



(12) 发明专利申请

(10) 申请公布号 CN 102292451 A

(43) 申请公布日 2011. 12. 21

(21) 申请号 200980125335. 3

(51) Int. Cl.

(22) 申请日 2009. 06. 30

C12Q 1/68 (2006. 01)

(30) 优先权数据

61/076, 785 2008. 06. 30 US

(85) PCT申请进入国家阶段日

2010. 12. 30

(86) PCT申请的申请数据

PCT/US2009/049244 2009. 06. 30

(87) PCT申请的公布数据

W02010/002883 EN 2010. 01. 07

(71) 申请人 生物纳米芯股份有限公司

地址 美国宾夕法尼亚州

(72) 发明人 肖明 帕里克希特·A·德什潘德

曹涵 迈克尔·奥斯汀

坎达斯瓦米·维贾彦

阿列克谢·Y·沙罗诺夫

迈克尔·博伊斯-亚契诺

(74) 专利代理机构 中原信达知识产权代理有限

责任公司 11219

代理人 张颖 樊卫民

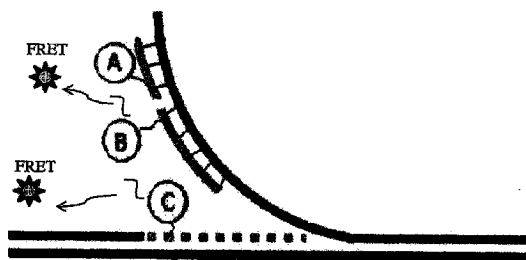
权利要求书 4 页 说明书 18 页 附图 11 页

(54) 发明名称

用于单分子全基因组分析的方法和装置

(57) 摘要

本发明提供了用于单分子基因组分析的方法和装置。在一个实施方案中,方法需要对双链核酸进行处理并对所述核酸进行表征。这些方法可用于例如确定个体之间的结构变异和拷贝数变异。



在掺入或杂交过程中标签剂 A、B 或 C 与被延伸的 DNA 相结合; 通过特异性荧光共振能量转移 (FRET) 信号可以检测共定位事件。

1. 一种对 DNA 进行表征的方法,所述方法包括:
对包含第一 DNA 链和第二 DNA 链的双链 DNA 进行处理,以产生第一 DNA 链的未杂交的悬片段和第二 DNA 链上的相应区域,
未杂交的悬片段包含约 1 到约 1000 个碱基;
沿着第二 DNA 链的相应区域延伸第一 DNA 链;以及
标记未杂交的悬片段的至少一部分、延伸的第一 DNA 链的至少一部分或两者。
2. 权利要求 1 的方法,其中处理包含产生第一 DNA 链的切口。
3. 权利要求 2 的方法,其中产生切口在一个或多个序列特异性位置处实现。
4. 权利要求 2 的方法,其中产生切口在一个或多个非特异性位置处实现。
5. 权利要求 2 的方法,其中产生切口通过将双链 DNA 暴露于产生切口的核酸内切酶、电磁辐射、自由基或其任意组合来实现。
6. 权利要求 1 的方法,其中沿着第二 DNA 链的相应区域延伸第一 DNA 链包括将第一 DNA 链与聚合酶、一种或多种核苷酸、连接酶或其任意组合相接触。
7. 权利要求 1 的方法,其中标记通过下列手段来实现:(a) 使至少一个互补探针与未杂交的悬片段的至少一部分结合,所述探针包含一个或多个标签,(b) 使用一种或多种包含一个或多个标签的核苷酸,沿着第二 DNA 链的相应区域延伸第一 DNA 链,或 (a) 和 (b) 的任意组合。
8. 权利要求 7 的方法,其还包括获得源自于一个或多个标签的序列信息。
9. 权利要求 8 的方法,其中序列信息源自于双链 DNA 上的两个或多个位置。
10. 权利要求 7 的方法,其中标记包括连接标签,所述标签包含荧光团、量子点、树枝状聚合物、纳米线、珠子、肽、蛋白、磁性珠、甲基、甲基转移酶、非切割型限制性酶、锌指蛋白、抗体、转录因子、DNA 结合蛋白、发夹状聚酰胺、形成三链体的寡脱氧核苷酸、肽核酸或其任意组合。
11. 权利要求 7 的方法,其还包括从一个或多个标签检测一种或多种信号。
12. 权利要求 11 的方法,其中信号包含荧光信号、化学发光信号、电磁信号、电信号、电势差、物理尺寸差或其任意组合。
13. 权利要求 11 的方法,其中信号源自于用于延伸第一 DNA 链的碱基上的标签与位于悬片段上的探针上的标签之间的荧光共振能量转移、位于悬片段上的探针上的两个或多个标签之间的荧光共振能量转移,或其任意组合。
14. 权利要求 1 的方法,其中悬片段包含双链 DNA 上目标碱基序列的至少一部分。
15. 权利要求 1 的方法,其中第一 DNA 链的未杂交的悬片段、第二 DNA 链的相应区域或两者,包含位于双链 DNA 上目标碱基序列两侧的双链 DNA 的至少一部分。
16. 权利要求 10 的方法,其还包括将包含未杂交的悬片段的双链 DNA 的至少一部分、第一 DNA 链的延伸区域的至少一部分或两者,至少部分线性化。
17. 权利要求 16 的方法,其还包括测量两个或多个标签之间的距离。
18. 权利要求 17 的方法,其还包括将所述距离与结构、序列组装、遗传或细胞遗传图谱、甲基化模式、CpG 岛的位置、表观基因组模式、生理特征或其任意组合相关联。
19. 权利要求 16 的方法,其中线性化通过将 DNA 限制在通道中来实现。
20. 权利要求 10 的方法,其还包括测量来自至少一个标签的至少一种信号的强度。

21. 权利要求 20 的方法, 其还包括将至少一种信号的强度与结构、序列组装、遗传或细胞遗传图谱、甲基化模式、生理特征、CpG 岛的位置、表观基因组模式或其任意组合相关联。

22. 权利要求 1 的方法, 其还包括从 DNA 获得表观基因组信息。

23. 一种对 DNA 进行表征的方法, 所述方法包括:

在第一双链 DNA 上标记第一 DNA 上的两个或多个序列特异性位置;

在第二双链 DNA 上标记第二 DNA 上的两个或多个相应序列特异性位置;

将第一双链 DNA 的至少一部分线性化;

将第二双链 DNA 的至少一部分线性化; 以及

将线性化的第一双链 DNA 上两个或多个标记物之间的距离与线性化的第二双链 DNA 上相应标记物之间的距离相比较。

24. 权利要求 23 的方法, 其中标记通过在双链 DNA 的第一链上产生切口, 以获得双链 DNA 的 (a) 第一链的未杂交的悬片段和 (b) 第二链中的相应区域来实现。

25. 权利要求 24 的方法, 其还包括将悬片段暴露于与所述探针的至少一部分互补的标记的探针, 使用一种或多种标记的碱基沿着第二链的相应区域延伸 DNA 的第一链, 或两者。

26. 权利要求 23 的方法, 其中标记通过将第一和第二双链 DNA 暴露于非切割型限制性酶、甲基转移酶、锌指蛋白、抗体、转录因子、DNA 结合蛋白、发夹状聚酰胺、形成三链体的寡脱氧核苷酸、肽核酸或其任意组合来实现。

27. 权利要求 26 的方法, 其中非切割型限制性酶包含标签。

28. 权利要求 23 的方法, 其还包括将所述距离与 DNA 的结构、序列组装、遗传或细胞遗传图谱、甲基化模式、生理特征、CpG 岛的位置、表观基因组模式或其任意组合相关联。

29. 权利要求 23 的方法, 其还包括从 DNA 获得表观基因组信息。

30. 一种对 DNA 进行表征的方法, 所述方法包括:

在第一双链 DNA 上标记第一双链 DNA 上的一个或多个序列特异性位置;

在第二双链 DNA 上标记第二双链 DNA 上相应的一个或多个序列特异性位置;

将第一双链 DNA 的至少一部分线性化;

将第二双链 DNA 的至少一部分线性化; 以及

将线性化的第一双链 DNA 的标记物的信号强度与线性化的第二双链 DNA 的标记物的信号强度相比较。

31. 权利要求 30 的方法, 其中标记通过在双链 DNA 的第一链上产生切口, 以获得双链 DNA 的 (a) 第一链的未杂交的悬片段和 (b) 第二链中的相应区域来实现。

32. 权利要求 31 的方法, 其还包括将悬片段暴露于与所述探针的至少一部分互补的标记的探针, 使用一种或多种标记的碱基沿着第二 DNA 链的相应区域延伸 DNA 的第一链, 或两者。

33. 权利要求 30 的方法, 其还包括将所述信号强度与 DNA 的结构、序列组装、遗传或细胞遗传图谱、甲基化模式、生理特征、CpG 岛的位置、表观基因组模式或其任意组合相关联。

34. 权利要求 30 的方法, 其还包括从 DNA 获得表观基因组信息。

35. 一种对大分子进行表征的方法, 所述方法包括:

将包含至少一个从其延伸出的悬片段的大分子, 沿着其中置有至少一个缢缩段的通道移位; 以及

- 检测对应于大分子的至少一个悬片段从通道的至少一个缢缩段通过的至少一种信号。
36. 权利要求 35 的方法,其中悬片段包含标记物。
37. 权利要求 35 的方法,其中大分子包含标记物。
38. 权利要求 35 的方法,其中信号是光信号、电信号或其任意组合。
39. 权利要求 35 的方法,其中大分子包含双链 DNA 或染色质纤维。
40. 权利要求 35 的方法,其还包括将至少一个悬片段移位通过至少一个缢缩段一次以上。
41. 一种对大分子进行表征的方法,所述方法包括:
标记大分子的至少一部分;
将大分子固定化;
将大分子的至少一部分置于通道中以使大分子的至少一部分在通道内被线性化;以及
检测与大分子的标记部分相关的至少一种信号。
42. 权利要求 41 的方法,其中大分子结合在至少一个珠子上,被至少一个珠子固定化的分子被横截面小于珠子的缢缩段捕获。
43. 权利要求 41 的方法,其中固定化通过将大分子化学限定到表面上、通过大分子的磁固定化、通过大分子的光捕获或其任意组合来实现。
44. 权利要求 41 的方法,其中大分子包含至少一种绝缘修饰,并通过光捕获进行固定化。
45. 权利要求 41 的方法,其中定制通道的尺寸以实现大分子的线性化。
46. 一种分析系统,其包含:
含有至少一个宽度在约 1 到约 100 纳米范围内的通道的基材;
所述基材包含至少一个能够相对于周围介质选择性固定化大分子的区域。
47. 权利要求 46 的分析系统,其中通道包含约 10 到约 50 纳米范围内的宽度。
48. 权利要求 46 的分析系统,其中固定化区域包含磁性区域、化学活性区域、缢缩段或其任意组合。
49. 权利要求 46 的分析系统,其还包含能够施加梯度的装置,以便将选择性固定化的大分子的至少一部分置于至少一个通道中。
50. 一种对核酸聚合物进行表征的方法,所述方法包括:
用一种或多种序列特异性基序标记物标记核酸聚合物的一个或多个区域;
将来自一个或多个序列特异性基序标记物的一种或多种信号,与核酸聚合物的一个或多个序列特异性基序标记物的位置相关联;
对核酸聚合物的一个或多个区段进行测序,所述一个或多个区段包括核酸聚合物的一个或多个序列特异性基序标记物;以及
将一个或多个被测序的区段的一种或多种信号与标记的核酸聚合物的一个或多个相应信号进行比较,以便获得两个或多个被测序区段在核酸聚合物中的相对位置。
51. 权利要求 50 的方法,其中标记通过在核酸聚合物中形成悬片段并标记悬片段、由悬片段空出的区域或其任意组合来实现。
52. 权利要求 50 的方法,其中关联包括将核酸聚合物的至少一个被标记部分线性化。
53. 权利要求 52 的方法,其中线性化包括将核酸聚合物的至少一个被标记部分置于能

够使核酸聚合物的至少一个被标记部分线性化的通道中。

54. 权利要求 48 的方法, 其还包括向核酸聚合物的至少一个被标记部分施加梯度, 以便使核酸聚合物的至少一个被标记部分线性化, 将核酸聚合物的至少一个被标记部分维持在线性形式下, 或其任意组合。

55. 权利要求 50 的方法, 其中关联包括确定两个或多个标记物之间的距离, 将从两个或多个标记物产生的信号强度进行比较, 或其任意组合。

56. 权利要求 50 的方法, 其中测序包括 Sanger 测序、Maxam-Gilbert 测序、染料终止物测序、体外克隆扩增、经合成测序、经杂交测序、经连接测序或其任意组合。

57. 权利要求 50 的方法, 其中区段包含约 5 到约 30 个千碱基。

58. 权利要求 47 的方法, 其中比较包括将一个或多个被标记、测序的区段与被标记的核酸聚合物进行比对, 以便将被标记、测序的区段的序列特异性基序标记物与被标记核酸聚合物的相应的序列特异性基序标记物配准。

用于单分子全基因组分析的方法和装置

[0001] 相关申请

[0002] 本申请要求 2008 年 6 月 30 日提交的美国专利申请 No. 61/076, 785 的优先权益, 其全部内容在此引为参考。

技术领域

[0003] 本发明涉及纳米流体领域和 DNA 测序领域。

[0004] 发明背景

[0005] 大分子是由许多化学单元彼此相连构成的长的聚合物链。多核苷酸是一类大分子, 其包括例如 DNA 和 RNA。多核苷酸由核苷酸的长序列构成。

[0006] 核苷酸的序列与生物体的基因组和后基因组的基因表达信息直接相关。对序列区域、基序和功能单元例如开放阅读框 (ORF)、非翻译区 (UTR)、外显子、内含子、蛋白因子结合位点、表观基因组位点例如 CpG 簇、microRNA 位点、小干扰 RNA (SiRNA) 位点、大型干涉性非编码 RNA (lincRNA) 位点和其他功能单元的直接测序和作图, 在评估个体的基因组组成中都是重要的。

[0007] 在许多情况下, 这些核苷酸序列在个体生命期间的复杂重排例如插入、缺失、倒位和移位, 导致了疾病状态例如遗传异常或恶性细胞。在其他情况下, 在个体之间以拷贝数变异 (CNV) 形式表现的序列差异反映出群体遗传组成的多样性及其对环境刺激和信号例如药物治疗的差异响应性。在其他情况下, 例如 DNA 甲基化、组蛋白修饰、染色质折叠这样的过程或其他改变 DNA 或 DNA-蛋白质相互作用的变化, 影响基因的调控、表达以及最终的细胞功能, 导致疾病和癌症。

[0008] 已经发现, 即使在健康个体之间, 基因组结构变异 (SV) 也比以前所想象的要广泛得多。因此, 对于人类健康和常见遗传疾病来说, 理解具有结构变异信息的基因组序列的重要性变得越来越明显。

[0009] 功能单元和共有结构变异被认为涵盖了从数十碱基到兆碱基以上。因此, 在多个个体的测序和精细规模作图计划中, 在沿着大的天然基因组分子上跨越从千碱基以下到兆碱基的分辨率尺度的直接、廉价且灵活的揭示序列信息和 SV 的方法, 是极为需要的, 以便对以前未表征的基因组特点进行编目。

[0010] 此外, 生物系统、特别是在多倍体生物例如人类中的表型多态性或疾病状态, 是从母系和父系遗传而来的两个单倍体基因组之间相互影响的结果。具体来说, 癌症通常是二倍体染色体病变中杂合性丧失的结果。

[0011] 常规的细胞遗传学方法例如核型分析、FISH (荧光原位杂交) 提供了少至单个细胞中基因组组成的全局视图, 它们能够有效揭示基因组的大体变化, 例如数千和数百万碱基对的大片段的非整倍性、获得、丧失或重排。但是, 这些方法的缺点在于在检测中等到小型序列基序或病变中相对低的灵敏度和分辨率。这些方法也是繁琐的, 限制了速度和不一致性。

[0012] 更近些的用于检测目标序列区域、序列基序和 SV 的方法, 例如 aCGH (阵列比较基

基因组杂交)、fiberFISH 或大规模双末端测序,已经在分辨率和通量方面有所改进。但是,这些方法不直接、繁琐、昂贵并且依赖于现有的参比数据库。此外,这些方法可能具有有限的固定分辨率,并提供依赖于返回到参比基因组作图进行重新组装所推断出的位置信息或比较强度比信息。因此,这样的方法不能揭示平衡的病变事件例如倒位或移位。

[0013] 目前的测序分析方法受到可用技术的限制,并主要基于源自于具有非常有限的单倍型信息的平均多倍性基因组材料的样品。目前用于从异源细胞群体提取混合的二倍体基因组材料的前端样品制备方法,有效地将材料碎裂成较小碎片,导致二倍体基因组天然的决定性重要结构信息的破坏。

[0014] 即使是更近些时候开发的第二代方法,尽管具有改进的通量,但是由于短得多的测序读取结果的组装更加困难,使复杂的基因组信息的草图绘制更加复杂化。

[0015] 总体来说,短的读取结果更难以单值地排列在复杂的基因组中,并且需要附加序列信息来解译短的靶区域的线性顺序。

[0016] 为了达到类似的组装置信度,需要将测序覆盖率提高到 25 倍的级别,而不是在常规 BAC 和所谓的鸟枪法 Sanger 测序中的 8-10 倍的覆盖率 (Wendl MC, Wilson RK, 医学 DNA 测序中的覆盖率情况 (Aspects of coverage in medical DNA 测序), BMC Bioinformatics, 16 May 2008 ;9 :239)。这种多倍的测序覆盖率必然需要高成本,有效上阻止了在本技术领域中将测序费用降低到 1000 美元的标杆以下的最高目标。

[0017] 因此,大的完整的基因组分子的单分子水平分析,通过不使用克隆过程或扩增而对序列基序进行原位精细作图,提供了保留精确的天然基因组结构的可能性。基因组片段越大,基因组样品中样品群体的复杂度越低,例如,在理想的情形下,为了覆盖整个正常二倍体人类基因组,在单分子水平上只需要分析 46 个染色体长度的片段,并且从这种方法产生的序列天然具有完整的单倍型信息。此外,可以从细胞提取兆碱基规模的基因组片段并将其保留用于直接分析,这极大地降低了复杂的算法和组装的负担,同时也将原始背景下的基因组和 / 或表观基因组信息与个体的细胞表型更直接地相关联。

[0018] 除了基因组学之外,在过去 20 年左右,由于在人类疾病例如癌症中的作用,表观基因组学领域突出的重要性已经越来越多地被认识到。随着基因组学和表观基因组学两个领域知识的积累,一个主要的挑战是理解基因组和表观基因组因素如何直接或间接与人类疾病和恶性肿瘤中多态性或病理生理状况的发生相关。全基因组分析的概念已经从划分在几个主要是独立进行研究的基因组测序、表观基因组甲基化分析和功能基因组学领域中的方法,向越来越多方面的整体方法演化。DNA 测序、结构变异作图、CpG 岛甲基化模式、组蛋白修饰、核小体改构、microRNA 功能和转录图谱已经被越来越多地以系统性方式更紧密地看待,但是,检测细胞分子状态的每种上述方面的技术通常是独立的、繁琐的和不相容的,严重阻碍了具有相关联的实验数据结果的整体分析。

[0019] 因此,在本技术领域,对于能够对大的、完整的天然生物样品进行单分子水平分析、以便能够确定靶样品的基因组和表观基因组信息的方法和装置,存在着需求。这样的方法和装置将为研究人员和临床医生等提供非常强有力的工具。

[0020] 发明简述

[0021] 在应对所述挑战中,要求保护的本发明首先提供了对 DNA 进行表征的方法,所述方法包含:对包含第一 DNA 链和第二 DNA 链的双链 DNA 进行处理,以产生第一 DNA 链的未杂

交的悬片段 (flap) 和第二 DNA 链上的相应区域,未杂交的悬片段包含约 1 到约 1000 个碱基;沿着第二 DNA 链的相应区域延伸第一 DNA 链;以及标记未杂交的悬片段的至少一部分、延伸的第一 DNA 链的一部分或两者。

[0022] 还提供了鉴定 DNA 之间的结构变异的方法,所述方法包含:在第一双链 DNA 上标记两个或多个第一 DNA 上的序列特异性位置;在第二双链 DNA 上标记两个或多个第二 DNA 上的相应序列特异性位置;将第一双链 DNA 的至少一部分线性化;将第二双链 DNA 的至少一部分线性化;以及比较线性化的第一双链 DNA 上两个或多个标记物之间的距离与线性化的第二双链 DNA 上相应标记物之间的距离。

[0023] 还公开了从 DNA 获得结构信息的方法,所述方法包含:在第一双链 DNA 上标记一个或多个第一 DNA 上的序列特异性位置;在第二双链 DNA 上标记第二双链 DNA 上相应的一个或多个序列特异性位置;将第一双链 DNA 的至少一部分线性化;将第二双链 DNA 的至少一部分线性化;以及比较线性化的第一双链 DNA 的至少一个标记物的信号强度与线性化的第二双链 DNA 的至少一个标记物的信号强度。

[0024] 还提供了从大分子获得结构信息的方法,所述方法包含:将包含至少一个从其伸出的悬片段的大分子,沿着其中置有至少一个缢缩段 (constriction) 的通道移位;以及检测对应于大分子的至少一个悬片段从通道的至少一个缢缩段通过的至少一种信号。

[0025] 还提供了从大分子获得结构信息的方法,所述方法包含:标记大分子的至少一部分;将大分子固定化;将大分子的至少一部分置于通道中以使大分子的至少一部分在通道内线性化;以及检测与大分子的标记部分相关的至少一种信号。

[0026] 还公开了分析系统,其包含:包含至少一个宽度在约 1 到约 100 纳米范围内的通道的基材;所述基材包含至少一个固定化区域。

[0027] 还提供了对核酸聚合物进行表征的方法,所述方法包含:用一种或多种序列特异性基序标记物标记核酸聚合物的一个或多个区域;将来自一个或多个序列特异性基序标记物的一种或多种信号,与核酸聚合物的一个或多个序列特异性基序标记物的位置相关联;对核酸聚合物的一个或多个区段进行测序,所述一个或多个区段包括核酸聚合物的一个或多个序列特异性基序标记物;以及将一个或多个被测序的区段的一种或多种信号与标记的核酸聚合物的一个或多个相应信号进行比较,以便获得两个或多个被测序区段在核酸聚合物中的相对位置。

附图说明

[0028] 当结合随附的图阅读时,发明简述以及下面的详细描述得到进一步理解。出于说明本发明的目的,在图中显示了本发明的示例性实施方案;但是,本发明不限于所公开的具体方法、组合物和装置。此外,图不是必须按比例绘制的。在图中:

[0029] 图 1 描绘了本发明的悬片段标记方法的示意图;

[0030] 图 2 描绘了与从第一 DNA 链产生的悬片段杂交的标记探针和驻留在第一链对应于悬片段的区域中的标记物;

[0031] 图 3 描绘了将 DNA “条形码”放置在聚核酸上的可选实施方案;

[0032] 图 4 描绘了沿着基因组区域测序;

[0033] 图 5 描绘了同时进行的平行测序和空间组装;

- [0034] 图 6 描绘了从核酸聚合物获得基因组组装信息；
- [0035] 图 7 是经历过图像分析的被标记 DNA 聚合物的软件图像；
- [0036] 图 8 描绘了本发明的光学和非光学检测方案；
- [0037] 图 9 描绘了在纳米通道或纳米轨道中线性化的被标记核酸聚合物；
- [0038] 图 10 描绘了通过各种不同方法固定化在纳米通道附近或其中的核酸聚合物；以及
- [0039] 图 11 描绘了置于纳米通道或纳米轨道中的核酸聚合物的磁和光捕获。

[0040] 说明性实施方式的详细描述

[0041] 通过与形成了本公开的一部分的随附的图和实施例相结合，参考下面的详细描述，可以更容易地理解本发明。应该理解，本发明不限于本文中描述和 / 或显示的具体装置、方法、应用、条件或参数，并且在本文中使用的术语时出于描述仅作为实例的具体实施方案的目的，而不打算对本发明进行限制。此外，当在说明书包括随附的权利要求书中使用时，除非另有指明，否则单数形式的指称包括复数，并且对具体数值的指称至少包括该具体数值。本文中使用的术语“多个”是指一个以上。当表述值的范围时，另一个实施方案包括了从一个具体值和 / 或到另一个具体值。类似地，当通过使用先行词“约”将值表示成近似值时，应该理解具体的值形成了另一个实施方案。所有范围是包含性的和可组合的。

[0042] 应该理解，为了清楚起见，本发明的某些特征在本文中描述在分开的实施方案的上下文中，它们也可以组合提供在单一实施方案中。相反，出于简便而描述在单一实施方案的上下文中本发明的各种不同特征，也可以独立或以任何子组合形式提供。此外，以范围形式陈述的对值的指称包括该范围内的每个和所有值。

[0043] 在第一种情况下，本发明提供了对 DNA 进行表征的方法，其包含对包含第一 DNA 链和第二 DNA 链的双链 DNA 进行处理，以产生第一 DNA 链的未杂交的悬片段和第二 DNA 链上的相应区域，未杂交的悬片段包含约 1 到约 1000 个碱基；沿着第二 DNA 链的相应区域延伸第一 DNA 链；以及标记未杂交的悬片段的至少一部分、延伸的第一 DNA 链的一部分或两者。

[0044] 适合情况下，悬片段长度为约 1 到约 1000 碱基。适合情况下，悬片段长度为约 20 到约 100 碱基，或甚至在约 30 到约 50 碱基的范围内。

[0045] 方法还包含将一个或多个置换碱基掺入到双链 DNA 的第一链中，以便延伸第一 DNA 链（悬片段从其揭下）以填补并消除由形成悬片段所留下的间隙（即目前第二 DNA 链的相应区域）。用户可以用一个或多个标签标记被处理的双链 DNA 的至少一部分（第一 DNA 链、第二 DNA 链、悬片段或其任意组合）。悬片段留下的填补空隙可以包括一个或多个被标记的部分。在某些实施方案中（未显示），可以使用悬片段移除酶切掉悬片段，留下掺入有一个或多个核苷酸的 dsDNA。

[0046] 适合情况下，处理通过在双链 DNA 的第一链上产生切口来完成。适合情况下，这种切口形成在一个或多个序列特异性位置处，尽管切口也可以在一个或多个非特异性位置、包括随机或非特异性位置处形成。

[0047] 适合情况下，切口形成通过将双链 DNA 聚合物暴露于切口核酸内切酶或切口酶来实现。适合情况下，切口酶具有高度序列特异性，意味着它们以高度特异性与特定碱基序列（基序）结合。切口酶可以从例如 New England BioLabs (www.neb.com) 获得。

[0048] 切口形成也可以通过其他在 DNA 链中执行断裂或切开的酶来完成。这样的断裂或

切口也可以通过暴露于电磁辐射（例如 UV 光）、一种或多种自由基等来实现。切口可以通过一种或多种这样的技术来实现。

[0049] 适合情况下，将置换碱基掺入双链 DNA 的第一链（即产生切口的链）包含将 DNA 与聚合酶、一种或多种核苷酸、连接酶或其任意组合相接触。其他用于代替悬片段中存在的“被剥离”碱基的方法，对于本技术领域的专业人员来说也是已知的。适合情况下，将第一 DNA 链沿着第二 DNA 的相应区域延伸，所述区域是通过形成悬片段而留下 / 暴露出的。在某些实施方案中，聚合酶与产生悬片段的切口酶同时起作用。

[0050] 这些置换碱基的掺入在概念上可以理解为填补由悬片段的形成和“剥离”所留下的间隙。通过填补间隙，以前由悬片段占据的位置被一组适合情况下与位于悬片段上的碱基具有相同序列的碱基占据。填充可以防止悬片段与悬片段以前所结合的 DNA 的第二链的重新杂交。

[0051] 适合情况下，标记可以通过 (a) 将至少一个互补探针与悬片段的至少一部分结合，所述探针包含一个或多个标签，(b) 利用包含一个或多个标签的核苷酸作为沿着第二 DNA 链的相应区域延伸的第一 DNA 链的一部分的置换碱基，或 (a) 和 (b) 的任意组合来实现。通过这种方式，可以对悬片段、填补空隙的碱基或两者进行标记。

[0052] 适合情况下，探针是包含本文别处所述的标签的核酸（单个或多个）。探针可以是序列特异性的（例如 AGGCTA，或某些其他特定碱基序列），尽管探针也可以随机产生。正如本文别处描述的，可以根据用户的意愿选择或构建探针，以使探针与目标序列结合，或在可选方案中，与特定 DNA 聚合物上目标序列的上游或下游序列或其他区域结合（即探针结合到目标区域的两侧或悬臂上）。探针可以与悬片段一样长（即长达 1000 碱基）。适合情况下，探针长度在 1 到约 100 碱基、或约 3 到 50 碱基的范围内，或长度甚至在约 5 到约 20 碱基的范围内。

[0053] 这些方法的示意图显示在图 1 中。在该图中，显示了悬片段的产生和得到的间隙的回填。回填可以使用所谓的“热”或标记碱基，并且悬片段可以与一个或多个与悬片段的至少一部分互补的探针相接触。序列特异性的切口核酸内切酶或切口酶，在双链 DNA 上产生单链切开间隙，聚合酶与切口位点结合并开始链延伸，并在同时产生了被置换的链或所谓的“剥离悬片段”。然后剥离悬片段产生了可用于测序与被标记探针序列特异性杂交以产生可检测和可分辨信号的区域（即核酸聚合物的第二 DNA 链上未杂交的相应区域）。

[0054] 图 1b 显示了在纳米通道中线性展开的标记的大基因组 DNA。正如在图的底部所示，荧光标记的悬片段能够使用户观察到较大的大分子背景中探针的位置。正如所示，带切口的标记大分子可以在纳米通道内线性化。来自标签的信号之间的空间距离是一致的，因此可以被定量，这反过来提供了独特的“条形码”标志样式，其反映出关于所分析区域的特异性基因组序列信息。作为实例，显示了由特定酶、包括但不限于 Nb. BbvCI、Nb. BsmI、Nb. BsrDI、Nb. BtsI、Nt. AlwI、Nt. BbvCI、Nt. BspQI、Nt. BstNBI、Nt. CviPII 以及任何上述酶的组组合消化产生的 λ dsDNA（总长度 48.5kbp）的多个切口位点。

[0055] 包含了线性化的单个 λ DNA 的图像，以显示与预期的切口酶产生位置杂交的荧光标记的寡核苷酸探针。这种沿着长的生物聚合物记录到的实际条形码，在本文中的其他地方被描述为观察到的条形码。

[0056] 通过将具有标记的悬片段、标记的间隙或两者的大分子线性化，用户能够确定标

记物彼此的相对位置。正如在本文别处所描述的,这样的相对距离信息可用于诊断应用和核酸聚合物的表征。

[0057] 在某些实施方案中,方法还包括获得从掺入到双链 DNA 的第一 DNA 链中的一个或多个置换碱基、从一个或多个与悬片段结合的探针、或从两者衍生的序列信息。该序列信息可以以各种不同方式获得。

[0058] 在一个实例中,将与特定碱基序列互补的被标记探针导入悬片段,然后用户确定该序列特异性探针是否与悬片段结合。这一过程可以使用具有不同序列特异性的探针重复几次,最终使用户能够确定存在于悬片段中的碱基序列。

[0059] 在另一个实例中,通过确定填补到悬片段留下的间隙中的碱基序列来获得序列信息。这可以通过用相同或不同的标记物标记一个或多个碱基,并测定当碱基被掺入到间隙中时或它们掺入到间隙中后所发出的信号。在其他实施方案中,用户可以监测由将碱基掺入到间隙中的聚合酶所产生的一种或多种信号,以确定碱基序列。

[0060] 序列信息的测定可以在自由溶液中进行,或者可以在纳米通道中进行,以便允许对单个 DNA 聚合物进行高分辨率分析。也可以将悬片段用适合的酶切下,然后也可以对切下的悬片段本身进行测序。

[0061] 序列信息可以从单一悬片段、单一间隙或两者获得。但是,在某些实施方案中,序列信息从两个或多个悬片段或间隙获得,从而能够对给定的靶进行更快测序。也可以通过使用序列特异性探针并确定这些探针与核酸聚合物的一部分结合的位置(以及是否结合),来确定序列信息。

[0062] 图 4 描绘了沿着相对长的基因组区域的测序。在该图中,在将“亲本”核酸聚合物用序列特异性切口内切核酸酶消化并在聚合物的第一链中进行聚合酶延伸后,产生了单链悬片段。该结构可以再次用切口内切核酸酶和切开悬片段与第一链的连接处(用箭头显示)的悬片段内切核酸酶消化,并且可以将得到的 dsDNA 在适合条件下变性,以产生跨越切口位点的单链间隙和悬片段内切核酸酶切割位点。然后可以将该间隙暴露于使用聚合酶延伸或使用特异性探针和酶的杂交和连接的测序反应。

[0063] 图 4b 描绘了示意显示了沿着长的 dsDNA 产生的多个切口位点、单链悬片段位点以及单链间隙位点。然后从一个或多个切口、悬片段序列位点或单链间隙位点处开始测序反应,其中测序通过聚合酶延伸或利用杂交或连接的测序来执行。

[0064] 各种不同物质可用作本发明方法的标签。标签可以是例如荧光团、量子点、树枝状聚合物、纳米线、珠子、肽、蛋白、磁性珠、甲基、甲基转移酶、非切割型限制性酶、锌指蛋白、抗体、转录因子、DNA 结合蛋白、发夹状聚酰胺、形成三链体的寡脱氧核苷酸、肽核酸等。方法可以包括使用两种或多种不同标签,因此单一分子可以包含多种标签。

[0065] 方法还包括从一个或多个标签检测一种或多种信号。这样的信号可以包括荧光信号、化学发光信号、电磁信号、电信号、电势差等。信号可以与两个物体之间的物理尺寸差异相关,其可以是例如当与 DNA 靶相连的珠子被捕获在横截面比珠子更小的缢缩段中时所产生的信号。荧光信号被认为是特别适合的,特别是在荧光分子与碱基、探针或两者相连的实施方案中。

[0066] 在某些实施方案中,信号可以源自于置换碱基上的标签与位于悬片段上的探针上的标签之间的转移能量(例如荧光能量转移“FRET”)、位于悬片段上的探针上的两个或多

个标签之间的荧光共振能量转移,或其任意组合。

[0067] 图 2 显示了按照本发明制备的核酸聚合物上的标记物和探针的示例性位置。该图描绘了位于悬片段上的探针(显示为 A 和 B),以及沿着 DNA 链延伸以便填补由悬片段的形成和剥离所留下的间隙的探针(显示为 C)。

[0068] 探针包括例如有机荧光团、量子点、树枝状聚合物、纳米线、珠子、Au 珠子、顺磁性珠子、磁性珠子、聚苯乙烯珠子、聚乙烯珠子、肽、蛋白、半抗原、抗体、链亲和素、亲和素、中性亲和素、生物素、核苷酸、寡核苷酸、序列特异性结合因子例如工程化的限制性酶、甲基转移酶、锌指结合蛋白等。正如所示,一个以上的探针可以位于悬片段上。在示例性实施方案中,间隙中的标签(或多个标签)由激发辐射激发。然后被激发的间隙-标签将能量转移到本身位于悬片段上的探针上所带的标签。

[0069] 间隙标签和悬片段标签中的一个或两个可以发出可以被用户检测的信号。在某些实施方案中,间隙标签、第一悬片段标签或两者可以激发第二悬片段标签。通过这种方式,用户可以通过选择只被特定波长或类型辐射所激发的标签构建高特异性检测系统,由此产生只有当一个或多个前体标签位于适合位置中时用户所检测的标签才被激发的系统。因此,可以通过两种或多种标记物之间的能量转移检测(例如可视化)共定位事件,这增加了结合事件测定的特异性。

[0070] 在某些情况下选择悬片段区,因为悬片段、间隙或两者包含双链 DNA 上的目标特定序列的至少一部分。这样的目标序列可以包括例如已知编码特定蛋白或特定情况的序列。

[0071] 在某些实施方案中,悬片段、间隙或两者包含位于双链 DNA 上的目标序列两侧的双链 DNA 的至少一部分。这在例如用户试图标记 DNA 上囊括目标特定基因或其他区域的区域以便突出该区域时是有用的。

[0072] 本发明的方法还包括将包含至少一个悬片段、一个间隙或两者的双链 DNA 的至少一部分至少部分线性化(例如解开)。用户也可以将包含至少两个悬片段、两个间隙或其任意组合的双链 DNA 的至少一部分至少部分线性化。这样的线性化可以通过例如将 DNA 移位通过通道或其他结构来实现,所述结构的维度使得利用在通道或其他结构中的物理限制使 DNA 线性化。

[0073] 在某些实施方案中,用户也可以测量两个悬片段之间、位于两个或多个悬片段附近的两个或多个标签之间、位于两个或多个间隙中的两个或多个标签之间的距离,或其任意组合。适合情况下,然后将该距离与 DNA 的结构、序列组装、遗传或细胞遗传图谱、甲基化模式、CpG 岛的位置、表观基因组模式、生理特征或其任意组合相关联。因为本发明能够研究结构和其他表观基因组因子(例如甲基化模式、CpG 岛的位置等),因此用户能够将结构与表观基因组模式相关的结果叠加,以获得完整的基因组图。

[0074] 本发明的一个特点是其提供关于核酸或其他遗传物质的基因组(序列)和表观基因组(supra-序列)信息的能力。更具体来说,本发明允许用户通过测序来确定特定基因是否存在,以及通过获得表观基因组信息确定该基因的活性。

[0075] 在一个非限制性实例中,用户可以获得关于核酸聚合物的基因组信息(通过在本文别处描述的标记方法),例如特定基因是否存在。然后用户也可以通过使用例如标记的甲基结合蛋白以鉴定沿着核酸聚合物的甲基的位置,来获得关于核酸聚合物的甲基化模式的

信息（其表明了位于甲基化附近的基因座的活性）。这样的甲基可以存在于胞嘧啶上和可能与功能性基因座的调控相关的所谓的 cpG 岛簇上。其他结合分子（例如与转录因子结合位点结合的分子等）也适合用于获得表观基因组信息。

[0076] 因此，在某些实施方案中，用户可以同时确定一个或多个功能性基因的存在，并通过基于甲基的表观基因组信息确定这些基因是否有活性。在一个实例中，用户可以用第一种颜色的标记物标记基因的序列信息并用第二种颜色的标记物标记甲基化区域，从而能够同时观察基因位置 / 序列和基因活性（即甲基化模式）。表观基因组信息也可以包括转录酶能够或不能结合的位置。

[0077] 表观基因组信息的用途是显而易见的。正如在本文别处描述的，基因组信息的用途在于基于寡聚物的探针（或包含条形码的探针组）提供了关于所研究的核酸聚合物序列的“静态”信息。表观基因组信息（例如关于甲基化或转录因子结合的信息）提供了关于基因序列的动态信息，有效地提供了关于基因的开 / 关信息。因此，本发明能够同时收集基因组和表观基因组信息两者。

[0078] 作为一个说明性的非限制性的实例，用户可以标记来自第一个患者的 DNA 上的位置（即悬片段、填补间隙或二者的某些组合），位置的选择使得它们位于 DNA 上特定基因例如乳腺癌基因的位置的上游和下游（即两侧）。在将标记的 DNA 线性化后，用户可以将这些标记物之间的距离与来自自己已知具有“适合”的乳腺癌基因拷贝数的对照对象的 DNA 上相应标记物之间的距离进行比较。如果第一个患者的标记物之间的距离大于对照对象的标记物之间的距离，那么可以知道患者具有附加的或额外的乳腺癌基因拷贝，因此可以设计治疗方案。

[0079] 该技术也可用于确定两个或多个都不是“对照”的个体之间的拷贝数变异，或甚至单一患者中的拷贝数变异（即通过比较在两个不同时间从患者获取的 DNA）。通过这种方式，本发明的方法便于对来自单一对象或来自较大群体区段的 DNA 或其他大分子进行快速分析和表征。

[0080] 用户也可以测量来自位于悬片段附近的至少一个标签、位于间隙中的标签或两者的至少一种信号的强度。然后用户可以将至少一种信号的强度与序列组装、遗传或细胞遗传图谱、生理特性或本文别处描述的其他特征（例如表观基因组模式）相关联。这使用户能够得出核酸聚合物来源的病理生理状态的整个图谱。

[0081] 这由非限制性的图 5c 所示。该图示意显示了使用标记的结合因子 (BF) 例如抗甲基抗体或甲基结合蛋白 (MBP) 沿着基因组区域定位一个或多个表观基因组位点，以产生表观基因组条形码模式。正如所示，在某些情况下，用户同时也使用本公开的方法绘制相同区域的“条形码”（使用例如序列特异性探针）以确定基因组区域的结构。基因组探针可以在与表观基因组分析所涉及的任何标记物不同的波长处发射或激发，或具有可以与所述标记物辨别开的信号。在一个实施方案中，表观基因组条形码包括（但不限于）从转录因子结合位点或 siRNA 或 lincRNA 结合位点衍生的模式。这证实了本发明在单分子水平上，使用直接成像在同一个视野中同时地、实时地将静态基因组序列和结构信息与动态的调控和功能信息相关联的能力。

[0082] 作为另一个非限制性的实例，用户可以标记对应于来自第一个患者的 DNA 的目标基因（例如乳腺癌基因）内的区域的一个或多个悬片段（或填补间隙）。然后用户测量从

这些标记物产生的一种或多种信号的强度。然后用户测量来自“对照”或第二个对象的 DNA 上的相应标记物产生的一种或多种信号的强度。如果来自第一个患者的信号的强度与来自对照的信号强度不同,用户将获得两个对象具有不同基因拷贝数的一些指示。强度信号也可以与在给定聚合物中单一碱基或特定碱基序列的普遍性相关联。信号的强度也可以提供关于与带有发射信号的标记物的探针互补的序列的空间密度的信息。

[0083] 图 7 显示了在本发明的核酸聚合物上进行的图像分析。具体来说,图显示了捕获的“原始”DNA 图像,其中端到端的等高线和强度信息被实时提取和测量。显示了尺寸分布的直方图以便证实得自异源 DNA 混合物的读取结果。

[0084] 本发明还提供了对多个 DNA 进行表征的方法。这些方法包括在第一双链 DNA 上标记第一 DNA 上的两个或多个位置(序列特异性、随机的或两者);在第二双链 DNA 上标记第二 DNA 上的两个或多个相应序列特异性位置;将第一双链 DNA 的至少一部分线性化;将第二双链 DNA 的至少一部分线性化;以及比较线性化的第一双链 DNA 上两个或多个标记物之间的距离与线性化的第二双链 DNA 上相应标记物之间的距离。

[0085] 在某些实施方案中,标记如本文别处所述,通过在双链 DNA 的第一链上产生切口,以获得 (a) 与双链 DNA 分离的第一链的悬片段和 (b) 由切口位点和悬片段与双链 DNA 第一链的相接位点所定义的在双链 DNA 的第一链中的间隙来实现。

[0086] 方法还可以包括将悬片段暴露于与探针的至少一部分互补的标记的探针,将一个或多个标记的碱基插入到间隙中,或两者。正如别处所述,适合情况下,标记通过将第一和第二双链 DNA 暴露于非切割型限制性酶、甲基转移酶、锌指蛋白、抗体、转录因子、DNA 结合蛋白、发夹状聚酰胺、形成三链体的寡脱氧核苷酸、肽核酸等来实现。非切割型限制性酶可以包含标签。如本文别处所述,在适合情况下,然后将距离与 DNA 的结构、序列组装、遗传或细胞遗传图谱、甲基化模式、生理特征、CpG 岛的位置、表观基因组模式或其任意组合相关联。

[0087] 这些方法的一个实施方案显示在图 5 中,其是示意图,显示了平行测序和同时的空间组装。沿着长的基因组区域可以序列基序特异性方式产生许多序列起始位点,在这种情况下是 GCTGAxxxx,并检测这些位点的物理位置并在物理图谱上登记。随后通过使用聚合酶延伸的测序或使用特异性探针杂交和连接的测序,通过测序化学方法记录读取结果。

[0088] 除了测序读取结果之外,在物理图谱上同时记录和组装这些多个测序读取结果的相应线性次序和空间距离与位置。这样的基于作图的测序方案与现有方法相比最终提供了更好的组装准确性、效率和成本降低。

[0089] 图 5b 是示意图,显示了 DNA 结合因子 (BF) 的使用,包括遗传工程化的非功能性限制性酶,其仅仅保留了限制性酶的结合结构域但缺乏这些酶的 DNA 切割功能。以序列特异性方式识别并结合 DNA 的 DNA 甲基转移酶也是有用的,其他酶也是同样,例如锌指蛋白、以序列基序特异性方式结合 DNA 的转录因子、与甲基化特异性位点结合的甲基结合蛋白或抗甲基抗体、其他特异性(次级结合)方式的 DNA 结合因子。例如, DNA 甲基转移酶 (MTase) 包括但不限于 M. BseCI (使 5' -ATCGAT-3' 序列中 N6 位的腺嘌呤甲基化)、M. TaqI (使 5' -TCGA-3' 序列中 N6 位的腺嘌呤甲基化) 和 M. HhaI (使 5' -GCGC-3' 序列中 C5 位的第一个半胱氨酸甲基化)。

[0090] 总的来说,该适合地结合的物体的名单包括(以例如序列特异性方式)与双链 DNA

结合并且不切开所述 dsDNA 的物体。在图中,各种不同的星表示不同标记标签,例如 QD(量子点)、荧光标记物等。这些标签之间的空间距离以及这些“串成线的点”条形码图形的强度,可用于研究其他生物功能例如活性转录位点、ORF(开放阅读框)、低和高甲基化位点等。

[0091] 另一方面,本发明提供了从 DNA 获得结构信息的方法。这些方法包括在第一双链 DNA 上标记第一 DNA 上的一个或多个序列特异性位置。方法还包括在第二双链 DNA 上标记第二双链 DNA 上相应的一个或多个序列特异性位置;使第一双链 DNA 的至少一部分线性化,使第二双链 DNA 的至少一部分线性化;以及比较线性化的第一双链 DNA 的至少一个标记物的信号强度与线性化的第二双链 DNA 的至少一个标记物的信号强度。

[0092] 正如本文别处所述,适合情况下,标记通过在双链 DNA 的第一链上产生切口,以获得 (a) 与双链 DNA 分离的第一链的悬片段和 (b) 双链 DNA 的第一链中对应于悬片段的间隙来实现,所述间隙由切口位点和悬片段与双链 DNA 第一链的相接位点所定义。适合情况下,将悬片段暴露于与探针的至少一部分互补的标记的探针,将一个或多个标记的碱基插入到间隙中,或两者,以便沿着第二 DNA 链的相应区域延伸第一链。然后将信号强度与核酸聚合物的供体的至少一种生理特征相关联。也可以将强度与核酸聚合物的结构特征、表观基因组模式或两者相关联。

[0093] 本发明为用户提供了从给定聚合物获得和分析结构和表观基因组信息两者的能力。正如在本文别处所述,本发明提供了“条形码”技术,通过其为核酸聚合物区域提供独特的签名。可以应用该条形码(如本文别处所述)以便提供关于结构(利用例如具有序列特异性基序的标记物,第一条形码)和表观基因组模式(利用对表观基因组指示物例如甲基化位点、CpG 岛等具有特异性的标记物,第二条形码)的信息。利用从第一和第二条形码收集的信息,用户可以获得关于给定核酸聚合物的结构和表观基因组信息。

[0094] 还提供了从大分子例如双链 DNA 获得结构信息的方法。这些方法包括将包含至少一个从其伸出的悬片段的大分子,沿着其中置有至少一个缢缩段的通道移位;以及检测对应于大分子的至少一个悬片段从通道的至少一个缢缩段通过的至少一种信号。在某些实施方案中,悬片段被标记,在其他实施方案中,悬片段未标记,并且信号与“裸露”悬片段通过缢缩段相关。

[0095] 适合的通道在本技术领域是已知的,例如在美国专利申请 10/484,293 中描述的通道,该专利申请在此以其全文引为参考。在某些实施方案中,悬片段或大分子与悬片段相邻的区域包含标记物。在某些实施方案中,正如本文别处所述,标记物置于当悬片段形成时留下的填补间隙中。

[0096] 适合情况下,信号是光信号、电信号、电磁信号或甚至其某些组合。信号可以与悬片段通过缢缩段相关,或可以与标记物通过缢缩段相关。悬片段可以被移位通过缢缩段一次以上。

[0097] 这些方法的示例性非限制性实施方案显示在图 8 中。第一个图(图 8a)描绘了利用光学和非光学检测方法从核酸聚合物获得标记的条形码信息的系统。

[0098] 正如所示,显示了在具有一个或多个狭窄的溢缩点(被称为纳米门控或纳米喷嘴;参见美国专利申请 12/374,141,其全部内容在此引为参考)的纳米通道中拉伸和线性化的被标记的长核酸分子。

[0099] 在某些实施方案中, DNA 移动和电流测量由与流体装置和外部储液器相连的电路控制。条形码模式的光学图像和标记物的非光学记录(即沿着均匀的聚合物的物理“隆起”的电记录), 示意显示在图 8b 和图 8c 中。光学和非光学结果可以彼此关联或对比, 以获得更好的数据准确性。

[0100] 图 8d 描绘了包含纳米门控的流体装置。这里显示了一系列通过前面描述的方法产生的“悬片段”, 所述悬片段可以包括附加的标记标签。悬片段、它们的标签或两者, 在通过纳米门控的过程中被直接检测, 在此过程中悬片段、标签或两者产生可检测的电子信号, 例如反映靶基因组区域的离子电流特征。正如本文别处所述, 在核酸聚合物中由悬片段空出的区域中也可以存在标记的碱基。这样的碱基当通过纳米门控时也可以被检测。

[0101] 还提供了从大分子获得结构信息的方法。这些方法包括标记大分子的至少一部分; 将大分子固定化; 将大分子的至少一部分置于通道中以使大分子的至少一部分在通道内线性化; 以及检测与大分子的被标记部分相关的至少一种信号。

[0102] 图 9 描绘了限定 (tethered) 在用于序列成像分析的基质表面上的纳米通道或纳米轨道内的一个末端或两个末端处的核酸。正如图中所示, 对核酸聚合物的区域进行修饰, 以便能够将具有被标记或待分析的序列 (R2) 的核酸聚合物限定在多个位置。

[0103] 作为非限制性的实例, 已知 R2 可以位于特定疾病的基因内, 并且聚合物中存在多个 R2 序列可以证实该序列的异常 (或正常) 拷贝数。可以将聚合物沿着通道从一个储液器移位到另一个储液器, 并可以在沿着其移位路径的任何位点处停止或固定化。

[0104] 固定化可以通过多种方式来实现。在一个实施方案中, 如图 10a 中所示, 将大分子结合到至少一个珠子上, 通过至少一个珠子被横截面小于珠子的缢缩段捕获, 使分子被固定化。固定化也可以通过将大分子化学限定到表面上、通过大分子的磁固定、通过大分子的光捕获或其任意组合来实现。

[0105] 在包含珠子的实施方案中, 珠子被选择成使其有效直径大于纳米通道的至少一个横截面维度。当修饰的核酸分子流入纳米通道时, 因为修饰的珠子比纳米通道的至少一部分大, 其流动被阻碍。然后核酸分子的未修饰部分能够被线性化并用于序列分析。珠子可以是聚合的、磁性的、半导体的、绝缘的、金属的或其任意组合, 并且核酸分子的修饰可以基于共价键或非共价相互作用, 包括蛋白质相互作用, 并可以包含中间连键。在所有限定或固定化方式中, 可以调节所施加的流动或梯度场, 以便能够形成或松开限定。

[0106] 可以选择用于限定的修饰物质, 使得将核酸分子结合在纳米通道中的性质是磁、电、光、化学、摩擦、基于流动、物理阻碍或其任意组合。

[0107] 在其他实施方案中, 核酸分子在分子的一个末端处或附近进行化学修饰, 如非限制性的图 10b 中所示。化学修饰被选择成使得在修饰物质与纳米通道之间发生的共价或非共价相互作用的强度足以限定核酸分子并阻止其流过纳米通道。

[0108] 化学修饰剂的实例包括巯基、硅烷基、羧基、胺基、烷基、磷酸基、可光切割基团、蛋白、生物素、氨基酸残基、金属基团或其任意组合。在某些情况下, 纳米通道的表面可以包括某些化学修饰以促进与修饰物质的相互作用。

[0109] 在另一个实施方案中, 核酸分子在分子的一个末端处或附近进行磁性修饰, 如图 11a 中所示。磁性修饰可以是磁性珠子、顺磁性珠子、超顺磁性粒子或其他能够在序列分析过程中维持磁性偶极的部分。在这样的情况下, 磁力可以整合在纳米通道装置内或附近, 或

者,也可以是外部施加的磁场的结果,如图 11a 中所示。

[0110] 在另一个实施方案中,核酸在分子的一个末端处或附近用能够在光学镊子存在下经历击穿力 (dielectric force) 梯度的粒子或部分修饰。这显示在非限制性的图 11b 中。

[0111] 正如所示,光学镊子被用于当粒子流过纳米通道时捕获限制在光束中的粒子,从而允许相连的核酸分子在纳米通道中被线性化。光学镊子可用于移动靶以及将其固定化。

[0112] 在另一个实施方案中,使用了多种力固定化或限定 DNA。例如,可以同时使用反向流体流动和电场,以将分子在分析区域中保持伸展和静止。

[0113] 适合情况下,线性化通过适合地定制通道的尺寸以执行大分子的线性化、适合情况下通过物理熵限制来实现。

[0114] 还提供了分析系统。本发明的系统包括含有至少一个宽度在约 1 到约 500 纳米范围内的通道的基材;所述基材包含至少一个固定化区域。适合情况下,通道具有约 10 到约 200nm、或约 20 到约 100nm 范围内或甚至约 50nm 的宽度。通道的深度可以在同样范围内,尽管具体通道的宽度和深度不是必须相同的。通道事实上可以是任何长度,从 10nm 直到数厘米。适合情况下,这样的通道具有毫米范围内的长度,尽管对于给定应用来说,最适长度对于本技术领域的普通专业用户来说是显而易见的。

[0115] 固定化区域能够固定化大分子。大分子可以包括一种或多种修饰,其可以包括悬片段、珠子、绝缘修饰、磁性粒子等。系统和大分子修饰可以协调地并在其彼此的亲和性的基础上选择。示例性的固定化区域包括磁性区域、化学活性区域、缢缩段等,如图 10 和图 11 中所示。

[0116] 在某些实施方案中,将聚合物固定化并施加梯度,以便将聚合物的至少一部分置于通道中,如图 10 和图 11 所示。通过这种方式,如本文别处所述可以被标记的聚合物可以被线性化,并通过将其限制在通道中可以保持在线性形式。

[0117] 尽管没有在图中显示,本发明还包括下列实施方案,其中被标记的聚合物被固定化或限定,然后通过施加梯度(压力、电等)线性化,以便位于聚合物上的一个或多个标记物(或悬片段)可以被检测,并与聚合物的特征相关联。可以通过持续施加梯度或在聚合物被梯度线性化后将其附着到基材上(即聚合物被线性化然后在其线性化形式下附着到基材上),将聚合物维持在线性形式下。

[0118] 还提供了对核酸聚合物进行表征的方法,这些方法包含用一种或多种序列特异性基序标记物标记核酸聚合物的一个或多个区域;将来自一个或多个序列特异性基序标记物的一种或多种信号,与核酸聚合物的一个或多个序列特异性基序标记物的位置相关联;对核酸聚合物的一个或多个区段进行测序,所述一个或多个区段包括核酸聚合物的一个或多个序列特异性基序标记物;以及将一个或多个被测序的区段的一种或多种信号与标记的核酸聚合物的一个或多个相应信号进行比较,以便获得两个或多个被测序区段在核酸聚合物中的相对位置。

[0119] 适合情况下,本公开的方法的标记方面通过本文别处描述的标记方法来实现,即在核酸聚合物中形成悬片段,并标记悬片段、由悬片段空出的区域或其任意组合。适合的标记物和标签在本文别处描述。

[0120] 适合情况下,关联需要将核酸聚合物的至少一个被标记部分线性化。线性化可以通过在尺寸适合的纳米通道中将聚合物的被标记部分线性化、通过向聚合物施加梯度(例

如流体、电)等来实现。在其他实施方案中,聚合物被限定或固定化,并通过施加梯度(压力、电等)线性化。区段可以通过核酸聚合物的随机或序列特异性切割开产生。

[0121] 关联可以包括例如测定两个或多个标记物之间的距离、比较从两个或多个标记物产生的信号的强度等。在某些情况下被称为“重叠群”的聚合物区段的测序,可以通过本技术领域已知的技术来实现。这些技术包括例如 Sanger 测序、Maxam-Gilbert 测序、染料终止物测序、体外克隆扩增、杂交测序等。适合情况下,区段的长度高达 30kb 或甚至 50kb,但是适合情况下在 kb 长度范围内。

[0122] 被标记区段的信号或多个信号与被标记核酸聚合物的相应信号的比较,通过例如将一个或多个被标记的测序区段针对被标记的核酸聚合物进行比对,使得被标记的测序区段的序列特异性基序标记物被放置成与被标记的核酸聚合物的相应序列特异性基序标记物配准来实现。这有效地允许用户使用区段上的标记物作为允许鉴定每个单独区段的“条形码”。因此,通过将条形码重叠群针对“亲本”核酸聚合物上的相应条形码进行匹配,用户可以确定条形码重叠群在“亲本”核酸聚合物中的位置(和取向)。

[0123] 采用这种方式,通过将来自区段上的标记物的一种或多种信号与“母本”聚合物上的相应标记物进行比对,用户可以确定区段的适合比对。通过对多个区段重复该过程,用户就能确定区段的适合次序和取向,允许核酸聚合物的大规模平行测序。

[0124] 该过程进一步描绘在图 6 中,该图描绘了本发明的从核酸聚合物获得基因组支架(例如序列)装配信息的方法。

[0125] 如图中所示,用户从聚合物中提取相对长的基因组 DNA 分子(例如 1kb 直到 100mb 或以上),并按照例如本文别处描述的标记方法标记分子,以便导致产生沿着线性化长聚合物而检测和记录到的序列特异性信号,以产生代表分子基因组的特定区域的特征性的“观察到的条形码”(显示成“观察到的条形码的原始图像”);分子可以表示基因组。然后可以将来自个体分子的观察到的条形码组装成相对长的支架,所述支架可以长达完整基因组的尺寸。

[0126] 离散的区段(在某些实施方案中是“重叠群”;从约 5 到约 30kb)可以根据由当前的测序技术产生的部分重叠的短碱基读出结果进行计算机组装。这样的重叠群可以是随机的,或在序列特异性的基础上产生。正如图 6 中所示,可以将基因组片段化成例如 50bp 到高达 1000bp 的重叠群。然后用户可以产生众多(数百万)约 35 到约 850bp 的短的读出结果。

[0127] 适合情况下,将一个或多个重叠群用与标记“亲本”核酸聚合物所使用的序列特异性基序相同的序列特异性基序(例如 Nb. BbvCI 位点, GCTGAGG)进行标记,以产生一系列条形码。当重叠群被虚拟标记(即通过计算机)时,条形码被当作计算机用(in silico)条形码。

[0128] 然后用户将重叠群(区段)的条形码针对实验构建的支架的相应的观察到的条形码进行比对,然后所述比对为用户提供了重叠群在支架中的物理位置以及重叠群在支架中的适合取向。这反过来产生了关于支架(以及相应基因组)的信息,例如支架中的序列拷贝数、结构信息(例如翻译)等。因此,每个重叠群被精确作图到基因组上,以便产生所分析的具体聚合物的真实、准确的基因组测序信息。

[0129] 这些方法与现有测序技术相比具有大量优点,包括提供关于拷贝数的信息的能力

和将重叠群相对于彼此放置在正确位置 / 次序的能力。这反过来提供了真实的序列信息；不适用本文描述的条形码技术，重叠群沿着被分析基因组的线性次序将是未知的，特别是，如果没有以前的参比数据库可供比较的情况下（从头测序）。由于具有拷贝数变异（CNV）和结构性变异（SV）的大基因组的高度复杂性，从随机的较短读出结果直接独立组装，特别是对于从头测序或高度不规则的癌症基因组来说，已变得越来越困难并容易出错。

[0130] 作为一个非限制性的实例，（已知序列的）第一个区段可以包含条形码 A、B 和 C，其每个对应于区段上序列特异性标记物的位置、序列特异性标记物的强度或两者。因此，基于 A、B 和 C 条形码，被标记的区段表现出独特的图形。（已知序列的）第二个标记的区段可以包含条形码 C、D 和 E。通过将第一和第二个区段针对从其切下区段的“母本”聚合物进行比对，用户能够确定两个区段在条形码 C 处重叠，并且通过合并两个区段的序列（没有两次计算对应于条形码 C 的序列），可以确定两个区段所源自的“母本”聚合物的序列。因此，通过将该过程规模放大以同时处理多个区段，本发明的方法能够确定长的核酸聚合物的序列信息。

[0131] 一个类似的实施方案显示在图 3 中。该图显示了使用 λ DNA 的实例，预测的切口酶 Nb. BbvC I 的切口位点显示在序列基序中，并用沿着长的 DNA 分子的箭头指示。将切口位点用荧光（Alexa）核苷酸 T 标记，所述荧光（Alexa）核苷酸 T 在天然 T 碱基被置换并取代时掺入到切口位点中（显示为绿色）。

[0132] 在该模型系统中，基于在低盐条件下在 80nm 乘 80nm 宽的通道中 100% 伸展的 λ DNA，观察到的标记的特征性“条形码”样式与在这里被称为计算机条形码的使用切口酶消化产生的基因组的计算机预测序列基序图谱一致，如图 3b 所示。在完全伸展的线性化人类 BAC 克隆 DNA 上（~ 170Kbp）显示了类似的条形码结果；也显示了超过 17 个被标记的位点（荧光颜色）。

[0133] 附加实施例和实施方案

[0134] 附加实施方案

[0135] 正如本文中别处所述，本发明尤其提供了与 DNA 作图和测序相关的方法，包括用于制造长的基因组 DNA 的方法、标记序列特异性标签并基于单个 DNA 分子的直接成像和在纳米通道（直径 < 500nm）内的单个 DNA 分子上定位多个序列基序或多态性位点的 DNA 条形码策略的方法。方法还提供了在 DNA 图谱的背景中提供连续的逐个碱基测序信息。与现有方法相比，本发明的 DNA 作图提供了增加的标记效率、更稳定的标记、高的灵敏度和更好的分辨率；我们的 DNA 测序方法提供了长的模板背景中的碱基读出结果，易于组装，并提供了不能从其他测序技术获得的信息，例如单倍型和结构变异。

[0136] 在 DNA 作图应用中，将单个基因组 DNA 分子或长范围 PCR 片段在特异性序列基序处用荧光染料标记。被标记的 DNA 分子在纳米通道内被拉伸成线性形式（在本文别处描述），并使用荧光显微术进行成像。通过确定荧光标记物相对于 DNA 骨架的位置和在某些情况下其颜色，可以精确地建立序列基序的分布，类似于包装物上的条形码。这种 DNA 条形码方法被用于鉴定 λ 噬菌体 DNA 分子和人类 bac 克隆。

[0137] 利用在 dsDNA 上的特异性序列位点处切口的一种实施方案包含下列步骤：

[0138] a) 使用在特异性序列基序处导入切口的一种或多种切口内切核酸酶，在长的（例如超过 2kb）双链基因组 DNA 分子的一条链上产生切口；

[0139] b) 使用 DNA 聚合酶在切口处掺入荧光染料标记的核苷酸；

[0140] c) 将被标记的 DNA 分子在纳米通道内拉伸成线性形式，分子流过通道或分子的一部分被固定化，使 DNA 的一个末端被置于通道内；

[0141] d) 使用荧光显微术确定荧光标记物相对于 DNA 骨架的位置，以获得 DNA 的图谱或条形码。

[0142] 另一个在序列特异性切口位点处带有悬片段序列的实施方案包括下列步骤：

[0143] a) 使用在特异性序列基序处导入切口的切口内切核酸酶，在长的 (> 2kb) 双链基因组 DNA 分子的一条链上产生切口；

[0144] b) 使用 DNA 聚合酶在切口处掺入荧光染料标记的核苷酸或没有荧光染料标记的核苷酸，置换下游链以产生悬片段序列；

[0145] c) 通过聚合酶掺入标记的核苷酸、或与荧光探针直接杂交或用连接酶连接荧光探针，以标记悬片段序列；

[0146] d) 如本文别处所述将被标记 DNA 分子拉伸成线性形式；

[0147] e) 使用荧光显微术确定荧光标记物相对于 DNA 骨架的位置，以获得 DNA 的图谱或条形码。

[0148] 另一个利用序列特异性切口位点处的 ssDNA 间隙的实施方案包括下列步骤：

[0149] a) 使用在特异性序列基序处导入切口的切口内切核酸酶，在长的 (> 2kb) 双链基因组 DNA 分子的一条链上产生切口；

[0150] b) 使用 DNA 聚合酶在切口处掺入荧光染料标记的核苷酸探针或没有荧光染料标记的核苷酸，置换下游链以产生一个或多个悬片段序列；

[0151] c) 使用切口内切核酸酶在新延伸的链上产生切口，并用悬片段内切核酸酶切下新形成的悬片段序列。分开的 ssDNA 可以通过例如增加温度以释放它们的键来移除；

[0152] d) 通过掺入标记的核苷酸、或与荧光探针直接杂交或用连接酶连接荧光探针，对 ssDNA 间隙（通过产生切口和随后形成悬片段所产生的）进行标记；

[0153] e) 如本文别处所述将被标记 DNA 分子拉伸成线性形式；

[0154] f) 使用荧光显微术确定荧光标记物相对于 DNA 骨架的位置，以获得 DNA 的图谱或条形码。

[0155] 在其他 DNA 测序应用中，将单个基因组 DNA 分子或长范围 PCR 片段在特异性序列基序处用荧光染料标记。然后将被标记 DNA 分子在纳米通道内线性化，然后使用荧光显微术成像。通过测定荧光标记物相对于 DNA 骨架的位置和颜色，可以与读取条形码相似的方式精确地建立序列基序的分布。在 DNA 图谱背景中获得了单个或多个碱基信息。

[0156] 这种可用于基因组 DNA 的测序方法的一个实施方案包括下列步骤：

[0157] a) 使用在特异性序列基序处导入切口的切口内切核酸酶，在长的 (> 2kb) 双链基因组 DNA 分子的一条链上产生切口；

[0158] b) 通过切口掺入、悬片段标记、ssDNA 间隙标记或其某些组合，使产生切口的位点带上荧光染料分子的标签；

[0159] c) 如本文别处所述将被标记 DNA 分子拉伸成线性形式；

[0160] d) 使用荧光显微术确定荧光标记物相对于 DNA 骨架的位置，以获得 DNA 的图谱或条形码；

[0161] e) 使用产生切口的位点作为测序反应的起始点。在 DNA 测序中有用的不同 DNA 结构,包括但不限于下列结构。

[0162] 在一个测序实施方案中,聚合酶在切口位点的 3' 末端掺入荧光核苷酸,相继地检测每个切口位点处掺入的标记物以获得序列信息。该过程被重复 / 循环,以连续获得许多碱基上的“读取结果”。

[0163] 在另一个实施方案中,将测序引物与悬片段序列杂交并使用聚合酶进行延伸,以掺入荧光核苷酸。通过读取这些各种被掺入的荧光核苷酸的颜色,就可以推断出序列信息。该过程被重复 / 循环,以连续获得许多碱基读取结果。

[0164] 在另一个实施方案中,将一个短的荧光寡核苷酸与悬片段序列直接杂交,序列信息可以从杂交的寡聚物的存在推断出来。该过程被重复 / 循环,以连续获得许多碱基读取结果。

[0165] 在另一个实施方案中,将两个短的寡核苷酸与彼此相邻的悬片段序列杂交,然后使用例如连接酶连接在一起。序列信息可以从连接产物推断出来。该过程被重复 / 循环,以连续获得许多碱基读取结果。

[0166] 在另一个实施方案中,将一个短的荧光寡核苷酸紧靠切口位点的 3' 末端直接杂交并连接。然后序列信息从连接的寡核苷酸的存在推断出来。该过程被重复 / 循环,以连续获得许多碱基读取结果。

[0167] 方法可以与纳米通道阵列相结合来执行。适合情况下,这样的阵列具有多个位于表面的材料中的通道,所述通道具有小于约 500 纳米的沟槽宽度和小于约 500 纳米的沟槽深度。适合情况下,至少某些通道被密封材料覆盖,使这些通道至少基本上是封闭的。

[0168] 在某些实施方案中,本发明包括筒或其他模块化装置。这样的筒可以包括,例如在本文中公开包含纳米流体芯片的筒。这样的筒能够被插入、使用并取出。可用于本发明的系统之外的分析系统的筒也在本发明的范围内。

[0169] 在某些实施方案中,纳米通道能够运输大分子通过其纵向。本发明的装置可以包含一种或多种可用于执行大分子运输的部件,所述运输可以通过跨过通道的压力或真空梯度、电渗作用和电动效应 (electrokinetic) 来执行。

[0170] 纳米通道的表面材料可以由几乎任何基材材料形成,例如导电材料、半导体材料或不导电材料。导电材料的实例包括金属例如铝、金、银和铬。半导体材料的实例包括掺杂的二氧化硅和砷化镓。不导电材料的实例包括熔融二氧化硅、二氧化硅、氮化硅、玻璃、陶瓷和合成聚合物。上述的仅仅是示例。

[0171] 在某些实施方案中,对核酸分子在一个末端处或附近进行修饰,然后将其置于纳米通道或纳米轨道(由限制流体通道的边界例如疏水边界所限定的区域)中。适合情况下,修饰允许核酸限定在纳米通道入口处或纳米通道内。

[0172] 然后,由于纳米通道,核酸受到限制以采用线性化形式。适合的核酸是 DNA 或 RNA,例如 dsDNA。纳米通道的长度优选 < 500nm、更优选 < 300nm、最优选 < 150nm,能够容纳超过 2000 碱基的线性化核酸。

[0173] 下面的实施方案也适用于纳米轨道,其是在化学或拓扑学预定义的表面图样上定义的线性区域。

[0174] 可以通过系统分析的流体包括来自哺乳动物的流体(例如 DNA、细胞、血液、活体

组织)、合成的大分子例如聚合物,以及在自然界中发现的材料(例如源自于植物、动物和其他生命形式的材料)。这样的流体可以使用本发明的自动或手动加样装置管理、加样和注射。

实施例

[0175] 实施例 1:在双链 DNA 分子上产生单链 DNA 悬片段

[0176] 将基因组 DNA 样品稀释到 50ng 以用于切口产生反应。将 10 μ L λ DNA (50ng/ μ L) 加入到 0.2mL PCR 离心管中,然后加入 2 μ L 10X NE 缓冲液 #2 和 3 μ L 切口内切核酸酶,包括但不限于 Nb. BbvCI、Nb. BsmI、Nb. BsrDI、Nb. BtsI、Nt. AlwI、Nt. BbvCI、Nt. BspQI、Nt. BstNBI、Nt. CviPII。将混合物在 37°C 下温育 1 小时。

[0177] 在切口反应完成后,在切口位点处使用有限聚合酶延伸继续实验,以置换 3' 下游链并形成单链悬片段。悬片段产生反应混合物由 15 μ l 切口产物和 5 μ l 掺入混合物构成,所述掺入混合物含有 2 μ l 10X 缓冲液、0.5 μ l 包括但不限于 vent(外切)、Bst 和 Phi29 聚合酶的聚合酶以及 1 μ l 从 1 μ M 到 1mM 的各种不同浓度的核苷酸。将悬片段产生反应混合物在 55°C 下温育。悬片段的长度由温育时间、使用的聚合酶和使用的核苷酸量来控制。

[0178] 实施例 2:在双链 DNA 分子上产生单链 DNA 间隙

[0179] 在悬片段产生后,使用原始的切口内切核酸酶在填充的双链 DNA 上产生切口,并使用包括但不限于 FEN1 的悬片段内切核酸酶切掉悬片段序列。通过升高温度,将带切口的单链 DNA 分子从双链 DNA 分子中除去,以产生双链 DNA 分子上的单链 DNA 间隙。

[0180] 实施例 3:产生长的单链 DNA 分子

[0181] 在切口产生反应完成后,在切口位点处使用完全聚合酶、包括但不限于 Phi29、Bst 聚合酶延伸继续进行实验,以置换 3' 下游链并产生单链 DNA 分子。

[0182] 实施例 4:在双链 DNA 分子上荧光标记序列特异性切口的方法

[0183] 将基因组 DNA 样品稀释到 50ng 以用于切口产生反应。将 10 μ L λ DNA (50ng/ μ L) 加入到 0.2mL PCR 离心管中,然后加入 2 μ L 10X NE 缓冲液 #2(New England BioLabs, www.neb.com) 和 3 μ L 包括但不限于 Nb. BbvCI、Nb. BsmI、Nb. BsrDI、Nb. BtsI、Nt. AlwI、Nt. BbvCI、Nt. BspQI、Nt. BstNBI、Nt. CviPII 的切口内切核酸酶。将混合物在 37°C 下温育 1 小时。

[0184] 在切口产生反应完成后,使用聚合酶延伸继续进行实验,以在切口位点上掺入染料核苷酸。在一个实施方案中,掺入单一荧光核苷酸终止物。在另一个实施方案中,掺入多个荧光核苷酸。

[0185] 掺入混合物由 15 μ l 切口产物和 5 μ l 掺入混合物构成,所述掺入混合物含有 2 μ l 10X 缓冲液、0.5 μ l 包括但不限于 vent(外切)的聚合酶、1 μ l 荧光染料核苷酸或包括但不限于 cy3、alexa 标记的核苷酸的核苷酸终止物。将掺入混合物在 55°C 下温育约 30 分钟。

[0186] 实施例 5:在双链 DNA 分子上序列特异性标记单链 DNA 悬片段的方法

[0187] 在悬片段序列产生后,可以将悬片段用荧光染料分子进行标记,标记方法包括但不限于下列方法:探针的杂交、使用聚合酶掺入荧光核苷酸和荧光探针的连接。

[0188] 实施例 6:在双链 DNA 分子上序列特异性标记单链 DNA 间隙的方法

[0189] 使用毛细作用,将宽度、深度或两者为 500nm 或以下的纳米流体芯片用含有染色

的基因组 DNA 的缓冲溶液进行填充,以便使用电场将 DNA 大分子拉入通道。将 λ 细菌噬菌体 DNA 分子 (48.5kb) 和人类 BAC 克隆 (170kb) 用染料 YOYO-1 染色。将该染色的 DNA 溶液在含有 0.1M 作为抗氧化剂的二硫苏糖醇和 0.1% 用作抗粘剂的线性丙烯酰胺的 0.5XTBE 中稀释到 0.5 μ g/mL。

[0190] 使用带有 100X(N. A. 1.35) 油浸物镜的 Olympus Ix-71 倒置显微镜,并使用固态激光器(例如二极管泵浦固态激光器),所述激光器可以具有不同的激发波长(例如对于 YOYO-1 染料来说 473nm)。其他激光器(例如用于 Alexa 系列染料 Cy3、Cy5 等的激光器)包括 532nm DPSS 激光器、635 激光二极管激光器、543nm 气体激光器、591nm DPSS 激光器和 633nm 气体激光器。使用具有 512X 512 像素阵列和 16 位数字输出的 ANDOR 冷却的 EMCCD 相机对分子进行成像。数字图像使用数据处理器通过 J-image 和其他分析软件进行分析。

[0191] 实施例 7:检测流程

[0192] 在检测流程的一个实例中,通过时间延迟和集成(TDI)相机捕获在流动状态下移动的 DNA 的视频图像。在这样的实施方案中,DNA 的移动与 TDI 同步。

[0193] 在检测流程的另一个实例中,通过 CCD 或 CMOS 相机捕获在流动状态下移动的 DNA 的视频图像,将帧通过软件或硬件进行整合,以鉴定和重构 DNA 的图像。

[0194] 在检测流程的另一个实例中,通过在一组分离的传感器上同时捕获不同波长来收集 DNA 的视频图像。这通过使用一个相机和双或多视图分割器、或使用滤光器和多个相机来实现。相机可以是 TDI、CCD 或 CMOS 检测系统。

[0195] 在另一个实例中,使用同时的多波长视频检测器,使用骨架染料来鉴定独特的 DNA 片段,并使用标记物作为标志物来追踪 DNA 的运动。这在 DNA 的长度大于相机视野的情况下是有用的,并且标志物可用于协助对 DNA 的重构图像进行作图。

首先通过产生切口和置换下游链产生悬片段序列

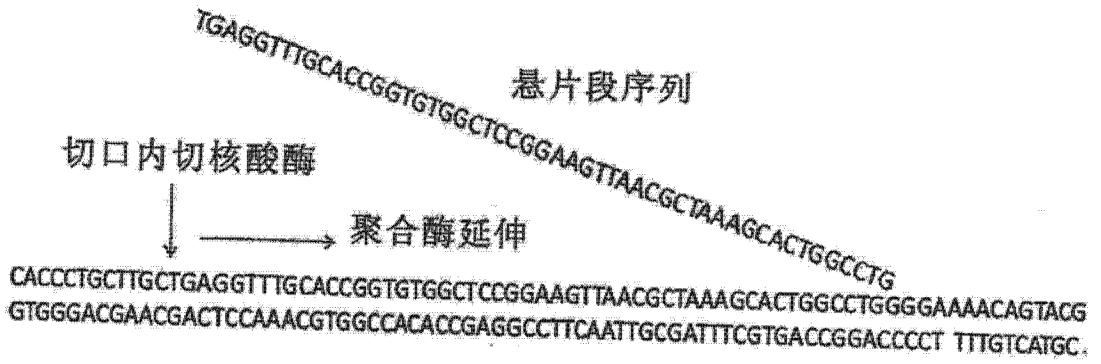
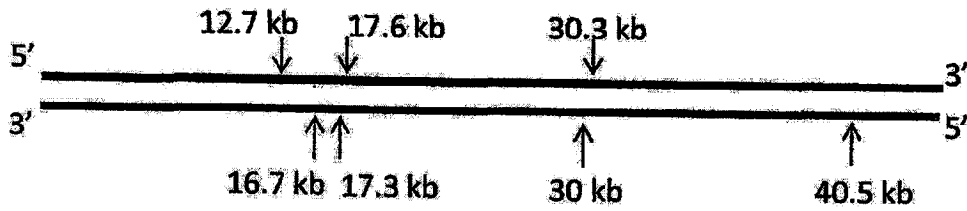
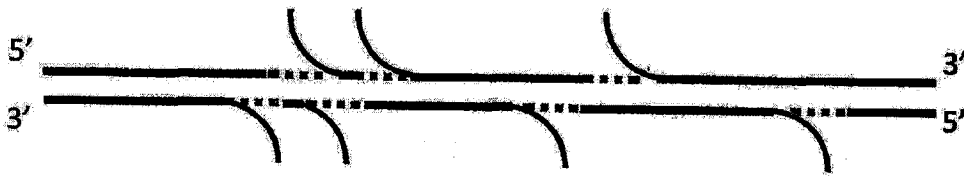


图 1a



悬片段产生



荧光序列特异性探针在一个位点处的杂交

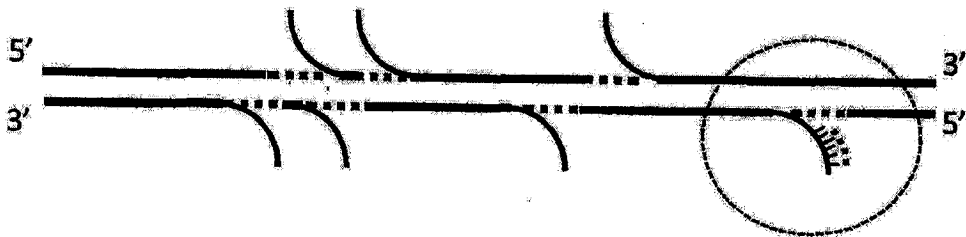
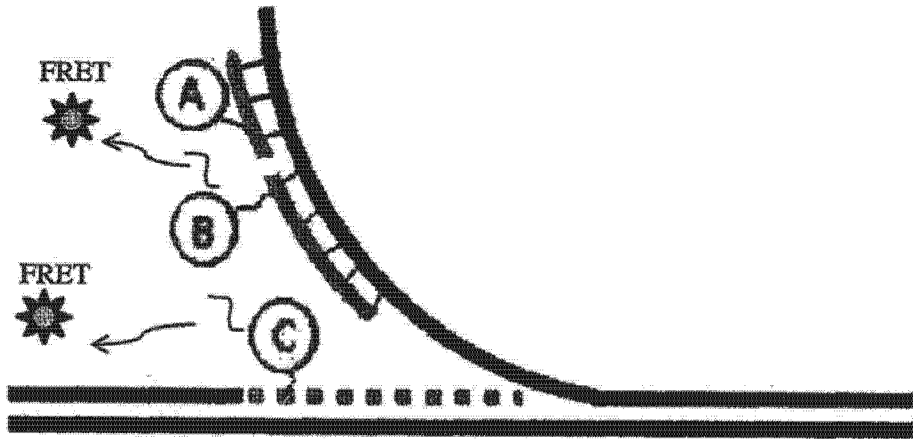
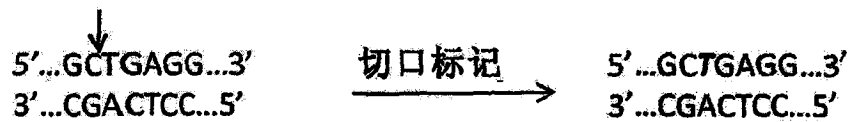


图 1b

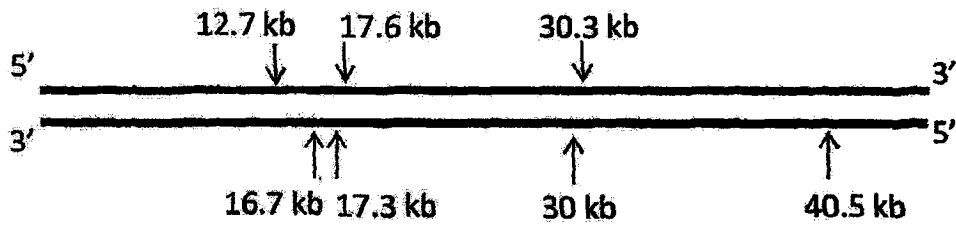


在掺入或杂交过程中标签剂 A、B 或 C 与被延伸的 DNA 相结合；通过特异性荧光共振能量转移 (FRET) 信号可以检测共定位事件。

图 2



在 λ ds-DNA (总长度 48.5 kbp) 上的计算机切口位点



在 λ ds-DNA (总长度 48.5 kbp) 上观察到的切口位点

图 3a



在完全伸展 (~170 kb) 的线性化人类 BAC 克隆 DNA 上显示的类似条形码结果；这里显示了超过 17 个被标记的位点 (荧光着色)。

图 3b

然后通过切口和悬片段内切核酸酶产生切口并切开模板，
以在 dsDNA 中留下 ssDNA 间隙

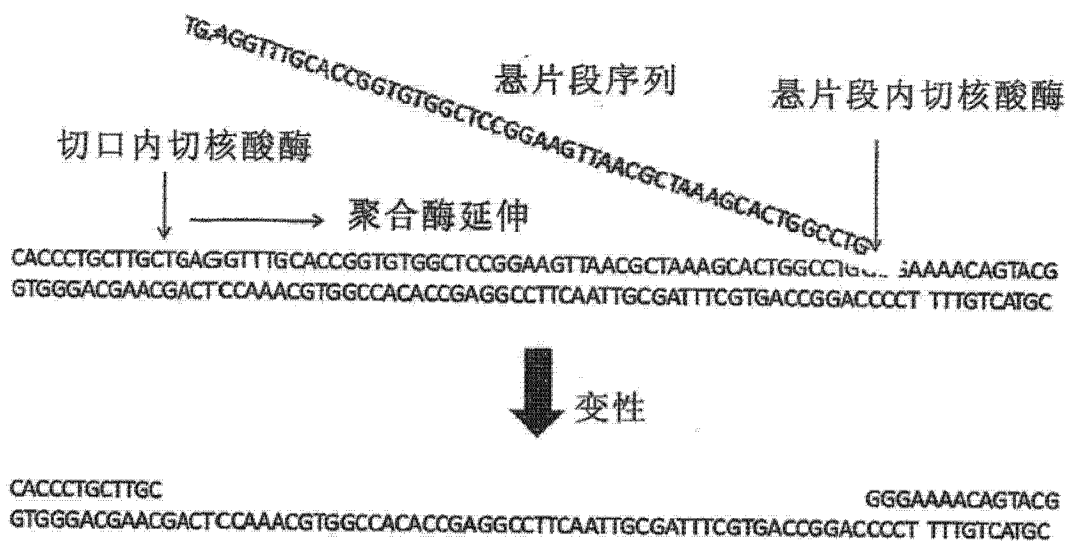


图 4a

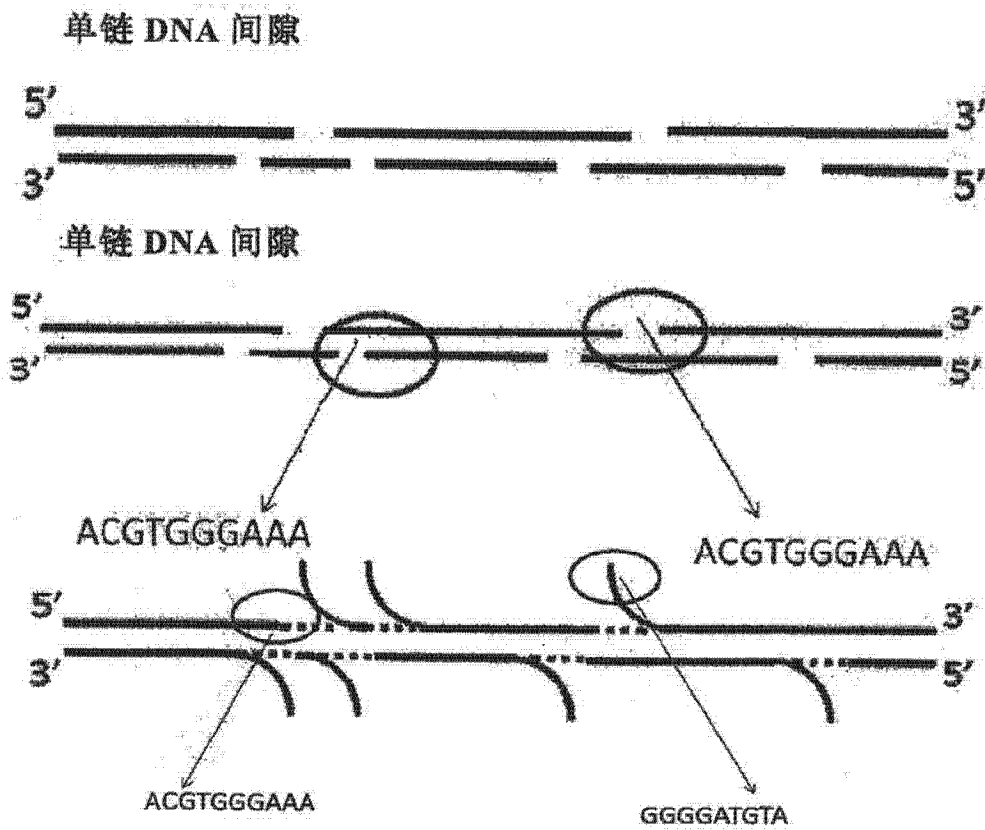


图 4b

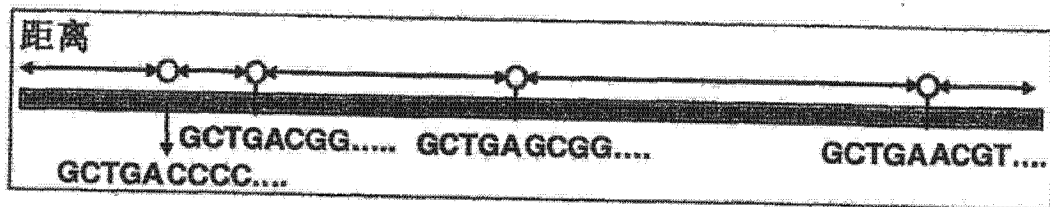


图 5a

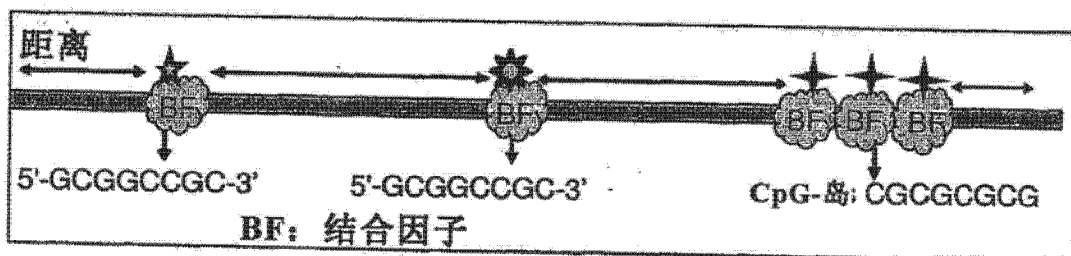


图 5b

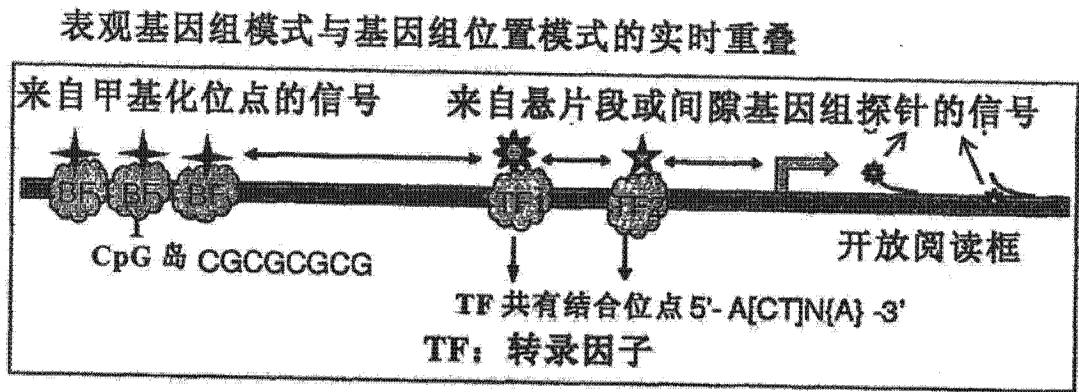


图 5c

NanoAnalyzer[®]基因组组装支架应用

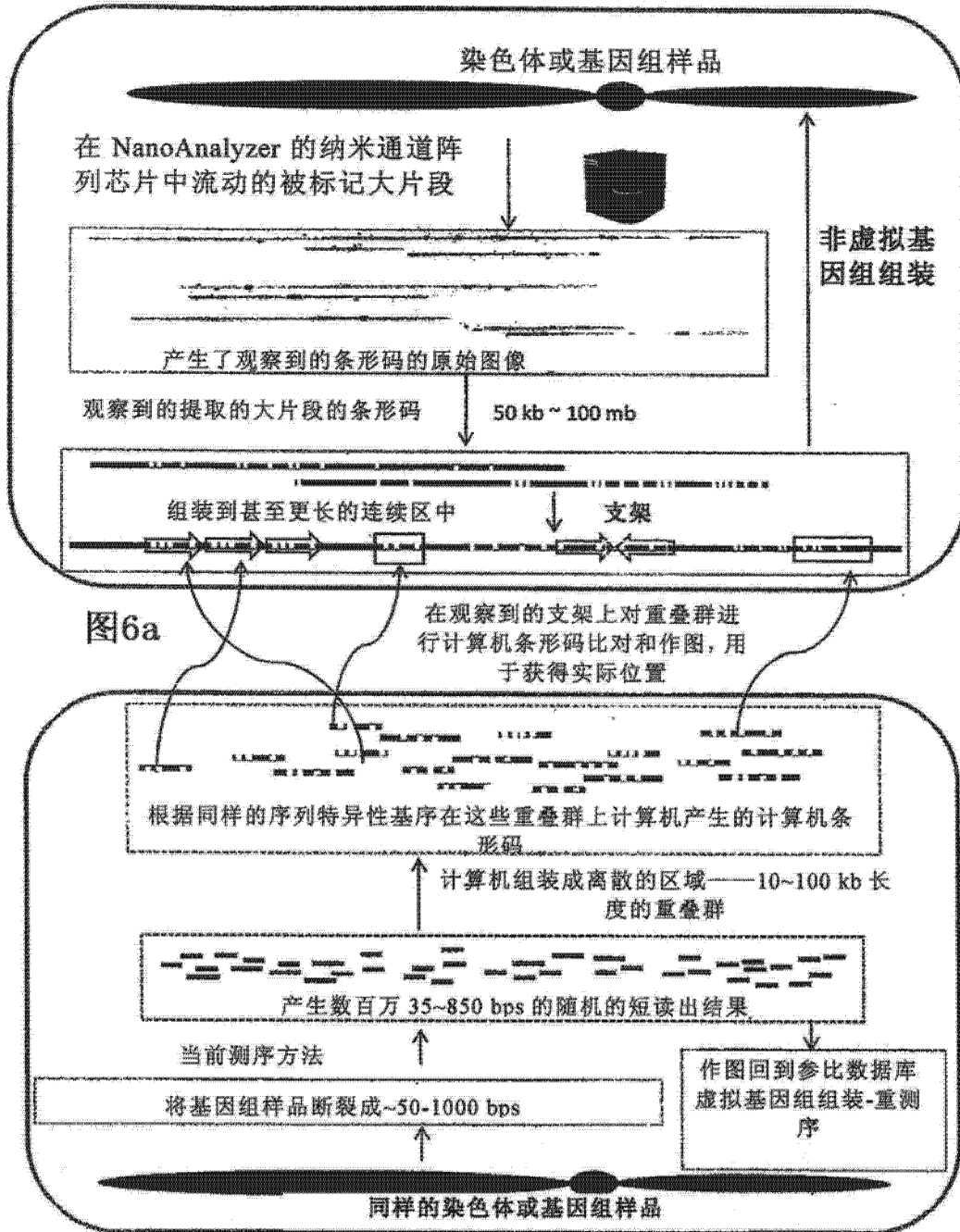


图6b

成像分析
自动 DNA 检测和尺寸确定

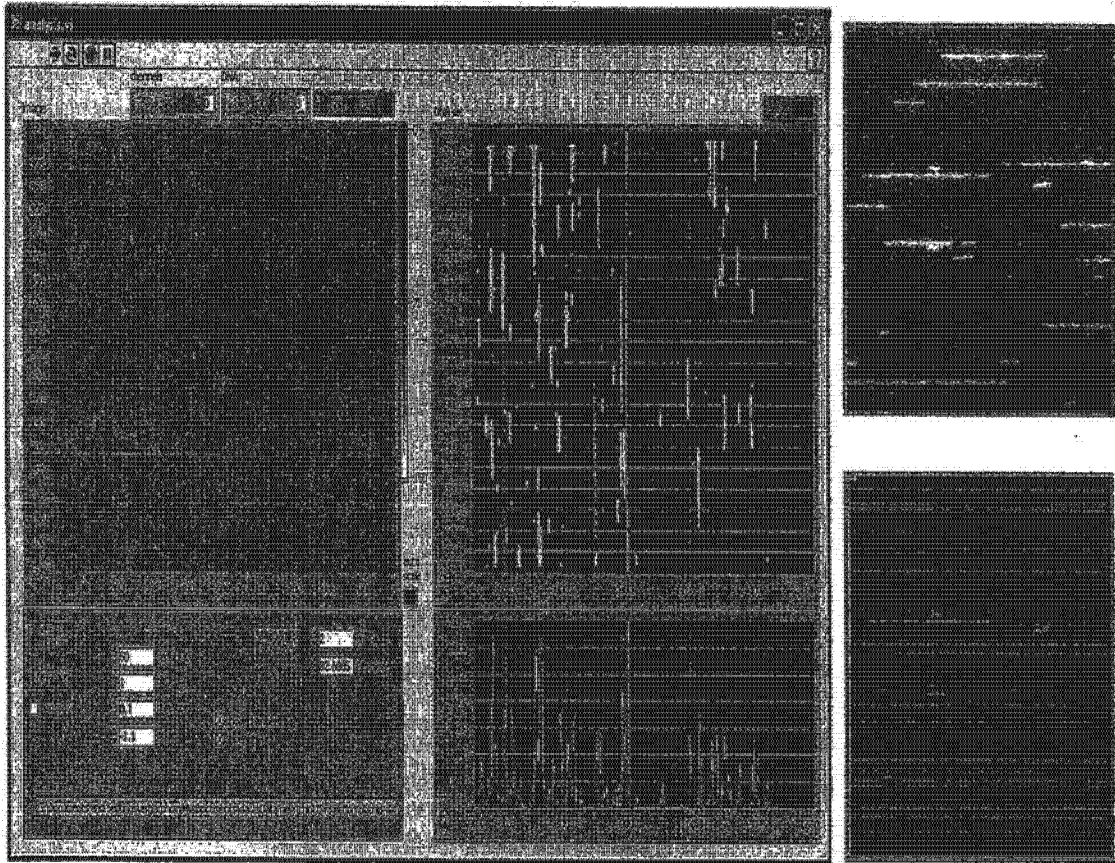


图 7

光学和非光学检测流程

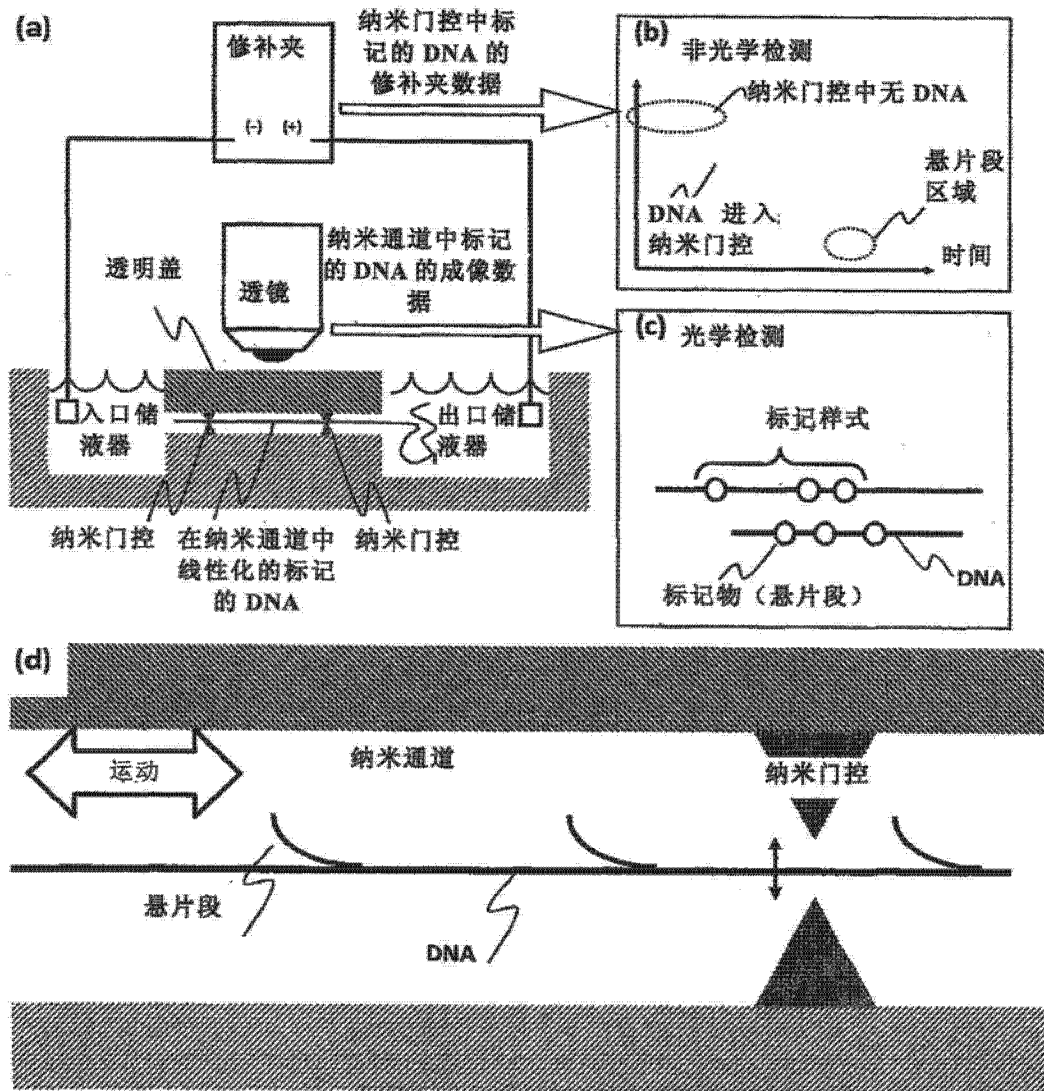
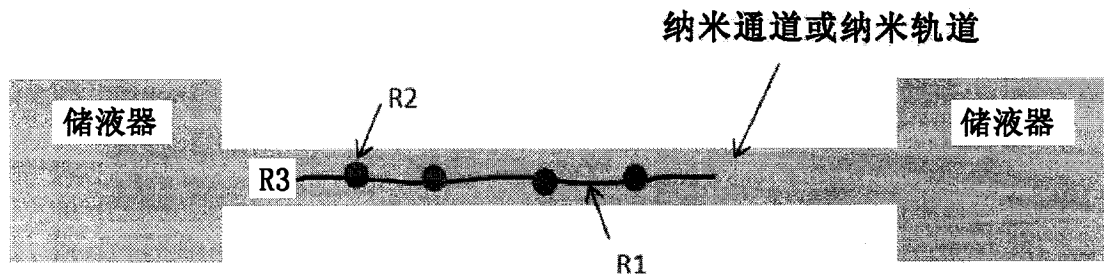


图 8



R1=核酸 (ss-DNA、ds-DNA、RNA)
R2=使用在多个点进行的分析所分析的序列
R3=用于进行限定的修饰

图 9

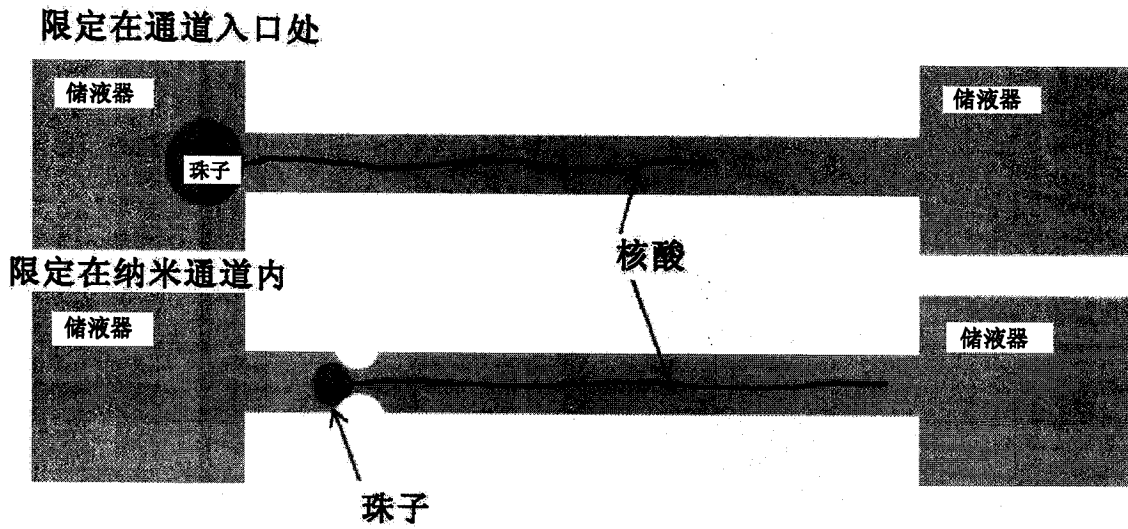


图 10a

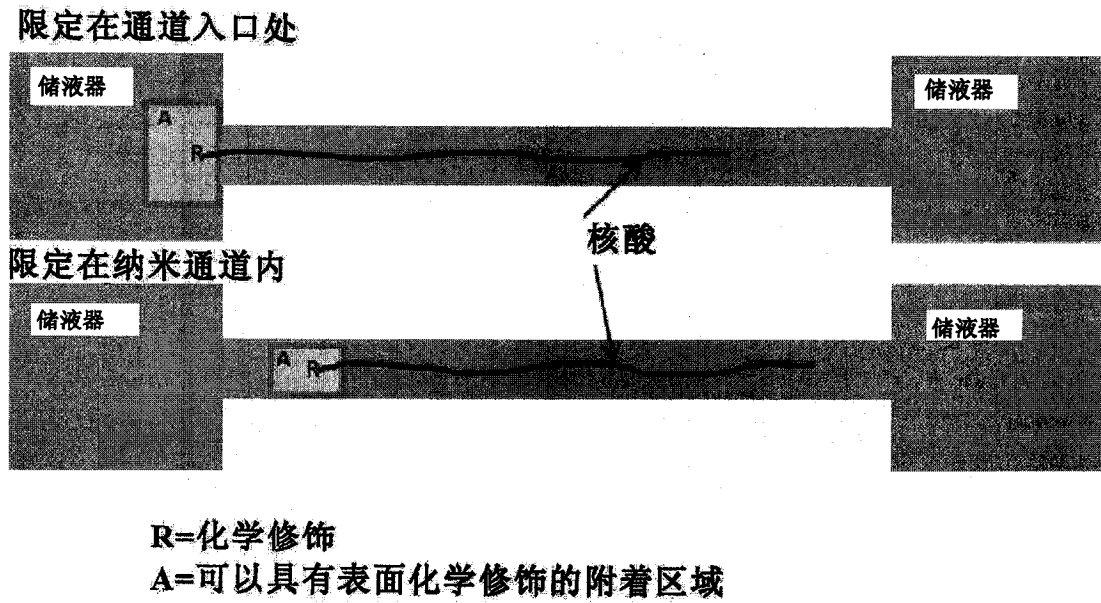


图 10b

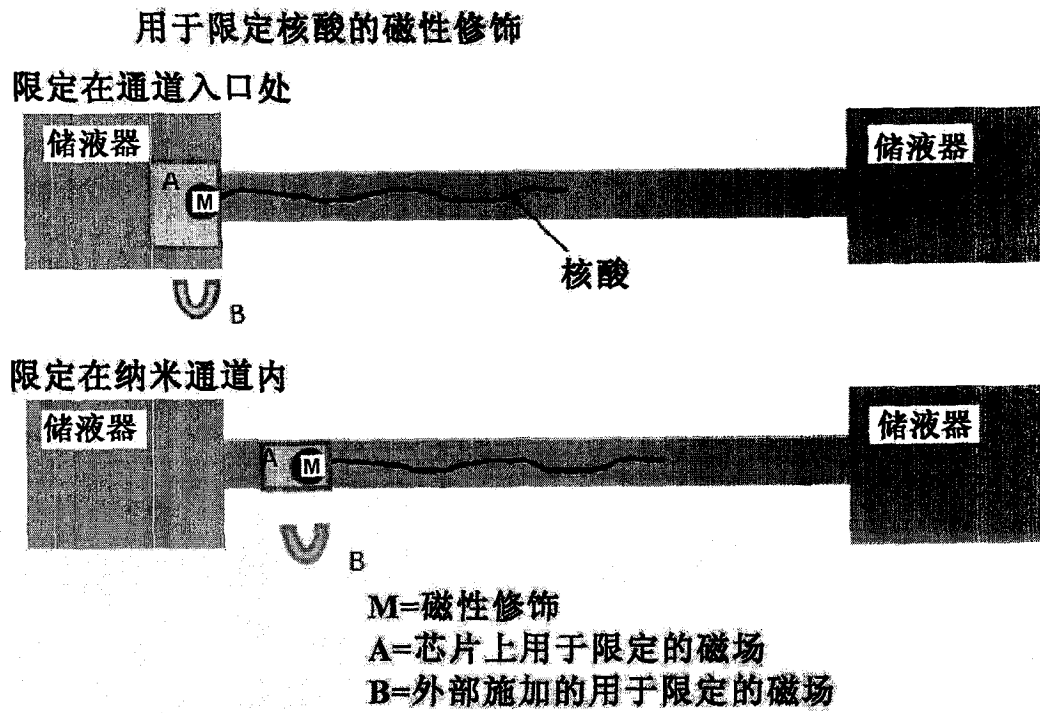


图 11a

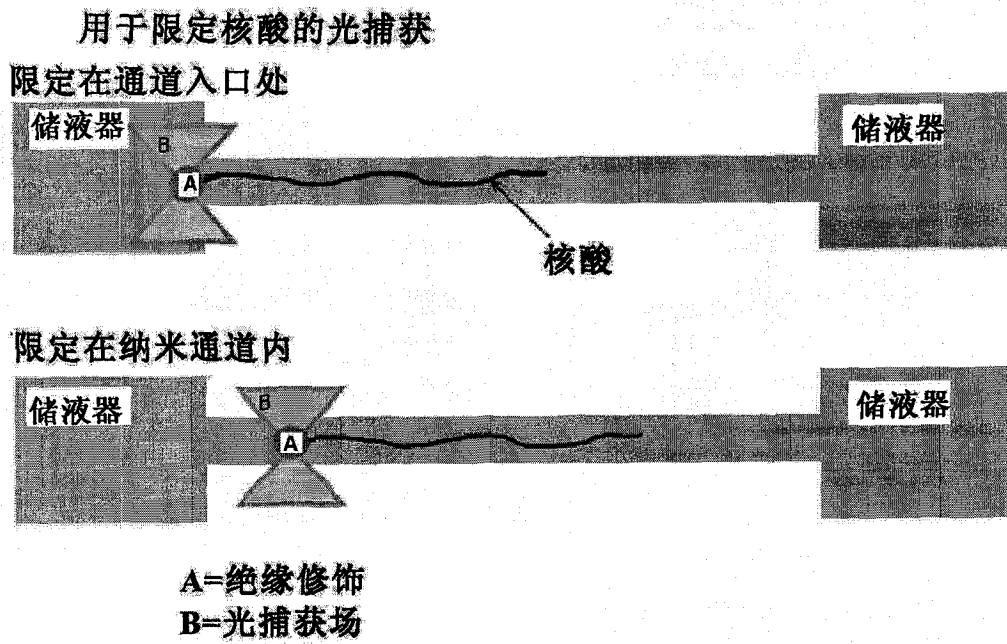


图 11b