



(12) 发明专利申请

(10) 申请公布号 CN 103562407 A

(43) 申请公布日 2014. 02. 05

(21) 申请号 201280017616. 9

(22) 申请日 2012. 02. 09

(30) 优先权数据

61/441, 209 2011. 02. 09 US

61/444, 539 2011. 02. 18 US

61/454, 373 2011. 03. 18 US

61/476, 115 2011. 04. 15 US

61/478, 777 2011. 04. 25 US

61/484, 197 2011. 05. 09 US

61/490, 055 2011. 05. 25 US

斯曼特·杜布 奥斯汀·苏

杰弗里·克拉克·梅伦

尼古拉斯·杰克·赫雷迪亚

凯文·奈斯 比尔·科尔斯顿

(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 高瑜 杨淑媛

(51) Int. Cl.

C12Q 1/68 (2006. 01)

(85) PCT国际申请进入国家阶段日

2013. 10. 09

(86) PCT国际申请的申请数据

PCT/US2012/024573 2012. 02. 09

(87) PCT国际申请的公布数据

W02012/109500 EN 2012. 08. 16

(71) 申请人 伯乐生命医学产品有限公司

地址 美国加利福尼亚州

(72) 发明人 约翰·弗雷德里克·里根

瑟奇·萨克森诺夫 迈克·卢塞罗

本·辛德森 菲尔·贝尔格莱德

权利要求书8页 说明书74页 附图51页

(54) 发明名称

核酸的分析

(57) 摘要

本发明提供了用于核酸分析的改进的方法、组合物和试剂盒。这些改进的方法、组合物和试剂盒可实现对样品中核酸拷贝数的估计。本发明还提供了用于确定样品中靶核酸的两个或多个拷贝的连接（例如，这两个或多个拷贝是在相同染色体上还是在不同染色体上）或用于对等位基因进行定相的方法、组合物和试剂盒。

1. 一种检测靶核酸拷贝数变异的方法,包括:
 - a. 使包含多种多核苷酸的样品与至少一种试剂接触,其中所述多核苷酸包括第一和第二靶核酸;
 - b. 使所述样品经受以下条件:当所述两种靶核酸位于同一多核苷酸上时,该条件使得所述试剂能够切割所述两种靶核酸之间特别选择的靶位点,从而分离所述两种靶核酸;
 - c. 将与所述试剂接触的样品分成多个空间上分离的分区;
 - d. 计算包含所述靶核酸的空间上分离的分区数;和
 - e. 基于所述计算,确定所述靶核酸的拷贝数。
2. 根据权利要求1所述的方法,其中所述空间上分离的分区为乳液内的小滴。
3. 根据权利要求2所述的方法,其中所述靶核酸以小于约五个拷贝/小滴的平均浓度存在。
4. 根据权利要求1所述的方法,其中所述特别选择的靶位点位于所述第一与第二靶核酸之间。
5. 根据权利要求1所述的方法,进一步包括使所述第一和第二靶核酸经历扩增反应。
6. 根据权利要求1所述的方法,其中所述第一和第二靶核酸具有相同的序列。
7. 根据权利要求1所述的方法,其中所述第一和第二靶核酸具有不同的序列。
8. 根据权利要求1所述的方法,其中所述特别选择的靶位点是能够用限制酶消化的位点。
9. 根据权利要求1所述的方法,其中所述试剂为一种或多种限制酶。
10. 根据权利要求1所述的方法,其中所述特别选择的位点和所述第一靶核酸位于同一基因内。
11. 根据权利要求1所述的方法,其中所述靶核酸与疾病或病症相关。
12. 根据权利要求1所述的方法,进一步包括计算含有参考核酸的空间上分离的分区数。
13. 根据权利要求12所述的方法,其中所述参考核酸在所述样品所来源于的基因组中以固定拷贝数存在。
14. 根据权利要求12所述的方法,其中所述参考核酸为管家基因。
15. 根据权利要求12所述的方法,其中所述参考核酸在所述样品所来源于的每个二倍体基因组中以两个拷贝存在。
16. 根据权利要求12所述的方法,其中所述确定所述靶核酸的拷贝数包括将计算的靶核酸数除以计算的参考核酸数。
17. 根据权利要求1所述的方法,其中所述试剂不切割所述靶核酸的序列。
18. 根据权利要求12所述的方法,其中所述试剂不切割所述参考核酸的序列。
19. 根据权利要求1所述的方法,其中所述靶核酸以多个拷贝存在于单一多核苷酸上。
20. 根据权利要求9所述的方法,其中所述一种或多种限制酶在所述两个靶序列之间具有超过一个识别序列。
21. 根据权利要求9所述的方法,其中所述一种或多种限制酶不展现出显著的星号活性。
22. 根据权利要求9所述的方法,其中所述一种或多种限制酶包含两种或更多种限制

酶。

23. 根据权利要求 9 所述的方法,其中使用软件选择所述一种或多种限制酶。

24. 根据权利要求 9 所述的方法,其中所述一种或多种限制酶在消化所述多核苷酸后被热灭活。

25. 根据权利要求 24 所述的方法,其中所述热灭活温度低于经限制酶切的靶片段的解链温度以便保持所述靶片段的双链性质。

26. 根据权利要求 9 所述的方法,其中进行对照限制酶消化来测定所述一种或多种限制酶消化核酸的效率。

27. 根据权利要求 1 所述的方法,其中确定所述样品中断裂的相连靶序列的百分比。

28. 一种检测核酸的拷贝数变异的方法,包括:

a. 提供包含多种多核苷酸的样品,所述多种多核苷酸包含位于多种多核苷酸中的至少一种内的第一和第二靶核酸;

b. 当所述第一和第二靶核酸存在于同一多核苷酸内时,在所述第一与第二靶核酸之间切割至少一种多核苷酸以形成切割样品;

c. 将所述切割样品分成多个空间上分离的区域;

d. 计算包含所述靶核酸的、空间上分离的区域的数目;和

e. 基于所述计算,确定所述靶核酸的拷贝数;

其中所述至少两种靶核酸中的两种位于同一多核苷酸的同区域内。

29. 根据权利要求 28 所述的方法,其中所述两种靶核酸相距小于一兆碱基。

30. 根据权利要求 28 所述的方法,其中所述两种靶核酸相距小于一千碱基。

31. 根据权利要求 28 所述的方法,其中利用限制酶完成所述切割。

32. 根据权利要求 28 所述的方法,其中所述多个空间上分离的区域为在乳液内的小滴。

33. 根据权利要求 28 所述的方法,进一步包括在步骤 d 的计算之前进行 PCR 反应。

34. 根据权利要求 28 所述的方法,其中所述方法不包括测序反应。

35. 一种检测靶核酸的拷贝数变异的方法,包括:

a. 获得样品,该样品包含 (i) 多种多核苷酸,其中至少一种所述多核苷酸包含第一靶核酸和该第一靶核酸的拷贝;(ii) 用于检测所述靶核酸的带有荧光标记物的探针;和 (iii) 能够进行 PCR 反应的试剂;

b. 将所述样品分成多个空间上分离的分区,其中所述空间上分离的分区平均包含少于五个靶核酸;

c. 使样品经历 PCR 扩增反应以检测所述靶核酸;

d. 在用于 PCR 反应的任何所述试剂变为限制性的之前,检测所述荧光标记物的荧光强度,其中分区中的较高荧光强度指示在多核苷酸上存在超过一种靶核酸;

e. 计算具有高于指示一个靶核酸拷贝的阈值的荧光强度的分区数;

f. 计算具有高于指示多个靶核酸拷贝的阈值的荧光强度的分区数;和

g. 或者 (i) 基于在步骤 e 和步骤 f 中得到的数值,计算所述靶核酸的拷贝数,或者 (ii) 确定两种靶核酸是否存在于同一多核苷酸上。

36. 根据权利要求 35 所述的方法,其中所述靶核酸的拷贝数的增加与疾病或病症相

关。

37. 一种确定多种靶核酸存在于同一多核苷酸上的方法,包括,

a. 将包含多种多核苷酸的样品分成至少两个子样品,其中所述多核苷酸包含第一和第二靶核酸;

b. 使所述第一子样品与如果所述第一靶核酸和所述第二靶核酸存在于同一多核苷酸上的话能够物理分离它们的试剂相接触;

c. 在步骤 b 之后,将所述第一子样品分成第一组分区;

d. 确定所述第一组分区中包含所述靶核酸的分区数;

e. 将第二子样品分成第二组分区;

f. 确定所述第二组分区中包含靶核酸的分区数;和

g. 将步骤 d 中获得的值与步骤 f 中获得的值进行比较,从而确定所述第一和第二靶核酸是否存在于同一多核苷酸内。

38. 根据权利要求 37 所述的方法,其中所述第一和第二靶核酸包含相同的序列。

39. 根据权利要求 37 所述的方法,其中所述试剂为限制酶。

40. 根据权利要求 37 所述的方法,其中当步骤 d 中获得的数值显著高于步骤 f 中获得的数值时,指示所述第一和第二靶核酸的连接。

41. 根据权利要求 39 所述的方法,其中所述限制酶识别所述第一与第二靶核酸之间的位点。

42. 根据权利要求 37 所述的方法,其中所述第一和第二靶序列定位于彼此相距小于一兆碱基以内。

43. 根据权利要求 39 所述的方法,其中在使所述样品与一种或多种限制酶接触后,在分离所述子样品之前,将所述样品储存少于 24 小时。

44. 根据权利要求 37 所述的方法,其中所述靶序列在所述第二子样品中在物理上不分离。

45. 根据权利要求 37 所述的方法,进一步包括确定所述第一和第二靶核酸是在相同的染色体上还是在不同的染色体上。

46. 根据权利要求 37 所述的方法,其中所述第一靶核酸的序列与所述第二靶核酸的序列不同。

47. 根据权利要求 46 所述的方法,其中所述第一靶核酸的序列是所述第二靶核酸的序列的遗传变异。

48. 根据权利要求 47 所述的方法,其中所述遗传变异是单核苷酸多态性。

49. 根据权利要求 46 所述的方法,其中所述第一和第二靶核酸位于同一基因内。

50. 根据权利要求 35 所述的方法,其中至少两种多核苷酸为染色体。

51. 根据权利要求 50 所述的方法,其中所述第一和第二靶核酸位于单独的染色体上。

52. 根据权利要求 50 所述的方法,其中第一和第二靶核酸位于同一染色体上。

53. 根据权利要求 50 所述的方法,其中至少一条所述染色体包含所述第一靶核酸的两个拷贝。

54. 根据权利要求 50 所述的方法,其中至少一条所述染色体包含所述第一靶核酸的至少三个拷贝。

55. 根据权利要求 39 所述的方法,其中一种或多种所述限制酶为甲基敏感的限制酶。
56. 一种确定样品中多核苷酸的断裂概率的方法,包括:
- 获取样品,该样品包含 (i) 多种多核苷酸,其中至少一种所述多核苷酸包含第一靶核酸和第二靶核酸;(ii) 用于检测所述第一靶核酸的第一标记探针;(iii) 用于检测所述第二靶核酸的第二标记探针;和 (iii) 能够进行 PCR 反应的试剂;
 - 将所述样品分成多个空间上分离的分区,其中所述空间上分离的分区平均包含少于五个靶多核苷酸;
 - 使样品经历 PCR 扩增反应以便检测所述第一和第二靶核酸;
 - 检测所述分区中的所述标记探针;
 - 基于步骤 d,计算同时包含所述第一和第二靶核酸的分区数;和
 - 将同时包含所述第一和第二靶核酸的分区数与所述样品的断裂程度相关联。
57. 根据权利要求 56 所述的方法,其中所述第一和第二靶核酸相距一兆碱基之内。
58. 根据权利要求 56 所述的方法,其中,在正常受试者中,所述第一和第二靶核酸相距大于一千碱基。
59. 根据权利要求 56 所述的方法,如果所述靶核酸是物理连接的,那么该方法进一步包括计算同时含有第一和第二靶核酸的分区的预期数目。
60. 根据权利要求 56 所述的方法,其中所述断裂概率与同时包含所述第一和第二靶核酸的分区数的减少呈正相关。
61. 一种用于确定多核苷酸样品中两个遗传连锁基因座的断裂概率的方法,包括:
- 对所述样品进行数字 PCR(dPCR),其中所述 dPCR 包括将多核苷酸分成单独的单元;
 - 确定具有指示存在第一基因座的信号的单元数与具有指示存在所述第一基因座和第二基因座的信号的单元数的第一总和;
 - 确定具有指示存在所述第二基因座的信号的单元数与具有指示存在所述第一基因座和所述第二基因座的信号的单元数的第二总和;和
 - 将所述第一和第二总和输入算法,从而确定样品中断裂的两个遗传连锁基因座的百分比。
62. 根据权利要求 56 或 61 所述的方法,其中所述样品中的所述多核苷酸部分降解。
63. 根据权利要求 56 或 61 所述的方法,其中所述多核苷酸为 DNA。
64. 根据权利要求 56 或 61 所述的方法,其中所述多核苷酸为 RNA。
65. 根据权利要求 56 或 61 所述的方法,其中所述多核苷酸包含 DNA 和 RNA 的混合物。
66. 一种单元型分析方法,包括:(A) 将含有核酸的水相分成多个离散体积;(B) 在所述体积中,扩增来自展示所述核酸中序列变异的第一多态基因座和第二多态基因座中每一个的至少一个等位基因序列;(C) 确定相同体积中来自两个基因座的等位基因序列的共扩增的至少一个测量值;和 (D) 基于所述共扩增的至少一个测量值,选择所述第一和第二基因座的单元型。
67. 根据权利要求 66 所述的方法,其中所述第一和第二基因座包含于所述核酸的靶区域内,并且其中所述分区步骤产生每体积小于数个靶区域拷贝的平均浓度。
68. 根据权利要求 67 所述的方法,其中所述分区步骤产生每体积小于约五个靶区域拷贝的平均浓度。

69. 根据权利要求 66 所述的方法,其中所述核酸获自或者来源于二倍体受试者。

70. 根据权利要求 66 所述的方法,其中所述确定至少一个测量值的步骤包括针对与来自相同体积的所述第二基因座的等位基因特异性扩增数据相关的所述第一基因座的等位基因特异性扩增数据,确定至少一个相关系数的步骤。

71. 根据权利要求 66 所述的方法,其中所述确定至少一个测量值的步骤包括针对分别与来自相同体积的所述第二基因座的等位基因特异性扩增数据相关的所述第一基因座的第一等位基因序列和第二等位基因序列的等位基因特异性扩增数据,确定第一相关系数和第二相关系数的步骤,并且其中所述选择单元型的步骤基于将所述第一和第二相关系数彼此进行比较的步骤。

72. 根据权利要求 66 所述的方法,其中所述确定至少一个测量值的步骤包括确定展示出所述第一基因座的特定等位基因序列与所述第二基因座的特定等位基因序列共扩增的体积数的步骤,并且其中所述选择单元型的步骤基于所述体积数。

73. 根据权利要求 66 所述的方法,其中所述确定体积数的步骤包括确定展示出所述第一基因座的第一等位基因序列或第二等位基因序列与所述第二基因座的特定等位基因序列分别共扩增的第一体积数和第二体积数的步骤,并且其中所述选择单元型的步骤基于第一和第二体积数。

74. 根据权利要求 66 所述的方法,其中所述分区步骤包括形成乳液的步骤,在该乳液中所述体积为小滴。

75. 一种单元型分析方法,包括:(A)将含有核酸的水相分成多个离散体积;(B)在所述体积中,扩增来自包含在所述核酸的靶区域中的第一多态基因座和第二多态基因座中每一个的至少一个等位基因序列;(C)从单个体积收集每个基因座的等位基因特异性扩增数据;(D)将所述第一基因座的等位基因特异性扩增数据与来自相同体积的所述第二基因座的等位基因特异性扩增数据进行关联;和(E)基于所述关联步骤,选择所述第一和第二基因座的靶区域的单元型。

76. 根据权利要求 75 所述的方法,其中所述分区步骤包括形成乳液的步骤,并且其中所述收集步骤包括从所述乳液的单个小滴收集数据的步骤。

77. 根据权利要求 75 所述的方法,其中所述形成乳液的步骤包括使所述水相通过孔口以便产生水相的单分散小滴的步骤。

78. 根据权利要求 75 所述的方法,其中所述分区步骤平均处理每个体积中少于约一个靶区域拷贝。

79. 根据权利要求 75 所述的方法,其中所述分区步骤平均处理每个体积中少于约一个基因组当量的所述核酸。

80. 根据权利要求 75 所述的方法,其中所述分区步骤包括形成至少约 1000 个体积的步骤。

81. 根据权利要求 75 所述的方法,其中所述分区步骤包括形成直径约为 10-1000 微米的小滴的步骤。

82. 根据权利要求 75 所述的方法,其中所述分区步骤包括对水相进行分区的步骤,该水相包括能够与扩增的每个等位基因序列特异性杂交的光学可辨别的荧光探针。

83. 根据权利要求 75 所述的方法,其中所述核酸是从受试者获得的遗传物质。

84. 根据权利要求 83 所述的方法,其中所述受试者是多细胞生物。

85. 根据权利要求 83 所述的方法,其中所述受试者是人。

86. 根据权利要求 75 所述的方法,其中所述核酸包括通过逆转录从受试者获得的 RNA 而得到的 cDNA。

87. 根据权利要求 75 所述的方法,其中所述扩增步骤包括扩增来自所述第一基因座的一对不同等位基因序列的步骤,并且其中所述收集数据的步骤包括收集以下数据的步骤,该数据区分单个小滴中这一对的每个等位基因序列的扩增。

88. 根据权利要求 75 所述的方法,其中所述关联步骤包括分别将所述第一基因座的每个等位基因序列的等位基因特异性扩增数据与所述第二基因座的等位基因序列的等位基因特异性扩增数据进行关联的步骤。

89. 根据权利要求 75 所述的方法,其中所述选择单元型的步骤是基于所述第一基因座的哪些等位基因特异性扩增数据展示出更高的与所述第二基因座的这些等位基因特异性扩增数据的相关性。

90. 根据权利要求 75 所述的方法,其中所述收集步骤包括从所述体积连续收集数据的步骤。

91. 根据权利要求 75 所述的方法,其中所述关联步骤包括确定所述第一和第二基因座的扩增数据的至少一个相关系数的步骤,并且其中所述选择单元型的步骤是基于所述至少一个相关系数。

92. 根据权利要求 91 所述的方法,其中所述选择单元型的步骤基于与从所述第一基因座扩增的各自不同等位基因序列相对应的相关系数。

93. 根据权利要求 75 所述的方法,进一步包括对所述等位基因特异性扩增数据应用阈值以将其转化成二进制形式的步骤,其中所述关联步骤利用所述数据的二进制形式进行。

94. 根据权利要求 93 所述的方法,其中所述关联步骤包括确定展示出来自两个基因座的特定等位基因序列共扩增的体积数的步骤。

95. 根据权利要求 94 所述的方法,其中所述关联步骤包括 (1) 确定展示出所述第二基因座等位基因序列与来自所述第一基因座的第一等位基因序列共扩增的第一体积数和展示出所述第二基因座等位基因序列与来自第一基因座的第二等位基因序列共扩增的第二小滴数的步骤,和 (2) 比较所述第一和第二体积数的步骤。

96. 根据权利要求 75 所述的方法,其中所述选择单元型的步骤是基于所述关联步骤指示相同体积中所述第一和第二基因座的特定等位基因序列的共扩增是呈负相关还是呈正相关。

97. 一种单元型分析方法,包括:(A) 将含有核酸的水相分成多个离散体积;(B) 在所述体积中,扩增来自所述核酸中第一多态基因座和第二多态基因座中每一个的等位基因序列;(C) 收集单个体积中每个等位基因序列的等位基因特异性扩增数据;和 (D) 至少部分地基于所述扩增数据指示相同体积中等位基因序列的扩增是呈负相关还是呈正相关来选择所述核酸的单元型。

98. 一种单元型分析方法,包括:(A) 将含有核酸的水相分成多个小滴,使得所述核酸的靶区域以每小滴少于约一个拷贝的平均浓度存在;(B) 在所述小滴中,扩增来自展示出所述核酸中的序列变异以及包含在所述靶区域中的第一多态基因座和第二多态基因座中

每一个的至少一个等位基因序列；(C) 确定相同小滴中来自两个基因座的等位基因序列的共扩增的至少一个测量值；和 (D) 基于所述共扩增的至少一个测量值选择所述第一和第二基因座的靶区域的单元型。

99. 根据权利要求 98 所述的方法，其中所述扩增步骤包括扩增来自所述第一基因座的第一等位基因序列和第二等位基因序列的步骤，并且其中所述确定步骤包括确定所述第一等位基因序列的共扩增的第一测量值和所述第二等位基因序列的共扩增的第二测量值的步骤，并且其中选择单元型的步骤是基于所述共扩增的第一和第二测量值的比较。

100. 一种用于单元型分析的系统，包含：

小滴发生器，其配置用于形成包含核酸的水相的小滴；

检测器，其配置用于从单个小滴收集每个基因座的等位基因特异性扩增数据；和

处理器，其配置用于将所述第一基因座的等位基因特异性扩增数据与来自相同体积的所述第二基因座的等位基因特异性扩增数据进行关联，并且基于所述等位基因特异性扩增数据的相关性，选择所述第一和第二基因座的核酸的单元型。

101. 一种用于确定第一靶核酸与第二靶核酸存在于同一多核苷酸上的概率的方法，包括：

a) 将多核苷酸样品分成至少两个子样品；

b) 在第一子样品中，用短周期 PCR 预扩增所述第一和第二靶核酸；

c) 将所述第一子样品分成第一组分区；

d) 计算来自所述第一子样品的、同时含有所述第一和第二靶核酸的分区数；

e) 将所述第二子样品分成第二组分区；

f) 计算来自所述第二子样品的、同时含有所述第一和第二靶核酸的分区数；和

g) 比较步骤 f 的值与步骤 d 的值，从而确定所述第一和第二靶核酸连接在同一多核苷酸上的概率。

102. 根据权利要求 101 所述的方法，其中所述短周期 PCR 包含少于 24 次 PCR 反应循环。

103. 根据权利要求 101 所述的方法，进一步包括使用算法来确定所述第一和第二靶核酸定相于同一多核苷酸上的概率。

104. 一种鉴定靶核酸从染色体上缺失的概率的方法，包括：

(a) 将样品细分成多个分区，其中所述样品包含：(i) 一对染色体，使得至少一条染色体含有第一靶核酸，并且其中两条染色体含有相同的标记核酸；(ii) 用于检测所述第一靶核酸的第一标记探针；(iii) 用于检测所述标记核酸的第二标记探针；和 (iii) 能够进行 PCR 反应的试剂；

(b) 进行扩增反应以检测所述分区中的所述第一靶核酸和所述标记核酸；

(c) 计算含有所述标记核酸而不含靶核酸的分区数；和

(d) 基于步骤 c 的值确定至少一条染色体上靶核酸缺失的概率，其中步骤 c 中的较高值与靶核酸从所述染色体对中的一条染色体上缺失的概率的增加有关。

105. 根据权利要求 104 所述的方法，其中所述靶核酸疑似以多个拷贝存在于至少一条染色体上。

106. 根据权利要求 104 所述的方法，其中所述标记核酸和所述靶核酸在至少一条染色

体上彼此相距很近。

107. 根据权利要求 106 所述的方法,其中所述标记核酸和所述靶核酸相距 5000 个碱基对以内。

108. 根据权利要求 104 所述的方法,其中所述标记核酸包含单核苷酸多态性。

109. 根据权利要求 104 所述的方法,进一步包括进行单独的实验以确定不同多核苷酸的断裂。

110. 根据权利要求 109 所述的方法,进一步包括使用所述单独实验的结果来辅助步骤 d 的所述确定,其中通过确定所述不同多核苷酸已断裂来加强对高缺失概率的确定。

111. 根据权利要求 104 所述的方法,进一步包括进行单独的实验以确定所述样品中高分子量 DNA 的存在。

112. 一种确定靶区域从多核苷酸上缺失或易位的概率的方法,包括:

(a) 将样品细分成多个分区,其中所述样品包含:(i) 疑似包含第一标记核酸和第二标记核酸的多核苷酸,其中所述第一和第二标记核酸相距至少一兆碱基,并且其中疑似所述靶区域位于所述第一和第二标记核酸之间;(ii) 用于检测所述第一标记核酸的第一标记探针;(iii) 用于检测所述第二标记核酸的第二标记探针;和 (iii) 能够进行 PCR 反应的试剂;

(b) 进行扩增反应以检测所述分区内的所述第一和第二标记核酸;

(c) 计算所述分区内同时包含所述第一和第二标记核酸的分区数;和

(d) 基于步骤 c 的值,确定所述靶区域缺失或易位的概率,其中步骤 c 中的较高值与所述靶核酸从多核苷酸上缺失的概率的增加有关。

113. 根据权利要求 112 所述的方法,其中所述多核苷酸为染色体。

114. 根据权利要求 112 所述的方法,其中所述靶区域是已知存在于野生型受试者中的染色体的特定区域。

115. 根据权利要求 112 所述的方法,进一步包括进行单独的实验以确定不同多核苷酸的断裂。

116. 根据权利要求 116 所述的方法,进一步包括使用所述单独实验的结果以辅助步骤 d 的所述确定,其中通过确定所述不同多核苷酸未断裂来加强对高缺失或易位概率的确定。

核酸的分析

交叉引用

[0001] 本申请要求提交于2011年2月9日的美国临时专利申请61/441,209;提交于2011年2月18日的61/444,539;提交于2011年3月18日的61/454,373;提交于2011年4月15日的61/476,115;提交于2011年4月25日的61/478,777;提交于2011年5月9日的61/484,197和提交于2011年5月25日的61/490,055的权益,所有这些申请均通过引用整体并入本文。

背景技术

[0002] 确定受试者样品中靶核酸的拷贝数可以为多种临床应用提供有用的信息。然而,如果样品中靶核酸的多个拷贝在同一多核苷酸上,那么一些用于确定靶核酸拷贝数的试验可能会低估靶核酸的拷贝数。例如,在数字聚合酶链反应(dPCR)中,可能会将包含两个靶序列拷贝的空间上分离的多核苷酸计数为只含有一个靶核酸拷贝。需要用于在实验如dPCR试验中对核酸靶序列进行拷贝数估计的改进的方法、组合物和试剂盒,其考虑了靶核酸序列的连接(如,样品中多个靶核酸序列拷贝是否在同一多核苷酸上)。

[0003] 对单元型或相邻多态性的定相(phasing)的了解在许多情况下可能是有用的。人类是二倍体生物,因为在人的每个体细胞中每种染色体类型都作为一对单个染色体而双份存在。该染色体对的一个拷贝遗传自此人的父亲,而另一个拷贝遗传自此人的母亲。因此,在每个二倍体细胞中大多数基因以两个拷贝存在。然而,这些拷贝通常具有许多基因座,其中在拷贝之间发生序列变异从而形成被称为等位基因的独特序列。

[0004] 通常,了解染色体对中的每个单个染色体的等位基因的模式,即单元型,可能是有用的。例如,如果人在基因中的两个不同基因座处具有失活突变,若该突变一起存在于同一单个染色体上,那么突变的后果可能是有限的,但是若突变分布在染色体对的两个单个染色体上,那么就可能会产生很大影响。在第一种情况下,基因的一个拷贝在两个不同的基因座上失活,但另一个拷贝能够提供有活性的基因产物。在第二种情况下,基因的每个拷贝上都有一个失活突变,所以两个基因拷贝都无法提供有活性的基因产物。取决于所讨论的基因,两个基因拷贝的突变可能会导致多种生理结果,包括非存活的表型、疾病风险性增加或者无法代谢一类药物等。

[0005] 关于单元型的信息对于生命科学应用、医学领域和诸如法医的应用市场会是很有用的信息。在许多人类(和非人类)遗传学环境中出现了单元型确定的问题。例如,许多遗传相关性依赖于单元型,并因此通过单元型进行预测。HLA(人白细胞抗原)区域是一个显著的实例,其中特定的遗传性疾病与主要组织相容性复合体的多种单元型有关。

[0006] 常规基因分型技术探询彼此分离的、诸如SNP(单核苷酸多态性)的序列变异的不同基因座。因此,这些技术可以确信地确定遗传物质样品中一对独特的等位基因存在于两个相连的基因座中的每一个上。然而,这些技术无法识别来自两个基因座的等位基因的哪种组合位于同一染色体拷贝上。需要一种新的确定单元型的方法来克服这个困难。

[0007] 类似地,需要新的方法来确定样品中两个靶核酸序列之间的断裂概率,测定核酸

(例如 DNA、RNA) 样品中的降解水平, 评估可变剪接, 或检测倒位、易位或缺失。

发明内容

[0008] 在一些方面, 本发明提供了一种检测靶核酸的拷贝数变异的方法, 包括 : a. 使包含多种多核苷酸的样品与至少一种试剂接触, 其中该多核苷酸包含第一和第二靶核酸 ; b. 使样品经受以下条件 : 当所述两种靶核酸位于同一多核苷酸上时, 该条件使得所述试剂能够切割这两种靶核酸之间特别选择的靶位点, 从而分离这两种靶核酸 ; c. 将与试剂接触的样品分成多个空间上分离的分区 ; d. 计算包含靶核酸的空间上分离的分区数 ; 和 e. 基于所述计算, 确定靶核酸的拷贝数。在一些实施方案中, 空间上分离的分区为乳液内的小滴。在一些实施方案中, 靶核酸以每个小滴少于约五个拷贝、每个小滴少于约四个拷贝、每个小滴少于约三个拷贝、每个小滴少于约两个拷贝或每个小滴少于约一个拷贝的平均浓度存在。在一些实施方案中, 特别选择的靶位点位于第一和第二靶核酸之间。在一些实施方案中, 该方法还包括使第一和第二靶核酸经历扩增反应。

[0009] 在该方面的一些实施方案中, 第一和第二靶核酸的序列相同或接近相同。在一些实施方案中, 第一和第二靶核酸具有不同的序列。在一些实施方案中, 特别选择的靶位点是能够用限制酶消化的位点。在一些实施方案中, 所述试剂为一种或多种限制酶。在一些实施方案中, 特别选择的位点和第一靶核酸位于同一基因内。在一些实施方案中, 靶核酸与疾病或病症相关。在一些实施方案中, 该方法还包括计算 (或计数) 包含参考核酸的、空间上分离的分区数目。在一些实施方案中, 参考核酸在样品所来源的基因组中以固定拷贝数存在。在一些实施方案中, 参考核酸为管家基因。在一些实施方案中, 参考核酸在样品所来源的每个二倍体基因组中以两个拷贝存在。在一些实施方案中, 确定靶核酸的拷贝数包括将已计算的靶核酸数除以已计算的参考核酸数。

[0010] 在该方面的一些实施方案中, 所述试剂不切割靶核酸的序列。在一些实施方案中, 该试剂不切割参考核酸序列。在一些实施方案中, 靶核酸以多个拷贝存在于单一多核苷酸上。在一些实施方案中, 一种或多种限制酶在两个靶序列之间具有超过一个识别序列。在一些实施方案中, 一种或多种限制酶不展现出显著的星号活性。在一些实施方案中, 一种或多种限制酶包含两种或更多种限制酶。在一些实施方案中, 使用软件来选择一种或多种限制酶。在一些实施方案中, 一种或多种限制酶在消化多核苷酸后被热灭活。在一些实施方案中, 为了保持靶片段的双链性质, 热灭活的温度低于限制酶切的靶片段的解链温度。在一些实施方案中, 进行对照限制酶消化来测定用一种或多种限制酶消化核酸的效率。在一些实施方案中, 确定样品中断裂的 (fragmented) 相连靶序列的百分比。

[0011] 另一方面, 本发明提供了一种检测核酸的拷贝数变异的方法, 包括 : a. 提供包含多种多核苷酸的样品, 所述多种多核苷酸包含位于多种多核苷酸中的至少一种内的第一和第二靶核酸 ; b. 当第一和第二靶核酸存在于同一多核苷酸内时, 切割第一与第二靶核酸之间的至少一种多核苷酸以形成切割样品 ; c. 将切割样品分离成多个空间上分离的区域 ; d. 计算包含靶核酸的、空间上分离的区域的数目 ; 和 e. 基于所述计算, 确定靶核酸的拷贝数 ; 其中至少两种靶核酸中的两种位于同一多核苷酸的同区域内。

[0012] 在一些实施方案中, 两种靶核酸相距少于一兆碱基。在一些实施方案中, 两种靶核酸相距少于一千碱基。在一些实施方案中, 用限制酶完成切割。在一些实施方案中, 多个空

间上分离的区域为乳液内的小滴。在一些实施方案中,该方法还包括在步骤 d 的计算之前进行 PCR 反应。在一些实施方案中,该方法不包括测序反应。

[0013] 另一方面,该方法包括一种检测靶核酸的拷贝数变异的方法,包括:a. 获得样品,该样品包含 (i) 多种多核苷酸,其中至少一种多核苷酸包含第一靶核酸和第一靶核酸的拷贝;(ii) 用于检测靶核酸的带有荧光标记物的探针;和 (iii) 能够进行 PCR 反应的试剂;b. 将样品分离成多个空间上分离的分区 (partitions),其中空间上分离的分区平均含有少于五种靶核酸;c. 使样品经历 PCR 扩增反应以检测靶核酸;d. 在用于 PCR 反应的任何试剂变为限制性的之前,检测荧光标记物的荧光强度,其中分区内的较高荧光强度指示多核苷酸上存在超过一种靶核酸;e. 计算具有高于指示一个靶核酸拷贝的阈值的荧光强度的分区数;f. 计算具有高于指示多个靶核酸拷贝的阈值的荧光强度的分区数;和 g. 或者 (i) 基于在步骤 e 和步骤 f 中得到的数值,计算靶核酸的拷贝数,或者 (ii) 确定两种靶核酸是否存在于同一多核苷酸上。

[0014] 在一些实施方案中,靶核酸拷贝数的增加与疾病或病症相关。

[0015] 再一方面,本发明提供了一种确定多种靶核酸存在于同一多核苷酸上的方法,包括,a. 将包含多种多核苷酸的样品分成至少两个子样品,其中该多核苷酸包含第一和第二靶核酸;b. 使第一子样品与能够将第一靶核酸与第二靶核酸(如果它们存在于同一多核苷酸上)物理分离的试剂相接触;c. 在步骤 b 之后,将第一子样品分成第一组分区;d. 确定第一组分区中包含靶核酸的分区的数目;e. 将第二子样品分成第二组分区;f. 确定第二组分区中包含靶核酸的分区的数目;和 g. 将步骤 d 中得到的值与步骤 f 中得到的值进行比较,以确定第一和第二靶核酸是否存在于同一多核苷酸内。

[0016] 在一些实施方案中,第一和第二靶核酸包含相同的序列。在一些实施方案中,所述试剂为限制酶。在一些实施方案中,当步骤 d 中得到的值显著高于步骤 f 中得到的值时,指示第一和第二靶核酸的连接。在一些实施方案中,限制酶识别第一与第二靶核酸之间的位点。在一些实施方案中,第一和第二靶序列位于彼此相距小于一兆碱基内。在一些实施方案中,在样品与一种或多种限制酶接触后,在分离子样品之前将样品储存不到 24 小时。在一些实施方案中,在第二子样品中靶序列不物理分离。在一些实施方案中,该方法还包括确定第一和第二靶核酸是在相同染色体上还是在不同染色体上。在一些实施方案中,第一靶核酸的序列与第二靶核酸的序列不同。在一些实施方案中,第一靶核酸的序列是第二靶核酸的遗传变异。在一些实施方案中,遗传变异是单核苷酸多态性。在一些实施方案中,其中第一和第二靶核酸在同一基因内。在一些实施方案中,至少两种多核苷酸为染色体。在一些实施方案中,第一和第二靶核酸位于单独的染色体上。在一些实施方案中,第一和第二靶核酸位于同一染色体上。在一些实施方案中,至少一条染色体包含第一靶核酸的两个拷贝。在一些实施方案中,至少一条染色体包含第一靶核酸的至少三个拷贝。在其它实施方案中,染色体包含第一靶核酸的至少 1、2、3、4、5、6、7、8、9 或 10 个拷贝。在一些实施方案中,一种或多种限制酶为甲基敏感的限制酶。

[0017] 再一方面,本发明提供了一种用于确定样品中多核苷酸的断裂 (fragmentation) 概率的方法,包括:a. 获取样品,该样品包含 (i) 多种多核苷酸,其中至少一种多核苷酸包含第一靶核酸和第二靶核酸;(ii) 用于检测第一靶核酸的第一标记探针;(iii) 用于检测第二靶核酸的第二标记探针;和 (iii) 能够进行 PCR 反应的试剂;b. 将样品分成多个空间

上分离的分区,其中空间上分离的分区平均包含少于五种靶多核苷酸;c. 使样品经历 PCR 扩增反应以便检测第一和第二靶核酸;d. 检测分区中的标记探针;e. 基于步骤 d,计算同时包含第一和第二靶核酸的分区的数目;和 f. 将同时包含第一和第二靶核酸的分区的数目与样品的断裂程度相关联。在一些实施方案中,第一和第二靶核酸相距一兆碱基之内。在一些实施方案中,在正常受试者中,第一和第二靶核酸相距大于一千碱基。在一些实施方案中,如果靶核酸为物理连接的,该方法还包括计算同时含有第一和第二靶核酸的分区的期望数。在一些实施方案中,断裂概率与同时含有第一和第二靶核酸的分区数的减少呈正相关。

[0018] 在另一不同但相关的方面,本发明提供了一种用于确定多核苷酸样品中两个遗传连锁基因座的断裂概率的方法,包括:a. 对所述样品进行数字 PCR(dPCR),其中 dPCR 包括将多核苷酸分离成单独的单元;b. 确定具有指示存在第一基因座的信号的单位数和具有指示存在第一基因座和第二基因座的信号的单位数的第一总和;c. 确定具有指示存在第二基因座的信号的单位数和具有指示存在第一基因座和第二基因座的信号的单位数的第二总和;和 d. 将第一和第二总和输入算法,以确定样品中断裂的两个遗传连锁基因座的百分比。

[0019] 在一些实施方案中,样品中的多核苷酸部分降解。在一些实施方案中,多核苷酸为 DNA。在一些实施方案中,多核苷酸为 RNA。在一些实施方案中,多核苷酸包含 DNA 和 RNA 的混合物。

[0020] 本发明还提供了一种用于确定第一靶核酸与第二靶核酸存在于同一多核苷酸上的概率的方法,包括:a) 将多核苷酸样品分成至少两个子样品;b) 在第一子样品中,利用短周期 PCR 预扩增第一和第二靶核酸;c) 将第一子样品分成第一组分区;d) 计算同时包含第一和第二靶核酸的第一子样品的分区数;e) 将第二子样品分成第二组分区;f) 计算同时包含第一和第二靶核酸的第二子样品的分区数;和 g) 将步骤 f 的值与步骤 d 的值进行比较,以确定第一和第二靶核酸连接在同一多核苷酸上的概率。

[0021] 在一些实施方案中,短周期 PCR 包括小于 24 个 PCR 反应循环。在一些实施方案中,该方法进一步包括使用算法来确定第一和第二靶核酸定相或存在于同一多核苷酸上的概率。

[0022] 另一方面,本发明提供了一种确定靶核酸从染色体上缺失的概率的方法,包括:(a) 将样品细分成多个分区,其中样品包含:(i) 一对染色体,使得至少一条染色体含有第一靶核酸,并且其中两条染色体都含有相同的标记核酸;(ii) 用于检测第一靶核酸的第一标记探针;(iii) 用于检测标记核酸的第二标记探针;和 (iii) 能够进行 PCR 反应的试剂;(b) 进行扩增反应,以检测分区内的第一靶核酸和标记核酸;(c) 计算含有标记核酸而不含靶核酸的分区数;和 (d) 基于步骤 c 的值确定靶核酸从至少一条染色体上缺失的概率,其中步骤 c 中的较高值与靶核酸从染色体对中的一条染色体上缺失的概率的增加有关。

[0023] 在一些实施方案中,靶核酸疑似以多个拷贝存在于至少一条染色体上。在一些实施方案中,标记核酸和靶核酸在至少一条染色体上彼此相距很近。在一些实施方案中,标记核酸和靶核酸相距 5000 个碱基对以内。在一些实施方案中,标记核酸包含单核苷酸多态性。在一些实施方案中,该方法进一步包括进行单独的测定以确定不同多核苷酸的断裂。在一些实施方案中,该方法进一步包括使用单独测定的结果来辅助步骤 d 的确定,其中通过

确定不同多核苷酸已断裂来加强对高缺失概率的确定。在一些实施方案中,该方法进一步包括进行单独的测定以确定样品中高分子量 DNA 的存在。

[0024] 再一方面,本发明提供了一种确定靶区域从多核苷酸上缺失或易位的概率的方法,包括:(a) 将样品细分成多个分区,其中样品包含:(i) 疑似包含第一标记核酸和第二标记核酸的多核苷酸,其中第一和第二标记核酸相隔至少一兆碱基,并且其中靶区域疑似位于第一和第二标记核酸之间;(ii) 用于检测第一标记核酸的第一标记探针;(iii) 用于检测第二标记核酸的第二标记探针;和 (iii) 能够进行 PCR 反应的试剂;(b) 进行扩增反应,以检测分区内的第一和第二标记核酸;(c) 计算分区内同时含有第一和第二标记核酸的分区数;和 (d) 基于步骤 c 的值,确定靶区域缺失或易位的概率,其中步骤 c 中的较高值与靶核酸从多核苷酸上缺失的概率的增加相关。在一些实施方案中,多核苷酸为染色体。在一些实施方案中,靶区域为已知存在于野生型受试者中的染色体的特定区域。在一些实施方案中,该方法进一步包括进行单独的实验以确定不同多核苷酸的断裂。在一些实施方案中,该方法进一步包括使用单独测定的结果来辅助步骤 d 的确定,其中通过确定不同的多核苷酸未断裂来加强对高缺失或易位概率的确定。

[0025] 一方面,本发明提供了单元型分析方法,包括:(a) 将含有核酸的水相分成多个离散 (discrete) 体积;(b) 在这些体积中,扩增来自展示出核酸中序列变异的第一多态基因座和第二多态基因座中每一个的至少一个等位基因序列;(c) 确定相同体积中来自两个基因座的等位基因序列的共扩增的至少一个测量值 (measure);和 (d) 基于共扩增的至少一个测量值来选择第一和第二基因座的单元型。

[0026] 在该方面的一些实施方案中,第一和第二基因座包含于核酸的靶区域中,其中分区步骤产生每体积少于数个靶区域拷贝的平均浓度。在一些实施方案中,分区步骤产生每体积少于约一个靶区域拷贝的平均浓度。

[0027] 在该方面的一些实施方案中,确定至少一个测量值的步骤包括针对与来自相同体积的第二基因座的等位基因特异性扩增数据相关的第一基因座的等位基因特异性扩增数据,确定至少一个相关系数的步骤。在一些情况下,确定至少一个测量值的步骤包括针对分别与来自相同体积的第二基因座的等位基因特异性扩增数据相关的第一基因座的第一等位基因序列和第二等位基因序列的等位基因特异性扩增数据,确定第一相关系数和第二相关系数的步骤,并且其中选择单元型的步骤是基于将第一和第二相关系数相互比较的步骤。在一些情况下,确定至少一个测量值的步骤包括确定展示出第一基因座的特定等位基因序列与第二基因座的特定等位基因序列共扩增的体积数的步骤,并且其中选择单元型的步骤是基于展示共扩增的体积数。确定体积数的步骤可进一步包括确定展示出第一基因座的第一等位基因序列或第二等位基因序列与第二基因座的特定等位基因序列分别共扩增的第一体积数和第二体积数的步骤,并且其中选择单元型的步骤基于第一和第二体积数。

[0028] 在一些实施方案中,确定至少一个测量值的步骤包括从单个体积收集每个基因座的等位基因特异性扩增数据,并且将第一基因座的等位基因特异性扩增数据与来自相同体积的第二基因座的等位基因特异性扩增数据相关联。

[0029] 在该方面的一些实施方案中,分区步骤包括形成乳液的步骤,其中体积为小滴。在一些实施方案中,收集步骤可包括从乳液的单个小滴收集等位基因特异性扩增数据。

[0030] 在一些实施方案中,形成乳液的步骤包括使水相通过孔口以便产生水相的单分散

小滴的步骤。在一些实施方案中,分区步骤包括形成约 10-1000 微米直径的小滴的步骤。在一些实施方案中,分区步骤包括形成至少约 1000 个体积的步骤。在一些情况下,分区步骤包括将包括能够与扩增的每个等位基因序列特异性杂交的光学可辨别的荧光探针的水相进行分区的步骤。

[0031] 在一些实施方案中,核酸获自或来源于二倍体受试者。在一些情况下,核酸是从受试者获得的遗传物质。在一些情况下,核酸包括通过从受试者获得的 RNA 的逆转录而得到的 cDNA。在一些情况下,受试者是多细胞生物。在一些情况下,受试者是人。

[0032] 在一些实施方案中,扩增步骤包括从第一基因座扩增一对不同的等位基因序列的步骤,并且收集数据的步骤包括收集区分单个小滴中等位基因序列对中每个等位基因序列的扩增的数据的步骤。

[0033] 在一些实施方案中,关联步骤包括分别将第一基因座的每个等位基因序列的等位基因特异性扩增数据与第二基因座的等位基因序列的等位基因特异性扩增数据进行关联的步骤。

[0034] 在一些情况下,关联步骤包括对等位基因特异性扩增数据应用阈值以将其转化成二进制形式 (binary form), 并且关联步骤利用数据的二进制形式进行。在一些情况下,关联步骤还包括确定展示来自两个基因座的特定等位基因序列的共扩增的体积数的步骤。在一些情况下,关联步骤还包括确定展示第二基因座等位基因序列与来自第一基因座的第一等位基因序列共扩增的第一体积数和展示第二基因座等位基因序列与来自第一基因座的第二等位基因序列共扩增的第二小滴数的第一步骤, 以及将第一和第二体积数进行比较的第二步骤。

[0035] 在一些实施方案中,选择单元型的步骤基于第一基因座的哪些等位基因特异性扩增数据展示出更高的与第二基因座的这些等位基因特异性扩增数据的相关性。在一些实施方案中,选择单元型的步骤基于与从第一基因座扩增的各自不同等位基因序列相对应的相关系数。在一些实施方案中,选择单元型的步骤基于关联步骤是指示相同体积中第一和第二基因座的特定等位基因序列的共扩增呈负相关还是正相关。

[0036] 另一方面,本发明提供了一种用于单元型分析的系统,包括:(a) 小滴发生器,其配置用于形成包含核酸的水相小滴;(b) 检测器,其配置用于从单个小滴收集每个基因座的等位基因特异性扩增数据;和(c) 处理器,其配置用于使第一基因座的等位基因特异性扩增数据与来自相同体积的第二基因座的等位基因特异性扩增数据相关联,并且基于等位基因特异性扩增数据的相关性选择第一和第二基因座的核酸的单元型。

[0037] 在一些方面,本发明提供了一种确定基因组 DNA 样品内的 CpG 岛的甲基化程度的方法,包括:a. 获得基因组 DNA 样品,其中基因组 DNA 包含通过 CpG 岛与第二靶核酸分隔的第一靶核酸;b. 在适于使基因组 DNA 内未甲基化的核酸能够被消化的条件下,使基因组 DNA 样品与甲基敏感的限制酶接触;c. 使样品与检测第一靶核酸的存在的探针接触;d. 使样品与检测第二靶核酸的存在的探针接触;e. 将样品分成多个空间上分离的分区;f. 在分区内,进行适合于检测第一和第二靶核酸的扩增反应;g. 确定同时包含第一和第二靶核酸的分区数;h. 将步骤 g 中得到的数值与包含第一或第二靶核酸的分区数进行比较;和 i. 由步骤 h 中的比较,确定基因组 DNA 样品内 CpG 岛的甲基化程度。在一些实施方案中,基因组 DNA 为人类基因组 DNA。在一些实施方案中,基因组 DNA 包含母体 DNA 和胎儿 DNA。在一些

实施方案中,基因组 DNA 包含胎儿 DNA。

援引并入

[0038] 本说明书中提到的所有出版物、专利和专利申请均通过引用而以相同程度并入本文,犹如特别地和单独地指出每个单独的出版物、专利或专利申请通过引用而并入。

附图说明

[0039] 本发明的新特征在随附的权利要求书中具体阐述。通过参考以下对在其中利用到本发明原理的说明性实施方式加以阐述的详细描述和附图,将会获得对本发明的特征和优点的更好的理解,在附图中:

[0040] 图 1 图示了用于估计靶序列的拷贝数的流程图。

[0041] 图 2 图示了两个靶序列在母本染色体上的实例和一个靶序列在母本染色体上而另一个在父本染色体上的实例。

[0042] 图 3a 图示了用于确定靶序列的连接的流程图的流程图。

[0043] 图 3b 图示了用于确定靶序列的连接的备选工作流程。

[0044] 图 4 是根据本发明的方面的流程图,列出了在样品分区中使用扩增进行单元型分析的示例性方法中可以进行的步骤。

[0045] 图 5 是根据本发明的方面,用于进行图 4 的方法的示例性系统的选定方面的示意图。

[0046] 图 6 是根据本发明的方面,可由位于受试者的遗传物质中相同染色体类型上的一对 SNP 创建的示例性单元型的示意图。

[0047] 图 7 是根据本发明的方面,阐述执行图 4 方法的示例性版本的流程图的示意图,其以小滴作为分区,并且分析来自图 6 受试者的遗传物质以区分图 6 中呈现的潜在的单元型。

[0048] 图 8 是根据本发明的方面,阐述关联图 7 的扩增数据的备选方法的图。

[0049] 图 9 图示了用于预测两个靶标之间的断裂的流程图。

[0050] 图 10 图示了连接的和未连接的靶标。图 10A 图示了未连接的靶标 T1 和 T2。图 10B 图示了连接的 T1 和 T2 与未连接的 T1 和 T2 的混合物。图 10C 图示了 T1 和 T2 之间的不同间距。

[0051] 图 11 图示了用于分析甲基化程度 (burden) 的流程图。

[0052] 图 12 图示了甲基化程度分析的实施方案。

[0053] 图 13 和 14 图示了在选择限制酶时可考虑的信息。

[0054] 图 15A 和图 15B 图示了可输入数据库的分析信息。

[0055] 图 16 图示了 ddPCR 实验的工作流程实例。

[0056] 图 17 图示了小滴产生中的最大拉伸。

[0057] 图 18 图示了作为样品流速的函数的最大拉伸。

[0058] 图 19 描述了未消化的样品 1-10 和已消化的样品 11-20 的小滴性质。

[0059] 图 20A 和图 20B 图示了储存的、已消化的 DNA 的 CNV 值的漂移。

[0060] 图 21 图示了个体间淀粉酶基因拷贝数的高变异性。

[0061] 图 22 图示了 CCL3L1 的拷贝数变异。

[0062] 图 23A 和图 23B 图示了 MRGPRX1 的拷贝数变异。

- [0063] 图 24A 和图 24B 图示了 CYP2D6 的拷贝数变异。
- [0064] 图 25 图示了使用简单梯度板确定的最佳共同退火温度。
- [0065] 图 26 图示了使用简单梯度板确定的最佳共同退火温度。
- [0066] 图 27 图示了 SM1 和 SMN2 拷贝数。
- [0067] 图 28 图示了 SMN1 和 SMN2 的拷贝数变异。
- [0068] 图 29 图示了 HER2 的分析。
- [0069] 图 30 图示了连接实验的结果。六个样品经历或不经历限制酶切消化,并确定靶序列的拷贝数。数据提示样品 4 和 5 在同一染色体上具有两个靶序列。
- [0070] 图 31 图示了对于 MRGPRX1 和 RPP30,在 25 个循环后实时 PCR 实验中的荧光强度分析。
- [0071] 图 32 图示了在实时 PCR 实验中,4 个样品的一式三份测量的平均荧光。疑似在一条染色体上具有两个靶标而在另一染色体上具有 0 个靶标的样品 (18853 和 19108) 比疑似在染色体上具有一个靶标拷贝的样品 (反式构型) (18507 和 BC106) 具有更强的荧光。
- [0072] 图 33 图示了对于 MRGPRX1 和 RPP30,在 28 个循环后实时 PCR 实验中的荧光强度分析。
- [0073] 图 34 图示了 ddPCR 浓度和拷贝数变异的比较。在样品 18853 和 19108 中,在 28 次循环时观察到未切割的样品的 CNV 下降。
- [0074] 图 35 图示了 28 次循环后,ddPCR 实验中 4 个样品的一式三份测量的平均荧光。疑似在一条染色体上具有两个靶标而在另一染色体上具有 0 个靶标的样品 (18853 和 19108) 比疑似在两条染色体的每一条上具有一个靶标拷贝的样品 (反式构型) (18507 和 BC106) 具有更强的荧光。
- [0075] 图 36 图示了在 31、34 和 40 次循环后, ddPCR 实验中 4 个样品的一式三份测量的平均荧光。在第 31 次循环时,疑似在一条染色体上具有两个靶标而在另一染色体上具有 0 个靶标的样品 (18853 和 19108) 比疑似在两条染色体的每一条上具有一个靶标拷贝的样品 (反式构型) (18507 和 BC106) 具有更强的荧光,但是第 34 次和 40 次循环时却不是这样。
- [0076] 图 37 图示了用于长片段 (long range) PCR 实验的实验设置。对于情形 A 和 B,预计长片段 PCR 失败,因为引物之间的距离太大。然而,对于含有具有 MRGPRX1 缺失的染色体的情形 C,PCR 应该成功。箭头指示引物。
- [0077] 图 38 图示了使用 6 个样品的长片段 PCR 实验。在 1500 处迁移的条带指示来自 MRGPRX1 内部引物的 PCR 产物,而在略低于 3000 处迁移的条带指示来自处于 MRGPRX1 基因侧翼的引物的 PCR 产物。
- [0078] 图 39 图示了 6 个样品的长片段 PCR 产物和 MRGPRX1CNV 结果,进一步通过限制酶切消化分析区分了在相同染色体上具有一个或两个 MRGPRX1 拷贝的样品。
- [0079] 图 40 图示了与具有估计的拷贝数的已消化样品的百分比差异。
- [0080] 图 41 是图示相隔 1K、10K 或 100K 个碱基的、由 FAM 和 VIC 探针识别的序列的示意图。
- [0081] 图 42 图示了核酸片段。T1 和 T2 为靶序列。图 42A 图示了 T1 和 T2 总是处于单独的核酸上的情形 (完全断裂)。图 42B 图示了 T1 和 T2 总是在核酸上连接的情形 (无断裂)。图 42C 图示了 T1 和 T2 在一些核酸上连接,并且也处于单独的核酸上的情形 (部分断

裂)。

[0082] 图 43 图示了 DNA 质量评估。

[0083] 图 44 图示了血浆中的大 DNA 污染影响 CNV 值。

[0084] 图 45 图示了用于滴定以进行 CNV 分析的限制酶。

[0085] 图 46 图示了用于 CNV 分析的限制酶滴定。

[0086] 图 47 图示了用于优化具有挑战性的 ddPCR CNV 试验的方案的开发。

[0087] 图 48 图示了利用 ddPCR 对限制酶活性的测量。

[0088] 图 49A 和图 49B 图示了通过共定位 (collocation) 的单元型分析。

[0089] 图 50 图示了可利用共定位试验分析的基因重排的实例。

发明详述

[0090] 概述

[0091] 通常,本文提供了用于分析核酸序列的方法、组合物和试剂盒,例如数字分区。例如,本文提供了用于估计样品如基因组中靶核酸序列的拷贝数的方法、组合物和试剂盒。本文还提供了用于确定样品如基因组中一种或多种靶序列的连接或单元型信息的方法、组合物和试剂盒。单元型分析信息可以是关于一个靶序列的多个拷贝是否在一条或多条染色体上的信息。利用在相同分区内不同靶序列共定位的概念,可以切实可行地推断相位,即一个突变或 SNP 的特定等位基因是否与另一个突变或 SNP 的等位基因物理连接。另一方面,本文提供了用于通过例如数字化分析共定位信号来确定核酸样品(例如,基因组 DNA 样品、RNA 样品、mRNA 样品、DNA 样品、cRNA 样品、cDNA 样品、miRNA 样品、siRNA 样品)中断裂或降解程度的方法、组合物和试剂盒。另一方面,本文提供了用于分析核酸甲基化状态,可变剪接,发现倒位、易位和缺失的方法。

[0092] 拷贝数变异的估计

[0093] 一种或多种靶序列的拷贝数变异可在许多疾病和病症中发挥作用。一种分析靶序列拷贝数变异的方法是通过数字分析,如数字 PCR 或小滴数字 PCR。然而,如果样品中靶核酸序列的多个拷贝在同一多核苷酸上,那么靶序列拷贝数的数字分析可能会低估样品中靶核酸序列的拷贝数。例如,在具有多个区室(例如,分区,空间上分离的区域)的数字 PCR 试验中,可将样品中的核酸进行分区,以使每个区室平均接收约 0、1、2 个或数个靶多核苷酸。每个分区(例如,小滴)可平均具有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下,至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个分区(如,小滴)具有 0 个靶核酸拷贝。可计算含有多核苷酸的区室的数目。然而,如果靶核酸序列的两个拷贝在单一多核苷酸上,那么可能将含有该多核苷酸的区室计数为只具有一个靶序列。

[0094] 本文提供了用于物理分离靶核酸序列的方法。通常,本文提供的方法避免由于在单一多核苷酸上存在靶序列的多个拷贝而低估靶序列的拷贝数。图 1 图示了拷贝数估计方法的实施方案的概述(101);该图和本公开内容中提供的其余附图仅用于说明性目的,并非意在限制本发明。图 1 中的步骤可以任何合适的顺序和组合进行,并且可与本公开内容的任何其它步骤结合。获得多核苷酸的第一样品(121);第一样品可以是,例如,基因组 DNA 样品。可将第一样品中的靶核酸序列进行物理分离(例如,通过使第一样品与一种或多种限制酶接触)(141)。可将第一样品分成多个分区(161)。可计算具有靶序列的分区数

(181)。然后可对靶序列的拷贝数进行估计 (201)。

[0095] 靶核酸可以是相同的 ;或者,在其它情况下,靶核酸可以是不同的。在一些情况下,靶核酸位于同一基因内。在一些情况下,靶核酸各自位于基因的不同拷贝(相同或接近相同的拷贝)中。在其它情况下,靶序列位于内含子内,或者位于基因之间的区域中。有时,一个靶序列位于基因中;而第二靶序列位于该基因外。在一些情况下,靶序列位于外显子内。

[0096] 在一些情况下,基因组包含一个靶序列。在一些情况下,基因组包含两个或更多个靶序列。当基因组包含两个或更多个靶序列时,靶序列可以大约或大于约 50%、55%、60%、65%、70%、75%、80%、85%、90%、95% 或 100% 相同。

[0097] 物理分离两个靶序列可包括通过切割核酸序列上的特定位点将靶序列进行物理分离。在一些情况下,物理分离靶核酸序列可包括使第一样品与一种或多种限制酶接触。物理分离靶核酸序列可包括在位于靶核酸序列之间的位点消化多核苷酸。在一些情况下,靶核酸序列各自位于基因内。在一些情况下,消化所针对的位点位于两个基因之间。在一些情况下,为消化选定的位点位于基因中;并且,在一些情况下,该基因是与含有靶序列的基因相同的基因。在其它情况下,为消化选定的位点位于与靶序列的基因不同的基因中。在一些情况下,靶序列和消化所针对的位点位于同一基因中;并且靶序列位于消化所针对的位点的上游。在其它情况下,靶序列和消化所针对的位点位于同一基因中;但是靶序列位于消化所针对的位点的下游。在一些情况下,可通过用一种或多种限制酶处理核酸样品而将靶核酸进行分离。在一些情况下,可通过剪切将靶核酸分离。在一些情况下,可通过超声处理将靶核酸分离。

[0098] 物理分离步骤(例如,利用一种或多种限制酶进行消化)后,样品可分为多个分区。多个分区中的每一个可包含约 0、1、2 个或数个靶多核苷酸。每个分区(例如,小滴)可平均具有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下,至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个小滴具有 0 个靶核酸拷贝。

[0099] 通常,在分区中扩增靶核酸。在一些情况下,扩增包括使用一种或多种 TaqMan 探针。

[0100] 在另一实施方案中,该方法进一步包括计算包含参考核酸序列的分区数的步骤。可知参考核酸序列在每个基因组中以特定拷贝数存在,并且可用于估计样品中靶核酸序列的基因组拷贝数。在另一实施方案中,估计拷贝数可包括将含有靶序列的分区数与含有参考核酸序列的分区数进行比较。在另一实施方案中,CNV 的估计是通过靶核酸序列与参考序列的浓度比来确定的。

[0101] 在另一实施方案中,该方法进一步包括分析第二样品的步骤,其中第二样品和第一样品来源于相同的样品(例如,将核酸样品分成第一样品和第二样品)。该方法可进一步包括不使第二样品与一种或多种限制酶接触。在一些情况下,该方法进一步包括将第二样品分成多个分区。该方法可进一步包括计算含有靶序列的第二样品的分区数。在另一实施方案中,该方法进一步包括计算含有参考序列的第二样品的分区数。在另一实施方案中,该方法包括估计第二样品中靶序列的拷贝数。在另一实施方案中,估计第二样品中靶序列的拷贝数包括比较具有靶序列的第二样品的分区数和具有参考序列的第二样品的分区数。

[0102] 可以将第一样品中靶序列的拷贝数与第二样品中靶序列的拷贝数进行比较,以确定是否低估了第二样品中靶序列的拷贝数。拷贝数被低估的程度可指示所探询的拷贝是否

全部在一条染色体上,或者是否至少一个拷贝在一条同源染色体上,而至少一个拷贝在另一条同源染色体上。如果每个二倍体基因组的值接近 1,可说明是第一种情况,而值接近 2,可说明是第二种情况。

[0103] 例如,在公开号为 20100203538 的美国专利申请中描述了另外的通过扩增来确定拷贝数差异的方法。在美国专利 6,180,349 和 Taylor 等人 (2008) PLoS One 3 (9):e3179 中描述了用于确定拷贝数变异的方法。

[0104] 当使用本文所述的方法时,可考虑许多特征:

[0105] 样品制备:将要考虑的核酸的性质可包括二级结构、扩增子长度和断裂程度。可以进行分析来确定核酸样品的断裂程度。如果核酸样品的断裂程度太高,那么可将该样品从分析中舍弃。可以采取的措施来消除样品中核酸的二级结构。可调整核酸的二级结构,例如,通过调节样品的温度或通过向样品中加入添加剂。可以确定潜在的扩增子是否太大而不能有效地扩增。在一个实施方案中,使用生物分析仪来评估核酸(如,DNA)的断裂。在另一实施方案中,使用大小排阻色谱法来评估核酸(如,DNA)的断裂。

[0106] 动态范围:增加分区或空间上分离的区域的数目可增大方法的动态范围。可将模板核酸稀释至动态范围内。

[0107] 准确度:如果使用同源样品,预期 CNV 值落在整数值(自我参考)上。漏码扩增(Drop-out amplification)可能会导致不准确的浓度测量,并因此导致不准确的 CNV 测定。可将添加剂(如,DMSO)加入到富含 GC 的试验中。

[0108] 多路复用(multiplexing):实验可以是多路复用的。例如,在本文提供的方法中可以使用两种颜色:FAM:BHQ 和 NFQ-MGB 试验;VIC:NFQ-MGB,TAMRA。HEX:BHQ。可以使用 5' 和 3' 标记,并且可以使用内部标记染料。在一些情况下,在本文提供的方法中使用的颜色数多于 2 种,如多于 3、4、5、6、7、8、9 或 10 种颜色。

[0109] 精度:可以多种方式实现精度的提高。在一些情况下,增加 dPCR 实验中的小滴数可以提高分辨靶核酸和参考核酸之间的小浓度差异的能力。软件可以通过合并来自单个孔的重复物实现“元孔(metawell)”分析。在一些情况下,本文提供的方法能够检测小于 30%、20%、15%、14%、13%、12%、11%、10%、9%、8%、7%、6%、5%、4%、3%、2% 或 1% 的拷贝数差异。

[0110] 试验概貌(assay landscape):本文所述的靶基因试验可与可购买到的或定制设计的靶基因试验相结合。

[0111] 本文所述的拷贝数变异可包括核酸序列的减少或增加。拷贝数变异可能是遗传的或可能是由新生(de novo)突变引起的。CNV 可为一个或多个不同的级别。参见,例如,Redon 等人,(2006) Global variation in copy number in the human genome. Nature 444 pp. 444-454。CNV 可由简单新生缺失、简单新生重复或缺失和重复两者引起。CNV 可由多等位基因变异的组合引起。CNV 可以是具有新生增加的复杂 CNV。CNV 可包括大约或大于约 0、1、2、3、4、5、6、7、8、9 或 10 个连续基因。CNV 可包括约 1 个至约 10 个、约 1 个至约 5 个、约 1 个至约 4 个、约 1 个至约 3 个、约 1 个至约 2 个、约 0 个至约 10 个、约 0 个至约 5 个或约 0 个至约 2 个连续基因。拷贝数变异可包括大约或大于约 100、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、500,000、750,000、一百万、五百万或一千万碱基对的增加或减少。在一些情况下,拷贝数变异可包括约 1,000 至约 10,000,000、约 10,000

至约 10,000,000、约 100,000 至约 10,000,000、约 1,000 至约 100,000 或约 1,000 至约 10,000 个核酸序列碱基对的增加或减少。拷贝数变异可能是核酸序列的缺失、插入或重复。在某些情况下,拷贝数变异可为串联重复。

[0112] 在另一实施方案中,可由通过分区样品的实时 PCR 或 ddPCR 产生的荧光信号评估 CNV 单元型。在实时 PCR 或 ddPCR 实验后期之前,当试剂可能变为限制性的时,具有较高靶序列拷贝数的分区可比具有较低靶序列拷贝数的分区具有更强的信号。在一个实施方案中,可将样品(例如,在连接实验中使用的样品的子样品)进行分区,并且可对分区(例如,小滴)进行 PCR。由于分区经历了靶核酸序列和/或参考核酸序列的指数扩增,所以可确定分区的平均荧光强度。平均强度可对应于靶序列的起始拷贝数。如果多个靶序列沿着单一多核苷酸链进行连接,那么捕获该链的分区(如,小滴)中的强度可高于捕获仅具有单一靶序列拷贝的链的分区(例如,小滴)的强度。具有较高平均振幅的阳性小滴的过量存在可提示存在具有多个 CNV 拷贝的单元型。相反,仅具有低平均振幅的阳性小滴的存在可提示在样品中仅存在具有单一 CNV 拷贝的单元型。在另一实施方案中,可基于分区的大小和分区中的试剂量来优化用于估计 CNV 的循环数。例如,具有较少试剂量的较小分区比预期具有较高试剂量的较大分区可能需要更少的扩增循环。

[0113] 该方法可以是有益的,因为其甚至可用于分析多核苷酸上彼此接近的,例如,相距小于约 10、9、8、7、6、5、4、5、2、1、0.7、0.5、0.3、0.2、0.1、0.05 或 0.01 兆碱基的靶序列拷贝;或多核苷酸上彼此非常接近的,例如,相距小于约 10、9、8、7、6、5、4、3、2 或 1 千碱基的靶序列拷贝。在一些情况下,该方法可用于分析多核苷酸上彼此非常接近的,例如,相距在约 1、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900 或 950 个碱基对以内的靶序列拷贝。在一些情况下,该方法可用于分析相隔零(0)个碱基对的靶序列拷贝。在一些情况下,该方法可用于相同、近似相同和完全不同的靶序列。

[0114] 本文描述了用于估计一种或多种靶序列的拷贝数的方法的其它实施方案。

[0115] 测定靶序列的连接

[0116] 本文所述的方法可以指示两个或多个靶序列是否连接在多核苷酸上(例如,该方法可用于测定靶序列的连接)。在一个实施方案中,提供了一种方法,该方法包括物理分离靶序列拷贝(例如,使用一种或多种限制酶),以便可将拷贝独立地分配至分区中用于数字读出,并且使用未消化的 DNA 的读数和已消化 DNA 的读数来估计靶序列拷贝是如何连接的。例如,本文所述的方法可用于确定靶序列是否存在于同一染色体上或者它们是否存在于不同染色体上(参见例如图 2)。图 2 图示了细胞核(左),其中母本染色体包含靶序列的两个拷贝,但是相应的父本染色体不包含拷贝;在右边的细胞核中,母本染色体和相应的父本染色体各包含靶序列的一个拷贝。

[0117] 图 3a 图示了一种方法的实施方案的工作流程,不限制于步骤的任何顺序。一方面,提供了一种方法(320),其包括 a) 将包含多种多核苷酸的样品分成至少两个子样品(322);b) 物理分离第一子样品中物理连接的靶序列(324);c) 将第一子样品分成第一组多个分区(326);d) 估计第一子样品中靶序列的拷贝数(328);e) 将第二子样品分成第二组多个分区(330);f) 估计第二子样品中靶序列的拷贝数(332);g) 比较第一子样品中估计的靶序列的拷贝数和第二子样品中估计的靶序列的拷贝数,以确定样品中靶序列的单元型(334)。

[0118] 在一个实施方案中,物理分离第一子样品中物理连接的靶序列包括使第一子样品与一种或多种限制酶接触。在另一实施方案中,使包含多核苷酸的样品与一种或多种限制酶接触包括在至少两个靶核酸序列之间消化核酸序列。在一些情况下,物理连接的靶核酸可通过使核酸样品与一种或多种限制酶接触而分离。在一些情况下,物理连接的靶核酸可通过剪切而分离。在一些情况下,物理连接的靶核酸可通过超声处理而分离。

[0119] 在另一实施方案中,第一和第二子样品的多个分区中的每一个都包含约 0、1、2 个或若干靶多核苷酸。每个分区(如,小滴)可平均具有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下,至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个分区(如,小滴)具有 0 个靶核酸拷贝。

[0120] 在另一实施方案中,在分区中扩增靶序列。

[0121] 在另一实施方案中,估计第一子样品中靶序列的拷贝数包括计算包含靶序列的第一子样品的分区数。在另一实施方案中,估计第一子样品中靶序列的拷贝数包括计算包含参考核酸序列的第一子样品的分区数。在另一实施方案中,估计第一子样品中靶序列的拷贝数包括比较包含靶序列的第一子样品的分区数和第一子样品中包含参考核酸序列的分区数。

[0122] 在另一实施方案中,该方法进一步包括不使第二子样品与一种或多种限制酶接触。在另一实施方案中,估计第二子样品中靶序列的拷贝数包括计算包含靶序列的第二子样品的分区数。在另一实施方案中,估计第二子样品中靶序列的拷贝数包括计算包含参考序列的第二子样品的分区数。在另一实施方案中,估计第二子样品中靶序列的拷贝数包括比较具有靶序列的第二子样品的分区数和具有参考序列的第二子样品的分区数。在另一实施方案中,第一和第二子样品的参考序列为相同序列或不同序列。

[0123] 在另一实施方案中,确定靶序列的单元型包括比较第一子样品中估计的靶序列的拷贝数和第二子样品中估计的靶序列的拷贝数。在另一实施方案中,单元型包含在单一多核苷酸上的靶序列的两个拷贝,并且在同源多核苷酸上无拷贝。在另一实施方案中,单元型分析包含第一多核苷酸上的靶序列的一个拷贝和第二(也许同源的)多核苷酸上的靶序列的第二拷贝。

[0124] 通常,第一子样品和第二子样品中拷贝数之间的差异越大,一条染色体不携带靶序列拷贝的可能性越大。

[0125] 图 3b 图示了一种方法的另一实施方案的工作流程,不限制于步骤的任何顺序。一方面,提供了一种方法(336),其包括 a) 获取包含多种多核苷酸的多核苷酸样品(338),将其分成至少两个子样品(340); b) 用短周期 PCR 预扩增第一子样品中的靶序列(342); c) 将第一子样品分成第一组多个分区(344); d) 估计第一子样品中靶序列的拷贝数(346); e) 将未预扩增的第二子样品(348)分成第二组多个分区(350); f) 估计第二子样品中靶序列的拷贝数(352); g) 比较第一子样品中估计的靶序列的拷贝数和第二子样品中估计的靶序列的拷贝数,以确定样品中靶序列的连接(354)。

[0126] 在一些情况下,用来分离靶序列的预扩增是特定靶序列扩增(STA)(Qin 等人(2008)Nucleic Acids Research 36e16),其需要进行简短的预扩增步骤,以产生针对靶核酸的单独的未连接的扩增子。

[0127] 在一个实施方案中,在第一子样品中预扩增靶序列包括使第一子样品与包含 DNA

聚合酶、核苷酸和靶序列特异性引物的反应混合物接触,以及将靶序列扩增有限次循环。任选地,该方法还包括使用参考序列的引物,并且任选地将参考序列扩增有限次循环。在一些实施方案中,循环数的数值可为4-25次循环。在一些情况下,循环数少于25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6、5或4次循环。循环数可根据小滴大小和可用试剂的量而变化。例如,对于较小大小的分区(如,小滴)可能需要较少的循环。

[0128] 可将预扩增的第一子样品分成多个分区,每个分区平均包含少于一个靶多核苷酸。每个分区(如,小滴)可能平均具有少于5、4、3、2或1个靶核酸拷贝。在一些情况下,至少0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175或200个分区(如,小滴)具有0个靶核酸拷贝。

[0129] 在另一实施方案中,估计第一子样品中靶序列的拷贝数包括计算包含参考核酸序列的第一子样品的分区数。在另一实施方案中,估计第一子样品中靶序列的拷贝数包括比较包含靶序列的第一子样品的分区数和第一子样品中包含参考核酸序列的分区数。

[0130] 在另一实施方案中,该方法进一步包括不使第二子样品经历预扩增步骤。在另一实施方案中,将第二子样品分为多个分区,每个分区平均含有约0、1、2个或若干靶多核苷酸。每个分区(如,小滴)可平均具有少于5、4、3、2或1个靶核酸拷贝。在一些情况下,至少0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175或200个分区(如,小滴)具有0个靶核酸拷贝。在另一实施方案中,估计第二子样品中靶序列的拷贝数包括计算包含靶序列的第二子样品的分区数。在另一实施方案中,估计第二子样品中靶序列的拷贝数包括计算包含参考序列的第二子样品的分区数。在另一实施方案中,估计第二子样品中靶序列的拷贝数包括比较具有靶序列的第二子样品的分区数和具有参考序列的第二子样品的分区数。在另一实施方案中,第一和第二子样品的参考序列为相同序列或不同序列。

[0131] 在另一实施方案中,确定靶序列的单元型包括比较第一子样品中估计的靶序列的拷贝数和第二子样品中估计的靶序列的拷贝数。在另一实施方案中,单元型包括单一多核苷酸上的靶序列的两个拷贝,而同源多核苷酸上无拷贝。在另一实施方案中,单元型包含单一多核苷酸上的靶序列的一个拷贝和第二(可能为同源的)多核苷酸上的靶序列的第二拷贝。

[0132] 第一子样品和第二子样品中拷贝数之间的差异越大,一条染色体上不携带靶序列拷贝的可能性越大。

[0133] 另一方面,本发明提供了一种确定多种靶核酸存在于同一多核苷酸上的方法,包括:a. 将包含多种多核苷酸的样品分成至少两个子样品,其中多核苷酸包含第一和第二靶核酸;b. 使第一子样品与能够将第一靶核酸与第二靶核酸(如果它们存在于同一多核苷酸上)进行物理分离的试剂接触;c. 步骤b后,将第一子样品分成第一组分区;d. 确定第一组分区中包含靶核酸的分区数;e. 将第二子样品分成第二组分区;f. 确定第二组分区中包含靶核酸的分区数;和g. 将步骤d中得到的值和步骤f中得到的值进行比较,以确定第一和第二靶核酸是否存在于同一多核苷酸内。

[0134] 样品可具有足够高的分子量,使得如果一对靶序列位于同一染色体上,它们在溶液中也大部分连接。如果样品中的核酸(如,DNA)完全未断裂,读数可能为0、1或2个靶序列拷贝(整数)。然而,由于核酸(如,DNA)可能部分降解,拷贝数可跨越非整数值以

及大于 2 的数值。可进行另一步骤以评估样品的核酸断裂,例如,通过使用凝胶、生物分析仪、大小排阻色谱法或数字 PCR 共定位方法(里程碑(milepost)试验)。如果发现核酸样品过度断裂,那么这就降低了可收集有关连接的信息的可能性。

[0135] 该方法可用于确定较小拷贝数状态,如 2、3、4。

[0136] 在另一实施方案中,提供了一种确定靶核酸序列的连接的方法,利用具有两种不同标记物(如 VIC 和 FAM)的探针检测相同的靶序列。例如,可将核酸序列分成多个空间上分离的分区,可扩增分区中的靶序列,并且可使用两种不同的探针检测靶序列。可将核酸样品分区,使得在每个分区中平均具有约 0、1、2 个或若干靶多核苷酸。每个分区(如,小滴)平均可具有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下,至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个分区(如,小滴)具有零个靶核酸拷贝。

[0137] 如果分区包含连接在多核苷酸上的两个靶序列,那么分区可具有仅 VIC(VIC/VIC)、仅 FAM(FAM/FAM) 或 VIC 和 FAM 的信号。与 FAM 和 VIC 靶序列随机分布所预期的相比,在分区中具有 VIC 和 FAM 信号的分区的过多存在可指示样品含有多核苷酸,该多核苷酸具有连接在多核苷酸上的至少两个靶序列。同时具有 VIC 和 FAM 信号的分区没有过多存在可指示样品中两个靶核酸序列不连接。

[0138] 共定位 (Colocalization)

[0139] 样品分区和分析分区中多个靶序列的能力可允许检测样品中的靶序列何时空间上聚集在一起。这可通过评估与如果靶序列随机分布在分区中所预期的相比,具有靶序列特定组合的分区数是否处于统计学过量来完成。这种分区的过量程度可用于估计靶序列组合的浓度。

[0140] 例如,可使用数字 PCR(如, ddPCR) 来测量两个靶序列:A 和 B。例如,将有四种类型的小滴:对两种靶序列均为阴性的小滴,对 A 为阳性的小滴,对 B 为阳性的小滴,和对两种均为阳性的小滴。在随机分布下,双阳性小滴的数目应该接近于(小滴总数)*(至少具有 B 的小滴的分数)*(至少具有 A 的小滴的分数)。如果双阳性小滴数显著超过期望值,则可推断两个靶序列在样品中是彼此接近的。该结果可意味着靶序列 A 和 B 凭借例如处于同一多核苷酸上而物理连接,它们是同一蛋白质/核酸复合物的一部分,它们是同一外体的一部分,或者它们是同一细胞的一部分。

[0141] 作为基于探针的 TaqMan 试验方案的一部分,分区中特定靶序列的存在可通过使用该靶序列特异性的荧光团进行评估。例如,当测量两个靶序列 A 和 B 时,可使用 FAM 测量 A,使用 VIC 测量 B。在一些实施方案中,可利用终点荧光,使用相同荧光团或嵌入染料评价不同靶序列,以将含有 A 的分区与含有 B 的分区以及含有 A 和 B 的分区区别开来。

[0142] 可变剪接

[0143] 可变剪接是真核生物的普遍特征。通常根据细胞类型、细胞状态或外部信号,通过使用不同构型的外显子,当由相同的基因组区域构建不同转录物时发生这种现象。例如,特定基因可具有两个外显子:外显子 A 和外显子 B,和三种可能的转录物:只含有 A 的转录物 1,只含有 B 的转录物 2,含有 A 和 B 的转录物 3。能够说明实际上表达了哪些转录物及其水平可对我们理解基础生物学、疾病、诊断学和治疗具有多重启示。

[0144] 共定位可用来通过测定具有两个或更多个目标外显子的 RNA 分子的浓度而检测

和量化可变剪接。例如,一个试验可针对一个外显子,而另一个试验可针对另一个外显子。两个试验的统计学上显著的共定位可提示含有这两个外显子的 RNA 分子的存在。可以这种方式针对并评估其它的外显子和在潜在的转录物分子上的位置。

[0145] 重排

[0146] 在另一实施方案中,可构建在多核苷酸上通常彼此相距甚远(例如,在染色体上相隔数百万 bp 的两个基因)的两种分析(扩增子)。一种分析在一个通道上(如,FAM),另一种在另一个通道上(如,VIC)。在数字扩增方法如 dPCR 或 ddPCR 中,通常不应该在基线统计期望之上观察到在同一分区(如,小滴)中的共定位。如果发生 FAM 和 VIC 的共定位(例如,本文所述的连接分析所测定的),那么可能指示两个基因座在基因组上彼此邻近。该结果可指示倒位或易位,这取决于基因座通常位于哪里。如果分析的终点荧光足够不同,分析也可以在同一通道上多路复用。可多路复用超过两种分析以捕捉多个倒位/易位事件或解释给定的易位可呈现不同断点的事实。

[0147] 重排的检测可用于诊断或预测多种病状,包括癌症和胎儿缺陷。重排的检测可用于为受试者选择一种或多种治疗性治疗。例如,t(9;22)(q34.1;q11.2)易位的检测可导致与慢性髓细胞性白血病(CML)相关的 BCR-ABL 融合蛋白的产生。可使用伊马替尼(Gleevec)治疗表达 BCR-ABL 的 CML 患者。

[0148] 可用本文所述的方法检测的重排包括,例如,倒位、易位、重复或缺失(参见例如图 50)。

[0149] 确认由数字实验产生的连接(单元型)信息

[0150] 在另一实施方案中,用样品的数字分析和限制酶消化所确定的连接信息可通过一种或多种其它试验来确认。在一个实施方案中,本文所述分区样品的实时 PCR 或 ddPCR 实验期间的信号产生可用于确认连接信息。在一个实施方案中,可将样品(例如,在连接实验中使用的样品的子样品)进行分区,并在分区(如,小滴)上进行 PCR。由于分区经历了靶核酸序列和/或参考核酸序列的指数扩增,所以可确定分区的平均荧光强度。含有具有靶核酸序列的多个(例如 2 个)连接拷贝的多核苷酸的分区比仅含有一个拷贝的靶核酸序列的小滴可具有更高的荧光强度。

[0151] 在另一实施方案中,长片段 PCR 可用于确认连接信息。例如,PCR 可用于检测同一染色体上两个串联排列的靶核酸序列拷贝(顺式构型)的存在,并且其可用于检测另一条染色体上靶核酸序列的缺失。在一个实施方案中,可以使用扩增区域(疑似具有靶序列串联拷贝的区域)之外的引物。在另一实施方案中,可将 DNA 多核苷酸分区成小滴。将 DNA 多核苷酸分区成小滴可能是有利的,因为这可允许检测两种类型的 DNA 物质:a) 具有串联排列的靶序列的 DNA 区段,和 b) 靶序列缺失的 DNA 区段。如果整体(in bulk)(例如,不对多核苷酸进行分区)进行类似的反应,那么代表靶序列缺失的 DNA 的较小 PCR 产物可超过代表具有串联排列的靶序列的 DNA 区段的 PCR 产物。因此,只能产生一种 PCR 产物。这些 PCR 产物的大小差别可使用例如凝胶电泳或生物分析仪进行估计。

[0152] 在一些情况下,具有串联排列的靶核酸序列拷贝的 DNA 可能太大而不能成功地被 PCR 扩增(例如,大小 >20KB)。在这些情况下,通常只扩增代表靶核酸序列缺失的 DNA 区段的较小 PCR 产物。如果靶核酸序列太长而不允许产生 PCR 产物,那么可在含有靶核酸序列缺失的染色体上进行 PCR。在这种情况下,如果 PCR 在该序列缺失的区域上进行,则可产生

产物,但是如果由于引物之间的距离太大而出现靶序列,则不会产生产物。

[0153] 在一些实施方案中,长片段 PCR 可用于分辨连接或确定拷贝数估计值。在一些情况下,长片段 PCR 与本文提供的方法联合使用。在一些情况下,可(单独或与本文提供的方法联合)使用亲本或其它亲属的基因型来推断目标个体的拷贝数状态。

在一些情况下,该方法可包括用重组 DNA 技术对染色体区域进行克隆,并对该染色体区域的单个拷贝进行测序。在一些实施方案中,该方法包括(a)使用下一代测序技术来鉴定与排列紧密的(例如,相距小于约 2000 核苷酸、小于约 1000 核苷酸、小于约 500 核苷酸、小于约 200 核苷酸或小于约 100 核苷酸)和一起存在于同一测序读取中的多态性相关的信息,和(b)使用本文提供的方法来鉴定与相距较远的(例如,相距大于约 5、10、50、100、150、200、250、300、400、450、500、550、600、650、700、750、800、850、900、950、1000、1250、1500、1750、2000、2500、3000、3500、4000、4500 或 5000 个核苷酸)多态性相关的信息。在一些实施方案中,该方法包括使用本文提供的方法来鉴定与相距较远的(例如,相距大于约 5、10、50、100、150、200、250、300、400、450、500、550、600、650、700、750、800、850、900、950、1000、1250、1500、1750、2000、2500、3000、3500、4000、4500 或 5000 个核苷酸)多态性相关的信息。在一些情况下,该方法包括使用本文提供的方法联合使用受试者双亲或其它近亲属的基因型信息来使用孟德尔遗传规律推断相位信息。然而,该方法不能对每个多态性进行定相。一些实施方案包括结合使用本文提供的方法和统计学方法来确定连接。

[0154] 单元型

[0155] 单元型可指同时存在于或连接在单一染色体上(例如,在同一染色体拷贝上)和/或同一片核酸和/或遗传物质上的两个或更多个等位基因。定相可以是确定等位基因是否同时存在于同一染色体上的过程。确定基因组中哪些等位基因是相连的对于考虑基因是如何遗传的可能是有用的。本发明通过扩增分区的样品为单元型分析提供了包括方法和装置的系统。

[0156] 图 4 示出了列有可在单元型分析的示例性方法(20)中进行的步骤的流程图。这些步骤可以任何合适的顺序和组合进行,并且可与本发明的任何其它步骤结合。通常,可从具有二倍体或更高染色体互补物的受试者获得样品(22)。可将样品进行分区(24)。将样品进行分区可包括将样品的包含核酸的水相分区或分开。可扩增一对(或更多)多态基因座(26)。可收集每个多态基因座的等位基因特异性扩增数据(28)。可将多态基因座的扩增数据与来自相同体积的扩增数据进行关联(30)。可选择多态基因座的单元型(32)。

[0157] 一种单元型分析方法可利用从受试者如人获得的样品来进行。可将样品的含有核酸的水相分区为多个离散体积,如小滴。每个体积可平均含有少于约一个基因组当量的核酸,使得每个体积平均含有少于约一个拷贝的第一多态基因座的等位基因和连接的第二多态基因座的等位基因。可扩增核酸中来自第一多态基因座和第二多态基因座中每一个的至少一个等位基因序列。可从单个体积收集每个基因座的可辨别的等位基因特异性扩增数据。第一基因座的等位基因特异性扩增数据可与来自相同体积的第二基因座的等位基因特异性扩增数据相关联。可基于等位基因特异性扩增数据的相关性来选择第一和第二基因座中每一个的核酸的单元型。总之,若等位基因序列构成了受试者的单元型,则该方法可依靠相同体积中来自不同基因座的等位基因序列的共扩增,而相反地,若等位基因序列不构成受试者的单元型,则不进行共扩增。

[0158] 用于单元型分析的系统能够执行本文公开的方法。该系统可包含配置用于形成包含核酸的水相小滴的小滴发生器。该系统还可包含配置用于从单个小滴收集每个基因座的等位基因特异性扩增数据的检测器。该系统还可包含处理器。该处理器可配置用于将第一基因座的等位基因特异性扩增数据与来自相同体积的第二基因座的等位基因特异性扩增数据进行关联,并基于等位基因特异性扩增数据的相关性选择核酸的单元型。

[0159] 任选地,可将样品分为子样品。任选地,可使第一子样品与切割多态基因座之间位点的限制酶接触;并且可任选地使第二子样品与限制酶接触。任选地,可将来自第一子样品的等位基因特异性扩增数据与来自第二子样品的等位基因特异性扩增数据进行关联。

[0160] 本发明的其它方面在以下章节中呈现:(I) 定义,(II) 系统概述,(III) 由连接的 SNP 创建的示例性潜在单元型,(IV) 利用在小滴中扩增的示例性单元型分析,和 (V) 选定的实施方案。

[0161] I. 定义

[0162] 本公开内容中使用的技术术语具有本领域技术人员公认的含义。然而,下列术语可能具有如下所述的附加含义。

[0163] 序列变异可以是在群体成员之间或受试者和 / 或样品的染色体类型的拷贝之间发现的基因组序列的任何趋异。序列变异也可称为多态性。

[0164] 基因座可为基因组的特定区域,通常为小于约一千碱基或小于约一百核苷酸的相对较短区域。

[0165] 多态基因座可为群体中和 / 或受试者和 / 或样品中在此处存在序列变异的基因座。多态基因座可由共同存在于基因组的同一位置上的两种或更多种不同序列产生。不同的序列可由于一个或多个核苷酸的置换、缺失 / 插入和 / 或任何数目的核苷酸(通常为相对少量的核苷酸,如少于约 50、10 或 5 个核苷酸)的重复而彼此不同。在示例性实施方案中,多态基因座由单核苷酸多态性(“SNP”)即在群体内变化的单核苷酸位置而创建。

[0166] 等位基因可为在多态基因座上共存的两种或多种形式之一。等位基因还可称为变体。等位基因可为多态基因座处存在的主要或优势形式,或者次要或甚至极罕见的形式。因此,来自同一多态基因座的一对等位基因可在群体中以任何合适的比例存在,如约 1:1、2:1、5:1、10:1、100:1、1000:1 等。

[0167] 等位基因序列可为表征、包含和 / 或重叠等位基因一串核苷酸。可利用等位基因序列的扩增来确定相应等位基因是否存在于样品分区中的多态基因座上。

[0168] 单元型可为同时存在于或连接在单一染色体上(例如,在同一染色体拷贝上)和 / 或在同一片核酸和 / 或遗传物质上的两个或更多个等位基因;单元型也可指同时存在于或连接在单一染色体上的两个或更多个靶核酸。靶核酸可以是相同的或不同的。

[0169] 连接可以是来自不同多态基因座的等位基因之间的连接,也可以是相同的或几乎相同的靶核酸之间的连接。显示连接(和 / 或已连接)的多态基因座通常包括同时存在于同一染色体拷贝上的各自的等位基因,并且可在同一拷贝上彼此相对靠近,如尤其在约 10、1 或 0.1 兆碱基以内。

[0170] II. 用于单元型分析的系统概述

[0171] 图 4 示出了列有可在单元型分析的示例性方法 20 中进行的步骤的流程图。这些步骤可以任何合适的顺序和组合进行,并且可与本发明的任何其它步骤结合。

[0172] 如 22 所示,可获得样品。可从受试者,通常为具有二倍体或更高染色体互补物的受试者获得样品。换句话说,受试者通常有至少两套染色体,且受试者细胞中有至少一对每种类型的染色体。例如,人的每个体细胞含有两个拷贝的染色体 1、2、3 等,得到 23 对染色体(两套染色体),总共 46 条染色体。

[0173] 如 24 所示,可将样品分区。将样品分区可包括将样品的含有核酸的水相分区或分开。分区将水相分成多个离散的且分开的体积,也可称为分区。体积可通过流体如连续相(例如,油)彼此分离。或者,体积可通过壁如样品架的壁彼此分离。可连续地或平行地形成体积。在一些实施方案中,体积为形成乳液的分散相的小滴。

[0174] 如 26 所示,可扩增一对(或更多)多态基因座。更特别地,可扩增来自每个多态基因座的至少一个等位基因序列。每个等位基因序列是基因座的相应等位基因所特有的。在一些实施方案中,仅可从每个基因座扩增一个等位基因序列,或可从至少一个基因座扩增一对等位基因序列。特定等位基因序列和扩增的不同等位基因序列的数目可由水相分区前水相中包含的特定引物组来确定。

[0175] 如 28 所示,可收集每个多态基因座的等位基因特异性扩增数据。该数据可与单个体积中每个等位基因序列的可分辨扩增(或其缺乏)相关。数据可由对应于并且能够特异性杂交每个扩增的等位基因序列的可分辨探针来检测。可平行地或连续地从体积收集数据。在示例性实施方案中,可通过扩增信号的光学检测收集数据。例如,光学检测可包括检测代表每个等位基因序列的可分辨扩增的荧光信号。

[0176] 如 30 所示,可将多态基因座的扩增数据和来自相同体积的扩增数据进行关联。关联通常确定哪些等位基因序列最有可能同时存在于单个体积中,因此最初在受试者遗传物质中的同一染色体拷贝上彼此连接。关联可包括确定与相同体积中不同等位基因序列的共扩增相对应的至少一个相关系数。在一些情况下,关联可包括确定与相同基因座的一对等位基因序列中的每一个与另一基因座的等位基因序列共扩增相对应的一对相关系数。关联还可包括将相关系数进行相互比较和/或与阈值进行比较,或者可包括确定相关系数是负还是正。在一些实施方案中,可用已经通过应用可区分扩增阳性信号和扩增阴性信号的阈值而转化为二进制形式的扩增数据进行关联。关联还可或备选地可包括比较展示出不同组等位基因序列共扩增的体积数,和/或比较展示出一组等位基因序列共扩增的体积数与仅展示出一个等位基因序列扩增的体积数。

[0177] 在一些实施方案中,28 和 30 所示的一个或两个步骤可替换为确定相同体积中来自两个基因座的等位基因序列共扩增的至少一个测量值的步骤。可使用共扩增的任何合适的测量值,如通过来自相同体积的多态基因座的等位基因特异性扩增数据的关联而获得的至少一个相关系数。在其它实例中,共扩增的测量值可以是代表来自每个基因座的等位基因序列共扩增的至少一个数值或频率的至少一个值。关联扩增数据和确定共扩增测量值的其它方面在本公开内容的其它地方如在第 IV 部分中描述。

[0178] 在一些实施方案中,可将含有多核苷酸的样品分为两个或更多个子样品。可使第一子样品与切割两个多态基因座之间的位点的限制酶接触。然后可将第一子样品分成多个分区。而后可收集如本文所述的每个多态基因座的等位基因特异性扩增数据。可将尚未与切割两个多态基因座之间的位点的限制酶接触的第二子样品分成多个分区。而后可收集每个多态基因座的等位基因特异性扩增数据。可将来自第一和第二子样品的扩增数据进行关

联以确定多态基因座的单元型。

[0179] 如 32 所示,可选择多态基因座的单元型。选择可基于扩增数据的相关性和 / 或基于共扩增的至少一个测量值。单元型可从正在研究的多态基因座的一组潜在单元型中进行选择。选定的单元型通常包括指定在受试者的相同染色体拷贝上可能彼此连接的至少一对特定等位基因。

[0180] 图 5 示出了用于执行图 4 的方法 20 的示例性系统 40 的选定方面的示意图。该系统可包括小滴发生器 (DG) 42、热循环仪 (TC) 44、检测器 (DET) 46 和处理器 (PROC) 48。箭头 50-54 在系统组件间延伸,分别指示小滴 (50 和 52) 和数据 (54) 的移动。

[0181] 小滴发生器 42 可形成含有核酸的水相的小滴。小滴可连续地或平行地形成。

[0182] 热循环仪 44 可使小滴经受多次加热和冷却循环从而驱动等位基因序列的扩增,如 PCR 扩增。热循环仪可以是间歇式热循环仪,其可平行地扩增所有小滴,或者可以是基于流动的热循环仪,其可连续地扩增小滴。

[0183] 检测器 46 收集扩增数据,如小滴的等位基因特异性扩增数据。检测器可以是,例如,荧光检测器,并且可连续地或平行地检测小滴。

[0184] 处理器 48,又可称为控制器,可与检测器 46 通讯,并可程序化以处理来自检测器的扩增数据。处理器(可以是数字处理器)可以程序化以处理来自检测器的原始数据,如减去背景和 / 或基于小滴大小将小滴数据标准化。处理器另外或备选地可以程序化以应用阈值将数据转化为二进制形式,从而进行扩增数据的关联,以计算和 / 或比较共扩增的一个或多个测量值,从而基于相关性和 / 或测量值或其任意组合选择单元型。

[0185] 小滴发生器、热循环仪、检测器和控制器的其它方面在 2010 年 7 月 8 日公开的、公开号为 2010/0173394A1 的美国专利申请中有所描述,其通过引用并入本文。

[0186] III. 由连接的 SNP 创建的示例性潜在单元型

[0187] 图 6 示意性地图示了由连接的 SNP 创建的单元型分析的情况,其中二倍体受试者 60 的遗传物质在两个不同的基因座的每一个中都有两种不同的核苷酸。单元型分析的目的在于确定在每个染色体拷贝上第一基因座处的哪种核苷酸与第二基因座处的哪种核苷酸结合。

[0188] 受试者 60 可具有由一对单核苷酸多态性 66、68 创建的两组备选单元型构型 62、64 中的任一种。每种构型代表两种单元型:构型 62 具有单元型 (G,C) 和 (A,T),而构型 64 具有单元型 (G,T) 和 (A,C)。受试者的细胞 70 包括一对相同类型的染色体拷贝 72、74。(未示出细胞中可能存在的其它染色体类型。)染色体拷贝 72、74 在序列上可能彼此大部分相同,但这些拷贝通常也具有许多序列变异的基因座,如多态基因座 76、78,这两个染色体拷贝在此处的序列是不同的。基因座 76、78 包含在基因组区域或靶区域 80 中,该区域在细胞 70 的细胞核中用虚线框勾画出,并且作为代表基因座 76、78 的基因型 82 的复合序列邻近细胞放大示出。(为简化描述,图 6(和图 7)中只示出了每个染色体拷贝和靶区域的一条链。)

[0189] 基因型 82 可通过任何适当的基因分型技术在单元型分析之前或作为单元型分析的一部分来确定。基因型 82 显示,基因座 76 的单一多态性核苷酸在染色体拷贝 72 和 74 上为“G”和“A”(反之亦然),而对于基因座 78 则为“C”和“T”。然而,基因型并没有指示两个基因座的单个核苷酸是如何结合在染色体拷贝 72、74 上的。因此,基因型可由备选的、

潜在的单元型构型 62、64 产生。本文所公开的单元型分析允许确定受试者的遗传物质中存在哪些潜在单元型。

[0190] IV. 在小滴中使用扩增进行的示例性单元型分析

[0191] 图 7 示意性地图示了图 4 方法的示例性版本 88 的执行。在此,分析来自图 6 的受试者的遗传物质,从而区分前述部分中所述的备选的、潜在的单元型构型。

[0192] 如 92 所示,获得样品 90。将该样品在包含受试者的核酸 96 的水相 94 中进行处理。在该视图中,为简便起见,只描述含有基因组区域 80 的片段 98。片段 98 足够长以致只有少数(例如,不完整片段 100、102)未包括来自两个基因座 76、78(也参见图 6)的等位基因序列 104-110。水相可配置用于等位基因序列 104-110 的 PCR 扩增。

[0193] 如 114 所示,形成小滴 112。小滴可以是包含将小滴彼此分离的连续相 118 的乳液 116 的一部分。小滴可以是单分散的,即,基本上大小相同。可能合适的示例性单分散性程度在 2010 年 7 月 8 日公开的公开号为 2010/0173394A1 的美国专利申请中有所描述,其通过引用并入本文。

[0194] 片段 98 可在其形成时随机分配至小滴中。在片段 98 在分区的水相中适当稀释下,并且通过恰当选择小滴大小,每个小滴中平均含有靶区域 80 的少于约一个拷贝或分子。因此,一些小滴如 120 所示的空小滴不含靶标的拷贝,许多只含有靶区域的一个拷贝,一些含有靶标的两个或更多个拷贝(如 122 所示的小滴),并且一些只含有靶区域的一个等位基因序列(如 124 所示的小滴)。

[0195] 如 126 所示,可扩增等位基因序列。在此,从基因座 76 扩增两个等位基因序列 104 和 108,并且只从基因座 78 扩增等位基因序列 110(也参见图 6)。扩增的每个等位基因的拷贝标记为 104'、108' 和 110'。在其它实施方案中,仅可从每个基因座扩增一个等位基因序列,或者尤其是,可从每个基因座扩增至少两个等位基因序列。(例如,可用与扩增等位基因 110 相同的引物扩增等位基因序列 106,但是为简化描述,此处未示出等位基因序列 106 的扩增。)

[0196] 如 130 所示,可从小滴收集等位基因特异性扩增数据。在该实例中,收集荧光数据,其中使用各自包含在不同等位基因特异性探针中的不同的、可辨别的荧光染料,为每个等位基因序列 104'、108'、110' 提供扩增信号。特别是,染料 FAM、VIC 和 ROX 发射分别与等位基因序列 104、108 和 110 的扩增相关的 FAM-、VIC- 和 ROX 信号 132-136。在其它实施方案中,可检测所有四个等位基因序列 104-110 或仅两个等位基因序列(每个基因座上 一个)的等位基因特异性扩增。

[0197] 如 140 所示,将扩增数据进行关联,和/或确定相同小滴中等位基因序列共扩增的至少一个测量值。图 142、144 示意性地示出了关联和/或确定共扩增测量值的方法。图 142 将单个小滴(由图中的点表示)的 FAM 和 ROX 信号强度作图,而图 144 将单个小滴的 VIC 和 ROX 信号强度作图。对于给定信号类型(因此给定的等位基因序列)代表扩增阴性(“-”)和扩增阳性(“+”)小滴的信号值在邻近图的各坐标轴处示出。

[0198] 线 146、148 代表每幅图的扩增数据与线性关系的最佳拟合。然而,这两种拟合具有相关联的相反极性的相关系数。图 142 中的扩增数据提供了负相关系数,因为相同小滴中等位基因序列 104(由 FAM 信号报告)和等位基因序列 110(由 ROX 信号报告)的共扩增为负相关。相反,图 144 中的扩增数据提供了正相关系数,因为相同小滴中等位基因序列

108(由 VIC 信号报告)和等位基因序列 110(由 ROX 信号报告)的共扩增为正相关。可相互比较相关系数从而选择单元型。例如,可基于哪个相关系数更大(例如,更接近 1.0)和/或哪个为正(如果只有一个为正)来选择单元型。在此,可基于 VIC 和 ROX 信号的正相关来选择包含等位基因序列 104 和 106 的第一单元型和包含等位基因序列 108 和 110 的第二单元型。在一些实施方案中,可仅基于一种相关性来选择单元型,如基于相关系数是负还是正,或基于相关系数与预先确定的值的比较。

[0199] 图 8 示出了柱状图 160,该柱状图图示了用于关联图 7 的扩增数据的备选方法。通过将每种类型的小滴信号(FAM、VIC 和 ROX)与可区分每个等位基因序列的扩增阳性小滴(指定为“1”)和扩增阴性小滴(指定为“0”)的阈值进行比较,已将图 7 的扩增数据转化成二进制形式。图 160 列出了数据的二进制形式,以单独或组合地呈现各种等位基因序列的扩增阳性小滴数。如 162 所示的左侧两个柱允许比较只含有等位基因序列 104(FAM)的小滴数与同时含有等位基因 104(FAM)和 110(ROX)的小滴数。左侧的数据显示等位基因序列 104 的扩增与等位基因序列 110 的扩增相关性不好。换句话说,等位基因序列 104 和 110 在相同小滴中不倾向于共扩增。如 164 所示的右侧两个柱允许比较只含等位基因序列 108(VIC)的小滴数和同时含有等位基因序列 108(VIC)和 110(ROX)的小滴数。右侧数据显示等位基因序列 108 的扩增与等位基因 110 的扩增相关性很好。换句话说,等位基因序列 108 和 110 在相同小滴中倾向于共扩增。将左侧的一对柱和右侧的一对柱分开考虑或共同考虑,指示了其中等位基因序列 108 与等位基因序列 110 相关联的单元型。

[0200] 在用本文所述的方法、组合物或试剂盒分析前,可使包含遗传连锁基因座的样品经受断裂。包含遗传连锁基因座的样品可通过,例如,机械剪切、使样品通过注射器、超声处理、热处理(例如,90°C 下 30 分钟)和/或核酸酶处理(例如,用 DNA 酶、RNA 酶、核酸内切酶、核酸外切酶、限制酶)进行断裂。在分析之前,包含遗传连锁基因座的样品可不经处理或经过有限处理。

[0201] 在另一实施方案中,使用小滴数字 PCR(ddPCR),可进行针对两个基因组基因座如共同染色体上的两个基因的双重反应(duplex reaction)。根据它们的荧光,小滴可分类为四个群体。例如,如果 FAM- 标记的探针用于检测一个基因座,而 VIC- 标记的探针用于检测另一基因座,则四个群体可以是 FAM+/VIC+、FAM+/VIC-、FAM-/VIC+ 和 FAM-/VIC-。通过比较具有这些群体中的每一个的小滴的数目,可以确定基因座共分离至同一小滴中的频率。利用泊松统计学,可以估计与两个分离的基因座碰巧在同一小滴中的情况相比,彼此实际连接的种类的百分比。

[0202] 可使用本文所述的方法、组合物和试剂盒检查以确定其在样品中是否仍连接或在样品中已分离的遗传连锁基因座的数目可以是大约、至少约或大于约 2、3、4、5、6、7、8、9 或 10。可使用本文所述的方法、组合物和试剂盒检查以确定其在样品中是否仍连接或在样品中已分离的遗传连锁基因座的数目可以是约 2 至约 10、约 2 至约 8、约 2 至约 6、约 2 至约 4、约 3 至约 10、约 3 至约 8、约 3 至约 6、约 4 至约 10 或约 4 至约 6。

[0203] 每个遗传连锁基因座之间的碱基对的数目可以为大约、至少约、或小于约 10bp、25bp、50bp、75bp、100bp、250bp、500bp、750bp、1000bp、2000bp、3000bp、4000bp、5000bp、6000bp、7000bp、8000bp、9000bp、10,000bp、15,000bp、20,000bp、33,000bp、50,000bp、75,000bp、100,000bp、250,000bp、500,000bp、750,000bp、1,000,000bp、1,250,000bp、

1, 500, 000bp、2, 000, 000bp、5, 000, 000bp 或 10, 000, 000bp。每个遗传连锁基因座之间的碱基对的数目可以为约 10 到约 10, 000, 000bp、约 100 到约 10, 000, 000bp、约 1, 000 到约 10, 000, 000bp、约 1, 000 到约 1, 000, 000bp、约 1, 000 到约 500, 000bp、约 1, 000 到约 100, 000bp、约 3000 到约 100, 000bp、约 1000 到约 33, 000bp、约 1, 000 到约 10, 000bp, 或约 3, 000 到约 33, 000bp。每个遗传连锁基因座之间的碱基对的数目可以为 0bp。

[0204] 在一些实施方案中,单元型分析方法包括检查两个不同基因座上的两个等位基因是否共定位于同一空间上分离的分区。在一个实施方案中,可分析两个基因座上额外的等位基因。例如,如果在数字实验中两个不同基因座上的两个等位基因没有共定位,那么可以分析两个基因座上的一个或多个其它等位基因,从而为共定位提供阳性对照。例如,假设母本遗传的染色体在基因座 1 上具有等位基因 A,而在与基因座 1 相距 2100bp 的基因座 2 上具有等位基因 Y。在相应的父本遗传的染色体上,假定等位基因 B 在基因座 1 上,而等位基因 Z 在基因座 2 上。如果将包含这些核酸的核酸样品分成空间上分离的分区,并对等位基因 A 和等位基因 Z 进行扩增,那么等位基因 A 和等位基因 Z 的扩增信号应该很少或从不共定位于单一分区中,因为等位基因 A 和等位基因 Z 不相连。可进行数字分析来确认等位基因 A 和等位基因 Y 连接在母本遗传的染色体上,或等位基因 B 和等位基因 Z 连接在父本遗传的染色体上。

[0205] 利用两种颜色进行的单元型分析

[0206] 虽然本文示出的实施方案展示的是使用三色系统来测量定相,但是也可使用两色系统测量定相。例如,如果需要对两种杂合 SNP (Aa 和 Bb) 进行定相,则可以设计针对 A 的 FAM 试验和针对 B 的 VIC 试验。同时含有 A 和 B 的分区过量将指示 A 和 B 之间连接,这提示两种单元型为 A-B 和 a-b。缺乏这种过量可能提示单元型的备选组合 :A-b 和 a-B。可以确定 DNA 具有足够高的分子量,从而作出后一推断。为了确认单元型的备选组合,可能需要在单独的孔中运行另一双重试验,其中针对等位基因的不同组合。例如,可运行针对 A 的 FAM 试验和针对 b 的 VIC 试验。同时含有 A 和 b 的分区过量将指示 A 与 b 之间连接,提示两种单元型为 A-b 和 a-B。

[0207] 参考序列

[0208] 在涉及拷贝数分析(或本文所述的其它应用)的方法中,计数例如在给定的基因组中发现特定序列(如,靶序列)的次数是有用的。在一些实施方案中,该分析是通过评估(或比较)已知在每个基因组中以一定固定拷贝数存在的靶核酸序列和参考核酸序列的浓度来完成的。对于参考序列,可以使用在每个二倍体基因组中以两个拷贝存在的管家基因(例如,维持细胞基本功能所需的基因)。将靶序列的浓度或量除以参考序列的浓度或量可以得到每个基因组的靶序列拷贝数的估计值。也可使用一种或多种参考序列确定靶序列的连接。

[0209] 可在本文所述的方法中用作参考序列的管家基因可包括编码以下的基因:转录因子、转录阻抑物、RNA 剪接基因、翻译因子、tRNA 合成酶、RNA 结合蛋白质、核糖体蛋白质、RNA 聚合酶、蛋白质加工蛋白质、热激蛋白质、组蛋白、细胞周期调节物、细胞凋亡调节剂、癌基因、DNA 修复/复制基因、碳水化合物代谢调节物、三羧酸循环调节物、脂类代谢调节物、氨基酸代谢调节物、核苷酸合成调节物、NADH 脱氢酶、细胞色素 c 氧化酶、ATP 酶、线粒体蛋白质、溶酶体蛋白质、变幻虫蛋白质、核糖核酸酶、氧化酶/还原酶、细胞骨架蛋白质、细胞粘

着蛋白、通道或转运蛋白、受体、激酶、生长因子、组织坏死因子等。可用于所述方法中的管家基因的具体实例包括,例如,HSP90、 β -肌动蛋白、tRNA、rRNA、ATF4、RPP30 和 RPL3。

[0210] 为了确定靶序列的连接,与另一个基因座遗传连锁的基因座之一可以是共同参照物,例如,RPP30。任何遗传连锁基因座可用于本文所述的方法中。

[0211] 可使用单拷贝参考核酸(例如,基因)来确定拷贝数变异。可使用多拷贝参考核酸(例如,基因)来确定拷贝数,从而扩大动态范围。例如,多拷贝参考基因可包含基因组中大约或大于约 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000 或 100,000 个拷贝。多种不同核酸(例如,多种不同基因)可用作参考。

[0212] 确定核酸断裂的概率

[0213] 可以进行数字分析来确定核酸样品中两个标记之间的断裂程度。图 9 图示了工作流程(900)。图 9 中的步骤可以任何合适的顺序和组合进行,并且可与本公开内容中的任何其它步骤结合。可获得多核苷酸样品(920)。可将该样品分成多个分区(940),使得每个分区平均只含有约 0、1、2 个或数个靶多核苷酸。每个分区(如,小滴)可平均具有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下,至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个分区(如,小滴)具有 0 个靶核酸拷贝。

[0214] 可对分区进行分析以计数含有第一靶序列和第二靶序列的分区(960),并且可使用算法预测第一和第二靶序列之间的断裂(980)。

[0215] 如果两个不同的基因座(T1 和 T2)在不同的多核苷酸上,那么含有多核苷酸的样品(920)将包含只含有 T1 和只含有 T2 的多核苷酸(参见图 10A)。然而,如果 T1 和 T2 在同一多核苷酸上,则包含含有 T1 和 T2 的多核苷酸的样品可以有三种:含有 T1 的断裂多核苷酸、含有 T2 的断裂多核苷酸和含有 T1 和 T2 的断裂多核苷酸(图 10B)。T1 和 T2 之间的距离越长,T1 和 T2 之间断裂的概率就越高。可将样品分区(图 9:940)。可进行数字分析,如数字 PCR 或小滴数字 PCR,并且可计数具有 T1、T2 以及 T1 和 T2 的信号的分区的分区(960)。可开发算法并用其确定 T1 和 T2 之间的断裂概率(980)。该算法可利用 T1 和 T2 之间的碱基或碱基对的数目(若已知)。该方法可用来确定 DNA 样品的断裂程度。如果含有 T1 和 T2 信号的分区的分区数大于 T1 和 T2 在同一分区中所预期的分区数,那么该观察结果可以指示 T1 和 T2 是连接的。

[0216] 对核酸(例如 DNA)样品使用上述方法来确保 DNA 具有足够高的分子量从而将连接信息保留在样品中是有益的。

[0217] 在本文所述的利用 DNA 的任何方法中,可进行试验来估计样品中 DNA 的断裂,并且该方法可引入有关 DNA 断裂的信息。在另一实施方案中,可基于样品中 DNA 的断裂程度将试验结果标准化。

[0218] 还可通过例如凝胶、生物分析仪或大小排阻色谱法测定核酸的断裂。

[0219] 甲基化程度的测定

[0220] 可使用里程碑试验来确定 CpG 岛的甲基化状态。在一个实施方案中,本文提供了一种方法,其包括使用可通过 ddPCR(里程碑试验)测量的扩增子的连接。扩增子之间的距离可跨越大于 10kb,完全覆盖对于 CpG 岛所观察到的大小范围。

[0221] CpG 岛可常见于人类基因的 5' 区域,包括在启动子序列中。在 X- 染色体失活和印刻过程中,CpG 岛的甲基化可能在基因沉默中发挥作用。启动子区域 CpG 岛的异常甲基化可能与瘤形成中肿瘤抑制基因的转录失活有关。例如,已观察到肿瘤抑制基因 p16、hMLH1 和 THBS1 的异常甲基化。微卫星不稳定性和启动子相关 CpG 岛的甲基化之间存在关联。甲基化状态的确定可在测定母体血浆中的胎儿 DNA 负载中起作用。CpG 序列的甲基化可以是癌症的标记。

[0222] 图 11 图示了确定 CpG 岛甲基化状态的方法的工作流程 (1100)。图 11 中的步骤可以任何合适的顺序和组合进行,并且可与本公开内容的任何其它步骤结合。可获得多核苷酸样品 (1110)。可用甲基化敏感酶处理该多核苷酸样品 (1120)。可将样品分成多个分区 (1130),使得每个分区平均只含有约 0、1、2 个或数个靶多核苷酸。每个分区(如,小滴)可平均含有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下,至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个分区(如,小滴)具有零个靶核酸拷贝。可扩增位于 CpG 岛 5' 和 3' 端侧翼的核酸序列 (1140)。可对分区进行分析以确定在 CpG 岛侧翼的序列是在相同分区中还是在不同分区中 (1150),从而确定甲基化状态 (1160)。如果 CpG 是甲基化的,那么它就不会被甲基化敏感酶切割,并且侧翼序列(里程碑)可在同一分区中。如果 CpG 不是甲基化的,那么 CpG 岛可被甲基化敏感酶切割,并且侧翼序列(里程碑)可在不同分区中。在另一实施方案中,甲基化的 CpG 岛可被能切割甲基化序列但不能切割未甲基化序列的甲基化敏感酶切割。在另一实施方案中,甲基化的 CpG 岛可被能识别(例如,结合)甲基化序列而不识别未甲基化序列的甲基化敏感酶切割。

[0223] 在另一实施方案中,里程碑(标记)可在 CpG 岛内。在另一实施方案中,里程碑(标记)可在单一 CpG 序列侧翼。在另一实施方案中,里程碑(标记)可在大约或至少约 2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400 或 500 个 CpG 的侧翼(例如,一个标记在 CpG 的 5',而另一个标记在 CpG 的 3')。

[0224] 在另一实施方案中,可确定 CpG 岛的脱甲基化。CpG 岛的甲基化状态可在一段时间内进行测定(例如,在不同时间点如日、月、年采集的样品中)。可以比较来自不同个体的样品之间的 CpG 岛的甲基化状态。

[0225] CpG 可以通过磷酸二酯键与鸟嘌呤连接的胞嘧啶,且 CpG 岛可以是一簇这样的二核苷酸序列。CpG 岛,或 CG 岛,可以是含有高频 CpG 位点的基因组区域。CpG 岛的长度可为大约、少于约或多于约 100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400、2500、2600、2700、2800、2900、3000、3100、3200、3300、3400、3500、3600、3700、3800、3900、4000、4100、4200、4300、4400、4500、4600、4700、4800、4900、5000、5100、5200、5300、5400、5500、5600、5700、5800、5900 或 6000 个碱基对。CpG 岛可为约 100 到约 6000bp、约 100 到约 5000bp、约 100 到约 4000bp、约 100 到约 3000bp、约 100 到约 2000bp、约 100 到约 1000bp、约 300 到约 6000bp、约 300 到约 5000bp、约 300 到约 4000bp、约 300 到约 3000bp、约 300 到约 2000bp 或约 300 到约 1000bp。在一个实施方案中,CpG 岛可以是长

度为至少 200bp 的一段 DNA,其中 C+G 含量至少为 50%,且观察到的 CpG/ 预期 CpG 的比例至少为 0.6 (Gardiner-Garden, M. 和 Frommer M. (1987) J. Mol. Biol. 196:261-282)。在另一实施方案中,CpG 岛可以是大于 500bp 的一段 DNA,其中 C+G 含量等于或大于 55%,且观察到的 CpG/ 预期 CpG 的比例至少为 0.65 (Takai D. 和 Jones P. (2001) PNAS 99:3740-3745)。

[0226] 在一种鉴定甲基化胞嘧啶的方法中,利用亚硫酸氢盐处理将胞嘧啶转化为尿嘧啶;然而,甲基化胞嘧啶受到保护。随后的 PCR 将尿嘧啶转化成胸腺嘧啶,导致相同的目标区域产生不同的序列输出。或者,通过亚硫酸氢盐处理导致的 C→T 转化可改变扩增子的解链温度,该解链温度可通过扩增后的熔解曲线分析来测量。然而,这种方法的灵敏限度约为 1-5%。此外,完全转化需要大量的亚硫酸氢盐处理,这可导致 DNA 的降解。

[0227] 在该方法的一个实施方案中,在目标 CpG 岛的侧翼设计两种扩增子(如 FAM 和 VIC 标记的)。DNA 样品用一种或多种甲基化敏感的限制酶消化,并且可通过数字 PCR 如 ddPCR 分析消化的和未消化的样品。未消化的样品应显示出 FAM 和 VIC 扩增子 100% 连接,而消化的样品将显示出 FAM 和 VIC 扩增子 0% 连接,从而指示甲基化状态的差异。该系统的扩展精度可检测两个样品之间至少约 1/10、1/50、1/100、1/200、1/300、1/400、1/500、1/600、1/700、1/800、1/900、1/1000、1/2000、1/3000、1/4000、1/5000、1/6000、1/7000、1/8000、1/9000、1/10,000 的甲基化状态差异。图 12 图示了该方法的一个实施方案。

[0228] 待扩增区域可在 CpG 岛 5' 或 3' 端的大约或至少约 1、10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、250、300、350、400、450 或 500bp 以内。

[0229] 一种或多种甲基化敏感的限制酶可用于本文提供的方法、组合物和试剂盒中。一种或多种甲基化敏感的限制酶可包括,例如, DpnI、Acc65I、KpnI、ApaI、Bsp120I、Bsp143I、MboI、Bsp0I、NheI、Cfr9I、SmaI、Csp6I、RsaI、Ecl136II、SacI、EcoRII、MvaI、HpaII、MspJI、LpnPI、FsnEI、DpnII、McrBc 或 MspI。在一个实施方案中,甲基化敏感的限制酶可切割未甲基化的核酸,但不能切割甲基化的核酸。一种或多种甲基化敏感的限制酶可以是约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24 或 25 种甲基化敏感酶。

[0230] 在一些实施方案中,细胞核经历 Dam 甲基化和 / 或 Dcm 甲基化。

[0231] 表 1 包括其限制酶切割 DNA 的能力可被 CpG 甲基化阻断或阻碍的限制酶的列表。

[0232] 表 1. 其切割能力可被 CpG 甲基化阻断或阻碍的限制酶的列表。

| 酶 | 序列 |
|----------|---------------------------|
| AatII | GACGT/C |
| Acc65I | G/GTACC |
| AccI | GT/MKAC |
| AciI | CCGC(-3/-1) |
| AclI | AA/CGTT |
| AfeI | AGC/GCT |
| AgeI | A/CCGGT |
| AgeI-HF™ | A/CCGGT |
| AhdI | GACNNN/NNGTC |
| AleI | CACNN/NNGTG |
| ApaI | GGGCC/C |
| ApaLI | G/TGCAC |
| AscI | GG/CGCGCC |
| AsiSI | GCGAT/CGC |
| AvaI | C/YCGRG |
| AvaII | G/GWCC |
| BaeI | (10/15)ACNNNNGTAYC(12/7) |
| BanI | G/GYRCC |
| BbvCI | CCTCAGC(-5/-2) |
| BceAI | ACGGC(12/14) |
| BcgI | (10/12)CGANNNNNTGC(12/10) |
| BcoDI | GTCTC(1/5) |
| BfuAI | ACCTGC(4/8) |
| BfuCI | /GATC |
| BglI | GCCNNNN/NGGC |
| BmgBI | CACGTC(-3/-3) |
| BsaAI | YAC/GTR |

| 酶 | 序列 |
|------------|---------------|
| BsaBI | GATNN/NNATC |
| BsaHI | GR/CGYC |
| BsaI | GGTCTC(1/5) |
| BsaI-HF™ | GGTCTC(1/5) |
| BseYI | CCCAGC(-5/-1) |
| BsiEI | CGRY/CG |
| BsiWI | C/GTACG |
| BsII | CCNNNNN/NNGG |
| BsmAI | GTCTC(1/5) |
| BsmBI | CGTCTC(1/5) |
| BsmFI | GGGAC(10/14) |
| BspDI | AT/CGAT |
| BspEI | T/CCGGA |
| BsrBI | CCGCTC(-3/-3) |
| BsrFI | R/CCGGY |
| BssHII | G/CGCGC |
| BssKI | /CCNGG |
| BstAPI | GCANNNN/NTGC |
| BstBI | TT/CGAA |
| BstUI | CG/CG |
| BstZ17I | GTA/TAC |
| BtgZI | GCGATG(10/14) |
| BtsIMutI | CAGTG(2/0) |
| Cac8I | GCN/NGC |
| ClaI | AT/CGAT |
| DpnI | GA/TC |
| DraIII | CACNNN/GTG |
| DraIII-HF™ | CACNNN/GTG |

| 酶 | 序列 |
|------------------------|--------------|
| DrdI | GACNNNN/NGTC |
| EaeI | Y/GGCCR |
| EagI | C/GGCCG |
| EagI-HF TM | C/GGCCG |
| EarI | CTCTTC(1/4) |
| EciI | GGCGGA(11/9) |
| Eco53kI | GAG/CTC |
| EcoRI | G/AATTC |
| EcoRI-HF TM | G/AATTC |
| EcoRV | GAT/ATC |
| EcoRV-HF TM | GAT/ATC |
| FauI | CCCGC(4/6) |
| Fnu4HI | GC/NGC |
| FokI | GGATG(9/13) |
| FseI | GGCCGG/CC |
| FspI | TGC/GCA |
| HaeII | RGCGC/Y |
| HgaI | GACGC(5/10) |
| HhaI | GCG/C |
| HincII | GTY/RAC |
| HinfI | G/ANTC |
| HinPII | G/CGC |
| HpaI | GTT/AAC |
| HpaII | C/CGG |
| Hpy166II | GTN/NAC |
| Hpy188III | TC/NNGA |
| Hpy99I | CGWCG/ |
| HpyAV | CCTTC(6/5) |

| 酶 | 序列 |
|-----------------------|---|
| HpyCH4IV | A/CGT |
| I-CeuI | CGTAACTATAACGGTCCTAAGGTAGCG AA(-9/-13) |
| I-SceI | TAGGGATAACAGGGTAAT(-9/-13) |
| KasI | G/GCGCC |
| MboI | /GATC |
| MluI | A/CGCGT |
| MmeI | TCCRAC(20/18) |
| MspAII | CMG/CKG |
| MwoI | GCNNNNN/NNGC |
| NaeI | GCC/GGC |
| NarI | GG/CGCC |
| Nb.BtsI | GCAGTG |
| NciI | CC/SGG |
| NgoMIV | G/CCGGC |
| NheI | G/CTAGC |
| NheI-HF TM | G/CTAGC |
| NlaIV | GGN/NCC |
| NotI | GC/GGCCGC |
| NotI-HF TM | GC/GGCCGC |
| NruI | TCG/CGA |
| Nt.BbvCI | CCTCAGC(-5/-7) |
| Nt.BsmAI | GTCTC(1/-5) |
| Nt.CviPII | (0/-1)CCD |
| PaeR7I | C/TCGAG |
| PhoI | GG/CC |
| PI-PspI | TGGCAAACAGCTATTATGGGTATTATG GGT(-13/-17) |

| 酶 | 序列 |
|----------|---|
| PI-SceI | ATCTATGTCGGGTGCGGAGAAAGAGGTAAT(-15/-19) |
| PleI | GAGTC(4/5) |
| PmeI | GTTT/AAAC |
| PmlI | CAC/GTG |
| PshAI | GACNN/NNGTC |
| PspOMI | G/GGCCC |
| PspXI | VC/TCGAGB |
| PvuI | CGAT/CG |
| PvuI-HF™ | CGAT/CG |
| RsaI | GT/AC |
| RsrII | CG/GWCCG |
| SacII | CCGC/GG |
| SalI | G/TCGAC |
| SalI-HF™ | G/TCGAC |
| Sau3AI | /GATC |
| Sau96I | G/GNCC |
| ScrFI | CC/NGG |
| SfaNI | GCATC(5/9) |
| SfiI | GGCCNNNN/NGGCC |
| SfoI | GGC/GCC |
| SgrAI | CR/CCGGYG |
| SmaI | CCC/GGG |
| SnaBI | TAC/GTA |
| StyD4I | /CCNGG |
| TfiI | G/AWTC |
| TliI | C/TCGAG |
| TseI | G/CWGC |

| 酶 | 序列 |
|-------|---------|
| TspMI | C/CCGGG |
| XhoI | C/TCGAG |
| XmaI | C/CCGGG |
| ZraI | GAC/GTC |

[0233] 在另一实施方案中,包含 CpG 序列或 CpG 岛的核酸与一种或多种 DNA 甲基转移酶 (MTases) 接触,并且进行本文所述的分析 CpG 序列的甲基化的方法。DNA 甲基转移酶可将甲基从 S-腺苷甲硫氨酸转移到胞嘧啶或腺嘌呤残基。该甲基转移酶可以是,例如,CpG MTase。DNA MTase 可以是 m6a MTase(可产生 N6-甲基腺嘌呤的 MTase);m4C MTase(可产生 N4-甲基胞嘧啶的 MTase);或 m5C MTase(可产生 C5-甲基胞嘧啶的 MTase)。MTase 可以是,例如, DNMT1、DNMT3A 和 DNMT3B。

[0234] 分离

[0235] 靶序列的物理分离可以序列特异性或非序列特异性的方式发生。分离靶序列的非序列特异性手段包括使用注射器、超声处理、热处理(例如,90°C 30 分钟)和某些类型的核酸酶处理(例如,用 DNA 酶、RNA 酶、核酸内切酶、核酸外切酶)。

[0236] 限制酶

[0237] 分离核酸序列的序列特异性方法可涉及一种或多种限制酶的使用。一种或多种限制酶可用于本文所述的任何方法中。例如,除了其它方法外,可以使用限制酶分离靶序列拷贝,以便准确估计拷贝数状态、评估定相、产生单元型或确定连接。可以选择一种或多种酶,以便对靶核酸序列之间的核酸(例如, DNA 或 RNA)进行限制酶切,但不限制酶切待扩增或分析的区域。在一些实施方案中,可选择限制酶,使得限制酶在靶序列内,例如,在靶序列的 5' 或 3' 末端内进行切割。例如,如果靶序列没有间隔序列而串联排列,那么靶序列的物理分离可包括靶序列内序列的切割。在数字分析(例如, ddPCR)反应中可使用已消化的样品进行拷贝数估计、连接测定、单元型分析、检查 RNA 或 DNA 降解或确定例如 CpG 岛的甲基化程度。

[0238] 可选择限制酶,并且可为了广泛应用而在大量样品和试验类型中确定和验证最佳条件,例如用于确定 CNV 的数字 PCR(例如, ddPCR)和本文所述的任何其它方法。

[0239] 可利用计算机软件为本文所述的方法、组合物和 / 或试剂盒选择一种或多种限制酶。例如,该软件可以是 Qtools 软件。

[0240] 在本文所述的方法、组合物和 / 或试剂盒中使用的一种或多种限制酶可以是任何限制酶,包括可从 New England **BioLabs**®, Inc. (参见 www.neb.com) 获得的限制酶。限制酶可以是,例如,限制性内切核酸酶、寻靶内切核酸酶、切口内切核酸酶或高保真 (HF) 限制酶。限制酶可为 I 型、II 型、III 型或 IV 型酶或寻靶内切核酸酶。在一些情况下,限制酶切消化在高星号活性条件下发生。在一些情况下,限制酶切消化在低星号活性条件下发生。

[0241] I 型酶可在远离识别位点的位点处进行切割;可能需要 ATP 和 S-腺苷-L-甲硫氨酸来发挥作用;并且可以是既有限制酶切活性又有甲基化酶活性的多功能蛋白质。I 型限

制性内切核酸酶的识别序列可以是二分的 (bipartite) 或中断的。限制性内切核酸酶的亚单位构型可以是五聚复合物。I 型限制性内切核酸酶的共活化剂和活化剂包括, 例如, 镁、AdoMet (S-腺苷甲硫氨酸; SAM、SAmE、SAM-e) 和 ATP。I 型限制性内切核酸酶可在远离识别位点并且可变的切割位点处进行切割。I 型限制性内切核酸酶的实例可包括, 例如, EcoKI、EcoAI、EcoBI、CfrAI、StyLTIII、StyLTIII 和 StySPI。

[0242] II 型酶可在识别位点内或与识别位点相距特定短距离处进行切割; 可能需要镁; 并且可不依赖于甲基化酶而起作用。II 型限制性内切核酸酶的识别序列可为回文序列或中断的回文序列。II 型限制性内切核酸酶的亚单位结构可为同源二聚体。用 II 型限制性内切核酸酶对切割位点进行切割可导致具有 3' 突出端、5' 突出端或平端的片段。II 型限制性内切核酸酶的实例包括, 例如, EcoRI、BamHI、KpnI、NotI、PstI、SmaI 和 XhoI。

[0243] II 型限制酶存在几种亚型, 包括 IIb 型、IIs 型和 IIe 型。

[0244] IIb 型限制性内切核酸酶可具有二分的或中断的识别序列。IIb 型限制性内切核酸酶的亚单位结构可以是异源三聚体。IIb 型限制性内切核酸酶的辅因子和活化剂可包括镁和 AdoMet (用于甲基化)。IIb 型限制性内切核酸酶可在两条链上的切割位点处, 在确定的、对称的、相隔短距离的识别位点的两侧切割, 并且产生 3' 突出端。IIb 型限制性内切核酸酶的实例包括, 例如 BcgI、Bsp24I、CjeI 和 CjePI。

[0245] IIe 型限制性内切核酸酶可具有回文的、具有不明确性的回文的或非回文的识别位点。IIe 型限制性内切核酸酶的亚单位结构可为同源二聚体或单体。IIe 型限制性内切核酸酶的辅因子和活化剂可包括镁, 并且可以顺式或反式作用于核酸内切酶的第二识别位点可以充当别构影响因素。IIe 型限制酶能以限定的方式在具有识别序列或相隔短距离的切割位点处切割。可使用活化剂 DNA 完成切割。IIe 型限制酶的实例包括, 例如, NaeI、NarI、BspMI、HpaII、SaiI、EcoRII、Eco57I、AtuBI、Cfr9I、SauBMKI 和 Ksp632I。

[0246] IIs 型限制酶可具有非回文的识别序列。识别序列可以是连续的且没有不明确性。IIs 型限制性内切核酸酶的亚单位结构可为单体。可以与 IIs 型限制酶共同使用的辅因子可为镁。IIs 型限制酶可以限定的方式在切割位点处切割, 其中至少一个切割位点在识别序列之外。IIs 型限制酶的实例包括, 例如, FokI、Alw26I、BbvI、BsrI、EarI、HphI、MboII、SfaNI 和 Tth111I。

[0247] III 型酶可在距识别位点较短距离处切割, 且可能需要 ATP。S-腺苷-L-甲硫氨酸可刺激利用 III 型酶的反应, 但不是必要的。III 型酶可作为与修饰甲基化酶的复合物的一部分而存在。III 型限制性内切核酸酶的识别序列可为非回文的。可与 III 型限制性内切核酸酶共同使用的辅因子和活化剂包括, 例如, 镁、ATP (未水解的) 和相反方向的、相隔可变距离的第二未修饰位点。III 型限制性内切核酸酶的实例包括, 例如, EcoP15I、EcoPI、HinfIII 和 StyLTI。

[0248] IV 型酶可针对甲基化的 DNA。IV 型限制酶的实例包括, 例如, 大肠杆菌 (*E. coli*) 的 McrBC 和 Mrr 系统。

[0249] 限制酶可为寻靶内切核酸酶。寻靶内切核酸酶可为双链 DNA 酶。寻靶内切核酸酶可具有大的、非对称的识别位点 (例如, 12-40 个碱基对)。寻靶内切核酸酶的编码序列可嵌入内含子或内含肽中。内含肽可为“蛋白质内含子”, 它可自我切除并利用肽键重新连接其余的部分 (外显肽)。寻靶内切核酸酶可容忍其识别序列内的某些序列简并。寻靶内切核

酸酶的特异性可为 10-12 个碱基对。寻靶内切核酸酶的实例包括 I-CeuI、I-SceI、I-PpoI、PI-SceI、PI-PspI 和 PI-SceI。

[0250] 在本文的方法、组合物和 / 或试剂盒中使用的限制酶可为二聚体、三聚体、四聚体、五聚体、六聚体等。

[0251] 在本文所述的方法、组合物和 / 或试剂盒中使用的一种或多种限制酶可为杂合蛋白或嵌合蛋白的组分。例如,含有酶活性(例如,内切核酸酶活性)的限制酶的结构域可与另一蛋白质如 DNA 结合蛋白质融合。DNA 结合蛋白质可将杂合蛋白靶向至 DNA 上的特定序列。具有酶活性的结构域的核酸切割活性可以是序列特异性的或非序列特异性的。例如, IIs 型限制性内切核酸酶 FokI 的非特异性切割结构域可用作杂合核酸酶的酶促(切割)结构域。具有酶活性的结构域可切割的序列可受限于杂合体经由 DNA 结合域与 DNA 的物理束缚。DNA 结合域可来自真核或原核转录因子。DNA 结合域可以识别连续核酸序列的大约或至少约 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24 或 25 个碱基对。在一些情况下,限制酶为 4- 碱基切割酶(cutter)、6- 碱基切割酶或 8- 碱基切割酶。DNA 结合域可识别序列的约 9 到约 18 个碱基对。DNA 结合域可为,例如,锌指 DNA 结合域。杂合体可为锌指核酸酶(例如,锌指核酸酶)。杂合蛋白质可作为多聚体(例如,二聚体、三聚体、四聚体、五聚体、六聚体等)起作用。

[0252] 可在本文所述的方法、组合物和 / 或试剂盒中使用的特定限制酶的实例包括 AaaI、AagI、AarI、AasI、AatI、AatII、AauI、AbaI、AbeI、AbrI、AccI、AccII、AccIII、Acc16I、Acc36I、Acc65I、Acc113I、AccB1I、AccB2I、AccB7I、AccBSI、AccEBI、AceI、AceII、AceIII、AciI、AclI、AclNI、AclWI、AcpI、AcpII、AcrII、AcsI、AcuI、AcvI、AcyI、AdeI、AeuI、AfaI、Afa22MI、Afa16RI、AfeI、AflI、AflIII、AflIII、AgeI、AgeI-HF、AglI、AhaI、AhaII、AhaIII、AhaB8I、AhdI、AhlI、AhyI、AitI、AjnI、AjoI、AleI、AlfI、AliI、AliAJI、AloI、AluI、AlwI、Alw21I、Alw26I、Alw44I、AlwNI、AlwXI、Ama87I、AcoI、AocII、AorI、Aor13HI、Aor51HI、AosI、AosII、ApaI、ApaB1、ApaCI、ApaLI、ApaORI、ApeKI、ApiI、ApoI、ApyI、AquI、AscI、AseI、AseIII、AsiSI、AvaI、AvaII、AvrII、BaeGI、BaeI、BamHI、BamHI-HF、BanI、BanII、BbsI、BbvCI、BbvI、BccI、BceAi、BcgI、BciVI、BclI、BcoDI、BfaI、BfuAI、BfuCI、BglI、BglII、BlpI、BmgBI、BmrI、BmtI、BpmI、Bpu10I、BpuEI、BsaAI、BsaBI、BsaHI、BsaI、BsaI-HF、BsaJI、BsaWI、BsaXI、BseRI、BseYI、BsgI、BsiEI、BsiHKAI、BsiWI、BslI、BsmAI、BSmBI、BsmFI、BsmI、BsoBI、Bsp1286I、BspCNI、BspDI、BspEI、BspHI、BspMI、BspQI、BsrBI、BsrDI、BsrFI、BsrGI、BsrI、BssHII、BssKI、BssSI、BstAPI、BstBI、BsteII、BstNI、BstUI、BstXI、BstYI、BstZ17I、Bsu36I、BtgI、BtgZI、BtsCI、BtsI、BtsIMutI、Cac8I、ClaI、CspCI、CviAI、CviKI-1、CviQI、DdeI、DpnI、DpnII、DraI、DraIII、DraIII-HFTM、DrdI、EaeI、EagI、EagI-HFTM、EarI、EciI、Eco53kI、EcoNI、Eco0109I、EcoP15I、EcoRI、EcoRI-HFTM、EcoRV、EcoRV-HFTM、FatI、FauI、Fnu4HI、FokI、FseI、FspEI、FspI、HaeII、HaeIII、HgaI、HhaI、HincII、HindIII、HindIII-HFTM、HinfI、HinP1I、HpaI、HpaII、HphI、Hpy166II、Hpy188I、Hpy188III、Hpy99I、HpyAV、HpyCH4III、HpyCH4IV、HpyCH4V、I-CeuI、I-SceI、KasI、KpnI、KpnI-HFTM、LpnPI、MboI、MboII、MfeI、MfeI-HFTM、MluCI、MluI、MlyI、MmeI、MnlI、MscI、MseI、MslI、MspA1I、MspI、MspJI、MwoI、NaeI、NarI、Nb. BbvCI、Nb. BsmI、Nb. BsrDI、Nb. BtsI、NciI、NcoI、NcoI-HFTM、NdeI、NgoMIV、NheI、NheI-HFTM、NlaIII、NlaIV、NmeAIII、NotI、NotI-HFTM、NruI、NsiI、

NspI、Nt. AlwI、Nt. BbvCI、Nt. BsmAI、Nt. BspQI、Nt. BstNBI、Nt. CviPII、PacI、PaeR7I、PciI、PflFI、PflMI、PhoI、PI-PspI、PI-SceI、PleI、PmeI、PmlI、PpuMI、PshAI、PsiI、PspGI、PspOMI、PspXI、PstI、PstI-HFTM、PvuI、PvuI-HFTM、PvuII、PvuII-HFTM、RsaI、RsrII、SacI、SacI-HFTM、SacII、SalI、SalI-HFTM、SapI、Sau3AI、Sau96I、SbfI、SbfI-HFTM、ScaI、ScaI-HFTM、ScrFI、SexAI、SfaNI、SfcI、SfiI、SfoI、SgrAI、SmaI、SmlI、SnaBI、SpeI、SphI、SphI-HFTM、SspI、SspI-HFTM、StuI、StyD4I、StyI、StyI-HFTM、SwaI、Taq α I、TfiI、TliI、TseI、Tsp45I、Tsp509I、TspMI、TspRI、Tth111I、XbaI、XcmI、XhoI、XmaI、XmnI 和 ZraI。

[0253] 在本文所述的方法、组合物和 / 或试剂盒中使用的一种或多种限制酶可来源于多种来源。例如，一种或多种限制酶可由重组核酸产生。一种或多种限制酶可由异源宿主（例如，细菌、酵母、昆虫或哺乳动物细胞）中的重组核酸产生。一种或多种限制酶可由异源宿主中的重组核酸产生并从异源宿主中纯化。一种或多种限制酶可从天然宿主如细菌或古细菌纯化而来。如果使用了超过一种限制酶，那么至少一种限制酶可来自重组来源，且超过一种限制酶中的至少一种可来自天然来源。

[0254] 一种或多种限制酶的识别位点可为多种序列中的任何一种。例如，一种或多种限制酶的识别位点可为回文序列。一种或多种限制酶的识别位点可为部分回文序列。在一些实施方案中，一种或多种限制酶的识别位点不是回文序列。一种或多种限制酶的识别位点可为大约或大于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个碱基或碱基对。限制酶的识别位点可为约 2 到约 20、约 5 到约 20、约 5 到约 15、约 5 到约 10、约 7 到约 20、约 7 到约 15 或约 7 到约 10 个碱基或碱基对。

[0255] 可用两种或更多种限制酶来消化多核苷酸。该两种或更多种限制酶可识别相同或不同的识别位点。在单一多核苷酸上的两个靶核酸序列之间可存在单一限制酶的一个或多个识别位点。在单一多核苷酸上的两个靶核酸序列之间可存在单一限制酶的大约或至少约 1、2、3、4、5、6、7、8、9、10 个或更多个识别位点。在单一多核苷酸上的两个靶核酸序列之间可存在两个或更多个不同的限制酶识别位点。在单一多核苷酸上的两个靶核酸序列之间可存在大约或至少约 1、2、3、4、5、6、7、8、9、10 个或更多个不同的限制酶识别位点。在单一多核苷酸的两个靶核酸序列之间可存在一个或多个不同的限制酶限制酶切位点。在单一多核苷酸上的两个靶核酸序列之间可存在大约或至少约 1、2、3、4、5、6、7、8、9、10 个或更多个限制酶限制酶切位点。

[0256] 限制酶消化可包含一种或多种同裂酶 (isoschizomer)。同裂酶是识别相同序列的限制性内切核酸酶。同裂酶可具有不同的切割位点；这些酶称为异裂酶 (neoschizomer)。

[0257] 在一些实施方案中，用限制酶切割产生平端。在一些实施方案中，用限制酶切割不产生平端。在一些实施方案中，用限制酶切割产生两个片段，每个片段具有一个 5' 突出端。在一些实施方案中，用限制酶切割产生两个片段，每个片段具有一个 3' 突出端。

[0258] 可设计用于一种或多种扩增反应的引物以扩增限制酶切割位点的上游和下游序列。

[0259] 在一个实施方案中，限制酶不切割靶核酸序列或参考扩增子。可使用参考序列，例如，基因组序列，来预测限制酶是否会切割核酸序列。在另一实施方案中，限制酶确实切割靶核酸序列。切割可发生在靶序列的 5' 或 3' 端附近（在约 5、10、15、25、50 或 100bp 内）、靶序列内。

[0260] 在另一实施方案中,限制酶不切割靶核酸序列或参考核酸序列或扩增子,即使该序列或扩增子含有一个或多个 SNP。SNP 信息可从几个数据库获得,最容易从 dbSNP(www.ncbi.nlm.nih.gov/projects/SNP/) 获得。

[0261] 一种或多种甲基化敏感的限制酶可用于本文提供的方法、组合物和试剂盒中。所述一种或多种甲基化敏感酶可包括,例如,DpnI、Acc65I、KpnI、ApaI、Bsp120I、Bsp143I、MboI、Bsp0I、NheI、Cfr9I、SmaI、Csp6I、RsaI、Ecl136II、SacI、EcoRII、MvaI、HpaII 或 MspI。在一个实施方案中,甲基化敏感的限制酶可切割未甲基化的核酸,但不能切割已甲基化的核酸。

[0262] 可对本发明中使用的限制酶进行选择,从而特异性地消化核酸序列的选定区域。在一个实施方案中,所述一种或多种限制酶在靶核酸序列或靶扩增子之间进行切割。可选择其识别序列在靶核酸序列或靶扩增子附近出现例如一次或多次的一种或多种酶。在一个实施方案中,需要注意要确保这些识别序列不受 SNP 的存在的影响。

[0263] 在一个实施方案中,限制酶是有效的但特异性(无星号活性)的切割酶。可通过进行适当的酶滴定实验预先确定这种性质与消化时间和酶浓度。在一个实施方案中,限制酶可以具有星号活性。星号活性可以是与限定的识别序列相似但不相同的序列的切割。

[0264] 限制酶的“单位”数与核酸(例如,DNA 或 RNA)量的比值可以是,例如,大约,或至少约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、300、350、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、12,000、14,000、15000、16,000、18,000 或 20,000 单位/ μ g 核酸。限制酶的单位数与核酸量的比值可为约 1 到约 20,000、约 1 到约 10,000、约 1 到约 5,000、约 100 到约 10,000、约 100 到约 1,000、约 50 到约 500 或约 50 到约 250 单位/ μ g。

[0265] 一种或多种限制酶可与包含多核苷酸的样品孵育大约或超过约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59 或 60 分钟。一种或多种限制酶可与包含多核苷酸的样品孵育大约或超过约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47 或 48 小时。一种或多种限制酶可与包含多核苷酸的样品孵育约 1min 到约 60min、约 1min 到约 48hr、约 1min 到约 24hr、约 1min 到约 20hr、约 1min 到约 16hr、约 0.5hr 到约 6hr、约 0.5hr 到约 3hr、约 1hr 到约 10hr、约 1hr 到约 5hr 或约 1hr 到约 3hr。

[0266] 限制酶消化可在大约或高于约 5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、

46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64 或 65℃ 的温度下进行。限制酶消化可在约 10 到约 65℃、约 20 到约 65℃、约 30 到约 65℃、约 37 到约 65℃、约 40 到约 65℃、约 50 到约 65℃、约 25 到约 37℃ 或约 30 到约 37℃ 的温度下进行。

[0267] 使用一种或多种限制酶的限制酶消化的 pH 可以是约 2、2.5、3、3.5、4、4.5、5、5.5、6、6.5、6.6、6.7、6.8、6.9、7、7.1、7.2、7.3、7.4、7.5、7.5、7.6、7.7、7.8、7.9、8、8.1、8.2、8.3、8.4、8.5、8.6、8.7、8.8、8.9、9、9.5、10、10.5、11、11.5、12 或 12.5。限制酶消化的 pH 可以是约 5 到约 9、约 5 到约 8、约 5 到约 7、约 6 到约 9 或约 6 到约 8。

[0268] 限制酶消化可包含一种或多种缓冲液。所述一种或多种缓冲溶液可为,例如, tris-HCl、双-tris-丙烷-HCl、TAP、N-二甘氨酸、tris、tris-乙酸盐、tris-HCl、三甲基甘氨酸、TAPSO、HEPES、TES、MOPS、PIPES、甲次膦酸盐、SSC、磷酸盐缓冲液、可力丁、乙酸佛罗罗拿、MES、ADA、ACES、氯化胆胺、乙酰氨基甘氨酸、甘氨酸、马来酸盐、CABS、哌啶、甘氨酸、柠檬酸盐、甘氨酸甘氨酸、苹果酸盐、甲酸盐、琥珀酸盐、乙酸盐、丙酸盐、吡啶、哌嗪、组氨酸、双-tris、乙醇胺、碳酸盐、MOPSO、咪唑、双-TRIS 丙烷、BES、MOBS、三乙醇胺 (TEA)、HEPPSO、POPSO、胍、氨基丁三醇 (tris)、EPPS、HEPPS、N-二甘氨酸、HEPBS、AMPSO、牛磺酸 (AES)、硼酸盐、CHES、2-氨基-2-甲基-1-丙醇 (AMP)、氢氧化铵或甲胺。溶液中缓冲液的浓度可以是,例如,约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或 100mM。溶液中缓冲液的浓度可以是约 10 到约 100mM、约 10 到约 75mM、约 25 到约 75mM 或约 10 到约 50mM。

[0269] 使用一种或多种限制酶的限制酶消化可包含牛血清白蛋白 (BSA)。限制酶切消化中 BSA 的浓度可以是大约或大于约 0.01、0.02、0.03、0.04、0.05、0.06、0.07、0.08、0.09、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、1.5、2、3、4、5、6、7、8、9 或 10mg/ml。限制酶切消化中 BSA 的浓度可以是约 0.01 到约 10mg/ml、约 0.01 到约 1mg/ml、约 0.05 到约 1mg/ml 或约 0.05 到约 0.5mg/ml。

[0270] 使用一种或多种限制酶的限制酶消化可包含甘油。甘油的浓度 (体积 / 体积) 可以是大约或至少约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24 或 25%。限制酶消化中甘油的浓度可以是约 1 到约 25%、约 1 到约 20%、约 1 到约 15%、约 1 到约 10% 或约 1 到约 5%。

[0271] 限制酶消化可包含一种或多种有机溶剂,例如, DMSO、乙醇、乙二醇、二甲基乙酰胺、二甲基甲酰胺或 suphalane。限制酶消化可不含一种或多种有机溶剂。

[0272] 限制酶消化可包含一种或多种二价阳离子。该一种或多种二价阳离子可以是,例如, Mg^{2+} 、 Mn^{2+} 、 Cu^{2+} 、 Co^{2+} 或 Zn^{2+} 。

[0273] 限制酶切消化可包含一种或多种盐。该一种或多种盐可包括,例如, 乙酸钾、氯化钾、乙酸镁、氯化镁、乙酸钠或氯化钠。一种或多种盐中每一种的浓度可以是,例如, 大约或大于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、

77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245 或 250mM。一种或多种盐中每一种的浓度可以是约 5 到约 250、约 5 到约 200、约 5 到约 150、约 5 到约 100、约 10 到约 100、约 10 到约 90、约 10 到约 80、约 10 到约 70、约 10 到约 60 或约 10 到约 50mM。

[0274] 限制酶切消化可以包含一种或多种还原剂。该一种或多种还原剂可以抑制蛋白质中二硫键的形成。还原剂可为,例如,二硫苏糖醇 (DTT)、2- 巯基乙醇 (BME)、2- 巯基乙胺 -HCl、三 (2- 羧乙基) 膦 (TCEP) 或半胱氨酸 -HCl。在限制酶消化中一种或多种还原剂的浓度可以是大约或大于约 0.01、0.02、0.03、0.04、0.05、0.06、0.07、0.08、0.09、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、1.1、1.2、1.3、1.4、1.5、1.6、1.7、1.8、1.9、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 或 25mM。在限制酶消化中一种或多种还原剂的浓度可以是约 0.01 到约 25mM、约 0.01 到约 15mM、约 0.01 到约 10mM、约 0.01 到约 5mM、约 0.1 到约 5mM 或约 0.5 到约 2.5mM。

[0275] 在核酸的限制酶消化中可以使用多于一种限制酶。例如,如果一种或多种限制酶不能有效切割核酸,或者如果它们不能全部很好地普遍作用于所有样品(例如,由于 SNP),那么可采用多重消化。在限制酶切消化中可使用的不同限制酶的数目可以是大约或至少约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20。在限制酶消化中可使用的不同限制酶的数目可以是,例如,约 1 到约 20、约 1 到约 15、约 1 到约 10、约 1 到约 7、约 1 到约 6、约 1 到约 5、约 1 到约 4、约 1 到约 3 或约 1 到约 2。

[0276] 有一些证据表明,当含有扩增子或靶序列的片段的大小相对较小时,PCR 表现更好。因此,选择切割位点在扩增子或靶序列附近的限制酶可能是理想的。例如,限制酶识别位点或切割位点可以在距多核苷酸上一个靶序列的 5' 端或 3' 端大约或小于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、300、350、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、12,000、14,000、15000、16,000、18,000 或 20,000 个碱基对之内。限制酶识别位点或切割位点可以在距靶核酸序列的 5' 端或 3' 端约 1 到约 10,000、约 1 到约 5,000、约 1 到约 2,500、约 1 到约 1,000、约 1 到约 100、约 100 到约 1000、约 100 到约 500 或约 100 到约 250bp 之内。

[0277] 可分析单一样品的多个 CNV。在这种情况下,选择对整套 CNV 有效的最小消化数可能是理想的。在一个实施方案中,可发现单一限制酶混合物在任何扩增子或靶核酸序列内都不切割,但在其中每一个附近都有识别位点或切割位点。限制酶识别位点或切割位点可

以在距多核苷酸上一个靶序列的 5' 端或 3' 端大约或小于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、101、102、103、104、105、106、107、108、109、110、111、112、113、114、115、116、117、118、119、120、121、122、123、124、125、126、127、128、129、130、131、132、133、134、135、136、137、138、139、140、141、142、143、144、145、146、147、148、149、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、300、350、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、12,000、14,000、15000、16,000、18,000 或 20,000 个碱基对之内。限制酶识别位点或切割位点可以在距多核苷酸上一个靶序列的 5' 端或 3' 端约 1 到约 20,000、约 10 到约 20,000、约 100 到约 20,000、约 1000 到约 20,000、约 10 到约 10,000、约 10 到约 1000、约 10 到约 100、约 50 到约 20,000、约 50 到约 1000、约 50 到约 500、约 50 到约 250、约 50 到约 150 或约 50 到 100 个碱基对之内。

[0278] 可以编写和 / 或使用合适的软件,从而使限制酶选择的过程自动化,并为用户如实验生物学家呈现界面,以考虑以上标准选择最合适的酶。软件可采用其它考虑因素,如酶的成本、酶的效率、限制酶的缓冲液相容性、甲基化敏感性、核酸区段中切割位点的数目或实用性。软件可在计算机上使用。可在计算机可读介质上生成算法,并用该算法选择用于消化核酸的一种或多种限制酶。计算机可与因特网连接,并且可用于访问能允许选择限制性内切核酸酶的网站。可使用网络工具来选择将在扩增子周围切割的限制酶,以便分离相连的基因拷贝以供 CNV 估计。例如,可将酶和分析存储于数据库中,并且限制酶的选择可以是自动化的。可以考虑的其它统计学包括,例如,最短片段的长度、%GC 含量、在扩增子周围 (或在其内) 切割的频率和酶的成本。可使用 QTools 来辅助一种或多种限制酶的选择。图 13 和图 14 图示了选择限制酶时可考虑的信息。

[0279] 对于用于数据分析的分析储存,研究者可按照位置或引物序列输入分析。QTools 可自动检索和储存扩增子序列及已知的 SNP,并计算热力学参数。随着研究者使用更多的分析,他们可以输入额外的数据,包括已确认的样品的 CNV 和退火温度。

[0280] 在单一试管中使用多于一种酶连续或一起进行的消化可以帮助确保对困难靶标的完全切割。一个样品的一系列限制酶消化可用不同的酶进行,例如约 1、2、3、4、5、6、7、8、9 或 10 种酶。

[0281] 在限制酶消化之后,可将消化物中的一种或多种限制酶灭活。在一些实施方案中,暴露于热不能将所述一种或多种限制酶灭活。在限制之后,通过升高限制反应的温度,可以将大多数限制酶热灭活。热灭活温度可以是,例如,大约或高于约 50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或 100 摄氏度。热灭活温度可以是约 50 到约 100、约 50 到约 90、约 60 到约 90、约 65 到约 90、约 65 到约 85 或约 65 到约 80 摄氏度。热灭活持续时间可以是,例如,大约或大于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、65、70、

80、90、100、110、120、180、240 或 300 分钟。热灭活持续时间可以是约 5 到约 300、约 5 到约 200、约 5 到约 150、约 5 到约 100、约 5 到约 75、约 5 到约 50、约 5 到约 40、约 5 到约 30、约 5 到约 35、约 5 到约 25、约 5 到约 20 或约 10 到约 20 分钟。热灭活温度可低于限制酶切靶片段的解链温度，以便保持双链模板拷贝。

[0282] 可通过向限制酶消化物中加入一种或多种螯合剂来停止限制酶消化。所述一种或多种螯合剂可以是，例如，EDTA、EGTA、柠檬酸或磷酸盐。限制酶消化物中的一种或多种螯合剂的浓度可以是，例如，大约或至少约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或 100mM。所述一种或多种螯合剂的浓度可以是约 1 到约 100mM、约 1 到约 75mM 或约 25 到约 75mM。

[0283] 可使用对照试验和模板来测量限制酶消化步骤的效率。

[0284] 样品

[0285] 将要用本文提供的方法、组合物和试剂盒分析的样品，可来自于包含核酸的非细胞实体（如，病毒）或来自于基于细胞的生物（如，古细菌、细菌或真核生物域的成员）。在一些情况下，样品可从医院、实验室、临床或医学实验室获得。样品可包含核酸，例如，RNA 或 DNA。样品可包含无细胞核酸。在一些情况下，样品从表面如门或工作台顶部的拭子获得。

[0286] 样品可来自受试者，例如，植物、真菌、真细菌、古细菌、原生生物 (protest) 或动物。受试者可以是生物体，单细胞生物体或多细胞生物体。受试者可以是培养的细胞，该细胞尤其可以是原代细胞或者来自自己确立的细胞系的细胞。样品最初可以任何适当形式从多细胞生物分离。动物可以是鱼，如斑马鱼。动物可以是哺乳动物。哺乳动物可以是，例如，狗、猫、马、牛、小鼠、大鼠或猪。哺乳动物可以是灵长类动物，如人、黑猩猩、猩猩或大猩猩。人可以是男性或女性。样品可来自人类胚胎或人类胎儿。人可以是婴儿、儿童、青少年、成年人或老年人。女性可以是怀孕的、可以是疑似怀孕的或计划怀孕的。

[0287] 样品可以来自健康受试者（例如，人类受试者）。在一些实施方案中，样品取自妊娠至少 4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25 或 2 周的受试者（例如，待产孕妇）。在一些实施方案中，受试者受遗传疾病的影响，是遗传疾病的携带者，或者有遗传或发展遗传疾病的危险，其中遗传疾病是任何可能与诸如突变、插入、添加、缺失、易位、点突变、三核苷酸重复疾病和 / 或单核苷酸多态性 (SNP) 等遗传变异相关的疾病。在其它实施方案中，样品取自生育年龄的女性患者，并且，在一些情况下，女性患者未怀孕或怀孕状态未知。在另外其它的情况下，受试者为男性患者、男性准父亲或处于特定遗传异常危险中、诊断有或患有该遗传异常的男性患者。在一些情况下，已知女性患者受遗传疾病或遗传变异的影响，或者是其携带者，或处于特定遗传异常的危险中、诊断有或患有该遗传异常。在一些情况下，女性患者有关遗传疾病或遗传变异的状态可能是未知的。在进一步的实施方案中，样品取自基因序列拷贝数变异状态已知或未知的任何儿童或成人患者。在一些情况下，已知儿童或成人患者受遗传疾病或遗传变异的影响，或者是其携带者。

[0288] 样品可以取自患有特定疾病、病症或病状的受试者，或疑似患有特定疾病、病症或病状（或处于患病危险中）的受试者。例如，样品可来自癌症患者，疑似患有癌症的患者，

或者处于患癌症危险中的患者。癌症可以是,例如,急性成淋巴细胞性白血病(ALL)、急性髓性白血病(AML)、肾上腺皮质癌、卡波西肉瘤、肛门癌、基底细胞癌、胆管癌、膀胱癌、骨癌、骨肉瘤、恶性纤维组织细胞瘤、脑干胶质瘤、脑癌、颅咽管瘤、室管膜母细胞瘤、室管膜瘤、髓母细胞瘤、髓上皮瘤(medullopithelioma)、松果体实质瘤、乳腺癌、支气管癌、伯基特淋巴瘤、非霍奇金淋巴瘤、类癌瘤、宫颈癌、脊索瘤、慢性淋巴细胞白血病(CLL)、慢性髓性白血病(CML)、结肠癌、结直肠癌、皮肤T-细胞淋巴瘤、原位导管癌、子宫内膜癌、食管癌、尤因肉瘤、眼癌、眼内黑素瘤、视网膜母细胞瘤、纤维性组织细胞瘤、胆囊癌、胃癌、神经胶质瘤、毛细胞白血病、头颈癌、心脏癌、肝细胞(肝)癌、霍奇金淋巴瘤、下咽癌、肾癌、喉癌、唇癌、口腔癌、肺癌、非小细胞癌、小细胞癌、黑素瘤、口腔癌、骨髓增生异常综合征、多发性骨髓瘤、髓母细胞瘤、鼻腔癌、鼻窦癌、神经母细胞瘤、鼻咽癌、口腔癌、口咽癌、骨肉瘤、卵巢癌、胰腺癌、乳头瘤病、副神经节瘤、甲状旁腺癌、阴茎癌、咽癌、垂体肿瘤、浆细胞肿瘤、前列腺癌、直肠癌、肾细胞癌、横纹肌肉瘤、唾液腺癌、赛泽里综合征、皮肤癌、非黑素瘤、小肠癌、软组织肉瘤、鳞状细胞癌、睾丸癌、喉癌、胸腺瘤、甲状腺癌、尿道癌、子宫癌、子宫肉瘤、阴道癌、外阴癌、瓦尔登斯特伦巨球蛋白血症或肾母细胞瘤。样品可来自癌症患者的癌组织和/或正常组织。

[0289] 样品可来自已知患有遗传疾病、病症或病状的受试者。在一些情况下,已知受试者的基因,或基因的一部分,例如CFTR、因子VIII(F8基因)、 β 珠蛋白、血色病、G6PD、神经纤维瘤病、GAPDH、 β 淀粉样蛋白或丙酮酸激酶基因为野生型或基因突变型。在一些情况下,受试者的状态为已知或未知的,并检测受试者基因如CFTR、因子VIII(F8基因)、 β 珠蛋白、血色病、G6PD、神经纤维瘤病、GAPDH、 β 淀粉样蛋白或丙酮酸激酶基因的突变或遗传变异的存在。

[0290] 样品可以是房水、玻璃体液、胆汁、全血、血清、血浆、乳汁、脑脊液、眵、内淋巴(enolymph)、外淋巴、胃液、粘液、腹膜液、唾液、皮脂、精液、汗液、眼泪、阴道分泌物、呕吐物、粪便或尿。样品可从医院、实验室、临床或医学实验室获得。样品可取自受试者。样品可包含核酸。核酸可以是,例如,线粒体DNA、基因组DNA、mRNA、siRNA、miRNA、cRNA、单链DNA、双链DNA、单链RNA、双链RNA、tRNA、rRNA或cDNA。样品可包含无细胞核酸。样品可为细胞系、基因组DNA、无细胞血浆、福尔马林固定石蜡包埋的(FFPE)样品或急骤冻结的样品。福尔马林固定石蜡包埋的样品可在提取核酸之前脱石蜡。样品可来自器官,例如,心脏、皮肤、肝脏、肺、乳腺、胃、胰腺、膀胱、结肠、胆囊、脑等。

[0291] 当核酸为RNA时, RNA的来源可为本文所述的任何来源。例如, RNA可以是无细胞mRNA, 可以来自组织活检、中心活检、细针抽出物、急骤冻结的或福尔马林固定石蜡包埋的(FFPE)样品。FFPE样品可在提取RNA之前脱石蜡。分析前可将提取的RNA加热到约30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98或99 $^{\circ}$ C。可将提取的RNA加热到任何前述温度,持续大约或至少约15min、30min、45min、60min、1.5hr、2hr、2.5hr、3hr、3.5hr、4hr、4.5hr、5hr、5.5hr、6hr、6.5hr、7hr、7.5hr、8hr、8.5hr、9hr、9.5hr或10hr。

[0292] RNA可用于多种下游应用。例如,可利用逆转录酶将RNA转化成cDNA,且cDNA可任选地经受PCR,如实时PCR。可在等温扩增反应如等温线性扩增反应中使用RNA或cDNA。

RNA、产生的 cDNA 或由此扩增的分子可用于微阵列实验、基因表达实验、Northern 分析、Southern 分析、测序反应、下一代测序反应等。可对特定 RNA 序列进行分析,或可对 RNA 序列进行全面分析。

[0293] 可通过本领域普通技术人员可利用的手段从样品中提取核酸。

[0294] 可处理样品以使其能够扩增。示例性的样品处理可包括,裂解样品的细胞以释放核酸,纯化样品(例如,将核酸与可能抑制扩增的其它样品组分分离),稀释/浓缩样品,和/或将样品与用于扩增的试剂结合,该试剂如 DNA/RNA 聚合酶(例如,用于 PCR 扩增的热稳定 DNA 聚合酶)、dNTPs(例如, dATP、dCTP、dGTP 和 dTTP(和/或 dUTP))、用于每个待扩增的等位基因序列或多态基因座的引物组、能够与每个待扩增的等位基因序列特异性杂交的探针(例如荧光探针,尤其如 TAQMAN 探针或分子信标探针)、 Mg^{2+} 、DMSO、BSA、缓冲液或其任意组合。在一些实例中,样品可以与限制酶、尿嘧啶 -DNA 糖苷酶(UNG)、逆转录酶或核酸处理的任何其它酶结合。

[0295] 靶多核苷酸

[0296] 本文所述的方法可用来分析或检测一种或多种靶核酸分子。术语多核苷酸,或其语法等同物,可以指共价连接在一起的至少两个核苷酸。本文所述的核酸可以含有磷酸二酯键,尽管在如下所示的一些情况中(例如在引物和探针如标记探针的构建中),包括可能具有交替骨架的核酸类似物,包含,例如,磷酰胺(Beaucage 等人, *Tetrahedron*49(10):1925(1993) 和其中的参考文献; Letsinger, J. *Org. Chem.* 35:3800(1970); Sprinzl 等人, *Eur. J. Biochem.* 81:579(1977); Letsinger 等人, *Nucl. Acids Res.* 14:3487(1986); Sawai 等人, *Chem. Lett.* 805(1984); Letsinger 等人, *J. Am. Chem. Soc.* 110:4470(1988); 和 Pauwels 等人, *Chemica Scripta*26:1419(1986))、硫代磷酸酯(Mag 等人, *Nucleic Acids Res.* 19:1437(1991); 和美国专利 5,644,048)、二硫代磷酸酯(Briu 等人, *J. Am. Chem. Soc.* 111:2321(1989))、O-甲基亚磷酰胺连接(参见 Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press)和肽核酸(此处又称“PNA”)骨架和连接(参见 Egholm, *J. Am. Chem. Soc.* 114:1895(1992); Meier 等人, *Chem. Int. Ed. Engl.* 31:1008(1992); Nielsen, *Nature*, 365:566(1993); Carlsson 等人, *Nature*380:207(1996), 均通过引用并入)。其它核酸类似物包括具有双环结构的核酸类似物,包括锁定核酸(此处又称“LNA”), Koshkin 等人, *J. Am. Chem. Soc.* 120. 132523(1998); 阳性骨架(positive backbone)(Denpcy 等人, *Proc. Natl. Acad. Sci. USA*92:6097(1995); 非离子骨架(美国专利 5,386,023、5,637,684、5,602,240、5,216,141 和 4,469,863; Kiedrowski 等人, *Angew. Chem. Intl. Ed. English*30:423(1991); Letsinger 等人, *J. Am. Chem. Soc.* 110:4470(1988); Letsinger 等人, *Nucleoside&Nucleotide*13:1597(1994); ASC Symposium Series580, 第2章和第3章,“Carbohydrate Modifications in Antisense Research”, Y. S. Sanghui 和 P. Dan Cook 编; Mesmaeker 等人, *Bioorganic&Medicinal Chem. Lett.* 4:395(1994); Jeffs 等人, *J. Biomolecular NMR*34:17(1994); *Tetrahedron Lett.* 37:743(1996))和无核糖骨架,包括美国专利 5,235,033 和 5,034,506 以及 ASC Symposium Series580, 第6章和第7章,“Carbohydrate Modifications in Antisense Research”, Ed. Y. S. Sanghui 和 P. Dan Cook 中描述的那些骨架。包含一个或多个环碳糖的核酸也包括在核酸的定义内(参见

Jenkins 等人, Chem. Soc. Rev. (1995) 169-176 页)。在 Rawls, C&E News, 1997 年 6 月 2 日, 第 35 页中描述了几种核酸类似物。“锁定核酸”也包含在核酸类似物的定义中。LANs 是其中核糖环由连接 2' - 氧原子和 4' - 碳原子的亚甲基桥“锁定”的一类核酸类似物。所有这些参考文献均通过引用特定地并入本文。可对核糖 - 磷酸骨架进行这些修饰以增加这类分子在生理环境下的稳定性和半衰期。例如, PNA :DNA 和 LNA-DNA 杂合体可显示出更高的稳定性, 并因此可以在一些实施方案中使用。如所说明的, 靶核酸可以是单链的或双链的, 或者同时含有双链或单链序列的部分。取决于应用, 核酸可以是 DNA (包括, 例如, 基因组 DNA、线粒体 DNA 和 cDNA)、RNA (包括, 例如 mRNA 和 rRNA) 或杂合体, 其中核酸包含脱氧核糖核苷酸和核糖核苷酸的任意组合, 以及包括尿嘧啶、腺嘌呤、胸腺嘧啶、胞嘧啶、鸟嘌呤、肌苷、黄嘌呤 (xanthine)、次黄嘌呤 (hypoxanthine)、异胞嘧啶、异鸟嘌呤等在内的碱基的任意组合。

[0297] 本文提供的方法和组合物可用于评估多核苷酸 (例如, DNA、RNA、线粒体 DNA、基因组 DNA、mRNA、siRNA、miRNA、cRNA、单链 DNA、双链 DNA、单链 RNA、双链 RNA、tRNA、rRNA、cDNA 等) 的量。所述方法和组合物可用来评估与第二多核苷酸的量相比的第一多核苷酸的量。所述方法可用来分析溶液中合成质粒的量; 检测从受试者或环境中获得的样品中的致病生物 (例如, 微生物、细菌、病毒、寄生虫、逆转录病毒、慢病毒、HIV-1、HIV-2、流感病毒等)。所述方法也可用于在较大的多核苷酸群体中存在罕见的多核苷酸群体的其它应用中。

[0298] 在使用本文提供的方法、组合物和试剂盒所分析的受试者的样品 (如基因组) 中, 靶核酸序列的拷贝数可以是 0, 或者大约或至少约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、5000、10,000、20,000、50,000 或 100,000。在使用本文提供的方法、组合物和试剂盒分析其样品的受试者的基因组中, 靶核酸序列的拷贝数可以是约 1 到约 20、约 1 到约 15、约 1 到约 10、约 1 到约 7、约 1 到约 5、约 1 到约 3、约 1 到约 1000、约 1 到约 500、约 1 到约 250、约 1 到约 100、约 10 到约 1000、约 10 到约 500、约 10 到约 250、约 10 到约 100、约 10 到约 50、约 10 到约 20、约 0 到约 100、约 0 到约 50、约 0 到约 25 或约 0 到约 10。

[0299] 靶核酸序列可在一条染色体上。如果靶核酸在来源于人类受试者的样品中, 那么靶核酸序列可在染色体 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、X 或 Y 中的一条或多条上。靶核酸可以在大约或多于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22 或 23 条染色体上。靶核酸序列的两个或更多个拷贝可以在相同或不同染色体上。在人类受试者中, 靶核酸序列的两个或更多个拷贝可在染色体 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、X 或 Y 上。靶核酸序列的两个或更多个拷贝可以在受试者的一条多核苷酸 (例如, 染色体) 上, 但由于样品处理 (例如, 通过断裂) 可将靶核酸在取自受试者的样品中进行分离。

[0300] 当靶核酸的两个拷贝在同一多核苷酸如同一条染色体上时, 两个拷贝在多核苷酸上可相距大约、大于约或小于约 1、2、3、4、5、6、7、8、9、10、25、50、75、100、200、300、400、500、

600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1 百万、2 百万、3 百万、4 百万、5 百万、6 百万、7 百万、8 百万、9 百万、1 千万、2 千万、3 千万、4 千万、5 千万、6 千万、7 千万、8 千万、9 千万或 1 亿个碱基或碱基对。两个靶核酸可相距约 100 到约 100,000、约 100 到约 10,000、约 100 到约 1,000、约 10 到约 10,000 或约 10 到约 1,000 个碱基或碱基对。

[0301] 靶序列可以是基因。例如,基因可以是 ERBB2、EGFR、BRCA1、BRCA2、APC、MSH2、MSH6、MLH1、CYP2D6、富含低拷贝重复 (LCR) 的序列 (参见例如 Balikova 等人 (2008) *Am J. Hum Genet.* 82:181-187)、TAS1R1、GNAT1、IMPDH1、OPN1SW、OR2A12、OR2A14、OR2A2、OR2A25、OR2A5、OR2A1、OR2A42、OR2A7、OR4F21、OR4F29、OR4C6、OR4P4、OR4S2、OR5D13、ROM1、TASOR14、TAS2R44、TAS2R48、TAS2R49、TAS2R50、OR6C2、OR6C4、OR6C68、OR6C70、OR4M1、OR4Q3、OR4K1、OR4K2、OR4K5、OR4N2、OR4K13、OR4K14、OR4K15、OR4M2、OR4N4、OR1F1、ACTG1、FSCN2、OR2Z1、OR11H1、MYH9、SKI、TP73、TNFRSF25、RAB3B、VAV3、RALB、BOK、NAT6、TUSC2、TUSC4、TAB3B、C6orf210、ESR1、MAFK、MAD1L1、MYC、VAV2、MAP3K8、CDKN1C、WT1、WIT-1、C1QTNF4、MEN1、CCND1、ORA0V1、MLL2、C13orf10、TNFAIP2、AXIN1、BCAR1、TAX1BP3、NF1、PHB、MAFG、C1QTNF1、YES1、DCC、SH3GL1、TNFSF9、TNFSF7、TNFSF14、VAV1、RAB3A、PTOV1、BAX、RRAS、BCAS4、HIC2、NROB2、TTN、SGCB、SMA3、SMA4、SMN1、LPA、PARK2、GCK、GPR51、BSCL2、A2M、TBXA2R、FKRP 或 COMT。

[0302] 靶序列可以编码微小 RNA,例如,hsa-let-7g、hsa-mir-135a-1、hsa-mir-95、hsa-mir-218-1、hsa-mir-320、has-let-7a-1、has-let-7d、has-let-7f-1、has-mir-202、has-mir-130a、has-mir-130a、has-mir-338、has-mir-199a-1、has-mir-181c、has-mir-181d、has-mir-23a、has-mir-24-2、has-mir-27a、has-mir-150、has-mir-499、has-mir-124a-3 或 has-mir-185。

[0303] 靶序列可以是 Wong 等人 (2007) *Am J of Hum Genetics*80:91-104 中所列的任何序列。

[0304] 扩增和检测

[0305] 本文所述的方法可利用核酸扩增。靶核酸的扩增可通过本领域已知的任何方法进行。扩增可通过热循环或等温进行。在示例性实施方案中,扩增可以通过聚合酶链反应 (PCR) 来完成。

[0306] 可使用的 PCR 技术的实例包括,但不限于,定量 PCR、定量荧光 PCR (QF-PCR)、多重荧光 PCR (MF-PCR)、实时 PCR (RT-PCR)、单细胞 PCR、限制性片段长度多态性 PCR (PCR-RFLP)、PCR-RFLP/RT-PCR-RFLP、热启动 PCR、巢式 PCR、原位聚合酶群落 (colony) PCR、原位滚环扩增 (RCA)、桥式 PCR、picotiter PCR、数字 PCR、小滴数字 PCR 和乳液 PCR。其它合适的扩增方法包括连接酶链反应 (LCR)、转录扩增、分子倒置探针 (MIP) PCR、自动维持序列复制、靶多核苷酸序列的选择性扩增、共有序列引物聚合酶链反应 (CP-PCR)、任意引物聚合酶链反应 (AP-PCR)、简并寡核苷酸引物 PCR (DOP-PCR) 和基于核酸的序列扩增 (NABSA)。本文可使用的其它扩增方法包括美国专利 5,242,794、5,494,810、4,988,617 和 6,582,938 中描述的那些方法。靶核酸的扩增可在珠上发生。在其它实施方案中,扩增不在珠上发生。扩增可通过等温扩增,例如,等温线性扩增。可进行热启动 PCR,其中在加入聚合酶之前将反应加热

到 95°C 维持两分钟,或者可在循环 1 中的第一加热步骤之前一直保持聚合酶无活性。可使用热启动 PCR 使非特异性扩增最小化。扩增的其它策略和方面在 2010 年 7 月 8 日公布的公开号为 2010/0173394A1 的美国专利申请中有所描述,其通过引用并入本文。

[0307] 靶序列和参考序列的扩增技术是本领域已知的,并且包括在美国专利 7,048,481 中描述的方法。简单地说,所述技术可包括将样品分成小滴的方法和组合物,在一些情况下,每个小滴平均含有少于约 5、4、3、2 或 1 个靶核酸分子(多核苷酸),在每个小滴中扩增核酸序列,并检测靶核酸序列的存在。在一些情况下,扩增的序列存在于针对基因组 DNA 的探针上,而不是基因组 DNA 本身上。在一些情况下,至少 200、175、150、125、100、90、80、70、60、50、40、30、20、10 或 0 个小滴含有零个靶核酸拷贝。

[0308] 可将有关扩增反应的信息输入数据库中。例如,图 15A 和图 15B 图示了可输入数据库的试验信息。

[0309] 引物

[0310] 可根据已知的参数设计引物以避免产生二级结构和自杂交。不同引物对可在大致相同的温度退火和解链,例如,与另一引物对相差约 1、2、3、4、5、6、7、8、9 或 10°C 之内。在一些情况下,起初使用大于约 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、100、200、500、1000、5000、10,000 个或更多的引物。这类引物可能与本文所述的遗传靶标杂交。在一些实施方案中,使用约 2 到约 10,000、约 2 到约 5,000、约 2 到约 2,500、约 2 到约 1,000、约 2 到约 500、约 2 到约 100、约 2 到约 50、约 2 到约 20、约 2 到约 10 或约 2 到约 6 个引物。

[0311] 引物可以通过许多方法制备,包括但不限于适当序列的克隆和使用本领域熟知的方法的直接化学合成(Narang 等人,Methods Enzymol. 68:90(1979);Brown 等人,Methods Enzymol. 68:109(1979))。引物也可以从商业来源获得,如 Integrated DNA Technologies、Operon Technologies、Amersham Pharmacia Biotech、Sigma 和 Life Technologies。引物可具有相同的解链温度。引物的解链温度可以是约 30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、81、82、83、84 或 85°C。在一些实施方案中,引物的解链温度是约 30 到约 85°C、约 30 到约 80°C、约 30 到约 75°C、约 30 到约 70°C、约 30 到约 65°C、约 30 到约 60°C、约 30 到约 55°C、约 30 到约 50°C、约 40 到约 85°C、约 40 到约 80°C、约 40 到约 75°C、约 40 到约 70°C、约 40 到约 65°C、约 40 到约 60°C、约 40 到约 55°C、约 40 到约 50°C、约 50 到约 85°C、约 50 到约 80°C、约 50 到约 75°C、约 50 到约 70°C、约 50 到约 65°C、约 50 到约 60°C、约 50 到约 55°C、约 52 到约 60°C、约 52 到约 58°C、约 52 到约 56°C 或约 52 到约 54°C。

[0312] 引物的长度可在 5' 端或 3' 端延长或缩短,从而生产有期望的解链温度的引物。引物对中的一个引物可以比另一个引物长。引物对内引物的 3' 退火长度可以不同。另外,可以设计每个引物对的退火位置,以使引物对的序列和长度产生期望的解链温度。用于确定小于 25 个碱基对的引物的解链温度的方程为 Wallace 规则($Td=2(A+T)+4(G+C)$)。还可以使用计算机程序设计引物对,包括但不限于阵列设计软件(Array Designer Software)(Arrayit Inc.)、用于遗传分析的寡核苷酸探针序列设计软件(Oligonucleotide Probe Sequence Design Software for Genetic Analysis)(Olympus Optical Co.)、NetPrimer 和来自 Hitachi Software Engineering 的 DNAsis。可使用软件程序,如 Net Primer(免

费的基于网络的程序, <http://www.premierbiosoft.com/netprimer/index.html>) 来计算每个引物的 T_M (解链温度或退火温度)。引物的退火温度可在任何扩增循环后重新计算并升高, 所述任何扩增循环包括但不限于约 1、2、3、4、5 次循环、约 6 次循环到约 10 次循环、约 10 次循环到约 15 次循环、约 15 次循环到约 20 次循环、约 20 次循环到约 25 次循环、约 25 次到约 30 次循环、约 30 次到约 35 次循环或约 35 次循环到约 40 次循环。在初始扩增循环后, 可将引物的 5' 一半引入来自每个目标基因座的产物中; 因此可基于每个引物的 5' 一半和 3' 一半的序列重新计算 T_M 。

[0313] 引物的退火温度可以在任何扩增循环后重新计算并升高, 所述任何扩增循环包括但不限于约 1、2、3、4、5 次循环、约 6 次循环到约 10 次循环、约 10 次循环到约 15 次循环、约 15 次循环到约 20 次循环、约 20 次循环到约 25 次循环、约 25 次到约 30 次循环、约 30 次到约 35 次循环或约 35 次循环到约 40 次循环。在初始扩增循环后, 可将引物的 5' 一半引入来自每个目标基因座的产物中, 因此可基于每个引物的 5' 一半和 3' 一半的序列重新计算 T_M 。

[0314] DNA 聚合酶

[0315] 可使用任何催化引物延伸的 DNA 聚合酶, 包括但不限于大肠杆菌 DNA 聚合酶、大肠杆菌 DNA 聚合酶 1 的 Klenow 片段、T7DNA 聚合酶、T4DNA 聚合酶、Taq 聚合酶、Pfu DNA 聚合酶、Pfx DNA 聚合酶、Tth DNA 聚合酶、Vent DNA 聚合酶、噬菌体 29、REDTaq™、基因组 DNA 聚合酶或测序酶。可使用热稳定的 DNA 聚合酶。DNA 聚合酶可具有 3' 到 5' 外切核酸酶活性。DNA 聚合酶可具有 5' 到 3' 外切核酸酶活性。DNA 聚合酶可同时具有 3' 到 5' 外切核酸酶活性和 5' 到 3' 外切核酸酶活性。

[0316] 热循环

[0317] 可用任何数目的 PCR 循环来扩增 DNA, 例如, 大约、大于约或小于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44 或 45 次循环。扩增循环数可以是约 1 到约 45、约 10 到约 45、约 20 到约 45、约 30 到约 45、约 35 到约 45、约 10 到约 40、约 10 到约 30、约 10 到约 25, 约 10 到约 20、约 10 到约 15、约 20 到约 35、约 25 到约 35、约 30 到约 35 或约 35 到约 40。

[0318] 热循环反应可以对包含于小滴中的样品进行。小滴可在热循环期间保持完整。小滴可在热循环期间以大于约 10,000 小滴/mL、100,000 小滴/mL、200,000 小滴/mL、300,000 小滴/mL、400,000 小滴/mL、500,000 小滴/mL、600,000 小滴/mL、700,000 小滴/mL、800,000 小滴/mL、900,000 小滴/mL 或 1,000,000 小滴/mL 的密度保持完整。在其它情况下, 两个或更多小滴可在热循环期间合并。在其它情况下, 多于 100 或多于 1,000 个小滴可在热循环期间合并。

[0319] 探针

[0320] 可通过本领域已知的方法设计通用探针。在一些实施方案中, 探针为随机序列。可选择通用探针以确保它在试验中不结合靶多核苷酸或结合样品中可能存在的其它非靶多核苷酸 (例如, 由靶多核苷酸占据的区域外的基因组 DNA)。

[0321] 在本文所述的方法中用于检测靶核酸序列或参考核酸序列的探针 (例如, Taqman 探针) 上使用的标记物 (荧光团、染料) 可以是, 例如, 6-羧基荧光素 (FAM)、四羧荧光素

(TET)、4, 7, 2' - 三氯 -7' - 苯基 -6- 羧基荧光素 (VIC)、HEX、Cy3、Cy3. 5、Cy5、Cy5. 5、Cy7、四甲基罗丹明、ROX 和 JOE。标记物可为 Alexa Fluor 染料, 例如, Alexa Fluor350、405、430、488、532、546、555、568、594、633、647、660、680、700 和 750。标记物可为级联蓝 (Cascade Blue)、海蓝 (Marina Blue)、俄勒冈绿 (Oregon Green) 500、俄勒冈绿 514、俄勒冈绿 488、俄勒冈绿 488-X、太平洋蓝 (Pacific Blue)、罗丹明绿、对甲氨基酚绿 (Rhodol Green)、罗丹明绿 -X、罗丹明红 -X 和德克萨斯红 -X。标记物可处于探针的 5' 端、探针的 3' 端、探针的 5' 端和 3' 端, 或在探针内部。实验中可使用独特的标记物检测每个不同的基因座。

[0322] 探针, 如 Taqman 探针, 可包含猝灭剂, 例如, 3' 猝灭剂。3' 猝灭剂可以是, 例如, TAMARA、DABCYL、BHQ-1、BHQ-2 或 BHQ-3。在一些情况下, 在本文提供的方法中使用的猝灭剂是黑洞猝灭剂 (BHQ)。在一些情况下, 猝灭剂是小沟结合物 (MGB)。在一些情况下, 猝灭剂是荧光猝灭剂。在其它情况下, 猝灭剂是非荧光猝灭剂 (NFQ)。

[0323] 探针可以是大约或至少约 5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39 或 40 个碱基长。探针可以是约 8 到约 40、约 10 到约 40、约 10 到约 35、约 10 到约 30、约 10 到约 25、约 10 到约 20、约 15 到约 40、约 15 到约 35、约 15 到约 30、约 15 到约 25、约 15 到约 20、约 18 到约 40、约 18 到约 35、约 18 到约 30、约 18 到约 25 或约 18 到 22 个碱基。

[0324] 试剂和添加剂

[0325] 用于进行 PCR 反应的溶液和试剂可包括缓冲液。所述缓冲溶液可包含大约、大于约或小于约 1、5、10、15、20、30、50、100 或 200mM Tris。在一些情况下, 氯化钾的浓度可以是大约、大于约或小于约 10、20、30、40、50、60、80、100、200mM。缓冲溶液可包含约 15mM Tris 和 50mM KCl。核苷酸可包含脱氧核苷酸三磷酸分子, 包括 dATP、dCTP、dGTP、dTTP, 各自的浓度是大约、大于约或小于约 50、100、200、300、400、500、600 或 700 μ M。在一些情况下, 将非标准核苷酸如 dUTP 加入到扩增反应中达到大约、大于约或小于约 50、100、200、300、400、500、600 或 700、800、900 或 1000 μ M 的浓度。在一些情况下, 将氯化镁 ($MgCl_2$) 以大约、大于约或小于约 1.0、2.0、3.0、4.0 或 5.0mM 的浓度加入到扩增反应中。 $MgCl_2$ 的浓度可为约 3.2mM。

[0326] 可使用非特异性封闭剂如 BSA 或来自牛皮的明胶, 其中明胶或 BSA 以约 0.1 到约 0.9%w/v 的浓度范围存在。其它可能的封闭剂可包括 β 乳球蛋白、酪蛋白、乳粉或其它常见封闭剂。在一些情况下, BSA 和明胶的优选浓度为约 0.1%w/v。

[0327] 在一些实施方案中, 扩增反应还可以包含一种或多种添加剂, 包括但不限于, 非特异性背景/封闭核酸 (如, 鲑精 DNA)、生物防腐剂 (如, 叠氮钠)、PCR 增强剂 (如, 甜菜碱、海藻糖等) 和抑制剂 (如, RNA 酶抑制剂)。一种或多种添加剂可包括, 例如, 2- 吡咯烷酮、乙酰胺、N- 甲基吡咯烷酮 (NMP)、B- 羟乙基吡咯烷酮 (HEP)、丙酰胺、NN- 二甲基乙酰胺 (DMA)、N- 甲基甲酰胺 (MMP)、NN- 二甲基甲酰胺、甲酰胺、N- 甲基乙酰胺 (MMA)、二甲亚砜 (DMSO)、聚乙二醇、甜菜碱、四甲基氯化铵 (TMAC)、7- 脱氮 -2' - 脱氧鸟苷、牛血清白蛋白 (BSA)、T4 基因 32 蛋白质、甘油或非离子型去污剂 (Triton X-100、吐温 20、Nonidet P-40 (NP-40)、吐温 40、SDS (例如, 约 0.1%SDS)、鲑精 DNA、叠氮钠、甜菜碱 (N, N, N- 三甲基甘氨酸; [羧甲基] 三甲基铵)、甲酰胺、海藻糖、二硫苏糖醇 (DTT)、 β - 巯基乙醇、植物多糖或 RNA 酶抑制剂。

[0328] 在一些实施方案中, 扩增反应包含一种或多种缓冲液。所述一种或多种缓冲液可

以包含,例如,TAPS、N-二甘氨酸、Tris、三甲基甘氨酸、TAPSO、HEPES、TES、MOPS、PIPES、甲次膦酸盐、SSC、ADA、ACES、氯化胆胺、乙酰氨基甘氨酸、甘氨酸、马来酸盐、磷酸盐、CABS、哌啶、甘氨酸、柠檬酸盐、双甘氨酸、苹果酸盐、甲酸盐、琥珀酸盐、乙酸盐、丙酸盐、吡啶、哌嗪、组氨酸、双-tris、乙醇胺、碳酸盐、MOPSO、咪唑、双-TRIS 丙烷、BES、MOBS、三乙醇胺 (TEA)、HEPPSO、POPSO、胍、氨基丁三醇 (tris)、EPPS、HEPPS、N-二甘氨酸、HEPBS、AMPSO、牛磺酸 (AES)、硼酸盐、CHES、2-氨基-2-甲基-1-丙醇 (AMP)、氢氧化铵、甲胺或 MES。

[0329] 在一些情况下,以约 0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9% 或 1.0% 的浓度,向扩增反应中加入非离子型环氧乙烷 / 环氧丙烷嵌段共聚物。常用的生物表面活性剂包括非离子型表面活性剂,如 Pluronic F-68、Tetronics、Zonyl FSN。Pluronic F-68 可以约 0.5%w/v 的浓度存在。

[0330] 在一些情况下,硫酸镁可以以相似的浓度替代氯化镁。来自不同供应商的众多常用的商业化 PCR 缓冲液可以替代缓冲的溶液。

[0331] 检测

[0332] 荧光检测可使用多种检测器装置完成,该检测器装置配备有用于产生可被荧光剂吸收的激发光的模块以及用于检测由荧光剂发射的光的模块。在一些情况下,可以整体检测样品(如小滴)。例如,可将样品分配到置于检测器中的塑料管中,该检测器可测量来自塑料管的本体荧光 (bulk fluorescence)。在一些情况下,可将一个或多个样品(如小滴)分配到板如 96-孔板或 384 孔板的一个或多个孔中,并且可使用荧光板读取器检测单个孔的荧光。

[0333] 在一些情况下,检测器进一步包含对小滴样品的处理能力,单个小滴输入检测器、进行检测、然后离开检测器。例如,流式细胞术装置可适用于检测来自小滴样品的荧光。在一些情况下,配备有控制小滴运动的泵的微流体装置用于依次检测来自小滴的荧光。在一些情况下,小滴排列在二维表面上,并且检测器相对于表面移动,在每个包含一个小滴的位置检测荧光。

[0334] 计算机

[0335] 获得荧光检测数据之后,可以使用计算机存储和处理数据。可使用计算机可执行逻辑行使诸如扣除背景荧光、靶序列和 / 或参考序列的分配和数据定量的功能。计算机可用于显示、存储、检索或计算来自分子谱分析的诊断结果;显示、存储、检索或计算来自基因组或核酸表达分析的原始数据;或显示、存储、检索或计算在本发明方法中有用的任何样品信息或患者信息。

[0336] 数字分析

[0337] 通过将样品中的靶序列进行分区和鉴定含有靶序列的分区,可利用数字读取试验如数字 PCR 计数靶序列(如,靶核酸序列)。数字读取是一种“全或无”的分析,因为它指定了给定分区是否含有目标靶序列,但不一定指示分区中有多少靶序列拷贝。例如,含有两个靶序列的单一多核苷酸可处于分区中,但是在正常分析条件下,该分区将只被认为含有一个靶序列。如果在同一多核苷酸上的靶序列被大量碱基对隔开,那么在样品纯化期间一些靶核酸序列可通过断裂而分离——一些连接的靶核酸序列在样品制备后可能不会保持物理连接。数字 PCR 通常,例如,在 Vogelstein 和 Kinzler (1999) PNAS96:9236-9241 中有所描述。该技术的应用包括,例如,高分辨率 CNV 测量、跟进 (follow-up) 至全基因组关联研

究、细胞遗传学分析、癌组织中的 CNV 改变和 CNV 连锁分析。

[0338] 通常, dPCR 可包括将来自样品的单个多核苷酸进行空间分离(或分区), 并对每个分区进行聚合酶链反应。分区可以是, 例如, 孔(例如, 微孔板的孔)、毛细管、乳液的分散相、室(例如, 小型化室阵列中的室)、小滴或核酸结合表面。可分配样品, 使得每个分区含有约 0、1 或 2 个靶多核苷酸。每个分区(如, 小滴)可平均含有少于 5、4、3、2 或 1 个靶核酸拷贝。在一些情况下, 至少 0、1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175 或 200 个分区(如, 小滴)含有零个靶核酸拷贝。PCR 扩增后, 可计算含有或不含 PCR 产物的分区数。分区的总数可以是大约或大于约 500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、150,000、200,000、500,000、750,000 或 1,000,000。分区的总数可以是约 500 到约 1,000,000、约 500 到约 500,000、约 500 到约 250,000、约 500 到约 100,000、约 1000 到约 1,000,000、约 1000 到约 500,000、约 1000 到约 250,000、约 1000 到约 100,000、约 10,000 到约 1,000,000、约 10,000 到约 100,000 或约 0,000 到约 50,000。

[0339] 在一个实施方案中, 数字 PCR 为小滴数字 PCR。在小滴数字 PCR 实验的一些实施方案中, 可检测少于约 0.00001、0.00005、0.00010、0.00050、0.001、0.005、0.01、0.05、0.1、0.5、1、2、2.5、3、3.5、4、4.5、5、6、7、8、9 或 10 个靶多核苷酸拷贝。在一些情况下, 可检测少于约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、150、200、250、300、350、400、450 或 500 个靶多核苷酸拷贝。在一些情况下, 本文所述的小滴以大于 1、2、3、4、5、10、50、100、200、300、400、500、600、700、800、900 或 1000 小滴/秒的速度产生。

[0340] 小滴数字 PCR(ddPCR™) 可提供用于验证由下一代测序仪和微阵列鉴定的拷贝数变异的实际方案。使用 ddPCR™ 的方法可以允许筛选许多样品, 例如数百个样品, 用于在单一工作班次中进行 CNV 分析。在一个实施方案中, 提供了 ddPCR™ 的工作流程, 该工作流程涉及在组装包括用来检测靶核酸序列(例如, 第一基因)和单拷贝参考核酸序列(例如, 第二基因)的试剂的双重 TaqMan® 试验之前, 使用一种或多种限制酶分离靶核酸序列的串联拷贝。当使用 ddPCR 时, 反应混合物然后可分成大约、少于约或多于约 500、1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、12,000、13,000、14,000、15,000、16,000、17,000、18,000、19,000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100,000、150,000、200,000、500,000、750,000、1,000,000、2,000,000、3,000,000、4,000,000、5,000,000、6,000,000、7,000,000、8,000,000、9,000,000 或 10,000,000 个纳升小滴, 在分析前可将这些纳升小滴热循环到终点。在一些情况下, 小滴大于一纳升; 在其它情况下, 小滴小于一纳升(例如, 皮升)。每个反应的小滴数可为约 1000 到约 1,000,000、约 1000 到约 750,000、约 1000 到约 500,000、约 1000 到约 250,000、约 1000 到约 100,000、约 1000 到约 50,000、约 1000 到约 30,000、约 1000 到约 10,000、约 10,000 到约 1,000,000、约 10,000 到约 750,000、约 10,000 到约 500,000、约 10,000 到约 250,000、约 10,000 到约 100,000、约 10,000 到约 50,000 或约 10,000 到约 30,000。每个反应的小滴数可以是约 20,000 到约 1,000,000、约 20,000 到约 750,000、约 20,000 到约 500,000、约 20,000 到约 250,000、约 20,000 到约 100,000、约 20,000 到约 50,000、约 20,000 到约 30,000。

50,000 到约 200,000 或约 50,000 到约 300,000。

[0341] 分析可在双色读取器上发生。正计数的小滴的分数可以使靶核酸序列和参考核酸序列（例如基因）的绝对浓度能够得到测定。可使用该信息来确定相对拷贝数。例如，每孔至少 20,000 个 PCR 重复可提供分辨高阶拷贝数差异的统计功效。这种低成本的方法可以可靠地产生具有 95% 置信区间且跨越整数而无相邻拷贝数状态重叠的拷贝数测量值。这种技术能够确定拷贝数变体的连接，并且可用于确定基因拷贝是在相同染色体上还是在不同染色体上。

[0342] 体积可具有任何合适的大小。在一些实施方案中，体积可具有约 10 到 1000 微米的直径或特征横截面尺寸。

[0343] 被分区的核酸可具有任何合适的特征。核酸尤其可包括受试者的遗传物质（例如，受试者的基因组 DNA 和 / 或 RNA）、受试者的信使 RNA 和 / 或来源于受试者 RNA 的 cDNA 等。核酸可具有任何合适的平均长度。通常，平均长度基本上大于染色体上待分析的多态基因座之间的距离。因为该平均长度，在受试者中连接的等位基因也常常连接在分离的核酸中，因此当对水相进行分区时，等位基因倾向于一起分布至相同体积中。在一些实施方案中，每个引物组能够从多态基因座扩增至少一对不同的等位基因。

[0344] 可将每个体积进行分区，从而含有任何合适的平均核酸浓度。通常，分区的过程，结合水相中核酸的合适起始浓度，产生每体积含有平均少于约数个基因组当量的核酸的体积。虽然可利用每体积平均多于一个基因组当量（例如，每体积约两个基因组当量）执行该方法，但通过将浓度限制到每体积平均少于约一个基因组当量，因而具有较低的背景，分析通常会变得更加有效和可靠。因此，每个体积可含有包括每个多态基因座的靶区域的平均少于约一个拷贝或分子，和 / 或每个多态基因座的任意等位基因序列的平均少于约一个拷贝。

[0345] 本文所述的方法中可使用集成的、快速、流通型热循环仪装置。参见，例如，2009 年 9 月 23 日提交的国际申请 PCT/US2009/005317。在这样的集成装置中，毛细管缠绕在保持 2、3 或 4 个温度区的圆柱体上。当小滴流经毛细管时，它们经过不同温度区实现热循环。随着小滴进入每个温度区，每个小滴的小体积会导致非常快速的温度转变。

[0346] 供本文所述的方法、组合物和试剂盒使用的数字 PCR 装置（例如，小滴数字 PCR 装置）可检测多种信号（参见，例如，2011 年 3 月 18 日提交的美国临时专利申请 61/454,373，其通过引用整体并入本文）。

[0347] 小滴数字 PCR 可涉及每秒数千个离散的、稳健的微滴反应器的产生。ddPCR 可涉及使用安装的基础仪器的标准热循环，其可使研究者立即获取数字数据。对每个小滴的快速探询可以产生初始样品中存在的靶分子的计数。

[0348] 图 16 图示了 ddPCR 实验的一般工作流程的实例。如图 16 中所示，该过程可从将样品分成多个分区（如，小滴）开始，然后在热循环仪中对样品进行热循环。然后可用读取器（例如，光学读取器）检测小滴的荧光。

[0349] 小滴产生

[0350] 本发明包括使用小滴数字 PCR 的组合物和方法。本文所述的小滴包括在美国专利 7,622,280 中描述的乳液组合物（或两种或更多种不混溶流体的混合物），和由在 2009 年 9 月 23 日提交的国际申请 PCT/US2009/005317 中所述的装置产生的小滴。本文所用的术

语乳液可指不混溶液体（如油和水）的混合物。油相和 / 或油包水乳液允许水性小滴内反应混合物的区室化。乳液可包含连续油相内的水性小滴。本文提供的乳液可以是水包油乳液，其中小滴为在连续水相中的油滴。本文提供的小滴设计为防止区室之间的混合，其中每个区室防止其内容物蒸发，以及防止其内容物与其它区室的内容物聚结 (coalescing)。

[0351] 本文所述的混合物或乳液可以是稳定的或不稳定的。乳液可以是相对稳定的且具有最小的聚结。当小的小滴结合逐渐形成较大的小滴时会发生聚结。在一些情况下，少于约 0.00001%、0.00005%、0.00010%、0.00050%、0.001%、0.005%、0.01%、0.05%、0.1%、0.5%、1%、2%、2.5%、3%、3.5%、4%、4.5%、5%、6%、7%、8%、9% 或 10% 的由小滴发生器产生的小滴与其它小滴聚结。乳液还可具有有限的絮凝，通过该过程分散相从悬浮液中呈片状析出。

[0352] 如本文所述将样品分成小反应体积可以实现使用减少量的试剂，从而降低分析的材料成本。由于在不同的区室中较高丰度的分子与较低丰度的分子相分离，所以通过分区减少样品的复杂性也提高了检测的动态范围，从而使较低丰度的分子有更高的比例与反应试剂接触，这转而增强较低丰度分子的检测。

[0353] 可产生平均直径为大约、小于约或大于约 0.001、0.01、0.05、0.1、1、5、10、20、30、40、50、60、70、80、100、120、130、140、150、160、180、200、300、400 或 500 微米的小滴。小滴可具有约 0.001 到约 500、约 0.01 到约 500、约 0.1 到约 500、约 0.1 到约 100、约 0.01 到约 100 或约 1 到约 100 微米的平均直径。已知使用微通道错流聚焦或物理搅拌产生乳液小滴的微流体方法产生单分散或多分散的乳液。小滴可以是单分散小滴。可产生小滴，使得所述小滴的大小变化不超过所述小滴的平均大小的 $\pm 5\%$ 。在一些情况下，产生小滴，使得所述小滴的大小变化不超过所述小滴的平均大小的 $\pm 2\%$ 。小滴发生器可从单一样品产生小滴群体，其中小滴的大小变化均不超过小滴总群的平均大小的 \pm 约 0.1%、0.5%、1%、1.5%、2%、2.5%、3%、3.5%、4%、4.5%、5%、5.5%、6%、6.5%、7%、7.5%、8%、8.5%、9%、9.5% 或 10%。

[0354] 较高机械稳定性对于微流体操作和较高剪切流体处理（例如，在微流体毛细管中或在流体路径中经过 90 度弯折部分，如阀门）可能是有用的。预先或后续热处理的小滴或胶囊对标准移液管操作和离心可能是机械稳定的。

[0355] 可通过使油相流过含水样品而形成小滴。水相可包含用于进行 PCR 反应的缓冲的溶液和试剂，包括用于荧光检测的核苷酸、引物、探针，模板核酸、DNA 聚合酶和任选的逆转录酶。

[0356] 水相可包含本文所述的一种或多种缓冲液和 / 或添加剂。

[0357] 水相中用于扩增的引物可具有大约、大于约或小于约 0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.2、1.5、1.7 或 2.0 μM 的浓度。水相中引物的浓度可以是约 0.05 到约 2、约 0.1 到约 1.0、约 0.2 到约 1.0、约 0.3 到约 1.0、约 0.4 到约 1.0 或约 0.5 到约 1.0 μM 。引物的浓度可以是约 0.5 μM 。水相可包含一种或多种用于荧光检测的探针，其浓度为大约、大于约或小于约 0.05、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1.0、1.2、1.4、1.6、1.8 或 2.0 μM 。水相可包含一种或多种用于荧光检测的探针，其浓度为约 0.05 到约 2.0、约 0.1 到约 2.0、约 0.25 到约 2.0、约 0.5 到约 2.0、约 0.05 到约 1、约 0.1 到约 1 或约 0.1 到约 0.5 μM 。用于荧光检测的探针的浓度可以是约 0.25 μM 。在 PCR 中，合适的靶核酸浓度范围为约 1pg 到约 500ng。

[0358] 油相可包含氟化基础油 (base oil)，该基础油可通过与氟化表面活性剂如全氟

化聚醚组合而额外得到稳定。在一些情况下,基础油可以是 HFE7500、FC-40、FC-43、FC-70 的一种或多种或另一种常用氟化油。在一些情况下,阴离子型表面活性剂为 Krytox 铵 (Krytox-AM)、Krytox FSH 的铵盐或 Krytox-FSH 的吗啉衍生物。Krytox-AS 可以大约、大于约或小于约 0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1.0%、2.0%、3.0% 或 4.0%w/w 的浓度存在。在一些优选的实施方案中,Krytox-AS 的浓度为 1.8%。在其它优选的实施方案中,Krytox-AS 的浓度为 1.62%。Krytox-FSH 的吗啉衍生物可以约 0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1.0%、2.0%、3.0% 或 4.0%w/w 的浓度存在。Krytox-FSH 的吗啉衍生物的浓度可为约 1.8%。Krytox-FSH 的吗啉衍生物的浓度可为约 1.62%。

[0359] 油相可进一步包含用于调节油的性质如蒸汽压或粘度或表面张力的添加剂。非限制性实例包括全氟辛醇和 1H, 1H, 2H, 2H- 全氟癸醇。可加入 1H, 1H, 2H, 2H- 全氟癸醇至大约、大于约或小于约 0.05%、0.06%、0.07%、0.08%、0.09%、1.00%、1.25%、1.50%、1.75%、2.00%、2.25%、2.50%、2.75% 或 3.00%w/w 的浓度。可加入 1H, 1H, 2H, 2H- 全氟癸醇至约 0.18%w/w 的浓度。

[0360] 可配制乳液以产生高度单分散的、具有液体样界面膜的小滴,这些小滴可通过加热转变成具有固体样界面膜的微胶囊;这些微胶囊能像生物反应器一样表现,在整个反应过程如 PCR 扩增中能够保持其内容物。加热时可发生向微胶囊形式的转变。例如,这种转变可在大于约 50、60、70、80、90 或 95 摄氏度的温度下发生。在一些情况下,使用热循环仪进行这种加热。在加热过程中,流体或矿物油覆盖层可用于防止蒸发。在加热前,可除去或不除去过多的连续相油。在众多的热处理和机械处理中,生物相容性胶囊可抵抗聚结和/或絮凝。

[0361] 转变之后,可将胶囊储存在大约、高于约或低于约 3、4、5、6、7、8、9、10、15、20、25、30、35 或 40 摄氏度。这些胶囊在生物医学应用中可能是有用的,如大分子的稳定、数字化封装,尤其是含有核酸或蛋白质或二者的混合物的水性生物流体;载药和疫苗递送;生物分子文库;临床成像应用等。

[0362] 微胶囊可含有一种或多种多核苷酸并且可以抗聚结,尤其是在高温下。因此,PCR 扩增反应可在非常高的密度下(例如,每单位体积的反应数)发生。在一些情况下,每毫升可发生多于约 100,000、500,000、1,000,000、1,500,000、2,000,000、2,500,000、5,000,000 或 10,000,000 个单独反应。在一些情况下,反应发生在单一孔内,例如,微量滴定板的孔,无需反应体积之间的相互混合。微胶囊还可含有使 PCR 反应能够发生的其它必需组分,例如,引物、探针、dNTPs、DNA 或 RNA 聚合酶等。这些胶囊在众多热处理和机械处理中显示出对聚结和絮凝的抗性。

[0363] 在一个实施方案中,当通过例如消化、热处理或剪切使 DNA 的大小减小后,小滴的生成可得到改善。

[0364] 图 17 展示了小滴的几个图像,其显示 a) 随着油从侧面流入将小滴压缩而形成小滴,和 b) 随着小滴从本体流体中脱离而伸长或收缩。

[0365] 图 18 示出了增加 DNA 负载的效果。图 18 将最大拉伸相对于流速作图。从截面的中心到小滴刚好破裂时的最远范围测量拉伸。一定的小滴拉伸是可容许的,但是如果拉伸变得过度,会拉出将小滴与本体流体连接起来的长“线”。随着小滴破裂,该线可塌缩成微

滴,从而导致不期望的多分散性。在极端情况下,小滴不破裂;相反水相作为连续相从通道中心流下,而油相沿着通道壁流动,并且没有小滴形成。

[0366] 减小拉伸的一个方法是降低流速。降低流速可具有较低的通量以及增加的小滴大小的不期望的副作用。紫色 (B)、青色 (E) 和绿色 (A) 曲线具有零 DNA。这些样品可容忍高流速而不会显著增加它们向通道内的拉伸。

[0367] 蓝色 (D)、橙色 (F) 和红色 (C) 曲线具有较高的 DNA 负载。在这些情况下,较高流速会导致小滴向通道内拉伸。可使用低流速来避免小滴的过度拉伸。

图 19 示出了在研究小滴性质的实验中未消化的样品 1-10 和已消化的样品 11-20。DNA 负载在最右边的列中示出;压力(大致与流速成正比)在第二行中示出。该表是用颜色与字母编码的:J(红色)表示喷射,E(黄色)表示拉伸,而 N(绿色)表示正常(无喷射或拉伸)小滴产生。可以看出,甚至在高 DNA 负载和高流速下,消化(用限制酶)导致改善的小滴产生。

[0368] 应用

[0369] 本文所述的方法可以用来诊断或预测病症或疾病。

[0370] 本文提供的方法和组合物对于人类和非人类受试者都是有用的。本文提供的方法和组合物的应用众多,例如,高分辨率 CNV 测量、跟进至全基因组关联研究、细胞遗传学分析、癌组织中的 CNV 改变、CNV 连锁分析以及单元型分析。

[0371] 本文提供的应用包括用于诊断、预测、确定或评估胎儿或胚胎的遗传特征的应用。在一些情况下,这些应用可用来诊断、预测、确定或评估通过体外受精或其它辅助生殖技术产生的胚胎中的核酸。此外,本文提供的方法可用于为未来父母(如,孕妇)提供信息,以便评估在发育中胎儿的基因组中的 CNV 或遗传定相。在其它情况下,本文提供的方法可用于帮助在关于未来后代的可能遗传特性方面给患者以建议。在一些情况下,这些方法可以与辅助生殖技术结合使用。例如,该信息可用于评估取自通过体外受精产生的胚胎的样品中的 CNV 或遗传定相。

[0372] 可在癌细胞中发现一种或多种 CNV。例如,在非小细胞肺癌中,EGFR 的拷贝数可能会增加。CNV 可以和治疗效果相关联。例如,在非小细胞肺癌晚期,增加的 HER2 基因拷贝数可以增强对吉非替尼治疗的响应。参见 Cappuzzo F. 等人 (2005) *J. Clin. Oncol.* 23:5007-5018。EGFR 基因的高拷贝数可以预测提高的对拉帕替尼和卡培他滨的敏感性。参见 Fabi 等人 (2010) *J. Clin. Oncol.* 28:15s (2010 年 ASCO 年会)。EGFR 基因的高拷贝数与提高的对西妥昔单抗和帕尼单抗的敏感性有关。

[0373] 在一个实施方案中,提供了一种方法,该方法包括使用本文所述的方法测定靶序列的拷贝数,以及基于所述测定设计疗法。在一个实施方案中,靶标是 EGFR,并且疗法包括施用西妥昔单抗、帕尼单抗、拉帕替尼和/或卡培他滨。在另一实施方案中,靶标是 ERBB2,并且疗法包含曲妥单抗(赫赛汀)。

[0374] 拷贝数变异可促进人类中的遗传变异。参见,例如,Shebat J. 等人 (2004) *Science* 305:525-528。

[0375] 与拷贝数变异有关的疾病可包括,例如,DiGeorge/腭-心-面综合征(22q11.2 缺失)、Prader-Willi 综合征(15q11-q13 缺失)、Williams-Beuren 综合征(7q11.23 缺失)、Miller-Dieker 综合征(MDLS)(17p13.3 微缺失)、Smith-Magenis 综合征(SMS)(17p11.2 微

缺失)、1型多发性神经纤维瘤(NF1)(17q11.2微缺失)、Phelan-McErmid综合征(22q13缺失)、Rett综合征(染色体Xq28上MECP2的功能缺失突变)、Merzbacher病(PLP1的CNV)、脊髓性肌萎缩(SMA)(5q13染色体上端粒SMN1的纯合缺乏)、Potocki-Lupski综合征(PTLS, 17p.11.2染色体的重复)。PMP22基因的额外拷贝可能与IA型Charcot-Marie-Tooth神经病(CMT1A)和易发生压迫性麻痹的遗传性神经病有关。本文所述的检测CNV的方法可用于诊断本文所述的以及通过引用而并入的出版物中所述的CNV病症。该疾病可以是在Lupski J. (2007)Nature Genetics39:S43-S47中描述的疾病。

[0376] 非整倍性,例如,胎儿非整倍性,可包括,例如,三体性13、三体性18、三体性21(唐氏综合征)、克莱恩费尔特综合征(XXY)、一条或多条染色体的单体性(X染色体单体性、特纳综合征)、三体性X、一条或多条染色体的三体性、一条或多条染色体的四倍性或五倍性(例如,XXXX、XXYY、XXXY、YYYY、XXXXX、XXXXY、XXXYY、YYYYY和XXYYY)、三倍性(每种染色体有三条,例如,在人类中的69条染色体)、四倍性(每种染色体有四条,例如,在人类中的92条染色体)和多倍性。在一些实施方案中,非整倍性可以是节段性非整倍性。节段性非整倍性可包括,例如,1p36重复、重复(17)(p11.2p11.2)综合征、唐氏综合征、Pelizaeus-Merzbacher病、重复(22)(q11.2q11.2)综合征和猫眼综合征。在一些情况下,异常基因型,例如,胎儿基因型,起因于性染色体或常染色体的一个或多个缺失,这会导致以下病状,如猫叫综合征、Wolf-Hirschhorn综合征、Williams-Beuren综合征、Charcot-Marie-Tooth病、易发生压迫性麻痹的遗传性神经病、Smith-Magenis综合征、多发性神经纤维瘤、Alagille综合征、腭-心-面综合征、DiGeorge综合征、类固醇硫酸酯酶缺乏症、Kallmann综合征、与线性皮肤缺陷相关的小眼畸形、肾上腺发育不全、甘油激酶缺乏症、Pelizaeus-Merzbacher病、Y染色体上的睾丸决定因子、精子缺乏(因子a)、精子缺乏(因子b)、精子缺乏(因子c)或1p36缺失。在一些实施方案中,染色体数的减少导致了XO综合征。

[0377] 在Li-Fraumeni癌症倾向综合征中发现过多的基因组DNA拷贝数变异(Shlien等人(2008)PNAS105:11264-9)。CNV与畸形综合征有关,包括CHARGE(缺损、心脏异常、鼻后孔闭锁、阻滞、生殖器(gential)和耳异常)、Peters-Plus、Pitt-Hopkins和血小板减少-桡骨缺失综合征(参见,例如,Ropers HH(2007)Am J of Hum Genetics81:199-207)。拷贝数变异与癌症之间的关系例如在Shlien A.和Malkin D. (2009)Genome Med. 1(6):62中有述。拷贝数变异与例如孤独症、精神分裂症和先天学习障碍有关。参见,例如,Sebat J.等人(2007)Science316:445-9;Pinto J.等人(2010)Nature466:368-72;Cook E. H.和Scherer S. W. (2008)Nature455:919-923。

[0378] 拷贝数变异可与癌症患者对某些治疗的抗性有关。例如,胸苷酸合酶的扩增可导致转移性结直肠癌患者对5-氟尿嘧啶治疗的抗性。参见Wang等人(2002)PNAS USA, vol. 99, pp. 16156-61。

[0379] CCL3L1的高拷贝数与对HIV感染的较低敏感性相关(Gonzalez E.等人(2005)Science307:1434-1440)。FCGR3B(CD16细胞表面免疫球蛋白受体)的低拷贝数可以增加对系统性红斑狼疮的敏感性(Aitman T. J.等人(2006)Nature439:851-855)。发现常染色体显性小耳畸形与在染色体4p16上的拷贝数变异区域的五个串联拷贝相关(Balikova I. (2008)Am J. Hum Genet. 82:181-187)。本文所述的方法、组合物和试剂盒可用于研究任

何这些病状。

[0380] 来自高淀粉饮食群体的个体通常比来自低淀粉饮食群体的个体具有更多的淀粉酶基因 (AMY1) 拷贝 (Perry H. 等人 (2007) Nature Genetics 39:1256-1260)。因此,在进化过程中拷贝数可经历正选择。本文所述的方法、组合物和试剂盒可用来研究进化。

[0381] 与疾病相关的拷贝数变异的其它实例包括,例如,三体性 21(唐氏综合征)、三体性 18(爱德华兹综合征)和三体性 13(帕塔综合征)。

[0382] 确定核酸是连接的还是分开的(断裂的)可以为多种应用提供有用的信息。例如,本文所述的方法可用来诊断或预测病症或疾病,例如,遗传性障碍。本文所述的方法可用来诊断和预测胎儿病症,例如,胎儿非整倍性。

[0383] 本文所述的方法对于评估感染如病毒或细菌感染可能是有用的。例如,这些方法可用于确定两个或多个突变是否位于单一病毒或细菌中,或两个或多个突变是否在不同的单个病毒或细菌中。

[0384] 本文所述的方法对于监测转基因小鼠的产生可能是有用的。例如,这些方法可用于确定转基因是否已被一次或多次地引入到转基因生物的基因组中。

[0385] 细胞凋亡

[0386] 确定核酸是连接的还是分离的(断裂的)可用来研究细胞凋亡或诊断或预测与细胞凋亡有关的疾病。细胞凋亡是程序性细胞死亡的过程。细胞凋亡的指征可以包括,例如,起泡、细胞膜非对称性和粘附性的丧失、细胞收缩、核断裂、染色质凝聚和染色体 DNA 断裂。在细胞凋亡过程中,内源性内切核酸酶(例如, Ca^{2+} 和 Mg^{2+} 依赖型内切核酸酶)可将染色质 DNA 切成约 180-200 个碱基对或约 180-200 个碱基对的倍数的核小体间片段。使用本文所述的方法、组合物和 / 或试剂盒可确定不同细胞凋亡阶段的不同形式的遗传物质的断裂状态。本文所述的方法可与一种或多种其它方法一起使用来分析细胞凋亡,包括,例如,流式细胞术、荧光分析或 (TUNEL) (末端脱氧核苷酸转移酶 dUTP 缺口末端标记)。

[0387] 细胞凋亡可包括外源的或死亡受体途径,或者内源的或线粒体途径,和 / 或涉及 T 细胞介导的细胞毒性和细胞的穿孔蛋白 - 粒酶依赖性死亡的途径。死亡受体途径可涉及死亡配体与死亡受体的结合(例如, FasL(脂肪酸合成酶配体、TNFSF6、Apo1、细胞凋亡抗原配体 1、CD95L、CD178、APT1LG1)/FasR(脂肪酸合成酶受体、TNFRSF6、ATP1、CD95); TNF- α (肿瘤坏死因子 α 、TNF 配体、TNFA、恶病质素)/TNFR1(肿瘤坏死因子受体 1、TNF 受体、TNFRSF1A、p55TNFR、CD120a); Apo3L(Apo3 配体、TNFSF12、Apo3 配体、TWEAK、DR3LG)/DR3(死亡受体 3、TNFRSF12、Apo3、WSL-1、TRAMP、LARD、DDR3); Apo2L(Apo2 配体、TNFSF10、TRAIL、TNF 相关细胞凋亡诱导配体)/DR4(死亡受体 4、TNFRSF10A、TRAILR1、APO2); 和 Apo2L/DR5(死亡受体 5、TNFRSF10B、TRAIL-R2、TRICK2、KILLER、ZTNFR9)。TNF 受体可包含富含半胱氨酸的胞外结构域和胞质“死亡结构域”。Fas 结合在 Fas 受体上可导致 FADD 衔接蛋白的结合。TNF 配体与 TNF 受体的结合可导致 TRADD(TNF 受体相关死亡结构域)衔接蛋白的结合以及 FADD(Fas 相关死亡结构域, MORT1) 和 RIP(受体相互作用蛋白质, RIPK1) 的募集。通过与死亡效应结构域二聚化以形成死亡诱导信号复合体(DISC), FADD 然后可与胱天蛋白酶原 -8 关联。该作用可导致胱天蛋白酶原 -8 的自动催化激活。胱天蛋白酶 -8 的激活可导致如下所述的细胞凋亡的“执行阶段”。

[0388] 死亡受体介导的细胞凋亡可受到 c-FLIP(Casper、I-FLICE、FLAME-1、CASH、CLARP、

MRIT) 的抑制, c-FLIP 可与 FADD 和胱天蛋白酶-8 (FLICE、FADD 样 Ice、Mach-1、Mch5) 结合。通过抑制胱天蛋白酶-8 的加工, Toso 可以阻断 T 细胞中 Fas 诱导的细胞凋亡。

[0389] 在穿孔蛋白 / 粒酶途径中, 细胞毒性 T 淋巴细胞 (CTL) 可通过分泌穿孔蛋白 (一种跨膜成孔分子) 杀死肿瘤细胞和病毒感染的细胞, 随后通过该孔释放胞质颗粒并进入到靶细胞中。丝氨酸蛋白酶粒酶 A 和粒酶 B 可以是颗粒的组分。粒酶 B 可切割胱天蛋白酶原-10 从而将其激活, 并且可以切割 ICAD (胱天蛋白酶激活的 DNA 酶的抑制剂)。粒酶 B 可通过切割 Bid 和诱导细胞色素 c 释放而在线粒体途径中发挥作用。粒酶 B 还可直接激活胱天蛋白酶-3。粒酶 A 可通过 DNA 酶 NM23-H1 激活 DNA 切口形成, DNA 酶 NM23-H1 可被 SET 复合物抑制。粒酶 A 可以切割 SET 复合物从而释放 NM23-H1 的抑制。

[0390] 内源性细胞凋亡途径可具有非受体介导的刺激, 并且该刺激可以正面或负面的方式起作用。例如, 某些因子 (例如, 生长因子、激素、细胞因子) 的缺乏可诱导细胞凋亡。能以正面方式起作用的刺激包括, 例如, 辐射、自由基、病毒感染、毒素、低氧和过热。这些刺激可以引起线粒体渗透性转变 (MPT) 孔的打开、线粒体跨膜电位的损失和促凋亡蛋白质从膜间隙向胞质溶胶中的释放。一组释放的蛋白质包括细胞色素 c、Smac/DIABLO 和丝氨酸蛋白酶 HtrA2/Omi, 它们可以激活胱天蛋白酶依赖的线粒体途径。细胞色素 c 可以结合并激活 Apaf-1 和胱天蛋白酶原-9 从而形成凋亡体, 并且可激活胱天蛋白酶-9。Smac/DIABLO 和 HtrA2/Omi 可通过抑制 IAP (细胞凋亡蛋白质的抑制剂) 来促进细胞凋亡。在细胞凋亡过程中, 线粒体可以释放第二组促凋亡蛋白质, AIF、内切核酸酶 G 和 CAD。AIF 可移动到细胞核并可导致 DNA 断裂成约 50 到约 300kb 的片段。AIF 可以引起外围核染色质的凝聚 (“阶段 I” 凝聚)。内切核酸酶 G 可移动到细胞核, 在细胞核内它可切割核染色质以形成寡核小体 DNA 片段。AIF 和内切核酸酶 G 可以不依赖胱天蛋白酶起作用。CAD 可由线粒体释放并且移动到细胞核, 在细胞核内 CAD 可导致寡核小体 DNA 断裂和晚期的核染色质凝聚 (“阶段 II” 凝聚)。

[0391] Bcl-2 家族的蛋白质可调节线粒体膜的通透性。Bcl-2 家族蛋白质可以是促凋亡的或抗凋亡的。抗凋亡蛋白质包括 Bcl-2、Bcl-x、Bcl-XL、Bcl-XS、Bcl-2 和 BAG。促凋亡蛋白质包括, 例如, Bcl-10、Bax、Bak、Bid、Bad、Bim、Bik 和 Blk。Bcl-2 家族的蛋白质可以通过改变线粒体膜的通透性来调节细胞色素 c 从线粒体的释放。蛋白质 14-3-3 可以基于 Bad 的磷酸化状态来调节 Bad。

[0392] Bad 可以与 Bcl-XI 或 Bcl-2 进行异二聚化, 这可以中和它们的保护效应并促进细胞死亡。当 Bcl-2 和 Bcl-XI 未被 Bad 隔离时, Bcl-2 和 Bcl-XI 可以防止细胞色素 c 从线粒体中释放。Aven 与 Bcl-XI 和 Apaf-1 的结合可以防止胱天蛋白酶原-9 的激活。

[0393] 其它 Bcl2 家族的成员包括 Puma 和 Noxa。Puma 和 Noxa 可在 p53 介导的细胞凋亡中起作用。而且 MYC 可以 p53 依赖型或非依赖型方式调节细胞凋亡。

[0394] 死亡受体途径 (外源途径) 和线粒体途径 (内源途径) 可在执行阶段终止。胱天蛋白酶-3、胱天蛋白酶-6 和胱天蛋白酶-7 可以作为执行胱天蛋白酶起作用, 并且可切割底物, 包括 PARP、细胞角蛋白、胞衬蛋白、NuMA、凝溶胶蛋白等。胱天蛋白酶-3 可由引发剂胱天蛋白酶如胱天蛋白酶-8、胱天蛋白酶-9 或胱天蛋白酶-10 激活。胱天蛋白酶-3 可通过切割其抑制剂 ICAD 来激活内切核酸酶 CAD。CAD 可以降解细胞核中的染色体 DNA, 并引起染色质凝聚。细胞凋亡可包括凋亡细胞的吞噬摄取, 后者可包括磷脂不对称性的识别和磷脂

酰丝氨酸的外化 (externalization)。

[0395] 细胞凋亡的概述可见例如 Elmore S. (2007) Apoptosis: A Review of Programmed Cell Death. *Toxicologic pathology* 35:495-516, 其通过引用整体并入本文。

[0396] 本文所述的方法、组合物和试剂盒可用来分析自体吞噬过程中 DNA 的断裂。

[0397] 检查点、DNA 损伤和细胞周期

[0398] 确定基因座是连接的还是分离的 (断裂的) 可以用来研究 DNA 损伤的修复、双链断裂修复、同源重组、微同源性介导的末端连接、单链退火 (SSA)、断裂诱导的复制或非同源末端连接 (NHEJ)。本文所述的方法可以用来诊断和预测与这些过程相关的疾病。

[0399] DNA 损伤可由环境因素和内源或正常代谢过程引起。可损伤 DNA 的内源因素包括, 例如, 活性氧和复制错误。生理双链 DNA 断裂可包括 V(D)J 重组断裂和类别转换断裂。病理性双链 DNA 断裂可由电离辐射、氧化自由基、跨越切口的复制、脆性位点处的无意酶作用、拓扑异构酶失效和机械应力引起。可引起 DNA 损伤的环境或内源因素包括紫外辐射、X-射线、 γ 射线、DNA 嵌入剂、某些植物毒素、病毒、热破坏和化学治疗。减数分裂细胞可具有额外的 DSB 来源, 包括酶 Spo11。

[0400] 双链 DNA 断裂可通过例如 NHEJ 进行修复。可参与 NHEJ 的因子包括, 例如, Ku70/86、DNA-PKcs、Artemis、 $\text{pol } \mu$ 和 λ 、XRCC4、DNA 连接酶 IV、XRCC4 和 XLF-Cernunnos。在双链 DNA 断裂形成后, Ku 可结合到该断裂形成 DNA 复合物。DNA 末端复合物可以募集核酸酶、聚合酶和连接酶活性。在 DNA 末端的 Ku 可与 DNA-PKcs 形成稳定复合物。DNA-PKcs 可包含 5' 内切核酸酶活性、3' 内切核酸酶活性和发夹打开活性。Artemis 可包含 5' 外切核酸酶活性。PALF (APLF) 的 3' 外切核酸酶可在 NHEJ 中发挥作用。聚合酶 μ 和 λ 可通过其 BRCT 结构域结合 Ku:DNA 复合物。DNA 连接酶 IV 可跨越切口连接、连接不相容 DNA 末端以及连接单链 DNA。NHEJ 可包括链切除。XRCC4 可以四聚化, 并且 PNK (多核苷酸激酶)、APTX (aprataxin, 一种可在中止连接产物的脱腺苷化中发挥作用的蛋白质) 和 PALF 可与 XRCC4 相互作用。通过 NHEJ 进行的双链 DNA 断裂修复在例如 Lieber, M (2011) *Annu. Rev. Biochem.* 79:181-211 中进行了综述, 其通过引用整体并入本文。NHEJ 可在细胞周期的任何时间发生。

[0401] NHEJ 蛋白质在 V(D)J 重组中发挥作用。蛋白质 RAG1 和 RAG2 可在 V(D)J 重组中发挥作用。在完成 V(D)J 重组后, 在 B 细胞中可发生类别转换重组, 并且其可用于改变免疫球蛋白重链基因。该过程可涉及激活诱导脱氨酶 (AID)、RNA 酶 H、尿嘧啶糖基化酶、APE1 和 Exo1。

[0402] 双链 DNA 酶断裂可通过同源性指导的修复 (例如, 同源重组或单链退火) 来进行修复。这些过程中可能涉及的因子的实例包括 RAD50、MRE11、Nbs1 (合称为 MRN 复合物); RAD51 (B、C、D)、XRCC2、XRCC3、RAD52、RAD54B 和 BRCA2。在细胞周期的 S 和 G2 期过程中, 存在两个离得很近的姐妹染色单体, 所以在这些阶段中同源性指导的修复可能更常见。

[0403] ATM 和 ATR 激酶可识别受损的 DNA。这些激酶与 DNA-PK 一起可以使 H2AX 磷酸化并产生 γ H2AX 灶。ATR 可被由复制叉停工 (stalling) 或大损伤处理产生的单链 DNA 区域激活。ATR 可与 ATRIP 相互作用。9-1-1 复合物 (Rad9、Hus1 和 Rad1) 可在 ATR 引起的底物磷酸化中发挥作用。RPA 可结合 ssDNA, 并可在 ATR 引起的底物磷酸化中发挥作用。

[0404] ATM 可通过 MRN 识别 DNA 末端。磷酸化的 H2AX 可募集 MDC1、泛素连接酶 RNF8 和

RNF168 以及 53BP1。ATM 可使 Chk2 和 p53 磷酸化。

[0405] 还可使用本文所述的方法、组合物和试剂盒分析检查点和细胞周期调控。细胞可经历细胞周期,且细胞周期可包含 G1 期、S 期 (DNA 合成)、G2 期和 M 期 (有丝分裂)。停止分裂的细胞可处在 G0 期 (静止期)。检查点可用于使细胞周期停止,并且在允许细胞周期继续前允许对 DNA 损伤的修复。DNA 损伤检查点可发生在 G1 和 S 期的边界以及 G2 和 M 期的边界。另一检查点为 S 期内检查点。

[0406] 其它过程

[0407] 确定核酸是连接的还是分离的 (断裂的) 可用来研究在诸如 DNA 复制和转录的过程中的聚合酶 (例如, DNA 聚合酶、RNA 聚合酶、逆转录酶)。例如,可确定聚合酶的持续合成能力 (例如,以便确定新生链的百分比,即全长相对于部分长度,可以通过计数基因的前一半拷贝数对比后一半拷贝数来测量存在多少截短形式的基因)。因为合成以 5' 到 3' 发生,所以预期会产生比待合成产物的后半 (3' 端) 更多的前一半 (5' 端)。

[0408] 确定样品中基因座是连接的还是分离的 (断裂的) 可用于研究一种或多种限制酶、RNA 酶、DNA 酶、外切核酸酶、内切核酸酶、RNA 酶、DNA 酶等,从而确定这些酶的切割效率 (例如,分离成两个连接的靶序列)。

[0409] 确定样品中基因座是连接的还是分离的 (断裂的) 对研究 RNA 剪接、基因重排、基因定位和癌症中的 DNA 重排可能是有用的。基因重排可以是,例如,染色体易位。易位可以是相互的 (非罗伯逊易位),其可涉及非同源染色体之间的物质交换。易位可以是罗伯逊易位。罗伯逊易位可涉及靠近着丝粒融合的两条近端着丝粒染色体的重排。与疾病相关的易位包括,例如, t (8;14) (q24;a32) (伯基特淋巴瘤;c-myc 与 IGH 的融合); t (11;14) (q13;q32) (套细胞淋巴瘤;细胞周期蛋白 D1 与 IGH 的融合); t (14;18) (q32;q21) (滤泡型淋巴瘤;IGH 与 Bcl-2 的融合); t (10;(各种))(q11;(各种)) (乳头状甲状腺癌;涉及染色体 10 上的 RET 原癌基因); t (2;3) (q13;p25) (滤泡性甲状腺癌;PAX8 与 PPAR γ 1 的融合); t (8;21) (q22;q22) (急性髓母细胞白血病); t (9;22) (q34;q11) 费城染色体 (慢性髓性白血病;急性淋巴细胞白血病;ETO 与 AML1 的融合); t (15;17) (急性早幼粒细胞白血病;PML 与 RAR- α 的融合); t (12;15) (p13;q25) (急性髓性白血病、先天性纤维肉瘤、分泌性乳腺癌;TEL 与 TrkC 受体的融合); t (9;12) (p24;p13) (CML、ALL;JAK 与 TEL 的融合); t (12;21) (p12;q22) (ALL;TEL 与 AML1 的融合); t (11;18) (q21;q21) (MALT 淋巴瘤;Bcl-2 与 MLT 的融合);和 t (1;11) (q42.1;q14.3) (精神分裂症)。

[0410] 本文所述的拷贝数变异分析可用来诊断产前状况,例如胎儿非整倍性,如三倍性 13、三倍性 18 或三倍性 21。

[0411] 确定法医用遗传物质的降解 (断裂) 程度可以帮助在浪费宝贵的样品之前确定什么样的分析可以成功进行。确定核酸是连接的还是分离的 (断裂的) 对于从完美整数拷贝数估计中确定由于 DNA 的随机剪切引起的预期缺陷可能是有用的。

[0412] 物质的共定位 (collocation)

[0413] 本文提供了用于确定微小 RNA 和其它 RNA 的共定位的方法。例如,可将微小 RNA 和其它类型的 RNA 包裹在血液中的外体中。外体可以是可包含蛋白质和 RNA 的 30-90nm 的囊泡。RNA,包括例如微小 RNA,可一起包装在蛋白质复合物中。在一个实施方案中,提供了一种用于确定转录物是共定位于外体中还是蛋白质复合物中的方法。

[0414] 在一个实施方案中,以保护感兴趣的外体或蛋白质复合物的方式,例如通过与甲醛交联,处理血浆或血浆的衍生物,然后将血浆分成多个空间上分离的分区。在分区内可破坏外体,可消化蛋白质,并且可进行 PCR 反应(与逆转录结合)(如,ddPCR)。PCR 反应可扩增至少 2、3、4、5、6、7、8、9、10 种或更多的靶核酸序列。如果靶核酸序列共定位于同一分区,那么靶序列转录可在同一蛋白质复合物中。在另一实施方案中,进行对照实验,其中样品没有经过保护外体或蛋白质复合物的过程。在一个实施方案中,外体的裂解(断裂)可通过温度调节来完成。在另一实施方案中,外体的裂解(断裂)可通过在携带外体或蛋白质复合物断裂剂的分区中释放内部乳液来完成。

[0415] 在另一实施方案中,提供了一种确定两种或多种无细胞 DNA 物质的共定位的方法。无细胞 DNA 分子可在血液或血浆样品中聚集。可将含有聚集的无细胞 DNA 分子的样品分成空间上分离的区域,可破坏聚集,可进行两种或多种不同无细胞 DNA 分子的扩增,并且可对分区进行分析以确定无细胞 DNA 分子是否在相同或不同分区中。

[0416] 本文提供的方法允许确定特定的 RNA、DNA 或蛋白质靶标是否一起在血浆中移动。附加的荧光通道可允许两种或多种(例如 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 种或更多种)靶标的同时共定位测量。例如,使用具有三种不同荧光团的探针可确定在三个靶核酸序列间的共定位频率。

[0417] 检测靶序列的缺失

[0418] 提供了一种用于通过共定位获得连接信息的方法。该方法可用来确定是否存在靶核酸序列的缺失,或用于 CNV 拷贝的单元型分析。在拷贝数变异区域,标记序列(例如用 VIC 标记探针检测)可在靶序列(例如用 FAM 标记探针检测)之外,但靠近靶序列。可将包含核酸的样品分区成多个空间上分离的区域,且可检测标记核酸序列和靶核酸序列(例如,通过扩增和用探针检测)。如图 49 中所示,可分析 VIC(标记序列)和 FAM(靶序列)的共定位。如果 VIC 和 FAM 总是共定位于分区中,那么很可能没有靶序列的缺失(图 49B)。如果存在只含有不与 FAM 共定位的 VIC 的分区,那么该结果提示靶序列的缺失(图 49A)。

[0419] 已消化的核酸的储存

[0420] 已消化的核酸(如,DNA)的储存时间长度可影响拷贝数变异测量值。延长存储可引起估计的拷贝数的减少。例如,延长存储可导致核酸降解。已消化的核酸样品的储存时间长度可以是大约,或小于约,例如 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或 100 小时。已消化的核酸样品的储存时间长度可以是大约,或小于约,例如 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、

63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99 或 100 年。

[0421] 图 20A 和图 20B 图示了储存的、已消化的 DNA 的 MRGPRX1CNV 值的漂移。图 20A 中，已消化的样品显示 CNV 值始终低于整数。图 20B 中，CNV 值更接近于整数（参见实施例 20）。

[0422] 在一个实施方案中，已消化的 DNA 在 4℃ 长时间储存会影响 CN 估计的质量（例如，随着时间的推移，拷贝数估计值会变得更小，例如，估计 CNV 为 6.0 的含有靶序列的样品，如果在 4℃ 储存 3 周，可得到 5.7 的 CNV）。

[0423] 核酸样品（例如，已消化的核酸样品）的储存温度可以是大约或小于约 4、0、-10、-20、-30、-40、-50、-60、-70、-80、-90、-100、-110、-120、-130、-140、-150、-160、-170、-180、-190 或 -200℃。

[0424] 在一个实施方案中，已消化的 DNA 可储存在缓冲溶液（例如，10mMtris, pH8.0）中。

[0425] 核酸长度对 CNV 分析的影响

[0426] 样品中长核酸的存在可影响拷贝数变异数值，即使靶核酸序列不相连（例如，若它们在不同染色体上）。通过例如限制酶切消化、热处理、剪切、超声处理、过滤等使样品中的核酸大小减小，可以改善拷贝数变异实验的结果。核酸长度的减小还可提高靶序列对于 PCR 的可及性。

[0427] 在高核酸负载时，核酸长度的减小可用于确保小滴数字 PCR 实验中一致的小滴形成。在具有长核酸的高核酸负载下，会减少或阻止小滴的形成，从而可产生细流。可通过例如超声处理、热处理、限制酶消化、过滤或剪切来减小核酸长度。

[0428] 小滴数字 PCR 可用于测量限制酶的效率和特异性。

[0429] 本申请出于所有目的通过引用整体并入以下材料：2006 年 5 月 9 日授予的美国专利 7,041,481；2010 年 7 月 8 日公开的公开号为 2010/0173394A1 的美国专利申请；和 Joseph R. Lakowicz, PRINCIPLES OF FLUORESCENCE SPECTROSCOPY (1999 年第二版)。

[0430] 试剂盒

[0431] 本文提供了用于实施本发明提供的方法的试剂盒。试剂盒可包含一种或多种限制酶、装置、缓冲液、试剂和使用说明书。试剂盒可包含限制酶、缓冲液、盐和使用说明书。试剂盒可包含一种或多种引物和一种或多种探针。在一个实施方案中，试剂盒包含至少一种限制酶、四种引物和两种探针。在另一实施方案中，试剂盒包含至少一种限制酶、至少四种引物和至少一种探针。在另一实施方案中，试剂盒包含至少一种限制酶、至少四种引物和至少两种探针。

[0432] 相关技术

[0433] 本文所述的方法中可以使用常规技术。这些常规技术可见于标准实验室手册，如 Genome Analysis:A Laboratory Manual Series (Vols. I-IV), Using Antibodies:A Laboratory Manual, Cells:A Laboratory Manual, PCR Primer:A Laboratory Manual, 和 Molecular Cloning:A Laboratory Manual (均来自 Cold Spring Harbor Laboratory Press)；Stryer, L. (1995) Biochemistry (第 4 版) Freeman, New York；Gait, "Oligonucleotide Synthesis:A Practical Approach" 1984, IRL Press, London, Nelson and Cox (2000), Lehninger, (2004) Principles of Biochemistry 第 4 版, W. H. Freeman Pub., New York, N. Y., 和 Berg 等人 (2006) Biochemistry, 第 6 版, W.

H. Freeman Pub., New York, N. Y., 全部出于所有目的通过引用整体并入本文。

[0434] 拷贝数变异可通过多种手段进行检测,例如,荧光原位杂交、比较基因组杂交、阵列比较基因组杂交、使用 SNP 阵列的虚拟核型分析和下一代测序。例如,在公开号为 20090239308 的美国专利申请中描述了通过数字 PCR 确定拷贝数变异的方法。可通过稀释核酸由数字 PCR 检测拷贝数变异。可通过使用可将单个 DNA 分子分成单独反应室的纳米流体芯片(数字阵列)(例如,Fluidigm 纳米流体芯片),由数字 PCR 检测拷贝数变异。可通过小滴数字 PCR 检测拷贝数变异。本文所述的方法可用于确认用一种或多种上述技术进行的拷贝数变异分析的结果。

[0435] 可用于确定拷贝数变异的下一代测序技术包括,例如,DNA 纳米球测序(使用滚环复制将基因组 DNA 的小片段扩增成 DNA 纳米球)(例如,Complete Genomics 使用的),纳米孔测序(例如,Oxford Nanopore Technologies 使用的)(Soni G. V. 和 Meller A. (2007) Clin. Chem. 53:1996-2001),离子半导体测序(Ion Torrent Systems)(公开号为 20090026082 的美国专利申请),SOLiD 测序(连接测序;例如,Applied Biosystems 使用的),Illumina(Solexa)测序(使用桥式扩增),454 焦磷酸测序(例如,Roche Diagnostics 使用的)(Margulies, M. 等人 2005 Nature, 437:376-380),真正单分子测序(例如,Helicos 使用的)(Harris T. D. 等人 (2008) Science 320:106-109),或由 Pacific Biosciences 使用的单分子实时测序(SMRT)。本文所述的方法、组合物和/或试剂盒可用于追踪通过这些方法中的一种方法进行的 CNV 分析。

[0436] 范围在本文中 can 表示为从“大约”一个特定值,和/或到“大约”另一个特定值。当表示这种范围时,另一实施方案包括从一个特定值和/或到另一特定值。同样,当通过使用先行词“约”将数值表示为近似值时,将理解成该特定值形成了另一实施方案。应当进一步理解,每个范围的端点有关于另一端点以及独立于另一端点都是有意义的。本文所用的术语“约”是指在特定使用环境内规定数值 $\pm 15\%$ 的范围。例如,约 10 包括从 8.5 到 11.5 的范围。

[0437] 除非另外限定,本文所用的所有技术和科学术语都具有与本发明所属领域普通技术人员通常理解的相同的含义。虽然与本文所述的方法和材料相似或相当的任何方法和材料也可用于本发明的实施或测试中,但现在对代表性示例性方法和材料进行描述。

实施例

[0438] 实施例 1:淀粉酶 CNV 分析

[0439] 本文所述的方法可用于确定淀粉酶基因的拷贝数变异。淀粉酶是可以水解多糖的键的酶。组间拷贝数变异提示基于饮食的进化变化。淀粉酶试验产生了对应于不同 CNV 状态的离散估计值。使用 15 个 Coriell 参考 DNA 样品获得的结果证明了人类该基因拷贝数的高变异性。参见图 21。

[0440] 使用来自 Coriell 医学研究所的样品。用 10U/ μ g DNA 的 AluI 在 37°C 消化样品 DNA 1 小时。将约 20-100 μ g 已消化的 DNA 加载到每个 ddPCR 反应中。一式三份处理每个样品,并且将数据合并得到如图 21 所示的数值。用于淀粉酶试验的正向引物的序列为 5'-TTCTGAGATTTATCTAGAGGCTGGGA-3',反向引物为 5'-CCCTGACAGACCGACAAGAC-3',且探针为 5'-6FAM-CTGGTTCAGGAGCCCT-MGB-NFQ-3'。将该试验双链体化于 (duplexed to) RPP30:正向

引物:5'-GATTTGGACCTGCGAGCG-3',反向引物5'-GCGGCTGTCTCCACAAGT-3'及探针5'-VIC-CTGACCTGAAGGCTCT-MGB-NFQ-3'。使用下述热循环条件:95℃ 10分钟,(94℃ 30秒+60℃ 1分钟)55个循环。

[0441] 实施例2:CCL3L1CNV分析

[0442] 本文所述的方法可用来确定CCL3L1基因的拷贝数变异。CCL3L1的拷贝数变异已与HIV-1易感性相关,但是由于使用标准qPCR对拷贝数评估的不准确性,这些关联出现不一致。CCL3L1基因簇包括5个假基因,它们会使特异性扩增变得困难。基于整数的CNV分析增加了正确定量的信心。Alkan等人((2009)Nat. Genetics. 41:1061-1067)发现NA18507具有5.7个CCL3L1拷贝。在此处的分析中,发现了5.94个拷贝,误差棒越过了“6”。参见图22。本文所述的方法提供了特别设计试验以针对基因和特别选择限制酶以释放该基因的能力,同时也切割假基因以降低抬高拷贝数估计的假阳性的可能性。

[0443] 使用57℃的退火温度可导致一些假基因被扩增,且CNV值比预期值约大1。

[0444] 对于CCL3L1,将各815ng纯化的人类基因组DNA样品(Coriell)用7.5单位的MseI(NEB)在10μL中在37℃下消化1小时。将上述消化物用TE缓冲液稀释3.5倍达到35μL,然后每20μL ddPCR反应分析69ng(3μL)。经修饰的CCL3L1分析序列19是(正向引物)5'-GGGTCCAGAAATACGTCAGT-3';(反向引物)5'-CATGTTCCCAAGGCTCAG-3';和(探针)6FAM-TTCGAGGCCAGCGACCTCA-MGBNFQ。将所有CNV试验都用RPP30参考试验双链体化(正向引物)5'-GATTTGGACCTGCGAGCG-3';(反向引物)5'-GCGGCTGTCTCCACAAGT-3';和(探针)VIC-CTGACCTGAAGGCTCT-MGBNFQ。热循环条件为95℃ 10分钟(1个循环),94℃ 30秒和60℃ 60秒(40个循环),98℃ 10分钟(1个循环),和12℃保持。也参见Sudmant, P. H.;Kitzman, J. O.;Antonacci, F.;Alkan, C.;Malig, M.;Tsalenko, A.;Sampas, N.;Bruhn, L.;Shendure, J.;Eichler, E. E. Science2010, 330, 641-646。

[0445] 实施例3:MRGPRX1CNV分析

[0446] 本文所述的方法可用来确定MRGPRX1基因的拷贝数变异。与Mas相关的G蛋白偶联受体成员X1是由MRGPRX1基因编码的人类蛋白质。MRGPRX1是与痒和痛觉有关的感觉神经元特异性受体。据认为MRGPRX1蛋白质参与阻断疼痛,因此它的缺失与慢性疼痛有关。图23A和图23B图示了MRGPRX1基因的拷贝数变异。

[0447] 对于MRGPRX1,各4.4μg纯化的人类基因组DNA样品(Coriell)用10单位的RsaI(NEB)在50μL中在37℃消化1小时。将消化物用TE缓冲液(pH8.0)稀释8倍到400μL,然后每20μL ddPCR反应分析33ng(3μL)。MRGPRX1分析序列是(正向引物)5'-TTAAGCTTCATCAGTATCCCCCA-3',(反向引物)5'-CAAAGTAGGAAAACATCATCACAGGA-3',和(探针)6FAM-ACCATCTCTAAAATCCT-MGBNFQ。将CNV试验用RPP30参考试验双链体化(正向引物)5'-GATTTGGACCTGCGAGCG-3',(反向引物)5'-GCGGCTGTCTCCACAAGT-3',和(探针)VIC-CTGACCTGAAGGCTCT-MGBNFQ。热循环条件为95℃×10分钟(1个循环),94℃×30秒和60℃×60秒(40个循环),98℃×10分钟(1个循环),和12℃保持(也参见Hindson B等人(2011)Anal. Chem. 83:8604-8610)。

[0448] 实施例4:CYP2D6CNV分析

[0449] 本文所述的方法可以用来确定CYP2D6基因的拷贝数变异。细胞色素P4502D6(CYP2D6),即细胞色素P450混合功能氧化酶体系的一个成员,参与体内生物异源

物质的代谢。CYP2D6 的拷贝数与代谢药物有关。CYP2D6 功能过低至无功能会导致药物代谢不良。高水平的 CYP2D6 功能会导致大量的药物代谢。CYP2D6 基因的多个拷贝可以表达并导致高于正常水平的 CYP2D6 功能（超速代谢型）。图 24A 和图 24B 图示了 CYP2D6 的拷贝数变异。

[0450] 对于 CYP2D6, 各 4.4 μ g 纯化的人类基因组 DNA 样品 (Coriell) 用 10 单位的 RsaI (NEB) 在 50 μ L 中 37 $^{\circ}$ C 消化 1 小时。将消化物用 TE 缓冲液 (pH8.0) 稀释 8 倍到 400 μ L, 然后每 20 μ L ddPCR 反应分析 33ng (3 μ L)。购买 CYP2D6 (Hs00010001_cn) 作为 20 \times 引物预混合物和 FAM-MGBNFQ 探针 (Applied Biosystems)。将 CNV 试验用 RPP30 参考试验双链体化 (正向引物) 5' -GATTTGGACCTGCGAGCG-3', (反向引物) 5' -GCGGCTGTCTCCACAAGT-3', 和 (探针) VIC-CTGACCTGAAGGCTCT-MGBNFQ。热循环条件为 95 $^{\circ}$ C \times 10 分钟 (1 个循环), 94 $^{\circ}$ C \times 30 秒和 60 $^{\circ}$ C \times 60 秒 (40 个循环), 98 $^{\circ}$ C \times 10 分钟 (1 个循环), 和 12 $^{\circ}$ C 保持。(也参见 Hindson B 等人 (2011) Anal. Chem. 83:8604-8610)。

[0451] 实施例 5: 脊髓性肌萎缩 CNV 分析

[0452] 脊髓性肌萎缩 (SMA) 可能是由脊髓区域的运动神经元损失引起的, 该运动神经元损失归因于生存运动神经元蛋白的低含量, 该运动神经元蛋白由 SMN1 和 SMN2 编码。SMN1 和 SMN2 经历缺失和扩增。除了一个核苷酸之外, 这些基因是相同的, 而这个核苷酸降低了由 SMN2 转录的全长转录物的水平。SMA 可能是由 SMN1 拷贝的损失引起的。疾病严重程度可能与 SMN2 补偿 SMN1 损失的程序相关。可以设计试验用于等位基因特异性基因检测和区别 (例如, 包括 ddPCR)。

[0453] 遗传可以是常染色体隐性的。出生人数的 1/6,000 可能受到 SMA 的影响; 1/40 的人是缺陷 SMN1 的携带者。0 和 1 型可导致 2 岁前死亡。2-4 型可导致 2 岁后发病 (重度至轻度的无力)。

[0454] 图 25 和图 26 图示了用简单梯度板确定的最佳共同退火温度。图 27 图示了 SMN1 和 SMN2 的拷贝数。

[0455] 独立双链体 (SMN1/Her2 和 SMN2/Tert) 可用来确定 SMN1 和 SMN2 的拷贝数。来自双链体的 SMN1/SMN2 拷贝的比值可用来确认测定的拷贝数的值。可通过三角测量法进行验证 (例如, 参见图 28 中的上方示意图)。三角测量法可意味着可通过使用两种不同的参照物评估 CNV 来验证靶序列的 CNV。例如, 可通过使两种不同的参照物一起形成双链体来测定它们的比值 (CNV), 从而避免使用不含预期的两个拷贝 / 二倍体基因组的参照物来估计 CNV 的可能性。在确定 CNV 的过程中可评估三种不同靶序列的拷贝数, 从而使得 CNV 估计中产生错误的程序大大降低。

[0456] 对于该实验, 使用用 Her2 (FAM) 试验双链体化的 SMN1 (VIC) 来评估 SMN1 的拷贝数。用 CviQI 消化约 50ng/ 试验的 DNA。执行标准 PCR 方案。使用用与试验双链体化的 SMN2 (FAM) 来确定 SMN2 的拷贝数, 从而检测 TERT (VIC)。SMN1/SMN2 的比值是通过进行 SMN1 (VIC) 和 SMN2 (FAM) 的双链体化来确定的。使用大致 50ng/ 试验, 用 CviQI 消化 DNA。热循环条件为 95 $^{\circ}$ C 10 分钟 (1 个循环), 94 $^{\circ}$ C 30 秒和 60 $^{\circ}$ C 60 秒 (40 个循环), 98 $^{\circ}$ C 10 分钟 (1 个循环), 和 12 $^{\circ}$ C 保持。

[0457] 图 28 图示了 SMN1 和 SMN2 的拷贝数变异。

[0458] 通过使用相同的引物对并改变探针, 可进行 SMN 基因的检测。常用

引物对为正向引物:5'-ATAGCTATTTTTTTAACTTCCTTTATTTTCC-3';反向引物5'-TGAGCACCTTCCTTCTTTTTGA-3'。SMN1探针序列是:5'-VIC-TTGTCTGAAACCCTG-MGB-NFQ-3',而SMN2探针序列是:5'-6FAM-TTTTGTCTAAAACCC-3'。Tert参考基因的正向引物是5'-GGCACACGTGGCTTTTCG-3';反向引物是5'-GGTGAACCTCGTAAGTTTATGCAA-3'。探针是5'-VIC-TCAGGACGTCGAGTGGACACGGTG-MGB-NFQ-3'。Her2分析序列是正向引物5'-CCCTGAGCAAAGAGTCACAGATAAA-3',反向引物5'-TGCCAGGGTCTGAGTCTCT-3',探针:5'-6FAM-ACTGCGTTTGTCTGG-MGB-NFQ-3'。

[0459] 实施例6:Her2CNV分析

[0460] 30%的乳腺癌具有Her2扩增。Her2+癌是极富侵袭性的,但可对赫塞汀有响应。对HER2水平进行了广泛测试,但是现有技术存在许多缺点。Her2,又称为ERBB2,距GRB720KB,GRB7在数种癌症中也经常过度表达。针对两个临床相关基因的试验可以验证每个拷贝数的测定结果。不出所料,一个样品只对Her2扩增,而另一个展示出镶嵌现象。参见图29。

[0461] 从每个正常和肿瘤乳腺组织样品(D8235086-1, Biochain)纯化的DNA(20ng)用0.2单位的NlaIII在10 μ L中37 $^{\circ}$ C消化1小时。将限制酶切的DNA以每20 μ L ddPCR反应8.8ng(4.4 μ L)直接加入到ddPCR主混合物中。ERBB2(Hs02803918_cn)和GRB7(Hs02139994_cn)分析作为20 \times 引物的预混合物和FAM-MGBNFQ探针(Applied Biosystems)购买,并用上述的RPP30参考试验双链体化。热循环条件为95 $^{\circ}$ C \times 10分钟(1个循环),94 $^{\circ}$ C \times 30秒和60 $^{\circ}$ C \times 60秒(40个循环),98 $^{\circ}$ C \times 10分钟(1个循环),和12 $^{\circ}$ C保持(也参见Hindson B等人(2011)Anal. Chem. 83:8604-8610)。

[0462] 实施例7:连锁分析

[0463] 如果对于消化后的特定分析,靶核酸序列的拷贝数(CN)是2.0,那么不确定是(A)在一条染色体上存在2个靶序列拷贝,而另一条上为0个;还是(B)在两条不同染色体的每一条上有1个靶核酸序列拷贝。用未消化的DNA进行的试验可用来分辨这两种可能。使用未消化的DNA,原则上,如果排列是(A)(两个靶核酸序列拷贝在一条染色体上),那么应该获得的CN值是1,如果排列是(B)(两条不同染色体的每一条上有一个拷贝),那么CN值为2。因为样品中的DNA可能是断裂的,所以结果可能不准确——(A)可能得到CN高于1.0的读数,但可能显著小于2.0;(B)应该精确地得到2.0。

[0464] 起始材料的更高断裂会使得(A)中的CN读数更接近2.0。作为一个例子,对于给定的试验,预测连接的拷贝相距约10kb,并且基于我们的断裂分析,10kb区段的30%是断裂的。在这种情况下,情形(A)应该产生1.3的读数,而情形(B)应该产生2.0的读数。

[0465] 实施例8:连锁分析

[0466] 如果消化后的CN读数是3.0,那么不确定是(A)在一条染色体上存在3个拷贝,而另一条上有0个拷贝;还是(B)一条上有2个拷贝,而另一条上有1个拷贝。如果对未消化的DNA进行相同的试验,并且如上假设相距10kb和30%断裂的相同参数,那么情形(A)会得到1.6个拷贝(=0.7 \times 0.3 \times 2+0.7 \times 0.3 \times 1+0.3 \times 0.7 \times 2+0.3 \times 0.3 \times 3)的读数,而(B)会得到2.3个拷贝的读数。

[0467] 对于数字PCR-实时杂合形式(例如,Life's Biotrove阵列),可尝试从未消化的DNA中提取额外的连接信息。例如,通过分析每个分区内的实时曲线,可以估计有多少分区含有两个、三个等靶序列拷贝。

[0468] 可以检查终点荧光,来确定含有两个靶核酸序列拷贝的区段是否比含有一个靶核酸序列拷贝的区段具有更高的振幅。在情形(A)下,应该有较少的阳性值,但是这些阳性值应该平均具有更高的荧光。

[0469] 可以调整循环数来获得最佳的分离(例如,通过使得携带一个拷贝的区段不能到达终点)。

[0470] 实施例 9: 确认 ddPCR 可用于对 CNV=2 的样品进行单元型分析(例如,区分顺式和反式构型的基因拷贝)

[0471] 染色体单元型分析的策略是 1) 比较已消化样品和未消化样品的 CNV(例如使用 ddPCR); 2) 当小滴经历指数扩增时(在任何试剂变为限制性的之前—这可能在 PCR 后期循环中发生),检测其平均荧光强度;和 3) 进行长片段 PCR 以检测相同染色体上(顺式构型)两个基因(或靶序列)拷贝的存在,或者,如果基因(或靶序列,或重复区域)过长,那么检查另一条染色体以确认目标基因(或靶序列)的缺失(例如,跨越缺失部分的 PCR 应该工作,但是基因(或靶序列)在适当位置的 PCR 却不会工作,因为引物之间的距离太大)。

[0472] 总方案指南

[0473] 1) ddPCR: 将限制酶消化的样品与未消化的样品进行比较,检查结果发现,在具有顺式构型拷贝的样品中,对于未消化的样品来说估计的 CNV 显著下降。DNA 可包含在泊松统计的最有效点(sweet spot)中(在该点,泊松校正因子有最小的统计误差)。对于 20,000 个小滴,最有效点可以是~ 1.6cpd(cpd:每个小滴的拷贝数)。

[0474] 2) 基于荧光的确认:在适当的一个循环或多个循环时(可凭经验确定,且根据试验效率每个试验可不同),可以检查含有顺式构型样品的小滴的平均强度是否高于含有反式构型样品的小滴的平均强度,如果在相同小滴中含有两个相连的拷贝则会发生这种情况。对于该实验,可以使用低 cpd。可以以 0.1cpd 为目标以使两个不相连的拷贝存在于同一小滴中的可能性最小化。使用良好的光学仪器,可以观察到荧光强度的条带(含有两个拷贝的小滴应该比仅含有一个拷贝的小滴明亮两倍)。可在例如约第 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49 或 50 次循环时检测荧光。可在多于一次循环时检测荧光。例如,可在 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49 或 50 次不同的循环时检测荧光。例如,可在第 25、28、31 和 40 次循环时测定荧光。

[0475] 3) 通过长片段 PCR 对连接的确认。通常, CNV 中扩增区域的大小可能太大而不允许两个串联定位的基因(或靶序列)的 PCR 扩增;因此,可能较容易检测基因(或靶序列)缺失而不是两个并排的靶序列拷贝的存在。对于该分析,可选择处于疑似被扩增的区域之外的彼此相对的引物。可以使用长片段聚合酶。可使用适当长度的延伸时间(注意时间太长或太短都会导致反应失败)进行 PCR。该长片段 PCR 可整体进行或在小滴中进行。

[0476] 图 30 图示了 6 个样品的拷贝数估计,每个用限制酶切割和不切割。数据提示样品 4 和 5 有两个靶序列连接在同一染色体上,而其它样品在两条染色体上各含有一个靶序列。

[0477] 样品 D1234090(结肠)、D1234106(食道)和 D1234152(肺)来自 BioChain Institute Inc。样品 NA18853、NA19108 和 NA18507 来自 Coriell 医学研究所。Coriell 的

样品包含 B- 淋巴细胞 DNA。

[0478] 每微克 DNA 使用约 10U 的 RsaI, 并且在 37°C 下持续消化 1 小时。在 40 个循环后用小滴读取器读取小滴的荧光。

[0479] MRGPRX1 分析序列可为 (正向引物) 5' -TTAAGCTTCATCAGTATCCCCCA-3' ; (反向引物) 5' -CAAAGTAGGAAAACATCATCACAGGA-3' ; 和 (探针) 6FAM-ACCATCTCTAAAATCCT-MGBNFQ。CNV 试验可用 RPP30 参考试验双链体化 (正向引物) 5' -GATTTGGACCTGCGAGCG-3' , (反向引物) 5' -GCGGCTGTCTCCACAAGT-3' , 和 (探针) VIC-CTGACCTGAAGGCTCT-MGBNFQ。热循环条件可以是 95°C × 10 分钟 (1 个循环), 94°C × 30 秒和 60°C × 60 秒 (40 个循环), 98°C × 10 分钟 (1 个循环), 和 12°C 保持 (也参见 Hindson B 等人 (2011) Anal. Chem. 83:8604-8610)。

[0480] 图 31 图示了在 25 个循环后, ddPCR 实验中对 MRGPRX1 和 RPP30 的荧光强度的分析。事件数指的是小滴计数。

[0481] 图 32 图示了在第 25 次循环时, ddPCR 实验中 4 个样品的一式三份测量的平均荧光。疑似在一条染色体上具有两个靶序列而在另一条上具有 0 个靶序列的样品 (NA18853 和 NA19108) 比疑似在一条染色体上具有一个靶序列拷贝的样品 (反式构型) (NA18507 和 D1234106) 具有更强的荧光。

[0482] 图 33 图示了在 28 次循环后, ddPCR 实验中对 MRGPRX1 和 RPP30 荧光强度的分析。

[0483] 图 34 图示了 ddPCR 浓度和拷贝数变异的比较。在样品 NA18853 和 NA19108 中, 在第 28 次循环时观察到未切割的样品的 CNV 下降。示出了 VIC (RPP30) 浓度、FAM (MRGPRX1) 浓度和 MRGPRX1 拷贝数的信号。

[0484] 图 35 图示了在 28 次循环后, ddPCR 实验中 4 个样品的一式三份测量的平均荧光。疑似在一条染色体上具有两个靶序列而在另一条上具有 0 个靶序列的样品 (18853 和 19108) 比疑似在两条染色体中的每一条上具有一个靶序列拷贝的样品 (反式构型) (18507 和 BC106) 具有更强的荧光。

[0485] 图 36 图示了在 31、34 和 40 次循环后, ddPCR 实验中 4 个样品的一式三份测量的平均荧光。在第 31 次循环时, 而不是在第 34 或 40 次循环时, 疑似在一条染色体上具有两个靶序列而在另一条上具有 0 个靶序列的样品 (NA18853 和 NA19108) 比疑似在两条染色体中的每一条上具有一个靶序列拷贝的样品 (反式构型) (NA18507 和 D1234106) 具有更强的荧光。

[0486] 总之, ddPCR 实验支持了这样的观点, 即样品 18853 和 19108 在单一染色体上具有两个相连的 MRGPRX1 的拷贝。

[0487] 还进行了长片段 PCR 实验来支持从小滴数字 PCR 实验中得到的有关靶序列连接的发现。CN=1 的样品可作为缺失区域扩增的对照。

[0488] 图 37 图示了用于长片段 PCR 实验的实验设置。对于情形 A 和 B, 预计长片段 PCR 会失败, 因为引物之间的距离太大。然而, 对于含有具有 MRGPRX1 缺失的染色体的情形 C, PCR 应该会成功。箭头指示引物。

[0489] 图 38 图示了长片段 PCR 实验的生物分析仪图像, 该实验利用 6 个样品通过 ddPCR 来确认连接 (设置示意图参见图 39)。每个泳道中在约 1500 处的条带是 MRGPRX1 PCR 产物 (MRGPRX1 序列的内部引物)。引物也设计为在 MRGPRX1 序列外部退火。样品 11994 和 18573 具有拷贝数 1: 一条染色体含有一个 MRGPRX1 拷贝, 而另一条含有缺失。3000 和 1500 之间

的 PCR 产物可能来自于从具有缺失的染色体的扩增。样品 19239 和 18507 在不同染色体上具有两个 MRGPRX1 拷贝。这些样品在 3000 到 1500 之间没有条带,这可能是因为 MRGPRX1 基因的存在使得 PCR 太长而不能生成产物 (25KB)。样品 19108 和 18553 在相同染色体和具有 MRGPRX1 缺失的染色体上具有两个 MRGPRX1 拷贝。3000 和 1500 之间的 PCR 产物可能来自从具有缺失的染色体的扩增。这些数据确认了小滴数字 PCR 实验的结果。

[0490] 图 39 图示了 6 个样品的长片段 PCR 结果和 MRGPRX1 拷贝数结果。顶端的示意图图示了 PCR 引物的位置和 PCR 产物的预期长度。样品 S1 和 S2 在一条染色体上具有一个 MRGPRX1 拷贝,而在另一条上具有缺失。引物在 MRGPRX1 基因的内部并且可以产生约 1.6kb 的产物。引物还在 MRGPRX1 基因的外部退火。样品 S3 和 S4 有两个反式的 MRGPRX1 拷贝 (在不同的染色体上)。来自最外层引物的 PCR 产物估计约为 24kb。样品 S5 和 S6 在一条染色体上具有两个 MRGPRX1 拷贝,而在另一条染色体上具有 MRGPRX1 缺失。使用具有缺失的染色体上的最外层引物产生的 PCR 产物的长度约为 2.7kb。在 PCR 设置示意图的下方是长片段 PCR 结果的图示。所有样品都含有 PCR 产物,该产物相当于在 1.6kb 由内部 MRGPRX1 引物得到的产物。样品 S1、S2、S5 和 S6 具有从具有染色体缺失的样品得到的约 2.7kb 的 PCR 产物。样品 S3 和 S4 不具有约 2.7kb 的 PCR 产物。使用包含一个或两个 MRGPRX1 拷贝的染色体上最外层引物的 PCR,由于引物之间的距离而失败。在长片段 PCR 图示下方是在使用或不使用限制酶消化的情况下通过 ddPCR 确定的 MRGPRX1 拷贝数的图示 (RE= 限制酶消化 ;ND= 未消化)。估计样品 S1 和 S2 具有 1 个 MRGPRX1 拷贝,无论样品是否用限制酶消化。估计样品 S3 和 S4 具有约 2 个 MRGPRX1 拷贝,不论样品是否被消化。如果样品被消化,那么估计样品 S5 和 S6 具有约 2 个 MRGPRX1 拷贝,但是如果样品未被消化,那么小于两个拷贝。在 S5 和 S6 中,拷贝数估计的差异取决于样品是否被消化,这提示两个 MRGPRX1 拷贝在同一染色体上。

[0491] 长片段 PCR 的方案如下 :用下面的热循环条件来使用 AgilentPfuUltraII Fusion HS DNA 聚合酶 :2 分钟 95°C (1 个循环),20 秒 95°C,20 秒 60°C,75 秒 72°C (40 个循环),3 分钟 72°C (1 个循环)。在 50 μ L 反应体积中包括 200nM 的引物、250uM 的 dNTPs、100ng 的 DNA、1uL PfuUltraII fusion HS DNA 聚合酶、Agilent 提供的 1X 主混合物。正向引物的序列为 5' -GATCTAGCTAAGAGACAGAGATAGACACATG-3',反向引物的序列为 5' -cagtatttttgcact gctttcctcat-3'。

[0492] 图 40 图示了与具有估计的拷贝数的已消化样品的百分比差异。

[0493] 实施例 10 :用于确定断裂的算法

[0494] 在该实施例中,分析了两种不同类型的靶核酸。一种利用 FAM 探针进行检测,一种利用 VIC 进行检测。假设两个靶核酸序列在同一多核苷酸上。在一个样品中,可以存在三种类型的 DNA 片段 :1) 连在一起的 Fam-Vic (未断裂),2)Fam 片段,和 3)Vic 片段。观察到一些可能性 (在 FAM-VIC 交会图中的计数),并且目的是推断浓度。首先完成正向。给定浓度,计算计数。然后进行反向,尝试不同的浓度值,并选择给出实际计数的一个。

[0495] N=20000 ;

[0496] A=10000 ;

[0497] B=20000 ;

[0498] AB=10000 ;% 连接起来

```

[0499] cA=A/N ;
[0500] cB=B/N ;
[0501] cAB=AB/N ;
[0502] fprintf(1, '%f%f%f\n', cAB, cA, cB) ;
[0503] pA=1-exp(-cA) ;
[0504] pB=1-exp(-cB) ;
[0505] pAB=1-exp(-cAB) ;
[0506] % 在交会图中 A 是 X 且 B 是 Y
[0507] p(2, 1)=(1-pA)*(1-pB)*(1-pAB) ;% 左下
[0508] p(2, 2)=pA*(1-pB)*(1-pAB) ;% 右下
[0509] p(1, 1)=(1-pA)*pB*(1-pAB) ;% 左上
[0510] p(1, 2)=1-p(2, 1)-p(2, 2)-p(1, 1) ;% 右上
[0511] disp(round(p*N)) ;
[0512] % 也直接计算边际值 (marginals)
[0513] cAorAB=(A+AB)/N ;%=c_A+c_AB ;
[0514] cBorAB=(B+AB)/N ;%=c_B+c_AB ;
[0515] pAorAB=1-exp(-cAorAB) ;% 也可由 p 计算
[0516] pBorAB=1-exp(-cBorAB) ;
[0517] % 反向
[0518] H=p*N ;% 我们得到一些命中 (hits)
[0519] %H=[08000 ;20000] ;
[0520] % 计算 prob
[0521] estN=sum(H(:)) ;
[0522] i_p=H/estN ;
[0523] i_pAorAB=i_p(1, 2)+i_p(2, 2) ;
[0524] i_pBorAB=i_p(1, 1)+i_p(1, 2) ;
[0525] i_cAorAB=-log(1-i_pAorAB) ;
[0526] i_cBorAB=-log(1-i_pBorAB) ;
[0527] maxVal=min(i_cAorAB, i_cBorAB) ;
[0528] delta=maxVal/1000 ;
[0529] errArr=[] ;
[0530] gcABArr=[] ;
[0531] 对于 gcAB=0 :delta :maxVal
[0532] gcA=i_cAorAB - gcAB ;
[0533] gcB=i_cBorAB - gcAB ;
[0534] gpA=1-exp(-gcA) ;
[0535] gpB=1-exp(-gcB) ;
[0536] gpAB=1-exp(-gcAB) ;
[0537] gp(2, 1)=(1-gpA)*(1-gpB)*(1-gpAB) ;% 左下

```

```
[0538] gp(2,2)=gpA*(1-gpB)*(1-gpAB);% 右下
[0539] gp(1,1)=(1-gpA)*gpB*(1-gpAB);% 左上
[0540] gp(1,2)=1-gp(2,1)-gp(2,2)-gp(1,1);% 右上
[0541] gH=gp*estN;
[0542] err=sqrt(sum((H(:)-gH(:)).^2));
[0543] errArr=[errArr;err];
[0544] gcABArr=[gcABArr;gcAB];
[0545] end
[0546] figure,plot(gcABArr,errArr);
[0547] minidx=find(errArr==min(errArr(:)));
[0548] minidx=minidx(1);
[0549] estAB=gcABArr(minidx);
[0550] estA=i_cAorAB - estAB;
[0551] estB=i_cBorAB - estAB;
[0552] fprintf(1,'%f%f%f\n',estAB,estA,estB);
[0553] gpA=1-exp(-estA);
[0554] gpB=1-exp(-estB);
[0555] gpAB=1-exp(-estAB);
[0556] gp(2,1)=(1-gpA)*(1-gpB)*(1-gpAB);% 左下
[0557] gp(2,2)=gpA*(1-gpB)*(1-gpAB);% 右下
[0558] gp(1,1)=(1-gpA)*gpB*(1-gpAB);% 左上
[0559] gp(1,2)=1-gp(2,1)-gp(2,2)-gp(1,1);% 右上
[0560] gH=gp*estN;
[0561] disp(round(gH));
[0562] % 使用模拟确认结果
[0563] numMolA=round(estA*estN);
[0564] numMolB=round(estB*estN);
[0565] numMolAB=round(estAB*estN);
[0566] A=unique(randsample(estN,numMolA,1));
[0567] B=unique(randsample(estN,numMolB,1));
[0568] AB=unique(randsample(estN,numMolAB,1));
[0569] U=1:estN;
[0570] notA=setdiff(U,A);
[0571] notB=setdiff(U,B);
[0572] notAB=setdiff(U,AB);
[0573] AorBorAB=union(A,union(B,AB));
[0574] none=setdiff(U,AorBorAB);
[0575] simcount(2,1)=length(none);
[0576] simcount(2,2)=length(intersect(A,intersect(notB,notAB)));
```

[0577] `simcount(1,1)=length(intersect(B,intersect(notA,notAB)))` ;

[0578] `simcount(1,2)=length(AorBorAB)-simcount(2,2)-simcount(1,1)` ;

[0579] `disp(simcount)` ;

[0580] 实施例 11 :里程碑试验分析—断裂概率

[0581] 问题陈述

[0582] 如果两个不同的基因座位于不同的分子上,那么可能存在两种物质(对应于 FAM 和 VIC 探针)。如果不同的基因座位于同一分子上,那么可能存在三种物质—断裂的 FAM、断裂的 VIC 和连接的 FAM-VIC。(参见图 411。

[0583] 由于存在两种染料,所以可能存在不明确性。需要计算所有三种物质的浓度。

[0584] 算法 :获得 FAM 和 VIC 计数的 2x2 表格。犹如存在一种物质一样,计算断裂的 FAM 和连接的 FAM-VIC 的浓度。犹如存在一种物质一样,计算断裂的 VIC 和连接的 FAM-VIC 的浓度。试用不同浓度的连接的 FAM-VIC(从中可以得出断裂的 FAM 和 VIC 的浓度),并且找到概率表与观察到的计数的最佳拟合 :

| | | |
|-------|-------------------|---------------|
| | FAM - | FAM+ |
| VIC+ | $(1-f)V(1-c)$ | 1 - 其它的总和 |
| VIC - | $(1-f)(1-V)(1-c)$ | $f(1-V)(1-c)$ |

断裂概率 (%)

| | | | |
|------------|------|------|------|
| | | | |
| 1k 未切割 | 6 | 6 | |
| 10K 未切割 | 29.4 | 29.8 | 29.5 |
| 100K 未切割 | 98.7 | 97.7 | 99.9 |
| 1K 注射器 | 11-4 | 11-1 | 11 |
| 10K 注射器 | 87.2 | 89.9 | 91.7 |
| 100K 注射器 | 100 | 100 | 100 |
| 1K Hae III | 100 | 100 | 100 |

[0585] 接下来的步骤可包括观察是否可容易地推导出闭公式和 / 或与 OTools 集成。

[0586] 实施例 12 :断裂分析

[0587] 使用 ddPCR,可进行针对两个基因组基因座例如在共同染色体上的两个基因的双重反应。根据小滴的荧光,可将小滴分成四个群体 (FAM+ / VIC+、FAM+ / VIC-、FAM- / VIC+ 和 FAM- / VIC-1。通过比较这些群体的小滴数,可能确定靶序列共同分离至同一小滴中的频率。使用泊松统计学,可以估计与两个分离的拷贝恰巧在同一小滴中的情况相比,彼此实际相连的种类的百分比。

[0588] 设计试验,其中基因座与共同的参照物 (RPP30) 相距 1K、3K、10K、33K 和 100K。已经进行了两个基因座相距 1K、10K 和 100K 的研究。通过用这三种双链体 (或仅一种双链体) 处理未切割 (未经限制酶消化) 的 DNA, 并且计数四种不同的小滴群体, 可使用统计分析来评估遗传物质的断裂状态。这些数据可用来帮助解释为什么拷贝数变异研究中的 95% 置信界限不总是跨越整数值。

[0589] 实施例 13 : 用于 DNA 断裂计算或用于数字 PCR 多路复用的算法

[0590] 靶序列之间的完全断裂

[0591] 两个 DNA 靶序列 T1 和 T2 分别对应于两种染料 FAM 和 VIC。在该实施例中, T1 和 T2 总是在单独的 DNA 片段上。具有 T1 和 T2 靶序列的 DNA 片段数分别是 M1 和 M2。参见图 42A。

[0592] 在具有多个分区的数字 PCR 实验中, FAM 和 VIC 阳性分区的计数分别为 N1 和 N2。N1 和 N2 将分别小于 M1 和 M2, 因为在分区中可能存在多个 DNA 片段。分区总数为 N。预计的分区计数在表 2 中示出。

[0593] 表 2. 分区计数

| | VIC 阴性 | VIC 阳性 | 总计 |
|--------|---------------------------|---------------------|--------|
| FAM 阳性 | $N1 * (N - N2) / N$ | $N1 * N2 / N$ | N1 |
| FAM 阴性 | $(N - N1) * (N - N2) / N$ | $(N - N1) * N2 / N$ | N - N1 |
| 总计 | N - N2 | N2 | N |

[0594] 如果将观察到 FAM 阳性分区的概率表示为 $p1 = N1 / N$, 并且将观察到 VIC 阳性分区的概率表示为 $p2 = N2 / N$, 那么相应的概率表为表 3。

[0595] 表 3. 概率表

| | VIC 阴性 | VIC 阳性 | 概率 |
|--------|-----------------------|-----------------|----------|
| FAM 阳性 | $p1 * (1 - p2)$ | $p1 * p2$ | $p1$ |
| FAM 阴性 | $(1 - p1) * (1 - p2)$ | $(1 - p1) * p2$ | $1 - p1$ |
| 总计 | $1 - p2$ | $p2$ | 1 |

[0596] 在这种情况下, T1 和 T2 之间存在 100% 的断裂。

[0597] 可分别如下计算 T1 和 T2 分子的数目 M1 和 M2 :

[0598] $M1 = -N \log(1 - p1)$

[0599] $M2 = -N \log(1 - p2)$

[0600] (如果 N 个数字分区中 P 个为阳性, 那么分子数为 $M = -N \log(1 - P / N)$)

[0601] 靶序列之间没有断裂

[0602] 如果 T1 和 T2 总是在同一 DNA 片段上, 那么它们是相连的 (可能由于它们的基因座染色体的相同部分上彼此距离很近, 并且限制酶消化不在 T1 和 T2 之间进行消化)。参见图 42B。因此, $N1 = N2$ 。

[0603] 表 4. 分区计数

| | VIC 阴性 | VIC 阳性 | 总计 |
|--------|--------|--------|------|
| FAM 阳性 | 0 | N1 | N1 |
| FAM 阴性 | N-N1 | 0 | N-N1 |
| 总计 | N-N1 | N1 | N |

表 5 概率表

| | VIC 阴性 | VIC 阳性 | 概率 |
|--------|--------|--------|------|
| FAM 阳性 | 0 | p1 | p1 |
| FAM 阴性 | 1-p1 | 0 | 1-p1 |
| 总计 | 1-p1 | p1 | 1 |

[0604] 在这种情况下,存在 0% 的断裂。

[0605] 可如下计算 T1 和 T2 分子的数目,其中 $p1=N1/N$:

[0606] $M1=-N \log(1-p1)$

[0607] $M2=-N \log(1-p1)$

[0608] 部分断裂

[0609] 在中间情况下,T1 和 T2 在一些片段上连接,但有时也恰巧在单独的片段上。参见图 42C。

[0610] 如果存在 M3 个连接的 T1 和 T2 片段的分子、M1 个单独 T1 片段分子和 M2 个单独 T2 片段分子,那么可制成以下分区计数表:

[0611] 表 6. 分区计数

| | VIC 阴性 | VIC 阳性 | 总计 |
|--------|--------|--------|------|
| FAM 阳性 | N01 | N11 | N1 |
| FAM 阴性 | N00 | N10 | N-N1 |
| 总计 | N-N2 | N2 | N |

[0612] 如果 $M1=M2=M3$,那么存在 50% 的断裂,因为 50% 的相连的分子断裂成单独的片段,而其余 50% 保持完整。

[0613] 实施例 14 :用里程碑试验评估血浆中的 DNA 质量

[0614] 似乎在具有较高 DNA 产率的样品中,额外的 DNA 的大小显著较大。如图 43 所示,当 DNA 产率在 2kGE(基因组当量)/ml 左右时,约半数 DNA 的大小小于 1Kb;当产率极高(10kGE/ml 或更高)时,90% 的 DNA 大于 1Kb。这提示小 DNA 的浓度相对恒定。这进一步提示较高的 DNA 产率是由于细胞 DNA 的污染。

[0615] 实施例 15 :血浆中的大 DNA 的污染影响 CNV 值

[0616] 大尺寸 DNA 污染会抑制测量的 CNV 值。顶部的图 (图 44) 显示了测量的 CNV 值与 DNA 大小之间的良好相关性。通过对 DNA 进行预处理,可以大大消除这种大小影响 (底部的图,菱形)。初步的数据已显示,在 CNV 测量之前适当的 DNA 断裂,使得在不同物质之间测量的 CNV 值是相同的 (此处数据未示出)。

[0617] 实施例 16 :用于 CNV 分析的限制酶滴定

[0618] 在测试试验中,可对限制酶 (RE) 进行滴定来确定它们在 CNV 分析中的应用。分析了 17 种限制酶 (主要是 4- 切割酶)。参见图 45。限制酶浓度包含 4 倍稀释系列 :20、5、1.25、0.31、0.08、0.02 和 0.005U/ μ g DNA ;未消化的对照 + 缓冲液 ;每个浓度 4 个重复的 ddPCR ;DNA 输入 :0.5cpd。在低至 0.02U/ μ g DNA 的浓度下,可达到有效 / 完全的限制酶切割。有效的 RE 浓度范围可随着靶序列和酶而变化。对于大部分限制酶来说,5U/ μ g DNA 似乎能给出最高 ddPCR 浓度,从而允许用于 CNV 分析的更加普遍的消化方案。测试的所有 17 种限制酶在 20U RE/ μ gDNA 浓度下都未观察到非特异性 RE 活性。参见图 46。

[0619] 实施例 17 :用于优化具有挑战性的 ddPCR CNV 试验的方案的开发

[0620] 使用温度梯度可允许确定最佳退火温度。使用 GC 添加剂可以减少二级结构,该二级结构会以其它方式阻断聚合酶活性或导致竞争性非特异性扩增产物的形成。添加剂的实例是富含 GC 的增强子溶液 (5X) (Roche)。参见图 47。

[0621] 实施例 18 :用 ddPCR 测定限制酶的活性

[0622] Alu I 和 Mse I 在宽浓度范围上可接受地运行。在应该发生完全切割的酶浓度下,仍有少量的残存靶序列。NEB 和 Fermentas 酶都表现良好,但是 Fermentas 酶稍微更加有效。Fermentas 酶具有易于使用的优点。

[0623] 图 48 的左区显示了使用从 NEW 购买的 Alu I 和 Fermentas 的试验。该试验中所用的引物可以是 :正向引物 5'-GGCCCTCATCCACCATAACAC-3',反向引物 5'-GGCCCTCATCCACCATAACAC-3'。探针可以是 5'-6FAM-TCCGAAAGAGCTGGTC-MGB-NFQ-3' ;Alu I 切割该扩增子序列两次。图 48 的右区显示了使用 Mse I 的试验。该试验中所用的引物可以是 :5'-CTCCCTCTCCATAGCTACTTAAGGA-3',反向引物 5'-CAAGGAGCCCTAACCAATGGA-3',而探针可以是 5'-6FAM-AAGGCAGAGATTAAAG-MGB-NFQ-3'。Mse I 切割该扩增子序列两次。如图 48 中所示,用 Alu I 或 MseI 切割参考基因。在该实验中,对加入到样品中的 RE 的量进行滴定。当 RE 以高水平存在时,其应该消化该实验设计用于扩增的靶序列。随着靶序列被消化,阳性小滴数减少。如果靶序列被消化,那么很少有或没有阳性小滴出现。参见图 48。

[0624] 实施例 19 :通过共定位的单元型分析

[0625] 提供了一种通过共定位获取单元型分析信息的方法。该方法可用来确定是否存在靶核酸序列的缺失。在拷贝数变异区域中,标记序列 (例如,用 VIC 标记的探针检测的) 可在靶序列 (如,用 FAM 标记的探针检测的) 外部但要在其附近。可将包含核酸的样品分成多个空间上分离的区域,并且可检测标记序列和靶核酸序列 (例如,通过扩增和用探针检测)。如图 49 中所示,可对 VIC (标记序列) 和 FAM (靶序列) 的共定位进行分析。如果 VIC 和 FAM 总是共定位于分区中,那么很可能不存在靶序列的缺失 (图 49B)。如果存在只含有不与 FAM 共定位的 VIC 的分区,那么该结果提示靶序列的缺失 (图 49A)。

[0626] 实施例 20 :DNA 样品的储存

[0627] 图 20A 和图 20B 图示了储存的已消化的 DNA 的 MRGPRX1CNV 值的漂移。在图 20A

中,已消化的样品显示始终在整数以下的 CNV 值。在图 20B 中,CNV 值更接近整数。分析的样品包括 Coriell 纯化的 DNA 样品 :NA11994、NA18507、NA18502、NA19221、NA19205、NA18916 和 NA19108。图 20A 中显示的样品与图 20B 中的样品在不同的日期进行分析。简言之,使用 RsaI 来消化样品,使用标准 ddPCR 热循环条件 :95°C 10 分钟 (1 个循环),94°C 30 秒和 60°C 1 分钟 (40 个循环),98°C 10 分钟 (1 个循环),12°C 保持。

[0628] MRGPRX1 分析序列为 (正向引物)5' -TTAAGCTTCATCAGTATCCCCCA-3' , (反向引物)5' -CAAAGTAGGAAAACATCATCACAGGA-3' ,和 (探针)6FAM-ACCATCTCTAAAATCCT-MGBNFQ。将样品双链体化于 RRP30(VIC 标记的);RPP30 参考试验:(正向引物)5' -GATTTGGACCTGCGAGCG-3' ,(反向引物)5' -GCGGCTGTCTCCACAAGT-3' ,和 (探针)VIC-CTGACCTGAAGGCTCT-MGBNFQ。立即 (在消化的 24 小时内)处理 DNA。

[0629] 虽然本文已经显示和描述了本发明的优选实施方案,但对于本领域中技术人员显而易见的是,这些实施方案仅通过举例的方式提供。在不偏离本发明的情况下,本领域技术人员现将想到众多更改、改变和替换。应当理解,在实施本发明的过程中可采用本文所述的本发明实施方案的各种替代方案。以下权利要求旨在限定本发明的范围,从而涵盖这些权利要求的范围内的方法和结构及其等效方案。

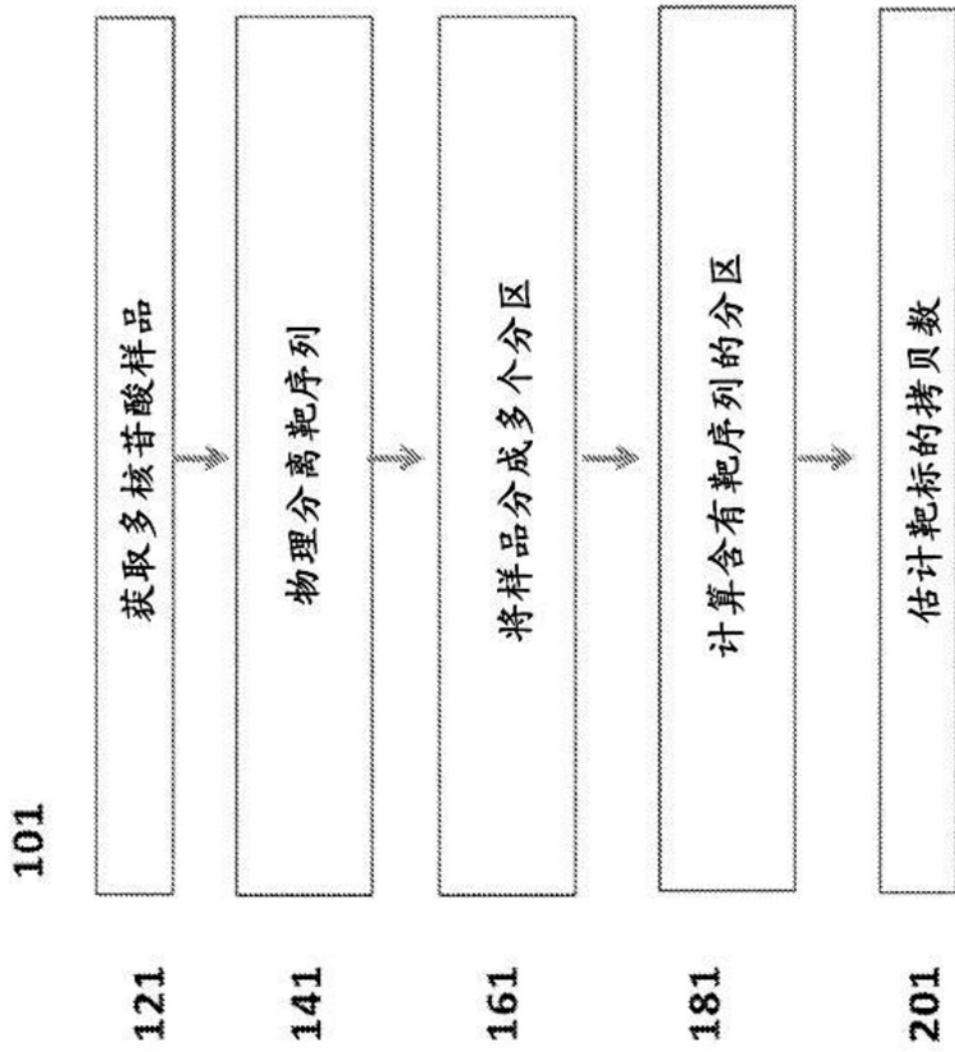


图 1

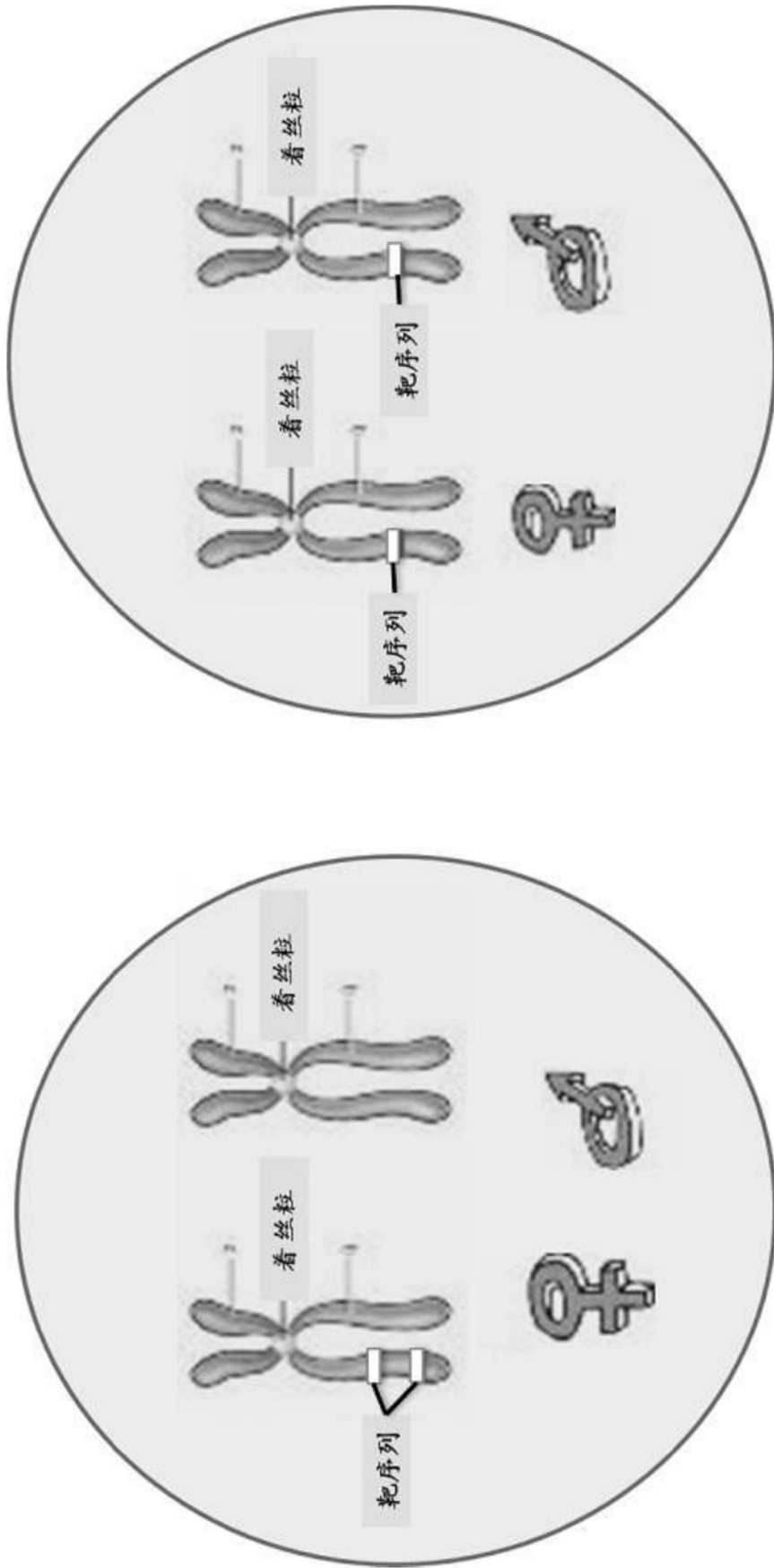


图 2

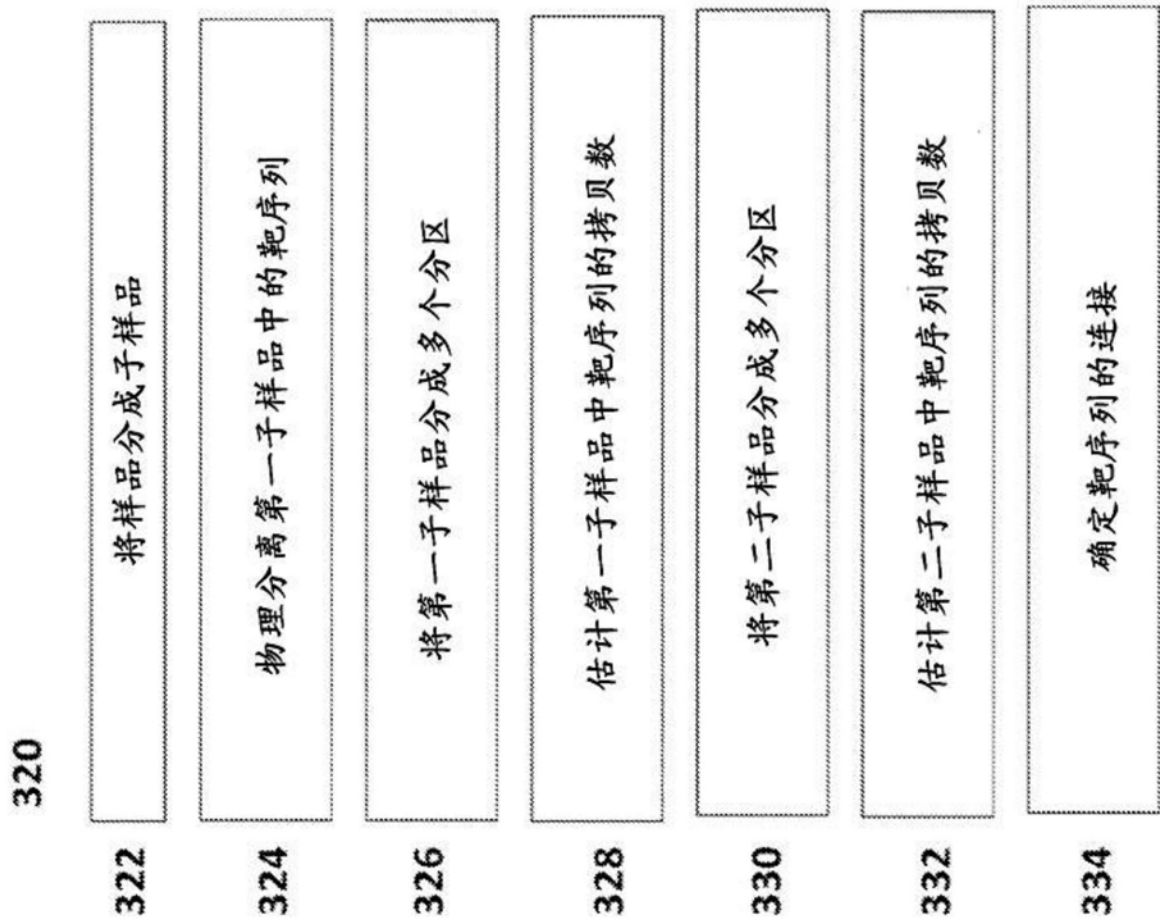


图 3A

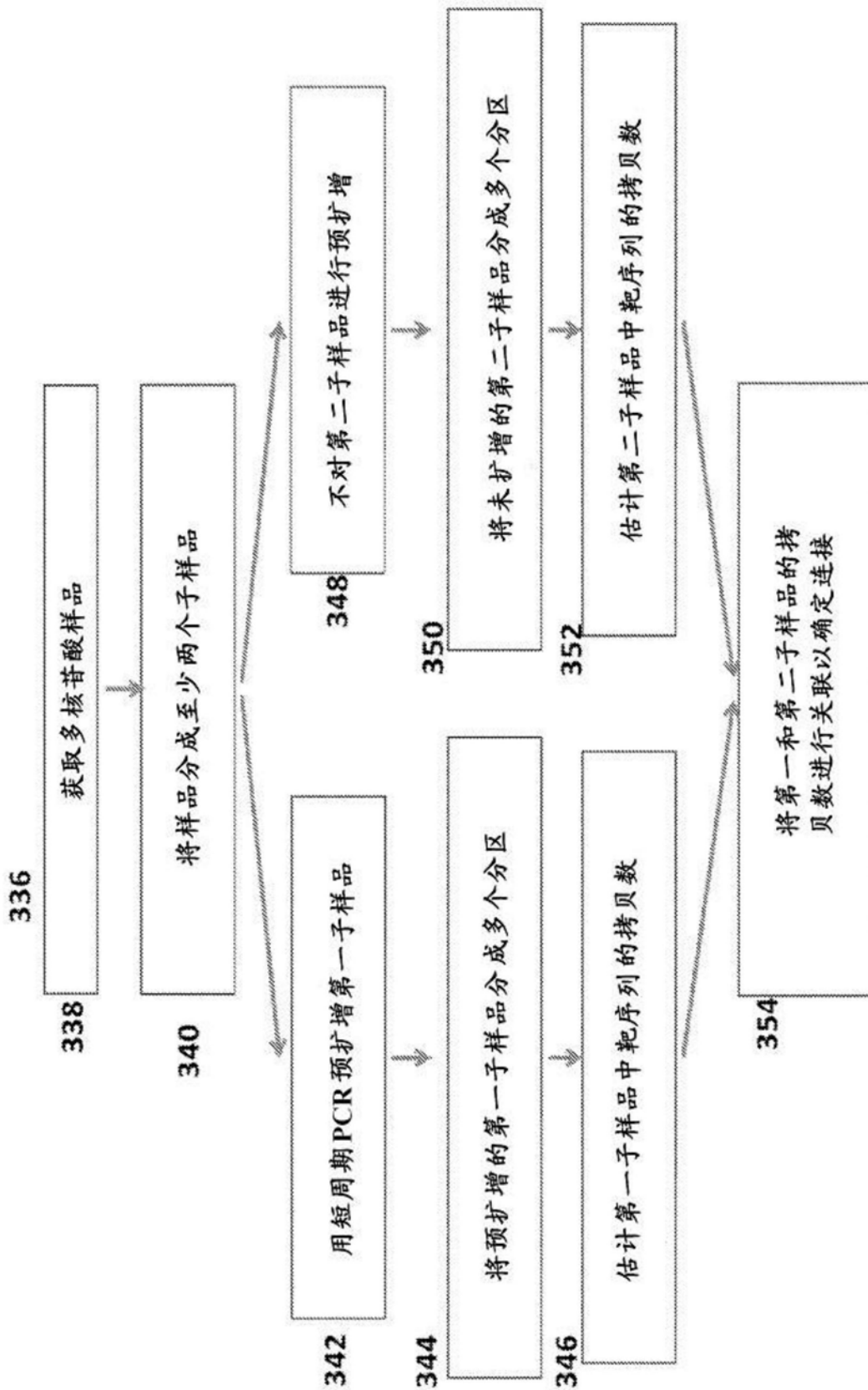


图 3B

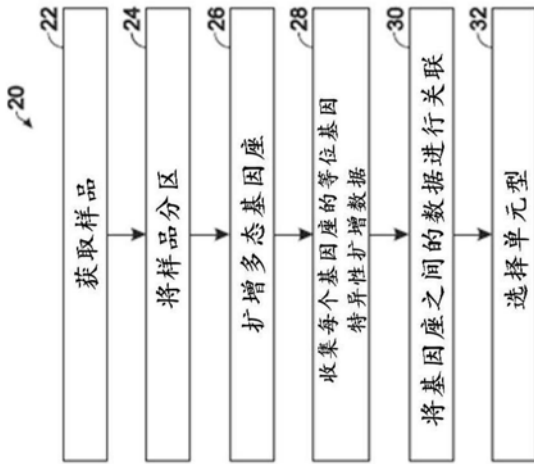


图 4

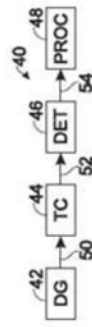


图 5

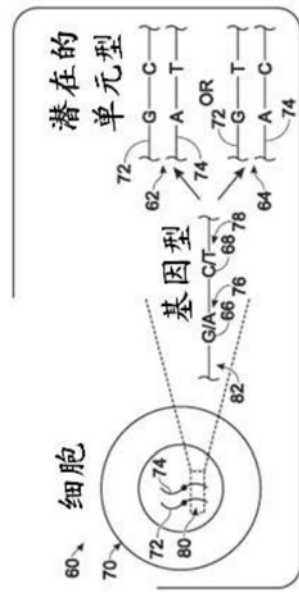


图 6

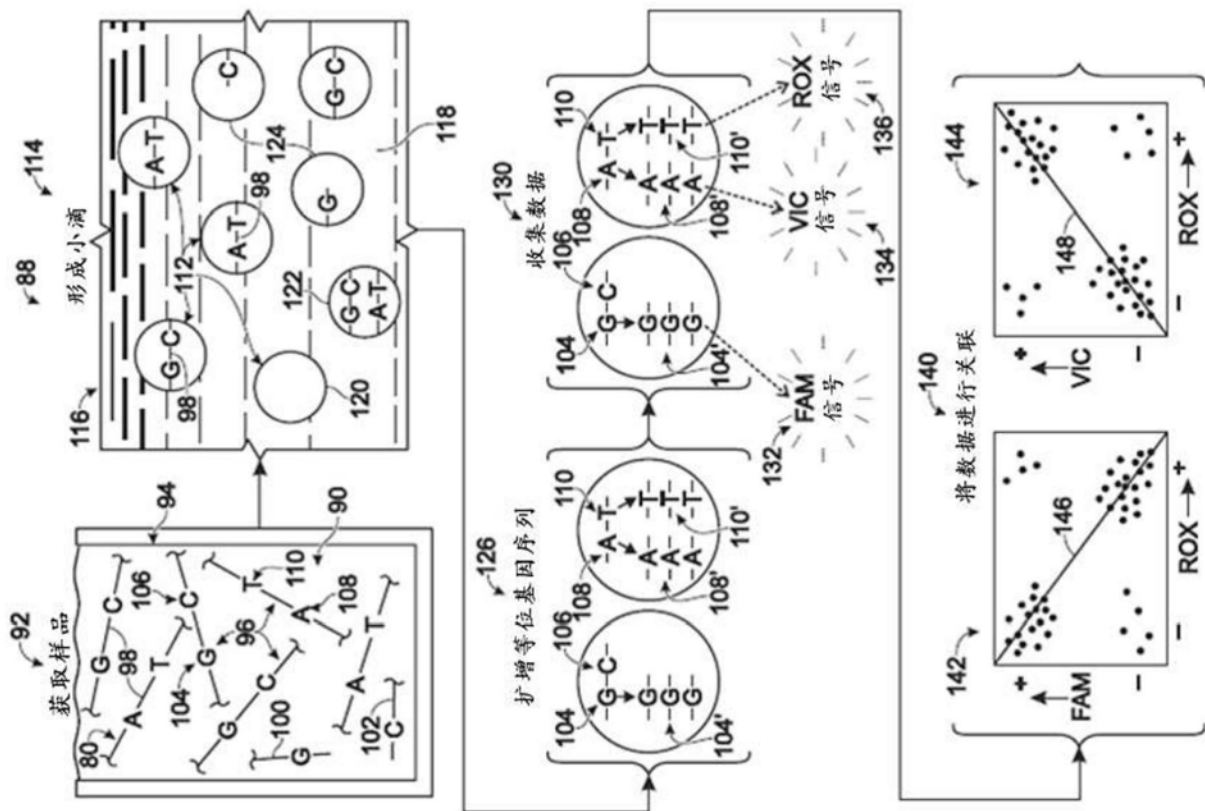


图 7

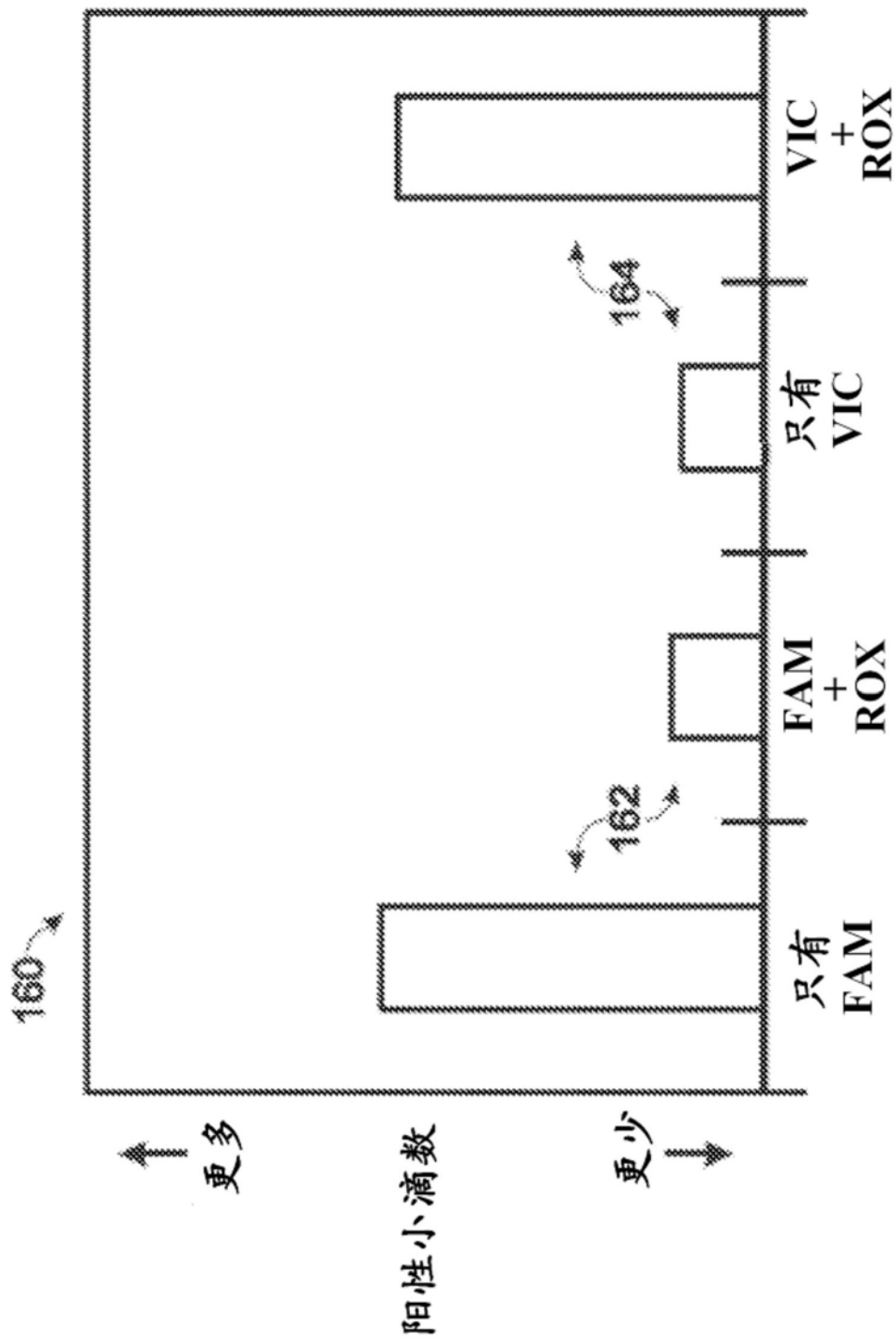


图 8

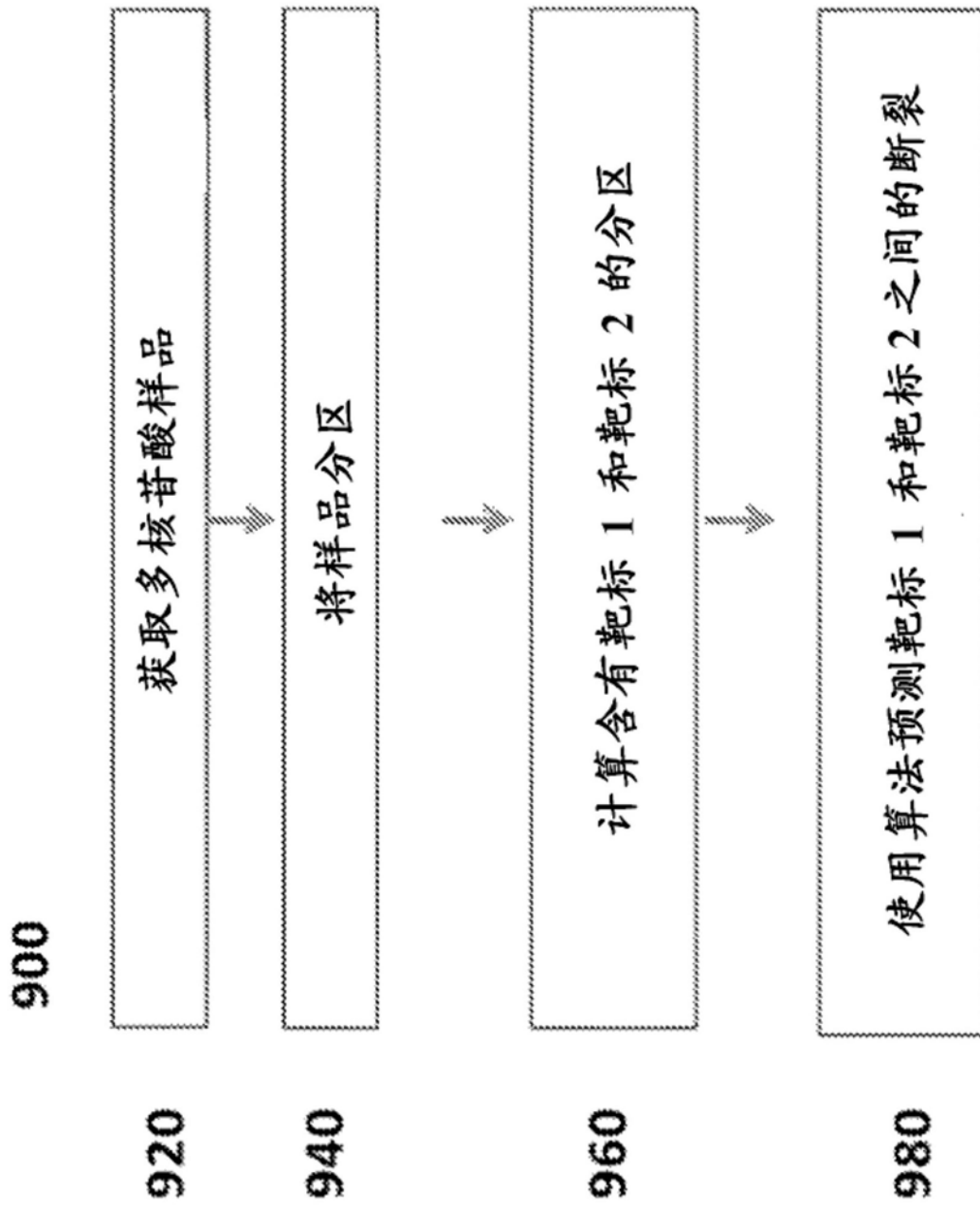
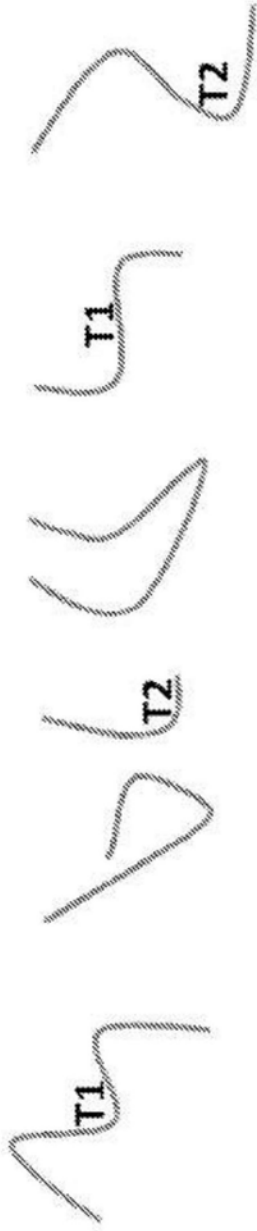


图 9

A T1 和 T2 不相连



B T1 和 T2 相连; T1 片段; T2 片段



C



图 10

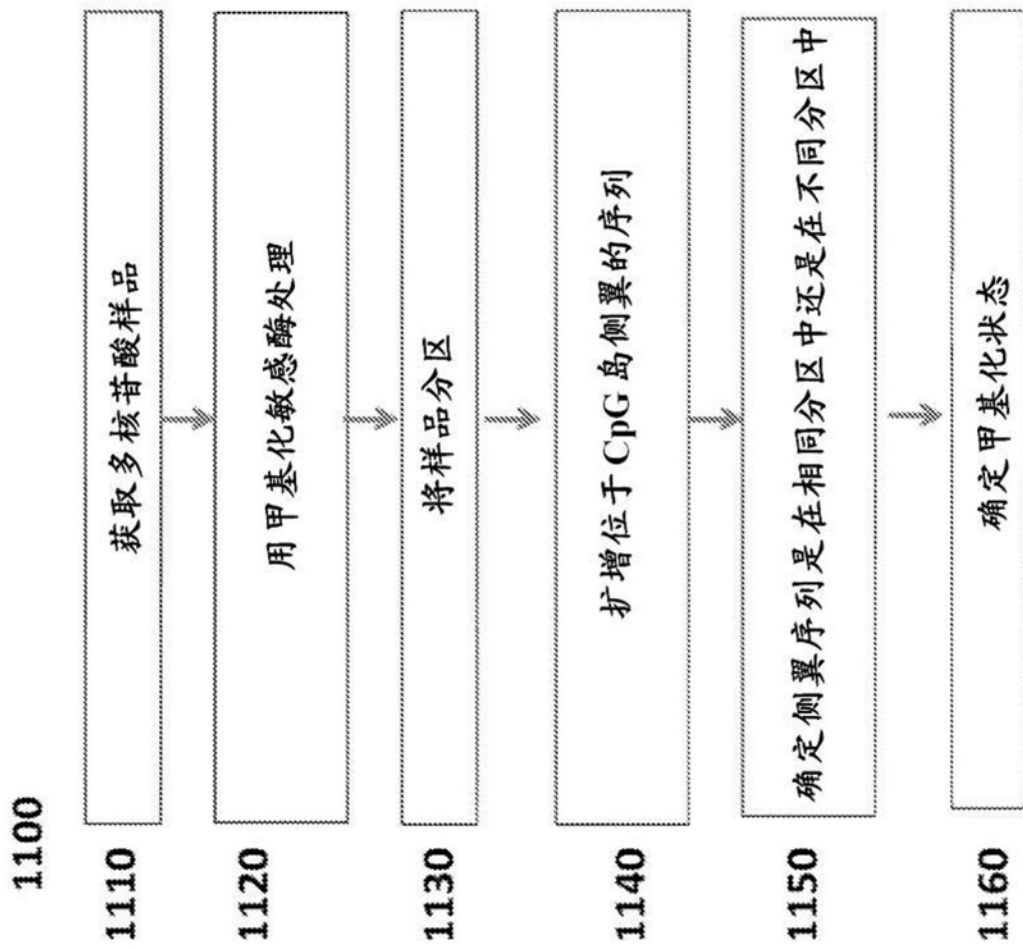


图 11

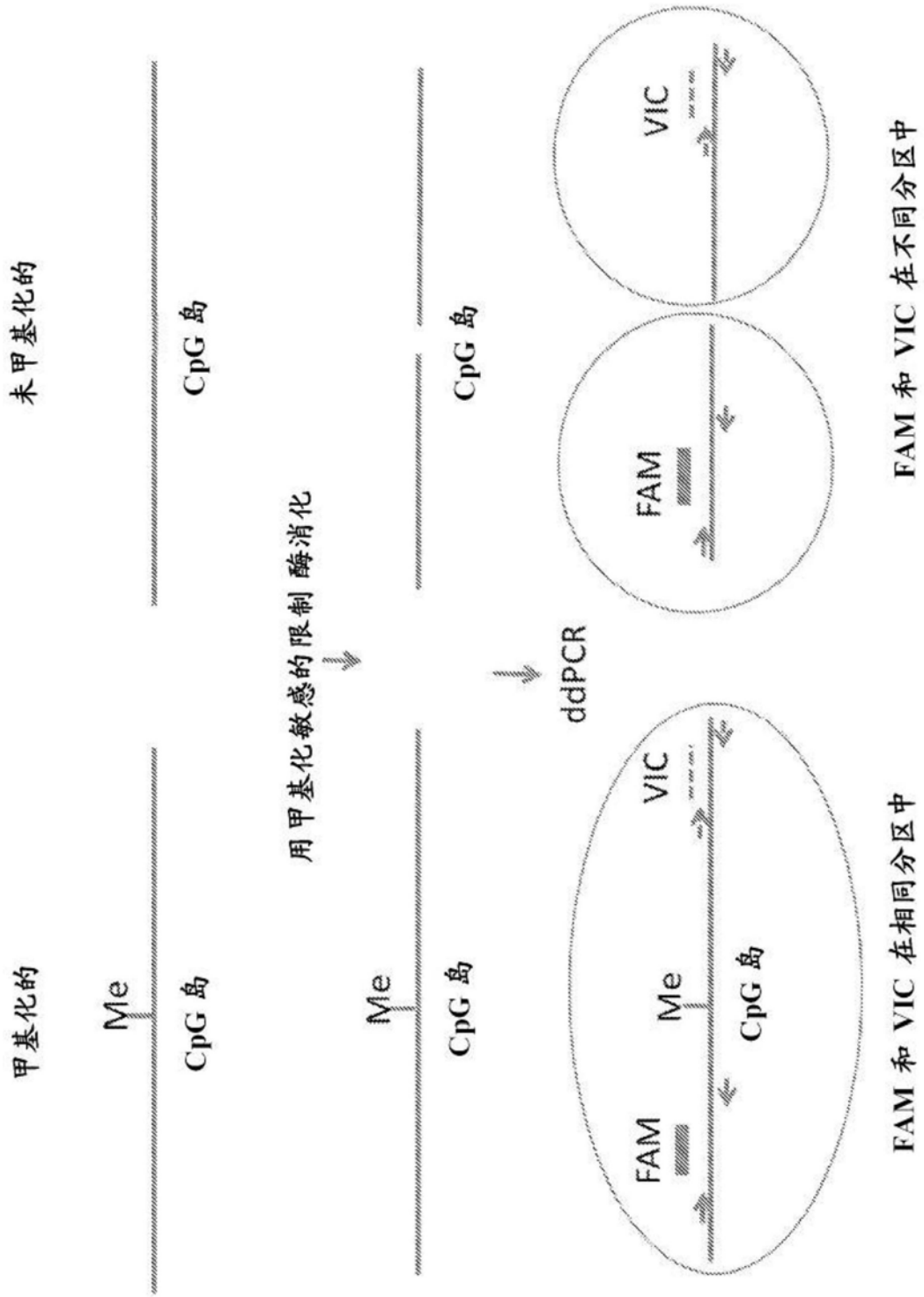


图 12

酶选择器：MRGPRXI CNV

试验：MRGPRXI CNV
对限制位点进行可视化

序列#1
 扩增子长度:70bp
 前级长度:1000bp
 后级长度:1000bp

单一消化物

| 名称 | # 切割 | 费用/100单位 |
|-----------|------|----------|
| NlaIII | 26 | \$12.20 |
| DdeI | 24 | \$6.10 |
| Fast DdeI | 24 | \$12.80 |
| HpyF3I | 24 | \$10.00 |
| MboI | 18 | \$13.20 |
| DpnII | 18 | \$6.30 |
| TseI | 16 | \$77.33 |

图 13

切割分析: MRGPRX1 CNV

选择酶并按下“cut”来确定酶是否在扩增子(绿色)或周围区域(红色、蓝色)内进行切割。可将鼠标悬停在颜色条上来查看序列本身以及切割是否靠近扩增子。红色条指示限制位点可能因突变而改变;绿色条为由于突变而出现的限制位点。

试验: MRGPRX1 CNV 选择限制酶切消化

扩增子长度:70bp
前级长度:300bp
后级长度:300bp
序列 FASTA 代码: chr11:18955710+18956379 chr11:18955710-18956379



测试切割

酶:

序列: GGCC

片段长度: 89

片段%GC: 39.33%

结果
切割两边
扩增子切割:0
前级切割:6
后级切割:4

图 14

编辑试验: MRGPRX1 CNV

“名称”:

基因:

输入者:

按引物输入 按位置输入 按SNP输入

正向引物:

反向引物:

探针序列:

染料:

猝灭剂:

二级结构:

最佳退火温度:

摄氏度

参考来源:

图 15A

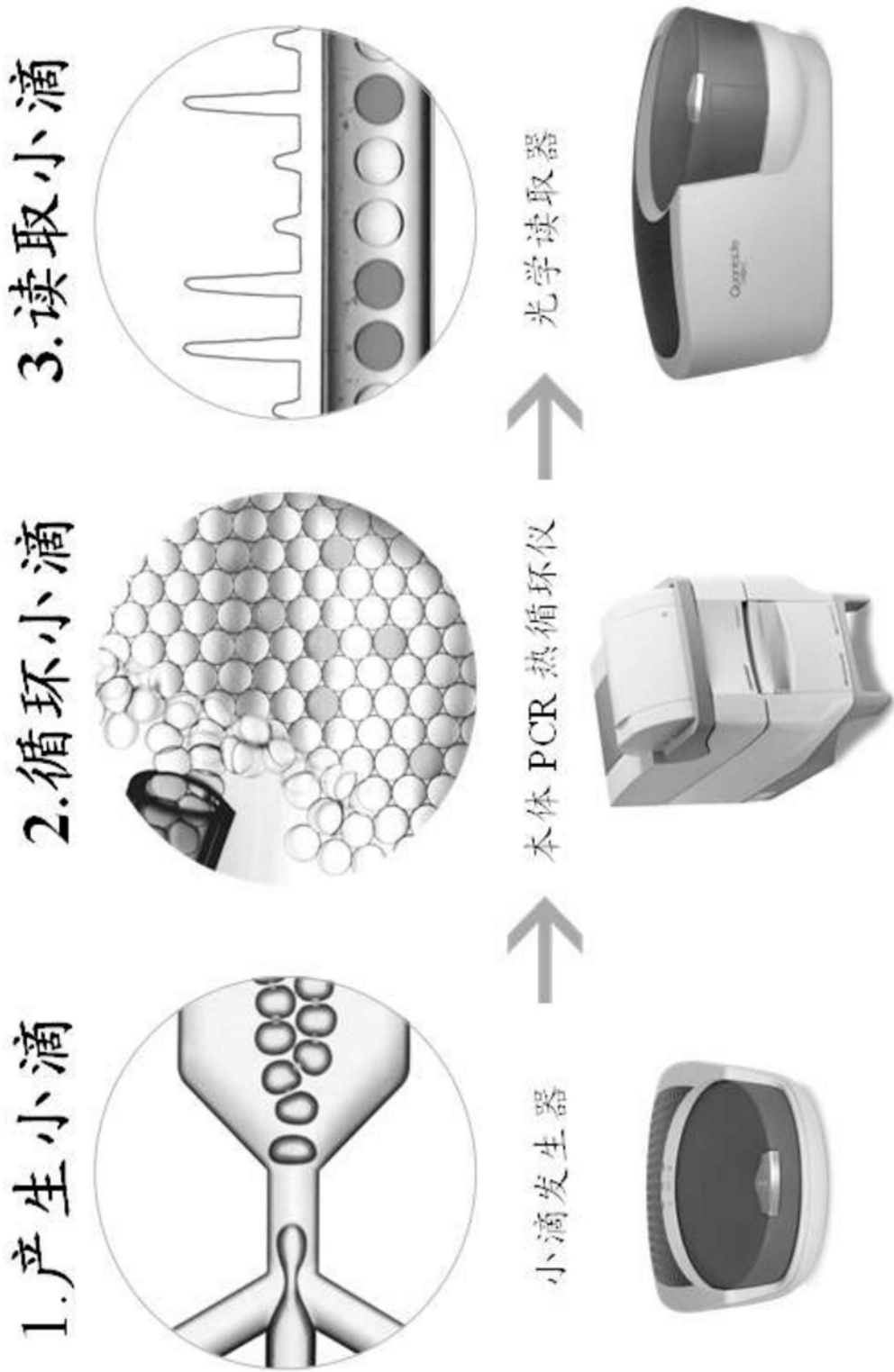


图 16

样品：主混合物/NTC

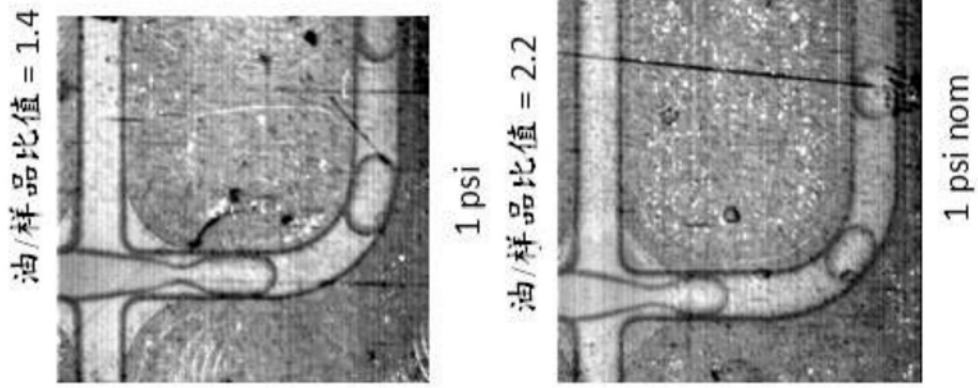
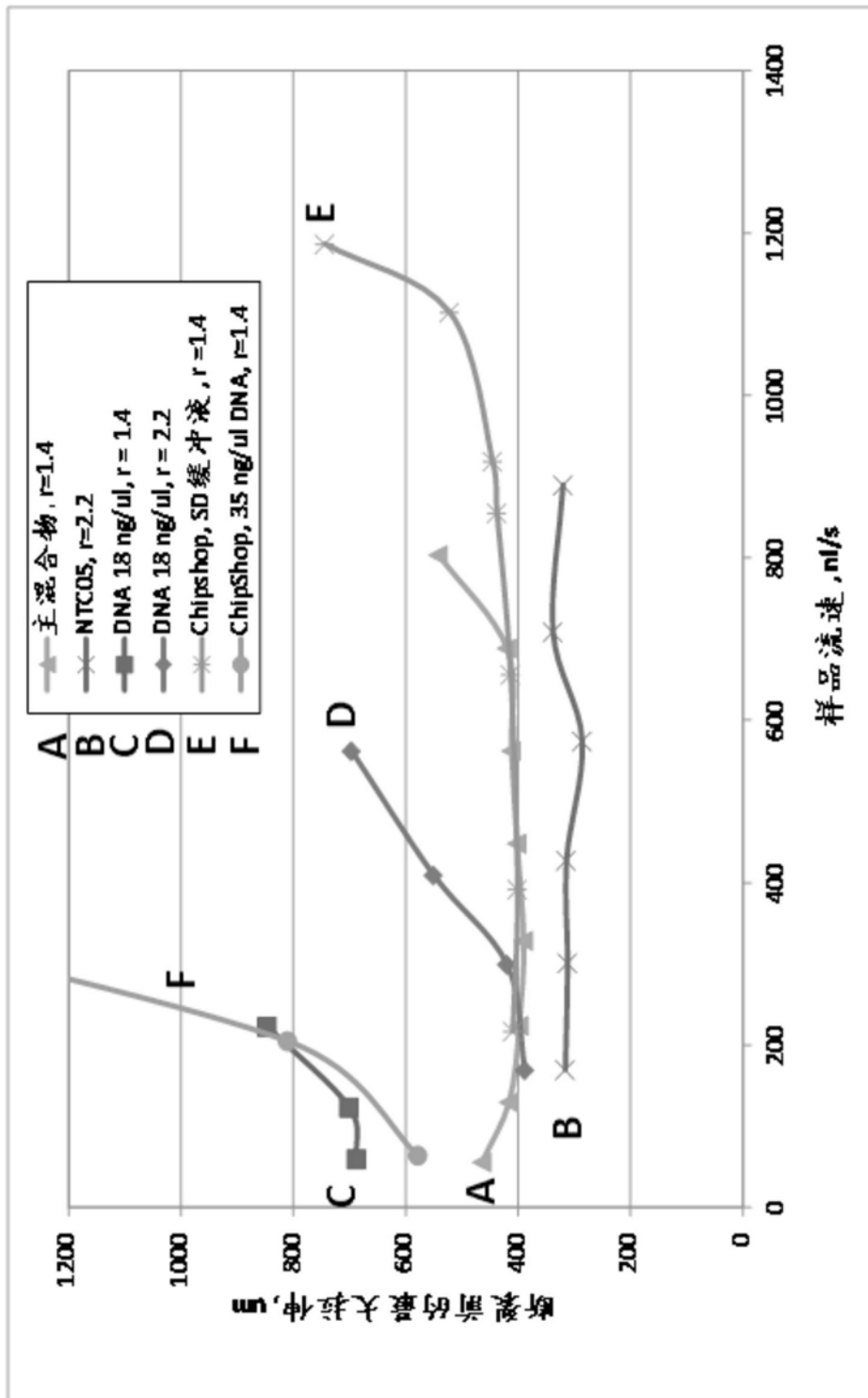


图 17



较高的油/样品比值使喷射阈值移位

图 18

| 样品编号 nom psi | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | | 样品类型 | 浓度 (ng/ul) |
|-----------------|-----|---|---|---|-----|---|---|---|-----|---|---|---|-----|---|-----|---|-----------|---------------|
| | 0.5 | 1 | 1 | 2 | 1.5 | 3 | 2 | 4 | 2.5 | 5 | 3 | 6 | 3.5 | 7 | 8 | | | |
| 1 | J | J | J | J | J | J | J | J | J | J | J | J | J | J | J | J | Raji 未消化 | 75 |
| 2 | N | N | E | E | J | J | J | J | J | J | J | J | J | J | J | J | Raji 未消化 | 37.5 |
| 3 | E | E | E | J | J | J | J | J | J | J | J | J | J | J | J | J | Raji 未消化 | 18.75 |
| 4 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Raji 未消化 | 3.75 |
| 5 | E | E | E | J | J | J | J | J | J | J | J | J | J | J | J | J | Raji 未消化 | 0 |
| 6 | E | J | J | J | J | J | J | J | J | J | J | J | J | J | J | J | 19205 未消化 | 75 |
| 7 | J | J | J | J | J | J | J | J | J | J | J | J | J | J | J | J | 19205 未消化 | 37.5 |
| 8 | E | E | J | J | J | J | J | J | J | J | J | J | J | J | air | J | 19205 未消化 | 18.75 |
| 9 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | E | E | 19205 未消化 | 3.75 |
| 10 | E | E | E | E | E | E | E | E | E | E | E | E | E | E | J | J | 19205 未消化 | 0 |
| 11 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Raji 已消化 | 75 |
| 12 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Raji 已消化 | 37.5 |
| 13 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Raji 已消化 | 18.75 |
| 14 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Raji 已消化 | 3.75 |
| 15 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Raji 已消化 | 0 |
| 16 | E | E | E | E | E | E | E | E | E | E | E | E | E | E | J | J | 19205 已消化 | 75 |
| 17 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | 19205 已消化 | 37.5 |
| 18 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | 19205 已消化 | 18.75 |
| 19 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | 19205 已消化 | 3.75 |
| 20 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | 19205 已消化 | 0 |

J = 喷射
 E = 拉伸
 N = 正常(无喷射或拉伸)

图 19

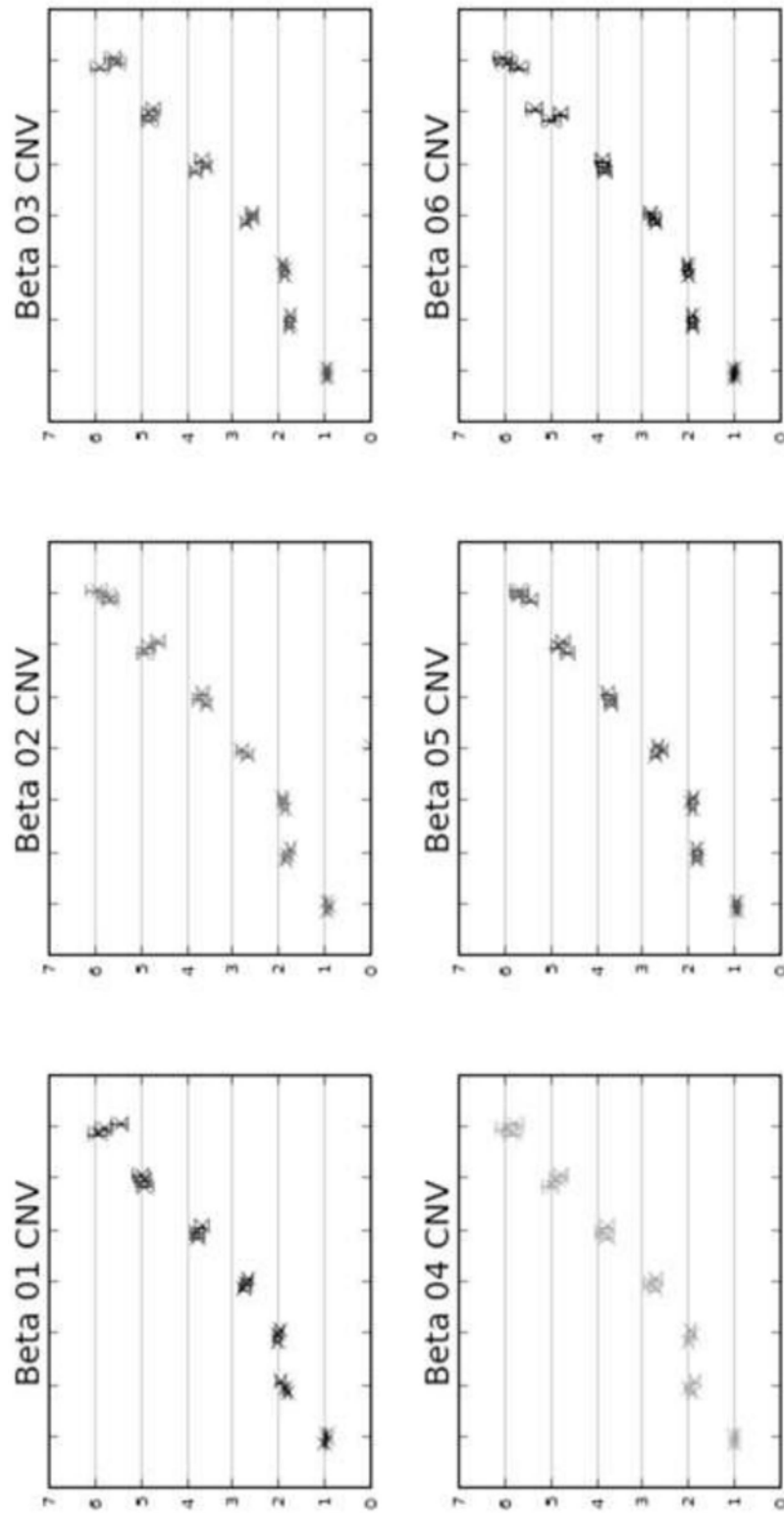


图 20A

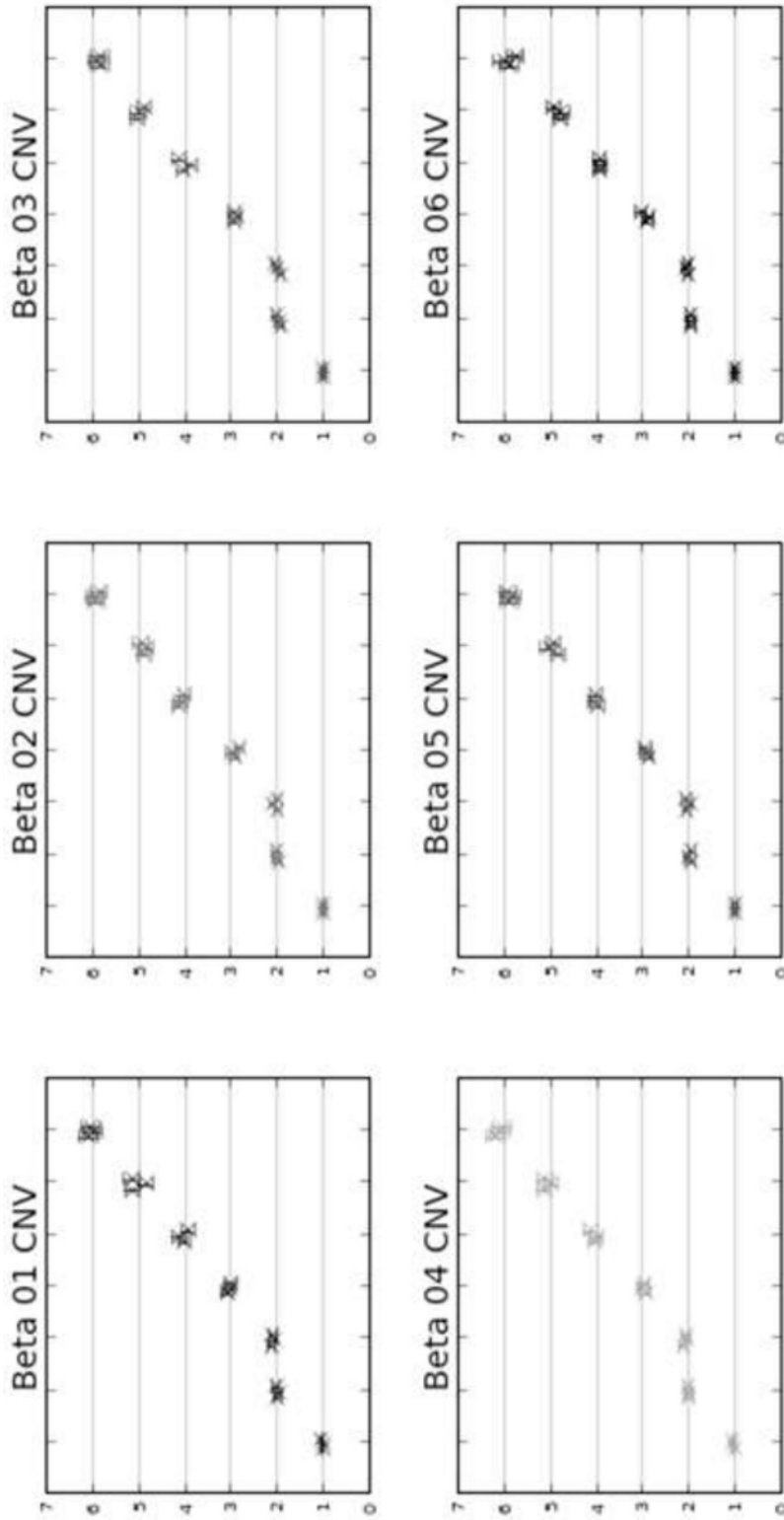


图 20B

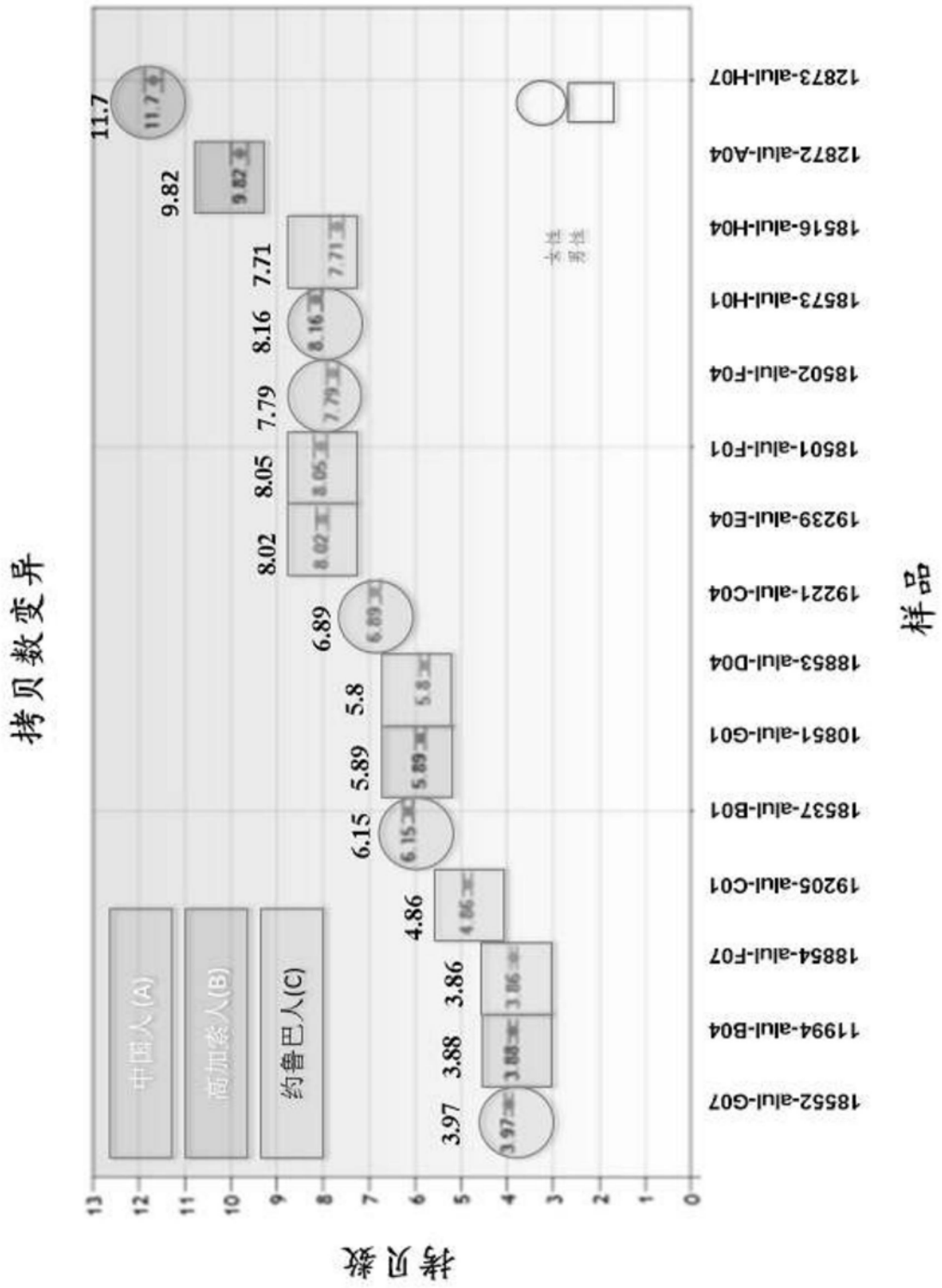


图 21

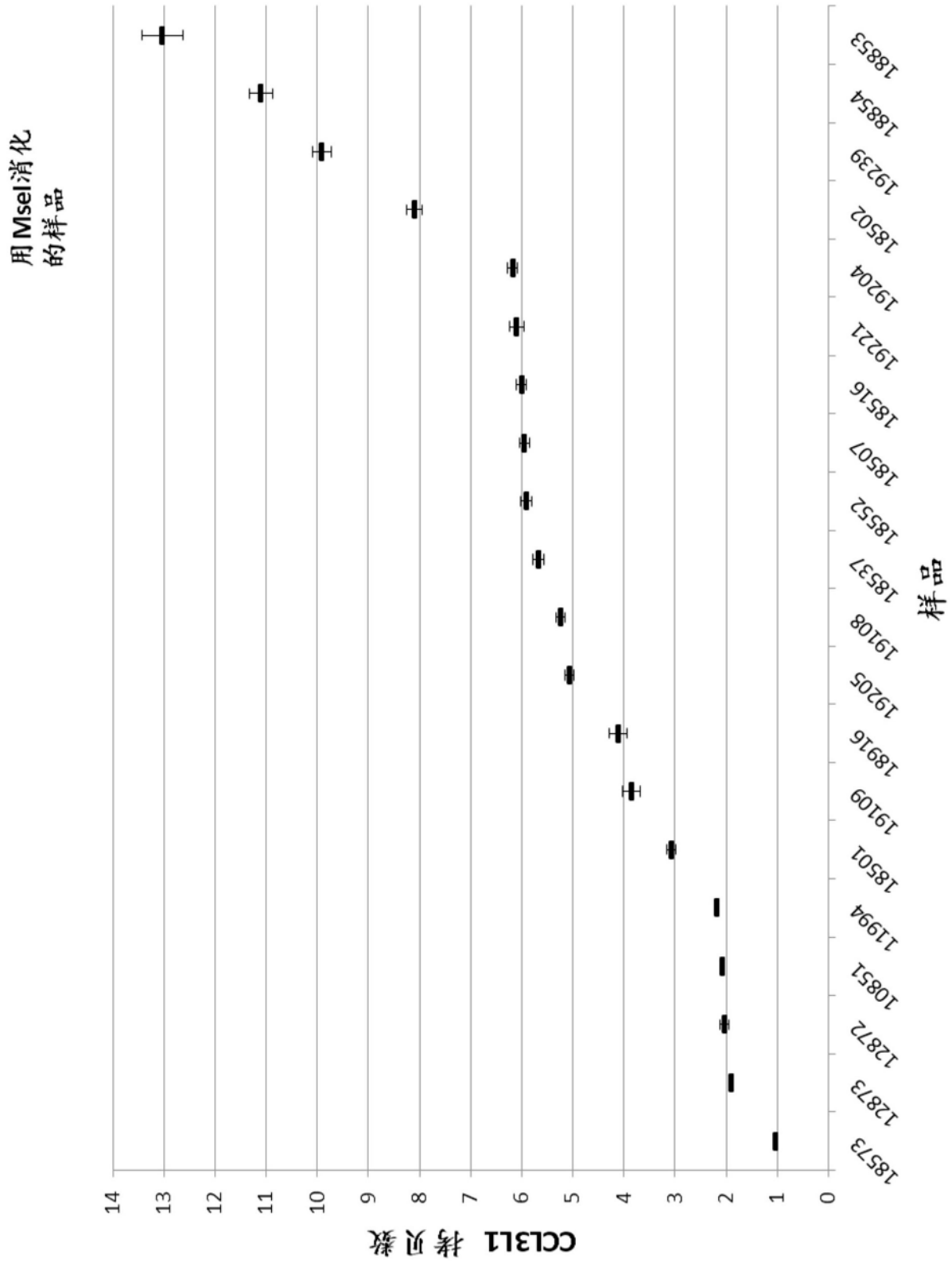


图 22

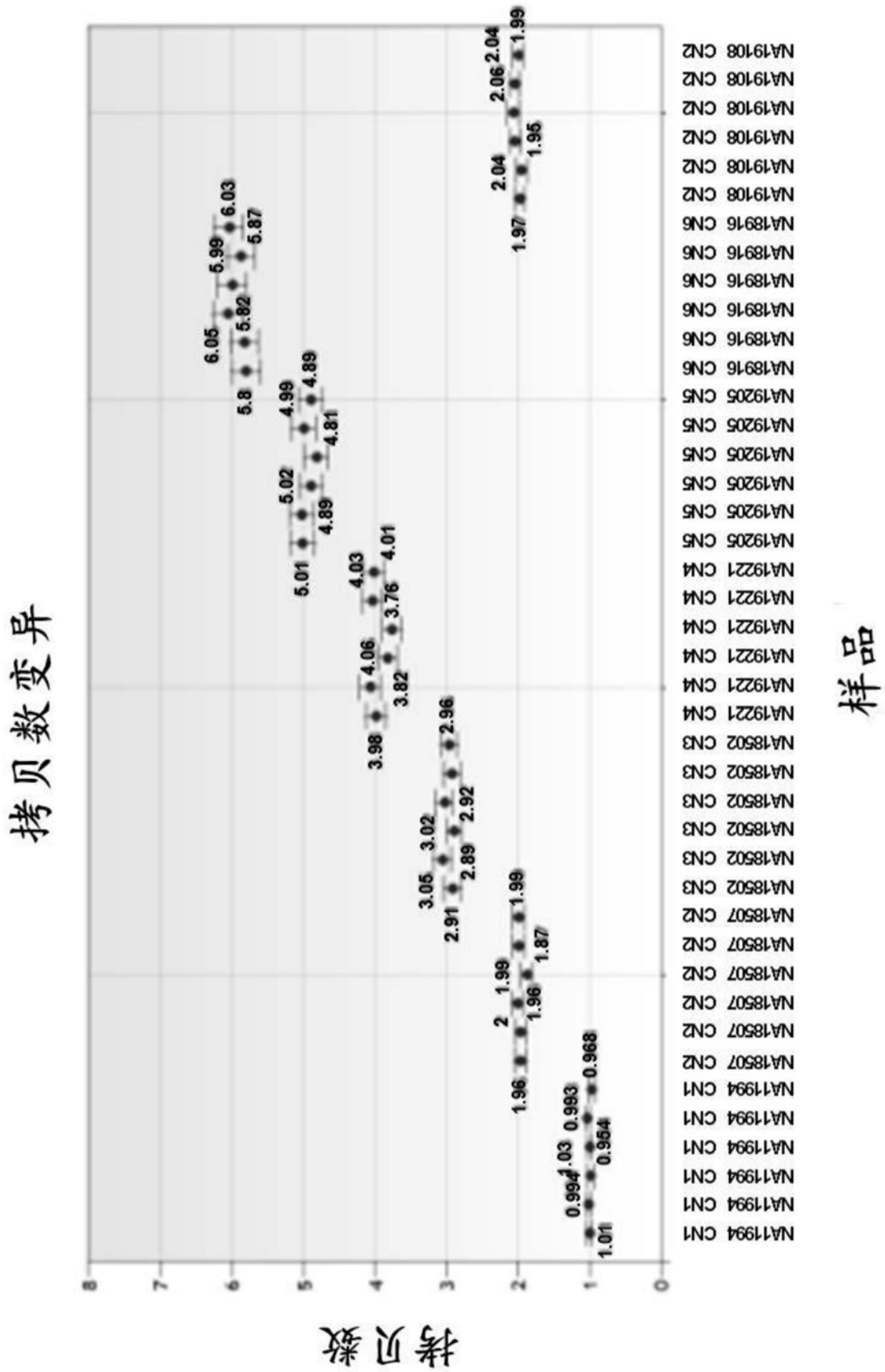


图 23A

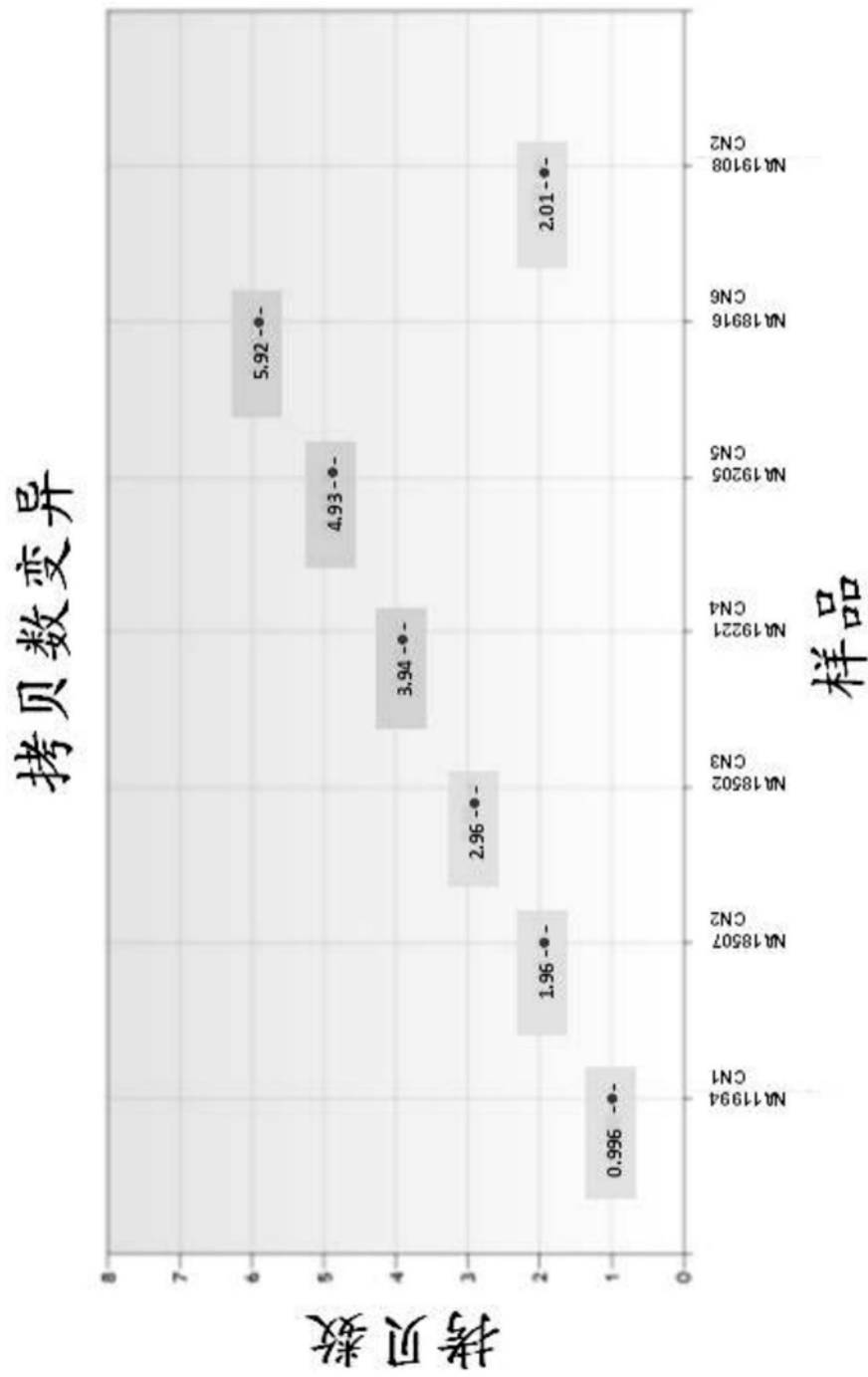


图 23B

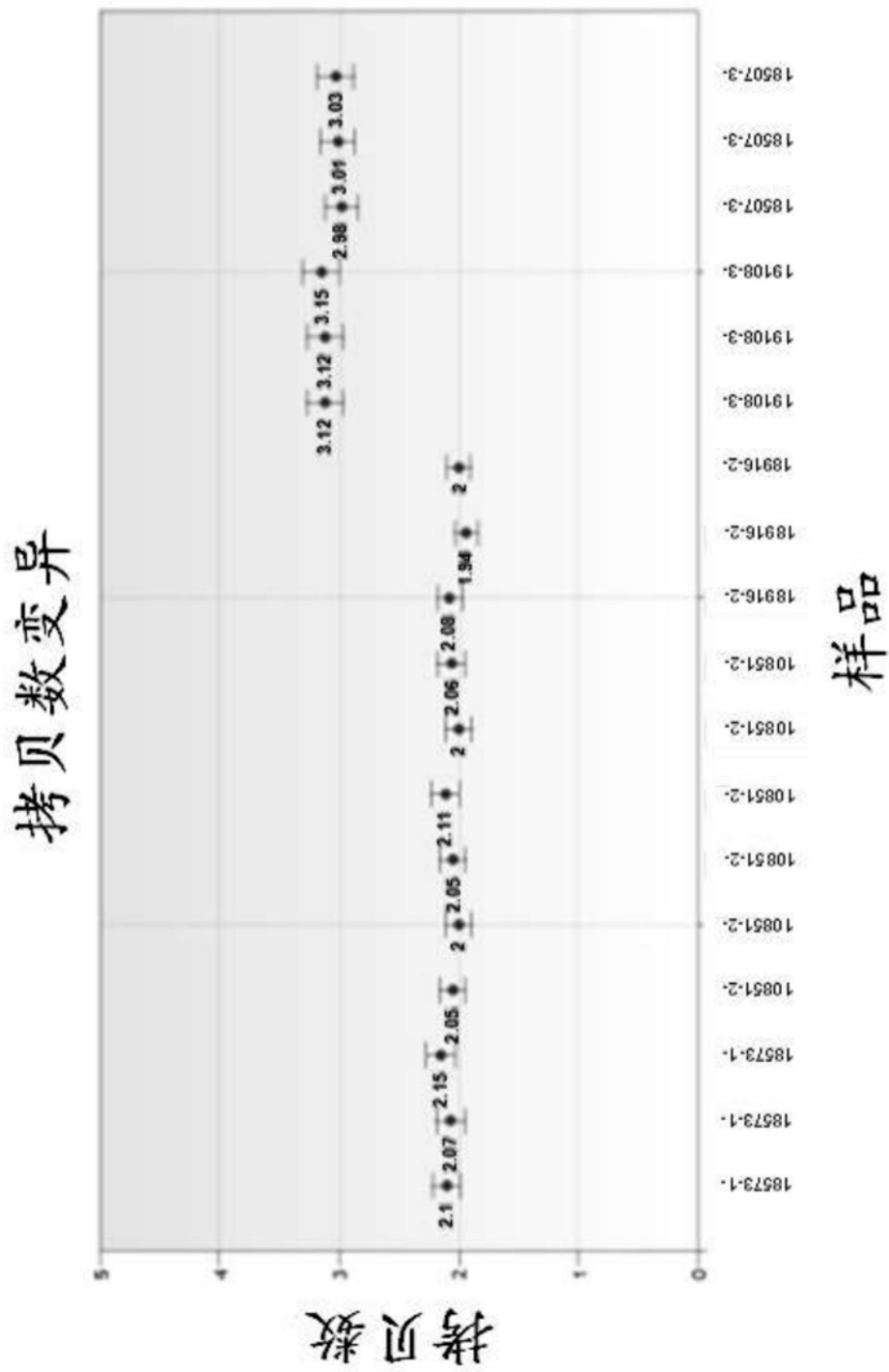


图 24A

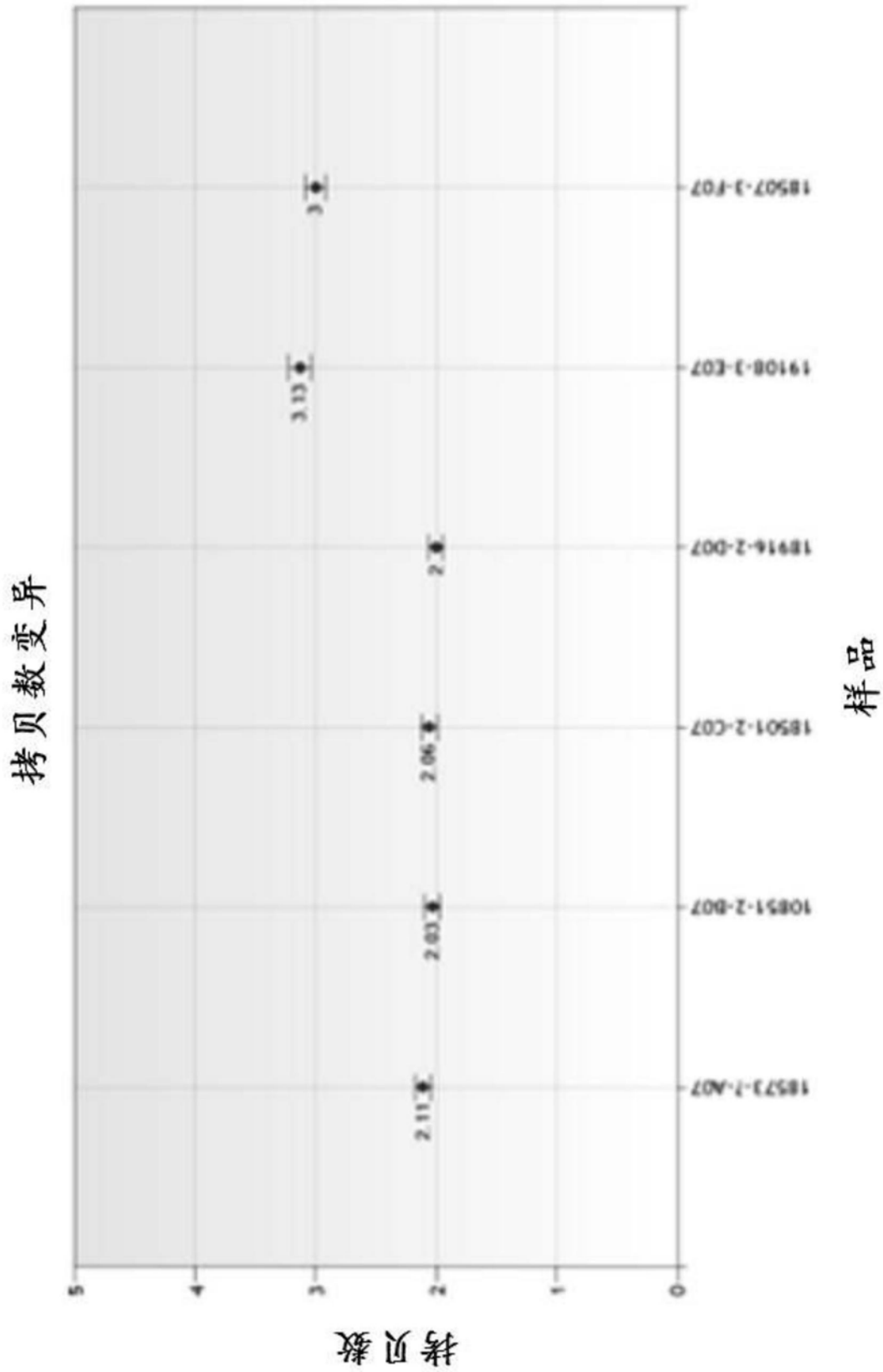
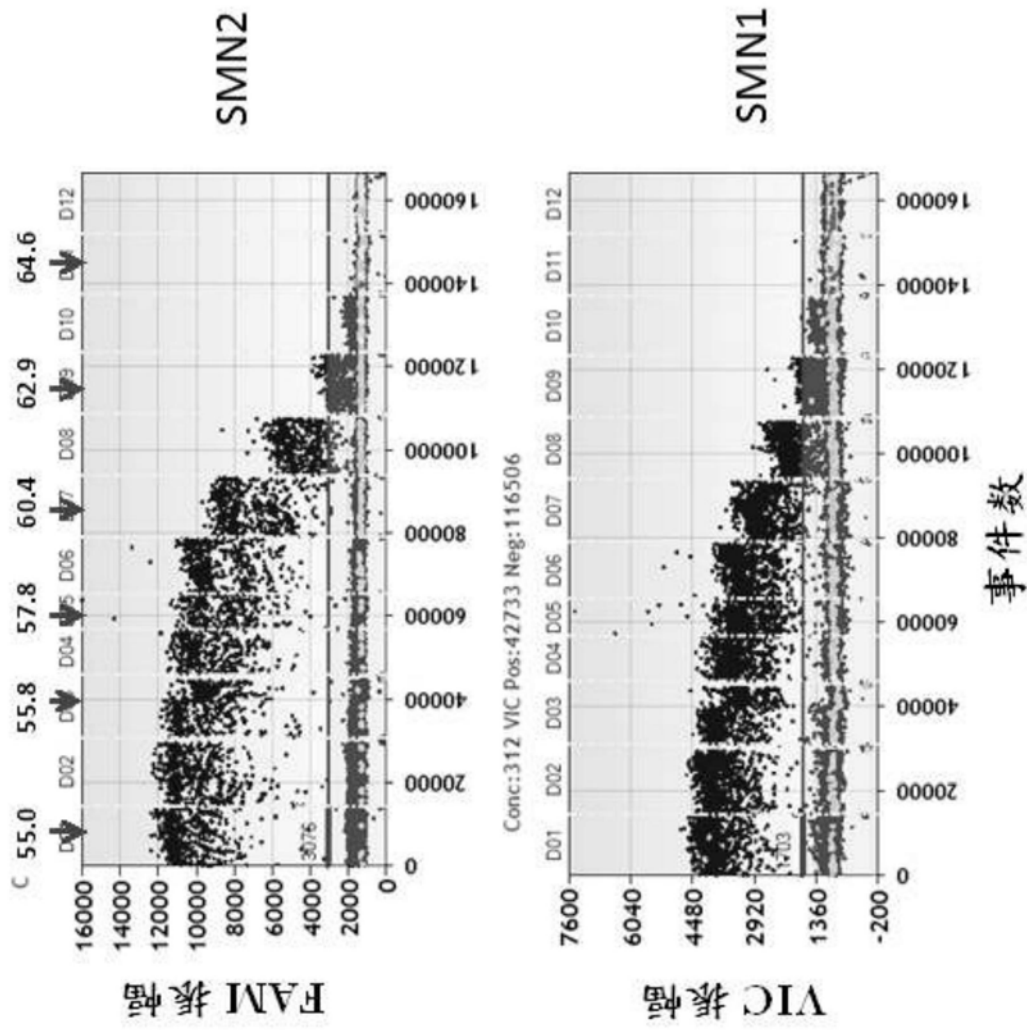


图 24B



用 CviQI 消化的 Coriell 样品 18507

图 25

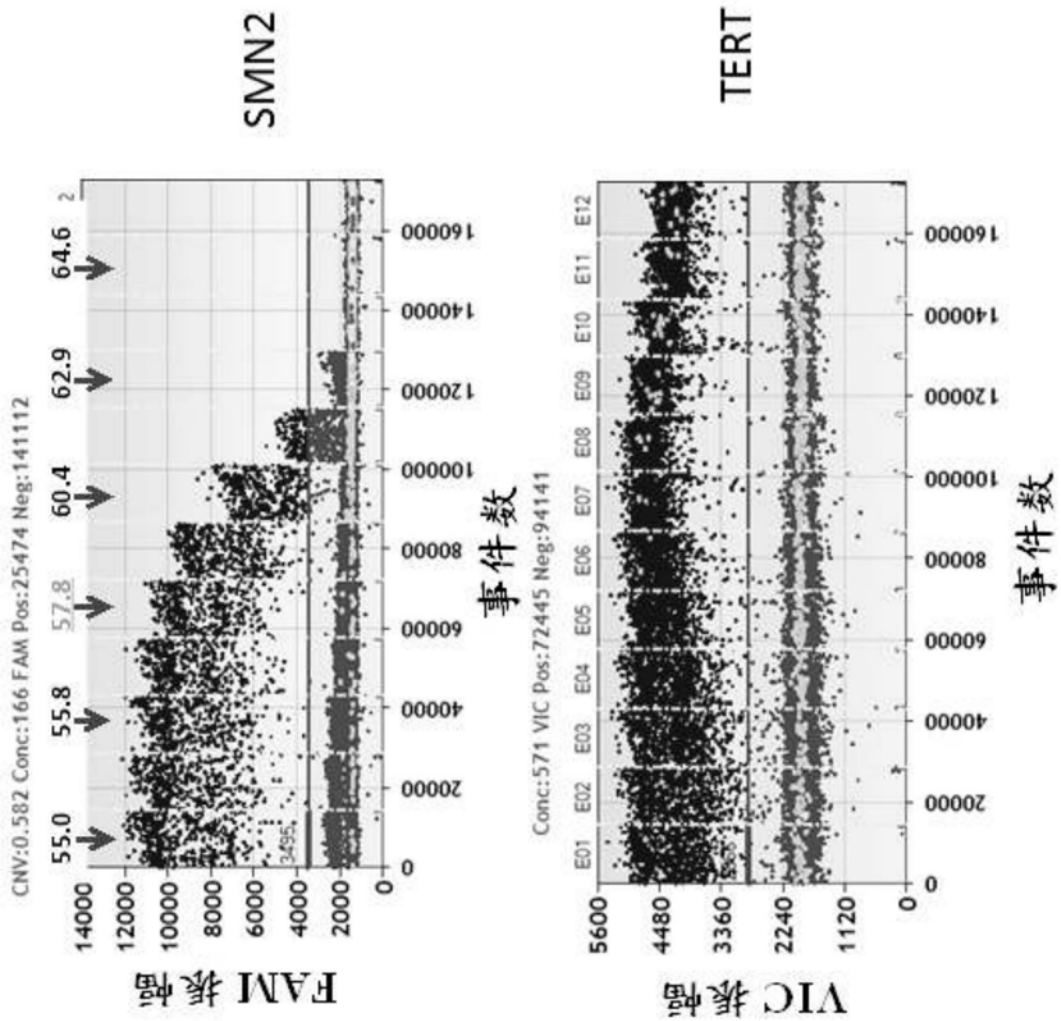


图 26

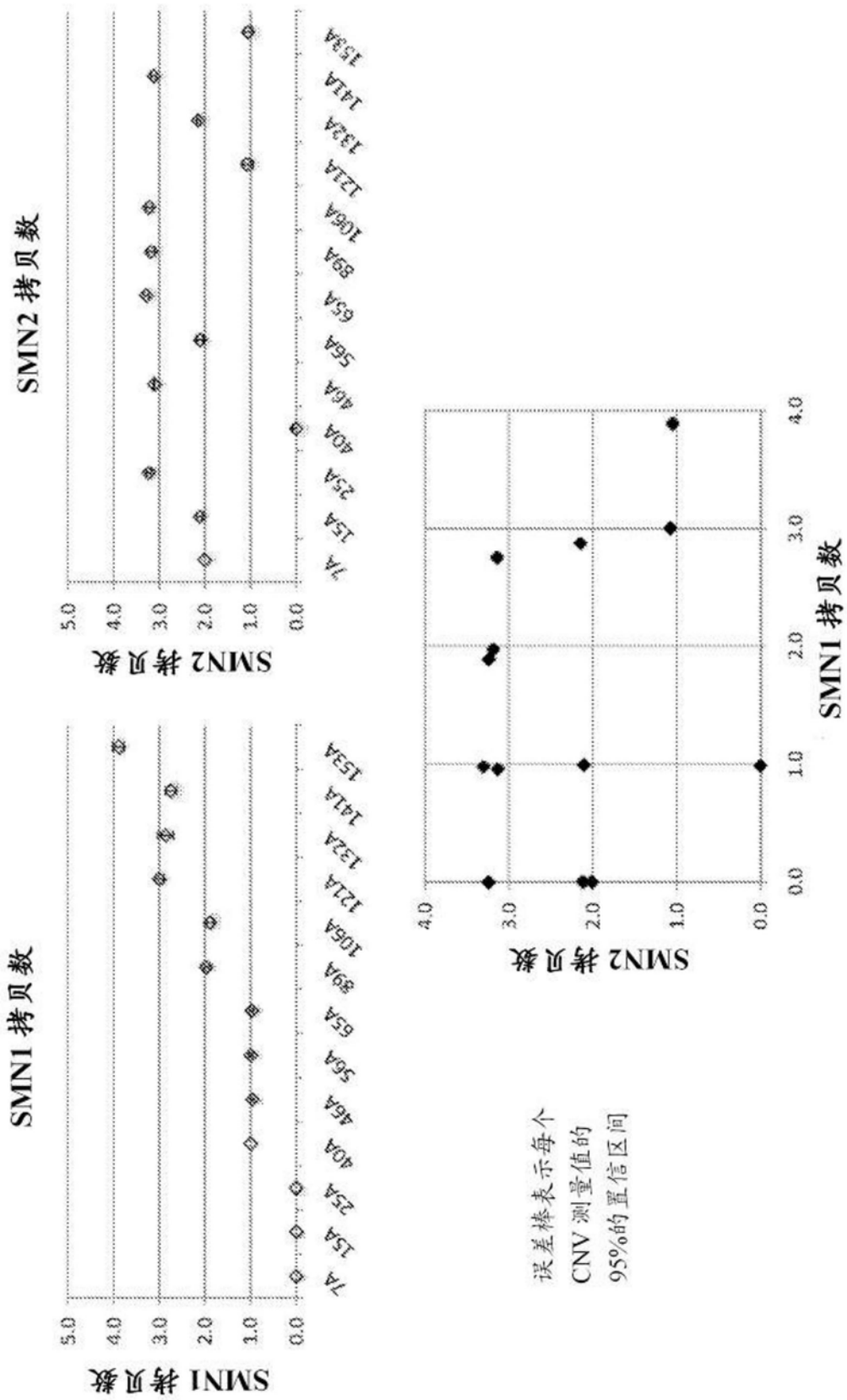


图 27

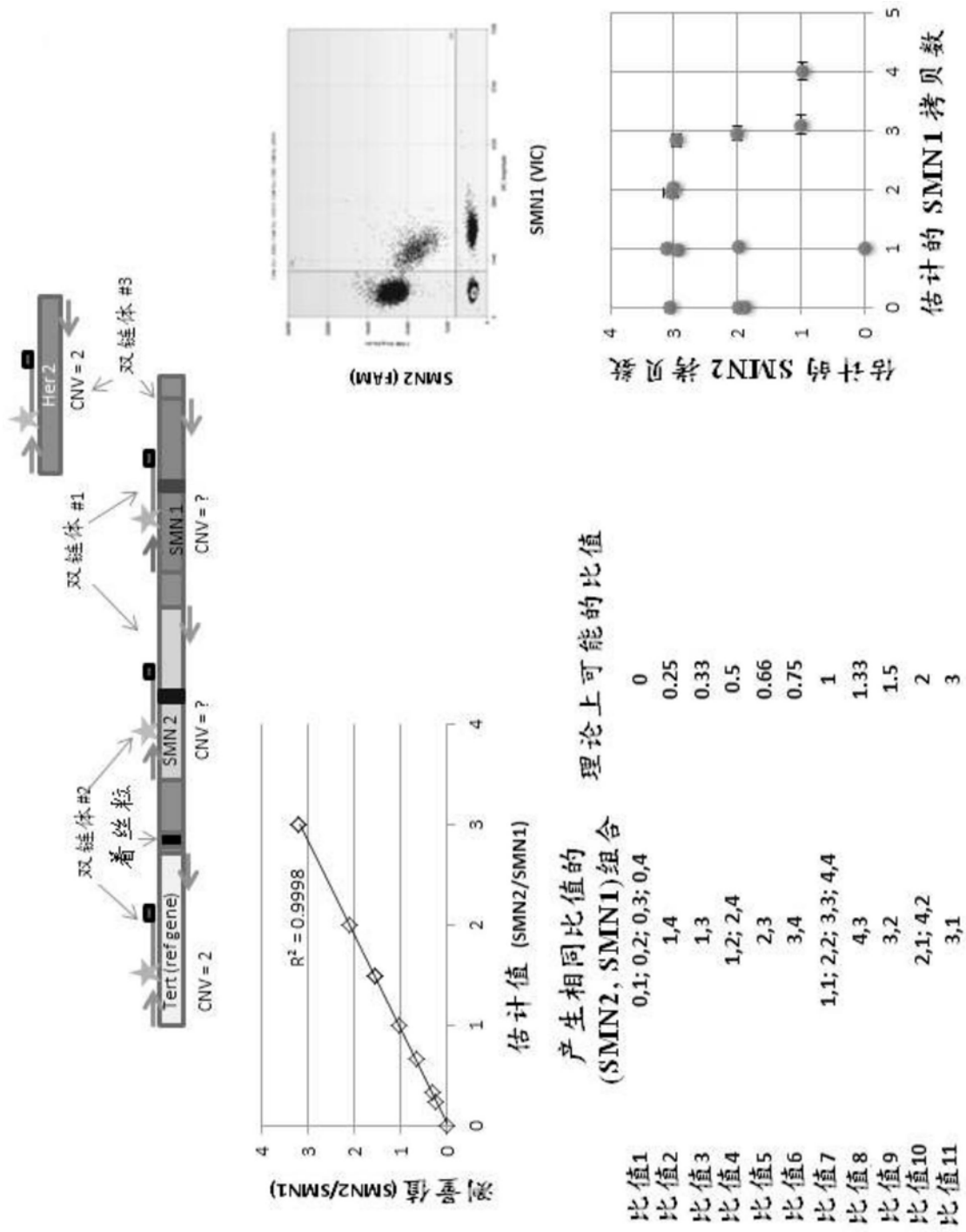


图 28

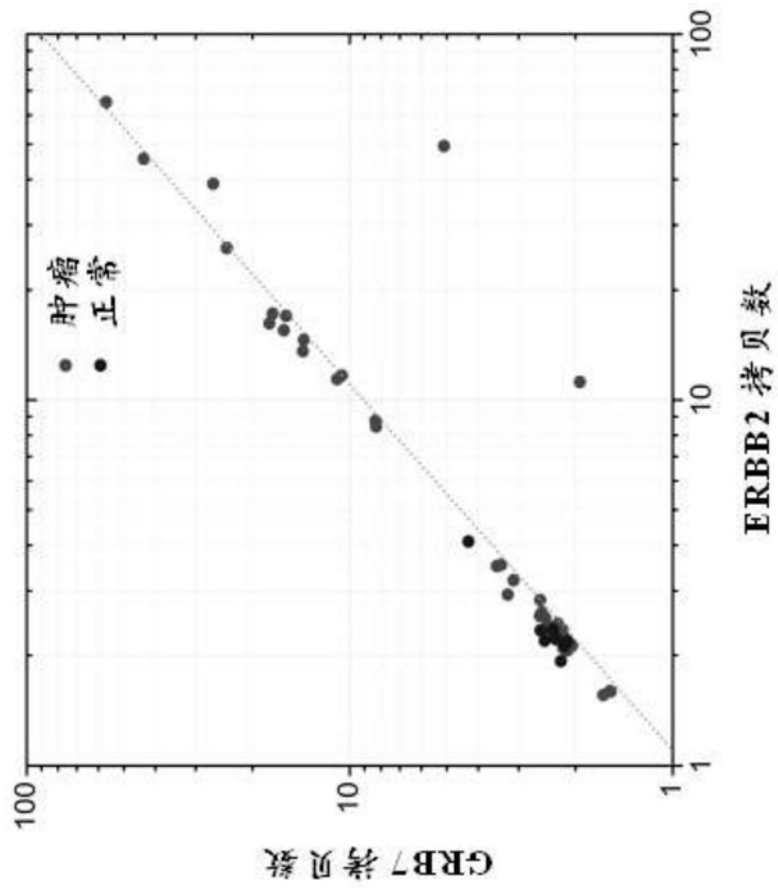


图 29

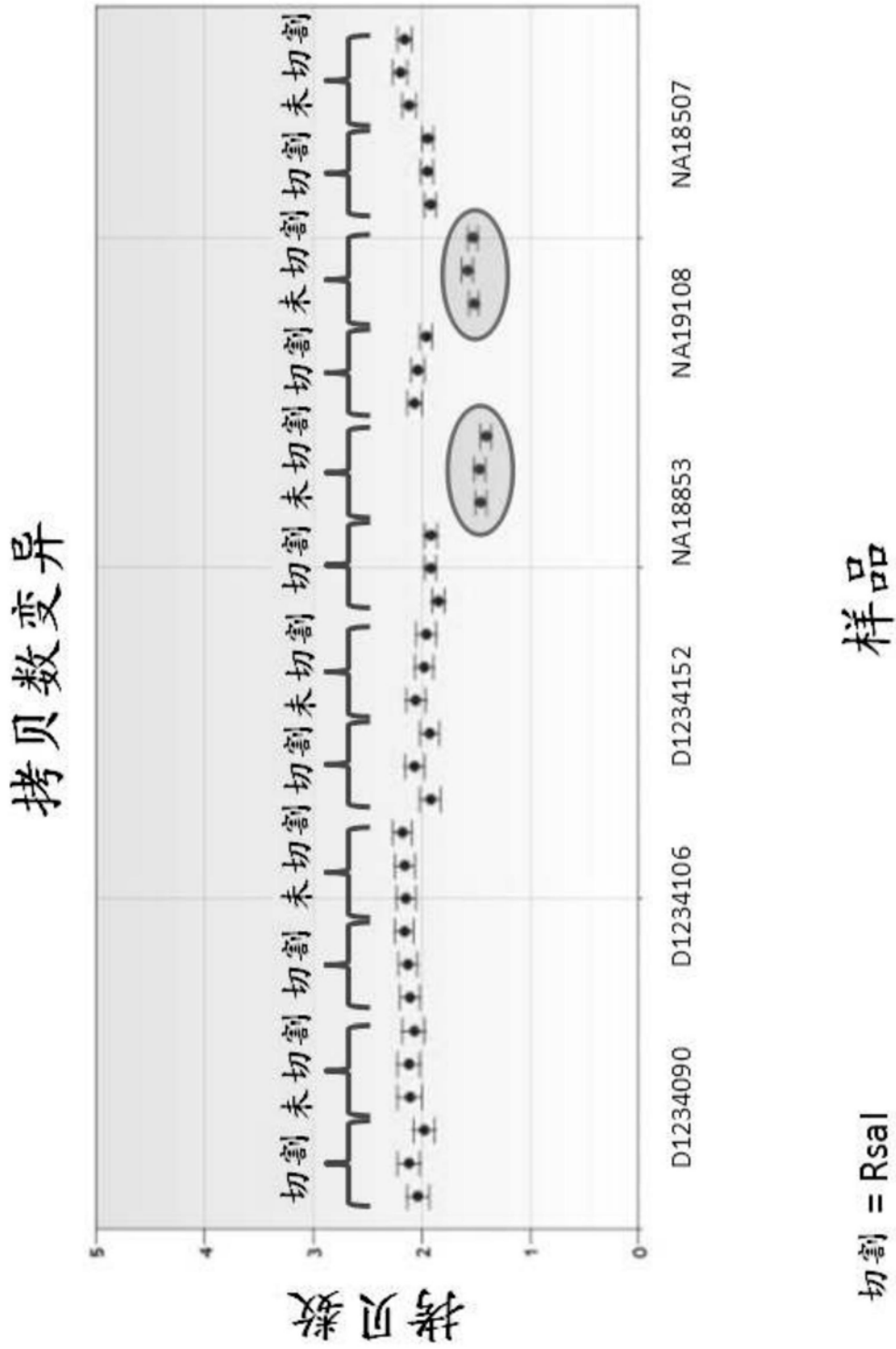


图 30

第 25 次循环的荧光强度

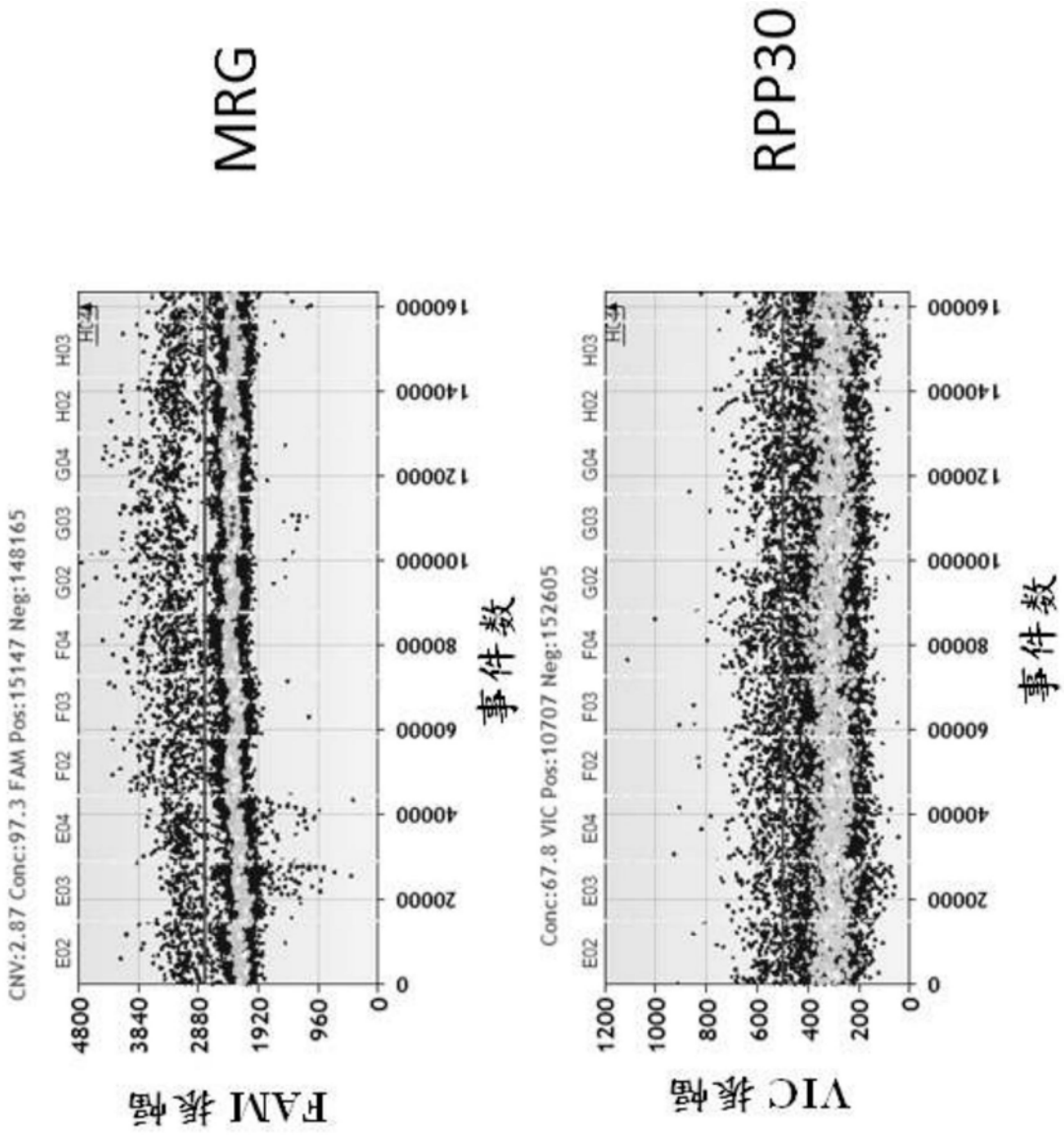


图 31

第 25 次循环时一式三份测量的平均荧光

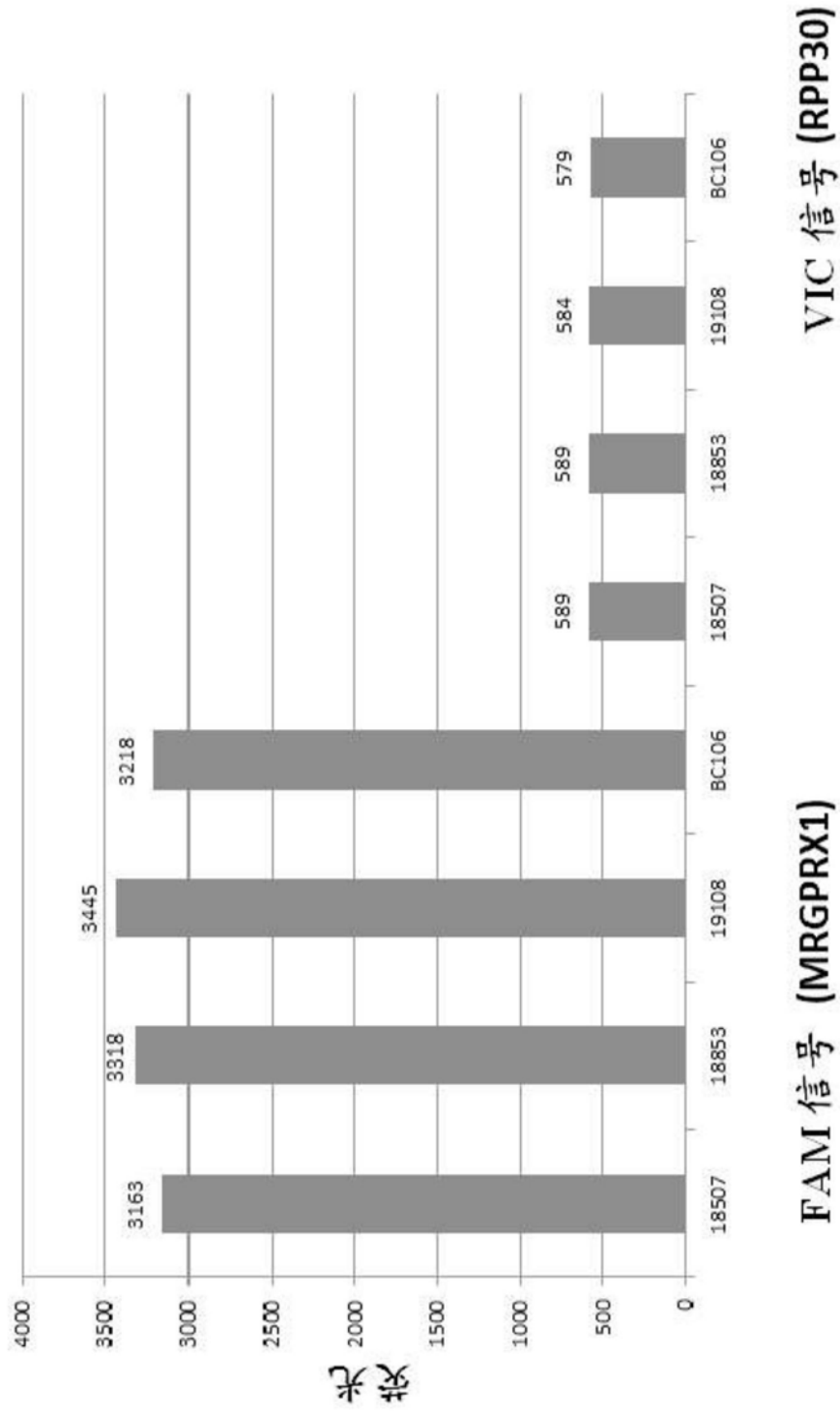


图 32

第 28 次循环的荧光强度

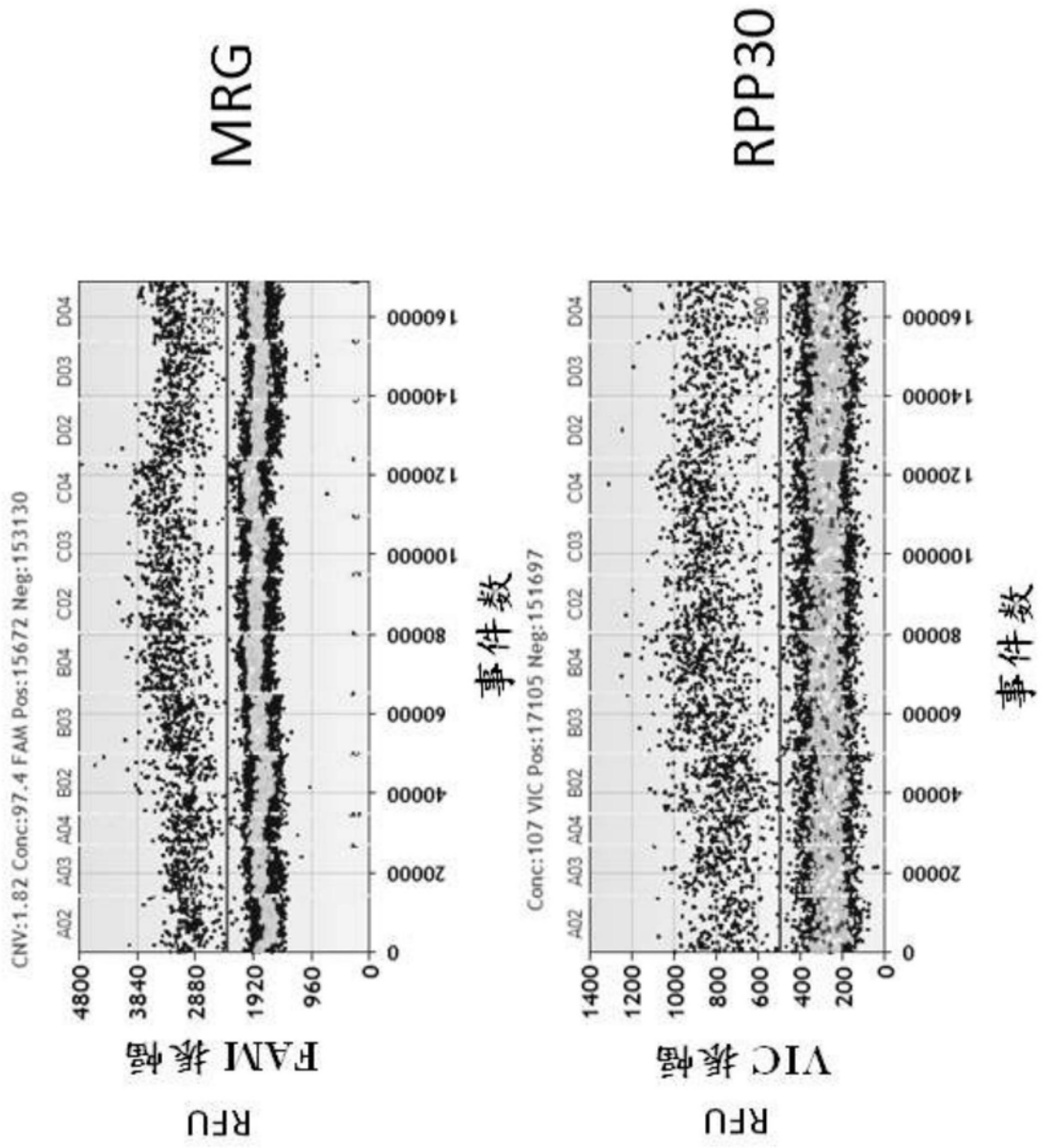


图 33

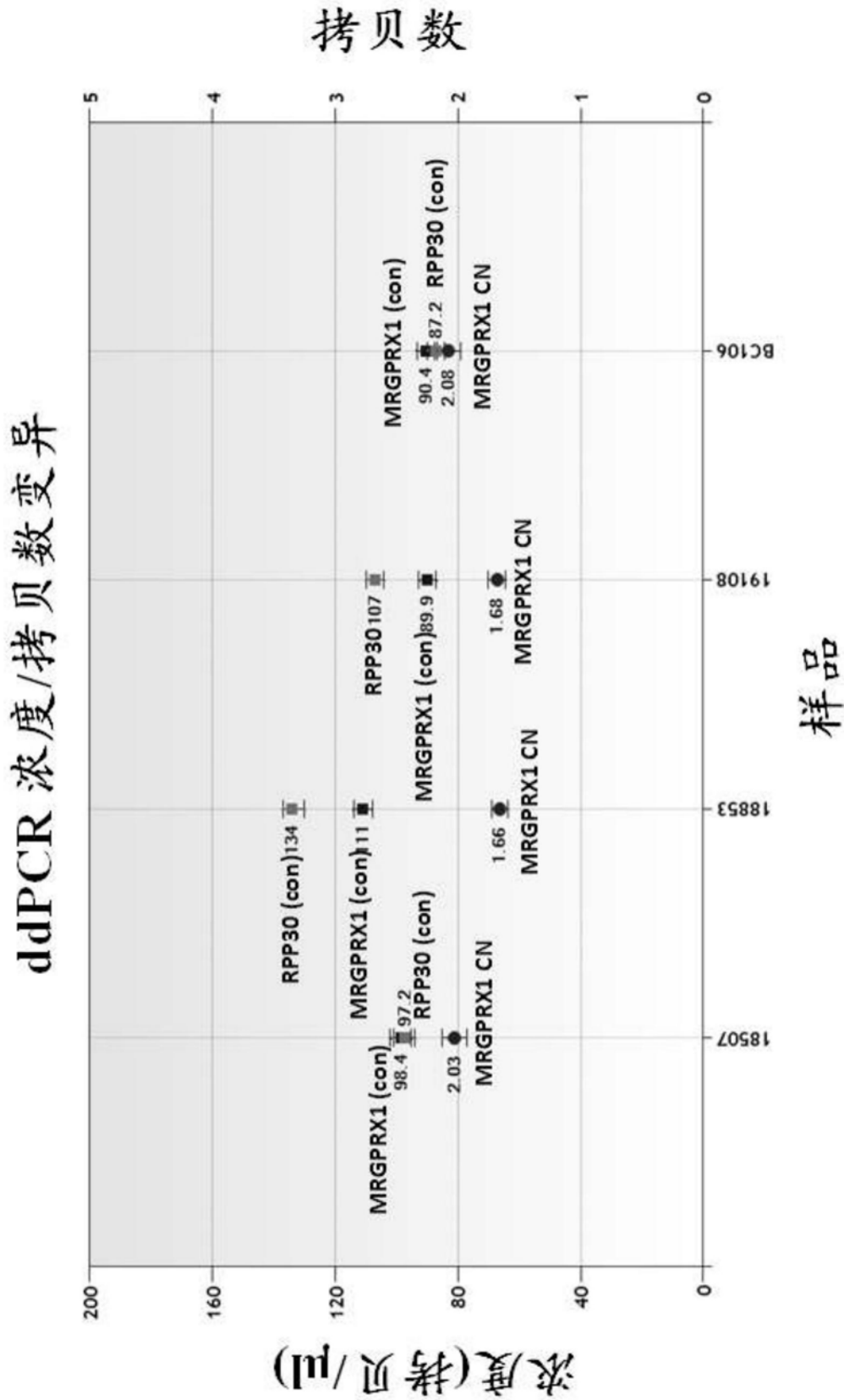


图 34

第 28 次循环时一式三份测量的平均荧光

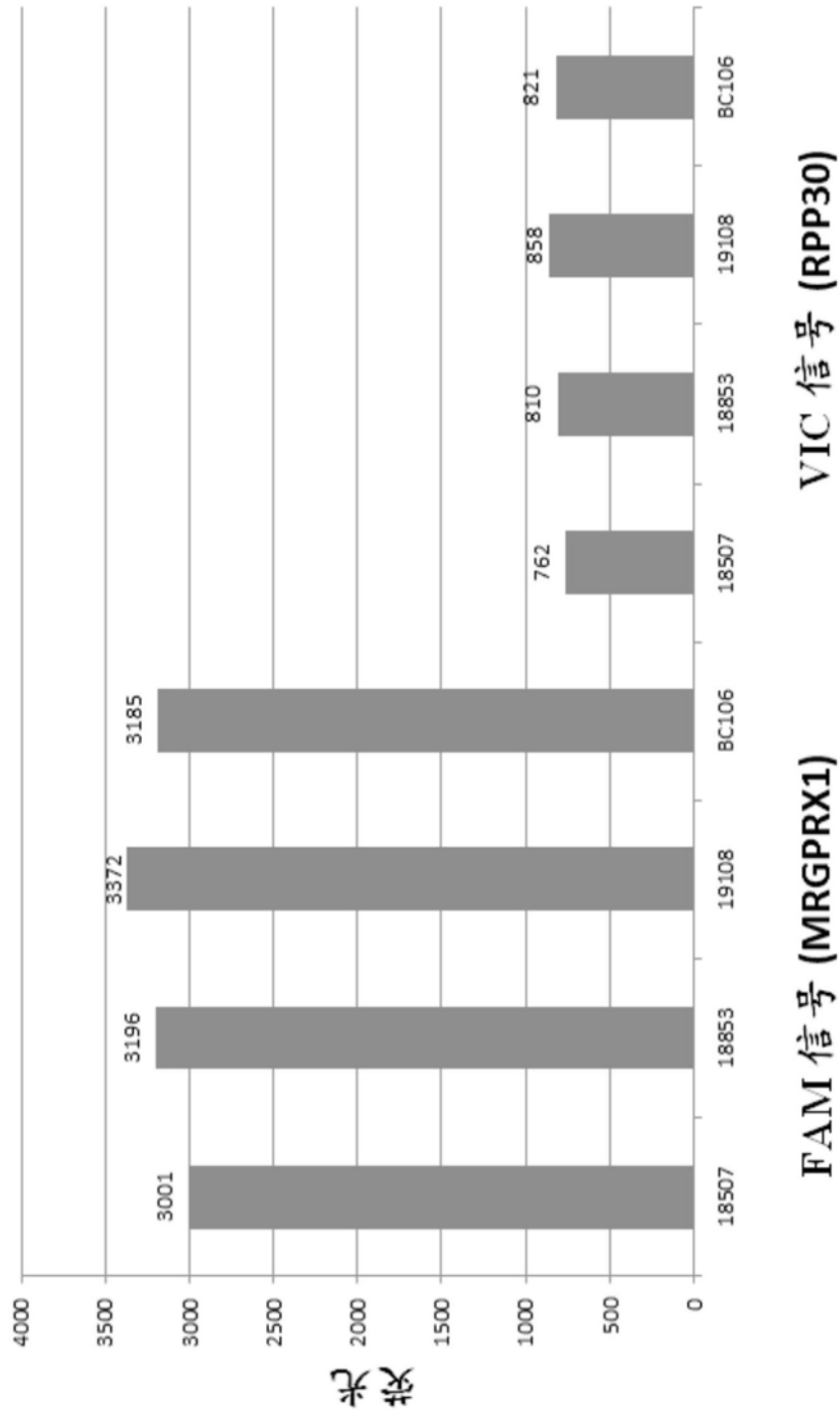
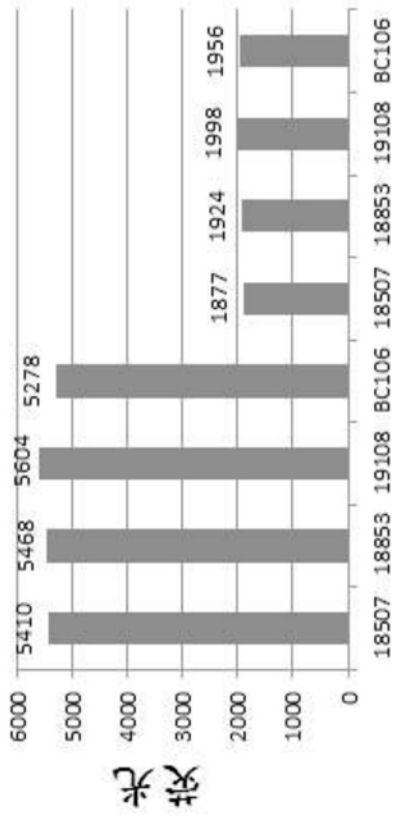


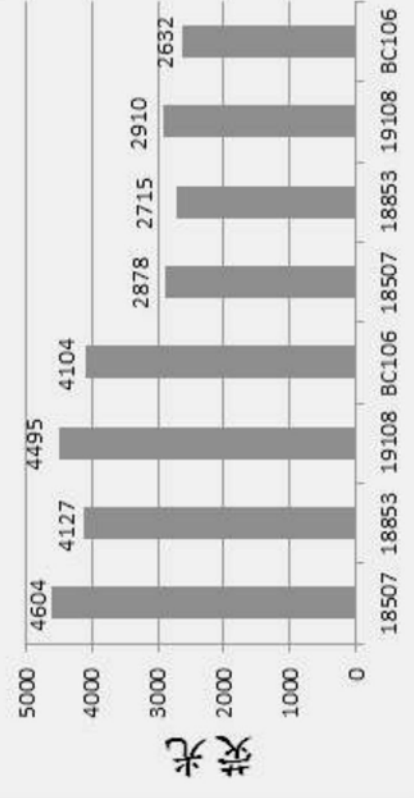
图 35

31 次循环 一式三份测量的平均荧光



40 次循环

一式三份测量的平均荧光



34 次循环

一式三份测量的平均荧光

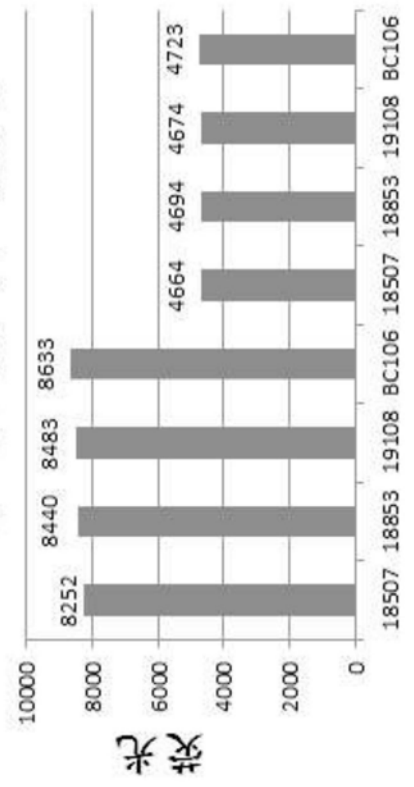


图 36

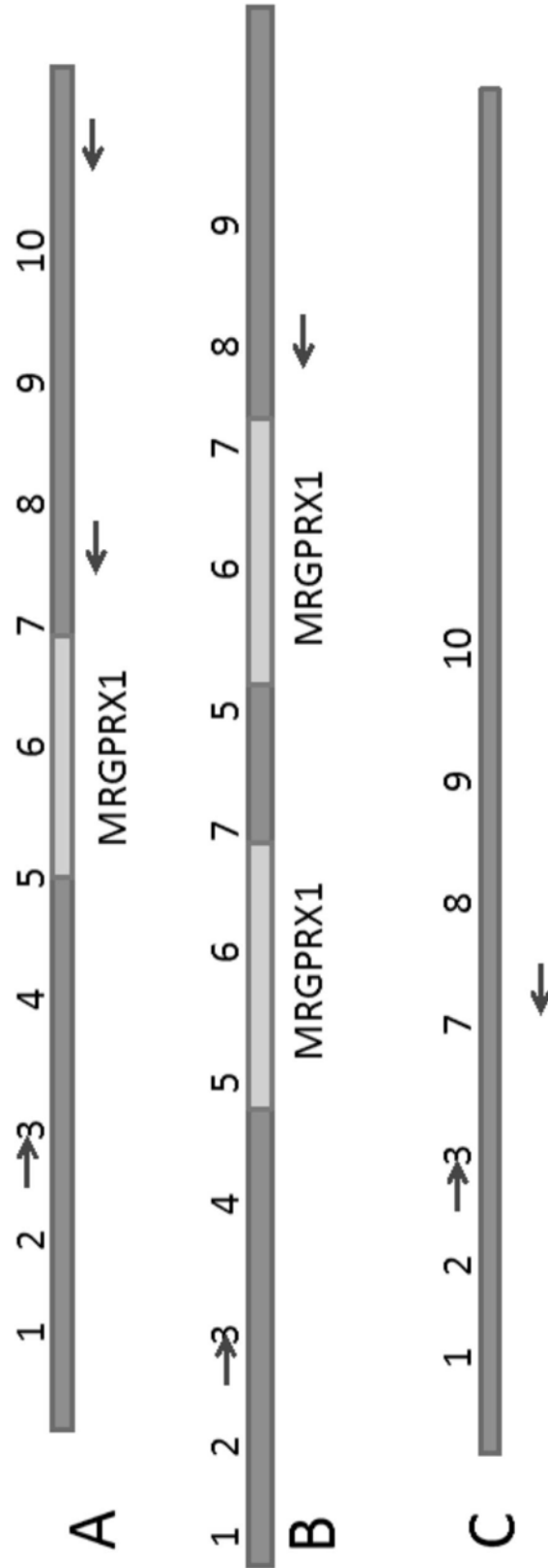


图 37

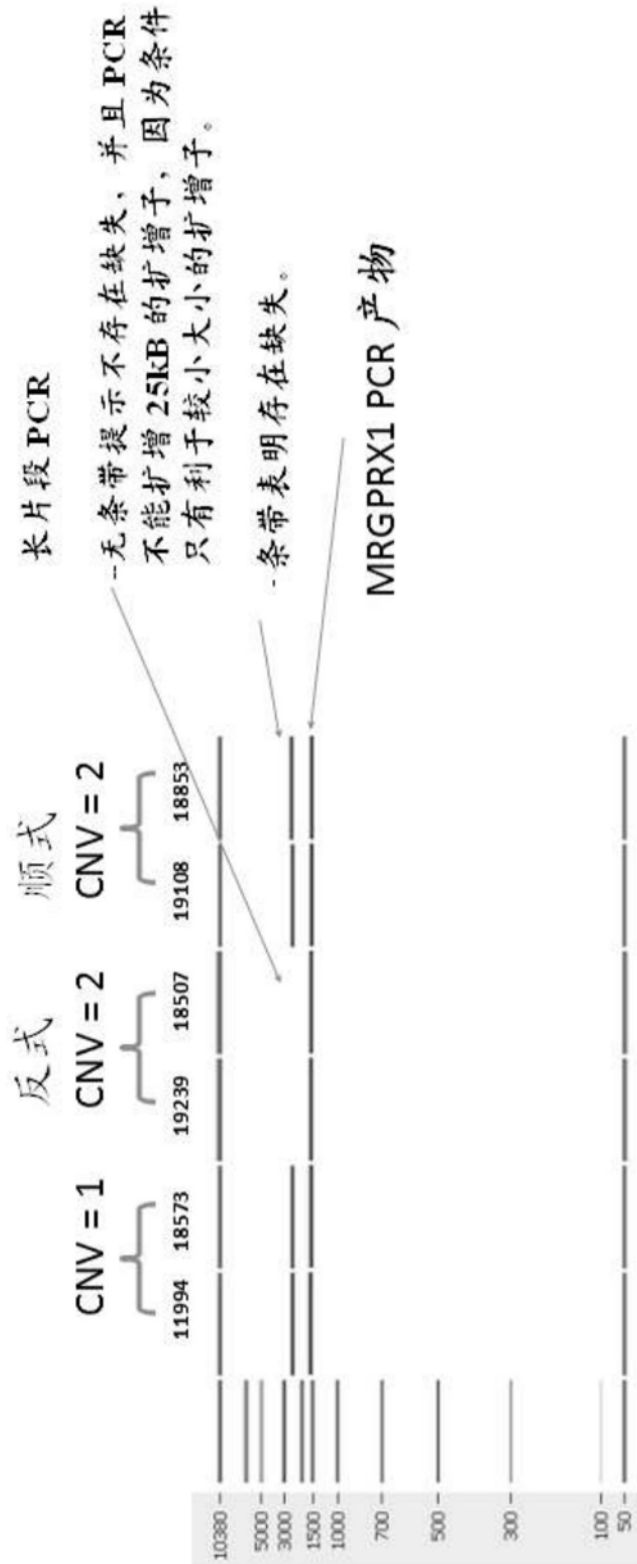


图 38

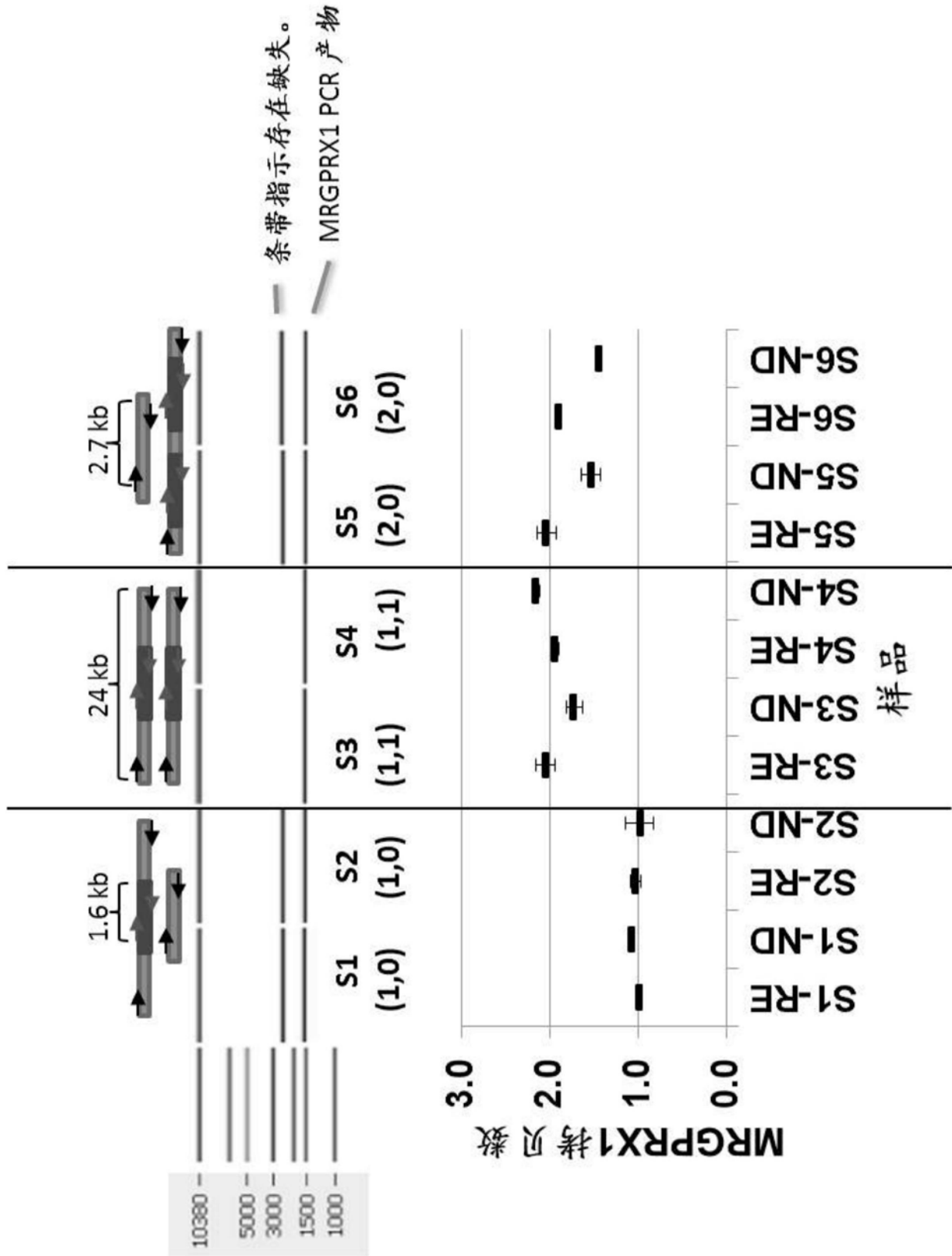


图 39

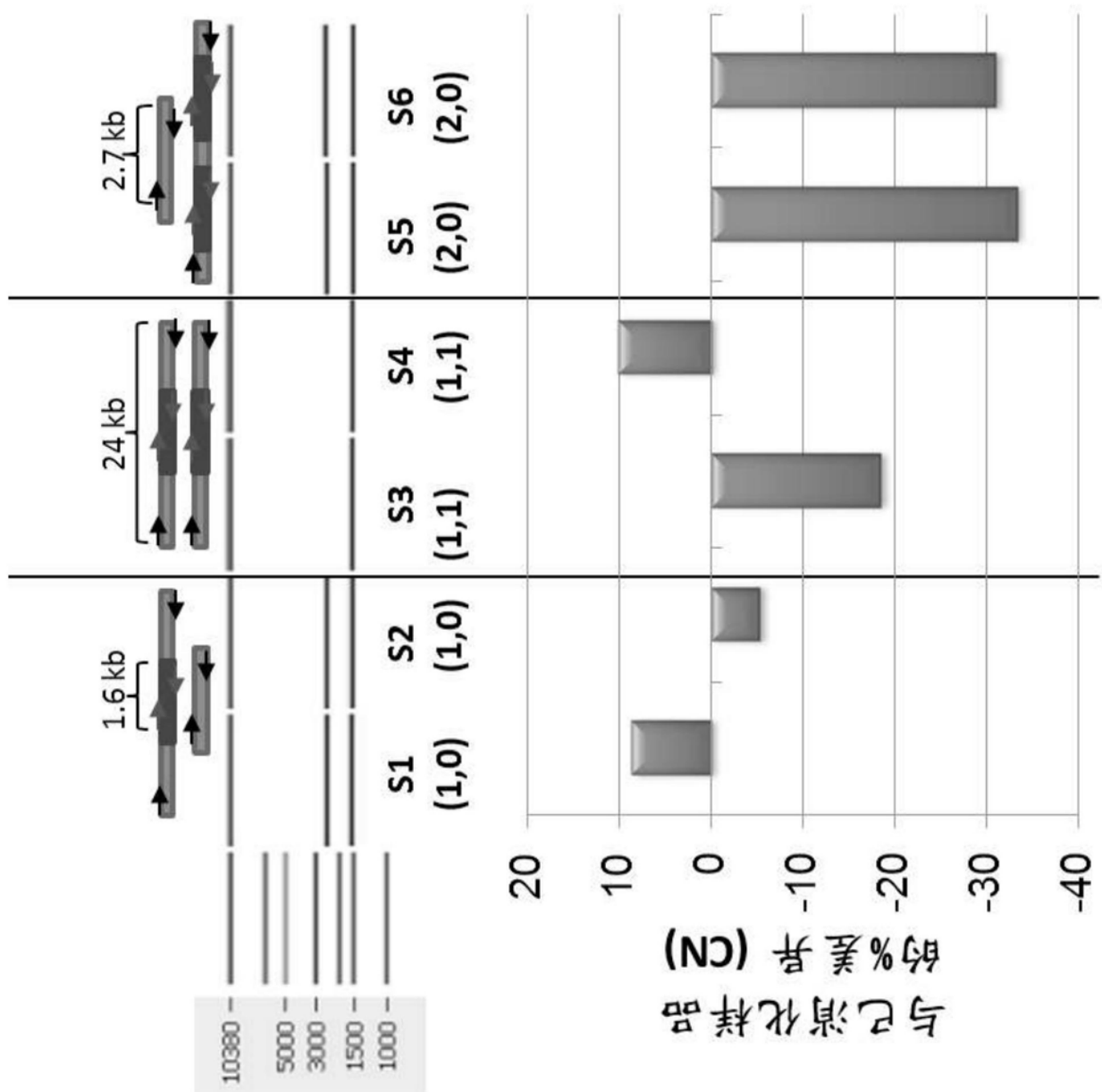


图 40

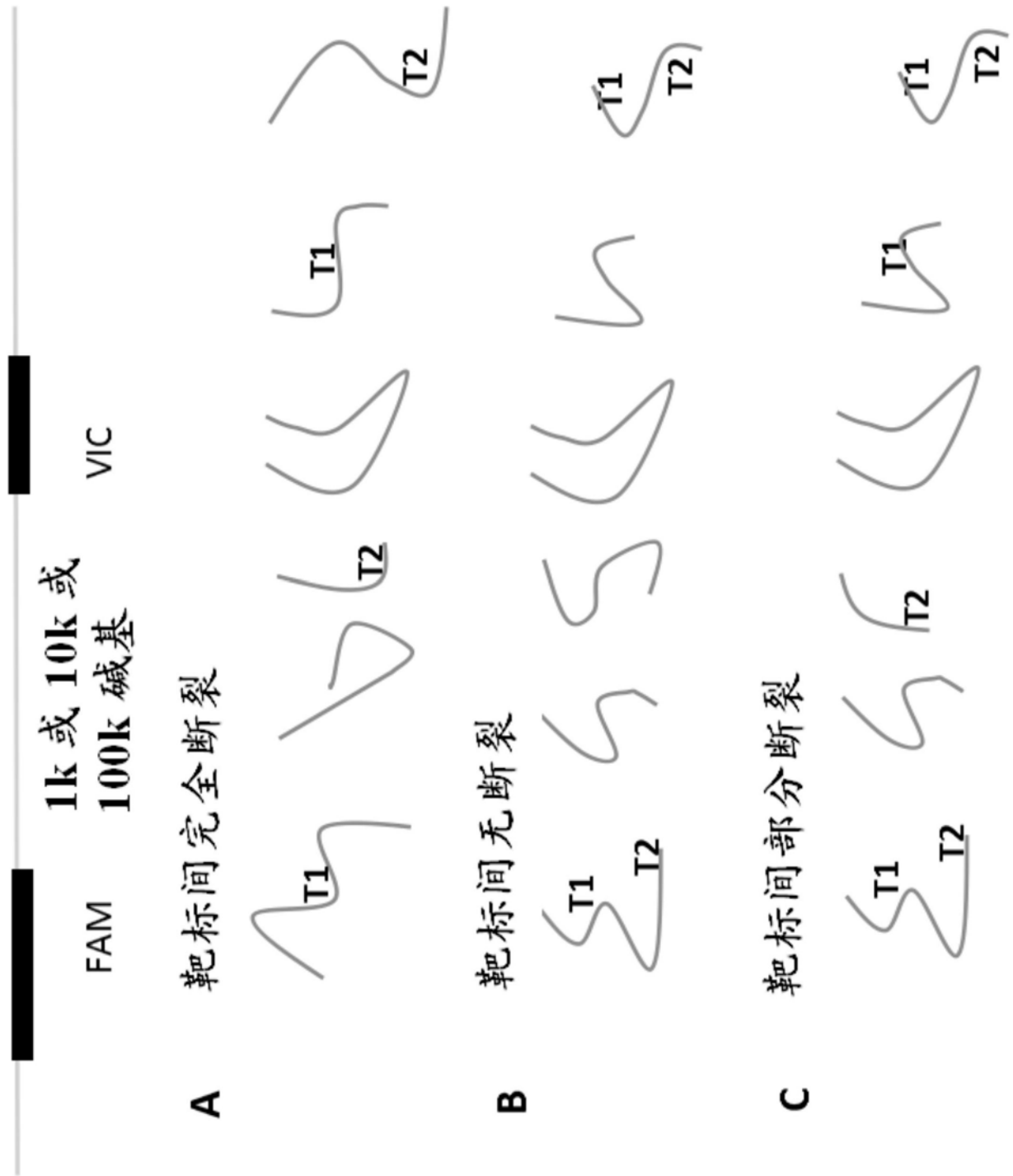


图 41

图 42

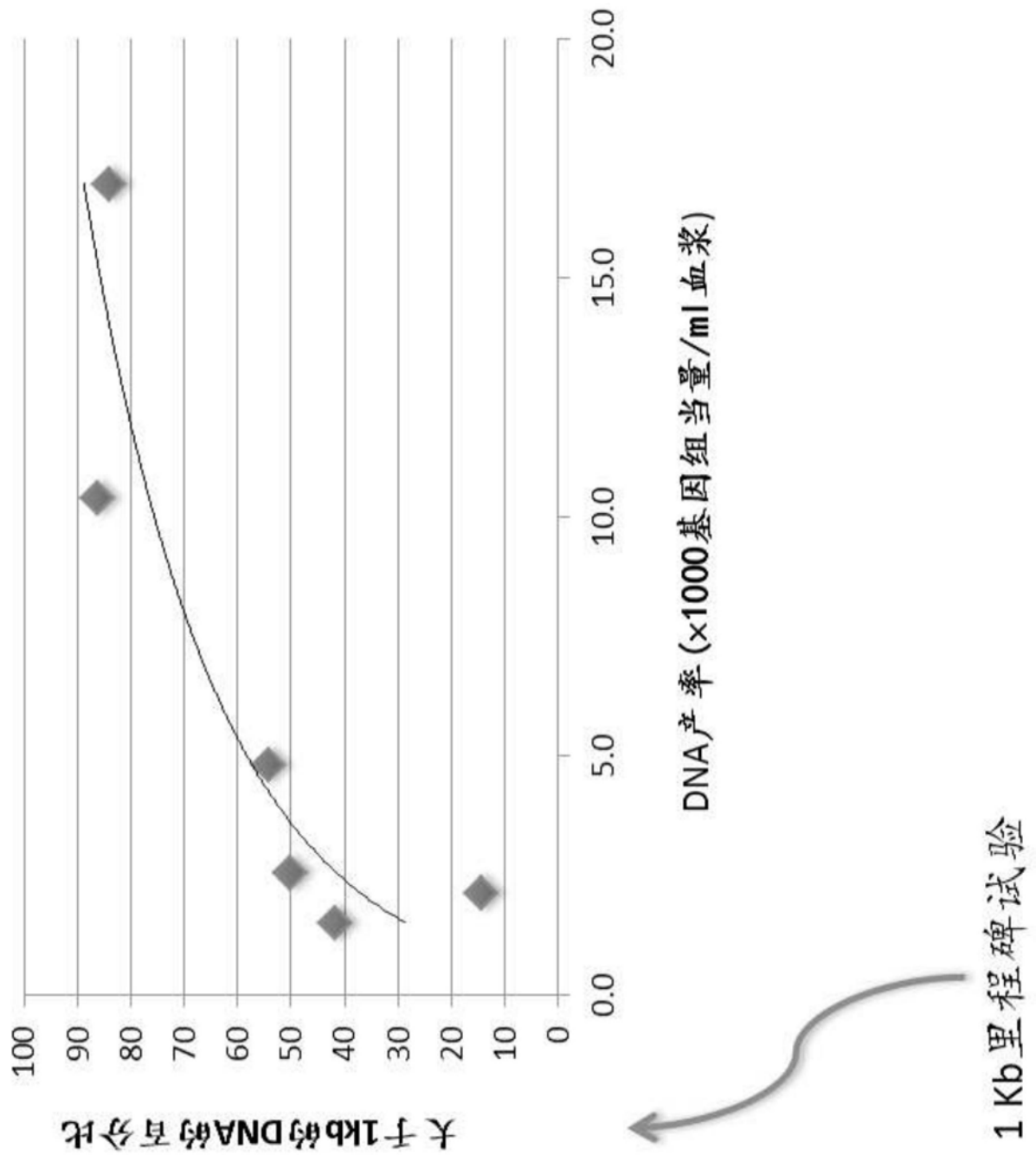


图 43

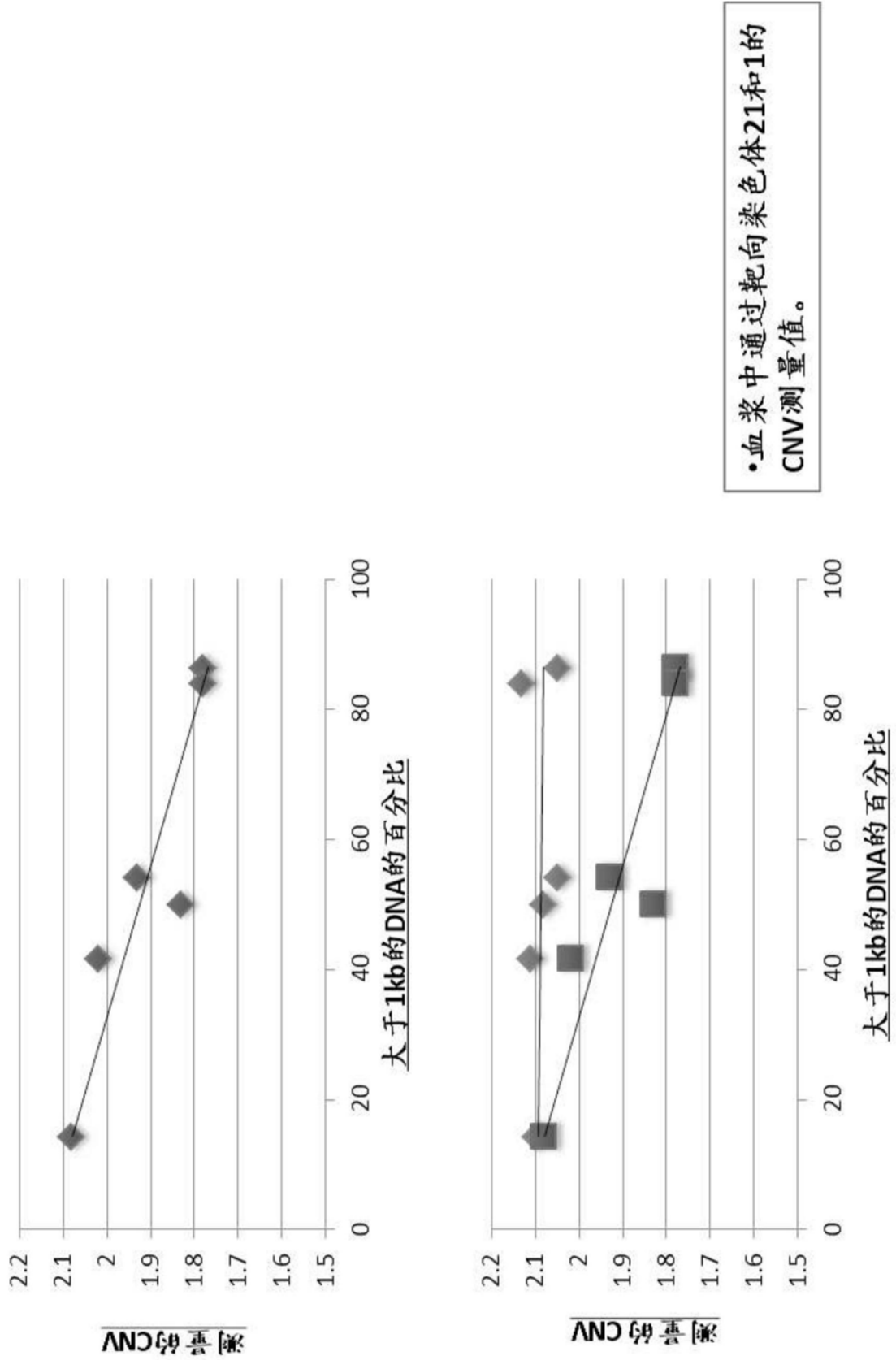


图 44

| 限制酶 | 甲基化敏感性 | 消化缓冲液 | Inc 温度 | 测试试验 | | |
|--------|-----------------|--------------|--------|----------|----------|----------|
| EcoRI | OL CpG | EcoRI 缓冲液 | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| RsaI | COL Dam | NEB 4 | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | Chmr 16 | QL_RPP30 |
| MseI | ok | NEB 4, BSA | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| MboI | OL Dam, CpG Imp | NEB 4 | 37 | MRGPRX1 | Chrm X | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| AluI | ok | NEB 4 | 37 | MRGPRX1? | SRY | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| TaqI | OL Dam | NEB 4, BSA | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| BsmI | ok | NEB 4 | 65 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| BstYI | ok | NEB 2 | 60 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| XhoI | CpG Imp | NEB 4, BSA | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| DpnII | Dam | DpnII Buffer | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| HaeIII | ok | NEB 4 | 37 | MRGPRX1 | Chrm X | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| HpaII | CpG | NEB 1 | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| HhaI | CpG | NEB 4, BSA | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | Chrm 14 | QL_RPP30 | Chrm 14 |
| MspI | ok | NEB 4 | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| TseI | COL CpG | NEB 4 | 65 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |
| BstUI | CpG | NEB 4 | 60 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | Chrm 14 | QL_RPP30 | Chrm 14 |
| NlaIII | ok | NEB 4, BSA | 37 | MRGPRX1 | CYP2D6 | CCL3L1 |
| | | | | QL_RPP30 | QL_RPP30 | QL_RPP30 |

图 45

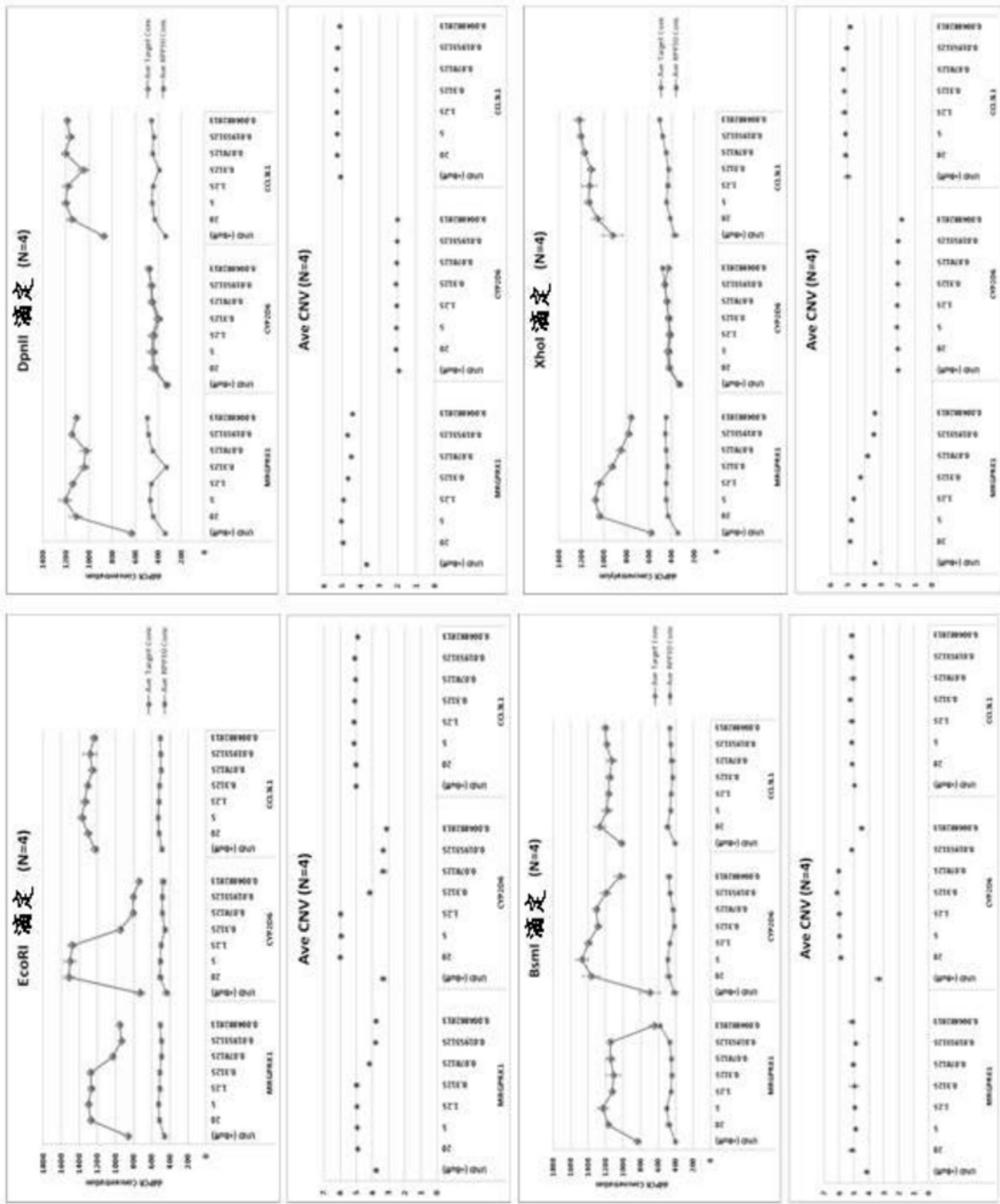


图 46

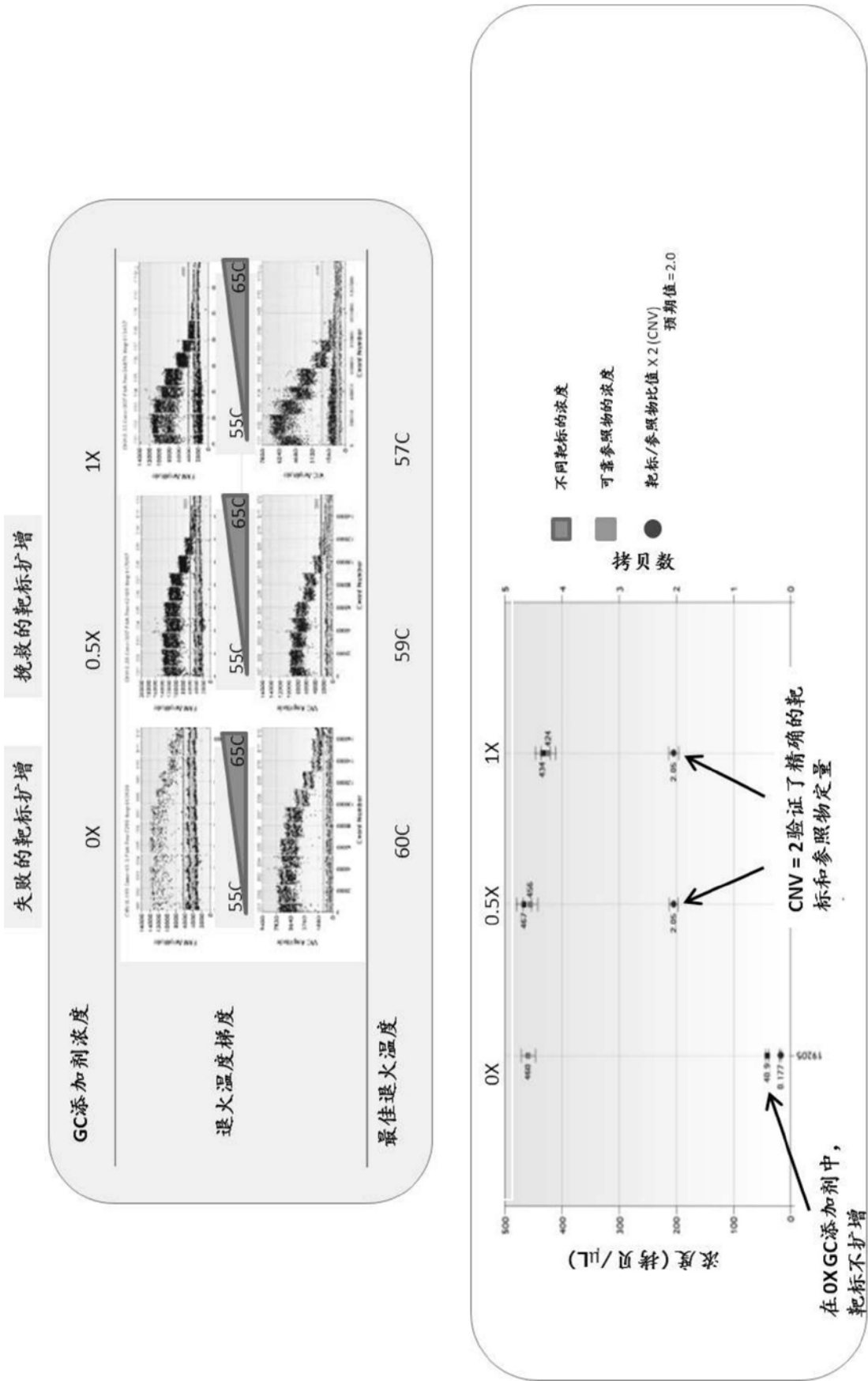


图 47

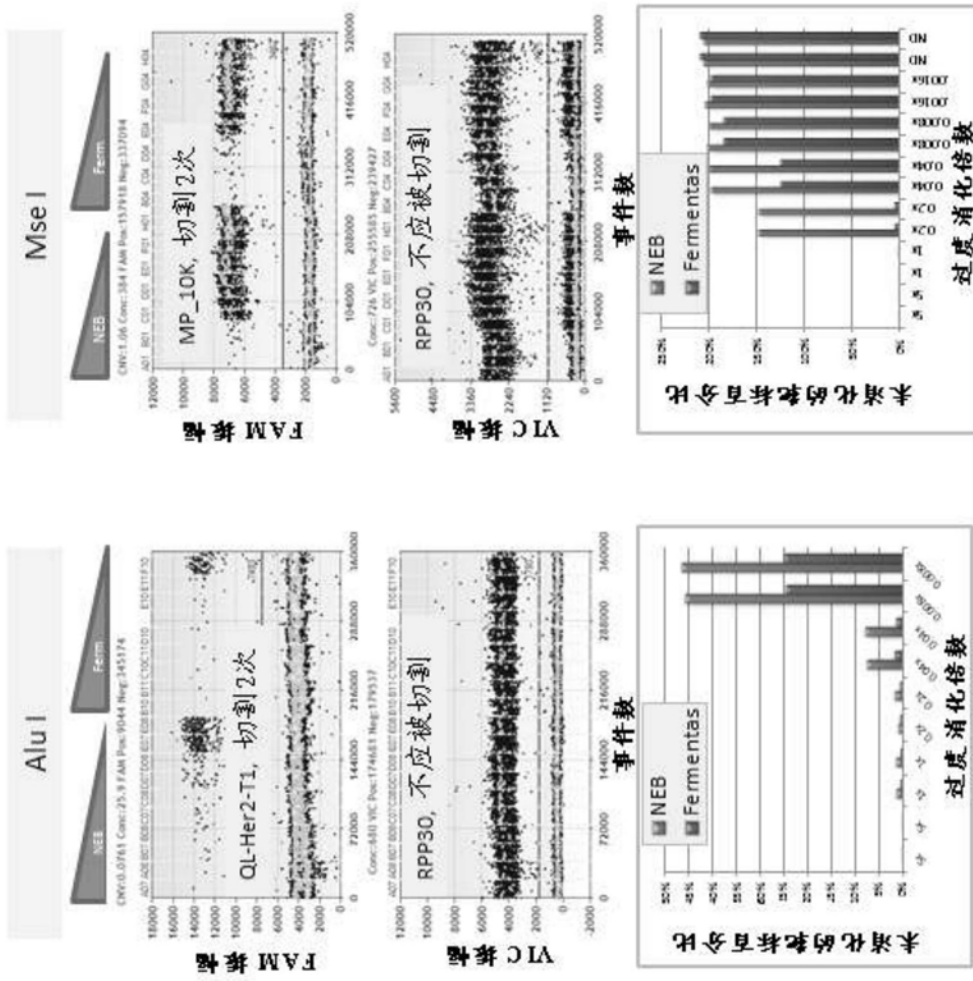


图 48

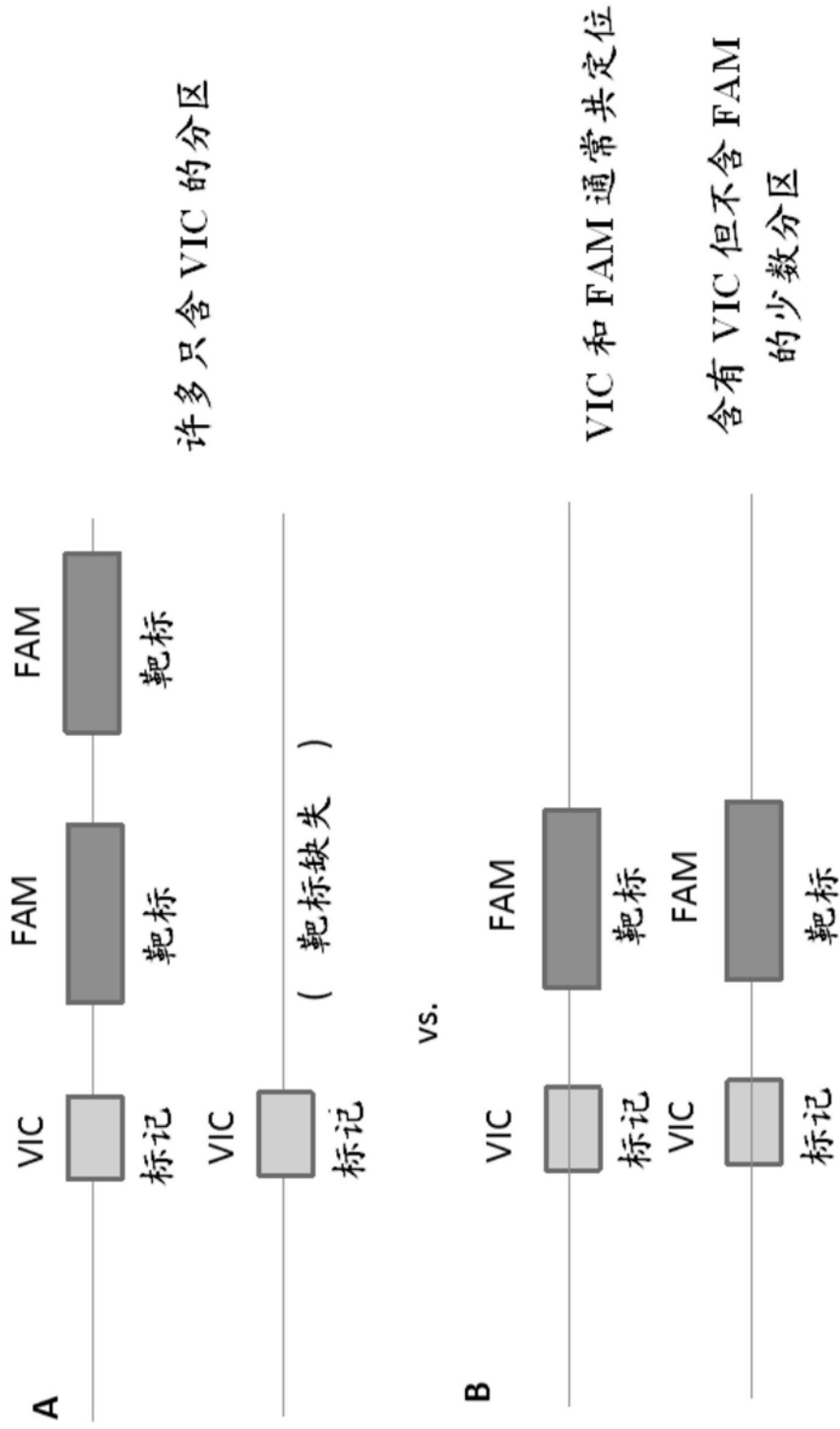


图 49

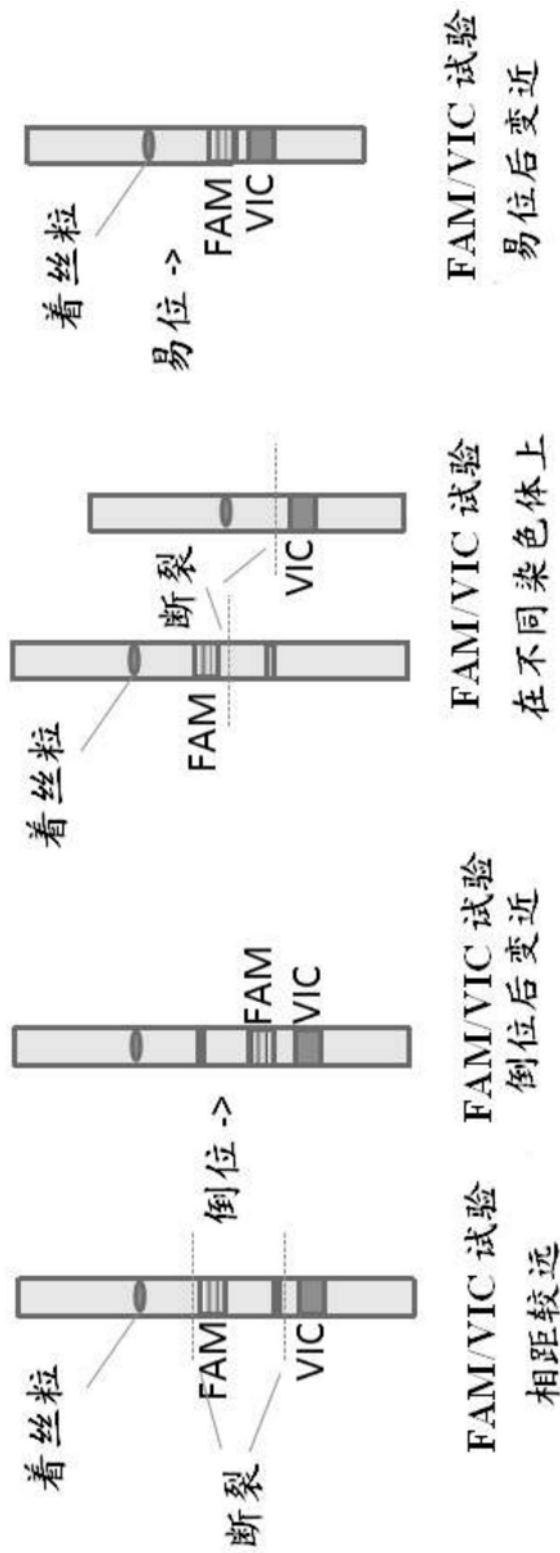


图 50