

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
G06F 19/00 (2006.01)



## [12] 发明专利申请公布说明书

[21] 申请号 200680036591.1

[43] 公开日 2008 年 10 月 1 日

[11] 公开号 CN 101278295A

[22] 申请日 2006.8.1

[21] 申请号 200680036591.1

[30] 优先权

[32] 2005. 8. 1 [33] US [31] 11/195,254

[86] 国际申请 PCT/US2006/030235 2006.8.1

[87] 国际公布 WO2007/086935 英 2007.8.2

[85] 进入国家阶段日期 2008.4.1

[71] 申请人 454 生命科学公司

地址 美国康涅狄格州

[72] 发明人 K·E·麦达德 J·M·费罗

J·R·奈特 J·乍鲁米林

J·H·利蒙 E·W·迈尔斯

J·W·辛普森 G·A·福尔克默

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 李 波 梁 谋

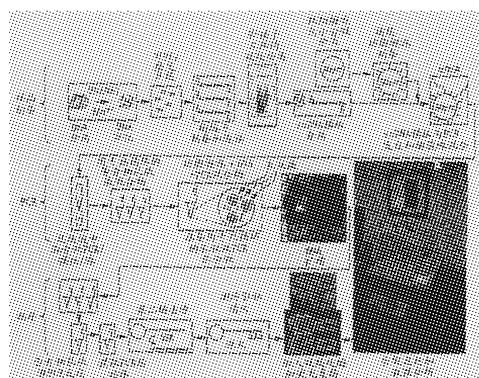
权利要求书 2 页 说明书 171 页 序列表 21 页  
附图 65 页

[54] 发明名称

对核酸进行扩增和测序的方法

[57] 摘要

本文提供了用于进行快速 DNA 测序，如基因组测序的装置和方法。该方法包括以下步骤：制备用于基因组测序的样品 DNA，以代表性的方式扩增制备的 DNA，以及在扩增的 DNA 上用仅仅一个引物杂交步骤进行多个测序反应。



1. 对靶核酸进行测序的方法，包括：

a) 将参照核酸序列转化为参照数目序列，其中参照数目序列包括理想流动图；

b) 当一个或多个核苷酸序贯流过一个或多个拷贝的靶核酸片段时检测信号，信号相应于表明所述片段的核酸序列的询问数目序列，其中询问数目序列包括询问流动图；

c) 在理想流动图的一个或多个位置上将询问流动图与理想流动图进行比对，其中获得询问流动图与理想流动图的近似匹配；

d) 比较询问流动图的数目序列与理想流动图的数目序列，得到质量评分，其中质量评分表明询问流动图与理想流动图之间比对的质量；

e) 对多个片段重复 b) - d) 的步骤；

f) 将多个询问流动图锚定在理想流动图上相应于最高质量比对的位置上；

g) 在一个或多个询问流动图所覆盖的理想流动图的每个位置上对一个或多个询问流动图的序列数目取平均值，从而产生包含共有区流动图的共有区数目序列；和

h) 将共有区流动图转化为核酸序列。

2. 权利要求 1 的方法，进一步包括将理想流动图分为预定长度的重叠的理想亚流动图，并且给理想亚流动图做索引。

3. 权利要求 2 的方法，进一步包括将每个询问流动图分为询问亚流动图，每个亚流动图的长度相应于理想亚流动图的预定长度。

4. 权利要求 3 的方法，进一步包括检索做索引的理想亚流动图，以确定将询问流动图与理想流动图进行比对的位置。

5. 权利要求 2 的方法，其中理想亚流动图是二进制编码的。

6. 权利要求 3 的方法，其中询问亚流动图是二进制编码的。

7. 对靶核酸进行测序的方法，包括：

a) 当一个或多个核苷酸序贯流过一个或多个拷贝的靶核酸片段时检测信号；

b) 将信号与表明片段的核酸序列的询问数目序列进行关联，其中询问数目序列包含询问流动图；

c) 重复步骤 a) 和 b)，得到多个询问流动图；

- d) 将多个询问流动图彼此进行比较,以鉴定多个询问流动图之间的重叠区域;
- e) 在重叠区域比对多个询问流动图;
- f) 基于比对得到质量评分, 其中质量评分表明比对的质量;
- g) 通过确定具有满足预定阈值的质量评分的比对, 鉴定逐对重叠询问流动图;
- h) 将逐对重叠询问流动图分组为一个或多个统一群;
- i) 对每个统一群内的一个或多个匹配位置中每一个位置上的询问流动图的序列数目取平均值, 从而产生包含统一群共有区流动图的共有区数目序列; 和
- j) 将每个统一群共有区流动图转化为统一群共有核酸序列。

8. 权利要求 7 的方法, 其中步骤 h)的一个或多个统一群包含询问流动图的最大重叠的一致链。

9. 权利要求 7 的方法, 进一步包括以下步骤:

k) 将统一群共有核酸序列彼此比较以鉴定序列重叠; 和

l) 连接具有共同重叠序列的统一群共有区, 从而形成包含重叠群核酸序列的一个或多个重叠群。

10. 权利要求 9 的方法, 进一步包括以下步骤:

m) 鉴定每个重叠群内的边界, 其中边界是这样的区域, 即统一群序列在此处从共同区域脱离出来; 并且

n) 在步骤 m)中鉴定出的边界上破坏重叠群。

11. 权利要求 10 的方法, 进一步包括以下步骤:

o) 连接任何两个末端通过相同片段核酸序列而重叠的重叠群, 其中如果鉴定了边界, 则任选破坏如此连接的重叠群。

12. 权利要求 11 的方法, 进一步包括以下步骤:

p) 鉴定片段核酸序列和重叠群之间的所有比对; 其中任选在少于 4 个片段核酸序列进行比对的任何位置破坏重叠群;

q) 通过对与在步骤 p)中与重叠群比对的片段核酸序列相关的询问流动图的序列数目取平均值, 计算重叠群共有区流动图; 和

r) 将重叠群共有区流动图转化为重叠群共有核酸序列。

13. 权利要求 12 的方法, 进一步包括用重叠群共有核酸序列重复步骤 p) - r), 直到基本没有共有区碱基调用改变, 从而计算最终重叠群共有序列。

---

## 对核酸进行扩增和测序的方法

### 发明领域

本发明涉及用于确定 DNA 的碱基序列的方法和装置。更具体地，本发明涉及可以用于自动化或半自动化扩增和确定基因组的碱基序列的方法和装置。

### 发明背景

利用自动化 DNA 测序仪的快速和灵敏核酸测序方法的开发彻底改革了现代分子生物学。通过一系列机器和一组技术人员的协同努力，目前可以分析植物、真菌、动物、细菌和病毒的完整基因组。但是，在短时间内对基因组进行快速和自动化或半自动化测序的目标目前还不能实现。仍然存在准确样品制备、扩增和测序的技术问题。

阻碍基因组序列分析的一个技术问题是研究者不能在短时间内快速制备多个包含完整基因组的核酸样品。

另一个技术问题是不能代表性地扩增基因到与目前的测序方法的灵敏度相容的水平。现代经济上可行的测序仪器尽管是灵敏的，但仍然需要超过 100 万拷贝的 DNA 片段来进行测序。目前用于提供高拷贝 DNA 测序的方法包括克隆或体外扩增的变化形式，其不能扩增经济地对完整基因组进行测序所需数目的单个克隆（600,000 或更多，对于人基因组是数千万）。

还有另一个技术问题是目前测序方法的限制，对于每次寡核苷酸引物杂交，其最多可以进行一次测序反应。测序引物的杂交通常是限速步骤，其限制 DNA 测序仪的通量。

在大多数情况下，聚合酶链反应（PCR; Saiki, R.K., et al., *Science* 1985, 230, 1350 – 1354; Mullis, K., et al., *Cold Spring Harb. Symp. Quant. Biol.* 1986, 51 Pt 1, 263 – 273）在获得 DNA 序列信息和扩增微小量特异性 DNA 以获得足够测序的浓度方面起主要作用。但是，将目前的 PCR 技术放大以满足现代遗传学不断增加的需求，既不经济，也不有效，特别是当需要考虑完整基因组测序的要求时更是如此。

使时间和经济效率最高的努力一般集中在两个方面：减少扩增所需

的反应体积和增加同时进行的扩增的数量。小型化赋予了采用更少样品和试剂、减少扩增时间和增加通量可测量性的优点。

尽管常规热循环仪由于热质量限制需要相对长的循环时间 (Woolley, A.T., et al., Anal. Chem. 1996, 68, 4081 – 4086), 较小的反应体积可以更快循环。连续流动 PCR 装置利用了蚀刻的微通道与固定温度区的组合, 以便将反应体积减少到亚微升样品水平 (Lagally, E.T., et al., Analytical Chemistry 2001, 73, 565 – 570; Schneegas, I., et al., Lab on a Chip – The Royal Society of Chemistry 2001, 1, 42 – 49)。

由空气 (Kalinina, O., et al., Nucleic Acids Res. 1997, 25, 1999 – 2004) 或红外光 (Oda, R.P., et al., Anal. Chem. 1998, 70, 4361 – 4368; Huhmer, A.F. and Landers, J.P., Anal. Chem. 2000, 72, 5507 – 5512) 加热的玻璃微毛细管也已经作为用于毫微升规模反应的有效容器。已经用微构造的硅热循环仪实现了相似的反应体积 (Burns, M.A., et al., Proc. Natl. Acad. Sci. USA 1996, 93, 5556 – 5561)。

在很多情况下, 这些小型化使总 PCR 反应时间减少, 对于改进的电热元件 (Kopp, M.U., et al., Science 1998, 280, 1046 – 1048; Chiou, J., Matsudaira, P., Sonin, A. and Ehrlich, D., Anal. Chem. 2001, 73, 2018 – 2021) 和热空气循环仪 (Kalinina, O., et al., Nucleic Acids Res. 1997, 25, 1999 – 2004) 减少到少于 30 分钟, 对于一些红外线控制的反应, 减少到 240 秒 (Giordano, B.C., et al., Anal. Biochem. 2001, 291, 124 – 132)。

某些技术同时利用增加的通量和小型化; 如 Sasaki 等人的 1536 凹陷系统设计 (Sasaki, N., et al., DNA Res. 1997, 4, 387 – 391), 其使反应体积保持在 1  $\mu$ l 以下。作为另一个实例, Nagai 等人 (Nagai, H., et al., Biosens. Bioelectron. 2001, 16, 1015 – 1019; Nagai, H., et al., Anal. Chem. 2001, 73, 1043 – 1047) 报道了在蚀刻在单个硅片中的 1 万个 86 pl 反应凹陷中扩增单个测试片段。不幸的是, 已经证明从这些方法回收和利用扩增子是有问题的, 需要通过选择性通透膜进行蒸发。

尽管存在反应体积和循环时间的这些显著改进, 以前的策略都没有提供显著将通量增加到分析完整人基因组所需的水平所需要的大规模平行扩增。DNA 测序仪仍然比需要的更慢和更昂贵。在纯研究情况下, 如果测序仪慢和昂贵, 可能是可接受的。但是当需要将 DNA 测序仪用于临床诊断情况下, 所述低效测序方法甚至对于财政充足的机构也是禁

止的。数千种克隆扩增的靶的大规模平行测序将显著促进大规模的全基因组文库分析，而不需要浪费时间的样品制备过程和昂贵的易错克隆过程。成功的高容量、固相克隆 DNA 扩增可以用于多种应用。因此，很明确需要制备基因组或大模板核酸，用于测序、用于扩增核酸模板，和用于对扩增的模板核酸进行测序，而不存在每次杂交进行一次测序反应的限制。此外，需要用一个系统将这些多个技术连接为一个可行的自动化或半自动化测序仪。

### 发明概述

本发明描述了一种集成的系统，包括新方法和新装置，用于(1)核酸样品制备，(2)核酸扩增，和(3)DNA 测序。

本发明提供了一种新方法，用于制备多个 DNA 序列的文库，特别是来源于大模板 DNA 或完整（或部分）基因组 DNA 的 DNA 序列的文库。单链 DNA 的序列通过单链 DNA 的断裂、磨光（Polishing）、衔接子连接、切口修复和分离，从大模板 DNA 或完整或部分 DNA 基因组的样品制备。该方法提供了制备连接于固体支持物的 ssDNA 文库的方法，包括(a)制备 ssDNA 模板的文库；(b)将 ssDNA 模板连接于固相；和(c)分离连接了 ssDNA 模板的固体支持物。

本发明也提供了通过以下步骤在单个反应管中扩增 DNA 文库的每个单独的成员的方法，所述步骤例如是：将多个 DNA 样品分别单独包封在乳状液的微胶囊中，同时进行多个包封的核酸样品的扩增，以及从微胶囊释放所述扩增的多个 DNA，用于后续反应。在一种实施方案中，使单拷贝的核酸模板种类与 DNA 捕获珠杂交，悬浮于完全的扩增溶液中，并且乳化到微反应器（典型是直径 100–200 微米）中，此后，用扩增（如 PCR）使最初模板种类的拷贝数克隆增加到超过 1,000,000 个拷贝的单核酸序列，优选 2 百万–2 千万个拷贝的单核酸。例如，扩增反应可以同时用每毫升反应混合物至少 3,000 个微反应器同时进行，并且可以在单个 100 $\mu$ l 体积试管（如 PCR 反应管）中用超过 300,000 个微反应器进行。本发明也提供了富集这些含有成功 DNA 扩增事件的珠子的方法（即，通过除去不连接 DNA 的珠子）。

本发明也提供了采用单个引物杂交步骤对来自多个引物的核酸进行测序的方法。使两个或更多测序引物与要测序的模板 DNA 杂交。然

后保护除一个测序引物之外的所有测序引物。通过延伸未保护的引物再次进行测序（例如焦磷酸测序）。使延伸完成（如果必要，用额外的聚合酶和 dNTPs）或终止（通过聚合酶和 ddNTPs）。除去链完成和/或终止试剂。然后使受保护的引物之一去保护，并且通过延伸新去保护的引物进行测序。继续进行该过程，直到所有测序引物都去保护并且测序。在一种优选实施方案中，用两个引物（一个受保护，一个未保护）对双链核酸的两个末端进行测序。

本发明也提供了用基于焦磷酸的测序方法对核酸进行测序的装置和方法。该装置具有电荷耦合器件(CCD)照相机、微射流室( microfluidics chamber )、样品药筒支持物 ( sample cartridge holder )、泵和流量阀。该装置采用化学发光作为检测方法，其对于焦磷酸测序具有固有的低背景。在一种优选实施方案中，用于测序的样品药筒称作“微微量滴定板 ( PicoTiter Plate )”，并且由商购的光导纤维面板形成，所述面板经过酸蚀刻，产生数十万个非常小的孔，每个孔体积是 75 pL。该装置包括适于与本文描述的阵列一起使用的新颖的试剂送递杯，以便给微微量滴定板提供流体试剂，并且包括一个与试剂送递杯连通的试剂送递工具。将来自微微量滴定板上每个孔的光子引导到 CCD 照相机的特定像素中，以检测测序反应。

本发明的一种实施方案涉及对核酸进行测序的方法，该方法包括：(a)提供置于平坦表面上的多个凹陷中的多个单链核酸模板，每个凹陷形成一个分析物反应室，其中反应室中心与中心的间距是 20 - 100 $\mu$ m，并且平坦表面上具有至少 10,000 个反应室；(b)通过以下步骤在所有反应室中同时进行基于焦磷酸的测序反应：使有效量的测序引物与核酸模板退火，并且用聚合酶和预定核苷酸三磷酸延伸测序引物，得到测序产物，并且，如果预定核苷酸三磷酸掺入所述测序引物的 3' 末端，则得到测序反应副产物；以及(c)鉴定测序反应副产物，从而确定每个反应室中的核酸的序列。

在本发明的任何方法中，可以在腺苷三磷酸双磷酸酶存在下进行测序反应。腺苷三磷酸双磷酸酶可以在溶液中，或者腺苷三磷酸双磷酸酶可以固定在表面上（例如，在分析物反应室中）。或者，腺苷三磷酸双磷酸酶可以固定在置于本发明的测序反应/方法的分析物反应室中的可移动固体支持物上。

在包括测序的本发明的任何实施方案中，测序可以通过基于焦磷酸的测序反应进行。此外，测序反应可以在 34°C - 36°C 下进行。在一种更优选的实施方案中，测序反应在 35°C 下进行。选择该温度，以便优化测序试剂，例如腺苷三磷酸双磷酸酶的功能。在一种最优选的实施方案中，测序反应在腺苷三磷酸双磷酸酶存在下进行。

本发明的另一实施方案涉及对核酸进行测序的方法，包括：(a)使大模板核酸分子断裂，产生多个断裂的核酸；(b)在油包水乳状液中将断裂的核酸送递到含水的微反应器中，使得多个含水微反应器包含单个拷贝的断裂的核酸、单个能够结合断裂的核酸的珠子，和含有进行核酸扩增所需的试剂的扩增反应溶液；(c)在微反应器中扩增断裂的核酸，以形成扩增的拷贝的核酸，并且使扩增的拷贝与微反应器中的珠子结合；(d)将珠子送递到平坦表面上的至少 10,000 个反应室的阵列中，其中所述多个反应室包含不超过单个珠子；和(e)同时在多个反应室中进行测序反应。

此外，本文描述的方法具有以下额外的特征：反应室的中心与中心的间距可以是 20 - 100 μm。断裂的核酸的长度可以是 30 - 500 个碱基。所述多个珠子结合至少 10,000 个扩增的拷贝。步骤(c)可以用聚合酶链反应进行。测序反应可以是基于焦磷酸的测序反应。例如，测序反应可以包括以下步骤：(a)使有效量的测序引物与扩增的拷贝的核酸退火，并且用聚合酶和预定核苷酸三磷酸延伸测序引物，得到测序产物，并且，如果预定核苷酸三磷酸掺入所述测序引物的 3' 末端，则得到测序反应副产物；以及(b)鉴定测序反应副产物，从而确定多个反应室中的核酸的序列。作为另一个实例，测序可以包括以下步骤：(a)使两个或多个测序引物与核酸分子的一个或多个单链杂交，其中除一个引物之外的所有引物都是可逆阻断的引物；(b)通过从未阻断的引物进行聚合酶延伸，将至少一个碱基掺入核酸分子；(c)防止未阻断的引物进一步延伸；(d)将可逆阻断的引物去阻断为未阻断的引物；以及(e)重复步骤(b) - (d)，直到至少一个可逆阻断的引物被去阻断，并且用于确定序列。可以通过蚀刻光导纤维束的一个末端而形成这些方法中反应室。

本发明的另一实施方案涉及包含具有多个凹陷的平坦表面的阵列，每个凹陷形成功能分析物反应室，其中反应室的中心与中心的间距是 20 - 100 μm，并且每个凹陷在至少一个维度的宽度是 20 μm - 70 μm，并且其

中有至少 10,000 个反应室。所述多个反应室可以含有至少 100,000 个拷贝的单种类的单链核酸模板。单链核酸模板可以固定在置于反应室中的可移动的固体支持物上。凹陷的中心与中心的间距是 40 - 60 $\mu\text{m}$ , 并且每个凹陷的深度可以是 20 $\mu\text{m}$  - 60 $\mu\text{m}$ 。

本发明的另一实施方案涉及包含平坦的上表面和平坦的下表面的阵列，其中平坦的上表面中具有至少 10,000 个凹陷，每个凹陷形成分析物反应室，平坦的下表面任选是传导性的，这样来自反应室的光学信号可以通过平坦的下表面进行检测，其中上表面和下表面之间的距离不超过 5 mm，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ，并且每个室在至少一个维度的宽度是 20 $\mu\text{m}$  - 70 $\mu\text{m}$ 。在该阵列中，上表面和下表面之间的距离不超过 2 mm。阵列上凹陷的数目可以超过 50,000，超过 100,000。每个反应室的形状可以基本上是六边形的。此外，每个凹陷可以具有至少一个不规则的壁表面。此外，凹陷可以具有光滑的壁表面。

阵列可以在熔凝的光导纤维束中形成。这可以通过例如蚀刻光导纤维束的一个末端而进行。每个凹陷可以含有用于分析核酸或蛋白的试剂。该阵列可以进一步包含位于平面阵列之外并且与其相对接触的第二个表面，使得在阵列上形成流动室。

本发明的另一实施方案涉及用于在含水环境中进行分开的平行共同反应的阵列工具，其中该阵列工具包含一种基质，该基质含有至少 10,000 个离散的反应室，反应室中含有能够与试剂反应的起始材料，每个反应室的尺寸使得当一种或多种含有至少一种试剂的流体送递到每个反应室时，试剂扩散到孔外的扩散时间超过起始材料与试剂反应形成产物所需要的时间。阵列的每个凹陷可以含有用于分析核酸或蛋白的试剂。阵列可以进一步包含置于反应室内的可移动的固体支持物群，每个可移动的固体支持物具有一个或多个与其连接的生物活性试剂。可以通过蚀刻、造型或显微机械加工，在基质上形成阵列中的凹陷。基质可以是光导纤维束。

在本发明的任何阵列中，至少 5% - 20%、至少 20% - 60% 或至少 50% - 100% 的反应室可以含有至少一个可移动的固体支持物，其上固定至少一种试剂。固定在可移动的固体支持物上的试剂可以是具有硫酸化酶活性的多肽、具有荧光素酶活性的多肽或具有固定的硫酸化酶和荧光素酶两种活性的多肽。所述多个反应室可以含有至少 100,000 个拷贝的

单种类的单链核酸模板。该阵列可以含有固定在置于反应室中的可移动的固体支持物上的单链核酸模板。本发明的阵列可以适用于焦磷酸测序反应。

本发明的另一实施方案涉及将生物活性试剂递送到阵列中的方法，包括在阵列上分散多个可移动的固体支持物，每个固体支持物具有至少一种固定在其上的试剂，其中该试剂适用于核酸测序反应中，其中该阵列包含一个平坦的表面，该表面具有多个置于其上的反应室，其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$ ，并且每个反应室在至少一个维度的宽度是至少 20 $\mu\text{m}$  – 70 $\mu\text{m}$ 。

本发明的另一实施方案涉及用于同时监测反应室阵列中表明在特定位点发生反应的光的装置，该装置包含以下组件：(a)从包含多个有凹陷的表面的平面基质形成的反应室阵列，每个有凹陷的表面形成适于容纳分析物的反应室，并且其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$ ，每个反应室的体积是 10 – 150 pL，该阵列包含超过 10,000 个离散的反应室；(b)光敏器件，其排列使得在使用中，来自特定反应室的光将撞击光敏器件的特定预定区；(c)用于确定撞击每个预定区的光水平的工具；和(d)记录每个反应室中随时间的光水平变化的工具。

本发明的另一个实施方案涉及一种分析传感器，包含以下组件：(a)从第一个光导纤维束形成的阵列，该光导纤维束的一端具有多个有凹陷的表面，每个有凹陷的表面形成适于容纳分析物的反应室，并且其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$ ，宽度是 20 – 70 $\mu\text{m}$ ，该阵列包含超过 10,000 个离散的反应室；(b)用于在反应室中产生光的酶或荧光工具；(c)包含光捕获工具的光检测工具和用于将光传输到光检测工具的第二个光导纤维束，所述第二个光导纤维束与阵列形成光学接触，使得各个反应室中产生的光被第二个光导纤维束的分开的纤维或分开的纤维的组捕获，用于传输到光捕获工具中。该传感器可以适用于生化测定或基于细胞的测定。光捕获工具可以是 CCD 照相机。反应室可以含有一个或多个可移动的固体支持物，其上固定了生物活性试剂。

本发明的另一实施方案涉及用于在含水环境中进行分开的平行共同反应的方法，包括以下步骤：(a)将含有至少一种试剂的流体递送到阵列，其中该阵列包含基质，该基质包含至少 10,000 个离散的反应室，每个反应室适于容纳分析物，并且其中反应室的体积是 10 – 150 pL，并且

含有能够与试剂反应的起始材料，每个反应室的尺寸使得当流体送递到每个反应室时，试剂扩散到凹陷外的扩散时间超过起始材料与试剂反应形成产物所需要的时间；以及(b)在以下时间段中从阵列洗去流体：(i)在起始材料与试剂反应从而在每个反应室中形成产物之后但(ii)在送递到任意一个反应室的试剂从该反应室扩散到任何其它反应室之前。一方面，在任意一个反应室中形成的产物不依赖于在任何其它反应室中形成的产物，但是是采用一种或多种共同试剂产生的。另一方面，起始材料是核酸序列，并且流体中的至少一种试剂是核苷酸或核苷酸类似物。另一方面，流体还包含能够与核酸序列和核苷酸或核苷酸类似物反应的聚合酶。另一方面，可以按顺序重复步骤(a)和(b)。

本发明的另一实施方案涉及将核酸测序酶送递到阵列的方法，该阵列具有平坦的表面，该表面具有多个凹陷，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ；该方法包括在阵列上散布多个固定了一种或多种核酸测序酶的固体支持物，使得所述多个反应室含有至少一种可移动的固体支持物。核酸测序酶之一是具有硫酸化酶活性、荧光素酶活性或这两种活性的多肽。

本发明的另一实施方案涉及将多个核酸模板送递到阵列中的方法，该阵列上具有平坦的表面，该表面具有多个凹陷，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$  并且该阵列具有至少 10,000 个反应室；该方法包括在阵列上散布多个可移动的固体支持物，每个可移动的固体支持物上固定了不超过单个种类的核酸模板，该散布导致任何一个反应室中放置不超过一个可移动的固体支持物。一方面，核酸序列可以是单链核酸。另一方面，至少 100,000 个拷贝的单个种类的核酸模板可以固定在多个可移动的固体支持物上。可以在微量滴定板上扩增每单个种类的核酸模板，在置于反应室后产生每孔至少 2,000,000 个拷贝的核酸模板。作为一个实例，扩增可以通过聚合酶链反应、连接酶链反应或等温 DNA 扩增而进行。

本发明的另一实施方案涉及对核酸进行测序的方法，该方法包括以下步骤：(a)提供置于平坦表面上的多个凹陷中的多个单链核酸模板，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$  并且该平坦表面具有至少 10,000 个反应室；(b)通过以下步骤在所有反应室中同时进行基于焦磷酸的测序反应：使有效量的测序引物与

核酸模板退火，并且用聚合酶和预定核苷酸三磷酸延伸测序引物，得到测序产物，并且，如果预定核苷酸三磷酸掺入所述测序引物的3'末端，则得到测序反应副产物；以及(c)鉴定测序反应副产物，从而确定每个反应室中的核酸的序列。在该方法中，测序反应副产物可能是 PP<sub>i</sub>，并且可以用偶联的硫酸化酶/荧光素酶反应产生用于检测的光。此外，硫酸化酶和荧光素酶中的一个或这两者固定在置于每个反应位点的一个或多个可移动的固体支持物上。

本发明的另一实施方案涉及在阵列上确定多个核苷酸的碱基序列的方法。该方法包括以下步骤：(a)提供至少 10,000 个 DNA 模板，每个模板独立地置于平坦表面上的多个凹陷中，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 μm 并且反应室的体积是 10 - 150 pL；(b)将一个已知的含氮碱基的激活的核苷 5' - 三磷酸前体加入每个反应室中的反应混合物，每个反应混合物包含针对模板的核苷酸聚合酶和与比模板短至少一个核苷酸残基的互补寡核苷酸引物链杂交的单链多核苷酸模板，以便在允许激活的核苷 5' - 三磷酸前体掺入引物链的 3' 末端的条件下在引物链的 3' 末端在每个模板中形成至少一个未配对的核苷酸残基，前提是激活的核苷 5' - 三磷酸前体的含氮碱基与模板的未配对核苷酸残基的含氮碱基互补；(c)检测核苷 5' - 三磷酸前体是否掺入引物链，其中核苷 5' - 三磷酸前体的掺入表明模板的未配对核苷酸残基具有与掺入的 5' - 三磷酸前体的含氮碱基组成互补的含氮碱基组成；(d)顺序重复步骤(b)和(c)，其中每次顺序重复添加并且检测已知含氮碱基组成的一种类型的激活的核苷 5' - 三磷酸前体的掺入；以及(e)从掺入的核苷前体的序列确定每个反应室中模板的未配对核苷酸残基的碱基序列。

本发明的另一实施方案涉及鉴定模板 DNA 的 DNA 序列中的靶位置中的碱基的方法，其中(a)至少 10,000 个独立的 DNA 模板独立地置于平坦表面上的多个凹陷中，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 μm，在置于反应室中之前或之后，使 DNA 成为单链；(b)提供延伸引物，其在紧接靶位置的位置上与固定的单链 DNA 杂交；(c)在预定的脱氧核苷酸或双脱氧核苷酸的存在下使固定的单链 DNA 进行聚合酶链反应，其中如果预定的脱氧核苷酸或双脱氧核苷酸掺入测序引物的 3' 末端，则形成测序反应副产物；以及(d)鉴定测序

反应副产物，从而确定与 10,000 个 DNA 模板中每一个 DNA 模板的靶位置互补的核苷酸。在该方法中，代替脱氧或双脱氧腺苷三磷酸(ATP)，可以使用能够作为聚合酶的底物但不能作为 ppi 检测酶的底物的 dATP 或 ddATP 类似物。

本发明的另一方面涉及用于分析核酸序列的装置，该装置包含以下组件：(a)试剂送递杯，其中该送递杯包含阵列，该阵列包含具有多个凹陷的平坦表面，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$ ，并且存在超过 10,000 个反应室，并且其中试剂送递杯含有用于测序反应的试剂；(b)与试剂送递杯连通的试剂送递工具；(c)与试剂送递室连通的成像系统；和(d)与成像系统连通的数据采集系统。

本发明的另一实施方案涉及用于在阵列上确定多个核苷酸的碱基序列的装置，该装置包含：(a)含有位于平坦表面上的多个凹陷的试剂杯，每个凹陷形成形成分析物反应室，其中存在超过 10,000 个反应室，每个反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$  并且反应室的体积是 10 – 150  $\text{pL}$ ；(b)用于将一个已知含氮碱基的激活的核苷 5' – 三磷酸前体同时加入每个反应室的试剂送递工具，每个反应混合物包含针对模板的核苷酸聚合酶和与比模板短至少一个核苷酸残基的互补寡核苷酸引物链杂交的单链多核苷酸模板，以便在允许激活的核苷 5' – 三磷酸前体掺入引物链的 3' 末端的条件下在引物链的 3' 末端在每个模板中形成至少一个未配对的核苷酸残基，前体是激活的核苷 5' – 三磷酸前体的含氮碱基与模板的未配对核苷酸残基的含氮碱基互补；(c)用于在每个反应室中检测核苷 5' – 三磷酸前体是否掺入引物链的检测工具，其中核苷 5' – 三磷酸前体的掺入表明模板的未配对核苷酸残基具有与掺入的 5' – 三磷酸前体的含氮碱基组成互补的含氮碱基组成；(d)用于顺序重复步骤(b)和(c)的工具，其中每次顺序重复添加并且检测已知含氮碱基组成的一种类型的激活的核苷 5' – 三磷酸前体的掺入；和(e)用于在每个反应室中同时从掺入的核苷前体的序列确定每个反应室中模板的未配对核苷酸残基的碱基序列的数据处理工具。

本发明的另一实施方案涉及用于处理多个分析物的装置，该装置包含以下组件：(a)具有置于其中的基质的流动室，该基质包含位于光导纤维束上的至少 50,000 个有凹陷的表面，每个有凹陷的表面形成适于容纳

分析物的反应室，并且其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$  并且反应室的直径是 20 - 70 $\mu\text{m}$ ；(b)用于将处理试剂从一个或多个储液器送递到流动室，从而使置于反应室中的分析物暴露于试剂的流动工具；和(c)用于同时检测来自每个反应室的一系列光学信号的检测工具，该系列的每个光学信号表明处理试剂和置于反应室中的分析物之间的相互作用，其中检测工具与有凹陷的表面连通。检测工具可以是 CCD 照相机。分析物可以是核酸。此外，分析物可以固定在一个或多个置于反应室中的可移动的固体支持物上。处理试剂可以固定在一个或多个可移动的固体支持物上。

本发明的另一实施方案涉及对核酸进行测序的方法，包括以下步骤：(a)在具有至少 50,000 个离散的反应位点的阵列中提供多个单链核酸模板；(b)使核酸模板与进行与发光偶联的基于焦磷酸的测序反应所必须的试剂接触；(c)检测从光敏装置上各自部分上的多个反应位点发出的光；(d)将撞击光敏器件的每个部分的光转化为电信号，该电信号可以与来自所以其它反应位点的信号区分；以及(e)根据相应的电信号，基于从每个所述离散反应位点的光发射，确定核酸模板的序列。该方法可以进一步包含以下步骤：(a)独特地标记来自不同生物来源文库的断裂核酸，以建立具有不同可检测序列标记的断裂核酸的文库；和(b)对断裂的核酸进行测序，并且检测来自每个标记的核酸片段的可检测序列标记。文库可以单独送递，或文库可以混合并且同时送递。可检测的序列标记可以包含 2 - 50 个碱基的寡核苷酸。

本发明的另一实施方案涉及对核酸进行测序的方法，包括以下步骤：(a)使大模板核酸分子断裂，产生多个断裂的核酸；(b)将多个断裂的核酸的一条链各自连接于珠子，得到各自连接于珠子的单链断裂核酸；(c)将各自连接于珠子的单链断裂核酸的群体送递到平坦表面上的至少 10,000 个反应室的阵列中，其中多个孔包含不超过 1 个具有一个单链断裂核酸的珠子；以及(d)在多个反应室中同时进行测序反应。在该方法中，反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ 。断裂的核酸的长度可以是 30 - 500 个碱基。此外，可以在步骤(d)之前，在反应中扩增断裂的核酸。可以用聚合酶链反应完成扩增。测序反应可以是例如基于焦磷酸的测序反应。作为另一个实例，测序反应可以包括以下步骤：(f)使有效量的测序引物与单链断裂核酸模板退火，并且用聚合酶和预定核苷酸三磷酸延

伸测序引物，得到测序产物，并且，如果预定核苷酸三磷酸掺入所述测序引物的3'末端，则得到测序反应副产物；以及(g)鉴定测序反应副产物，从而确定多个反应室中的核酸的序列。作为另一个实例，测序反应可以包括以下步骤：(a)使两个或更多测序引物与核酸分子的一个或多个单链杂交，其中除一个引物之外的所有引物都是可逆阻断的引物；(b)通过从未阻断的引物进行聚合酶延伸，将至少一个碱基掺入核酸分子；(c)阻止未阻断的引物进一步延伸；(d)使可逆保护的引物之一去保护为未阻断的引物；以及(e)重复步骤(b)-(d)，直到至少一个可逆阻断的引物去保护并且用于确定序列。反应室可以是通过蚀刻光导纤维束的一端而形成的凹陷。

另一方面，本发明提供了对靶核酸进行测序的方法，包括将参照核酸序列转化为参照数目序列，其中参照数目序列包括理想流动图（flowgram）；当一个或多个核苷酸序贯流过一个或多个拷贝的靶核酸片段时检测信号，信号相应于表明所述片段的核酸序列的询问数目序列，其中询问数目序列包括询问流动图；在理想流动图的一个或多个位置上将询问流动图与理想流动图进行比对，其中获得询问流动图与理想流动图的近似匹配；比较询问流动图的数目序列与理想流动图的数目序列，得到质量评分，其中质量评分表明询问流动图与理想流动图之间比对的质量；对多个片段重复前面的步骤；将多个询问流动图锚定在理想流动图上相应于最高质量比对的位置上；在一个或多个询问流动图所覆盖的理想流动图的每个位置上对一个或多个询问流动图的序列数目取平均值，从而产生包含共有区流动图的共有区数目序列；并且将共有区流动图转化为核酸序列。

在一个实施方案中，本发明的方法进一步包括将理想流动图分为预定长度的重叠的理想亚流动图，并且给理想亚流动图做索引。在其它实施方案中，该方法进一步包括将每个询问流动图分为询问亚流动图，每个亚流动图的长度相应于理想亚流动图的预定长度。在其它实施方案中，该方法进一步包括检索做索引的理想亚流动图，以确定将询问流动图与理想流动图进行比对的位置。理想亚流动图、询问亚流动图、或这两者，可以是二进制编码的。

本发明也提供了对靶核酸进行测序的方法，包括当一个或多个核苷酸序贯流过一个或多个拷贝的靶核酸片段时检测信号；将信号与表明片

段的核酸序列的询问数目序列进行关联，其中询问数目序列包含询问流动图；重复前面的步骤，得到多个询问流动图；将多个询问流动图彼此进行比较，以鉴定多个询问流动图之间的重叠区域；在重叠区域比对多个询问流动图；基于比对得到质量评分，其中质量评分表明比对的质量；通过确定具有满足预定阈值的质量评分的比对，鉴定逐对重叠询问流动图；将逐对重叠询问流动图分组为一个或多个统一群（unitig）；对每个统一群内的一个或多个匹配位置中每一个位置上的询问流动图的序列数目取平均值，从而产生包含统一群共有区流动图的共有区数目序列；并且将每个统一群共有区流动图转化为统一群共有核酸序列。

在一个实施方案中，本文描述的一个或多个统一群包含询问流动图的最大重叠的一致链。在另一实施方案中，该方法进一步包括以下步骤：将统一群共有核酸序列彼此比较以鉴定序列重叠；和连接具有共同重叠序列的统一群共有区，从而形成包含重叠群（contig）核酸序列的一个或多个重叠群。在其它实施方案中，本发明的方法进一步包含鉴定每个重叠群内的边界的步骤，其中边界是这样的区域，即统一群序列在此处从共同区域脱离出来；并且在前面步骤中鉴定出的边界上破坏重叠群。在另一实施方案中，该方法进一步包括连接任何两个末端通过相同片段核酸序列而重叠的重叠群的步骤，其中如果鉴定了边界，则任选破坏如此连接的重叠群。在进一步的实施方案中，本发明的方法进一步包括鉴定片段核酸序列和重叠群之间的所有比对的步骤；其中任选在少于4个片段核酸序列进行比对的任何位置破坏重叠群；通过对与在前面步骤中与重叠群比对的片段核酸序列相关的询问流动图的序列数目取平均值，计算重叠群共有区流动图；并且将重叠群共有区流动图转化为重叠群共有核酸序列。该方法可以进一步包括用重叠群共有核酸序列重复前面的步骤，直到基本没有共有区碱基调用改变，从而计算最终重叠群共有序列。

### 附图简述

图1描绘了整个文库制备过程的示意图，包括以下步骤：模板DNA断裂（图1A）、末端磨光（图1B）、衔接子连接（图1C）、切口修复、链延伸和凝胶分离（图1D）。图1E描绘了对模板DNA进行扩增和测序的阶段的示意图（图1E）。图1F描绘了代表性琼脂糖凝胶，其含有

本发明方法的 180 - 350 碱基对腺病毒 DNA 文库的样品制备物。图 1G 描述了文库制备、扩增和测序的详细示意图。

图 2A 描绘了本发明的通用衔接子设计的示意图。每个通用衔接子由两个互补的 ssDNA 寡核苷酸产生，所述寡核苷酸设计为含有用作 PCR 引物的 20 bp 核苷酸序列、用作测序引物的 20 bp 核苷酸序列，和包含非重复核苷酸序列（即 ACGT、CAGT 等）的独特 4 bp 区分序列。图 2B 描绘了用于本发明的代表性通用衔接子序列对。衔接子 A 有义链：SEQ ID NO:1；衔接子 A 反义链：SEQ ID NO:2；衔接子 B 有义链：SEQ ID NO:3；衔接子 B 反义链：SEQ ID NO:4。图 2C 描绘了用于本发明的通用衔接子设计的示意图。

图 3 描绘了本发明的有切口的双链 DNA 片段的链置换和延伸。在连接从合成的寡核苷酸产生的通用衔接子后，将产生双链 DNA 片段，其在 T4 DNA 连接酶处理后含有 2 个有切口的区域（图 3A）。加入链置换酶（即 Bst DNA 聚合酶 I）将结合切口（图 3B），对有切口的链进行链置换，并且完成链的核苷酸延伸（图 3C），从而产生无切口的双链 DNA 片段（图 3D）。

图 4 描绘了根据本发明，采用链霉亲和素包被的珠子分离按方向连接的单链 DNA。用通用衔接子 A 和 B（所述两个不同的衔接子有时称作“第一”和“第二”通用衔接子）连接后，双链 DNA 将含有四种可能的组合的衔接子：AA, BB, AB 和 BA。当通用衔接子 B 含有 5' - 生物素时，用链霉素包被的磁性固体支持物捕获并分离 AB, BA 和 BB 群体（洗掉群体 AA）。BB 群体保持在珠子上，因为双链 DNA 的每端都连接于珠子，并且没有释放。但是，当在低盐缓冲液存在下洗脱时，仅有群体 AB 和 BA 将释放与结合的链互补的单链 DNA 片段。从上清液分离单链 DNA 片段，并且用作后续扩增和测序的模板。该方法在下文更详细描述。

图 5 描绘了 DNA 捕获珠的结构的示意图。

图 6 描绘了珠乳状液扩增过程的一个实施方案的示意图。

图 7 描绘了除去没有连接任何 DNA 的珠子的富集过程的示意图。

图 8A - B 描绘了本发明的双末端测序反应的示意图。

图 9 描绘了本发明的焦磷酸测序装置上的双末端测序示范。

图 10A - F 描绘了示例的双末端测序过程。SEQ ID NO:44:

atgcacatggttgacacagtgg; SEQ ID NO:45: atgcacatggttgacacagtgg; SEQ ID NO:46: atgccaccgacacctgtctcaaactt.

图 11A - D 描绘了采用锚定引物，基于滚环的扩增的示意图。在图片 B 中，序列如下： gacctcacac gatggctgca gctt (SEQ ID NO:71) 和 tcgtgtgagg ttcagcatc ttatgtatat ttacttctat ttcaggtgc ccaagctgca gcc (SEQ ID NO:72)。在图片 C 中，序列如下： gacctcacac gatggctgca gctt (SEQ ID NO:71), acttctattc tcagttgcct aagctgcagc cattgtgtga ggtctcagca tcttatgtat attt (SEQ ID NO:73) 和 gtcctagaat agaagtaaat atacatgctc ga (SEQ ID NO:74)。在图片 D 中，序列如下： gacctcacac gagtagcatg gctgcagctt (SEQ ID NO: 75), tcgtgtgagg ttcagcatc ttatgtatat ttacttctat ttcagttgc ctaagctgca gcc (SEQ ID NO:76) 和 tgctac (SEQ ID NO:77)。

图 12 描绘了本发明的测序装置图。

图 13 描绘了本发明的试剂送递/灌注室的示意图。

图 14 描绘了本发明的有凹陷的光导纤维束的显微图，其称作 PicoTiterPlate<sup>TM</sup>。

图 15 描绘了铺有珠子的微量滴定板的显微图，所述珠子上固定了 DNA 模板，并且固定了硫酸化酶和荧光素酶。

图 16 描绘了试剂流动室和 FOR A (PicoTiterPlate<sup>TM</sup>) 的示意图。

图 17 描绘了本发明的分析仪器的示意图。

图 18 描绘了 PicoTiterPlate<sup>TM</sup> 内的显微平行测序反应的示意图。

图 19 描绘了单孔反应的显微图。

图 20 描绘了 PicoTiterPlate<sup>TM</sup> 加样药筒。“A”表示具有朝向药筒中的微孔的 PicoTiterPlate<sup>TM</sup>，PicoTiterPlate<sup>TM</sup> 孔的开口侧与加样药筒的壁之间的距离是 0.3 mm；“B”表示硅密封垫；“C”表示入口，“D”表示入口加样管；“E”表示出口，“F”表示出口管。所述 PicoTiterPlate<sup>TM</sup> 用塑料夹保持在药筒中。液体通过入口加样管 D 填充，并且通过入口 C 进入 PicoTiterPlate<sup>TM</sup> 孔的开口侧和加样药筒壁之间的空间。由硅密封垫 B 限度的区域是填充的，过量的液体通过出口 E 和出口管 F 离开药筒。

图 21 描绘了 PicoTiterPlate<sup>TM</sup> 扩增室的剖视图。“A”表示具有 6 个保持螺钉的扩增室盖；“B”表示封闭的小室泡沫隔离垫；“C”表示 25 mm × 75 mm 标准显微镜载玻片；“D”表示 0.25 mm 厚的硅片；“E”表示 PicoTiterPlate<sup>TM</sup>；“F”表示扩增室的底；“G”表示第二个 0.25 mm

厚的硅片。

图 22 描绘了固相 PicoTiterPlate<sup>TM</sup> PCR 的示意图。圆柱结构代表各个 PicoTiterPlate<sup>TM</sup> 孔。灰色球体代表具有固定的引物的珠子。正向 “F” (红色) 和反向 “R” (蓝色) 引物以箭头所示的 5' 到 3' 方向表示。与正向和反向引物互补的合成序列如暗红色 (F 互补序列) 和暗蓝色 (R 互补序列) 条表示。单链模板 DNA 如实心灰色线条表示，新合成的 DNA 链如灰色虚线表示。荧光标记的杂交探针如绿色条所示。

图 23A - C 描述了与珠子固定的测试 DNA 片段的荧光探针杂交。图 23A (左上) 和 23B (右下) 说明了分别与固定在对照珠上的片段 A 和片段 B 杂交的混合探针群体的特异性。片段 B 珠显示 Alexa Fluor 647 信号 (红色)，片段 A 珠显示 Alexa Fluor 488 信号 (绿色)。图 23C (下图) 描述了 PTPCR 后来自 DNA 捕获珠的探针荧光。珠子显示同质的片段 A 和片段 B 信号，以及显示为不同程度黄色的模板混合物。

图 24 描述了来自单链 DNA 文库分析的代表性 BioAnalyzer 输出。

图 25 描述了侧翼于 PCR 引物和测序引物的插入片段。

图 26 描述了由位于交叉杂交区 (CHR) 的 PCR 引物错配产生的截短的产物。

图 27 描述了基于解链温度进行的引物候选物的计算。

图 28A - D 描述了用于本发明的方法的喷雾器的组装。喷雾器顶部放置管帽 (图 7A)，并且用喷雾器钳组合件固定该帽 (图 7B)。喷雾器的底部与氮供应连接 (图 7C)，并且整个器件包裹在石蜡膜 (parafilm) 中 (图 7D)。

图 29A 描述了雾化和磨光后，单链 DNA 文库的 LabChip 分析的代表性结果。

图 29B 描述了雾化、磨光和凝胶纯化后，衔接子连接的单链 DNA 文库的代表性大小分布结果。

图 30 描述了用于将管保持在垂直注射器泵下面的搅拌板上的夹具。夹具经过改进，以支持三组珠乳状液扩增反应混合物。在注射器中加入 PCR 反应混合物和珠子。

图 31 描述了垂直注射器泵中注射器的最佳放置，和注射器出口下的乳状液管的方向。

图 32 描述了靠着注射器泵的注射器泵推动器块的最佳放置，以及

搅拌板上夹具的最佳方向。采用这种排列，将注射器内容物排出到搅拌的乳状液油中。

图 33 描述了按照本发明的方法悬浮于各个微反应器中的珠子（参见箭头）。

图 34 描述了双末端测序结果，显示确定了 DNA 模板的两个末端的序列。SEQ ID NO:44: atgcacatgggtgacacagtgg; SEQ ID NO:45: atgcacatgggtgacacagtgg; SEQ ID NO:46: atgccaccgaccctagtctcaaactt。

图 35 说明了用于双链测序的、包含 2 个寡核苷酸序列的珠子的封装。

图 36 说明了液相 PCR 和驱动到珠子程序 - 双末端测序的一个优选实施方案中的一个步骤。

图 37 说明了乳状液破裂和回收珠子上的扩增的模板 DNA - 双末端测序的一个优选实施方案中的一个步骤。

图 38 说明了双末端测序的一个优选方法的示意图。

图 39 说明了金黄色葡萄球菌基因组的测序结果。

图 40 说明了一个涉及双末端测序的实验中的平均读出结果长度。

图 41 说明了双末端测序实验中针对每个基因组跨度的孔的数目。

图 42 说明了双末端测序程序的典型输出和比对串。从上到下按顺序显示序列：SEQ ID NO:47 - SEQ ID NO:60。

图 43 样品制备。(A)顺时针从左上开始：(i)分离基因组 DNA，断裂，连接于衔接子并且分离为单链；(ii)在有利于每个珠子一个片段的条件下使片段与珠子结合，将珠子捕获在油包 PCR 反应乳状液的液滴中，在每个液滴中发生 PCR 扩增，使每个珠子携带 1 千万个拷贝的独特 DNA 模板；(iii)乳状液破裂，使 DNA 链变性，将携带单链 DNA 克隆的珠子沉积在光导纤维截面的孔中；(iv)将携带焦磷酸测序所需的固定的酶的较小的珠子沉积在每个孔中。(B) 显示含有珠子的液滴和空液滴的乳状液的显微镜照片。细箭头指向 28 $\mu\text{m}$  的珠子，粗箭头指向大约 100 $\mu\text{m}$  的液滴。(C)光导纤维截面的一部分的 SEM 照片，显示珠子沉积前的光导纤维覆层和孔。

图 44 测序仪。该测序仪由以下主要子系统组成：射流组合件 (A)，包括含孔的光导纤维截面的流动室 (B)，基于 CCD 照相机的成像组合件 (C) 和提供必要的用户界面和仪器控制的计算机。

图 45：来自生殖道支原体运行的 113 个碱基的读出结果的流动图。核苷酸以 T, A, C, G 的顺序流动。序列示于流动图上。相应于各个均聚物的信号值间距示于右边。最前面的四个碱基（红色，在流动图上面）构成“关键”序列，用于鉴定含有携带 DNA 的珠子的孔。

图 46 生殖道支原体数据。(A)生殖道支原体测序运行的 306,178 个高质量读出结果的读出结果长度分布。该分布反映了各个测序模板的碱基组成。(B)单读出结果水平的平均读出结果准确度，作为针对同一次运行的 238,066 个作图的读出结果的碱基位置的函数。

图 47：非磷酸化的 A 和 B 衔接子连接于磷酸化的、磨光的、双链基因组 DNA 片段末端。A 和 B 衔接子的核苷酸序列以及 B 衔接子上 5' 生物素标记的存在均不同。切口存在于每个衔接子的 3' 接合处，通过 Bst DNA 聚合酶的链置换活性填充文库片段。链霉亲和素 - 生物素相互作用用于除去纯合衔接子组(A/A 和 B/B)侧翼的片段，并且用于制备单链文库模板。片段与链霉亲和素珠结合；洗去未结合的材料（包含纯合 A/A 衔接子组，其缺乏生物素）。然后使固定的片段变性；B/B 片段的两条链都通过生物素化的 B 衔接子保持固定，而 A/B 片段被洗下来，并且用于随后的测序步骤。观察到重复的文库制备物产生基因组的覆盖和过量取样，CV 为 5% 或更少。

图 48 雾化的 DNA 样品的大小分布。尖锐的侧峰是上和下参照标志物。

图 49：单孔的动力学建模。假定：每个珠子 1 千万个 DNA 拷贝， $[DNA] = 0.3 \mu M$ 。

图 50：化学干扰建模。 $t=0$ ,  $[DNA]_{孔1} = 0.3 \mu M$ ,  $[DNA]_{孔2} = 0$ 。

图 51：对 6 个测试片段的混合物进行测序中的详细错误率，作为均聚物长度的功能。参照测序的单碱基总数计算单碱基错误率。对于每种均聚物，参照测序的属于该长度的均聚物的碱基的总数计算错误率。

图 52：阴性和阳性流动的信号强度的典型直方图。

图 53：归因于各种均聚物的流动信号的平均值，其用于文章中讨论的生殖道支原体运行的作图的读出结果。

图 54：对生殖道支原体文库进行测序的详细错误率，作为均聚物长度的函数。对于测试片段，参照测序的单碱基总数计算单碱基错误率；对于均聚物，参照测序的属于每个长度的均聚物的碱基的总数计算错误

率。示出了各个读出结果的错误率，并且在用所有读出结果形成共有序列之后，无 Z - 评分限制。

图 55：作为生殖道支原体运行的基因组位置的函数的覆盖深度。分离区的略微较低的覆盖是由于存在作图中排除的重复区。

图 56：空肠弯曲菌的测序运行中预测和观察到的质量评分的关联（数据未示出）。

图 57：成对末端读出结果的读出结果长度。注意这是对 21 轮运行进行的，所以平均长度与较小数目循环相当。

### 发明详述

本文描述了一个新的平台，其允许同时以低至 39.5 微升体积的 30 万个离散的 PCR 反应(PTPCR)。可以通过洗涤步骤回收来自整个反应的合并的 PTPCR 产物，并且通过实时 PCR 测定特异性模板的存在与否。更令人感兴趣的是，本文中显示了可以驱动这些 PTPCR 产物到达固体支持物，并且通过与两色荧光探针杂交而进行检测，使得能够进行高容量、固相、克隆 DNA 扩增和大规模平行测序。

本发明涉及用于实施基因组测序的方法和装置，其满足以下目标：(1)以快速和有效的方式制备核酸（如基因组），用于进行测序，(2)以代表性方式扩增核酸，和(3)仅仅用一次引物杂交进行多个测序反应。本发明特别适于以经济的方式从核酸的小样品进行基因型分析、检测和诊断。这些目标列于下文。

#### 定义：

除非另外指出，本文用到的所有技术和科学术语具有本发明所属领域普通技术人员通常理解的相同含义。类似或等同于本文描述的方法和材料的那些方法和材料可以用于实施本发明，并且下文描述了示例的合适方法和材料。例如，可以描述包括两个以上步骤的方法。在所述方法中，并不是所有步骤都是实现确定目的所必须的，并且本发明预想采用分离的步骤来实现这些不同的目的。在此全文引入本文提到的所有公开文献、专利申请、专利和其它参考文献作为参考。此外，材料、方法和实施例仅仅是说明性的，不意欲作为限制。

如本文中用到的，术语“通用衔接子”是指经过设计而包含用作 PCR 引物的核苷酸序列和用作测序引物的核苷酸序列的两个互补和退火的

寡核苷酸。任选地，通用衔接子可以进一步包含独特的区分关键序列，其包含非重复的核苷酸序列（即 ACGT, CAGT 等）。一组通用衔接子包含两个独特和不同的双链序列，它们可以连接于双链 DNA 的末端。因此，可以将相同的通用衔接子或不同的通用衔接子连接于 DNA 分子的任一末端。当包含在单链的较大 DNA 分子时，或当作为寡核苷酸存在时，通用衔接子可以称作单链通用衔接子。

“靶 DNA”应该表示这样的 DNA，其序列要通过本发明的方法和装置确定。

结合对应该表示一对分子，其通过依赖于涉及的分子的三维结构的特异性非共价相互作用而进行相互作用。典型的特异性结合配偶体对包括抗原 - 抗体、半抗原 - 抗体、激素 - 受体、核酸链 - 互补核酸链、底物 - 酶、底物类似物 - 酶、抑制剂 - 酶、碳水化合物 - 凝集素、生物素 - 亲和素和病毒 - 细胞受体。

如本文中用到的“区分关键序列”是指由四种脱氧核糖核苷酸（即 A, C, G, T）中的至少一种组成的序列。相同的区分序列可以用于 DNA 片段的整个文库。或者，不同的区分关键序列可以用于对来源于不同生物的 DNA 片段的文库进行示踪。

如本文中用到的，术语“多个分子”是指从相同来源分离的 DNA，从而可以通过相同方法分开制备不同的生物。在一种实施方案中，多个 DNA 样品来源于 DNA 的大片段、完整基因组 DNA、cDNA、病毒 DNA 或来自病毒 RNA 的逆转录物。该 DNA 可以来源于任何来源，包括哺乳动物（即人、非人灵长类动物、啮齿类动物或犬科动物）、植物、鸟类、爬行动物、鱼类、真菌、细菌或病毒。

如本文中用到的，术语“文库”是指由单 DNA 模板产生的一个亚组的较小 DNA 种类，所述模板是分段的或完整的基因组。

如本文中用到的，“独特 PCR 引物区”中的“独特”是指不存在于要扩增或测序的 DNA 分子内，或以非常低的拷贝水平存在于要扩增或测序的 DNA 分子内的序列。

如本文中用到的，术语“相容的”是指衔接子分子可以连接的双链 DNA 的末端（即平端或粘端）。

如本文中用到的，术语“断裂”是指较大 DNA 分子转化为较小 DNA 片段的过程。

如本文中用到的，“大模板 DNA”是超过 25kb，优选超过 500kb，更优选超过 1MB，最优选 5MB 或更大的 DNA。

如本文中用到的，“严格杂交条件”是指仅有互补序列能彼此杂交的条件。

本文描述的发明通常用于处理核酸的系统和方法。该系统和方法可以用于以多种利用核酸测序的途径来处理核酸。可以进行所述测序，以确定核酸序列的同一性，或用于核酸片段中的单核苷酸多态性检测，用于核酸表达谱分析（比较两种或多种状态之间的核酸表达谱，例如比较病变和正常组织，或比较未处理组织和用药物、酶、放疗或化疗处理的组织），用于单倍型分析（比较人受试者中存在的两个等位基因中每一个等位基因上的基因或基因变异），用于核型分析（怀孕前诊断比较典型地来自胚胎/胎儿的测试组织中的一种或多种基因，以检测先天缺陷 - 与来自“正常”核型受试者的相同基因进行比较），以及用于基因型分析（比较一个物种的第一个体中的一个或多个基因与相同物种的其它个体中的相同基因）。

该系统具有一些组件。包括(1)要处理的核酸模板，(2)用于容纳核酸模板的微量滴定板，(3)流动室和流体送递工具，其允许核酸处理试剂在核酸模板上流动，其中随着核酸被处理，处理试剂产生光，(4)光捕获工具，其检测随着核酸被处理而发射的光，并且将捕获的光转化为数据，和(5)数据处理工具，其处理数据，产生关于处理的核酸的有意义的信息。下文将详细讨论该系统的每个上述组件。

## 1. 核酸模板及其制备

### 核酸模板

可以根据本发明进行测序的模板，例如核酸文库，一般可以包括开环或闭环核酸分子。“闭环”是共价闭合的环状核酸分子，如环状 DNA 或 RNA 分子。“开环”是线性单链核酸分子，其具有 5' 磷酸基团和 3' 羟基。

在一种实施方案中，单链核酸含有至少 100 个拷贝的特定核酸序列，每个拷贝末端与末端共价连接。在一些实施方案中，开环是从线性双链核酸分子原位形成。给定开环核酸分子的末端可以通过 DNA 连接酶连接。开环分子的 5' 和 3' 末端的序列与第二核酸分子中相邻核苷酸的两个

区域，如锚定引物的衔接子区（有时称作衔接子）互补，或与第二 DNA 分子中几乎邻接的两个区域互补。因此，可以用 DNA 连接酶连接开环分子的末端，或通过 DNA 聚合酶在缺口填充反应中延伸。开环详细描述于 Lizardi, 美国专利 No. 5,854,033，在此全文引入作为参考。在例如开环退火于锚定引物后，在 DNA 连接酶（用于 DNA）或 RNA 连接酶存在下开环可以转化为闭环。

如果需要，核酸模板可以提供为扣锁探针。扣锁探针是线性寡核苷酸，其包括位于每个末端的靶 - 互补序列，并且通过连接序列分开。连接序列可以连接于已经例如用限制性内切酶物理剪切或消化的核酸序列文库的成员的末端。与靶序列杂交后，使这些线性寡核苷酸的 5' 和 3' 末端区域并列。这种并列使得这两个探针片段（如果合适杂交）能够由酶促连接（如用 T4 DNA 连接酶）进行共价结合，由此将探针转化为闭环分子，所述分子连接于特异性靶序列（参见例如 Nilsson, et al., 1994. Science 265: 2085 – 2088）。得到的探针适于同时分析很多基因序列，这是由于它们对基因序列变体的特异性和选择性（参见例如 Lizardi, et al., 1998. Nat. Genet. 19: 225 – 232; Nilsson, et al., 1997. Nat. Genet. 16: 252 – 255），以及由于得到的反应产物保持定位于特异性靶序列。此外，预期很多不同探针的分子内连接相对于基于多路 PCR 的方法较不易发生非特异性交叉反应性，在基于多路 PCR 的方法中，非相关引物对可以产生无关扩增产物（参见例如 Landegren and Nilsson, 1997. Ann. Med. 29: 585 – 590）。

可以构建包含单链或双链核酸分子的起始核酸模板文库，前提是核酸序列包括可以用于退火（如果在文库中存在），或可以使其用于退火于锚定引物序列的区域。例如，当用作滚环扩增的模板时，双链模板的一个区域需要至少瞬时是单链的，以便作为延伸锚定引物的模板。

文库模板可以包括多个元件，包括但不限于一个或多个与锚定引物互补的区域。例如，模板文库可以包括与测序引物互补的区域、控制核苷酸区域和包含随后要表征的测序模板的插入序列。如下文更详细解释的，控制核苷酸区域用于校准副产物量和掺入的核苷酸数目之间的关系。本文用到的术语“互补序列”是指能够与特异性核苷酸序列杂交形成匹配的双链体的核苷酸序列。

在一种实施方案中，文库模板包括：(i)两个与锚定引物互补的独特

的区域，(ii)一个与测序引物同源的区域，(iii)一个任选的控制核苷酸区域，(iv)要测序的插入序列，其具有例如 30 - 500, 50 - 200 或 60 - 100 个核苷酸。当然，模板可以包括这些特征中的 2 个、3 个或全部 4 个。

可以从任何核酸来源，如任何细胞、组织或生物构建模板核酸，并且可以通过任何本领域公知的方法制备。合适的方法包括例如基因组 DNA 的超声处理和用一种或多种限制性核酸内切酶 (RE) 消化，从最初的核酸分子群体产生需要的长度范围的片段。优选地，一种或多种所述限制酶具有独特的四碱基识别序列。所述酶的实例包括例如 Sau3A1, MspI 和 TaqI。优选地，酶与锚定引物一起使用，该引物具有含有相应限制酶的识别序列的区域。在一些实施方案中，锚定引物的一个或所有两个衔接子区含有额外的连接已知限制酶识别序列的序列，从而能够捕获或退火于感兴趣的特定限制片段的锚定引物。在其它实施方案中，限制酶与 IIS 型限制酶一起使用。

或者，可以通过从 RNA，例如信使 RNA(mRNA) 制备互补 DNA(cDNA) 文库，制备模板文库。如果需要，cDNA 文库可以进一步用限制性内核酸内切酶处理，以获得特定 RNA 的特征性 3' 末端、内部片段或包括分离 RNA 的 3' 末端的片段。锚定引物中的衔接子区可以与认为存在于模板文库中的感兴趣的序列互补，所述感兴趣的序列例如通过内切核酸酶消化产生的片段内的已知或怀疑的序列多态性。

在一种实施方案中，可以使索引寡核苷酸与模板文库的成员连接，使得能够随后将模板核酸与产生模板核酸的核酸群体进行关联。例如，可以采用任何以前公开的方法（例如限制消化、超声处理）分开使一个或多个起始 DNA 群体的样品断裂。每个样品的特异性索引寡核苷酸序列附着于，例如连接于断裂的群体的成员的末端。索引寡核苷酸可以作为环化、扩增和任选作为测序的区域，其允许它用于对核酸做索引或编码，从而从来源样品中鉴定起始样品。

用多种可区分的索引引物制备的独特模板文库可以混合在一起，用于随后的反应。确定文库成员的序列，允许鉴定相应于索引寡核苷酸的序列。基于该信息，给定片段的来源可以推断出来。

本发明包括样品制备过程，其得到固体或可移动的固体基质阵列，其中含有多个与模板核酸连接的锚定引物或衔接子。

当模板核酸是环状时，优选通过以下方法形成共价连接的锚定引物

和一个或多个拷贝的靶核酸：使锚定引物退火于环状核酸的互补区，然后用聚合酶延伸退火的锚定引物，导致形成含有一个或多个拷贝的与环状核酸互补的序列的核酸。

锚定引物与固体或可移动的固体基质的连接可以在延伸退火的锚定引物之前、期间或之后发生。因此，在一种实施方案，一种或多种锚定引物与固体或可移动的固定基质连接，此后锚定引物退火于靶核酸，并且在聚合酶存在下延伸。或者，在第二个实施方案中，锚定引物首先与靶核酸退火，用聚合酶延伸退火的锚定引物的3'OH末端。然后将延伸的锚定引物连接于固体或可移动的固体基质。通过改变锚定引物的序列，可以特异性扩增核酸群体中存在的独特靶核酸。

下文概括了制备用于扩增和测序反应的模板核酸的优选实施方案。本发明包括制备样品DNA的方法，包括7个概括步骤：(a)使大模板DNA或完整基因组DNA样品断裂，产生多个消化的DNA片段；(b)在所述多个消化的DNA样品上建立相容的末端；(c)将一组通用衔接子序列连接到断裂的DNA分子末端，制备多个衔接子连接的DNA分子，其中每个通用衔接子序列具有已知和独特的碱基序列，其包含共同PCR引物序列、共同测序引物序列和区别的四碱基关键序列，并且其中一个衔接子连接于生物素；(d)分开和分离所述多个连接的DNA片段；(e)除去所述多个连接的DNA片段的任何部分；(f)切口修复和链延伸所述多个连接的DNA片段；(g)将每个所述连接的DNA片段连接于固体支持物；和(h)分离包含单链的衔接子连接的DNA片段的群体，对于所述DNA片段，每个末端具有独特的衔接子（即提供方向性）。

以下讨论概括了包含在本发明的方法中的基本步骤。但是，如本领域技术人员可以理解的，虽然以特定顺序描述各个步骤，但可以调整步骤的顺序达到相同的结果。这种调整是发明人考虑的。此外，一些步骤可以减少到最少，这也是本领域技术人员可以了解的。

## 断裂

在本发明方法的实施中，可以通过本领域技术人员已知的方法进行DNA样品的断裂。优选地，通过酶促或机械手段进行断裂。机械手段可以是超声处理或物理剪切。可以通过用核酸酶（如脱氧核糖核酸酶I(DNA酶I)）或一种或多种限制性内切核酸酶消化，进行酶促手段。

在一种优选实施方案中，断裂得到序列未知的末端。

在一种优选实施方案中，酶促工具是 DNA 酶 I。DNA 酶 I 是一种通用酶，其非特异性切割双链 DNA(dsDNA)，以释放 5' - 磷酸化的二核苷酸、三核苷酸和寡核苷酸产物。DNA 酶 I 在含有  $Mn^{2+}$ 、 $Mg^{2+}$  和  $Ca^{2+}$ ，但是不含其它盐的缓冲液中具有最佳活性。DNA 酶 I 消化步骤的目的是使大 DNA 基因组断裂为包含文库的较小的种类。DNA 酶 I 的切割特征导致在基于锰的缓冲液存在下使用时模板 DNA 的随机消化（即没有序列偏倚）和平端 dsDNA 片段的优势(Melgar, E. and D.A. Goldthwait. 1968. Deoxyribonucleic acid nucleases. II. The effects of metal on the mechanism of action of deoxyribonuclease I. J. Biol. Chem. 243: 4409)。DNA 酶 I 处理基因组模板后产生的消化产物的范围依赖于三种因素：i)使用的酶量(单位)；ii)消化的温度( $^{\circ}C$ )；和 iii)温育时间(分钟)。下文概括的 DNA 酶 I 消化条件经过优化，产生大小范围是 50 – 700 碱基对 (bp) 的基因组文库。

在一种优选实施方案中，DNA 酶 I 消化大模板 DNA 或完整基因组 DNA 1 – 2 分钟，产生多核苷酸群体。在另一优选实施方案中，DNA 酶 I 消化在 10  $^{\circ}C$  – 37  $^{\circ}C$  的温度下进行。在另一优选实施方案中，消化的 DNA 片段长度是 50 bp – 700 bp。

## 磨光

在  $Mn^{2+}$  存在下用 DNA 酶 I 消化基因组 DNA (gDNA) 模板，将产生平端 DNA 片段，或具有长度为 1 个或 2 个核苷酸的突出末端的 DNA 片段。在一种优选实施方案中，用 Pfu DNA 聚合酶产生了数量增加的平端。在其它实施方案中，可以用效力较低的 DNA 聚合酶，如 T4 DNA 聚合酶或克列诺 DNA 聚合酶产生平端。Pfu “磨光” 或平端化用于增加用 DNA 酶 I 进行基因组模板消化后产生的平端种类的量。用 Pfu DNA 聚合酶进行片段磨光，将导致 5' 突出端的填充。此外，Pfu DNA 聚合酶不展示 DNA 延伸酶活性，但是具有 3' → 5' 外切核酸酶活性，将导致去除单核苷酸和双核苷酸延伸，以进一步增加可用于衔接子连接的平端 DNA 片段的量(Costa, G.L. and M.P. Weiner. 1994a. Protocols for cloning and analysis of blunt - ended PCR - generated DNA fragment s. PCR Methods Appl 3(5):S95; Costa, G.L., A. Grafsky and M.P. Weiner. 1994b.

Cloning and analysis of PCR - generated DNA fragments. PCR Methods Appl 3(6):338; Costa, G.L. and M.P. Weiner. 1994c. Polishing with T4 or Pfu polymerase increases the efficiency of cloning of PCR products. Nucleic Acids Res. 22(12):2423).

### 衔接子连接

如果核酸文库将连接于固体基质，则优先选用公知的技术将核酸模板退火于锚定引物序列（参见，例如 Hatch, et al., 1999. Genet. Anal. Biomol. Engineer. 15: 35 – 40; Kool, 美国专利 No. 5,714,320 and Lizardi, 美国专利 No. 5,854,033）。概言之，只要导致锚定引物序列的衔接子区和存在于模板文库中的序列之间的特异性，即完美或几乎完美互补的形式，则任何使锚定引物与模板核酸序列退火的程序都是合适的。

在一种优选实施方案中，在DNA文库断裂和平端化之后，在每个DNA片段中加入通用衔接子序列。通用衔接子设计为包括一组独特的PCR引物区，其长度典型是20 bp，邻接一组独特的测序引物区，该测序引物区长度典型是20 bp，其后任选是由四种脱氧核糖核苷酸（即A, C, G, T）中的至少一种组成的独特的区别关键序列。在一种优选实施方案中，区别关键序列的长度是4个碱基。在另一实施方案中，区别关键序列可以是1–4个碱基的组合。在另一实施方案中，每种独特通用衔接子的长度是四十四 bp (44 bp)。在一种优选实施方案中，用T4 DNA连接酶将通用衔接子连接到DNA片段的每个末端上，以便总共在每个DNA片段上添加88 bp。对于每次DNA文库制备，特别设计了不同的通用衔接子，因此对每种生物提供了独特标识符。通用衔接子的大小和序列可以进行调整，这是本领域技术人员可以理解的。

例如，为了制备两种独特的通用衔接子（即“第一”和“第二”），可以从销售商（即 Integrated DNA Technologies, IA 或 Operon Technologies, CA）订购单链寡核苷酸。在一种实施方案中，在合成过程中调整通用衔接子寡核苷酸序列，其中用两个或三个硫代磷酸酯键代替5'和3'末端的磷酸二酯键。未修饰的寡核苷酸通过核酸酶进行快速降解，因此用途受限。核酸酶是通过切割核苷酸碱基之间的磷酸二酯键而催化多核苷酸的水解裂解的酶。因此，用于寡核苷酸应用中的一种简单和广泛使用的核酸酶抗性化学是硫代磷酸酯修饰。在硫代磷酸酯中，硫原子

代替了寡核苷酸主链中的非桥接氧，使寡核苷酸主链抗所有形式的核酸酶消化（即抗内切核酸酶和外切核酸酶消化）。对每个寡核苷酸进行HPLC纯化，以确保合成的寡核苷酸制备中不存在污染或多余的寡核苷酸。设计通用衔接子，以便能够有方向地连接于平端的断裂DNA。每组双链通用衔接子设计为具有PCR引物区，其含有非互补5'四碱基突出端，所述突出端不能连接于平端DNA片段并且防止这些末端彼此连接。因此，结合仅仅能够发生在衔接子的3'末端和DNA片段的5'末端之间，或DNA片段的3'末端和衔接子的5'末端之间。通过使用设计具有主要使互补寡核苷酸退火，并且防止两个非互补寡核苷酸之间交叉杂交的序列的单链寡核苷酸，产生双链通用衔接子序列。在一种实施方案中，95%的通用衔接子从互补寡核苷酸的退火形成。在一种优选实施方案中，97%的通用衔接子从互补寡核苷酸的退火形成。在一种更优选的实施方案中，99%的通用衔接子从互补寡核苷酸的退火形成。在一种最优选的实施方案中，100%的通用衔接子从互补寡核苷酸的退火形成。

两个衔接子之一可以连接于支持物结合部分。在一种优选实施方案中，将5'生物素连接于第一个通用衔接子，使得随后能够分离ssDNA模板，并且能够将通用衔接子非共价偶联于用生物素结合蛋白（即链霉亲和素、中性亲和素(neutravidin)或亲和素）饱和的固体支持物表面。其它键也是本领域公知的，并且可以用于代替生物素-链霉亲和素（例如抗体/抗原-表位、受体/配体和寡核苷酸配对或互补）。在一种实施方案中，固体支持物是珠子，优选聚苯乙烯珠。在一种优选实施方案中，珠子的直径是约2.8 μm。如本文中用到的，该珠子称作“样品制备珠”。

可以通过使两个ssDNA寡核苷酸混合和退火而制备，其中一个含有有义序列，另一个含有反义(互补)序列。通用衔接子设计的示意图概括于图2。

### 分离连接产物

通用衔接子连接导致形成一端具有衔接子的断裂DNAs、未结合的单衔接子，和衔接子二聚体。在一种优选实施方案中，琼脂糖凝胶电泳用作从未连接的单衔接子和衔接子二聚体群体中分开和分离衔接的DNA文库群体的方法。在其它实施方案中，可以通过大小排阻层析或蔗糖沉降，分离片段。DNA的DNA酶I消化程序典型地产生50-700 bp

的文库群体。在一种优选实施方案中，在 DNA 标志物存在下进行琼脂糖凝胶电泳时，加入 88 bp 通用引物衔接子组，将使 DNA 文库群体转变为更大的大小，并且导致大小范围为大约 130 – 800 bp 的迁移谱；衔接子二聚体将以 88 bp 迁移；未连接的衔接子将以 44 bp 迁移。因此，可以从琼脂糖凝胶中物理分离大小范围是 200 – 800 bp 的许多双链 DNA 文库，并且用标准提取技术纯化。在一种实施方案中，衔接子连接的 DNA 文库的凝胶分离将导致大小为 200 – 400 bp 的文库群体的回收。其它区分衔接子连接的片段的方法是本领域技术人员已知的。

### 切口修复

由于用于通用衔接子的 DNA 寡核苷酸不是 5' 磷酸化的，在连接酶处理后，断裂的 DNAs 的 3' 接合处将存在缺口（参见图 3A）。可以使用能够结合、链置换和延伸有切口的 DNA 片段的 DNA 聚合酶，填充这些“缺口”或“切口”。缺乏 3' → 5' 外切核苷酸活性，但是具有 5' → 3' 外切核酸酶活性的 DNA 聚合酶具有识别切口、置换有切口的链和以导致切口修复的形式和形成未切口的双链 DNA 的方式延伸链的能力（参见图 3B 和 3C）(Hamilton, S.C., J.W. Farchaus and M.C. Davis. 2001. DNA polymerases as engines for biotechnology. BioTechniques 31:370)。

将一些修饰酶用于切口修复步骤，包括但不限于聚合酶、连接酶和激酶。可以用于此应用的 DNA 聚合酶包括例如大肠杆菌 DNA pol I、热硫化氢热厌氧杆菌 pol I 和噬菌体 phi 29。在一种优选实施方案中，链置换酶热硫化氢热厌氧杆菌 pol I (Bst DNA 聚合酶 I) 用于修复有切口的 dsDNA，得到未切口的 dsDNA（参见图 3D）。在另一种优选实施方案中，连接酶是 T4，并且激酶是多核苷酸激酶。

### 分离单链 DNA

产生未切口的 dsDNA 后，分离包含第一和第二衔接子分子的 ssDNAs（需要的群体在下文用星号表示；“A”和“B”相应于第一和第二衔接子）。双链 DNA 文库将具有按以下配置结合的衔接子：

1. 通用衔接子 A - DNA 片段 - 通用衔接子 A
2. 通用衔接子 B - DNA 片段 - 通用衔接子 A\*
3. 通用衔接子 A - DNA 片段 - 通用衔接子 B\*

#### 4. 通用衔接子 B - DNA 片段 - 通用衔接子 B

设计通用衔接子，使得仅仅一个通用衔接子具有 5' 生物素部分。例如，如果通用衔接子 B 具有 5' 生物素部分，则可以将链霉亲和素包被的样品制备珠用于使所有双链 DNA 文库种类与通用衔接子 B 结合。含有两个通用衔接子 A 种类的基因组文库群体将不含 5' 生物素部分，并且不与含链霉亲和素的样品制备珠结合，由此可以洗掉。仅有的保持与珠子连接的种类是具有衔接子 A 和 B 的那些，以及具有两个通用衔接子 B 序列的那些。具有两个通用衔接子 B 序列的 DNA 种类（即每个 5' 末端具有生物素部分）将在每个末端结合于链霉亲和素包被的样品制备珠，因为双链中包含的每条链都将被结合。具有通用衔接子 A 和通用衔接子 B 的双链 DNA 种类将含有单个 5' 生物素部分，因此仅仅在一个末端结合于链霉亲和素包被的珠子。样品制备珠是磁性的，因此，当磁化时，样品制备珠将保持与固体支持物偶联。因此，在低盐（“解链”或变性）溶液存在下，仅有含单个通用衔接子 A 和单个通用衔接子 B 序列的 DNA 片段将释放互补的未结合链。可以收集这种单链 DNA 群体，并且通过例如焦磷酸测序、实时定量 PCR、琼脂糖凝胶电泳或毛细凝胶电泳进行定量。

#### 模板与珠的连接

在一种实施方案中，对根据本发明的方法制备的 ssDNA 文库进行定量，以计算每单位体积的分子数目。使这些分子与固体支持物（珠）退火，所述支持物含有与 ssDNA 种类的通用衔接子末端的 PCR 引物区互补的寡核苷酸捕获引物。然后使珠子转移到扩增方案。然后可以对捕获在 DNA 珠上的单个种类的克隆群体进行测序。在一种实施方案中，固体支持物是珠子，优选琼脂糖珠。如本文中用到的，该珠子称作“DNA 捕获珠”。

此处用到珠子可以是任何方便的大小并且可以由任何数目的已知材料制作。所述材料的实例包括：无机物、天然聚合物和合成聚合物。这些材料的具体实例包括：纤维素、纤维素衍生物、丙烯酸树脂、玻璃；硅胶、聚苯乙烯、明胶、聚乙烯吡咯烷酮、乙烯和丙烯酰胺的共聚物、与二乙烯基苯交联的聚苯乙烯等（参见 Merrifield Biochemistry 1964, 3, 1385 – 1390）、聚丙烯酰胺、乳胶凝胶、聚苯乙烯、右旋糖苷、橡胶、硅、

塑料、硝酸纤维素、纤维素、天然海绵、硅胶、玻璃、金属塑料、纤维素、交联右旋糖昔(例如 Sephadex<sup>TM</sup>)和琼脂糖凝胶(Sepharose<sup>TM</sup>)以及本领域技术人员已知的固相支持物。在一种实施方案中，DNA 捕获珠的直径是 20 - 70μm。在一种优选实施方案中，DNA 捕获珠的直径是 20 - 50μm。在一种更优选的实施方案中，DNA 捕获珠的直径是约 30μm。

一方面，本发明包括制备固体支持的文库的方法，包括：(a)根据本文公开的方法制备 ssDNA 模板的群体；(b)将每个 DNA 模板连接于固体支持物，使得每个固体支持物存在一个分子的 DNA；(c)扩增单链模板的群体，使得扩增在每个固体支持物上产生每个 DNA 片段的克隆群体；(d)对珠子的克隆群体进行测序。

在一种实施方案中，固体支持物是 DNA 捕获珠。在另一实施方案中，DNA 是基因组 DNA、cDNA 或病毒 RNA 的逆转录物。DNA 可以连接于固体支持物，例如，通过生物素 - 链霉亲和素键、共价键或通过互补寡核昔酸杂交而连接。在一种实施方案中，每个 DNA 模板与一组通用衔接子连接。在另一实施方案中，通用衔接子对包含共同 PCR 引物序列、共同测序引物序列和区分关键序列。分离单链 DNA，其提供独特末端；然后将单链分子连接于固体支持物，并且暴露于用于群体的克隆扩充的扩增技术。DNA 可以通过 PCR 扩增。

另一方面，本发明提供了通过下文描述的方法制备的固体支持物的文库。

通过该方法制备的核酸模板（如 DNA 模板）可以用于很多分子生程序，例如线性延伸、滚环扩增、PCR 和测序。这种方法可以通过例如使用高摩尔比的珠：DNA，在连接反应中实现。单链 DNA 分子的捕获将按照泊松分布，并且将导致一个亚组的珠子不连接 DNA，而一个亚组的珠子连接两个分子的 DNA。在一种优选实施方案中，一个珠子连接一个分子的 DNA。此外，可以在衔接子中引入能够用于分离文库的额外操作的额外成分。

## 2. 核酸模板扩增

为了根据本发明的方法对核酸模板进行测序，必须扩增拷贝数，产生足够数目的模板拷贝，以产生通过光检测方法可以检测的信号。可以使用任何合适的核酸扩增方法。

已经描述了许多体外核酸扩增技术。这些扩增方法可以区分为以下方法：(i)需要温度循环 - 聚合酶链反应 (PCR) (参见例如 Saiki, et al., 1995. *Science* 230: 1350 – 1354), ligase chain reaction (参见例如 Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189 – 193; Barringer, et al., 1990. *Gene* 89: 117 – 122)和基于转录的扩增(参见例如 Kwok, et al., 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173 – 1177)的方法，和(ii)等温扩增系统 - 自我持续、序列复制 (参见例如 Guatelli, et al., 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874 – 1878);QB 复制酶系统 (参见例如 Lizardi, et al., 1988. *BioTechnology* 6: 1197 – 1202); 链置换扩增( *Nucleic Acids Res.* 1992 Apr 11;20(7):1691 – 6 )，以及描述于 *PNAS* 1992 Jan 1;89(1):392 – 6; 和 *NASBA J Virol Methods.* 1991 Dec;35(3):273 – 86 中的方法。

在一种实施方案中，使用等温扩增。等温扩增也包括基于滚环的扩增(RCA)。RCA 讨论于例如 Kool, 美国专利 No. 5,714,320 和 Lizardi, 美国专利 No. 5,854,033; Hatch, et al., 1999. *Genet. Anal. Biomol. Engineer.* 15: 35 – 40。RCA 的结果是从锚定引物的 3'末端延伸的单链 DNA (因此连接于固体支持物基质)，并且所述单链 DNA 包括含有多个拷贝的与引物序列退火的环状模板的多联体。典型地，可以用 RCA 获得 1,000 – 10,000 个或更多个拷贝的环状模板，每个模板具有例如大约 30 – 500, 50 – 200 或 60 – 100 个核苷酸的大小范围。

环状核酸分子与锚定引物退火后的 RCA 扩增产物系统地示于图 11A。环状模板核酸 102 与锚定引物 104 退火，所述锚定引物 104 在其 5'末端与表面 106 连接，并且具有游离 3'OH 可用于延伸。环状模板核酸 102 包括两个衔接子区 108 和 110，它们与锚定引物 104 中的序列区互补。环状模板核酸 102 中也包括插入片段 112 和与测序引物同源的区域 114，所述测序引物用于下文描述的测序反应中。

退火后，可以用模板核酸 102 内的序列延伸锚定引物 104 上的游离 3' – OH。可以沿模板延伸锚定引物 102 多次，每次重复在从锚定引物延伸的序列上添加一个与环状模板核酸互补的序列。图 11A 示出了 4 次重复，或 4 轮滚环复制，延伸的锚定引物扩增产物是 114。锚定引物的延伸导致扩增产物共价或其它方式物理连接于基质 106。可以用许多体外核酸扩增技术延伸锚定引物序列。扩增典型地在聚合酶如定向于 DNA 或 RNA 的 DNA 聚合酶，和一、二、三或四种类型的核苷酸三磷酸以及

任选在辅助的结合蛋白存在下进行。概言之，任何能够延伸引发的 3' - OH 基团的聚合酶都能够使用，只要它缺乏 3' 到 5' 外切核酸酶活性。合适的聚合酶包括例如来自嗜热脂肪芽孢杆菌、水生栖热菌、激烈火球菌、Thermococcus litoralis 和嗜热栖热菌的 DNA 聚合酶、噬菌体 T4 和 T7，以及大肠杆菌 DNA 聚合酶 I 克列诺片段。合适的定向于 RNA 的 DNA 聚合物包括例如来自禽类成髓细胞性白血病病毒的逆转录酶、来自莫罗尼鼠白血病病毒的逆转录酶和来自人免疫缺陷病毒 - I 的逆转录酶。

环状模板和锚定引物的其它实施方案在图 11B - 11D 中更详细显示。图 11B 说明了在连接后可以作为锚定引物延伸模板的退火的开环线性底物。使具有序列 5' - tcg tgt gag gtc tca gca tct tat gta tat tta ctt cta ttc tca gtt gcc taa gct gca gcc a - 3'(SEQ ID NO:5) 的模板分子与 5' 末端具有生物素连接体的锚定引物和序列 5' - gac ctc aca cga tgg ctg cag ctt - 3'(SEQ ID NO:6) 退火。模板的退火导致模板分子 5' 和 3' 模板的并列。锚定引物的 3'OH 可以用环状引物延伸。

用环状模板和锚定引物鉴定单核苷酸多态性示于图 11C。示出了具有序列 5' - gac ctc aca cga tgg ctg cag ctt - 3'(SEQ ID NO:7) 的通用锚定引物。锚定引物与具有序列 5' - ttt ata tgt att cta cga ctc tgg agt gtg cta ccg acg tcg aat ccg ttg act ctt atc ttc a - 3'(SEQ ID NO:8) 的 SNP 探针退火。SNP 随后与具有序列 5' - cta gct cgt aca tat aaa tga aga taa gat cct g - 3'(SEQ ID NO:9) 的基因的含 SNP 区杂交。含有多态性的核酸序列与 SNP 探针复合物的杂交使得随后能够连接和环化 SNP 探针。设计 SNP 探针，使得其 5' 和 3' 末端与基因组区退火，以便在多态性位点的区域中相邻，如图 11C 所示。随后可以用本文描述的方法延伸环状的 SNP 探针，并且测序。缺乏多态性的核酸不杂交，因此导致 SNP 探针的 5' 和 3' 末端并列。在这种情况下，SNP 探针不能连接形成随后延伸所需的环状底物。

图 11D 说明了一起使用缺口的寡核苷酸和环状模板分子。通过生物素连接体将具有序列 5' - gac ctc aca cga gta gca tgg ctg cag ctt - 3'(SEQ ID NO:10) 的锚定引物与表面连接。使具有序列 5' - tcg tgt gag gtc tca gca tct tat gta tat tta ctt cta ttc tca gtt gcc taa gct gca gcc a - 3'(SEQ ID NO:11) 的模板分子与锚定引物退火，得到侧翼于双链区的锚定引物中的部分单链、或有缺口的区域。然后使具有序列 5' - tgc tac - 3' 的缺口分

子退火于锚定引物。将缺口寡核苷酸的两个末端与模板分子连接，导致形成可以作为滚环扩增模板的环状核酸分子。

当从起始处开始双链分子的复制时，可以发生 RCA。随后，切口打开链之一，通过 DNA 聚合酶的作用延伸由切口产生的游离 3' - 末端羟基部分。新合成的链最终置换最初的亲本 DNA 链。这种前述复制类型称作滚环复制 (RCR)，因为复制的点可以看作在环状模板链上“滚动”，并且理论上它能够继续无限如此进行下去。此外，由于新合成的 DNA 链与最初的模板共价结合，置换的链在 5' 末端加工最初的基因组序列（如感兴趣的基因或其它序列）。在 RCR 中，最初的基因组序列后面是任何数目的与最初模板序列互补的“复制单位”，其中通过继续旋转所述最初的模板序列合成每个复制单位。因此，每个随后的旋转置换了在前面的复制循环中合成的 DNA。

通过使用 RCA 反应，可以产生一种链，该链代表环状分子的很多串联拷贝的互补序列。例如，最近已经利用 RCA 在体外获得了环状扣锁探针的等温级联扩增反应，以便检测人基因组 DNA 样品中的单拷贝基因（参见 Lizardi, et al., 1998. Nat. Genet. 19: 225 – 232）。此外，也已经利用 RCA 在基于固相的测定中检测单个 DNA 分子，尽管当该技术用于原位杂交时产生困难（参见 Lizardi, et al., 1998. Nat. Genet. 19: 225 – 232）。

如果需要，可以在升高的温度下进行 RCA，例如在超过 37°C、42°C、45°C、50°C、60°C 或 70°C 的温度下。此外，最初可以在更低的温度下，例如室温下进行 RCA，然后转变到升高的温度。升高温度下的 RCA 优选可以用耐热核酸聚合酶进行，并且采用能够在升高的温度下稳定退火和具有特异性的引物。

也可以用非天然寡核苷酸，如肽核酸进行 RCA。此外，可以在辅助蛋白，如单链结合蛋白存在下进行 RCA。

扩增固定于固体支持物的短 DNA 分子的方法，即所称的 RCA 的开发最近已经在文献中描述（参见例如 Hatch, et al., 1999. Genet. Anal. Biomol. Engineer. 15: 35 – 40; Zhang, et al., 1998. Gene 211: 277 – 85; Baner, et al., 1998. Nucl. Acids Res. 26: 5073 – 5078; Liu, et al., 1995. J. Am. Chem. Soc. 118: 1587 – 1594; Fire and Xu, 1995. Proc. Natl. Acad. Sci. USA 92: 4641 – 4645; Nilsson, et al., 1994. Science 265: 2085 –

2088)。通过杂交和 DNA 连接酶反应, RCA 靶定特定 DNA 序列。然后, 环状产物用作滚环复制反应中的模板。

等温扩增系统的其它实例包括例如(i)自我持续的序列复制(参见例如 Guatelli, et al., 1990. Proc. Natl. Acad. Sci. USA 87: 1874 – 1878), (ii)Q $\beta$ 复制酶系统(参见例如 Lizardi, et al., 1988. BioTechnology 6: 1197 – 1202)和(iii)基于核酸复制的测序(NASBA ;参见 Kievits, et al., 1991. J. Virol. Methods 35: 273 – 286)。

### 核酸模板的 PCR 扩增

在一种优选实施方案中, 用聚合酶链反应(“PCR”)制备额外拷贝的模板核酸。可以在将核酸模板分布在微微量滴定板之前或可以在核酸模板分布在微微量滴定板之后进行 PCR 扩增。

### 珠乳状液 PCR 扩增

在一种优选实施方案中, 在核酸模板分布在微微量滴定板上之前进行 PCR 扩增步骤。

在一种特别优选的实施方案中, 通过将待扩增的模板核酸(如 DNA)连接于固体支持物, 进行本文中称作“珠乳状液扩增”的新扩增系统, 所述固体支持物优选是通常球形的珠子。根据本发明的样品制备方法制备的单链模板 DNA 的文库是要连接于珠子, 从而用于扩增方法中的起始核酸模板文库的一个合适来源的实例。

将珠子与大量互补于模板 DNA 的区域的单引物种类(即图 6 中的引物 B)连接。模板 DNA 与珠结合的引物退火。将珠子悬浮液含水反应混合物中, 然后包封在油包水乳状液中。乳状液包含离散的水相微滴, 直径大约 60 – 200 $\mu\text{m}$ , 包封在耐热油相中。每个微滴优选含有扩增反应溶液(即核酸扩增所必须的试剂)。扩增的实例是 PCR 反应混合物(聚合酶、盐、dNTPs)和一对 PCR 引物(引物 A 和引物 B)。参见图 6A。一个亚组的微滴群体也包含 DNA 珠, 其中含有 DNA 模板。该亚组的微滴是扩增的基础。不在该亚组范围内的微胶囊不具有模板 DNA, 并且不参与扩增。在一种实施方案中, 扩增技术是 PCR, 并且 PCR 引物以 8:1 或 16:1 存在(即 8 或 16 个第一引物: 1 个第二引物)以进行不对称 PCR。

在此概括内容中, DNA 与固定在珠子上的寡核苷酸(引物 B)退火。

在热循环中（图 6B），单链 DNA 模板与珠子上的固定 B 引物之间的键断裂，将模板释放到周围的微胶囊包封的溶液中。扩增溶液，在该情况下是 PCR 溶液，包含额外的液相引物 A 和引物 B。液相 B 引物容易地结合于模板的互补 b' 区，因为液相引物的结合动力学比固定的引物更快。在早期阶段 PCR 中，A 和 B 链扩增同样好（图 6C）。

通过中间阶段 PCR（即第 10 - 30 个循环之间），消除 B 引物，使指数扩增停止。然后反应进入不对称扩增，扩增群体变主要是 A 链（图 6D）。在晚期阶段 PCR 中（图 6E），在 30 - 40 个循环后，不对称扩增增加了溶液中 A 链的浓度。过量 A 链开始退火于珠固定的 B 引物。随后，耐热聚合酶利用 A 链作为模板，合成扩增子的固定的、结合珠的 B 链。

在最后阶段 PCR 中（图 6F），继续的热循环迫使与珠结合的引物进行额外的退火。此阶段液相扩增可能是最少的，但固定的 B 链的浓度增加。然后，乳状液破裂，通过除去互补 A 链的变性（通过热、pH 等）使固定的产物成为单链。A 引物与固定链的 A' 区退火，用测序酶和任何必须的辅助蛋白加载固定链。然后用公知的焦磷酸技术（例如描述于美国专利 6,274,320, 6258,568 和 6,210,891，在此全文引入作为参考）对珠子进行测序。

## 模板设计

在一种优选实施方案中，要通过珠乳状液扩增进行扩增的 DNA 模板可以是 DNA 的群体，例如基因组 DNA 文库或 cDNA 文库。优选群体的每个成员在第一个末端具有共同的核酸序列，并且在第二末端具有共同的核酸序列。这可以通过例如将第一个衔接子 DNA 序列连接于 DNA 群体的一个末端，并且将第二个衔接子 DNA 序列连接于 DNA 群体的第二个末端而实现。很多 DNA 和 cDNA 文库，通过克隆载体的性质（例如 Bluescript, Stratagene, La Jolla, CA），满足这种描述，即在每个成员 DNA 的第一个末端具有共同的序列，并且在每个成员 DNA 的第二个末端具有第二种共同的序列。DNA 模板可以是适于体外扩增的任何大小（包括 PCR 和不对称 PCR 的优选扩增技术）。在一种优选实施方案中，DNA 模板的大小是约 150 - 750，例如，大小是约 250 bp）。

## 核酸模板与捕获珠结合

在第一步中，待扩增的单链核酸模板与捕获珠连接。核酸模板可以通过本领域已知的任何方式与固体支持物捕获珠连接。本领域存在许多用于使 DNA 连接于固体支持物的方法，例如优选的显微珠。根据本发明，DNA 与珠子的共价化学连接可以通过采用例如水溶性碳二亚胺的标准偶联剂实现，以便通过磷酰胺（phosphoamidate）键将 DNA 上的 5'-磷酸与胺包被的捕获珠连接。另一种备选方案是用相似的化学首先将特定寡核苷酸连接体与珠子偶联，然后用 DNA 连接酶将 DNA 连接于珠子上的连接体。用于连接寡核苷酸和珠子的其它连接化学包括使用 N-羟基琥珀酰胺（NHS）及其衍生物。在这样的方法中，寡核苷酸的一个末端可以含有与固体支持物形成共价键的反应基团（如酰胺基团），而连接体的另一个末端含有可以与待固定的寡核苷酸键合的第二反应基团。在一种优选实施方案中，寡核苷酸通过共价键与 DNA 捕获珠结合。但是，非共价键，如螯合或抗原 - 抗体复合物，也可以用于将寡核苷酸连接于珠子。

可以采用与 DNA 片段末端的独特序列特异性杂交的寡核苷酸连接体，所述末端例如来自限制酶位点的重叠末端或基于噬菌体入的克隆载体的“粘端”，但也可以有益地使用平端连接。这些方法详细描述于 US 5,674,743。优选的是任何用于固定珠子的方法将继续通过本发明方法中的步骤结合固定的寡核苷酸。

在一种实施方案中，将每个捕获珠设计为具有多个识别（即互补于）核酸模板的一部分的核酸引物，因此核酸模板与捕获珠杂交。在本文描述的方法中，模板系列的克隆扩增是需要的，所以优选的是仅仅一个独特核酸模板与任何一个捕获珠连接。

此处用到珠子可以是任何方便的大小并且可以由任何数目的已知材料制作。所述材料的实例包括：无机物、天然聚合物和合成聚合物。这些材料的具体实例包括：纤维素、纤维素衍生物、丙烯酸树脂、玻璃、硅胶、聚苯乙烯、明胶、聚乙烯吡咯烷酮、乙烯和丙烯酰胺的共聚物、与二乙烯基苯交联的聚苯乙烯等（如 Merrifield Biochemistry 1964, 3, 1385 - 1390 中的描述）、聚丙烯酰胺、乳胶凝胶、聚苯乙烯、右旋糖昔、橡胶、硅、塑料、硝酸纤维素、天然海绵、硅胶、控制孔玻璃、金属、交联右旋糖昔（例如 Sephadex<sup>TM</sup>）和琼脂糖凝胶（Sephadose<sup>TM</sup>）以及本领域

技术人员已知的固相支持物。在一种优选实施方案中，捕获珠是直径为约 25 – 40 $\mu\text{m}$  的琼脂糖珠。

## 乳化

具有连接的单链模板核酸的捕获珠乳化为耐热的油包水乳状液。可以根据本领域已知的任何合适方法形成乳状液。制备乳状液的一种方法描述于下文，但可以使用任何制备乳状液的方法。这些方法是本领域已知的，包括佐剂方法、逆流方法、横流方法、转鼓方法和膜方法。此外，可以通过改变成分的流速和速度，调整微胶囊的大小。例如，在滴加中，液滴的大小和递送的总时间可以改变。优选地，乳状液含有的珠“微反应器”的密度是每毫升大约 3,000 个珠子。

乳状液优选通过将模板连接的珠子悬浮于扩增溶液中而制备。如本文中用到的，术语“扩增溶液”表示进行模板 DNA 扩增所必须的试剂的足够的混合物。扩增溶液的一个实例，即 PCR 反应溶液，在下文实施例中提供。应该理解，可以对 PCR 溶液进行各种改变。

在一种实施方案中，将珠/扩增溶液混合物滴加到生物相容油（如轻矿物油，Sigma）的旋转中的混合物中，使其乳化。使用的油可以补加一种或多种生物相容的乳状液稳定剂。该乳状液稳定剂可以包括 Atlox 4912、Span 80 和其它已知和可商购的合适稳定剂。优选地，形成的液滴大小是 5 微米 – 500 微米，更优选地，是约 50 – 300 微米，最优选是 100 – 150 微米。

微反应器的大小没有限制。微反应器应该足够大，以包括对于需要的扩增程度来说足够的扩增试剂。但是，微反应器应该足够小，以便微反应器的群体（每个包含 DNA 文库的一个成员）能够通过常规实验室设备（如 PCR 热循环设备、试管、温育箱等）扩增。

根据上文描述的限制，微反应器的最佳大小可以是直径 100 – 200 微米。这种大小的微反应器使得能够在体积小于 10 ml 的微反应器悬浮液中扩增包含约 600,000 个成员的 DNA 文库。例如，如果 PCR 是选择的扩增方法，10 ml 将装在具有 96 管容量的常规热循环仪的 96 个管中。在一种优选实施方案中，600,000 个微反应器的悬浮液的体积是小于 1 ml。可以在常规 PCR 热循环仪的约 10 个管中扩增小于 1 ml 的悬浮液。在一种最优选的实施方案中，600,000 个微反应器的悬浮液将具有小于

0.5 ml 的体积。

### 扩增

包封后，可以通过任何合适的 DNA 扩增方法来扩增模板核酸，所述方法包括基于转录的扩增系统(Kwoh D. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86:1173 (1989); Gingeras T. R. et al., PCT 申请 WO 88/10315; Davey, C. et al., 欧洲专利申请公开 No. 329,822; Miller, H. I. et al., PCT appl. WO 89/06700)和“race”(Frohman, M. A., In: PCR Protocols: A Guide to Methods and Applications, Academic Press, NY (1990))以及“一侧 PCR”(Ohara, O. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86:5673 – 5677 (1989))。本发明中也可以使用其它较不常见的方法，如“二 - 寡核苷酸”扩增、等温扩增(Walker, G. T. et al., Proc. Natl. Acad. Sci. (U.S.A.) 89:392 – 396 (1992))和滚环扩增（综述于 5,714,320）。

在一种优选实施方案中，通过 PCR 进行 DNA 扩增。可以通过用包含 PCR 的所有必须试剂的 PCR 溶液包封结合于珠子的靶核酸，进行本发明的 PCR。然后，可以通过将乳状液暴露于本领域已知的任何合适热循环方案，完成 PCR。在一种优选实施方案中，进行 30 – 50 个循环，优选约 40 个循环的扩增。理想地，但并不是必须的，在扩增程序之后，在扩增循环后存在一个或多个杂交和延伸循环。在一种优选实施方案中，进行 10 – 30 个循环，优选约 25 个循环的杂交和延伸（例如，如实施例中的描述）。常规地，扩增模板，直到每个珠子典型地固定至少 2 百万 – 5 百万个拷贝，优选约 1 千万 – 3 千万个拷贝的模板 DNA。

### 使乳状液破裂和回收珠子

扩增模板后，使乳状液“破裂”（在本领域中也称作“破乳作用”）。存在很多使乳状液破裂的方法（参见例如美国专利 5,989,892 及其中引用的参考文献），并且本领域技术人员能够选择合适的方法。在本发明中，一种使乳状液破裂的方法是加入额外的油，导致乳状液分为两相。然后除去油相，加入合适的有机溶解（如己烷）。混合后，除去油/有机溶剂相。该步骤可以重复数次。最后，除去珠子上面的水相。然后用有机溶剂/退火缓冲液混合物（如一种合适的退火缓冲液描述于实施例中）洗涤珠子，然后在退火缓冲液中再次洗涤。合适的有机溶剂包括醇，如

甲醇、乙醇等。

然后可以将扩增的含模板的珠子重悬于水溶液中，用于例如根据已知技术的测序反应中。（参见 Sanger, F. et al., Proc. Natl. Acad. Sci. U.S.A. 75, 5463 – 5467 (1977); Maxam, A. M. & Gilbert, W. Proc Natl Acad Sci USA 74, 560 – 564 (1977); Ronaghi, M. et al., Science 281, 363, 365 (1998); Lysov, I. et al., Dokl Akad Nauk SSSR 303, 1508 – 1511 (1988); Bains W. & Smith G. C. J.TheorBiol 135, 303 – 307(1988); Drnanac, R. et al., Genomics 4, 114 – 128 (1989); Khrapko, K. R. et al., FEBS Lett 256, 118 – 122 (1989); Pevzner P. A. J Biomol Struct Dyn 7, 63 – 73 (1989); Southern, E. M. et al., Genomics 13, 1008 – 1017 (1992).)如果珠子用于基于焦磷酸的测序反应（描述于例如美国专利 6,274,320, 6258,568 和 6,210,891, 在此全文引入作为参考），则必须除去 PCR 产物的第二条链，并且使测序引物与结合于珠子的单链模板退火。

简言之，用任何公知的方法如 NaOH、低离子(如盐)强度或热处理，使得第二条链解链离开。该解链步骤后，沉淀珠子，弃去上清液。将珠子重悬于退火缓冲液中，加入测序引物，并且用标准退火循环退火于珠子连接的单链模板。

### 纯化珠子

此时，可以直接在珠子上或者在不同反应容器中对珠子上的扩增 DNA 进行测序。在本发明的一种实施方案中，通过将珠子转移到反应容器并且使 DNA 进行测序反应（例如焦磷酸或 Sanger 测序），直接在珠子上对 DNA 进行测序。或者，可以分离珠子，可以每个珠子上除去 DNA 并且测序。在任何一种情况下，可以在每个单独的珠子上进行测序步骤。但是，该方法尽管是商业上可获得并且是技术上可行的，但可能不是最经济的，因为很多上述珠子将是阴性珠（不连接扩增的 DNA 的珠子）。因此，可以在分布在微微量滴定板上之前，用以下任选方法除去不含核酸模板的珠子。

如果最初DNA连接的目的是使具有两个不同拷贝的DNA的珠子减少到最小，则高百分比的珠子可能是“阴性的”（即不连接扩增的核酸模板）。对于有用的焦磷酸测序，每个珠子应该含有多个拷贝的单种类的DNA。通过使具有单个结合的DNA片段的珠子总数最大（在扩增前），

最接近到达这种要求。这一目的可以通过观察数学模型而实现。

对于在 M 个珠子上随机分布的“N”个 DNA 的一般情况，含有任何数目 DNA 的相对珠群体依赖于 N/M 比值。含有 N 个 DNA 的珠子的比例 R(N) 可以用泊松分布计算：

$$R(N) = \exp - (N/M) \times (N/M)^N / N! \quad (\text{其中 } \times \text{ 是乘号})$$

下表 1 显示了针对多个 N/M (DNA 片段与珠子的比值的平均值) 和 N (实际与珠子结合的片段数目)。

表 1

N/M	0.1	0.5	1	2
R(0)	0.9	0.61	0.37	0.13
R(1)	0.09	0.3	0.37	0.27
R(N>1)	0.005	0.09	0.26	0.59

在表中，最上面一行表示多个 N/M 比值。R(0) 表示不具有 DNA 的珠子的比例，R(1) 表示连接一个 DNA 的珠子的比例（扩增前），R(N>1) 表示连接一个以上 DNA 的珠子的比例（扩增前）。

该表表明含有单个 DNA 片段的珠子的最大比例是 0.37 (37%)，并且发生在片段与珠子的比值为 1 时。在该混合物中，约 63% 的珠子对于测序是无用的，因为它们不具有 DNA，或具有超过单种类的 DNA。此外，控制片段与珠子的比值，需要进行复杂的计算，并且由于变异，会产生具有显著更小有用珠子比例的珠子批次。

如果含有扩增子的珠子（来自于至少一个片段的结合）可以与不含扩增子的那些（来自不具有结合的片段的珠子）分开，则这种低效率可以显著改善。扩增子定义为由体外核酸扩增技术产生的任何核酸分子。结合可以在平均片段与珠子比值较低( $N/M < 1$ ) 时进行，使结合超过一个 DNA 的珠子的比值最小。分离步骤可以除去大多数或所有不具有 DNA 的珠子，得到具有一个种类的扩增 DNA 的珠子的富集群体。这些珠子可以用于任何测序方法，例如焦磷酸测序。由于具有一个扩增子的珠子 ( $N=1$ ) 的比例已经富集，任何测序方法都将是有效的。

作为一个实例，由于片段与珠子的比值的平均值是 0.1，90% 的珠子将没有扩增子，9% 的珠子将是具有一个扩增子的有用珠子，0.5% 的

珠子将具有一个以上扩增子。本发明的富集过程将除去 90% 的零扩增子珠，得到珠子群体，其可测序比例如下(N=1):

$$1 - (0.005/0.09) = 94\%.$$

片段稀释到珠混合物，伴随含有扩增子的珠子分开，可以产生相对于最佳未富集方法 2.5 倍的富集。 $94\%/37\%$  (参见上表 N/M=1) = 2.5。本发明的富集程序的额外益处是可测序的珠子的最终比例对 N/M 的变异是相对不敏感的。因此，推导最佳 N/M 比值的复杂计算是不必要的，或可以在较低水平精确度下进行。这将最终使该程序更适于受较少训练的个人或自动化的执行。该程序的额外益处是零扩增子珠可以再循环和再利用。尽管再循环不是必须的，但它可以减少成本或试剂的总量，使本发明的方法更适于一些目的，例如，便携式取样、远程自动化取样等。此外，该程序的所有益处（即受较少训练的个人、自动化、试剂再循环）将减少程序的花费。该程序在下文更详细描述。

富集程序可以用于处理已经在上文珠乳状液方法中扩增的珠子。设计扩增，使得每个扩增的分子在其 3' 末端含有相同 DNA 序列。核苷酸序列可以是 20 聚体，但可以是从 15 个碱基或更多，如 25 个碱基、30 个碱基、35 个碱基或 40 个碱基或更长的任何序列。天然地，尽管较长的寡核苷酸末端是功能性的，它们不是必须的。该 DNA 序列可以由本领域技术人员导入扩增的 DNA 末端。例如，如果用 PCR 扩增 DNA，则该序列可以是 PCR 引物对的一个成员的一部分。

富集过程的一个示意图示于图 7。此处，与 4 个空珠子混合的扩增子结合的珠子代表片段稀释的扩增珠混合物。在步骤 1，使与扩增子的 3' 末端互补的生物素化的引物与扩增子退火。在步骤 2，将 DNA 聚合酶和 4 种天然脱氧核苷酸三磷酸(dNTPs)加入珠混合物，并且延伸生物素化的引物。该延伸是为了增强生物素化的引物和珠结合的 DNA 之间的结合。如果生物素化的引物 - DNA 键是很强的（如，在高离子环境中），则该步骤可以省略。在步骤 3，将容易受到磁场吸引的链霉亲和素包被的珠子（本文中称作“磁性链霉亲和素珠”）引入珠混合物。磁珠是可商购的，例如，从 Dynal (M290) 购买。链霉亲和素捕获部分结合与扩增子杂交的生物素，其随后将扩增子结合的珠子特异性固定于磁性链霉亲和素珠。

在步骤 5，将磁场（由磁铁代表）施加在反应混合物附近，这导致

所有“磁性链霉亲和素珠/扩增子结合的珠复合物”位于最接近磁场的管的一侧。未连接扩增子的结合的珠的磁珠预期也位于相同侧。不含扩增子的珠子保留在溶液中。洗涤珠混合物，除去并且弃去未通过磁铁固定的珠子（即空珠子）。在步骤 6，通过“解链”步骤可以通过例如热或 pH 改变而完成。热可以是低盐条件下（如在低离子环境，如 0.1X SSC）的 60 °C。pH 的改变可以通过加入 NaOH 而实现。然后洗涤混合物，回收含有扩增子结合的珠的上清液，同时通过磁场保留目前未结合的磁珠。得到的富集的珠子可以用于 DNA 测序。注意到 DNA 捕获珠上的引物可以与上文步骤 2 的引物相同。在此情况下，扩增子 - 引物互补链（有或无延伸）的退火是靶 - 捕获物亲和力的来源。

可以用多种捕获物 - 靶对代替生物素 - 链霉亲和素对。两个种类是这样的对，其结合可以随后被裂解，并且它们在实际上可以达到的条件下可逆结合。如果需要靶 - 捕获物复合物的裂解，可裂解的对包括硫醇 - 硫醇、洋地黄毒昔/抗洋地黄毒昔 - Captavidin<sup>TM</sup>。

如上文的描述，步骤 2 是任选的。如果省略步骤 2，可能不需要从扩增子结合的珠子分离磁珠。扩增子结合的珠，其连接了磁珠，可以直接用于测序。如果测序要在微孔中进行，如果扩增子结合的珠 - 磁珠复合物可以置入微孔中，则不是必须进行分离。

尽管磁性捕获珠的使用是方便的，捕获部分可以与其它表面结合。例如，链霉亲和素可以与表面，例如，管的内表面结合。在此情况下，扩增的珠混合物可以流过。扩增子结合的珠将保留，直到“解链”，此时空珠将流过。这种安排对于使珠制备过程自动化可能是特别有利的。

尽管上文描述的实施方案是特别有用的，也可以预计用其它方法分离珠子。例如，可以用使靶 - 捕获珠复合物发荧光的荧光部分标记捕获珠。可以通过流式细胞术或荧光细胞分选仪分离靶捕获珠复合物。采用大捕获珠，能够通过过滤或其它颗粒大小分离技术进行分离。由于捕获珠和靶珠都能够与许多其它珠子形成复合物，可以粘着大量交联的捕获珠 - 靶珠。粘着团块很大，使得能够通过简单洗掉未粘着的空珠而进行分离。该方法详细描述于例如 Bauer, J.; J. Chromatography B, 722 (1999) 55 - 69 和 Brody et al., Applied Physics Lett. 74 (1999) 144 - 146。

然后，根据上述方法制备的各自含有多个拷贝的单种类核酸模板的

DNA 捕获珠可以适用于分布在微微量滴定板上。

### 微微量滴定板上的核酸扩增

在一种替代述实施方案中，在扩增前将核酸模板分布在微微量滴定板上，然后在微微量滴定板上原位扩增。该方法详细描述于实施例。

### 3. 对核酸模板进行测序

根据本发明的方法用焦磷酸测序对核酸模板进行测序。该技术基于在 DNA 合成中检测释放的焦磷酸(Ppi)。参见例如 Hyman, 1988. A new method of sequencing DNA. Anal Biochem. 174:423 – 36; Ronaghi, 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. 11:3 – 11。

在酶促反应的级联中，与掺入的核苷酸数目成比例地产生可见光。该级联以核酸聚合反应开始，其中随着通过聚合酶掺入核苷酸，释放无机 Ppi。释放的 Ppi 通过 ATP 硫酸化酶转化为 ATP，其给荧光素酶氧化荧光素提供能量，并且产生光。由于添加的核苷酸是已知的，可以确定模板的序列。固相焦磷酸测序利用三酶系统中的固定的 DNA (参见附图)。为了增加信噪比，天然 dATP 被 dATP  $\alpha$  S 替代。典型地，dATP  $\alpha$  S 是两种异构体(Sp 和 Rp)的混合物；焦磷酸测序中纯 2' - 脱氧腺苷 - 5' - O' - (1 - 硫代三磷酸酯) Sp - 异构体的使用允许显著更长的读出结果，最多达到读出结果长度的加倍。

### 4. 用于对核酸进行测序的装置

本发明提供了用于对核酸进行测序的装置，其通常包括一个或多个用于进行测序反应的反应室、用于将试剂送入或送出反应室的工具，和用于检测测序反应事件的工具。在另一种实施方案中，该装置包括试剂递杯，其含有位于平坦表面上的多个凹陷。在一种优选实施方案中，该装置连接于至少一台计算机，所述计算机用于控制装置的各个组件，并且用于存储和/分析从序列反应事件的检测获得的信息。

本发明也提供了在惰性基质材料，本文也称作“固体支持物”上排列的一个或多个反应室，其使得核酸模板能够离散定位，并且测序反应中的试剂能离散定位于限定的空间，并且使得能够检测测序反应事件。因此，如本文用到的，术语“反应室”或“分析物反应室”表示位于基

质材料上的、促使例如核酸测序反应中的试剂相互作用的定位的区域。如下文更完全的讨论，本发明考虑的测序反应优选存在于串联的多个单独的核酸样品中，特别是对来自基因组和染色体核酸模板（如 DNA）的多个核酸样品同时进行测序。

因此，本发明的装置优选包含足够数目的反应室，用于进行所述多个单独的测序反应。在一种实施方案中，存在至少 10,000 个反应室，优选至少 50,000 个反应室，更优选超过 100,000 个反应室，甚至更优选超过 200,000 个反应室。

由于同时进行的测序反应的数目受到反应室数目的限制，通过制作孔密度增加的板而增加通量。下表 2 显示了分别来自 25 x 75 mm 和 40 x 75 mm 阵列的 14 x 43 mm 和 30 x 60 mm 活性区域的该级数。

表 2：更高孔技术阵列的开发

孔距 (um)	孔 直径(um)	孔数目 (14 x 43 mm)	孔数目 (30 x 60 mm)
50	44	275K	800K
43	38	375K	1.2M
35	31	575K	1.6M
25	22	1.1M	3.2M

阵列上的反应室典型地采取基质材料中的凹陷或孔的形式，其具有宽度和深度，试剂可以沉积在其中。典型地，核酸模板分布在反应室中的一个或多个固体支持物或珠上；试剂位于促进反应并且从反应室流过的介质中。当形成凹陷或孔时，反应室优选具有足够的尺寸，以便允许(i)将必须的试剂引物室中，(ii)在室内发生反应，和(iii)抑制室间试剂的混合。孔或凹陷的形状优选是圆形或圆柱形，但可以是多边形，以便近似于圆形或圆柱形。在一种优选实施方案中，孔或凹陷的形状基本是六边形。凹陷可以具有平滑壁表面。在一种额外的实施方案中，凹陷可以具有至少一个不规则的壁表面。凹陷可以具有平坦的底面或凹陷的底面。

反应室的间距可以是 5μm - 200μm。通过测量两个相邻反应室之间中心与中心的距离，确定间距。典型地，反应室的间距可以是 10μm -

150 $\mu\text{m}$ , 优选 20 $\mu\text{m}$  – 100 $\mu\text{m}$ , 最优选 40 – 60 $\mu\text{m}$ 。在一种实施方案中, 反应室在一个维度上的宽度(直径)是 0.3 $\mu\text{m}$  – 100 $\mu\text{m}$ , 更优选 20 $\mu\text{m}$  – 70 $\mu\text{m}$ , 最优选约 30 – 50 $\mu\text{m}$ 。反应室的深度是优选 10 $\mu\text{m}$  – 100 $\mu\text{m}$ , 优选 20 $\mu\text{m}$  – 60 $\mu\text{m}$ 。或者, 反应室的深度是反应室在一个维度上的宽度的 0.25 – 5 倍, 在另一实施方案中, 是反应室在一个维度上的宽度的 0.3 – 1 倍。

在一种优选实施方案中, 阵列由切片的光导纤维束(即一束熔凝的光导纤维缆)制作, 并且反应室是通过蚀刻光导纤维反应器阵列的一面而形成的。凹陷也可以通过蚀刻、造型或显微机械加工在基质上形成。

每个凹陷或反应室的深度典型是 10 $\mu\text{m}$  – 100 $\mu\text{m}$ ; 或者, 深度是凹陷的宽度尺寸的 0.25 – 5 倍, 优选是凹陷宽度尺寸的 0.3 – 1 倍。

在一种实施方案中, 本文描述的阵列典型地包括平坦的上表面和平坦的下表面, 下表面任选是传导性的, 使得可以通过平坦的下表面检测来自反应室的光信号。在这些阵列中, 典型地, 上表面和下表面之间的距离不超过 10 cm, 优选不超过 2 cm, 通常是 0.5 mm – 5 mm, 最优选约 2 mm。

在一种特别优选的实施方案中, 固体支持物称作微微量滴定板, 反应室中心与中心的间距是大约 43 $\mu\text{m}$  – 50 $\mu\text{m}$ , 孔直径是大约 38 $\mu\text{m}$  – 44 $\mu\text{m}$ , 孔体积是 10 – 150 pL, 优选 20 – 90 pL, 更优选 40 – 85 pL, 最优选约 75 pL。

在一种实施方案中, 阵列的每个凹陷或反应室含有用于分析核酸或蛋白的试剂。典型地, 这些含有核酸的反应室(不是阵列中的所有反应室都需要含有核酸)仅仅含有单种类的核酸(即单个感兴趣的核酸)。在任何特定反应室中可能存在单拷贝的这种种类的核酸, 或可能具有多个拷贝。通常优选的是反应室含有至少 100,000 个拷贝的核酸模板序列, 优选至少 1,000,000 个拷贝, 更优选 2,000,000 – 20,000,000 个拷贝, 最优选 5,000,000 – 15,000,000 个拷贝的核酸。本领域技术人员可以理解, 任何一个反应室中核酸种类拷贝数目的改变将影响焦磷酸测序反应中产生的光子数目, 并且可以常规进行调整, 以根据需要提供更多或更少的光子信号。在一种实施方案中, 用 PCR、RCA、连接酶链反应、其它等温扩增或其它常规核酸扩增方法扩增核酸种类, 以提供需要的拷贝数目。在一种实施方案中, 核酸是单链的。

## 固体支持物材料

任何材料都可以用作固体支持物材料，只要其表面允许稳定连接引物和检测核酸序列。固体支持物材料可以是平板的或可以是有凹陷的，例如，在有凹陷的光导纤维末端或在蚀刻、造型或其它方式显微机械加工在平坦表面中的孔中，例如，它们是采用微电子机械系统的构造中通常采用的技术形成。参见例如 Rai - Choudhury, HANDBOOK OF MICROLITHOGRAPHY, MICROMACHINING, AND MICROFABRICATION, VOLUME I: MICROLITHOGRAPHY, Volume PM39, SPIE Press (1997); Madou, CRC Press (1997), Aoki, Biotech. Histochem. 67: 98 - 9 (1992); Kane et al., Biomaterials. 20: 2363 - 76 (1999); Deng et al., Anal. Chem. 72:3176 - 80 (2000); Zhu et al., Nat. Genet. 26:283 - 9 (2000)。在一些实施方案中，固体支持是透光的，例如玻璃。

透光的固体支持物上的连接位点的阵列可以采用集成电路构造中常用的平板印刷技术构造，所述技术描述于例如美国专利 5,143,854, 5,445,934, 5,744,305 和 5,800,992 中描述的连接技术；Chee et al., Science 274: 610 - 614 (1996); Fodor et al., Nature 364: 555 - 556 (1993); Fodor et al., Science 251: 767 - 773 (1991); Gushin, et al., Anal. Biochem. 250: 203 - 211 (1997); Kinosita et al., Cell 93: 21 - 24 (1998); Kato - Yamada et al., J. Biol. Chem. 273: 19375 - 19377 (1998); 和 Yasuda et al., Cell 93: 1117 - 1124 (1998)。光刻和电子束平板印刷术用允许连接修饰的生物分子（如蛋白或核酸）的连接基团使固体支持物或基质敏化。参见例如 Service, Science 283: 27 - 28(1999); Rai - Choudhury, HANDBOOK OF MICROLITHOGRAPHY, MICROMACHINING, AND MICROFABRICATION, VOLUME I: MICROLITHOGRAPHY, Volume PM39, SPIE Press (1997)。或者，可以用 Zasadzinski et al., Science 263: 1726 - 1733 (1994) 中描述的薄膜技术产生敏化位点的阵列。

基质材料优先用促进反应事件的检测的材料制造。例如，在典型的测序反应中，可以通过检测光子，监测 dNTP 与要测序的样品核酸的结合，所述光子是通过酶对测序反应中释放的磷酸的作用而产生的。因此，具有由透明或光导材料制造的基质材料，促进了光子的检测。

在一些实施方案中，固体支持物可以与用于检测和传导光产物的光导纤维束偶联。束内的光导纤维总数可以不同，使得匹配用于测序反应

中的阵列中的各个反应室数目。掺入束中的光导纤维数目设计为匹配检测器件的分辨率，使得能够 1:1 成像。选择束的总体大小，以便优化检测器件的可用区域，同时保持反应室中的理想试剂（流动）特征。因此，对于具有  $15\mu\text{m}$  像素的  $4096 \times 4096$  像素 CCD（电荷偶联的器件）阵列，纤维束选择为大约  $60\text{ mm} \times 60\text{ mm}$ ，或具有大约  $90\text{ mm}$  的直径。所需数目的光导纤维最初熔凝成束或光导纤维阵列，其末端随后可以被切割，并且磨光，以形成具有需要的厚度（如  $1.5\text{ mm}$ ）的“薄片”。得到的光导纤维薄片与玻璃平面具有相似的加工特性。各个纤维可以具有任何大小的直径（例如  $3\mu\text{m} - 100\mu\text{m}$ ）。

在一些实施方案中，采用两个光导纤维束：第一束直接连接于检测器件（在此也称作纤维束或连接器），第二束用作反应室基质（薄片或基质）。在这种情况下，这两者置于直接接触中，任选采用光耦合流体，使得反应室中心成像到检测器件上。如果 CCD 用作检测器件，薄片可以略微更大，以使 CCD 区域的使用最大，或略微更小，以匹配典型显微镜玻片的形式— $25\text{ mm} \times 75\text{ mm}$ 。选择束内各个纤维的直径，以便在本领域现有技术的限制内，使单个反应将成像到检测器件中的单个像素中的可能性最大。示例的直径对于纤维束是  $6 - 8\mu\text{m}$ ，对于薄片是  $6 - 50\mu\text{m}$ ，但  $3 - 100\mu\text{m}$  的任何直径都可以使用。可以从 CCD 照相机生产商商购纤维束。例如，可以从 Incom, Inc. (Charlton, MA) 获得薄片，并且从典型  $2\text{ mm}$  厚的光导纤维的大熔凝物切割或磨制，但  $0.5 - 5\text{ mm}$  的厚度都是可以的。薄片与玻璃平面或玻璃显微镜玻片具有相似加工特性的。

可以在由光导纤维材料制造的基质中形成反应室。通过处理纤维素的末端，例如用酸处理，在光导纤维材料中形成凹痕，使光导纤维表面有凹陷。因此，在一种实施方案中，从光导纤维束形成凹陷，优选可以通过蚀刻光导纤维束的一个末端，形成凹陷。每个有凹陷的表面可以形成反应室。所述阵列在本文中称作光导纤维反应室阵列或 PORA。凹痕的深度是各个光导纤维直径的大约一半到纤维直径的  $2 - 3$  倍。可以通过将光导纤维薄片的一侧置于酸浴中不同的时间，将凹陷引入纤维末端。时间的量可以根据需要的反应凹陷的总体深度而改变（参见例如 Walt, et al., 1996. Anal. Chem. 70: 1888）。宽通道凹陷可以具有大约  $14\text{mm} \times 43\text{mm}$  的统一流速尺寸。因此，由于该大约的尺寸和大约  $4.82 \times$

$10^{-4}$  个凹陷/ $\mu\text{m}^2$  的密度，该装置可以具有大约 290,000 个流体可进入的凹陷。一些方法在本领域中已知可以用于将分子连接于光导纤维束末端中蚀刻的凹陷中（并且检测连接的分子）。参见例如 Michael, et al., Anal. Chem. 70: 1242 – 1248 (1998); Ferguson, et al., Nature Biotechnology 14: 1681 – 1684 (1996); Healey and Walt, Anal. Chem. 69: 2213 – 2216 (1997)。也可以用类似于制备平面支持物上的反应垫图样中使用的技术的照相平板印刷技术，在微孔中建立反应位点的图样。参见 Healey, et al., Science 269: 1078 – 1080 (1995); Munkholm and Walt, Anal. Chem. 58: 1427 – 1430(1986), and Bronk, et al., Anal. Chem. 67: 2750 – 2757 (1995)。

光导纤维薄片的对侧（即未蚀刻侧）典型地是高度磨光的，以便能够光耦合（例如通过浸入油或其它光耦合流体）于第二光导纤维束。该第二光导纤维束准确匹配含有反应室的光导薄片的直径，并且作为将光产物传导到连接的检测器件，如 CCD 成像系统或照相机的管道。

在一种优选实施方案中，彻底清洁光导纤维薄片，例如通过连续洗涤：在体积：体积比为 15%  $\text{H}_2\text{O}_2$ /15%  $\text{NH}_4\text{OH}$  的水溶液中洗涤，然后用去离子水冲洗 6 次，然后 0.5M EDTA，然后是去离子水 6 次，然后 15%  $\text{H}_2\text{O}_2$ /15%  $\text{NH}_4\text{OH}$ ，然后去离子水 6 次（每次洗涤温育进行 1 个半小时温育）。

优选对光导纤维薄片的表面进行涂布，以促进其在测序反应中的使用。涂布的表面优选是透光的，使得能够容易连接蛋白和核酸，并且对固定的蛋白的活性没有不利影响。此外，该表面优选使大分子的非特异性吸附最少，并且增加连接的大分子（例如连接的核酸和蛋白）的稳定性。

用于涂布阵列的合适材料包括例如塑料（如聚苯乙烯）。塑料可以优选是旋转涂布或溅射的（ $0.1 \mu\text{m}$  厚度）。其它用于涂布阵列的材料包括金层，例如 24K 金， $0.1 \mu\text{m}$  厚，具有吸附的长链硫醇烷的自组装单层。然后将生物素共价偶联到表面上，用生物素结合蛋白（如链霉亲和素或亲和素）饱和。

涂布材料可以额外包括用于将引物锚定于基质的系统。使蛋白通过氨基、巯基或羧基直接共价偶联的有机硅烷试剂，也可以用于涂布阵列。其它涂布物质包括光活性连接体，如光生物素 (Amos et al., "Biomaterial Surface Modification Using Photochemical Coupling Technology," in

Encyclopedic Handbook of Biomaterials and Bioengineering, Part A: Materials, Wise et al. (eds.), New York, Marcel Dekker, pp. 895926, 1995)。

其它涂布材料包括亲水聚合物凝胶（聚丙烯酰胺、多糖），其优选直接聚合在表面上或在聚合后共价连接聚合物链(Hjerten, J. Chromatogr. 347,191 (1985); Novotny, Anal. Chem. 62,2478 (1990), 以及 pluronic 聚合物(三嵌段共聚物如 PPO - PEO - PPO,也称作 F - 108), 其特异性吸附于聚苯乙烯或硅烷化的玻璃表面(Ho et al., Langmuir 14:3889 – 94, 1998), 以及被动吸附的生物素结合蛋白层。也可以用使得试剂能够通过胺键偶联的环氧物涂布表面。

此外，上述任何材料都可以用本领域公知可以用于固定酶和核酸的一种或多种官能团，例如金属螯合基团（例如次氨基三乙酸、亚氨基二乙酸、五配位基化合物）进行衍生，所述基团将结合 6 xHis 标记的蛋白和核酸。

可以使用表面涂层，其增加可以用于随后处理，例如酶的连接（下文讨论）的可获得的结合位点的数目，使其超过 2D 表面的理论结合能力。

在一种优选实施方案中，用于制备熔凝的光导纤维束/薄片的各个光导纤维的直径（即 6  $\mu\text{m}$  – 12  $\mu\text{m}$ ）大于用于光学成像系统中的直径（即 3  $\mu\text{m}$ ）。因此，一些光学成像纤维可以用于使单个反应位点成像。

在一种特别优选的实施方案中，从商购的光导纤维面板形成称作‘微微量滴定板’的用于核酸模板测序的样品药筒，所述面板进行酸蚀刻，以产生孔结构。每个光导纤维核心的直径是约 44 微米，具有 2 – 3 微米覆层，通过酸蚀刻形成每个孔，形成约 65 pL – 85 pL，优选约 75 pL 的反应孔体积。光导纤维面板表面蚀刻的孔的使用具有三重目的；i)延迟从阵列不同区域的发射光进行荧光漫射；ii)分离含有扩增的模板分子的反应室(“试管”);和 iii)非常有效的、高数值光圈( high numerical aperture )与 CCD 的光耦合。最后，测序模板固定在孔内的量越大，能够达到更多的光学信号。

## 送递工具

用于将试剂送递到反应室中的工具的一个实例是图 13 所示的本发明的灌注室。灌注室包括具有透明的上和下侧的密封腔室。其设计为使

溶液流过基质表面，并且使得能够迅速交换试剂。因此，其适于进行例如焦磷酸测序反应。可以调整室的形状和尺寸，以优化试剂交换，以便将集流交换、扩散流交换，或这两者引入层流或湍流方案中。

灌注室优选在制备时与成像系统分离，而仅仅当进行测序反应时，其置于成像系统上。在一种实施方案中，固体支持物（即 DNA 芯片或载玻片）通过金属或塑料架固定，其可以组装或解开，以允许更换所述固体支持物。灌注室的固体支持物的下侧携带反应室阵列，并且由于具有常规基于光的聚焦系统，用高数值光圈物镜使反应室中心阵列的图像聚焦到 CCD 成像系统上。

因此，可以平行分析很多样品。采用本发明的方法，可以通过使含有酶和一个核苷酸的溶液流过表面，然后检测由每个样品产生的信号，分析很多核酸模板。然后可以重复该程序。或者，一些与模板互补的不同寡核苷酸可以分布在表面上，然后与模板杂交后。可以通过用各种寡核苷酸作为引物产生的信号监测针对每种寡核苷酸的脱氧核苷酸或双脱氧核苷酸的掺入。通过组合来自表面的不同区域的信号，可以通过采用多种双脱氧核苷酸进行的 4 轮聚合酶反应进行基于序列的分析。

当支持物是有凹陷的阵列形式，如微微量滴定板的末端或其它微孔阵列形式，试剂的合适送递工具包括流动和洗涤，并且也包括例如流动、喷雾、电喷雾、喷墨送递、冲压、超声自动化(Sonotek Corp., Milton, NY)和滚压。当使用喷雾时，试剂可以以同质薄层送递到微微量滴定板中，所述薄层由工业型喷雾嘴(Spraying Systems, Co., Wheaton, IL)或用于薄层色谱(TLC)中的雾化喷头，如 CAMAG TLC Sprayer (Camag Scientific Inc., Wilmington, NC)产生。这些喷雾器将试剂喷雾成大小为 0.3 – 10 μm 的气溶胶喷雾颗粒。

连续的试剂送递步骤优先用本领域公知的技术通过洗涤步骤隔开。这些洗涤可以例如用上文描述的方法，包括高流速喷雾器或通过微微量滴定板或微孔阵列表面上的液体流进行。洗涤可以在原料与试剂反应以在每个反应室中形成产物之后，但在送递到任何一个反应室的试剂从该反应室扩散到其它反应室之前的任何时间段进行洗涤。在一种实施方案中，任何一个反应室都不依赖于在任何其它反应室中形成的产物，而是用一种或多种试剂产生。

完整装置的一个实施方案说明于图 12。该装置包括入口导管 200，

其与可拆卸的灌注室连通。该入口导管 200 使得测序试剂能够进入多个管 202 - 212，所述每个管与多个测序分配试剂容器 214 - 224 连通。

用加压系统或泵驱动正向流，从而通过导管 200 将试剂引入灌注室 226。典型地，试剂的流速是 0.05 - 50 ml/分钟（如 1 - 50ml/分钟），体积是 0.100 ml 到连续流（用于洗涤）。阀门在计算机控制下，以便核苷酸和洗涤试剂能够循环。测序试剂，如聚合酶，可以与核苷酸预先混合，或在液流中加入。一个歧管使所有 6 个管 202 - 212 成为 1 个用于给灌注室进料的管。因此，一些试剂递口使得能够进入灌注室。例如，可以用一个口输入含水测序试剂，而另一个口使这些试剂（以及任何反应产物）从灌注室除去。

在一种优选实施方案中，将一种或多种试剂递送到固定或连接于可移动的固体支持物，如珠子或微球的群体的阵列。珠子或微球不需要是球形的，不规则形状的珠子也可以使用。它们典型地从许多物质，如塑料、玻璃或陶瓷制造，并且珠子的大小从纳米到微米，这取决于反应室的宽度。可以使用多种珠化学，如甲基苯乙烯、聚苯乙烯、丙烯酸聚合物、乳胶、顺磁材料、氧化钛溶胶、石墨和二氧化钛。可以选择珠子的构造和化学，以促进需要的试剂的连接。

在另一种实施方案中，首先合成生物活性试剂，然后与珠子共价连接。如本领域技术人员理解的，可以依赖于生物试剂和珠子的组成而实现。诸如某些聚合物的固体支持物表面与化学反应基团如硫醇、胺、羧基等的官能化，是本领域公知的。

在一种优选实施方案中，将核酸模板递送到珠子上的微微量滴定板。同样将荧光素酶和硫酸化酶递送到珠子上的每个孔（参见附图），如同 DNA 聚合酶。注意到核酸模板、荧光素酶和硫酸化酶中的一个或多个，可以各自递送到分开的珠子上，或一起递送到相同的珠子上。为了能够在升高的温度下对 DNA 进行测序，我们克隆并修饰了来自嗜热脂肪芽孢杆菌的耐热硫酸化酶。我们也克隆和修饰了一些用于固相酶活性的荧光素酶，包括 *P. pennsylvanica* 和 *P. pyralis* 的荧光素酶。*P. pyralis* 荧光素酶用于优选实施方案中。

可以使用“空白”珠，其具有促进用户连接需要的官能度的表面化学。这些用于空白珠的表面化学的额外实例包括但不限于氨基，包括脂族和芳香胺、羧酸、醛、酰胺、氯甲基、酰肼、羟基、磺酸酯和硫酸酯。

这些官能团可以用于通常采用已知的化学将任何数目的不同候选试剂添加到珠子上。例如，含有碳水化合物的候选试剂可以连接于氨基官能化的支持物，用标准技术制备碳水化合物的醛，然后使醛与表面上的氨基反应。在一种替代的实施方案中，可以使用巯基连接体。存在一些本领域已知的巯基反应连接体，例如 SPDP、马来酰亚胺、 $\alpha$ -卤代乙酰基和吡啶基二硫化物(参见例如 1994 Pierce 化学公司目录，关于交联剂的技术部分, 155 – 200 页,在此引入作为参考)，它们可以用于将含有半胱氨酸的蛋白试剂连接于支持物。或者，候选试剂上的氨基可以用于将氨基连接在表面上。例如，大量稳定的双功能基团是本领域公知的，包括同双功能和异双功能连接体(参见 Pierce 目录和手册, 155 – 200 页)。在一种额外的实施方案中，可以用公知的连接体(参见 Pierce 目录)衍生羧基(来自表面或来自候选试剂)。例如，碳二亚胺激活羧基，用于受到良好亲核试剂如胺的攻击(参见 Torchilin et al., Critical Rev. Therapeutic Drug Carrier Systems, 7(4):275 – 308 (1991))。也可以用本领域已知的其它技术，连接蛋白候选试剂，例如用于将抗体连接于聚合物；参见 Slinkin et al., Bioconj. Chem. 2:342 – 348 (1991); Torchilin et al., supra; Trubetskoy et al., Bioconj. Chem. 3:323 – 327 (1992); King et al., Cancer Res. 54:6176 – 6185 (1994); 和 Wilbur et al., Bioconjugate Chem. 5:220 – 235 (1994)。应该理解，候选试剂可以通过多种方式连接，包括上文列出的那些。优选地，连接方式不显著改变候选试剂的官能度；即，候选试剂应该以允许其与靶相互作用的灵活方式连接。

用于将酶固定在珠子上的特定技术是现有技术已知的。在一种情况下，使用  $\text{NH}_2$  表面化学珠。采用含 2.5% 低聚甲醛的磷酸缓冲液(10 mM)，其提供 6.9 的 pH (138 mM NaCl, 2.7 mM KCl)，实现表面活化。室温下在搅拌床上将该混合物搅拌大约 2 小时。然后用超纯净水 + 0.01% Tween 20 (表面活性剂) – 0.02% 洗涤珠子，用 pH 7.7 PBS + 0.01% tween 20 再次冲洗。最后，将酶加入溶液，优选在用 0.45  $\mu\text{m}$  Amicon Micropure 滤器预先过滤后加入。

将可移动的固体支持物的群体至于反应室中。在一些实施方案中，5% – 20% 的反应室可以具有可移动的固体支持物，其上固定至少一种试剂、20% – 60% 的反应室可以具有可移动的固体支持物，其上固定至少一种试剂，或者 50% – 100% 的反应室可以具有可移动的固体支持物，

其上固定至少一种试剂。优选地，至少一个反应室可以具有可移动的固体支持物，其上固定至少一种试剂，并且试剂适用于核酸测序反应中。

在一些实施方案中，固定于可移动的固体支持物的试剂可以是具有硫酸化酶活性的多肽、具有荧光素酶活性的多肽、或具有硫酸化酶和荧光素酶活性的嵌合多肽。在一种实施方案中，它可以是 ATP 硫酸化酶和荧光素酶融合蛋白。由于硫酸化酶的产物被荧光素酶消耗，可以通过以融合蛋白形式共价连接这两种酶而实现这两种酶的邻近。本发明不仅仅可以用于基质通道，也可以用于减少生产花费，并且可能使链霉亲和素珠上的结合位点数目加倍。

在另一种实施方案中，硫酸化酶是耐热 ATP 硫酸化酶。在一种优选实施方案中，耐热硫酸化酶在高于室温的温度下（到至少 50°C）是有活性的。在一种实施方案中，ATP 硫酸化酶来自嗜热生物。在一种额外的实施方案中，可移动的固体支持物可具有固定在其上的第一试剂和第二试剂，第一试剂是具有硫酸化酶活性的多肽，第二试剂是具有荧光素酶的多肽。

在另一实施方案中，固定于可移动的固体支持物的试剂可以是核酸；优选地，所述核酸是单链多联体。在一种优选实施方案中，核酸可以用于对核酸进行测序，例如焦磷酸测序反应。

本发明也提供了采用可移动的固体支持物对 ATP 活性进行检测和定量的方法；优选地，可以作为核酸测序反应的一部分，对 ATP 进行检测或定量。

用可移动的固体支持物“覆盖”的微量滴定板示于图 15，所述可移动的固体支持物上连接了核酸或试剂酶。

## 5. 对核酸进行测序的方法

随后进行基于焦磷酸的测序。然后在核苷酸三磷酸存在下使样品 DNA 序列和延伸引物进行聚合酶反应，从而核苷酸三磷酸仅仅在与靶位置上的碱基互补的情况下掺入，且释放焦磷酸 (PPi)，核苷酸三磷酸加入样品 - 引物混合物的分开的等分物中或连续加入相同的样品 - 引物混合物。然后检测 PPi 的释放，从而表明掺入了哪种核苷酸。

在一种实施方案中，使测序引物与模板核酸的一个区域退火，然后使测序引物与 DNA 聚合酶和已知的核苷酸三磷酸，即 dATP, dCTP,

dGTP, dTTP 或其中之一的类似物接触，从而确定序列产物的一个区域。可以通过按照下文的描述检测序列反应副产物，确定序列。

序列引物可以是任何长度或碱基组成，只要它能够特异性退火于扩增的核酸模板的一个区域。不需要测序引物具有特定结构，只要它能够特异性引发扩增的模板核酸上的一个区域。优选地，测序引物与模板上位于要鉴定的序列和可以与锚定引物杂交的序列之间的模板区域互补。测序引物用 DNA 聚合酶延伸，形成序列产物。在一种或多种类型的核苷酸三磷酸存在下，如果需要，在辅助结合蛋白存在下进行延伸。

优选通过测定测序副产物的存在，确定 dNTP 的掺入。在一种优选实施方案中，通过测量当 dNMP 掺入延伸的序列引物时从核苷酸三磷酸 (dNTP) 释放的无机焦磷酸 (PP<sub>i</sub>)，确定测序产物的核苷酸序列。该测序方法称作 Pyrosequencing™ 技术 (PyroSequencing AB, Stockholm, Sweden)，可以在溶液中 (液相) 进行，或作为固相技术。基于 PP<sub>i</sub> 的测序方法总体描述于例如 WO9813523A1, Ronaghi, et al., 1996. Anal. Biochem. 242: 84 - 89, Ronaghi, et al., 1998. Science 281: 363 - 365 (1998) 和 USSN 2001/0024790。在此全文引入这些关于 PP<sub>i</sub> 测序的公开内容作为参考。也参见例如美国专利 6,210,891 和 6,258,568，在此全文引入作为参考。

可以酶促检测这些条件下释放的焦磷酸 (例如通过在荧光素酶 - 荧光素反应中产生光)。所述方法使得能够在给定靶位置鉴定核苷酸，并且简单和快速对 DNA 测序，同时避免对电泳的需要和潜在有危险的放射性标记的使用。

可以通过多种不同的方法检测 PP<sub>i</sub>，并且已经描述了多种酶促方法 (参见例如 Reeves, et al., 1969. Anal. Biochem. 28: 282 - 287; Guillory, et al., 1971. Anal. Biochem. 39: 170 - 180; Johnson, et al., 1968. Anal. Biochem. 15: 273; Cook, et al., 1978. Anal. Biochem. 91: 557 - 565; 和 Drake, et al., 1979. Anal. Biochem. 94: 117 - 120)。

由于通过聚合酶掺入 dNTP 而释放的 PP<sub>i</sub> 可以用例如 ATP 硫酸化酶转化为 ATP。已经鉴定出该酶参与硫代谢。还原和氧化形式的硫是植物和动物生长的关键矿物营养 (参见例如 Schmidt and Jager, 1992. Ann. Rev. Plant Physiol. Plant Mol. Biol. 43: 325 - 349)。在植物和微生物中，硫酸盐活性摄取后还原为硫化物。由于相对于可获得的细胞还原剂，硫酸盐具有非常低的氧化/还原可能，同化的主要步骤需要其通过 ATP 依

赖性反应活化(参见例如 Leyh, 1993. Crit. Rev. Biochem. Mol. Biol. 28: 515 – 542)。ATP 硫酸化酶(ATP: 硫酸腺苷基酰转移酶; EC 2.7.7.4)催化无机硫酸根( $\text{SO}_4^{2-}$ )的代谢中的最初的还原; 参见例如 Robbins and Lipmann, 1958. J. Biol. Chem. 233: 686 – 690; Hawes and Nicholas, 1973. Biochem. J. 133: 541 – 550)。在该反应中,  $\text{SO}_4^{2-}$  活化为腺苷酰硫酸(APS)。

已经从一些来源高度纯化了 ATP 硫酸化酶, 所述来源例如啤酒糖酵母(参见例如 Hawes and Nicholas, 1973. Biochem. J. 133: 541 – 550); 产黄青霉(参见例如 Renosto, et al., 1990. J. Biol. Chem. 265: 10300 – 10308); 大鼠肝(参见例如 Yu, et al., 1989. Arch. Biochem. Biophys. 269: 165 – 174); 和植物(参见例如 Shaw and Anderson, 1972. Biochem. J. 127: 237 – 247; Osslund, et al., 1982. Plant Physiol. 70: 39 – 45)。此外, 从原核生物(参见例如 Leyh, et al., 1992. J. Biol. Chem. 267: 10405 – 10410; Schwedock and Long, 1989. Mol. Plant Microbe Interaction 2: 181 – 194; Laue and Nelson, 1994. J. Bacteriol. 176: 3723 – 3729); 真核生物(参见例如 Charest, et al., 1987. Mol. Gen. Genet. 210: 307 – 313; Mountain and Korch, 1991. Yeast 7: 873 – 880; Foster, et al., 1994. J. Biol. Chem. 269: 19777 – 19786); 植物(参见例如 Leustek, et al., 1994. Plant Physiol. 105: 897 – 90216); 和动物(参见例如 Li, et al., 1995. J. Biol. Chem. 270: 29453 – 29459)克隆了 ATP 硫酸化酶基因。该酶是同寡聚物或异二聚物, 这取决于特定来源(参见例如 Leyh and Suo, 1992. J. Biol. Chem. 267: 542 – 545)。

在一些实施方案中, 使用耐热硫酸化酶。可以从例如古生球菌或热球菌种获得耐热硫酸化酶。耐热硫酸化酶的序列可以从数据库 Acc. No. 028606, Acc. No. Q9YCR4, 和 Acc. No. P56863 获得。

已经将 ATP 硫酸化酶用于很多不同的应用, 例如在高浓度 ATP 存在下进行 ADP 的生物发光计检测(参见例如 Schultz, et al., 1993. Anal. Biochem. 215: 302 – 304); 连续监测 DNA 聚合酶活性(参见例如 Nyrbn, 1987. Anal. Biochem. 167: 235 – 238); 和 DNA 测序(参见例如 Ronaghi, et al., 1996. Anal. Biochem. 242: 84 – 89; Ronaghi, et al., 1998. Science 281: 363 – 365; Ronaghi, et al., 1998. Anal. Biochem. 267: 65 – 71)。

已经开发了一些测定, 用于检测正向 ATP 硫酸化酶反应。比色钼裂

解(molybdolysis)测定是基于磷酸检测(参见例如 Wilson and Bandurski, 1958. J. Biol. Chem. 233: 975 - 981), 而连续的分光光度计钼裂解测定是基于 NADH 氧化的检测(参见例如 Seubert, et al., 1983. Arch. Biochem. Biophys. 225: 679 - 691; Seubert, et al., 1985. Arch. Biochem. Biophys. 240: 509 - 523)。后一种测定需要一些检测酶的存在。此外, 文献中也描述了一些放射性测定(参见例如 Daley, et al., 1986. Anal. Biochem. 157: 385 - 395)。例如, 一种测定是基于检测从<sup>32</sup>P - 标记的 ATP 释放的<sup>32</sup>PPi(参见例如 Seubert, et al., 1985. Arch. Biochem. Biophys. 240: 509 - 523), 另一种是基于<sup>35</sup>S 在[<sup>35</sup>S] - 标记的 APS 中的掺入(该测定也需要纯化的 APS 激酶作为偶联酶参见例如 Seubert, et al., 1983. Arch. Biochem. Biophys. 225: 679 - 691); 第三种反应依赖于<sup>35</sup>SO<sub>4</sub><sup>-2</sup>从[<sup>35</sup>S] - 标记的 APS 中的释放(参见例如 Daley, et al., 1986. Anal. Biochem. 157: 385 - 395)。

为了检测反向 ATP 硫酸化酶反应, 以前已经描述了连续分光光度计测定(参见例如 Segel, et al., 1987. Methods Enzymol. 143: 334 - 349); 生物发光计测定(参见例如 Balharry and Nicholas, 1971. Anal. Biochem. 40: 1 - 17);<sup>35</sup>SO<sub>4</sub><sup>-2</sup> 释放测定(参见例如 Seubert, et al., 1985. Arch. Biochem. Biophys. 240: 509 - 523); 和<sup>32</sup>PPi 掺入测定(参见例如 Osslund, et al., 1982. Plant Physiol. 70: 39 - 45)。

可用酶促反应水解由 ATP 硫酸化酶产生的 ATP, 以产生光。发光化学反应(即化学发光)和生物反应(即生物发光)广泛用于分析生物化学中, 用于灵敏测量多种代谢物。在生物发光反应中, 用酶催化导致发光的化学反应。例如, 荧光素 - 荧光素酶系统使得能够特异性测定 ATP, 并且细菌荧光素酶 - 氧化还原酶系统可以用于监测 NAD(P)H。这两种系统都延伸到通过参与 ATP 或 NAD(P)H 的制备或利用的偶联的反应分析多种物质(参见例如 Kricka, 1991. Chemiluminescent and bioluminescent techniques. Clin. Chem. 37: 1472 - 1281)。

新试剂的开发使得能够获得与 ATP(参见例如 Lundin, 1982. Applications of firefly luciferase In; Luminescent Assays (Raven Press, New York))或 NAD(P)H(参见例如 Lovgren, et al., Continuous monitoring of NADH - converting reactions by bacterial luminescence. J. Appl. Biochem. 4: 103 - 111)浓度成比例的稳定发光。采用所述稳定的发光试

剂，可以产生终点测定，并且通过加入已知量的 ATP 或 NAD(P)H 校准每个单独的测定。此外，稳定的发光系统也使得能够连续监测 ATP - 或 NAD(P)H - 转化系统。

用于将 ATP 转化为光的合适的酶包括荧光素酶，例如昆虫荧光素酶。荧光素酶产生作为催化的终末产物的光。最了解的发光酶是萤火虫 *Photinus pyralis* (鞘翅目) 的酶。已经在细菌(参见例如 de Wet, et al., 1985. Proc. Natl. Acad. Sci. USA 80: 7870 – 7873)和植物(参见例如 Ow, et al., 1986. Science 234: 856 – 859)，以及昆虫(参见例如 Jha, et al., 1990. FEBS Lett. 274: 24 – 26)和哺乳动物细胞(参见例如 de Wet, et al., 1987. Mol. Cell. Biol. 7: 725 – 7373; Keller, et al., 1987. Proc. Natl. Acad. Sci. USA 82: 3264 – 3268)中克隆和表达了相应基因。此外，最近克隆和部分表征了来自牙买加叩甲虫，即 *Pyroplorus plagiophihalamus* (鞘翅目) 的许多荧光素酶基因(参见例如 Wood, et al., 1989. J. Biolumin. Chemilumin. 4: 289 – 301; Wood, et al., 1989. Science 244: 700 – 702)。不同的荧光素酶有时可以产生不同波长的光，其使得能够同时监测不同波长的发光。因此，这些前述特征是独特的，并且在利用目前的报道系统方面加入了新的元素。

萤火虫的荧光素酶在荧光素、腺昔 5' - 三磷酸(ATP)、镁离子和氧存在下催化生物发光，导致 0.88 的量子产率(参见例如 McElroy and Selinger, 1960. Arch. Biochem. Biophys. 88: 136 – 145)。萤火虫荧光素酶生物发光反应可以用作检测 ATP 的测定，所述 ATP 的检测的检测限是大约  $1 \times 10^{-13}$  M (参见例如 Leach, 1981. J. Appl. Biochem. 3: 473 – 517)。此外，荧光素酶介导的检测系统的总体灵敏度和方便性在开发基于萤火虫荧光素酶的生物传感器方面产生了显著的兴趣(参见例如 Green and Kricka, 1984. Talanta 31: 173 – 176; Blum, et al., 1989. J. Biolumin. Chemilumin. 4: 543 – 550)。

采用上文描述的酶，将序列引物暴露于聚合酶和已知的 dNTP。如果 dNTP 摊入引物序列的 3'末端，则切割 dNTP，并且释放 PPi 分子。然后用 ATP 硫酸化酶将 PPi 转化为 ATP。优选地，ATP 硫酸化酶以足够高的浓度存在，以便 PPi 转化随着关于 PPi 的一级动力学进行。在荧光素酶存在下，ATP 水解产生光子。反应优选具有足够浓度的存在于反应混合物中的荧光素酶，以便反应  $\text{ATP} \rightarrow \text{ADP} + \text{PO}_4^{3-} + \text{光子(光)}$  随

着关于 ATP 的一级动力学进行。可以用下文描述的方法和装置测量光子。在一种实施方案中，用 PPi 和偶联的硫酸化酶/荧光素酶反应产生用于检测的光。在一些实施方案中，硫酸化酶和荧光素酶之一或这两者固定在置于每个反应位点的一个或多个可移动的固体支持物上。

因此，本发明使得能够在产生实时信号的聚合酶反应过程中检测 PPi 释放。可以实时连续监测测序反应。因此本发明使得能够进行快速检测 PPi 释放的程序。估计反应在不到 2 秒的时间内发生(Nyren and Lundin, *supra*)。限速步骤是通过 ATP 硫酸化酶将 PPi 转化为 ATP，而荧光素酶反应很快，估计在不到 0.2 秒的时间内发生。已经通过多种方法估计了聚合酶的掺入速度，并且发现例如在克列诺聚合酶的情况下，完整掺入一个碱基可以在不到 0.5 秒的时间内发生。因此，估计用于掺入一个碱基并且通过该酶促测定检测的总时间是大约 3 秒。因此可以发现，非常快的反应时间是可能的，使得能够进行实时检测。反应时间可以通过采用更多的耐热萤光酶而进一步减少。

对于大多数应用，理想的是使用不含例如 ATP 和 PPi 的污染物的试剂。可以通过使试剂流过含有与树脂结合的腺苷三磷酸双磷酸酶和/或焦磷酸酶的预备柱而除去这些污染物。或者，腺苷三磷酸双磷酸酶或焦磷酸酶可以结合于磁珠，并且用于除去试剂中存在的污染性 ATP 和 PPi。此外，理想的是用洗涤缓冲液洗去可扩散的测序试剂，例如未掺入的 dNTPs。可以使用任何用于焦磷酸测序中的洗涤缓冲液。

在一些实施方案中，测序反应中的试剂浓度包括 0.2 ml 缓冲液中的 1 pmol DNA, 3 pmol 聚合酶, 40 pmol dNTP。参见 Ronaghi, et al., *Anal. Biochem.* 242: 84 – 89 (1996)。

如果需要，可以用四种预定核苷酸中的每一种进行测序反应。“完整”循环一般包括以预定的顺序，序贯使用用于核苷酸 dATP, dGTP, dCTP 和 dTTP(或 dUTP)中的每一种核苷酸的测序试剂。在每次添加核苷酸之间，洗掉未掺入的 dNTPs。或者，通过腺苷三磷酸双磷酸酶降解未掺入的 dNTPs (参见下文)。按照需要重复循环，直到获得需要量的测序产物的序列。在一些实施方案中，从一个退火测序引物的延伸获得序列的约 10 – 1000, 10 – 100, 10 – 75, 20 – 50 或约 30 个核苷酸的信息。

在一些实施方案中，修饰核苷酸，使其含有半抗原如生物素的二硫化物衍生物。通过聚合后步骤分析修饰的核苷酸在退火于锚定的底物的

初期引物中的添加，所述步骤包括 i) 在修饰是生物素的例子中，序贯结合与酶分子连接的亲和素或链霉亲和素缀合的部分，ii) 洗去过量的亲和素或链霉亲和素连接的酶，iii) 在适于酶活性的条件下流过合适的酶底物，和 iv) 检测酶底物反应产物。在该实施方案中，通过加入还原剂除去半抗原。一些方法使得能够在给定靶位置鉴定核苷酸，并且简单和迅速对 DNA 进行测序，同时避免对电泳和使用可能有危险的放射性标记的需要。

用于检测半抗原的优选酶是辣根过氧化物酶。如果需要，可以在添加多种本文的试剂之间使用洗涤缓冲液。腺苷三磷酸双磷酸酶可以用于除去未反应的 dNTP，所述 dNTP 用于延伸测序引物。洗涤缓冲液可以任选包含腺苷三磷酸双磷酸酶。

示例的半抗原，如生物素、洋地黄毒昔、荧光染料分子 cy3 和 cy5，以及荧光素，以各种效率掺入延伸的 DNA 分子。半抗原的连接可以通过与核苷酸上的糖、碱基和磷酸部分的键进行。信号放大的示例手段包括荧光、电化学和酶促。在一种采用酶促放大的优选实施方案中，例如碱性磷酸酶(AP)、辣根过氧化物酶(HRP)、 $\beta$ -半乳糖苷酶、荧光素酶的酶可以包括发光底物已知的那些酶，并且用于检测这些发光(化学发光)底物的工具可以包括 CCD 照相机。

在一种优选方式中，加入修饰的碱基，发生检测，除去半抗原缀合的部分，或通过使用裂解剂或灭活剂灭活。例如，如果可裂解的连接体是二硫化物，那么裂解剂可以是还原剂，例如二硫苏糖醇(DTT)， $\beta$ -巯基乙醇等。灭活的其它实施方案包括热、冷、化学变性剂、表面活性剂、疏水试剂和自杀抑制剂。

荧光素酶可以直接水解 dATP，伴随释放光子。这导致假阳性信号，因为水解不依赖于 dATP 在延伸的测序引物中的掺入而发生。为了避免该问题，可以使用掺入 DNA 的 aATP 类似物，即它是 DNA 聚合酶的底物，但不是荧光素酶的底物。一种所述的类似物是  $\alpha$ -硫-dATP。因此， $\alpha$ -硫-dATP 的使用避免了当 dATP 被水解而不掺入生长中的核酸链时过多光子产生。

典型地，通过测量将对照核苷酸加入测序反应混合物后释放的光，校准基于 PPi 的检测，所述对照核苷酸是在加入测序引物后立即加入的。这使得能够对反应条件进行标准化。通过释放的光量的相应增加，发现

连续掺入 2 个或更多相同的核苷酸。因此，释放的光相对于对照核苷酸增加 2 倍，表明在延伸的引物中掺入了 2 个连续的 dNTPs。

如果需要，腺苷三磷酸双磷酸酶可以“洗涤”或“流过”固体支持物的表面，以便促进测序反应混合物内任何保留的、未掺入的 dNTPs 的降解。腺苷三磷酸双磷酸酶也降解产生的 ATP，因此“关闭”从反应产生的光。用腺苷三磷酸双磷酸酶处理后，在准备随后的 dNTP 温育和光子检测步骤中，洗去任何保留的反应物。或者，腺苷三磷酸双磷酸酶可以与固体或可移动的固体支持物结合。

### 双末端测序

在一种优选实施方案中，我们提供了从核酸模板的两个末端开始测序的方法。传统地，双链 DNA 分子的两个末端的测序需要很少的引物杂交、一个末端的测序、第二引物的杂交和另一末端的测序。替代的方法是分开双链核酸的各个链，并且每条链单独测序。本发明提供了第三种选择，它比前两种方法更迅速并且劳动强度更小。

本发明提供了从多个引物对核酸进行序贯测序的方法。本申请中提到的 DNA 测序涉及用聚合酶进行测序，其中随着核苷酸三磷酸 (NTP) 掺入测序引物的生长链而确定序列。这种类型的测序的一个实例是焦磷酸测序检测方法（参见例如美国专利 6,274,320, 6258,568 和 6,210,891，在此全文引入每篇上述专利作为参考）。

在一种实施方案中，本发明提供了对模板双链核酸的两个末端进行测序的方法。双链 DNA 由两个单链 DNA 组成；在此称作第一单链 DNA 和第二单链 DNA。使第一引物与第一单链 DNA 杂交，第二引物与第二单链 DNA 杂交。第一引物是未保护的，而第二引物是保护的。“保护”和“保护的”在本公开内容中定义为在引物上的反应位点中加入防止引物通过 DNA 聚合酶进行聚合的化学基团。此外，所述化学保护基团的加入应该是可逆的，使得在逆转之后，目前受保护的引物再次能够作为测序引物。通过采用常规方法如焦磷酸测序，用 DNA 聚合酶延伸第一引物，从而在一个方向上（如从模板的一个末端开始）确定核酸序列。然后使第二引物去保护，通过用 DNA 聚合酶和常规方法如焦磷酸测序，以另一方向（例如，从模板的另一个末端开始）延伸第二引物，从而确定序列。第一和第二引物的序列特别设计为与双链 DNA 的两个末端，

或任何沿着该方法的模板的位置杂交。

在另一实施方案中，本发明提供了从多个引物对核酸测序的方法。在该方法中，使许多测序引物与要测序的模板核酸杂交。除了一个以外，所有测序引物都是可逆保护的。保护的引物是不能用聚合酶和通常用于DNA测序反应中的dNTPs延伸的寡核苷酸引物。可逆保护的引物是能够去保护的保护的引物。本发明中提到的所有保护的引物都是可逆保护的。去保护后，可逆保护的引物作为正常测序引物起作用，并且能够参与正常测序反应。

本发明提供了从多个引物对核酸进行序贯测序的方法。本发明包括以下步骤：第一，提供一个或多个要测序的模板核酸。第二，使多个测序引物与模板核酸杂交。测序引物的数目可以由数字n表示，其中n可以是任何大于1的正数。该数字可以是例如2, 3, 4, 5, 6, 7, 8, 9, 10或更大。所以，例如，如果n是2, 3, 4, 5, 6, 7, 8, 9或10，则n-1分别是1, 2, 3, 4, 5, 6, 7, 8, 9。其余的引物（如n个引物 - (n-1)个保护的引物 = 1个其余的引物）是未保护的。第三，延伸未保护的引物，通过常规方法，例如焦磷酸测序，确定模板DNA序列。第四，对第一个引物测序后，使其余的保护的引物之一去保护。第五，延伸未保护的引物，通过常规，例如焦磷酸测序，确定模板DNA序列。任选地，可以重复该方法，直到在所有保护的引物上进行了测序。

另一方面，本发明包括对核酸进行序贯测序的方法，包括以下步骤：(a)使2个或更多测序引物与核酸杂交，其中除一个之外，所有引物都是可逆保护的；(b)通过从未保护的引物进行聚合酶延伸，确定核酸的一条链的序列；(c)使可逆保护的引物之一去保护为未保护的引物；(d)重复步骤(b)和(c)直到所有可逆保护的引物都去保护，并且用于确定序列。在一种实施方案中，该方法包括步骤(b)和(c)之间的一个额外步骤，即，以下步骤：通过使未保护的引物接触DNA聚合酶和一个或多个核苷酸三磷酸或双脱氧核苷酸三磷酸，终止未保护的引物的延伸。在另一种实施方案中，该方法进一步包括步骤(b)和(c)之间的一个额外步骤，即，以下步骤：通过使未保护的引物接触DNA聚合酶和选自ddATP, ddTTP, ddCTP, ddGTP或其组合的双脱氧核苷酸三磷酸，终止未保护的引物的延伸。

另一方面，本发明包括对核酸进行测序的方法，包括：(a)使未保护的第一引物与核酸的第一条链杂交；(b)使保护的第二引物与第二条链杂

交；(c)使第一和第二条链暴露于聚合酶，使得受保护的第一引物沿着第一条链延伸；(d)完成第一测序引物的延伸；(e)使第二测序引物去保护；和(f)使第一和第二条链暴露于聚合酶，使得第二测序引物沿第二条链延伸。在一种优选实施方案中，完成包括加帽或终止延伸。

在另一实施方案中，本发明提供了用于对包括第一和第二单链DNA的模板双链核酸的两个末端进行测序的方法。在该实施方案中，在相同步骤中，使第一引物与第一单链DNA杂交，第二引物与第二单链DNA杂交。第一引物是未保护的，而第二引物是保护的。

在杂交后，采用常规方法，如焦磷酸测序，用DNA聚合酶延伸第一引物，从而在一个方向上（如，从模板的一个末端开始）确定核酸序列。在一种优选实施方案中，聚合酶不含 $3' \rightarrow 5'$ 外切核酸酶活性。然后使第二引物去保护，采用常规方法，如焦磷酸测序，用DNA聚合酶在另一个方向上（如，从模板的另一个末端开始）延伸第二引物，从而确定其序列。如前面描述的，设计第一引物和第二引物的序列，使其与双链DNA的两个末端或沿模板的任何位置杂交。该技术对于在两个末端含有独特测序引物杂交位点的许多模板DNAs的测序特别有用。例如，很多克隆载体提供了侧翼于插入位点的独特测序引物杂交位点，以促进任何克隆的序列（如Bluescript, Stratagene, La Jolla, CA）的后续测序。

本发明的该方法的一个益处是两个引物可以在单个步骤中杂交。这种方法和其它方法的益处在平行测序系统中特别有用，其中杂交比正常情况下更多涉及。平行测序系统的实例公开于共同未决的美国专利申请序列号10/104,280，在此全文引入其公开内容作为参考。

可以通过常规技术，例如采用商购寡核苷酸合成仪和/或通过将如此合成的亚片段连接在一起，合成本发明的寡核苷酸引物。

在本发明的另一实施方案中，可以确定双链靶核酸的长度。确定双链核酸长度的方法是本领域已知的。可以在核酸测序之前或之后进行长度确定。核酸分子长度确定的已知方法包括凝胶电泳、脉冲凝胶电泳、质谱等。由于平端双链核酸由两个相同长度的单链组成，核酸的一条链的长度的确定足够确定相应双链的长度。

本发明的序列反应也使得能够确定模板核酸长度。第一，从核酸的一个末端到另一个末端的完整序列使得能够确定长度。第二，两个末端的序列确定可能在中间重叠，使得两个序列能够连接。可以确定完整序

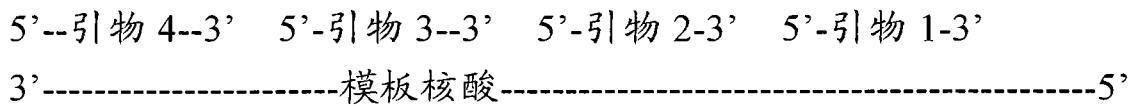
列，并且可以发现长度。例如，如果模板的长度是 100 bps，从一个末端开始的测序可以确定碱基 1 - 75；从另一个末端开始的测序可以确定碱基 25 - 100；因此，在中间从碱基 25 - 碱基 75 存在 51 个碱基的重叠；根据该信息，可以确定从 1 - 100 的完整序列，并且通过完整序列可以发现 100 个碱基的长度。

本发明的另一方法涉及包括以下步骤的方法。首先使各自具有不同序列的多个测序引物与要测序的 DNA 杂交。测序引物的数目可以是任何大于 1 的值，例如 2, 3, 4, 5, 6, 7, 8, 9, 10 或更多。除一个以外，所有这些引物都是可逆保护的。在测序反应中延伸这一个未保护的引物，并且确定其序列。通常，当完全延伸一个引物时，它不能再延伸，并且不会影响从另一个引物开始的后续测序。如果需要，可以用过量聚合酶和 dNTP 或用 ddNTPs 终止测序的引物。如果进行终止步骤，应该在该步骤后除去终止试剂（dNTPs 和 ddNTPs）。然后，对可逆保护的引物之一去保护，并且进行从第二引物开始的测序。重复使引物去保护、从去保护的引物开始测序，和任选终止从引物开始的测序的步骤，直到所有保护的引物都去保护并且用于测序。

应该用不同化学基团保护可逆保护的引物。通过选择合适的去保护方法，可以对一个引物去保护，而不影响其它引物的保护基。在一种优选实施方案中，保护基是 PO<sub>4</sub>。也就是说，第二个引物受 PO<sub>4</sub> 保护，并且通过 T4 多核苷酸激酶完成去保护（利用其 3' - 磷酸酶活性）。在另一种优选实施方案中，保护基是硫醇基或磷酸硫醇基。

模板核酸可以是 DNA、RNA 或肽核酸（PNA）。尽管 DNA 是优选的模板，RNA 和 PNA 可以通过常规技术转化为 DNA，所述技术例如随机引物 PCR、逆转录、RT - PCR 或这些技术的组合。此外，本发明的方法可以用于对序列未知和已知的核酸进行测序。已知序列的核酸的测序可以用于例如证实合成 DNA 的序列，或用于用已知核酸序列证实怀疑的病原体的身份。核酸可以是超过一个核酸群体的混合物。已知具有足够特异性的测序引物（例如 20 个碱基、25 个碱基、30 个碱基、35 个碱基、40 个碱基、45 个碱基或 50 个碱基）可以用于对长核酸或无关核酸群体中的一个亚组的序列进行测序。因此，例如，模板可以是一个 10 Kb 的序列或 10 个均为 1 Kb 的序列。在一种优选实施方案中，模板 DNA 的长度是 50 bp - 700 bp。DNA 可以是单链或双链的。

在模板核酸为单链的情况下，许多引物可以与模板核酸杂交，如下文所示：



在此情况下，优选的是最初未保护的引物是在模板的最 5' 末端杂交的引物。参见上文举例说明的引物 1。在该方向中，引物 1 的延伸将不会置换（通过链置换）引物 2、3 或 4。当完成从引物 1 开始的测序时，引物 2 可以是未保护的，并且可以开始核酸测序。从引物 2 开始的测序可以置换引物 1 或引物 1 的延伸形式，但对其余的保护的引物（引物 3 和 4）无作用。采用该顺序，每个引物可以序贯使用，并且从一个引物开始的测序反应不会影响从随后的引物开始的测序。

本发明的一个特征是在一个或多个核酸上使用多个测序引物的能力和仅仅用一个杂交步骤从多个引物开始测序的能力。在杂交步骤中，所有测序引物（如 n 个测序引物）可以同时与模板核酸杂交。在常规测序中，从一个引物开始的测序通常需要一个杂交步骤。本发明的一个特征是可以通过单个杂交步骤进行从 n 个引物（如上文定义）开始的测序。这有效去除了 n - 1 杂交步骤。

在一种优选实施方案中，n 个引物的序列是足够不同的，使得引物不交叉杂交或自杂交。交叉杂交是指一个引物与另一个引物由于序列互补性导致的杂交。一种形式的交叉杂交通常称作“引物二聚体”。在引物二聚体的情况下，两个引物的 3' 末端是互补的，并且形成当延伸时大约为两个引物长度总和的结构。自杂交是指引物的 5' 末端与引物的 3' 末端互补的情况。在该情况下，引物具有自杂交形成发夹样结构的倾向。

引物可以与模板分子相互作用或与其特异性结合。术语“相互作用”或“结合”在此表示两种物质或化合物（如引物和模板；化学部分和核苷酸）彼此充分结合（例如连接、结合、杂交、相连、退火、共价连接或其它方式结合），从而能够进行预期测定。术语“特异性”或“特异性地”在本文中表示两种成分选择性地彼此结合。需要实现特异性相互作用的参数可以常规确定，例如，采用本领域的常规方法确定。

为了获得更高灵敏度或辅助分析复合物结构，可以用设计产生清楚的独特信号的化学部分修饰（如衍生）保护的引物。例如，可以用通过酰胺键与寡核苷酸链在沿链的杂交部分的一个或多个位置上连接的不

同的天然或合成氨基酸衍生每个保护的引物。当然，可以在从靶核酸切割后，或与靶核酸结合时检测化学修饰。通过以可区分的方式鉴定每个保护的靶核酸，可以在单个测定中测定（例如筛选）大量不同的靶核酸。可以快速和容易地进行很多这样的测定。因此，可以以本文定义的高通量效率进行所述测定或一组测定。

在本发明的方法中，在延伸第一引物并且确定了模板 DNA 的序列后，使第二引物去保护并且测序。在第一引物的测序反应与第二引物（目前未保护）的测序反应之间没有干扰，因为第一引物已经完全延伸或终止。由于第一引物完全延伸，用常规方法如焦磷酸测序从第二引物开始的测序将不会受到延伸的第一引物的存在的影响。本发明也提供了从第一引物减少任何可能的信号污染的方法。信号污染表示第一引物未完全延伸的发生。在该情况下，当随后的引物去保护并且延伸时，第一引物将继续延伸。第一和第二引物的延伸都会干扰 DNA 序列的确定。

在一种优选实施方案中，从一个引物开始的测序反应（如链延伸反应）首先终止或完成，然后才从第二引物开始测序反应。通过使模板 DNA 与 DNA 聚合酶和双脱氧核苷酸三磷酸(ddNTPs)如 ddATP, ddTTP, ddGTP 和 ddCTP 接触，可以终止 DNA 的链延伸反应。终止后，可以通过用不含 ddNTPs 的溶液洗涤反应，除去双脱氧核苷酸三磷酸。防止引物的进一步延伸的第二种方法是在反应中添加核苷酸三磷酸 (dNTPs，如 dATP, dTTP, dGTP 和 dCTP) 和 DNA 聚合酶，以完全延伸任何没有完全延伸的引物。完全延伸后，在下一个引物去保护前，除去 dNTPs 和聚合酶。通过在另一个引物的去保护前完成或终止一个引物，可以改进测序反应（如焦磷酸测序）的信噪比。

可以重复以下步骤：(a)任选终止或完成测序，(b)使新引物去保护，和(c)从去保护的引物测序，直到从每个引物的延伸确定序列。在该方法中，杂交步骤包括“n”个引物和一个未保护的引物。未保护的引物首先测序，并且可以重复上述步骤(a)、(b)和(c)。

在一种优选实施方案中，焦磷酸测序用于根据本发明的方法进行的所有测序中。

在另一优选实施方案中，根据图 10 概括的方法进行双末端测序。该方法可以分成 6 个步骤：(1)建立捕获珠（图 10A）；(2)驱动到珠(DTB) PCR 扩增（图 10B）；(3)SL 报道系统制备（图 10C）；(4)对第一链的

测序（图 10D）；(5)制备第二链（图 10E 和图 10F）；和(6)分析每条链（图 10G）。该示例的方法概括于下文。

在步骤 1，使 N - 羟基琥珀酰亚胺(NHS)活化的捕获珠(如 Amersham Biosciences, Piscataway, NJ)与正向引物和反向引物偶联。NHS 偶联形成与含有伯氨基的配体的化学稳定的酰胺键。捕获珠也与生物素偶联（图 10A）。此处使用的珠子（即固体核酸捕获支持物）可以是任何方便的大小，并且用任何已知的材料制作。所述材料的实例包括：无机物、天然聚合物和合成聚合物。这些材料的具体实例包括：纤维素、纤维素衍生物、丙烯酸树脂、玻璃；硅胶、聚苯乙烯、明胶、聚乙烯吡咯烷酮、乙烯和丙烯酰胺的共聚物、与二乙烯基苯交联的聚苯乙烯等（参见 Merrifield Biochemistry 1964, 3, 1385 – 1390）、聚丙烯酰胺、乳胶凝胶、聚苯乙烯、右旋糖苷、橡胶、硅、塑料、硝酸纤维素、纤维素、天然海绵、硅胶、玻璃、金属塑料、纤维素、交联右旋糖苷（例如 Sephadex<sup>TM</sup>）和琼脂糖凝胶(Sepharose<sup>TM</sup>)以及本领域技术人员已知的固相支持物。在一种优选实施方案中，捕获珠是直径为大约 25 – 40μm 的琼脂糖珠。

在步骤 2，添加与正向和反向引物杂交的模板 DNA，通过 PCR 扩增策略，扩增 DNA（图 10B）。在一种实施方案中，通过乳状液聚合酶链反应、驱动珠聚合酶链反应、滚环扩增或环介导的等温扩增进行 DNA 扩增。在步骤 3，在添加要与链霉亲和素偶联的硫酸化酶和荧光素酶之前，添加链霉亲和素（图 10C）。在测序方法中添加辅助的酶已经公开于 U.S.S.N. 10/104,280 和 U.S.S.N. 10/127,906，在此全文引入作为参考。在一种实施方案中，模板 DNA 具有与 5' 和 3' 末端连接的 DNA 衔接子。在一种优选实施方案中，通过 DNA 衔接子之一与 DNA 捕获珠上的互补序列的杂交，DNA 与 DNA 捕获珠偶联。

在第一步中，要扩增的单链核酸模板与捕获珠连接。核酸模板可以通过本领域已知的任何方式与捕获珠连接。本领域存在许多用于将 DNA 连接于显微珠的方法。可以采用标准偶联剂，如水溶性碳二亚胺将 DNA 与珠子共价化学连接，以便通过磷酰胺键将 DNA 上的 5' 磷酸与胺包被的微球连接。另一种备选方案是用相似的化学首先将第一偶联特异性寡核苷酸连接体与珠子偶联，然后用 DNA 连接酶将 DNA 连接于珠子上的连接体。其它连接化学包括使用 N - 羟基琥珀酰亚胺 (NHS) 及其衍生物，以便将寡核苷酸连接于珠子。在这样的方法中，寡核苷酸的一个末端可

以含有与固体支持物形成共价键的反应基团（如酰胺基团），而连接体的另一个末端含有可以与待固定的寡核苷酸键合的另一反应基团。在一种优选实施方案中，寡核苷酸通过共价键与 DNA 捕获珠结合。但是，非共价键，如螯合或抗原 - 抗体复合物，也可以用于将寡核苷酸连接于珠子。

可以采用与 DNA 片段末端的独特序列特异性杂交的寡核苷酸连接体，所述末端例如来自限制酶位点的重叠末端或基于噬菌体入的克隆载体的“粘端”，但也可以有益地使用平端连接。这些方法详细描述于 US 5,674,743，其公开内容在此全部引入作为参考。优选的是任何用于固定珠子的方法将继续通过本发明方法中的步骤结合固定的寡核苷酸。在一种优选实施方案中，寡核苷酸通过共价键与 DNA 捕获珠结合。但是，非共价键，如螯合或抗原 - 抗体复合物，也可以用于将寡核苷酸连接于珠子。

在步骤 4，通过将捕获珠沉积在微量滴定板(PTP)，并且通过本领域普通技术人员已知的方法（例如焦磷酸测序）进行测序，对 DNA 的第一条链进行测序（图 10D）。测序后，按顺序加入 dNTPs 和 ddNTPs 的混合物，以便对测序过程“加帽”或终止（图 10E）。在步骤 5，通过加入腺苷三磷酸双磷酸酶从阻断的引物链除去 ddNTPs 和加入多核苷酸激酶(PNK)从阻断的引物链除去 3' 磷酸基团，制备核酸的第二条链（图 10F）。然后，加入聚合酶引发第二条链，然后根据本领域普通技术人员已知的标准方法对第二条链进行测序（图 10G）。在步骤 7，分析第一条链和第二条链的序列，使得确定连续 DNA 序列。

## 检测工具

固体支持物光学连接于成像系统 230，所述系统包括与常规光学器件或光导纤维束联合的 CCD 系统。在一种实施方案中，灌注室基质包括光导纤维阵列薄片，使得水界面附近产生的光通过光导纤维直接传递到基质或室的外部。当 CCD 系统包括光导纤维连接器时，可以通过使灌注室基质与连接器直接接触，实现成像。或者，可以用常规光学器件对光进行成像，例如，采用 1-1 放大高数值光圈透镜系统，从光导纤维基质外部直接成像到 CCD 传感器上。当基质不提供光导纤维偶联时，也可以按照上文描述使用透镜系统，在该情况下基质或灌注室的盖子是

透光的。上文描述了示例的 CCD 成像系统。

成像系统 230 用于从基质表面的反应器上采集光。例如，可以用本领域已知的高灵敏度低噪装置将光成像到 CCD 上。对于基于光导纤维的成像，优选将光导纤维直接引入盖玻片，或用于 PORA，以便使形成微孔的光导纤维也作为将光传送到检测器中的光导纤维。

成像系统连接于计算机控制和数据采集系统 240。概言之，可以使用任何公众可得到的硬件和软件包。计算机控制和数据采集系统也连接于导管 200，以控制试剂递送。

仅仅当通过聚焦器件（如光学透镜或光学纤维）时，由焦磷酸测序反应产生的光子被 CCD 捕获，并且聚焦在 CCD 部件上。但是，发射的光子将在所有方向上相等地逃逸。为了在利用平面阵列（如 DNA 芯片）时使随后的“捕获”和定量最大，优选在距离产生光子的点尽可能近的位置采集光子，例如，直接在平面固体支持物上。这是通过以下任一方式实现的：(i) 在盖玻片和传统光学透镜或光导纤维束之间利用光学浸渍油，或优选(ii) 将光导纤维直接引入盖玻片。类似地，当使用薄的、透光的平面时，也可以将光导纤维束放置在其背面，消除通过整个反应室/灌注室的深度进行成像的需要。

可以用多种检测装置，如光电信增管、CCD、CMOS、吸光度光度计、照度计、电荷注入器件(CID)，或其它固态检测器，以及本发明描述的装置，对反应事件，如荧光素酶产生的光子进行检测。在一种优选实施方案中，通过使用安装了熔凝光导纤维束的 CCD 照相机实现对发射的光子的定量。在另一优选实施方案中，通过使用安装了微通道板增强器的 CCD 照相机实现对发射的光子的定量。可以用薄型背照式 (back-thinned)CCD 增加灵敏度。CCD 检测器描述于例如 Bronks, et al., 1995. Anal. Chem. 65: 2750 – 2757。

示例的 CCD 系统是 Spectral Instruments, Inc. (Tucson, AZ) 600 系列 4 端口照相机，其具有 Lockheed – Martin LM485 CCD 芯片和 1 - 1 光导纤维连接器（束），其单个纤维直径是 6 – 8 $\mu\text{m}$ 。该系统具有 4096 x 4096 或更高的 1600 万像素，并且量子效率是 10% 到 > 40%。因此，根据波长，将多至 40% 的成像到 CCD 传感器上的光子转化为可检测的电子。

在其它实施方案中，荧光部分可以用作标记，可以用共焦扫描显微镜扫描检测反应事件，所述显微镜采用激光或其它可获得、能够进行更

小的光学分辨的技术如扫描近场光学显微术(SNOM)扫描阵列的表面，从而使得能够使用“更密集”的阵列。例如，采用 SNOM，当隔开小于 100 nm 的距离，如 10nm x 10nm 时，可以区分各个多核苷酸。此外，可以使用扫描隧道显微术 (scanning tunneling microscopy) (Binning et al., Helvetica Physica Acta, 55:726 – 735, 1982) 和原子力显微术(Hanswa et al., Annu Rev Biophys Biomol Struct, 23:115 – 139, 1994)。

### 单倍型应用

采用本发明的方法和装置，可以实现基本任何测序应用。在一种实施方案中，我们考虑单倍型作图。人基因多样性是对药物的患者反应的多样性的重要因素。该多样性的最精确测量是单倍型，其是染色体上存在的多态性变异的组构。最近，美国、加拿大和欧洲的主要政府和学术基因组研究者同意了单倍型是可以将遗传信息的复杂度减少到实际形式的有力工具。单倍型可以用于药物发现中，改进靶证实和药物筛选研究的结果，和用于药物开发中，以改进临床试验的设计和可靠性。单倍型标志物可以用于预测新药和批准药物的有效性和安全性，并且将作为通过来自临床标志物 - 关联数据库的指导使匹配患者的药物个体化为正确剂量下的正确药物的新规范的基础。

许多经验性研究证明附近的 SNP 等位基因通常彼此处于连锁不平衡(LD)，使得一个 SNP 等位基因的状态通常与另一个邻近 SNP 的等位基因高度相关。这些相关性的存在是因为共有的紧密连锁的 SNP 的历史，它们从一代共同传递到下一代。因此，人序列变异(单倍型)的模式代表祖先的 DNA 片段。历史的减数分裂缓慢使等位基因与祖先染色体上的相邻等位基因分离，紧密连锁的变体除外。具有最近的瓶颈的建立者群体中的连锁不平衡程度是许多研究的目的，特别是在简单孟德尔病症的克隆中，所述病症如囊性纤维化 (16)、亨廷顿病 (11)，畸形发育不良(DTD) (8)。尽管这些克隆研究从显示跨长距离的 LD 的大染色体片段(通常是兆碱基范围)获益，直到最近，仅仅获得了非常少的关于跨世界群体的人基因组的 LD 的经验数据。

我们致力于大规模 LD (和单倍型) 研究的最近三个实例 (参见例如 Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R. & Lander, E. S.

2001. Linkage disequilibrium in the human genome. *Nature* 411, 199 – 204.26)。我们对 19 个染色体区的 SNP 含量进行了取样。首先在高加索人样本中对跨 2 – 160 kb 间隔的高频率的 SNP 进行了基因型分析。在所有区域中，在大约 60kb 的距离可以检测到 LD，区域之间有显著差异，因为一个基因座上范围短至 6kb，而另一个基因座上长至 155 kb。不出意料，LD 与估计的局部重组率显著相关。在黑人样本中的进一步分析提供了在该群体中 LD 较短的证据，但短距离上的等位基因重组与高加索样本类似。总体上，该研究提供证据证明了 LD 的大区块在人基因组中是共有的，并且对疾病基因进行基因组范围的 LD 作图是可行的。

## 试剂盒

本发明也包括用于本发明的方法中的试剂盒，其可以包括一个或多个以下成分：(a)与样品 DNA 杂交的特异性测试引物，使得靶位置直接邻接于引物的 3' 末端；(b)聚合酶；(c)用于鉴定 PPi 释放的检测酶工具；(d)脱氧核苷酸，包括替代 dATP 的 dATP 类似物，其能够作为聚合酶的底物，但是不能作为所述 PPi 检测酶的底物；和(e)任选包括双脱氧核苷酸，ddATP 任选被 ddATP 类似物替代，所述类似物能够作为聚合酶的底物，但是不能作为所述 PPi 检测酶的底物。如果试剂盒用于 PCR 扩增，则它可以包括以下成分：(i)一对用于 PCR 的引物，至少一个引物具有允许所述酶固定的工具；(ii)优选耐热的聚合酶，例如 Taq1 聚合酶；(iii)用于 PCR 反应的缓冲液；和(iv)脱氧核苷酸。当酶标记用于评估 PCR 时，试剂盒将有利地含有酶的底物和检测系统的其它成分。

本发明的一种实施方案涉及对核酸进行测序的方法。该方法包括使大模板核酸分子断裂，产生多个断裂的核酸。然后，在油包水乳状液中将断裂的核酸递送到含水的微反应器中，使得多个含水微反应器包含单个拷贝的断裂的核酸、单个能够结合断裂的核酸的珠子，和含有进行核酸扩增所需的试剂的扩增反应溶液。在随后的步骤中，在微反应器中扩增断裂的核酸，以形成扩增的拷贝的核酸，并且使扩增的拷贝与微反应器中的珠子结合。随后，将珠子递送到平坦表面上的至少 10,000 个反应室的阵列中，其中所述多个反应室包含不超过单个珠子。最后，同时在多个反应室中进行测序反应。

本发明的另一实施方案涉及包含具有多个凹陷的平坦表面的阵列，

每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ，并且每个凹陷在至少一个维度的宽度是至少 20 $\mu\text{m}$  - 70 $\mu\text{m}$ 。此外，阵列中至少有 10,000 个反应室。每个反应室可以含有至少 100,000 个拷贝的单种类的单链核酸模板。

本发明的另一实施方案涉及包含平坦的上表面和平坦的下表面的阵列，其中平坦的上表面中具有至少 10,000 个凹陷，每个凹陷形成分析物反应室，平坦的下表面任选是传导性的，这样来自反应室的光学信号可以通过下表面进行检测，其中上表面和下表面之间的距离不超过 5 mm，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ，并且每个室在至少一个维度的宽度是至少 20 $\mu\text{m}$  - 70 $\mu\text{m}$ 。在该阵列中，上表面和下表面之间的距离不超过 2 mm。

本发明的另一实施方案涉及用于在含水环境中进行分开的平行共同反应的阵列工具。该阵列工具可以包含一种基质，该基质含有至少 10,000 个离散的反应室，反应室中含有能够与试剂反应的起始材料，每个反应室的尺寸使得当一种或多种含有至少一种试剂的流体送递到每个反应室时，试剂扩散到孔外的扩散时间超过起始材料与试剂反应形成产物所需要的时间。

本发明的另一实施方案涉及将生物活性试剂送递到阵列中的方法。该方法包括在阵列上分散多个可移动的固体支持物，每个可移动的固体支持物具有至少一种固定在其上的试剂，其中该试剂适用于核酸测序反应中，其中该阵列包含一个平坦的表面，该表面具有多个置于其上的反应室。反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ，并且每个反应室在至少一个维度的宽度是 20 $\mu\text{m}$  - 70 $\mu\text{m}$ 。

本发明的另一实施方案涉及用于同时监测反应室阵列中表明在特定位点发生反应的光的装置。该装置包含：(a)从包含多个有凹陷表面的平面基质形成的反应室阵列，每个有孔表面形成适于容纳分析物的反应室，并且其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ，每个反应室的体积是 10 - 150 pL，该阵列包含超过 10,000 个离散的反应室；(b)光敏器件，其排列使得在使用中，来自特定反应室的光将撞击光敏器件的特定预定区；(c)用于确定撞击每个预定区的光水平的工具；和(d)记录每个反应室中随时间的光水平变化的工具。

本发明的另一个实施方案涉及一种分析传感器，包含：(a)从第一个

光导纤维束形成的阵列，该光导纤维束的一端具有多个有凹陷的表面，每个有凹陷的表面形成适于容纳分析物的反应室，并且其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ ，宽度是 20 - 70 $\mu\text{m}$ ，该阵列包含超过 10,000 个离散的反应室；(b)用于在反应室中产生光的酶或荧光工具；(c)包含光捕获工具的光检测工具和用于将光传输到光检测工具的第二个光导纤维束，所述第二个光导纤维束与阵列形成光学接触，使得各个反应室中产生的光被第二个光导纤维束的分开的纤维或分开的纤维的组捕获，用于传输到光捕获工具中。

本发明的另一实施方案涉及用于在含水环境中进行分开的平行共同反应的方法。第一个步骤涉及将含有至少一种试剂的流体送递到阵列，其中该阵列包含基质，该基质包含至少 10,000 个离散的反应室，每个反应室适于容纳分析物，并且其中反应室的体积是 10 - 150 pL，并且含有能够与试剂反应的起始材料，每个反应室的尺寸使得当流体送递到每个反应室时，试剂扩散到孔外的扩散时间超过起始材料与试剂反应形成产物所需要的时间。第二个步骤涉及在以下时间段中从阵列洗去流体：(i)在起始材料与试剂反应从而在每个反应室中形成产物之后但(ii)在送递到任意一个反应室的试剂从该反应室扩散到任何其它反应室之前。

本发明的另一实施方案涉及将核酸测序酶送递到阵列的方法。该阵列具有平坦的表面，该表面具有多个凹陷，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$ 。该方法包括在阵列上散布多个固定了一种或多种核酸测序酶的固体支持物，使得所述多个反应室含有至少一种可移动的固体支持物。

本发明的另一实施方案涉及将多个核酸模板送递到阵列中的方法。该阵列上可以具有平坦的表面，该表面具有多个凹陷，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 - 100 $\mu\text{m}$  并且该阵列具有至少 10,000 个反应室；该方法包括在阵列上散布多个可移动的固体支持物的步骤，每个可移动的固体支持物上固定了不超过单个种类的核酸模板，该散布导致任何一个反应室中放置不超过一个可移动的固体支持物。

本发明的另一实施方案涉及对核酸进行测序的方法，该方法包括以下步骤：提供置于平坦表面上的多个凹陷中的多个单链核酸模板，每个

凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$  并且该平坦表面具有至少 10,000 个反应室。随后的步骤涉及通过以下步骤在所有反应室中同时进行基于焦磷酸的测序反应：使有效量的测序引物与核酸模板退火，并且用聚合酶和预定核苷酸三磷酸延伸测序引物，得到测序产物，并且，如果预定核苷酸三磷酸掺入所述测序引物的 3' 末端，则得到测序反应副产物。第三个步骤涉及鉴定测序反应副产物，从而确定每个反应室中的核酸的序列。

本发明的另一实施方案涉及在阵列上确定多个核苷酸的碱基序列的方法。第一个步骤涉及提供至少 10,000 个 DNA 模板，每个模板独立地置于平坦表面上的多个凹陷中，每个凹陷形成形成分析物反应室，其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$  并且反应室的体积是 10 – 150 pL。第二个步骤涉及将一个已知的含氮碱基的激活的核苷 5' – 三磷酸前体加入每个反应室中的反应混合物，每个反应混合物包含针对模板的核苷酸聚合酶和与比模板短至少一个核苷酸残基的互补寡核苷酸引物链杂交的单链多核苷酸模板，以便在允许激活的核苷 5' – 三磷酸前体掺入引物链的 3' 末端的条件下在引物链的 3' 末端在每个模板中形成至少一个未配对的核苷酸残基，前提是激活的核苷 5' – 三磷酸前体的含氮碱基与模板的未配对核苷酸残基的含氮碱基互补。第三个步骤涉及检测核苷 5' – 三磷酸前体是否掺入引物链，其中核苷 5' – 三磷酸前体的掺入表明模板的未配对核苷酸残基具有与掺入的 5' – 三磷酸前体的含氮碱基组成互补的含氮碱基组成。第四个步骤涉及顺序重复步骤(b)和(c)，其中每次顺序重复添加并且检测已知含氮碱基组成的一种类型的激活的核苷 5' – 三磷酸前体的掺入。第五个步骤涉及从掺入的核苷前体的序列确定每个反应室中模板的未配对核苷酸残基的碱基序列。

本发明的另一实施方案涉及鉴定模板 DNA 的 DNA 序列中的靶位置中的碱基的方法。第一个步骤涉及提供至少 10,000 个独立的 DNA 模板独立地置于平坦表面上的多个凹陷中，每个凹陷形成形成分析物反应室，其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$ ，在置于反应室中之前或之后，使 DNA 成为单链。第二个步骤涉及提供延伸引物，其在紧接靶位置的位置上与固定的单链 DNA 杂交。在预定的脱氧核苷酸或双脱氧核苷酸的存在下使固定的单链 DNA 进行聚合酶链反应，其中如果预定的脱氧核苷酸或双脱氧核苷酸掺入测序引物的 3' 末端，则形成测序

反应副产物。第四个步骤涉及鉴定测序反应副产物，从而确定与 10,000 个 DNA 模板中每一个模板的靶位置的碱基互补的核苷酸。

本发明的另一方面涉及用于分析核酸序列的装置，该装置包含：(a)试剂递送杯，其中该递送杯包含阵列，该阵列包含具有多个凹陷的平坦表面，每个凹陷形成分析物反应室，其中反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$ ，并且存在超过 10,000 个反应室，并且其中试剂递送杯含有用于测序反应的试剂；(b)与试剂递送杯连通的试剂递送工具；(c)与试剂递送室连通的成像系统；和(d)与成像系统连通的数据采集系统。

本发明的另一实施方案涉及用于在阵列上确定多个核苷酸的碱基序列的装置。该装置包含：(a)含有位于平坦表面上的多个凹陷的试剂杯，每个凹陷形成分析物反应室，其中存在超过 10,000 个反应室，每个反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$  并且反应室的体积是 10 – 150 pL；(b)用于将一个已知含氮碱基的激活的核苷 5' – 三磷酸前体同时加入每个反应室的试剂递送工具，每个反应混合物包含针对模板的核苷酸聚合酶和与比模板短至少一个核苷酸残基的互补寡核苷酸引物链杂交的单链多核苷酸模板，以便在允许激活的核苷 5' – 三磷酸前体掺入引物链的 3' 末端的条件下在引物链的 3' 末端在每个模板中形成至少一个未配对的核苷酸残基，前体是激活的核苷 5' – 三磷酸前体的含氮碱基与模板的未配对核苷酸残基的含氮碱基互补；(c)用于在每个反应室中检测核苷 5' – 三磷酸前体是否掺入引物链的检测工具，其中核苷 5' – 三磷酸前体的掺入表明模板的未配对核苷酸残基具有与掺入的 5' – 三磷酸前体的含氮碱基组成互补的含氮碱基组成；(d)用于顺序重复步骤(b)和(c)的工具，其中每次顺序重复添加并且检测已知含氮碱基组成的一种类型的激活的核苷 5' – 三磷酸前体的掺入；和(e)用于在每个反应室中同时从掺入的核苷前体的序列确定每个反应室中模板的未配对核苷酸残基的碱基序列的数据处理工具。

本发明的另一实施方案涉及用于处理多个分析物的装置，该装置包含：(a)具有置于其中的基质的流动室，该基质包含位于光导纤维束上的至少 50,000 个有凹陷的表面，每个有凹陷的表面形成适于容纳分析物的反应室，并且其中每个反应室的中心与中心的间距是 20 – 100 $\mu\text{m}$  并且反应室的直径是 20 – 70 $\mu\text{m}$ ；(b)用于将处理试剂从一个或多个储液器递送到流动室，从而使置于反应室中的分析物暴露于试剂的流动工具；和(c)

用于同时检测来自每个反应室的一系列光学信号的检测工具，该系列的每个光学信号表明处理试剂和置于反应室中的分析物之间的相互作用，其中检测工具与有凹陷的表面连通。

本发明的另一实施方案涉及对核酸进行测序的方法。第一个步骤涉及在具有至少 50,000 个离散的反应位点的阵列中提供多个单链核酸模板。第二个步骤涉及使核酸模板与进行与发光偶联的基于焦磷酸的测序反应所必须的试剂接触。第三个步骤涉及检测从光敏器件上各自部分上的多个反应位点发出的光。第四个步骤涉及将撞击光敏器件的每个部分的光转化为电信号，该电信号可以与来自所以其它反应位点的信号区分。第五个步骤涉及根据相应的电信号，基于从每个所述离散反应位点的光发射，确定核酸模板的序列。

本发明的另一实施方案涉及对核酸进行测序的方法。第一个步骤涉及使大模板核酸分子断裂，产生多个断裂的核酸。第二个步骤涉及将多个断裂的核酸的一条链各自连接于珠子，得到各自连接于珠子的单链断裂核酸。第三个步骤涉及单独与珠子连接的单链的断裂核酸群体送递到平坦表面上的至少 10,000 个反应室的阵列中，其中所述多个孔包含不超过一个具有单链的断裂核酸的珠子。第四个步骤涉及同时在多个反应室中进行测序反应。测序反应可以具有以下步骤：(a)使有效量的测序引物与单链断裂核酸模板退火，并且用聚合酶和预定核苷酸三磷酸延伸测序引物，得到测序产物，并且，如果预定核苷酸三磷酸掺入所述测序引物的 3' 末端，则得到测序反应副产物；以及(b)鉴定测序反应副产物，从而确定多个反应室中的核酸的序列。或者，测序反应可以包括以下步骤：(a)使两个或多个测序引物与核酸分子的一个或多个单链杂交，其中除一个引物之外的所有引物都是可逆阻断的引物；(b)通过从未阻断的引物进行聚合酶延伸，将至少一个碱基掺入核酸分子；(c)防止未阻断的引物进一步延伸；(d)将可逆阻断的引物去阻断为未阻断的引物；以及(e)重复步骤(b) - (d)，直到至少一个可逆阻断的引物被去阻断，并且用于确定序列。

其它材料和方法可以参见以下共同未决的美国专利申请：2003 年 1 月 29 日提交的 USSN 60/443,471，2003 年 4 月 23 日提交的 USSN 60/465,071，2004 年 1 月 28 日提交的 USSN 10/767,894，2004 年 1 月 28 日提交的 USSN 10/767,899，和 2004 年 1 月 28 日提交的 USSN 10/768,729。在此引入本公开内容中引用的所有专利、专利申请和参考

文献作为参考。

## 实施例

### 实施例 1：样品制备

#### DNA 样品：

DNA 应该是高质量的，并且不含污染物，如蛋白、核酸、脂质和其它化学物质（如来自制备物的残留 EDTA）以及盐。优选的基因组 DNA 应该具有 1.8 或更高的 260/280 比。如果需要仅仅对一种生物的基因组进行测序，则应该对 DNA 进行质量检验，以确保不含污染 DNA。例如，可以通过 PCR 检验人 DNA 的制备物，以确保其不受细菌 DNA 分子的污染。另一种检验污染的方法是通过用已知对已知生物（如人或小鼠）特异的合适的探针和已知对可能的污染生物（如大肠杆菌）特异的第二探针检验限制酶消化模式，特别是限制酶消化后的 Southern 印迹。如果需要，DNA 应该来自生物的单克隆（例如，如果来自细菌，则是菌落）。

#### 步骤 1：DNA 酶 I 消化

DNA 酶 I 消化步骤的目的是使 DNA 的大片段，如完整基因组或基因组的大部分断裂为较小的种类。从单 DNA 模板得到的较小的 DNA 种类的这种群体称作“文库”。脱氧核酸酶 I (DNA 酶 I) 是一种内切核酸酶，其切割双链模板 DNA。DNA 酶 I 的切割特征使得能够随机消化模板 DNA (即最小序列偏倚)，并且将导致当存在基于锰的缓冲液 (Melgar and Goldthwait 1968) 的条件下使用时，平端、双链 DNA 片段的优势。DNA 酶 I 对基因组模板的消化依赖于三种因素：i)使用的酶的量 (单位)；ii)消化的温度 (°C)；和 iii)温育时间 (分钟)。下文概括的 DNA 酶 I 消化条件进行优化，以产生大小范围为 50 – 700 个碱基对 (bp) 的 DNA 文库。

1. 获得 DNA，并且在 Tris - HCl (10mM, pH 7 – 8) 中制备为 0.3 mg/ml 的浓度。该制备需要总共 134 µl DNA (15 µg)。推荐不使用含有 EDTA 的缓冲液 (即 TE, Tris/EDTA) 稀释的 DNA 制备物。EDTA 的存在对于 DNA 酶 I 的酶消化是抑制性的。如果 DNA 制备物含 EDTA，重要的是 DNA 从溶液中“盐析”出来，并且用合适的 Tris - HCl 缓冲液 (10 mM, pH 7 – 8) 或毫微纯 (nanopure) 的 H<sub>2</sub>O (pH 7 – 8) 重配。

2. 在 0.2 ml 试管中, 制备 DNA 酶 I 缓冲液, 其中含有 50 $\mu$ l Tris pH 7.5 (1M), 10 $\mu$ l MnCl<sub>2</sub> (1M), 1 $\mu$ l BSA (100 mg/ml) 和 39 $\mu$ l 水。

3. 在分开的 0.2 ml 试管中, 加入 15 $\mu$ l DNA 酶 I 缓冲液和 1.5 $\mu$ l DNA 酶 I (1U/ml)。将反应管置于设置为 15°C 的热循环仪中。

4. 将 134 $\mu$ l DNA (0.3 mg/ml) 加入 DNA 酶 I 反应管中, 所述反应管置于设置为 15°C 的热循环仪中。关上盖子, 将样品准确温育 1 分钟。温育后, 加入 50 $\mu$ l 50 mM EDTA, 以终止酶消化。

5. 使用 QiaQuick PCR 纯化试剂盒纯化消化的 DNA。然后将消化反应分为四等分, 用四个旋转柱纯化每份等分物 (每个旋转柱 37.5 $\mu$ l)。根据制造商的方案, 用 30 $\mu$ l 洗脱缓冲液(EB)洗脱每个柱子。然后合并洗脱液, 得到 120 $\mu$ l 的最终反应体积。

6. 保留消化反应的一个 3 $\mu$ l 等分物, 用于采用 BioAnalzyer DNA 1000 LabChip 进行分析。

## 步骤 2: Pfu 磨光

用 DNA 酶 I 消化 DNA 模板, 产生 DNA 片段, 其中主要是平端的, 但一些片段具有含长度为 1 个或 2 个核苷酸的突出端的末端。采用 Pfu 磨光, 通过填充 (即“平端化”) 5' 突出端, 增加平端种类的量。此外, Pfu DNA 聚合酶具有 3' → 5' 外切核酸酶活性, 其将导致除去单和双核苷酸延伸。Pfu 磨光增加可以用于衔接子连接的平端 DNA 片段的量(Costa 1994a, 1994b, 1994c)。使用以下 Pfu 磨光程序。

1. 在 0.2 ml 试管中, 按顺序加入 115 $\mu$ l 纯化的、DNA 酶 I 消化的 DNA 片段, 15 $\mu$ l 10X 克隆的 Pfu 缓冲液, 5 $\mu$ l dNTPs (10 mM) 和 15 $\mu$ l 克隆的 Pfu DNA 聚合酶(2.5 U/ $\mu$ l)。

2. 充分混合磨光反应成分, 72°C 下温育 30 分钟。

3. 温育后, 除去反应试管并且置于冰上 2 分钟。

4. 然后将磨光反应混合物分为四等分, 用 QiaQuick PCR 纯化柱纯化 (每个柱 37.5 $\mu$ l)。根据制造商的方案, 用 30 $\mu$ l 缓冲液 EB 洗脱每个柱子。然后合并洗脱液, 得到 120 $\mu$ l 的最终反应体积。

5. 保留最终磨光反应的一个 3 $\mu$ l 等分物, 用于采用 BioAnalzyer DNA 1000 LabChip 进行分析。

### 步骤 3：将通用衔接子连接到断裂的 DNA 文库

断裂和基因组 DNA 文库磨光后，将引物序列加入每个 DNA 片段的末端。这些引物序列称作“通用引物”并且包含含有特异性引物区的双链寡核苷酸，其提供 PCR 扩增和核苷酸测序。通用衔接子设计为包含一组独特的 PCR 引物区，其长度为 20 个碱基对，邻接一组独特的测序引物区，该测序引物区长度典型是 20 bp，其后是由每种脱氧核糖核苷酸（即 A, C, G, T）中的一种组成独特的 4 碱基“关键序列”。每种独特通用衔接子称作“通用衔接子 A”和“通用衔接子 B”的长度是四十四个碱基对 (44 bp)。用 T4 DNA 连接酶将通用衔接子连接到 DNA 片段的每个末端上，以便总共在每个 DNA 片段上添加 88 bp。对于每次 DNA 文库制备，特别设计了不同的通用衔接子，因此对每种生物提供了独特标识符。

为了制备一对通用衔接子，单链寡核苷酸是内部设计的，并且通过销售商制造。通用衔接子 DNA 寡核苷酸设计为具有在每个寡核苷酸末端具有两个硫代磷酸酯键，起保护作用，使其抗核酸酶活性(Samini, T.D., B. Jolles, and A. Laigle. 2001. Best minimally modified antisense oligonucleotides according to cell nuclease activity. Antisense Nucleic Acid Drug Dev. 11(3):129., 其公开内容在此全文引入作为参考)。每种寡核苷酸进行 HPLC 纯化，以确保最终制备物不含污染或多余的 DNA 寡核苷酸序列。

设计通用衔接子，使得能够方向性连接于平端的、断裂的基因组 DNA。对于每个通用衔接子对，PCR 引物区包含 5' 四碱基突出端和平端 3' 关键区。由于通用衔接子的平端侧连接于平端 DNA 片段，而衔接子的 5' 突出端不能连接于平端 DNA 片段，实现了方向性。此外，将 5' 生物素加入通用衔接子 B，使得随后能够分离 ssDNA 模板 (步骤 8)。通过在单个试管中使两个单链互补 DNA 寡核苷酸 (即含有有义序列的一个寡核苷酸和含有反义序列的第二寡核苷酸) 退火，制备每种通用衔接子。使用以下连接方案。

1. 在 0.2ml 管中，按顺序加入 39 $\mu$ l nH<sub>2</sub>O (分子生物级水)，25 $\mu$ l 消化的、磨光的 DNA 文库，100 $\mu$ l 2X 快速连接酶反应缓冲液，20 $\mu$ l MMP1 (10 pm/ $\mu$ l) 衔接子组，100:1 比例，和 16 $\mu$ l 快速连接酶。充分混合连接反应，室温下温育 20 分钟。

2. 随后除去连接反应，纯化连接反应的 10 $\mu$ l 等分物，用于 BioAnalyzer 上。使用 Qiagen Min - Elute 试剂盒的单旋转柱。按照制造商的方案，用 10 $\mu$ l EB 洗脱柱子。用 BioAnalyzer DNA 1000 LabChip 加入纯化的连接反应的 1 $\mu$ l 等分物。该纯化步骤推荐为未纯化的连接反应，所述反应含有大量盐和 PEG，将抑制样品在 BioAnalyzer 上合适地流动。

3. 将其余的连接反应物 (190 $\mu$ l) 用于步骤 4 的凝胶分离。

步骤 3a: Microcon 过滤和衔接子构建。总制备时间是大约 25 分钟。

通用衔接子连接反应需要 100 倍过量的衔接子。为了避免除去这些过量的衔接子，通过 Microcon YM - 100 过滤装置过滤双链 gDNA 文库。Microcon YM - 100 膜可以用于除去小于 125 bp 的双链 DNA。因此，可以从连接的 gDNA 文库群体除去未结合的衔接子(44 bp)和衔接子二聚体(88 bp)。使用以下过滤方案：

1. 将来自步骤 4 的 190  $\mu$ L 连接反应物加入组装的 Microcon YM - 100 装置。

2. 将该装置置于离心机中，以 5000  $\times$  g 旋转大约 6 分钟，或直到膜基本上干燥。

3. 为了洗涤，加入 200 $\mu$ l 1X TE。

4. 使样品以 5000  $\times$  g 旋转额外的 9 分钟，或直到膜基本上干燥。

5. 为了回收，将储液器插入新的管形瓶，并且以 3000  $\times$  g 旋转 3 分钟。弃去储液器。回收的体积是大约 10 $\mu$ l。随后，加入 80 $\mu$ l TE。

使用前 HPLC 纯化衔接子 (A 和 B)，用硫代磷酸酯键修饰。对于衔接子“A”(10 $\mu$ M)，将 10 $\mu$ l 100  $\mu$ M 衔接子 A (44 bp, 有义)与 10 $\mu$ l 100  $\mu$ M 衔接子 A (40 bp, 反义)混合，并且混合 30 $\mu$ l 1X 退火缓冲液( $V_f = 50\mu$ l)。用 ANNEAL 程序在 Sample Prep Lab 热循环仪上进行引物退火 (参见下文)。对于衔接子 “B” (10 $\mu$ M)，将 10 $\mu$ l 100  $\mu$ M 衔接子 B(40 bp, 有义)与 10 $\mu$ l 100  $\mu$ M 衔接子 B (44 bp, 反义)和 30 $\mu$ l 1X 退火缓冲液( $V_f = 50\mu$ l) 混合。用 ANNEAL 程序在 Sample Prep Lab 热循环仪上进行引物退火。衔接子组可以在 -20 °C 下储存备用。

用于引物退火的 ANNEAL - A 程序：

1. 95 °C 下温育 1 分钟；
2. 以 0.1 °C / 秒将温度降低到 15 °C；以及

### 3. 保持在 15°C。

对于基因组 DNA 插入片段和衔接子，不需要方向性。可以在任一末端连接片段。通用衔接子组中包括四个单链 DNA 寡核苷酸。以 1 微摩尔规模合成每个单链寡核苷酸，并且进行 HPLC 纯化。每个单链寡核苷酸在每个末端包括四个硫代磷酸酯键。

### 步骤 4：凝胶电源和衔接的 DNA 文库的提取

通用衔接子连接方案产生：1) 在每个末端具有衔接子的断裂的 DNAs；2) 未结合的单衔接子；或 3) 衔接子引物的形成。琼脂糖凝胶电泳用作从未连接的单衔接子和衔接子二聚体群体分开和分离衔接的 DNA 文库群体的方法。基因组 DNA 的 DNA 酶 I 消化产生 50 – 700 bp 的文库群体（步骤 1）。88 – bp 通用衔接子的加入使群体变为较大的大小，并且导致大小范围为大约 130 – 800 bp 的迁移谱。衔接子二聚体将在 88 bp 迁移，未连接的衔接子将在 44 bp 迁移。因此，可以从琼脂糖凝胶物理分离大小 > 200 bp 的基因组 DNA 文库，用标准凝胶提取技术纯化。衔接的 DNA 文库的凝胶分离将导致大小范围 ≥ 200 bp 的文库群体（文库的大小范围可以根据应用而不同）的回收。使用以下电泳和提取方案。

1. 制备 2% 琼脂糖凝胶。
2. 将 10μl 10X Ready - Load 染料加入其余的 90μl DNA 连接混合物。
3. 将染料/连接反应混合物加入使用 4 个相邻的泳道的凝胶中（每个泳道 25μl）。
4. 将 10μl 100 bp 序列梯（0.1μg/μl）加入离开连接反应泳道的两个泳道中。
5. 100V 下凝胶电泳 3 小时。
6. 当完成凝胶电泳时，从凝胶盒取出凝胶，转移到用塑料套覆盖的平面上。用手动长波 UV 光显现 DNA 条带。采用无菌、单用途刮刀，从琼脂糖凝胶上切下大小为 200 – 400 bp 的片段。采用该方法，可以分离任何大小范围的文库。也可以分离超过 1 个大小范围。当文库大小范围是 200 – 900 bp 时，可以从单个孔分离一些大小范围（即 200 – 400 bp 和 500 – 700 bp）。

7. 按照制造商的说明书,用 Qiagen MinElute 凝胶提取试剂盒分离包埋在琼脂糖凝胶中的 DNA。简言之,加入缓冲液 QG,以覆盖试管中的琼脂糖。使琼脂糖完全溶解。通过根据 Qiagen 说明书调节 pH 使样品损失最小,保持缓冲液 QG 的颜色。用两个 MinElute 旋转柱(Qiagen)进行纯化。溶解的琼脂糖的大体积需要每个柱加样数次。用 55°C 下预热的 10 μl 缓冲液 EB 洗脱柱子。合并洗脱液,产生 20μl gDNA 文库。

8. 用 BioAnalyzer DNA 1000 LabChip 分析每个分离的 DNA 文库的 1μl 等分物,以评估 DNA 文库群体的准确分布。

#### 步骤 5: 有切口的双链 DNA 文库的链置换和延伸

由于用于通用衔接子的 DNA 寡核苷酸不是磷酸化的,在断裂的 gDNAs 的 3'接合处存在缺口。这两个“缺口”或“切口”可以通过采用链置换 DNA 聚合酶而进行填充。聚合酶识别切口,置换有切口的链,并且以导致切口修复和形成无切口的双链 DNA 的方式延伸链。使用的链置换酶是 Bst DNA 聚合酶的大片段。

1. 在 0.2 ml 管中,按顺序加入 19μl 凝胶提取的 DNA 文库,40μl nH<sub>2</sub>O,8μl 10X ThermoPol 反应缓冲液,8μl BSA(1 mg/ml),2μl dNTPs (10 mM) 和 3μl Bst I 聚合酶 (8U/μl)。

2. 充分混合样品,置于热循环仪中,用标准链置换温育程序“BST”进行温育。有切口的双链 DNA 的链置换和延伸的 BST 程序:

1. 65°C 下温育 30 分钟;
2. 80°C 下温育 10 分钟;
3. 58°C 下温育 10 分钟; 以及
4. 保持在 14°C。

3. 用 BioAnalyzer DNA 1000 LabChip 分析 Bst 处理的 DNA 文库的 1μl 等分物。

#### 步骤 6: 制备链霉亲和素珠

制备无切口的双链基因组 DNA 后,必须分离含有侧翼于通用衔接子序列的单链基因组 DNAs。该步骤概括了生物素标记的双链 DNA 与链霉亲和素珠的结合。为了制备链霉亲和素珠,采用以下方案。

1. 通过将磁珠用于 MPC,用 200μl 1X 结合缓冲液(1 M NaCl, 0.5

mM EDTA, 5 mM Tris, pH 7.5) 将 100 $\mu$ l Dynal M - 270 链霉亲和素珠洗涤 2 次。

2. 将珠子重悬于 100 $\mu$ l 2X 结合缓冲液，然后加入其余的 79 $\mu$ l Bst 处理的 DNA 样品（来自步骤 5）和 20 $\mu$ l 水。

3. 将珠溶液充分混合，室温下放置在管旋转器上 20 分钟。用 MPC 洗涤珠混合物，其中用 100 $\mu$ l 1X 结合缓冲液洗涤 2 次，然后用 nH<sub>2</sub>O 洗涤 2 次。结合&洗涤(B&W)缓冲液(2X 和 1X):通过混合 10 mM Tris• HCl (pH 7.5), 1 mM EDTA 和 2 M NaCl 制备 2X B&W 缓冲液。按照上文的描述合并试剂，并且充分混合。溶液可以在室温下保持 6 个月；通过 1: 1 混合 2X B&W 缓冲液和 nH<sub>2</sub>O，制备 1X B&W 缓冲液。终浓度是上述的一半，即 5 mM Tris• HCl (pH 7.5), 0.5 mM EDTA 和 1 M NaCl。

#### 步骤 7: 用链霉亲和素珠分离单链 DNA 文库

将双链 gDNA 文库结合于链霉亲和素珠后，优选从连接的合并物仅仅分离含有通用衔接子 A 和通用衔接子 B 的单链 gDNAs（需要的群体在下文用星号表示）。双链基因组 DNA 片段合并物将具有以下面的构型结合的衔接子：

通用衔接子 A - gDNA 片段 - 通用衔接子 A

通用衔接子 B - gDNA 片段 - 通用衔接子 A\*

通用衔接子 A - gDNA 片段 - 通用衔接子 B\*

通用衔接子 B - gDNA 片段 - 通用衔接子 B

由于仅仅通用衔接子 B 具有 5' 生物素部分，可以用含有链霉亲和素的磁珠结合所有具有通用衔接子 B 的 gDNA 文库种类。含有两个通用衔接子 A 种类（或非连接种类）的基因组文库群体不结合于含链霉亲和素的珠子，并且在洗涤过程中除去。在洗涤后保持结合于珠子的种类包括具有衔接子 A 和 B 的那些或具有两个通用衔接子 B 末端的那些。

具有两个通用衔接子 B 序列的基因组 DNA 种类可以在两个末端都结合于含有链霉亲和素的珠子，所述衔接子 B 序列带有两个生物素分子。具有只有单个生物素分子的 A 和 B 衔接子的种类仅仅可以在“B”末端结合于珠子。为了分离单链群体，用破坏互补 DNA 链之间的氢键的氢氧化钠溶液处理珠结合的双链 DNA。如果 DNA 片段在每个末端（通用衔接子 B 末端）都具有生物素，则得到的两个单链都保持与珠子结合。

如果片段仅仅具有单个生物素(通用衔接子A和B), 则互补链与DNA-珠复合物分开。

从液相收集得到的单链基因组DNA文库, 并且例如用焦磷酸测序(PyroSequence)或用RNA Pico 6000 LabChip(Agilent, Palo Alto, CA)定量。通过计算每单位体积的分子数, 对单链基因组DNA文库进行定量。然后使单链gDNA分子退火于含有DNA捕获引物(PCR引物B)的25-30 μm琼脂糖珠(每个珠子一半拷贝, 以获得每个珠子一个有效拷贝)。然后用乳状液聚合酶链反应方案扩增模板。随后可以用已知技术进行测序。为了分离单链文库, 使用以下方案。

1. 将250μl解链溶液(0.125 M NaOH, 0.1 M NaCl)加入来自上文步骤6的洗涤的珠子。
2. 充分混合珠溶液, 在管旋转器上室温下将珠混合物温育10分钟。
3. 使用Dynal MPC(磁颗粒浓缩器), 小心取出沉淀的珠子, 留出上清液。250μl上清液包括单链DNA文库。
4. 在分开的管中, 加入1250μlPB(来自QiaQuick纯化试剂盒), 并且通过加入9μl20%醋酸, 中和溶液。
5. 采用Dynal MPC, 沉淀来自包含单链gDNA文库的250μl上清液的珠子, 小心取出上清液, 并且转移到新制备的PB/醋酸溶液。
6. 用单QiaQuick纯化旋转柱纯化1500μl溶液(以每次加样750μl, 通过相同柱子加样2次)。用50μlEB洗脱单链DNA文库。

步骤8a: 用焦磷酸测序进行单链gDNA定量。总制备时间是大约1小时。

1. 在0.2 ml管中, 按顺序加入以下试剂:

25μl 单链DNA  
1μl MMP2B测序引物  
14μl 文库退火缓冲液  
总共40μl

2. 用ANNEAL-S程序使DNA退火(参见下文附录)。
3. 在PSQ(焦磷酸测序夹具)上运行样品, 以确定每个样品中模板的皮摩尔数(参见下文)。测序方法可以参见美国专利6,274,320; 美国专利4,863,849; 美国专利6,210,891; 和美国专利6,258,568, 在此全文

引入其公开内容作为参考。进行计算，以确定每皮摩尔的单链 gDNA 模板分子数目。剩余的 25 $\mu$ l 制备的单链 gDNA 文库用于扩增和后续的测序（大约  $1 \times 10^6$  个反应）。

步骤 8b：用 RNA Pico 6000 LabChip 进行单链 gDNA 定量。总制备时间是大约 30 分钟。

1. 在 BioAnalyzer (软件版本 2.12) 上选择 mRNA Pico 测定选项。
2. 根据制造商的指导，在 BioAnalyzer 上制备 RNA Pico 6000 LabChip。
3. 根据制造商 (Ambion) 的说明书，制备 RNALabChip 序列梯 (RNA 6000 序列梯)。简言之，将溶液中的 RNA LabChip 序列梯加热到 70°C 2 分钟。在冰上将溶液冷却 5 分钟，以便突然冷却序列梯。短暂离心溶液，以便从管壁清除任何冷凝物。在冰上储存 RNA LabChip 序列梯，并且在一天内使用。
4. 采用 3 个 1 $\mu$ l 等分物，在相邻的泳道上以三份进行 ssDNA 文库分析。
5. 用 BioAnalyzer 软件计算每个 ssDNA 文库泳道的浓度（参见下表 3 和图 24）。采用下文概括的程序用所有三个泳道的平均值计算文库的 DNA 浓度。
  - a. 峰积分下限线（图 24 的宽虚线）移动到文库峰前（参见下文）。
  - b. 峰积分上限线（图 24 的宽虚线）移动到文库峰后。以此方式，连接下积分线和上积分线的峰积分线在背景的斜坡之后。
  - c. 用小鼠箭头确定基底中的峰（通常邻接峰最高点）的平均大小，或根据软件选择使用限定的峰。
  - d. 积分值用于峰中物质的量。获得的回收的皮克值转化为回收的分子数（参见下表 3）。然后确定文库浓度（每微升的分子数）。

表 3

	1	2	3	4	5 平均值	6 平均	7 平均	8 平均	9 平均值
	pg/ $\mu$ L (1)	pg/ $\mu$ L (2)	pg/ $\mu$ L (3)	pg/ $\mu$ L	大小(bp) 1	大小(bp) 2	大小(bp) 3	大小(bp)	
样品	1633	1639	1645	1639	435	435	432	434	

	10 平均分子量(g/摩尔)	11 平均分子量	12 文库	13	14	15
	核糖核苷酸	(g/摩尔)	g/ $\mu$ L	摩尔/g	摩尔/ $\mu$ L	分子数/ $\mu$ L
	328.2	1.42E+05	1.64E-09	7.02E-06	1.15E-14	6.93E+09

如上表 3 所示，文库 1 的浓度计算为 1639 pg/ $\mu$ L (第 5 列)，平均片段大小是 434 个核苷酸 (第 9 列)。这些值是按照上文步骤(a) - (d) 的描述从 Agilent 2100 软件获得的。核糖核苷酸的平均分子量(MW)是 328.2 g/摩尔 (第 10 列)。通过用平均片段长度 (434) 乘以平均核糖核苷酸(328.2)，计算平均文库片段分子量( $1.42 \times 10^5$  g/摩尔，第 11 列)。将定量的文库(1639 pg/ $\mu$ L)转化为每微升的克数 ( $1.64 \times 10^{-9}$  g/ $\mu$ L，第 12 列)。通过用每微升的克数 ( $1.64 \times 10^{-9}$  g/ $\mu$ L，第 12 列)除以文库片段的平均分子量 ( $1.42 \times 10^5$ ,第 11 列)，计算每微升的摩尔数 ( $1.15 \times 10^{-14}$  摩尔/ $\mu$ L，第 14 列)。最后，通过用每微升的摩尔数( $1.15 \times 10^{-14}$  摩尔/ $\mu$ L，第 14 列)乘以阿佛加德罗常数( $6.02 \times 10^{23}$  个分子/摩尔)，得到每微升的分子数目( $6.93 \times 10^9$  个分子/ $\mu$ L，第 15 列)。

预期最终文库浓度大于  $1 \times 10^8$  个分子/ $\mu$ L。用于文库定量的更重要因素是衔接子二聚体浓度。在图 24 中，确定文库峰的高度比衔接子二聚体峰 (标志物后的第一个峰) 高大约 10 倍。预期高质量文库的峰高度至少比二聚体峰高 2 倍。应该注意，RNA Pico 6000 LabChip 提供的估计值在单链 gDNA 浓度的 500% 准确度范围内。因此，重要的是用模板滴定进行最初的测序运行，以确定输入 gDNA 中每个珠子的拷贝数 (cpb)。推荐的输入 DNA 是 2.5 cpb, 1 cpb, 0.5 cpb 和 0.1 cpb。在 14 x 43 PTP 上使用 4 裂隙珠加样室，可以容易检查该滴度。

### 步骤 9：单链 gDNA 文库的稀释和储存

洗脱单链 gDNA 文库，并且在缓冲液 EB 中定量。为了防止降解，在 EDTA 存在于 -20°C 冷冻储存单链 gDNA 文库。定量后，在文库储存液中加入等体积的 10 mM TE。所有随后的稀释都是在 TE 中。产率如下：

PSQ 分析后剩余的 ssDNA 文库终体积 = 25 μL。

LabChip 分析后剩余的 ssDNA 文库终体积 = 47 μL。

对于最初的储存液稀释，在 1X Library - Grade Elution 缓冲液中将单链 gDNA 文库稀释为 1 亿个分子/μL。制备单链 gDNA 的等分物用于共同使用。为此，将 200,000 个分子/μL 在 Library - Grade Elution 缓冲液稀释，并且测量 20 μL 等分物。将单用途文库等分物在 -20°C 下储存。

### 步骤 10：乳状液聚合酶链反应

当优选 cpb 数目增加时，按照 2003 年 6 月 6 日提交的美国专利申请系列 No. 06/476,504 中的描述进行珠乳状液 PCR，在此全文引入该申请作为参考。

#### 试剂制备

终止溶液(50 mM EDTA)包括与 900 μl nH<sub>2</sub>O 混合的 100 μL 0.5 M EDTA，以获得 1.0 ml 50 mM EDTA 溶液。对于 10 mM dNTPs, 10 μl dCTP (100 mM), 10 μl dATP (100 mM), 10 μl dGTP (100 mM) 和 10 μl dTTP (100 mM) 与 60 μl 分子生物级水混合。所有 4 种 100 mM 核苷酸储存液在冰上解冻。然后，将 10 μL 每种核苷酸与 60 μL nH<sub>2</sub>O 合并，达到 100 μL 的终体积，并且彻底混合。随后，将 1 ml 等分物分散到 1.5 ml 微量离心管中。储存液可以在 -20°C 下储存一年。

10 X 退火缓冲液包括 200 mM Tris (pH 7.5) 和 50 mM 醋酸镁。为了制备该溶液，将 24.23 g Tris 加入 800 ml nH<sub>2</sub>O，将混合物调节为 pH 7.5。在该溶液中加入 10.72 g 醋酸镁，完全溶解。使溶液达到 1000 ml 的终体积，并且可以在 4°C 下储存 1 个月。10 X TE 包括 100 mM Tris·HCl (pH 7.5) 和 50 mM EDTA。将这些试剂添加在一起，并且彻底混合。溶液可以在室温下储存 6 个月。

## 实施例 2：引物设计

如上文的讨论，通用衔接子设计为包括：1)一组独特的 PCR 引物区，其长度典型是 20 bp (邻接(2))；2)一组独特的测序引物区，其长度典型是 20 bp；和 3)任选后面是独特的区分关键序列，其由四种脱氧核糖核苷酸（即 A, C, G, T）中的至少一种组成。引物和感兴趣的基因组的未预期区域之间可能的交叉杂交随着基因组大小的增加和与引物的完整匹配长度减少而增加。但是，由于下文阐述的原因，与交叉杂交区(CHR)的这种潜在的相互作用预期不会产生问题。

在本发明的在一种优选实施方案中，单链 DNA 文库用于 PCR 扩增和随后的测序。测序方法需要将给定的基因组随机消化为 150 – 500 个碱基对的片段，其后两个独特的双向引物（包括 PCR 和测序区）连接在片段的 5' 和 3' 末端（图 25）。与典型的 PCR 扩增（其中基于解链温度 ( $T_m$ )、基因组内引物序列的独特性和与感兴趣的特定区域的邻接，基因组的现有区段选作引物位点）不同，公开的方法利用合成的引物位点，其使得必须小心从头设计引物。

### 四聚体选择：

从头设计引物的测序可以参见关于以下内容的文献：在分子标记上进行杂交实验（参见 Hensel, M. and D.W. Holden, Molecular genetic approaches for the study of virulence in both pathogenic bacteria and fungi. Microbiology, 1996. 142(Pt 5): p. 1049 – 58; Shoemaker, D.D., et al., Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar - coding strategy. Nat Genet, 1996. 14(4): p. 450 – 6）和 PCR/LDR（聚合酶链反应/连接检测反应）杂交引物（参见 Gerry, N.P., et al., Universal DNA microarray method for multiplex detection of low abundance point mutations. Journal of Molecular Biology, 1999. 292: p. 251 – 262; Witowski, N.E., et al., Microarray - based detection of select cardiovascular disease markers. BioTechniques, 2000. 29(5): p. 936 – 944.）。

PCR/LDR 工作特别相关，并且致力于设计寡核苷酸“代码 (zipcodes)”，即由六个具有相似最终  $T_m$  的特别设计的四聚体组成的 24 个碱基的引物（参见 Gerry, N.P., et al., Universal DNA microarray

method for multiplex detection of low abundance point mutations. Journal of Molecular Biology, 1999. 292: p. 251 – 262; 美国专利 No. 6,506,594)。基于以下标准选择四聚体成分：每个四聚体与其它四聚体有至少两个碱基不同，排除诱导自我配对或发夹形成的四聚体，以及省略回文(AGCT)或重复四聚体(TATA)。256 ( $4^4$ ) 种可能的排列中的 36 种满足必须的要求，并且随后进行可接受的 PCR 引物设计所需进一步限制（表 4）。

表 4

	TT	TC	TG	TA	CT	CC	CG	CA	GT	GC	GG	GA	AT	AC	AG	AA
TT	<b>TTTT</b>	TTTC	TTTG	TTTA	TTCT	TTCC	TTCG	TTCA	TTGT	TTGC	TTGG	<b>TTGA</b>	TTAT	TTAC	<b>TTAG</b>	TTAA
TC	TCTT	<b>TCTC</b>	<b>TCTG</b>	TCTA	TCCT	<b>TCCT</b>	TOOG	TOCA	<b>TGTT</b>	TOGC	TOGG	TGGA	TCAT	TCAC	TCAG	TCAA
TG	TGTT	<b>TGTC</b>	<b>TGTC</b>	TGTA	TGCT	TGCC	<b>TGCC</b>	TGCA	TGGT	TGGC	TGGG	TGGA	<b>TGAT</b>	TGAC	TGAG	TGAA
TA	TATT	TATC	TATG	<b>TATA</b>	TACT	TACC	TAOG	<b>TACA</b>	TAGT	TAGC	TAGG	TAGA	TAAT	TAAC	TAAG	TAAA
CT	CTTT	CTTC	<b>CTTG</b>	CTTA	<b>CTCT</b>	CTCC	CTOG	<b>CTCA</b>	<b>CTGT</b>	CTGC	CTGG	CTGA	CTAT	CTAC	CTAG	CTAA
CC	CCTT	OCTC	OCTG	<b>CCTA</b>	OCTT	<b>CCCC</b>	OOCG	OCCA	OOGT	OOGC	OOGG	OOGA	<b>CCAT</b>	CCAC	CCAG	CCAA
CG	<b>CGTT</b>	OCTC	OCTG	CGTA	OCTT	OCCG	<b>CGCC</b>	OCCA	OOGT	CGGC	OOGG	OOGA	CGAT	CGAC	CGAG	<b>CGAA</b>
CA	CATT	CATC	CATG	CATA	CACT	CAOC	<b>CAOG</b>	<b>CACA</b>	CAGT	<b>CAGC</b>	CAGG	CAGA	CAAT	CAAC	CAAG	CAA
GT	GTTT	GTTC	GTG	GTTA	<b>GTC</b>	GTOG	GTCA	<b>GTGT</b>	<b>GTGC</b>	GTGG	GTGA	GTAT	GTAC	GTAG	GTAA	
GC	<b>GCTT</b>	GCTC	GCTG	GCTA	OCTT	GOOC	GOOG	OCCA	GOGT	<b>GCCC</b>	GOGG	GOGA	GCAT	GCAC	GCAG	<b>GCAA</b>
GG	GGTT	GGTC	GGTG	<b>GGTA</b>	GGCT	GGCC	GGOG	OCCA	GGGT	GGGC	<b>GGCG</b>	GGGA	GGAT	<b>GGAC</b>	GGAG	GGAA
GA	GATT	GATC	<b>GATG</b>	GATA	GAAT	<b>GAOC</b>	GAOG	CACA	<b>GAGT</b>	GAGC	GAGG	<b>GAGC</b>	GAAT	GAAC	GAAG	GAAA
AT	ATTT	ATTC	ATTG	ATTA	ATCT	ATCC	<b>ATOG</b>	ATCA	ATGT	ATGC	ATGG	ATGA	<b>ATAT</b>	<b>ATAC</b>	ATAG	ATAA
AC	ACTT	ACTC	ACTG	ACTA	<b>ACCT</b>	ACCC	AOCG	ACCA	AOGT	AOGC	<b>AOGG</b>	AOGA	ACAT	<b>ACAC</b>	ACAG	ACAA
AG	AGTT	AGTC	<b>AGTG</b>	AGTA	AGCT	<b>AGOC</b>	AGOG	ACCA	AGGT	AGGC	AGGG	<b>AGGA</b>	AGAT	AGAC	<b>AGAC</b>	AGAA
AA	AATT	<b>AATC</b>	AATG	AATA	AACT	AACC	AAOG	AACA	AAGT	AAGC	AAGG	AAGA	AAAT	AAAC	<b>AAAG</b>	<b>AAAA</b>

表 4 显示了一个矩阵，其证明基于 Gerry et al. 1999. J. Mol. Bio. 292: 251 – 262 概括的标准进行的四聚引物成分选择。每种四聚体必须与其它四聚体有至少两个碱基不同。四聚体不能是回文的或与任何其它四聚体互补。选择了 36 种四聚体（粗体，下划线）；斜体序列表明排除考虑的回文四聚体。

### 引物设计：

设计 PCR 引物，以满足通常引物设计共同的标准（参见 Rubin, E. and A.A. Levy, A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res, 1996. 24(18): p. 3538 – 45; Buck, G.A., et al., Design strategies and performance of custom DNA sequencing primers. Biotechniques, 1999. 27(3): p. 528 – 36），通过计算机程序 MMP 进行实际选择。引物限制为 20 个碱基的长度（5 个四聚体），用于有效合成总的双向 PCR/测序引物。每种引物在 5' 末端包含 2 个碱基的 GC 锚，在 3' 末端包含单个 GC 锚（表 5），并且所有引物具有相

似的  $T_m (+/- 2^\circ\text{C})$  (图 27)。引物内不允许有发夹结构 (内部发夹茎  $\Delta G > -1.9 \text{ kcal/mol}$ )。也控制二聚；允许 3 个碱基的最大可接受二聚体，但它可以最终以 6 个 3' 碱基存在，并且 3' 二聚体的最大允许  $\Delta G$  是  $-2.0 \text{ kcal/mol}$ 。此外，对引物应用罚分，其中 3' 末端与组中其它 3' 末端太相似，从而防止一个引物和另一个引物的反向互补序列之间的交叉杂交。

表 5

	1-pos	2-pos	3-pos	4-pos	5-pos
1	CCAT	TGAT	TGAT	TGAT	ATAC
2	CCTA	CTCA	CTCA	CTCA	AAAG
3	CGAA	TACA	TACA	TACA	TTAG
4	CGTT	AGCC	AGCC	AGCC	AATC
5	GCAA	GACC	GACC	GACC	TGTC
6	GCTT	TCCC	TCCC	TCCC	AGTG
7	GGAC	ATCG	ATCG	ATCG	CTTG
8	GGTA	CACG	CACG	CACG	GATG
9		TGCG	TGCG	TGCG	TCTG
10		ACCT	ACCT	ACCT	
11		GTCT	GTCT	GTCT	
12		AGGA	AGGA	AGGA	
13		TTGA	TTGA	TTGA	
14		CAGC	CAGC	CAGC	
15		GTGC	GTGC	GTGC	
16		ACGG	ACGG	ACGG	
17		CTGT	CTGT	CTGT	
18		GAGT	GAGT	GAGT	
19		TCGT	TCGT	TCGT	

表 5 显示提供两个 5' 和单个 3' G/C 钳的 36 种选定的多联体。内部位置包含其余的四联体。这得到  $8 \times 19 \times 19 \times 19 \times 9$  种排列，或 493,848 种可能的组合。图 27 显示第一个通过，基于  $T_m$  的可接受的引物选择，

使 493,848 种引物减少到  $T_m$  为 64 – 66°C 的 56,246 种候选物。

表 6

引物完美序列匹配的可能性随着匹配长度需要的减少和感兴趣的基因组大小的增加而增加				
匹配长度	完美匹配的可能性 (1/(4 <sup>长度</sup> ))	腺病毒中匹配的几率% ~35K 碱基	NCBI 细菌数据库中 匹配的几率% ~488M 碱基	人类中匹配的几率% ~3B 碱基
20	9.1E-13	0.00%	0.04%	0.27%
19	7.3E-12	0.00%	0.65%	4.32%
18	4.4E-11	0.00%	5.76%	34.37%
17	2.3E-10	0.00%	35.69%	99.17%
16	1.2E-09	0.02%	97.52%	>100%
15	5.6E-09	0.12%	>100%	>100%
14	2.6E-08	0.64%	>100%	>100%
13	1.2E-07	3.29%	>100%	>100%
12	5.4E-07	15.68%	>100%	>100%
11	2.4E-06	58.16%	>100%	>100%
10	1.0E-05	99.35%	>100%	>100%
9	4.6E-05	99.77%	>100%	>100%
8	2.0E-04	>100%	>100%	>100%
7	8.5E-04	>100%	>100%	>100%
6	3.7E-03	>100%	>100%	>100%
5	1.6E-02	>100%	>100%	>100%
4	6.4E-02	>100%	>100%	>100%
3	2.5E-01	>100%	>100%	>100%
2	7.1E-01	>100%	>100%	>100%
1	1.0E+00	>100%	>100%	>100%

除报道的 PCR 对复合物样品群体中的错配的耐受性（参见例如 Rubin, E. and A.A. Levy, A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res, 1996. 24(18): p. 3538 – 45）外，感兴趣的基因组内存在的互补区的可能性不是引物设计过程中的主要考虑。尽管发现与 20 个碱基的引物的完美匹配的可能性非常低( $4^{20}$ )（表 6），发现较少非连续匹配的可能性随着感兴趣的基因组的大小显著增加。因此，对于腺病毒基因组，发现 20 个碱基中至少 10 个碱基的完美匹配的可能性是 99.35%。对于 NCBI 数据库中的序列，发现 16 个碱基的完美匹配的可能性是 97%（比腺病毒基因组高大约 100 倍）。对于人基因组中的序列，发现与 20 个碱基的引物的 17 个碱基的完美匹配的可能性是 99%（30 亿个碱基）。

由于用于产生模板片段的随机 DNA 消化，引物与基因组区域的交叉杂交的高可能性比预期产生较少的问题。因此，交叉杂交区(CHR)的效果是相当良性的。很可能 CHR 不能成功与溶液中的 PCR 引物和模板之间的完美匹配竞争。此外，在 3'末端包含错配的引物将具有显著的竞

争性缺点。即使 CHR 应该竞争预期的 PCR 引物，它将产生截短的 PCR 产物，其没有测序引物的下游位点。如果截短的产物可以驱动到捕获珠并且固定，可以导致两种情况中的一种。如果 CHR 竞争液相引物，则固定的产物将缺乏测序引物结合位点，并且将导致空的微量滴定板(PTP)孔。如果 CHR 竞争珠结合的引物，则测序引物仍然将存在，仅有的效果将是插入片段更短。上述结构都不会过度破坏序列质量。考虑到样品制备过程中使用大量基因组物质(目前 25 $\mu$ g, 含  $5.29 \times 10^{16}$  个拷贝的 35 Kb 腺病毒基因组)，可以用过量取样来提供缺乏完整 CHR 的片段，并且允许考虑的区域进行标准 PCR 扩增。

### 实施例 3：通过雾化进行样品制备

#### 通过雾化进行 DNA 制备

雾化步骤的目的是使 DNA 的大片段，如完整基因组或基因组的大部分断裂为容易进行 DNA 测序的较小分子种类。从单 DNA 模板产生的这种较小的 DNA 种类的群体称作文库。雾化可以将双链模板 DNA 剪切成 50 - 900 个碱基对的片段。剪切的文库含单链末端，所述末端可以通过 T4 DNA 聚合酶、大肠杆菌 DNA 聚合酶 I (克列诺片段) 和 T4 多核苷酸激酶的组合进行末端修复。T4 和克列诺 DNA 聚合酶都用于通过它们的 5' - 3' 聚合酶活性“填充”3' 凹进的末端 (5' 突出端)。T4 和克列诺聚合酶的单链 3' - 5' 外切核酸酶活性将除去 3' 突出端，T4 多核苷酸激酶的激酶活性将在 5' 羟基末端加上磷酸。

如下制备样品：

1. 获得 15  $\mu$ g gDNA (基因组 DNA)，并且在 10 mM TE( 10 mM Tris, 0.1 mM EDTA, pH 7.6; 参见本节末的试剂列表) 中调节到 100 $\mu$ l 的终体积。通过测量 O.D. <sub>260/280</sub> 比，分析 DNA 中的污染，所述比值是 1.8 或更高。预计最终 DNA 浓度是大约 300 $\mu$ g/ml。
2. 将 1600 $\mu$ l 冰冷的雾化缓冲液 (参见本节末) 加入 gDNA。
3. 将反应混合物置于冰冷的喷雾器(CIS - US, Bedford, MA)中。
4. 将具有保护帽的 15 ml falcon 管的帽置于喷雾器顶部(图 28A)。
5. 用 Nebulizer Clamp 组合件固定帽，所述组合件由固定的盖 (用于 falcon 管盖) 和两个橡胶 O 形环组成 (图 28B)。
6. 将喷雾器的底连接于氮供应装置，整个器件包裹在石蜡膜

( parafilm ) 中 ( 图 28C 和 28D )。

7. 使喷雾器保持垂直的同时 (图 28D 所示) , 施加 50 psi ( 磅/每平方英寸 ) 的氮 5 分钟。每隔数秒将喷雾器的底轻拍在硬表面上, 以迫使冷凝的液体到达底部。

8. 五分钟后关掉氮。压力标准化后 (30 秒), 从喷雾器除去氮源。

9. 除去石蜡膜, 打开喷雾器顶部。取出样品。转移到 1.5 ml 微量离心管中。

10. 重新安装喷雾器顶部, 使喷雾器在 500 rpm 离心 5 分钟。

11. 收集喷雾器中其余的样品。总回收是大约 700 $\mu$ l。

12. 根据制造商的说明, 用 QIAquick 柱 (Qiagen Inc., Valencia, CA) 纯化回收的样品。大体积需要柱子数次加样。用 55°C 下预热的 30  $\mu$ l 缓冲液 EB (10 mM Tris HCl, pH 8.5; 在 Qiagen 试剂盒中供应) 洗脱样品。

13. 通过 UV 光谱对样品定量 (2 $\mu$ l 在 198 $\mu$ l 水中, 1:100 稀释)。

## 酶促磨光

DNA 模板的雾化产生很多具有磨损的末端的 DNA 片段。用三种酶, 即 T4 DNA 聚合酶、大肠杆菌 DNA 聚合酶 (克列诺片段) 和 T4 多核苷酸激酶使这些末端成为平端, 并且用于连接于衔接子片段。

如下制备样品:

1. 在 0.2ml 管中按顺序加入以下试剂:

28 $\mu$ l 纯化的、雾化的 gDNA 片段

5 $\mu$ l 水

5 $\mu$ l 10 X T4 DNA 聚合酶缓冲液

5 $\mu$ l BSA (1mg/ml)

2 $\mu$ l dNTPs (10 mM)

5 $\mu$ l T4 DNA 聚合酶(3 个单位/ $\mu$ l)

50 $\mu$ l 终体积

2. 充分混合步骤 1 的溶液, 在 MJ 热循环仪 (可以使用任何准确的温育器) 上 25°C 下温育 10 分钟。

3. 加入 1.25 $\mu$ l 大肠杆菌 DNA 聚合酶 (克列诺片段) (5 个单位/ml)。

4. 充分混合反应物, 在 MJ 热循环仪上 25°C 下温育 10 分钟, 并且 16°C 下额外温育 2 小时。

5. 用 QiaQuick 柱纯化处理过的 DNA，用 55°C 下预热的 30 μl 缓冲液 EB (10 mM Tris HCl, pH 8.5)洗脱。
6. 在 0.2ml 管中合并以下试剂：  
30μl Qiagen 纯化的、磨光的、雾化的 gDNA 片段  
5μl 水  
5μl 10 X T4 PNK 缓冲液  
5μl ATP (10 mM)  
5μl T4 PNK (10 个单位/ml)  
50 μl 终体积
7. 混合溶液，置于采用 T4 PNK 程序的 MJ 热循环仪上，37°C 下温育 30 分钟，65°C 下温育 20 分钟，然后 14°C 下储存。
8. 用 QiaQuick 柱纯化样品，在 55°C 下预热的 30μl 缓冲液 EB 中洗脱。
9. 用 BioAnalyzer DNA 1000 LabChip 分析最终磨光反应的 2μl 等分物（参见下文）。

### 连接衔接子

如下进行连接衔接子的程序：

1. 在 0.2 ml 管中按顺序添加以下试剂：  
20.6μl 分子生物级水  
28μl 消化的、磨光的 gDNA 文库  
60μl 2 X 快速连接酶反应缓冲液  
1.8μl MMP (200 pmol/μl) 通用衔接子组  
9.6μl 快速连接酶  
总共 120μl

上述反应是针对 5μg 设计的，并且根据使用的 gDNA 量放大。

2. 充分混合试剂，25°C 下温育 20 分钟。将管置于冰上，直到凝胶制备用于琼脂糖凝胶电泳。

### 凝胶电泳和衔接的 gDNA 文库的提取

基因组 DNA 的雾化产生 50 – 900 bp 的文库群体。88 – bp 通用衔接子的加入使群体变为较大的大小，并且导致较大小范围（大约 130 –

800 bp) 的迁移谱。衔接子二聚体将在 88 bp 迁移, 未连接的衔接子将在 44 bp 迁移。因此, 可以从琼脂糖凝胶物理分离大小  $\geq 250$  bp 的基因组 DNA 文库, 用标准凝胶提取技术纯化。衔接的 DNA 文库的凝胶分离将导致大小范围  $\geq 250$  bp 的文库群体 (文库的大小范围可以根据应用而不同) 的回收。衔接子连接后的文库大小范围是 130 – 980 bp。应该注意, 该程序可以通过切割凝胶的不同区域而进行调整, 用于任何条带大小范围, 例如 130 – 200 bp, 200 – 400 bp, 250 – 500 bp, 300 – 600 bp, 500 – 700 bp 等的分离。用下文描述的程序用于分离 250 bp – 500 bp 的片段。

制备 150 ml 琼脂糖凝胶, 使其包含 2% 琼脂糖、1X TBE 和 4.5 $\mu$ l 溴化乙锭(10 mg/ml 储存液)。将连接的 DNA 与 10X Ready Load 染料混合, 并且加入凝胶。此外, 也在离开侧翼于样品的连接反应的两个泳道中加入 10 $\mu$ l 100 – bp 序列梯(0.1 $\mu$ g/ $\mu$ l)。100 V 下进行 3 小时凝胶电泳。当完成凝胶电泳后, 从凝胶盒取出凝胶, 转移到 GelDoc 中, 用塑料套覆盖。用 Prep UV 光显现 DNA 条带。用无菌、单用途刮刀, 从琼脂糖凝胶上切下文库群体, 其片段大小是 250 – 500 bp。尽可能迅速完成该过程, 以防止 DNA 切口。将凝胶片放置在 15 ml falcon 管中。用 Qiagen MinElute 凝胶提取试剂盒分离琼脂糖包埋的 gDNA 文库。用 BioAnalyzer DNA 1000 LabChip 分析每个分离的 gDNA 文库的等分物, 以评估 gDNA 文库群体的确切分布。

gDNA 文库的链置换和延伸, 以及用链霉亲和素珠分离单链 gDNA 文库

按照实施例 1 的描述进行有切口的双链 gDNA 文库的链置换和延伸, 不同之处是在热循环仪中 65°C 下将 Bst - 处理的样品温育 30 分钟, 并且放置在冰上备用。按照实施例 1 的描述制备链霉亲和素珠, 不同之处是用 200  $\mu$ l 1X 结合缓冲液洗涤两次和用 200  $\mu$ l nH<sub>2</sub>O 洗涤两次, 进行最终洗涤。按照下文用链霉亲和素珠分离单链 gDNA 文库。从洗涤的珠子除去水, 加入 250 $\mu$ l 解链溶液 (参见下文)。充分混合珠悬浮液, 在管旋转器上室温下温育 10 分钟。在分开的管中, 加入 1250 $\mu$ l PB (来自 QiaQuick 纯化试剂盒), 并且混合 9 $\mu$ l 20 % 醋酸。采用 Dynal MPC, 沉淀 250 $\mu$ l 解链溶液中的珠子, 小心取出上清液, 并且转移到新制备的 PB/醋酸溶液。用单 QiaQuick 纯化旋转柱纯化 1500 $\mu$ l 溶液中的 DNA。这是通过以每次加样 750 $\mu$ l, 通过相同柱子加样 2 次而进行的。用 55°C 下预

热的 15 $\mu$ l 缓冲液 EB 洗脱单链 gDNA 文库。

### 单链 gDNA 定量和储存

按照实施例 1 的描述，用 RNA Pico 6000 LabChip 对单链 gDNA 进行定量。在一些情况下，通过第二次测定对单链文库进行定量，以确保最初的 Agilent 2100 定量是准确进行的。为此，按照描述（通过荧光计进行 ssDNA 定量）进行 RiboGreen 定量，以证实 Agilent 2100 定量。如果两个估计值的差别大于 3 倍，则重复每次分析。如果定量显示两个程序之间的差别超过 3 倍，则使用更宽范围的模板与珠子比例。

按照实施例 1 中的描述进行单链 gDNA 文库的洗脱和储存。产率如下：

LabChip 分析后保留的 ssDNA 文库的终体积 = 12 $\mu$ l.

RiboGreen 分析后保留的 ssDNA 文库的终体积 = 9 $\mu$ l.

添加 TE 后 ssDNA 文库的终体积 = 18 $\mu$ l.

将等体积的 TE 加入单链 gDNA 文库储存液中。单链 gDNA 文库在缓冲液 TE 中达到  $1 \times 10^8$  个分子/ $\mu$ l。在 TE 中将储存液稀释(1/500)到 200,000 个分子/ $\mu$ l，并且制备 20 $\mu$ l 等分物。

### 雾化后的文库片段大小分布

1 $\mu$ l 材料雾化和磨光后的 Agilent 2100 DNA 1000 LabChip 分析的典型结果示于图 29A。大多数产物的大小范围分布预期落在 50 – 900 个碱基对。预期平均大小（峰顶部）是大约 450 bp。衔接子连接的文库片段的凝胶纯化的典型结果示于图 29B。

### 试剂

除非另外指出，列于实施例中的试剂代表可商购的标准试剂。例如，克列诺、T4 DNA 聚合酶、T4 DNA 聚合酶缓冲液、T4 PNK、T4 PNK 缓冲液、Quick T4 DNA 连接酶、Quick 连接缓冲液、Bst DNA 聚合酶（大片段）和 ThermoPol 反应缓冲液可以从 New England Biolabs (Beverly, MA) 获得。dNTP 混合物可以从 Pierce (Rockford, IL) 获得。琼脂糖、UltraPure TBE、BlueJuice 凝胶加样缓冲液和 Ready - Load 100bp DNA 序列梯可以从 Invitrogen (Carlsbad, CA) 购买。溴化乙锭和 2 - 丙烷可以从 Fisher (Hampton, NH) 购买。DNA 序列梯可以从 Ambion (Austin, TX)

购买。其它试剂是公知的和/或列于下表 7:

表 7: 解链溶液:

成分	需要的量	销售商	批号
NaCl (5 M)	200μl	Invitrogen	24740 - 011
NaOH (10 N)	125μl	Fisher	SS255 - 1
分子生物级水	9.675 ml	Eppendorf	0032 - 006 - 205

解链溶液包含 100 mM NaCl 和 125 mM NaOH。合并列出的试剂，并且重复混合。溶液可以在室温下储存 6 个月。

表 8:结合& 洗涤(B&W)缓冲液(2X 和 1X):

成分	需要的量	销售商	批号
UltraPure Tris - HCl(pH 7.5, 1 M)	250μl	Invitrogen	15567 - 027
EDTA (0.5 M)	50μl	Invitrogen	15575 - 020
NaCl (5 M)	10 ml	Invitrogen	24740 - 011
分子生物级水	14.7 ml	Eppendorf	0032 - 006 - 205

2X B&W 缓冲液包含 10 mM Tris - HCl (pH 7.5), 1 mM EDTA 和 2 M NaCl 的终浓度。通过合并和充分混合，合并列出的试剂。溶液可以在室温下储存 6 个月。1X B&W 缓冲液是通过混合 2X B&W 缓冲液与 picopure H<sub>2</sub>O 1:1 混合而制备的。终浓度是上文列出的浓度的一半，即 5 mM Tris - HCl (pH 7.5), 0.5 mM EDTA 和 1 M NaCl。

其它缓冲液包括以下这些：1X T4 DNA 聚合酶缓冲液: 50 mM NaCl, 10 mM Tris - HCl, 10 mM MgCl<sub>2</sub>, 1 mM 二硫苏糖醇 (pH 7.9 @ 25 °C)。TE: 10 mM Tris, 1 mM EDTA。

#### 特定试剂制备:

表 9: TE (10 mM):

成分	需要的量	销售商	批号
TE (1M)	1 ml	Fisher	BP1338 - 1
分子生物级水	99 ml	Eppendorf	0032 - 006 - 205

混合试剂，溶液可以在室温下储存 6 个月。

表 10: 雾化缓冲液:

成分	需要的量	销售商	批号
甘油	53.1 ml	Sigma	G5516
分子生物级水	42.1 ml	Eppendorf	0032 - 006 - 205
UltraPure Tris - HCl (pH 7.5, 1M)	3.7 ml	Invitrogen	15567 - 027
EDTA (0.5M)	1.1 ml	Sigma	M - 10228

将所有试剂加入（甘油最后加入）Stericup 和混合的孔中。给溶液贴标签，并且可以在室温下储存 6 个月。

表 11: ATP (10 mM):

成分	需要的量	销售商	批号
ATP (100 mM)	10μl	Roche	1140965
分子生物级水	90μl	Eppendorf	0032 - 006 - 205

混合试剂，溶液可以在 -20℃ 下储存 6 个月。

表 12: BSA (1 mg/ml):

成分	需要的量	销售商	批号
BSA (10 mg/ml)	10μl	NEB	M0203 kit
分子生物级水	90μl	Eppendorf	0032 - 006 - 205

混合试剂，溶液可以在 4℃ 下储存 6 个月。

表 13: 文库退火缓冲液, 10X:

成分	需要的量	销售商	批号
UltraPure Tris - HCl (pH 7.5, 1 M)	200 ml	Invitrogen	15567 - 027
醋酸镁, 酶级(1 M)	10.72 g	Fisher	BP - 215 - 500
分子生物级水	约 1 L	Eppendorf	0032 - 006 - 205

10 X 退火缓冲液包含 200 mM Tris (pH 7.5)和 50 mM 醋酸镁。对于该缓冲液，将 200 ml Tris 加入 500 ml picopure H<sub>2</sub>O。随后，将 10.72 g 醋酸镁加入溶液中，并且完全溶解。将溶液调节到 1000 ml 的终体积。溶液可以在 4°C 储存 6 个月。为了避免文库污染的可能性，等分缓冲液，用于单次或短期使用。

衔接子：

表 14: 衔接子 “A” (400 μM):

成分	需要的量	销售商	批号
衔接子 A (有义; HPLC - 纯化的, 硫代磷酸酯键, 44 bp, 1000 pmol/μl)	10.0μl	IDT	客户
衔接子 A (反义; HPLC - 纯化的, 硫代磷酸酯键, 40 bp, 1000 pmol/μl)	10.0μl	IDT	客户
退火缓冲液 (10X)	2.5μl	454 Corp.	前面的表
分子生物级水	2.5μl	Eppendorf	0032-006-205

为了制备该溶液，将 10μl 1000 pmol/μl 衔接子 A (44 bp, 有义)与 10μl 1000 pmol/μl 衔接子 A(40 bp, 反义)、2.5μl 10X 文库退火缓冲液和 2.5μl 水(V<sub>f</sub> = 25μl)混合。在 Sample Prep Lab 热循环仪上用 ANNEAL - A 程序 (参见下文附录)使衔接子退火。衔接子设计的更多详细内容提供于附录。

表 15: 衔接子 “B” (400 μM):

成分	需要的量	销售商	批号
衔接子 B (有义; HPLC - 纯化的, 硫代磷酸 酯键, 40 bp, 1000 pmol/μl))	10μl	IDT	客户
衔接子 B (反义; HPLC - 纯化的, 硫代磷酸 酯键, 5'生物素化的, 44 bp, 1000 pmol/μl)	10μl	IDT	客户
退火缓冲液(10X)	2.5μl	454 Corp.	前面的表
分子生物级水	2.5μl	Eppendorf	0032-006-205

为了制备该溶液，将 10 $\mu$ l 1000 pmol/ $\mu$ l 衔接子 B(40 bp,有义)与 10 $\mu$ l 1000 pmol/ $\mu$ l 衔接子 B(44 bp,反义)、2.5 $\mu$ l 10X 文库退火缓冲液和 2.5 $\mu$ l 水( $V_f = 25\mu$ l)混合。在 Sample Prep Lab 热循环仪上用 ANNEAL - A 程序(参见下文附录)使衔接子退火。退火后，合并衔接子“A”和衔接子“B”( $V_f = 50\mu$ l)。衔接子组可以在-20℃下储存备用。

表 16: 20%醋酸:

成分	需要的量	销售商	批号
冰醋酸	2 ml	Fisher	A35 - 500
分子生物级水	8 ml	Eppendorf	0032 - 006 - 205

为了制备该溶液，将冰醋酸加入水中。溶液可以在室温下储存 6 个月。

#### 衔接子退火程序:

用于引物退火的 ANNEAL - A 程序:

1. 95℃下温育 1 分钟；
2. 以 0.1℃/秒使温度降低到 15℃；和
3. 保持在 14℃。

用于末端修复的 T4 聚合酶/克列诺磨光程序:

1. 25℃下温育 10 分钟；
2. 16℃下温育 2 小时；和
3. 保持在 4℃。

用于末端修复的 T4 PNK 程序:

1. 37℃下温育 30 分钟；
2. 65℃下温育 20 分钟；和
3. 保持在 14℃。

用于链置换和有切口的双链 gDNA 延伸的 BST 程序:

1. 65℃下温育 30 分钟；和
2. 保持在 14℃。

#### 步骤 9: 单链 DNA 文库的稀释和储存

EB 缓冲液中的单链 DNA 文库：保持最终体积 = 25 $\mu$ l。

最初的储存液稀释如下。采用焦磷酸测序(Pyrosequencing AB, Uppsala, Sweden)结果，在 1X 退火缓冲液中将单链 DNA 文库稀释到 100M 分子/ $\mu$ l (通常是 1:50 稀释)。

通过在 1X 退火缓冲液中稀释到 200,000 个分子/ $\mu$ l 并且制备 30 $\mu$ l 等分物，制备单链 DNA 文库的等分物。在 -20°C 下储存。将样品用于乳状液 PCR。

#### 试剂制备：

终止溶液(50 mM EDTA): 将 100 $\mu$ l 0.5 M EDTA 与 900 $\mu$ l nH<sub>2</sub>O 混合，制备 1.0 ml 50 mM EDTA 溶液。

10 mM dNTPs 的溶液包括 10  $\mu$ l dCTP (100 mM), 10  $\mu$ l dATP (100 mM), 10  $\mu$ l dGTP (100 mM) 和 10  $\mu$ l dTTP (100 mM), 60  $\mu$ l 分子生物级水, (nH<sub>2</sub>O)。在冰上解冻所有 4 种 100 mM 核苷酸储存液。将每种核苷酸各 10 $\mu$ l 与 60 $\mu$ l nH<sub>2</sub>O 合并到终体积为 100 $\mu$ l，并且充分混合。将 1 ml 等分物分散到 1.5 ml 微量离心管中，-20°C 下储存不超过 1 年。

退火缓冲液，10X: 10 X 退火缓冲液包括 200 mM Tris (pH 7.5) 和 50 mM 醋酸镁。为制备该溶液，将 24.23 g Tris 加入 800 ml nH<sub>2</sub>O，调节到 7.5。在其中加入 10.72 g 醋酸镁，并且完全溶解。使溶液达到 1000 ml 的终体积。溶液能够在 4°C 下储存 1 个月。

10x TE: 10 X TE 包括 100 mM Tris·HCl (pH 7.5) 和 50 mM EDTA。一起加入这些试剂，充分混合。溶液可以在室温下储存 6 个月。

#### 实施例 4: 珠乳状液 PCR

可以在单个管中进行以下程序，包括捕获模板 DNA，DNA 扩增，和回收与扩增的模板结合的珠。乳状液形式确保珠子物理分离到该单管内的 100 - 200 $\mu$ m “微反应器” 中，由此多个模板能够克隆扩增。通过沿着与 DNA 捕获珠结合的寡核苷酸延伸模板，实现扩增产物的固定。典型地，固定的模板的拷贝数是每个珠子 1 千万 - 3 千万个拷贝。固定了多个拷贝的单种类核酸模板的 DNA 捕获珠准备用于分布到 PTPs 上。

PTP 表面上蚀刻的 300,000 个 75 微升的孔提供了用于以大量平行、有效和经济的方式对短 DNA 模板测序的独特测定。但是，这要求

每个反应孔中具有相当大量（数百万个拷贝）的克隆模板。本发明的方法使得用户能够通过在标准管或微量滴定板中进行的 PCR 反应，克隆扩增单链基因组模板种类。单拷贝的模板种类可以与捕获珠混合，重悬于完全的 PCR 扩增溶液中，并且乳化到微反应器（直径 100 – 200 $\mu\text{m}$ ）中，此后 PCR 扩增产生最初模板种类的  $10^7$  倍扩增。该程序比以前的方法要简单得多，并且更经济。

### 使核酸模板与捕获珠结合

本实施例描述了制备优选连接了仅仅一个独特核酸模板的珠的群体。成功的克隆扩增依赖于给每个珠子递送受控数目的模板种类(0.5 – 1)。过量种类的递送会导致混合的模板群体的 PCR 扩增，防止产生有意义的序列数据，而种类的缺乏会导致含有用于测序的模板的孔较少。这会减少测序阶段提供的基因组覆盖程度。因此，优选的是通过重复定量准确确定模板浓度，以及按照下文描述的结合方案。

### 模板质量控制

乳状液 PCR 反应的成功与模板种类的质量相关。不论对扩增阶段的小心和细节如何，不良质量的模板将阻碍成功的扩增和有意义的序列数据的产生。为了防止时间和金钱的不必要的损失，重要的是在开始该过程的乳状液 PCR 阶段之前检查模板材料的质量。优选地，在用于乳状液 PCR 之前，文库应该经过两个质量控制步骤。应该确定其含有的产物的浓度和分布。理想地，文库应该作为具有很少或不具有可见的衔接子二聚体（如约 90 个碱基）的片段的异质群体存在。同样，用 PCR 引物进行的扩增应该导致例如 300 – 500 bp 的产物污染范围。扩增产物的缺乏可能反映将衔接子适当连接于模板的失败，而任何大小的单条带的存在可能反映模板的污染。

### 制备 PCR 溶液

该阶段的主要考虑是防止偏离的扩增子对 PCR 反应混合物的污染。残留扩增子对 PCR 反应的污染是能够导致测序失败的关键问题之一。为了减少污染的可能性，应该遵循合适的实验室技术，并且应该在洁净室内在 UV 处理的层流罩中进行反应混合物制备。

### PCR 反应混合物:

为了得到 200 $\mu$ l PCR 反应混合物 (对于扩增 600,000 个珠子是足够的)，在 0.2 ml PCR 管中合并以下试剂：

表 17

	储存液	最终	微升
HIFI 缓冲液	10 X	1 X	20
处理的核苷酸	10 mM	1 mM	20
Mg	50 mM	2 mM	8
BSA	10 %	0.1 %	2
Tween 80	1 %	0.01 %	2
Ppase	2 U	0.003 U	0.333333
引物 MMP1a	100 $\mu$ M	0.625 $\mu$ M	1.25
引物 MMP1b	10 $\mu$ M	0.078 $\mu$ M	1.56
Taq 聚合酶	5 U	0.2 U	8
水			136.6
总共			200

彻底涡旋试管，在冰上储存，直到珠子与模板退火。

### DNA 捕获珠:

1. 将 600,000 个 DNA 捕获珠从储存管转移到 1.5 ml 微量离心 (microfuge)管。使用的精确量取决于标准化试剂的珠浓度。
2. 在台式微量离心机中沉淀珠子，除去上清液。
3. 在 PCR 洁净室中进行步骤 4 - 11。
4. 用 1 mL 1X 退火缓冲液洗涤珠子。
5. 在微量离心机中沉淀捕获珠。将管转 180°，再次离心。
6. 从含有珠子的管中除去除大约 10  $\mu$ l 之外的全部上清液。珠子没有被扰动。
7. 加入 1 mL 1X 退火缓冲液，将混合物温育 1 分钟。按照步骤 5 沉淀珠子。

8. 从管中除去除大约 100  $\mu\text{l}$  之外的全部材料。
9. 其余的珠子和溶液转移到 PCR 管中。
10. 通过上下吸取数次，用 150  $\mu\text{l}$  1X 退火缓冲液洗涤 1.5 mL 管。将其加入含有珠子的 PCR 管。
11. 按照步骤 5 沉淀珠子，除去除 10  $\mu\text{l}$  之外的全部上清液，小心不要扰动珠沉淀。
12. 取出定量的单链模板 DNA(sstDNA) 的等分物。终浓度是 200,000 - sst DNA 分子/ $\mu\text{l}$ 。
13. 将 3  $\mu\text{l}$  稀释的 sstDNA 加入含有珠子的 PCR 管。这等同于 600,000 个拷贝的 sstDNA。
14. 轻柔涡旋试管，以混合各内容物。
15. 在 PCR 热循环仪中，使用储存在 MJ 热循环仪上 EPCR 文件夹中的程序 80Anneal，使 sstDNA 与捕获珠退火，其中采用以下方案：
  - 65°C 下 5 分钟；
  - 以 0.1°C/秒 降低到 60°C；
  - 保持在 60°C 1 分钟；
  - 以 0.1°C/秒 降低到 50°C；
  - 保持在 50°C 1 分钟；
  - 以 0.1°C/秒 降低到 40°C；
  - 保持在 40°C 1 分钟；
  - 以 0.1°C/秒 降低到 20°C；和
  - 保持在 10°C，直到用于下一个步骤。
- 在大多数情况下，在模板结合后立即将珠子用于扩增。如果不能立即使用珠子，则应该将它们在 4°C 下储存在模板溶液中直到需要。储存后，如下处理珠子：
16. 如步骤 6，从热循环仪中取出珠子，离心，在不扰动珠子的条件下除去退火缓冲液。
17. 将珠子储存在冰桶上，直到进行乳化（实施例 2）。
18. 捕获珠包括每个珠子平均 0.5 - 1 个拷贝的与每个珠子结合的 sstDNA，并且准备用于乳化。

#### 实施例 5：乳化

下文描述了适用于此步骤的 PCR 溶液。为了得到 200 $\mu$ l PCR 反应缓冲液（足够用于扩增 600K 珠子），将以下物质加入 0.2 ml PCR 管：

表 18

	储存液	最终	微升
HIFI 缓冲液	10 X	1 X	20
处理的 Nukes	10 mM	1 mM	20
Mg	50 mM	2 mM	8
BSA	10 %	0.1 %	2
Tween 80	1 %	0.01 %	2
Ppase	2 U	0.003 U	0.333333
引物 MMP1a	100 $\mu$ M	0.625 $\mu$ M	1.25
引物 MMP1b	10 $\mu$ M	0.078 $\mu$ M	1.56
Taq	5 U	0.2 U	8
水			136. 6
总共			200

本实施例描述了如何制备含有每微升大约 3,000 个 PCR 微反应器的耐热油包水乳状液。下文概括了用于制备乳状液的方案。

1. 将 200 $\mu$ l PCR 溶液加入 600,000 个珠子（两种成分都来自实施例 1）。

2. 上下吸取溶液数次，以便将珠子重悬。  
3. 使 PCR - 珠混合物室温下温育 2 分钟，以便用 PCR 溶液平衡珠子。

4. 将 400 $\mu$ l Emulsion Oil 加入 UV 照射过的 2ml 微量离心管。

5. 将“不含扩增子的” 1/4" 磁力搅棒加入 Emulsion Oil 的试管。

不含扩增子的搅棒如下制备。用大搅棒支持 1/4" 搅棒。该搅棒随后：

- 用 DNA - Off(滴或喷雾) 洗涤；
- 用 picopure 水冲洗；
- 用 Kimwipe 边缘干燥；和
- UV 照射 5 分钟。

6. 取出 Dynal MPC - S 管支架的磁力插件。将 Emulsion Oil 的管

置于管支架中。将管放在设置为 600 rpm 的搅拌盘中心。

7. 充分涡旋管，以使珠子重悬。这确保珠子的簇集最少。

8. 采用 P - 200 移液管，将 PCR - 珠混合物以大约每 2 秒 1 滴的速度滴加到旋转的油中，使每个液滴沉积到磁力搅棒的水平，并且在加入下一滴之前乳化。溶液变成同质乳状白色液体，其粘度与蛋黄酱相似。

9. 一旦加入了全部 PCR - 珠混合物，轻轻敲打微量离心管几次，以便在具有乳状液的表面混合任何油。

10. 搅拌再持续 5 分钟。

11. 重复步骤 9 和 10。

12. 通过用更大的搅棒拖出管外，从乳化的材料除去搅棒。

13. 取出 10  $\mu\text{L}$  乳状液，置于显微镜载玻片上。用盖玻片覆盖乳状液，以 50X 放大(10X 目镜和 5X 物镜)检查乳状液。预期“良好的”乳状液主要包括油中的 PCR 溶液的分离液滴(微反应器)中的单个珠子。

14. 如下制备具有乳状液稳定剂的合适的乳状液油混合物。乳状液混合物的成分示于表 19。

表 19

成分	需要的量	来源	参考编号
Sigma 轻矿物油	94.5 g	Sigma	M - 5904
Atlox 4912	1 g	Uniqema	NA
Span 80	4.5 g	Uniqema	NA

通过在水浴中将 Atlox 4912 预热到 60°C，制备乳状液油混合物。然后，将 4.5 克 Span 80 加入 94.5 克矿物油，以形成混合物。然后，将 1 克预热的 Atlox 4912 加入混合物中。将溶液放置在密闭容器中，通过振荡和反转进行混合。通过将混合物加热到 60°C，然后再振荡，纠正 Atlox 沉降或固化的任何迹象。

#### 实施例 6：扩增

本实施例描述了在珠乳状液中扩增模板 DNA。根据本发明的方案，该过程的 DNA 扩增阶段需要 3 - 4 小时。扩增完成后，可以将乳状液放置在热循环仪上最多 12 小时，然后开始分离珠子的过程。通过将 50 -

100  $\mu\text{l}$  乳化的反应混合物放置到各个 PCR 反应室（即 PCR 管）中，进行 PCR 热循环。如下进行 PCR：

1. 用单个移液管尖，以 50 – 100  $\mu\text{L}$  的量，将乳状液转移到大约 10 个分开的 PCR 管或 96 孔板。对于该步骤，油包水乳状液是高度粘稠的。
2. 密封平板，或盖上 PCR 管的盖子，将容器置于含有或不含 96 孔板适配器的 MJ 热循环仪中。
3. 对 PCR 热循环仪编程，允许以下程序：
  - 1 循环(94°C 下 4 分钟) - 热启动；
  - 40 个循环(94°C 下 30 秒, 58°C 下 30 秒, 68°C 下 90 秒)；
  - 25 个循环(94°C 下 30 秒, 58°C 下 6 分钟)；和
  - 14°C 下储存。
4. 完成 PCR 反应后，取出扩增的材料，以进行乳状液破裂和珠回收。

#### 实施例 7：乳状液破裂和珠回收

本实施例描述如何使乳状液破裂并且回收具有扩增的模板的珠子。优选地，PCR 后的乳状液应该保持完整。通过肉眼检查，乳状液的下面的相保持乳状白色悬浮液。如果溶液是澄清的，则乳状液可以部分分成其水相和油相，并且很可能很多珠子将具有模板的混合物。如果乳状液在一个或两个管中破裂，这些样品应该不与其它样品混合。如果乳状液在所有管中破裂，则该程序不应该继续。

1. 用单个移液管尖，将来自最初 600  $\mu\text{l}$  样品的所有 PCR 反应物合并到单个 1.5 ml 微量离心管中。如上文所述，乳状液粘度非常高。在一些情况下，对于每个管，充分吸取数次。将尽可能多的材料转移到 1.5 ml 管中。
2. 通过将 50  $\mu\text{l}$  Sigma 矿物油加入每个样品中，从每个 PCR 管回收保持乳化的材料。采用单个移液管尖，将每个管上下吸取数次，以使其余的材料重悬。
3. 将该材料加入含有大量乳化材料的 1.5 ml 管中。
4. 将样品涡旋 30 秒。
5. 在 Eppendorf 微量离心机中，以 13.2K rpm 在台式微量离心管中将样品旋转 20 分钟。

6. 将乳状液分成具有大的白色界面的两相。尽可能多地除去上面的清澈的油相。在管中保留絮状材料。通常白色层分开油层和水层。通常观察到珠子沉淀在管底部。

7. 取出珠子上的水层，保留用于分析（凝胶分析，Agilent 2100，和 Taqman）。如果白色材料的界面在水层上持续，取出 20 微升下面的水层。这是通过用移液管尖穿透界面材料并且从下面吸入溶液而进行的。

8. 在 PTP 制作和表面化学室排烟罩中，将 1 ml 己烷加入其余的乳状液。

9. 将样品涡旋 1 分钟，并且全速旋转 1 分钟。

10. 在 PTP 制作和表面化学室排烟罩中，除去上面的油/己烷相，并且置于有机废物容器中。

11. 在 80% 乙醇中将 1 ml 1X 退火缓冲液加入其余的水相、界面和珠。

12. 将样品涡旋 1 分钟，直到白色物质溶解。

13. 将样品高速离心 1 分钟。将管旋转 180 度，再离心 1 分钟。取出上清液而不扰动珠沉淀。

14. 用含有 0.1% Tween 20 的 1 ml 1X 退火缓冲液洗涤珠子，并且重复该步骤。

#### 实施例 8：单链取出和引物退火

如果珠子要用于基于焦磷酸的测序反应，则必须除去 PCR 产物的第二条链，并且使测序引物与结合于珠子的单链模板退火。本实施例描述了用于实现其的方案。

1. 用 1 ml 水洗涤珠子，离心两次，每次 1 分钟。离心之间将管旋转 180°。离心后，除去水相。

2. 用 1 ml 1 mM EDTA 洗涤珠子。按照步骤 1 将管离心，并且除去水相。

3. 加入 1 ml 0.125 M NaOH，将样品温育 8 分钟。

4. 短暂涡旋样品，置于微量离心机中。

5. 6 分钟后，按照步骤 1 沉淀珠子，除去尽可能多的溶液。

6. 完成 8 分钟的 NaOH 温育后，加入 1 ml 1X 退火缓冲液。

7. 短暂涡旋样品，按照步骤 1 沉淀珠子。除去尽可能多的上清液，加入另外的 1 ml 1X 退火缓冲液。
8. 短暂涡旋样品，按照步骤 1 沉淀珠子，除去 800 $\mu$ l 1X 退火缓冲液。
9. 将珠子转移到 0.2 ml PCR 管中。
10. 转移珠子，除去尽可能多的退火缓冲液，而不扰动珠子。
11. 加入 100 $\mu$ l 1X 退火缓冲液。
12. 加入 4 $\mu$ l 100 $\mu$ M 测序引物。涡旋样品，然后立即退火。
13. 在 MJ 热循环仪中用“80Anneal”程序进行退火。
14. 用 200 $\mu$ l 1X 退火缓冲液洗涤珠子 3 次，用 100 $\mu$ l 1X 退火缓冲液重悬。
15. 在 Hausser 血细胞计数器上对珠子进行计数。典型地，回收了 300,000 – 500,000 个珠子(3,000 – 5,000 个珠子/ $\mu$ L)。
16. 4°C 下储存珠子，可以用于 1 周的测序。

#### 实施例 9：任选的富集步骤

可以用以下程序对珠子进行富集，得到含扩增子的珠子。富集不是必须的，但可以用于更有效地进行随后的分子生物技术，如 DNA 测序。

将 55 微升 10  $\mu$ M (总共 500 皮摩尔) 生物素 - 测序引物加入来自实施例 5 的含有扩增子的琼脂糖珠。将珠子置于热循环仪中。通过实施例 2 的热循环仪退火程序，使引物退火于珠子上的 DNA。

退火后，用含有 0.1% Tween 20 的退火缓冲液将琼脂糖珠洗涤 3 次。通过离心浓缩目前含有与生物素 - 测序引物退火的 ssDNA 片段的珠子，并且重悬于 200  $\mu$ l BST 结合缓冲液中。然后将 10 微升 50,000 个单位/ml 的 Bst - 聚合酶加入重悬的珠子，将容纳珠子的容器置于旋转器上 5 分钟。加入 2 微升 10 mM dNTP 混合物 (即 10 mM dATP, dGTP, dCTP 和 dTTP 各 2.5  $\mu$ l)，室温下将混合物再温育 10 分钟。用含有 0.1% Tween 20 的退火缓冲液将珠子洗涤 3 次，重悬于最初体积的退火缓冲液中。

用含有 0.1% Tween 20 的退火缓冲液将 50 微升 Dynal 链霉亲和素珠 (Dynal Biotech Inc., Lake Success, NY; 10 mg/ml 的 M270 或 MyOne<sup>TM</sup> 珠) 洗涤三次，并且以最初体积重悬于含有 0.1% Tween 20 的退火缓冲液中。然后将 Dynal 珠混合物加入重悬的琼脂糖珠。涡旋混合物，室温下

置于旋转仪中 10 分钟。

通过以 2300 g 离心(500 rpm, Eppendorf Centrifuge 5415D), 将珠子收集在试管底部。将珠子重悬于最初体积的含有 0.1% Tween 20 的退火缓冲液中。将试管中的混合物置于磁力分离器 (Dynal) 中。用含有 0.1% Tween 20 的退火缓冲液将珠子洗涤 3 次, 以最初体积重悬于相同的缓冲液中。通过上文描述的洗涤步骤除去不含扩增子的珠子。仅仅保留含有合适的 DNA 片段的琼脂糖珠。

通过加入 500  $\mu$ l 0.125 M NaOH, 从琼脂糖珠上分离磁珠。涡旋混合物, 通过磁力分离, 取出磁珠。将保留在溶液中的琼脂糖珠转移到另一个试管中, 用 400  $\mu$ l 50 mM Tris 醋酸盐洗涤, 直到 pH 稳定在 7.6。

#### 实施例 10: 用珠乳状液 PCR 进行核酸测序

进行以下实验, 以测试珠乳状液 PCR 的效率。对于该方案, 将平均直径为 25 - 35  $\mu$ m 的 600,000 个琼脂糖珠 (由制造商提供) 以每个珠子 3 千万 - 5 千万个拷贝的比例共价连接于捕获引物。将具有共价连接的捕获引物的珠子与 1200 万个拷贝的单链腺病毒文库混合。文库构建体包括与珠子上的捕获引物互补的序列。

用实施例 1 描述的程序使腺病毒文库退火于珠子。然后, 将珠子重悬于完全的 PCR 溶液中。用与实施例 2 的描述相同的程序将 PCR 溶液和珠子乳化于 2 体积的旋转中的乳化油中。通过实施例 3 描述的 PCR, 使乳化的(包封的)珠子进行扩增。按照实施例 4 的描述使乳状液破裂。用实施例 5 的程序使珠子上的 DNA 成为单链, 使测序引物退火。

随后, 通过采用来自 454 Life Sciences (New Haven, CT) 的焦磷酸测序仪进行焦磷酸测序, 同时对 70,000 个珠子进行测序 (参见与本文同时提交的 Lohman 等人的共同未决的申请, 名称为“对核酸进行扩增和测序的方法”, 即 2003 年 6 月 6 日提交的 USSN 60/476,592)。对多个批次的 70,000 个珠子进行测序, 数据列于下表 20。

表 20

比对误差耐受	比对				覆盖	推定的读出结果误差
	无	单个	多个	独特的		
0%	47916	1560		1110	54.98%	0.00%
5%	46026	3450		2357	83.16%	1.88%
10%	43474	6001	1	3742	95.64%	4.36%

该表显示了通过比较获自焦磷酸测序仪的序列与腺病毒序列，从 BLAST 分析获得的结果。第一列显示 BLAST 程序中使用的误差耐受。最后一列显示通过与已知序列直接比较确定的实际误差。

### 用于双末端测序的珠乳状液 PCR

#### 实施例 11：模板质量控制

如前面指出的，发现乳状液 PCR 反应的成功与单链模板种类的质量相关。因此，在开始乳状液 PCR 方案之前，用两个分开的质量控制评估模板材料的质量。首先，在 2100 BioAnalyzer (Agilent) 上运行单链模板的等分物。用 RNA Pico Chip 证实样品包含大小为大约 200 – 500 个碱基的片段的异质群体。其次，在 Bio-Tek FL600 板荧光计上用 RiboGreen 荧光测定对文库进行定量。确定 DNA 浓度低于 5 ng/μl 的样品被认为对于使用来说太稀了。

#### 实施例 12：DNA 捕获珠合成

从 1 mL N - 羟基琥珀酰亚胺酯 (NHS) 活化的 Sepharose HP 亲和柱(Amersham Biosciences, Piscataway, NJ)上取出塞满的珠子。通过连续通过 30 和 25 μm 孔滤网片(Sefar America, Depew, NY, USA)，选择 30 – 25 μm 大小的珠子。收集通过第一个滤网，但保留在第二个滤网上的珠子，并且按照产品文献 (Amersham Pharmacia Protocol # 71700600AP) 中的描述活化。获得两个不同的胺标记的 HEG (六甘醇) 长捕获引物，相应于要扩增的模板的有义和反义链的 5' 末端(5' – 胺 – 3' HEG 间隔基 gcttacacctgaccgtccatcccctgttgtgcgtgtc – 3'; SEQ ID NO:12; 和 5' – 胺 – 3' HEG 间隔基 ccattccccagctcgatctgtttccctccctgtc – 3'; SEQ ID

NO:13) (IDT Technologies, Coralville, IA, USA)。引物设计用于捕获扩增产物的两条链，以允许双末端测序，即对扩增产物的第一和第二条链进行测序。将捕获引物溶解于 20 mM 磷酸缓冲液, pH 8.0, 获得 1mM 的终浓度。结合 3 微升每种引物，以筛选 30 – 25 $\mu$ m 的珠子。然后将珠子储存于珠储存缓冲液(50 mM Tris, 0.02% Tween 和 0.02% 叠氮化钠, pH 8)中。用血细胞计数器(Hausser Scientific, Horsham, PA, USA)对珠子进行定量，并且在 4°C 下储存直到需要。

#### 实施例 13：PCR 反应混合物制备和制剂化

对于任何单分子扩增技术，来自其它实验的外来或残留扩增子对反应的污染会干扰测序运行。为了减少污染的可能性，在 PCR 洁净室内在 UV 处理的层流罩中进行 PCR 反应混合物制备。对于每个 600,000 个珠的乳状液 PCR 反应，将以下试剂混合在 1.5 ml 试管中：225  $\mu$ l 反应混合物(1X Platinum HiFi 缓冲液(Invitrogen)), 1 mM dNTPs, 2.5 mM MgSO<sub>4</sub> (Invitrogen), 0.1% BSA, 0.01% Tween, 0.003 U/ $\mu$ l 耐热 PPi - ase (NEB), 0.125  $\mu$ M 正向引物(5' – gcttacctgaccgacctctg – 3'; SEQ ID NO:14)和 0.125  $\mu$ M 反向引物(5' – ccattccccagctcgtctg – 3'; SEQ ID NO:15) (IDT Technologies, Coralville, IA, USA)以及 0.2 U/ $\mu$ l Platinum Hi - Fi Taq 聚合酶 Invitrogen)。取出 25 微升反应混合物，储存在单独的 200  $\mu$ l PCR 管中，用作阴性对照。反应混合物和阴性对照都储存在冰上备用。

#### 实施例 14：将模板种类结合于 DNA 捕获珠

用于测序的成功克隆 DNA 扩增与向每个珠子进行受控数目的模板种类送递相关。对于下文描述的实验，确定典型的靶模板浓度是每个捕获珠 0.5 个模板。在此浓度下，泊松分布表明 61% 的珠子没有结合的模板，30% 具有一个种类的模板，9% 具有两个或多个模板种类。过量种类的送递会导致单个珠子上混合的群体（2 个或更多种类）的结合和随后的扩增，防止产生有意义的序列数据。但是，送递太少的种类会导致含有模板的孔（每个珠子 1 个种类）较少，这会减少测序覆盖程度。因此，认为单链文库模板浓度是重要的。

通过在 UV 处理的层流罩中进行的以下方法，使模板核酸分子与 DNA 捕获珠上的互补引物退火。将 60 万个悬浮于珠储存缓冲液中的

DNA 捕获珠(参见上文实施例 9)转移到 200  $\mu\text{l}$  PCR 管中。将管在台式微型离心机中离心 10 秒，旋转 180 度，再离心 10 秒以确保均匀沉淀形成。除去上清液，用 200  $\mu\text{l}$  退火缓冲液(20 mM Tris, pH 7.5 和 5 mM 醋酸镁)洗涤珠子。将管涡旋 5 秒，以使珠子重悬，按照上文沉淀珠子。除去珠子上的除大约 10  $\mu\text{l}$  以外的全部上清液，加入额外的 200  $\mu\text{l}$  退火缓冲液。将珠子再次涡旋 5 秒，静置 1 分钟，然后按照上文沉淀。弃去除 10  $\mu\text{l}$  以外的全部上清液。

随后，在珠子中加入 1.5  $\mu\text{l}$  300,000 个分子/ $\mu\text{l}$  的模板文库。将管涡旋 5 秒以混合内容物，使模板在 MJ 热循环仪中进行的受控变性/退火程序中与珠子退火。程序是在 80°C 下温育 5 分钟，然后以 0.1°C/秒降低到 70°C，70°C 下温育 1 分钟，以 0.1°C/秒降低到 60°C，60°C 下保持 1 分钟，以 0.1°C/秒降低到 50°C，50°C 下保持 1 分钟，以 0.1°C/秒降低到 20°C，20°C 下保持。完成退火程序后，从热循环仪取出珠子，按照上文离心，小心倒去退火缓冲液。捕获珠包括与每个珠子结合的平均 0.5 个拷贝的单链模板 DNA，并且在冰上储存直到需要。

### 实施例 15：乳化

乳化过程产生了含有每微升 10,000 个离散的 PCR 微反应器的耐热油包水乳状液。其作为用于进行靶文库的各个分子的单分子克隆扩增的基质。按照以下方式对单个反应的反应混合物和 DNA 捕获珠进行乳化。在 UV 处理的层流罩中，将 200  $\mu\text{l}$  PCR 溶液(来自实施例 10)加入含有 600,000 个 DNA 捕获珠(来自实施例 11)的管中。通过反复吸取使珠子重悬。此后，在室温下将 PCR - 珠混合物温育至少 2 分钟，使得珠子在 PCR 溶液中平衡。此时，将 450  $\mu\text{l}$  Emulsion Oil(轻矿物油(Sigma)中的 4.5 % (w:w) Span 80, 1% (w:w) Atlox 4912 (Uniqema, Delaware))等分在含有无菌 1/4 英寸磁力搅拌(Fischer)的平顶 2 ml 离心管(Dot Scientific)中。然后将管放置在客户制作的塑料管支持夹具上，随后放置在设置为 450 RPM 的 Fisher Isotemp 数字搅拌烤盘(Fisher Scientific)中心。

将 PCR - 珠溶液涡旋 15 秒，以使珠子重悬。然后将溶液抽取到固定了塑料安全注射器针头(Henry Schein)的 1ml 一次性塑料注射器(Benton - Dickenson)中。该注射器置于注射器泵(Cole - Parmer)中，该泵

用铝底单位改进，使得垂直泵出而不是水平泵出（图 30）。将具有乳状液油的管排列在搅拌盘上，使得其在塑料注射器针头下集中，并且磁力搅棒正确旋转。注射器泵设置为以 5.5 ml/小时分配 0.6 ml。以逐滴方式将 PCR - 珠溶液加入乳状液油。小心确保液滴在落入旋转中的油中时不接触管壁。

一旦形成了乳状液，就必须小心使乳化过程和乳化后等分步骤中乳状液的搅动最少。发现涡旋、快速移液或过度混合，会导致乳状液破裂，破坏离散的微反应器。在乳状液的形成中，两种溶液成为同质乳状白色混合物，具有蛋黄酱的粘度。将注射器的内容物转移到旋转中的油中。然后，从支持夹具上取下乳状液管，用食指轻轻敲打，直到乳状液上的任何残留油层消失。将管重新放置在支持夹具中，用磁力搅拌再搅拌 1 分钟。通过沿试管外移动磁力回收工具，从乳状液中取出搅棒，弃去搅棒。

用 P100 移液管从试管中央取出 20 微升乳状液，放置在显微镜载玻片上。用较大的移液管尖使剪切力最小。用 50X 放大观察乳状液，以确保其主要在油包 PCR 溶液的 30 - 150 微米直径微反应器中包含单个珠子（图 33）。肉眼观察后，立即扩增乳状液。

#### 实施例 16：扩增

将乳状液等分到 7 - 8 个分开的 PCR 管中。每个管包含大约 75  $\mu$ l 乳状液。将管密封，与上文描述的 25 $\mu$ l 阴性对照一起放置在 MJ 热循环仪中。采用以下循环时间：1 个循环的 94°C 下温育 4 分钟(热启动)，30 个循环的 94°C 下 30 秒和 68°C 下 150 秒（扩增），和 40 个循环的 94°C 下温育 30 秒和 68°C 下 360 秒（杂交和延伸）。完成 PCR 程序后，取出管，立即使乳状液破裂，使反应在 10°C 维持最多 16 小时，然后起始破裂过程。

#### 实施例 17：乳状液破裂和珠回收

扩增后，检查乳状液破裂（分开油和水相）。未破裂的乳状液合并到单个 1.5 ml 的微量离心管中，而弃去偶然破裂的乳液。由于乳状液非常粘稠，相当大的量保留在每个 PCR 管中。通过在每个 PCR 管中加入 75  $\mu$ l 矿物油并且吸取混合物，回收保留在管中的乳状液。将混合物加入

含有大量乳化材料的 1.5 ml 管中，然后将 1.5 ml 管涡旋 30 秒。此后，在台式微量离心机中以 13.2K rpm（全速）将管离心 20 分钟。

离心后，乳状液分成具有大的白色界面的两相。弃去透明的、上层油相，而絮状界面材料保留在管中。在化学排烟罩中，将 1 ml 己烷加入下面的相和界面层中。将混合物涡旋 1 分钟，在台式微量离心机中全速离心 1 分钟。取出和弃去上面的油/己烷相。此后，将 1 ml 80% 乙醇/1X 退火缓冲液加入剩余的水相、界面和珠子。将该混合物涡旋 1 分钟，或直到来自界面的白色材料溶解。然后在台式微量离心机中将样品全速离心 1 分钟。将管旋转 180 度，再次离心 1 分钟。然后小心除去上清液而不扰动珠沉淀。

用含有 0.1% Tween 20 的 1 ml 退火缓冲液洗涤白色珠沉淀 2 次。弃去洗涤溶液，在上文描述的每次洗涤后使珠子沉淀。用 1 ml Picopure 水洗涤沉淀。用前面使用的离心 - 旋转 - 离心方法沉淀珠子。小心除去水相。然后按照上文描述用 1 ml 1 mM EDTA 洗涤珠子，不同之处是在沉淀和除去上清液之前以中等设置使珠子短暂涡旋 2 秒。

处理固定在捕获珠上的扩增的 DNA，获得单链 DNA。通过在碱性解链溶液中温育，除去第二链。随后在珠子中加入 1 ml 解链溶液(0.125 M NaOH, 0.2 M NaCl)。通过以中等设置涡旋 2 秒，使沉淀重悬，将管置于 Thermolyne LabQuake 管旋转器中 3 分钟。然后按照上文沉淀珠子，小心取出上清液并且弃去。通过加入 1 ml 退火缓冲液中和残留的解链溶液。此后，以中等速度将珠子涡旋 2 秒。沉淀珠子，按照上文描述除去上清液。重复退火缓冲液洗涤，不同之处是在离心后仅仅除去 800  $\mu$ l 退火缓冲液。将珠子和剩余的退火缓冲液转移到 0.2 ml PCR 管中。珠子立即使用，或在继续进入富集程序之前，在 4°C 下储存最多 48 小时。

### 实施例 18：任选的珠富集

珠团块包括具有扩增的、固定的 DNA 链的珠子和空或无效珠。如前面提到的，计算出扩增过程中 61% 的珠子缺乏模板 DNA。用富集过程选择性分离具有模板 DNA 的珠子，从而使测序效率最大。富集过程在下文中详细描述。

用离心 - 旋转 - 离心方法沉淀来自实施例 14 的单链珠，除去尽可能多的上清液，而不干扰珠子。在珠子中加入 15 微升退火缓冲液，然

后加入 2 $\mu$ l 100  $\mu$ M 生物素化的、40 个碱基的富集引物(5' - 生物素 - 四 - 乙二醇间隔基 ccattccccagctcgcttgcacatctgtccctccctgtctcag - 3'; SEQ ID NO:16)。引物与珠 - 固定的模板的 3'末端上的合并的扩增和测序位点(每个的长度都是 20 个碱基)互补。通过以中间设置涡旋 2 秒而混合溶液，在 MJ 热循环仪中用受控的变性/退火程序使富集引物与固定的 DNA 链退火。该程序由以下循环时间和温度组成：65°C 下温育 30 秒，以 0.1°C/秒降低到 58°C，58°C 下温育 90 秒，保持在 10°C。

当引物退火时，通过轻柔涡旋，使 Dynal MyOne™链霉亲和素珠重悬。随后，将 20  $\mu$ l MyOne™珠加入含有 1 ml 增强液(2 M NaCl, 10 mM Tris - HCl, 1 mM EDTA, pH 7.5)的 1.5 ml 微量离心管中。将 MyOne 珠混合物涡旋 5 秒，将管置于 Dynal MPC - S 磁铁中。顺磁珠沿微量离心管壁沉淀。小心除去上清液，并且弃去，而不扰动 MyOne™珠。从磁铁上取下管，加入 100  $\mu$ l 增强液。将管涡旋 3 秒，使珠子重悬，并且在冰上储存直到使用。

完成退火程序后，立即将 100  $\mu$ l 退火缓冲液加入含有 DNA 捕获珠和富集引物的 PCR 管中。将管涡旋 5 秒，将内容物转移到新的 1.5 ml 微量离心管中。用 200  $\mu$ l 退火缓冲液洗涤 PCR 管（其中富集引物与捕获珠退火），将洗涤溶液加入 1.5 ml 管中。用 1 ml 退火缓冲液洗涤珠子 3 次，涡旋 2 秒，按照上文沉淀。小心除去上清液。第三次洗涤后，用 1 ml 冰冷的增强液洗涤珠子 2 次。涡旋珠子，沉淀，按照上文的描述除去上清液。将珠子重悬于 150  $\mu$ l 冰冷的增强液中，将珠溶液加入洗涤的 MyOne™珠中。

将珠混合物涡旋 3 秒，在 LabQuake 管旋转器上室温下温育 3 分钟。链霉亲和素包被的 MyOne™珠结合于与 DNA 捕获珠上的固定的模板退火的生物素化的富集引物。然后以 2,000 RPM 将珠子离心 3 分钟，此后将用 2 秒的脉冲涡旋珠子，直到重悬。将重悬的珠子放置在冰上 5 分钟。此后，将 500  $\mu$ l 冷的增强液加入珠子，将管插入 Dynal MPC - S 磁铁。使珠子保持不搅动 60 秒，使沿着磁铁沉淀。此后，小心取出具有过量 MyOne™和无效 DNA 捕获珠的上清液，并弃去。

将管从 MPC - S 磁铁上取下，在珠子中加入 1 ml 冷的增强液。通过用手指轻柔敲打，使珠子重悬。重要的是此时不涡旋珠子，因为强力混合将破坏 MyOne™和 DNA 捕获珠之间的连接。珠子返回到磁铁，除

去上清液。再重复洗涤3次，以确保除去所有无效捕获珠。为了除去退火的富集引物和 MyOne<sup>TM</sup>珠，将DNA捕获珠重悬于400 μl解链溶液中，涡旋5秒，用磁铁沉淀。将具有富集的珠子的上清液转移到分开的1.5 ml微量离心管中。为了最大回收富集的珠子，将第二份400 μl解链溶液等分物加入含有MyOne<sup>TM</sup>珠的管中。按照上文涡旋和沉淀珠子。除去来自第二次洗涤的上清液，与第一团富集的珠子合并。弃去失效的MyOne<sup>TM</sup>珠的管。

将富集的DNA捕获珠的微量离心管置于Dynal MPC-S磁铁上，以沉淀任何残留的MyOne<sup>TM</sup>珠。将上清液中的富集的珠子转移到第二个1.5 ml微量离心管中并且离心。除去上清液，用1 m退火缓冲液洗涤珠子3次，以中和残留的解链溶液。第三次洗涤后，除去800 μl上清液，将其余的珠子和溶液转移到0.2 ml PCR管中。以2,000 RPM将富集的珠子洗涤3分钟，倒去上清液。随后，加入20 μl退火缓冲液和3 μl两种不同的100 μM测序引物(5' - ccatctgtccctccctgtc - 3'; SEQ ID NO:17; 和5' - cctatccccctgttgcgtgtc - 3'磷酸; SEQ ID NO:18)。将管涡旋5秒，置于MJ热循环仪中，进行以下4阶段退火程序：在65°C下温育5分钟，然后以0.1°C/秒降低到50°C，50°C下温育1分钟，以0.1°C/秒降低到40°C，40°C下保持1分钟，以0.1°C/秒降低到15°C，15°C下保持。

完成退火程序后，从热循环仪取出珠子，通过离心10秒进行沉淀。将管旋转180度，再离心10秒。倒出上清液并且弃去，在管中加入200 μl退火缓冲液。用5秒涡旋重悬珠子，按照上文沉淀。除去上清液，将珠子重悬于100 μl退火缓冲液中。此时，用Multisizer 3 Coulter计数器(Beckman Coulter)对珠子进行定量。4°C下储存珠子，并且在至少1周内稳定。

### 实施例19：双链测序

对于双链测序，使用两种不同的测序引物；即未修饰的引物MMP7A和3'磷酸化引物MMP2Bp。该过程有多个步骤。该过程在图38中图示。

1. 第一条链测序。第一条链的测序包括用DNA聚合酶通过在预定数目的循环中序贯添加核苷酸而延伸未修饰的引物。

2. 加帽：通过流动含有25 mM Tricine, 5 mM 醋酸镁, 1 mM DTT, 0.4 mg/ml PVP, 0.1 mg/ml BSA, 0.01% Tween 和每种双脱氧核苷酸各2

$\mu\text{M}$  和每种脱氧核苷酸各  $2 \mu\text{M}$  的加帽缓冲液终止第一条链的测序。

3. 清除：通过流入含 25 mM Tricine, 5 mM 醋酸镁, 1 mM DTT, 0.4 mg/ml PVP, 0.1 mg/ml BSA, 0.01% Tween 和 8.5 个单位/L 的腺苷三磷酸双磷酸酶的腺苷三磷酸双磷酸酶缓冲液，除去残留的脱氧核苷酸和双脱氧核苷酸。

4. 切割：通过流动含 5 个单位/ml 牛肠磷酸酶的切割缓冲液，从修饰的 3' 磷酸化引物的 3' 末端除去磷酸基团，解除第二个阻断的引物的阻断。

5. 继续：通过使 1000 个单位/ml 的 DNA 聚合酶流动，加入聚合酶，以捕获所有可利用的引物位点，从而激活第二个未阻断的引物。

6. 第二条链测序：通过在预定数目的循环中序贯添加核苷酸，通过 DNA 聚合酶进行第二条链的测序。

采用上文描述的方法，对金黄色葡萄球菌的基因组 DNA 进行测序。结果示于图 39。基于第一条链的 15770 个读出结果和第二条链的 16015 个读出结果，获得总共 31,785 个读出结果。其中，总共 11,799 个读出结果是配对的，8187 个读出结果不是配对的，总共覆盖 38%。

读出结果长度是 60 – 130，平均是  $95 +/ - 9$  个碱基（图 40）。基因组跨度的分布和每个基因组跨度的孔数目示于图 41。该基因组测序的代表性比对串示于图 42。

#### 实施例 20：模板 PCR

将 30 微米 NHS 琼脂糖珠与以下引物各 1 mM 偶联：

MMP1A: cgttccccctgtgtgccttg (SEQ ID NO:19)

MMP1B: ccatctgtgcgtgcgtgtc (SEQ ID NO:20)

通过以 1: 1 的体积：体积比在 PCR 主混合物中加入 50  $\mu\text{l}$  洗涤的引物偶联的珠子，在 MJ 热循环仪上的管中进行驱动到珠子的 PCR。PCR 主混合物包含：

1X PCR 缓冲液；

每种 dNTP 各 1 mM；

0.625  $\mu\text{M}$  引物 MMP1A；

0.625  $\mu\text{M}$  引物 MMP1B；

1  $\mu\text{l}$  1 个单位/ $\mu\text{l}$  Hi Fi Taq (Invitrogen, San Diego, CA)；和

约 5 – 10 ng 模板 DNA (要测序的 DNA)。

通过对 MJ 热循环仪进行以下编程，进行 PCR 反应： 94℃下温育 3 分钟； 39 个循环的 94℃下温育 30 秒， 58℃下 30 秒， 68℃下 30 秒；然后 94℃下温育 30 秒，以及 58℃下 10 分钟； 10 个循环的 94℃下温育 30 秒， 58℃下 30 秒；和 10℃下储存。

### 实施例 21：模板 DNA 制备和测序引物退火

用蒸馏水洗涤来自实施例 1 的珠子两次；用 1 mM EDTA 洗涤一次，并且用 0.125 M NaOH 温育 5 分钟。这除去了不与珠子连接的 DNA 链。然后，用 50 mM Tris 醋酸缓冲液洗涤珠子一次，用退火缓冲液： 200mM Tris - 醋酸， 50mM 醋酸镁， pH 7.5 洗涤两次。随后，在珠子中加入 500 皮摩尔测序引物 MMP7A (ccatctgttccctccctgtc; SEQ ID NO:21) 和 MMP2B - phos (cctatccccgttgcggtgc; SEQ ID NO:22)。在 MJ 热循环仪上用以下程序使引物退火： 60℃下温育 5 分钟；以每秒降低 0.1 度使温度降低到 50℃； 50℃下温育 5 分钟；以每秒降低 0.1 度使温度降低到 40℃； 40℃下温育 5 分钟；以每秒降低 0.1 度使温度降低到 10℃。然后用标准焦磷酸测序对模板进行测序。

### 实施例 22：第一条链的测序和终止

以 3000 rpm 10 分钟，将珠子旋转到 55 μm 微微量滴定板 (PTP) 中。将 PTP 放置在台架上，并且用预定数目的循环从头测序运行。通过给第一条链加帽而终止测序。通过加入 100 μl 1X AB (50 mM 醋酸镁， 250 mM Tricine)， 1000 个单位/ml BST 聚合酶， 0.4 mg/ml 单链 DNA 结合蛋白， 1 mM DTT， 0.4 mg/ml PVP (聚乙烯吡咯烷酮)， 每种 ddNTP 各 10 uM， 和每种 dNTP 各 2.5 μM， 给第一条链加帽。然后使腺苷三磷酸双磷酸酶流过，以通过加入 1X AB， 0.4 mg/ml PVP， 1 mM DTT， 0.1 mg/ml BSA， 0.125 单位/ml 腺苷三磷酸双磷酸酶除去过量的核苷酸， 温育 20 分钟。

### 实施例 23：制备第二条链用于测序

通过加入 100 μl 1X AB， 0.1 单位/ml 多核苷酸激酶， 5mM DTT， 解除第二条链的阻断。用标准焦磷酸测序 (描述于例如美国专利 6,274,320, 6258,568 和 6,210,891，在此引入作为参考) 对得到的模板进行测序。测

序方法的结果可以参见图 10F，其中用焦磷酸测序和这些实施例中描述的方法在两个末端对 174 bp 的片段进行测序。

#### 实施例 24：微微量滴定板上的核酸序列分析

将含有实施例 2 描述的扩增的核酸的微微量滴定板置于灌注室中，然后，将硫酸化酶、腺苷三磷酸双磷酸酶和荧光素酶递到微微量滴定板。

测序引物引发 DNA 合成，延伸到怀疑具有多态性的插入片段中，如图 11A - 11D 所示。首先通过连续将洗涤溶液，DNA 聚合酶和 dTTP、dGTP、dCTP 或  $\alpha$  硫代 dATP (一种 dATP 类似物)之一送递到灌注室中，延伸测序引物。连接于末端的硫酸化酶、荧光素酶和腺苷三磷酸双磷酸酶将作为测序反应的一部分释放的 PPi 转化为可检测的光。存在的腺苷三磷酸双磷酸酶降解任何未反应的 dNTP。典型地通过连接于纤维成像束的 CCD 照相机收集光 3 秒 (但 1 - 100，如 2 - 10 秒也是合适的)，此后将额外的洗涤溶液加入灌注室中，以除去过量的核苷酸和副产物。然后与聚合酶一起添加下一个核苷酸，从而重复循环。

洗涤过程中，收集的光图像从 CCD 照相机转移到计算机。通过计算机分析光发射，并且用于确定相应的 dNTP 是否掺入了延伸的测序引物。重复添加 dNTPs 和焦磷酸测序试剂，直到获得了含有怀疑的多态性的插入区的序列。

#### 实施例 25：微微量滴定板上的 PCR 扩增

微微量滴定板制备：在进一步的实施方案中，将连接于珠子的单链文库直接分布到微微量滴定板上，然后扩增每个珠子上的核酸模板（用 PCR 或其它已知扩增技术），产生足够拷贝数的模板，其将在本文公开的基于焦磷酸的测序方法中产生可检测的信号。

#### 实施例 26：PTP 上的核酸序列分析

用于序列分析和作为对照的试剂是四种核苷酸和在底物溶液中制备的  $0.1\mu\text{M}$  焦磷酸 (PPi)。底物溶液是指  $300\mu\text{M}$  荧光素和  $4\mu\text{M}$  腺苷酰硫酸 (APS) 的混合物，它们是涉及 PPi、荧光素酶和硫酸化酶的反应级联的底物。底物在测定缓冲液中制备。用于测试酶并且确定通过室的试

剂的背景水平的 PPi 浓度是 0.1  $\mu\text{M}$ 。核苷酸 dTTP、dGTP、dCTP 的浓度是 6.5  $\mu\text{M}$ ,  $\alpha\text{dATP}$  的浓度是 50  $\mu\text{M}$ 。每种核苷酸与浓度为 100 U/mL 的克列诺 DNA 聚合酶混合。

将 PTP 置于实施的仪器的下室中，将流动室与 CCD 照相机的面板连接。通过使底物（每分钟 3 ml, 2 分钟）流过该室，洗涤 PTP。此后，通过与驱动器连接的泵使一系列试剂流过该室，该驱动器经过编程而变换在不同试剂中插入管的位置。试剂的顺序、流速和流动时间是确定的。照相机设置在快速采集模式，曝光时间 = 2.5 秒。

从垫的信号输出确定为垫内所有像素上的计数的平均值。帧数等于实验中经过的时间。用图表代表不同试剂的流动。

### 实施例 27：用于微微升规模 PCR 反应的基于板的平台

#### 材料和方法

除非另外之处，所有常用的实验室化学物质是购自 Sigma (Sigma - Aldrich Corporation, St. Louis, MI) 或 Fisher (Fisher Scientific, Pittsburgh, PA)。

PicoTiterPlates<sup>TM</sup> (微微量滴定板) (25 x 75 x 2 mm) 是通过按照类似于以前描述的方式(Pantano, P. and Walt, D.R., Chemistry of Materials 1996, 8, 2832 – 2835) 在光导纤维面板上各向异性蚀刻而制造的。以三个不同的微孔深度，即 26、50 和 76  $\mu\text{m}$  对板进行蚀刻。微孔的中心与中心间距是 50  $\mu\text{m}$ ，孔直径是 39 – 44  $\mu\text{m}$  (参见图 14)，计算的孔密度是 480 个孔/ $\text{mm}^2$ 。

寡核苷酸引物的固相固定：从 1 mL NHS 活化的 Sepharose HP 亲和柱(Amersham Biosciences, Piscataway, NJ)上取出塞满的珠子，按照制造商的说明书 (Amersham Pharmacia Protocol # 71700600AP) 活化。使 20 mM 磷酸缓冲液 pH 8.0 中的 25 微升 1 mM 胺标记的 HEG 捕获引物(5' – 胺 – 3' 六甘醇间隔基 ccatctgtgcgtgcgtgtc – 3'; SEQ ID NO:23) (IDT Technologies, Coralville, IA)与珠子结合。此后，通过连续通过 30 和 25  $\mu\text{m}$  孔滤网片(Sefar America, Depew, NY, USA)，选择 36 – 25  $\mu\text{m}$  的珠子。通过第一个滤网但保留在第二个滤网上的 DNA 捕获珠收集在珠储存缓冲液(50 mM Tris, 0.02% Tween, 0.02% 叠氮化钠, pH 8)中，用血细胞计数器(Hausser Scientific, Horsham, PA)定量，4°C 下储存直到需要。

制备测试 DNA 片段：扩增测试片段来源于可商购的腺病毒血清型 5 载体 pAdEasy (Stratagene, La Jolla, CA)。用双向 PCR 引物扩增片段，所述引物的 5' 末端含有 20 个碱基的扩增区和 20 个碱基的 3' 区，其与腺病毒基因组的特定区域互补。采用这些引物，从腺病毒基因组的 12933 – 13070 和 5659 – 5767 位置扩增了两个片段，并且分别赋予标记片段 A 和片段 B。

片段 A 的正向和反向引物的序列如下：斜线(/)表示引物两个区域的分隔：正向(5' – cgttccctgtgccttg / catttgtccactaggctc – 3'; SEQ ID NO:24 – SEQ ID NO:25), 和 反 向 (5' – ccatctgtgcgtgcgtgc / accagcactcgaccacc – 3'; SEQ ID NO:26 – SEQ ID NO:27)。片段 B 的引物包括：正向(5' – cgttccctgtgccttg / tacctctccgcgtaggcg – 3'; SEQ ID NO:28 – SEQ ID NO:29), 和 反 向 (5' – ccatctgtgcgtgcgtgc / ccccgacgagacgcag – 3'; SEQ ID NO:30 – SEQ ID NO:31)。

反应条件包括 50 mM KCl, 10 mM Tris – HCl (pH 9.0), 0.1% Triton X – 100, 2.5 mM MgCl<sub>2</sub>, 0.2 mM dNTP, 正向和反向引物各 1 μM, 0.1 U/μl Taq (Promega, Madison, WI) 和 50 nmol 模板 DNA。两种模板都用包括 35 个循环的 94°C 下温育 3 秒、56°C 下 30 秒和 72°C 下 90 秒的 PCR 程序扩增。采用 PCR 引物，扩增的片段的总长度对于片段 A 是 178 bp, 对于片段 B 是 148 bp。

为了制备荧光探针，通过按照上文所述从 pAdEasy 载体进行 PCR 扩增，制备生物素化的双链荧光探针。但是，改变引物序列，以防止测试片段和探针引物区的杂交。此外，这两个片段的反向引物利用 5' 生物素和其后的 3 × 六甘醇间隔基，以便在单链探针的洗脱之前将产物固定在珠子上。

荧光片段 A 探针的正向引物的序列如下。斜线(/)表示引物两个区域的分隔：(5' – atctctgcctactaaccatgaag / catttgtccactaggctc – 3'; SEQ ID NO:32 – SEQ ID NO:33)。反向引物的序列是 5' – 生物素 3X 六甘醇间隔基 gtttctctccagcctctaccga / accagcactcgaccacc – 3'; SEQ ID NO:34 – SEQ ID NO:35。片段 B 的引物如下：正向(5' – atctctgcctactaaccatgaag / tacctctccgcgtaggcg – 3'; SEQ ID NO:36 – SEQ ID NO:37), 和反向(5' – 生物素 3X 六甘醇间隔基 gtttctctccagcctctaccga / ccccgacgagacgcag – 3'; SEQ ID NO:38 – SEQ ID NO:39)。

将荧光部分掺入核苷酸混合物中。对于片段 A，该混合物包含 0.2 mM dATP/dGTP/dCTP, 0.15 mM TTP 和 0.05 mM Alexa Fluor 488 - dUTP (Molecular Probes, Eugene, OR)。或者，用 0.2 mM dATP/dGTP/TTP, 0.15 mM dCTP 和 0.05 mM Alexa Fluor 647 - dCTP (Molecular Probes, Eugene, OR)扩增片段 B。用 QIAquick PCR 纯化试剂盒(Qiagen, Valencia, CA)纯化荧光产物。随后，室温下使生物素化的 DNA 与 1×结合洗液(5 mM Tris HCl pH 7.5, 1 M NaCl, 0.5 mM EDTA, 0.05% Tween - 20)中的 100  $\mu$ l (大约 810 万)链霉亲和素琼脂糖高性能珠(Amersham Biosciences)结合 2 小时。温育后，在 TE 缓冲液(10 mM Tris, 1 mM EDTA, pH 8.0)中将珠子洗涤 3 次，与 250  $\mu$ l 解链溶液(0.125 N NaOH/0.1 M NaCl)一起温育 2 分钟，从珠子上释放单链探针。

通过在台式离心机中短暂离心，沉淀珠子，用含有 1.9  $\mu$ l 冰醋酸的 1.25 ml 缓冲液 PB (Qiagen)中和上清液。在 QiaQuick 柱(Qiagen)上重新纯化该混合物，采用 BioRad iCycler (BioRad, Hercules, CA)，通过 TaqMan 定量，确定纯化的探针的浓度。

如下进行液相 PTPCR。将 PCR 反应混合物加入单个 14 mm x 43 mm PicoTiterPlate<sup>TM</sup>的各个孔中。为此，在 1.5 ml 微量离心管中合并 500  $\mu$ l PCR 反应混合物(1X Platinum HiFi 缓冲液 (Invitrogen, Carlsbad, CA), 2.5 mM MgSO<sub>4</sub>, 0.5% BSA, 1mM dNTPs (MBI Fermentas, Hanover, MD), 1  $\mu$ M 正向(5' - cgtttccctgtgtgccttg - 3'; SEQ ID NO:40)和反向(5' - ccatctgttgcgtgcgtg - 3'; SEQ ID NO:41)引物, 0.05% Tween - 80, 1 U/ $\mu$ l Platinum High Fidelity DNA 聚合酶(Invitrogen), 0.003 U/ $\mu$ l 耐热焦磷酸酶(USB, Cleveland, OH),和计算的每孔 5 个拷贝的片段 B 模板)。彻底涡旋该管，在冰上储存，直到组装了 PicoTiterPlate<sup>TM</sup>加样药筒。

将内部 PicoTiterPlate<sup>TM</sup>加样药筒连接于具有两个塑料夹的 PicoTiterPlate<sup>TM</sup>，所述塑料夹将硅药筒垫圈牢固地固定在 PicoTiterPlate<sup>TM</sup>表面（参见图 20）。将 PCR 反应混合物抽入 1 ml 一次性注射器，将注射器嘴插入加样药筒的输入管。加样药筒放置在末端，使得输入端口定向在药筒底，PCR 混合物缓慢加样到室中。加样时，通过 PicoTiterPlate<sup>TM</sup>的透明背板观察，以确保均匀、无气泡的送递。

加样后，将 PCR 混合物温育 5 分钟，此时从 PicoTiterPlate<sup>TM</sup>加样药筒中吸出反应混合物。从加样药筒取下 PicoTiterPlate<sup>TM</sup>，随后放置在扩

增室中(参见图 21)。PicoTiterPlate™表面用 0.25 mm 厚的 Silpad A - 2000 硅片(The Bergquist Company, Chanhassen, MN)覆盖。其上放置 25 mm x 75 mm 标准玻璃载玻片(Fisher)。将密闭的小室泡沫绝缘垫(Wicks Aircraft Supply, Highland, IL)放置在载玻片上。通过 6 个 25 mm 的闩将铝盖连接于室底部, 以密封扩增室。

一旦密封, 将扩增室放置在装有 Flat Block Alpha 单位的热循环仪 MJ PTC 225 Tetrad (MJ Research, Waltham, MA)上。扩增程序包括 94℃ 下温育 3 分钟(热启动), 然后是 40 个循环的 94℃ 下温育 12 秒, 58℃ 下 12 秒, 68℃ 下 12 秒和 10℃ 最终保持。完成 PCR 程序后, 从扩增室取出 PicoTiterPlate™, 重新连接加样药筒。用 1 ml H<sub>2</sub>O, 用一次性注射器填充药筒室, 并且室温下 10℃ 温育 20 分钟。

温育完成后, 从加样药筒取出回收溶液, 转移到 1.5 ml 的微量离心管中。用 iCycler RealTime PCR 单位(BioRad)和 FAM 标记的报道探针(Epoch Biosciences, Bothell, WA)对 PCR 产物进行定量。TaqMan 通用 PCR 主混合物(Applied Biosystems, Foster City, CA)与加入 96 孔 PCR 板的每个孔中的 0.3 μM 正向和反向引物、0.15 μM FAM 标记的探针和 27 μl 反应混合物合并。

用纯化的片段建立标准曲线(6 个标准,  $1 \times 10^9 - 1 \times 10^4$  个分子/孔), 其三份进行。用以下参数进行 PCR 扩增: 94℃ 下温育 5 分钟(热启动), 然后是 60 个循环的 94℃ 下温育 15 秒, 68℃ 下 45 秒, 4℃ 最终保持。用 iCycler Optical Systems 软件 2.3 版(BioRad)分析数据, 采用 iCycler 数据和 Microsoft Excel (Microsoft, Redmond, WA)对 PCR 产率进行定量。

类似于液相 PTPCR, 进行固相 PTPCR, 不同之处是按照下文所述通过离心扩增之前, 将 DNA 捕获珠加样到 PicoTiterPlate™孔中。此外, 完成珠沉积后, 将 PCR 混合物加样到微孔中。为了促进洗涤步骤中捕获珠的保留, 固相实验利用 50 μm 深的 PicoTiterPlate™。将 PicoTiterPlate™放置在内部建立的树脂玻璃珠加样夹具中。这类似于图 20 描述的 PicoTiterPlate™加样夹具, 不同之处是 PicoTiterPlate™夹在底部树脂玻璃板和含有入口和出口的夹具顶板之间, 并且通过具有塑料螺丝的硅垫圈密封。

通过 80℃ 下温育 3 分钟, 以每个珠子 5 个模板拷贝使模板 DNA 预先退火于 DNA 捕获珠, 此后使珠子冷却到室温 15 分钟。然后在加入 PCR

混合物之前将珠子离心到 PicoTiterPlate<sup>TM</sup>孔中。通过入口之一移液到夹具中,注入含有 10 万个琼脂糖 DNA 捕获珠(大约每 3 个 PicoTiterPlate<sup>TM</sup>孔 1 个珠子)的珠加样缓冲液(450  $\mu$ l; 1X Platinum HiFi PCR 缓冲液(Invitrogen), 0.02% Tween - 80)。每个入口孔随后用环形粘合垫(3M VHS, St. Paul, MN)密封。夹具保持 PicoTiterPlate<sup>TM</sup>的孔朝上, 并且用珠悬浮液覆盖。室温下在使用微量滴定转子的 Allegra 6 离心机 (Beckman Coulter, Fullerton, CA)中以 2000 rpm 离心 5 分钟。

离心后, 从夹具取下 PicoTiterPlate<sup>TM</sup>。按照上文对液相 PCR 所述, 将 PCR 反应混合物加入 PicoTiterPlate<sup>TM</sup>。但是, 固相 PCR 混合物省略模板, 因为模板预先与 DNA 捕获珠退火。固相 PCR 扩增程序包括额外的杂交/延伸循环, 以补偿固定的引物的较慢动力学。该程序包括 94°C 下温育 3 分钟, 进行热启动, 40 个循环的 94°C 下温育 12 秒, 58°C 下 12 秒, 68°C 下 12 秒, 然后是 10 个循环的 94°C 下温育 12 秒, 68°C 下杂交和延伸 10 分钟, 10°C 下最终保持。

完成 PCR 程序后, 从扩增室取出 PicoTiterPlate<sup>TM</sup>, 按照对液相 PCR 的描述用 1 ml H<sub>2</sub>O 洗涤。然后准备 PicoTiterPlate<sup>TM</sup>, 用于固定的 PCR 产物的杂交检测。

如下用荧光标记的探针进行杂交。完成 PTPCR 后, 除去与固定的链互补的链。为此, 室温下将整个 PicoTiterPlate<sup>TM</sup>在 0.125 M NaOH 中温育 8 分钟。通过在 50 ml 20 mM Tris - 醋酸 pH 7.5 中进行两次 5 分钟的洗涤, 中和该溶液。然后将 PicoTiterPlate<sup>TM</sup>放置在客户制备的 800  $\mu$ l 杂交室中, 用杂交缓冲液(3.5X SSC, 3.0% SDS, 20X SSC 缓冲液是 3 M NaCl; 0.3 M 柠檬酸钠)65°C 下阻断 30 分钟。用含有探针: 20 nM 荧光片段 A (Alexa - 488)和片段 B (Alexa - 647)的新的杂交缓冲液替换室的内容物。使探针与它们的靶杂交。65°C 下温育 4 小时, 同时在轨道振荡器上以 200 RPM 振荡(Barnstead International, Dubuque, IA)。

杂交后, 37°C 下用 2X SSC, 0.1% SDS 洗涤 PicoTiterPlate<sup>TM</sup>, 然后 37°C 下用 1x SSC 洗涤 15 分钟, 最后 37°C 下用 0.2x SSC 洗涤两次, 每次 15 分钟。杂交后洗涤之后, 在空气中干燥 PicoTiterPlate<sup>TM</sup>, 放置在 FLA - 8000 荧光图像分析仪(Fujifilm Medical Systems USA, Stamford, CT)中, 在 635 和 473nm 波长下扫描。将得到的 16 位 tiff 图像输入到 Genepix 4.0 (Axon Instruments, Union City, CA)中。在感兴趣的区域画出

100 个分析特征的区块，对于每种特征记录 635 和 473 荧光强度。然后将数据输出到 Microsoft Excel 进行进一步分析。

对照珠如下制备。通过从 pAdEasy 载体 PCR 扩增、纯化、固定在链霉亲和素琼脂糖高性能珠上，以及按照“荧光探针的制备”部分的描述进行链分离，制备生物素化的模板 A 和 B。但是，PCR 反应中省略荧光表达的 dNTPs。用 TE 缓冲液将沉淀的珠子洗涤 3 次，4℃下在 TE 中储存，直到沉积到 PicoTiterPlate<sup>TM</sup> 上。

## 结果

通过将含有计算的每个 PicoTiterPlate<sup>TM</sup> 孔 5 个模板拷贝的 PCR 主混合物加入 PicoTiterPlate<sup>TM</sup>，证明液相扩增。反应在具有 26、50 和 76  $\mu\text{m}$  深的孔的 PicoTiterPlate<sup>TM</sup> 中双份进行。按照材料和方法部分的描述进行 40 个循环的 PTPCR 扩增。加入添加剂，以防止二氧化硅反应容器中常规报道的有害表面效应(Kalinina, O., et al., Nucleic Acids Res. 1997, 25, 1999 – 2004; Wittwer, C.T. and Garling, D.J., Biotechniques 1991, 10, 76 – 83; Taylor, T.B., et al., Nucleic Acids Res. 1997, 25, 3164 – 3168)。

在反应混合物中引入 0.5% BSA 和 0.05% Tween – 80 不仅有效减少表面效应，也促进扩增。减少任何一种所述试剂的相对浓度对扩增有不良影响。此外，由于二氧化硅表面的聚合酶灭活特性(Taylor, T.B., et al., Nucleic Acids Res. 1997, 25, 3164 – 3168; Shoffner, M.A., Cheng, J., Hvichia, G.E., Kricka, L.J. and Wilding, P., Nucleic Acids Res. 1996, 24, 375 – 379)，升高的 Taq 浓度提供益处。高于 1 U/ $\mu\text{l}$  的浓度对于增强扩增子产率是最佳的。

PTPCR 后，回收来自 PicoTiterPlate<sup>TM</sup> 的溶液，并且通过 TaqMan 测定，对每种溶液的三份样品进行定量。用稀释的模板的标准曲线（从  $1 \times 10^9$  –  $10^4$  个分子是线性的， $r^2 = 0.995$ ）确定扩增的产物的浓度。通过用扩增的产物量除以 PicoTiterPlate<sup>TM</sup> 中的孔的总数（372,380），获得每个孔的扩增分子数。通过用该数目除以每个孔的最初模板浓度，计算每个孔的扩增量。PTPCR 扩增在所有 PicoTiterPlate<sup>TM</sup> 中都是成功的，产率是 39.5 pl 孔中的  $2.36 \times 10^6$  倍 – 50 pl 孔中的  $1.28 \times 10^9$  倍（参见下表 21）。

表 21

微微量滴定板深度 [μm]	孔体积 [pl]	平均扩增倍数 N=6	扩增倍数 SD	终产物浓度 [M]
26	39.5	2.36E+06	1.02E+06	4.96E - 07
50	76.0	1.28E+09	1.03E+09	1.40E - 04
76	115.6	9.10E+08	4.95E+08	6.54E - 05

上表显示了通过 TaqMan 测定确定的 PicoTiterPlate™ PCR 扩增。数值反映了从双份 PicoTiterPlate™ 取的三份测量值。(N=6); SD = 标准差。

通过孔体积反映产率。对于 50 μm 深的孔获得的终产物浓度( $1.4 \times 10^{-4}$  M) 显著高于 76 μm 深的孔获得的终产物浓度( $6.54 \times 10^{-5}$  M)(ANOVA 的 p 值 = 0.023), 这两者都比 26 μm 深的孔获得的终产物浓度( $4.96 \times 10^{-7}$  M)高两个数量级。50 μm 深的微孔产率代表了与低体积 PCR 相关的花费和益处的最佳平衡。在此情况下, 获得了有效浓度的最大升高和试剂的低热质量, 但面积与体积比仍然足够低, 以防止有害表面效应显著减少扩增效率。

在每个不同孔深度获得的 PTPCR 产物的最终浓度( $4.96 \times 10^{-7} - 1.4 \times 10^{-4}$  M)超过了在发生 PCR 平台效应前典型报道的最大可获得的  $10^{-8}$  M 浓度(Sardelli, A., Amplifications 1993, 9, 1 - 5)。由低微孔体积导致的引物和模板分子有效浓度较高, 增加了总体反应效率, 并且推迟了平台阶段的开始, 直到获得更高摩尔产率。或者, 该作用由 PTPCR 反应中使用的高浓度 Taq 导致, 因为升高的聚合物浓度在延迟平台效应中也显示出有效(Kainz, P., Biochim. Biophys. Acta 2000, 1494, 23 - 27; Collins, F.S., et al., Science 2003, 300, 286 - 290)。40 个循环中的扩增效率对于 26、50 和 76 μm 深的孔分别是 44.3、68.9 和 67.5%, 提供了高扩增子终浓度。在 50 μm 深的孔中观察到了最高产率。但是, 应该认识到没有进行循环数优化; 采用少得多的循环已经实现了相似的扩增产率, 从而增加 PTPCR 扩增的效率。

克隆固相 PTPCR 的实验策略从单个有效拷贝的单链 DNA 片段开始, 以通过荧光探针杂交检测的特异性珠固定的 DNA 扩增子结束, 其描述于图 22, 并且在下文详述:

第 1 阶段: 每个 PicoTiterPlate™ 孔含有由单链模板分子(单链或与

本文所示 DNA 捕获珠退火，或自由漂浮在溶液中）、溶液中的正向“F”（红色）和反向“R”（蓝色）引物，以及与 DNA 捕获珠连接的 R 引物组成。液相引物以 8:1 的摩尔比存在，F 引物过量。箭头表示 5' -> 3'DNA 方向。

第 2 阶段：最初的热循环使 DNA 模板变性，使得溶液中的 R 引物与模板分子上的互补区结合。耐热聚合酶起始引物位点的延伸（虚线），并且在随后的循环中，继续发生液相指数扩增。珠固定的引物不认为是此阶段扩增的主要导致因素。

第 3 阶段：早期 PCR。在早期指数扩增中（1 - 10 个循环），尽管溶液中的 F 引物是过量的，F 和 R 引物都相同地扩增模板。

第 4 阶段：中期 PCR。在循环 10 - 30 之间，R 引物耗尽，停止指数扩增。反应随后进入不对称扩增阶段，扩增子群体中 F 链越来越占优势。

第 5 阶段：晚期 PCR。30 - 40 个循环之后，不对称扩增继续增加溶液中 F 链的浓度。过量不含 R 链互补序列的 F 链开始退火于珠固定的 R 引物。耐热聚合酶利用 F 链作为模板，合成扩增子的固定的 R 链。

第 6 阶段：终末期 PCR。继续的热循环迫使额外退火于珠结合的引物。液相扩增此时可能最少，但固定的 R 链的浓度继续增加。

第 7 阶段：通过碱变性，除去与固定的 R 链互补的非固定 F 链。现在通过扩增子的单链 R 链群集 DNA 捕获珠。

第 8 阶段：使与 R 链互补的荧光标记的探针（绿色条）与固定的链退火。用独特的荧光团标记对特定链具有特异性的探针，得到一系列同质和异质的荧光信号，这取决于给定的 PicoTiterPlate<sup>TM</sup>孔内扩增的离散模板的数目。

最初通过使生物素化的片段 A 或片段 B 测试 DNA 片段与链霉亲和素琼脂糖珠结合，通过离心将珠子加入 50 μm 深的 PicoTiterPlate<sup>TM</sup>，并且与片段 A 和片段 B 片段的荧光标记的探针的混合群体杂交，证实荧光标记的探针的特异性。没有观察到混合信号或非特异性杂交；具有片段 A 产物的珠子展示 488 nm 信号，而片段 B 珠展示 635 nm 信号（参见图 23A 和 23B）。图 23A 和 23B 的仔细观察发现片段 B 垫中存在少量片段 A 珠，反之亦然。考虑到这些游离的珠子展示的信号的纯度，很可能它们是加样过程中一些交叉污染的产物，或在随后的洗涤步骤中从一个垫

洗涤到另一个垫中。

如图 23C 所示，荧光探针检测出片段 A 和片段 B 模板的成功固相 PTPCR 扩增。杂交探针产生的信号取决于探针内染料掺入的相对效率，反应对不同量模板 DNA 的敏感性，以及每个珠上存在的扩增产物的总量和相对量。此外，可能各个孔之间产生和保留在 DNA 捕获珠上的模板量不同，并且与每个珠结合的捕获引物的数目也可能由于珠大小分布而不同。因此，由探针杂交产生的非标准化比值应该看作是半定量的而不是定量数据。然而，杂交探针产生的荧光信号是同质片段 B 信号（红色）到同样同质的片段 A 信号（绿色），而两种信号的异质混合物（黄色程度）也是可见的。

由于对照所展示的探针特异性，以及 PicoTiterPlate<sup>TM</sup> 上同质红色和绿色珠的相当大的数目，可能非特异性探针杂交导致了异质信号。任一模板的同质珠的紧密邻近，表明异质珠由扩增期间孔之间的扩增子渗漏而导致；如果是由于孔内干扰，则可以预期观察到位于任一模板的同质珠之间的异质珠，以及同质信号通常斑驳的分布。相反，可能模板分子从它们最初的珠子解离，并且在开始离心到微孔中之前重新退火到 PicoTiterPlate<sup>TM</sup> 加样混合物中的新珠子，或随着 PCR 混合物加入 PicoTiterPlate<sup>TM</sup>，从一个珠子洗涤到另一个珠子。无论混合的模板珠的原因如何，杂交结果显示 PicoTiterPlate<sup>TM</sup> 微孔中的 PCR 扩增可以驱动足够产物到达 DNA 捕获珠，使得能够进行荧光探针杂交和检测。

## 讨论

该实施例的结果证明基于 PicoTiterPlate<sup>TM</sup> 的 PCR 减弱了很多与 DNA 扩增过程相关的因素，如试剂的高消耗、大量反应和反应时间长，这给 PCR 技术带来了“革命性的飞跃”。即使在低至 39.5 微升的反应体积，单个 PicoTiterPlate<sup>TM</sup> 上的微孔可以作为最多 370,000 个离散的反应容器，实现高产率 ( $2.3 \times 10^6 - 1.2 \times 10^9$  倍) 扩增。因此，增加了通量，并且减少了 PTPCR 的总试剂消耗；整个 26 或 76  $\mu\text{m}$  深的 PicoTiterPlate<sup>TM</sup> 中包含的反应体积分别是 15.3 和 43  $\mu\text{l}$ 。PicoTiterPlate<sup>TM</sup> 大小的增加可以进一步增加最大通量。例如，将 PicoTiterPlate<sup>TM</sup> 的尺寸增加到 40 mm x 75 mm，提供大约  $1.4 \times 10^6$  个离散的反应容器，并且，与商购 96 孔 PCR 板 (85.47 mm x 127.81 mm) 具有相同外周尺寸的

PicoTiterPlate<sup>TM</sup>可以含有多至  $5.24 \times 10^6$  个孔。

无论进行的数目和体积，液相 PCR 扩增只有有限的用途，除非可以容易和有效地回收产物。以前在平行 PCR 中的努力(Nagai, H., et al., Anal. Chem. 2001, 73, 1043 – 1047)需要蒸发液体反应混合物，留下扩增子干燥到微反应器的壁上，此后可以回收用于进一步的操作。本文公开的方法避免了引入固相扩增（PCR 产物固定于 DNA 捕获珠）而回收产物的问题。因此，PicoTiterPlate<sup>TM</sup>微孔反应的产物不是含有液相 PCR 产物的 370,000 个孔，而是结合了固定的 PCR 产物的最高达 370,000 个珠子。这些 PCR 产物适于许多核酸研究的固相方法，包括支持大量平行方法的可能能力，所述方法对含有最多至数亿个碱基的完整基因组进行测序。公开的方法的简便性将显著减少测序和其它目前需要自动装置来维持大规模克隆和 PCR 的应用的花费。

在说明书中阐述了本发明的一个或多个实施方案的公开内容。尽管与此处的描述相似或等同的方法和材料可以用于实施或测试本发明，目前描述的是优选的方法和材料。从说明书和权利要求中可以明确本发明的其它特征、目的和优点。在说明书和所附权利要求中，单数形式包括复数指标，除非上下文明确指出相反的意思。除非另外定义，本文使用的所有技术和科学术语具有与本发明所属领域技术人员通常的理解相同的含义。除非明确指出相反的意思，本文使用或考虑的技术是本领域技术人员公知的标准方法。实施方案的实施例仅仅是出于说明的目的。说明书中引用的专利和公开文献在此引入作为参考。

## 实施例 28：台架测序方法

### 步骤 1：制备 pAdEasy PCR DNA 珠

该程序用于腺病毒克隆的 384 孔板 PCR。通过用 2 M NaCl 溶液洗涤一次并且重悬于 288 ml 2 M NaCl，制备链霉亲和素 - 琼脂糖珠(12 ml)用于结合 PCR 片段。将洗涤过的珠予以 200 $\mu$ l 珠悬浮液/孔转移到 15 个 96 孔板。用 Tecan TeMo 自动设备，将 PCR 产物(25 $\mu$ l)转移到 384 深孔板中。为了使 DNA 结合于固体支持物，在每个 384 深孔板的每个孔中用 TeMo 自动设备加入 25 $\mu$ l 珠悬浮液(15,000 个珠子)，并且混合。结合反应中 NaCl 的终浓度是 1M。室温振荡下在振荡器上将结合反应温育 3 小时。通过将 384 孔板倒置在储液器上，合并微量滴定板的内容物，并

且在 Beckman Allegra 台式离心机上以 1000g 离心。将合并的珠子转移到 50 ml Falcon 管中，1000g 下离心，除去上清液。

用 100 $\mu$ l 2 M NaCl 将一百万个珠子(可移动的固体支持物)洗涤一次，然后用蒸馏水洗涤 2 次(每次 100 $\mu$ l)。在 300 $\mu$ l 解链试剂(0.1 M NaCl 和 0.125 M NaOH)中在转子上将洗涤过的珠子温育 10 分钟，以除去非生物素化的 DNA 链。以最大速度离心管子，以沉淀珠子，并且取出并弃去解链溶液。用 100 $\mu$ l 解链溶液洗涤珠子，随后用 1×退火缓冲液再洗涤三次。洗涤后，将珠子重悬于 25 $\mu$ l 1X 退火缓冲液中。

将引物 P2 (500 皮摩尔)加入珠混合，并且混合。将管中的珠混合物置于具有以下温度谱的自动化温育器(在此情况下是 PCR 热循环仪)：60℃下温育 5 分钟，以 0.1℃/秒降低到 50℃，50℃下温育 5 分钟，以 0.1℃/秒降低到 40℃，40℃下温育 5 分钟，以 0.1℃/秒降低到 4℃，4℃下保持。

退火后，小心洗涤珠子，并且重悬于 200 $\mu$ l Bst DNA 聚合酶结合溶液中。然后，处理珠悬浮液的 10 $\mu$ l 等分物(50,000 个珠子)，用于在下文描述的仪器上测序。

## 步骤 2：制备对照 DNA 珠

将 6 个对照 DNA 序列 TF 2, 7, 9, 10, 12 和 15 克隆到 pBluescript II KS +载体中，并且将质粒 DNA 用作 PCR 模板，所述 PCR 用生物素化的引物进行扩增子的固相固定。

将以下试剂加入 1.7 ml 管，以建立 PCR 混合物。

表 22

10 X HIFI 缓冲液	100 $\mu$ l
10 mM dNTP 混合物	100 $\mu$ l
50 mM MgSO <sub>4</sub>	60 $\mu$ l
5' - Bio - 3HEG - MMP1B	10 $\mu$ l
MMP1A	10 $\mu$ l
HIFI Taq 聚合酶	10 $\mu$ l
分子生物级水	690 $\mu$ l

加入 20 微升质粒模板 DNA, 将混合物等分为 50 $\mu$ l, 加入 0.2 ml PCR 管。用以下程序进行热循环: 94℃下温育 4 分钟, 39 个循环的 94℃下温育 15 秒, 58℃下 30 秒, 68℃下 90 秒和 68℃下 120 秒; 10℃下保持。

按照制造商的说明书, 用 Qiagen MinElute PCR Clean - Up 试剂盒纯化每个测试片段的扩增 DNA。用 Agilent 2100 Bioanalyzer 和 DNA 500 试剂盒和芯片评估每个测试片段 DNA 的纯度和产率。以 1 千万个 DNA 拷贝/珠将生物素化的 PCR 产物固定在琼脂糖链霉亲和素珠上。

用 2 M NaCl 溶液将珠子洗涤一次。这是通过以下步骤进行的: 加入 100 $\mu$ l, 短暂涡旋以重悬珠子, 在最大速度下离心 1 分钟以沉淀珠子, 然后除去上清液。随后用 2 M NaCl 进行第二次洗涤。然后将珠子重悬于 30 $\mu$ l 2 M NaCl 中。将 PCR 产物加入珠子。涡旋混合物, 将珠子重悬于溶液中, 然后放置在滴定板振荡器的架子上, 以速度 7, 室温下振荡 1 小时。

通过室温下在高架转子上与碱性解链溶液(0.1 M NaOH / 0.15 M NaCl)一起温育 10 分钟, 除去非生物素化的第二条链。此后用 100 $\mu$ l 解链溶液洗涤珠子一次, 然后用 100 $\mu$ l 1X 退火缓冲液(50 mM Tris - 醋酸, pH 7.5; 5 mM MgCl<sub>2</sub>)洗涤 3 次。通过以最大速度离心 1 分钟, 使测序引物退火于固定的单链 DNA。除去上清液, 使珠子重悬于 25 $\mu$ l 1X 退火缓冲液中。随后, 在珠悬浮液中加入 5 $\mu$ l 测序引物 MMP7A (100 pmol/ $\mu$ l), 用以下温度谱使测序引物杂交:

- 60℃下温育 5 分钟;
- 以 0.1℃/秒降低到 50℃;
- 50℃下温育 5 分钟;
- 以 0.1℃/秒降低到 40℃;
- 40℃下温育 5 分钟;
- 以 0.1℃/秒降低到 4℃; 和
- 4℃下保持。

用 100 $\mu$ l 1X 退火缓冲液洗涤珠子两次, 然后用 1X 退火缓冲液重悬于 200 $\mu$ l 的终体积, 并且在贴标签的管中以 10 $\mu$ l 等分物在 4℃的冰箱中储存。

### 步骤 3: 测序化学

室温下在转子上将具有固定的单链 DNA 模板和退火的测序引物的琼脂糖珠与大肠杆菌单链结合蛋白(Amersham Biosciences) (5  $\mu$ l 2.5 $\mu$ g/ $\mu$ l ssb 储存液/50,000 个珠子)和 200 $\mu$ l Bst 聚合酶结合溶液(25 mM Tricine pH 7.8; 5 mM 醋酸镁; 1 mM DTT; 0.4 mg/ml PVP MW 360,000)中的 500 U (10  $\mu$ l 50U/ $\mu$ l) Bst DNA 聚合酶(NEB)一起温育 30 分钟。此后，将 DNA 珠与 SL 珠混合，并且按照下文沉积在微微量滴定板的孔中。在 454 仪器上进行测序运行所需的试剂包括 1)底物洗涤溶液；2)含腺苷三磷酸双磷酸酶的洗涤溶液；3)100 nM 无机焦磷酸校准标准物；4)各个核苷酸三磷酸溶液。

所有溶液都是在具有酶底物(25 mM Tricine pH 7.8; 5 mM 醋酸镁; 0.4 mg/ml PVP MW 360,000; 0.01% Tween 20; 300  $\mu$ M D - 荧光素; 4  $\mu$ M APS)的硫酸化酶荧光素酶测定缓冲液中制备的。底物洗涤溶液与荧光素酶测定缓冲液相同。含腺苷三磷酸双磷酸酶的洗涤溶液是基于荧光素酶测定缓冲液，不同之处是没有加入酶底物 (APS 和 D - 荧光素)，并且该洗涤溶液含有终浓度为 8.5 U / 1 的腺苷三磷酸双磷酸酶 (Sigma St. Louis, MO; Pyrosequencing AB, Pyrosequencing, Inc. Westborough, MA)。

通过将焦磷酸钠十水合物(Sigma, St. Louis, MO)加入荧光素酶测定缓冲液中达到 100 nM 的终浓度，制备焦磷酸钠(PP<sub>i</sub>)标准物。在荧光素酶测定缓冲液中将核苷酸三磷酸(dCTP, dGTP, TTP; 最小二磷酸级) (Amersham Biosciences AB, Uppsala, Sweden)稀释到 6.5  $\mu$ M 的终浓度。脱氧腺苷三磷酸类似物，即 2' - 脱氧腺苷 5'O(1 - 硫代三磷酸)，Sp - 异构体(Sp - dATP -  $\alpha$  - S, Biolog Life Science Institute, Bremen, Germany)在荧光素酶测定缓冲液中稀释到 50  $\mu$ M 的终浓度。

#### 步骤 4：克隆 His6 - BCCP - 硫酸化酶和 His6 - BCCP - 荧光素酶

将嗜热脂肪芽孢杆菌(Bst) ATP 硫酸化酶 (E.C. 2.7.7.4)和萤火虫(Photinus pyralis) 荧光素酶(E.C. 1.13.12.7)克隆到 Nhe I - BamH I 消化的 pRSET - A 载体(Invitrogen)中。用 BCCP (生物素羧基载体蛋白)基因的编码序列(Alix,J.H., DNA 8 (10), 779 - 789 (1989); Muramatsu, S. and Mizuno, T., Nucleic Acids Res. 17 (10), 3982 (1989), Jackowski, S. and Alix, J.H., J. Bacteriol. 172 (7), 3842 - 3848 (1990); Li, S.J. and Cronan, J.E. Jr., J. Biol. Chem. 267 (2), 855 - 863 (1992), Genbank 登录号 M80458)

设计用于扩增相应于 BCCP 蛋白的氨基酸 87 – 165 的片段。正向引物是 5' – ctagctagcatggaagcgccagcagca – 3'; SEQ ID NO:42, 反向引物是 5' – ccggatccctcgatgacgaccagcggc – 3'; SEQ ID NO:43。PCR 鸡尾酒制备为混合物 1 和混合物，各 25  $\mu$ l。混合物 1 包括 75 皮摩尔引物，100 ng 大肠杆菌基因组 DNA 和 5 微摩尔 dNTPs。混合物 2 包括 1 个单位的 Fidelity Expand DNA 聚合酶 (Boehringer Mannheim/Roche Diagnostics Corporation, Indianapolis, IN, Cat. No. 1 732 641) 和 5  $\mu$ l 10X Fidelity Expand 缓冲液(Boehringer Mannheim/Roche Diagnostics Corporation, Indianapolis, IN)。为了进行 PCR 热启动，分别在 96°C 下将混合物 1 和混合物 2 加热 20 秒，然后合并。合并的反应如下循环：96°C 下温育 3 分钟，10 个循环的 96°C 下 30 秒，55°C 下 1 分钟，和 68°C 下 2 分钟，然后是 20 个循环的 96°C 下温育 30 秒，60°C 下 1 分钟，和 68°C 下 2 分钟，然后是 72°C 下温育 7 分钟的磨光步骤。PCR 后，获得单个 250 bp 片段。用 Nhe I 和 BamH I 消化 BCCP 片段，亚克隆到 Nhe I – BamH I 消化的 pRSET – A 中。

#### 步骤 5：表达硫酸化酶和荧光素酶

采用分别含有 Pst I/Hind III 和 BamH I/Xho I 位点（第一个酶在 5' 末端，第二个酶在 3' 末端）的引物通过 PCR 扩增 Bst ATP 硫酸化酶和 *P. pyralis* 荧光素酶的读码框。这产生了 6X His 和 BCCP 结构域与 ATP 硫酸化酶和荧光素酶的 N 末端融合体。用补充生物素的生长培养基在大肠杆菌中表达酶，使得能够通过 BCCP 结构域进行体内生物素化。用 IMAC 和大小排阻柱层析的组合，将酶纯化到几乎同质。用 Agilent 2100 Bioanalyzer 在 Protein 200 Plus 芯片上通过电泳评估纯化。

#### 步骤 6：荧光素酶和硫酸化酶的固相固定

通过 ATP 硫酸化酶和荧光素酶的 1:3 混合物的温育，分别将酶固定在 Dynal M – 280 链霉亲和素包被的磁性微粒(Dynal, Oslo, Norway)和 Bangs 微球 (300 nm) 上。通过将 50  $\mu$ g ATP 硫酸化酶和 150  $\mu$ g 荧光素酶与 1 mg Dynal M – 280 珠或 0.6 mg Bangs 微球在 TAGE 缓冲液(25 mM Tris – 醋酸 pH 7.8, 200 mM 硫酸铵, 15% v/v 甘油和 30% v/v 乙二醇) 中混合，进行结合。4°C 下在转子上将混合物温育 1 小时。结合后，可以将

珠子在 -20°C 下在酶溶液中储存 3 个月。使用前，在含有 0.1 mg/ml 牛血清白蛋白(Sigma, St Louis, MO)的荧光素酶测定缓冲液中彻底洗涤珠子。用光度计(Turner, Sunnyvale, California)测定固定的酶的活性。将洗涤过的珠子储存在冰上，直到沉积在 PTP 载玻片上。

#### 步骤 7: PicoTiterPlates™ (PTPs)

PicoTiterPlates™ (25 x 75 x 2 mm)是通过按照类似于文献描述的方式在光导纤维面板上各向异性蚀刻而制造的。以三个不同的微孔深度，即 26、50 和 76 mm 对板进行蚀刻。微孔的中心与中心间距是 50 μm，孔直径是 39 – 44 μm (参见图 14)，计算的孔密度是 480 个孔/mm<sup>2</sup>。

#### 步骤 8: PTP 加样

用基于离心的方法，将携带 DNA 模板的琼脂糖珠和具有固定的硫酸化酶和荧光素酶的 Dynal M - 280 / Bangs 0.3 μm 珠混合物沉积在微微量滴定板的各个孔中。该程序使用内部聚碳酸酯固定设备（夹具），其包括底板（具有载玻片固定栓）、弹性体密封垫圈和具有两个加样端口的顶板。将 PTP 片放置在底板上，蚀刻侧朝上，具有适当位置的密封垫圈的顶板夹在 PTP 片的顶部。用四个塑料螺丝拉紧整个组合件，以便提供防水密封。密封垫圈设计用于遮蔽珠沉积，得到覆盖大约 270,000 PTP 个孔的六边形区域(14 x 43 mm)。

珠子沉积于有序的层中。将 PTP 转移离开珠洗涤缓冲液的温育。在第 1 层中沉积了 DNA 和酶珠的混合物。离心后，从 PTP 吸出第 1 层上清液，第 2 层中沉积 Dynal 酶珠。

通过将 120 μl ssb/Bst pol 结合混合物 (参见上文) 中的携带 150,000 个 DNA 的琼脂糖珠与 270 μl Dynal - SL 和 Bangs - SL 珠(都是 10 mg/ml) 在含有 0.1 mg/ml 的牛血清白蛋白的总体积为 500 μl 的荧光素酶测定缓冲液中混合，制备珠悬浮液。涡旋珠浆，并且通过移液端口流入珠沉积夹具中。小心避免引入气泡。在装有 4 位置板转出 (swing - out) 转子的 Beckman Allegra 6 离心机中以 2000 rpm 将夹具/PTP 组合件离心 8 分钟。离心后，用移液管从夹具室中小心取出上清液。沉积仅有 Dynal - SL 珠的第二层。该层包括 1.5 ml 管(2.5 mg/ml Dynal 珠)中的 125 μl Dynal - SL ( 10 mg/ml)和 375 μl 珠洗涤缓冲液。将 Dynal 珠混合物移液

到 PTP 主活跃区，并且 2000 rpm 下离心 8 分钟。吸出第二层混合物，将 PTP 放回珠洗涤缓冲液(含有 0.1 mg/ml 牛血清白蛋白和 8.5 U/l 腺苷三磷酸双磷酸酶的荧光素酶测定)中直到准备加入测序仪。

#### 步骤 9：测序仪器

内部测序仪器包括三个主要组合件：射流子系统 (fluidics subsystem)、PTP 药筒/流动室和成像子系统。射流子系统包括试剂储存器、试剂入口管、多阀门歧管和蠕动泵。其允许试剂送递到流动室，一次一种试剂，以预先编程的流速和持续时间流动。PTP 药筒/流动室的设计方式使得在连接 PTP 后，在 PTP 顶(蚀刻侧)和室顶之间存在 300 μm 的间隙。其包括用于控制试剂和 PTP 温度的工具，如不透光的遮蔽物。PTP 的磨光侧暴露于 PTP 药筒的背侧，并且直接与成像系统接触。成像系统包括具有 1-1 成像束的 CCD 照相机，以及用于照相机的低温冷却系统，和照相机控制电子元件。使用的照相机是具有 Fairchild Imaging LM485 CCD (1600 万像素, 15 μm 像素大小)的 Spectral Instruments (Tucson, AZ)600 系列照相机。其直接连接于具有 6 μm 纤维距的成像纤维束。照相机冷却到 -70 °C，以帧转移模式操作。以此方式，用 CCD 的中心部分进行成像，而 CCD 的外部用于图像储存和读出。读出是通过 CCD 每个角的 4 个端口进行的。数据采集速度设置为每 30 秒 1 帧。帧转移转换时间是大约 0.25 秒。所有照相机图像以 UTIFF 16 形式储存在计算机硬盘(IBM eServer xSeries 335, IBM, White Plains, NY)中。

#### 步骤 10：测序运行条件

通过射流系统的预先编程的操作，实现测序试剂循环送递到 PTP 孔，和从孔中洗去测序反应副产物。以 Microsoft Excel 写入形式写程序，指出每个写入步骤中试剂名称(洗涤, dATP $\alpha$ S, dCTP, dGTP, dTTP, PP<sub>i</sub> 标准物)、流速和持续时间。对于所有试剂，流速设置为 3 ml/分钟，流动室中的线性速率是近似的。最初洗涤步骤 (5 分钟) 后是 PP<sub>i</sub> 标准流动 (2 分钟)，然后是 21 或 42 个循环的 (洗涤 - C - 洗涤 - A 洗涤 - G - 洗涤 T)，其中每个核苷酸流动是 0.5 分钟，洗涤步骤是 2 分钟。在所有的核苷酸添加和洗涤循环后，送递第二个 PP<sub>i</sub> 标准流 (2 分钟)，然后是最终的 5 分钟洗涤步骤。总运行时间是 4 小时。完成该运行写入

所需的试剂体积如下：300 ml 每种洗涤溶液，50 ml 每种核苷酸溶液，20 ml PP<sub>i</sub> 标准溶液。运行过程中，所有试剂都保持在室温。由于流动室和流动室入口管保持在 3°C，所有进入流动室的试剂都是 30°C。

### 实施例 29. 在开放的显微制作的高密度微微升反应器中的基因组测序

我们描述了一种可放大的、高度平行的测序系统，其总通量显著高于现有技术的毛细管电泳仪器。该装置利用一种新的 60x60 mm<sup>2</sup> 光导纤维片，其含有 1,600,000 个孔，并且能够在 4 小时运行中以 99 % 或更好的准确度(phred 20)对 2 千 5 百万个碱基进行测序。为了提供测序模板，我们在乳状液的液滴中的珠子上克隆扩增了 DNA 片段。携带模板的珠子加入孔中，将每个孔转变为微微升规模测序反应器。我们通过用优化用于固体支持物的焦磷酸测序方案和开放反应器的小尺寸进行合成，从而进行测序。此处我们通过鸟枪测序和从头组装生殖道支原体基因组显示了该系统的用途、通量、准确度和耐用性，其中机器的一次运行以 99.96 % 的准确度覆盖 96 %。

DNA 测序显著改变了生化研究和医学的性质。对大量 DNA 测序所需的花费、复杂性和时间的减少，包括对细菌和真核生物基因组进行测序的能力的改进，将具有显著的科学、经济和文化影响。大规模测序项目，包括完整基因组测序，通常需要将 DNA 片段克隆到细菌载体、扩增和纯化各个模板，然后是用荧光链终止核苷酸类似物(Prober, J. M. et al., Science 238, 336 (1987))和厚凝胶或毛细管电泳进行 Sanger 测序 (Sanger, F., et al. Proc. Natl. Acad. Sci. USA 74, 5463 (1977))。目前估计人基因组测序花费是 1000 万 - 2500 万美元(NIH News Release, October 14, 2004)。已经描述了替代的测序方法(Nyren, P., Pettersson ,B., Uhlen, M., Anal. Biochem. 208, 171 (1993); Ronaghi, M. et al., Anal. Biochem. 242, 84 (1996); Jacobson, K. B. et al., GATA 8, 223 (1991) ; Bains, W. and Smith, G. C., J. Theor. Biol. 135, 303 (1988); Jett, J. H. et al., Biomol. Struct. Dynamics 7, 301 (1989))。但是，还没有技术能够替代利用细菌载体和 Sanger 测序作为序列信息的主要发生器。

在本文中，我们描述了一种集成系统，它们的通量通常使得能够进行需要数百万个序列信息碱基的应用，包括完整基因组测序。我们的注意力集中在共同开发基于乳状液的方法(Tawfik, D. S., Griffiths, A. D.,

Nat. Biotechnology 16, 652 (1998); Ghadessy, F. J., Ong, J. L., Holliger, P., Proc. Nat. Acad. Sci. USA 98, 4552 (2001); Dressman, D., Yan, H., Traverso, G., Kinzler, K. W., Vogelstein, B., Proc. Nat. Acad. Sci. USA 100, 8817 (2003))来体外分离和扩增 DNA 片段，和制作在微微升大小的孔中进行基于焦磷酸测序(“pyrosequencing”；Ronaghi, M. et al., Anal. Biochem. 242, 84 (1996)；Ronaghi, M., Uhlen, M., Nyren, P., Science 281, 363 (1998)) 的基质和仪器。

在一种典型的运行中，我们建立了具有 phred 20 或更好的质量的评分的超过 2500 万个碱基(预测具有 99 % 或更高的准确度)。尽管该 phred 20 质量通量显著高于通过毛细管电泳进行的 Sanger 测序，但目前具有显著更短的读出结果和更低的平均单个读出结果准确度的缺点(目前的基于 Sanger 的毛细管电泳测序系统产生来自 96 个 DNA 模板中每个模板的最多 700 bp 的序列信息，1 小时内的平均读出结果准确度是 99.4%，或每小时 6 万 7 千个碱基，基本所有碱基都具有 phred 20 或更好的质量；Applied Biosystems 3730xl DNA Analyzer Specification Sheet, 2004))。我们通过对已知的细菌基因组，即生殖道支原体(580 kbp)进行测序，并且将我们的鸟枪测序和从头组装与最初获自该基因组的结果(Fraser, C. M. et al., Science 270, 397 (1995)进行比较，进一步鉴定了系统的性能，并且证明了可以从相对短的读出结果从头组装细菌基因组。鸟枪测序和从头组装较大细菌基因组，即肺炎链球菌(Tettelin, H. et al., Science 293, 498 (2001); 2.1 Mbp)的结果描述于表 28。

### 基于乳状液的样品制备

我们通过剪切完整基因组，并且通过有限稀释(补充方法：文库制备)分离单 DNA 分子，制备 DNA 片段的随机文库。具体地，我们使完整基因组随机断裂，在片段中加入特化的通用衔接子，将各个片段捕获在它们自身的珠子上，并且在乳状液的液滴内，克隆扩增各个片段(图 43A 和 43B)。与目前的测序技术不同，我们的方法不需要在细菌中亚克隆或操作单个克隆；模板在乳状液中批量操作(Tawfik, D. S., Griffiths, A. D., Nat. Biotechnology 16, 652 (1998); Ghadessy, F. J., Ong, J. L., Holliger, P., Proc. Nat. Acad. Sci. USA 98, 4552 (2001); Dressman, D., Yan, H., Traverso, G., Kinzler, K. W., Vogelstein, B., Proc. Nat. Acad. Sci. USA 100,

8817 (2003))。

### 在制作的微微升大小的反应容器中测序

通过用设计用于利用小规模孔的改进的焦磷酸测序方案，我们通过在光导纤维片的开孔中同时合成，而进行测序。光导纤维片是通过将通过重复牵引和熔凝光导纤维获得的光导纤维块切片而制造的。在每次重复时，随着六边形地包装成横切面积大小不断增加的束，各个纤维的直径减少。每个光导纤维核心的直径是  $44\mu\text{m}$ ，由  $2 - 3\mu\text{m}$  的覆层围绕；每个核心的蚀刻产生深度为大约  $55\mu\text{m}$  的反应孔，中心与中心的距离是  $50\mu\text{m}$  (图 43C)，导致计算的孔大小是  $75 \text{ pL}$ ，孔密度是  $480 \text{ 个孔/mm}^2$ 。含有大约 1600 万个孔的该光导纤维片 (Leamon, J. H. et al., Electrophoresis 24, 3769 (2003)) 中加入珠子，并且封固在设计用于产生  $300\mu\text{m}$  高的通道的流动室中，所述通道在孔开口上面，测序试剂通过所述开口流动 (图 44, A 和 B)。光导纤维片的未蚀刻的底部与第二个光导纤维成像束接触，所述成像束连接于 CCD 传感器，使得能够从每个单独的孔的底部捕获发生的光子 (图 44C 和补充方法：成像系统)。

我们开发了一种三个珠的系统，并且优化各个组件，从而在固体支持物上实现高效率。微微升大小的孔、由小珠导致的酶加样统一性、和增强的固体支持物化学的组合使得我们能够开发一种方法，该方法将通过合成进行的测序的有用读出结果长度延伸到 100 bp (补充方法：测序)。

在流动室中，循环送递的试剂垂直流入孔中。这种构型使得能够在开放的孔中在携带模板的珠子上同时进行延伸反应，并且依赖对流和扩散输送，来控制试剂和副产物的添加或去除。目前的构型中扩散进入孔中和扩散出孔外的时间规模是在 10 秒的数量级，并且取决于孔深度和流动通道高度。产生信号的酶促反应的时间规模是在 0.02 – 1.5 秒的数量级 (补充方法：孔间扩散)。目前的反应主要是通过质量转移作用，并且基于试剂更快送递的改进是可能的。基于许多补充要求，选择孔深度：(i)孔需要足够深，以便在存在通过孔的对流输送时携带 DNA 的珠子保持在孔中，(ii)它们必须足够深，以提供与副产物从发生掺入的孔扩散到未发生掺入的孔的足够分离，和(iii)它们必须足够浅，以允许核苷酸快速扩散到孔中，并且在每个流动循环末，使剩余的核苷酸快速洗出，

以便能够实现高测序通量和试剂使用的减少。每种核苷酸的流动后，用含有核酸酶的洗液确保在引入下一个核苷酸之前任何孔中都不保留核苷酸。

### 各个珠子的碱基调用

通过无机焦磷酸(PPi)的相关释放和光子的产生，检测核苷酸掺入(Ronaghi, M. et al., *Anal. Biochem.* 242, 84 (1996); Ronaghi, M., Uhlen, M., Nyren, P., *Science* 281, 363 (1998))。通过检测读出结果开始时的已知四核苷酸“关键”序列，鉴定含有携带模板的珠子的孔（补充方法：图像处理）。扣除原始信号的背景，标准化，并且校正。特定孔的每个核苷酸流动的标准化信号强度表明掺入（如果有）的核苷酸数目。将信号的线性至少保留到长度 8 的均聚物（图 52）。在通过合成进行的测序中，每个珠子上非常少的模板失去了同步性（即在顺序上在其它模板之前或之后；Ronaghi, M., *Genome Research* 11, 3 (2001))。该作用主要是由于孔中残留的核苷酸（产生“结转”）或由于不完全延伸。典型地，我们观察到 1 – 2% 的结转率和 0.1 – 0.3% 的不完全延伸率。这些转变的校正是关键的，因为同步性的失去是降低以更长读出结果长度进行测序的质量的累计效应。我们基于作为基础的物理现象的详细模型开发了一些算法，使得我们能够确定并且校正各个孔中存在的结转和不完全延伸的量（补充方法：信号处理）。图 45 显示了处理的结果，即在下文讨论的生殖道支原体运行中产生了 113 bp 长读出结果。为了独立于基于乳状液的样品制备过程中引入的伪迹而评估该校正算法的性能和有效性，我们建立了具有难以测序段的测试片段，其具有长度增加的相同片段（均聚物）（补充方法：测试片段以及图 46）。采用这些测试片段，我们证实了在各个读出结果水平，我们在超过 100 bp 的读出结果长度，实现了大约 99.4% 的碱基调用准确度（表 23）。

### 高质量读出结果和共有区准确度

在碱基调用或读出结果比对之前，我们不依赖于要测序的基因组或模板的已知信息，选择了高质量读出结果（补充方法：高质量读出结果）。这种选择是基于以下观察结果，即低质量读出结果具有高比例的不允许明确区分未掺入核苷酸的流动和掺入一个或多个核苷酸的流动之间的

信号。当碱基调用各个读出结果时，可能发生错误，这是因为具有含糊值的信号（图 51）。为了改进我们的读出结果的可用性，我们也开发了一种公制，其允许我们从头开始估计读出结果的每个碱基的质量（或校正碱基调用的可能性），类似于目前的 Sanger 测序使用的 phred 评分 (Ewing, B., Hillier, L., Wendl, M.C., Green, P., Genome Research 8, 175 (1998))（补充方法：质量评分和图 54）。

通过利用我们的系统提供的高度过量取样和建立共有序列，可以实现高质量序列。可以用每个核苷酸流动，而不是各个碱基调用的信号强度使序列彼此比对，以确定最优化比对（补充方法：流动间隙作图，共有区准确度和基因组覆盖）。然后对相应的信号取平均值，此后进行碱基调用。这种方法显著改进了序列的准确度（图 53），并且提供了共有区碱基的质量估计。我们将该质量测量值称作 Z - 评分；它是在一个位置上的所有读出结果中的信号跨度和平均信号和最近的碱基调用阈值之间的距离的测量值。在重新测序和从头测序中，由于最小 Z - 评分使共有区准确度增加，而覆盖范围减少；随着 Z - 评分增加，大约一半的排除碱基属于长度 4 及以上的均聚物。Sanger 测序仪通常需要任何碱基覆盖深度为 3 或更多，以实现 99.99% 的共有区准确度。为了实现典型基因组的 95% 独特部分的最少三倍覆盖，需要大约 7 - 8 倍的过量取样。由于我们的较高错误率，我们观察到用 4 或更多的覆盖深度，在相似比例的基因组上实现了相似的共有区准确度，其中需要大约 10 - 12 倍过量取样。

### 生殖道支原体(580,069 bp)

按照前面的描述使支原体基因组 DNA 断裂，并且制备为测序文库（这是通过单个个体在 4 小时内完成的）。在乳状液 PCR 和珠沉积到  $60 \times 60 \text{ mm}^2$  光导纤维片上之后，在仪器的自动化 4 小时运行中使利用 1 个单独的 6 小时、4 个核苷酸的 42 个循环的过程从测序系统上流通。结果概括于表 29.2。为了测量各个读出结果的质量，我们用类似于以前在评估其它碱基调用仪的准确度时采用的流动 - 间隙作图和标准(Ewing, B., Hillier, L., Wendl, M.C., Green, P., Genome Research 8, 175 (1998)), 以 70 % 严格性，将每个高质量读出结果与参照基因组进行比对。当评估测序质量时，仅仅包括作图于参照基因组中独特位置的读出结果。由于该

过程排除重复区（基因组中相应流动图彼此 70% 相似的部分），选择的读出结果不完全覆盖基因组。图 46A 说明了该运行的读出结果长度分布。平均读出结果长度是 110 bp，导致的 40 倍过量取样和 84,011 个读出结果(27.4%)是完美的。图 46B 概括了作为碱基位置的函数的平均错误。非重复区的覆盖与不偏倚的样品制备和乳状液一致（图 54）。在各个读出结果水平，我们观察到插入和缺失错误率是大约 3.3%；取代错误具有低得多的比率，是 0.5% 的数量级。当使用这些没有任何 Z - 评分限制的读出结果时，我们以 99.97% 的共有区准确度在 10 个连续区域中覆盖了基因组的 99.94%。在共有序列中，均聚物中的错误率显著减少（图 53）。在该共有序列未覆盖的碱基（366 bp）中，所有都属于排除的重复区。将最小 Z - 评分设置为等于 4，覆盖减少到基因组的 98.1%，而共有区准确度增加到 99.996%。我们通过重复生殖道支原体完整基因组测序额外的 8 次，进一步证明了该系统的可重复性，在 8 次独立的仪器运行中的每一次运行中实现了基因组的 40 倍覆盖（表 27）。

我们将来自单次运行的生殖道支原体读出结果组装为 25 个重叠群，其平均长度为 22.4 kbp。由于 60 bp 的塌陷的串联重复区，这些重叠群中的一个是一个是错误组装的，并且人工校正。在用于完成序列的定向测序前，最初的生殖道支原体测序得到 28 个重叠群(Fraser, C. M. et al., Science 270, 397 (1995))。我们的组装覆盖 96.54% 的基因组，并且实现了 99.96% 的共有区准确度。不可分辨的重复区占基因组的 3%：因此，我们覆盖了基因组的 99.5% 独特区。重叠群之前的 16 个断裂是由于不可分辨的重复区，2 个是由于丢失的重叠读出结果（我们的读出结果过滤器和计时器不是完美的，并且我们用于进行流动图的模式匹配的算法偶尔丢失有效重叠），其余的是由于小的读出结果覆盖。将最小 Z - 值设定为 4，覆盖减少到基因组的 95.27%（基因组的不可分辨部分的 98.2%），共有区准确度增加到 99.994%。

## 讨论

我们在本文中证明了采用新开发的体外样品制备方法和测序技术，在仪器的单次运行中，以 96% 的平均准确度同时获得数十万个长度为 80 – 120 个碱基的序列读出结果。采用 phred 20 作为截止值，我们证明了我们的仪器能够从测试片段产生超过 4700 万个碱基，从基因组文库产

生 2500 万个碱基。我们用测试片段将我们的样品制备方法从我们的测序技术中解偶联。单次读出结果准确度从测试片段的 99.4% 减少到基因组文库的 96% 主要是由于在乳状液中的一部分基因组模板缺乏克隆性，并且不是测序技术的固有限制。由信号分布加宽导致的剩余的误差，特别是对于大均聚物（7 或更多），导致含糊的碱基调用。最近关于校正孔间干扰的测序化学和算法的研究表明，信号分布将变窄，伴随误差的减少和读出结果长度的增加。在用基因组文库进行的也包括乳状液方案的改进的初步实验中，我们能够用 84 个循环实现 200bp 的读出结果长度，准确度类似于本文对 100 bp 证明的准确度。有时，用 168 个循环，我们建立了各个读出结果，其在超过 400bp 的范围是 100 % 准确的。

采用生殖道支原体，我们证明了短片段理论上不阻止细菌基因组的从头组装。实际上，由我们的系统的通量提供的更大过量取样在显著减少工作量的条件下得到了具有比 Sanger 读出结果更少的重叠群的选择序列。通过利用过量取样的优点，对于该基因组实现了大于 99.96% 的共有区准确度。通过对该组装进行进一步的质量过滤，可以以超过 99.99 % 的准确度选择共有序列，而仅仅导致基因组覆盖的微小损失。当我们对肺炎链球菌的 2.1 Mbp 基因组(Tettelin, H. et al., Science 293, 498 (2001); 表 28)进行鸟枪测序和从头组装时，观察到了相似结果。从头组装比细菌更复杂的基因组，包括哺乳动物基因组，可能需要与为 Sanger 测序开发的方法类似的方法的开发，用于对可以跨这些基因组中的重复区的成对末端文库进行制备和测序。为了促进成对末端文库的利用，我们开发了在单个孔中从基因组模板的两个末端开始测序的方法，并且计划在我们的组装程序中加入成对末端读出结果能力（补充方法：双末端测序）。

通量的进一步增加和每个碱基的花费的一致减少，可能是来自光导纤维反应器的持续小型化，使每单位面积能够产生更多序列 - 类似于在开始开发时能够预测集成电路中显著改进的缩放比例特征(Moore, G. E., Electronics 38, Number 8, April 19, (1965))。

## 方法

基于乳状液的克隆扩增。通过在通过建立油包 PCR 反应混合物乳状液而制备的约  $100\mu\text{m}$  含水液滴中分离各个携带 DNA 的珠子 ( $2 \times 10^6/\text{mL}$ )

个数量级），实现片段的同时扩增。（图 43B 和补充方法：制备 DNA 捕获珠、将模板种类与 DNA 捕获珠结合、PCR 反应混合物制备和制剂化、乳化和扩增）。液滴作为分开的微反应器，其中进行平行的 DNA 扩增，产生每个珠子大约个  $10^7$  拷贝的模板；在标准的 2 mL 试管中制备含有 1500 万个珠子的 800  $\mu\text{L}$  乳状液。将每种乳状液等分到 8 个 PCR 管中用于扩增。PCR 后，使乳状液破裂以释放珠子，其中包括具有扩增的、固定的 DNA 模板的珠子和空珠子（补充方法：使乳状液破裂和回收珠子）。我们然后富集携带模板的珠子（补充方法：珠子的富集）。典型地，大约 30 % 的珠子将具有 DNA，每个乳状液反应产生 450,000 个携带模板的珠子。制备的乳状液的数目取决于基因组的大小和实现足够的过量取样所需的预计运行次数。在一个  $60 \times 60 \text{ mm}^2$  光导纤维片上测序的 580 kbp 生殖道支原体基因组需要 1.6 mL 乳状液。过量取样 10 倍的人基因组需要大约 3000 mL 乳状液。

珠子加样到微微升孔中。通过离心到开放的孔中沉积富集的携带模板的珠子（图 43C），其沿着  $60 \times 60 \text{ mm}^2$  光导纤维片的一面排列。确定珠子的大小（直径约  $28 \mu\text{m}$ ），以确保大多数孔中具有不超过 1 个珠子（我们观察到 2 - 5% 填充的孔含有超过 1 个珠子）。实验发现，将 450,000 个珠子（来自一个乳状液制备物）加入  $60 \times 60 \text{ mm}^2$  板的每一半，将珠子限制为占据所有孔的大约 35%，从而减少了孔之间的化学和光学干扰。也将从游离焦磷酸产生光所需的携带固定的 ATM 硫酸化酶和荧光素酶的较小珠子的混合物加入孔中，以产生各个测序反应器（补充方法：珠沉积、制备酶珠和微粒过滤器）。

图像捕获。对于每个掺入的核苷酸，携带 1 千万个拷贝的模板的珠子在 CCD 传感器上产生大约 10,000 个光子。产生的光通过光导纤维片的底传播，并且通过大型 CCD（4095x4096 像素）检测。处理图像，同时产生所有携带珠 - 模板的孔的序列信息。成像系统设计适于大量小孔和在每个核苷酸流动期间从各个孔产生的大量光学信号。一旦封固，光导纤维片的位置不会变动；这使得图像分析软件能够基于每个测序运行前焦磷酸溶液流动期间产生的光确定每个孔的位置（无论它是否含有携带 DNA 的珠子）。通过大约 9 个  $15 \mu\text{m}$  的像素，对单个孔成像。对于每次核苷酸流动，通过覆盖特定孔的像素收集的光强度进行总计，产生特定核苷酸流动时针对该特定孔的信号。由 CCD 捕获的每个图像产生 32

兆字节数据。为了实时进行所有必须的处理信号，给控制计算机安装辅助板（补充方法：现场可编程门阵列），其带有 6 百万门 FPGA (Mehta, K., Rajesh, V. A., Veeraswamy, S., Biomed Sci Instrum. 29, 507 (1993); Fagin, B., Watt, J. G., Gross, R., Comput Appl Biosci. 9, 221 (1996)。

从头鸟枪序列组装程序。开发了从头流动 - 间隙组装程序，用于捕获最初的基于流动的信号轨迹中含有的信息。其也解决了现有的组装程序对于 80 - 120 bp 读出结果并非最佳的事实，特别是针对由于达到等同基因组覆盖所需的测序读出结果的数目增加导致的内存控制（覆盖 100 bp 读出结果的完全随机基因组需要多 50 % 的读出结果来产生与用 700 bp 读出结果实现的数目相同数目的连续区（重叠群），假定读出结果之间需要 30 bp 重叠）(Lander, E. S., Waterman, M. S., Genomics 2, 231 (1988))。该组装程序由一系列模块组成：Overlapper，其发现和产生读出结果之间的重叠，Unitigger，其构建重叠序列读出结果的更大重叠群，和 Multialigner，其产生每个重叠群内碱基的共有调用和质量评分（补充方法：从头序列组装程序）。（软件模块的名称是基于 Myers 开发的其它组装程序中行使相关功能的软件模块的名称(Myers, E. W., J Comput Biol. 2, 275 1(1995); Hamilton, S.C., J.W. Farchaus and M.C. Davis. 2001. BioTechniques 31:370)。

表 23: 测试片段的测序统计的概括

光导纤维片的大小	60x60 mm <sup>2</sup> Sy
运行时间/循环数	243 分钟/42
测试片段读出结果	497,893
平均读出结果长度	108
测试片段中的碱基数目	53,705,267
Phred 20 和更高的碱基	47,181,792
单独读出结果插入错误率	0.44%
单独读出结果缺失错误率	0.15%
单独读出结果取代错误率	0.004%
所有错误	0.60%

表 24: 生殖道支原体的概括统计

测序概括	生殖道支原体 1
仪器运行次数	1
光导纤维片的大小	60x60 mm <sup>2</sup>
运行时间/循环数	243 min/42
高质量读出结果	306,178
平均读出结果长度	110 b
高质量读出结果中的碱基数目	33,655,553
Phred 20 和更高的碱基	26,753,540
重新测序	
作图到单个位点的读出结果	238,066
作图的读出结果中的碱基数目	27,687,747
单独读出结果插入错误率	1.67%
单独读出结果缺失错误率	1.60%
单独读出结果取代错误率	0.68%
重新测序共有区	
平均过量取样	40x
覆盖, 所有( $Z \geq 4$ )	99.9% (98.2%)
共有区准确度, 所有( $Z \geq 4$ )	99.97% (99.996%)
共有区插入错误率, 所有( $Z \geq 4$ )	0.02% (0.003%)
共有区缺失错误率, 所有( $Z \geq 4$ )	0.01% (0.002%)
共有区取代错误率, 所有( $Z \geq 4$ )	0.001% (0.0003%)
重叠群数目	10
从头组装	
覆盖, 所有( $Z \geq 4$ )	96.54% (95.27%)
共有区准确度, 所有( $Z \geq 4$ )	99.96% (99.994%)
重叠群数目	25
平均重叠群大小	22.4 kb

单独读出结果错误率是指作图的读出结果中的碱基总数。

## 实施例 30 补充材料和方法

### 文库制备 (图 47)

DNA 断裂。从范围是细菌菌落到获自销售商的冻干样品的不同来源获得基因组 DNA 样品。收到后，用 1.8 - 2.0 的 OD<sub>260</sub>/OD<sub>280</sub> 比，证实浓度(>300 μg/mL)。将 15 微克基因组 DNA 在 2.0 mL 试管中，在 1X TE 缓冲液(10 mM Tris, 1 mM EDTA, pH 7.6)中稀释到 100 μL 的终体积。通过加入 1.6 mL 冰冷的喷雾缓冲液(53.1% 甘油, 37 mM Tris - HCl, 5.5 mM EDTA, pH 7.5)并且通过重复的相反移液作用轻柔混合，进一步稀释样品。

用 Aeromist 喷雾器(Alliance Medical, Russleville, MO)使 DNA 溶液断裂，该喷雾器按照下文的描述在向实验室外通风的 PCR 罩(Labconco, Kansas City, MO, USA)中改进。简言之，将来自具有保护帽的 15 ml falcon 管的帽置于喷雾器顶端上。为了减少喷雾期间样品雾化导致的损失，将 0.50" OD x 0.31" ID x 1.5" 长的硅酮管段组成的喷雾器冷凝管固定在存在的喷雾器加样管中。将 DNA 样品混合物转移到喷雾器室的底部，将喷雾器的顶部紧密丝扣连接在室上。设计松散固定、客户定制的迭尔林帽，用于覆盖喷雾器的顶部，并且在喷雾器的外面提供侧沟，用于固定一对 34 号丁纳橡胶 - N O 环，它们用于固定帽。然后，将整个喷雾器组合件紧密包裹在石蜡膜(American Nat'l Can, Menasha, WI)中。然后将喷雾器连接于具有供应管的氮气罐，将管连接件包裹在石蜡膜中。

将组装的喷雾器垂直放置在冰桶上，该单位的下半部浸入冰中。以 50 psi 供应氮气 5 分钟；喷雾器壁上的冷凝物通过偶然轻扣，敲击到室的底部。关上气体，使压力标准化 30 秒，然后从喷雾器取下管子。小心拆开喷雾器，将样品转移到 1.5 mL 的微量离心管中。回收的体积一般超过 900 μL。

根据制造商的说明书，通过 Qiaquick PCR 纯化柱(Qiagen, Valencia, CA)离心，纯化雾化的 DNA。由于大体积，以一些等分物在相同的柱上进行 DNA 加样和纯化。用 30 μL 55 °C 缓冲液 EB (在 Qiagen 试剂盒中提供)洗脱纯化的 DNA。通过在使用 DNA 1000 LabChip 的 Agilent 2100 BioAnalyzer (Agilent, Palo Alto, CA.) 上分辨雾化的材料的 2 μL 等分物，确定雾化的片段的大小分布 (代表性的轨迹参见图 48)。回收的材

料表现出 50 – 900 bp 的大小范围，平均片段大小是  $325 \pm 50$  bp。

酶促磨光。DNA 的雾化产生磨损的末端占优势的片段(Pan, H. et al., GATA 11, 181 (1994); Bankier, A. T., Weston, K. M. and Barrell, B. G., Meth. Enzymol. 155, 51 (1987) )。使这些末端成为平端，并且通过三种酶，即 T4 DNA 聚合酶、大肠杆菌 DNA 聚合酶(克列诺片段)(New England Biolabs, Beverly, MA)和 T4 多核苷酸激酶(New England Biolabs)的活性进行磷酸化。

在 0.2 mL 试管中，将剩余的 28  $\mu$ L 纯化的、雾化的 DNA 片段与 5  $\mu$ L 分子生物级水 (Eppendorf, Hamburg, Germany), 5  $\mu$ L 10X NEBuffer 2 (New England Biolabs), 5  $\mu$ L 1 mg/mL BSA (New England Biolabs), 2  $\mu$ L 10mM dNTPs (Pierce, Rockford, IL.), 和 5  $\mu$ L 3u/ $\mu$ L T4 DNA 聚合酶 (New England Biolabs)合并。将磨光反应充分混合，25°C 下在热循环仪 (MJ Research, Waltham, MA) 中温育 10 分钟。温育后，加入 1.25  $\mu$ L 5u/ $\mu$ L 大肠杆菌 DNA 聚合酶(克列诺片段)(New England Biolabs)，将反应充分混合，25°C 下额外温育 10 分钟，然后 16°C 下温育 2 小时。

然后在 Qiaquick PCR 柱上纯化磨光反应，用 30  $\mu$ L 55°C 缓冲液 EB 洗脱，并且转移到 0.2 mL 试管中进行磷酸化。通过加入 5  $\mu$ L 分子生物级水, 5  $\mu$ L 10X T4 PNK 缓冲液(New England Biolabs), 5  $\mu$ L 10mM ATP (Pierce) 和 5  $\mu$ L 10u/ $\mu$ L T4 PNK (New England Biolabs)，将 DNA 稀释到 50  $\mu$ L。混合反应，37°C 下温育 30 分钟，然后 65°C 下温育 20 分钟。然后如前面所述，在 Qiaquick PCR 纯化柱上纯化磷酸化的片段，在 30  $\mu$ L 55°C 缓冲液 EB 中洗脱。通过采用 Turner TBS - 380 微型荧光计(Turner Biosystems, Sunnyvale, CA)的荧光测量法，在 2  $\mu$ L 等分物中进行 DNA 浓度定量。

基因组 DNA 文库的断裂和磨光后，将引物序列添加到 DNA 片段的每个末端。44 个碱基的引物序列（下文称作“衔接子”）是双链寡核苷酸，包含一个 5' 20 碱基 PCR 扩增引物，其后是 20 个碱基的测序引物，和 3', 4 碱基、非回文测序“关键”序列，其包含每种脱氧核糖核苷酸（例如 AGTC）之一。每个反应中使用两类衔接子，称作“衔接子 A”和“衔接子 B”。A 和 B 衔接子的核苷酸序列和 B 衔接子上 5' 生物素标记的存在不同。设计衔接子对，使得能够方向性连接于平端的、断裂的基因组 DNA ( 衔接子 A: CCATCTCATCCCTGCGTGTCCCATCTGT

TCCCTCCCTGTCTCAG SEQ ID NO:61。衔接子 B:/5BioTEG/CCTAT CCCCTGTGTGCCTTGCCTATCCCCTGTTGCGTGTCTCAG SEQ ID NO:62）。对于每个衔接子对，PCR 引物区包含 5'四碱基突出端和平端的 3'关键区。当衔接子的 3'平端侧与平端基因组 DNA 片段连接，而 5'突出端防止与衔接子的 PCR 引物区连接时，实现方向性。

将剩余的 28  $\mu$ L 雾化的、磨光的 DNA 转移到 0.2 mL 试管，并且与 20.6  $\mu$ L 分子生物级水, 60  $\mu$ L 2X Quick 连接酶反应缓冲液(New England Biolabs), 1.8  $\mu$ L 衔接子 A 和 B 的等摩尔混合物(每种衔接子 200 pmol/ $\mu$ L), 9.6  $\mu$ L 2000 U/ $\mu$ L Quick 连接酶(New England Biolabs)合并。将试管的内容物彻底混合，25°C 下温育 20 分钟，在 Qiaquick PCR 纯化柱上纯化两次，并且在每次离心后在 30  $\mu$ L 55°C 的缓冲液 EB 中洗脱。

凝胶纯化。将 4.5  $\mu$ L 10mg/mL 的溴化乙锭储存液(Fisher Scientific, Pittsburgh, PA)加入熔化的琼脂糖溶液，制备 2% 琼脂糖(Invitrogen, Carlsbad, CA) /TBE 厚片凝胶。将 3 微升 10X Ready - Load 染料(Invitrogen)加入 30  $\mu$ L 连接的 DNA 文库，将染料/连接反应加入凝胶的两个相邻孔中(大约每个泳道 16.5  $\mu$ L)。将 10 微升(1 $\mu$ g)100 - bp 序列梯(Invitrogen)加入文库样品任一侧的侧翼孔中，用两个空泳道分开文库和序列梯样品。100V 下凝胶电泳 3 小时，此后将凝胶转移到 GelDoc (BioRad, Hercules, CA) UV 盒中，所述盒用塑料套遮盖，以减少污染的机会。用无菌的一次性刮刀切下 DNA 序列梯中 250 和 500 碱基对标志物之间迁移的每个文库样品的区域，然后将凝胶切片放置在 15 mL Falcon 试管中。用来自 MinElute 凝胶提取试剂盒(Qiagen)的 2 个柱从每个凝胶塞提取文库，每个样品一个柱。根据制造商的说明书进行该过程，采用以下改变。由于溶解的琼脂糖的大体积，使每个文库破裂成几个等分物，并且连续通过各自的柱。也将缓冲液 PE 旋转后干燥旋转的持续时间延长到 2 分钟(而不是 1 分钟)，以确保完全除去乙醇，并且合并来自每个柱的洗脱液，以获得 20  $\mu$ L 的最终文库体积。在 BioAnalyzer DNA 1000 LabChip 上分析 1 微升的分离文库，以证实文库群体的大小分布是 250 – 500 bp。

切口修复。通过 Bst DNA 聚合酶大片段的链置换活性，修复位于 3' - 接合处的两个切口。将剩余的 19  $\mu$ L 大小分级文库与 40  $\mu$ L 分子生物级水, 8  $\mu$ L 10X ThermoPol 反应缓冲液(New England Biolabs), 8  $\mu$ L 1mg/mL BSA (New England Biolabs), 2  $\mu$ L 10 mM dNTPs (Pierce), 和 3

$\mu\text{L}$  8U/ $\mu\text{L}$  Bst DNA 聚合酶大片段(New England Biolabs)合并, 65°C下温育 30 分钟。

分离单链 AB 衔接的文库。通过以下步骤, 在含有 200  $\mu\text{L}$  1X B&W 缓冲液(5 mM Tris - HCl (pH 7.5), 0.5 mM EDTA, 1 M NaCl)的 1.5mL 微量离心管中将 100 微升储存 M - 270 链霉亲和素珠(Dynal, Oslo, Norway)洗涤两次: 在洗涤溶液中涡旋珠子, 用磁性颗粒浓集器(MPC) (Dynal) 固定珠子, 将溶液从固定的珠子吸出, 并且重复。第二次洗涤后, 将珠子重悬于 100  $\mu\text{L}$  2X B&W 缓冲液(10 mM Tris - HCl (pH 7.5), 1 mM EDTA, 2 M NaCl)中, 随后在其中加入完整的 80  $\mu\text{L}$  Bst 聚合酶处理的文库和 20  $\mu\text{L}$  分子生物级水。然后通过涡旋混合样品, 并且室温下放置在水平试管转子上 20 分钟。然后用 200  $\mu\text{L}$  1X B&W 缓冲液洗涤珠混合物两次, 然后用 200  $\mu\text{L}$  分子生物级水洗涤两次。

用 MPC 从珠堆中去除最终的涤液体, 加入 250  $\mu\text{L}$  解链溶液(100 mM NaCl 和 125 mM NaOH)。通过在解链溶液中充分混合, 使珠子重悬, 将珠悬浮液室温下在管转子中室温下温育 10 分钟。

在一个单独的 1.5mL 离心管中, 通过加入 9  $\mu\text{L}$  20% 的含水乙酸, 使 1250  $\mu\text{L}$  缓冲液 PB (来自 QiaQuick PCR 纯化试剂盒) 雾化。采用 Dynal MPC, 使解链溶液中的珠子沉淀; 小心倒去 250  $\mu\text{L}$  上清液 (含有目前是单链的文库), 并且转移到新制备的雾化缓冲液 PB 的试管中。

在来自 MinElute PCR 纯化试剂盒(Qiagen)的单个柱中浓缩 1500  $\mu\text{L}$  中和的、单链文库, 温热到室温备用。由于体积限制, 以两个 750  $\mu\text{L}$  等分物加入和浓缩样品。根据制造商关于使用微量离心管的旋转柱的说明书, 进行每个等分物的浓缩, 其中具有以下改进: 缓冲液 PE 旋转后的干燥旋转延长到 2 分钟 (而不是 1 分钟), 以确保完全除去乙醇, 并且 55°C 下在 15  $\mu\text{L}$  缓冲液 EB (Qiagen) 中洗脱单链文库样品。

文库定量和质量评估。用 Agilent 2100 和荧光板读数器评估得到的单链 DNA 文库的数量和质量。由于文库由单链 DNA 组成, 根据制造商的指导, 使用和制备了用于 Agilent 2100 的 RNA Pico 6000 LabChip。分析了三份 1  $\mu\text{L}$  等分物, 将 Agilent 分析软件报道的平均值用于评估 DNA 浓度。最终文库浓度典型地超过  $10^8$  个分子/ $\mu\text{L}$ 。在 -20°C 下储存浓缩的文库样品, 直到需要。

## 制备 DNA 捕获珠

从 1 mL-N-羟基琥珀酰亚胺酯 (NHS) 活化的 Sepharose HP 亲和柱 (Amersham Biosciences, Piscataway, NJ) 上取出塞满的珠子，按照产品说明 (Amersham Pharmacia Protocol # 71700600AP) 活化。使 20 mM 磷酸缓冲液 pH 8.0 中的 25 微升 1 mM 胺标记的 HEG 捕获引物 (5' - 胺 - 3' 连续 18 - 原子六甘醇间隔基 CCATCTGTTGCGTGCCTGTC - 3' SEQ ID NO:63) (IDT Technologies, Coralville, IA, USA) 与珠子结合。此后，通过连续通过 36 和 25 μm 孔滤网片 (Sefar America, Depew, NY, USA)，选择 25 - 36 μm 的珠子。通过第一个滤网但保留在第二个滤网上的 DNA 捕获珠收集在珠储存缓冲液 (50 mM Tris, 0.02% Tween, 0.02% 叠氮化钠, pH 8) 中，用 Multisizer 3 Coulter 计数器 (Beckman Coulter, Fullerton, CA, USA) 定量，4°C 下储存直到需要。

## 将模板种类与 DNA 捕获珠结合

在 UV 处理的层流罩中使模板分子与 DNA 捕获珠上的互补引物退火。将 150 万个悬浮于珠储存缓冲液中的 DNA 捕获珠转移到 200 μl PCR 管中，在台式微型离心机中离心 10 秒，将管旋转 180 度，再离心 10 秒以确保均匀沉淀形成。除去上清液，用 200 μl 退火缓冲液 (20 mM Tris, pH 7.5 和 5 mM 醋酸镁) 洗涤珠子，涡旋 5 秒，以使珠子重悬，按照上文沉淀。除去珠子上的除大约 10 μl 以外的全部上清液，加入额外的 200 μl 退火缓冲液。将珠子再次涡旋 5 秒，静置 1 分钟，然后按照上文沉淀。弃去除 10 μl 以外的全部上清液，在珠子中加入 1.2 μl  $2 \times 10^7$  个分子/μl 的模板文库。将管涡旋 5 秒以混合内容物，其后使模板在 MJ 热循环仪中进行的受控变性/退火程序中与珠子退火 (80°C 下温育 5 分钟，然后以 0.1°C/秒 降低到 70°C，70°C 下温育 1 分钟，以 0.1°C/秒 降低到 60°C，60°C 下保持 1 分钟，以 0.1°C/秒 降低到 50°C，50°C 下保持 1 分钟，以 0.1°C/秒 降低到 20°C，20°C 下保持)。完成退火程序后，珠子在冰上储存直到需要。

## PCR 反应混合物制备和制剂化

为了减少污染的可能性，在 PCR 洁净室内的 UV 处理的层流罩中进行 PCR 反应混合物制备。对于每个 1,500,000 个珠的乳状液 PCR 反应，

将以下试剂混合在 1.5 ml 试管中：225  $\mu$ l 反应混合物(1X Platinum HiFi 缓冲液(Invitrogen)), 1 mM dNTPs, 2.5 mM MgSO<sub>4</sub> (Invitrogen), 0.1% 乙酰化的、分子生物级 BSA(Sigma, St. Louis, MO), 0.01% Tween - 80 (Acros Organics, Morris Plains, NJ), 0.003 U/ $\mu$ l 耐热焦磷酸酶(NEB), 0.625  $\mu$ M 正向引物(5' - CGTTTCCCCCTGTGTGCCTTG - 3' SEQ ID NO:64)和 0.039  $\mu$ M 反向引物(5' - CCATCTGTTGCG TGC GTGTC - 3' SEQ ID NO:65) (IDT Technologies)以及 0.15 U/ $\mu$ l Platinum Hi - Fi Taq 聚合酶 Invitrogen)。取出 25 微升反应混合物，储存在单独的 200  $\mu$ l PCR 管中，用作阴性对照。反应混合物和阴性对照都储存在冰上备用。此外，在 1.5 mL 试管中制备针对每种乳状液的 240  $\mu$ L 假扩增混合物(1X Platinum HiFi 缓冲液(Invitrogen), 2.5 mM MgSO<sub>4</sub> (Invitrogen), 0.1% BSA, 0.01% Tween)，并且类似地在室温下储存备用。

### 乳化和扩增

乳化过程产生每微升具有大约 1,000 个离散的 PCR 微反应室的耐热油包水乳状液，其作为靶文库的各个分子的单分子克隆扩增的基质。按照以下方式乳化用于单个反应的反应混合物和 DNA 捕获珠：在 UV 处理的层流罩中，将 160  $\mu$ l PCR 溶液加入含有 1,500,000 个 DNA 捕获珠的管中。通过反复吸取使珠子重悬，此后，在室温下将 PCR - 珠混合物静置至少 2 分钟，使得珠子在 PCR 溶液中平衡。此时，将 400  $\mu$ l 乳状液油 (60 % (w/w) DC 5225C 制剂辅料(Dow Chemical Co., Midland, MI), 30% (w/w) DC 749 流体 (Dow Chemical Co.), 和 30% (w/w) Ar20 Silicone Oil (Sigma))等分在平顶 2 ml 离心管(Dot Scientific, Burton, MI) 中。然后将 240  $\mu$ L 假扩增混合物加入 400  $\mu$ L 乳状液油，给管安全加盖，并且放置在 TissueLyser MM300 (Retsch GmbH & Co. KG, Haan, Germany) 的 24 孔 TissueLyser 适配器(Qiagen)中。以 25 次振动/秒，将乳状液匀浆 5 分钟，以产生非常小的乳状液，或“微粒 (microfines)”，其给反应赋予额外的稳定性。

短暂涡旋合并的珠子和 PCR 反应混合物，使其平衡 2 分钟。微粒形成后，将扩增混合物、模板和 DNA 捕获珠加入乳化的材料。使 TissueLyser 速度减少到 15 次振动/秒，将反应混合物匀浆 5 分钟。较低的匀浆速度在油混合物中产生了水滴，平均直径是 100 – 150  $\mu$ m，其对于容纳 DNA

捕获珠和扩增混合物是足够大的。

乳状液的总体积是在一个 2 mL 平顶离心管中含大约 800 $\mu$ L。将乳状液等分到 7 – 8 个分开的 PCR 管中，每个管包含大约 100  $\mu$ l 乳状液。将管密封，与上文制备的 25 $\mu$ l 阴性对照一起放置在 MJ 热循环仪中。采用以下循环时间：1X(4 分钟 @ 94°C) – 热启动, 40X (30 秒 @ 94°C, 60 秒 @ 58°C, 90 秒 @ 68°C) – 扩增, 13X (30 秒 @ 94°C, 58°C 下 360 秒) 杂交延伸。完成 PCR 程序后，取出反应物，立即使乳状液破裂（如下文所述）或在起始破裂过程前将反应物在 10°C 下储存最多 16 小时。

#### 使乳状液破裂和回收珠子

在每个含有扩增材料的乳状液的 PCR 管中加入 50 微升异丙醇 (Fisher)，涡旋 10 秒，以降低乳状液的速度。在微量离心机中将管离心几秒，以除去任何捕集在管盖中的乳化的材料。将每个管中的乳状液 – 异丙醇混合物吸入安装了 16 号钝针头(Brico Medical Supplies, Metuchen, NJ)的 10 mL BD – 一次性注射器(Fisher Scientific)中。将额外的 50  $\mu$ L 异丙醇加入每个 PCR 管，涡旋，按照前面所述离心，加入注射器的内容物中。用异丙醇使注射器内的体积增加到 9 mL，此后倒置注射器，将 1 mL 空气吸入注射器，以促进异丙醇和乳状液的混合。取下钝针头，将含有 15 $\mu$ m 孔 Nitex 筛选结构 (Sefar America, Depew, NY, USA) 的 25 mm Swinlock 滤器支架 (Whatman, Middlesex, United Kingdom) 连接于注射器腔，将钝针头固定于 Swinlock 单元的对侧。

通过 Swinlock 滤器单位和针头将注射器的内容物轻柔但完全地排出到具有漂白作用的废料容器中。通过钝针头和 Swinlock 滤器单位将 6 微升新的异丙醇收回注射器，将注射器倒置 10 次，以混合异丙醇、珠和剩余的乳状液成分。将注射器的内容物再次排出到废料容器中，每次洗涤用 6 mL 额外的异丙醇，将洗涤过程重复 2 次。用 6 mL 80% 乙醇 / 1X 退火缓冲液(80% 乙醇, 20 mM Tris - HCl, pH 7.6, 5 mM 醋酸镁)重复洗涤步骤。然后用含有 0.1% Tween 的 6 mL 1X 退火缓冲液(0.1% Tween - 20, 20 mM Tris - HCl, pH 7.6, 5 mM 醋酸镁)洗涤珠子，然后用 6 mL picopure 水洗涤。

将最终的洗液排出到废料容器中，将 1.5 mL 1 mM EDTA 抽入注射器，取下 Swinlock 滤器单位，并且静置。将注射器的内容物连续转移到

1.5 mL 的离心管中。在微型离心机中周期性将管离心 20 秒，以沉淀珠子，并且除去上清液，此后，将注射器的剩余内容物加入离心管。将 Swinlock 单元重新连接到滤器，将 1.5 mL EDTA 抽入注射器。最后一次除去 Swinlock 滤器，将珠子和 EDTA 加入离心管中，沉淀珠子，并且在必要时除去上清液。

## 第二条链的去除

通过在基本解链溶液中温育，除去第二条链，从而使固定在捕获珠上的扩增的 DNA 成为单链。将 1mL 新制备的解链溶液(0.125 M NaOH, 0.2 M NaCl)加入珠子，通过在中间设置下涡旋 2 秒，使沉淀重悬，并且将管置于 Thermolyne LabQuake 管旋转器中 3 分钟。然后按照上文使珠子沉淀，小心除去并弃去上清液。然后通过加入 1 mL 退火缓冲液(20 mM Tris - Acetate, pH 7.6, 5 mM 醋酸镁)稀释残留的解链溶液，此后在中间速度将珠子涡旋 2 秒，沉淀珠子，按照上文除去上清液。重复退火缓冲液洗涤，不同之处是离心后仅仅除去 800  $\mu$ L 退火缓冲液。将珠子和剩余的退火缓冲液转移到 0.2 mL PCR 管，立即使用，或者在继续进行随后的富集过程前在 4°C 下储存最多 48 小时。

## 珠子的富集

此时，珠质量包括具有扩增的、固定的 DNA 链的珠子，和不含扩增的产物的无效珠。利用富集过程选择性捕获具有可测序量的模板 DNA 的珠子，而排除无效珠。

通过在台式微型离心机中离心 10 秒，沉淀来自前面步骤的单链珠，此后，将管旋转 180 度，再离心 10 秒，以确保均匀的沉淀形成。然后，除去尽可能多的上清液而不扰动珠子。将 15 微升退火缓冲液加入珠子，然后加入 2  $\mu$ L 100  $\mu$ M 生物素化的、40 个碱基的 HEG 富集引物(5 生物素 18 - 碳六甘醇间隔基 - CGTTTCCCCCTGTGTGCCTTGCCATCTGTT CCCTCCCTGTC - 3'，IDT Technologies SEQ ID NO:66)，该引物与珠固定的模板 3' 末端上的合并的扩增和测序位点（长度均 20 个碱基）互补。通过在中间设置下涡旋 2 秒，混合溶液，在 MJ 热循环仪中用受控的变性/退火程序使富集引物退火于固定的 DNA 链。(30 秒@ 65°C，以 0.1 °C /秒降低到 58°C, 90 秒@ 58°C, 和 10°C 下保持)

当引物退火时，通过轻柔涡旋，使 SeraMag - 30 磁性链霉亲和素珠 (Seradyn, Indianapolis, IN, USA) 的储存液重悬，并且，将 20  $\mu$ l SeraMag 珠加入含有 1 ml 富集液(2 M NaCl, 10 mM Tris - HCl, 1 mM EDTA, pH 7.5)的 1.5 ml 微量离心管中。将 SeraMag 珠混合物涡旋 5 秒，将管置于 Dynal MPC - S 磁铁中，顺磁珠沿微量离心管壁沉淀。小心除去上清液，并且弃去，而不扰动 SeraMag 珠。从磁铁上取下管，加入 100  $\mu$ l 增强液。将管涡旋 3 秒，使珠子重悬，并且在冰上储存直到使用。

完成退火程序后，将 100  $\mu$ l 退火缓冲液加入含有 DNA 捕获珠和富集引物的 PCR 管中。将管涡旋 5 秒，将内容物转移到新的 1.5 ml 微量离心管中。用 200  $\mu$ l 退火缓冲液洗涤 PCR 管（其中富集引物与捕获珠退火）一次，将洗涤溶液加入 1.5 ml 管中。用 1 ml 退火缓冲液洗涤珠子 3 次，涡旋 2 秒，按照上文沉淀，小心除去上清液。第三次洗涤后，用 1 ml 冰冷的增强液洗涤珠子 2 次，涡旋，沉淀，按照上文的描述除去上清液。将珠子重悬于 150  $\mu$ l 冰冷的增强液中，将珠溶液加入洗涤的 SeraMag 珠中。

将珠混合物涡旋 3 秒，在 LabQuake 管旋转器上室温下温育 3 分钟，而链霉亲和素包被的 SeraMag 珠结合于与 DNA 捕获珠上的固定的模板退火的生物素化的富集引物。然后以 2,000 RPM 将珠子离心 3 分钟，此后轻柔“敲打”珠子，直到珠子重悬。然后将重悬的珠子放置在冰上 5 分钟。在冰上温育后，将冷的增强液加入珠子，达到 1.5 mL 的终体积。将管插入 Dynal MPC - S 磁铁。使珠子保持不扰动 120 秒，使沿着磁铁沉淀。此后，小心取出（含有过量 SeraMag 和无效 DNA 捕获珠）的上清液，并弃去。

将管从 MPC - S 磁铁上取下，在珠子中加入 1 ml 冷的增强液，通过用手指轻柔敲打，使珠子重悬。重要的是此时不涡旋珠子，因为涡旋将破坏 SeraMag 和 DNA 捕获珠之间的连接。珠子返回到磁铁，除去上清液。再重复洗涤 3 次，以确保除去所有无效捕获珠。为了从 DNA 捕获珠除去退火的富集引物和 SeraMag 珠，将珠子重悬于 1 mL 解链溶液中，涡旋 5 秒，用磁铁沉淀。将具有富集的珠子的上清液转移到分开的 1.5 ml 微量离心管中，沉淀珠子，除去上清液。然后将富集的珠子重悬于含有 0.1% Tween - 20 的 1 X 退火缓冲液中。在 MPC 上再次沉淀珠子，将上清液转移到新的 1.5 mL 管中，确保最大程度去除剩余的 SeraMag

珠。离心珠子，此后除去上清液，用 1 mL 1X 退火缓冲液洗涤珠子 3 次。第三次洗涤后，除去 800  $\mu$ L 上清液，将剩余的珠子和溶液转移到 0.2 mL PCR 管中。富集过程的平均产率是最初加入乳状液的珠子的 30%，或每个乳化反应大约 450,000 个富集的珠子。由于  $60 \times 60 \text{ mm}^2$  的片需要 900,000 个富集的珠子，按照上文处理两份 1,500,000 个珠子的乳状液。

### 测序引物退火

以 2,000 RPM 将富集的珠子洗涤 3 分钟，倒去上清液，随后，加入 15  $\mu$ l 退火缓冲液和 3  $\mu$ l 100  $\mu$ M 测序引物 (5' - CCATCTGTTCCCTCCCTGTC - 3', IDT Technologies SEQ ID NO:67)。然后将管涡旋 5 秒，置于 MJ 热循环仪中，进行以下 4 阶段退火程序：在 65°C 下 5 分钟，以 0.1°C/秒降低到 50°C，50°C 下 1 分钟，以 0.1°C/秒降低到 40°C，40°C 下保持 1 分钟，以 0.1°C/秒降低到 15°C，15°C 下保持。

完成退火程序后，从热循环仪取出珠子，通过离心 10 秒进行沉淀。将管旋转 180 度，再离心 10 秒。倒出上清液并且弃去，加入 200  $\mu$ l 退火缓冲液。用 5 秒涡旋重悬珠子，按照上文沉淀。除去上清液，将珠子重悬于 100  $\mu$ l 退火缓冲液中。此时，用 Multisizer 3 Coulter 计数器对珠子进行定量。4°C 下储存珠子，并且在至少 1 周内稳定。

### 用 Bst DNA 聚合酶大片段和 SSB 蛋白温育 DNA 珠

通过在含有 0.1% BSA 的 1x 测定缓冲液中加入腺苷三磷酸双磷酸酶 (Biotage, Uppsala Sweden) (最终活性 8.5 单位/升)，制备珠洗涤缓冲液 (100 ml)。从 picopure 水中取出光纤片，在珠洗涤缓冲液中温育。离心 90 万个以前制备的 DNA 珠，小心除去上清液。然后将珠子在含有 0.4 mg/mL 聚乙烯吡咯烷酮(MW 360,000), 1 mM DTT, 175  $\mu$ g 大肠杆菌单链结合蛋白(SSB) (United States Biochemicals Cleveland, OH) 和 7000 单位 Bst DNA 聚合酶大片段(New England Biolabs) 的 1290  $\mu$ l 珠洗涤缓冲液中温育。室温下在转子上将珠子温育 30 分钟。

### 制备酶珠和微粒过滤器

室内将 UltraGlow 荧光素酶(Promega Madison WI)和 Bst ATP 硫酸化酶制备为生物素羧基载体蛋白(BCCP)融合蛋白。87 个氨基酸的 BCCP

区含有融合蛋白在大肠杆菌体内表达时共价连接于生物素的赖氨酸残基。按照制造商的说明书，预先混合生物素化的荧光素酶 (1.2 mg)和硫酸化酶 (0.4 mg)，4°C下结合于 2.0 mL Dynal M280 顺磁珠(10 mg/mL, Dynal SA)。在 2000 μL 珠洗涤缓冲液中将酶结合珠洗涤 3 次，重悬于 2000 μL 珠洗涤缓冲液中。

如下制备 Seradyn 微粒(Powerbind SA, 0.8 μm, 10 mg/mL, Seradyn Inc, Indianapolis, IN): 用 1000 μL 含 0.1% BSA 的 1X 测定缓冲液洗涤 1050 μL 储存液。以 9300 g 将微粒离心 10 分钟，除去上清液。再重复洗涤 2 次，将微粒重悬于 1050 μL 含 0.1% BSA 的 1X 测定缓冲液中。将珠子和微粒储存在冰上备用。

### 珠沉积

将 Dynal 酶珠和 Seradyn 微粒涡旋 1 分钟，将各 1000 μL 在新的微量离心管中混合，短暂涡旋，在冰上储存。将酶/ Seradyn 珠(1920 μl)与 DNA 珠(1300 μl)混合，用珠洗涤缓冲液将最终体积调节到 3460 μL。珠予以有序的层沉积。从珠洗涤缓冲液中取出光导纤维片，沉积第一层，即 DNA 和酶/Seradyn 珠的混合物。离心后，从光导纤维片吸出第一层上清液，并且沉积第二层，即 Dynal 酶珠。这部分详细描述了如何离心不同的层。

第一层。小心将在 60x60 mm<sup>2</sup> 光导纤维片表面上建立两个 30x60 mm<sup>2</sup> 的活性区的垫圈安装到夹具顶部分配的不锈钢销。光导纤维片放置在夹具中，其平滑的未蚀刻侧朝下，并且夹具顶部/垫圈固定在光导纤维片的蚀刻侧。然后通过拉紧对侧使其紧密指接，用提供的螺丝合适地固定夹具顶部。通过夹具顶部提供的两个入口，将 DNA - 酶珠混合物加入光导纤维片上。在珠混合物加入过程中，非常小心使气泡尽可能少。通过一次轻柔连续推动移液管的柱塞，完成每次沉积。整个组合件在具有 GH 3.8 - A 转子的 Beckman Coulter Allegra 6 离心机中以 2800 rpm 离心 10 分钟。离心后，用移液管除去上清液。

第二层。将 Dynal 酶珠(920 μL)与 2760 μL 珠洗涤缓冲液混合，并且按照前面的描述将 3400 μL 酶 - 珠悬浮液加入光导纤维片上。光导纤维片组合件在 2800 rpm 下离心 10 分钟，倒去上清液。从夹具上取下光导纤维片，在珠洗涤液中储存，直到准备加入仪器中。

## 在 454 仪器上测序

在含有 0.4 mg/mL 聚乙烯吡咯烷酮(MW 360,000), 1 mM DTT 和 0.1% Tween 20 的 1x 测定缓冲液中制备所有流动试剂。在含有 0.4 mg/mL 聚乙烯吡咯烷酮(MW 360,000), 1 mM DTT 和 0.1% Tween 20 的 1x 测定缓冲液中制备底物(300 μM D - 荧光素(Regis, Morton Grove, IL)和 2.5 μM 腺苷酰硫酸 (Sigma))。通过在含有 0.4 mg/mL 聚乙烯吡咯烷酮(MW 360,000), 1 mM DTT 和 0.1% Tween 20 的 1x 测定缓冲液中加入腺苷三磷酸双磷酸酶达到 8.5 单位的最终活性，制备腺苷三磷酸双磷酸酶洗液。在底物缓冲液中将脱氧核苷酸 dCTP、dGTP 和 dTTP (GE Biosciences Buckinghamshire, United Kingdom) 制备为 6.5 μM 的终浓度，将 α - 硫代脱氧腺苷三磷酸(dATPαS, Biolog, Hayward, CA) 和 焦磷酸钠(Sigma) 分别制备为 50 μM 和 0.1 μM 的终浓度。

454 测序仪由三个主要的组合件组成：射流子系统、光导纤维片药筒/流动室和成像子系统。试剂入口管、多阀门歧管和蠕动泵形成射流子系统的一部分。各个试剂连接于合适的试剂入口管，使试剂能够送递到流动室中，一次一种试剂，以预先编程的流速和持续时间流动。光导纤维片药筒/流动室在光导纤维片的蚀刻侧和流动室顶之间存在 300 μm 的间隙。流动室也包括用于控制试剂和光导纤维片温度的工具，以及不透光的遮蔽物。光导纤维片的磨光（未蚀刻）侧直接与成像系统接触。

通过预先编程的射流系统操作，实现测序试剂向光导纤维片孔的循环送递和从孔中洗去测序反应副产物。该程序写为界面控制语言 (ICL) 形式，指出每个步骤的试剂名称（洗液，dATPαS, dCTP, dGTP, dTTP 和 PPi 标准物）、流速和持续时间。对于所有试剂，流速设置为 4 mL/分钟，流动室内的线性速率是大约 1 cm/s。测序试剂的流动顺序组织成核心，其中第一个核心由以下部分组成：Ppi 流动(21 秒)，然后是 14 秒的底物流动，28 秒的腺苷三磷酸双磷酸酶洗涤，和 21 秒的底物流动。第一次 PPi 流动后是 21 个循环的 dNTP 流动(dC - 底物 - 腺苷三磷酸双磷酸酶洗涤 - 底物 dA - 底物 - 腺苷三磷酸双磷酸酶洗涤 - 底物 - dG - 底物 - 腺苷三磷酸双磷酸酶洗涤 - 底物 - dT - 底物 - 腺苷三磷酸双磷酸酶洗涤 - 底物)，其中每个 dNTP 流动包括 4 个单独的核心。每个核心是 84 秒长 (dNTP - 21 秒，底物流动 - 14 秒，腺苷三磷酸双磷酸酶洗涤

- 28 秒，底物流动 - 21 秒)；21 秒后和 63 秒后捕获图像。21 个循环的 dNTP 流动后，引入 PPi 核心，然后是另一 21 个循环的 dNTP 流动。测序运行的终末后是第三个 PPi 核心。总运行时间是 244 分钟。完成该运行需要的试剂体积如下：500 mL 每种洗涤溶液，100 mL 每种核苷酸溶液。在运行过程中，所有试剂保持在室温。流动室的和流动室入口管的温度控制在 30°C，所有进入流动室的试剂预热到 30°C。

## 成像系统

照相机是 Spectral Instruments (Tucson, AZ)600 系列照相机，其具有 Fairchild Imaging LM485 CCD (4096x4096 15 $\mu\text{m}$  像素)，直接连接于 1 - 1 成像纤维束。该照相机冷却到 -20°C，可以用两种模式之一操作：(i)帧转移模式，其中 CCD 的中心部分用于成像，而 CCD 的外部用于图像存储和缓慢读出（该模式用于较小的光导纤维片），或(ii)全帧模式，其中将整个 CCD 用于在每个流动循环的洗涤（即暗）部分发生的成像和读出（该模式用于 60x60 mm<sup>2</sup> 片）。该数据通过 4 个端口读出，CCD 的每个角具有一个端口。帧转移模式中信号集成设置在每帧 28 秒，帧转变时间为大约 0.25 秒；在全帧模式中，信号集成（帧持续时间）设置为 21 秒（洗涤捕获帧）和 63 秒（核苷酸捕获帧）。所有照相机图像以 UTIFF 16 形式储存在计算机硬盘(IBM eServer xSeries 337, IBM, White Plains, NY)中。

## 孔间扩散

为了评估我们的系统对从一个孔扩散到邻近孔的反应副产物的灵敏度，我们开发了一种简化的一维孔间扩散行为模式。我们发现在目前孔与孔的距离为 50  $\mu\text{m}$  时，ATP 的扩散诱导了在紧邻的孔中 10% 或更少数量级的背景信号。我们开发了正确的计算机算法来抑制该噪音来源。

我们建立了一种光导纤维面板的一维模型（即模拟孔的线性阵列），其中孔表示为集中的化学反应器，其在测序反应过程中产生焦磷酸和 ATP。在每个孔中，反应副产物的产生可以通过如下一组偶联的动力学公式来模拟：

$$\frac{d}{dt} \begin{bmatrix} \text{DNA}_n \\ \text{dNTP} \\ \text{PPI} \\ \text{ATP} \end{bmatrix}_{(1)} = \begin{bmatrix} -R_{bst(1)} \\ -R_{bst(1)} - k_c ([\text{dNTP}]_{(1)} - [\text{dNTP}]_{(0)}) \\ R_{bst(1)} - R_{sulf(1)} + R_{luc(1)} - k_c [\text{PPI}]_{(1)} \\ R_{sulf(1)} - R_{luc(1)} - k_c [\text{ATP}]_{(1)} \end{bmatrix}$$

该组公式的数字解示于图 49。

当考虑两个相邻孔时，可以添加下组公式：

$$\frac{d}{dt} \begin{bmatrix} \text{DNA}_n \\ \text{dNTP} \\ \text{PPI} \\ \text{ATP} \end{bmatrix}_{(2)} = \begin{bmatrix} -R_{bst(2)} \\ -R_{bst(2)} - k_c ([\text{dNTP}]_{(2)} - [\text{dNTP}]_{(0)}) \\ R_{bst(2)} - R_{sulf(2)} + R_{luc(2)} - k_c ([\text{PPI}]_{(2)} - \theta[\text{PPI}]_{(1)}) \\ R_{sulf(2)} - R_{luc(2)} - k_c ([\text{ATP}]_{(2)} - \theta[\text{ATP}]_{(1)}) \end{bmatrix}$$

孔间干扰的特征在于通过流动条件和孔几何学确定的质量转移系数  $k_c$  和混合比  $\theta$ 。这些参数 ( $k_c$  和  $\theta$ ) 是通过用有限要素方法解决完全三维的两孔问题获得的；然后将它们的值延伸到相似流动和孔几何学的多孔模拟。转运和化学反应现象的这种分离使得我们能够以高光导纤维面板占据数目刺激测序，并且探测相邻孔之间化学污染的影响。显示该孔间影响的该公式的数字解保持很低，即使在显著减少的孔距( $8\mu\text{m}$ )时也是如此（图 50）。

### 现场可编程门阵列(FPGA)

板载计算机安装有附加的 RC2000 PCI 板(Celoxica, Abingdon, UK)，其具有 6 百万个门 Virtex II FPGA (Field Programmable Gate Array) 芯片(Xilinx, San Jose, CA)。我们开发了软件，下载到在硬件中编码进行连续图像处理步骤的算法的 FPGA 二进制模块中。用 Handel-C (Celoxica, Abingdon, UK)设计 FPGA 硬件逻辑。在测序运行结束时，板载计算机可以获得所有数据，执行最终的信号调整，并且将片段与指定的基因组进行比对或进行鸟枪组装。在不使用 FPGA 的情况下，本文描述的测序运行的图像处理在板载计算机中进行额外的 6 小时。

### 图像处理

一旦用于成像系统，光导纤维片的位置不再改变；这使得基于每个测序运行前的 PPI 标准流动期间产生的光，图像分析软件能够确定每个孔的位置（在 CCD 像素座标上）。在操作中，通过照相机对整个片同时成像。通过大约 9 个像素对单个孔成像。处理数据中的第一步是在像素水平对每个采集的图像进行背景扣除，其中采用自动化确定每个像素的局部背景的“侵蚀 - 扩张”算法。然后，对于每次核苷酸流动，综合通过覆盖特定孔的像素在整个流动期间收集的光强度，在该特定流动中产生该特定孔的信号。我们校正了采集的图像，去除由于光渗漏（光导纤维覆层不是完全不透光的，并且将孔内产生的一小部分光传递到相邻的孔）和 ATP 或 Ppi（在合成中产生）从一个孔扩散到更下游的另一个孔而导致的孔间干扰。为了进行该校正，我们通过经验确定了低占据条件下干扰的程度和推导的解卷积矩阵，以便从每个孔的信号除去由相邻孔导致的贡献。为了解释每个孔中携带酶的珠子的数目变异和结合于每个珠子的模板拷贝数目的变异，进行了两类标准化：(i)首先参照测序运行前和后的 Ppi 标准物流动，对原始信号进行标准化，(ii)通过参照每个模板中包含的已知“关键”序列的前三个碱基掺入期间测量的信号，对这些信号进一步标准化。

## 信号处理

我们校正了每次流动和每个孔中的信号，以解释结转和不完全延伸。假定结转和不完全延伸的特定水平，可以容易计算任何已知序列的同步性损失程度。表 25，即模型计算的结果，说明了这些作用对序列准确度的影响；其显示了可以耐受的不完全延伸和结转的程度，假定没有进行校正，以便实现各种读出结果长度下大约 99% 的读出结果准确度。或者，通过使用反向转化来校正原始信号的同步性损失，可以用相似的不完全延伸和结转值实现更高水平的准确度，或，通过校正信号，可以在相同水平的准确度容纳更高水平的不完全延伸和结转。由于结转和不完全延伸的量以及作为基础的序列是现有技术未知的，我们的方法是基于重复技术和二维最小化，以实现适合测量的信号和模型输出之间的最小平方。结转和不完全延伸的影响特别接近读出结果的末端，这是因为这些误差的累积效应。

## 测试片段

我们建立了难以测序的片段，其包括长度不断增加的升和降段的相同碱基（均聚物）(2N, 3N, 4N, 5N, 6N, 5N, 4N, 3N, 2N), 其中散布单个核苷酸，以研究仪器的测序性能。这些片段使得我们能够从我们的评估中消除任何可能导致额外误差的样品制备或 PCR 伪迹。总体测序准确度示于表 25，通过图 51 的均聚物进一步细分。

纯化测试片段质粒 DNA。将各个测试片段克隆到 pBluescript II KS + 载体(Stratagene, La Jolla, CA)中, 转染到大肠杆菌培养物中，在甘油中 -80℃下储存，直到需要。将各自含有 6 个单独的测试片段之一的大肠杆菌培养物的各个管形瓶铺板，并且在 LB Amp / X-gal 琼脂培养皿板上生长。通过蓝/白色筛选，选择含有质粒的菌落，并且在含有氨苄青霉素的液体 LB 培养液中 37℃下饱和过夜。收获质粒，按照制造商的说明书用 QiaFilter Midi 质粒纯化试剂盒(Qiagen)从 25 mL 培养物中纯化。将纯化的质粒在 1X TE (10 mM Tris, 1 mM EDTA, pH 7.5) 中稀释到 10 ng/μL，在 -20℃下储存。

测试片段的 PCR 扩增。通过用一对 PCR 引物扩增，使测试片段生物素化，所述一对引物中的一个含有 5' 生物素。在 1.5 mL 试管中制备 980 微升 PCR 主混合物(1X Platinum HiFi 缓冲液(Invitrogen), 1mM dNTPs (Pierce), 2.5 mM MgSO<sub>4</sub> (Invitrogen), 1 μM 正向 (5' - CGTTTCCCCTGTGTGCCTTG - 3' SEQ ID NO:68) 和 1 μM 生物素化反向引物 (5' - 生物素 - 3' 连续 18 - 碳六甘醇间隔基 CCATCTGTT GCGTGCGTGTC - 3' SEQ ID NO:69) (IDT Technologies) 和 0.02 U/μL Platinum Hi-Fi Taq 聚合酶(Invitrogen)，通过涡旋彻底混合，除去 50 μL 阴性对照。将 20 微升给定测试片段加入其余部分，混合溶液，以 50 μL 等分物分散到 0.2mL PCR 试管中。对于 5 个剩余的测试片段中的每一个，重复该过程。将 PCR 反应物和相应的阴性对照置于 MJ 热循环仪中，在以下条件下扩增：94℃下 4 分钟热启动起始，然后是 39 个扩增循环的 94℃下 15 秒，58℃下 30 秒，68℃下 90 秒，和 68℃下单次延伸 120 秒。扩增以一直保持在 10℃下结束。通过根据制造商的说明书用 MinElute PCR Clean - Up 试剂盒(Qiagen)处理，纯化生物素化的 PCR 片段，不同之处是每个测试片段的 950 μL PCR 反应在 6 个 MinElut 柱上分开，并且最终步骤后合并。采用根据制造商的指导制备的 DNA 500 LabChip，用

Agilent 2100 BioAnalyzer 评估 PCR 产物的数量和质量。分析三份 1 $\mu$ L 等分物；纯化的 PCR 产物的浓度典型是 1-3pmol/ $\mu$ L。

使生物素化的 PCR 产物与链霉亲和珠结合。如下以 1 千万个 DNA 拷贝/珠，将生物素化的 PCR 产物固定在筛选的琼脂糖链霉亲和素包被的颗粒(Amersham)上。通过 28  $\mu$ m N/28/17/65 尼龙网(Sefar America, Depew, NY, USA)筛选 5 个 50 ml 瓶的琼脂糖链霉亲和素颗粒，以排除大珠。然后，使通过该过滤器的珠子通过孔径为 25  $\mu$ m 的 N25/19/55 尼龙网(Sefar America)。然后在 Multisizer 3 Coulter 计数器(Beckman)上对由滤器保留的、大小范围为直径 27 - 32 $\mu$ m 的珠子，随后用于结合生物素化的测试片段。用 100 $\mu$ L 2 M NaCl 溶液将 700,000 个筛选的珠子的等分物洗涤一次，短暂涡旋以重悬，然后在微量离心机中以最大速度离心 1 分钟，以沉淀珠子。然后除去上清液，此后再次用 2M NaCl 洗涤珠子，并且重悬于 30  $\mu$ L 2 M NaCl 中。将总共 11.6 皮摩尔生物素化的 PCR 产物加入珠子，涡旋，使珠子重悬于溶液，并且使其室温下在速度 7 的滴定板振荡器上与链霉亲和素珠结合 1 小时。通过在水平管旋转器上在碱性解链溶液(0.1 M NaOH / 0.15 M NaCl)中温育 10 分钟，除去非生物素化的第二链。弃去含有变性的、非生物素化链的上清液，用 100 $\mu$ L 解链溶液洗涤珠子一次，并且用 100 $\mu$ L 1 X 退火缓冲液(50 mM Tris - 醋酸, pH 7.5; 5 mM MgCl<sub>2</sub>)洗涤三次。然后在微量离心机上以最大速度将珠子离心 1 分钟，弃去上清液，将珠子重悬于 25 $\mu$ L 1 × 退火缓冲液中。在珠悬浮液中加入 5 微升 100 $\mu$ M 测序引物(5' - CCATCTGTTCCCTCCCTGTC - 3', IDT Technologies SEQ ID NO:70)。然后将珠/引物混合物涡旋 5 秒，置于 MJ 热循环仪中进行以下 4 个阶段的退火程序：60°C 下温育 5 分钟，以 0.1°C/秒降低到 50°C，50°C 下 1 分钟，以 0.1°C/秒降低到 40°C，40°C 下保持 1 分钟，以 0.1°C/秒降低到 15°C，15°C 下保持。退火步骤后，用 100 $\mu$ L 1X 退火缓冲液(20 mM Tris, pH 7.5 和 5 mM 醋酸镁)将珠子洗涤两次，用 1X 退火缓冲液重悬于 200 $\mu$ L 的终体积。在 4°C 冰箱中在贴标签的管中以 10  $\mu$ L 等分物储存珠子直到需要。

## 高质量珠

每个孔中的每次流动导致未掺入、插入 1 个或 2 个或 3 个核苷酸。对于任何测序运行，可以组装这些组中每一组的信号强度直方图（当用

已知序列处理时）。如图 52 所示，各个组的信号强度略微重叠。通常，良好读出结果（即以很少的误差作图到参照基因组的那些）的大多数信号接近等于掺入的核苷酸数目的积分值。图 53 显示了对连续长度的均聚物测量的所有信号的平均值随着均聚物长度线性增加到非常高的准确度。我们发现了这样的读出结果，其中显著数目的落入阴性流动（其中没有掺入核苷酸）和阳性流动（其中掺入至少一个核苷酸）之间的重叠区 ( $0.5 < \text{信号} < 0.7$ ) 的信号是低质量的（即，不作图到基因组的任何位置，或作图到基因组的位置但具有大量误差），主要是因为所述读出结果来自携带两个或多个模板的拷贝的珠子。这使得我们能够推理地开发用于选择“高质量读出结果”的过滤器：对于每个读出结果，我们对落入重叠区的流动数进行计数，并且仅仅选择所述流动数目小于流动总数的 5% 的那些读出结果。对于不满足该标准的读出结果，我们通过消除流动，从读出结果末端开始逐渐修正读出结果，直到满足该标准（未确定区中的流动数目  $<$  剩余流动的 5%），或流动数目减少到小于 84(21 个循环)，此时认为读出结果从高质量读出结果库中过滤出来。

## 碱基调用

原则上，观察到的信号强度直接表明掺入的核苷酸数目。但是，如图 52 所示，各个均聚物的信号强度分布略微重叠。如果不是由于该重叠，可以对任何给定的信号序列明确进行碱基调用。在基于焦磷酸的测序中，两类直接误差是过度调用（调用比基因组中实际存在的碱基多一个碱基）或调用不足（调用比基因组中实际存在的碱基少一个碱基）。碱基的标识是不考虑的，因为它是通过一次加入一个已知核苷酸而确定的。取代错误（错误调用一个碱基代替另一个）由两个连续错误（调用不足，随后是过度调用，或反之亦然）导致，并且因此显著更少。我们观察到，在单个读出结果水平，文库读出结果的平均错误率比测试片段高（比较图 51 和图 54）。我们开发了预期的信号的计算机模型，用于证实我们的测量结果和较高错误率是与以下假设一致的，即，当对文库测序时，一些珠子携带一个以上模板的拷贝。这些读出结果的大多数是通过上文描述的选择过程过滤出来的。但是，它们（对于它们，混合物显著有利于一个模板）可能不能过滤出来，并且对总体错误率起主要作用。

在各个读数水平,表 25 和 26 报道了对于比对的碱基总数的错误率。这些数目类似于目前的测序仪报道的错误率;但是,它们不能最佳表征仪器的内在性能,因为在阴性流动时也可以出现误差。每个流动,无论阴性或阳性,都可以赋予错误率。例如,对于在表 26 中分析的 238,066 个生殖道支原体读数,针对流动总数的插入率 1.53% (当针对比对的碱基时是 2.01%) ;类似地,针对流动总数的缺失错误率是 1.48% (当针对比对的碱基时是 1.94%) 。

### 质量评分

与给定的信号值相关的任何特定碱基调用的置信度(或“质量”)是对于给定均聚物长度,该信号落入信号分布中的位置的函数。基于大量运行(其中我们对多种已知基因组(腺病毒、金黄色葡萄球菌、生殖道支原体)以及测试片段进行了测序)和测试片段,并且对得到的读出结果作图,我们确定了符合对数分布的阴性流动,而所有阳性流动是正态分布的,其平均值(图 53)和标准差与作为基础的均聚物长度成比例;此外,这些分布在不同基因组和测试片段中保持显著的变异。该观察结果使得能够计算调用的每个单独碱基的质量评分。为了评估特定碱基调用的质量评分,必须确定测量的信号来自长度至少等于调用长度的均聚物的概率。例如,如果对于特定信号调用两个 A,则第二个 A 的质量评分是通过观察到的信号来自长度为 2 个更多的均聚物的概率而给出的。由于给定均聚物长度下测量信号的概率是经验建立的,可以用 Baye 定理确定特定均聚物长度产生观察到的信号的概率,如下:

$$P(n|s) = \frac{P(s|n)P(n)}{\sum_j P(s|j)P(j)}$$

其中 s 是观察到的信号, n 是产生信号的均聚物的长度。如上文的描述,给定均聚物长度 n 时测量信号 s 的概率  $P(s|n)$  满足高斯分布。对于随机核苷酸序列,遇到长度为 n 的均聚物的概率简单地是  $1/4^n$  (忽略乘法标准化常数)。对于每个片段,赋予调用的每个碱基的质量评分可以报道为 phred - 当量,其中采用以下转换:

$$Q = -10 \log_{10} [P(\geq n | s)]$$

我们通过将计算的 phred 评分与观察的 phred 评分进行关联，对除用于建立信号分布的基因组以外的已知基因组进行测序，证实了该方法的有效性（图 56）。我们的关联显示了与最多 phred 50 的良好相应性，并且对 Sanger 测序和毛细管电泳建立的相应性更有利(Li, M., Nordbord, M. and Li, L. M., Nucleic Acids Research 32, 5183 (2004))。

### 流动 - 间隙作图，共有区准确度和基因组覆盖

考虑到核苷酸流动的顺序，给定的参照基因组暗示了理想信号值的已知延续。这种理想流动图分为特定长度（默认长度是 24 次流动）的连续、重叠、亚流动图，给其做索引，以便允许非常迅速的检索（每个亚流动图在阳性流动开始）。为了将询问流动图作图到靶，我们将询问流动图分成变化的亚流动图，其具有用于做索引步骤的长度，并且检索做索引的理想亚流动图的间隙。完美的匹配将询问流动图锚定于参照基因组。然后，从 5' 末端开始，沿着读出结果的整个长度移动，评估读出结果的比对。选择满足用户指定的总错配阈值的最长片段，在此处终止比对，并且平衡读出结果。在非常低的严格性水平将读出结果与参照比对，以检测突变或其它基因组变异。一旦进行了所述比对，对来自相应于靶中相同位置的各个读出结果的流动信号取算术平均值，此后进行各个碱基调用。如图 54 所示，该程序在减少错误率方面非常有效；无论是对从头组装进行重新测序还是共有区碱基调用，它都是同样有用的。通过测量与相应均聚物的最近信号阈值的距离的绝对值，并且除以特定基因组位置测量的所有信号的标准化的标准差，我们估计了平均信号的质量（而不依赖于对作为基础的序列的了解）。我们将该比值称作 Z - 评分。为了增强观察到的变异的可靠性，通过强加最小 Z - 评分过滤了共有序列，以得到高质量共有序列。通过采用确切了解的序列，我们确定了产生共有区调用质量的估计值和最小 Z - 评分和共有区准确度之间的关联的错误数。我们报道了基于共有序列准确度 99.99% 或更佳的区域的基因组覆盖，其典型是通过选择等于 4 的最小 Z - 值而实现的。没有 Z - 值限制，我们自然以略低的共有区准确度获得了更大覆盖。

## 从头序列组装程序

我们选择了高质量读出结果（如上文描述），以确保要处理的流动图最可能由来自原始样品的序列数据组成。Overlapper 进行完整的全对全片段比较，以鉴定片段之间的所有可能的重叠。为了组装仪器产生的读出片段，Overlapper 通过直接比较每个读出结果的流动图，评估读出结果相似性；我们目前利用分级产物评估流动图之间的相似性：

$$\text{评分} = \sum_i S_{1i} \bullet S_{2i}$$

其中  $S_{1i}$  和  $S_{2i}$  是信号强度（进行了标准化，使得每个“载体”的长度等于 1），并且在推定的重叠区上进行求和。我们发现 0.85 – 0.90 的阈值提供了最优预测性和选择性。如果两个流动图区之间观察到的重叠评分超过选择的最小严格性值，对该读出结果对设置重叠标记。（重叠确定中也考虑反向互补读出结果的可能性）。为了增加效率，Overlapper 使用散列索引法，以迅速鉴定可能考虑为潜在重叠候选物的片段。考虑到通过 Overlapper 确定的读出结果之间的该组所有逐对重叠，Unitigger 模块将这些读出结果分组为统一群。统一群是一些读出结果的集合，所述读出结果彼此之间的重叠是一致的，并且不受统一群外部的读出结果的竞争。统一群的末端代表组装的基因组中或来自完整未测序区的重复区的入口或出口。从最大深度重叠的一致链构建统一群（即，彼此最大重叠的逐对读出结果）。最后，Multialigner 采用组成统一群的所有读数结果，并且对所有读出信号进行比对。其通过首先对给定区域的信号取平均值以获得用于进行实际碱基调用的单个平均信号，进行共有区调用。

然后，将 Multialigner 产生的统一群送入重叠群优化过程，其中修复由重叠检测或最大深度重叠的链使用的缺陷导致的断裂。Multialigner 统一群具有它们的末端伸入重复区或通过一个或多个可能使最大重叠区的链断裂的错误读出结果而伸入“断裂”为多个重叠群的区域的性质。重叠群优化过程包括三个步骤。第一个步骤进行全对全统一群比较，并且连接在统一群之间检测到的任何重叠区。在核苷酸间隙中进行的该比较之后是分支点分析，其基于重叠群序列从共同序列分支的位置鉴定重复区边界。重叠群在这些边界破裂，输出任何大于 500 个碱基的非重复重叠群。

重叠群优化过程的第二步采用来自第一步的重叠群，并且进行“重新缝合”，其中用跨两个重叠群末端的任何读出结果连接这些重叠群。

至于第一步，这是在核苷酸间隙中进行，并且用分支点分析程序鉴定任何重复区连接。最后一步是质量控制步骤，其中将所有读出结果作图到得到的重叠群序列，重叠群在小于 4 个跨越读出结果的位置破裂，并且仅仅输出大于 500 个碱基的重叠群。

最后，进行共有区产生步骤，以计算最终重叠群共有序列。该步骤利用前面章节描述的相同的流动间隙作图和共有区产生程序，不同之处是进行重复程序，其中新的共有区重新用作程序的输入值，直到 Z - 值为 4 或更多的碱基都不再改变。然后通过组装程序处理输出得到的重叠群和共有序列。

### 双末端测序

为了在单独的孔中从单个模板的两个末端开始进行测序（“双末端测序”），改变了乳状液 PCR 程序，使两个寡核苷酸引物（每个一个方向）连接于琼脂糖 DNA 捕获珠。构建用于 ssDNA 文库制备的衔接子序列，使得两个独特的测序引物掺入文库片段（每条链一个）。在双末端测序中，使用两个测序引物，第二个测序引物用 3' - 磷酸保护。如同单末端测序一样，以一个方向进行测序。通过使含有 25 mM Tricine, 5 mM 醋酸镁, 1 mM DTT, 0.4 mg/mL PVP, 0.1 mg/mL BSA, 0.01% Tween 和每种双脱氧核苷酸各 2  $\mu$ M 以及每种脱氧核苷酸各 2  $\mu$ M 的加帽缓冲液流动，终止第一条链测序。通过使含有 25 mM Tricine, 5 mM 醋酸镁, 1 mM DTT, 0.4 mg/mL PVP, 0.1 mg/mL BSA, 0.01% Tween 和 8.5 单位/L 腺苷三磷酸双磷酸酶的腺苷三磷酸双磷酸酶缓冲液流过，除去残留的脱氧核苷酸和双脱氧核苷酸。通过流过含有 25 mM Tricine 中的 5 个单位/mL 的牛肠碱性磷酸酶，5 mM 醋酸镁，1 mM DTT, 0.4 mg/mL PVP, 0.1 mg/mL BSA, 0.01% Tween 的切割缓冲液，从修饰的 3' 磷酸化引物的 3' 末端除去磷酸基团，从而使第二个阻断的引物去除阻断。通过流过 1000 个单位/mL 的 Bst DNA 聚合酶大片段而加入聚合酶，捕获所有可用的引物位点，从而使第二个未阻断的引物激活。通过 Bst DNA 聚合酶大片段进行的第二条链的测序是通过像单末端测序那样在预定数目的循环中连续添加核苷酸而进行的。在证明理论的实验中，我们证明了双末端测序确实产生了成对的末端读出结果，而第二条链的测序质量没有显著损失。图 57 显示了金黄色葡萄球菌 COL 的双末端测序运行(de Lencastre, H.,

Tomasz, A., Antimicrob Agents Chemother. 38, 2590 (1994) 21 cycles followed by 21 cycles)中来自扩增的片段的作图的成对读出结果的读出结果长度；表 26 概括了对于两个读出结果在各个读出结果水平的测序统计。

表 25 作为不同读出结果长度时不完全延伸和结转的函数的预测的测序准确度

读出结果长度	100	100	200	200	400	400
不完全延伸	0.0	0.0025	0.0	0.0013	0.0	0.0007
结转	0.01	0.0	0.005	0.0	0.003	0.00
预测的准确度	约 99%	约 99%	约 99%	约 99%	约 99%	约 99%

表 26 金黄色葡萄球菌的双末端测序运行的测序统计

测序运行概括	金黄色葡萄球菌 读出结果 1	金黄色葡萄球菌 读出结果 2
光导纤维片的大小	30x60 mm <sup>2</sup>	30x60 mm <sup>2</sup>
运行时间/循环数	126 min/21	126 min/21
高质量读出结果	162,261	148,187
平均读出结果长度	44 b	44 b
高质量读出结果中的碱基数目	7,115,310	6,575,545
单独读出结果		
作图到单个位点的读出结果	63,184	56,027
单独读出结果插入错误率	0.98%	1.49%
单独读出结果缺失错误率	2.57%	4.41%
单独读出结果取代错误率	0.23%	0.47%

表 27 生殖道支原体的 8 个额外的独立测序运行的概括统计

测序运行概括	生殖道 支原体 2	生殖道 支原体 3	生殖道 支原体 4	生殖道 支原体 5	生殖道 支原体 6	生殖道 支原体 7	生殖道 支原体 8	生殖道 支原体 9
高质 量读出结果	294,529	301,482	285,905	302,193	287,364	294,451	288,613	287,506
平均读出结果长度	111 b	111 b	111 b	112 b	111 b	112 b	111 b	112 b
高质 量读出结果中的碱基数目	32,633,064	33,426,447	31,683,321	33,930,424	31,927,352	32,176,323	32,125,411	32,131,739
Phred 20 和更高的碱基	25,576,988	26,492,794	24,201,381	26,632,816	24,608,146	24,982,885	24,696,409	25,181,881
重新测序								
作图到单个位点的读出结果	228,420	235,770	221,401	234,679	221,025	222,527	224,651	224,478
作图的读出结果中的碱基数目	26,553,470	27,423,465	25,734,641	27,191,266	25,720,866	25,845,112	26,139,994	26,103,403
插入错误率	1.85%	1.64%	2.07%	2.41%	1.85%	2.00%	1.91%	1.76%
缺失错误率	1.65%	1.59%	1.73%	2.12%	1.66%	1.74%	1.63%	1.65%
取代错误率	0.69%	0.69%	0.70%	0.73%	0.68%	0.68%	0.68%	0.67%
重新测序共有区								
平均过量取样 (独特读出结果)	38x	39x	37x	38x	37x	37x	38x	37x
基因组覆盖, 所有 ( $Z \geq 4$ )	98.1%	98.3%	98.1%	98.2%	98.1%	98.1%	98.1%	98.2%
共有区准确度, 所有 ( $Z \geq 4$ )	99.996%	99.996%	99.998%	99.996%	99.994%	99.998%	99.996%	99.996%
共有区插入错误率, 所有 ( $Z \geq 4$ ) (插入的碱基数)	0.003% (17 b)	0.001% (17 b)	0.001% (6 b)	0.002% (12 b)	0.005% (27 b)	0.000% (3 b)	0.002% (13 b)	0.002% (12 b)
共有区缺错误差率, 所有 ( $Z \geq 4$ ) (缺失的碱基数)	0.001% (8 b)	0.001% (8 b)	0.001% (8 b)	0.002% (9 b)	0.001% (6 b)	0.002% (11 b)	0.002% (9 b)	0.002% (10 b)

所有运行 (42 个循环) 都是在 60×60 mm<sup>2</sup> 光导纤维片上进行的。为了清楚, 仅仅示出了  $Z \geq 4$  的共有序列的统计。从断裂和衔接子连接的基因组 DNA 开始, 我们建立了每次运行 2 份珠乳状液。8 次测序运行产生了平均 292,755 个高质量读出结果 (CV=2.2%), 得到平均 32,504,260 个碱基 (CV=2.4%), 其平均值 25,296,663 (CV=3.5%) 是 phred 20 或更佳质量的。平均 226,619 个读出结果 (CV=2.56%) 作图于单个位置, 产生平均 26,339,027 个作图的碱基 (CV=2.5%)。将最小 Z - 值设置为等于 4, 平均覆盖为 98.15% (CV=0.08%) 时, 平均共有区准确度为 99.996% (CV=0.001%)。

表 28 肺炎链球菌的概括统计(2 次运行)

测序概括	
光导纤维片的大小	60x60 mm <sup>2</sup>
运行时间/循环数	243 min/42
高质量读出结果	529,077
平均读出结果长度	100 b
高质量读出结果中的碱基数目	52,692,071
Phred 20 和更高的碱基	40,014,199
重新测序	
作图到单个位点的读出结果	382,829
作图的读出结果中的碱基数目	43,129,894
单独读出结果插入错误率	1.17%
单独读出结果缺失错误率	2.50%
单独读出结果取代错误率	0.83%
重新测序共有区	
平均过量取样	17x
覆盖, 所有 ( $Z \geq 4$ )	96.3% (91.2%)
共有区准确度, 所有 ( $Z \geq 4$ )	99.85% (99.981%)
共有区插入错误率, 所有 ( $Z \geq 4$ )	0.11% (0.018%)
共有区缺失错误率, 所有 ( $Z \geq 4$ )	0.04% (0.001%)
共有区取代错误率, 所有 ( $Z \geq 4$ )	0.003% (0.0009%)
重叠群数目	184
从头组装	
覆盖, 所有 ( $Z \geq 4$ )	92.0% (90.44%)
共有区准确度, 所有 ( $Z \geq 4$ )	99.97% (99.992%)
重叠群数目	255
平均重叠群大小	7.8 kb

第一次运行具有 110,516 个完美读出结果(40.8%)，第二次运行具有 88,068 个完美读出结果(34.1%)。单独读出结果错误率是针对作图的读出结果中碱基的总数。存在 12 个错误组装的重叠群 (3 个是由于串联重复失败, 9 个是由于分开的基因组区的错误连接)；在评估组装结果前手工校正。至于生殖道支原体，大多数(88%)未覆盖的碱基属于基因组的不可分辨的重复区，其覆盖基因组的 7%。当我们选择最低 Z- 值为 4 的碱基时，覆盖了 90.44% 的基因组，共有区准确度为 99.992%。相比较而言，肺炎链球菌的原始测序产生了 390 个大于 1.5 kbp 的重叠群 (Tettelin, H. et al., Science 293, 498 (2001))。

<110> Keith E. McDade

Joseph M. Fierro

James R. Knight

Jerry Charumilind

John H. Leamon

Eugene W. Myers

John W. Simpson

Greg A. Volkmer

<120> 对核酸进行扩增和测序的方法

<130> 21465-510-CIP-061

<140> 11/195, 254

<141> 2005-08-01

<150> 10/767779

<151> 2004-01-28

<150> 60/476602

<151> 2003-06-06

<150> 60/476504

<151> 2003-06-06

<150> 60/443471

<151> 2003-01-29

<150> 60/476313

<151> 2003-06-06

<150> 60/476592

<151> 2003-06-06

<150> 60/465071

<151> 2003-04-23

<150> 60/497985

<151> 2003-08-25

<160> 77

<170> PatentIn version 3.3

<210> 1  
<211> 44  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 1  
cgttccct gtgtgccttg ccatctgttc cctccctgtc atgc 44

<210> 2  
<211> 40  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 2  
gcatgacagg gagggAACAG atggcaaggc acacagggga 40

<210> 3  
<211> 40  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 3  
gcatgacacg caacagggga tagggacacg cacgcaacag 40

<210> 4  
<211> 44  
<212> DNA  
<213> 人工序列

<220>  
<223> 引物

<400> 4  
ccatctgttg cgtgcgtgtc cctatcccct gttgcgtgtc atgc 44

<210> 5  
 <211> 64  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 5  
 tcgtgtgagg tctcagcatc ttatgtatat ttacttctat tctcagttgc ctaagctgca 60  
 gccat 64

<210> 6  
 <211> 24  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 6  
 gacctcacac gatggctgca gctt 24

<210> 7  
 <211> 24  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 7  
 gacctcacac gatggctgca gctt 24

<210> 8  
 <211> 64  
 <212> DNA  
 <213> 人工序列

<220>  
 <223> 引物

<400> 8  
 tttatatgtt ttctacgact ctggagtggtt ctaccgacgt cgaatccgtt gactcttatac 60

ttca	64
<210> 9	
<211> 34	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 9	
ctagctcgta catataaatg aagataagat cctg	34
<210> 10	
<211> 30	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 10	
gacctcacac gagtagcatg gctgcagtt	30
<210> 11	
<211> 64	
<212> DNA	
<213> 人工序列	
<220>	
<223> 引物	
<400> 11	
tcgtgtgagg tctcagcate ttatgtatat ttacttctat tctcagttgc ctaagctgca	60
gccca	64
<210> 12	
<211> 40	
<212> DNA	
<213> 人工序列	
<220>	

<223> 引物

<400> 12

gcttacctga ccgacctctg cctatccctt gttgcgtgtc

40

<210> 13

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 13

ccattccccca gtcgtttt ccatctgttc cttccctgtc

40

<210> 14

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 14

gcttacctga ccgacctctg

20

<210> 15

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 15

ccattccccca gtcgtttt

20

<210> 16

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 16

ccattccccca gctcggtttg ccatctgttc cctccctgtc tcag

44

<210> 17

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 17

ccatctgttc cctccctgtc

20

<210> 18

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 18

cctatccccc gttgcgtgtc

20

<210> 19

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 19

cgtttccccc gtgtgccttg

20

<210> 20

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 20

ccatctgttg cgtgcgtgtc

20

<210> 21

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 21

ccatctgttc cctccctgtc

20

<210> 22

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 22

cctatccccct gttgcgtgtc

20

<210> 23

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 23

ccatctgttg cgtgcgtgtc

20

<210> 24

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 24

cgtttccctt gtgtgccttg

20

<210> 25

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 25

catcttgtcc actaggctct

20

<210> 26

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 26

ccatctgttg cgtgcgtgac

20

<210> 27

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 27

accagcactc gcaccacc

18

<210> 28

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 28

cgtttccccc gtgtgccttg

20

<210> 29

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 29

tacctctccg cgttagcgc

18

<210> 30

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 30

ccatctgttg cgtgcgtgtc

20

<210> 31

<211> 17

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 31

ccccggacga gacgcag

17

<210> 32

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 32

atctctgcct actaaccatg aag

23

<210> 33

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 33

catcttgtcc actaggctct

20

<210> 34

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 34

gtttctctcc agcctctcac cga

23

<210> 35

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 35

accagcactc gcaccacc

18

<210> 36

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 36

atctctgcct actaaccatg aag

23

<210> 37

<211> 18

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 37

tacctctcccg cgtaggcg

18

<210> 38

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 38

gtttctctcc agcctctcac cga

23

<210> 39

<211> 17

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 39

ccccggacga gacgcag

17

<210> 40

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 40

cgtttccctt gtgtgccttg

20

<210> 41

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 41

ccatctgttg cgtgcgtgtc

20

<210> 42

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 42

ctagcttagca tggaaggcgcc agcagca

27

<210> 43

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 43

ccgggatccc tcgatgacga ccagcggc

28

<210> 44

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 44

atgcacatgg ttgacacagt ggt

23

<210> 45

<211> 22

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 45

atgcacatgg ttgacacagt gg

22

<210> 46

<211> 25

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 46

atgccaccga cctagtcctca aacctt

25

<210> 47

<211> 51

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 47

tattgttgat gctgtaaaaa gaagctactg gtgttagtatt tttatgaagt t

51

<210> 48

<211> 47

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 48

tgctcaaaga attcatttaa aatatgacca tatttcattt tatcttt

47

<210> 49

<211> 48

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 49

aagcgaacag tcaagtacca cagtcagttg acttttacac aagcgat

48

<210> 50

<211> 47

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 50

tacagggttt ggtatgccat ttgcgatttg ttgcgcttgg ttagccg

47

<210> 51

<211> 52

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 51

aacatataaaa catcccstat ctcaatttcc gttccatgt aacaaaaaaaaa gc

52

<210> 52

<211> 39

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 52

tagatatac ttgcgtgtta ctggtaatgc aggcatgag

39

<210> 53

<211> 41

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 53

attcaactct ggaaatgctt tcttgatacg cctcgatgat g

41

<210> 54

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 54

gataggaggac tgcaatggca atgggttaaa ggcacatcatcg

40

<210> 55

<211> 45

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 55

tgtatctcga tttggattag ttgcttttg catcttcatt agacc

45

<210> 56

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 56

cattaacatc tgcaccagaa atagcttcta atacgattgc

40

<210> 57

<211> 46

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 57

gcgacgacgt ccagctaata acgctgcacc taaggctaat gataat

46

<210> 58

<211> 43

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 58

aaaccatgca gatgctaaca aagctcaagc attaccagaa act

43

<210> 59

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 59

tgttgctgca tcataattta atactacatc atttaattct ttgg

44

<210> 60

<211> 51

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 60

gcagatggtg tgactaacca agttggtaaa aatgccctaa atacaaaaga t

51

<210> 61

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 61

ccatctcata cctgcgtgtc ccatctgttc cctccctgtc tcag

44

<210> 62

<211> 44

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 62

cctatcccct gtgtgccttg cctatcccct gttgcgtgtc tcag

44

<210> 63

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 63

ccatctgttg cgtgcgtgtc

20

<210> 64

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 64

cgtttcccccgtgtgccttg

20

<210> 65

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 65

ccatctgttg cgtgcgtgtc

20

<210> 66

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 66

cgtttcccccgtgtgccttg ccatctgttc cctccctgtc

40

<210> 67

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 67

ccatctgttc cctccctgtc

20

<210> 68

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 68

cgtttccctt gtgtgccttg

20

<210> 69

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 69

ccatctgttg cgtgcgtg

20

<210> 70

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 70

ccatctgttc cctccctgtc

20

<210> 71

<211> 24

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 71

gacctcacac gatggctgca gctt

24

<210> 72

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 72

tcgtgtgagg tctcagcatc ttatgtatat ttacttctat tctcaggtgc ccaagctgca 60

gccat

64

<210> 73

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 73

acttctatttc tcagttgcct aagctgcagc cattgtgtga ggtctcagca tcttatgtat 60

attt

64

<210> 74

<211> 32

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 74

gtcctagaat agaagtaaat atacatgctc ga 32

<210> 75

<211> 30

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 75

gacctcacac gagtagcatg gctgcagctt 30

<210> 76

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 76

tcgttgagg tctcagcatc ttatgtatat ttacttctat tctcagttgc ctaagctgca 60

gccat

64

<210> 77

<211> 6

<212> DNA

<213> 人工序列

<220>

<223> 引物

<400> 77

tgctac

6



图 1A

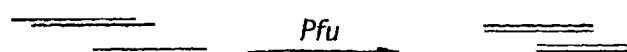


图 1B

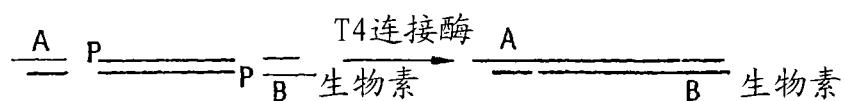


图 1C

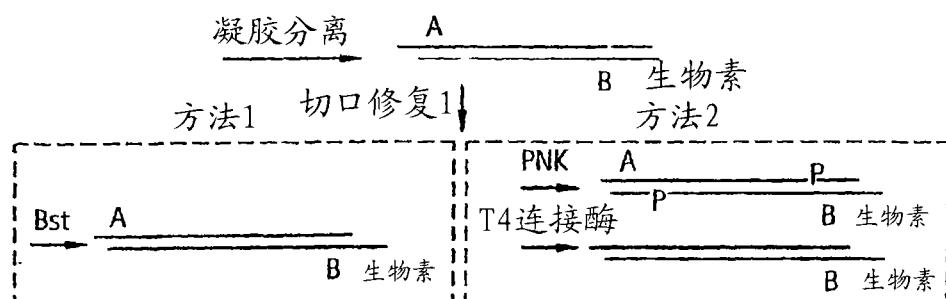


图 1D

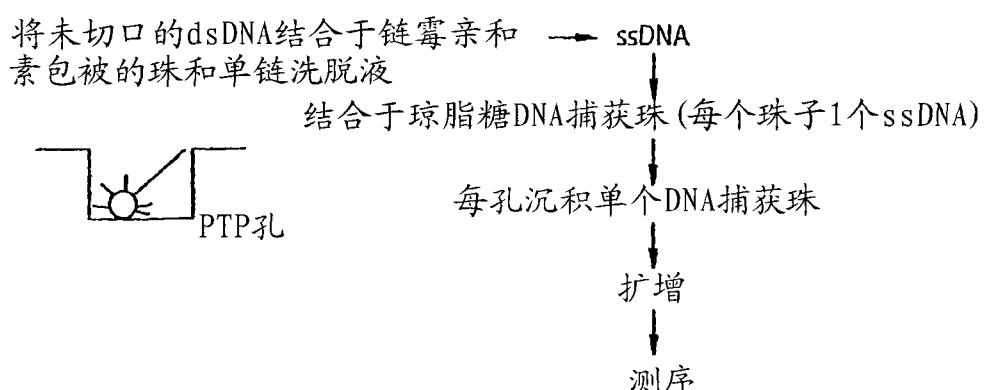


图 1E

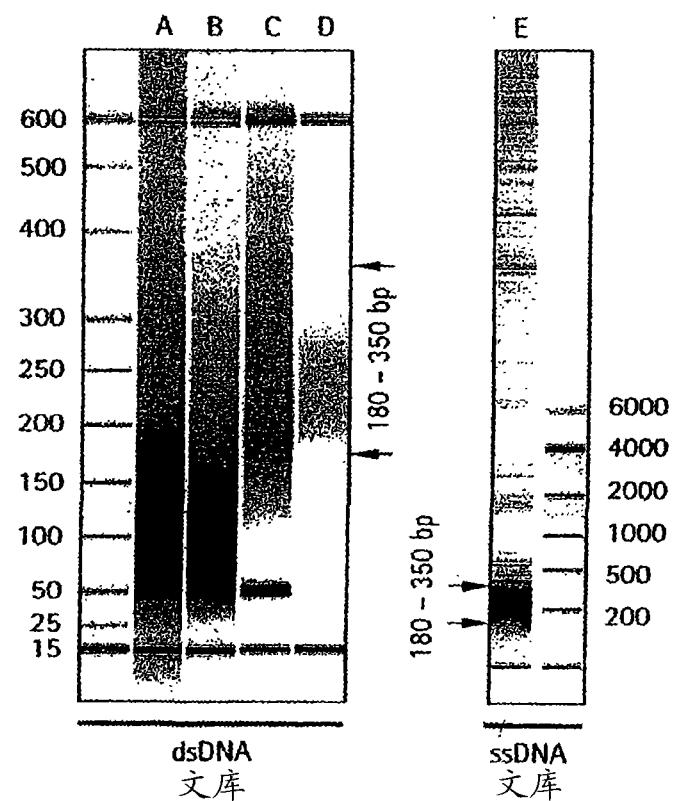


图 1F

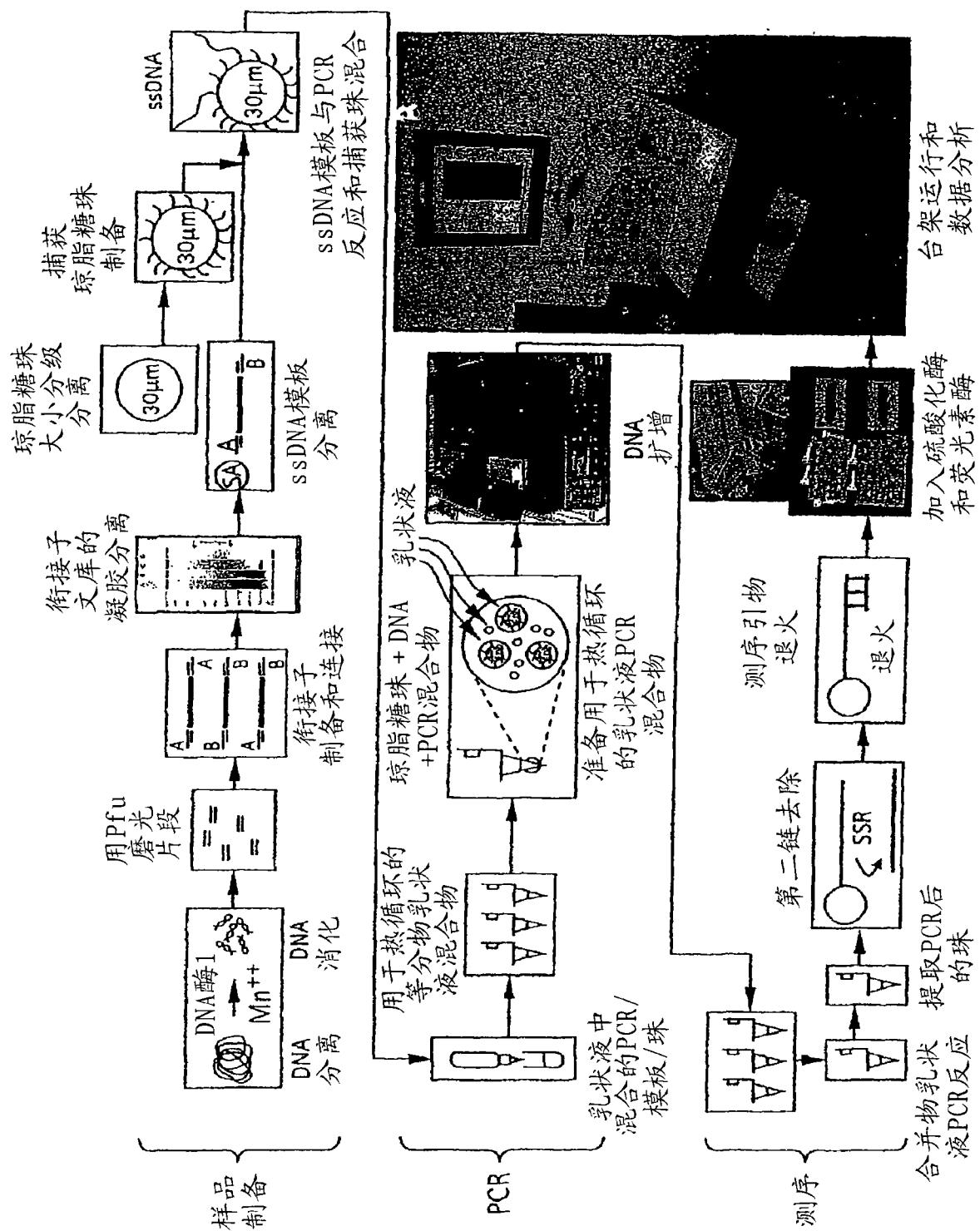


图 1G

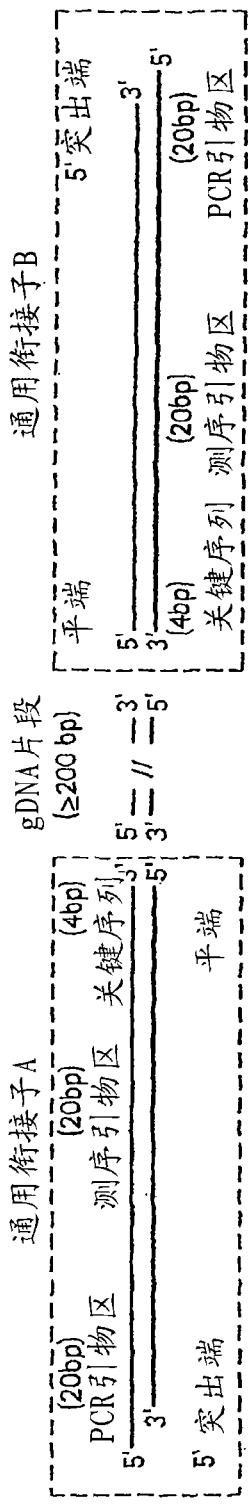


图 2A

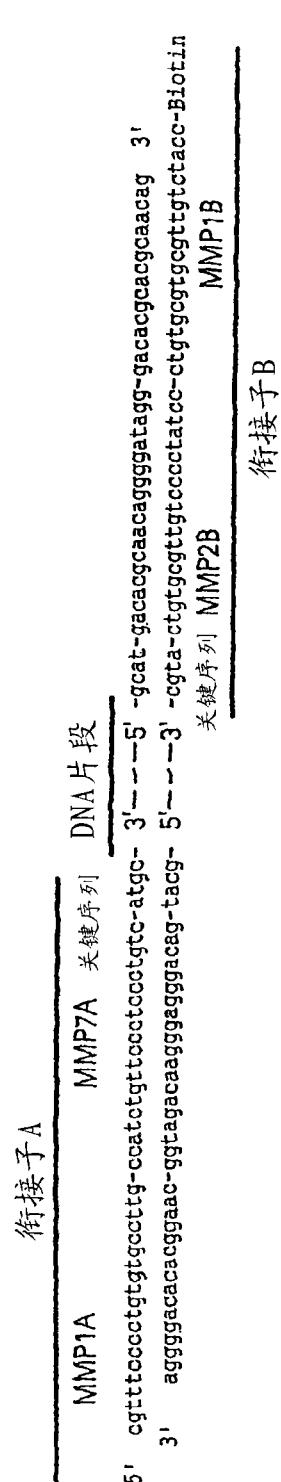


图 2B

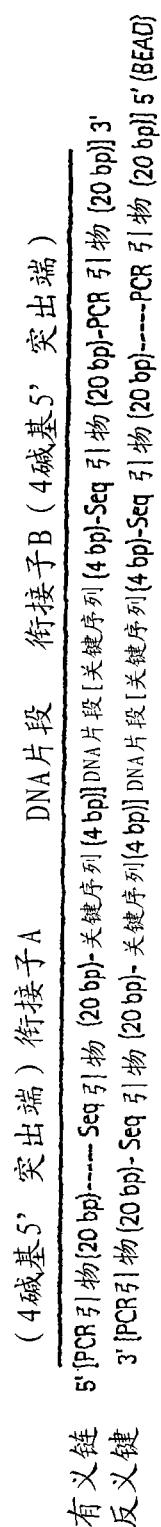
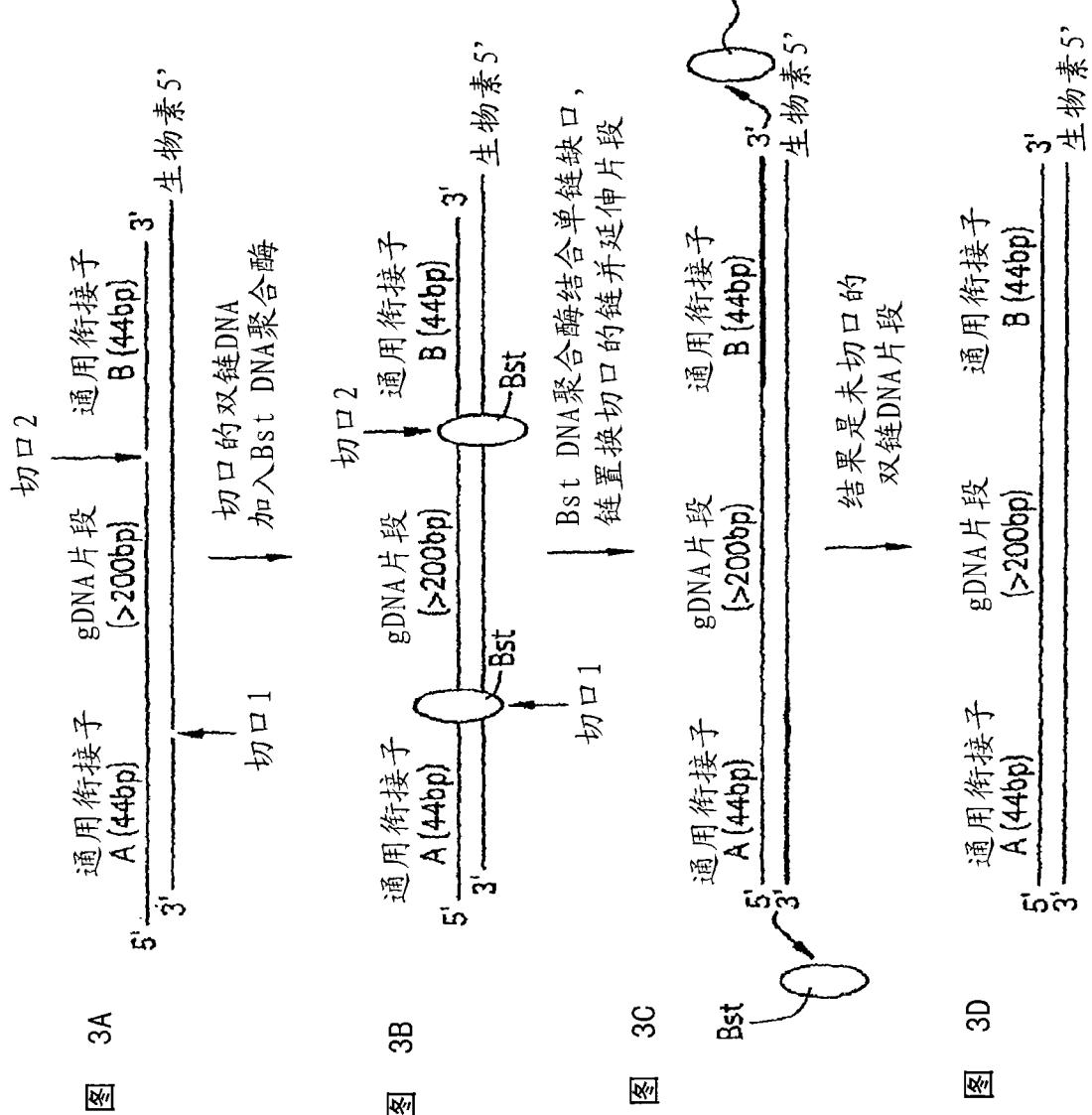


图 2C



可能的双链基因组DNA文库种类：

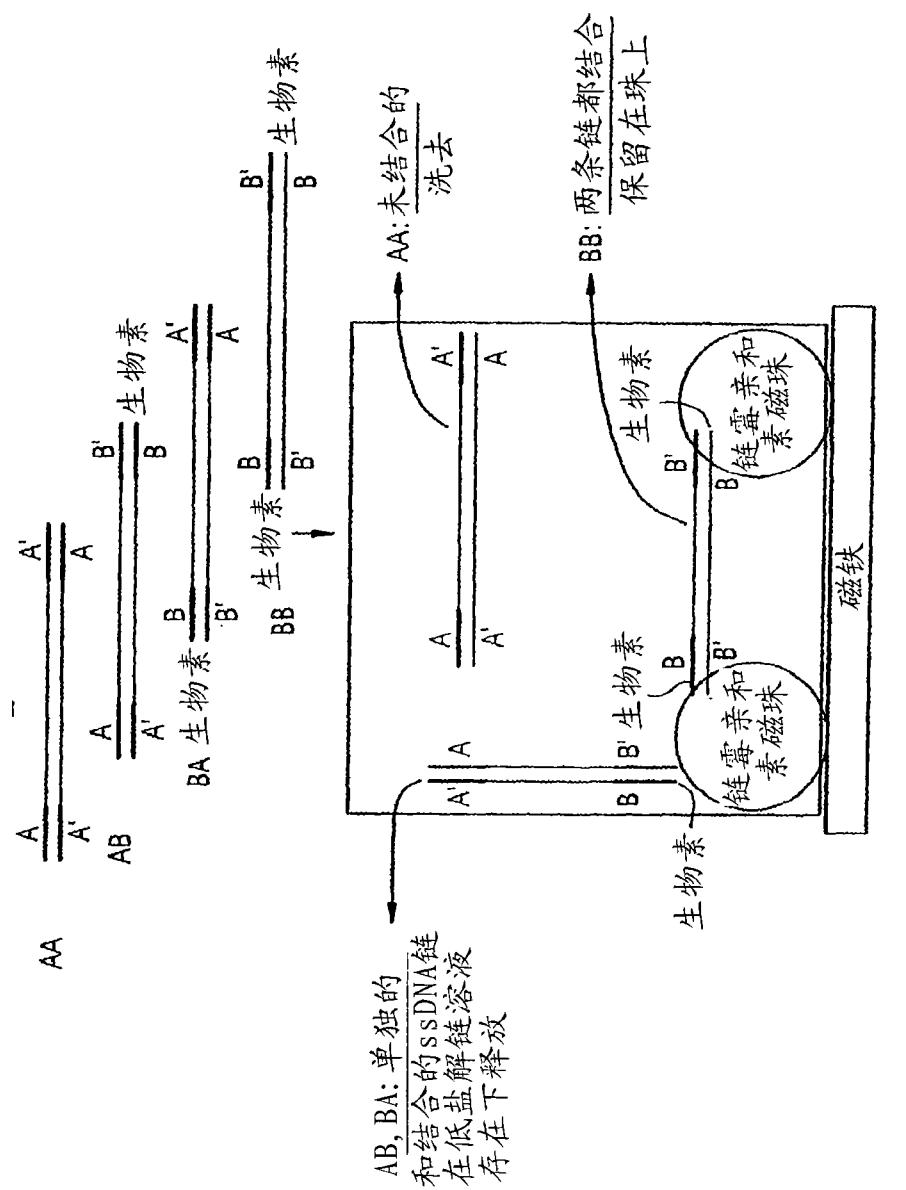


图 4

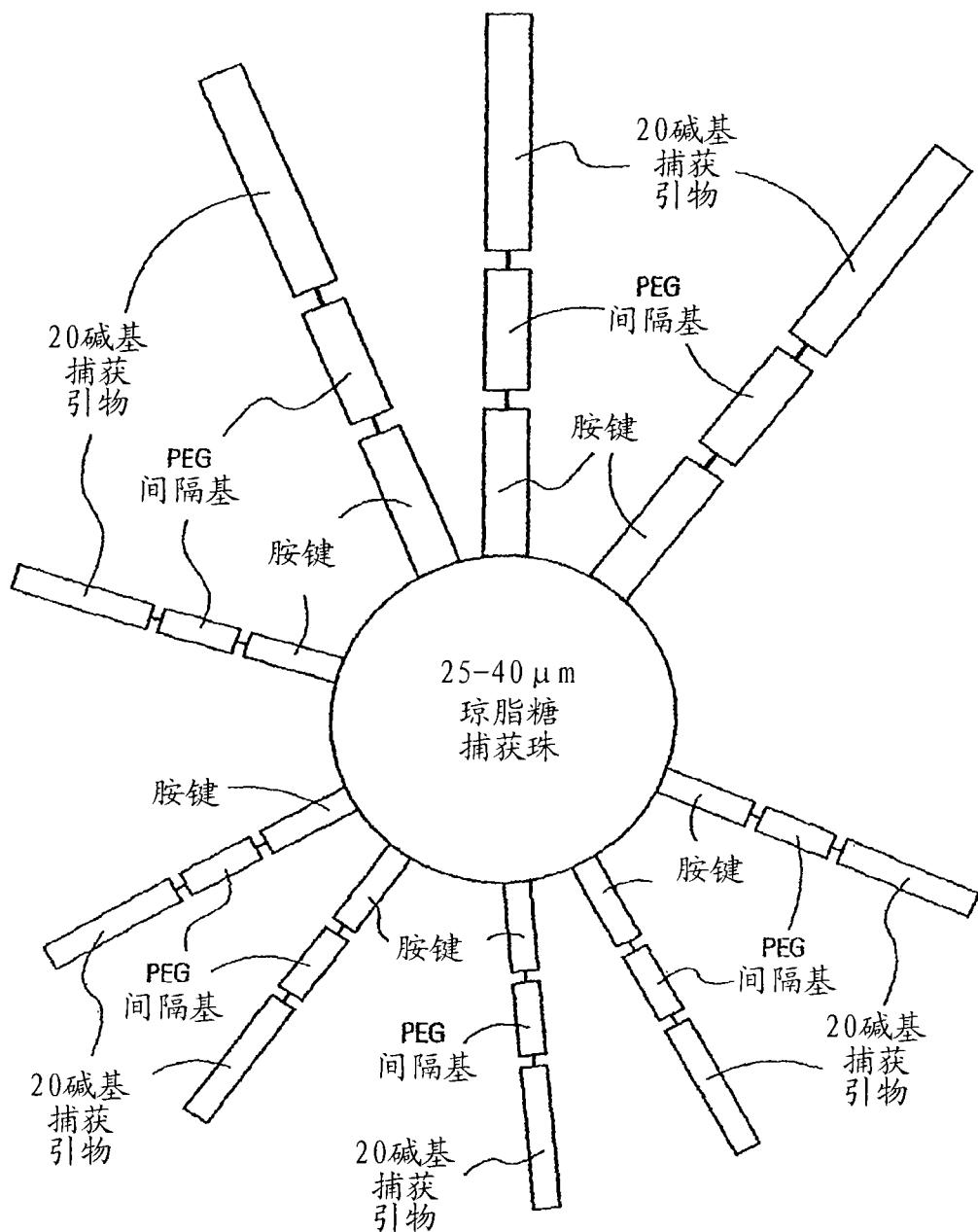


图 5

图 6A

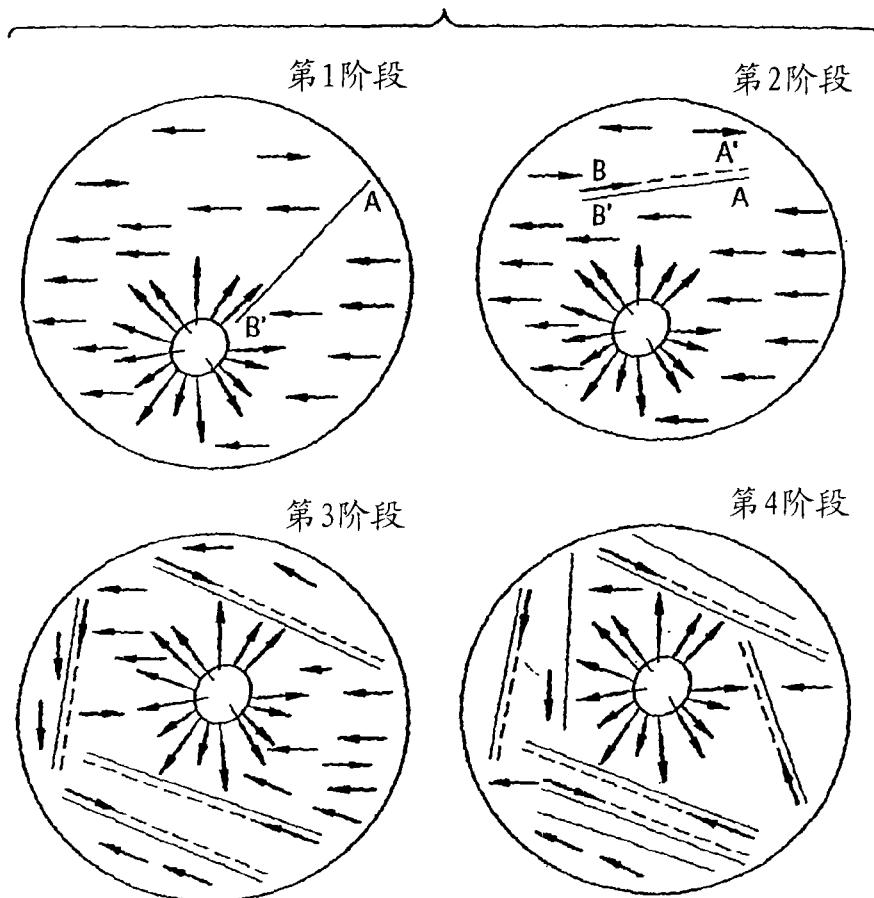
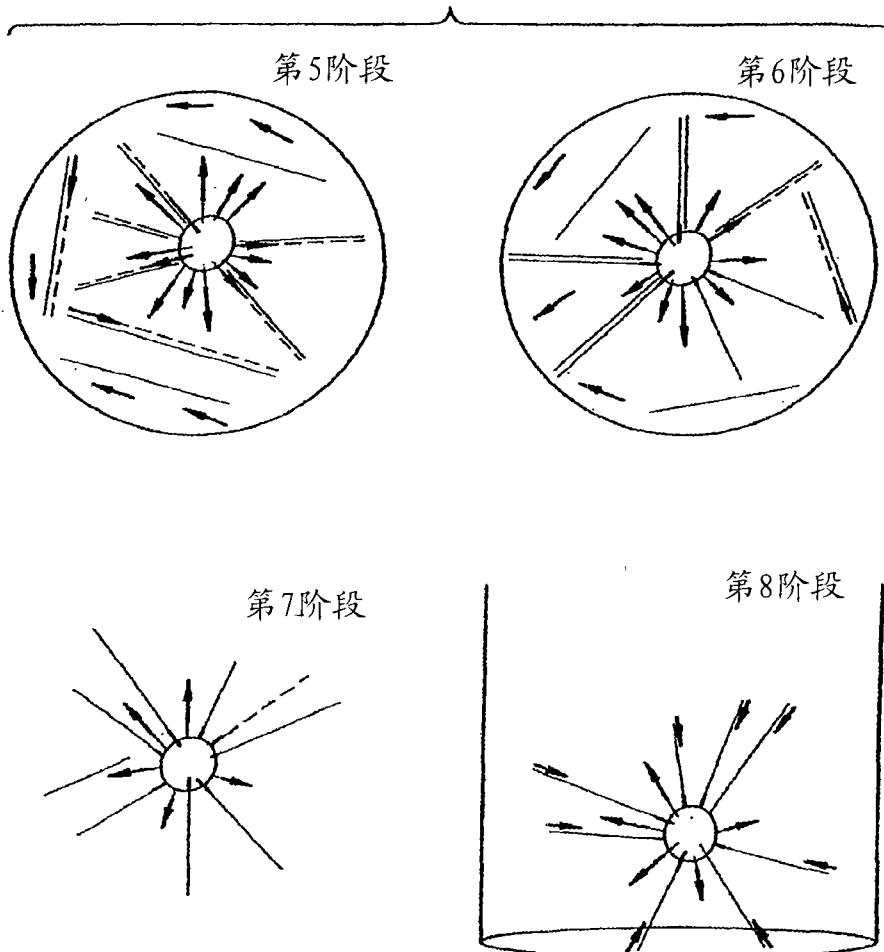


图 6A



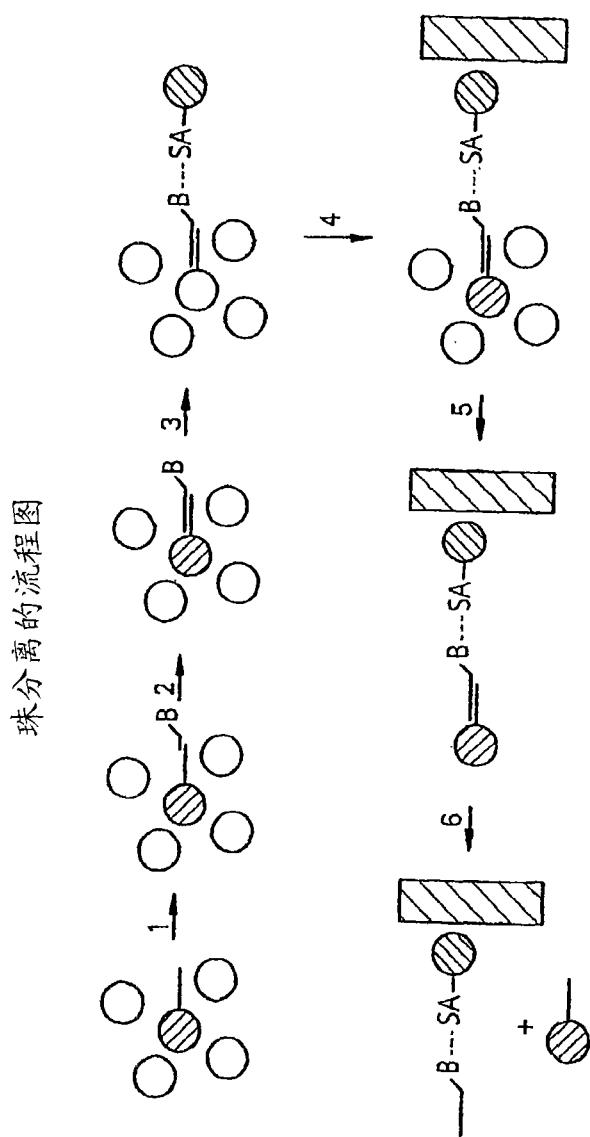


图 7

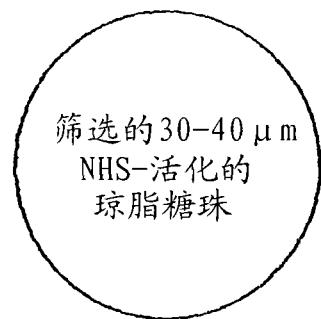


图 8A

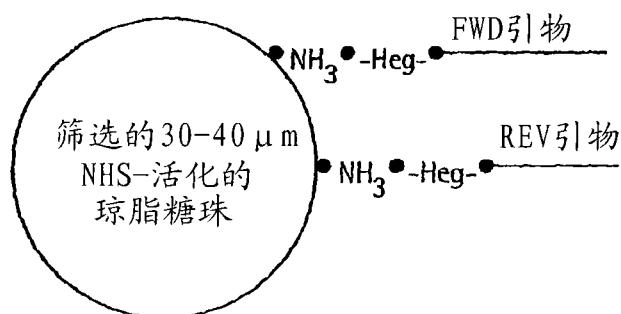


图 8B

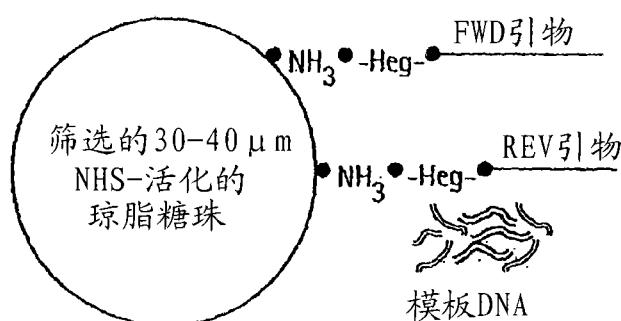


图 8C

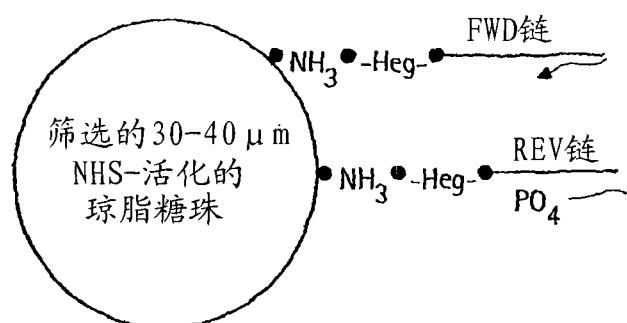


图 8D

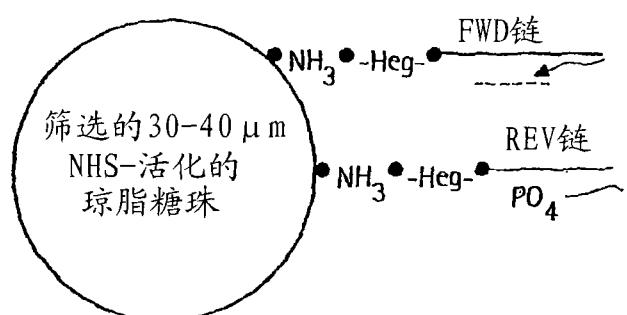


图 8E

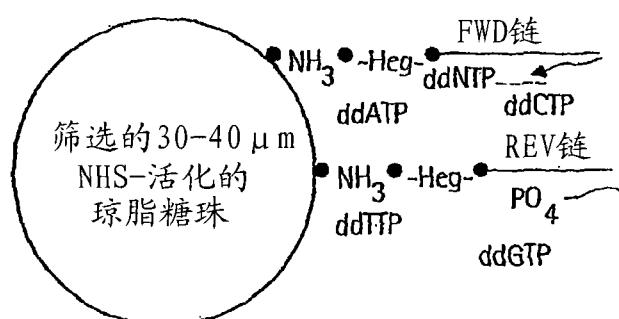
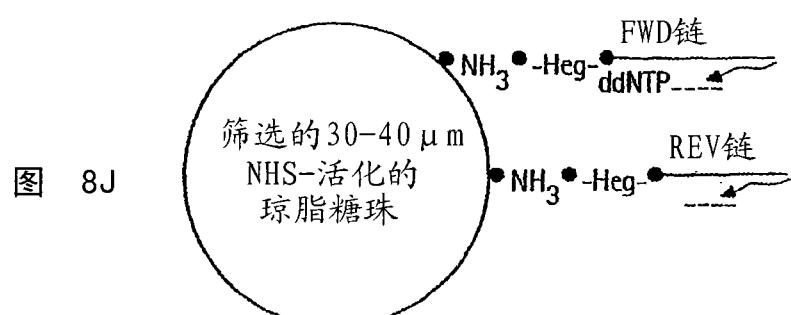
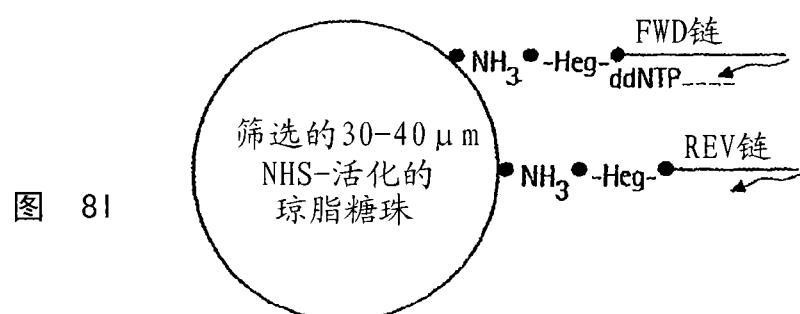
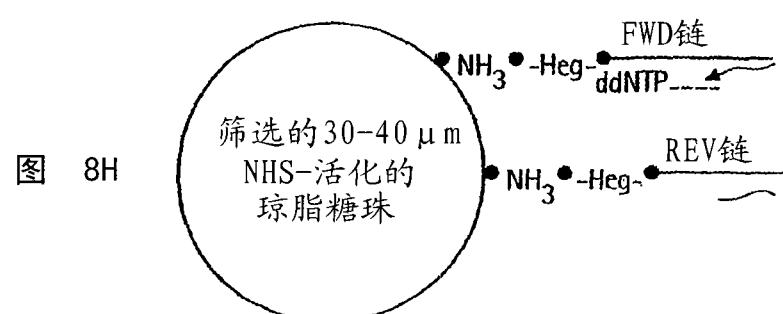
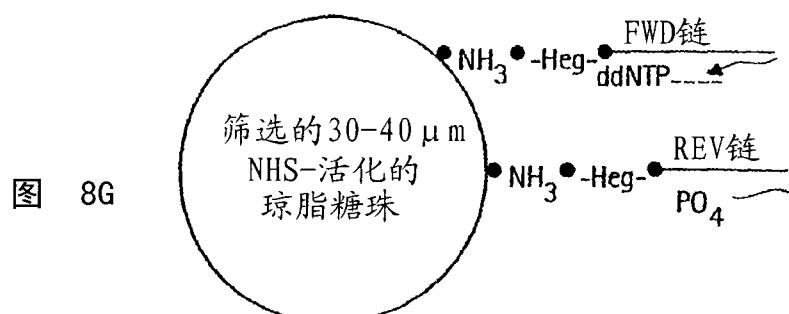


图 8F



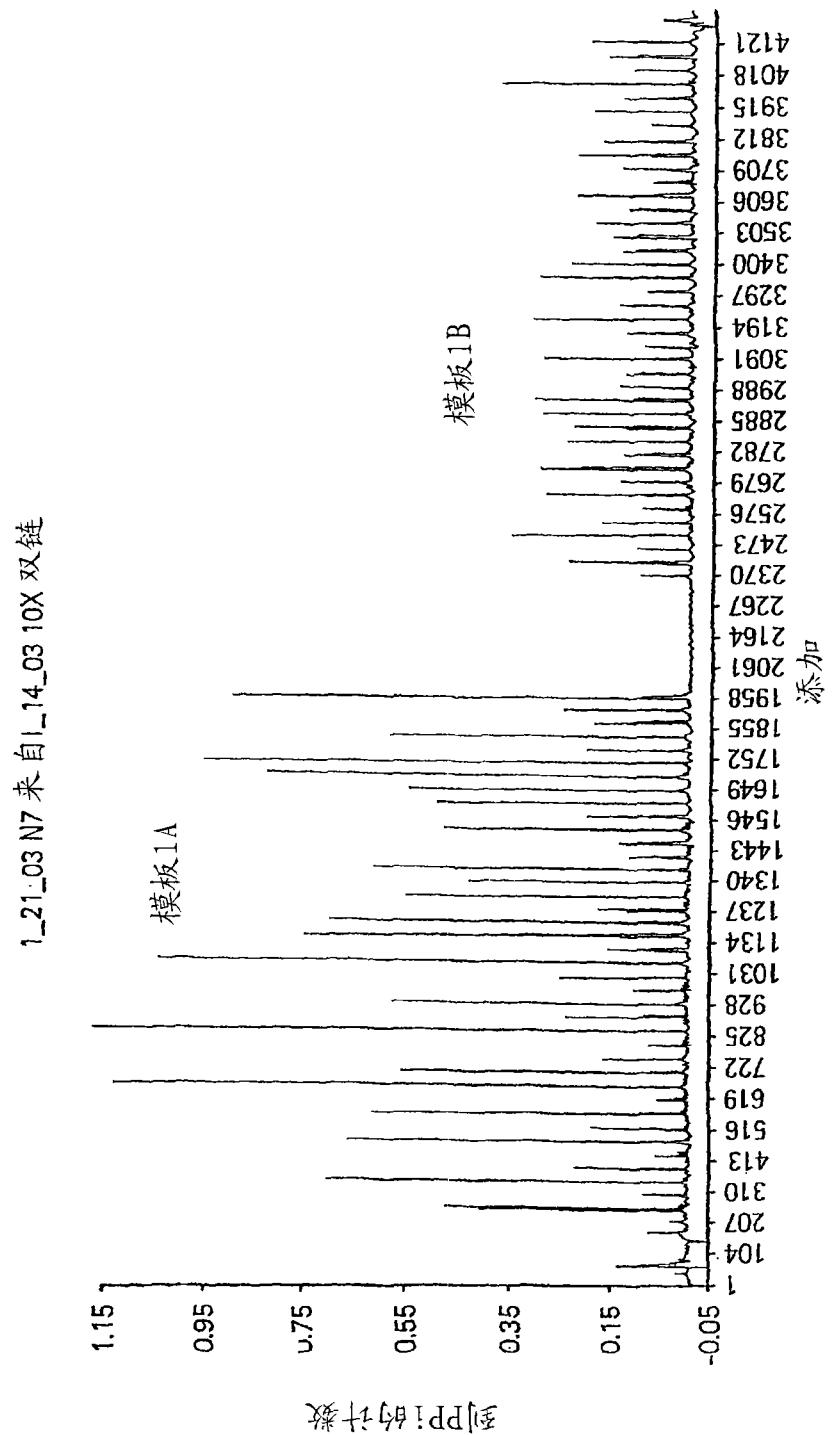


图 9

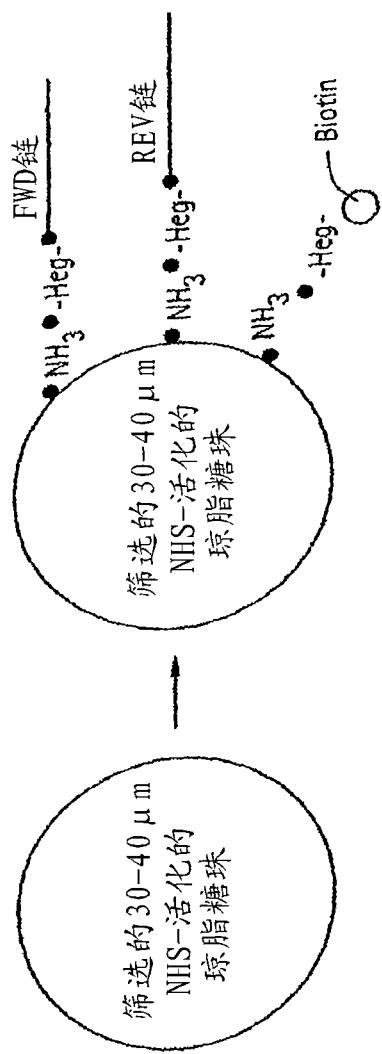


图 10A

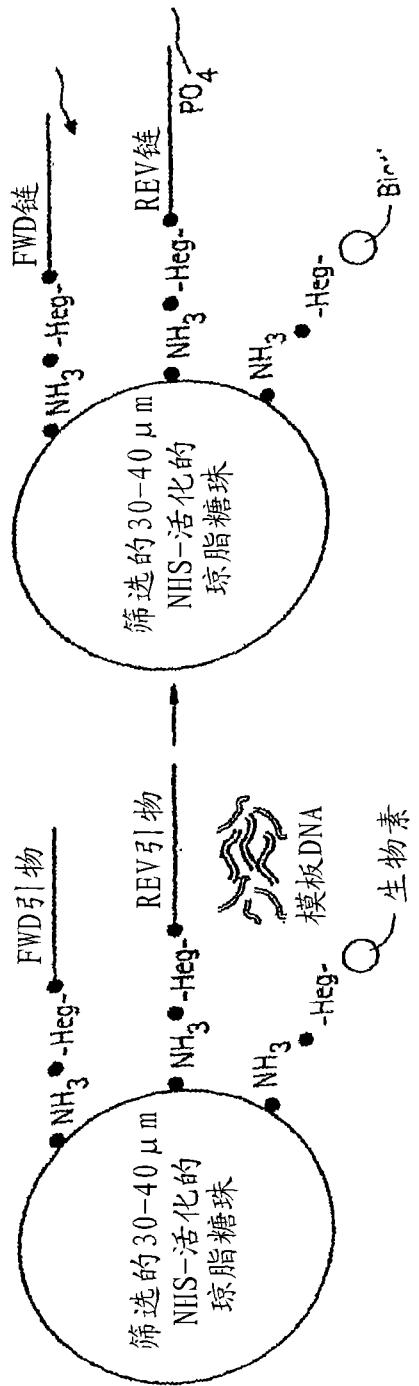


图 10B

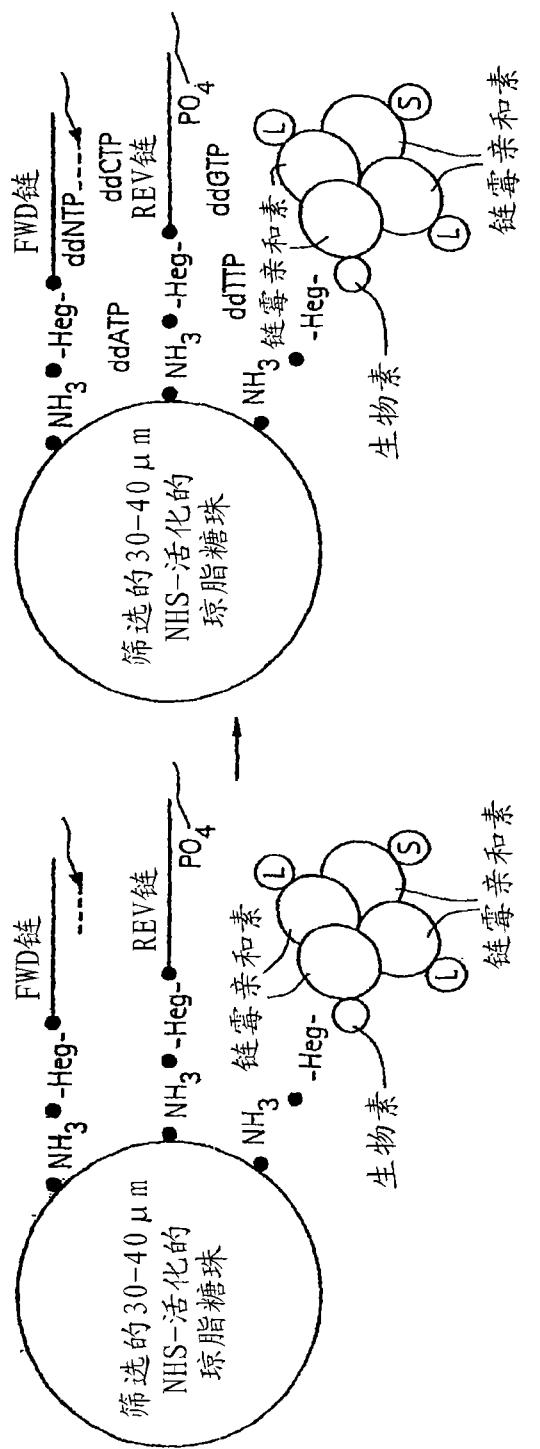


图 10C

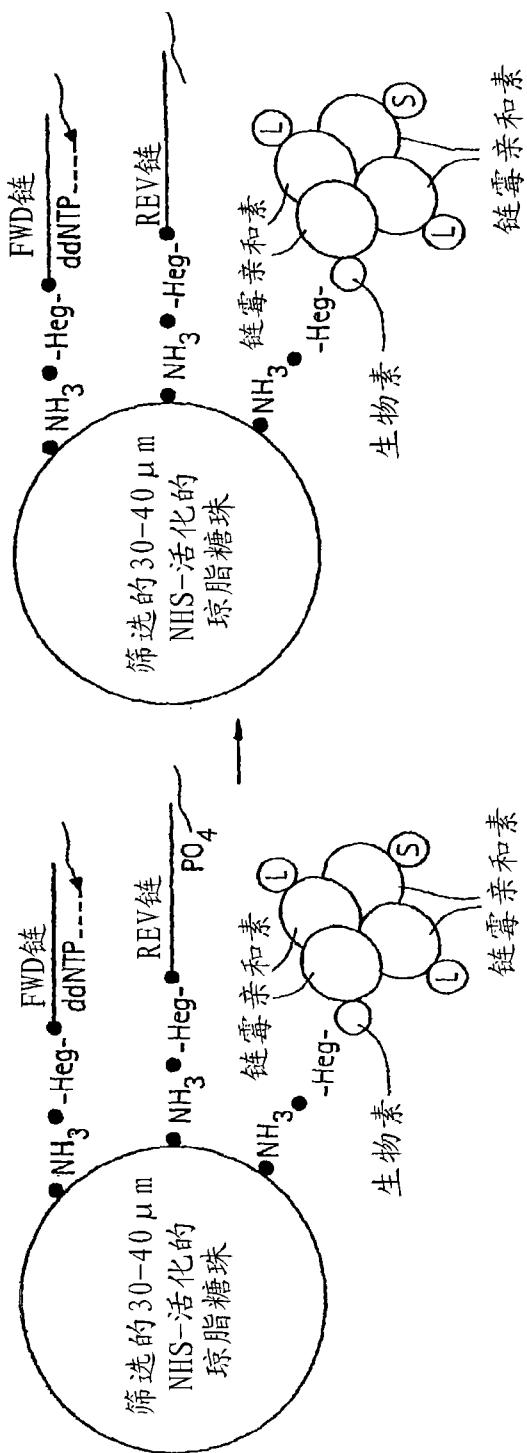


图 10D

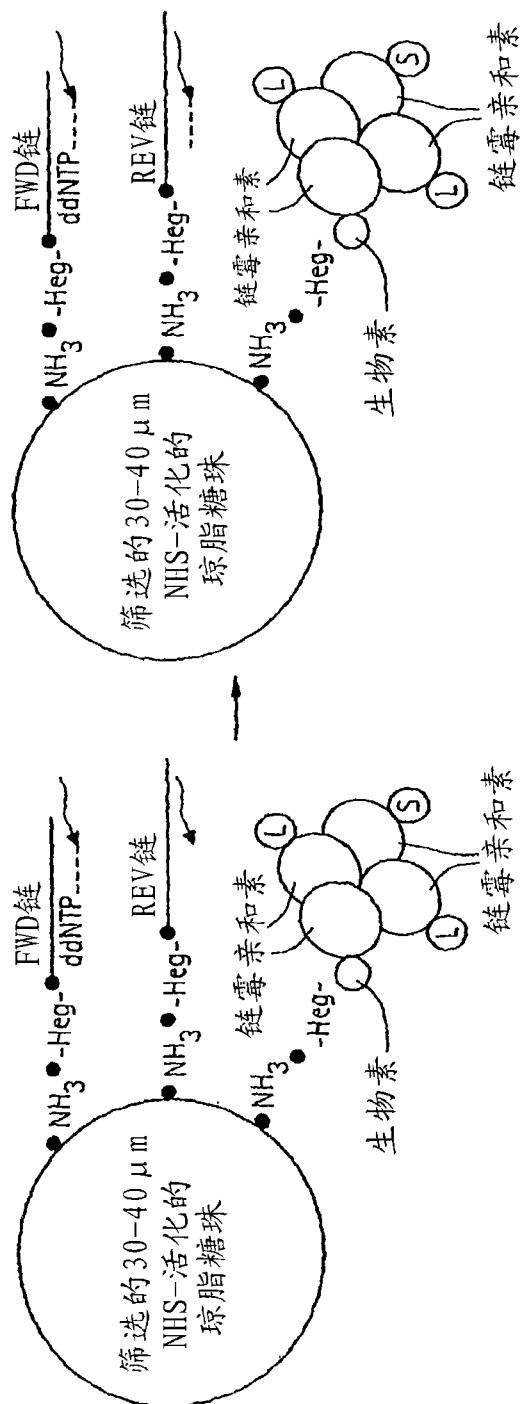
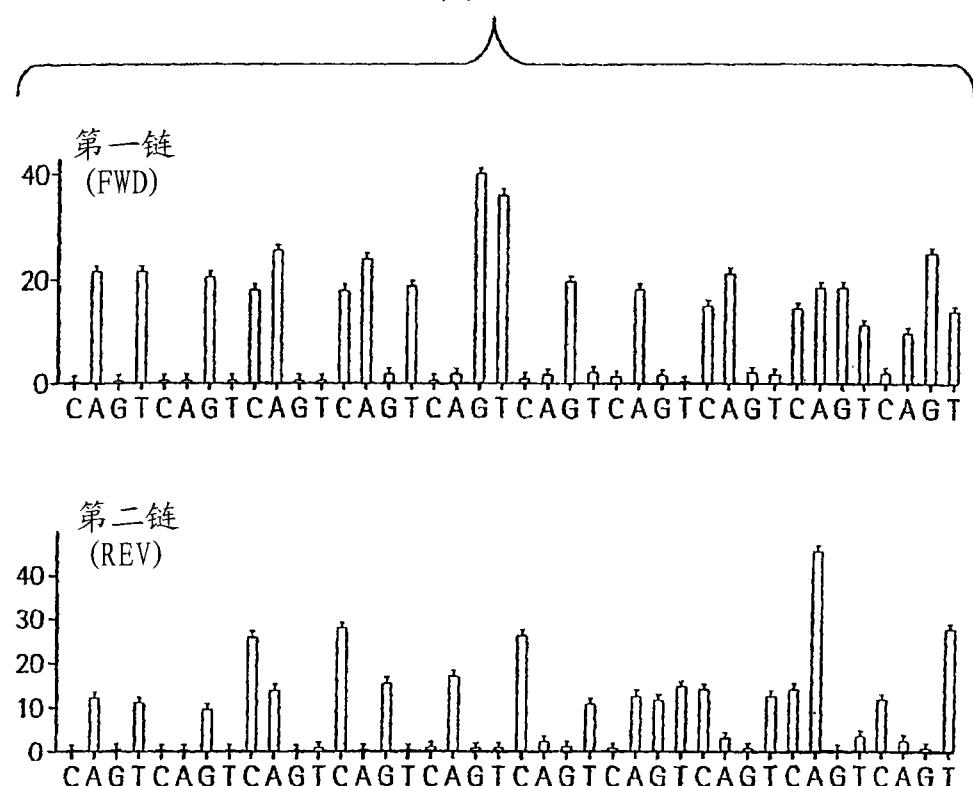


图 10E

图 10F

第一链  
(FWD)

样口	孔位置	序列	完整匹配长度
F6_14_1	00001_1362_1660.孔	ATGCACATGGTTGACACAGTGGT	22

ATGC ACATGGTTGACACAGTGG

ATGC CACCGACCTAGTCACMACT

第二链  
(REV)

样口	孔位置	序列	完整匹配长度
F6_14_1	00003_1363_1660.孔	ATGCAACCGACCTAGTCACAACTT	25

图 11A

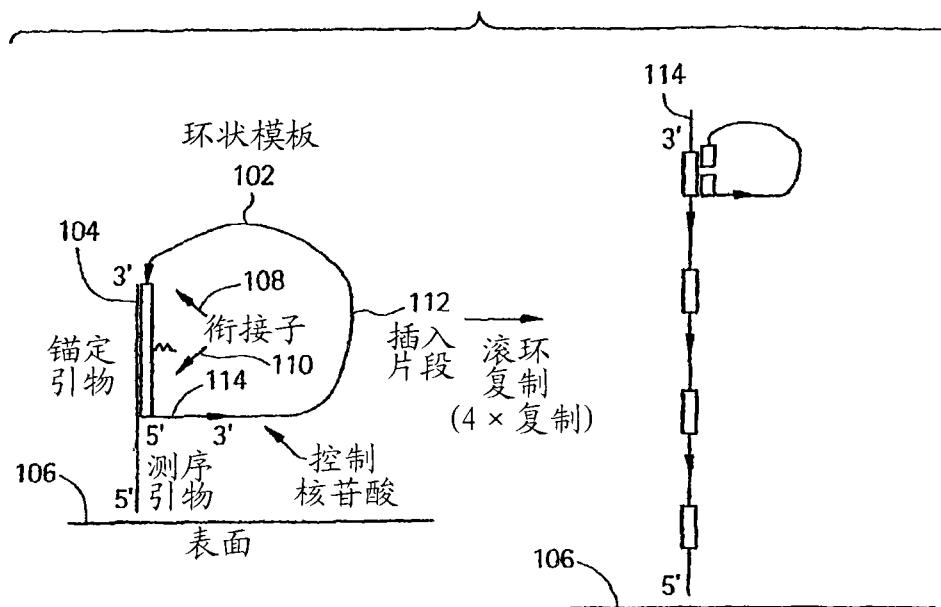
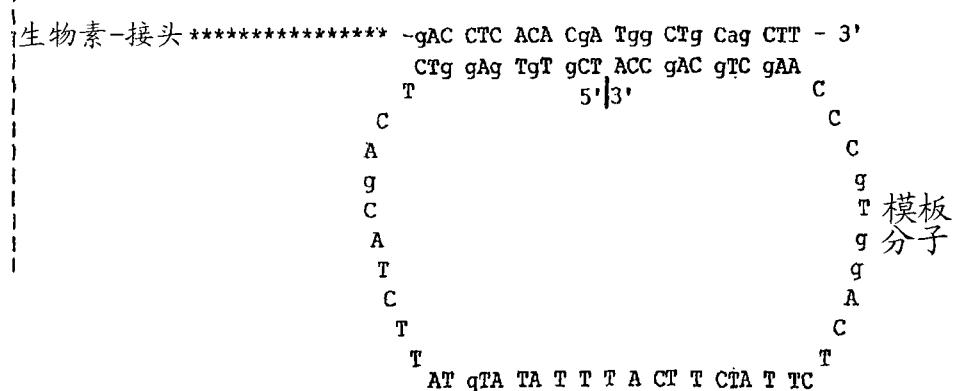


图 11B



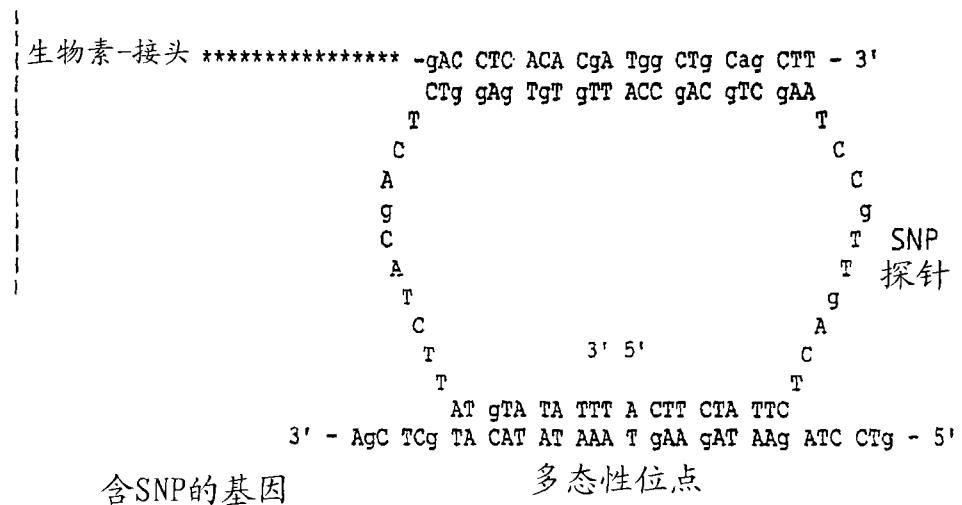
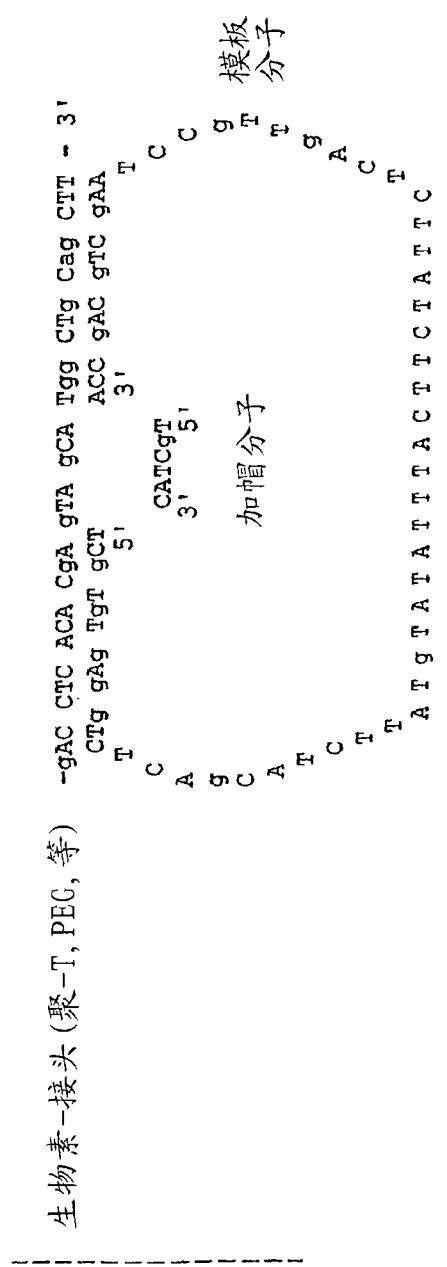


图 11C



11D

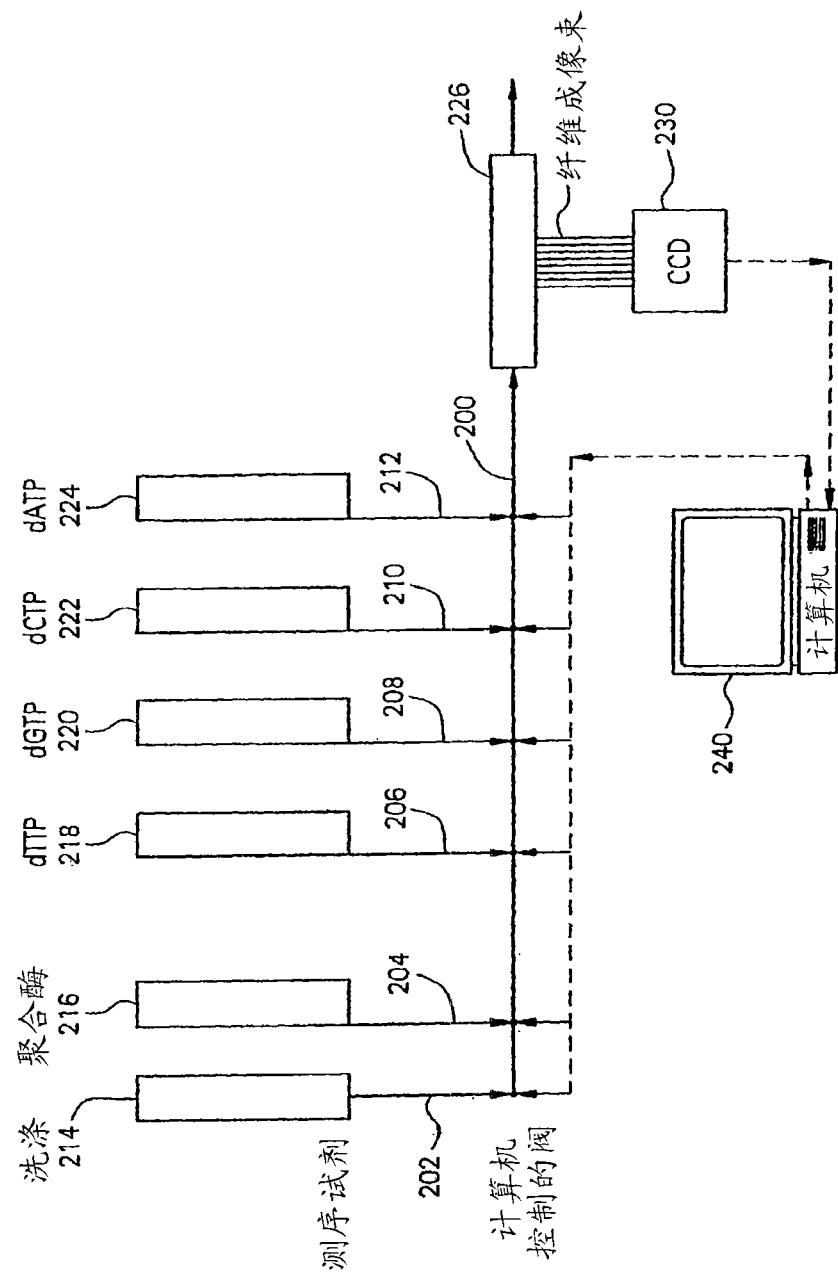


图 12

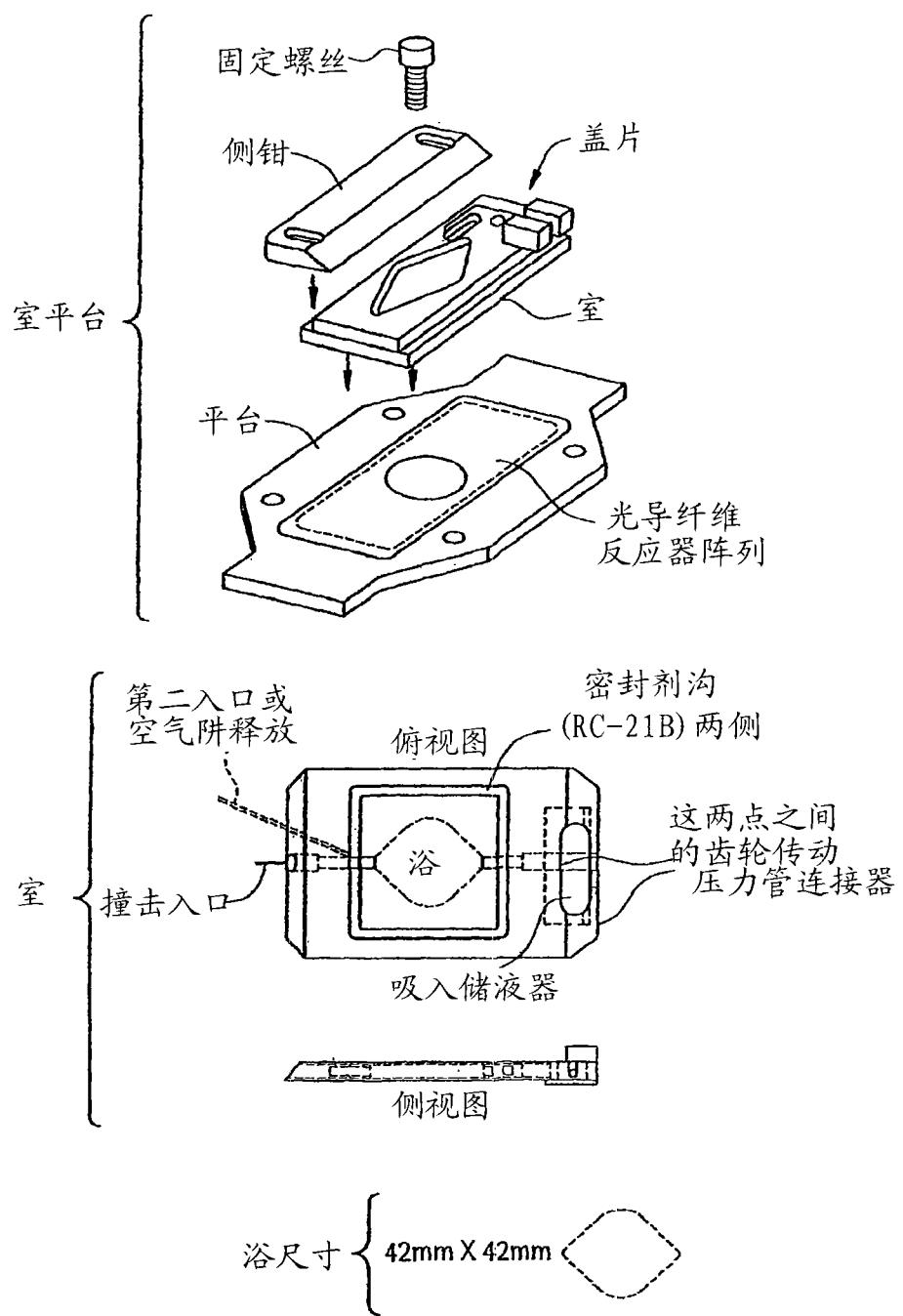
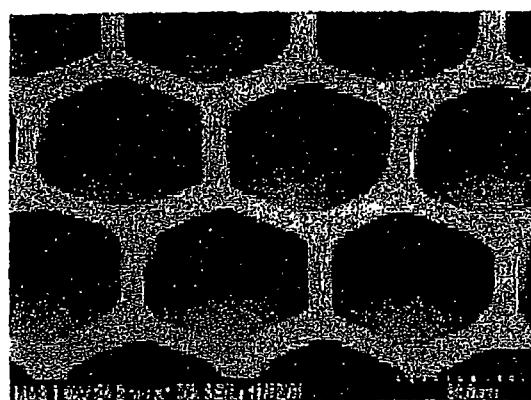
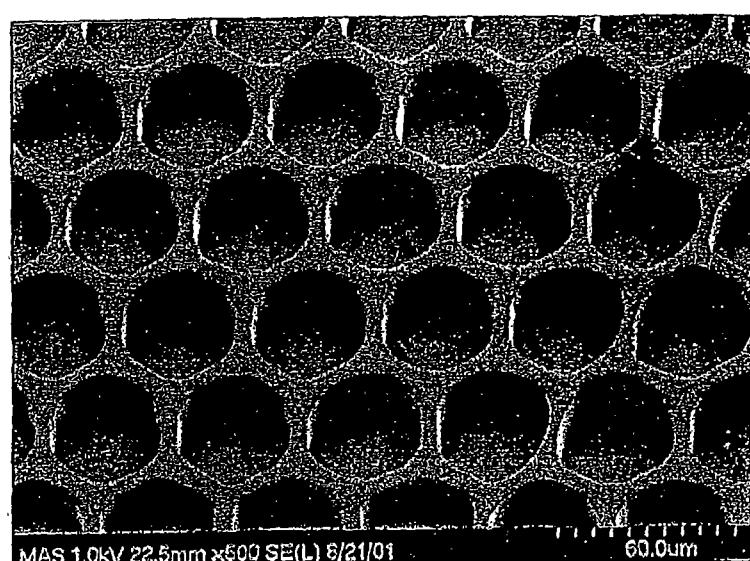


图 13

图 14



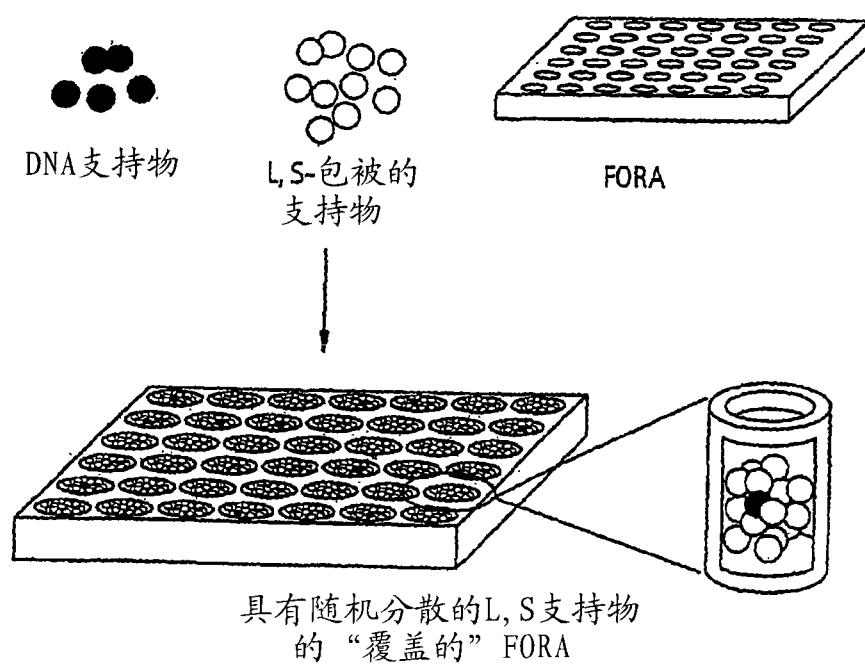


图 15

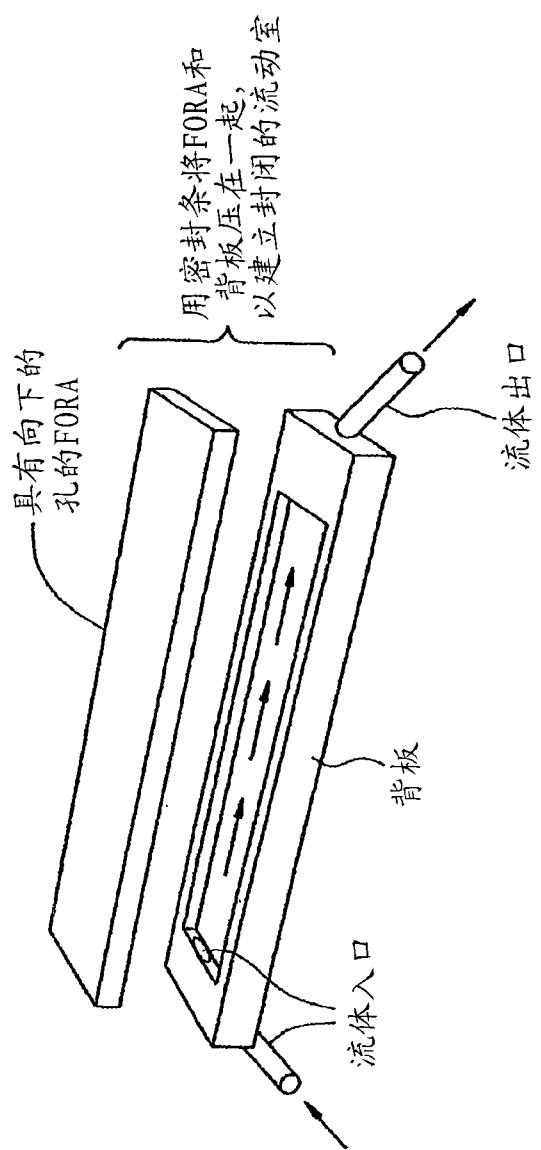


图 16

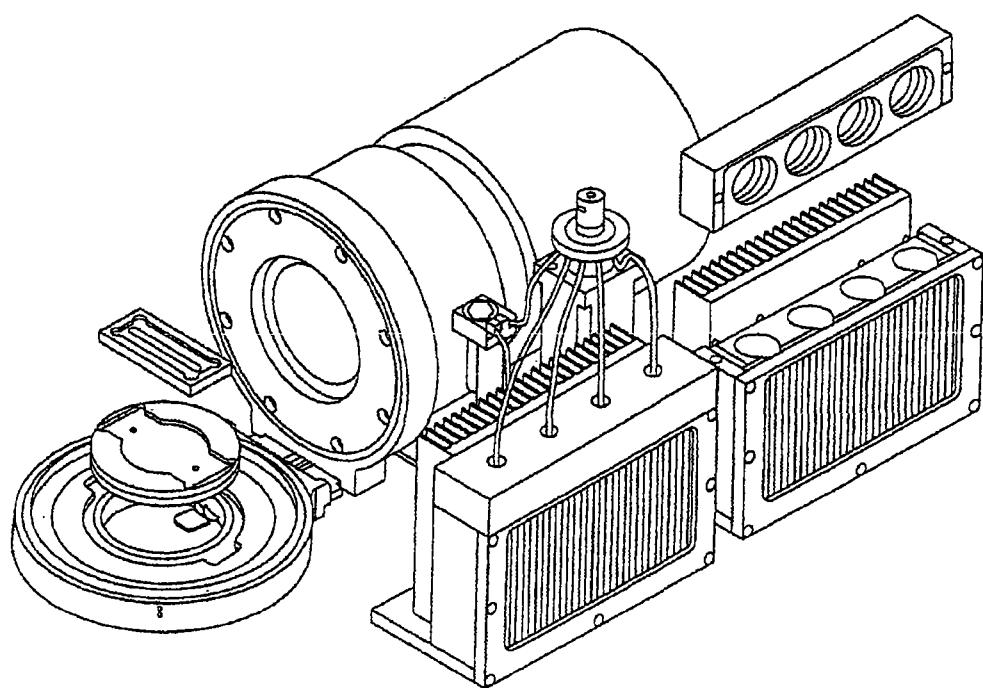
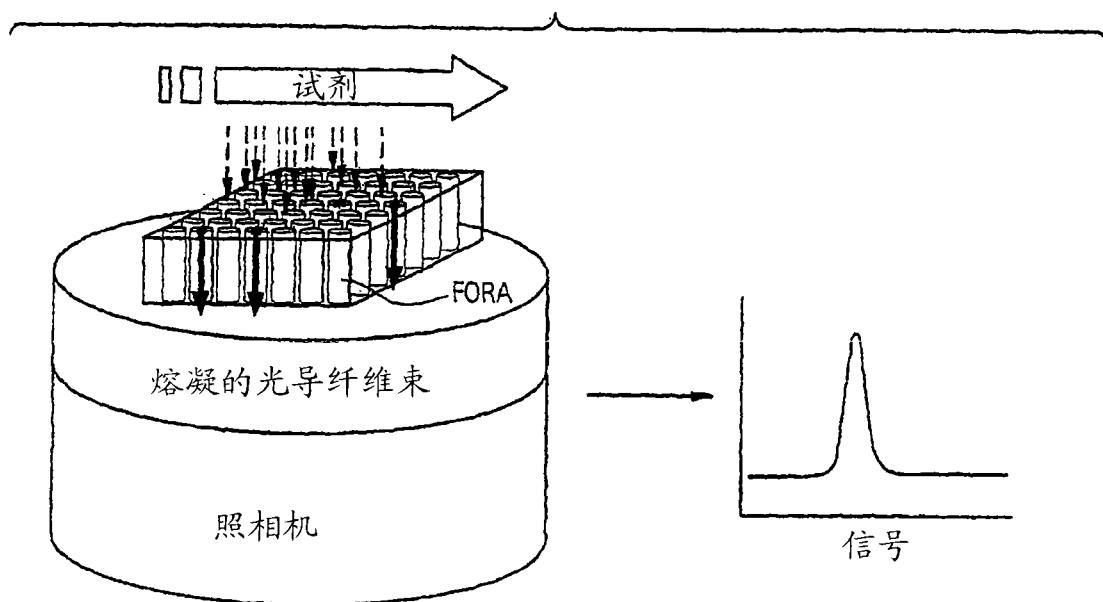


图 17

图 18



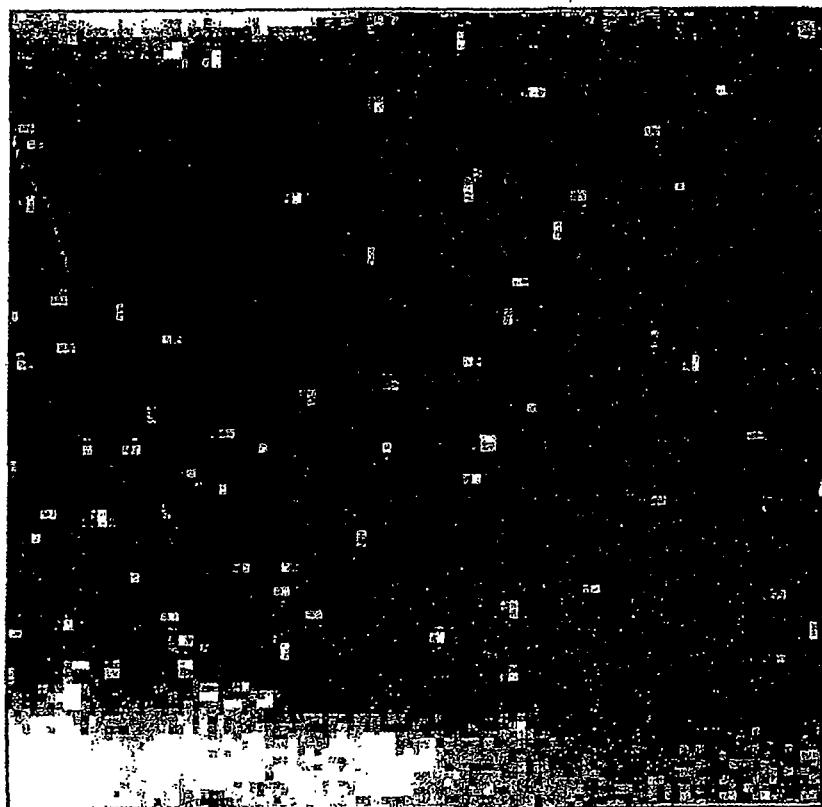


图 19

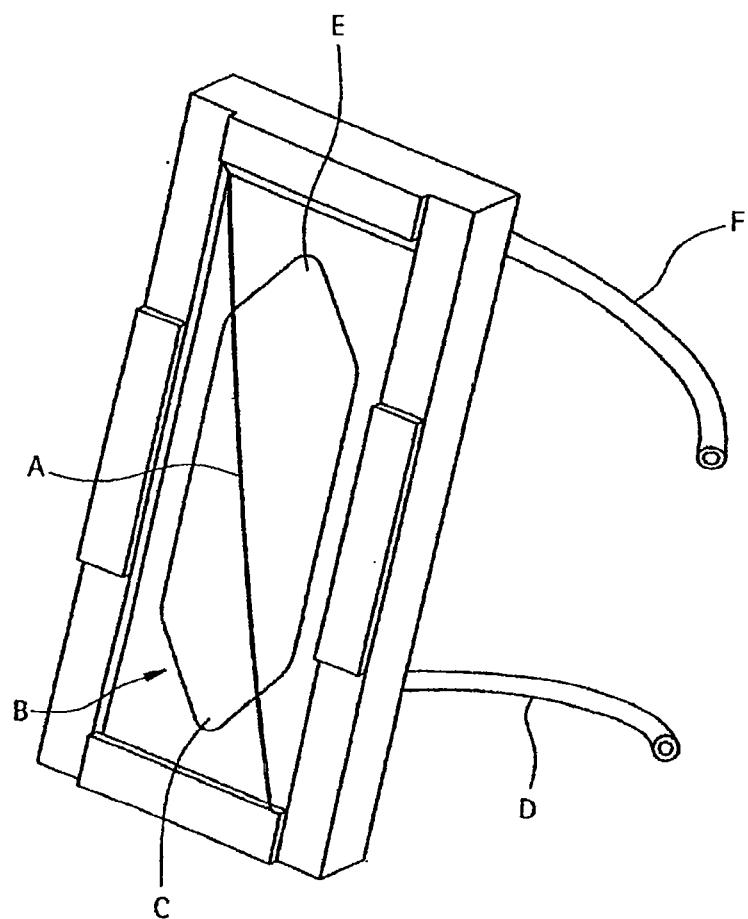


图 20

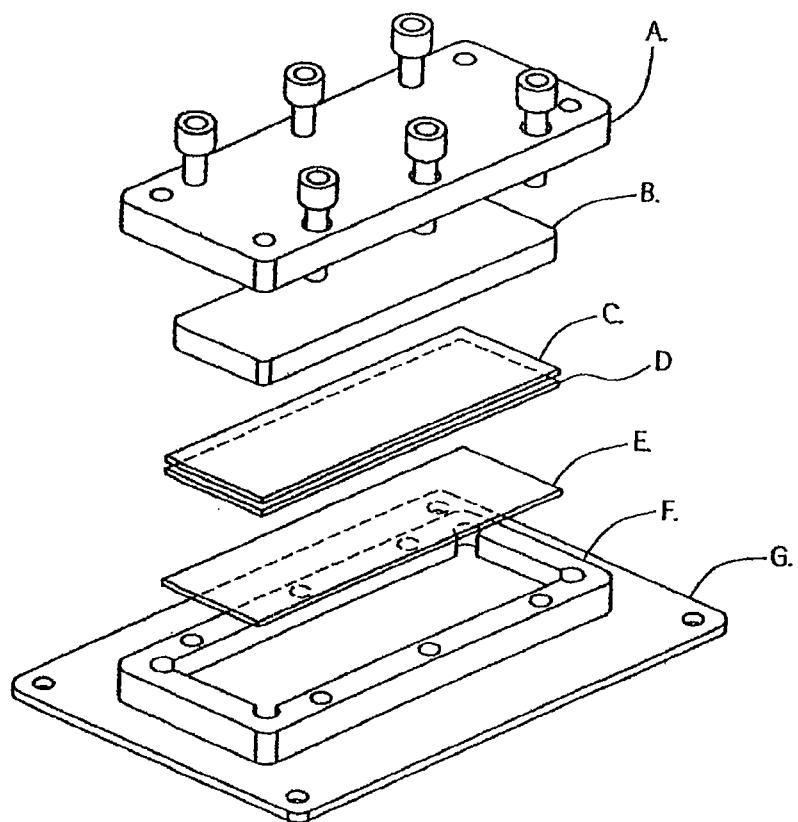
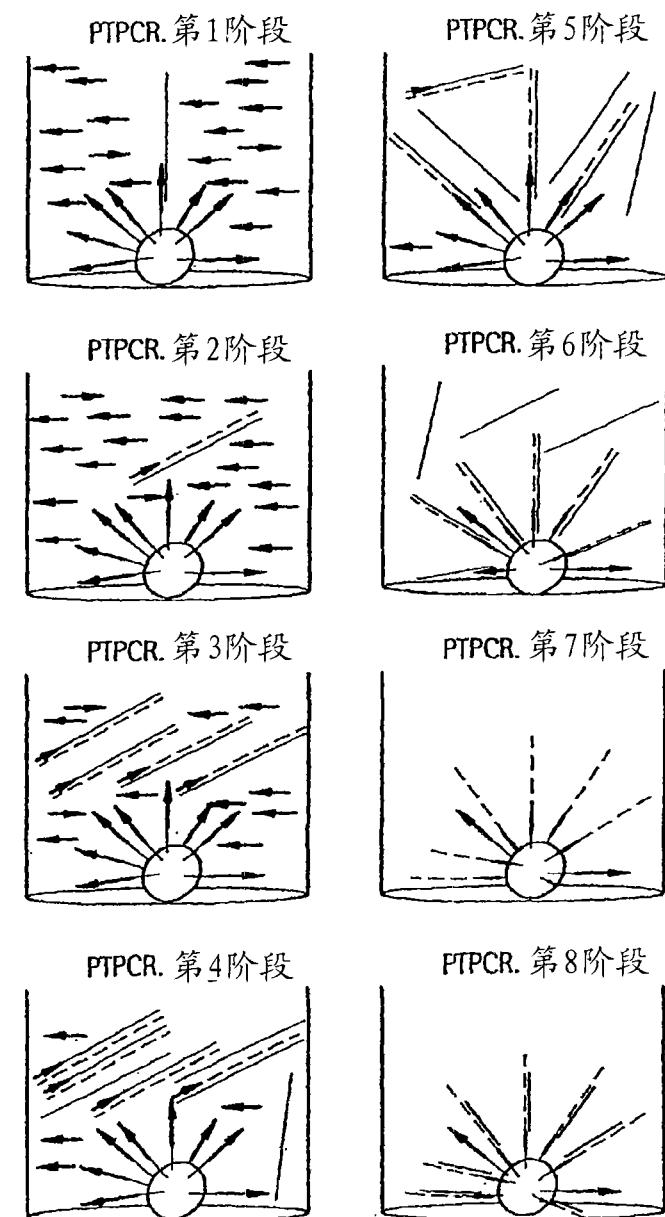


图 21

图 22



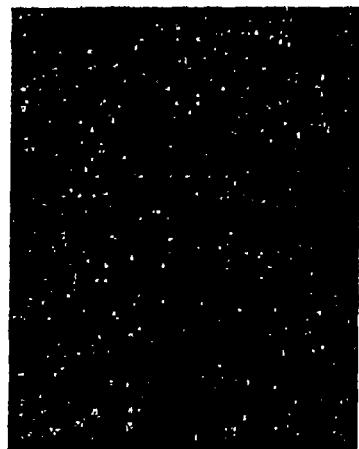


图 23A

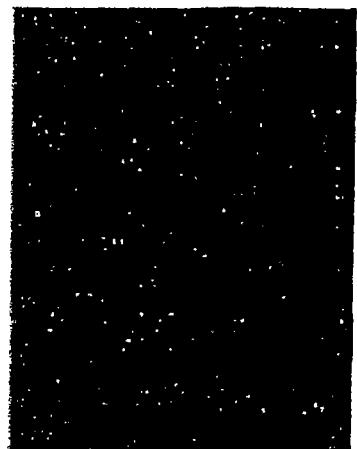


图 23B



图 23C

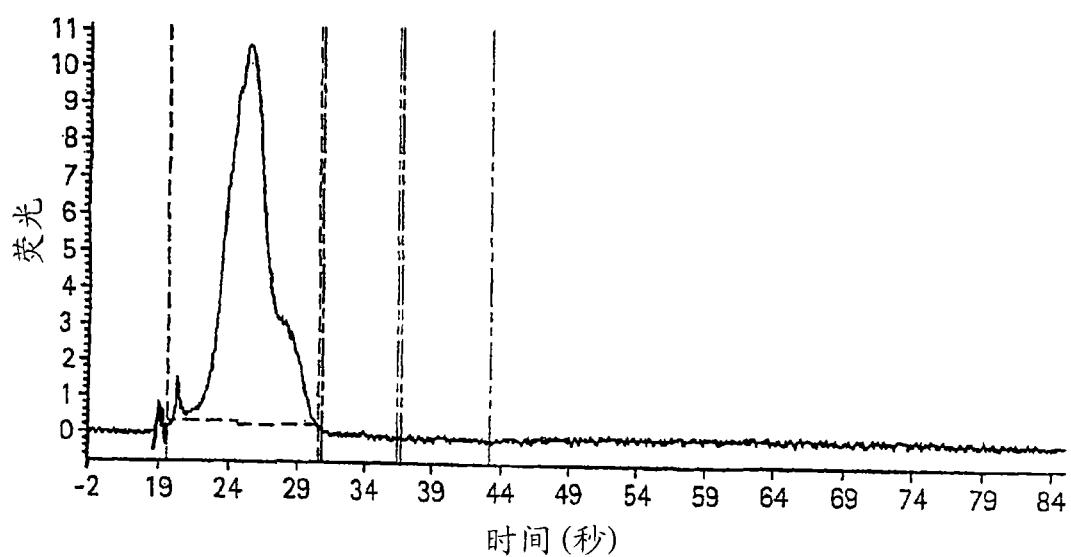
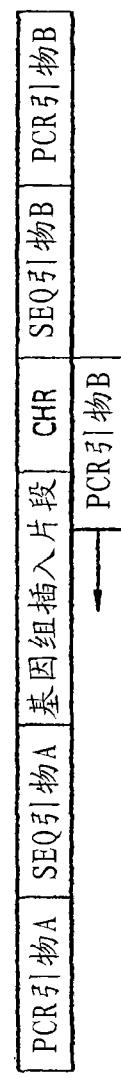


图 24

图 25



图 26



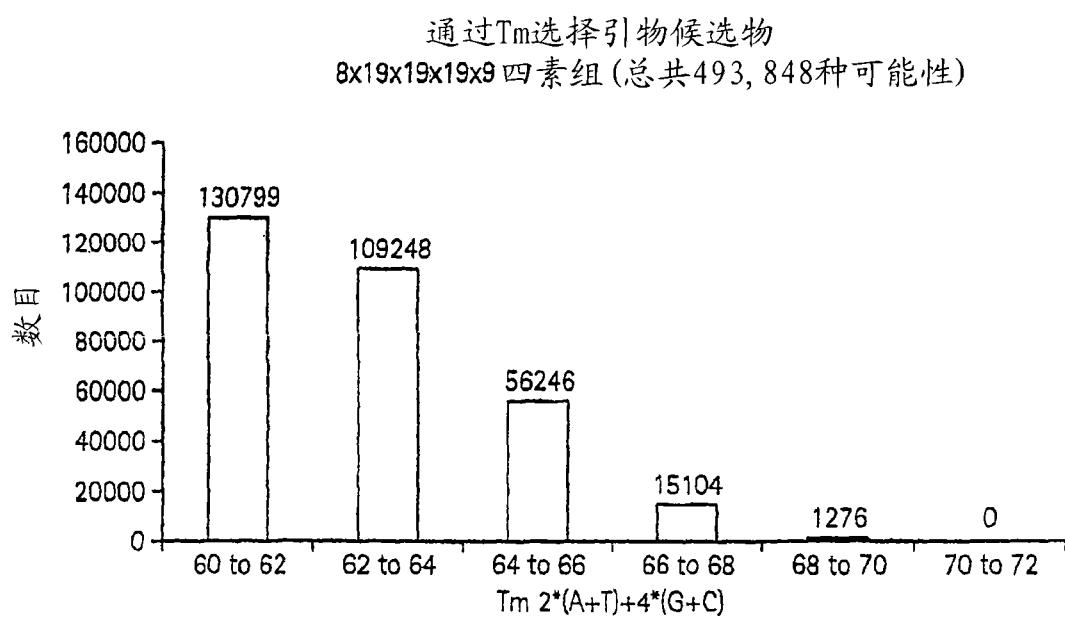


图 27



图 28D



图 28C



图 28B

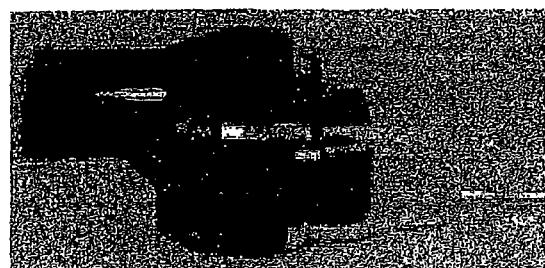


图 28A

FIG. 29A

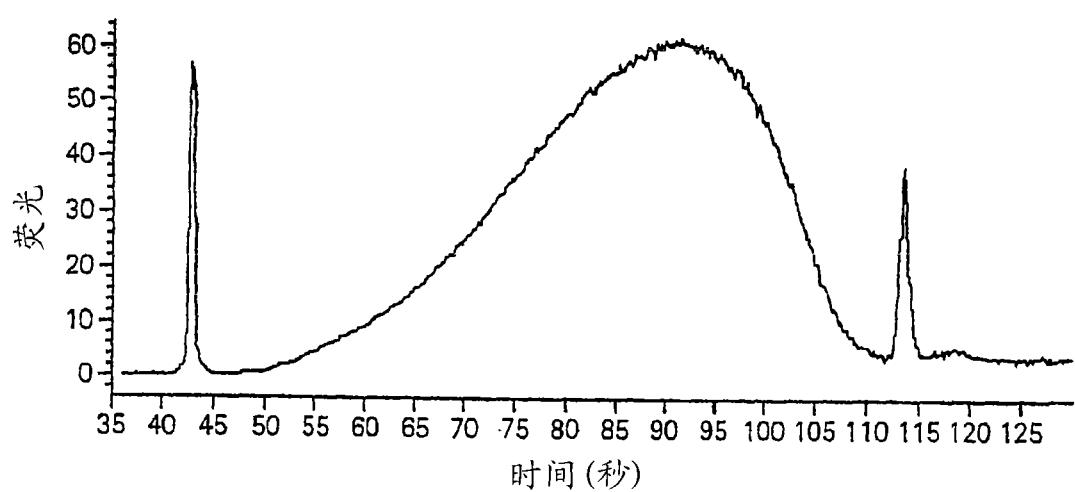


图 29A

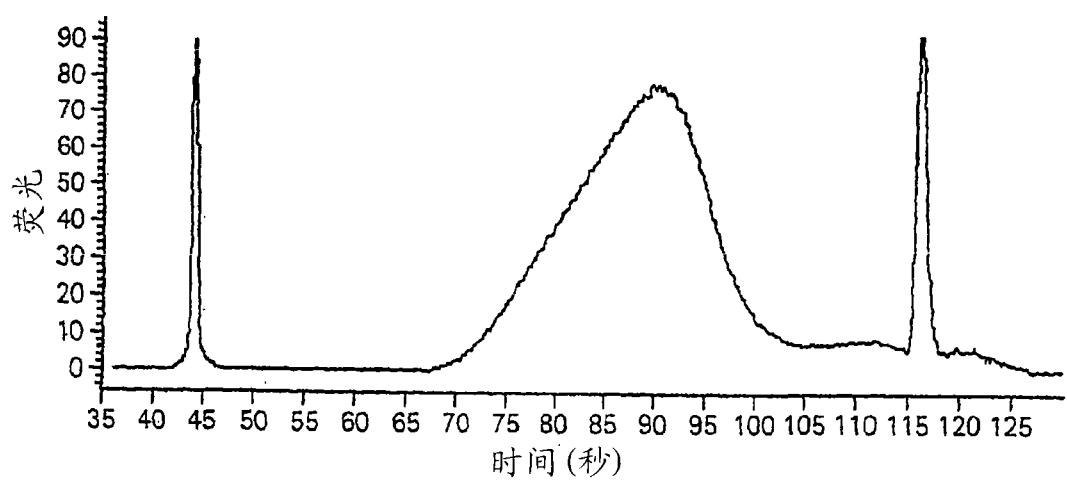


图 29B

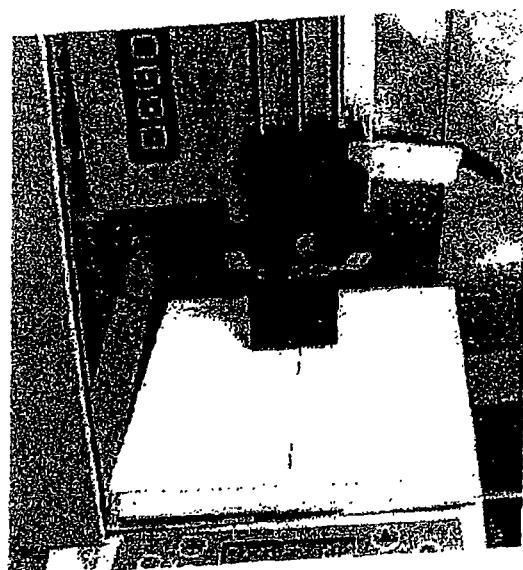


图 30

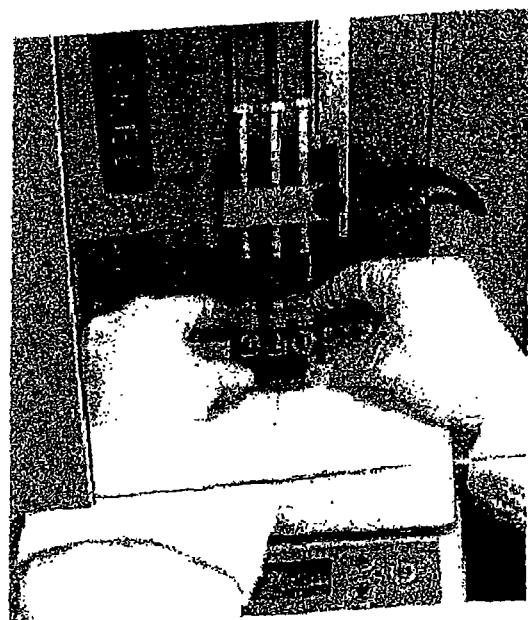


图 31

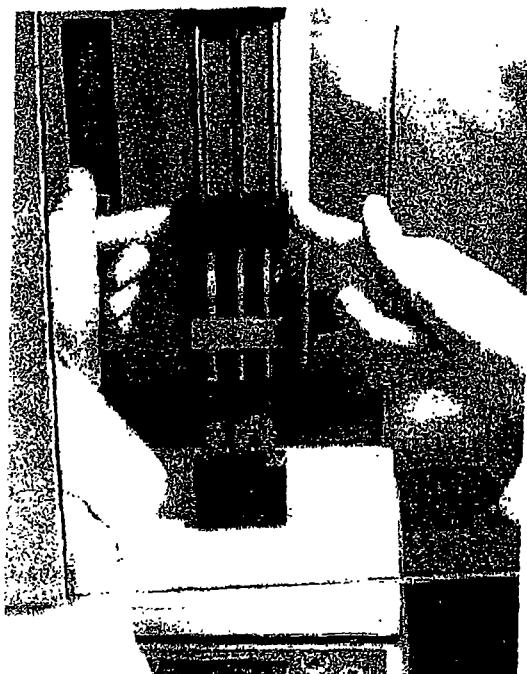


图 32

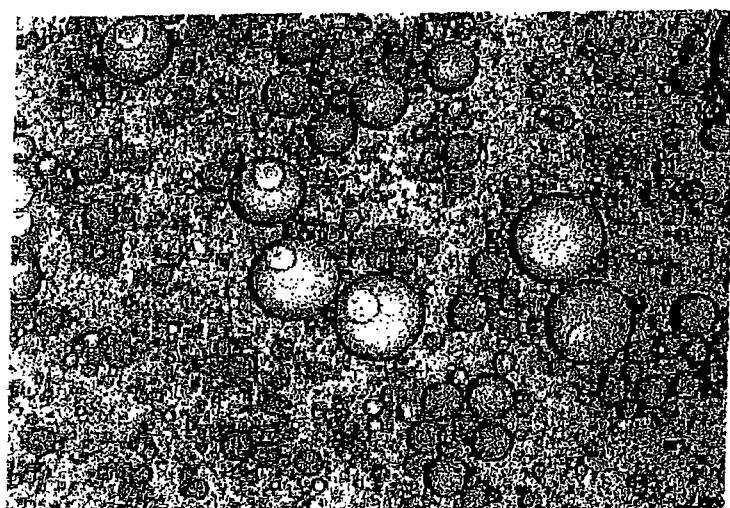
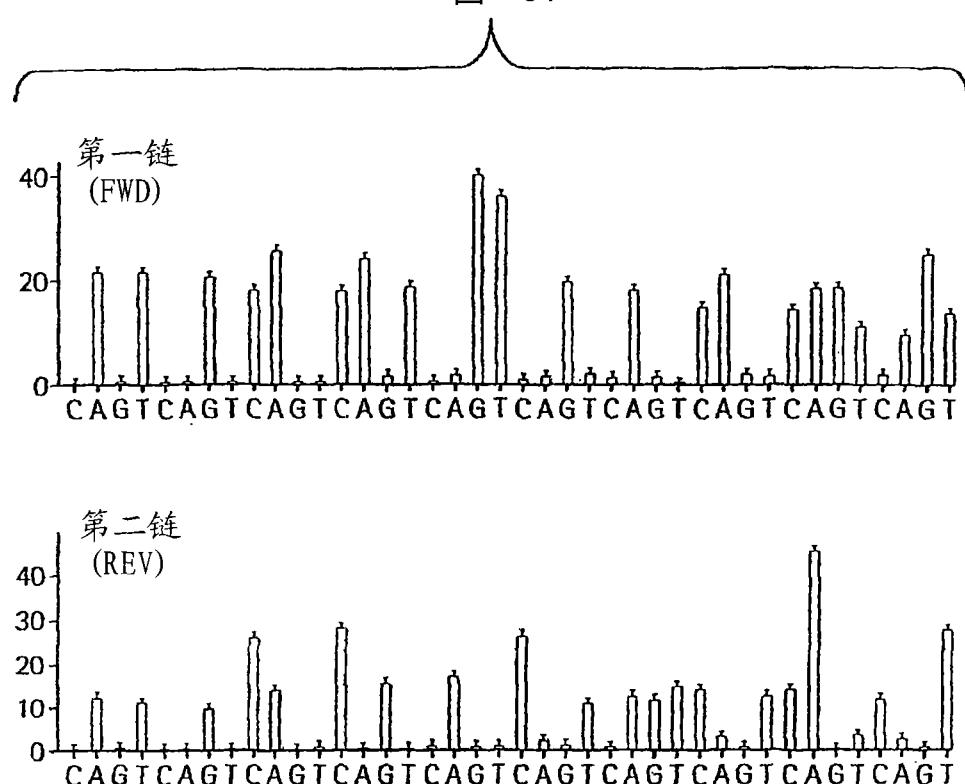


图 33

冬 34



第一链 (FWD)	样口	孔位置	序列	完整匹配长度
	F6_14_1	00001_1362_1660. 孔	ATGCACATGGTGTGACACAGTGTT	22

ATGC ACATGGTGTGACACAGTGG

ATGC CACCGACCTAGTCTGAACTT

第二链 (REV)	样口	孔位置	序列	完整匹配长度
	F6_14_1	00003_1363_1660.孔	ATGCCAACGACCTAGTCCTAAACTT	25



图 35A

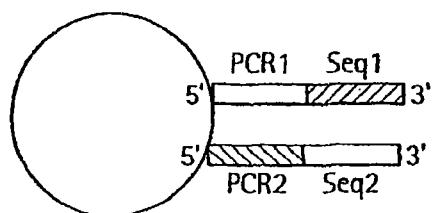


图 35B

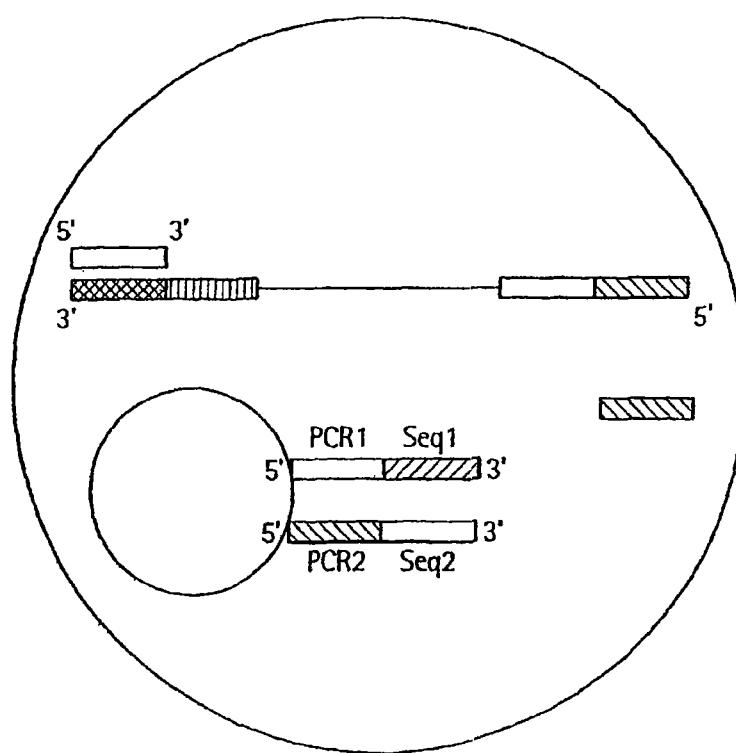


图 35C

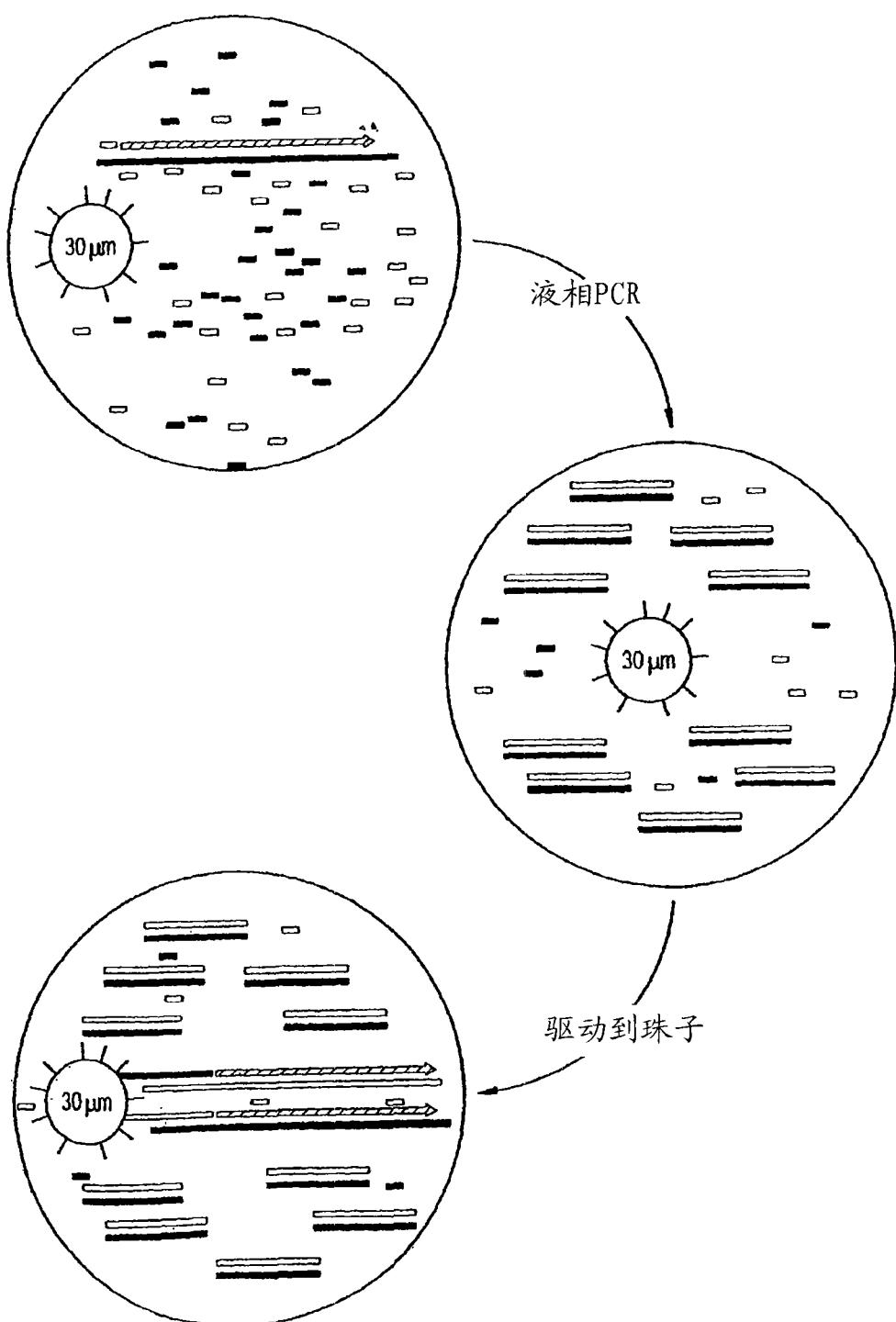


图 36

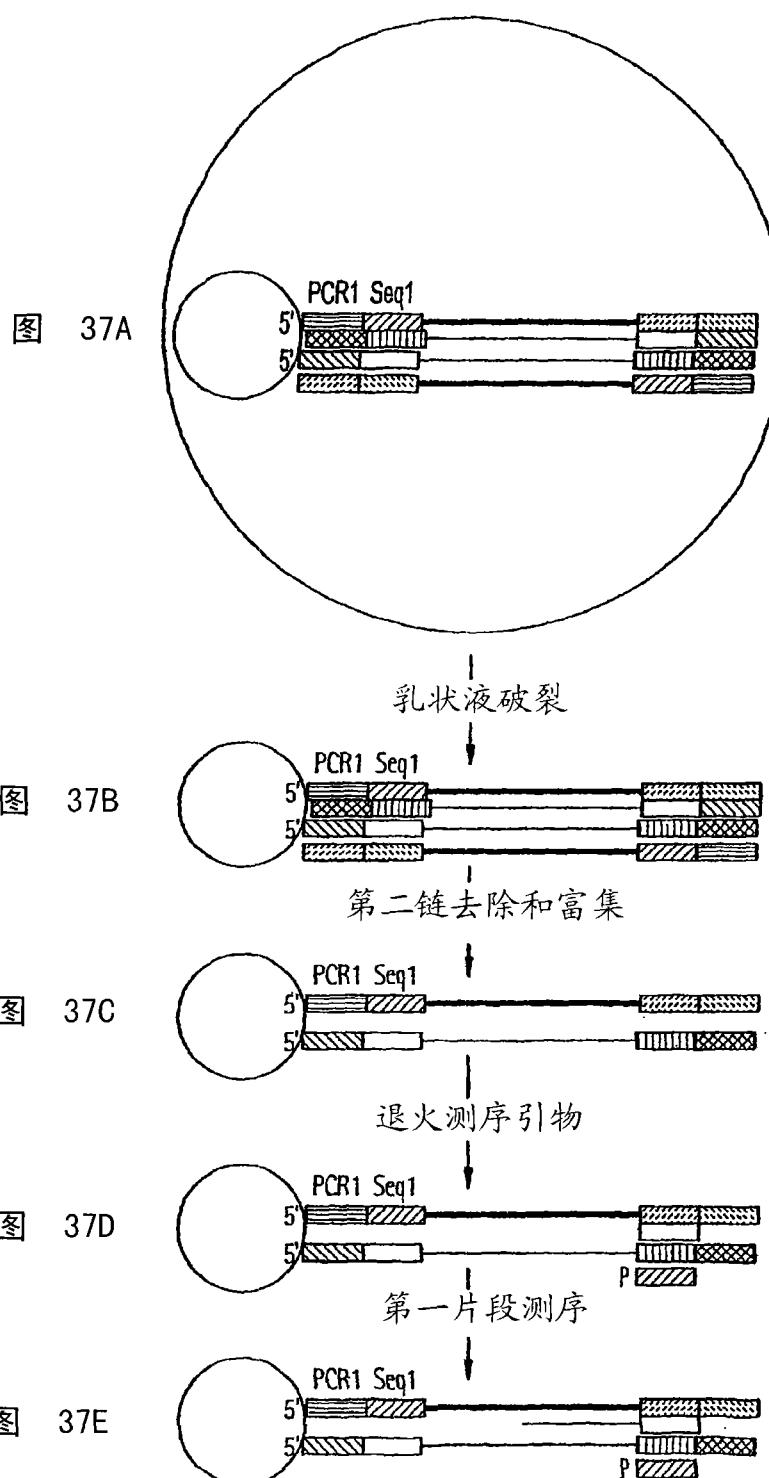
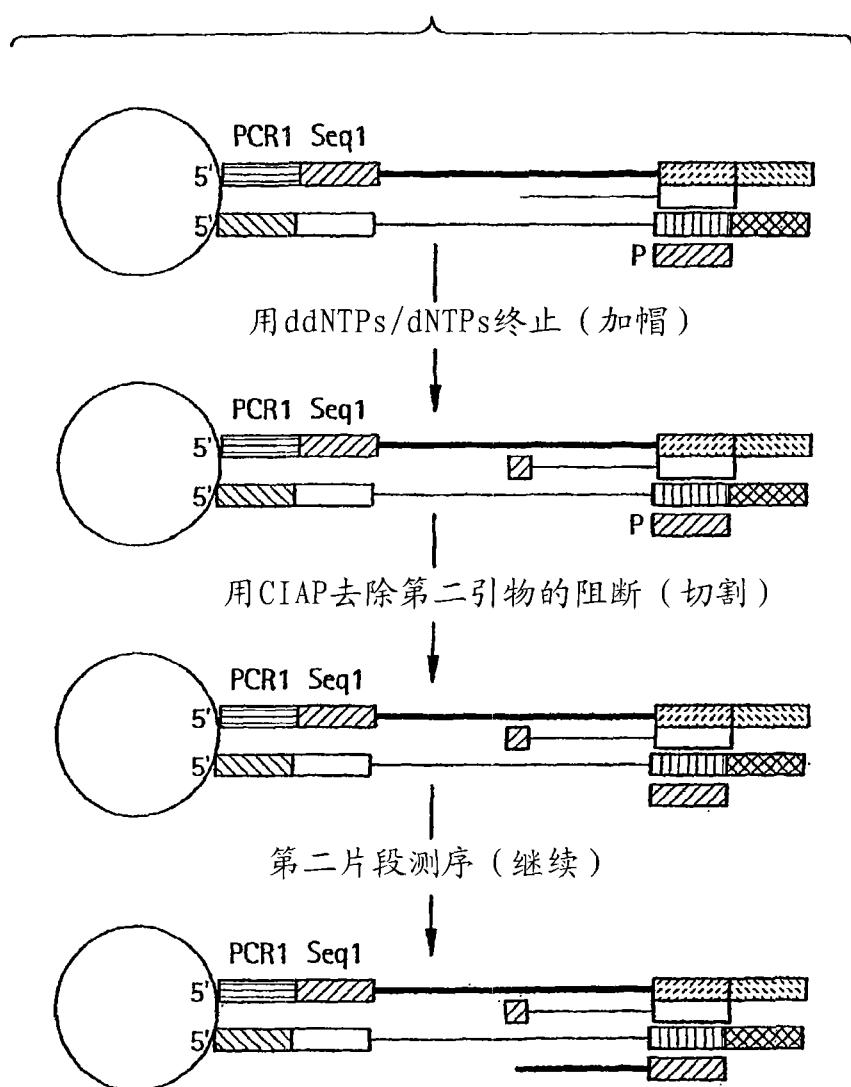


图 38



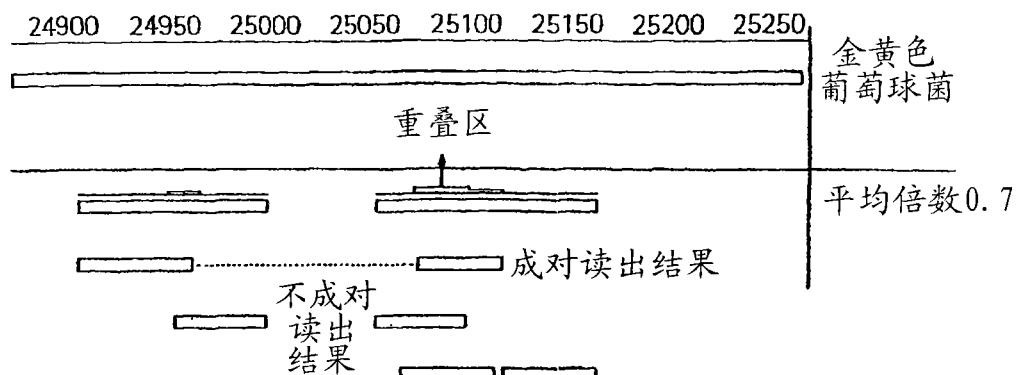


图 39A

总读出结果	31,785
总第一链	15,770
总第二链	16,015
成对	11,799
不成对读出结果	8,187
总覆盖	38%

图 39B

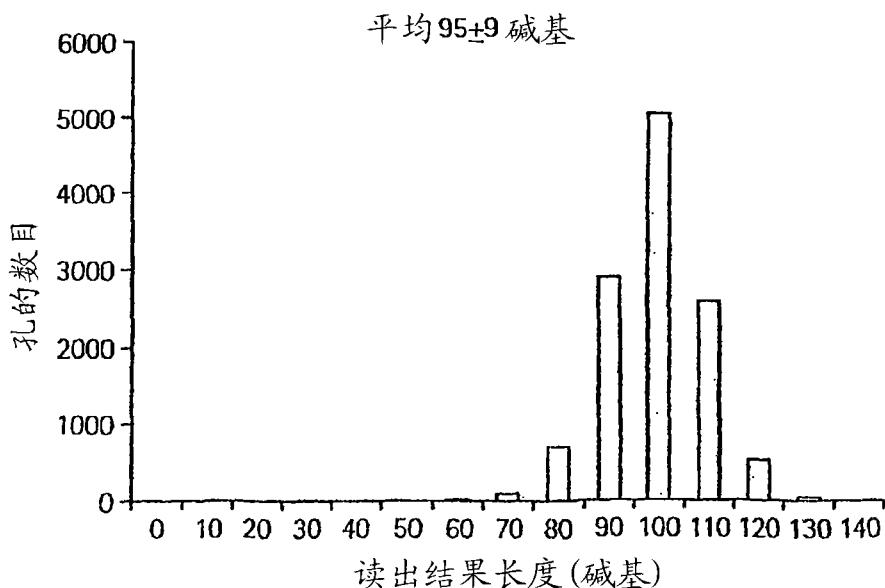


图 40

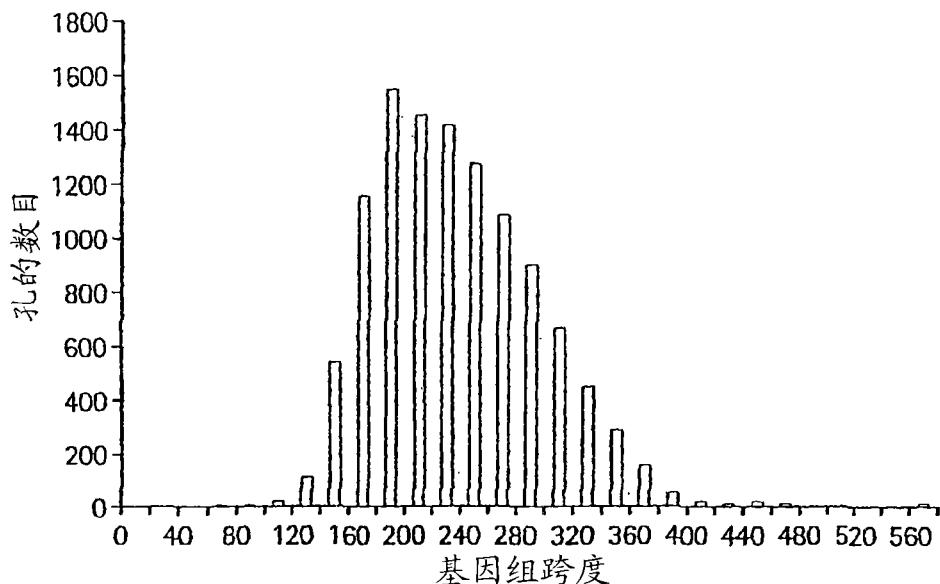


图 41

孔	基因组位置	方向	比对串
00364_0548_2509	571366	F	TATTGTTGATGCTGTAAAAaGAAGCTACTGGTGTAGtATTTTATGAAGTT
00364_0548_2509_D2	571512	R	TGCTCAAAGAACATCAATTAAATATGACCATAATTCTATGATCTTT
00383_0985_2232	1487890	R	AAGCGAACAGTCAGTACCCACAGTCAGTGCACCTTACACAAGCGGAT
00383_0985_2232_D2	1487769	F	TACAGGGTGTGGTATGCCATTGCGATTGGTGGCCTGGTAGCCG
00397_0940_2923	2611033	F	AACATATAAACATCCCCTATCTCAATTCCGCTTCCATGTAAcAAAAAAAGC
00397_0940_2923_D2	2611164	R	TAGATAATCACITGGCTGTACTGGTAATGCAGGCATGAG
00417_0611_1933	122001	R	ATTCAACTCTGAAATGCCTCTGATAACGCCCTGATGATG
00417_0611_1933_D2	121930	F	GATGAGGGAGCTGCAATGGCAATGGGTTAAGGCATCATCG
00434_0595_0993	2022591	R	TGTATCTCGAATTGGATTAGTTGCTTTGCTCTTCATTAGACC
00434_0595_0993_D2	2022473	F	CATTAACATCTGCACCAAGAAATAGCTCTAATACGATTGC
00443_1003_0754	107373	F	GCAGACGACGTCAGCTAATAACGCTGCACCTAAGGCTAATGATAAT
00443_1003_0754_D2	107502	R	AAACCATGCAGATGCTAACAAAGCTCAAGCATTACAGAAACT
00454_1257_3047	59038	R	TGTTGCTGCATCATTAATTAATCTACATCATTTAAATCTTGG
00454_1257_3047_D2	58880	F	GCAGATGGTGTGACTAACCAAGTGGTCAAATGCCCTAAATACAAAAGAT

图 42

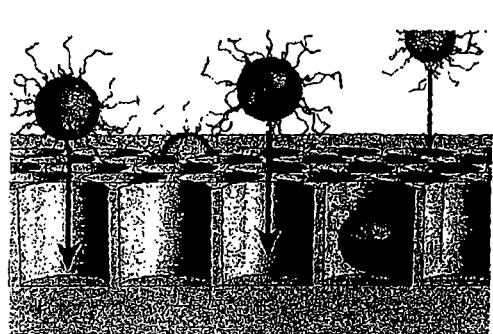
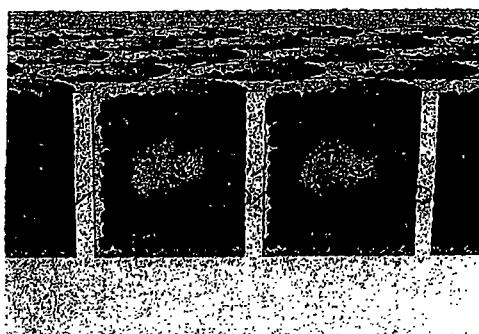
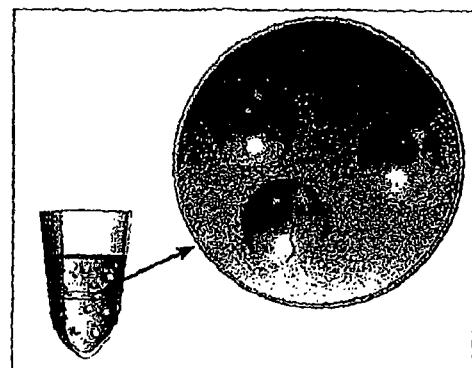
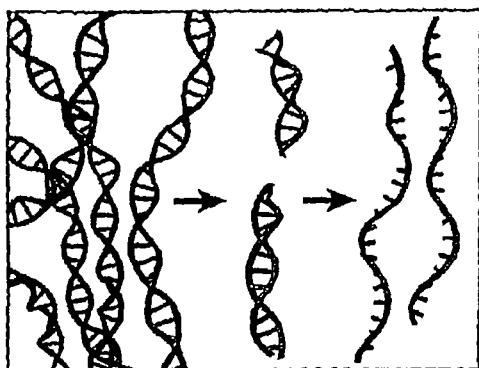


图 43A

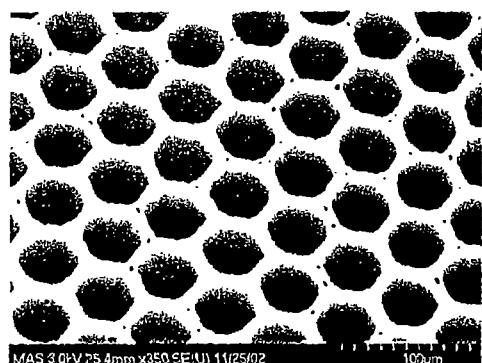
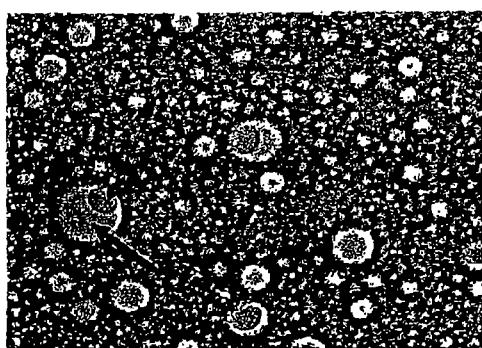


图 43B

图 43C

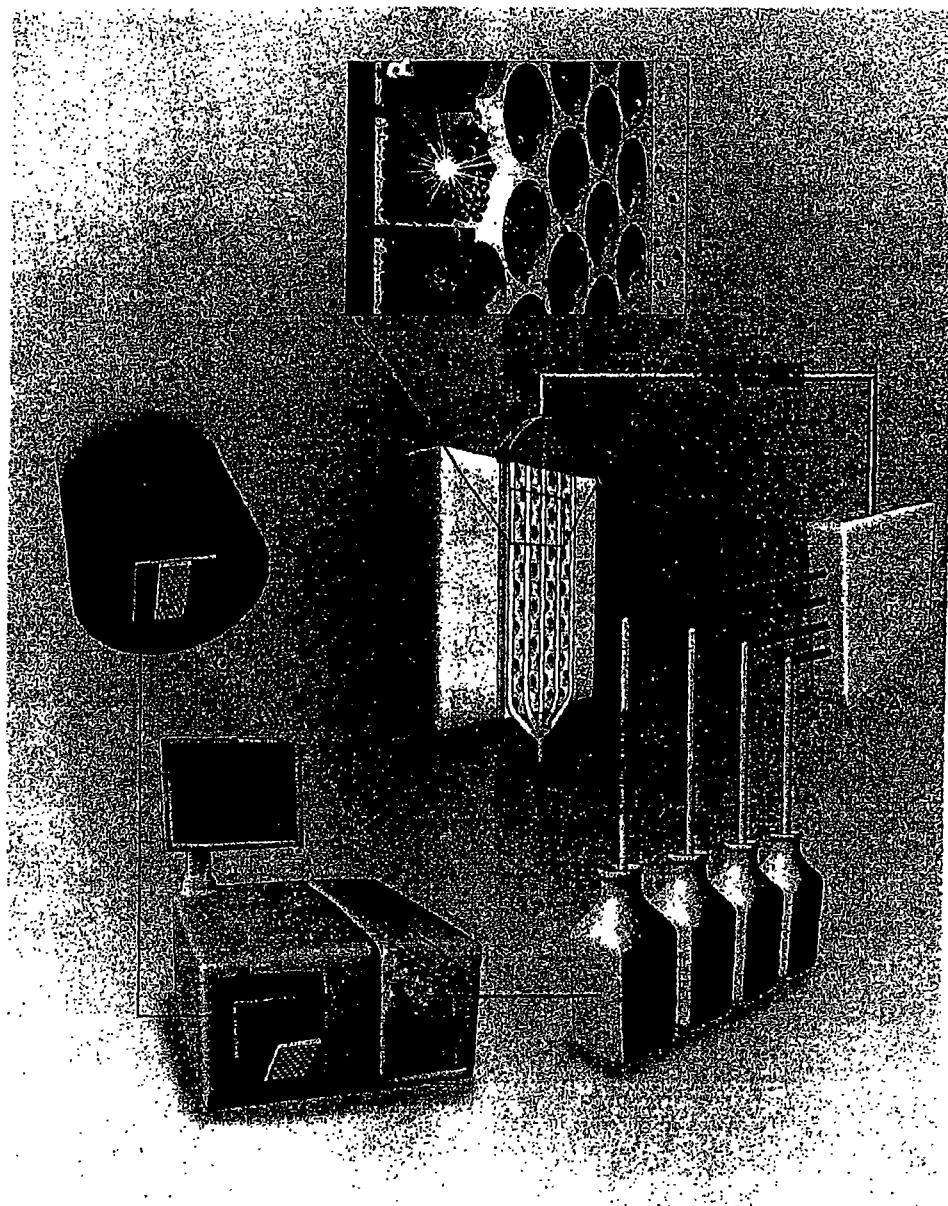


图 44

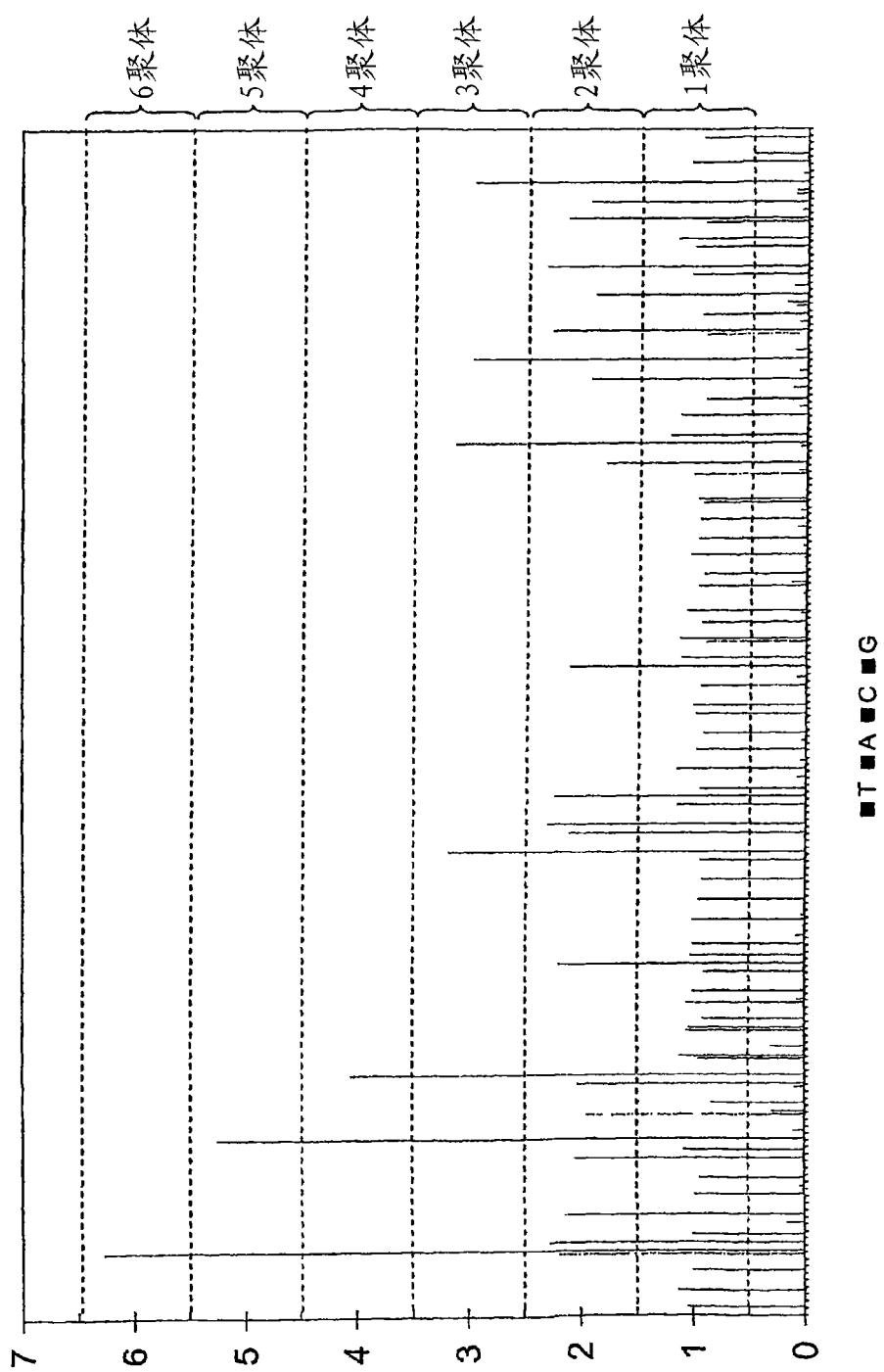


图 45

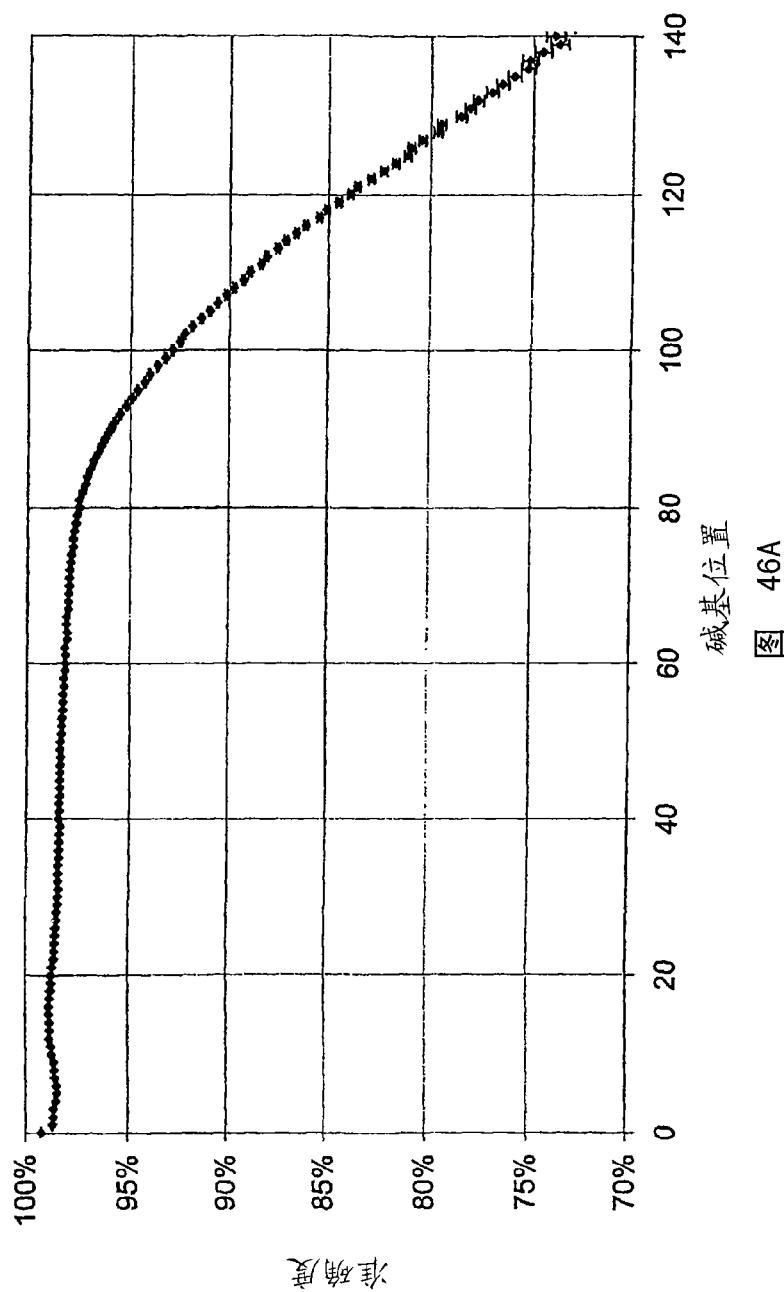


图 46A

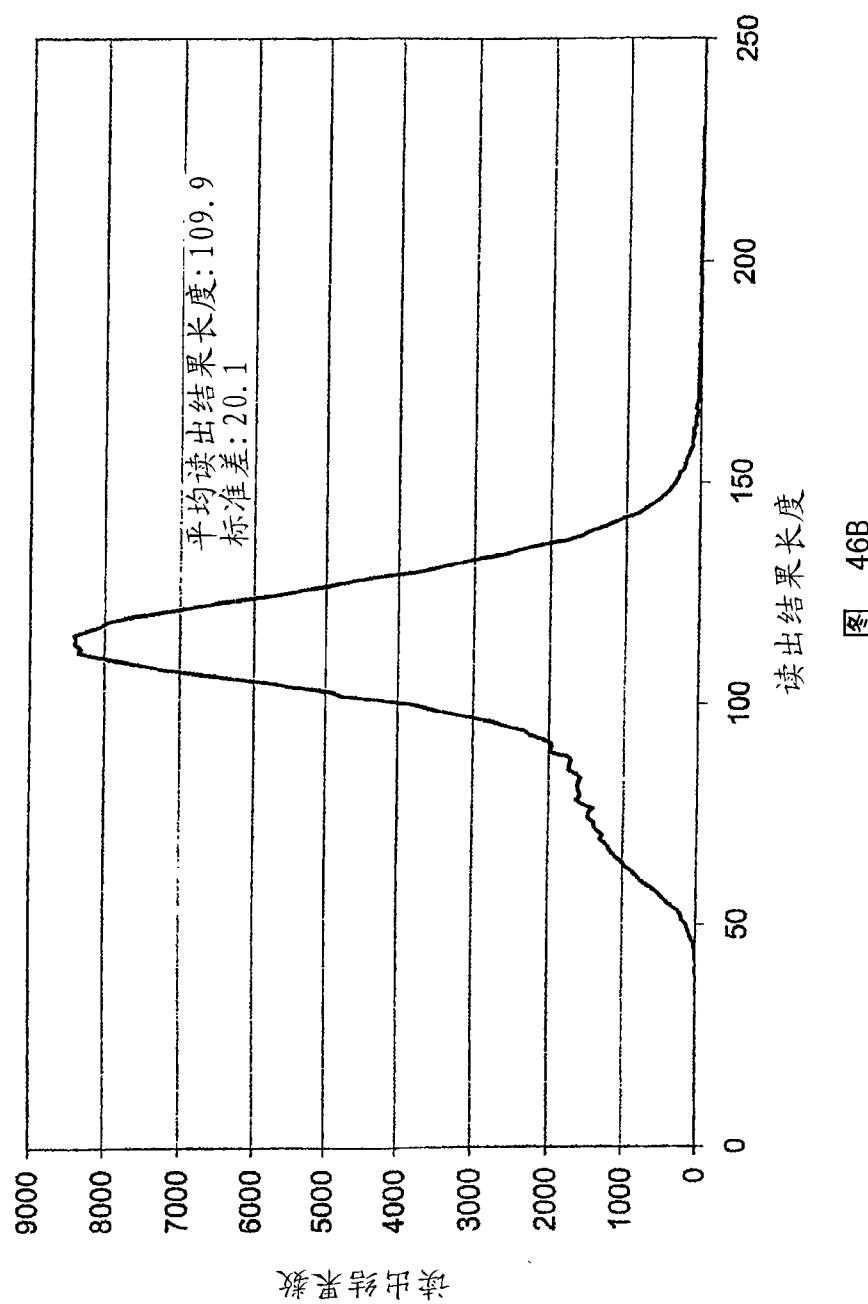


图 46B

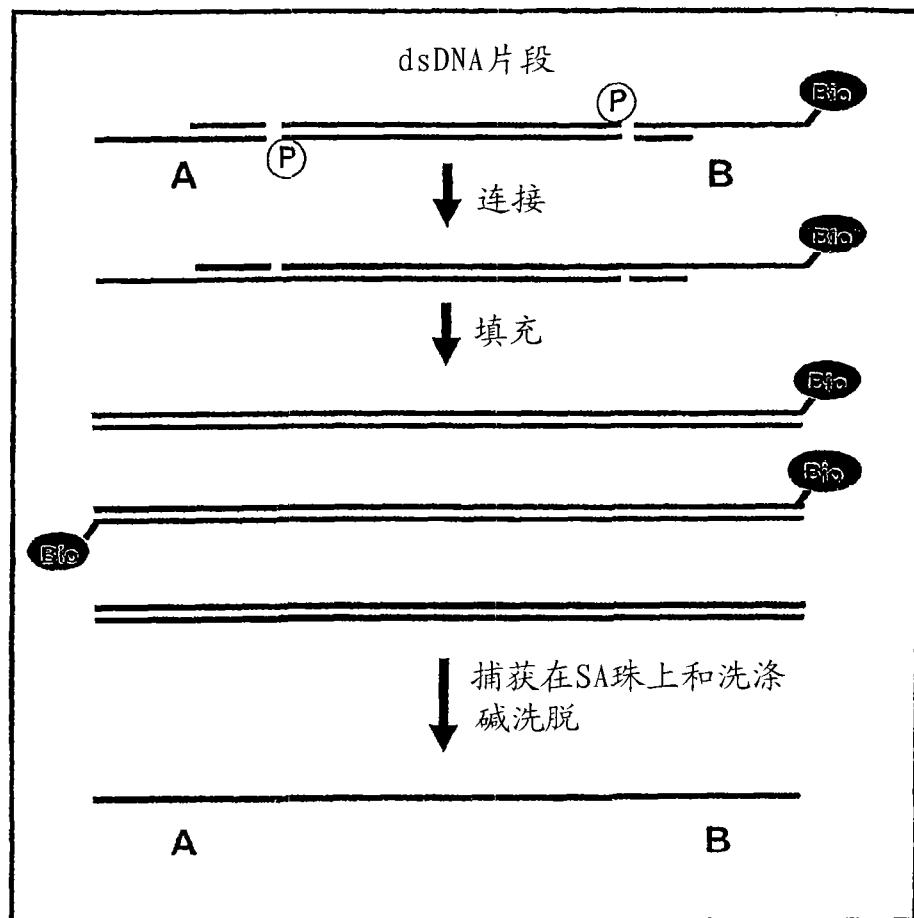


图 47

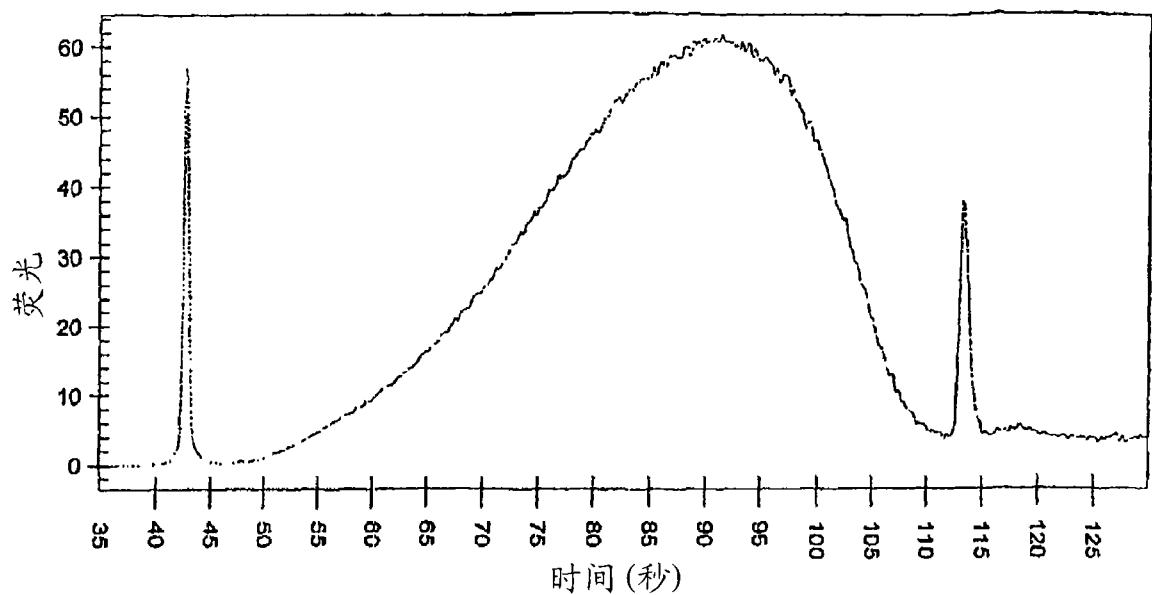


图 48

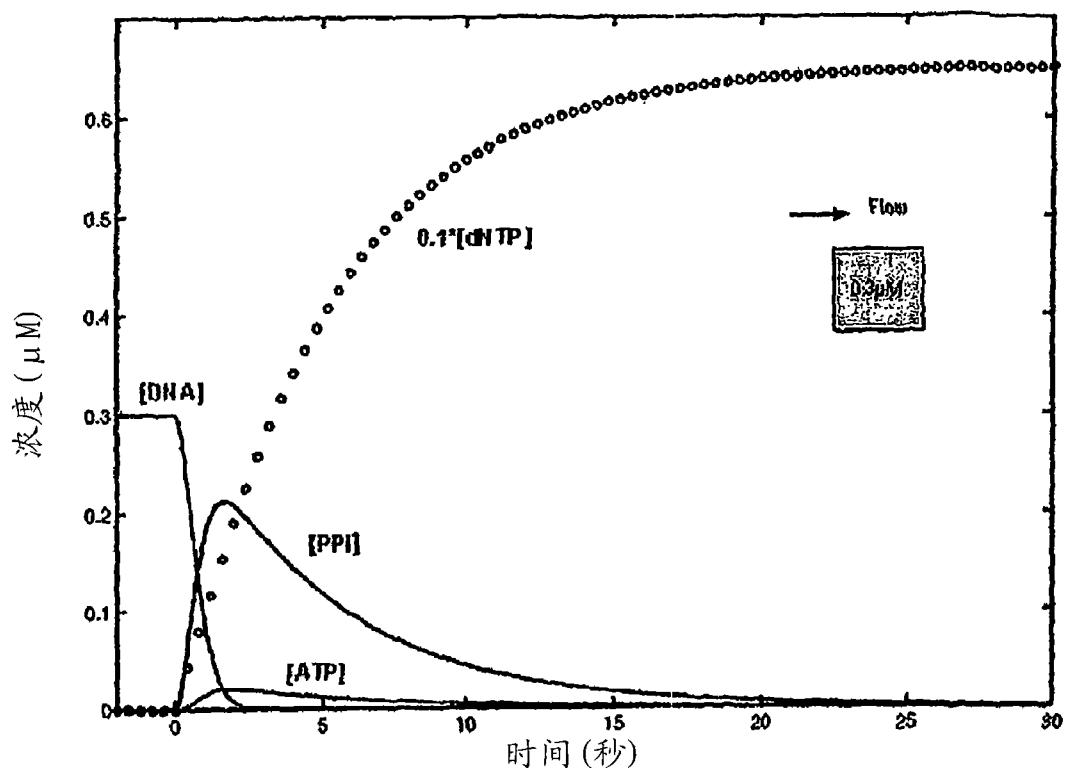


图 49

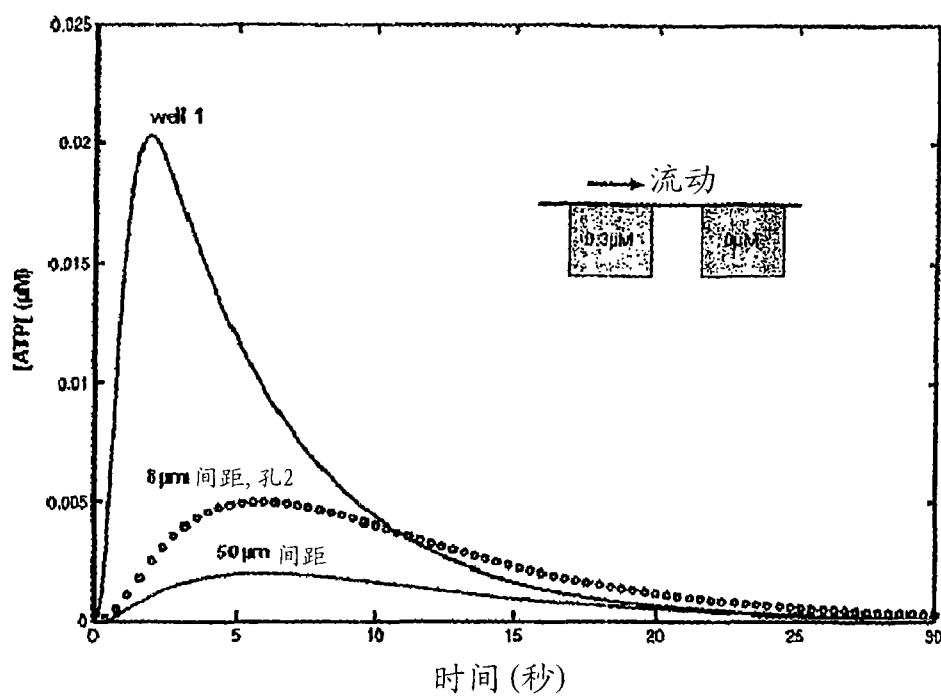


图 50

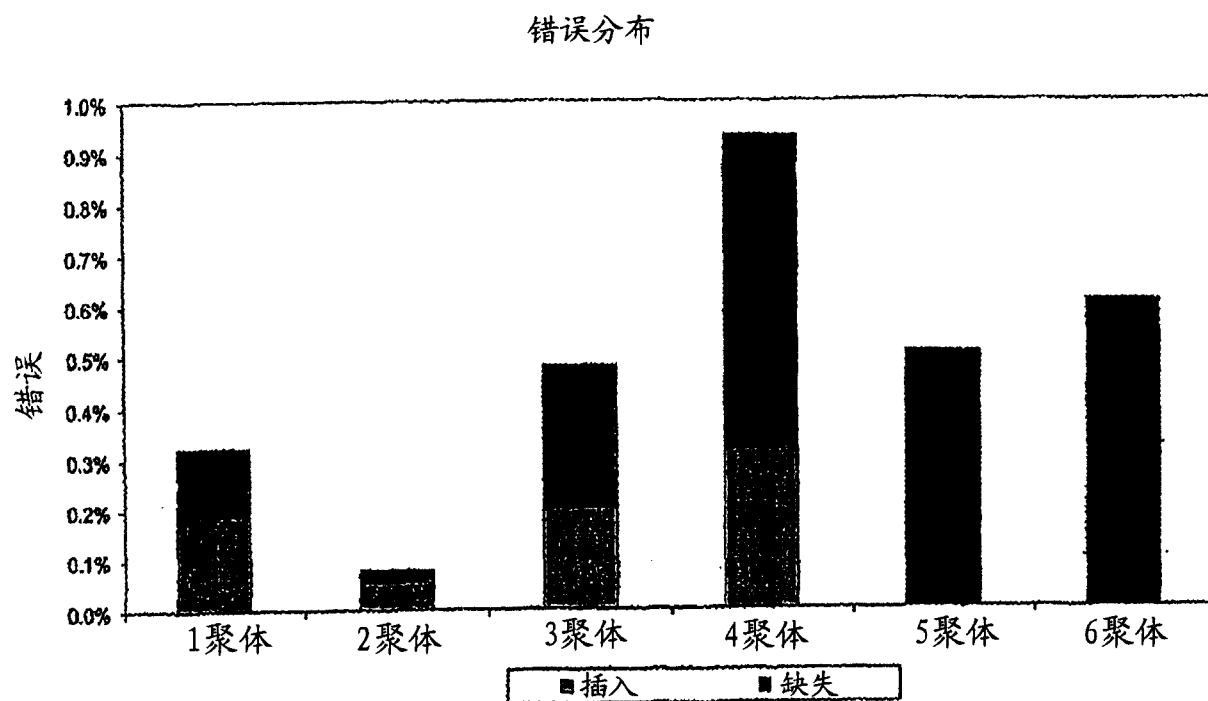


图 51

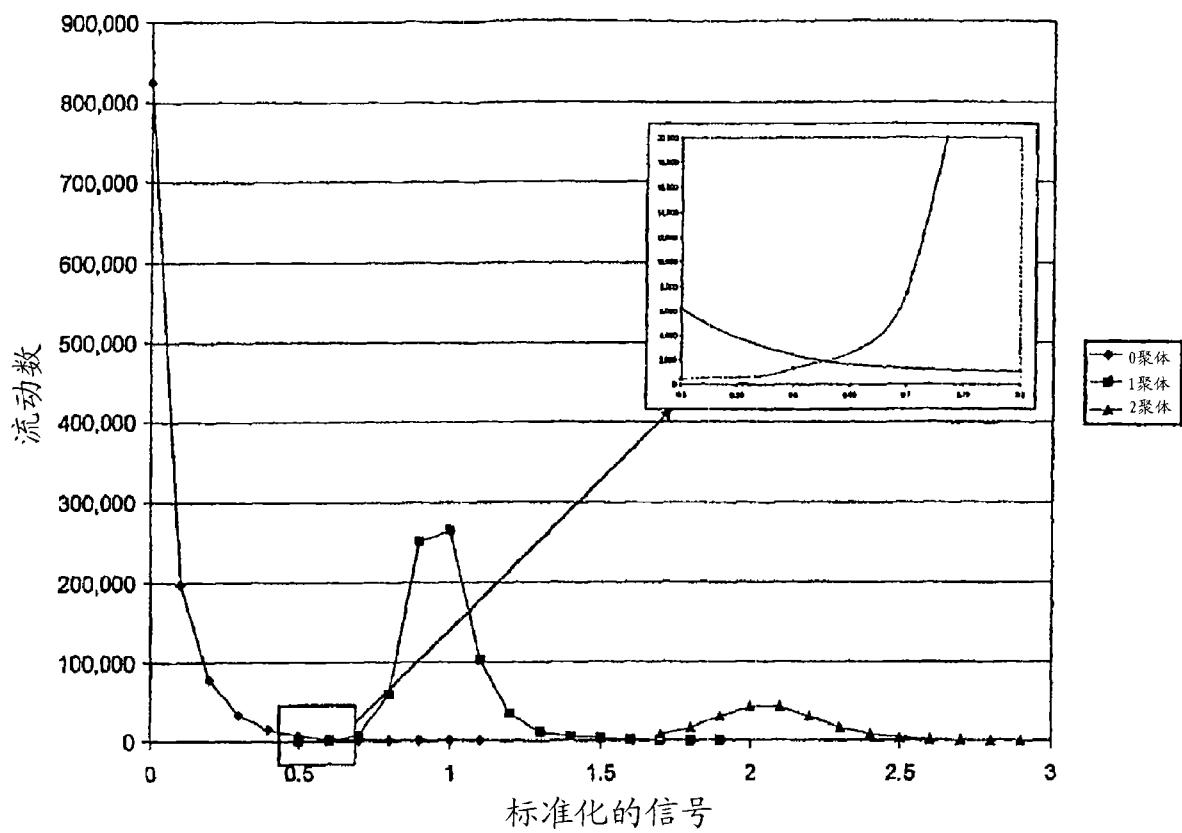


图 52

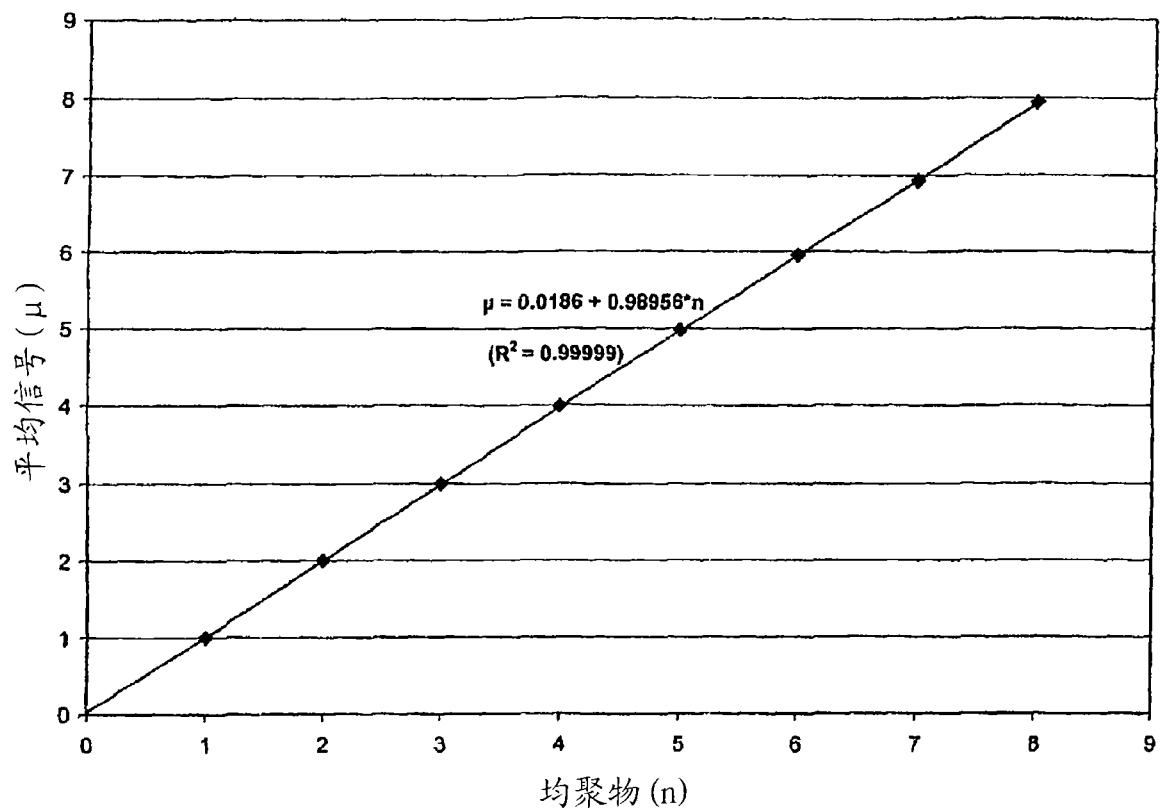


图 53

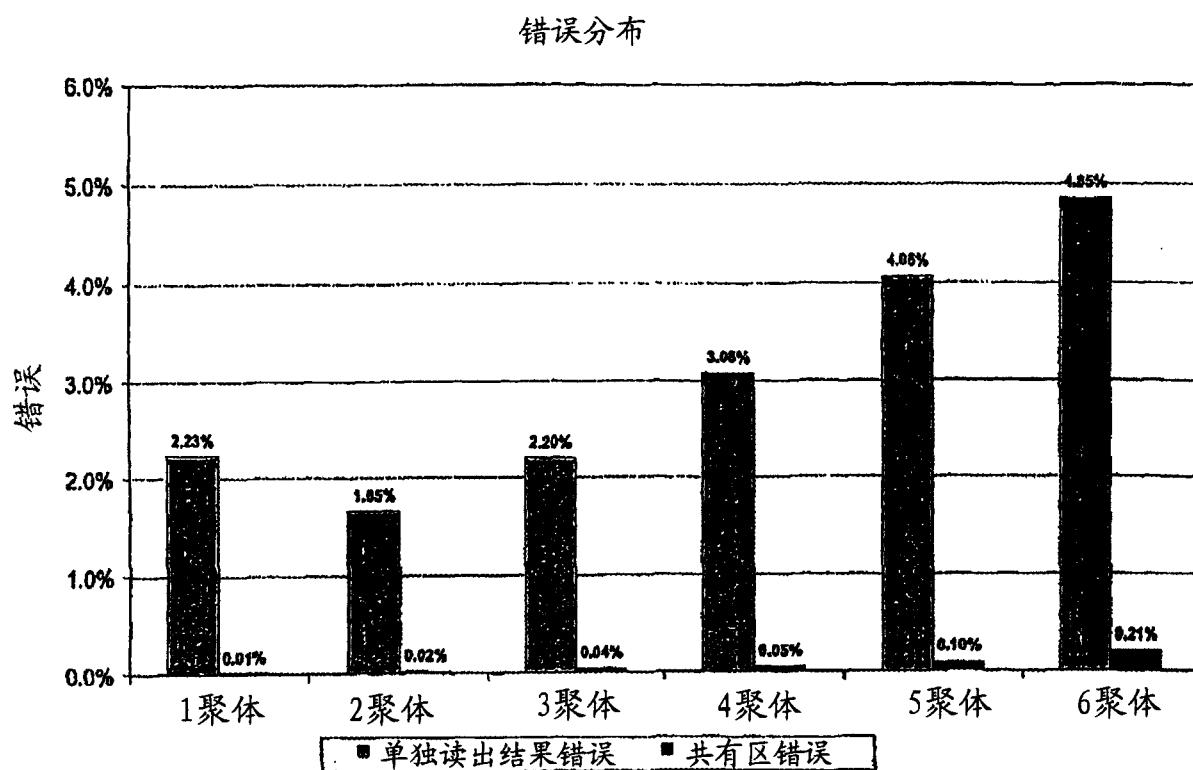


图 54

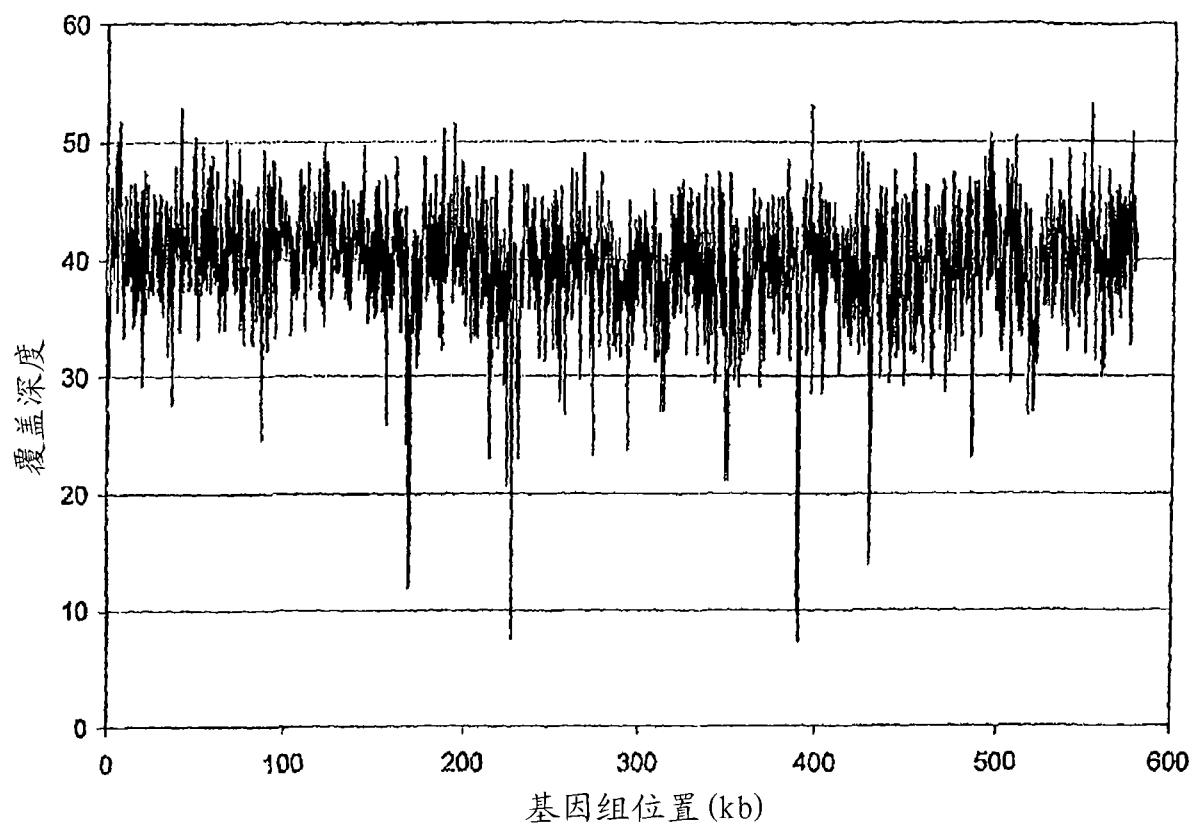


图 55

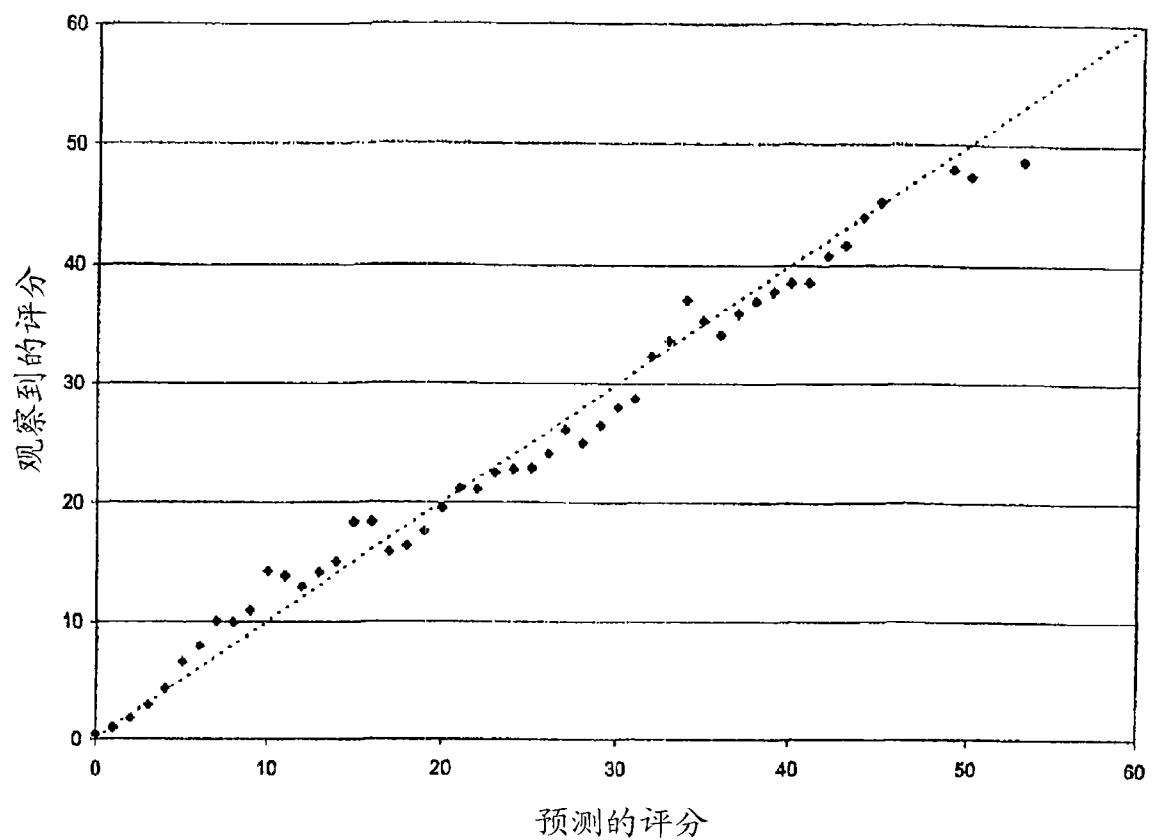


图 56

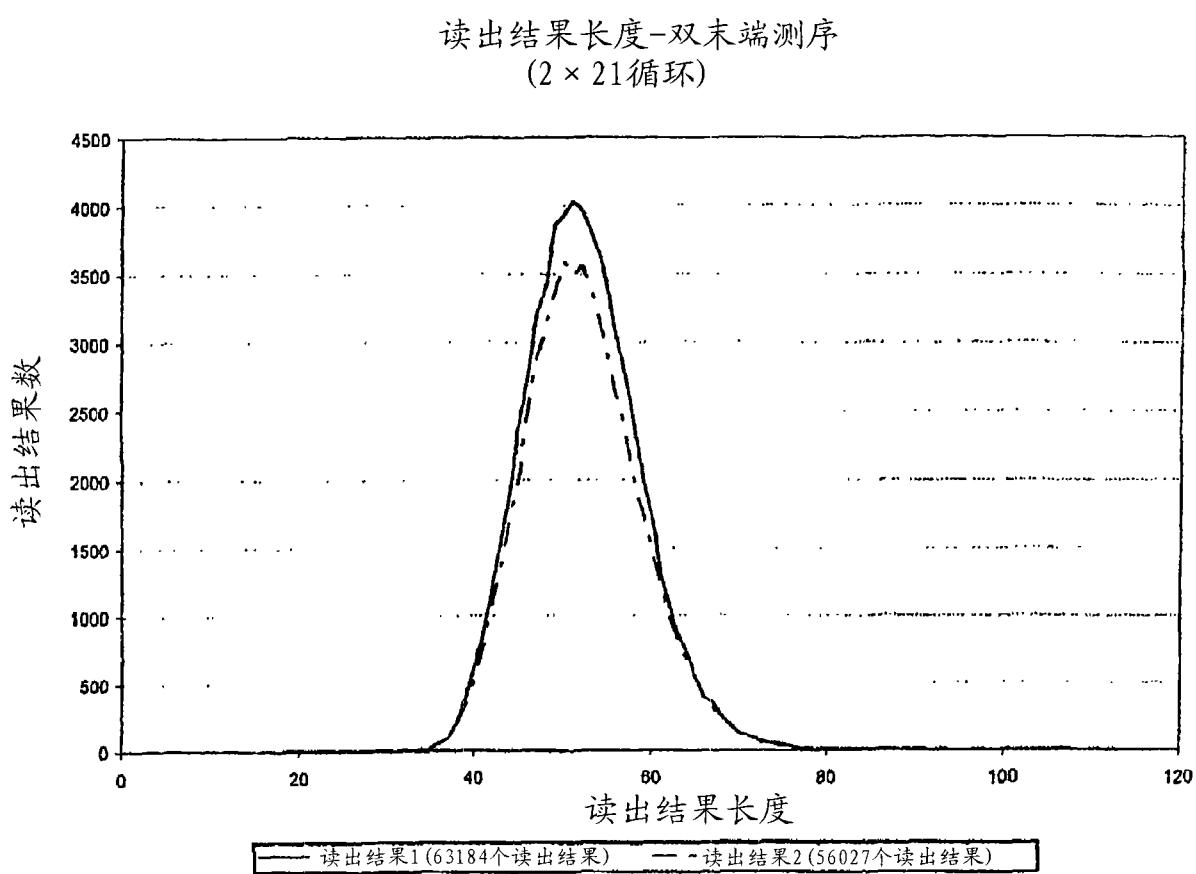


图 57