

[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 200680051470.4

[51] Int. Cl.

*C12Q 1/68 (2006.01)*  
*G01N 33/53 (2006.01)*  
*C12N 15/09 (2006.01)*  
*G06F 17/00 (2006.01)*

[43] 公开日 2009年2月4日

[11] 公开号 CN 101360834A

[22] 申请日 2006.11.21

[21] 申请号 200680051470.4

[30] 优先权

[32] 2005.11.21 [33] US [31] 60/738,019

[86] 国际申请 PCT/AU2006/001740 2006.11.21

[87] 国际公布 WO2007/056825 英 2007.5.24

[85] 进入国家阶段日期 2008.7.21

[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 马尔科姆·J·西蒙斯

[74] 专利代理机构 北京市柳沈律师事务所  
代理人 封新琴

权利要求书3页 说明书24页 附图12页

[54] 发明名称

识别核苷酸序列的方法与探针

[57] 摘要

本发明提供识别目的核苷酸序列集的方法，所述核苷酸序列能识别相关核苷酸序列组的成员；该方法包含将组中每个成员的核苷酸序列分成多个子序列的步骤，其中至少两个子序列重叠。所述方法可用于产生能在 HLA 或 KIR 基因座分配等位基因的探针集。

1. 用于识别目的核苷酸序列集的方法，所述核苷酸序列能识别相关核苷酸序列组的成员；该方法包含将组中每个成员的核苷酸序列分成多个子序列的步骤，其中至少两个子序列重叠。

2. 根据权利要求1的方法，其中至少三个子序列互相重叠。

3. 根据权利要求1的方法，其中至少四个子序列互相重叠。

4. 根据权利要求1的方法，其中至少五个子序列互相重叠。

5. 根据权利要求1的方法，其中重叠为完全重叠。

6. 根据权利要求1的方法，其包含分析至少一部分子序列的冗余性的步骤。

7. 根据权利要求1的方法，其中一个或多个子序列在其5'和/或3'末端或末端附近不包含一个或多个多态位点。

8. 根据权利要求1的方法，其中一个或多个子序列在其中心或中心附近包含一个或多个多态位点。

9. 根据权利要求1的方法，其中一个或多个子序列在其中心包含一个多态位点。

10. 根据权利要求1的方法，其中相关序列的区别在于存在一个或多个核苷酸多态性。

11. 根据权利要求10的方法，其中核苷酸多态性是单核苷酸多态性。

12. 根据权利要求1的方法，其中所述子序列为探针长度。

13. 根据权利要求1的方法，其中所述子序列长约10至约50个核苷酸。

14. 根据权利要求1的方法，其中子序列长约15至约35个核苷酸。

15. 根据权利要求1的方法，其中子序列长度为约25个核苷酸。

16. 根据权利要求1的方法，其中所有子序列的长度相同或相近。

17. 根据权利要求1的方法，其中相关核苷酸序列具有至少50%、60%、70%、80%、90%、95%或99%的序列同一性。

18. 根据权利要求1的方法，其中相关序列表现出高密度的单核苷酸多态性(SNP)。

19. 根据权利要求1的方法，其中相关序列是蛋白编码序列、非编码序列，或蛋白编码和非编码序列的组合。

20. 根据权利要求1的方法, 其中相关序列指向基因组的相同区域。
21. 根据权利要求1的方法, 其中相关核苷酸序列是基因的等位基因。
22. 根据权利要求1的方法, 其中相关核苷酸序列组中相关核苷酸序列的数目多于100、200、300、400、500、600、700、800、900或1000个。
23. 根据权利要求1的方法, 其中相关核苷酸序列是免疫系统中涉及的基因座的一部分。
24. 根据权利要求23的方法, 其中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、杀伤细胞抑制性受体或免疫球蛋白的基因座。
25. 根据权利要求23的方法, 其中所述基因座是人白细胞抗原(HLA)系统的基因座。
26. 根据权利要求23的方法, 其中所述基因座是I类或II类MHC跨膜蛋白。
27. 根据权利要求23的方法, 其中所述基因座是DR、DQ或DP基因座。
28. 根据权利要求6的方法, 其包含从目的核苷酸序列集中去除或不包括至少一个冗余序列。
29. 根据权利要求28的方法, 其中所述方法将目的核苷酸序列集中的序列数目从理论预期的探针数目减少到至少约1/5、1/10或1/20。
30. 根据权利要求28的方法, 其中所述方法将探针数目减少至少约50%、60%、70%、80%、90%或95%。
31. 根据权利要求28的方法, 其中在探针集中, 去除或不包括基本上所有冗余序列。
32. 根据权利要求1的方法, 其中所述方法适于自动化。
33. 根据权利要求1的方法, 其中所述方法能识别相关序列中新的多态性位点或多态性位点的新组合。
34. 探针集, 其能够特异性杂交至由根据权利要求1的方法识别的目的核苷酸序列。
35. 根据权利要求34的探针集, 其中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。
36. 固体基质, 其包括固定化的根据权利要求34的探针集。
37. 根据权利要求36的固体基质, 其中固体基质是微阵列芯片。

- 
38. 使用根据权利要求 34 的探针集识别相关核苷酸序列组的成员的方法。
  39. 确定的等位基因分配的方法,其包括使用根据权利要求 34 的探针集。
  40. 基于 HLA 系统进行移植组织分型的方法,其包括使用根据权利要求 34 的探针集。
  41. 识别新等位基因的方法,其包括使用根据权利要求 34 的探针集。
  42. 能够执行根据权利要求 1 的方法的计算机可执行代码。
  43. 参照任何附图或实施例,基本上如本文所述的根据权利要求 1 的方法。

## 识别核苷酸序列的方法与探针

### 发明领域

本发明针对分子生物学领域。更特定地，本发明针对用于产生寡核苷酸探针的方法，及其在识别相关核苷酸组的成员中的用途。可以使用所述方法和探针识别个体中基因的等位基因。

### 发明背景

人类基因组计划强调了基因组中单核苷酸多态性(SNP)的重要性。在整个基因组中平均每 100 至 300 个碱基就会出现这种多态性。已知所有人类的基因有超过 99%的基因相同，而 SNP 的存在提供了种属中基因多样性的主要部分。基因的不同等位基因可以赋予个体非常不同的表型，包括如疾病抗性、对药物化合物的响应能力、运动能力等多种多样的特性。

植物基因组也包含 SNP，其能够导致不同的性质。在遗传分析中，SNP 越来越多地成为特别选定的标记，并且在农业育种项目中常规地用作标记。SNP 不仅能用来把特别的基因型和表型联系起来。在识别如细菌、病毒等具有多样性的生物时，它们还能用作“指纹”。

由于多种原因，将基因型与单独个体联系起来的能力非常重要。从广义来讲，这涉及识别相关生物受试基因(subject gene)的核苷酸序列。提供这种信息的最直接方式是将受试基因测序。尽管自动测序成为可能已有数年的时间，但是过程仍然是耗时而且昂贵的。

因为直接测序的广泛应用受到了限制，结果导致发展了很多间接的方法来识别等位基因。一种最简单的方法是使用限制性片段长度多态性(RFLP)。这种方法依赖于限制性内切酶对特定核苷酸序列的特异性。因此，如果存在特定序列，内切酶将切割多核苷酸，而如果不存在特定序列，则不会产生切割。通过不同模式的限制性片段来检测不同的基因型，如用凝胶电泳检测。这种方法的缺点是如果没有针对等位基因范围内每个 SNP 的特异性内切酶，则不能用 RFLP 识别所有的等位基因。这是经常出现的情况，所以 RFLP 的使用受到了明显的限制。

另一种检测等位基因的方法涉及使用寡核苷酸探针，所述寡核苷酸探针特异性地与一个等位基因中存在的序列结合，而不同其它等位基因中存在的序列结合。通过使用标签，比如荧光化合物或放射性同位素，可以检测探针与目的等位基因的结合。基于寡核苷酸探针的方法的问题是，要精确确定基因型，可能有必要使用非常大量的探针。因为多核苷酸杂交的生物物理学限定探针长度是受到限制的(通常不超过约 65 个核苷酸)，所以当受试基因长于最大探针长度时，必须设计一系列不同的探针，以覆盖基因的全长。当受试基因具有大量等位基因和大量 SNP 时，当 SNP 的密度高时，或出现这些因素中任何因素的组合时，不同探针的数目会增长得非常多。

本领域中问题的实例是人白细胞抗原 HLA-DRB 基因座，经常在用于器官移植的组织分型中分析所述基因座。这个基因座现在有 483 个已识别的等位基因，并且在可变的第二外显子中有 270 个核苷酸。对于在这个基因座解析基因型将需要的探针，简单的乘法可得到 130,410 种不同的核苷酸序列变化。产生如此大量的不同寡核苷酸探针，然后评价每个探针与测试样品杂交的能力，这明显是严重的负担。此外，会继续发现原来未识别的等位基因，从而使提供能够解析个体的 HLA 类型的探针集这一问题更加严重。

通过能使几千个探针与“芯片”结合形成“微阵列”的固相技术的发展，已经部分克服了使用大量探针的内在问题。然而，微阵列技术仍然需要使用很多探针来识别基因的所有等位基因，而且只是提供更方便的形式来处理大的探针集。现在用于 SNP 检测的探针指向目的 DNA 分子中在物理上分开的区域，而且经常选择 SNP 的侧翼序列为单态的位置。像这样的探针的使用方法以“重测序(resequencing)”在本领域中为人所知。

重测序取决于能够识别所有可能的 SNP 的、特别设计的探针的使用。Guo 等(2002, Genome Research 12: 447-457)论述了为 HLA-分型提供探针的问题，通过制作 20-mer 的探针，每个探针经过设计，代表 SNP 的特定组合，而不是单一的 SNP。这个方法的问题是它不是系统的方法，必需由人通过周密考虑来设计探针。假定在这个过程中确实可能存在错误，那么对于在探针设计过程结束时，探针集是否将识别出所有的等位基因，则仍然存在不确定性。

Guo 等提供的方法的更深入的问题是必需在探针全长范围内从头到尾包括 SNP 位点。参考 Guo 等的表 1 表明，在 20-mer 的探针中，从 5' 末端到 3' 末端存在多态性位点。在本领域中已知杂交的精确性沿着探针的侧翼而降低，

所以可以预期,使用 Guo 等的方法,杂交反应中将存在不精确性。要特别注意的是,由 Guo 等设计的探针集在 100 个杂交中产生了 32 个假阳性反应。

因此,本发明的一个方面是,通过提供设计探针集的系统方法克服或减轻在先技术中的问题,所述探针集能有力地(robustly)识别核苷酸序列中所有已知的多态性。

本说明书中包括的对文献、法令、材料、设备、文章等的讨论,仅旨在提供本发明的背景。这些内容中的任何内容或全部内容不因其存在于本申请各项权利要求的优先权日之前,而暗示或代表其构成了现有技术基础的一部分,或是与本发明相关的领域中的公知常识。

### 发明概述

在第一方面,本发明提供用于识别目的核苷酸序列集的方法,所述核苷酸序列能识别相关核苷酸序列组的成员;所述方法包含将组中每个成员的核苷酸序列分成多个子序列(subsequence)的步骤,其中至少两个子序列重叠。申请人已经发现,通过将研究的目的核苷酸序列分成重叠的子序列,有可能识别目的核苷酸序列集,所述核苷酸序列可用作与寡核苷酸探针杂交的靶标。优选地,子序列中至少一个与多于一个的其它子序列重叠。更优选地,子序列中至少一个与多于 2、3、4 或 5 个其它子序列重叠。

具有优势地,所述方法适于(is amenable to)自动化,而且预计适用于提供能解析具有大量等位基因和/或高密度 SNP 的基因的探针,所述基因如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、杀伤细胞抑制性受体(killer inhibitor receptor; KIR)等的基因。

在所述方法的一个实施方案中,应用中所需的探针数目可以通过识别冗余探针,并在探针集中去除或不包括冗余探针而显著减少。在本领域中尚未认识到的是,在为了设计寡核苷酸探针集而分析相关序列时,一个成员序列中的多态性不必要存在于另一个成员序列中。因此,不必要提供覆盖每一种多态性的每一种组合的探针,因为相关序列组中不必要存在所有的组合。

在所述方法的另一个实施方案中,一个或多个子序列(和源自子序列的任何探针)在所述一个或多个子序列的 5'和/或 3'末端或末端附近,不包含一个或多个多态位点。在所述方法的另一个实施方案中,一个或多个子序列在所述一个或多个子序列的中心或中心附近包含一个或多个多态位点。避免多态

位点接近探针侧翼，并且将位点集中在探针中心，可以克服 Guo 等(2002)提供的探针的问题，即明显的不精确结合，而产生大量假阳性杂交反应。

在另一方面，本发明提供能够与目的核苷酸序列特异性杂交的探针集，所述序列由本文所述方法识别。优选地，探针指向(direct to)多个外显子的覆盖范围(coverage)，而且能够提供总的等位基因分配(total allele assignment)。

在另一方面，本发明提供使用本文所述探针集识别相关核苷酸序列组成员的方法。方法将通常使用固定于微阵列芯片上的探针。

在另一方面，本发明提供能够执行本文所述方法的计算机可执行程序(软件)。

### 附图简述

图 1 表示选择探针集方法的假想应用。在这种情况下，有三个相关的 19-mer 序列(#1、#2 和#3)。在外显子中将第一个核苷酸设为 1 (即第五个核苷酸)，那么外显子在位置 6 和 11 有两个 SNP(下划线表示)。图 1A 表示分成 9-mer 子序列的相关序列，子序列之间完全重叠。图 1B 表示从相关序列#1、#2 和 #3 中汇聚的所有子序列。图 1C 表示在去除冗余子序列后源自图 1B 中的子序列集。要强调的是，这个假想实例不一定展示了本发明的所有优势，而仅仅旨在表明操作所述方法的优选形式。

图 2 表示由本发明识别的、用于分配 HLA-A\*0201 (外显子 2 和 3)的探针序列。选择探针长度为 25-mer，并且探针间具有最大的重叠。

### 发明详述

申请人提出用于设计探针的系统方法，所述探针能识别相关核苷酸序列组的成员。因此，在第一方面，本发明提供用于识别目的核苷酸序列集的方法，所述目的核苷酸序列能识别相关核苷酸序列组的成员，该方法包含将所述组中每个成员的核苷酸序列分成多个子序列的步骤，其中至少两个子序列重叠。

申请人已经发现，通过将研究的序列分成重叠的子序列，能够识别可用于与寡核苷酸探针杂交的目的核苷酸序列集。因此，子序列的相关组可以覆盖特定的基因座，而相关组的每个成员具有不同的核苷酸序列。在本发明方法的一种形式中，将相关序列组的每个成员分成多个子序列。在指定的成员

序列中，子序列互相重叠，从而可以产生潜在的大量子序列。这种方法与现有技术的方法截然不同，现有技术的方法是以使用连续的子序列为基础的。

优选地，至少一个子序列与多于一个的其它子序列重叠。更优选地，至少一个子序列与多于2、3、4或5个其它的子序列重叠。

用来产生重叠的探针长度子序列系列的重叠的程度可以是可能的最小值。一系列25-mer子序列的最小重叠的实例是，其中第一个子序列覆盖核苷酸1至25，第二个子序列覆盖核苷酸25至50，第三个子序列覆盖核苷酸50至75，等等。

重叠可以是最大可能程度的重叠。具有最大可能重叠的一系列25-mer子序列的实例是，其中第一个子序列覆盖核苷酸1至25，第二个子序列覆盖核苷酸2至26，第三个子序列覆盖核苷酸3至27，等等。

本发明包括最小和最大可能之间的任何中间程度的重叠。然而，优选使用基本上最大的重叠，因为这在设计探针集的个体方面要求的判断量最小。所用的重叠程度越高，覆盖相关序列中存在的SNP的更多组合的能力就越强。

对于具有任何指定成员的组，使用所述方法不必固定重叠的数目。也不必固定子序列的长度。对于技术人员，有可能可以按照常规方法来研究改变子序列长度和子序列之间重叠程度的影响，以确定是否获得了任何优势。

可以理解的是，当使用高度的重叠时，将产生非常大量的子序列。因此，探针集中将包括非常大量的探针。虽然微阵列芯片能够携带大量探针，但为了经济原因，至少期望限制对于指定分析所需的探针数目。在所述方法的一个实施方案中，应用中所需的探针数目能够通过识别冗余探针和在探针集中去除或不包括所述冗余探针而显著减少。在本领域中尚未认识到的是，当为了设计寡核苷酸探针集而分析相关序列时，一个成员序列中的多态性不必在另一个成员序列中存在。因此，不必提供覆盖每一个多态性的每一种组合的探针，因为在相关序列的组中不必存在所有组合。当相关序列多态性高时这种方法特别有用，而且按照本技术领域的现状预计，需要多于必需数目的探针来识别组中所有的理论成员。因此，在优选实施方案中，所述方法包括步骤：分析至少部分子序列的冗余性，和将识别为冗余的任何子序列中的至少部分去除。

可以使用相减方法实现冗余水平的降低，例如，假设组中所有成员都存在所有多态性，并根据该假设产生多个子序列。然后，分析所述多个子序列

中冗余序列是否存在，然后去除冗余序列，保留独特目的核苷酸序列集。预期的是，所述目的核苷酸子序列集与根据所有成员都存在所有多态性的假设而产生的较大的子序列集具有相同的识别组中每个成员的能力。

可选地，可以使用相加的方法，其中多个探针长度的序列一个一个渐增地产生，而且根据所有前面产生的子序列来分析每个新产生的子序列的冗余性。如果发现新产生的子序列是冗余的，就不将其加入目的核苷酸序列集中，否则就将其包括在目的核苷酸序列集中。无论使用相加还是相减的方法，最终结果是一样的：产生没有冗余，或者冗余水平降低的子序列集，其能识别相关序列组中的所有成员。

由于多种原因，期望限制识别成员序列所需的探针数目。在执行识别核苷酸序列的方法的经济可行性上，合成探针和产生携带那些探针的微阵列芯片的成本是主要考虑因素。无论单纯为了研究目的，或者为了高通量商业应用，比如在病理学实验室中，都会是这种情况。特别地，当核苷酸序列能具有很多可选形式时(即当相关序列组中的成员数目高时)，现有技术方法要求数目相当的大量不同的特异性探针。因此，要筛选是否存在单个成员核苷酸序列，根据要查询的序列的长度，可能必需要使用上百或甚至上千个单独探针。

限制识别成员核苷酸序列所必需的探针数目的另一个原因与特定探针杂交方法在实践中的局限性有关。例如，标准的点印迹仪器可能只有 64 个孔用于样品应用；从而限制了使用者只能用 64 个不同的探针，因此限制了识别能力，每次运行只能识别 64 个不同的核苷酸序列。进一步的实例是使用微阵列系统识别非常大量的可选形式的核苷酸序列。目前，标准微阵列芯片能装载多达 500,000 个不同的寡核苷酸探针。尽管这可能看起来已经很充足，但是对于一些应用，这个数目仍然不够，还有必要制备多个芯片来容纳所有探针。

在本方法的一个实施方案中，一个或多个子序列(和源自子序列的任何探针)在所述一个或多个子序列的 5'和/或 3'末端或末端附近不包含一个或多个多态性位点。在本方法的另一个实施方案中，一个或更多的子序列在所述一个或多个的子序列的中心或中心附近包含一个或多个多态位点。避免多肽位点接近探针侧翼，并将位点集中在探针中心，这可以克服由 Guo 等(2002)提供的探针的问题，即明显结合得不准确，从而产生大量假阳性杂交反应。

相关核酸序列可以是基因组核酸序列、RNA、cDNA 或 cRNA。在将基因组 DNA 样品应用到阵列之前，通常使用目的区域侧翼的引物对其进行扩

增。可以从实质上任何组织来源(除纯的红细胞之外)获得基因组 DNA。例如,合适的组织样品包括全血、精液、唾液、眼泪、尿液、排泄物料、汗、口腔组织样品、皮肤和毛发。如果从中获得样品的个体是纯合的,则扩增包含多态位点的基因组 DNA 将产生单一种类的目的核酸;或者如果所述个体是杂合的,则将产生两种目的分子。

可以用技术人员已知的任何合适的方法制备用于分析的 DNA,包括使用合适的引物的 PCR。当期望分析全基因组时,可以使用全基因组扩增的方法(WGA)。对于这种方法,可以方便地使用商业试剂盒,包括 Sigma-Aldrich Corp (St. Louis, MO, USA)制造的 GenoPlex<sup>®</sup> Complete WGA 试剂盒。这个试剂盒基于将基因组随机地分为片段形成系列模板。得到的较短的 DNA 链产生 DNA 片段文库,所述 DNA 片段具有确定的 3'和 5'末端。在初始阶段使用线性等温扩增,然后用有限循环的几何(PCR)扩增来复制文库。WGA 方法适合使用多种来源的纯化基因组 DNA,所述来源包括血液测试卡(blood card)、全血、口腔拭取物(buccal swaps)、粪便(soil)、植物和用甲醛溶液固定的石蜡包埋组织。

也经常扩增 mRNA 样品。在这种情况下,扩增通常在反转录之后进行。扩增所有表达的 mRNA 可以如 WO 96/14839 和 WO 97/01603 所述进行。如果从中取样的个体在表达的 mRNA 中存在的多肽位点是杂合的,则从二倍体样品扩增的 RNA 样品能产生两种目的分子。

显而易见的是,由本方法识别的核苷酸子序列将在后来用于设计探针集,所述探针集能识别相关序列组中所有现在已识别的成员。如本文所使用的,术语“目的核苷酸序列”指这样的序列,针对该序列可以产生基本上特异性的探针。将在下文进一步讨论探针的产生,然而探针通常是能与目的核苷酸序列杂交的寡核苷酸探针。

申请人发现,即使当相关序列组有大量成员时,和/或当成员有大量多态碱基时,和/或当多态碱基有多于两种的可选形式时,可能产生探针集,其能够使用明显少于本领域原来认为必需的数目的探针准确地识别组中的任何成员。所述方法可以用来,例如,产生能识别基因座的任何指定等位基因的探针集,而且当等位基因的数目非常大时特别有用。相反,Guo 等(2002)没有公开用于设计具有多外显子覆盖率、能提供全部等位基因分配的探针的实用而有力的方法。

技术人员将理解的是，探针长度的子序列的长度可以是任何长度，只要其可提供区分相关序列组中成员的能力即可。

用于微阵列应用的探针通常长约 25 个核苷酸，但是在本发明的上下文中，预计更长或更短的探针是有用的。需要足够的核苷酸提供结合的特异性，由此可以确定更短的有用长度，可以从约 10 个核苷酸至约 15 个核苷酸。当测试“亚基因组(sub-genome)”时，可能期望小于 15 个核苷酸的探针。这种情况的实例是当测试单个单倍体染色体的时候，而且序列检测特异性不要要求探针长度达到分析人类全基因组的近似 30 亿个核苷酸时所需的长度。可以通过与熔融双链区域和使多核苷酸的单链退火所需要相关的物理限制来确定上限。上限可以从约 30 至约 50 个核苷酸。上限可以随着 C/G 碱基的比例变化，与 A/T 碱基对相比，要分离双链中的这些碱基，需要更高的熔融温度。探针长度可以有实际的上限和下限，这些限制随着应用的特异性而变化，技术人员将能够通过常规的经验性实验识别长度最合适的探针。

可以理解的是，所述方法可以应用于有必要区别多个相关核苷酸序列的任何状况。如本文所使用的，术语“核苷酸序列”及其变化期望包括脱氧核糖核酸(DNA)和核糖核酸(RNA)序列。相关核苷酸序列可以是任何核苷酸序列组，其显示出最低水平的序列同一性。优选地，序列具有至少 50%、60%、70%、80%、90%、95%或 99%的同一性。当，例如相关序列较长，并且有一系列 SNP 散布于整个序列中时，同一性甚至可以高于 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与非蛋白质编码序列的组合。

相关序列可以源自二倍体、单倍体、三倍体或多倍体材料，或提供关于二倍体、单倍体、三倍体或多倍体状态的信息。

当搜索关于单倍体状态的信息时，本方法可用于提供探针，所述探针能对单倍型分层提供明确的 DNA 等位基因分配。基因座等位性的概念在本领域是已知的，然而原先并未认识到与区域结合的基因座的等位性，包括与同义变化有关的等位基因，有助于单倍体分层。因此，用于基因组(二倍体)DNA 的探针能报告关于单倍型(顺式连锁相(*cis phase*))多等位基因的分配。具体而言，同义等位基因是多基因座染色体单倍型片段中的单元。由本文所述方法产生的表征基因座等位性的探针有助于揭示多基因座共等位性的模式，其为单倍型。用端粒 G 和 F 基因座示例性说明这个概念。在 HLA-G 有 23 个等位

基因，在 HLA-F 有 20 个等位基因。这 43 个等位基因，与着丝粒 DPB1 基因座的 120 个等位基因结合，再连同其间的很多等位基因，帮助将有限的多基因座等位变化分配作为跨越 <4 Mb MHC 的区域中的单倍型。

相关序列可以是天然的或合成的。它们可以来自任何生物，包括动物、植物、微生物、细菌或病毒。

在本发明的一种形式里，相关序列指向基因组的同样区域。例如，从某个外显子的第一个核苷酸到该外显子的最后一个核苷酸的区域。在这种情况下，当使用 25-mer 探针时，可以如下设计探针，使探针的第 13 个核苷酸(即中心核苷酸)指向外显子的第一个核苷酸。因此，当第一个核苷酸是 G 时，探针的第 13 个核苷酸是 C。显而易见的是，探针的侧翼 12-mer 区域在一边将指向外显子前区域，而在另一边将进一步指向外显子中。

本发明的一个实施方案的常规操作可以用图 1 所示的极为简化的实例来证明。这种证明涉及 3 个相关核苷酸序列(# 1、# 2 和 # 3)，外显子从左边第 5 个核苷酸或 5' 末端(即 A)开始。将外显子中第一个核苷酸设为 1，则外显子在 6 和 11 位置有两个 SNP(下划线标出)。使用 9 个核苷酸的子序列，子序列中存在完全重叠。因此，第一个子序列从 -4 位置开始，在 +5 位置终止。

从图 1A 可以明显看出，每个相关序列分成了 11 个 9-mer 子序列。这提供了总共 33 个子序列(图 1B)。去除重复的子序列，剩余 17 个独特的子序列(图 1C)。技术人员将理解的是，如果初始目的分子是双链分子，则探针序列不需要互补。在这种情况下，可以直接使用核苷酸序列作为探针序列，或使其与 ACAGGGGTGTCGTGCAAAGAACCTC 互补，视由技术人员选择的序列产生策略而定。因此，如果在最终目的序列产生时使用 dsDNA，探针可以指向阵列上两条链中的任一条链，或同时指向两条链。

应该理解的是，这个实例仅用来证明产生探针集所需的步骤，所述探针集能根据本发明的一种形式区别相关核苷酸序列组的成员。在这种情况下，实现了将探针数目降低约 50%。在更复杂的系统中，探针数目的降低将更显著，可能会超过 95%。

本发明的方法可以分析核苷酸序列中的很多变化，包括缺失、取代、添加等。在本发明的一种形式中，除了存在 SNP 之外，相关核苷酸序列是一样的。

尽管 SNP 可能以任何密度存在，但是当 SNP 以高密度存在时，本方法提

供更大的优势。优选地，密度为在核苷酸序列的探针长度的区域中存在两个或更多的 SNP。对包括高密度 SNP 的相关核苷酸序列进行区别的能力在先前一直是个问题，因为至今为止一直认为必需提供大量探针来覆盖指定区域中 SNP 的每种组合。在设计探针集用于 HLA 分型时，这尤其会成为问题，其中 HLA 外显子中 20%至 50%的核苷酸是多态的，而且多态位点通常是成束的。这使现有技术提出预测，为了精确指定个体的 HLA 类型，需要的不同探针的数目是在实际上不可行的。

应该清楚的是，尽管组中相关核苷酸序列的数目可以低至两个，但是在相关核苷酸序列数目较多时，所述方法提供的优势将增加。在本方法的优选形式中，相关核苷酸序列组中的相关核苷酸序列数目多于 100、200、300、400、500、600、700、800、900 或 1000 个。本发明尤其可以应用于相关核苷酸序列数高而且 SNP 密度高时。

在所述方法的优选形式中，相关核苷酸序列是基因的等位基因。已知的是，编码同样蛋白的人类基因可以在不同的个体中具有不同的序列(等位基因)。分析的基因的比例可以是能够提供等位基因特异性信息的任何比例。例如，多态位点经常非随机性地分布于外显子全长中。因此，可能有必要将探针仅指向基因中的某些离散区域。

虽然大多数基因只有几个等位基因，但是一些基因的等位基因数目非常高。具有大量等位基因的基因实例主要涉及免疫系统，其中超变异性(hypervariability)是普遍特征。示例性基因包括主要组织相容性复合物(MHC)、T-细胞受体、B-细胞受体、免疫球蛋白、杀伤细胞抑制性受体(KIR)的基因等。然而可以理解的是，本文所述方法对于相关核苷酸序列的任何组都有用，但是当相关核苷酸序列是超变异性时，可以获得更大的优势。当超变异性作为高密度 SNP 存在时，还能提供更大的优势。

如上所指出的，MHC 基因是极端多态的。I 类和 II 类 MHC 跨膜蛋白组成了人白细胞抗原(HLA)系统，其用于以评价移植相容性为目的的组织分型。I 类蛋白由三个基因座编码：HLA-A、HLA-B 和 HLA-C，目前分别识别出 309、563 和 167 个等位基因。

II 类蛋白具有  $\alpha$  和  $\beta$  链，由基因座 DR、DQ 和 DP 编码。DR 基因座包含 3 个  $\alpha$  链的等位基因和 483 个  $\beta$  链的等位基因。DQ 基因座包含 25 个  $\alpha$  链的等位基因和 56 个  $\beta$  链的等位基因。DP 基因座包含 20 个  $\alpha$  链的等位基因和

107个 $\beta$ 链的等位基因。因此要注意的是,仅是I类区域,就有等位基因的很多组合来提供个体的HLA类型。

过去,使用对那些已经在人类群中识别出来的HLA抗原特异性的探针,根据血清学进行基于HLA的组织分型。现在,大多数HLA分型通过DNA方法进行,通过测序或相当于测序的方法进行高水平等位基因分配。这种DNA分型,有望改进组织分型的灵敏度和特异性。然而,通过基于DNA的方法(涉及寡核苷酸序列作为探针)识别所有HLA等位基因的尝试出现的问题是,需要非常大量的探针来覆盖所有可能的等位基因。本发明通过提供数量上可操作,同时仍能识别所有已知等位基因的探针集,从而减轻这个问题。

现在认为HLA-DR $\beta$ 基因座包含483个等位基因,直到将每个等位基因理解为是跨越全部外显子核苷酸分布的SNP的独特组合/模式,就会发现只需要483个探针(每个等位基因一个探针)。本领域通常认为甚至是双等位的SNP的存在都是探针设计中的显著问题,鉴于现在的微阵列SNP检测实践中使用了25-mer的寡核苷酸探针,那么有12-mer将在第13个位点SNP等位基因的侧翼。因此,当侧翼领域为非单态时,本领域至今认为必需要包括覆盖25-mer区域中每一种已知组合的每一个SNP的探针,即使并非每一种组合的每一个SNP都是天然存在的。本领域公认的是,任何多态位点需要的探针数为4的幂,幂次为在那个位点出现的已知等位基因的数量。因此,如果在两个侧翼中的侧翼12-mer中的每段各包含两个SNP,则将第13个位置的SNP分型需要的探针数目至少为4的2次幂=16。

申请人的方法不同,该方法基于这样的认识:并非在任何探针长度的子序列中为多态的所有位点都存在于HLA基因座的所有等位基因中。

不希望以任何方式受到理论的限制,提出对于HLA基因座,理论的可能性是观察到的等位序列的大概5-20倍高。复杂的高密度SNP基因座的实例是HLA-DRB域基因座(表达的DRB1、DRB3、DRB4、DRB5;假基因-不表达的DRB2、DRB6、DRB7、DRB8、DRB9)。在这个区域中的两个种类的基因中间有(大约)483个已经识别的等位基因。在可变的第二外显子中有270个核苷酸。简单的乘法得到,要解析这个基因座的基因型,需要130,410种不同的探针。这种观察结果可能有两个主要原因:(i)SNP的组合表现出连锁不平衡性,因为它们在染色体的长度上遗传,保证了SNP关联的不随机性;和(ii)群体经过了“瓶颈”,使一些多SNP等位基因消失,而其它等位基因的出现

频率相对增加，这受到群体遗传因子的影响，比如自然选择、重组倾向性等。

本发明使降低高度多态系统(比如 HLA 基因座)中识别基因型所必需探针的数目成为可能，使识别每个等位基因所需的所有探针可以固定于一块常规微阵列芯片上。

可以理解的是，精确识别等位基因所需探针的最终数量将依赖于所考察的基因座。然而，在本方法的优选形式中，期望相对于认为必需的探针的理论数目，有可能使探针数有超过 50%、60%、70%、80%、90%或 95%的减少。

尽管期望在去除所有冗余子序列的情况下，通过将探针数目最小化可以获得最大优势，但对于本发明来说，不需要去除所有冗余子序列。事实上，在一些例子中，具有优势的是保留子序列中的一些冗余，其中将产生内部质量控制机制。探针集中的冗余可能由于基因座之间发生冗余而导致。因此，为了进行质量控制，在通过本发明获得的探针集中，可以将与基因座之间的冗余相关的冗余探针保留。作为实例，当为了在 HLA 的 I 类和 II 类基因座分配等位基因类型，和在 KIR 基因座分配基因和等位基因类型而产生探针列表时，识别了大约 34,500 个探针。该列表可以识别在 HLA 的 I 类基因座(A、B、C)的超变外显子 2 和 3 以及 II 类基因座(DRB、DQB、DPB)的外显子 2 中涉及的变化，还能识别 KIR 基因座的多达 10 个外显子的所有已知变化。在探针列表中，当比较 HLA-A、-B 和 -C 时，或比较 DPB、DQB 和 DRB 时，因为序列存在直接重复，所以有 2167 个重复的序列，例如

探针标签(Probe Tag)	探针序列
5522A_E3_232_2_25	TCCGCAGATACCTGGAGAACAGGAA
15458C_E3_232_4_25	TCCGCAGATACCTGGAGAACAGGAA

探针标签	探针序列
9492B_E3_13_17_25	TCCAGAGGATGTTTGGCTGCGACCT
13765C_E3_13_10_25	TCCAGAGGATGTTTGGCTGCGACCT

探针标签	探针序列
22138R_E2_155_21_25	TGTCGCCGAGTACTGGAACAGCCAG
17957Q_E2_155_9_25	TGTCGCCGAGTACTGGAACAGCCAG

探针标签	探针序列
21088R_E2_105_3_25	TTCGACAGCGACGTGGGGGAGTTCC
17442Q_E2_105_3_25	TTCGACAGCGACGTGGGGGAGTTCC
16011P_E2_99_1_25	TTCGACAGCGACGTGGGGGAGTTCC

其中以下述方式标记探针

a=连续探针数

F=A、B、C、P、Q、R、K中的一种

E=外显子

c=外显子数

d=25-mer 在外显子中的第一个碱基

e=1-30, 1 是参考(共有), 后面是连续的独特等位基因类型

f=探针长度。

在本发明的一种形式中, 保留重复的探针序列, 对质量保证的技术和遗传成分都(technical and genetic component)有帮助。具体而言, 当与一个探针发生真实杂交时, 所述探针与识别第一个基因座的等位基因的所有其它探针的反应性一致, 但其中相同的探针序列不是第二个基因座上任何一种等位基因的完整组成成分, 则在该重复中的反应性与那些反映第二个基因座的等位基因的探针序列的反应性截然不同。

作为这种内部质量控制机制的操作实例, 最低解析水平是等位基因谱系或家族。考虑 DRB 时, 有 13 个谱系(lineage) (\*01、\*03、\*04、\*07、\*08、\*09、\*10、\*11、\*12、\*13、\*14、\*15、\*16)。包括用于所有四个表达 DRB 的基因座的探针, 由此 DRB3、DRB4 和 DRB5 是否存在可以提供关于 DRB1 等位基因谱系类型的信息, 其与 DRB1 探针的反应性无关。

在本发明的上下文中, 术语“冗余”用于表示如果将序列从子序列的第一个集合中去除, 而识别相关核苷酸序列组的成员的能力没有明显差别。可以将冗余作为完全(即两个子序列的核苷酸序列相同)或不完整的冗余(例如两个子序列在物理上不同, 但是功能相同)考虑。因此, 根据所用的杂交条件, 两个不同的探针可能结合到一个核苷酸序列上, 因此功能相同。当杂交条件具有相对低的严紧性时预计会出现这种情况。

可以根据原来使用 DNA 测序识别的等位基因, 产生无冗余或冗余性降低

的序列。如果识别出包含新多态性的新等位基因，那么可能需要在探针集中包含额外的目的序列，以确保能检测到这个新多态性。如果新多态性存在于原先认为是冗余的目的序列中，那么根据对该新多态性的认识，这个目的序列将成为必需的探针靶标，因此不是冗余的。

在本方法的一种形式中，方法易于自动化。现有技术的方法，比如 Guo 等(2002)提出的方法，根据对所有相关核苷酸序列的认真考虑来设计探针，尽量识别能覆盖 SNP 的所有观察到的组合的探针。这当然非常耗费人力，而且成功还是失败依赖于进行分析的个人的专业水平。如果相关序列的数目非常大，或者等位基因的数目非常大，那么设计探针的任务可能变得实际上不可行。相反，本方法特别易于在计算机上以基于软件的探针集设计形式执行。

本方法可以包括不同子序列长度和子序列间不同重叠水平的组合。在本发明的高度优选的形式中，子序列长约 25 个核苷酸，而且重叠程度最大。

相关序列可以包括来自基因的所有已知等位基因的序列。可选地，相关序列可以包括已知和迄今未知的序列。例如，可能已知在基因中指定位置发现多态性，而且所述位置可能有一或两种可选形式(例如 A 或 T)。则可以将 G 或 C 存在于该位置的“假设”序列包括在内。可选地，当不知道指定位置具有任何多态性但怀疑其有时，可以将涉及三种可选形式的探针包括在探针集中。此外，本发明将允许检测导致新等位基因的 SNP 新组合。这些方法非常需要探针，而本发明的使用大量减少所需的探针数目，使该方法实际可行。当子序列之间使用最大重叠时，发现新等位基因的机会将更大。

可以预期的是，通过内部质量控制机制，如前所讨论的，还可以发现迄今为止未识别的等位基因。与已知等位基因不同的探针反应性将标志着方法中存在错误，或者存在新的等位基因。

如前所讨论的，分析的等位基因可以唯一地指向蛋白编码域，或者唯一地指向非编码域。可选地，可以使用非编码和蛋白编码域的组合。

在另一个方面，本发明提供能特异性地与用本文所述方法识别的目的核苷酸序列杂交的探针集。在本发明的一种形式中，探针集的冗余水平低于用领域中已知方法设计的探针集。

给定了目的子序列，技术人员将能够合成能与每个目的子序列杂交的探针。探针与识别的非冗余序列基本上互补。如果从双链模板产生目的序列，则探针可以是正义或反义的。可以通过技术人员已知的任何方法生成探针，

尽管探针的最终用途很可能决定制备其最合适的方法。例如当探针用于微阵列环境中时，它们可以在形成阵列固体支撑基质的玻璃或尼龙片上原位合成。对于其它应用，探针可以在自动设备，比如 Beckman 1000M DNA 合成仪上合成，然后用比如 PCR 的方法检测等位基因。可选地，探针可以在制造后与固体支撑体偶联。

在技术人员能力范围内，可以很好地研究通过在用即时方法(instant method)设计的探针中使用修饰的核苷酸，比如锁定核酸(locked nucleic acid)，是否具有任何优势。

为了保证质量，探针集可选地在阵列上为每个完美地匹配目的序列的探针包括了与其成对的“错配”探针。错配探针包含直接位于 25 碱基的探针序列中间的单个错配。当完美匹配探针在样品与其结合时提供可测量的荧光时，使用与其成对的错配探针检测并消除测量中的任何错误或污染的荧光。错配探针作为对其完美匹配的配偶体(partner)的内部控制，因为它可以像它的对应匹配探针一样有效地结合非特异性序列，使得假的信号，例如从交叉杂交中产生的假的信号，可以被有效地定量，并从基因型响应(genotype call)或基因表达测量中减去。

探针可以包括标记以方便检测。示例性的标记包括 Cy5、Cy3、FITC、罗丹明(rhodamine)、生物素、DIG 和各种放射性同位素。

可以根据本发明产生的探针序列列表扩展至包括下述位置的其它等位基因变化：在 mRNA 转录物内的其它外显子处，在插入外显子或外显子侧翼的序列处，其中包括内含子、5'和 3'末翻译区域和基因间的区域。

在另一方面，本发明提供使用本文所述的探针集识别相关核苷酸序列组中成员的方法。实现这点的一种方法是使用微阵列技术。因此，另一方面，本发明提供固定于固体基质上的如本文所述的探针集。本发明的这种形式的示例性实施方案可参见由 Affymetrix®市场化的 GeneChip®技术。这种技术依赖于影印光刻法(photolithographic process)，其用光敏型化合物涂覆 5"×5"的石英片，所述化合物防止石英片与产生的 DNA 探针的第一个核苷酸偶联。使用印刷蒙片(lithographic mask)阻止或传送光到达片表面的特定位置。然后用包含腺嘌呤、胸腺嘧啶、胞嘧啶或鸟嘌呤的溶液淹没表面，在玻璃上只有在通过照射已经脱保护的那些区域中发生偶联。偶联的核苷酸还带有光敏保护基团，使循环可以重复。很多公司，包括 Oxford Gene Technology (Oxford,

U.K.)、Agilent Technologies (Palo Alto, CA, U.S.A.)和 Nimblegen Systems Inc (Madison, WI, U.S.A), 提供固定探针的其它方法。

在另一方面, 本发明提供计算机可执行程序(软件), 其能执行本文所述的方法。虽然本发明可以手动进行, 但是优选在合适软件的指导下, 在个人计算机上执行。本文给出了本公开, 技术人员将能够写出合适的代码来执行本方法。用于 0101 等位基因 DRB1 基因座的实例伪代码(pseudo-code)遵循:

[等待用户输入]

(IF) 在网格界面检测到鼠标点击动作;

[确定]点击的网格行与网格列;

/\* 因为所有序列以表格形式显示, 所以也根据下述以表格形式作为内存对象储存:

参照名称阵列[位置 0] = “DRB10101”;

参照基础阵列[位置 0] = “TGTCCCCA.....”;

其在内存中形成表格结构, 如:

参照名称阵列

参照基础阵列

索引0: “DRB01\*010101”

“TGTCCCCA.....”

索引1: “DRB01\*010102”

“TGTCCCCC.....”

索引2: “DRB01\*010103”

“TGTCCCCC.....”

\*/

[确定]使用点击的网格列数值作为索引值, 确定参照基础阵列的碱基范围 (25 个碱基)。

[确定]使用点击的网格行作为索引值, 确定参照名称阵列

/\*\*

如何确定 25 个碱基的范围?

如果参照基础阵列(即: 所有碱基的阵列)包含 150 个碱基, 则使用网格列点击值确定中间点。

即:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	20	21	..n	
DRB01*010101:	G	T	G	T	C	C	C	C	A	C	A	G	...							

如果用户点击第 12 列, 那么我们的范围就变成, 最小值 = 中间点 - 12,

最大值 = 中间点 + 12;

\*\*/

[提取] 从每个参照基础阵列记录中提取 25 个碱基

(IF)碱基与参照记录不同

[隐藏/去除] 行

else

[显示行]

软件可以方便地研究参数范围对解析特异性等位基因所需探针数目的影响。用这种方法，可能进一步减少所需探针的数目。例如，软件可以允许用户确定探针长度子序列的长度、子序列重叠的程度、确定两个子序列是否冗余的规则等。事实上，软件可以包括算法，自动尝试每个参数的范围，给出探针长度子序列的最小数目(因此给出探针集中探针的数目)。如果认为探针很可能具有显著的二级结构，或者熔融温度太高或太低使该探针不能可靠地与相关靶标杂交，也可以将该探针从探针集中去除。基于证明缺乏适合性的经验性探针优化实验，可将探针从探针集中去除。

可以预期的是，本发明将在广泛的技术领域中得到应用。可以期望的是，医药领域将获得特别的优势，在医药领域中，本方法可以用于个体的基因分型。在移植组织分型中(例如使用 HLA 基因、KIR 基因、次要组织相容性基因座等)，以及药物基因组学、DNA “指纹分析” 等中，本方法特别有用。探针可以用于任何应用，包含原位杂交、狭缝印迹(slot blot)、斑点印迹(dot blot)、菌落杂交、噬菌斑杂交、Northern 印迹、Southern 印迹，以及微阵列应用。

可以预期的是，本发明将可以用于任何应用，所述应用必需或者期望减少分析核苷酸序列所需的独特探针的数目，而且不仅在微阵列分析领域。甚至在执行从大量其它序列中识别特定核苷酸序列这一任务时，当所需的探针数目并没有超出芯片容量时，本发明也将适用。将探针数目最小化使一块芯片中可以包括对其它基因座的测试，从而将一块芯片能够测试的基因座的数目增加。和运行 20 块芯片相比，运行一块芯片的成本当然更低。

可以预期的是，应用将扩展到非人类动物中，比如灵长类动物，例如对候选药物的临床前药物基因组学评价。本发明还预期可以用于测试具有经济重要性的动物(比如牲畜、家禽等)，例如在育种项目中改善参数，比如瘦肌肉(lean muscle)含量。

现在将参照下述非限定性实例进一步描述本发明。技术人员将理解的是，HLA 基因座是自然界中发现的变化最多的一些基因座。可以赞同的是，能针对 HLA 基因座操作的方法，则可以操作任何其它基因座。

实施例 1: 识别用于 HLA-DRB 基因座的精确基因分型的寡核苷酸探针集

#### 步骤概要

用本方法分析 HLA 的 DRB 基因座，识别能识别所述基因座的任何已知等位基因的探针集。DRB 基因座具有下述已知的等位基因：

DRB1\*010101、DRB1\*010102、DRB1\*010103、DRB1\*010201、DRB1\*010202、DRB1\*010203、DRB1\*010204、DRB1\*0103、DRB1\*0104、DRB1\*0105、DRB1\*0106、DRB1\*0107、DRB1\*0108、DRB1\*0109、DRB1\*0110、DRB1\*0111、DRB1\*0112、DRB1\*0113、DRB1\*030101、DRB1\*030102、DRB1\*030201、DRB1\*030202、DRB1\*0303、DRB1\*0304、DRB1\*030501、DRB1\*030502、DRB1\*0306、DRB1\*0307、DRB1\*0308、DRB1\*0309、DRB1\*0310、DRB1\*0311、DRB1\*0312、DRB1\*0313、DRB1\*0314、DRB1\*0315、DRB1\*0316、DRB1\*0317、DRB1\*0318、DRB1\*0319、DRB1\*0320、DRB1\*0321、DRB1\*0322、DRB1\*0323、DRB1\*0324、DRB1\*0325、DRB1\*0326、DRB1\*0327、DRB1\*0328、DRB1\*040101、DRB1\*040501、DRB1\*040502、DRB1\*040503、DRB1\*040504、DRB1\*0406、DRB1\*040701、DRB1\*040702、DRB1\*040703、DRB1\*0408、DRB1\*0409、DRB1\*0410、DRB1\*0411、DRB1\*0412、DRB1\*0413、DRB1\*0414、DRB1\*0415、DRB1\*0416、DRB1\*0417、DRB1\*0418、DRB1\*0419、DRB1\*0420、DRB1\*0421、DRB1\*0422、DRB1\*0423、DRB1\*0424、DRB1\*0425、DRB1\*0426、DRB1\*0427、DRB1\*0428、DRB1\*0429、DRB1\*0430、DRB1\*0431、DRB1\*0432、DRB1\*0433、DRB1\*0434、DRB1\*0435、DRB1\*0436、DRB1\*0437、DRB1\*0438、DRB1\*0439、DRB1\*0440、DRB1\*0441、DRB1\*0442、DRB1\*0443、DRB1\*0444、DRB1\*0445、DRB1\*0446、DRB1\*0447、DRB1\*0448、DRB1\*0449、DRB1\*0450、DRB1\*0451、DRB1\*0452、DRB1\*070101、DRB1\*070102、DRB1\*0703、DRB1\*0704、DRB1\*0705、DRB1\*0706、

DRB1\*0707、DRB1\*0708、DRB1\*0709、DRB1\*080101、DRB1\*080102、  
DRB1\*080201、DRB1\*080202、DRB1\*080203、DRB1\*080302、DRB1\*080401、  
DRB1\*080402、DRB1\*080403、DRB1\*080404、DRB1\*0805、DRB1\*0806、  
DRB1\*0807、DRB1\*0808、DRB1\*0809、DRB1\*0810、DRB1\*0811、  
DRB1\*0812、DRB1\*0813、DRB1\*0814、DRB1\*0815、DRB1\*0816、  
DRB1\*0817、DRB1\*0818、DRB1\*0819、DRB1\*0820、DRB1\*0821、  
DRB1\*0822、DRB1\*0823、DRB1\*0824、DRB1\*0825、DRB1\*0826、  
DRB1\*0827、DRB1\*0828、DRB1\*0829、DRB1\*090102、DRB1\*0902、  
DRB1\*0903、DRB1\*0904、DRB1\*100101、DRB1\*100102、DRB1\*110101、  
DRB1\*110102、DRB1\*110103、DRB1\*110104、DRB1\*110105、DRB1\*1102、  
DRB1\*1103、DRB1\*110401、DRB1\*110402、DRB1\*1105、DRB1\*110601、  
DRB1\*110602、DRB1\*1107、DRB1\*110801、DRB1\*110802、DRB1\*1109、  
DRB1\*1110、DRB1\*1111、DRB1\*111201、DRB1\*111202、DRB1\*1113、  
DRB1\*1114、DRB1\*1115、DRB1\*1116、DRB1\*1117、DRB1\*1118、  
DRB1\*111901、DRB1\*111902、DRB1\*1120、DRB1\*1121、DRB1\*1122、  
DRB1\*1123、DRB1\*1124、DRB1\*1125、DRB1\*1126、DRB1\*112701、  
DRB1\*112702、DRB1\*1128、DRB1\*1129、DRB1\*1130、DRB1\*1131、  
DRB1\*1132、DRB1\*1133、DRB1\*1134、DRB1\*1135、DRB1\*1136、DRB1\*1137、  
DRB1\*1138、DRB1\*1139、DRB1\*1140、DRB1\*1141、DRB1\*1142、DRB1\*1143、  
DRB1\*1144、DRB1\*1145、DRB1\*1146、DRB1\*1147、DRB1\*1148、DRB1\*1149、  
DRB1\*1150、DRB1\*1151、DRB1\*1152、DRB1\*1153、DRB1\*1154、  
DRB1\*120101、DRB1\*120102、DRB1\*120201、DRB1\*120202、DRB1\*120302、  
DRB1\*1204、DRB1\*1205、DRB1\*1206、DRB1\*1207、DRB1\*1208、  
DRB1\*1209、DRB1\*1210、DRB1\*1211、DRB1\*130101、DRB1\*130102、  
DRB1\*130103、DRB1\*130201、DRB1\*130202、DRB1\*130301、DRB1\*130302、  
DRB1\*1304、DRB1\*1305、DRB1\*1306、DRB1\*130701、DRB1\*130702、  
DRB1\*1308、DRB1\*1309、DRB1\*1310、DRB1\*1311、DRB1\*1312、  
DRB1\*1313、DRB1\*131401、DRB1\*131402、DRB1\*1315、DRB1\*1316、  
DRB1\*1317、DRB1\*1318、DRB1\*1319、DRB1\*1320、DRB1\*1321、  
DRB1\*1322、DRB1\*1323、DRB1\*1324、DRB1\*1325、DRB1\*1326、  
DRB1\*1327、DRB1\*1328、DRB1\*1329、DRB1\*1330、DRB1\*1331、

DRB1\*1332、DRB1\*1333、DRB1\*1334、DRB1\*1335、DRB1\*1336、  
DRB1\*1337、DRB1\*1338、DRB1\*1339、DRB1\*1340、DRB1\*1341、  
DRB1\*1342、DRB1\*1343、DRB1\*1344、DRB1\*1345、DRB1\*1346、  
DRB1\*1347、DRB1\*1348、DRB1\*1349、DRB1\*1350、DRB1\*1351、  
DRB1\*1352、DRB1\*1353、DRB1\*1354、DRB1\*1355、DRB1\*1356、  
DRB1\*1357、DRB1\*1358、DRB1\*1359、DRB1\*1360、DRB1\*1361、  
DRB1\*1362、DRB1\*1363、DRB1\*1364、DRB1\*1365、DRB1\*1366、  
DRB1\*140101、DRB1\*140102、DRB1\*1402、DRB1\*140301、DRB1\*140302、  
DRB1\*1404、DRB1\*140501、DRB1\*140502、DRB1\*1406、DRB1\*140701、  
DRB1\*140702、DRB1\*1408、DRB1\*1409、DRB1\*1410、DRB1\*1411、  
DRB1\*1412、DRB1\*1413、DRB1\*1414、DRB1\*1415、DRB1\*1416、  
DRB1\*1417、DRB1\*1418、DRB1\*1419、DRB1\*1420、DRB1\*1421、  
DRB1\*1422、DRB1\*142301、DRB1\*142302、DRB1\*1424、DRB1\*1425、  
DRB1\*1426、DRB1\*1427、DRB1\*1428、DRB1\*1429、DRB1\*1430、  
DRB1\*1431、DRB1\*1432、DRB1\*1433、DRB1\*1434、DRB1\*1435、  
DRB1\*1436、DRB1\*1437、DRB1\*1438、DRB1\*1439、DRB1\*1440、  
DRB1\*1441、DRB1\*1442、DRB1\*1443、DRB1\*1444、DRB1\*1445、  
DRB1\*1446、DRB1\*1447、DRB1\*1448、DRB1\*150101、DRB1\*150102、  
DRB1\*150103、DRB1\*150104、DRB1\*150105、DRB1\*150201、DRB1\*150202、  
DRB1\*150203、DRB1\*1503、DRB1\*1504、DRB1\*1505、DRB1\*1506、  
DRB1\*1507、DRB1\*1508、DRB1\*1509、DRB1\*1510、DRB1\*1511、  
DRB1\*1512、DRB1\*1513、DRB1\*1514、DRB1\*1515、DRB1\*160101、  
DRB1\*160102、DRB1\*160201、DRB1\*160202、DRB1\*1603、DRB1\*1604、  
DRB1\*160501、DRB1\*160502、DRB1\*1607、DRB1\*1608、DRB2\*0101、  
DRB3\*010101、DRB3\*01010201、DRB3\*01010202、DRB3\*010103、  
DRB3\*010104、DRB3\*0102、DRB3\*0103、DRB3\*0104、DRB3\*0105、  
DRB3\*0106、DRB3\*0107、DRB3\*0108、DRB3\*0109、DRB3\*0110、  
DRB3\*0111、DRB3\*0201、DRB3\*020201、DRB3\*020202、DRB3\*020203、  
DRB3\*020204、DRB3\*0203、DRB3\*0204、DRB3\*0205、DRB3\*0206、  
DRB3\*0207、DRB3\*0208、DRB3\*0209、DRB3\*0210、DRB3\*0211、  
DRB3\*0212、DRB3\*0213、DRB3\*0214、DRB3\*0215、DRB3\*0216、

DRB3\*0217、DRB3\*0218、DRB3\*0219、DRB3\*030101、DRB3\*030102、DRB3\*0302、DRB3\*0303、DRB4\*01010101、DRB4\*0102、DRB4\*010310101、DRB4\*01030102N、DRB4\*010302、DRB4\*010303、DRB4\*010304、DRB4\*0104、DRB4\*0105、DRB4\*0106、DRB4\*0107、DRB4\*0201N、DRB4\*0301N、DRB5\*010101、DRB5\*010102、DRB5\*0102、DRB5\*0103、DRB5\*0104、DRB5\*0105、DRB5\*0106、DRB5\*0107、DRB5\*0108N、DRB5\*0109、DRB5\*0110N、DRB5\*0111、DRB5\*0112、DRB5\*0113、DRB5\*0202、DRB5\*0203、DRB5\*0204、DRB5\*0205、DRB6\*0101、DRB6\*0201、DRB6\*0202、DRB7\*010101、DRB7\*010102、DRB8\*0101 和 DRB9\*0101。

选择 25 个核苷酸的子序列长度，使用最大连续重叠提供子序列系列。选择第二外显子作为分析的起始点，定位第一个 25-mer 子序列，使所述子序列的第 13 个核苷酸(下划线标出，见下文)与第二外显子的第一个碱基对齐。下面使用很多 DRB 等位基因中的典型参照序列表示如下：

内含子 1 \_\_\_\_\_ 外显子 2 \_\_\_\_\_ ...  
GTGTCCCCACAGCACGTTTCTTGTG...

步骤 1: 确定子序列，其用于选择中心位于第二外显子的第一个核苷酸的探针。

第一个受试子序列是内含子 1 和外显子 2 交界处的 DRB 基因座的 25 个核苷酸子序列。针对外显子 1 中的第一个核苷酸(用下划线标出的“C”残基)，产生第一个子序列：GTGTCCCCACAGCACGTTTCTTGTG (这个序列是 26 个等位基因中发现的参照序列)。

步骤 2: 确定子序列，其用于选择中心位于第二外显子的第二个核苷酸的探针。

重复步骤 1 中的过程，但是 25-mer 子序列的中心在第二个核苷酸。此外，鉴于参照序列，25-mer 是：TGTCTCCCCACAGCACGTTTCTTGTGG。

步骤 3 至 284。确定子序列，其用于选择中心位于第二外显子的第 3 个至第 284 个核苷酸的探针

对外显子中的每个核苷酸，重复步骤 1 中的过程。

步骤 285: 汇集 25-mer 子序列

将用于基因座的每个等位基因的所有 25-mer 子序列组合，形成能识别基

因座的所有等位基因的目的核苷酸序列集。

#### 步骤 286: 去除冗余子序列

分析所有子序列, 去除冗余序列(完全匹配), 只保留独特的子序列。据估计, 如果对第二外显子的所有 270 个核苷酸进行这个过程, 大约只产生 5,500 个独特的子序列。这使现有技术中预计的探针数目显著减少。

#### 实施例 2: 产生微阵列芯片

由 Affymetrix Inc 提供客户基因芯片阵列服务, 直接在微阵列芯片上合成集合中的 5,500 个目的核苷酸序列。

#### 实施例 3: 使用探针为个体分配识别 DRB 等位基因

##### 患者样品

外周血和颊膜涂片的 DNA 提取按标准实施。推荐在微阵列测试中使用约 1,000 ng 的 DNA。

##### 长 PCR。

引物可以位于内含子、外显子或其组合中。例如, 对于 HLA-DRB 分型, 在内含子 1 的上游和外显子 6 的下游选择引物。扩增子约 5.1 kb。使用内含子序列作为引物位点的不足之处在于与相应的外显子序列相比, 通常序列数据较少, 而且缺乏对应于外显子等位基因的数据。对于 HLA-DRB, 公开的数据为引物选择提供了足够的内含子 1 数据。然而, 即使在这种情况下, 进一步的测序几乎必然会揭示新的 SNP。如果 SNP 在引物序列中, 则可以预见将使扩增携带所述新变体的序列复杂化。可选的是使用外显子序列, 因为这些序列得到了更深入的研究。对于 HLA-DRB, 在更上游处有适于做引物的位点, 位于外显子 1 中。因为使用外显子 1 和外显子 6 引物得到的扩增子包括了 8 kb 内含子 1 的全长, 所以得到的扩增子长度超过 13 kb。申请人确认了用于扩增 17 kb 的商业长 PCR 试剂盒的适合性, 所以只用外显子作为引物引导的扩增子也是适用的。

##### 扩增子的片段化(fragmentation)

过程是非特异性的, 结果将扩增子剪切成与附着于芯片的探针充分杂交

所需的几十至几百个(百位数字较小) (low hundreds)核苷酸的片段。下述文献 GeneChip® CustomSeq™ Resequencing (Array Protocol) 2.0 版, 701231 Rev. 3 中给出了细节; 其全部内容通过引用并入本文。这个文献可以从 Affymetrix Inc(技术支持)3380 Central Expressway Santa Clara, CA 95051 U.S.A.获得。

### 杂交

GeneChip® CustomSeq™ Resequencing (Array Protocol) 2.0 版, 701231 Rev. 3 中给出了细节。

### 等位基因分配

通过将探针杂交模式与等位序列变化相关联, 用迭代还原算法(iterative reduction algorithm) (Helmberg W, Lanzer G, Zahn R, Weinmayr B, Wagner T, Albert E. Virtual DNA analysis-a new tool for combination and standardised evaluation of SSO, SSP and sequencing-based typing results. Tissue Antigens. 1998 Jun; 51(6): 587-92)实现等位基因分配。

实施例 4: 产生用于在 HLA-A\*0201 (外显子 2 和 3)分配等位基因的探针集

使用 HLA\*0201 的下述外显子序列来产生用于 HLA-A\*0201 的分配的探针集。就产生探针而言, 将外显子序列在 5'和 3'两个方向均延伸 12 个核苷酸, 进入相邻的内含区域。

#### 外显子 2:

```
GCTCCCACTCCATGAGGTATTTCTTCACATCCGTGTCCCGGCCCGGCCGCGGGGAGCCC
CGCTTCATCGCCGTGGGCTACGTGGACGACACGCAGTTCGTGCGGTTGACAGCGACGC
CGCGAGCCAGAGGATGGAGCCGCGGGCGCCGTGGATAGAGCAGGAGGGTCCGGAGTATT
GGGACGGGGAGACACGAAAGTGAAGGCCCACTCACAGACTCACCGAGTGGACCTGGGG
ACCCTGCGCGGCTACTACAACCAGAGCGAGGCCG
```

#### 外显子 3

```
GTTCTCACACCGTCCAGAGGATGTATGGCTGCGACGTGGGGTCGGACTGGCGCTTCCTC
CGCGGGTACCACCAGTACGCTACGACGGCAAGGATTACATCGCCCTGAAAGAGGACCT
GCGCTCTTGACCGCGGCGGACATGGCAGCTCAGACCACCAAGCACAAGTGGGAGGCCG
```

---

CCCATGTGGCGGAGCAGTTGAGAGCCTACCTGGAGGGCACGTGCGTGGAGTGGCTCCGC  
AGATACCTGGAGAACGGGAAGGAGACGCTGCAGCGCACGG

选择子序列长度为 25，使用最大重叠。

图 2 显示了能识别上述超变外显子 2/3 区域的探针集。当期望识别上述之外的超变区域时，对每个超变区域重复探针产生过程。然后可以去除冗余的探针序列。

最后，可以理解的是，在不背离如本文所指出的本发明精髓的情况下，可以做出各种其它的修正和/或改变。

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

图 1A

ATCGATCGA  
TCGATCGAT  
CGATCGATC  
GATCGATCG  
ATCGATCGA  
TCGATCGAT  
CGATCGATC  
GATCGATCG  
ATCGATCGA  
TCGATCGAT  
CGATCGATC

ATCGATCGA  
TCGATCGAT  
CGATCGATG  
GATCGATGG  
ATCGATGGA  
TCGATGGAT  
CGATGGATC  
GATGGATCG  
ATGGATCGA  
TGGATCGAT  
GGATCGATC

ATCGATCGA  
TCGATCGAT  
CGATCGATC  
GATCGATCG  
ATCGATCGA  
TCGATCGAT  
CGATCGATC  
GATCGATCC  
ATCGATCCA  
TCGATCCAT  
CGATCCATC

图 1B

ATCGATCGA  
TCGATCGAT  
CGATCGATC  
GATCGATCG  
CGATCGATG  
GATCGATGG  
ATCGATGGA  
TCGATGGAT  
CGATGGATC  
GATGGATCG  
ATGGATCGA  
TGGATCGAT  
GGATCGATC  
GATCGATCC  
ATCGATCCA  
TCGATCCAT  
CGATCCATC

图 1C

**外显子 2**

CGAGGGTGAGGTACTCCATAAAGAA  
 GAGGGTGAGGTACTCCATAAAGAAG  
 AGGGTGAGGTACTCCATAAAGAAGT  
 GGGTGAGGTACTCCATAAAGAAGTG  
 GGTGAGGTACTCCATAAAGAAGTGT  
 GTGAGGTACTCCATAAAGAAGTGT  
 TGAGGTACTCCATAAAGAAGTGTAG  
 GAGGTACTCCATAAAGAAGTGTAGG  
 AGGTACTCCATAAAGAAGTGTAGGC  
 GGTACTCCATAAAGAAGTGTAGGCA  
 GTACTCCATAAAGAAGTGTAGGCAC  
 TACTCCATAAAGAAGTGTAGGCACA  
 ACTCCATAAAGAAGTGTAGGCACAG  
 CTCCATAAAGAAGTGTAGGCACAGG  
 TCCATAAAGAAGTGTAGGCACAGGG  
 CCATAAAGAAGTGTAGGCACAGGGC  
 CATAAAGAAGTGTAGGCACAGGGCC  
 ATAAAGAAGTGTAGGCACAGGGCCG  
 TAAAGAAGTGTAGGCACAGGGCCGG  
 AAAGAAGTGTAGGCACAGGGCCGGG  
 AAGAAGTGTAGGCACAGGGCCGGGC  
 AGAAGTGTAGGCACAGGGCCGGGCC  
 GAAGTGTAGGCACAGGGCCGGGCCG  
 AAGTGTAGGCACAGGGCCGGGCCGG  
 AGTGTAGGCACAGGGCCGGGCCGGC  
 GTGTAGGCACAGGGCCGGGCCGGCG  
 TGTAGGCACAGGGCCGGGCCGGCGC  
 GTAGGCACAGGGCCGGGCCGGCGCC  
 TAGGCACAGGGCCGGGCCGGCGCCC  
 AGGCACAGGGCCGGGCCGGCGCCCC  
 GGCACAGGGCCGGGCCGGCGCCCCCT  
 GCACAGGGCCGGGCCGGCGCCCCCTC  
 CACAGGGCCGGGCCGGCGCCCCCTCG  
 ACAGGGCCGGGCCGGCGCCCCCTCGG  
 CAGGGCCGGGCCGGCGCCCCCTCGGG  
 AGGGCCGGGCCGGCGCCCCCTCGGGG  
 GGGCCGGGCCGGCGCCCCCTCGGGGC  
 GGCCGGGCCGGCGCCCCCTCGGGGCG  
 GCCGGGCCGGCGCCCCCTCGGGGCGA  
 CCGGGCCGGCGCCCCCTCGGGGCGAA  
 CGGGCCGGCGCCCCCTCGGGGCGAAGT  
 GGCCGGCGCCCCCTCGGGGCGAAGTA  
 GCCGGCGCCCCCTCGGGGCGAAGTAG  
 CCGGGCGCCCCCTCGGGGCGAAGTAGC  
 CGGGCGCCCCCTCGGGGCGAAGTAGCG  
 GCGCCCCCTCGGGGCGAAGTAGCGG  
 GCGCCCCCTCGGGGCGAAGTAGCGGC  
 CGCCCCCTCGGGGCGAAGTAGCGGCA  
 GCCCCTCGGGGCGAAGTAGCGGCAC  
 CCCCTCGGGGCGAAGTAGCGGCACC  
 CCCTCGGGGCGAAGTAGCGGCACCC

图 2

CCTCGGGGCGAAGTAGCGGCACCCG  
CTCGGGGCGAAGTAGCGGCACCCGA  
TCGGGGCGAAGTAGCGGCACCCGAT  
CGGGGCGAAGTAGCGGCACCCGATG  
GGGGCGAAGTAGCGGCACCCGATGC  
GGGCGAAGTAGCGGCACCCGATGCA  
GGCGAAGTAGCGGCACCCGATGCAC  
GCGAAGTAGCGGCACCCGATGCACC  
CGAAGTAGCGGCACCCGATGCACCT  
GAAGTAGCGGCACCCGATGCACCTG  
AAGTAGCGGCACCCGATGCACCTGC  
AGTAGCGGCACCCGATGCACCTGCT  
GTAGCGGCACCCGATGCACCTGCTG  
TAGCGGCACCCGATGCACCTGCTGT  
AGCGGCACCCGATGCACCTGCTGTG  
GCGGCACCCGATGCACCTGCTGTGC  
CGGCACCCGATGCACCTGCTGTGCG  
GGCACCCGATGCACCTGCTGTGCGT  
GCACCCGATGCACCTGCTGTGCGTC  
CACCCGATGCACCTGCTGTGCGTCA  
ACCCGATGCACCTGCTGTGCGTCAA  
CCCGATGCACCTGCTGTGCGTCAAG  
CCGATGCACCTGCTGTGCGTCAAGC  
CGATGCACCTGCTGTGCGTCAAGCA  
GATGCACCTGCTGTGCGTCAAGCAC  
ATGCACCTGCTGTGCGTCAAGCACG  
TGCACCTGCTGTGCGTCAAGCACGC  
GCACCTGCTGTGCGTCAAGCACGCC  
CACCTGCTGTGCGTCAAGCACGCCA  
ACCTGCTGTGCGTCAAGCACGCCAA  
CCTGCTGTGCGTCAAGCACGCCAAG  
CTGCTGTGCGTCAAGCACGCCAAGC  
TGCTGTGCGTCAAGCACGCCAAGCT  
GCTGTGCGTCAAGCACGCCAAGCTG  
CTGTGCGTCAAGCACGCCAAGCTGT  
TGTGCGTCAAGCACGCCAAGCTGTC  
GTGCGTCAAGCACGCCAAGCTGTGCG  
TGCGTCAAGCACGCCAAGCTGTGCG  
GCGTCAAGCACGCCAAGCTGTGCGT  
CGTCAAGCACGCCAAGCTGTGCGTG  
GTCAAGCACGCCAAGCTGTGCGTGC  
TCAAGCACGCCAAGCTGTGCGTGCG  
CAAGCACGCCAAGCTGTGCGTGCGG  
AAGCACGCCAAGCTGTGCGTGCGGC  
AGCACGCCAAGCTGTGCGTGCGGCG  
GCACGCCAAGCTGTGCGTGCGGCGC  
CACGCCAAGCTGTGCGTGCGGCGCT  
ACGCCAAGCTGTGCGTGCGGCGCTC  
CGCCAAGCTGTGCGTGCGGCGCTCG  
GCCAAGCTGTGCGTGCGGCGCTCGG  
CCAAGCTGTGCGTGCGGCGCTCGGT  
CAAGCTGTGCGTGCGGCGCTCGGTC  
AAGCTGTGCGTGCGGCGCTCGGTCT  
AGCTGTGCGTGCGGCGCTCGGTCTC  
GCTGTGCGTGCGGCGCTCGGTCTCC  
CTGTGCGTGCGGCGCTCGGTCTCCT  
TGTCGCTGCGGCGCTCGGTCTCCTA

图 2(续)

GTGCTGCGGCGCTCGGTCTCCTAC  
TCGCTGCGGCGCTCGGTCTCCTACC  
CGCTGCGGCGCTCGGTCTCCTACCT  
GCTGCGGCGCTCGGTCTCCTACCTC  
CTGCGGCGCTCGGTCTCCTACCTCG  
TGCGGCGCTCGGTCTCCTACCTCGG  
GCGGCGCTCGGTCTCCTACCTCGGC  
CGGCGCTCGGTCTCCTACCTCGGCG  
GGCGCTCGGTCTCCTACCTCGGCGC  
GCGCTCGGTCTCCTACCTCGGCGCC  
CGCTCGGTCTCCTACCTCGGCGCCC  
GCTCGGTCTCCTACCTCGGCGCCCG  
CTCGGTCTCCTACCTCGGCGCCCGC  
TCGGTCTCCTACCTCGGCGCCCGCG  
CGGTCTCCTACCTCGGCGCCCGCGG  
GGTCTCCTACCTCGGCGCCCGCGC  
GTCTCCTACCTCGGCGCCCGCGGCA  
TCTCCTACCTCGGCGCCCGCGGCAC  
CTCCTACCTCGGCGCCCGCGGCACC  
TCCTACCTCGGCGCCCGCGGCACCT  
CCTACCTCGGCGCCCGCGGCACCTA  
CTACCTCGGCGCCCGCGGCACCTAT  
TACCTCGGCGCCCGCGGCACCTATC  
ACCTCGGCGCCCGCGGCACCTATCT  
CCTCGGCGCCCGCGGCACCTATCTC  
CTCGGCGCCCGCGGCACCTATCTCG  
TCGGCGCCCGCGGCACCTATCTCGT  
CGGCGCCCGCGGCACCTATCTCGTC  
GGCGCCCGCGGCACCTATCTCGTCC  
GCGCCCGCGGCACCTATCTCGTCCT  
CGCCCGCGGCACCTATCTCGTCCTC  
GCCC CGCGGCACCTATCTCGTCCTCC  
CCCGCGGCACCTATCTCGTCCTCCC  
CCGCGGCACCTATCTCGTCCTCCA  
CGCGGCACCTATCTCGTCCTCCCAG  
GCGGCACCTATCTCGTCCTCCCAGG  
CGGCACCTATCTCGTCCTCCCAGGC  
GGCACCTATCTCGTCCTCCCAGGCC  
GCACCTATCTCGTCCTCCCAGGCCT  
CACCTATCTCGTCCTCCCAGGCCTC  
ACCTATCTCGTCCTCCCAGGCCTCA  
CCTATCTCGTCCTCCCAGGCCTCAT  
CTATCTCGTCCTCCCAGGCCTCATA  
TATCTCGTCCTCCCAGGCCTCATAA  
ATCTCGTCCTCCCAGGCCTCATAAC  
TCTCGTCCTCCCAGGCCTCATAACC  
CTCGTCCTCCCAGGCCTCATAACCC  
TCGTCTCCCAGGCCTCATAACCCT  
CGTCCTCCCAGGCCTCATAACCCTG  
GTCTCCCAGGCCTCATAACCCTGC  
TCCTCCCAGGCCTCATAACCCTGCC  
CCTCCCAGGCCTCATAACCCTGCCC  
CTCCCAGGCCTCATAACCCTGCCCC  
TCCCAGGCCTCATAACCCTGCCCCT  
CCCAGGCCTCATAACCCTGCCCCCT  
CCAGGCCTCATAACCCTGCCCCCT  
CAGGCCTCATAACCCTGCCCCCTG

图 2(续)

AGGCCTCATAACCCTGCCCTCTGT  
GGCCTCATAACCCTGCCCTCTGTG  
GCCTCATAACCCTGCCCTCTGTGC  
CCTCATAACCCTGCCCTCTGTGCC  
CTCATAACCCTGCCCTCTGTGCCT  
TCATAACCCTGCCCTCTGTGCCTT  
CATAACCCTGCCCTCTGTGCCTTT  
ATAACCCTGCCCTCTGTGCCTTTC  
TAACCCTGCCCTCTGTGCCTTTCA  
AACCTGCCCTCTGTGCCTTTTCA  
ACCCTGCCCTCTGTGCCTTTTCACT  
CCCTGCCCTCTGTGCCTTTTCACTT  
CCTGCCCTCTGTGCCTTTTCACTTC  
CTGCCCTCTGTGCCTTTTCACTTCC  
TGCCCTCTGTGCCTTTTCACTTCCG  
GCCCTCTGTGCCTTTTCACTTCCGG  
CCCTCTGTGCCTTTTCACTTCCGGG  
CCTCTGTGCCTTTTCACTTCCGGGT  
CTCTGTGCCTTTTCACTTCCGGGTG  
TCTGTGCCTTTTCACTTCCGGGTGAG  
CTGTGCCTTTTCACTTCCGGGTGAGT  
TGTGCCTTTTCACTTCCGGGTGAGTG  
GTGCCTTTTCACTTCCGGGTGAGTGT  
TGCTTTTCACTTCCGGGTGAGTGTC  
GCCTTTTCACTTCCGGGTGAGTGTCT  
CCTTTTCACTTCCGGGTGAGTGTCTG  
CTTTTCACTTCCGGGTGAGTGTCTGA  
TTTCACTTCCGGGTGAGTGTCTGAG  
TTCCTTCCGGGTGAGTGTCTGAGT  
TCACTTCCGGGTGAGTGTCTGAGTG  
CACTTCCGGGTGAGTGTCTGAGTGG  
ACTTCCGGGTGAGTGTCTGAGTGGC  
CTTCCGGGTGAGTGTCTGAGTGGCT  
TTCCGGGTGAGTGTCTGAGTGGCTC  
TCCGGGTGAGTGTCTGAGTGGCTCA  
CCGGGTGAGTGTCTGAGTGGCTCAC  
CGGTGAGTGTCTGAGTGGCTCAC  
GGGTGAGTGTCTGAGTGGCTCACCT  
GGTGTGAGTGTCTGAGTGGCTCACCTG  
GTGAGTGTCTGAGTGGCTCACCTGGA  
GAGTGTCTGAGTGGCTCACCTGGAC  
AGTGTCTGAGTGGCTCACCTGGACC  
GTGTCTGAGTGGCTCACCTGGACCC  
TGTCTGAGTGGCTCACCTGGACCCC  
GTCTGAGTGGCTCACCTGGACCCCT  
TCTGAGTGGCTCACCTGGACCCCTG  
CTGAGTGGCTCACCTGGACCCCTGG  
TGAGTGGCTCACCTGGACCCCTGGG  
GAGTGGCTCACCTGGACCCCTGGGA  
AGTGGCTCACCTGGACCCCTGGGAC  
GTGGCTCACCTGGACCCCTGGGACG  
TGGCTCACCTGGACCCCTGGGACGC  
GGCTCACCTGGACCCCTGGGACGCG  
GCTCACCTGGACCCCTGGGACGCGC  
CTCACCTGGACCCCTGGGACGCGCC

图 2(续)

TCACCTGGACCCCTGGGACGCGCCG  
CACCTGGACCCCTGGGACGCGCCGA  
ACCTGGACCCCTGGGACGCGCCGAT  
CCTGGACCCCTGGGACGCGCCGATG  
CTGGACCCCTGGGACGCGCCGATGA  
TGGACCCCTGGGACGCGCCGATGAT  
GGACCCCTGGGACGCGCCGATGATG  
GACCCCTGGGACGCGCCGATGATGT  
ACCCCTGGGACGCGCCGATGATGTT  
CCCTGGGACGCGCCGATGATGTTG  
CCCTGGGACGCGCCGATGATGTTGG  
CCTGGGACGCGCCGATGATGTTGGT  
CTGGGACGCGCCGATGATGTTGGTC  
TGGGACGCGCCGATGATGTTGGTCT  
GGGACGCGCCGATGATGTTGGTCTC  
GGACGCGCCGATGATGTTGGTCTCG  
GACGCGCCGATGATGTTGGTCTCGC  
ACGCGCCGATGATGTTGGTCTCGCT  
CGCGCCGATGATGTTGGTCTCGCTC  
GCGCCGATGATGTTGGTCTCGCTCC  
CGCCGATGATGTTGGTCTCGCTCCG  
GCCGATGATGTTGGTCTCGCTCCGG  
CCGATGATGTTGGTCTCGCTCCGGC

### 外显子 3

CAAGAGTGTGGCAGGTCTCCTACAT  
AAGAGTGTGGCAGGTCTCCTACATA  
AGAGTGTGGCAGGTCTCCTACATAC  
GAGTGTGGCAGGTCTCCTACATACC  
AGTGTGGCAGGTCTCCTACATACCG  
GTGTGGCAGGTCTCCTACATACCGA  
TGTGGCAGGTCTCCTACATACCGAC  
GTGGCAGGTCTCCTACATACCGACG  
TGGCAGGTCTCCTACATACCGACGC  
GGCAGGTCTCCTACATACCGACGCT  
GCAGGTCTCCTACATACCGACGCTG  
CAGGTCTCCTACATACCGACGCTGC  
AGGTCTCCTACATACCGACGCTGCA  
GGTCTCCTACATACCGACGCTGCAC  
GTCTCCTACATACCGACGCTGCACC  
TCTCCTACATACCGACGCTGCACCC  
CTCCTACATACCGACGCTGCACCCC  
TCCTACATACCGACGCTGCACCCCA  
CCTACATACCGACGCTGCACCCAG  
CTACATACCGACGCTGCACCCAGC  
TACATACCGACGCTGCACCCAGCC  
ACATACCGACGCTGCACCCAGCCT  
CATACCGACGCTGCACCCAGCCTG  
ATACCGACGCTGCACCCAGCCTGA  
TACCGACGCTGCACCCAGCCTGAC  
ACCGACGCTGCACCCAGCCTGACC  
CCGACGCTGCACCCAGCCTGACCG  
CGACGCTGCACCCAGCCTGACCGC  
GAOGCTGCACCCAGCCTGACCGCG  
ACGCTGCACCCAGCCTGACCGCGA  
CGCTGCACCCAGCCTGACCGCGAA

图 2(续)

GCTGCACCCAGCCTGACCGCGAAG  
CTGCACCCAGCCTGACCGCGAAGG  
TGCACCCAGCCTGACCGCGAAGGA  
GCACCCAGCCTGACCGCGAAGGAG  
CACCCAGCCTGACCGCGAAGGAGG  
ACCCAGCCTGACCGCGAAGGAGGC  
CCCCAGCCTGACCGCGAAGGAGGCG  
CCAGCCTGACCGCGAAGGAGGCGCC  
CAGCCTGACCGCGAAGGAGGCGCCC  
AGCCTGACCGCGAAGGAGGCGCCCA  
GCCTGACCGCGAAGGAGGCGCCCAT  
CCTGACCGCGAAGGAGGCGCCCATG  
CTGACCGCGAAGGAGGCGCCCATGG  
TGACCGCGAAGGAGGCGCCCATGGT  
GACCGCGAAGGAGGCGCCCATGGTG  
ACCGCGAAGGAGGCGCCCATGGTGG  
CCGCGAAGGAGGCGCCCATGGTGGT  
CGCGAAGGAGGCGCCCATGGTGGTC  
GCGAAGGAGGCGCCCATGGTGGTCA  
CGAAGGAGGCGCCCATGGTGGTCAT  
GAAGGAGGCGCCCATGGTGGTCATG  
AAGGAGGCGCCCATGGTGGTCATGC  
AGGAGGCGCCCATGGTGGTCATGCG  
GGAGGCGCCCATGGTGGTCATGCGG  
GAGGCGCCCATGGTGGTCATGCGGA  
AGGCGCCCATGGTGGTCATGCGGAT  
GGCGCCCATGGTGGTCATGCGGATG  
GCGCCCATGGTGGTCATGCGGATGC  
CGCCCATGGTGGTCATGCGGATGCT  
GCCCATGGTGGTCATGCGGATGCTG  
CCCATGGTGGTCATGCGGATGCTGC  
CCATGGTGGTCATGCGGATGCTGCC  
CATGGTGGTCATGCGGATGCTGCCG  
ATGGTGGTCATGCGGATGCTGCCGT  
TGGTGGTCATGCGGATGCTGCCGTT  
GGTGGTCATGCGGATGCTGCCGTT  
GTGGTCATGCGGATGCTGCCGTTCC  
TGGTCATGCGGATGCTGCCGTTCC  
GGTCATGCGGATGCTGCCGTTCCCTA  
GTCATGCGGATGCTGCCGTTCCCTAA  
TCATGCGGATGCTGCCGTTCCCTAAT  
CATGCGGATGCTGCCGTTCCCTAATG  
ATGCGGATGCTGCCGTTCCCTAATGT  
TGCGGATGCTGCCGTTCCCTAATGTA  
GCGGATGCTGCCGTTCCCTAATGTAG  
CGGATGCTGCCGTTCCCTAATGTAGC  
GGATGCTGCCGTTCCCTAATGTAGCG  
GATGCTGCCGTTCCCTAATGTAGCGG  
ATGCTGCCGTTCCCTAATGTAGCGGG  
TGCTGCCGTTCCCTAATGTAGCGGGA  
GCTGCCGTTCCCTAATGTAGCGGGAC  
CTGCCGTTCCCTAATGTAGCGGGACT  
TGCCGTTCCCTAATGTAGCGGGACTT  
GCCGTTCCCTAATGTAGCGGGACTTT  
CCGTTCCCTAATGTAGCGGGACTTTC  
CGTTCCCTAATGTAGCGGGACTTTC

图 2(续)

GTTCTAATGTAGCGGGACTTTCTC  
TTCCTAATGTAGCGGGACTTTCTCC  
TCCTAATGTAGCGGGACTTTCTCCT  
CCTAATGTAGCGGGACTTTCTCCTG  
CTAATGTAGCGGGACTTTCTCCTGG  
TAATGTAGCGGGACTTTCTCCTGGA  
AATGTAGCGGGACTTTCTCCTGGAC  
ATGTAGCGGGACTTTCTCCTGGACG  
TGTAGCGGGACTTTCTCCTGGACGC  
GTAGCGGGACTTTCTCCTGGACGCG  
TAGCGGGACTTTCTCCTGGACGCGA  
AGCGGGACTTTCTCCTGGACGCGAG  
GCGGGACTTTCTCCTGGACGCGAGA  
CGGGACTTTCTCCTGGACGCGAGAA  
GGGACTTTCTCCTGGACGCGAGAAC  
GGACTTTCTCCTGGACGCGAGAAC  
GACTTTCTCCTGGACGCGAGAACCT  
ACTTTCTCCTGGACGCGAGAACCTG  
CTTTCTCCTGGACGCGAGAACCTGG  
TTTCTCCTGGACGCGAGAACCTGGC  
TTCTCCTGGACGCGAGAACCTGGCG  
TCTCCTGGACGCGAGAACCTGGCGC  
CTCCTGGACGCGAGAACCTGGCGCC  
TCCTGGACGCGAGAACCTGGCGCCG  
CCTGGACGCGAGAACCTGGCGCCGC  
CTGGACGCGAGAACCTGGCGCCGCC  
TGGACGCGAGAACCTGGCGCCGCCCT  
GGACGCGAGAACCTGGCGCCGCCCTG  
GACGCGAGAACCTGGCGCCGCCCTGT  
ACGCGAGAACCTGGCGCCGCCCTGTA  
CGCGAGAACCTGGCGCCGCCCTGTAC  
GCGAGAACCTGGCGCCGCCCTGTACC  
CGAGAACCTGGCGCCGCCCTGTACCG  
GAGAACCTGGCGCCGCCCTGTACCGT  
AGAACCTGGCGCCGCCCTGTACCGTC  
GAACCTGGCGCCGCCCTGTACCGTCG  
AACCTGGCGCCGCCCTGTACCGTCGA  
ACCTGGCGCCGCCCTGTACCGTCGAG  
CCTGGCGCCGCCCTGTACCGTCGAGT  
CTGGCGCCGCCCTGTACCGTCGAGTC  
TGGCGCCGCCCTGTACCGTCGAGTCT  
GGCGCCGCCCTGTACCGTCGAGTCTG  
GCGCCGCCCTGTACCGTCGAGTCTGG  
CGCCGCCCTGTACCGTCGAGTCTGGT  
GCCGCCCTGTACCGTCGAGTCTGGTGG  
CGCCTGTACCGTCGAGTCTGGTGGT  
GCCTGTACCGTCGAGTCTGGTGGTT  
CCTGTACCGTCGAGTCTGGTGGTTC  
CTGTACCGTCGAGTCTGGTGGTTCG  
TGTACCGTCGAGTCTGGTGGTTCGT  
GTACCGTCGAGTCTGGTGGTTCGTG  
TACCGTCGAGTCTGGTGGTTCGTGT  
ACCGTCGAGTCTGGTGGTTCGTGTT  
CCGTCGAGTCTGGTGGTTCGTGTTTC  
CGTCGAGTCTGGTGGTTCGTGTTCA  
GTCGAGTCTGGTGGTTCGTGTTCAC

图 2(续)

TCGAGTCTGGTGGTTCGTGTTACCC  
CGAGTCTGGTGGTTCGTGTTACCC  
GAGTCTGGTGGTTCGTGTTACCCCT  
AGTCTGGTGGTTCGTGTTACCCCTC  
GTCTGGTGGTTCGTGTTACCCCTCC  
TCTGGTGGTTCGTGTTACCCCTCCG  
CTGGTGGTTCGTGTTACCCCTCCGC  
TGGTGGTTCGTGTTACCCCTCCGCC  
GGTGGTTCGTGTTACCCCTCCGCCG  
GTGGTTCGTGTTACCCCTCCGCCGG  
TGGTTCGTGTTACCCCTCCGCCGGG  
GGTTCGTGTTACCCCTCCGCCGGGT  
GTTTCGTGTTACCCCTCCGCCGGGTA  
TTCGTGTTACCCCTCCGCCGGGTAC  
TCGTGTTACCCCTCCGCCGGGTACA  
CGTGTTACCCCTCCGCCGGGTACAC  
GTGTTACCCCTCCGCCGGGTACACC  
TGTTACCCCTCCGCCGGGTACACCG  
GTTACCCCTCCGCCGGGTACACCGC  
TTCACCCCTCCGCCGGGTACACCGCC  
TCACCCCTCCGCCGGGTACACCGCCT  
CACCCCTCCGCCGGGTACACCGCCTC  
ACCCTCCGCCGGGTACACCGCCTCG  
CCCTCCGCCGGGTACACCGCCTCGT  
CCTCCGCCGGGTACACCGCCTCGTC  
CTCCGCCGGGTACACCGCCTCGTCA  
TCCGCCGGGTACACCGCCTCGTCAA  
CCGCCGGGTACACCGCCTCGTCAAC  
CGCCGGGTACACCGCCTCGTCAACT  
GCCGGGTACACCGCCTCGTCAACTC  
CCGGGTACACCGCCTCGTCAACTCT  
CGGGTACACCGCCTCGTCAACTCTC  
GGGTACACCGCCTCGTCAACTCTCG  
GGTACACCGCCTCGTCAACTCTCGG  
GTACACCGCCTCGTCAACTCTCGGA  
TACACCGCCTCGTCAACTCTCGGAT  
ACACCGCCTCGTCAACTCTCGGATG  
CACCGCCTCGTCAACTCTCGGATGG  
ACCGCCTCGTCAACTCTCGGATGGA  
CCGCCTCGTCAACTCTCGGATGGAC  
CGCCTCGTCAACTCTCGGATGGACC  
GCCTCGTCAACTCTCGGATGGACCT  
CCTCGTCAACTCTCGGATGGACCTC  
CTCGTCAACTCTCGGATGGACCTCC  
TCGTCAACTCTCGGATGGACCTCCC  
CGTCAACTCTCGGATGGACCTCCCG  
GTCAACTCTCGGATGGACCTCCCGT  
TCAACTCTCGGATGGACCTCCCGTG  
CAACTCTCGGATGGACCTCCCGTGC  
AACTCTCGGATGGACCTCCCGTGCA  
ACTCTCGGATGGACCTCCCGTGAC  
CTCTCGGATGGACCTCCCGTGACGC  
TCTCGGATGGACCTCCCGTGACGC  
CTCGGATGGACCTCCCGTGACGC  
TCGGATGGACCTCCCGTGACGC  
CGGATGGACCTCCCGTGACGC  
GGATGGACCTCCCGTGACGC

图 2(续)

GATGGACCTCCCGTGCACGCACCTC  
ATGGACCTCCCGTGCACGCACCTCA  
TGGACCTCCCGTGCACGCACCTCAC  
GGACCTCCCGTGCACGCACCTCACC  
GACCTCCCGTGCACGCACCTCACCG  
ACCTCCCGTGCACGCACCTCACCGA  
CCTCCCGTGCACGCACCTCACCGAG  
CTCCCGTGCACGCACCTCACCGAGG  
TCCCGTGCACGCACCTCACCGAGGC  
CCCGTGCACGCACCTCACCGAGGCG  
CCGTGCACGCACCTCACCGAGGCGT  
CGTGCACGCACCTCACCGAGGCGTC  
GTGCACGCACCTCACCGAGGCGTCT  
TGCACGCACCTCACCGAGGCGTCTA  
GCACGCACCTCACCGAGGCGTCTAT  
CACGCACCTCACCGAGGCGTCTATG  
ACGCACCTCACCGAGGCGTCTATGG  
CGCACCTCACCGAGGCGTCTATGGA  
GCACCTCACCGAGGCGTCTATGGAC  
CACCTCACCGAGGCGTCTATGGACC  
ACCTCACCGAGGCGTCTATGGACCT  
CCTCACCGAGGCGTCTATGGACCTC  
CTCACCGAGGCGTCTATGGACCTCT  
TCACCGAGGCGTCTATGGACCTCTT  
CACCGAGGCGTCTATGGACCTCTTG  
ACCGAGGCGTCTATGGACCTCTTGC  
CCGAGGCGTCTATGGACCTCTTGCC  
CGAGGCGTCTATGGACCTCTTGCCC  
GAGGCGTCTATGGACCTCTTGCCCT  
AGGCGTCTATGGACCTCTTGCCCTT  
GGCGTCTATGGACCTCTTGCCCTTC  
GCGTCTATGGACCTCTTGCCCTTCC  
CGTCTATGGACCTCTTGCCCTTCCT  
GTCTATGGACCTCTTGCCCTTCCTC  
TCTATGGACCTCTTGCCCTTCCTCT  
CTATGGACCTCTTGCCCTTCCTCTG  
TATGGACCTCTTGCCCTTCCTCTGC  
ATGGACCTCTTGCCCTTCCTCTGCG  
TGGACCTCTTGCCCTTCCTCTGCGA  
GGACCTCTTGCCCTTCCTCTGCGAC  
GACCTCTTGCCCTTCCTCTGCGACG  
ACCTCTTGCCCTTCCTCTGCGACGT  
CCTCTTGCCCTTCCTCTGCGACGTC  
CTCTTGCCCTTCCTCTGCGACGTG  
TCTTGCCCTTCCTCTGCGACGTGCG  
CTTGCCCTTCCTCTGCGACGTGCGG  
TTGCCCTTCCTCTGCGACGTGCGGT  
TGCCCTTCCTCTGCGACGTGCGGTG  
GCCCTTCCTCTGCGACGTGCGGTGC  
CCCTTCCTCTGCGACGTGCGGTGCC

图 2(续)

专利名称(译)	识别核苷酸序列的方法与探针		
公开(公告)号	<a href="#">CN101360834A</a>	公开(公告)日	2009-02-04
申请号	CN200680051470.4	申请日	2006-11-21
[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)	哈普罗米克技术有限公司		
[标]发明人	马尔科姆J西蒙斯		
发明人	马尔科姆·J·西蒙斯		
IPC分类号	C12Q1/68 G01N33/53 C12N15/09 G06F17/00		
CPC分类号	C12Q1/6881 G06F19/22 C12N15/1093 G06F19/20 Y10T436/143333 C12Q2600/156 G16B25/00 G16B30/00		
优先权	60/738019 2005-11-21 US		
其他公开文献	CN101360834B		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

本发明提供识别目的核苷酸序列集的方法，所述核苷酸序列能识别相关核苷酸序列组的成员；该方法包含将组中每个成员的核苷酸序列分成多个子序列的步骤，其中至少两个子序列重叠。所述方法可用于产生能在HLA或KIR基因座分配等位基因的探针集。

<p><b>ATCGATCGATCGATCGATC</b>            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC</p> <p><b>ATCGATCGATCGATCGATC</b>            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TGGATCGAT            GGATCGATC</p> <p><b>ATCGATCGATCGATCGATC</b>            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC            GATCGATCG            ATCGATCGA            TCGATCGAT            CGATCGATC</p>	<p><b>相关序列 #1</b>            子序列 (1) 1            子序列 (1) 2            子序列 (1) 3            子序列 (1) 4            子序列 (1) 5            子序列 (1) 6            子序列 (1) 7            子序列 (1) 8            子序列 (1) 9            子序列 (1) 10            子序列 (1) 11</p> <p><b>相关序列 #2</b>            子序列 (2) 1            子序列 (2) 2            子序列 (2) 3            子序列 (2) 4            子序列 (2) 5            子序列 (2) 6            子序列 (2) 7            子序列 (2) 8            子序列 (2) 9            子序列 (2) 10            子序列 (2) 11</p> <p><b>相关序列 #3</b>            子序列 (3) 1            子序列 (3) 2            子序列 (3) 3            子序列 (3) 4            子序列 (3) 5            子序列 (3) 6            子序列 (3) 7            子序列 (3) 8            子序列 (3) 9            子序列 (3) 10            子序列 (3) 11</p>
--	--