



(12) 发明专利申请

(10) 申请公布号 CN 103080338 A

(43) 申请公布日 2013. 05. 01

(21) 申请号 201180041659. 6

(22) 申请日 2011. 08. 25

(30) 优先权数据

61/402, 350 2010. 08. 27 US

(85) PCT申请进入国家阶段日

2013. 02. 27

(86) PCT申请的申请数据

PCT/US2011/049151 2011. 08. 25

(87) PCT申请的公布数据

W02012/027572 EN 2012. 03. 01

(71) 申请人 弗·哈夫曼-拉罗切有限公司

地址 瑞士巴塞尔

(72) 发明人 S·塞沙基里

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 黄革生 林柏楠

(51) Int. Cl.

C12Q 1/68 (2006. 01)

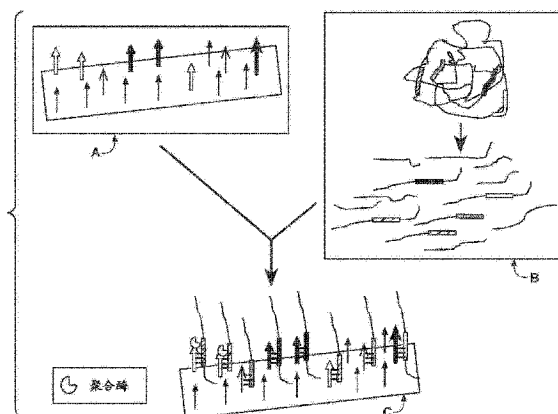
权利要求书1页 说明书11页 附图6页

(54) 发明名称

核酸捕获和测序方法

(57) 摘要

本发明提供了捕获和测序靶核酸分子的方法。并提供了测定基因组 DNA 的甲基化状态的方法。



1. 捕获和测序靶核酸分子的方法,其包括:
 - (a) 在杂交条件下将固相支持物暴露于包含靶核酸分子的核酸混合物,其中靶核酸分子与固定在固相支持物上的处于能够引发延伸的构型中的引物形成特异性杂交复合物;
 - (b) 从固相支持物上分离未结合的和非特异性结合的核酸;
 - (c) 在聚合条件下将固相支持物暴露于聚合酶和核苷酸;和
 - (d) 通过检测使用靶核酸分子作为模板,通过聚合酶从固定引物起始的核酸聚合,测定靶核酸分子的核酸序列。
2. 权利要求 1 的方法,其中所述靶核酸分子来自基因组 DNA 区域。
3. 权利要求 2 的方法,其中所述靶核酸包含外显子的全部或部分。
4. 权利要求 1 的方法,其中所述靶核酸分子是 RNA,所述聚合酶是逆转录酶,且所述引物包含 3' 多聚 T 序列。
5. 权利要求 1 的方法,其中所述靶核酸分子是 DNA,且所述聚合酶是 DNA 聚合酶。
6. 权利要求 1 的方法,其中所述核苷酸在其末端磷酸上被标记。
7. 权利要求 6 的方法,其中所述聚合酶用 FRET 供体标记,所述核苷酸用 FRET 受体标记。
8. 权利要求 7 的方法,其中所述 FRET 供体是荧光纳米颗粒。
9. 测定基因组 DNA 片段的甲基化状态的方法,所述方法包括:
 - (a) 在固相支持物上固定基因组 DNA 片段;
 - (b) 通过检测使用固定的基因组 DNA 片段作为模板,经由聚合酶的核酸聚合,测定固相支持物上固定的基因组 DNA 片段的核酸序列;
 - (c) 对固定的基因组 DNA 片段进行亚硫酸氢盐处理;
 - (d) 通过检测使用固定的经亚硫酸氢盐处理的基因组 DNA 片段作为模板,经由聚合酶的核酸聚合,测定固相支持物上固定的经亚硫酸氢盐处理的基因组 DNA 片段的核酸序列;
 - (e) 比较在 (b) 中测定的核酸序列和在 (d) 中测定的序列,其中在基因组 DNA 片段中胞嘧啶残基的转化指示,这些残基在亚硫酸氢盐处理前在基因组 DNA 片段中是未甲基化的,其中在基因组 DNA 片段中不存在胞嘧啶残基转化指示这些残基在亚硫酸氢盐处理前在基因组 DNA 片段中是甲基化的。
10. 权利要求 9 的方法,其中所述基因组 DNA 片段通过接头固定至固相支持物。
11. 权利要求 10 的方法,其中所述接头包含引物结合位点,且在引物结合位点中的胞嘧啶是受保护的。
12. 权利要求 11 的方法,其中所述 (b) 和 / 或 (d) 的聚合酶从与引物结合位点退火的引物起始聚合核酸链。
13. 权利要求 9 的方法,其中通过检测标记核苷酸的掺入,检测 (b) 和 / 或 (d) 中的核酸聚合。
14. 权利要求 13 的方法,其中所述标记的核苷酸在其末端磷酸上被标记。
15. 权利要求 14 的方法,其中所述 (b) 和 / 或 (d) 的聚合酶用 FRET 供体标记,所述核苷酸用 FRET 受体标记。
16. 权利要求 15 的方法,其中所述 FRET 供体是荧光纳米颗粒。

核酸捕获和测序方法

[0001] 相关申请

[0002] 本申请根据 35USC119(e) 要求 2010 年 8 月 27 日提交的美国临时申请号 61/402, 350 的优先权, 其内容引入本文作为参考。

发明领域

[0003] 本发明涉及从核酸混合物中分离靶核酸分子和测定这些分子的序列。

[0004] 背景

[0005] 自全部人基因组序列完成后, 基因组研究已转向“再测序(resequencing)”研究, 其中在基因组中和基因组间鉴定变异, 例如疾病相关突变。富含特定目的区域(例如外显子)的“分区”基因组的再测序需要涉及“捕获”和测序这些区域的若干步骤。例如, 在某些基于微阵列的方法中, 基因组 DNA 被剪切为特定大小范围的片段; 末端修复所述片段, 与独特的接头(adaptor)连接, 并扩增; 然后使用包含与参考目的基因组序列互补的探针的微阵列捕获扩增的片段; 洗脱和扩增捕获的(杂交的)片段; 测序扩增的片段, 例如使用“下一代(next generation)”测序技术或再测序阵列。见, 例如 W02008/115185; Okou 等人 (2007) Nature Methods 4:907-909; Hodges 等人 (2007) Nature Genetics 39:1522-1527。减少分离和测序目的核酸所需的步骤将提高效率和准确性, 且可能降低成本。本发明满足了此要求并提供了额外的益处。

[0006] 一般在基因组中的 CpG 二核苷酸处出现的胞嘧啶甲基化在基因调控和表观遗传(epigenetic inheritance)中起重要作用。测定基因组区域的甲基化状态的某些现有方法利用亚硫酸氢盐(bisulfite)处理。在这些方法中, 将变性的基因组 DNA 暴露于亚硫酸氢根离子导致胞嘧啶脱氨基为尿嘧啶, 而甲基化胞嘧啶受保护免于此转化。可例如通过下一代测序方法或通过使用探针阵列检测转化事件的存在与否。见, 例如, W02010/085343。这样的方法经常需要若干步骤, 例如扩增、捕获、洗脱和测序亚硫酸氢盐处理的 DNA, 或者备选地需要设计探针, 以检测每个目的基因组区域处的转化事件的存在与否, 这可能是昂贵和费力的。此外, 这样的方法不能在个体 DNA 分子水平上评估甲基化状态, 而是检查对应特定目的基因组区域的核酸分子群。本发明提供了更有效和准确的评估基因组 DNA 的甲基化状态的方法, 因此满足了本领域的需要并提供了其他益处。

[0007] 概述

[0008] 在一个方面, 本发明提供了捕获和测序靶核酸分子的方法, 所述方法包括 (a) 在杂交条件下将固相支持物暴露于包含靶核酸分子的核酸混合物, 其中靶核酸分子与固定在固相支持物上的处于能够引发延伸

[0009] (priming-competent) 的构型中的引物形成特异性杂交复合物; (b) 从固相支持物上分离未结合的和非特异性结合的核酸; (c) 在聚合条件下将固相支持物暴露于聚合酶和核苷酸; 和 (d) 通过检测使用靶核酸分子作为模板, 通过聚合酶从固定引物起始的核酸聚合, 测定靶核酸分子的核酸序列。

[0010] 在一个实施方案中, 靶核酸分子来自基因组 DNA 区域。在一个这样的实施方案中,

靶核酸包含外显子的全部或部分。在另一个实施方案中,靶核酸分子是 RNA,聚合酶是逆转录酶,引物包含 3' 多聚 T 序列。在另一个实施方案中,靶核酸分子是 DNA,聚合酶是 DNA 聚合酶。在另一个实施方案中,在核苷酸的末端磷酸上标记核苷酸。在一个这样的实施方案中,用 FRET 供体标记聚合酶,用 FRET 受体标记核苷酸。在一个这样的实施方案中,FRET 供体是荧光纳米颗粒。

[0011] 在另一个方面,提供了测定基因组 DNA 片段的甲基化状态的方法,所述方法包括 (a) 在固相支持物上固定基因组 DNA 片段;(b) 通过检测使用固定的基因组 DNA 片段作为模板,通过聚合酶的核酸聚合,测定固相支持物上固定的基因组 DNA 片段的核酸序列;(c) 对固定的基因组 DNA 片段进行亚硫酸氢盐处理;(d) 通过检测使用固定的亚硫酸氢盐处理的基因组 DNA 片段作为模板,通过聚合酶的核酸聚合,测定固相支持物上固定的亚硫酸氢盐处理的基因组 DNA 片段的核酸序列;(e) 比较在 (b) 中测定的核酸序列和在 (d) 中测定的序列,其中在基因组 DNA 片段中胞嘧啶残基的转化指示这些残基在亚硫酸氢盐处理前在基因组 DNA 片段中是未甲基化的,且其中在基因组 DNA 片段中不存在胞嘧啶残基转化指示这些残基在亚硫酸氢盐处理前在基因组 DNA 片段中是甲基化的。

[0012] 在一个实施方案中,通过接头将基因组 DNA 片段固定在固相支持物上。在一个这样的实施方案中,接头包含引物结合位点,且在引物结合位点中的胞嘧啶是受保护的。在一个这样的实施方案中,(b) 和 / 或 (d) 的聚合酶从与引物结合位点退火的引物起始聚合核酸链。在另一个实施方案中,通过检测标记核苷酸的掺入,检测 (b) 和 / 或 (d) 中的核酸聚合。在一个这样的实施方案中,标记核苷酸在其末端磷酸上被标记。在一个这样的实施方案中,(b) 和 / 或 (d) 的聚合酶用 FRET 供体标记,核苷酸用 FRET 受体标记。在一个这样的实施方案中,FRET 供体是荧光纳米颗粒。

[0013] 附图简述

[0014] 图 1 描述了使用互补的寡核苷酸在寡核苷酸阵列上直接捕获剪切的 DNA 的选定区域,和通过使用用于捕获 DNA 的寡核苷酸的游离 3' 末端、允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和聚合酶的标记的聚合酶和标记的 dNTP 直接测序捕获的 DNA。

[0015] 图 2 描述了使用互补的寡核苷酸包被的小珠直接捕获剪切的 DNA 的选定区域,随后阵列小珠和通过使用用于捕获 DNA 的寡核苷酸的游离 3' 末端、允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和聚合酶的标记的聚合酶和标记的 dNTP 测序捕获的 DNA。

[0016] 图 3 描述了在多聚 dT 寡核苷酸阵列上直接捕获 RNA,和通过使用多聚 dT 寡核苷酸的游离 3' 末端和逆转录酶将其转化为 cDNA 测序捕获的 RNA。在 cDNA 转化后,使用多聚 A 引物,允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和聚合酶的标记的聚合酶和标记的 dNTP 测序 cDNA。备选地,可使用允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和逆转录酶的标记的逆转录酶和标记的 dNTP 直接测序捕获的 RNA 以得到序列。

[0017] 图 4 描述了使用固定在小珠上的多聚 dT 寡核苷酸直接捕获 RNA,在表面上阵列小珠,然后通过使用多聚 dT 寡核苷酸的游离 3' 末端和逆转录酶将其转化为 cDNA 测序捕获的 RNA。在 cDNA 转化后,使用多聚 A 引物,允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和逆转录酶的标记的聚合酶和标记的 dNTP 测序 cDNA。备选地,可使用允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和逆转录酶的标记的逆转录酶和标记的 dNTP 直接测序捕获的 RNA 以得到序列。

[0018] 图 5 描述了通过测序在连续反应中在阵列上的相同的 DNA 分子,测定核酸甲基化状态,其中第二轮测序在亚硫酸氢盐处理后进行,以允许甲基化胞嘧啶转化为尿嘧啶。合适的引物、标记的聚合酶和标记的 dNTP 允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和聚合酶。

[0019] 图 6 描述了通过测序在连续反应中在小珠上的相同的 DNA 分子,测定核酸甲基化状态,其中第二轮测序在亚硫酸氢盐处理后进行,以允许甲基化胞嘧啶转化为尿嘧啶。合适的引物、标记的聚合酶和标记的 dNTP 允许在 DNA 的实时合成期间直接监测加入的碱基(序列)和聚合酶。

[0020] 发明实施方案详述

[0021] I. 定义

[0022] “亚硫酸氢盐处理”指将核酸暴露于足以将未受保护的胞嘧啶转化为尿嘧啶的浓度的亚硫酸氢根离子(例如,亚硫酸氢镁或亚硫酸氢钠)。“亚硫酸氢盐处理”也指将核酸暴露于合适浓度的可用于将未受保护的胞嘧啶转化为尿嘧啶的其他试剂,例如亚硫酸氢盐(disulfite 和 hydrogensulfite)。在暴露于亚硫酸氢根离子或其他试剂之后,“亚硫酸氢盐处理”一般包括将核酸暴露于碱,例如 NaOH。

[0023] “胞嘧啶残基的转化”指由于亚硫酸氢盐处理,胞嘧啶残基至尿嘧啶残基的转化。

[0024] “外显子”指基因组的编码区。

[0025] “杂交条件”指允许互补核酸链杂交的条件。

[0026] “测定核酸序列”指测定靶核酸分子的至少一个核苷酸的身份(identity),在一些实施方案中,测定靶核酸分子的多个核苷酸的身份。

[0027] “固定的”和“固定”指通过互补碱基配对以外的方式将核酸直接或间接连接至固相支持物。如果在特异性杂交复合物中的两条核酸链的至少一条如上面定义的“固定”在固相支持物上,则认为特异性杂交复合物固定在固相支持物上。

[0028] “标记物”指可直接或间接检测的任意部分。

[0029] “核酸”指任意长度的核苷酸聚合物。

[0030] “核苷酸”指能够通过聚合酶掺入延伸的核酸链的核苷酸及其类似物。核苷酸包括但不限于通常掺入 DNA 的 4 种核苷酸(腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶);通常掺入 RNA 的 4 种核苷酸(腺嘌呤、鸟嘌呤、胞嘧啶和尿嘧啶);具有修饰的碱基的核苷酸,例如肌苷;和标记的或修饰的核苷酸。

[0031] “聚合酶”指能够在聚合条件下将核苷酸掺入延伸的核酸链的天然或非天然存在的酶,或其酶活性片段,包括但不限于 DNA 聚合酶、RNA 聚合酶和逆转录酶。

[0032] “聚合条件”指允许聚合酶将核苷酸掺入延伸的核酸链的条件。

[0033] “引物”指可通过聚合酶添加核苷酸的核酸。“添加”指通过聚合酶直接对引物添加核苷酸,以及随后对从引物起始的延伸的核酸链添加核苷酸。

[0034] “能够引发延伸的构型”指引物具有可用的反应基,聚合酶可对其添加核苷酸。

[0035] “固相支持物”指任意固相基质。

[0036] “特异性杂交复合物”指在严格杂交条件和/或严格洗涤条件下能够形成或基本维持的杂交复合物。

[0037] “靶核酸分子”指任意目的核酸分子。

[0038] “模板”指可被聚合酶利用以合成互补核酸链的单链核酸,或双链核酸的变性区。

[0039] II. 靶核酸分子的捕获和测序

[0040] 在一个方面,本发明涉及使用固定在固相支持物上的互补核酸,例如引物捕获靶核酸分子的方法。在某些实施方案中,然后通过检测通过使用靶核酸分子作为模板的聚合酶从引物起始的核酸聚合,测定靶核酸分子的核酸序列。检测在单分子水平上实时或近似实时进行。

[0041] 靶核酸分子

[0042] 在多个实施方案中,靶核酸分子是 DNA。在一个实施方案中,靶核酸分子可对应基因组,例如人基因组或来自任意其他生物的基因组的任意区域(“靶区域”)。靶区域可为几百万碱基(megabase)的一个或多个连续的区块,或若干较小的连续或不连续的区域,例如来自一个或多个染色体的所有外显子,或已知包含 SNP 的位点。包含靶区域的基因组可为部分或完整的。基因组可源自任意生物来源,例如患者样品或合并的患者样品;细胞系或细胞培养物;活检材料;正常组织样品或来自肿瘤或其他患病组织的样品;和本领域技术人员会考虑的其他生物来源。在一个实施方案中,通过例如超声处理或水动力将包含靶核酸的基因组 DNA 剪切为一般约 200-600 碱基对的片段,从片段或其分级部分中捕获靶核酸分子。在另一个实施方案中,靶核酸分子可为编码或非编码序列。在一个这样的实施方案中,靶核酸分子是外显子或其部分。

[0043] 在多个实施方案中,靶核酸分子是 RNA。在一个实施方案中,靶核酸是 mRNA 转录物或其部分。在一个这样的实施方案中,靶核酸是具有多聚 A 尾巴的 mRNA 转录物或其部分。多聚 A 尾巴的存在可允许与包含足够长度的多聚 T 序列(一般在引物的 3' 末端)的探针或引物杂交。在另一个实施方案中,靶核酸分子是来自 mRNA 的 cDNA,例如通过使用逆转录酶产生。

[0044] 捕获

[0045] 在多个实施方案中,从例如 RNA、DNA(例如基因组 DNA)或 cDNA 分子的核酸混合物中捕获靶核酸分子。在一个实施方案中,在捕获靶核酸分子前扩增混合物中的核酸。这可通过例如将包含通用引发位点的接头连接至混合物中的核酸分子的末端实现,其中所述末端可任选地在连接前进行末端修复。从而通用引物可用于扩增混合物中的核酸。

[0046] 在多个实施方案中,使用固定在固相支持物上的互补核酸捕获靶核酸分子。互补核酸不需要与靶核酸分子完全互补,而是可包含错配,只要靶核酸分子和互补核酸能够形成特异性杂交复合物即可。在一个实施方案中,互补核酸是引物。在一个这样的实施方案中,引物以能够引发延伸的构型固定在固相支持物上。例如,在固相支持物上固定具有 3' -OH 的引物,其中所述 3' -OH 可被聚合酶利用,用于在引物的 3' 末端添加核苷酸。这可例如通过引物的 5' 末端或通过引物的内部区域将引物固定在固相支持物上实现,只要引物是能够引发延伸的即可。引物可为任意长度,只要其能够与靶核酸分子形成特异性杂交复合物,在某些实施方案中,引物长度为至少 10, 15, 20, 25, 50, 75, 100, 200, 300, 400 或 500 碱基对。

[0047] 本领域熟知将核酸固定在固相支持物上的方法。例如,可通过共价或非共价连接将核酸,例如上面提供的互补核酸固定在固相支持物上。合适的化学接头和其他连接是本领域技术人员公知的。例如,在一个实施方案中,待固定的核酸是生物素化的(例如,包含

一个或多个生物素化核苷酸), 固相支持物在其表面上具有链霉抗生物素, 其中核酸的生物素部分与链霉抗生物素结合, 从而固定核酸。在另一个实施方案中, 通过在固相支持物上合成核酸实现上述实施方案中提供的固定。例如, 可通过从 5' 至 3' 的方向聚合核苷酸在固相支持物上合成引物, 在远离固相支持物的引物末端留下可用的 3'-OH。在固相支持物(例如高密度微阵列) 上从 5' 至 3' 的方向合成寡核苷酸的化学方法是本领域公知的, 并可用于本文描述的目的。见, 例如 Albert 等人 (2003) "Light directed 5' → 3' synthesis of complex oligonucleotide microarrays," *Nucleic Acids Res.* 31(7):e35, 以其整体引入本文作为参考。

[0048] 在多个实施方案中, 固相支持物是可固定核酸的任意基质。这样的基质包括但不限于玻璃(例如, 玻璃显微镜载玻片)、金属、陶瓷、聚合小珠和其他基质。在某些实施方案中, 固相支持物为阵列的形式, 例如, 微阵列。在某些实施方案中, 可将核酸固定在固相支持物(例如小珠) 上, 所述固相支持物进而被捕获或固定在另一种固相支持物(例如, 载玻片或微阵列) 上。

[0049] 在某些实施方案中, 在杂交条件下将其上固定了互补核酸的固相支持物暴露于包含靶核酸分子的核酸混合物。靶核酸分子从而与互补核酸形成特异性杂交复合物。在其他实施方案中, 洗涤固相支持物以去除未结合的和非特异性结合的核酸, 因此从混合物中的其他核酸中分离靶核酸分子(包含在特异性杂交复合物中)。在某些实施方案中, 分别在严格杂交条件和/或严格洗涤条件下将固相支持物暴露于核酸混合物和/或洗涤固相支持物。

[0050] 如本文使用的, "杂交" 指互补核酸链的配对。杂交和杂交强度(即核酸链之间结合的强度) 受这些因素的影响, 例如核酸之间的互补程度, 条件的严格性, 杂交复合物的 T_m , 和核酸的 G:C 比。尽管本发明不限于特定杂交条件的集合, 但可应用严格杂交条件。本领域技术人员可使用常规方法经验性确定严格杂交条件。严格杂交条件是序列依赖的, 并且也依赖于环境因素, 例如盐浓度和有机溶剂的存在。一般地, 选择严格杂交条件使得在确定的离子强度和 pH 下比特定核酸序列的热熔点 (T_m) 低约 5°C 至 20°C。在某些实施方案中, 严格杂交条件比与互补核酸结合的特定核酸的热熔点低约 5°C 至 10°C。 T_m 是 50% 的核酸(例如, 靶核酸分子) 与完全匹配的引物杂交的温度(在确定的离子强度和 pH 下)。

[0051] 类似地, 本领域技术人员可使用常规方法经验性确定严格洗涤条件。例如, 可确定允许非特异性结合的核酸与固相支持物(例如阵列) 上固定的特异性杂交复合物分离的严格洗涤条件。在一个实施方案中, 使阵列暴露于杂交条件(例如, 严格杂交条件), 然后用包含连续降低的盐浓度, 和/或升高的洗涤剂浓度, 和/或处于渐增的温度的缓冲液洗涤, 直到特异性与非特异性杂交的信噪比足够高, 以便于检测特异性杂交, 例如共有完全互补性或基本完全的互补性的核酸链之间的杂交。在某些实施方案中, 严格洗涤条件将包括约 30°C, 37°C, 42°C, 45°C, 50°C 或 55°C 的温度。在某些实施方案中, 严格洗涤条件将包括 $\leq 1M$, $\leq 500mM$, $\leq 250mM$, $\leq 100mM$, $\leq 50mM$ 或 $\leq 25mM$, 但 $\geq 10mM$ 的盐浓度。严格杂交条件的实例如下: 50% 甲酰胺, 5x SSC (0.75M NaCl, 0.075M 柠檬酸钠), 50mM 磷酸钠 (pH6.8), 0.1% 焦磷酸钠, 5x Denhardt 溶液, 超声处理的鲑精 DNA (50 $\mu g/ml$), 0.1% SDS 和 10% 硫酸葡聚糖, 在 42°C 下。严格洗涤条件的实例如下: 包含 EDTA 的 0.1x SSC, 在 55°C 下。

[0052] 序列测定

[0053] 在靶核酸分子（包含在特异性杂交复合物中）从混合物的其他核酸中分离后，可对靶核酸分子测序。在多个实施方案中，此步骤一般被称为“再测序”，其中来自参考基因组的靶区域的序列是已知的。再测序的现有方法要求从特异性杂交复合物中洗脱靶核酸分子，之后扩增靶核酸分子，然后使用“下一代”测序技术（例如，能够平行、高通量测序的测序平台）或再测序阵列（包含特异性检测在核酸的离散区段中存在突变与否的探针的微阵列）对扩增的靶核酸分子测序。对洗脱和扩增步骤的要求是费时费力的，且其可能导致样品的损失，或偏向展示靶核酸分子群中的个体靶核酸分子。

[0054] 因此，在多个实施方案中，进行靶核酸分子的再测序而无需将其从特异性杂交复合物中洗脱。在一个实施方案中，这可如下实现，其中包含在特异性杂交复合物中的互补核酸是以能够引发延伸的构型固定在固相基质上的引物。例如，引物可与靶核酸分子的特定区域杂交，靶核酸分子的其余部分为单链形式。因此，聚合酶将能够使用靶核酸分子的单链（未杂交）部分作为模板对引物的可用 3' -OH 添加核苷酸。使用通过聚合酶合成的核酸链测定靶核酸分子的序列。

[0055] 因此，可将通过引物在其上固定特异性杂交复合物的固相支持物暴露于聚合反应混合物。在一个实施方案中，聚合反应混合物包含聚合酶和核苷酸。聚合酶可为 DNA 聚合酶、RNA 聚合酶或逆转录酶。当靶核酸分子为 RNA 时，聚合酶可为逆转录酶。在其他实施方案中，当靶核酸分子为 DNA 时，聚合酶为 DNA 聚合酶。某些示例性 DNA 聚合酶包括但不限于，细菌 DNA 聚合酶（例如，大肠杆菌 DNA pol I, II, III, IV 和 V, 和 DNA pol I 的 Klenow 片段）；病毒 DNA 聚合酶（例如，T4 和 T7DNA 聚合酶）；古细菌 DNA 聚合酶（例如，栖热水生菌（*Thermus aquaticus*）(Taq) DNA 聚合酶，强烈火球菌（*Pyrococcus furiosus*）(Pfu) DNA 聚合酶，“Deep Vent” DNA 聚合酶（New England BioLabs））；真核 DNA 聚合酶；及其改造或修饰的变体。某些示例性 RNA 聚合酶包括但不限于，T7, T3 和 SP6RNA 聚合酶，及其改造或修饰的变体。某些示例性逆转录酶包括但不限于，来自 HIV, MMLV 和 AMV 的逆转录酶，以及市售的逆转录酶，例如 SUPERScript (Invitrogen, Carlsbad, CA)。

[0056] 在某些实施方案中，标记聚合反应混合物中的核苷酸和 / 或聚合酶。在一个实施方案中，差异标记 1、2、3 或 4 种核苷酸。在一个这样的实施方案中，用 4 种不同的标记物标记 4 种不同的核苷酸。例如，在 dNTP 的情况下，用不同的标记物，例如不同的荧光团分别标记腺嘌呤（或其功能等同类似物），鸟嘌呤（或其功能等同类似物），胞嘧啶（或其功能等同类似物）和胸腺嘧啶（或其功能等同类似物）。类似地，在 rNTP 的情况下，用不同的标记物，例如不同的荧光团分别标记腺嘌呤（或其功能等同类似物），鸟嘌呤（或其功能等同类似物），胞嘧啶（或其功能等同类似物）和尿嘧啶（或其功能等同类似物）。合适的标记物包括但不限于，发光、光致发光（photoluminescent）、电致发光、生物发光、化学发光、荧光和 / 或磷光标记物。荧光标记物包括但不限于，黄嘌呤染料、荧光素、青色素、罗丹明、香豆素、吖啶、德克萨斯红染料、BODIPY、ALEXA、GFP，及其修饰物。标记物可直接连接至核苷酸，或可通过合适的接头连接。标记物可在不显著干预聚合酶将核苷酸掺入延伸的核酸链的能力的任意位置连接至核苷酸。在一个实施方案中，标记物连接至核苷酸的磷酸上，例如，核苷酸的末端磷酸上，其中，核苷酸的磷酸链连带着标记物在核苷酸掺入延伸的核酸链时被切割。标记的核苷酸和 / 或聚合酶可允许检测聚合酶合成的核酸链。

[0057] 在某些实施方案中，通过鉴定经由聚合酶掺入延伸的核酸链中的核苷酸，测定靶

核酸分子的序列。一个实施方案包含直接或间接依掺入的顺序检测掺入延伸的核酸链中的核苷酸的标记物,将检测的标记物与核苷酸的身份相联系,由此确定延伸的核酸链的序列。这样测定靶核酸分子(或其互补物,取决于引物是与靶核酸分子的正义链还是反义链杂交)的序列。在这样的实施方案中,标记物的检测、和原则上靶核酸分子的测序实时地在“单分子”水平上进行。上文提到且将在下文进一步例证,可在核苷酸掺入延伸的核酸链时同时去除标记物,使得到的核酸链不被标记。

[0058] 在 W02010/002939 中描述了在核苷酸掺入延伸的核酸链时检测标记的核苷酸的某些方法。这样的方法依赖于当分子彼此足够接近时,“供体”分子(FRET 供体)和“受体”分子(FRET 受体)之间的 Forster 共振能量转移(Forster Resonance Energy Transfer, FRET)。在特定的实施方案中,用 FRET 供体荧光团标记聚合酶,用 FRET 受体荧光团标记核苷酸。当聚合酶将核苷酸掺入延伸的核酸链时,使 FRET 供体和受体荧光团靠近,允许能量从 FRET 供体荧光团至 FRET 受体荧光团转移。能量转移减少 FRET 供体荧光团的发射强度,增加 FRET 受体荧光团的发射强度。检测 FRET 受体的发射光谱指示掺入的核苷酸的身份。在一个实施方案中,连接至聚合酶的 FRET 供体是荧光纳米颗粒,例如,纳米晶体,更具体地,如在 W02010/002939 中描述的量子点(quantum dot)。可用激发源例如激光使 FRET 供体发光,其中产生供体发射。在其他实施方案中,将不同的 FRET 受体分别连接至一种或多种核苷酸,特别地分别连接至 3 种或 4 种核苷酸。FRET 受体可为上述任意荧光标记物。

[0059] 可使用任意合适方法或装置检测标记物,包括但不限于电荷耦合装置(charge couple device)和全内反射显微镜法。

[0060] 在某些实施方案中,靶核酸分子是多聚 A-RNA,使用含多聚 T 序列的引物捕获靶核酸分子。使用逆转录酶从引物合成 cDNA(但不测序)。然后在固相支持物上变性(分离)特异性杂交复合物中的靶核酸分子,留下新合成的 cDNA 链,其通过引物固定在固相支持物上。然后将固相支持物暴露于包含多聚 A 的引物,所述引物与新合成的 cDNA 杂交。通过将固相支持物暴露于如上提供的聚合反应混合物对新合成的 cDNA 测序,其中 DNA 聚合酶从多聚 A 引物合成核酸链。

[0061] 使用上述方法,通过选择特异性针对各个目的靶核酸分子的引物,和在固相支持物的分立区域,例如微阵列上固定引物,可分离和测序多个靶核酸分子。这样,可以高通量方式分离和测序目的靶核酸分子。

[0062] III. 甲基化状态的测定

[0063] 在另一个方面,本发明涉及测定基因组 DNA 片段中 CpG 二核苷酸的甲基化状态的方法,所述方法通过将基因组 DNA 片段固定于固相支持物;通过检测使用基因组 DNA 片段作为模板,通过聚合酶的核苷酸聚合,测定固定的基因组 DNA 片段的核酸序列;从新合成的核酸链中变性固定的基因组 DNA 片段;将固相支持物暴露于亚硫酸氢盐;和通过检测使用靶核酸分子作为模板,通过聚合酶的核苷酸聚合,测定基因组 DNA 片段的核酸序列。

[0064] 基因组 DNA 片段

[0065] 可从任意基因组,例如人基因组或来自任何其他生物的基因组获得基因组 DNA 片段。基因组可为部分或完整的。基因组可源自任意生物来源,例如患者样品或合并的患者样品;细胞系或细胞培养物;活检材料;正常组织样品或来自肿瘤或其他患病组织的样品;和本领域技术人员会考虑的其他生物来源。在一个实施方案中,通过例如超声处理或水动

力将基因组 DNA 剪切为一般约 200-600 碱基对的片段。在其他实施方案中,通过酶消化片段化基因组 DNA。

[0066] 基因组 DNA 片段的固定

[0067] 可通过任意不同方法将基因组 DNA 片段固定至固相支持物。可通过共价或非共价连接将基因组 DNA 片段直接或间接固定至固相支持物。合适的化学接头和其他连接是本领域技术人员公知的。在特定的实施方案中,基因组 DNA 片段是变性的,其以单链形式固定至固相支持物。通过连接核酸,或接头将基因组 DNA 片段,例如单链基因组 DNA 片段固定至固相支持物。例如,接头可与基因组 DNA 片段的一端或两端连接,接头固定至固相支持物。接头可为单链的,例如寡核苷酸。在某些实施方案中,接头首先与基因组 DNA 片段连接,然后将接头固定至固相支持物,或备选地,首先将接头固定至固相支持物,然后将基因组片段与固相支持物上的接头连接。在某些其他实施方案中,接头是生物素化的(例如,包含一个或多个生物素化核苷酸),固相支持物在其表面上具有链霉抗生物素,其中生物素部分与链霉抗生物素结合,从而固定接头。

[0068] 固定的基因组 DNA 片段的方向可为从固相支持物起 5' → 3',或从固相支持物起 3' → 5'。在一个实施方案中,固定的基因组 DNA 片段的方向为从固相支持物起 3' → 5'。在其他实施方案中,固定的基因组 DNA 片段是单链的。在另一个实施方案中,单链基因组片段通过接头,例如寡核苷酸固定至固相支持物。在特定实施方案中,基因组 DNA 片段是单链的并与固定至固相支持物的接头连接,其中接头和基因组 DNA 片段的方向为从固相支持物起 3' → 5'。

[0069] 测序

[0070] 在固相支持物上测序固定的基因组 DNA 片段。在某些实施方案中,基因组 DNA 片段以单链形式固定在固相支持物上,或以双链形式固定在固相支持物上,其中所述双链形式能够全部或部分转化为保持固定至固相支持物的单链形式。

[0071] 在某些实施方案中,在杂交条件下将固相支持物暴露于引物,其中引物和基因组 DNA 片段形成特异性杂交复合物。在某些其他实施方案中,在杂交条件下将固相支持物暴露于引物,其中引物和接头形成特异性杂交复合物。在一个这样的实施方案中,接头是连接基因组 DNA 片段的寡核苷酸,其中接头固定至固相支持物。在前述实施方案中,引物结合的基因组 DNA 片段或接头中的核酸序列中的胞嘧啶例如通过具有保护基受到保护,而免于由亚硫酸氢盐处理导致的脱氨。保护基可为甲基,例如,受保护的胞嘧啶可为 5-甲基胞嘧啶。

[0072] 在其他实施方案中,将固相支持物暴露于聚合反应混合物。在一个实施方案中,聚合反应混合物包含 DNA 聚合酶和核苷酸,其中 DNA 聚合酶使用基因组 DNA 片段作为模板,从引物开始合成核酸链。在一个这样的实施方案中, DNA 聚合酶从与接头形成特异性杂交复合物的引物开始合成核酸链,其中使用接头(任选地)和基因组 DNA 片段作为模板。例如,如果接头连接基因组 DNA 片段与固相支持物,其中接头和基因组 DNA 片段的方向为以从固相支持物起 3' → 5',那么引物可以从固相支持物起 5' → 3' 的方向与接头杂交,从而通过聚合酶以 5' → 3' 的方向引发合成,(任选地)使用接头作为模板和使用基因组 DNA 片段作为模板。在另一个这样的实施方案中, DNA 聚合酶从与基因组 DNA 片段形成特异性杂交复合物的引物开始合成核酸链,其中使用基因组 DNA 片段作为模板。

[0073] 合适的 DNA 聚合酶包括但不限于细菌 DNA 聚合酶(例如,大肠杆菌 DNA pol

I, II, III, IV 和 V, 和 DNA pol I 的 Klenow 片段); 病毒 DNA 聚合酶 (例如, T4 和 T7DNA 聚合酶); 古细菌 DNA 聚合酶 (例如, 栖热水生菌 (*Thermus aquaticus*) (Taq) DNA 聚合酶, 强烈火球菌 (*Pyrococcus furiosus*) (Pfu) DNA 聚合酶, “Deep Vent” DNA 聚合酶 (New England BioLabs)); 真核 DNA 聚合酶; 及其改造或修饰的变体。

[0074] 在某些实施方案中, 标记聚合反应混合物中的核苷酸和 / 或聚合酶。在一个实施方案中, 差异标记 1、2、3 或 4 种核苷酸。在一个这样的实施方案中, 用 4 种不同的标记物标记 4 种不同的核苷酸。例如, 在 dNTP 的情况下, 用不同的标记物, 例如不同的荧光团分别标记腺嘌呤 (或其功能等同类似物), 鸟嘌呤 (或其功能等同类似物), 胞嘧啶 (或其功能等同类似物) 和胸腺嘧啶 (或其功能等同类似物)。合适的标记物包括但不限于, 发光、光致发光、电致发光、生物发光、化学发光、荧光和 / 或磷光标记物。荧光标记物包括但不限于, 黄嘌呤染料、荧光素、青色素、罗丹明、香豆素、吖啶、德克萨斯红染料、BODIPY、ALEXA、GFP, 及其修饰物。标记物可直接连接至核苷酸, 或可通过合适的接头连接。标记物可在不显著干预聚合酶将核苷酸掺入延伸的核酸链的能力的任意位置连接至核苷酸。在一个实施方案中, 标记物连接至核苷酸的磷酸上, 例如, 核苷酸的末端磷酸上, 其中, 核苷酸的磷酸链连带标记物在核苷酸掺入延伸的核酸链时被切割。标记的核苷酸和 / 或聚合酶可允许检测聚合酶合成的核酸链。

[0075] 在某些实施方案中, 通过依掺入的顺序鉴定经聚合酶掺入延伸的核酸链中的核苷酸, 测定基因组 DNA 片段的序列。一个实施方案包含依掺入的顺序直接或间接检测掺入延伸的核酸链中的核苷酸的标记物, 将检测的标记物与核苷酸的身份相联系, 由此确定延伸的核酸链的序列。这样确定基因组 DNA 片段 (或其互补物, 取决于固定的是基因组 DNA 片段的正义还是反义链) 的序列。在这样的实施方案中, 标记物的检测, 和原则上基因组 DNA 片段的测序实时地在“单分子”水平上进行。上文提到且将在下文进一步例证, 可在核苷酸掺入延伸的核酸链的同时去除标记物, 使新合成的核酸链不被标记。

[0076] 在 W02010/002939 中描述了在核苷酸掺入延伸的核酸链时检测标记的核苷酸的某些方法。这样的方法依赖于当分子彼此足够接近时, “供体” 分子 (FRET 供体) 和 “受体” 分子 (FRET 受体) 之间的 Forster 共振能量转移 (FRET)。在特定的实施方案中, 用 FRET 供体荧光团标记聚合酶, 用 FRET 受体荧光团标记核苷酸。当聚合酶将核苷酸掺入延伸的核酸链时, 使 FRET 供体和受体荧光团靠近, 允许能量从 FRET 供体荧光团转移至 FRET 受体荧光团。能量转移减少 FRET 供体荧光团的发射强度, 增加 FRET 受体荧光团的发射强度。检测 FRET 受体的发射光谱指示掺入的核苷酸的身份。在一个实施方案中, 连接至聚合酶的 FRET 供体是荧光纳米颗粒, 例如, 纳米晶体, 更具体地, 如在 W02010/002939 中描述的量子点。可用激发源例如激光激发 FRET 供体, 其中产生供体发射。在其他实施方案中, 将不同的 FRET 受体分别连接至一种或多种核苷酸, 特别地分别连接至 3 种或 4 种核苷酸。FRET 受体可为上述任意荧光标记物。

[0077] 可使用任意合适方法或装置检测标记物, 包括但不限于电荷耦合装置和全内反射显微镜法。

[0078] 使用上述方法, 通过在固相支持物的分立区域, 例如微阵列上固定基因组 DNA 片段, 可分离和测序多个基因组 DNA 片段。这样, 可以高通量方式固定和测序目的基因组 DNA 片段。

[0079] 亚硫酸氢盐处理和处理后测序

[0080] 在基因组 DNA 片段的测序后,变性(例如,通过热或碱变性)由新合成的 DNA 链和基因组 DNA 片段组成的特异性杂交复合物,分离新合成的 DNA 链和基因组 DNA 片段。对固相支持物上的基因组 DNA 片段进行亚硫酸氢盐处理。进行亚硫酸氢盐处理的方法是本领域已知的,在例如 Herman 等人 (1996)Proc. Natl. Acad. Sci. USA93:9821-9826 中描述。在基因组 DNA 片段中未受保护(未甲基化)的胞嘧啶因此被转化为尿嘧啶。然后对基因组 DNA 片段进行如上文概述的测序步骤。将得到的序列与亚硫酸氢盐处理前得到的序列比较,以鉴定在基因组 DNA 片段中已转化为尿嘧啶残基的胞嘧啶残基,例如,通过在测序步骤中产生的新合成链中替换鸟嘌呤的胸腺嘧啶的存在指示。转化的残基指示基因组 DNA 片段中的未甲基化状态,而未转化的残基指示基因组 DNA 片段中的受保护,即甲基化状态。这样确定基因组 DNA 片段的甲基化状态。

实施例**[0081] [a]DNA 分子的直接捕获和实时测序**

[0082] 将包含目的靶核酸分子的 DNA(例如,具有目的蛋白质编码区的基因组 DNA) 剪切为合适大小,如在图 1B 和 2B 中所示。使用包含与目的区域互补的寡核苷酸的基质,例如,阵列(如图 1A 中所示)或小珠(如图 2A 中所示)。目的区域可为,例如外显子。通过与寡核苷酸的杂交在基质上捕获包含来自剪切的 DNA 的相关区域的核酸。在图 1 中,通过在阵列上与互补寡核苷酸的杂交捕获包含来自剪切的 DNA 的相关区域的核酸。在图 2 中,于溶液中小珠上捕获包含来自剪切的 DNA 的相关区域的核酸。洗掉未结合的和非特异结合的 DNA。在图 2C 中,将小珠置于另一种基质,例如有序或无序的阵列上。对捕获的 DNA 进行直接测序,如在图 1C 和 2C 中所示。这是通过使用寡核苷酸(其作为引物)的游离 3' 端,允许在单分子水平上在 DNA 的实时合成期间直接监测加入的碱基和聚合酶的标记的聚合酶和标记的 dNTP 实现的。

[0083] [b]RNA 分子的直接捕获和实时测序

[0084] 可使用包含多聚 A 尾巴的 RNA,如在图 3B 和 4B 中所示。同时使用包含含有多聚 dT 的寡核苷酸的基质,例如,阵列(如图 3A 中所示)或小珠(如图 4A 中所示)。通过多聚 A 尾巴与含有多聚 dT 的寡核苷酸的杂交在基质上捕获 RNA,如图 3C 和 4C 中所示。洗掉未结合的和非特异结合的 RNA,使用多聚 dT 的游离 3' 端和逆转录酶将捕获的 RNA 转化为 cDNA。RNA 转化为 cDNA 后,测序 cDNA。在测序 cDNA 的一个实施方案中,将寡核苷酸接头与新合成的 cDNA 的游离 3' 端连接,然后使引物与接头退火。使用引物,允许在单分子水平上在 DNA 的实时合成期间直接监测加入的碱基和聚合酶的标记的聚合酶和标记的 dNTP 对 cDNA 测序。可选地,可使用允许在单分子水平上在 cDNA 的实时合成期间直接监测加入的碱基和逆转录酶的标记的逆转录酶和标记的 dNTP 对捕获的 RNA 直接测序。(见图 3C 和 4C。)

[0085] [c] 通过相同模板的循环测序测定甲基化状态

[0086] 为测定 DNA(例如,基因组 DNA) 的甲基化状态,首先将其剪切为合适大小,并优选地转化为单链,如在图 5B 和 6B 中所示。使用包含甲基化寡核苷酸(即含有甲基化胞嘧啶的寡核苷酸)的基质,例如,阵列(如图 5A 中所示)或小珠(如图 6A 中所示)。将剪切的 DNA 与甲基化寡核苷酸连接,如在图 5C 和 6C 中所示。然后使用与阵列或小珠上的甲基化寡

核苷酸互补的引物,允许在单分子水平上在 DNA 的实时合成期间直接监测加入的碱基和聚合酶的标记的聚合酶和标记的 dNTP 对连接的 DNA 测序。在对连接的 DNA 测序后,去除新合成的链,用亚硫酸氢盐处理原始 DNA 以允许将甲基化胞嘧啶转化为尿嘧啶,如在图 5D 和 6D 中所示。使用引物,允许在单分子水平上在 DNA 的实时合成期间直接监测加入的碱基和聚合酶的标记的聚合酶和标记的 dNTP 对处理的 DNA 测序。比较相同分子在用亚硫酸氢盐处理前和后得到的序列将允许测定 DNA 的甲基化状态。

[0087] 引入本文引用的所有专利和出版物作为参考。

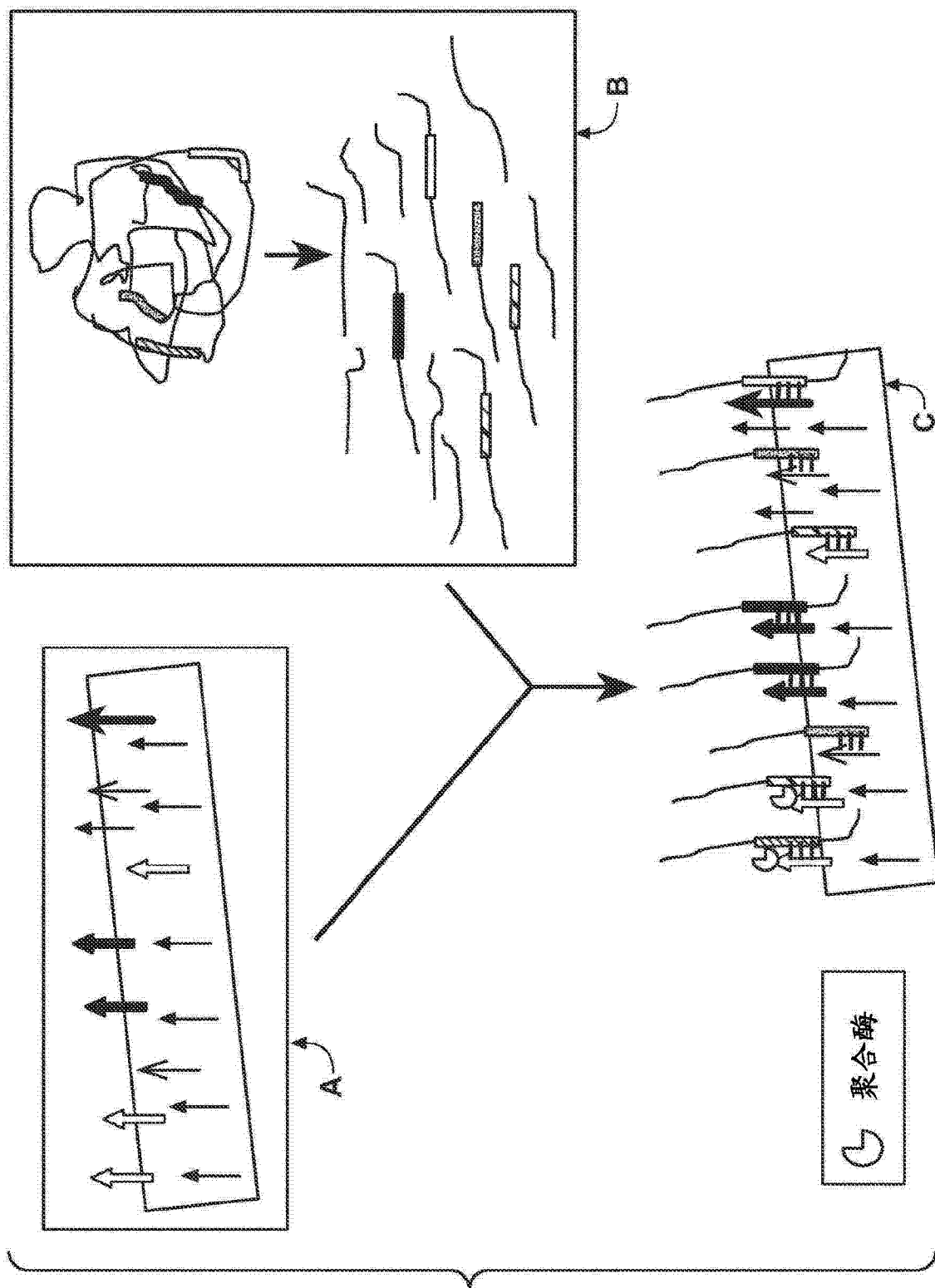


图 1

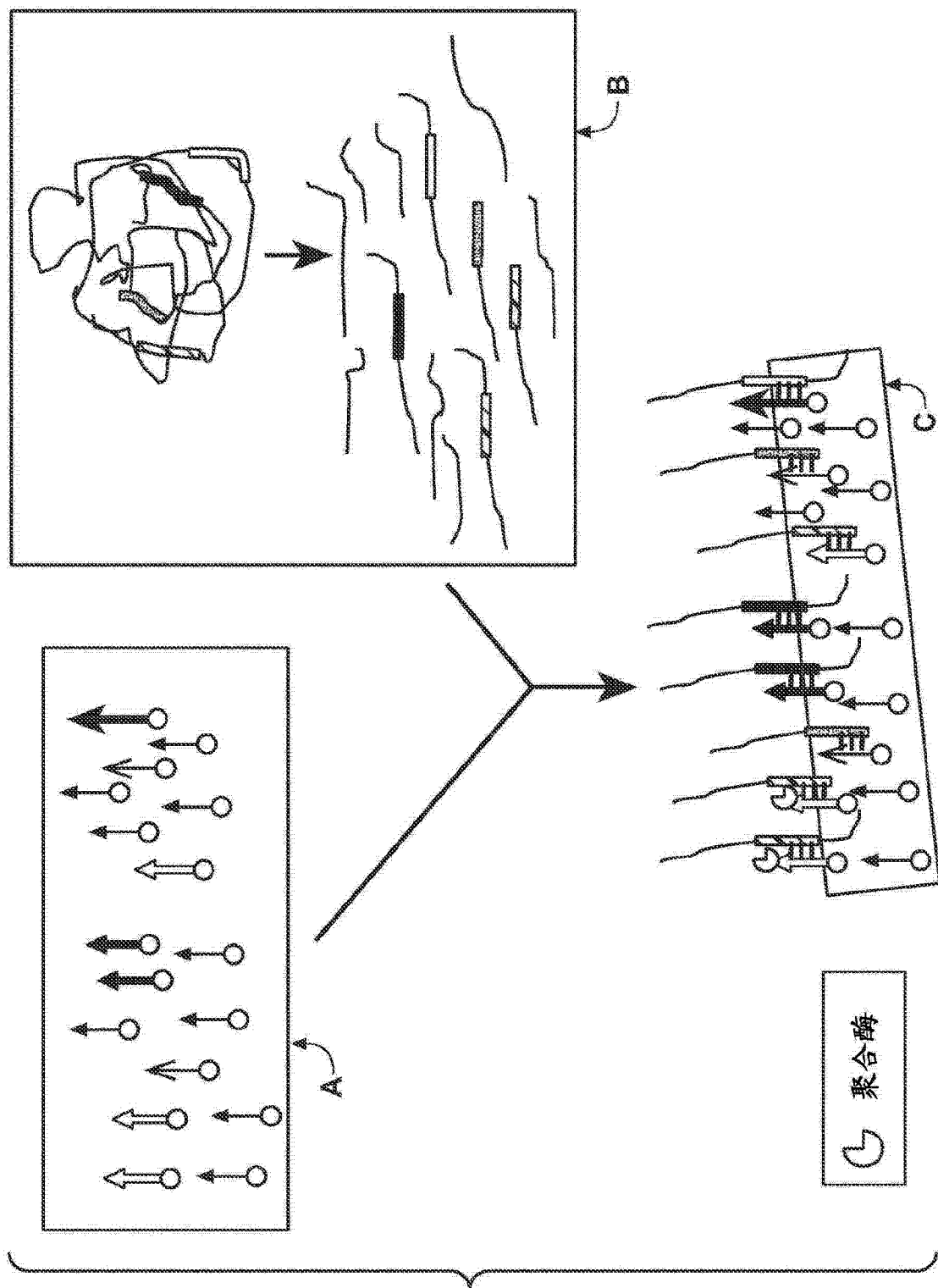


图 2

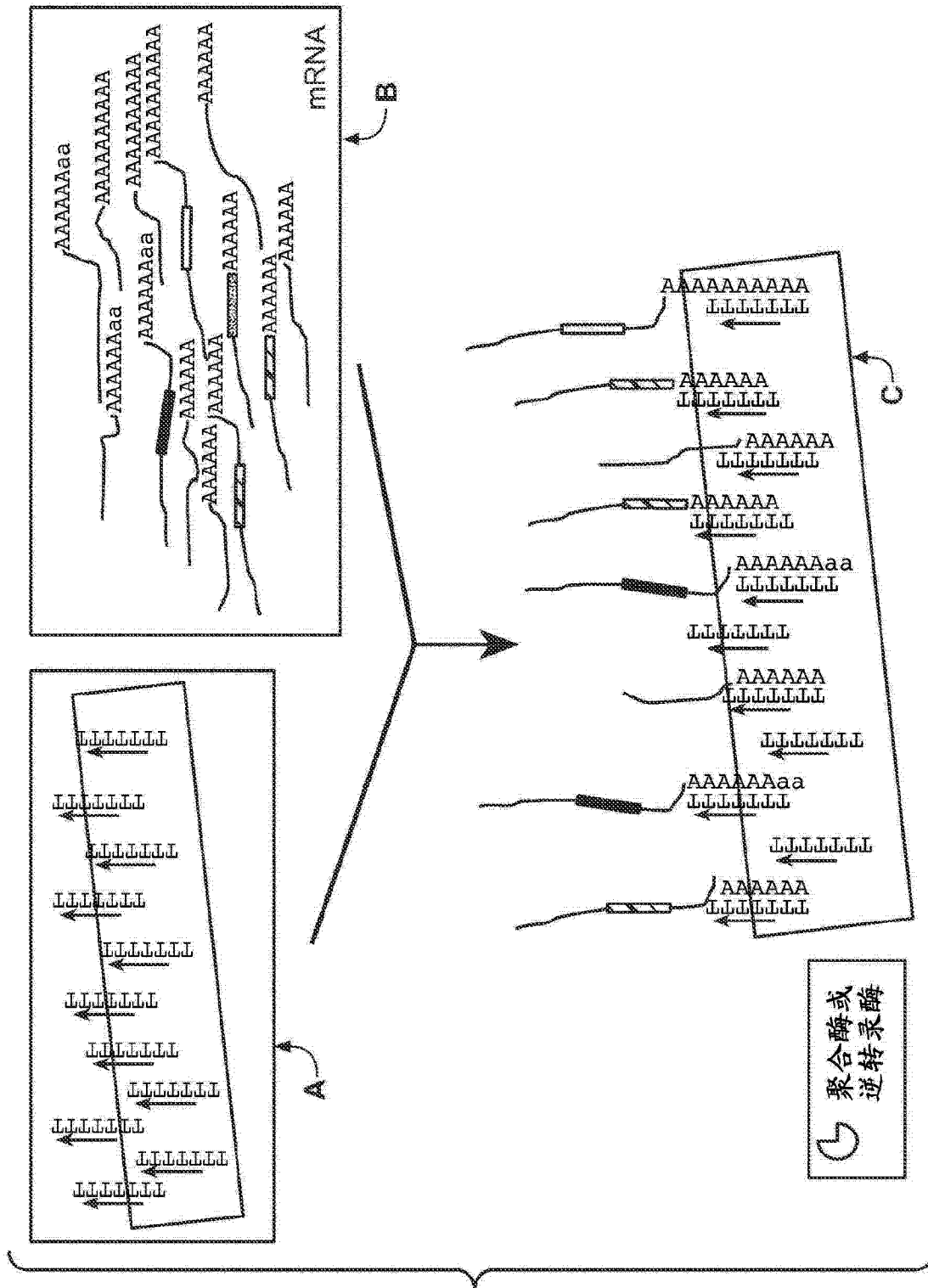


图 3

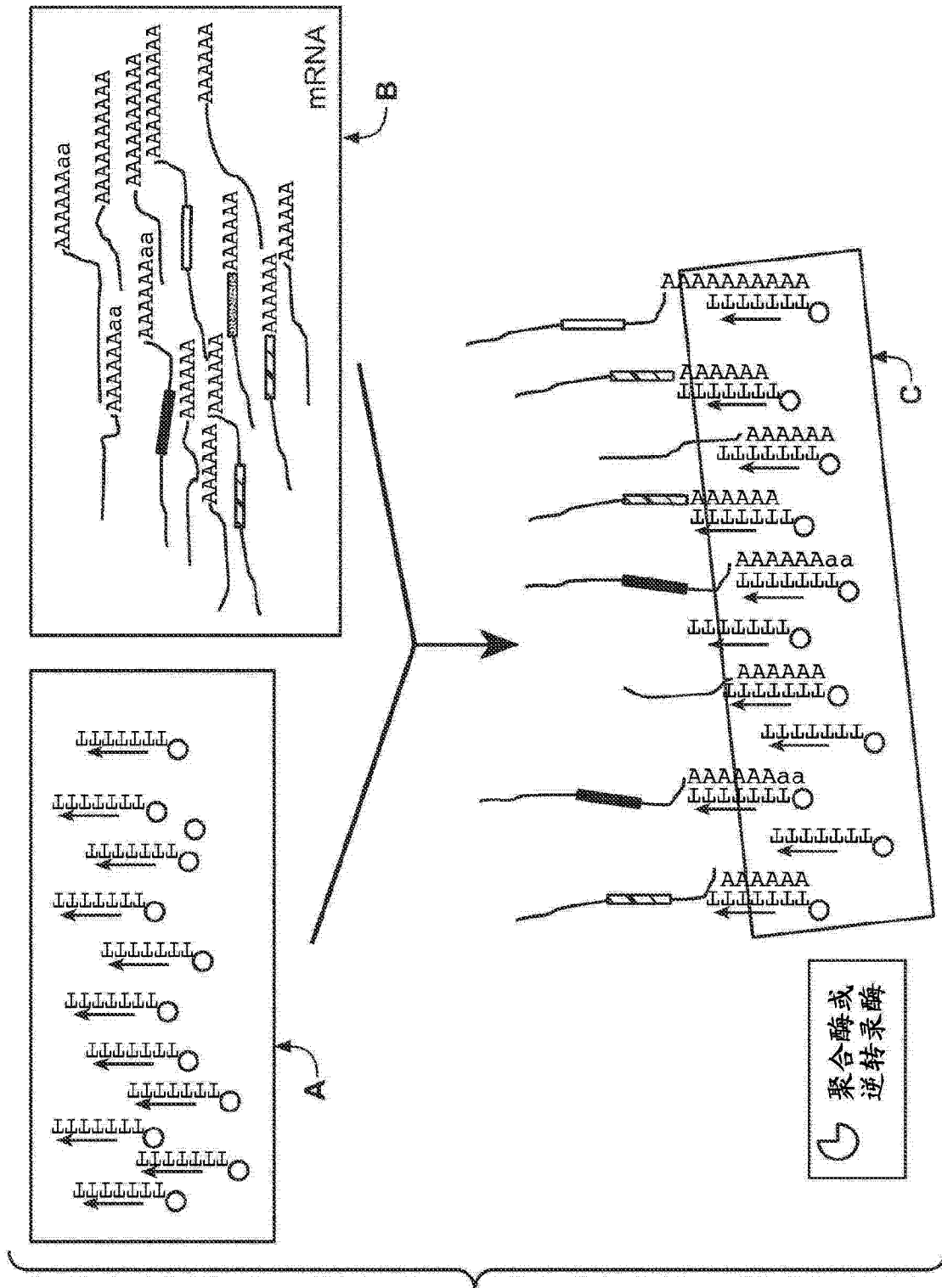


图 4

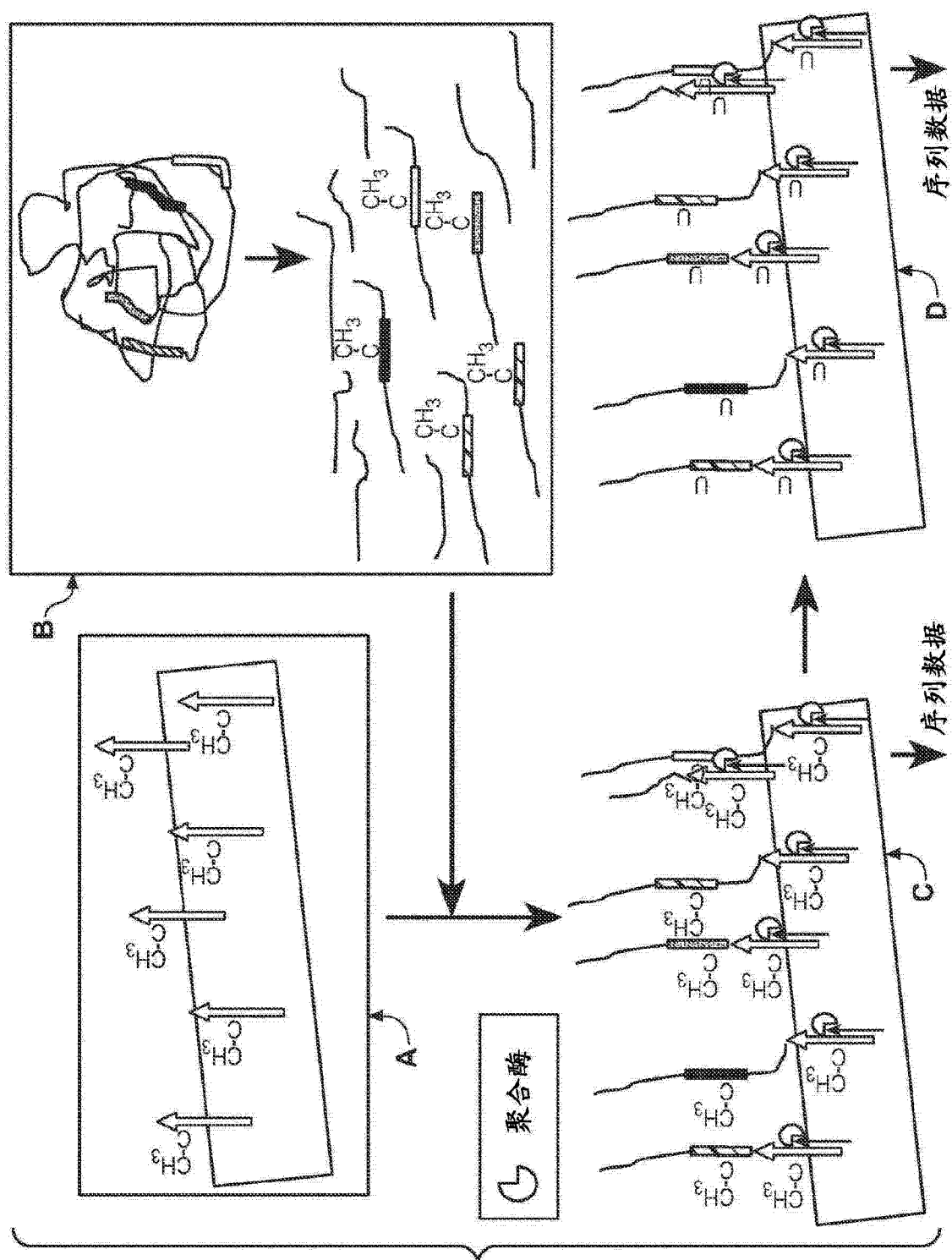


图 5

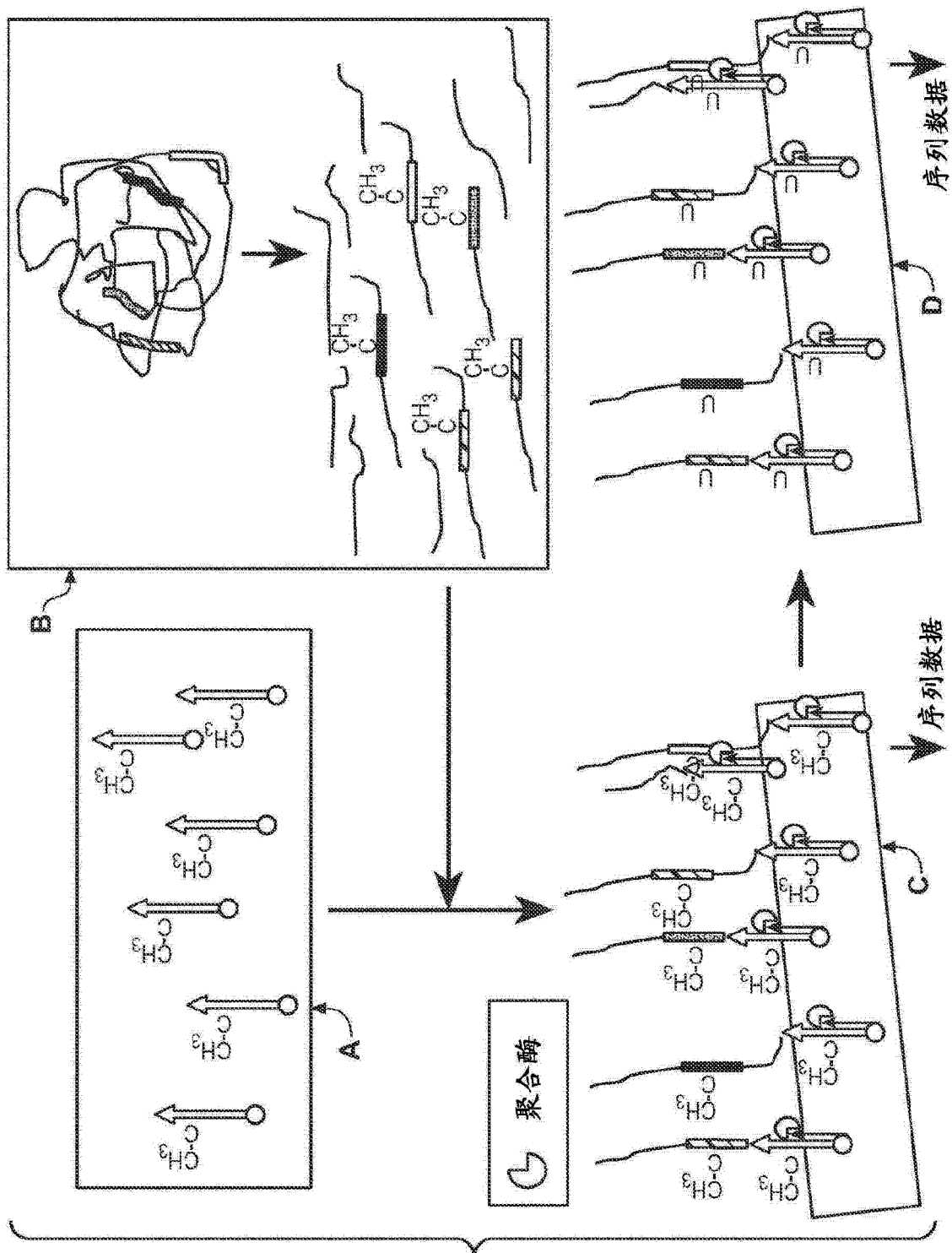


图 6