[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
C12Q 1/68
C07H 21/04

[12] 发明专利申请公开说明书

[21] 申请号 03820373.1

[43] 公开日 2005 年 10 月 26 日

[11] 公开号 CN 1688716A

[22] 申请日 2003.6.27 [21] 申请号 03820373.1

[30] 优先权

[32] 2002. 6.28 [33] US [31] 60/392,629

[86] 国际申请 PCT/US2003/020504 2003.6.27

[87] 国际公布 WO2004/003233 英 2004.1.8

[85] 进入国家阶段日期 2005.2.28

[71] 申请人 罗斯塔英法美蒂克斯有限责任公司

地址 美国华盛顿州

[72] 发明人 M·马顿 M·迈尔 A·琼斯

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 郭广迅 刘 玥

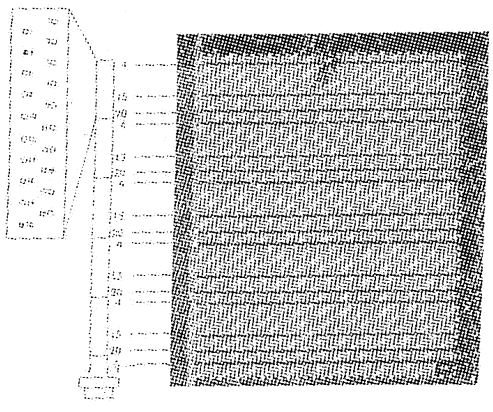
权利要求书 5 页 说明书 43 页 序列表 1 页

附图 20 页

[54] 发明名称 评估微阵列质量的方法

[57] 摘要

本发明涉及评估微阵列质量的方法和组合物。具体地，本发明涉及在逐步合成中通过单体在微阵列上合成的质量控制探针的用途。 通过评估来自质量控制探针的信号程度和确定它们与预期的信号强度的偏离，可以确定微阵列合成的质量。 本发明还涉及检测微阵列的保存和处理过程中发生的缺陷的方法。 本发明还涉及使用计算机鉴定具有合成、保存或处理过程中的一个或多个缺陷的微阵列。



ISSN1008-4274

知识产权出版社出版

1. 位置可寻址阵列，其含有粘附许多不同生物聚合物探针的基质，所述许多探针中所述不同生物聚合物探针位于所述表面上的不同位置并且是所述基质上所述生物聚合物探针的逐步合成产物，所述许多不同结合探针含有许多质量控制探针，所述许多质量控制探针中的每个质量控制探针含有(i)相同的预定结合序列或(ii)具有相同结合特异性的不同的预定结合序列，每个所述质量控制探针中所述预定结合序列的合成在序贯合成周期的所述逐步合成过程中被启动。

2. 权利要求 1 的阵列，其中所述许多质量控制探针中的每个所述质量控制探针的序列由所述预定结合序列组成。

3. 权利要求 1 的阵列，其中所述许多质量控制探针含有由与所述预定结合序列邻接的化学结构组成的第二序列，其中所述质量控制探针的至少一些与所述质量控制探针的其他探针在所述化学结构的长度上不同。

4. 权利要求 3 的阵列，其中所述化学结构是与所述预定结合序列邻接的 0 到 N 个单体的序列，并且其中 N 是等于或大于 1 的整数。

5. 确定位置可寻址生物聚合物阵列是否具有合成缺陷的方法，该方法按陈述顺序包括下面的步骤:

a) 将权利要求 1-2 中任何一项的阵列与含有结合所述预定结合序列的结合伴侣的样品接触;

b) 检测或测量两种或多种所述质量控制探针与样品中所述结合伴侣间的结合; 和

c) 比较两种或多种所述所述质量控制探针的结合，其中如果所述结合是类似的，则表明所述阵列的所述序贯合成周期之间没有合成缺陷。

6. 确定位置可寻址的生物聚合物阵列是否具有合成缺陷的方法，该方法按陈述顺序包括下面的步骤:

a) 将权利要求 3 的阵列与含有结合所述预定结合序列的结合伴侣的样品接触;

b) 检测或测量(i)在所述单体数目上不同的两种或多种所述质量控制探针; 和(ii)样品中所述结合伴侣间的结合; 和

c) 比较两种或多种所述质量控制探针的结合，其中如果所述结

合是类似的，则表明用于合成两种或多种所述质量探针的所述序贯合成周期之间没有合成缺陷。

7. 权利要求 5 的方法，其中所述比较包括确定两种或多种所述质量控制探针中的两种的结合比率，其中所述结合比率是所述两种质量控制探针的第一种与所述结合伴侣结合的量与所述两种质量控制探针的第二种与所述结合伴侣结合的量的商，并且其中所述结合比率在 0.5 到 2.0 之间表明不存在所述合成缺陷。

8. 权利要求 6 的方法，其中所述比较包括确定两种或多种所述质量控制探针中的两种的结合比率，其中所述结合比率是所述两种质量控制探针的第一种与所述结合伴侣结合的量与所述两种质量控制探针的第二种与所述结合伴侣结合的量的商，并且其中所述结合比率在 0.5 到 2.0 之间表明不存在所述合成缺陷。

9. 权利要求 5 或 6 的方法，其还包括在步骤(a)之前的合成所述阵列的步骤。

10. 权利要求 5 或 6 的方法，其中所述样品含有(i)来自一种或多种细胞的总细胞 RNA 或 mRNA 或由其来源的多种核酸，和(ii)所述结合伴侣，其中所述结合伴侣不被所述细胞表达。

11. 权利要求 2、3 或 4 的阵列，其中所述生物聚合物探针是核酸。

12. 权利要求 11 的阵列，其中所述预定结合序列长为 10-40 个核苷酸。

13. 权利要求 11 的阵列，其中所述生物聚合物探针由 20-100 个核苷酸的序列组成。

14. 权利要求 12 的阵列，其中所述预定结合序列长为 25 个核苷酸。

15. 权利要求 14 的阵列，其中所述预定结合序列是 SEQ ID NO:1 或其互补序列。

16. 权利要求 2、3 或 4 的阵列，其中所述生物聚合物探针为蛋白质。

17. 权利要求 16 的阵列，其中所述蛋白质是抗体。

18. 权利要求 2 的阵列，其中所述质量控制生物聚合物探针的所述预定结合序列的长度是阵列上不是所述质量控制探针的生物聚合物

探针长度的 10-75%。

19. 权利要求 18 的阵列，其中所述预定结合序列由 25 个单体组成，并且其中阵列上不是所述质量控制探针的所述生物聚合物探针由 60 个单体组成。

20. 权利要求 4 的阵列，其中 N 不大于阵列上不是所述质量控制生物聚合物探针的所述生物聚合物探针中单体的数目与所述预定结合序列中单体数目的差。

21. 权利要求 4 的阵列，其中 N 大于阵列上不是所述质量控制生物聚合物探针的所述生物聚合物探针中单体的数目与所述预定结合序列中单体数目的差。

22. 权利要求 4 的阵列，其包含 N 不同的三种所述质量控制探针。

23. 权利要求 22 的阵列，其中不同质量控制探针的 N 分别为 0、20 和 35。

24. 制备许多不同生物聚合物探针的位置可寻址阵列的方法，该方法包括使用逐步合成方法在基质上从单体合成所述许多不同生物聚合物探针，从而所述不同生物聚合物探针的每一个都粘附到所述基质上不同位置，其中所述许多不同生物聚合物探针含有许多质量控制探针，所述许多质量控制探针中的每一个质量控制探针含有相同预定结合序列，其中所述质量控制探针每一个中所述预定结合序列的合成在所述序贯合成周期的逐步合成中启动。

25. 权利要求 24 的方法，其中所述许多质量控制探针的每一个的序列由所述预定结合序列组成。

26. 权利要求 24 的方法，其中所述许多质量控制探针含有与所述预定结合序列邻接的 0 到 N 个单体的第二序列，其中所述质量控制探针的至少一些与所述质量控制探针的其他探针在所述单体的数目上不同，并且其中 N 是等于或大于 1 的整数。

27. 权利要求 1 的阵列，其中所述许多质量控制探针含有

（i） 序列由所述预定序列组成的质量控制探针；和

（ii） 含有与所述预定结合序列邻接的 0 到 N 个单体的第二序列的质量控制探针，其中至少一些所述质量控制探针与其他所述质量控制探针在所述单体数目上不同，并且其中 N 是等于或大于 1 的整数。

28. 权利要求 27 的阵列，其中所述生物聚合物探针是寡核苷酸，

所述预定序列由 25 个核苷酸组成，并且不是所述质量控制探针的所述生物聚合物探针由 60 个核苷酸组成。

29. 含有 SEQ ID NO:1 或 SEQ ID NO:2 的核苷酸序列或其互补序列的寡核苷酸。

30. 位置可寻址阵列，其含有粘附许多不同生物聚合物探针的基质，所述许多探针中所述不同生物聚合物探针位于所述表面上的不同位置并且是在所述基质上向所述生物聚合物探针逐步添加单体的产物，所述许多不同生物聚合物探针含有许多质量控制探针，所述许多质量控制探针中的每个质量控制探针含有至少一个被标记的单体，所述被标记单体向所述质量控制探针的加入在序贯合成周期的逐步合成中被启动。

31. 确定权利要求 30 的位置可寻址生物聚合物阵列是否具有合成缺陷的方法，该方法包括比较两种或多种所述质量控制探针的所述至少一个被标记单体的信号，其中如果所述信号是类似的，则表明所述阵列的所述序贯合成周期间不存在合成缺陷。

32. 权利要求 31 的方法，其中所述比较包括确定两种或多种所述质量控制探针中的两种的信号比率，其中所述信号比率是所述两种质量控制探针的第一种发出的信号量与所述两种质量控制探针的第二种发出的信号量的商，并且其中所述信号比率在 0.5 到 2.0 之间表明不存在所述合成缺陷。

33. 权利要求 30 的阵列，其中所述生物聚合物探针是核酸。

34. 权利要求 33 的阵列，其中所述生物聚合物探针由 20-100 个核苷酸序列组成。

35. 权利要求 30 的阵列，其中所述生物聚合物探针是蛋白质。

36. 权利要求 35 的阵列，其中所述蛋白质是抗体。

37. 权利要求 5、6、31 中任一项的方法，其中所述合成缺陷是喷嘴故障。

38. 权利要求 37 的方法，其中所述阵列含有以周期 P 排列的所述质量控制探针的至少一部分并且其中使用具有 P 个喷嘴的喷墨印刷头通过逐步合成合成所述阵列，其中 P 是等于或大于 1 的整数。

39. 权利要求 38 的方法，其中 P 等于 20。

40. 使用权利要求 1 或 2 的位置可寻址阵列检测喷嘴故障的方

法，该方法按陈述顺序包括下面的步骤：

a） 将权利要求 1 或 2 的阵列与含有结合所述预定结合序列的结合伴侣的样品接触，其中所述许多质量控制探针的至少一部分以周期 P 排列并且其中使用具有 P 个喷嘴的喷墨印刷头通过逐步合成合成所述阵列，其中 P 是等于或大于 1 的整数；

b） 检测或测量两种或多种所述质量控制探针与样品中所述结合伴侣间的结合；和

c） 比较周期为 P 的两种或多种所述质量控制探针的结合，其中如果所述结合是类似的，则表明没有喷嘴缺陷。

41. 使用权利要求 30 的位置可寻址阵列检测喷嘴故障的方法，该方法包括比较周期为 P 的两种或多种所述质量控制探针的所述至少一个被标记单体的信号，其中所述许多质量控制探针的至少一部分以周期 P 排列并且其中使用具有 P 个喷嘴的喷墨印刷头通过逐步合成合成所述阵列，其中如果所述信号是类似的，则表明没有喷嘴缺陷，其中 P 是等于或大于 1 的整数。

评估微阵列质量的方法

本发明要求 2002 年 6 月 28 日提交的美国临时专利申请序列号 60/392,629 的优先权，该临时专利申请在这里被完整并入作为参考。

1. 发明领域

本发明涉及评估微阵列合成的质量的方法和组合物。本发明还涉及检测微阵列的保存或处理过程中发生的缺陷的方法。具体地，本发明涉及在微阵列上合成的质量控制探针用于评估微阵列质量的用途。本发明还涉及使用计算机鉴定具有例如在合成、保存或处理过程中发生的一个或多个缺陷的微阵列。

2.发明背景

DNA 阵列技术已经使得可能一次监控许多基因转录产物的表达水平(见，例如，Schena 等，1995，Science 270: 467-470; Lockhart 等，1996，Nature BioTechnology 14:1675-1680；Blanchard 等，1996，Nature BioTechnology 14: 1649;Shoemaker 等，美国专利申请序列号 09/724,538，2000 年 11 月 28 日提交)。DNA 阵列还应用于基因发现中，例如用于鉴定基因的外显子结构(见，例如，Shoemaker 等，美国专利申请序列号 09/724,538，2000 年 11 月 28 日提交;Meltzer,2001, Curr. Opin. Genet. Dev. 11 (3):258-63; Andrews 等，2000，Genome Res. 10(12): 2030-43; Abdellatif, 2000, Circ. Res. 86(9): 919-20; Lennon, 2000, Drug Discov. Today 5(2): 59-66; Zweiger, 1999, Trends Biotechnol. 17(11): 429-36)。

通过同时监控数万基因，微阵列技术已经允许对细胞或细胞型或任何生物样品中 mRNA 的表达进行基因组范围的研究。通过复杂的数据管理和分析方法的帮助，可以在 mRNA 水平上表征细胞或细胞型的转录状态以及响应外部干扰(包括但不限于药物干扰)转录状态的改变(见，例如，美国专利号 6,203,987; Stoughton 等，国际公开号 WO 00/24936(2000 年 5 月 4 日公布)；Stoughton 等，国际公开号 WO 00/39336(2000 年 7 月 6 日公布)；Friend 等，国际公开号 WO

00/24936(2000 年 5 月 4 日公布))。这些技术的应用包括，例如，在各种生理状态，尤其是疾病状态中被上调或下调的基因的鉴定。DNA 阵列的额外的示例性用途包括分析信号途径成员，和鉴定各种药物的靶标，见例如，Friend 和 Hartwell，国际公开号 WO/9838329(1998 年 9 月 3 日公布)；Friend 和 Stoughton，国际公布号 WO 99/59037(1999 年 11 月 18 日公布)；美国专利号 6,132,969、5,965,352、6,218,122。

微阵列是载体上位置可寻址结合(例如通过杂交)位点的阵列。这些结合位点的每一个含有结合到载体上预定区域的探针的许多生物聚合物分子。可以用多种方法制造微阵列，包括将预合成的探针固定在载体上或者在载体上原位合成探针。例如，预合成的探针的固定可以如 DeRisi 等(1997, Science 278(5338) :680-6)中描述的通过机器人进行或者通过喷墨(inkjet)进行。原位合成可通过不同方法完成，包括使用喷墨技术或者通过光活化的合成(Holmes 等，1995, Biopolymers 37 (3): 199-211; Jacobs 等，1994, Trends Biotechnol. 12(1) : 19-26; Fodor 等，1991, Science 251 (4995): 767-73)。在任一种原位合成中，都在载体上发生化学反应，其中一种或多种单体被加到生物聚合物。然而，随着生物聚合物链的增长，可能一个或多个合成周期会失败(完全地或部分地)，从而产生缺少一种或多种所期望单体的探针。合成效率依赖于多种因素，包括试剂纯度、反应时间、喷墨头的正确调准。这些过程任一种中的缺陷都可导致一种或多种单体向正增长的生物聚合物链的无效加入。

此外，对于喷墨-合成的微阵列，当喷墨头的喷嘴之一不能正确递送试剂(例如，如果喷嘴被暂时或者永久堵塞)时也可以发生合成缺陷。喷嘴故障指单个喷墨喷嘴的故障。如果喷嘴不能递送生物聚合物加入所需的期望的溶液，其有时被称为被“阻塞”。喷嘴故障可以在微阵列合成的任何时间点发生。合成开始时的故障可以是由于引导新试剂穿过喷嘴的不足。如果，例如喷嘴中有被捕获的气泡或者微粒，那么喷嘴故障也可以在一组微阵列的印刷开始之后发生。可以在微阵列合成之前检测并修正喷嘴故障。每批合成开始之前和每批合成结束时可以试验印刷头上的每个喷嘴以确定其正确地运行。这可以通过在迫使每个喷嘴挤出少量液体之前将干净基质置于喷头组合上实现。如果所有喷嘴都正确工作，将有相应于每个喷嘴的一滴液体。然而，如

果一个或多个喷头故障，将缺少相应于那个喷嘴位置的液滴。因为该液滴的小尺寸，喷嘴故障偶尔可由于人为误差而被忽视，从而将合成显示出喷嘴故障证据的阵列。现在需要更可靠方法以确定是否合成故障已经发生并且，如果发生了，在微阵列合成过程中何处何时发生这些故障。虽然通过常规 DNA 测序、质谱或通过其他方法可能在预合成的探针上进行质量控制，但是缺少对原位合成的探针质量的评估方法。

本申请描述了为评价微阵列合成的质量而设计的方法。此处公开的方面描述了设计和生产微阵列上质量控制探针的方法和分析从微阵列处理所得信息的方法，该方法允许确定合成的总体质量以及鉴定最可能已经有缺陷的合成周期。本发明还包括含有关于微阵列上质量控制探针的位置和身份的信息的数据库。

此处的引用或参考文献的讨论不应被理解为承认这些引用或参考文献的讨论是本发明的已有文献。

3. 发明概述

本发明涉及评估微阵列质量的方法和组合物，在微阵列中，生物聚合物探针在逐步合成中通过单体在阵列基质单体上合成。具体地，通过将质量控制探针包含在微阵列上来检测微阵列的单个合成周期的沉积中的故障或无效。质量控制探针与其他生物聚合物探针同时被合成到微阵列上，从而将可能遭受可能发生的合成故障或无效。通过评估来自质量控制探针的信号的程度和确定它们与所预期信号强度的偏差，可以确定微阵列合成的质量。

在一个实施方案中，每组质量控制探针含有相同的预定结合序列，其结合伴侣存在于样品中或者被导入样品中以接触用于分析的微阵列。每个质量控制探针中预定结合序列的合成在序贯合成周期的逐步合成过程中被启动。通过评估能够结合质量控制探针的预定结合序列的生物聚合物的结合程度，可以确定微阵列合成的质量。在另一个实施方案中，质量控制探针不含有预定结合序列。通过质量控制探针自身而不是结合到预定结合序列的被标记的结合伴侣产生可检测信号。这可以通过例如，将一种或多种标记单体掺入到质量控制探针，用荧光染料对质量控制探针染色等完成。

在优选的实施方案中，本发明涉及检测寡核苷酸微阵列上合成故障的方法。在更优选的实施方案中，本发明涉及检测喷墨寡核苷酸微阵列合成过程中合成故障(包括喷嘴故障)的方法。除了合成故障，也可以检测影响微阵列质量的其他缺陷，例如微阵列的保存或处理过程中由于探针的降解造成的缺陷。

本发明提供了含有粘附许多不同生物聚合物探针的基质的位置可寻址阵列，所述许多探针中所述不同生物聚合物探针位于所述表面的不同位置并且是所述基质上所述生物聚合物探针的逐步合成的产物，所述许多不同结合探针含有许多质量控制探针，所述质量控制探针的合成已经在所述序贯合成周期的逐步合成过程中被启动。含有预定结合序列的所述许多探针中每种质量对照探针优选含有相同的预定结合序列或者备选地不同的预定结合序列但是具有相同结合特异性或者类似结合特征(例如，在相同结合条件下结合到它们的各自具有类似强度的结合伴侣)。在一个实施方案中，可以使用不同长度的预定结合序列(例如，25 聚物和 24 聚物)。

在阵列的一个特定实施方案中，所述许多探针的每种所述质量控制探针的序列由所述预定结合序列组成。

在另一特定实施方案中，许多质量控制探针含有由与所述预定结合序列邻接的化学结构组成的第二序列，其中至少一些质量控制探针与其他质量控制探针在所述化学结构的长度上不同。在特定实施方案中，该化学结构是与所述预定结合序列邻接的 0 到 N 个单体的序列，并且其中 N 是等于或大于 1 的整数。在特定实施方案中，生物聚合物探针是寡核苷酸，所述预定序列由 25 个核苷酸组成，不是所述质量控制探针的所述生物聚合物探针由 60 个核苷酸组成。在特定实施方案中，N 不大于阵列上非所述质量控制生物聚合物探针的所述生物聚合物探针中的单体数目减去所述预定结合序列中单体数目。在另一特定实施方案中，质量控制探针含有比阵列上非所述质量控制生物聚合物探针的生物聚合物探针更多的单体。在另一特定实施方案中，阵列含有 3、10、30、60 或更多所述质量控制探针，这些探针的 N 不同。一个特定实施方案是对于不同质量控制探针，其中 N 分别为 0、20 和 35。

在另一特定实施方案中，许多质量控制探针含有

(i) 序列由所述预定序列组成的质量控制探针；和

(ii) 含有与所述预定结合序列邻接的 0 到 N 个单体的第二序列的质量控制探针，其中至少一些所述质量控制探针与其他所述质量控制探针在所述单体数目上不同，并且其中 N 是等于或大于 1 的整数。

在各种特定实施方案中，生物聚合物探针是核酸、蛋白质，或抗体。优选地，预定结合序列是长度为 10-40 个核苷酸，更优选地，长为 25 个核苷酸。在特定实施方案中，预定结合序列是 SEQ ID NO：1 或 SEQ ID NO: 2 或它们的互补序列。

在一个实施方案中，生物聚合物探针由 20 到 100 个核苷酸序列组成。

优选地，质量控制生物聚合物探针的预定结合序列的长度是阵列上非质量控制探针的生物聚合物探针长度的 10-75%。

在特定实施方案中，预定结合序列由 25 个单体组成，并且阵列上非质量控制探针的生物聚合物探针由 60 个单体组成。

本发明还提供了确定位置可寻址生物聚合物阵列是否有合成缺陷的方法，该方法按陈述顺序包括下列步骤：

a) 将本发明的阵列与含有结合所述预定结合序列的结合伴侣的样品接触；

b) 检测或测量两种或多种所述质量控制探针与样品中所述结合伴侣间的结合；和

c) 比较两种或多种所述所述质量控制探针的结合，其中如果所述结合是类似的，则表明所述阵列合成的所述序贯合成周期之间没有合成缺陷。

本发明还提供了确定位置可寻址的生物聚合物阵列是否具有合成缺陷的方法，其按陈述顺序包括下面的步骤：

a) 将含有包含 0 到 N 个单体邻接序列的质量控制探针的如上描述的阵列与含有结合所述预定结合序列的结合伴侣的样品接触；

b) 检测或测量(i)在所述单体数目上不同的两种或多种所述质量控制探针；和(ii)样品中所述结合伴侣间的结合；和

c) 比较两种或多种所述质量控制探针的结合，其中如果所述结

合是类似的，则表明用于合成两种或多种所述质量探针的所述序贯合成周期之间没有合成缺陷。

本发明还提供了确定是否位置可寻址的生物聚合物阵列具有喷嘴故障导致的合成缺陷的方法，其按陈述顺序包括如下步骤：

a) 将本发明的阵列与含有结合所述预定结合序列的结合伴侣的样品接触，其中所述许多质量控制探针的至少一部分以周期性 P 排列并且其中所述阵列使用具有 P 个喷嘴的喷墨印刷头通过逐步合成方法合成，其中 P 是等于或大于 1 的整数；

b) 检测或测量两种或多种所述质量控制探针与样品中所述结合伴侣间的结合；和

c) 比较周期性为 P 的两种或多种所述质量控制探针的结合，其中如果所述结合是类似的，则表明不存在喷嘴故障。

在前面的方法中，比较步骤可以包括确定两种或多种所述质量控制探针的结合比率，其中所述结合比率是所述两种质量控制探针的第一种与所述结合伴侣结合的量与所述两种质量控制探针的第二种与所述结合伴侣结合的量的商，并且其中所述结合比率在 0.5 到 2.0 之间表明不存在所述合成缺陷。

在特定实施方案中，前面的方法还包括在步骤(a)之前合成所述阵列的步骤。

在特定实施方案中，样品含有(i)来自一种或多种细胞的总细胞 RNA 或 mRNA 或从它们得到的多种核酸，和(ii)所述结合伴侣，其中所述结合伴侣不被所述细胞表达。

本发明还提供了制备许多不同生物聚合物探针的位置可寻址阵列的方法，该方法包括使用逐步合成方法在基质上从单体合成所述许多不同生物聚合物探针，从而所述生物聚合物探针的每一个都粘附到所述基质上不同位置，其中所述许多不同生物聚合物探针含有许多质量控制探针，所述许多质量控制探针中的每一个质量控制探针含有相同预定结合序列，其中所述质量控制探针每一个中所述预定结合序列的合成在所述序贯合成周期的逐步合成中启动。从而所制备的阵列可以具有上述特征。

本发明还提供了含有 SEQ ID NO:1 或 SEQ ID NO:2 的核苷酸序列或其互补序列的寡核苷酸。

4. 附图描述

图 1 阐明了用具有三个不正常工作的喷嘴合成的喷墨寡核苷酸微阵列。相应于喷嘴 4、15 和 20 的整行由于喷嘴故障而没有合成。

图 2A-2B 示意性阐明了使用具有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A)25 个核苷酸长的探针或者直接合成到微阵列上或者附着到不定长度(即 20 个核苷酸或 35 个核苷酸)的间隔区。(B)描绘了合成周期 24 中的合成错误并从而影响仅仅头两种所示质量控制探针中预定结合序列中单体的序列。实线描述了质量控制探针，虚线描述了间隔区。

图 3A-3B 示意性阐明了使用交错开始质量控制探针确定寡核苷酸微阵列的合成质量。(A)从合成周期 1 开始直到合成周期 36 在微阵列上直接合成一组 25 个核苷酸质量控制探针。这些质量控制探针的唯一差异是合成开始的合成周期。(B)描述了合成周期 29 中的合成错误并从而仅仅影响了质量控制探针，其中合成周期 29 实际用于将单体加到质量控制探针(即，在合成周期 5-29 开始合成的那些质量控制探针)的序列。粗线描述了质量控制探针，细线描述了没有单体沉积的合成周期。

图 4A-4B 阐明了当寡核苷酸微阵列合成期间没有已知的或可检测到的合成缺陷时，使用含有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A)杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B)(A)中微阵列的更高放大倍数，其描述了 25 聚体、40 聚体和 60 聚体的位置。

图 5A-5B 描述了当寡核苷酸微阵列合成过程中第一个合成周期被故意跳过时，使用含有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A)杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B)(A)中微阵列的更高放大倍数，其描述了 25 聚体、40 聚体和 60 聚体的位置。

图 6A-6B 描述了当寡核苷酸微阵列合成过程中第一和第二个合成周期被故意跳过时，使用含有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A)杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B)(A)中微阵列的更高放大倍数，其描述了

25 聚体、40 聚体和 60 聚体的位置。

图 7A-7B 描述了当寡核苷酸微阵列合成过程中第三十六个合成周期被故意跳过时，使用含有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A) 杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B) (A) 中微阵列的更高放大倍数，其描述了 25 聚体、40 聚体和 60 聚体的位置。

图 8A-8B 描述了当寡核苷酸微阵列合成过程中第三十四和三十五个合成周期被故意跳过时，使用含有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A) 杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B) (A) 中微阵列的更高放大倍数，其描述了 25 聚体、40 聚体和 60 聚体的位置。

图 9A-9B 描述了当寡核苷酸微阵列合成过程中头二十二个合成周期中存在无效合成时，使用含有间隔区的质量控制探针确定寡核苷酸微阵列的合成质量。(A) 杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B) (A) 中微阵列的更高放大倍数，其描述了 25 聚体、40 聚体和 60 聚体的位置。

图 10A-10B 描述了当寡核苷酸微阵列合成过程中第一和第二个合成周期中存在无效合成时，使用交错开始质量控制探针确定寡核苷酸微阵列的合成质量。(A) 杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B) 每个合成周期质量控制探针的平均荧光强度图。

图 11A-11B 描述了当寡核苷酸微阵列合成过程中头五个合成周期中存在无效合成时，使用交错开始质量控制探针确定寡核苷酸微阵列的合成质量。(A) 杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B) 每个合成周期质量控制探针的平均荧光强度图。

图 12A-12B 描述了当寡核苷酸微阵列合成过程中头八个合成周期中存在无效合成时，使用交错开始质量控制探针确定寡核苷酸微阵列的合成质量。(A) 杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B) 每个合成周期质量控制探针的平均荧光强度图。

图 13A-13B 描述了当寡核苷酸微阵列合成过程中第四十五到第六

十个合成周期中存在无效合成时，使用交错开始质量控制探针确定寡核苷酸微阵列的合成质量。(A)杂交到被杂交到质量控制探针的荧光标记的寡核苷酸后的微阵列成像。(B)每个合成周期质量控制探针的平均荧光强度图。

图 14A-14B 阐明了单个缺失质量控制探针的增加的敏感性。用质量控制探针合成了在第三十四和三十五个合成周期中有合成缺陷的微阵列，在预定结合序列中(A)没有或(B)有故意的单个缺失。使用全长 25 个核苷酸预定结合序列的标记的反向互补序列与每个微阵列杂交。为每个微阵列确定每个合成周期时质量控制探针的平均荧光强度图。

图 15A-15C 阐明了在合成中跳过了头二十二个合成周期的微阵列(A)；没有合成缺陷的微阵列(B)；和在合成中跳过了头二十二个合成周期的微阵列与没有合成缺陷的微阵列(C)的荧光反向对之间的相关性。

图 16A-16D 阐明了没有合成缺陷的寡核苷酸微阵列与在合成中第一个(A)、第一个和第二个(B)、第三十六个(C)或第三十四和三十五个(D)合成周期被跳过的寡核苷酸微阵列之间的相关性。

图 17A-17D 示意性阐明了质量控制探针粘附到基质的微阵列。(A)外部的网格子线、(B)对角网格线、(C)内部簇、(D)角落簇。

5．发明详述

本发明的目的是评估阵列的微阵列合成质量，其中生物聚合物探针在阵列基质单体上通过单体在逐步合成中合成。该目的通过在将要评估的微阵列上合成质量控制探针完成。质量控制探针以与微阵列上其他生物聚合物探针相同的方式合成并且与这些生物聚合物探针相邻。

质量控制探针可以含有预定的结合序列。该预定结合序列具有结合伴侣，其可用于检测微阵列处理过程中预定结合序列的存在。在一些实施方案中，质量控制探针还含有与预定结合序列相邻的化学结构(这些化学结构在这里称为“间隔区”)。间隔区优选为附着(邻接)到预定结合序列的额外单体的聚合物(例如，序列)。微阵列合成完成时，通过结合到标记的结合伴侣检测质量控制探针。为每个质量控制探针定量结合程度并与其他质量控制探针的结合强度比较。相似的结合强

度表明在整个合成中合成同等地有效。

在另一特定实施方案中，质量控制探针不含有预定结合序列。在这种实施方案中，用这种质量控制探针观察到的信号或者通过 1)组成质量控制探针的单体直接发出或 2)与组成质量控制探针的单体相互作用或附着到这些单体的标记(例如，染料)发出。与期望的结合强度的背离表明阵列中存在缺陷，该缺陷例如是由于合成缺陷，或者保存或处理中的降解。

尽管本发明就一组质量控制探针的使用作一般性描述，但是应该理解在单个微阵列上也可以使用质量控制探针的不同组。质量控制探针的不同组可能具有不同的预定结合序列或者可以是质量控制探针与和不与预定结合序列的混合物。质量控制探针也可以是不同长度的混合物(例如，含有 25 聚体或 24 聚体的预定结合序列的质量控制探针的混合物)。

5.1 具有预定结合序列的质量控制探针

5.1.1 预定结合序列

具有预定结合序列的质量控制探针是含有预定结合序列的生物聚合物并且不干扰计划的微阵列处理的结果。目的是避免结合中的交叉反应性，所要分析的样品的生物聚合物不应该结合微阵列上的质量控制探针。此外，用于结合和检测质量控制探针的预定结合序列的反向互补序列不应该结合微阵列上的试验探针(即，设计用于结合样品中的生物聚合物的微阵列上的探针)。在本发明方法中，根据将要合成的微阵列上所要分析的样品的起源、制备和处理的具体要求的组合制备质量控制探针。优选地，其中将要在微阵列上分析的样品含有天然发生的核酸或蛋白质，质量控制探针的预定结合序列分别不存在或者不知道存在于天然发生的核酸中或者不知道编码天然发生的蛋白质。在另一实施方案中，质量控制探针的预定结合序列不存在于或不知道存在于样品中。这样做是为了减少预定结合序列将具有交叉反应性的可能性。交叉反应性表明微阵列处理过程中生物聚合物能够与一种以上的其他生物聚合物相互作用(例如，杂交或结合)。例如，在寡核苷酸微阵列处理过程中，如果预定结合序列与其互补核酸以及与生物样品中的不同序列杂交，那么就说该探针是交叉反应的。探针中的交叉

反应性是不希望的，因为其可以改变从样品处理观察到的信号强度并影响微阵列合成质量的评估。

在一个实施方案中，组成预定结合序列的单体的潜在序列可以从随机合成序列库鉴定。然后可以分析这些潜在的预定结合序列与所要处理的生物样品或者设计用于检测处理过程中生物样品的天然发生序列的探针的交叉反应性。优选地，在质量控制探针中使用基本不与样品中所分析的生物聚合物交叉反应的预定结合序列。从而，在微阵列合成时，质量控制探针的序列是已知的，尽管该序列是随机的，因为其最初是随机合成的产物。随机序列是生物聚合物残基(例如，核苷酸或氨基酸序列)，它们是关于实际所得序列不进行预先计划的特定设计而产生的序列，即，当说单体(例如，核苷酸、氨基酸)是随机的时候，哪个单体将在该残基处发生是不可预知的。随机序列可以通过无偏性合成方案合成，其中每种可能的残基具有相等机会被掺入到生物聚合物的每个位置。备选地，随机序列可通过有偏性合成方案合成，其中生物聚合物的一些位置具有一种残基比另一种残基的机会高。此外，可以组合使用无偏性和有偏性合成方法合成任一生物聚合物。在一个实施方案中，任一末端或内部位置上的序列可被加入到预定结合序列中以方便标准分子生物学操作。一旦产生，就确定了随机产生的预定结合序列的序列。优选地，试验该序列的交叉反应性，并记录备用。对于每个阵列，在微阵列上合成经验确定为非交叉反应的预定结合序列的一种或多种以允许将来评估合成质量或阵列中的其他非合成缺陷。

在另一实施方案中，预定结合序列可以是天然发生的序列，其不是将要在微阵列上处理的样品内源的。例如，如果该样品来自真核生物来源，那么可以将细菌的序列(或其片段)用作预定结合序列。可以作为预防措施评估交叉反应性。

因此，当质量控制探针的预定序列的结合伴侣不内源性存在于将要分析与微阵列结合的样品中时，在样品与微阵列接触前的任何时候或接触过程中将质量控制探针中的预定结合序列的结合伴侣导入到样品中。在一个实施方案中，结合伴侣在样品处理过程中被加入到样品中。在更优选的实施方案中，在样品与微阵列接触前将结合伴侣加到样品。

预定结合序列可以由任何类型的生物大分子制备；优选地，质量控制探针的分子性质与微阵列上其他生物聚合物探针的性质一致。例如，预定结合序列可以由核苷酸(即 DNA 或 RNA)、氨基酸、聚糖、糖类或小有机分子组成。

在一个实施方案中，预定结合序列是核酸，优选寡核苷酸，在易于杂交的条件下，核酸微阵列与含有包含与预定结合序列互补的序列的核酸的样品接触，并比较杂交到质量控制探针的量。

在另一实施方案中，预定结合序列是蛋白质(多肽或肽)，并且在易于结合的条件下蛋白质微阵列与含有所述蛋白质的结合伴侣的样品接触，并比较结合到质量控制探针的量。在一个实施方案中，结合部分是被抗体，优选单克隆抗体识别的表位。优选地，表位是唯一的(即，不在为样品提供蛋白质材料的细胞或组织中内源性表达)以在检测过程中使针对预定结合序列表位的抗体与样品表位的交叉反应性最小化。

预定结合序列的长度可以根据用于检测所要评估的样品中结合伴侣的微阵列上其他生物聚合物探针的长度而变。典型地，预定结合序列由比微阵列上其他生物聚合物探针更少数目的单体组成。这允许预定结合探针仅代表组成微阵列上其他生物聚合物探针的全部单体的子集。这样，需要许多预定结合序列来代表每个全长生物聚合物探针。这允许根据合成的哪些周期用于合成预定结合序列通过不同质量控制探针靶向合成的不同周期以进行检查。可以比较不同预定结合序列的结合强度以确定关于全长生物聚合物探针的不同部分的信息。预定结合序列优选为微阵列上其他生物聚合物探针长度的 5-95%、10-75%、25-65%、35-60%、 40-55%或 41-48%。在另一实施方案中，当微阵列上其他探针长为 60 个生物聚合物残基时，预定结合序列为 15 个生物聚合物残基。在另一实施方案中，当微阵列上其他探针长为 60 个生物聚合物残基时，预定结合序列为 25 个生物聚合物残基。微阵列上非质量控制探针的生物聚合物探针当为核酸时，长度为 10-500 个核苷酸，更优选地，10-250、20-100、40-80、50-70 或 60 个核苷酸。

5.1.1.1 具有故意缺失的预定结合序列

在一些实施方案中，相对于结合用于检测微阵列处理过程中质量

控制探针的结合伴侣的序列，预定结合序列具有一个或多个单体的故意缺失。从而，在特定实施方案中，相对于与用于通过杂交检测质量控制探针的核酸完美互补的序列，预定结合序列具有内部缺失。尽管由于有缺陷的结合对而降低信号强度，但是仍然可以观察到信号。由于微阵列合成中的故障导致的任何额外的缺失将增加预定结合序列和结合伴侣之间的差异并从而剧烈减小微阵列处理过程中观察到的信号。在一个实施方案中，在 oligo 微阵列上，预定结合序列是 24 聚体(即，一个单体故意缺失)并且结合伴侣是 25 聚体。

在一个实施方案中，微阵列上每个质量控制探针含有包含一个或多个这些故意缺失的预定结合序列。在另一实施方案中，微阵列上质量控制探针是含有包含一个或多个故意缺失的预定结合序列的那些质量控制探针和含有没有故意缺失的预定结合序列的那些质量控制探针的混合物。

5.1.2 间隔区

在一些实施方案中，质量控制探针含有与预定结合序列相邻的化学结构。该化学结构在此处称为间隔区。该间隔区优选由 0 到 N 个单体(例如，核苷酸、氨基酸残基)组成，其中 N 是等于或大于 1 的整数。优选地，所加入的间隔区小于 75%，小于 50%、小于 25%、小于 20%、小于 15%、小于 10%、小于 5%、或小于 1%的质量控制探针的总序列。间隔区可以在预定结合序列的一边或两边。对于核酸探针，间隔区可以在预定结合序列的 5'或 3'或 5'和 3'。在一个实施方案中，间隔区仅在预定结合序列的 3'。对于蛋白质探针，间隔区可以在预定结合序列的氨基或羧基末端或者氨基和羧基末端。在特定实施方案中，间隔区是核苷酸或蛋白质序列。

在一个实施方案中，通过在微阵列上合成的不是质量控制探针的生物聚合物探针(即，不含有预定结合序列的探针)的长度确定 N 的上限值。质量控制探针的总长度优选不大于微阵列上其他生物聚合物探针的总长度。因此，在特定实施方案中，N 加上预定结合序列中单体数目将等于阵列上不是质量控制探针的生物聚合物探针中的单体总数。

在另一实施方案中，N 的值不受阵列上合成的不是质量控制探针

的生物聚合物探针(即，不含有预定结合序列的探针)长度的限制。在该实施方案中，质量控制探针可以比微阵列上其他生物聚合物探针更短或更长。

间隔区优选地不与样品中所分析的生物聚合物交叉反应。在微阵列处理过程中，优选地从间隔区检测不到信号。此外，微阵列处理过程中间隔区不应干扰从与其结合伴侣结合的预定结合序列产生的信号。在一个实施方案中，对该信号的干扰被防止了，因为组成间隔区的化学结构被修饰从而该间隔区不能结合结合伴侣。例如，不能杂交的被修饰的核酸可以用于间隔区并且将是非交叉反应的，例如，无碱基核苷酸(即，缺少核苷酸基础，但是具有糖和磷酸盐部分的部分)(一般见美国专利6,248, 878; Takeshita 等,1987, R Biol. Chem. 262: 10171; 无碱基核苷酸可通过商业途径从 Virginia, Sterling 的 Glen Research 得到)。在另一实施方案中，间隔区可以由与微阵列上其他生物聚合物探针中存在的单体不同的化学部分制成，该化学部分不专用于质量控制和/或组成预定结合序列的单体。例如，核苷酸微阵列上，间隔区可以从非核苷酸部分如聚醚、聚胺、聚酰胺，或聚烃化合物制造。特定实例包括Seela和Kaiser, 1990, Nucleic Acids Res. 18: 6353; Seela 和 Kaiser, 1987, NucleicAcids Res. 1987, 15: 3113;Cload 和 Schepartz, 1991, J. Am. Chem. Soc. 113: 6324; Richardson 和 Schepartz, 1991, JAm. Chem. Soc. 113: 5109; Ma 等, 1993, Nucleic Acids Res. 21: 2585 ; Ma 等, 1993, Biochemistry 32:1751 ; Durand 等, 1990, NucleicAcids Res. 18: 6353; McCurdy 等, 1991, Nucleosides & Nucleotides 10: 287;Jaschke 等, 1993, Tetrahedron Lett. 34: 301; Ono 等, 1991, Biochemistry 30: 9914; Ferentz 和 Verdie, 1991, J. Am. Chem. Soc. 113: 4000; 美国专利6,362, 323;国际公布号 WO 89/02439 所描述的那些。

优选地，一旦产生，就确定完整质量控制探针序列，试验交叉反应性并记录备用。

5.2 没有预定结合序列的质量控制探针

在一些实施方案中，质量控制探针没有预定结合序列但是仅由一个间隔区构成。用该类型的质量控制探针观察到的信号 1)从组成质量

控制探针的化学结构(例如单体)直接发出或者 2)通过使用与组成质量控制探针的化学结构(例如单体)相互作用的标记染料间接发出。这些类型的质量控制探针可以发出信号而不用使用标记的结合伴侣。

从而，在一个实施方案中，用标记的单体合成质量控制探针。标记的单体可以是，例如，荧光标记的(例如，Cy3，Cy5)核苷酸或荧光标记的氨基酸。可用的其他标记包括，但不限于，富电子分子和放射活性同位素。在合成过程中每个质量控制探针掺入一个或多个被标记的单体。

在一个特定实施方案中，其中被标记单体被掺入质量控制探针的合成周期随着每种质量控制探针而变。合成的每个周期由至少一个，但是优选地一个以上的质量控制探针代表，该质量控制探针具有在那个合成周期中沉积的单体中的标记。如果发生合成缺陷，那么没有标记单体被掺入并可以检测该缺陷。在优选的方面，每个质量控制探针长度相同。

在另一特定实施方案中，质量控制探针由与组成微阵列上试验探针(即不是质量控制探针的微阵列上的那些探针)的单体数目相同的单体组成，这些质量控制探针的单体的一种被标记。

在另一特定实施方案中，质量控制探针为不定长度从而有至少一种，但是优选一种以上的在合成的每个周期终止的质量控制探针。在这种质量控制探针中，每种质量控制探针的最后一个单体是被标记的单体。

在另一实施方案中，使用未标记单体合成没有预定结合序列的质量控制探针。所产生的信号依赖于单体的内在能力以产生信号，例如，荧光。可以合成不定长度的核酸质量控制探针并且在通过杂交到标记探针处理前可以扫描微阵列。所观察的荧光程度由于更长探针中单体(核苷酸)数增加而应该与质量控制探针的长度相关。

在另一实施方案中，可以使用直接结合质量控制探针的单体的标记染料产生可检测信号。例如，对于核酸微阵列，可以使用各种荧光核酸染色剂，如 POPO、SYBR GreenI、SYBR GreenII、SYTO 59 和 SYTO 61 (从 Molecular Probes, Inc. in Eugene, OR 得到)。评估微阵列合成效率后，在用试验样品孵育微阵列之前除去染料。

5.3 微阵列上质量控制探针的合成

在逐步生物聚合物探针合成到微阵列基质上的过程中，在微阵列的一个或多个区域在合成的一个或多个合成周期中存在错误的单体加入。为了辨别是否发生了这种合成错误，在微阵列的不同位置合成本发明的质量控制探针并比较不同质量控制探针的信号。与所预期信号的显著信号偏离表明合成缺陷(见部分 5.4)。

5.3.1 垂直放置

在一个实施方案中，通过结合到预定结合序列或者通过掺入标记单体产生可检测信号的质量控制探针可以互相垂直放置以评估合成的所有周期的效率。在一个实施方案中，在合成的不同周期逐步单体加入过程中启动预定结合序列的合成。因此，尽管一组质量控制探针的每个预定结合序列是相同的，但是在微阵列上产生预定结合序列的合成周期以垂直的方式相互放置。在另一实施方案中，其中标记单体被掺入到质量控制探针的合成周期改变从而合成的每个周期应该已经在至少一种质量控制探针中掺入一种标记单体。可以使用这些方法精确确定受错误单体加入影响的合成周期。

在一个实施方案中，该垂直放置通过使用间隔区实现。例如，通过改变间隔区中单体数，预定结合序列的微阵列(合成在其上开始)合成周期也将变化。因此，这使得每种预定结合序列易遭受合成的不同周期发生的单体加入缺陷。如果微阵列合成中没有合成缺陷，那么预定结合序列的结合伴侣将同等好(即类似地)地结合到所有质量控制探针上的预定结合序列。在确定是否不同质量控制探针上预定结合序列的结合伴侣类似结合，应该明白，当质量控制探针含有间隔区时，结合的不同可能部分由于预定结合序列与微阵列的距离(见部分 5.4)。从而，当确定是否不同质量控制探针是结合“类似的”时，所预期的由于不同间隔区长度导致的结合差异优选被忽视。

在特定实施方案中，一个组的质量控制探针都含有相同的预定结合序列但是由于含有间隔区的不定数目的单体而在质量控制探针中单体的总数中不同。

在另一实施方案中，通过改变标记单体掺入质量控制探针的微阵列合成周期实现该垂直放置。因此，这使得每个标记单体加入易遭受

在合成的不同周期发生的单体加入中的缺陷。如果在微阵列合成中没有合成缺陷，那么每种质量控制探针将已经被掺入并且具有相等数目的标记单体并从而给出可比较的信号。

在另一实施方案中，该垂直放置通过交错开始信号完成。如上，每个预定结合序列在其合成开始时通过单体加入的一个或多个序贯周期相互放置。在一个实施方案中，使用间隔区完成该放置。在更优选的实施方案中，不使用间隔区完成该放置。相反地，微阵列上该位置的单体加入被延迟以被预定结合序列占据直到微阵列合成已经达到某一周期，在该周期预定结合序列的合成将被启动。在该实施方案中，所有质量控制探针含有相同数目的单体但是在微阵列的不同位置(相应于质量控制探针)上使用这些单体的合成被暂时分开。

5.3.2 水平放置

可以在微阵列基质上许多不同位置中合成本发明的质量控制探针以组成许多不同模式。这些模式可用于鉴定微阵列合成缺陷的区域以及在处理过程中将位置信息传递给微阵列。微阵列上质量控制探针的数目应该足够充分代表整个微阵列的合成。例如，微阵列上为质量控制探针的探针数目应该是微阵列上总探针的约 0.5%或更多、1%或更多、2 或更多、3%或更多、5%或更多、10%或更多、20%或更多。

在一个实施方案中，可以在全部微阵列的不同位置合成一或多排质量控制探针(称为网格线)。微阵列的每个部分可以含有一个网格线以确保已经对所有部分评估了令人满意的合成。在一个实施方案中，微阵列边缘生物聚合物探针合成的完整性可以通过使用外部(或外周)网格线，例如质量控制探针的 1-5 邻边来监控(图 17A)。边缘或接近边缘的微阵列部分可专用于质量控制探针从而如果存在任何缺陷就可检测到该缺陷。在另一实施方案中，微阵列中心的生物聚合物探针合成的完整性可以通过使用对角网格线监控(图 17B)。可以在对角穿过该阵列从而穿过微阵列的代表部分的位置中合成质量控制探针。在优选的实施方案中，网格线模式由含有间隔区的质量控制探针组成。

在另一实施方案中，可以在微阵列的部分中合成质量控制探针群以评估合成质量。在另一实施方案中，在阵列整个中央的随机位置中合成质量控制探针(图 17C)。在另一实施方案中，可以在微阵列的角

落合成质量控制探针(图 17D)。

在另一实施方案中，当通过喷墨技术合成微阵列时，质量控制探针可以在微阵列上排列从而可以检测一个或几个特定喷嘴的故障。与通过特定喷嘴印刷的周期性一致的质量控制探针中信号强度的减少可以表示那个喷嘴已经反复有缺陷。当喷墨头中有 N 个喷嘴时，周期为 N 的质量控制探针强度的减少表示被堵塞或有缺陷的喷嘴(其中 N 为 1 或更大的整数)。在一个实施方案中，N 为 20。在另一实施方案中，用对角网格线(图 17B)评估喷嘴堵塞或缺陷。

在另一实施方案中，质量控制探针模式可用于告知关于微阵列的位置信息。因为微阵列合成过程中质量控制探针合成的位点是已知的，所以可以使用探针调准处理过程中的微阵列。

5.4 微阵列上缺陷的检测

含有相同预定结合序列的所有质量控制探针应该类似地结合到结合伴侣。然而，在许多情况中，研究人员已经发现增加预定结合序列与微阵列距离的间隔区实际上增强了结合给定的预定结合序列时的信号强度，该信号强度的增强通过将其与从直接粘附到微阵列的相同的预定结合序列观察到的信号比较得到。不被特定机理所限制，该增强的信号强度可能是由于预定结合序列由于距离微阵列更远(例如通过将间隔区直接附着到含有与该预定结合序列邻接的数目增加的单体的微阵列)而更易接近其结合伴侣所导致。大于由于预定结合序列与微阵列的不同距离导致的不同质量控制探针和结合伴侣之间的结合量的偏离可表明微阵列质量中的存在问题。微阵列质量中的缺陷可以是全局的(即，该缺陷影响整个微阵列)或者局部的(即，该缺陷影响微阵列的一个或多个区域并且不影响其他区域)。

在特定实施方案中，本发明的质量控制探针的使用允许检测微阵列合成缺陷(例如，流槽梯度，其中流槽中气泡或其他问题导致一些合成周期中微阵列的非均一试剂覆盖。然而，通过使用本发明的质量控制探针也可以检测影响微阵列质量的其他类型的缺陷。微阵列中的缺陷可以是由于除了合成缺陷之外的事件。质量控制探针也可用于检测这些类型的缺陷。在一个实施方案中，通过本发明的方法可检测的微阵列缺陷在微阵列的保存过程中发生。亚最佳条件(例如，不适宜

的温度或湿度水平)可导致微阵列质量恶化。可通过本发明的方法检测的其他缺陷包括但不限于导致微阵列上局部缺陷的磨损。这种磨损可以在保存或微阵列处理过程中发生。在微阵列处理过程中可以发生缺陷。这种缺陷可导致信号的非均一性，其可通过比较横跨微阵列的信号强度来检测。可以以许多方法完成结合强度的比较。

在一个实施方案中，可以计算每组质量控制探针的结合比率。对于含有预定结合序列且不含有间隔区的质量控制探针，微阵列处理过程中关于特定质量控制探针产生的信号应该等于关于另一不同质量控制探针产生的信号。两种信号的比例应该接近 1。与 1 偏离表明计算中所用的两种质量控制探针之一对其结合伴侣的结合减弱。如果合成缺陷导致质量控制探针中预定结合序列是有缺陷的并从而不能以正常水平结合其结合伴侣，那么将出现这种情况。对于含有预定结合序列和间隔区的质量控制探针，微阵列处理过程中关于特定质量控制探针产生的信号由于与微阵列距离的不同而可能会或不会等于关于另一不同质量控制探针产生的信号。从距离微阵列类似距离的预定结合序列(例如，每种预定结合序列的合成在相互的 3 个合成周期内启动)得到的两种信号的比例应该接近 1。然而，从距离微阵列不同距离的预定结合序列(例如，每种预定结合序列的合成在彼此合成 3 个周期以上启动)得到的两种信号的比例可以偏离 1。在该实施方案中，可以使用从已知没有缺陷的微阵列得到的数据确定预期的比例。可以鉴定这些微阵列，例如通过制备许多阵列(优选地至少 5 个)并比较结果以鉴定没有缺陷的阵列)进行鉴定。与该确定的预期比例的偏离可以用于检测微阵列中的缺陷。

对于每种类型的微阵列(例如，寡核苷酸、蛋白质，等)，可以经验地确定表明没有缺陷的结合比率值的范围。例如，已知没有缺陷的各种预定结合序列可以结合它们的结合伴侣并记录信号。这可用作用于确定预期的结合比率的基线值。通过改变微阵列上质量控制探针的水平和垂直放置，可以确定一系列可接受的比例。与这些通过经验确定的比例的偏离表明有缺陷微阵列。在特定实施方案中，当微阵列是寡核苷酸微阵列时，0.25 到 2.25、0.5 到 2.0、或 0.75 到 1.25 的结合比率表明没有合成缺陷。

在特定实施方案中，质量控制探针是含有包含相对于结合用于检

测微阵列处理过程中质量控制探针的结合伴侣的序列的一个或多个故意缺失的预定结合序列的那些质量控制探针和含有无故意缺失的预定结合序列的那些质量控制探针的质量控制探针的混合物，对于使用该质量控制探针混合物的微阵列，可以计算结合比率并将其用于评估微阵列质量。通过将标记的结合伴侣结合到每种预定结合序列(即，有或没有故意缺失)产生的信号将必定不同。从已知没有缺陷的微阵列确定的结合强度可用于计算预期的结合比率。可以鉴定这些微阵列，例如，通过制备许多阵列(优选地至少 5 个)并比较结果以鉴定没有缺陷的阵列。与预期比例的偏离表明存在缺陷。

在另一实施方案中，结合强度的比较可通过统计分析完成。通过将为每个质量控制探针观察到的信号(例如，荧光)值平均，计算一组质量控制探针的平均结合强度。然后，为每个单独的质量控制探针所观察的信号量可以与该组的平均值比较。在一个实施方案中，在离平均值两个标准误以内的那些质量控制探针表明微阵列中没有质量缺陷，例如，没有在它们的合成过程中的缺陷，或者在处理、保存或其他方面招致的缺陷。在更优选的实施方案中，在离平均值一个标准误以内的那些质量控制探针表明没有缺陷。

在另一实施方案中，可以使用一种以上的荧光染料标记结合到预定结合序列的结合伴侣。例如，结合伴侣的一个子集可以用 Cy3 标记，一个子集可以用 Cy5 标记。可以为所用的每种类型荧光染料确定从一种质量控制探针检测的信号的比例。通过改变质量控制探针在微阵列上的水平和垂直放置，可以确定一系列可接受的比例。与该凭经验确定的比例的偏离表明微阵列缺陷。

对于使用没有预定结合序列并用标记的单体合成的质量控制探针的微阵列，可以使用类似方法检测缺陷。信号不是从预定结合序列的被标记的结合伴侣产生，而是来自粘附到微阵列的质量控制探针自身。比例和平均信号的标准误可用于评估微阵列的完整性。

对于使用没有预定结合序列或被标记单体的质量控制探针的微阵列，可以使用类似方法检测质量缺陷。然而，在这些微阵列中，可检测信号与质量控制探针的长度成比例；从而，对于不同长度的每种质量控制探针，信号强度将不类似。相反，预期从更长的质量控制探针将得到更强的信号。与预期看到的探针间差异的偏离表明微阵列中存

在缺陷。

在一个实施方案中，当使用质量控制探针的混合物时，可以凭经验确定预期的结合比率或信号强度。已知不含有缺陷的微阵列可用于为结合到其结合伴侣的预定结合序列得到基线值或者为质量控制探针的每种不同的类型得到信号强度。可以从这些数据确定比例并用作预期比例。与这些比例的偏离表明有缺陷的微阵列。

5.5 微阵列合成和处理

微阵列上的探针可以是许多不同生物聚合物的任一种，例如，DNA、RNA、肽核酸(PNAs)(见例如，Egholm 等，1993，Nature 363:566-568；美国专利号 5,539，083)，或者蛋白质。通过单体的逐步加入在固体载体上合成本发明的微阵列。每个这种单体都是一个合成周期中加入的一单位生物聚合物。在一个实施方案中，每个合成周期加入的生物聚合物单位自身由不超过一个基本生物聚合物单位(例如，核苷酸、氨基酸，等)组成。在另一实施方案中，每个合成周期加入的生物聚合物单位由一个以上的基本生物聚合物单位(例如，二核苷酸、二肽、共价连接到另一部分的核苷酸或氨基酸，等)组成。在另一实施方案中，每个合成周期加入的生物聚合物的单位随着不同合成周期而变。

5.5.1 核苷酸微阵列

在本发明的一个优选的实施方案中，样品的处理是通过在核苷酸微阵列上杂交。在更优选的实施方案中，微阵列是寡核苷酸阵列。在最优选的实施方案中，寡核苷酸阵列是喷墨合成的寡核苷酸微阵列。优选地，该微阵列含有 20 到 50,000 个核酸探针。这些探针可以以各种模式排列。例如，探针可以以行和列，多边形(例如，六边形)，或圆形模式等排列。

杂交水平优选使用多核苷酸探针阵列或微阵列检测。在多核苷酸阵列上，含有目标序列的多核苷酸探针被固定到载体，例如，固体载体的表面上。例如，探针可含有 DNA 序列，RNA 序列，或 DNA 和 RNA 的共聚物序列。探针的多核苷酸序列还可以含有 DNA 和/或 RNA 类似物(例如，肽核酸)，或它们的组合。例如，探针的多核苷酸序列可以

是来自细胞的基因组 DNA 或 mRNA 的全部或部分序列，或者可以是从中得到的 cDNA 或 cRNA 序列。

本发明方法中使用的一种或多种探针优选被固定到固体载体或表面，该固体载体或表面可以是有孔的或无孔的。例如，本发明的探针可以是附着到硝酸纤维素或尼龙膜或滤器的多核苷酸序列。这些杂交探针是本领域中熟知的(见例如，Sambrook 等，编者，1989，分子克隆：实验指南，1-3 卷，第二版，冷泉港实验室，冷泉港，纽约)。备选地，固体载体或表面可以是玻璃或塑料表面。

5.5.1.1 使用微阵列杂交分析

微阵列是载体上位置可寻址的结合(例如，杂交)位点的阵列。这些结合位点的每一个含有结合到载体的预定区域的探针的许多多核苷酸分子。可以用多种方法制造微阵列，下面描述了其中的一些方法(例如，例如，Meltzer，2001，Curr. Opin. Genet. Dev. 11(3)：258-63；Andrews 等，2000，Genome Res. 10(12)：2030-43；Abdellatif，2000，Circ. Res. 86 (9)：919-20；Lennon，2000，Drug Discov. Today 5 (2)：59-66；Zweiger，1999，Trends Biotechnol. 17(11)：429-36)。然而生产后，微阵列有一些共同特征。阵列优选是可复制的，允许生产给定阵列的多个拷贝并容易相互比较。优选地，微阵列由在结合(例如，核酸杂交)条件下稳定的材料制造。微阵列优选为 1 cm^2 到 25 cm^2，优选地约 10 cm^2 到 15 cm^2。然而，也考虑更大和更小(例如，0.5 cm^2 或更小)的阵列，例如用于同时评价非常大数目的不同探针。

在尤其优选的实施方案中，测量由固相组成的探针微阵列的杂交水平，在该固相上固定着一群多核苷酸，如一群 DNA 或 DNA 模拟物，或者备选地，一群 RNA 或 RNA 模拟物。固相可以是无孔的，或者，任选地，有孔材料如凝胶。可以使用微阵列，例如，用于分析细胞的转录状态，如暴露于目标药物的梯度水平或者对目标生物学途径的梯度干扰的细胞的转录状态。可以使用微阵列同时筛选许多不同探针以评价，例如每种探针的敏感性和对特定目标多核苷酸的特异性。

优选地，微阵列上给定结合位点或一组独特结合位点将特异结合(例如，杂交)单个基因产物或者细胞或生物体的基因转录产物(例如，结合特定 mRNA 或者由其得到的特定 cDNA)。然而，通常，其他相关的

或类似的序列可能交叉杂交到给定结合位点。

用于本发明方法和组合物中的微阵列包括一种或多种试验探针，其每一种都有与所要检测的 RNA 或 DNA 的亚序列互补的多核苷酸序列。每种探针优选地具有不同核酸序列，并且阵列的固体表面上的每个探针的位置优选是已知的。实际上，微阵列优选为可寻址的阵列，更优选地为位置可寻址的阵列。更特别地，阵列的每个探针优选地位于固体载体上已知的预定位置，从而每个探针的身份(例如，序列)可以从其在阵列上(例如，载体或表面上)的位置确定。

优选地，微阵列上探针的密度为每 $1cm^2$ 约 100 个不同(即，非-相同的)探针或更高。更优选地，本发明中使用的微阵列将具有每 $1cm^2$ 至少 550 个探针、每 $1cm^2$ 至少 1000 个探针、每 $1cm^2$ 至少 1500 个探针、或者每 $1cm^2$ 至少 2000 个探针。在尤其优选的实施方案中，微阵列是高密度阵列，优选地具有每 $1cm^2$ 至少约 2500 个不同探针的密度。本发明中使用的微阵列因此优选地含有至少 2500、至少 5000、至少 10000、至少 15000、至少 20000、至少 25000、至少 50000 或至少 55000 个不同(例如，非-相同的)探针。这些探针的一个子集将相应于可能已经加入到样品的刺入的标签 (spike-in tags)。

这些多核苷酸优选长为 15 到 200 个碱基，更优选地长为 20 到 100 个碱基，最优选 40 到 60 个碱基。应该理解每个探针序列可以除了与其靶序列互补的序列可以还含有一个衔接子(例如，间隔区)。如此处所用的，衔接子指与其靶序列互补的序列和表面之间的化学结构。衔接子不必是核苷酸序列。例如，衔接子可以由核苷酸序列，或者肽核酸、烃链等组成。

在一个实施方案中，微阵列为阵列(例如，基质)，其中每个位置代表被基因编码的转录产物(例如，mRNA 或从 mRNA 得到的 cDNA)的不同结合位点。例如，在各种实施方案中，本发明的微阵列可以含有被生物基因组中少于 50%基因编码产物的结合位点。备选地，本发明的微阵列可以具有生物基因组中至少 50%、至少 75%、至少 85%、至少 90%、至少 95%、至少 99%或 100%，或者至少 50、100、500、1000、或 10000 个基因编码的产物的结合位点。在其他实施方案中，本发明的微阵列可以具有生物的一个细胞表达的基因的少于 50%、至少 50%、至少 75%、至少 85%、至少 90%、至少 95%、至少 99%或至少 100%所编码的产物的

结合位点。结合位点可以是特定 RNA 可以特异杂交的 DNA 或 DNA 类似物。该 DNA 或 DNA 类似物可以是例如，合成的寡聚体或基因片段，例如相应于外显子的基因片段。

优选地，本发明中使用的微阵列具有与目标药物的作用相关或者目标生物学途径中的一个或多个基因的几组基因的结合位点(例如，探针)。如上所讨论的，“基因”被识别为被 RNA 聚合酶转录的 DNA 的一部分，其可以包括 5’非翻译区(UTR)、内含子、外显子和 3’UTR。基因组中基因数目可以从细胞或生物表达的 mRNA 分子数目估计，或者通过对基因组的良好表征部分的外推估计。当目标生物的基因组已经被测序时，可以确定开放阅读框(ORFs)的数目并通过分析 DNA 序列鉴定 mRNA 编码区。例如酿酒酵母(Saccharomyces cerevisiae)的基因组已经被完全测序并据报导有约 6275 个编码氨基酸残基长度大于 99 的 ORF 编码序列。对这些 ORF 的分析指出有 5,885 个可能编码蛋白质产物的 ORF(Goffeau 等, 1996, Science 274 : 546-567)。相比，估计人基因组含有约 30000 到 130000 个基因(见 Crollius 等, 2000, Nature Genetics 25: 235-238; Ewing 等, 2000, Nature Genetics 25: 232-234)。还完成或接近完成其他生物的基因组序列，这些生物包括但不限于，果蝇、线虫(C. elegans)、植物，例如，稻和拟南芥，和哺乳动物，例如，小鼠和人。从而，在本发明的优选实施方案中，提供了含有生物基因组中所有基因探针的阵列组。

将明白当制备了靶核酸分子，例如与细胞的 RNA 互补的 cDNA 的样品并在适宜杂交条件下杂交到微阵列时，杂交到阵列中位点的水平将反映样品中相应互补序列的优势。例如，当可检测被标记(例如，用荧光团)cDNA 被杂交到微阵列时，相应于不在样品中的核苷酸序列的阵列上位点将有很小的或者没有信号(例如，荧光信号)，并且在样品中占优势的核苷酸序列将具有相对强的信号。样品中不同核苷酸序列的相对丰度可以通过微阵列上探针的信号强度模式确定。

使用两色方案，处于两种不同条件下的来自两种不同细胞的样品的核酸可以杂交到微阵列的结合位点。对于药物应答，一种细胞样品暴露于药物，相同类型的另一种细胞样品不暴露于药物。来自两种细胞类型的每一种的 cDNA 被不同标记(例如，用 Cy3 和 Cy5)从而它们可被区别。在一个实施方案中，例如，使用荧光素标记的 dNTP 合成来

自药物处理的(或者具有突变或疾病，等)细胞的 cDNA，使用罗丹明标记的 dNTP 合成没有暴露药物的第二种细胞的 cDNA。当两种 cDNA 分子被混合并杂交到微阵列时，为阵列上每个位点确定来自每个 DNA 组的信号的相对强度，并检测特定基因丰度的相对差异。

在上述实施例中，当荧光团被激发时，来自药物处理的细胞的核酸将发出绿色荧光而来自未处理细胞的核酸将发出红色荧光。结果，当药物处理对细胞中特定基因的转录直接或间接无效时，两种细胞中的表达模式将是不可区别的，反转录时，红色标记的和绿色标记的核酸将具有相等优势。当杂交到微阵列时，那种核酸的结合位点将发出两种荧光团的特征波长。相反，当用直接或间接改变该细胞中特定基因的转录的药物处理暴露于药物的细胞时，通过每个结合位点的绿色荧光与红色荧光的比例代表的表达模式将改变。当药物增加一种 mRNA 的优势时，该 mRNA 的每个结合位点的比例将增加，而当药物减少了 mRNA 的优势时，该 mRNA 中每个结合位点的每个比例将减小。

使用两色荧光标记和检测方案定义基因表达中的变化连同 mRNA 分子的检测已经在例如 Shena 等，1995，用互补 DNA 微阵列定量监控基因表达模式，Science 270: 467-470 中描述。使用用两种不同荧光团标记的 cDNA 的一个优点是可以进行两种细胞状态中相应于每种排列的基因的 mRNA 或外显子表达水平的直接和内部受控的比较。并且由于实验条件(例如，杂交条件)中微小不同造成的变化将不会影响随后的分析。然而，应该认识到还可能使用来自单个细胞的 cDNA，并比较，例如，在例如，药物处理的或途径-被干扰的细胞和未处理细胞中特定外显子的绝对量。此外，本发明中也考虑用两种以上的颜色标记。在本发明的一些实施方案中，至少 5、10、20 或 100 种不同颜色的染料可用于标记。该标记允许可区分的被标记 cDNA 群体同时杂交到相同阵列，从而测量，并任选比较来自两种以上样品的 mRNA 分子的表达水平。可使用的染料包括，但不限于，荧光素及其衍生物、罗丹明及其衍生物、得克萨斯红、5'羧基-荧光素(FMA)、2',7'-二甲氧基-4',5'-二氯-6-羧基-荧光素(JOE)、N,N,N',N'-四甲基-6-羧基-罗丹明(TAMRA)、6'羧基-X-罗丹明(ROX)、HEX、TET、IRD40 和 IRD41、cyamine 染料，包括但不限于 Cy3、Cy3.5 和 Cy5; BODIPY 染料包括但不限于 BODIPY-FL、BODIPY-TR、BODIPY-TMR、BODIPY-630/650、和

BODIPY-650/670;和 ALEXA 染料，包括但不限于 ALEXA-488、ALEXA-532、ALEXA-546、ALEXA-568、和 ALEXA- 594；以及本领域技术人员公知的其他荧光染料。

5.5.1.2 制备微阵列的探针

如上所提到的，特定多核苷酸分子特异杂交的探针是互补的多核苷酸序列。通常，微阵列上每个探针为 20 个碱基到 600 个碱基，并且通常长为 30 到 200 个碱基。

产生微阵列的多核苷酸探针的方法是通过合成合成的多核苷酸或寡核苷酸，例如，使用 N-膦酸盐或亚磷酰胺化学合成(Froehler 等，1986, Nucleic Acid Res. 14: 5399-5407; McBride 等，1983, Tetrahedron Lett. 24: 246-248)。合成序列通常长为约 15 到约 600 个碱基，更典型地约 20 到约 100 个碱基，最优选地长为约 40 到约 70 个碱基。

微阵列上探针为粘附到微阵列的固体载体的大分子。在本发明中，探针优选为核酸序列(或其片段)。

5.5.1.3 粘附探针到固体表面

本发明的方法使用直接在载体上合成的多核苷酸探针形成阵列。探针附着到固体载体或表面，其可以用例如玻璃、塑料(例如，聚丙烯、尼龙)、聚丙烯酰胺、硝酸纤维素、凝胶、或其他有孔或无孔材料制造。

制造微阵列的一种方法是制造高密度寡核苷酸阵列。有各种公知的技术用于生产在表面上限定的位置含有数千与限定序列互补的寡核苷酸的阵列。例如，可以使用用于原位合成的照相平板印刷技术(见，Fodor等 ，1991, Science 251: 767-773; Pease等，1994, Proc. Natl. Acad. Sci. U. S. A. 91: 5022-5026; Lockhart 等,1996, Nature BioTechnology 14: 1675; 美国专利号5,489, 678; 5,578, 832; 5,556, 752; 5,510, 270; 6,197, 506; 和 6,346, 413)或用于快速合成和限定寡核苷酸的沉积的其他方法(Blanchard 等，Biosensors & Bioelectronics 11: 687-690)。

也可以使用例如通过掩蔽制造微阵列(Maskos 和 Southern,

1992,Nucl. Acids. Res. 20: 1679-1684)的其他方法。原则上，并且如前述，可以使用任何类型的阵列，例如，尼龙杂交膜上的点印迹(见 Sambrook，如前)。然而，如本领域技术人员将要认识到的，通常将优选非常小的阵列，因为杂交体积将更小。

在尤其优选的实施方案中，通过用于寡核苷酸合成的喷墨印刷设备，例如，使用 Blanchard 在国际专利公布号 WO98/41531(1998 年 9 月 24 日公布)；Blanchard 等.，1996，生物传感器和生物电子学 11: 687-690; Blanchard, 1998, Synthetic DNA Arrays in Genetic Engineering, Vol. 20, J. K. Setlow,编者, Plenum Press, New York, 111-123 页; Hughes 等, 2001, Nature BioTechnology 19: 342-347; 和 Blanchard 的美国专利号 6,028, 189 描述的方法和系统制造本发明的微阵列。特别地，优选在阵列中，例如，在玻璃载玻片上，通过将单个核苷酸碱基连续沉积在高表面张力溶剂如碳酸丙烯的微滴中合成这些微阵列中的寡核苷酸探针。微滴具有小的体积(例如，100pL 或更小，更优选地 50pL 或更小)并且在微阵列上相互分离(例如，通过疏水结构域)以形成圆形表面张力孔，其限定了阵列元件(即，不同探针)的位置。多核苷酸探针共价粘附到该多核苷酸 3'末端处的表面。

当使用这些方法时，在表面如衍生化玻璃载玻片上直接合成已知序列的寡核苷酸(例如，60 聚体)。产生的阵列可以是冗余的，每个基因有几个寡核苷酸分子。

5.5.1.4 目标多核苷酸分子

目标多核苷酸是在微阵列上被处理的生物样品的多核苷酸。目标多核苷酸可以是 RNA 分子如，但绝不限于信使 RNA(mRNA)分子、核糖体 RNA(rRNA)分子、cRNA 分子(即，体外转录的从 cDNA 分子制备的 RNA 分子)和其片段。此外，目标多核苷酸还可以是，但不限于，DNA 分子如基因组 DNA 分子、cDNA 分子，和其片段，包括寡核苷酸、ESTs、STSs 等。在特定实施方案中，样品含有 1000、5000、10000、50000、100000、250000 或 1000000 以上的不同核苷酸序列的核酸分子。

目标多核苷酸可以来自任何来源。例如，目标多核苷酸分子可以是天然发生的核酸分子，如从生物分离的基因组或基因组外 DNA 分子，或者从生物分离的 RNA 分子，如 mRNA 分子。备选地，多核苷酸分子

可以是合成的，包括，例如，通过酶促方法在体内或体外合成的核酸分子，如 cDNA 分子，或通过 PCR 合成的多核苷酸分子，通过体外转录合成的 RNA 分子，等。目标多核苷酸的样品可以含有，例如，DNA、RNA 或 DNA 和 RNA 共聚物的分子。在优选的实施方案中，本发明的目标多核苷酸将相应于特定基因或特定基因转录产物(例如，细胞中表达的特定 mRNA 或从这些 mRNA 序列得到的 cDNA 序列)。然而，在许多实施方案中，尤其是其中多核苷酸分子来自哺乳动物细胞的那些实施方案中，目标多核苷酸可以相应于基因转录产物的特定片段。例如，目标多核苷酸可以相应于相同基因的不同外显子，例如，从而可以检测和/或分析该基因的不同剪接变体。

在优选的实施方案中，从细胞提取核酸体外制备所要分析的目标多核苷酸。例如，在一个实施方案中，从细胞提取 RNA(例如，总细胞 RNA，poly(A)+信使 RNA，其片段)并从总提取的 RNA 纯化 mRNA。制备总的和 poly(A) + RNA 的方法是本领域熟知的，并且一般地，例如在 Sambrook 等(如前)中描述。在一个实施方案中，使用硫氰酸胍裂解然后通过 CsCl 离心和寡 dT 纯化从本发明的各种类型的目标细胞提取 RNA(Chirgwin 等，1979，Biochemistry 18: 5294- 5299)。在另一实施方案中，使用硫氰酸胍裂解然后通过 Rneasy 柱(Qiagen)纯化从细胞提取总 RNA。然后使用例如，寡聚-dT 或随机引物从纯化的 mRNA 合成 cDNA。在优选的实施方案中，目标多核苷酸为从 mRNA 或自细胞提取的总 RNA 制备的 cDNA 制备的 cRNA。如此处所用的，cRNA 可以与样品 RNA 互补(反义)或者与样品 RNA 的序列相同(有义)。使用一种方法扩增提取的 RNA 分子，该方法中，使用连接到 RNA 聚合酶启动子的引物从样品 RNA 分子合成双链 cDNA 分子。结果，RNA 聚合酶启动子可以掺入到 cDNA 链的一条或两条。使用 cDNA 分子的第一条链上的 RNA 聚合酶启动子，可以合成与样品 RNA 序列相同的 cRNA。为了合成与样品 RNA 互补的 cRNA，使用 RNA 聚合物可以从双链 cDNA 分子的第二条链上的 RNA 聚合酶启动子启动转录(见，例如，美国专利号 5,891，636、5，716，785；5,545，522 和 6,132，997；还见，美国专利号 6,271，002 和美国临时专利申请序列号 60/253,641，由 Ziman 等在 2000 年 11 月 28 日提交)。可以使用含有 RNA 聚合酶启动子或其互补序列的寡聚-dT 引物(美国专利号 5,545，522 和 6,132，997)或随机引物美国临

时专利申请序列号 60/253,641,由 Ziman 等在 2000 年 11 月 28 日提交)。优选地,目标多核苷酸为短的和/或片段化多核苷酸分子,它们代表细胞的最初核酸群体。在一个实施方案中,总 RNA 用作 cRNA 合成的进料。含有 T7 RNA 聚合酶启动子序列的寡聚-dT 引物可用于引发第一条链 cDNA 合成。当期望第二条链合成时,可以使用随机六聚体通过反转录酶引发第二条链 cDNA 合成。该反应产生在 3'末端含有 T7 RNA 聚合酶启动子的双链 cDNA。然后可以通过 T7 RNA 聚合酶将双链 cDNA 转录成 cRNA。

将要分析的目标多核苷酸优选为可检测地标记。例如,cDNA 可以例如用核苷酸类似物直接标记,或者例如,通过第二条标记的 cDNA 链用第一条链作为模板间接标记。备选地,双链 cDNA 可以被转录为 cRNA 并被标记。

优选地,可检测的标记是荧光标记,例如,通过掺入核苷酸类似物。适宜用于本发明的其他标记包括,但不限于,生物素、imminobiotin、抗原、辅因子、二硝基酚、硫辛酸、烯烃化合物、可检测的多肽、富电子分子、能够通过作用于底物产生可检测信号的酶、和放射性同位素。优选的放射性同位素包括 ^{32}P、^{35}S、^{14}C、^{15}N 和 ^{125}I。适于 本发明的荧光分子包括,但不限于,荧光素及其衍生物、罗丹明及其衍生物、得克萨斯红、5'羧基-荧光素(FMA)、2',7'-二甲氧基-4',5'-二氯-6-羧基-荧光素(JOE)、N,N,N',N'-四甲基-6-羧基-罗丹明(TAMRA)、6'羧基-X-罗丹明(ROX)、HEX、TET、IRD40 和 IRD41。适于本发明的荧光分子还包括: cyamine 染料,包括但不限于 Cy3、Cy3.5 和 Cy5; BODIPY 染料包括但不限于 BODIPY-FL、BODIPY-TR、BODIPY-TMR、BODIPY-630/650、和 BODIPY-650/670;和 ALEXA 染料,包括但不限于 ALEXA-488、ALEXA-532、ALEXA-546、ALEXA-568、和 ALEXA-594;以及本领域技术人员公知的其他荧光染料。适于本发明的富电子指示剂分子包括,但不限于,铁蛋白、血蓝蛋白,和胶体金。备选地,在较不优选的实施方案中,目标多核苷酸可以通过将一个基团特异配位到多核苷酸而标记。第二基团共价连接到指示剂分子并且其具有对第一基团的亲和性,该第二基团可用于间接检测目标多核苷酸。在这种实施方案中,适于用作第一基团的化合物包括,但不限于,生物素和 imminobiotin。适于用作第二基团的化合物包括,但不限于,抗生物

素蛋白和抗生物素蛋白链菌素。

质量控制探针的预定结合序列的结合伴侣可被加到目标分子后接触微阵列。在一个实施方案中，结合伴侣在目标分子处理的过程中加到目标分子。在更优选的实施方案中，在将要接触微阵列前结合伴侣被加到目标分子。

5.5.1.5 杂交到微阵列

如前所述，选择核酸杂交和洗涤条件使得所要分析的多核苷酸分子(或者目标多核苷酸分子)特异结合或特异杂交到阵列的互补多核苷酸序列，优选一个或多个特定阵列位点，其中其互补序列位于该位点。

含有位于其上的双链探针 DNA 的阵列优选地被置于变性条件使得 DNA 单链化之后接触目标多核苷酸分子。含有单链探针 DNA(例如，合成的寡脱氧核酸)的阵列在与目标多核苷酸分子接触前可能需要变性，例如，以除去由于自身互补序列形成的发夹或二聚体。

最优杂交条件将依赖于探针和目标核酸的长度(例如，寡聚体对大于 200 个碱基的多核苷酸)和类型(例如，RNA，或 DNA)。在 Sambrook 等，(如前)，和 Ausubel 等(1987, Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, 纽约)中描述了核酸的特异(即，严格)杂交条件的一般参数。例如，当使用 cDNA 微阵列时，典型的杂交条件是在 5×SSC 加 0.2% SDS 中 65℃杂交 4 小时，然后在低严格洗涤缓冲液(1×SSC+0.2%SDS)中 25℃洗涤，接着在更严格的洗涤缓冲液(0.1×SSC+0.2% SDS)中 25℃10 分钟(Hughes 等，2001, Nature BioTechnology 19: 342-347)。在例如，Tijessen, 1993, 用核酸探针杂交，Elsevier Science Publishers B. V. 和 Kricka, 1992, 非同位素 DNA 探针技术，Academic Press, San Diego, CA 中也提供了有用的杂交条件。

本发明的筛选和/或信号芯片使用的尤其优选的杂交包括在处于或接近探针的平均熔解温度(例如，5℃内，更优选地，2℃内)在 1 M NaCl、50 mM MES 缓冲液(pH 6.5)、0.5%肌氨酸钠和 30%甲酰胺中杂交。

5.5.1.6 信号检测和数据分析

将明白当制备了靶序列，例如，与细胞的 RNA 互补的 cDNA 或 cRNA 并在适宜条件下杂交到微阵列时，杂交到相应于特定基因的阵列中位点的水平将反映含有那个基因的转录产物的 mRNA 或 mRNA 分子在细胞中的优势。例如，当与总细胞 mRNA 互补的可检测标记的(例如，用荧光团)cDNA 杂交到微阵列时，相应于未在细胞中转录的基因的阵列上位点(即，能够特异结合基因表达的一种或多种产物)将几乎没有或者没有信号(例如，荧光信号)，并且所编码的 mRNA 表达转录产物占优势的基因将具有相对强的信号。

在优选的实施方案中，来自两个不同细胞的目标序列，例如，cDNA 分子或 cRNA 分子，被杂交到微阵列的结合位点。对于药物应答，一个细胞样品被暴露于药物而相同类型的另一个细胞样品未暴露于该药物。对于途径应答，一个细胞暴露于途径干扰而相同类型的另一细胞未暴露于途径干扰。来自两种细胞类型的每一种的 cDNA 或 cRNA 被不同地标记从而它们能够被区别。在一个实施方案中，例如，用荧光素标记的 NTP 合成来自药物处理的(或其他干扰的)细胞的 cDNA，用罗丹明-标记的 dNTP 合成来自另一种细胞(未暴露于药物的细胞)的 cDNA。当两种 cDNA 分子被混合并杂交到微阵列时，为阵列上每个位点确定来自每个 cDNA 组的信号的相对强度，并检测特定转录产物丰度的相对差异。

在前面段落中上述实例中，当荧光团被激发时来自药物处理的(或其他干扰的)细胞的 cDNA 将发出绿色荧光，来自未处理细胞的 cDNA 将发出红色荧光。结果，当药物处理对细胞中特定基因的转录没有效果(直接或间接地)时，表达模式在两种细胞中将不可区别，当反转录时，红色标记的和绿色标记的 cDNA 将具有相等优势。当杂交到微阵列时，那种 RNA 的结合位点将发出两种荧光团的特征波长。相反，当暴露于药物的细胞被直接或间接改变细胞中特定基因的转录剪接的药物处理时，通过每个转录结合位点的红色与绿色荧光的比例代表的表达模式将改变。当药物增加 mRNA 的优势时，以 mRNA 表达的每种转录产物片段的比例将增加，而当药物减少 mRNA 的优势时，以该 mRNA 表达的每种外显子的比例将减少。

使用两色荧光标记和检测方案定义基因表达中的变化连同 mRNA 分子的检测已经在例如 Shena 等，1995，用互补 DNA 微阵列定量监控

基因表达模式，Science 270: 467-470 中描述。使用用两种不同荧光团标记的目标序列，例如，cDNA 分子或 cRNA 分子的一个优点是可以进行两种细胞状态中相应于每种排列的基因的 mRNA 表达水平的直接和内部受控的比较。并且由于实验条件(例如，杂交条件)中微小不同造成的变化将不会影响随后的分析。然而，应该认识到还可能使用来自单个细胞的 cDNA，并比较，例如，在例如，药物处理的或其他被干扰的细胞和未处理细胞中特定外显子的绝对量。

在其他优选的实施方案中，使用了单通道检测方法，例如，使用单色荧光标记(见美国专利申请序列号 09/781,814，2001 年 2 月 12 日提交)。在该实施方案中，设计并产生了含有反向互补(RC)探针的阵列。因为 DNA 序列的反向互补具有序列复杂性，该序列复杂性和与目标序列互补的相应前导链(FS)探针关于各种参数(例如，参数如 GC 含量和 GC 趋势在反向互补下是不变的)的复杂性相当，所以用 RC 探针作为对照探针以确定非-特异性交叉杂交到相应 FS 探针的水平。通过比较 FS 探针的原始强度测量和 RC 探针的相应原始强度测量以及各自的测量误差确定目标序列的 FS 探针强度的显著性。在优选的实施方案中，如果 FS 探针和相应的 RC 探针之间的强度差异是显著的，那么称存在转录产物。更优选地，如果 FS 探针强度也显著高于背景水平，那么称存在转录产物。单通道检测方法可以与多色标记一起使用。在一个实施方案中，每个样品用不同颜色标记的许多不同样品被杂交到阵列。使用 FS 和 RC 探针关于每种颜色的差异确定相应样品的杂交水平。

当使用荧光标记的探针时，可以优选地通过扫描共聚焦激光显微镜方法检测转录阵列的每个位点的荧光发射。在一个实施方案中，为所用的两种荧光团的每一种使用适宜的激发线进行分开扫描。备选地，可以使用允许在两种荧光团特异波长同时样本照射的激光并同时分析两种荧光团的发射(见 Shalon 等，1996,Genome Res. 6: 639-645)。在优选的实施方案中，使用带计算机控制的 X-Y 镜台和显微镜物镜的激光荧光扫描仪扫描阵列。用多线的混合气体激光实现两种荧光团的顺次激发，通过波长分开发射的光线并用两个光电倍增管检测。这些荧光激光扫描设备在例如 Schena 等，1996,Genome Res. 6: 639-645 中描述。备选地，可以使用 Ferguson 等，1996，Nature

BioTechnology 14: 1681-1684 描述的纤维光束同时监控许多位点的mRNA 丰度水平。

记录信号并且，在优选的实施方案中，通过计算机，例如使用 12 位或 16 位数字板模拟器分析。在一个实施方案中，扫描的图像使用图像程序(例如，Hijaak Graphics Suite)去斑然后使用图像网格化程序分析，该程序产生每个位点每种波长下的平均杂交的电子数据表。如果必要，可以进行两种荧光染料的通道之间的串话(或重叠)的实验确定的校正。对于转录阵列上的任何特定杂交位点，可以计算两种荧光团的发射比例。该比例独立于同源基因的绝对表达水平，但是对一些基因有用，这些基因的表达被药物施用、基因缺失，或任何其他试验的事件显著调节。

两种细胞或细胞系中 mRNA 的相对丰度优选记分为被干扰的(即，两种来源中所检验的 mRNA 的丰度不同)或未被干扰的(即，相对丰度相同)。如此处所用的，如果 RNA 的两种来源之间的差异至少为约 25%(即，一种来源中 RNA 的丰度比另一种来源的多 25%)，更通常地约 50%，甚至更通常地约 2(即，丰度为 2 倍)，3(丰度为 3 倍)或 5(丰度为 5 倍)，那么该差异优选被记分为干扰。

然而，还有利的是确定两种细胞或两种细胞系中以 mRNA 表达的 mRNA 丰度的相对差异的量值。如上所指出的，这可以通过计算用于差别标记的两种荧光团的发射比例，或者通过本领域技术人员容易明白的类似方法进行。

5.5.2 蛋白质微阵列

在本发明的一个实施方案中，微阵列为蛋白质微阵列。结果，该实施方案中的质量控制探针为多肽或肽。蛋白质质量控制探针优选地具有可以得到的相应结合伴侣，从而该探针与所述结合伴侣的接触可允许特异的和可定量的结合。

在蛋白质微阵列上，能够结合目标蛋白质的蛋白质探针被固定到基质表面，例如，固体支持物(见，例如，Goffeau 等，1996，Science 274: 546-567; Aebersold 等，1999，Nature BioTechnology 10: 994-999; Haab 等，2001,Genome Biology 2: RESEARCH0004. 1-RESEARCH0004. 13; 美国专利号 6,346, 413)。例如，可以使用用于

肽合成的标准固相技术制备多肽探针。如通常所知的，可以使用商业途径可得到的设备和试剂，按照生产商的关于封闭干扰基团、保护要反应的氨基酸、偶联、去保护和未反应残基的帽化的使用说明制备多肽。蛋白质探针可以含有非-肽连接和/或修饰的或非天然发生的氨基酸，例如，D-氨基酸、氨基酸的磷类似物，如α-氨基磷酸和β-氨基磷酸。

用于本发明方法的探针优选地在固体载体或者表面上合成，该固体载体或者表面可以是有孔或无孔的。例如，本发明的探针可以是粘附到硝酸纤维素或尼龙膜或过滤器的多肽序列。备选地，固体载体或表面可以是玻璃或塑料表面。

可以在位置可寻址的阵列上合成蛋白质，许多蛋白质附着到基质上，每种蛋白质位于固体载体上的不同位置。优选地，这许多蛋白质含有在单个生物样品中表达的至少 10、50、100、250、500、1000、1500、2000、2500、3000、4000、5000、6000、7000、8000、9000、10000、50000、或 100000 种不同多肽，加上质量控制探针。使用氨基酸单体在逐步合成中将蛋白质探针合成到基质上。

在一个实施方案中，质量控制探针是抗体或其片段。在另一实施方案中，质量控制探针的结合伴侣是抗体或其片段。在优选的实施方案中，抗体是单克隆抗体或其片段(例如，Fab 片段)(见，例如，Zhu 等，2001, Science 293:2101-2105; MacBeath 等, 2000, Science 289:1760-63; de Wildt 等, 2000, Nature BioTechnology 18: 989-994)。

将明白当蛋白质样品在适宜条件下结合到蛋白质微阵列时，结合到阵列的特定位点的水平将反应样品中相应结合伴侣的优势。微阵列上多肽质量控制探针和其蛋白质结合伴侣之间的水平优选地通过信号化合物指示。例如，当蛋白质样品结合到蛋白质微阵列时，相应于蛋白质探针与不在样品中的相应结合伴侣的阵列上位点将几乎没有或者无信号，多肽探针与样品中占优势的相应结合伴侣将具有相对强的信号。样品中不同蛋白质的相对丰度可以通过阵列上探针的信号强度模式确定。在一个实施方案中，一种或多种信号化合物(例如，荧光染料)直接粘附到质量控制探针的蛋白质结合伴侣。在另一实施方案中，一种或多种信号化合物间接粘附到质量控制探针的蛋白质结合伴侣(例如，通过使用荧光标记的抗体)。

5.6 实现系统和方法

优选使用计算机系统，如本章节描述的计算机系统，根据下面的程序和方法可以实现本发明的分析方法。这种计算机系统还可优选地储存和操作本发明的数据库，该数据库含有关于微阵列上质量控制探针的位置的位置信息概略以及它们在合成的哪个顺次周期中合成(即，微阵列中的垂直放置)并且该数据库可被计算机系统用于实现本发明的分析方法。因此，认为这些计算机系统也是本发明的部分。在特定实施方案中，质量控制位置信息以数字形式储存在数据库中。

在特定实施方案中，计算机系统含有一个或多个处理器和连接到所述一个或多个处理器的一个或多个存储器所述一个或多个存储器含有一个或多个程序，其导致所述一个或多个处理器执行比较本发明的阵列上两种或多种质量控制探针对它们的结合伴侣的结合。结果优选以质量控制探针的结合比率输出。在特定实施方案中，计算机程序导致所述一个或多个处理器执行步骤:

（a） 接收含有所处理微阵列上质量控制探针的结合强度的第一数据结构；和

（b） 将所述第一数据结构与数据库中许多数据结构比较，每个数据结构含有关于与所述微阵列结合的质量控制探针的位置信息，以鉴定所述微阵列上相关位置以比较和评估合成完整性；和

（c） 比较两种或多种质量控制探针的结合。

在特定实施方案中，计算机系统包括导致处理器比较适宜的质量控制探针结合强度并从而确定微阵列是否被正确合成的程序。

在另一实施方案中，计算机系统执行样品质量控制的一个或多个方面。例如，计算机可以直接从扫描的微阵列图像的 TIFF 文件中代表的原始数据读出微阵列质量控制探针强度并比较适宜的强度，和确定阵列的合成是否有缺陷，从而导致可疑数据。如果鉴定了合成缺陷，计算机可以产生非一致性报告并防止将可疑数据自动加到含有微阵列处理数据的数据库中直到质量控制问题被进一步提出。在一个实施方案中，如果质量控制探针的结合比率不在 0.5 和 2.0 之间，那么计算机可以产生非一致性报告。

适于实现本发明的分析方法的一个示例性计算机系统优选地包含

连接到外部元件的内部元件。该计算机系统的内部元件包括与主存储器相互连接的处理器元件。例如，计算机系统可以是基于 Intel Pentium®的处理器，其时钟频率为 200 MHz 或更大，具有 32 MB 或更大的主存储器。在优选的实施方案中，计算机系统为含有一个头“节点”和 8 个兄弟“节点”的许多计算机群，每个节点具有一个中央处理器(CPU)。此外，该机群还包括头节点上的至少 128 MB 随机存取存储器(RAM)和 8 个兄弟节点的每一个上的至少 256 MB RAM。因此，本发明的计算机系统不限于由单个存储器或单个处理器组成的那些系统。

外部元件可以包括大容量存储器。该大容量存储器可以是一个或多个硬盘，它们通常与处理器和存储器封装在一起。这种硬盘通常为 1GB 或更大存储容量并且更优选地具有至少 6GB 存储容量。例如，在上述优选的实施方案中，其中本发明的计算机系统包括几个节点，每个节点可以有其自己的硬盘驱动器。头节点优选地具有至少 6GB 存储容量的硬盘驱动器，而每个兄弟节点优选有至少 9GB 存储容量的硬盘驱动器。本发明的计算机系统可以还含有其他大容量存储器包括，例如，一个或多个软盘驱动器，一个或多个 CD-ROM 驱动器、一个或多个 DVD 驱动器或一个或多个 DAT 驱动器。

其他外部元件通常包括一个用户界面设备，其最典型地为监视器和键盘以及图像输入设备如“鼠标”。计算机系统还通常连接到网络连接，其可以是例如，局域网(LAN)或其他的局部计算机系统的一部分和/或广域网(WAN)，如 Internet 的一部分，该广域网连接到其他的远程计算机系统。例如，在上面讨论的优选的实施方案中，其中计算机系统含有许多接点，每个接点优选连接到网络，优选 NFS 网络，从而计算机系统的接点通过网络相互通信，任选地与其他计算机系统通信，并从而相互共享数据和处理任务。

在这种计算机系统运行中加载到存储器的是一些软件组件。软件组件包括本领域中的标准软件组件和本发明特有的组件。这些软件组件典型地保存在大容量存储器如硬盘驱动器上，但是也可以保存在其他计算机可读介质上，包括，例如，一个或多个软盘、一个或多个 CD-ROM、一个或多个 DVD 或一个或多个 DAT。软件组件代表负责管理计算机系统和其网络相互连接的操纵系统。操纵系统可以是，例如，

Microsoft Windows 家族的操纵系统，如 Windows 95、Windows 98、Windows NT 或 Windows2000。备选地，操作软件可以是 Macintosh 操作系统、UNIX 操作系统或 LINUX 操作系统。软件组件含有优选存在于系统中以帮助程序实现本发明特定方法的普通语言和功能。可用于对本发明的分析方法编程的语言包括，例如，C 和 C++、FORTRAN、 PERL、HTML、JAVA，和 UNIX 或 LINUX shell 命令语言的任一种，如 C shell 脚本语言。本发明的方法也可以在数学软件包中编程或建模，这些数学软件包允许方程的符号性输入和处理的高级说明，包括所要用的特定算法，从而不需要用户程序性地编程单个方程和算法。这些软件包包括，例如，Mathworks(Natick, MA)的 Matlab、Wolfram Research (Champaign, IL)的 Mathematica 或 MathSoft (Seattle, WA)的 S-Plus。

软件组件包含本发明的分析方法，其优选地以程序化语言或符号包编程。例如，软件组分优选地包括导致处理器实现接受每个微阵列上每个质量控制探针的许多位置数据并将数据保存在存储器中的步骤。例如，计算机系统可以接受用户手工输入的数据(例如，通过用户界面)。然而，备选地，程序导致计算机系统从数据库得到质量控制探针信息。这种数据库可以保存在大容量存储器(例如，硬盘驱动器)或其他计算机可读介质上并加载到计算机存储器中，或者数据库可以被计算机系统通过网络存取。

在一个实施方案中，计算机可读介质含有编码的数据结构，其含有:

(a) 微阵列上质量控制探针的位置的数字表示; 和

(b) 每个质量控制探针在哪一个合成周期合成的数字表示。

在另一实施方案中，可以处理具有故意缺陷的对照微阵列并保存信号强度模式和比例。本发明还包括一种方法，通过该方法将来自样品微阵列的信号强度和/或所得比例与含有已知误差的概要的数据库比较。如果在数据库中发现匹配，那么可以确定样品微阵列中的缺陷。

除了此处描述的示例性程序结构和计算机系统，其他的备选的程序结构和计算机系统对于技术人员将是显而易见的。这些备选系统无论在精神上还是在范围中都不背离上述计算机系统和程序结构，因此这些备选系统旨在包括在所附权利要求书中。

下面的实施例为了阐明本发明而给出，并不旨在限制本发明。

6.实施例 1：使用质量控制探针的质量控制

6.1 合成错误的阐明

喷墨打字机使用两个喷墨头将 phosphoramidites 或活化剂分布到阵列的玻璃基质上。每个喷墨头含有三组每组 20 个喷嘴，每 20 个喷嘴组用于沉积一种 phosphoramidite 或活化剂。一组中 20 个喷嘴以两个交错的列排列，每列 10 个(见图 1)。该模式允许喷墨头每经过一次沉积 20 行碱基，每个喷嘴负责一行。因为每个喷嘴负责特定行，任何堵塞或其他喷嘴故障都可导致所有或部分行以 20 行周期性地被除去或者无效率地合成(通过受影响的质量控制探针中强度的减少来检测)。图 1 显示了用三个被堵塞喷嘴(即喷嘴 4、15 和 20)合成的 25,000 个寡核苷酸阵列。

6.2 合成故障检测

在图 2A 中示意性描述了被堵塞的质量控制探针。直接在微阵列上(从而序列在合成周期 1-25 产生)或间隔区上(通过虚线描述)合成 25 个核苷酸预定结合序列(通过实线描绘)。间隔区长为 20 个核苷酸(从而序列在合成周期 21-45 产生)或者 35 个核苷酸(从而序列在合成周期 36-60 产生)。如果在寡核苷酸微阵列合成过程中没有合成缺陷，那么预定结合序列的反向互补序列将同样好地杂交到所有质量控制探针上的预定结合序列并给出可比信号。

图 2B 示意性描绘了寡核苷酸微阵列的合成周期 24 中的合成缺陷(通过带纹条描绘)。因为当预定结合序列不在间隔区上或者在 20 个核苷酸的间隔区上时，这影响预定结合序列的序列，当与没有观察到合成错误的结合水平相比较时，对预定结合序列的反向互补序列的杂交将减弱。35 个核苷酸间隔区上的预定结合序列不受影响；然而，其应该与没有合成错误存在时相同程度地杂交到其反向互补序列。

在没有间隔区(总长度 25 个核苷酸)、有 20 个寡核苷酸间隔区(总长度 45 个核苷酸)或 35 个寡核苷酸间隔区(总长度 60 个核苷酸)的喷墨寡核苷酸阵列上合成具有 SEQ ID NO:1 的序列的质量控制探针。

5′ATCATCGTAGCTGGTCAGTGTATCC 3′ (SEQ ID NO : 1)

SEQ ID NO:1 的荧光标记的反向互补序列用于杂交到寡核苷酸微阵列。当寡核苷酸微阵列合成过程中没有合成缺陷时，所有质量控制探针同样好地杂交到它们的反向互补序列(图 4)。这通过微阵列处理后对荧光标记的反向互补核苷酸的杂交的类似水平显示(见图 4A-4B)。在两个微阵列上一式两份地进行每个质量控制探针的数据定量荧光强度并在表 1 中给出。25 聚体的平均荧光强度与 45 聚体或 60 聚体的平均荧光强度的比例接近 1 并指出所有的都同等地结合到它们的反向互补序列。

进行在微阵列合成中各种合成周期有缺陷的类似实验，目的是确定质量控制探针的敏感性。当合成中第一(图 5)或第一和第二(图 6)个合成周期被跳过时，仅仅 25 聚体对其互补的荧光标记的寡核苷酸的杂交受到影响(图 5A-5B 和 6A-6B)。表 1 中两种比例相对于没有合成周期被跳过时所见到的比例都减少了。当第 36(图 7)或第 34 和 35(图 8)个合成周期被跳过时，45 聚体和 60 聚体对它们的互补荧光标记的寡核苷酸的杂交都受到影响(图 7A-7B 和 8A-8B)。表 1 中两种比例相对于没有合成周期被跳过时所见到的比例都增加了。当第一个 22 个合成周期中有无效合成时(图 9)，仅仅 25 聚体对其互补的荧光标记的寡核苷酸的杂交被严重影响(图 9A-9B)。表 1 中两种比例相对于没有合成周期被跳过或者无效时所见到的比例都减少了。

表 1

受影响的合成周期		阵列 1	阵列 2	平均	25 聚体/45 聚体的比例	25 聚体/60 聚体的比例
无	25 聚体	0.0628	0.0495	0.0562	1.28	1.03
	45 聚体	0.0413	0.0399	0.0406		
	60 聚体	0.535	0.0555	0.0545		
1	25 聚体	0.0133	0.0149	0.0141	0.28	0.20
	45 聚体	0.0461	0.0536	0.0499		
	60 聚体	0.0656	0.0770	0.0713		
1-2	25 聚体	0.0056	0.0044	0.005	0.10	0.07
	45 聚体	0.0532	0.0442	0.0476		
	60 聚体	0.0793	0.0675	0.0734		
36	25 聚体	0.0692	0.0730	0.0711	5.47	2.30
	45 聚体	0.0120	0.0140	0.013		
	60 聚体	0.0278	0.0339	0.0309		
34-35	25 聚体	0.1028	0.0644	0.0836	42.0	1.87
	45 聚体	0.0020	0.0019	0.0195		
	60 聚体	0.0471	0.0427	0.0449		
1-22	25 聚体	0.0024	0.0020	0.0022	0.008	0.008
	45 聚体	0.2165	0.3682	0.2924		
	60 聚体	0.2640	0.2962	0.2801		

实施例 2: 使用没有间隔区的质量控制探针的质量控制

7.1 合成故障检测

在图 3A 中示意性描述了延迟开始的质量控制探针。直接在微阵列上合成一系列 25 个核苷酸预定结合序列(通过粗线描绘)，在每个合成周期开始单个探针的合成(从合成周期 1-36)。不像上面的策略，不使用间隔区从而所有质量控制探针仅由长为 25 个核苷酸的预定结合序列组成。质量控制探针间的仅有差异是合成开始的周期(粗线描

绘了质量控制探针，细线描绘了没有单体沉积的合成周期)。组成每个质量控制探针的合成周期在图 3A 中每个探针上列出。如果在寡核苷酸微阵列合成过程中没有合成缺陷，那么探针序列的反向互补序列将同等好地杂交到所有质量控制探针并给出相当的信号。

图 3B 示意性描绘了寡核苷酸微阵列的合成周期 29 中的合成缺陷(通过灰色条给出)。因为这影响将合成周期 29 作为它们的序列(即，在合成周期 5-29 开始的那些质量控制探针)的所有预定结合序列。这些质量控制探针中反向互补序列的杂交与没有观察到合成错误的结合水平相比将降低。然而，在合成周期 29 中不含有沉积单体的质量控制探针(即在周期 1-4 或 30-35 开始合成的那些质量控制探针)不受影响，这样它们应该与没有合成错误存在时同样程度地杂交到它们的反向互补序列。

使用延迟开始在喷墨寡核苷酸微阵列上合成具有 SEQ ID NO:1 的序列的质量控制探针。在微阵列合成过程中质量控制探针在 1 到 35 的每个渐进合成周期开始。SEQ ID NO:1 的荧光标记的反向互补序列用于杂交到寡核苷酸微阵列。

当寡核苷酸微阵列合成中第一和第二合成周期中有无效合成时，仅仅头两个延迟开始的质量控制探针受影响(图 10)。绘制每个合成周期的质量控制探针的平均荧光强度并且仅仅在微阵列的第一和/或第二合成周期含有它们的质量控制序列部分的探针处显示出强度的降低(图 10B)。在第二个合成周期随后开始合成的所有质量控制探针都未受影响并且同等好的杂交到它们的反向互补序列。在寡核苷酸微阵列合成中，当头 5 个合成周期(图 11)、头 8 个合成周期(图 12)，或最后 15 个合成周期(图 13)中有无效合成时看到类似结果。在每种情况中，仅仅对于在微阵列的受影响的合成周期沉积的单体构成其序列的部分的质量控制探针荧光强度减弱。

8. 实施例 3: 质量控制探针的增强的敏感性

8.1 使用缺失

寡核苷酸微阵列合成过程中的合成故障导致一个或多个合成周期受损，该合成故障降低了质量控制探针对其荧光标记的反向互补寡核苷酸的结合程度(见例如，上面的段落 6.2 和 7.1)。然而，对于仅仅

少数(例如，一个或两个)合成周期受损从而质量控制探针现在比全长稍短(即，相对于全长 25 聚体为 24 聚体或 23 聚体)的情况，与其反向互补寡核苷酸的结合仍然相对牢固。为了增加合成故障检测的敏感性，在本发明的方法中使用具有已经含有单个缺失的预定结合序列的质量控制探针。这些质量控制探针具有合成的预定结合序列，该序列具有 SED ID NO:1 的第 19 个残基(从 5'末端)的缺失。微阵列合成中由于故障的任何额外缺失将加剧该缺陷并导致更加不能结合全长 25 聚体序列的反向互补序列。图 14 显示具有第 34 和 35 合成周期中合成缺陷的微阵列上单个缺失质量控制探针比没有缺失的质量控制探针更敏感。

5′ATCATCGTAGCTGGTCAGGTATCC 3′ (SEQ ID NO : 2)

全长 25 个核苷酸预定结合序列的标记的反向互补序列用于与每个微阵列上质量控制探针杂交。为每个微阵列确定每个合成周期质量控制探针的平均荧光强度图。全长质量控制探针显示出从第 15 个合成周期开始的合成缺陷(图 14A)。当缺失质量控制探针显示出在第 11 个合成周期开始的合成错误(图 14B)。从而单缺陷质量控制探针是微阵列质量的更敏感测量工具。

8.2 相关性图的比较

这些实验表明使用含有一个或多个缺陷的微阵列可以提供表面上看起来可接受的数据。然而，当该数据与没有缺陷的微阵列比较时，问题变得明显起来。相关性图通过检查实验的可复现性(例如，使用荧光染料-反向对分析)来评估数据的质量。为在头 22 个合成周期(图 15A)具有缺陷的微阵列和没有合成缺陷(图 15B)的微阵列绘制荧光染料反向对之间的相关性。寡核苷酸用红色或绿色荧光染料标记并将混合物用于杂交每个微阵列。将红色对绿色荧光信号的比例的 log10 对复制芯片的红色对绿色荧光信号的比例的 log10 作图。当来自头 22 个合成周期被跳过的微阵列的数据与其自己比较时，没有检测到缺陷(图 15A)。类似地，当来自无缺陷微阵列的数据与其自身比较时没有检测到缺陷(图 15B)。然而，当来自头 22 个合成周期被跳过的微阵列的数据与来自无缺陷微阵列的数据比较时，存在差异(图 15C)。甚至有缺陷的微阵列将导致数据。因为事先不知道该数据将看起来像什

么，来自缺陷阵列的数据将最初看起来是可接受的。根据本发明的质量控制探针的使用防止使用差质量数据。

用合成中第一(图 16A)、第一和第二(图 16B)、第 36(图 16C)或第 34 和 35(图 16D)个合成周期被跳过的寡核苷酸微阵列进行类似实验。来自杂交到有缺陷微阵列的寡核苷酸的数据对来自无缺陷微阵列的数据作图。这些图再次都看起来类似并且本应该不会检测到合成缺陷。这证明用相关性图进行的微阵列分析对于鉴定有缺陷微阵列不够敏感。

9. 引用的参考文献

所引用的所有参考文献在此处被完整并入作为参考并且相同程度地为了所有目的就像每个单独的出版物或专利或专利申请被特别地和个别地指出以为了所有的目的完整地被并入作为参考一样。

可以进行本发明的许多修饰和改变而不背离其精神和范围，如将对于本领域技术人员显而易见的。此处描述的特定实施方案将仅仅被所附权利要求书以及这些权利要求有资格的等价权利要求的全部范围所限制。

<110> Rosetta Inpharmatics LLC
Marton, Matthew
Meyer, Michael
Jones, Allan

<120> 评估微阵列质量的方法

<130> 9301-159-228

<140>
<141>

<150> 60/392,629,
<151> 2002-06-28

<160> 2

<210> 1
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：引物

<400> 1
atcatcgtag ctggtcagtg tatcc 25

<210> 2
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：引物

<400> 2
atcatcgtag ctggtcaggt atcc 24

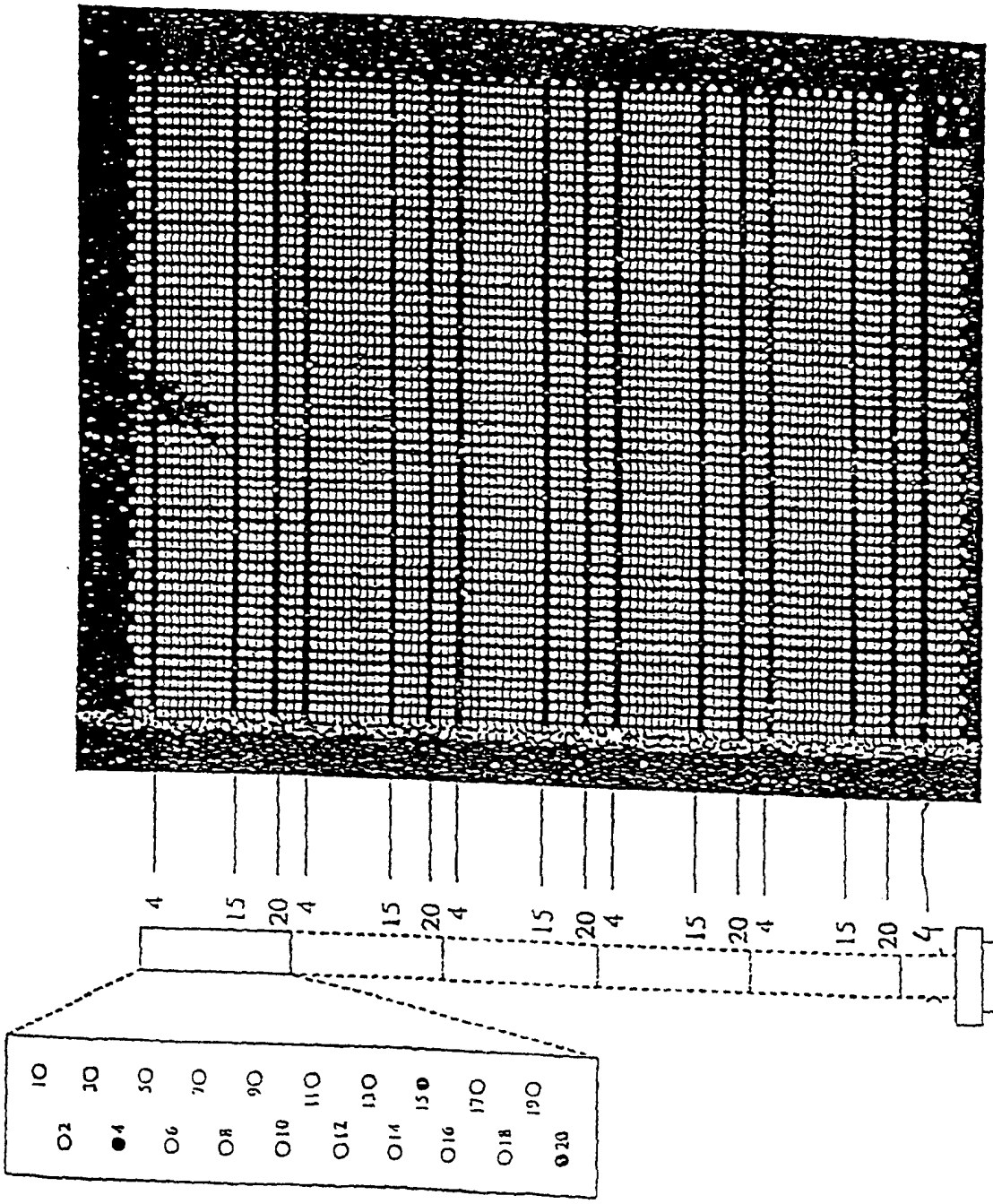

4
15
20
4
15
20
4
15
20
4
15
20
4
15
20
4
1O
3O
5O
7O
9O
11O
13O
15O
17O
19O
O2
O4
O6
O8
O10
O12
O14
O16
O18
O20

图 1

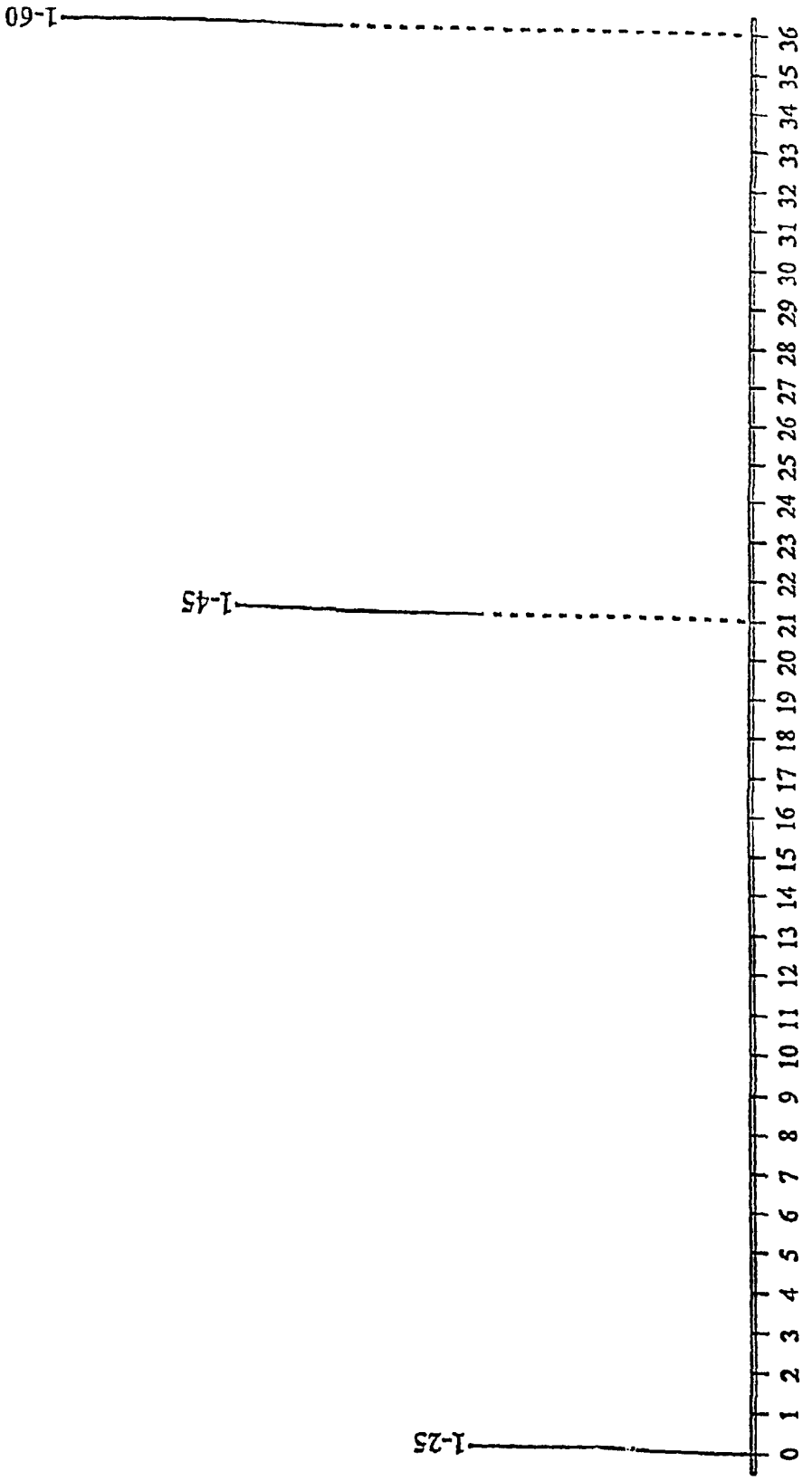

1-60
1-45
1-25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

图 2A

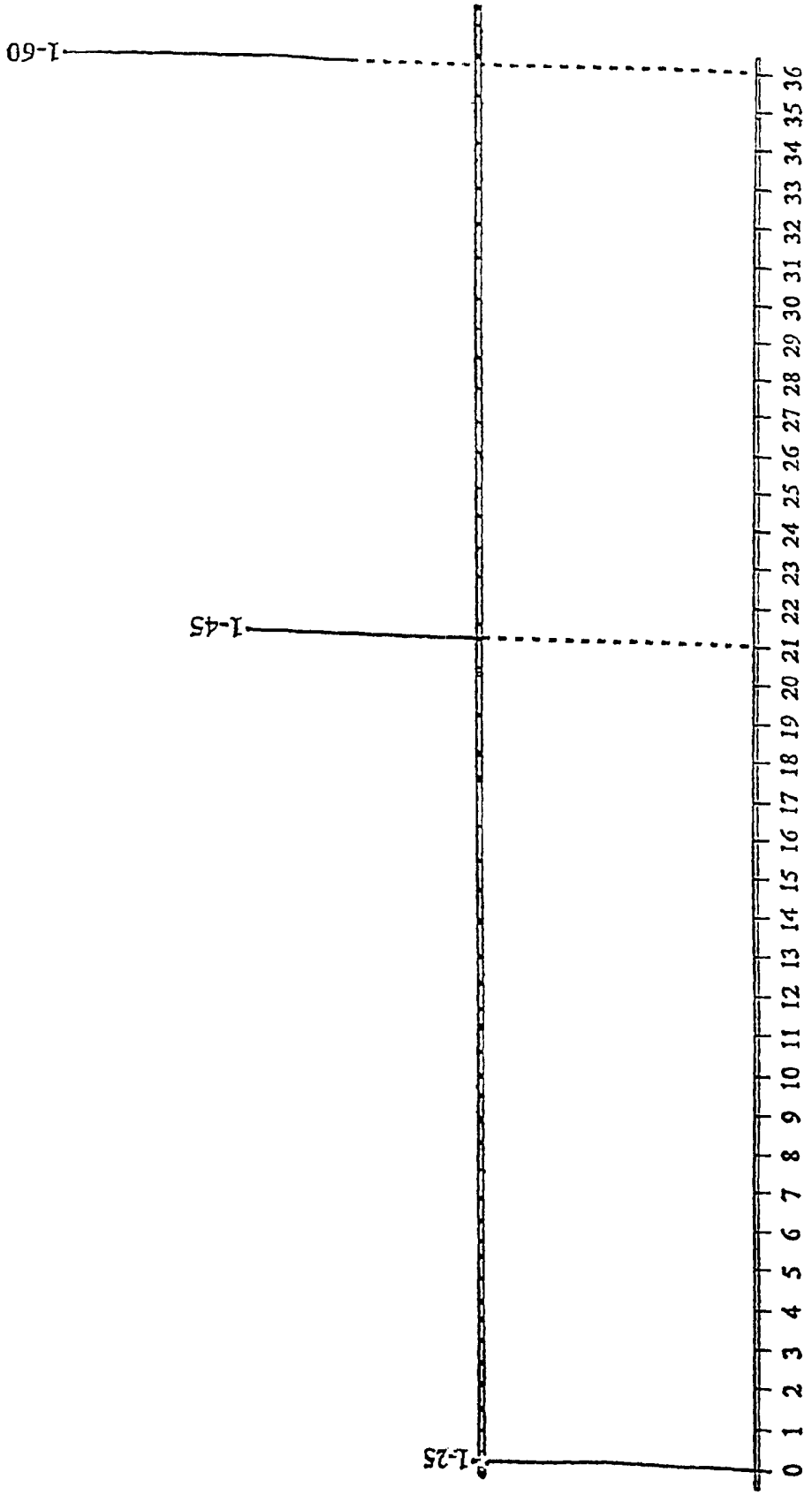

1-60
1-45
1-25
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

图 2B

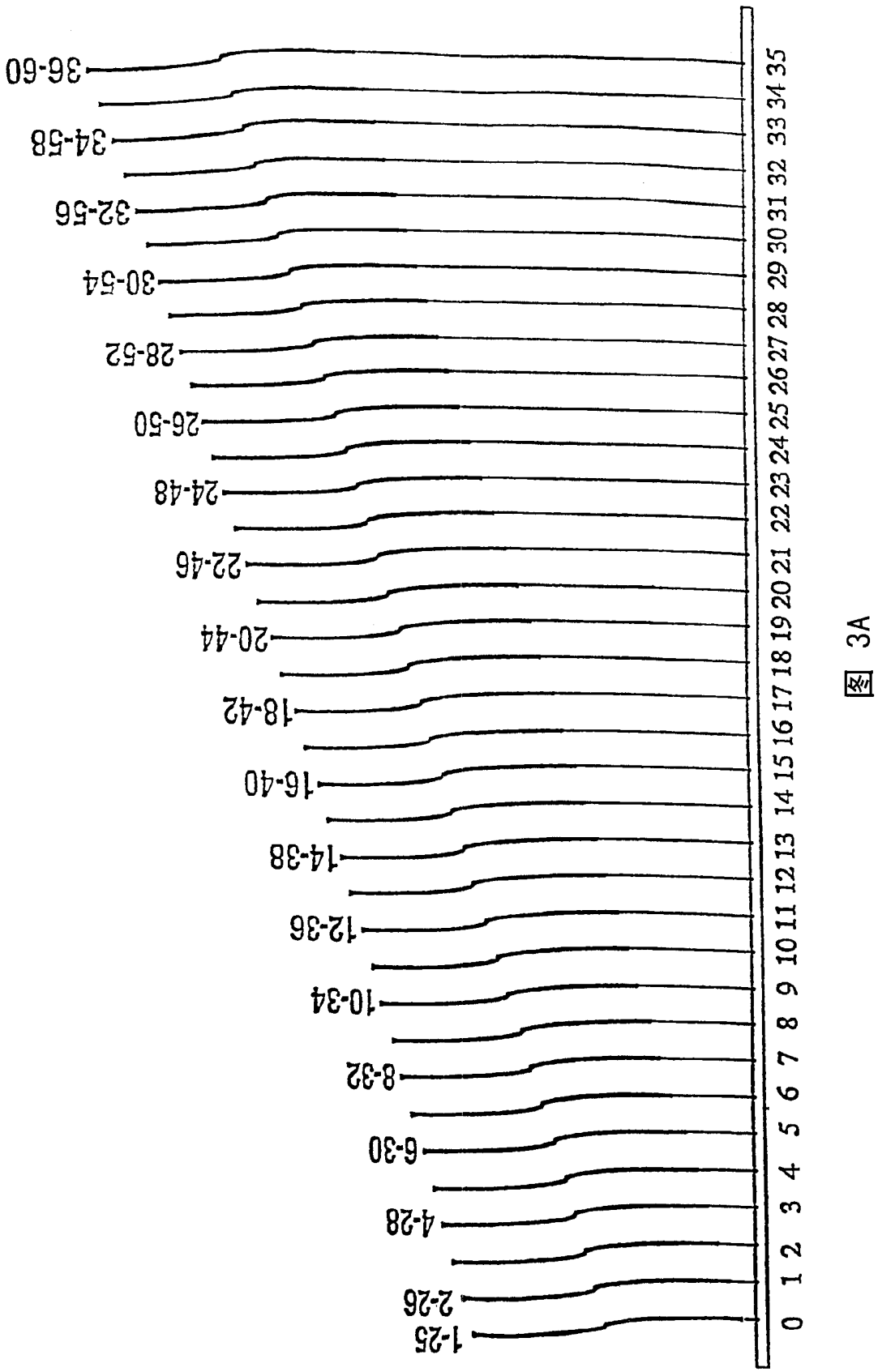

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
1-25
2-26
4-28
6-30
8-32
10-34
12-36
14-38
16-40
18-42
20-44
22-46
24-48
26-50
28-52
30-54
32-56
34-58
36-60

图 3A

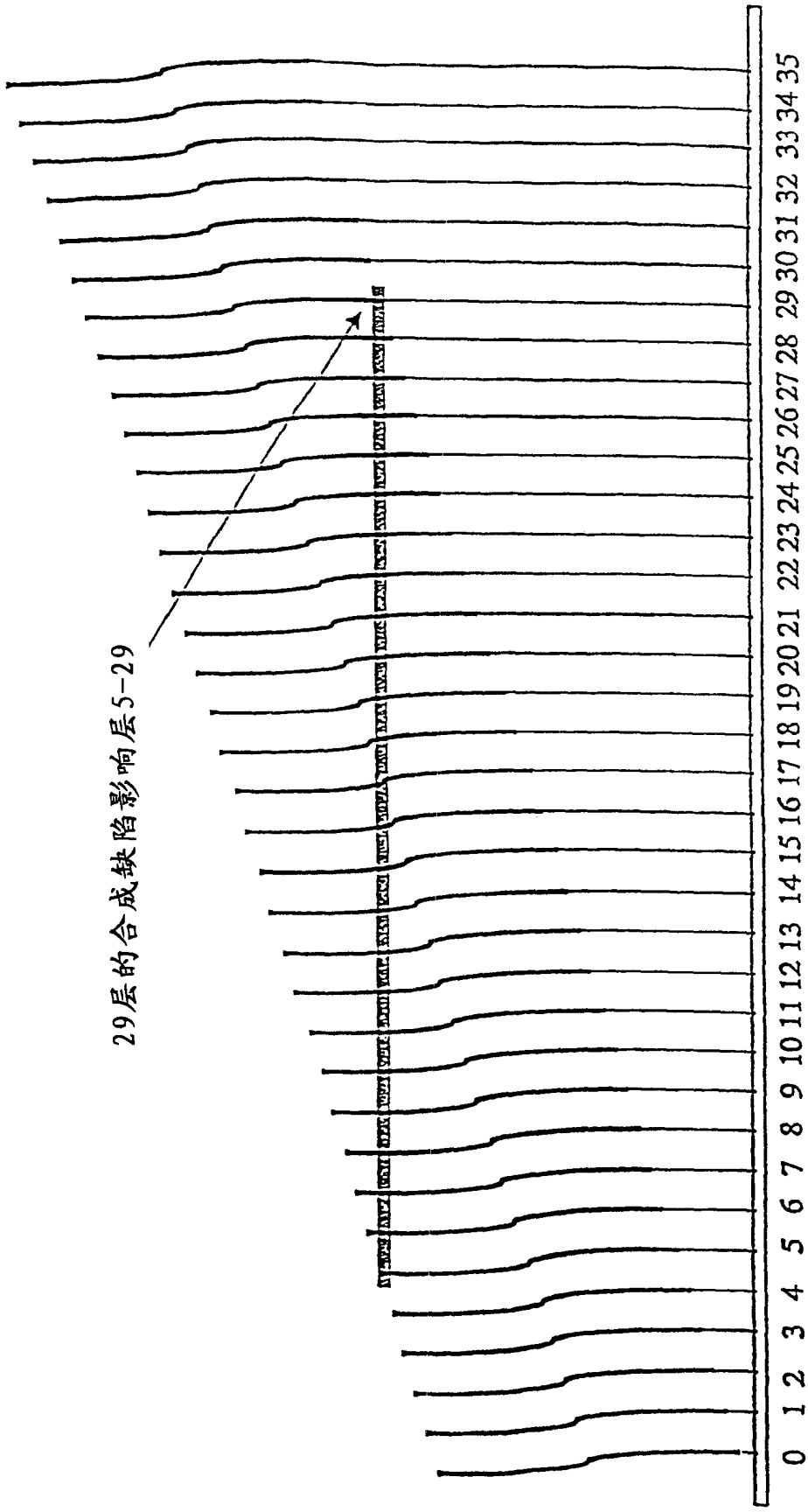

29层的合成缺陷影响层5-29
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

图 3B

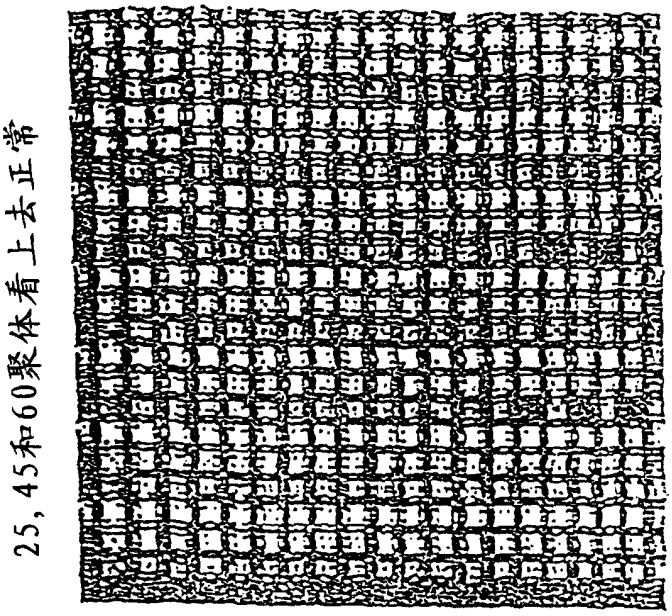


图 4B

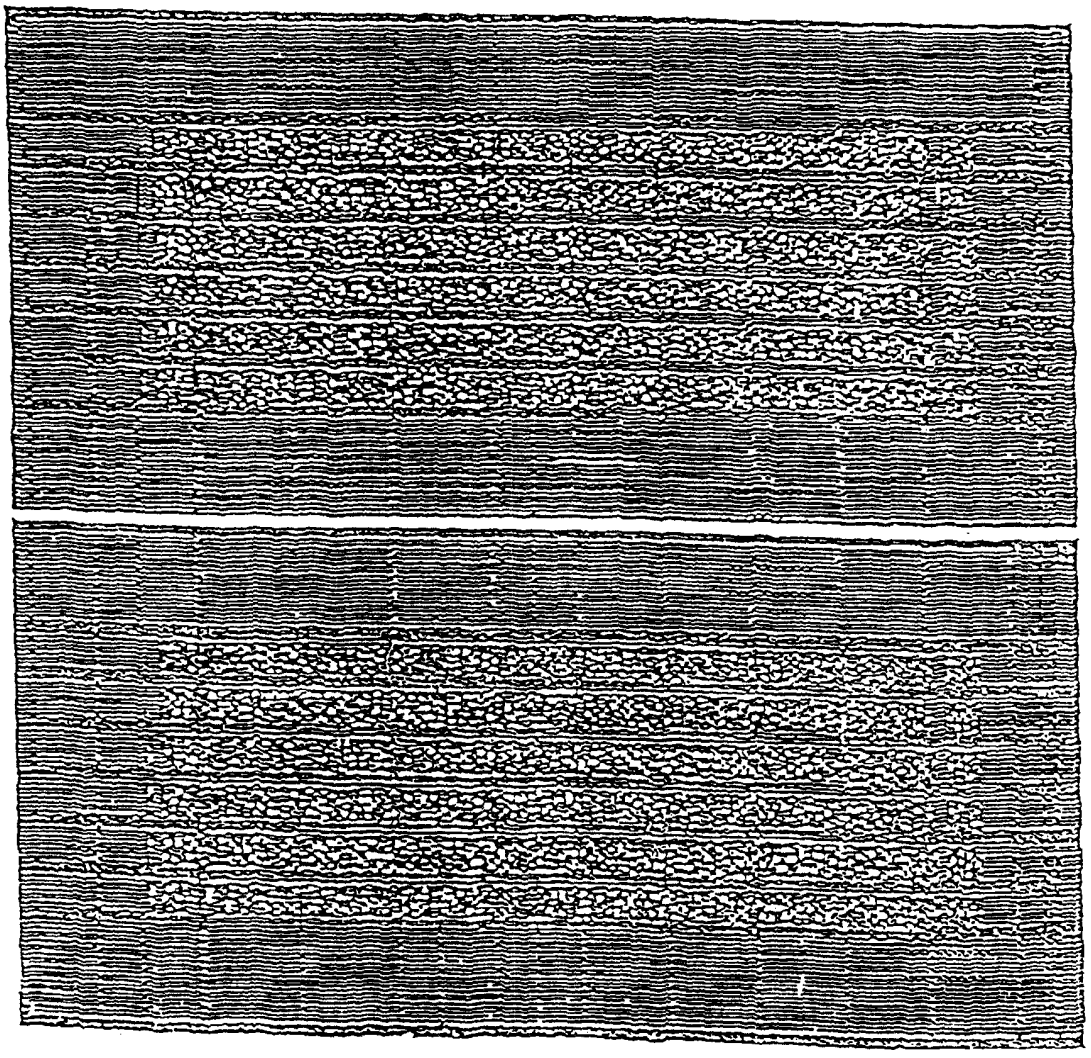


图 4A

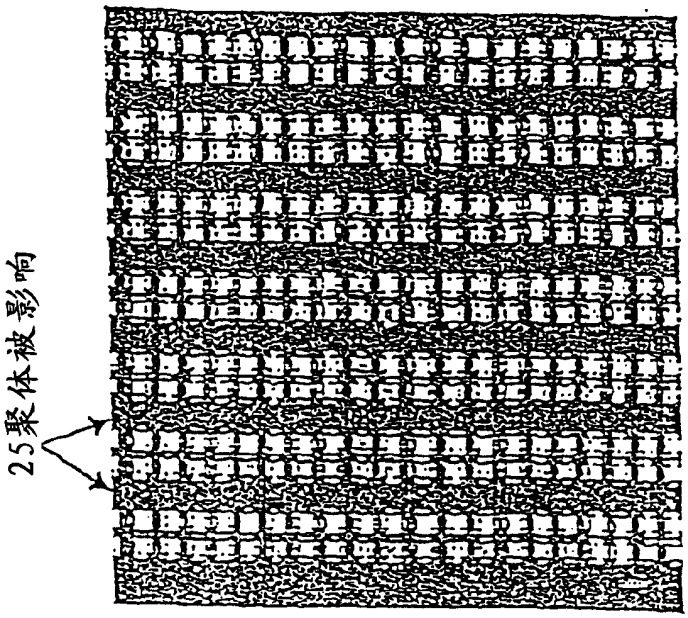


图 5B

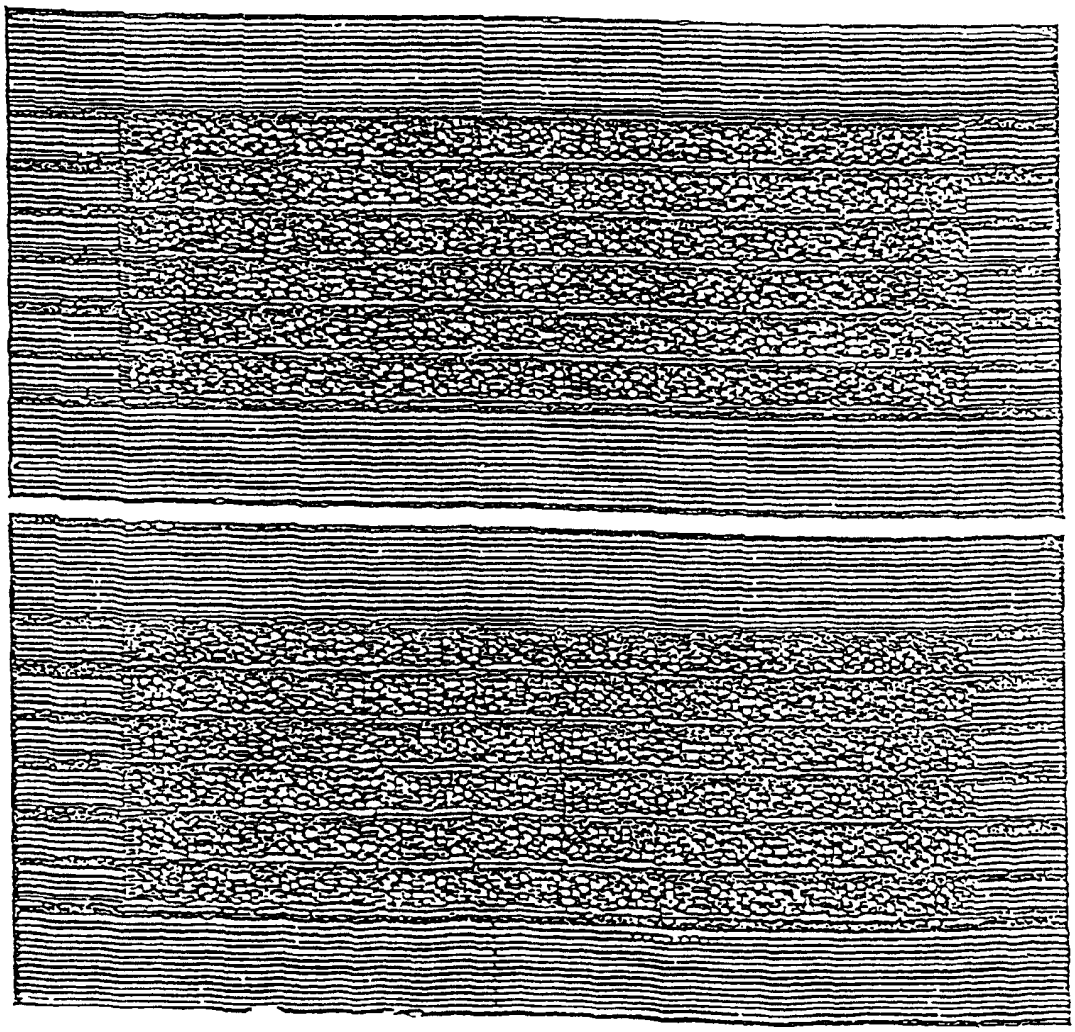


图 5A

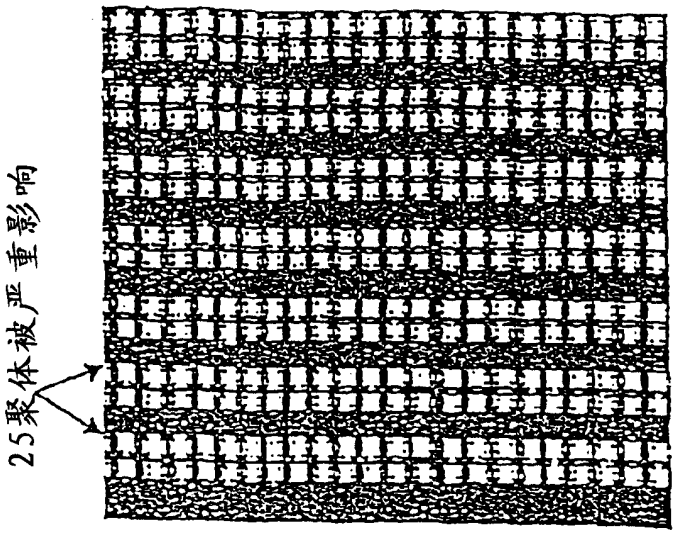


图 6B

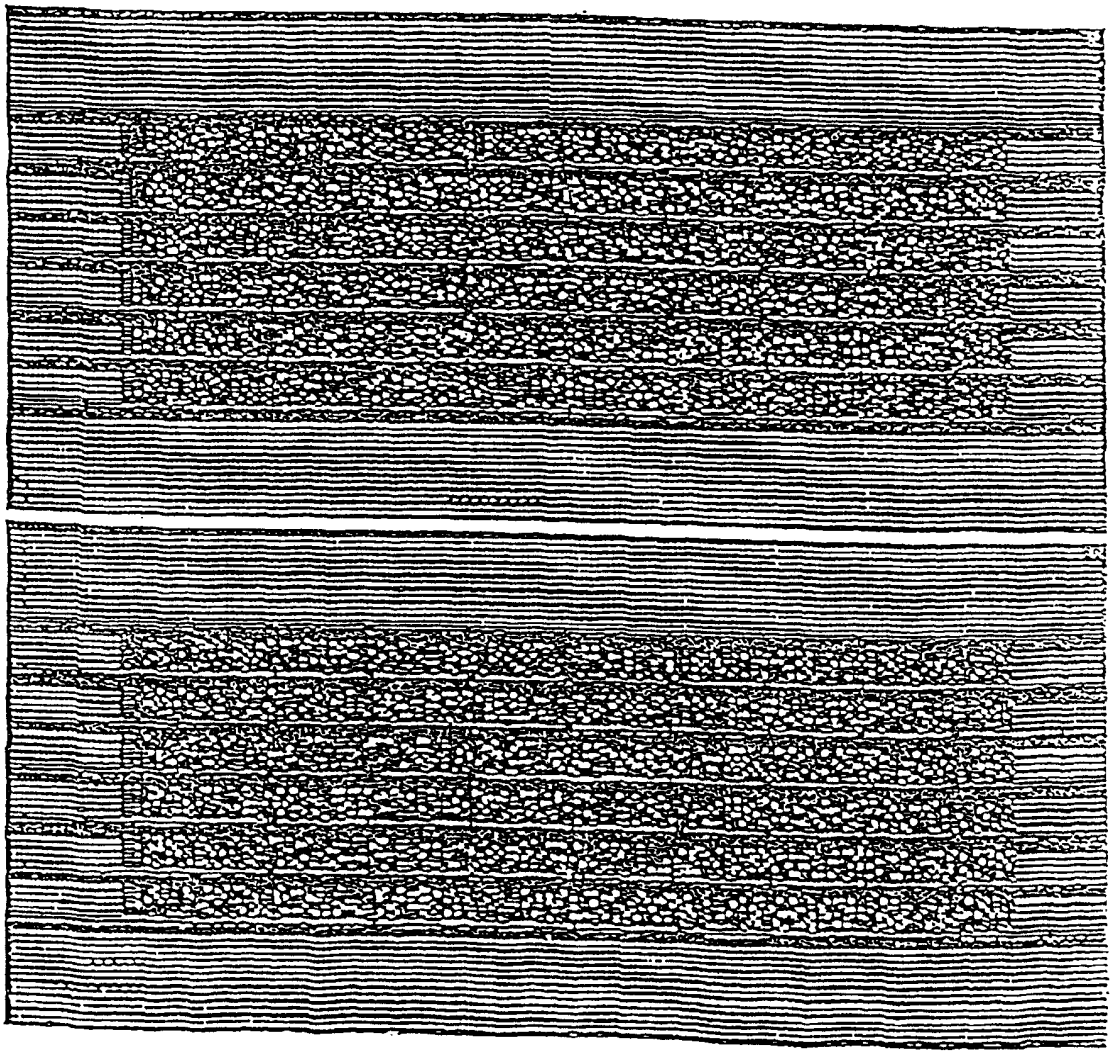


图 6A

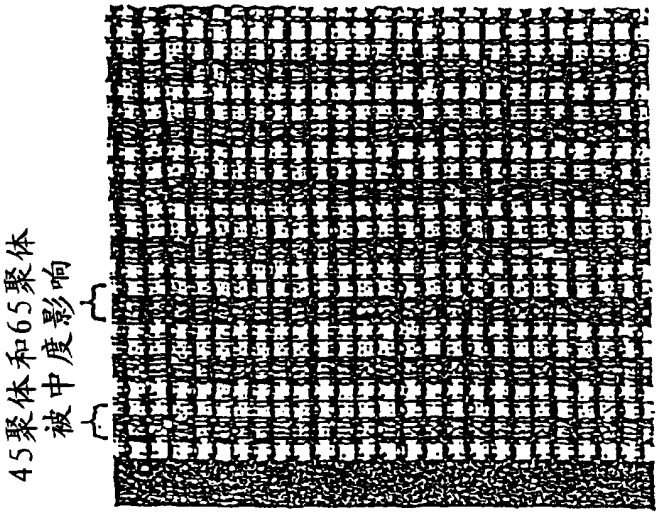


图 7B

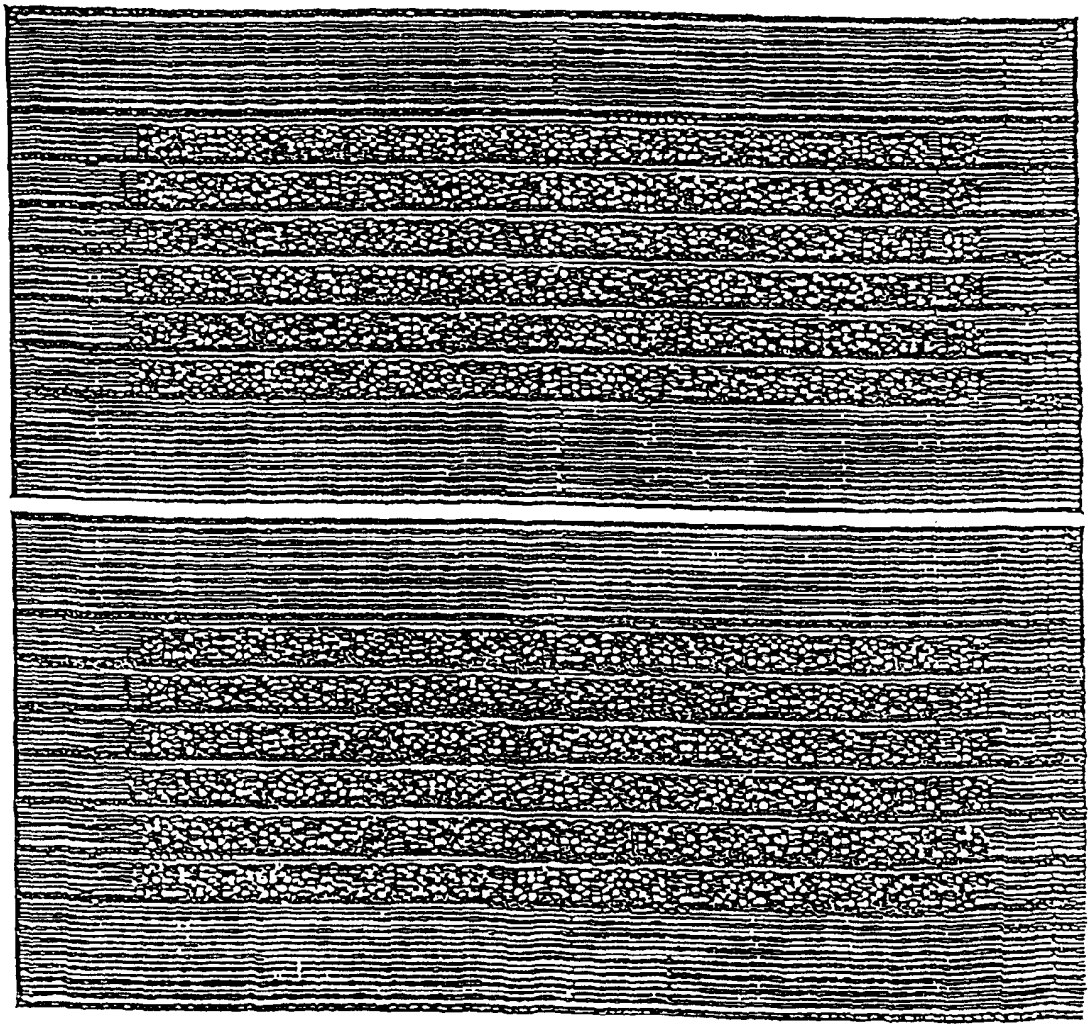


图 7A

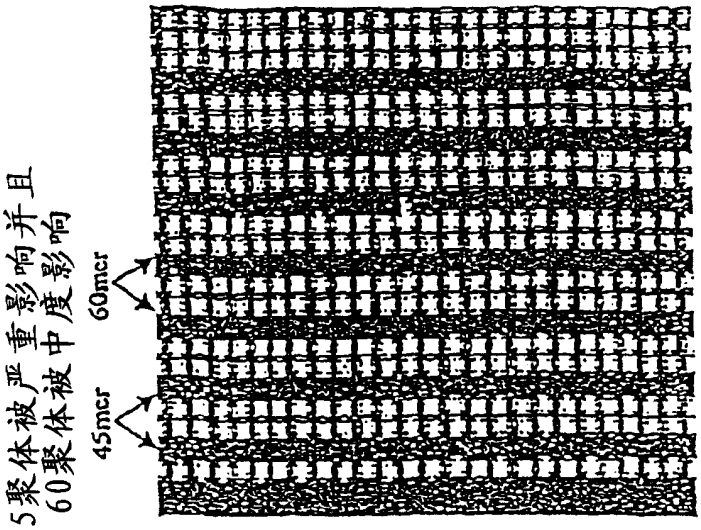


图 8B

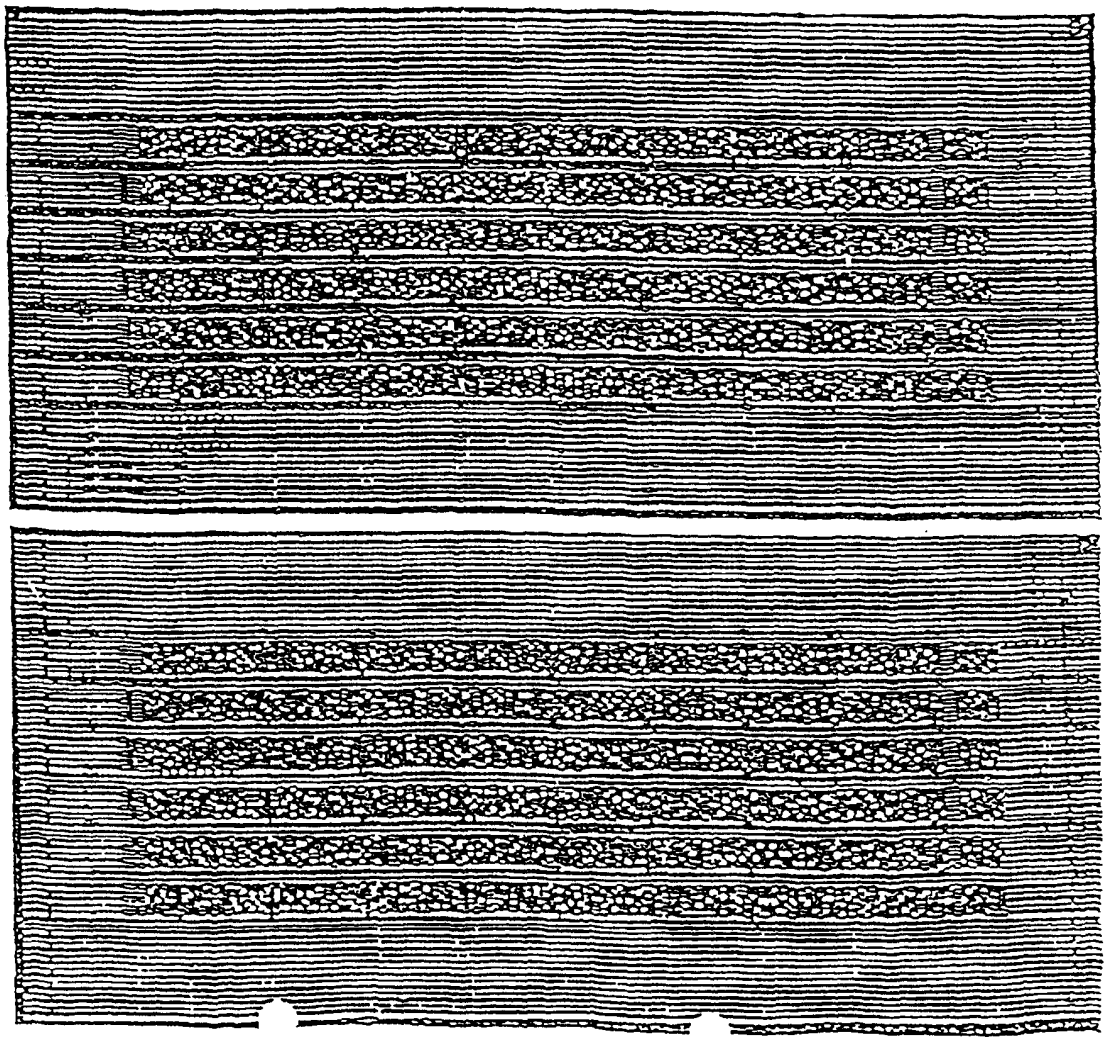


图 8A

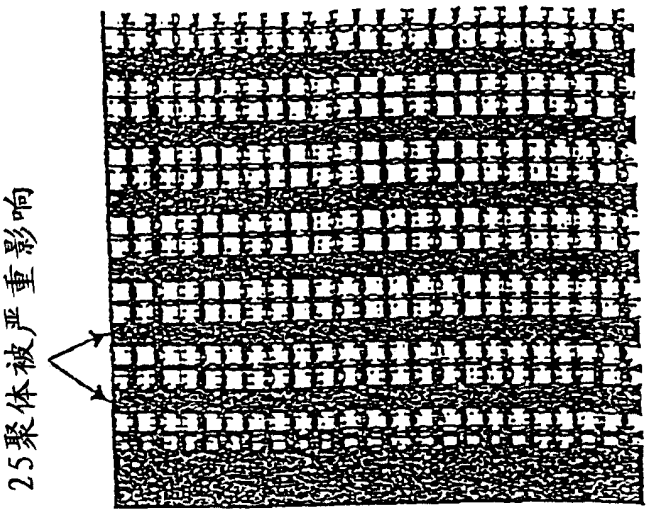


图 9B

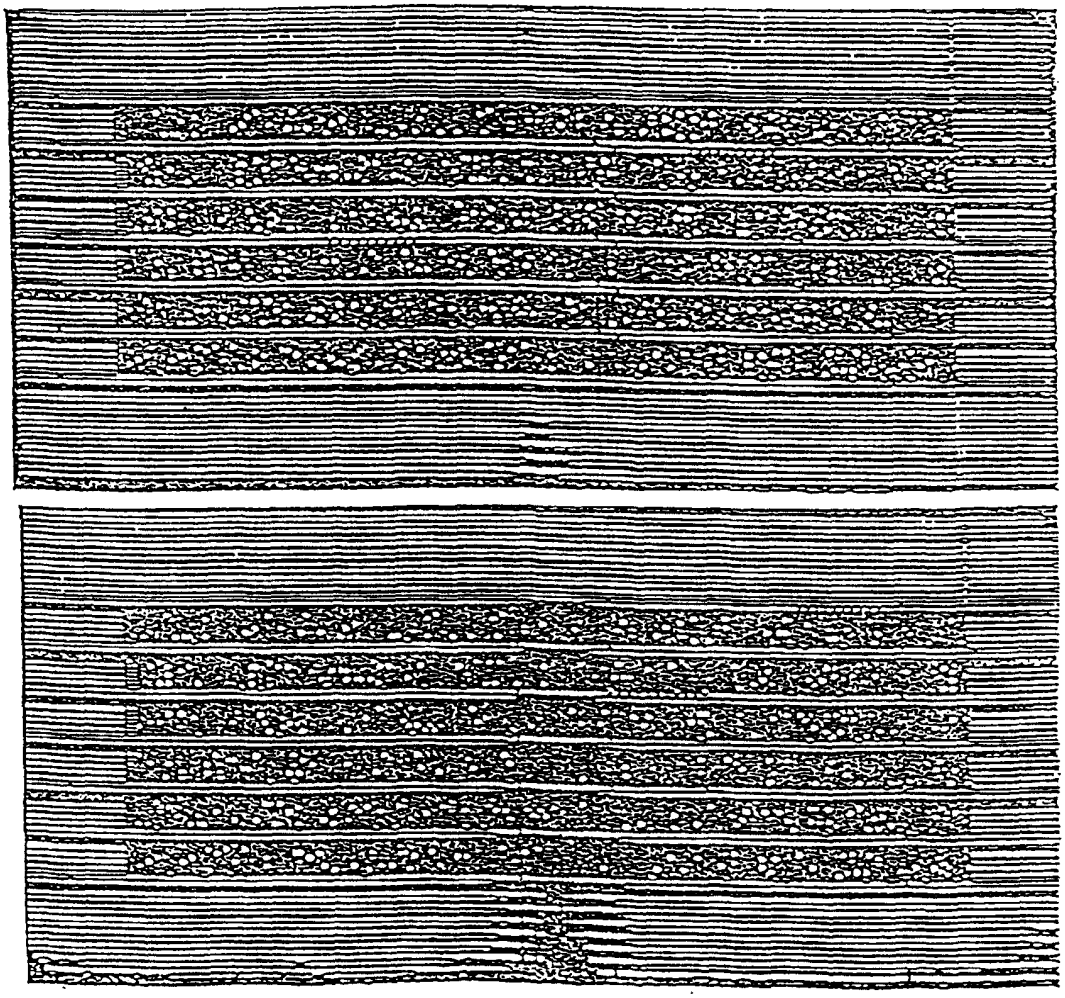


图 9A

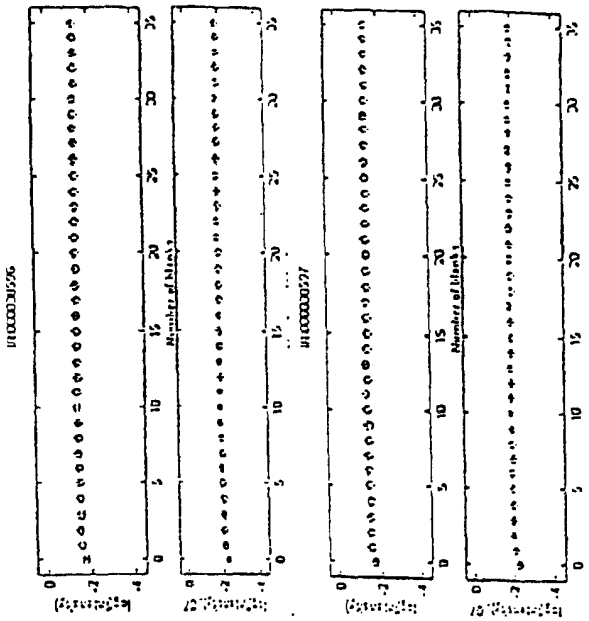


图 10B

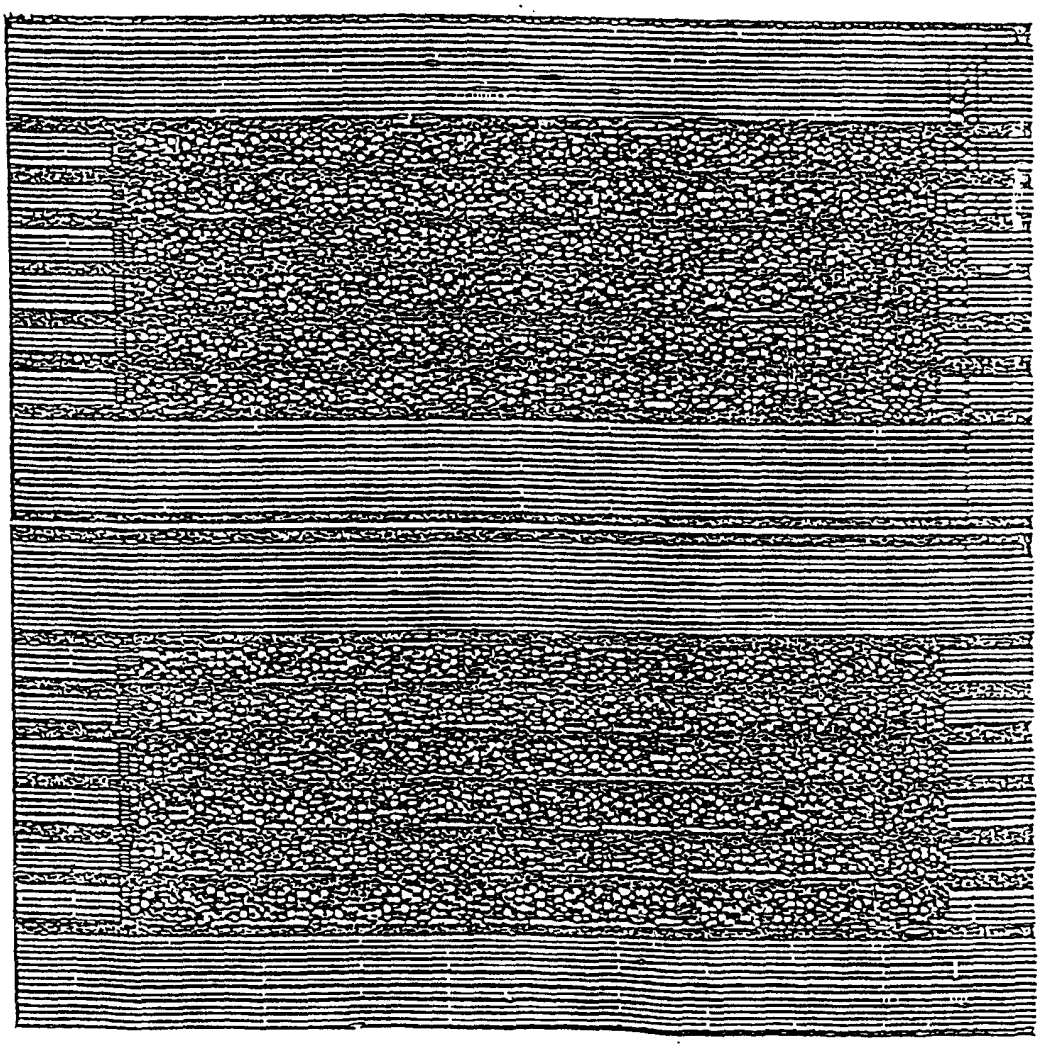


图 10A

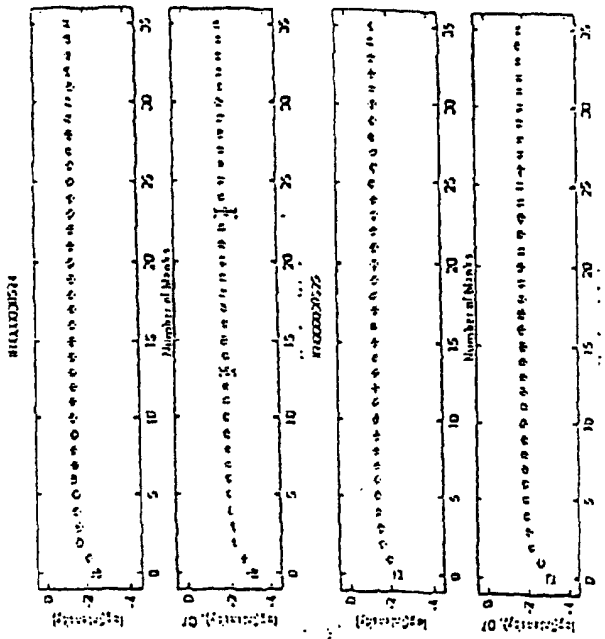


图 11B

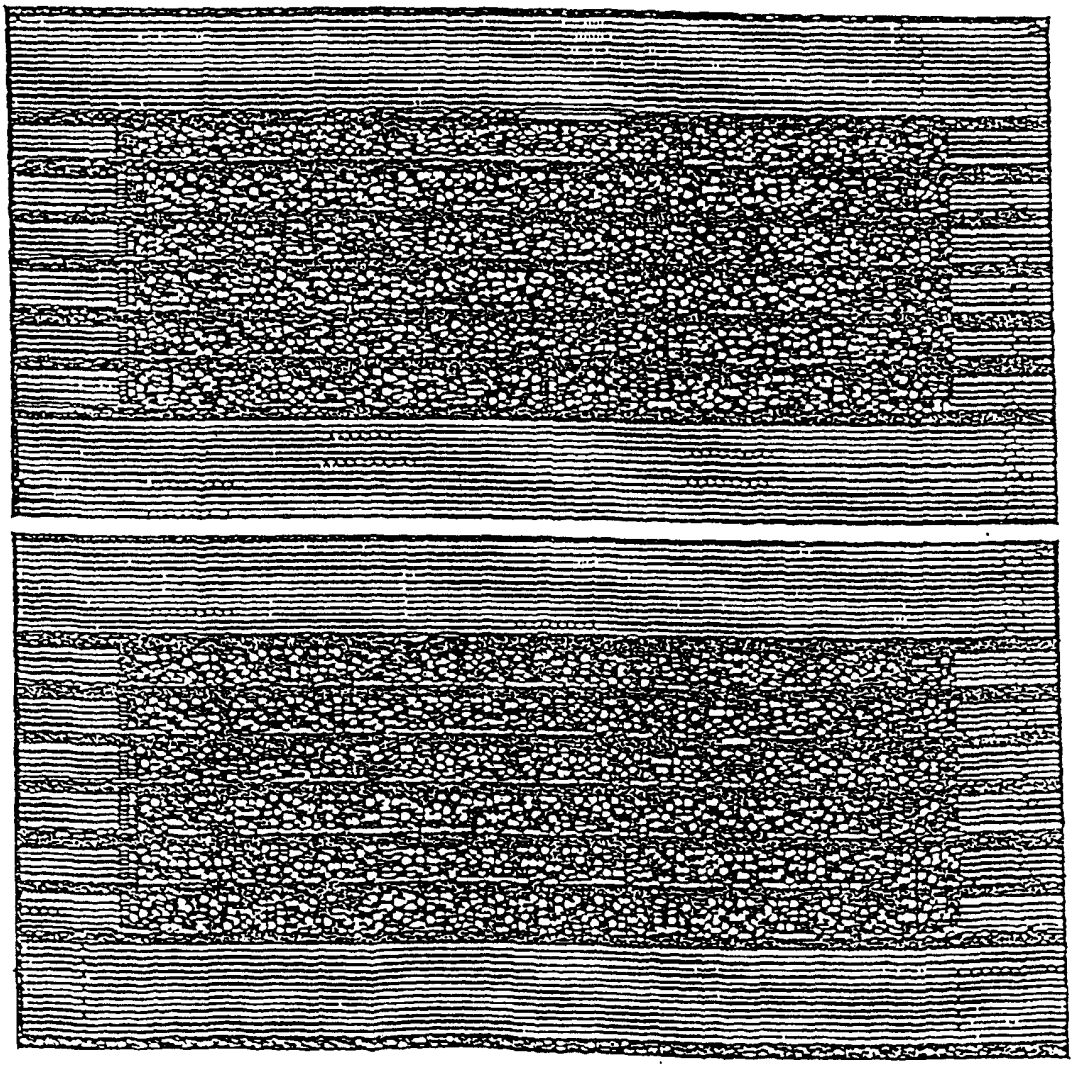


图 11A

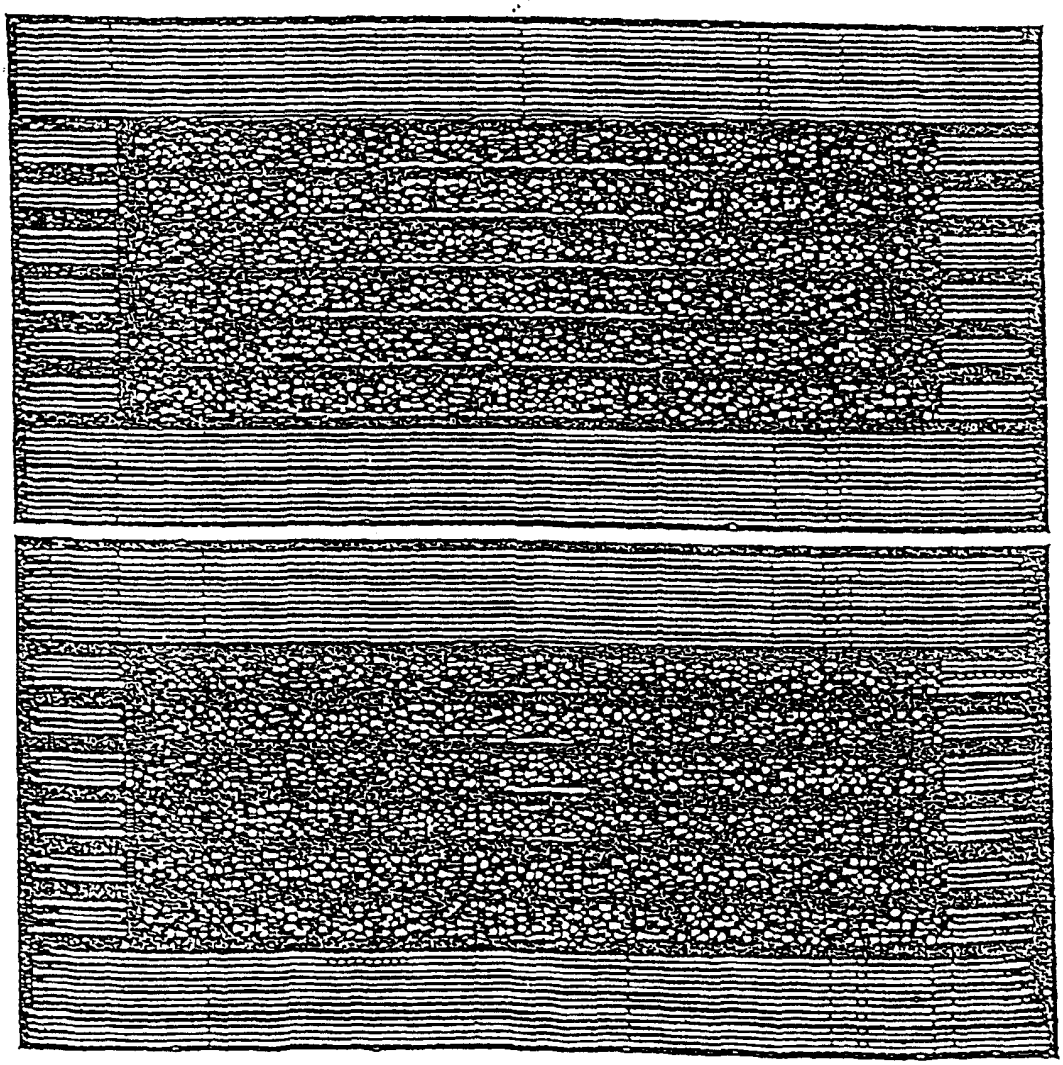


图 12A

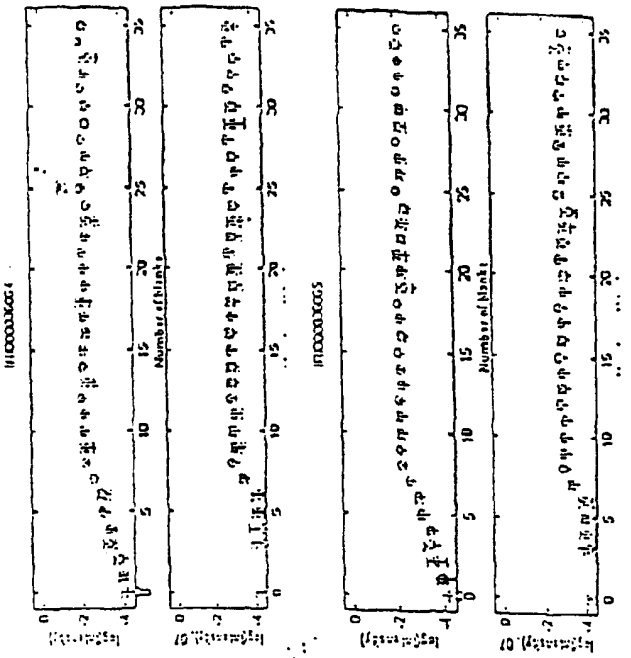


图 12B

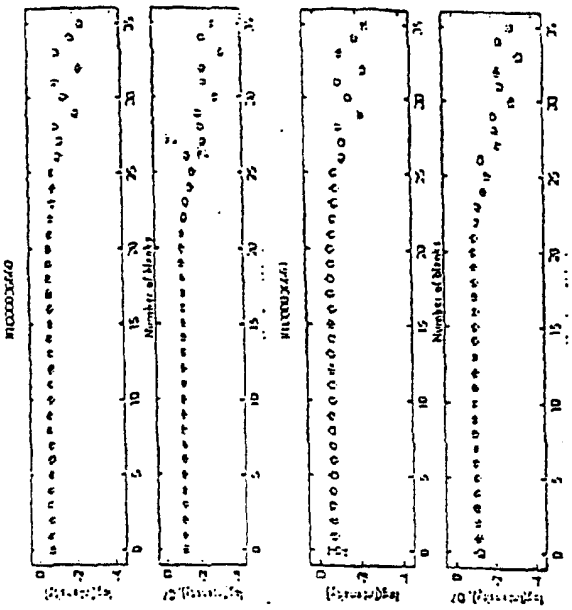


图 13B

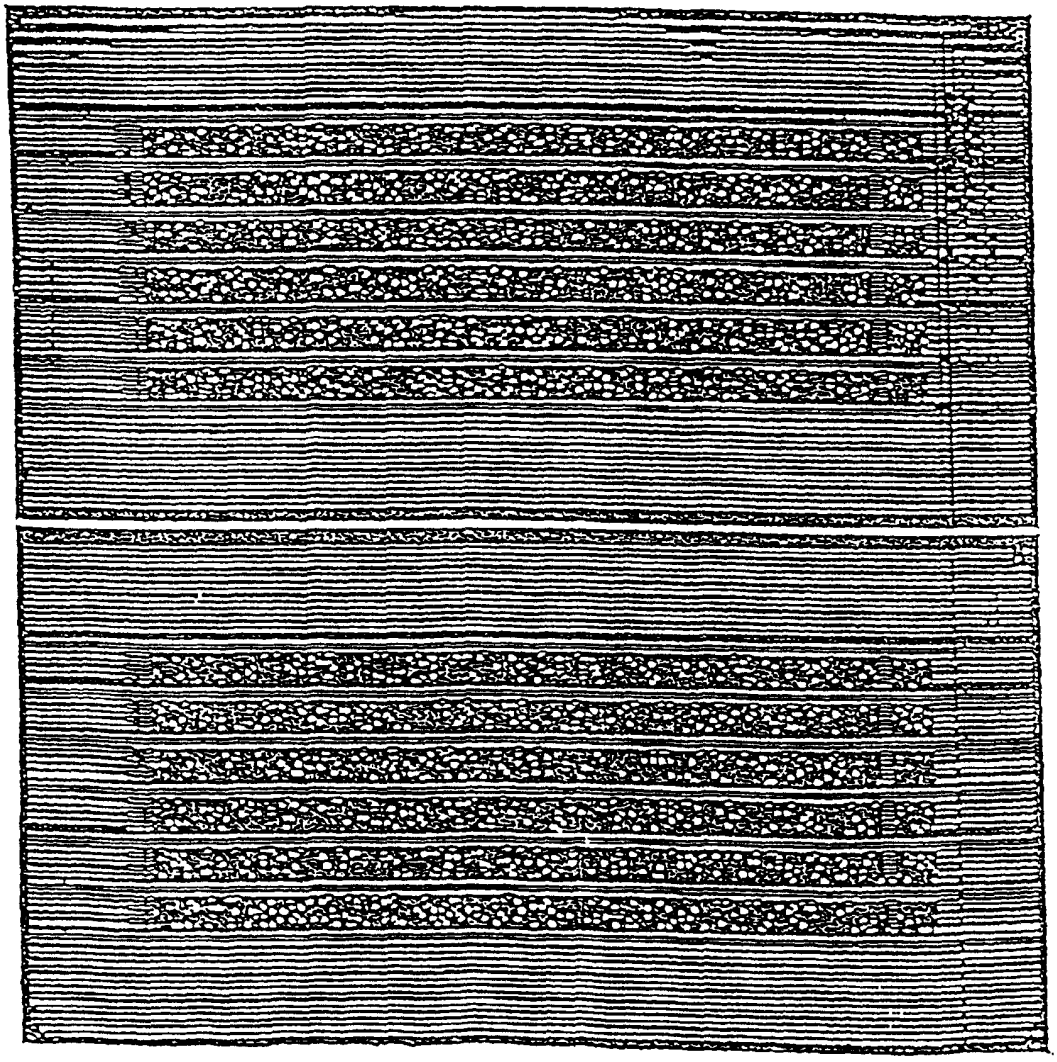


图 13A

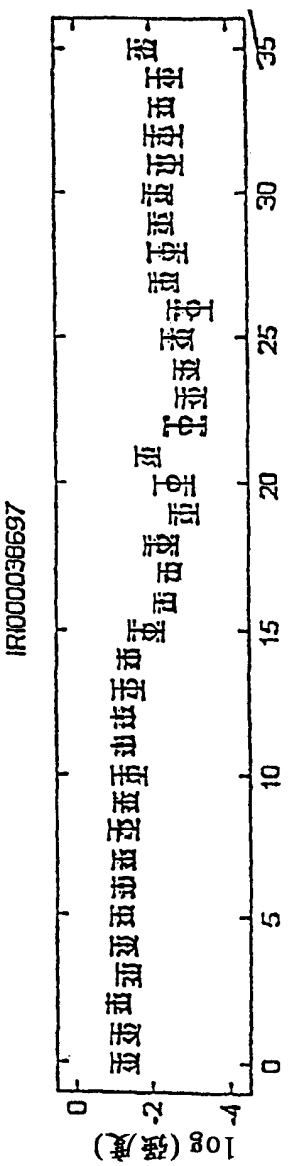

IR0000038697
log (强度)
0
-2
-4
0
5
10
15
20
25
30
35

图 14A

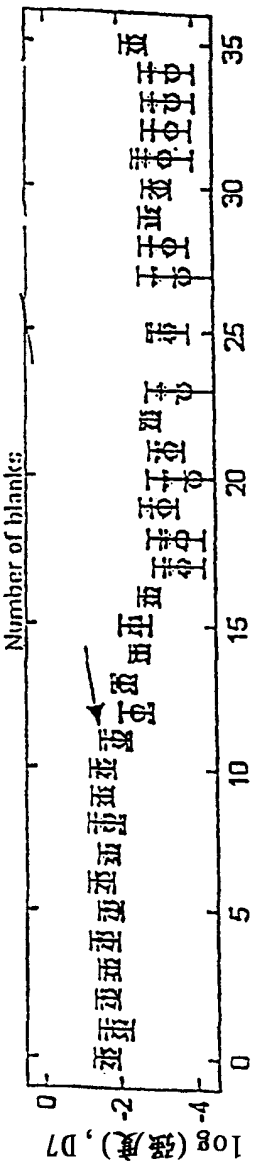

Number of blanks
log (强度), D7
0
-2
-4
0
5
10
15
20
25
30
35

图 14B

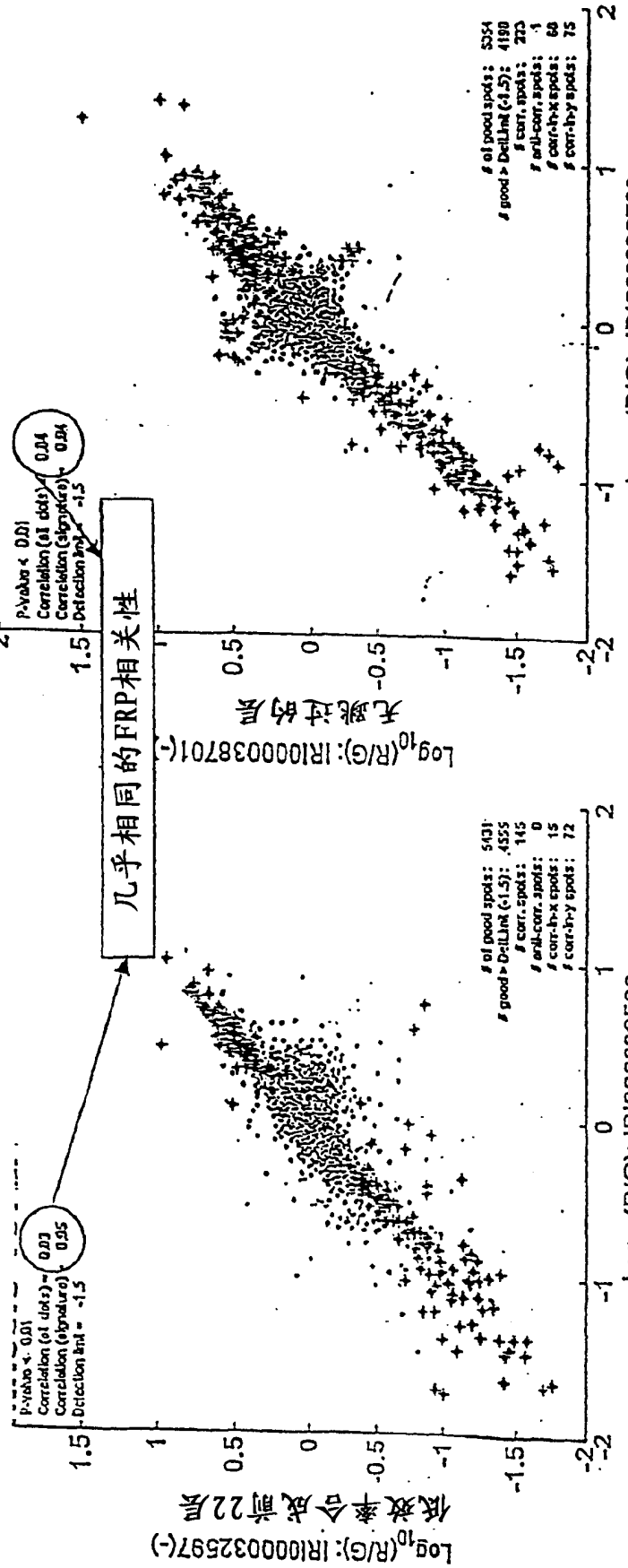

几乎相同的FRP相关性
低效率合成前22层
无跳过的层
Log10(R/G): IRI000032596
Log10(R/G): IRI000038700
P-Value < 0.01
Correlation (all dots) = 0.03
Correlation (signature) = 0.95
Detection limit = -1.5
P-Value < 0.01
Correlation (all dots) = 0.84
Correlation (signature) = 0.84
Detection limit = -1.5
# of good spots: 5431
good > DelLink (-1.5): 4555
corr. spots: 145
anti-corr. spots: 0
corr-h-x spots: 15
corr-h-y spots: 72
of good spots: 5354
good > DelLink (-1.5): 4190
corr. spots: 225
anti-corr. spots: 1
corr-h-x spots: 68
corr-h-y spots: 75

图 15A

图 15B

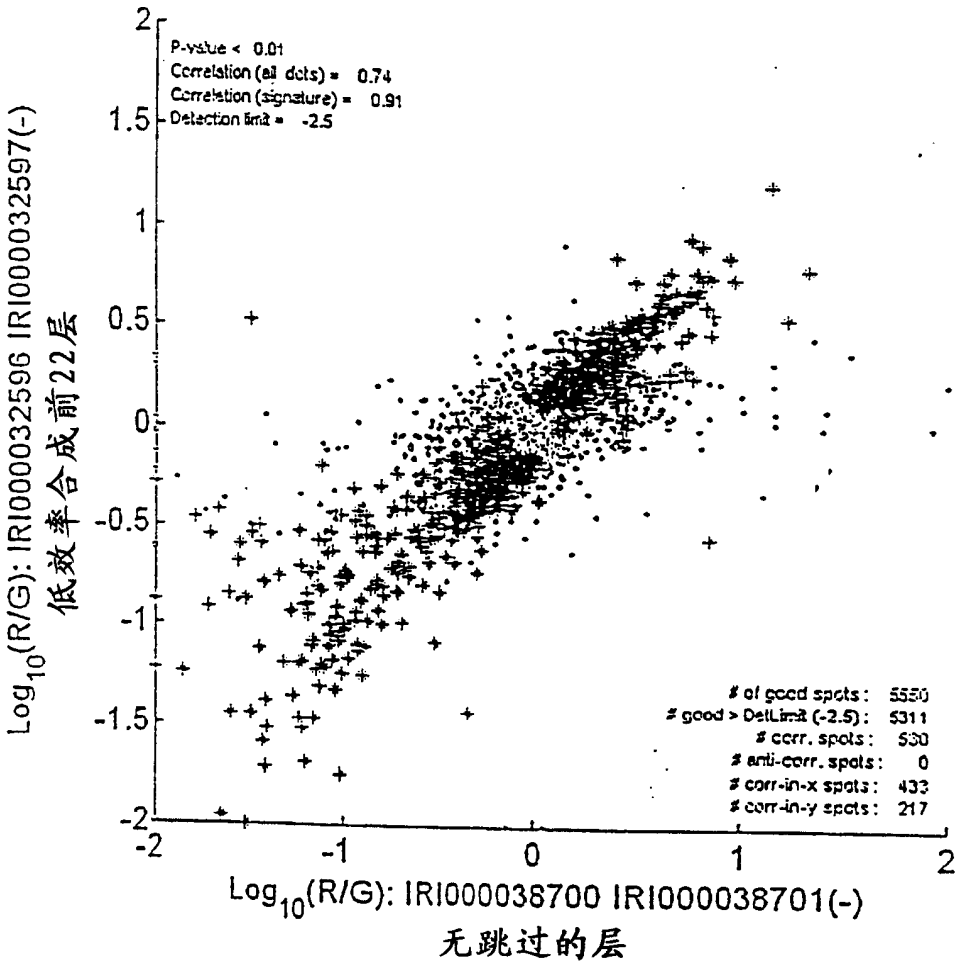

P-value < 0.01
Correlation (all dots) = 0.74
Correlation (signature) = 0.91
Detection limit = -2.5
# of good spots: 5550
# good > DetLimit (-2.5): 5311
# corr. spots: 530
# anti-corr. spots: 0
# corr-in-x spots: 433
# corr-in-y spots: 217
Log10(R/G): IRI000032596 IRI000032597(-)
低效率合成前22层
Log10(R/G): IRI000038700 IRI000038701(-)
无跳过的层

图 15C

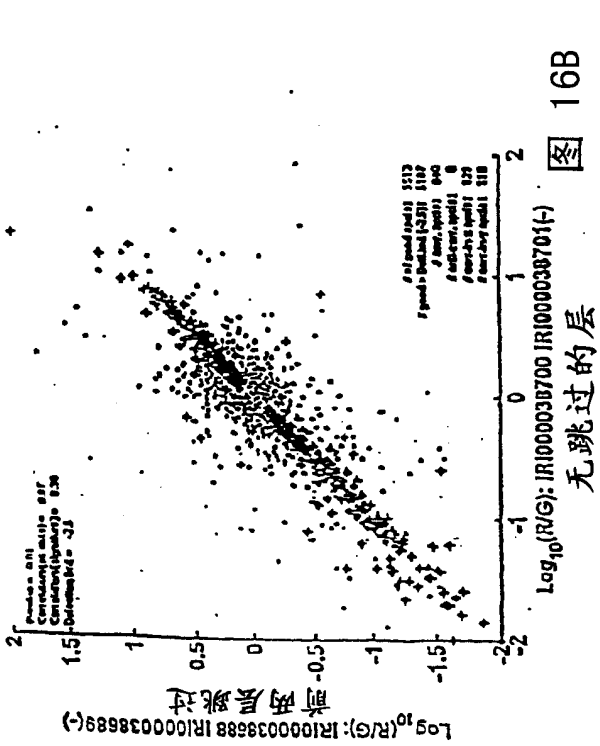


图 16B

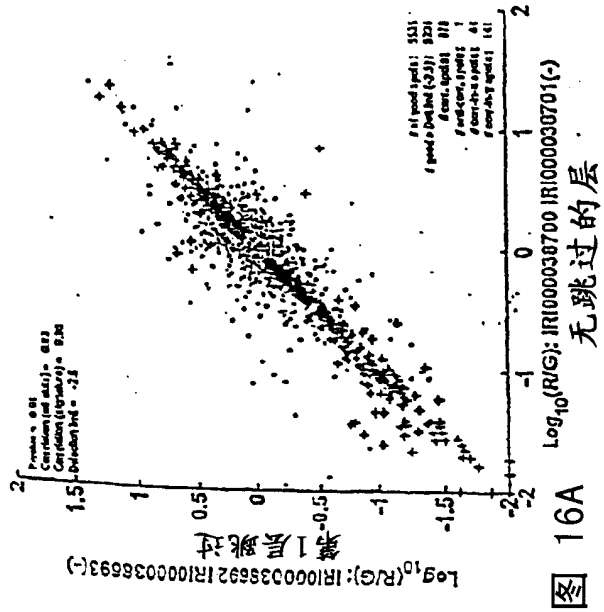


图 16A

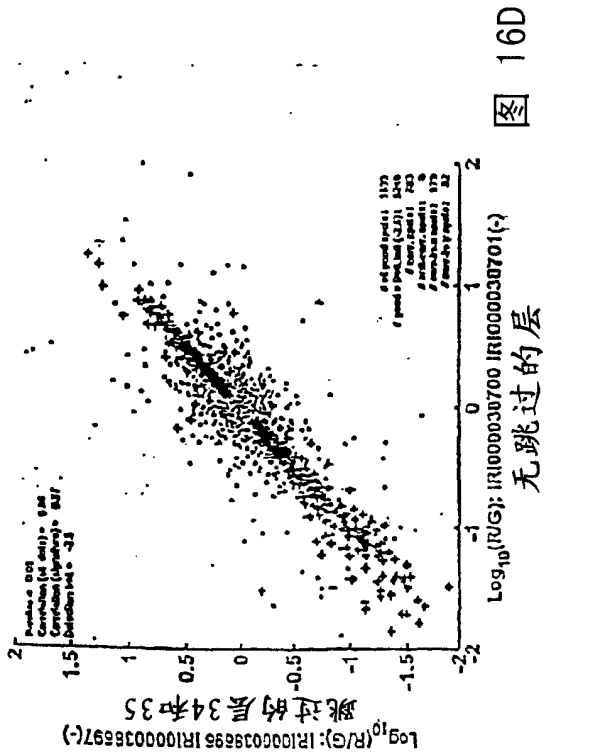


图 16D

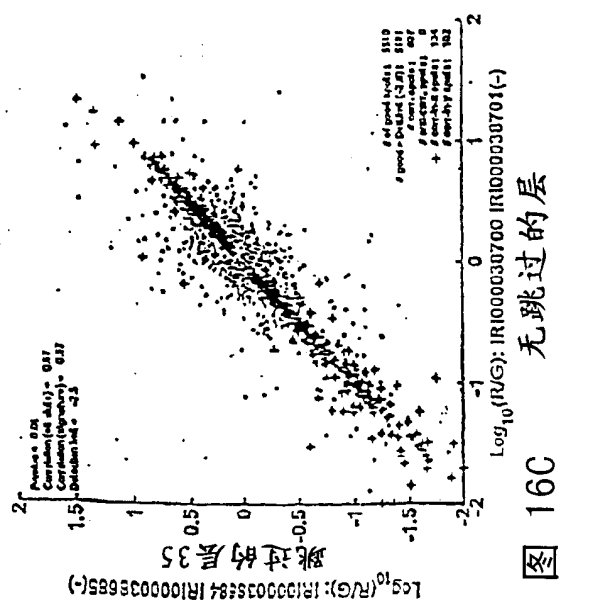


图 16C

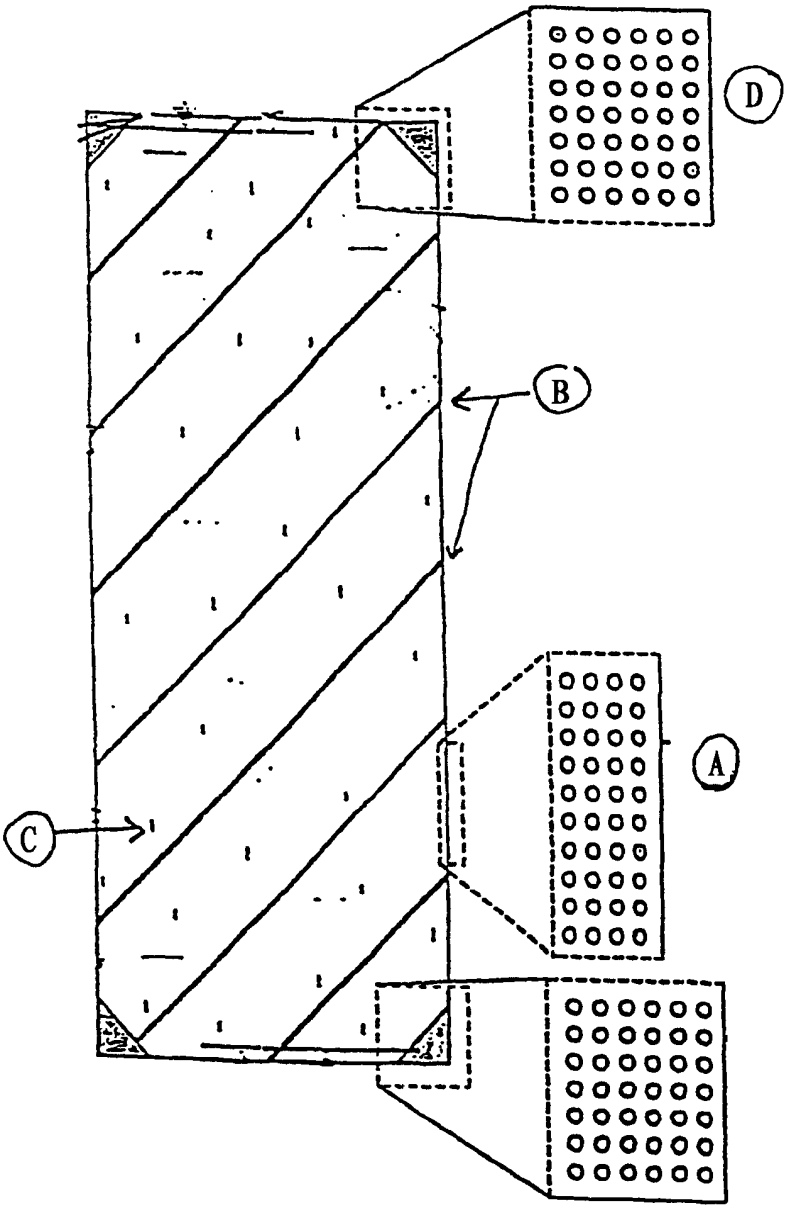


图 17A-17D