由靶特异性结构域与靶核酸杂交,而c5探针复合物经由靶特异性结构域与靶核酸杂交,使得c5探针复合物对c3探针复合物的5'杂交。c5探针复合物的单链突出端结构域和靶核酸分子形成5'突出的瓣结构。在两种捕获探针杂交后,使靶核酸-捕获探针复合物与FEN1和连接酶一起温育。FEN1去除5'突出的瓣结构,并且靶核酸的3'末端通过连接酶连接到包含基底特异性结构域的c3探针复合物的链,如图35中所示。图36中所示的所得的复合物与F样珠结合,所述F样珠与c5探针复合物中存在的F样标签杂交。洗涤珠并加入USER酶。USER酶切割位于c3探针复合物和c5探针复合物两者的靶特异性结构域和双链体结构域之间的可切割部分,从而从F样珠中释放靶核酸,如图37中所示。如图38中所示的洗脱复合物使用SPRI珠进行进一步纯化。然后使纯化的复合物与F珠结合,所述F珠与c3探针复合物中存在的F标签杂交。在洗涤后,通过将珠暴露于UV光,从F珠中洗脱靶核酸,从而切割c3探针复合物中位于基底特异性结构域和F标签之间的光可切割部分,如图39中所示。然后通过使连接的c3探针复合物的基底特异性结构域与附着至基底的互补核酸杂交,使所得的复合物与基底结合,如图40中所示。

[0357] 图41-46中显示了本公开内容的另一种样品制备方法的实例。在该非限制性实例中,首先例如通过基于CRISPR的片段化,使从生物样品获得的gDNA片段化。在片段化后,使靶核酸与两种捕获探针杂交,如图41中所示。在该非限制性实例中,两种捕获探针是如上所述的c3.2探针复合物和c5.2探针复合物。c3.2探针复合物和c5.2探针复合物沿着靶核酸在不重叠的位置处与靶核酸杂交。c5.2探针复合物经由靶特异性结构域与靶核酸杂交,使得c5.2探针复合物对c3.2探针复合物的5'杂交。在两种捕获探针杂交后,使靶核酸与c3.2探针复合物的一条链和c5.2复合物的一条链连接,如图42中所示。连接可以包括酶促连接、自连接、化学连接或其任何组合。在包括酶促连接的方面,可以使用高保真度、模板指导的切口连接酶来执行酶促连接。然后可以使图42中所示的所得的复合物与包含至少一种寡核苷酸的珠结合,所述至少一种寡核苷酸与单链纯化序列杂交。可以洗涤珠,并且可以切割位于样品特异性结构域和单链纯化序列之间的可切割部分,从而从珠中释放靶核酸,如图43中所示。然后可以通过使基底特异性结构域与附着至基底的寡核苷酸杂交,将所得的复合物固定到基底上,如图44中所示。基底/寡核苷酸复合物可以是本公开内容的任何阵列。

[0358] 前述方法可以进一步包括使至少一种报告探针与样品特异性结构域杂交,其中所述报告探针包含第一可检测标记和第二可检测标记。然后可以鉴定第一可检测标记和第二可检测标记,从而基于第一可检测标记和第二可检测标记的身份,来鉴定靶核酸源于其的样品。

[0359] 可替代地,前述方法可以进一步包括使第一报告探针与样品特异性结构域杂交,其中所述报告探针包含第一可检测标记和第二可检测标记。然后可以鉴定第一可检测标记和第二可检测标记。然后可以去除第一可检测标记和第二可检测标记,并且可以使包含第三可检测标记和第四可检测标记的第二报告探针与样品特异性结构域杂交。然后,可以鉴定第三可检测标记和第四可检测标记,从而基于第一可检测标记、第二可检测标记、第三可检测标记和第四可检测标记的身份,来鉴定靶核酸源于其的样品。

[0360] 在鉴定靶核酸源于其的样品之后,可以切割位于样品特异性结构域和双链体结构域之间的可切割部分,如图45中所示,从而释放样品特异性结构域。

[0361] 本公开内容的方法

[0362] 本公开内容的测序方法包括使本文公开的至少一种测序探针与靶核酸可逆地杂交。

[0363] 用于对核酸进行测序的方法可以包括(1)使本文所述的测序探针与靶核酸杂交。靶核酸可以任选地在一个或多个位置处固定至基底。示例性的测序探针可以包含靶结合结构域和条形码结构域;其中所述靶结合结构域包含表1中所述的任何构建体。示例性的靶结合结构域包含与靶核酸杂交的至少八个核苷酸,其中所述靶结合结构域中的至少六个核苷酸可以鉴定靶核酸分子中的相应核苷酸(例如,这六个核苷酸鉴定它与之杂交的靶分子的互补六个核苷酸),并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸(例如,这至少两个核苷酸并不鉴定它与之杂交的靶分子的互补两个核苷酸);其中所述靶结合结构域中的至少六个核苷酸中的任何核苷酸可以是修饰的核苷酸或核苷酸类似物,并且其中所述靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何碱基、或者通用碱基或简并碱基。示例性的条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列能够由互补核酸分子结合,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中所述至少三个附着位置的每个位置的核酸序列,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0364] 在其它方面,示例性的靶结合结构域可以包含与靶核酸杂交的至少六个核苷酸,其中所述靶结合结构域中的至少六个核苷酸可以鉴定靶核酸分子中的相应核苷酸(例如,当靶结合结构域序列恰好是六个核苷酸时,这六个核苷酸鉴定它与之杂交的靶分子的互补六个核苷酸);其中所述靶结合结构域中的至少六个核苷酸无一或至少六个核苷酸中的任一个,可以是修饰的核苷酸或核苷酸类似物。

[0365] 在测序探针与靶核酸杂交之后,该方法包括(2)使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置

结合；(3) 检测结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记；(4) 鉴定固定的靶核酸中的至少两个核苷酸的位置和身份。例如，当第一互补核酸分子包含两种可检测标记时，两种可检测标记鉴定固定的靶核酸中的至少两个核苷酸。

[0366] 在检测到至少两种可检测标记之后，从第一互补核酸分子中去除至少两种可检测标记。因此，该方法进一步包括(5) 使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合，从而解除包含可检测标记的第一互补核酸分子的结合，或者使包含可检测标记的第一互补核酸分子，与足以释放第一可检测标记和至少第二可检测标记的力接触。因此，在步骤(5)之后，没有可检测标记与第一附着位置结合。该方法进一步包括(6) 使包含第三可检测标记和至少第四可检测的第二互补核酸分子，与条形码结构域的至少三个附着位置的第二附着位置结合；(7) 检测结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记；(8) 鉴定任选固定的靶核酸中的至少两个核苷酸的位置和身份；(9) 重复步骤(5)至(8)，直至条形码结构域中的至少三个附着位置的每个附着位置，已由包含两种可检测标记的互补核酸分子结合，并且已检测到结合的互补核酸分子的两种可检测标记，从而鉴定由测序探针的靶结合结构域杂交的、固定的靶核酸的至少第一区域的至少六个核苷酸的线性次序；并且(10) 从任选固定的靶核酸中去除测序探针。

[0367] 该方法可以进一步包括(11) 使第二测序探针与任选地在一个或多个位置处固定至基底的靶核酸杂交，并且其中第一测序探针和第二测序探针的靶结合结构域是不同的；(12) 使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子，与条形码结构域的至少三个附着位置的第一附着位置结合；(13) 检测结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记；(14) 鉴定任选固定的靶核酸中的至少两个核苷酸的位置和身份；(15) 使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合，从而解除包含可检测标记的第一互补核酸分子或复合物的结合，或者使包含可检测标记的第一互补核酸分子或复合物，与足以释放第一可检测标记和至少第二可检测标记的力接触；(16) 使包含第三可检测标记和至少第四可检测标记的第二互补核酸分子，与条形码结构域的至少三个附着位置的第二附着位置结合；(17) 检测结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记；(18) 鉴定固定的靶核酸中的至少两个核苷酸的位置和身份；(19) 重复步骤(15)至(18)，直至条形码结构域中的至少三个附着位置的每个附着位置，已由包含两种可检测标记的互补核酸分子结合，并且已检测到结合的互补核酸分子的两种可检测标记，从而鉴定由第二测序探针的靶结合结构域杂交的、固定的靶核酸的至少第二区域的至少六个核苷酸的线性次序；并且(20) 从任选固定的靶核酸中去除第二测序探针。

[0368] 该方法可以进一步包括组装固定的靶核酸的至少第一区域和至少第二区域中的每个鉴定的核苷酸线性次序，从而鉴定固定的靶核酸的序列。

[0369] 步骤(5)和(6)可以序贯或同时发生。第一可检测标记和至少第二可检测标记可以具有相同的发射光谱，或可以具有不同的发射光谱。第三可检测标记和至少第四可检测标记可以具有相同的发射光谱，或可以具有不同的发射光谱。

[0370] 第一互补核酸分子可以包含可切割接头。第二互补核酸分子可以包含可切割接头。第一互补核酸分子和第二互补核酸分子可以各自包含可切割接头。优选地，可切割接头是光可切割的。释放力可以是光。优选地，UV光。光可以由选自弧光灯、激光器、聚焦UV光源和发光二极管的光源提供。

[0371] 第一互补核酸分子和缺乏可检测标记的第一杂交核酸分子,可以包含相同的核酸序列。例如,缺乏可检测标记的第一杂交核酸分子,可以包含与第一互补核酸分子的一部分相同的核酸序列,所述部分与条形码结构域的至少三个附着位置的第一附着位置结合。缺乏可检测标记的第一杂交核酸分子,可以包含与侧翼单链多核苷酸互补的核酸序列,所述侧翼单链多核苷酸与条形码结构域中的第一附着位置相邻。

[0372] 第二互补核酸分子和缺乏可检测标记的第二杂交核酸分子,可以包含相同的核酸序列。缺乏可检测标记的第二杂交核酸分子,可以包含与侧翼单链多核苷酸互补的核酸序列,所述侧翼单链多核苷酸与条形码结构域中的第二附着位置相邻。

[0373] 本公开内容还提供了用于对核酸进行测序的方法,其包括(1)使本文所述的测序探针与靶核酸杂交。靶核酸可以任选地在一个或多个位置处固定至基底。示例性的测序探针可以包含靶结合结构域和条形码结构域;其中所述靶结合结构域包含表1中所述的任何构建体。示例性的靶结合结构域包含与靶核酸杂交的至少八个核苷酸,其中所述靶结合结构域中的至少六个核苷酸可以鉴定靶核酸分子中的相应核苷酸(例如,这六个核苷酸鉴定它与之杂交的靶分子的互补六个核苷酸),并且其中所述靶结合结构域中的至少两个核苷酸并不鉴定靶核酸分子中的相应核苷酸(例如,这至少两个核苷酸并不鉴定它与之杂交的靶分子的互补两个核苷酸);其中所述靶结合结构域中的至少六个核苷酸中的任何核苷酸可以是修饰的核苷酸或核苷酸类似物,并且其中所述靶结合结构域中并不鉴定靶核酸分子中的相应核苷酸的至少两个核苷酸,可以是对由靶结合域中的至少六个核苷酸决定的靶并非特异性的四种规范碱基中的任何碱基、或者通用碱基或简并碱基。示例性的条形码结构域包括合成主链,该条形码结构域包含至少三个附着位置,每个附着位置包括包含至少一个核酸序列的至少一个附着区域,所述至少一个核酸序列能够由互补核酸分子结合,其中所述至少三个附着位置的每个附着位置对应于靶结合结构域中的至少六个核苷酸的两个核苷酸,并且所述至少三个附着位置各自具有不同的核酸序列,并且其中所述至少三个附着位置的每个位置的核酸序列,确定靶核酸中由靶结合结构域结合的至少六个核苷酸的相应两个核苷酸的位置和身份。

[0374] 在其它方面,示例性的靶结合结构域可以包含与靶核酸杂交的至少六个核苷酸,其中所述靶结合结构域中的至少六个核苷酸可以鉴定靶核酸分子中的相应核苷酸(例如,当靶结合结构域序列恰好是六个核苷酸时,这六个核苷酸鉴定它与之杂交的靶分子的互补六个核苷酸);其中所述靶结合结构域中的至少六个核苷酸无一或至少六个核苷酸中的任一个,可以是修饰的核苷酸或核苷酸类似物。

[0375] 在测序探针与靶核酸杂交之后,该方法包括(2)使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置结合;(3)检测且记录结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记。

[0376] 在检测和记录至少两种可检测标记之后,从第一互补核酸分子中去除至少两种可检测标记。因此,该方法进一步包括(4)使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合,从而解除包含可检测标记的第一互补核酸分子的结合,或者使包含可检测标记的第一互补核酸分子,与足以释放第一可检测标记和至少第二可检测标记的力接触。因此,在步骤(4)之后,没有可检测标记与第一附着位置结合。该方法进一步包括(5)使包含第三可检测标记和至少第四可检测的第二互补核酸分子,与条形码结构域的至少三个附着位置

的第二附着位置结合；(6) 检测且记录结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记；(7) 重复步骤 (4) 至 (6)，直至条形码结构域中的至少三个附着位置的每个附着位置，已由包含两种可检测标记的互补核酸分子结合，并且已检测且记录结合的互补核酸分子的两种可检测标记；(8) 使用步骤 (3)、步骤 (6) 和步骤 (7) 中记录的可检测标记，鉴定与测序探针的靶结合结构域杂交的、固定的靶核酸的至少第一区域的至少六个核苷酸的位置和身份；并且 (9) 从任选固定的靶核酸中去除测序探针。

[0377] 该方法可以进一步包括 (10) 使第二测序探针与任选地在一个或多个位置处固定至基底的靶核酸杂交，并且其中第一测序探针和第二测序探针的靶结合结构域是不同的；(11) 使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子，与条形码结构域的至少三个附着位置的第一附着位置结合；(12) 检测且记录结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记；(13) 使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合，从而解除包含可检测标记的第一互补核酸分子或复合物的结合，或者使包含可检测标记的第一互补核酸分子或复合物，与足以释放第一可检测标记和至少第二可检测标记的力接触；(14) 使包含第三可检测标记和至少第四可检测标记的第二互补核酸分子，与条形码结构域的至少三个附着位置的第二附着位置结合；(15) 检测且记录结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记；(16) 重复步骤 (13) 至 (15)，直至条形码结构域中的至少三个附着位置的每个附着位置，已由包含两种可检测标记的互补核酸分子结合，并且已检测且记录结合的互补核酸分子的两种可检测标记；(17) 使用步骤 (12)、步骤 (15) 和步骤 (16) 中记录的可检测标记，鉴定由第二测序探针的靶结合结构域杂交的、固定的靶核酸的至少第二区域的至少六个核苷酸的位置和身份；并且 (18) 从任选固定的靶核酸中去除第二测序探针。

[0378] 该方法可以进一步包括组装固定的靶核酸的至少第一区域和至少第二区域中的每个鉴定的核苷酸线性次序，从而鉴定固定的靶核酸的序列。

[0379] 步骤 (4) 和 (5) 可以序贯或同时发生。第一可检测标记和至少第二可检测标记可以具有相同的发射光谱，或可以具有不同的发射光谱。第三可检测标记和至少第四可检测标记可以具有相同的发射光谱，或可以具有不同的发射光谱。

[0380] 第一互补核酸分子可以包含可切割接头。第二互补核酸分子可以包含可切割接头。第一互补核酸分子和第二互补核酸分子可以各自包含可切割接头。优选地，可切割接头是光可切割的。释放力可以是光。优选地，UV光。光可以由选自弧光灯、激光器、聚焦UV光源和发光二极管的光源提供。

[0381] 第一互补核酸分子和缺乏可检测标记的第一杂交核酸分子，可以包含相同的核酸序列。例如，缺乏可检测标记的第一杂交核酸分子，可以包含与第一互补核酸分子的一部分相同的核酸序列，所述部分与条形码结构域的至少三个附着位置的第一附着位置结合。缺乏可检测标记的第一杂交核酸分子，可以包含与侧翼单链多核苷酸互补的核酸序列，所述侧翼单链多核苷酸与条形码结构域中的第一附着位置相邻。

[0382] 第二互补核酸分子和缺乏可检测标记的第二杂交核酸分子，可以包含相同的核酸序列。缺乏可检测标记的第二杂交核酸分子，可以包含与侧翼单链多核苷酸互补的核酸序列，所述侧翼单链多核苷酸与条形码结构域中的第二附着位置相邻。

[0383] 前述方法可以进一步包括适合于记录可检测标记的介质。这种介质可以是合适的

计算机可读介质。

[0384] 本公开内容进一步提供了利用本文公开的多种测序探针,对核酸进行测序的方法。例如,靶核酸与多于一种测序探针杂交,并且每种探针可以对它与之杂交的靶核酸的一部分进行测序。

[0385] 本公开内容还提供了用于对核酸进行测序的方法,其包括(1)使包含本文所述的多种测序探针的第一测序探针的至少一个第一群体,与任选地在一个或多个位置处固定至基底的靶核酸杂交;(2)使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置结合;(3)检测结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记;(4)鉴定固定的靶核酸中的至少两个核苷酸的位置和身份;(5)使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合,从而解除包含可检测标记的第一互补核酸分子的结合,或者使包含可检测标记的第一互补核酸分子,与足以释放第一可检测标记和至少第二可检测标记的力接触;(6)使包含第三可检测标记和至少第四可检测的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置结合;(7)检测结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记;(8)鉴定任选固定的靶核酸中的至少两个核苷酸的位置和身份;(9)重复步骤(5)至(8),直至条形码结构域中的至少三个附着位置的每个附着位置,已由包含两种可检测标记的互补核酸分子结合,并且已检测到结合的互补核酸分子的两种可检测标记,从而鉴定由测序探针的靶结合结构域杂交的、固定的靶核酸的至少第一区域的至少六个核苷酸的线性次序;并且(10)从任选固定的靶核酸中去除第一测序探针的至少一个第一群体。

[0386] 该方法可以进一步包括(11)使包含本文公开的多种测序探针的第二测序探针的至少一个第二群体,与任选地在一个或多个位置处固定至基底的靶核酸杂交,并且其中第一测序探针和第二测序探针的靶结合结构域是不同的;(12)使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置结合;(13)检测结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记;(14)鉴定任选固定的靶核酸中的至少两个核苷酸的位置和身份;(15)使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合,从而解除包含可检测标记的第一互补核酸分子或复合物的结合,或者使包含可检测标记的第一互补核酸分子或复合物,与足以释放第一可检测标记和至少第二可检测标记的力接触;(16)使包含第三可检测标记和至少第四可检测标记的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置结合;(17)检测结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记;(18)鉴定固定的靶核酸中的至少两个核苷酸的位置和身份;(19)重复步骤(15)至(18),直至条形码结构域中的至少三个附着位置的每个附着位置,已由包含两种可检测标记的互补核酸分子结合,并且已检测到结合的互补核酸分子的两种可检测标记,从而鉴定由测序探针的靶结合结构域杂交的、固定的靶核酸的至少第二区域的至少六个核苷酸的线性次序;并且(20)从任选固定的靶核酸中去除第二测序探针的至少一个第二群体。

[0387] 该方法可以进一步包括组装固定的靶核酸的至少第一区域和至少第二区域中的每个鉴定的核苷酸线性次序,从而鉴定固定的靶核酸的序列。

[0388] 步骤(5)和(6)可以序贯或同时发生。第一可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。第三可检测标记和至少第四可检测标

记可以具有相同的发射光谱,或可以具有不同的发射光谱。

[0389] 第一互补核酸分子可以包含可切割接头。第二互补核酸分子可以包含可切割接头。第一互补核酸分子和第二互补核酸分子可以各自包含可切割接头。优选地,可切割接头是光可切割的。释放力可以是光。优选地,UV光。光可以由选自弧光灯、激光器、聚焦UV光源和发光二极管的光源提供。

[0390] 第一互补核酸分子和缺乏可检测标记的第一杂交核酸分子,可以包含相同的核酸序列。缺乏可检测标记的第一杂交核酸分子,可以包含与侧翼单链多核苷酸互补的核酸序列,所述侧翼单链多核苷酸与条形码结构域中的第一附着位置相邻。

[0391] 第二互补核酸分子和缺乏可检测标记的第二杂交核酸分子,可以包含相同的核酸序列。缺乏可检测标记的第二杂交核酸分子,可以包含与侧翼单链多核苷酸互补的核酸序列,所述侧翼单链多核苷酸与条形码结构域中的第二附着位置相邻。

[0392] 本公开内容还提供了用于对核酸进行测序的方法,其包括(1)使包含本文所述的多种测序探针的第一测序探针的至少一个第一群体,与任选地在一个或多个位置处固定至基底的靶核酸杂交;(2)使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置结合;(3)检测且记录结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记;(4)使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合,从而解除包含可检测标记的第一互补核酸分子的结合,或者使包含可检测标记的第一互补核酸分子,与足以释放第一可检测标记和至少第二可检测标记的力接触;(5)使包含第三可检测标记和至少第四可检测标记的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置结合;(6)检测且记录结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记;(7)重复步骤(4)至(6),直至条形码结构域中的至少三个附着位置的每个附着位置,已由包含两种可检测标记的互补核酸分子结合,并且已检测且记录结合的互补核酸分子的两种可检测标记;(8)使用步骤(3)、步骤(6)和步骤(7)中记录的可检测标记,鉴定由测序探针的靶结合结构域杂交的、固定的靶核酸的至少第一区域的至少六个核苷酸的位置和身份;并且(9)从任选固定的靶核酸中去除第一测序探针的至少一个第一群体。

[0393] 该方法可以进一步包括(10)使包含本文公开的多种测序探针的第二测序探针的至少一个第二群体,与任选地在一个或多个位置处固定至基底的靶核酸杂交,并且其中第一测序探针和第二测序探针的靶结合结构域是不同的;(11)使包含第一可检测标记和至少第二可检测标记的第一互补核酸分子,与条形码结构域的至少三个附着位置的第一附着位置结合;(12)检测且记录结合的第一互补核酸分子的第一可检测标记和至少第二可检测标记;(13)使缺乏可检测标记的第一杂交核酸分子与第一附着位置结合,从而解除包含可检测标记的第一互补核酸分子或复合物的结合,或者使包含可检测标记的第一互补核酸分子或复合物,与足以释放第一可检测标记和至少第二可检测标记的力接触;(14)使包含第三可检测标记和至少第四可检测标记的第二互补核酸分子,与条形码结构域的至少三个附着位置的第二附着位置结合;(15)检测且记录结合的第二互补核酸分子的第三可检测标记和至少第四可检测标记;(16)重复步骤(13)至(15),直至条形码结构域中的至少三个附着位置的每个附着位置,已由包含两种可检测标记的互补核酸分子结合,并且已检测且记录结合的互补核酸分子的两种可检测标记;(17)使用步骤(12)、步骤(15)和步骤(16)中记录的

可检测标记,鉴定由第二测序探针的靶结合结构域杂交的、固定的靶核酸的至少第二区域的至少六个核苷酸的位置和身份;并且(18)从任选固定的靶核酸中去除第二测序探针的至少一个第二群体。

[0394] 该方法可以进一步包括组装固定的靶核酸的至少第一区域和至少第二区域中的每个鉴定的核苷酸线性次序,从而鉴定固定的靶核酸的序列。

[0395] 步骤(4)和(5)可以序贯或同时发生。第一可检测标记和至少第二可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。第三可检测标记和至少第四可检测标记可以具有相同的发射光谱,或可以具有不同的发射光谱。

[0396] 第一互补核酸分子可以包含可切割接头。第二互补核酸分子可以包含可切割接头。第一互补核酸分子和第二互补核酸分子可以各自包含可切割接头。优选地,可切割接头是光可切割的。释放力可以是光。优选地,UV光。光可以由选自弧光灯、激光器、聚焦UV光源和发光二极管的光源提供。

[0397] 第一互补核酸分子和缺乏可检测标记的第一杂交核酸分子,可以包含相同的核酸序列。缺乏可检测标记的第一杂交核酸分子,可以包含与侧翼单链多核苷酸互补的核酸序列,所述侧翼单链多核苷酸与条形码结构域中的第一附着位置相邻。

[0398] 第二互补核酸分子和缺乏可检测标记的第二杂交核酸分子,可以包含相同的核酸序列。缺乏可检测标记的第二杂交核酸分子,可以包含与侧翼单链多核苷酸互补的核酸序列,所述侧翼单链多核苷酸与条形码结构域中的第二附着位置相邻。

[0399] 前述方法可以进一步包括适合于记录可检测标记的介质。这种介质可以是合适的计算机可读介质。

[0400] 在本文中进一步描述了测序方法。

[0401] 图10显示了本公开内容的单个示例性测序循环的示意性概图。尽管本方法不要求在测序之前固定靶核酸,但在该实例中,该方法始于已使用捕获探针捕获并结合至流动室表面的靶核酸,如最左上图中所示。然后使测序探针池流入流动室内,以允许测序探针与靶核酸杂交。在该实例中,测序探针是图1中描绘的那些。这些测序探针包含在靶结合结构域内与靶核酸杂交的6聚体序列。该6聚体在任一侧上侧翼为(N)个碱基,其可以是通用/简并碱基,或者由对通过碱基 b_1 - b_2 - b_3 - b_4 - b_5 - b_6 决定的靶并非特异性的四种规范碱基中的任何组成。使用6聚体序列,4096(4^6)种测序探针的集合使得能够对任何靶核酸进行测序。对于该实例,使4096种测序探针的集合在各512种测序探针的8个池中与靶核酸杂交。测序探针的靶结合结构域中的6聚体序列将沿着靶核酸的长度,在其中6聚体与靶核酸之间存在完全互补匹配的位置处杂交,如图10的中上图中所示。在该实例中,单一测序探针与靶核酸杂交。从流动室中洗出任何未结合的测序探针。

[0402] 这些测序探针还包含如上所述具有三个附着位置 R_1 、 R_2 和 R_3 的条形码结构域。在附着位置 R_1 内的附着区域包含一个或多个核苷酸序列,其对应于测序探针的6聚体的第一二核苷酸。因此,只有包含对应于测序探针的靶结合结构域中存在的第二二核苷酸身份的互补核酸的报告探针与附着位置 R_1 杂交。同样地,在测序探针的附着位置 R_2 内的附着区域对应于靶结合结构域中存在的第二二核苷酸,并且在测序探针的附着位置 R_3 内的附着区域对应于靶结合结构域中存在的第二二核苷酸。

[0403] 该方法以图10的最右上图继续。使报告探针池流入流动室内。报告探针池中的每

个报告探针包含以双色组合形式的可检测标记,以及可以与测序探针的附着位置 R_1 内的相应附着区域杂交的互补核酸。如上所述,特定报告探针的双色组合和互补核酸对应于16种可能的二核苷酸之一。这样设计每个报告探针池,使得在测序之前建立对应于特异性二核苷酸的双色组合。例如,在图10中所示的测序实验中,对于与附着位置 R_1 杂交的第一报告探针池,双色组合黄色-红色可以对应于二核苷酸腺嘌呤-胸腺嘧啶。在报告探针与附着位置 R_1 杂交后,如图10的右上图中所示,然后从流动室中洗出任何未结合的报告探针,并且记录结合的报告探针的可检测标记,以确定6聚体的第一二核苷酸的身份。

[0404] 去除归于与附着位置 R_1 杂交的报告探针的可检测标记。为了去除可检测标记,报告探针可以包括可切割接头,并且可以添加适当的切割试剂。可替代地,使缺乏可检测标记的互补核酸与测序探针的附着位置 R_1 杂交,并且置换具有可检测标记的报告探针。不管去除可检测标记的方法如何,附着位置 R_1 不再发出可检测信号。先前发出可检测信号的条形码结构域的附着位置通过其致使不再能够发出可检测信号的过程,在本文中称为“变暗”。

[0405] 使第二报告探针池流入流动室内。报告探针池中的每个报告探针包含以双色组合形式的可检测标记,以及可以与测序探针的附着位置 R_2 内的相应附着区域杂交的互补核酸。特定报告探针的双色组合和互补核酸对应于16种可能的二核苷酸之一。特定的双色组合在第一报告探针池的背景下对应于一种二核苷酸,且在第二报告探针池的背景下对应于不同的二核苷酸是可能的。在报告探针与附着位置 R_2 杂交后,如图10的右下图中所示,然后从流动室中洗出任何未结合的报告探针,并且记录可检测标记,以确定测序探针中存在的6聚体的第二二核苷酸的身份。

[0406] 为了去除在位置 R_2 处的可检测标记,报告探针可以包括可切割接头,并且可以添加适当的切割试剂。可替代地,使缺乏可检测标记的互补核酸与测序探针的附着位置 R_2 杂交,并且置换具有可检测标记的报告探针。不管去除可检测标记的方法如何,附着位置 R_2 不再发出可检测信号。

[0407] 然后使第三报告探针池流入流动室内。第三报告探针池中的每个报告探针包含以双色组合形式的可检测标记,以及可以与报告探针的附着位置 R_3 内的相应附着区域杂交的互补核酸。特定报告探针的双色组合和互补核酸对应于16种可能的二核苷酸之一。在报告探针与位置 R_3 杂交后,如图10的中下图中所示,然后从流动室中洗出任何未结合的报告探针,并且记录可检测标记,以确定测序探针中存在的6聚体的第三二核苷酸的身份。以这种方式,鉴定了靶结合结构域的所有三个二核苷酸,并且可以将其组装在一起,以揭示靶结合结构域的序列并因此揭示靶核酸的序列。

[0408] 为了继续对靶核酸进行测序,可以从靶核酸中去除任何结合的测序探针。即使报告探针仍与条形码结构域的位置 R_3 杂交,也可以从靶核酸中去除测序探针。可替代地,可以在从靶结合结构域去除测序探针之前,从条形码结构域去除与位置 R_3 杂交的报告探针,例如通过使用如上文对于在位置 R_1 和 R_2 处的报告物所述的变暗程序。

[0409] 图10中所示的测序循环可以重复任何次数,每个测序循环始于同一测序探针池与靶核酸分子的杂交、或不同测序探针池与靶核酸的杂交。第二测序探针池与靶核酸结合的位置与在第一测序循环期间第一测序探针或测序探针池在其处结合的位置重叠是可能的。由此,靶核酸内的某些核苷酸可以测序多于一次且使用多于一种测序探针。

[0410] 图11描绘了本公开内容的测序方法的一个完整循环的示意图,以及在该循环期间

收集的相应成像数据。在该实例中,使用的测序探针是图1中描绘的那些,并且测序步骤与图10中描绘的和上文描述的那些相同。在测序探针的测序结构域与靶核酸杂交之后,使报告探针与测序探针的第一附着位置(R_1)杂交。然后对第一报告探针成像以记录彩色点。在图11中,彩色点用虚线圆圈标记。彩色点对应于在完整循环期间记录的单一测序探针。在该实例中,记录了7种测序探针(1至7)。然后使条形码结构域的第一附着位置变暗,并且使双重荧光报告探针与测序探针的第二附着位置(R_2)杂交。然后对第二报告探针成像以记录彩色点。然后使条形码结构域的第二附着位置变暗,并且使双重荧光报告探针与测序探针的第三附着位置(R_3)杂交。然后对第三报告探针成像以记录彩色点。然后按次序排列来自每种测序探针1至7的三个彩色点。然后使用解码矩阵将每个色斑映射至特异性二核苷酸,以揭示测序探针1至7的靶结合结构域的序列。

[0411] 在单个测序循环期间,确定与靶核酸结合的任何测序探针的靶结合结构域序列所需的报告探针池的数目,等同于条形码结构域中的附着位置的数目。因此,对于具有三个位置的条形码结构域,使三个报告探针池循环经过测序探针。

[0412] 测序探针池可以包含在序列中全部等同的多种测序探针,或在序列中并不全部等同的多种测序探针。当测序探针池包括在序列中并不全部等同的多种测序探针时,每种不同的测序探针可以以相同的数目存在,或者不同的测序探针可以以不同的数目存在。

[0413] 图12显示了本公开内容的示例性的测序探针池配置,其中当测序探针含有以下时,上文指定的8种颜色组合用于设计8个不同的测序探针池:(a)具有与靶核酸特异性结合的6个核苷酸(6聚体)的靶结合结构域,以及(b)条形码结构域中的三个附着位置(R_1 、 R_2 和 R_3)。存在可能的4096个独特的6聚体序列($4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4096$)。鉴于条形码结构域中的三个附着位置各自可以与由8种不同颜色组合之一结合的互补核酸杂交,存在可能的3色组合的512个独特集合($8 \times 8 \times 8 = 512$)。例如,其中 R_1 与颜色组合GG结合的互补核酸杂交、 R_2 与颜色组合BG结合的互补核酸杂交,并且 R_3 与颜色组合YR结合的互补核酸杂交的探针,3色组合的集合相应地为GG-BG-YR。在测序探针池内,3色组合的每个独特集合对应于靶结合结构域内的独特6聚体。鉴于每个池含有512种独特的6聚体,并且存在总共4096种可能的6聚体,需要8个池来对所有可能的6聚体进行测序($4096/512=8$)。确定置于8个池各自中的特异性测序探针,以确保每种测序探针与靶核酸的最佳杂交。为了确保最佳杂交,采取包括以下的几项预防措施:(a)将完全的6聚体互补体分到不同的池内;(b)将具有高 T_m 和低 T_m 的6聚体分到不同的池内;并且(c)基于凭经验得知的杂交模式,将6聚体分到不同的池内。

[0414] 图13显示了在美国专利公开号20160194701中描述的测序探针与本公开内容的测序探针之间的差异。如图13的左图上所示,美国专利公开号20160194701描述了具有条形码结构域的测序探针,所述条形码结构域包含与互补核酸杂交的六个附着位置。每种互补核酸与四种不同的荧光染料之一结合。在该配置中,每种颜色(红色、蓝色、绿色、黄色)对应于靶结合结构域中的一个核苷酸(A、T、C或G)。这种探针设计产生了4096种独特探针(4^6)。如图13的右图中描绘的,在本公开内容的一个实例中,每种测序探针的条形码结构域包含与互补核酸杂交的3个附着位置,如图13的右图中描绘的。与美国专利公开号20160194701不同,这些互补核酸由8种不同颜色组合(GG、RR、GY、RY、YY、RG、BB和RB)之一结合。每种颜色组合对应于靶结合结构域中的特异性二核苷酸。这种配置产生了512种独特探针(8^3)。为了覆盖靶结合结构域内的所有可能的六聚体组合(4096),需要这512种独特探针的8个分开

池,以对整个靶核酸进行测序。由于使用8种颜色组合来标记互补核酸,但存在16种可能的二核苷酸,某些颜色组合将对应于不同的二核苷酸,这取决于待使用哪个测序探针池。例如,在图13中,在第1、第2、第3和第4测序探针池中,颜色组合BB对应于二核苷酸AA,而颜色组合GG对应于二核苷酸AT。在第5、第6、第7和第8测序探针池中,颜色组合BB对应于二核苷酸CA,而颜色组合CT对应于二核苷酸AT。

[0415] 多种测序探针(即,多于一种测序探针)可以在测序窗口内杂交。在测序期间,记录与多种杂交的测序探针中的每种测序探针结合的可检测标记的身份和空间位置。这允许随后鉴定多个二核苷酸的位置和身份两者。换言之,通过使多种测序探针同时与单个靶核酸分子杂交,可以同时沿着靶核酸的多重位置进行测序,提高了测序速度。

[0416] 在一些方面,单一测序探针可以与捕获的靶核酸分子杂交。在一些方面,多种测序探针可以与捕获的靶核酸分子杂交。在两个杂交的5'和3'捕获探针之间的测序窗口,可以允许单一测序探针或多种测序探针沿着靶核酸分子的长度的杂交。通过沿着靶核酸分子的长度杂交多种测序探针,靶核酸分子上的多于一个位置可以同时测序,提高了测序速度。可以在空间上分辨来自沿着靶核酸的长度结合的多种探针的个别探针的荧光信号。

[0417] 在一些方面,测序探针可以沿着靶核酸的长度以均匀的间隔结合。在一些方面,测序探针无需沿着靶核酸的长度以均匀的间隔结合。可以在空间上分辨来自沿着靶核酸的长度结合的多种测序探针的信号,以同时获得在靶核酸的多重位置处的测序信息。

[0418] 探针沿着靶核酸的长度的分布对于可检测信号的分辨是至关重要的。有时某个区域中的探针太多可以导致其可检测标记的重叠,从而阻止两种附近探针的分辨。这解释如下。鉴于一个核苷酸长度为0.34 nm,且鉴于测序装置的横向(x-y)空间分辨率为约200nm,测序装置的分辨率极限为约588个碱基对(即,1个核苷酸/0.34nm x 200nm)。也就是说,当两种探针在彼此约588个碱基对内时,上文提到的测序装置将不能分辨来自与靶核酸杂交的两种探针的信号。因此,取决于测序装置的分辨率,两种探针将需要间隔开大约600bp,之后其可检测标记可以被分辨为不同的“位点”。因此,在最佳间隔下,每600bp靶核酸应该存在单一探针。优选地,探针群体中的每种测序探针将彼此不靠近于600个核苷酸结合。各种软件方法(例如利用荧光强度值和波长依赖性比率),可以用于监测、限制且潜在地解卷积在靶核酸的可分辨区域内杂交的探针数目,并且相应地设计探针群体。此外,可以选择提供更多离散信号的可检测标记(例如荧光标记)。此外,文献(例如,Small和Parthasarthy:“Superresolution localization methods.” *Annu. Rev. Phys Chem.*, 2014;65:107-25)中的方法,描述了结构化照明和各种超分辨率方法,其将测序显微镜的分辨率极限降低直至10纳米。使用更高分辨率的测序装置允许使用具有更短靶结合结构域的探针。

[0419] 如上文提到的,探针 T_m 的设计可以影响与靶核酸杂交的探针数目。可替代地或另外地,可以增加群体中的测序探针浓度,以增加靶核酸的特异性区域中的探针覆盖率。可以减少测序探针的浓度,以降低靶核酸的特异性区域中的探针覆盖率,例如高于测序装置的分辨率极限。

[0420] 虽然关于两种可检测标记的分辨率极限是大约600个核苷酸,但这并不妨碍本公开内容的强大的测序方法。在某些方面,任何群体中的多种测序探针不被靶核酸上的600个核苷酸分开。然而,在统计学上(遵循泊松分布),存在仅具有与其结合的一种测序探针的靶核酸,并且该测序探针是光学可分辨的那种。对于具有在600个核苷酸内结合的多重探针

(且因此不是光学可分辨的)的靶核酸,关于这些不可分辨的测序探针的数据可能被丢弃。重要的是,本公开内容的方法提供了多轮结合并检测多种测序探针。因此,可能在几轮中检测到来自所有测序探针的信号,在几轮中仅检测到来自一部分测序探针的信号,并且在几轮中未检测到来自任何测序探针的信号。在一些方面,可以操纵(例如,通过控制浓度或稀释)与靶核酸结合的测序探针的分布,使得每个靶核酸仅结合一种测序探针。

[0421] 随机地,但部分地取决于靶结合结构域的长度、探针的 T_m 和所施加的探针浓度,群体中的两种不同的测序探针在彼此600个核苷酸内结合是可能的。

[0422] 可替代地或另外地,可以减少群体中的测序探针浓度,以降低靶核酸的特异性区域中的探针覆盖率,例如高于测序装置的分辨率极限,从而从分辨率受限的位点产生单个读数。

[0423] 如果在使用本公开内容的方法对靶核酸进行测序之前,已知靶核酸的序列或序列的部分,则可以这样设计且选择测序探针,使得没有两种测序探针在彼此600个核苷酸内与靶核酸结合。

[0424] 在使测序探针与靶核酸杂交之前,一个或多个互补核酸分子可以由第一可检测标记结合,并且至少第二可检测标记可以与测序探针的条形码结构域内的一个或多个附着位置杂交。例如,在与靶核酸杂交之前,可以使由第一可检测标记和至少第二可检测标记结合的一个或多个互补核酸分子,与每种测序探针的第一附着位置杂交。因此,当与其靶核酸接触时,测序探针能够从第一附着位置发出可检测的信号,并且不需要提供指向条形码结构域上的第一位置的互补核酸或报告探针的第一池。在另一个实例中,可以使由第一可检测标记和至少第二可检测标记结合的一个或多个互补核酸分子,与测序探针的条形码结构域内的所有附着位置杂交。因此,在该实例中,可以读取六个核苷酸序列,而无需序贯地替换互补核酸。由于省略了所述方法的许多步骤,因此使用这种预杂交的测序探针-报告探针复合物将减少获得序列信息的时间。然而,这种探针将获益于非重叠的可检测标记,例如荧光团被非重叠波长的光激发或荧光团发出非重叠波长的光。

[0425] 在本公开内容的方法的一些方面,来自记录的彩色点的信号强度可以用于更准确地对靶核酸进行测序。在一些方面,彩色点内的特定颜色的位点强度,可以用于确定特异性彩色点对应于其为一种颜色的重复(即,BB、GG、YY或RR)的颜色组合的概率。

[0426] 条形码结构域内的位置变暗可以通过在与该位置杂交的报告探针内存在的可切割接头修饰处的链切割来实现。图14描绘了在测序循环期间,使用可切割接头修饰使条形码位置变暗。在图14的最左侧图上描绘的第一步,包括使报告探针的一级核酸与测序探针的第一附着位置杂交。一级核酸与条形码结构域的第一位置的附着区域内的特异性互补序列杂交。一级核酸的第一结构域和第二结构域通过可切割接头修饰共价连接。在第二步中,然后记录可检测标记,以确定测序探针的靶结合结构域中的特异性二核苷酸的身份和位置。在第三步中,通过在可切割接头修饰处切割报告探针,使条形码结构域的第一位置变暗。这释放一级核酸的第二结构域,从而释放可检测标记。使现在缺乏任何可检测标记的一级核酸分子的第一结构域与条形码结构域的第一附着位置保持杂交,从而条形码结构域的第一位置不再发出可检测信号,并且在随后的测序步骤中将不能与任何其它报告探针杂交。在图14最右侧图中描绘的最后一步中,使报告探针与条形码结构域的第二位置杂交,以继续测序。

[0427] 条形码结构域的附着位置可以通过以下而变暗:置换报告探针中由可检测标记结合的任何二级核酸或三级核酸,同时仍允许报告探针的一级核酸分子保持与测序探针杂交。这种置换可以通过与未由可检测标记结合的一级核酸、二级核酸或三级核酸杂交来实现。图15是本公开内容的示例性测序循环的说明性实例,其中条形码结构域内的位置通过标记的二级核酸的置换而变暗。图15的最左侧图描绘了测序循环的开始,其中报告探针的一级核酸分子与测序探针的条形码结构域的第一附着位置杂交。然后使与可检测标记结合的二级核酸分子与一级核酸分子杂交,并记录可检测标记。为了使条形码结构域的第一位置变暗,与可检测标记结合的二级核酸分子由缺乏可检测标记的二级核酸分子置换。在测序循环的下一步中,使包含可检测标记的报告探针与条形码结构域的第二位置杂交。

[0428] 通过在核酸未由可检测标记结合的相应条形码结构域附着位置处与测序探针杂交,置换报告探针的任何一级核酸分子,可以使条形码结构域的附着位置变暗。在其中条形码结构域包含邻近或侧接至少一个附着位置的至少一个单链核酸序列的情况下,通过与侧翼序列以及由一级核酸分子占据的条形码结构域的一部分杂交,未由可检测标记结合的核酸可以置换一级核酸分子。需要时,通过掺入加速可检测标记的交换速率的小单链寡核苷酸(例如“立足点”探针;参见例如,Seeling等人,“Catalyzed Relaxation of a Metastable DNA Fuel”;*J. Am. Chem. Soc.* 2006,128(37),第12211-12220页),可以加速可检测标记交换的速率。

[0429] 可以从附着区域去除包含可检测标记的互补核酸或报告探针,而不是替换为缺乏可检测标记的杂交核酸。例如,这可以通过添加离液剂、升高温度、改变盐浓度、调整pH和/或施加水动力来发生。在这些实例中,需要更少的试剂(即,缺乏可检测标记的杂交核酸)。

[0430] 本公开内容的方法可以用于同时捕获且测序来自同一样品的RNA和DNA分子,包括mRNA和gDNA。可以在同一流动室中执行来自同一样品的RNA和DNA分子两者的捕获和测序。在一些方面,本公开内容的方法可以用于同时捕获、检测且测序来自FFPE样品的gDNA和mRNA两者。

[0431] 本公开内容的测序方法进一步包括以下步骤:对于固定的靶核酸的每个区域组装每个鉴定的核苷酸线性次序,从而鉴定固定的靶核酸的序列。组装步骤使用非暂时性计算机可读存储介质,在其上存储有可执行程序。程序指示微处理器对于靶核酸的每个区域排列每个鉴定的核苷酸线性次序,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后或在完成数据采集之后。

[0432] 本公开内容的测序方法的原始特异性为大约94%。通过用多于一种测序探针对靶核酸中的相同碱基进行测序,可以使本公开内容的测序方法的准确率增加至大约99%。图16描绘了本公开内容的测序方法如何允许用不同的测序探针对靶核酸的相同碱基进行测序。在该实例中的靶核酸是NRAS外显子2(SEQ ID NO:1)的片段。特定的目的碱基是在靶核酸中突出显示的胞嘧啶(C)。目的碱基将与两种不同的测序探针杂交,每种探针具有与靶核酸杂交的独特足迹。在该实例中,测序探针1至4(条形码1至4)结合目的碱基左侧的3个核苷酸,而测序探针5至8(条形码5至8)结合目的碱基左侧的5个核苷酸。因此,目的碱基将通过两种不同的探针进行测序,从而增加对于该特异性位置的碱基识别量,且从而增加在该特异性位置处的总体准确率。图17显示了可以如何组合从一种或多种测序探针记录的、靶核苷酸上的特异性核苷酸位置的多重不同的碱基识别,以形成一致序列(SEQ ID NO:2),从而增加

最终碱基识别的准确率。

[0433] 术语“Hyb & Seq化学”、“Hyb & Seq测序”和“Hyb & Seq”指上述本公开内容的方法。

[0434] 本公开内容的阵列和使用所述阵列的方法

[0435] 本公开内容提供了用于固定核酸分子的组合物和方法,包括阵列和使用阵列的方法,如本文详细描述。

[0436] 本公开内容提供了包含以下的组合物:平面固体支撑基底;在平面固体支撑基底上的第一层;在第一层上的第二层;其中所述第二层包含多个纳米孔,其中每个纳米孔提供了对第一层的暴露部分的接近,其中每个纳米孔包含共价附着到第一层的暴露部分的多个第一寡核苷酸。

[0437] 本公开内容提供了包含以下的组合物:平面固体支撑基底;与平面固体支撑基底的第一表面接触,在平面固体支撑基底上的第一层;与第一层的第二表面接触的在第一层上的第二层,其中所述第一层的第二表面不与平面固体支撑基底的表面接触;其中所述第二层包含多个纳米孔,其中每个纳米孔提供了对第一层的暴露部分的接近,其中每个纳米孔包含共价附着到第一层的暴露部分的多个第一寡核苷酸。

[0438] 第一层可以包括与平面固体支撑基底的表面接触的第一表面,以及与第二层接触但不与平面固体支撑基底的表面接触的第二表面。

[0439] 第二层可以包括与第一层的表面接触的第一表面,以及暴露于环境的第二表面。

[0440] 图47是本发明的示例性阵列的示意性横截面。该阵列包括平面固体支撑基底101,在平面固体支撑基底101上的第一层102,以及在第一层102上的第二层103。第二层103包括多个纳米孔104。每个纳米孔104在两侧上开口,从而暴露了每个纳米孔105中的第一层的一部分。多个第一寡核苷酸106共价附着到每个纳米孔中的暴露的第一层105。

[0441] 在一些方面,平面固体支撑基底可以是表面、膜、珠、多孔材料或电极。例如,平面固体支撑基底可以包含但不限于聚合物材料、金属、硅、玻璃或石英。

[0442] 在一些方面,第一层102可以包含氧化物膜,例如但不限于二氧化硅。

[0443] 在一些方面,第一层102可以具有约50至约150 nm的厚度。第一层102可以具有约90 nm的厚度。

[0444] 在一些方面,第二层103可以包含但不限于双(三甲基甲硅烷基)胺,也称为六甲基二硅氮烷(HMDS或HDMS)。

[0445] 在一些方面,第二层103可以包含并非化学反应性的材料,使得第二层不结合生物大分子。

[0446] 在一些方面,第二层103可以具有约1 nm至约10 nm的厚度。第二层103可以具有约3 nm至约4 nm的厚度。

[0447] 在一些方面,平面固体支撑基底包含硅,第一层包含二氧化硅,而第二层包含HMDS。

[0448] 在一些方面,平面固体支撑基底包含玻璃,第一层包含二氧化硅,而第二层包含HMDS。

[0449] 在一些方面,第二层可以包含约 0.1×10^5 至约 100×10^7 个纳米孔/平方毫米。第二层可以包含约 0.1×10^6 至约 100×10^6 个纳米孔/平方毫米。第二层可以包含约 1×10^6 至约 10×10^6 个纳米孔/平方毫米。第二层可以包含约 2×10^6 至约 5×10^6 个纳米孔/平方毫米。第二

层可以包含约 3×10^6 个纳米孔/平方毫米。

[0450] 如本文使用的,“纳米孔的密度”指在指定表面积内存在的纳米孔数目。例如,具有 1.0 mm^2 的表面积并且包括 1.0×10^6 个纳米孔的第二层,被说成具有 1.0×10^6 纳米孔/ mm^2 的纳米孔密度。

[0451] 在一些方面,纳米孔的密度可以为约 0.1×10^5 至约 100×10^7 个纳米孔/ mm^2 。纳米孔的密度可以为约 0.1×10^6 至约 100×10^6 个纳米孔/ mm^2 。纳米孔的密度可以为约 1×10^6 至约 10×10^6 个纳米孔/ mm^2 。纳米孔的密度可以为约 2×10^6 至约 5×10^6 个纳米孔/ mm^2 。纳米孔的密度可以为约 3×10^6 个纳米孔/ mm^2 。

[0452] 在一些方面,纳米孔中的第一层的暴露部分的表面积可以为约200至约50,000 nm^2 。每个纳米孔中的第一层的暴露部分的表面积可以为约300至约40,000 nm^2 。每个纳米孔中的第一层的暴露部分的表面积可以为约700至约8,000 nm^2 。每个纳米孔中的第一层的暴露部分的表面积可以为约2,000至约3,000 nm^2 。

[0453] 在一些方面,每个纳米孔中的第一层的暴露部分是圆形的。在一些方面,每个纳米孔中的第一层的暴露部分是椭圆形的。在一些方面,每个纳米孔中的第一层的暴露部分是矩形的。在一些方面,每个纳米孔中的第一层的暴露部分是正方形的。在一些方面,每个纳米孔中的第一层的暴露部分是六边形或八边形的。在一些方面,每个纳米孔中的第一层的暴露部分具有规则多边形的形状。在一些方面,每个纳米孔中的第一层的暴露部分具有不规则多边形的形状。

[0454] 在其中纳米孔中的第一层的暴露部分是圆形的一些方面,第一层的暴露部分可以具有约10 nm至约200 nm的直径。第一层的暴露部分可以具有约20 nm至约200 nm的直径。第一层的暴露部分可以具有约30 nm至约100 nm的直径。第一层的暴露部分可以具有约50 nm至约60 nm的直径。第一层的暴露部分可以具有约60 nm的直径。

[0455] 在一些方面,纳米孔可以是圆柱形的。纳米孔可以是矩形的。纳米孔可以是立方形的。纳米孔可以是多面体的。纳米孔可以具有直圆柱体的形状。纳米孔可以具有椭圆柱体的形状。纳米孔可以具有矩形棱柱的形状。纳米孔可以具有立方体的形状。纳米孔可以具有三棱柱的形状。纳米孔可以具有圆锥的形状。纳米孔可以具有棱锥的形状。纳米孔可以具有四角锥的形状。

[0456] 在一些方面,纳米孔的一部分可以具有与同一纳米孔的另一部分不同的尺寸。在图48中所示的非限制性实例中,本公开内容的阵列包括平面固体支撑基底201、在平面固体支撑基底201上的第一层202、以及在第一层202上的第二层203。第二层203包括多个纳米孔204。每个纳米孔204在两侧上开口,从而暴露了每个纳米孔205中的第一层的一部分。最靠近第一层202的纳米孔部分具有的直径206小于距第一层最远的纳米孔部分的直径207。

[0457] 在一些方面,多个纳米孔可以以随机模式排列。如本文使用的,术语“以随机模式排列”、“随机模式化”或“随机”指特征例如纳米孔或垫的非有序、非笛卡尔分布(换言之,未排列在沿着网格的x轴和y轴的预定点处,或者未排列在距径向模式中心的限定的“时钟位置”、度或半径处),其无法通过有意的设计(或通过其可以实现此类设计的程序)或特征的特异性放置来实现。可以通过将溶液、乳状液、气雾剂、蒸气或干制剂滴、喷、镀或散布到支撑物上,而无需将任何特征导向在其上的特定位点的任何方式,来实现特征的“随机模式化”集合。图49显示了本公开内容的示例性阵列的示意图解,所述阵列包括以随机模式排列

的多个圆柱形纳米孔301。

[0458] 在一些方面,多个纳米孔可以以有序模式排列。如本文使用的,术语“以有序模式排列”或“有序模式”指这样的特征分布,其沿着网格的x轴和y轴,沿着规则的预定点排列,或者在距径向模式中心的限定的“时钟位置”、度或半径处排列。在一些方面,多个纳米孔可以排列在有序网格中。有序网格可以具有约50 nm至约3500 nm的间距。有序网格可以具有约100 nm至约3200 nm的间距。有序网格可以具有约300 nm至1000 nm的间距。有序网格可以具有约440 nm至约710 nm的间距。有序网格具有约575 nm的间距。有序网格具有约577 nm的间距。如本文使用的,术语“间距”指排列在有序网格上的两个相邻特征的中心之间的距离。图50显示了本公开内容的示例性阵列的示意图解,所述阵列包括以恒定间距402排列在有序网格中的纳米孔401。

[0459] 附着至阵列和/或基底的寡核苷酸在本文中也可以称为草坪寡核苷酸。

[0460] 在一些方面,第一寡核苷酸的核酸长度可以是约10至约100个核苷酸。第一寡核苷酸的核酸长度可以是约20至约40个核苷酸。

[0461] 在一些方面,第一寡核苷酸可以包含至少一个天然碱基。第一寡核苷酸可以不包含天然碱基。第一寡核苷酸可以包含至少一个修饰的核苷酸或核酸类似物。第一寡核苷酸可以不包含修饰的核苷酸或核酸类似物。第一寡核苷酸可以包含至少一个通用碱基。第一寡核苷酸可以不包含通用碱基。第一寡核苷酸可以包含至少一个简并碱基。第一寡核苷酸可以不包含简并碱基。

[0462] 在一些方面,第一寡核苷酸可以包含RNA、D-DNA、L-DNA、LNA、异鸟嘌呤、异胞嘧啶、无碱基核苷酸或其任何组合。

[0463] 在一些方面,多个第一寡核苷酸中的每个第一寡核苷酸可以包含相同的序列。多个第一寡核苷酸中的每个第一寡核苷酸可以包含不同的序列。

[0464] 在一些方面,多个纳米孔中的每个纳米孔包含相同的多个第一寡核苷酸。在其它方面,多个纳米孔中的每个纳米孔包含不同的多个第一寡核苷酸。

[0465] 多个第一寡核苷酸可以包含2、3、4、5、6、7、8、9、10、50、100、200、300、400、500、600、700、800、900、1,000、10,000、100,000或1,000,000个第一寡核苷酸。

[0466] 可以使用本领域已知的方法,包括但不限于光刻法等,来制造本公开内容的阵列。用于制造阵列的方法在美国专利号7,250,371和US 2006/0134917中进行描述,所述专利的内容整体引入作为参考。

[0467] 在一些方面,每个纳米孔中的第一层的暴露部分可以用反应性部分官能化,以吸引并结合存在于生物大分子,例如第一寡核苷酸或靶核酸复合物上的特异性化学基团。众所周知,这些官能团能够通过各种缀合化学特异性吸引并结合生物大分子。示例性的反应性部分包括但不限于包含以下的表面:环氧基、醛、金、酰肼、巯基、NHS-酯、胺、炔、叠氮化物、硫醇、羧酸酯、马来酰亚胺、羟甲基膦、(3-氨基丙基)三甲氧基硅烷、亚氨酸酯、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。

[0468] 在一些方面,可以使用光反应性部分,将生物大分子包括但不限于核酸分子附着至每个纳米孔中的第一层的暴露部分。光反应性部分的一些实例包括芳基叠氮化物,例如N-((2-吡啶基二硫代)乙基)-4-叠氨基水杨酰胺;氟化芳基叠氮化物,例如4-叠氨基-2,3,

5,6-四氟苯甲酸;基于二苯甲酮的试剂,例如4-苯甲酰基苯甲酸的琥珀酰亚胺酯;和5-溴-脱氧尿苷。

[0469] 本公开内容提供了固定来自样品的至少一个靶核酸的方法,其包括:a)提供本公开内容的组合物;并且b)使组合物与至少一个靶核酸接触,其中所述至少一个靶核酸与第一寡核苷酸杂交,从而固定至少一个靶核酸,使得在单个纳米孔内固定不多于一个靶核酸。

[0470] 图51显示了图47中所示的示例性阵列的示意图解,其中一个靶核酸复合物501固定在每个纳米孔中。

[0471] 本公开内容的固定方法具有以下优点:阵列上的纳米孔的大小(例如体积)防止多于一个靶核酸复合物在单个纳米孔内杂交。因此,在固定后,阵列上的每个纳米孔含有一个或零个靶核酸复合物。这对于应用如单分子测序是有利的。通过仅将一个靶核酸复合物固定在单个纳米孔内,可以例如使用本领域已知的荧光探针,个别地查询各个靶核酸复合物(例如参见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、U.S. 2014/0162251和U.S. 2016/0194701,其各自整体引入本文作为参考)。

[0472] 图52显示了图47中所示的示例性阵列的示意图解,其中一个靶核酸复合物601固定在每个纳米孔中。靶核酸复合物602无法在最左边的纳米孔603中结合,因为纳米孔603已经被靶核酸复合物601占据。

[0473] 在一些方面,纳米孔可以具有的体积大约等于、等于或小于靶核酸的排除体积,使得仅一个靶核酸复合物可以在给定的时间在物理上适合单个纳米孔。

[0474] 在一些方面,纳米孔这样间隔,使得来自与固定在纳米孔内的生物大分子杂交/结合的荧光探针的信号,可与来自与固定在相邻纳米孔中的生物大分子杂交/结合的荧光探针的信号在光学上分辨。

[0475] 在一些方面,靶核酸包含约10个核苷酸至约100,000个核苷酸。在一些方面,靶核酸可以包含约100、或约250、或约500、或约750、或约1,000、或约5,000个核苷酸、或约10,000个核苷酸、或约100,000个核苷酸、或约1,000,000个核苷酸、或约10,000,000个核苷酸。

[0476] 在一些方面,靶核酸可以包括但不限于可以具有各种长度的聚合形式的核苷酸(脱氧核糖核苷酸或核糖核苷酸)或其类似物。靶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质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、分支多核苷酸、质粒、载体、序列的分离DNA、序列的分离RNA、核酸探针和引物。靶核酸的身份和/或序列可以是已知的。可替代地,靶核酸的身份和/或序列可以是未知的。靶核酸的一部分序列是已知的也是可能的。

[0477] 靶核酸可以得自核酸的任何样品或来源,例如任何细胞、组织或生物、体外、化学合成仪等。靶核酸可以通过任何领域公认的方法获得。靶核酸可以从临床受试者的血液样品获得。靶核酸可以从生物样品中分离。可以使用本领域众所周知的方法和试剂盒,从来源

或样品中提取、分离或纯化靶核酸。在一些方面,组织样品是活组织检查的肿瘤或其一部分,即临床上相关的组织样品。例如,肿瘤可以来自乳腺癌。样品可以是切除的淋巴结。

[0478] 靶核酸可以在固定之前通过本领域已知的任何手段进行片段化。优选地,通过酶促或机械手段执行片段化。机械手段可以是超声处理或物理剪切。可以通过用核酸酶(例如,脱氧核糖核酸酶I (DNase I))、或者一种或多种限制性核酸内切酶消化来执行酶促手段。靶核酸可以使用基于CRISPR的片段化步骤进行片段化。基于CRISPR的片段化步骤允许靶片段化和产生所需大小的靶核酸。在非限制性实例中,基于CRISPR的片段化可以产生这样的靶核酸,其具有的大小防止多于一个靶核酸复合物被固定在单个纳米孔中。

[0479] 在本公开内容的方法的一些方面,靶核酸可以经由第二寡核苷酸与第一寡核苷酸间接地杂交,其中所述第二寡核苷酸包含与第一寡核苷酸杂交的第一区域、以及与至少一个靶核酸杂交的第二区域。

[0480] 第二寡核苷酸可以包含至少一个天然碱基。第二寡核苷酸可以不包含天然碱基。第二寡核苷酸可以包含至少一个修饰的核苷酸或核酸类似物。第二寡核苷酸可以不包含修饰的核苷酸或核酸类似物。第二寡核苷酸可以包含至少一个通用碱基。第二寡核苷酸可以包含至少一个通用碱基。第二寡核苷酸可以不包含通用碱基。第二寡核苷酸可以包含至少一个简并碱基。第二寡核苷酸可以不包含简并碱基。

[0481] 第二寡核苷酸可以包含D-DNA、L-DNA、LNA、异鸟嘌呤、异胞嘧啶、无碱基核苷酸或其组合。第二寡核苷酸可以包含L-DNA。第二寡核苷酸可以包含鉴定样品的序列。

[0482] 在本公开内容的方法的一些方面,靶核酸可以经由第二寡核苷酸和第三寡核苷酸与第一寡核苷酸间接地杂交,其中所述第二寡核苷酸包含与第一寡核苷酸杂交的第一区域、以及与第三寡核苷酸上的第一区域杂交的第二区域,并且其中所述第三寡核苷酸包含与第二寡核苷酸的第二区域杂交的第一区域、以及与靶核酸杂交的第二区域。

[0483] 第三寡核苷酸可以包含至少一个天然碱基。第三寡核苷酸可以不包含天然碱基。第三寡核苷酸可以包含至少一个修饰的核苷酸或核酸类似物。第三寡核苷酸可以不包含修饰的核苷酸或核酸类似物。第三寡核苷酸可以包含至少一个通用碱基。第三寡核苷酸可以包含至少一个通用碱基。第三寡核苷酸可以不包含通用碱基。第三寡核苷酸可以包含至少一个简并碱基。第三寡核苷酸可以不包含简并碱基。

[0484] 第三寡核苷酸可以包含D-DNA、L-DNA、LNA、异鸟嘌呤、异胞嘧啶、无碱基核苷酸或其组合。第三寡核苷酸可以包含L-DNA。第三寡核苷酸可以包含鉴定样品的序列。

[0485] 在本公开内容的方法的一些方面,靶核酸可以与第一寡核苷酸直接杂交。靶核酸可以经由与靶核酸的3'末端连接的核酸探针,与第一寡核苷酸直接杂交。

[0486] 核酸探针可以包含至少一个天然碱基。核酸探针可以不包含天然碱基。核酸探针可以包含至少一个修饰的核苷酸或核酸类似物。核酸探针可以不包含修饰的核苷酸或核酸类似物。核酸探针可以包含至少一个通用碱基。核酸探针可以包含至少一个通用碱基。核酸探针可以不包含通用碱基。核酸探针可以包含至少一个简并碱基。核酸探针可以不包含简并碱基。

[0487] 靶核酸复合物的核酸探针可以包含RNA、D-DNA、L-DNA、LNA、异鸟嘌呤、异胞嘧啶、无碱基核苷酸或其任何组合。

[0488] 核酸探针可以包含可切割接头。可切割接头可以是光可切割接头或酶促可切割接

头。

[0489] 在本公开内容的方法的一些方面,在固定靶核酸之后,可以对靶核酸进行测序。可以使用本领域已知的任何方法对靶核酸进行测序。优选地,通过直接检测方法对靶核酸进行测序(例如,参见WO 2016/081740、WO 2018/094385,其各自整体引入本文作为参考)。

[0490] 如本文使用的,术语“阵列”以其最广泛的含义使用,以指包括多个特征的基底,其中特征包含一种或多种固定的生物大分子,和/或其中特征能够捕获且固定一种或多种生物大分子。因此,在一些方面,本发明的组合物可以被视为阵列。

[0491] 术语“特征”、“垫”、“位点”和“纳米孔”在本文中可互换使用,以指包含一种或多种固定的生物大分子、和/或能够捕获且固定一种或多种生物大分子的结构和/或区域。

[0492] 如本文使用的,术语“生物大分子”以其最广泛的含义使用,以指有机分子,例如但不限于碳水化合物、脂质、蛋白质、肽或核酸分子。

[0493] 如本文使用的,术语“固定的”指核酸分子与表面之间的键合。键合可以是非共价的。例如,可以通过与共价连接至表面的寡核苷酸杂交,将靶核酸固定至固体支撑基底。

[0494] 如本文使用的,“排除体积”指由特定分子占据以排除其它此类分子的空间的体积。

[0495] 任何上述方面都可以与本文公开的任何其它方面组合。

[0496] 定义

[0497] 如本文使用的,术语“退火”和“杂交”可互换使用,以意指稳定的双链体的形成。在一个方面,稳定的双链体意指双链体结构通过在以下条件下的严格洗涤不被破坏:例如比双链体的链的 T_m 低约 5°C 或高约 5°C 的温度,以及低单价盐浓度例如低于 0.2 M 、或低于 0.1 M 或本领域技术人员已知的盐浓度。当提及双链体使用时,术语“完全匹配的”意指构成双链体的多核苷酸和/或寡核苷酸链彼此形成双链结构,使得每条链中的每一个核苷酸经历与另一条链中的核苷酸的沃森-克里克碱基配对。术语“双链体”包含但不限于可以采用的核苷类似物(例如脱氧肌苷、具有2-氨基嘌呤碱基的核苷、PNA等等)的配对。双链体中的两个寡核苷酸之间的“错配”意指双链体中的一对核苷酸未能经历沃森-克里克键合。

[0498] 如本文使用的,术语“杂交条件”通常包括小于约 1 M ,更通常小于约 500 mM ,且甚至更通常小于约 200 mM 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下执行,所述严格条件例如在其下探针与其靶子序列特异性杂交的条件。严格条件是序列依赖性的,并且在不同情况下是不同的。更长的片段可能需要更高的杂交温度用于特异性杂交。由于其它因素可能影响杂交的严格性,包括碱基组成和互补链的长度、有机溶剂的存在以及碱基错配的程度,因此参数的组合比单独任何一种的绝对测量更重要。

[0499] 一般地,严格条件选择为比特异性序列在限定的离子强度和pH下的 T_m 低约 5°C 。示例性的严格条件包括在pH 7.0至8.3和至少 25°C 的温度下,至少 0.01 M 至不多于 1 M 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、 5 mM EDTA ,pH 7.4)和 $25-30^\circ\text{C}$ 温度的条件适合于等位基因特异性探针杂交。关于严格条件,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)。如本文使用的,术语“特异性与……杂交”

或“与……特异性杂交”或类似术语,指分子在严格条件下基本上与一种或多种特定核苷酸序列结合、双链体化或杂交。

[0500] 与探针的特定位置相关的可检测标记可以被“读出”(例如,检测到其荧光)一次或多次;“读出”可以与术语“碱基识别”同义。多重读取改善准确率。当检测到衍生自单个原始靶分子的邻接段的序列信息时,靶核酸序列被“读取”;通常,这经由多次通过一致(如下文定义)生成。如本文使用的,术语“覆盖率”或“覆盖深度”指靶区域已被测序(经由离散读取)并与参考序列比对的次数。读取覆盖率是映射到特异性参考靶序列的读取总数;碱基覆盖率是在特异性基因组位置处进行的碱基识别总数。

[0501] “读数”是测序仪输出的单位。邻接段的序列信息衍生自单个原始靶分子。每个读数具有与读数内的碱基识别的置信水平相关联的质量度量。测序仪输出的单位。邻接段的序列信息衍生自单个原始靶分子。在Hyb & Seq中,所有读数都经由多次通过一致生成。

[0502] “读取长度”是描述来自每个读数的序列长度(以bp为单位)的度量。该度量由测序技术确定。

[0503] 如本文使用的,“Hyb & Seq循环”指检测特定探针或探针群体上的每个附着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“Hyb & Seq循环”至少包括使探针与靶核酸杂交,使互补核酸/报告探针与在探针的条形码结构域上的六个位置各自处的附着区域杂交,并且检测与六个位置各自相关联的可检测标记。

[0504] 术语“k聚体探针”与本公开内容的测序探针同义。k聚体读出是Hyb&Seq数据的基本单位。单个k聚体读出得自单个靶分子/Hyb & Seq循环。执行多重Hyb & Seq循环,以从单个靶分子生成足够的离散k聚体读出,以致使离散k聚体明确比对成邻接段的序列。

[0505] 当比对来自离散读数的两个或更多个序列时,可以组合重叠部分以产生单个一致序列。在其中重叠部分具有相同碱基(单列比对)的位置中,这些碱基变成一致序列。对于其中重叠序列中存在不一致的位置,可以使用各种规则来生成一致序列。简单的多数规则使用列中最常见的碱基作为一致序列。“多次通过一致”是来自单个靶分子的所有离散探针读出的比对。根据所应用的探针群体/投票的循环总数,可以用不同的冗余或重叠水平查询单个靶分子内的每个碱基位置;一般地,冗余增加碱基识别的置信水平。

[0506] “一致”是当比对来自离散读数的两个或更多个DNA序列时,可以组合重叠部分以产生单个一致序列。在其中重叠部分具有相同碱基(单列比对)的位置中,这些碱基变成一致序列。对于其中重叠序列中存在不一致的位置,可以使用各种规则来生成一致序列。简单的多数规则使用列中最常见的碱基作为一致序列。

[0507] “原始准确率”是系统正确鉴定碱基的固有能力的量度。原始准确率取决于测序技术。“一致准确率”是系统利用另外的读数和统计功效正确鉴定碱基的能力的量度。“特异性”指每次运行的总读数中映射到预期靶的读数百分比。“统一性”指跨越靶区域的序列覆盖率中的可变性;高统一性与低可变性相关。这个特征通常报告为跨越所有靶向区域,由 $\geq 20\%$ 的平均覆盖深度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;给予足够的通过次数,可以实现基本上‘完全一致’或‘无错误’测序。

[0508] 可以使用能够实施该方法和/或记录结果的任何装置,来实施本文所述的方法和/或记录结果。可以使用的装置的实例包括但不限于电子计算装置,包括所有类型的计算机。

当本文描述的方法以计算机实施和/或记录时,可以用于配置计算机以实施方法的步骤的计算机程序,可以包含在能够包含计算机程序的任何计算机可读介质中。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装置。可以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,也可以经电子网络例如经互联网、内联网或其它网络提供。

[0509] “可消耗的测序卡”可以并入本领域已知的荧光成像装置内。具有许多不同特征的任何荧光显微镜均能够执行这种测序读出。例如:宽场灯、激光器、LED、多光子、共焦或全内反射照明可以用于激发和/或检测。具有基于滤波器或基于光栅的光谱分辨率(一种或多种光谱分辨的发射波长)的相机(单个或多个)和/或光电倍增管(单个或多个),在荧光显微镜的发射检测通道上是可能的。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。

[0510] 测序数据可以通过任何数目的标准下一代测序装配器来分析(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)。在显微镜的单个衍射限制区域内获得的测序数据被“局部组装”,以从衍射点内的多个读数生成一致序列。然后将多个衍射位点组装的读数映射到一起,以生成代表整个靶向基因集或整个基因组的从头组装的邻接序列。

[0511] 与本公开内容有关的另外教导在下述一个或多个中进行描述: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、U.S. 2014/0162251和U.S. 2016/0194701,其各自整体引入本文作为参考。

实施例

[0512] 实施例1 - 使用Hyb & Seq化学的单分子长读

[0513] 本文公开的测序探针和利用该测序探针的方法便利地称为Hyb & Seq。该术语在说明书自始至终用于描述所公开的测序探针和方法。Hyb & Seq是无文库、无扩增的单分子测序技术,其使用荧光分子条形码的核酸杂交循环到天然靶上。

[0514] 用下述关键步骤,对长33千碱基(kb)的单分子DNA靶证实了使用Hyb & Seq的长读:(1)捕获长DNA分子,并将其水动力拉伸到测序流动室上;(2)使多重完全匹配的测序探针跨越长单分子靶杂交;(3)使荧光报告物与测序探针中的条形码区域杂交,以鉴定所有结合的序列;和/或(4)使用空间分辨的荧光数据确定单分子靶内的序列的相对位置。

[0515] 使用Hyb & Seq的长读取的关键优点包括但不限于:读取长度由分子长度决定,不受化学限制;简单、有限的样品制备导致较少的片段化;与测序探针相关的位置信息帮助组装;和/或使变体定相至长范围单倍型的能力。

[0516] Hyb & Seq化学设计 - 测序探针包含与单分子靶碱基配对的靶结合结构域,以及

具有至少三个位置 (R_1 、 R_2 和 R_3) 的条形码结构域,所述至少三个位置对应于靶结合结构域中存在的六聚体序列。4096种测序探针的集合使得能够对任何靶序列进行测序。报告探针:三个报告探针与条形码结构域的位置序贯地结合。每个报告复合物对应于特异性二核苷酸。杂交驱动功能性。

[0517] 本公开内容的长读和短读测序方法可以使用相同的简单探针杂交工作流程,用于核酸的靶向捕获。多种测序探针可以同时与靶核酸杂交,并且光学分辨率允许每个长靶的几个位点个别地加以区分。通过同时杂交且记录多种测序探针,增加了单次读取的信息含量。长范围单倍型是单分子分析中固有的,并且可以通过实际物理位置而不是计算重建来组装。使用本公开内容的测序方法,高达数百的千碱基的长测序读取是可行的。

[0518] 测序探针可以在预计的序列特异性位置和相对物理距离处与拉伸的靶(优选为水动力学拉伸的靶)杂交。与短读技术相比,本公开内容的测序方法具有增加的信息含量,允许每个循环中读出更多的碱基。本公开内容的测序方法还记录测序读出的相对位置,其帮助长读数的组装。使用本公开内容的测序方法,读取长度=一致序列长度=捕获的靶分子的长度。

[0519] 在使用本公开内容的方法的一个实验中,捕获了33千碱基DNA片段,拉伸,与测序探针和报告探针杂交并进行检测。本公开内容的测序方法与高达33千碱基及以上的DNA片段相容。读取长度仅受靶核酸片段的初始长度限制,而不受酶或测序化学限制。

[0520] 本公开内容的方法具有关于靶向定相长读取的另外能力。长范围定相单倍型是数据中固有的,并且对于变体的定相容易地鉴定。不需要对整个长靶分子进行测序,因为“阻断物寡核苷酸”可以用于将测序循环限于目的测序窗口。

[0521] 实施例1的结果显示了,本公开内容的测序方法能够以长读取长度进行单分子测序。特别地,结果显示了:15千碱基和33千碱基单链DNA分子的成功捕获和水动力拉伸;空间分辨的荧光数据准确地对应于跨越长单分子的实际相对位置;以及每个测序循环10+碱基序列的同时读出。

[0522] 实施例2 - 组装算法:用于靶向测序的Hyb & Seq读数的准确、参考引导的组装,以分辨短核苷酸变体和InDel

[0523] 组装算法是开放源代码算法,其设计为执行Hyb & Seq的独特六聚体读出(六聚体谱)的组装。组装算法也可以称为ShortStack或HexSembler™ 分析软件。该算法是利用来自每个成像特征的六聚体读数进行靶鉴定,并且伴随错误校正基于单个分子,将六聚体读出组装成一致序列的统计方法。

[0524] 使用Hyb & Seq化学和组装算法的单分子测序如下执行:在使用Hyb & Seq化学的每个杂交循环后,生成单分子靶的六聚体读出;在多个杂交循环后,产生了覆盖每个单分子靶区域的六聚体谱;并且六聚体谱与每个靶核酸分子的参考序列一起使用,以推导每个单分子靶的一致序列。

[0525] 使用Hyb & Seq技术与组装算法的靶测序的结果显示了:使用六聚体谱的单分子靶鉴定算法具有100%的成功率;参考引导的组装算法在5x覆盖率下产生了>99%(~QV 32)的单分子一致性准确率;使用预先表征的参考gDNA样品证实了一致的体细胞变体检测($R^2 \sim 90\%$);和/或使用所有六聚体和组装算法的计算机模拟(*in silico*)实验,确认了跨越较大的靶组的平均QV >90。

[0526] 组装算法可以准确地组装Hyb & Seq数据。图18显示了来自使用本公开内容的测序方法获得并使用组装算法分析的测序实验的结果。在该实验中,测序的靶核酸包括以下基因的片段: BRAF (SEQ ID NO: 3)、EGFRex18 (SEQ ID NO: 4)、KRAS (SEQ ID NO: 5)、PIK3CA (SEQ ID NO: 6)、EGFRex20 (SEQ ID NO: 7) 和NRAS (SEQ ID NO: 8)。图18显示了碱基覆盖率和变体识别两者。覆盖率图显示了FFPE (福尔马林固定的石蜡包埋的) gDNA中的碱基覆盖率。结果显示了,跨越各种靶的大多数碱基被可用的测序探针覆盖。误差图显示了跨越各种靶,在FFPE gDNA样品中的查询位置处的错误率相对于覆盖率。结果显示了,在8x覆盖率下,错误率<1%。频率图显示了测序的Horizon gDNA样品中,变体的预期频率与已知频率之间的关联性。该表提供了测序的Horizon Genomic Reference gDNA,并且显示了变体分子的分数与参考样品的已知频率一致。

[0527] 实施例2中的结果显示了,组装算法是用于使用本公开内容的测序方法获得的六聚体谱的子组装的准确算法。特别地,结果显示了:使用模拟数据,靶鉴定中的100%准确率以及每碱基平均质量值>30;在5x覆盖率下,实验Hyb & Seq数据中的碱基识别中的>99%准确率;以及与已知值一致的频率 ($R^2 \sim 90\%$),来自基因组DNA的变体检测;并且计算性能是有效的,并且随着组装的六聚体的数目线性地缩放,能够在~15分钟内在个人计算机上组装69k分子。

[0528] 实施例3 - 使用Hyb & Seq™技术-基于杂交的单分子测序系统-对来自FFPE样品的天然gDNA进行无文库的靶向测序

[0529] 使用本公开内容的测序方法 (Hyb & Seq),对来自FFPE样品的天然gDNA进行靶向癌组测序,以证实:具有准确碱基识别的癌基因靶的靶向单分子测序;已知的致癌单核苷酸变体 (SNV) 和插入/缺失 (InDel) 的准确检测;来自FFPE提取的gDNA (中值DNA片段大小200个碱基) 的癌基因靶的多路复用捕获;和/或在高级原型器械上执行的端到端自动测序。

[0530] Hyb & Seq化学和 workflow 如下演示:使目的基因组靶直接捕获到测序流动室上;使含有数百种六聚体测序探针的池流入测序腔室内;使荧光报告探针与测序探针的条形码区域序贯地杂交,以经过3个报告物交换循环鉴定六聚体碱基;一旦鉴定了碱基,就洗掉测序探针;并且用新的测序探针池重复该循环,直至靶区域已被读取至足够的深度。

[0531] Hyb & Seq的关键优点:简单快速的FFPE workflow - 临床样本在60分钟内开始测序;无酶或扩增/无文库构建;15分钟的总动手时间;高准确率 - 低化学错误率 + 固有错误校正;和/或长读和短读两者 - 读取长度由输入样品确定,不受化学限制。

[0532] Hyb & Seq化学设计如实施例1中所述。用于处理FFPE组织的Hyb & Seq样品制备由三个简单步骤组成:(1)单管脱石蜡和裂解;(2)使用注射器式过滤器的微粒去除;以及(3)任选的DNA片段化和靶捕获。该过程要求每个样品使用一至三个10微米FFPE卷状物 (cur1)。整个过程在60分钟内完成,并且仅需要常见的实验室设备:加热块、移液器、过滤器和试剂。

[0533] 图19显示了来自FFPE样品的癌基因靶的多路复用捕获和测序的实验设计的示意图。设计且构建了总共425种测序探针,以对11个致癌基因靶 (SEQ ID NO:3-13) 的一部分进行测序。关于每个基因靶的已知变体的基因座由许多测序探针覆盖 (完全匹配+单个错配)。跨越这些区域测量了碱基覆盖率和逐个碱基的准确率。使用预先表征的参考样品,获得了变体检测的准确率。图19的上图显示了,测序探针 (蓝色) 与围绕已知变体位置 (红色) 的靶

序列(灰色)比对。对于每个变体位置(红色),对于每个(A、G、C、T)碱基变体,包括4种探针序列。在测序期间,跟踪单个靶DNA分子用于800个条形码交换循环,提供了通过组装算法重新组装的多重六聚体读数,如实施例2中所述。

[0534] 图18显示了测序结果,其包括每个靶的平均覆盖率、单个碱基错误率、以及观察到的相对于预计的变体频率。实施例3中的结果显示了,Hyb & Seq测序可以用于对FFPE和参考gDNA样品中的11个靶区域执行多路复用测序,其中以低误差检测到单核苷酸变异。

[0535] 实施例4 - 使用Hyb & Seq™化学的直接单分子RNA测序,而无需cDNA转换

[0536] 使用Hyb & Seq化学的直接单分子RNA测序如下演示:直接捕获天然RNA分子而无需cDNA转换,并且将其固定到测序流动室上;使含有数百种六聚体测序探针的池流入测序流动室内;使完全匹配的测序探针在单分子RNA靶上随机杂交;使荧光报告探针与测序探针的条形码区域序贯地杂交,以鉴定六聚体碱基;并且鉴定碱基,然后洗掉测序探针;重复循环直至靶已被读取至足够的深度。

[0537] 关键结果:对靶向的单分子RNA进行测序,显示了与DNA相似的覆盖率谱;在多于200个Hyb & Seq循环自始至终,RNA分子稳定地维持在流动室上;从单个FFPE切片中同时捕获且定量mRNA和基因组DNA;和/或使用低至10 ng的总RNA,多路复用捕获八种转录物且进行定量。

[0538] Hyb & Seq化学设计如实施例1中所述。图20的左图显示了与使用cDNA转换执行的常规RNA测序相关的步骤相比,与直接RNA测序相关的实验步骤的说明性示意图。图20的中图和左图显示了,来自测试RNA分子与本公开内容的测序方法的相容性的实验的结果。在该实验中,对4种靶RNA分子进行测序(SEQ ID NO:14-17)。结果显示了,RNA分子可以被捕获并检测至少200个测序循环,证实了本公开内容的测序方法与RNA分子的相容性。

[0539] 图21显示了来自使用本公开内容的测序方法,验证直接单分子RNA测序的实验的结果。直接捕获编码NRASex2的片段(SEQ ID NO:18)的天然RNA分子而无需cDNA转换,并且将其固定到测序流动室上,并且使用本方法进行测序。还使用捕获的DNA分子代替RNA重复了该实验。图21显示了关于DNA和RNA的测序覆盖率是可比较的,证实了使用本公开内容的测序方法,可以直接对RNA进行测序而无需转换为cDNA。

[0540] 在一些方面,本公开内容提供了来自FFPE样品的RNA和DNA的整合捕获。使用实施例3中所述的相同FFPE工作流程制备样品。使用相同的捕获方案,但用RNA和DNA特异性捕获探针。用相同的测序探针,在同一流动室中,同时对DNA和RNA分子进行测序。在一些方面,需要特异性的RNA和DNA捕获探针,以同时捕获RNA和DNA。

[0541] 还用各种输入量的总RNA(0 ng、1 ng、10 ng、100 ng、1000 ng),执行了对人类通用参考RNA的8种中至高表达转录物的多路复用捕获。将多路复用捕获的RNA分子固定到流动室上,并且使特异性测序探针和报告探针与固定的RNA分子杂交用于定量。关于每个特异性RNA靶的计数的定量显示了随着输入RNA的总量增加的增加。

[0542] 实施例4中的结果显示了,用Hyb & Seq化学实现单分子RNA测序。特别地,结果证实了:(1)直接RNA测序而无需cDNA转换;(2)在Hyb & Seq循环过程自始至终,RNA分子是稳定的;(3)RNA和DNA分子两者均可以在一个Hyb & Seq工作流程中捕获且测序;以及(4)mRNA组的靶捕获可以用少至10 ng的总RNA输入来执行。

[0543] 实施例5 - 由Hyb & Seq测序平台生成的高通量分子水平的短读数的整合生物信

息学算法

[0544] 组装算法软件被设计为执行基于标准测序的生物信息学分析任务,例如比对、错误校正、突变识别和读数组装。组装算法软件管线的步骤可以包括:六聚体的比对和覆盖率估计;突变序列鉴定;图形数据结构构建;和/或分子水平序列重建和错误校正。

[0545] 所有算法都严格地在从单个分子获得的信息内执行,确保最终的突变识别结果不因样品的突变频率而偏差。根据组结合位置,将六聚体分组成不同的分子。为了将分子分配给靶,将六聚体每个分子与所有不同的靶区域进行比对,并且选择最匹配的基因靶。

[0546] 测量统计度量以评价分子鉴定的质量。针对N个不同靶区域的比对对于每个靶产生N个总计覆盖率值的分布。选择最高的总计覆盖率分数匹配作为正确匹配。测量所选择的最高匹配分数针对所有N个不同靶的分数分布的Z分数统计。过滤掉低置信度分子鉴定(低于 2.5σ 的z分数)。

[0547] 组装算法的关键优点包括:通过实施分层哈希索引设计来准确处理可能的序列歧义;和/或高级算法设计结构通过优先级确保映射质量,并防止突变的高估。

[0548] 另外,突变图形数据结构使得能够对各种类型的突变(取代、插入和缺失)进行计算建模,并且产生输出用于序列重建和变体识别:取代变体在图形中表示为与原始序列相同长度的另外节点;插入可以通过添加任何长度的连接节点来建模;缺失被建模为在图形中用空碱基对字符串添加人工节点;在盲目突变搜索(即,关于突变耐受的序列比对的搜索)中,从每一个参考序列位置测量汉明距离,并且将新节点加入代表搜索到的突变的图形中;和/或使用分层哈希表执行关于突变的六聚体的覆盖率估计。

[0549] 所构建的图形数据结构使得能够进行分子水平序列重建和器械误差校正。在构建的图形中,应用动态程序化算法以发现最佳评分路径,其中将分数定义为标准化的碱基覆盖率。图形的最佳评分路径代表分子水平序列重建。包括正确的突变序列,同时弃去六聚体中的器械误差。

[0550] 模拟数据集确认了该软件能够提供高度准确的分子水平序列组装和突变识别结果。这些结果显示了10个随机突变的突变识别准确率。在中等器械误差数据集中,准确率平均显示为99.39%(靶向搜索)和98.02%(盲目搜索)。在升高的器械误差模拟下,性能平均显示为97.19%(靶向搜索)和93.53%(盲目搜索)。当分子水平碱基覆盖率阈值增加至2x时,结果改善至99.5%(2x覆盖率)和99.9%(3x覆盖率)。

[0551] 组装算法软件可以处理广泛范围的各种突变,包括选择在1 bp和15 bp之间的插入和缺失,其突变识别准确率为94.4%(1x覆盖率)、97.7%(2x覆盖率)至98.5%(3x覆盖率)。

[0552] 实施例6 -用于处理FFPE组织用于Hyb & Seq的样品制备

[0553] 福尔马林固定的石蜡包埋的(FFPE)组织是对于传统测序平台具有挑战性的样品输入类型。Hyb & Seq的样品制备方法成功地处理了FFPE组织输入物,用于下游测序。首先,在一步法中,从福尔马林固定、石蜡包埋的(FFPE)组织中提取待测序的核酸。在水基核酸提取缓冲液中加热一个或多个10 μ m厚的FFPE卷状物,以同时熔化石蜡、分解组织并从细胞中释放核酸。合适的提取缓冲液是本领域已知的,并且通常包括蛋白酶、去污剂例如Triton-100、螯合剂例如EDTA和铵离子。使FFPE卷状物和提取缓冲液在56°C下温育30分钟,以使石蜡与组织分开,并且允许蛋白酶K消化组织结构,并使包埋的细胞暴露于去污剂以使得细胞裂解。将溶液以8分钟的间隔倒置三次,以帮助在组织脱石蜡和消化过程期间的试剂混合。

在此步骤之后,将溶液加热到98℃,以促进甲醛交联的逆转,以进一步帮助核酸的提取。

[0554] 一旦已从FFPE组织中提取了核酸,就使用2.7 μm孔径的玻璃纤维过滤器(Whatman)过滤溶液,以去除组织碎片和凝结的石蜡。所得到的溶液是含有核酸的均质、半透明溶液,所述核酸由于福尔马林固定过程和贮存条件而是高度片段化的。如果需要进一步的片段化,则可以使用Covaris聚焦超声发生器对DNA进行机械剪切。由于缓冲条件,需要延长的超声处理以剪切核酸。使用50W峰值入射功率、20%占空比、200个循环/脉冲的标准设置进行超声处理600秒,以实现捕获靶中的最大增加。为了实现更短的片段长度,可以通过在21,000 g和4℃下离心15分钟,从过滤的溶液中沉淀出乳化的石蜡。这允许DNA剪切至约225 bp。

[0555] 接下来,在快速杂交步骤期间,通过使捕获探针对与靶核酸分子结合来执行靶捕获。5'捕获探针含有3'生物素部分,其在靶沉积过程期间,允许靶与链霉抗生物素蛋白包被的流动室表面结合。3'捕获探针含有5'标签序列(G序列),其使得能够在纯化过程期间与珠结合。反应速率受捕获探针浓度驱动,所述捕获探针浓度以低纳摩尔范围加入,以使反应速率达到最大。捕获探针以与目的区域侧接的方式与靶杂交,以便生成测序窗口。对于每个DNA靶,捕获探针集合还包括由与测序窗口相同的序列组成的寡核苷酸,以与靶的反义链杂交并防止再退火。将含有捕获探针的溶液加热至98℃共3分钟,以使基因组DNA变性,随后为在65℃下的15分钟温育。在400 mM至600 mM范围内的NaCl浓度用于这种杂交反应。表3列出了已进行实验验证的超过100种靶的组,所述表3详细列出了靶向DNA区域的基因和外显子。

[0556] 表3:靶向DNA区域的基因和外显子

[0557]

基因	靶	基因	靶	基因	靶
ABL1	ABL1_ex4	EZH2	EZH2_ex8	NOTCH1	NOTCH1_ex26
	ABL1_ex6		EZH2_ex11	NRAS	NRAS_ex2
	ABL1_ex7		EZH2_ex15		NRAS_ex3
AKT1	AKT1_ex6	FBXW7	FBXW7_ex2	NRAS_ex3	NRAS_ex4
ALK	ALK_ex26		FBXW7_ex5	PDGFRA	PDGFRA_ex1
APC	APC_ex5		FBXW7_ex7		PDGFRA_ex4
	APC_ex16		FBXW7_ex8		PDGFRA_ex7
	APC_ex17	FBXW7_ex9	PDGFRA_ex10		
	APC_ex17	FBXW7_ex10	PDGFRA_ex11		
	APC_ex17	FGFR1	FGFR1_ex6	PDGFRA_ex14	
ATM	ATM_ex8	FGFR2	FGFR2_ex7	PDGFRA_ex15	
	ATM_ex9	FLT3	FLT3_ex11	PDGFRA_ex16	
	ATM_ex11		FLT3_ex12	PDGFRA_ex18	
	ATM_ex26	FLT3_ex21	PIK3CA	PIK3CA_ex2	
	ATM_ex34	GNAQ		GNAQ_ex5	PIK3CA_ex3
	ATM_ex39	IDH1		IDH1_ex4	PIK3CA_ex7
			IDH1	IDH1_ex10	
		IDH2	IDH2_ex4		

	ATM_ex49 ATM_ex49 ATM_ex55 ATM_ex59	JAK2	JAK2_ex3 JAK2_ex7 JAK2_ex14 JAK2_ex20		PIK3CA_ex10 PIK3CA_ex14 PIK3CA_ex21 PIK3CA_ex21
BRAF	BRAF_ex8 BRAF_ex11 BRAF_ex13 BRAF_ex15	KDR	KDR_ex7 KDR_ex7 KDR_ex9 KDR_ex11 KDR_ex27 KDR_ex30	PTEN	PTEN_ex5 PTEN_ex7 PTEN_ex8
CDH1	CDH1_ex9			PTENP1	PTENP1_ex1
CSF1R	CSF1R_ex3 CSF1R_ex22	KIT	KIT_ex5 KIT_ex9 KIT_ex14 KIT_ex14 KIT_ex17 KIT_ex18	RB1	RB1_ex10 RB1_ex17 RB1_ex17 RB1_ex20 RB1_ex22
CTNNB1	CTNNB1_ex3 CTNNB1_ex6 CTNNB1_ex16				RET
EGFR	EGFR_ex3 EGFR_ex10 EGFR_ex15 EGFR_ex18 EGFR_ex20 EGFR_ex21	KRAS	KRAS_ex2 KRAS_ex3 KRAS_ex4	SMAD4	SMAD4_ex3 SMAD4_ex8 SMAD4_ex9 SMAD4_ex10 SMAD4_ex11
		MEK	MEK_ex3		
ERBB2	ERBB2_ex7	MET	MET_ex2 MET_ex3 MET_ex11 MET_ex14 MET_ex16 MET_ex19	SMARCB1	SMARCB1_ex5
ERBB4	ERBB4_ex4 ERBB4_ex5 ERBB4_ex7 ERBB4_ex8 ERBB4_ex23 ERBB4_ex25		MLH1	MLH1_ex12 MLH1_ex16	TP53

[0558]

[0559] 在靶向DNA区域与捕获探针结合后,将它们从基因组DNA的剩余部分纯化,以产生富集的靶溶液。使由针对3'捕获探针的结合序列的反义寡核苷酸(抗G序列)包被的珠,与捕获反应混合物一起在室温下温育15分钟。在结合步骤后,将珠用0.1x SSPE洗涤三次,以去除非靶DNA和含生物素的5'捕获探针。在洗涤之后,使珠重悬浮于14 μ L的0.1x SSPE中,然后在45°C下加热10分钟,以从珠中洗脱纯化的DNA靶。在洗脱后,添加1 μ L的5 M NaCl,以确保捕获探针保持与DNA靶结合。

[0560] 样品制备过程的最后步骤是将DNA靶沉积到流动室表面上,在其中可以使用如本文公开的本公开内容的探针对它们进行分析。利用注射泵,以控制在其下将靶装载到流动室流体通道内的速率,使得所有靶都有时间跨越通道的高度扩散,并且与链霉抗生物素蛋白表面结合。这种装载方法生成靶的密度梯度,其中每单位面积的最高分子数目在流体通道入口处最大,并且沿着通道长度在朝向出口的流体流动方向上减小。对于1.6 mm的通道宽度和40 μ m的高度,0.35 μ L/秒的流速实现了在约10 mm的通道长度内的定量捕获。一旦靶通过生物素化的5'捕获探针与表面结合,就注入生物素化的寡核苷酸(G钩)溶液,其是3'

捕获探针的结合序列的反向互补体,以固定靶的自由末端,以产生桥接结构,其中中间的ssDNA区域是目的测序窗口。接下来,添加G序列寡核苷酸的溶液,以与表面上的过量G钩杂交,以减少表面上的ssDNA量。图8显示了使用本公开内容的双捕获探针系统的靶核酸捕获。

[0561] 实施例7 - Hyb & Seq的多色报告图像处理

[0562] 图像处理管线包括下述步骤:背景减除、配准、特征检测和分类。在背景减除中,任何给定通道的平均背景是散粒噪声和曝光的函数。在我们的系统中,蓝色通道具有最高的背景水平加上更大的方差。应用具有半径7像素的圆形结构元素的简单顶帽过滤器,以执行局限性背景减除。对于配准,必须将目的特征对于多色和多循环特征分析进行完全比对。这种系统需要两种配准形式。对于第一种形式,局部仿射变换应用于单个采集堆栈内的所有图像通道。这种变换是光学系统的函数,并且因此对于给定的器械是一致的。这种函数对于每一次运行预先计算,并且应用于所获取的每一个图像。对于第二种形式,使用标准化的互相关来计算以刚性移位形式的总体变换,以捕获在运行期间机械台架的漂移。下一步为特征检测。

[0563] 一旦所有图像被配准,就使用匹配的滤波器经由LoG(高斯拉普拉斯)滤波器来检测特征。滤波器以固定内核尺寸(与特征的衍射极限相匹配)和不同的标准差(与相应通道的波长相匹配)应用,以增强位点应答。局部最大值用于鉴定潜在的报告物位置。检索关于每个鉴定特征的关联强度值用于分类。最后一步是分类。使用高斯朴素贝叶斯模型,对多色报告物强度进行分类。该模型假设报告物强度是独立的,并且遵循正态分布。然后,该模型使用最大后验或MAP规则,计算特异性特征 \hat{y} (由所有通道的强度 \hat{x}_i 指定)属于某个类别(C_k)的概率:

$$[0564] \quad \hat{y} = \underset{k \in \{1, \dots, K\}}{\operatorname{argmax}} p(C_k) \prod_{i=1}^n p(x_i | C_k)$$

[0565] 在使用两种染料(蓝色和红色)的编码方案的实例中,在2色编码方案中存在可能的六个类别(包括背景)。在实施的系统中,四种颜色的选择导致14个潜在类别。应注意,单半染料分布相对于全染料分布之间存在一些重叠。因而,这些类别之间的分类呈现了较高的错误率,其中'xG'和'GG'之间的最大错误分类率为11.8%。关于10类模型的错误分类率小于0.2%。由于每种报告物需要最多八个类别,因此选择具有最少分类错误的类别很简单。基于查找表(look up table),将检测到的颜色代码转换成鉴定的碱基对。使用如本文公开的本公开内容的探针,跨越多个循环跟踪特征。

[0566] 实施例8 - 使用捕获探针的靶核酸纯化和沉积。

[0567] 为了捕获靶核酸分子,使用双捕获探针系统用于高度特异性富集。捕获探针被设计为在靶向目的区域侧翼的位置处与靶核酸结合,产生“测序窗口”。称为CapB的5'捕获探针含有3'生物素部分。称为CapA的3'捕获探针含有称为G序列的5'亲和标签序列。平均而言,捕获探针为长度大约40个核苷酸,并且基于Tm和序列背景进行设计。测序窗口为长度约70个核苷酸,并且易于调整。图8显示了双捕获探针系统的示意图。

[0568] CapB上的生物素部分将靶核酸栓系至链霉抗生物素蛋白包被的流动室表面用于测序。CapA上的亲和标签允许在纯化期间靶核酸分子与磁珠的可逆结合。CapA和CapB的使用允许高度严格的靶富集,因为两种探针均保持与单个靶核酸分子的结合,以使该靶经磁

珠纯化和表面沉积过程两者而存活。多路复用捕获已被证实同时具有高达100种靶。为了在短时间内实现有效捕获,以1 nM至10 nM的浓度范围添加捕获探针。

[0569] 在实验测试中,使用G珠和双探针捕获系统纯化了约10种靶核酸分子的组。首先使CapA和CapB探针与靶核酸杂交。然后使结合的CapA探针的G序列部分与G珠上的G钩杂交,从而使靶核酸分子连接到G珠。使用0.1x SSPE进行一系列严格洗涤,以去除非靶向的DNA和未结合的CapB。为了从G珠释放靶核酸分子,进行低盐、45℃洗脱以使G序列变性,同时仍允许CapA和CapB保持与靶核酸杂交。

[0570] 测试显示了,当纯化约10种靶核酸分子的组时,非特异性/背景信号显著增加。背景的增加可能是由于包括以下的几种因素:(1) CapA和CapB探针种类之间的相互作用增加,其导致通过纯化携带的游离CapB探针的量增加;以及(2) CapB探针与G钩或G珠之间的相互作用增加,其导致不需要的靶核酸的纯化。此外,随着组的大小增加,CapB种类、CapA种类和测序探针之间可能的相互作用呈指数增加。这些相互作用可以干扰密集沉积靶的能力,并且导致浪费的测序读数。

[0571] 为了减少由于游离探针种类和不需要的靶核酸分子的纯化导致的非特异性和背景信号,可以对纯化程序进行一些修改。首先,在靶核酸分子与G珠结合期间使用的缓冲液中包括浓度为30% v/v的甲酰胺,使背景计数减少两倍(如通过缺乏靶分子的对照中的计数测量的),可能是通过干扰游离捕获探针与G钩的不完全杂交,允许过量探针被洗掉。其次,在G珠上的G钩中包括4种异-dG碱基(异-G钩)、以及在CapA G序列中包括互补的异-dC碱基,使背景计数减少三倍(如通过缺乏靶分子的对照中的计数测量的)。异-dC和异-dG是天然dC和dG碱基的异构体变体。由于异碱基将与其它异碱基而不是天然碱基碱基配对,捕获探针与异-G钩之间的不完全杂交可以仅存在于G序列的非异碱基和异-G钩之间。在严格洗涤期间,这些不完全的相互作用更易于破坏。最后,异-G珠洗出液用Ampure® XP (Agencourt Biosciences Company) 珠的随后纯化,将背景计数进一步减少至少20倍(如通过缺乏靶分子的对照中的计数测量的)。在Ampure® XP珠纯化期间,将DNA样品与羧基化磁珠在聚乙二醇(PEG)和NaCl的溶液中的悬浮液混合。可以这样滴定PEG和NaCl的浓度,使得仅高于分子量阈值的分子沉淀并与珠结合。与捕获探针杂交的Hyb&Seq靶为大约81 kDa,而游离探针为大约17kDa或更小。通过将Ampure® XP珠悬浮液与异-G珠洗出液以1.8:1的体积比混合,使杂交的靶与珠结合,并且可以在最终洗脱之前洗掉显著部分的游离探针。

[0572] 因此,模型纯化工作流程由下述步骤组成:(1) 使捕获探针-靶核酸组装物在5x SSPE/30%甲酰胺中与异-G珠杂交;(2) 用0.1x SSPE洗涤异-G珠;(3) 在45℃下在0.1x SSPE中洗脱捕获探针-靶核酸组装物;(4) 使异-G珠洗出液与1.8x体积的Ampure® XP珠结合;(5) 用75%乙醇洗涤Ampure® XP珠;以及(6) 以0.1x SSPE洗脱捕获探针-靶核酸组装物,使得靶以7.5 μL洗脱,随后加入0.5 μL的5 M NaCl。

[0573] 在纯化后,使用输注注射泵,将捕获探针-靶核酸组装物沉积到测序表面上,以使纯化的靶缓慢注入通过流动室。为了确定沉积梯度,流动室在沿着通道长度的各个位置处成像。对于20 μm的通道高度,以0.167 μL/分钟的流速装载样品使靶集中,使得所有靶的80%在沿着通道长度的5.1 mm内结合,这代表Gen2成像仪的~240个FOV,其中FOV为0.0357 mm²且流动室通道宽度为1.7 mm。梯度可以通过在沉积期间调整流速来调节。

[0574] 上述程序用于测试100plex靶核酸组的纯化和沉积,其中基因组DNA剪切至~300个

碱基对的大小。一式三份执行一系列实验,其中DNA输入范围在25 ng至500 ng之间。通过对沉积梯度成像来外推流动室上的总靶数目,以获得平均计数的数目。捕获效率为6.6%,并且在DNA质量输入的范围是一致的。

[0575] 实施例9 - 测序探针的设计和特征

[0576] 测序探针经由靶结合结构域与靶核酸分子杂交。在本实施例中,靶结合结构域长8个核苷酸,并且含有侧翼为(N)个碱基的锁核酸(LNA)六聚体,所述碱基可以是通用/简并碱基或规范碱基($N_1-B_1-B_2-B_3-B_4-B_5-B_6-N_2$,其中 B_1 至 B_6 是LNA,并且 N_1 和 N_2 是不依赖于(6聚体)序列 $B_1-B_2-B_3-B_4-B_5-B_6$ 的核酸序列的规范碱基或通用/简并碱基)。4096种测序探针的完整集合编码所有可能的六聚体,并且使得能够对任何靶核酸进行测序。每种测序探针还包括条形码结构域,其编码靶结合结构域中存在的六聚体序列。每个条形码结构域含有3个位置(R1、R2和R3)。条形码结构域中的每个位置对应于靶结合结构域的六聚体中的特异性二核苷酸,并且含有可以与特异性标记的报告复合物结合的独特序列。测序探针的示意性概述显示于图1中。条形码结构域中的每个位置编码八种“颜色组合”,其使用四种荧光染料产生:蓝色(B);绿色(G);黄色(Y);以及红色(R)。在每个测序循环期间,报告复合物与条形码结构域中的三个位置之一结合,指示了靶结合结构域的六聚体中的相应二核苷酸的身份。在三个序贯测序循环期间,记录三个“颜色组合”,条形码结构域中的每个位置一个,允许鉴定靶结合结构域的整体六聚体。将4096种测序探针拆分到8个池内,并且每个池与512种可能的条形码之一相关联。

[0577] 实施例10 - 报告复合物的设计、纯化和结合条件。

[0578] 在该实施例中,每个报告复合物是37种DNA寡聚物分支结构,其设计为容纳总共30种荧光染料,其中对于颜色组合的每种颜色具有15种染料。构成报告复合物的37种DNA寡聚物可以按其大小分类。称为一级核酸的最大寡聚物共价附着至长度为12或14个核苷酸的互补核酸。一级核酸长96个核苷酸。一级互补核酸与测序探针的条形码结构域上的位置R1、R2或R3结合。下一个最大的DNA寡聚物长89个核苷酸,并且称为二级核酸。存在六个二级核酸/报告复合物,其中对于颜色组合的每种颜色具有三个二级核酸。每个二级核酸包含长14个核苷酸的序列,其允许二级核酸与一级核酸杂交。最小的DNA寡聚物长15个核苷酸,并且称为三级核酸。存在30个三级核酸/双色探针,其中每种颜色具有15个三级核酸。五个三级核酸与每个二级核酸结合。37种DNA寡聚物分支结构的示意图显示于图4中。

[0579] 三级核酸包括以荧光染料形式的可检测标记。存在四种荧光染料:蓝色(B);绿色(G);黄色(Y);以及红色(R)。在报告复合物中将染料组合在一起导致十种可能的双色组合(BB、BG、BR、BY、GG、GR、GY、RR、YR、YY)。为了防止不同荧光染料之间的颜色交换或交叉杂交,对应于特异性荧光染料的每个二级核酸和三级核酸含有独特序列。例如,用Alexa 488荧光团或蓝色标记的每个三级核酸包含仅与蓝色二级核酸互补的序列。蓝色二级核酸进一步具有独特的序列,其仅与对应于包括蓝色的颜色组合的一级核酸分子互补。

[0580] 每个互补核酸含有在测序探针的条形码结构域的位置R1、R2和R3之间独特的序列。因此,即使相同条形码结构域的位置R1和R2编码相同的二核苷酸,鉴定在位置R1处的二核苷酸的互补核酸分子的结合也不与位置R2结合。同样地,鉴定在位置R2处的二核苷酸的互补核酸分子不与位置R1结合。这样设计互补核酸,使得它们可以使用竞争性的立足点交换(对于长度为12个核苷酸的互补核酸)、或UV切割(对于长度为14个核苷酸的互补核酸),

有效地解除与测序探针的结合。

[0581] 报告复合物的制备在两个序贯的杂交步骤中发生：(1) 三级核酸与二级核酸，然后(2) 三级核酸+二级核酸与一级核酸。通过在4.2X SSPE缓冲液中，在室温下，将100 μM 的二级核酸和600 μM 的三级核酸组合30分钟，来制备四个分开的三级核酸-二级核酸反应。然后使用2 μM 的一级核酸、7.2 μM 的二级核酸+染料#1三级核酸和7.2 μM 的二级核酸+染料#2三级核酸，在4.8X SSPE中分开制备24种报告探针。将这些反应在45 $^{\circ}\text{C}$ 下加热5分钟，然后在室温下冷却30分钟。然后将24个反应合并到对应于条形码结构域(即 R_1 、 R_2 和 R_3)的三个不同池内。例如，将与 R_1 条形码结构域结合的八种不同报告探针(各2 μM)合并在一起，稀释十倍至每种报告复合物200 nM的最终工作浓度。可以使用高效液相层析(HPLC)纯化报告复合物。HPLC纯化可以去除游离寡聚物和畸形探针，以产生报告探针。

[0582] 报告复合物制备之后是用于质量保证的标准测试。在三个分开的流动室中，测试三个报告探针池各自与其相应的条形码区域(R_1 、 R_2 和 R_3)的结合。对修饰的测序探针构建体(仅存在条形码结构域并固定到流动室上)进行测试。代表每种颜色的所有八种12聚体都是多路复用的，并且预计所有八种报告探针都以高颜色计数鉴定。

[0583] 为了改善报告探针和测序探针的条形码结构域杂交的效率和准确率，测试了各种缓冲添加剂。来自实验的结果指示了，含有5%硫酸葡聚糖(500K)和15%甲酰胺或15%碳酸乙烯酯的缓冲液，允许报告探针和测序探针在短杂交时间内的最有效和最准确的杂交。然而，来自其它实验的结果指示了，碳酸乙烯酯对测序玻片的表面具有负面影响，导致靶核酸随着时间过去的高度丧失。因此，含有5%硫酸葡聚糖(500K)和15%甲酰胺的缓冲液，对于报告探针和测序探针的有效和准确杂交是优良的。

[0584] 实施例11 - 互补核酸序列的设计和验证

[0585] 报告探针含有与测序探针的条形码结构域上的特异性位置(R_1 、 R_2 或 R_3)结合的互补核酸。设计并测试了含有12个核苷酸(12聚体)或14个核苷酸(14聚体)的互补核酸，以确定用于杂交的最佳序列。对于筛选，使用下述标准来确定最佳序列：序列必须展示高结合效率，如通过报告探针和测序探针在10个测序循环中以> 80%效率结合所定义的；序列必须展示在15秒至30秒内的快速杂交动力学；并且序列必须展示高特异性，在报告物池中具有<5%交叉杂交错误。

[0586] 表4显示了鉴定的二十四种12聚体序列(SEQ ID NO:19-42)。由于每个条形码结构域含有三个位置，因此可以将二十四种12聚体序列分成三组，以产生8x8x8 12聚体报告物集合。

[0587] 表4

[0588]

报告物位置	12聚体序列	报告物名称	颜色	SEQ ID NO
1	AGGACAGATGAC	R1BB-07	BB	19
1	GTATCGGATGAC	R1BG-07d (R1RR-06)	BG	20
1	AGGAGTGATGAC	R1BR-07	BR	21
1	AGGGGTGAGGAG	R1GG-07c (R1YR-07)	GG	22
1	AGAGGGGATGAC	R1GR-07	GR	23
1	AGTGGGGAGGAG	R1GY-07c (R1BY-07)	GY	24
1	AGCCGAGATGAC	R1RR-07	RR	25

1	AGGGTGGATGAC	R1YY-07	YY	26
2	TGGATGAAAAG	R2 BB (用于GRv5)	BB	27
2	GAAGGAGAAAAG	R2 BG (用于GYv5)	BG	28
2	GGGGATGAAAAG	R2 BR (用于GRv4)	BR	29
2	GTGAGGGAAAAG	R2 BY (用于YYv5)	BY	30
2	AGCCGAGAAAAG	R2 GG	GG	31
2	CGAGAGGAAAAG	R2 GY (用于GGv5)	GY	32
2	GAGGGCGAAAAG	R2 RR (用于GGv4)	RR	33
2	AGCGTGGAAAAG	R2 YY	YY	34
3	TGAGAAGGGTAG	RPTR12-BG_Screen3_D2	BG	35
3	GTTGTTATTGTG	RPTR12-BR_RC_D4	BR	36
3	TTTGGGTTTAGG	RPTR12-BY_RC_D3	BY	37
3	GTTAGTGGGAAA	RPTR12-GR_RC_D7	GR	38
3	ATGGGAAAAAGT	RPTR12-GY_RC_D6	GY	39
3	GAGTTGGATGAG	RPTR12-RR_RC_D10	RR	40
3	ATGTTGTGGGTA	RPTR12-YR_RC_D9	YR	41
3	GAGGGTTTTAAG	RPTR12-YY_RC_D8	YY	42

[0589] 以相似的方式设计14聚体序列,但在三个方面不同于12聚体序列。首先,14聚体序列含有较长的杂交序列,鉴于14聚体序列含有与条形码结构域上的特异性位置结合的14个单链核苷酸,而不是12聚体中存在的12个单链核苷酸。其次,14聚体序列含有更多的序列多样性,因为它们并非设计为适应立足点介导的去除。由于14聚体序列更强烈地与测序探针杂交,因此立足点介导的去除效率降低。因此,对于14聚体序列探索了序列不依赖性的去除策略,减轻了在筛选期间的序列约束。使用包括下述规则集合的算法设计用于筛选的序列:缺乏“G”或“C”的核苷酸组成(即低复杂性序列);在40%至60%之间的GC含量;在35°C至37°C之间的解链温度(T_m);发夹折叠能量(dG) > 2 ;以及与其它测序探针的相容性(汉明距离 $>= 7$)。为了最小化14聚体序列与可以存在于靶核酸中的基因组序列的杂交,使用External RNA Controls Consortium序列作为指导,来过滤潜在序列。第三,14聚体序列设计为在其中14聚体互补核酸与报告复合物的一级核酸附着的位置处,通过使用可切割接头修饰的链切割,从测序探针的条形码结构域中去除。14聚体序列的去除导致报告复合物信号的“变暗”,允许下一个循环的测序和信号检测发生。测试了各种可切割接头修饰,包括UV光可切割接头、还原剂(例如TCEP)可切割接头、以及酶促可切割接头(例如通过USER™酶切割的尿嘧啶)。发现所有这些可切割接头修饰促进有效的报告复合物变暗。通过将可切割接头修饰引入二级核酸内,进一步增强变暗。将这些可切割接头修饰置于与一级核酸杂交的序列以及与三级核酸杂交的序列之间。图7显示了关于报告探针内的可切割接头修饰的可能位置。

[0590] 潜在的14聚体序列的筛选导致两组可接受序列的鉴定。表5显示了第一组,其含有24种序列(SEQ ID NO:43-66)。可以将这24种序列拆分成三组,以产生8x8x8 14聚体报告物集合。

[0591] 表5

[0592]

报告物位置	14聚体序列	报告物名称	颜色	SEQ ID NO
-------	--------	-------	----	-----------

A	ATCTTTTCCCCACT	R14-BG_RC-Sc3_B2	BG	43
A	CCCCACTATTTCTT	RPTR14-BY_Screen4_I2	BY	44
A	CTACCCACAACATA	RPTR14-YR_Screen3_D9	YR	45
A	CCATATAAACCCCA	R14-GG_RC-Sc3_B5	GG	46
A	AAACTCCAATCTCC	R14-GR_RC-Sc3_B7	GR	47
A	CTATTCTCAACCTA	RPTR14-YY_RS0255_H8	YY	48
A	CCCCCTCTTTTAAA	R14-BB_RC-Sc3_B1	BB	49
A	CCAATCTTACCTCA	RPTR14-RR_Screen3_B10	RR	50
B	CCCTCACATAACTT	RPTR14-BG_Screen4_I1	BG	51
B	CTCCTCTACTTTCC	RPTR14-BB_ERCC_00014.1_10	BB	52
B	CCCTAAACCCAAAA	RPTR14-BY_Screen3_D3	BY	53
B	CACTTTTTCCCATC	RPTR14-GY_Screen3_D6	GY	54
B	CATCTGATTCCTCC	R14-RR_ERCC_00042.1_150RC	RR	55
B	CTAAACCCCCTACT	R14-BR_RC-Sc3_B4	BR	56
B	CCTTTACAAACACA	RPTR14-GR_RS0247_H7	GR	57
B	ATACCACCCTCTTT	RPTR14-YY_Screen3_B8	YY	58
C	TATTCTTCTACCCC	RPTR14-YR_Screen4_I5	YR	59
C	TCTACCCTTCTCAT	R14-BG_RC-Sc3_D2	BG	60
C	CCACAATAACAACC	RPTR14-BR_Screen3_D4	BR	61
C	ACCTTAACATTCCC	R14-GG_RC-Sc3_D5	GG	62
C	ATTTCCCCTAACC	RPTR14-GR_Screen3_D7	GR	63
C	ACTTAAAACCCTCC	RPTR14-YY_Screen3_D8	YY	64
C	TACCTATTCTCCA	RPTR14-BB_Screen3_D1	BB	65
C	CCCCTTTCTCTAAG	RPTR14-RR_ERCC_00051.1_220	RR	66

[0593] 表6显示了另一组,其含有30种序列(SEQ ID NO:67-96)。可以将这30种序列拆分成三组,以产生10x10x10 14聚体报告物集合。

[0594] 表6

[0595]

报告物位置	14聚体序列	报告物名称	颜色	SEQ ID NO
A	GATGATGGTAGGTG	R14_PC_J2_BB_v2	BB	67
A	ATGAGAAGGGTAGA	R14_PC_D2_BG_v2	BG	68
A	GTTTTGTTGGTGAG	R14_PC_K2_BY_v2	BY	69
A	TTAGTGTGTTGGAG	R14_PC_K5_BR_v2	BR	70
A	ATGTAGGAGAGAGA	R14_PC_L1_GG_v2	GG	71
A	GGGAATGTTAAGGT	R14_PC_D5_GY_v2	GY	72
A	GGTAGTGGGAAAT	R14_PC_rcD7_GR_v2	GR	73
A	GGAGGGTTTTAAGT	R14_PC_rcD8_YY_v2	YY	74
A	GTAGTGTGGATGTT	R14_PC_J5_YR_v2	YR	75
A	CTTAGAGAAAGGGG	R14_PC_ERCC51_RR_v2	RR	76
B	GGAAGAGGATGAAA	R14_PC_K1_BB_v2	BB	77

B	AAGTTATGTGAGGG	R14_PC_spB_BG_v1	BG	78
B	GGAAAGTAGAGGAG	R14_PC_spB_BY_v1	BY	79
B	TTTTGGGTTTAGGG	R14_PC_spB_BR_v1	BR	80
B	AGATGTATGGGTGA	R14_PC_L2_GG_v2	GG	81
B	GATGGGAAAAAGTG	R14_PC_spB_GY_v1	GY	82
B	GGAGGAATCAGATG	R14_PC_spB_GR_v1	GR	83
B	AGAGGGATTGATGA	R14_PC_J4_YY_v2	YY	84
B	TGTGTTTGTAAGG	R14_PC_spB_YR_v1	YR	85
B	AAGGAGTGATAGGA	R14_PC_J1_RR_v2	RR	86
C	TGGTGATTTAGAGG	R14_J3_BB_v2	BB	87
C	GGGGTAGAAGAATA	R14_rcI5_BG_v2	BG	88
C	AAGAAAAGTGGGG	R14_PC_spA_BY_v1	BY	89
C	TATGTTGTGGGTAG	R14_PC_spA_BR_v1	BR	90
C	GTTAAAGGGAGGTT	R14_K3_GG_v2	GG	91
C	TGGGGTTTATATGG	R14_PC_spA_GY_v1	GY	92
C	AGGGAATATGGAGA	R14_K6_GR_v2	GR	93
C	TAGGTTGAGAATAG	R14_PC_spA_YY_v1	YY	94
C	TTTAAAAGAGGGGG	R14_PC_spA_YR_v1	YR	95
C	TGAGGTAAGATTGG	R14_PC_spA_RR_v1	RR	96

[0596] 在筛选后,在实验上验证了8x8x8 12聚体、8x8x8 14聚体和10x10x10 14聚体报告物集合。对于8x8x8 12聚体结合方案,使用Hyb & Seq原型执行验证,以记录10个测序循环。在长和短工作流程方法两者中使用三个报告探针池。测试所有512种可能的测序探针条形码结构域。表7显示长和短工作流程方法的实验步骤。

[0597] 表7

一个循环中的步骤	长工作流程： 不含立足点的报告物杂交	短工作流程： 含有立足点的报告物杂交
1	报告物 1 共 15s、30s 或 60s	报告物 1 共 30s
2	成像	成像
3	立足点 1 共 60s 以变暗	报告物 2 + 立足点 1 共 15s
4	成像	成像
[0598] 5	报告物 2 共 15s、30s 或 60s	报告物 3 + 立足点 2 共 15s
6	成像	成像
7	立足点 2 共 60s 以变暗	洗涤
8	成像	成像
9	报告物 3 共 15s、30s 或 60s	
10	成像	
11	洗涤	
12	成像	

[0599] 长工作流程实验导致>97%的变暗效率。对于短工作流程实验,假设变暗差不多有效,然而预计小频率的非变暗报告物将在每个图像中遗留,并且被误识别为新报告物。事实上,短工作流程实验中的最高条形码计数为YYYYYY,其很可能是非变暗和背景的假象。与长工作流程相比,8x8x8 12聚体报告物集合的性能在短工作流程中一般更低。与长工作流程相比,报告复合物一(其与条形码结构域的位置R1结合)和报告复合物三(其与条形码结构域的位置R3结合)在短工作流程中具有更低的效率。这对于报告复合物三是预计的,因为它包括八个另外的立足点寡核苷酸,以各2.5 μM的高浓度,其可以干扰报告物杂交。报告复合物一应该在两种工作流程之间表现类似,因为在短或长工作流程中没有使用立足点来去除第一报告复合物。对于所有三种报告探针,与长工作流程相比,总误差在短工作流程中也更高(1.3至2倍)。

[0600] 通过测试与所有512个可能的测序探针条形码结构域杂交的效率、特异性和速度,来验证8x8x8 14聚体报告物集合。将测序探针条形码结构域直接固定到Hyb & Seq测序盒的玻璃上。8x8x8 14聚体报告探针在仅15秒内杂交,伴随88%的平均效率与5.1%的平均错误率。这种错误的大部分是由于报告物的不正确鉴定,而不是由于不正确的杂交。报告物的错分类误差仍然是报告物错误的最大组分。

[0601] 通过测试与30个互补、截短的测序探针条形码结构域杂交的效率、特异性和速度,来验证10x10x10 14聚体报告物集合。每个条形码结构域仅含有一个报告物结合位点。将这些条形码结构域直接固定到Hyb & Seq测序盒的玻璃上。10x10x10 14聚体报告物集合在仅15秒内杂交,伴随90%的平均效率与5.0%的平均错误率。再次,绝大多数的错误是由于报告物的不正确鉴定,而不是由于不正确的杂交。

[0602] 实施例12 - 标准和三部分测序探针的设计和测试

[0603] 测序探针的靶结合结构域和条形码结构域由双链“茎”分开。图2显示了经过实验测试的两种测序探针结构。在标准测序探针上,靶结合结构域和条形码结构域存在于同一

寡核苷酸上,所述寡核苷酸与茎寡核苷酸结合,以产生36个核苷酸长的双链区域。使用这种结构,探针池中的每种测序探针都使用相同的茎序列。在三部分探针上,靶结合结构域和条形码结构域是分开的DNA寡核苷酸,其由36个核苷酸茎寡核苷酸结合在一起。为了防止条形码结构域的可能交换,每个条形码具有独特的茎序列,并且在合并测序探针之前分开进行杂交。

[0604] 图22显示了将三部分测序探针与标准测序探针比较所执行的一系列实验的结果。这些实验确认了,三部分测序探针经整个测序循环而存活,其中两种配置的所有读数的约80%包括第三报告探针的检测。当与标准测序探针相比时,三部分探针显示了少约12%的计数。为了研究条形码结构域寡核苷酸的交换倾向,将高浓度的含有相同茎序列的短替代寡核苷酸加入反应中。结果指示了,约13%的检测到的三部分测序探针已交换条形码寡核苷酸。需要通过掺入独特的茎序列来减轻寡核苷酸交换。尽管性能略有下降,但三部分探针提供了设计灵活性、快速寡核苷酸合成和降低成本的益处。

[0605] 实施例13 - 锁核酸取代在靶结合结构域中的效应

[0606] 如下测试了将锁核酸(LNA)取代到测序探针的靶结合结构域内的效应。使测序探针与报告探针在溶液中杂交,并且纯化适当形成的测序探针-报告探针。然后使测序探针-报告探针与合成靶核酸在溶液中杂交,并且装载到原型测序盒上。合成靶核酸由50个核苷酸组成并且是生物素化的。测序探针个别地或在九种测序探针的池中进行测试。对于九种测序探针的池,探针设计为沿着靶核酸的长度结合。为了分析,通过试验板器械将整个反应沉降到链霉抗生物素蛋白包被的盖玻片上,然后进行流动拉伸。然后使用适当的器械和软件,例如用NanoString nCounter[®]器械和软件,对报告探针进行成像且计数。

[0607] 每种测序探针含有10个核苷酸的靶结合结构域(SEQ ID NO:97)。使靶结合结构域内的LNA取代包括在图23中所示的位置处的2、3或4个LNA碱基。图23显示了,个别测序探针对于靶核酸的结合亲和力随着LNA碱基的数目增加而增加。重要的是,图23显示了LNA碱基的掺入并未降低序列探针结合的特异性。测试了九种测序探针的池,以确定当探针可以竞争靶结合时的碱基覆盖率。图24显示了,当将单个LNA探针引入池内时,受影响碱基的覆盖率增加,伴随对周围探针的结合的很小作用。这些结果指示了,LNA碱基取代可以改善碱基敏感性而不减少特异性。

[0608] 实施例14 - 修饰的核苷酸和核酸类似物取代在靶结合结构域中的效应

[0609] 如下测试各种修饰的核苷酸和核酸类似物,包括锁核酸(LNA)、桥接核酸(BNA)、丙炔修饰的核酸、拉链核酸(ZNA[®])、异鸟嘌呤和异胞嘧啶,取代到测序探针的靶结合结构域内的效应。将长度为50个核苷酸的生物素化的靶核酸装载到原型测序盒的链霉抗生物素蛋白盖玻片上。然后将测序探针和报告探针序贯地引入样品腔室内,并且使用Hyb & Seq原型器械进行成像。处理图像以比较每种不同测序探针的计数。使测序探针的10个核苷酸(SEQ ID NO: 99)的靶结合结构域中的取代包括图25中所示的位置处的LNA、BNA、丙炔和ZNA[®]碱基。图25显示了含有LNA和BNA的探针显示结合亲和力中的最大增加,同时维持了特异性,如通过对于匹配靶和错配靶检测到的计数数目所指示的。这些结果指示了,LNA或BNA碱基取代可以改善碱基敏感性,而不减少特异性。

[0610] 实施例15 - 确定本公开内容的测序方法的准确率

[0611] 图26描绘了来自量化本公开内容的测序方法的原始特异性的实验的结果。在该实

验中,执行测序反应,其中将4种不同测序探针的池与包括NRAS外显子2(SEQ ID NO:1)的片段的靶核酸杂交。每种测序探针(条形码1至4)具有等同的靶结合结构域,除了靶结合结构域的六聚体在 b_5 位置处不同之外,如图26的上图中描绘的。在该实例中,条形码4是正确的测序探针。在测序探针杂交后,使报告探针与条形码结构域的三个位置(R_1 、 R_2 和 R_3)各自序贯地杂交,并记录相应的荧光数据。图26的中图描绘了对于三个条形码结构域位置记录每种颜色组合的次数,以及记录正确组合的时间百分比。在 R_1 处的颜色组合在96%的时间被正确鉴定,在 R_2 处的颜色组合在97%的时间被正确鉴定,而在 R_3 处的正确颜色组合在94%的时间被正确鉴定。如图26的下图中描绘的,这导致94%的总体原始特异性。可以解释错误识别的条形码结构域位置的错误来源包括:(a)报告探针与流动室表面的非特异性结合,以及(b)报告探针的不正确杂交。报告物杂交错误的估计量为大约2至4%。

[0612] 图27显示了当通过多于一种测序探针对靶核酸中的核苷酸进行测序时,来自确定本公开内容的测序方法的准确率的实验的结果。如图27的上图中所示,在该实例中的靶核酸是NRAS外显子2(SEQ ID NO:1)的片段。特定的目的碱基是在靶核酸中突出显示的胞嘧啶(C)。目的碱基将与两种不同的测序探针杂交,每种探针具有与靶核酸杂交的独特足迹。在该实例中,测序探针1至4(条形码1至4)结合目的碱基左侧的3个核苷酸,而测序探针5至8(条形码5至8)结合目的碱基左侧的5个核苷酸。图27的中图显示了在测序探针的条形码结构域的每个位置处记录特异性颜色组合的次数。在图像量化并使用图17中描绘的碱基识别技术后,可以记录到约98.98%的平均准确率。

[0613] 实施例16:使用基于CRISPR的片段化捕获靶核酸

[0614] 如下使用c3探针复合物(上文所述)和c5探针复合物(上文所述),捕获靶核酸并将其固定到基底上:使用基于CRISPR的片段化,使基因组DNA(gDNA)片段化。使包含c3探针复合物和c5探针复合物的溶液与片段化的gDNA混合温育,以使c3和c5探针复合物与靶核酸杂交,使得形成图34中描绘的复合物。然后使靶核酸-捕获探针复合物与HiFi Taq连接酶和FEN1的混合物一起在65°C下温育。通过FEN1去除5'突出的瓣结构,并且将靶核酸分子的3'末端连接到包含基底特异性结构域的c5探针复合物的链。在温育后,使靶核酸-捕获探针复合物与20 μ l F样珠一起伴随搅动温育10分钟。用磁体捕获F状珠并洗涤三次。为了从F样珠中洗脱靶核酸-捕获探针复合物,将F样珠悬浮于20 μ l的USER酶混合物中,并且在25°C下温育15分钟。用磁体去除珠,并且使用SPRI珠,进一步纯化所得的含有洗脱的靶核酸-捕获探针复合物的上清液。将36 μ l的SPRI珠与上清液混合,并且在25°C下温育10分钟。然后使珠与磁体结合,并且用100 μ l的80%乙醇洗涤两次。然后去除乙醇,并且允许珠在25°C下干燥五分钟。在干燥后,将珠重悬浮于10 μ l 0.1X SSPE中。从SPRI珠中洗脱纯化的核酸-捕获探针复合物10分钟。在洗脱后,将衍生自不同生物样品的靶核酸-捕获探针复合物合并在一起,并且加入F珠中。使合并的复合物与F珠一起在25°C下温育10分钟。将F珠洗涤三次,且重悬浮于8 μ l中。然后将F珠转移至包含核酸分子的卡,所述核酸分子与c5探针复合物的基底特异性结构域互补,并且附着至基底。使附着至基底的互补核酸分子与基底特异性结构域杂交,从而将靶核酸固定到基底上。然后将基底暴露于UV光,以切割c3探针复合物上的光可切割接头,从而从F珠中释放靶核酸-捕获探针复合物。

[0615] 实施例17 - 基于CRISPR的gDNA片段化

[0616] 下述方案可以用于使用基于CRISPR的片段化步骤,使基因组DNA片段化。该方案使

得能够在体外切割双链靶DNA,以生成可以用于本公开内容的方法中的靶核酸分子。

[0617] 下述方案使用了Cas9 RNP复合物,其含有多重单引导RNA (sgRNA) 和V3-Cas9核酸酶两者。在该非限制性实例中,引导RNA在分子的5' 末端处含有与靶序列相对应的17-20个核苷酸,其与期望被切割的dsDNA恰好匹配。

[0618] 该方案在无RNA酶的环境中执行:(1) 使含有各sgRNA分子的管离心,以确保在管的底部处收集到干燥的RNA沉淀。在该实例中,154种单独的sgRNA用于进行154次切割,导致77种靶核酸的生成;(2) 将各sgRNA沉淀溶于无核酸酶的1X TE缓冲液中,使得达到50 μM 的最终浓度;(3) 然后通过将5 μl 来自步骤(2)的每种样品混合在一起,将所有sgRNA合并在一起。

[0619] 在将sgRNA合并在一起之后,该方案进一步包括:(4) 以表9和表10中所述的等摩尔量组合sgRNA和Cas9酶。为了确保关于每个切割位点的良好切割效率,对于待切割的每一种1000 ng dsDNA,使用1.0 pmol的sgRNA和Cas9。

表 9

[0620]	量/1 个切割位点	sgRNA	1 pmol
		Cas9	1 pmol
		10xCas9 缓冲液	0.3 μL
		100xBSA	0.1 μL
		DTT(1M)	0.2 μL
		dH ₂ O(加入至 10 μL 的最终体积)	X μL

表 10

[0620]	用于 154 个切割位点的量	sgRNA1-154(各 50 μM)	3.1 μL
		Cas9(61 μM)	2.5 μL
		10xCas9 缓冲液	0.3 μL
		100xBSA	0.1 μL
		DTT(1M)	0.2 μL
		dH ₂ O	3.8 μL

[0621] 在将sgRNA和Cas9酶混合后,该方案进一步包括:(5) 使sgRNA-Cas9混合物在37°C下温育10分钟,以允许RNP复合物的形成;(6) 然后,根据表11中所述的量,使1000 ng提取的人类基因组DNA (hgDNA) 与步骤(5)中形成的RNP复合物在室温下混合。

[0622] 表11.

[0623]	1 μg hgDNA(77 个靶核酸)	NEB3 缓冲液	4 μL
		DTT	2 μL
		CRISPR154	10 μL
		hgDNA	1000 ng
		ddH ₂ O(加入至 10 μL 的最终体积)	X μL

[0624] 在将hgDNA与RNP复合物混合后,该方案进一步包括:(7) 使在步骤(6)中组装的反应在37°C下温育60分钟;(8) 通过使来自步骤(7)中的反应在70°C下温育10分钟,使CRISPR

切割失活;并且(9)将1 μ L的蛋白酶K(20 mg/mL储备溶液)加入来自步骤(8)的反应中,并且使该混合物在56°C下温育10分钟,以从Cas9核酸内切酶中释放DNA底物。可替代地,可以将2 μ L Qiagen蛋白酶(1.07 AU/mL)加入来自步骤(8)的反应中,随后为在37°C下的30分钟温育,随后为在70°C下15分钟的第二次温育,以使Qiagen蛋白酶加热失活。

[0625] 实施例18 - 从已使用基于CRISPR的片段化步骤和USER介导的切割片段化的DNA中纯化靶核酸

[0626] 在基于CRISPR的片段化后,下述方案可以用于使用本公开内容的探针和方法来捕获且纯化特异性靶核酸。

[0627] 在从样品中提取的DNA基于CRISPR的片段化之后,将片段化的DNA与捕获探针混合。在该非限制性实例中,将片段化的DNA与两种捕获探针混合。第一捕获探针包含与特异性靶核酸的5'末端互补的结构域和亲和力部分。在该实例中,亲和力部分是F标签。可切割部分位于F标签以及与靶核酸5'末端互补的结构域之间。在该实例中,可切割部分是USER序列。第二捕获探针包含与靶核酸的3'末端互补的结构域。第二捕获探针可以任选地包括亲和力部分。使1 nM的捕获探针与20 fM的靶核酸在5X SSPE缓冲液中混合至100 μ l的最终反应体积。

[0628] 为了使捕获探针与靶核酸杂交,首先使片段化的DNA和探针的溶液在98°C下温育3分钟,然后在65°C下温育15分钟,且最后在22°C下温育,直至方案中的下一步。

[0629] 为了在杂交之后捕获靶核酸-捕获探针复合物,使CRISPR片段化的DNA和捕获探针的溶液,与F珠(其能够结合位于第一捕获探针上的F-标签)一起,在室温下伴随转动或手动混合温育10分钟。在温育后,将珠用1 mL 0.1X SSPE缓冲液洗涤三次,以去除与珠并未特异性结合的DNA。将珠重悬浮于100 μ l的1X SSPE中,转移到PCR管中,并且离心。去除上清液,并且将珠重悬浮于10 μ l的USER主混合物中,所述USER主混合物含有1X Cutsmart缓冲液和1 μ l的USER酶。使珠在USER主混合物中在室温下温育15分钟,从而切割第一捕获探针中的USER序列,从珠中释放靶核酸。使珠离心,并且收集含有靶核酸的上清液。

[0630] 为了进一步纯化靶核酸,执行两次AMPure纯化。在AMPure纯化中,将含有靶核酸的上清液与AMPure XP珠(固相可逆固定珠,SPRI)以1:1.8的样品/珠比率混合。使该混合物温育10分钟,以允许靶核酸与珠的结合。然后将珠用200 μ l的75-80%乙醇洗涤两次。在每次洗涤过程中,使用磁体将珠拉过乙醇两次。然后使珠在磁体上干燥5分钟。然后将珠重悬浮于10 μ l的0.1X SSPE溶液中,且温育10分钟。最后,从珠中洗脱纯化的靶核酸。

[0631] 实施例19 - 从已使用基于CRISPR的片段化步骤和USER介导的切割片段化的DNA中纯化和沉积靶核酸

[0632] 在基于CRISPR的片段化后,下述方案可以用于使用本公开内容的探针和方法,将特异性靶核酸捕获、纯化且沉积到合适的基底上。

[0633] 在从样品中提取的DNA基于CRISPR的片段化之后,将片段化的DNA与捕获探针混合。在该非限制性实例中,如图34中描绘的,将片段化的DNA与c3探针复合物和c5复合物混合,不同之处在于将位于C5探针上的F样标签替换为G标签。将1 nM的捕获探针、20 fM的靶核酸、1 μ l的HiFi DNA连接酶和2 μ l的FEN1酶,在1X HiFi缓冲液中混合至100 μ l的最终反应体积。

[0634] 为了使捕获探针与靶核酸杂交,切割由靶核酸和c5探针复合物形成的5'-突出的

瓣结构,并且将靶分子的3'末端连接到包含基底特异性结构域的c3探针的链,首先使片段化的DNA和探针的溶液在98°C下温育3分钟,然后在65°C下温育60分钟,且最后在22°C下温育,直至方案中的下一步。在65°C温育的最后十分钟内,添加10 μ l的5M NaCl,以使盐浓度达到500 mM,并且帮助稳定捕获探针的杂交。

[0635] 为了在杂交、切割和连接之后捕获靶核酸-捕获探针复合物,使片段化的DNA和捕获探针的溶液与G珠(其能够结合G标签)一起,在室温下伴随转动或手动混合温育10分钟。在温育后,然后将珠用1 mL 0.1X SSPE缓冲液洗涤三次,以去除与珠并未特异性结合的DNA。将珠重悬浮于100 μ l的1X SSPE中,转移到PCR管中,并且离心。去除上清液,并且将珠重悬浮于10 μ l的USER主混合物中,所述USER主混合物含有1X Cutsmart缓冲液和1 μ l的USER酶。使珠在USER主混合物中在室温下温育15分钟,从而切割第一捕获探针中的USER序列,从珠中释放靶核酸。使珠离心,并且收集含有靶核酸的上清液。

[0636] 为了进一步纯靶核酸,执行两次AMPure纯化。在AMPure纯化中,将含有靶核酸的上清液与AMPure XP珠(固相可逆固定珠,SPRI)以1:1.8的样品/珠比率混合。使该混合物温育10分钟,以允许靶核酸与珠的结合。然后将珠用200 μ l的75-80%乙醇洗涤两次。在每次洗涤过程中,使用磁体将珠拉过乙醇两次。然后使珠在磁体上干燥5分钟。然后将珠重悬浮于10 μ l的0.1X SSPE溶液中,且温育10分钟。最后,从珠中洗脱纯化的靶核酸。然后可以将纯化的靶核酸沉积到能够与连接到靶核酸上的基底特异性结构域结合的基底上。

[0637] 实施例20 - 从已使用基于CRISPR的片段化步骤片段化的DNA中纯靶核酸

[0638] 首先,制备100 μ l杂合-捕获反应,其中使用DEPC处理/无核酸酶的水,将预片段化的gDNA与捕获探针在5X SSPE缓冲液中混合。在将预片段化的gDNA加入杂合-捕获反应之前,使用Qubit dsDNA BR测定试剂盒(Thermo Scientific),测量预片段化的gDNA的浓度。Qubit测定的定量范围为2-1000 ng。优选地,应该使用Qubit测定来测定10-100ng的gDNA样品。如果认为gDNA样品具有大于1000 ng/ μ l的浓度,则应该首先将样品稀释至约10-1000 ng/ μ l。总的来说,应该将100至1000 ng的gDNA加入杂合-捕获反应中,优选1000 ng。

[0639] 在该非限制性实例中,两种捕获探针用于每种待纯化的靶核酸。第一捕获探针包含与特异性靶核酸的5'末端互补的结构域和亲和力部分。在该实例中,亲和力部分是F标签。第二捕获探针包含与靶核酸的3'末端互补的结构域。第二捕获探针可以任选地包括亲和力部分。首先,制备1 μ M/探针的储备溶液。将储备溶液稀释至50 nM/探针工作溶液。将2 μ l的50 nM/探针工作溶液加入最终的100 μ l杂合-捕获反应中。

[0640] 为了使捕获探针与靶核酸杂交,使杂合-捕获反应在98°C下温育3分钟,随后为在65°C下的15分钟温育。在65°C温育后,反应可以贮存于22°C下,直至方案中的下一步。

[0641] 通过用1 ml的5X SSPE洗涤两次来制备能够结合F标签的F珠。在洗涤后,使珠离心,去除上清液,并且将珠重悬浮于包含5X SSPE、60%甲酰胺和0.1% Tween-20的缓冲液中。

[0642] 大致上,预计F珠在10 mg/ml样品浓度下具有1.25 pmol/ μ L的结合能力。优选使用1.5X的F珠结合能力,如通过反应中存在的第二捕获探针的总摩尔确定的。例如,在旨在捕获77种不同靶核酸的反应中,使用0.1 pmol的每种第二探针(对应于每种靶核酸),这意味着存在7.7 pmol的总第二探针。7.7 pmol的1.5X为11.55 pmol,并且由于F珠的结合能力计算为1.25 pmol/ μ l,因此应该使用9.24 μ L的珠。

[0643] 在捕获探针杂交后,使100 μ l杂合-捕获反应与100 μ l重悬浮的F珠一起在室温下

温育10分钟,同时进行转动。在使捕获探针-靶核酸复合物与珠结合后,洗涤珠以去除未捕获的gDNA。为了洗涤珠,首先将珠用磁体浓缩,然后用1 ml的0.1X SSPE洗涤三次。在最后一次洗涤后,去除900 μ l的最终洗涤溶液,并且将珠重悬浮于剩余的100 μ l中。将重悬浮的珠转移到净PCR管中,使用磁体浓缩并去除上清液。

[0644] 为了从珠中洗脱捕获探针-靶核酸复合物,将珠重悬浮于12 μ l的0.1X SSPE中,并且在45°C下温育7.5分钟。在温育后,珠快速聚集在磁体上,并且取出12 μ l洗出液。

[0645] 使用两次SPRI (AMPure) 纯化进一步纯靶核酸。将12 μ l洗出液与AMPure珠以1:1.8的洗出液/珠比混合,然后在室温下在旋转器上温育10分钟。然后将珠用磁体聚集,并且用100 μ l的80%乙醇洗涤两次。每次洗涤使珠移动穿过磁体两次,以确保乙醇洗涤整个珠沉淀。在第二次乙醇洗涤后,去除尽可能多的乙醇,并且允许珠在无盖管中的磁体上干燥5分钟。然后将珠重悬浮于12 μ l的0.1X SSPE中,并且在室温下洗脱10分钟。将珠聚集并取出洗出液。在第二次SPRI纯化后,将1 μ l 20X SSPE缓冲液加入最终洗出液中,以稳定捕获探针-靶核酸复合物。

[0646] 实施例21 - 使用本公开内容的测序探针对靶核酸进行测序

[0647] 下述是描述使用两种不同的测序探针设计对合成靶核酸进行测序的实例。

[0648] 第一测序探针设计,在本文中称为LG-间隔的测序探针,显示于图53中。从5' 末端开始,测序探针包含靶结合结构域。靶结合结构域包含六核苷酸(6聚体)长的序列((b_1 - b_2 - b_3 - b_4 - b_5 - b_6)),其与靶核酸中的六个互补核苷酸特异性杂交。该6聚体序列在任一侧上侧翼为简并碱基(N)。测序探针还包含条形码结构域,其经由指定为L-DNA茎的双链DNA间隔区共价附着至靶结合结构域。条形码结构域和双链DNA间隔区两者均完全由L-DNA组成,而靶结合结构域完全由D-DNA组成。L-DNA茎为长度25个核苷酸。条形码结构域为长度27个核苷酸,并且包含三个附着区域(称为位点1、位点2和位点3)。每个附着位置包含9个核苷酸。每个附着区域的3'末端核苷酸是L-dG核苷酸。

[0649] 对于这些实验,设计了由512个不同种类的LG-间隔的测序探针组成的池(LG-间隔的池),其中每种测序探针包含位点1、位点2和位点3核酸序列的独特组合。对于附着区域设计了总共30种独特的、低复杂性的核酸序列,其中10种设计用于位点1中,10种设计用于位点2中,且10种设计用于位点3中。LG-间隔的测序探针的每个不同种类还包含在靶结合结构域中不同的6聚体序列。

[0650] 第二测序探针设计,在本文中称为D-口袋测序探针,显示于图54中。从5' 末端开始,测序探针包含靶结合结构域。靶结合结构域包含六核苷酸(6聚体)长的序列((b_1 - b_2 - b_3 - b_4 - b_5 - b_6)),其与靶核酸中的六个互补核苷酸特异性杂交。该6聚体序列在任一侧上侧翼为简并碱基(N)。测序探针还包含条形码结构域,其经由指定为D-DNA茎的双链DNA间隔区共价附着至靶结合结构域。D-DNA茎为长度25个的核苷酸。条形码结构域为长度74个的核苷酸,并且包含三个附着区域(称为位点1、位点2和位点3)。每个附着位置包含8个核苷酸。位点1和位点2通过双链DNA区域分开,所述双链DNA区域由与25个核苷酸长的口袋寡核苷酸杂交的条形码结构域组成。位点2和位点3也通过双链DNA区域分开,所述双链DNA区域由与25个核苷酸的口袋寡核苷酸杂交的条形码结构域组成。D-口袋测序探针完全由D-DNA组成。

[0651] 对于这些实验,设计了512个不同种类的D-口袋测序探针的两个不同池(D-口袋池1和D-口袋池3)。测序探针的每个不同种类包含在靶结合结构域中不同的6聚体序列。

[0652] 对于这些实验,D-口袋测序探针池中的一系列测序探针的靶结合结构域,等同于LG-间隔的测序探针池中的一系列测序探针的靶结合结构域,并且关于报告物附着位置1、2和3各自的颜色分配,在D-口袋测序探针和LG-间隔的测序探针之间是相同的。换言之,对于每种D-口袋测序探针,存在相应的LG-间隔的测序探针,具有相同的靶结合结构域序列和相应的三色报告物读出。

[0653] 为了测试LG-间隔的测序探针和D-口袋测序探针的效率和准确率,设计了合成靶核酸。从5'末端开始,合成靶核酸包含长度为25个核苷酸到45个核苷酸的单链测序窗口。这些测序窗口包含合成核苷酸序列,以及来自已知人类基因(例如P53、ALX1、SMO1-1)的核苷酸序列两者。测序窗口随后为35个核苷酸长的通用序列,在本文中称为35-8序列。在35-8序列之后为3'生物素-TEG修饰。合成靶核酸完全由D-DNA组成。合成靶核酸的示意图显示于图55中。

[0654] 为了捕获并固定用于下述实验的合成靶核酸,使用了70个核苷酸长的捕获探针。捕获探针包含与35-8序列互补的序列,允许捕获探针与合成靶核酸杂交。与35-8序列互补的序列完全由D-DNA组成。捕获探针还包含35个核苷酸长的聚L-dA序列。如图55中所示,在使捕获探针与合成靶核酸杂交后,捕获探针-靶核酸复合物可以固定至包括包含聚L-dT序列的草坪寡核苷酸的表面。另外,草坪寡核苷酸可以包含生物素分子。因此,如图55中所示,在捕获探针的聚L-dA序列与草坪寡核苷酸的聚L-dT序列杂交后,可以添加中性抗生物素蛋白,以在合成靶核酸和草坪寡核苷酸之间产生蛋白质锁,因为中性抗生物素蛋白结合了合成靶核酸的3'生物素以及位于草坪寡核苷酸上的生物素两者。

[0655] 在下述实验中,使合成靶核酸与捕获探针杂交,并且固定到有序阵列上。有序阵列含有由草坪寡核苷酸功能化的200 nm直径的垫。有序阵列的间距为800 nm。通过在固定的捕获探针-靶核酸复合物上移取0.2 mg/mL中性抗生物素蛋白溶液,并温育5分钟,可以执行合成靶核酸和草坪寡核苷酸之间的蛋白质锁定。随后通过使固定和蛋白质锁定的靶核酸在1 μ M生物素化的PEG溶液中温育一分钟,用双重生物素化的PEG封闭开放的中性抗生物素蛋白结合位点。最后,使用低盐(0.0033x SSPE)洗涤,从有序阵列中去除任何非蛋白质锁定的靶核酸。

[0656] 在靶核酸固定和蛋白质锁定之后,将测序探针池以200 nM/测序探针的浓度(在3.75xSSPE和15%甲酰胺缓冲液中)在8°C下流动到阵列上1分钟,以允许测序探针与靶核酸杂交。在使测序探针与靶核酸杂交后,在三轮序贯杂交中,在8.75xSSPE的缓冲条件下,使测序探针与报告探针杂交。这些实验中的报告探针包含在与附着区域结合的报告探针部分与包含可检测标记的部分之间的UV可切割接头。

[0657] 在第一轮中,使第一报告探针与测序探针的位点1杂交。在鉴定与位点1杂交的第一报告探针的可检测标记后,通过将样品暴露于UV光来切割第一报告探针,从而去除可检测标记。在第二轮中,使第二报告探针与测序探针的位点2杂交。在鉴定与位点2杂交的第二报告探针的可检测标记后,通过将样品暴露于UV光来切割第二报告探针,从而去除可检测标记。在最后的第三轮中,使报告探针与测序探针的位点3杂交。在鉴定与位点3杂交的第三报告探针的可检测标记后,使用低盐洗涤(0.0033x SSPE),使测序探针与固定的合成靶核酸脱离杂交(de-hybridized)。在脱离杂交后,可以使新的测序探针池与固定的合成靶核酸杂交。这些步骤构成了一个测序循环。

[0658] 在第一组实验中, LG-间隔的池用于在50个循环的测序运行中, 对几种不同的固定的合成靶核酸进行测序。在第二组实验中, D-口袋池用于在50个循环的测序运行中, 对几种不同的固定的合成靶核酸进行测序。这两组实验的结果显示于表12。

[0659] 表12. 测序结果

池	%条形码效率	净 3 观察者	净 2 观察者	净 1 观察者			
LG-间隔的 (2044)	41.1	12.3	10.4	13.9			
D-口袋(2073)	32.0	10.1	10.0	10.2			
池	净位点 1 效率	净位点 2 效率	净位点 3 效率	净% 3 位点 读出	%变暗的 多色	%有效 的	%无效的
LG-间隔的 (2044)	76.4	70.0	65.4	29.8	11.2	86.8	13.2
D-口袋(2073)	74.0	69.2	63.6	31.5	5.5	91.7	8.4

[0660] 存活截断且表决的

池	%条形码效率	净 3 观察者	净 2 观察者	净 1 观察者			
LG-间隔的 (2044)	46.9	15.2	11.8	14.4			
D-口袋(2073)	29.3	10.5	8.7	8.7			
池	净位点 1 效率	净位点 2 效率	净位点 3 效率	净% 3 位点 读出	%变暗的 多色	%有效 的	%无效的
LG-间隔的 (2044)	82.3	71.7	65.4	32.5	11.6	86.9	13.1
D-口袋(2073)	75.6	69.9	66.8	35.9	4.6	93.5	6.5

[0661] 表12显示了, 与D-口袋池相比, LG-间隔的池展示增加的条形码效率(47%相对于30%)。此外, 与D-口袋池相比, LG-间隔的池也展示增加的净位点1效率(82%相对于76%)。对于LG-间隔的池和D-口袋池两者, 净位点2和净位点3效率大致相同。

[0662] 此外, 图56-58显示了LG-间隔的池和D-口袋池显示了在预测的有问题的靶区域中增加的覆盖率。在图56-58中, 底部x轴显示了待测序的靶核酸的特异性核苷酸/位置。顶部图表显示了在LG-间隔和D-口袋测序实验中的理论和观察到的多样性/碱基。碱基的测序多样性是不同的独特探针种类的数目, 所述探针种类包含可以与靶核酸中的特异性核苷酸杂交的靶结合结构域。因此, 碱基的测序多样性是可以使用多少种不同的测序探针来查询(例如测序或鉴定)单个位置的量度。不受理论的束缚, 特定碱基的测序多样性越高, 在多重测序循环期间的预计覆盖率就越高, 因为存在碱基在每个循环中由探针结合且测序的更大机会。图56-58的顶部图表还显示了在LG-间隔和D-口袋测序实验中观察到的每个碱基的覆盖率。红色框指示了预测的有问题的区域。相对于D-口袋测序探针, LG-间隔的测序探针显示了在这些有问题的区域中的覆盖率增加, 其中在一些情况下, LG-间隔的测序探针在其中D-口袋测序探针未显示覆盖率的区域中显示了覆盖率。此外, 图59-61显示了来自使用LG-间隔的测序探针和D-口袋测序探针的更多实验的测序结果。如图59-61中所示, 与当使用D-口袋测序探针时相比, 当使用LG-间隔的测序探针时, 观察到的测序覆盖率和观察到的测序多样性两者均增加。

[0663] 图62中显示了比较LG-间隔的测序探针和D-口袋测序探针性能的更多实验结果。图62中每个直方图的x轴指出了在实验中记录的观察到的条形码事件数目(即,某种形式的可检测信号)。直方图的黑色区段具体指出了有效的3位点读出(即,其中有效鉴定了测序探针的三个附着区域中的每个)。Y轴指示了阵列上的特征(即,个别的固定的靶核酸)数目,其对于特定靶展示相应的条形码事件数目。图62显示了,与D-口袋池相比,LG-间隔的池显示出增加数目的总条形码事件和有效的3位点读出。

[0664] 图63中显示了比较LG-间隔的测序探针和D-口袋测序探针性能的更多实验结果。总的来说,使用LG-间隔的测序探针或D-口袋测序探针,对20种不同类型的合成靶核酸进行测序。图63中的图表显示了中靶事件,无效事件,脱靶,以及在靶结合结构域的任何个别碱基位置 b_1 、 b_2 、 b_3 、 b_4 、 b_5 和/或 b_6 处的1/6、2/6、3/6、4/6、5/6或6/6错配的总数。如图63中所示,LG-间隔的测序探针展示的中靶事件/脱靶事件比、以及中靶事件/无效事件比,是D-口袋测序探针的大约两倍。如图64中所示,当对各种不同的特异性合成靶核酸进行测序时,LG-间隔的测序探针的这种增加的性能是一致的。

[0665] 为了直接比较D-口袋测序探针和LG-间隔的测序探针,使用D-口袋测序探针和LG-间隔的测序探针的组合,对合成靶核酸进行测序。D-口袋池(D-口袋池1)用于前50个测序循环。然后,将相同的固定的合成靶核酸用LG-间隔的池进行50个循环测序。然后比较来自使用D-口袋测序探针的前50个循环和使用LG-间隔的测序探针的后50个循环的结果。这些结果显示于表13。

[0666] 表13. 测序结果

[0667]

池	%条形码效率	净3观察者	净2观察者	净1观察者			
D-口袋(循环1-50)	26.9	6.8	8.4	10.6			
LG-间隔的(循环51-100)	37.8	12.8	11.6	11.0			
池	净位点1效率	净位点2效率	净位点3效率	净% 3位点读出	%变暗的多色	%有效的	%无效的
D-口袋(循环1-50)	70.0	62.6	58.0	25.4	4.2	87.3	12.7
LG-间隔的(循环51-100)	73.3	74.5	66.7	33.9	6.4	86.2	13.8

[0668] 表13显示了,与D-口袋池相比,LG-间隔的池展示增加的条形码效率、净位点1效率、净位点2效率、净位点3效率和净3位点读出。

[0669] 此外,图65显示了1个观察者(可能的三个报告探针中仅一个被成功地记录)、2个观察者(可能的三个报告探针中仅两个被成功地记录)、以及3个观察者(所有三个可能的报告探针都被成功地记录)事件数目/循环。如图65中所示,1个观察者、2个观察者和3个观察者事件的数目都从循环51(对应于转换到LG-间隔的测序探针)开始增加。特别地,最大的增加在3个观察者事件的数目中发生。

[0670] 图66显示了在上述实验的每个循环中记录的中靶、新六聚体、冗余六聚体、脱靶和无效事件的数目。中靶事件是其中测序探针与靶核酸正确结合,并且所有三个报告探针都成功且正确鉴定的事件。新六聚体事件是其中新六聚体在单个固定的靶核酸中得到鉴定/测序的事件,其中该特定的六聚体在先前的循环中未得到鉴定。冗余六聚体事件是其中特定六聚体在单个固定的靶核酸中再次鉴定的事件。脱靶和无效事件是其中在测序探针和报告探针的杂交和/或检测中发生错误的事件。如图66最左侧的图中所示,中靶、新六聚体和冗余六聚体事件的数目都从循环51(对应于转换到LG-间隔的测序探针)开始增加。相应地,新六聚体事件的数目在循环51-100之间减少,因为在每个相继循环中存在靶核酸中尚未

得到鉴定的较少六聚体,并且在冗余六聚体事件的数目中可见相应的增加。

[0671] 在另一组实验中,单个LG-间隔的池和两个D-口袋池(D-口袋池1和D-口袋池3),用于对固定的合成靶核酸进行测序。LG-间隔的池用于前50个测序循环。然后,将相同的固定的合成靶核酸用D-口袋池1和D-口袋池3进行100个循环测序。将两个池交替使用:D-口袋池1用于循环51,D-口袋池3用于循环52,D-口袋池1用于循环53,依此类推等等。来自150个测序循环的结果显示于表14中。

[0672] 表14. 测序结果

[0673]

池	%条形码效率	净3观察者	净2观察者	净1观察者			
LG-间隔的池1(循环1-50)	36.6	11.3	12.4	11.7			
D-口袋池1(51-100的奇数循环)	13.1	2.8	4.1	5.9			
D-口袋池3(51-100的偶数循环)	11.8	2.6	3.7	5.4			
池	净位点1效率	净位点2效率	净位点3效率	净% 3位点读出	%变暗的多色	%有效的	%无效的
LG-间隔的池1(循环1-50)	66.6	69.9	67.0	30.9	3.3	88.1	12.0
D-口袋池1(51-100的奇数循环)	64.2	58.4	55.0	21.4	1.8	89.5	10.6
D-口袋池3(51-100的偶数循环)	64.1	58.3	56.1	21.8	2.0	88.7	11.3

[0674] 表14中显示的结果证实了,与D-口袋池相比,LG-间隔的池展示增加的条形码效率、净位点1效率、净位点2效率、净位点3效率和净3位点读出。

[0675] 总之,上述结果证实了,LG-间隔的测序探针和D-口袋测序探针两者均可以用于查询固定的靶核酸,其中与D-口袋测序探针相比,LG-间隔的探针显示了增加的覆盖率、效率和有效的条形码事件。

序列表

[0001]	<p><110> 纳米线科技公司 Dunaway, Dwayne Hoang, Margaret Gregory, Mark Kim, Dae Manrao, Elizabeth A Beechem, Joseph M Walsh, Matthew Khafizov, Rustem Korukonda, Sanghamithra Deng, Yi Meredith, Gavin McElwain, Mark Skene, Peter Burke, Cassandra</p> <p><120> 化学组合物及其使用方法</p> <p><130> NATE-039/001W0</p> <p><150> US 62/671,091</p> <p><151> 2018-05-14</p> <p><150> US 62/836,327</p> <p><151> 2019-04-19</p> <p><160> 114</p> <p><170> PatentIn version 3.5</p> <p><210> 1</p> <p><211> 38</p> <p><212> DNA</p> <p><213> 智人(Homo sapiens)</p> <p><400> 1</p> <p>accgtgacat gagaagaaca ggtcgacata ggtcatac</p> <p><210> 2</p> <p><211> 10</p> <p><212> DNA</p> <p><213> 人工序列</p> <p><220></p> <p><223> 实例一致序列</p> <p><400> 2</p> <p>aacaccacct</p> <p><210> 3</p> <p><211> 24</p> <p><212> DNA</p> <p><213> 智人(Homo sapiens)</p> <p><400> 3</p>	<p>38</p> <p>10</p>
--------	--	---------------------

	gtctagctac agtgaaatct cgat	24
	<210> 4	
	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 4	
	tcaaagtgct gggctccggt gcg	23
	<210> 5	
	<211> 27	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 5	
	tagttggagc tggtagcgta ggcaaga	27
	<210> 6	
	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 6	
	gaatgatgca catcatggtg gct	23
	<210> 7	
	<211> 23	
	<212> DNA	
[0002]	<213> 智人(Homo sapiens)	
	<400> 7	
	gcagctcatc acgcagctca tgc	23
	<210> 8	
	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 8	
	atacagctgg acaagaagag tac	23
	<210> 9	
	<211> 25	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 9	
	tggetcgaga tatcatgagt gattc	25
	<210> 10	
	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 10	
	tcatcatagg tcgtcatgct tat	23
	<210> 11	

	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 11	
	atggagtatg tgtctgtgga gac	23
	<210> 12	
	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 12	
	gtgcaactct ccgtacatcg tgg	23
	<210> 13	
	<211> 23	
	<212> DNA	
	<213> 智人(Homo sapiens)	
	<400> 13	
	cagccgctg ctgcacacca acg	23
	<210> 14	
	<211> 30	
	<212> RNA	
	<213> 智人(Homo sapiens)	
	<400> 14	
[0003]	gcuucagcac cacuccacuc cacauucua	30
	<210> 15	
	<211> 30	
	<212> RNA	
	<213> 智人(Homo sapiens)	
	<400> 15	
	uguccucaug uauuggucuc ucauggcacu	30
	<210> 16	
	<211> 30	
	<212> RNA	
	<213> 智人(Homo sapiens)	
	<400> 16	
	cuuucgcaa aguggaggag accaucgccg	30
	<210> 17	
	<211> 30	
	<212> RNA	
	<213> 智人(Homo sapiens)	
	<400> 17	
	gaaggugaug uuugggucag gugccuuagu	30
	<210> 18	
	<211> 50	
	<212> DNA	

	<213> 智人(Homo sapiens)	
	<400> 18	
	caaactggtg gtggttggag caggtggtgt tgggaaaagc gcactgacaa	50
	<210> 19	
	<211> 12	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 12 聚体附着位置	
	<400> 19	
	aggacagatg ac	12
	<210> 20	
	<211> 12	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 12 聚体附着位置	
	<400> 20	
	gtatcggatg ac	12
	<210> 21	
	<211> 12	
[0004]	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 12 聚体附着位置	
	<400> 21	
	aggagtgatg ac	12
	<210> 22	
	<211> 12	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 12 聚体附着位置	
	<400> 22	
	aggggtgagg ag	12
	<210> 23	
	<211> 12	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 12 聚体附着位置	
	<400> 23	
	agaggggatg ac	12

<210>	24	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	24	
agtggggagg ag		12
<210>	25	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	25	
agccgagatg ac		12
<210>	26	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	26	
agggatgatg ac		12
<210>	27	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	27	
tggatgaaa ag		12
<210>	28	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	28	
gaaggagaaa ag		12
<210>	29	
<211>	12	
<212>	DNA	
<213>	人工序列	

[0005]

	<220>		
	<223>	12 聚体附着位置	
	<400>	29	
	ggggatgaaa ag		12
	<210>	30	
	<211>	12	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	12 聚体附着位置	
	<400>	30	
	gtgagggaaa ag		12
	<210>	31	
	<211>	12	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	12 聚体附着位置	
	<400>	31	
	agccgagaaa ag		12
	<210>	32	
[0006]	<211>	12	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	12 聚体附着位置	
	<400>	32	
	cgagaggaaa ag		12
	<210>	33	
	<211>	12	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	12 聚体附着位置	
	<400>	33	
	gaggcgaaa ag		12
	<210>	34	
	<211>	12	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	12 聚体附着位置	
	<400>	34	
	agcgtgaaa ag		12

<210>	35	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	35	
tgagaagggt ag		12
<210>	36	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	36	
gttgttattg tg		12
<210>	37	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	37	
tttgggttta gg		12
<210>	38	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	38	
gtagtgggga aa		12
<210>	39	
<211>	12	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	12 聚体附着位置	
<400>	39	
atgggaaaaa gt		12
<210>	40	
<211>	12	
<212>	DNA	
<213>	人工序列	

[0007]

	<220>		
	<223> 12 聚体附着位置		
	<400> 40		
	gagttggatg ag		12
	<210> 41		
	<211> 12		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 12 聚体附着位置		
	<400> 41		
	atgttgtggg ta		12
	<210> 42		
	<211> 12		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 12 聚体附着位置		
	<400> 42		
	gagggtttta ag		12
	<210> 43		
[0008]	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 43		
	atcttttccc cact		14
	<210> 44		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 44		
	ccccactatt tctt		14
	<210> 45		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 45		
	ctaccacaa cata		14

<210>	46	
<211>	14	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	14 聚体附着位置	
<400>	46	
ccatataaac ccca		14
<210>	47	
<211>	14	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	14 聚体附着位置	
<400>	47	
aaactccaat ctcc		14
<210>	48	
<211>	14	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	14 聚体附着位置	
<400>	48	
ctattctcaa ccta		14
<210>	49	
<211>	14	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	14 聚体附着位置	
<400>	49	
ccccctcttt taaa		14
<210>	50	
<211>	14	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	14 聚体附着位置	
<400>	50	
ccaatcttac ctca		14
<210>	51	
<211>	14	
<212>	DNA	
<213>	人工序列	

[0009]

	<220>	
	<223> 14 聚体附着位置	
	<400> 51	
	ccctcacata actt	14
	<210> 52	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 52	
	ctcctctact ttcc	14
	<210> 53	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 53	
	ccctaaaccc aaaa	14
	<210> 54	
[0010]	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 54	
	cactttttcc catc	14
	<210> 55	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 55	
	catctgattc ctcc	14
	<210> 56	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 56	
	ctaaaccccc tact	14

	<210> 57	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 57	
	cctttacaaa caca	14
	<210> 58	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 58	
	ataccaccct cttt	14
	<210> 59	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0011]	<223> 14 聚体附着位置	
	<400> 59	
	tattetteta cccc	14
	<210> 60	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 60	
	tctacccttc tcac	14
	<210> 61	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 61	
	ccacaataac aacc	14
	<210> 62	
	<211> 14	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223> 14 聚体附着位置		
	<400> 62		
	accttaacat tccc		14
	<210> 63		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 63		
	atttcccact aacc		14
	<210> 64		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 64		
	acttaaaacc ctcc		14
	<210> 65		
[0012]	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 65		
	tacctattcc tcca		14
	<210> 66		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 66		
	cccccttctc taag		14
	<210> 67		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 67		
	gatgatggta ggtg		14

	<210> 68	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 68	
	atgagaaggg taga	14
	<210> 69	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 69	
	gtttgttg tgag	14
	<210> 70	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0013]	<223> 14 聚体附着位置	
	<400> 70	
	ttagtggtt ggag	14
	<210> 71	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 71	
	atgtaggaga gaga	14
	<210> 72	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 72	
	gggaatgta aggt	14
	<210> 73	
	<211> 14	
	<212> DNA	
	<213> 人工序列	

	<220>	
	<223> 14 聚体附着位置	
	<400> 73	
	ggttagtggg aaat	14
	<210> 74	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 74	
	ggagggtttt aagt	14
	<210> 75	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 75	
	gtagtggtgga tggt	14
	<210> 76	
	<211> 14	
[0014]	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 76	
	cttagagaaa gggg	14
	<210> 77	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 77	
	ggaagaggat gaaa	14
	<210> 78	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 78	
	aagttatgtg aggg	14

	<210> 79	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 79	
	ggaaagtaga ggag	14
	<210> 80	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 80	
	ttttgggttt aggg	14
	<210> 81	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0015]	<223> 14 聚体附着位置	
	<400> 81	
	agatgtatgg gtga	14
	<210> 82	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 82	
	gatgggaaaa agtg	14
	<210> 83	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 83	
	ggaggaatca gatg	14
	<210> 84	
	<211> 14	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223> 14 聚体附着位置		
	<400> 84		
	agaggattg atga		14
	<210> 85		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 85		
	tgtgtttgta aagg		14
	<210> 86		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 86		
	aaggagtgat agga		14
	<210> 87		
[0016]	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 87		
	tggtgattta gagg		14
	<210> 88		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 88		
	gggtagaag aata		14
	<210> 89		
	<211> 14		
	<212> DNA		
	<213> 人工序列		
	<220>		
	<223> 14 聚体附着位置		
	<400> 89		
	aagaaatagt gggg		14

	<210> 90	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 90	
	tatgttgtgg gtag	14
	<210> 91	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 91	
	gttaaaggga gggt	14
	<210> 92	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0017]	<223> 14 聚体附着位置	
	<400> 92	
	tggggtttat atgg	14
	<210> 93	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 93	
	agggaatatg gaga	14
	<210> 94	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 94	
	taggttgaga atag	14
	<210> 95	
	<211> 14	
	<212> DNA	
	<213> 人工序列	

	<220>	
	<223> 14 聚体附着位置	
	<400> 95	
	tttaaaagag gggg	14
	<210> 96	
	<211> 14	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 14 聚体附着位置	
	<400> 96	
	tgaggtaaga ttgg	14
	<210> 97	
	<211> 10	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 靶结合结构域	
	<400> 97	
	agcttcttat	10
	<210> 98	
[0018]	<211> 50	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 靶核酸	
	<400> 98	
	gagactagca gaataggagg ataagaagct cgaagagtga ggacaaatgg	50
	<210> 99	
	<211> 10	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 靶结合结构域	
	<400> 99	
	ttcactgtag	10
	<210> 100	
	<211> 35	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 实例性草坪寡核苷酸	
	<220>	
	<221> misc_feature	

	<222> (1)..(1)	
	<223> 用具有 6 碳接头的 5' 胺修饰	
	<400> 100	
	tggtgagggtt gttggtagta gtgagtttgt aggggt	35
	<210> 101	
	<211> 25	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 实例性草坪寡核苷酸	
	<220>	
	<221> misc_feature	
	<222> (1)..(1)	
	<223> 用具有 6 碳接头的 5' 胺修饰	
	<400> 101	
	tggtgagggtt gttggtagta gtgag	25
	<210> 102	
	<211> 35	
	<212> DNA	
	<213> 人工序列	
	<220>	
[0019]	<223> 实例性草坪寡核苷酸	
	<220>	
	<221> misc_feature	
	<222> (1)..(1)	
	<223> 用具有 6 碳接头的 5' 胺修饰	
	<400> 102	
	tttttttttt tttttttttt tttttttttt ttttt	35
	<210> 103	
	<211> 36	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 实例性草坪寡核苷酸	
	<220>	
	<221> misc_feature	
	<222> (1)..(1)	
	<223> 用具有 6 碳接头的 5' 胺修饰	
	<400> 103	
	catctcaaac accttctaca atatgacctt acacac	36
	<210> 104	
	<211> 36	
	<212> DNA	
	<213> 人工序列	

	<220>		
	<223>	实例性草坪寡核苷酸	
	<220>		
	<221>	misc_feature	
	<222>	(1)..(1)	
	<223>	用具有 6 碳接头的 5' 胺修饰	
	<400>	104	
		gtgatgggta taagagggtg tgatatatt atagta	36
	<210>	105	
	<211>	36	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	实例性草坪寡核苷酸	
	<220>		
	<221>	misc_feature	
	<222>	(1)..(1)	
	<223>	用具有 6 碳接头的 5' 胺修饰	
	<400>	105	
		tattgatatt gagaaagcgt ttgatgatgt attgat	36
	<210>	106	
	<211>	36	
[0020]	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	实例性草坪寡核苷酸	
	<220>		
	<221>	misc_feature	
	<222>	(1)..(1)	
	<223>	用具有 6 碳接头的 5' 胺修饰	
	<400>	106	
		tagttatgta gtagtttgcg aaagagttat agttat	36
	<210>	107	
	<211>	36	
	<212>	DNA	
	<213>	人工序列	
	<220>		
	<223>	实例性草坪寡核苷酸	
	<220>		
	<221>	misc_feature	
	<222>	(1)..(1)	
	<223>	用具有 6 碳接头的 5' 胺修饰	
	<400>	107	
		actaccctac tctacccttc taagatatac atatac	36

- <210> 108
 <211> 35
 <212> DNA
 <213> 人工序列
 <220>
 <223> 实例性草坪寡核苷酸
 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> 用具有 6 碳接头的 5' 胺修饰
 <220>
 <221> misc_feature
 <222> (3)..(3)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature
 <222> (7)..(7)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature
 <222> (11)..(11)
 <223> 包含异鸟嘌呤碱基
 <220>
 [0021] <221> misc_feature
 <222> (15)..(15)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature
 <222> (18)..(18)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature
 <222> (21)..(21)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature
 <222> (25)..(25)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature
 <222> (29)..(29)
 <223> 包含异鸟嘌呤碱基
 <220>
 <221> misc_feature

	<222> (33).. (33)	
	<223> 包含异鸟嘌呤碱基	
	<400> 108	
	tggtgagggtt gttggtagta gtgagtttgt aggggt	35
	<210> 109	
	<211> 25	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> OA-序列	
	<400> 109	
	cgaaagccat gacctccgat cactc	25
	<210> 110	
	<211> 60	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 连接子寡核苷酸	
	<220>	
	<221> misc_feature	
	<222> (28).. (28)	
[0022]	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (32).. (32)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (36).. (36)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (40).. (40)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (43).. (43)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (46).. (46)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	

	<221> misc_feature	
	<222> (50)..(50)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (54)..(54)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<220>	
	<221> misc_feature	
	<222> (58)..(58)	
	<223> 包含 5-甲基-异胞嘧啶碱基	
	<400> 110	
	gagtgatcgg aggtcatggc tticgaccct acaaactcac tactaccaac aacctcacca	60
	<210> 111	
	<211> 63	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 捕获探针	
	<400> 111	
	ccggtcaacc gttttgtaga acaactcccg tcccctcaet cactagcctc cagtaccgaa	60
[0023]	agc	63
	<210> 112	
	<211> 15	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 单链纯化序列	
	<400> 112	
	aacatcacac agacc	15
	<210> 113	
	<211> 15	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 单链纯化序列	
	<400> 113	
	gtctatcadc acagc	15
	<210> 114	
	<211> 34	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 草坪寡核苷酸	

[0024] <220>
<221> misc_feature
<222> (1).. (34)
<223> L-DNA
<400> 114
tttttttttt tttttttttt tttttttttt tttt 34

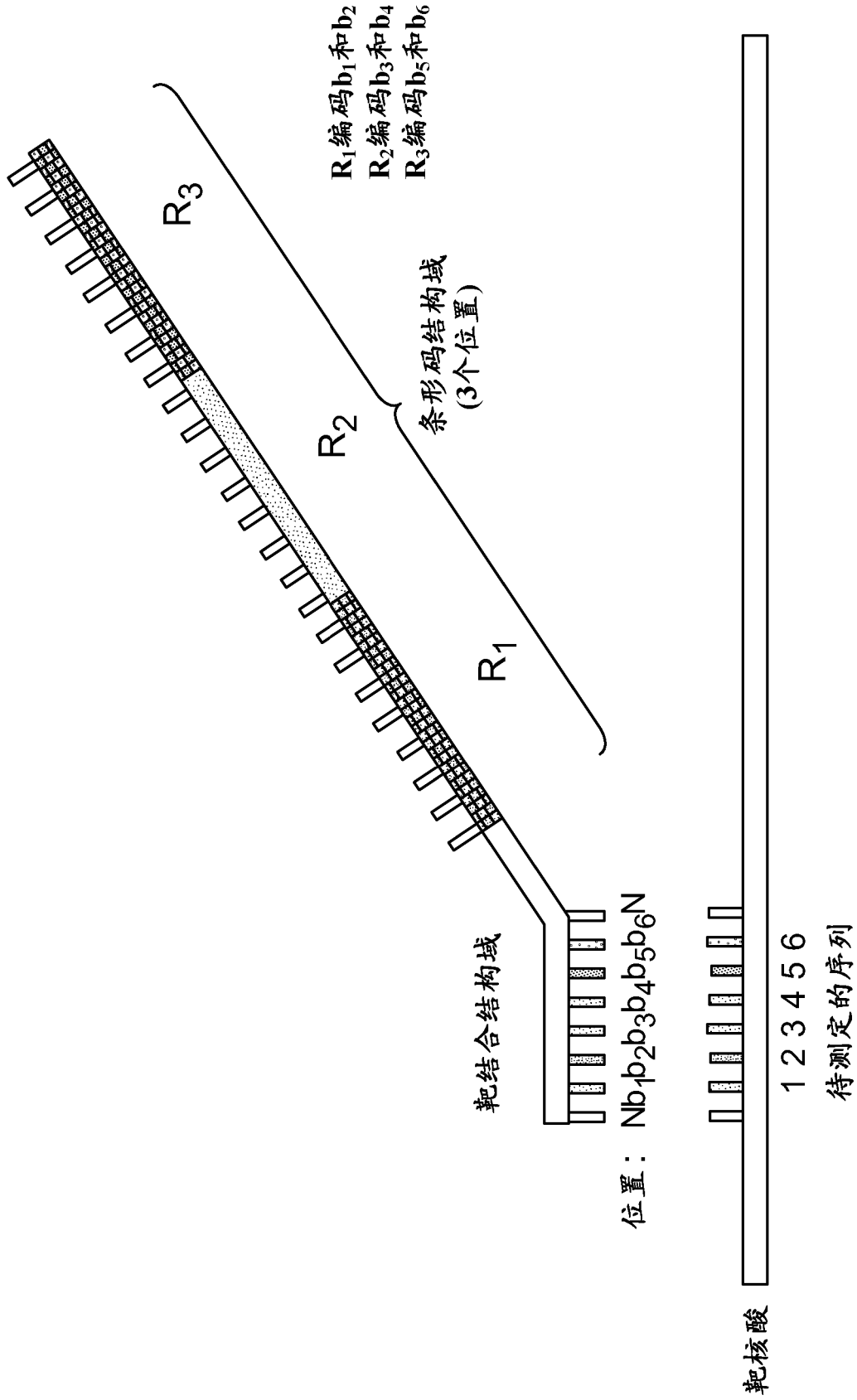


图 1

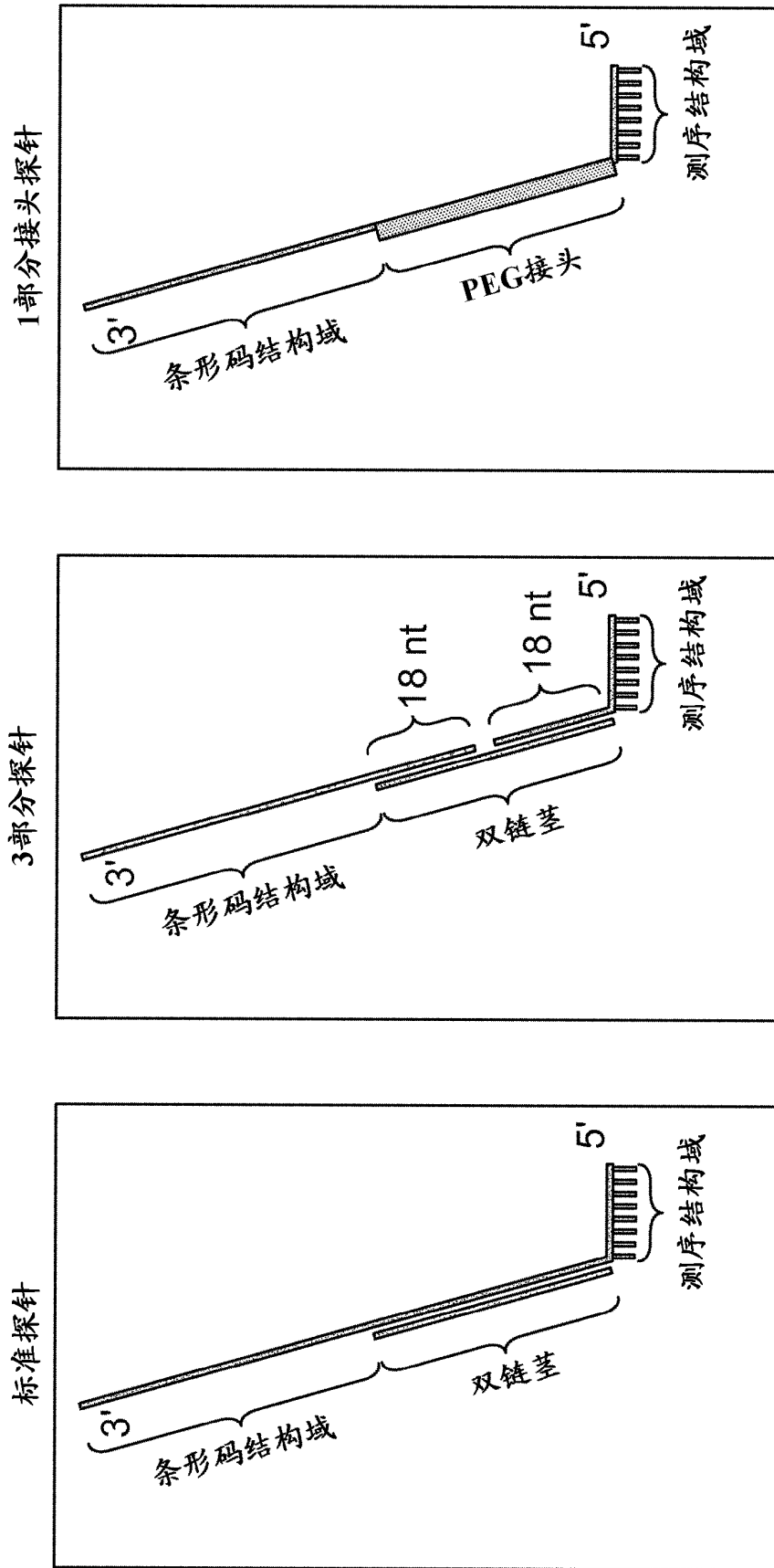


图 2

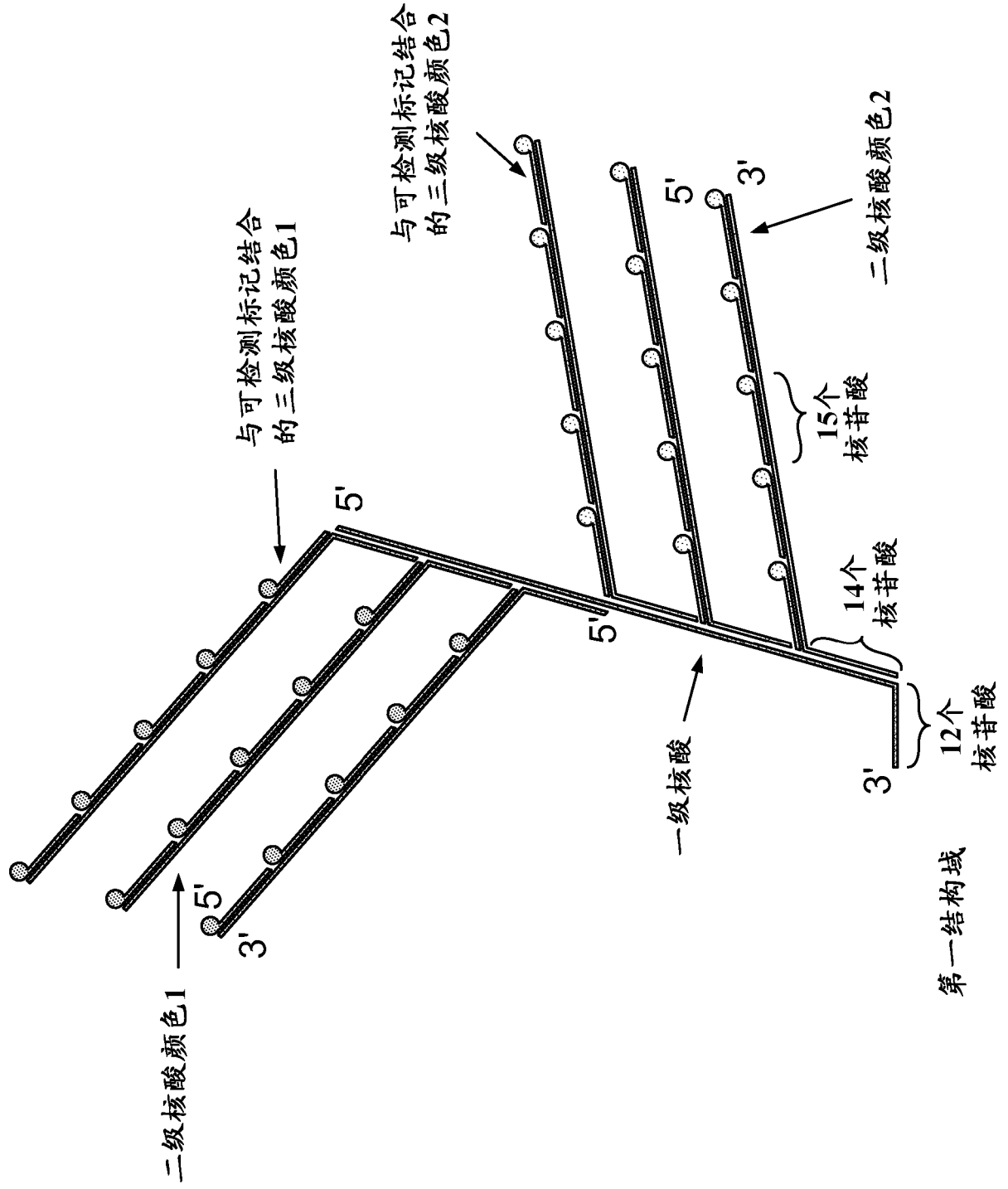


图 4

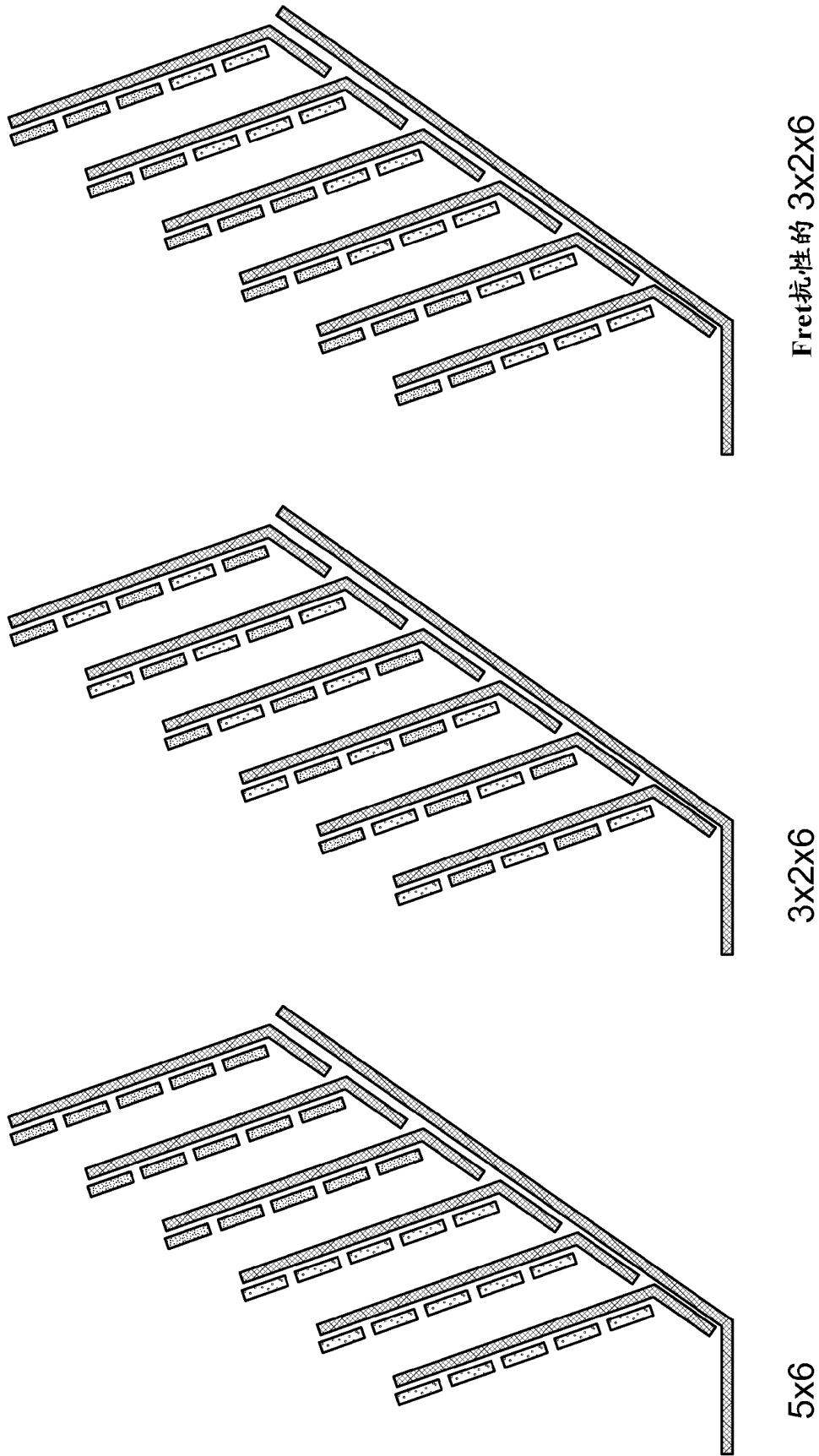


图 5

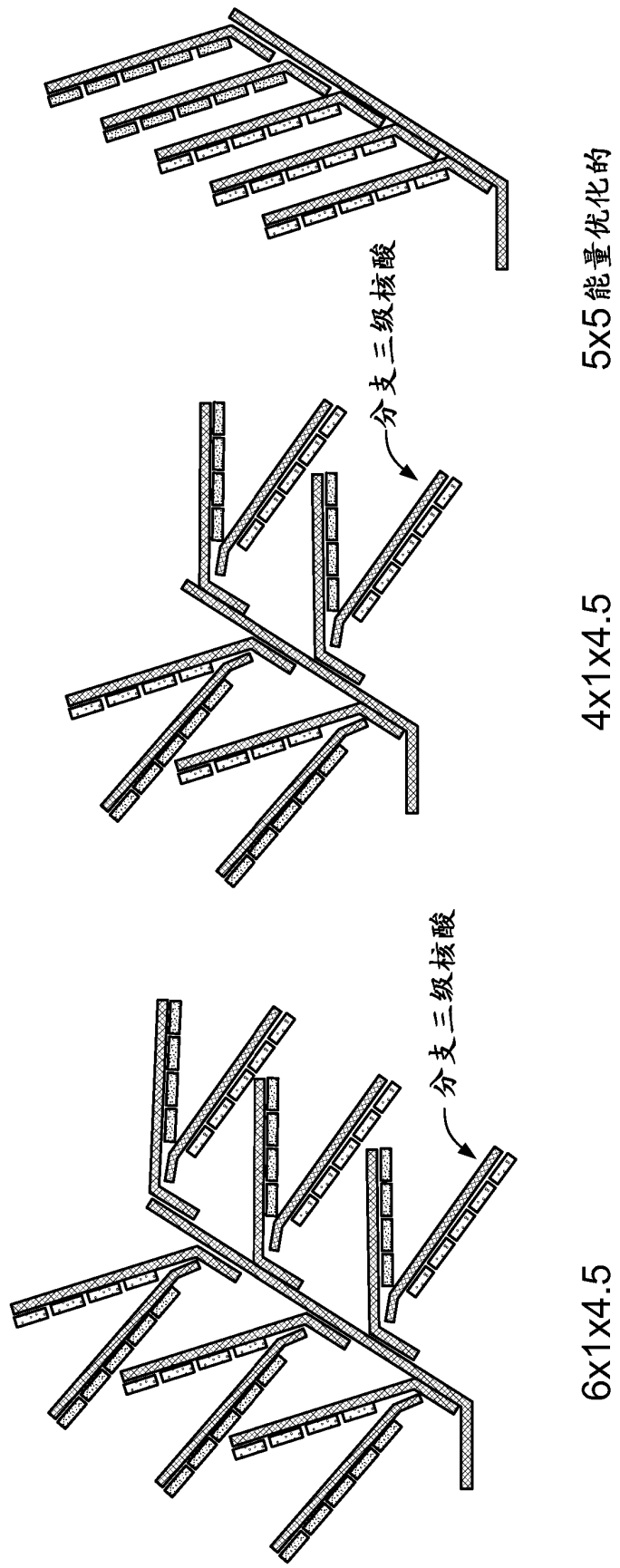
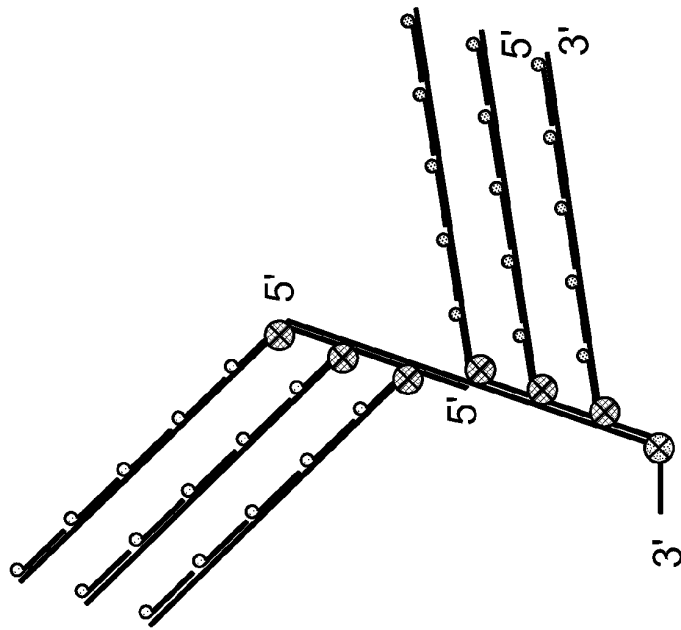


图 6

报告物设计

- ⊗ = 一级核酸切割位点
- ⊙ = 二级核酸切割位点



实例切割修饰

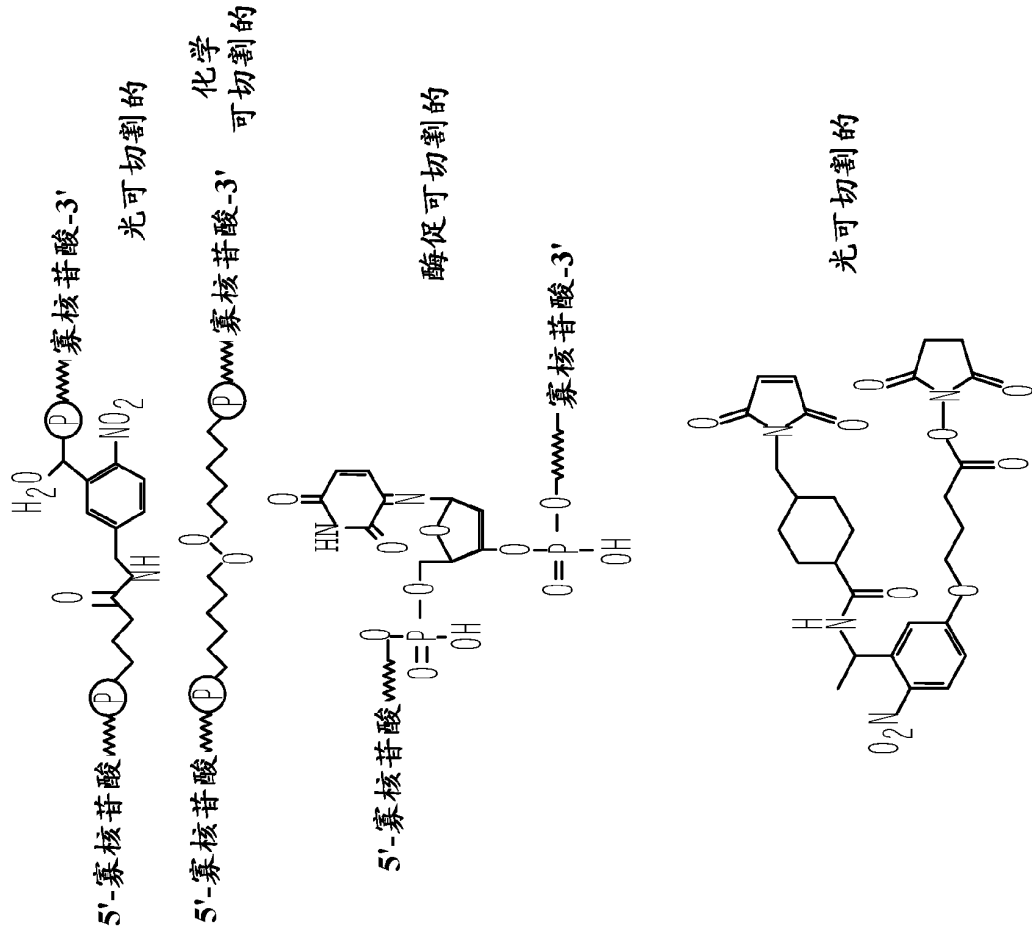


图 7

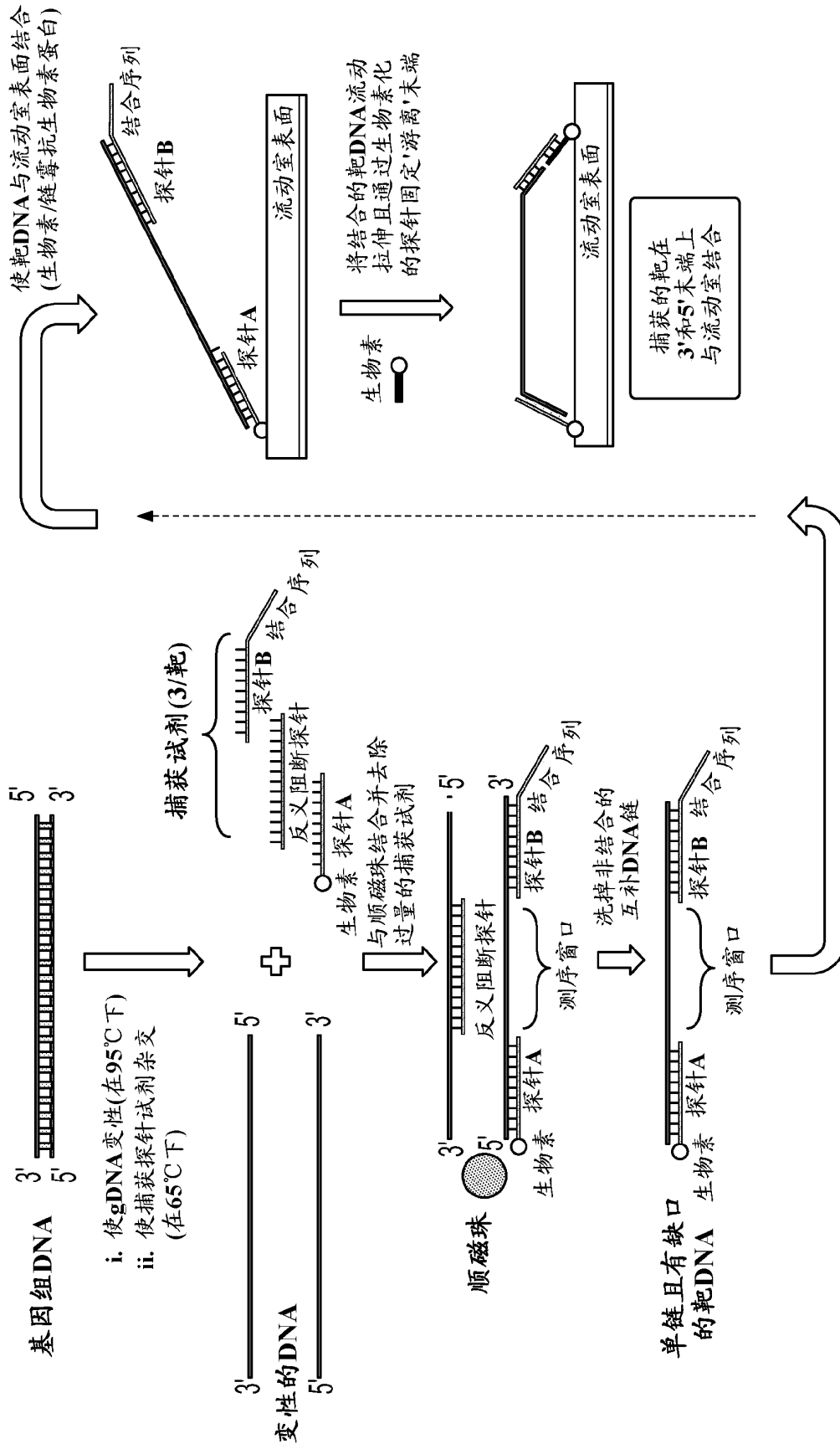


图 8

来自FFPE组织的癌症组(100种靶)的数字定量和鉴定

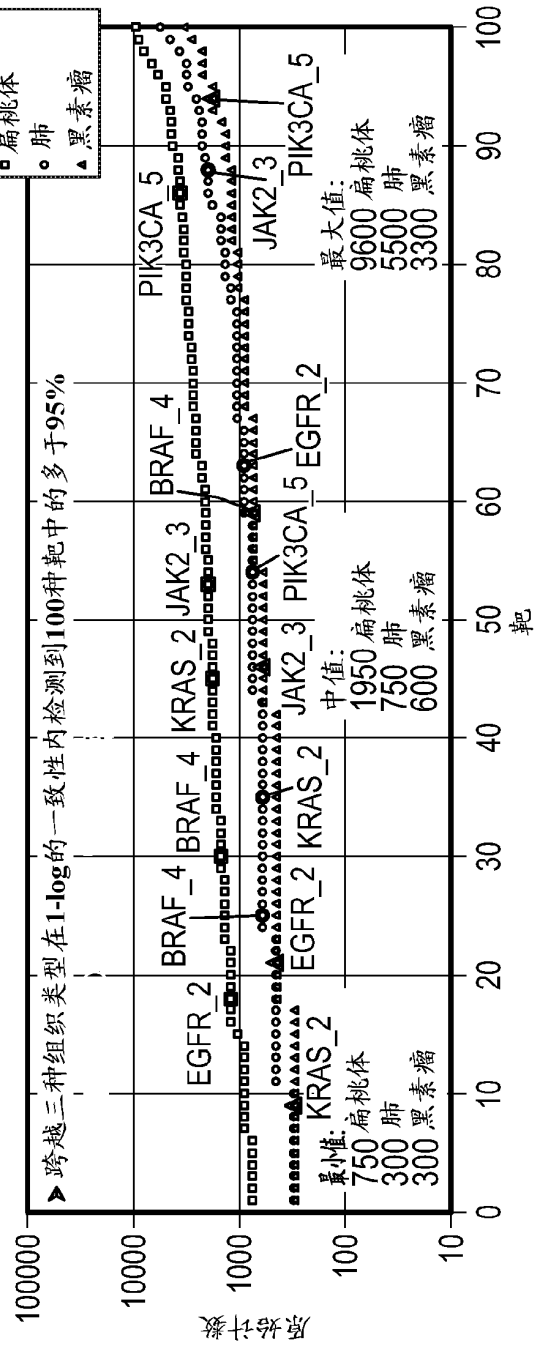
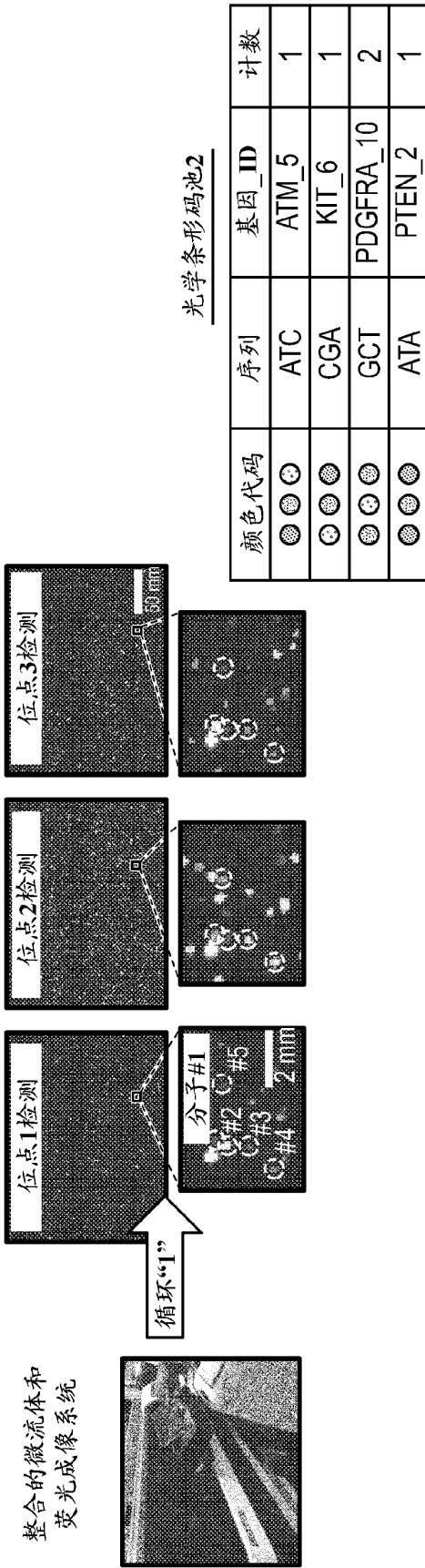


图 9

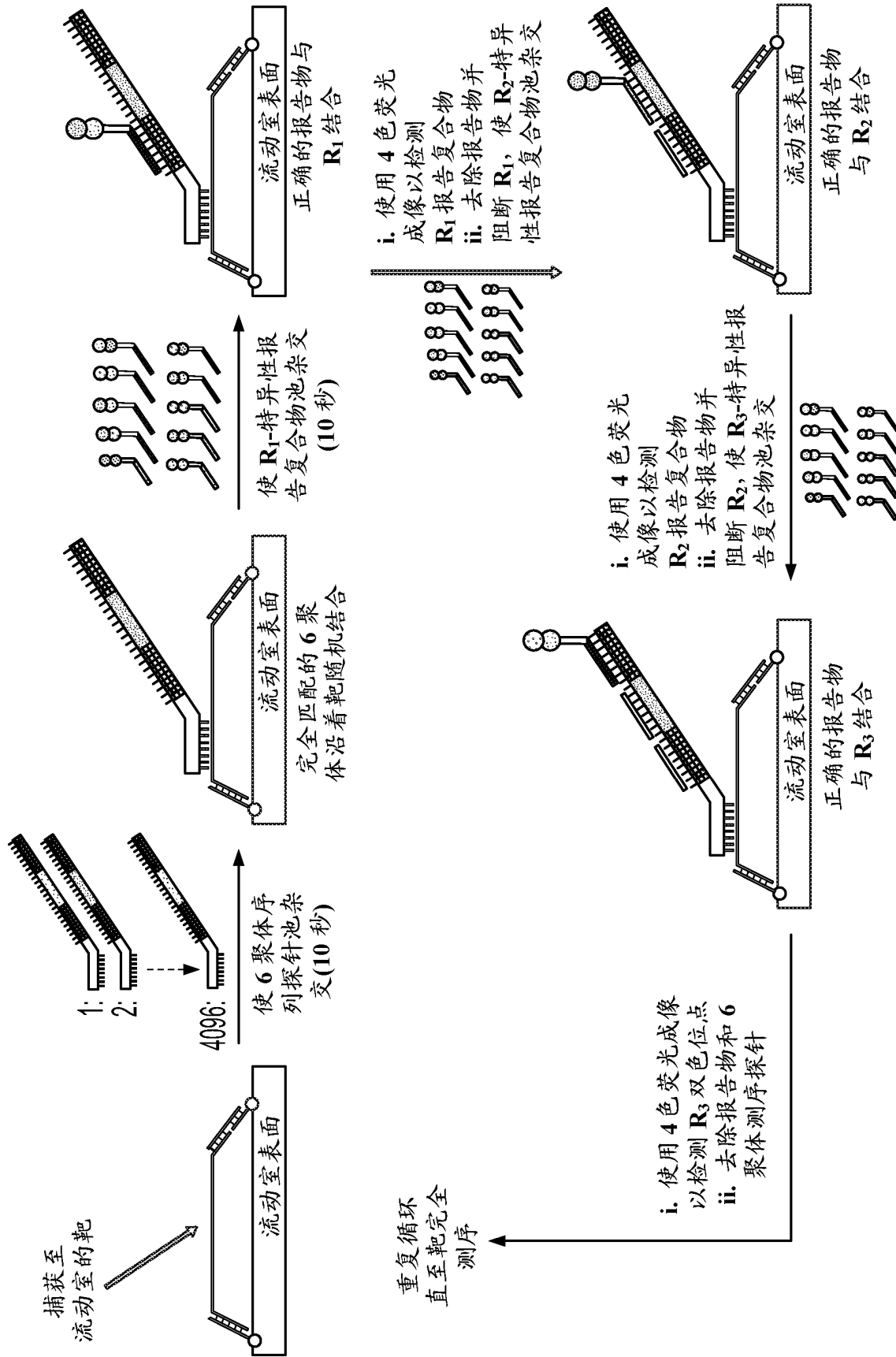


图 10

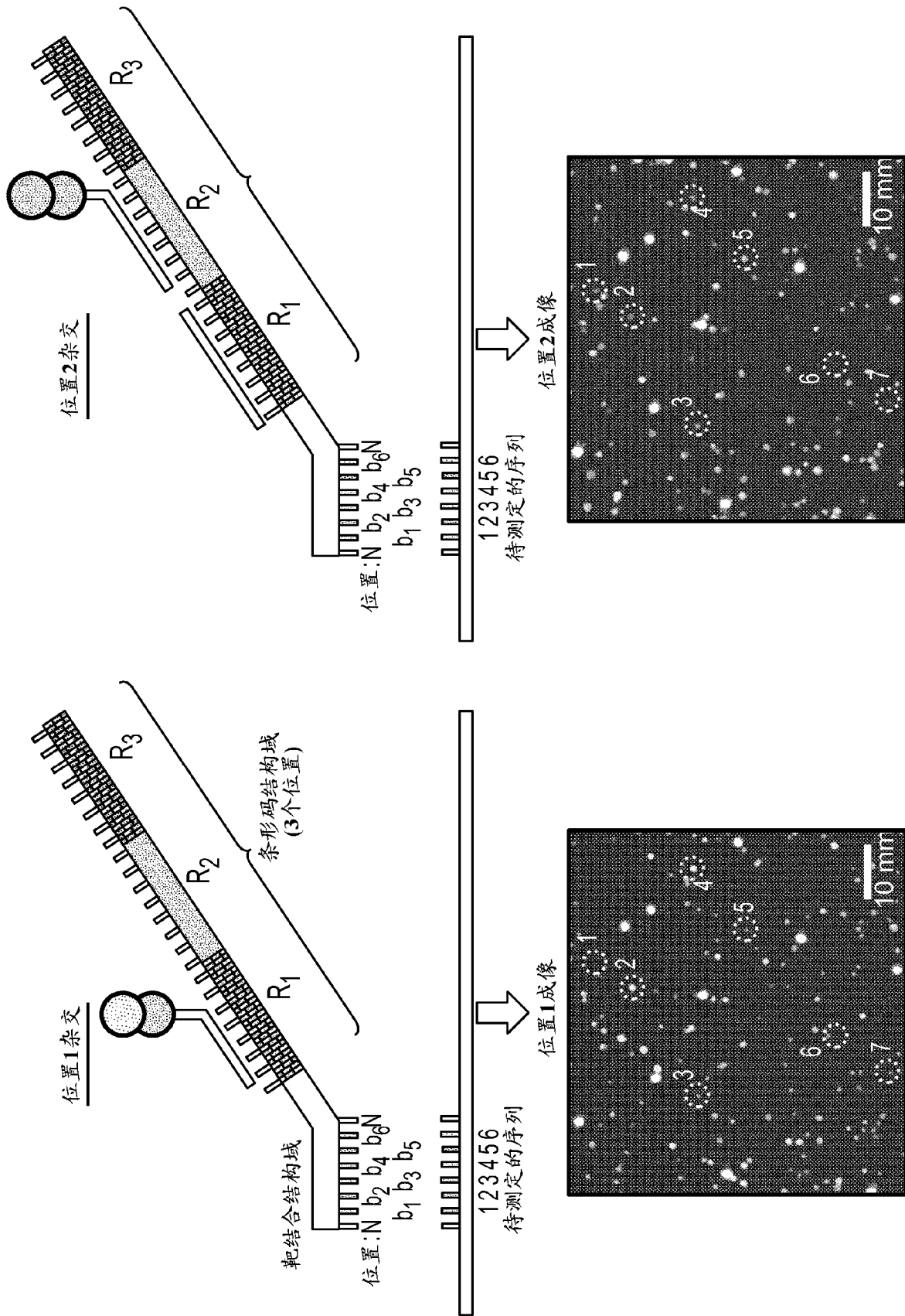


图 11

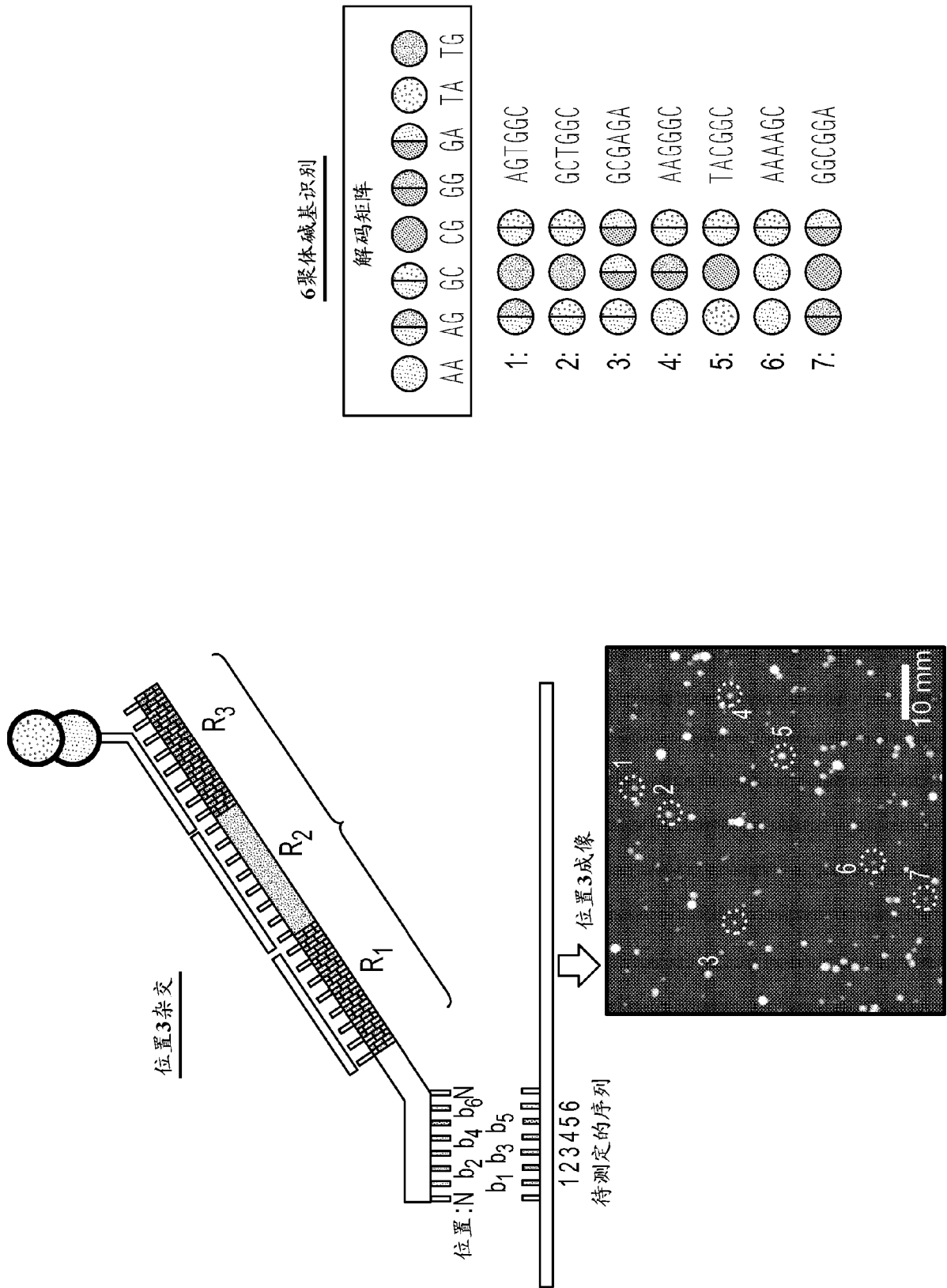


图 11 (续)

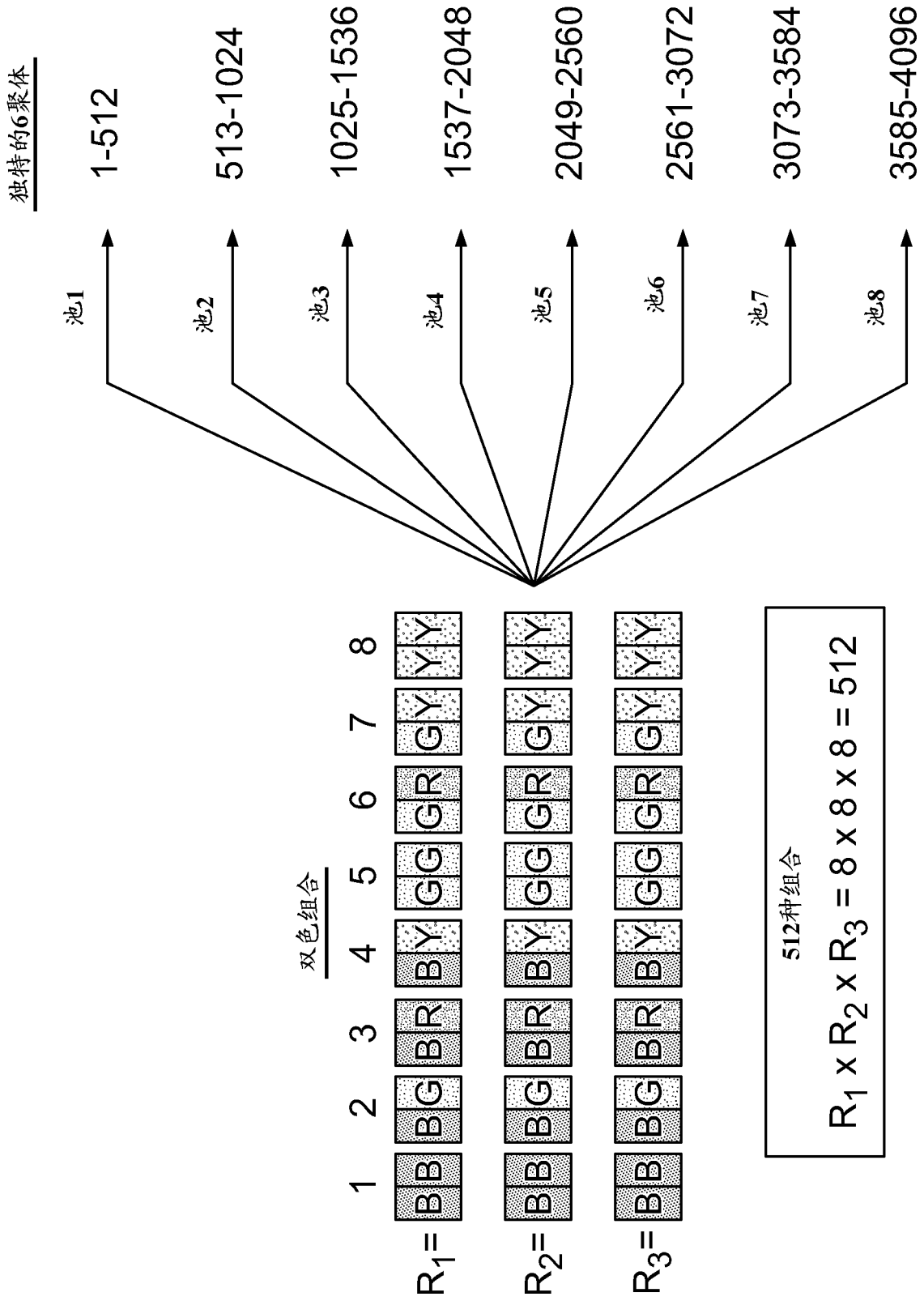
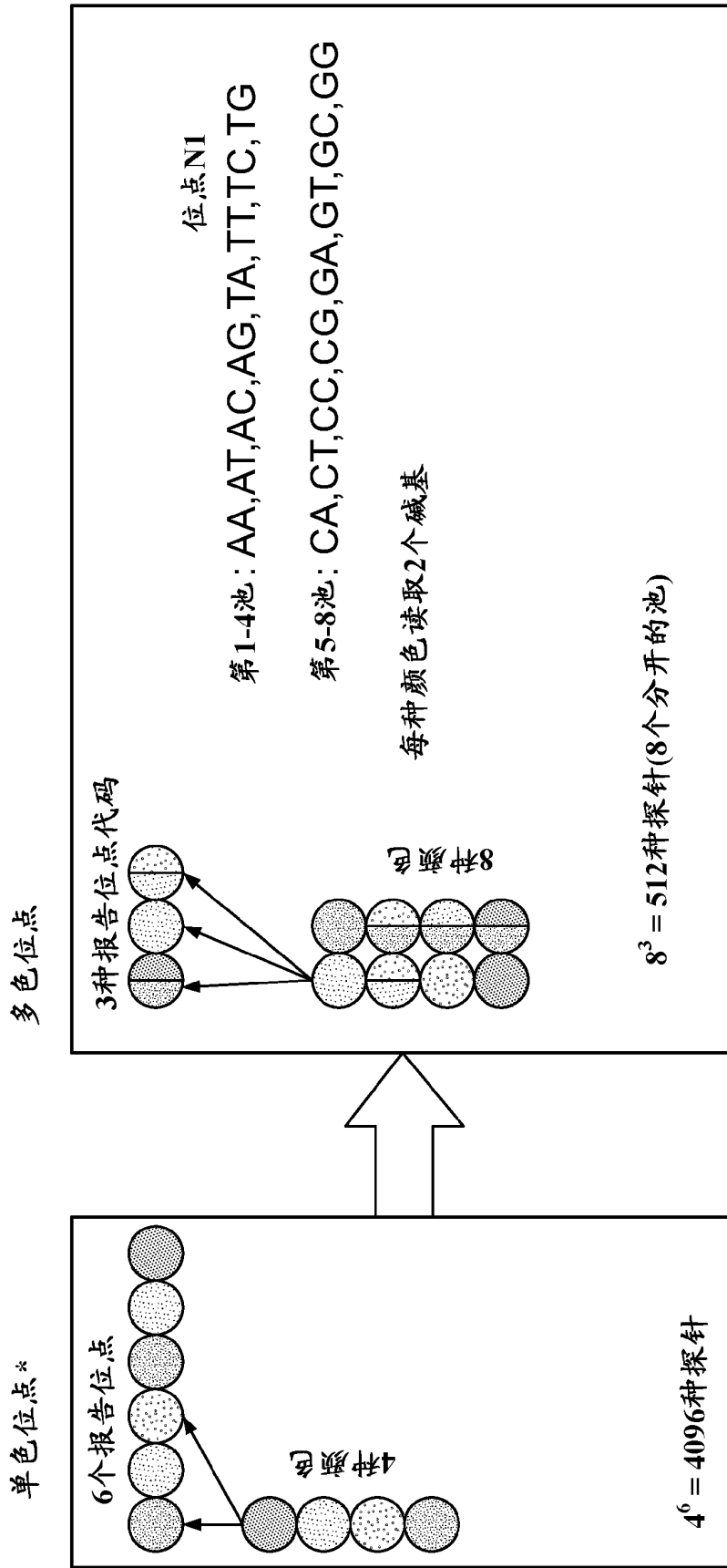


图 12



*如US 2016-0194701中公开的，其整体引入本文作为参考

图 13

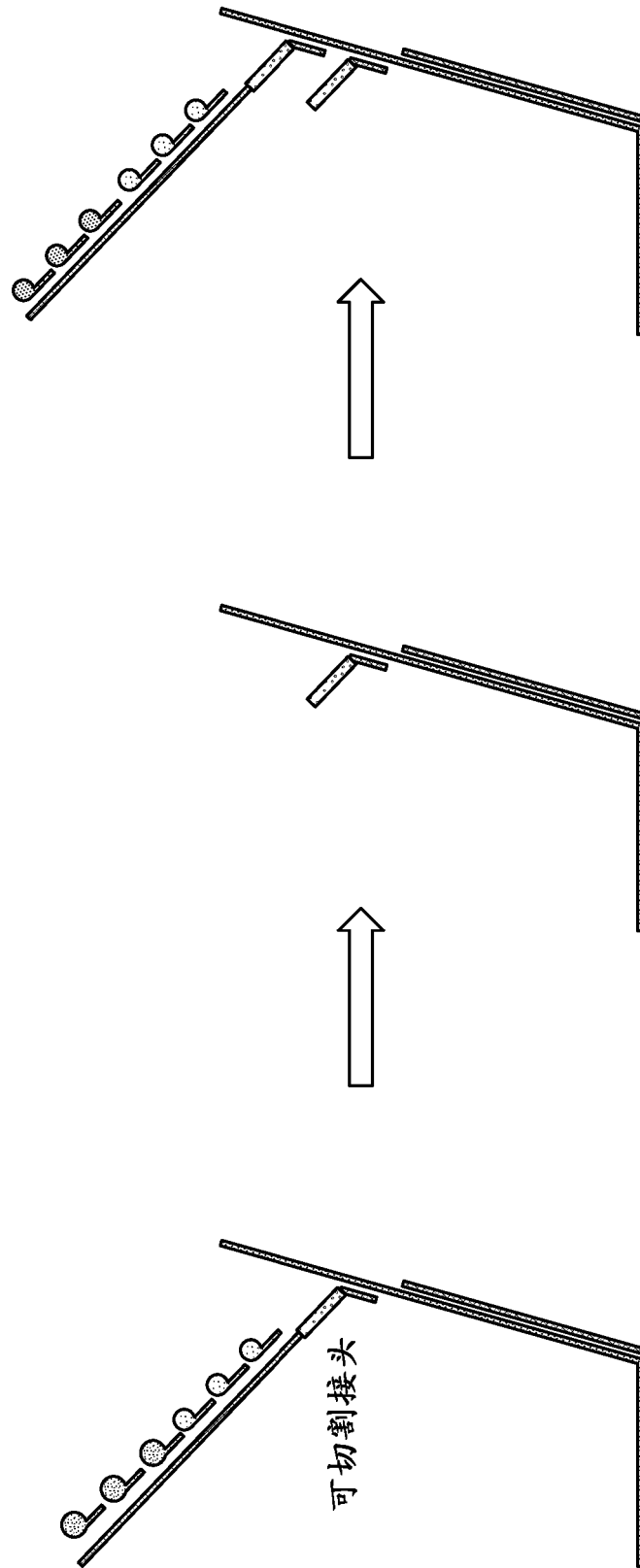


图 14

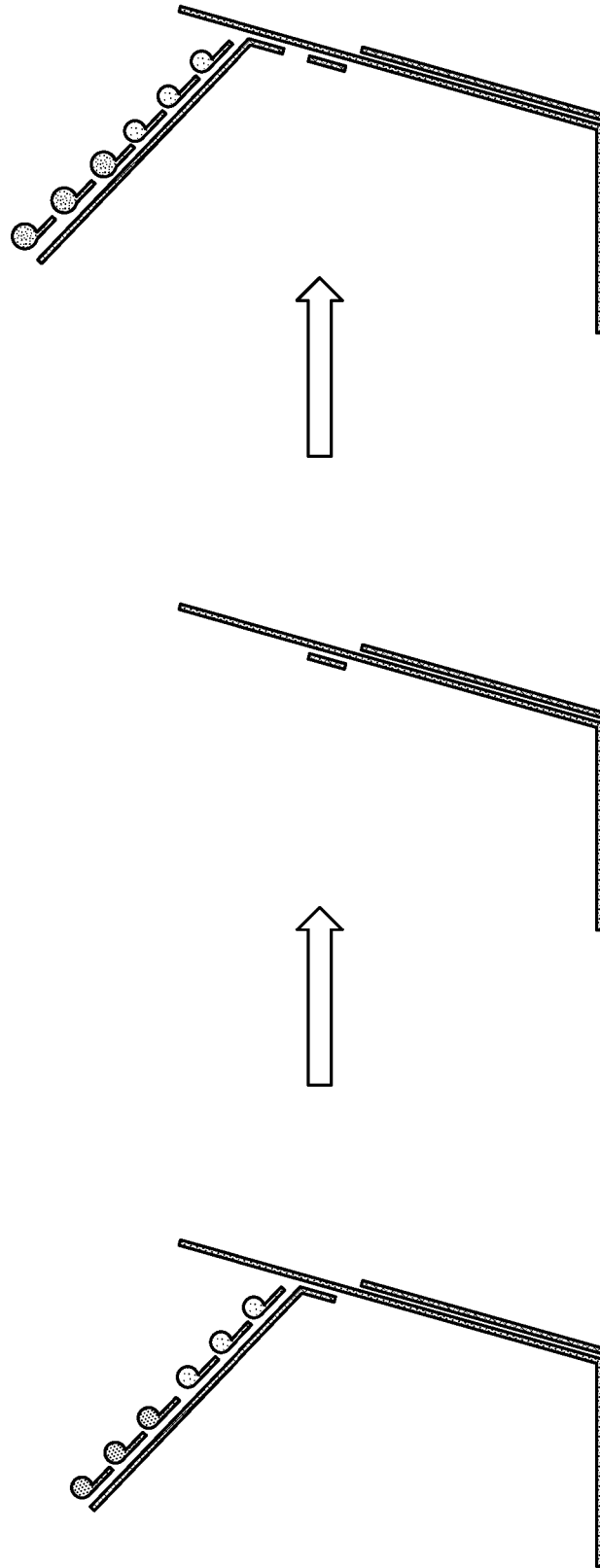
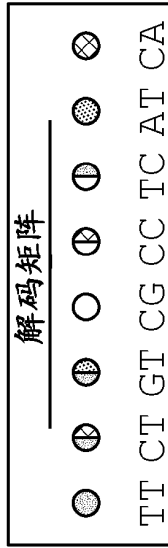


图 15



观察到的(平均%)

1.6%
1.9%
1.7%
94.8%

事实

●●⊖
●⊖⊖
●●⊖
●●⊖

R1	R2	R3
T	A	C
T	C	C
T	T	C
T	G	C

条形码 1:
条形码 2:
条形码 3:
条形码 4:

R1	R2	R3
T	A	C
T	C	C
T	T	C
T	G	C

条形码 5:
条形码 6:
条形码 7:
条形码 8:

NRAS外显子2

3'-ACCGTGACATGAGAGACAAGGTCGACATAGGTCATAC-5'

(SEQ ID NO: 1)

图 16

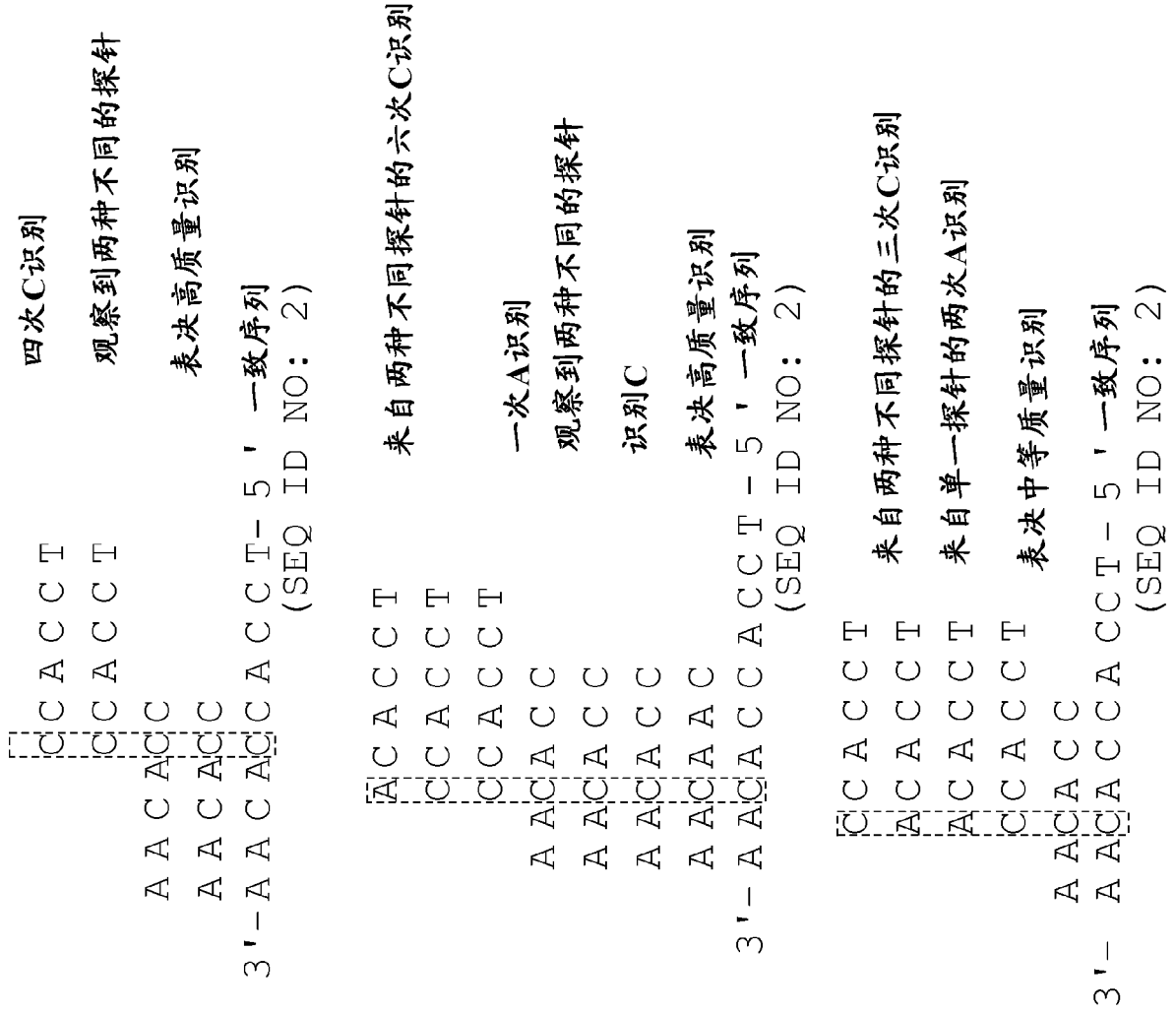


图 17

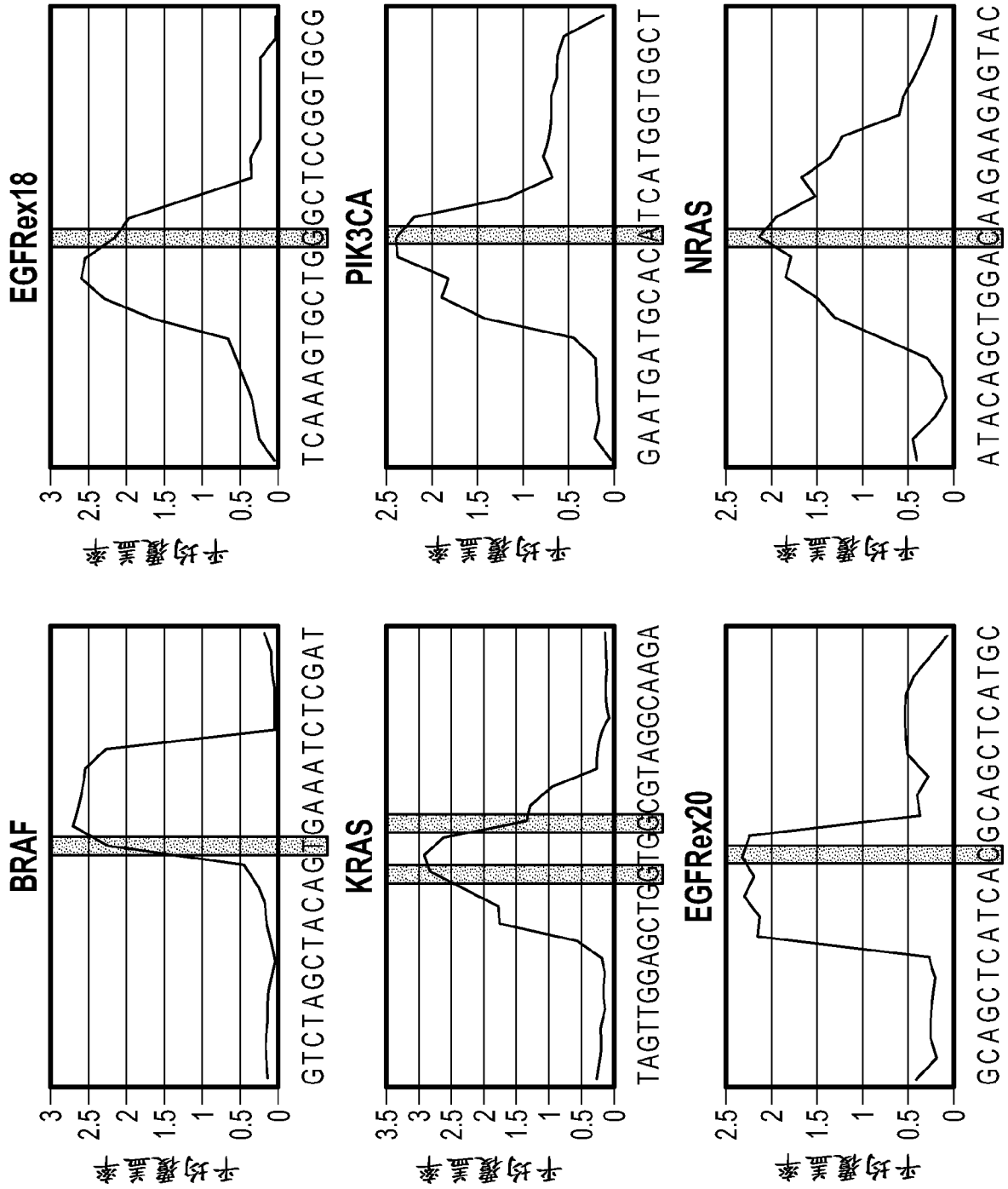


图 18

基因	变体	观察到的频率(%)	预计的频率(%)
BRAF	ref V600E	91.0% 9.0%	89.5% 10.5%
外显子18	ref G719S	88.1% 11.9%	75.5% 24.5%
外显子20	ref T790M	98.9% 1.1%	99.0% 1.0%
NRAS	ref Q61K	63.0% 37.0%	87.5% 12.5%
PICK3CA	ref H1047R	92.1% 7.9%	82.5% 17.5%
KRAS	ref G12D G13D	99.5% 0.46% 0.005%	79.0% 6.0% 15.0%

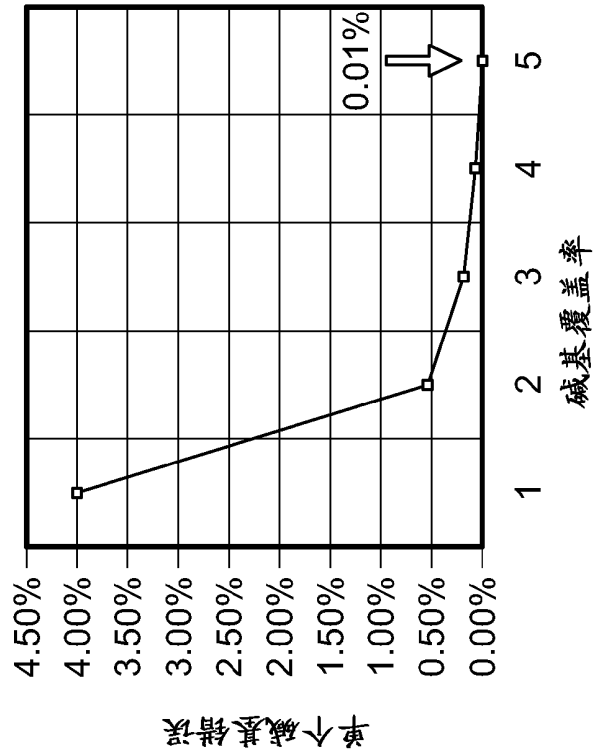
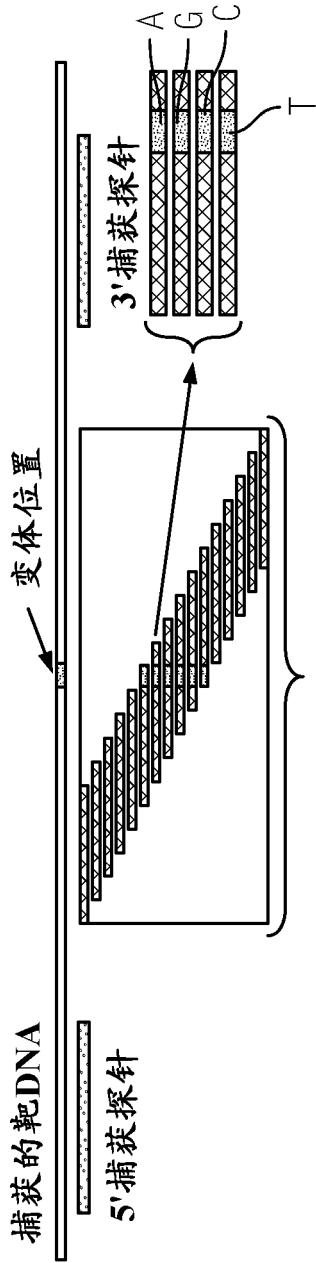


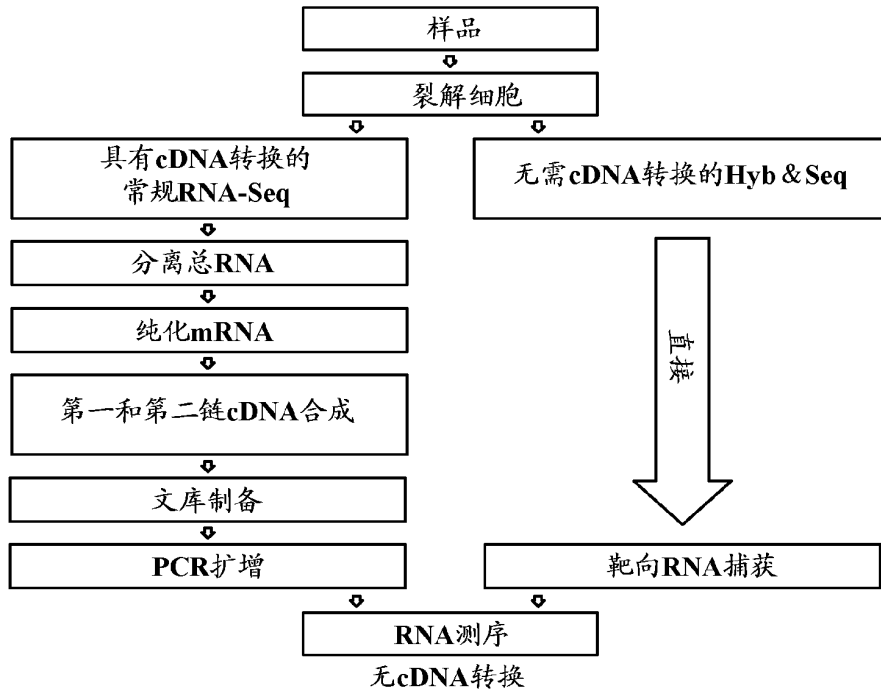
图 18(续)



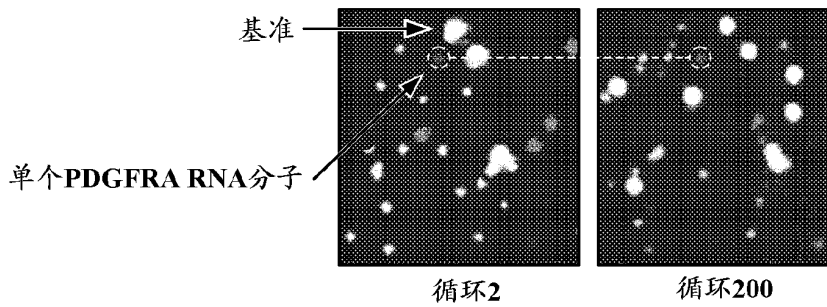
测序区域
捕获11种DNA靶的多路复用
列出目的变体

基因	变体	测序区域 5' → 3'	SEQ ID NO
BRAF	V600E V600K	GTCTAGCTACA[G]GAAATCTCGAT	3
EGFR	G719S	TCAAAGTGT[G]GGTCCGGTGCG	4
KRAS	G12A G12R G13D	TAGTTGGAGCT[G]TGGGTTAGGCAAGA	5
PIK3CA	H1047R	GAATGATGCACATCATGGTGGCT	6
EGFR	T790M	GCAGCTCATCAGGCAGCTCATGC	7
NRAS	Q61K	ATACAGCTGGAAAGAAGAGTAC	8
FLT3	ΔI836	TGGCTCGAGATATGATGAGTGATTC	9
IDH1	R132C	TCATCATAGGTGGTTCATGCTTAT	10
JAK2	V617F	ATGGAGTATGTCTCTGTGGAGAC	11
MEK1	P124L	GTGCAACTCTCGTACATCGTGG	12
NOTCH1	L1601P	CAGCCGGTGGGCACACCAACG	13

图 19



通过流动室上的循环跟踪单个RNA分子(PDGFRA)



mRNA转录物	测试的RNA序列 (检测序列是 加下划线的, 5'→3')	报告探针颜色	SEQ ID NO
PDGFRA	GCUUCAG <u>C</u> ACCACUCCACUCCACAUUCUCA	蓝	14
EZH2	UGUCCUCAUGUAU <u>L</u> GGUCUCUCAUGGCACU	绿	15
NRAS	CUUUCGCCAAAGUGGAGGAGACCAUCGCCG	黄	16
SMAD4	GAAGGUGAUGUUUGGGUCAGGUGCCUUAGU	红	17

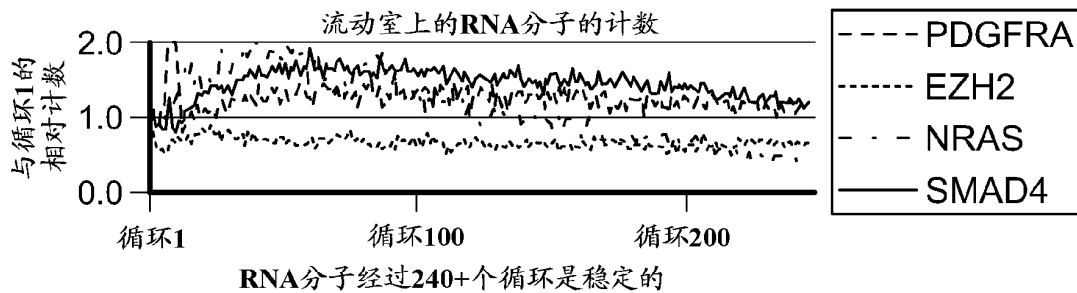


图 20

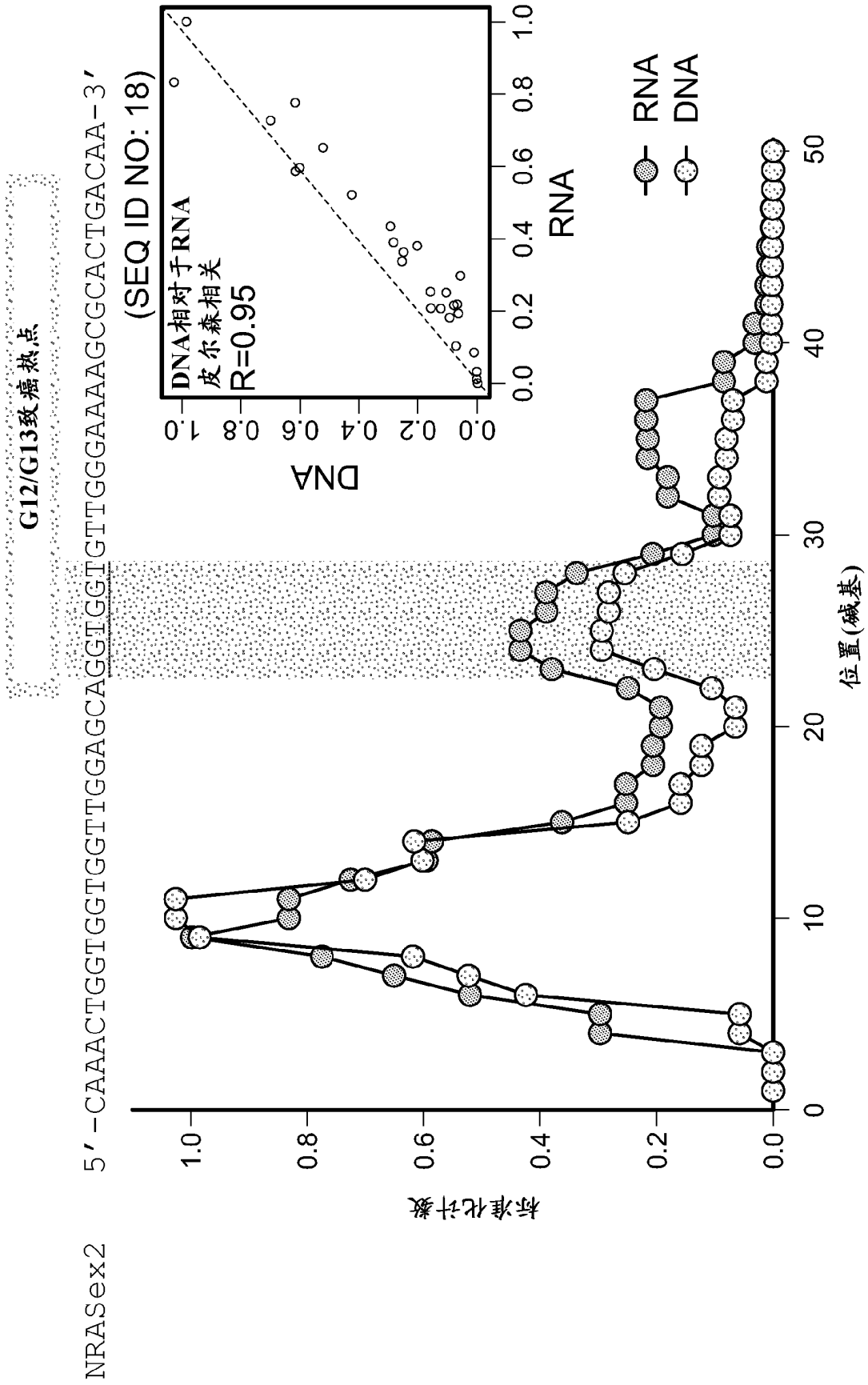


图 21

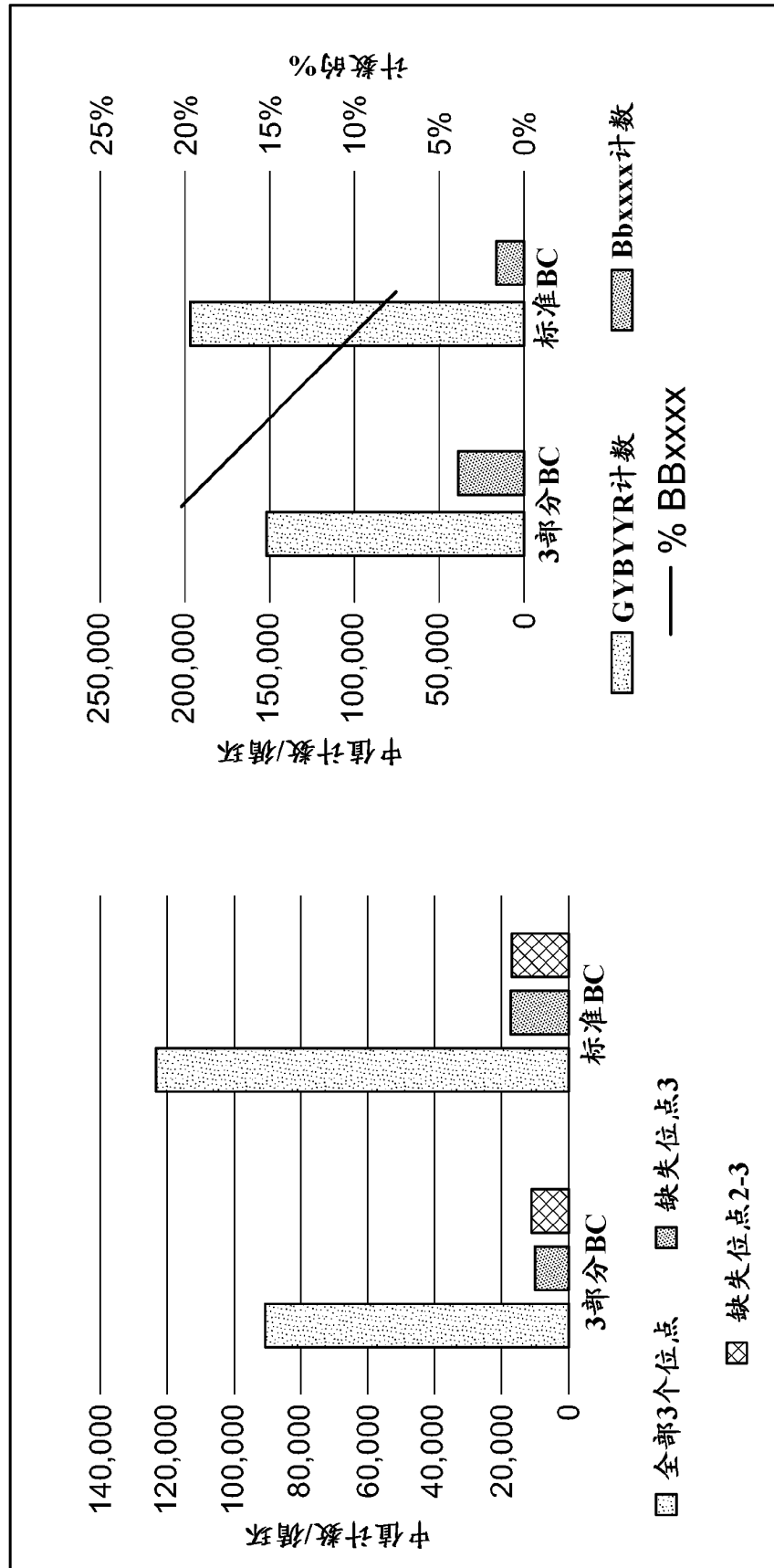


图 22

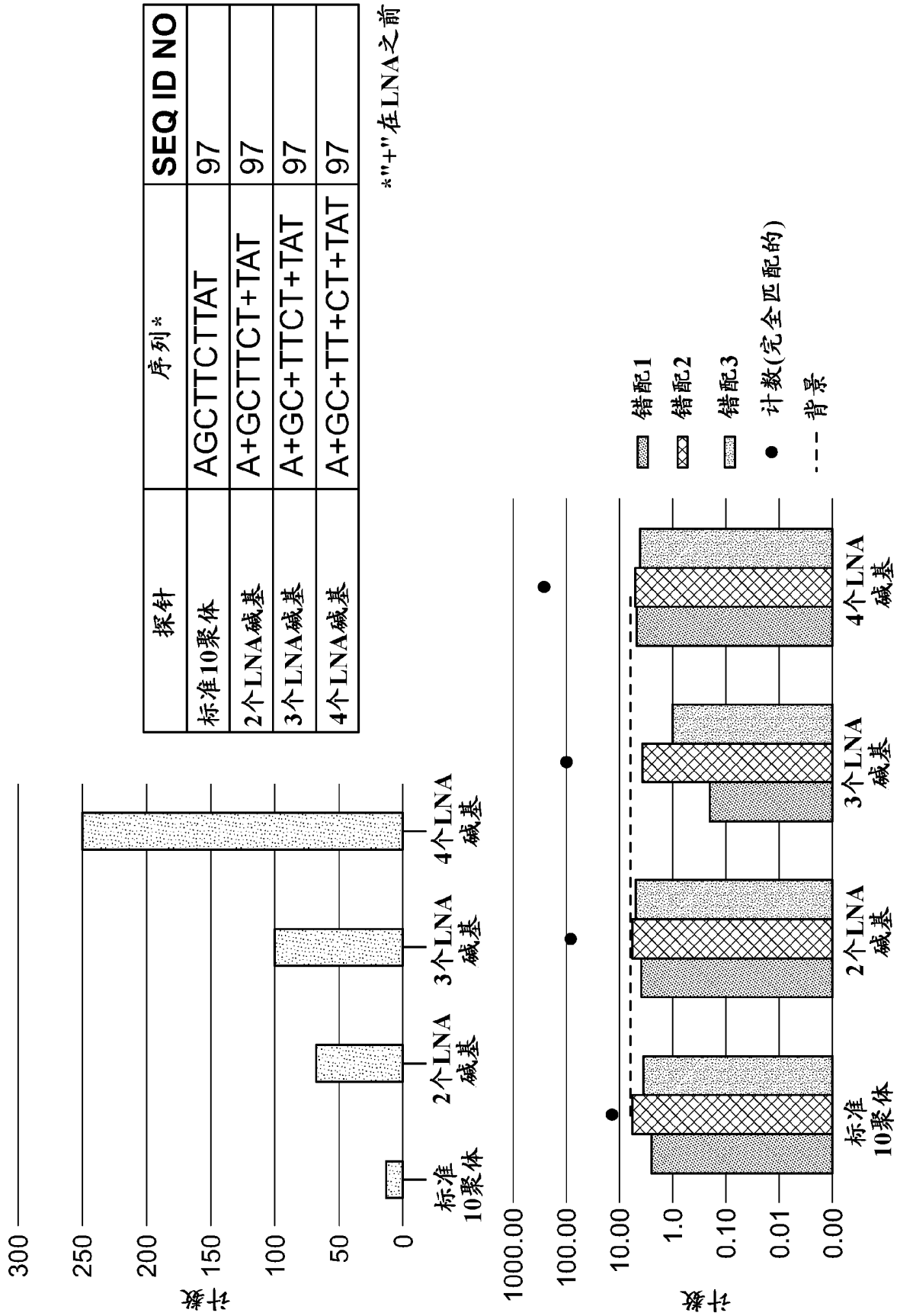


图 23

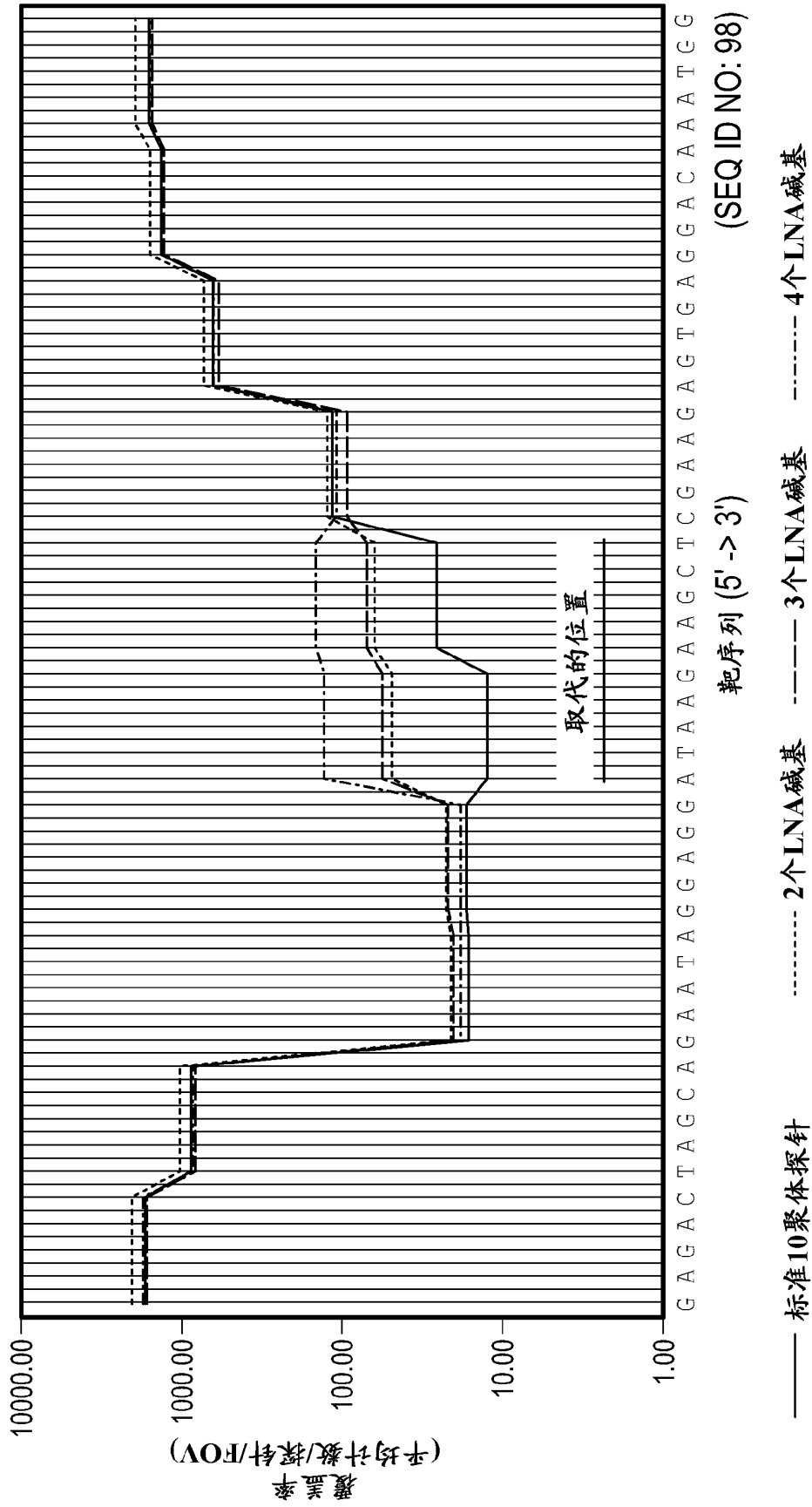


图 24

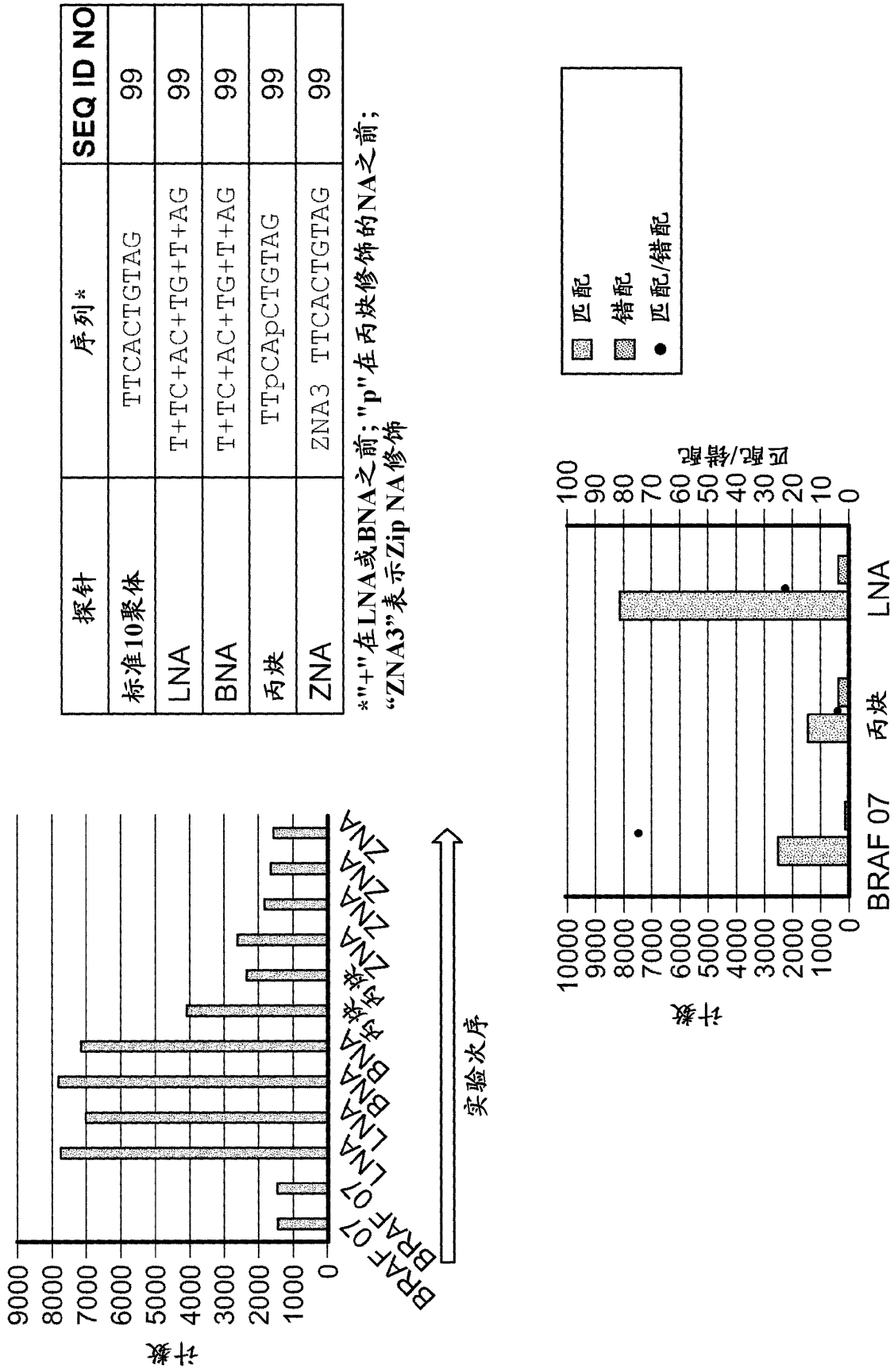
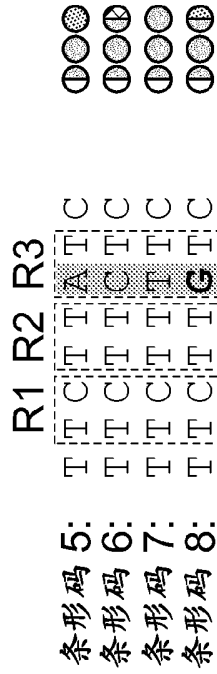
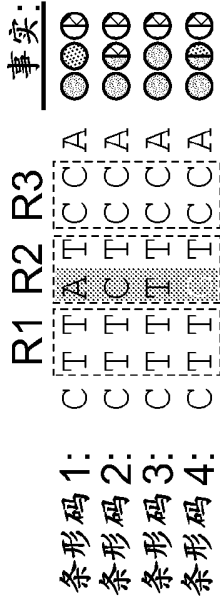
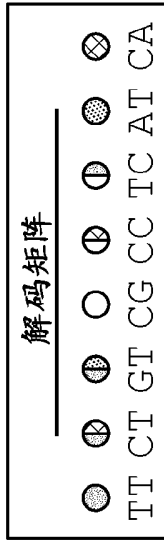


图 25

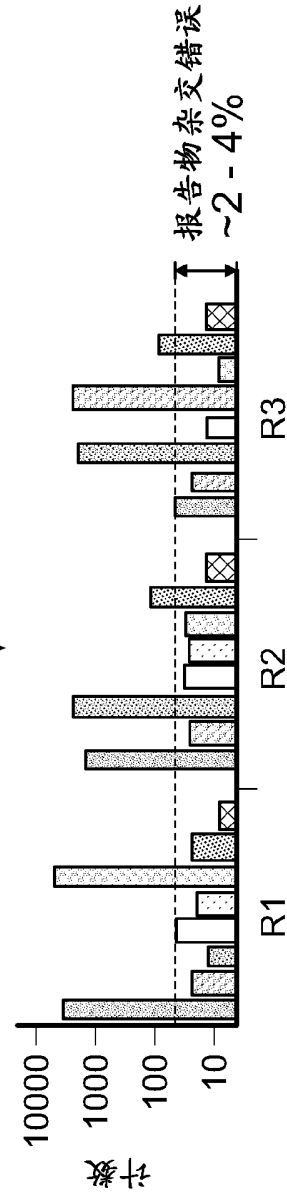
使用双色报告物的Hyb & Seq化学具有低原始错误率，
并且错误来源可以分类成两个区域



NRAS外显子2

3'-ACCGTGACATGAGAGAAGACAGGTCGACATAGGTCAATAC-5'

(SEQ ID NO: 1)



平均约99.98%准确率(具有3次通过读数)

图 27

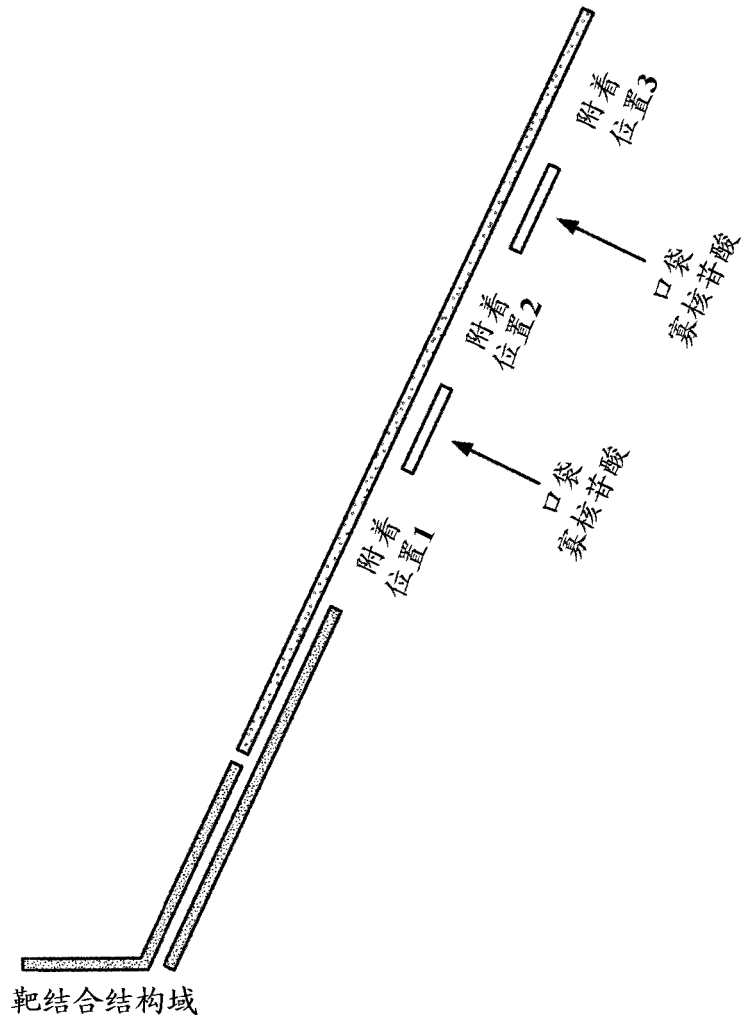


图 28

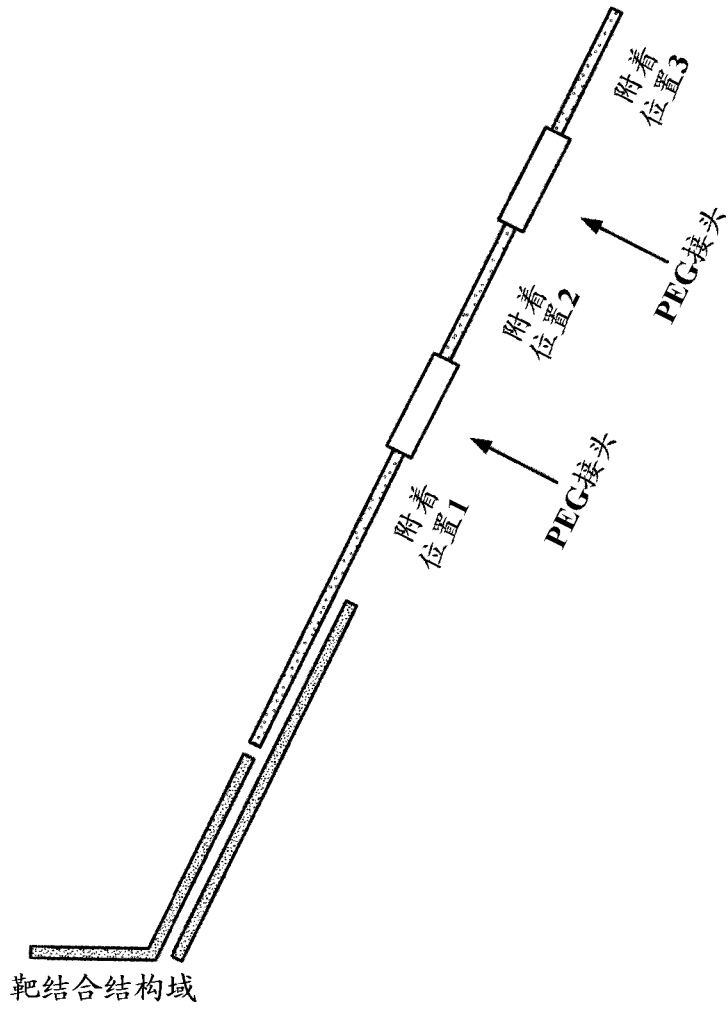


图 29

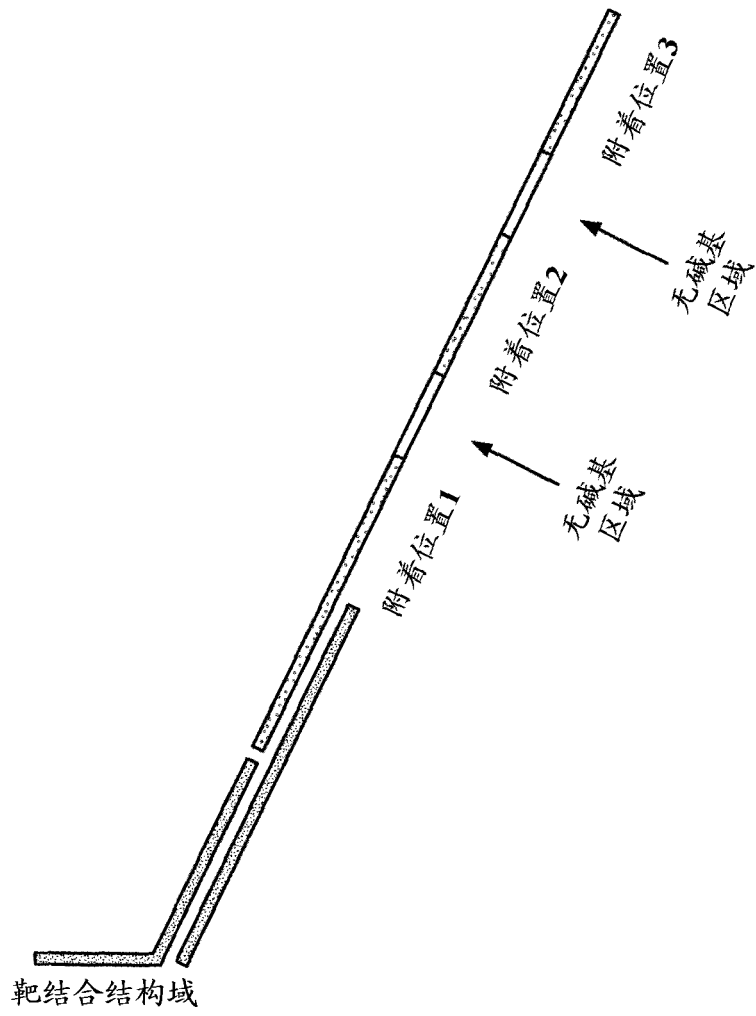


图 30

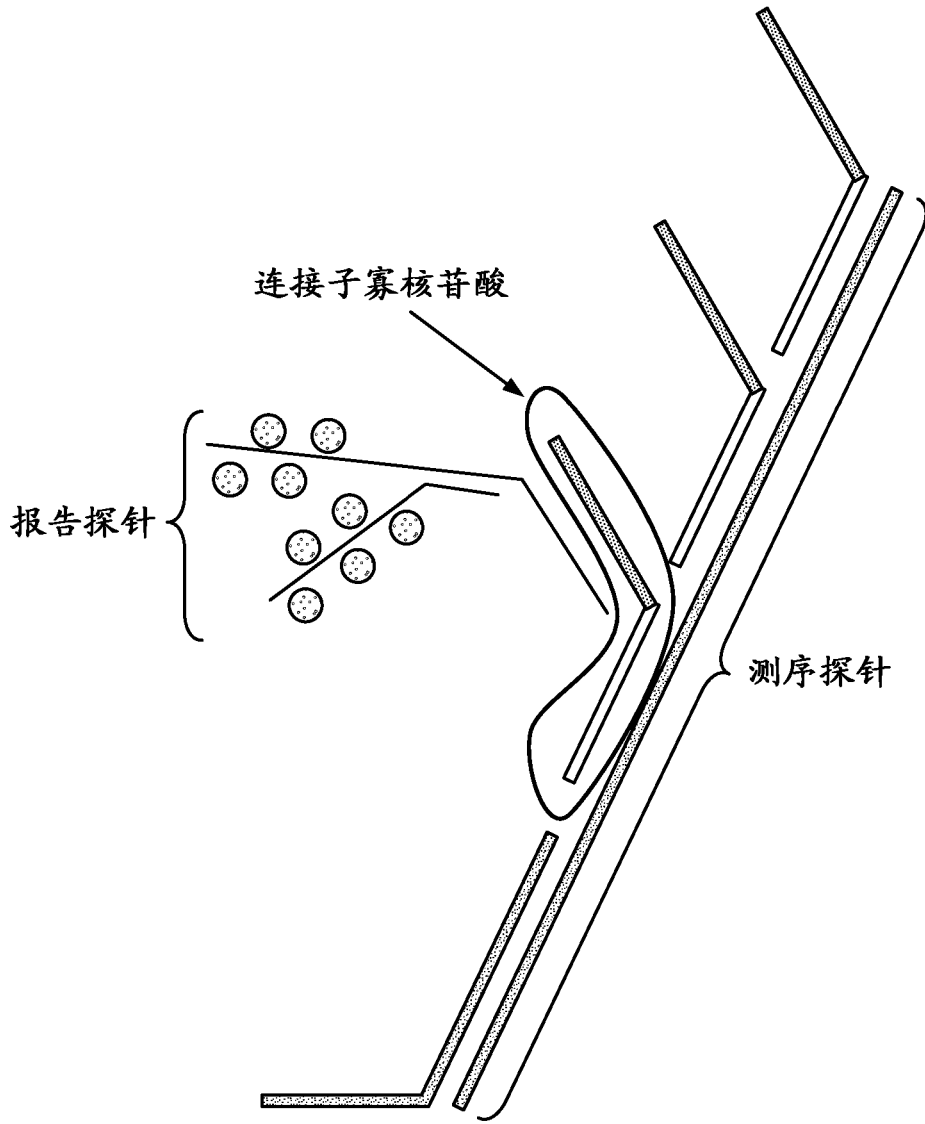


图 31

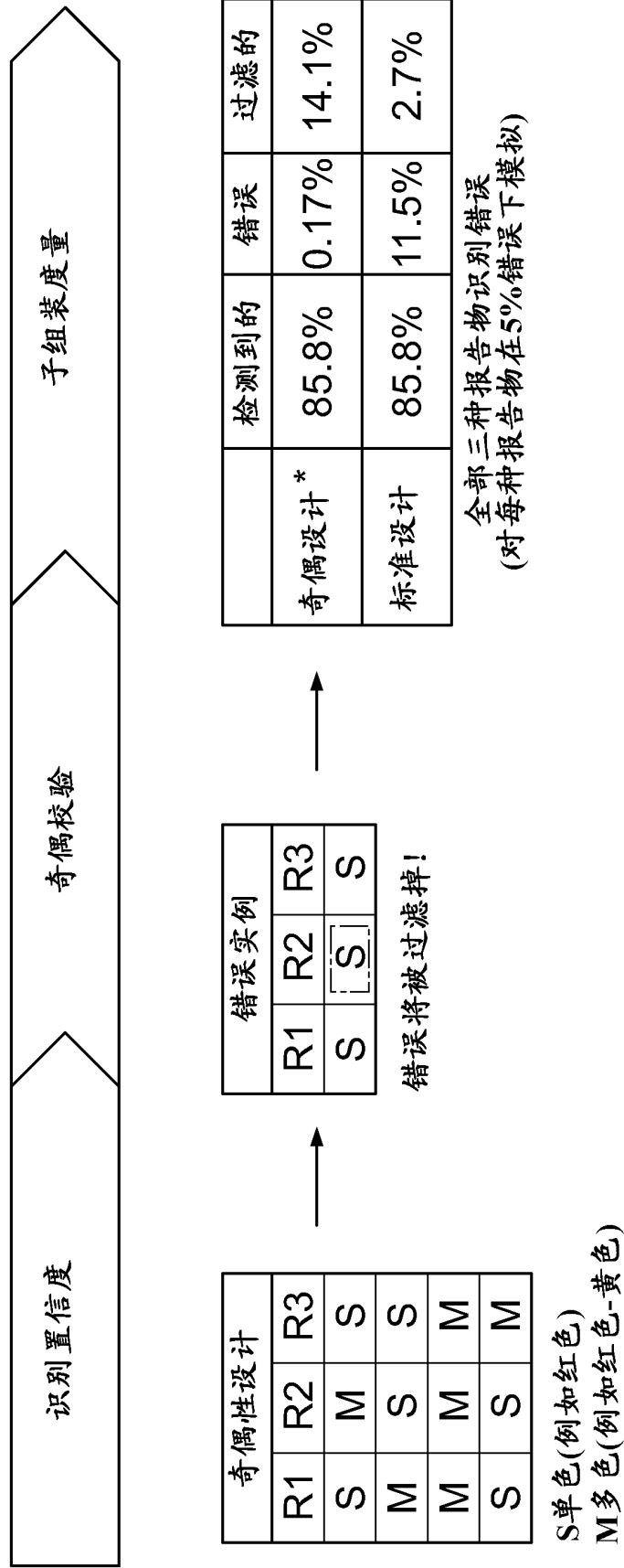


图 32

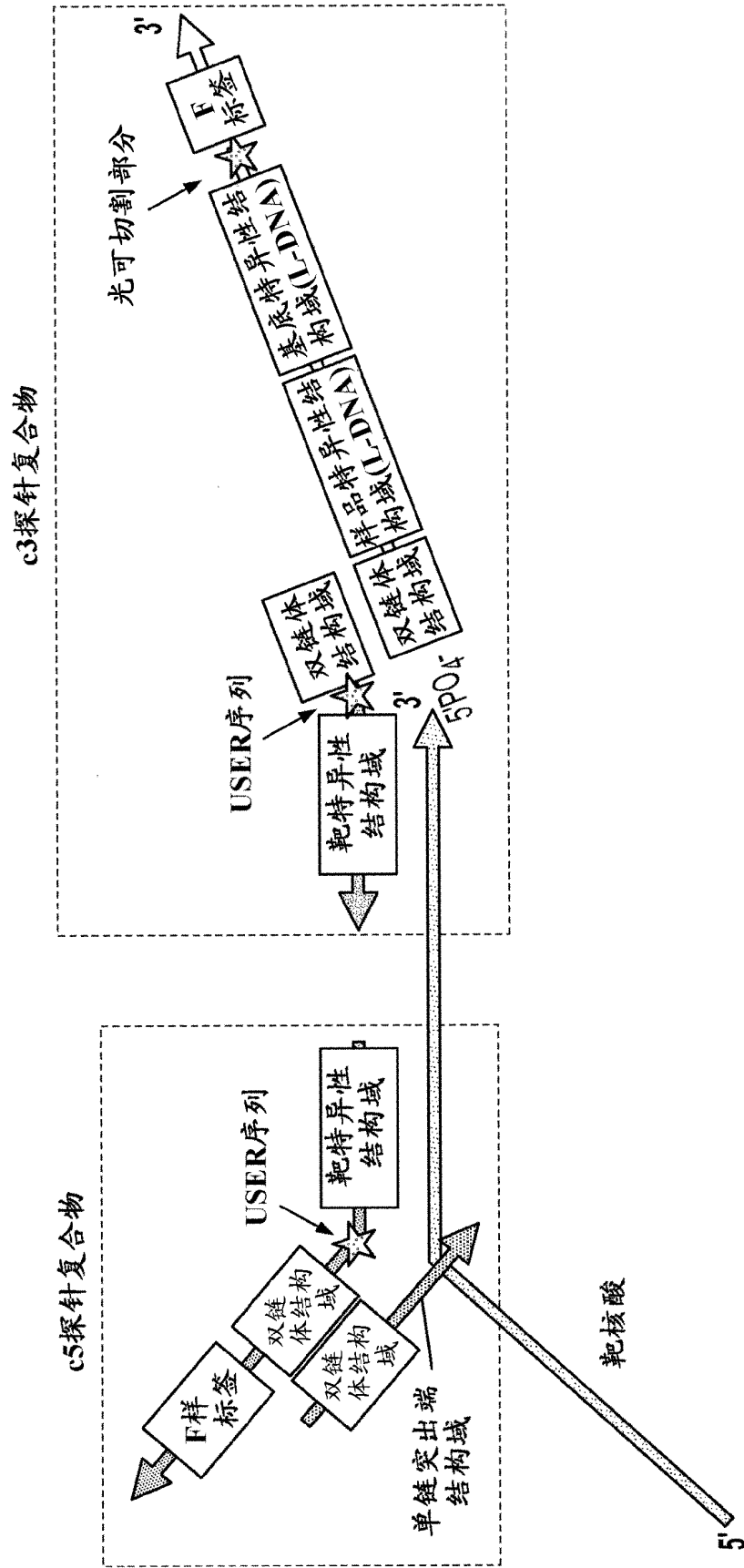


图 34

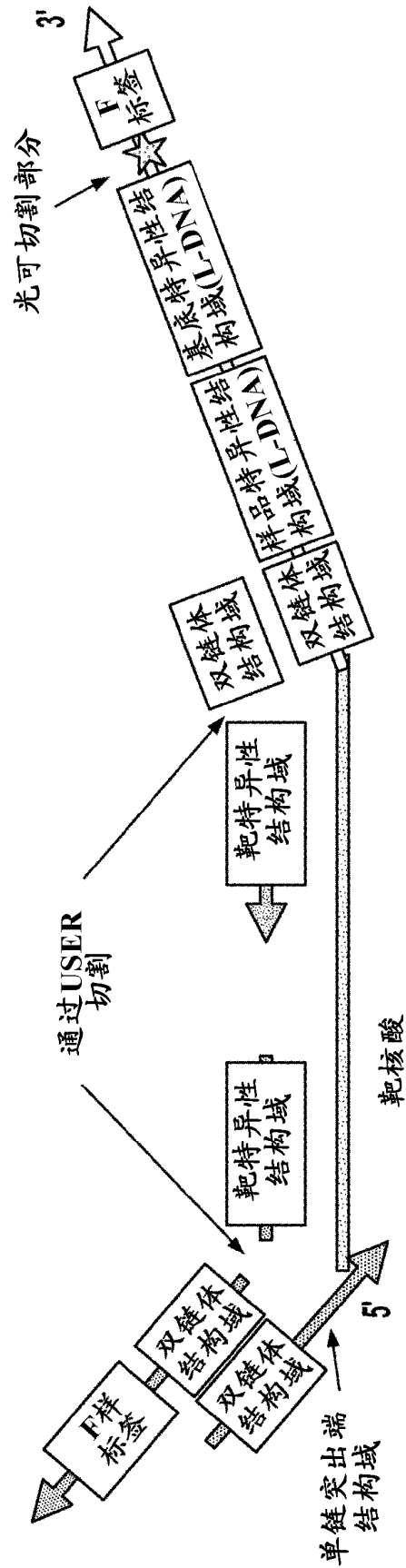


图 37

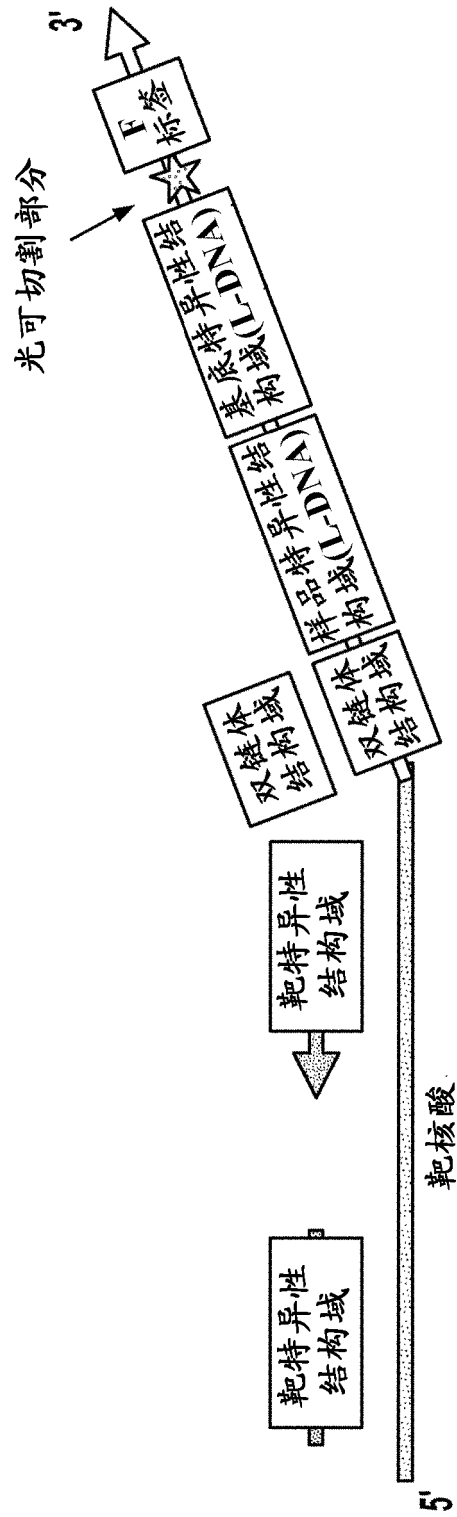


图 38

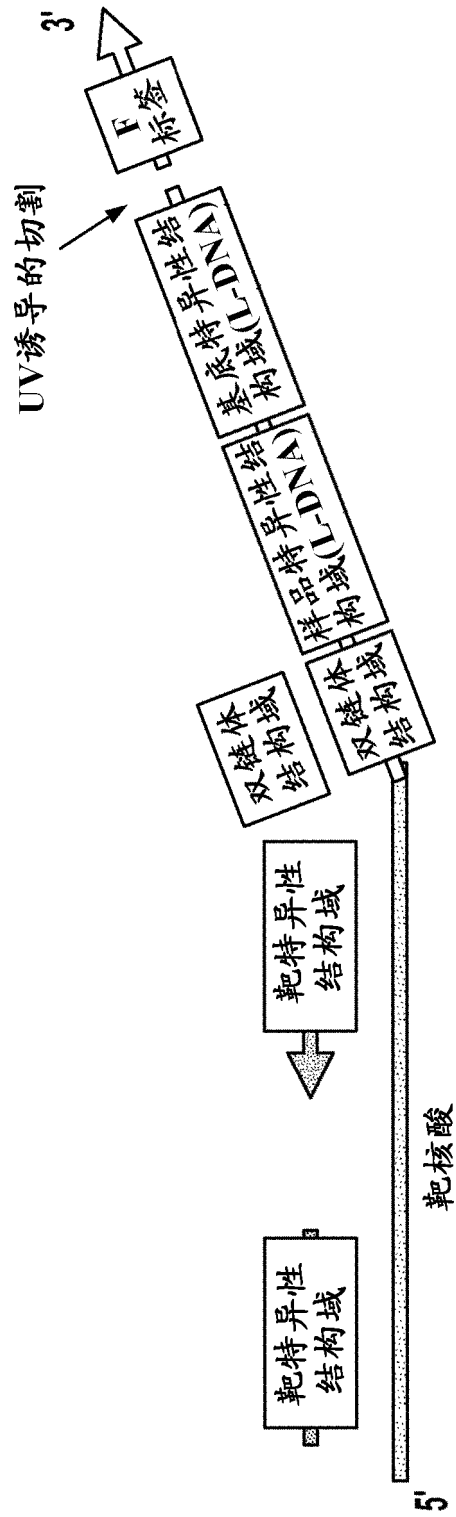


图 39

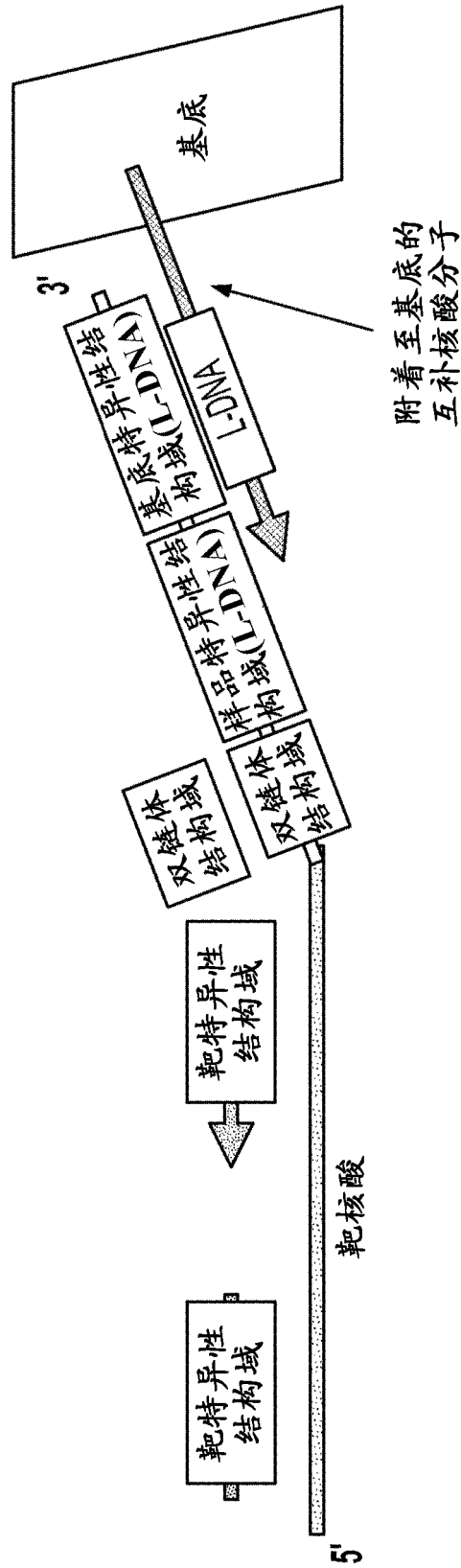


图 40

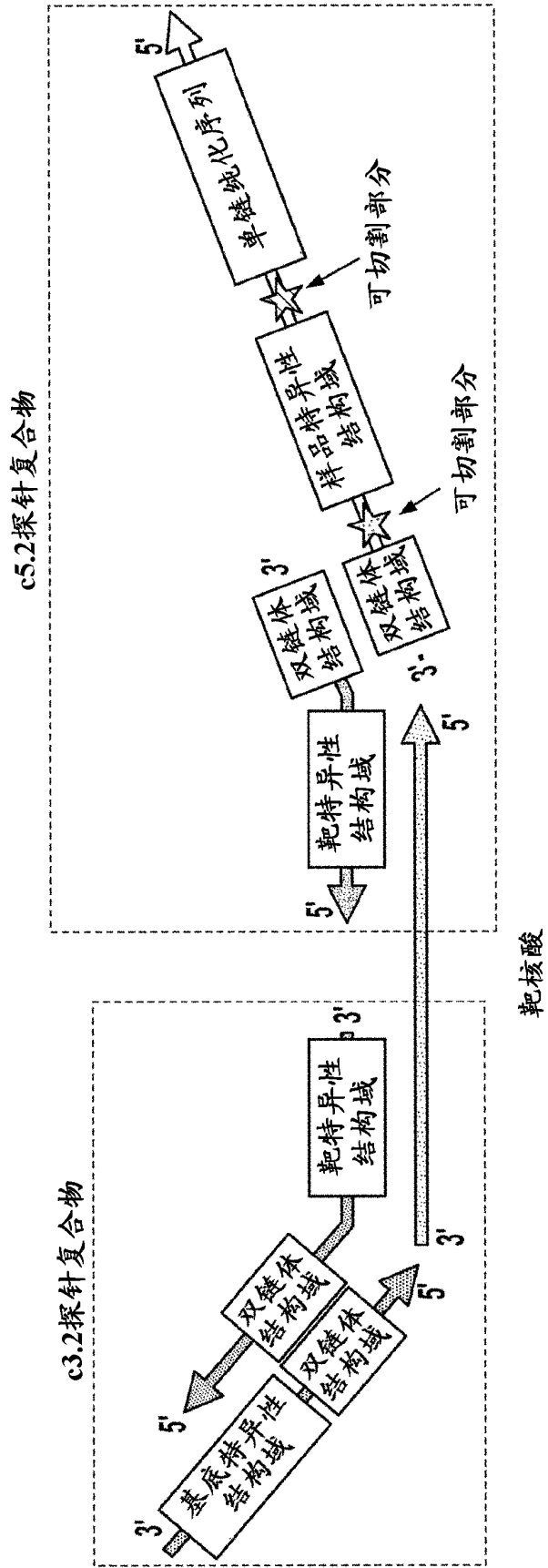


图 41

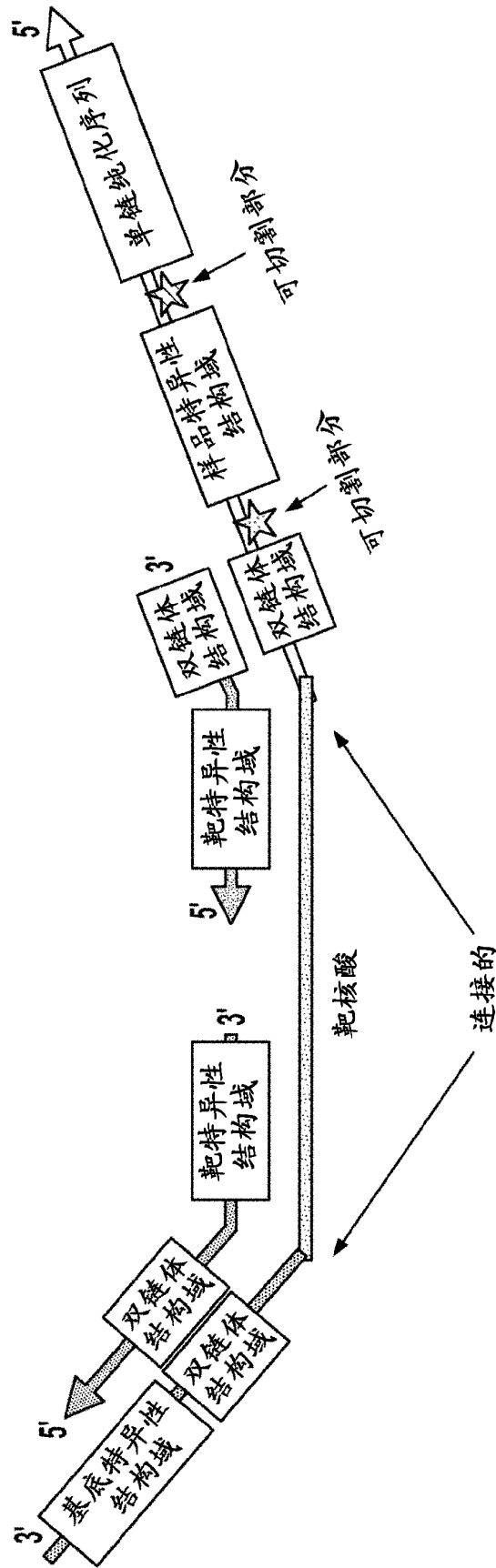


图 42

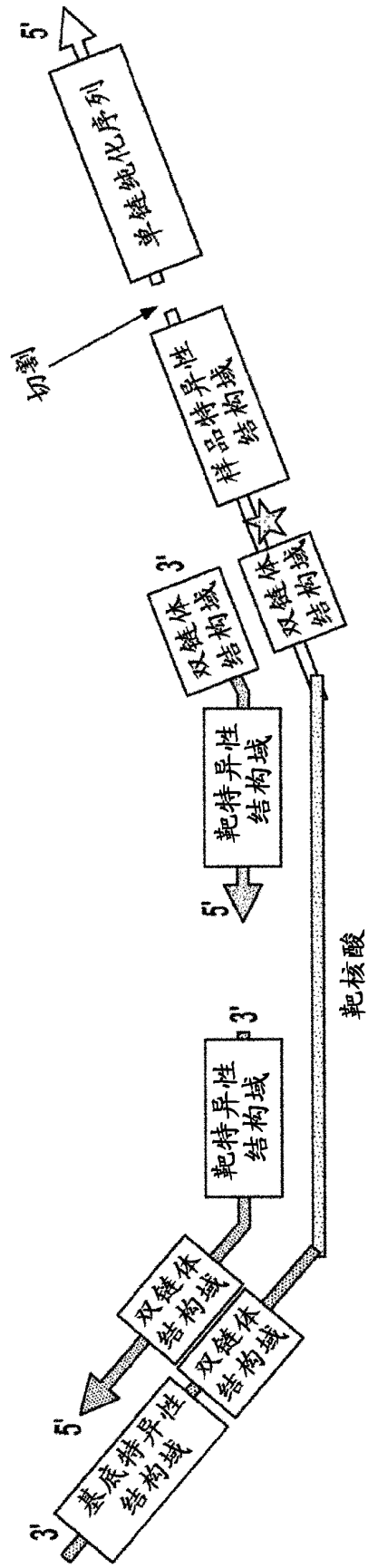


图 43

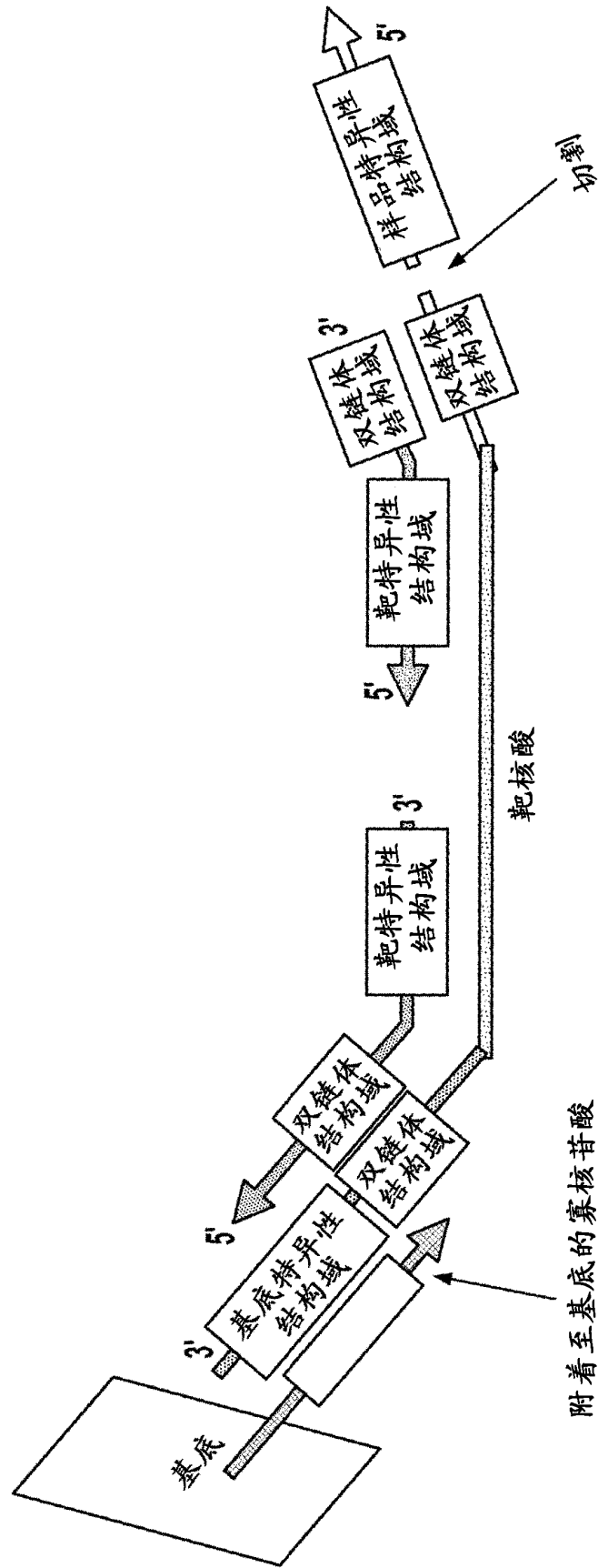


图 45

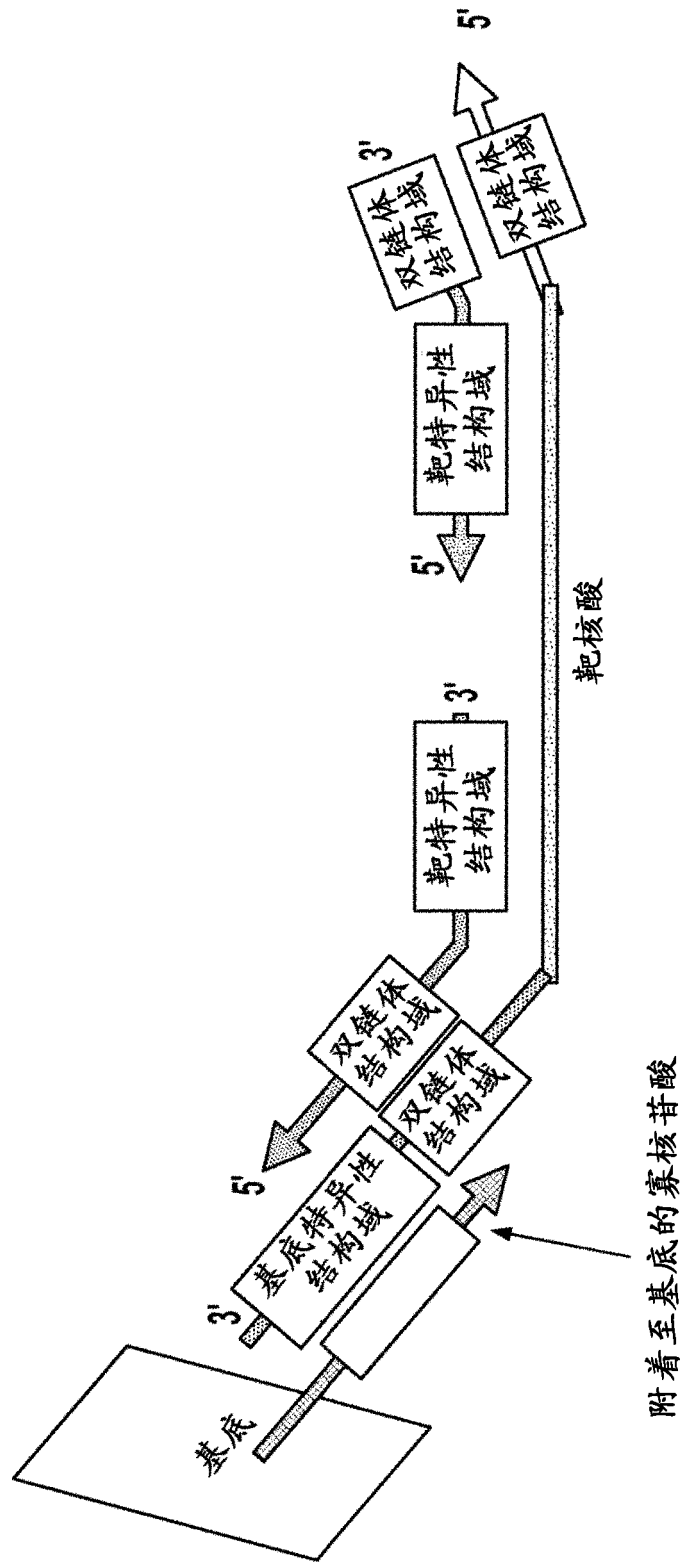


图 46

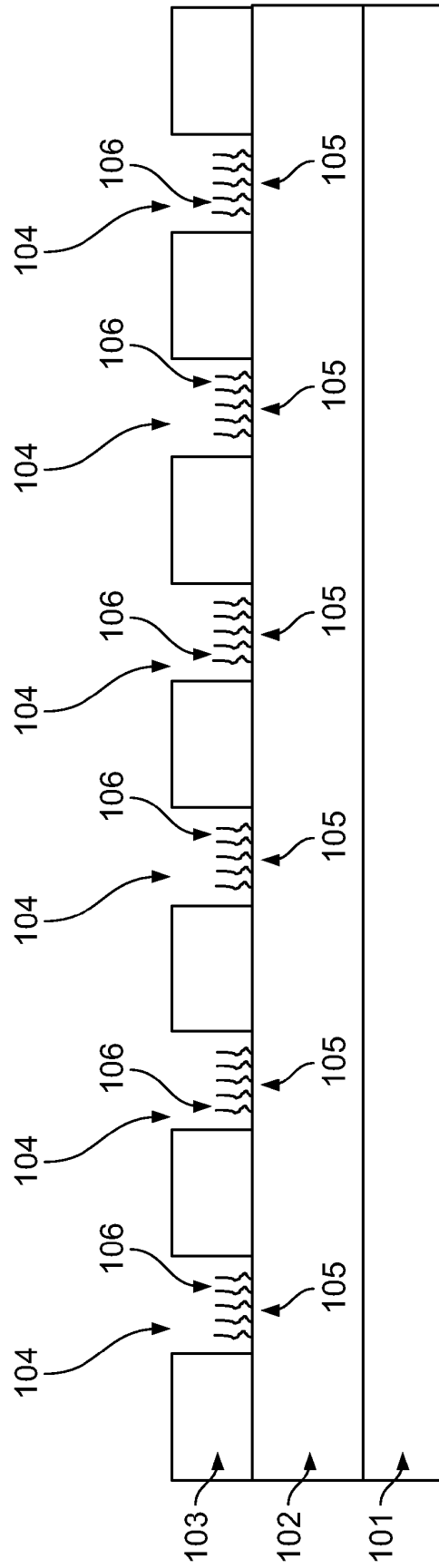


图 47

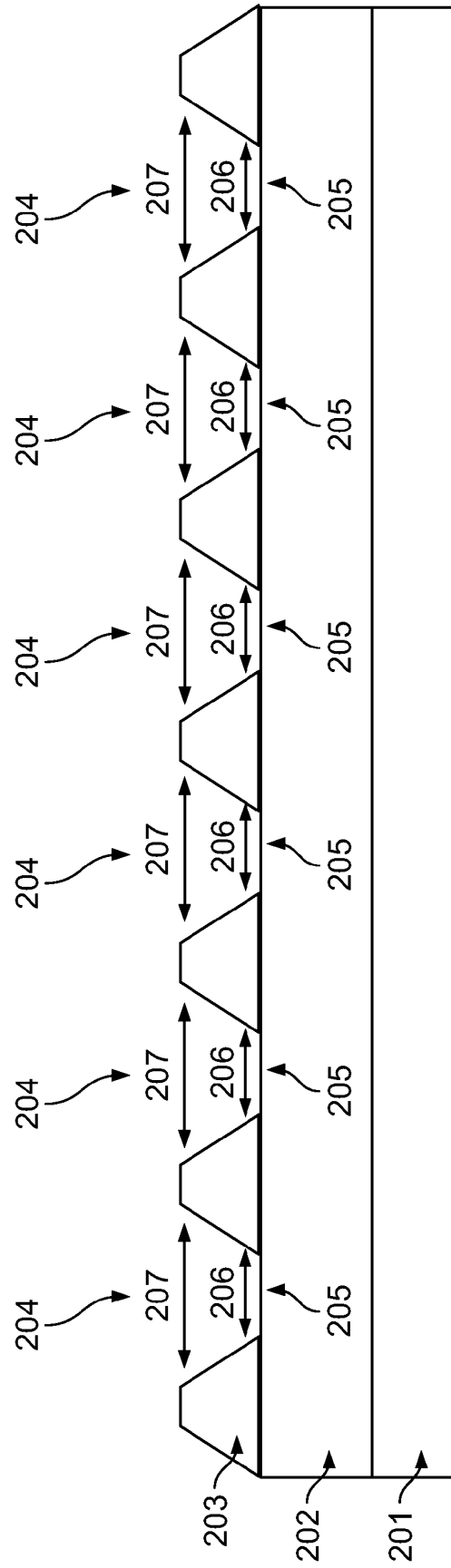


图 48

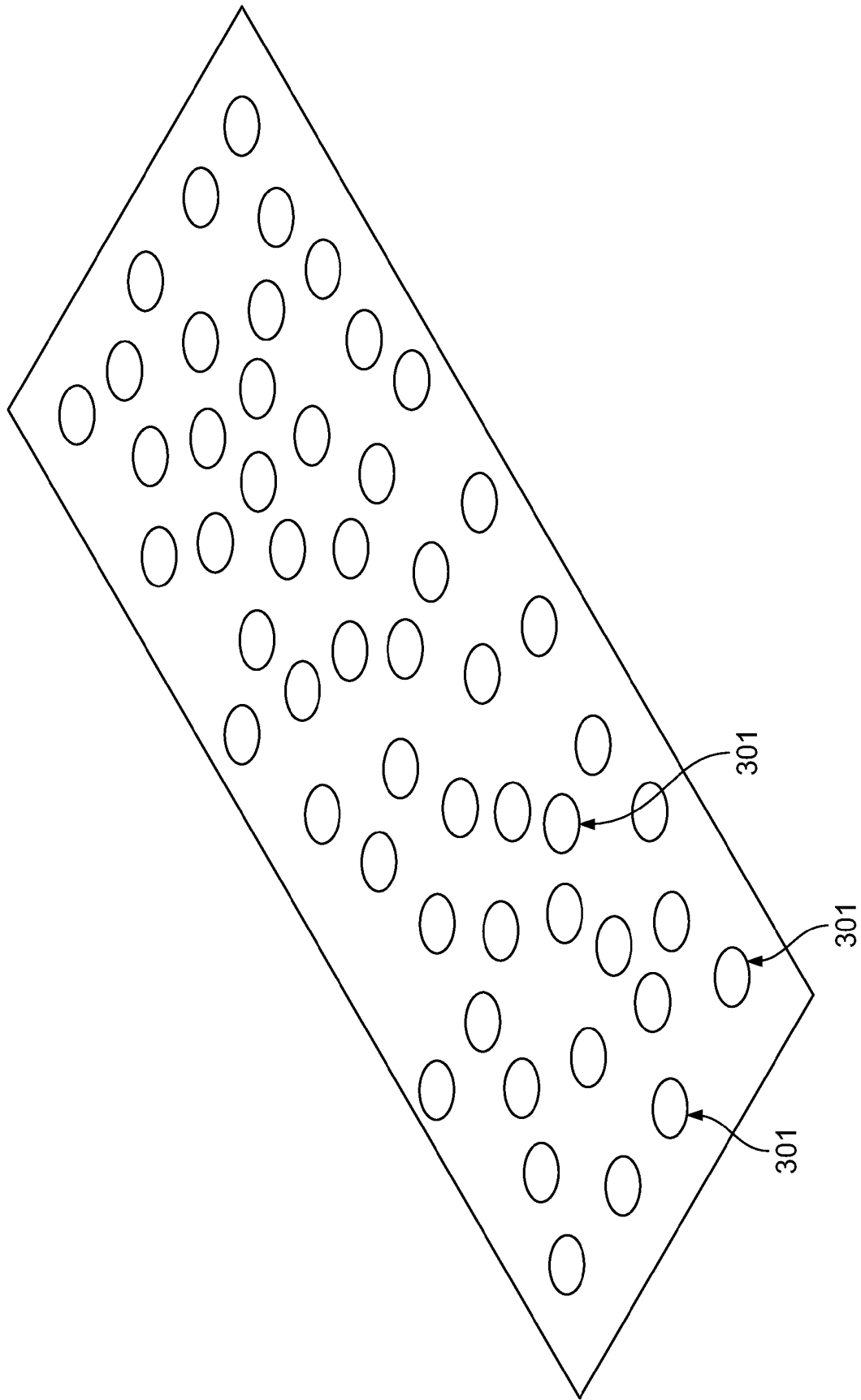


图 49

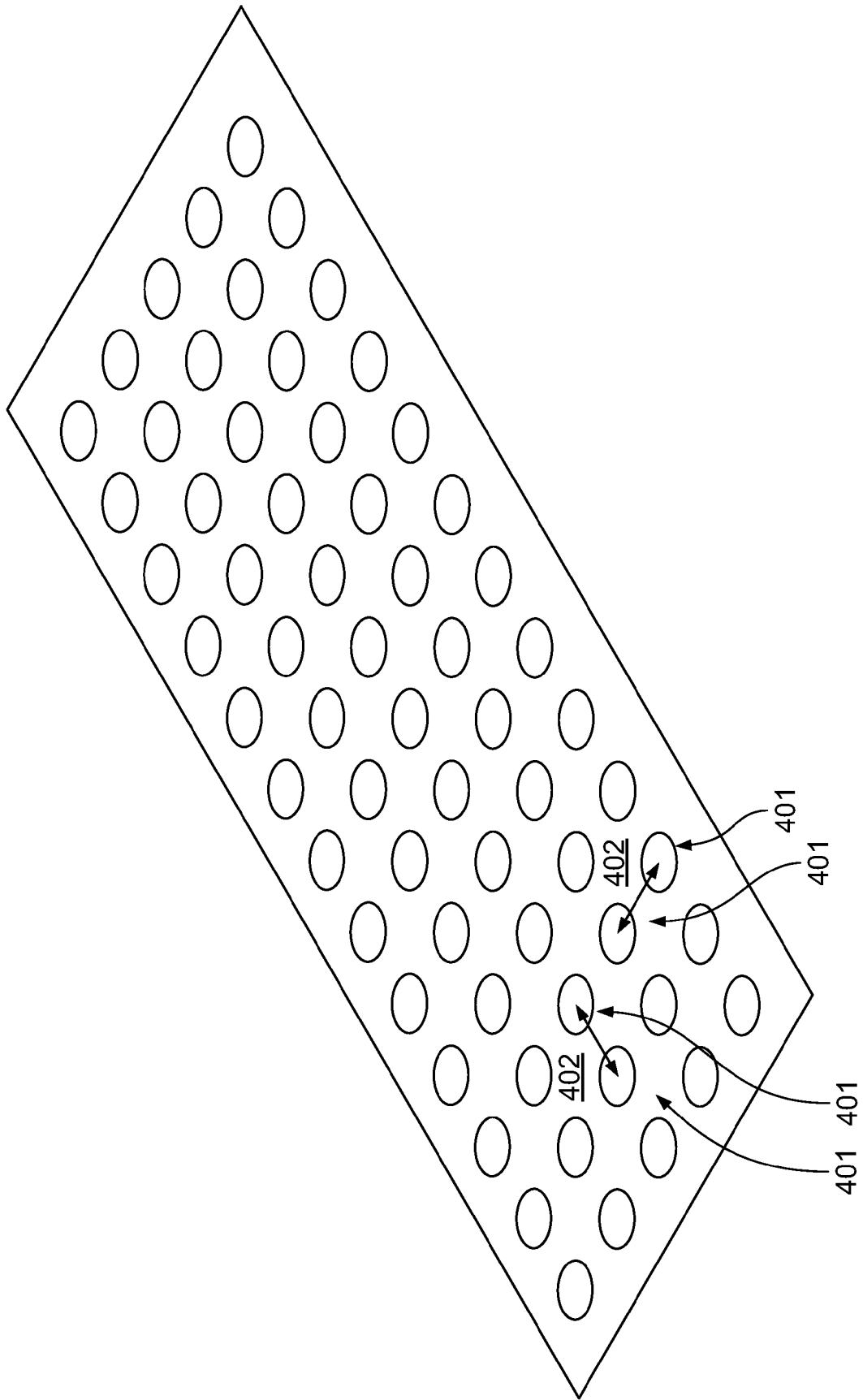


图 50

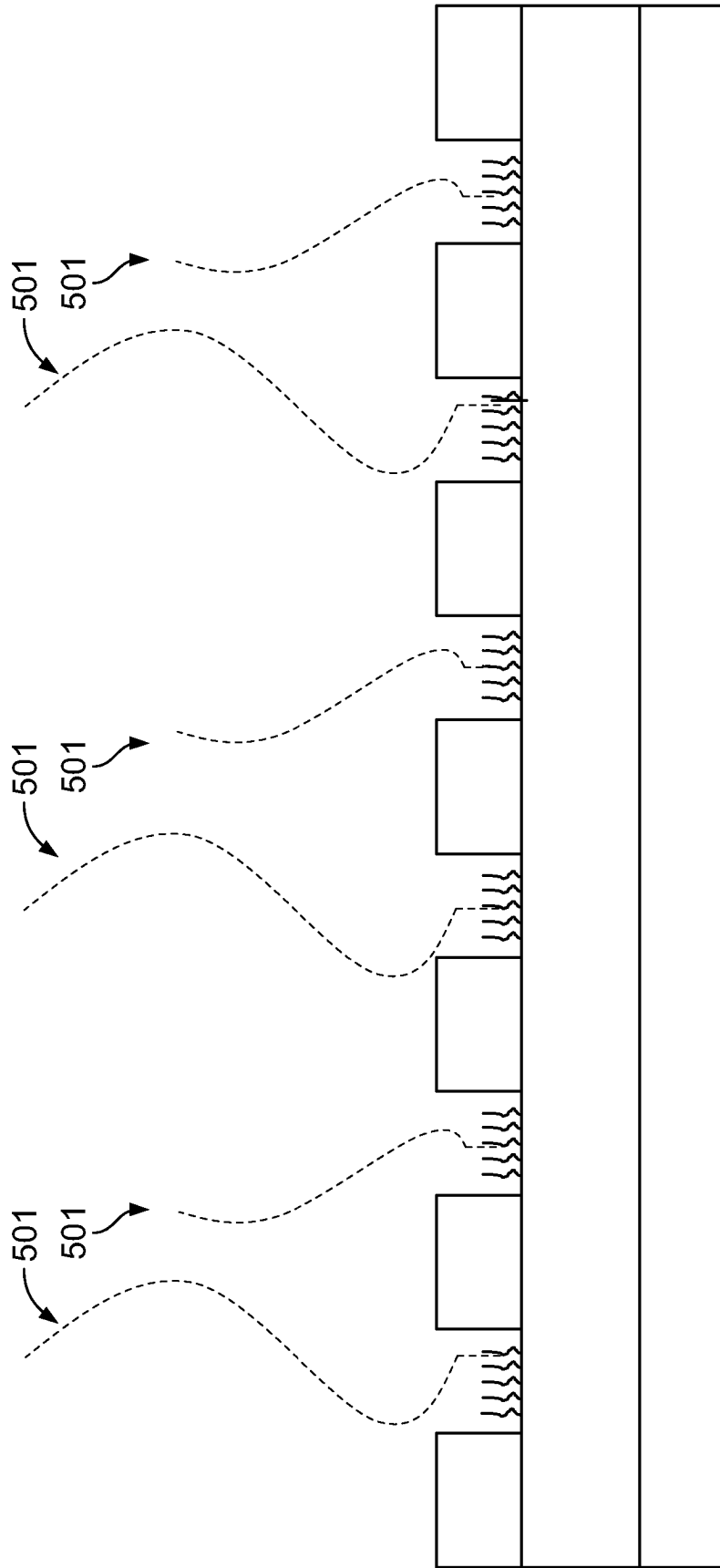


图 51

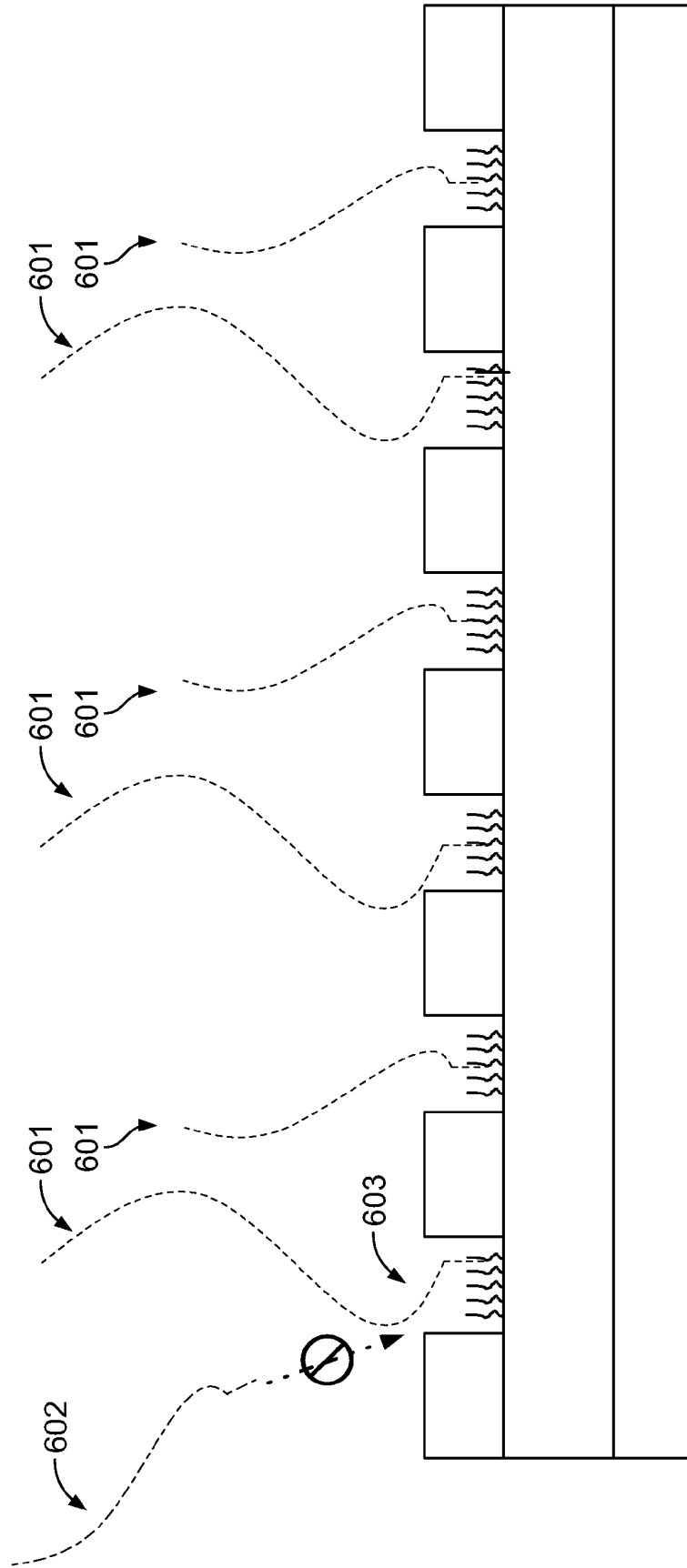


图 52

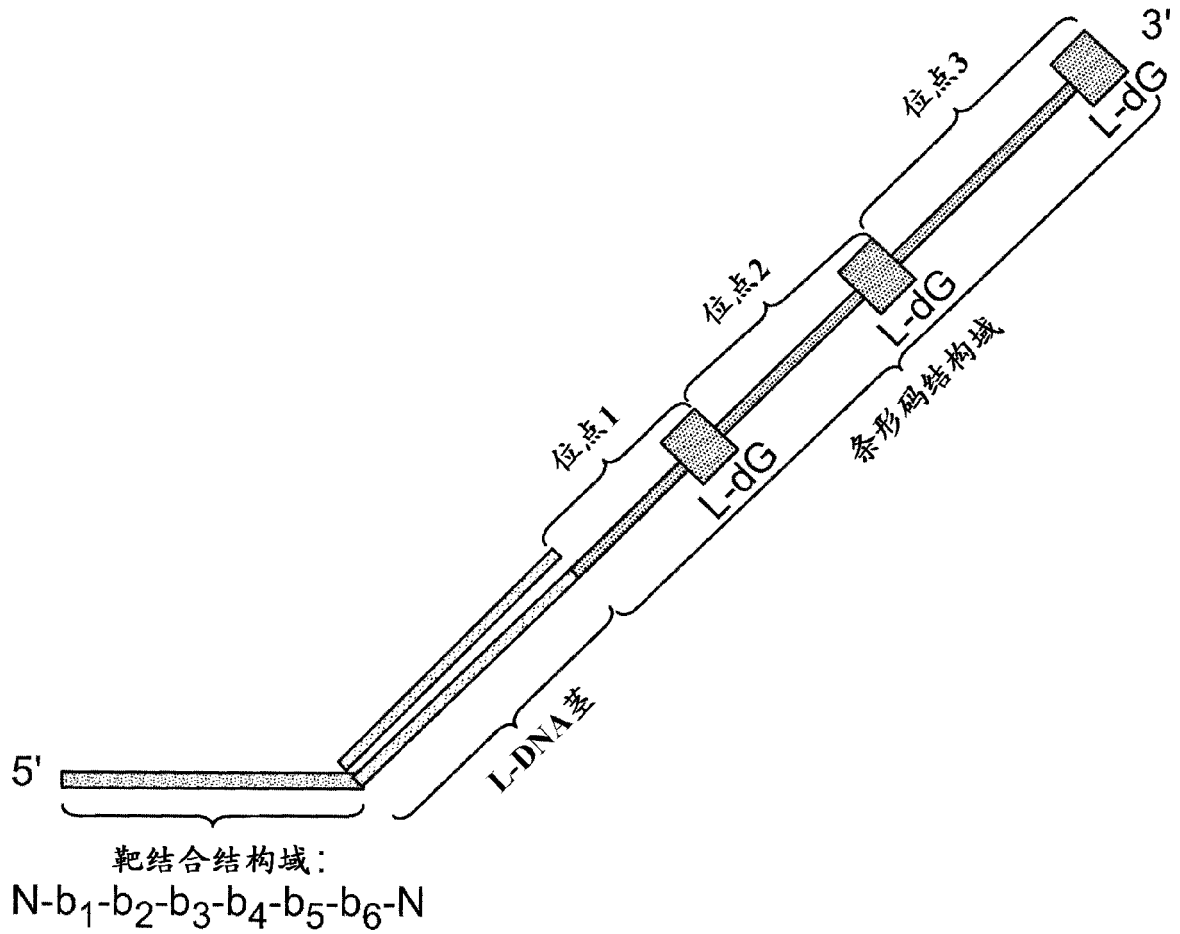
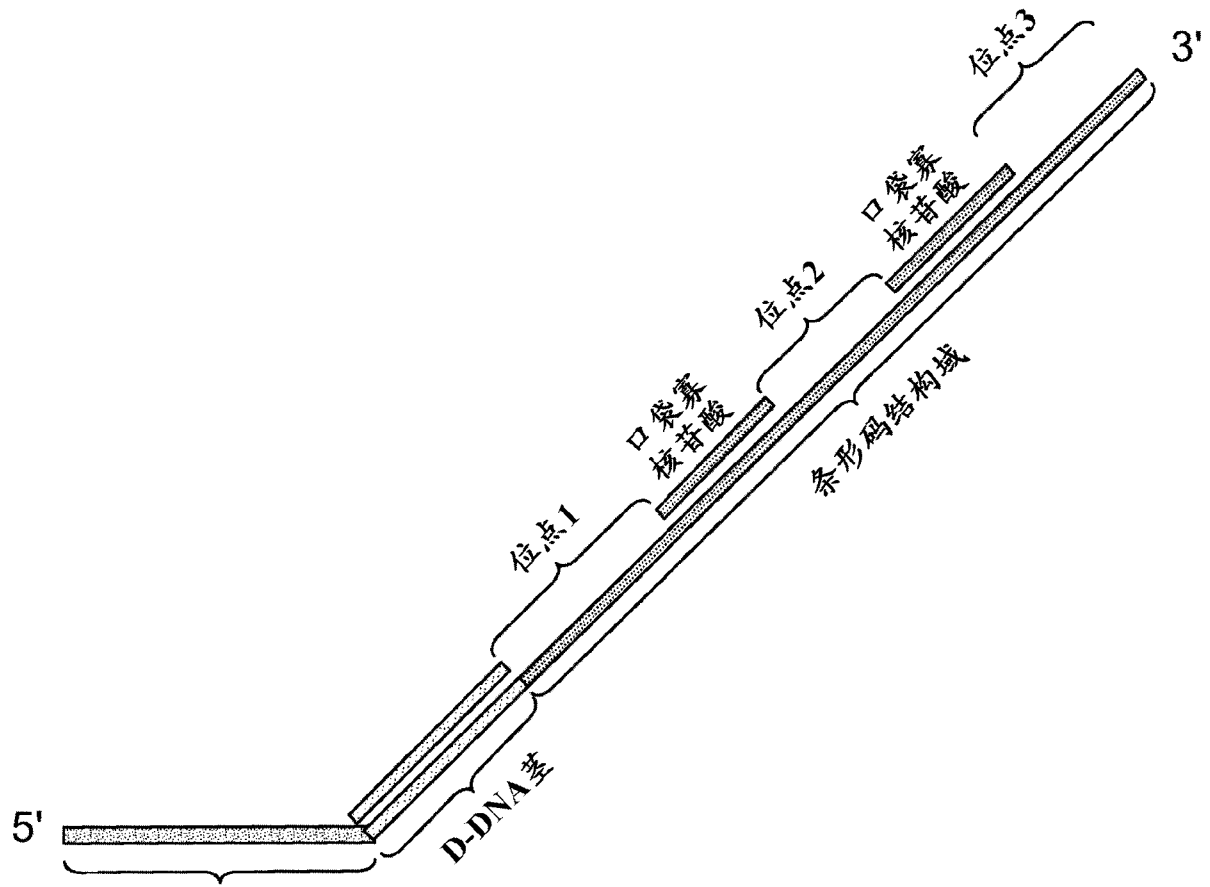


图 53



靶结合结构域：
N-b₁-b₂-b₃-b₄-b₅-b₆-N

图 54

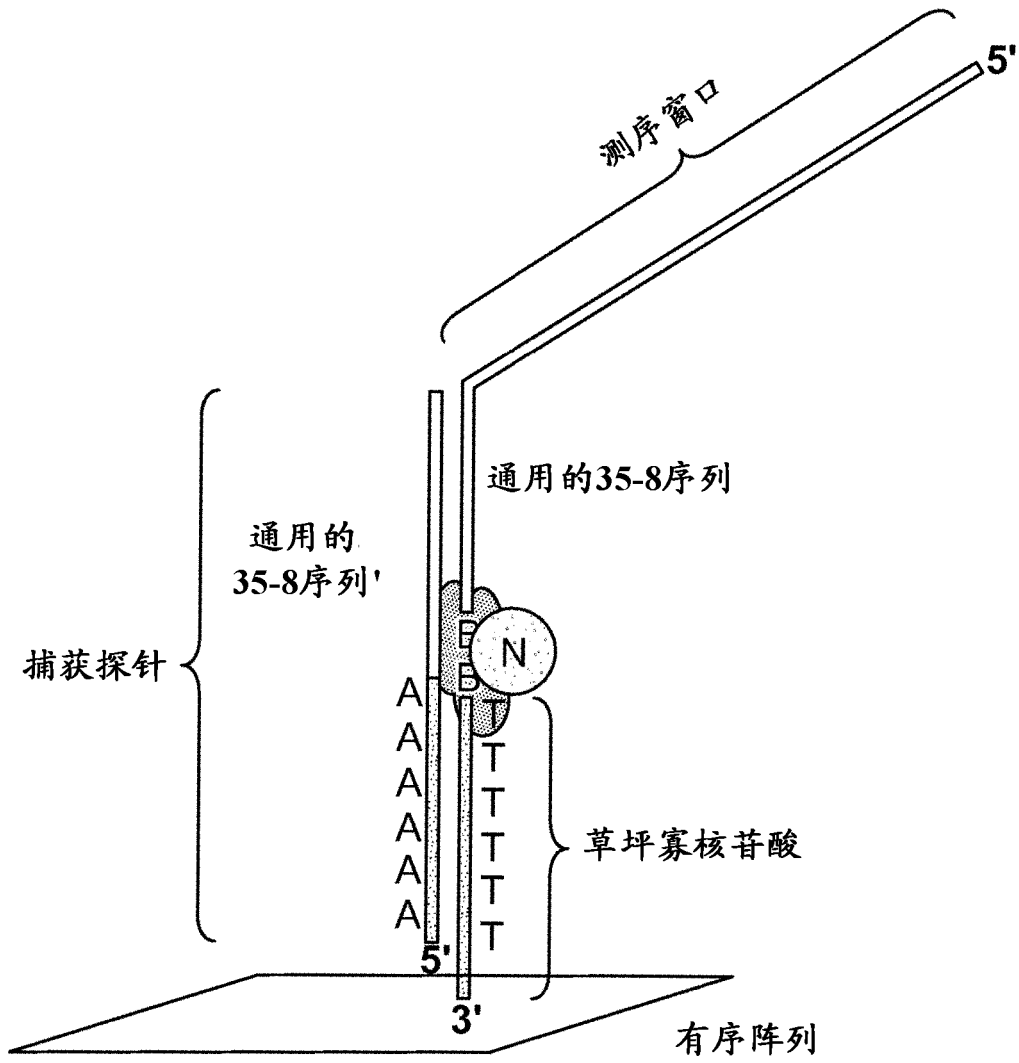


图 55

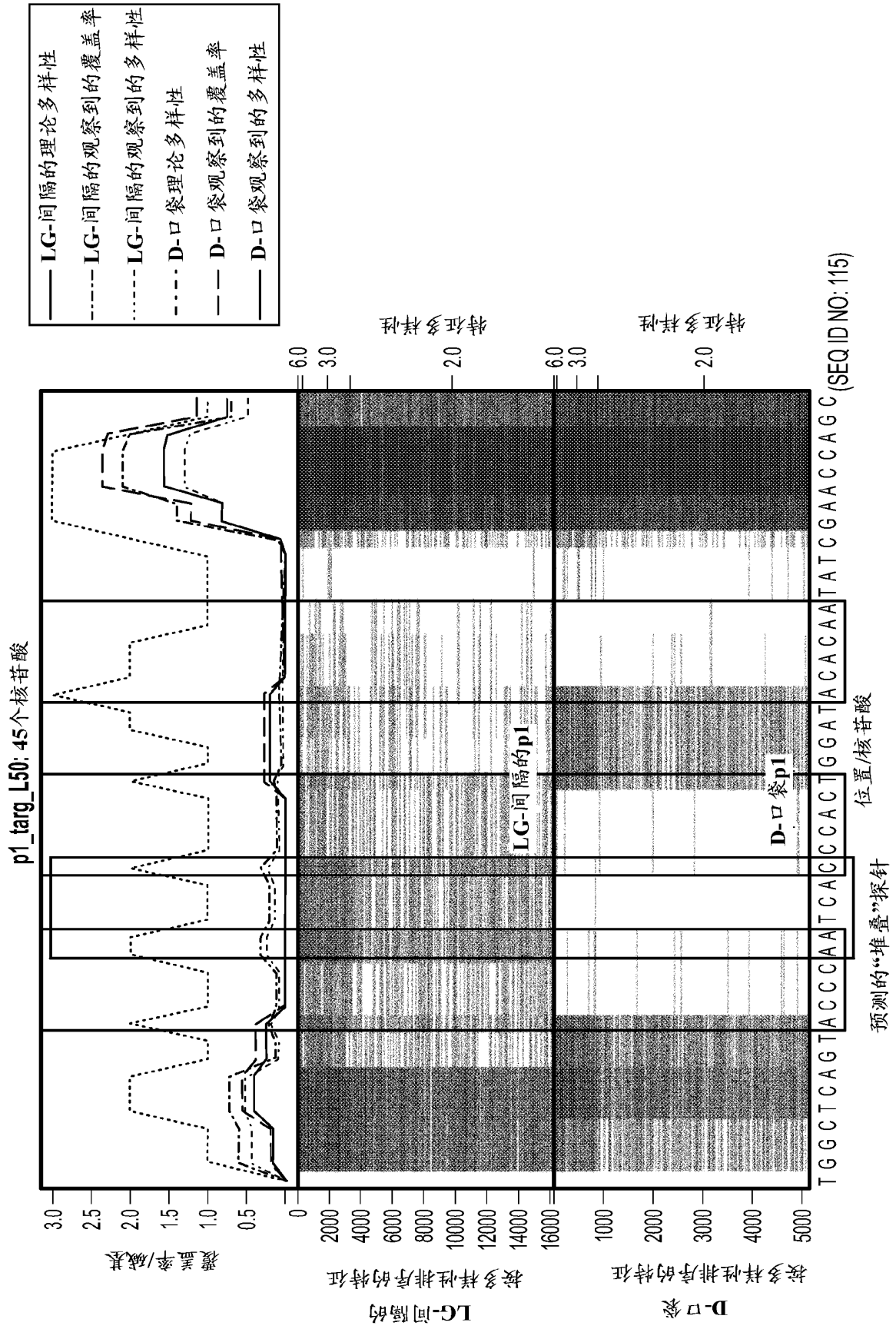


图 56

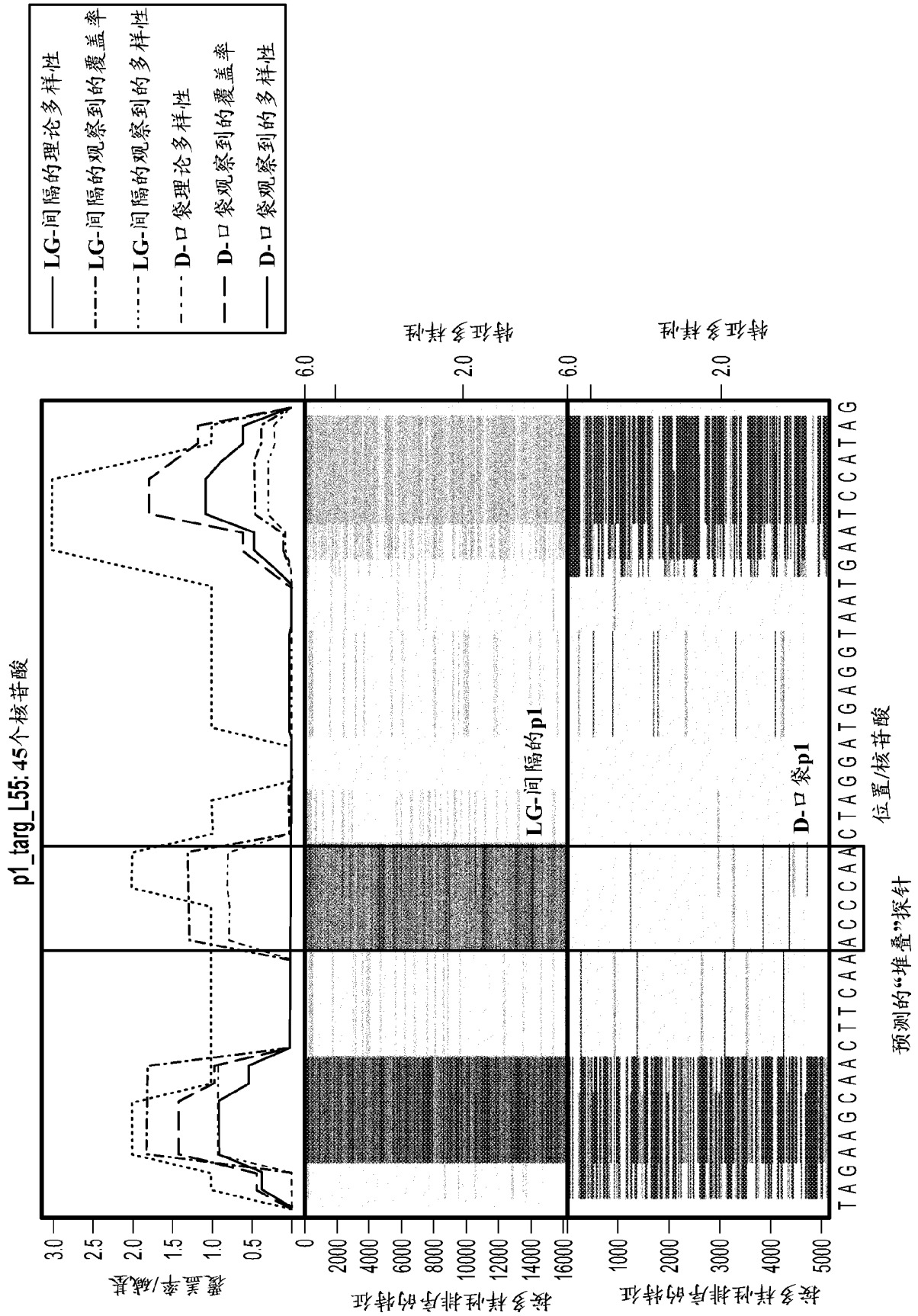


图 57

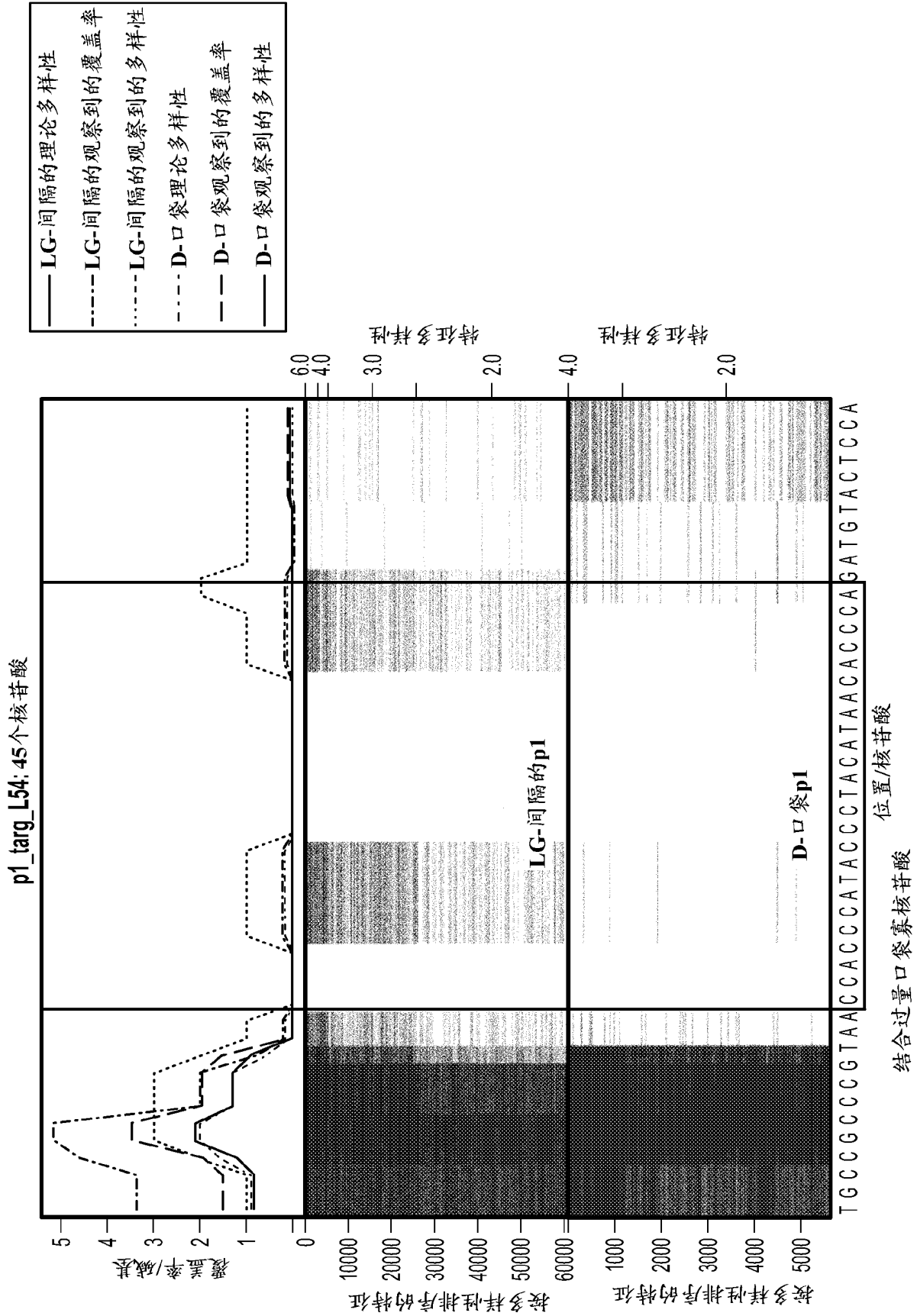


图 58

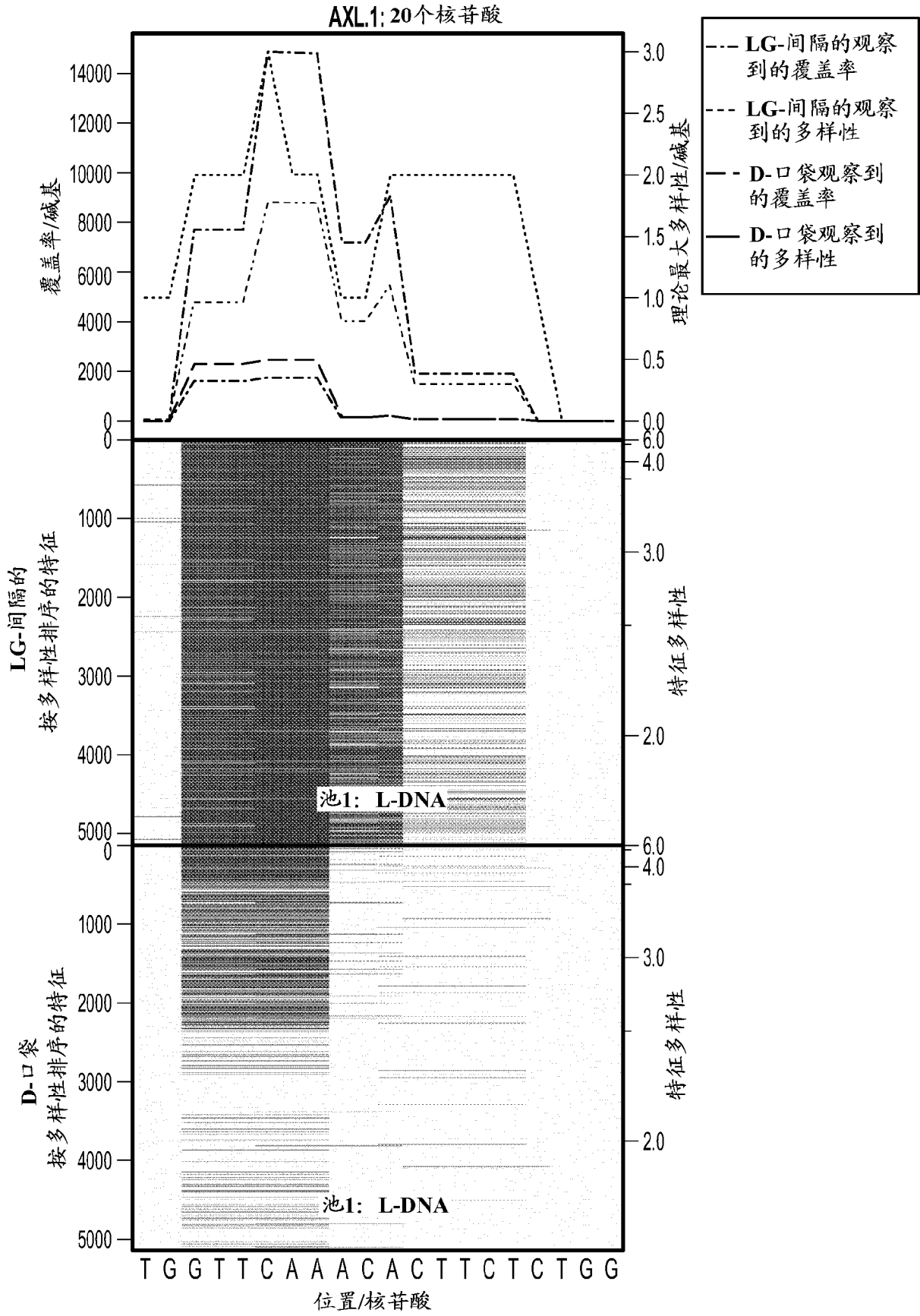


图 59

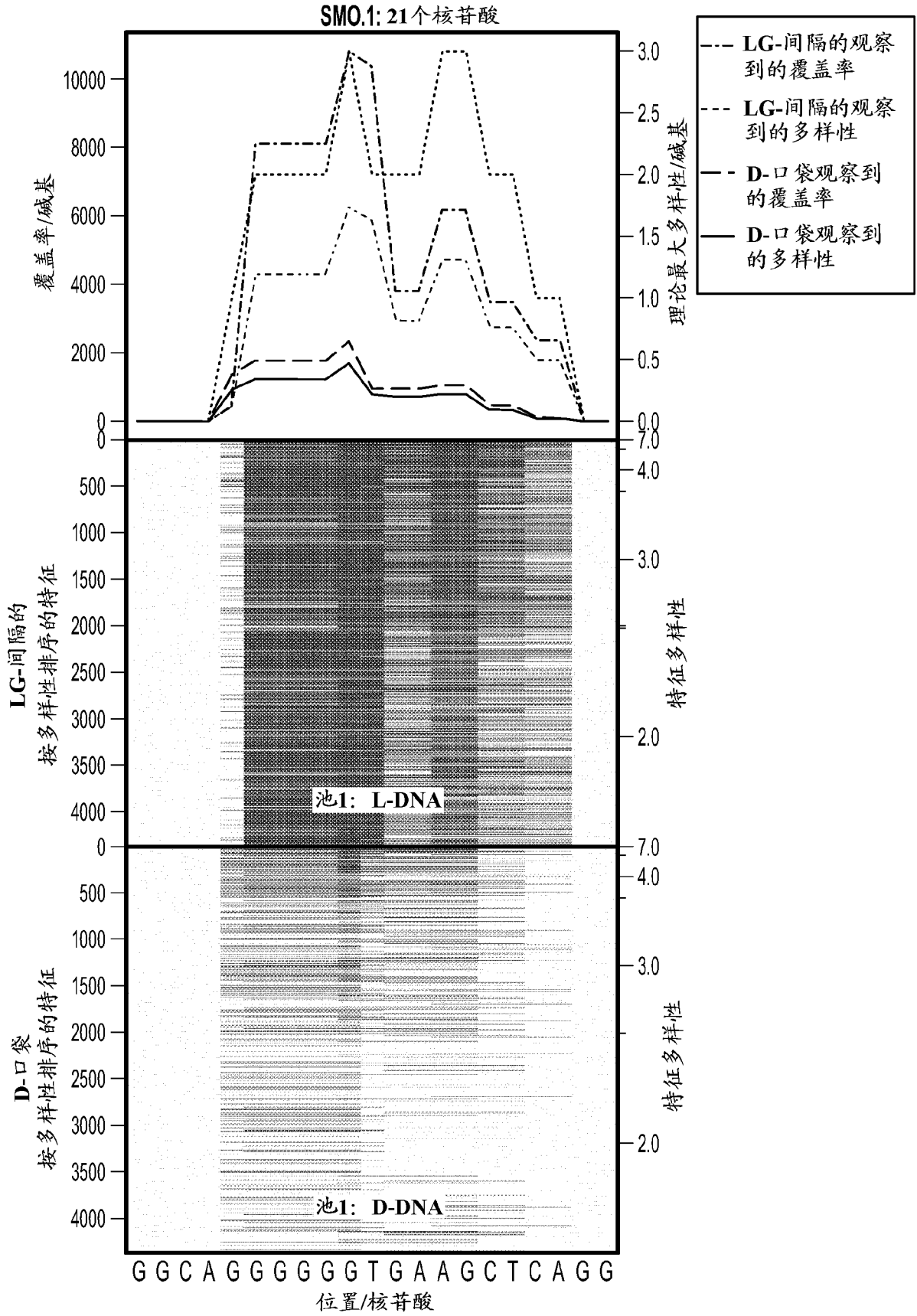


图 60

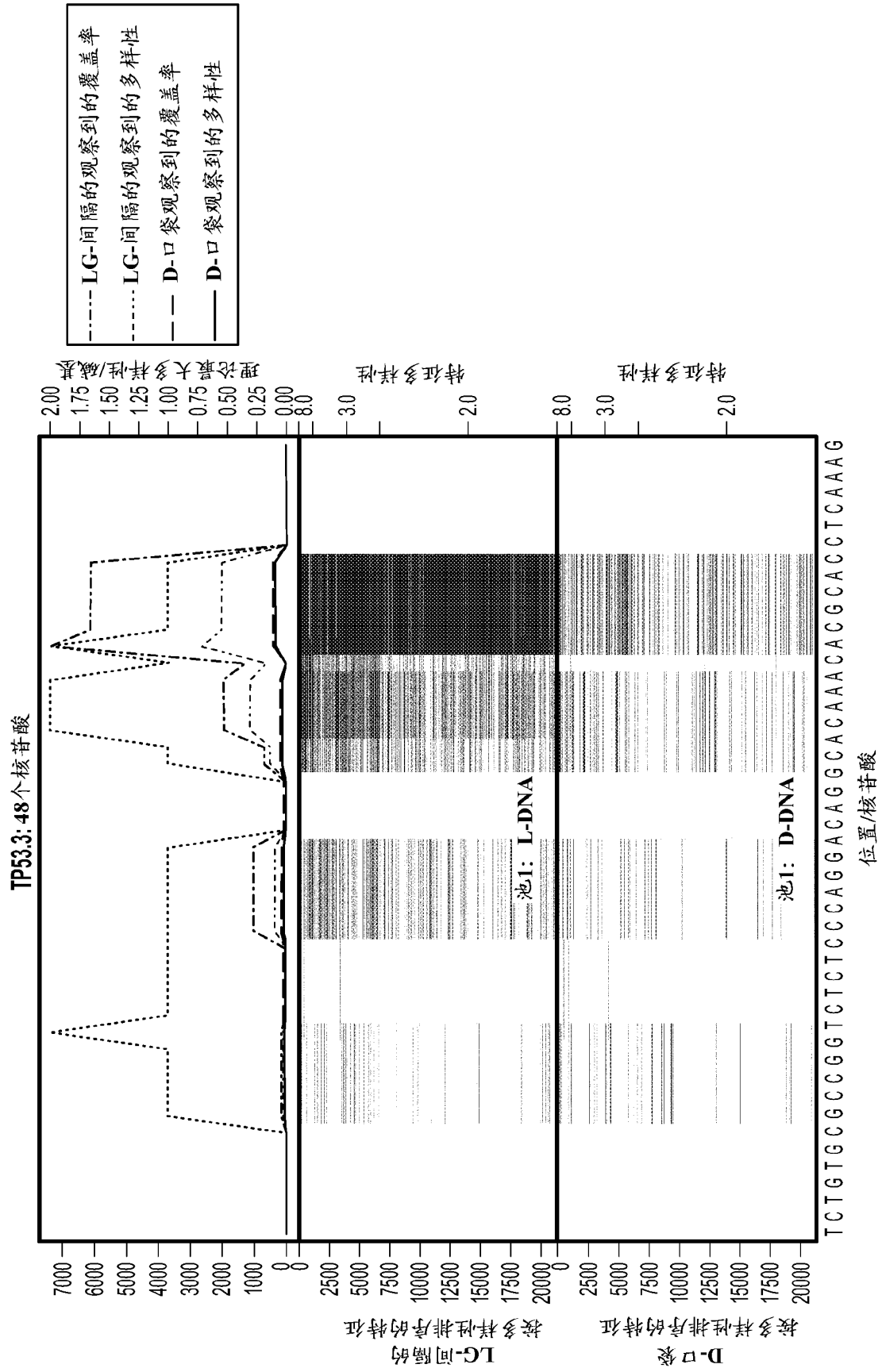


图 19

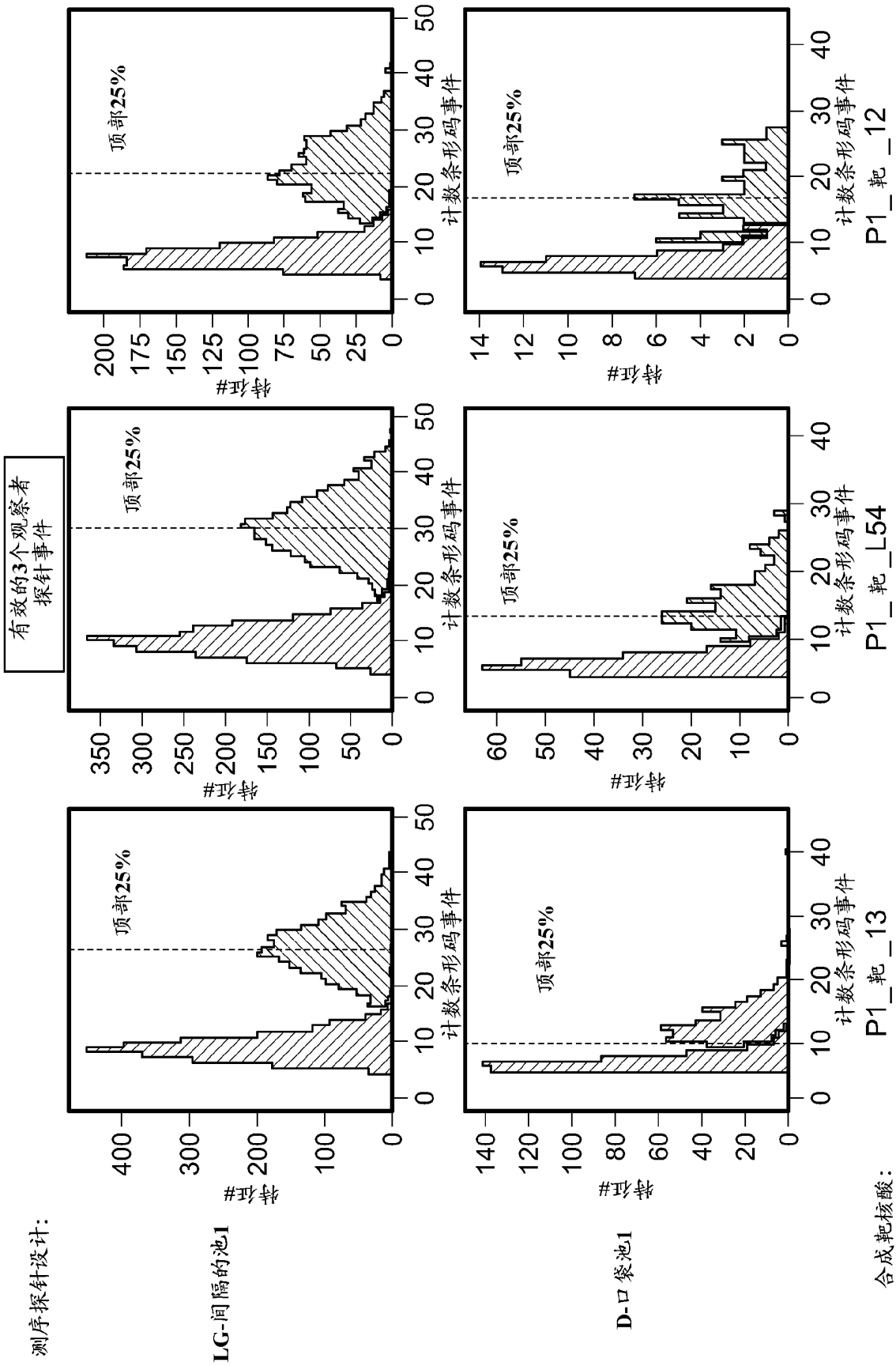


图 62

所有合成靶核酸的总和

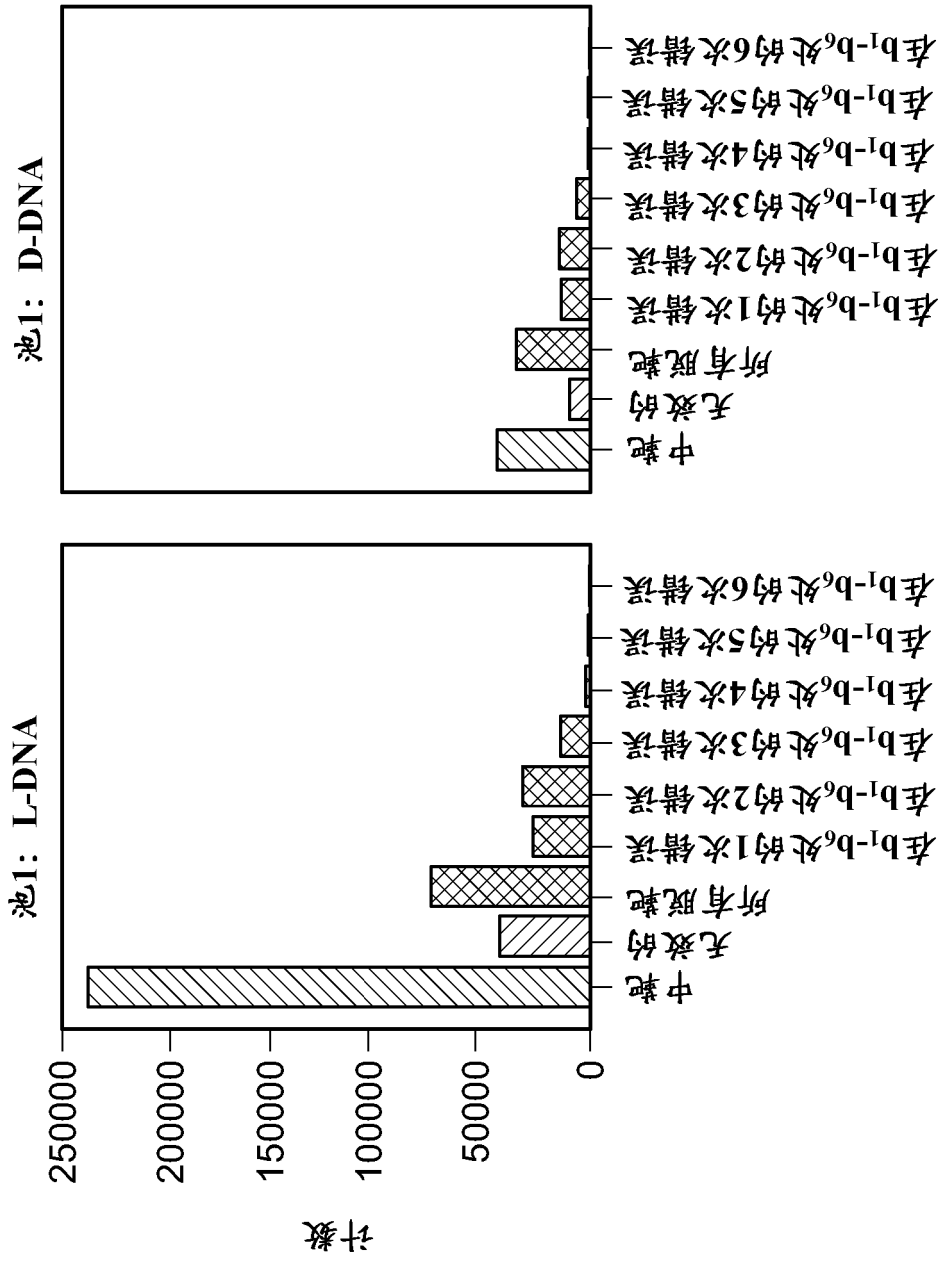


图 63

TP53.3

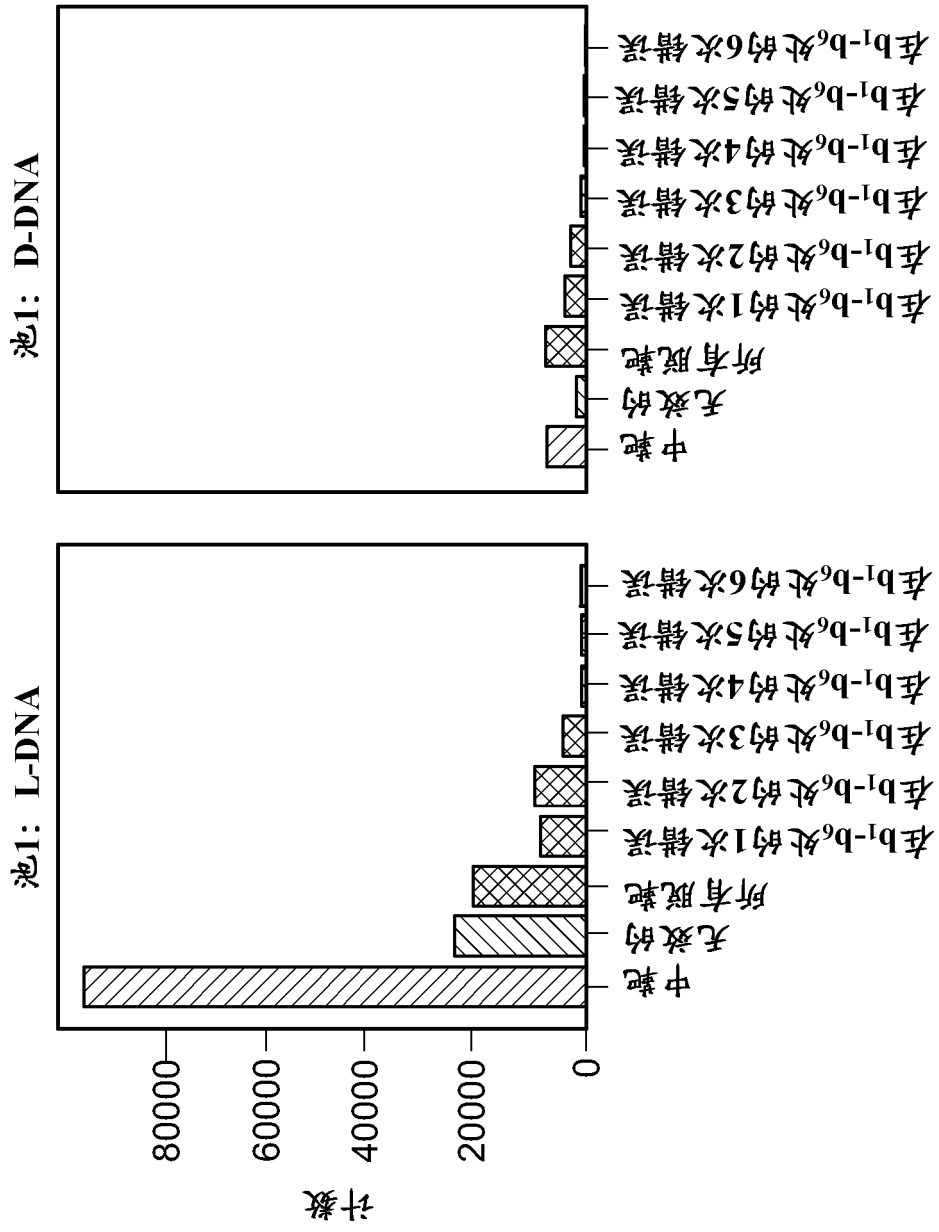


图 64

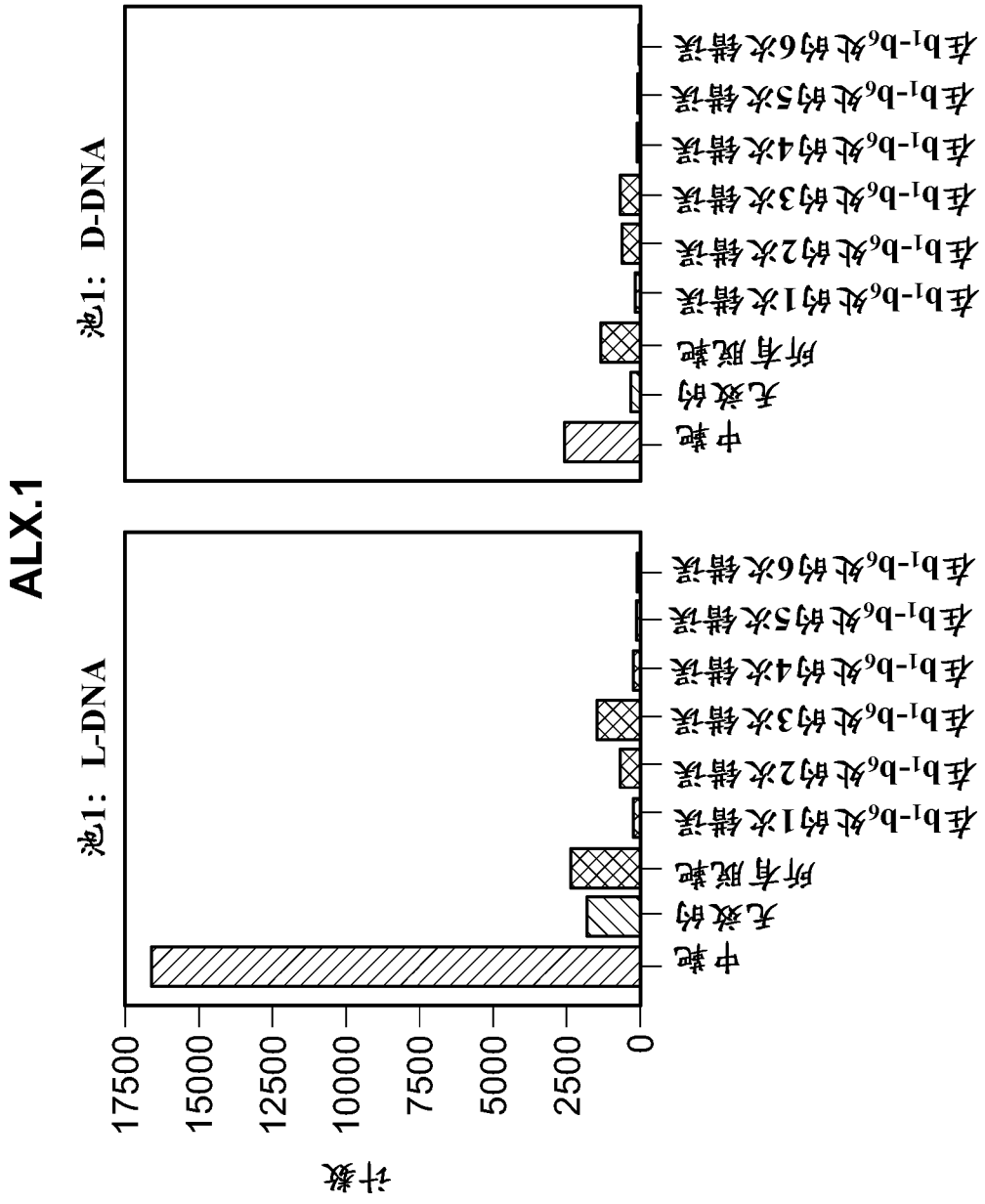


图 64(续)

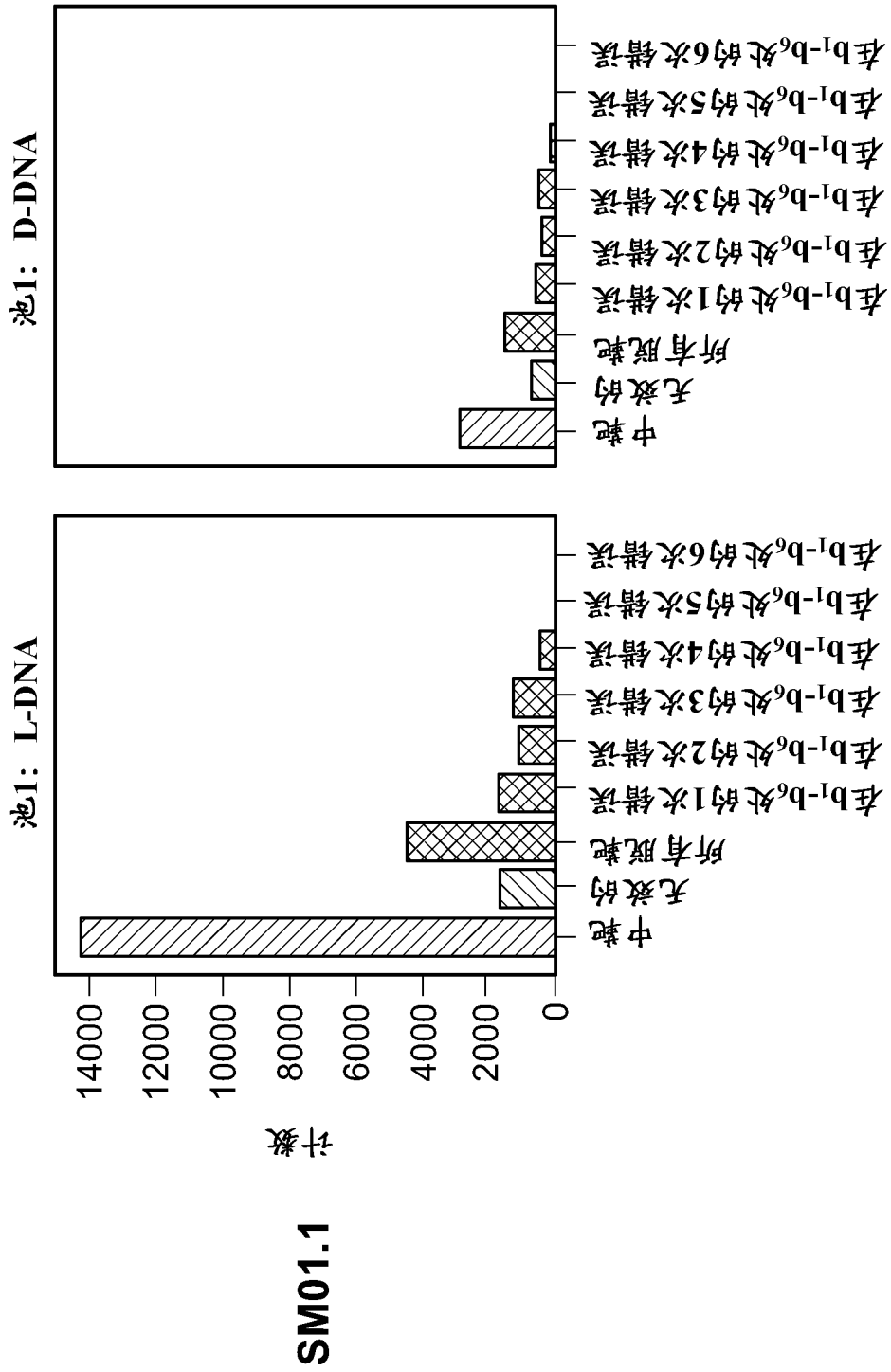


图 64(续)

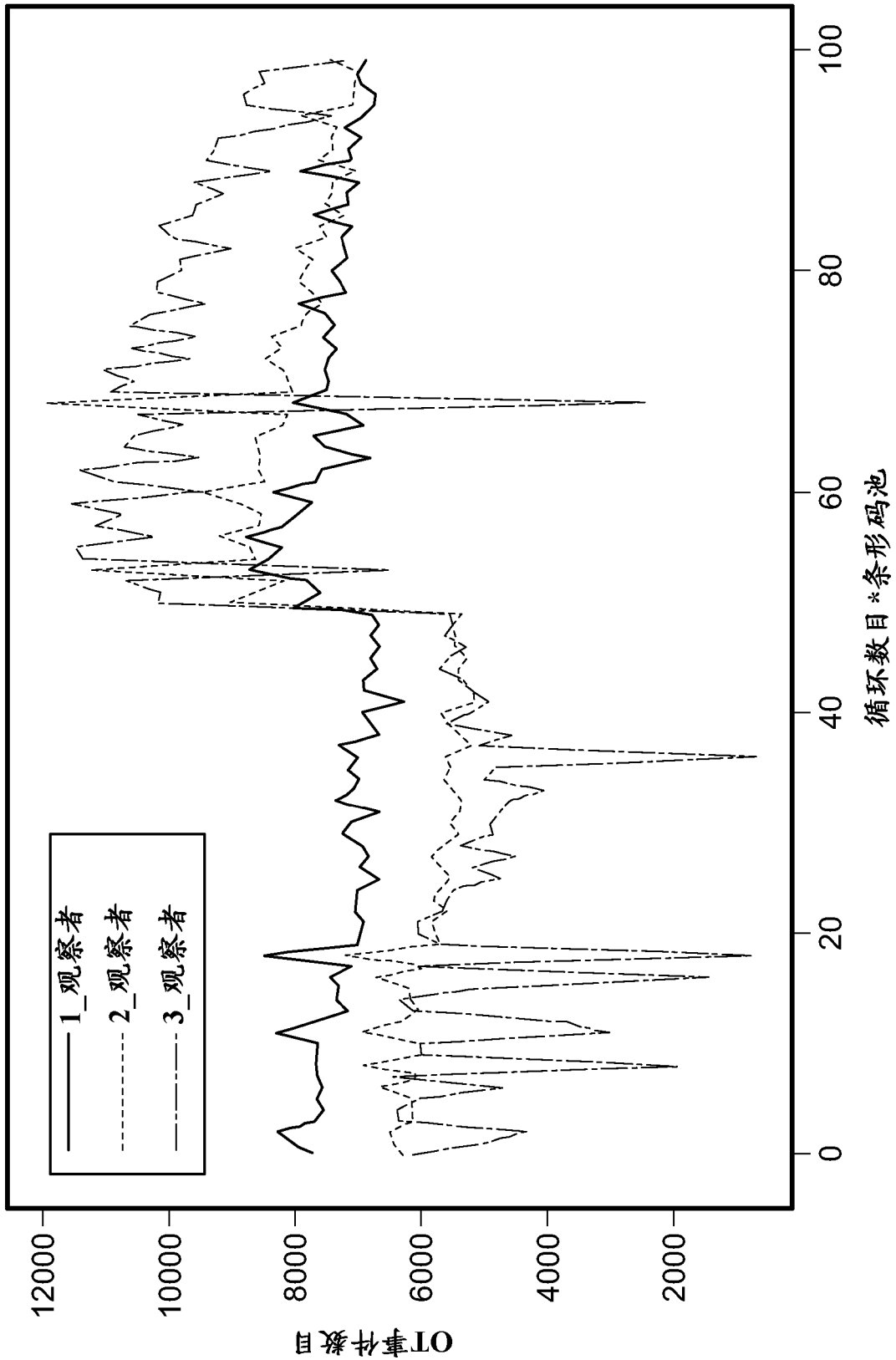


图 65

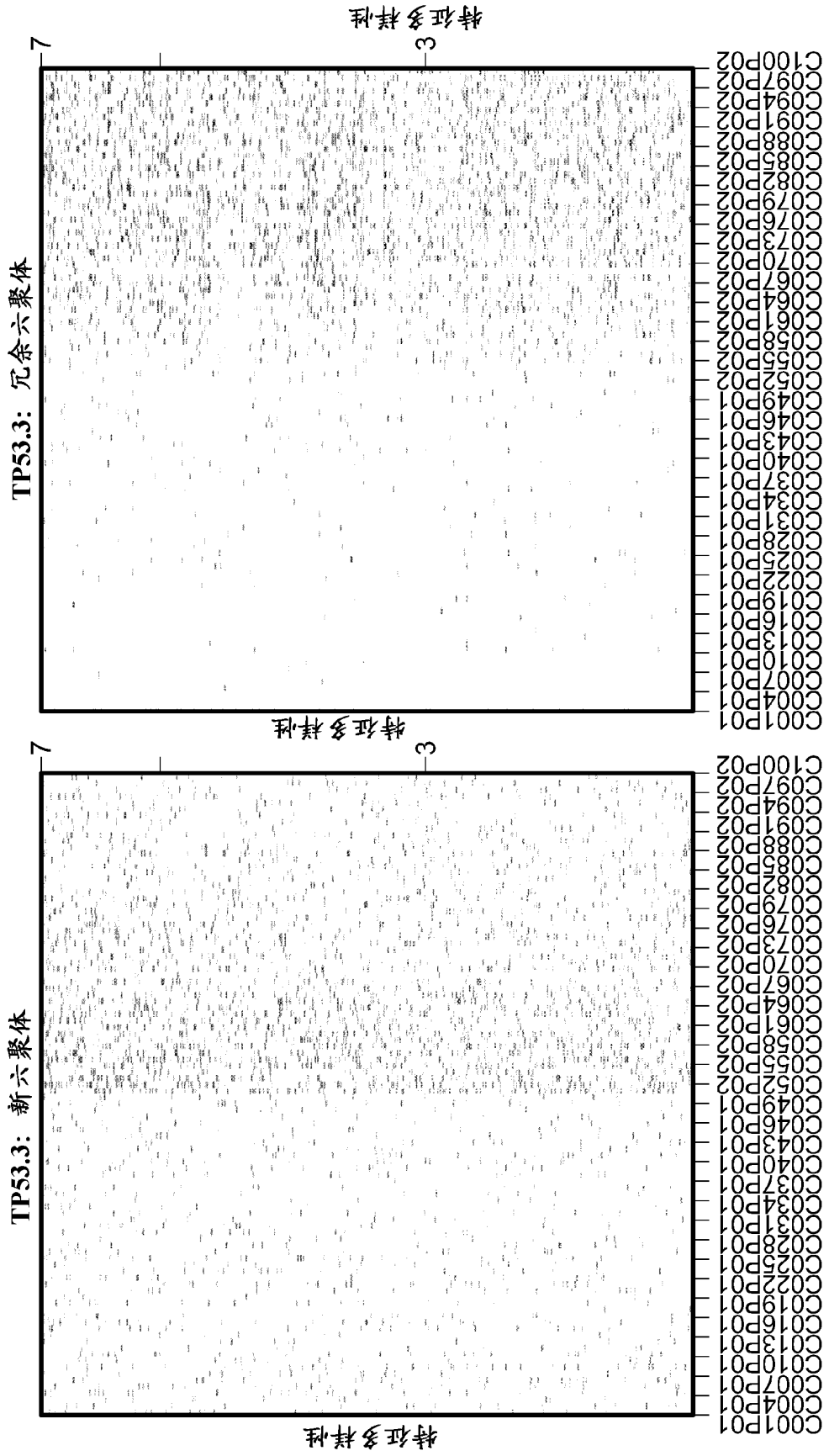


图 66 (续)

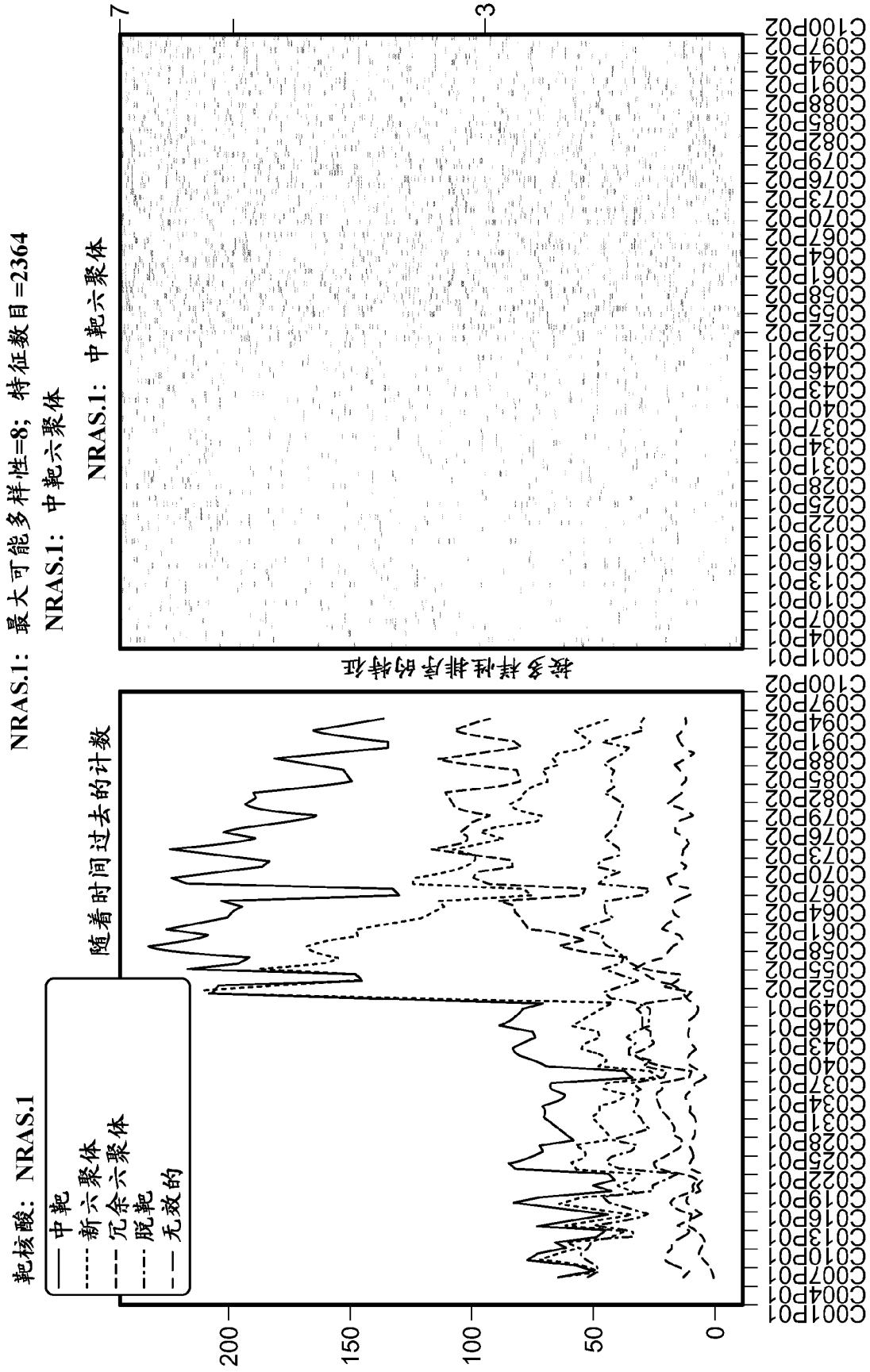


图 66 (续)

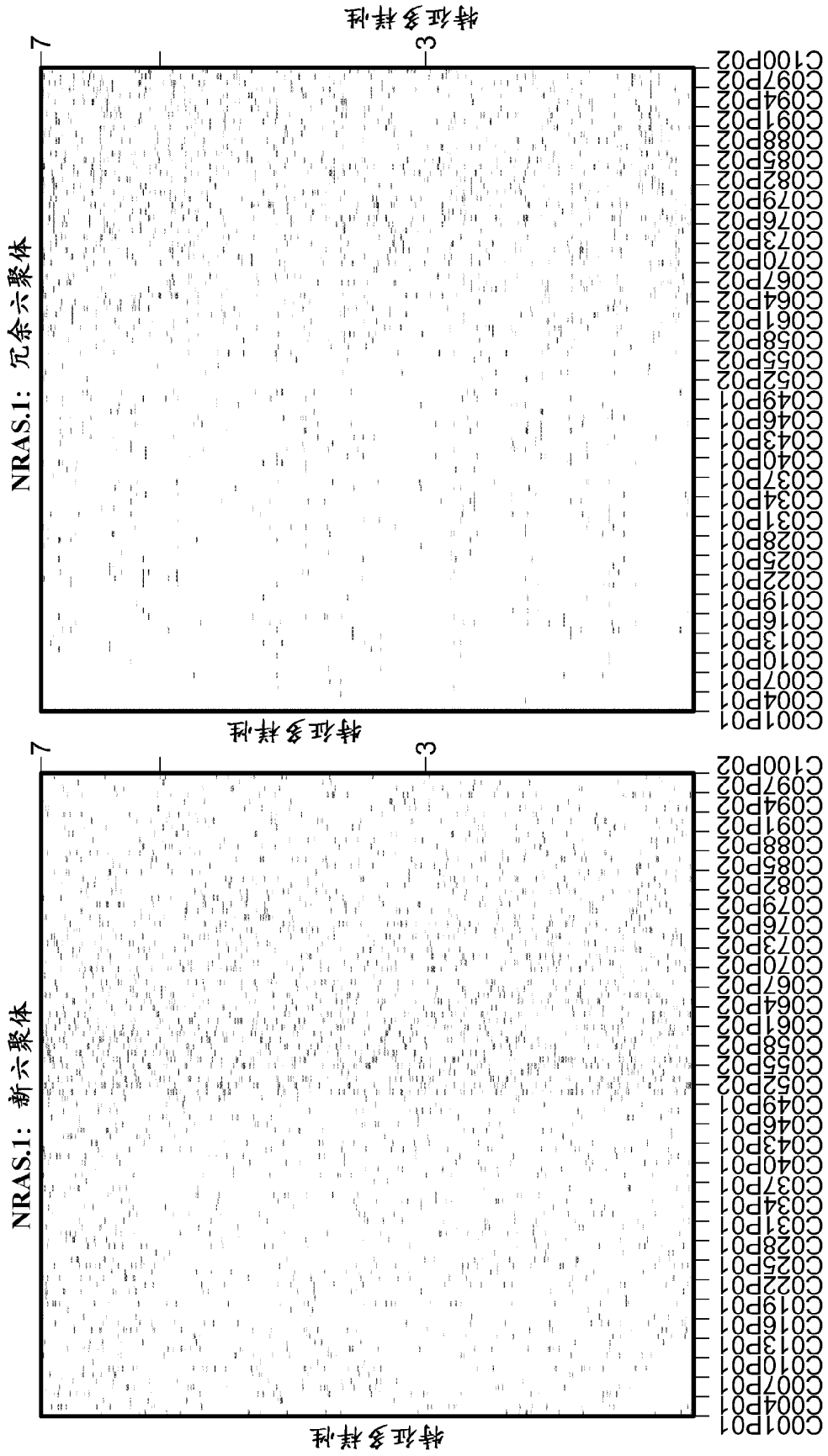


图 66 (续)

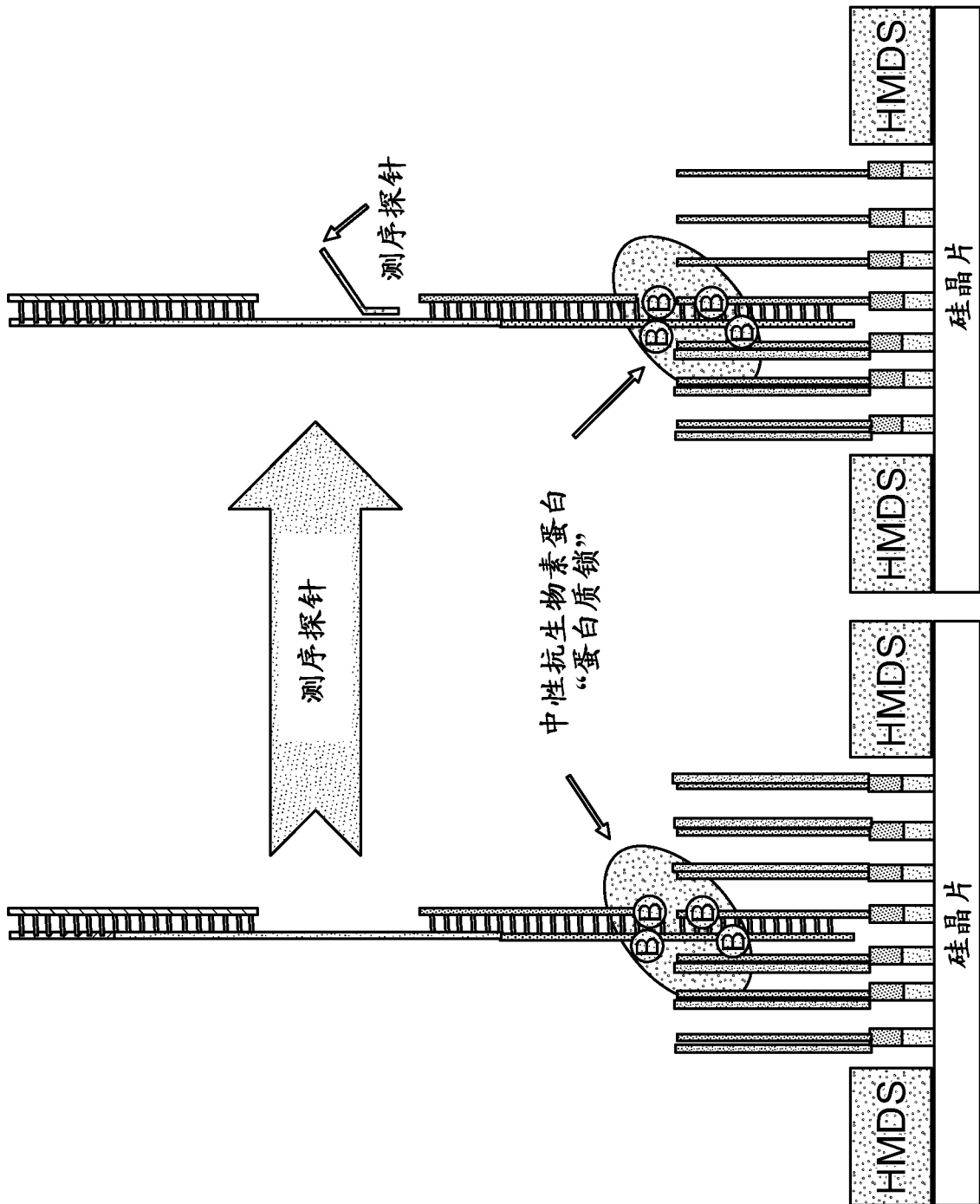


图 67