

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200680052374.1

[51] Int. Cl.

C12Q 1/68 (2006.01)

G01N 33/53 (2006.01)

G06F 17/00 (2006.01)

[43] 公开日 2008年12月31日

[11] 公开号 CN 101336301A

[22] 申请日 2006.12.29

[21] 申请号 200680052374.1

[30] 优先权

[32] 2006.1.5 [33] AU [31] 2006900060

[86] 国际申请 PCT/AU2006/001977 2006.12.29

[87] 国际公布 WO2007/076577 英 2007.7.12

[85] 进入国家阶段日期 2008.8.5

[71] 申请人 西蒙斯单倍体有限公司

地址 中国香港湾仔轩尼诗路48-62 上海工业投资大厦12层1024室

[72] 发明人 M·J·西蒙斯

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 韦东

权利要求书3页 说明书11页 附图2页

[54] 发明名称

微阵列方法

[57] 摘要

本发明提供了一种能鉴定一组相关核苷酸序列中某成员的微阵列探针组的鉴定方法，该方法包括以下步骤：提供候选探针组，其包含至少一种能与该相关核苷酸序列组中的两个或多个成员差异杂交的探针；检测该探针组与相关核苷酸序列组中两个或多个成员的反应性；以及观察该探针组与相关核苷酸序列组中两个或多个成员反应性模式的差异程度。

1. 一种能鉴定一组相关核苷酸序列中某成员的微阵列探针组的鉴定方法，该方法包括以下步骤：提供候选探针组，其包含至少一种能与该相关核苷酸序列组中的两个或多个成员差异杂交的探针；检测该探针组与相关核苷酸序列组中两个或多个成员的反应性；以及观察该探针组与该相关核苷酸序列组中两个或多个成员反应性模式的差异程度。

2. 如权利要求 1 所述的方法，其特征在于，所述反应性模式的差异程度足够充分，从而该相关核苷酸序列组的几乎所有成员都显示出独特反应性模式，所述候选探针组是信息探针组。

3. 如权利要求 1 或权利要求 2 所述的方法，其特征在于，该相关核苷酸序列组的各成员获自或衍生自两种或多种生物。

4. 如权利要求 1-3 中任一项所述的方法，其特征在于，有关该相关核苷酸序列组成员或者获得或衍生出该成员的生物的遗传特征或遗传连锁特征是已知的，从而所述独特反应性模式为存在或不存在所述遗传特征或遗传连锁特征提供了信息。

5. 如权利要求 1-4 中任一项所述的方法，其特征在于，所述候选探针组是通过包括将该相关核苷酸序列组中各成员的核苷酸序列分成多个亚序列的方法制造的，其中至少 2 个亚序列重叠且其中所述候选探针组针对该亚序列。

6. 如权利要求 5 所述的方法，其特征在于，至少有 3 个亚序列相互重叠。

7. 如权利要求 5 所述的方法，其特征在于，至少有 4 个亚序列相互重叠。

8. 如权利要求 1 所述的方法，其特征在于，至少有 5 个亚序列相互重叠。

9. 如权利要求 5 所述的方法，其特征在于，所述重叠是完全重叠。

10. 如权利要求 1-9 中任一项所述的方法，其特征在于，所述相关序列的不同之处在于存在一个或多个核苷酸多态性。

11. 如权利要求 10 所述的方法，其特征在于，所述核苷酸多态性为单核苷酸多态性。

12. 如权利要求 1-11 中任一项所述的方法，其特征在于，所述亚序列为探针长度。

13. 如权利要求 1-12 中任一项所述的方法, 其特征在于, 所述亚序列的长度约 10-50 个核苷酸。

14. 如权利要求 1-12 中任一项所述的方法, 其特征在于, 所述亚序列的长度约 15-35 个核苷酸。

15. 如权利要求 1-12 中任一项所述的方法, 其特征在于, 所述亚序列的长度约 25 个核苷酸。

16. 如权利要求 1-15 中任一项所述的方法, 其特征在于, 所有亚序列的长度相同或类似。

17. 如权利要求 1-16 中任一项所述的方法, 其特征在于, 所述相关核苷酸序列的序列相同性为至少 50%、60%、70%、80%、90%、95%或 99%。

18. 如权利要求 1-17 中任一项所述的方法, 其特征在于, 所述相关序列显示含有高密度的 SNP。

19. 如权利要求 1-18 中任一项所述的方法, 其特征在于, 所述相关序列是蛋白质编码序列、非编码序列、或蛋白质编码序列或非编码序列的组合。

20. 如权利要求 1-19 中任一项所述的方法, 其特征在于, 所述相关序列针对基因组的同一区域。

21. 如权利要求 1-20 中任一项所述的方法, 其特征在于, 所述相关核苷酸序列是某基因的等位基因。

22. 如权利要求 1-21 中任一项所述的方法, 其特征在于, 所述相关核苷酸序列组中相关核苷酸序列的数目大于 100、200、300、400、500、600、700、800、900 或 1000。

23. 如权利要求 1-22 中任一项所述的方法, 其特征在于, 所述相关核苷酸序列是涉及免疫系统的基因座的一部分。

24. 如权利要求 23 所述的方法, 其特征在于, 所述基因座是主要组织相容性复合体(MHC)、T 细胞受体、B 细胞受体、杀伤细胞抑制受体或免疫球蛋白的基因座。

25. 如权利要求 23 所述的方法, 其特征在于, 所述基因座是人白细胞抗原(HLA)系统的基因座。

26. 如权利要求 1-25 中任一项所述的方法, 其特征在于, 所述方法是自动

化方法。

27. 一种用权利要求 1-26 中任一项所述方法产生的信息探针组或部分信息探针组。

28. 如权利要求 27 所述的探针组,其特征在於,至少一个探针包含选自下组的标记: Cy5、Cy3、FITC、罗丹明、生物素、DIG 和放射性同位素。

29. 一种固体基质,其包括如权利要求 27 或权利要求 28 所述的固定的探针组。

30. 如权利要求 29 所述的固体基质,其特征在於,所述固体基质是微阵列芯片。

31. 一种用权利要求 27 或权利要求 28 所述的探针组鉴定一组相关核苷酸序列中某成员的微阵列方法。

32. 一种鉴定生物的遗传特征或遗传连锁特征的方法,所述方法包括采用如权利要求 27 或权利要求 28 所述的探针组。

33. 一种明确等位基因分配的方法,所述方法包括采用权利要求 27 或权利要求 28 所述的探针组。

34. 一种基于 HLA 系统的移植组织分型方法,所述方法包括采用权利要求 27 或权利要求 28 所述的探针组。

35. 一种鉴定新的等位基因的方法,所述方法包括采用权利要求 27 或权利要求 28 所述的探针组。

36. 一种如权利要求 5 所述的基本如上文参照附图描述的方法。

微阵列方法

发明领域

本发明涉及检测样品中特定核酸序列的方法，所述样品含有大量此类序列，如衍生自原核或真核细胞的 mRNA 或基因的完全遗传互补物的那些序列。具体地说，本发明涉及筛选用于微阵列应用的探针的方法。

发明背景

微阵列技术已经彻底改革了遗传学领域，它提供了用大量独特探针筛选含有核酸分子的待测混合物的方法。微阵列分析在本领域中被认为是一项经典的‘精确进入-精确离开 (precision in-precision out)’技术。基于合理的实验设计、优化的程序、正确设计的阵列元件、完美设计的探针组、纯净样品、强大的制造业和表面化学、高质量筛选、以及正确采用的样品示踪、量化和数据发掘工具的实验能够产生有价值的结果。

业已对微阵列所用的探针设计作了大量研究。尽管该技术正在改进中并变得更加强大，但为样品中要检测的各靶核苷酸序列筛选独特的能提供信息的探针仍存在困难。一个原因是，无论如何小心地选择探针组，至少一部分探针由于熟知的交叉杂交现象仍会结合一种以上的靶序列。诸如形成二级结构和探针的解链温度等其他因素也会造成杂交误差，从而降低了实验的精确度。

现有技术提供了许多筛选信息探针(informative probe)的理论模型(常以基于软件的算法实施)，这些探针在微阵列环境中引起错误杂交的可能性较低。艾飞麦趣公司(Affymetrix Inc)研究出的第一种探针设计程序可产生用于微阵列平台的 20-25 个碱基的短探针(Chee 等, Science, 1996/10/25; 274 (5287):610-4)。该研究指定了许多筛选强大和提供信息的探针的标准，分别为：

1. 均一性，以确保探针在实验温度下能结合靶分子，
2. 灵敏度，以确保探针不形成二级结构(这种结构将阻止探针结合靶 cDNA)，和

3. 特异性，以确保探针甚至在一些碱基改变之后仍保持其独特性。

本领域认为，基于这些原则的算法能极大降低实验室检测各靶分子对探针反应时所消耗的时间和精力。然而，也认为这种方法有缺陷，其中一个主要问题是由于所有探针都位于单个微阵列芯片上，对于任何给定的探针组只能采用一组杂交条件。由于不能为各个探针单独优化杂交条件(如离子强度和温度)，这种缺陷增加了交叉杂交和杂交误差的可能性。显然，微阵列平台的探针筛选领域需要作进一步改进。

本发明一方面提供了鉴定用于微阵列的探针的方法，该方法克服或减轻了现有技术的问题。

本说明书中关于文献、方法、材料、装置、论文等的讨论仅出于为本发明提供内容的目的。这并不说明或表示这些事物中的任何一种或其所有由于在本申请各项权利要求的优先权日之前就存在便形成了现有技术基础的一部分或是本发明所述领域的公知常识。

发明概述

在本发明的一个方面中，提供了一种能鉴定一组相关核苷酸序列中某成员的微阵列探针组的鉴定方法，该方法包括以下步骤：提供候选探针组，其包含至少一种能与该相关核苷酸序列组中的两个或多个成员差异杂交的探针；检测该探针组与相关核苷酸序列组中两个或多个成员的反应性；以及观察该探针组与相关核苷酸序列组中两个或多个成员反应性模式的差异程度。与现有技术的方法不同，本发明的方法不需要根据对靶序列的详细了解来仔细考虑和系统设计探针组。相反，本发明方法依据下述论点，即观察到的经验性探针反应性模式可提供足够辨别力以测定样品中是否存在预定的核苷酸序列。优选这种反应性模式的差异程度足够充分，从而该相关核苷酸序列组的几乎所有成员都能显示出其独特的反应性模式，所述候选探针组是信息探针组。

在本发明的一种形式中，有关该相关核苷酸序列组各成员或者获得或衍生出所述成员的生物的遗传特征(如突变的存在)或遗传连锁特征(如表型)是已知的，从而这种独特反应性模式为存在或不存在这种遗传特征或遗传连锁特征提供了信息。

在本发明的优选形式中，通过包括下述步骤的方法产生此候选探针组：将该相关核苷酸序列组中各成员的核苷酸序列分成多个亚序列，其中至少有 2 个亚序列重叠且其中所述候选探针组是针对该亚序列的。

在本发明的另一方面，提供了通过本文所述方法产生的信息探针组或部分信息探针组。通常，所述探针将是长度约 25 个核苷酸的寡核苷酸。该探针可与固体基质结合以便用于微阵列模式。

在该说明书的描述或权利要求中，单词“包括”及其变化形式，如“包含”或“涵盖”不表示排除其他添加物或组分或整数。

附图简述

图 1 显示了选择探针组优选方法的假想应用。该例中，有三个相关的 19 聚序列(#1、#2 或#3)。将外显子中的第一个核苷酸作为 1 (即其中的第 5 个核苷酸)，则该外显子在 6 或 11 位(下划线表示)含有两个 SNP。将该相关序列分成 9 聚亚序列，使亚序列之间完全重叠。图 1B 显示了从相关序列#1、#2 和#3 收集的所有亚序列。

发明详述

本申请提出了一种设计能够鉴定生物的遗传特征(如单核苷酸多态性; SNP)或生物的遗传连锁特征(如表型)的微阵列探针组的替代方法。该方法明显不同于技术人员设计适用于通过微阵列筛选样品的探针时采用的计算机(*in silico*)方法和体内方法。现有技术的方法涉及详细考虑例如基因的等位基因中发现的核苷酸序列的差异。一旦鉴定出这种差异便可设计出与基因内的某靶核苷酸序列特异性杂交的探针。与探针的杂交模式则提供了等位基因信息。鉴于寡核苷酸探针与靶序列的杂交不理想，因此只需要考虑简单的 Watson-Crick 碱基配对，故而这种方法会造成问题。因此，探针的核苷酸序列无需确定探针与靶核苷酸序列结合的能力。这种非理想行为的基础还未完全理解，但认为与探针和/或靶序列中存在二级结构有关。

不同于现有技术的方法，本发明的方法不要求根据对靶序列的详细了解来仔细考虑和系统设计探针组。而是，本发明方法依据观察到的经验性探针反应

性模式所提供的足够辨别力以测定样品中是否存在预定的核苷酸序列。因此，本发明提供了一种能鉴定一组相关核苷酸序列中某成员的微阵列探针组的鉴定方法，该方法包括以下步骤：提供候选探针组，其包含至少一种能与该相关核苷酸序列组中的两个或多个成员差异杂交的探针；检测该探针组与相关核苷酸序列组中两个或多个成员的反应性；以及观察该探针组与该相关核苷酸序列组中两个或多个成员反应性模式的差异程度。

该相关核苷酸序列组可获自细胞—单细胞生物的细胞或多细胞生物的细胞。例如，该相关核苷酸序列组中的各成员可获自不同生物，各生物展现出不同表型。成员序列不需要直接获自生物。成员序列也可以衍生自生物，例如通过在体外合成可复制的成员序列。

所述差异程度可以是为是：该相关核苷酸序列组的一些、大多数或所有成员显示出独特反应性模式。根据该方法的应用，不必使所有成员显示出独特反应性模式，但在本发明的优选方式中，反应性模式的差异程度足够充分以致该相关核苷酸序列组的基本上所有成员都显示出其独特的反应性模式。此时，由于采用同一探针组并标明了探针的反应性模式，从而可提供未知试样的确定信息，因此认为候选探针组是信息探针组。需要强调的是，不必要求探针组是完全信息探针组(即能够分辨所有成员的序列)，部分信息探针组(即只能分辨所有成员序列中的一部分)也包括在本发明的范围之内。

在所述方法的一个实施方式中，关于该成员序列或者获得或衍生出该成员序列的生物的遗传特征或遗传连锁特征是已知的，从而这种独特反应性模式为存在或不存在这种遗传特征或遗传连锁特征提供了信息。因此，可将标明的任何给定样品的反应性模式反过来与获得或衍生出该相关核苷酸序列组成员的生物的已知遗传特征或遗传连锁特征相关联。一个非限制性的例子是，需要鉴定某个体引起囊性纤维化的基因(CFTR 基因)中是否含有一个或多个突变。该基因有 1200 种突变(包括 SNP 或框内缺失)，其中已知只有 200 种多态性与基因功能障碍无关。本领域目前认为实际上不可能鉴定到具有可实现的杂交条件狭窄窗口的探针而对各个探针作最佳+/-解释。这种想法得自分离时各个 SNP-杂交探针必需提供信息的吝啬假设。可用本发明方法选择候选探针组，检测其对含有所有 1200 种可能突变的个体，包括含有认为与囊性纤维化表型无关的

所有 200 种多态性的个体的 DNA 的反应性。所选的探针组和(杂交)条件能使正常个体的反应性模式不同于含有功能障碍性 CFTR 基因的个体的反应性模式。在现有技术中,进行该过程时设计的探针组中,要单独设计各探针以鉴定 1200 种突变中的每一种。虽然在诸突变已被充分表征的前提下这也是可能的,但这种方法迄今还未成功过。不希望局限于理论的限制,我们认为这种失败是由于从理论上讲无论多么小心仔细地选择探针,但由于此法固有的不能为各个探针确定各自的杂交条件的缺陷,仍将有一部分探针不能发挥预期作用。

对微阵列用途的目前看法是,各个探针的作用是提供单一信息,这只能在对所有探针都是共同的具体杂交方案下进行。还认为不可能预测单个探针在具体杂交方案中的行为。本申请人首次提出任何一个探针是否按照理论预期发挥作用并不重要。此外,本发明人还提出,所用的杂交条件对于微阵列上所用多种探针中的任何一个不是最佳无关紧要。实际上,当杂交条件对于探针组上任何一个探针都不是最佳时本发明也可行。

该方法要求提供候选探针组的步骤。文中,术语“候选探针组”包括技术人员预测的能提供相关核苷酸序列组成员之间反应性模式充分差异程度的一组探针。技术人员不可能确信必然会预测任何给定的候选探针组能否提供反应性模式的所需差异。因此,本发明方法可包括对许多候选探针组进行试验然后鉴定出一组可接受的探针。庆幸的是,本文描述的方法更多是经验性选择探针的方法,而现有技术的方法则依赖于杂交理论。然而,理想的是可最大程度减小试验次数,为实现该目的,本文提到了各种示范性策略。应理解,本发明的方法不限于本文所述的任何策略。实际上,在本发明的一种形式中,不必考虑靶序列而采用随机或半随机探针组座位候选探针组。通过检测随机或半随机探针组根据经验可知,反应性模式“A”只见于具有等位基因“X”的个体,而反应性模式“B”只见于具有等位基因“Y”的个体。因此,通过回顾鉴定差异性反应探针的序列,将能够重新产生信息探针组。

在本发明的另一种形式中,候选探针组部分依据于对靶分子核苷酸序列的了解。此时,探针可针对靶分子中的多个重叠的亚序列。尽管这种方法需要知晓靶序列,但不需要精细考虑设计探针以覆盖所有预期的核苷酸序列。该方法可产生大量探针,优点是使微阵列芯片能够容纳非常大数目的探针。

在本发明的一种形式中，考虑将靶序列分成重叠的亚序列来鉴定一组候选探针。为提高鉴定反应的差异能力，将相关序列组的各成员分成许多亚序列。在给定的成员序列中，亚序列相互重叠，从而可产生相当大数量的亚序列。

优选地，至少一个亚序列与一个以上其他亚序列重叠。更优选地，至少一个亚序列与 2、3、4 或 5 个以上其他亚序列重叠。

用来产生一系列重叠探针长度亚序列的重叠程度应尽可能小。一系列 25 聚亚序列的最小重叠的一个例子是，第一个亚序列包括核苷酸 1-25，第二个亚序列包括核苷酸 25-50，第三个亚序列包括核苷酸 50-75，依此类推。

重叠可以是可能的最大重叠程度。具有最大可能重叠的一系列 25 聚亚序列的一个例子是，第一个亚序列包括核苷酸 1-25，第二个亚序列包括核苷酸 2-26，第三个亚序列包括核苷酸 3-27，依此类推。

本发明包括介于可能的最小值和最大值之间的任何中间重叠程度。然而，采用几乎最大的重叠程度是优选的，因为这样每次设计探针组时需要判断的量最少。

该组任何给定成员采用该方法时不需要固定重叠的数量。也无需固定亚序列的长度。技术人员可以常规研究不同长度的亚序列或亚序列间不同重叠程度的影响以确定可否从反应性差异获得任何益处。

需要强调的是，本发明的方法不限于包括针对靶核苷酸序列重叠亚序列的探针的候选探针组。采用至少反应性程度差异最小的探针的候选探针组也包括在本发明范围之内。

所述相关核酸序列可以是基因组 RNA、cDNA 或 cRNA。通常对基因组 DNA 样品进行扩增再施加于阵列，采用的引物侧接感兴趣区域。基因组 DNA 实际上可获自任何组织来源(纯红血细胞除外)。例如，方便的组织样品包括全血、精液、唾液、泪液、尿液、粪便、汗液、口腔(刮取物)、皮肤或毛发。如果获得样品的个体某多态性位点是纯合的，扩增包含该多态性位点的基因组 DNA 可产生一种靶核酸，如果个体是杂合的则得到两种靶分子。

可用技术人员已知的任何合适方法制备供分析的 DNA，包括采用合适引物的 PCR 法。当需要分析整个基因组时，可采用全基因组扩增(WGA)法。可方便获得的用于该方法的市售试剂盒包括由美国密苏里州圣路易斯西格玛阿

尔德里奇公司(Sigma-Aldrich Corp, St Louis, MO, USA)制造的 GenoPlex®完全 WGA 试剂盒(GenoPlex® Complete WGA kit)。该试剂盒基于将基因组随机片段化成一系列模板。所得较短 DNA 链形成含有明确的 3 引导末端和 5 引导末端的 DNA 片段文库。在最初阶段采用线性等温扩增,然后经有限轮次的几何(PCR)扩增复制该文库。另一种市售试剂盒是由德国希尔顿的恰根公司(Qiagen GmbH, Hilden, Germany)制造的 REPLI-g。WGA 方法适用于各种来源的纯化的基因组 DNA,所述来源包括血卡(blood card)、全血、口腔刮取物、土壤、植物以及福尔马林固定的石蜡包埋组织。

mRNA 样品通常要被扩增。此时,扩增通常是通过逆转录进行的。可按照 WO 96/14839 或 WO 97/01603 的描述扩增所有表达的 mRNA。如果获得样品的个体表达的 mRNA 内多态性位点是杂合的,则扩增二倍体样品的 RNA 样本可产生两种靶分子。

显然,可利用该方法鉴定的核苷酸亚序列随后设计能够鉴定相关序列组目前鉴定到的所有成员的探针组。文中,术语“靶核苷酸序列”表示随后可产生的针对其的特异性探针的序列。下文将进一步讨论探针的产生,然而,所述探针通常是能够与靶核苷酸序列杂交的寡核苷酸探针。

技术人员将理解,探针长度亚序列的长度可以是任何长度,只要能够区分相关序列组的各成员即可。用于微阵列应用的探针通常约 25 个核苷酸长,但较长或较短的探针也可用于本发明。当需要足够的核苷酸以提供结合特异性时可决定使用较低的有用长度,且该长度可以从约 10 个核苷酸到 15 个核苷酸。在检测“亚基因组”时可考虑使用长度少于 15 个核苷酸的探针。一个例子是检测单个单倍体染色体,且序列检测特异性不要求探针长度能够分析整个人类基因组中的大约 30 亿个核苷酸时。上限可由与解链双链区域和退火多核苷酸单链的需求有关的物理限制条件决定。其值可以为约 30-50 个核苷酸。上限可根据 C/G 碱基(假设需要较高解链温度来分离双链体中的这些碱基)与 A/T 配对的比例而不同。虽然探针长度实施时可以有上限和下限,但这些程度限制将根据实际应用而不同,技术人员将通过常规的经验性实验鉴定最合适的探针长度。

预计当需要检测大量候选探针组时,可能需要检测大量杂交条件以确保忠

实检测该相关核苷酸序列组的所有成员。例如，一旦选出了最终的探针组便可优化杂交条件以使某些边缘探针能提供较佳的信噪比。当候选探针组不能充分检测某一组相关核苷酸序列的所有成员时也可优化杂交条件。改变杂交条件可使该探针组能够鉴定该组的所有成员。

初始杂交条件可能一开始具有低严格性，包括低温、低离子强度和低去污剂浓度。低严格性杂交的典型缓冲剂包括 1X SSC 和 0.2% SDS。低严格性的典型温度约为 42°C。

如果低严格性条件、无法提供探针反应性的所需差异程度，则可使用含有 0.1X SSC 和 0.2% SDS 的严格性较高的缓冲剂。也可尝试约 65°C 的温度。也可在缓冲剂中加入变性剂如甲酰胺以改变严格性条件，较高浓度的变性剂能降低核酸分子的熔点。

技术人员通过常规实验将能够鉴定高出用于特定杂交条件的特定探针组，从而实现所需的探针反应性模式的差异程度。

应理解，所述方法可用于需要区分许多相关核苷酸序列的任何情况。文中，术语“核苷酸序列”及其变化形式包括脱氧核糖核酸(DNA)和核糖核酸(RNA)序列。相关核苷酸序列可以是显示序列相同性水平最低的任何核苷酸序列组。优选地，所述序列的相同性水平至少为 50%、60%、70%、80%、90%、95% 或 99%。例如，当相关序列较长且其中散布有一系列 SNP 时这种相同性甚至可高于 99%。

相关序列可以是蛋白质编码序列、非蛋白质编码序列、或蛋白质编码序列或非蛋白质编码序列的组合。相关序列可来源于二倍体、单倍体、三倍体或多倍体材料，或在二倍体、单倍体、三倍体或多倍体状态上提供信息。相关序列可以是天然的或合成的。它们可来自任何生物，其中包括动物、植物、微生物、细菌或病毒。

在本发明的一种形式中，所述相关序列针对基因组的同一区域。例如，从某外显子第一个核苷酸到该外显子最后一个核苷酸之间的区域。此时，当采用 25 聚探针时，可将该探针设计成其第 13 个核苷酸(即当中的核苷酸)针对该外显子的第一个核苷酸。因此，当第一个核苷酸是 G 时，该探针的第 13 个核苷酸将是 C。显然，该探针两侧的 12 聚区域在一种情况中将针对该外显子之前

的区域，而在另一种情况中将针对该外显子中的区域。

用于本发明方法优选实施方式的重叠策略的一般操作方法可见图1所示的大为简化的实施例。该实施例涉及3个相关核苷酸序列(#1、#2和#3)，外显子开始于从左手或5'端开始的第5个核苷酸(即A)。将外显子中的第一个核苷酸作为1，则该外显子在6或11位(下划线表示)含有两个SNP。采用含有9个核苷酸的亚序列，使亚序列之间完全重叠。因此，第一个亚序列开始于-4位终止于+5位。

由图1A可见，各相关序列被分成11个9聚亚序列。这总共提供了33个亚序列(图1B)。技术人员将理解，如果原始的靶分子是双链(ds)分子则探针序列不需要互补。此时，核苷酸序列可直接用作探针序列或与ACAGGGGTGTCGTGCAAAGAACCTC互补，这取决于技术人员所选择的产生靶序列的策略。因此，如果用dsDNA来产生最终的靶序列时，探针可针对阵列上的一条链或两条链。

本发明的方法将能够分析核苷酸序列中的许多变化，包括缺失、取代、添加等等。在本发明的一种形式中，除了存在SNP外，所述相关核苷酸序列相同。

尽管SNP能够以任何密度存在，但当SNP以高密度存在时该方法可提供更大的优点。优选所述密度为该核苷酸序列在探针长度区域内存在两个或多个SNP。在区分包含高密度SNP的相关核苷酸序列时先前存在的问题，是由于迄今为止一直认为需要提供大量探针才能覆盖某给定区域内SNP的每一种组合。在设计用于HLA分型的探针组时这尤其是一个问题，HLA外显子中20%-50%的核苷酸为多态性，且多态性位点常成簇出现。这导致需要事实上无法实现数量的不同探针的现有技术预计能明确将一种HLA类型归于某个体。

显然，尽管一组中相关核苷酸序列的数目可低至2个，但该方法在相关核苷酸序列的数目较高时可提供更多优点。在该方法的优选形式中，所述相关核苷酸序列组中相关核苷酸序列的数目大于100、200、300、400、500、600、700、800、900或1000。本发明尤其适用于相关核苷酸序列的数目较高且SNP密度较高的情况。

在该方法的优选形式中，所述相关核苷酸序列是某基因的等位基因。已知

编码相同蛋白质的人基因在不同个体中可能具有不同序列(等位基因)。具有较多数目等位基因的基因例子主要是那些涉及免疫系统的基因,它们以超变性(hypervariability)为共同特征。示例性的基因包括主要组织相容性复合体(MHC)、T细胞受体、B细胞受体、免疫球蛋白、杀伤细胞抑制受体(KIR)等的基因。然而应理解,本文所述的方法将可用于任何相关核苷酸序列组,但当相关核苷酸序列具有超变性时将获得较大益处。当超变性为高密度 SNP 时仍能提供较大益处。

如上所述,MHC 基因尤其具有多态性。I 或 II 类 MHC 跨膜蛋白构成人白细胞抗原(HLA)系统,可利用该系统对组织分型以评价移植相容性。I 类蛋白由以下三个基因座编码:HLA-A、HLA-B 和 HLA-C,目前已分别识别各有 309、563 或 167 个等位基因。

II 类蛋白含有 α 或 β 链,由基因座 DR、DQ 和 DP 编码。DR 基因座 α 链包括 3 个等位基因, β 链包括 483 个等位基因。DQ 基因座 α 链包括 25 个等位基因, β 链包括 56 个等位基因。DP 基因座 α 链包括 20 个等位基因, β 链包括 107 个等位基因。因此需要注意,仅就 I 类区域而言,由许多等位基因组合提供某个体的 HLA 类型。

在该方法的一种形式中,该方法是自动进行的。现有技术的方法,如 Guo 等(2002)设计的探针,是基于仔细考虑了所有相关的核苷酸序列而努力鉴定能覆盖所有观察到的 SNP 组合的探针。这是一项劳动强度非常高的过程,其成功或失败取决于进行该分析的个体的技能。如果相关序列的数目非常大,或者等位基因的数目非常大,完成设计探针的任务实际上是不可能的。

所述方法可包括不同长度的亚序列以及亚序列间不同重叠水平的组合。在本发明的高度优选形式中,亚序列的长度约 25 个核苷酸,且重叠程度最大。

将了解,本发明还能够发现是否存在迄今仍未识别的等位基因。可发现具有新反应性模式的个体。将能够鉴定显示出新反应性的探针并推导出新的等位基因。

如上所述,分析到的等位基因可能只针对蛋白质编码区,或只针对非编码区。或者,可采用蛋白质编码区和非编码区的组合。

给定靶亚序列之后,技术人员将能够合成出能与各靶亚序列杂交的探针。

所述探针与所鉴定的非冗余序列基本互补。如果靶序列是由双链模板产生的，则这些探针可以是正义或反义的。

技术人员可通过即时方法(*instant method*)如锁定核酸都能够研究在探针设计中使用修饰的核苷酸是否能获得任何益处。

探针可包括标记以便于检测。示例性的标记包括 Cy5、Cy3、FITC、罗丹明、生物素、DIG 和各种放射性同位素。

在本发明的另一个方面，提供了一种用上述一组探针鉴定一组相关核苷酸序列中某成员的微阵列方法。因此，本发明的另一个方面提供了一组上述固定在固体基质上的探针。本发明这种形式的一个示例性的实施方式可见 Affymetrix® 销售的 GeneChip® 技术。该技术依赖于光版印制过程，即用光敏化合物涂布 5"×5" 的石英晶片(*wafer*)，从而阻止晶片与所产生的 DNA 探针的第一个核苷酸之间发生偶联。用平板印制面罩来阻止光线进入或使光线进入晶片表面的特定部位。然后用含有腺嘌呤、胸腺嘧啶、胞嘧啶或鸟嘌呤的溶液浸没该表面，该玻璃上只有那些已通过光照而去保护的区域才发生偶联。偶联的核苷酸还带有光敏保护基，如此重复进行该循环。许多公司提供了固定探针的其他方法，其中包括英国牛津的牛津基因技术公司(Oxford Gene Technology, Oxford, U.K.)、美国加州帕洛阿托的艾吉兰特科技公司(Agilent Technologies, Palo Alto, CA, U.S.A.)以及美国威斯康星州麦迪逊的宁布莱金系统有限公司(Nimblegen Systems Inc., Madison, WI, U.S.A.)。

将会理解，本发明可用于许多技术领域。预计医学领域尤其会受益，该领域用此方法对个体的基因分型。所述方法将尤其可用于移植组织分型(如采用 HLA 基因、KIR 基因、最小组织相容性基因座，等等)，以及生药学、DNA “指纹识别” 等等。所述探针可用于任何应用，包括原位杂交、狭线印迹、斑点印迹、集落杂交、噬菌斑杂交、Northern 印迹、Southern 印迹以及微阵列应用。

预计这种应用将延伸至用于非人动物，如灵长类，例如用于候选药物的临床前药物基因组学评价。本发明还考虑可用于检测具有经济价值的动物(如牛、家禽等)，例如在繁殖过程中改进诸如瘦肉率等参数。

	ATCGATCGATCGATCGATC	相关序列#1
5	ATCGATCGA	subsequence (1) 1
	TCGATCGAT	subsequence (1) 2
	CGATCGATC	subsequence (1) 3
	GATCGATCG	subsequence (1) 4
	ATCGATCGA	subsequence (1) 5
10	TCGATCGAT	subsequence (1) 6
	CGATCGATC	subsequence (1) 7
	GATCGATCG	subsequence (1) 8
	ATCGATCGA	subsequence (1) 9
	TCGATCGAT	subsequence (1) 10
15	CGATCGATC	subsequence (1) 11
20	ATCGATCGAT<u>G</u>GATCGATC	相关序列#2
	ATCGATCGA	subsequence (2) 1
	TCGATCGAT	subsequence (2) 2
	CGATCGATG	subsequence (2) 3
	GATCGATGG	subsequence (2) 4
25	ATCGATGGA	subsequence (2) 5
	TCGATGGAT	subsequence (2) 6
	CGATGGATC	subsequence (2) 7
	GATGGATCG	subsequence (2) 8
	ATGGATCGA	subsequence (2) 9
30	TGGATCGAT	subsequence (2) 10
	GGATCGATC	subsequence (2) 11
35	ATCGATCGATCGAT<u>C</u>CATC	相关序列#3
	ATCGATCGA	subsequence (3) 1
	TCGATCGAT	subsequence (3) 2
	CGATCGATC	subsequence (3) 3
40	GATCGATCG	subsequence (3) 4
	ATCGATCGA	subsequence (3) 5
	TCGATCGAT	subsequence (3) 6
	CGATCGATC	subsequence (3) 7
	GATCGATCC	subsequence (3) 8
45	ATCGATCCA	subsequence (3) 9
	TCGATCCAT	subsequence (3) 10
	CGATCCATC	subsequence (3) 11

图

1A

5 ATCGATCGA
TCGATCGAT
CGATCGATC
GATCGATCG
ATCGATCGA
10 TCGATCGAT
CGATCGATC
GATCGATCG
ATCGATCGA
TCGATCGAT
15 CGATCGATC
ATCGATCGA
TCGATCGAT
CGATCGATC
20 ATCGATCGA
TCGATCGAT
CGATCGATG
GATCGATGG
ATCGATGGA
TCGATGGAT
CGATGGATC
GATGGATCG
25 ATGGATCGA
TGGATCGAT
GGATCGATC
ATCGATCGA
30 TCGATCGAT
CGATCGATC
GATCGATCG
ATCGATCGA
TCGATCGAT
35 CGATCGATC
GATCGATCC
ATCGATCCA
TCGATCCAT
CGATCCATC
40

图 1B

专利名称(译)	微阵列方法		
公开(公告)号	CN101336301A	公开(公告)日	2008-12-31
申请号	CN200680052374.1	申请日	2006-12-29
[标]申请(专利权)人(译)	西蒙斯单倍体有限公司		
申请(专利权)人(译)	西蒙斯单倍体有限公司		
当前申请(专利权)人(译)	西蒙斯单倍体有限公司		
[标]发明人	MJ西蒙斯		
发明人	M·J·西蒙斯		
IPC分类号	C12Q1/68 G01N33/53 G06F17/00		
CPC分类号	C12Q1/6881 C12Q2600/156		
代理人(译)	韦东		
优先权	2006900060 2006-01-05 AU		
外部链接	Espacenet SIPO		

摘要(译)

本发明提供了一种能鉴定一组相关核苷酸序列中某成员的微阵列探针组的鉴定方法，该方法包括以下步骤：提供候选探针组，其包含至少一种能与该相关核苷酸序列组中的两个或多个成员差异杂交的探针；检测该探针组与相关核苷酸序列组中两个或多个成员的反应性；以及观察该探针组与相关核苷酸序列组中两个或多个成员反应性模式的差异程度。

5	ATCGATCGATCGATCGATC ATCGATCGA TCGATCGAT CGATCGATC GATCGATCG ATCGATCGA TCGATCGAT CGATCGATC GATCGATCG ATCGATCGA TCGATCGAT CGATCGATC	相关序列#1 subsequence (1) 1 subsequence (1) 2 subsequence (1) 3 subsequence (1) 4 subsequence (1) 5 subsequence (1) 6 subsequence (1) 7 subsequence (1) 8 subsequence (1) 9 subsequence (1) 10 subsequence (1) 11
20	ATCGATCGATCGATCGATC ATCGATCGA TCGATCGAT CGATCGATG GATCGATCG ATCGATGGA TCGATCGAT CGATCGATC GATCGATCG ATCGATCGA TGGATCGAT CGATCGATC	相关序列#2 subsequence (2) 1 subsequence (2) 2 subsequence (2) 3 subsequence (2) 4 subsequence (2) 5 subsequence (2) 6 subsequence (2) 7 subsequence (2) 8 subsequence (2) 9 subsequence (2) 10 subsequence (2) 11
35	ATCGATCGATCGATCGATC ATCGATCGA TCGATCGAT CGATCGATC GATCGATCG ATCGATCGA TCGATCGAT TCGATCGAT CGATCGATC GATCGATCC ATCGATCGA TCGATCCAT CGATCCATC	相关序列#3 subsequence (3) 1 subsequence (3) 2 subsequence (3) 3 subsequence (3) 4 subsequence (3) 5 subsequence (3) 6 subsequence (3) 7 subsequence (3) 8 subsequence (3) 9 subsequence (3) 10 subsequence (3) 11