

[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 200580036019.0

[51] Int. Cl.

C12Q 1/68 (2006.01)

C12P 19/34 (2006.01)

C12P 19/24 (2006.01)

C07H 21/04 (2006.01)

[43] 公开日 2007年11月14日

[11] 公开号 CN 101072882A

[22] 申请日 2005.9.8

[21] 申请号 200580036019.0

[30] 优先权

[32] 2004.9.10 [33] US [31] 60/608,712

[86] 国际申请 PCT/US2005/032441 2005.9.8

[87] 国际公布 WO2006/031745 英 2006.3.23

[85] 进入国家阶段日期 2007.4.20

[71] 申请人 塞昆纳姆股份有限公司

地址 美国加利福尼亚州

[72] 发明人 迪尔克·约翰内斯·范登博姆

塞巴斯蒂安·伯尔克

[74] 专利代理机构 北京律盟知识产权代理有限责任
公司

代理人 王允方 刘国伟

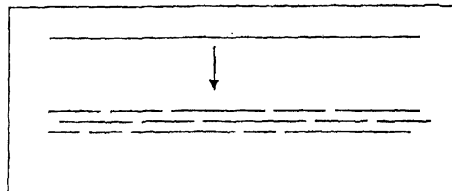
权利要求书6页 说明书95页 附图2页

[54] 发明名称

用于核酸长程序列分析的方法

[57] 摘要

本发明提供一种通过下述步骤对目标核酸测序的方法：使目标核酸断裂；将片段与捕获寡核苷酸的阵列杂交；测定所述杂交片段的质量；和由所述质量测量值构建所述目标核酸的核苷酸序列。



1. 一种对目标核酸测序的方法，其包含：
 - a) 产生目标核酸的重叠片段；
 - b) 使所述片段与捕获寡核苷酸阵列在不排除所述片段与所述捕获寡核苷酸的错配杂交的条件下接触；
 - c) 通过质谱法测量每一个阵列位点处杂交片段的质量；和
 - d) 由所述质量测量值构建所述目标核酸的核苷酸序列。
2. 一种对目标核酸测序的方法，其包含：
 - a) 产生目标核酸的重叠片段；
 - b) 使所述片段与捕获寡核苷酸阵列接触，其中所述捕获寡核苷酸中的一者或一者以上为部分简并寡核苷酸；
 - c) 通过质谱法测量每一个阵列位置处与所述捕获寡核苷酸杂交的片段的质量；和
 - d) 由所述质量测量值构建所述目标核酸的核苷酸序列。
3. 根据权利要求1或2所述的方法，其中所述构建步骤d)包含：

试验性地构建在核苷酸位点处含有假定核苷酸的核苷酸序列；

预测所述试验性核苷酸序列的断裂，预测哪一些预测片段与捕获寡核苷酸杂交；

和

预测已杂交的预测片段的质量；

将所预测的片段质量与实验观察的质量相比较；和

如果所述预测质量与所述观察质量相匹配，那么将所述目标核酸分子中的所述核苷酸位点鉴定为含有所述假定核苷酸。
4. 根据权利要求3所述的方法，其中所述试验性地构建步骤另外包括试验性地构建在核苷酸位点处含有四个典型核苷酸中的每一个核苷酸的核苷酸序列；和对所有试验性核苷酸序列进行所述预测和比较步骤；和将所述预测质量与所述观察质量最紧密匹配的试验性核苷酸序列鉴定为所述目标核酸分子中的核苷酸序列。
5. 根据权利要求3或4所述的方法，其中所述试验性构建、预测、比较和鉴定步骤重复进行，其中每一次重复包括试验性地构建在核苷酸位点处含有假定核苷酸的不增长的核苷酸序列。

6. 根据权利要求1或2所述的方法，其中所述构建步骤d)包含：
 - 确立核酸断裂的片段产物的界限；
 - 确立可与特定捕获寡核苷酸杂交的核酸片段的界限；
 - 预测可在与所述捕获寡核苷酸杂交的核苷酸片段的质谱中观察到的可能质量；
 - 将观察质量与所述可观察到的预测质量相比较以鉴定可能存在的可能序列和/或鉴定不存在的序列；和
 - 对一种或一种以上其它的捕获寡核苷酸重复所述比较、确立、预测和比较步骤，
由此减少可能存在的可能序列的数量，
由此鉴定所述目标核酸分子的核苷酸序列的至少一部分。
7. 根据权利要求1至6中任一权利要求所述的方法，其中所述重叠片段是随机产生的。
8. 根据权利要求1至6中任一权利要求所述的方法，其中所述重叠片段是非特异性产生的。
9. 根据权利要求1至6中任一权利要求所述的方法，其中所述片段使用选自自由酶断裂、物理断裂、化学断裂和其组合组成的群组的断裂方法产生。
10. 根据权利要求1至6中任一权利要求所述的方法，其中所述片段是通过使用一种或一种以上酶进行酶断裂产生，并且其中所述用于酶断裂的一种或一种以上酶选自自由非特异性RNA酶、非特异性DNA酶、至少两种双碱基切割酶、优先切割的核酸内切酶、限制性核酸内切酶、单碱基切割酶、双碱基切割酶和其组合组成的群组。
11. 根据权利要求1至6中任一权利要求所述的方法，其中所述片段是通过物理断裂产生，其中所述物理断裂方法选自自由流体动力、搅动、超声波处理和喷雾法组成的群组。
12. 根据权利要求1至6中任一权利要求所述的方法，其中所述片段是通过化学断裂产生，其中所述化学断裂方法选自自由酸水解、碱水解、烷基化和辐射组成的群组。
13. 根据权利要求1至12中任一权利要求所述的方法，其中所述片段的统计学尺寸范围选自自由5个至50个碱基、10个至40个碱基、11个至35个碱基和12个至30个碱基组成的尺寸范围群组。
14. 根据权利要求1至12中任一权利要求所述的方法，其中所述片段的统计学尺寸范围选自自由20个至50个碱基、30个至60个碱基、40个至70个碱基和50个至80个碱基组成的尺寸范围群组。
15. 根据权利要求1至14中任一权利要求所述的方法，其中所述目标核酸为单链。
16. 根据权利要求1至15中任一权利要求所述的方法，其中所述目标核酸是单链RNA。

17. 根据权利要求 1 至 14 中任一权利要求所述的方法，其中所述目标核酸为双链。
18. 根据权利要求 2 至 17 中任一权利要求所述的方法，其中所述杂交步骤是在不排除错配杂交的条件下进行的。
19. 根据权利要求 1 至 18 中任一权利要求所述的方法，其中所述杂交步骤是在低严格度下进行的。
20. 根据权利要求 1 至 19 中任一权利要求所述的方法，其中所述阵列上存在的组合少于捕获寡核苷酸序列的所有理论组合。
21. 根据权利要求 1 和 3 至 20 中任一权利要求所述的方法，其中所述捕获寡核苷酸中的一者或一者以上为部分简并寡核苷酸。
22. 根据权利要求 1 至 21 中任一权利要求所述的方法，其中所述捕获寡核苷酸全部为部分简并寡核苷酸。
23. 根据权利要求 2 至 22 中任一权利要求所述的方法，其中所述部分简并寡核苷酸包含选自至少 10%、至少 20%、至少 30%、至少 40%和至少 50%组成的群组的分率的简并位置。
24. 根据权利要求 2 至 23 中任一权利要求所述的方法，其中所述部分简并寡核苷酸包含选自 1、2、3、4、5、6、7、8、9 和 10 组成的群组的个数的简并位置。
25. 根据权利要求 24 所述的方法，其中各简并位置都包含选自通用碱基和半通用碱基组成的群组的简并碱基。
26. 根据权利要求 25 所述的方法，其中所述通用碱基选自以下物质组成的群组：肌苷、黄苷、3-硝基吡咯、4-硝基吡咯、5-硝基吡咯、6-硝基吡咯、硝基咪唑、4-硝基咪唑、5-氨基吡咯、4-硝基苯并咪唑、4-氨基苯并咪唑、苯基 C-核糖核苷、苯并咪唑、5-氟吡咯、吡咯；无环糖类似物，次黄嘌呤衍生物、咪唑 4,5-二甲酰胺、3-硝基咪唑、5-硝基咪唑；芳族类似物，苯、萘、菲、芘、吡咯、二氟甲苯；异喹诺酮核苷衍生物，MICS、ICS；和氢键合类似物，N8-吡咯并吡啶。
27. 根据权利要求 25 所述的方法，其中所述半通用碱基选自以下物质组成的群组：优先与嘌呤 A 和 G 杂交的碱基、优先与嘧啶 C 和 T 杂交的碱基、优先与嘧啶 C 和 U 杂交的碱基、6H,8H-3,4-二氢嘧啶并[4,5-c][1,2]恶嗪-7-酮和 N6-甲氧基-2,6-二氨基嘌呤。
28. 根据权利要求 25 至 27 中任一权利要求所述的方法，其中大部分所述简并碱基定位于所述捕获寡核苷酸的 3'端上。
29. 根据权利要求 25 至 27 中任一权利要求所述的方法，其中大部分所述简并碱基定位

于所述捕获寡核苷酸的 5'端上。

30. 根据权利要求 1 至 29 中任一权利要求所述的方法，其中所述阵列含有选自由以下数量组成的群组的数量的不同捕获寡核苷酸：不超过 5,000、不超过 4096、不超过 4,000、不超过 3,000、不超过 2500、不超过 2100、不超过 2000、不超过 1536、不超过 1500、不超过 1400、不超过 1300、不超过 1200、不超过 1100、不超过 1000、不超过 900、不超过 800、不超过 700、不超过 600、不超过 500、不超过 400、不超过 384、不超过 300、不超过 200、不超过 100、不超过 96 和不超过 64。
31. 根据权利要求 30 所述的方法，其中所述捕获寡核苷酸阵列含有 4096 个捕获寡核苷酸并且每个捕获寡核苷酸基本上由 12 个碱基组成。
32. 根据权利要求 1 至 31 中任一权利要求所述的方法，其中所述捕获寡核苷酸阵列固定于选自由以下物质组成的群组的固体载体上：杂交芯片、针具、珠粒、聚苯乙烯、聚碳酸酯、聚丙烯、尼龙、玻璃、右旋糖酐、几丁质、砂、浮石、琼脂、多糖、树枝状聚合物、巴基球、聚丙烯酰胺、硅、金属、橡胶、微量滴定盘、微量滴定孔、载玻片、硅芯片、硝基纤维素薄片和尼龙网。
33. 根据权利要求 1 至 32 中任一权利要求所述的方法，其另外包含用酶处理所述捕获片段阵列以降低所述杂交片段的全长。
34. 根据权利要求 33 所述的方法，其中所述酶选自由单链特异性 RNA 酶、单链特异性 DNA 酶、碱基特异性 RNA 酶和碱基特异性 DNA 酶组成的群组。
35. 一种控制目标核酸片段的质谱的复杂度的方法，其包含：
 - (a) 调节与捕获寡核苷酸探针杂交的目标核酸片段第一区中不同核苷酸序列的数量，由此使所述各自第一区中含有不同核苷酸序列的两个或两个以上目标核酸片段与所述捕获寡核苷酸探针杂交；和
 - (b) 通过质谱法测量与所述捕获寡核苷酸探针杂交的所述目标核酸片段的质量；由此控制所述质谱的复杂度。
36. 根据权利要求 35 所述的方法，其另外包含在测量所述目标核酸片段的质量之前控制所述目标核酸片段的长度的步骤。
37. 根据权利要求 35 至 36 中任一权利要求所述的方法，其中所述捕获寡核苷酸探针含有一个或一个以上简并碱基。
38. 根据权利要求 37 中所述的方法，其中所述简并碱基选自由通用碱基和半通用碱基组成的群组。
39. 根据权利要求 35 至 38 中任一权利要求所述的方法，其中一个或一个以上所述目标

核酸片段另外含有不与所述捕获寡核苷酸探针杂交的第二区。

40. 根据权利要求 39 所述的方法，其中，在含有第二区的所述一个或一个以上目标核酸片段中至少两个在其各自第二区中含有不同的核苷酸序列。
41. 根据权利要求 35 至 40 中任一权利要求所述的方法，其中所述目标核酸片段与所述捕获寡核苷酸探针在选自由中等严格度杂交条件和低严格度杂交条件组成的群组的杂交条件下杂交。
42. 根据权利要求 35 至 41 中任一权利要求所述的方法，其中一个或一个以上所述目标核酸片段的所述第一区含有选自由 3'端和 5'端组成的群组的所述目标核酸片段末端。
43. 根据权利要求 39 至 42 中任一权利要求所述的方法，其中所述一个或一个以上目标核酸片段的第二区在选自由 3'端和 5'端组成的群组的所述目标核酸片段末端的核苷酸位上含有一个或一个以上已知核苷酸。
44. 根据权利要求 35 至 43 中任一权利要求所述的方法，其中所述控制所述目标核酸片段长度的步骤另外包括碱基特异性切割。
45. 根据权利要求 35 至 44 中任一权利要求所述的方法，其中所述目标核酸片段与捕获寡核苷酸探针阵列杂交，其中所述阵列含有多个位置，并且每一阵列位置处的所述捕获寡核苷酸探针的核苷酸序列都不同于所有其它阵列位置处捕获寡核苷酸探针的核苷酸序列。
46. 一种鉴定目标核酸的一部分的方法，其包含：
 - (a) 根据权利要求 35 至 45 中任一权利要求所述的方法，收集具有受控复杂度的质谱；和
 - (b) 将所述一个或一个以上目标核酸片段质量与一种或一种以上参考核酸的一个或一个以上质量相比较；

其中一个或一个以上目标核酸片段质量与一个或一个以上参考质量之间的相关性将所述目标核酸的一部分鉴定为与所述参考核酸对应或与所述参考核酸的一部分对应。
47. 根据权利要求 46 所述的方法，其中所述至少一种参考核酸的一个或一个以上参考质量经计算得到。
48. 根据权利要求 46 至 47 中任一权利要求所述的方法，其中所述至少一种参考核酸的一个或一个以上参考质量是根据实验测得。
49. 根据权利要求 46 至 48 中任一权利要求所述的方法，其中所述目标核酸片段使用选

自序列特异性断裂和非特异性断裂的方法形成。

50. 根据权利要求 46 至 49 中任一权利要求所述的方法,其中所述经鉴定的目标核酸部分含有 SNP。
51. 一种鉴定目标核酸的一部分的组合,其包含:
 - (a) 在固体载体上的两种或两种以上捕获寡核苷酸的阵列,其中至少一种捕获寡核苷酸为部分简并寡核苷酸;和
 - (b) 可操作地连接至所述阵列的质谱仪。
52. 根据权利要求 51 所述的组合,其另外包含计算机程序,以供由一组从与所述捕获寡核苷酸杂交的核酸分子获取的质量信号构建所述目标核酸的核苷酸序列。
53. 根据权利要求 52 所述的组合,其另外包含一组一个或一个以上参考质量峰。

用于核酸长程序列分析的方法

相关申请案

本申请案主张 2004 年 9 月 10 日申请的 60/608,712 的权利,所述专利涉及下述专利: 2003 年 4 月 11 日申请的 Lin 等人的标题为“METHOD AND DEVICE FOR PERFORMING CHEMICAL REACTION ON A SOLID SUPPORT”的美国申请案第 10/412,801 号; 2003 年 3 月 24 日申请的 Lin 等人的标题为“METHOD AND DEVICE FOR PERFORMING CHEMICAL REACTION ON A SOLID SUPPORT”的美国临时申请案第 60/457,847 号; 2002 年 4 月 11 日申请的 Lin 等人的标题为“METHOD AND DEVICE FOR PERFORMING CHEMICAL REACTION ON A SOLID SUPPORT”的美国临时申请案第 60/372,711 号; 2003 年 11 月 27 日申请的 van den Boom 等人的标题为“FRAGMENTATION-BASED METHODS AND SYSTEMS FOR SEQUENCE VARIATION DETECTION AND DISCOVERY”的美国申请案第 10/723,365 号; 2002 年 11 月 27 日申请的 van den Boom 等人的标题为“FRAGMENTATION-BASED METHODS AND SYSTEMS FOR SEQUENCE VARIATION DETECTION AND DISCOVERY”的美国临时申请案第 60/429,895 号; 2004 年 4 月 22 日申请的 Böcker 等人的标题为“FRAGMENTATION-BASED METHODS AND SYSTEMS FOR DENOVO SEQUENCING”的美国临时申请案第 10/830,943 号; 和 2003 年 4 月 25 日申请的 Böcker 等人的标题为“FRAGMENTATION-BASED METHODS AND SYSTEMS FOR DE NOVO SEQUENCING”的美国临时申请案第 60/466,006 号。这些非临时和临时申请案中每一个申请案的主题和内容都全部并入本文用作参考。

技术领域

本发明提供一种核酸分析方法。

背景技术

分析各种生物聚合物结构是医药和研究中极其重要的方面。分子遗传学取决于 DNA 或 RNA 分子的核苷酸序列的知识。蛋白质的氨基酸序列提供可用于研究蛋白质功能和

调控的信息。存在多种用于分析生物聚合物序列的策略。确定核酸序列的最常用的方法，双脱氧法 (dideoxy method)，包括：建立四组在四个碱基的每一个处终止的 DNA 分子的子序列；通过聚丙烯酰胺凝胶电泳 (gel electrophoresis, PAGE) 分离所述片段；和读取所得谱带以确定所述序列。凝胶电泳可能较慢且存在错误。

已提出的用于克服凝胶电泳测序的缺点的方法是一种称为杂交测序的方法，例如参看 Bains 和 Smith, *J. Theoret. Biol.*, 135:303-307 (1998); Lysov 等人, *Dokl. Acad. Sci. USSR* 303:1508-1511 (1988); Drmanac 等人, *Genomics* 4:114-128 (1989); Pevzner, *J. Biomolec. Struct. Dynamics* 7(1):63-73 (1989); Pevzner 和 Lipschutz, *Nineteenth Symp. on Math. Found. of Comp. Set*, LNCS-841: 143-258 (1994); Waterman, *Introduction to Computational Biology*, Chapman 和 Hall, London, 1995。杂交测序 (Sequencing by hybridization, SBH) 是一种使短核苷酸序列 (探针) 的阵列 (SBH 芯片) 与目标 DNA 序列 (复制品) 的溶液接触而进行的 DNA 测序技术。生化方法确定与目标序列结合的探针子集 (所述序列的图谱)，并且使用组合方法由所述图谱重建 DNA 序列。由于技术上限制 SBH 芯片上探针的数目，因此具有挑战性的组合问题在于设计出可对指定长度的任意随机 DNA 链测序的最少探针组。

SBH 的实施使用了“经典”探测流程，即，芯片提供所有 4^k k -mer 寡核苷酸 (无间隙的“实心”探针)，所述符号为众所周知的 DNA 碱基 {A、C、G、T}；且 k 为与技术相关的整数参数。据说，“杂交测序的主要问题在于，可靠地检测正确的双链体并使其与含有错配碱基对的双链体相区别 ([t]he main challenge for sequencing by hybridization is to reliably detect the perfect duplexes and discriminate them from duplexes containing mismatched base pairs)” (Chechetkin 等人, *J. of Biomolecular Structure & Dynamics* 18(1):83-101 (2000))。因此，杂交测序方法力图避免和最小化碱基对错配，碱基对错配将导致假阳性或假阴性结果，最终将导致测序方法的失败。

SBH 方法依靠避免错配杂交来排除假阳性和/或假阴性读数。因此，需要允许错配杂交而获得新生核酸序列信息的基于杂交的方法。由此，在本文的众多目的中，一个目的在于提供允许错配杂交而获得新生核酸序列信息的方法。

发明内容

本文所提供的方法为允许错配杂交而获得新生核酸序列信息的方法。本文提供用于核酸序列分析 (包括从头测序) 的方法，其包含产生目标核酸的重叠片段；在不排除错配杂交的条件下，使所述片段与固体载体上的捕获寡核苷酸阵列杂交以形成所捕获片段

的阵列；通过例如质谱分析法测定其质量来确定所述阵列中每一位点处所捕获片段的质量；和由从各阵列位置获取的一组质量信号构建所述目标核酸的核苷酸序列或一组核苷酸序列。本文还提供对核酸测序的方法，其包含产生目标核酸的重叠片段；使所述片段与固体载体上的捕获寡核苷酸阵列杂交以形成所捕获片段的阵列，其中至少一分组所述捕获寡核苷酸为部分简并寡核苷酸；通过例如质谱分析法测定其质量来确定所述阵列中各位点处所捕获片段的质量；和由从各阵列位置获取的一组质量信号构建所述目标核酸的核苷酸序列或一组核苷酸序列。在一实施例中，所述重叠片段是随机产生。

使用本发明的方法由样本所获得的序列信息可用于基因分型和单倍体分型、多重基因分型和单倍体分型、核酸混合物分析、长程再测序、序列变异和突变的长程检测、多重测序、长程甲基化模式分析、生物体鉴定、病原鉴定和分型等。

因此，本文所提供的方法有利地合并基于固相杂交的方法与基于算法的杂交产物组成分析，以显著增强使用质谱法进行的基于固相杂交的序列分析。本文所提供的方法的一个优势在于，可达成与前述方法相比显著增加的目标核酸序列读取长度的数量和准确性。较高（长程）序列读取长度是使用经非特异性切割或部分特异性切割并随后与固相上的捕获寡核苷酸结合的目标核酸的质谱分析实现，部分或全部所述捕获寡核苷酸可为部分简并寡核苷酸。例如，本文所提供的方法能够在在一个反应/实验中对至少 250 个、500 个、600 个、700 个、800 个、900 个、1,000 个、1,500 个、2,000 个、3,000 个、4,000 个、5,000 个、6,000 个、7,000 个、8,000 个、9,000 个、多达 10,000 个或更多个核苷酸测序。为实现这一目标，最终定制所产生的供本文所提供的方法分析的片段以便提供较大目标核酸的序列。

在另一实施例中，通过本文所提供的方法对多种具有较短长度的较短目标核酸片段进行测序或分析。当已知特定序列的一部分时，这些多重较短序列组（例如）可用于再测序方法中。这些多重较短序列组也可用于多重基因分型、单倍体分型、SNP 和甲基化检测方法中。

所述片段可通过整体或部分非特异性切割和/或通过部分特异性切割产生，并且通常获得重叠片段以供分析。可使用单一非特异性切割反应和/或互补或部分碱基特异性切割反应获得重叠片段，从而获得相同目标生物分子序列的其它重叠片段。切割方式可为酶切割、化学切割、物理切割或其组合，并且通常产生重叠片段。因此，视选择用于产生重叠片段的特定方法而定，所述重叠片段可为随机产生或不随机产生。

可使用此项技术中已知的方法测定经切割和未经切割的目标序列片段的质量，所述方法包括（但不限于）质谱法和凝胶电泳法。在典型实施例中，使用 MALDI-TOF 质谱

法测定所述片段的质量。用于进行高通量质谱分析的芯片和试剂盒可从 SEQUENOM, INC 以商标 MassARRAY7 购得。用于本文中的另一示范性芯片为 2002 年 4 月 11 日申请的相关美国申请案第 60/372,711 号、2003 年 3 月 24 日申请的第 60/457,847 号和 2003 年 4 月 11 日申请的第 10/412,801 号中所述的“h 芯片 (h-chip)”，所述专利全部并入本文用作参考。

因此，在一实施例中，本文所提供的方法组合固相杂交与在固相上分选的重叠切割产物的质谱检测和鉴定的高通量。本文所提供的方法也改进鉴定由非特异性断裂或部分特异性断裂所产生的片段信号的准确性和清晰度，并且还通过使用在一个目标核酸或一组目标核酸内重建序列的算法增加这些信号的分析速度。

附图说明

图 1 描述重叠片段的产生。

图 2 显示与固体载体上的简并捕获寡核苷酸杂交的多个片段。

图 3 描述对杂交捕获寡核苷酸:目标片段双链体进行的“修剪”。

具体实施方式

A. 定义

B. 核酸分子测序方法

C. 目标核酸分子

1. 来源

2. 制备

3. 目标核酸分子的尺寸和组成

4. 扩增

D. 断裂

1. 聚核苷酸的酶断裂

a. 聚核苷酸的核酸内切酶断裂

b. 核酸酶断裂 (Nuclease Fragmentation)

c. 核酸酶断裂 (Nucleic Acid Enzyme Fragmentation)

d. 碱基特异性断裂

2. 聚核苷酸的物理断裂

3 聚核苷酸的化学断裂

4.断裂方法的组合

5.杂交后断裂

E.捕获寡核苷酸

1.控制目标核酸片段的复杂度

a.控制复杂度的方法

b.片段的区域

c.部分单链的捕获寡核苷酸

2.捕获寡核苷酸的组成

a.核苷酸类型

i.通用碱基

ii.半通用碱基

b.其它特征

c.制备捕获寡核苷酸

F.固体载体和阵列

G.特异性或非特异性杂交

H.修剪

I.有关目标核酸片段的信息

1.分子质量

a.质谱分析

b.其它测量方法

2.质量峰特征

3.捕获寡核苷酸和杂交条件

4.断裂条件

J.核苷酸序列的构建

K.通过质量图鉴定核苷酸序列

L.鉴定一部分目标核酸

M.应用

1.长程再测序

2.突变/序列变异的长程检测

3.多重测序

- 4.长程甲基化模式分析
- 5.生物体鉴定
- 6.病原鉴定和分型
- 7.分子育种和定向进化
- 8.作为标记的目标核酸片段
- 9.检测指示感染的病毒或细菌核酸序列的存在
- 10.抗生素谱
- 11.鉴定疾病标记
- 12.单倍体分型
- 13.DNA 重复序列
- 14.检测等位基因变异
- 15.测定等位基因频率
- 16.表观遗传学 (Epigenetics)

实例

A.定义

除非另作定义，否则本文所使用的所有科技术语都具有与本发明所属领域技术人员通常所理解的含义相同的含义。除非另作说明，否则本文的全部公开内容通篇所提及的所有专利、专利申请案、公开申请案和公告、GENBANK 序列、网页和其它公开资料均全部并入本文用作参考。在本文中的术语存在多种定义的情况下，以本节定义为准。当提到 URL 或类似标志符或地址时，应了解这些标志符会改变，并且互联网上的特定信息也会变化不定，但等同的信息是已知的并且例如可通过搜寻互联网和/或合适数据库而容易地获得。所附参考文献证明此类信息的可用性和公共传播。

如本文所使用的“阵列”是指元素（例如核酸）的集合。阵列通常含有三个或三个以上成员。可编址的阵列为例如可通过固体载体上的位置鉴定所述阵列的成员的阵列。因此，阵列成员可固定于固相表面上可鉴定的不连续位点处，或另外例如可通过粘附上包括电子标签和化学标签在内的标签或由这些标签做标记来进行鉴定。阵列包括（但不限于）单个固相表面上元素的集合，例如芯片上寡核苷酸的集合。

如本文所使用的“特异性杂交”是指通常在高严格度杂交条件下探针或引物仅优先与目标序列而不是非目标序列杂交。例如，特异性杂交包括探针和与所述探针 100%互补的目标序列杂交。所属领域技术人员熟悉影响杂交的参数，例如温度、探针或引物长度和组成、缓冲液组成和盐浓度；并且可容易地调整这些参数以达成核酸与目标序列的

特异性杂交。

如本文所使用的杂交严格度是指除去捕获寡核苷酸与目标核酸片段的非特异性结合的洗涤条件。示范性杂交条件如下：

- 1) 高严格度：0.1 × SSPE，0.1% SDS，65EC
- 2) 中等严格度：0.2 × SSPE，0.1% SDS，50EC
- 3) 低严格度：1.0 × SSPE，0.1% SDS，50EC。

所属领域技术人员已知选择用于稳定杂合体的洗涤步骤，也了解 SSPE 的成份（例如，参看 Sambrook, E.F. Fritsch, T. Maniatis, in: *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989), 第 3 卷, 第 B.13 页；也参看描述常用实验室溶液的多个目录）。SSPE 是 pH 7.4 经磷酸盐缓冲的 0.18 M NaCl。而且，所属领域技术人员认识到，通过 T_m 确定杂合体的稳定性， T_m 是钠离子浓度和温度的函数 ($T_m = 81.5EC - 16.6(\log_{10}[\text{Na}^+]) + 0.41(\%G + C) - 600/l$)，因此，在洗涤条件中对杂合体稳定性极为重要的参数是 SSPE（或 SSC）中钠离子的浓度和温度。特异性杂交通常是在高严格度条件下发生。应了解，可使用其它缓冲液、盐和温度达到相同的严格度。

如本文所使用的“核酸”或“核酸分子”是指聚核苷酸，例如脱氧核糖核酸（DNA）和核糖核酸（RNA）。所述术语也应理解为包括由核苷酸类似物制造的 RNA 或 DNA 的等效物、衍生物、变异体和类似物；单链（有义链或反义链）和双链聚核苷酸。脱氧核糖核苷酸包括脱氧腺苷、脱氧胞苷、脱氧鸟苷和脱氧胸苷。对于 RNA 而言，尿嘧啶碱基为尿苷。

如本文所使用的“质谱法”涵盖所属领域技术人员已知的任何适当的质谱格式。所述格式包括（但不限于）基质辅助激光解吸/电离（Matrix-Assisted Laser Desorption/Ionization）、飞行时间（Time-of-Flight, MALDI-TOF）、电喷射（Electrospray, ES）、IR-MALDI（例如，参看已公开的国际 PCT 申请案第 99/57318 号和美国专利第 5,118,937 号）、正交飞行时间（Orthogonal-TOF, O-TOF）、轴向飞行时间（Axial-TOF, A-TOF）、线性/反射飞行时间（Linear/Reflectron, RETOF）、离子回旋共振（Ion Cyclotron Resonance, ICR）、傅立叶转换（Fourier Transform）和这些方法的组合。MALDI、尤其 UV 和 IR 为此项技术中已知的格式。也参看 Aebersold 和 Mann, 2003 年 3 月 13 日, *Nature*, 422:198-207（例如，图 2）中有关适用于本文所提供的方法中的质谱法的示范性方法的综述，所述文献全部并入本文用作参考。MALDI 方法通常包括 UV-MALDI 或 IR-MALDI。

如本文所使用的短语“质谱分析”是指原子、分子或分子片段荷质比的测定。

如本文所使用的质谱是指由通过质谱法分析生物聚合物或其片段获得的以图形表

示或数字编码或其它方式提供的数据呈现。

如本文所使用的关于质谱或质谱分析的图案是指信号、其峰或数字表示的特征分布和数量。

如本文在有关质谱和其分析的上下文中所使用的信号、峰或测量值是指，可反映原子、分子或分子片段荷质比且也可反映所述原子、分子或其片段的量的输出数据。荷质比可用于测定原子、分子或分子片段的质量，且所述量可用于定量或半定量方法中。例如，在一些实施例中，信号峰或测量值可反映具有特定荷质比的分子的数量或相对数量。信号或峰包括输出数据的可视、图形和数字表示。

如本文中当提及测量质量时所使用的强度是指，样本或组合物中所存在的分析物与其它样本或组合物组分相比时的相对量的反映。例如，第一个质谱峰或信号的强度可相对于第二个质谱峰报道，或可相对于所有峰强度的总和报道。所属领域技术人员可认识到多种报道峰的相对强度的方式。强度可以峰高、半高峰宽、峰下面积、信噪比或此项技术中已知的其它表示法表示。

如本文所使用的比较测量质量或质量峰是指，分析一个或一个以上样本的一个或一个以上测量样本质量峰或参考质量峰。例如，可通过与计算的质量峰图案相比较来分析测量的样本质量峰，并且可以确定测量的质量峰与计算的质量峰之间的任何重叠，进而鉴定样本质量或分子。参考质量峰表示参考原子、分子或分子片段的质量。

如本文所使用的参考质量为可与测量的样本质量进行比较的质量。样本质量与参考质量之间的比较可鉴定样本质量与参考质量相同或不同。所述参考质量可经计算；可提供于数据库中；或可以实验法测定。计算的参考质量可基于核酸的预测质量。例如，计算的参考质量可基于所预测的已知或预测序列的目标核酸分子的断裂模式。以实验法得到的参考质量可由任何核酸样本的测量质量得到。例如，以实验法得到的质量可为在断裂条件下处理核酸分子并使所述片段与捕获寡核苷酸接触后所测量的质量。参考质量的数据库可含有一个或一个以上参考质量，其中所述参考质量可经计算或以实验法测定；数据库可含有与目标核酸分子的计算或以实验法测定的断裂模式对应的参考质量；数据库可含有与两个或两个以上目标核酸分子的计算或以实验法测定的断裂模式对应的参考质量。

如本文所使用的参考核酸分子是指，已知核苷酸序列或已知身份（identity）（例如，无已知序列但具有已知疾病相关性的位点）的核酸分子。参考核酸可用于计算或以实验法得出参考质量。用于计算参考质量的参考核酸通常为含有已知核苷酸序列的核酸。用于以实验法得出参考质量的参考核酸可具有（但非必须具有）已知序列；即使当参考核

酸不具有已知序列时，仍可使用如本文所公开的方法或此项技术中已知的其它方法等鉴定参考核酸的核苷酸序列。

如本文所使用的一个或一个以上样本质量（或一个或一个以上样本质量峰特征）与一个或一个以上参考质量（或一个或一个以上参考质量峰特征）之间的相关性和其语法派生词是指，一个或一个以上样本质量（或一个或一个以上样本质量峰特征）与一个或一个以上参考质量（或一个或一个以上参考质量峰特征）之间的比较，其中质量相似性的增加表明目标核酸分子或其片段的核苷酸序列与参考核酸的核苷酸序列相同的可能性增加。

如本文所使用的一个或一个以上样本质量峰与一个或一个以上参考质量峰之间的相关性和其语法派生词是指，一个或一个以上样本质量峰与一个或一个以上参考质量峰之间的关系，其中所述一个或一个以上样本质量峰与所述一个或一个以上参考质量峰之间一个或一个质量峰特征相似性的增加表明，样本目标核酸的至少一部分与参考核酸的至少一部分相同的可能性增加；或表明目标核酸的一个或一个以上核苷酸位置处的核苷酸序列与参考核酸的一个或一个以上核苷酸位置处的核苷酸序列相同的可能性增加。

如本文所使用的目标核酸分子核苷酸序列与参考核苷酸序列之间的相关性是指，目标核酸分子的核苷酸序列与参考的核苷酸序列的相似性或同一性。

如本文所使用的“分析”是指测定单一寡核苷酸或寡核苷酸混合物的特定特性。这些特性包括（但不限于）单一寡核苷酸或寡核苷酸混合物的核苷酸组成和完整序列；单一核苷酸多态性和一个以上寡核苷酸之间其它突变的存在；寡核苷酸的质量和长度；和样本分子内分子或序列的存在。

如本文所使用的“多重”、“多重反应”或其语法派生词是指，在单一反应或单一质谱法或其它序列测量（即，读取序列的单一质谱法或其它方法）中同时评定或分析一个以上的分子，例如生物分子（例如，寡核苷酸分子）。

如本文所使用的扩增是指增加生物聚合物、尤其是核酸的量的方式。基于所选择的5'和3'引物，扩增也起到限制和界定经历分析的基因组的区域的作用。可通过所属领域技术人员已知的任何方式进行扩增，包括使用聚合酶链反应（PCR）等。当需要测定多态性的频率时，扩增（例如，PCR）必须定量进行。

如本文所使用的短语“统计学尺寸范围”是指使用部分切割所产生的大部分片段的尺寸范围，因此，部分片段可实质小于或大于特定尺寸范围内的大部分其它片段。例如，12-30个碱基的统计学尺寸范围内也可包括一些小到1个核苷酸或大到300个核苷酸或更多个核苷酸的寡核苷酸，但这些特殊尺寸在统计学上相对罕见。片段的统计学范围可

包括：60%片段在所需尺寸范围内的情形；60%或60%以上片段在所需尺寸范围内的情形；70%或70%以上片段在所需尺寸范围内的情形；80%或80%以上片段在所需尺寸范围内的情形；90%或90%以上片段在所需尺寸范围内的情形；或95%或95%以上片段在所需尺寸范围内的情形。

如本文所使用的短语“杂交”或其语法派生词是指核酸序列与其完全或部分互补链的结合。如本文所使用的术语杂交可适用于完全互补链的结合，也适用于不是完全互补的链的结合。因此，杂交可包括第一个核酸与第二个核酸结合且其中所述第一个核酸与第二个核酸具有一个或一个以上错配碱基的情形。

如本文所使用的短语“在不排除错配杂交的条件下”是指，允许具有一个或一个以上碱基对错配的捕获寡核苷酸结合的杂交条件。在一些实施例中，所允许的错配数量选自不超过5个、不超过4个、不超过3个、不超过2个和不超过1个碱基对错配。

如本文所使用的短语“捕获片段”是指与捕获寡核苷酸（例如固相上的捕获寡核苷酸）结合的目标核酸片段。

如本文所使用的“简并位”是指核苷酸上含有代替四个常见碱基中的一个碱基、与一个以上核苷酸结合的取代基的位置。例如，核苷酸的简并位可为含有通用碱基或半通用碱基的核苷酸的位置。部分简并核苷酸是指含有至少一个简并位和至少一个非简并位（例如，含有通用碱基或半通用碱基和非简并碱基，例如A、G、C或T/U）的核苷酸，或含有至少一个相对于其它核苷酸优先与一些核苷酸结合的简并位（例如，含有至少一个半通用碱基）的核苷酸。在本文的某些实施例中，部分简并寡核苷酸含有至少10%、20%、30%、40%、多达50%个简并位。例如，对于长度为20个核苷酸的捕获寡核苷酸而言，这些部分简并寡核苷酸可含有1、2、3、4、5、6、7、8、9、多达10个简并位。在其它实施例中，简并寡核苷酸可含有50%以上简并位，包括100%简并位。例如，长度为20个核苷酸的寡核苷酸可含有20个半通用核苷酸，或10个通用核苷酸和10个半通用核苷酸。

如本文所使用的固体载体微粒是指离散微粒形式的材料。所述微粒具有任何形状和维度，但通常至少一个维度尺寸为100 mm或100 mm以下、50 mm或50 mm以下、10 mm或10 mm以下、1 mm或1 mm以下、100 μm 或100 μm 以下、50 μm 或50 μm 以下，并且通常尺寸为100 mm^3 或100 mm^3 以下、50 mm^3 或50 mm^3 以下、10 mm^3 或10 mm^3 以下和1 mm^3 或1 mm^3 以下、100 μm^3 或100 μm^3 以下，且可为立方微米；所述微粒直径通常为大于约1.5微米且小于约15微米，例如约4-6微米。所述微粒可统称为“珠粒”。

如本文所使用的“固体载体”是指能够提供可进行反应和/或反应产物可保留在可鉴

定的位点处的表面的不溶性载体。载体可由实质上任何不溶性材料或固体材料制造。例如，硅胶、玻璃（例如，可控孔径玻璃（controlled-pore glass, CPG）、尼龙、王氏树脂（Wang resin）、Merrifield 树脂、葡聚糖凝胶（Sephadex）、琼脂糖凝胶（Sepharose）、纤维素、金属表面（例如，钢、金、银、铝和铜）、硅和塑料材料（例如，聚乙烯、聚丙烯、聚酰胺、聚酯、聚偏二氟乙烯（polyvinylidenedifluoride, PVDF））。示范性固体载体包括（但不限于）平坦载体，例如玻璃纤维过滤器、玻璃表面、金属表面（钢、金、银、铝、铜和硅）和塑料材料。固体载体为适于安装于芯筒基底（cartridge base）上的任何所需形式，包括（但不限于）：板、膜、晶片、有孔晶片（wafer with pit）、多孔三维载体（porous three-dimensional support）和所属领域技术人员已知的其它几何形状和形式。示范性载体为经设计以接收或连接不连续位点处的样本的平坦表面，例如具有包围用于接收、容纳或结合样本的亲水位点的疏水区的平坦表面。

如本文在有关核酸断裂的内容中所使用的短语“非特异性切割”或“非特异性断裂”是指，始终在任意位置处断裂目标核酸分子从而随机产生多个具有不同尺寸和核苷酸序列含量的片段。如本文所使用，在任意位置处断裂并不需要纯数学上的随机性，而只是在断裂时缺乏强烈的基于序列的优先选择。例如，通过辐射或剪切方式断裂可在几乎任何位置处切割 DNA；然而，所述方法可导致断裂在一些位置处比其它位置处略为频繁。尽管如此，出于本文的目的，仍认为仅具有极小序列优先选择性在几乎所有位置处进行的断裂为随机的。使用本文所述的方法进行的非特异性切割导致产生重叠核苷酸片段。

如本文所使用的术语部分或不完整切割或者部分或不完整断裂或其语法派生词是指其中在特定断裂条件下仅各自切割位点的一部分实际被切割的反应。断裂条件可为（但不限于）存在酶、化学力或物理力。如本文所述，一种达成部分断裂的方式为：在制造目标生物分子期间使用可切割或不可切割核苷酸或氨基酸的混合物，从而使特定切割位点含有不可切割的核苷酸或氨基酸，此使得甚至当切割反应完成时仍使目标生物分子受到部分切割。例如，如果未经切割的目标生物分子中具有 4 个潜在切割位点（例如，核酸的切割碱基），那么由部分切割所得到的产物混合物可具有由下述方式得到的目标生物分子的片段的任何组合：在第一个、第二个、第三个或第四个切割位点处进行的单次切割；在 2 个切割位点的任一个或一个以上组合处进行的双切割；或在 3 个切割位点的任一个或一个以上组合处进行的三次切割。部分切割的产物可与全部切割产物存在于同一混合物中。

如本文所使用的短语“重叠片段”是指共有天然目标核酸的一个或一个以上核苷酸位置的片段。如本文所使用的“统计学重叠片段”是指所界定尺寸的亚群体与至少一个

其它片段重叠的一组片段。例如，统计学重叠片段可指其中至少 50%、至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 95%或至少 98%的片段与至少一个其它片段重叠的一组片段。

如本文所使用的“非特异性 RNA 酶”是指不考虑切割位点处的核苷酸序列切割 RNA 分子的酶。示范性非特异性 RNA 酶为 RNA 酶 I。

如本文所使用的“非特异性 DNA 酶”是指不考虑切割位点处所存在的核苷酸序列切割 DNA 分子的酶。示范性非特异性 DNA 酶为 DNA 酶 I。

如本文所使用的术语“单碱基切割酶 (single-base cutter)”是指识别并切割特定碱基 (例如，对于 DNA 而言为 A、C、T 或 G；或对于 RNA 而言为 A、C、U 或 G) 或特定类型的碱基 (例如，嘌呤或嘧啶) 的限制性酶。

如本文所使用的术语“1-1/4-切割酶 (1-1/4-cutter)”是指识别并切割核酸中的 2 个碱基段的限制性酶，其中，一个碱基位置的身份固定而另一个碱基位置的身份为四个常见碱基中的任何三个。

如本文所使用的术语“1-1/2-切割酶”是指识别并切割核酸中的 2 个碱基段的限制性酶，其中一个碱基位置的身份固定而另一个碱基位置的身份为四个常见碱基中的任何两个。

如本文所使用的术语“双碱基切割酶”或“2 切割酶”是指识别并切割两个碱基长的特定核酸位点的限制性酶。

如本文所使用的短语“质量信号组”是指对两个或两个以上核酸片段所作的两次或两次以上质量测定结果。

如本文所使用的记分或得分是指对特定序列变异候选者实际存在于目标核酸或蛋白质序列中的概率的计算。得分值用于确定与实际目标序列相对应的序列变异候选者。通常，在一组目标序列样本中，最高得分表示目标分子中最有可能存在序列变异；但也可以使用其它规则进行选择，例如当存在单一目标序列时检测正得分。

如本文所使用的模拟是指，基于核酸或蛋白质序列和核酸或蛋白质序列中对于特定特异性切割试剂的预测切割位点计算断裂模式。可将断裂模式模拟为数字表 (例如，作为与参考生物分子片段的质量信号相对应的峰值列表)、质谱、凝胶上的谱带图案或测量质量分布的任何技术的表示。在大部分情况下，模拟可通过计算机程序进行。

如本文所使用的模拟切割是指实质切割目标分子或参考分子的电子方法 (*in silico process*)。

如本文所使用的电子方法 (*in silico*) 是指使用计算机进行研究和实验。电子方法包

括（但不限于）分子模拟研究（molecular modelling study）、生物分子对接实验（biomolecular docking experiment）和分子结构和/或过程（例如分子相互作用）的虚拟展示（virtual representation）。

如本文所使用的短语“构建核苷酸序列”是指，使用可经设计用于所述构建的各种算法说明目标核酸分子的核苷酸序列的方法。

如本文所使用的受检者包括（但不限于）动物、植物、细菌、病毒、寄生虫和具有核酸的任何其他生物体或实体。受检者为哺乳动物时，优选（但非必需）为人类。患者是指受疾病或病症折磨的受检者。

如本文所使用的表型是指包括生物体的任何可辨别的特质在内的一组参数。表型可为物理特质，且在受检者为动物的情况下可为心理特质，例如情感特质。

如本文所使用的“指定（assignment）”是指对核酸或蛋白质片段的位置表明特定分子重量和特定末端核苷酸或氨基酸的确定。

如本文所使用的“一”是指一个或一个以上。

如本文所使用的“多个”是指两个或两个以上。例如，多个聚核苷酸或多肽是指两个或两个以上聚核苷酸或多肽，所述聚核苷酸或多肽各自具有不同序列。这一差异可归因于所述序列中天然存在的变异，例如，可归因于核苷酸或经编码氨基酸中的等位基因变异；或可归因于各个序列中特定修饰的引入，例如，将质量修饰的核苷酸有差别地并入多个核酸或蛋白质中的每一个核酸或蛋白质中。

如本文所使用的“确定的（unambiguous）”是指对于与目标分子中特定序列变异（例如突变）对应的峰或信号的唯一指定，且在大量分子或突变为多重的情况下，可仅将表示特定序列变异的峰分配至各突变或各分子。

如本文所使用的数据处理程序是指可在软件中具体表现的程序，用于确定所获得的数据（即，阵列的最终结果）的生物意义。例如，数据处理程序可基于所收集的数据确定基因型。在本文的系统和方法中，数据处理程序也可基于所测定的结果控制仪器和/或数据收集程序。数据处理程序和数据收集程序可相结合并提供反馈以通过仪器操纵数据的采集，并因此提供本文所提供的基于阵列的判断方法。

如本文所使用的多个基因包括至少 2 个、5 个、10 个、25 个、50 个、100 个、250 个、500 个、1000 个、2,500 个、5,000 个、10,000 个、100,000 个、1,000,000 个或更多个基因。多个基因可包括一个生物体或甚至多个生物体的完整或部分基因组。选择生物体类型确定从中选择基因调控区的基因组。供基因筛选的示范性生物体包括动物，例如哺乳动物，包括人类和啮齿动物（例如小鼠）；昆虫；酵母；细菌；寄生虫；和植物。

如本文所使用的“样本”是指含有待检测的物质的组合物。在优选实施例中，样本为“生物样本”。术语“生物样本”是指由活体来源获得的任何物质，所述活体来源例如为动物，如人类或其他哺乳动物；植物；细菌；真菌；原生生物；或病毒。生物样本可为任何形式，包括固体物质，例如组织、细胞、细胞小球、细胞提取物或活组织检查；或生物体液，例如尿液、血液、血浆、血清、唾液、痰液、羊水、感染或发炎区的渗出液或含有口腔细胞的漱口液、脑脊髓液、滑液、器官、精液、眼内液、粘液、如胃液或乳汁的分泌液；和病理样本，例如石蜡包埋的福尔马林固定样本。优选使固体材料与流体混合。具体说来，在本文中，当进行生物材料（例如核酸）的质谱分析时，可使样本与基质混合。源自意思是可例如通过纯化或分离和/或扩增核酸分子处理样本。

如本文所使用的组合物是指任何混合物。可为溶液、悬浮液、液体、粉末、糊状物、含水、无水或上述物质的任何组合。

如本文所使用的组合是指两个或两个以上项目之间的任何联合。

如本文所使用的术语“扩增子”是指可复制的 DNA 区域。

如本文所使用术语“完整切割”或“全部切割”是指完全切割由特定切割试剂所识别的所有切割位点的切割反应。

如本文所使用的术语“假阳性”是指超出背景噪音且不是由预期的事物产生的信号。例如，当观察到不是反映目标核酸核苷酸序列的质量峰时，或当通过不同于核酸或蛋白质的特异性实际切割或模拟切割的方法形成片段时，会出现假阳性。

如本文所使用的术语“假阴性”是指实际测量遗漏但另外预期的实际信号。例如，当实际质谱中未观察到的质量信号经计算存在于对应的模拟图谱中时，会出现假阴性。

如本文所使用的断裂或切割意思是将核酸或蛋白质分子分成较小段的任何方式。断裂或切割方法包括物理切割、酶切割、化学切割和产生较小核酸段的任何其它的方式。

如本文所使用的断裂条件或切割条件是指一种或一种以上断裂试剂、缓冲液或可用于进行实际或模拟切割反应的其它化学或物理条件组。所述条件包括反应参数，例如时间、温度、pH 值；或缓冲液的选择。

如本文所使用的未切割的切割位点是指，作为已知切割试剂的识别位点但在反应（例如时间、温度）条件下或对切割识别位点处的碱基进行修饰以防止试剂切割的条件下未经切割试剂切割的切割位点。

如本文所使用的互补切割反应是指，使用不同切割试剂或通过改变相同切割试剂的切割特异性从而产生相同目标核酸或参考核酸或蛋白质的交替切割模式于相同目标核酸或参考核酸或蛋白质上进行或模拟的切割反应。

如本文所使用的流体是指可流动的任何组合物。因此，流体涵盖半固体、糊状物、溶液、水性混合物、凝胶、洗液、乳液形式的组合物和其它所述组合物。

如本文所使用的细胞提取物是指由溶解或破坏细胞得到的制剂或片段。

如本文所使用的试剂盒为视情况将组分与使用说明书和/或结合使用的试剂和装置一起包装的组合。

如本文所使用的系统是指元件与软件和用于控制和指导本文所提供的方法的任何其它元件的组合。

如本文所使用的软件是指当通过计算机执行时进行计算机操作的计算机可读程序指令。通常，将软件提供于含有记录于计算机可读媒体上的程序指令的程序产品上，所述计算机可读媒体例如为（但不限于）：磁性媒体，包括软盘、硬盘和磁带；和光学媒体，包括 CD-ROM 光盘、DVD 光盘、磁光盘；和可记录程序指令的其它所述媒体。

如本文所使用的短语目标核酸或目标核酸分子是指有兴趣进行分析的核酸分子。目标核酸分子可为单链分子或双链分子。

如本文所使用的短语“部分消化”意思是仅一分组限制性位点被切割。

如本文所使用的“控制复杂度”和其语法派生词是指，操纵具有不同核苷酸序列的核酸分子的数量、变异性或数量和变异性的方法。例如，控制与捕获寡核苷酸杂交的目标核酸片段的复杂度是指操纵实验条件以控制与特定捕获寡核苷酸探针序列杂交的具有不同核苷酸序列的目标核酸片段的数量、变异性或数量和变异性。与捕获寡核苷酸探针杂交的不同目标核酸序列的数量是指与捕获寡核苷酸探针的至少一部分特定核苷酸序列杂交的不同目标核酸或目标核酸片段的数量。例如，具有彼此不同的序列的两个或两个以上目标核酸片段可与单一阵列位置杂交，其中所述单一阵列位置的所有捕获寡核苷酸探针具有相同的核苷酸序列。在一实例中，具有不同序列的两个目标核酸可与捕获寡核苷酸杂交，其中杂交反应需要捕获寡核苷酸与目标核酸片段的两个不同核苷酸序列之间碱基配对。因此，在本文所公开的方法的一个实施例中，捕获寡核苷酸能够与两个或两个以上不同核苷酸序列碱基配对。与捕获寡核苷酸探针杂交的不同目标核酸序列的变异性是指序列同一性程度，这是就与捕获寡核苷酸探针杂交的不同目标核酸序列的长度与核苷酸序列而言。

如本文所使用的“调节”与捕获寡核苷酸探针杂交的序列的数量是指，设置或改变条件以便设置或改变与捕获寡核苷酸探针杂交的目标核酸片段的序列的数量、变异性或数量和变异性。可经设置或改变的示范性条件已于上文提供。因此，可通过调节与捕获寡核苷酸探针杂交的目标核酸序列的数量控制与捕获寡核苷酸探针杂交的目标核酸片

段的复杂度，这可通过设置或改变影响与捕获寡核苷酸探针杂交的目标核酸片段的数量、变异性或数量和变异性的条件实现。

如本文所使用的短语“半特异性捕获”是指两个或两个以上不同目标核酸片段与可为部分简并或可能不含有任何简并核苷酸碱基的单一捕获寡核苷酸序列结合。半特异性捕获不包括结合所有目标核酸片段或随机结合核酸片段，而是指优先于至少一个其它目标核酸片段结合两个或两个以上目标核酸片段。

在描述阵列中捕获寡核苷酸的核苷酸序列时，使用的术语“唯一”和短语“相同序列”是指严格同一性；因此，当第一个寡核苷酸具有序列 ATCG 且第二个寡核苷酸具有序列 ATCGA 时，这两个寡核苷酸是唯一的，且不具有相同序列。类似地，除非另外指出，否则如本文所使用的，与捕获寡核苷酸杂交的一个或一个以上目标核酸或目标核酸片段是指，单独与多个具有相同序列的捕获寡核苷酸探针中的一个结合的一个或一个以上目标核酸或目标核酸片段的每一个核酸或核酸片段。通常，一个或一个以上目标核酸或目标核酸片段与特定阵列位置处的捕获寡核苷酸杂交。

如本文所使用的短语“部分简并捕获寡核苷酸”是指，与具有类似特异性的至少两个不同核苷酸序列杂交但不结合具有类似特异性的所有可能核苷酸序列的寡核苷酸。例如，部分简并捕获寡核苷酸可为含有通用碱基的寡核苷酸。

如本文所使用的短语“所有理论组合”是指，具有指定长度的寡核苷酸的完整组群，因此能够表示具有所述长度的所有可能核苷酸序列。

如本文所使用的“简并碱基”是指“通用碱基”或“半通用碱基”或可以类似特异性与目标核酸或目标核酸片段的两个或两个以上碱基进行碱基配对的其它碱基。

如本文所使用的“通用碱基”是指可与基因组 DNA 中存在的四个核苷酸中的任何一个结合而无任何实质区别的碱基。用于本文中的示范性通用碱基包括：肌苷；黄苷；3-硝基吡咯（Bergstrom 等人, *Abstr. Pap. Am. Chem. Soc.* 206(2):308 (1993); Nichols 等人, *Nature* 369:492-493; Bergstrom 等人, *J. Am. Chem. Soc.* 117:1201-1209 (1995)); 4-硝基吡咯（Loakes 等人, *Nucleic Acids Res.*, 22:4039-4043 (1994)); 5-硝基吡咯（Loakes 等人 (1994)、6-硝基吡咯（Loakes 等人 (1994)); 硝基咪唑（Bergstrom 等人, *Nucleic Acids Res.* 25:935-1942 (1997)); 4-硝基吡啶（Bergstrom 等人 (1997)); 5-氨基吡啶（Smith 等人, *Nucl. Nucl.* 17:555-564 (1998)); 4-硝基苯并咪唑（Seela 等人, *Helv. Chim. Acta* 79:488-498 (1996)); 4-氨基苯并咪唑（Seela 等人, *Helv. Chim. Acta* 78:833-846 (1995)); 苯基 C-核糖核苷（Millican 等人, *Nucleic Acids Res.* 12:7435-7453(1984); Matulic-Adamic 等人, *J. Org. Chem.* 61:3909-3911 (1996)); 苯并咪唑（Loakes 等人, *Nucl. Nucl.* 18:2685-2695

(1999); Papageorgiou 等人, *Helv. Chim. Acta* 70:138-141 (1987)); 5-氟吡啶 (Loakes 等人, (1999)); 吡啶 (Girgis 等人, *J. Heterocycle Chem.* 25:361-366 (1988)); 无环糖类似物 (acyclic sugar analog) (Van Aerschot 等人, *Nucl. Nucl.* 14:1053-1056 (1995); Van Aerschot 等人, *Nucleic Acids Res.* 23:4363-4370 (1995); Loakes 等人, *Nucl. Nucl.* 15:1891-1904 (1996)), 包括次黄嘌呤衍生物, 咪唑 4,5-二甲酰胺、3-硝基咪唑、5-硝基吡啶; 芳族类似物 (Guckian 等人, *J. Am. Chem. Soc.* 118:8182-8183 (1996); Guckian 等人, *J. Am. Chem. Soc.* 122:2213-2222 (2000)), 包括苯、萘、菲、蒽、吡咯、二氟甲苯; 异噻诺酮核苷 (isocarbostyryl nucleoside) 衍生物 (Berger 等人, *Nucleic Acids Res.* 28:2911-2914 (2000); Berger 等人, *Angew. Chem. Int. Ed. Engl.* 39:2940-2942 (2000)), 包括 MICS、ICS; 氢键合类似物, 包括 N8-吡咯并吡啶 (Seela 等人, *Nucleic Acids Res.* 28:3224-3232 (2000)); 和 LNA, 例如芳基-P-C-LNA (Babu 等人, *Nucleosides, Nucleotides & Nucleic Acids* 22:1317-1319 (2003); WO 03/020739)。

如本文所使用的短语“半通用碱基”是指以相同或相似的特异性优先与 2 个或 3 个脱氧核糖核苷酸结合而不与所有 4 个常见核苷酸 (即, DNA 中为 A、C、G 和 T; RNA 中为 A、C、G 和 U) 结合的碱基。例如, 半通用碱基与 2 个或 3 个常见核苷酸以比与至少一个其它常见核苷酸结合的程度高得多的程度结合。

如本文所使用的“固体载体” (也称为不溶性载体或固体载体) 是指与所关注的分子 (通常为生物分子、有机分子或生物特异性配体) 连接或接触的任何固体或半固体或不溶性载体。所述材料包括用作亲和基质或化学和生物分子合成与分析的载体, 例如 (但不限于) 聚苯乙烯、聚碳酸酯、聚丙烯、尼龙、玻璃、右旋糖酐、几丁质、砂、浮石、琼脂、多糖、树枝状聚合物 (dendrimer)、巴基球 (buckyball)、聚丙烯酰胺、硅、橡胶和用作固相合成、亲和分离和纯化、杂交反应、免疫测定和类似应用的载体的其它材料。

如本文所使用的“部分”核酸 (例如目标核酸或参考核酸) 是指, 不涵盖完整核酸的核苷酸序列或核酸区域。例如, 一部分可为短核苷酸序列, 例如核酸的 SNP、甲基化 C 或微卫星 (microsatellite)。一部分也可例如为具有已知或未知核苷酸序列的核酸的特定片段, 其中所述片段可例如因生物体、品系或物种之间的变异引起的序列差异而产生, 并且其中所述片段是使用本文所公开的方法形成。一部分也可为相对于另一区域不同地相互作用或经不同处理的核酸区域。

B. 核酸分子测序方法

本文提供通过下述步骤对核酸测序的方法:

a) 产生目标核酸的重叠片段;

b) 在不排除错配杂交的条件下, 使所述片段与固体载体上的捕获寡核苷酸阵列杂交以形成捕获片段的阵列;

c) 使用质谱分析法测定各阵列位置处捕获片段的的质量; 和

d) 由一组从各阵列位置获取的质量信号构建所述目标核酸的核苷酸序列。

本文还提供包含以下步骤的核酸测序方法:

a) 产生目标核酸的重叠片段;

b) 使所述片段与固体载体上的捕获寡核苷酸阵列杂交以形成捕获片段的阵列, 其中至少一分组所述捕获寡核苷酸为部分简并寡核苷酸;

c) 使用质谱分析法测定各阵列位置处捕获片段的的质量; 和

d) 由一组从各阵列位置获取的质量信号构建所述目标核酸的核苷酸序列。

本文还提供包含以下步骤的核酸测序方法:

a) 产生目标核酸的重叠片段;

b) 使所述片段与固体载体上的捕获寡核苷酸阵列杂交以形成捕获片段的阵列, 其中至少一个捕获寡核苷酸与两个或两个以上片段杂交;

c) 使用质谱分析法测定各阵列位置处捕获片段的的质量; 和

d) 由一组从各阵列位置获取的质量信号构建所述目标核酸的核苷酸序列。

在本文所提供的每一种方法的某些实施例中, 目标核酸的重叠片段是随机产生。

在本文所提供的每一种方法的另一实施例中, 在步骤 c) 测定捕获片段的的质量之前, 将已杂交的片段再溶解于溶液中。所述再溶解步骤允许众所周知地使用(例如)针型阵列(pin array), 即, 将其浸入含有已再溶解的片段的溶液中以将所述片段转移到用于质谱分析的适当芯片中。

如上文所述, 本文所提供的方法允许比使用与固相芯片结合的目标核酸的 SBH 和/或质谱分析可达成的读取长度长的目标核酸序列读取长度。在另一实施例中, 可通过本文所提供的方法对多种具有较短长度(例如, 200 个、300 个、400 个、500 个、600 个、700 个、800 个、900 个、1,000 个、1,500 个碱基)的目标核酸片段进行测序或分析。本文的方法包括分析 5 个、10 个、15 个、20 个、50 个、100 个、200 个、500 个或更多个核酸片段。当已知特定序列的一部分时, 这些多样较短序列组可例如用于重测序方法中。这些多样较短序列组也可用于多重基因分型、单倍体分型、SNP 和甲基化检测方法中。

C. 目标核酸分子

目标核酸分子可为单链或双链核酸分子。在特定实施例中, 当使用 MALDI-TOF MS 分析时, 或当基于 RNA 转录的方法将增加杂交于芯片上的片段的产量时, 或当与 DNA

捕获寡聚物杂交的 RNA 将于杂交后允许进一步修饰时，使用 RNA 而非 DNA。在另一实施例中，使用 DNA 并使其与 DNA 捕获寡聚物杂交；也可实现对 DNA:DNA 杂合体的杂交后进一步修饰。

1.来源

目标核酸可选自单链 DNA、双链 DNA、cDNA、单链 RNA、双链 RNA、DNA/RNA 杂合体和 DNA/RNA 嵌合核酸 (mosaic nucleic acid)。目标核酸也可包括修饰核酸，例如甲基化 DNA 和含有例如假尿苷的 RNA。可直接从生物样本中分离出目标核酸，或可通过由生物样本扩增或克隆核酸片段得到目标核酸。用作克隆或扩增的模板的目标核酸可为完整无缺的目标核酸或目标核酸片段，其中所述目标核酸片段可具有杂交或质量测量所需的长度，或可具有首先扩增目标核酸片段随后使其经历一个或一个以上其它断裂步骤的中间长度。

用于本文所述的方法中的样本可根据应用本方法的目的进行选择。例如，样本可来自单一个体，其中对个体的样本进行检查以确定一个或一个以上位点处的核苷酸序列。所属领域技术人员可使用本文所述的方法确定待检查的所需样本。

样本可来自任何受检者，包括动物、植物、细菌、病毒、寄生虫、鸟、爬行动物、两栖动物、真菌、鱼和其它植物和动物。受检者为哺乳动物时，通常为人类。来自受检者的样本可为任何形式，包括固体材料，例如组织、细胞、细胞小球、细胞提取物、活组织检查；或生物体液，例如尿液、血液、间质液、腹腔液、血浆、淋巴液、腹水、汗液、唾液、滤泡液、乳汁、非乳汁乳房分泌液 (non-milk breast secretion)、血清、脑脊髓液、粪便、精液、肺部痰液、羊水、感染或发炎区的渗出液、含有口腔细胞的漱口液、滑液或由受检者产生的任何其它流体样本。此外，样本可为所收集的组织，包括骨髓、上皮细胞、胃、前列腺、肾、膀胱、乳房、结肠、肺、胰腺、子宫内膜、神经元和肌肉。样本可包括组织、器官和病理样本，例如石蜡包埋的福尔马林固定样本。

2.制备

所属领域技术人员认识到，一些样本可直接用于本文所提供的方法中。例如，可使用本文所述的方法检查样本，而无需任何纯化或处理步骤来提高所需细胞或核酸分子的纯度。

需要时，可使用已知技术制备样本，例如 Maniatis 等人 (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, N.Y., 第 280-281 页 (1982)) 所述的技术。例如，可在一个或一个以上纯化步骤中处理使用本文所述的方法检查的样本，以提高样本中所需细胞或核酸的纯度。需要时，可使固体材料与流体混合。

众所周知使样本中的核酸与任何生物体或体内的组织或器官以及经培养细胞基本分离的方法。例如，可处理样本以使器官、组织或细胞样本均质化，并且可使用已知溶解缓冲液、超声波处理、电穿孔和已知方法和其组合溶解细胞。如所属领域技术人员所了解，视需要可进行进一步纯化。此外，样本制备可包括随后步骤中可包括的各种试剂。这些试剂包括多种试剂，例如盐、缓冲液、中性蛋白（例如，白蛋白）、清洁剂，和可用于促进最佳杂交或酶反应和/或降低非特异性或背景相互作用的试剂。同样，视样本制备方法和目标核酸分子的纯度而定，可使用另外改进测定效率的试剂，例如蛋白酶抑制剂、核酸酶抑制剂和抗菌剂。

3.目标核酸分子的尺寸和组成

可使用的目标核酸分子的长度可根据目标核酸分子的序列、用于断裂的特定方法、用于杂交的特定捕获寡核苷酸、待测定的核苷酸序列中总目标核酸分子的百分含量、序列测定中所需的准确程度和测序种类（例如，从头测序与再测序）而变化。例如，目标核酸分子的长度可限于能使用本文所公开的断裂和检测方法测定的目标核酸分子的至少约 1%、至少约 3%、至少约 5%、至少约 10%、至少约 20%、至少约 30%、至少约 40%、至少约 50%、至少约 60%、至少约 70%、至少约 80%、至少约 85%、至少约 90%、至少约 95%、至少约 98%、至少约 99%或 100%的核苷酸序列的长度。例如，目标核酸分子可为至少约 20、25、30、35、40、50、60、70、80、90、100、120、140、160、180、200、225、250、275、300、350、400、450、500、550、600、700、800、900、1000、1200、1400、1600、1800、2000、2500 或 3000 个碱基的长度。通常，目标核酸分子不超过约 10,000、5000、4000、3000、2500、2000、1500、1000、900、800、700、600、500、450、400、350、280、260、240、220、200、190、180、170、160、150、140、130、120、110 或 100 个碱基的长度。

4.扩增

在一些实施例中，可扩增目标核酸分子以增加在随后步骤中可处理和测量的核酸分子的数量，并且视情况处理目标核酸序列。可通过聚合酶链反应（PCR）、反转录-聚合酶链反应（RT-PCR）、滚环扩增（rolling circle amplification）、全基因组扩增、链置换扩增（strand displacement amplification, SDA）和基于转录的方法达成扩增。在可产生多种不同扩增产物的多种不同的扩增方法中，扩增方法可具有不同反应条件和/或反应物。

a.反应参数

可进行扩增步骤，其中分离互补链（如果存在），使引物与所述链杂交，并且引物已加入核苷酸中以形成新的互补链。链分离可作为单独步骤或与引物延长产物的合成同

时实现。所述链分离可使用多种适当的变性条件（包括物理、化学或酶方式）完成，词语“变性”包括所有所述方式。一种分离核酸链的物理方法包括加热目标核酸分子直至其变性。典型的加热变性可包括在约 80EC 至 105EC 范围内的温度下历时约 1 至 10 分钟范围内的时间。链分离也可通过化学方式完成，包括高盐条件或强碱性条件。还可通过来自称作解螺旋酶（helicase）的酶类中的酶或具有解螺旋酶活性的酶 RecA 且在 riboATP 存在下诱发链分离，这被称为使 DNA 变性。Kuhn Hoffmann-Berling, *CSH-Quantitative Biology*, 43:63 (1978)中描述了适合用解螺旋酶使核酸链分离的反应条件，并且在 Radding, *Ann. Rev. Genetics* 16:405-437 (1982)中评论了使用 RecA 的技术。

每一个扩增步骤之后，扩增产物通常为双链并且各链彼此互补。可分离互补链，并且可将两条已分离的链用作合成其它核酸链的模板。这一合成可在允许引物与模板发生杂交的条件下进行。合成一般在缓冲水溶液中于通常为约 7-9（例如约 pH 8）的 pH 值下发生。通常，可将摩尔过量的两个寡核苷酸引物加入到含有已分离的模板链的缓冲液中。在一些实施例中，目标核酸的量未知（例如，当将本文所公开的方法用于诊断应用时），因此相对于互补链的量的引物的量无法明确确定。

在示范性方法中，可将脱氧核糖核苷三磷酸 dATP、dCTP、dGTP 和 dTTP 单独或与引物一起加入到合成混合物中，并且可将所得溶液加热到约 90EC-100EC 历时约 1 至 10 分钟，通常历时 1 至 4 分钟。所述加热阶段之后，可使溶液冷却到约室温。可向已冷却的混合物中加入适当的酶用于实现引物延长反应（在本文中称为“聚合反应酶”），并且可使反应在此项技术中已知的条件下发生。这一合成（或扩增）反应可在室温到高于聚合反应酶不再起作用的温度的温度下发生。例如，如果聚合反应酶为热稳定酶，那么此酶也可以在高于室温的温度下使用。在一实施例中，扩增方法是通过如本文所述且如所属领域技术人员常用的 PCR 进行。已描述并且可使用扩增的备选方法。此项技术中已知用于这一目的的各种合适的酶，包括，例如，大肠杆菌（*E. coli*）DNA 聚合酶 I、*E. coli* DNA 聚合酶 I 的 Klenow 片段、T4 DNA 聚合酶、其它可用的 DNA 聚合酶、聚合酶突变蛋白（polymerase mutein）、反转录酶和其它酶，包括热稳定酶（即，在高温下进行引物延长的酶，所述温度通常为引起待扩增的核酸变性的温度）。

b. 修饰核苷

在一实施例中，使用修饰核苷（例如经修饰的核苷三磷酸）扩增目标核酸。一些修饰可通过相应的切割方法赋予或改变目标核酸序列切割特异性。其它修饰（例如质量修饰）可改变目标核酸扩增核酸和其片段的质量。其它核苷可改变聚核苷酸的功能特性，包括（但不限于）增加聚核苷酸对断裂的敏感性、降低进一步延长聚核苷酸的能力。修

饰核苷不必为非天然存在的，而仅仅是通常未被并入特定聚核苷酸中的核苷（例如，当形成 DNA 时为除 A、C、T 和 G 外的核苷；或当形成 RNA 时为除 A、C、U 和 G 外的核苷）。

在一实施例中，使用天然存在但非目标核酸的标准前体（normal precursor）的核苷三磷酸扩增目标核酸。例如，可将一种 rNTP 和三种 dNTP 并入经扩增的聚核苷酸（例如，rCTP、dATP、dTTP 和 dGTP）中。在另一实例中，可通过在标准 DNA 前体核苷酸（例如，dCTP、dATP 和 dGTP）和 dUTP 存在下扩增 DNA 来将 DNA 中通常不存在的脱氧尿苷三磷酸并入经扩增的 DNA 分子中。所述将尿苷并入 DNA 中的步骤可促进 DNA 的碱基特异性切割。例如，当用尿嘧啶-DNA 糖基化酶（uracil-DNA glycosylase, UDG）处理经扩增的含尿苷 DNA 时，尿嘧啶残基被切割。随后化学处理 UDG 反应的产物导致对磷酸酯骨架的切割和核酸碱基（nucleobase）特异性片段的产生。而且，在糖基化酶处理之前，分离扩增产物的互补链使得能够产生互补的断裂模式。因此，使用 dUTP 和尿嘧啶 DNA 糖基化酶允许互补链产生 T 特异性片段，提供指定序列内有关 T 以及 A 位置的信息。

可使用能够终止延长的核苷酸类似物（例如双脱氧核苷酸）进行扩增或诸如转录等其它核苷酸合成反应。在一个实施例中，反应条件包括通常以双脱氧核苷酸形式并入寡核苷酸中的四个核苷酸单体中的一个。在其它实施例中，反应条件包括四个双脱氧核苷酸形式的核苷酸单体中的两个、四个中的三个或所有四个所述核苷酸单体。反应条件可包括核糖核苷酸、脱氧核苷酸和/或双脱氧核糖核苷酸形式的特定核苷酸单体的任何可能的混合物。例如，腺苷（A）可在反应混合物中以 10%核糖核苷酸形式、80%脱氧核苷酸形式和 10%双脱氧核苷酸形式存在。扩增或诸如转录等其它反应无需进行至完成。例如，PCR 中的扩增步骤可在所有引物都已充分延长之前中止，产生具有多种不同长度的目标核酸片段。因此，在一个实施例中，反应可以得到目标核酸异质池的形式进行，提供在延长期间于不同位置处终止的寡核苷酸。

在一个实施例中，一个或一个以上的核苷三磷酸可经在核苷酸之间产生选择性不可水解键的类似物取代。例如，核苷可经 α -巯基-底物（ α -thio-substrate）取代，且随后可通过使用例如烷基卤化物（例如，碘代乙酰胺、碘乙醇）或 2,3-环氧基-1-丙醇等试剂进行烷基化反应来修饰硫代磷酸酯核苷间连键（phosphorothioate internucleoside linkage）。可选择性不可水解的其它示范性核苷包括 2'氟核苷、2'脱氧核苷和 2'氨基核苷。

质量修饰核苷可选自经质量修饰的脱氧核苷三磷酸、经质量修饰的双脱氧核苷三磷酸和经质量修饰的核糖核苷三磷酸。经质量修饰的核苷三磷酸可于碱基、糖和/或磷酸酯

部分上经修饰，并且通过酶步骤、化学步骤或两种步骤的组合引入。一方面，修饰可包括非羟基的 2'取代基。另一方面，可修饰核苷间连键，例如硫代磷酸酯键；或使硫代磷酸酯键进一步与烷基化剂反应。另一方面，可用甲基修饰经修饰的核苷三磷酸，例如 5-甲基胞嘧啶或 5-甲基尿苷。

其它已知的质量修饰部分包括以下述物质取代 H：卤素，如 F、Cl、Br 和/或 I；或假卤素，例如 SCN、NCS；或通过使用不同的烷基、芳基或芳烷基部分，例如甲基、乙基、丙基、异丙基、叔丁基、己基、苯基、经取代苯基、苯甲基；或以下官能团：例如 CH_2F 、 CHF_2 、 CF_3 、 $\text{Si}(\text{CH}_3)_3$ 、 $\text{Si}(\text{CH}_3)_2(\text{C}_2\text{H}_5)$ 、 $\text{Si}(\text{CH}_3)(\text{C}_2\text{H}_5)_2$ 、 $\text{Si}(\text{C}_2\text{H}_5)_3$ 。另一种质量修饰可通过经由核酸分子（例如，检测剂（D））或核苷三磷酸连接同源肽（homopeptide）或异源肽（heteropeptide）获得。可用于产生质量增量为 57 的质量修饰物质的一个实例为连接寡聚甘氨酸（oligoglycine），例如达成 74 ($r = 1, m = 0$)、131 ($r = 1, m = 2$)、188 ($r = 1, m = 3$)、245 ($r = 1, m = 4$) 的质量修饰。也可使用简单的酰胺低聚物(oligoamide)，例如可获得 74 ($r = 1, m = 0$)、88 ($r = 2, m = 0$)、102 ($r = 3, m = 0$)、116 ($r = 4, m = 0$) 等质量修饰。

质量修饰部分可连接到（例如）寡核苷酸的 5'端、核酸碱基（或碱基）、磷酸酯骨架、核苷的 2'位和/或末端 3'位。质量修饰部分的实例包括例如卤素；叠氮基；或具有 XR 型，其中 X 为连接基，且 R 为质量修饰官能团。例如，质量修饰官能团可用于将界定的质量增量引入寡核苷酸分子中，如本文所述。例如以 α -硫基核苷三磷酸在磷酸二酯键处引入修饰的优势在于，这些修饰并不会干扰准确的 Watson-Crick 碱基配对，并且另外允许例如通过烷基化反应对完整核酸分子进行一步骤合成后位点特异性修饰（one-step post-synthetic site-specific modification）（例如参看 Nakamaye 等人., *Nucl. Acids Res.* 16:9947-9959 (1988)）。示范性质量修饰官能团为经硼修饰的核酸，可通过聚合酶将其有效地并入核酸中（例如，参看 Porter 等人. *Biochemistry* 34:11963-11969 (1995)；Hasan 等人, *Nucl. Acids Res.* 24:2150-2157 (1996)；Li 等人 *Nucl. Acids Res.* 23:4495-4501 (1995)）。

而且，例如可通过将质量修饰官能团连接到核苷三磷酸中糖环的 3'位来添加质量修饰官能团，以便影响链终止。对于所属领域技术人员而言，显然许多组合可用于本文所提供的方法中。同样，所属领域技术人员应认识到，链延长的核苷三磷酸也可利用官能团和连接位的多种变异和组合以类似方式进行质量修饰。

可使用不同的质量修饰核苷酸同时检测多种不同的核酸片段。在一实施例中，质量修饰可在扩增过程中并入。在另一实施例中，可通过质量修饰一个或一个以上目标核酸

分子来进行不同目标核酸分子的多重作用，其中视需要可对每一个不同的目标核酸分子进行不同的质量修饰。

c. 扩增方法

可根据所需分析设计使用扩增方法产生多种不同的扩增产物。

在一实施例中，本文提供扩增或其它反应（例如转录）的核苷酸产物，其中产物核苷酸的尺寸可不同，甚至当提供单一模板尺寸时其尺寸也不同。例如，产物核苷酸可重叠，使得两个或两个以上产物核苷酸共用来自天然目标核酸的一个或一个以上核苷酸位。所述重叠核苷酸包括“梯状”核苷酸，其中一系列具有不同尺寸的核苷酸共用同一个核心序列，并且连续较大核苷酸通常在一个或一个以上核苷酸位的增量中仅在核苷酸的3'或5'端含有另外的核苷酸。可使用多种方法形成所述产物，包括（但不限于）以存在于双脱氧核苷与非双脱氧核苷的组合中的四个核苷中的一个进行的核酸合成反应。

在其它实施例中，可使用一种或一种以上与模板目标核酸或模板目标核酸片段中的恒定区与可变区杂交的引物进行扩增或其它核苷酸合成反应。例如，可使用本文所公开的方法断裂目标核酸分子；所述目标核酸片段可已与一种或一种以上接头寡核苷酸接合，由此使具有相同序列的接头寡核苷酸与具有不同序列的两个或两个以上目标核酸片段的相同端（即，3'端或5'端）接合。每种接合产物都含有目标核酸片段和接头寡核苷酸。由于目标核酸片段的部分随片段不同而不同，因此引物可通过与接头寡核苷酸区的至少一部分和一些（但非全部）目标核酸片段的至少一部分杂交而与一些（但非全部）接合产物杂交。随后仅对与接合片段可变区中的引物杂交的一分组目标核酸片段进行扩增或其它核苷酸合成反应。以此方式，一组一种或一种以上引物可用于扩增所有目标核酸片段亚群体，据此，目标核酸片段的可变序列与引物杂交。在一实施例中，仅使用一种引物序列与目标核酸片段的3'端、5'端或3'端与5'端接合。在另一实施例中，使用两种引物与目标核酸片段接合：第一种与3'目标核酸片段接合；且第二种与5'目标核酸片段接合。在另一实施例中，使用两种或两种以上引物与3'端或5'端接合。例如，可使用识别不同恒定区的多种引物，使得第一组引物与第一群体目标核酸片段杂交，且第二组引物与第二群体目标核酸片段杂交；通常，第一群体目标核酸和第二群体目标核酸不具有重叠成员。

选择性核苷酸合成也可与断裂一起进行。通过多个核酸合成循环扩增的目标核酸使用与目标核酸分子的两个单独区杂交的引物。两个引物杂交位点之间的中心区中的目标核酸分子断裂会防止目标核酸分子扩增。因此，即使用于核酸合成反应中的引物不具选择性或不具高选择性，核酸分子中心区的选择性断裂也可导致目标核酸分子的选择性扩

增。

在一实例中，在用核酸合成条件进行处理之前，可用断裂条件处理样本。在这一实例中，断裂条件可选择性切割特定核苷酸序列。例如，样本中可已加入如 EcoRI 的限制性核酸内切酶。这一步骤产生含有含 EcoRI 识别位点的经切割目标核酸分子的样本，和不含 EcoRI 识别位点的完整目标核酸分子。随后，可使用经设计以扩增仅未经切割的目标核酸分子的引物利用核酸合成条件处理样本。作为切割的结果，根据限制性核酸内切酶识别位点的存在，扩增对所有核酸分子的分组具选择性。可用于本文所提供的方法中的断裂条件包括可选择性切割核酸分子的任何断裂条件，包括限制性核酸内切酶。可使用的其它断裂条件包括可通过序列特异性切割的任何断裂条件。

在另一实施例中，可只以核酸扩增方法或加上其它核酸扩增方法进行转录。使用模板 DNA 分子形成 RNA 分子的转录方法可用于扩增目标核酸分子，并将 DNA 形式的目标核酸分子变为 RNA 形式。示范性模板 DNA 包括经扩增的目标核酸分子产物和经处理、未经扩增的目标核酸分子。

如本文所述，使经处理目标核酸分子经历一个或一个以上核酸合成反应。核酸合成反应可用于扩增经处理的目标核酸分子和/或改变核酸分子的形式。在一实施例中，经处理的目标核酸分子或 PCR 产物经过转录。

可对模板 DNA 的一条链或模板 DNA 的两条链进行模板 DNA 的转录，所述模板 DNA 例如为目标核酸分子或其扩增产物。在一实施例中，待转录的核酸分子含有可与能够进行转录的酶结合的部分；所述部分例如可为转录启动子序列。

可使用此项技术中已知的各种方法利用此项技术中已知的各种酶进行转录反应。例如，可将具有并入 dNTP 与 rNTP 的能力的突变体 T7 RNA 聚合酶（T7 R&DNA 聚合酶；Epicentre, Madison, WI）用于转录反应中。转录反应可在此项技术中已知的标准反应条件下执行，所述反应条件例如为：40 mM Tris-Ac (pH 7.5)、10 mM NaCl、6 mM MgCl₂、2 mM 亚精胺、10 mM 二硫苏糖醇、1 mM 各 rNTP、5 mM dNTP（当使用时）、40 mM DNA 模板和 5 U/μL T7 R&DNA 聚合酶，在 37°C 下培育 2 小时。转录后，可将虾碱性磷酸酶（shrimp alkaline phosphatase, SAP）加入切割反应中以降低环单磷酸酯副产物的量。此项技术中已知 T7 R&DNA 聚合酶的使用，如以下文献中所举例说明：美国专利第 5,849,546 号、第 6,107,037 号；和 Sousa 等人, EMBO J. 14:4609-4621 (1995)；Padilla 等人, Nucl. Acid Res. 27:1561-1563 (1999)；Huang 等人, Biochemistry 36:8231-8242 (1997)；以及 Stanssens 等人, Genome Res., 14:126-133 (2004)。

除用四个常规核糖核苷酸底物（rCTP、rATP、rGTP 和 rUTP）进行转录外，还可用

核苷类似物（例如本文所提供和此项技术中已知的那些核苷类似物）或用对应的脱氧核糖核苷三磷酸（例如，用 dCTP 替换 rCTP，或用 dUTP 或 dTTP 替换 rUTP）替换一个或一个以上核糖核苷三磷酸来进行反应。在一实施例中，用核苷或核苷类似物替换一个或一个以上 rNTP，其中所述核苷或核苷类似物在并入经转录核酸后在施加于经转录核酸的断裂条件下不可切割。

在一实施例中，转录是在一个或一个以上核酸合成反应后进行。例如，可在扩增目标核酸分子后进行扩增产物的转录。在另一实施例中，经处理目标核酸分子是在无任何前述核酸合成步骤的情况下进行转录。

在一些方法中，涉及核酸的反应也可包括使双链核酸变性产生单链分子的步骤。例如，可在反应混合物温度超过特定双链核酸的溶解温度的条件下达成变性。

许多核酸反应（例如扩增反应）涉及使核酸杂合体链变性和退火的温度升高和降低的重复循环。于 2002 年 4 月 11 日申请的第 60/372,711 号、2003 年 3 月 24 日申请的第 60/457,847 号和 2003 年 4 月 11 日申请的第 10/412,801 号中提供的装置通过直接、快速和有效地加热和冷却小室内具有相对低质量和高导热性的固体载体底部并通过避免将反应物转移到单独热循环仪中的任何步骤来促进小室内反应混合物温度的变化。

D. 断裂

使用已知方法产生足量目标核酸后，可将目标核酸序列切割成核酸片段。可使用将核酸分子切割成片段的各种方法产生核酸片段。例如，可使用非特异性随机断裂。在一些情况下，断裂方法产生适当的片段尺寸分布。此项技术中已知聚核苷酸的断裂，并且可用多种方式达成。例如，可用物理、化学或酶法断裂由 DNA、RNA、DNA 和 RNA 类似物构成的聚核苷酸。在一实例中，使用物理断裂产生具有不同尺寸的随机目标核酸片段。在另一实例中，可利用一个或一个以上特异性和/或非特异性切割位点处的部分酶切割产生本文所使用的随机目标核酸片段。

在特定实施例中，制备尺寸在 5-50 个碱基、10-40 个碱基、11-35 个碱基和 12-30 个碱基的统计学范围内的目标核酸片段用于本文中。在其它实施例中，例如预期在质谱分析之前“修剪”捕获寡核苷酸:目标-片段复合物的实施例中，所述目标核酸片段可相当大且其尺寸可在包括下述的尺寸范围群组的统计学范围内：20-50 个碱基、30-60 个碱基、40-70 个碱基、50-80 个碱基、60-90 个碱基、70-100 个碱基和更多个碱基。预期本文所使用的其它尺寸范围包括介于约 50 个至约 150 个碱基之间、约 25 个至约 75 个碱基或约 12 个至 30 个碱基。在一特定实施例中，使用具有约 12 个至约 30 个碱基的片段。一般地，选择片段尺寸范围使得较短片段能够以足够强度与捕获寡核苷酸结合，并且以

足够特异性杂交，并且较长片段以足够效率杂交从而使其能够得到表现。同样，在一些实施例中，选择尺寸范围以便促进 MALDI-TOF MS 中所需的解吸效率。

片段尺寸长度和片段尺寸范围可通过本文所提供的各种断裂方法达成。例如，当使用物理断裂方法时，调整施加物理力/应力的参数可产生不同的片段尺寸和范围。在另一实例中，当使用限制性酶时，可利用所使用的限制性酶的量 and 类型和所选择的特定反应条件控制所产生的片段的平均长度。片段的尺寸可变化，且用于本文中的适当片段通常为小于约 500 个、小于约 400 个、小于约 300 个、小于约 200 个核苷酸的长度。

在统计学重叠片段池中，片段与其它片段重叠；例如，重叠片段可与 1 个或 1 个以上、2 个或 2 个以上、3 个或 3 个以上、4 个或 4 个以上、5 个或 5 个以上、6 个或 6 个以上、8 个或 8 个以上、10 个或 10 个以上、15 个或 15 个以上、20 个或 20 个以上其它片段重叠，并且通常与至少 2 个、至少 3 个、至少 4 个、至少 5 个、至少 6 个、至少 8 个、至少 10 个、至少 15 个或至少 20 个其它片段重叠。

重叠片段为与未经断裂的目标核酸分子共有 1 个或一个以上核苷酸位的片段。因此，重叠片段包括第一个片段含有位于第二个片段中的所有核苷酸位并且第一个片段的 5'端、3'端或 5'端与 3'端含有另外的核苷酸位的片段。重叠片段还包括第一个片段的 3'端与第二个片段的 5'端重叠的片段。重叠片段仅需要重叠一个核苷酸位；然而，统计学重叠片段池也可重叠至少 2 个、至少 3 个、至少 4 个、至少 5 个、至少 6 个、至少 8 个、至少 10 个、至少 15 个或至少 20 个核苷酸位。

1. 聚核苷酸的酶断裂

核酸分子片段可由酶切割单链或多链核酸分子产生。多链核酸分子包括含有一个以上核酸分子链的核酸分子复合物，包括例如双链和三链核酸分子。视所使用的酶而定，可非特异性地或在特定核苷酸序列处切割核酸分子。可以使用能够切割核酸分子的任何酶，包括（但不限于）核酸内切酶、核酸外切酶、单链特异性核酸酶、双链特异性核酸酶、核糖酶和 DNA 酶。此项技术中已知断裂核酸分子的多种酶，并且市面有售，例如核酸酶 BAL-31、绿豆核酸酶（mung bean nuclease）、核酸外切酶 I、核酸外切酶 III、核酸外切酶 VIII、 λ 核酸外切酶、T7 核酸外切酶、核酸外切酶 T、RecJ、RNA 酶 I、RNA 酶 III、RNA 酶 A、RNA 酶 U2、RNA 酶 T1、RNA 酶 H、Shortcut RNA 酶 III、Acc I、BasA I、BtgZ I、Mfe I、Sac I、N.BbvC IA、N.BbvC IB、N.BstNBI、I-CeuI、I-SceI、PI-PspI、PI-SceI、McrBC 和其它已知酶（例如，参看 New England Biolabs, Inc. Catalog; Sambrook, J., Russell, D.W., *Molecular Cloning: A Laboratory Manual*, 第 3 版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001）。也可使用酶将大核酸分子降解成

较小片段。可单独使用或组合使用本文所提供的酶而产生重叠目标核酸分子片段。重叠片段的产生可通过多种不同方法达成。例如，可使用非特异性 RNA 酶（RNA 酶 I）或非特异性 DNA 酶（DNA 酶 I）进行有限消化/部分消化。

a.核酸内切酶断裂

核酸内切酶是用于断裂核酸分子的示范性酶类。核酸内切酶切割核酸分子链内的键。核酸内切酶可对双链或单链核酸分子具有特异性。切割可于核酸分子内随机发生或于特定序列处发生。随机切割双链核酸分子的核酸内切酶通常与核酸分子的骨架相互作用。核酸分子的特异性断裂可于连续反应中或同时使用一种或一种以上酶实现。可切割同源或异源核酸分子。核酸内切酶也可切割单链核酸；例如，S1 或绿豆核酸酶可使单链 DNA（绿豆）或 DNA 或 RNA（S1）降解产生平端双链核酸分子。

限制性核酸内切酶为识别双链核酸分子内的特定序列并通常切割识别序列内或其附近的两条链的核酸内切酶亚类。DNA 分析中一种常用的酶是 HaeIII，于序列 5'-GGCC-3'处切割 DNA。其它示范性限制性核酸内切酶包括 Acc I、Afl III、Alu I、Alw44 I、Apa I、Asn I、Ava I、Ava II、BamH I、Ban II、Bcl I、Bgl I、Bgl II、Bln I、Bsm I、BssH II、BstE II、Cfo I、Cla I、Dde I、Dpn I、Dra I、EclX I、EcoR I、EcoR I、EcoR II、EcoR V、Hae II、Hae III、Hind II、Hind III、Hpa I、Hpa II、Kpn I、Ksp I、Mlu I、MluN I、Msp I、Nci I、Nco I、Nde I、Nde II、Nhe I、Not I、Nru I、Nsi I、Pst I、Pvu I、Pvu II、Rsa I、Sac I、Sal I、Sau3A I、Sca I、ScrF I、Sfi I、Sma I、Spe I、Sph I、Ssp I、Stu I、Sty I、Swa I、Taq I、Xba I、Xho I。此项技术中已知这些酶的切割位点。也涵盖从识别位点的下游切割的 IIS 型限制性核酸内切酶。

视所使用的酶而定，对核酸分子的切割可产生一条悬突于另一条的链，又称“粘”端。例如，BamH I 产生粘性 5'悬端，而 Kpn I 产生粘性 3'悬端。另外，切割可产生不具有悬端的“平”端。例如，Dra I 切割产生平端。限制性酶可切割含有特定核苷酸序列的核酸分子，而不切割不含有所述核苷酸序列的核酸分子。在一些情况下，可通过甲基化遮蔽切割识别位点。

限制性核酸内切酶可用于产生各种核酸分子片段尺寸。例如，CviJ I 是识别两碱基与三碱基 DNA 序列的限制性核酸内切酶。用 CviJ I 完全消化可产生平均长度为 16 至 64 个核苷酸的 DNA 片段。因此，用 CviJ I 部分消化可以与剪切或超声波处理类似的“准”随机方式（"quasi" random fashion）断裂 DNA。CviJ I 通常切割 G 与 C 之间的 RGCY 位点，留下易于克隆的平端，其中 R 为任何嘌呤且 Y 为任何嘧啶。在 1 mM ATP 和 20% 二甲亚砜存在下，切割特异性松弛，CviJ I 也切割 RGCN 和 YGCY 位点。在这些“星级

(star)”条件下，CviJ I 切割产生准随机消化 (quasi-random digest)。可在此时选择经消化或经剪切 DNA 的尺寸。

此项技术中普遍已知使用限制性核酸内切酶断裂核酸分子的方法。在一示范性方案中，制备 20-50 μl 含有以下成分的反应混合物：DNA 1-3 μg ；限制性酶缓冲液 1 \times ；和用于 1 μg DNA 的限制性核酸内切酶 2 单位。此项技术中也已知适当的缓冲液，并且包括适当的离子强度、辅助因子和视情况 pH 缓冲液以提供酶活性的最佳条件。特定酶可能需要一般从所述酶的生产商购得的特定缓冲液。示范性缓冲液为谷氨酸钾缓冲液 (KGB)。Hannish, J.和 M. McClelland, "Activity of DNA modification and restriction enzymes in KGB, a potassium glutamate buffer," *Gene Anal. Tech* 5:105 (1988); McClelland, M.等人 "A single buffer for all restriction endonucleases," *Nucl. Acids Res.* 16:364 (1988)。在 37EC 下培育反应混合物 1 小时或为产生具有所需尺寸或尺寸范围的片段所需的任何时间。视需要，可通过在 65EC 或 80EC 下加热混合物来终止反应。或者，可通过用例如 EDTA 螯合如 Mg^{2+} 的二价阳离子终止反应。

在特定实施例中，可使用一种以上酶断裂核酸分子。可将多种酶用于同一反应中，只要所述酶在类似条件下具有活性即可，所述条件例如为离子强度、温度或 pH 值；或可于连续反应中使用多种酶。通常，多种酶是与例如 KGB 的标准缓冲液一起使用。当使用限制性酶时，可部分或完全消化核酸分子。

DNA 酶也可用于产生核酸分子片段。Anderson, S., "Shotgun DNA sequencing using cloned DNase I-generated fragments," *Nucl. Acids Res.* 2:3015-3027 (1981)。DNA 酶 I (脱氧核糖核酸酶 I) 是一种将双链和单链 DNA 非特异性地消化成聚核苷酸和单核苷酸的核酸内切酶。这类酶能够对单链以及双链 DNA 和染色质起作用。

II 型脱氧核糖核酸酶在核酸研究中用于许多应用，包括在酸性 pH 值下的 DNA 测序和消化。来自猪脾的脱氧核糖核酸酶 II 具有 38,000 道尔顿 (dalton) 的分子量。这类酶是具有二聚体结构的糖蛋白核酸内切酶。在 0.15 M 离子强度下，最佳的 pH 值范围为 4.5-5.0。脱氧核糖核酸酶 II 水解天然和变性 DNA 中的脱氧核糖核苷酸键，产生具有 3'-磷酸的产物。在 pH 值 5.6-5.9 下，也对对硝基苯基磷酸二酯起作用。Ehrlich, S.D.等人 "Studies on acid deoxyribonuclease. IX. 5'-Hydroxy-terminal and penultimate nucleotides of oligonucleotides obtained from calf thymus deoxyribonucleic acid," *Biochemistry* 10(11):2000-2009 (1971)。

核酸内切酶可对特定类型的核酸分子具有特异性。例如，核酸内切酶可对 DNA 或 RNA 或单链或双链核酸分子具有特异性。核酸内切酶可具有序列特异性或非序列特异

性。例如，核糖核酸酶 H 是一种特异性降解 RNA-DNA 杂合体中的 RNA 链的内切核糖核酸酶。核糖核酸酶 A 是一种特异性攻击 C 和 U 残基处的单链 RNA 的内切核糖核酸酶。核糖核酸酶 A 催化切割核苷酸 5'-核糖与连接到相邻嘧啶核苷酸 3'-核糖的磷酸基团之间的磷酸二酯键。所得 2',3'-环磷酸可水解成相应的 3'-核苷磷酸。RNA 酶 T1 仅消化 G 核糖核苷酸处的 RNA，于鸟苷酸残基的 3'-羟基与侧翼核苷酸 5'-羟基之间进行切割。RNA 酶 U₂ 仅消化 A 核糖核苷酸处的 RNA。碱基特异性消化的实例可见于 Stanssens 等人的公开案 WO 00/66771 中。

BenzonaseJ、核酸酶 P1 和磷酸二酯酶 I 是适于产生 200 个碱基对或更少个碱基对的范围内的核酸分子片段的非特异性核酸内切酶。BenzonaseJ (Novagen, Madison, WI) 是一种降解所有形式的 DNA 和 RNA (单链、双链、线形和环形) 并且可在广泛范围的操作条件下使用的经基因工程改造的核酸内切酶。这类酶将核酸完全消化成 2-5 个碱基长的以 5'-单磷酸封端的寡核苷酸。BenzonaseJ 的核苷酸和氨基酸序列提供于美国专利第 5,173,418 号中。如本文所提供的方法中核酸的断裂也可由二核苷酸 (“2 切割酶 (2 cutter)”) 或松弛态二核苷酸 (“1-1/2 切割酶” 或 “1-1/4 切割酶”) 切割特异性达成。二核苷酸特异性切割试剂已为所属领域技术人员所知 (例如参看 WO 94/21663; Cannistraro 等人, *Eur. J. Biochem.* 181:363-370 (1989); Stevens 等人, *J. Bacteriol.* 164:57-62 (1985); Marotta 等人, *Biochemistry* 12:2901-2904 (1973))。

使用限制性核酸内切酶进行的切割可为部分切割, 和/或使用随机并入限制性核酸内切酶识别位点中的修饰核苷酸进行改变。这些修饰核苷酸相对于标准核苷酸显示出对于切割的不同敏感性。这一不同敏感性可包括增加的切割倾向, 也可包括降低的切割倾向, 包括对切割的完全抵抗。例如, 可将对酶切割具有抗性的脱氮杂核苷酸 (deaza nucleotide) 部分且随机地并入限制性核酸内切酶的识别位点中, 引起部分切割, 即使执行限制性核酸内切酶反应至完成时也为部分切割。在另一实例中, 可将脱氧尿苷并入 DNA 核苷酸中, 并且可使用尿嘧啶-DNA 糖基化酶除去尿嘧啶, 随后在这一位置处切割 DNA; 因此, 将尿苷并入 DNA 中可显示出增加的切割倾向。在另一实例中, 可以常规与 α -巯基-底物的混合物合成所关注的目标核酸分子的转录本, 随后可通过使用例如烷基卤化物 (例如, 碘代乙酰胺、碘乙醇) 或 2,3-环氧基-1-丙醇的试剂进行烷基化来修饰硫代磷酸酯核苷间连键。预期通过所述修饰形成的硫代磷酸酯键不为 RNA 酶的底物。不被 RNA 酶切割的其它示范性核苷酸包括 2'氟核苷酸、2'脱氧核苷酸和 2'氨基核苷酸。在使用此程序的一个实例中, 视所需切割特异性而定, 可通过并入例如 2'-修饰形式的 C 核苷酸或 U 核苷酸的不可水解核苷酸而将 RNA 酶 A 的切割特异性限于 CpN 或 UpN 二核苷酸。因此,

在一实例中，可通过将 α S-dUTP、 α S-ATP、 α S-CTP 和 GTP 核苷酸并入转录本中制备转录本（目标分子）。可通过使用例如 RNA 酶-U2 和 RNA 酶-T1 的其它 RNA 酶来进一步扩充有用二核苷酸特异性切割试剂的清单。在例如 RNA 酶-T1 的单特异性 RNA 酶的情况下，使用不可切割核苷酸可将 GpN 键的切割限于四个可能 GpN 键的任何三个、两个或一个，这视所选择的不可切割核苷酸而定。也可使用这些选择性修饰策略通过选择性修饰均聚物系统内的某些核苷酸使经修饰核苷酸对切割具有较弱或较强抗性来防止对均聚物系统中每一个碱基处的切割。

b. 核酸外切酶断裂

可使用将不同长度的碱基从聚核苷酸末端除去的核酸酶（称为核酸外切酶）来将聚核苷酸断裂成小聚核苷酸。核酸外切酶可断裂双链核酸或可断裂单链核酸。可断裂单链或双链核酸的示范性核酸外切酶为 Bal 31 核酸酶。

核酸外切酶可以从多种聚核苷酸的末端切割核苷酸。例如，存在 5'核酸外切酶（从 DNA 链的 5'端切割 DNA）和 3'核酸外切酶（从所述链的 3'端切割 DNA）。不同的核酸外切酶可水解单链或双链 DNA。例如，核酸外切酶 III 为 3'至 5'核酸外切酶，将 5'-单核苷酸从 DNA 链的 3'端释放出来；其为 DNA 3'-磷酸酶，水解 3'-末端磷酸单酯；且其为 AP 核酸内切酶，切割无嘌呤或无嘧啶位点处的磷酸二酯键，产生无碱基的脱氧核糖 5'-磷酸残基的 5'端。此外，所述酶具有 RNA 酶 H 活性；其可能以核酸外切方式优先降解 DNA-RNA 杂交双链中的 RNA 链。在 S1 哺乳动物细胞中，主要的 DNA 3'-核酸外切酶是 DNA 酶 III（也称为 TREX-1）。因此，可通过使用核酸外切酶降解聚核苷酸末端形成片段。

c. 核酸酶断裂（Nucleic Acid Enzyme Fragmentation）

此项技术中已知催化性 DNA 和 RNA 且可用于切割核酸分子产生核酸分子片段。Santoro, S. W.和 Joyce, G. F. "A general purpose RNA-cleaving DNA enzyme," *Proc, Natl. Acad. Sci. USA* 94:4262-4266 (1997)。呈单链分子形式的 DNA 可折叠成与 RNA 类似的三维结构，而 2'-羟基对于催化作用是不必要的。作为核糖酶，也可通过选择使 DNA 酶依赖于辅助因子。这已由用于 RNA 水解的组氨酸依赖性 DNA 酶得到证实。美国专利第 6,326,174 号和第 6,194,180 号公开了能够切割核酸序列或分子（尤其 RNA）的脱氧核糖核酸酶，即催化活性和酶活性 DNA 分子。

此项技术中已知使用核糖酶切割核酸分子。核糖酶为催化化学反应（例如切割共价键）的 RNA。Uhlenbeck 说明了一种小的活性核糖酶，即锤头核糖酶（hammerhead ribozyme），其中催化链和底物链被分离（Uhlenbeck, *Nature* 328:596-600 (1987)）。所述

核糖酶通过碱基配对相互作用与底物 RNA 结合，切割已结合的目标 RNA，释放切割产物，并进行再循环，使得可重复这一过程多次。Haseloff 和 Gerlach 列举了有关能够反式起作用的简单锤头核糖酶的通用设计规则（Haseloff 等人, *Nature*, 334:585-591 (1988)）。已研发出多种具有高切割特异性的不同锤头核糖酶，并且此项技术中已知有关设计具有所需底物特异性的锤头核糖酶的通用方法，如美国专利第 5,646,020 号和第 6,096,715 号中所举例说明。另一类具有反式切割活性的核糖酶为得自肝炎 δ 病毒基因组的 δ 核糖酶。Ananvoranich 和 Perrault 已描述关于 δ 核糖酶切割的底物特异性的因子（Ananvoranich 等人, *J. Biol. Chem.* 273:13812-13188 (1998)）。发夹核糖酶（hairpin ribozyme）也可用于反式切割，并且也已知有关发夹核糖酶的底物特异性的原则（例如，参看 Perez-Ruiz 等人, *J. Biol. Chem.* 274:29376-29380 (1999)）。所属领域技术人员可使用已知的底物特异性原则选择核糖酶并设计核糖酶序列以达成所需的核酸分子切割特异性。

可使用 DNA 切口酶或 DNA 酶识别和切割 DNA 双链中的一条链。已知多种切口酶。这些酶例如为具有以下切割位点的 NY2A 切口酶和 NYS1 切口酶（百万碱基（Megabase））：

NY2A: 5'...R AG...3'

3'...Y TC...5', 其中 R = A 或 G 且 Y = C 或 T;

NYS1: 5'... CC[A/G/T]...3'

3'... GG[T/C/A]...5'。

随后化学处理切口酶反应的产物导致对磷酸骨架的切割和片段的产生。

Fen-1 断裂方法涉及 Fen-1 酶，这种酶是称作“瓣状”核酸内切酶的位点特异性核酸酶（U.S. 5,843,669、5,874,283 和 6,090,606）。此酶识别并切割由与目标 DNA 链杂交的两个寡核苷酸的重叠所产生的 DNA “瓣”。这一切割具有高度特异性，并且可识别单个碱基变异，允许检测所关注的核苷酸位点处的单个甲基化碱基。Fen-1 酶可以是类 Fen-1 核酸酶（Fen-1 like nuclease），例如人类、鼠类和爪蟾（*Xenopus*）XPG 酶和酵母 RAD2 核酸酶；或来自例如詹氏甲烷球菌（*M. jannaschii*）、嗜热古细菌（*P. furiosus*）和伍斯氏火球菌（*P. woesei*）的 Fen-1 核酸内切酶。

可使用的另一种技术为 DNA 嵌合体的切割。使三重 DNA-RNA-DNA 探针与例如结核分支杆菌（*M. tuberculosis*）特异性序列的目标核酸分子杂交。加入 RNA 酶 H 后，嵌合探针的 RNA 部分被降解，释放出 DNA 部分（Yule, *Bio/Technology* 12:1335 (1994)）。

d. 碱基特异性断裂

可使用于特定碱基（例如，对于 DNA 而言为 A、C、T 或 G 且对于 RNA 而言为 A、

C、U 或 G) 或碱基类型 (即, 嘧啶或嘌呤) 处选择性切割的核酸酶断裂目标核酸分子。在一实施例中, 可使用特异性切割 3 种 RNA 核苷酸 (例如, U、G 和 A)、2 种 RNA 核苷酸 (例如, C 和 U) 或 1 种 RNA 核苷酸 (例如, A) 的 RNA 酶对目标核酸分子的转录本进行碱基特异性切割。例如, RNA 酶 T1 切割 G 核糖核苷酸处的 ssRNA (单链 RNA); RNA 酶 U2 消化 A 核糖核苷酸处的 ssRNA; RNA 酶 CL3 和黄瓜酶 (cusativin) 切割 C 核糖核苷酸处的 ssRNA; PhyM 切割 U 和 A 核糖核苷酸处的 ssRNA; 且 RNA 酶 A 切割嘧啶核糖核苷酸 (C 和 U) 处的 ssRNA。此项技术中已知例如 RNA 酶 T₁ (G 特异性) 和 RNA 酶 U₂ (A 特异性) 的单特异性 RNA 酶的使用 (Donis-Keller 等人, *Nucleic Acids Res.* 4:2527-2537 (1977); Gupta 和 Randerath, *Nucleic Acids Res.* 4:1957-1978 (1977); Kuchino 和 Nishimura, *Methods Enzymol.* 180:154-163 (1989); 和 Hahner 等人, *Nucl. Acids Res.* 25(10): 1957-1964 (1997))。已报道另一种酶, 鸡肝核糖核酸酶 (RNA 酶 CL3), 优先在胞苷处切割, 但已报道所述酶对此碱基的倾向受反应条件影响 (Boguski 等人, *J. Biol. Chem.* 255:2160-2163 (1980))。报道还提出从黄瓜 (*Cucumis sativus L*) 干种子分离出来的另一种核糖核酸酶黄瓜酶的胞苷特异性 (Rojo 等人, *Planta* 194:328-338 (1994))。另外, 已证实通过使用 RNA 酶 PhyM (A 和 U 特异性) (Donis-Keller, *H. Nucleic Acids Res.* 8:3133-3142 (1980)) 和 RNA 酶 A (C 和 U 特异性) (Simoncsits 等人, *Nature* 269:833-836 (1977); Gupta 和 Randerath, *Nucleic Acids Res.* 4:1957-1978 (1977)) 鉴定嘧啶残基。Stanssens 等人, WO 00/66771 中提供了所述切割模式的实例。

此外, 可例如通过将修饰核苷酸并入核酸中并切除所述核苷酸的碱基来靶向碱基; 随后, 在适当条件下或以酶处理核酸可引起已切除碱基位点处核酸断裂。例如, 可将 dUTP 并入 DNA 中, 并且可通过使用 UDG 除去尿嘧啶碱基随后在已知切割条件下切割 DNA 来实现碱基特异性断裂。在另一实例中, 可将甲基-胞嘧啶并入 DNA 中, 并且可使用甲基胞嘧啶糖基化酶除去甲基胞嘧啶随后在已知条件下进行处理以导致 DNA 断裂来实现碱基特异性断裂。碱基特异性断裂可用于部分切割反应 (包括当目标核酸分子含有已并入的不可切割核苷酸时进行至完成的部分切割反应) 和完全切割反应中。

此项技术中已知使用 RNA 酶的碱基特异性切割反应条件, 例如可包括 4 mM Tris-Ac (pH 8.0)、4 mM KAc、1 mM 亚精胺、0.5 mM 二硫苏糖醇和 1.5 mM MgCl₂。

在一实施例中, 可将扩增产物转录成单链 RNA 分子, 随后通过内切核糖核酸酶进行碱基特异性切割。在一实施例中, 目标核酸分子转录可得到可使用特异性 RNA 核酸内切酶切割的 RNA 分子。例如, 可使用两种不同的内切核糖核酸酶 (例如 RNA 酶 T1 和 RNA 酶 A) 对 RNA 分子进行碱基特异性切割。RNA 酶 T1 特异性切割 G 核苷酸, 且

RNA 酶 A 特异性切割嘧啶核糖核苷酸（即，胞嘧啶和尿嘧啶残基）。在一实施例中，当使用切割一种以上核苷酸的酶（例如 RNA 酶 A）进行切割时，可在目标核酸分子或扩增产物转录期间并入不可切割核苷，例如 dNTP。例如，可在扩增产物转录期间并入 dCTP，并且可使所得经转录核酸经历由 RNA 酶 A 于 U 核糖核苷酸处进行的切割，但对于 RNA 酶 A 于 C 脱氧核糖核苷酸处进行的切割具有抗性。在另一实例中，可在目标核酸分子转录期间并入 dTTP，并且可使所得经转录核酸经历由 RNA 酶 A 于 C 核糖核苷酸处进行的切割，但对于 RNA 酶 A 于 T 脱氧核糖核苷酸处进行的切割具有抗性。通过选择性使用例如 dNTP 的不可切割核苷且通过使用例如 RNA 酶 A 和 RNA 酶 T1 的 RNA 酶进行碱基特异性切割，可对相同目标核酸序列的不同转录本进行对三种不同核苷酸碱基具有特异性的碱基切割。例如，可使特定目标核酸分子的转录本经历使用 RNA 酶 T1 进行的 G 特异性切割；可使转录本在转录反应中经历使用 dTTP 进行的 C 特异性切割，随后经历用 RNA 酶 A 进行的消化；并且可使转录本在转录反应中经历使用 dCTP 进行的 T 特异性切割，随后经历用 RNA 酶 A 进行的消化。

在另一实施例中，使用 dNTP、不同 RNA 酶和目标核酸分子的两个方向可允许六种不同的切割方案。例如，双链目标核酸分子可得到两种不同的单链转录产物，可称之为目标核酸分子的正链转录产物和目标核酸分子的反链转录产物。两种不同转录产物中的每一种可经历三种单独的碱基特异性切割反应（例如 G 特异性切割、C 特异性切割和 T 特异性切割，如本文所述），产生六种不同的碱基特异性切割反应。六种可能的切割方案列于表 1 中。使用四种不同的碱基特异性切割反应可得到有关一条目标核酸分子链中所有四个核苷酸碱基的信息。考虑到可通过切割反链上的互补碱基来模仿正链的切割，可参考反链的切割达成对正链的四个核苷酸中每一个的碱基特异性切割。例如，可对目标核酸分子正链的转录本执行三种碱基特异性切割反应，产生目标核酸分子正链的 G、C 和 T 特异性切割；第四个碱基特异性切割反应可为目标核酸分子反链转录本的 T 特异性切割反应，结果与目标核酸分子正链转录本的 A 特异性切割相同。所属领域技术人员将了解，可使用可能的碱基特异性切割反应（包括表 1 中关于 RNA 酶 T1 和 A 所提供的切割反应）的多种不同组合实现碱基特异性切割以便得到有关一条目标核酸分子链所有四个核苷酸碱基的信息，并且可利用此项技术中已知或本文所公开的其它碱基特异性 RNA 酶执行正链或反链和/或使用不可水解核苷酸的其它切割反应。

表 1

	正向引物	反向引物
RNA 酶 T1	G 特异性切割	G 特异性切割
RNA 酶 A; dCTP	T 特异性切割	T 特异性切割
RNA 酶 A; dTTP	C 特异性切割	C 特异性切割

在一实例中，RNA 酶 U2 可用于碱基特异性切割目标核酸分子转录物。RNA 酶 U2 可于 A 核苷酸处碱基特异性切割 RNA。因此，通过使用 RNA 酶 T1、U2 和 A 且通过使用适当 dNTP（与 RNA 酶 A 一起使用），可通过碱基特异性切割仅一条目标核酸分子链的转录本来检查目标核酸分子所有四个碱基的位置。在一些实施例中，当使用仅碱基特异性切割四个核糖核苷酸中的一个核糖核苷酸的 RNA 酶进行碱基特异性切割时，无需不可切割的核苷三磷酸。例如，使用 RNA 酶 T1、RNA 酶 CL3、黄瓜酶或 RNA 酶 U2 进行碱基特异性切割不需要目标核酸分子转录本中存在不可切割的核苷酸。使用例如 RNA 酶 T1 和 RNA 酶 U2 的 RNA 酶可得到有关目标核酸分子中所有四个核苷酸碱基的信息。例如，可合成目标核酸分子或扩增产物的正链与反链的转录本，且可使用 RNA 酶 T1 和 RNA 酶 U2 对各转录本进行碱基特异性切割。所得四个切割反应的切割模式产生有关目标核酸分子一条链中所有四个核苷酸碱基的信息。在所述实施例中，可进行两种转录反应：第一种为目标核酸分子正链的转录且第二种为目标核酸分子反链的转录。

也涵盖用于所述方法中的多种不同的碱基特异性切割方法。此项技术中已知多种不同的碱基特异性切割方法且已于本文中描述，包括酶催化的 RNA 的碱基特异性切割、酶催化的经修饰 DNA 的碱基特异性切割和以化学方法进行的 DNA 碱基特异性切割。例如，此项技术中已知酶催化的碱基特异性切割，例如使用尿嘧啶-去糖基化酶（uracil-deglycosylase, UDG）或甲基胞嘧啶去糖基化酶（methylcytosine deglycosylase, MCDG）切割，且已于本文中描述，且其可与本文所述的酶催化的 RNA 酶介导的碱基特异性切割反应一起进行。本文中另外涵盖使用碱基特异性切割反应断裂含有不可水解碱基的核酸，例如 RNA，由此产生部分全碱基特异性切割反应（partially complete base specific cleavage reaction）。

2. 聚核苷酸的物理断裂

核酸分子的断裂可使用物理或机械力（包括机械剪切力和超声波处理）达成。可例如使用流体动力（hydrodynamic force）实现核酸分子的物理断裂。通常，通过将含有核酸分子的溶液反复推入和抽出配备有针的注射器来剪切溶液中的核酸分子。Thorstenson,

Y.R.等人 "An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing," *Genome Research* **8**:848-855 (1998); Davison, P. F. *Proc. Natl. Acad. Sci. USA* **45**:1560-1568 (1959); Davison, P. F. *Nature* **185**:918-920 (1960); Schriefer, L. A.等人 "Low pressure DNA shearing: a method for random DNA sequence analysis," *Nucl. Acids Res.* **18**:7455-7456 (1990)。例如用皮下注射针 (hypodermic needle) 剪切 DNA 通常产生大量在 1-2 kb 范围内的片段, 但少量片段可小到 300 bp。

剪切核酸分子 (包括例如基因组 DNA) 的装置市面有售。示范性装置使用注射泵通过将 DNA 样本推过小的突然收缩部分来产生流体剪切力。Thorstenson, Y.R.等人 "An Automated Hydrodynamic Process for Controlled, Unbiased DNA Shearing," *Genome Research* **8**:848-855 (1998)。剪切的体积通常为 100-250 μL , 且处理时间小于 15 分钟。可通过计算机控制使样本的剪切完全自动化。

由 Oefner 等人所研发的流体动力点汇剪切法 (hydrodynamic point-sink shearing method) 是一种利用流体动力剪切核酸分子的方法。Oefner, P. J.等人 "Efficient random subcloning of DNA sheared in a recirculating point-sink flow system," *Nucl. Acids Res.* **24**(20):3879-3886 (1996)。“点汇 (point-sink)”是指所述系统中流体流动的理论模型。应变率张量 (rate-of-strain tensor) 描述施加于分子上并因此使其断裂的力。DNA 断裂是由所述张量的“剪切”形式引起, 这一类断裂方法称为剪切。断裂可通过剪切形式 (当流体在狭窄管或孔内时) 与拉伸应变形式 (当流体接近孔时) 引起。点汇剪切是通过用泵 (例如 HPLC 泵) 施加压力迫使核酸分子 (例如 DNA) 通过极小直径的管道实现。所得片段具有严格的尺寸范围, 其中最长片段大约是最短片段的两倍长。片段尺寸与流动速率成反比。

也可例如通过混合、掺合、搅拌或涡旋溶液来搅动溶液中的大核酸分子而获得核酸分子片段。Hershey, A. D.和 Burgi, E. *J. Mol. Biol.* **2**:143-152 (1960); Rosenberg, H. S.和 Bendich, A. *J. Am. Chem. Soc.* **82**:3198-3201 (1960)。可搅动溶液不同时间长度直至获得具有所需尺寸或尺寸范围的片段。将珠粒或微粒加入溶液中可帮助断裂核酸分子。

一种物理断裂核酸分子的适当方法是基于超声波处理核酸分子。Deininger, P. L. "Approaches to rapid DNA sequence analysis," *Anal. Biochem.* **129**:216-223 (1983)。通常, 通过将含有经缓冲的核酸分子的微离心管放入超声波仪 (例如杯型破碎头超声波仪 (cup-horn sonicator)) 中的冰水浴中, 并使用最大输出和连续功率超声波处理历经不定次数的短脉冲来进行超声波处理而产生核酸分子片段。短脉冲可持续约 10 秒。例如, 参看 Bankier, A.T.等人 "Random cloning and sequencing by the M13/dideoxynucleotide

chain termination method," *Meth. Enzymol.* **155**:51-93 (1987)。在一个示范性超声波处理方案中，超声波处理大核酸分子视超声波处理的条件而定产生在 300-500 bp 或 2-10 kb 范围内的片段，所述处理条件例如持续时间和超声波强度。Kawata, Y.等人 "Preparation of a Genomic Library Using TA Vector," *Prep. Biochem & Biotechnol.* **29**(1):91-100 (1999)。

超声波处理期间，温度上升会导致片段分布型式不均匀，为此，可小心监测冰水浴温度，并且需要时可加入新鲜冰水。确定超声波处理的具体条件的示范性超声波处理方案包括：将 350 μ l 适当缓冲液中的大约 100 μ g 核酸分子样本分成十份 35 μ l 的等分试样，五份经历超声波处理递增次数的 10 秒脉冲。通过在每 10 秒脉冲之间将所述管放置于冰水浴中至少 1 分钟来冷却核酸分子样本。需要时，可在各样本之间更换超声波仪中的冰水浴。可离心样本以回收浓缩物，并与尺寸标记 (size marker) 相比较，对等分试样进行琼脂糖凝胶电泳测定。因此，基于由琼脂糖凝胶电泳所检测的片段尺寸范围，可超声波处理剩余 5 个管从而获得所需片段尺寸。

也可使用喷雾器达成核酸分子的断裂。Bodenteich, A., Chissoe, S., Wang, Y.-F.和 Roe, B. A. (1994) In Adams, M. D., Fields, C.和 Venter, J. C. (编辑) Automated DNA Sequencing and Analysis, Academic Press, San Diego, CA。此项技术中已知喷雾器且市面有售。使用喷雾器断裂核酸分子的示范性方案包括：将 2 ml 经缓冲的含有 25-50%甘油的核酸分子溶液 (约 50 μ g) 放入冰水浴中，并使所述溶液经受 8-10 psi 压力下的气流 (例如氮气) 2.5 分钟。应了解，可使用任何气体，尤其惰性气体。气压是片段尺寸的主要决定因素。改变压力可产生各种片段尺寸。于喷雾处理时使用冰水浴可产生均匀分布的片段。类似地，可使用高压喷射器产生片段。Cavalieri, L. F.和 Rosenberg, B. H., *J. Am. Chem. Soc.* **81**:5136-5139(1959)。

断裂核酸分子的另一种方法采用反复冷冻和解冻经缓冲的核酸分子溶液。可视需要冷冻和解冻核酸分子样本以产生具有所需尺寸或尺寸范围的片段。另外，可用离子或微粒轰击核酸分子以产生具有不同尺寸的片段。例如，可在真空下将核酸分子暴露于离子引出束流 (ion extraction beamline)。以 7 kV \times q 从电子束离子阱引出离子，并将其导向到目标核酸分子上。可辐照核酸分子任何长度的时间，通常数小时，直至例如达成每平方微米 100 个离子的总注量。

还可通过辐照核酸分子来达成核酸分子的断裂。通常，例如 γ 射线或 x 射线辐射的辐射足以断裂核酸分子。片段尺寸可通过调整暴露于辐射的强度和持续时间来进行调整。也可使用紫外线辐射。也可调整暴露强度和持续时间以将辐射对核酸分子的不良作用减到最少。

煮沸核酸分子也可产生片段。通常，在持续搅动下煮沸核酸分子的溶液数小时。可达成约 500 bp 的片段。片段尺寸可随煮沸持续时间变化。

3.核酸分子的化学断裂

可使用化学断裂以碱基特异性或不以碱基特异性断裂核酸分子。可通过化学反应断裂核酸分子，所述化学反应包括例如水解反应，包括碱水解和酸水解。因为 RNA（或未配对碱基）在碱性条件下不稳定，因此碱性条件可用于断裂含有切口的核酸分子或 RNA。参看 Nordhoff 等人 "Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry," *Nucl. Acids Res.* 21 (15):3347-3357 (1993)。可在酸、通常强酸（例如 6 M HCl）存在下水解 DNA。温度可升高到室温以上以促进水解。视条件和反应时间长度而定，核酸分子可被断裂成不同尺寸，包括单碱基片段。在严苛条件下，水解可破坏磷酸酯键和脱氧核糖与嘌呤和嘧啶碱基之间的 N-糖苷键。

已知产生核酸分子片段的示范性酸/碱水解方案（例如，参看 Sargent 等人 *Meth. Enz* 152:432 (1988)）。简单而言，将 1 g DNA 溶解于 50 mL 0.1 N NaOH 中。加入 1.5 mL 浓盐酸，并迅速搅拌溶液。DNA 立即沉淀，不应搅拌超过数秒以防止形成大的聚集体。在室温下培育样本 20 分钟以使 DNA 部分脱嘌呤。随后，加入 2 mL 10 N NaOH（OH⁻浓度到 0.1 N），并搅拌样本直至 DNA 完全再溶解。随后在 65°C 下培育样本 30 分钟以水解 DNA。典型尺寸在约 250-1000 个核苷酸的范围内，但可视水解条件而变低或变高。

化学断裂也可具有特异性。例如，可通过烷基化切割所选核酸分子，尤其硫代磷酸酯修饰的核酸分子（例如，参看 K.A. Browne, "Metal ion-catalyzed nucleic Acid alkylation and fragmentation," *J. Am. Chem. Soc.* 124(27):7950-7962 (2002)）。硫代磷酸酯修饰处的烷基化使核酸分子的修饰位点处易于切割。I.G.Gut 和 S. Beck 描述烷基化 DNA 以用于质谱法检测的方法。I.G. Gut 和 S. Beck, "A procedure for selective DNA alkylation and detection by mass spectrometry," *Nucl. Acids Res.* 23(8):1367-1373 (1995)。

此项技术中已知碱基特异性和非碱基特异性化学切割寡核苷酸的其它各种化学制品和方法，且预期其用于本文所提供的断裂方法中。例如，碱基特异性切割可使用化学制品实现，所述化学制品例如为哌啶甲酸（piperidine formate）、哌啶、硫酸二甲酯、胍和氯化钠、胍。例如，可使用硫酸二甲酯和哌啶在 G 核苷酸处对 DNA 进行碱基特异性切割；可使用硫酸二甲酯、哌啶和酸在 A 和 G 核苷酸处对 DNA 进行碱基特异性切割；可使用胍和哌啶在 C 和 T 核苷酸处对 DNA 进行碱基特异性切割；可使用胍、哌啶和氯化钠在 C 核苷酸处对 DNA 进行碱基特异性切割；以及可使用强碱在 A 核苷酸处并以对 C 核苷酸的较低特异性对 DNA 进行碱基特异性切割。在另一实例中，可将核糖核苷酸

和脱氧核糖核苷酸并入目标核酸分子中，并且根据目标核酸分子的组成，可使目标核酸分子与特异性切割 RNA 或 DNA 的条件接触，引起碱基特异性切割（部分或完全切割）。

4.断裂方法的组合

可使用本文所述的断裂方法的任何组合形成片段，例如使用不同酶断裂方法的组合、不同化学断裂方法的组合、不同物理断裂方法的组合或酶和化学断裂方法的组合、酶和物理断裂方法的组合、化学和物理断裂方法的组合或酶和化学和物理断裂方法的组合。几个具体实例包括（但不限于）不同碱基特异性切割方法的组合和剪切与序列特异性酶的组合。产生特定片段的方法可与产生随机片段的方法组合。另外，可组合产生随机片段的不同方法，并且可组合产生特定片段的不同方法。例如，一种或一种以上在特定位点处切割核酸分子的酶可与一种或一种以上在不同位点处特异性切割核酸分子的酶组合使用。在另一实例中，切割特定种类的核酸分子的酶可组合使用，例如，RNA 酶与 DNA 酶可组合使用，或单链特异性核酸酶可与双链特异性核酸酶组合使用，或核酸外切酶可与核酸内切酶组合使用。在又一实例中，随机切割核酸分子的酶可与特异性切割核酸分子的酶组合使用。使用组合断裂是指对核酸分子先后或同时执行一种或一种以上的方法。

如本文所预期，组合使用也涵盖对第一部分核酸分子样本使用第一种断裂方法，对第二部分核酸分子样本使用第二种断裂方法。在随后的检测和质量测量方法中可单独分析两个样本；或可将两个样本汇集起来并在随后的检测和质量测量方法中同时进行分析。断裂方法组合可包括两种或两种以上断裂方法、三种或三种以上断裂方法或四种或四种以上断裂方法。

5.杂交后断裂

也可在目标核酸与捕获寡核苷酸探针杂交后断裂目标核酸。在一实施例中，目标核酸在与捕获寡核苷酸探针杂交之前经历一个或一个以上断裂步骤，且随后在与捕获寡核苷酸探针杂交后经历一个或一个以上另外的断裂步骤。在另一实施例中，目标核酸分子在与捕获寡核苷酸探针杂交之前不经历任何断裂步骤，但在与捕获寡核苷酸探针杂交之后经历一个或一个以上断裂步骤。目标核酸与捕获寡核苷酸探针杂交之后发生反应的实例包括酶断裂和化学断裂。在一实施例中，所述杂交后断裂步骤选择性断裂单链核酸而不断裂双链核酸。在另一实施例中，杂交后断裂包括碱基特异性切割。

E.捕获寡核苷酸

本文所提供的方法和组合物中也包括一种或一种以上可与目标核酸片段杂交的捕获寡核苷酸。在通常一些目标核酸片段与捕获寡核苷酸杂交而一些目标核酸片段不与捕

获寡核苷酸杂交的条件下，可使本文所提供的捕获寡核苷酸与目标核酸片段接触。可将与捕获寡核苷酸杂交的目标核酸片段与不与捕获寡核苷酸杂交的目标核酸片段分离开来。与捕获寡核苷酸杂交的目标核酸片段和与不与捕获寡核苷酸杂交的目标核酸片段可在与捕获寡核苷酸接触后和/或在分离开已杂交与未杂交片段之后经历分离处理步骤。目标核酸片段与捕获寡核苷酸接触之后，可测量目标核酸片段的质量。由于目标核酸片段与捕获寡核苷酸接触可导致核酸片段的分离，因此与捕获寡核苷酸接触的目标核酸片段的质谱相对于未与捕获寡核苷酸接触的片段的质谱可具有较低质量（例如，在不同质量处具有较少峰）。尽管可使用捕获寡核苷酸与仅单一序列杂交，但本文预期，也可通过使用（例如）简并碱基或低严格度或中等严格度杂交条件使用捕获寡核苷酸有意地与一种以上捕获寡核苷酸序列杂交。与捕获寡核苷酸杂交的不同目标核酸片段的数量和种类可决定通过质谱法测量的不同片段的数量和种类。

因此，本文所提供的一种示范性方法是一种测量目标核酸片段的质量的方法，其包含：

（a）控制与捕获寡核苷酸探针杂交的目标核酸片段的复杂度，其中所述目标核酸片段中各片段都含有与所述捕获寡核苷酸探针杂交的至少一个第一区；和

（b）使用质谱法测量与所述捕获寡核苷酸探针杂交的目标核酸片段的质量；

其中所述控制复杂度的步骤包括调节所述与捕获寡核苷酸探针杂交的目标核酸片段第一区中不同序列的数量，由此使个别第一区中含有不同核苷酸序列的两个或两个以上目标核酸片段与所述捕获寡核苷酸探针杂交。

1.控制目标核酸片段的复杂度

如本文别处所述，本文所提供的方法包括测量目标核酸片段的质量的步骤。视在特定测定中测量质量（例如，在单一质谱中测量质量）的目标核酸片段的数量和/或变异性而定，不同片段的质量可能容易或不容易辨别，以特定质量表示的不同核苷酸序列的数量可大可小，且不存在的的质量（例如，可能的但未呈现质量峰）可能容易或不容易鉴定。当片段复杂度极低时，质谱仅具有少数存在/不存在质量，这会限制由序列测定方法所提供的稳健程度（例如，当通过质量测量确定仅单一片段存在或不存在时，提供极少传统杂交测序方法尚未获得的信息）。当片段复杂度极高时，质谱可具有大量存在/不存在的质量，且每个质量可表示许多不同的核苷酸序列，这会限制可使用特定观察（例如，质量存在或不存在）以高概率指定核苷酸序列的范围（例如，当可存在/不存在过多片段时，提供不同于未使用捕获寡核苷酸杂交情况下质谱方法的极少的复杂度降低）。因此，控制目标核酸片段的复杂度可用于“调整”质谱，从而使质谱可提供大量可解析的观察结

果（例如，解析质量的存在或不存在），并且视情况，所述观察结果表示允许序列测定的足够少量不同序列。

在一实施例中，在测量目标核酸片段的质量之前控制目标核酸片段的复杂度。在另一实施例中，控制复杂度包括控制目标核酸片段的一个区，其中至少部分目标核酸片段另外含有未控制复杂度或以不同方式控制复杂度的第二区。

a.控制复杂度的方法

如本文所预期，目标核酸的断裂以及目标核酸与连接至固体载体的捕获寡核苷酸的杂交可用于控制或降低质量待分析的目标核酸混合物的复杂度。

在控制复杂度的实例中，通过断裂控制目标核酸片段的长度，也可控制目标核酸片段中的一部分序列，包括目标核酸片段 3'端、5'端或 3'端与 5'端一个或一个以上核苷酸位的身份。在另一实例中，目标核酸与捕获寡核苷酸杂交可控制与捕获寡核苷酸探针杂交的区域中目标核酸序列的复杂度。在一实施例中，当目标核酸的第一区与捕获寡核苷酸探针杂交时，可独立于目标核酸第二、非杂交区的复杂度而单独地控制目标核酸第一区的复杂度。

例如，当捕获探针为 5 个核苷酸长且目标核酸序列为 8 个核苷酸长时，可使用例如杂交条件和仅允许两种不同的目标核酸序列与捕获寡核苷酸探针序列杂交的捕获寡核苷酸探针序列控制复杂度，致使与特定捕获寡核苷酸杂交的不同目标核酸片段的可能数量限于不超过 512 个。可使用序列特异性断裂条件（例如使用如上文所讨论的序列特异性核酸内切酶或碱基特异性切割）进一步限制复杂度。

一般可通过控制目标核酸片段的长度、控制目标核酸片段的统计学尺寸范围内不同长度的数量、控制所分析的目标核酸的全长、使用序列特异性或非特异性断裂方法以及控制捕获寡核苷酸探针与目标核酸片段 5'端或 3'端处的核苷酸位杂交的能力来控制与捕获寡核苷酸探针杂交的目标核酸片段的杂交区与非杂交区的复杂度。此外，可通过改变使目标核酸暴露于捕获寡核苷酸的条件（例如，低严格度杂交条件、中等严格度杂交条件或高严格度杂交条件）和通过改变核苷酸的数量和/或捕获寡核苷酸探针的核苷酸的简并性（例如，通过使用通用或半通用核苷酸）来进一步控制杂交区的复杂度。例如，可通过减小目标核酸片段的长度、减少目标核酸片段的统计学尺寸范围内不同长度的数量、减小所分析的目标核酸的全长、使用序列特异性或碱基特异性断裂方法、使用倾向于与目标核酸片段 5'端或 3'端处的核苷酸位杂交的捕获寡核苷酸探针、使用高严格度杂交条件以及在捕获寡核苷酸中包括较多序列特异性核苷酸来降低与捕获寡核苷酸探针杂交的目标核酸片段的复杂度。在另一实例中，可通过增加目标核酸片段的长度、增加

目标核酸片段的统计学尺寸范围内不同长度的数量、增加所分析的目标核酸的全长、使用非特异性断裂方法、使用不倾向于与目标核酸的特定区杂交的捕获寡核苷酸探针、使用低严格度杂交条件以及在捕获寡核苷酸中包括较少和/或较低序列特异性核苷酸（例如，通用或半通用碱基）来增加与捕获寡核苷酸探针杂交的目标核酸分子片段的杂交区与非杂交区的复杂度。

在一实施例中，在测量目标核酸片段的质量的步骤之前控制与捕获寡核苷酸探针杂交的目标核酸片段的复杂度。例如，控制目标核酸片段的复杂度可在目标核酸片段与捕获寡核苷酸探针杂交之前（例如，在断裂步骤中）进行，和/或控制目标核酸片段的复杂度可包括使目标核酸片段与捕获寡核苷酸探针杂交，和/或控制目标核酸片段的复杂度可在目标核酸片段与捕获寡核苷酸探针杂交之后但在测量目标核酸片段的质量之前（例如，在随后的断裂步骤中，例如“修剪”）进行。

目标核酸的断裂产物可以多种方式捕获于固相上。例如，与一种或一种以上断裂产物特异性或半特异性杂交的捕获寡核苷酸可与固体载体连接以用于所述产物的特异性或“半特异性”捕获。

所属领域技术人员可根据本文所提供的教导和此项技术中的知识来评估与特定捕获寡核苷酸结合的目标核酸片段的预期复杂度。例如，当含有特定序列的捕获寡核苷酸含有包含通用核苷酸（例如，肌苷）的单一简并位时，多达四个与捕获寡核苷酸具有相同长度和相同序列组成（除与通用碱基互补位置处的核苷酸外）的不同目标核酸片段可以粗略相同的结合亲和力与所述特定捕获寡核苷酸结合。如果还存在较大的目标核酸片段并且比捕获寡核苷酸长 1 至 5 个核苷酸，那么多达 30,948 种不同的目标核酸片段可与一个捕获寡核苷酸序列结合（参看图 2）。类似地，当捕获寡核苷酸中具有与通用寡核苷酸对应的 2 个简并位时，多达 16 个具有相同长度和序列组成（除与通用碱基互补位置处的核苷酸外）的不同目标核酸片段可以粗略相同的结合亲和力与所述特定捕获寡核苷酸结合。

在一实施例中，可完全除去目标核酸片段的非杂交区。这可例如通过产生与捕获寡核苷酸探针具有相同尺寸的目标核酸片段或通过产生比捕获寡核苷酸探针长的目标核酸片段使目标核酸与捕获寡核苷酸探针杂交且随后使用单链特异性核酸酶切割未杂交的核苷酸实现。

在一些实施例中，可获得有关与特定捕获探针杂交的不同序列的最小数量的信息。例如，当使用低严格度杂交条件或简并捕获寡核苷酸探针时，一种以上目标核酸序列可与同一个捕获寡核苷酸探针序列杂交。在此情况下，如果所有目标核酸片段都与捕获寡

核苷酸探针具有相同尺寸并且所有目标核酸片段都具有不同组成（即，A、C、T 和 G 的数量不同），那么质量峰的数量将对应于与捕获寡核苷酸探针杂交的不同目标核酸序列的数量。由于具有不同序列的目标核酸片段可能具有相同组成（即，A、C、T 和 G 的数量相同），因此一些不同序列可具有相同的质量测量值，因而质量峰的数量提供所存在的不同序列的最小数量。

也可基于碱基组成通过例如序列特异性切割（例如单碱基特异性切割）修饰非杂交端（例如，5'端或3'端）。例如，如果所使用的目标核酸片段为 RNA，并且所述 RNA 首先与捕获探针杂交且随后暴露于 RNA 酶 T1（其特异性切割单链 RNA 中 G 的 3'端），那么不同目标探针的非杂交端的长度将根据与目标核酸的杂交端最接近的 G 的位置而变化。因此，例如碱基特异性切割非杂交端的方法可允许控制非杂交端，而无需在碱基特异性切割之前预先界定非杂交端的长度。

可对通常出现于核酸中的四个碱基中的任一个进行非杂交端的碱基特异性切割。在一实施例中，将目标核酸的样本分成四个单独的样本，且每一个单独样本都与一个或四个相同芯片上的捕获探针杂交。与捕获探针杂交后，使四个芯片（或一个芯片上的四个不同位置）的目标核酸各自经历四种不同碱基特异性切割反应中的一个。最后，测量杂交目标核酸的质量。这种四倍碱基特异性切割（four-fold base-specific cleavage）也可连续进行，其中使四个分开的样本连续与同一个芯片杂交，以四种碱基特异性切割反应中的一个进行处理，并测量质量。通过测量来自四种不同碱基特异性切割反应的与同一个捕获探针杂交的目标核酸的质量，在一次碱基特异性切割后可能具有相同组成（且因此具有相同质量）的非杂交端的不同序列在一次或一次以上不同的碱基特异性切割后具有不同组成（且因此具有不同质量）。

如所属领域技术人员所认识到的，可执行断裂、杂交和视情况再断裂的各种组合以便达到所需复杂度。

b. 片段的区域

目标核酸片段可含有至少一个、至少两个或至少三个区。例如，仅含有一个区的目标核酸片段可为目标核酸的每一个核苷酸都与捕获寡核苷酸探针杂交的目标核酸；含有至少两个区的目标核酸可为目标核酸中仅一分组核苷酸与捕获寡核苷酸探针杂交的目标核酸（例如，含有两个区的目标核酸可为目标核酸的 3'端与捕获寡核苷酸探针杂交而非 5'端不杂交的目标核酸，反之亦然）；含有至少三个区的目标核酸可为目标核酸的中心区而非 5'端和 3'端与捕获寡核苷酸探针杂交的目标核酸，或可为 5'端和 3'端而非中心区与捕获寡核苷酸探针杂交的目标核酸；具有三个以上区的目标核酸可为具有两个或两个以

上与捕获寡核苷酸探针杂交的实体分离区 (physically separated region) 的目标核酸。

类似地, 捕获寡核苷酸探针可具有一个或一个以上区。例如, 具有两个区的捕获寡核苷酸可具有与目标核酸片段杂交的第一区和不与至少一种目标核酸杂交的第二区。

c. 部分单链的捕获寡核苷酸

在另一实施例中, 固体载体上的捕获寡核苷酸可以是具有单链悬突的部分双链。捕获寡核苷酸中单链悬突的长度通常为 5-6 个核苷酸, 也可在 4 至 10 个核苷酸的范围内或更多个。当捕获寡核苷酸为部分双链并且例如具有 5 核苷酸单链悬突时, 具有 1024 个不连续位点的固体载体可含有与所有可能的目标核酸的 5 个核苷酸互补的捕获探针。另外, 使用具有单链悬突的双链捕获寡核苷酸通过允许捕获寡核苷酸探针与目标核酸一端之间的碱基堆积相互作用 (base-stacking interaction) 而增加目标核酸对捕获寡核苷酸的亲和力。通过目标核酸的一端与捕获寡核苷酸探针的碱基堆积, 可独立于另一端的复杂度而单独地控制目标核酸一端的复杂度。

例如, 当捕获探针具有从一条链的 3'端延伸的 5 核苷酸单链悬突时, 目标核酸 3'端处的 5 个核苷酸可与捕获探针的单链悬突杂交。如果捕获探针不具有简并位, 那么目标核苷酸中仅一个 3'端 5 碱基序列以最高互补性与所述探针杂交。如果捕获探针具有一个通用或半通用碱基, 那么目标核酸中分别仅 4 个或 2 个 3'端 5 碱基序列以最高互补性与所述探针杂交。

另外, 在所述实例中, 当捕获探针具有从一条链的 3'端延伸的 5 核苷酸单链悬突时, 目标核苷酸的长度可大于 5 个碱基; 在本实例中, 为简单起见, 目标核苷酸的长度可为 5 个碱基到 7 个碱基。因此, 具有 3 种不同长度 (5 个碱基、6 个碱基和 7 个碱基) 的核苷酸可以最高互补性与非简并捕获寡核苷酸探针杂交。假定捕获寡核苷酸探针为非简并寡核苷酸, 并且由于目标核酸的每一个位置都可具有四个不同碱基中的任一个, 因此多达 21 种 ($4^2 + 4^1 + 4^0$) 不同目标核酸可与每一个非简并捕获寡核苷酸探针杂交。如果捕获探针单链区中的 5 个碱基中的一个为通用碱基, 那么多达 21×4 种或 84 种目标核酸可与每一个捕获探针杂交。如果不使用通用碱基, 而是操纵杂交条件以允许目标核苷酸与捕获探针相互作用的 5 个位置中的任一个位置处出现 1 处错配, 那么多达 $21 \times 4 \times 5$ 种或 420 种目标核酸可与每一个捕获探针杂交。如所属领域技术人员所了解, 可基于其它各种探针和杂交严格度进行类似计算, 以模拟目标核酸片段一个区的复杂度或整个片段的复杂度。

由上述三个实例可见独立于 5'端复杂度进行的对于 3'端复杂度的单独控制。在所述实例中, 5'端序列仅受目标核酸的长度控制, 因此 5'端可具有多达 21 种不同序列, 或者

如果增加长度和/或长度的可变性,那么 5'端可具有更多种不同序列。在本实例中,可通过使用简并位和/或杂交条件来控制 3'端序列,从而使 3'端的复杂度可在 1 与 20 种不同序列之间变化,或者,如果进一步放松杂交严格度或捕获探针中包括其它简并位,那么 3'端的复杂度将更高。另外,3'端的复杂度也可受捕获探针中所存在的单链悬突碱基的数量控制。

2.捕获寡核苷酸的组成

根据捕获寡核苷酸的所需特性,捕获寡核苷酸可具有各种组成。例如,捕获寡核苷酸可为单链或含有单链与双链区,捕获寡核苷酸可含有通用和/或半通用碱基,而且捕获寡核苷酸可为各种长度。

a.核苷酸的类型

捕获寡核苷酸可含有天然存在与非天然存在的各种核苷酸的任一种。通常,捕获寡核苷酸含有一种或一种以上相对于目标核酸的第二组核苷酸更倾向于与目标核酸的第一组核苷酸杂交的核苷酸。例如,捕获寡核苷酸可含有 A、G、C 或 T/U 中的一者或一者以上。

在一些实施例中,捕获寡核苷酸可为部分简并寡核苷酸,且含有一个或一个以上简并碱基。例如,一个或一个以上简并碱基可“位于捕获寡核苷酸的 3'端”。而在其它实施例中,一个或一个以上简并碱基可“位于捕获寡核苷酸的 5'端”。例如,于捕获寡核苷酸一端处安置一个或一个以上通用碱基可有益于在不改变捕获寡核苷酸的碱基特异性的情况下增强捕获寡核苷酸与目标核酸之间的杂交;然而,可使用所述安置改变捕获寡核苷酸优先结合的目标核酸的长度。

在其它实施例中,一个或一个以上简并碱基(例如通用和半通用碱基)位于捕获寡核苷酸探针中的特定、非简并碱基之间。以此方式,相对于捕获寡核苷酸探针识别序列中第二分组的核苷酸位置,捕获寡核苷酸探针识别序列中第一所选分组的核苷酸位置对特定核苷酸具有增加的特异性。如所属领域技术人员所认识到的,非简并碱基之间简并碱基的分布可采取各种形式。因此,一个或一个以上相邻的简并碱基可分布于识别序列的一个或一个以上单独位置中,其中所述简并碱基位于非简并碱基之间。

i.通用碱基

捕获寡核苷酸的简并性可使用通用碱基达成,所述通用碱基可以类似的亲和力结合 DNA 或 RNA 的四个常见碱基中的任一个。用于本文中的示范性通用碱基包括肌苷、黄苷、3-硝基吡咯(Bergstrom 等人, *Abstr. Pap. Am. Chem. Soc.* 206(2):308 (1993); Nichols 等人, *Nature* 369:492-493; Bergstrom 等人, *J. Am. Chem. Soc.* 117:1201-1209 (1995))、4-

硝基吡啶 (Loakes 等人, *Nucleic Acids Res.*, 22:4039-4043 (1994))、5-硝基吡啶 (Loakes 等人 (1994))、6-硝基吡啶 (Loakes 等人 (1994)); 硝基咪唑 (Bergstrom 等人, *Nucleic Acids Res.* 25:935-1942 (1997))、4-硝基吡啶 (Bergstrom 等人 (1997))、5-氨基吡啶 (Smith 等人, *Nucl. Nucl.* 17:555-564 (1998))、4-硝基苯并咪唑 (Seela 等人, *Helv. Chim. Acta* 79:488-498 (1996))、4-氨基苯并咪唑 (Seela 等人, *Helv. Chim. Acta* 78:833-846 (1995))、苯基 C-核糖核苷 (Millican 等人, *Nucleic Acids Res.* 12:7435-7453 (1984); Matulic-Adamic 等人, *J. Org. Chem.* 61:3909-3911 (1996))、苯并咪唑 (Loakes 等人, *Nucl. Nucl.* 18:2685-2695 (1999); Papageorgiou 等人, *Helv. Chim. Acta* 70:138-141 (1987))、5-氟吡啶 (Loakes 等人 (1999))、吡啶 (Girgis 等人, *J. Heterocycle Chem.* 25:361-366 (1988)); 无环糖类似物 (Van Aerschot 等人, *Nucl. Nucl.* 14:1053-1056 (1995); Van Aerschot 等人, *Nucleic Acids Res.* 23:4363-4370 (1995); Loakes 等人, *Nucl. Nucl.* 15:1891-1904 (1996)), 包括次黄嘌呤衍生物, 咪唑 4,5-二甲酰胺、3-硝基咪唑、5-硝基吡啶; 芳族类似物 (Guckian 等人, *J. Am. Chem. Soc.* 118:8182-8183 (1996); Guckian 等人, *J. Am. Chem. Soc.* 122:2213-2222 (2000)), 包括苯、萘、菲、蒽、吡咯、二氟甲苯; 异喹诺酮核苷衍生物 (Berger 等人, *Nucleic Acids Res.* 28:2911-2914 (2000); Berger 等人, *Angew. Chem. Int. Ed. Engl.* 39:2940-2942 (2000)), 包括 MICS、ICS; 氢键合类似物, 包括 N8-吡咯并吡啶 (Seela 等人, *Nucleic Acids Res.* 28:3224-3232 (2000)); 和 LNA, 例如芳基-P-C-LNA (Babu 等人, *Nucleosides, Nucleotides & Nucleic Acids* 22:1317-1319 (2003); WO 03/020739)。

ii. 半通用碱基

半通用碱基优先与 2 个或 3 个常见 (即, DNA 中的 A、C、G 和 T; 和 RNA 中的 A、C、G 和 U) 核苷酸结合, 但不以相同或相似的特异性与所有 4 个常见核苷酸结合。例如, 半通用碱基以比与至少一个其它常见核苷酸结合的亲和力高的亲和力与 2 个或 3 个常见核苷酸结合。用于本文中的示范性半通用碱基优先与嘌呤 A 和 G 或嘧啶 C 和 T 杂交。例如, 嘧啶类似物 6H,8H-3,4-二氢嘧啶并[4,5-c][1,2]恶嗪-7-酮优先与 A 或 G 杂交; 嘌呤类似物 N6-甲氧基-2,6-二氨基嘌呤优先与 C、T 或 U 杂交 (例如, 参看 Bergstrom 等人, *Nucleic Acids Res.* 25:1935-1942 (1997))。

b. 其它特征

捕获寡核苷酸的序列、长度和组成根据所属领域技术人员已知的多种因素变化, 所述因素包括 (但不限于) 目标核酸分子的长度、断裂方法、杂交条件、待使用的不同捕获寡核苷酸的数量和所希望与特定捕获寡核苷酸杂交的不同核苷酸组成和/或序列的所需数量。

在本文的特定实施例中，一分组捕获寡核苷酸可为部分简并寡核苷酸。例如，本文涵盖至少 10%、至少 20%、至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 90%、至少 95%捕获寡核苷酸为部分简并寡核苷酸的实施例。此外，本文涵盖不超过 10%、不超过 20%、不超过 30%、不超过 40%、不超过 50%、不超过 60%、不超过 70%、不超过 80%、不超过 90%、不超过 95%捕获寡核苷酸为部分简并寡核苷酸的实施例。在本文的其它实施例中，所有捕获寡核苷酸都为部分简并寡核苷酸。在其它实施例中，捕获寡核苷酸都不为部分简并寡核苷酸。

部分简并捕获寡核苷酸中可含有一个或一个以上非简并核苷酸（例如，对于 DNA 而言为 A、C、G、T；对于 RNA 而言为 A、C、G、U）和一个或一个以上简并核苷酸（例如，并入捕获寡核苷酸中的通用碱基或半通用碱基）的组合。在另一实施例中，部分简并寡核苷酸仅含有简并核苷酸，其中所述部分简并寡核苷酸仍保持以相对于结合第二组核苷酸序列而言更高的特异性结合第一组核苷酸序列的能力。例如，部分简并寡核苷酸可仅含有半通用碱基或半通用碱基与通用碱基的组合，并且半通用碱基的优先结合赋予部分简并寡核苷酸结合特异性。

使用部分简并捕获寡核苷酸允许一种以上特定目标核酸序列与个别部分简并捕获寡核苷酸结合，由此允许比所有理论捕获寡核苷酸序列组合少的捕获寡核苷酸存在于阵列上，以便捕获所有理论目标核酸组合。选择特定捕获寡核苷酸上所使用的简并位的数量以使单个捕获寡核苷酸能够优先与切割步骤中所产生的各种片段中的两个或两个以上不同目标核酸片段杂交。

如本文别处所提供，使用比所有理论捕获寡核苷酸组合少的捕获寡核苷酸也涵盖降低或放松杂交条件的严格度以允许错配结合，由此允许一种以上特定目标核酸序列与个别部分简并或非简并捕获寡核苷酸结合，从而允许比所有理论捕获寡核苷酸序列组合少的捕获寡核苷酸序列存在于阵列上，以便捕获所有理论目标核酸的组合。

捕获寡核苷酸可对每一种目标核酸断裂产物具有特异性，或者捕获寡核苷酸可与目标核酸中两种或两种以上不同片段的共用区互补。例如，在特定杂交反应试验中，经固相固定的捕获寡核苷酸可与具有不同尺寸、包括共用亚片段序列的断裂产物杂交。此外，可使用单个捕获寡核苷酸通过使用较低严格度杂交条件和/或通过使用捕获寡核苷酸内的一个或一个以上简并核苷酸捕获多个目标核酸片段，所述目标核酸片段中与捕获寡核苷酸互补的区域的序列彼此有一个或一个以上核苷酸不同。换句话说，可根据经验选择捕获寡核苷酸和严格度条件以允许单个捕获寡核苷酸序列与一种以上目标核酸片段序列结合。同样，可根据经验选择捕获寡核苷酸和严格度条件以控制与捕获寡核苷酸杂交

的具有不同序列的不同核苷酸片段或具有不同组成的核苷酸片段的数量。

因此，本文所使用的捕获寡核苷酸含有具有足够长度和足够互补性的核苷酸序列以在接触或组合步骤条件下与本文所制备的目标核酸片段半特异性杂交。所述杂交之前、期间或之后（杂交可在溶液或固相中发生），固定捕获寡核苷酸且将其排列于固体载体上对应的不连续、不重叠元件上，使得每一个元件都含有不同的捕获寡核苷酸。此项技术中已知用于将寡核苷酸排列于固体载体的不连续元件上的多种材料和方法，所述材料例如为玻璃、硅、塑料、尼龙膜、多孔材料等；所述方法包括：接触沉积，例如美国专利第 5,807,522 号、第 5,770,151 号等；基于光刻法的方法，例如参看美国专利第 5,861,242 号、第 5,858,659 号、第 5,856,174 号、第 5,856,101 号、第 5,837,832 号等；基于流动通道（flow path-based）的方法，例如美国专利第 5,384,261 号；基于蘸笔纳米光刻法（dip-pen nanolithography-based）的方法，例如 Piner 等人, *Science Jan.* 29:661-663 (1999)。在特定实施例中，捕获寡核苷酸排列于对应的不连续位置（位点）处，一般为每个固相阵列（例如，芯片）不超过 20,000 个、不超过 15,000 个、不超过 10,000 个、不超过 7,000 个、不超过 5,000 个、不超过 4,000 个、不超过 3,000 个、不超过 2500 个、不超过 2100 个、不超过 2000 个、不超过 1500 个、不超过 1400 个、不超过 1300 个、不超过 1200 个、不超过 1100 个、不超过 1000 个、不超过 900 个、不超过 800 个、不超过 700 个、不超过 600 个、不超过 500 个、不超过 400 个、不超过 300 个、不超过 200 个、不超过 100 个不连续元件（位点）。

如本文所述，本文所提供的方法中所使用的固相阵列可含有具有若干个简并核苷酸的捕获寡核苷酸。这会降低为捕获原始目标核酸序列中所密封的信息所需的寡核苷酸的总数。因此，最初切割目标核酸期间所产生的类似序列的多个片段可与同一个捕获寡核苷酸的个别位置杂交。如果多种物质具有完全不同的核苷酸组成，那么质谱分析允许通过分子质量对其进行鉴定。

在本文所涵盖的一个特定实施例中，使用通用或半通用碱基允许芯片与少至 4096 个或更少的捕获位置杂交以用于测序。特殊应用可能需要甚至更少的寡核苷酸数量。例如，在本文所涵盖的一个实施例中，4096 个捕获寡核苷酸将允许产生长度为 12 的所有捕获寡核苷酸以供简并嘌呤/嘧啶杂交碱基（即，含有 12 个半通用碱基的 12 碱基捕获寡核苷酸），或捕获具有 6 个非简并碱基（A、C、G、T）和 6 个通用碱基的寡聚物或其组合（例如，2 个非简并碱基、8 个半通用碱基和 2 个通用碱基）。本实施例不需要阵列中的每一个捕获寡核苷酸都具有相同的非简并、半通用和通用碱基含量以产生所有捕获寡核苷酸。例如，一些捕获寡核苷酸可仅含有半通用碱基，而其它捕获寡核苷酸可含有非

简并碱基、通用碱基和半通用碱基；而仍有其它一些仅含有非简并碱基和通用碱基。各类碱基的相对量可由所属领域技术人员根据所需的捕获寡核苷酸特异性程度进行确定。

在另一实施例中，杂交结构可具有少至（例如）1024 个捕获位置。可使用所述芯片杂交多个样本，例如，各自在特异性切割不同碱基的条件下经单独处理的四个样本（例如，样本 1 是在 A 特异性切割条件下经处理；样本 2 是在 C 特异性切割条件下经处理；样本 3 是在 G 特异性切割条件下经处理；而样本 4 是在 T 特异性切割条件下经处理）。在一实施例中，经四种不同切割条件处理的同一个核苷酸的四个样本同时与杂交结构杂交，并测量目标核酸的质量。在另一实施例中，经四种不同切割条件处理的同一个核苷酸的四个样本是在四个单独杂交步骤中与杂交结构杂交，其中在四个单独杂交步骤中的每一个杂交步骤之后测量目标核酸的质量。在另一实施例中，所述碱基特异性切割可对单链核酸具有选择性，使得目标核酸中未与捕获寡核苷酸探针结合的部分经历碱基特异性切割，产生比与目标核酸杂交的捕获寡核苷酸探针长的目标核酸（即，使捕获寡核苷酸探针悬突），其中悬突的长度由相对于目标核酸的杂交部分最接近的经特异性切割的碱基的位置决定。

c. 制备捕获寡核苷酸

可单独合成寡核苷酸且随后将其连接至固体载体，或其合成可在固体载体表面上原位进行。寡核苷酸可购自多个公司，包括 Integrated DNA Technology (IDT)、Fidelity Systems、Proligo、MWG、Operon、MetaBION 等。

寡核苷酸和寡核苷酸衍生物可通过此项技术中已知的标准方法合成，例如通过使用自动化 DNA 合成仪（例如购自 Biosearch (Novato, CA)、Applied Biosystems (Foster City, CA) 和其它公司的合成仪）并结合固体载体（例如可控孔径玻璃 (CPG) 或聚苯乙烯）和其它树脂和化学方法（例如亚磷酰胺方法、氢亚磷酸酯 (H-phosphonate) 或磷酸三酯方法）进行合成。寡核苷酸也可在溶液中或可溶载体上合成。例如，可通过 Stein 等人 (Nucl. Acids Res. 16:3209 (1988)) 的方法合成硫代磷酸酯寡核苷酸，并且可通过使用可控孔径玻璃聚合物载体 (Sarin 等人, Proc. Natl. Acad. Sci. U.S.A. 85:7448-7451 (1988)) 制备磷酸甲酯寡核苷酸 (methylphosphonate oligonucleotide)。如本文所公开和此项技术中所已知，还可使用用于扩增的酶法（例如，PCR 或转录）产生寡核苷酸。

表面结合的捕获寡核苷酸为与目标核酸片段上的互补区杂交的核酸。捕获寡核苷酸一般实质上不涉及产生目标核酸片段而发生的任何反应，例如于 2002 年 4 月 11 日申请的相关申请案第 60/372,711 号、2003 年 3 月 24 日申请的第 60/457,847 号和 2003 年 4 月 11 日申请的第 10/412,801 号中所公开的芯片室中发生的反应。优选的寡核苷酸具有

足以允许与目标核苷酸序列特异性或半特异性杂交的数量的核苷酸。

捕获寡核苷酸可为各种长度，并且可包括与目标核酸核苷酸序列结合的核苷酸和与目标核酸核苷酸序列结合的核苷酸。例如，捕获寡核苷酸可含有与将捕获寡核苷酸锚定至固体载体的核苷酸序列杂交的部分，或结合目标核酸片段的引物序列的部分（例如，不为目标核酸核苷酸序列部分的转录起始位点）。捕获寡核苷酸也含有可与目标核酸核苷酸序列结合的核苷酸。与目标核酸序列结合的捕获寡核苷酸部分可根据本文所提供和所属领域技术人员已知的因素而为各种长度。通常，这部分捕获寡核苷酸具有 5 至 30 个碱基长度。因此，预期本文中使用的寡核苷酸的特定长度包括 4 个、5 个、6 个、7 个、8 个、9 个、10 个、11 个、12 个、13 个、14 个、15 个、16 个、17 个、18 个、19 个、20 个、21 个、22 个、23 个、24 个、25 个、26 个、27 个、28 个、29 个、30 个核苷酸，或视需要更多个核苷酸。如本文所述，寡核苷酸可由天然核苷酸、修饰核苷酸或核苷酸类似物（例如，通用或半通用碱基）制得，以改变与互补序列杂交的特异性或改变所形成的杂合体的稳定性。

可通过将简并碱基或位点并入捕获寡核苷酸序列中控制捕获寡核苷酸的特异性。以肌苷取代序列内的碱基可例如导致对目标核酸产物中多态性位点的普遍杂交[例如，参看 Ohtsuka 等人 *J. Biol. Chem.* 260:2605 (1985); Takahashi 等人 *Proc. Natl. Acad. Sci. U.S.A.* 82:1931 (1985)]。可通过在捕获寡核苷酸序列或目标核酸序列内完全或部分使用例如 RNA（如果针对 DNA 目标）、锁核酸（locked nucleic acid, LNA）[Braasch 等人 *Chemistry & Biology* 8:1-7 (2001)]、肽核酸（PNA）[Armitage 等人 *Proc. Natl. Acad. Sci. U.S.A.* 94:12320-12325 (1997)]或其它经修饰核酸衍生物来显著增加双链核酸杂合体的稳定性。也可通过并入一个或几个无碱基位点、非杂交碱基衍生物或引起较低熔解温度的核酸修饰（例如硫代磷酸酯）来降低稳定性。例如所述方法的各种已知方法可用于调节几乎任何序列的熔解温度和用于所需熔解温度的序列长度。

寡核苷酸的合成

此项技术中众所周知在溶液或固体载体上合成寡核苷酸的方法[例如，参看 Beaucage 等人 *Tetrahedron Lett.* 22:1859-1862 (1981); Sasaki 等人 (1993) *Technical Information Bulletin T-1792*, Beckman Instrument; Reddy 等人, 美国专利 5,348,868; Seliger 等人 *DNA and Cell Biol.* 9:691-696 (1990)]。

寡核苷酸的原位合成

此项技术中众所周知使用光导合成于玻璃和硅表面上进行的寡核苷酸原位合成[例如，参看 McGall 等人 *J. Am. Chem. Soc.* 119:5081-5090 (1997); Wallraff 等人 *Chemtech*

27:22-32 (1997); McGall 等人 *Proc. Natl. Acad. Sci. U.S.A.* **93**:13555-13560 (1996); Lipshutz 等人 *Curr. Opin. Structural Biol.* **4**:376-380 (1994);和 Pease 等人 *Proc. Natl. Acad. Sci. U.S.A.* **91**:5022-5026 (1994)]。

寡核苷酸可连接至经过化学衍生化的固体载体或例如具有官能团的聚合物或塑料的固体载体。可通过多种方法使寡核苷酸与固体载体结合,包括光刻法、共价键连接或通过例如离子相互作用、范德华 (Van der Waal) 和氢键的非共价相互作用被动连接。可通过 5'端或 3'端修饰使寡核苷酸共价连接至固体载体表面。通常使用键联剂 (linker) 以便使寡核苷酸远离表面。例如,如果寡核苷酸将通过其 5'端连接,那么键联剂将在 5'端上直接进行 5'修饰。所使用的典型键联剂包括己基乙二醇 (hexylethyleneglycol) (一个或一个以上单元) 和寡聚脱氧胸苷 (oligodeoxythymidine) dTn (其中 $n = 5 - 20$)。

可使用各种方法将寡核苷酸连接至经过用反应性官能团化学衍生化的表面。例如,经氨基修饰的寡聚物可与环氧化物活化的表面反应形成共价键[例如,参看 Lamture 等人 *Nuc. Acids Res.* **22**:2121-2125 (1994)]。类似地,可于经羧酸改质的表面[Stoother 等人 *J. Am. Chem. Soc.* **122**:1205-1209 (2000)]、经异硫氰酸酯、胺、硫醇改质的表面[Penchovsky 等人 *Nuc. Acids Res.* **28**:e98 1-6 (2000); Lenigk 等人 *Langmuir* **17**:2497-2501 (2001)]、经异氰酸酯改质的表面[Lindroos 等人 *Nuc. Acids Res.* **29**:e69 1-7 (2001)]和经醛改质的表面[Zammatteo 等人 *Anal. Biochem.* **280**:143-150 (2000)]上达成经氨基修饰的寡核苷酸的共价连接。

通常,可如本文所述对硅表面进行化学衍生化,随后固定寡核苷酸[也参看 Benters 等人 *Nuc. Acids Res.* **30**:e10 1-7 (2002)]。例如,洗涤表面后,用氨基丙基三甲氧基硅烷处理表面以于表面上得到氨基硅氧烷层。用双功能交联剂对苯二异硫氰酸酯 (1,4-phenylenediisothiocyanate) 活化表面。交联剂的一个异硫氰酸酯基团与表面上的氨基官能团反应,形成稳定的硫脲键。第二个,即现与表面结合的异硫氰酸酯基团打开,与具有氨基的其它分子发生共价反应。在随后的步骤中,一种树枝状聚合的聚胺(例如,具有 64 个氨基末端的 4 代 Starburst (PAMAM) 树枝状聚合物) 与活化表面反应,于固体载体上形成具有密集量的共价连接氨基的均匀夹层。所述表面上的这些功能再次被对苯二异硫氰酸酯活化。未反应的胺由 4-硝基-苯异硫氰酸酯 (4-nitro-phenylene isothiocyanate) 阻断。经氨基修饰的寡核苷酸现通过相同类型的反应与经活化树枝状聚合物夹层共价交联。在最后的步骤中,未反应的异硫氰酸酯由小分子伯胺(如己胺)阻断。

捕获寡核苷酸连接于固体载体的多个不连续的已知位置或阵列位置处。每个位置都

可含有多个具有相同序列的寡核苷酸的拷贝。例如，捕获寡核苷酸探针阵列可在特定位置处具有多个寡核苷酸拷贝，其中所述特定位置处的所有寡核苷酸都具有相同的核苷酸序列，且其中相对于阵列上其它位置处捕获寡核苷酸的核苷酸序列，所述特定位置处捕获寡核苷酸的核苷酸序列是唯一的。因此，可排列阵列使得特定阵列位置处的所有寡核苷酸具有相同序列，且不同阵列位置处所有寡核苷酸序列为唯一的。

或者，每一位置可具有序列不同的寡核苷酸。可例如在多重反应中使用这种寡核苷酸排列。在相同位置处具有不同序列的寡核苷酸可混合在一起或分隔为相似序列组。例如，在相同位置处可具有两种、三种、四种或四种以上不同的寡核苷酸。所利用的不同寡核苷酸的数量仅受解析与一个位置内各不同序列结合的产物的能力限制。

固体载体上的不同位置通常含有具有不同序列的寡核苷酸。寡核苷酸在一个位置处通常占据 0.0025 mm^2 至 1.0 mm^2 的面积且寡核苷酸的量在 10 amol 与 10 pmol 之间。在某些实施例中，典型格式为固体载体，尺寸为 $20 \times 30 \text{ mm}$ ，具有 96 个、384 个或 1536 个位置， 8×12 、 16×24 或 32×48 模式和与反应板上的间距相同的间距（ 2.25 mm 、 1.125 mm 或 0.5625 mm 的中心距）。其它实施例可采用多达 4096 个位置。在一实施例中，位置为约在一类质谱分析中所使用的激光的直径，例如，一些位置不大于激光的直径。固体载体的尺寸、位置总数和排列位置的模式可与在固体载体上产生阵列、液体处理和/或分析所使用的设计方面和装置相适应。例如，间距和斑点尺寸可为能够由产生阵列的仪器的准确性和/或液滴尺寸所指示的尺寸。固体载体上行或列中安置寡核苷酸的位置的数量可为使 MALDI-TOF 质谱仪的激光不同时包围一个以上位置的数量。

捕获寡核苷酸组可以任何排列方式定位于固体载体表面上。例如，可将寡核苷酸安置于固体载体内所制的个别孔或小室中。固体载体上所存在的孔的数量可视固体载体的尺寸而变化，通常使用 96 或 384 格式，并且多达 4096 或更多的容易得到的格式。通常，所述孔或小室保持分离并维持其完整性。在一实例中，可将寡核苷酸放于固体载体上共用覆盖试剂通道（overlying reagent channel）的行或列中的不连续已知位置处。在另一实例中，寡核苷酸也可在所述不连续已知位置中以任何排列方式排列于完全平坦表面顶部上。所述位置也可再分成具有个别寡核苷酸或寡核苷酸混合物的较小区域。试剂小室或孔可具有用固体载体顶部上安置的相同或不同材料制成的障板（mask）。而且，可设计固体载体上的孔和小室使其例如根据尺寸定位或甚至分离和分选珠粒。在此设计中，所述珠粒为用于捕获反应产物核酸片段和衍生物的寡核苷酸的载体。

F. 固体载体和阵列

本文所提供的方法可利用将待测序目标核酸片段捕获于固体载体上的步骤。固体载

体可由用作亲和基质或化学和生物分子合成和分析的载体的任何材料形成，例如（但不限于）聚苯乙烯、聚碳酸酯、聚丙烯、尼龙、玻璃、金属、磁性珠粒、乳胶、右旋糖酐、几丁质、砂、浮石、琼脂、多糖、树枝聚合物、巴基球、聚丙烯酰胺、硅、橡胶和用作固相合成、亲和分离和纯化、杂交反应、免疫测定和类似应用的载体的其它材料。本文的固体载体可为微粒状；或可为连续表面的形式，例如经涂布的针具、微量滴定盘或孔、载玻片、金属、塑料或硅芯片、硝基纤维素薄片、尼龙网、例如多孔三维凝胶的多孔三维结构或类似材料。当为微粒状时，微粒通常具有至少一个在 5-10 mm 范围内或更小的维度尺寸。这类微粒（在本文中统称为“珠粒”）通常（但非必需）为球形。然而，此说法并不是约束固体载体的几何形状，其可为任何形状，包括随机形状、针状、纤维状和细长状。也涵盖大致呈球形的“珠粒”，尤其是可用于液相中的微球。“珠粒”可包括额外组分，例如使用磁体进行分离的磁性或顺磁性微粒（例如，参看 Dynabeads7 (Dyna, Oslo, Norway)），只要所述额外组分不干扰本文的方法和分析即可。

例如，在特定实施例中，将 2002 年 4 月 11 日申请的相关美国申请案第 60/372,711 号、2003 年 3 月 24 日申请的第 60/457,847 号和 2003 年 4 月 11 日申请的第 10/412,801 号中所述的杂交芯片用作捕获寡核苷酸阵列的固体载体，例如，通过小室内底表面上固相固体载体表面上的捕获寡核苷酸捕获目标核酸片段，在所述固体载体表面上进行产生目标核酸片段的反应。在特定实施例中，在含有固体载体的小室中或者小室底部即为固体载体的小室中执行断裂反应，其中所述固体载体能够与目标核酸断裂产物特异性杂交并且在除去或洗去小室中其它分子的过程中仍保持断裂产物连接于固体载体。所述相互作用可在目标核酸断裂产物与已固定于固体载体（例如，衍生化或功能化的固体载体）上的捕获寡核苷酸之间进行。可使用任何类型的能够达成特异性捕获目标核酸断裂产物目的的固体载体。

例如，固体载体可以是平坦的二维表面或三维表面，或者可以是珠粒。在平坦固体载体的情况下，小室可由延伸出固体载体表面外的壁（例如，如本文所提供的装置实施例中所述以“障板”提供者）形成，或通过将孔或柱或通道蚀刻于固体载体中以产生不连续和分离小室而形成。可制造固体载体的可能材料包括（但不限于）硅、具有顶部氧化物层的硅、玻璃、如铂或金的金属、如聚丙烯酰胺的聚合物和塑料。在特定实施例中，固体载体为硅芯片或晶片。

平坦固体载体也可经改质以含有导热材料，以促进小室中反应混合物的温度调控。在特定实施例中，固体载体为涂布有金属材料的平坦硅芯片。示范性固体载体已于本文中描述且可与本文所述和提供的设备和方法一起使用。

如上文所述，捕获寡核苷酸排列于对应不连续元件的多个位置（位点）处，一般为每个固相载体（例如芯片）不超过 20,000 个、不超过 15,000 个、不超过 10,000 个、不超过 7,000 个、不超过 5,000 个、不超过 4,000 个、不超过 3,000 个、不超过 2500 个、不超过 2100 个、不超过 2000 个、不超过 1500 个、不超过 1400 个、不超过 1300 个、不超过 1200 个、不超过 1100 个、不超过 1000 个、不超过 900 个、不超过 800 个、不超过 700 个、不超过 600 个、不超过 500 个、不超过 400 个、不超过 300 个、不超过 200 个、不超过 100 个不连续元件。在其它实施例中，阵列含有 4096 个或更少、1536 个或更少、384 个或更少、96 个或更少、64 个或更少个具有捕获寡核苷酸的不连续位置。在特定实施例中，捕获寡核苷酸阵列含有 4096 个捕获寡核苷酸。在阵列含有 4096 个寡核苷酸的一个实施例中，捕获寡核苷酸可为 12 个碱基长。在使用 4096 个寡核苷酸阵列的其它实施例中，捕获寡核苷酸可为 30 个碱基长、25 个碱基长、20 个碱基长、15 个碱基长、10 个碱基长、9 个碱基长、8 个碱基长、7 个碱基长和 6 个碱基长。

在特定实施例中，固体载体上的所有捕获寡核苷酸为全部或部分简并寡核苷酸，例如，含有至少一个通用或半通用碱基。在其它实施例中，固体载体中可含有全部简并、部分简并和/或非简并捕获寡核苷酸的组合。非简并捕获寡核苷酸为一种不含有任何简并碱基（通用或半通用碱基）的寡核苷酸。

可根据捕获寡核苷酸的所需特性以多种方式设计捕获寡核苷酸阵列。组成阵列的捕获寡核苷酸的长度、序列、组成或双链部分的存在与否和其组合可变化。例如，可将阵列设计成具有 12 个碱基长的所有单链捕获寡核苷酸且每个捕获寡核苷酸包括 6 个通用碱基的阵列。或者，可对阵列进行设计以使其含有 50%单链和 50%部分双链的具有各种长度和/或各种组成（例如，不同的通用碱基和/或半通用碱基数）或二者的寡核苷酸。例如，可对阵列进行设计以使其含有长度在 6 至 18 个碱基内的捕获寡核苷酸，且另外或其它，可对阵列进行设计以使其含有具有 6 个与 12 个之间数量的通用碱基或半通用碱基的捕获寡核苷酸。

通常，捕获寡核苷酸探针阵列含有 4 个或 4 个以上核苷酸长、5 个或 5 个以上核苷酸长、6 个或 6 个以上核苷酸长、7 个或 7 个以上核苷酸长、8 个或 8 个以上核苷酸长、10 个或 10 个以上核苷酸长、12 个或 12 个以上核苷酸长或 15 个或 15 个以上核苷酸长的捕获寡核苷酸探针。此外，典型的捕获寡核苷酸探针阵列含有不超过 50 个碱基长、不超过 40 个碱基长、不超过 35 个碱基长、不超过 30 个碱基长、不超过 25 个碱基长、不超过 20 个碱基长、不超过 18 个碱基长、不超过 16 个碱基长、不超过 14 个碱基长、不超过 12 个碱基长、不超过 10 个碱基长或不超过 8 个碱基长的捕获寡核苷酸探针。另

外，捕获寡核苷酸探针可在 3'端、5'端或 3'端与 5'端具有一个或一个以上额外的简并碱基。

所设计的阵列中捕获寡核苷酸的尺寸、组成和双链部分的存在与否可根据根据各种目的加以选择。在一实施例中，可对阵列进行设计以使其含有各自在相同严格度条件下与大约相同数量的不同目标核酸序列杂交的阵列。例如，可对阵列进行设计以使其含有各自在相同杂交条件下（例如，具有相同的熔解温度）与完全互补序列杂交的捕获寡核苷酸。这可例如通过设计具有相同(A+T)/(C+G)比率的引物、通过制造比富含 A/T 的捕获寡核苷酸短的富含 C/G 的捕获寡核苷酸、改变捕获寡核苷酸的长度、包括通用或半通用碱基或包括具有双链区的捕获寡核苷酸来实现。在另一实例中，可对阵列进行设计以使其具有熔解温度不同但在特定条件下与相同数量的不同目标核酸杂交的捕获寡核苷酸。例如，相对于具有较低熔解温度的捕获寡核苷酸而言，具有较高熔解温度的捕获寡核苷酸可具有较短长度或含有较多通用或半通用碱基。因此，在某些杂交条件下，捕获寡核苷酸可与大约相同数量的不同目标核酸序列杂交。例如，第一捕获寡核苷酸中与目标核酸片段杂交的部分可仅含有少量核苷酸，而所述核苷酸可主要是 G 和 C 核苷酸，由于目标核酸中不与第一捕获寡核苷酸杂交的部分的目标核酸序列未受约束，因此导致多个不同的目标核酸片段被结合；对于第二捕获寡核苷酸而言，与目标核酸片段杂交的部分可含有较多核苷酸，而所述核苷酸可包括比 G 和 C 核苷酸杂交能力弱的通用或半通用碱基，由于与捕获寡核苷酸结合的目标核酸序列可根据捕获寡核苷酸中简并碱基的数量变化，因此导致产生多个不同的目标核酸片段被结合；因此，在任何特定杂交条件下与第一和第二捕获寡核苷酸杂交的不同目标核酸序列的总数可大致相同。

或者，可对所设计的阵列中捕获寡核苷酸的尺寸和组成加以选择，使得在所选择的杂交条件下不同的捕获寡核苷酸与不同数量的不同目标核酸杂交。例如，可设计第一捕获寡核苷酸使其在引起第二捕获寡核苷酸与 10 种不同目标核酸杂交的相同条件下与 20 种不同的目标核酸杂交。例如，第一捕获寡核苷酸可含有 6 个非简并碱基和 6 个通用碱基，而第二捕获寡核苷酸可含有与第一捕获寡核苷酸相同的 6 个非简并碱基，和两个另外的非简并碱基；因此，仅一分组与第一捕获寡核苷酸结合的目标核酸也与第二捕获寡核苷酸结合。

也可选择所设计的阵列中捕获寡核苷酸的尺寸、组成和核苷酸序列，以便满足下述标准中的一项或多项：靶向序列的特定类型，例如 SNP 或微卫星；靶向随机或未知序列；控制不同区域处目标核酸的复杂度（例如，通过具有一些双链捕获寡核苷酸以控制一些目标核酸的末端序列部分的复杂度）；和增加或降低与特定捕获寡核苷酸杂交的重叠片

段的数量（例如，通过使用高百分含量的通用或半通用碱基来降低；或通过使用无双链区域并且在视情况除一端或两端外的任何位置处无通用碱基的较短、特定序列来增加）。

G. 特异性或非特异性杂交

本文所提供的方法通常包括杂交两个或两个以上核酸分子的步骤。在本方法中，捕获寡核苷酸可与一个或一个以上目标核酸分子或其片段杂交以形成“捕获寡核苷酸:目标片段复合物”或“捕获寡核苷酸:目标核酸复合物”。所述复合物通常为双链复合物（即，双链体），但也可以是三链复合物。

杂交的程度和特异性随反应条件、尤其关于温度和盐浓度的变化而变化。杂交反应条件通常用严格度来提及，例如，低、中等和高严格度，这些条件是在所属领域技术人员已知和本文举例说明说明的不同温度和盐浓度下达成。因此，在一实施例中，例如，为降低杂交核酸之间不完全匹配的量，可使用较高严格度条件，例如较高温度和/或较低盐浓度。相反，为增加杂交核酸之间所允许的不完全匹配的量，可使用较低严格度条件，例如，较低温度和/或较高盐浓度。

在特定实施例中，用于与目标核酸片段杂交的捕获寡核苷酸不以完全碱基特异性杂交，因此不会排除错配杂交或杂交简并性。这使得杂交严格度降低，因此并非所有理论核苷酸捕获序列的组合都需要呈现于芯片阵列上。如本文所述，可根据经验改变捕获寡核苷酸的简并性和杂交严格度条件，以允许固体载体上存在少至 4096 个或更少个捕获寡核苷酸。错配片段的组成和序列可通过在随后的质谱分析中获取分子量来进行鉴定。

有利地用于本文所提供的方法中的错配杂交的量显著大于在尝试排除所述错配杂交的条件下典型 SBH 方法中发生的不合需要的错配杂交的量。例如，根据本文所提供的方法使用的捕获寡核苷酸可具有两个或两个以上与其杂交的目标核酸片段。在某些情况下，两个或两个以上目标核酸片段可以完全互补性与捕获寡核苷酸杂交；所述情况的实例为与含有两个或两个以上简并核苷酸的捕获寡核苷酸杂交的两个或两个以上目标核酸片段，或比捕获寡核苷酸长且序列根据不与捕获寡核苷酸杂交的片段部分变化的两个或两个以上目标核酸片段。在其它情况下，可选择已降低严格度的杂交条件，从而使两个或两个以上目标核酸片段可与捕获寡核苷酸杂交；在所述情况下，可能需要一个或一个以上目标核酸片段以低于完全互补性的互补性与捕获寡核苷酸杂交。与捕获寡核苷酸杂交的目标核酸片段的示范性所得混合物包括其中没有特定目标核酸片段以大于混合物中目标核酸片段的 95%、90%、85%、80%、75%、70%、65%、60%、55%、50%、45%、40%、35%、30%或 25%的量存在于与捕获寡核苷酸杂交的目标核酸片段的混合物中的目标核酸片段的混合物。在另一实例中，所得混合物包括其中至少两个、至少三个、

至少四个或至少五个目标核酸片段以大于与捕获寡核苷酸杂交的目标核酸分子的 5%、10%、15%或 20%的量存在的目标核酸片段的混合物。在另一实例中，没有目标核酸片段以比与捕获寡核苷酸杂交的目标核酸片段的混合物中至少一个其它目标核酸片段的量大 2 倍、大 3 倍、大 4 倍或大 5 倍的量存在（即，相对于最丰富的目标核酸片段，至少一个其它片段的的存在量为最丰富片段的量的至少 50%、33%、25%或 20%）。

在特定实施例中，设计捕获寡核苷酸，使得每个芯片位置（通常具有多个相同捕获寡核苷酸拷贝）与两个或两个以上目标核酸片段结合。例如，本文中涵盖使 2 个至 500 个、2 个至 400 个、2 个至 300 个、2 个至 250 个、2 个至 200 个、2 个至 150 个、2 个至 100 个、2 个至 75 个、2 个至 50 个、2 个至 40 个、2 个至 30 个、2 个至 25 个、2 个至 20 个、2 个至 15 个、2 个至 10 个或 2 个至 5 个不同目标核酸片段与一种捕获寡核苷酸结合的条件。在所述情况下，不同的目标核酸片段包括作为其它片段的亚片段的片段的结合（例如，产生片段梯）；以及具有相同或不同长度且对特定芯片位置和捕获寡核苷酸具有相似杂交特性但具有不同组成的片段的结合。

在一些实施例中，包括两种或两种以上不同杂交反应（例如，具有两个或两个以上与目标核酸片段接触的不连续位点的阵列）的方法不需要两种或两种以上杂交反应（例如，阵列位置）都产生具有两个或两个以上与之杂交的目标核酸片段的捕获寡核苷酸。在一些情况下，一些反应（例如，阵列位置）可不含有与其杂交的目标核酸片段。在其它情况下，一些反应（例如，阵列位置）可仅含有一个与其杂交的目标核酸片段。通常，所有反应中至少 50%、至少 55%、至少 60%、至少 65%、至少 70%、至少 75%、至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%或至少 99%产生两个或两个以上与捕获寡核苷酸杂交的寡核苷酸，其中所述两个或两个以上捕获寡核苷酸的相对量如本文所提供的含量存在。

为提高杂交效率，可利用通用碱基延长捕获寡核苷酸。例如，捕获寡核苷酸可含有两个区：仅含有通用碱基的第一区；和含有至少一个常见或半通用碱基的第二区。第二区含有用于与目标核酸特异性或半特异性杂交的碱基，而第一区的通用碱基用于稳定捕获寡核苷酸与目标核酸之间的杂交。

此外，由于多种目标核酸可与单一捕获寡核苷酸杂交，故捕获寡核苷酸可将简并碱基并入捕获寡核苷酸的序列识别部分中，产生简并捕获寡核苷酸。如果保持低的芯片阵列位置总数，那么简并捕获寡核苷酸序列识别部分的长度和/或特异性会受到限制。在一实施例中，目标长度为 12 个核苷酸的捕获寡核苷酸将被安置于 4096 个位置中。因此，将其它通用碱基添加至捕获寡核苷酸的一端将在不改变捕获寡核苷酸序列特异性的情

况下显著增加杂交复合物的稳定性，并提高总效率。视进一步的修饰而定，在一实施例中，这些额外的通用核苷酸可朝着捕获寡核苷酸的 3'端安置。在另一实施例中，这些额外的通用核苷酸可朝着捕获寡核苷酸的 5'端安置。在另一实施例中，所述额外的通用核苷酸可安置在捕获寡核苷酸的两端。

对杂交片段的进一步修饰可能增加信息含量和系统的灵活性和稳健性，或降低系统的组成复杂度。例如，以单链特异性 RNA 酶或 DNA 酶（“修剪反应”）处理固相阵列上捕获寡核苷酸:目标片段双链体将杂交片段的总长降低至更均匀的长度。使用修剪会影响初始断裂条件的选择。例如，可放松初始随机断裂方法期间强加的限定，并且可增加断裂尺寸的上限。尺寸为 35 个碱基或更多个碱基的杂交片段可朝着捕获寡聚物的长度缩短和/或缩短至容易由 MALDI-MS 检测的尺寸。本文中涵盖放松断裂参数以改进系统用于各种序列的灵活性。此外，可使用碱基特异性 RNA 酶或 DNA 酶（“碱基特异性修剪”），不必将杂交片段缩短至捕获寡聚物的精确长度，而可将目标核酸片段缩短至最接近捕获寡聚物的目标碱基。所述碱基特异性切割可靶向核苷酸中四个碱基的任一个，因此可根据特定碱基特异性切割反应导致相同的杂交片段被改变成四个不同片段中的一个片段。

使捕获寡核苷酸与目标片段杂交的步骤涉及选择性控制捕获寡核苷酸对对应目标核酸片段的相对亲和力，以足以提供所需程度的捕获寡核苷酸与对应目标核酸片段的杂交，同时排除捕获寡核苷酸对非对应目标核酸片段的相对亲和力。如本文所述，在一实施例中，选择严格度条件以允许捕获寡核苷酸:目标片段双链体中存在一个或一个以上错配。因此，对应于特定捕获寡核苷酸的目标片段中不仅包括含有严格互补序列的片段，而且也可包括具有至少一个或一个以上核苷酸错配的目标核酸片段。在聚集体中，一般将捕获寡核苷酸对错配目标核酸的相对亲和力测量为与一个或一个以上错配目标核酸片段结合的捕获寡核苷酸（例如，捕获寡核苷酸与目标核酸之间具有至少一个单碱基错配）相对于与完全互补的目标核酸片段结合的捕获寡核苷酸的比率。所述比率增加是指，捕获寡核苷酸与错配目标核酸片段的结合相对于捕获寡核苷酸与完全匹配寡核苷酸的结合增加。本文所使用的比率可相应变化，一般为至少约 0.5 倍（即，每结合两个完全互补的目标核酸片段，捕获寡核苷酸探针结合 1 个错配目标核酸）、至少约 1 倍、至少约 1.5 倍、至少约 2 倍、至少约 3 倍、至少约 5 倍、至少约 7 倍、至少约 10 倍、至少约 15 倍或至少约 20 倍。所属领域技术人员可依据多种因素选择所述比率，这些因素包括所研究的目标核酸的长度、不同目标核酸片段的长度和数量、解析所测量得到的质量峰的能力和/或使用所测量得到的质量峰确定目标核酸的核酸序列的能力。

可使用多种方法或测定条件调节每个捕获寡核苷酸对对应目标核酸（例如，捕获寡

聚物以特异性或半特异性亲和力结合的目标核酸)的相对亲和力。在一特定实施例中,通过一种方法在至少部分程度上增加每个捕获寡核苷酸对对应目标核酸的相对亲和力,所述方法包含以下步骤:在杂交步骤中包括一种将与测定探针形成的杂合体的熔解温度标准化、尤其将目标核酸与捕获寡核苷酸之间形成的杂合体的熔解温度标准化以足以提供所述对应目标核酸与其它非对应目标核酸之间的所需区别的试剂。多种适当的标准化试剂包括清洁剂(例如,十二烷基硫酸钠、吐温(Tween))、变性剂(例如,胍、季铵盐)、聚阳离子(例如,聚赖氨酸、精胺)、小沟结合剂(minor groove binder)(例如,偏端霉素(distamycin)、CC-1065,参看Kutyavin等人,1998,美国专利第5,801,155号)等,所述标准化试剂和其用途已描述于本文中和/或另外由此项技术中已知。有效浓度和适当测定条件可根据经验容易地确定(例如参看下文的实例)。

在特定实施例中,变性剂为季铵盐,例如四甲基氯化铵、四乙基氯化铵、四甲基氟化铵或四乙基氟化铵。熔解温度的标准化可通过任何常规方式证实,例如熔解温度的变异系数(coefficient of variance, CV)或标准偏差的降低。例如,可通过将CV或标准偏差降低至少20%、至少40%、至少60%或至少80%将熔解温度标准化。完全匹配的信号与单碱基错配信号之间比率的增加表明,可能需要要求较低的CV。本文中涵盖使用产生下述示范性匹配与错配比率的严格度条件,包括2:1的匹配与错配比率、3:1、4:1、5:1、6:1、7:1、8:1、9:1、10:1、15:1、20:1等的匹配与错配比率。就5:1的匹配与错配示范性比率而言,需要20%或更低的CV以及10%或更低的CV;而对于50:1的匹配与错配比率而言,需要50%或更低的CV。

可通过使用通用碱基或半通用碱基或通过改变杂交条件或二者实现对于与特定捕获寡核苷酸探针杂交的目标核酸序列数量的控制。使用通用碱基组合物和杂交代表两种控制与特定寡核苷酸探针杂交的目标核酸序列数量的单独和独立的方法。所属领域技术人员可基于与捕获寡核苷酸杂交的目标核酸片段的所需复杂度选择使用通用碱基或半通用碱基或改变杂交条件或二者。

可使用通用碱基控制能以相同或相似亲和力与捕获寡核苷酸碱基配对的不同目标核酸序列的理论数量,也可用于确定目标核酸中与捕获寡核苷酸无序列特异性碱基配对的的部分的位置。例如,在捕获探针中使用两个通用碱基允许多达16个不同目标核酸序列以相似亲和力与所述捕获探针碱基配对,并且可已知捕获寡核苷酸上非通用碱基的位置。因此,可控制与捕获寡核苷酸碱基配对的目标核酸序列的数量,并且可已知目标核酸上核苷酸序列可变的核苷酸位置。

操纵杂交条件使得容易地改变杂交条件以便达成实际与捕获寡核苷酸探针杂交的

不同目标核酸序列的所需数量。例如，可用实验法测定在特定杂交条件下与捕获寡核苷酸探针杂交的不同目标核酸序列的数量。所述实验测定后，如果需要，那么可放松杂交条件以允许各种不同目标核酸片段与捕获寡核苷酸探针更多地杂交；或使所述杂交条件更严格，以减少与捕获寡核苷酸杂交的不同目标核酸片段的数量。可改变所述杂交条件数次，以便选择得到与捕获寡核苷酸探针杂交的不同目标核酸片段的所需数量的杂交条件。

除去捕获寡核苷酸与目标核酸片段的非特异性结合的严格度条件，和与高、中等或低严格度实质相同的条件包括：

- 1) 高严格度：0.1 × SSPE, 0.1% SDS, 65EC
- 2) 中等严格度：0.2 × SSPE, 0.1% SDS, 50EC
- 3) 低严格度：1.0 × SSPE, 0.1% SDS, 50EC;

其中 SSPE 一般含有约 150 mM NaCl、10 mM NaH₂PO₄、1 mM EDTA, pH 7.0, 或与其等效的组分。

应了解，可使用备选缓冲液、盐和温度达成等效的严格度。在特定实施例中，为在一个或一个以上捕获寡核苷酸上捕获 1 个以上特定目标核酸片段序列，对于具有极少或无简并核苷酸的捕获寡核苷酸，可将杂交严格度条件放松至中等或低严格度。同样，当所述捕获寡聚物中含有若干个简并寡核苷酸时，可使杂交条件更严格，例如，杂交条件可为高严格度条件。可根据经验选择所述条件，使得不完全排除错配杂交，但同时，仅一分组经断裂目标核酸可与特定捕获寡聚物结合；可改变严格度条件以达到所结合目标核酸片段分组的所需尺寸。

在一实施例中，可从初始杂交条件改变杂交条件。所述改变可为降低或升高杂交条件的严格度。例如，杂交最初可在低严格度杂交条件下进行；随后，所述杂交条件可升至中等或高严格度杂交条件。在另一实例中，可最初在高严格度条件下进行杂交；随后，可将所述杂交条件降低至中等或低严格度杂交条件。

在一实施例中，可改变杂交条件以改变与捕获寡核苷酸探针杂交的目标核酸的数量。例如，可升高杂交条件的严格度以减少与捕获寡核苷酸探针杂交的目标核酸的数量。或者，可降低杂交条件的严格度以增加与捕获寡核苷酸探针杂交的目标核酸的数量。因此，如本文所预期，可改变杂交条件以达成所需的与捕获寡核苷酸探针杂交的目标核酸的数量。

与捕获寡核苷酸探针杂交的目标核酸的数量可通过此项技术中已知的测量与寡核苷酸阵列结合的核酸的任何方法测定，包括：例如可对寡核苷酸阵列（例如寡核苷酸芯

片)进行的光学测量法,例如荧光性或吸光率;散射、放射性、化学发光、量热或磁性标记的检测;一个或一个以上阵列位置的质谱;或此项技术中已知的其它方法,例如美国专利第 6,045,996 号中公开的那些方法。

可使用关于与一个或一个以上捕获寡核苷酸探针杂交的目标核酸数量的一个或一个以上测量值,以将与捕获寡核苷酸探针杂交的目标核酸的实际数量和与所述捕获寡核苷酸探针杂交的目标核酸的所需数量相比较。测量与所述一个或一个以上捕获寡核苷酸探针杂交的目标核酸的数量时,可改变杂交条件以视需要增加或减少与所述捕获寡核苷酸探针杂交的目标核酸的数量。所述过程可反复进行直至达成与所述一个或一个以上捕获寡核苷酸探针杂交的目标核酸的所需数量。

H.修剪

在一些实施例中,可将捕获寡核苷酸:目标片段双链体的单链悬突部分修剪至促进随后的双链体质谱分析并降低组成复杂度的尺寸。例如,当目标核酸片的平均尺寸相对较大时,或当存在大范围不同尺寸的目标核酸片时,可进行修剪。可通过修剪来减小准备用质谱法测量的目标核酸片的尺寸。也可通过修剪来减小准备用质谱法测量的目标核酸片的不同尺寸的范围;和/或减小准备用质谱法测量的片的质量。

修剪方法可通过各种已知方法进行。例如,可通过用酶或化学制品进一步处理捕获片段的阵列来进行修剪以除去未杂交的核苷酸。酶可例如为此项技术中已知的任何核酸外切酶,或“单链特异性 RNA 酶或 DNA 酶”或“碱基特异性 RNA 酶或 DNA 酶”或序列特异性核酸酶。在另一实例中,可使用核酸内切酶(例如单链特异性核酸内切酶)修剪未杂交的核苷酸;在所述修剪反应中,不必除去所有的未杂交核苷酸。单链特异性核酸内切酶可具有序列特异性或不具有序列特异性。例如,酶可以是碱基特异性 RNA 酶或 DNA 酶,并且比捕获寡核苷酸长的杂交片可根据 A、C、G 或 T/U 中一者或一者以上的存在而对 3'端或 5'端或两端进行修剪。

I.有关目标核酸片的信息

重建目标核酸的核酸序列的方法和本文所公开的其它方法(包括鉴定目标核酸的一部分)可利用本文的方法中所提供的各种有关目标核酸和目标核酸片的信息,以重建目标核酸的序列或鉴定所述目标核酸的一部分。所述信息包括质量测量值、质量峰特征、与目标核酸杂交的捕获寡核苷酸的序列、杂交条件和所使用的断裂方法。

1.分子质量

如本文所述,重建目标核酸的核酸序列的步骤和本文所公开的其它方法(包括鉴定目标核酸的一部分)可利用对与捕获核酸杂交的目标核酸片或捕获寡核苷酸:目标片

双链体的分子质量的测定，由此确定所述目标核酸片段的质量。

a. 质谱分析

可使用质谱分析法测定特定分子的质量。所述格式包括（但不限于）基质辅助激光解吸/电离、飞行时间（MALDI-TOF）、电喷射电离（ESI）、IR-MALDI（例如参看已公开的国际 PCT 申请案第 99/57318 号和美国专利第 5,118,937 号）、正交飞行时间（O-TOF）、轴向飞行时间（A-TOF）、离子回旋共振（ICR）、傅立叶转换（Fourier Transform）、线性/反射飞行时间（RETOF）和其组合。也参看 Aebersold 和 Mann, 2003 年 3 月 13 日, *Nature*, 422:198-207（例如，图 2）中有关适用于本文所提供的方法中的质谱法的示范性方法的综述，所述文献全部并入本文用作参考。MALDI 方法通常包括 UV-MALDI 或 IR-MALDI。可通过依赖于质谱法的检测方法和实验方案分析核酸（例如，参看美国专利第 5,605,798 号、第 6,043,031 号、第 6,197,498 号、第 6,428,955 号、第 6,268,131 号和国际专利申请案第 WO 96/29431 号、国际 PCT 申请案第 WO 98/20019 号）。可使这些方法自动化（例如，参看美国公开案 2002 0009394，其描述自动化生产线）。包括（但不限于）曲线场反射器（curved field reflectron）或延时引出飞行时间质谱仪（delayed extraction time-of-flight MS instrument）的中分辨率仪器（medium resolution instrumentation）也能改进用于测序或诊断的 DNA 检测。所述每种仪器都能够检测大于等于 30-mer 链中 9 Da（ $\Delta m(A - T)$ ）的位移。

当使用质谱法（例如 MALDI）进行分析时，可将纳升（nanoliter）体积的样本装载于芯片上。使用所述体积可获得定量或半定量质谱结果。例如，所得质谱中的峰下面积与样本中组分的相对浓度成比例。此项技术中已知制备和使用所述芯片的方法，如美国专利第 6,024,925 号、美国公开案 2001 0008615 和 PCT 申请案第 PCT/US97/20195 号（WO 98/20020）中所举例说明；制备和使用所述芯片的方法也提供于共同待决的美国申请案第 08/786,988 号、第 09/364,774 号和第 09/297,575 号中。进行这些分析的芯片和试剂盒是从 SEQUENOM 以商标 MassARRAY7 购得。MassARRAY7 系统含有可用于 MALDI-TOF（基质辅助激光解吸电离-飞行时间）质谱法的小型化阵列（例如 SpectroCHIP7 阵列）以快速传递结果。其在无标签的情况下准确地辨别与遗传变异体有关的 DNA 片段尺寸的单碱基改变。

i. 所测量的核酸分子的特征

在一实施例中，测量断裂步骤中所形成的所有核酸分子片段的质量。目标核酸分子片段或扩增产物片段的测量质量也可称为“样本”测量质量，与由参考核酸片段得到的“参考”质量相对。

在另一实施例中，使用质谱法测量质量的核酸分子片段的长度不超过 75 个核苷酸长、不超过 60 个核苷酸长、不超过 50 个核苷酸长、不超过 40 个核苷酸长、不超过 35 个核苷酸长、不超过 30 个核苷酸长、不超过 27 个核苷酸长、不超过 25 个核苷酸长、不超过 23 个核苷酸长、不超过 22 个核苷酸长、不超过 21 个核苷酸长、不超过 20 个核苷酸长、不超过 19 个核苷酸长或不超过 18 个核苷酸长。

在另一实施例中，使用质谱法测量质量的核酸分子片段的长度为至少 3 个核苷酸长、至少 4 个核苷酸长、至少 5 个核苷酸长、至少 6 个核苷酸长、至少 7 个核苷酸长、至少 8 个核苷酸长、至少 9 个核苷酸长、至少 10 个核苷酸长、至少 12 个核苷酸长、至少 15 个核苷酸长、至少 18 个核苷酸长、至少 20 个核苷酸长、至少 25 个核苷酸长、至少 30 个核苷酸长或至少 35 个核苷酸长。

在一实施例中，经测量质量的核酸分子片段为 RNA。在另一实施例中，经测量质量的目标核酸片段为 DNA。在又一实施例中，经测量质量的目标核酸片段含有一个经修饰或非典型的核苷酸（即，不同于 DNA 中的脱氧-C、脱氧-T、脱氧-G 或脱氧 A 或不同于 RNA 中的 C、U、G 或 A 的核苷酸）。例如，转录反应的核酸分子产物可含有核糖核苷酸与脱氧核糖核苷酸的组合。在另一实例中，核酸分子可含有常见核苷酸和质量修饰的核苷酸，或可含有常见核苷酸和非天然存在的核苷酸。

ii. 调节 (conditioning)

在质谱分析之前，可处理核酸分子以改进解析度。所述方法称为分子的调节。可“调节”分子，以例如减少挥发所需的激光能量和/或使断裂减至最少。此项技术中已知多种用于核酸分子调节的方法。调节的实例为修饰核酸分子的磷酸二酯骨架（例如，通过阳离子交换），此调节可用于排除因每个核苷酸单元所结合的阳离子的异质性而造成的峰加宽。在另一实例中，使核酸分子与烷基化剂接触可将核酸分子的单硫代磷酸二酯键转换成磷酸三酯键，所述烷基化剂例如为烷基碘、碘代乙酰胺、 β -碘乙醇或 2,3-环氧基-1-丙醇。同样，可使用例如三烷基氯硅烷 (trialkylsilyl chloride) 将磷酸二酯键转换成不带电荷的衍生物。其它调节可包括并入降低对脱嘌呤作用 (MS 期间断裂) 的敏感性的核苷酸，例如，嘌呤类似物 (如 N7-或 N9-脱氮杂嘌呤核苷酸)，或 RNA 结构单元；或使用寡核苷酸三酯；或并入已烷基化的硫代磷酸酯官能；或使用寡核苷酸模仿物，例如 PNA。

iii. 多重作用

对于某些应用而言，可同时检测一种以上核酸分子片段。在其它应用中，可使用例如多种固体载体上的寡核苷酸或寡核苷酸模仿物阵列进行平行加工。“多重作用”可通

过若干不同的方法达成。例如，来自若干不同核酸分子的片段可同时经历质谱测量方法。通常，在多重质量测量中，核酸分子片段应当是足以辨别的，使得可能同时检测多重核酸分子片段。可通过确保所述片段的质量可由待使用的质量测量方法辨别来使核酸分子片段成为可辨别的。这可通过序列本身（组成或长度）或通过质量修饰官能团引入一个或一个以上核酸分子中达成。

b.其它测量方法

此项技术中已知的其它质量测量方法可用于质量测量方法中，包括电泳法，例如凝胶电泳和毛细管电泳；和色谱法，包括尺寸排除色谱法和反相色谱法。

2.质量峰特征

使用质量分析方法（例如本文所述的那些方法），可获得有关目标核酸片段的质量的信息。可由质量测量获得的质量峰的信息包括峰的信噪比、峰面积（例如，以峰下面积或以半高峰宽表示）、峰高、峰宽、相对于一个或一个以上其它质量峰的峰面积、相对于一个或一个以上其它质量峰的峰高和相对于一个或一个以上其它质量峰的峰宽。可通过将扩增片段的至少一个质量峰特征与一种或一种以上参考核酸的一个或一个以上质量峰特征相比较来将所述质量峰特征用于本序列测定方法中，例如，用于鉴定目标核酸分子的核苷酸序列的方法中。

3.捕获寡核苷酸和杂交条件

在包括与捕获寡核苷酸杂交的方法中，所述捕获寡核苷酸通常具有已知核苷酸序列。而且，通常也已知当目标核酸片段与捕获寡核苷酸接触时所使用的杂交条件的严格度。可使用有关捕获寡核苷酸序列的知识和有关杂交条件的知识提供关于与捕获寡核苷酸杂交的目标核酸片段的核苷酸序列的信息。

在构建目标核酸分子的核苷酸序列的方法中，可使用捕获寡核苷酸探针的序列减少以特定观察质量表示的可能目标核酸序列的数量。当已知捕获寡核苷酸序列时，所属领域技术人员可预测在特定杂交条件下与捕获寡核苷酸杂交的目标核酸片段的核苷酸序列。此外，所属领域技术人员可预测在特定杂交条件下可能不与捕获寡核苷酸杂交的目标核酸片段的核苷酸序列。

可能存在的一些核苷酸序列和可能不存在的其它核苷酸序列可辅助解释质量观察值。特定质量的观察值可用于确定以所述质量表示的目标核酸片段的组成（例如，DNA片段中C、G、A和T的数量），但通常在没有更多信息的情况下，不能用于确定以所述质量表示的目标核酸片段的核苷酸序列。因此，通常，特定质量观察值可表示各种不同的目标核酸片段核苷酸序列。质量观察值可补充有杂交信息（捕获寡核苷酸和杂交条件），

可限制或减少以特定质量观察值表示的可能核苷酸序列的数量。经限制或减少的可能核苷酸序列的数量可用于序列构建方法中，或用于与参考相比较，如本文所提供。

在一实例中，四核苷酸捕获寡核苷酸可具有核苷酸序列 5'ACTG3'，且可使目标核酸片段与所述捕获寡核苷酸在高严格度条件下接触，从而使仅与所述捕获寡核苷酸完全互补的目标核酸片段与所述捕获寡核苷酸杂交。此外，在这一实例中，测量与所述捕获寡核苷酸杂交的目标核酸片段的质量，并确定所述片段的组成，其中确定一个质量具有组成 A₃CTG。当组合质量（且因此组成）与杂交信息时，预测 A₃CTG 质量含有一个或一个以上具有核苷酸序列 AAAGT、AAGTA 或 AGTAA 的片段。因此，所述目标核酸分子可含有核苷酸序列 AAAGT、AAGTA 或 AGTAA 中的一者或一者以上。

在具有相同捕获寡核苷酸和杂交条件的类似实例中，未观察到与组成 A₃CTG 对应的质量峰。当与杂交信息组合时，这一观察表明，所述目标核酸分子可能不含有核苷酸序列 AAAGT、AAGTA 或 AGTAA 中的任一个。

在包括比较观察质量特征与参考质量特征的方法中，捕获寡核苷酸序列和杂交条件可以是有关匹配样本图形与参考图形的额外信息来源。例如，可测量阵列中多个捕获寡核苷酸的质量。可观察或计算参考序列以具有所述多个捕获寡核苷酸中每一个的质量特征的特殊图形，产生质量对捕获寡核苷酸的二维图形。根据本文所提供的方法，可将一个或一个以上参考图形与样本的图形相比较，以鉴定目标核酸或鉴定核苷酸序列。

4.断裂方法

用于断裂目标核酸分子的方法可提供可用于核苷酸序列构建或本文所提供的其它方法中的信息。在一实例中，可通过断裂得到具有已知统计学尺寸范围的目标核酸片段。在另一实例中，在与捕获寡核苷酸杂交后，可对片段进行“修剪”以具有与捕获寡核苷酸相同的长度或通常仅略大于捕获寡核苷酸的长度（例如，当碱基特异性断裂时进行修剪）。断裂方法也可限制片段中一个或一个以上核苷酸位点处的核苷酸序列；通常，这种情况在进行序列特异性切割（例如，使用碱基特异性 RNA 酶或限制性核酸内切酶）时发生。因此，可在所产生的片段具有已知尺寸（或尺寸范围）、一些已知核苷酸序列信息或二者的情况下进行断裂方法。

除关于可基于所使用的断裂方法已知的目标核酸片段的信息外，本文所提供的核苷酸序列构建方法可利用通过断裂方法产生重叠片段时所提供的信息。重叠片段的存在提供可用于构建核酸序列或增加核酸序列构建准确性的信息冗余。例如，第一个和第二个目标核酸片段可由在目标核酸中彼此相邻的核苷酸部分产生；第三个目标核酸片段可含有所述第一个目标核酸片段的一部分核苷酸序列和所述第二个目标核酸片段的一部分

核苷酸序列，并且可用于将所述第一个和第二个目标核酸片段鉴定为相邻核苷酸序列，由此用于构建目标核酸的核苷酸序列。

J.核苷酸序列的构建

可使用与目标核酸片段有关的信息构建所述目标核酸分子的核苷酸序列，所述信息例如为断裂方法、质量测量值、质量峰特征和与所述目标核酸片段杂交的捕获寡核苷酸（和杂交条件）。例如，序列构建方法可利用质谱方法根据组分质量分离和测量样本组分的能力。所述序列构建方法也可利用本文所提供的杂交方法在视情况仍产生具有两个或两个以上核酸片段的样本的同时，降低样本中核酸片段的复杂度（例如，核酸片段的数量和/或变异性）。所述序列构建方法还可利用由断裂方法形成的核酸片段的尺寸和/或序列，并且可利用重叠核酸片段的的存在。通过利用这些信息来源，可测定核酸分子的部分或全部核苷酸序列。可将核苷酸序列构建方法用于以下方法中：长程从头测序、长程再测序、长程 SNP 发现、长程突变发现、使用较长序列区进行的细菌分型（例如，使用基于全长 16S rRNA 基因的方法进行的细菌分型）、多重测序（例如，一个实验中的多个较短扩增子）、长程甲基化分析（使用例如甚至具有较少芯片位置的专用甲基化芯片）、人类鉴定（使用例如一个长区或多个短区）、生物体鉴定（使用例如一个长区或多个短区）、病原体和非病原体混合物的分析以及异源核酸混合物的定量。

1.与目标核酸片段有关的信息的作用

本文所提供的关于构建核苷酸序列的方法可基于预测或界定质谱中质量的核苷酸序列界限的能力。例如，关于质谱中质量的预测序列或序列界限可基于例如以下信息：

（1）断裂方法；（2）捕获寡核苷酸；和（3）质量测量值。

如本文所提供，可使用断裂方法产生各种核酸片段，例如，核苷酸长度在特定范围内（例如，长度在 15 至 30 个核苷酸的范围内）的片段；在特定碱基处经切割（例如，碱基特异性切割）的片段；在一个或一个以上特定核苷酸序列处经切割的片段（例如，以序列特异性核酸内切酶消化形成的片段）；或与捕获寡核苷酸具有相同长度的片段（例如，“经修剪”片段）。所得片段具有与断裂方法相关的降低的复杂度。例如，核苷酸长度在特定范围内（例如，长度在 15 至 30 个核苷酸的范围内）的片段池相对于核苷酸长度不在特定范围内的片段池（例如，任何长度的片段）具有降低的复杂度。核苷酸片段的降低的复杂度可用于预测或界定所述片段核苷酸序列的界限。例如，在碱基特异性切割中，所有片段在一端都具有单个特定的核苷酸（经碱基特异性切割的核苷酸）且所述片段的剩余部分具有剩余三个核苷酸中的任一个。所述核苷酸片段降低的复杂度可用于限制与特定捕获寡核苷酸杂交的不同核苷酸片段的数量，和/或限制通过质谱法测量的不

同核苷酸片段的数量。例如，如果所有片段都与捕获寡核苷酸具有相同的长度，那么可将与捕获寡核苷酸杂交的片段的数量和通过质谱法测量的片段的数量限制于仅与所述捕获寡核苷酸互补的片段的数量。

如本文所提供，捕获寡核苷酸可含有具有各种长度的寡核苷酸，且可包括通用碱基和/或半通用碱基。与各捕获寡核苷酸杂交的不同核苷酸片段的数量可根据各捕获寡核苷酸的长度和组成进行控制。例如，仅含有典型核苷酸（例如，A、C、G和T）的较长捕获寡核苷酸相对于仅含有典型核苷酸的较短捕获寡核苷酸可具有较少的与其杂交的不同核苷酸片段。在另一实例中，仅含有典型核苷酸的捕获寡核苷酸相对于含有一个或一个以上通用或半通用碱基具有相同长度的捕获寡核苷酸可具有较少的与其杂交的不同核苷酸片段。对于与特定捕获寡核苷酸杂交的不同核苷酸片段数量的约束可用于预测或界定所述片段核苷酸序列的界限。对于与特定捕获寡核苷酸杂交的不同核苷酸片段数量的约束也可用于限制通过质谱法测量的不同核苷酸片段的数量。

质量测量可用于测定一个或一个以上核苷酸片段的组成。例如，质量测量可用于测定DNA片段中所存在的A、T、G和C的数量。核苷酸片段的组成可用于预测或界定所述片段核苷酸序列的界限。

2. 序列构建方法

例如由断裂、捕获寡核苷酸杂交和质量测量所提供的信息可用于本文所提供的各种方法中，以构建目标核酸分子的核苷酸序列。为构建目标核酸分子的核苷酸序列，本文所提供的教导可指导所属领域技术人员使用通过杂交测序进行核苷酸序列分析的已知技术以及通过质谱法进行核苷酸序列分析的已知技术。例如，可通过已知方法将实验数据转换成 de Bruijn 图的子图；例如，参看 Pevzner, *J. Biomol. Struct. Dyn.*, 7:63-73 (1989)。如此项技术中已知，可在这一图中搜寻欧拉路径（Eulerian path），其中循环（cycle）和凸起（bulge）需预先断裂；例如，参看 Pevzner 等人, *Proc. Natl. Acad. Sci. USA* 98:9748-9753 (2001)。可使用质谱通过此项技术中已知的方法独特地鉴定核酸片段的核苷酸组成；例如，参看 Böcker, *Lect. Notes Comp. Sci.* 2812:476-487 (2003)。如此项技术中已知，可使用测定复合体（compomer）的核苷酸序列的方法，例如分枝限界法（branch-and-bound method），并且于 Böcker, *Lect. Notes Comp. Sci.* 2812:476-487 (2003) 中举例说明。分枝限界法中因假阴性峰引起的问题可通过此项技术中已知的方法解决，如 S. Böcker, "Sequencing from compomers in the presence of false negative peaks" Technical Report 2003-07, Technische Fakultät der Universität Bielefeld, Abteilung Informationstechnik, 2003 中所举例说明；也可访问

http://www.cebitec.uni-bielefeld.de/groups/ims/download/Preprint_2003-07_WeightedSC_SBoecker.pdf。

在一示范性方法中，可构建目标核酸或其片段的假定核苷酸序列，可预测所述片段的断裂/杂交/质量，并且可将预测质量与观察质量相比较以测试所述假定核苷酸序列是否可能存在。在另一实例中，可以使用有关断裂/杂交方法的知识预测所有可能观察到的质量并鉴定与特定质量对应的序列，随后，可将这一信息与观察质量相比较以限制可存在于目标核酸分子中的不同核苷酸序列的数量。下文提供使用这一信息构建核苷酸序列的示范性方法。

a. 假定序列的测试

在使用断裂、杂交和质量测量信息的一种示范性方法中，可构建目标核酸或其片段的假定核苷酸序列，可预测所述片段的断裂/杂交/质量，并且可将预测质量与观察质量相比较以测试所述假定核苷酸序列是否可能存在。这一方法可通过以下步骤进行：构建一部分目标核酸分子（例如，一个核苷酸片段）的假定核苷酸序列，且在测定所述部分的核苷酸序列时，将一个或一个以上额外的假定核苷酸添加到所述部分中，并测试所述额外的假定核苷酸是否可能存在。

在一实例中，目标核酸分子在一端或两端（例如，3'端或5'端，或两端）处可具有已知的核苷酸序列。例如当用具有已知核苷酸序列的引物扩增目标核酸分子时，可出现这一情况。可将一个或一个以上假定核苷酸加入到已知序列中，并且可通过参考观察质谱测试所述假定核苷酸的存在。假定核苷酸与实际核苷酸之间的错配导致实验观察质谱中不存在的假定质量的存在，和/或实验观察质谱中存在的假定质量的不存在。因此，可将产生与实验观察质量最紧密匹配的预测片段质量的假定核苷酸鉴定为存在于目标核酸分子中对应位置处的核苷酸。

各质谱中多个质量的存在或不存在可用于确定存在四个核苷酸中的哪一个核苷酸，并提供信息冗余，由此增加准确测定序列的可能性。例如，可通过比较单一质谱的预测质量与观察质量来确定特定核苷酸位置处核苷酸的身份；除此确定外，还可参考一个或一个以上额外的质谱获得证实或反驳所述确定的其它信息。多个质谱可增加用于鉴定特定核苷酸的观察的数量，因此也可增加准确鉴定核苷酸的可能性。

基于核苷酸假说测试进行序列构建的一种示范性方法如下：

- (1) 在一个或一个以上特定位置处指定假定核苷酸；
- (2) 根据所述断裂方法预测含有所述核苷酸的片段；
- (3) 对于每个捕获寡核苷酸，预测是否存在所述预测片段与所述捕获寡核苷酸的

杂交；

(4) 计算每个捕获寡核苷酸的杂交片段的质量/组成；和

(5) 比较预测质量与观察质量；

预测质量与观察质量之间的匹配可将所述假定核苷酸鉴定为目标核酸分子核苷酸序列中的实际核苷酸。

可视需要在每个核苷酸位置处对所有四个常见核苷酸（例如，对于 DNA 而言为 A、G、C 和 T）重复这一方法，并且可将预测质量与观察质量最紧密匹配的核苷酸选为目标核酸分子中所述位置处存在的核苷酸。可同时通过这一方法测试单个或多个核苷酸位，且根据观察的数量（例如，所存在的的质量的数量和不存在的的质量的数量）、质谱（例如，可存在于质谱中的不同序列的数量）和目标核酸分子的长度，根据本文所提供的指导方针和此项技术中已知的方法可确定将同时测试的核苷酸位置的数量。

在基于核苷酸假说测试进行序列构建的特定说明性实例中，可断裂具有（未知）核苷酸序列 ACATGAGCTTACAAC (SEQ ID NO: 1) 的目标寡核苷酸以得到 5-7 个核苷酸长的片段。接下来，所述核酸片段可与具有四个半通用碱基（例如，仅结合嘧啶 (Y) 或仅嘌呤 (R) 的碱基）的杂交区的捕获寡核苷酸杂交。接着，可通过质谱法检测杂交片段。出于这一实例的目的，已知目标寡核苷酸的前七个核苷酸的序列为 ACATGAG。可试验性地将第八个核苷酸指定为四个可能常见核苷酸中的任一个，例如“T”。基于含有序列 ACATGAGT 的寡核苷酸，可根据所测量的每个不同捕获寡核苷酸序列的每个质谱预测质量。例如，当将所述核苷酸位置试验性地指定为“T”时，预测具有序列 RYYY 的捕获寡核苷酸探针的质谱含有与组成 T_2G_2A 、 $T_2G_2A_2$ 和 $T_2G_2A_2C$ 对应的质量。对核苷酸序列 ACATGAGCTTACAAC (SEQ ID NO: 1) 而言，在实验中对所述捕获寡核苷酸仅观察到 $T_2G_2A_2C$ 。类似地，“G”的存在将得到三个预测质量，所有所述预测质量在实验中都不存在于这一捕获寡核苷酸中。当将八个位置预测为“A”时，三个预测质量中的两个在实验中存在；而当将八个位置预测为“C”时，观察到所有对应的实验质量。因此，“C”提供最接近的匹配。为进一步证实这一位置处存在“C”，可比较来自一个或一个以上其它捕获寡核苷酸的图谱的质量。例如，如果“A”存在，那么具有序列 YYYY 的捕获寡核苷酸的质谱包括对应于 TG_2A_2 的质量。实验中未观察到所述质量；但捕获寡核苷酸 YYR 的质谱具有对应于组成 TG_2AC 的质量，表明“C”（可能）存在于这一位置处。

在本实例中，可使用 16 种不同的捕获寡核苷酸，且每一捕获寡核苷酸都可与若干个含有重叠序列的核酸片段杂交（例如，当片段为 5-7 个核苷酸长时，具有重叠序列的

9种不同片段可与相同的4个核苷酸长的捕获寡核苷酸杂交)。因此,在本实例中,单个质谱的多达9种不同质量可提供有关特定核苷酸位置处核苷酸身份的信息,并且可收集16种不同的质谱。因此,可使用大量信息鉴定这一目标寡核苷酸的每个核苷酸位置处的核苷酸。

b.限制可能的序列

在一实例中,可使用断裂方法和捕获寡核苷酸的组成界定或限制与捕获寡核苷酸杂交的核苷酸片段的质谱中以特定质量呈现的可能核苷酸序列的数量,也可用于界定或限制与捕获寡核苷酸杂交的核苷酸片段的质谱中可存在的可能质量的数量。例如,将所有片段切割成8个核苷酸长的断裂方法限制可出现在 4^8 的不同核苷酸序列的数量,甚至进一步限制质谱中可能的不同质量的数量。与特定4核苷酸序列在核苷酸片段的3'端处杂交的捕获寡核苷酸进一步限制可出现(特定捕获寡核苷酸位置处)在 4^4 的可能核苷酸序列的数量,甚至进一步限制质谱中可能的不同质量的数量。

这些界限可应用于实验测量的质谱中,以得到目标核酸分子的可能核苷酸序列的界限。所述界限可为正(例如,特定核苷酸序列存在或可能存在于所述目标核酸分子中)或负(例如,特定核苷酸序列不存在于所述目标核酸分子中)。例如,可限制由上述示范性断裂和捕获寡核苷酸条件所得的片段质量以与24个或更少个可能核苷酸序列对应,导致将所述目标核酸分子的8核苷酸区段限制于24个或更少个核苷酸序列中的一个。同样,不存在任何具有特定质量的片段可表明,目标核酸分子中不存在得到这一质量的核苷酸序列。在进一步求精中,可比较多种不同捕获寡核苷酸的质谱,并且多个质谱的负界限和正界限可减少可存在于特定观察质量处的可能序列的数量。

当观察的数量(包括特定质量存在或特定质量不存在的观察)足够大,而质谱(例如可存在于每一质谱中的不同序列的数量)相对于待构建的核苷酸序列足够简化(如可通过已知方法根据本文所提供的教导确定)时,可部分或完整构建目标核酸分子的核苷酸序列。例如,在一些情况下,所观察的核苷酸片段的组成(其可例如由观察质量确定)可具有所指定的核苷酸序列;且当足量的核苷酸片段、尤其重叠片段具有所指定的核苷酸序列时,可由此构建目标核酸分子的整个核苷酸序列。在另一实例中,所观察的核苷酸片段组成不具有所指定的核苷酸序列;尽管如此,仍可使用对所述片段中可能核苷酸序列的限制,例如通过提供足够的界限以确定片段之间的重叠并提供足够的界限以基于片段之间的重叠确定所述片段的序列,确定目标核酸分子的序列。在另一实例中,具有指定核苷酸序列的片段可与具有非指定核苷酸序列但具有其核苷酸序列界限的片段一起使用。

基于限制核苷酸片段和/或目标核酸分子的可能序列进行序列构建的一种示范性方法可根据以下步骤进行：

- (1) 界定或确立核酸断裂的片段产物的界限；
- (2) 界定或确立可与每个特定捕获寡核苷酸杂交的核酸片段的界限；
- (3) 预测可于与捕获寡核苷酸杂交的核苷酸片段的质谱中观察到的可能质量；
- (4) 建立可能存在于特定观察质量中的可能核苷酸序列的限制规则集；和
- (5) 将观察质量与上述规则集相比较以鉴定可能存在的可能序列和/或鉴定不存在的序列。

3.确定方法稳健性的指导方针

根据随本文所提供的方法而变化的要素，所属领域技术人员可确定构建序列的目标核酸分子的长度，和/或正确测定序列的可能性程度。此外，根据要构建序列的目标核酸分子的长度和/或所需的正确测定序列的可能性程度，所属领域技术人员可设计本文所提供的方法。例如，本文所提供的方法可控制可用于序列构建的实验信息量和实验信息表示目标核酸分子中存在或不存在独特核苷酸序列的程度。

例如，本文所提供的方法可控制可用于核苷酸序列构建中的不同观察质量的数量。所观察的质量可例如为存在于质谱中的质量，或不存在于质谱中的质量（例如，在可能的核苷酸片段的质量处不存在峰）。质谱中观察质量的数量可受所使用的断裂方法和所使用的杂交方法（例如，杂交条件和捕获挂核苷酸的序列）影响。例如，相对于得到 5-15 个核苷酸长的片段的目标核酸分子的断裂，仅得到 10 个核苷酸长的片段的目标核酸分子的断裂可减少观察质量的数量。观察质量的数量也可受对不同杂交反应（例如，不同杂交条件和/或不同捕获寡核苷酸序列）收集的质谱的数量影响。

本文所提供的方法也可控制具有可呈现于相同质谱中的相同质量的核苷酸序列的数量和/或变异性。例如，本文所提供的断裂和杂交方法可影响具有相同核苷酸组成并且可存在于相同质谱中并因此呈现于质谱的相同质量峰中的不同核苷酸序列的数量。

所属领域技术人员已知测定可获得的有关例如观察数量和可呈现于相同观察中的不同核苷酸序列的数量的实验信息的方法。测定可获得的实验信息后，所属领域技术人员可估计核酸分子的长度和/或核苷酸序列测定的可能性程度。或者，基于所需目标核酸分子的长度和/或核苷酸序列测定的所需可能性程度，所属领域技术人员可设计实现所需结果的断裂方法和/或杂交反应的数量和类型。

K.通过质量图鉴定核苷酸序列

在另一实施例中，本文提供一种鉴定目标核酸分子的核苷酸序列的方法，包含：

(a) 使目标核酸分子的片段与捕获寡核苷酸探针杂交，其中两种或两种以上不同核酸片段与所述捕获寡核苷酸探针杂交；

(b) 测量与所述捕获核酸探针杂交的所述目标核酸片段的质量；

(c) 将样本质量与一个或一个以上参考质量图形相比较；

(d) 鉴定与所述样本质量匹配的参考质量图形；

由此，所述样本质量与参考质量图形之间的匹配将所述目标核酸分子中的核苷酸序列鉴定为与参考研苷酸序列对应。在所述方法中，质量峰的两个或两个以上特征可用于鉴定目标核酸中的序列。在所述鉴定方法中，质量峰的两个或两个以上特征的集合称为“图形”。

在本文所提供的方法中，特定核苷酸序列可产生用作所述核苷酸序列的独特特征的质量图形。例如，特定核苷酸序列可产生仅当目标核酸含有该核苷酸序列时形成的质量图形。在所述情况下，无需通过核苷酸序列构建鉴定所述核苷酸序列—而是可简单地通过将观察图形与参考图形（其中所述参考图形对应于一特定核苷酸序列）相匹配来鉴定所述核苷酸序列。

质量图形可存在于单一质谱中，或可存在于两种或两种以上不同杂交反应的质谱中。参考图形可为经计算的图形或实验观察的图形。在参考图形是实验观察图形的情况下，核苷酸序列的鉴定不受重复性误差（例如，质谱中的误差，其中计算为存在或不存在的峰分别重复不存在或存在）存在的影响。

在一些实施例中，可使通过图形匹配进行的序列鉴定与本文所提供的核苷酸序列构建组合。例如，可通过图形匹配测定一部分目标核酸分子的核苷酸序列；并且可通过核苷酸序列构建方法确定所述目标核酸中所述部分的位置和/或所述目标核酸分子剩余部分的核苷酸序列。在其它实施例中，通过图形匹配进行的序列鉴定可用于鉴定目标核酸分子的全部核苷酸序列。

在一些情况下，例如再测序和 SNP 分析，目标核酸分子中可能会存在先前已知的序列（例如，公共数据库的序列），然而所关注的特定目标核酸的序列尚未知。在其它情况下，目标核酸片段中特定核苷酸序列的质量图形可以是已知的。在任一种情况下，可能通过测量与一个或一个以上捕获寡核苷酸杂交的目标核酸片段的质量图形并将所述图形与经计算或根据实验测定的质量图形相比较来鉴定目标核酸中的核苷酸序列。

待鉴定的质量峰可具有三个或三个以上鉴定特征，包括捕获寡核苷酸阵列上的位置（即，与目标片段杂交的特定捕获寡核苷酸，和当已知捕获寡核苷酸序列时，与目标核酸片段杂交的序列）、测量质量和质量测量的信噪比。本文预期，在通过质量图形匹配

进行的核苷酸序列测定的方法中可以使用少至 1 个或少至 2 个质量峰鉴定特征。

在分析已知序列（例如，再测序或基因分型方法）时，可使用经计算的质量图形或根据实验测定的质量图形鉴定可鉴定目标核酸中的核苷酸序列的一个或一个以上质量峰特征。例如，可通过测定指示所论及的 SNP 位置处是否存在特定核苷酸的一个或一个以上峰来进行 SNP 分析。因此，鉴定一个或一个以上指示质量峰是否存在可用于鉴定所论及的 SNP 位置处的核苷酸，而无需核苷酸序列构建方法测定目标核酸分子的所有或任一种核苷酸序列。

断裂和杂交图形的计算可鉴定可用于预测质量图形或质量峰特征图形的质量峰。所述方法可产生任一个或所有的质量峰特征，包括在捕获寡核苷酸阵列上的特定位置处是否片段、片段质量和质量峰的信噪比。在一些情况下，通过重复这些对所论及的相同位置中不同核苷酸序列的计算，可能产生在目标核酸上的一个或一个以上核苷酸位置处指示不同核苷酸序列的一个或一个以上质量峰的多个不同（并且互斥）的集合。

对样本目标核酸片段的实验分析可产生可与一个或一个以上指示计算序列的质量峰的集合相比较的质量峰，且可将所述一个或一个以上指示理论计算序列质量峰的集合与实验质量峰相关联。随后，视情况，只要相关性在用户界定的临界量以上，就可将样本核酸的完整序列或部分序列鉴定为参考序列，所述参考序列对应于与实验质量峰最紧密相关的指示计算序列质量峰的集合。可于实验得到的参考质量图形与样本目标核酸分子的质量图形之间进行类似关联。

样本峰与参考峰的关联可用所属领域技术人员已知的任何方式进行。在简单实例中，为特定捕获寡核苷酸提供的一个参考质量可仅以多个参考质量峰图形中的一个存在。如果检测到样本目标核酸分子的这一相同质量，那么可将所述目标核酸分子的至少部分核苷酸序列鉴定为与参考质量峰对应的核苷酸序列。样本峰与参考峰之间的关联也可使用考虑多个峰的统计学方法和使用已知用于数据关联的其它方法进行，所述统计学方法包括回归法，例如线性回归或非线性回归。

在一实施例中，用户可界定临界值，用于设定参考核酸以足够的可能性鉴定目标核酸中的核苷酸序列所需的最小相关性。当关联不是发生于临界值以上时，无参考核酸能以足够的可能性鉴定目标核酸中的核苷酸序列。

在一实施例中，与阵列单一位置中的捕获探针杂交的目标核酸片段的质量图形可用于鉴定目标核酸的一个或一个以上序列或部分目标核酸。例如，当样本目标核酸为来自生物体的染色体，并且测试目标核酸的特定基因或序列以确定（例如）基因表达、基因型、物种（species）和变种（variety）时，与阵列单一位置中的捕获探针杂交的目标核

酸片段（例如，所有目标核酸片段都与全部具有相同核苷酸序列的捕获寡核苷酸探针杂交）的质量图形可指示特定基因表达、基因型、物种或变种，或可指示目标核酸不与特定基因表达、基因型、物种或变种对应。

在其它实施例中，与多个捕获探针阵列位置杂交的目标核酸片段的质量图形可用于鉴定目标核酸中的核苷酸序列，其中所述目标核酸片段与位于阵列中的 500 个或更少个位置中、阵列中的 250 个或更少个位置中、阵列中的 100 个或更少个位置中、阵列中的 75 个或更少个位置中、阵列中的 50 个或更少个位置中、阵列中的 25 个或更少个位置中、阵列中的 20 个或更少个位置中、阵列中的 15 个或更少个位置中、阵列中的 10 个或更少个位置中、阵列中的 8 个或更少个位置中、阵列中的 6 个或更少个位置中、阵列中的 5 个或更少个位置中、阵列中的 4 个或更少个位置中、阵列中的 3 个或更少个位置中或阵列中的 2 个或更少个位置中的捕获探针杂交。

在不需要核苷酸序列构建的方法中，可使用产生重叠的目标核酸片段，但非必需。例如，在再测序方法中或鉴定 SNP 序列的方法中，可产生不重叠的目标核酸片段，并且可测定所有或部分核苷酸序列。在例如 SNP 鉴定的应用中，可使用少至一个目标核酸片段指示所述 SNP 位置处目标核酸的核苷酸序列。

L. 鉴定一部分目标核酸

在另一实施例中，本文提供一种鉴定一部分目标核酸的方法，包含：

(a) 使所述目标核酸的片段与捕获寡核苷酸探针杂交，其中两个或两个以上不同目标核酸片段与所述捕获寡核苷酸探针杂交；

(b) 测量与所述捕获寡核苷酸探针杂交的所述目标核酸片段的质量；和

(c) 将所述质量与参考核酸分子的片段的质量相比较；

由此，一个或一个以上样本质量与一个或一个以上参考质量之间的相关性将一部分目标核酸鉴定为与所述参考核酸分子对应。在所述鉴定方法中，两个或两个以上质量峰特征的集合称为“图形”。

在一实施例中，可能使用与一个或一个以上捕获寡核苷酸杂交的目标核酸片段的质量图形鉴定目标核酸的一个或一个以上部分，而无需测定所述目标核酸的整个核苷酸序列。在另一实施例中，在不测定目标核酸的任一核苷酸序列的情况下鉴定目标核酸的一个或一个以上部分。

在一些情况下，可已知说明目标核酸分子或其片段位于何处的参考核酸质量图形，即使所述目标核酸的序列尚未知。例如，染色体可具有目标核酸片段图（类似于 RFLP 或 AFLP 图），但所有染色体或仅一分组染色体可能具有已知的核苷酸序列。不管核苷

酸序列已知或未知，都可能通过测量与一个或一个以上捕获寡核苷酸杂交的目标核酸片段的质量图并将所述图形与经计算（在已知序列的情况下）或根据实验测量的质量图形相比较来鉴定一部分目标核酸分子。

当未知所论及的区域的序列时，仍然可通过将目标核酸片段的一个或一个以上质量峰与来自一个或一个以上参考核酸的一个或一个以上质量峰相比较实现对于目标核酸的一个部分或一个以上部分的鉴定。本方法可类似于传统的 DNA 指纹法，其中将未知样本的一个或一个以上凝胶电泳谱带与一个或一个以上已知或参考样本的一个或一个以上凝胶电泳谱带相比较。在本方法中，例如，由样本目标核酸测量的质量峰的三个特征（即，阵列上的位置、质量和信噪比）中的一个或一个以上特征可与由一个或一个以上参考核酸测量的质量峰的一个或一个以上特征比较，且可将所述一个或一个以上参考质量峰与所述样本目标核酸质量峰相关联。随后，将所述样本目标核酸部分鉴定为与具有一个或一个以上与所述样本目标核酸质量峰最紧密相关的质量峰的参考核酸部分对应，并且视情况，只要所述相关性在用户界定的临界量以上。因此，即使所论及的所述部分的序列和位置都未知，也可通过将特定参考核酸鉴定为具有相同质量图形来实现对目标核酸的一个或一个以上部分的鉴定。

在一实施例中，与阵列中单一位置中的捕获探针杂交的目标核酸片段的质量图形可用于鉴定一部分目标核酸。例如，当样本目标核酸是来自生物体的染色体，并且测试目标核酸的（例如）基因表达、基因型、物种和变种时，与阵列单一位置中的捕获探针杂交的目标核酸片段的质量图形可指示特定基因表达、基因型、物种或变种，或可指示目标核酸不与特定基因表达、基因型、物种或变种对应。

在其它实施例中，与多个捕获探针杂交的目标核酸片段的质量图形可用于鉴定一部分目标核酸，其中所述目标核酸片段与位于阵列中的 500 个或更少个位置中、阵列中的 250 个或更少个位置中、阵列中的 100 个或更少个位置中、阵列中的 75 个或更少个位置中、阵列中的 50 个或更少个位置中、阵列中的 25 个或更少个位置中、阵列中的 20 个或更少个位置中、阵列中的 15 个或更少个位置中、阵列中的 10 个或更少个位置中、阵列中的 8 个或更少个位置中、阵列中的 6 个或更少个位置中、阵列中的 5 个或更少个位置中、阵列中的 4 个或更少个位置中、阵列中的 3 个或更少个位置中或阵列中的 2 个或更少个位置中的捕获探针杂交。

在不需要核苷酸序列构建的方法中，可使用产生重叠的目标核酸片段，但非必需。例如，可使用目标核酸片段图形鉴定生物体、品系（strain）或物种，其中所述图形中所使用的两个或两个以上质量峰特征的每一个都源于表示目标核酸中的不相邻序列的目

标核酸片段；可将这一图形与一个或一个以上参考核酸图形相比较，并且通过将所述样本图形与所述一个或一个以上参考图形相关联来鉴定生物体、品系或物种。

M.应用

可使用本文所公开的方法得到有关目标核酸分子的信息而用于各种目的。下文所公开的应用提供本文所公开的方法的示范性用途。所属领域技术人员应了解，下文所述的应用可使用构建目标核酸的核苷酸序列的方法进行，也可使用鉴定一部分目标核酸的方法（例如要求分析目标核酸质量峰图形的方法）进行。

1.长程再测序

除上文所述的长程从头测序方法外，本文所提供的测序方法也可用于长程再测序。来自各种生物体的可用基因组序列信息量的急剧增长增加了对于允许大规模比较序列分析以便将序列信息与功能、基因型或同一性相关联的技术的需求。所述用于比较序列分析的技术的应用可极为广泛，包括例如，SNP发现和病原体的序列特异性鉴定。因此，再测序和高通量突变筛选技术对于鉴定潜在的疾病突变以及潜在的差异药物反应和对治疗方案差别反应遗传变异性至关重要。

已研发出多种方法来满足这些需要。高通量 DNA 测序的技术包括使用电泳和激光诱导荧光检测的 DNA 序列分析仪。基于电泳的测序方法对于检测杂合子（heterozygote）具有先天限制并且受 GC 压缩影响。因此，在不使用电泳的情况下产生数字信号的 DNA 测序平台可克服这些问题。基质辅助激光解吸/电离飞行时间质谱（MALDI-TOF-MS）利用数字数据输出测量 DNA 片段。本文所提供的特异性切割断裂分析方法允许相对于参考序列以高通量、高速和高准确性说明核酸序列。本方法使得可能常规使用 MALDI-TOF MS 测序用于准确的序列校正以及突变检测，例如筛选 BRCA1 和 BRCA2 中的原始基因突变（founder mutation），此突变与乳癌的发展相关。

再测序方法可使用本文所公开的各种用于目标核酸分析的方法进行。例如，可使用可用于测定大核酸区段的核苷酸序列的序列构建方法进行再测序。在另一实例中，可使用鉴定一部分目标核酸的方法；例如，当所述目标核酸可与已知或参考核酸仅有小百分含量（例如，5%或更小）不同时，可使用例如质量峰图形分析的方法鉴定所述不同的核苷酸位置 and 不同核苷酸位置处核苷酸的身份。因此，例如，当公共数据库核苷酸序列含有错误时，本文所公开的各种方法可用于校正一个或一个以上所述错误。

2.突变/序列变异的长程检测

本文的目的是提供可用于鉴定遗传性疾病和其标记的经改进的比较核酸测序方法。由本文所提供的方法鉴定的序列变异候选者包括含有多态性序列变异的序列。多态性包

括天然存在、体细胞序列变异与由突变引起的变异。多态性包括（但不限于）：序列微变异，包括 SNP，其中局部区域中的一个或一个以上核苷酸随个体不同而不同；插入和缺失，使尺寸由一个核苷酸变化成数百万个碱基；和微卫星或核苷酸重复，其重复数可变化。核苷酸重复包括同源重复，例如二核苷酸、三核苷酸、四核苷酸或较大重复，其中相同序列重复数次；和异核苷酸（heteronucleotide）重复，其中发现序列基序（sequence motif）重复。对于给定位点而言，核苷酸重复的数量可视个体而变化。

多态性标记或位点是出现分歧的位点。所述位点可小至一个碱基对（例如，SNP）。多态性标记包括（但不限于）限制性片段长度多态性（restriction fragment length polymorphism, RFLP）；可变数目串联重复（variable number of tandem repeat, VNTR's）；高变区；微卫星；二核苷酸重复；三核苷酸重复；四核苷酸重复；和其它重复模式，例如卫星和小卫星、简单序列重复和插入因子（例如 Alu）。基因的多态性也表现为不同的孟德尔等位基因（mendelian allele）。可通过蛋白质差异、蛋白质修饰、RNA 表达修饰、表观基因组差异（epigenomic difference）、DNA 和 RNA 甲基化、改变基因表达和 DNA 复制的调控因子以及和基因组核酸或细胞器核酸中的任何其它变化表现来观察多态性。

此外，许多基因具有多态区。由于个体具有多个多态区等位基因变异体中的任一个，故可基于基因多态区等位基因变异体的类型鉴定个体。这可用于（例如）法医目的。在其它情况下，了解个体具有的等位基因变异体的身份尤为重要。例如，在移植物排斥或移植物抗宿主疾病（例如骨髓移植）中涉及例如主要组织相容性复合体（major histocompatibility complex, MHC）基因的某些基因中的等位基因差异。因此，特别需要开发出用于确定基因或基因损伤多态区中等位基因变异体的身份的快速、灵敏且准确的方法。如本文所提供的方法或试剂盒可通过确定受检者体内一个或一个以上基因或染色体中一个或一个以上多态区的一个或一个以上等位基因变异体的身份来对所述受检者进行基因分型。使用本文所提供的一种或一种以上方法对受检者进行基因分型可用于法医或身份测试目的，并且所述多态区可存在于（例如）线粒体基因中，或可为短的串联重复。

单核苷酸多态性（SNP）一般为双等位基因系统（biallelic system），也就是说，个体可具有两个用于任何特定标记的等位基因。这意味着，当与可具有多达 10 个等位基因的微卫星标记相比较时，每个 SNP 标记的信息含量相对较低。SNP 也倾向于极具种群特异性（population-specific）；一个种群的多态性标记可能在另一种群中并不十分具有多态性。大约每千碱基可见的 SNP（参看 Wang 等人 *Science* 280:1077-1082 (1998)）可能

会产生极高密度的遗传图谱，而可用于研究所关注的基因或区域的单倍体分型系统，并且由于 SNP 的性质，使得其实际上可以是与所研究的疾病表型相关的多态性。SNP 的低突变率也使其成为研究复杂遗传特质的优良标记。

基因组学更多关注于 SNP 的鉴定，其重要性来自各种原因。SNP 允许间接测试（单倍型的相关性）和直接测试（功能性变异体）。SNP 是最丰富、最稳定的遗传标记。一般疾病最好通过一般基因改变加以解释，并且人类群体的天然变异帮助了解疾病、治疗和环境相互作用。

3.多重测序

本文也涵盖以高通量说明来自多个目标核酸序列的核酸序列的方法。多重反应是指同时说明一个以上目标核酸序列。已知进行多重反应、尤其与质谱法一起进行多重反应的方法（例如，参看美国专利第 6,043,031 号、第 5,547,835 号和国际 PCT 申请案第 WO 97/37041 号）。

在一个实验中，可使用目标核酸的多个较短扩增子进行相同目标核酸序列多个较短区域的多重反应。多重反应提供的优势在于，与需要对每个目标核酸序列进行单独质谱分析相比，可在少至一个质谱中对多种核酸测序。本文所提供的方法使其自身成为以高速度和准确性说明核酸序列的高通量、高度自动化方法。

多重反应可用于测定目标核酸的完整序列；测定目标核酸的至少一个核苷酸而非所有核苷酸的序列；鉴定目标核酸的一个或一个以上部分；或鉴定含有多种目标核酸的样本中一个或一个以上特定目标核酸的存在，或鉴定其存在和相对浓度。在一实施例中，目标核酸为两个或两个以上 mRNA 核酸，或使用两个或两个以上 mRNA 核酸的模板所形成的经扩增核酸。在所述方法中，可检查一个或一个以上细胞的基因表达概况，所述细胞包括组织样本或血液或骨髓样本。例如，两个或两个以上质量峰可指示两个或两个以上 mRNA 的表达，并且所述两个或两个以上质量峰的测量值可揭示是否所述 mRNA 中每一个都存在于目标核酸样本中，和目标核酸样本中存在的 mRNA 的含量。所述方法可用于检查各种 mRNA 的表达水平，所述 mRNA 包括例如，指示细胞的肿瘤状态（neoplastic）或转移（metastatic）的致癌基因和其它基因；编码细胞表面蛋白的基因；与遗传病症相关的基因；指示受病原体感染或细胞其他疾病状态的 mRNA；和与活化细胞毒性细胞相关的基因。所述方法也可用于测定各种不同样本中一个或一个以上基因的表达水平，所述不同样本包括例如不同细胞类型、不同组织类型、不同生物体、不同品系、不同物种或新细胞类型、新组织类型、新生物体、新品系和新物种。不同样本中表达水平的测定可用于例如确定细胞的转移状态，诊断受检者，包括患有遗传性、感染性、

自体免疫性或肿瘤疾病的患者；辨别细胞类型、组织类型、品系类型或生物体类型；确定两个或两个以上基因之间表达的联系；或确定基因表达与细胞形态（例如细胞的有丝分裂或减数分裂状态）之间的相关性。

来自任何两种或两种以上生物分子来源的生物样本的混合物可汇集于单一混合物中供本文分析。例如，本文所提供的方法可用于对不同来源的目标核酸或氨基酸的多个拷贝测序，并因此检测生物样本中核酸混合物中目标核酸或氨基酸的序列变异。生物样本混合物也可包括（但不限于）来自个体组合的核酸，或来自一个或一个以上个体的不同区域的核酸，或得自单一组织或细胞类型的同源肿瘤样本，或含有一种以上组织类型或细胞类型的异源肿瘤样本，或得自原发性肿瘤的细胞系。也涵盖例如单倍体分型方法的方法，其中于相同基因中检测到两种突变。

4.长程甲基化模式分析

本文所提供的方法可用于说明核酸序列变异，其为目标序列的基因外改变（epigenetic change），例如所述目标序列的甲基化模式改变。细胞甲基化分析是一种新兴的研究学科。甲基于胞嘧啶中的共价添加主要出现在 CpG 二核苷酸处（微卫星）。尽管不位于启动子区中的 CpG 岛的功能有待探索，但启动子区中的 CpG 岛由于其甲基化状态调控相关基因的转录和表达而显得特别重要。启动子区甲基化导致基因表达沉默。这一沉默是持久的且持续到有丝分裂和减数分裂过程中。尤其其于基因表达中的重要作用，DNA 甲基化对发育过程、印记过程（imprinting）和 X 染色体失活以及肿瘤发生、老化和对寄生 DNA 的抑制都具有影响。人们认为，许多分布广泛的肿瘤的肿瘤形成中涉及甲基化，例如肺肿瘤、乳房肿瘤和结肠癌和白血病。甲基化与蛋白质功能障碍（长 Q-T 综合症）或代谢疾病（新生儿暂时性糖尿病、2 型糖尿病）之间也存在联系。

可利用亚硫酸氢盐处理基因组 DNA 以分析所述 DNA 内甲基化胞嘧啶残基的位置。用亚硫酸氢盐处理核酸使胞嘧啶残基脱氨基成为尿嘧啶残基，而甲基化胞嘧啶保持不变。因此，例如，通过将未经亚硫酸氢盐处理的目标核酸的序列与在本文所提供的方法中经亚硫酸氢盐处理的核酸序列相比较，可推断出核酸中的甲基化程度以及胞嘧啶经甲基化的位置。所述经处理目标核酸与未经处理目标核酸之间的比较可通过各种方法实现。例如，未经处理的目标核酸可为先前已知的序列，其中由未经处理的目标核酸产生的质量峰已经过计算且未经实验测定。此外，可在不用亚硫酸氢盐处理的情况下通过进行断裂和质量峰分析根据实验确定未经处理的目标核酸序列质量峰。在另一方法中，相同经处理目标核酸的互补链可用于鉴定甲基化胞嘧啶。本方法是基于当使用亚硫酸氢盐将胞嘧啶转化成尿嘧啶时引起的碱基对错配。经亚硫酸氢盐处理后，甲基化双链目标核

酸含有一个或一个以上 G-U 错配。通过测定两条互补链的序列，G-U 错配的存在可用于指示在尿嘧啶位置处未经甲基化的胞嘧啶的存在，且 G-C 匹配碱基对的存在可用于指示甲基化胞嘧啶的存在。

通过使用具有甲基化特异性识别位点的限制性酶（例如 HpaII 和 MSP I）使得可能通过限制性核酸内切酶反应进行甲基化分析。基本原理为，某些酶被识别序列中的甲基化胞嘧啶阻断。实现此区分后，可使用如本文所提供的方法对所得片段进行随后的分析。

这些方法可在结合亚硫酸氢盐限制性分析（combined bisulfite restriction analysis, COBRA）中一起使用。用亚硫酸氢盐处理引起经扩增的 PCR 产物中 BstU I 识别位点的丧失，与未经处理的样本相比，会引起新的可检测片段出现在分析中。本文所提供的基于断裂的测序方法可与甲基化位点的特异性切割组合使用以提供有关目标核酸序列中甲基化模式的快速、可靠的信息。

5.生物体鉴定

本文所提供的方法可用于鉴定生物体或辨别不同于其他生物体的生物体。在一实施例中，可鉴定人类样本（例如，一个长区域或多个短区域）。基因的多态 STR 位点和其它多态区是作为对以下鉴定极其有用的标记的序列变异：人类鉴定、父亲和母亲身份鉴定、遗传图谱构建、移居和遗产纠纷、双生子卵性鉴定（zygosity testing in twins）、人类近亲结婚鉴定、人类培养细胞的质量控制（quality control）、人类遗体的鉴定和精液样本、血渍与法医学中其它物质的鉴定。所述位点也是经济动物育种与血统分析和经济植物育种中的有用标记。植物作物和动物中重要的经济学特质可通过使用多态 DNA 标记的连锁分析鉴定。有效且准确的基于断裂的核酸测序方法和本文所提供的鉴定一部分目标核酸的方法可用于测定所述位点的身份。目标核酸（例如，基因组 DNA）可从一个长目标核酸区和/或多个短目标核酸区获得。

在其它实施例中，所述方法可用于鉴定非人类生物体，例如非人类哺乳动物、鸟类、植物、真菌和细菌。

6.病原体鉴定和分型

本文也涵盖一种使用本文所提供的基于断裂和杂交的方法鉴定微生物品系的过程或方法。微生物选自多种生物体，包括（但不限于）细菌、真菌、原生动物、纤毛虫和病毒。所述微生物不限于特定属、种、品系或血清型。可通过相对于一个或一个以上参考序列测定目标微生物序列中的核酸序列和/或序列变异来鉴定微生物。参考序列可例如从来自相同或不同属、种、品系或血清型或来自宿主原核生物体或真核生物体的其他微生物获得。

细菌病原体的鉴定和分型对于感染性疾病的临床管理至关重要。微生物的准确身份不仅用于区分疾病状态与健康状态，而且也是确定抗生素或其它抗菌疗法是否适用于治疗和何种抗生素或其它抗菌疗法最适用于治疗的基础。传统的病原体分型方法已使用多种表型特征鉴定细菌，包括生长特征、颜色、细胞或菌落形态、抗生素易感性、染色、气味和对特定抗体的反应性。所有这些方法都需要对疑似病原体进行培养，这引起许多严重问题，包括高材料和劳动成本、工人暴露的危险、由不正确操作引起的假阳性和由活细胞数量少或许多病原体苛刻的培养要求引起的假阴性。此外，培养方法需要相对长的时间完成诊断，并且由于所述感染可能危及生命，因此通常在可获得结果之前开始抗菌疗法。

在许多情况下，病原体与组成正常菌群的生物体极其相似，且无法通过上文所提及的表型方法将其与无害菌株辨别开。在这些情况下，确定致病菌株的存在可能需要由本文所提供的基于断裂和杂交的方法所提供的较高分辨度。例如，PCR 扩增目标核酸序列，随后使用基质辅助激光解吸/电离飞行时间质谱进行断裂和杂交测序，接着如本文所提供筛选序列变异，由此将允许可靠地区别仅有一个核苷酸不同的序列，并且组合以 MALDI-TOF MS 的速度产生的序列信息的区别能力。类似地，通过将一个个以上质量峰与质量峰图形相比较鉴定一部分目标核酸的方法可用于检测序列变异。

例如，可使用本文所提供的断裂和杂交测序方法（包括以比较格式进行的断裂测序方法）实现使用更可靠的较长序列区（例如全长 16S rRNA 基因）进行细菌分型。出于说明的目的，可获得一种或一种以上已知细菌类型的序列并将其与未知细菌类型的序列相比较。

7. 分子育种和定向进化

在一实施例中，当目标核酸可表示经过修饰的核酸、病毒或生物体时，本文所公开的方法可用于测定目标核酸的序列或部分目标核酸。可使用所述方法将生物分子的特性或生物体或病毒的表型与生物分子、生物体或病毒的基因型相关联。例如，本文所公开的方法可用于鉴定核苷酸序列、质量峰或质量峰图形，此与目标核酸的特定特性、由目标核酸所编码的蛋白质或含有目标核酸的病毒或生物体有关。

例如，本文的方法可用于鉴定与目标核酸序列、质量峰或质量峰图形有关的特定蛋白质特性。在本实例中，可通过使用此项技术中已知的各种基因修饰方法修饰编码蛋白质的一种或一种以上基因重新设计一种或一种以上蛋白质，所述基因修饰方法包括 DNA 改组（美国专利第 6,117,679 号和第 6,537,746 号）、易错 PCR（error-prone PCR）（Caldwell, R. C. 和 Joyce, G. F. (1992) PCR Methods and Applications 2:28-33）、盒式突变法（cassette

mutagenesis) (Goldman, E R 和 Youvan D C (1992) *Bio/Technology* 10:1557-1561; Delagrave 等人 *Protein Engineering* 6:327-331 (1993)) 和密码子随机突变方法 (美国专利第 5,264,563 号和第 5,723,323 号)。可使用本文所公开的方法检查编码经重新设计而具有一个或一个以上特定特性的蛋白质的基因的序列或部分, 并将一个或一个以上质量峰鉴定为与所述经重新设计蛋白质的所述一个或一个以上特定特性有关。示范性蛋白质特性包括结合能力、催化能力、热稳定性、对蛋白酶的敏感性、表达水平、溶解性、膜插入或缔合、翻译后修饰、光学特性、电子转移特性、细胞器官靶向、被分泌的能力、对肝脏降解的易感性、免疫原性和跨越生物屏障转运 (包括从内脏吸收到血流中和跨越血脑屏障) 的能力。

鉴定与所述经重新设计蛋白质的一个或一个以上特性有关的一个或一个以上质量峰的方法包括分析编码一种或一种以上具有所述一个或一个以上特定特性的经重新设计的蛋白质的基因的质量峰图形; 和鉴定与那些特定特性有关的核苷酸序列或一个或一个以上质量峰或质量峰特征。可通过测定两个或两个以上编码具有特定特性的蛋白质的基因共有的序列或质量峰来测定与特定特性有关的序列或质量峰, 并且通常所述序列或质量峰为编码具有特定特性的蛋白质的基因的至少 50%、至少 70%、至少 85%、至少 90%或至少 90%所共有。即使仅一种所述蛋白质具有所述特定特性, 也可通过测定仅编码所述蛋白质的基因的序列或质量峰实现对于与特定特性有关的序列或质量峰的测定。

根据上述方法, 另一实施例包括鉴定编码具有一个或一个以上特定特性的蛋白质的基因的方法, 其中所述方法包括断裂基因; 使所述基因片段与一个或一个以上捕获寡核苷酸探针杂交, 其中与具有相同核苷酸序列的捕获寡核苷酸探针杂交的两个或两个以上基因具有不同的核苷酸序列; 和测量所述两个或两个以上基因片段的质量。在一实施例中, 当测量质量峰时, 可将一个或一个以上测量质量峰与一个或一个以上参考质量峰相比较, 其中所述一个或一个以上参考质量峰与经重新设计蛋白质的一个或一个以上特定特性有关。参考质量峰可使用例如上文所讨论的方法根据实验测定, 或可根据理论确定。在另一实施例中, 可构建所述目标核酸的核苷酸序列, 并且可将含有与一个或一个以上特定蛋白质特性有关的序列的目标核酸鉴定为编码具有所述特性的蛋白质的基因。

此外, 根据本实施例, 可进一步使用本文所述的方法分析与经重新设计蛋白质的一个或一个以上特定特性有关的一个或一个以上质量峰, 以提供有关编码所述经重新设计蛋白质的目标核酸基因的核苷酸序列信息。例如, 可通过将一个或一个以上质量峰特征与一个或一个以上参考质量峰特征相比较获得目标核酸序列信息, 其中所述一个或一个以上参考质量峰特征与目标核酸上一个或一个以上核苷酸位置处的特定核苷酸序列对

应。在另一实例中，可根据所测量的质量峰特征或通过使用本文所提供的序列构建方法测定一个或一个以上目标核酸片段的核苷酸序列。在另一实例中，可使用本文所提供的序列构建方法测定整个目标核酸序列或其部分。

在另一实例中，可通过使用各种方法修饰病毒基因组来对一种或一种以上病毒进行重新设计，所述方法包括病毒基因组改组（美国专利第 6,596,539 号）和病毒突变与选择方法。产生一种或一种以上具有一个或一个以上特定特性的病毒的经修饰病毒基因组可使用本文所公开的方法进行检查，并且可将一个或一个以上质量峰鉴定为与所述经修饰病毒的所述一个或一个以上特定特性有关。示范性病毒特性包括病毒感染性、复制、宿主范围、嗜性、基因功能、转录调控序列功能、在非允许细胞中复制的能力、宿主范围和/或细胞嗜性、病毒滴度（例如病毒性）、致病性或引起疾病的能力、感染性、包装能力、病毒微粒的物理/化学稳定性、细胞内稳定性、一种或一种以上病毒基因的表达、染色体整合、组织特异性和优先感染特定器官的能力、宿主（例如，人类）中的免疫原性或病毒或病毒蛋白、生物佐剂作用（例如共同表达病毒编码的人类细胞因子）和治疗作用（例如，诱导基本抗病毒宿主反应的能力-例如产生干扰素）。

鉴定一个或一个以上与所述经重新设计蛋白质的一个或一个以上特定特性有关的质量峰的方法包括分析一种或一种以上具有所述一个或一个以上特定特性的经重新设计病毒的病毒序列的质量峰图形，和鉴定与所述特定特性有关的核苷酸序列或一个或一个以上质量峰或质量峰特征。可通过测定两个或两个以上具有特定特性的病毒序列共有的序列或质量峰实现对于与特定特性有关的序列或质量峰的测定，并且通常所述序列或质量峰为具有特定特性的病毒序列的至少 50%、至少 70%、至少 85%、至少 90%或至少 90%所共有。即使仅一种所述病毒具有所述特定特性，也可通过测定仅所述病毒序列的序列或质量峰实现对于与特定特性有关的序列或质量峰的测定。

根据上述方法，另一实施例包括鉴定具有一个或一个以上特定特性的一种或一种以上病毒序列的方法，其中所述方法包括断裂病毒核酸；使所述病毒核酸片段与一个或一个以上捕获寡核苷酸探针杂交，其中与具有相同核苷酸序列的捕获寡核苷酸探针杂交的两个或两个以上病毒核酸片段具有不同的核苷酸序列；和测量所述两个或两个以上病毒核酸片段的质量。在一实施例中，当测量质量峰时，可将一个或一个以上测量质量峰与一个或一个以上参考质量峰相比较，其中所述一个或一个以上参考质量峰与经重新设计病毒的一个或一个以上特定特性有关。参考质量峰可使用例如上文所讨论的方法根据实验测定，或可根据理论确定。在另一实施例中，可构建所述病毒核酸的核苷酸序列，并且可将含有与一个或一个以上特定蛋白质特性有关的序列的病毒核酸鉴定为编码具有

所述特性的蛋白质的病毒序列。

此外,根据本实施例,可进一步使用本文所述的方法分析与经重新设计病毒的一个或一个以上特定特性有关的一个或一个以上质量峰,以提供有关所述经重新设计病毒的病毒核酸的核苷酸序列信息。例如,可通过将一个或一个以上质量峰特征与一个或一个以上参考质量峰特征相比较获得病毒核酸序列信息,其中所述一个或一个以上参考质量峰特征与病毒核酸上一个或一个以上核苷酸位置处的特定核苷酸序列对应。在另一实例中,可根据所测量的质量峰特征或通过使用本文所提供的序列构建方法测定一个或一个以上病毒核酸片段的核苷酸序列。在另一实例中,可使用本文所提供的序列构建方法测定整个病毒核酸序列或其部分。

本文进一步涵盖鉴定与生物体(例如经遗传修饰的生物体)的一个或一个以上特定特性有关的一个或一个以上质量峰的方法。示范性生物体包括植物,例如农作物,包括玉米、稻米、小麦、黑麦、燕麦、大麦、豌豆、菜豆、小扁豆、花生、豆薯、豇豆、黎豆、大豆、苜蓿、紫苜蓿、羽扇豆蓝、野豌豆、荷花、草木犀、柴藤、香豌豆、高粱、小米、向日葵和油菜;鸟类,包括火鸡和鸡;鱼;昆虫;线虫;非人类哺乳动物,包括家畜,例如猪、母牛、马和其他家畜。此项技术中已知修饰各种生物体的基因组的方法,包括DNA改组(美国专利第6,379,964号和第6,500,617号),也包括通过有性生殖进行的传统育种。所述生物体的特性可根据生物体而变化,但一般包括成活力、对疾病的抗性、生长速率、生殖能力、营养需求、水需求、温度敏感性和对环境压力的抗性。可使用上文关于病毒所述的方法鉴定一个或一个以上与生物体(例如经遗传修饰的生物体)的一个或一个以上特定特性有关的一个或一个以上质量峰。

8.作为标记的目标核酸片段

在其它实施例中,可将目标核酸片段用作标记或较大目标核酸的序列或部分的指示器。所述实施例不需要测定目标核酸的完整序列,但可包括测定部分目标核酸的序列,或简单测定目标核酸片段的质量峰图形。这些实施例也不需要目标核酸片段重叠,因此,对于这些实施例而言,目标核酸片段可为重叠片段或非重叠片段。所述方法可包括例如指纹法和与指纹法有关的方法,和包括使用非重叠DNA片段作为目标核酸的序列或部分的指示器的其它方法。在本文所公开的方法中可使用包括扩增步骤的指纹法,例如扩增核糖体DNA限制性分析(ARDRA)、随机扩增多态DNA分析(RAPD)和扩增片段长度多态性(AFLP)。

在一实施例中,可形成目标核酸的片段,使其与捕获核酸的阵列杂交,并测定所述片段的质量以产生以一个、两个、三个或三个以上特征为特征的质量峰图形,所述特征

例如与目标核酸杂交的捕获寡核苷酸探针的位置、质量和质量峰的信噪比。所述质量峰图形可用作目标核酸的序列或部分的指示器。

在一实施例中，经特别设计的引物和扩增方法能够以仅扩增一分组目标核酸片段的方式控制扩增，并且随后这一分组片段可与捕获寡核苷酸探针阵列杂交并分析质量。本实施例可用作目标核酸：基因、染色体片段、酵母人工染色体（YAC）、细菌人工染色体（BAC）、整个染色体、整个基因组或任何适当的其它核酸分子；或来自例如物种或品系集群的一种或一种以上不同生物体的多个基因、染色体片段、YAC、BAC、整个染色体和整个基因组。此项技术中已知扩增核酸片段分组的方法，例如扩增片段长度多态性（AFLP）方法（例如，参看美国专利第 6,045,994 号）。

根据本实施例，使用一种或一种以上限制性酶产生目标核酸的片段。通常，使用两种切割不同核苷酸序列的限制性酶。例如，可使用稀切酶（rare cutter）（一种识别例如 6 个核苷酸的长核苷酸序列且因此在核酸上较少位点处进行切割的限制性酶）和密切酶（common cutter）（一种识别例如 4 个核苷酸的短核苷酸序列且因此在核酸上更多位点处进行切割的限制性酶）。在其它实例中，可使用两种稀切酶或两种密切酶。可根据目标核酸的长度和所需目标核酸片段的数量和长度选择限制性酶数量和所述酶的特异性。

在不考虑限制性片段末端的核苷酸序列是否已知的情况下，可进行限制性片段的 PCR 扩增。这可通过首先将已知序列的合成寡核苷酸（接头）与限制性片段的两端接合，由此向每一限制性片段提供两种可与 PCR 扩增中所使用的引物互补的共用标签来达成。

通常，限制性酶产生两条链的末端核苷酸都碱基配对的平端，或两条链的一条伸出以提供短的单链区的“粘”端。在具有平端的限制性片段的情况下，将接头与所述平端的一条链接合。在具有粘端的限制性片段的情况下，所述接头具有与限制性片段的单链区互补的区域。所述接头首先以一定方式与限制性片段的单链区互补部分杂交，所述方式为：使接头端与所述限制性片段的一条链的末端相邻；随后使所述接头与相邻限制性片段端接合。

因此，对于每种类型的限制性切割，可设计不同的接头以允许所述接头的一端与特别对应的限制性片段接合。通常，所述接头大约为 10 至 30 个核苷酸长，通常为 12 至 22 个核苷酸长。使用接合酶，可将接头与限制性片段的混合物接合。当使用相对于限制性片段大摩尔过量的接头时，几乎所有限制性片段与接头在两端接合。用本方法制备的限制性片段称为“经标记限制性片段（tagged restriction fragment）”。

每个经标记的限制性片段都具有以下通用结构：可变 DNA 序列在所述经标记限制性片段的每一端侧接恒定 DNA 序列。恒定 DNA 序列含有限制性核酸内切酶的部分或全

部识别序列，也含有与经标记限制性片段的每一端连接的接头的序列。限制性片段的可变序列位于恒定 DNA 序列之间，因此包括限制性片段中不含有限制性核酸内切酶识别序列的部分。可变序列可为已知或未知，通常在限制性片段之间变化。因此，与恒定 DNA 序列侧接的核苷酸序列可为不同序列的较大混合物。

在一实施例中，所述接头可以是 PCR 引物的严格互补体。例如，限制性片段可在其两端具有相同的接头，并且一个 PCR 引物可在不与限制性片段序列的任何部分杂交的情况下与接头杂交，并且可用于扩增限制性片段。在另一实例中，例如使用两种不同的限制性酶切割 DNA，可将两种不同的接头与所述限制性片段的末端接合。在这一情况下，可使用一种或两种不同的 PCR 引物扩增所述限制性片段。在本实施例中，使用 PCR 引物扩增所有经标记限制性片段，而不考虑所述限制性片段的可变序列。

不管在上述步骤中是否已扩增经标记限制性片段，随后都使用含有第一核苷酸序列部分和第二序列部分的可变序列特异性 PCR 引物扩增所述经标记的限制性片段。对所述第一序列部分进行设计以使其与所述经标记限制性片段的恒定 DNA 序列完全碱基配对。第二序列部分可含有任何所选择的序列或随机序列，且长度在 1 至约 10 个核苷酸的范围内。第二序列部分仅与一分组经标记限制性片段杂交，仅引起所述已杂交的经标记限制性片段分组扩增。在一实施例中，可使用在第二序列部分中具有不同序列的若干种不同的序列特异性 PCR 引物，以便扩增较大分组的经标记限制性片段。

将第二序列部分加入到所述序列特异性引物的 3'端可确定在 PCR 步骤中哪些经标记限制性片段被扩增：序列特异性引物将仅引起其中所述序列特异性 PCR 引物的第二部分可与经标记限制性片段碱基配对的那些经标记限制性片段上的 DNA 合成。

序列特异性扩增一分组经标记限制性片段后，需要时，可根据本文所公开的方法进一步断裂所述限制性片段（也可称其为目标核酸片段）。例如，可使目标核酸片段（限制性片段）经历另外的序列特异性切割、碱基特异性切割或非特异性切割。随后使目标核酸片段与捕获寡核苷酸探针的阵列杂交。杂交后，需要时，可根据本文所公开的方法进一步断裂目标核酸片段。例如，可使目标核酸片段经历碱基特异性切割。可在杂交前或杂交后进行切割（例如）以达成与一个或一个以上捕获寡核苷酸探针杂交的目标核酸片段的所需复杂程度，或达成所需的目标核酸片段长度以（例如）用于使用质谱法以所需准确度进行质量测定。

9.检测指示感染的病毒或细菌核酸序列的存在

可通过相对于一种或一种以上参考序列鉴定病毒或细菌核酸序列中所存在的序列变异，而使用本文所提供的方法确定指示感染的病毒或细菌核酸序列的存在。参考序列

可包括（但不限于）从相关未感染生物体获得的序列或从宿主生物体获得的序列。

病毒、细菌、真菌和其他感染生物体含有截然不同的核酸序列，包括多态性，其不同于宿主细胞中所含的序列。目标 DNA 序列可以是外来基因序列的一部分，例如入侵微生物的基因组，包括例如细菌和其噬菌体、病毒、真菌和原生动植物。本文所提供的方法尤其适用于辨别微生物的不同变异体或品系，以便（例如）选择适当的治疗干预（therapeutic intervention）。感染人类和动物且可通过公开方法检测的致病性病毒的实例包括（但不限于）反转录病毒科（*Retroviridae*）（例如，人类免疫缺陷病毒，例如 HIV-1（也称为 HTLV-III、LAV 或 HTLV-III/LAV；Rattier 等人, *Nature* 113:227-284 (1985); Wain Hobson 等人, *Cell* 40:9-17 (1985))、HIV-2（Guyader 等人, *Nature*, 328:662-669 (1987); 欧洲专利公开案第 0 269 520 号; Chakrabarti 等人, *Nature* 328:543-547 (1987); 欧洲专利公开案第 0 655 501 号); 和其他分离菌，例如 HIV-LP（国际公开案第 WO 94/00562 号)); 小 RNA 病毒科（*Picomaviridae*）（例如，脊髓灰质炎病毒（poliovirus）、甲型肝炎病毒（Gust 等人, *Intervirolgy*, 20:1-7 (1983))、肠道病毒、人类柯萨奇病毒（human coxsackie virus）、鼻病毒、艾柯病毒（echovirus)); 杯状病毒科（*Calciviridae*）（例如，引起肠胃炎的菌株); 披膜病毒科（*Togaviridae*）（例如，马脑炎病毒（equine encephalitis virus）、风疹病毒（rubella virus)); 黄病毒科（*Flaviridae*）（例如，登革病毒（dengue virus）、脑炎病毒（encephalitis virus）、黄热病毒（yellow fever virus)); 冠状病毒科（*Coronaviridae*）（例如，冠状病毒（coronaviruse)); 弹状病毒科（*Rhabdoviridae*）（例如，水泡性口炎病毒（vesicular stomatitis virus）、狂犬病病毒（rabies virus)); 丝状病毒科（*Filoviridae*）（例如，伊波拉病毒（ebola virus)); 副粘液病毒科（*Paramyxoviridae*）（例如，副流感病毒（parainfluenza viruse）、腮腺炎病毒（mumps virus）、麻疹病毒（measles virus）、呼吸道合胞病毒（respiratory syncytial virus)); 粘液病毒科（*Orthomyxoviridae*）（例如，流感病毒（influenza virus)); 布尼亚病毒科（*Bungaviridae*）（例如，汉滩病毒（Hantaan virus）、野菰病毒（bunga virus）、白蛉热病毒（phlebovirus）和内罗病毒（Nairo virus)); 沙状病毒科（*Arenaviridae*）（出血热病毒（hemorrhagic fever virus)); 呼肠孤病毒科（*Reoviridae*）（例如，呼吸道肠道孤儿病毒（reovirus）、环状病毒属（orbivirus）和轮状病毒（rotavirus)); 双核糖核酸病毒科（*Birnaviridae*); 肝病毒科（*Hepadnaviridae*）（乙型肝炎病毒); 细小病毒科（*Parvoviridae*）（细小病毒属（parvovirus)); 乳头多瘤空泡病毒科（*Papovaviridae*); 肝病毒科（*Hepadnaviridae*）（乙型肝炎病毒); 细小病毒科（*Parvoviridae*）（大多数腺病毒); 乳头多瘤空泡病毒科（*Papovaviridae*）（乳头瘤病毒（papilloma virus）、多瘤病毒（polyoma virus)); 腺病毒科（*Adenoviridae*）（大多数腺

病毒); 疱疹病毒 (*Herpesviridae*) (1 型单纯疱疹病毒 (*herpes simplex virus type 1*) (HSV-1) 和 HSV-2、水痘带状疱疹病毒 (*varicella zoster virus*)、巨细胞病毒 (*cytomegalovirus*)、疱疹病毒 (*herpes virus*)); 痘病毒科 (*Poxviridae*) (天花病毒 (*variola virus*)、牛痘病毒 (*vaccinia virus*)、痘病毒 (*pox virus*)); 虹彩病毒科 (*Iridoviridae*) (例如, 非洲猪瘟病毒 (*African swine fever virus*)); 和未分类病毒 (例如, 海绵状脑病 (*Spongiform encephalopathies*) 的病原体、丁型肝炎病原体 (认为是乙型肝炎的缺陷卫星)、非甲型 (non-A)、非乙型肝炎病原体 (第 1 类=内部传染; 第 2 类=不经肠传染, 即, 丙型肝炎); 诺瓦克和相关病毒 (*Norwalk and related virus*) 和星状病毒 (*astrovirus*)。

感染性细菌的实例包括 (但不限于) 幽门螺杆菌 (*Helicobacter pyloris*)、伯氏疏螺旋体 (*Borelia burgdorferi*)、肺炎军团菌 (*Legionellapneumophilia*)、分支杆菌 (*Mycobacteria sp.*) (例如, 结核分支杆菌 (*M. tuberculosis*)、鸟分支杆菌 (*M. avium*)、腿内杆菌 (*M. intracellulare*)、坎沙西分枝杆菌 (*M. kansasii*)、戈氏分枝杆菌 (*M. gordonae*))、金黄色葡萄球菌 (*Staphylococcus aureus*)、奈瑟淋球菌 (*Neisseria gonorrhoeae*)、脑膜炎奈瑟菌 (*Neisseria meningitidis*)、单核细胞增多性李斯特菌 (*Listeria monocytogenes*)、化脓性链球菌 (*Streptococcus pyogenes*) (A 群链球菌 (Group A Streptococcus))、无乳链球菌 (*Streptococcus agalactiae*) (B 群链球菌)、链球菌属 (*Streptococcus sp.*) (草绿色链球菌群 (*viridans group*))、粪链球菌 (*Streptococcus faecalis*)、牛链球菌 (*Streptococcus bovis*)、链球菌属 (*Streptococcus sp.*) (厌氧型菌)、肺炎链球菌 (*Streptococcus pneumoniae*)、致病性弯曲菌属 (*pathogenic Campylobacter sp.*)、肠球菌属 (*Enterococcus sp.*)、流感嗜血杆菌 (*Haemophilus influenzae*)、炭疽杆菌 (*Bacillus anthracis*)、白喉杆菌 (*Corynebacterium diphtheriae*)、棒杆菌属 (*Corynebacterium sp.*)、红斑丹毒丝状菌 (*Erysipelothrix rhusiopathiae*)、产气荚膜梭菌 (*Clostridium perfringens*)、破伤风梭菌 (*Clostridium tetani*)、产气大肠杆菌 (*Enterobacter aerogenes*)、克雷伯氏杆菌 (*Klebsiella pneumoniae*)、巴斯得菌 (*Pasturella multocida*)、拟杆菌属 (*Bacteroides sp.*)、具核梭杆菌 (*Fusobacterium nucleatum*)、念珠状链杆菌 (*Streptobacillus moniliformis*)、梅毒密螺旋体 (*Treponema pallidum*)、极细密螺旋体 (*Treponema pertenue*)、钩端螺旋体 (*Leptospira*) 和衣氏放线菌 (*Actinomyces israeli*)。

感染性真菌的实例包括 (但不限于) 新型隐球菌 (*Cryptococcus neoformans*)、荚膜组织胞浆菌 (*Histoplasma capsulatum*)、粗球孢子菌 (*Coccidioides immitis*)、皮炎芽生菌 (*Blastomyces dermatitidis*)、沙眼衣原体 (*Chlamydia trachomatis*)、白色念珠菌 (*Candida albicans*)。其他感染性生物体包括原生生物, 例如恶性疟原虫 (*Plasmodium falciparum*)

和刚地弓形虫 (*Toxoplasma gondii*)。

10. 抗生素谱 (Antibiotic Profiling)

如本文所提供的目标核酸片段的质量分析可改进药物抗性 (包括抗生素抗性) 中所涉及的核苷酸改变检测的速度和准确性。对异烟肼 (isoniazid)、利福平 (rifampin)、链霉素 (streptomycin)、氟喹诺酮 (fluoroquinolones) 和乙硫异酰胺 (ethionamide) 的抗性中所涉及的遗传位点已得到鉴定[Heym 等人, *Lancet* 344:293 (1994)和 Morris 等人, *J. Infect. Dis.* 171:954 (1995)]。常规上将异烟肼 (inh) 和利福平 (rif) 以及吡嗪酰胺 (pyrazinamide) 和乙胺丁醇 (ethambutol) 或链霉素的组合用作对抗结核分枝杆菌 (*M. tuberculosis*) 确诊病例的疾病发作的首要选择[Banerjee 等人, *Science* 263:227 (1994)]。所述抗性菌株不断增加的发病率使得有必要研发出快速测定法, 以检测所述菌株并由此降低继续无效且可能有害的治疗的费用和公众健康损害。对药物抗性中所涉及的一些遗传位点的鉴定促进了采用快速筛选引起药物抗性的核苷酸改变的突变检测技术。

11. 鉴定疾病标记

本文提供快速且准确地鉴定作为疾病遗传标记的序列变异的方法, 可用于诊断或确定疾病的预后。以遗传标记为特征的疾病可包括 (但不限于) 动脉硬化症、肥胖症、糖尿病、自体免疫病症和癌症。无论是遗传或由身体对环境压力 (例如病毒和毒素) 起反应所引起, 所有生物体的疾病都具有遗传成分。正在进行的基因组研究的最终目标是使用这一信息研发出鉴定、治疗和可能治愈这些疾病的方式。首先是筛选疾病组织和鉴定个体样本层面上基因组的改变。这些“疾病”标记的鉴定取决于检测基因组标记的改变从而鉴定错误基因或多态性的能力。基因组标记 (包括单核苷酸多态性 (SNP)、微卫星和其它非编码基因组区、串联重复、内含子和外显子的所有遗传位点) 都可用于鉴定所有生物体, 包括人类。这些标记提供一种不仅能够鉴定种群而且也允许根据种群对疾病、药物治疗、对环境作用和其它因素的反应将其分层的方式。

12. 单倍体分型

本文所提供的方法可用于检测单倍型。在任何二倍体细胞中, 在任何基因或其它染色体区段处都存在两种含有至少一种可辨别的变异的单倍型。在许多充分研究的遗传系统中, 单倍型与单核苷酸变异相比更强烈地与表型相关。因此, 测定单倍型对于了解各种表型的遗传基础是颇有价值的, 所述表型包括疾病倾向或易感性、对治疗干预的反应和医药、动物饲养和农业中所关注的其它表型。

如本文所提供的单倍体分型过程允许从个体的两个同源染色体中的一个中选择一部分序列并对所述序列部分上已连锁的 SNP 进行基因分型。直接解析单倍型可得到增加

的信息含量，从而改进对任何相关疾病基因的诊断或鉴定与这些疾病有关的连锁。

13.DNA 重复序列

本文所提供的基于断裂的方法允许快速检测 DNA 重复序列中的序列变异。各种 DNA 重复序列都可能与疾病有关 (Thangavelu 等人, *Prenat. Diagn.* 18:922-25 (1998); Bennett 等人, *J. Autoimmun.* 9:415-21 (1996))。DNA 重复序列包括卫星、小卫星和微卫星。卫星的单位尺寸可在 2 碱基单元重复至约 100 碱基单元重复的范围内，或更多，通常所述重复单元在约 1000 个重复序列至约 10,000 个重复序列的范围内。小卫星（也称为短串联重复序列（或 STR））的单位尺寸可在 3 碱基单元重复至约 100 碱基单元重复的范围内，通常所述重复单元在约 2 个重复序列至约 100 个重复序列的范围内，或更多；从而使小卫星的最小长度通常为约 500 个碱基。微卫星的单位尺寸可在 1 碱基单元重复至约 7 碱基单元重复的范围内，通常所述重复单元在约 5 个重复序列至约 100 个重复序列的范围内。微卫星可位于染色体基因附近并且可在基因表达中起作用。卫星、小卫星或微卫星中变异的检测可用作变异体或疾病倾向的标记。

微卫星（有时称为可变数量的串联重复序列或 VNTR）是一个至七个或更多个碱基的短串联重复核苷酸单元，其中最突出的是二核苷酸重复单元、三核苷酸重复单元和四核苷酸重复单元。微卫星在基因组 DNA 中每 100,000 bp 处存在 (J. L. Weber 和 P. E. Can, *Am. J. Hum. Genet.* 44:388 (1989); J. Weissenbach 等人, *Nature* 359:794 (1992))。例如 CA 二核苷酸重复单元组成约 0.5% 人类线粒体外基因组；CT 和 AG 重复单元一起组成约 0.2%。CG 重复单元极为罕见，这很可能是因为 CpG 岛的调控功能。微卫星的长度具有高度多态性，且广泛分布于整个基因组中，其中主要是在非编码区中含量丰富，而其在所述基因组中的功能尚不明确。

微卫星对于法医应用尤为重要，因为种群保持所述种群的各种微卫星特征且与其它种群截然不同，不会混种。

微卫星内的许多改变会是沉默的，但一些可导致基因产物或表达水平显著变化。例如，在一些肿瘤中基因编码区中所见的三核苷酸重复单元受到影响 (C. T. Caskey 等人, *Science* 256:784 (1992))，并且微卫星的变化可引起遗传不稳定，进而导致癌症倾向 (P. J. McKinnen, *Hum. Genet.* 1(75): 197 (1987); J. German 等人, *Clin. Genet.* 35:57 (1989))。

也可使用本文所提供的方法相对于（例如）不含 STR 区的基因组的参考基因组序列鉴定基因组一些目标序列中的微卫星或短串联重复序列 (STR)。STR 区为与任何疾病或病况无关的多态区。人类基因组的许多位点中都含有多态性短串联重复序列 (STR) 区。STR 位点含有 3 至 100 个碱基长的短、重复序列元件。据估计，存在 200,000 个预期三

聚和四聚 STR, 在人类基因组中以每 15 kb 一次频繁出现 (例如, 参看国际 PCT 申请案第 WO 9213969 A1 号, Edwards 等人, *Nucl. Acids Res.* 19:4791 (1991); Beckmann 等人 *Genomics* 12:627-631 (1992))。这些位点有近一半具多态性, 提供遗传标记的丰富来源。特定位点处重复单元数量的变化是造成可变核苷酸串联重复序列 (VNTR) 位点 (Nakamura 等人 *Science* 235:1616-1622 (1987)) 以及含有较长重复单元的小卫星位点 (Jeffreys 等人 *Nature* 314:67-73 (1985)) 和微卫星或二核苷酸重复序列位点 (Luty 等人 *Nucleic Acids Res.* 19:4308 (1991); Lift 等人 *Nucleic Acids Res.* 18:4301 (1990); Litt 等人 *Nucleic Acids Res.* 18:5921 (1990); Luty 等人 *Am. J. Hum. Genet.* 46:776-783 (1990); Tautz *Nucl. Acids Res.* 17:6463-6471 (1989); Weber 等人 *Am. J. Hum. Genet.* 44:388-396 (1989); Beckmann 等人 *Genomics* 12:627-631 (1992)) 处可观察到的多态性的原因。

STR 位点的实例包括 (但不限于) 人类 CD4 位点中的五核苷酸重复序列 (Edwards 等人, *Nucl. Acids Res.* 12:4791 (1991)); 人类细胞色素 P-450 芳香化酶基因中的四核苷酸重复序列 (CYP19; Polymeropoulos 等人, *Nucl. Acids Res.* 19:195 (1991)); 人类凝血因子 XIII A 亚基基因中的四核苷酸重复序列 (F13A1; Polymeropoulos 等人, *Nucl. Acids Res.* 19:4306 (1991)); F13B 位点中的四核苷酸重复序列 (Nishimura 等人, *Nucl. Acids Res.* 20:1167 (1992)); 人类 c-les/fps, 即原癌基因中的四核苷酸重复序列 (FES; Polymeropoulos 等人, *Nucl. Acids Res.* 19:4018 (1991)); LFL 基因中的四核苷酸重复序列 (Zuliani 等人, *Nucl. Acids Res.* 18:4958 (1990)); 人类胰腺磷脂酶 A-2 基因处三核苷酸重复多态性 (PLA2; Polymeropoulos 等人, *Nucl. Acids Res.* 18:7468 (1990)); VWF 基因中四核苷酸重复多态性 (Ploos 等人, *Nucl. Acids Res.* 18:4957 (1990)); 和人类甲状腺过氧化物酶 (hTPO) 位点中的四核苷酸重复序列 (Anker 等人, *Hum. Mol. Genet.* 1:137 (1992))。

14. 检测等位基因变异

本文所提供的方法允许高通量、快速和准确地检测等位基因变异体。有关等位基因变异的研究不仅涉及在复合物背景中检测特定序列, 而且还涉及区别具有极少或单核苷酸差异的序列。一种通过 PCR 检测等位基因特异性变异体的方法是基于下述事实: 当模板链与引物 3' 端之间存在错配时, Taq 聚合酶难以合成 DNA 链。等位基因特异性变异体可通过使用仅与可能的等位基因中的一个完全匹配的引物进行检测; 与其它等位基因错配阻止引物延长, 因此阻止所述序列扩增。这一方法的主要限制在于, 错配的碱基组成影响阻止跨越所述错配延长的能力, 并且某些错配不会阻止延长或仅具有极小作用 (Kwok 等人, *Nucl. Acids Res.* 18:999 [1990])。本文所提供的基于断裂和杂交的方法克服了所述引物延长方法的局限。

15.测定等位基因频率

本文所述的方法可用于鉴定一种或一种以上遗传标记，所述遗传标记的频率在种群内随年龄、种族、性别或一些其它标准而改变。例如，此项技术中已知年龄依赖型 ApoE 基因型分布（参看，Schächter 等人 *Nature Genetics* 6:29-32 (1994)）。已知在一定程度上与疾病有关的多态性的频率可用于检测或监测疾病状态的进程。例如，导致丝氨酸取代氨基酸密码子 291 位处的天冬酰胺的脂蛋白脂酶基因的 N291S 多态性（N291S），引起高密度脂蛋白胆固醇（HDL-C）含量降低，这与男性患动脉硬化、尤其是心肌梗塞的高风险有关（参看，Reymer 等人 *Nature Genetics* 10:28-34 (1995)）。此外，测定等位基因频率的改变可允许鉴定先前未知的多态性并最终鉴定疾病发生和发展中所涉及的基因或途径。

16.表观遗传学

本文所提供的方法可用于相对于参考核酸研究目标核酸或蛋白质中的变异。所述方法不是基于序列，例如作为所述核酸中天然存在的单体单元的碱基的身份。例如，本文所提供的方法中所使用的特异性切割试剂可识别目标分子与参考分子之间不依赖于序列的特征（例如甲基化模式、经修饰碱基的存在或高序结构（higher order structure）差异）的差异以产生在不依赖于序列的位点处被切割的片段。表观遗传学是基于基因表达的差异而非基因序列的差异对遗传信息进行研究。基因外改变是指有丝分裂和/或减数分裂方面基因功能可遗传的改变，或无法通过核酸序列的改变解释的高序核酸结构的改变。经历基因外变异或改变的特征的实例包括（但不限于）动物体内的 DNA 甲基化模式、组蛋白修饰和 Polycomb-trithorax 组群（Pc-G/trx）蛋白质复合物（例如，参看 Bird, A., *Genes Dev.*, 16:6-21 (2002)）。

基因外改变通常（但非一定）导致通常（但非一定）不可遗传的基因表达的改变。例如，如上文所讨论，甲基化模式的改变是癌症和其它疾病发生和发展中的早期事件。在许多癌症中，某些基因由于异常甲基化而不适当地切断或开启。甲基化模式压制或活化转录的能力可得到遗传。类似甲基化的 Pc-G/trx 蛋白质复合物可以可遗传的方式压制转录。使 Pc-G/trx 多蛋白组装体靶向所述基因组的特定区域，在所述区域中，无论基因具活性或不具活性，所述组装体都有效冷冻基因的胚胎基因表达状态，并通过发育稳定传播所述状态。Pc-G/trx 蛋白质组靶向基因组并与基因组结合的能力仅影响所述基因组中所含基因的表达水平，而不影响基因产物的特性。本文所提供的方法可与特异性切割试剂一起使用以基于不依赖于序列的改变（例如基因外改变）相对于参考序列鉴定目标序列中的变异。

实例 1

为重建潜在 DNA 序列 (underlying DNA sequence), 可使用本实例所述和举例说明的方法利用通过杂交后测序进行的核苷酸序列分析技术以及通过质谱法进行核苷酸序列分析技术。具体来说, 可将实验数据转换成 de Bruijn 图的子图, 参看 Pevzner, J. *Biomol. Struct. Dyn.*, 7:63-73 (1989)。随后, 可搜寻这一图中的欧拉路径, 其中循环和凸起需预先断裂, 参看 Pevzner 等人, *Proc. Natl. Acad. Sci. USA* 98:9748-9753 (2001)。

例如, 将 ACATGAGCTTACAAC (SEQ ID NO: 1) 作为研究中的 DNA 序列。切割反应非特异性地将这一 DNA (或 RNA) 分子切割成 5-7 nt 的片段。最后, 使所得片段与含有 16 个具有 4 个简并碱基的位置的杂交芯片结合, 每一简并碱基都与嘌呤 (字母 R、A 或 G) 或嘧啶 (字母 Y、C 或 T) 结合。在这一简并字母表中, 研究中的序列变为 RYRYRRRYYYRYRYY。随后, 在芯片上出现以下结合模式:

简并模式	与杂交斑点连接的片段
RRRR	(无片段)
RRRY	CATGAGC, ATGAGC, ATGAGCT, TGAGC, TGAGCT, GAGCTT, GAGCT, GAGCTT, GAGCTTA
RRYR	(无片段)
RYYR	ATGAGCT, TGAGCT, TGAGCTT, GAGCT, GAGCTT, GAGCTTA, AGCTT, AGCTTA, AGCTTAC
RYRR	ACATGA, ACATGAG, CATGA, CATGAG, CATGAGC, ATGAG, ATGAGC, ATGAGCT, CTTACAA, TTACAA, TTACAAC
RYRY	ACATG, ACATGA, ACATGAG
RYRR	(无片段)
RYYY	TGAGCTT, GAGCTT, GAGCTTA, AGCTT, AGCTTA, AGCTTAC, GCTTA, GCTTAC, GCTTACA
YRRR	ACATGAG, CATGAG, CATGAGC, ATGAG, ATGAGC, ATGAGCT, TGAGC, TGAGCT, TGAGCTT
YRRY	TTACAAC
YRYR	ACATG, ACATGA, ACATGAG, CATGA, CATGAG, CATGAGC, GCTTACA, CTTACA, CTTACAA, TTACA, TTACAA, TTACAAC
YRYR	(无片段)
YYRR	(无片段)
YYRY	AGCTTAC, GCTTAC, GCTTACA, CTTAC, CTTACA, CTTACAA, TTACA, TTACAA, TTACAAC
YYYY	GAGCTTA, AGCTTA, AGCTTAC, GCTTA, GCTTAC, GCTTACA, CTTAC, CTTACA, CTTACAA
YYYY	(无片段)

使用质谱分析, 可测定片段的组成, 例如, 参看 Böcker, *Lect. Notes Comp. Sci.* 2812:476-487 (2003)。随后, 测量与以下复合体对应的质谱。

简并模式	杂交斑点上检测到的复合体
RRRR	(无峰)
RRRY	$A_2C_2G_2T_1, A_2C_1G_2T_1, A_2C_1G_2T_2, A_1C_1G_2T_1, A_1C_1G_2T_2, A_1C_1G_2T_3,$ $A_1C_1G_2T_1, A_1C_1G_2T_2, A_2C_1G_2T_1$
RRYR	(无峰)
RRYY	$A_2C_1G_2T_2, A_1C_1G_2T_2, A_1C_1G_2T_3, A_1C_1G_2T_1, A_1C_1G_2T_2, A_2C_1G_2T_2,$ $A_1C_1G_1T_2, A_2C_1G_1T_2, A_2C_2G_1T_2$
RYRR	$A_3C_1G_1T_1, A_3C_1G_2T_1, A_2C_1G_1T_1, A_2C_1G_2T_1$ (两次), $A_2C_2G_2T_1, A_2G_2T_1,$ $A_2C_1G_2T_1, A_2C_1G_2T_2, A_3C_2T_2$ (两次), $A_3C_1T_2$
RYRY	$A_2C_1G_1T_1, A_3C_1G_1T_1, A_3C_1G_2T_1$
RYRR	(无峰)
RYYY	$A_1C_1G_2T_3, A_1C_1G_2T_2, A_2C_1G_2T_2, A_1C_1G_1T_2$ (两次), $A_2C_1G_1T_2, A_2C_2G_1T_2$ (两次), $A_1C_2G_1T_2$
YRRR	$A_3C_1G_2T_1, A_2C_1G_2T_1$ (两次), $A_2C_2G_2T_1, A_2G_2T_1, A_2C_1G_2T_2, A_1C_1G_2T_1,$ $A_1C_1G_2T_2, A_1C_1G_2T_3$
YRRY	$A_3C_2T_2$
YRYR	$A_2C_1G_1T_1$ (两次), $A_3C_1G_1T_1, A_3C_1G_2T_1, A_2C_1G_2T_1, A_2C_2G_2T_1, A_2C_2G_1T_2,$ $A_2C_2T_2, A_3C_2T_2$ (两次), $A_2C_1T_2, A_3C_1T_2$
YRYR	(无峰)
YYRR	(无峰)
YYRY	$A_2C_2G_1T_2$ (两次), $A_1C_1G_1T_2, A_1C_2T_2, A_2C_2T_2, A_3C_2T_2$ (两次), $A_2C_1T_2,$ $A_3C_1T_2$
YYR	$A_2C_1G_2T_2, A_2C_1G_1T_2, A_2C_2G_1T_2$ (两次), $A_1C_1G_1T_2, A_1C_2G_1T_2, A_1C_2T_2,$ $A_2C_2T_2, A_3C_2T_2$
YYYY	(无峰)

这一信息如下用于分支限界研究中：假定 ACATGAG 是正确序列的已知前缀 (prefix)。下一个碱基的身份可随机指定，随后将其与一个或一个质量峰比较。如果指定所述下一个碱基为 A，那么预测在若干个不同质量峰中以下片段和复合体的峰：

片段：	复合体：	与以下对应的图谱：
CATGAGA	$A_3C_1G_2T_1$	YRYR, RYRR, YRRR, RRRR
ATGAGA	$A_3G_2T_1$	RYRR, YRRR, RRRR
TGAGA	$A_2G_2T_1$	YRRR, RRRR

质谱与上述假设矛盾：如果 ACATGAGA 是所述位点处的正确核苷酸，那么对应于杂交位 RRRR 的质谱将含有至少三个峰。但在这一图谱上并未检测到一个峰。此结论是基于在 4 个质谱中观察到或未观察到 9 个峰，因而极为稳健。类似推理表明 G 与 T 都不能与前缀 ACATGAG 连接。

相比之下，将碱基 C 附加于前缀 ACATGAG 中将在若干不同质谱中产生以下片段和

复合体:

片段:	复合体:	与以下对应的质谱
CATGAGC	$A_2C_2G_2T_1$	YRYR, RYRR, YRRR, RRRY
ATGAGC	$A_2C_1G_2T_1$	RYRR, YRRR, RRRY
TGAGC	$A_1C_1G_2T_1$	YRRR, RRRY

由于在 4 个截然不同的质谱中观察到全部 9 个峰, 因此 C 为所连接的正确字符。也可通过上述方法分析更复杂的切割模式, 并且在这些复杂的设置中所述方法也具有稳健性。

由于所属领域技术人员对于修改将显而易见, 故希望本发明仅受随附权利要求书的保护范围所限制。

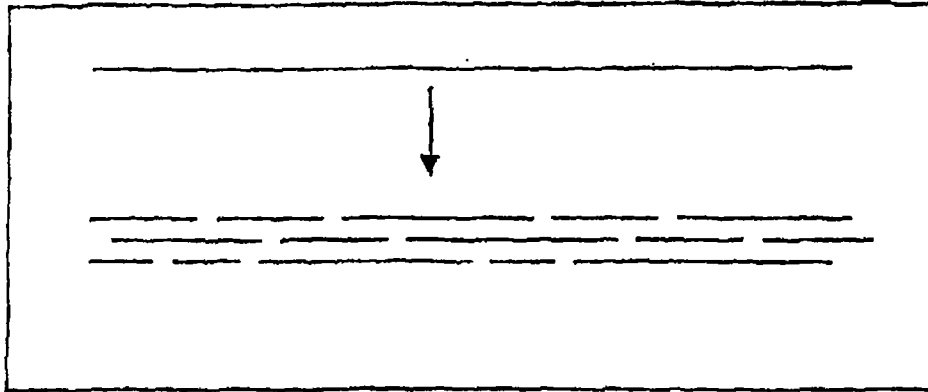


图1

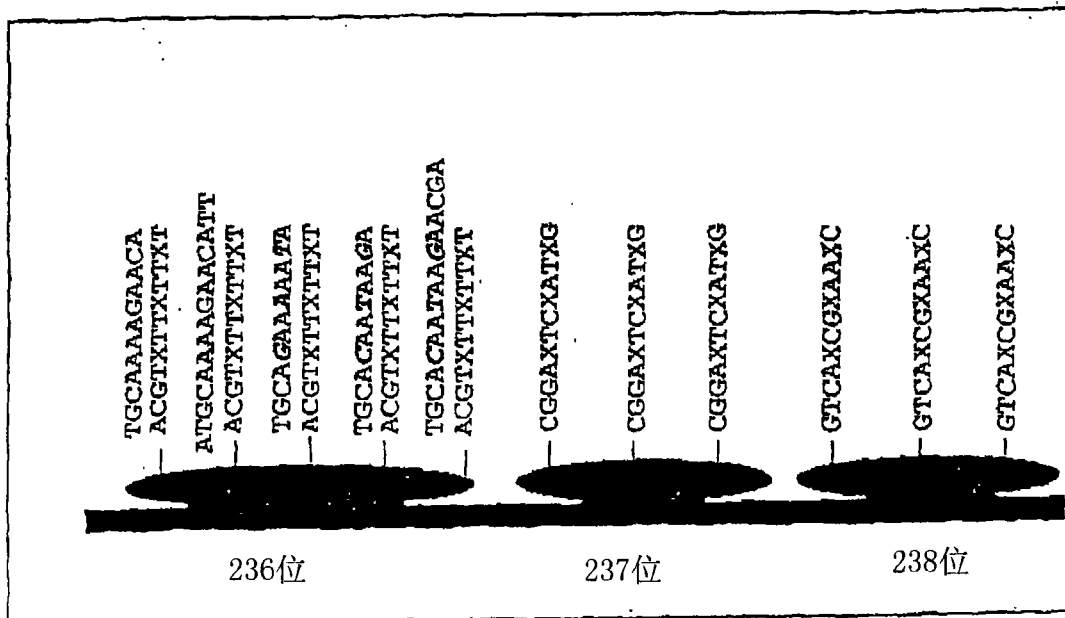


图2

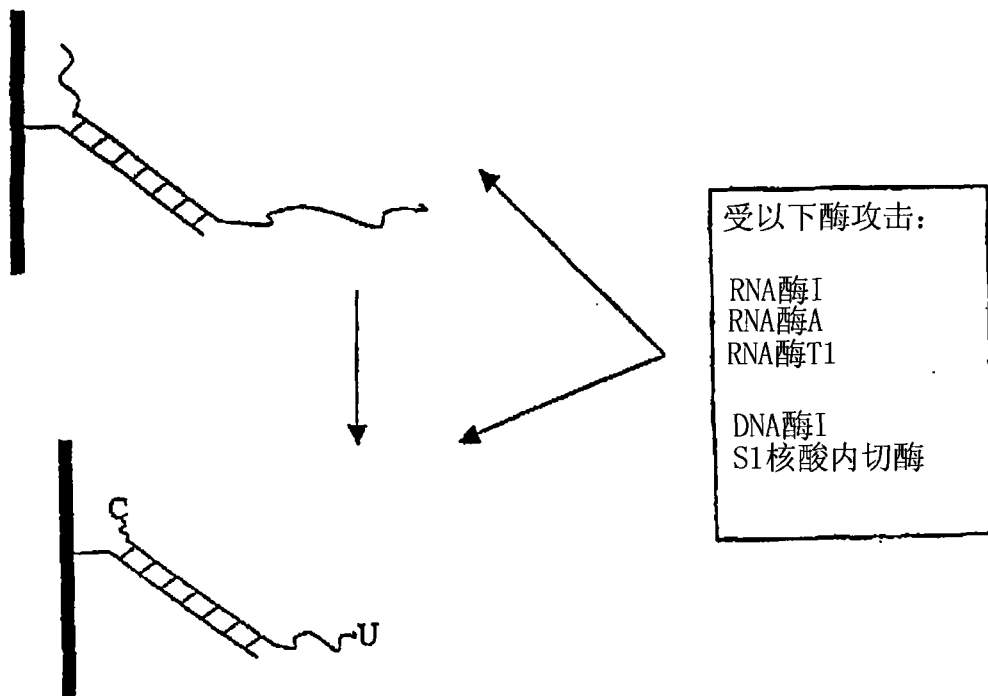


图3