

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12Q 1/68 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200910135532.5

[43] 公开日 2009年10月21日

[11] 公开号 CN 101560569A

[22] 申请日 2003.9.4

[21] 申请号 200910135532.5

分案原申请号 03824835.2

[30] 优先权

[32] 2002.9.9 [33] US [31] 10/238,439

[71] 申请人 佰尔瑞溶液有限公司

地址 美国新泽西州

[72] 发明人 喀扎拉·哈士密 迈克尔·休

乔金·梅辛

[74] 专利代理机构 北京安信方达知识产权代理有限公司

代理人 郑霞 陶贻丰

权利要求书 3 页 说明书 38 页 附图 18 页

[54] 发明名称

遗传分析和鉴定

[57] 摘要

本发明提供了用于有机体遗传测试，并使遗传测试的结果与能明确鉴定有机体的唯一标记相关的组合物和方法。该标记可以是内部标记，如例如单核苷酸多态性(SNPs)，短串联重复(STR)，或基因组基因座内的其它位点。可选择地，标记可以是外部的，以便它们能在测试之前单独添加到遗传样本中。

1. 一种组合物，其用于分析患者基因组样本中的一种或多种核酸序列，并用于鉴定提供所述样本的所述患者，所述组合物包含下述混合物：

第一组寡核苷酸，其中所述第一组的成员具有不同的序列，并且所述成员连接到珠子上，所述珠子与鉴定连接到其上的寡核苷酸序列的光学可区分特性缔合，所述第一组寡核苷酸用于鉴定所述患者样本中的靶核酸序列；

第二组寡核苷酸，其中所述第二组的成员具有不同的序列，并且所述成员连接到珠子上，所述珠子与鉴定连接到其上的寡核苷酸序列的光学可区分特性缔合，所述第二组寡核苷酸用于鉴定所述患者基因组样本中的许多标记序列，每个所述标记序列包括至少一个多态性标记，对所述标记序列的鉴定提供对所述患者的鉴定。

2. 如权利要求 1 所述的组合物，其中所述珠子混合物排列于平面阵列中，并且所述阵列中各个珠子的位置不是预定的。

3. 如权利要求 2 所述的组合物，其中所述珠子阵列包含子阵，并且第二组寡核苷酸与第一组寡核苷酸位于不同的子阵中。

4. 如权利要求 2 所述的组合物，其中所述珠子阵列置于电极上。

5. 如权利要求 2 所述的组合物，其中所述珠子阵列置于硅片上。

6. 如权利要求 1 所述的组合物，其中所述光学可检测特性是化学标记。

7. 如权利要求 1 所述的组合物，其中所述靶核酸序列被分析以确定它是否含有突变，并且其中所述靶核酸和所述多态性标记位于相同的基因上。

8. 如权利要求 1 所述的组合物，其中所述靶核酸序列被分析以确定它是否含有突变，并且其中所述靶核酸和所述多态性标记位于不同的基因

上。

9. 如权利要求 1 所述的组合物, 其中所述靶核酸序列包含突变位点, 并提供了一组寡核苷酸, 所述寡核苷酸能共同地鉴定所述突变位点, 其中所述组内的寡核苷酸的序列具有差异, 但都与所述靶核酸序列至少部分互补, 所述序列差异包括对应于所述突变位点或与所述突变位点互补的寡核苷酸内的核苷酸。

10. 如权利要求 9 所述的组合物, 其中所述组包括寡核苷酸, 所述寡核苷酸中对应于所述突变位点或与所述突变位点互补的核苷酸是 A、C、G 和 T 或 U。

11. 如权利要求 9 所述的组合物, 其中所述突变位点是多态性位点, 并且是 A、C、G 和 T/U。

12. 如权利要求 11 所述的组合物, 其中所述多态性位点包含单核苷酸多态性。

13. 如权利要求 3 所述的组合物, 其中所述多态性位点包含 STR。

14. 如权利要求 1 所述的组合物, 其中所述寡核苷酸与衍生自所述靶核酸序列和所述标记序列的核酸序列结合。

15. 如权利要求 1 所述的组合物, 其中所述第一和第二组中的寡核苷酸能与衍生自所述患者基因组样本的核酸序列退火。

16. 如权利要求 15 所述的组合物, 其中所述核酸序列是通过扩增所述患者基因组样本内的一种或多种核酸子序列得到的。

17. 如权利要求 16 所述的组合物, 其中所述扩增是使用聚合酶链式反应完成的。

18. 如权利要求 1 所述的组合物, 其中所述第一组寡核苷酸包括能与所述靶核酸序列退火的成员, 并且所述第二组寡核苷酸包括能与所述标记序列退火的成员。

19. 如权利要求 17 所述的组合物, 其中所述第一组寡核苷酸或所述第二组寡核苷酸包括, 或所述第一和第二组都包括能在退火后被延伸的成

员，所述延伸是通过掺入一种或多种类型的脱氧核糖核苷三磷酸或二脱氧核糖核苷三磷酸。

20. 如权利要求 1 所述的组合物，其中所述靶核酸序列包括突变位点，并且所述第一组包括设计用来检测在所述突变位点的不同核苷酸的成员。

21. 如权利要求 1 所述的组合物，其中所述标记序列包括多态性位点，并且所述第一组包括设计用来检测在所述多态性位点的不同核苷酸的成员。

22. 如权利要求 1 所述的组合物，其中所述靶序列包括突变位点，并且所述第一组寡核苷酸包括能与靶序列内的子序列退火的成员，所述靶序列包括所述突变位点，或在所述突变位点的邻近范围内。

23. 如权利要求 1 所述的组合物，其中所述标记序列包括多态性位点，并且所述第二组寡核苷酸包括能与标记序列内的子序列退火的成员，所述标记序列包括所述多态性位点，或在所述多态性位点的邻近范围内。

24. 如权利要求 22 所述的组合物，其中所述多态性位点是单核苷酸多态性。

25. 如权利要求 22 所述的组合物，其中所述多态性位点是短期串联重复位点。

26. 如权利要求 18 所述的组合物，其中至少一种类型的脱氧核糖核苷三磷酸或二脱氧核糖核苷三磷酸的一部分被标记，以产生与掺入后延伸产物相关的光学可检测特征。

27. 如权利要求 19 所述的组合物，其中所述光学可检测标记是荧光染料。

28. 如权利要求 1 所述的组合物，其中所述光学可检测特性允许所述第一和第二组中各个珠子的鉴别。

29. 如权利要求 2 所述的组合物，其中所述珠子阵列的图象显示了各个珠子在随机平面阵列内的位置。

30. 如权利要求 2 所述的组合物，其中与所述第一和第二组寡核苷酸都缔合的珠子随机地置于平面阵列中。

遗传分析和鉴定

本申请是申请日为 2003 年 9 月 4 日、申请号为 03824835.2、发明名称为“遗传分析和鉴定”的原申请的分案申请。

发明背景

遗传分析广泛用于基础和应用研究以及诊断学中，以对患者进行筛选，模式和基因型分析。临床实验目前提供了对超过 300 种疾病或情况的遗传测试，包括在 **BRCA1** 和 **BRCA2** 基因，以及 P53, N-, C-和 K-RAS, 细胞色素 P450, CFTR, HLA I 和 II 类, 杜兴氏肌营养不良和 β - 球蛋白基因中进行突变分析。测试项目继续增加，因为人类基因组计划的发展使得可识别在引起疾病中起作用的遗传定子。

遗传测试包括基因和/或染色体分析，来检测可遗传的或其它突变以及染色体异常，以提供对疾病易感性的诊断。此外，监测蛋白水平来获得疾病发展或对治疗应答的指示。遗传测试用于诊断和检测癌症，以及评估个体发生该疾病的症状前风险。目前，例如，诊断为几种疾病如运动失调性毛细血管扩张症 (Ataxia-telangiectasia), 布卢姆氏综合症 (Bloom's syndrome), 范可尼贫血 (Fanconi's anemia) 或 着色性干皮病 (Xeroderma Pigmentosum) 的家族成员，能用于测试在各自的基因中突变的发生。此外，调节基因 p53 中的几种突变也与发展为不同类型癌症的风险相关。那些遗传了 p53 突变的个体处于发生肉瘤，脑瘤或白血病的高风险之中。用于遗传分析的 DNA 分型和 DNA 指纹分析的标准分析方法包括，(1) 串联重复序列可变数 (VNTR) 分析 (如, Nakamura et al., Science, Vol. 235, pp. 1616-1622 (1987)), (2) 短串联重复序列 (STR) 分析 (如, Edwards et al., Am. J.

Hum. genet. Vol. 49, pp. 746-756 (1991); Ricciardone et al., Biotechniques, Vol. 23, pp. 742-747 (1997), (3) 单核苷酸多态性(SNP)分析 (如, Nickerson et al., Proc. Natl. Acids. Sci. U. S. A., Vol. 87, pp. 8923-8927 (1990); Nikiforov et al. Nucleic Acids Res. Vol. 22, pp. 4167- 4175 (1994); Ross et al., Anal. Chem. Vol. 69, pp. 4197-4202), (d) 限制性片断长度多态性(RFLPS) 分析(如, Botstein et al., Am. J. Hum. Genet. Vol. 32, pp. 314-331 (1980)), 和(4) 线粒体 DNA 序列分析。VNTR 和 STR 分析使用简单或多重聚合酶链式反应(PCR) 技术(如, Mullis et al., Cold Spring Harbor Symp. Quant. Biol., Vol. 51, pp. 263- 273 (1986); Mullis et al., Science, Vol. 239, Vol. 487-491 (1988))。RFLP 分析利用限制性酶消化 DNA, 接着用标记探针进行 DNA 杂交技术; 和线粒体 DNA 序列分析在诸如循环测序的已知方法中利用 PCR 技术和常规双脱氧测序的组合。在特异性遗传定位中 STR 数量上个体间的差异, 显示与几种常见的遗传疾病相关。例如, 不稳定双重复已知与疾病状态如囊肿性纤维化和结肠直肠癌相关。某些不稳定三重复已知与几种遗传疾病包括肯尼迪氏病, 脆性 X 染色体综合征和肌强直性营养不良相关。尤其亨廷顿(氏)舞蹈病已经广泛研究, 而且对一段基因进行了 STR 作图, 以鉴定 51 个跨越 1.86Mbp DNA 区段的三重复。高阶重复, 如四聚物, 也与特定疾病状态包括亨廷顿(氏)舞蹈病和 1 型脊髓小脑性共济失调相关。

基于标准实验方法的 DNA 分型, 要求大量的样本制备和重要的 PCR 后处理。后者包括限制性酶消化, 琼脂糖/丙烯酰胺凝胶电泳, 序列分析或这些方法的组合的步骤。这些多步骤方案在数据中引入了相当大的误差, 而且劳动强度大并耗费时间。

DNA 指纹分析, 也称为身份测试, 依靠高多态性遗传基因座的分析来提供明确的个体分子鉴定。多种多态性标记包括限制性片断长度多态性(RFLPs), 单核苷酸多态性(SNPs), STRs/微卫星和串联重复序列可变数(VNTRS)/小卫星可用于这种目的。RFLP 分析要求对基因组 DNA 进行酶切消化, 接着进行凝胶电泳并使放射性标记的探针与凝胶杂交。该程序的

复杂性阻碍了 RFLP 分析广泛用于鉴定测试。SNPs, 其中一个等位基因与另一个等位基因在单一位置不同, 在编码和非编码区出现的平均频率为 1000 个碱基中出现 1 个, 并组成了人基因组中所有多态性的 90% (Brooker, GENE, 234: 177-186 (1999))。它们已经用于与疾病如癌症相关基因的作图、骨髓移植供体的分类和群体遗传学范围内的遗传研究(Kwok at al. , Mol.Med. Today, 538-543, (1999))。然而, 在开发合适的 SNPs 组以提供明确的 DNA 指纹的时候, 要求仔细的确认那些新的标记。此外, 与目前常用的 STR 标记相比, 用于确保排除不明确的既定可能性所需的 SNP 标记组较大。SNPs 和 STR 多态性都可用作标记, 但是需要每个 STR 多态性中大约 7 到 12 个 SNPs 得到 99.73% 的排斥力。STRs 和 VNTRs 是多信息的多态性标记。许多遗传基因座含有由短、重复序列单元组成的多态性 STR 区域, 长度一般为 3 到 7 个碱基。三聚和四聚 STRs 出现频率为每一个 15,000 碱基的既定序列出现一次, 并广泛用于在家系和法医检定法中进行身份分类。相较于 SNPs 的情况, 其中需要大量的基因座用于排除, 只需要 9 个特异性 STR 基因座就可以提供 99.73% 的组合平均排斥力(Alford et. al Current Opinion in Biotechnology, 29-33 (1994), Latour et al, 829-37 (2001))。STRs 可以使用针对旁临串联重复区域的特异性引物序列, 通过聚合酶链式反应 (PCR) 来进行扩增。

由于在等位基因中重复单元数的差异而引起的其它多态性, 包括串联重复序列可变数(VNTRS)/小卫星, 其是含有 9 到 60 个碱基的短序列的串联重复, 以及含有 1 到 5 个碱基的微卫星。小卫星和微卫星通常被认为是 VNTRs 的亚类。由于估计大约 500,000 个微卫星重复分布于人基因组, 平均间隔为 7,000 个碱基, VNTR 区域也可用于身份测试。

在常规实验室实践中, 使用放射性标记的或荧光标记的引物通过 PCR 对 STRs 和 VNTRs 进行扩增。PCR 产物通过凝胶电泳或毛细管电泳进行分离用于鉴定。

在遗传测试的常规执行中, 与样本和患者鉴定相关的信息是人工记录

的，一般涉及制作贴到样本收集容器上的条形码标记。这种标记程序表示可能的重要误差来源，包括违反运行规程，贴错标签和转换样本。

因此，需要存在一个机制，由此收集的已知生物学样本在收集的时候就被明确标记和确认。这将保护以防在分析过程中违反运行规程，贴错标签和转换样本。

发明概述

本发明提供了平行分析 STRs 和相关重复序列单元的方法，以明确地把样本与遗传测试结果和患者身份联系起来。特别地，本发明提供了与完成遗传分析同时记录分子鉴定(ID)的方法，通过把患者的遗传模式与患者的分子指纹联系起来，从而使样本的疏忽违反运行规程的发生率最小化，并通过与以前记录或后面记录的分子鉴定进行比较来进行明确的鉴定。

本发明的一个方面，提供了一种用于分析从患者样本获得的靶核酸序列，并同时提供患者的遗传指纹的组合物。所述组合物包括一第一组探针和一第二组探针。所述第一组探针包括能与从用于遗传测试的患者样本获得的靶核酸序列杂交的寡核苷酸探针，而所述第二组包括用于与多个多态性标记杂交的寡核苷酸探针。与这些标记的所述杂交提供了能确认患者的遗传指纹。这两组探针都连接到珠子上，所述珠子带有能用于唯一确定连接到所述珠子上的所述探针的化学或物理可区分特性。

优选地，所述组合物中的珠子排列于平面阵列中。更进一步优选地，所述组合物中的珠子阵列包括子阵，而且第二组探针与第一组探针位于不同的子阵中。

优选地，所述组合物中的平面珠子阵列置于电极上。另外优选地，所述组合物中的平面珠子阵列置于硅片上。

优选地，所述组合物中的珠子具有能唯一确定连接到所述珠子上的探针的化学标记，其中的化学标记包括一种或多种荧光染料。

优选地，所述靶核酸序列被分析以确定它是否含有突变，而且其中的靶核酸和多态性标记位于相同的基因上。

另外优选地，所述靶核酸序列被分析以确定它是否含有突变，而且其中的靶核酸和多态性标记位于不同的基因上。

优选地，所述靶序列包括突变位点，并提供了多个相应的探针，所述的探针与所述靶序列互补，除对应于该突变位点的核苷酸之外，而且位于所述位点的核苷酸一起达到了位于所述位点的所有突变。

优选地，所述靶序列包括突变位点，并提供了多个相应的探针，所述的探针与所述靶序列互补，除对应于该突变位点的核苷酸之外，以致位于所述位点的核苷酸一起提供了所有四种碱基 A,C,G 和 T/U。

优选地，所述每个标记都含有多态性位点，而且对于每个标记，提供了多个对应的探针，所述探针与所述标记序列互补，除了位于该多态性位点的核苷酸之外，以致位于该多态性位点的核苷酸一起达到了位于所述位点的所有已知多态性。

更进一步优选地，所述每个标记都含有多态性位点，而且对于每个标记，提供了多个对应的探针，所述探针与所述标记序列互补，除了位于多态性位点的核苷酸之外，以致位于该多态性位点的核苷酸一起提供了所有的四种碱基 A,C,G 和 T/U。

优选地，所述多态性标记包括单核苷酸多态性。

优选地，所述多态性标记包括 STR。

本发明的另一个方面，提供了一种用于分析从患者样本获得的靶核酸序列，并同时提供患者的遗传指纹的方法。该方法包括提供第一组探针和第二组探针。所述第一组探针包括能与从用于遗传测试的患者样本获得的靶核酸序列杂交的寡核苷酸探针，而所述第二组探针包括用于与多个多态性标记杂交的寡核苷酸探针。与这些标记的所述杂交提供了能确认患者的遗传指纹。所述第一组和第二组探针都连接到珠子上，所述珠子带有能用

于唯一确定连接到所述珠子上的所述探针的化学或物理可区分特性。该方法进一步包括使靶序列和多个多态性标记与所述第一组和第二组探针接触，然后检测所述第一组探针与所述靶序列之间的杂交，并检测所述第二组探针与所述多态性标记之间的杂交。

优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述具有连接到其上的探针的珠子排列于平面阵列中。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述珠子阵列包括子阵，而且第二组探针与第一组探针位于不同的子阵中。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述平面珠子阵列置于电极上。

另外更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述平面珠子阵列置于硅片上。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述珠子具有能唯一确定连接到所述珠子上的探针的化学标记，其中的化学标记包括一种或多种荧光染料。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述靶核酸序列被分析以确定它是否含有突变，而且其中的靶核酸和多态性标记位于相同的基因上。

另外更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述靶核酸序列被分析以确定它是否含有突变，而且其中的靶核酸和多态性标记位于不同的基因上。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述靶序列包括突变位点，并提供了多个相应的探针，所述的探针与所述靶序列互补，除对应于该突变位点的核苷酸之外，而且位于所述位点的核苷酸一起达到了位于所述位点的所有突变。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述靶序列包括突变位点，并提供了多个相应的探针，所述的探针与所述靶序列互补，除对应于该突变位点的核苷酸之外，以致位于所述位点的核苷酸一起提供了所有四种碱基 A,C,G 和 T/U。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述每个标记都含有多态性位点，而且对于每个标记，提供了多个对应的探针，所述探针与所述标记序列互补，除了位于该多态性位点的核苷酸之外，以致位于该多态性位点的核苷酸一起达到了位于所述位点的所有已知多态性。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述每个标记都含有多态性位点，而且对于每个标记，提供了多个对应的探针，所述探针与所述标记序列互补，除了位于多态性位点的核苷酸之外，以致位于该多态性位点的核苷酸一起提供了所有的四种碱基 A,C,G 和 T/U。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述多态性标记包括单核苷酸多态性。

更进一步优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，所述多态性标记包括 STR。

优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，其进一步包括在使所述标记和序列于探针接触之前，同时扩增多态性标记和靶序列的步骤。

优选地，在所述分析靶核酸序列并同时提供患者的遗传指纹的方法中，其包括分析多个靶序列，而且第一组探针包括用于与所述多个靶序列杂交的寡核苷酸探针。

本发明的另一个方面，提供了一种用于分析从患者样本获得的靶核酸序列的方法。该方法包括提供一种唯一联系序列分析与样本的装置，并包

括提供一组探针，所述探针包括能与从患者样本获得的靶核酸序列杂交的寡核苷酸探针。所述探针连接到珠子上，所述珠子带有能唯一确定连接到所述珠子上的所述探针的化学或物理可区分特性。所述方法进一步包括使寡核苷酸探针与含有靶核酸序列的溶液接触，以使靶序列与相应的探针杂交，并检测所述探针与所述靶序列的杂交。该溶液用能唯一确定靶溶液的分子标记进行标记，以便通过探测标记来确定患者的身份。所述标记可以在溶液加入到寡核苷酸之前或之后或同时添加给样本。

优选地，在所述分析靶核酸序列的方法中，其进一步包括检测分子标记来鉴定患者的步骤。

优选地，在所述分析靶核酸序列的方法中，所述样本的标记包括添加一种或多种荧光标签，所述标签因此成为与靶溶液相关。

优选地，在所述分析靶核酸序列的方法中，所述靶溶液的标记包括添加寡核苷酸标签。

优选地，在所述分析靶核酸序列的方法中，其进一步包括设计一组探针与对应的序列标签杂交，所述序列标签探针连接到珠子上，其中所述的珠子被编码能唯一确定连接到所述珠子上的探针。

本发明的另一个方面，提供了一种测定靶核酸序列中串联核苷酸重复数的方法，其中串联重复在每侧通过一个非重复旁临序列旁临。所述方法包括，提供一组连接到珠子上的寡核苷酸探针，其中每个珠子都带有能唯一确认连接到所述珠子上的所述探针的化学或物理可区分特性。每个探针都能与靶序列退火，而且含有探测位点。设计这组探针以使这些探针具有不同的重复核苷酸数。当探针与靶序列退火形成杂交复合物时，每个探针的探测位点都与一个位于串联重复里面或其外面的靶位点连接。所述方法进一步包括使靶序列与寡核苷酸探针接触，以使所述靶序列与所述探针形成杂交复合物。所述靶序列和这组探针之间的杂交复合物，被平行探测以确定探针的探测位点结束于靶重复外或靶重复里面。还确定了靶序列中的

重复数。优选地，所述平行探测杂交复合物的步骤包括使所述杂交复合物与序列特异性聚合酶和标记的核苷三磷酸接触，以致核苷三磷酸整合到一个或多个探针序列中，该标记根据探针-靶复合物的构象而不同。优选地，所述珠子以空间编码的方式装配排列到平面中，并通过直接成象检测光学信号的改变和确定颗粒身份。

本发明的另一个方面，提供了一种在靶核酸序列分析中产生的检测信号的序列特异性扩增的方法。该方法可以实时监控扩增到的信号，而且该方法包括提供一温度控制的样本密封装置，其能实时记录在该装置中产生的光学检测信号。该方法进一步包括提供一种温度控制装置，用于控制该装置的温度，并在样本密封装置内，提供一组探测寡核苷酸探针。这些探针能与靶核酸形成杂交复合物并连接到珠子上。每个珠子都带有能确定连接到所述珠子上的所述探针的化学或物理可区分特性。使寡核苷酸探针与靶序列接触，从而所述探针与所述靶序列之间形成杂交复合物。这种杂交复合物再与第二寡核苷酸探针接触，所述第二寡核苷酸探针含有标记，并能与所述杂交复合物内包含的探测探针连接。该方法还包括提供一装置，使所述第二标记的寡核苷酸探针与所述杂交复合物内的所述探测探针连接，然后实时检测来自固定探针组的光学信号。进行一个或多个退火-连接-检测-变性循环，每个循环都能以算术级数增加延伸探针的数量，并包括以下步骤：

(i)提供用于形成杂交复合物的第一温度；

(ii)提供用于发生探测探针和第二标记探针的连接酶催化连接的第二温度，其中连接与珠子的光学信号变化相关，所述珠子带有连接的探针；

(iii)描绘和/或记录来自探针的光学信号；和

(iv)提供用于使全部的杂交复合物变性的第三温度。

当联系下述的详细描述和伴随的附图时，可以更清楚地理解本发明的目的，特征和优点，其中的详细描述应当理解为只是示例说明。

附图说明

图 1 为显示产生嵌入的遗传 ID 的方法的示意图。

图 2 为显示了 CFTR 区域的限制性图谱的示意图 (X 是 SNP 标记, D7S122 和 D7S8 是 STRs, MET 是甲硫氨酸, NOT1 是限制性位点, IRP 是基因)。

图 3 为显示了在 CFTR 基因的外显子内突变的示意图。

图 4 是一个 DNA 序列, 其显示了 CFTR 基因的外显子 7 内的多态性标记 (大写字母表示已知的多态性)。

图 5A 是一个 DNA 序列, 其显示了 CFTR 基因的外显子 10 内的多态性标记。

图 5B 为设计用于 CFTR 基因外显子 10 内 SNP 鉴定的探针的示意图。

图 6 为显示用磷酸化引物进行 PCR 分析的方法的示意图。

图 7 为显示了分析肌营养不良蛋白基因缺失的方法的示意图。大多数多态性标记在该基因中缺失。可以设计引物来扩增发生缺失的外显子的旁临序列。缺失的序列和多态性标记可以同时被鉴定。

图 8 为显示了具有位于环外面的基因, 以及环内各种疾病导致的突变和多态性的线粒体基因组的示意图。

图 9A 为探针设计的示意图, 其用于带有使用标记的 ddNTPs 的锚定序列的 STR 长度多态性。

图 9B 为使用 STR 多态性对 THO1 基因座的结果的示意图。

图 10A 为探针设计的示意图, 其用于不带锚定序列的 STR 长度多态性。

图 10B 显示了使用不带有锚定的探针对 THO1 基因座的结果。

图 11A 为带有两种不同标记的 ddNTPs 的芯片上的 STR 长度多态性的

示意图。在该情况中，ddNTPs用不同的颜色进行不同的标记。

图 11B 显示了使用两种不同标记的 ddNTPs 获得的结果，其中绿色 ddNTP 整合有正确的匹配，橙色 ddNTP 当探针终止于重复内时整合。

图 12 显示了锚定长度对使用 6 个或 11 个碱基的两种锚定进行 STR 多态性分析的 SBE 的影响。

图 13 是用于 STR 多态性分析的杂合引物的示意图。

图 14 说明了退火温度对用于 STR 多态性分析的 SBE 的影响。

图 15A 说明了对内部探测探针的荧光能量转移。

图 15B 说明了对外部探测探针的荧光能量转移。

图 16 示例了用于鉴定 CFTR 基因内含子 8 中的聚-T 变异体的探针序列设计。

图 17 示例了使用 CFTR 基因内含子 8 中的聚-T 变异体鉴定各种靶的结果。

图 18 示例了探针序列设计用于鉴定较长的重复，如那些 10 到几百个碱基长度的重复。

图 19A 示例了通过连接探测探针来鉴定重复序列(T_1 是起始检测温度， T_2 是终止检测温度，其中 T_2 大于 T_1)。

图 19B 示例了通过连接和芯片上循环来鉴定重复序列 (T_1, T_2 和 T_3 表示 3 种检测不同温度，其中 $T_3 > T_2 > T_1$)。

详细描述

患者的遗传模式分析不仅在基础和应用临床研究中，而且在疾病诊断和对疾病诱因的评估中的作用越来越重要。一种安全，可靠的遗传测试方法优选合并个体测试中涉及患者鉴定的所有相关信息。本发明提供了用于联系从患者样本分析获得的遗传模式与患者身份的方法和组合物。在记录遗传指纹或分子识别 (ID) 的基础上，在遗传测试或者任何诊断或预兆

测试的同时建立了患者遗传模式与身份之间的这种相关性。

本发明的方法可用于预防在遗传测试过程中的违反运行规程，贴错标签和转换样本。这些方法可用于父性和母性测试，迁移和遗传争论，双胞胎中的接合性测试，近亲交配测试，骨髓移植成功的评估，人遗体的确认以及诸如包括精液或血液的法医检定测试。本发明通过整合遗传指纹或分子识别到遗传或其它测试的记录中，其以例如这里详细描述的形象的形式获得，预防或纠正了与违反运行规程，贴错标签和转换样本相关的鉴定错误。这样，建立了记录与患者身份之间的明确关系。分子识别可用于追踪并确认样本的身份，从而提供鉴定的方法。本发明的方法提供了组合物和用于产生遗传 ID，这里也称作 ID 的方法，同时完成遗传或者其它诊断或预兆测试。在分析遗传基因座如含有多个突变和广泛分布标记的 CFTR 的情况下，本发明提供了记录位于靶内的已经扩增的 ID 标记的方法，其用于遗传分析目的。在分析中，其中的遗传基因座只含有很少几个突变，位于其它基因组区域内的 ID 标记可通过提供另外的引物来进行扩增。

一种遗传指纹分析常用的方法包括，某些基因座内许多重复序列单元的多态性分析。为了促进重复序列多态性的整合，这里提供的方法用于通过 STR/VNTR 分析进行阵列形式的鉴定测试。本发明的方法最小化了样本操作和处理所需的步骤数，从而最小化了测定过程影响结果的可能性，这种可能性是开发用于患者鉴定的数据库要关注的。尽管这里描述的遗传指纹分析方法尤其可用于在遗传模式分析的情况中提供患者身份，它们也可用于联合其它遗传分析，如基因分型，单倍体分型或 HLA 分子分型。本发明方法发现也可用于遗传分析的独立进行地遗传鉴定的情况中。

而且，通过把 ID 指派给特定的样本容器或载体，本发明的方法产生了载体 ID 与遗传 ID 之间的明确关联，从而不仅最小化了样本操作错误的可能性，而且能确认检测结果。载体 ID 和遗传 ID 与一个数据库联系，进行患者身份的数据确认和鉴定。

一组已知序列的 DNA 片断用作本发明所述的外部标记的用途，在至

少两个方面是比现有技术有利的。第一，标签的确定，及因此鉴定被标签的样本，可使用如杂交，延伸和连接等方法与感兴趣的遗传分析同时进行。第二，得到的样本 ID 通常可嵌入到由遗传分析产生的图像或数据记录中。也就是说，样本 ID 和遗传分析的结果保持相关联。相反，通过现有技术的方法进行的标签鉴定通常要求完成单独的分析程序如用于确定外部标记的 DNA 片断长度的电泳，以及遗传测试本身。

如这里所用的，术语“多态性”指基因中的序列变异，“突变”指与或被认为与一种表型相关的基因中的序列变异。术语“基因”指编码功能性产物蛋白调控区域的基因组的片断。本发明中所用的用于患者确认的多态性标记，可位于基因组的编码区或非编码区。术语“患者”，如这里所用的，指提供测试样本的个体，从该测试样本获得的靶核酸用于遗传测试目的。

术语“ID”，“ID 标记”和“标记”这里可互换使用，指内部标记和外部标记，它们的高可变性使它们适于作为特殊个体的分子识别。为了本发明的目的，内部标记，特别包括用于 DNA 指纹分析的遗传 ID 标记，可以是位于感兴趣的基因内的“内在的”标记，或可以是非内在的，或“外来的”标记。

内部标记包括位于基因组的基因座内的 SNPs,STRs,VNTRs 和其它多态性位点。外部标记包括化学的，荧光的，磁性的，分子的和一般在样本收集时能整合到样本中的其它标签。它们可用于使遗传或其它测试如预兆和诊断测试中获得的结果与患者身份唯一相关。外部标记的例子包括生物标记如由寡核苷酸，肽核酸，DNA，RNA，蛋白质，ABO 血型等等组成的标签，以及化学标记如光学活性颗粒，染料等等。

随机 DNA 序列标签的收集，如那些构建用于某些杂交检测的标签，表示一类能提供很大的编码能力的外部标记。例如，最近公布的一组 164 个序列标签 (<http://waldo.wi.mit.edu/publications/SBE-TAGS>)，衍生自噬菌体的基因组，能提供 2^{164} 不同的组合。在一个优选的实施方案中，选自

文库的每个 DNA 序列标签的给定样本中存在或缺失，可通过提供颜色编码珠子的鉴定阵列来确定，其中每种类型的珠子表示一种能唯一与一种序列标签匹配的寡核苷酸探针。来自鉴定阵列中的珠子的这组信号，用相同类型的珠子进行平均，组成了患者样本的分子条形码。

通过本领域已知的方法合成的无机荧光毫微颗粒，对以颗粒大小确定的波长进行的单个波长激发产生应答而发光，表示另一类外部标记。特定亚类的毫微颗粒标签在样本中的存在，可通过样本的光谱分析并同时完成感兴趣的 DNA 分析来进行确定。现有技术中已经描述了两种类型的无机毫微颗粒标记，即半导体 Q-点（量子点）颗粒和 RLS（谐振光散射）金属毫微颗粒。量子点是毫微米（ 10^{-9} 米）大小的颗粒，由半导体材料如硒化镉(CdSe)，碲化镉(CdTe)或砷化铟(InAs)制成。它们的组成和小的体积（几百个到几千个原子）给予这些点特别的光学特性，其能通过改变这些点的体积和组成而容易地定制。量子点吸收光，然后迅速地以不同的波长发射光。尽管其它的有机染料和无机材料显示这种现象，但是量子点具有窄的、对称发射光谱的明亮和非漂白的优点，而且具有使用单个激发波长可同时激发的多种可分辨的颜色（Bruchez et al., “Semiconductor Nanocrystals as Fluorescent Biological Labels, Science, 281, 2013(1998), Alivisatos, “Semiconductor Clusters, Nanocrystals, and Quantum Dots Science, 271, 933(1996)”).

谐振光散射技术（RLS 技术）是基于“毫微大小的”金属（例如金或银）胶体颗粒，当用简单的白光源照射时其能以散射光的形式辐射能量。单个 RLS 颗粒产生的单色光信号比从最灵敏的荧光团获得的信号高 10^4 到 10^6 倍，因此这些毫微颗粒能在多种分析生物检测和测试形式中用作超灵敏的生物标记。（Yguerabide, J. Analytical Biochemistry, 262, 137-156 (1998), Yguerabide, J. Analytical Biochemistry, 262, 157-176 (1998)）。

为了使用本发明的外部标记，收集样本如血液，唾液，毛发或骨髓样本。外部标记被整合作为样本的部分，并且当样本进行处理和分析时保持

与样本整合或其他关联。这种标记在包括群体载体筛选，基因分型，单倍体分型或诸如细胞因子，抗原或血清中抗体等蛋白质分析的模式分析的检测中可以被整合。

在某些实施方案中，一种或多种外部标记可用于一种或多种内部标记的场合。在其它的实施方案中，一种或多种外部标记可用于与一种或多种内部标记组合，从而提供另一种进行样本确认和鉴定的方法。外部标记也可整合到检测中，该检测包含于设计并提供用于收集和/或储存患者样本的柱体中，从而建立样本，检测容器或载体与遗传测试之间的物理联系。

用于遗传测试和遗传指纹分析的靶序列或靶核酸，可以是基因，调控序列，基因组 DNA，cDNA 和 RNA（包括 mRNA 和 rRNA）的部分。基因组 DNA 样本通常在与探针接触之前进行扩增。基因组 DNA 可从任何组织来源或循环细胞（除了纯的红血球）获得。例如，基因组 DNA 方便的来源包括全血，精液，唾液，眼泪，尿，粪便，汗，口腔细胞，皮肤和毛发。含有多态性位点的基因组 DNA 的扩增，如果从其获得样本的个体在多态性位点是纯合的就产生单一种类的靶核酸，或者如果个体是杂合的就产生两种靶分子。RNA 样本也经常进行扩增。在这种情况下，扩增通常先于反转录进行。所有表达的 mRNA 的扩增可如例如 WO 96/14839 和 WO 97/01603 中所述的进行，其全部内容引入这里作为参考。如果提供样本的个体在表达的 RNA 内出现的多态性位点是杂合的，对来自二倍体样本的 RNA 样本进行扩增，会产生两种靶分子，或者如果 RNA 的种类进行可变剪接则可能更多。通常使用本领域已知的 PCR 方法进行扩增。靶样本中的核酸可在扩增的过程中通过混入一种或多种标记的核苷酸到扩增混合物中来标记。标记也可以在扩增之后连接到扩增产物上（如通过末端-标记）。扩增产物可以是 RNA 或 DNA，取决于扩增反应中使用的酶和底物。

串联重复，如例如短的串联重复 (STRs)，可用作遗传 ID 标记。通常，进行基因分型和相关的遗传测试的感兴趣遗传基因座，含有由在个体之间

高可变的很多重复序列单元组成的重复序列。本发明的一个方面，其中的标记是串联重复，靶核酸序列中的串联重复数可通过这里所述的平行探测来确定。

在本发明一个优选的实施方案中，为了在遗传测试过程中给突变打分，通常设计针对感兴趣的每个可变靶位的探针对。成对的两个探针都与靶核酸序列互补，除了突变位点，其中成对探针中的至少一个与靶互补。探针的长度足以与靶进行杂交，优选长度为 10 到 50 个碱基，更优选长度为 10 到 20 个碱基。探针也可通过接头连接到固体支持物上。突变分析包括一种或多种靶序列的探测，并可用便于高通量筛选的多种方式进行。例如，针对不同靶序列的探针可以通过如这里所述的把探针连接到编码的珠子上而排列于平面随机阵列中。在另一个实施例中，可以形成相同基质（如硅电极）上的不同亚阵列，而且编码和定位珠子亚阵列的珠子都提供关于位于各个珠子上的探针的身份的信息。如对重复的序列的分析，平行探测包括形成靶与序列特异性探针之间的杂交复合物。在优选的实施方案中，探针延伸提供了通过荧光标记的 dNTPs 整合进行延伸产物直接标记的方法，包括但不限于，组合延伸和芯片上温度循环的方法。标记的延伸产物可通过这里所述的成像进行检测。

在本发明优选的实施方案中，能“引导”延伸反应的延伸探针组，以能保持它们的特性并减少延伸产物鉴定中的不明确的方式，固定于固相载体上。例如，这可以通过空间分散不同探针和/或通过化学编码探针特性来完成。

当探针与靶链在允许杂交复合物形成的条件下接触时，位于每个探针 3' 末端上或其附近的探测位点通常以两种构象中的一种，即以“重复-内部”构象或以“重复-外部”构象，与靶排列。前者中，探测位点与靶的重复序列中的位点并列；后者中，探测位点与靶的前导序列中的位点并列。

探针通常设计成具有与目的基因组区域互补的尾巴和前导序列。为了制备通用的尾巴和前导序列，杂交引物可以设计成具有扩增产物的部分的

序列。例如，正向和反向 PCR 引物可以设计成在它们各自的 5' 末端含有富含 GC 的标签。基因组 DNA 的 PCR 扩增在产物的 5' 末端引入该标签（图 13）。寡核苷酸探针然后被设计成含有与引物标签互补的 5' 锚定序列。这种设计机动性便于改正探针与扩增的靶之间的连接以使滑动最小化（也就是，靶在几个不同的位置杂交），并增强检测中的区分。例如，尾巴序列以从一个重复序列中不存在的核苷酸（两个-和三个-核苷酸重复）开始的方式设计，并能通过单个碱基延伸进行检测。

通过一组合含有重复可变数的探测探针进行靶重复序列探测的步骤，被设计成把对应于两种构象中的一种的两个值中的一个指派给每个探针，即匹配，（数字表示为 1），或不匹配，（数字表示为 0）。由针对重复序列多态性的一组探针产生的探测结果的二进制序列，确定了该多态性内的靶重复数。如在很多实施例中详细描述，多种探针结构和检测方法的变化都是可能的，包括但不限于，用以荧光能量转移方式进行检测的直接杂交，及模板介导的探针延伸，其包括单个碱基延伸或连接。

在某些实施方案中，探针序列可以构造成含有补偿序列，位于第一个探针重复的 5' 端，而且该补偿序列与靶重复的第一个或多个（但不是所有的）核苷酸匹配。含有补偿的探针与靶以最佳的排列退火，又在两种可能构象之一产生了杂交复合物。在第一种构象中，探针重复被数量等于补偿大小的靶重复在探针的 3' 方向置换，以便探针的 3' 末端与一种靶重复的内部位置连接。在第二种构象中，探针序列的 3' 末端与前导序列内的位置连接，该位置可通过补偿的大小确定。

对于每个感兴趣的重复序列多态性，探针组可以构建成含有相同的（随意的）锚定序列和相同（随意的）补偿序列，但接着具有更高的重复数，以便探针组能跨越这里详细描述的可能的靶重复范围。在一个实施方案中，探针可以设计成含有与靶的尾巴和前导序列互补的 5' -或 3' -锚定序列，以稳定探针与靶的所需连接。目的组合物的前导和尾巴序列可以通过本领域已知的 PCR 方法 (Innis et al, Academic Press, San Diego,

CA(1990),Mattila et al., Nucleic Acid Research, 4967 (1991))在患者 DNA 扩增的位点引入。前导和尾巴序列可以转换以确定互补 DNA 链上的重复。

在其它的实施方案中,既不含 5' -也不含 3' -锚定序列的寡核苷酸探针可用于竞争杂交。在一个这种实施方案中,标记的或未标记的靶首先与设计成含有足够大重复数的“封闭”DNA 链形成杂交组合物,以超过最大预期的靶重复数。在该杂交复合物中,所有的靶重复因而是双重构象的。使这种“封闭”靶的溶液现在与显示具有不同重复数探针的编码珠子的阵列接触。反应混合物进行孵育使探针与封闭链竞争以结合靶重复。在该设计中,只有那些含有的重复数与靶重复数相等或超过靶重复数的探针,将从靶中取代(部分)封闭链并从而获得检测信号。以增加的严紧度洗涤,并调整本文所述的退火温度,增强区分的水平。

在另一个优选的实施方案中,在升高的温度记录检测图象,而且信号被记录作为升高的温度的函数。在所有探针-靶杂交复合物的强度-对-温度图中的连续峰,表示各自的解链温度,而且探针含有与具有最高解链温度的靶重复数相等或超过其的重复数。不同颜色标记的靶可以在相同反应中进行分析。

与固定的探针形成杂交复合物的靶,可通过这里前面所述的检测方法观察。例如,与靶链退火的探针可以用标记的 dNTPs 延伸,以便当探针与靶中的重复数最佳匹配时出现延伸。几种用于产生阳性检测信号的其它构象也可以容易地构建。

如通常所述的序列特异性探针,用于平行探测重复序列的探针可以通过接头部分固定到固体支持物上,该接头部分的应用是本领域熟知的。作为总的原则,探针应当足够长以避免与不相关的 DNA 靶序列退火。探针的长度可以为大约 10 到 50 个碱基,更优选大约 15 到 25 个碱基,甚至更优选 18 到 20 个碱基。在多重检测中,可使一种或多种溶液形式的靶与多种固定的探针在允许进行退火和延伸反应的条件下接触。因此,本发明具有超过现有的通过凝胶电泳分析重复序列多态性的方法的优点,现有方法

是一种不适于高通量操作的方法。

本发明还包括用于平行探测单核苷酸多态性和单位点突变，以及用于检测其它类型突变和多态性如多核苷酸多态性（例如，两个，三个等等），和通常观察到的并可用于遗传测试的小的插入和缺失的方法。

为了使完成分析所需的劳动，材料和时间最小化，要求在一个单一反应中同时扩增和分析多个基因座。多重扩增方法尤其可用于遗传疾病的分析，其中包括，杜兴氏肌营养不良，自毁容貌综合征和囊性纤维化。此外，几种自身免疫性疾病（如糖尿病）已经与人白细胞抗原（HLA）系统中的多态性相关。HLA复合体的多态性基因座显示了强的连锁不平衡，以致特定的单倍型在染色体上比预期的更常出现。对胰岛素依赖性糖尿病（IDDM）的易感性已经发现与DQ基因座编码的特定II类等位基因相关。基因座DQ α 可用于分子分型，并可用于包括疾病测序和分子分型的同时分析。

遗传学上定义的线粒体疾病，其大部分是由线粒体（mt）DNA中的突变引起的，提供了另一个例子。平均的人细胞含有成千个mtDNA分子，其是编码37个基因，13个氧化磷酸化（OXPHOS）多肽，rRNAs和22 tRNAs的16,586核苷酸对的闭合循环分子。mtDNA分子具有比核基因组更高的突变率。

几个mtDNA突变与大脑，心脏，骨骼肌和肾的变性疾病相关（图8）。线粒体脑肌病形成了与受损的氧化磷酸化相关的紊乱的遗传异源组。患者可能显示从肌肉衰弱到视力降低和大脑变性紊乱的大范围临床症状，而对其目前没有有效的治疗。线粒体基因组中发生的碱基取代和重排突变，与眼肌病，Kearn-Sayre或Pearson综合征和成人型糖尿病相关。线粒体疾病起因于雌性种系中的突变或获得的突变。ATP合成酶中的碱基取代突变与肌肉衰弱，共济失调，色素性视网膜炎（NARP），Leigh综合征，中央视觉降低（LHON）肌张力障碍和MELAS相关。三种mtDNA突变已经证实与阿尔茨海默氏病和帕金森病相关。

几个多态性标记已经在 mtDNA 中鉴定，并广泛用于群体遗传学研究中。这些标记中的大多数都位于含有导致疾病的突变的基因内，而且能与其一起扩增。特异性地，32 序列多态性，位于 tRNA 基因内，适合作为单独的 ID 标记，其能嵌入到突变模式中。(Garboczi, et al, Mol. Cell BioChem 107: 21-29 (1991))。

本发明的方法包括靶核酸序列与遗传 ID 标记的同时探测。这可通过提供两组或多组核酸探针来完成，如单链或双链形式的 DNA 或 RNA，或带有合成主链的核酸类似结构如肽核酸。根据本发明，第一组探针设计成与感兴趣的靶核酸序列互补，第二组探针设计成与一种或多种指定的 ID 标记互补。这些第一与第二组探针可以连接到固相载体上，例如，芯片，晶片，玻片，膜，颗粒，珠子或任何能与考虑的检测兼容的表面。

如这里所用的，术语“珠子”，“微球体”，“微粒”和“颗粒”可以互换使用。珠子成分可包括，但不限于，塑胶，陶瓷，玻璃，聚苯乙烯，甲基苯乙烯，丙烯酸聚合物，顺磁物质，碳石墨，二氧化钛，乳胶或交叉连接右旋糖苷如琼脂糖，纤维素，尼龙，交叉连接胶束和聚四氟乙烯。

珠子可以带有物理或化学可区分的特性。例如，珠子可用光学可区分的标签组染色，如那些含有一种或多种荧光团或发色团染料的标签，可通过激发波长，发射波长，激发态寿命或发射强度来区分。以某种摩尔比率组合的光学可区分染料可用于根据本领域已知的方法对珠子进行染色。用于外和内表面的组合的色码描述于国际申请号 PCT/US98/10719, 在此引入作为参考。能在物理或化学可区分特性的基础上被鉴定的珠子被认为是“编码的”。

每组珠子的化学或物理可区分特性的检测，以及使用这些珠子在遗传或其它测试(如诊断或预后测试)中产生的这种珠子上的光学信号的确定，可分别通过记录一组珠子或这些珠子的阵列的译码图象和检测图象并对两个图象进行比较来进行。例如，在某些实施方案中，可以使用带有图象检测器和计算机化影像捕捉和分析装置的系统。获得译码图象以确定能唯

一鉴定显示于珠子表面的探针的化学和/或物理可区分特性。这样，阵列中每个颗粒上的探针的身份可通过可区分的特性来提供。获得阵列的检测图象以对如这里下面详细描述的检测中产生的光学信号进行检测。

除了被编码外，具有特异性寡核苷酸探针或引物的珠子可以某种方式空间分散，以便珠子的位置能提供关于珠子和进而探针或引物身份的信息。在一个实施方案中，空间编码可通过把珠子置于两个或多个空间分散的子阵列中来提供。

在一个优选的实施方案中，在译码和分析之前，珠子可以排列在基质上的平面阵列中。珠子阵列可通过 PCT/US01/20179 中所述的方法制备，其全部内容引入这里作为参考。珠子阵列也可以使用美国专利号 6,251,691 中所述的方法形成，其全部内容引入这里作为参考。例如，光控动电力可用于装配称为“LEAPS”过程中的珠子阵列，如美国专利号 6,251,691 所述。可选择地，如果使用了顺磁珠子，可以通过应用垂直于表面的磁场而在基质表面形成阵列。珠子阵列也可以通过把珠子机械地置于基质表面的约束结构（如凹口）中而形成。在某些实施方案中，珠子阵列在形成后可通过物理方法，如，例如，通过使珠子嵌入到凝胶中形成凝胶颗粒膜来进行固定。

使用随机编码珠子阵列进行多重分子分析的例子，可通过 ABO（和 RH）血型的遗传分析和测试提供。在该实施方案中，编码珠子阵列装配于给定芯片的分散位置中，以允许同时进行遗传分析和测试。该分析可通过在编码珠子上显示对应于 ABO 型和 RH 因子的各自的抗原，并把这些珠子装配到用于在多重免疫检测中检测患者血清中抗体模式的阵列中进行。

ABO 血液分型是基于人类和大多数脊椎动物都携带有连接到某些膜蛋白的丝氨酸侧链上的复杂寡糖的事实。氧连接的多糖复合物经常暴露于人细胞的外表面，而且当携带它们的细胞被注射到不含相同细胞表面抗原的个体中时，会引发特异性免疫应答。这种失配个体中的相反的免疫应答

模式形成了 ABO 血组分类的基础，其中相同血型的个体可以接受另一个体的输血，而具有不同血型的个体则不能。在与 Rh 因子兼容或不兼容的基础上定义了其它血型，并用于联系血型与不同个体的遗传指纹。

血型反映了确定 A 和 B 血型的两个基因的表达。A 基因编码糖基转移酶，其催化末端 N-乙酰基-半乳糖胺 (Gal Nac) 残基添加到核心多糖上，而 B 基因编码使半乳糖 (Gal) 残基添加到相同位置的类似酶。当 A 和 B 基因存在时，两种结构都发现了，但当仅仅存在 O 基因时，暴露寡糖上的位点。AA 或 AO 个体的细胞携带 A 抗原，BB 或 BO 个体的细胞携带 B 抗原，而 AB 个体的细胞携带 A 和 B 抗原，OO 个体的细胞两种抗原都不携带。

因此，生物化学标记，其构成了患者病史档案的部分，如限定个体血型的细胞表面抗原组，能用于把遗传模式与患者身份联系起来。这种信息可通过芯片上的遗传测试获得，并与同时记录的生化 ID 标记联系起来，该生化 ID 标记可依次与现有的患者记录相互参照来确保真实性。

根据本发明的方法，靶核酸序列与遗传指纹的同时探测可通过以下方法进行，首先如这里的实施例中所详细描述，确定遗传 ID 标记或多种感兴趣的标记，然后，设计多种寡核苷酸探针：第一组针对感兴趣的靶核酸序列，第二组针对感兴趣的标记或多个标记。也就是说，第一组探针用于为遗传测试或模式分析所设计的检测中，第二组探针用于与遗传测试同时进行的分子指纹测定中。当使用内在标记时，在靶序列中内在标记与其它指定的多态性突变一起进行扩增。当使用外在标记时，需要用于分散但同时扩增的分散引物组。

含有多态性位点的扩增靶核酸序列，或多种用于遗传模式分析与遗传 ID 标记的这种序列的探测包括，通过使靶与编码的，序列特异性寡核苷酸探针退火形成杂交复合物，来测定探针与靶片断之间序列互补的程度。

靶基因与标记的探测，可用本发明范围内的各种方法来完成。例如，

在一个实施方案中（这里称为“直接杂交”），荧光标记的靶片断可用于把可检测的光学信号如荧光赋予杂交的探针-靶复合物。标记的靶可使用标记的引物通过前面的靶扩增容易地生产。在另一个实施方案中，直接杂交可使用标记的可检测探针。例如，如这里提供的实施例中所详细描述，荧光能量转移可用于通过形成三成员的杂交复合物来产生可检测的光学信号。在另一个实施方案中，靶序列组合物可从用于提供的探针组的酶介导的连接和延伸反应的信号模式测定。在延伸介导的检测中，聚合酶介导的延伸反应产生了反映设计用作引物功能的序列特异性探针的延伸或延长的信号模式。通过本发明的任何方法产生的信号模式包含连同“指纹”一起的唯一一个体基因型。

如这里所使用的，“杂交”是指第一核酸分子优选与特定的第二核苷酸分子结合，退火，双链或杂交。杂交复合物的稳定性随序列组成，长度和外部条件而改变。杂交方法包括那些依赖于控制反应条件的严紧性来破坏在混合物中形成的一些但是不是所有的杂交复合物的方法。使用这些方法，使形成杂交复合物的探针与靶序列之间完全互补的从部分互补的中区分开是可能的。

为了便于检测，杂交复合物可以进行修饰，使其含有一种或多种标记。这些标记可以通过本领域技术人员熟知的任何方法整合。可用于本发明的可检测标记包括通过分光镜，光化学，生物化学，免疫化学，电学，光学或化学方法可检测的任何组合物。本发明中有用的标记包括，高亲和力结合标记如用于使用标记的抗生蛋白链菌素或其共扼物进行染色的生物素，磁珠，荧光染料（例如，荧光素，Texas 红，若丹明，绿色荧光蛋白，等等），放射性同位素标记（例如， ^3H ， ^{125}I ， ^{35}S ， ^{14}C 或 ^{32}P ），酶（例如，辣根过氧化物酶，碱性磷酸酶和其它常用于 ELISA 的酶），表位标记和颜色标记如胶体金，着色的玻璃或塑料珠子（例如，聚苯乙烯，聚丙烯，乳胶，等等）。检测这些标记的方法是本领域技术人员熟知的。因此，例如，放射性同位素标记可使用照相软片或闪烁计数器进行检测，荧光标记可使用

光电探测器探测发射的光来检测。酶标记一般通过给酶提供底物并检测酶对底物作用所产生的反应产物来进行检测，颜色标记可简单地目测着色的标记来进行检测。一种方法使用胶体金作为标记，其可通过测定来自金的光散射来进行检测。标记可在杂交之前或之后添加给扩增产物。

“直接标记”是可检测的标记，其在杂交之前直接连接，或整合到核酸中。相反地，“间接标记”附着，或整合到杂交后形成的杂交复合物中。通常，间接标记在杂交之前连接到已经连接到扩增核酸上的结合部分上。因此，例如，扩增核酸可在杂交之前进行生物素化。杂交之后，共扼荧光团的抗生物素蛋白或抗生蛋白链霉素将结合含有生物素的杂交双链体，提供容易检测的标记。

检测与阵列中的探针杂交的标记核酸的方法是本领域技术人员已知的。例如，当使用颜色标记时，对标记进行简单的目测就足够了。当使用放射性同位素标记的探针时，检测放射性（例如，使用照相软片或固态检测器）就足够了。荧光标记靶核酸的检测可通过荧光显微镜方法来完成。杂交复合物阵列可使用光源以选择的特定荧光标记的激发波长来激发，并检测在发射波长得到的荧光。激发光源可以是，例如，适于荧光标记激发的激光。

在优选的实施方案中，探测步骤包括，靶退火的探针的延伸。该反应，由聚合酶催化，通过以反映靶序列在已存在的杂交复合物中的组成的次序，给探针序列添加一种或多种三磷酸核苷，来产生延伸的杂交复合物。为了进行该延伸反应，探针的长度必须包含末端延伸起始（“TEI”）序列。该 TEI 序列依次含有探测位点，其优选与 3' 末端相符，但可能被引物序列中的 3-4 个碱基从 3' 末端取代。如果探测位点的组成与指定的靶位点的组成相匹配就进行延伸。

通过添加选择的脱氧核苷三磷酸（dNTPs）以复制突出的单互补链并检测反应的特异性，来延伸双链 DNA 的 3' 凹缺末端的方法，在现有技术中已经有报道（Wu, Proc. Natl Acad. Sci, 57 (1) : 170-171 (1967), Wu, J. Mol.

Biol, 14: 35 (3): 523-37 (1968))。它们同时整合一种 dNTP, 达到在每个位置含有 4 个 dNTP, 并使用具有放射性标记的 dNTP 来检测成功的整合。

相较于现有技术, 本发明提供了平行模式的重复序列分析, 其中所有的延伸反应在由位点特异性探测探针与靶形成的多拷贝双链 DNA 中同时发生。在本发明的一些实施方案中, 为了便于平行检测成功的延伸, 可以使延伸反应产生的光学信号成像。

在本发明的实施方案中, 可提供两种或多种探针用于特异性指定位点的探测, 这些探针构建用于预期位于探测位点和接近指定多态性位点一定范围内的非指定多态性位点的多态性或突变。在优选的实施方案中, 这些多种探针序列含有至少一种与特异性靶序列在接近确保延伸范围内的所有位置都匹配的探针。

此外, 在本发明的一些实施方案中, 使用了覆盖探针组。覆盖探针组, (描述于美国临时专利申请 60/364,416, 其全部内容引入这里作为参考) 包含允许同时探测位于核酸序列内给定的多个指定多态性位点的探针, 并包括对于每个位点, 至少一种能与靶退火以允许在后续延伸反应的基础上分配两个可能值之一, “匹配的”(延伸) 或 “非匹配”(没有延伸) 给该位置的探针。

与每个指定位置相关的覆盖探针组可能含有两种或多种在一个或多个位置不同的探针。在某些实施方案中, 探针序列可含有能与 DNA 中的任何核苷酸形成碱基对的通用核苷酸。在某些实施方案中, 探针可连接到编码的微粒上, 而且特异性地, 覆盖探针组中的两种或多种不同类型的探针可连接到相同类型的微粒上。把两种或多种不同类型的探针连接到珠子上的过程称为 “探针合并”。

TEI 区域内一个位置的错配, 或在双链锚定 (“DA”) 区域 (也即, 退火的结果) 内三个或多个位置的错配足以阻止延伸。因此, 显示这种组成不同的两种探针的延伸, 通常产生不同的延伸模式。所有这些探针只要它

们是分离的，也即单独编码的，就可以在平行延伸反应中倍增。

显示相同 TEI 结果与显示至多两个位置不同的 DA 结果的探针，通常产生与最佳匹配相当或较低产率（和信号强度）的延伸反应。在第一种情况中，表示容许错配，由被讨论的探针匹配的等位基因组将扩展至包括那些 DA 区域内显示容许错配序列构象的等位基因。在第二种情况中，表示仅仅部分容许，这里描述了三种方法以进一步说明等位基因匹配模式。第一种方法中，那些在它们各自的 DA 区域显示一个或两个核苷酸多态性的探针包括于覆盖组中，并通过定量比较由覆盖组内不同探针所产生的信号强度来获得关于靶序列的信息。在第二种方法中，包括由系链（tether）连接的分离 TEI 和 DA 区域的探针，用于使 DA 区域更远离 TEI 区域以避免靶多态性。第三种方法中，探针被合适的合并，在该情况中只提供匹配等位基因组的适度延伸。

而这种提供或鉴定非指定的多态性位点的方法，尤其可用于联系序列特异性探针的倍增延伸，它也可用于小的测序反应（参见，如 Pastinen ,et al. *Genome Res.* 7: 606-614 (1997)，引入这里作为参考）。

在某些实施方案中，催化引物延伸的聚合酶是缺乏 3' 到 5' 核酸外切酶活性的 DNA 聚合酶。这种酶的例子包括 T7 DNA 聚合酶，T4 DNA 聚合酶，ThermoSequenase 和 Taq 聚合酶。当靶含有 RNA 序列时可使用反转录酶。除了聚合酶之外，也可以添加三磷酸核苷，优选所有四种碱基。例如，可添加 dNTPs 或类似物。在某些其它实施方案中，可以添加 ddNTPs。

成功的探针延伸，可通过与延伸的引物相关的固相载体的光学信号的变化来表示。这可通过本领域熟知的直接或间接标记方法来完成。（如标记核酸并检测标记的杂交核酸的方法的综述，参见 *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 24, P. Tigran Ed., Elsevier, NY (1993)）。

直接标记中，延伸反应产生了具有对应的光学信号的产物。在某些实

施方案中，荧光团或发色团染料可以连接到在延伸过程中被添加的一个或多个核苷酸上，以便延伸的引物获得特有的光学信号。前面已经描述了成功的延伸，其包括使用标记的脱氧核苷三磷酸(dNTPs)如 Cy3-dUTP 或者双脱氧核苷三磷酸 (ddNTP)。(Wu, 1968, 参见上述)。

在间接标记中，光学信号产生于延伸反应之后进行的附加步骤中。本发明还提供了提供用于检测成功延伸反应的光学信号的新方法，从而不需要标记的 dNTPs 或 ddNTPs，其是有利的，因为如果 dNTPs 或 ddNTPs 是标记的，可用的聚合酶的效率在调节 dNTP 或 ddNTP 过程中会降低。

本发明的方法进一步包括，形成三成员的杂交复合物，以及它们用于串联重复的平行探测中的应用，包括但不限于，组合连接和芯片上温度循环的方法。

在本发明的一个实施方案中，探测步骤利用形成的三成员杂交复合物。除了靶和探测探针外，该复合物包括修饰的探针，其在与退火的探测探针很接近的位置与靶退火。合适地，该修饰的探针可与探测探针连接。

在一个实施方案中，形成的三成员的杂交复合物包括，探测探针和设计成荧光能量转移对的读出探针。该读出探针的序列与靶的选择结果匹配，覆盖接近靶退火的探测探针 3' 端的一个或多个重复。换句话说，只有当探测探针的 3' 端与重复序列多态性内部的靶位点连接时，才形成三成员的杂交复合物。为了使三成员的杂交复合物中的探测探针和读出探针形成荧光能量转移对，探测探针被构建成在它的 3' 端或其附近含有第一个荧光团，称为供体，而读出探针被构建成在它的 5' 端或其附近含有第二个荧光团，称为受体。当供体和受体与靶退火时，一个通常不超过 3-4 个核苷酸的缺口使供体从受体中分隔开。在这些条件下，当三成员的杂交复合物形成时就发生从供体到受体的荧光能量转移，而当不形成三成员的复合物时就不发生。

荧光能量转移的例子如图 15a 和 15b 所示。两组寡核苷酸探针设计成

与重复序列之前的扩增产物中的序列互补。在该设计中，三成员的杂交复合物以这样的仅仅把含有正确重复数的探测探针置于很接近的读出探针中的方式形成，从而使荧光能量从供体转移到受体。

在该方式中，通过探测探针和靶形成的杂交复合物的“重复内末端”和“重复外末端”构象，在使用双重染色检测进行的平行探测检测中很容易区分。前者允许形成三成员的杂交复合物，从而允许荧光能量转移，从而以供体的激发波长照射产生受体的发射波长的荧光。相反，后者不形成三成员的杂交复合物，从而以供体的激发波长照射产生供体的发射波长的荧光。

本发明的探测方法，当使用覆盖多个靶序列重复的读出探针时，尤其可用于分析不需要单个重复分辨的长重复序列多态性。亨廷顿基因提供了一个例子，其含有位于蛋白编码序列的起始的 CAG 三核苷酸重复的多态性延伸。该疾病是由该基因的一个拷贝中该重复异常大的扩展而导致。本发明可用于鉴定患者样本中重复的扩展。

另一个实施方案包括，在第一个温度 T_1 ，形成在探测探针与读出探针间不存在缺口的三成员的杂交复合物，读出探针被设计成 3' 端或其附近含有荧光染料。形成三成员的复合物之后，读出探针于第二个温度 T_2 连接到探测探针上，而且杂交复合物于第三个温度 $T_3 > T_m$ (其中 T_m 是解链温度) 变性，以留下表示读出探针终止于起始杂交复合物的重复区段内部的荧光链。该方法可被推广到双重染色检测形式，其中具有不同荧光染料的第二个读出探针与重复序列多态性外部区域中的序列退火 (图 19)。

这种探测方法允许通过温度 T_1, T_2 和 T_3 之间的重复性循环，而进行线性的芯片上扩增，从而每个靶序列在连接产物形成中用作模板。

某种程度上，用于 STR 分析的方法，如本发明所述，是分析核酸中的重复序列单元和其它多态性的一般方法，它可用于分析来自除了人类来源之外的 DNA。

例如,可分析微生物 DNA (如细菌 DNA) 来进行包括鉴定药物抗性菌株的菌株分类,以及指导抗生素治疗的选择。DNA 标记如限制性长度多态性 (RFLPs) 和多态性串联重复 (STRs, VNTRs) 已经用于细菌和酵母菌株的鉴定。当菌株单独通过形态和生化标记不能区别或区分时,这尤其有帮助。一种特别的应用是炭疽杆菌 (炭疽) 的菌株分类,其是已知最单一同态的细菌种类之一。在位于 *vrrA* 基因的可变区的串联重复序列可变数的基础上已经鉴定了 5 个已知的菌株。

本发明 STR 分析方法的另一个应用,是在鉴定和筛选显示所需特性的特异性遗传品种的情况中。例如,这些遗传标记可用于标签感兴趣的特性,该特性通过未表征的遗传因子确定并具有紧密联系的明确的多态性基因座。已经用于标记辅助筛选 (MAS), 鉴定和植物品种保护,作物育种和指纹识别目的的几类 DNA 标记包括,单位点改变 (如 SNPs), 以及单和多基因座重复标记如 VNTRs, STRs, 与简单序列重复 (SSRs)。

从下面的实施例可以更好地理解本发明。不过,本领域的技术人员可以很容易地理解这里的特定方法和讨论的结论只是对后面的权利要求中所述的发明的解释。

实施例

实施例 1: 具有嵌入的遗传指纹的囊肿性纤维化突变模式

在该实施例中,对 CFTR 基因中的突变进行分析,以产生具有遗传标识嵌入板作为内部标记的遗传模式 (图 1)。大多数感兴趣的 CF 突变位于外显子 3 到 21 上 (图 2 和表 1)。如果 CFTR 基因内的多态性标记的频率对于 CF 患者相较于总的群体而言没有显著差异,这些标记的探测在与正常染色体相对的突变体的分析中没有产生偏差。

选择合适的引物确保位于 CFTR 基因内的序列多态性能与指定的突变同时扩增,以通过后面的杂交进行探查。在包括由编码的珠子组成的阵列的优选实施方案中,一组珠子用与感兴趣的 CFTR 突变互补的探针进行修

饰，第二组探针用与选择的内在多态性 ID 标记互补的探针进行修饰。（扩增的）患者样本的等分试样置于装配的珠子阵列中，以产生并记录，在单一步骤中，具有嵌入遗传 ID 的遗传模式。在包括随机编码珠子阵列的优选实施方案中，下述的是典型的检测条件。

用两种荧光染料的不同组合进行染色的珠子，用中性链亲和素（neutravidin）和生物素化的寡核苷酸探针进行功能化，后一个步骤在 TBS（10 mM Tris HCl, 0.5 M NaCl w/v）中于室温进行 45min。使用用 Cy5 标记 5' 的正向引物和用磷酸盐基团修饰的反向引物对靶进行扩增。引物序列设计成与特异性外显子最旁临的内含子序列，或与标记某些外显子的起始或末端的序列同源。这些引物的使用，允许靶基因与位于由选择的引物限定的基因座内的多态性标记同时进行 PCR 扩增。

外显子 7 的多态性:当含有 3 个 CF 突变（334W, 347P 和 1078T）的外显子 7，用它的末端旁临的引物组进行扩增时，位于 dbSNP 100083 C/T, dbSNP 100084 C/G, dbSNP 100085 A/G, dbSNP 1799834 C/G, dbSNP 1800086 C/G, dbSNP 1800087 A/G 位置的序列多态性也被扩增了（图 3 和 4）。这些多态性表示内在的 ID 标记，其通过与位于随机阵列内的编码珠子上的对应寡核苷酸探针组杂交而进行探测。

外显子 10 的多态性:最常见的 CF 突变， δ 508，位于外显子 10 上。位于外显子 10 上的 SNPs 包括 dbSNP 100089 C/T, dbSNP 100090 C/T, dbSNP 213950 A/G, dbSNP 100092 C/G, dbSNP 1801178 A/G dbSNP 180093 G/T, dbSNP 180094 A/G 和 dbSNP 1900095 G/T(图 5), 可与突变一起扩增，并通过与位于随机阵列中的编码珠子上的对应寡核苷酸组杂交而进行探测。

从患者样本提取的基因组 DNA，使用引物组在多重 PCR（mPCR）反应中进行扩增。进行囊肿性纤维化分析的 mPCR 方案的优选实施方案(L. McCurdy, Thesis, Mount Sinai School of Medicine, 2000, 其引入这里作为参考)如下所述。使用标签的嵌合引物进行多重 PCR，在其各自的 5' 末端，

用一个通用序列来缩小各自的解链温度范围。合成的 PCR 引物在 5' 末端含有 Cy5 或 Cy55(Amersham)标记。使用 Perkin Elmer 9600 热循环仪, 进行 28 个扩增循环, 每个循环由 10 秒的变性步骤, 于 94 ° C 以 48 秒倾斜 (ramp), 10 秒的退火步骤, 于 60 ° C 以 36 秒倾斜, 和 40 秒的延伸步骤, 于 72 ° C 以 38 秒倾斜组成, 每 50 μ l 的反应混合物含有 500ng 基因组 DNA, 1X PCR 缓冲液 (10 mM Tris HCl, 50 mM KCl, 0.1% Triton X-100), 1.5mM MgCl₂, 200 μ M 每种 PCR 级的 dNTPs 和 5 个单位的 Taq DNA 聚合酶。对每个引物对确定最佳的引物浓度。

扩增以后, 对产物进行纯化以除去所有的试剂, 并通过分光光度分析测定 DNA 的浓度。为了产生单链的靶, PCR 产物 (200ng) 用 2.5 单位的 (λ) 核酸外切酶在 1X 缓冲液中于 37° C 孵育 20min, 并于 75° C 灭活 10min。在这些条件下, 酶从 5' -磷酸化末端消化双链 DNA 的一条链, 并释放 5' 磷酸单核苷 (图 6)。产物直接用于杂交混合物 (2.25 M TMAC, 0.15% SDS, 3 mM EDTA) 中。

对于每个杂交反应, 5 μ l 的杂交混合物置于硅片的表面, 每个芯片携带显示寡核苷酸探针的编码珠子的阵列。芯片置于带盖子的盘中, 并置于 55° C 孵育器中的振动表面 (~200rpm) 中 15min。芯片通过用 1X TMAC 缓冲液冲洗阵列三次来进行洗涤。记录来自阵列中每个珠子的 Cy5 荧光信号, 并使用荧光显微镜、CCD 照相机和图像分析软件进行分析, 以确定突变模式。

实施例 2: 具有嵌入遗传指纹的杜兴氏肌营养不良突变模式

该实施例解释了用于杜兴氏肌营养不良 (DMD) 的遗传 ID 的设计, X-连接的隐性特性大部分出现在雄性中, 其特征在于肌肉力量的逐步衰退。尽管肌肉的 DMD 蛋白 (抗肌营养不良蛋白) 分析, 提供了精确的诊断测试, 但是它是有侵袭性的并且成本和风险高。而且, 由于蛋白在羊水或绒毛膜绒毛组织中不表达, 蛋白测试不适于产前诊断。另一方面, 已经克隆了 DMD 基因, 而且使用要求与 10 个 cDNA 探针杂交的 Southern 印

迹，鉴定了几种缺失（图 7）。(Kunkel, et al, Nature, 322: 73-77 (1986))。

多重 PCR 方法已经描述用于同时分析这些基因缺失。(Monaco et al, Nature, 323: 646-650 (1986))。内在遗传 ID 可以得自位于 DMD 基因内多个位点的几个二核苷酸重复多态性组。图 7 阐述了用于分析 DMD 基因内的基因缺失和多态性 ID 标记的方法。PCR 引物设计成位于外显子序列的旁临，以致旁临序列内存在缺失而阻断了靶的扩增，导致后续探测中无效的结果。

实施例 3: 遗传模式与遗传指纹记录的匹配

提供通过已知的 DNA 指纹识别方法记录的遗传指纹，如，例如，用作部分新生儿测试或那些用于整个选择的群体的方法（例如，防卫部队的成员或监狱犯人），本发明的方法提供了确认遗传指纹并同时进行遗传测试的方法。例如，标记，如 STRs，其被记录用于亲子关系和法医检定分析。此外，可以使用得自 SNP 基因分型或单倍体分型中的标记。从而，具有嵌入遗传 ID 的遗传模式通过匹配遗传 IDs 很明确地与特定的患者记录相关联。例如，根据本发明的方法，人的遗传指纹可以与其它的遗传数据一起记录到数据库中。如果个体接着进行后续的遗传测试（例如，为了任何遗传疾病或单倍体分型），第二个测试的结果可通过比较与第一个和第二个测试相关的遗传指纹而明确证实。

实施例 4: TH01 中 STR 长度多态性的芯片上鉴定

HUMTH01 基因座在酪氨酸羟化酶基因中含有四核苷酸重复(CATT)。为了测定这些 STR 多态性的长度，合成的寡核苷酸探针，含有设计成与靶的尾巴序列互补的六个核苷酸的 5' 锚定序列，以及设计成与靶的前导序列互补的六个核苷酸的 3' 锚定序列，如下：

5' ttc cct... .cac cat 3'

为了测定靶中重复序列的长度，提供了一组寡核苷酸探针，其含有一个长度匹配的探针，因此在重复 CATT 序列单元的数量内，与任何预期

的互补靶重复数匹配。在该设计中，只有含有正确数量的四聚物重复的探针，将与靶形成杂交复合物，其中 5' 锚定和 3' 锚定都能分别与靶的尾巴和前导序列适当连接，而且只有正确连接的探针被延伸从而产生阳性检测信号。

合成寡核苷酸探针，其含有 17 个核苷酸而且生物素标记通过 12-C 间隔区方式连接到 5' 末端（生物素-TEG）（图 9a），并使用商业销售商提供的反相 HPLC 方法（Synthegen TX）进行纯化。使用生物素部分，探针连接到涂有抗生物素蛋白链霉素的编码珠子上，该珠子装配到硅片中的平面阵列中。使含有 10mM Tris-HCl (pH 7.4) 1mM EDTA, 0.2 M NaCl 和 0.1% Triton X-100 的杂交混合物中的 1 微摩尔含有 6 个 CATT 重复的单链靶，与随机编码的珠子阵列接触并于 50° C 孵育 20min。杂交混合物然后被含有 3U 的 Thermo Sequenase (Amersham Pharmacia Biotech NJ)，和与具有 TAMRA 标记的双脱氧核苷酸 (ddNTP) 类似物 (NEW Life Sciences) 的 1X 酶缓冲液的混合物取代，并于 60° C 进行探针延伸 3min。退火和延伸反应可在单个步骤中通过添加 ddNTPs 或 dNTPs 到混合物中并于两个温度进行反应来进行。珠子阵列然后用蒸馏水洗涤 15min，含有来自阵列中每个珠子的荧光信号的图象使用荧光显微镜和 CCD 照相机进行记录。分析图象来确定每个延伸（和荧光）探针的身份。如图 9b 所示，记录的最强的信号来自显示 6 个四聚物的探针 (ST-6) 的珠子，表明该探针含有与靶重复数相匹配的正确重复数。

实施例 5: 使用含有 3' 锚定（“钩状”）序列的探针在芯片上确定 STR 长度多态性

设计的含有短 3' 锚定（“钩状”）序列（但不是 5' 锚定序列）的寡核苷酸，被连接到编码的珠子上（图 10a）。特异性地，分别含有 3, 4 和 6 个四聚物重复的探针，被设计用于探测含有由 5' 前导序列和 3' 尾巴序列旁临的 6 个四聚物重复的靶片断。

在该设计中，所有的探针将与靶形成杂交复合物，其中探针的 3' 锚

定序列与靶的前导序列适当地连接。但是，通过设定检测温度为超过除最长的探针以外的所有探针的解链温度的值，只有含有该探针的杂交化合物保留，而且仅该探针被延伸以产生阳性检测信号。

如这里前面所述的，在存在 DNA 聚合酶和 ddGTP 时进行单碱基延伸。如所期望的，记录的最强的信号来自显示 6 个四聚物探针的珠子(图 10b)。这些结果证实通过本发明的组合物和方法能产生高水平的区分。

实施例 6:使用双重染色检测在芯片上鉴定 STR 长度多态性

寡核苷酸的设计如实施例 4，但是单碱基延伸反应在存在分别用 TAMRA(绿色)和荧光素(橙色)(NEW Life Sciences)标记的两种 ddNTPs 时进行(图 11a)。如果退火的探针终止于重复序列的外部(外部标记)，绿色 ddNTP 被整合，而如果退火的探针终止于重复序列的内部(内部标记)橙色 ddNTP 则整合。含有 6 个四聚物重复的靶被添加到溶液中，并如前所述的进行检测和综合分析。如前，记录在由 6 个四聚物探针 ST-6 的延伸所产生的绿色信道中的强信号，表明该探针含有正确的重复数。此外，由 3 个四聚物探针 ST3 和 4 个四聚物探针 ST4 的延伸所产生的橙色信道中的强信号，表明这些探针终止于重复序列的内部而且含有错误的重复数(图 11b)。这种双重染色形式通过产生组中所有探针的阳性信号提供了检测结果的附加的确认。

实施例 7:锚定长度对探针滑动的影响

合成的两组寡核苷酸探针，除了 3, 4 和 6 个四聚物重复外，含有两组探针分别有 6 和 11 个核苷酸的 5' 锚定序列，以及相同的 3' 锚定序列，如下：

5' ttc cct... .. cac cat 3'

5' ctt att tcc ct... .. cac cat 3'

如前所述，5' 锚定序列设计成与重复旁临的靶尾巴序列互补。为了测定靶中重复序列的长度，提供的一组寡核苷酸探针含有一个长度匹配的

探针,从而在重复 CATT 序列单元数内,与任何预期的互补靶重复数匹配。

使用如前所述的反应条件(实施例4),含有较长的5'锚定序列的探针产生了高于含有短的5'锚定序列的信号水平,表明较长的探针形成的杂交复合物的稳定性较高。(图12)。锚定序列可以改变以适应实验的需要。

实施例8:退火温度对 STR 多态性分析的特异性的影响

设计寡核苷酸探针,单碱基延伸的进行如实施例5所述,但是于两个不同的退火温度,即于37 °C和50 °C。对50 °C获得的结果进行的图像分析,显示在含有正确重复数的ST-6探针与含有不正确重复数的探针之间区别较大,而对37 °C获得的结果进行的图像分析显示基本上没有区别(图14)。选择正确的检测温度能相当大地增强检测的特异性。

实施例9:同义和反义探针

本实施例示例了使用既不含5'也不含3'锚定序列的寡核苷酸探针进行同义和反义DNA链的分析。含有这种锚定序列与同义和反义DNA两者中的尾巴和前导序列互补的寡核苷酸探针,也可用于其它实施方案中。两组探针都连接到编码的珠子上,该编码的珠子分别装配到置于相同芯片载体上的两个硅片表面。用两组引物扩增靶,每个靶含有6个重复以及一个3'尾巴序列和一个5'前导序列。如实施例5中所述的,单碱基延伸当存在DNA聚合酶和ddNTPs时进行。在该设计中,只有探针含有与靶重复数相匹配的正确重复数时才发生延伸。

实施例10:聚T鉴定

设计寡核苷酸探针用于CFTR基因中内含子8聚胸苷(T)束(tract)变异体(5T,7T,9T)的鉴定。合成每个探针并使用由商业销售者(IDT)提供的反相HPLC进行纯化,以含有通过12C-间隔区(生物素-TEG)的方式连接至5'末端的生物素标记。探针被设计成含有与靶的3'尾巴序列匹配的5'锚定序列,以及与靶的5'前导序列匹配的3'锚定(“钩状”)序列。锚定序列长度在4和10个碱基间变化;更长的锚定序列长度也是可

能的(图16)。含有可变重复数的探针固定于前面所述的编码珠子上。

在该设计中,当DNA聚合酶存在时,只有于靶序列连接并与靶重复数匹配的探针被延伸。在延伸步骤中,提供dNTPs,至少一种荧光标记的dNTP来产生荧光标记的延伸产物,或者一种或多种标记的ddNTPs用于进行如前面实施例所述的单碱基延伸。通过如本申请中所述的阵列的瞬时成像来记录信号。使用含有不同长度聚-T变异体的靶得到的实验结果,证实各种聚-T变异体的鉴定是可以的。

实施例 11:CFTR 基因聚 T 变异体的鉴定

解链温度,含有nT变异体的给定的杂交复合物的 $T_m^{(n)}$,反映了变异体的长度以及探针与靶之间不匹配的程度。也就是说,通过设定检测温度使除了那些含有与靶中的聚-T序列长度匹配或超过其的探针的杂交复合物外的所有杂交复合物不稳定,从而达到最佳的区分。例如,通过设定检测温度T以满足 $T_m^{(5)} < T < T_m^{(7)}, T_m^{(9)}$ 的条件,很容易鉴定含有7T变异体的靶。

为了进行检测,1 μ mol的靶被添加到退火和延伸混合物中,其含有10 mM Tris-HCl (pH 7.4) 1mM EDTA, 0.2 M NaCl, 0.1% Triton X-100, 3 U的Thermo Sequenase (Amersham Pharmacia Biotech NJ), 以及TAMRA-标记的双脱氧核苷酸(ddNTP)类似物(NEN Life Sciences)。

退火和延伸之后,珠子阵列用蒸馏水洗涤15min。使用荧光显微镜和CCD照相机记录含有来自阵列中每个珠子的荧光信号的图象,并对图象进行分析以确定每个延伸引物的身份。图17中的结果证实通过本发明的组合和方法完全能鉴定靶中的聚T变异体。

实施例 12: 设计梯级探针用于长重复的分析

本实施例解释了设计探针用于分析很多遗传疾病中普遍存在的长靶重复的方法。该方法包括“基数-补偿(base-offset)”结构,其中第一组“基数”探针构建成含有1N,2N,3N,...(N>1)重复,而且这些“基数”探针连

接到分离的编码珠子上。第二组“补偿-探针”构建成含有全部 $N+n$ 重复，其中 $n < N$ 。重要地，为了使所需的编码数最少，含有相同补偿的所有探针都连接到相同类型的编码珠子上。图 18 描述了具有公因子如 5' 锚定序列的两种分离的探针设计。

当诊断不需要进行单重复分析时，该方法尤其有用。例如，在诊断对某种疾病状态（如亨廷顿氏疾病）的诱因时，关键是要使用单重复分析来确定仅仅在 35 与 41 之间的重复数，确定病理学可能性的临界范围。也即，具有超过 41 个重复的患者将发生该疾病，而具有少于 35 个重复的患者则不会，病理学的可能性随 35 与 41 之间的重复数而增加。

例如，为了测定为 7（即 $N=7$ ）的分析中靶重复数达到 42，6 个“基数-探针”被构建成分别含有 7, 14, 21, ..., 42 个重复。这些探针连接到编码珠子上。在前面的实施例中所述的检测条件下，使除了那些含有重复数与靶重复数匹配或超过靶重复数的探针的杂交复合物外所有的杂交复合物都不稳定。例如，通过设定检测温度， T ，到超过 $T_m^{(35 \text{ 个重复})}$ 的适当高值，在检测之前可以容易地确定很精确的数值，含有 32 个靶重复的样本产生仅针对含有 35 和 42 个重复的基数探针的信号，而含有 39 个靶重复的样本产生仅针对含有 42 个重复的基数探针的信号。在第一种情况中，不需要进一步的分析。在第二种情况中，靶被确定含有至少 35 个重复，使用一组“补偿探针”来确定确切的数量。因此，补偿， $n < 7$ ，在 1 与 6 之间变化，显示相同补偿数的探针分为一组。也即，具有 $0+1, 7+1, 14+1, \dots, 35+1$ 个重复的那些探针，具有 $0+2, 7+2, 14+2, \dots, 35+2$ 个重复的那些探针，等等直到具有 $0+6, 7+6, 14+6, \dots, 35+6$ 个重复的那些探针被分组（“集中”），并连接到相同类型的编码珠子上。在上述的检测条件下，含有 39 个重复的样本产生对于补偿探针 $n = 4, 5$ 和 6 ，但不针对补偿探针 $n = 1, 2$ 和 3 的信号，因为前面 3 组分别含有具有 $35+4, 35+5, 35+6$ 个重复的探针，而后 3 组中表示的最大重复数， $35+3$ ，与靶重复数不匹配或不超过靶重复数。这组检测读数确定了确切的靶重复数。

相关的设计使用 2 组“基数探针”，第二组相对于第一组而改变。在上述的实施例中，改变为 $\Delta N=3$ 的第二组基数探针，含有具有 $0+3$, $7+3$, $14+3$, ... $35+3$ 个重复的探针，每个都连接到编码珠子上。对于具有 39 个重复的患者样本，第一组基数探针产生仅针对具有 42 个重复的基数探针的信号，从而把靶重复数设为 >35 ；第二组基数探针根本不产生信号，从而把靶重复数设为 >38 。这种使用倍数改变补偿探针组的可选择设计对确定靶的重复数尤其有用。

实施例 13:通过连接和芯片上循环鉴定多态性重复

本实施例中，设计用于鉴定 TH01 基因座中的重复数的实验，在由 2 个探针与靶形成的三成员的杂交复合物中，使用连接来把标记的检测探针连接到邻近的固定捕捉探针上。

设计的检测探针与靶的前导序列（如实施例 1）和部分重复序列互补。在第一种温度， T_1 ，靶将与固定的捕捉探针退火。添加检测探针，于第一种温度，或优选的第二种，更高的温度， $T_2 > T_1$ ，并允许与靶的前导序列退火，但是只有当它含有正确的靶重复数时才被连接。选择在第三种温度， $T_3 > T_1, T_2$ ，以使非连接的三成员杂交复合物不稳定，只有在那些保留连接的检测探针的珠子上才留有信号（图 19a）。这些珠子鉴定了具有正确重复数的探针。

在上述设计的变化中，该检测通过 $T_1 < T_2 < T_3$ 的顺序进行多次循环，以使每个单独靶链介导多个连接反应。在确保捕捉和检测探针过剩的情况下，循环导致了线性的信号扩增。该检测可在具有瞬间成象的轻便热循环仪中于任何温度进行。

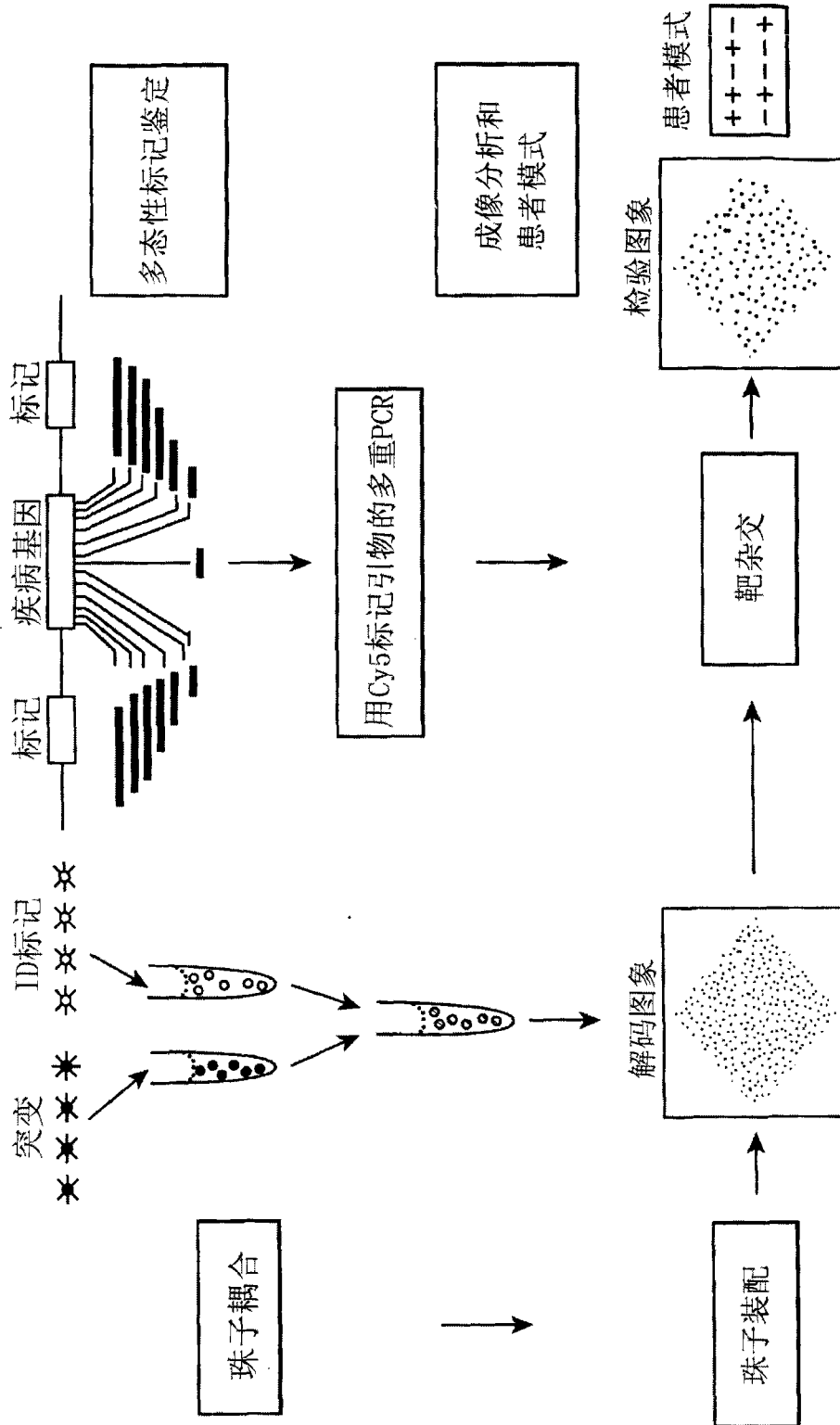


图1

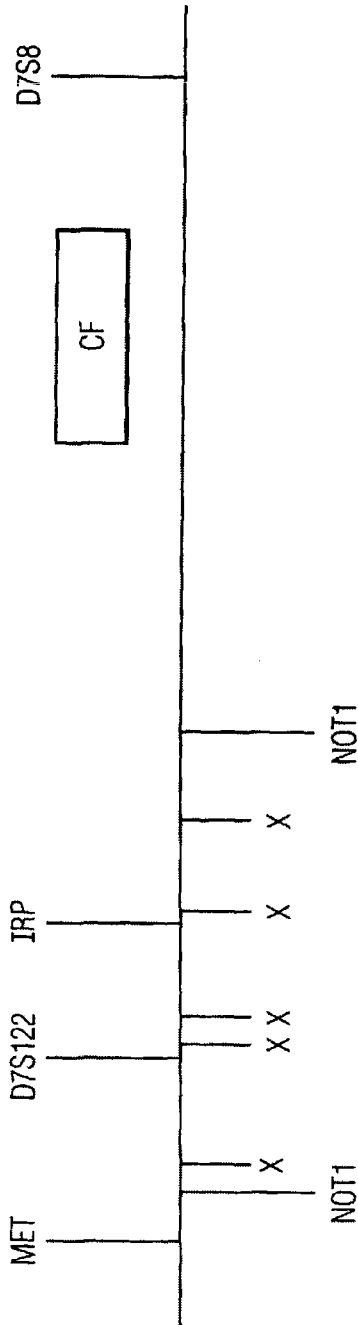


图2

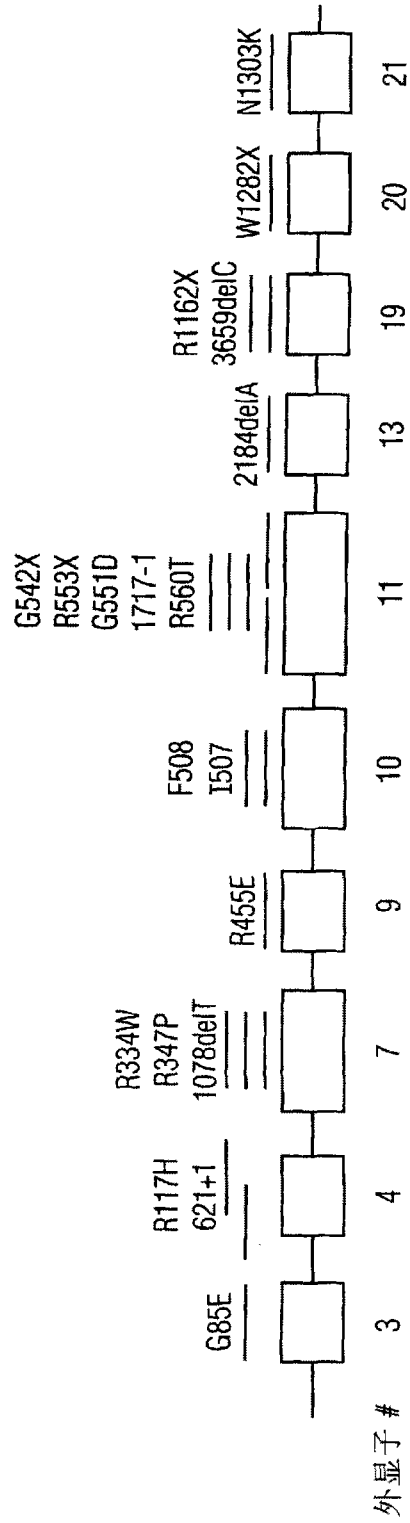


图3

1 ttacaagta ctacaagcaa aacactggta ctttcattgt tatctttca tataaggtaa
 61 ctgaggccca gagagattaa ataacatgcc caaggtcaca caggcatat gatgtggagc
 121 caggtaaaa atataggcag aaagactcta gagaccatgc tcagatcttc cattccaaga
 181 tccctgatat ttgaaaaata aaataacatc ctgaatttta ttgtattgt tttttataga
 241 acagaactga aactgactcg gaaggcagcc tatgtgagat acitcaatag ctacagCcttc
 301 ttCttct caG GgttCTttgT ggtg tttTtA tctgTgcttc cctatgcacT aatcaaAGga
 361 atcatcctcC GgaaaATatt caCcacAtc TCattctgCa ttgttcTgCG catggcggtc
 421 actCGgCaAt ttccctGGgc tgaCAaaCa TgGtatgaCT ctcTtggagc aataaacaaa
 481 atacaggtaa tgtaccataa tgctgcatta tatactatga tttaaataat cagtcaatag
 541 atcagttcta atgaactttg caaaaatgtg cgaaaagata gaaaaagaaa tttcctcac
 601 taggaagtta taaaagttgc cagctaatac taggaatgtt cacctaaac ttttcctagc
 661 atttctctgg acagtatgat ggatgagagt ggcatttatg caaattacct taaaatccca
 721 ataatactga ttagctagc agctttgaga aa

图4

1 cactgtagct	gtactacctt	ccatctctc	aacctattcc	aactatctga	atcatgtgcc
61 cttctctgtg	aacctctatc	ataaactgtg	tcacactgta	ttgtaaltgt	ctcttttact
121 ttcccttgta	tcttttgtgc	atagcagagt	acctgaaaca	ggaagtattt	taaataattt
181 gaatcaaatg	agttaataga	atctttacaa	ataagaatat	acacttctgc	ttaggatgat
241 aattggagggc	aagtgaatcc	tgagcgtgat	ttgataatga	cctaataatg	atgggtttta
301 ttccagact	tcaCttctaa	tgAtgattat	gggagaactg	gagccttcag	agggtaaaat
361 taagcacagt	ggaagaattt	cattctgttc	tcagtittcc	tggattatgc	ctggcaccat
421 taaagaaaat	AtCAtctTtg	gtgtttccta	tgatgaatat	agatacagaa	gcgctatcaa
481 agcatgccaa	ctagaAgagG	taagaaacta	tgtgaaaact	ttttgattat	gcatatgaac
541 ccttcacact	acccaaatta	tatatttggc	tccatattca	atcggttagt	ctacatatat
601 ttatgtttcc	tctatgggta	agctactgtg	aatggatcaa	ttaataaac	acatgaccta
661 tgcttaaga	agcttgcaaa	cacatgaaat	aatgcaatt	tatttttaa	ataatgggtt
721 catttgatca	caataaatgc	attttatgaa	atggtgagaa	ttttgtcac	tcattagtga
781 gacaaacgtc	tcaatgggta	tttatatggc	atgcatatag	tgatatgtgg	t

图5A

野生型

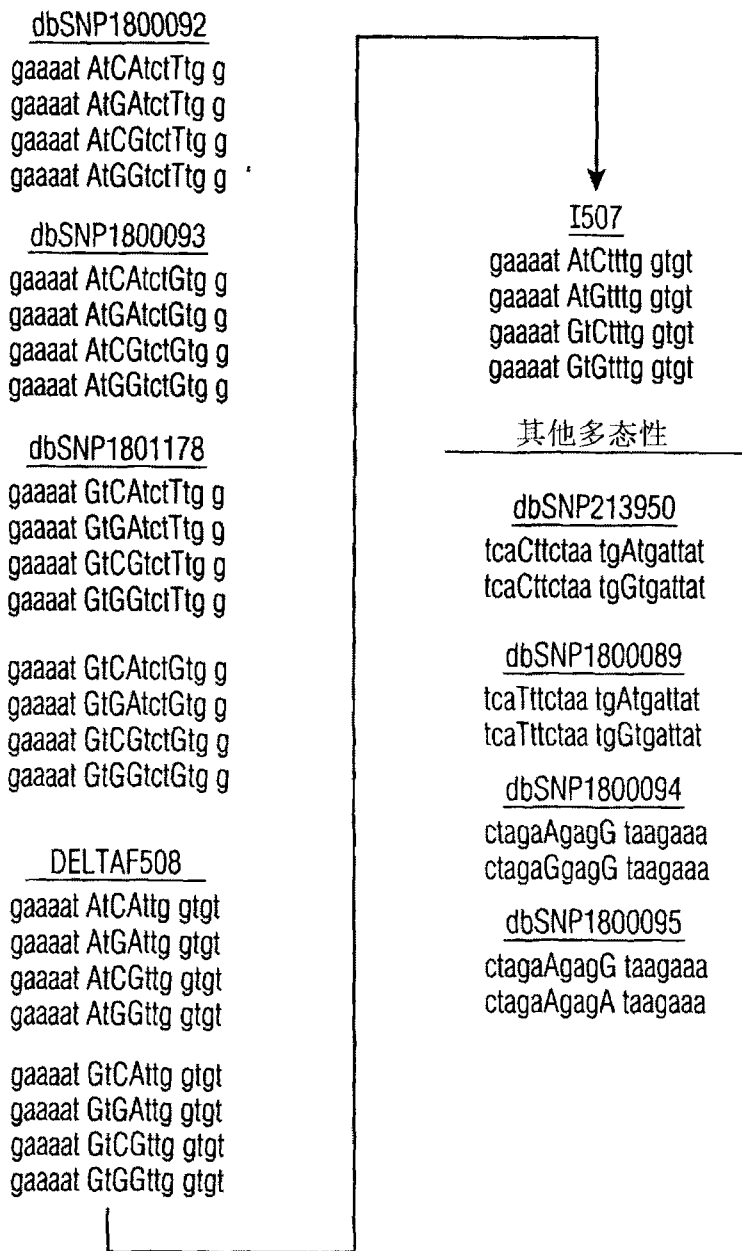


图5B

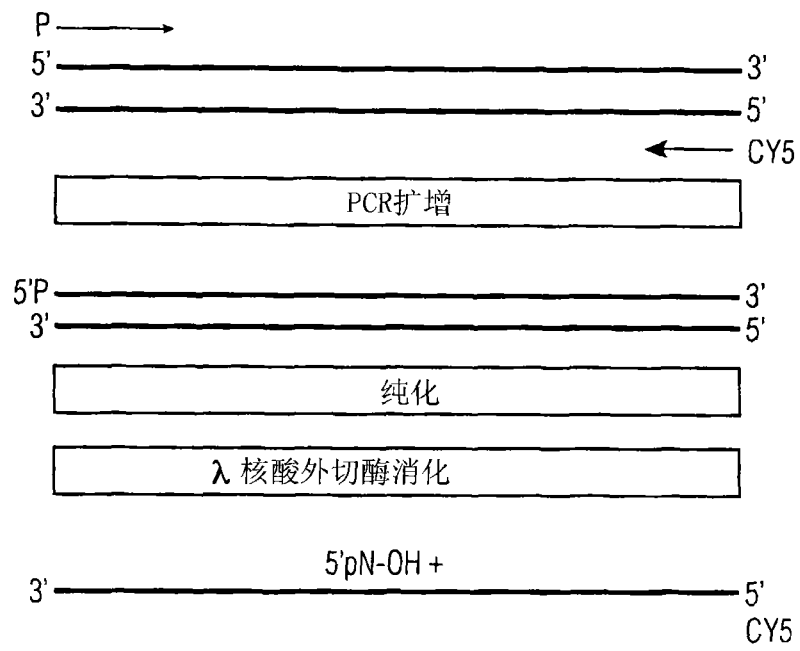


图6

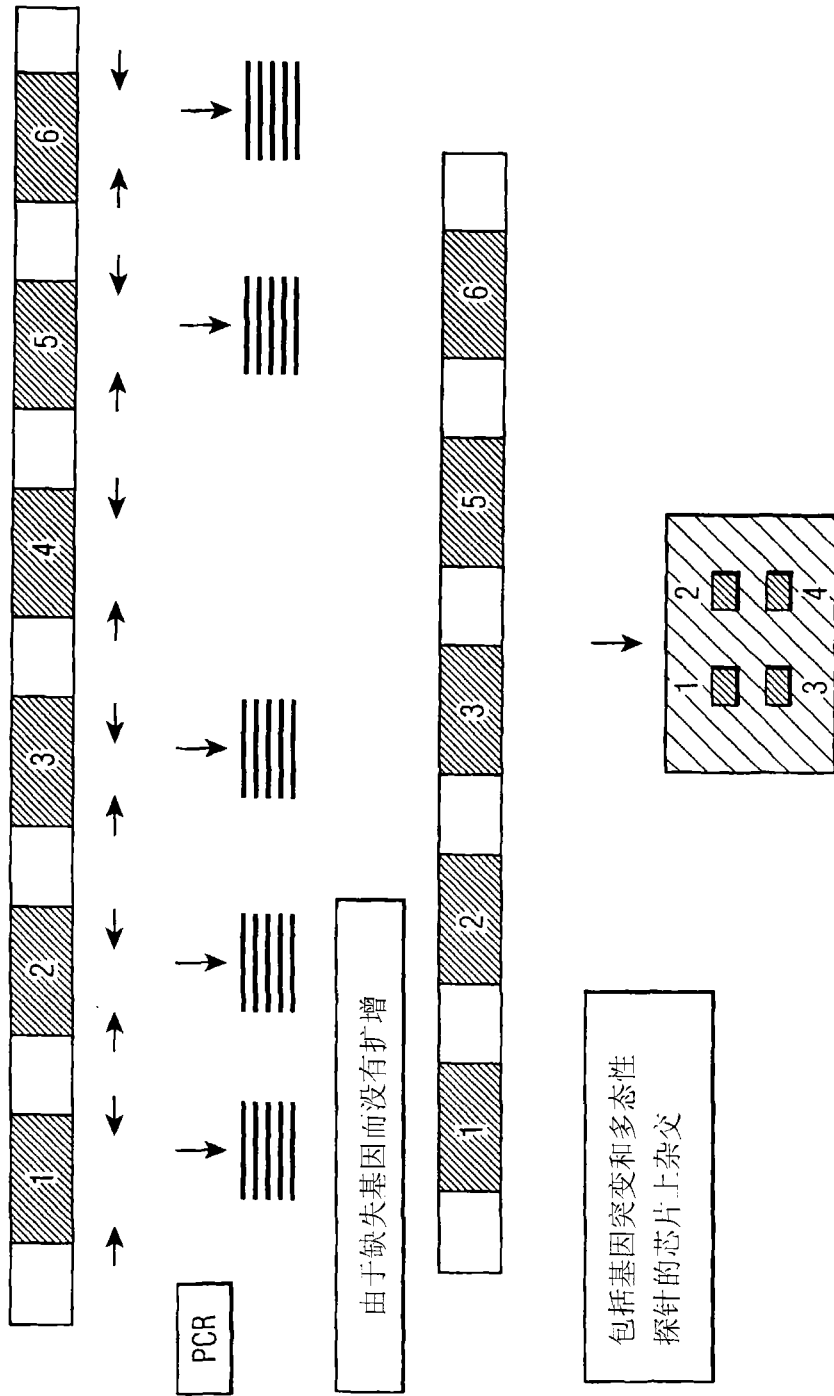


图7

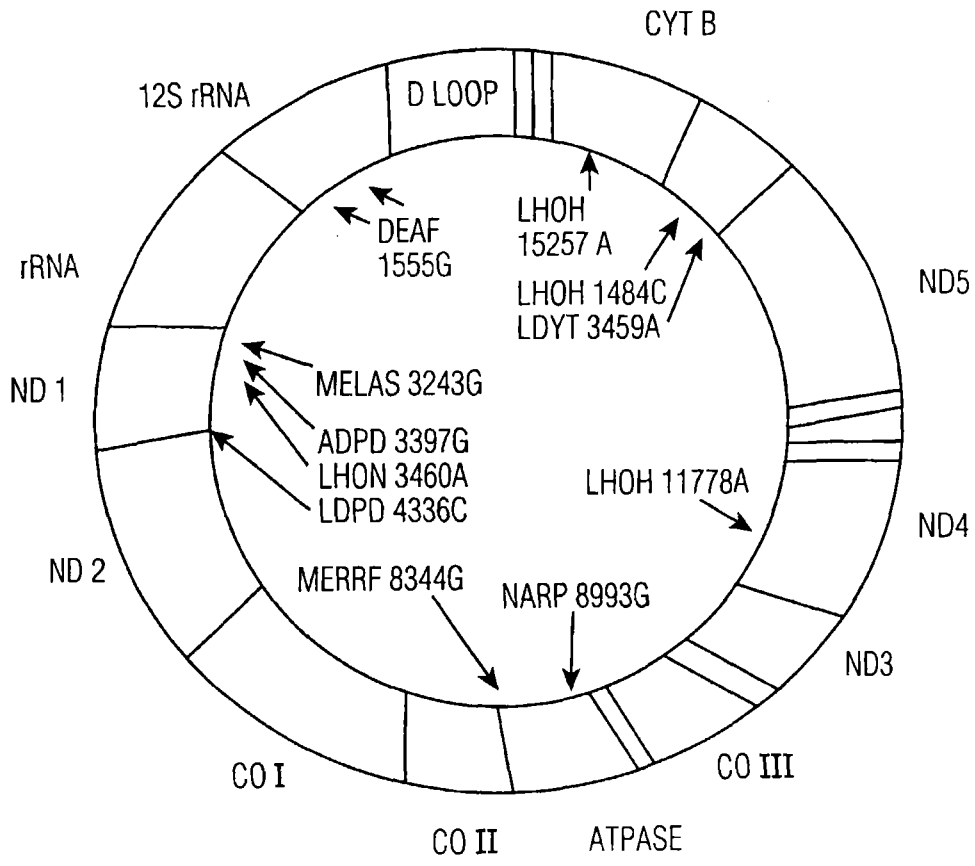


图8

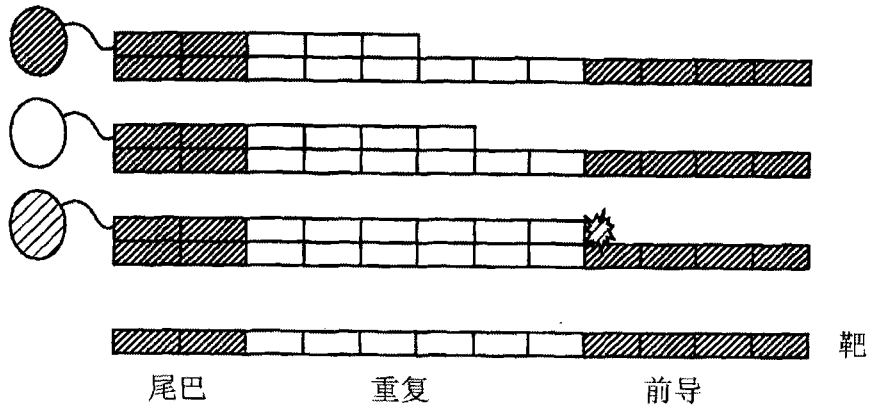


图9A

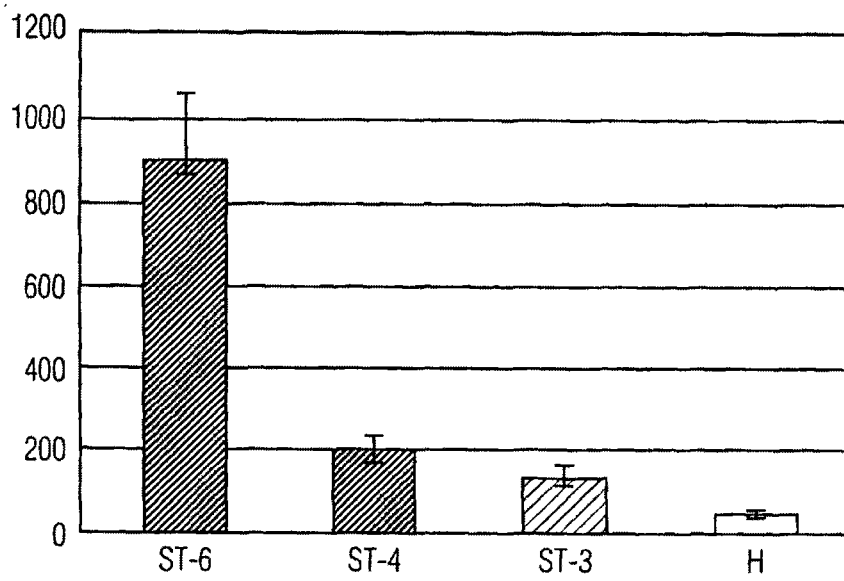


图9B

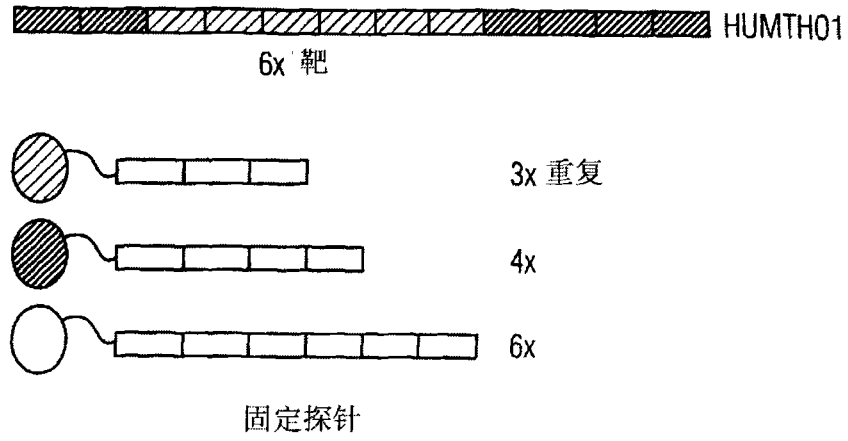


图10A

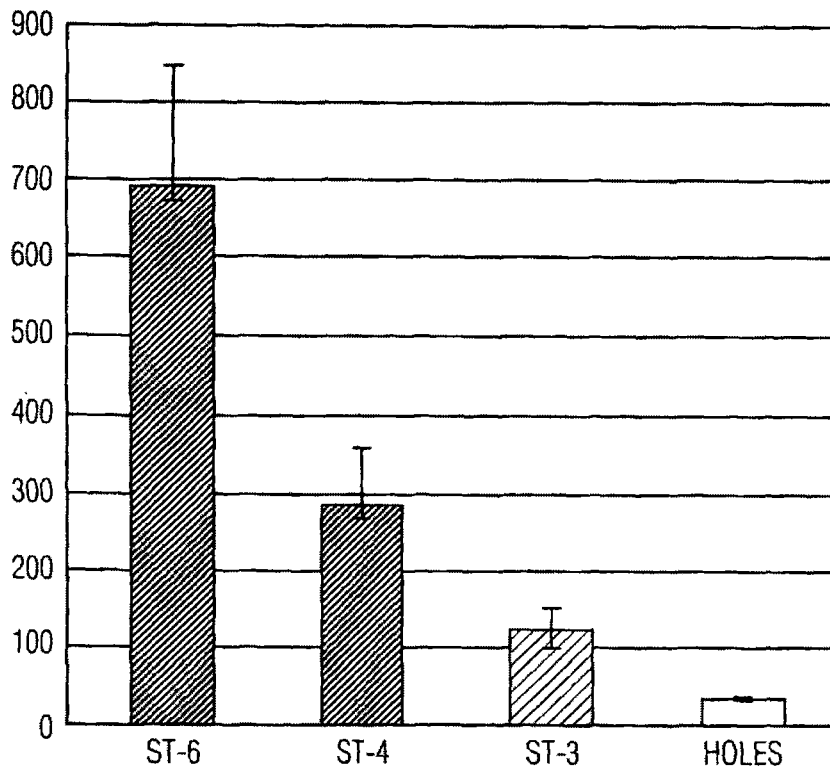


图10B

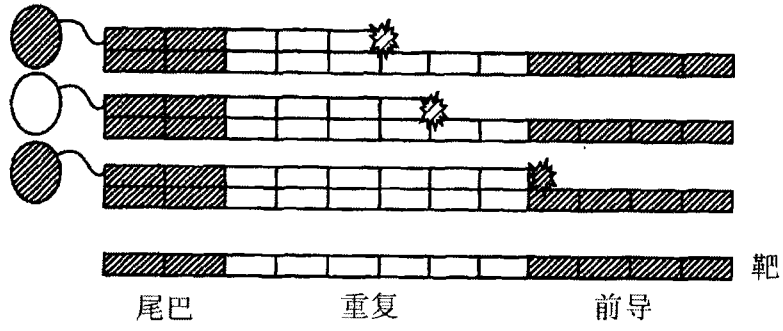


图11A

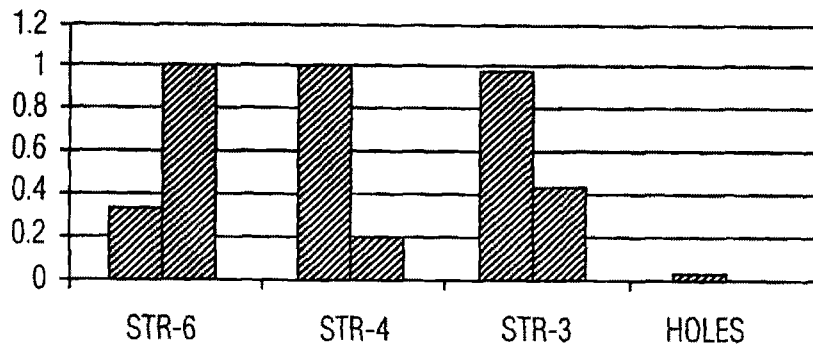
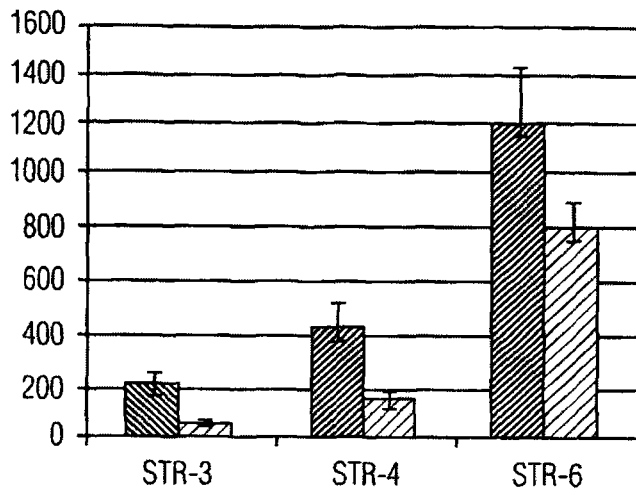


图11B



带有6和11个核苷的探针序列

长锚定
 5' ctt att tcc ct-----cac cat 3'

短锚定
 5' ttc cct-----cac cat 3'

图12

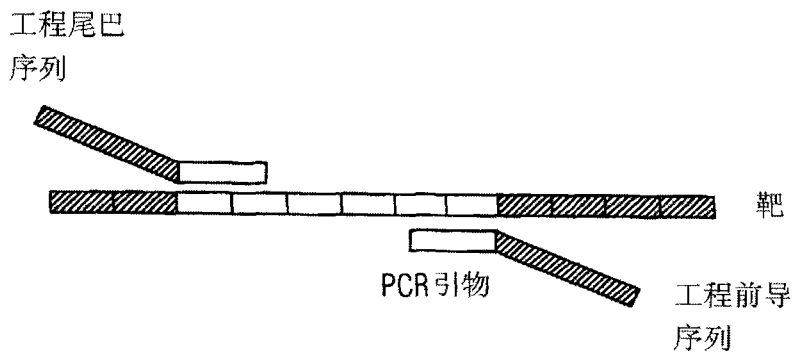


图13

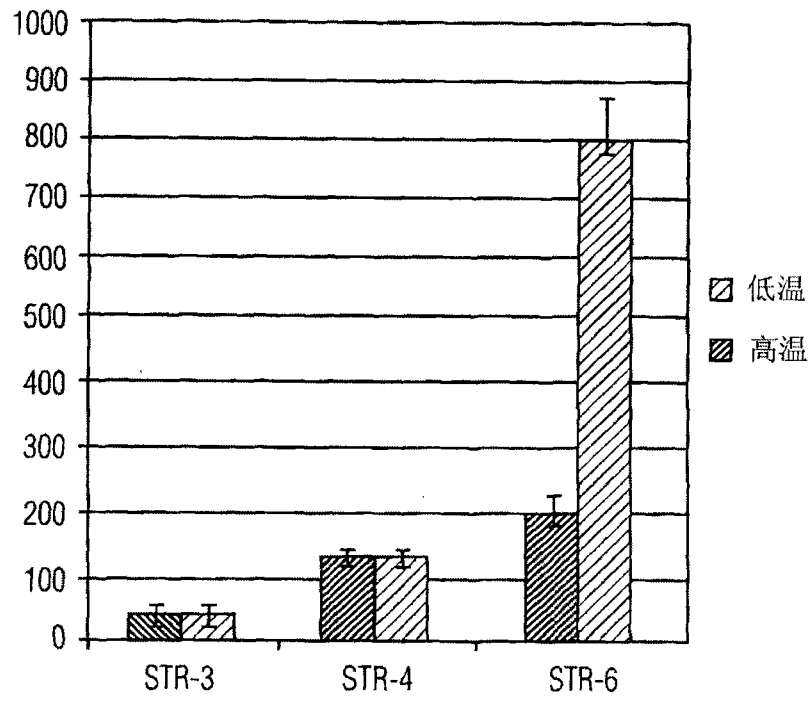


图14

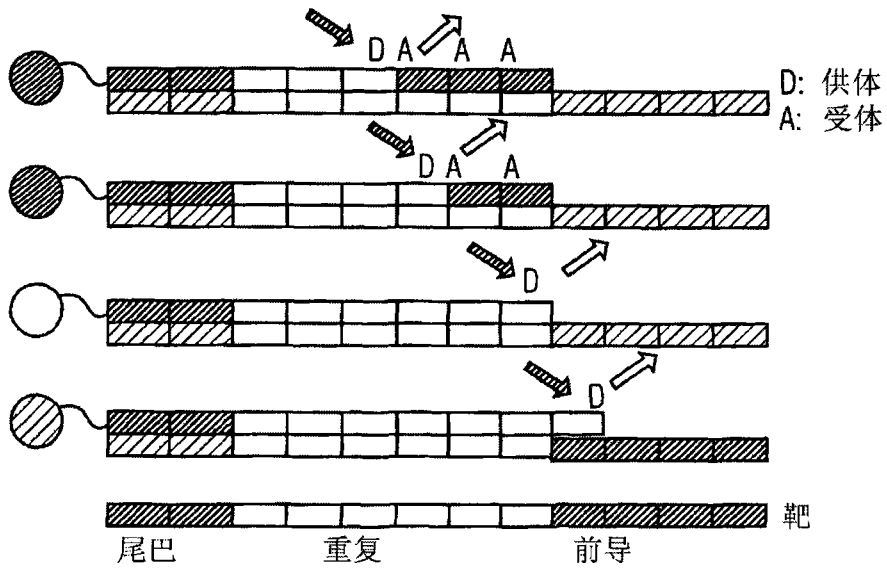


图15A

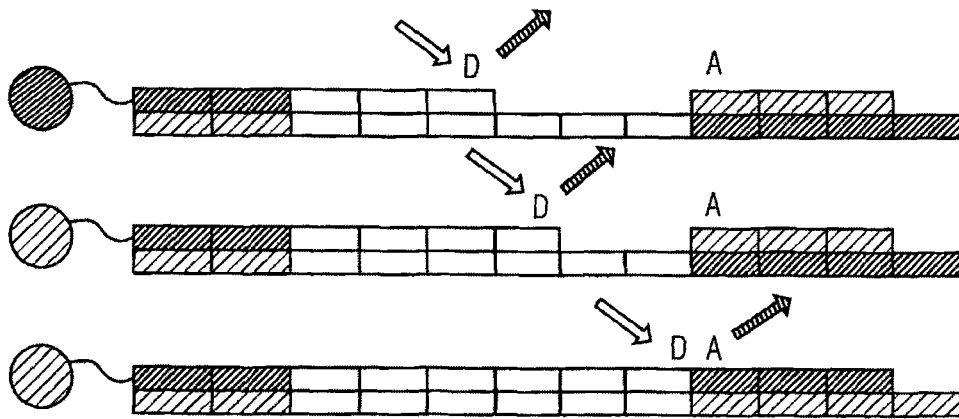


图15B

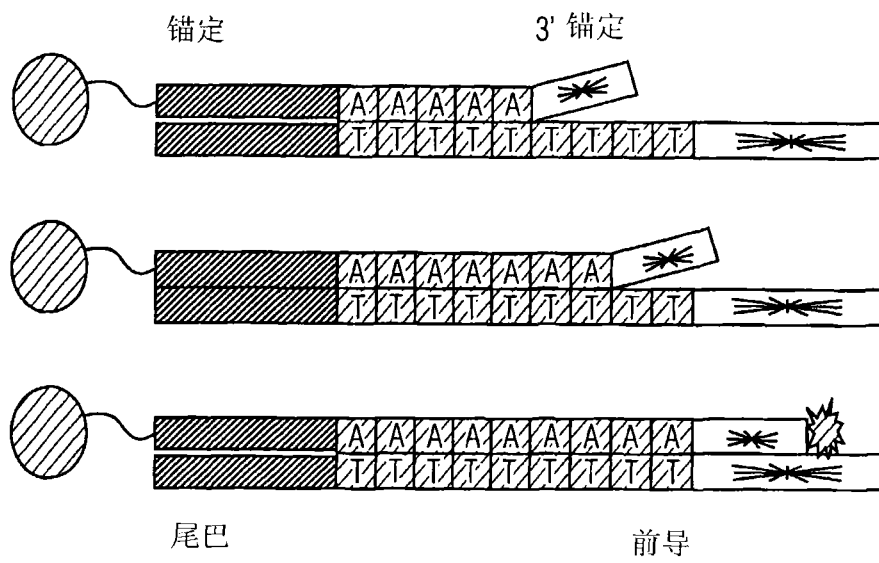


图16

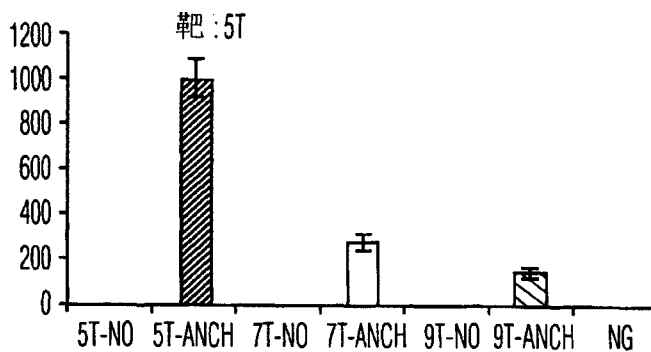


图17A

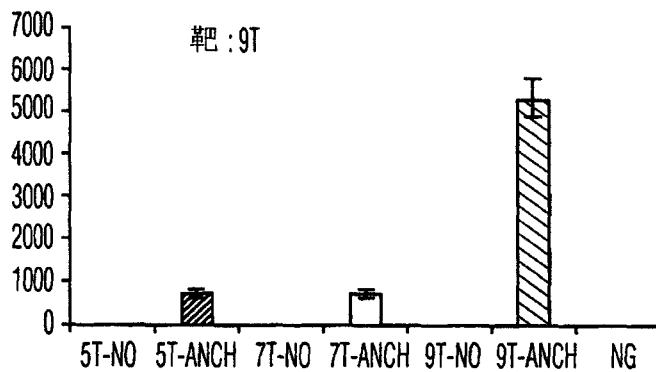


图17B

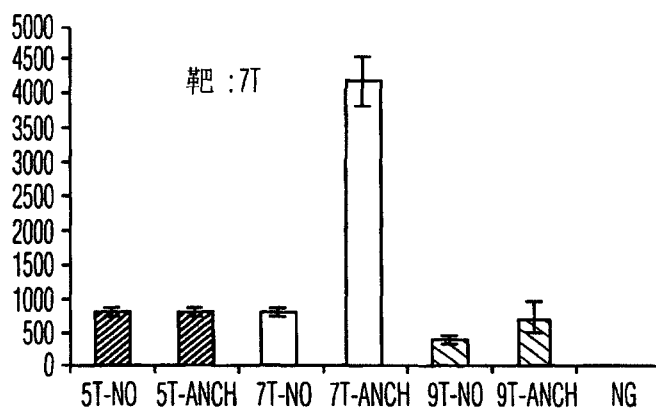


图17C

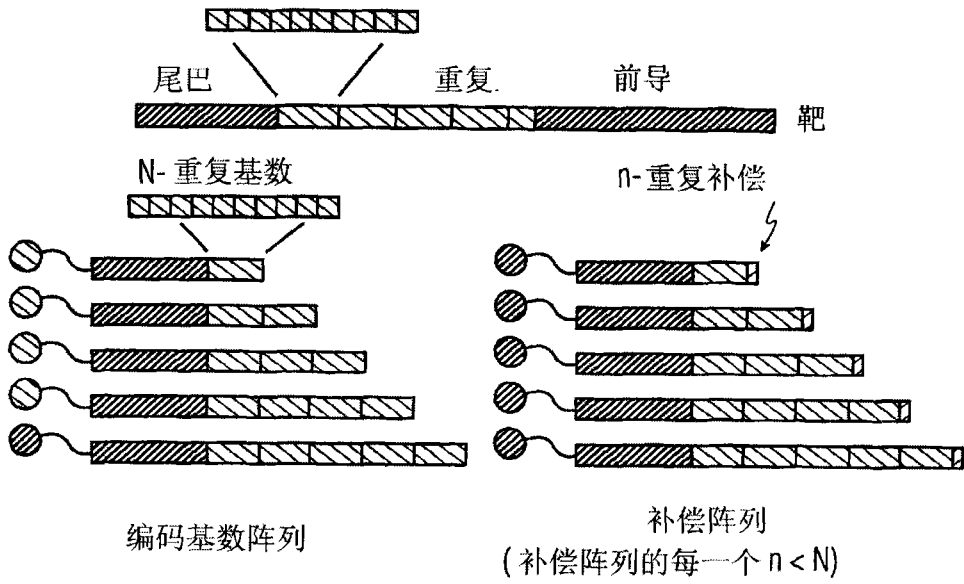


图18

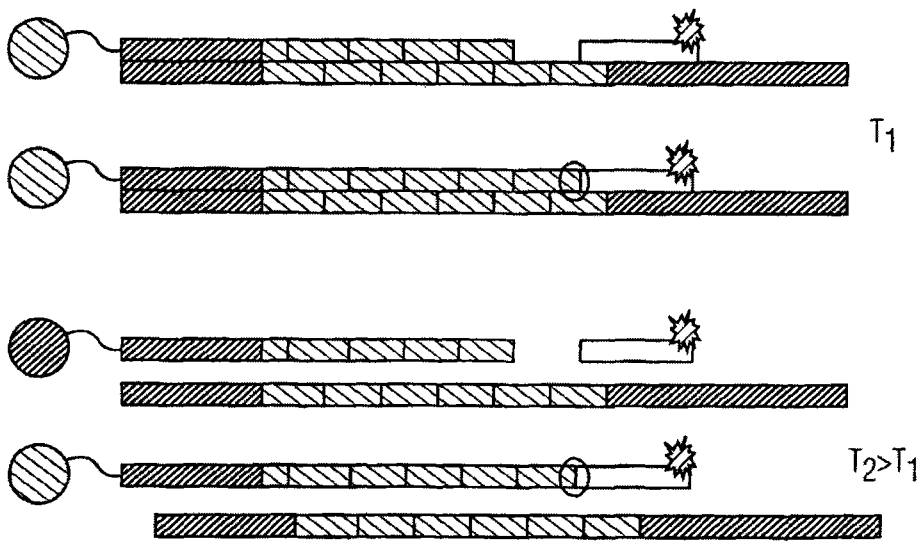


图19A

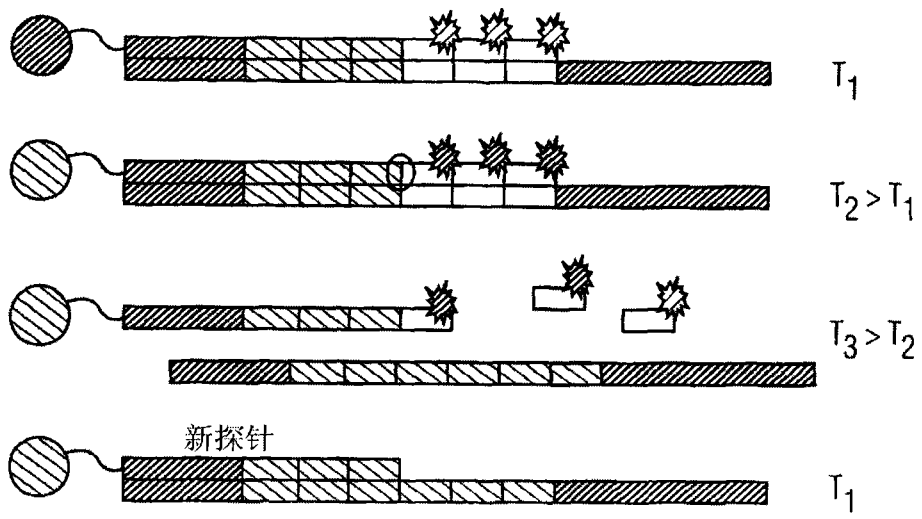


图19B