



(12) 发明专利

(10) 授权公告号 CN 101896623 B

(45) 授权公告日 2013. 12. 18

(21) 申请号 200880119985. 2

(74) 专利代理机构 隆天国际知识产权代理有限公司 72003

(22) 申请日 2008. 10. 21

代理人 吴小瑛

(30) 优先权数据

0720675. 8 2007. 10. 22 GB

0815113. 6 2008. 08. 19 GB

(51) Int. Cl.

C12Q 1/68(2006. 01)

(85) PCT申请进入国家阶段日

审查员 黄琦

2010. 06. 09

(86) PCT申请的申请数据

PCT/GB2008/003555 2008. 10. 21

(87) PCT申请的公布数据

W02009/053679 EN 2009. 04. 30

(73) 专利权人 LGC 有限公司

地址 英国米德尔塞克斯郡

(72) 发明人 尼特亚·加雷 保罗·德贝纳姆

戴维·约翰·弗伦奇

丽贝卡·霍华德

戴维·戈登·麦克道尔 汤姆·布朗

权利要求书4页 说明书40页

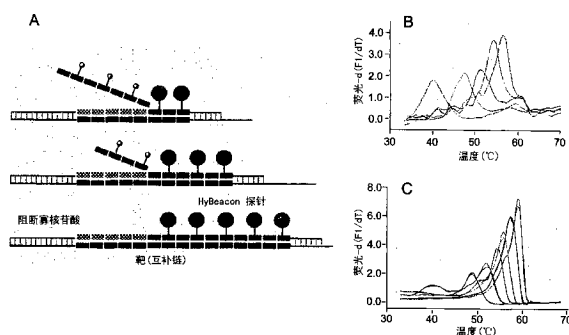
序列表34页 附图13页

(54) 发明名称

寡核苷酸及其应用

(57) 摘要

本发明涉及用于确定靶多核苷酸内串联重复的数目的方法,所述方法包括:(a) 提供包含所述靶多核苷酸的样品,其中所述靶多核苷酸内的一个或多个串联重复是单链形式,(b) 使经标记的探针寡核苷酸与靶多核苷酸的单链部分杂交,其中所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少一个串联重复互补,并且所述探针寡核苷酸中的至少5个核苷酸与所述靶多核苷酸的单链部分中的串联重复互补,和(c) 基于所述探针寡核苷酸与所述靶多核苷酸的单链部分的杂交,确定所述靶多核苷酸内串联重复的数目。



1. 用于确定靶多核苷酸内串联重复的数目的方法,其中已知所述靶多核苷酸在给定的多态性基因座具有不同数目的串联重复的多个等位基因,

所述方法包括:

(a) 提供包含所述靶多核苷酸的样品,其中所述靶多核苷酸内的两个或更多个串联重复是单链形式,

(a1) 使阻断寡核苷酸与步骤(a)中提供的所述靶多核苷酸中的至少一个但并非全部串联重复杂交,从而使所述靶多核苷酸中的一个或多个串联重复在与阻断寡核苷酸杂交后保持单链形式,

(b) 使经标记的探针寡核苷酸与所述靶多核苷酸的单链部分杂交,其中所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少一个串联重复互补,并且所述探针寡核苷酸中的至少5个核苷酸与所述靶多核苷酸的单链部分中的串联重复互补,其中所述探针寡核苷酸允许根据其解链温度(T_m)区分所述靶多核苷酸的单链部分中串联重复的数目,并且,其中所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少一个串联重复完全或部分互补,和

(c) 基于所述探针寡核苷酸与所述靶多核苷酸的单链部分的杂交及其解链温度,确定所述靶多核苷酸内串联重复的数目,

其中,所述方法用于非诊断和治疗目的。

2. 如权利要求1所述的方法,其中,所述方法用于法医科学领域、亲子和亲缘关系测定、连锁制图、微生物分型或食物链内的示踪。

3. 如权利要求1所述的方法,其中,所述靶多核苷酸是DNA。

4. 如权利要求1所述的方法,其中,在步骤(c)中使用解链峰分析或高分辨率解链曲线分析进行所述确定。

5. 如权利要求1-4中任一项所述的方法,其中,在步骤(a)中,所述靶多核苷酸内的三个或更多个串联重复为单链形式。

6. 如权利要求1-5中任一项所述的方法,其中,在步骤(b)中,所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少两个串联重复互补。

7. 如权利要求6所述的方法,其中,在步骤(b)中,所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少三个串联重复互补。

8. 如前述权利要求中任一项所述的方法,其中,在步骤(a)中,所述包含串联重复的靶多核苷酸的单链部分包含至少8个核苷酸。

9. 如前述权利要求中任一项所述的方法,其中,在步骤(a)中,所述包含串联重复的靶多核苷酸的单链部分包含至少16个核苷酸。

10. 如权利要求1所述的方法,其中,在步骤(a1)中,所述阻断寡核苷酸与所述靶多核苷酸的单链部分中的至少两个串联重复杂交。

11. 如权利要求10所述的方法,其中,在步骤(a1)中,所述阻断寡核苷酸与所述靶多核苷酸的单链部分中的至少三个串联重复杂交。

12. 如权利要求1所述的方法,其中,在步骤(a1)中,两个或更多个串联重复在与所述阻断寡核苷酸杂交后保持单链形式。

13. 如权利要求12所述的方法,其中,在步骤(a1)中,三个或更多个串联重复在与所述

阻断寡核苷酸杂交后保持单链形式。

14. 如权利要求 1 所述的方法,其中,在步骤 (a1) 中,将两个或更多个阻断寡核苷酸与步骤 (a) 中提供的至少一个但并非全部串联重复杂交,从而使所述靶多核苷酸中的一个或多个串联重复在与阻断寡核苷酸杂交后保持单链形式。

15. 如权利要求 14 所述的方法,其中一个或多个阻断寡核苷酸与部分数目的串联重复杂交。

16. 如前述权利要求中任一项所述的方法,其中,在步骤 (a) 中的样品在扩增反应期间或之后产生。

17. 如权利要求 16 所述的方法,其中所述扩增反应是 PCR。

18. 如前述权利要求中任一项所述的方法,其中所述探针寡核苷酸是荧光标记的。

19. 如权利要求 1 所述的方法,其中所述阻断寡核苷酸与所述靶多核苷酸中的至少一个串联重复完全地互补。

20. 如权利要求 1 所述的方法,其中所述阻断寡核苷酸与所述靶多核苷酸中的至少一个串联重复部分地互补。

21. 如前述权利要求中任一项所述的方法,其中,在步骤 (a) 中,在所述靶多核苷酸中侧翼于串联重复的一个区域或两个区域都为单链形式。

22. 如权利要求 21 所述的方法,其中,在步骤 (b) 中,所述探针寡核苷酸包含锚定部分,该锚定部分与所述靶多核苷酸中侧翼于串联重复的区域互补。

23. 如权利要求 1 所述的方法,其中,在步骤 (a) 中,在所述靶多核苷酸中侧翼于串联重复的一个或两个区域都为单链形式,并且所述阻断寡核苷酸包含锚定部分,该锚定部分与侧翼于所述串联重复的区域互补。

24. 如权利要求 23 所述的方法,其中,在步骤 (b) 中,所述探针寡核苷酸包含锚定部分,该锚定部分与侧翼于所述串联重复的区域互补,但该侧翼于所述串联重复的区域不与所述阻断寡核苷酸的锚定部分互补。

25. 如权利要求 1 所述的方法,其中所述阻断寡核苷酸在其 3' 或 5' 端包含钳部分,该钳部分与在所述探针寡核苷酸的 5' 或 3' 端的钳部分互补。

26. 如权利要求 25 所述的方法,其中使用至少两对不同的阻断寡核苷酸和探针寡核苷酸,其中每对阻断寡核苷酸和探针寡核苷酸各自的钳部分彼此互补。

27. 如权利要求 26 所述的方法,其中选择所述钳部分的核苷酸序列,以使每对互补的钳部分具有不同的 T_m 。

28. 如权利要求 1 所述的方法,其中所述步骤 (a) 和 (a1) 同时进行。

29. 如权利要求 28 所述的方法,其中 PCR 产生靶多核苷酸,其中,所述靶多核苷酸是 DNA,所述靶 DNA 内的两个或更多个串联重复为单链形式,并且在所述 PCR 中使用的引物也包含所述阻断寡核苷酸。

30. 如权利要求 29 所述的方法,其中所述引物在其 3' 端包含与靶 DNA 内串联重复的 3' 端区域互补的部分,并且在其 5' 端包含与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分。

31. 如权利要求 30 所述的方法,其中所述引物从 3' 向 5' 方向包含:(i) 与所述靶 DNA 内串联重复的 3' 端区域互补的部分,(ii) 任选的间隔子部分,(iii) 与通过所述引物合成

的 PCR 产物的链中的侧翼部分互补的锚定部分,和 (iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分。

32. 如权利要求 31 所述的方法,其中所述引物进一步包含 (v) 钳部分,该钳部分与探针寡核苷酸的 3' 端的钳部分互补。

33. 如权利要求 31 所述的方法,其中使用至少两对不同的 PCR 引物和探针寡核苷酸,其中每对阻断寡核苷酸和探针寡核苷酸各自的钳部分彼此互补。

34. 如权利要求 29 所述的方法,其中所述 PCR 中使用的引物还包含探针寡核苷酸。

35. 如权利要求 31 所述的方法,其中所述引物进一步包含 (v) 间隔子部分和 (vi) 探针寡核苷酸。

36. 如权利要求 1 所述的方法,其中由所述探针寡核苷酸和包含所述串联重复的靶 DNA 的单链部分之间形成的杂交体的 T_m 之间的解链温度差异 (ΔT_m) 为至少 0.5°C 。

37. 如前述权利要求中任一项所述的方法,其中所述杂交步骤 (b) 在 30°C 至 75°C 的温度范围内进行。

38. 用于确定靶多核苷酸内串联重复的数目的系统,其中已知所述靶多核苷酸在给定的多态基因座具有不同数目的串联重复的多个等位基因,并且,其中所述靶多核苷酸内的一个或多个串联重复为单链形式,所述系统包括:

(a) 经标记的探针寡核苷酸,其与所述靶多核苷酸的单链部分中的至少一个串联重复互补,且所述探针寡核苷酸的至少 5 个核苷酸与所述靶多核苷酸的单链部分中的串联重复互补,其中所述探针寡核苷酸允许根据其解链温度 (T_m) 区分所述靶多核苷酸的单链部分中串联重复的数目,并且,其中所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少一个串联重复完全或部分互补,和

(b) 阻断寡核苷酸,其与所述靶多核苷酸中至少一个但并非全部串联重复互补。

39. 寡核苷酸引物,其用于参与 PCR 反应以扩增包含串联重复的靶 DNA,所述引物从 3' 向 5' 方向包含:(i) 与靶 DNA 内串联重复的 3' 端区域互补的部分,(ii) 任选的间隔子部分,(iii) 与通过所述引物合成的 PCR 产物的链中的侧翼部分互补的锚定部分,和 (iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分。

40. 如权利要求 39 所述的寡核苷酸引物,其进一步包含 (v) 间隔子部分和 (vi) 探针寡核苷酸。

41. 用于确定靶多核苷酸内串联重复的数目的系统,所述系统包括:标记的寡核苷酸,其包含含有与锚定部分结合的至少两个串联重复的部分,其中所述锚定部分的序列和所述至少两个串联重复在靶多核苷酸中连续地存在,和权利要求 39 所述的寡核苷酸引物。

42. 用于确定靶多核苷酸内串联重复的数目的系统,所述系统包括:标记的寡核苷酸,其包含含有与锚定部分结合的至少两个串联重复的部分,其中所述锚定部分的序列和所述至少两个串联重复在靶多核苷酸中连续地存在,其进一步在其 3' 端包含钳部分,和权利要求 39 所述的寡核苷酸引物,其进一步包含 (v) 钳部分,其中所述钳部分是彼此互补的。

43. 用于确定靶多核苷酸内串联重复的数目的系统,其包含两对或更多对权利要求 42 所述的标记的寡核苷酸和寡核苷酸引物,其中所述寡核苷酸对具有互补的钳部分。

44. 权利要求 39 所述的用于参与 PCR 反应以扩增包含串联重复的靶 DNA 的寡核苷酸引物的制备方法,所述方法包括:

- (a) 选择包含串联重复的靶 DNA,
- (b) 获得串联重复的序列以及一个或多个侧翼区域的序列, 和
- (c) 合成寡核苷酸, 该寡核苷酸从 3' 向 5' 方向包含 : (i) 与靶 DNA 内串联重复的 3' 端区域互补的部分, (ii) 任选的间隔子部分, (iii) 与通过所述引物合成的 PCR 产物的链中的侧翼区域互补的锚定部分, (iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分, 和任选的 (v) 钳部分, 或任选的 (v) 间隔子部分, 和 (vi) 探针寡核苷酸。

45. 权利要求 41 所述的系统的制备方法, 所述方法包括 :

- (a) 选择包含串联重复的靶 DNA,
- (b) 获得串联重复的序列以及一个或多个侧翼区域的序列, 和
- (c) 合成权利要求 41 中所述的标记的寡核苷酸, 该寡核苷酸包含含有与锚定部分结合的至少两个串联重复的部分和任选的在其 3' 端的钳部分, 其中所述锚定部分的序列和所述至少两个串联重复在靶多核苷酸中连续地存在, 其中标记所述寡核苷酸, 并且通过权利要求 44 所述的方法制备权利要求 39 或 40 所述的寡核苷酸引物。

46. 如权利要求 1-37、44 或 45 中任一项所述的方法, 其中所述靶多核苷酸或 DNA 是人基因组 DNA。

47. 如权利要求 38、41 或 43 中任一项所述的系统, 其中所述靶多核苷酸或 DNA 是人基因组 DNA。

48. 如权利要求 39 或 40 所述的寡核苷酸引物, 其中所述靶多核苷酸或 DNA 是人基因组 DNA。

寡核苷酸及其应用

[0001] 本发明涉及寡核苷酸,特别涉及寡核苷酸在检测 DNA 内串联重复中的应用。

[0002] 已知多种用于检测特异 DNA 序列并评价已知单核苷酸多态性 (SNP) 的技术和探针系统,包括均相聚合酶链式反应 (PCR)、TaqMan™ 探针、Eclipse 探针、分子信标、Scorpion 引物、简便杂交探针、ResonSense 探针、GenePin 探针和杂交信标(HyBeacons®)。这些内容例如在本申请人较早期的专利申请 WO 01/73118 和 WO 2007/010268 中被讨论。

[0003] 尽管许多 DNA 多态性和突变是 SNP,但很多是由于 DNA 序列的串联重复造成的。目前,通常通过如在 PCR 扩增靶 DNA 序列后的大小分析或 DNA 测序来分析基因组 DNA 中的此类重复。尽管曾尝试使用寡核苷酸探针系统来鉴定 DNA 中的串联重复(例如,参见 Radtkey et al (2000) Nucl. Acids Res. 28, e17(i-vi)),但迄今为止尚没有使用杂交探针分析这些重复序列的实践方法。

[0004] Alec Jeffreys 根据其有关人类基因组内重复 DNA 序列的发现 (Jeffreys 1985a, 1985b),发明了 DNA 指纹或 DNA 图谱分析 (DNA profiling)。发现所谓 STR(短串联重复)或 VNTR(可变数目串联重复)的重复序列在不同个体内的长度不同。在首次描述的 DNA 图谱分析中,通过使用限制性酶(在特异的核苷酸序列处切割 DNA 的酶)分型了很多此类基因座,其中该限制性酶基于酶识别序列之间重复序列的重复数目释放长度不同的片段。在那些早期的 DNA 图谱分析中,在琼脂糖凝胶上分离片段,并且通过使用放射性 DNA 探针鉴定包含目标 STR 的特异片段。这是复杂且耗时的方法,而自从首次描述 DNA 图谱分析和特性确定(identity determination)后,大量的改进和研发已经对此方法产生了极大的影响,提高了分析的容易性,并由此提高了可处理样品的数量,以及辨识能力和成本。

[0005] 在该方法的最新形式中,使用聚合酶链式反应 (PCR) 扩增 STR 基因座,随后通过如毛细管电泳 (CE) 的技术将其分离。在英国,每年处理包括从常规口腔拭样到严重犯罪现场样品在内的数以万计的样品,用于与目前拥有超过 200 万此类图谱的国家 DNA 数据库进行比较。

[0006] 尽管在紧急情况下,实验室具有在短至单个工作日内由例如口腔拭样的简单样品生成图谱的能力,但是难以想象如何能对于所有的样品在常规基础上实现这一点。此外,如果考虑从样品收集至到达实验室的时间,显然无法及时测定大部分图谱。结果,常常在 DNA 图谱被获得并分析前在没有其他证据的情况下而不得不释放被捕的嫌疑人。这样做的结果可能是,某个因相对较小的犯罪而被逮捕并随后又被释放的个体在此后可能被确定为严重犯罪的主犯,对于此类主犯,如果真的能确定其位置,则必须将其重新逮捕。这个过程不仅耗时多且成本高,而且还存在该个体实施其他潜在的严重犯罪的风险,而如果能够在该个体仍被拘留期间获得所述图谱分析,则完全可能防范这种风险。

[0007] 目前基于 PCR 的方法包括 DNA 提取、STR 扩增、基于 CE 的大小分离和分析,此方法的各个方面以及所用设备的成本和复杂性意味着 STR 图谱分析主要局限于专业实验室,其结果又制约了最终完成的周转时间。然而,如果要使图谱分析变得足够快速以至于能够在个体仍被拘留期间通过常规方式获得结果,就必须找到能够进行原位分析的方法。

[0008] 在其他非法 DNA 分析中,已有使用均相 PCR 的适合方法,其中通过荧光的简单

变化在同一试管中进行 DNA 的扩增和分析。现已将多种此类基于荧光的探针系统,包括 HyBeacon 和 TaqMan 探针(参见 WO 01/73118 和 WO 2007/010268 的综述)用于测定 DNA 序列中的变异,例如单核苷酸多态性(SNP)插入和删除。然而,此类探针尚未用于测定对于如 STR 或其他 DNA 序列分析所要求的较长的长度多态性,实际上也不知道使用均相探针系统如何能够测定这些长度多态性。

[0009] 已有多个小组描述了均相探针系统,其能够使在单个管中进行 PCR 产物分析,该分析使用例如扩增首先被检测的循环以提供一条 DNA 序列在另一条中的相对定量(例如用于食品分析中各类动物样品的混合物),或者提供一个等位基因变体在另一个中的相对定量(例如用于关于单核苷酸多态性或 SNP 的基因分型以及载体状态)。本发明的发明人此前描述了新型的荧光探针系统 HyBeacons,其能够鉴定微小的序列差异,例如 SNP,或者小插入或缺失,因为这些差异显著地影响解链温度,HyBeacons 结构的新性质使其能够完成这一点。此类方法可通过 SNP 分型以均相且可能便携的方式鉴别个体。然而,现有的数据库通常基于 STR,这是因为与在每个基因座通常仅具有 2 个等位基因的 SNP 相比,在每个基因座平均具有 10 个等位基因的 STR 的可变性明显大很多。因此,英国法医 DNA 图谱分析的工业标准为来自 AppliedBiosystems 的包含 10 个 STR 的 SGM+ 试剂盒,而性别测试优先可选的 SNP 分型组(alternative SNP typing panel),SNP 分型组需要约 50–100 个 SNP 的区域以获得个体鉴定的可比较水平。显然,尽管 SNP 组可用于比较个体样品,但是目前只有基于 STR 的数据库,个体样品可与之进行比较。

[0010] 在使 STR 分型能够在专业实验室之外的背景中应用的尝试中,曾试图通过使用微型化 CE 而将标准图谱分析方法的各个步骤组合成均相的形式。此类微型化系统的主要优点是由较短的毛细管获得更快速的分析。通常将此类毛细管蚀刻在显微镜载玻片中,并且通常深 10–50 μm ,宽 50 μm ,长度在几个厘米的范围,并且由玻璃盖玻片密封(Wooley and Mathies,1994)。还描述过其他可选形式,例如使用注入成型的塑料(McCormick et al,1997)。

[0011] 与更为常见的柱长 36cm 的 ABI 系统相反,用于 Promega PowerPlex™1.1 STR 试剂盒的长数个厘米的微型化 CE 柱能够将分离时间从约 45 分钟减少至仅 2.5 分钟(Schmalzing et al,1999)。还曾描述过可在仅 2 分钟内同时分析多达 96 个样品的毛细管阵列系统(Shi et al,1999),以及最近开发的多色系统,不过其分离时间相对较长(Goedecke et al,2004)。

[0012] 还需要在分离前扩增 DNA 的便携式 PCR 系统,现已证明了多种此类系统,范围涉及从已证实用于 4 个基因座的由电池供电的系统(Belgrader et al,1998),到集成微芯片 CE 装置的系统(Lagally et al,2001)。今后面临的挑战之一将是使用具有快速扩增技术的多 STR 试剂盒。多重 PCR 需要仔细的优化,并且必须在不同引物对的反应优化之间进行折衷和平衡。快速 PCR 的反应条件通常更为严谨,因此,其对多重 PCR 系统特别需要的任何折衷的容忍性都较小。

[0013] 在进一步简化 STR 分析的尝试中,Nanogen Inc. (San Diego,CA) 试图使用杂交的形式,其中将 PCR 扩增的靶与固定在微型硅芯片不同位置的长度不同的捕获探针杂交。随后使用带有末端标记的长度恒定的探针检测靶序列的其余部分,所述探针主要靶向非重复序列。在需要由电场严格控制的严谨条件以及用于控制重复单元在杂交期间滑移的温度

下,只有能够使探针和捕获序列在彼此靠近时立即杂交的靶才能允许末端碱基参与碱基堆叠并稳定此结构。此类复合物是足够稳定的,以允许通过微芯片上荧光的位置确定给定基因座的不同等位基因 (Radtkey et al,2000,Westin et al,2000),不过从较短的捕获探针序列也产生信号。此类探针由单个荧光团进行末端标记,并且其荧光不随其杂交状态而改变。因此,为了检测给定位点的给定序列,必须依赖探针的捕获并且洗去任何未杂交的探针。

[0014] 长度多态性,特别是重复长度多态性的鉴定为法医提供了大量情报。然而,此类多态性还有很多其他的应用。在整个人类基因组中有 15,000 多个 STR 标记物,并且基于该信息 (informity),可用于在绘图 (mapping) 研究中排除直到数个厘摩的基因组 (Weissenbach et al,1992)。举例而言,STR 绘图研究已经用于鉴定与肥厚型心肌病相关的新基因座 (Watkins et al,1993,Carrier et al,1993,Thierfelder et al,1993)。

[0015] 已知其他长度多态性,特别是三联体重复扩增导致疾病,特别是神经变性疾病和其他类型的疾病,包括强直性肌营养不良、脆性 X 综合征、亨廷顿病,多种脊髓小脑型共济失调症和弗里德里希共济失调症 (Friedreich ataxia) (Sinden 1999)。还有其他长度多态性与疾病易感性相关。例如,在胰岛素基因上游 600 碱基由 14bp 重复单元构成的微小卫星影响个体患糖尿病的风险 (Bell et al,1982),而微卫星不稳定性杂合性的损失也是包括肺癌在内的很多癌症的特征 (Ionov et al,1993)。现已将微卫星不稳定性与高突变率及 DNA 修复过程相关连 (Loeb, L. A. 1994, Frayling, I. M. 1999)。

[0016] 在很多其他物种中发现长度多态性,并可有利地用于分型的目的。例如,巴西副球孢子菌 (*Paracoccidioides brasiliensis*) 是拉丁美洲地方性疾病,副球孢子菌病的病原因子,在拉丁美洲估计有 1000 万人受此病影响 (Restrepo-Moreno,2003)。为进一步了解该具有临床重要性的生物体而进行的个体分离物和系统分类物种的研究直到最近仍受到缺少用于分型目的的分子标记物而被阻碍,但近期 Matute 基于其他人的成功对其进行了整理 (Matute et al,2006)。微卫星标记物系统为大量其他生物体的 DNA 图谱分析提供了高度有效的方法,并已经成功地用于真菌,例如酿酒酵母 (*Saccharomyces cerevisiae*) (Hennequin 2001)、烟曲霉 (*Aspergillus fumigatus*) (Bart-Delabesse et al,2001) 和假丝酵母 (*Candida* spp.) (Foulet et al,2005) 的分型。

[0017] 现已公开用于植物物种的类似分型方法。现已从测序的水稻基因组鉴定出多于 2,000 个简单序列重复标记物 (McCouch et al,2002),还在小麦基因组 (Roder et al,1998) 和水稻基因组 (Brondani et al,1998,2003) 等中鉴定出其他的标记物。

[0018] 不应将本说明书列出或讨论的明显先公开文件视为承认所述文件是现有技术的一部分或承认所述文件是公知常识。

[0019] 本发明提供了寡核苷酸,以及使用寡核苷酸的方法,其可用于检测并区分多核苷酸序列中不同数目的串联重复。因此,本发明可应用于医学诊断学和法医科学领域,并且可以应用于亲子和亲缘关系测定,连锁制图、微生物分型,以及食物链内的示踪能力等。

[0020] 本发明的第一方面提供了用于确定靶多核苷酸内串联重复的数目的方法,所述方法包括:

[0021] (a) 提供包含所述靶多核苷酸的样品,其中所述靶多核苷酸内的一个或多个串联重复是单链形式,

[0022] (b) 使经标记的探针寡核苷酸与靶多核苷酸的单链部分杂交,其中所述探针寡核苷酸与所述靶多核苷酸的单链部分中的至少一个串联重复互补,并且所述探针寡核苷酸中的至少 5 个核苷酸与所述靶多核苷酸的单链部分中的串联重复互补,和

[0023] (c) 基于所述探针寡核苷酸与所述靶多核苷酸的单链部分的杂交,确定所述靶多核苷酸内串联重复的数目。

[0024] 此方法也可被视为确定靶多核苷酸内串联重复的数目的测定法。

[0025] 所述靶多核苷酸可以是 DNA 或可以是 RNA。通常为 DNA。

[0026] 众所周知,很多天然存在的多核苷酸,特别是 DNA 分子包含串联重复,例如 (ABCD...)_n 形式的重复,其中 A、B、C 和 D 为核苷酸,n 是此核苷酸序列重复的次数。重复可能比该形式更复杂,并且特定的重复可能散布在其他特定重复中。通常,每个串联重复包含 2、3、4、5 或 6 个核苷酸,并可重复 2-50 次或更多,即 n 可以是 2-50 或更大。部分重复可能出现在重复的靶序列中。此外,STR 等位基因可包含多于一种类型的重复,并可包含位于重复元件之间的非重复序列,例如,对于 D251338 和 FGA STR 基因座分别为 (TGCC)_m(TTCC)_n 和 (TTTC)₃ TTTTTTCT(CTTT)_nCTCC(TTCC)₂(SEQ ID No. 1 和 SEQ ID No. 2)。下表 21 给出了常见的串联重复序列。

[0027] 常见的串联重复的实例为 (GATA)_n、(CAG)_n 和 (TCTTA)_n。

[0028] 在某些串联重复中,重复序列可具有多态性,致使例如一个或多个重复序列与核心重复序列略有不同。例如,STR TH01 具有特异的核心重复序列 AATG。然而,在被称作 9.3 的一个等位基因中,第七个重复缺少 A 碱基。如实施例 12 中的详细讨论,某些 STR 包含可变的 TCTA 和 TCTG 重复。

[0029] 在此方法的步骤 (a) 中,提供包含所述靶多核苷酸的样品,其中所述靶多核苷酸内的一个或多个串联重复是单链形式。如下文详述,这允许探针寡核苷酸与靶的单链区杂交。靶多核苷酸可以是合成的 DNA 分子的一部分,或者是可以是天然 DNA 分子的一部分。当靶多核苷酸为 DNA 时,通常通过扩增另一个 DNA 分子例如天然 DNA 分子的一部分而产生。在此实施方案中,通过扩增方法本身或者扩增产物的进一步加工产生其中一个或多个串联重复是单链形式的靶 DNA。例如,可使用 PCR 反应,其中采用与天然 DNA(例如人基因组 DNA)内侧翼于串联重复的区域杂交的引物。如本领域所熟知的,可通过例如其中一种引物相对于另一条引物摩尔过量的不对称 PCR 而将 PCR 产物制成单链,或者可通过例如解链将双链 PCR 产物制成单链。

[0030] 应理解,给定的串联重复可出现在天然 DNA 的多于一个的区域。因此,通常可期望从天然 DNA(例如基因组 DNA)扩增特定的目标靶多核苷酸(如 DNA),例如使用与独特区域杂交的 PCR 引物,所述独特区域侧翼于在天然 DNA 内存在的目标串联重复区域。

[0031] 当所述靶多核苷酸是 RNA 分子时,通常通过转录由适合的 DNA 模板产生。例如,靶 RNA 可通过 DNA 分子的体外转录产生,其中所述 DNA 分子包含位于待分析基因座上游的启动子。适合用于产生靶 RNA 的 RNA 聚合酶包括 SP6 和 T7 RNA 聚合酶。因此,例如,可使用 PCR 从天然 DNA 扩增 DNA,其中引物之一包含 RNA 聚合酶的识别位点。在存在 RNA 聚合酶和适合的核苷酸条件下,可产生单链 RNA。

[0032] 在优选的实施方式中,所述方法包括在步骤 (a) 后或同时进行的另外的步骤:

[0033] (a1) 使阻断寡核苷酸与步骤 (a) 中提供的靶多核苷酸中的至少一个但并非全部

串联重复杂交,从而使所述靶多核苷酸中的一个或多个串联重复在与阻断寡核苷酸杂交后保持单链的形式。

[0034] 在一个实施方式中,在步骤(a1)中,将两个或更多个阻断寡核苷酸与步骤(a)中提供的靶多核苷酸中的至少一个但并非全部串联重复杂交,从而使所述靶多核苷酸中的一个或多个串联重复在与阻断寡核苷酸杂交后保持单链的形式。

[0035] 应理解,通过使用一个或多个阻断寡核苷酸,可限制样品中靶寡核苷酸内存在的单链形式串联重复的数目。例如,如果靶多核苷酸(在没有阻断寡核苷酸时)在靶多核苷酸内包含12个单链形式的串联重复,而阻断寡核苷酸能够与其中3个杂交(因为其包含与3个串联重复互补的多核苷酸部分),那么在存在阻断寡核苷酸的条件下,在样品中靶多核苷酸存在的单链形式重复的数目是9。

[0036] 通常,一个或多个阻断寡核苷酸与靶DNA单链部分中的至少两个串联重复杂交,例如3或4或5或6或7或8或9或10或20或30或40个。通过可存在于待分析的具体基因座的串联重复的可能数目确定阻断寡核苷酸内串联重复的数目。通常,阻断寡核苷酸与存在于靶DNA单链部分的串联重复中的至少8个核苷酸(或至少12个,或至少16个,或至少20个核苷酸)互补。利用现有技术能够合成长达250个核苷酸单元的寡核苷酸,而预期在未来能够合成更长的寡核苷酸。

[0037] 优选阻断寡核苷酸(沿其长度与靶DNA中的单链DNA互补时)的长度为12-150个核苷酸,例如12-120,12-100或12-90个核苷酸。当阻断寡核苷酸具有如下文所述的其他功能时,寡核苷酸可更长,例如20-180个核苷酸,通常其长度为30-150个核苷酸,例如30-120、30-100或30-80个核苷酸的长度。

[0038] 所述一个或多个阻断寡核苷酸无需是串联重复中碱基的整数倍。实际上,它们可包括部分重复,这样阻断寡核苷酸以“偏置(off-set)”的方式与串联重复结合。

[0039] 在使用两个或更多个阻断寡核苷酸时,特别优选使用“偏置”型阻断寡核苷酸,而且此类“偏置”型阻断寡核苷酸也可用于例如以下情况中。

[0040] 在存在7个重复的阻断剂的条件分析包含类型8、10个重复的样品时,表现出1和3个重复单元的组合供探针与之杂交。然而,这似乎与存在10个重复的阻断剂的条件分析包含11、13个重复的样品时获得的结果类似。因此,当两种阻断剂同时存在时,将难以确定样品中存在哪种重复组合。

[0041] 在上述实例中,如果使用10.2个重复的阻断剂(其中“.2”代表下一个重复的最开始的两个碱基)代替10个重复的阻断剂,那么包含11、13个重复的样品将仅具有供杂交的0.2和2.2个重复序列,这将导致解链峰移动。因此,偏置型阻断方法在解链峰可能重叠的情况下特别有用。例如,偏置型阻断方法可用于将多至4个阻断剂与4个碱基重复序列组合(通过使用由添加1个碱基而具有长度偏移的大小增加的阻断剂),或用于5个碱基重复的5个阻断剂,以此类推。

[0042] 这样,阻断寡核苷酸可被“重叠”以提供“完全在同一试管中进行”的测定法,其可用于区分相同样品内不同数目的重复。通常选择具有部分重复的可重叠的偏置型阻断寡核苷酸,这样在(串联重复数目中的)各个不同长度变体之间具有至少0.5°C的解链峰 T_m 差异。通常,通过使用不同的偏置型阻断寡核苷酸,可以在单个PCR中分析可能包含串联重复数目不同的多个等位基因的单个可变基因座。此实施方案在实施例12中更详细地被描述。

[0043] 为了避免争议,可通过使用能够与 y 串联重复杂交的一个或多个阻断寡核苷酸,将包含单链形式的 x 串联重复的本发明的靶 DNA 转化成包含单链形式的 (x-y) 串联重复的本发明的靶 DNA。

[0044] 通常,在靶多核苷酸中有单链形式的 2 或 3 或 4 或 5 或 6 或 7 或 8 或 9 或 10 或 11 或 12 或 13 或 14 或 15 个串联重复。优选其在 2-10 个之间,更优选 4-8 个之间。靶多核苷酸中单链形式的这些串联重复的通常数目可在阻断寡核苷酸存在或不存在下获得。包含串联重复的靶多核苷酸的单链部分通常包含至少 8 个,优选至少 12 个,更优选至少 16 或至少 20 个核苷酸。

[0045] 众所周知,多态性的常见类型涉及在具体基因座的串联重复的数目,其可发生显著的变化。应理解,本发明的方法是用于测定串联重复的数目,并且优选在任何给定的测定中,小范围的串联重复是单链形式。因此,为了确定可能存在于具体基因座的串联重复的数目,有必要进行不使用阻断寡核苷酸的方法,并且进行分别使用一个或多个阻断寡核苷酸的方法,其中所述阻断寡核苷酸能够与不同数目的串联重复杂交。这样,对于已知具有串联重复数目不同的多个等位基因的任何给定基因座,可产生具有类似大小的单链部分的靶多核苷酸(例如,单链形式的重复的数目的可变性降低)。应理解,可以使用多于一个的阻断寡核苷酸,只要在与所述阻断寡核苷酸杂交后,在靶多核苷酸中保留单链形式的一个或多个串联重复。

[0046] 在串联重复序列具有多态性(例如 STR TH01 或实施例 12 中描述的 STR)的一个实施方式中,可期望阻断寡核苷酸阻断具有可变性的区域,从而使单链形式的靶多核苷酸仅包含相同的重复。例如,如实施例 12 所述,用于 STRD8S1179 的阻断寡核苷酸仅留下可与探针寡核苷酸杂交的 TCTA 重复。

[0047] 优选地,除非按照上文使用偏置型阻断剂,优选通过阻断剂生成在与阻断寡核苷酸杂交后包含保持单链形式的一个或多个串联重复的不同靶多核苷酸,从而避免不同的等位基因具有相同的 T_m 。

[0048] 如下文详述,所述阻断寡核苷酸也可以是用于生产靶多核苷酸的寡核苷酸,其中所述靶多核苷酸中的一个或多个串联重复是单链形式,例如作为 PCR 反应中使用的引物部分。

[0049] 步骤 (b) 中的经标记的探针寡核苷酸可以是任何适合的探针寡核苷酸。所述探针寡核苷酸通常与至少 2 个或 3 个或 4 个或 5 个或 6 个或 7 个或 8 个串联重复互补,这可基于待分析的基因座而确定。优选探针寡核苷酸与 4-10 个串联重复互补,不过对于双核苷酸重复,可优选探针寡核苷酸与至少 5 个或 6 个串联重复互补。通常,探针寡核苷酸中 5-40 个核苷酸部分与靶多核苷酸中单链形式的串联重复互补。探针寡核苷酸通常与靶多核苷酸内以单链部分存在的串联重复中的至少 8 个核苷酸互补;更优选其与存在的串联重复中的至少 10 个、12 个、15 个或 20 个核苷酸互补。

[0050] 如下文详述,优选寡核苷酸探针能够与单链靶多核苷酸结合,且其 T_m 在 40°C -70°C 的范围。

[0051] 在优选的实施方式中,所述寡核苷酸探针是荧光标记的。其可包含单个荧光标记物,或者可包含多个荧光标记物。所述荧光标记物通常与内部残基相连。

[0052] 寡核苷酸通常具有 2 或 3 或 4 或 5 或 6 或 7 或 8 或 9 或 10 个由荧光团标记的内

部残基。其数目可取决于寡核苷酸的长度。通常使用荧光团标记约三分之一的内部残基，但更少的标记也是可以的。

[0053] 优选本发明的探针寡核苷酸的长度适合与靶多核苷酸的单链部分杂交，从而提供稳定的杂交体，此杂交体的解链温度取决于靶的精确序列以及靶单链部分内串联重复的数目，但通常在 40℃ 至 70℃ 的范围内。在很多情况下，包含小于 15 个核苷酸残基的寡核苷酸不形成足够稳定的杂交体，特别是在两条杂交序列不完全地互补的情形下，不过它们可以在某些情况下使用。长度大于约 40 个核苷酸残基的寡核苷酸可形成杂交体，其解链温度对于可能存在的单核苷酸错配相对较不敏感，不过它们可以在某些情况下使用。对于具有长度多态性的靶，长寡核苷酸探针与长重复序列的杂交产生的 T_m 的差异小。如下文详述，使用与靶多核苷酸中大于约 30-40 个的核苷酸杂交的寡核苷酸探针通常需要使用高度敏感的方法进行解链曲线和解链峰的分析。

[0054] 探针寡核苷酸的长度通常为 10-60 个核苷酸残基，优选长度为 15-50 个核苷酸残基，更优选长度为 15-40 个核苷酸残基。因此，寡核苷酸的长度为从 10 或 11 或 12 或 13 或 14 或 15 个核苷酸残基至（并包括）35 或 36 或 37 或 38 或 39 或 40 或 45 个核苷酸残基。因此，本发明包括大小在上述任意范围内的探针寡核苷酸的应用。

[0055] 长度范围在 15-60 个核苷酸的寡核苷酸可以是单个标记的，或者具有至多约 14 个（对于 60mer 的寡核苷酸）由荧光团标记的内部核苷酸残基，但长度在此范围内的寡核苷酸通常具有 2 或 3 或 4 或 5 或 6 个由荧光团标记的内部残基。通常在非末端荧光标记的核苷酸之间有大约 3 个碱基。

[0056] 核苷酸碱基通常衍生自天然存在的核苷 A、C、G、T 和 U。然而，可在用于本发明的探针寡核苷酸的一个或多个位点使用核苷酸类似物，例如核苷酸类似物在碱基部分和 / 或糖部分和 / 或磷酸连接处被修饰。碱基修饰如丙炔基 dU (dT- 类似物) 和 2- 氨基 dA (dA 类似物) 通常改变杂交性质，并可使得具有小于 15 个核苷酸残基的寡核苷酸的应用更具吸引力。对于包含丙炔基 dU 的寡核苷酸的情况，其长度为约 10 个残基，这取决于与所需靶序列的解链温度。也可使用如 N4- 乙基 -dC (dC 类似物) 的碱基修饰使长寡核苷酸探针去稳定，从而增大在与长靶序列的解链温度上的差异。因此，在一些实施方式中，可使用碱基修饰获得适合的 T_m 变化。

[0057] 或者，也可以使用包含肽核酸 (PNA)、锁核酸 (LNA)、2'-O- 甲基 RNA、亚磷酰胺 DNA、硫代磷酸酯 DNA、甲基膦酸酯 DNA 或磷酸三酯 DNA 的寡核苷酸，或者由肽核酸 (PNA)、锁核酸 (LNA)、2'-O- 甲基 RNA、亚磷酰胺 DNA、硫代磷酸酯 DNA、甲基膦酸酯 DNA 或磷酸三酯 DNA 组成的寡核苷酸，以形成化学或酶促更稳定的与靶序列的相互作用。

[0058] 优选在用于本发明的探针核苷酸中使用同一荧光团。然而，在寡核苷酸被复用时，在该同一寡核苷酸中使用两种或多种不同荧光团可能特别有利。例如，可在不同探针上使用谱性不同的荧光团，从而在单个反应管中同时分析多个 STR 等位基因。只要不阻止寡核苷酸与其靶序列的杂交，可使用能够与核苷酸残基相连的任何荧光团。（也可以通过使用长度不同的寡核苷酸探针，并使用或不使用不同的阻断寡核苷酸来增强多倍分析，特别是对于极长 STR 等位基因）。

[0059] 适合的荧光团包括基于荧光素的荧光团，例如 FAM (6- 羧基荧光素)、TET (四氯荧光素)、HEX (六氯荧光素)；基于罗丹明的荧光团，例如 ROX (6- 羧基 -X- 罗丹明)

和 TAMRA (6-羧基四甲基罗丹明);Cy 类染料,特别是 Cy3 和 Cy5,上述均可得自于 Glen Research,22825 Davis Drive, Sterling, VA 20164, USA。

[0060] 也可使用其他荧光素染料,例如具有不同发射谱的那些,例如 NED 和 JOE。也可使用其他荧光团,例如 Alexa、Atto、Dyomics、Dyomics Megastokes 和 Thilyte 染料类中的那些,详见下表 14-20。

[0061] 在本发明优选的实施方式中,分别使用 C6 FAM dU(可得自于 University of Southampton, UK) 或荧光素 dT(可得自于 Glen Research, Sterling, VA) 标记探针寡核苷酸内部尿嘧啶/胸腺嘧啶碱基的 5-位(在本文中,dT 和 dU 的结构相同,因此这两个术语可互换)。可将 FMOC-保护的亚磷酰胺加入寡核苷酸的内部 T 位点,并可将其用作各种荧光染料的连接位点,所述荧光染料包括但不限于 FAM、TET、HEX、ROX、TAMRA、Cy3 和 Cy5,上述所有染料均由 Glen Research 获得。在合成寡核苷酸后,可从 2'-保护的尿苷除去 FMOC 基团,并且将荧光亚磷酰胺,例如适当保护的 6-羧酸荧光素亚磷酰胺偶联到游离的 2'-羟基基团。在另一个实施方式中,可在内部的 A、C 或 G 位点标记寡核苷酸,其中在固相寡核苷酸合成期间加入作为亚磷酰胺的标记核苷酸,或者在合成寡核苷酸后使用受保护的亚磷酰胺(例如,8-氨基烷基-dA、7-氨基烷基-7-脱氮-dA、N(4)-氨基烷基 dC 和 5-氨基烷基-dC)来连接荧光团。

[0062] 特别优选的标记探针寡核苷酸为 WO 01/73118 公开的 HyBeacon[®] 探针,或者如 WO 2007/010268 公开的寡核苷酸探针。

[0063] 应理解,在另一个实施方式中,可使用淬灭分子标记探针寡核苷酸,并且靶多核苷酸包含荧光分子,当探针寡核苷酸与靶多核苷酸结合时,所述荧光分子的荧光被淬灭。或者,可使用荧光分子标记探针寡核苷酸,并且靶多核苷酸包含淬灭分子,当探针寡核苷酸与靶多核苷酸结合时,所述淬灭分子使探针寡核苷酸上荧光分子的荧光淬灭。在另一个实施方式中,可使用荧光团和淬灭部分标记探针和阻断寡核苷酸,以在与相邻靶序列杂交(或分离)后增强/淬灭荧光发射。

[0064] 在另一个实施方案中,使用一个荧光团标记探针寡核苷酸,且靶多核苷酸包含另一个荧光团,这样在探针寡核苷酸与靶多核苷酸结合时有能够被测量的荧光共振能量转移(FRET)。在另一个实施方式中,可使用供体和受体荧光团标记探针和阻断寡核苷酸,以在与相邻靶序列杂交后促进 FRET。

[0065] 在使用步骤(a1)并且由此使用一个或多个阻断寡核苷酸的本发明的方法的实施方式中,如果阻断寡核苷酸包含荧光团且探针寡核苷酸包含淬灭剂,则是方便的,反之亦然。适合的荧光团/淬灭剂对在本领域是众所周知的。类似地,如果使用能够相互参与 FRET 的荧光团来标记探针寡核苷酸和阻断寡核苷酸,则是方便的。适合的荧光团对在本领域是也公知的。

[0066] 应理解,当携带荧光团的一条或多条寡核苷酸以某种构型杂交时,能够参与 FRET 的两个荧光团应当在适合发生 FRET 的距离之内,以指示特定的期望结果。类似地,荧光团和淬灭剂应当在适合发生淬灭的距离之内,当淬灭发生时指示特定的期望结果。在阻断寡核苷酸包含淬灭剂且探针寡核苷酸包含荧光团的实施方式中(或者阻断寡核苷酸和探针寡核苷酸分别包含能够进行 FRET 的荧光团的实施方式中),所述阻断寡核苷酸和探针寡核苷酸相邻地与串联重复杂交,从而使荧光团/淬灭剂(或荧光团/荧光团)对彼此相邻。

[0067] 应理解,在优选的实施方式中,在与适当标记的阻断寡核苷酸结合后,标记物(例如,淬灭分子或荧光分子)可与靶多核苷酸结合并成为靶多核苷酸的一部分。

[0068] 在优选的实施方式中,探针寡核苷酸与靶多核苷酸的单链部分中的至少一个串联重复完全地互补。

[0069] 换句话说,在靶多核苷酸的单链部分与探针寡核苷酸之间的整个杂交区域上存在 Watson-Crick 碱基配对。然而,应理解,可存在某些错配或非 Watson-Crick 碱基配对,但探针仍然能够杂交。在这种情况下,探针寡核苷酸与单链靶多核苷酸部分地互补。在一些情况下,可能期望探针寡核苷酸与靶多核苷酸的单链部分部分地互补,这是因为错配能够降低寡核苷酸-靶杂交体的 T_m ,从而在一定程度上控制 T_m (以及 ΔT_m ,即取决于与探针寡核苷酸杂交的串联重复的数目而出现的 T_m 差异)。

[0070] 与此类似,阻断寡核苷酸可以与单链靶多核苷酸完全地互补(优选),或部分地互补(次优选)。

[0071] 在另一个优选实施方式中,靶多核苷酸内的串联重复由一个或两个单链区域侧翼。在这种情况下,阻断寡核苷酸(如果使用)或探针寡核苷酸或这二者(如果靶多核苷酸内的两个侧翼区域均为单链)可包含与串联重复侧翼区域互补的锚定部分。应理解,如果所使用的阻断寡核苷酸和探针寡核苷酸均包含锚定部分,那么寡核苷酸之一中的锚定部分将与一个侧翼区域互补,而另一条寡核苷酸内的锚定部分将与另一个侧翼区域互补。

[0072] 阻断寡核苷酸或探针寡核苷酸的锚定部分可包含 3-40 个核苷酸;通常为 4-30 个核苷酸,例如 4-20 个核苷酸。对于探针寡核苷酸,锚定部分通常为 5-10 个核苷酸。锚定部分通常与侧翼区域完全地互补。应理解,阻断寡核苷酸(如果使用)和探针寡核苷酸的锚定部分的大小和组成可影响阻断寡核苷酸和探针寡核苷酸的 T_m ,因此通过改变探针寡核苷酸锚定区域,可以在一定程度上控制 T_m (和 ΔT_m)。

[0073] 此外,还应理解,锚定区域可用于减少或防止“滑移”(即,其锚定寡核苷酸与侧翼区域的杂交)。这可以改进寡核苷酸与靶多核苷酸杂交的选择性,并且改进与寡核苷酸(特别是探针寡核苷酸)结合的不同数目的串联重复之间的差异。

[0074] 在优选的实施方式中,同时进行此方法的步骤(a)和(a1)。因此,在此实施方式中,在提供包含靶多核苷酸的样品的同时提供一条或多条阻断寡核苷酸,这样在一条或多条阻断寡核苷酸的存在下,样品中的靶多核苷酸包含被一条或多条阻断寡核苷酸阻断的一些串联重复,以及有(保持)单链形式的一些串联重复,从而可以自由地与探针寡核苷酸杂交。在一个实施方式中,阻断寡核苷酸与靶多核苷酸分离(即没有共价结合)。在一个实施方式中,在图 6 中以图示的方式例示了此实施方式。至少在靶多核苷酸的产生和阻断方面,可将其视为双分子反应。

[0075] 在特别优选的实施方式中,PCR 产生靶 DNA,其中所述靶 DNA 内的一个或多个串联重复为单链形式,并且在 PCR 中使用的引物也包含阻断寡核苷酸。在一个实施方式中,在图 8 中以图示的方式例示了此实施方式。至少在靶多核苷酸的产生和阻断方面,可将其视为单分子反应。优选还包含阻断寡核苷酸的 PCR 引物不具有基本自身互补的任何区域。

[0076] 在优选的实施方式中,在 PCR 中使用的引物(此引物还包含阻断寡核苷酸)在其 3' 端包含与靶 DNA 中待分析的串联重复的 3' 端区域互补的部分,并且在其 5' 端包含与通过所述引物合成的 PCR 产物的链中至少一个串联重复互补的部分。因此,此引物是在 PCR

反应中产生 DNA 链的引物,其中所述 DNA 链包含靶 DNA 中的串联重复。因为此引物还包含与一个或多个这些串联重复互补的部分,所以其由引物的 5' 部分(在此实施方式中,其还是阻断寡核苷酸)与通过在 PCR 中使用此引物合成的一个或多个串联重复之间的分子内杂交形成发卡结构。引物的 5' 部分不与引物或寡核苷酸探针的任何部分互补,并且在通过 PCR 扩增靶序列之前不参与杂交。应理解,如果引物在其 5' 端包含与 y 串联重复互补的部分, y 将被阻断。如果在 PCR 扩增的位点存在 x 串联重复,那么在所产生的单链形式的靶 DNA 中将存在 x-y 串联重复。

[0077] 在本发明的此实施方式中,特别优选引物从 3' 向 5' 方向包含:(i) 与靶 DNA 内串联重复 3' 端区域互补的部分,(ii) 任选的间隔子部分,(iii) 与通过所述引物合成的 PCR 产物的链中的侧翼区域互补的锚定部分,和 (iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分。

[0078] 在另一个实施方式中,此引物进一步包含 (v) 与探针寡核苷酸 3' 端的锚部分互补的锚部分。所述锚部分位于 (iv) (与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分) 的 5' 端。

[0079] 在另一个实施方式中,使用至少两对不同的 PCR 引物和探针寡核苷酸,其中每对阻断寡核苷酸和探针寡核苷酸各自的锚部分彼此互补。优选地,选择锚部分的核苷酸序列,以使每对互补的锚部分具有不同的 T_m 。不同 T_m 之间的差别通常为至少 0.5°C。

[0080] 在 PCR 引物包含锚部分的实施方式中,所述锚部分具有如下文详述的阻断寡核苷酸的锚部分的性质。

[0081] 图 15 中以图示的方式例示此实施方式。

[0082] 特别优选使用不对称 PCR 增强阻断剂和探针与合成的靶的杂交。例如,与可在 PCR 中以有限浓度使用的其他引物相比,单分子阻断寡核苷酸/引物以过量的浓度存在,使得在 PCR 中早早地耗尽其他引物,从而产生单链 DNA 靶。

[0083] 图 8 中以图示的方式显示了此实施方式的实例(在此实施方式中,PCR 引物和探针寡核苷酸不包含锚部分)。

[0084] 在不同的实施方式中,PCR 中使用的引物(此引物还包含阻断寡核苷酸)还包含探针寡核苷酸。图 16 中以图示的方式例示此实施方式。通常在探针寡核苷酸和阻断寡核苷酸之间还存在间隔子部分。因此,在典型的实施方式中,所述引物从 3' 向 5' 方向包含:(i) 与靶 DNA 内串联重复的 3' 端区域互补的部分,(ii) 任选的间隔子部分,(iii) 与通过所述引物合成的 PCR 产物的链中的侧翼区域互补的锚定部分,(iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分,(v) 第二间隔子部分,和 (vi) 探针寡核苷酸。探针寡核苷酸 (vi) 包含在 PCR 引物中,并可被视为探针部分。其通常具有荧光标记,并且与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补。与靶 DNA 内串联重复的 3' 端区域互补的引物部分能够在例如 PCR 条件下,与串联重复的 3' 端杂交。通常,此引物部分包含与 3' 侧翼区域互补的 10-30 个核苷酸,例如 15-30 或 15-20 个核苷酸。优选其包含 18-25 个核苷酸。3' 侧翼区域通常位于靶序列 3' 端的 1-200 个核苷酸。

[0085] 优选此部分与靶 DNA 内串联重复的 3' 端区域完全地互补;不过其可以是部分互补的,只要其能够杂交并参与 PCR。引物的 3' 端自由参与链延伸反应,并且因此包含 3' OH 基团。引物还可包含“通用”碱基,例如 5-硝基吡啶和肌苷,从而“中和”在人群中鉴定的

已知单核苷酸多态性。

[0086] 间隔子部分 (ii) 是任选的。当存在时,其可包含任何适合的间隔子,其含义包括占据约 1-20 个核苷酸残基的长度,但不参与碱基配对的化学单元。所述间隔子通常可以是六甘醇 (HEG) 或四甘醇 (TEG) 基团。在另一个实施方式中,所述间隔子可以是一个或多个无碱基残基。无碱基残基保留核苷酸残基的间隔,但不参与碱基配对 (因为没有碱基)。在另一个实施方式中,存在核苷酸残基,但这些残基与靶链中的核苷酸残基错配,因此不参与碱基配对。

[0087] 所述间隔子通常占据约 1、2、3、4 或 5 个核苷酸残基的长度。

[0088] 在探针寡核苷酸包含在 PCR 引物中 (图 16) 的实施方式中,探针寡核苷酸和阻断寡核苷酸之间的间隔子 (v) 通常具有以下性质。其充当允许探针寡核苷酸与阻断寡核苷酸独立杂交的物理间隔子,而没有对稳定性和探针 T_m 的累计作用。适合地,使用如碳水化合物或肽聚合物的简单线性分子,从而具有最小的三维转动限制,并且长度足以在空间中完成环形结构 (这样其末端能够返回到非常接近)。优选间隔子不具有对核酸的内在亲和力,并且不与核酸结合。优选间隔子对靶序列的杂交过程保持中立。显然,间隔子必须能够与阻断寡核苷酸和探针寡核苷酸化学结合。

[0089] 引物中与通过所述引物合成的 PCR 产物的链中的侧翼区域互补的锚定部分可包含 5-25 个核苷酸,通常为 5-20 个核苷酸,例如 10-20 个核苷酸。这可以随侧翼区域的序列组成而改变。方便地,引物的锚定部分具有的 T_m 高于探针寡核苷酸锚定部分。锚定部分通常与侧翼区域完全地互补。

[0090] 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分 (其是引物的一部分) 通常与至少一个,或两个,或三个,或四个或五个串联重复互补。其通常与 2-20 个之间,例如 4-8 个串联重复互补。优选此部分与串联重复完全地互补,但也可部分地互补,只要其仍能够与通过所述引物合成的 PCR 产物内的串联重复杂交。

[0091] 上述间隔子部分也可存在于探针寡核苷酸内或阻断寡核苷酸内。在这些情况下,优选所述间隔子的长度约为 1 或 2 或 3 或 4 或 5 个或更多个核苷酸。

[0092] 应理解,在阻断寡核苷酸不是 PCR 引物一部分的实施方式中,优选其 3' 端不包含 OH 基团。这样使其不能参与链延伸反应。因此,通常使用例如磷酸基团或辛二醇封闭其 3' 端。

[0093] 类似地,应理解,通常可期望探针寡核苷酸不包含 3' -OH,并且通过例如磷酸或辛二醇封闭以防止链延伸。这样,在使用 PCR 实施本发明的方法时,探针寡核苷酸 (如果在 PCR 期间存在) 不会被存在的 DNA 聚合酶延伸。

[0094] 在另一个实施方式中,阻断寡核苷酸在其 3' (或 5') 端包含与存在于探针寡核苷酸 5' (或 3') 端的钳部分互补的钳部分。在此实施方式中,当阻断寡核苷酸和探针寡核苷酸与靶多核苷酸的单链部分杂交时,阻断寡核苷酸的钳部分与探针寡核苷酸的钳部分杂交在一起。在一个具体实施方式中,图 12 中以图示的方式例示了此实施方式。钳部分通常具有 3-10 个核苷酸,例如 4-8 个,如 6-8 个核苷酸。钳部分通常主要包含 G 或 C 残基;优选大于 75% 的碱基为 G 或 C。钳部分的序列优选不与 STR 的任何部分互补。方便地,钳部分对热稳定性贡献 10°C 到 30°C,通常确保探针与正确的序列杂交并且防止滑移。钳部分的 T_m 不应将探针寡核苷酸的 T_m 提高到阻止区分相似长度的靶重复的程度。在一个实施方式

中,探针寡核苷酸的钳部分包含荧光标记物,且阻断寡核苷酸的钳部分包含淬灭分子(反之亦然),从而使荧光团和淬灭剂在与靶多核苷酸的单链部分结合后发生相互作用。或者,探针寡核苷酸和阻断寡核苷酸的钳部分均包含荧光团,从而能够在与靶多核苷酸杂交后参与 FRET(参见图 17)。

[0095] 在优选的实施方式中,使用至少两对不同的阻断寡核苷酸和探针寡核苷酸,其中每对阻断寡核苷酸和探针寡核苷酸各自的钳部分彼此互补。通常,不同的阻断寡核苷酸和探针寡核苷酸对具有不同的钳部分。因此,一对可具有表示为 A1:A1' (A1 和 A1' 互补) 的钳部分,而另一对可具有表示为 B1:B1' 的钳部分,等等,其中,例如 A1 存在于阻断寡核苷酸上,而 A1' 存在于探针寡核苷酸上。通常选择钳部分的核苷酸序列,以使每对互补的钳部分具有不同的 T_m 。

[0096] 带有钳(clamp)部分的阻断寡核苷酸和带有钳部分的探针寡核苷酸可被视为接头寡核苷酸(junction oligonucleotide)。

[0097] 在双分子形式中,阻断寡核苷酸通常将包含(i)与紧邻于串联重复 5' 或 3' 端的侧翼区域互补的锚定部分,(ii)与所合成的 PCR 产物中的至少一个串联重复互补的部分,和(iii)与连接到待使用的探针寡核苷酸 5' 或 3' 端的钳序列互补的富含 G/C 的短钳序列。

[0098] 在单分子模式中,阻断寡核苷酸/引物从 3' 至 5' 通常包含:(i)与靶 DNA 内串联重复的 3' 端区域互补的部分,(ii)任选的间隔子部分,(iii)与通过所述引物合成的 PCR 产物的链中的侧翼区域互补的锚定部分,(iv)与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分,和(v)与连接至待使用的探针寡核苷酸 5' 端的钳序列互补的富含 G/C 的短钳序列。

[0099] 可以方便地实施本发明的方法,其中选择探针寡核苷酸以使可以根据其解链温度(T_m)区分靶多核苷酸的单链部分中的串联重复的数目。因此,相同的探针寡核苷酸能够区分存在的不同数目的串联重复。通常,探针寡核苷酸可区分靶多核苷酸中以单链形式存在的 2 个、3 个和 4 个串联重复,或者 3 个、4 个和 5 个,或者 4 个、5 个和 6 个,或者 5 个、6 个和 7 个,或者 6 个、7 个和 8 个,或者 7 个、8 个和 9 个,或者 8 个、9 个和 10 个,或者 9 个、10 个和 11 个串联重复,等等。探针寡核苷酸可以方便地区分靶多核苷酸中以单链形式存在的 2 个、3 个、4 个、5 个和 6 个串联重复。然而,通过使用适合的阻断寡核苷酸,可以分析带有很多种可能的串联重复的 STR。

[0100] 优选连续数目的重复之间的 ΔT_m (即,取决于与探针寡核苷酸杂交的串联重复的数目的 T_m 差异)不小于 0.5°C 。更优选 ΔT_m 为至少 1°C ,通常为至少 2°C ,例如至少 3°C 。可通过改变其长度、组成、与单链多核苷酸靶的互补程度、锚定部分(如果存在)或钳部分(如果存在)来调整探针寡核苷酸的杂交性质。

[0101] 优选此方法的杂交步骤(b)在接近由靶多核苷酸的单链部分和探针寡核苷酸之间形成的一个或多个杂合体的一个或多个 T_m 的指定温度下进行。可以通过参考靶多核苷酸的单链部分中待分析的串联重复的数目,是否存在单链侧翼区域,以及上述探针寡核苷酸的形式来选择预定的温度。

[0102] 更优选此方法的杂交步骤(b)在包括由靶多核苷酸的单链部分和探针寡核苷酸之间形成的一个或多个杂合体的 T_m 的一定温度范围内进行。

[0103] 方便地,所述温度范围可以为 30℃ -75℃。形成的杂交体通常具有在 40℃ -65℃ 范围内的 T_m 。解链峰分析通常应用一套算法以平滑数据(即除去荧光噪声),其在 ΔT_m 较小时降低区分等位基因的潜能。通过解链峰的 T_m 确定重复的数目。高分辨率解链曲线分析将温度范围的低温区域和高温区域内荧光发射水平进行平均,以进行背景校正。具有小 ΔT_m 的多态性等位基因更可能通过使用高分辨率解链曲线分析来区分。可通过解链曲线的形状区分 STR 等位基因,但确定重复数目可能需要一整套互补的长度标准品进行比较。

[0104] 可获得具有低热分辨率限的各种仪器,可将其用于区分 T_m 非常接近的分子的解链。例如,来自 Idaho Technologies 的高分辨率解链 (HRM) HR-1 仪器能够区别差别小于 1℃,例如 0.5℃,乃至仅差别 0.1℃ 的 T_m 。其他适合的仪器包括 LightScanner (Idaho Technologies)、Light Cycler 480 (RocheDiagnostics) 和 Rotor-Gene 6000 (Corbett Life Sciences),据报道,其可处理超过 1.0℃ 的热分辨率。Corbett 声称其 Rotor-Gene 6000 具有 0.02℃ 的热分辨率,不过这是基于解链曲线,而非解链峰。

[0105] 靶多核苷酸可以是包含串联重复的任何靶多核苷酸。靶多核苷酸通常是由待分析的天然 DNA 生成的 DNA 或 RNA。天然 DNA 通常是来自植物、动物或微生物的基因组 DNA。如引言部分所述,在很多基因组中发现串联重复序列,其分析可用于很多情况。对人和其他哺乳动物 DNA 内串联重复的分析可用于医学诊断、法医学、亲子和亲缘关系测试,以及绘制连锁图。

[0106] 对微生物基因组内串联重复的分析可用于疾病控制(例如菌株鉴定)和工业情况(例如,鉴定酿酒酵母和面包酵母株)。

[0107] 对食品中 DNA 内串联重复的分析可用于跟踪食物链中的材料,例如所使用的植物材料的类型等。

[0108] 特别优选的天然 DNA 是人基因组 DNA。如上所述,靶 DNA 通常通过扩增反应,例如 PCR 生成。

[0109] 本发明的第二方面提供了用于确定靶多核苷酸内串联重复的数目的系统,其中所述靶多核苷酸内一个或多个串联重复为单链形式,所述系统包括:

[0110] (a) 与靶多核苷酸的单链部分中的至少一个串联重复互补的标记探针寡核苷酸,且所述探针寡核苷酸的至少 5 个核苷酸与靶多核苷酸的单链部分中的串联重复互补,和

[0111] (b) 与靶多核苷酸中至少一个但并非全部串联重复互补的阻断寡核苷酸。

[0112] 所述标记探针寡核苷酸和阻断寡核苷酸优选具有如本发明第一方面的方法中所述的性质。

[0113] 方便的系统包括但不限于:

[0114] (1) (a) 包含钳部分的标记探针寡核苷酸,和 (b) 包含互补钳部分的阻断寡核苷酸;

[0115] (2) (a) 标记探针寡核苷酸,和 (b) 包含上述阻断寡核苷酸的 PCR 引物;

[0116] (3) (a) 包含钳部分的标记探针寡核苷酸,和 (b) 包含上述阻断寡核苷酸和钳部分的 PCR 引物;和

[0117] (4) 包含上述阻断寡核苷酸和探针寡核苷酸的 PCR 引物。

[0118] 可将此系统视为包含两条寡核苷酸(或者在某些实施方式中为一条寡核苷酸)的“组分试剂盒”。方便地,此试剂盒也可包含用于从天然存在的 DNA 分子产生靶多核苷酸的

PCR 成分。此试剂盒还可包含用于检测探针寡核苷酸与靶多核苷酸的杂交的方法。

[0119] 本发明的第三方面提供了寡核苷酸,其包含含有与锚定部分结合的至少两个串联重复的部分,其中所述锚定部分的序列和所述至少两个串联重复在靶多核苷酸中连续地存在。

[0120] 优选所述寡核苷酸是标记的。优选寡核苷酸的标记如上所述。在本文第三方面,靶 DNA 通常是全部或部分的人 STR 及其侧翼区域。优选所述靶 DNA 是表 21 所示任何 STR 和侧翼区域的全部或部分。

[0121] 优选所述锚定部分具有与上述阻断寡核苷酸或探针寡核苷酸的优选性质相同的性质。

[0122] 优选本发明此方面的寡核苷酸还包含间隔子部分。优选所述间隔子部分具有上述间隔子部分的性质。

[0123] 本发明的第四方面提供了参与 PCR 反应以扩增包含串联重复的靶 DNA 的寡核苷酸引物,所述引物从 3' 向 5' 方向包含:(i) 与靶 DNA 内串联重复的 3' 端区域互补的部分,(ii) 任选的间隔子部分,(iii) 与通过所述引物合成的 PCR 产物的链中的侧翼部分互补的锚定部分,和 (iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分。

[0124] 任选地,所述寡核苷酸还进一步包含 (v) 钳部分。此寡核苷酸可用于上述本发明的适合的实施方式中。

[0125] 任选地,在不同的实施方式中,寡核苷酸可进一步包含 (v) 间隔子部分和 (vi) 探针寡核苷酸。此寡核苷酸可用于上述本发明的适合的实施方式中。

[0126] 优选寡核苷酸引物不具有基本自身互补的任何区域。锚定部分和阻断部分(即 (iv) 部分)在靶 DNA 合成之前不参与杂交。

[0127] 在本文第四方面,靶 DNA 优选如上述第三方面所述。

[0128] 与此类似,锚定部分和间隔子部分优选如上述第三方面所述。

[0129] 本发明还包括用于测定靶 DNA 内串联重复的数目的系统,所述系统包含本发明第三方面和第四方面的寡核苷酸。

[0130] 本发明还包括用于制备本发明第三方面的寡核苷酸的方法,所述方法包括:

[0131] (a) 选择包含串联重复的靶 DNA,

[0132] (b) 获得串联重复的序列以及一个或多个侧翼区域的序列,

[0133] (c) 合成寡核苷酸,该寡核苷酸包含含有与锚定部分结合的至少两个串联重复的部分和任选的在其 3' 端的钳部分,其中所述锚定部分的序列和所述至少两个串联重复在靶 DNA 中连续地存在。

[0134] 本发明还包括用于制备本发明第四方面的寡核苷酸的方法,所述方法包括:

[0135] (a) 选择包含串联重复的靶 DNA,

[0136] (b) 获得串联重复的序列以及一个或多个侧翼区域的序列,

[0137] (c) 合成寡核苷酸,其从 3 向 5' 方向包含:(i) 与靶 DNA 内串联重复的 3' 端区域互补的部分,(ii) 任选的间隔子部分,(iii) 与通过所述引物合成的 PCR 产物的链中的侧翼区域互补的锚定部分,(iv) 与通过所述引物合成的 PCR 产物的链中的至少一个串联重复互补的部分,和任选的 (v) 钳部分,或任选的 (v) 间隔子部分,和 (vi) 探针寡核苷酸。

[0138] 应理解,如果此寡核苷酸是具有阻断重复的引物,则优选其中存在间隔子,从而能够形成用于杂交的茎环。

[0139] 本发明这些方面选择的靶 DNA 可以是包含串联重复的任何 DNA。靶 DNA 通常是基因组 DNA。优选其为哺乳动物基因组 DNA,更优选人基因组 DNA。然而,也可以是来自植物、酵母或细菌的 DNA。

[0140] 优选所述靶 DNA 含有包括侧翼序列的人 STR。更优选所述 STR 是表 21 中所述之一。

[0141] 本文引用的所有文献均并入本说明书中。

[0142] 现在将参考以下附图和实施例更详细地描述本发明。

[0143] 图 1:由 STRV2 寡核苷酸与寡核苷酸靶的杂交产生的解链峰。由左向右的各个解链峰分别是由具有 5、7、8、10 和 11 个 TATC 串联重复的靶产生的。

[0144] 图 2:A) 150nM STRV2 寡核苷酸与 75nM RC7 靶寡核苷酸及 75nMRC11 靶寡核苷酸的杂交的解链曲线分析。产生 ΔT_m 为 6.6°C 的两个清晰解链峰。B) 150nM STRV2 探针与 75nM RC10 靶寡核苷酸及 75nM RC11 靶寡核苷酸的杂交的解链曲线分析。产生的宽解链峰阻碍了对等位基因组成的清楚鉴定。C) STRV2 探针与 RC8 和 RC10 寡核苷酸混合物的杂交产生了 ΔT_m 为 3.7°C 的两个清晰解链峰。

[0145] 图 3:A) 由 STRV3 探针与 RC10d 和 RC11d 靶寡核苷酸的杂交产生的解链峰。观察到两个解链峰。B) 由于没有锚定序列, HYBSTR 探针与 RC10e 靶寡核苷酸的杂交产生宽且多噪音的解链峰。

[0146] 图 4:从 NCBI 数据库获得的 D16S539 STR 序列(登录号 G07925)(SEQ IDNo.:70)。

[0147] 图 5:A) 使用 LightCycler 仪器和 HYBSTR 寡核苷酸探针对 D16S539 序列进行的实时扩增和检测。B) 使用 11 和 13 个重复的杂合样品产生的解链峰。C) 使用 8 和 13 个重复的杂合样品产生的解链峰。

[0148] 图 6:A) 使用双分子阻断剂策略分析 STR 靶。STR 靶以黑色方格表示;阻断寡核苷酸中的重复以网纹方格表示;荧光探针中的重复以灰色方格表示。当重复数目小于 15 时,阻断剂的长锚定部分以及所使用的摩尔过量(相对于探针)阻止探针与靶完全杂交。B) 在 B1 阻断剂存在下,使用 FL1 探针与靶寡核苷酸产生的解链峰。由左向右的各个解链峰分别是由具有 8、9、10、11 和 12 个 GATA 重复的靶产生的。C) 在没有 B1 阻断剂时,使用 FL1 探针与靶寡核苷酸产生的解链峰。由左向右的各个解链峰分别是由具有 5、6、7、8 和 9 个 GATA 重复的靶产生的。

[0149] 图 7:A) 使用 LightCycler 仪器、FL1 探针和 B1 阻断剂对来自提取 DNA 和唾液样品的 STR 靶进行的实时扩增。双分子 B1 阻断剂降低了 PCR 的效率,从而将检测循环(Cts)推迟到约 38-40 个循环。省略 PCR 分析中的阻断剂获得约 32 个循环的 Cts。B) 使用 9 和 12 个重复的等位基因的杂合样品产生的解链峰。

[0150] 图 8:A) 使用单分子阻断剂/引物策略分析 STR 靶。STR 靶以黑色方格表示;阻断寡核苷酸中的重复以网纹方格表示;荧光探针中的重复以灰色方格表示。阻断剂重复、锚定序列和正向引物均包含在单个寡核苷酸内。阻断剂和扩增的靶序列成为同一 DNA 链的部分,因此单分子阻断剂不会通过妨碍 Taq 聚合酶的行进而抑制 PCR。在热力学上,阻断剂与扩增的重复之间的单分子结合比探针杂交更为有利,从而在重复数目小于 15 时,阻止探针

全长结合。提供了与具有 8 和 13 个重复的 D16S539 等位基因的探针杂交的例示。

[0151] 图 9 :A) 使用 LightCycler 仪器、FL1 探针和单分子 BP1 阻断剂 / 引物对来自提取 DNA 和唾液样品的 STR 靶进行的实时扩增。扩增产品的检测 Cts 为约 32 个循环。B) 由左向右, 由 11、12 和 13 个重复的等位基因的纯合样品产生的解链峰。C) 使用 8 和 13 个重复的等位基因的杂合样品产生的解链峰。D) 具有 9 和 12 个重复的等位基因的杂合样品。D) 具有 11 和 13 个重复的等位基因的杂合样品。D) 具有 11 和 12 个重复的等位基因的杂合样品。D) 具有 12 和 13 个重复的等位基因的杂合样品。

[0152] 图 10 :由 FL4 探针和 BP1 阻断剂产生的解链峰。A) 提取的具有 9 和 12 个重复的等位基因的杂合 DNA 样品。B) 未纯化的 11/13 基因型唾液样品的直接分析。C) 未纯化的 8/13 基因型唾液样品的直接分析。D) 提取的 10/11 基因型 DNA 样品。E) 提取的 10/13 基因型 DNA 样品。

[0153] 图 11 :使用 HR-1 高分辨率解链仪器重新分析包含扩增的靶序列、BP1 单分子阻断剂 / 探针和 FL1 寡核苷酸的 LightCycler 毛细管。A) 由左向右, 由 11、12 和 13 个重复的等位基因产生的解链峰。纯合的 11/11、12/12 和 13/13 基因型产生单个解链峰, 而杂合的 11/12 和 12/13 基因型产生带有肩峰的较宽的峰型。B) 高分辨率解链曲线数据证实 11、12 和 13 个重复的等位基因的可靠区别。C) 11/12 和 12/13 基因型与 11/11、12/12 和 13/13 基因型的区别。D) 使用 OriginPro 7.5 软件 (OriginLab, Massachusetts, USA) 重新分析 LightCycler 数据。在 BP1 阻断剂存在下, FL1 探针对 11/13 基因型唾液样品产生宽解链峰, 对 11 个重复的等位基因仅表现出小的肩峰。OriginPro 软件将此肩峰鉴定为真峰 (true peak), 并且绘制了 11 和 13 个重复的等位基因的两个组分的解链峰。

[0154] 图 12 :合成具有 5 和 7 个 TATC 重复、锚定序列和富含 GC 的钳的阻断寡核苷酸。合成具有 10 和 8 个 TATC 重复, 并且带有与阻断剂中的钳互补的 GC 钳的寡核苷酸。A) 在 5LSB1 阻断剂的存在下, 10SFL1 寡核苷酸中仅有 5 个重复与 10 个重复的靶杂交。B) 在 7LSB2 阻断剂的存在下, 8SFL2 寡核苷酸中仅有 3 个重复与 10 个重复的靶杂交。C) 使用具有 8-14 个 GATA 重复的互补寡核苷酸由 5LSB1 和 10SFL1 产生的解链峰。D) 使用具有 10-15 个 GATA 重复的互补寡核苷酸由 7LSB2 和 8SFL2 产生的解链峰。

[0155] 图 13 :从 NCBI 数据库获得的 TH01STR 序列 (登录号 NT_009237) (SEQID No. :71)。

[0156] 图 14 :使用 HYBTH01 探针和双分子阻断剂 TH01_BL, 由扩增的 TH01STR 等位基因获得的解链峰。显示了代表 8、9、9.3 和 10 个重复的等位基因的解链峰。

[0157] 图 15 :单分子阻断剂可具有长度和 / 或组成不同的钳部分 (A1、A1、A3 等), 探针寡核苷酸可具有与 PCR 引物上钳部分互补的钳部分 (A1'、A2'、A3' 等)。

[0158] 图 16 :与单分子构建物中 PCR 引物相连的探针寡核苷酸。

[0159] 图 17 :包含能够进行 FRET 的荧光团, 或荧光团 :淬灭剂对的互补钳部分。

[0160] 图 18 :使用两个“偏置”D8S1179 阻断剂的结果, 允许检测 8、9、10、11、12 和 13 个重复的等位基因。标为“+”的解链峰来自于全长探针与未阻断的靶重复的杂交。

[0161] 图 19 :11 和 15 个重复的 D8S1179 等位基因的检测。分别使用 HYBD8 探针与 7.2 个重复和 10.3 个重复的阻断剂同时检测 11 和 15 个重复的等位基因, 所有的寡核苷酸均包含在同一个管中。

[0162] 实施例 1

[0163] 用于实施例的材料和方法

[0164] 寡核苷酸探针的设计和合成

[0165] 标准 DNA 亚磷酰胺 (DNA phosphoramidite)、固相支持物和其他试剂购自于 Link Technologies 或 Applied Biosystems Ltd。补骨脂素 C6 亚磷酰胺购自于 Glen Research Inc。所有的寡核苷酸均使用 0.2 微摩尔亚磷酰胺进行的酸催化脱三苯甲基作用、偶联、加帽和碘氧化循环在 Applied Biosystems 394 自动 DNA/RNA 合成仪上合成。使正常单体 (A、G、C 和 T) 偶联 25 秒,并允许所有其他单体再偶联额外的 300 秒。通过测量三苯甲基阳离子导电性来测定分步偶联效率以及带有 DMT 保护的单体的总产量,在所有情况下均 > 98.0%。在密封试管中在 55℃ 下于浓氨水 (33%) 中将寡核苷酸从固相支持物上断裂去除,反应进行 5 小时。使用 C6FAM dU (University of Southampton, UK) 或荧光素 dT (Glen Research, Sterling, VA) 将荧光团连接到探针序列的内部残基上。在 C6 FAM dU 的情况下,通过本领域技术人员熟知的 DNA 合成方法将 6-羧基荧光素 (FAM) 连接到尿嘧啶碱基的 5-位。本发明的寡核苷酸具有 3'-磷酸组分或其他阻断剂以阻止将探针加入实时 PCR 测定中时由 Taq 介导的延伸。通过将寡核苷酸探针等分试样溶解在指定体积的水中并测量 260nm 处的紫外吸收来测量合成获得的探针的量。由寡核苷酸的 UV 吸收及其在 260nm 处的消光系数计算探针的浓度。通过将构成此寡核苷酸的未修饰的和荧光标记的核苷酸的单个消光系数的加和来计算寡核苷酸的消光系数。

[0166] 寡核苷酸的纯化

[0167] 使用由 Gilson 7.12 软件控制的 ABI Aquapore 柱 (C8), 8mm x 250mm, 孔径 300 Å, 在 Gilson 系统上通过反相 HPLC 进行寡核苷酸的纯化。使用以下程序:运行时间:30 分钟;流速:3mL/分钟;二元系统:时间按分钟计(缓冲液 B%);0(0);3(0);5(20);22(100);25(100);27(0);30(0)。洗脱缓冲液 A:0.1M 乙酸铵, pH7.0;缓冲液 B:0.1M 乙酸铵和 25% 乙腈, pH7.0。通过 310nm 处(荧光素寡聚物)或 295nm 处(所有其他寡聚物)检测洗脱过程。HPLC 纯化后,按照厂商的说明,使用一次性 NAP 10 Sephadex 柱 (Pharmacia) 对寡核苷酸进行脱盐,等分取样至 Eppendorf 管中,并在 -20℃ 下贮存于蒸馏去离子水中。

[0168] 聚合酶链式反应

[0169] PCR 的体积通常为 20 μl,一般包含 2 μl 样品,1x QIAGEN PCR 缓冲液、0.5 μM 正向引物、0.1 μM 反向引物、1 单位 Taq HotStarTaq 聚合酶、3mM 总 MgCl₂、5ng/μl BSA (Roche Diagnostics)、1mM dNTPs (GE Healthcare) 和 150nM 探针。使用 LightCycler 仪器 (Roche Diagnostics) 进行靶的均相扩增和检测,其中在初步的变性反应步骤 (95℃, 15min) 后,使用包括变性 (95℃, 5s)、引物退火 (55℃, 10s) 和产物延伸 (72℃, 10s) 的 50 个循环来扩增靶。在每个循环中的每个引物退火步骤结束时进行荧光采集。在 LightCycler 扩增后,立即进行解链曲线分析,其中快速地将样品变性 (95℃, 5 秒) 并冷却 (35℃, 30 秒),然后使用 0.1℃/秒的转变率将温度从 35℃ 升高到 95℃,并进行连续的荧光采集。通过绘制荧光相对于温度的负导函数 (y 轴上的 -dF/dT) 与温度 (x 轴) 的曲线,使用 LightCycler 软件 (3.5 版) 获得解链峰。使用探针峰的解链温度 (T_m) 检测并鉴定靶。

[0170] 还使用 384 孔 PCR 白板 (Bio-Rad) 和 384 孔 Tetrad 热循环仪 (MJ Research Inc) 进行靶序列的扩增。热程序通常由以下组成:用于激活热启动酶的初步变性阶段 (95℃, 15 分钟),随后进行包含变性 (95℃, 15s)、引物退火 (55℃, 30s) 和产物延伸 (72℃, 30s) 的 50

个 PCR 循环。在扩增后,立即使用 LightTyper 仪器 (Roche Diagnostics) 进行解链曲线分析,使用 0.1℃ / 秒或 0.05℃ / 秒的转变率将样品从 35℃ 逐步加热至 75℃。

[0171] 短串联重复的分析

[0172] 使用一系列寡核苷酸靶研究寡核苷酸探针分析短串联重复 (STR) 的可能性。合成 3 种探针,用于检测和区分 D16S539 等位基因 (表 1),D16S539 等位基因可包含 5-15 个 GATA 重复。所有探针均包含 5' -GGTG 锚定序列,发现此锚定序列降低探针沿重复序列滑移的可能性,从而防止产生宽且多噪音的解链峰。在没有锚的情况下,STR 探针的 5' 重复可能与靶重复中的任意一个相互作用,从而可能出现全长杂交或部分杂交的事件。紧侧翼于重复序列的锚定序列促使探针在特异位点杂交,并且有助于防止 DNA 滑移的现象。锚的稳定性和有效性在很大程度上由其长度和序列组成决定。稳定性不足将导致一定程度的 DNA 滑移,而 T_m 过大的锚可能妨碍对 STR 等位基因的区分。因此,锚序列有助于优先于探针在任何数目的位点中结合之前完成靶序列的第一重复与寡核苷酸探针的第一同源重复的杂交。

[0173] STRV2 和 STRV3 探针还包含试图将探针分隔成两个分离组分的六甘醇 (HEG) 修饰,从而降低整个探针的 T_m 并改进用于区分等位基因的可能性。探针由两个荧光素部分标记并具有 3' - 磷酸以阻止其在包括于 PCR 测定中时被 Taq 聚合酶的延伸。

[0174] 表 1:寡核苷酸和寡核苷酸靶序列,其中 5、(HEG) 和 3P 分别代表荧光团 C6FAM dU、六甘醇和 3' - 磷酸。

[0175]

寡聚物	SEQ ID	序列	标注
HYBSTR	3	GGTGGATAGA5AGATAGA5AGATAGATAGATAGAT AGATAGATAGATAGATAGATAGATA3P	14 个重 复
STRV2	4	GGTGGATAGATAGATA (HEG) GATGA5AGATAGA5 AGATAGATAGATAGATA3P	11 个重 复
STRV3	5	GGTGGATAGATAGATA (HEG) (HEG) GATAGA5AGAT AGA5AGATAGATAGATAGATAGATAGATAGATAGAT A3P	16 个重 复
RC5	6	TATCTATCTATCTATCTATCCACC	5 个重 复
RC7	7	TATCTATCTATCTATCTATCTATCTATCCACC	7 个重 复
RC8	8	TATCTATCTATCTATCTATCTATCTATCTATCCACC	8 个重 复
RC10	9	TATCTATCTATCTATCTATCTATCTATCTATCTATCTA TCCACC	10 个重 复

RC11	10	TATCTATCTATCTATCTATCTATCTATCTATCTATCTA TCTATCCACC	11 个重 复
RC10b	11	TATCTGTCTATCTGTCTATCTGTCTATCTGTCTATCT GTCCACC	5G/T 错配
RC10c	12	TGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT CTGTCCACC	10G/T 错配
RC10d	13	TATCTATCTATCTATCCCCCTATCTATCTATCTATCTA TCCACC	“敲除” 个重复 6
RC10e	14	TATCTATCTATCTATCTATCTATCTATCTATCTATCTA TC	除去 4bp 锚
RC11b	15	TATCTGTCTATCTGTCTATCTGTCTATCTGTCTATCT GTCTATCCACC	5G/T 错配
RC11c	16	TGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT CTGTCTGTCCACC	11G/T 错配
RC11d	17	TATCTATCTATCTATCTATCCCCCTATCTATCTATCTA TCTATCCACC	“敲除” 个重复 6
RC11e	18	TATCTATCTATCTATCTATCTATCTATCTATCTATCTA TCTATC	除去 4bp 锚

[0177] 探针 HYBSTR 为研发的起点。对探针 ΔT_m 和等位基因鉴别的改进应当再参考此探针。

[0178] 合成寡核苷酸靶用于模拟各种 D16S539 等位基因,该 D16S539 等位基因具有 5、7、8、10 和 11 个 TATC 重复以及与探针锚互补的 3' -CACC 序列 (表 1)。在 TaKaRa PCR 缓冲液和总计 3mM 的 $MgCl_2$ 中,将 150nM 的 D16 寡核苷酸探针与 150nM 的各靶寡核苷酸杂交。使用 LightCycler 仪器进行解链曲线分析,其中在将初步变性 (95°C, 5 秒) 并冷却 (35°C, 30 秒) 后,使用 0.1°C / 秒的转变率将反应从 35°C 加热到 95°C。表 2 中详细列出了由合成靶序列观察到的探针解链温度。图 1 中显示了由 STRV2 探针和各寡核苷酸靶产生的解链峰。

[0179] 另外,还使用靶寡核苷酸的组合分析了探针,其中所述组合模拟了具有不同重复数目的等位基因的杂合基因型。对杂合样品进行基因分型的能力取决于所存在的两个 D16 等位基因的 ΔT_m 。LightCycler 仪器通常需要至少 3°C 的 T_m 差异以产生具有清晰组分峰的杂合解链印迹 (然而,可获得提供更好分辨率的其他仪器,例如高分辨率解链仪器,如来自 Idaho Technologies 的 HR-1)。较短的 D16 靶 (例如 5 和 7 个重复) 显示出比较长的序列 (例如 10 和 11 个重复) 大很多的 ΔT_m , 因此更易于区分。图 2 显示了使用 STRV2 可清楚

地区分 7/11 和 8/10 基因型,而 10 和 11 个重复的 T_m 过于类似而不能同时检测。(然而,如上所述,使用阻断寡核苷酸 (BP1) 和不同的探针寡核苷酸 (FL1) 实现了 10 和 11 个重复的同时检测,显示 ΔT_m 为 3.5°C)。

[0180] 表 2:探针 T_m 和由与寡核苷酸靶序列的杂交获得的 ΔT_m

[0181]

重复的组合	探针 T_m & (ΔT_m)		
	HYBSTR	STRV2	STRV3
5/5		47.66°C	
7/7		55.09°C	
8/8		56.83°C	
10/10	63.62°C	60.60°C	59.84°C
11/11	64.54°C	61.70°C	61.83°C
5/7		(7.4°C)	
5/8		(9.2°C)	
5/10		(12.9°C)	
5/11		(14.0°C)	
7/8		(1.7°C)	
7/10		(5.5°C)	
7/11		(6.6°C)	
8/10		(3.7°C)	
8/11		(4.9°C)	
10/11	(0.92°C)	(1.1°C)	(1.99°C)

[0183] 实施例 2

[0184] 使用实施例 1 中所述的寡核苷酸设计不是总可以准确地分辨出包含长重复序列的杂样品。由于探针 ΔT_m 的量级依赖于 STR 靶的长度和组成,使杂交不稳定化可能提高对长度多态性的敏感度。可通过如下策略完成探针的不稳定化:

[0185] • 引入核苷酸错配以降低探针 T_m 。错配的数目取决于探针的长度和序列组成以及所采用的错配的类型。可在沿探针长度上的多个重复中使用较大量的稳定错配,例如 G/T 以降低 T_m 。可能需要较少量高度不稳定的错配,例如 C/A 以完成相同的 T_m 降低。沿探针长度上分配错配的另一方式是将错配成簇,从而从寡核苷酸上除去整个重复,例如,将 15 个重复的探针分成 5 个重复和 9 个重复组分。

[0186] • 加入碱基类似物,例如 N4- 乙基 -dC7- 脱氮 -dG、7- 脱氮 -dC、C-5 丙炔基 -dC、C-5 丙炔基 -dU、5- 甲基 -dC、2- 氨基 -dA、G- 钳 1 (三环氨基乙基吩噻嗪 2' -dC 类似物)、锁核酸 (LNA)、5' - 三甲氧基苄帽、5' - 苣帽。使用这些类似物在提高 T_m 方面具有优点,可进一步区分不同探针区域对整体解链温度的贡献,从而进一步增强在不同重复长度之间的观测的 ΔT_m ,以改进探针 T_m 。

[0187] • 包含较长的 HEG、TEG (或其他) 间隔子以将探针拆分成单独的组分。

[0188] • 使用两个以上的荧光团标记寡核苷酸。

[0189] 为了研究核苷酸错配对于探针 T_m 和 ΔT_m 的影响,将碱基取代加入寡核苷酸靶而非探针中,以便于试验评估。表 1 中详细列出了对包含 10 和 11 个重复的靶序列的修饰。

[0190] RC10b、RC10c、RC11b 和 RC11c 中加入了有规则地分布在整个探针 / 靶双链中的 5 和 10 个 G/T 核苷酸错配。碱基取代显著地降低了探针 / 靶双链的 T_m (比较表 2 和表 3), 并且还提高了 HYBSTR、STRV2 和 STRV3 探针的 ΔT_m 。然而, ΔT_m 增加的量级不足以使用 LightCycler 软件对 10 和 11 个 STR 重复的等位基因进行可靠的区分。

[0191] RC10d 和 RC11d 寡核苷酸在 STR 靶共有的第六个重复中加入了 4 个核苷酸错配。这些错配降低了 D16 寡核苷酸的 T_m , 并且提高了 10 和 11 个重复的寡核苷酸之间的 ΔT_m (表 3)。使用 STRV3 探针可见到 10 和 11 个重复的两个峰 (图 3A)。然而, 还需要进一步增强探针 ΔT_m , 从而能够使用 LightCycler 软件可靠地鉴定并区分长 STR 靶, 下文对此进行了详述。

[0192] RC10e 和 RC11e 靶寡核苷酸证明 HYBSTR 探针需要 4bp 的锚。寡核苷酸对所有其他修饰和未修饰的靶产生确定的平滑曲线, 但在没有锚时产生宽且多噪音的印迹 (图 3B)。峰质量的下降是由于探针沿靶序列滑移造成的。值得注意的是, STRV2 在不存在锚的情况下没有显示出此类峰质量的下降, 这可能是由于存在内部 HEG 修饰造成的。

[0193] 表 3: 探针与错配寡核苷酸靶的 T_m 和 ΔT_m

[0194]

探针	RC10b	RC11b	ΔT_m	RC10c	RC11c	ΔT_m	RC10d	RC11d	ΔT_m
HYBSTR	56.33	57.44	1.11	48.34	49.37	1.03	57.80	59.46	1.66
STRV2	52.18	53.47	1.29	46.03	45.02	1.01	53.2	55.3	2.1
STRV3	51.05	53.54	2.49	43.22	44.54	1.32	54.37	56.03	1.66

[0195] 实施例 3

[0196] 使用引物 STRF2 和 STRR2, 通过聚合酶链式反应 (PCR) 扩增 D16S539 靶序列。扩增子的大小随重复数目而变化, 范围从 133bp 至 173bp (基因序列参见图 4)。

[0197] STRF2 CAGATCCCAAGCTCTTCCTCTTCCCTAG (SEQ ID :19)

[0198] STRR2 ACGTTTGTGTGTGCATCTGTAAGCATGTATC (SEQ ID :20)

[0199] 未经 DNA 纯化, 使用 HYBSTR 探针和 LightCycler 仪器直接分析 7 份唾液样品。唾液样品均为具有 9/12、13/15、9/13、11/14、9/12、11/13 和 8/13 个重复基因型的杂合样品。实时荧光的增加以及产生的解链峰确认了从唾液样品有效地扩增出 D16 靶。具有相差 2、3 或 4 个重复的 D16 等位基因的杂合基因型没有对每个组成等位基因产生清晰的个体峰。反而产生了由各 D16 重复的峰组合而成的宽解链图谱。8/13D16 基因型的 5 个重复差异足以对每个等位基因产生清晰的峰 (图 5)。

[0200] 实施例 4

[0201] 使用既具有 5' - 又具有 3' - 锚定序列的探针以及仅具有 3' - 锚定序列的探针

来研究无碱基位点对双链稳定性的影响（表 4）。根据 DNA 的负链（minus）设计探针，增加用于连接荧光团的可能位点的数目（表 5）。探针与具有 5、7、10、12 和 15 个重复的互补寡核苷酸杂交（表 5）。

[0202] 表 4：用于分析无碱基位点的影响的寡核苷酸探针，其中 4 和 5 分别代表无碱基位点和荧光素 dT 荧光标记。

[0203]

探针	SEQ ID	序列
02466R	21	CAATGATA5CTATCTA5CTATCTA5CTATCTA5CTATCTA5CTATCT A5CTATCTATCTATCTATCCACC
02467R	22	CAATGATA5CTATCTA5CTATCTA5CTATCTA5CTATCTA5CTATCT A5C4ATCTATCTATCTATCCACC
02468R	23	TA5CTATCTA5CTATCTA5CTATCTA5CTATCTA5CTATCTATCTAT CTATCTATCTATCCACC
02469R	24	TA5CTATCTA5CTATCTA5CTATCTA5CTATCTA5CTATC4ATCTAT CTATCTATCTATCCACC

[0204] 表 5：用于探针评估的作为靶使用的寡核苷酸，其中 (GATA)_n 代表 STR 重复的数目。

[0205]

寡核苷酸	SEQ ID	序列
C5	25	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₅ TCATTGAAAG
C7	26	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₇ TCATTGAAAG
C8	27	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₈ TCATTGAAAG
C9	28	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₉ TCATTGAAAG
C10	29	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₁₀ TCATTGAAAG
C11	30	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₁₁ TCATTGAAAG
C12	31	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₁₂ TCATTGAAAG
C13	32	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₁₃ TCATTGAAAG
C14	33	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₁₄ TCATTGAAAG

C15	34	CCCTAGATCAATACAGACAGACAGACAGGTG (GATA) ₁₅ TCATTGAAAG
-----	----	---

[0206] 包含无碱基位点通常将探针杂交的 T_m 降低约 1-2°C。然而,没有获得 ΔT_m 的改进(表 6)。使用包含 5 和 6 个荧光标记的 T 碱基的探针产生了高质量的解链峰。其他探针设计在具有多至 8 个荧光标记碱基时产生高质量的解链峰。迄今为止尚未测试具有超过 8 个荧光团的探针,但预期在适合的探针长度和标记之间的间隔时,此类探针具有功能性。

[0207] 表 6:无碱基位点对探针 T_m 的影响

[0208]

重复	02466R	02467R	02468R	02469R
5			47.2°C	49.0°C
7	60.6°C	58.6°C	54.9°C	54.0°C
10	64.8°C	63.2°C	63.0°C	57.5°C
12	67.0°C	64.3°C	67.4°C	66.1°C
15	68.9°C	66.8°C	68.9°C	68.0°C
7/15 的 ΔT_m	8.3°C	8.2°C	14.0°C	14.0°C

[0209] 实施例 5

[0210] D16S539 的等位基因总是具有至少 5 个 GATA 重复。因此,如果将 STR 重复分成两个寡核苷酸,可以降低探针杂交的 T_m (表 7)。第一寡核苷酸没有荧光,作为阻断剂,其包含前面 5 个共有重复。第二寡核苷酸是仅与其他重复杂交的荧光探针,即仅检测 10 个重复的等位基因中的 5 个重复(图 6A)。阻断剂的目的是减少荧光探针可利用的 STR 重复的数目,从而提高长度类似的等位基因之间的 ΔT_m 。

[0211] 阻断寡核苷酸(B1)由 5 个 d(TATC) 重复、用于防止滑移的 31-mer 锚定序列(其与侧翼序列完全地互补)和用于防止探针在 PCR 期间延伸的 3' - 磷酸组成。荧光寡核苷酸探针(FL1)具有 10 个 d(TATC) 重复、5 个内部荧光素 dT 碱基(Glen Research)、6 个核苷酸的短锚(其与另一个侧翼序列完全地互补)和位于 3' - 端的辛二醇 PCR 阻断剂(表 7)。

[0212] 表 7:使用双分子阻断剂和寡核苷酸探针的 STR 分析,其中 P、5 和 6T 分别代表 3' - 磷酸、荧光素 dT 和辛二醇。

[0213]

寡核苷酸	SEQ ID	序列
B1	35	TATCTATCTATCTATCTATCCACCTGTCTGTCTGTCTGTATTGATCTAGGGP
FL1	36	5ATCTATC5ATCTATC5ATCTATC5ATCTATC5ATCTATC6T

[0214] 使用摩尔过量的阻断寡核苷酸有利于与 5 个共有重复的杂交。使用 100nM FL1 探针和 600nM B1 阻断剂分析具有 8、9、10、11 和 12 个 GATA 重复的互补寡核苷酸 (图 6B)。表 8 中列出了对每个靶的探针 T_m 和 ΔT_m 。将 STR 探针分成两个组分的设计显著改进了 ΔT_m 以及区分长度类似的等位基因的能力。例如,在使用全长 HYBSTR 探针进行分析时,10/11 基因型的 ΔT_m 为 0.92°C,但使用此双分子阻断策略,可将 ΔT_m 提高至 2.2°C。

[0215] 检测 5 和 7 个重复的等位基因可能需要省略反应中的阻断寡核苷酸。图 6C 说明了在没有 B1 阻断剂时,使用 FL1 探针产生的解链峰。需要平行进行具有和没有阻断寡核苷酸的分析,以检测和鉴定所有范围的 D16S539 等位基因。

[0216] 表 8:在双分子阻断剂 B1 的存在下,由 FL1 探针与寡核苷酸靶序列的杂交获得的探针 T_m 和 ΔT_m 。

[0217]

重复组合	T_m & (ΔT_m)
8/8	40.0°C
9/9	47.5°C
10/10	51.8°C
11/11	54.0°C
12/12	57.0°C
8/9	(7.5°C)
8/10	(11.8°C)
8/11	(14.0°C)
8/12	(17.0°C)
9/10	(4.3°C)
9/11	(6.5°C)
9/12	(9.5°C)
10/11	(2.2°C)
10/12	(5.2°C)
11/12	(3.0°C)

[0218] 实施例 6

[0219] 使用引物 STRF2 和 STRR2 从提取的 DNA 扩增 D16S539 靶序列,并直接从唾液样品

扩增 D16S539 靶序列。需要优化 B1 阻断剂和 FL1 探针寡核苷酸的相对浓度,从而能够有效地扩增靶,同时防止探针与具有不到 15 个重复的靶全长杂交。通过分析阻断寡核苷酸和探针寡核苷酸的各种浓度和比例完成优化。发现阻断寡核苷酸与探针寡核苷酸之间的比例为 6 : 1 是有用的。B1 阻断剂不足导致在约 59°C 时产生常规峰,其中 FL1 探针没有被完全阻断,从而可以与全部 10 个重复杂交。反之,过量的 B1 阻断剂抑制靶的扩增,增加了检测靶所需的循环数(图 7A)。发现最佳浓度为使用 37.5nM FL1 探针和 225nM B1 阻断剂。使用具有 11/11、10/11、12/12、10/13、11/12、13/13、12/13 和 9/12 基因型的提取 DNA 和唾液样品进行分析。对应于 9、10、11、12 和 13 个重复的解链峰 T_m 分别为 47.0°C、51.5°C、53.5°C、56.0°C 和 57.5°C。9/12 基因型的样品是唯一获得两个清晰峰的解链图谱(图 7B)。使用任何样品都不能有效地检测 10 个重复的峰。与 11 和 13 个重复相比,10 个重复的峰具有降低的高度。小 ΔT_m 意味着 10 个重复的峰被隐藏在 11 和 13 个重复的印迹中,从而导致连肩峰也观察不到。如 FL4 所示提高 ΔT_m 或者试着标准化或改进较短重复的峰高度应克服这个问题。11/12 和 12/13 杂合样品产生 T_m 介于纯合峰之间的单个解链峰(即分别约 54.5°C 和 57.0°C)。HRM 仪器的应用应当有助于使用这些中间峰 T_m 鉴定杂合体。

[0220] 通过加入修饰碱基丙炔基 dU 和丙炔基 dC,以及通过加入 G- 钳和 5' - 三甲氧基芪修饰提高了阻断剂的稳定性。

[0221] 实施例 7

[0222] 将 STRF2 正向引物和 B1 阻断剂序列组合到单条非荧光寡核苷酸中,使用六甘醇(HEG)作为连接子和 PCR 阻断剂。单分子引物/阻断剂寡核苷酸(BP1)包含 5 个 d(TATC)重复、用于防止滑移的 10-mer 锚定序列、HEG 和正向引物(表 9)。

[0223] 表 9:使用单分子阻断剂/引物和寡核苷酸探针的 STR 分析,其中 P、5、6T 和 (HEG) 分别代表 3' - 磷酸、荧光素 dT、辛二醇和六甘醇。

[0224]

寡核苷酸	SEQ ID	序列
BP1	37	TATCTATCTATCTATCTATCCACCTGTCTG (HEG) GATCCCAAGCTCTTCCTCTT
FL1	38	CAATGA5ATCTATC5ATCTATC5ATCTATC5ATCTATC5ATCTATC6T
FL2	39	CAATGATATC5ATCTATC5ATCTATCTATCTATCTATC5ATCTATCP
FL3	40	CAATGA5ATCTA5CTATC5ATCTA5CTATC5ATCTA5CTATC5ATCP
FL4	41	CAATGATATC5ATCTA5CTATCTATCTATCTATCTATC5ATCTATCP
FL5	42	CAATGATATCTA5CTATCTA5CTATCTA5CTATCTA5CTATCTATC6T

[0225] 此方法的优点是阻断剂和靶序列在扩增后成为同一 DNA 链的一部分。在热力学上,阻断剂和靶之间的单分子相互作用比探针杂交更为有利。此外,由于阻断剂在靶序列扩增后才能杂交,所以 PCR 效率没有损失(图 8)。

[0226] 使用 0.5 μ M BP1 和 0.05 μ M STRR2 反向引物从提取的 DNA 和直接从唾液样品扩增 D16S539 靶序列。使用 50nM 的 FL1 探针检测并鉴定从 11/11、12/12、13/13、10/11、10/13、11/12、12/13、9/12、11/13 和 8/13 基因型的样品扩增的靶序列。实时 LightCycler 荧光数据证明靶扩增和检测的效率高于使用双分子阻断剂方法获得的效率(图 9A)。另外,还使用 384 孔 PCR 板进行靶扩增,随后使用 LightTyper 仪器进行解链分析。对应于 8、9、11、12 和 13 个重复的解链峰的 T_m 分别为 40.0°C、46.0°C、53.0°C、55.5°C 和 57.5°C(图 9)。在这些样品的任一个中都没能可靠地检测到 10 个重复的峰,这是因为由于 10 个重复的峰的小 ΔT_m 和降低的峰高度使其被隐藏在 11 和 13 个重复的印迹中。这可以通过上文讨论的方式克服。8/13 和 9/12 基因型的样品产生了代表组分等位基因的两个清晰解链峰(图 9C-D)。另一方面,11/12、12/13 基因型的杂合样品产生 T_m 介于纯合等位基因 T_m 之间的单个解链峰(即约 54.5°C 和 56.0°C,分别显示在图 9F-G 中)。

[0227] 使用 BP1 和寡核苷酸探针 FL2、FL3、FL4 和 FL5 重复对 DNA 和唾液样品的分析,其中上述探针具有不同数目的荧光团标记碱基和在荧光团之间不同的间隔子(表 9)。由扩增的 STR 靶产生的解链峰 T_m 显示在表 10 中。与 FL1 探针类似,FL3 和 FL5 构建物不能可靠地检测杂合基因型例如 10/11 和 10/13 中具有 10 个重复的等位基因。荧光标记碱基的数目和间隔影响探针的信噪比和解链峰质量。探针 FL2 和 FL4 仅具有 3 个荧光团标记的碱基,并且能够检测 10/11 和 10/13 基因型中的 10 个重复的等位基因(图 10D-E)。线性的未修饰 HYBSTR 探针对 10 和 11 个重复的等位基因显示出 0.92°C 的 T_m 。使用单分子阻断剂 BP1 和 FL4HyBeacon 获得了 3.5°C 的 ΔT_m 。

[0228] 表 10:使用单分子 BP1 阻断剂/引物从提取的 DNA 和未提取的唾液样品扩增出的靶产生的解链峰的 T_m s。

[0229]

重复	峰 T_m				
	FL1	FL2	FL3	FL4	FL5
8	40.0	42.5	39.0	41.0	42.5
9	46.0	48.0	44.5	47.5	47.0
10		52.5		52.5	
11	53.0	55.5	51.0	56.0	53.5
12	55.5	58.0	52.5	58.0	55.5
13	57.5	60.0	54.5	60.0	57.0

[0230] 实施例 8

[0231] 使用高分辨率解链曲线仪器(HR-1, Idaho Technology Inc, Utah, USA)重新分析包含扩增的 D16S539 靶、BP1 阻断剂和 F6A 寡核苷酸的 LightCycler 毛细管。HR-1 能够清楚地区分由 11、12 和 13 个重复序列产生的解链峰(图 11A)。然而,如上所述,解链峰不能用于鉴定杂合基因型,例如 11/12。高分辨率解链曲线软件的确能够清楚地区分 9/12、11/11、

11/12、12/12、12/13 和 13/13 基因型 (图 11B-C)。高分辨率分析在区分长度类似的等位基因时不需要超过 3°C 的 ΔT_m 。此外,改进的数据分析方法能够将宽峰和峰肩清楚地分成组分等位基因 (图 11D)。

[0232] 实施例 9

[0233] 重复阻断方法的另一个实施方式是将探针锚定至阻断剂而不是锚定至侧翼于重复靶的 DNA 序列。此方法提高了在阻断剂存在下探针杂交的稳定性,从而提高了阻断效率。在双分子形式中,阻断寡核苷酸可包含指定数目的序列重复、用于防止滑移的适合的锚定序列和富含 GC 的钳 (图 12)。阻断寡核苷酸的钳与探针寡核苷酸中的互补钳杂交。在单分子形式中,可通过适合的间隔子 /PCR 阻断剂 (例如 HEG) 将阻断重复、锚定序列和富含 GC 的钳连接到 PCR 引物之一上。用于这些阻断剂的荧光探针可具有与阻断剂中富含 GC 的序列互补的富含 GC 的序列。当阻断剂与邻近的探针寡核苷酸杂交时,富含 GC 的钳将形成接头 (junction)。对于 D16S539 的情况,当靶重复数目小于 15 时,可防止探针全长杂交 (图 12A-B)。

[0234] 分别将具有 5 个 D16 重复 (5LSB1) 和 7 个重复 (7LSB2) 的阻断寡核苷酸与具有 10 个重复 (10SFL1) 和 8 个重复 (8SFL2) 的荧光探针组合。使这些接头探针和阻断剂 (表 11) 与具有 8-15 个重复的互补靶寡核苷酸 (表 5) 杂交,并使用 LightCycler 仪器分析 (图 12C-D)。表 12 中列出了探针的 T_m 。由于探针长度减小,7LSB2 的 ΔT_m 增大。在没有阻断剂分子时,可分析包含 9 个或更少重复的等位基因。

[0235] 表 11:接头探针和阻断剂的序列,其中 6 为吡咯-脱氧胞嘧啶 (pdC),5 为荧光素 dT, P 为 3' - 磷酸,而 (HEG) 为六甘醇。

[0236]

寡核苷酸	SEQ ID	序列
5LSB1	43	GCGGCTATCTATCTATCTATCTATCCACCTGTCTGTCTGTCTGTA (HEG) GA TCCAAGCTCTTCCTCTT
7LSB2	44	GCGGCTATCTATCTATCTATCTATCTATCTATCCACCTGTCTGTCTGTCTGT ATTGATCTAGGGP
10SFL1	45	TATCTATCTA56TATCTATCTA56TATCTAT65ATCTATCGCCGCP
8SFL2	46	TATCTAT65ATCTAT65ATCTAT65ATCTATCGCCGCP

[0237] 表 12:使用寡核苷酸同源物由接头探针和阻断剂产生的解链峰的 T_m 。

[0238]

重复	T_m 5SBP1/10SFL1	T_m 7LSB2/8SFL2
8	36.1°C	

9	44.6°C	
10	51.3°C	34.9°C
11	55.2°C	44.4°C
12	57.6°C	50.5°C
13	60.2°C	55.0°C
14	61.4°C	57.6°C
15	62.1°C	59.7°C

[0239] 作为上述实施方式的变式,单分子阻断剂(PCR引物)可具有长度和/或组成不同的钳部分(A1、A1、A3等),探针寡核苷酸可包含与PCR引物上钳部分互补的钳部分(A1'、A2'、A3'等)。这样,例如在PCR分析中可包含3种或更多探针,并可将钳部分的序列用于修饰峰 T_m 。图15中给出了此变式的图示说明。

[0240] 实施例10

[0241] 研究了第二个STR基因座,以证明通过解链曲线分析来检测并区分重复序列并非仅适用于D16S539。TH01基因座包含(AATG)_n的重复,并且报道了具有3-14个重复的等位基因(基因序列参见图13)。TH01序列可包含部分重复,例如在9.3等位基因(AATG)₆ATG(AATG)₃(SEQ ID NO. :47)中。在设计验证原理的TH01分析时,仅考虑最常见的重复等位基因,其中99.5%的等位基因包含6个重复(23%)、7个重复(24.1%)、8个重复(11.6%)、9个重复(20%)、9.3个重复(17.9%)或10个重复(2.9%)。TH01等位基因分析的其他挑战是不仅要区分长度相差4个碱基的一倍或多倍的完整重复等位基因,而且还要区分长度仅相差单个核苷酸的9.3个部分重复和10个完整重复的等位基因。

[0242] 设计探针HYBTH01以用于检测和区分TH01等位基因。此探针包含6.1个AATG重复(表13)和5'-TGGFG锚定序列。使用相隔7个核苷酸的两个荧光素dT染料标记探针,其中一个标记位于锚定序列中。使用图6所示的双分子阻断策略。设计双分子阻断剂TH01_BL以防止HYBTH01探针在靶等位基因内的重复数目小于10时与靶序列完全杂交。TH01_BL阻断剂包含3.3个AATG重复和3'-AGGGAAATAAGGG锚定序列(表13)。

[0243] 使用引物TH01_F和TH01_R(表13)从纯化DNA样品以及未纯化的唾液扩增TH01靶序列。使用不对称PCR增强探针与扩增靶序列杂交的效率,其中引物TH01_F和TH01_R的浓度分别为0.1 μM和1 μM。所使用的探针和阻断寡核苷酸的浓度分别为75nM和375nM。使用5倍摩尔过量的阻断寡核苷酸以防止探针与TH01等位基因共有的3.1个AATG重复杂交。阻断寡核苷酸限制了可用于探针杂交的靶序列的量,有效地将6-10个重复的等位基因的长度缩短至2.1-6.1个重复,从而改进了基于解链峰 T_m 区分TH01等位基因的能力。

[0244] 表13:TH01重复的分析,其中5和P分别代表荧光素dT和3'-磷酸。

[0245]

寡核苷酸	SEQ ID	序列

HYBTH01	48	TGG5GAATGAA5GAATGAATGAATGAATGAP
TH01_BL	49	ATGAATGAATGAATGAGGGAAATAAGGGP
TH01_F	50	GGCTTCCGAGTGCAGGTCA
TH01_R	51	GGTGATTCCCATTTGGCCTG

[0246] HYBTH01 探针和 TH01_BL 阻断剂的组合能够可靠地检测并鉴定扩增的 8、9、9.3 和 10 个重复的等位基因, 显示出分别为约 53°C、57°C、60.5°C 和 61.5°C 的解链峰 T_m (图 14)。由于 6 和 7 个重复的等位基因的解链峰太低, 不能可靠地通过这一设计用于检测在此基因座内更具挑战性的较长重复序列的具体探针结构形式进行检测。显而易见地, 通过减少阻断剂中重复的数目, 增加探针中重复的数目或提高探针的 T_m (通过增加锚定序列的长度, 或通过加入 DNA 碱基类似物或亚磷酰胺帽, 例如 5' - 三甲氧基芪) 能够应用本发明所述的相同原理检测 6 和 7 个重复的等位基因。与此类似, 预期在使用包含通过六甘醇 (HEG) 修饰分隔的 TH01_BL 阻断剂和 TH01_R 引物序列的长寡核苷酸的单分子模式中, 该 TH01 分析也有效地发挥功能。

[0247] 结论

[0248] 可使用寡核苷酸探针分析重复序列, 其中通过靶长度和杂交探针的比例确定重复的数目。需要超过 60 个核苷酸的长寡核苷酸探针分析短串联重复。长重复之间的探针 T_m 差异通常小, 妨碍了某些等位基因的可靠鉴定。可通过使探针不稳定化或通过将重复序列分隔成长度减小的非荧光阻断剂和荧光探针来提高 ΔT_m 。本文描述的各种阻断策略均具有增加的探针 ΔT_m , 以及区分具有大量重复的 D16S539 和 TH01 等位基因的能力。本发明的实施例可延伸至其他 STR, 包括但不限于 D19S433 和 D18S51。

[0249] 实施例 11

[0250] 预防探针与阻断剂的交叉杂交

[0251] D8S1179 和 D3S1358 基因座均包含 (TCTA)_n 重复序列。探针和阻断剂可能交叉杂交, 妨碍有效的靶检测或导致错误的结果。将阻断剂与引物相连 (即单分子阻断) 防止了交叉杂交。将探针与单分子构建物相连防止了对错误靶等位基因的检测。出于杂交稳定性的目的使用将阻断剂和探针组分分成两个单独部分的寡核苷酸间隔子来实现这一点 (参见图 16)。

[0252] 实施例 12

[0253] 将多个阻断剂包含在单个管中, 利用解链峰 T_m 的移动来检测其他重复等位基因。这通过使用部分重复 (偏置) 阻断剂而实现, 该阻断剂允许 (长度 / 序列) 不等的探针杂交。

[0254] 尽管已报道 D8S1179 等位基因包含 TCTA 和 TCTG 元件, 但 TCTG 重复仅限于具有 13 个或更多个重复的等位基因并且仅位于第二、第三和第四个重复位置 (GenBank 登录号 AF216671)。因此, 可变的 TCTA 和 TCTG 重复位于阻断寡核苷酸中, 仅留下 TCTA 重复可与探针杂交。已报道 D8S1179 基因座具有 6-25 个四核苷酸重复, 然而, 在设计 D8S1179 分析时, 仅考虑了最常见的重复等位基因, 其中 99.9% 的等位基因包含 8 个重复 (0.9%)、9 个重复

(0.8%)、10个重复(8.8%)、11个重复(7.3%)、12个重复(12.6%)、13个重复(27.6%)、14个重复(22.6%)、15个重复(14.1%)、16个重复(4.3%)或17个重复(0.9%)。

[0255] 设计探针 HYBD8 以用于检测和区分 D8S1179 等位基因。此探针包含 8 个 TCTA 重复(表 14)和 3' -FTCCCCP 锚定序列。使用相隔 5 个核苷酸的两个荧光素 dT 染料标记探针,其中一个标记位于锚定序列中。使用图 6 所示的双分子阻断策略,其中使用 3 个阻断寡核苷酸来检测和区分全长的常见 D8S1179 等位基因。设计双分子阻断剂 D8BL5、D8BL8 和 D8BL11(表 14),分别用于防止 HYBD8 探针在靶等位基因内重复的数目小于 13、16 和 19 时与靶序列完全杂交。

[0256] 使用引物 D8F 和 D8R(表 14)从纯化 DNA 样品以及未纯化的唾液扩增 D8S1179 靶序列。使用不对称 PCR 增强探针与扩增靶序列杂交的效率,其中引物 D8F 和 D8R 的浓度分别为 0.1 μ M 和 1 μ M。所使用的探针和阻断剂寡核苷酸的浓度分别为 75nM 和 375nM。使用 5 倍摩尔过量的阻断寡核苷酸以防止全长探针与不恰当的靶等位基因杂交。

[0257] 在 D8BL5 存在下的解链曲线分析能够可靠地检测并鉴定 8、9 和 10 个重复的等位基因(表 15)。D8BL8 阻断剂用于检测 11、12 和 13 个重复的等位基因,而 D8BL11 能够鉴定 14、15、16 和 17 个重复的等位基因。

[0258] 在分别与 D8BL5、D8BL8 和 D8BL11 阻断剂组合使用时, HYBD8 探针与 9 个重复、12 个重复和 15 个重复靶等位基因的杂交产生了 44.73°C、44.4°C 和 44.4°C 的解链峰 T_m 。由于 D8BL5、D8BL8 和 D8BL11 各留下与探针杂交的 3 个靶重复,预期 9、12 和 15 个重复的等位基因产生类似的解链峰 T_m 。因此,在单一试管中同时检测 9、12 和 15 个重复的等位基因时,不能将 D8BL5、D8BL8 和 D8BL11 阻断寡核苷酸一起使用。

[0259] 表 14 :D8S1179 重复的分析,其中 5、P 和 X 分别代表荧光素 dT、3' -磷酸和 3' -氨基 C7

[0260]

寡核苷酸	SEQ ID	序列
D8F	52	CGGCCTGGCAACTTATATGT
D8R	53	GCCTTAATTTATTTACCTATCCTGTAGA
HYBD8	54	TCTATCTATCTATCTATCTATCTATC5ATCTA5TCCCCP
D8BL5	55	GTATTTTCATGTGTACATTCGTA(tcta) ₅ x
D8BL8	56	GTATTTTCATGTGTACATTCGTA(tcta) ₈ x
D8BL11	57	GTATTTTCATGTGTACATTCGTA(tcta) ₁₁ x
D8BL7.2	58	GTATTTTCATGTGTACATTCGTA(tcta) ₇ tctx
D8BL10.3	59	GTATTTTCATGTGTACATTCGTA(tcta) ₁₀ tctx

[0261] 使用带有部分重复的阻断寡核苷酸可导致解链峰 T_m 的移动(“偏置”),增加可

同时检测的 STR 等位基因的数目。使用具有 11-17 个重复的扩增靶评估包含 7.2 个 (即 $(TCTA)_7TC$) 和 10.3 个 (即 $(TCTA)_{10}TCT$) 重复的部分重复阻断剂。相对于 D8BL8 和 D8BL11, 减少阻断剂的长度能够使更长的探针杂交, 从而提高了解链峰 T_m (表 15)。例如, 在 D8BL7.2 阻断剂存在下, 探针与 12 个重复的靶的杂交产生 46.88°C 的解链峰 T_m , 这样, D8BL5 和 D8BL7.2 阻断剂可用于同时检测 9 和 12 个重复的等位基因 (图 18)。图 19 说明使用 D8BL7.2 和 D8BL10.3 阻断剂在单一试管中同时检测 11 和 15 个重复的等位基因。

[0262] 表 15

[0263]

阻断剂	靶	T_m	偏置阻断剂	靶	预期 T_m	T_m	ΔT_m
D8BL5	8	37.12					
D8BL5	9	44.73					
D8BL5	10	49.36					
D8BL5	11	52.67					
D8BL8	11	37.12	D8BL7.2	11	40.76	40.43	3.31
D8BL8	12	44.4	D8BL7.2	12	46.88	46.38	1.98
D8BL8	13	49.36	D8BL7.2	13	51.02	50.68	1.32
D8BL8	14	52.67	D8BL7.2	14	53.83	53.66	0.99
D8BL11	14	36.79	D8BL10.3	14	39.33	39.1	2.31
D8BL11	15	44.4	D8BL10.3	15	46.27	45.72	1.32
D8BL11	16		D8BL10.3	16	50.89		
D8BL11	17	52.67	D8BL10.3	17	53.44	53.66	0.99

[0264] 适合的染料列表

[0265] 表 14 : 用于寡核苷酸标记的通用荧光染料

[0266]

染料	λ - 激发	λ - 发射	颜色
荧光素	494nm	525nm	绿色
四氯荧光素 TET	521nm	536nm	橙色
JOE	525nm	555nm	绿色

Yakima Yellow	530nm	549nm	黄色
六氯荧光素 HEX	535nm	556nm	粉色
Cy3(也称作 Quasar 570)	546nm	563nm	红色
5-TAMRA	541nm	568nm	玫瑰色
6-TAMRA	547nm	573nm	玫瑰色
Redmond Red	579nm	595nm	红色
Cy3. 5	588nm	604nm	紫色
ROX	585nm	610nm	红色
Pulsar 650	490nm	650nm	紫色
Cy5(也称作 670)	646nm	662nm	蓝紫色
Cy5. 5	683nm	707nm	深蓝色

[0267] 表 15 :Alexa 染料 (Invitrogen)

[0268]

Alexa 染料	λ - 激发	λ - 发射
Alexafluor 350	350nm	442nm
Alexafluor 405	405nm	421nm
Alexafluor 430	430nm	540nm
Alexafluor 488	488nm	518nm
Alexafluor 500	502nm	524nm
Alexafluor 514	518nm	542nm
Alexafluor 532	534nm	553nm
Alexafluor 546	546nm	565nm
Alexafluor 555	552nm	567nm
Alexafluor 568	578nm	603nm

Alexafluor 594	591nm	618nm
Alexafluor 610	612nm	628nm
Alexafluor 633	633nm	650nm
Alexafluor 647	647nm	662nm
Alexafluor 660	663nm	690nm
Alexafluor 680	679nm	702nm
Alexafluor 700	696nm	719nm
Alexafluor 750	752nm	779nm

[0269] 表 16 :ATTO 染料 (ATTO-TEC GmbH)

[0270]

ATTO 染料	λ - 激发	λ - 发射
ATTO 425	436nm	484nm
ATTO 465	453nm	508nm
ATTO 488	501nm	523nm
ATTO 495	495nm	527nm
ATTO 520	525nm	545nm
ATTO 532	532nm	553nm
ATTO 550	554nm	576nm
ATTO 565	563nm	592nm
ATTO 590	594nm	624nm
ATTO 610	615nm	634nm
ATTO 620	619nm	643nm
ATTO 635	635nm	659nm
ATTO 647	645nm	669nm

ATTO 655	633nm	684nm
ATTO 680	680nm	700nm
ATTO 700	700nm	719nm

[0272] 表 17 :Dyomics 染料 (Dyomics GmbH)

[0273]

Dyomics 染料	λ - 激发	λ - 发射
DY415	418nm	465nm
DY495	495nm	520nm
DY505	505nm	530nm
DY547	557nm	574nm
DY548/549	558nm	572nm
DY550	553nm	578nm
DY555	555nm	580nm
DY556	548nm	573nm
DY560	559nm	578nm
DY590	580nm	599nm
DY610	609nm	629nm
DY615	62nm	641nm
DY630	636nm	657nm
DY631	637nm	658nm
DY632/633/634	637nm	657nm
DY635	647nm	671nm
DY636	645nm	671nm
DY647	652nm	673nm

DY648	653nm	674nm
DY650	653nm	674nm
DY651	653nm	678nm
DY652	654nm	675nm
DY675/676	674nm	699nm
DY677	673nm	694nm
DY680/682	690nm	709nm
DY700	702nm	723nm
DY701	706nm	731nm
DY730	734nm	750nm
DY731/734	736nm	759nm
DY732	736nm	759nm
DY750	747nm	776nm
DY751	751nm	779nm
DY752	748nm	772nm
DY776	771nm	801nm
DY781	783nm	800nm
DY782	782nm	800nm

[0275] 表 18 :Dyomics Megastokes 染料 (Dyomics GmbH)。均能在 488nm 激发。

[0276]

Dyomics 染料	λ - 激发	λ - 发射
DY475XL	493nm	514nm
DY480XL	500nm	630nm
DY485XL	485nm	560nm

DY500XL	505nm	555nm
DY510XL	509nm	590nm
DY600XL	603nm	634nm
DY520XL	520nm	664nm

[0277] 表 19 :Hilyte 染料 (Cambridge Bioscience)

[0278]

染料	λ - 激发	λ - 发射
Hilyte Fluor 488	502nm	527nm
Hilyte Fluor 555	552nm	569nm
Hilyte Fluor 647	649nm	674nm
Hilyte Fluor 680	678nm	699nm

[0279] 表 20 :低激发波长 (UV) 荧光团

[0280]

衍生物	Abs [*]	Em [*]
Alexa Fluor 350	346	442
Alexa Fluor 405	402	412
苯氨基萘	326	462
Bimane	375	456
丹磺酰	328	563
Dapoxyl	374	572
3,5- 双溴甲基 -2,6- 二甲基吡 唑并 [1,2-a] 吡唑 -1,7- 二酮 (Dibromobimane)	394	490
二乙基氨基香豆素	384	470
二甲基氨基香豆素	376	465
二甲基氨基萘	391	500

Monobromobimane	394	490
Monochlorobimane	394	490
萘	336	490
芘	339	384
芪	329	408

[0281] 表 21 : 常见 STR

[0282]

基因座名称	染色体定位	共有序列基序	报道的等位基因
D3S1358	3p	TCTA (TCTG) ₁₋₃ (TCTA) _n (SEQ ID NO :60-62)	9, 10, 11, 12, 13, 14, 15, 15. 2, 16, 16. 2, 17, 17. 1, 18, 19, 20
vWA	12p12-pter	TCTA (TCTG) ₃₋₄ (TCTA) _n (SEQ ID NO :63-64)	10, 11, 12, 13, 14, 15, 15. 2, 16, 17, 18, 18. 2, 18. 2, 19, 19. 2, 20, 21, 22, 23, 24, 25
D16S539	16q24-qter	(AGAT) _n	5, 8, 9, 10, 11, 12, 13, 14, 15
D2S1338	2q35-37. 1	(TGCC) _n (TTCC) _n (SEQ ID NO :65)	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28
釉原蛋白 (Amelogenin)	X : p22. 1-22. 3 Y :p11. 2	—	X, Y
D8S1179 ^a	8	(TCTR) _n ^b	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20. 2, 21. 2
D21S11	21q11. 2-q21	(TCTA) _n (TCTG) _n [(TCTA) ₃ TA (TCTA) ₃ TCA (SEQ ID NO :66)	24, 24. 2, 25, 25. 2, 26, 26. 1, 27, 27. 2, 27. 3,
		(TCTA) ₂ TCCA TA] (TCTA) _n (SEQ ID NO :67)	28, 28. 2, 28. 3, 29, 29. 1, 29. 2, 29. 3, 30, 30. 1, 30. 2, 31, 31. 1, 31. 2, 31. 3, 32, 32. 1, 32. 2, 32. 2, 33, 33. 3, 34, 34. 2, 34. 3, 35, 35. 2, 35. 3, 36, 36. 2, 36. 3, 37, 37. 2, 38, 38. 2
D18S51	18q21. 3	(AGAA) _n	7, 8, 9, 9. 2, 10, 10. 2, 11, 12, 12. 2, 13, 13. 2, 14, 14. 2, 15, 15. 1, 15. 2, 15. 3, 16, 16. 2, 16. 3, 17, 17. 1, 17. 3, 18, 18. 1, 18. 2, 19, 19. 2, 20, 20. 1, 20. 2, 21, 21. 2, 22, 23, 23. 1, 24, 25, 26, 27

D19S433	19q12-13.1	(AAGG) ₃ (AAAG) (AAGG)(TAGG) (SEQ ID NO :68)	9,10,11,12,12.2,13,13.2,14,14.2, 15,
		(AAGG) _n	15.2,16,16.2,17,17.2,18,18.2,19
TH01	11p15.5	(AATG) _n	3,4,5,5.3,6,6.1,6.3,7,7.1,7.3,8, 8.3,9,9.3,10,10.3,11,13.3,14
FGA	4q28	(TTTC) ₃ TTTTTCT (CTTT) _n CTCC (SEQ ID NO :69)	12.2,13,15,16,16.1,16.2,17,17.1, 17.2,
		(TTCC) ₂	18,18.1,18.2,19,19.1,19.2,19.3,20, 20.1,20.2,20.3,21,21.2,22,22.1, 22.2,22.3,23,23.2,23.3,24,24.1, 24.2,24.3,25,25.1,25.2,25.3,26, 26.1,26.2,27,27.1,27.3,28,28.1,29, 29.2,30,30.2,31,31.2,32,32.2,33.2, 35.2,42.2,43.2,44.2,45.2,46.2,47.2, 48.2,49.2,50.2,51.2

[0284] a. 在一些文献中,将此基因座命名为 D6S502。

[0285] b. R 可代表 A 或 G 核苷酸。

[0286] 常用的 STR

[0287] 在英国使用的 SGM+ 基因座为：

[0288] D3S1358、VWA、D16S539、D2S1338、D8S1179、D21S11、D18S51、D19S433、TH01 FGA。

[0289] 在美国使用的 13CODIS 基因座为：

[0290] CSF1PO、FGA、TH01、TPOX、VWA、D3S1358、D5S818、D7S820、D8S1179、D13S317、D16S539、D18S51 和 D21S11。

[0291] 世界各地的实验室广泛使用一套核心 Y- 染色体 STR(Y-STR) 基因座进行个人身份鉴定和遗传谱系学。在 20 世纪 90 年代晚期从可用的 Y-STR 小集合中选择了最小的单倍型基因位点 (MHL)。MHL 包括 DYS19、DYS389I、DYS389II、DYS390、DYS391、DYS392、DYS393 和多态性多拷贝标记 DYS385。2003 年,美国 DNA 分析方法科学工作组 (Scientific Working Group on DNA Analysis Methods, SWGDAM) 的 Y- 染色体小组推荐了命名为 DYS438 和 DYS439 的另外两个 Y-STR。

[0292] 参考文献

[0293] Bart-Delabesse, E., Sarfati, J., Debeaupuis, J. P., van Leeuwen, W., van Belkum, A., Bretagne, S., Latge, J. P. (2001). Comparison of restriction fragment length polymorphism, microsatellite length polymorphism, and random amplification of polymorphic DNA analyses for fingerprinting *Aspergillus fumigatus* isolates. J. Clin. Microbiol. 39 :2683-6.

[0294] Belgrader, P., Smith, J. K., Weedn, V. W., Northrup, M. A. (1998). Rapid PCR for identity testing using a battery-powered miniature thermal cycler. J. Forensic Sci. 43 (2) :315-9.

- [0295] Bell, G.I., Selby, M. J., Rutter, W. J. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature* 295 :31-5.
- [0296] Brondani, C., Borba, T. C., Rangel, P. H. N., and Brondani, R. P. V. (1998). Determination of genetic variability of traditional varieties of Brazilian rice using microsatellite markers. *Genetics* 149 :2007-2023.
- [0297] Brondani, C., Rangel, P. H., Borba, T. C., Brondani, R. P. (2003). Transferability of microsatellite and sequence tagged site markers in *Oryza* species. *Hereditas*. 138(3) :187-92.
- [0298] Carrier, L., Hengstenberg, C., Beckmann, J. S., Guicheney, P., Dufour, C., Bercovici, J., Dausse, E., Berebby-Bertrand, I., Wisnewsky, C., Pulvenis, D. et al (1993). Mapping of a novel gene for familial hypertrophic cardiomyopathy to chromosome 11. *Nat. Genet.* 4 :311-3.
- [0299] Frayling, I. M. (1999). Microsatellite instability. *Gut* 45 :1-4.
- [0300] Goedecke, N., McKenna, B., El-Difrawy, S., Carey, L., Matsudaira, P., Ehrlich D. (2004). A high-performance multilane microdevice system designed for the DNA forensics laboratory. *Electrophoresis* 25 :1678-86.
- [0301] Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363 :558-561.
- [0302] Jeffreys AJ, Wilson V, Thein SL. (1985a). *Nature* 314 :67-73. Hypervariable ' minisatellite' regions in human DNA.
- [0303] Jeffreys AJ, Wilson V, Thein SL. *Nature*. (1985b). 316 :76-9. Individual-specific ' fingerprints' of human DNA.
- [0304] Lagally, E. T., Emrich, C. A., Mathies, R. A. (2001). Fully integrated PCR-capillary electrophoresis microsystem for DNA analysis. *Lab Chip* 1 :102-7.
- [0305] Loeb, L. A. (1994). Microsatellite instability : marker of a mutator phenotype in cancer. *Cancer Res.* 54 :5059-5063.
- [0306] Matute, D. R., Sepulveda, V. E., Quesada, L. M., Goldman, G. H., Taylor, J. W., Restrepo, A., McEwen, J. G. (2006). Microsatellite analysis of three phylogenetic species of *Paracoccidioides brasiliensis*. *J. Clin. Microbiol.* 44 :2153-7.
- [0307] McCormick, R. M., Nelson, R. J., Alonso-Amigo, M. G., Benvegna, D. J., Hooper, H. H. (1997). Microchannel electrophoretic separations of DNA in injection-molded plastic substrates. *Anal. Chem.* 69 :2626-30.
- [0308] McCouch, S. R., Teytelman, L., Xu, Y., Lobos K. B., Clare K., Walton, M., Fu, B., Maghirang R., Li, Z., Xing, Y., Zhang, Q., Kono, I., Yano M., Fjellstrom, R., DeClerk, G., Schneider, D., Cartinhour, S., Ware, D. and Stein, L. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* 9 :199-207.

- [0309] Radtkey, R., Feng, L., Muralhida, M., Duhon, M., Canter, D., DiPierro, D., Fallon, S., Tu, E., McElfresh, K., Nerenberg, M., Sosnowski, R. (2000). Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNAmicrochip. *Nucleic Acids Res.* 28 :E17.
- [0310] Roder, M. S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M. H., Leroy, P., Ganal, M. W. (1998). *Genetics* 149 :2007-23. A microsatellite map of wheat.
- [0311] Schmalzing, D., Koutny, L., Chisholm, D., Adourian, A., Matsudaira, P., Ehrlich, D. (1999). Two-color multiplexed analysis of eight short tandem repeat loci with an electrophoretic microdevice. *Anal. Biochem.* 270 :148-52.
- [0312] Shi, Y., Simpson, P. C., Scherer, J. R., Wexler, D., Skibola, C., Smith, M. T., Mathies, R. A. (1999). Radial capillary array electrophoresis microplate and scanner for high-performance nucleic acid analysis. *Anal. Chem.* 71 :5354-61.
- [0313] Sinden, R. R. (1999). Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am. J. Hum. Genet.* 64 :346-53.
- [0314] Thierfelder, L., MacRae, C., Watkins, H., Tomfohrde, J., Williams, M., McKenna, W., Bohm K., Noeske, G., Schlepper, M., Bowcock, A. et al (1993). A familial hypertrophic cardiomyopathy locus maps to chromosome 15q2. *Proc. Natl. Acad. Sci. U S A.* 90 :6270-4.
- [0315] Watkins, H., MacRae, C., Thierfelder, L., Chou, Y. H., Frenneaux, M., McKenna, W., Seidman, J. G., Seidman, C. E. (1993). A disease locus for familial hypertrophic cardiomyopathy maps to chromosome 1q3. *Nat. Genet.* 3 :333-7.
- [0316] Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature* 359 :794-801.
- [0317] Westin, L., Xu, X., Miller, C., Wang, L., Edman, C. E., Nerenberg, M. (2000). Anchored multiplex amplification on a microelectronic chip array. *Nature Biotechnol.* 18 :199-204.
- [0318] Wooley, A. T. and Mathies, R. A. (1994). *Proc. Natl. Acad. Sci. USA* 91 :11348-11352.

序列表

<110> 南安普顿大学

<110>LGC 有限公司

<120> 寡核苷酸及其应用

<130>LGCBW/P41789PC

<150>GB 0720675.8

<151>220ctober 2007

<160>71

<170>SeqWin99

<210>1

<211>16

<212>DNA

<213> 人工序列

<220>

<223>STR 重复

<400>1

tgccctgcctt ccttcc

16

<210>2

<211>40

<212>DNA

<213> 人工序列

<220>

<223>STR 重复

<400>2

tttttttttt ttttttttct ctttttttct ccttccttcc

40

<210>3

<211>61

<212>DNA

<213> 人工序列

<220>

<223>HYBSTR

<220>

<221>modified_base

<222>(11)... (11)

<223>C6FAM dU

<220>

<221>modified_base

<222>(19)... (19)

<223>C6FAM dU

<220>

<221>misc_feature

<222>(61)... (61)

<223>3 末端磷酸封闭基团

<400>3

ggtggataga nagatagana gatagataga tagatagata gatagataga tagatagata 60

n 61

<210>4

<211>50

<212>DNA

<213> 人工序列

<220>

<223>STRV2

<220>

<221>misc_feature

<222>(17)... (17)

<223> 六甘醇间隔子

<220>

<221>modified_base

<222>(24)... (24)

<223>C6FAM dU

<220>

<221>modified_base

<222>(32)... (32)

<223>C6FAM dU

<220>

<221>misc_feature

<222>(50)... (50)

<223>3 末端磷酸封闭基团

<400>4

ggtggataga tagatangat aganagatag anagatagat agatagatan 50

<210>5

<211>67

<212>DNA

<213> 人工序列

<220>

<223>STRV3

<220>

<221>misc_feature

<222>(17)... (18)

<223> 六甘醇间隔子

<220>

<221>modified_base

<222>(25)... (25)

<223>C6FAM dU

<220>

<221>modified_base

<222>(33)... (33)

<223>C6FAM dU

<220>

<221>misc_feature

<222>(67)... (67)

<223>3 末端磷酸封闭基团

<400>5

ggtggataga tagatannga taganagata ganagataga tagatagata gatagataga 60
tagatan 67

<210>6

<211>24

<212>DNA

<213> 人工序列

<220>

<223>RC5

<400>6

tatctatcta tctatctatc cacc 24

<210>7

<211>32

<212>DNA

<213> 人工序列

<220>

<223>RC7

<400>7

tatctatcta tctatctatc tatctatecca cc 32

<210>8

<211>36

<212>DNA

<213> 人工序列

<220>

<223>RC8

<400>8

tatctatcta tctatctate tatctatcta tccacc 36

<210>9

<211>44

<212>DNA

<213> 人工序列

<220>

<223>RC10

<400>9

tatctatcta tctatctate tatctatcta tctatctate cacc 44

<210>10

<211>48

<212>DNA

<213> 人工序列

<220>

<223>RC11

<400>10

tatctatcta tctatctate tatctatcta tctatctate tatccacc 48

<210>11

<211>44

<212>DNA

<213> 人工序列

<220>

<223>RC10b

<400>11

tatctgtcta tctgtctate tgtctatctg tctatctgtc cacc 44

<210>12

<211>44

<212>DNA

<213> 人工序列

<220>

<223>RC10c

<400>12

tgtctgtctg tctgtctgtc tgtctgtctg tctgtctgtc cacc 44

<210>13

<211>44

<212>DNA

<213> 人工序列

<220>

<223>RC10d

<400>13

tatctatcta tctatccct tatctatcta tctatctatc cacc 44

<210>14

<211>40

<212>DNA

<213> 人工序列

<220>

<223>RC10e

<400>14

tatctatcta tctatctatc tatctatcta tctatctatc 40

<210>15

<211>48

<212>DNA

<213> 人工序列

<220>

<223>RC11b

<400>15

tatctgtcta tctgtctatc tgtctatctg tctatctgtc tatccacc 48

<210>16

<211>48

<212>DNA

<213> 人工序列

<220>

<223>RC11c

<400>16

tgtctgtctg tctgtctgtc tgtctgtctg tctgtctgtc tgtccacc 48

<210>17

<211>48

<212>DNA

<213> 人工序列

<220>

<223>RC11d

<400>17

tatctatcta tctatctatc cccttatcta tctatctatc tatccacc 48

<210>18

<211>44

<212>DNA

<213> 人工序列

<220>

<223>RC11e

<400>18

tatctatcta tctatctatc tatctatcta tctatctatc tate 44

<210>19

<211>28

<212>DNA

<213> 人工序列

<220>

<223>STRF2

<400>19

cagatcccaa gctcttcctc ttcctag

28

<210>20

<211>31

<212>DNA

<213> 人工序列

<220>

<223>STRR2

<400>20

acgttttgtgt gtgcatctgt aagcatgtat c

31

<210>21

<211>70

<212>DNA

<213> 人工序列

<220>

<223>02466R

<220>

<221>modified_base

<222>(9)... (9)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(17)... (17)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(25)... (25)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(33)... (33)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(41)... (41)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(49)... (49)

<223> 荧光素 dT

<400>21

caatgatanc tatctancta tctanctatc tanctatcta nctatctanc tatctatcta 60

tctatccacc

70

<210>22

<211>70

<212>DNA

<213> 人工序列

<220>

<223>02467R

<220>

<221>modified_base

<222>(9)... (9)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(17)... (17)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(25)... (25)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(33)... (33)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(41)... (41)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(49)... (49)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(51)... (51)

<223> 无碱基位点

<400>22

caatgatanc tatctancta tctanctatc tanctatcta nctatctanc natctatcta 60

tctatccacc 70

<210>23

<211>64

<212>DNA

<213> 人工序列

<220>

<223>02468R

<220>

<221>modified_base

<222>(3)... (3)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(11)... (11)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(19)... (19)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(27)... (27)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(35)... (35)

<223> 荧光素 dT

<400>23

tancatatcta nctatctanc tatctancta tctanctatc tatctatcta tctatctatc 60

cacc 64

<210>24

<211>64

<212>DNA

<213> 人工序列

<220>

<223>02469R

<220>

<221>modified_base

<222>(3)... (3)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(11)... (11)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(19)... (19)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(27)... (27)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(35)... (35)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(41)... (41)

<223> 无碱基位点

<400>24

tancatctcta nctatctanc tatctancta tctanctatc natctatcta tctatctatc 60

cacc 64

<210>25

<211>61

<212>DNA

<213> 人工序列

<220>

<223>C5

<400>25

ccctagatca atacagacag acagacaggt ggatagatag atagatagat atcattgaaa 60

g

61

<210>26

<211>69

<212>DNA

<213> 人工序列

<220>

<223>C7

<400>26

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatat 60

cattgaaag 69

<210>27

<211>73

<212>DNA

<213> 人工序列

<220>

<223>C8

<400>27

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60

atatcattga aag 73

<210>28

<211>77

<212>DNA

<213> 人工序列

<220>

<223>C9

<400>28

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60

atagatatca ttgaaag 77

<210>29

<211>81

<212>DNA

<213> 人工序列

<220>

<223>C10

<400>29

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60
atagatagat atcattgaaa g 81

<210>30

<211>85

<212>DNA

<213> 人工序列

<220>

<223>C11

<400>30

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60
atagatagat agatatcatt gaaag 85

<210>31

<211>89

<212>DNA

<213> 人工序列

<220>

<223>C12

<400>31

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60
atagatagat agatagatat cattgaaag 89

<210>32

<211>93

<212>DNA

<213> 人工序列

<220>

<223>C13

<400>32

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60
atagatagat agatagatag atatcattga aag 93

<210>33

<211>97

<212>DNA

<213> 人工序列

<220>

<223>C14

<400>33

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60
atagatagat agatagatag atagatatca ttgaaag 97

<210>34

<211>101

<212>DNA

<213> 人工序列

<220>

<223>C15

<400>34

ccctagatca atacagacag acagacaggt ggatagatag atagatagat agatagatag 60
atagatagat agatagatag atagatagat atcattgaaa g 101

<210>35

<211>52

<212>DNA

<213> 人工序列

<220>

<223>B1

<220>

<221>misc_feature

<222>(52)... (52)

<223>3 末端磷酸封闭基团

<400>35

tatctatcta tctatctatc cacctgtctg tctgtctgta ttgatctagg gn

52

<210>36

<211>41

<212>DNA

<213> 人工序列

<220>

<223>FL1

<220>

<221>modified_base

<222>(1)... (1)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(9)... (9)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(17)... (17)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(25)... (25)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(33)... (33)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(41)... (41)

<223> 辛二醇封闭基团

<400>36

natctatcna tctatcnate tatenatcta tcnatctate n 41

<210>37

<211>51

<212>DNA

<213> 人工序列

<220>

<223>BP1

<220>

<221>misc_feature

<222>(31)... (31)

<223> 六甘醇间隔子

<400>37

tatctatcta tctatctate cacctgtctg ngatcccaag ctcttcctct t 51

<210>38

<211>47

<212>DNA

<213> 人工序列

<220>

<223>FL1

<220>

<221>modified_base

<222>(7)... (7)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(15)... (15)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(23)... (23)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(31)... (31)

<223> 荧光素 dT

<220>

<221>modified base

<222>(39)... (39)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(47)... (47)

<223> 辛二醇封闭基团

<400>38

caatganatc tatcnatcta tcnatctatc natctatcna tctatcn 47

<210>39

<211>47

<212>DNA

<213> 人工序列

<220>

<223>FL2

<220>

<221>modified_base

<222>(11)... (11)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(19)... (19)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(39)... (39)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(47)... (47)

<223>3 末端磷酸封闭基团

<400>39

caatgatatc natctatcna tctatctatc tatctatcna tctatcn 47

<210>40

<211>47

<212>DNA

<213> 人工序列

<220>

<223>FL3

<220>

<221>modified_base

<222>(7)... (7)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(13)... (13)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(19)... (19)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(25)... (25)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(31)... (31)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(37)... (37)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(43)... (43)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(47)... (47)

<223>3 末端磷酸封闭基团

<400>40

caatganatc tancatcna tctancatc natctancta tcnatcn 47

<210>41

<211>47

<212>DNA

<213> 人工序列

<220>

<223>FL4

<220>

<221>modified_base

<222>(11)... (11)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(17)... (17)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(39)... (39)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(47)... (47)

<223>3 末端磷酸封闭基团

<400>41

caatgatatc natctancta tctatctatc tatctatcna tctatcn 47

<210>42

<211>47

<212>DNA

<213> 人工序列

<220>

<223>FL5

<220>

<221>modified_base

<222>(13)... (13)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(21)... (21)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(29)... (29)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(37)... (37)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(47)... (47)

<223> 辛二醇封闭基团

<400>42

caatgatatc tancatctcta nctatctanc tatctancta tctatcn

47

<210>43

<211>66

<212>DNA

<213> 人工序列

<220>

<223>5LSB1

<220>

<221>misc_feature

<222>(46)... (46)

<223> 六甘醇间隔子

<400>43

gcggctatct atctatctat ctatccacct gtctgtctgt ctgtangatc ccaagctctt 60

cctctt

66

<210>44

<211>65

<212>DNA

<213> 人工序列

<220>

<223>7LSB2

<220>

<221>misc_feature

<222>(65)... (65)

<223>3 末端磷酸封闭基团

<400>44

gcggctatct atctatctat ctatctatct atccacctgt ctgtctgtct gtattgatct 60

agggn 65

<210>45

<211>46

<212>DNA

<213> 人工序列

<220>

<223>10SFL1

<220>

<221>modified_base

<222>(11)... (11)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(12)... (12)

<223> 吡咯-脱氧胞嘧啶

<220>

<221>modified_base

<222>(23)... (23)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(24)... (24)

<223> 吡咯 - 脱氧胞嘧啶

<220>

<221>modified_base

<222>(32)... (32)

<223> 吡咯 - 脱氧胞嘧啶

<220>

<221>modified_base

<222>(33)... (33)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(46)... (46)

<223>3 末端磷酸封闭基团

<400>45

tatctatcta nntatctatc tanntatcta tnnatctatc gccgcn 46

<210>46

<211>38

<212>DNA

<213> 人工序列

<220>

<223>8SFL2

<220>

<221>modified_base

<222>(8)... (8)

<223> 吡咯 - 脱氧胞嘧啶

<220>

<221>modified_base

<222>(9)... (9)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(16)... (16)
<223> 吡咯 - 脱氧胞嘧啶

<220>
<221>modified_base
<222>(17)... (17)
<223> 荧光素 dT

<220>
<221>modified_base
<222>(24)... (24)
<223> 吡咯 - 脱氧胞嘧啶

<220>
<221>modified_base
<222>(25)... (25)
<223> 荧光素 dT

<220>
<221>misc_feature
<222>(38)... (38)
<223>3 末端磷酸封闭基团

<400>46
tatctatnna tctatnnatc tatnnatcta tcgccgcn 38

<210>47
<211>39
<212>DNA
<213> 人工序列

<220>
<223>TH019.3 等位基因

<400>47
aatgaatgaa tgaatgaatg aatgatgaat gaatgaatg 39

<210>48
<211>31

<212>DNA

<213> 人工序列

<220>

<223>HYBTH01

<220>

<221>modified_base

<222>(4)... (4)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(12)... (12)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(31)... (31)

<223>3 末端磷酸封闭基团

<400>48

tggngaatga angaatgaat gaatgaatga n 31

<210>49

<211>29

<212>DNA

<213> 人工序列

<220>

<223>TH01_BL

<220>

<221>misc_feature

<222>(29)... (29)

<223>3 末端磷酸封闭基团

<400>49

atgaatgaat gaatgaggga aataagggn 29

<210>50

<211>19

<212>DNA

<213> 人工序列

<220>

<223>TH01_F

<400>50

ggcttccgag tgcaggtca

19

<210>51

<211>19

<212>DNA

<213> 人工序列

<220>

<223>TH01_R

<400>51

ggtgattccc attggcctg

19

<210>52

<211>20

<212>DNA

<213> 人工序列

<220>

<223>D8F

<400>52

cggcctggca acttatatgt

20

<210>53

<211>28

<212>DNA

<213> 人工序列

<220>

<223>D8R

<400>53

gccttaattt atttacctat cctgtaga

28

<210>54

<211>39

<212>DNA

<213> 人工序列

<220>

<223>HYBD8

<220>

<221>modified_base

<222>(27)... (27)

<223> 荧光素 dT

<220>

<221>modified_base

<222>(33)... (33)

<223> 荧光素 dT

<220>

<221>misc_feature

<222>(39)... (39)

<223>3 末端磷酸封闭基团

<400>54

tctatctatc tatctatcta tctatcnatc tantccccc

39

<210>55

<211>43

<212>DNA

<213> 人工序列

<220>

<223>D8BL5

<220>

<221>misc_feature

<222>(43)... (43)

<223>3 末端氨基 C7 封闭基团

<400>55

gtatttcacg tgcacattcg tatctatcta tctatctatc tan

43

<210>56

<211>55

<212>DNA

<213> 人工序列

<220>

<223>D8BL8

<220>

<221>misc_feature

<222>(55)... (55)

<223>3 末端氨基 C7 封闭基团

<400>56

gtatttcacg tgcacattcg tatctatcta tctatctatc tatctatcta tctan

55

<210>57

<211>67

<212>DNA

<213> 人工序列

<220>

<223>D8BL11

<220>

<221>misc_feature

<222>(67)... (67)

<223>3 末端氨基 C7 封闭基团

<400>57

gtatttcacg tgcacattcg tatctatcta tctatctatc tatctatcta tctatctatc

60

tatctan

67

<210>58

<211>53

<212>DNA

<213> 人工序列

<220>

<223>D8BL7.2

<220>

<221>misc_feature

<222>(53)... (53)

<223>3 末端氨基 C7 封闭基团

<400>58

gtatttcacg tgtacattcg tatctatcta tctatctatc tatctatcta tcn

53

<210>59

<211>66

<212>DNA

<213> 人工序列

<220>

<223>D8BL10.3

<220>

<221>misc_feature

<222>(66)... (66)

<223>3 末端氨基 C7 封闭基团

<400>59

gtatttcacg tgtacattcg tatctatcta tctatctatc tatctatcta tctatctatc
tatctn

60

66

<210>60

<211>16

<212>DNA

<213> 人工序列

<220>

<223>D3S1358-1

<400>60

tctatctgtc tatcta

16

<210>61

<211>20

<212>DNA

<213> 人工序列

<220>

<223>D3S1358-2

<400>61

tctatctgtc tgtctatcta

20

<210>62

<211>24

<212>DNA

<213> 人工序列

<220>

<223>D3S1358-3

<400>62

tctatctgtc tgtctgtcta tcta

24

<210>63

<211>24

<212>DNA

<213> 人工序列

<220>

<223>vWA-3

<400>63

tctatctgtc tgtctgtcta tcta

24

<210>64

<211>28

<212>DNA

<213> 人工序列

<220>

<223>vWA-4

<400>64

tctatctgtc tgtctgtctg tctatcta

28

<210>65

<211>16

<212>DNA

<213> 人工序列

<220>

<223>D2S1338

<400>65

tgccctgcctt ccttcc

16

<210>66

<211>45

<212>DNA

<213> 人工序列

<220>

<223>D21S11

<400>66

tctatctatc tgtctgtcta tctatctata tctatctatc tatca

45

<210>67

<211>22

<212>DNA

<213> 人工序列

<220>

<223>D21S11-2

<400>67

tctatctatc catatctatc ta

22

<210>68

<211>16

<212>DNA

<213> 人工序列

<220>

<223>D19S433

<400>68

aaggaaagaa ggtagg

16

<210>69

<211>32

<212>DNA

<213> 人工序列

<220>

<223>FGA

<400>69

tttctttctt tcttttttct ctttctttct cc

32

<210>70

<211>420

<212>DNA

<213> 人 (Homo sapiens)

<220>

<221>unsure

<222> (38)... (38)

<223> 未知核苷酸

<220>

<221>unsure

<222> (390)... (390)

<223> 未知核苷酸

<220>

<221>unsure

<222>(397)... (397)

<223> 未知核苷酸

<220>

<221>unsure

<222>(402)... (402)

<223> 未知核苷酸

<220>

<221>unsure

<222>(413)... (413)

<223> 未知核苷酸

<400>70

atggctgccc tcacggctgc accgggagga tgactgtntt cccactctca gtcctgccga	60
ggtgcctgac agccctgcac ccaggagctg ggggggtctaa gagcttgtaa aaagtgtaca	120
agtgccagat gctcgttgtg cacaaatcta aatgcagaaa agcactgaaa gaagaatcca	180
gaaaaccaca gtteccattt ttatatggga gcaaacaag gcagatccca agctcttcct	240
cttccttaga tcaatacaga cagacagaca ggtggataga tagatagata gatagataga	300
tagatagata gatagatatc attgaaagac aaaacagaga tggatgatag atacatgctt	360
acagatgcac acacaaacgt aaatggtatn aaaaatngga tncactcttg tanggttgtt	420

<210>71

<211>239

<212>DNA

<213> 人

<400>71

ccagcctggc ccacacagtc ccctgtacac agggcttccg agtgcaggtc acagggaaca	60
cagactccat ggtgaatgaa tgaatgaatg aatgaatgaa tgagggaat aaggaggaa	120
caggccaatg ggaatcacc cagagcccag ataccctttg aattttgccc cctatttgcc	180
caggaccccc caccatgagc tgctgctaga gcctgggaag ggccttgggg ctgcctccc	239

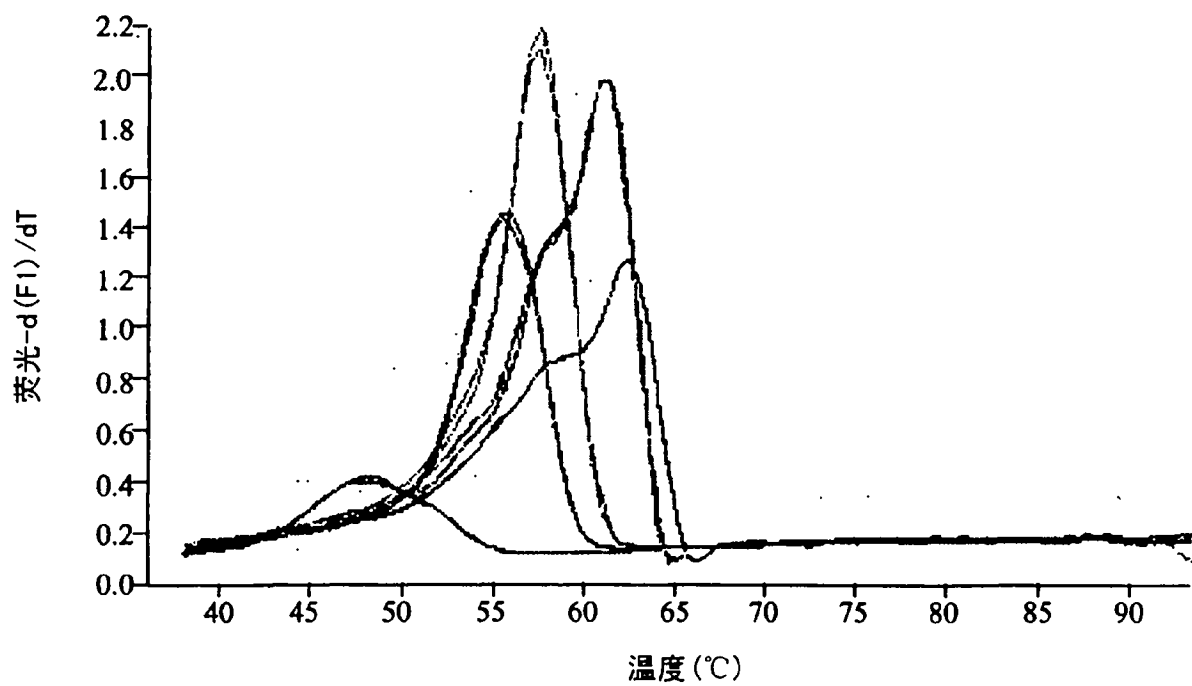


图 1

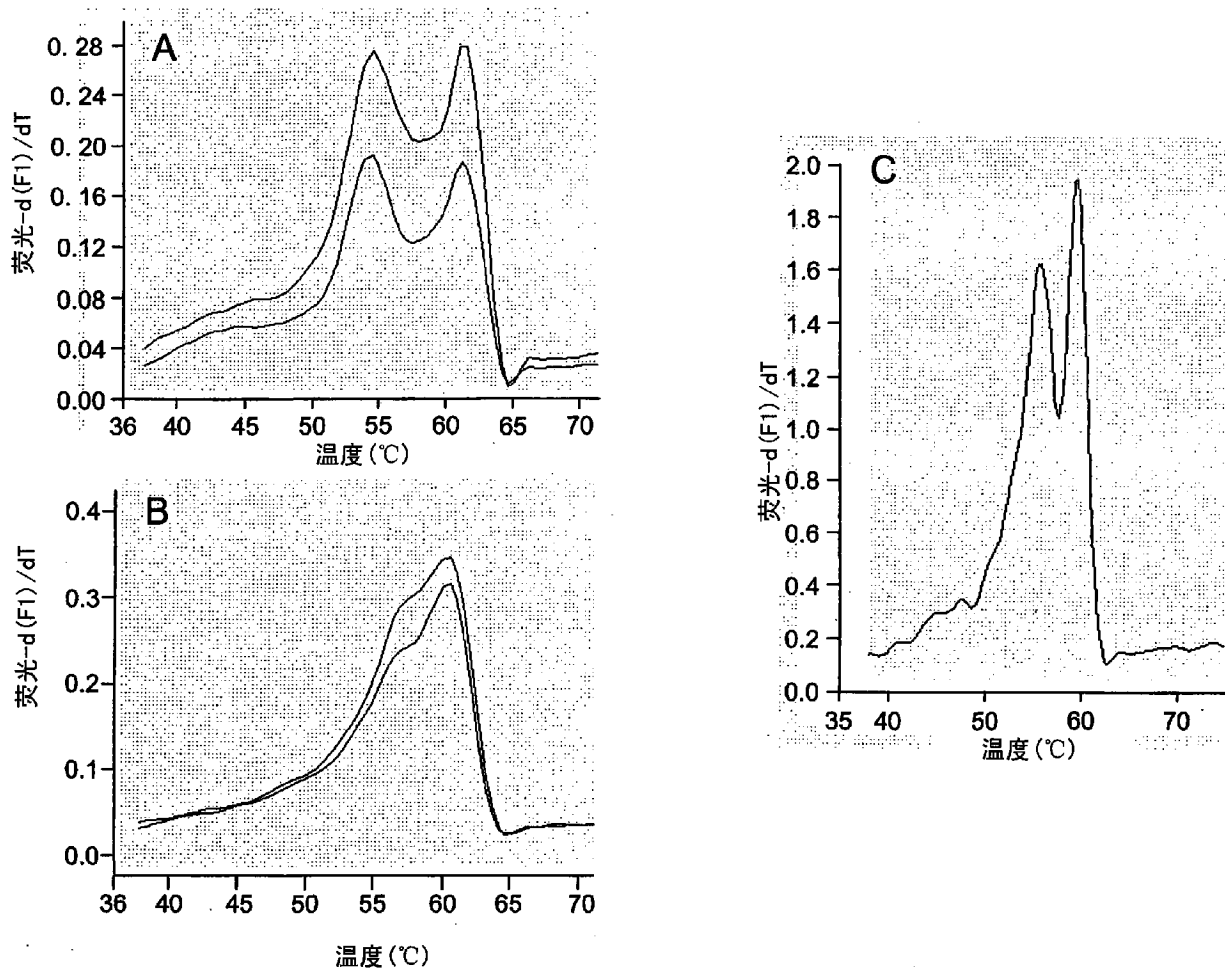


图 2

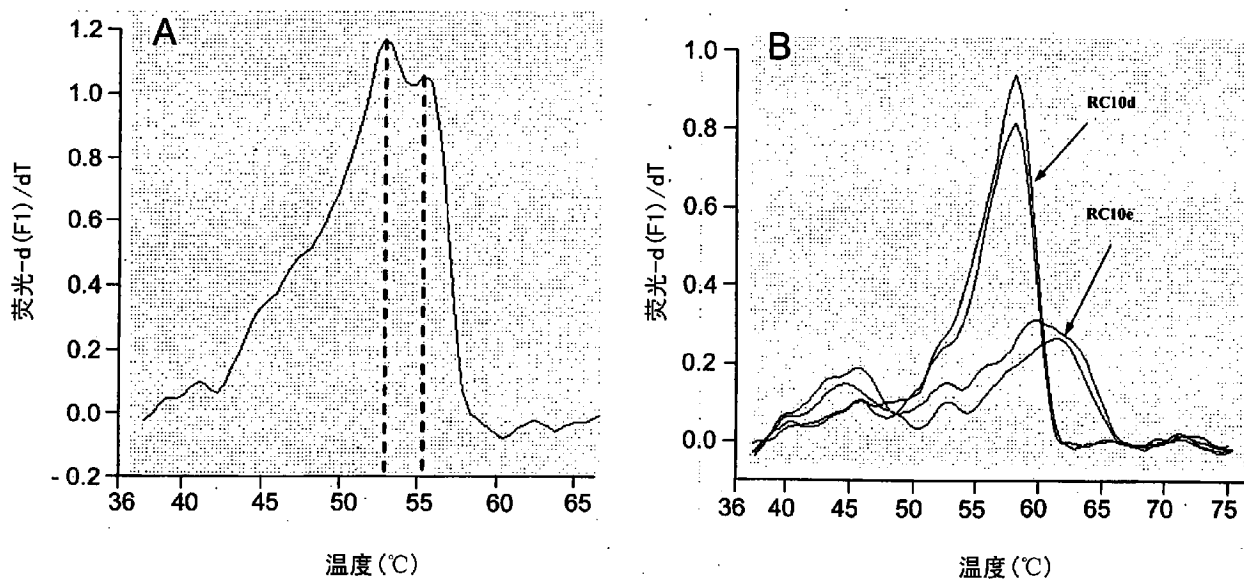


图 3

```

1   atggctgccc tcacggtgc accgggagga tgactgtntt cccactctca gtccctgccga
61  ggtgcctgac agccctgcac ccaggagctg gggggtctaa gagcttgtaa aaagtgtaca
121 agtgccagat gctcgttgtg cacaatcta aatgcagaaa agcactgaaa gaagaatcca
181 gaaaaccaca gttcccattt ttatatggga gcaaacaaag gcagatccca agctcttcct
241 cttccctaga tcaatacaga cagacagaca ggtggataga tagatagata gatagataga
301 tagatagata gatagatatc attgaaagac aaaacagaga tggatgatag atacatgctt
361 acagatgcac acacaaacgt aaatggtatn aaaaatngga tncactcttg tanggttggtt

```

图 4

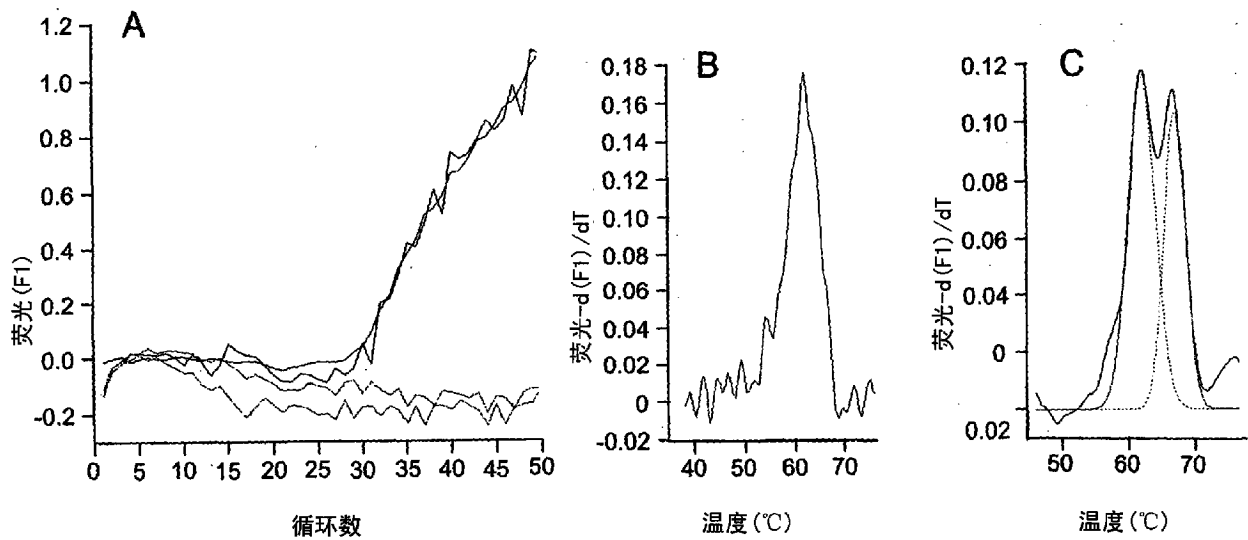


图 5

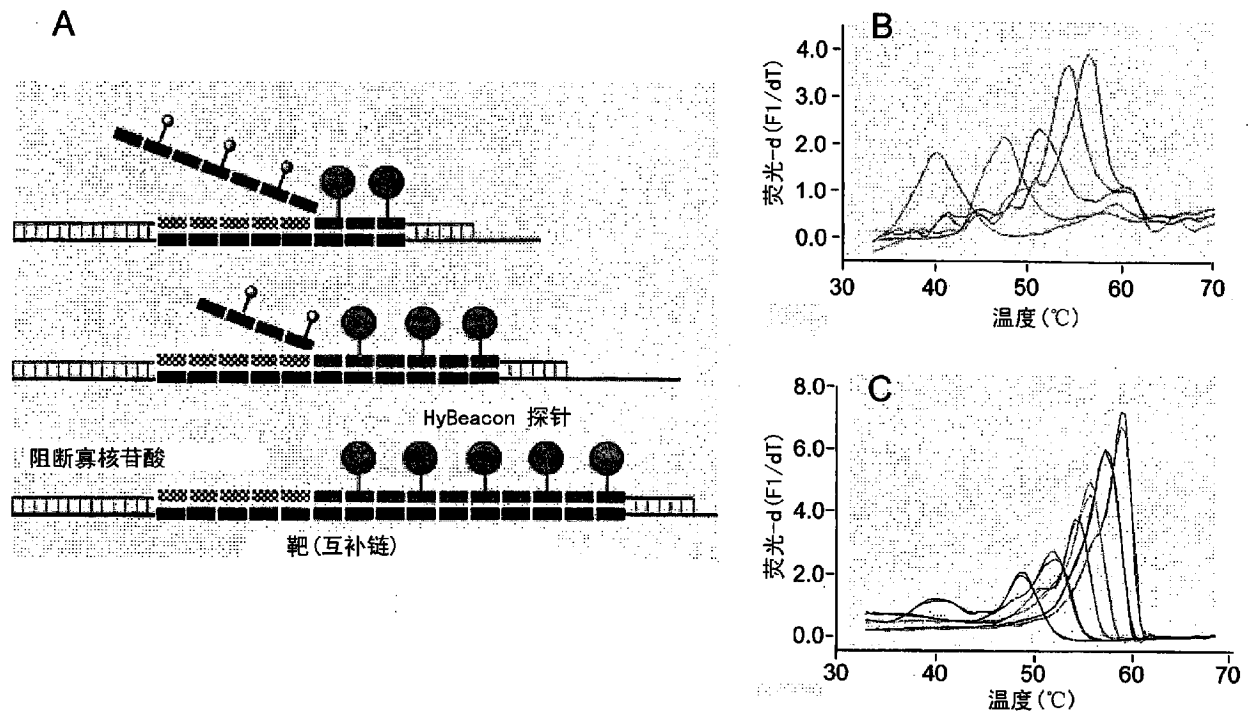


图 6

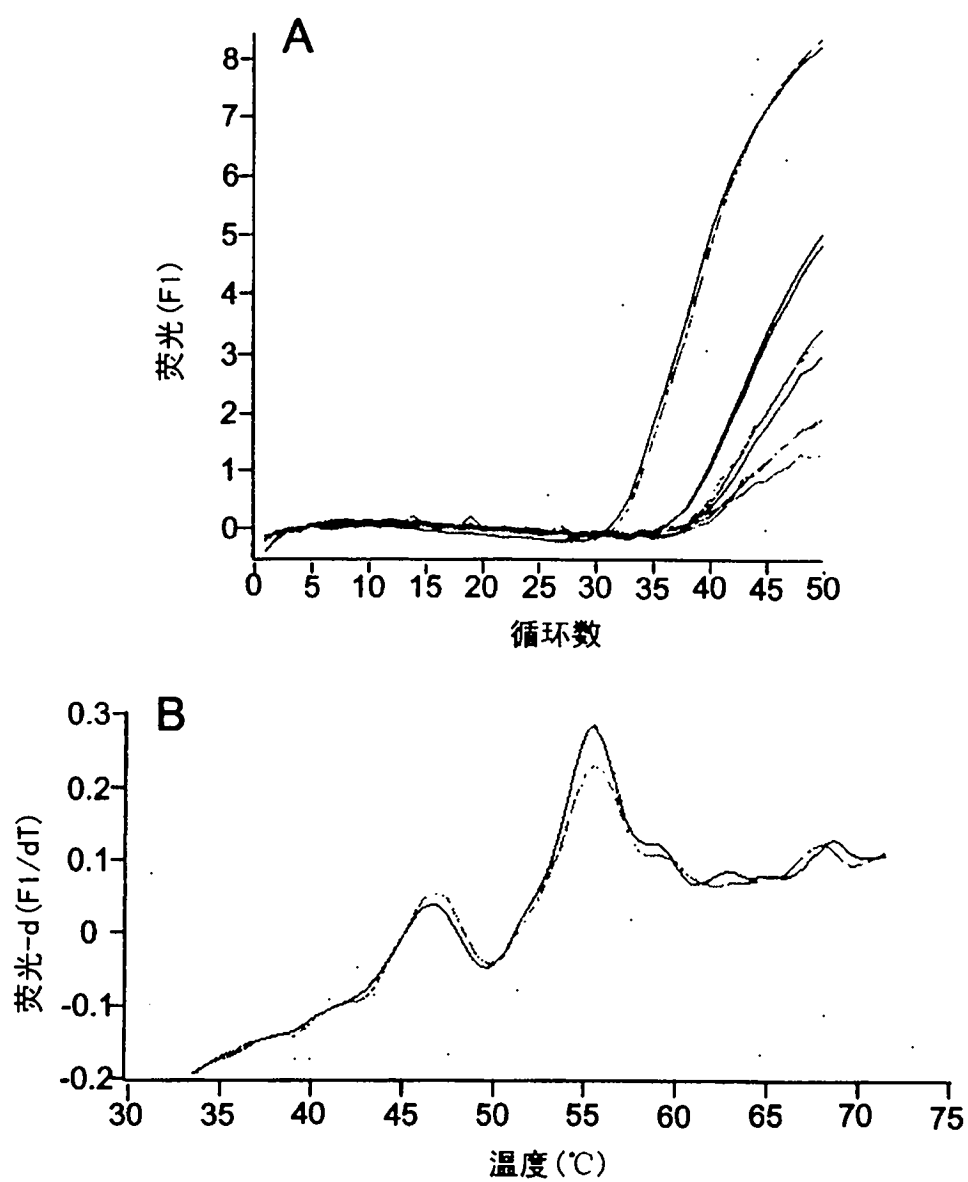


图 7

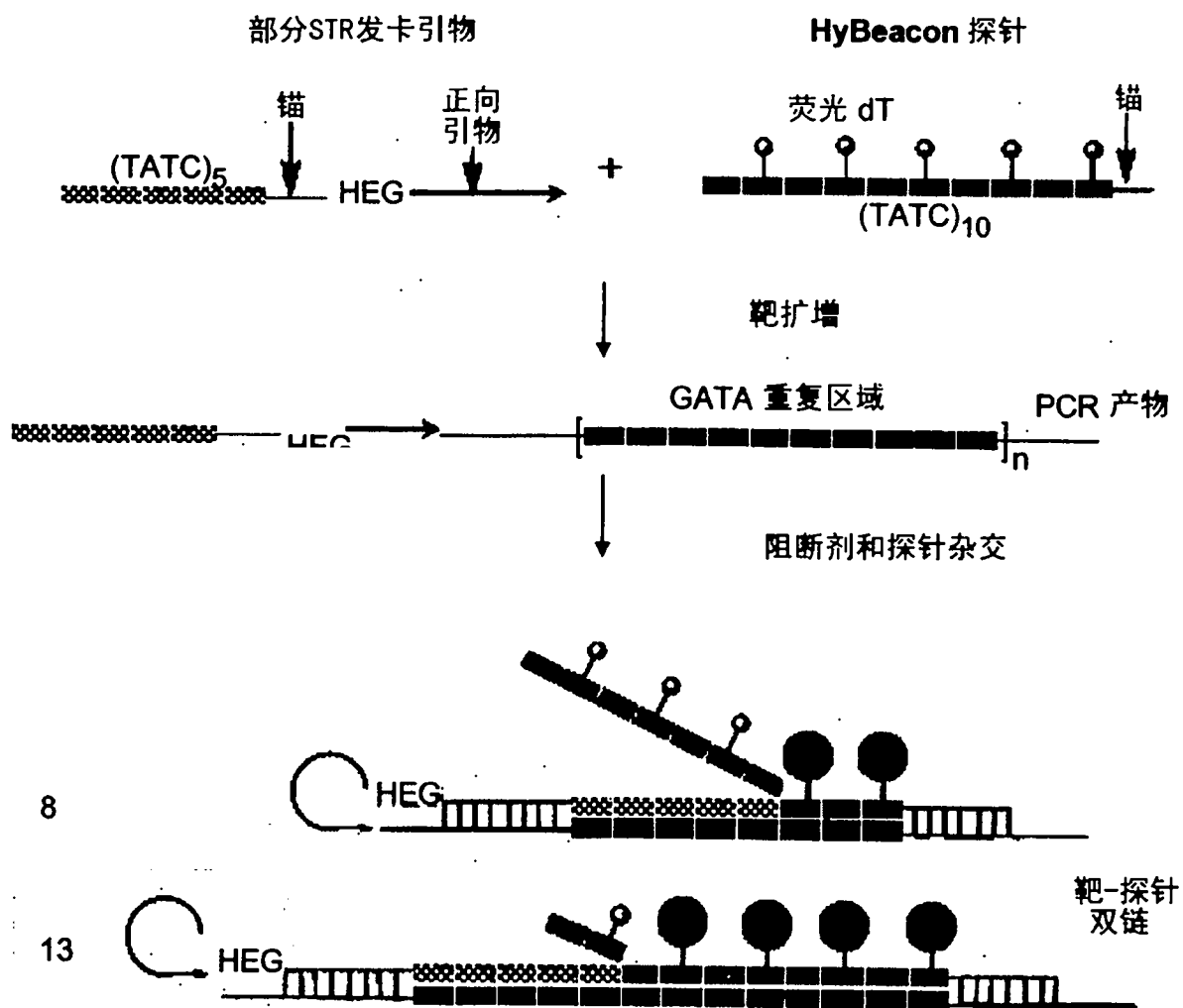


图 8

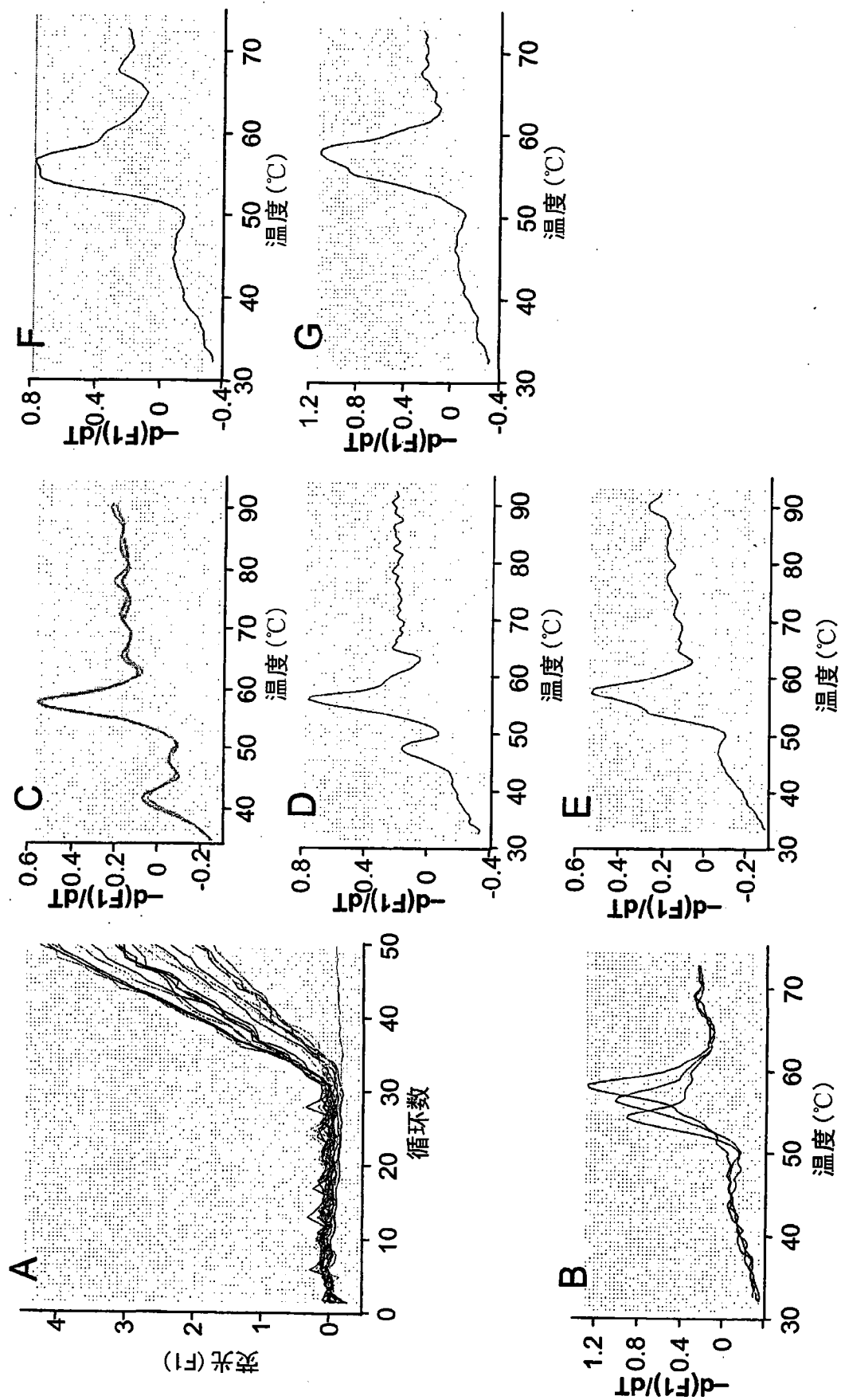


图 9

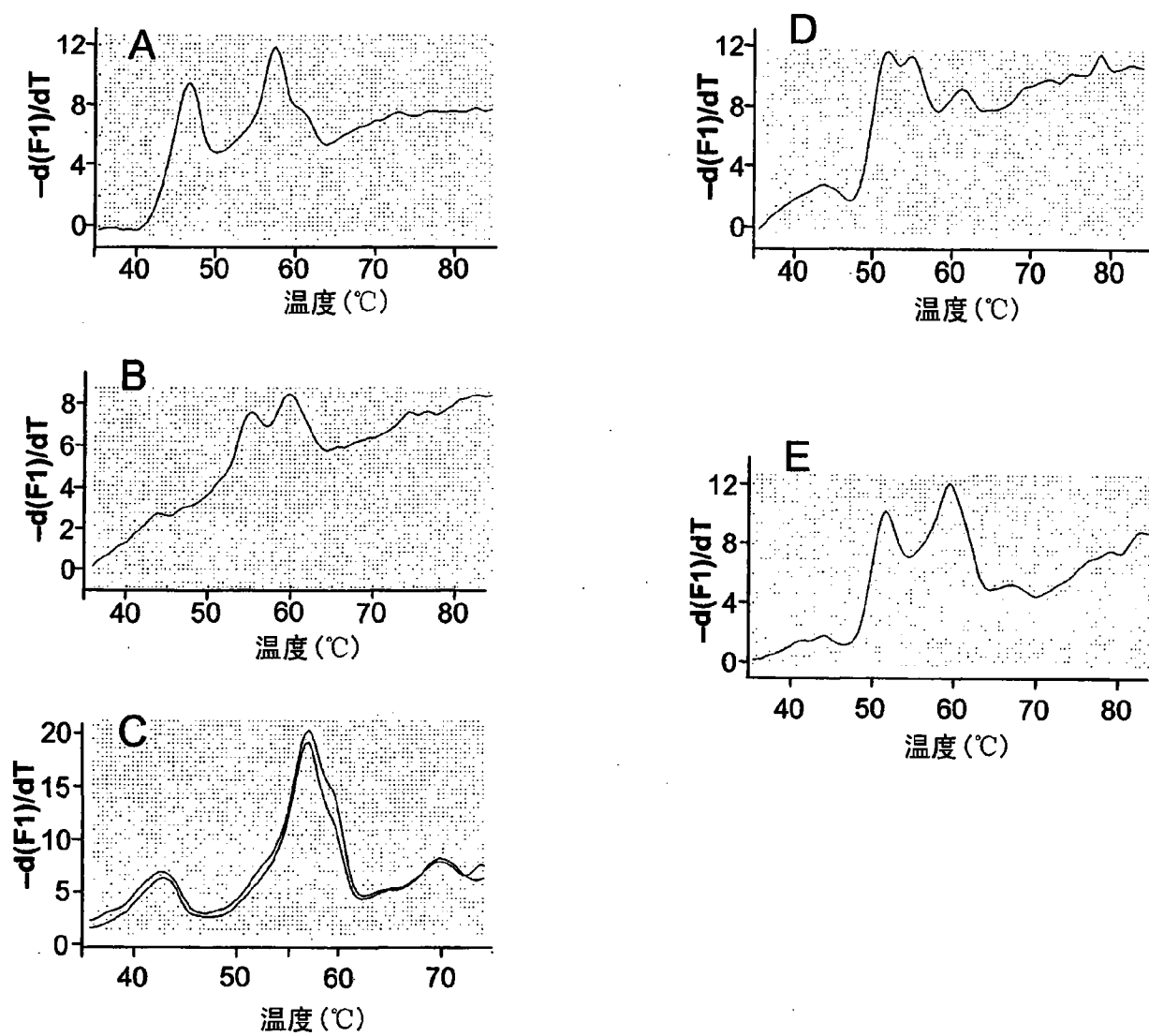


图 10

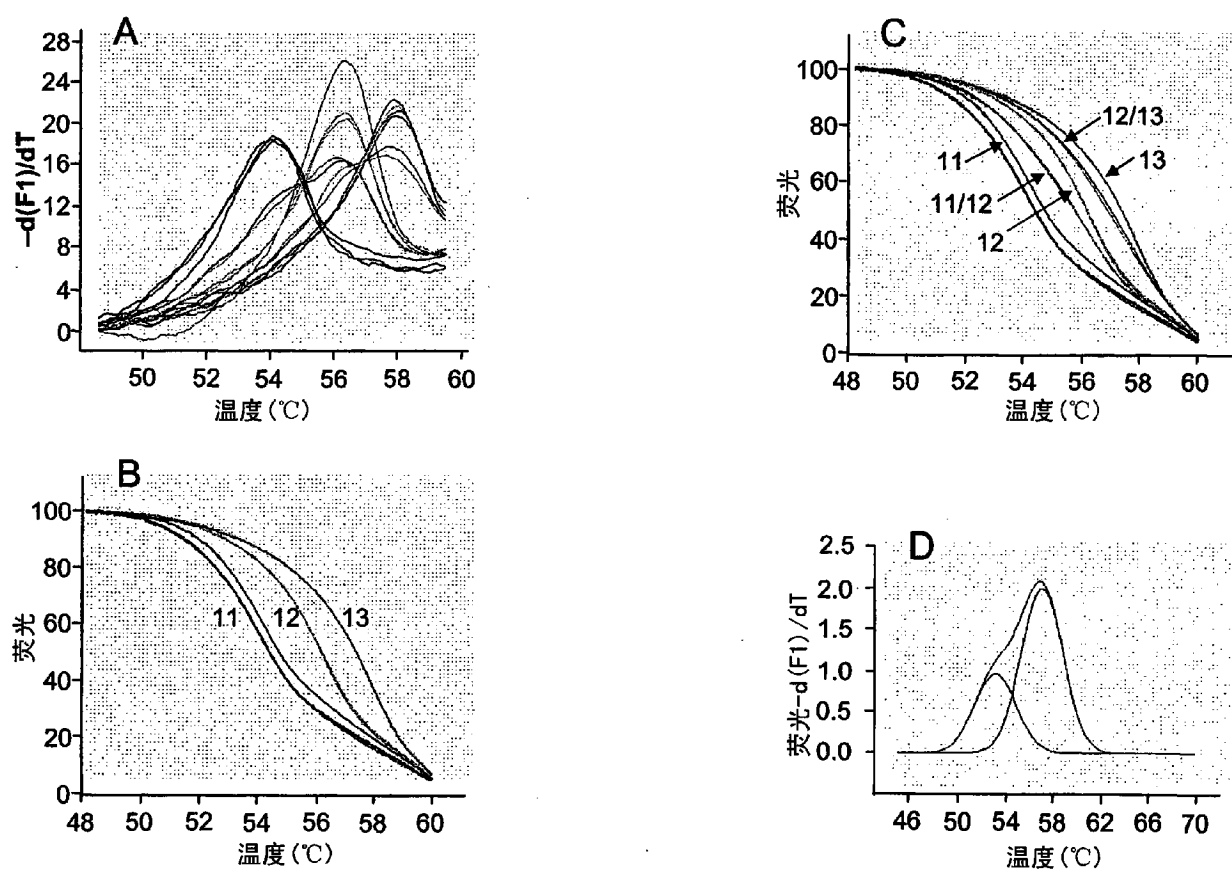


图 11

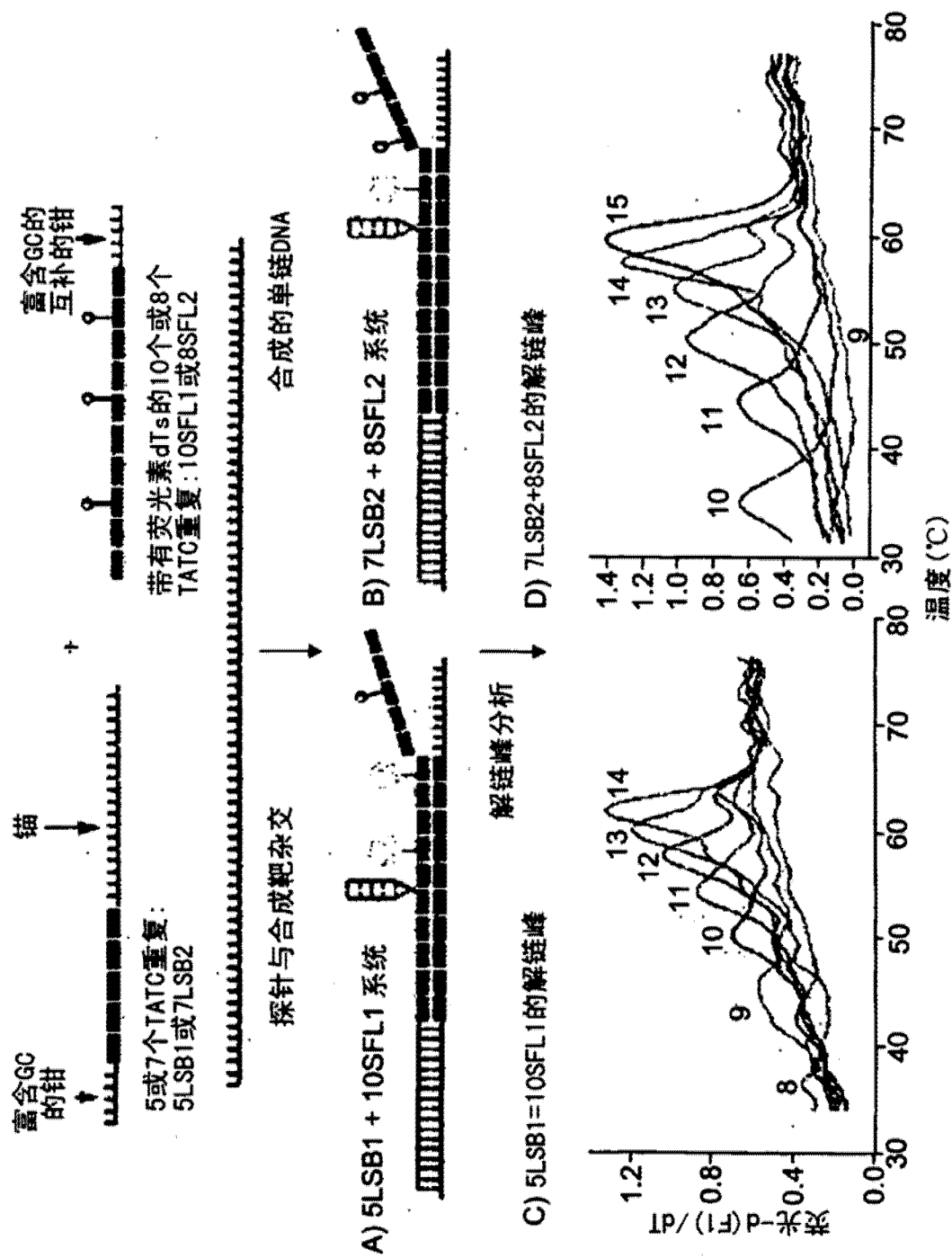


图 12

1	CCAGCCTGGC	CCACACAGTC	CCCTGTACAC	AGGGCTTCCG	AGTGCAGGTC
51	ACAGGGAACA	CAGACTCCAT	GGTGAATGAA	TGAATGAATG	AATGAATGAA
101	TGAGGGAAAT	AAGGGAGGAA	CAGGCCAATG	GGAATCACCC	CAGAGCCCAG
151	ATACCCTTTG	AATTTTGCCC	CCTATTTGCC	CAGGACCCCC	CACCATGAGC
201	TGCTGCTAGA	GCCTGGGAAG	GGCCTTGGGG	CTGCCTCCC	

图 13

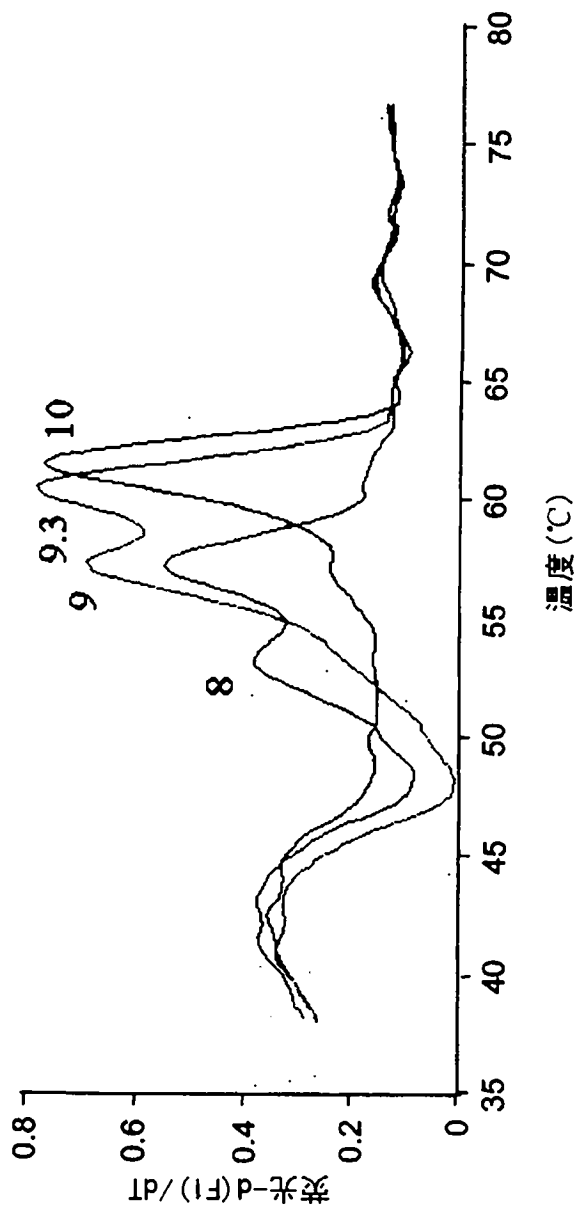


图 14

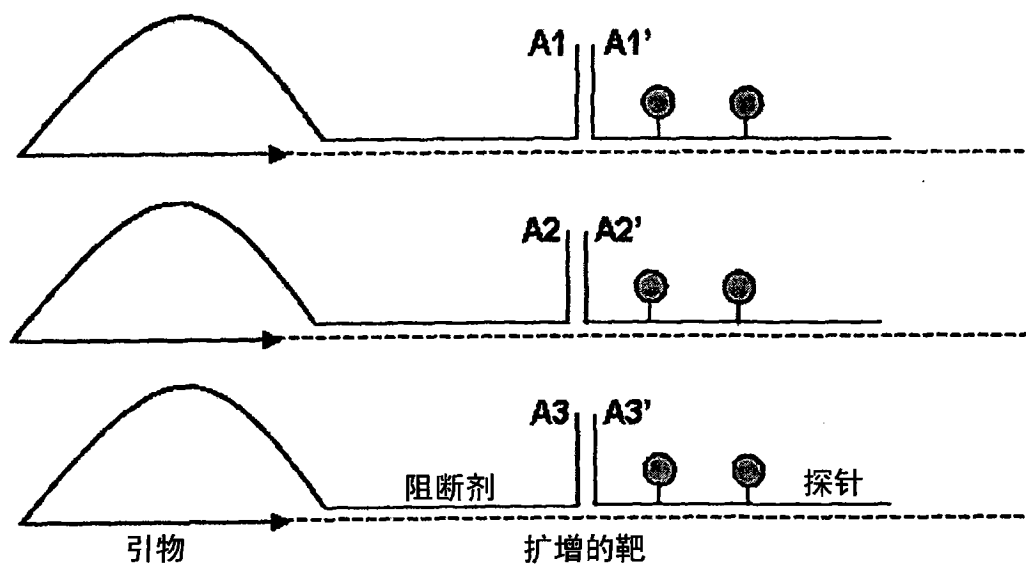


图 15

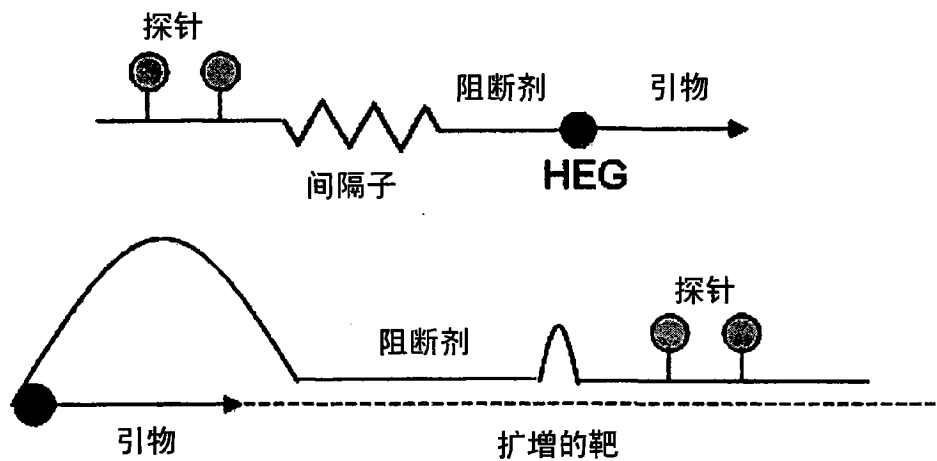


图 16



图 17

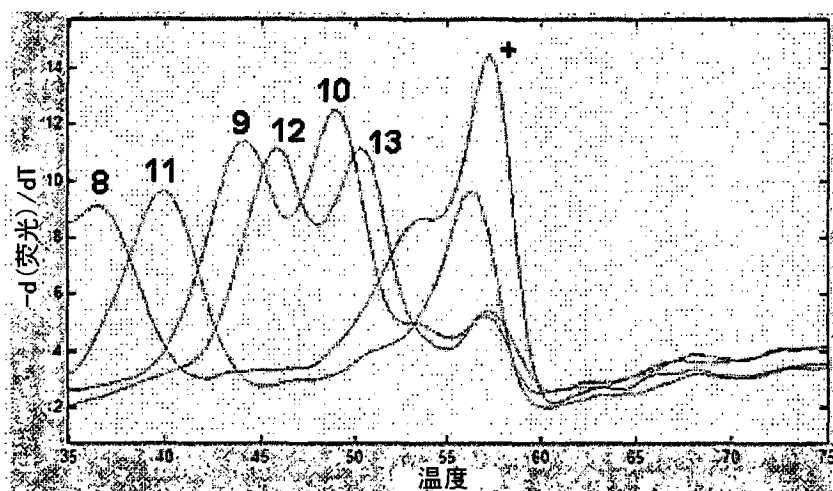


图 18

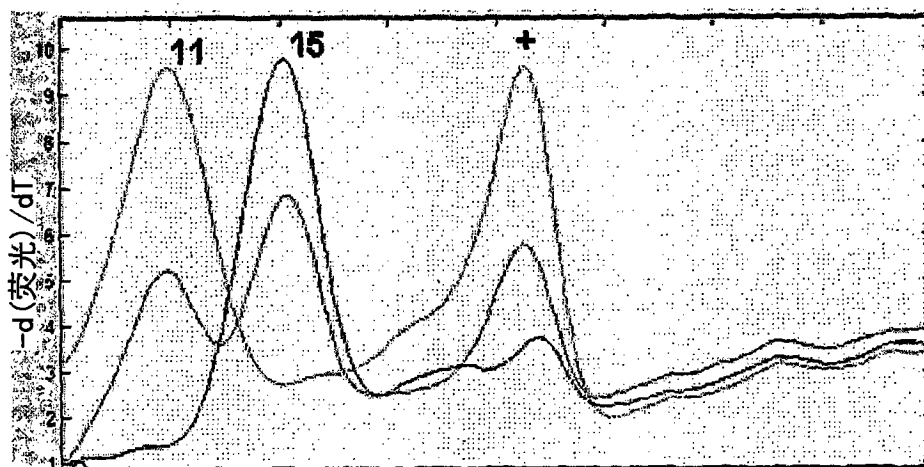


图 19