



(12) 发明专利申请

(10) 申请公布号 CN 101889074 A

(43) 申请公布日 2010. 11. 17

(21) 申请号 200880119334. 3

(22) 申请日 2008. 10. 06

(30) 优先权数据

60/997, 427 2007. 10. 04 US

61/132, 960 2008. 06. 23 US

(85) PCT申请进入国家阶段日

2010. 06. 04

(86) PCT申请的申请数据

PCT/US2008/078986 2008. 10. 06

(87) PCT申请的公布数据

W02009/046445 EN 2009. 04. 09

(71) 申请人 哈尔西恩莫尔丘勒公司

地址 美国加利福尼亚州

(72) 发明人 威廉·安德雷格 迈克尔·安德雷格

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 刘蕾 沙捷

(51) Int. Cl.

C12M 1/00 (2006. 01)

G01N 23/04 (2006. 01)

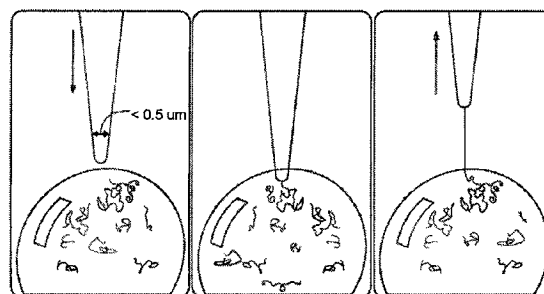
权利要求书 5 页 说明书 23 页 附图 16 页

(54) 发明名称

采用电子显微镜对核酸聚合物测序

(57) 摘要

本发明涉及使用电子显微镜通过直接检视来对标记的拉伸的 DNA 进行测序。该方法比目前的 DNA 测序方法具有更高的准确性、更低的成本和更长的阅读长度。



1. 一种获取 DNA 聚合链的序列信息的方法,所述方法包括以下步骤:
 - a) 提供包含所述 DNA 聚合链的溶液,其中已对所述 DNA 聚合链进行处理从而使得多个 DNA 碱基已用具有碱基特异性或碱基选择性的造影剂进行标记;
 - b) 将 DNA 结合工具引入所述溶液并使所述 DNA 聚合物的一部分与该工具结合;
 - c) 将所述工具从溶液中除去;
 - d) 将标记的 DNA 聚合链在空间内拉伸从而使该标记的 DNA 聚合链悬挂于空气 / 溶剂界面与所述工具之间;
 - e) 将经拉伸的标记 DNA 聚合链沉积于基片上;
 - f) 使用电子显微镜使标记的 DNA 聚合链成像由此确定经标记的碱基和未经标记的碱基的位置;和
 - g) 将经标记的碱基和未经标记的碱基的位置与 DNA 聚合物的序列进行相关。
2. 如权利要求 1 所述的方法,其中所述经拉伸的标记 DNA 聚合链为单链。
3. 如权利要求 1 所述的方法,所述方法还包括使所述 DNA 聚合链变性以便在用造影剂进行标记之前产生单链核酸的步骤。
4. 如权利要求 3 所述的方法,其中所述变性步骤是热变性。
5. 如权利要求 1 所述的方法,所述方法还包括在所述成像步骤之前将碳层沉积于 DNA 聚合链上方的步骤。
6. 如权利要求 1 所述的方法,其中所述 DNA 聚合链是大于约 100kb 的高分子量 DNA。
7. 如权利要求 1 所述的方法,其中所述拉伸步骤在 DNA 聚合链中产生恒定的碱基与碱基间距。
8. 如权利要求 7 所述的方法,其中所述碱基与碱基间距为约 3Å~约 7Å。
9. 如权利要求 8 所述的方法,其中所述碱基与碱基间距为约 5Å。
10. 如权利要求 1 所述的方法,其中所述处于空气 / 溶剂界面与工具之间的 DNA 聚合链的悬挂区的长度至少为约 2 μm。
11. 如权利要求 1 所述的方法,其中仅对所述 DNA 聚合链的碱基的子集进行标记。
12. 如权利要求 11 所述的方法,其中所述碱基子集包括胸腺嘧啶和胞嘧啶。
13. 如权利要求 11 所述的方法,其中所述碱基子集包括腺嘌呤和鸟嘌呤。
14. 如权利要求 1 所述的方法,其中所述引入步骤中的工具是具有采用与 DNA 链结合的第一涂层进行功能化的尖端的针。
15. 如权利要求 14 所述的方法,其中所述针由选自玻璃、金、钨、PMMA、聚苯乙烯、PVC 和硅的材料构成。
16. 如权利要求 14 所述的方法,其中所述第一涂层是选自 PMMA、聚苯乙烯、PVB 和寡核苷酸的化合物。
17. 如权利要求 14 所述的方法,其中采用不与 DNA 结合的第二涂层对所述针进行功能化。
18. 如权利要求 17 所述的方法,其中所述第二涂层是选自辛基硫醇、己基硫醇、壬基硫醇、癸基硫醇和庚基硫醇的化合物。
19. 如权利要求 14 所述的方法,其中所述针尖的曲率半径小于约 200nm。
20. 如权利要求 14 所述的方法,其中使所述针以约 1 纳米 / 秒~约 100 米 / 秒的速率

移进和移出溶液。

21. 如权利要求 14 所述的方法,其中所述针以约 $1\text{ \AA} \sim 20\text{ }\mu\text{m}$ 的深度进入溶液。

22. 如权利要求 1 所述的方法,其中所述造影剂包含高 Z 原子。

23. 如权利要求 22 所述的方法,其中所述造影剂是 Os-bipy、乙酸汞或二甲基硫化铂。

24. 如权利要求 1 所述的方法,其中所述造影剂是高 Z 原子簇标记。

25. 如权利要求 1 所述的方法,其中所述沉积步骤包括通过采用搁板拉线将标记的 DNA 聚合链附着于成像基片。

26. 如权利要求 1 所述的方法,其中所述沉积步骤包括通过采用搁板拉线将标记的 DNA 聚合链附着于载体基片,其后通过采用转印将该载体基片上的标记的 DNA 聚合链转移至成像基片。

27. 如权利要求 1 所述的方法,其中所述沉积步骤包括通过采用间隙拉线将标记的 DNA 聚合链附着于载体基片。

28. 如权利要求 1 所述的方法,其中所述沉积步骤包括通过采用间隙拉线将标记的 DNA 聚合链附着于载体基片,其后通过采用刷印将该载体基片上的标记的 DNA 聚合链转移至成像基片。

29. 如权利要求 1 所述的方法,其中将多个单独的 DNA 聚合链在空间内拉伸并沉积于基片上然后将每个链成像。

30. 如权利要求 29 所述的方法,其中将多个 DNA 聚合链作为平行的链阵列安置于基片上。

31. 如权利要求 1 所述的方法,所述方法还包括以下步骤:

在步骤(a)之前,将包含来自同一个体的多个 DNA 聚合链的溶液分为至少两个子样,并对所述子样进行处理从而以不同的碱基特异性或选择性来标记 DNA 聚合链,和

在步骤(g)之后,汇集从所述成像步骤获得的多个 DNA 聚合链的数据以获取序列信息。

32. 一种获取核酸序列信息的方法,所述方法包括通过电子显微镜确定多个标记的核酸链中的每一个的至少 1000 个毗连碱基区域内经标记的碱基和未经标记的碱基的位置。

33. 如权利要求 32 所述的方法,其中针对至少 1000 个毗连碱基确定位置。

34. 如权利要求 32 所述的方法,其中针对至少 10,000 个毗连碱基确定位置。

35. 如权利要求 32 所述的方法,其中确定至少 20 个个体链的经标记的碱基和未经标记的碱基的位置。

36. 如权利要求 32 所述的方法,其中确定至少 100 个个体链的经标记的碱基和未经标记的碱基的位置。

37. 如权利要求 32 所述的方法,其中确定至少 1000 个个体链的经标记的碱基和未经标记的碱基的位置。

38. 如权利要求 32 所述的方法,其中确定至少 10,000 个个体链的经标记的碱基和未经标记的碱基的位置。

39. 如权利要求 35 所述的方法,其中针对经不同标记的核酸聚合链确定核酸链的至少 200 个毗连碱基的位置。

40. 如权利要求 35 所述的方法,其中针对经不同标记的核酸聚合链确定核酸链的至少 500 个毗连碱基的位置。

41. 如权利要求 35 所述的方法,其中针对经不同标记的核酸聚合链确定核酸链的至少 10,000 个毗连碱基的位置。

42. 如权利要求 32 所述的方法,其中所述核酸链在伸长时具有至少约 100 μm 的长度。

43. 如权利要求 32 所述的方法,其中所述核酸链是 DNA 序列。

44. 如权利要求 43 所述的方法,其中所述 DNA 是单链的。

45. 如权利要求 32 所述的方法,其中所述核酸在碱基之间的碱基与碱基间距为约 3 $\text{\AA}$ ~ 约 7 $\text{\AA}$ 。

46. 如权利要求 45 所述的方法,其中所述核酸的其中所述核酸在碱基之间的碱基与碱基间距为约 5 $\text{\AA}$ 。

47. 一种将至少一个核酸链受控地安置于基片上的方法,所述方法包括以下步骤:

提供含有多个核酸链的溶液;

将针尖插入溶液;

将针尖从含有多个核酸链的溶液中提出,其中已用与核酸结合的第一涂层将所述针尖功能化;

将核酸链在空白空间内拉伸从而使单个核酸链悬挂于空气 / 溶剂界面与每个针尖之间;

将至少一个拉伸的核酸链附着于基片。

48. 如权利要求 47 所述的方法,其中所述浸提步骤中的针由选自玻璃、金、钨、PMMA、聚苯乙烯、PVC 和硅的材料制造。

49. 如权利要求 47 所述的方法,其中所述第一涂层是选自 PMMA、聚苯乙烯、PVB 和寡核苷酸的化合物。

50. 如权利要求 47 所述的方法,其中用不与核酸结合的第二涂层将所述针功能化。

51. 如权利要求 50 所述的方法,其中所述第二涂层是选自辛基硫醇、己基硫醇、壬基硫醇、癸基硫醇和庚基硫醇的化合物。

52. 如权利要求 47 所述的方法,其中所述浸提步骤中的针尖的曲率半径小于约 200nm。

53. 如权利要求 47 所述的方法,其中使所述浸提步骤中的针尖以约 1 纳米 / 秒 ~ 约 100 纳米 / 秒的速率移进和移出溶液。

54. 如权利要求 47 所述的方法,其中所述浸提步骤中的针尖以约 10nm ~ 20nm 的深度进入溶液。

55. 如权利要求 47 所述的方法,其中将所述浸提步骤中的针置于单纳米定位器驱动的载体上。

56. 如权利要求 47 所述的方法,其中利用纳米定位器驱动的载体将多个针同时插入溶液。

57. 如权利要求 47 所述的方法,其中所述核酸是高分子量 DNA。

58. 如权利要求 47 所述的方法,其中所述拉伸步骤在核酸链中产生恒定的核酸与核酸间距。

59. 如权利要求 58 所述的方法,其中所述碱基与碱基间距为约 3 $\text{\AA}$ ~ 约 7 $\text{\AA}$ 。

60. 如权利要求 59 所述的方法,其中所述碱基与碱基间距和标记与标记间距为约 5 $\text{\AA}$ 。

61. 如权利要求 47 所述的方法,其中将所述拉伸步骤中的核酸拉伸至约 100 μm 的长

度。

62. 如权利要求 47 所述的方法,其中使附着于基片的至少两个核酸链定向来彼此基本平行。

63. 如权利要求 47 所述的方法,其中所述附着步骤包括采用搁板拉线来将所述至少一个核酸链附着于基片。

64. 如权利要求 47 所述的方法,其中所述附着步骤包括采用间隙拉线来将所述至少一个核酸链附着于基片。

65. 如权利要求 47 所述的方法,其中所述附着步骤中的基片是成像基片。

66. 如权利要求 47 所述的方法,其中所述附着步骤中的基片是载体基片。

67. 一种制造品,所述制造品包含:

包含多个核酸聚合链的液体;

工具;和

具有处于所述液体中的第一末端和与所述工具相连的第二末端的单个核酸聚合链,其中所述单个核酸聚合链的至少一部分悬挂于所述工具与所述液体之间的空间中。

68. 如权利要求 67 所述的制造品,其中将所述核酸聚合链的碱基伸长从而令碱基具有恒定的碱基与碱基间距。

69. 如权利要求 68 所述的制造品,其中碱基之间的碱基与碱基间距为约 $3\text{\AA}$ ~约 $7\text{\AA}$ 。

70. 如权利要求 68 所述的制造品,其中将核酸聚合链伸长以使其在悬挂区域中是线性的。

71. 如权利要求 67 所述的制造品,其中在空间中伸长的核酸链部分的长度至少为约 $100\text{ }\mu\text{m}$ 。

72. 一种制造品,所述制造品包含:

平面基片;和

位于所述平面基片上的至少一个伸长核酸聚合链,其中所述至少一个伸长核酸聚合链在约 1000 个碱基对的长度上具有恒定的碱基与碱基间距。

73. 如权利要求 72 所述的制造品,其中所述基片选自 PDMS 块、PMAA 块、薄膜、超平硅网、成像基片、载体基片和 TEM 网。

74. 如权利要求 73 所述的制造品,所述超平硅网覆盖有碳膜、多孔网膜或连续碳膜。

75. 如权利要求 72 所述的制造品,所述制造品还包含安置于所述至少一个伸长核酸聚合物上方的膜,从而使所述至少一个伸长核酸聚合物夹在所述平面基片与所述膜之间。

76. 如权利要求 75 所述的制造品,其中所述膜由碳、铍和低 Z 元素构成。

77. 如权利要求 72 所述的制造品,其中所述至少一个伸长核酸链的碱基与碱基间距为约 $3\text{\AA}$ ~约 $7\text{\AA}$ 。

78. 如权利要求 72 所述的制造品,其中所述平面基片由选自 PDMS、碳、硼、锂、铍、铝、氮化物、氮氧化物及其组合的材料构成。

79. 如权利要求 72 所述的制造品,其中将所述至少一个伸长核酸链用至少一个高 Z 标记化合物进行标记。

80. 如权利要求 79 所述的制造品,其中所述高 Z 原子选自由钨、汞和铂。

81. 如权利要求 72 所述的制造品,所述制造品包含多个置于基片上的伸长核酸链,其

中所述核酸链彼此基本平行。

82. 如权利要求 81 所述的制造品,所述多个伸长核酸链包约 1×10^6 个核酸链。

83. 如权利要求 72 所述的制造品,所述将至少一个伸长核酸链拉伸至至少约 $2 \mu\text{m}$ 的长度。

84. 一种包括对储存于存储器上的核酸序列进行分析的方法,其中所述序列通过权利要求 1 所述的方法确定。

85. 如权利要求 84 所述的方法,其中所述核酸序列是人类受试对象的基因组序列。

86. 如权利要求 85 所述的方法,其中所述分析包括对一个或多个单核苷酸多态性中的至少一个的存在或缺失、拷贝数、变异体、得失位、重排或全基因组比较进行确定。

87. 如权利要求 84 所述的方法,其中所述存储器是选自硬盘或软盘、光学介质、光盘(CD)、数字视频光盘(DVD)、半导体介质和闪存的磁性介质。

88. 一种针,所述针包含:

远端;

近端;和

在与所述远端和所述近端接触的液体内延伸的轴,

其中所述近端包含曲率半径小于约 200nm 的尖端部件,且其中已用与核酸末端结合的化合物将所述尖端部件功能化。

89. 如权利要求 88 所述的针,所述针涂覆有不具有与核酸结合的亲和力的第二化合物。

90. 如权利要求 88 所述的针,其中用选自 PMMA、聚苯乙烯、PVB、硅烷化物和寡核苷酸的化合物将所述尖端部件功能化。

91. 如权利要求 88 所述的针,其中所述核酸是单链 DNA。

92. 如权利要求 88 所述的针,其中所述针由选自玻璃、金、钨、PMMA、聚苯乙烯、PVC 和硅的化合物构成。

93. 如权利要求 88 所述的针,其中将所述针置于单纳米定位器驱动的载体上。

采用电子显微镜对核酸聚合物测序

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2007 年 10 月 4 日提交的美国临时申请第 60/997,427 号和于 2008 年 6 月 23 日提交的美国临时申请第 61/132,960 号的优先权,本文通过参考特地并入其全部公开内容。

技术领域

[0003] 本发明涉及核酸测序方法。

背景技术

[0004] 目前,DNA 测序大多通过 Sanger 法和其它合成测序方法来完成。这些方法的缺点在于成本高、阅读长度短和处理量不足。

[0005] 已经对电子显微镜测序进行了探索。电子显微镜测序的想法并不新。它在 DNA 结构被发现仅 6 年后就由 Richard Feynman 提出。然而,尚未将其成功用以产生有意义的序列信息。

[0006] 透射电子显微镜 (TEM) 通过将电子束穿过样品送至检测器或屏幕上而工作。样品中的部分阻碍或反射电子束,从而到达检测器的电子图案形成图像。在多数情形中,低原子数 (Z) 的原子产生极小的反差并且在电子显微镜中基本不可见。包含低 Z 的氢、碳、氮、氧和磷原子的普通 DNA 在电子显微镜下几乎不显示出反差,并且几乎不可能在承载背景下显现。为了使用现有电子显微技术使 DNA 可视化,可以用高 Z 原子标记碱基或使得其可由 TEM 检测。

[0007] 最初的严肃的电子显微镜测序研究由 Beer, M. 和 Moudrianakis, E. N., 48 (3) PNAS 409-416 (1962) 进行。他们最初研究集中于 DNA 的重原子标记和试图使重原子在电子显微镜下可视化。后来的研究且可能是迄今最佳的其它研究由 Whiting 和 Ottensmeyer 进行。据 Ottensmeyer 报道:“开发了腺嘌呤和鸟嘌呤的重原子标记物并在模型序列上进行了尝试,结果仅表明虽然电子显微镜和化学不再是主要障碍,但样本制备仍然是主要障碍。标记的单链核酸聚合物在样本载体上的不可控安置造成相当不规则的碱基与碱基间距。因此,无法从单个分子的容易且精确地进行阅读。” (Ottensmeyer, F. P. (1979) " Molecular Structure Determination by High Resolution Electron Microscopy " Ann. Rev. Biophys. Bioeng. 9:129-144 (略去内部参考文献))

[0008] 对个体基因组和其它序列进行快速测序的能力对于生物医学极具重要性,且如上所述,需要改进的方法。

发明内容

[0009] 本发明提供了使用电子显微镜通过直接检视来对标记的拉伸的核酸如 DNA 进行测序的方法。本发明的方法、装置和组合物使得能够可控地将核酸置于基片上从而具有恒定的碱基与碱基间距,这使得能够使用电子显微镜获取准确的核酸序列信息。本发明可以

以多种方式实施。

[0010] 根据本发明的一个方面, DNA 聚合链的序列信息的获取方法可以包括: 提供包含所述 DNA 聚合链的溶液, 其中已对所述 DNA 聚合链进行处理从而已用具有碱基特异性或碱基选择性的造影剂将多个 DNA 碱基标记; 将 DNA 结合工具引入所述溶液并使所述 DNA 聚合物的一部分与该工具结合; 将所述工具从溶液中除去; 将标记的 DNA 聚合链在空间内拉伸从而使该标记的 DNA 聚合链悬挂于空气 / 溶剂界面与所述工具之间; 将经拉伸的标记 DNA 聚合链沉积于基片上; 使用电子显微镜使标记的 DNA 聚合链成像由此确定经标记的碱基和未经标记的碱基的位置; 和将经标记的碱基和未经标记的碱基的位置与 DNA 聚合物的序列进行相关。

[0011] 所述方法还可包括使 DNA 链变性以便在用造影剂进行标记之前产生单链 DNA 的步骤。变性步骤通过热变性或化学变性进行。所述方法还可以包括在所述成像步骤之前将碳层沉积于与基片相连的 DNA 聚合链上方。所述方法还可以包括用不同标记对来自 DNA 聚合物样品的多个链进行标记并汇集从该 DNA 聚合物的多个链的成像获取的数据以便获得完整序列信息。

[0012] DNA 可以为大于约 100kb 的高分子量 DNA。拉伸步骤可能在 DNA 聚合链中产生恒定的碱基与碱基间距。碱基之间的碱基与碱基间距可以为约 $3\text{\AA}$ ~ 约 $7\text{\AA}$, 且特别是约 $5\text{\AA}$ 。DNA 聚合链可以被拉伸至至少约 $2\text{ }\mu\text{m}$ 的长度。

[0013] 可以仅对 DNA 聚合链的碱基的子集进行标记。碱基子集包括胸腺嘧啶和胞嘧啶。碱基子集包括腺嘌呤和鸟嘌呤。

[0014] 所述工具可以为针, 从而针的尖端已用与 DNA 聚合链结合的第一涂层功能化。针可以由玻璃、金、钨、PMMA、聚苯乙烯、PVC 或硅制成。所述涂层可以为如 PMMA、聚苯乙烯、PVB 和寡核苷酸等化合物。可以采用不与核酸结合的第二涂层对针进行功能化。第二涂层可以为辛基硫醇、己基硫醇、壬基硫醇、癸基硫醇和庚基硫醇中的至少一种。针尖的曲率半径小于约 200nm。可以使所述针以约 1 纳米 / 秒 ~ 约 100 米 / 秒的速率移进和移出溶液。具体而言, 可以使针以约 $1\text{\AA}$ ~ 约 $20\text{ }\mu\text{m}$ 的深度进入溶液。

[0015] 造影剂可以为高 Z 原子标记化合物。高 Z 原子可以为 Os-bipy (Os-bipy)、乙酸汞和二甲基硫化铂。造影剂可以为高 Z 原子簇标记。

[0016] 可以通过采用搁板拉线 (shelf threading) 将 DNA 聚合链附着于基片。DNA 聚合链可以与载体基片或成像基片相连。可以通过采用搁板拉线将 DNA 聚合链附着于载体基片, 其后通过采用转印 (transfer printing) 将该载体基片上的标记的 DNA 聚合链转移至成像基片。可以通过采用间隙拉线 (gap threading) 将标记的 DNA 聚合链附着于载体基片, 其后通过采用刷印 (swipe printing) 将该载体基片上的标记的 DNA 聚合链转移至成像基片。DNA 聚合链可以为多个 DNA 聚合链, 且可以作为平行的链阵列被安置于基片上。拉伸的标记 DNA 聚合链可以为单链的。

[0017] 根据本发明的一个方面, 核酸序列信息的获取方法可以包括通过电子显微镜确定多个标记的核酸链的每一个的至少 1000 个毗连碱基区域内经标记的碱基和未经标记的碱基的位置。特别地, 可以确定核酸链的至少 1000 个毗连碱基的位置, 且更特别地可以确定核酸链的至少 10,000 个毗连碱基的位置。可以确定至少约 20 个个体链的经标记的碱基和未经标记的碱基的位置, 可以确定至少约 100 个多条链的位置, 可以确定至少约 5,000 个多

条链的位置,且可以确定至少约 10,000 个多条链的位置。可以确定来自相同样品的经不同标记的核酸聚合链的位置。可以确定至少 200 个毗连碱基、至少 500 个毗连碱基、至少 1000 个毗连碱基或至少 10,000 个毗连碱基的经不同标记的核酸聚合链的位置。

[0018] 核酸序列可以以至少 1,000 碱基 / 秒的速率获得。核酸链在伸长时可能具有约 100 μm 的长度。核酸在碱基之间的碱基至碱基间距可以为 3 $\text{\AA}$ ~约 7 $\text{\AA}$,且特别是约 5 $\text{\AA}$ 。核酸序列可以为 DNA 序列。

[0019] 根据本发明的另一个方面,一种制造品可以包含具有多个核酸聚合链的液体、工具和具有处于该液体中的第一末端和与该工具相连的第二末端的单个核酸聚合链,其中所述单个核酸聚合链的至少一部分悬挂于所述工具与所述液体之间的空间中。可以将所述核酸聚合链的碱基伸长从而令碱基具有恒定的碱基与碱基间距。碱基之间的碱基与碱基间距可以为 3 $\text{\AA}$ ~约 7 $\text{\AA}$ 。可以将核酸聚合链伸长以使该链为线性的。

[0020] 根据本发明的又一个方面,一种制造品可以包含固体平面基片和位于该平面基片上至少一个伸长核酸聚合链。所述至少一个伸长核酸聚合链在约 1000 个碱基对的长度上具有恒定的碱基与碱基间距。所述制造品还包含安置于所述至少一个伸长核酸聚合物上方的膜,从而使所述至少一个伸长核酸聚合物夹在平面基片与所述膜之间。所述膜可以由碳或低 Z 元素构成。平面基片可以由如 PDMS、碳、硼、锂、氢、铍、铝、氮化物、氮氧化物及其组合等材料构成。

[0021] 所述至少一个伸长核酸链的碱基与碱基间距可以为 3 $\text{\AA}$ ~约 7 $\text{\AA}$ 。多个伸长核酸链可以基本彼此平行。多个伸长核酸链可以包含约 1×10^6 核酸链。所述至少一个伸长核酸链可以拉伸至至少约 2 μm 的长度。所述至少一个核酸聚合链可以用至少一个 Z 标记化合物进行标记。Z 标记化合物可以为高 Z 原子化合物,例如,Os-bipy、乙酸汞和二甲亚砷合铂或簇标记。

[0022] 根据本发明的再一个方面,一种使用电子显微镜对核酸链的至少 200 个毗连碱基进行测序的方法可以包括:用至少一个 Z 标记化合物对多个核酸链进行标记;将来自含有多个标记核酸链的溶液的单个标记核酸链结合至工具上;将所述单个标记核酸链在空间内拉伸从而使该单个标记核酸链悬挂于空气 / 溶剂界面与所述工具的尖端之间;将经拉伸的标记核酸链附着于基片;和使用电子显微镜使标记的核酸链成像。

[0023] 所述方法还可以包括在所述标记步骤之前用亚硫酸氢盐对含有多个核酸链的溶液进行预处理以便将未甲基化的胞嘧啶碱基转化为尿嘧啶碱基。所述方法还可以包括在所述标记步骤之前使多个核酸链变性以便产生单链核酸。变性步骤可以通过热变性或化学变性进行。所述方法还可以包括在所述成像步骤之前将碳层沉积于与基片相连的标记核酸链的上方。

[0024] 核酸可以为高分子量 DNA。拉伸步骤在核酸链中产生恒定的核酸与核酸间距。核酸与核酸间距和标记与标记间距可以为约 3 $\text{\AA}$ ~约 7 $\text{\AA}$,且特别可以为 5 $\text{\AA}$ 。核酸链可以拉伸至至少约 25 μm 的长度。仅对核酸链的碱基子集进行标记。碱基子集可以包括胸腺嘧啶和胞嘧啶。碱基子集可以包括腺嘌呤和鸟嘌呤。

[0025] 所述工具可以为针从而已将针尖用与核酸链结合的第一涂层功能化。该针可以由如玻璃、金、钨、聚甲基丙烯酸甲酯 (PMMA)、聚苯乙烯、PVC 和硅等材料制成。所述涂层可以为如 PMMA、聚苯乙烯、PVB 和寡核苷酸等化合物。可以采用不与核酸结合的第二涂层对针进

行功能化。第二涂层可以为如辛基硫醇、己基硫醇、壬基硫醇、癸基硫醇和庚基硫醇等化合物。

[0026] 针尖的半径可以为小于约 200nm。可以使所述针以约 1 纳米 / 秒 ~ 约 100 米 / 秒的速率移进和移出溶液。可以使针以约 0nm ~ 约 20 μm 的深度进入溶液。

[0027] 所述至少一个标记化合物可以为高 Z 原子标记化合物。高 Z 原子可以为如 Os-bipy、乙酸汞和二甲亚砷合铂等一种或多种化合物。所述至少一个标记化合物可以为高 Z 原子簇标记。

[0028] 附着步骤可以包括通过采用搁板拉线将至少一个标记的核酸链附着于成像基片。附着步骤可以包括通过采用搁板拉线将至少一个标记的核酸链附着于载体基片,其后通过采用转印将该载体基片上的至少一个标记的核酸链转移至成像基片。附着步骤可以包括通过采用间隙拉线将至少一个标记的核酸链附着于载体基片。附着步骤可以包括通过采用间隙拉线将至少一个标记的核酸链附着于载体基片,其后通过采用刷印将该载体基片上的至少一个标记的核酸链转移至成像基片。

[0029] 根据本发明的另一个方面,将至少一个核酸链受控地安置于基片上的方法可以包括:提供含有多个核酸链的溶液;将针尖插入溶液;将针尖从含有多个核酸链的溶液中提出,其中已用与核酸结合的第一涂层将所述针尖功能化;将核酸链在空白空间内拉伸从而使单个核酸链悬挂于空气 / 溶剂界面与每个针尖之间;并将至少一个拉伸的核酸链附着于基片。核酸为高分子量 DNA。

[0030] 针可以由如玻璃、金、钨、PMMA、聚苯乙烯、PVC 和硅等材料制成。第一涂层可以为如 PMMA、聚苯乙烯、PVB 和寡核苷酸等化合物。可以采用不与核酸结合的第二涂层对针进行功能化。第二涂层可以为如辛基硫醇、己基硫醇、壬基硫醇、癸基硫醇和庚基硫醇等化合物。

[0031] 针尖的曲率半径可以为小于约 200nm。可以使针尖以约 1nm/s ~ 约 100m/s 的速率移进和移出溶液。可以使针尖以约 0nm ~ 约 20 μm 的深度进入溶液。可以将针置于单纳米定位器驱动的载体上。可以用纳米定位器驱动的载体将多个针同时插入溶液。

[0032] 拉伸步骤在核酸链中产生恒定的核酸与核酸间距。核酸与核酸间距和标记与标记间距可以为约 3 $\text{\AA}$ ~ 约 7 $\text{\AA}$ 。核酸与核酸间距和标记与标记间距可以为 5 $\text{\AA}$ 。核酸链可以拉伸至约 100 μm 的长度。

[0033] 将附着于基片的所述至少两个核酸链定位来基本彼此平行。附着步骤可以包括采用搁板拉线来将至少一个核酸链附着于基片。附着步骤可以包括使用间隙拉线来将至少一个核酸链附着于基片。基片可以为成像基片。所述附着步骤中的基片可以为载体基片。

[0034] 根据本发明的再一个方面,提供了一种对储存于存储器中的核酸序列进行分析的方法,其中通过下述方式来确定序列:用至少一个标记化合物对多个核酸链进行标记;将来自含有多个标记核酸链的溶液的单个标记核酸链结合至工具上;将所述单个标记核酸链在空间内拉伸从而使该单个标记核酸链悬挂于空气 / 溶剂界面与所述工具的尖端之间;将经拉伸的标记核酸链附着于基片;和使用电子显微镜使标记的核酸链成像。核酸序列是人类受试对象的基因组序列。

[0035] 所述分析可以包括对一个或多个单核苷酸多态性中的至少一个的存在或缺失、拷贝数、变体、得失位、重排或全基因组比较进行确定。所述存储器是选自硬盘或软盘、光学介质、光盘 (CD)、数字视频光盘 (DVD)、半导体介质和闪存的介质。

[0036] 根据本发明的又一个方面,一种针可以包含远端、近端和在与所述远端和所述近端接触的液体内延伸的轴,其中所述近端包含曲率半径小于约 200nm 的尖端部件,且其中已用与核酸末端结合的化合物将尖端部件功能化。可以用不具有与核酸结合的亲和力的第二化合物来涂覆针。可以用如 PMMA、聚苯乙烯、PVB、硅烷化物、寡核苷酸、端粒和限制位点伸出物 (restriction site overhang) 等化合物将尖端部件功能化。针可由如玻璃、金、钨、PMMA、聚苯乙烯、PVC 和硅等化合物构成。可将针安置于单纳米定位器驱动的载体上。

[0037] 考虑到以下详细描述、附图和权利要求,可以使本发明的其它特点、优点和实施方式得到展示或显而易见。此外,可以理解,以上对本发明的概述和以下的详细描述都是示例性的并且旨在提供进一步的说明而不限所要求的本发明的范围。

附图说明

[0038] 所包含的提供对本发明的进一步理解的附图并入且构成本说明书的一部分,其对本发明的实施方式进行图示说明,并与详细描述共同起到解释本发明的原理的作用。并未试图以超出对于本发明及可以将其实施的各种方式的基本理解所必须的细节来展示本发明的结构细节。

[0039] 图 1 是显示根据本发明的原理使用电子显微镜准确地对核酸进行测序的方法的流程图。

[0040] 图 2 是显示用涂层 1 和涂层 2 功能化的针的示意图,其中涂层 1 能与核酸链结合而涂层 2 不与核酸链结合。

[0041] 图 3 显示尖针的曲率半径的示意图。

[0042] 图 4 是显示根据本发明的原理将核酸链在空白空间内伸长的方法的示意图。图片 I 显示针尖和含多个核酸链的溶液的液滴。图片 II 显示移进含有核酸链的溶液的液滴内并与单个核酸链结合的针尖。图片 III 显示在空白空间内被伸长的核酸链。

[0043] 图 5 是显示针尖在含核酸链的溶液液滴内的最大浸入深度 (图片 I) 和最小浸入深度 (图片 II) 的示意图。

[0044] 图 6 是显示针尖在含核酸链的溶液液滴内的最小浸入深度的示意图。放大图示意性显示了与单一核酸链特异性结合的溶液液滴中的针尖,所述单一核酸链可以处在溶液中或在空气 / 溶液界面处。

[0045] 图 7 是显示根据本发明的原理的搁板拉线方法的示意图。图片 I 显示了浸入含标记核酸链的溶液内的尖针。图片 II 显示了从溶液中提离并将所附着的核酸拉伸至空白空间内的尖针。图片 III 显示了开始与 TEM 网格接触的伸长核酸。图片 IV 和 V 显示了拉回以将核酸从尖针的尖端释放的尖针。

[0046] 图 8 是本发明的搁板拉线方法的示意性图示。图片 I 显示了采用单针的搁板拉线,而图片 II 显示了采用多针的搁板拉线。

[0047] 图 9 是显示经调整方向而与液滴表面正交的伸长核酸链的示意图。

[0048] 图 10 是显示伸长于空白空间中的核酸在该链被置于载体基片上时基本沿其长度与该基片接触的示意图。

[0049] 图 11 是显示通过近端放下 (set down) 和随后的更多拉伸而使核酸链延长的示意图。

[0050] 图 12 是显示间隔较近的核酸链（如 DNA 链）的大平行阵列的示意图，所述阵列可以通过依次重复受压电执行器（piezo-actuator）控制的浸入、浸出、放下、拖动、提起和平移的针运动而形成。该图中所绘制的核酸链没有实际实践中的链那么直。

[0051] 图 13 是显示下述故障模式的示意图：其中溶液表面 / 载体基片 / 针运动角度的不正确考虑将引起载体基片与悬挂的链在针尖与基片接触之前的不受控接触，而这反过来会通过过度拉伸而造成链的断裂。

[0052] 图 14 是显示下述故障模式的示意图：其中错误校正的溶液表面 / 载体基片 / 针运动角度会使得链无法与基片实质上接触，从而使得该链的相当大的一部分留在溶液表面与针与基片的接触点之间的空白空间中。

[0053] 图 15 是显示根据本发明的原理的间隙拉线方法的示意图。

[0054] 图 16 是显示在将核酸沉积至载体基片上之后的根据本发明的原理的转印方法的示意图。

[0055] 图 17 是显示在通过电子显微镜成像之前基片上被嵌入顶层的核酸链的示意图。

[0056] 图 18 是显示用于根据本发明的原理对核酸测序的系统的示意图。

[0057] 图 19 是显示用于根据本发明的原理对核酸测序的假定系统的示意图，使用本发明的原理产生了单一钡标记的 ssDNA 分子的模拟图像。

[0058] 图 20 是显示替代性制备方法中固有的图像模糊性的示意图。

[0059] 图 21 是显示简单模式液滴保持器的示意图。

具体实施方式

[0060] 应该理解，本发明不限于本文所述的特定方法学、方案和试剂等，因为如本领域资深技术人员可以认识到的那样，这些都可以进行变化。还应理解，本文所用的术语仅用于描述特定实施方式的目的，而并非旨在限制本发明的范围。还应注意，如本文和所附权利要求书中所用，除非上下文明确另外指明，单数形式“a”、“an”和“the”包括了复数的指代物。因此，例如，对一个分子的指代即是对一个或多个分子及其本领域技术人员已知的等价物的指代。

[0061] 除非另外限定，本文使用的所有技术和科学术语具有与本发明所属领域的普通技术人员所通常理解的相同的含义。本发明的实施方式及其各种特点和有利细节可以通过参考非限制性实施方式而得到更全面地说明，和 / 或在附图中得到图示说明和在以下描述中得到详细说明。应该注意，附图中所示特点不一定是按比例绘制的，并且即使在本文中未明确指出，一个实施方式的特点也可以如本领域资深技术人员所认识的那样由其它实施方式所采用。

[0062] 本文所述的任何数值包括从低值至高值的以一个单位递增的所有值，条件是在任何低值和任何高值之间具有至少两个单位的间隔。举例而言，如果据说明组分的浓度或如尺寸、角度大小、压力和时间等过程变量值是例如 1 ~ 90、特别是 20 ~ 80、更特别是 30 ~ 70，其意图在于将诸如 15 ~ 85、22 ~ 68、43 ~ 51、30 ~ 32 等值明确在本说明书中列出。对于小于 1 的值而言，根据情况认为一个单位是 0.0001、0.001、0.01 或 0.1。这些仅仅是特别计划的实例，而在所列出的最低值与最高值之间的所有可能的数值组合都应被认为是以相似的方式明确在本申请中得到陈述。

[0063] 此外,以下紧接着提供了“定义”部分,其中具体定义了与本发明相关的某些术语。对特定的方法、装置和材料进行了描述,但与本文所述的那些方法和材料相似或等价的任何方法和材料都能用于实践或测试本发明。本文通过参考并入本文所提及的所有参考文献的全部内容。

[0064] 定义

[0065] A 是腺嘌呤。

[0066] C 是胞嘧啶。

[0067] G 是鸟嘌呤。

[0068] T 是胸腺嘧啶。

[0069] U 是尿嘧啶。

[0070] ssDNA 是单链 DNA。

[0071] AFM 是原子力显微镜。

[0072] CCD 是电荷耦合器件。

[0073] CMOS 是互补金属氧化物半导体。

[0074] DMSO 是二甲亚砜。

[0075] EDTA 是乙二胺四乙酸。

[0076] HAADF 是高角度环形暗场。

[0077] IMPREST 是单个分子布局空白空间拉线。

[0078] PFGE 是脉冲场凝胶电泳 PFGE。

[0079] Os-bipy 是四氧化钼 2,2'-二吡啶。

[0080] PDMS 是聚二甲基硅氧烷。

[0081] PLD-UHV 是脉冲激光沉积-超高真空。

[0082] PMMA 是聚甲基丙烯酸甲酯。

[0083] PMT 是光电倍增管。

[0084] PVB 是聚乙烯醇缩丁醛。

[0085] SEM 是扫描电子显微镜。

[0086] STEM 是扫描透射电子显微镜。

[0087] TE 是 Tris EDTA。

[0088] TEM 是透射电子显微镜。

[0089] UHV 是超高真空。

[0090] 本文所用术语“Z”是指原子核中的质子数,也称作原子数。“高 Z”是指原子数较大,然后薄膜成像,但对实际测序而言意味着高于约 30、优选高于约 70 或更优选高于约 90。

[0091] 本文所用术语“恒定”通常是指在拉伸的核酸链的整个长度上,拉伸的核酸链的碱基之间的间距相对而言相同。碱基“之间”的间距通常是从中心至中心的距离,或从磷酸根至磷酸根的距离。当可以在例如约 50 个碱基、约 100 个碱基、约 1000 个碱基或约 10,000 个碱基的特定长度上确定核酸链的电子显微图像中标记的碱基和未经标记的碱基的顺序时,碱基的间距是恒定的。

[0092] 本文所用术语“毗连”通常是指核酸链中的碱基均处于公共边界内并且没有间断地相连。

[0093] 本文所用术语“连续”通常是指核酸链中的碱基以未中断的连续性或顺序彼此相接。

[0094] 短语“多个链”、“多个核酸链”和“多个 DNA 链”等是指 2 个链、5 个链、10 个链、100 个链和 1000 个链等等。

[0095] 本文所用术语“载体基片”是指伸长和 / 或拉伸的核酸聚合物可以与其相连的任何基质。

[0096] 本文所用术语“成像基片”是指将被直接用于电子显微镜成像的基片。成像基片可以被置于显微镜内,并且必要时可以支撑成像薄膜。成像基片可以包括多孔或微孔弗姆瓦网 (lacey formvar mesh)、其它聚合物网、带孔氮化硅薄膜和其它合适类型的网格。成像基片通常比成像薄膜厚,但必须可支撑纤巧的成像薄膜。本文所用的成像基片可以指成像薄膜和保持其的载体结构体,或仅指载体基片。参见描述电子显微镜的一般方法的 M. Hayat, Principles and Techniques of Electron Microscopy Biological Applications, 第 4 版。

[0097] 本文所用术语“成像薄膜”是指以下物质的薄层:碳、硼、锂、氢、铍、铝或其它低 Z 元素和 / 或其氮化物和氧化物以及它们的任何组合。该薄层的厚度为约 0.2nm ~ 约 30nm。成像薄膜可以被支撑在成像基片或其它表面上。

[0098] 术语“基本平行”通常是指两条直线从不相交的几何概念。本文所用的基本平行是指伸长的核酸聚合物在其中序列数据待确定的区域内不相交或交叉。

[0099] 本文所用术语“核酸”包括寡核苷酸和多核苷酸,以及源自基因组、重组或合成的可以为单链或双链并且代表正义或反义链的 DNA 或 RNA,或指源自天然、重组或合成的任何 DNA 样或 RNA 样材料。

[0100] 本文所用术语“互补”包括在许可的盐和温度条件下通过碱基配对而成的多核苷酸的天然氢键。例如,序列“A-G-T”与互补序列“T-C-A”结合。

[0101] 两个单链分子之间可以是部分互补的,其中仅核酸的某些部分结合,或者在单链分子之间存在完全互补性是可以是完全互补的。核酸链之间的互补程度对于核酸链之间的杂交的效率和强度具有显著影响。

[0102] 本文所用术语“样品”是指包含核酸的生物材料,例如来自人或动物的组织或液体,包括但不限于:血浆、血清、脊髓液、淋巴液、皮肤的外部部分、呼吸道、肠道和泌尿生殖道、泪液、唾液、血细胞、肿瘤、器官、组织,以及来自植物、真菌、细菌、病原体 and 体外细胞培养物的样品。样品可以获自任何含有核酸的物种,在系统发生学上涵盖了所有病毒、原核生物和真核生物。样品还可以包含已使用如固相合成等本领域内的已知技术进行人工合成或使用例如 PCR 进行体外合成的核酸。

[0103] 本文所用术语“大约”是用于描述值的范围,可应用于该范围的上限和下限。例如,短语“约 10 ~ 100”与“约 10 ~ 约 100”具有相同含义。而且,当指代距离时,术语“大约”通常是指 $\pm 10\%$ 。例如,短语“约 5nm”是指 $5\text{nm} \pm 10\%$ 。

[0104] 本文所用术语“造影剂”、“标记”或“标记化合物”通常是指其原子数 (Z) 和 / 或密度和 / 或积分电子散射比成像薄膜材料和未标记的 DNA 更高的原子、分子、簇或材料。“造影剂”、“标记”或“标记化合物”可以是具有与核酸链自身的碱基不同的对比度并且与该核酸链相连的化合物。

[0105] 本文所用术语“簇”通常是指包含两个以上的高 Z 原子且与核酸链以碱基选择性

或碱基特异性方式相连的化学结构。

[0106] 综述

[0107] 本发明广泛地涉及通过使用电子显微镜直接检视标记的拉伸核酸来确定核酸序列的方法、装置和制造品。在一个具体实施方式中,本发明涉及包括使用工具将单核酸链从溶液中拉出而受控地将核酸置于基片或载体上。本发明的方法使得可以获得比现有测序技术更高的准确率、更低的成本和更长的阅读长度。例如,本发明的测序方法使得可以使用电子显微镜准确地确定核酸样品的至少约 20 个连续核酸碱基、优选至少约 50 个连续核酸碱基、更优选至少约 1,000 个连续核酸碱基、再更优选至少约 10,000 个连续核酸碱基、进而更优选至少约 100,000 个连续核酸碱基且再更优选至少约 1,000,000 个连续核酸碱基。在此上下文中的“连续碱基”是指所分析的 DNA 起始材料中的碱基的顺序。

[0108] 使用本发明的方法还可以比使用合成方法更快地产生序列。本发明的方法(与高速 EM 组合)可以允许至少约 10,000 碱基/秒、优选至少约 100,000 碱基/秒且更优选至少约 200,000 碱基/秒的成像。例如,使用 TEM 成像可以以约 $1\mu\text{m}^2/\text{秒}$ 的高分辨速率使根据本发明排列的 DNA 链样品成像。含核酸链的 $1\mu\text{m}^2$ 区域可以在 1 秒中成像,对应于约 500,000 碱基/秒的成像速率。

[0109] 图 1 是显示根据本发明的原理使用电子显微镜准确地对核酸进行测序的方法的流程图。该图供以图示说明而非意在限制本发明。在步骤 102 中,从含有核酸序列的受试对象获取含核酸样品。在步骤 104 中,使用本领域技术人员已知的技术将感兴趣的核酸从样品分离。在步骤 106 中,采用例如高 Z 原子对所分离的核酸的特定碱基进行标记以产生经高 Z 原子标记的核酸聚合物。在步骤 108 中,将核酸聚合物拉伸至空白空间内以便确保恒定的核苷酸的碱基至碱基间距。在步骤 110 中,将核酸聚合物与载体附着并以非重叠图案铺设(例如,基本相对彼此平行)。在步骤 112 中,通过电子显微镜将所附着的核酸聚合物成像以便沿该聚合物确定标记的位置。通过检测器(例如,CCD 或 CMOS 照相机或者 PMT)捕获图像并记录标记的位置,通常将其记录于计算机可读介质上。在步骤 114 中,采用算法来使用间距信息以确定来自核酸聚合物的碱基序列信息。下文将对这些步骤进行更详细的描述。

[0110] 核酸的制备

[0111] 在步骤 104 中,可以使用本领域公知的方法将感兴趣的核酸分离,具体方法的选择取决于核酸的来源和性质以及相似的因素。感兴趣的核酸可以为天然存在和/或源自基因组、不是源自合成或重组,并且可以包括双链或单链的寡核苷酸或多核苷酸。替代性地,感兴趣的核酸链可以源自重组或合成,其可以为单链或双链。在一个具体实施方式中,核酸是 DNA,且具体而言是大于约 100 千碱基的极高分子量 DNA。在某些实施方式中,高分子量 DNA 的长度至少为 300 千碱基。极高分子量的分离用方法是已知的(参见例如, Murry 和 Thompson, 1980, Nucleic Acids Research 10:4321-4325; 和 Kovacic, R. 等, 1995, 23(19) Nuc. Acids. Res. 23(19)3999-4000)

[0112] 根据一种方法,将极高分子量 DNA 从嵌入于琼脂糖块中的真核细胞中分离以使剪切最小化。通过 PFGE 将 DNA 与其他细胞组分分开,随后从琼脂糖中以约 $0.01\text{ng}/\mu\text{l}$ ~ 约 $0.5\text{ng}/\mu\text{l}$ ~ 电洗脱至 TE 缓冲液。可以使用本领域技术人员已知的热变性或化学变性方法将 DNA 从双链形式变性为单链形式。参见 Barnes, W., 91PNAS 2216-2220 (1994)。例如,可

以通过在 2 分钟内将 DNA 样品加热至 94℃ 来进行双链 DNA 的热变性。变性步骤可以在标记之前进行,也可在将核酸拉线 (threading) 之前进行。理想的是,如果使用 ssDNA 测序则在标记之前将 dsDNA 转化为 ssDNA,但这不是必须的,并且可以如本领域资深技术人员所理解的来标记 dsDNA。

[0113] 核酸标记

[0114] 在步骤 106 中,采用如高 Z 原子等造影剂来标记核酸的特定碱基以便通过电子显微镜进行高效检测。应该使核酸与电子致密原子以至少部分碱基特异性方式结合。部分碱基特异性的造影剂或标记 (即“碱基选择性”) 会与 4 种 DNA 或 RNA 碱基的一种或多种比其它碱基更优选地结合 (例如,将 A 强染色,将 G 较弱染色,完全不将 T 染色)。如果造影剂基本仅与一种碱基 (例如 A) 或序列 (例如,特定的二聚体序列) 结合,则其为完全碱基特异性的。已知有相当多的 DNA 标记用方法并且它们可以用于本发明中。例如,Whiting 等,474BIOCHIMICA ET BIOPHYSICA ACTA334-348(1977),Jelen 等,10GEN PHYSIOL BIOPHY 461-473(1991 和 Dale 等,14(11)BIOCHEMISTRY 2447-2457(1975) 描述了碱基特异性和/或碱基选择性重金属染色方案,本文通过参考特意并入其全部内容。本发明不应被视为受任何特定的核酸标记方法的限制,且如本领域技术人员所理解,许多方法已在科学文献中得到描述。

[0115] 根据所用方法,可以用不同的高 Z 标记化合物将每种核酸碱基专有性标记,可以用不同的高 Z 标记化合物来标记核酸碱基的选定子集,或可以用如 Os-bipy 等相同高 Z 标记化合物来标记核酸碱基的选定子集 (如下所述)。根据一个具体实施方式,用不同的 Z 标记化合物来专用地或优选地标记每种 DNA 碱基 (例如,A、G、C 和 T)。在本发明的方法中可以采用各种各样的高 Z 标记化合物,包括但不限于包含 Pt、Hg、I、Rh、Au、Ir、Ag 和 Os 等的化合物。可以例如基于高 Z 标记剂的差异或高 Z 试剂和未经标记的碱基之间的差异来区分不同碱基,这几乎应该是不可见的。例如,Os-bipy 和碘可以基于其散射截面来区分。此外,例如,可以将与所有 A 碱基结合的 Au 原子和与所有 G 碱基结合的 3 个 Au 簇区分开。

[0116] 由于在本发明的方法中 TEM 成像检测具有恒定碱基间距的各种非重叠 DNA 链,因而标记的保真度不是非常重要。因此,本发明的方法的一个优点在于,可以对核酸进行化学标记,并且对于给定的反应批次不必将每个链的所有靶标碱基都标记 (例如,当使用 T 特异性标记时,不必将链中的每个 T 均标记)。由于可以将相同链的多个图像组合来确定基本序列 (这是由于可以确定标记碱基之间的未经标记的碱基的数目),因而完全的标记准确性不是必要的。举例而言,考虑在一批 DNA 中出现 20 次的具体基因组序列 (即,在该批次中存在共享对于该位置共有的区域的 20 个分子)。例如,可以将该批次调节来含有总 DNA 的约 20 个基因组等价物。假设序列中的特定位置的碱基身份是 T,即所有 20 个在该位置均含 T。使含有这些分子的 DNA 批次进行特定的反应条件,所述反应条件已知会标记 90%~100%的任何给定 T 并且标记 0%~10%的任何给定 A、C 或 G。在成像时,在此情形下可将该位置视为在 19 个分子中被标记,而一个分子中未经标记。同样,可以使一个分子中的“A”被标记,而在 19 个分子中未标记。机率性处理可以以极高的准确性将对基因组中的所述位置指认出正确的身份。在该情形中,19/20 个分子中标记的位置的身份是“T”,而 1/20 个分子中标记的位置的身份不是“T”。此外,通过本领域技术人员已知的机率性处理可以实现序列的指认和联接。

[0117] 本发明的另一个优点在于,可以推导出互补链的序列信息,这也为给定碱基的确定的有效性提供了额外的统计支持,即,一个分子上的T的位置的高置信度与互补链上A的位置的高置信度相互密切联系。因此,例如,在来自就标记化学而言仅允许对T的良好区分的化合物/反应条件批次的序列中,可以肯定地确定所有A和T的位置。每条链上标记的T限定了互补链上的A的位置。

[0118] 因此,在某些实施方式中,仅将核酸碱基的子集标记。例如,以至少部分碱基特异性的方式将DNA的等分试样分别标记。在一种方法中,使含多个核酸分子的溶液在以下条件下反应:至少约70%、有时至少约80%且有时至少约90%~约100%的一种或多种特定核苷酸碱基(A、T、G或C)被标记,而少于20%、优选少于10%的至少一种核苷酸碱基被标记。例如,使溶液在以下条件下反应:约90%~约100%的T和C被标记而仅很小百分比的A和G被标记。

[0119] 在一个实施方式中,可以使用不同条件以Os-bipy将DNA等分试样标记以便获取不同的碱基特异性标记密度(参见下文实施例1)。在该方法中,将感兴趣的DNA分离并分为两个溶液,即溶液1和2,每个溶液中的浓度为约0.01ng/ μ l~约1ng/ μ l。使用不同条件(如下文所进一步详细描述)使每个溶液与Os-bipy反应以便获取不同的碱基特异性标记密度。此外,对选定的溶液在与Os-bipy反应之前进行预处理,例如,亚硫酸氢盐处理,如以下和下文实施例1中所述。

[0120] 使溶液1在26°C与pH 8.0的TE缓冲液(具有100mM Tris和10mMEDTA)过量4倍摩尔的四氧化锇和2,2'-二吡啶反应20小时,这些条件标记了约100%的T、约85%的C、约7%的G和约0%的A。使溶液2在与溶液1相同的条件下反应,不同之处在于反应仅进行15分钟,且仅使用过量2.5倍摩尔的四氧化锇和2,2'-二吡啶,这些条件标记了约90%的T、约8%的C、约5%的G和约0%的A。如下所述,使用本发明的方法,可以将这些结果与经较短温育时间反应获得的那些结果相比,由此确定C的碱基特异性模式,甚至于确定互补链的G的碱基特异性模式。甚至还可以使用本发明的方法来确定T的碱基特异性模式乃至互补链的A的碱基特异性模式。因而,可以使用采用两种反应条件的单一标记化合物来确定T、A、G和C的模式。

[0121] 还可理解,本发明允许使用Os-bipy来确定胞嘧啶甲基化的模式。在此情形中,使用4个等分试样(溶液1、2、3和4)。将溶液1和2如上文溶液1那样处理,将溶液3和4如上文溶液2那样处理。然而,在进行Os-bipy反应之前,首先对溶液2和4进行亚硫酸氢盐处理以便将未甲基化的C残基转化为U。这使得可以通过对来自用亚硫酸氢盐处理过的溶液与来自未经处理的溶液的序列的比较而确定甲基化模式。参见Jelen等,10Gen. PHYSIO. AND BIOPHYS 461-473(1991)。在与Os-bipy的反应中的不同条件下,甲基胞嘧啶和U都具有不同于4种典范碱基的标记效率的标记效率。因此,为了推导出基因组样品的后成修饰(epigenetic modification)信息,可以在未用亚硫酸氢盐处理的DNA的环境中确定甲基胞嘧啶的碱基特异性模式,或可在亚硫酸氢盐处理之后的测序环境中确定U的模式。在进行标记反应后,通过超滤或透析除去未标记的锇以使成像过程中的外源性重原子污染最小化,并将DNA聚合物在pH 8的TE缓冲液中稀释至约0.1ng/ μ l。

[0122] 许多其它高原子的碱基特异性高原子标记方法是本领域技术人员所已知的。一些实例如下:Beer和Moudrianakis描述了使用与铀酰粒子偶合的重氮盐化合物来进行鸟

嘌呤残基的标记 (Beer, M. 和 Moudrianakis, E., Determination of Base Sequence in Nucleic Acids with the Electron Microscope 48(3)PNAS, 409-416(1962))。Robert Whiting 使用 Pt-DMSO[二甲亚砷合铂, KPtCl₃(DMSO)] 来实现适于鉴定腺嘌呤核苷酸的不同标记 (Studies of Nucleic Acid Sequences by Dark Field Electron Microscopy, 博士论文, (1975), University of Toronto, School of Graduate Studies)。Dale 及其合作者报道了使用乙酸汞对胞嘧啶残基进行汞化 (Direct covalent mercuriation of nucleotides and polynucleotides, Dale, RMK 等, 14BIOCHEMISTRY, 2447-2457(1975))。另一种 DNA 标记方法是首先将其以碱基特异性方式进行共价修饰从而其在经修饰的碱基处接受重金属标记。该方法的一个实例由 Seth Rose 描述, 他用氯乙醛修饰腺嘌呤和胞嘧啶残基从而乙酸汞或四氧化钨能随后与这些残基结合 (Rose, SD., Mercuriation of Modified Nucleotides Chemical Methods Toward Nucleic Acid Sequencing by Electron Microscopy, 361 BIOCHIM. BIOPHYS. ACT., 231-235(1974))。

[0123] 在另一个实施方式中, 可以在本发明的方法中采用簇标记。簇标记 (cluster label) 是含有多于一个重原子的化合物。簇标记可以在利用可提供足够的碱基与碱基间隔的拉伸方法的方案中使用。为了获得不受邻近的附着的簇的空间位阻的限制的序列数据, 足够的间隔是必需的。可以将簇与寡聚物相连, 然后使寡聚物与 DNA 通过互补碱基配对而杂交。以这种方式, 可以使用电子显微镜对互补序列进行定位。为了有效地获取完全序列数据, 可以将来自不同批次的与未知序列杂交的极短的簇标记寡聚物 (三聚物和四聚物) 的成像的信息合并。具体而言, 极短的簇标记寡聚物的长度可以为约 5 ~ 约 20 个碱基。能直接标记未经修饰的 DNA (即具有天然碱基组成而没有非天然碱基的 DNA, 所述非天然碱基含有可以增加簇标记的效率的官能团) 的簇包括如 Rosenberg 等, 689J. ORGANOMETAL. CHEM. 4729-4738(2004) 所述的三钨化合物。在 DNA 链的簇标记中, 特定的经修饰的核苷酸具有如氨基烯丙基或巯基等官能团, 这些官能团使得可以与如来自 Nanoprobe Inc. (Yaphank, NY) 的单马来酰亚氨基十一金等商购簇标记试剂反应。可以用以碱基特异性或碱基选择性方式与 DNA 反应的连接子功能化的商购簇标记试剂包括十二烷羰基三钨、十二烷羰基三钨和十二烷羰基四钨 (三种试剂均可获自 Sigma Aldrich Corp., St. Louis, MO) 和单马来酰亚氨基十一金 (Nanoprobe, 见上)。通过在溶液中拉伸的 DNA 上进行标记反应, 可以更有效地进行具有受到空间位阻的簇的簇标记。还可以在细管 (直径为约 30nm ~ 约 1 μm) 内进行标记, 其中用已知技术将 DNA 在管或毛细管内在溶液中伸长 (Chan, E. 和 Goncalves, N., 14(6)GENOME RES. 1137-1146(2004)) 并与标记反应。该方法的优点是避免具有超过一个结合位点的标记 (例如, 纳米金) 的交联。

[0124] 在一个实施方式中, 可以将簇标记用作造影剂, 而使用已知可碱基选择性或碱基特异性地修饰核酸聚合物的化学连接结构, 可以将已知簇以碱基特异性或碱基选择性方式与核酸聚合物相连。该簇标记方法可以称作“搭载 (piggybacking)”。搭载可以以如下方式进行, 例如, 可以使用乙酸汞来将胞嘧啶汞化, 如 Dale, R. 等, 14BIOCHEMISTRY 2447-2457(1975) 中所述, 并且可以通过使经汞桥连的三钨簇 (mu₃-eta²-c²-t-Bu) Os₃(CO)₉(mu-Hg)I 乙酰化来用乙酸汞部分制备可与胞嘧啶连接的簇化合物, 如 Rosenberg, E. 等, 10ORGANOMETALLICS 203-210(1991) 中所述。作为搭载的另一个实例, 已知邻菲罗啉可与四氧化钨形成复合物, 该复合物可用作碱基选择性标记 (Pacek, E. 等,

J. 13(3)BIOMOL. STRUCT. DYN. 537-546(1995))。5-氨基邻菲罗啉 (Polysciences, Inc., Warrington PA) 可经由其环外氨基与二苯基膦丙酸的琥珀酰亚胺酯 (Argus Chemicals SRL, Vernio, 意大利) 相连。可以通过本领域技术人员已知的技术来将膦与数种簇化合物中的任一种连接 (Cheng 等, 127J STRUCT. BIO. 169-176(1999))。作为搭载的再一个实例, 可以用如磺酸酯等烷基化部分来衍生簇化合物。相关合成步骤描述于 Susan Ermer 等, J. ORGANOMET. CHEM., (1980), 187, 81-90 中。用以对核酸聚合物进行烷基化的磺酸酯如 Yi-Zhang 等, 32(31)BIOCHEMISTRY 7954-7965(1993) 中所述。

[0125] 本发明提供了通过确定核酸 (例如, DNA) 链的指定区域中的选定碱基的位置序列来获取核酸聚合物的序列信息的方法。序列信息是指标记的碱基和未经标记的碱基的一种或多种核酸碱基的位置是已知的, 而位置序列是指至少一种碱基 (例如, T) 相对于其它碱基的位置是确定的。例如, 在其中 T 被标记的 25 个碱基的 DNA 链中, 25 碱基区内的位置序列可以描述如下:

[0126] T000T0000TTT00T00T000T00T

[0127] 其中, “0”是除 T 以外的碱基。因此, 对 T 和非 T 的位置的检测使得可以确定 T 的位置序列。从本说明书可以理解, T 和 / 或 C 和 / 或 G 和 / 或 A 或其组合的位置序列可以得到确定。本发明的方法提供了以至少 70% 的准确性、替代性地至少 80% 的准确性且通常至少 90% 的准确性来确定单个核酸链中至少一种碱基的位置序列的方法。可以确定包含至少 100 个~至多 1,000,000 个碱基、有时 200 个~1,000,000 个碱基、有时 1,000 个~1,000,000 个碱基、有时 10,000 个~1,000,000 个碱基的区域中的位置序列。在某些实施方式中, 确定了链的至少 200 个、至少 1000 个、至少 5000 个、至少 10,000 个、至少 100,000 个或至少 1,000,000 个碱基的区域中的位置序列。在一个实施方式中, 确定了包含 1000 个~100,000 个碱基的区域中的至少一种碱基的位置序列。所述方法的准确性可以通过对已知序列的 DNA 的再测序而确定。

[0128] “位置序列”是序列信息的一种类型。从本公开显而易见, 通过比较各个碱基 (或碱基组合) 的位置序列可以获得更完整的序列信息, 包括链或基因组的区域内的所有 4 种碱基的位置序列 (即完整序列)。

[0129] 有利的是, 本发明的测序方法不依赖于将修饰的核苷酸引入 DNA 或核酸链。尽管该方法与经酶促引入的标记 (例如, 在聚合期间引入) 相容且可以与其共同使用, 但该方法更常与经过分离并直接标记 (使用例如上文和文献中所述的标记) 的天然存在的 DNA 链共同使用。此外, 本方法使得可以 (但不是必须) 针对单链 DNA 而非双链分子来确定序列, 由此消除了采用其它方法可能产生的模糊性。

[0130] 核酸悬挂

[0131] 在步骤 108 中, 将各个核酸聚合物拉伸至空白空间中以确保核酸链内的恒定的碱基与碱基间距。根据一个实施方式, 各 DNA 链在空间内被伸长 (即, 链的相当大部分长度未受基片支撑或者未悬浮于溶液或缓冲液中)。然后可以将基本不含溶液或缓冲液的悬挂的 DNA 转移至成像基片进行成像。该过程可以称作 DNA 拉线 (DNA threading), 下文将详细讨论。

[0132] DNA 在空白空间内的悬挂导致了核酸聚合物的恒定的碱基与碱基间距。在概念上, 这类似于用两手抓住弹簧的两端并拉伸以使弹簧的每一环与相邻环因其经受相同量的力

而具有相同的距离。这导致电子显微镜中可见的重原子标记以如下方式被间隔开：该间距对应于其沿核酸聚合物的实际间距并且对应于其所连接的特定碱基的间距。而且，可以基于所述恒定间距来确定未经标记的碱基的位置。

[0133] 在一个具体实施方式中，使用附着着 DNA 链的末端的工具将 DNA 悬挂。可以将该工具浸入含多个核酸链或单个链的溶液（通常为小液滴），核酸链会优选与工具的尖端结合，并随着将工具提拉出溶液，核酸分子就被悬挂于空间中，从而核酸链的第一末端处于溶液中而核酸的第二末端与所述工具相连，并且末端之间的区域（“悬挂区”）被悬挂于空间中。可以理解，DNA 分子的“末端”不必受溶液中该分子的物理末端（例如，5' 磷酸根和 3' 羟基）的限定。例如，140kb 的分子其一端的 20kb 可以处于液滴中，100kb 处于悬挂区中，而另一端的 20kb 与针结合。将核酸链悬挂以使核酸聚合链的碱基伸长，从而具有恒定的碱基的碱基与碱基间距。具体而言，碱基与碱基的间距或周期性在碱基之间为约 3 Å~约 7 Å，其可从每个磷酸根的中心至中心测定。

[0134] 在一个实施方式中，液滴的体积可以为约 0.5 μl ~ 约 50 μl，有时体积为约 1 μl ~ 约 25 μl，有时体积为约 1 μl ~ 约 15 μl，有时体积为约 1 μl ~ 约 10 μl，而有时体积为约 1 μl ~ 约 5 μl。

[0135] 在一个实施方式中，可以将单个核酸聚合链悬挂于工具与液体之间的空间中。可以使核酸聚合链的碱基伸长以便具有恒定的碱基的碱基与碱基间距。碱基与碱基间距在碱基之间可以为约 3 Å~约 7 Å，且特别为约 5 Å。可以将核酸聚合链伸长以使链为线性。

[0136] 用于将 DNA 分子牵拉至空白空间中（例如，从液滴中牵拉出）的工具可以是各种装置的任何一种，条件是它可用于结合单一核酸链并将其悬挂于空间中。在某些实施方式中，该工具可以是尖针、中空针、或磁性探针所采用的具有结合核酸的亲合力的小的（即直径小于约 300nm）磁性颗粒。

[0137] 在一种方法中，所述工具是尖针。尖针可由如玻璃、金、钨、PMMA、聚苯乙烯、PVC、硅或任何其它适合的可制成极尖的针的物质等材料构成。针尖可以通过本领域普通技术人员已知的技术容易地制备，例如，使用标准拉针器（pipette puller）、微细加工（Handbook of Microlithography, Micromachining & Microfabrication, P. Rai-Choudhury, SPIE OptiactEngineering Press, 1997）、生长或模塑和浇铸。例如，玻璃针可以通过在乙醇火焰上加热玻璃纤维（直径大约与微毛细管相同）的中部并从该纤维的任何一端进行拉伸而制成。替代性地，可以使用标准拉针器。在一个实施方式中，如图 2 所示，针 200 具有包括直径小于约 200nm 的尖端 206 的近端 202、远端 204 和伸展于近端与远端之间的轴 208。尖端 206 的直径为小于 1 μm、至多约 1 μm 或大于约 1 μm。尖端直径可以使用 SEM 确定。端部曲率半径可以以几何一致性方式测定，例如，可以鉴定出尖端处的圆形区域并可插入与该曲率匹配的假想圆，然后通过测定从该圆的圆心至圆周的距离来测定其半径。在某些情况下，针的末端处的圆弧仅大致近似于圆的外边缘（图 3）。在其它情况下，针的末端被折断而类似于阔约 20nm ~ 约 500nm 的平面台面。优选直径可以为小于约 300nm。

[0138] 可以通过涂覆优选与核酸末端结合的材料来将针尖功能化，所述材料可以包括但不限于 PMMA、聚苯乙烯、PVB、如硅烷化等化学处理、与基因组序列或限制位点伸出物互补的寡核苷酸或多核苷酸（寡聚物可以拥有如肌苷等简并匹配碱基，从而允许更大的选择范围）、具有对特定序列或结构的高亲和力的适体、具有对如生物素等分子的高亲和力的抗生

物素链霉亲和蛋白或其它蛋白、或者任何其它适合的可特异性地与核酸末端相连的材料。在一个具体实施方式中,可以通过将针尖浸入约 0.5% 的 PMMA 丙酮溶液并在丙酮饱和的空气中干燥来用 PMMA 涂覆针尖。

[0139] 在另一个实施方式中,可以用第二涂层来涂覆针尖以限制核酸可以结合的区域,且该第二涂层可以包括如辛基硫醇、壬基硫醇、十六烷基硫醇或其它线性烷基硫醇链等材料。在图 2 中,涂层 1 与 DNA 的末端结合,而涂层 2 不结合或较不积极地结合。例如,在一种方法中,可以将金针尖浸入聚苯乙烯溶液,然后使聚苯乙烯仅在极尖端处在电子束中交联。随后,可以通过本领域技术人员已知的方法(例如,浸入丙酮或氯仿)来将未交联的聚苯乙烯从针的其余部分除去。然后可将针浸入辛基硫醇溶液,该辛基硫醇溶液在金上而不在聚苯乙烯上形成多个单层。DNA 的末端不会与辛基硫醇区域结合。

[0140] 在再一个实施方式中,可以使用磁性纳米颗粒寡核苷酸来特异性地标记核酸聚合链的末端。在该方法中,将具有磁化末端的工具浸入含有经磁性标记的核酸聚合链的溶液中并与经纳米颗粒标记的核酸聚合链的单个分子的末端结合。随着磁化工具被拉出溶液,核酸分子被伸长并悬挂于空间中从而核酸链的第一末端处于溶液中而核酸的第二末端与该工具相连。核酸链被悬挂来使得核酸聚合链的碱基被伸长从而具有恒定的碱基的碱基与碱基间距,且具体而言,该碱基与碱基间距在碱基之间为约 3 Å~约 7 Å,且特别为约 5 Å。

[0141] 在一个实施方式中,尖工具可以是中空针。特别地,可以采用本领域已知技术来制造中空微型针,从而可以经由内孔(bore)泵入核酸溶液。随后,可以将该微型针与具有对核酸末端的亲和力的载体基片接触或与更尖的聚合物涂覆固体针接触,以便直接从中空内孔中拉线。因此,中空微型针可以充当直接拉线工具并充当精确溶液控制用通道。所述内孔可以通过本领域已知技术制成,从而仅一条链会沿内孔程度方向进入并协助控制核酸线(nuclear acid thread)浓度。这种使用中空针的技术与使用尖工具将 DNA 从溶液中拉出相一致。使用中空针的区别在于溶液的表面和形状恰巧极小并受到中空针壁的限制。

[0142] 将尖针功能化后,将核酸“拉线”或附着于尖针上。在一个实施方式中,通过将功能化尖针浸入和浸出 DNA 聚合物溶液来进行 DNA 拉线,从而如图 4 所示将 DNA 链拉出至空白空间中。针尖可以以约 1nm/小时~约 10m/s 的速率、特别是 1 μm/s~约 10mm/s 的速率移入和移出溶液。应该理解,被拉出的 DNA 链保持与 DNA 溶液表面正交。纳米定位器、包括压电执行器(如 AFM 中所用那些)可以用于以次埃级精度来控制针的位置和运动(使用高品质反馈机制)。纳米定位器是本领域中已知的,例如美国专利第 5,903,085 中所述,且可商购自 Piezosystem Jena(德国)。影响有效的单链拉线的变量包括溶液温度和 pH、周围环境的湿度、溶液中的标记 ssDNA 浓度、针浸入和浸出溶液的速度、针在溶液中的停留时长、针的尖锐度以及针浸入溶液时的深度和进入角度。

[0143] “浸入”距离是可以控制与针相连的 DNA 量的变量(图 5)。例如,对于浓度为约 0.01ng/μl~约 0.5ng/μl 的高分子量 DNA 而言,针浸入的深度可以为进入溶液约 1 Å~约 20 μm。在一个实施方式中,采用浅浸入(图 6)。采用浅浸入可以将针以极短的距离(例如,小于约 1 μm)量置于溶液中以便限制可供链结合的针表面积。可以将多个针和/或自动系统用于针的拉线。

[0144] 另一个变量是针在溶液中停留的时间量。与尖端相连的核酸分子的数目可以通过控制允许尖端在溶液中保持的时间量来调节。对于给定溶液而言,更长的针留置时间会导

致更多的附着,而更短的时间会导致更少的附着,所述附着是分子在溶液中扩散并与针相连所需的时间的函数。该时间范围可以从少于约 1ns 至数秒或数分钟变化。

[0145] 例如,可以将大量超尖针置于单个纳米定位器驱动的载体上并间隔紧密从而单次浸入运动能同时拉出大量链并具有足以避免链间的干扰的间隔。并联的针或“针床”也可以通过本领域技术人员已知的微细加工技术制成。

[0146] 一旦核酸附着于针上,即将针拉出溶液,由此将核酸碱基拉伸至空白空间中以确保碱基之间的恒定间距。伸长的核酸的碱基与碱基间距应为约 $3\text{ \AA}$ ~约 $7\text{ \AA}$,且特别为约 $5\text{ \AA}$ 。然而,不应认为碱基的间距被专门限制于该范围,因为适当的碱基间距将取决于多种因素,诸如是否通过使用搁板而将核酸进一步拉伸(如下所述)。作用于空气-水界面与针尖之间的力自身可以导致约 $5\text{ \AA}$ 的碱基与碱基间距或约 65pN 的力。空白空间中特定伸长链的总长度将对应于链中碱基数乘以平均碱基与碱基间距。例如,假设碱基与碱基间距为 $5\text{ \AA}$,长约 10,000,000 碱基的链可以被拉伸至约 5mm 的长度。可以将核酸链拉伸至约 20nm 的长度,有时为约 100nm 的长度,有时为约 200nm 的长度,有时为约 $1\text{ }\mu\text{m}$ 的长度。应该注意,必须考虑到具体的核酸库中的长度分布。如此才能使不会将链从溶液中完全拉出,从而空气-溶液界面处的弯月面力能继续在远离针尖的链末端处施加张力。对于含有长度较长的核酸聚合物的溶液而言,相对于含有长度较短的核酸聚合物的溶液,核酸链可以被拉出至离溶液更远的距离。可获得恒定碱基与碱基间距的足够伸长度是核酸分子的分子量的线性函数。在具体实施方式中,可以将核酸链拉伸并伸长至至少约 $2\text{ }\mu\text{m}$ 的长度,更特别地是至少约 $25\text{ }\mu\text{m}$ 的长度,再更特别地是至少约 $50\text{ }\mu\text{m}$ 的长度,进而更特别地是至少约 $100\text{ }\mu\text{m}$ 的长度。

[0147] 由于参与核酸拉线的力不会破坏共价键,因而核酸不会在拉线过程中断裂。作用于空气-水界面处的弯月面力强至足以瓦解二级结构,但尚未强至足以破坏共价键。如果将拉线装置安置并编程来使得施加超出将 DNA 拉出溶液所需的额外力,则必须谨慎而勿使 DNA 断裂。

[0148] 在许多实施方式中,理想的是使蒸发最小化和/或控制液滴的位置和角度。在最简模式中,将液滴保持静止并小心地安置和配置 PDMS 保持器部件。在图 21 中,图片 I 显示了液滴向网格保持器部件的靠近。图片 II 显示了保持于两个 PDMS 部件之间而使液滴会保持拉线表面前端的液滴。即使液滴蒸发,拉线表面前端也会保持不变,而仅是滚动表面(receding surface)前端移动(图 III)。在另一种简单模式中,在拉线装置周围构建湿度腔从而实现高湿度(例如,接近 100%)状态且液滴不会蒸发。

[0149] 替代性地,可以通过使用光学或磁性珠将核酸链核酸链在溶液中附着于两点而将核酸链从溶液中除去并悬挂于两点之间(Bustamante, C. 等,421(6921)NATURE 423-427(2003))。然后将该链冻干并使冰升华,从而留下悬挂的单链,其后可将该单链转移至成像基片。在另一个实施方式中,使个体核酸链与单个磁性颗粒结合,然后将该磁性颗粒从液滴中取出(例如,使用磁性探针)以使 DNA 链延伸至空间内。

[0150] 与成像基片的附着

[0151] 在步骤 110 中,一旦将核酸链拉线于空白空间中的针与液滴之间且核酸链不再由溶液或缓冲液包围,则可以将该链直接置于基片(如载体基片或成像基片)上以通过电子显微镜进行成像。一种实现上述过程的方法被称作“搁板拉线”,其中将伸长的 DNA 链置于

成像基片上或在分开的步骤中置于载体基片上并转移至成像基片上,参见图 7A。相关的方法被称作“间隙拉线”,其中在分开的步骤中将 DNA 链悬挂穿过载体基片中的间隙并转移至成像基片,参见图 15。

[0152] 其后,可将链以各种各样的取向进行放置,包括二维和三维取向,条件是链在“悬挂区”中不(或极少发生)重叠或交叉。最方便和标准的 EM 样品制备方法是二维的,通常使多条链彼此平行安置。参见例如,图 8。在该设置下,可以易于使样品与空白或“非样品”空间之比最大化。一般而言,链间间距越近越好,条件是链不会相距过近而使其干扰其相邻链的可见度和鉴定。次纳米精度的定位器可商购,且非常方便的配置是将链以相距约 2nm ~ 约 10nm 的平行线放置。在另一个实施方式中,将链放射状安置。

[0153] 在一个实施方式中,可以将多个悬挂链以基本显性的取向放置。图 12 显示了通过针 706 从含核酸链 704(如 DNA)的溶液 702 的液滴中将核酸链 704 置于基片 1202 上。可以通过依次重复受基本编程的纳米定位器控制的浸入、浸出、放下、必要时沿载体基片拖动针尖以便将链沉积、提起和在链间的所需距离中平移的针运动而形成紧密相间的核酸链(如 DNA 链)的大平行阵列。在工业高通量规模上,可以通过这种方式、特别是使用大的平行针阵列而将上百万的链置于一个载体基片上。

[0154] 由于所述方法比其它方法提供的 DNA 链更直(更加线性),可以使用拥挤的阵列并且可确定准确的序列。阵列可以包含约 2,000,000 ~ 10,000,000 个基本线性链和/或平行链且具有长度为约 1 μm ~ 约 5mm 的悬挂区和约 1nm ~ 约 10nm 的间距,从而产生约 1 碱基/nm² ~ 1 碱基/5nm² 的密度。阵列可以具有例如大于 5 个、大于 10 个、大于 100 个、大于 1000 个或大于 10,000 个 DNA 链。

[0155] 显然,本发明的方法能产生线性(直线)双链和单链核酸链。线性(linearity)可以依据围住链(或其线性部分)的假想方框、或更方便的是依据围住链在承载基片上的二维投影(或链自身)的假想矩形的维度来描述。链的线性部分通常长度为至少约 2 μm ,更通常长度为至少约 5 μm ,进而更通常长度为至少约 10 μm 、20 μm 、30 μm 、50 μm 或 100 μm 。在某些情况下,链的线性部分将接近链的整个长度。在一个实施方式中,所述矩形如下:如果绘制最小的可能假想矩形来围住链或线性部分从而所有部分均处在矩形之内,则该矩形的长宽比为不小于 100 : 1、或优选至少为 160 : 1、或优选至少为 200 : 1、更优选至少为 500 : 1 且再更优选至少为 2000 : 1。例如,对于长 5 μm 的线性部分,绘制来围住平面表面上的线性拉伸链的方框将宽约 30nm(比例 1 : 160)或宽 2.5nm(比例 1 : 2000)。对于长 10 μm 的链,边界方框可以为 10 μm × 62.5nm 从而使比例为 160 : 1。对于长 3 μm 的线性部分,边界方框可以为 4nm 从而使比例为 500 : 1。在三维中,方框的第三(深度)维度将在宽度维度的 5 倍以内、优选 2 倍以内、且最优选与宽度大致相同。由于使用本发明的方法安置的链是直的(通常基本沿着悬挂区的整个长度),可以将阵列中的相邻链彼此非常紧密地放置,例如,与其它相邻链相距约 2nm、相距约 3nm、相距约 4nm、相距约 10nm、相距约 20nm 或相距约 50nm。在某些实施方式中,多个相邻链相距小于 50nm,优选相距小于 20nm,更优选相距小于 10nm 或相距小于 4nm(例如,相距 2nm ~ 50nm、相距 2nm ~ 20nm、相距 2nm ~ 10nm、相距 4nm ~ 50nm、相距 4nm ~ 20nm 或相距 4nm ~ 10nm)。在一个具体实施方式中,将线性链放置以使其不交叉。

[0156] 搁板拉线

[0157] 在一个方法中,可以采用搁板拉线来将 DNA 直接置于成像基片上。参见图 7,图片 I ~ V 所示。成像基片可以制成成像薄膜,成像薄膜由如石墨烯单片、超薄碳膜、氧化铍或氮化铍、水冰(需要冷冻电子技术)和其它合适的固体低 Z(电子透明)固体形式。由于这些超薄膜非常纤巧,它们被置于网格或网的顶部或由其制成,所述网格或网由 TEM 网格上的标准弗姆瓦微孔膜、SiN 膜(DuraSiN by Protochips, Inc.) 中的微细加工孔或简单制造的多孔网格构成。图 7 显示了含有多个标记的核酸链 704(为清楚起见仅显示出一个)的溶液 702 的液滴、针 706、处于 PDMS 块 710 顶部的 TEM 网格 708。如图 7 的图片 I 所示,将针 706 浸入溶液 702 的液滴中。在图片 II 中,将针 706 从溶液 702 拉出从而将核酸链 704 伸长至空白空间中。在图片 III 中,使伸长的核酸链 704 与 TEM 网格 708 接触。核酸链 704 从针 706 释放,如图片 IV ~ V 所示。

[0158] 在另一种搁板拉线方法中,如图 8 所示,将载体基片 710 放置来紧邻核酸聚合物溶液 702 的液滴,从而使悬挂于液滴与尖针之间的核酸链向下接触载体基片。然后,通过如下所述的“转印”将多个核酸链转移至成像基片。图 8 的图片 I 示意性显示了将核酸链 704 沉积于 PDMS 载体基片 710 上的单个针,而图 8 的图片 II 示意性显示了将多个核酸链 704 一次一个地沉积于经载体膜覆盖的预制 TEM 网格 708 上的单个针 706(显示了处在其运动时间轴路径的两个位置的针)。

[0159] 如图 9 所示,核酸链总是正交于液滴表面。对溶液表面相对于载体基片的角度和针的运动进行控制从而使空白空间中的核酸 704 在被置于基片 708 上时沿其长度与该基片基本接触(图 10)。在核酸被“拉出”溶液后,可以将核酸伸长。图 11 的图片 I ~ IV 显示了通过近端放下接着进行进一步拉伸来伸长核酸。在图 11 中,核酸具有夸大的松弛性,以显示通过机械力比单纯弯月面力的更大拉伸。“A”和“B”可以代表载体的截面图,例如,EM 网格中的两条(微米规模)或 PDMS 的两个带(微米规模)。“A”和“B”也可代表平面基片上的两个位置。

[0160] 图 13 和 14 显示出在伸长和放置 DNA 链时可以考虑的几个参数。在一个故障模式中,不正确的溶液表面 / 载体基片 / 针运动角度的考虑将引起载体基片与悬挂的链在针尖与基片接触之前的不受控接触,而这反过来会通过过度拉伸而造成链的断裂(图 13, 图片 II)。在另一个故障模式中,错误校正的溶液表面 / 载体基片 / 针运动角度会使得链无法与基片实质上接触,从而使得该链的相当大的一部分仍悬挂在溶液表面与针与基片的接触点之间的空白空间中(图 14)。

[0161] 间隙拉线

[0162] 在另一个实施方式中,可以采用间隙拉线来将核酸链置于载体基片上。间隙拉线可以以如图 15 中示意性图示的下述方式进行。图 15 显示了具有间隙 1504 的 PDMS 块 1502、跨越间隙 1504 的核酸链 1506 和成像基片——在该情形中为覆有碳膜或其它低 Z 膜的超平硅网格 1508。首先,将核酸液滴置于小 PDMS 块中的间隙旁。通常,PDMS 块的长度为约 3mm 且宽约 3mm。通过如上所述的尖针拉线使核酸跨越该间隙。核酸跨越间隙 1504。将覆有碳膜、多孔金膜或连续碳铍膜、其它低 Z 膜或成像薄膜(如上所述)的超平硅网格 1508 放置而与拉伸跨过间隙 1504 的核酸接触,从而使得核酸被转移至网格 1508(也称作“刷印”)。如本文所用,术语“刷印”是指使跨越载体结构体上的间隙的一个或多个核酸链与可以为另一载体基片或者成像基片或成像载体薄膜的结构体接触。具有跨越间隙 1504 的 DNA 1506

的 PDMS 块 1502 还可以被翻转并通过如下所述的转印方法转移至成像载体膜（未显示）。

[0163] 转印

[0164] 如上所述,可以将 DNA 链置于载体基片上并随后转移至成像基片或成像薄膜。完成上述操作的一种方式是通过“转印”。转印是本领域内已知的,且由 Nakao, H. 等,125J. Am. Chem. Soc. 7162-7163(2003) 一般性描述。可以采用转移来将核酸链置于载体基片或成像基片上,如图 16 所示。图 16 示意性展示了长 3mm 宽 3mm 的 PDMS 块 1602、含有核酸 1606 的液滴 1604 和针 1608。将针 1608 浸入液滴 1604 内,核酸 1606 与针 1608 的尖端结合。将针 1608 从液滴 1604 中拉出,从而将核酸链拉伸至空白空间。伸长的核酸链 1610 与 PDMS 块 1602 相连。重复这一过程直到将多个核酸链置于 PDMS 块上（小图 6）。必须格外谨慎（以较大的链间间距）以免一个拉伸的链干扰（即接触或交叉）此前拉伸的链。以待已将所需数目的核酸链放置于 PDMS 块 1602 上,则将 PDMS 块 1602 翻转并置于成像薄膜或其它转移基片 1612 上。将 PDMS 块 1602 从膜 1612 除去,而 DNA 留在膜 1612 上,由此多个核酸链被转移至膜 1612。该膜可以为碳、铍或其它低 Z 载体,其在核酸转移时被沉积于新切割的盐晶体或云母片上,然后置于 TEM 网格上以进行成像。

[0165] 替代性地,可将核酸链置于载体基片上并转移至另一个载体基片上以用于储存、运输或其它目的。

[0166] 成像薄膜可以是成像基片,它是由但不限于碳、硼、铍、铝或其它低 Z 元素和 / 或其氮化物和氧化物构成的薄膜或如前定义的成像薄膜。这些膜可以通过已知技术制备,例如,沉积于切割的盐晶体或云母上。在一个具体方面,采用超薄（约 1.5nm）碳膜。在另一个方面,成像基片是覆有薄（约 1.5nm 厚）支撑碳膜的超平硅 TEM 网格。也可将成像薄膜置于弗姆瓦微网涂覆的 TEM 网格或具有规则或不规则孔或孔隙的机加工硅网格上。

[0167] 在一个实施方式中,可以将至少一个伸长的核酸聚合链置于平面基片上。所述至少一个伸长核酸聚合链可具有在约 1000 个碱基对的长度上的恒定的碱基与碱基间距。可将膜置于所述至少一个伸长核酸聚合物的上方从而该至少一个伸长核酸链聚合物被夹在平面基片与膜之间。膜可以由碳或低 Z 元素构成。平面基片可以由如 PDMS、碳、硼、锂、氢、铍、铝、氮化物、氮化物氧化物及其组合等材料构成。

[0168] 上文所述的方法不限于核酸聚合物,而是也可用于各种各样的其它长的（无支链）高分子量分子。例如,可以将包括但不限于纳米管（例如,氮化碳、硼和氮化硼等）、氨基酸链微管、肌动蛋白微丝、具有重复单元의 其它长线性聚合物和其它聚合物的其它高分子量聚合物拉线于合适工具上并附着于合适的基片。具体而言,所述方法可以用于在其末端（端基）处与针或其它结合工具差异性结合的线性聚合物。在某些实施方式中,可以将线性分子在其一端或两端修饰从而末端优选与工具结合。例如,可以使用本领域技术人员已知的技术使聚合物末端与 DNA 复合,然后其可在其末端如上所述的工具优选结合。

[0169] 稳定化

[0170] 核酸可能受到电子显微镜产生的电子束的损伤。为此在某些实施方式中,一旦将标记的核酸链置于成像基片上,则可在成像之前将该标记核酸链稳定化。例如,可以通过蒸发、溅射或直接沉积预制膜来在样品上放置另外的碳或其它低 Z 原色或聚合物。可以通过碳或铍的超快 PLD-UHV 来完成碳或其它低 Z 材料的间接蒸发。这种附加层（即,面层（topcoat））的存在增加了下层的核酸的稳定性。图 17 示意性显示了具有多孔网 1704 的

TEM 网格 1702、基底层 1706、核酸 1708 和顶层 1710。

[0171] 在一个实施方式中,通过将面层蒸发至核酸上来使核酸稳定,所述面层通过在使沉积原子冷却的气体氛围中进行脉冲激光沉积而制备,但使压力、靶标与样品距离、脉冲长度、脉冲频率和脉冲注量 (pulse fluence) 等条件优化来给出将标记的核酸聚合物嵌入的固体均质膜。这些条件会使得沉积原子在到达基片的途中彼此之间发生最少的反应从而产生完全将标记嵌入的更致密的膜。面层的用途在于通过可使序列数据得到确定的方式使标记和 / 或 DNA 稳定化并避免运动或损伤。

[0172] 成像薄膜必须极薄,以便在 TEM 中得到良好的单原子图像。通过标准技术制成的膜在暴露于实验室空气中后往往受到污染。这种污染的累积使得膜变厚,这会妨碍单个原子和簇的成像。除了使膜变厚以外,来自空气的污染物还能导致膜在与电子束相互作用时的结构不稳定性。必须进行严格清洁和维护以确保这种累积不会发生。对于本文所述的所有方法而言,所有步骤都应优选在受控的无烃环境 (例如,纯氮气、氩气或其它惰性气体) 下进行。一旦 DNA 被悬挂并置于成像基片或转移基片 (如 PDMS) 上并从溶液中移出,所有其它步骤最好在受控环境中继续进行,所述受控环境优选为清洁的无烃气体,但更优选为 UHV (10^{-10} 托)。这可以包括在受控气氛环境下进行拉线、将 DNA 链放置于 UHV 中的膜上,和在纳米技术、半导体制造等领域内的技术人员已知的其它清洁技术 (K π shnan, S. ;Laparra, O “Contamination issues in gas delivery for semiconductor processing” Semiconductor Manufacturing, IEEE Transactions, 第 10 卷, 第 2 期, 1997 年 5 月, 第 273-278 页 ;William Whyte, Clean room Technology :Fundamentals of Design Testing and Operation ;Dorothy Hoffman, Handbook of Vacuum Science and Technology)。

[0173] 来自显微镜自身的污染也是可能限制单个原子的可见性的关键因素。必须谨慎地保持非常清洁、干燥的成像用系统。这包括具有冷阱、多重离子泵、涡轮泵、钛升华泵、吸气泵、阶段加热器 (stage heater) 和加样锁 (Dorothy Hoffman, Handbook of Vacuum Science and Technology)。

[0174] 步骤 112 显示了本发明的一个实施方式,其中通过电子显微镜将 TEM 网格上的标记核酸成像。本发明不应被视为在此步骤限于 TEM 网格,而可以是本领域技术人员已知的任何成像基片。此外,可以使用任何适合的电子显微镜来使标记的位置可视化,所述电子显微镜优选具有合适的使用 HAADF STEM 的像差校正器。

[0175] 在步骤 114 中,产生并分析了核酸序列数据。根据本发明的一个实施方式,图 18 显示了用于分析核酸序列的系统,其可以包含电子显微镜 2202、处理器模块 2204、至少一个存储器模块 2206、分析器模块 2208、用户界面 2210 和网络接口 2212。可以将电子显微镜 2202 设置来产生代表电子致密区的电信号。将分析器模块 2208 设置来基于由电子显微镜 2202 产生的电信号分析核酸序列。所述至少一个存储器模块 2206 被改造来存储代表电子显微镜和 / 或分析产生的核酸信号的电信号中的至少一个。将用户界面设置来使得用于能与分析互动。在一个具体实施方式中,所述至少一个存储器模块包含用于储存分析和用于储存代表核酸序列的电信号的不同存储器。在另一个实施方式中,分析器 2208 和所述至少一个存储器模块 1806 远离电子显微镜,并且通过网络与电子显微镜相连。一个或多个模块位于同一装置 (例如,计算机或处理器) 中以便执行各种功能。替代性地,所述模块可以执

行不同功能的不同结构部件。

[0176] 在一个更具体的实施方式中,如图 19 所示,可以从商购系统或定制系统采集数据分析。图 19 中的系统可以包括电子显微镜 1902(商购或定制显微镜)、处理器检测器 1904、图像识别计算机 1906、至少一个存储器模块 1908、分析器模块 1910、用户界面 1912(概念性视图或屏幕)和可选的网络接口(互联网或局域网)。可以将电子显微镜设置来产生代表核酸序列的电信号。将分析器模块 1910 设置来基于由电子显微镜 1902 产生的电信号分析核酸序列。在某些实施方式中,存储器是选自硬盘或软盘、光学介质、光盘(CD)、数字多用光盘(DVD)、半导体介质和闪存的介质。

[0177] 手动或优选地使用图像识别系统来组装最终序列信息。核酸间距信息产生自高数据输出检测器,其可以是例如 CCD 检测器、CMOS 或 PMT。采用算法来从接收自高数据输出 CCD 检测器的信息确定例如钽标记的间距,并相应地对于给定标记反应批次而言确定特定碱基的序列。从 CCD 接收的信息被储存于存储器上并分析。存储器可以包括例如,如常规硬盘或软盘等磁性介质、如光盘(CD)或数字多用光盘(DVD)等光学介质和/或如闪存等半导体介质。序列组装用算法(计算机程序)是公知的。可以在本发明中使用或改编使用的程序包括如 DNA Naser 和 Cap3,这是最常见的本领域中使用的序列组装软件程序,例如,题为“具有提高的准确性和阅读长度的 DNA 序列比对用方法(Method for Alignment of DNASEquences with Enhanced Accuracy and Read Length)”的美国专利第 6,760,668 号和题为“序列比对重要性确定方法(Method for Determining SequenceAlignment Significance)”的美国专利第 6,988,039 号中公开的那些程序,本文通过参考特地并入其全部公开内容。也参见互联网 dnabaser.com/index.html (Heracle Software, Lihenthal, 德国)及 Huang, X. 和 Madan, A., 9GENOMERES. 868-877(1999)。具体的组装参数将部分取决于所用标记的性质。在一个实施方式中,例如,对于成像的每条链,用标记的特异性(例如,是否所有 T 被标记,是否所有 T 和所有 C 被标记)对包括每个标记的碱基和每个未经标记的碱基的相对位置的信息进行关联以便确定标记的碱基的位置序列和所存储的信息。将对于链及其推导互补物的存储信息与对于其它链及其互补物产生的信息进行比较并鉴定匹配。鉴定出相邻序列并组装而产生完整序列。

[0178] 当待测序的核酸来自此前已被测序的基因组(例如,小鼠、人、细菌、病毒)时,可以将初始序列数据(例如,不同链内的位置序列)与已知的参照序列进行匹配以加快分析。

[0179] 在一方面,本发明包括对储存于存储器中的核酸序列(A 的序列、T 的序列、G 的序列和 C 的序列)进行分析,其中所述序列通过上文所述的方法确定。在一方面,本发明包括接收成像数据(例如,标记碱基和未经标记的碱基的位置)、位置序列(可选的是其中至少一种碱基未确定的位置序列)或其它序列信息并处理数据以确定核酸样品的核苷酸序列。通常,数据以电子形式接收。在一个实施方式中,核酸序列是人类受试对象的基因组序列。在一个实施方式中,所述分析包括对一个或多个单核苷酸多态性的存在或缺失、拷贝数、变异体、得失位、重排或全基因组序列中的至少一者的确定。

[0180] 图 19 显示了带有标记的 T 的单链 DNA 链的说明性图像,简图说明了重标记的图案如何对应于所成像区域的碱基特异性部分序列信息。合并了用不同标记和/或反应条件对相同基本序列的多次成像的信息使得可以进行极为准确的序列确定。

[0181] 由于链已被伸直,故不必跟踪核酸主链,从而没有必要用独特的重标记对每个碱

基进行标记。图 20 显示了替代性制备方法所固有的模糊性。使用上文和下文具体实例中所述的本发明的方法将不会引入同步误差 (phasing error), 这是因为缺少的标记仅会被“阅读”作为空白点。

[0182] 可以通过生物信息学将多个标记图案合并来确定基本核酸序列。作为单分子放置控制的结果, 本发明提供了高效通量。由于以高度可预测方式安置的链不仅单独地 (即恒定的碱基与碱基间距) 而且可预测地将单条链的致密阵列并行安置, 图像分析将极为高效。本发明的方法还使得拉伸力可受控制, 由此导致对碱基与碱基间距的程度的优化。

[0183] 无须赘述, 据信本领域技术人员可以使用前述说明来将本发明利用至其最完全的范围。以下实例仅为说明性, 而非以任何方式限制本公开。

[0184] 实施例

[0185] 实施例 1

[0186] 使用本领域普通技术人员已知的技术从样品分离感兴趣的 DNA。然后将感兴趣的 DNA 分在 4 个溶液中, 即溶液 1、2、3 和 4。使每个溶液与 Os-bipy 以不同时长和不同的 Os-bipy 浓度反应, 以便获得不同的碱基特异性标记密度。

[0187] 使溶液 1 和 2 在 26°C 与 pH 8.0 的 TE 缓冲液 (具有 100mM Tris 和 10mM EDTA) 中的过量 4 倍摩尔的四氧化锇和 2,2'-二吡啶反应 20 小时, 这些条件标记了约 100% 的 T、约 85% 的 C、约 7% 的 G 和约 0% 的 A。使溶液 3 和 4 在与溶液 1 和 2 相同的条件下反应, 不同之处在于反应仅进行 15 分钟, 且仅使用过量 2.5 倍摩尔的四氧化锇和 2,2'-二吡啶, 这些条件标记了约 90% 的 T、约 8% 的 C、约 5% 的 G 和约 0% 的 A。然而, 在 Os-bipy 反应之前, 首先使溶液 2 和 4 进行亚硫酸氢盐处理以将未甲基化的 C 残基转化为 U。亚硫酸氢盐方案是本领域技术人员已知的, 而此处描述了此种方案的极简版本。将最终浓度为 0.05mM 的氢醌、33M 的亚硫酸氢钠和 3ng/微升的变性 DNA 加入离心管内, 封住该管并用铝箔屏蔽光, 在 55°C 温育 8 小时, 通过标准方法纯化 DNA, 并将 DNA 重悬于 TE 缓冲液 (pH 8.0, 100mM Tris, 10mM EDTA) 中以用于后续重原子标记。这使得可以通过对来自亚硫酸氢钠处理的溶液的序列与未处理的那些序列的比较来确定甲基化模式。根据参考文献: Jelen 等, 10GENPHYS AND BIOPHYS, 461。标记反应进行后, 通过超滤除去未标记的锇以使成像过程中的外源性重原子污染最小化, 并将 DNA 聚合物在 pH 8 的 TE 缓冲液中稀释至约 0.1ng/ μ l。

[0188] 通过在乙醇火焰中加热玻璃纤维并牵拉分开来制备尖针, 由此产生两个在其末端处的曲率半径小于约 200nm 的尖针。然后将一根针通过浸入 0.5% PMMA 丙酮溶液并在丙酮气氛中干燥而以 PMMA 涂覆。然后将该针粘至保持部件上或者夹在定位器如压电执行器的臂上 (例如, 可编程 AFM 硅悬臂) 以便控制针的位置和运动。

[0189] 将针浸入含有感兴趣的 DNA 聚合物的 DNA 聚合物溶液随后拉出以将 DNA 链伸长或从 DNA 聚合物溶液中“拉出”至空白空间中。伸长的 DNA 聚合链应该保持与 DNA 聚合物溶液的表面垂直。

[0190] 将 DNA 聚合物附着至覆有厚度为约 1.5nm ~ 约 5nm 的碳薄膜的超平硅 TEM 网格。硅 TEM 网格通过如下方式制备: 使用微细加工工业普通技术人员已知的技术, 将碳蒸发于硅片的一侧上, 然后蚀刻背面侧并确保碳膜平整。

[0191] 将网格置于 DNA 聚合物溶液的液滴旁边, 从而使悬挂于液滴与尖针之间的 DNA 聚合链可以在被完全“拉出”DNA 聚合物溶液之前与网格接触。仅当 DNA 聚合链被置于 TEM 网

格上时,使伸长至空白空间中的 DNA 聚合物与网格接触。考虑 DNA 聚合物溶液相对于网格的角度和针的运动以确保这一点。使用该拉线技术,可以将成千上万的链彼此平行置于 TEM 网格上。

[0192] 使用移液器尖或微定位器将 DNA 聚合物溶液拖动离开跨越位置,从而移动液滴所停留的碱基。然后蒸发约 0.5nm ~ 约 6nm 的碳以使 DNA 聚合物稳定。在受控环境室中进行跨越以使蒸发和灰尘最小化。

[0193] 在 Z 反差 STEM 模式的带有像差校正器的商购 Titan 80-300 (FEI Company, Hillsboro, OR) 中或其它合适的高分辨电子显微镜中进行成像。使用来自检测器的信息以确定钨标记的间距并相应确定感兴趣的 DNA 的序列。

[0194] 实施例 2

[0195] 以与上文具体实施例 1 中相同的方式执行相同方法,不同之处在于使 DNA 聚合物向下搭架跨越至 PDMS 片上,如图 17 所示。然后使 PDMS 与云母上的碳膜接触从而将 DNA 聚合物转移至碳上。然后将 PDMS 片提走,留下 DNA 聚合物。在云母上的碳膜上的 DNA 聚合物的上方蒸发约 1nm ~ 约 5nm 的碳。使碳膜漂浮于水上并由 TEM 网格拾取。然后如具体实施例 1 或具体实施例 2 中所述对聚合物进行成像,以确定序列信息。

[0196] 上文给出的实例仅为说明性而非旨在作为本发明的所有可能实施方式、应用或变形的详尽列表。因此,在不背离本发明的范围和主旨的情况下,对本发明所述方法和系统的各种修改和变化将对本领域技术人员显而易见。尽管已经结合具体实施方式对本发明进行了描述,但应理解,所要求的发明不应过分地受到这些具体实施方式的限制。事实上,对于分子生物学、免疫学、化学、生物化学或相关领域中的技术人员显而易见的对本发明的所述实施方式的各种改变应该包括在所附权利要求的范围之内。

[0197] 通过参考将上文所引用的所有参考文献和出版物的全部公开内容并入,其程度等同于通过参考将每个参考文献和出版物单独并入。

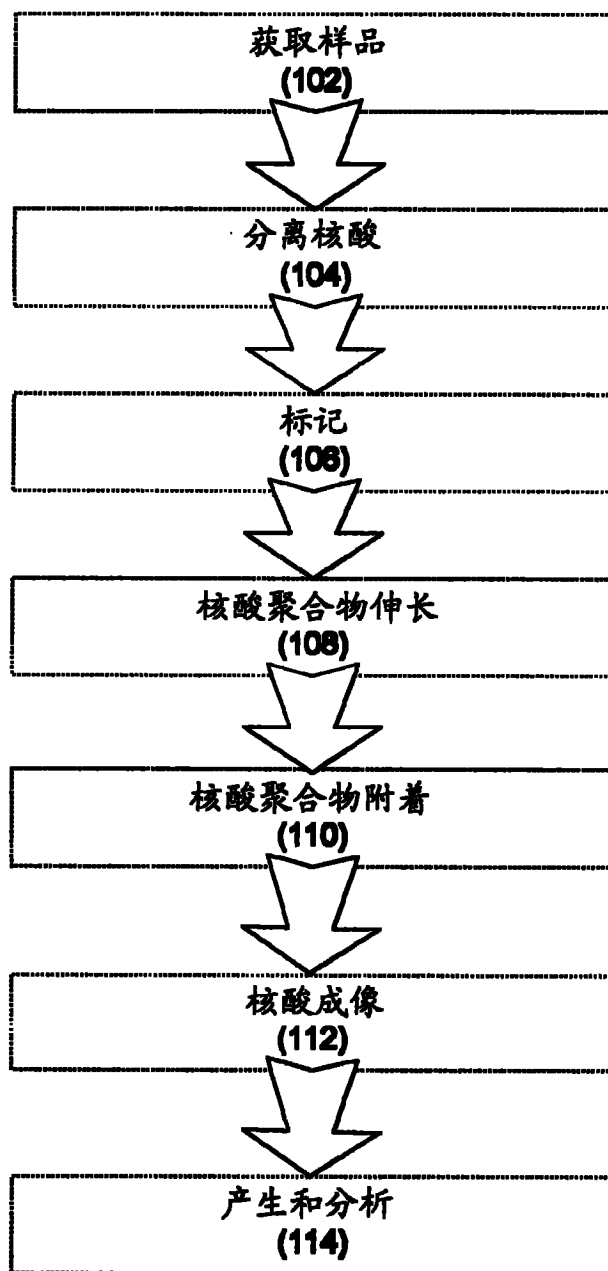


图 1

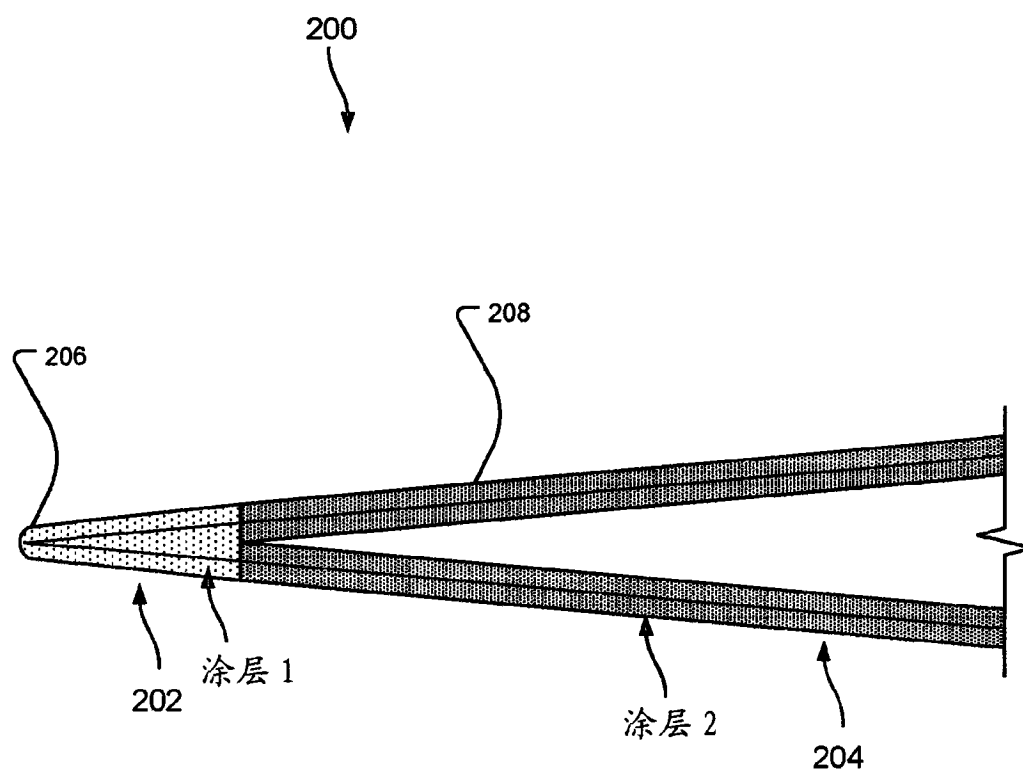


图 2

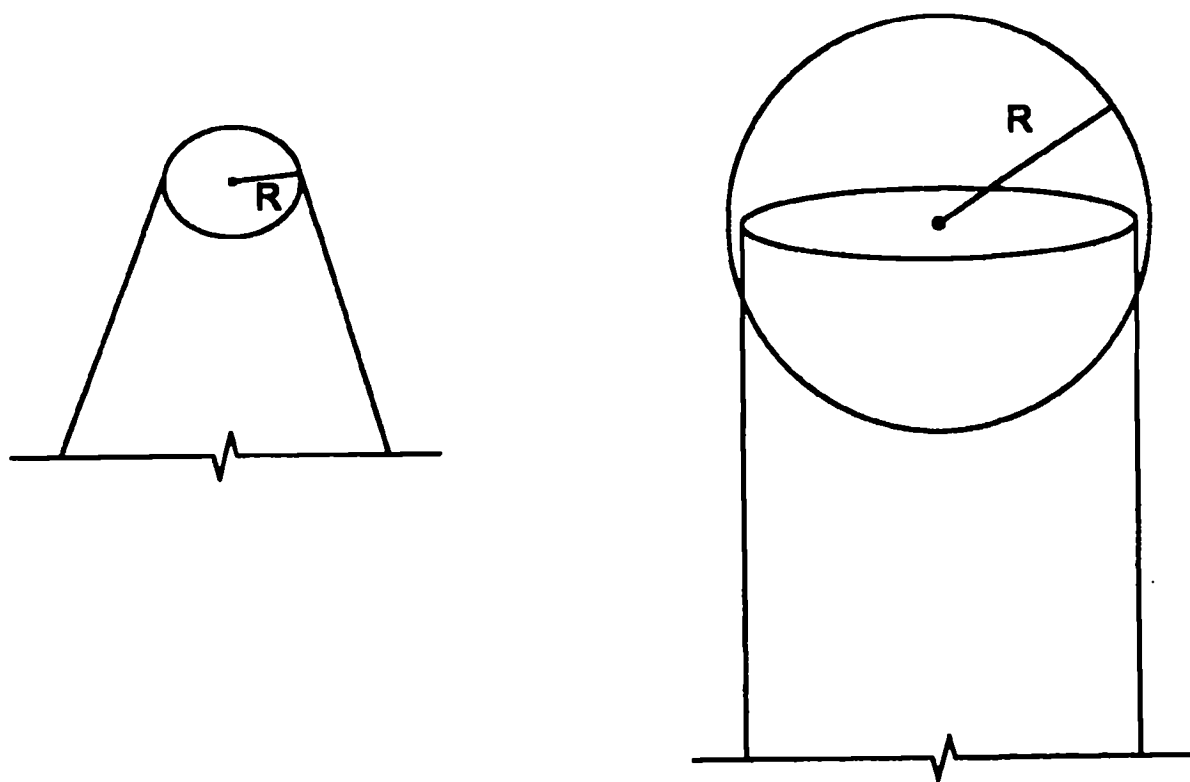


图 3

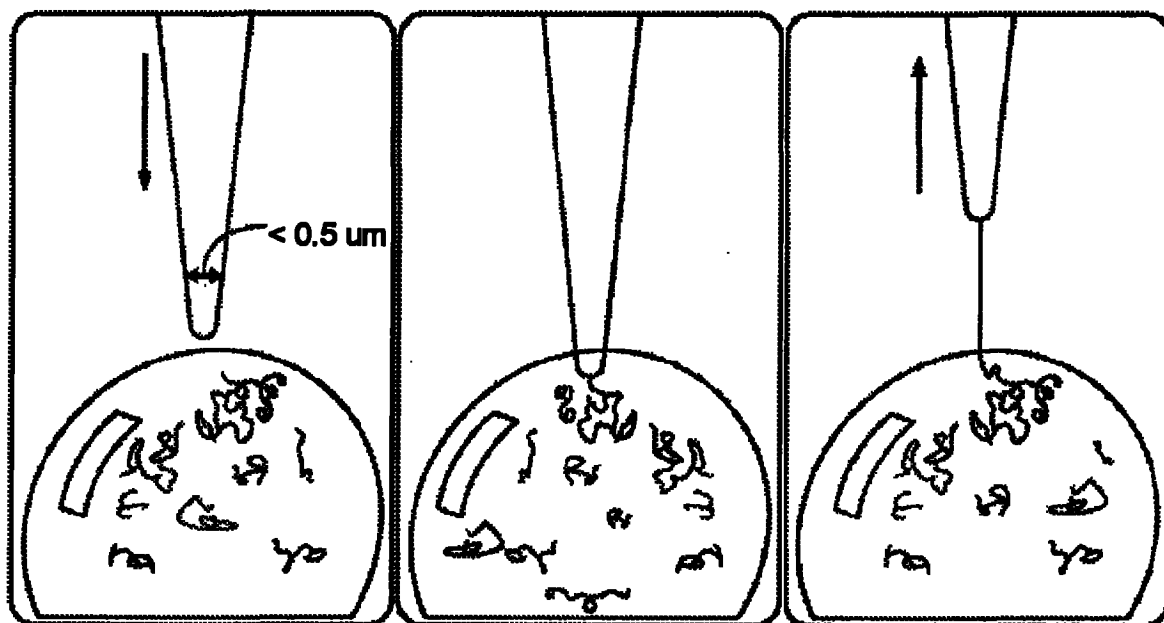


图 4

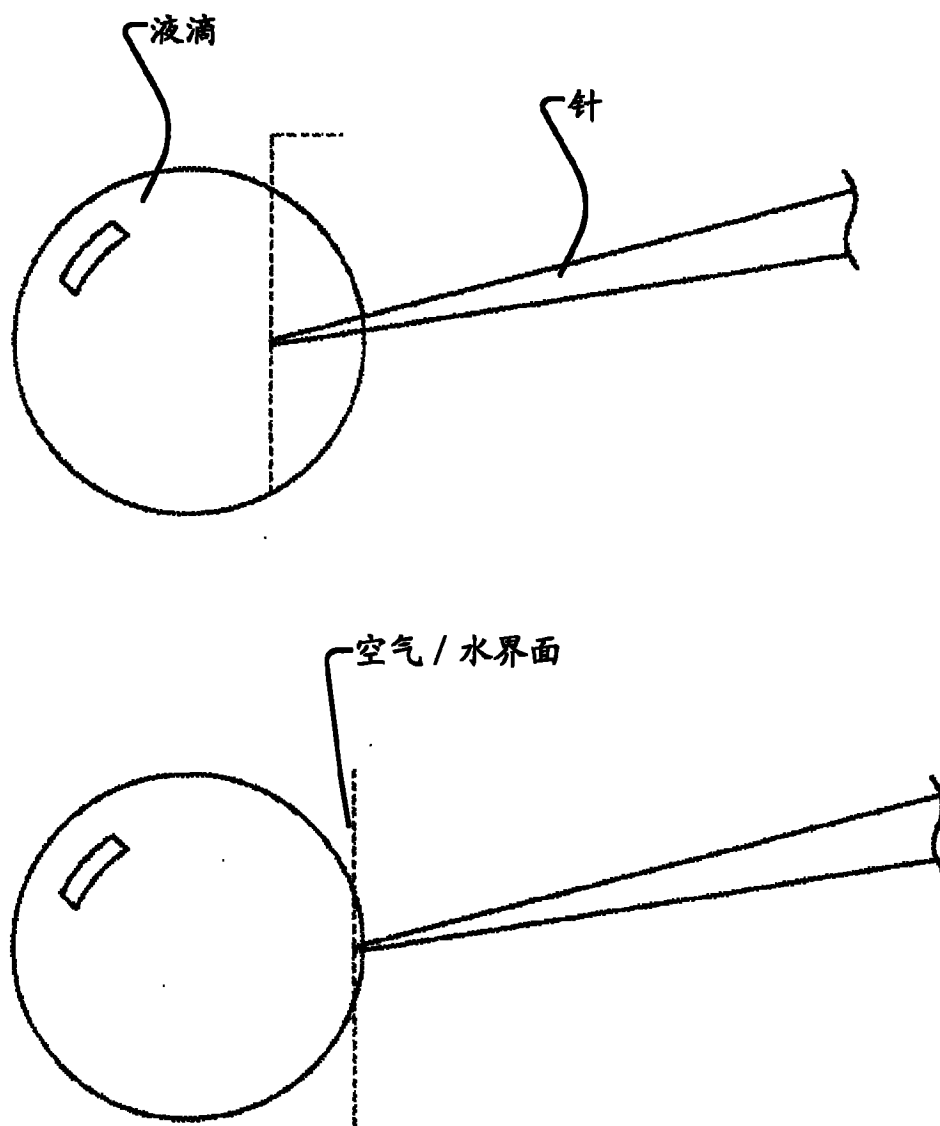


图 5

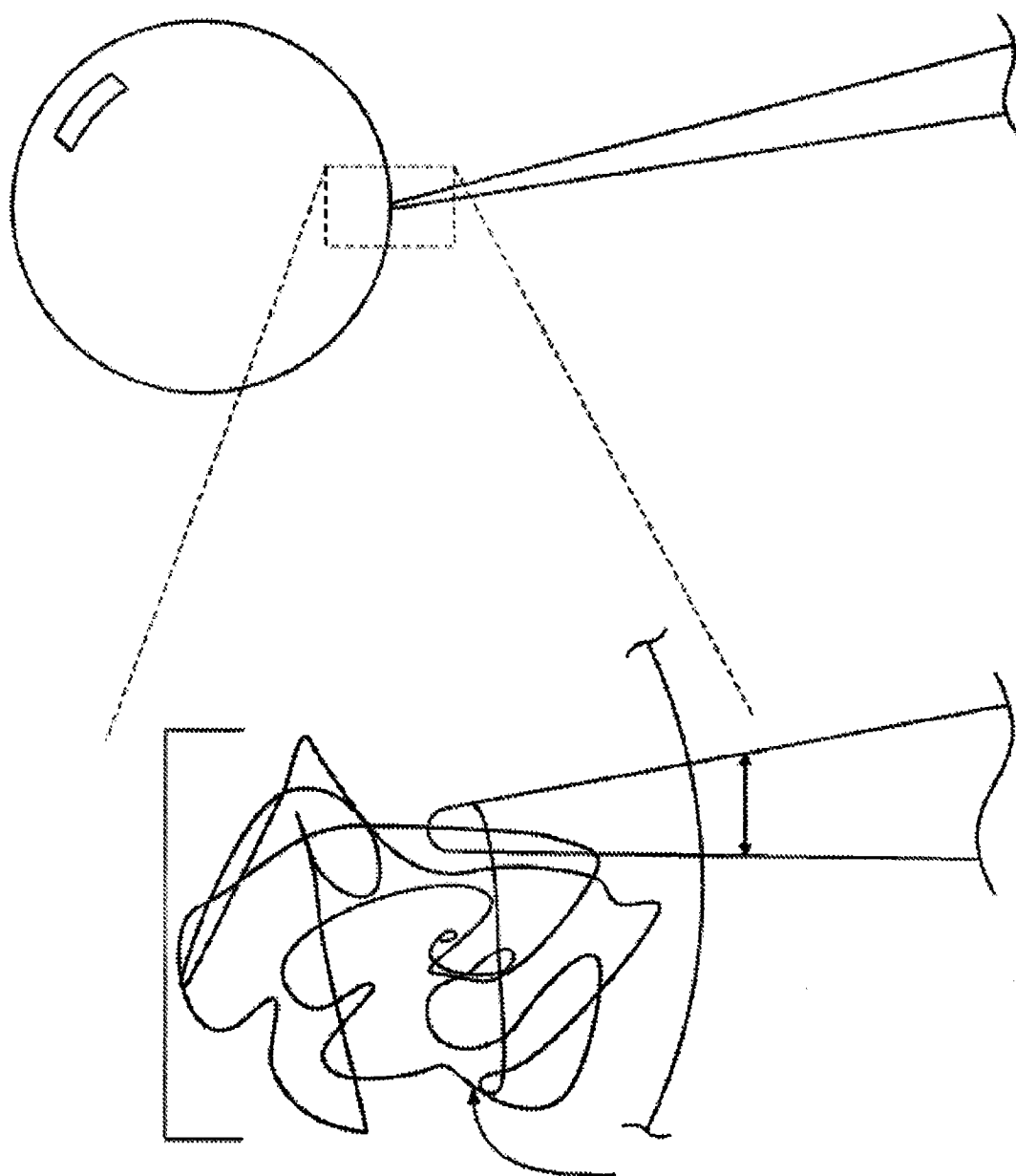


图 6

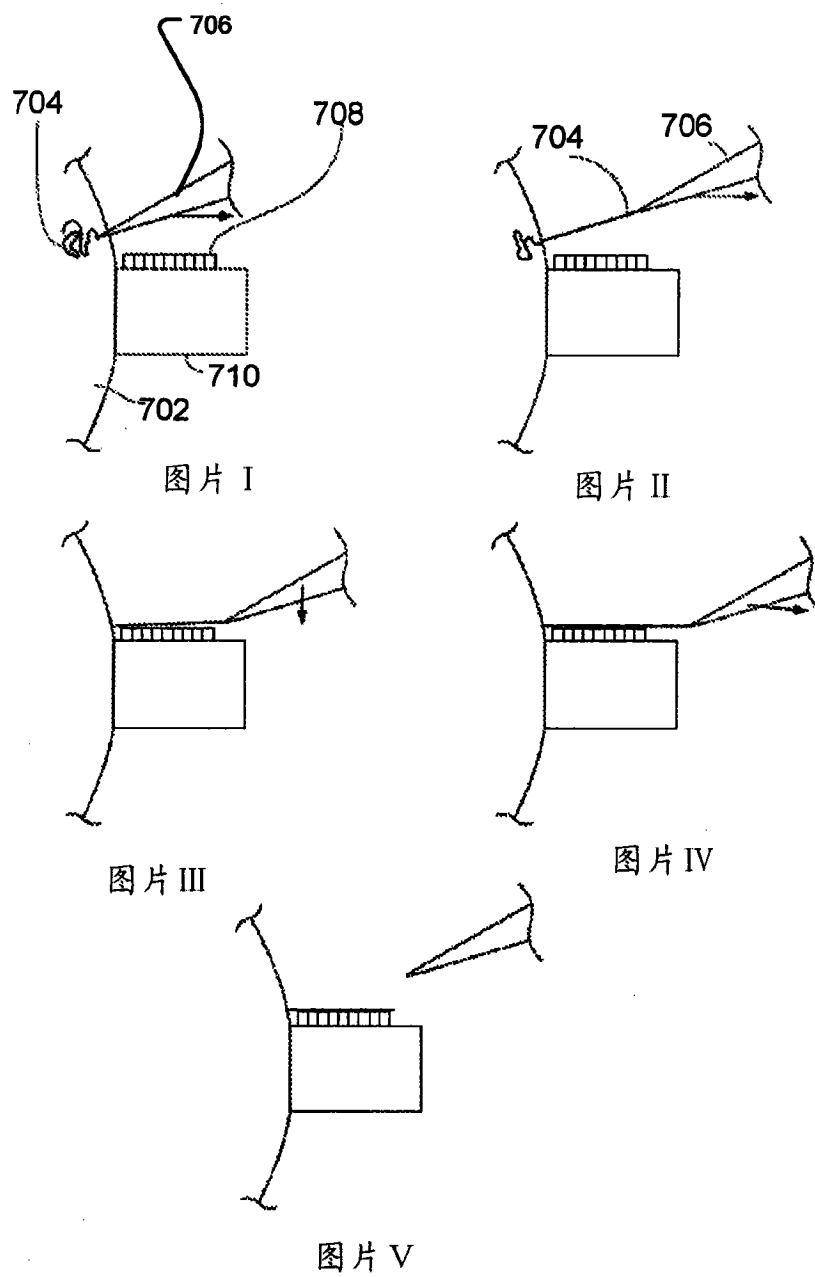


图 7

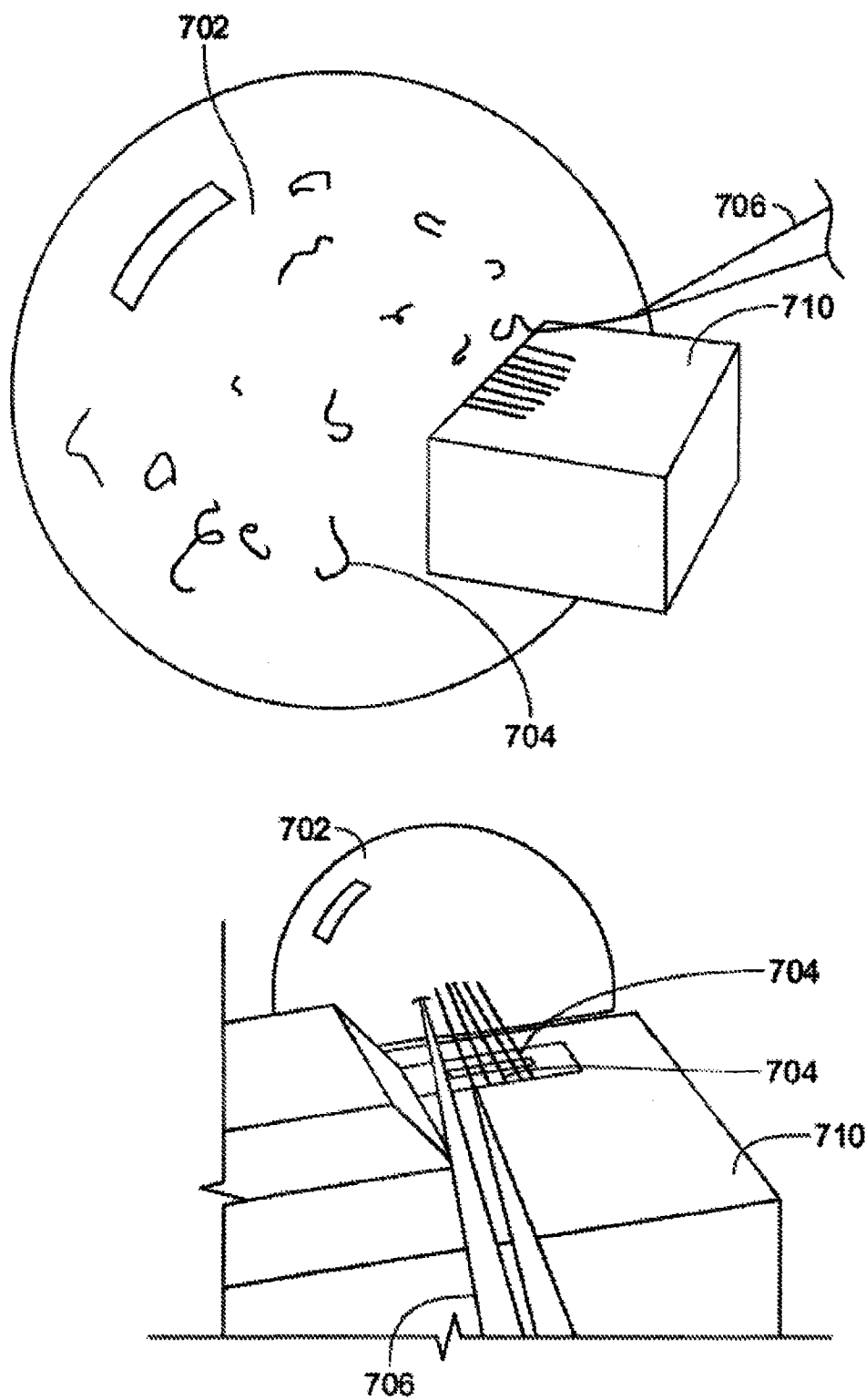


图 8

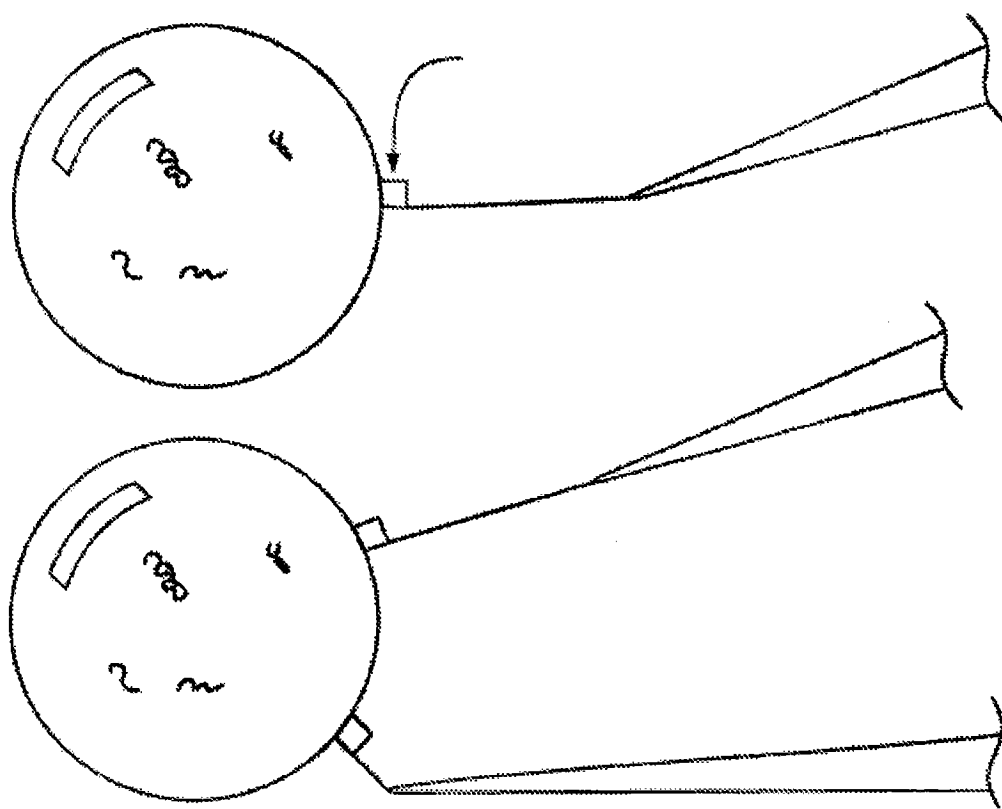


图 9

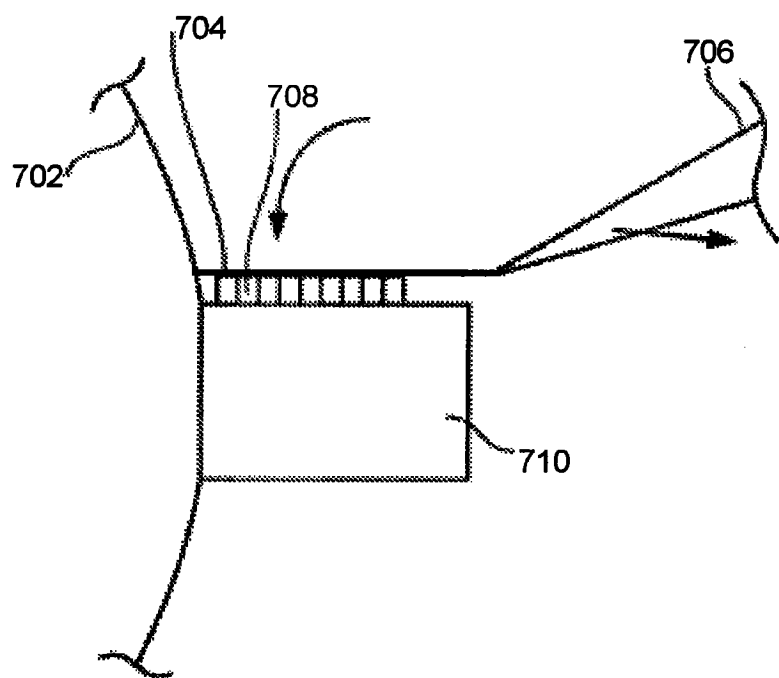


图 10

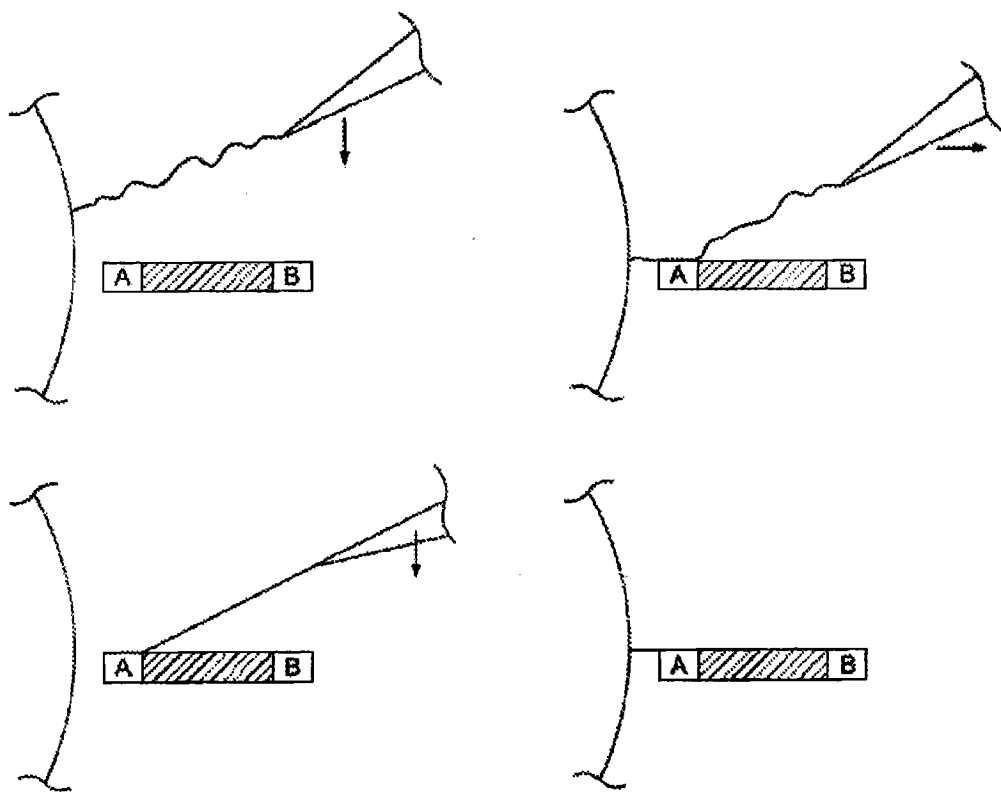


图 11

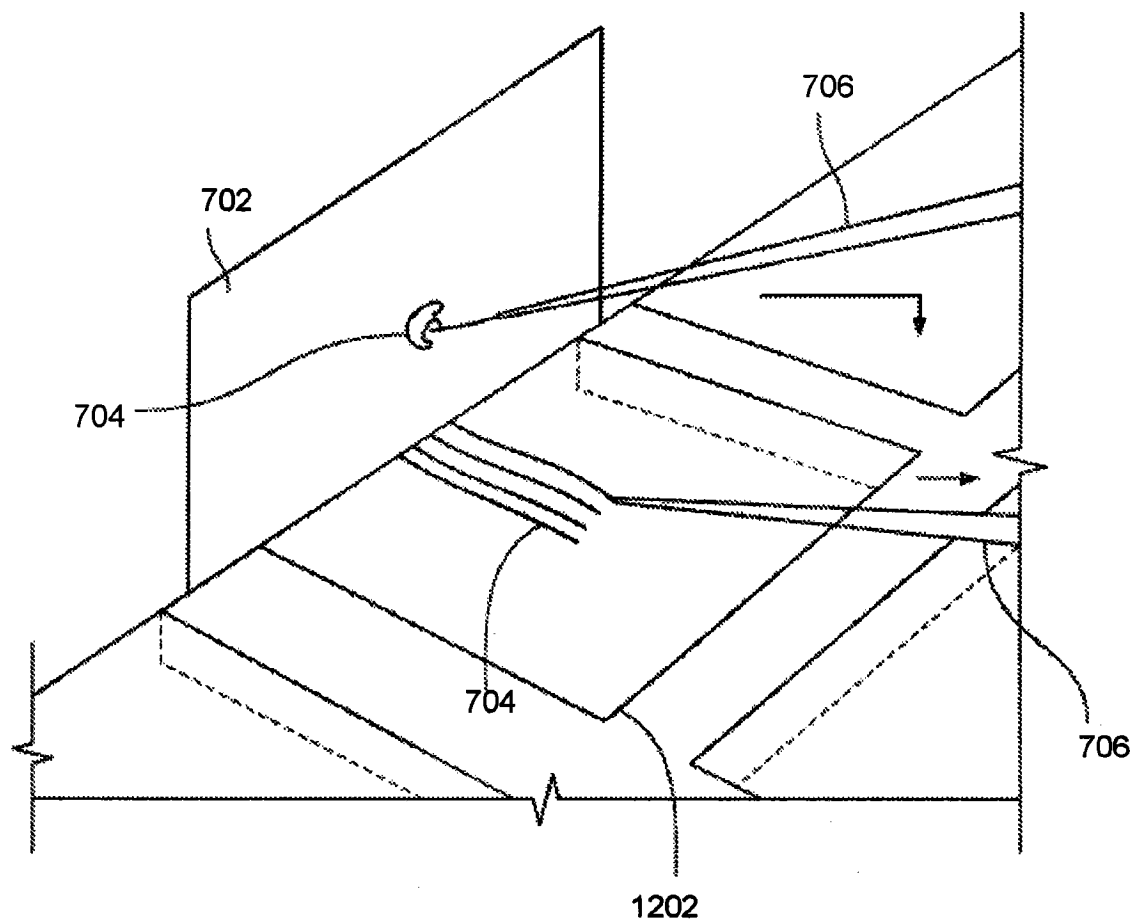


图 12

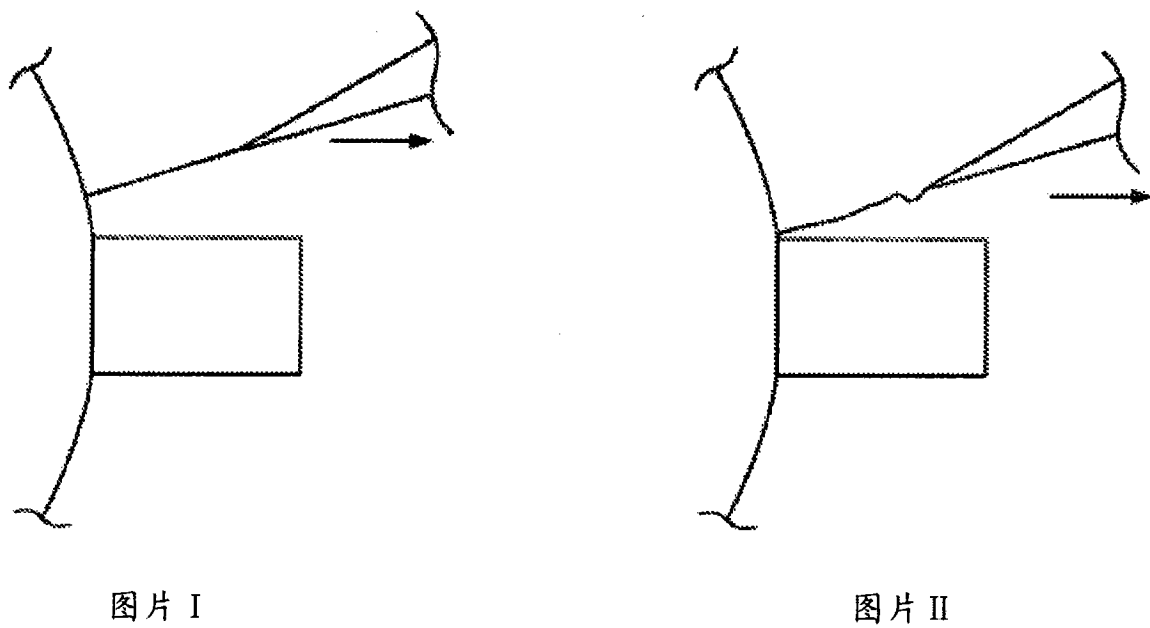
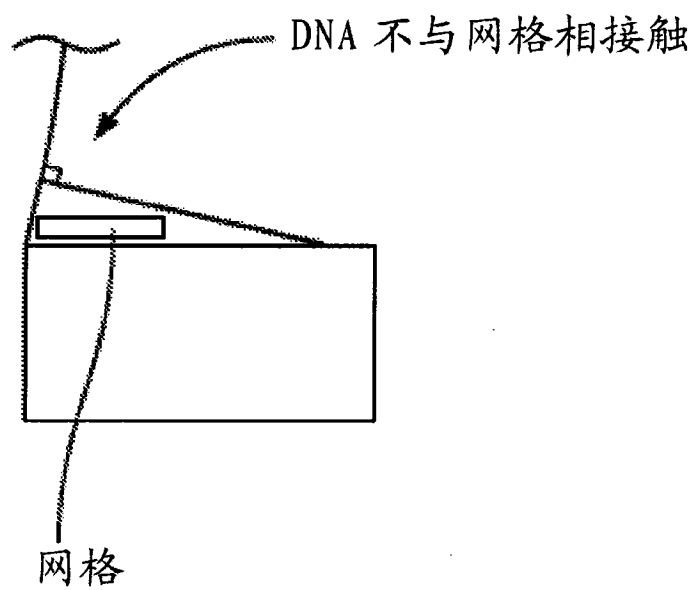


图 13



网格

图 14

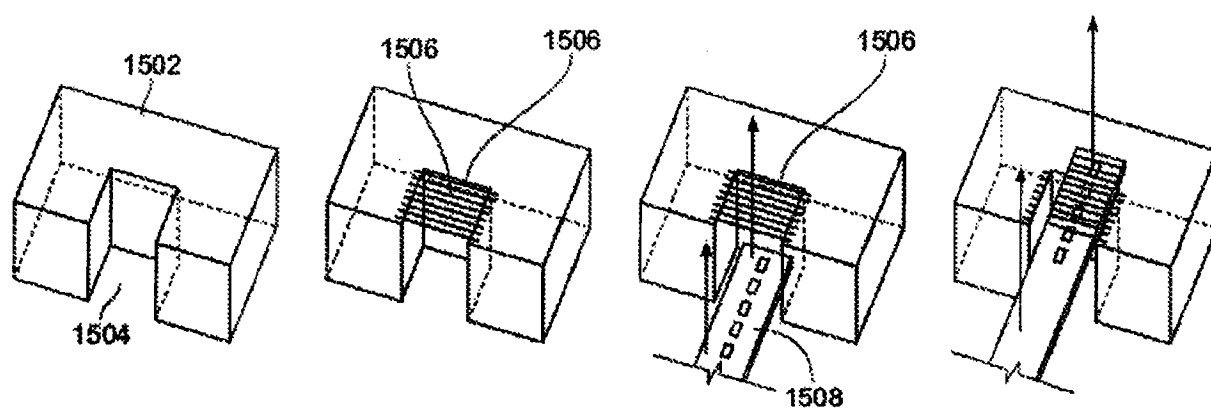


图 15

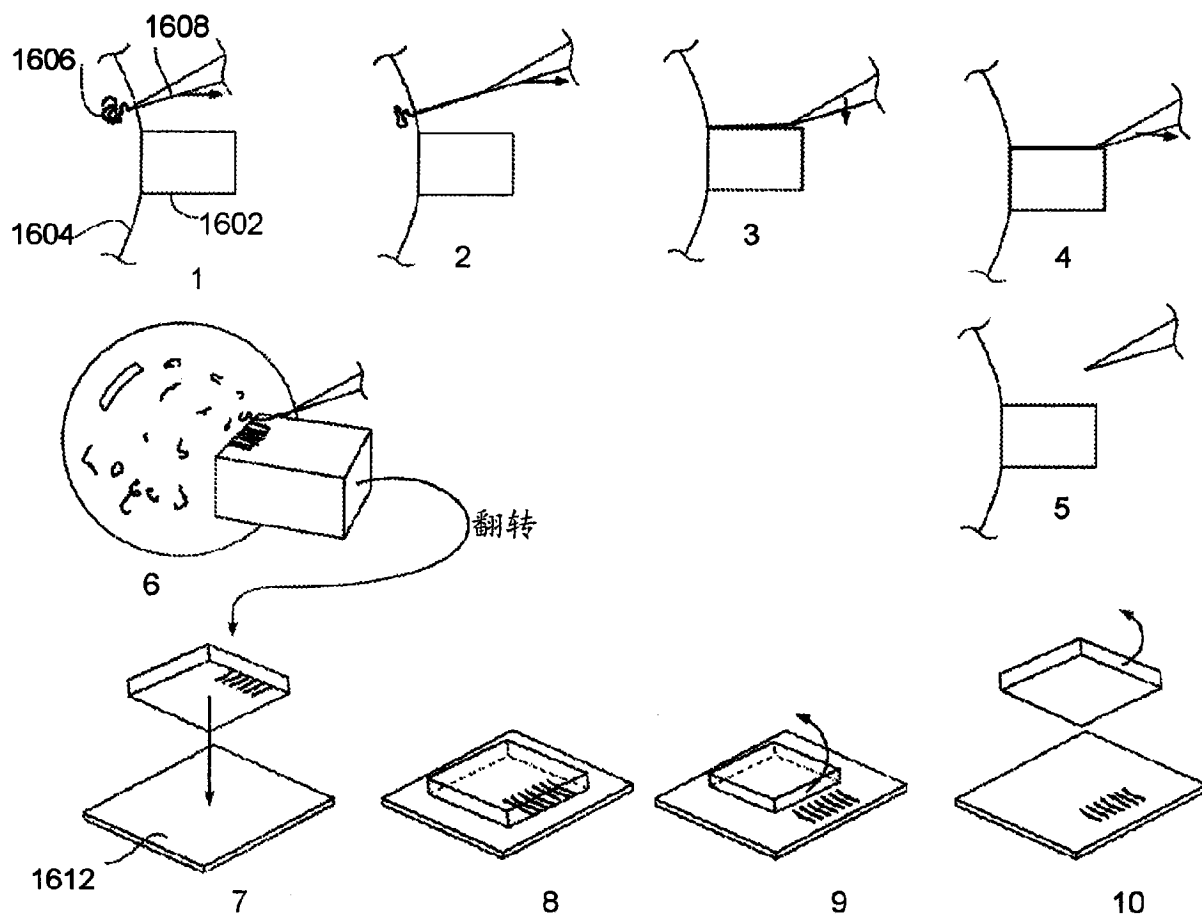


图 16

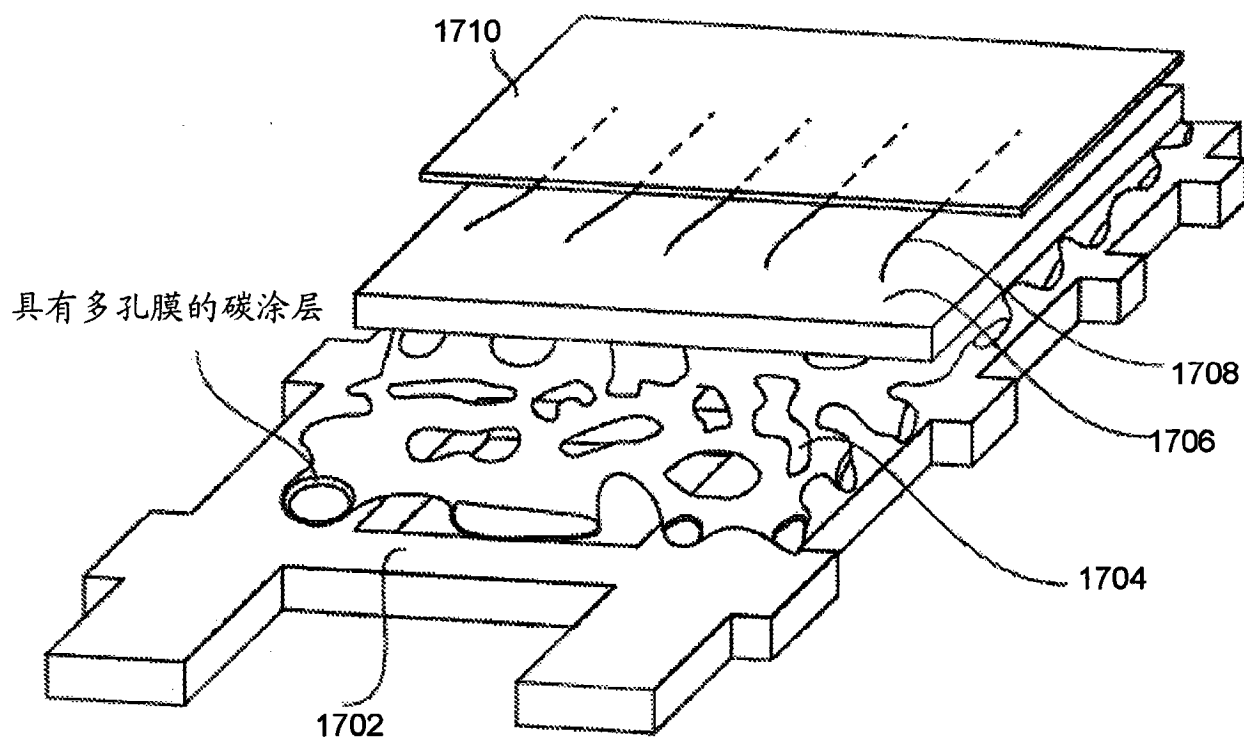


图 17

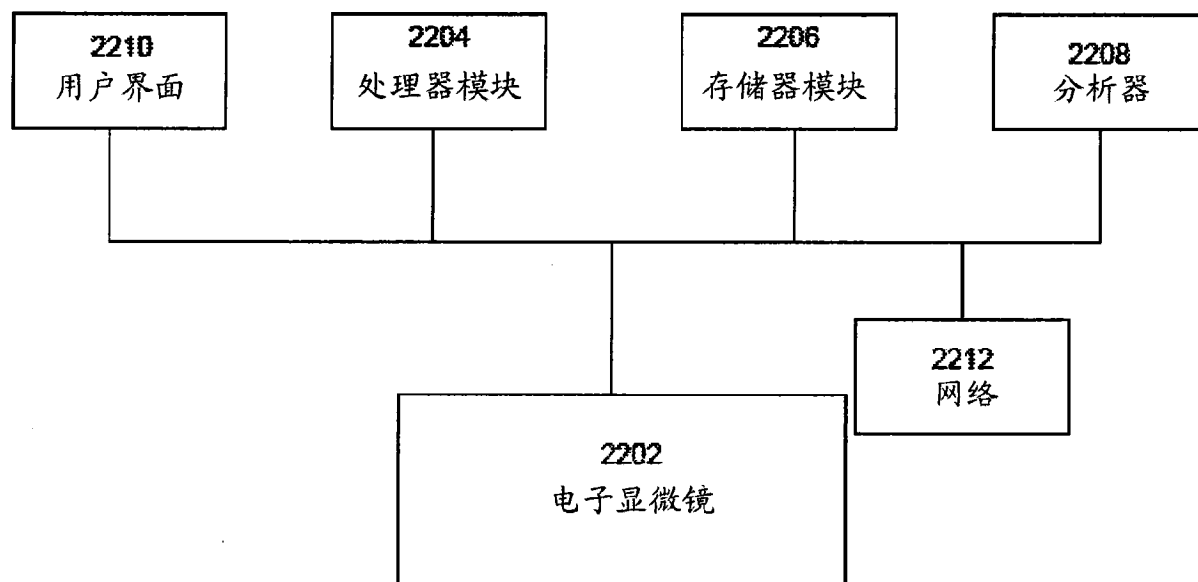


图 18

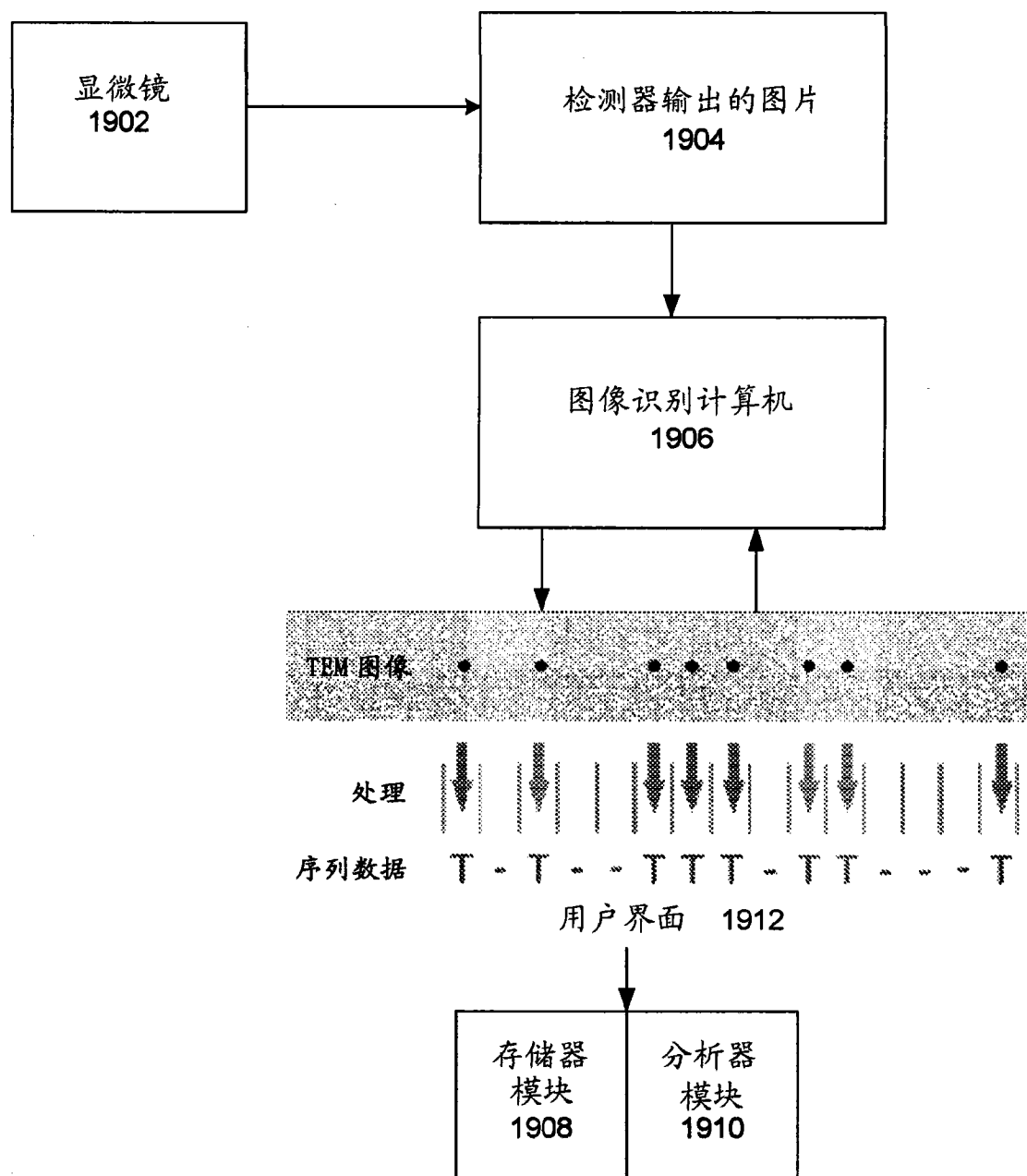


图 19

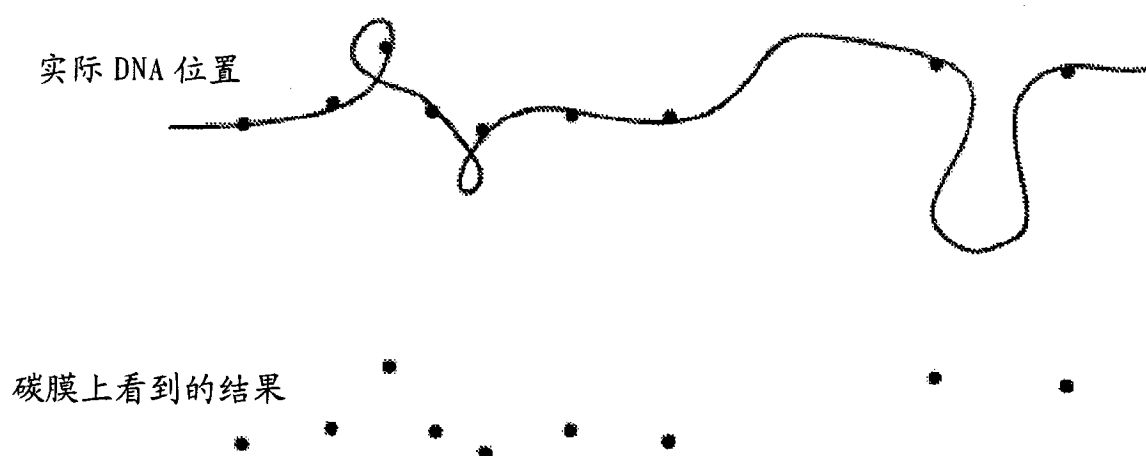
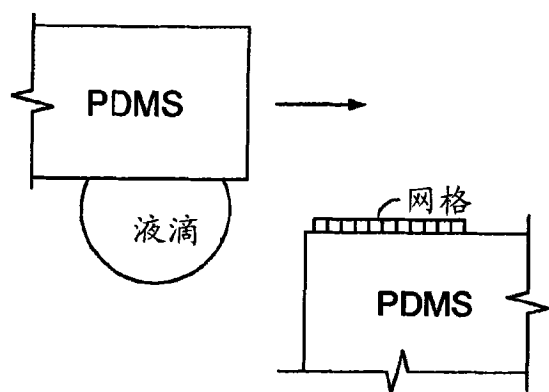
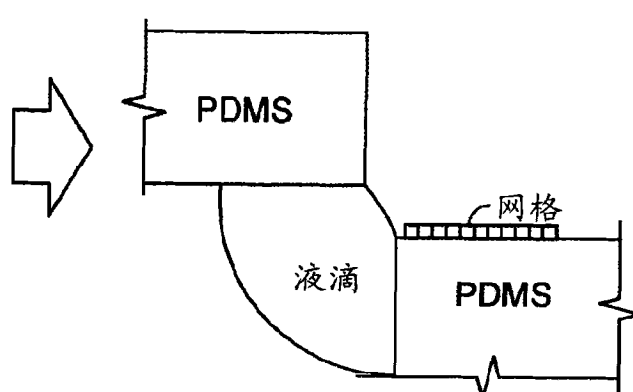


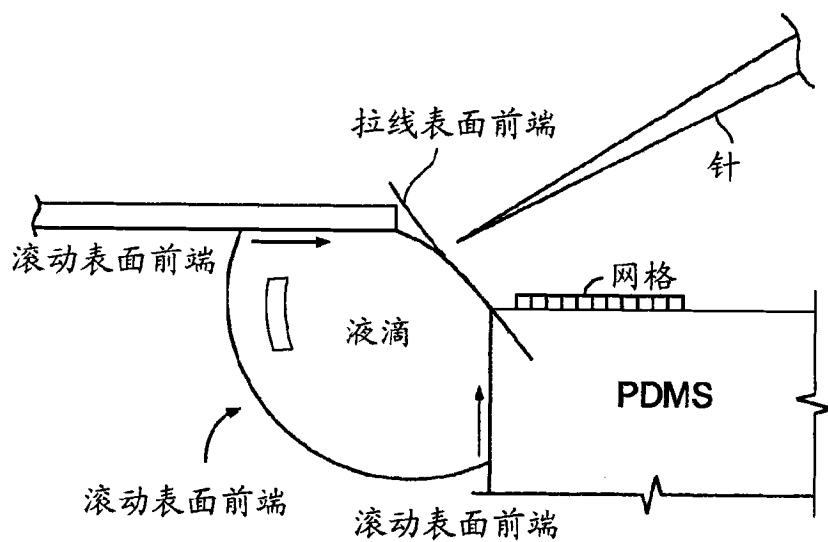
图 20



图片 1



图片 2



图片 3

图 21