



## (12)发明专利申请

(10)申请公布号 CN 109844133 A

(43)申请公布日 2019.06.04

(21)申请号 201780063576.4

(22)申请日 2017.08.15

(30)优先权数据

62/375,396 2016.08.15 US

(85)PCT国际申请进入国家阶段日

2019.04.15

(86)PCT国际申请的申请数据

PCT/US2017/047029 2017.08.15

(87)PCT国际申请的公布数据

W02018/035170 EN 2018.02.22

(71)申请人 安可济控股有限公司

地址 开曼群岛大开曼岛

(72)发明人 孙朝辉 翁莉 林盛榕

(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 贺淑东

(51)Int.Cl.

C12Q 1/6806(2018.01)

C12Q 1/6869(2018.01)

C12P 19/34(2006.01)

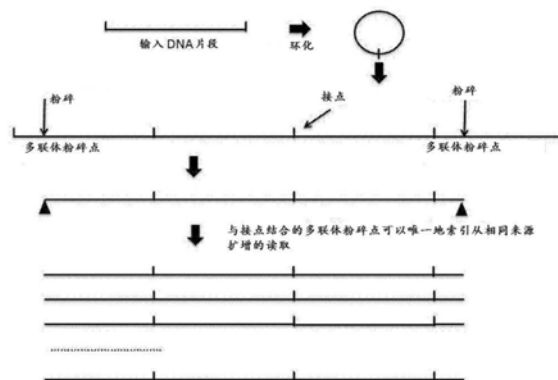
权利要求书6页 说明书105页 附图65页

### (54)发明名称

检测罕见序列变体的组合物和方法

### (57)摘要

在一些方面,本公开内容提供了用于鉴定核酸样品中的序列变体的方法。在一些实施方案中,方法包括鉴定测序读取与参考序列之间的序列差异,以及将出现在至少两种不同的环状多核苷酸例如具有不同接点的两种环状多核苷酸中,或出现在两种不同的剪切多核苷酸中的序列差异判定为序列变体。在一些方面,本公开内容提供了可用于所述方法的组合物和系统。



1. 一种进行滚环扩增的方法,所述方法包括:

(a) 使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中所述多种多核苷酸中的每种多核苷酸在连接前具有5'端和3'端;

(b) 将所述连接酶降解;以及

(c) 在将所述连接酶降解后扩增所述环状多核苷酸,以产生扩增的多核苷酸;

其中不在步骤(a)与(c)之间纯化或分离多核苷酸。

2. 如权利要求1所述的方法,其进一步包括在步骤(a)与(c)之间将线性多核苷酸降解。

3. 如权利要求1所述的方法,其中所述多种多核苷酸为单链。

4. 如权利要求1所述的方法,其中个体环状多核苷酸具有接点,所述接点在环化的多核苷酸中各不相同。

5. 如权利要求1所述的方法,其中环化包括将衔接子多核苷酸接合至所述多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。

6. 如权利要求1所述的方法,其中扩增包括使所述环状多核苷酸经历包含随机引物的扩增反应混合物。

7. 如权利要求1所述的方法,其中扩增包括使所述环状多核苷酸经历包含一种或多种引物的扩增反应混合物,每种所述引物均通过序列互补性与不同的靶序列特异性杂交。

8. 如权利要求1所述的方法,其中所述样品是来自受试者的样品。

9. 如权利要求8所述的方法,其中所述样品是尿液、粪便、血液、唾液、组织或体液。

10. 如权利要求8所述的方法,其中所述样品包含肿瘤细胞。

11. 如权利要求8所述的方法,其中所述样品是福尔马林固定、石蜡包埋的(FFPE)样品。

12. 如权利要求8所述的方法,其进一步包括基于所述判定步骤来诊断和任选地治疗所述受试者。

13. 如权利要求1所述的方法,其中所述序列变体是因果遗传变体。

14. 如权利要求1所述的方法,其中所述序列变体与癌症的类型或阶段相关。

15. 如权利要求1所述的方法,其中所述多种多核苷酸包含无细胞多核苷酸。

16. 如权利要求15所述的方法,其中所述无细胞多核苷酸包含循环肿瘤DNA。

17. 如权利要求1所述的方法,其进一步包括对所述扩增的多核苷酸进行测序,以产生多个测序读取。

18. 如权利要求17所述的方法,其进一步包括鉴定所述测序读取与参考序列之间的序列差异。

19. 如权利要求18所述的方法,其进一步包括仅在以下情况下才将序列差异判定为所述多种多核苷酸中的序列变体:(i)在双链输入分子的两条链上都鉴定到所述序列差异;(ii)所述序列差异出现在通过滚环扩增形成的多联体的共有序列中;和/或(iii)所述序列差异出现在两种不同的分子中。

20. 如权利要求19所述的方法,其中当所述序列差异出现在具有在5'端与3'端之间形成的不同接点的至少两种环状多核苷酸中时,序列差异被鉴定为出现在两种不同的分子中。

21. 如权利要求19或20所述的方法,其中当对应于所述两种不同分子的读取具有不同的5'端和不同的3'端时,序列差异被鉴定为出现在两种不同的分子中。

22. 如权利要求1所述的方法,其进一步包括剪切所述扩增的核苷酸。
23. 一种鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,所述多种多核苷酸中的每种多核苷酸具有5'端和3'端,所述方法包括:
- (a) 将所述多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,每种环状多核苷酸在5'端与3'端之间具有接点;
  - (b) 扩增(a)的所述环状多核苷酸以产生扩增的多核苷酸;
  - (c) 将所述扩增的多核苷酸剪切以产生剪切的多核苷酸,每种剪切的多核苷酸包含在5'端和/或3'端处的一个或多个剪切点;
  - (d) 对所述剪切的多核苷酸进行测序,以产生多个测序读取;
  - (e) 鉴定测序读取与参考序列之间的序列差异;以及
  - (f) 当序列差异出现在至少两种不同的剪切多核苷酸中时,将所述序列差异判定为所述序列变体。
24. 如权利要求23所述的方法,其中当(i)所述序列差异出现在具有不同接点的至少两种环状多核苷酸中;(ii)在双链输入分子的两条链上都鉴定到所述序列差异;和/或(iii)所述序列差异出现在由包括滚环扩增在内的扩增形成的多联体的共有序列中时,进一步出现将所述序列差异判定为所述序列变体。
25. 如权利要求23所述的方法,其中所述多种多核苷酸为单链。
26. 如权利要求23所述的方法,其中通过使所述多种多核苷酸经历连接反应来实现环化。
27. 如权利要求23所述的方法,其中所述序列变体是单核苷酸多态性。
28. 如权利要求23所述的方法,其中所述参考序列是通过将所述测序读取彼此比对而形成的共有序列。
29. 如权利要求23所述的方法,其中所述参考序列是测序读取。
30. 如权利要求23所述的方法,其中环化包括将衔接子多核苷酸接合至所述多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。
31. 如权利要求23所述的方法,其中通过使用具有链置换活性的聚合酶实现扩增。
32. 如权利要求23所述的方法,其中扩增包括使所述环状多核苷酸经历包含随机引物的扩增反应混合物。
33. 如权利要求23所述的方法,其中扩增包括使所述环状多核苷酸经历包含一种或多种引物的扩增反应混合物,每种所述引物均通过序列互补性与不同的靶序列特异性杂交。
34. 如权利要求23所述的方法,其中在未富集的情况下使所述扩增的多核苷酸经历所述测序步骤。
35. 如权利要求23所述的方法,其进一步包括通过在测序之前进行富集步骤来富集所述扩增的多核苷酸中的一种或多种靶多核苷酸。
36. 如权利要求23所述的方法,其中基于所述判定步骤鉴定微生物污染物。
37. 如权利要求23所述的方法,其中所述样品是来自受试者的样品。
38. 如权利要求37所述的方法,其中所述样品是尿液、粪便、血液、唾液、组织或体液。
39. 如权利要求37所述的方法,其中所述样品包含肿瘤细胞。
40. 如权利要求37所述的方法,其中所述样品是福尔马林固定、石蜡包埋的(FFPE)样

品。

41. 如权利要求37所述的方法,其进一步包括基于所述判定步骤来诊断和任选地治疗所述受试者。

42. 如权利要求23所述的方法,其中所述序列变体是因果遗传变体。

43. 如权利要求23所述的方法,其中所述序列变体与癌症的类型或阶段相关。

44. 如权利要求23所述的方法,其中所述多种多核苷酸包含无细胞多核苷酸。

45. 如权利要求44所述的方法,其中所述无细胞多核苷酸包含循环肿瘤DNA。

46. 一种用于进行如权利要求1至43中任一项所述的方法的反应混合物,其中所述反应混合物包含 (a) 多种多联体,其中所述多种多联体中的个体多联体包含通过将具有5' 端和3' 端的个体多核苷酸环化而形成的不同接点; (b) 包含序列A' 的第一引物,其中所述第一引物通过序列A与序列A' 之间的序列互补性与靶序列的序列A特异性杂交; (c) 包含序列B的第二引物,其中所述第二引物通过B与B' 之间的序列互补性与包含所述靶序列的互补序列的互补多核苷酸中存在的序列B' 特异性杂交; 以及 (d) 使所述第一引物和所述第二引物延伸以产生扩增的多核苷酸的聚合酶; 其中所述靶序列中序列A的5' 端与序列B的3' 端之间的距离为75nt或更小。

47. 如权利要求46所述的反应混合物,其中所述第一引物包含在序列A' 的5' 侧的序列C,所述第二引物包含在序列B的5' 侧的序列D,并且在扩增反应中的第一扩增步骤期间,序列C和序列D均不与两种或更多种多联体杂交。

48. 一种用于检测序列变体的系统,其包含 (a) 计算机,所述计算机被配置用于接收关于对样品进行检测反应的用户请求; (b) 扩增系统,所述扩增系统响应于所述用户请求,对所述样品或其部分进行核酸扩增反应,其中所述扩增反应包括以下步骤: (i) 使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中所述多种多核苷酸中的每种多核苷酸在连接前具有5' 端和3' 端; (ii) 将所述连接酶降解; 以及 (iii) 在将所述连接酶降解后扩增所述环状多核苷酸,以产生扩增的多核苷酸; 其中不在步骤 (i) 与 (iii) 之间纯化或分离多核苷酸; (c) 测序系统,所述测序系统产生由所述扩增系统扩增的多核苷酸的测序读取,鉴定测序读取与参考序列之间的序列差异,并将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为所述序列变体; 以及 (d) 报告生成器,所述报告生成器向接收者发送报告,其中所述报告包含所述序列变体的检测结果。

49. 如权利要求48所述的系统,其中所述接收者是所述用户。

50. 一种包含代码的计算机可读介质,所述代码在由一个或多个处理器执行时实施检测序列变体的方法,所实施的方法包括: (a) 接收关于对样品进行检测反应的客户请求; (b) 响应于所述客户请求,对所述样品或其部分进行核酸扩增反应,其中所述扩增反应包括以下步骤: (i) 使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中所述多种多核苷酸中的每种多核苷酸在连接前具有5' 端和3' 端; (ii) 将所述连接酶降解; 以及 (iii) 在将所述连接酶降解后扩增所述环状多核苷酸,以产生扩增的多核苷酸; 其中不在步骤 (i) 与 (iii) 之间纯化或分离多核苷酸; (c) 进行包括以下步骤的测序分析: (i) 产生在所述扩增反应中扩增的多核苷酸的测序读取; (ii) 鉴定测序读取与参考序列之间的序列差异; 以及 (iii) 将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为所述序列变体; 以及 (d) 生成包含所述序列变体的测序结果的报告。



51. 一种鉴定包含多种多核苷酸的核酸样品中的序列变体的方法, 所述多种多核苷酸中的每种多核苷酸具有5'端和3'端, 所述方法包括:

(a) 将所述多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸, 其中所述多种多核苷酸中的给定环状多核苷酸具有由所述环化产生的接合序列;

(b) 扩增(a)的环状多核苷酸, 以产生多种扩增的多核苷酸, 其中所述多种扩增多核苷酸中的第一扩增多核苷酸和所述多种扩增多核苷酸中的第二扩增多核苷酸包含所述接合序列, 但在其各自的5'端和/或3'端包含不同的序列;

(c) 对所述多种扩增多核苷酸或其扩增产物进行测序, 以产生对应于所述第一扩增多核苷酸和所述第二扩增多核苷酸的多个测序读取; 以及

(d) 当所述序列差异出现在对应于所述第一扩增多核苷酸和所述第二扩增多核苷酸的测序读取中时, 将所述测序读取中检测到的序列差异判定为所述序列变体。

52. 如权利要求51所述的方法, 其中(a)中的将个体多核苷酸环化通过连接酶实现。

53. 如权利要求52所述的方法, 其中在(b)之前, 所述连接酶被降解。

54. 如权利要求51所述的方法, 其中未环化的多核苷酸在(b)之前被降解。

55. 如权利要求51-54中任一项所述的方法, 其中所述多种环状多核苷酸在(b)之前未纯化或分离。

56. 如权利要求51所述的方法, 其中(a)中的环化包括将衔接子多核苷酸接合至所述多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。

57. 如权利要求51所述的方法, 其中(b)中的将所述环状多核苷酸扩增通过具有链置换活性的聚合酶实现。

58. 如权利要求51所述的方法, 其中(b)中的将所述环状多核苷酸扩增包括滚环扩增(RCA)。

59. 如权利要求51所述的方法, 其中(b)中的扩增包括使所述环状多核苷酸经历包含随机引物的扩增反应混合物。

60. 如权利要求59所述的方法, 其中个体随机引物包含在其各自的5'和/或3'端的彼此不同的序列。

61. 如权利要求51所述的方法, 其中(b)中的扩增包括使所述环状多核苷酸经历包含靶标特异性引物的扩增反应混合物。

62. 如权利要求51所述的方法, 其中扩增包括变性、引物结合和引物延伸的多个循环。

63. 如权利要求51所述的方法, 其中在未富集的情况下使所述扩增的多核苷酸经历(c)的所述测序。

64. 如权利要求51所述的方法, 其进一步包括通过在(c)的所述测序之前进行富集步骤, 在所述扩增的多核苷酸或其扩增产物中富集一种或多种靶多核苷酸。

65. 如权利要求51所述的方法, 其中所述多种多核苷酸包含单链多核苷酸。

66. 如权利要求51所述的方法, 其中所述序列变体是单核苷酸多态性。

67. 如权利要求51所述的方法, 其中所述样品是来自受试者的样品。

68. 如权利要求67所述的方法, 其中所述样品包括尿液、粪便、血液、唾液、组织或体液。

69. 如权利要求67所述的方法, 其中所述样品包含肿瘤细胞。

70. 如权利要求67所述的方法, 其中所述样品包含福尔马林固定、石蜡包埋的(FFPE)样

品。

71. 如权利要求51所述的方法,其中所述多种多核苷酸包含无细胞多核苷酸。

72. 如权利要求71所述的方法,其中所述无细胞多核苷酸包含无细胞DNA。

73. 如权利要求71所述的方法,其中所述无细胞多核苷酸包含无细胞RNA。

74. 如权利要求71所述的方法,其中所述无细胞多核苷酸包含循环肿瘤DNA。

75. 如权利要求71所述的方法,其中所述无细胞多核苷酸包含循环肿瘤RNA。

76. 如权利要求51所述的方法,其中在(d)中,当序列差异出现在来自所述第一扩增多核苷酸的至少50%的测序读取和来自所述第二扩增多核苷酸的至少50%的测序读取中时,将所述测序读取中所检测到的所述序列差异判定为所述序列变体。

77. 一种鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,所述多种多核苷酸中的每种多核苷酸具有5'端和3'端,所述方法包括:

(a) 将所述多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,其中所述多种环状多核苷酸中的给定环状多核苷酸具有由所述环化产生的接合序列;

(b) 扩增(a)的环状多核苷酸,以产生多种扩增的多核苷酸;

(c) 将所述扩增的多核苷酸剪切,以产生剪切的多核苷酸,每种剪切的多核苷酸在5'端和/或3'端包含一个或多个剪切点;

(d) 对所述剪切的多核苷酸的扩增产物进行测序,以产生多个测序读取;以及

(e) 当序列差异出现在对应于第一剪切多核苷酸和第二剪切多核苷酸的测序读取中时,将所述测序读取中所检测到的所述序列差异判定为所述序列变体。

78. 如权利要求77所述的方法,其中(a)中的将个体多核苷酸环化通过连接酶实现。

79. 如权利要求78所述的方法,其中在(b)之前,所述连接酶被降解。

80. 如权利要求77所述的方法,其中未环化的多核苷酸在(b)之前被降解。

81. 如权利要求77-80中任一项所述的方法,其中所述多种环状多核苷酸在(b)之前未纯化或分离。

82. 如权利要求77所述的方法,其中(a)中的环化包括将衔接子多核苷酸接合至所述多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。

83. 如权利要求77所述的方法,其中(b)中的将所述环状多核苷酸扩增通过具有链置换活性的聚合酶实现。

84. 如权利要求77所述的方法,其中(b)中的将所述环状多核苷酸扩增包括滚环扩增(RCA)。

85. 如权利要求77所述的方法,其中(b)中的扩增包括使所述环状多核苷酸经历包含随机引物的扩增反应混合物。

86. 如权利要求85所述的方法,其中个体随机引物包含在其各自的5'和/或3'端的彼此不同的序列。

87. 如权利要求77所述的方法,其中(b)中的扩增包括使所述环状多核苷酸经历包含靶标特异性引物的扩增反应混合物。

88. 如权利要求77所述的方法,其中在未富集的情况下使所述剪切的多核苷酸的所述扩增产物经历所述测序。

89. 如权利要求77所述的方法,其进一步包括通过在(d)的所述测序之前进行富集步

骤,在所述剪切的多核苷酸的所述扩增产物中富集一种或多种靶多核苷酸。

90.如权利要求77所述的方法,其中所述多种多核苷酸包含单链多核苷酸。

91.如权利要求77所述的方法,其中所述序列变体是单核苷酸多态性。

92.如权利要求77所述的方法,其中所述样品是来自受试者的样品。

93.如权利要求92所述的方法,其中所述样品包括尿液、粪便、血液、唾液、组织或体液。

94.如权利要求92所述的方法,其中所述样品包含肿瘤细胞。

95.如权利要求92所述的方法,其中所述样品包含福尔马林固定、石蜡包埋的 (FFPE) 样品。

96.如权利要求77所述的方法,其中所述多种多核苷酸包含无细胞多核苷酸。

97.如权利要求96所述的方法,其中所述无细胞多核苷酸包含无细胞DNA。

98.如权利要求96所述的方法,其中所述无细胞多核苷酸包含无细胞RNA。

99.如权利要求96所述的方法,其中所述无细胞多核苷酸包含循环肿瘤DNA。

100.如权利要求96所述的方法,其中所述无细胞多核苷酸包含循环肿瘤RNA。

101.如权利要求77所述的方法,其中在(e)中,当序列差异出现在来自所述第一剪切多核苷酸的至少50%的测序读取和来自所述第二剪切的多核苷酸的至少50%的测序读取中时,将所述测序读取中所检测到的所述序列差异判定为所述序列变体。

102.如权利要求77-101中任一项所述的方法,其中(c)中的剪切包括使所述扩增的多核苷酸经历超声处理。

103.如权利要求77-101中任一项所述的方法,其中(c)中的剪切包括使所述扩增的多核苷酸经历酶促切割。

## 检测罕见序列变体的组合物和方法

### 交叉引用

[0001] 本申请要求于2016年8月15日提交的美国临时专利申请号62/375,396的权益,该临时申请通过引用以其全文并入于此用于所有目的。

### 背景技术

[0002] 在复杂群体内鉴定序列变异是一个正在蓬勃发展的领域,特别是随着大规模平行核酸测序技术的出现。然而,由于常用技术的固有误差频率比群体内许多实际序列变异的频率更大,大规模平行测序具有显著的局限性。例如,在标准的高通量测序中已报道了0.1-1%的误差率。当变体频率低,如等于或低于该误差率时,对罕见序列变体的检测具有高的假阳性率。

[0003] 经常存在检测罕见序列变体的迫切需要。例如,检测罕见特征性序列可用于鉴定和区分有害环境污染物如细菌分类群的存在。表征细菌分类群的一种常见方法是鉴定高度保守序列如rRNA序列的差异。然而,迄今为止典型的基于测序的方法面临与给定样品中不同基因组的绝对数量和成员之间的同源性程度相关的挑战,从而为本已繁琐的方法带来复杂的问题。改进的方法将会有在各种情境下增强污染检测的潜力。例如,可以用本系统和方法来调查用于组装卫星和其他航天器组件的洁净室,以了解存在哪些微生物群落,并开发更好的去污和清洁技术来防止将陆地微生物引入其他行星或其样品中,或开发用于将假定的外星微生物所产生的数据与污染陆地微生物所产生的数据相区分的方法。食品监测应用包括针对食源性病原体定期检测食品加工厂的生产线,调查屠宰场,检查餐馆、医院、学校、惩戒设施和其他机构的厨房和食品储存区。还可以类似地监测水储备和加工厂。

[0004] 罕见变体检测对于病理性突变的早期检测也可能很重要。例如,在临床样品中检测癌症相关的点突变可以改善化疗期间极少残留疾病的鉴定,并检测复发患者的肿瘤细胞的出现。罕见点突变的检测对于评估环境诱变剂的暴露、监测内源性DNA修复以及研究衰老个体中体细胞突变的积累也很重要。此外,更灵敏的检测罕见变体的方法可以增强产前诊断,能够实现对母体血液中存在的胎儿细胞的表征。

### 发明内容

[0005] 鉴于前述内容,存在对检测罕见序列变体的改进方法的需求。本公开内容的组合物和方法满足了该需求,并且还提供了另外的益处。特别地,本公开内容的各个方面提供对罕见或低频核酸序列变体(有时称为突变)的高灵敏度检测。这包括在可能含有在正常序列背景下的少量变异序列的样品中鉴定和阐明低频核酸变异(包括置换、插入和缺失),以及在测序错误背景下鉴定低频变异。

[0006] 在一方面,本公开内容提供了进行滚环扩增的方法,该方法包括(a)使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中多种多核苷酸中的每种多核苷酸在连接前具有5'端和3'端;(b)将连接酶降解;(c)在将连接酶降解后扩增环状多核苷酸,以产生扩增的多核苷酸;其中不在步骤(a)与(c)之间纯化或分离多核苷酸。在

一些实施方案中,该方法进一步包括在步骤(a)与(c)之间将线性多核苷酸降解。在一些实施方案中,多种多核苷酸包含单链多核苷酸。在一些实施方案中,个体环状多核苷酸具有接点,该接点在环化的多核苷酸中各不相同。在一些实施方案中,环化包括将衔接子多核苷酸接合至多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。在一些实施方案中,扩增包括使环状多核苷酸经受包含随机引物的扩增反应混合物。在一些实施方案中,扩增包括使环状多核苷酸经受包含一种或多种引物的扩增反应混合物,每种引物通过序列互补性与不同的靶序列特异性杂交。在一些实施方案中,样品是来自受试者的样品。在一些实施方案中,样品是尿液、粪便、血液、唾液、组织或体液。在一些实施方案中,样品包含肿瘤细胞。在一些实施方案中,样品是福尔马林固定、石蜡包埋的(FFPE)样品。在一些实施方案中,该方法进一步包括基于判定步骤来诊断和任选地治疗所述受试者。在一些实施方案中,序列变体是因果遗传变体。在一些实施方案中,序列变体与癌症的类型或阶段相关。在一些实施方案中,多种多核苷酸包含无细胞多核苷酸。在一些实施方案中,无细胞多核苷酸包含循环肿瘤DNA。在一些实施方案中,无细胞多核苷酸包含循环肿瘤RNA。在一些实施方案中,该方法进一步包括对扩增的多核苷酸进行测序,以产生多个测序读取。在一些实施方案中,该方法进一步包括鉴定测序读取与参考序列之间的序列差异。在一些实施方案中,该方法进一步包括仅在以下情况下才将序列差异判定为多种多核苷酸中的序列变体:(i)在双链输入分子的两条链上都鉴定到序列差异;(ii)序列差异出现在通过滚环扩增形成的多联体的共有序列中;和/或(iii)序列差异出现在两种不同的分子中。在一些实施方案中,当序列差异出现在具有在5'端与3'端之间形成的不同接点的至少两种环状多核苷酸中时,序列差异被鉴定为出现在两种不同分子中。在一些实施方案中,当对应于两种不同分子的读取具有不同的5'端和不同的3'端时,序列差异被鉴定为出现在两种不同的分子中。

[0007] 在另一方面,本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,多种多核苷酸中的每种多核苷酸具有5'端和3'端,该方法包括:(a)环化多种多核苷酸中的个体多核苷酸以形成多种环状多核苷酸,每种环状多核苷酸在5'端与3'端之间具有接点;(b)扩增(a)的环状多核苷酸以产生扩增的多核苷酸;(c)将扩增的多核苷酸剪切以产生剪切的多核苷酸,每个剪切的多核苷酸包含在5'端和/或3'端处的一个或多个剪切点;(d)对剪切的多核苷酸进行测序,以产生多个测序读取;(e)鉴定测序读取与参考序列之间的序列差异;以及(f)当序列差异出现在至少两种不同的剪切多核苷酸中时,将序列差异判定为序列变体。在一些实施方案中,当(i)序列差异出现在具有不同接点的至少两种环状多核苷酸中;(ii)在双链输入分子的两条链上都鉴定到序列差异;和/或(iii)序列差异出现在由包括滚环扩增在内的扩增形成的多联体的共有序列中时,进一步出现将序列差异判定为序列变体。在一些实施方案中,多种多核苷酸包含单链多核苷酸。在一些实施方案中,通过使多种多核苷酸经受连接反应来实现环化。在一些实施方案中,序列变体是单核苷酸多态性。在一些实施方案中,参考序列是通过将测序读取彼此比对而形成的共有序列。在一些实施方案中,参考序列是测序读取。在一些实施方案中,环化包括将衔接子多核苷酸接合至多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。在一些实施方案中,通过使用具有链置换活性的聚合酶实现扩增。在一些实施方案中,扩增包括使环状多核苷酸经受包含随机引物的扩增反应混合物。在一些实施方案中,扩增包括使环状多核苷酸经受包含一种或多种引物的扩增反应混合物,每种引物通过序列互补性与不同的靶序列特异性

杂交。在一些实施方案中,在未富集的情况下使扩增的多核苷酸经受测序步骤。在一些实施方案中,该方法进一步包括通过在测序之前进行富集步骤来富集扩增的多核苷酸中的一种或多种靶多核苷酸。在一些实施方案中,基于判定步骤鉴定微生物污染物。在一些实施方案中,样品是来自受试者的样品。在一些实施方案中,样品是尿液、粪便、血液、唾液、组织或体液。在一些实施方案中,样品包含肿瘤细胞。在一些实施方案中,样品是福尔马林固定、石蜡包埋的 (FFPE) 样品。在一些实施方案中,该方法进一步包括基于判定步骤来诊断和任选地治疗所述受试者。在一些实施方案中,序列变体是因果遗传变体。在一些实施方案中,序列变体与癌症的类型或阶段相关。在一些实施方案中,多种多核苷酸包含无细胞多核苷酸。在一些实施方案中,无细胞多核苷酸包含循环肿瘤DNA。图14提供了示例工作流程的说明性示意图。

[0008] 在另一方面,本公开内容提供了反应混合物,用于实施根据本文的任何方法的方法,其中反应混合物包含 (a) 多种多联体,其中多种多联体中的个体多联体包含通过将具有 5' 端和 3' 端的个体多核苷酸环化而形成的不同接点; (b) 包含序列A' 的第一引物,其中第一引物通过序列A与序列A' 之间的序列互补性与靶序列的序列A特异性杂交; (c) 包含序列B的第二引物,其中第二引物通过B与B' 之间的序列互补性与包含靶序列的互补序列的互补多核苷酸中存在的序列B' 特异性杂交; 以及 (d) 使第一引物和第二引物延伸以产生扩增的多核苷酸的聚合酶; 其中靶序列中序列A的 5' 端与序列B的 3' 端之间的距离为 75nt 或更小。在一些实施方案中,第一引物包含在序列A' 的 5' 侧的序列C,第二引物包含在序列B的 5' 侧的序列D,并且在扩增反应中的第一扩增步骤期间,序列C和序列D均不与两个或更多种多联体杂交。

[0009] 在另一方面,本公开内容提供了一种用于检测序列变体的系统,其包含 (a) 计算机,该计算机被配置用于接收关于对样品进行检测反应的用户请求; (b) 扩增系统,该扩增系统响应于用户请求,对样品或其部分进行核酸扩增反应,其中扩增反应包括以下步骤: (i) 使用连接酶将多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,其中多种多核苷酸中的每种多核苷酸在连接前具有 5' 端和 3' 端; (ii) 将连接酶降解; 以及 (iii) 在将连接酶降解后扩增环状多核苷酸,以产生扩增的多核苷酸; 其中不在步骤 (i) 与 (iii) 之间纯化或分离多核苷酸; (c) 测序系统,该测序系统产生由扩增系统扩增的多核苷酸的测序读取,鉴定测序读取与参考序列之间的序列差异,并将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体; 以及 (d) 报告生成器,该报告生成器向接收者发送报告,其中该报告包含序列变体的检测结果。在一些实施方案中,接收者是用户。

[0010] 在另一方面,本公开内容提供了一种包含代码的计算机可读介质,该代码在由一个或多个处理器执行时实施检测序列变体的方法,所实施的方法包括: (a) 接收关于对样品进行检测反应的客户请求; (b) 响应于客户请求,对样品或其部分进行核酸扩增反应,其中扩增反应包括以下步骤: (i) 使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中多种多核苷酸中的每种多核苷酸在连接前具有 5' 端和 3' 端; (ii) 将连接酶降解; 以及 (iii) 在将连接酶降解后扩增环状多核苷酸,以产生扩增的多核苷酸; 其中不在步骤 (i) 与 (iii) 之间纯化或分离多核苷酸; (c) 进行包括以下步骤的测序分析: (i) 产生在扩增反应中扩增的多核苷酸的测序读取; (ii) 鉴定测序读取与参考序列之间的序列差异; 以及 (iii) 将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序

列变体;以及(d)生成包含序列变体的检测结果的报告。

[0011] 在另一方面,本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,多种多核苷酸中的每种多核苷酸具有5'端和3'端,该方法包括:(a)将所述多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,每种环状多核苷酸在5'端和3'端之间具有接点;(b)将连接酶降解;(c)使用随机引物扩增(a)的环状多核苷酸,以产生扩增的多核苷酸;(d)将扩增的多核苷酸剪切;(e)对剪切的多核苷酸进行测序,以产生多个测序读取;(f)鉴定测序读取与参考序列之间的序列差异;以及(g)在以下情况下将序列差异判定为序列变体:(i)在双链输入分子的两条链上都鉴定到序列差异;(ii)序列差异出现在通过滚环扩增形成的多联体的共有序列中;(iii)当序列差异出现在至少两种不同的剪切多核苷酸中时,将序列差异称为序列变体;和/或(iv)序列差异出现在两种不同的分子中;其中当序列差异出现在具有在5'端与3'端之间形成的不同接点的至少两种环状多核苷酸中时,将序列差异鉴定为出现在两种不同分子中。

[0012] 在另一方面,本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,多种多核苷酸中的每种多核苷酸具有5'端和3'端,该方法包括:(a)将所述多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,每种环状多核苷酸在5'端和3'端之间具有接点;(b)将连接酶降解;(c)使用一种或多种引物扩增(a)的环状多核苷酸以产生扩增的多核苷酸,每种引物通过序列互补性与不同的靶序列特异性杂交;(d)将扩增的多核苷酸剪切;(e)对剪切的多核苷酸进行测序,以产生多个测序读取;(f)鉴定测序读取与参考序列之间的序列差异;以及(g)在以下情况下将序列差异判定为序列变体:(i)在双链输入分子的两条链上都鉴定到序列差异;(ii)序列差异出现在通过滚环扩增形成的多联体的共有序列中;和/或(iii)序列差异出现在两种不同的分子中;其中当序列差异出现在具有在5'端与3'端之间形成的不同接点的至少两种环状多核苷酸中时,将序列差异鉴定为出现在两种不同分子中。

[0013] 在一方面,本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,多种多核苷酸中的每种多核苷酸具有5'端和3'端,该方法包括(a)将多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,其中多种环状多核苷酸中的给定环状多核苷酸具有由环化产生的接合序列;(b)扩增(a)的环状多核苷酸,以产生多种扩增的多核苷酸,其中该多种扩增多核苷酸中的第一扩增多核苷酸和该多种扩增多核苷酸中的第二扩增多核苷酸包含接合序列,但在其各自的5'端和/或3'端包含不同的序列;(c)对多种扩增多核苷酸或其扩增产物进行测序,以产生对应于第一扩增多核苷酸和第二扩增多核苷酸的多个测序读取;以及(d)当在对应于第一扩增多核苷酸和第二扩增多核苷酸的测序读取中出现序列差异时,将测序读取中所检测到的序列差异判定为序列变体。

[0014] 在一些实施方案中,(a)中的将个体多核苷酸环化通过连接酶实现。在一些实施方案中,在(b)之前,连接酶被降解。在一些实施方案中,未环化的多核苷酸在(b)之前被降解。在一些实施方案中,多种环状多核苷酸在(b)之前未纯化或分离。

[0015] 在一些实施方案中,(a)中的环化包括将衔接子多核苷酸接合至多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。

[0016] 在一些实施方案中,(b)中的将环状多核苷酸扩增通过具有链置换活性的聚合酶实现。在一些实施方案中,(b)中的将环状多核苷酸扩增包括滚环扩增(RCA)。在一些实施方

案中, (b) 中的扩增包括使环状多核苷酸经历包含随机引物的扩增反应混合物。在一些实施方案中, 个体随机引物包含在其各自的5' 和/或3' 端的彼此不同的序列。在一些实施方案中, (b) 中的扩增包括使环状多核苷酸经历包含靶标特异性引物的扩增反应混合物。在一些实施方案中, 扩增包括变性、引物结合和引物延伸的多个循环。

[0017] 在一些实施方案中, 在未富集的情况下使扩增的多核苷酸经历(c) 的测序。在一些实施方案中, 该方法进一步包括通过在(c) 的测序之前进行富集步骤, 在扩增的多核苷酸或其扩增产物中富集一种或多种靶多核苷酸。

[0018] 在一些实施方案中, 多种多核苷酸包含单链多核苷酸。在一些实施方案中, 序列变体是单核苷酸多态性。在一些实施方案中, 样品是来自受试者的样品。在一些实施方案中, 样品包括尿液、粪便、血液、唾液、组织或体液。在一些实施方案中, 样品包含肿瘤细胞。在一些实施方案中, 样品包含福尔马林固定、石蜡包埋的 (FFPE) 样品。在一些实施方案中, 多种多核苷酸包含无细胞多核苷酸。在一些实施方案中, 无细胞多核苷酸包含无细胞DNA。在一些实施方案中, 无细胞多核苷酸包含无细胞RNA。在一些实施方案中, 无细胞多核苷酸包含循环肿瘤DNA。在一些实施方案中, 无细胞多核苷酸包含循环肿瘤RNA。

[0019] 在一些实施方案中, 在(d) 中, 该方法包括当序列差异出现在来自第一扩增多核苷酸的至少50% 的测序读取和来自第二扩增多核苷酸的至少50% 的测序读取中时, 将测序读取中所检测到的序列差异判定为序列变体。

[0020] 在一方面, 本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法, 多种多核苷酸中的每种多核苷酸具有5' 端和3' 端, 所述方法包括 (a) 将多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸, 其中多种环状多核苷酸中的给定环状多核苷酸具有由环化产生的接合序列; (b) 扩增 (a) 的环状多核苷酸, 以产生多种扩增的多核苷酸; (c) 将扩增的多核苷酸剪切, 以产生剪切的多核苷酸, 每个剪切的多核苷酸在5' 端和/或3' 端包含一个或多个剪切点; (d) 对剪切的多核苷酸的扩增产物进行测序, 以产生多个测序读取; 以及 (e) 当序列差异出现在对应于第一剪切多核苷酸的测序读取和对应于第二剪切多核苷酸的测序读取中时, 将测序读取中所检测到的序列差异判定为序列变体。

[0021] 在一些实施方案中, (a) 中的将个体多核苷酸环化通过连接酶实现。在一些实施方案中, 在 (b) 之前, 连接酶被降解。在一些实施方案中, 未环化的多核苷酸在 (b) 之前被降解。在一些实施方案中, 多种环状多核苷酸在 (b) 之前未纯化或分离。

[0022] 在一些实施方案中, (a) 中的环化包括将衔接子多核苷酸接合至多种多核苷酸中的多核苷酸的5' 端、3' 端或5' 端和3' 端二者的步骤。

[0023] 在一些实施方案中, (b) 中的将环状多核苷酸扩增通过具有链置换活性的聚合酶实现。在一些实施方案中, (b) 中的将环状多核苷酸扩增包括滚环扩增 (RCA)。在一些实施方案中, (b) 中的扩增包括使环状多核苷酸经历包含随机引物的扩增反应混合物。在一些实施方案中, 个体随机引物包含在其各自的5' 和/或3' 端的彼此不同的序列。在一些实施方案中, (b) 中的扩增包括使环状多核苷酸经历包含靶标特异性引物的扩增反应混合物。

[0024] 在一些实施方案中, 在未富集的情况下使剪切的多核苷酸的扩增产物经历测序。在一些实施方案中, 该方法进一步包括通过在 (d) 的测序之前进行富集步骤, 在剪切的多核苷酸的扩增产物中富集一种或多种靶多核苷酸。在一些实施方案中, (c) 中的剪切包括使扩增的多核苷酸经历超声处理。在一些实施方案中, (c) 中的剪切包括使扩增的多核苷酸经历



酶促切割。

[0025] 在一些实施方案中,多种多核苷酸包含单链多核苷酸。在一些实施方案中,序列变体是单核苷酸多态性。在一些实施方案中,样品是来自受试者的样品。在一些实施方案中,样品包括尿液、粪便、血液、唾液、组织或体液。在一些实施方案中,样品包含肿瘤细胞。在一些实施方案中,样品包含福尔马林固定、石蜡包埋的 (FFPE) 样品。在一些实施方案中,多种多核苷酸包含无细胞多核苷酸。在一些实施方案中,无细胞多核苷酸包含无细胞DNA。在一些实施方案中,无细胞多核苷酸包含无细胞RNA。在一些实施方案中,无细胞多核苷酸包含循环肿瘤DNA。在一些实施方案中,无细胞多核苷酸包含循环肿瘤RNA。

[0026] 在一些实施方案中,在 (e) 中,该方法包括当序列差异出现在来自第一剪切多核苷酸的至少50%的测序读取和来自第二剪切的多核苷酸的至少50%的测序读取中时,将测序读取中所检测到的序列差异判定为序列变体。

#### 援引并入

[0027] 本说明书中所提及的所有出版物、专利和专利申请均通过引用并入本文,其程度如同特别地且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

[0028] 本发明的新颖特征在所附权利要求书中进行了具体阐述。通过参考以下对利用了本发明原理的说明性实施方案加以阐述的详细描述及其附图,将会获得对本发明的特征和优点的更好的了解,在附图中:

[0029] 图1描绘了根据本公开内容的方法的一个实施方案的示意图。将DNA链环化,并添加对应于所研究基因的靶标特异性引物,以及聚合酶、dNTP、缓冲液等,使得发生滚环扩增 (RCA),以形成模板DNA (例如“单体”) 的多联体 (例如“多聚体”)。对多联体进行处理以合成相应的互补链,然后添加衔接子以构建测序文库。随后会使用标准技术来测序的所得文库通常将含有三种物质:不含罕见序列变体 (例如突变) 的nDNA (“正常”DNA), 包含酶测序错误的nDNA,以及包含在扩增前预先存在于样品多核苷酸中的“真实”或实际序列变体的多聚体的DNA。有效罕见突变的多个拷贝的存在使检测和鉴定序列变体成为可能。

[0030] 图2描绘了与图1类似的策略,但添加了衔接子以促进多核苷酸环化。图2还显示了靶标特异性引物的使用。

[0031] 图3类似于图2,只是扩增中使用了衔接子引物。

[0032] 图4A-4C描绘了与环化单链 (ss) DNA的形成相关的三个实施方案。在顶部,单链DNA (ssDNA) 在没有衔接子的情况下环化,中间的方案描绘了衔接子的使用,而底部的方案使用两个衔接子寡核苷酸 (在每一端产生不同的序列) 并且可以进一步包括与两个衔接子杂交以使两端接近的夹板寡核苷酸。

[0033] 图5描绘了通过使用“分子钳”以使单链DNA的两端在空间上接近以供连接,而将特定靶标环化的实施方案。

[0034] 图6A和6B描绘了使用核酸的封闭端添加衔接子的两种方案。

[0035] 图7A-7C描绘了引发滚环扩增 (RCA) 反应的三种不同方式。图7A显示了靶标特异性引物例如感兴趣的特定靶基因或靶序列的使用。这通常导致仅扩增靶序列。图7B描绘了使用随机引物进行全基因组扩增 (WGA),这通常将所有样品序列扩增,然后在处理过程中对其

进行生物信息学分选。图7C描绘了在使用衔接子时衔接子引物的使用,这也导致总体上的非靶标特异性扩增。

[0036] 图8描绘了根据实施方案进行双链DNA环化和扩增,使得两条链都被扩增的实例。

[0037] 图9A-9D描绘了多种用于实现互补链合成以供后续测序的方案。图9A描绘了使用靶链的随机引发,随后进行连接。图9B描绘了使用靶链的衔接子引发,类似地随后进行连接。图9C描绘了“环”衔接子的使用,其中衔接子具有互补的两段序列,使得其彼此杂交以产生环(例如茎-环结构)。在连接到多联体末端时,环的游离端用作互补链的引物。图9D显示了使用超支化随机引物来实现第二链合成。

[0038] 图10显示了根据实施方案的PCR方法,该方法使用在靶序列的单体内比对时彼此呈远离定向的一对引物(也称为“背靠背”,例如在两个方向上定向但不处于待扩增的结构域的末端上),促进含有至少两个拷贝的靶核酸序列的环状多核苷酸或链的测序。在一些实施方案中,在形成多联体后,使用这些引物组来促进扩增子成为靶序列的更高的多聚体,例如,二聚体、三聚体等。任选地,该方法可以进一步包括大小选择,以去除比二聚体小的扩增子。

[0039] 图11A-11D描绘了一个实施方案,其中背靠背(B2B)引物与“递升(touch up)”PCR步骤一起使用,使得不大利于短产物(如单体)的扩增。在这种情况下,引物具有两个结构域;第一结构域与靶序列杂交(灰色或黑色箭头),第二结构域是“通用引物”结合结构域(弯曲矩形;有时也称为衔接子),其不与原始靶序列杂交。在一些实施方案中,第一轮PCR通过低温退火步骤完成(图11A),使得基因特异性序列结合。低温运行产生各种长度的PCR产物,包括短产物(图11B)。在少量轮次之后,升高退火温度,使得有利于整个引物(两个结构域)的杂交(图11C);如图所示,在模板的末端可见这些杂交,而内部结合则不太稳定。因此,相比于较低温度或仅单个结构域,较高温度和两个结构域较不利于产生较短的产物(图11D)。

[0040] 图12A和12B描绘了两种不同的测序文库构建方法。图12A示出了Illumina® Nextera样品制备系统的实例,通过该系统可将DNA在单一步骤中同时片段化并用测序衔接子标记。如图12B所示,通过超声处理使多联体片段化,然后向两端添加衔接子(例如通过使用KAPA Biosystems的试剂盒)并进行PCR扩增。可使用其他方法。

[0041] 图13A-C提供了背靠背(B2B)引物设计与传统PCR引物设计相比的示例优势的图示。传统的PCR引物设计(左)将引物(箭头,A和B)置于靶序列侧翼的区域,该区域可能是突变的热点(黑色星形),并且它们通常至少相距60个碱基对(bp),从而产生约100bp的典型足迹(footprint)。在该图示中,B2B引物设计(右)将引物置于靶序列的一侧。两个B2B引物面向相反的方向,可以有任意重叠(例如约或少于约12bp、10bp、5bp或更少)。根据B2B引物长度,本图示中的总足迹可介于28-50bp之间。无论是对于线性片段(图13A)、环化DNA(图13B)还是扩增产物(图13C),由于足迹较大,片段化事件更可能破坏传统设计中的引物结合,导致序列信息丢失。此外,如图13C所示,B2B引物设计捕获可用于区分不同多核苷酸的接合序列(也称为“天然条形码”)。

[0042] 图14示出了根据实施方案的产生用于检测序列变体的模板的方法(例如,使用环化多核苷酸的方法的示例实施方式)。将DNA输入变性为ssDNA,通过连接而环化,并且通过外切核酸酶消化来降解未环化的DNA。通过定量PCR(qPCR),比较输入DNA量与环化DNA量将连接效率定量,通常得到至少约80%的连接效率。将环化DNA纯化以交换缓冲液,然后用随

机引物和Phi29聚合酶进行全基因组扩增(WGA)。纯化WGA产物,并将产物片段化(例如通过超声处理)为约或小于约400bp的短片段。通过qPCR,比较相同量的参考基因组DNA与扩增DNA来定量扩增DNA的中靶率(on-target rate),通常显示平均中靶率为约或大于约95%。

[0043] 图15A-15C示出了利用有尾B2B引物扩增的进一步实施方式,以及在更高温度下实施PCR的“上升”第二阶段。B2B引物含有序列特异性区域(粗黑线)和衔接子序列(空心框)。在较低的第一阶段退火温度下,靶标特异性序列退火到模板以产生初始单体,并且PCR产物含有串联重复(图15A)。在较高温度下的第二扩增阶段,靶标特异性杂交和衔接子序列杂交二者一起优于单独的靶标特异性序列杂交,降低了优先产生短产物的程度(图15B)。在不利于整个引物的情况下,采用靶标特异性序列的内部退火迅速增加单体的比例(图15C,左)。

[0044] 图16示出了通过靶标测序方法检测的背景噪声(变体的频率)之间的比较,该靶标测序方法使用Q30过滤器,要求(底部线)及不要求(顶部线)在将被计数为变体的两个不同多核苷酸(例如,通过不同的接点来鉴定)上出现序列差异。将人基因组DNA(12878, Coriell Institute)片段化至100-200bp,并且其包含含有已知SNP(CYP2C19)的基因组DNA(19240, Coriell Institute)的2%掺入(spike-in)。真正的变体信号(标出的峰值)未明显高于背景(上方,浅灰色图)。通过应用验证过滤器(下方,黑色图)将背景噪声降低至约0.1。

[0045] 图17示出了应用本公开内容的方法对在多核苷酸群体(2%、0.2%和0.02%)中以不同的较低频率掺入的序列变体的检测,该变体仍显著高于背景值。

[0046] 图18A和18B示出了本公开内容的实施方案的连接效率和中靶率的分析结果。

[0047] 图19示出了根据本公开内容的实施方案的方法中等位基因频率的保留和偏差的基本消失。

[0048] 图20示出了根据实施方案检测小输入样品中序列变体的结果。

[0049] 图21示出了根据标准测序方法,在未要求在两种不同多核苷酸上出现序列差异的情况下获得的序列变体检测结果中的高背景的实例。

[0050] 图22提供了示出基因组的GC含量分布与根据本公开内容的实施方案的方法产生的测序结果(本公开内容的方法;左)、使用替代测序文库构建试剂盒的测序结果(Rubicon, Rubicon Genomics;中)和如文献报道的32ng无细胞DNA(cfDNA)(右)的GC含量分布之间的比较的图。

[0051] 图23提供了示出从根据实施方案的方法的测序读取获得的输入DNA的大小分布的图。

[0052] 图24提供了示出通过根据实施方案的随机引发方法对多个靶标的均匀扩增的图。

[0053] 图25A和25B示出了在没有环化的情况下形成具有可鉴定接点的多核苷酸多聚体的实施方案。多核苷酸(如多核苷酸片段或无细胞DNA)接合形成具有非天然接点的多聚体,该接点可用于根据本公开内容的实施方案区分独立的多核苷酸(在本文中也称为“自动标签”)。在图25A所示,多核苷酸通过平端连接彼此直接接合。在图25B中,多核苷酸通过一个或多个间插的衔接子寡核苷酸接合,该衔接子寡核苷酸可以进一步包含条形码序列。然后通过多种方法中的任一种,如通过随机引物(全基因组扩增)、衔接子引物或一种或多种靶标特异性引物或引物对,使多聚体经历扩增。

[0054] 图26示出了图25的过程的示例变化。多核苷酸(例如cfDNA或其他多核苷酸片段)是末端修复、A尾和衔接子连接的(例如使用标准试剂盒,如KAPA Biosystems的试剂盒)。可

以补充用内部尿嘧啶 (U) 标记的载体DNA, 以将总DNA输入提高至所需水平 (例如至约或大于约20ng)。待检测的序列变体由“星形”表示。当连接完成时, 可以通过添加尿嘧啶特异性切除试剂 (USER) 酶将载体DNA降解, 该酶是尿嘧啶DNA糖基化酶 (UDG) 和DNA糖基化酶-裂解酶内切核酸酶VIII的混合物。将产物纯化以消除载体DNA的片段。将纯化的产物扩增 (例如通过PCR, 使用针对衔接子序列的引物)。由于降解和在至少一端与衔接子分离, 任何残留的载体DNA都不大可能被扩增。可以将扩增产物纯化, 以除去短DNA片段。

[0055] 图27A-27E示出了图25的过程的示例变化。靶标特异性扩增引物包含用作衔接子的常见5'“尾”(灰色箭头)。初始扩增 (例如通过PCR) 进行几个循环 (例如至少约5、10或更多个循环)。PCR产物也可以用作引物, 从而退火到其他PCR产物 (例如, 当退火温度在第二阶段降低时), 以产生具有可鉴定接点的多联体。第二阶段可包括多个循环 (例如5、10、15、20或更多循环), 并且可包括有利于多联体形成和扩增的条件选择或变化。根据该示意图的方法也称为“中继扩增测序 (Relay Amp Seq)”, 其可以特别用于区室化设置 (例如小液滴中)。

[0056] 图28A-28E示出了用于多核苷酸环化的方法的非限制性实例。在图28A中, 双链多核苷酸 (例如dsDNA) 变性为单链, 然后直接环化 (例如通过CircLigase自接合连接)。在图28B中, 多核苷酸 (例如DNA片段) 被末端修复和A尾化 (将腺苷的单基延伸添加至3'端) 以提高连接效率, 然后变性为单链并环化。在图28C中, 多核苷酸被末端修复和A尾化 (如果是双链), 接合到具有胸苷 (T) 延伸的衔接子, 变性为单链并环化。在图28D中, 多核苷酸被末端修复和A尾化 (如果是双链), 两端连接到具有三个元件 (用于连接的T延伸、衔接子之间的互补性和3'尾) 的衔接子, 使链变性, 并将单链多核苷酸环化 (通过衔接子序列之间的互补性促进)。在图28E中, 将双链多核苷酸变性为单链形式, 并在分子钳的存在下环化, 该分子钳使多核苷酸的末端彼此更加接近, 以促进接合。

[0057] 图29示出了用于根据本公开内容的方法鉴定序列变体的扩增系统的示例工作流程设计示例, 其特别有关于环化多核苷酸。

[0058] 图30示出了用于根据本公开内容的方法鉴定序列变体的扩增系统的示例工作流程设计, 其特别有关于无环化步骤的线性多核苷酸输入。

[0059] 图31提供了用于根据本公开内容的方法鉴定序列变体的示例工作流程的概要图示。沿“线性多核苷酸分析”(上)分支, 分析可包括数字PCR (例如数字小液滴PCR、ddPCR)、实时PCR、通过探针捕获的富集 (捕获测序) 以及接合序列 (自动标记) 的分析、基于插入衔接子序列 (条形码插入) 的测序或中继扩增测序。沿“环化多核苷酸分析”(下)分支, 分析可包括数字PCR (例如数字小液滴PCR、ddPCR)、实时PCR、通过探针捕获的富集 (捕获测序) 以及接合序列 (自然条形码) 的分析、通过探针捕获或靶向扩增 (例如B2B扩增) 的富集, 以及带有验证步骤的序列分析, 该验证步骤是将序列变体鉴定为出现在两种不同多核苷酸 (例如, 具有不同接点的多核苷酸) 中的差异。

[0060] 图32是根据实施方案的系统的图示。

[0061] 图33示出了根据实例的捕获效率和沿靶区域的覆盖。>90%的靶向碱基被覆盖超过20倍, >50%的靶向碱基具有>50倍的覆盖。

[0062] 图34A和34B示出了示例性单反应测定工作流程。

[0063] 图35示出了使用接点信息的序列变体判定。

[0064] 图36A-36H示出了本发明的各实施方案的步骤。

[0065] 图37A和37B示出了靶多核苷酸包括单链多核苷酸的实施方案。

[0066] 图38示出了来自本文描述的各工作流程的文库复杂性。

[0067] 图39示出了示例示意图,其中多联体断裂点(shredding point)和接合序列被用于唯一地索引从相同来源扩增的读取。

[0068] 图40示出了示例示意图,其中在扩增期间由随机引发产生的多联体5'和3'序列和接合序列被用于唯一地索引从相同来源扩增的读取。

[0069] 图41A-41C示出了示例示意图,其中扩增多核苷酸的接合序列和5'/3'端被用于生成读取家族。在图41A中,因为组中的所有读取都显示具有多联体确认的变体('x'),所以该组被计为“变体”。在图41B中,变体被拒绝,并且读取家族的共有序列被分类为野生型,因为该家族中的大多数读取未显示出变体('x')。在图41C中,当在至少两个不同的读取家族中检测到相同的序列差异('x')时,判定出变体。未在至少两个不同的读取家族中检测到的序列差异(圆圈)不被判定为变体。

### 具体实施方式

[0070] 除非另有说明,否则本文公开的一些实施方案的实施采用了本领域技术范围内的免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组DNA的常规技术。参见,例如,Sambrook和Green,Molecular Cloning:A Laboratory Manual,第四版(2012);Current Protocols in Molecular Biology系列(F.M.Ausubel等编著);Methods In Enzymology系列(Academic Press,Inc.),PCR 2:A Practical Approach(M.J.MacPherson,B.D.Hames和G.R.Taylor编著(1995));Harlow和Lane编著(1988)Antibodies,A Laboratory Manual,and Culture of Animal Cells:A Manual of Basic Technique and Specialized Applications,第6版(R.I.Freshney编著(2010))。

[0071] 术语“约”或“大约”意指如本领域普通技术人员所确定的特定值在可接受的误差范围内,其部分地取决于该值是如何测量或确定的,即测量系统的限制。例如,“约”可以是指根据本领域的实践,在1个或大于1个标准差内。或者,“约”可以是指给定值的高达20%、高达10%、高达5%或高达1%的范围。或者,尤其是对于生物系统或过程,该术语可以是指在数值的一个数量级内,优选地在5倍以内,更优选地在2倍以内。在本申请和权利要求书中描述特定值时,除另有说明外,否则术语“约”应该被认为是指特定值的可接受的误差范围内。

[0072] 术语“多核苷酸”、“核苷酸”、“核苷酸序列”、“核酸”和“寡核苷酸”是可以互换使用的。它们是指任意长度的核苷酸(脱氧核糖核苷酸或核糖核苷酸)或其类似物的聚合形式。多核苷酸可以具有任意三维结构,并且可以行使任何已知的或未知的功能。以下是多核苷酸的非限制性实例:基因或基因片段的编码或非编码区,通过连锁分析确定的基因座(座位)、外显子、内含子、信使RNA(mRNA)、转运RNA(tRNA)、核糖体RNA(rRNA)、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微小RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任意序列的分离的DNA、任意序列的分离的RNA、核酸探针和引物。多核苷酸可以包含一个或多个修饰的核苷酸,例如甲基化核苷酸和核苷酸类似物。如果存在,对核苷酸结构的修饰可以在聚合物组装之前或之后赋予。核苷酸序列可以被非核苷酸组分打断。多核苷酸可以在聚合后被进一步修饰,例如,通过与标记组分偶联。

[0073] 通常,术语“靶多核苷酸”是指具有靶序列的核酸分子起始群体中的核酸分子或多核苷酸,该靶序列的存在、量和/或核苷酸序列或其中一个或多个的改变希望加以确定。通常,术语“靶序列”是指在核酸单链上的核酸序列。靶序列可以是基因的一部分,调节序列,基因组DNA,cDNA,包括mRNA、miRNA、rRNA在内的RNA,等等。靶序列可以是来自样品的靶序列或次级靶标如扩增反应的产物。

[0074] “杂交”是指这样的反应,在该反应中,一个或多个多核苷酸发生反应以形成复合体,该复合体通过在核苷酸残基的碱基之间的氢键键合而得到稳定。该氢键键合可以通过Watson Crick碱基配对、Hoogsteen结合或根据碱基互补性以任意其他序列特异性方式而发生。该复合体可以包含形成双链体结构的两条链、形成多链复合体的三条或更多条链、自杂交的单链或其任意组合。杂交反应可以构成更广泛的过程中的步骤,例如PCR的起始,或内切核酸酶对多核苷酸的酶切。与第一序列互补的第二序列称为第一序列的“互补序列”。应用于多核苷酸的术语“可杂交的”是指多核苷酸形成复合体的能力,该复合体通过杂交反应中核苷酸残基的碱基之间的氢键而得到稳定。

[0075] “互补性”是指核酸通过传统的Watson-Crick或其他非传统类型的作用与另一核酸序列形成氢键的能力。互补性百分比表示核酸分子中可以与第二核酸序列形成氢键(例如,Watson-Crick碱基配对)的残基百分比(例如,10个中的5、6、7、8、9、10个分别为50%、60%、70%、80%、90%和100%互补)。“完全互补”是指核酸序列的所有连续残基都将与第二核酸序列中相同数目的连续残基进行氢键结合。如本文所用的,“基本上互补”是指在具有8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50个或更多个核苷酸的区域上至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的互补程度,或指在严格条件下杂交的两个核酸。序列同一性,例如为了评估互补性百分比,可以通过任何合适的比对算法进行测量,包括但不限于Needleman-Wunsch算法(参见,例如,EMBOSS Needle比对器,可从[www.ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)获得,任选地具有默认设置)、BLAST算法(参见,例如,BLAST比对工具,可从[blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)获得,任选地具有默认设置)或者Smith-Waterman算法(参见,例如,EMBOSS Water比对器,可从[www.ebi.ac.uk/Tools/psa/emboss\\_water/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html)获得,任选地具有默认设置)。最优算法可以使用所选定算法的任意合适的参数(包括默认参数)进行评估。

[0076] 通常,用于杂交的“严格条件”是指这样的条件,在该条件下与靶序列互补的核酸主要与靶序列杂交而基本上不与非靶序列杂交。严格条件通常是序列依赖性的,并且随许多因素变化。通常,序列越长,序列与其靶序列特异性杂交的温度越高。在Tijssen(1993), *Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes* (第一部分,第二章) “Overview of principles of hybridization and the strategy of nucleic acid probe assay”, Elsevier, N.Y. 中详细描述了严格条件的非限制性实例。

[0077] 在一方面,本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,多种多核苷酸中的每种多核苷酸具有5'端和3'端。在一些情况下,该方法包括(a)将多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸,其中多种多核苷酸中的给定环状多核苷酸具有由所述环化产生的接合序列;(b)扩增(a)的环化多核苷酸以产生多种

扩增的多核苷酸；(c) 将扩增的多核苷酸剪切以产生剪切多核苷酸，每个剪切多核苷酸在5'端和/或3'端包含一个或多个剪切点；(d) 对剪切多核苷酸和/或剪切多核苷酸的扩增产物进行测序，以产生多个测序读取；以及(d) 当在对应于第一剪切多核苷酸和第二剪切多核苷酸的测序读取中出现序列差异时，将测序读取中所检测到的序列差异判定为序列变体。

[0078] 在一些情况下，该方法包括(a) 将所述多种多核苷酸中的个体多核苷酸环化，以形成多种环状多核苷酸，每种环状多核苷酸在5'端和3'端之间具有接点；(b) 扩增(a)的环状多核苷酸，以产生扩增的多核苷酸；(c) 将扩增的多核苷酸剪切以产生剪切多核苷酸，每个剪切多核苷酸在5'端和/或3'端包含一个或多个剪切点；(d) 对剪切多核苷酸进行测序，以产生多个测序读取；(e) 鉴定测序读取与参考序列之间的测序差异；以及(f) 当序列差异出现在至少两种不同的剪切多核苷酸中时，将序列差异判定为序列变体。

[0079] 通常，将多核苷酸的末端彼此接合以形成环状多核苷酸(直接接合，或通过一个或多个中间衔接子寡核苷酸接合)产生具有接合序列的接点。当多核苷酸的5'端和3'端通过衔接子多核苷酸接合时，术语“接点”可以指多核苷酸与衔接子之间的接点(例如，5'端接点或3'端接点之一)，或指多核苷酸的5'端和3'端之间由此形成且包含衔接子多核苷酸的接点。当多核苷酸的5'端和3'端在没有间插衔接子的情况下接合时(例如单链DNA的5'端和3'端)，术语“接点”指这两个末端接合的点。接点可以通过构成该接点的核苷酸序列(也称为“接合序列”)来鉴定。

[0080] 在一些实施方案中，样品包含具有通过以下过程形成的末端混合物的多核苷酸：自然降解过程(如细胞裂解、细胞死亡和多核苷酸如DNA和RNA从细胞释放到其周围环境的其他过程，多核苷酸在该周围环境中可进一步降解，如无细胞多核苷酸，如无细胞DNA和无细胞RNA)、作为样品处理的副产物的片段化(诸如固定、染色和/或储存过程)，以及通过切割不限于特定靶序列的DNA的方法进行的片段化(例如，机械片段化，如通过超声处理；非序列特异性核酸酶处理，如DNA酶I、片段化酶)。当样品包含具有末端混合物的多核苷酸时，两个多核苷酸具有相同5'端或3'端的可能性较低，并且两个多核苷酸独立地具有相同5'端和3'端的可能性更低。因此，在一些实施方案中，即使两个多核苷酸包含具有相同靶序列的部分，也可以使用接点来区分不同的多核苷酸。当多核苷酸末端在没有间插衔接子的情况下接合时，可以通过与参考序列比对来鉴定接合序列。例如，在两个组成序列的顺序相对于参考序列似乎颠倒的情况下，似乎发生颠倒的点可以表示在该点处的接点。当多核苷酸末端通过一个或多个衔接子序列接合时，可以通过与已知衔接子序列的接近度鉴定接点，或者如果测序读取的长度足以从环化多核苷酸的5'和3'端获得序列，则通过如上所述的比对来鉴定接点。在一些实施方案中，特定接点的形成是非常罕见的事件，使得其在样品的环化多核苷酸中是独特的。

[0081] 在一些实施方案中，(a) 中的将个体多核苷酸环化通过使多种多核苷酸经历连接反应来实现。连接反应可包含连接酶。在一些实施方案中，连接酶在(b) 中的扩增之前降解。在(b) 中的扩增之前连接酶的降解可以提高可扩增多核苷酸的回收率。在一些实施方案中，多种环化多核苷酸在(b) 之前未纯化或分离。在一些实施方案中，未环化的线性多核苷酸在扩增之前降解。

[0082] 在一些情况下，(a) 中的环化包括将衔接子多核苷酸接合到多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。如前所述，当多核苷酸的5'端和/或3'端通

过衔接子多核苷酸接合时,术语“接点”可以指多核苷酸与衔接子之间的接点(例如,5'端接点或3'端接点之一),或指多核苷酸的5'端和3'端之间由此形成且包含衔接子多核苷酸的接点。

[0083] 可以例如在连接酶降解后扩增环化的多核苷酸,以产生扩增的多核苷酸。(b)中的将环状多核苷酸扩增可以通过具有链置换活性的聚合酶实现。在一些情况下,聚合酶是Phi29DNA聚合酶。在一些情况下,扩增包括滚环扩增(RCA)。由RCA产生的扩增多核苷酸可包括线性多联体,或包含来自模板多核苷酸的两个或更多个拷贝的靶序列(例如,亚基序列)的多核苷酸。在一些实施方案中,扩增包括使环状多核苷酸经历包含随机引物的扩增反应混合物。在一些情况下,扩增包括使环状多核苷酸经历包含一种或多种引物的扩增反应混合物,每种引物通过序列互补性与不同的靶序列特异性杂交。

[0084] 在一些情况下,将扩增的多核苷酸剪切,以产生长度相对于未剪切多核苷酸更短的剪切多核苷酸。源自相同线性多联体的两个或更多个剪切的多核苷酸可具有相同的接合序列,但可具有不同的5'和/或3'端(例如剪切末端)。

[0085] 可以使用多种方法剪切扩增的多核苷酸,例如,但不限于,物理片段化、酶促方法和化学片段化。可用于将扩增多核苷酸片段化的物理片段化方法的非限制性实例包括声学剪切、超声处理和流体动力学剪切。在一些情况下,声学剪切和超声处理可能是优选的。可用于将扩增的多核苷酸片段化的酶促片段化方法的非限制性实例包括使用酶如DNA酶I和其他限制性内切核酸酶,包括非特异性核酸酶和转座酶。可用于将扩增的多核苷酸片段化的化学片段化方法的非限制性实例包括使用热和二价金属阳离子。

[0086] 与未剪切的多核苷酸相比长度更短的剪切多核苷酸(也称为片段化多核苷酸)可能是期望的,以与用于产生测序读取(也称为序列读取)的测序仪器的功能相匹配。例如,可以将扩增的多核苷酸片段化(例如剪切)至下游测序平台所确定的最佳长度。本文进一步描述的各种测序仪器可以适应不同长度的核酸。在一些情况下,扩增的多核苷酸在附接可用于下游测序平台的衔接子的过程中,例如在流动池附着或测序引物结合中被剪切。在一些情况下,剪切的多核苷酸在测序之前经历扩增,以产生剪切多核苷酸的扩增产物。可能需要额外的扩增,例如,以产生足量的多核苷酸以供下游分析,例如测序分析。所得的扩增产物可包含多个拷贝的个体剪切多核苷酸。

[0087] 在测序期间,可以对来自相同扩增多核苷酸的剪切多核苷酸或其扩增产物进行测序。可以将由测序产生的测序读取分组为读取家族。读取家族可包含任何合适数目的序列读取。在一些情况下,读取家族包含至少5、10、15、20、25、50、75或100个序列读取。在一些情况下,除非存在最小数目的序列读取,否则可能不会将一组序列读取鉴定为读取家族。例如,读取家族可包含至少2、3、4、5、7、8、9或10个序列读取。在一些情况下,读取家族包含至少25个读取序列。在一些情况下,可以基于共享的接合序列和5'和3'端的共享序列将序列读取划分为读取家族。在一些实施方案中,读取家族的序列读取具有相同的接合序列。在一些实施方案中,读取家族的序列读取在5'和3'端具有相同的序列,例如,在5'和3'端各有至少5个碱基、6个碱基、7个碱基、8个碱基、9个碱基或10个碱基上的序列可以是相同的。在一些情况下,由于扩增和/或测序错误导致的错误,5'和3'端的序列在读取家族的序列读取中不尽相同。当例如通过比对来比较时,读取家族的测序读取可能表现出重叠。在一些情况下,当最佳比对时,读取家族的测序读取表现出至少75%的同一性。术语“同一性百分比



(%)”是指在对序列进行比对和引入空位(如有需要)以实现最大百分比同一性之后,在两个序列(例如,候选序列和参考序列)之间共享的相同残基的百分比(即可以在候选序列和参考序列中的一者或两者中引入空位,以获得最佳比对,并且在一些情况下,可以为了比较而忽略非同源序列)。为了确定同一性百分比,可以以各种方式实现比对,例如,使用公共可用的计算机软件,如BLAST、ALIGN或Megalign (DNASTAR) 软件。可以通过使用BLAST对测试序列与比较序列进行比对、确定经比对的测试序列中的氨基酸或核苷酸与比较序列的相同位置中的氨基酸或核苷酸相同的数目,并将相同氨基酸或核苷酸的数目除以比较序列中氨基酸或核苷酸的数目,来计算两个序列的同一性百分比。当最佳比对时,家族的两个测序读取可以在任何合适长度的碱基上表现出至少75%的同一性(例如,至少80%、85%、90%或95%的同一性)。读取家族中的第一对测序读取可以表现出与第二对测序读取不同的同一性百分比。在一些情况下,为在至少50个碱基(例如,至少60个碱基、70个碱基、80个碱基、90个碱基、100个碱基、110个碱基、120个碱基、130个碱基、140个碱基或150碱基)的长度上的比对确定同一性百分比。在一些情况下,在约25-250个碱基、约50-200个碱基、约75-175个碱基或约100-150个碱基的长度上比对。在一些情况下,在测试序列或比较序列的整个长度上比对。在一些实施方案中,当最佳比对时,读取家族的两个测序读取在至少50个碱基(例如,至少60个碱基、70个碱基、80个碱基、90个碱基、100个碱基、110个碱基、120个碱基、130个碱基、140个碱基或150个碱基)的长度上表现出至少75%的同一性(例如,至少80%、85%、90%或95%的同一性)。

[0088] 包含环状多核苷酸模板的线性多联体的扩增多核苷酸可包含环状多核苷酸模板序列的多个重复或拷贝。由扩增的多核苷酸产生的剪切多核苷酸可具有环状多核苷酸模板序列的多个拷贝。剪切的多核苷酸可具有重复序列的少于一个拷贝,重复序列的至少一个拷贝,重复序列的至少两个拷贝,或重复序列的至少三个拷贝。剪切的多核苷酸中的重复数目可取决于重复序列的长度。例如,对于大小大致相同的剪切片段,与具有长度更长的重复的多联体相比,具有长度相对较短的重复的多联体可以产生具有更多拷贝的重复序列的剪切片段。

[0089] 在一些情况下,剪切多核苷酸或其扩增产物的测序读取可包含重复序列的至少一个拷贝。在一些情况下,测序读取包含重复序列的至少两个拷贝(例如,至少三个拷贝、四个拷贝或五个拷贝)。来自读取家族的序列读取的重复序列的平均拷贝数可取决于核酸样品的多核苷酸的长度。

[0090] 可以通过首先鉴定多联体中重复区段的长度和/或序列,将测序读取分组为读取家族,该多联体对应于环状多核苷酸模板的序列。在一些情况下,鉴定重复区段的长度和/或序列包括将读取与其他读取比对或与参考序列比对。随后,可以例如通过与参考序列比对来鉴定接合序列。可以确定多核苷酸的5' 和3' 端的序列以及它们与接点的相对距离(例如,以碱基计)。在5' 和3' 端具有相同接合序列和共享序列的读取可以被分组为读取家族,表示源自相同剪切多核苷酸的扩增产物的测序读取。

[0091] 在一些情况下,通过确认序列差异出现在具有相同接合序列但在各自的5' 和3' 端具有不同序列的第二读取家族(例如,至少两种剪切多核苷酸)中,可以将读取家族中所观察到的序列差异判定为真实序列差异而不是扩增和/或测序错误的结果。具有相同接合序列但具有不同的5' 和/或3' 端的两个读取家族可以对应于相同线性多联体的两种剪切多核

核苷酸。观察与相同扩增多核苷酸的两种剪切多核苷酸相对应的两个读取家族中的序列差异可以是一种确认序列差异真实存在于其他环状多核苷酸上,而不是一种剪切多核苷酸的扩增和/或测序错误的结果的方法。

[0092] 在一些情况下,如果序列差异出现在读取家族的大多数测序读取中,则认为在读取家族的序列读取中观察到的序列差异是序列差异。在一些情况下,如果序列差异出现在读取家族的至少50%的测序读取(例如,至少60%、70%、80%、90%或95%的测序读取)中,则认为读取家族的序列读取中观察到的序列差异是序列差异。在一些情况下,如果序列差异出现在读取家族的100%的测序读取中,则认为读取家族的序列读取中观察到的序列差异是序列差异。在一些情况下,当序列差异出现在来自第一剪切多核苷酸的大多数测序读取和来自第二剪切多核苷酸的大多数测序读取中时,将测序读取中检测到的序列差异判定为序列变体。在一些情况下,当序列差异出现在来自第一剪切多核苷酸的至少50%的测序读取(例如,至少60%、70%、80%、90%或95%的测序读取)和来自第二剪切多核苷酸的至少50%的测序读取(例如,至少60%、70%、80%、90%或95%的测序读取)中时,将测序读取中检测到的序列差异判定为序列变体。在一些情况下,当序列差异出现在来自第一剪切多核苷酸的100%的测序读取和来自第二剪切多核苷酸的100%的测序读取中时,将测序读取中检测到的序列差异判定为序列变体。

[0093] 通过使用两种不同的剪切多核苷酸,即具有相同接合序列但剪切末端不同的两种剪切多核苷酸,来确认从样品中的测序读取鉴定的序列差异的存在,可以改善序列变体检测。预期真正的序列变体会见于源自相同扩增多核苷酸的至少两种剪切多核苷酸中,而预期错误会见于少于两种剪切多核苷酸中。在一些情况下,变体检测的错误率降低。在一些实施方案中,变体检测的错误率降低至少5%、10%、15%、20%、25%、30%、35%、40%、45%或50%。在一些情况下,变体检测的灵敏度和/或特异性增加。在一些实施方案中,变体检测的灵敏度增加至少5%、10%、15%、20%、25%、30%、35%、40%、45%或50%。在一些实施方案中,变体检测的特异性增加至少5%、10%、15%、20%、25%、30%、35%、40%、45%或50%。在一些情况下,假阳性率降低。

[0094] 在一些情况下,当(i)序列差异出现在具有不同接点的至少两种环状多核苷酸中;(ii)在双链输入分子的两条链上都鉴定到序列差异;和/或(iii)序列差异出现在由包括滚环扩增(RCA)在内的扩增形成的多联体的共有序列中时,进一步出现将序列差异判定为序列变体。在一些情况下,参考序列是测序读取。在一些情况下,参考序列是通过将测序读取彼此比对而形成的共有序列。

[0095] 在一些情况下,在未富集的情况下使剪切的多核苷酸经历测序。然而,如果需要,可以在测序之前在富集步骤中进行扩增多核苷酸和/或剪切多核苷酸中的一种或多种靶多核苷酸的富集。示例性富集步骤可包括使用具有与靶序列互补的序列的核酸。

[0096] 如本文进一步描述的,序列变体可以是关于参考序列的任何变异。可使用本文方法检测的序列变体的非限制性实例包括单核苷酸多态性(SNP)、缺失/插入多态性(DIP)、拷贝数变体(CNV)、短串联重复(STR)、简单序列重复(SSR)、可变数目串联重复(VNTR)、扩增片段长度多态性(AFLP)、基于反转录转座子的插入多态性、序列特异性扩增多态性,以及可检测为序列变体的表观遗传标记的差异(例如甲基化差异)。在一些情况下,序列变体是多态性,例如单核苷酸多态性。在一些情况下,序列变体是因果遗传变体。在一些情况下,序列变

体与癌症的类型或阶段相关。

[0097] 核酸样品可以是来自受试者的样品。在一些情况下,样品来自人受试者。在一些情况下,样品包括来自受试者(例如人受试者)的尿液、粪便、血液、唾液、组织或体液。在一些情况下,样品包含肿瘤细胞。在一些情况下,样品包括福尔马林固定、石蜡包埋的样品。在一些情况下,样品的多种多核苷酸包含无细胞多核苷酸。无细胞多核苷酸可包含无细胞DNA,并且在一些情况下,可包含循环肿瘤DNA和/或循环肿瘤RNA。无细胞多核苷酸可包含无细胞RNA。在一些实施方案中,该方法进一步包括基于序列变体的判定来诊断和任选地治疗受试者。在一些情况下,样品中的微生物污染物基于序列变体的判定来鉴定。在这种情况下,样品可以来自受试者,但也可以来自非受试者样品,例如土壤样品或食物样品。

[0098] 多种多核苷酸可以是单链的。在一些情况下,多核苷酸是双链形式,并通过例如变性进行处理以在进行环化之前产生单链。在一些情况下,将双链多核苷酸环化以产生双链环,并且双链环通过例如变性进行处理以产生单链环。

[0099] 在另一方面,本公开内容提供了鉴定包含多种多核苷酸的核酸样品中的序列变体的方法,多种多核苷酸中的每种多核苷酸具有5'端和3'端。在一些实施方案中,该方法包括:(a)将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中给定的环状多核苷酸具有由所述环化产生的接合序列;(b)扩增(a)的环状多核苷酸以产生多种扩增的多核苷酸,其中多种扩增多核苷酸中的第一扩增多核苷酸和多种扩增多核苷酸中的第二扩增多核苷酸包含接合序列,但在其各自的5'端和/或3'端包含不同的序列;(c)对多种扩增多核苷酸和/或其扩增产物进行测序,以产生对应于第一扩增多核苷酸和第二扩增多核苷酸的多个测序读取;以及(d)当序列差异出现在对应于第一扩增多核苷酸和第二扩增多核苷酸的测序读取中时,将测序读取中所检测到的序列差异判定为序列变体。在一些实施方案中,(a)中的将个体多核苷酸环化通过连接酶实现。在一些实施方案中,连接酶在(b)中的扩增之前降解。在(b)中的扩增之前连接酶的降解可以提高可扩增多核苷酸的回收率。在一些实施方案中,多种环化多核苷酸在(b)之前未纯化或分离。

[0100] 在一些情况下,(a)中的环化包括将衔接子多核苷酸接合到多种多核苷酸中的多核苷酸的5'端、3'端或5'端和3'端二者的步骤。如前所述,当多核苷酸的5'端和/或3'端通过衔接子多核苷酸接合时,术语“接点”可以指多核苷酸与衔接子之间的接点(例如,5'端接点或3'端接点之一),或指多核苷酸的5'端和3'端之间由此形成且包含衔接子多核苷酸的接点。

[0101] 在环化后,将环状多核苷酸扩增。(b)中的将环状多核苷酸扩增可以通过具有链置换活性的聚合酶实现。在一些情况下,聚合酶是Phi29DNA聚合酶。在一些情况下,扩增(b)中的环状多核苷酸包括滚环扩增(RCA)。滚环扩增可以产生包含模板环状多核苷酸序列的线性多联体的扩增多核苷酸。在一些情况下,(b)中的扩增包括使环状多核苷酸经历使用随机引物的扩增反应混合物。随机引物在(b)的扩增过程中可以与环状多核苷酸非特异性地(例如,随机地)杂交。可以与环状多核苷酸非特异性杂交的随机引物可以与常见的环状多核苷酸、多种环状多核苷酸或两者杂交。在一些情况下,两个或更多个随机引物与同一环状多核苷酸(例如,同一环状多核苷酸的不同区域)杂交,并产生具有相同靶序列(或亚基序列)的重复的扩增多核苷酸。相同模板(例如,环状多核苷酸)的扩增多核苷酸可具有相同的接合序列。在一些实施方案中,个体随机引物包含在其各自的5'和/或3'端的彼此不同的序列,

并且所产生的扩增多核苷酸可以具有在其各自的5' 和/或3' 端的彼此不同的序列。在一些情况下,相同模板的扩增多核苷酸具有不同的5' 和/或3' 端,这取决于引物最初结合的位置和核苷酸掺入终止的位置。在一些情况下,(b)中的扩增包括使环状多核苷酸经历包含靶标特异性引物的扩增反应混合物。靶标特异性引物可以指靶向特定基因序列的引物,或者在一些情况下是指靶向衔接子多核苷酸序列的引物。使用靶标特异性引物产生的扩增多核苷酸可以共享共同的第一末端(例如引物),并且可以不共享第二末端,这取决于核苷酸掺入终止的位置。扩增可包括变性、引物结合和引物延伸的多个循环。在一些情况下,可以使扩增的多核苷酸经历进一步扩增,以产生扩增多核苷酸的扩增产物。可能需要额外的扩增,例如,以产生足量的多核苷酸以供下游分析,例如测序分析。所得到的扩增产物可包含个体扩增多核苷酸的多个拷贝。

[0102] 随后可对扩增的多核苷酸和/或其扩增产物进行测序,以产生测序读取。在一些情况下,在未富集的情况下使扩增的多核苷酸和/或扩增产物经历测序。然而,如果需要,可以在测序之前在富集步骤中进行扩增的多核苷酸和/或扩增产物中的一种或多种靶多核苷酸的富集。

[0103] 可以将测序读取分组为读取家族。读取家族可包含任何合适数目的序列读取。在一些情况下,读取家族包含至少5、10、15、20、25、50、75或100个序列读取。在一些情况下,除非存在最小数目的序列读取,否则可能不会将一组序列读取鉴定为读取家族。例如,读取家族包含至少2、3、4、5、6、7、8、9或10个序列读取。在一些情况下,读取家族包含至少25个读取序列。在一些实施方案中,读取家族的序列读取具有相同的接合序列。在一些实施方案中,读取家族的序列读取在5' 和3' 端具有相同的序列,例如,在5' 和3' 端各有至少5个碱基、6个碱基、7个碱基、8个碱基、9个碱基或10个碱基上的序列可以是相同的。在一些情况下,由于扩增和/或测序产生的错误,5' 和3' 端的序列在读取家族的序列读取中不尽相同。当例如通过比对来比较时,读取家族的测序读取可能表现出重叠。在一些情况下,当最佳比对时,读取家族的测序读取表现出至少75%的同一性。当最佳比对时,家族的两个测序读取可以在任何合适长度的碱基上表现出至少75%的同一性(例如,至少80%、85%、90%或95%的同一性)。读取家族中的第一对测序读取可以表现出与读取家族中的第二对测序读取不同的同一性百分比。在一些情况下,为在至少50个碱基(例如,至少60个碱基、70个碱基、80个碱基、90个碱基、100个碱基、110个碱基、120个碱基、130个碱基、140个碱基或150碱基)长度上的比对确定同一性百分比。在一些情况下,在约25-250个碱基、约50-200个碱基、约75-175个碱基或约100-150个碱基的长度上比对。在一些情况下,在测试序列或比较序列的整个长度上比对。在一些实施方案中,当最佳比对时,读取家族的两个测序读取在至少50个碱基(例如,至少60个碱基、70个碱基、80个碱基、90个碱基、100个碱基、110个碱基、120个碱基、130个碱基、140个碱基或150个碱基)的长度上表现出至少75%的同一性(例如,至少80%、85%、90%或95%的同一性)。

[0104] 包含共享环状多核苷酸模板的线性多联体的扩增多核苷酸可以产生具有相同环状多核苷酸序列但在多个的个体分子上的多个线性多联体。在一些情况下,扩增的多核苷酸或其扩增产物的测序读取可包含重复序列的至少一个拷贝。在一些情况下,测序读取包含重复序列的至少两个拷贝(例如,至少三个拷贝、四个拷贝或五个拷贝)。来自读取家族的序列读取的重复序列的平均拷贝数可取决于核酸样品的多核苷酸的长度。例如,如果多联

体的长度相似,与包含相对较短的多核苷酸的样品相比,包含相对较长的多核苷酸的样品可以产生重复较少的多联体。

[0105] 可以通过首先鉴定多联体中重复区段的长度和/或序列,将测序读取分组为读取家族,该多联体对应于环状多核苷酸模板的序列。在一些情况下,鉴定重复区段的长度和/或序列包括将读取段与其他读取比对或与参考序列比对。随后,可以例如通过与参考序列比对来鉴定接合序列。可以多核苷酸的5' 和3' 端的序列以及它们与接点的相对距离(例如,以碱基计)。在5' 和3' 端具有相同接合序列和共享序列的读取可以被分组为读取家族,表示源自相同扩增多核苷酸的扩增产物的测序读取,或环状多核苷酸的相同分子拷贝。

[0106] 在一些情况下,通过确认序列差异出现在具有相同接合序列但在各自的5' 和3' 端具有不同序列的第二读取家族,可以将读取家族中所观察到的序列差异判定为真实序列差异而不是扩增和/或测序错误的结果。具有相同接合序列但具有不同的5' 和/或3' 端的两个读取家族可以对应于相同环状多核苷酸的两种扩增多核苷酸。观察与相同环状多核苷酸相对应的两个读取家族中的序列差异可以是一种确认序列差异确实存在于环状多核苷酸上,而不是一种扩增多核苷酸的扩增和/或测序错误的结果的方法。

[0107] 在一些情况下,如果序列差异出现在读取家族的大多数测序读取中,则认为在读取家族的序列读取中观察到的序列差异是序列差异。在一些情况下,如果序列差异出现在读取家族的至少50%的测序读取(例如,至少60%、70%、80%、90%或95%的测序读取)中,则认为读取家族的序列读取中观察到的序列差异是序列差异。在一些情况下,如果序列差异出现在读取家族的100%的测序读取中,则认为读取家族的序列读取中观察到的序列差异是序列差异。在一些情况下,当序列差异出现在来自第一扩增多核苷酸的大多数测序读取和来自第二扩增多核苷酸的大多数测序读取中时,将测序读取中检测到的序列差异判定为序列变体。在一些情况下,当序列差异出现在来自第一扩增多核苷酸的至少50%的测序读取(例如,至少60%、70%、80%、90%或95%的测序读取)和来自第二扩增多核苷酸的至少50%测序读取(例如,至少60%、70%、80%、90%或95%的测序读取)中时,将测序读取中检测到的序列差异判定为序列变体。在一些情况下,当序列差异出现在来自第一扩增多核苷酸的100%的测序读取和来自第二扩增多核苷酸的100%的测序读取中时,将测序读取中检测到的序列差异判定为序列变体。

[0108] 在实施本文所述的方法中,可以改善包含多种多核苷酸的样品中的变体检测。在一些情况下,变体检测的错误率降低。在一些实施方案中,变体检测的错误率降低至少5%、10%、15%、20%、25%、30%、35%、40%、45%或50%。在一些情况下,变体检测的灵敏度和/或特异性增加。在一些实施方案中,变体检测的灵敏度增加至少5%、10%、15%、20%、25%、30%、35%、40%、45%或50%。在一些实施方案中,变体检测的特异性增加至少5%、10%、15%、20%、25%、30%、35%、40%、45%或50%。在一些情况下,假阳性率降低。

[0109] 如本文进一步描述的,序列变体可以是关于参考序列的任何变异。可使用本文方法检测的序列变体的非限制性实例包括单核苷酸多态性(SNP)、缺失/插入多态性(DIP)、拷贝数变体(CNV)、短串联重复(STR)、简单序列重复(SSR)、可变数目串联重复(VNTR)、扩增片段长度多态性(AFLP)、基于反转录转座子的插入多态性、序列特异性扩增多态性,以及可检测为序列变体的表观遗传标记的差异(例如甲基化差异)。在一些情况下,序列变体是多态性,例如单核苷酸多态性。

[0110] 核酸样品可以是来自受试者的样品。在一些情况下,样品来自人受试者。在一些情况下,样品包括来自受试者(例如人受试者)的尿液、粪便、血液、唾液、组织或体液。在一些情况下,样品包含肿瘤细胞。在一些情况下,样品包括福尔马林固定、石蜡包埋的样品。在一些情况下,样品的多种多核苷酸包含无细胞多核苷酸。无细胞多核苷酸可包含无细胞DNA,并且在一些情况下,可包含循环肿瘤DNA。无细胞多核苷酸可包含无细胞RNA,在一些情况下,还可包含循环肿瘤RNA。

[0111] 如前所述,多种多核苷酸可以是单链的。在一些情况下,多核苷酸是双链形式,并通过例如变性进行处理以在进行环化之前产生单链。在一些情况下,将双链多核苷酸环化以产生双链环,并且双链环通过例如变性进行处理以产生单链环。

[0112] 在另一方面,本公开内容提供了例如在包含多种多核苷酸的核酸样品中进行滚环扩增的方法。在一些实施方案中,多种多核苷酸中的每种多核苷酸具有5'端和3'端,并且该方法包括:(a)使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,每种多核苷酸具有在5'端和3'端之间的接点;(b)将连接酶降解;以及(c)在将连接酶降解后扩增(a)的环状多核苷酸,其中在步骤(a)和(c)之间未纯化或分离多核苷酸。在一些实施方案中,该方法包括另外的步骤(d)对扩增的多核苷酸进行测序,以产生多个测序读取;(e)鉴定测序读取与参考序列之间的序列差异;以及(f)将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体。在一些实施方案中,该方法包括鉴定测序读取与参考序列之间的序列差异,并将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体,其中:(a)测序读取对应于至少两种环状多核苷酸的扩增产物;并且(b)至少两种环状多核苷酸中的每一种包含通过连接相应多核苷酸的5'端和3'端而形成的不同接点。

[0113] 在另一方面,本公开内容提供了例如在包含多种多核苷酸的核酸样品中进行滚环扩增的方法。在一些实施方案中,多种多核苷酸中的每种多核苷酸具有5'端和3'端,并且该方法包括:(a)使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,每种多核苷酸具有在5'末端和3'端之间的接点;(b)将连接酶降解;(c)在将连接酶降解后扩增(a)的环状多核苷酸,以产生扩增的多核苷酸,其中在步骤(a)和(c)之间未纯化或分离多核苷酸;(d)将扩增的多核苷酸剪切,以产生剪切多核苷酸,每种剪切多核苷酸在5'端和/或3'端包含一个或多个剪切点。在一些实施方案中,该方法包括另外的步骤(e)对剪切的多核苷酸进行测序,以产生多个测序读取;(f)鉴定测序读取与参考序列之间的序列差异;以及(g)当序列差异出现在至少两种不同的剪切多核苷酸中时,将序列差异判定为序列变体。连接酶在(c)中的扩增之前降解可以提高可扩增多核苷酸的回收率。

[0114] 在一些实施方案中,该方法包括鉴定测序读取与参考序列之间的序列差异,并将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体,其中:(a)测序读取对应于至少两种环状多核苷酸的扩增产物;并且(b)至少两种环状多核苷酸中的每一种包含通过连接相应多核苷酸的5'端和3'端而形成的不同接点。在一些实施方案中,该方法包括当(i)序列差异出现在具有不同接点的至少两种环状多核苷酸中;(ii)在双链输入分子的两条链上都鉴定到序列差异;和/或(iii)序列差异出现在由包括滚环扩增在内的扩增形成的多联体的共有序列中时,进一步将序列差异判定为序列变体。

[0115] 通常,术语“序列变体”是指序列中相对于一个或多个参考序列的任何变异。一般

而言,对于参考序列已知的个体的给定群体,序列变体以比参考序列更低的频率发生。例如,特定细菌属可具有16S rRNA基因的共有参考序列,但该属内的个体物种可在基因(或其部分)内具有一个或多个序列变体,该变体可用于在细菌群中鉴定出该物种。作为另一实例,相同物种的多个个体的序列(或同一个体的多个测序读取)在最佳比对时可以产生共有序列,并且关于该共有序列的序列变体可用于鉴定指示危险污染的群体中的突变体。通常,“共有序列”是指反映序列中每个位置处的最常见碱基选择的核苷酸序列,其中系列相关核苷酸已经历大量的数学和/或序列分析,例如根据各种序列比对算法中的任一种的最佳序列比对。有多种比对算法可用,其中一些在本文中进行了描述。在一些实施方案中,参考序列是单个已知参考序列,例如单个个体的基因组序列。在一些实施方案中,参考序列是通过对多个已知序列,如作为参考群体的多个个体的基因组序列,或来自相同个体的多核苷酸的多个测序读取进行比对而形成的共有序列。在一些实施方案中,参考序列是通过最佳地比对来自分析中的样品的序列而形成的共有序列,使得序列变体表示相对于同一样品中的对应序列的变异。在一些实施方案中,序列变体在群体中以低频率发生(也称为“罕见”序列变体)。例如,序列变体可以以约为或低于约5%、4%、3%、2%、1.5%、1%、0.75%、0.5%、0.25%、0.1%、0.075%、0.05%、0.04%、0.03%、0.02%、0.01%、0.005%、0.001%或更低的频率发生。在一些实施方案中,序列变体以约为或低于约0.1%的频率发生。

[0116] 序列变体可以是相对于参考序列的任何变异。序列变异可以由一个核苷酸或多个核苷酸(例如2、3、4、5、6、7、8、9、10个或更多个核苷酸)的变化、插入或缺失组成。当序列变体包含两个或更多个核苷酸差异时,不同的核苷酸可以是彼此相邻近的,或不连续的。序列变体类型的非限制性实例包括单核苷酸多态性(SNP)、缺失/插入多态性(DIP)、拷贝数变体(CNV)、短串联重复(STR)、简单序列重复(SSR)、可变数目串联重复(VNTR)、扩增片段长度多态性(AFLP)、基于反转录转座子的插入多态性、序列特异性扩增多态性和可检测为序列变体的表观遗传标记的差异(例如,甲基化差异)。

[0117] 可经历本文所述方法的核酸样品可衍生自任何合适的来源。在一些实施方案中,使用的样品是环境样品。环境样品可以来自任何环境来源,例如,天然存在的或人造的大气、水系统、土壤或任何其他感兴趣的样品。在一些实施方案中,环境样品可以从,例如大气病原体收集系统、地下沉积物、地下水、地下深处的古代水、草地的植物根-土界面、沿海水和污水处理厂获得。

[0118] 来自样品的多核苷酸可以是多种多核苷酸中的任一种,包括但不限于,DNA、RNA、核糖体RNA(rRNA)、转运RNA(tRNA)、微小RNA(miRNA)、信使RNA(mRNA),其中任何一种的片段,或其中任何两种或更多种的组合。在一些实施方案中,样品包含DNA。在一些实施方案中,样品包含基因组DNA。在一些实施方案中,样品包含线粒体DNA、叶绿体DNA、质粒DNA、细菌人工染色体、酵母人工染色体、寡核苷酸标签或其组合。在一些实施方案中,样品包含通过扩增产生的DNA,例如通过使用引物和DNA聚合酶的任何合适的组合的引物延伸反应,包括但不限于聚合酶链反应(PCR)、逆转录及其组合。当引物延伸反应的模板是RNA时,逆转录产物称为互补DNA(cDNA)。用于引物延伸反应的引物可包含对一种或多种靶标特异的序列、随机序列、部分随机序列及其组合。通常,样品多核苷酸包含样品中存在的任何多核苷酸,其可以包含或不包含靶多核苷酸。多核苷酸可以是单链、双链或其组合。在一些实施方案中,可经历本公开内容方法的多核苷酸是单链多核苷酸,其可以存在或不存在双链多核苷酸。



酸。在一些实施方案中,多核苷酸是单链DNA。单链DNA(ssDNA)可以是以单链形式分离的ssDNA,或以双链形式分离并随后出于本公开内容方法中的一个或多个步骤的目的而制成单链的DNA。

[0119] 在一些实施方案中,在没有提取步骤和/或没有纯化步骤的情况下使多核苷酸经历后续步骤(例如环化和扩增)。例如,可以在没有提取步骤的情况下处理流体样品以除去细胞,从而产生纯化的液体样品和细胞样品,然后从纯化的流体样品中分离DNA。有多种分离多核苷酸的方法可用,例如通过沉淀,或与基底非特异性结合并随后洗涤该基底以释放结合的多核苷酸。在没有细胞提取步骤的情况下从样品中分离多核苷酸时,多核苷酸将主要是细胞外或“无细胞”多核苷酸,如无细胞DNA和无细胞RNA,其可以对应于死亡或受损的细胞。这些细胞的特性可用于表征其所源自的细胞或细胞群,如肿瘤细胞(例如癌症检测时)、胎儿细胞(例如产前诊断时)、来自移植组织的细胞(例如移植失败的早期检测时)或微生物群落的成员。

[0120] 如果处理样品以例如从样品中的细胞提取多核苷酸,则可使用多种提取方法。例如,可以通过用苯酚、苯酚/氯仿/异戊醇或类似制剂(包括TRIzol和TriReagent)进行有机提取来纯化核酸。提取技术的其他非限制性实例包括:(1)有机萃取然后乙醇沉淀,例如,使用苯酚/氯仿有机试剂(Ausubel等人,1993),使用或不使用自动核酸提取器,例如可从Applied Biosystems(Foster city, Calif)获得的341DNA型提取器;(2)固定相吸附法(美国专利号5,234,809;Walsh等人,1991);和(3)盐诱导核酸沉淀法(Miller等人,1988),该沉淀方法一般被称作“盐析”法。核酸分离和/或纯化的另一个实例包括使用核酸能够特异性或非特异性结合的磁性颗粒,然后使用磁体分离珠子,并洗涤和从珠子中洗脱核酸(参见,例如,美国专利号5,705,628)。在一些实施方案中,上述分离方法之前可先进行酶消化步骤以帮助从样品中去除不需要的蛋白质,例如用蛋白酶K或其他类似的蛋白酶进行消化。参见,例如,美国专利号7,001,724。如果需要,可向裂解缓冲液中添加RNase抑制剂。对于特定的细胞或样品类型,可能需要在方案中增加蛋白质变性/消化步骤。纯化方法可以针对分离DNA、RNA或此两者。当DNA和RNA在提取程序过程中或之后被一起分离时,可使用进一步的步骤来与另一种分开地纯化一种或两者。也可生成提取的核酸的亚级分,例如,根据大小、序列或其他物理或化学特性进行纯化。除了初始核酸分离步骤,核酸的纯化还可以在所公开的方法的任意步骤之后进行,例如用于去除过量的或不需要的试剂、反应物或产物。多种用来确定样品中的核酸量和/或核酸纯度的方法是可用的,例如通过吸光度(例如,在260nm、280nm处的光吸收,和其比值)和标记物的检测(例如,荧光染料和嵌入剂,例如SYBR绿、SYBR蓝、DAPI、碘化丙锭、Hoechst染色剂、SYBR金、溴化乙锭)。

[0121] 如果需要,可以在进一步处理之前将来自样品的多核苷酸片段化。片段化可以通过多种方法中的任一种完成,包括化学、酶促和机械的片段化。在一些实施方案中,片段的平均长度或中值长度为约10至约1,000个核苷酸,例如10-800、10-500、50-500、90-200或50-150个核苷酸。在一些实施方案中,片段的平均或中值长度为约或小于约100、200、300、500、600、800、1000或1500个核苷酸。在一些实施方案中,片段为约90-200个核苷酸,并且/或者具有约150个核苷酸的平均长度。在一些实施方案中,片段化是机械地完成的,包括使样品多核苷酸经历声学超声处理。在一些实施方案中,片段化包括在适合于一种或多种酶的条件下,用一种或多种酶处理样品多核苷酸,以产生双链核酸断裂。可用于产生多核苷酸



片段的酶的实例包括序列特异性核酸酶和非序列特异性核酸酶。核酸酶的非限制性实例包括DNA酶I、片段化酶、限制性内切核酸酶、其变体及其组合。例如,在不存在 $Mg^{++}$ 且存在 $Mn^{++}$ 的情况下,用DNA酶I消化可以诱导DNA中的随机双链断裂。在一些实施方案中,片段化包括用一种或多种限制性内切核酸酶处理样品多核苷酸。片段化可产生具有5'突出端、3'突出端、平端或其组合的片段。在一些实施方案中,如当片段化包括使用一种或多种限制性内切核酸酶时,样品多核苷酸的裂解留下具有可预测序列的突出端。可以通过标准方法(如柱纯化或从琼脂糖凝胶中分离)使片段化的多核苷酸经历对片段进行大小选择的步骤。

[0122] 根据一些实施方案,来自样品的多种多核苷酸中的多核苷酸被环化。环化可包括将多核苷酸的5'端接合到同一多核苷酸的3'端,接合到样品中的另一多核苷酸的3'端,或接合到来自不同来源的多核苷酸(例如,人工多核苷酸,如寡核苷酸衔接子)的3'端。在一些实施方案中,多核苷酸的5'端接合到同一多核苷酸的3'端(也称为“自接合”)。在一些实施方案中,选择环化反应的条件以利于在特定长度范围内的多核苷酸的自接合,以便生成具有特定平均长度的环化多核苷酸群体。例如,可以选择环化反应条件以利于长度短于约5000、2500、1000、750、500、400、300、200、150、100、50个或更少的核苷酸的多核苷酸的自接合。在一些实施方案中,偏向于长度为50-5000个核苷酸、100-2500个核苷酸或150-500个核苷酸的片段,以使得环化多核苷酸的平均长度落入各自的范围内。在一些实施方案中,80%或更多的环化片段的长度为50-500个核苷酸,例如长度为50-200个核苷酸。可以优化的反应条件包括分配给接合反应的时间长度、各种试剂的浓度和待接合的多核苷酸的浓度。在一些实施方案中,环化反应保持在环化前存在于样品中的片段长度的分布。例如,环化前样品中的片段长度以及环化多核苷酸的平均值、中值、众数(mode)和标准差中的一个或多个彼此在75%、80%、90%、95%或更高的百分比以内。

[0123] 在一些情况下,使用一种或多种衔接子寡核苷酸,使得样品中多核苷酸的5'端和3'端通过一个或多个间插衔接子寡核苷酸接合形成环状多核苷酸,而不是优先形成自接合环化产物。例如,多核苷酸的5'端可以接合至衔接子的3'端,而同一衔接子的5'端可以接合至同一多核苷酸的3'端。衔接子寡核苷酸包括具有某种序列的任意寡核苷酸,该序列的至少一部分是已知的,它能够与样品多核苷酸连接。衔接子寡核苷酸可以包含DNA、RNA、核苷酸类似物、非典型核苷酸、标记的核苷酸、修饰的核苷酸或其组合。衔接子寡核苷酸可以是单链的、双链的或部分双链体。通常,部分双链体衔接子包含一个或多个单链区域和一个或多个双链区域。双链衔接子可包含彼此相互杂交的两个分离的寡核苷酸(也称作“寡核苷酸双链体”),并且杂交可留下一个或多个平端、一个或多个3'突出端、一个或多个5'突出端、一个或多个由错配的和/或不配对的核苷酸所导致的凸起,或其任意组合。当衔接子的两个杂交区域被非杂交区域彼此分开时,会产生“气泡”结构。不同种类的衔接子,例如具有不同序列的衔接子,可以组合使用。不同的衔接子可以在顺序的反应中或同时地接合到样品多核苷酸。在一些实施方案中,将相同的衔接子添加至靶多核苷酸的两端。例如,可以将第一和第二衔接子添加到相同的反应中。可以在与样品多核苷酸组合之前操作衔接子。例如,可以添加或去除末端磷酸盐。

[0124] 在使用衔接子寡核苷酸的情况下,衔接子寡核苷酸可包含多种序列元件中的一个或多个,包括但不限于,一个或多个扩增引物退火序列或其互补体、一个或多个测序引物退火序列或其互补体、一个或多个条形码序列、一个或多个在多个不同衔接子或不同衔接子

的子集之间共有的共同序列、一个或多个限制酶识别位点、一个或多个与一个或多个靶多核苷酸突出端互补的突出端、一个或多个探针结合位点(例如,用于附接至测序平台,如用于大规模平行测序的流动池,如Illumina, Inc.开发的流动池)、一个或多个随机的或接近随机的序列(例如,在一个或多个位置处从一组两个或更多个不同核苷酸中随机选择的一个或多个核苷酸,其中在一个或多个位置处选择的不同核苷酸中的每一个均表现在包含随机序列的一组衔接子中),或其组合。在一些情况下,衔接子可用于纯化这些含有衔接子的环,例如通过使用以包含衔接子互补序列的寡核苷酸涂覆的珠子(为了易于处理,特别是磁珠),该珠子可以通过与之杂交而“捕获”具有正确衔接子的闭合环,洗掉那些不包含衔接子和任何未连接的组分的环,然后从珠子上释放所捕获的环。此外,在一些情况下,杂交的捕获探针和目标环的复合体可直接用于生成多联体,例如通过直接滚环扩增(RCA)。在一些实施方案中,环中的衔接子还可用作测序引物。两个或更多个序列元件可以是彼此不邻近的(例如被一个或多个核苷酸隔开)、彼此邻近的、部分重叠的或完全重叠的。例如,扩增引物退火序列还可作为测序引物退火序列。序列元件可位于或靠近3'端、位于或靠近5'端或在衔接子寡核苷酸内部。序列元件可以是任意合适的长度,例如约为或少于约3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50个或更多个核苷酸的长度。衔接子寡核苷酸可具有任意合适的长度,至少足以容纳其所包含的一个或多个序列元件。在一些实施方案中,衔接子的长度约为或少于约10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100、200个或更多个核苷酸。在一些实施方案中,衔接子寡核苷酸的长度在约12至40个核苷酸的范围内,例如长度为约15至35个核苷酸。

[0125] 在一些实施方案中,与来自一个样品的片段化多核苷酸接合的衔接子寡核苷酸包含一个或多个所有衔接子寡核苷酸所共有的序列和对于与该特定样品的多核苷酸接合的衔接子而言独特的条形码,以使得该条形码序列可用于区分来源于一个样品或衔接子接合反应的多核苷酸与来源于另一个样品或衔接子接合反应的多核苷酸。在一些实施方案中,衔接子寡核苷酸包含与一个或多个靶多核苷酸突出端互补的5'突出端、3'突出端或此两者。互补突出端可以是一个或多个核苷酸的长度,包括但不限于1、2、3、4、5、6、7、8、9、10、11、12、13、14、15个或更多个核苷酸的长度。互补突出端可包含固定的序列。衔接子寡核苷酸的互补突出端可以包含一个或多个核苷酸的随机序列,以使得在一个或多个位置处从一组两个或更多个不同核苷酸中随机选择一个或多个核苷酸,其中在一个或多个位置处选择的不同核苷酸中的每一个都表现在包含随机序列的一组具有互补突出端的衔接子中。在一些实施方案中,衔接子突出端与通过限制性内切核酸酶消化产生的靶多核苷酸突出端互补。在一些实施方案中,衔接子突出端由腺嘌呤或胸腺嘧啶组成。

[0126] 多种环化多核苷酸的方法是可用的。图28A-28E示出了多核苷酸环化方法的非限制性实例。在一些实施方案中,环化包含酶促反应,例如使用连接酶(例如RNA或DNA连接酶)。多种连接酶是可行的,包括但不限于,Circligase™(Epicentre;Madison,WI)、RNA连接酶、T4RNA连接酶1(ssRNA连接酶,其作用于DNA和RNA两者)。此外,如果不存在dsDNA模板,T4DNA连接酶也可以连接ssDNA,尽管这通常是缓慢的反应。连接酶的其他非限制性实例包括:NAD-依赖性连接酶,包括Taq DNA连接酶、丝状栖热菌(*Thermus filiformis*)DNA连接酶、大肠杆菌(*E. coli*)DNA连接酶、Tth DNA连接酶、水管致黑栖热菌(*Thermus scotoductus*)DNA连接酶(I和II)、热稳定的连接酶、Ampligase热稳定的DNA连接酶、VanC-

型连接酶、9°N DNA连接酶、Tsp DNA连接酶和通过生物勘探发现的新型连接酶;ATP-依赖性连接酶,包括T4RNA连接酶、T4DNA连接酶、T3DNA连接酶、T7DNA连接酶、Pfu DNA连接酶、DNA连接酶I、DNA连接酶III、DNA连接酶IV和通过生物勘探发现的新型连接酶;以及野生型、突变同工型,及其遗传工程变体。当需要自接合时,可调节多核苷酸和酶的浓度以促进分子内环而非分子间结构的形成。反应温度和时间也可调整。在一些实施方案中,使用60°C来促进分子内环的形成。在一些实施方案中,反应时间为12-16小时。反应条件可以是所选择的酶的制造商所规定的条件。在一些实施方案中,可以包括外切核酸酶步骤以在环化反应后消化任何未连接的核酸。也就是说,闭合环不含游离5'或3'端,因此引入5'或3'外切核酸酶不会消化闭合环但会消化未连接的组分。这尤其可用于多路系统中。

[0127] 通常,多核苷酸的末端彼此接合以形成环状多核苷酸(直接地,或者使用一个或多个居间衔接子寡核苷酸)会产生具有接合序列的接点。当多核苷酸的5'端和3'端通过衔接子多核苷酸接合时,术语“接点”指该多核苷酸与衔接子之间的接点(例如,5'端接点或3'端接点之一),或指多核苷酸的5'端与3'端之间由此形成且包含衔接子多核苷酸的接点。当多核苷酸的5'端和3'端在不使用间插衔接子的情况下接合时(例如,单链DNA的5'端和3'端),术语“接点”可以指这两个末端相接合的点。接点可通过包含接点的核苷酸的序列(也被称为“接合序列”)进行鉴定。在一些实施方案中,样品包含具有通过以下过程形成的末端混合物的多核苷酸:自然降解过程(如细胞裂解、细胞死亡和DNA从细胞释放到其周围环境的其他过程,DNA在该周围环境中可进一步降解,如从无细胞多核苷酸中,如无细胞DNA和无细胞RNA)、作为样品处理的副产物的片段化(诸如固定、染色和/或储存过程),以及通过切割不限于特定靶序列的DNA的方法进行的片段化(例如,机械片段化,如通过超声处理;非序列特异性核酸酶处理,如DNA酶I、片段化酶)。当样品包含具有末端混合物的多核苷酸时,两个多核苷酸具有相同的5'端或3'端的可能性很低,并且两个多核苷酸独立地具有相同的5'端和3'端两者的可能性极低。因此,在一些实施方案中,甚至在两个多核苷酸包含具有相同靶序列的部分时,可以使用接点来区分不同的多核苷酸。当多核苷酸末端在不使用间插衔接子的情况下接合时,接合序列可通过与参考序列进行比对来鉴定。例如,当两种组分序列的顺序相对于参考序列似乎反转时,显示发生反转的点可以指示在该点处的接点。当多核苷酸末端通过一个或多个衔接子序列接合时,接点可通过与已知衔接子序列的邻近进行鉴定,或者在测序读取的长度足以从环化多核苷酸的5'和3'端获得序列的情况下,则通过如上所述的比对进行鉴定。在一些实施方案中,特定接点的形成是十分罕见的事件,使得其在样品的环化多核苷酸之间是独特的。

[0128] 图4A-4C示出了多核苷酸环化方法的三个非限制性实例。在顶部(图4A),多核苷酸在无衔接子的情况下环化,而中间的方案(图4B)描绘了衔接子的使用,而底部的方案(图4C)使用了两种衔接子。当使用两种衔接子时,一种可以接合到多核苷酸的5'端,而第二衔接子可以接合到同一多核苷酸的3'端。在一些实施方案中,衔接子连接可包括使用两种不同的衔接子以及与两种衔接子互补以促进连接的“夹板”核酸。也可以使用叉形或“Y”形衔接子。在使用两种衔接子的情况下,由于自退火,在后续步骤中可以除去两端具有相同衔接子的多核苷酸。图1-3描绘了根据本公开内容的方法的实施方案,其中多核苷酸在不存在衔接子(图1)和存在衔接子(图2和3)的情况下环化。具有衔接子的环化多核苷酸(图2和3)可以使用靶标特异性引物(图2)或与衔接子序列杂交的引物(图3)通过滚环扩增(RCA)扩增。

[0129] 图6A-6B示出了多核苷酸(如单链DNA)环化的其他非限制性示例方法。衔接子可以不对称地添加到多核苷酸的5' 或3' 端。如图6A所示,单链DNA(ssDNA)在3' 端可以具有游离羟基基团,并且衔接子可以有封闭的3' 端,使得在连接酶存在的情况下,优选的反应将ssDNA的3' 端与衔接子的5' 端接合。在该实施方案中,在分子内连接形成环之前,使用诸如聚乙二醇(PEG)等试剂来驱动单个ssDNA片段和单个衔接子的分子间连接可能是有用的。也可以实现相反的末端顺序(封闭的3',游离的5' 等)。一旦完成线性连接,可以用酶处理经连接的段,以除去封闭部分,例如通过使用激酶或其他合适的酶或化学法。一旦除去封闭部分,添加环化酶,如CircLigase,允许发生分子内反应以形成环化的多核苷酸。如图6B所示,通过使用一条具有封闭5' 或3' 端的链的双链衔接子,可以形成双链结构,该结构在连接后产生具有切口的双链片段。然后可以分离两条链,除去封闭部分,并将单链片段环化以形成环化的多核苷酸。在一些情况下,如图8所示,将双链DNA(dsDNA)环化,以产生环化的双链环。可以使双链环变性,以允许引物结合和两条链的扩增。

[0130] 在一些实施方案中,使用分子钳将多核苷酸的两个末端(例如单链DNA)在一起,以增强分子内环化的速率。图5提供了一个这样的过程的示例说明。这可以使用或不使用衔接子完成。在平均多核苷酸片段长度大于约100个核苷酸的情况下,分子钳的使用可能特别有用。在一些实施方案中,分子钳探针包含三个结构域:第一结构域、间插结构域和第二结构域。第一和第二结构域将会通过序列互补性与靶多核苷酸中的相应序列杂交。分子钳探针的间插结构域可能不与靶序列显著杂交。因此,该钳与靶多核苷酸的杂交可使靶序列的两端更接近,这有利于在环化酶存在的情况下靶序列的分子内环化。在一些实施方案中,这还有额外用途,因为分子钳也可以用作扩增引物。

[0131] 环化后,采用蛋白质降解步骤从反应产物中除去连接酶。在一些实施方案中,蛋白质降解包括将环化反应中所用的连接酶去除或降解的处理。在一些实施方案中,将连接酶降解的处理包括用蛋白酶如蛋白酶K的处理。蛋白酶K处理可遵循制造商方案或标准方案(例如Sambrook和Green,Molecular Cloning:A Laboratory Manual,第4版(2012)中提供的)。在一些实施方案中,蛋白质降解包括用低pH值或酸性溶液或缓冲液的处理。在一些实施方案中,蛋白质降解包括将该反应加热,例如将反应加热至高于55℃、高于60℃、高于65℃、高于70℃或更高。在一些实施方案中,线性多核苷酸在环化后降解。在一些实施方案中,使用外切核酸酶来降解线性多核苷酸。在一些实施方案中,外切核酸酶包括λ外切核酸酶。在一些实施方案中,外切核酸酶包括RecJf核酸酶。在一些实施方案中,外切核酸酶选自ExoI、ExoIII、ExoV、ExoVII和ExoT中的至少一种。

[0132] 环化后可直接对环化多核苷酸进行测序。或者,可以在测序之前进行一个或多种扩增反应。通常,“扩增”是指由靶多核苷酸或其部分形成一个或多个拷贝的过程。多种扩增多核苷酸(例如DNA和/或RNA)的方法是可用的。扩增可以是线性的,指数式的,或在多阶段扩增过程中涉及线性和指数阶段。扩增方法可包括温度的改变,例如热变性步骤,或者可以是不需要热变性的等温过程。聚合酶链反应(PCR)采用变性、引物对与相反链的退火和引物延伸的多个循环,以指数式增加靶序列的拷贝数。退火的核酸链的变性可以通过施加热、提高局部金属离子浓度(例如,美国专利号6,277,605)、超声辐射(例如,W0/2000/049176)、施加电压(例如,美国专利号5,527,670、美国专利号6,033,850、美国专利号5,939,291和美国专利号6,333,157)和与结合到磁响应性材料上的引物相组合地施加电磁场(例如,美国专

利号5,545,540)来实现。在称为RT-PCR的变化中,使用逆转录酶(RT)从RNA制备互补DNA(cDNA),然后通过PCR扩增cDNA,以产生多拷贝的DNA(例如美国专利号5,322,770和美国专利号5,310,652)。等温扩增方法的一个实例是链置换扩增,通常称为SDA,其使用以下过程的循环:引物序列对与靶序列的相反链的退火,在dNTP存在下的引物延伸以生成双链体半硫代磷酸化的引物延伸产物,内切核酸酶介导的半修饰限制性内切核酸酶识别位点的切口形成,和聚合酶介导的从切口3'端的引物延伸以取代已存在的链并产生用于下一轮引物退火、切口形成和链置换的链,从而导致产物的几何扩增(例如,美国专利号5,270,184和美国专利号5,455,166)。嗜热SDA(tSDA)在基本相同的方法中在更高的温度下使用嗜热内切核酸酶和聚合酶(欧洲专利号0 684 315)。其他扩增方法包括滚环扩增(RCA)(例如Lizardi, "Rolling Circle Replication Reporter Systems"美国专利号5,854,033);解旋酶依赖性扩增(HDA)(例如,Kong等人, "Helicase Dependent Amplification Nucleic Acids"美国专利申请公开号US 2004-0058378A1);和环介导的等温扩增(LAMP)(例如Notomi等人, "Process for Synthesizing Nucleic Acid"美国专利号6,410,278)。在一些情况下,等温扩增采用通过RNA聚合酶从启动子序列起进行转录,例如可以引入寡核苷酸引物中。基于转录的扩增方法包括基于核酸序列的扩增,也称为NASBA(例如美国专利号5,130,238);依赖于使用RNA复制酶(通常称为QB复制酶)扩增探针分子本身的方法(例如,Lizardi, P.等(1988) BioTechnol. 6, 1197-1202);自动维持序列复制(例如,Guatelli, J.等人(1990) Proc. Natl. Acad. Sci. USA 87, 1874-1878; Landgren(1993) Trends in Genetics 9, 199-202;和HELEN H. LEE等人, NUCLEIC ACID AMPLIFICATION TECHNOLOGIES(1997));和生成额外的转录模板的方法(例如,美国专利号5,480,784和美国专利号5,399,491)。另外的等温核酸扩增方法包括使用含有非典型核苷酸(例如,尿嘧啶或RNA核苷酸)的引物并结合使用在非典型核苷酸处切割核酸的酶(例如, DNA糖基化酶或RNaseH),以暴露针对额外的引物的结合位点(例如,美国专利号6,251,639、美国专利号6,946,251和美国专利号7,824,890)。等温扩增过程可以是线性的或指数式的。

[0133] 在一些实施方案中,扩增包括滚环扩增(RCA)。典型的RCA反应混合物包含一种或多种引物、聚合酶和dNTP,并且生成多联体。一般来说,RCA反应中的聚合酶是具有链置换活性的聚合酶。多种这样的聚合酶是可用的,其非限制性实例包括外切核酸酶-DNA聚合酶I大(Klenow)片段、Phi29DNA聚合酶、Taq DNA聚合酶等。通常,多联体是多核苷酸扩增产物,其包含来自模板多核苷酸的靶序列的两个或更多个拷贝(例如,靶序列的约或大于约2、3、4、5、6、7、8、9、10或更多个拷贝;在一些实施方案中,约或大于约2个拷贝)。扩增引物可具有任意合适的长度,诸如约或至少约5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100个或更多个核苷酸,其任何部分或全部可与该引物所杂交的相应靶序列互补(例如,约或至少约5、10、15、20、25、30、35、40、45、50或更多个核苷酸)。图7A-7C描绘了合适引物的三个非限制性实例。图7A显示了不使用衔接子且使用靶标特异性引物,其可用于检测特定靶序列内是否存在序列变体。在一些实施方案中,在同一反应中使用针对多个靶标的多个靶标特异性引物。例如,可以在一个扩增反应中使用针对约为或至少约10、50、100、150、200、250、300、400、500、1000、2500、5000、10000、15000个或更多个不同靶序列的靶标特异性引物,以便平行地扩增相应数目的靶序列(如果存在)。多个靶序列可以对应于相同基因的不同部分、不同基因或非基因序列。当多个引物以单一基因中的多个靶序列为目标时,引物可

以沿着基因序列间隔(例如,间隔开约或至少约50个核苷酸,每50-150个核苷酸,或每50-100个核苷酸),以覆盖靶基因的全部或指定部分。图7C示出了与衔接子序列(在一些情况下可能是衔接子寡核苷酸本身)杂交的引物的使用。

[0134] 图7B示出了通过随机引物扩增的实例。通常,随机引物包含一个或多个随机或近随机的序列(例如,在一个或多个位置处从一组两个或更多个不同核苷酸中随机选择的一个或多个核苷酸,其中在一个或多个位置处选择的不同核苷酸中的每一个均表现在包含随机序列的衔接子的池中)。通过这种方法,多核苷酸(例如所有或基本上所有环化多核苷酸)可以以序列非特异性的方式扩增。这样的过程可称为“全基因组扩增”(WGA);然而,典型的WGA方案(其不涉及环化步骤)不能有效扩增短多核苷酸,如本公开内容设想的多核苷酸片段。关于WGA程序的进一步说明讨论,请参见例如Li等人(2006) *J Mol. Diagn.* 8(1):22-30。

[0135] 在测序之前将环化的多核苷酸扩增的情况下,可以在未富集的情况下直接使扩增产物经历测序,或者在一个或多个富集步骤之后使其经历测序。富集可包括纯化一种或多种反应组分,例如通过保留扩增产物或除去一种或多种试剂。例如,扩增产物可以通过与附着于基底的多个探针杂交,然后例如通过洗涤步骤释放捕获的多核苷酸来纯化。或者,扩增产物可以用结合对中的成员标记,然后与附着于基底的结合对中的另一成员结合,并洗涤以释放扩增产物。可能的基底包括但不限于玻璃和改性或功能化玻璃、塑料(包括丙烯酸类塑料、聚苯乙烯,和苯乙烯与其他材料的共聚物、聚丙烯、聚乙烯、聚丁烯、聚氨酯、特氟龙™等)、多糖、尼龙或硝化纤维素、陶瓷、树脂、二氧化硅或二氧化硅基材料,包括硅和改性硅、碳、金属、无机玻璃、塑料、光纤束和各种其他聚合物。在一些实施方案中,基底是珠子或其他小的离散颗粒的形式,它们可以是磁珠或顺磁珠,以便通过施加磁场来促进分离。通常,“结合对”是指第一和第二部分之一,其中第一和第二部分对彼此具有特异性结合亲和力。合适的结合对包括但不限于抗原/抗体(例如,洋地黄毒苷/抗洋地黄毒苷、二硝基苯基(DNP)/抗DNP、丹磺酰基-X-抗丹磺酰基、荧光素/抗荧光素、荧光黄/抗荧光黄,和罗丹明/抗罗丹明);生物素/抗生物素蛋白(或生物素/链霉抗生物素蛋白);钙调蛋白结合蛋白(CBP)/钙调蛋白;激素/激素受体;凝集素/碳水化合物;肽/细胞膜受体;蛋白A/抗体;半抗原/抗半抗原;酶/辅因子;以及酶/底物。

[0136] 在一些实施方案中,环化多核苷酸扩增后的富集包括一个或多个额外的扩增反应。在一些实施方案中,富集包括在扩增反应混合物中扩增包含序列A和序列B(以5'至3'方向取向)的靶序列,该混合物包含(a)扩增的多核苷酸;(b)包含序列A'的第一引物,其中第一引物通过序列A与序列A'之间的序列互补性与靶序列的序列A特异性杂交;(c)包含序列B的第二引物,其中第二引物通过B与B'之间的序列互补性与包含靶序列的互补序列的互补多核苷酸中存在的序列B'特异性杂交;以及(d)延伸第一引物和第二引物以产生扩增的多核苷酸的聚合酶;其中靶序列中序列A的5'端与序列B的3'端之间的距离为75nt或更小。图10示出了在单重复(除非为环状,否则通常不会被扩增)和包含多个拷贝的靶序列的多联体的情况下,第一和第二引物相对于靶序列的示例排列。考虑到引物相对于靶序列单体的取向,该排列可以称为“背对背”(B2B)或“倒置”引物。用B2B引物扩增有助于富集环状和/或串联扩增产物。此外,这种取向结合相对较小的足迹(由一对引物跨越的总距离)允许在靶序列周围的更多种片段化事件的扩增,因为相比于出现在典型的扩增反应(彼此面对,跨越靶序列)中发现的引物排列中,接合出现在引物之间的可能性更小。背对背引物的其他实施方

案和优点如图13A-13C所示。

[0137] 在一些实施方案中,序列A的5'端与序列B的3'端之间的距离为约或小于约200、150、100、75、50、40、30、25、20、15或更少的核苷酸。在一些实施方案中,序列A是序列B的互补序列。在一些实施方案中,针对多个不同靶序列的多对B2B引物用于同一反应中,以平行扩增多个不同的靶序列(例如,约或至少约10、50、100、150、200、250、300、400、500、1000、2500、5000、10000、15000或更多不同的靶序列)。引物可具有任何合适的长度,如本文其他地方所述。扩增可以包括在适当条件下的任何合适的扩增反应,如本文所述的扩增反应。在一些实施方案中,扩增是聚合酶链反应。

[0138] 在一些实施方案中,B2B引物包含至少两个序列元件,通过序列互补性与靶序列杂交的第一元件,和在第一元件进行杂交的第一杂交温度下的第一扩增阶段期间不与靶序列杂交的5'“尾”(例如,由于尾与紧邻第一元件所结合的位置3'侧的靶序列部分之间缺少序列互补性)。例如,第一引物包含在序列A'的5'侧的序列C,第二引物包含在序列B的5'侧的序列D,并且在第一杂交温度下的第一扩增阶段期间,序列C和序列D都不与多种多联体杂交。在使用这种有尾引物的一些实施方案中,扩增可包括第一阶段和第二阶段;第一阶段包括在第一温度下的杂交步骤,在此期间第一和第二引物与多联体(或环化多核苷酸)杂交并引物延伸;第二阶段包括在高于第一温度的第二温度下的杂交步骤,在此期间第一和第二引物与包含延伸的第一或第二引物或其互补序列的扩增产物杂交,并引物延伸。较高的温度有利于引物延伸产物中引物的第一元件与尾元件之间的杂交,而不利引物中的仅第一元件与多联体内靶序列之间的杂交形成的较短片段。因此,可以使用两阶段扩增以减少可能对短扩增产物有利的程度,从而保持相对较高比例的具有两个或更多个靶序列拷贝的扩增产物。例如,在5个循环的第二温度下杂交和引物延伸(例如,至少5、6、7、8、9、10、15、20或更多个循环)后,反应混合物中至少5%(例如至少5%、6%、7%、8%、9%、10%、15%、20%、25%、30%或更多)的扩增多核苷酸包含两个或更多个靶序列拷贝。根据这种两阶段有尾B2B引物扩增过程的实施方案的说明如图11A-11D所示。

[0139] 在一些实施方案中,富集包括在倾斜(skewed)的条件下的扩增,以增加来自多联体扩增子的长度。例如,可以降低引物浓度,使得不是每个引发位点都会与引物杂交,从而延长PCR产物。类似地,在循环期间减少引物杂交时间也会减少引物杂交,从而增加平均PCR扩增子大小。此外,增加循环的温度和/或延伸时间可以类似地增加PCR扩增子的平均长度。可以使用这些技术的任何组合。

[0140] 在一些实施方案中,特别是在使用B2B引物进行扩增的情况下,处理扩增产物,以基于大小过滤所得的扩增子,从而减少和/或消除包含多联体混合物中的单体数目。这可以使用多种可用的技术完成,包括但不限于,从凝胶中切除片段和凝胶过滤(例如,以富集长度大于约300、400、500个或更多个核苷酸的片段);以及通过微调结合缓冲液浓度进行SPRI珠子(Agencourt AMPure XP)的大小选择。例如,在与DNA片段混合期间使用0.6x结合缓冲液可用于优先结合大于约500个碱基对(bp)的DNA片段。

[0141] 在扩增产生单链多联体的一些实施方案中,在为测序反应产生的测序文库的形成之前或作为其一部分,将单链转化为双链构建体。从单链核酸产生双链构建体可以使用多种合适的方法。一些可能的方法如图9A-9D所示,而且也可以使用许多其他方法。例如,如图9A所示,使用随机引物、聚合酶、dNTP和连接酶将产生双链。图9B描绘了当多联体含有衔接



子序列时的第二链合成,该衔接子序列可以用作反应的引物。图9C描绘了“环”的使用,其中环衔接子的一个末端被添加到多联体的末端,其中环衔接子具有一小段自杂交核酸。在这种情况下,环衔接子的连接产生自杂交并用作聚合酶引物模板的环。图9D示出了超支化引物的使用,通常最常用于靶序列已知、形成多个链,特别是使用具有强链置换功能的聚合酶的情况。

[0142] 根据一些实施方案,使环化的多核苷酸(或其扩增产物,可任选地经富集)经历测序反应以产生测序读取。这些方法产生的测序读取可以根据本文公开的其他方法使用。多种测序方法是可用的,尤其是高通量测序方法。实例包括但不限于Illumina制造的测序系统(诸如HiSeq®和MiSeq®的测序系统)、Life Technologies制造的测序系统(Ion Torrent®、SOLiD®等)、Roche的454Life Sciences系统、Pacific Biosciences系统等。在一些实施方案中,测序包括使用HiSeq®和MiSeq®系统产生长度约为或多于约50、75、100、125、150、175、200、250、300个或更多个核苷酸的读取。在一些实施方案中,测序包括合成测序过程,其中随着单独的核苷酸被添加至生长的引物延伸产物上,该核苷酸被迭代地鉴定。焦磷酸测序是合成测序法的一个实例,其通过分析所产生的合成混合物中测序反应副产物即焦磷酸的存在而鉴定核苷酸的掺入。特别是,引物/模板/聚合酶复合体与一种类型的核苷酸接触。如果该核苷酸掺入,则聚合反应切割三磷酸链的 $\alpha$ 和 $\beta$ 磷酸之间的三磷酸核苷,从而释放焦磷酸。然后使用化学荧光酶报告系统鉴定所释放的焦磷酸的存在,该系统将含有AMP的焦磷酸转化为ATP,之后用萤光素酶测量ATP以生成可测量的光信号。当检测到光时,碱基已掺入,当未检测到光时,碱基未掺入。在适当的洗涤步骤后,使各种碱基周期性地与该复合体接触,以连续地鉴定模板序列中的后续碱基。参见,例如,美国专利号6,210,891。

[0143] 在相关的测序过程中,将引物/模板/聚合酶复合体固定在基底上,并且该复合体与标记的核苷酸接触。复合体的固定可通过引物序列、模板序列和/或聚合酶进行,并且可以是共价的或非共价的。例如,复合体的固定可以通过聚合酶或引物与基底表面之间的连接而实现。在可替代的构型中,核苷酸具有以及不具有可去除的终止基团。在掺入后,标记物与复合体偶联,因此是可检测的。在携带终止子的核苷酸的情况下,携带可单独鉴定的标记物的全部四种不同的核苷酸与复合体相接触。标记的核苷酸的掺入由于终止子的存在而阻止了延伸,并将标记物添加到复合体中,从而允许鉴定掺入的核苷酸。然后将标记物和终止子从掺入的核苷酸上去除,并在适当的洗涤步骤后重复该过程。在未终止的核苷酸的情况下,如焦磷酸测序那样,将一种类型的标记的核苷酸添加到复合体中以确定其是否将会掺入。在去除核苷酸上的标记基团和适当的洗涤步骤之后,各种不同的核苷酸在同一过程中通过反应混合物进行循环。参见,例如,美国专利号6,833,246,其为了所有目的通过引用而整体并入本文。例如,Illumina基因组分析系统(Illumina Genome Analyzer System)是基于W0 98/44151中描述的技术,其中DNA分子通过锚定探针结合位点(另外被也称为流动池结合位点)结合到测序平台(流动池)上,并且在载玻片上原位扩增。在其上扩增DNA分子的固体表面一般包含多个第一和第二结合寡核苷酸,第一个与靠近或位于靶多核苷酸的一个末端的序列互补,而第二个与靠近或位于靶多核苷酸的另一个末端的序列互补。这种排列允许进行桥式扩增,例如US20140121116中所描述的。DNA分子然后与测序引物退火,并且使用可逆终止子方法逐个碱基地平行测序。在测序引物的杂交之前,可在锚定双链桥的结



合寡核苷酸之一中的切割位点处切割双链桥多核苷酸的一条链,从而留下一条单链不与固体基底结合,其可通过变性去除,而另一条链结合于并可用来与测序引物杂交。通常,Illumina基因组分析仪系统利用具有8个通道的流动池,产生长度为18至36个碱基的测序读取,每次运行产生>1.3Gbp的高质量数据(参见www.illumina.com)。

[0144] 在另一个合成测序过程中,随着模板依赖性合成的进行而实时观察不同标记的核苷酸的掺入。具体而言,随着荧光标记的核苷酸的掺入,观察单独的固定的引物/模板/聚合酶复合物,从而允许随着碱基的添加实时地鉴定每种添加的碱基。在该过程中,标记基团连接于在掺入过程中被切割的核苷酸的一部分上。例如,通过将标记基团连接于在掺入过程中去除的磷酸链的一部分,即,核苷多磷酸上的 $\alpha$ 、 $\beta$ 、 $\gamma$ 或其他末端磷酸基团上,该标记物不掺入新生链中,而是生成天然DNA。对单独的分子观察一般涉及将复合物光学限制在非常小的照明体积内。通过对复合物进行光学限制,产生了受监测的区域,其中随机分散的核苷酸在非常短的时间内存在,而掺入的核苷酸在掺入时能够在观察体积内保留更久的时间。这产生了与掺入事件相关的特征信号,其还可通过所加入的碱基所特有的信号谱进行表征。在相关的方面,在聚合酶或复合物其他部分和掺入的核苷酸上提供相互作用的标记组分,例如荧光共振能量转移(FRET)染料对,从而掺入事件使得标记组分相互邻近,并产生特征信号,这也是所掺入的碱基所特有的(参见,例如,美国专利号6,917,726、7,033,764、7,052,847、7,056,676、7,170,050、7,361,466和7,416,844;和US 20070134128)。

[0145] 在一些实施方案中,样品中的核苷酸可通过连接进行测序。该方法一般使用DNA连接酶来鉴定靶序列,例如,如在聚合酶克隆(polony)方法和在SOLiD技术(Applied Biosystems,目前为Invitrogen)中所使用的。通常,提供一组固定长度的所有可能的寡核苷酸,按照测序位置进行标记。将寡核苷酸退火并连接;DNA连接酶对匹配序列的优先连接会产生对应于该位点的互补序列的信号。

[0146] 在一些实施方案中,在测序分析之前,从扩增的DNA多联体构建测序文库。扩增的DNA多联体可以同时片段化并用测序衔接子标记,如图12A所示。在一些情况下,扩增的DNA多联体被片段化,例如通过超声处理,并且将衔接子添加到片段的两端,如图12B所示。

[0147] 根据一些实施方案,如果测序读取与参考序列之间的序列差异出现在至少两种不同的多核苷酸(例如,两种不同的环状多核苷酸,其可以由于具有不同的接点而被区分)中,则将该序列差异判定为真正的序列变体(例如,存在于扩增或测序前的样品中,并且不是这些过程中的任何一个过程的结果)。因为作为扩增或测序错误的结果的序列变体不可能在包含相同靶序列的两种不同多核苷酸上准确地重复(例如,位置和类型),所以添加该验证参数大大减少了错误序列变体的背景,同时伴有检测样品中的实际序列变异的灵敏度和准确性的提高。在一些实施方案中,频率约为或低于约5%、4%、3%、2%、1.5%、1%、0.75%、0.5%、0.25%、0.1%、0.075%、0.05%、0.04%、0.03%、0.02%、0.01%、0.005%、0.001%或更低的序列变体充分高于背景,从而允许准确地判定。在一些实施方案中,序列变体以约为或小于约0.1%的频率出现。在一些实施方案中,当序列变体的频率在统计学上显著地高于背景错误率(例如,p值约为或小于约0.05、0.01、0.001、0.0001或更低)时,该频率充分高于背景。在一些实施方案中,当序列变体的频率约为或至少约为背景错误率的2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、25倍、50倍、100倍或更多倍(例如,至少5倍)时,该频率充分高于背景。在一些实施方案中,在精确确定给定位置处的序列时的背景错误率约为或低

于约1%、0.5%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%或更低。在一些实施方案中,错误率低于0.001%。

[0148] 在一些实施方案中,鉴定真正的序列变体(也被称为“判定”或“作出判定”)包括将一个或多个测序读取与参考序列进行最优比对,以鉴定这二者之间的差异,以及鉴定接点。通常,比对包括将一个序列沿另一个序列放置,迭代地沿每一个序列引入缺口,对两个序列匹配的好坏如何进行打分,并且优选地沿着参考序列对各个位置进行重复。具有最佳得分的匹配被视为对准(alignment),并且代表了关于序列间的关系程度的推断。在一些实施方案中,与测序读取进行比较的参考序列是参考基因组,例如与受试者属于同一种的成员的基因组。参考基因组可以是完整的或不完整的。在一些实施方案中,参考基因组仅由包含靶多核苷酸的区域组成,例如该区域源自参考基因组或源自所分析的测序读取生成的共有序列。在一些实施方案中,参考序列包含一个或多个生物体的多核苷酸序列,例如,来自一个或多个细菌、古菌、病毒、原生生物、真菌或其他生物体的序列,或由该序列组成。在一些实施方案中,参考序列仅由参考基因组的一部分,例如与一个或多个所分析的靶序列相对应的区域(例如,一个或多个基因,或其部分)组成。例如,为检测病原体(例如在检测污染的情况下),参考基因组是该病原体(例如HIV、HPV或有害的细菌菌株,如大肠杆菌)的完整基因组,或是其一部分,该部分可用于鉴定,例如鉴定特定的菌株或血清型。在一些实施方案中,将测序读取与多个不同的参考序列进行比对,如用于筛选多种不同的生物体或菌株。

[0149] 在典型的比对中,测序读取中的碱基旁边有参考序列中的不匹配碱基指示在该点发生了置换突变。类似地,当一个序列在另一序列中的碱基旁边包括缺口时,推断发生了插入或缺失突变(“插入/缺失(indel)”)。当希望指明一个序列与另一个彼此比对时,该比对有时被称为配对比对(pairwise alignment)。多序列比对通常是指两个或更多个序列的比对,包括,例如通过一系列配对比对。在一些实施方案中,给比对打分涉及为置换和插入/缺失的可能性设定数值。当单独的碱基进行比对时,匹配或不匹配导致根据置换可能性的比对得分,其可能是,例如,1为匹配而0.33为不匹配。插入/缺失从比对得分中扣除空位罚分,其可以是,例如-1。空位罚分和置换可能性可以基于经验知识或基于关于序列如何突变的先验假设。它们的值影响产生的比对。用于进行比对的算法的实例包括但不限于Smith-Waterman(SW)算法、Needleman-Wunsch(NW)算法、基于Burrows-Wheeler转换的算法(BWT)以及散列函数比对器如Novoalign(Novocraft Technologies;可从www.novocraft.com获得)、ELAND(Illumina, San Diego, Calif)、SOAP(可从soap.genomics.org.cn获得)和Maq(可从maq.sourceforge.net获得)。执行BWT方法的一个示例性比对程序是Burrows-Wheeler比对器(BWA),其可从Geeknet(Fairfax, Va.)维护的SourceForge网站获得。BWT一般每个核苷酸占据2比特的内存,使得用典型的台式或膝上型计算机对长达4G碱基对的核苷酸序列进行索引化(index)成为可能。预处理过程包括BWT的构建(例如,参考序列的索引化)和支持辅助数据结构。BWA包括两种不同的算法,均基于BWT。通过BWT进行的比对能够使用bwa-短(bwa-short)算法进行,该算法是针对高达约200bp的短查询而设计的,具有低误差率(<3%)(Li H.和Durbin R. *Bioinformatics*, 25:1754-60 (2009))。第二个算法, BWA-SW, 是针对长读取而设计的,具有更多的错误(Li H.和Durbin R. (2010). *Fast and accurate long-read alignment with Burrows-Wheeler Transform*. *Bioinformatics*, Epub.)。bwa-sw比对器有时称为“bwa-长”、“bwa长算法”或类似的名称。执行Smith-

Waterman算法版本的一个比对程序是MUMmer,其可从Geeknet (Fairfax, Va.) 维护的SourceForge网站获得。MUMmer是一个用来快速比对完整基因组的系统,无论它是完整的形式还是草图形式(Kurtz, S. 等人, *Genome Biology*, 5:R12 (2004); Delcher, A.L. 等, *Nucl. Acids Res.*, 27:11 (1999))。例如, MUMmer 3.0在2.4GHz Linux台式计算机上,使用78MB的内存,能够在13.7秒内找出在一对5兆碱基基因组之间的全部20-碱基对或更长的精确匹配。MUMmer也可以比对不完整的基因组;它能够容易地处理来自鸟枪法测序项目的100s或1000s的叠连群,并且使用该系统内包含的NUCmer程序将其与另一组叠连群或基因组进行比对。比对程序的其他非限制实例包括:来自Kent Informatics (Santa Cruz, Calif.) 的BLAT (Kent, W.J., *Genome Research* 4:656-664 (2002)); 来自北京基因组研究所(北京, Conn.) 或BGI Americas Corporation (Cambridge, Mass.) 的SOAP2; Bowtie (Langmead等人, *Genome Biology*, 10:R25 (2009)); 高效大规模核苷酸比对数据库 (ELAND) 或序列及变异共同评估 (CASAVA) 软件的ELANDv2组件 (Illumina, San Diego, Calif.); 来自Real Time Genomics, Inc. (San Francisco, Calif.) 的RTG Investigator; 来自Novocraft (Selangor, Malaysia) 的Novoalign; Exonerate, 欧洲生物信息学研究所 (Hinxton, UK) (Slater, G. 和 Birney, E., *BMC Bioinformatics* 6:31 (2005)); Clustal Omega, 来自University College Dublin (Dublin, Ireland) (Sievers F. 等人, *Mol Syst Biol* 7, 文章 539 (2011)); 来自University College Dublin (Dublin, Ireland) 的ClustalW或ClustalX (Larkin M.A. 等人, *Bioinformatics*, 23, 2947-2948 (2007)); 和FASTA, 欧洲生物信息学研究所 (Hinxton, UK) (Pearson W.R. 等人, *PNAS* 85 (8):2444-8 (1988); Lipman, D.J., *Science* 227 (4693):1435-41 (1985))。

[0150] 在图36A-36H中提供了根据一些实施方案,特别是对于采用3' 加尾反应的实施方案的方法的说明。图36A示出了样品的无细胞双链多核苷酸1、2、3……K (101), 其每个无细胞双链多核苷酸含有由单个核苷酸组成的遗传基因座 (100), 该遗传基因座可被“G”或罕见变体“A”占据。含有这类多核苷酸的样品可以是患者组织样品,如血液或血浆样品等。通常,参考序列(例如在人基因组数据库中)可用于比较多核苷酸序列。每种多核苷酸具有四个对应于每个末端的两条互补链的序列的序列区。因此,例如,图36A的靶多核苷酸1在链的每个末端具有序列区n1 (110) 和n2 (112), 并且在互补链 (120) 的末端具有互补序列区n1' (116) 和n2' (108)。尽管各种多核苷酸链的序列区被图示为链的小部分,但是序列区可以包含从链末端到遗传基因座 (100) 的整个区段。

[0151] 在一些实施方案中,向样品的靶多核苷酸添加3' 加尾活性以及核酸单体和/或其他反应组分以实施用一个或多个A延伸3' 端的加尾反应 (125)。在该实施方案中,预定核苷酸的延伸显示为“A……A”以指示添加一个或多个核苷酸,但是添加至每条链的确切数目可能是不确定的(除非使用外切聚合酶,如下文所列)。通过“A……A”表示添加的核苷酸并非旨在将添加的核苷酸的种类仅限制为A。添加的核苷酸是预定的,在某种意义上加尾反应中使用的核苷酸前体的种类是已知的并且被选择作为试验设计选择。例如,考虑到所选择的核苷酸的种类,选择特定实施方案的一种预定核苷酸的因素可以是环化步骤的效率。在一些实施方案中,核苷酸前体可以是四种核苷酸中的任何一种的核苷三磷酸,或者单独地,以便产生均聚物尾,或者以混合物形式,以便产生双核苷酸或三核苷酸尾。在一些情况下,除了四种天然DNA碱基之外或代替四种天然DNA碱基,可以使用尿嘧啶和/或核苷酸类似物。在

使用CircLigase<sup>TM</sup>酶的一些实施方案中,预定核苷酸可以是A和/或T。在一些实施方案中,外切聚合酶用于加尾反应,并且仅将单个脱氧腺苷酸添加至3'端。

[0152] 在加尾和任选地从反应混合物中分离反应产物之后,使用环化反应将单独的链环化,如图36B所示,以产生环(132),每个环包含形式为“ $n_j-A \cdots A-n_{j+1}$ ”(133)的序列元件。在环化和任选地从反应混合物中分离环(132)之后,将引物(134)退火至环(132)的一种或多种引物结合位点,之后将它们延伸以产生每个包含它们相应的 $n_j-A \cdots A-n_{j+1}$ 序列元件的拷贝的多联体,如图36E所示。在测序之后,可以通过将序列元件组分 $n_j$ 和 $n_{j+1}$ 与它们相应的互补体 $n_j'$ 和 $n_{j+1}'$ 进行匹配来鉴定互补链,如(136)和(138)。在环(132)上选择引物结合位点是设计选择的问题,或者可以使用随机序列引物。在一些实施方案中,选择与遗传基因座(100)相邻的单个引物结合位点;在其他实施方案中,选择多个引物结合位点,每种引物结合位点用于单独的引物,以确保扩增,即使在一个引物结合位点中发生边界。在一些实施方案中,具有单独的引物结合位点的两个引物用于产生多联体。

[0153] 在鉴定含有互补链的多联体对之后,可以比对多联体序列并且可以比较两条链的匹配位置处的碱基判定。在多联体对的一些位置处,如图36F中的(140)所示,在一对中的一个成员中的给定位置处判定的碱基可能不与在该对中的另一个成员上判定的碱基互补,这表明由于例如扩增错误、测序错误等进行了错误判定。在这种情况下,可以通过检查在多联体对内的其他拷贝的相应位置处的碱基判定来解决给定位置处的不确定性。例如,给定位置处的碱基判定可以被认为是对一对多联体中的单独的拷贝进行的碱基判定的共识或多数。用于进行此类测定的其他方法对于本领域普通技术人员而言是可用的,其可以代替这些方法或除了这些方法之外使用,以补充当互补链之间的序列信息不互补时解决碱基判定的工作。在一些情况下,在源自相同双链分子的互补链中特定位置的碱基(例如,如由3'和5'端序列所鉴定的)不是互补的,解决碱基判定有利于比较样品序列的参考序列,使得差异不被鉴定为关于这样的参考序列的真实序列变体。

[0154] 在其他情况下,相同的错误可出现在多联体内的靶多核苷酸的每个拷贝中,如图36G中的(145)所示。这样的数据表明靶多核苷酸在扩增或测序之前被破坏。

[0155] 在其他情况下,可以仅鉴定单个多联体;即,基于边界信息找不到匹配的多联体,例如,预定核苷酸区段的长度,相邻3'和5'端的序列等。这类情况在图37A和图37B中示出。其中,靶多核苷酸(201)包含单链多核苷酸1和双链多核苷酸2,每个均包含遗传基因座(200)。预定核苷酸(例如,腺苷酸)可以在加尾反应(225)中附接至多核苷酸1和2两者,以形成3'加尾的多核苷酸(220)。如上所述,然后将多核苷酸(220)环化,通过RCA扩增,并对其进行测序以得到多联体序列(230),如图37B所示。如果观察到的变异在DNA损伤中很常见,例如C到T或G到T,来自未配对的多联体的这样的信息仍然有助于确定它是否是真正的突变而不仅仅是DNA损伤。

[0156] 在一些实施方案中,如图36C和图36D所示,可将每个含有分子标签的引物(例如MT1(150)、MT2等)退火至预定引物结合位点,以便产生每个具有独特标签的多联体。独特分子标签的存在将区分碰巧具有相同边界的单链环的产物,或 $n_j-A \cdots A-n_{j+1}$ 序列元件。这样的标签也可用于对分子进行计数以确定遗传基因座处的拷贝数变异,例如,根据在通过引用并入本文的Brenner等人,美国专利7,537,897等中描述的方法。在一些实施方案中,可以选择具有分子标签的引物,其仅在靶多核苷酸的一条链上具有结合位点,使得具有分子标

签的多联体仅代表靶多核苷酸的两条链中的一条。在其他实施方案中,来自靶多核苷酸的互补链的环可各自使用具有分子标签的引物扩增(如图36C中所示)。

[0157] 在一些实施方案中,用于鉴定靶多核苷酸的互补链的上述步骤可以并入用于检测遗传基因座处的罕见变体的方法中。在一些实施方案中,所述方法包括以下步骤:(a)通过多核苷酸的一个或多个预定核苷酸3'端延伸;(b)使多核苷酸的单独的链环化以形成单链多核苷酸环,所述一个或多个预定核苷酸限定每个单链多核苷酸环的3'序列与5'序列之间的边界;(c)通过滚环复制(RCR)扩增单链多核苷酸环以形成多联体;(d)对多联体进行测序;(e)通过与一个或多个预定核苷酸相邻的3'序列和5'序列的同一性鉴定含有多核苷酸的互补链的多联体对;以及(f)从包含相同多核苷酸的互补链的所述多联体对的序列确定遗传基因座的序列。在其他实施方案中,通过RCR扩增单链环的步骤包括将具有5'-非互补加尾的引物退火至单链环,其中这样的引物在5'-非互补尾中包含独特分子标签并且根据RCR方案延伸这样的产物。得到的产物是含有独特分子标签的多联体,其可以与附接至来自相同基因座的环的其他分子标签一起计数,以提供基因座的拷贝数测量。

[0158] 在一些实施方案中,延伸步骤可以通过在加尾反应中通过多核苷酸的一个或多个预定核苷酸3'端加尾来实施。在一些实施方案中,这样的加尾可以通过未模板化的3'核苷酸添加活性(如TdT活性、外切聚合酶活性等)来实施。

[0159] 使用上述步骤,可以从多核苷酸序列中鉴定多联体序列。在大规模平行测序(也称为“下一代测序”或NGS)中,可以鉴定含有多联体的读取并用于进行错误校正和发现序列变体。原始输入分子的接点(DNA/RNA序列的起始和结束)可以通过将其与参考序列比对而从多联体重建;并且接点可用于鉴定原始输入分子并去除测序重复以进行更准确的计数。可以通过将读取与参考序列比对并检查如图36A中所述的序列元件组分 $n_j$ 和 $n_{j+1}$ 来计算可包含多联体的每个读取的链同一性。在标记为互补链的两个多联体中发现的变体具有更高的统计置信水平,其可用于进行进一步的错误校正。使用链同一性的变体确认可以通过(但不限于)以下步骤进行:a)在具有互补链同一性的读取中发现的变体被认为是更确信的;b)携带变体的读取可通过其接点鉴定进行分组,当在具有相同接点鉴定的一组读取的读取中发现互补链同一性时,该变体被认为是更确信的;c)携带变体的读取可通过它们的分子条形码或分子条形码和接点鉴定的组合进行分组。当在具有相同分子条形码和/或接点鉴定的一组读取的读取中发现互补链同一性时,变体被认为是更确信的。

[0160] 采用分子条形码和接点鉴定的错误校正可以单独地或者与先前步骤中描述的多联体测序的错误校正组合使用。a)可将具有不同分子条形码(或接点鉴定)的读取分组成不同的读取家族中,该读取家族代表源自不同输入分子的读取;b)可从读取家族构建共有序列;c)可使用共有序列进行变体判定;d)可结合分子条形码和接点鉴定以形成读取的复合ID,其将有助于鉴定原始输入分子。在一些实施方案中,在不同的读取家族中发现的碱基判定(例如,相对于参考序列的序列差异)被赋予更高的置信度。在一些情况下,如果序列差异通过一个或多个增加碱基判定置信度的过滤器,则序列差异仅被鉴定为代表起始源多核苷酸的真实序列变体(与样品处理或分析的错误相反),如上文描述的那些。在一些实施方案中,如果(a)序列差异在双链输入分子的两条链上鉴定出,(b)序列差异出现在其起源的多联体的共有序列中(例如,在包含序列差异的多联体内大于50%、80%、90%或更多的重复);和/或(c)序列差异出现在两种不同的分子中(例如,如由不同的3'和5'端点和/或由外

源标签序列所鉴定的), 则序列差异仅被鉴定为真正的序列变体。

[0161] 确定链同一性: 1) 通过将序列与参考序列比对, 可从可能含有多联体序列的读取重建原始输入分子的接点; 2) 可使用比对来定位读取中的接点; 3) 可从基于读取中的接点位置的序列提取如图36A所示的序列元件组分 $n_j$ 和 $n_{j+1}$ , 其代表链同一性; 在多联体的情况下, 可以在多联体序列的接点之间找到序列; 4) 可使用读取与之比对的参考序列的链(阳性或阴性), 与步骤3中鉴定的读取内的链同一性序列组合, 来鉴定掺入序列文库并对其测序的原始链, 并鉴定序列变体起源于哪条链。例如, 假设将链同一性序列“AA”添加至原始输入DNA片段链的末端; 测序后, 将DNA片段的读取与参照的“+”链进行比对, 并且读取中的链同一性序列为“AA”, 我们知道原始输入链是“+”; 如果链同一性序列为“TT”, 则读取与原始输入链反向互补, 并且原始输入链为“-”链。链同一性测定允许将序列变体与其反向互补对应物区分开, 例如, 将C>T置换与G>A置换区分开。等位基因变化的精确鉴定可用于在变体判定中进行等位基因特异性错误减少。例如, 当某些等位基因改变时一些DNA损伤更频繁地发生, 并且可以进行等位基因特异性错误减少以抑制这种损伤; 这样的错误减少可通过各种统计方法来完成, 例如, 1) 计算测序数据(基线)中不同等位基因变化的分布, 随后进行2) z-检验或其他统计学测试以确定观察到的等位基因变化是否不同于基线分布。

[0162] 在一些实施方案中, 本公开内容提供了一种通过将测量的序列或一个或多个核苷酸的频率与导致相同序列或一个或多个核苷酸(如测量的序列)的核苷酸损伤的基线频率进行比较来鉴定遗传基因座处的特定链上的遗传变体的方法。在一些实施方案中, 这样的方法可包括以下步骤: (a) 通过多核苷酸的一个或多个预定核苷酸3'端延伸; (b) 扩增延伸的多核苷酸的单独的链; (c) 对延伸的多核苷酸的扩增的单独的链进行测序; (d) 通过与一个或多个预定核苷酸相邻的3'序列和/或5'序列的同一性鉴定多核苷酸的互补链, 并且鉴定遗传基因座处的每条链的核苷酸; (e) 从所鉴定的多联体中确定遗传基因座处的一个或多个核苷酸中的每一个的频率, 以用于鉴定遗传变体。在一些实施方案中, 该方法可用于通过以下步骤将遗传变体与核苷酸损伤区分开: 每当显示至少一个核苷酸的链的所述频率超过预定因子的具有产生相同核苷酸的核苷酸损伤的链的基线频率时, 将由所述一个或多个预定核苷酸鉴定的所述链上的所述遗传基因座处的所述一个或多个核苷酸中的至少一个判定为所述遗传变体。

[0163] 如上所述, 在一些实施方案中, 扩增步骤可通过以下进行: (i) 使多核苷酸的单独的链环化以形成单链多核苷酸环, 一个或多个预定核苷酸限定每个单链多核苷酸环中的多核苷酸的3'序列与5'序列之间的边界; 以及(ii) 通过滚环复制扩增单链多核苷酸以形成单链多核苷酸环的多联体。

[0164] 具有核苷酸损伤的链的基线频率可以基于对来自通过该方法测试的相同个体的样品的先前测量, 或者基线频率可以基于对除所测试的个体之外的个体群体的先前测量。基线频率还可以取决于和/或特定于在通过本公开内容的方法制备用于分析的样品中使用的步骤或方案的种类。通过将测量的频率与基线频率进行比较, 可以获得测量或确定的序列是真正的遗传变体而不是由于处理导致的损伤或错误的可能性(或置信水平)的统计测量。

[0165] 一般而言, 测序数据从大规模平行测序反应获得。许多新一代高通量测序系统将数据输出为FASTQ文件, 但是也可使用其他格式。在一些实施方案中, 一般通过序列比对, 分

析序列以鉴定重复单元长度(例如单体长度)、通过环化形成的接点以及任何相对于参考序列的真正变异。鉴定重复单元长度可以包括计算重复单元的区域,找出序列的参考基因座(例如,当一个或多个序列尤其以扩增、富集和/或测序为目标时)、每一个重复区域的边界和/或每次测序运行中的重复序列的数目。序列分析可以包括分析双链体的两条链的序列数据。如上所述,在一些实施方案中,出现来自样品的不同多核苷酸(例如具有不同接点的环化多核苷酸)的读取序列的相同变体被视为是确认的变体。在一些实施方案中,如果序列变体在相同多核苷酸的多于一个重复单元中发生,则该序列变体也可被认为是确认的或真正的变体,因为相同的序列变体同样不可能在相同多联体内的重复靶序列的相同位置上发生。序列的质量得分可在鉴定变体和确认变体时考虑,例如,可以过滤掉质量得分低于阈值的序列和碱基。可以使用其他生物信息学方法进一步提高变体判定的灵敏性和特异性。

[0166] 在一些实施方案中,统计分析可用于变体(突变)的确定并量化变体在总DNA样品中的比率。特定碱基的总测量值可以使用测序数据计算。例如,根据先前步骤中计算的比对结果,可以计算“有效读取”的数目,即每个基因座的确认读取的数目。变体的等位基因频率可以通过基因座的有效读取计数来标准化。可以计算总噪声水平,即所有基因座上观测到的变体的平均比率。变体的频率和总噪声水平以及其他因素可用于确定变体判定的置信区间。诸如泊松分布等统计模型可用于评估变体判定的置信区间。变体的等位基因频率也可用作总样品中变体相对量的指标。

[0167] 在一些实施方案中,基于判定步骤鉴定微生物污染物。例如,特定的序列变体可以指示潜在感染性微生物的污染。序列变体可以在高度保守的多核苷酸内鉴定,以鉴定微生物。用于系统发育表征和微生物鉴定的示例性高度保守的多核苷酸包括在16S rRNA基因、23S rRNA基因、5S rRNA基因、5.8S rRNA基因、12S rRNA基因、18S rRNA基因、28S rRNA基因、gyrB基因、rpoB基因、fusA基因、recA基因、cox1基因和nifD基因中发现的核苷酸序列。对于真核生物,rRNA基因可以是核rRNA基因、线粒体rRNA基因或两者。在一些实施方案中,16S-23S rRNA基因内转录间隔区(ITS)中的序列变体可用于在使用或未使用其他rRNA基因的情况下区分和鉴定密切相关的分类群。由于16S rRNA的结构限制,整个基因的特定区域具有高度保守的多核苷酸序列,尽管非结构区段可具有高度可变性。鉴定序列变体可用于鉴定代表亚属、属、亚科、科、亚目、目、亚类、类、亚门、门、亚界或界的操作分类单位(OTU),并可选地确定它们在群体中的频率。特定序列变体的检测可用于检测指示污染的微生物的存在和任选的量(相对量或绝对量)。示例应用包括针对粪便或其他污染物的水质检测,检测动物或人类病原体,定位水污染源,检测再生水或循环水,检测包括海洋排放羽流在内的污水排放流,监测水产养殖设施的病原体,监测海滩、游泳区域或其他与水有关的娱乐设施,以及预测有毒水华。食品监测应用包括针对食源性病原体定期检测食品加工厂的生产线,调查屠宰场,检查餐馆、医院、学校、惩戒设施和其他机构的厨房和食品储存区,该食源性病原体是例如大肠杆菌菌株0157:H7或0111:B4,单核细胞增生利斯特氏菌(*Listeria monocytogenes*)或肠沙门氏菌肠亚种肠炎血清变种(*Salmonella enterica* subsp.*enterica* serovar *Enteritidis*)。可以调查贝类和贝类生产水域中造成麻痹性贝毒、神经毒性贝毒、痢性贝毒和记忆丧失性贝毒的藻类。此外,可在海关放行前筛查进口食品,以确保食品安全。植物病原体监测应用包括园艺和苗圃监测,例如监测栎树猝死疫霉(*Phytophthora ramorum*)——导致栎树猝死病的微生物,作物病原体监测与病害管理,以



及林业病原体监测与病害管理。可以调查微生物污染是重大安全问题的药物、医疗设备和其他消耗品或关键组件的制造环境中特定病原体如铜绿假单胞菌 (*Pseudomonas aeruginosa*) 或金黄色葡萄球菌 (*Staphylococcus aureus*) 的存在,与人类相关的更多常见微生物、与水相关的微生物或表现出先前在该特定环境或类似环境中鉴定的生物负荷的微生物的存在。类似地,可以监测包括航天器在内的敏感设备的构造和组装区域中已知栖息至或最常引入至这种环境中的先前鉴定的微生物。

[0168] 在一些实施方案中,该方法包括鉴定包含少于50ng多核苷酸的核酸样品中的序列变体,每种多核苷酸具有5'端和3'端。在一些实施方案中,该方法包括:(a)用连接酶环化所述样品中的个体多核苷酸以形成多个环状多核苷酸;(b)在将所述连接酶与所述环状多核苷酸分离后,扩增环状多核苷酸以形成多联体;(c)对多联体进行测序以产生多个测序读取;(d)鉴定多个测序读取与参考序列之间的序列差异;以及(e)在来自少于50ng多核苷酸的所述核酸样品的所述多个读取中将以0.05%或更高的频率出现的序列差异判定为序列变体。

[0169] 样品中多核苷酸的原始量可以很小。在一些实施方案中,原始多核苷酸的量小于100ng。在一些实施方案中,原始材料的量小于75ng。在一些实施方案中,原始材料的量小于50ng,如小于45ng、40ng、35ng、30ng、25ng、20ng、15ng、10ng、5ng、4ng、3ng、2ng、1ng、0.5ng、0.1ng或更低。在一些实施方案中,原始多核苷酸的量在0.1-100ng的范围内,如1-75ng、5-50ng或10-20ng。通常,较低的原始材料增加了从各种处理步骤中提高回收率的重要性。减少样品中参与后续反应的多核苷酸方法降低了可检测到罕见突变的灵敏度。例如,Lou等人,(PNAS,2013,110(49))描述的方法预计仅能回收原始材料的10-20%。对于大量的原始材料(例如从实验室培养的细菌中纯化),这可能不是实质性障碍。然而,对于原始材料显著较低的样品,在该低范围内的回收可能是检测足够罕见变体的实质性障碍。因此,在一些实施方案中,本公开内容的方法中,样品在从一个步骤回收至另一个步骤,例如,输入到可用于输入后续扩增步骤(或测序步骤)中的环化步骤的质量分数约或大于约50%、60%、75%、80%、85%、90%、95%或更多。从特定步骤的回收可能接近100%。回收可以关于特定形式,如从非环状多核苷酸的输入中回收环状多核苷酸。

[0170] 多核苷酸可以来自任何合适的样品,如本文关于本公开内容的多个方面描述的样品。来自样品的多核苷酸可以是多种多核苷酸中的任意种,包括但不限于,DNA、RNA、核糖体RNA(rRNA)、转移RNA(tRNA)、微小RNA(miRNA)、信使RNA(mRNA),其中任一种的片段,或其中任两种或更多种的组合。在一些实施方案中,样品包含DNA。在一些实施方案中,多核苷酸在获得时是单链,或者通过处理(例如变性)成为单链。本文描述了合适的多核苷酸的其他实例,如关于本公开内容的多个方面中的任一方面所述。在一些实施方案中,多核苷酸在没有提取步骤和/或没有纯化步骤的情况下经历后续步骤(例如环化和扩增)。例如,可以处理流体样品以在没有提取步骤的情况下除去细胞,从而产生纯化的液体样品和细胞样品,然后从纯化的流体样品中分离DNA。有多种分离多核苷酸的方法可用,例如通过沉淀,或与基底非特异性结合,然后洗涤该基底以释放结合的多核苷酸。在没有细胞提取步骤的情况下从样品中分离多核苷酸时,多核苷酸将主要是细胞外或“无细胞”多核苷酸,如无细胞DNA和无细胞RNA,其可以对应于死亡或受损的细胞。这些细胞的特性可用于表征其所源自的细胞或细胞群,如微生物群落中。如果处理样品以提取多核苷酸,例如从样品中的细胞提取多核苷



酸,有多种提取方法可用,本文提供了其中的实例(例如关于本公开内容的多个方面中的任一方面)。

[0171] 核酸样品中的序列变体可以是多种序列变体中的任何一种。本文描述了序列变体的多个非限制性实例,如关于本公开内容的对个方面中的任一方面所述。在一些实施方案中,序列变体是单核苷酸多态性(SNP)。在一些实施方案中,序列变体在群体中以低频率出现(也称为“罕见”序列变体)。例如,序列变体可以以约或低于约5%、4%、3%、2%、1.5%、1%、0.75%、0.5%、0.25%、0.1%、0.075%、0.05%、0.04%、0.03%、0.02%、0.01%、0.005%、0.001%或更低的频率出现。在一些实施方案中,序列变体以约或低于约0.1%的频率出现。

[0172] 根据一些实施方案,样品的多核苷酸被环化,如通过使用连接酶。环化可包括将多核苷酸的5'端接合到同一多核苷酸的3'端,接合到该样品中的另一多核苷酸的3'端,或接合到来自不同来源的多核苷酸(例如,人工多核苷酸,如寡核苷酸衔接子)的3'端。在一些实施方案中,多核苷酸的5'端接合到同一多核苷酸的3'端(也称为“自接合”)。本文提供了环化过程(例如,使用和不使用衔接子寡核苷酸)、试剂(例如衔接子的类型、连接酶的使用)、反应条件(例如有利于自接合)和任选的额外加工(例如反应后纯化)的非限制性实例,如关于本公开内容的多个方面中的任一方面所述。

[0173] 如前所述,将多核苷酸的末端彼此接合以形成环状多核苷酸(直接地,或者使用一个或多个居间的衔接子寡核苷酸)通常产生具有接合序列的接点。当多核苷酸的5'端和3'端通过衔接子多核苷酸接合时,术语“接点”指该多核苷酸与衔接子之间的接点(例如,5'端接点或3'端接点之一),或指多核苷酸的5'端和3'端之间由此形成且包含衔接子多核苷酸的接点。当多核苷酸的5'端和3'端在不使用间插衔接子的情况下接合时(例如,单链DNA的5'端和3'端),术语“接点”可以指这两个末端相接合的点。接点可通过包含接点的核苷酸的序列(也被称为“接合序列”)进行鉴定。在一些实施方案中,样品包含具有通过以下过程形成的末端混合物的多核苷酸:自然降解过程(如细胞裂解、细胞死亡和DNA从细胞释放到其周围环境的其他过程,DNA在该周围环境中可进一步降解,如在不细胞多核苷酸中,如无细胞DNA)、作为样品处理的副产物的片段化(诸如固定、染色和/或储存过程),以及通过切割不限于特定靶序列的DNA的方法进行的片段化(例如,机械片段化,如通过超声处理;非序列特异性核酸酶处理,如DNA酶I、片段化酶)。当样品包含具有末端混合物的多核苷酸时,两个多核苷酸具有相同的5'端或3'端的可能性很低,并且两个多核苷酸独立地具有相同的5'端和3'端两者的可能性极低。因此,在一些实施方案中,甚至在两个多核苷酸包含具有相同靶序列的部分时,可以使用接点来区分不同的多核苷酸。当多核苷酸末端在不使用间插衔接子的情况下连接时,接合序列可通过与参考序列进行比对来鉴定。例如,当两种组分序列的顺序相对于参考序列似乎反转时,显示发生反转的点可以指示在该点处的接点。当多核苷酸末端通过一个或多个衔接子序列连接时,接点可通过与已知衔接子序列的邻近进行鉴定,或者在测序读取的长度足以从环化多核苷酸的5'和3'端获得序列的情况下,则通过如上所述的比对进行鉴定。在一些实施方案中,特定接点的形成是十分罕见的事件,使得其在样品的环化多核苷酸之间是独特的。

[0174] 环化之后,反应产物可在扩增或测序之前进行纯化以提高可参与后续步骤的环化多核苷酸的相对浓度或纯度(例如,通过环状多核苷酸的分离或反应中一种或多种其他分

子的去除)。例如,可处理环化反应或其组分以去除单链(未环化的)多核苷酸,例如通过外切核酸酶处理。作为进一步的实例,环化反应或其部分可进行大小排阻色谱法,借此保留及丢弃小试剂(例如未反应的衔接子),或在单独的体积中保留并释放环化产物。多种用于清理连接反应的试剂盒是可用的,例如由Zymo Research制造的Zymo寡核苷酸纯化试剂盒所提供的试剂盒。在一些实施方案中,纯化包括用于去除或降解在环化反应中使用的连接酶和/或将环化多核苷酸从该连接酶中纯化的处理。在一些实施方案中,用于降解连接酶的处理包括用蛋白酶进行的处理。合适的蛋白酶可从原核生物、病毒和真核生物中获得。蛋白酶的实例包括蛋白酶K(来自白色念珠菌(*Tritirachium album*))、链霉菌蛋白酶E(来自灰色链霉菌(*Streptomyces griseus*))、多粘芽孢杆菌(*Bacillus polymyxa*)蛋白酶、嗜热菌蛋白酶(来自嗜热菌)、胰蛋白酶、枯草杆菌蛋白酶、弗林蛋白酶等。在一些实施方案中,蛋白酶是蛋白酶K。蛋白酶处理可遵循制造商的方案,或经历标准条件(例如Sambrook和Green, *Molecular Cloning: A Laboratory Manual*, 第4版(2012)所提供的)。蛋白酶处理之后还可进行提取和沉淀。在一个实例中,环化多核苷酸如下纯化:在0.1% SDS和20mM EDTA的存在下进行蛋白酶K(Qiagen)处理,用1:1苯酚/氯仿和氯仿抽提,并用乙醇或异丙醇沉淀。在一些实施方案中,沉淀在乙醇中进行。

[0175] 如关于本公开内容的其他方面所述,环化之后可直接对环化的多核苷酸进行测序。或者,在测序之前可以进行一个或多个扩增反应。可以使用多种扩增多核苷酸(例如DNA和/或RNA)的方法。扩增可以是线性的、指数的或在多相扩增过程中涉及线性和指数两相。扩增方法可以涉及温度变化,例如热变性步骤,或者可以是不需要热变性的等温过程。本文描述了合适的扩增过程的非限制性实例,如关于本公开内容的多个方面中的任一方面的所述。在一些实施方案中,扩增包括滚环扩增(RCA)。如本文其他地方所述,典型的RCA反应混合物包含一种或多种引物、聚合酶和dNTP,并且产生多联体。一般来说,RCA反应中的聚合酶是具有链置换活性的聚合酶。多种这样的聚合酶是可用的,其非限制性实例包括外切核酸酶DNA聚合酶I大(Klenow)片段、Phi29 DNA聚合酶、Taq DNA聚合酶等。通常,多联体是多核苷酸扩增产物,其包含来自模板多核苷酸的靶序列的两个或更多个拷贝(例如,靶序列的约或大于约2、3、4、5、6、7、8、9、10或更多个拷贝;在一些实施方案中,约或大于约2个拷贝)。扩增引物可具有任意合适的长度,诸如约或至少约5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100个或更多个核苷酸,其任何部分或全部可与该引物所杂交的相应靶序列互补(例如,约或至少约5、10、15、20、25、30、35、40、45、50或更多个核苷酸)。本文描述了各种RCA方法的实例,如使用随机引物、靶特异性引物和衔接子靶向引物,其中一些如图7A-7C所示。

[0176] 在测序之前将环化的多核苷酸扩增(例如以产生多联体)的情况下,可以在未富集的情况下直接使扩增产物经历测序,或者在一个或多个富集步骤之后使其经历测序。本文描述了合适富集过程的非限制性实例,如关于本公开内容的各个方面中的任一方面所述(例如,使用B2B引物进行第二扩增步骤)。根据一些实施方案,使环化的多核苷酸(或其扩增产物,可任选地经富集)经历测序反应以产生测序读取。这些方法产生的测序读取可以根据本文公开的其他方法使用。多种测序方法是可用的,尤其是高通量测序方法。实例包括但不限于Illumina制造的测序系统(诸如HiSeq®和MiSeq®的测序系统)、Life Technologies制造的测序系统(Ion Torrent®、SOLiD®等)、Roche的454Life

Sciences系统、Pacific Biosciences系统等。在一些实施方案中,测序包括使用HiSeq®和MiSeq®系统产生长度约为或多于约50、75、100、125、150、175、200、250、300个或更多个核苷酸的读取。本文描述了扩增平台和方法的其他非限制性实例,如关于本公开内容的各个方面中的任一方面所述。

[0177] 根据一些实施方案,如果测序读取与参考序列之间的序列差异出现在至少两种不同的多核苷酸(例如,两种不同的环状多核苷酸,其可以由于具有不同的接点而被区分,或者具有不同的5'端和/或3'端的两种不同多核苷酸)中,则将该序列差异判定为真正的序列变体(例如,存在于扩增或测序前的样品中,并且不是这些过程中的任何一个过程的结果)。因为作为扩增或测序错误的结果的序列变体不可能在包含相同靶序列的两种不同多核苷酸上准确地重复(例如,位置和类型),所以添加该验证参数大大减少了错误序列变体的背景,同时伴有检测样品中的实际序列变异的灵敏度和准确性的提高。在一些实施方案中,频率约为或低于约5%、4%、3%、2%、1.5%、1%、0.75%、0.5%、0.25%、0.1%、0.075%、0.05%、0.04%、0.03%、0.02%、0.01%、0.005%、0.001%或更低的序列变体充分高于背景,从而允许准确地判定。在一些实施方案中,序列变体以约为或小于约0.1%的频率出现。在一些实施方案中,该方法包括判定为真正的序列变体,那些序列差异的频率范围为约0.0005%至约3%,如0.001%-2%或0.01%-1%。在一些实施方案中,当序列变体的频率在统计学上显著地高于背景错误率(例如,p值约为或小于约0.05、0.01、0.001、0.0001或更低)时,该频率充分高于背景。在一些实施方案中,当序列变体的频率约为或至少约为背景错误率的2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、25倍、50倍、100倍或更多倍(例如,至少5倍)时,该频率充分高于背景。在一些实施方案中,在精确确定给定位置处的序列时的背景错误率约为或低于约1%、0.5%、0.1%、0.05%、0.01%、0.005%、0.001%、0.0005%或更低。在一些实施方案中,错误率低于0.001%。本文描述了用于确定频率和错误率的方法,如关于本公开内容的各个方面中的任一方面所述。

[0178] 在一些实施方案中,鉴定真正的序列变体(也被称为“判定”或“作出判定”)包括将一个或多个测序读取与参考序列进行最优比对,以鉴定这二者之间的差异,以及鉴定接点。通常,比对包括将一个序列沿另一个序列放置,迭代地沿每一个序列引入缺口,对两个序列匹配的好坏如何进行打分,并且优选地沿着参考序列对各个位置进行重复。具有最佳得分的匹配被视为对准,并且代表了关于序列间的关系程度的推断。多种比对算法和实现它们的比对器是可用的,本文描述了其非限制性实例,如关于本公开内容的多个方面中的任一方面所述。在一些实施方案中,与测序读取进行比较的参考序列是已知的参考序列,如参考基因组(例如与受试者属于相同物种的成员的基因组)。参考基因组可以是完整的或不完整的。在一些实施方案中,参考基因组仅由包含靶多核苷酸的区域组成,例如该区域源自参考基因组或源自所分析的测序读取生成的共有序列。在一些实施方案中,参考序列包含一个或多个生物体的多核苷酸序列,例如,来自一个或多个细菌、古菌、病毒、原生生物、真菌或其他生物体的序列,或由该序列组成。在一些实施方案中,参考序列仅由参考基因组的一部分,例如与一个或多个所分析的靶序列相对应的区域(例如,一个或多个基因,或其部分)组成。例如,为检测病原体(例如在检测污染的情况下),参考基因组是该病原体(例如HIV、HPV或有害的细菌菌株,如大肠杆菌)的完整基因组,或是其一部分,该部分可用于鉴定,例如鉴

定特定的菌株或血清型。在一些实施方案中,将测序读取与多个不同的参考序列进行比对,如用于筛选多种不同的生物体或菌株。本文描述了与之相比可鉴定序列差异(以及判定序列变体)的参考序列的其他非限制性实例,如关于本公开内容的多个方面中的任一方面所述。

[0179] 在一方面,本公开内容提供了在反应混合物中扩增包含两个或更多个靶序列拷贝的多种不同多联体的方法,其中靶序列包含以5'至3'方向取向的序列A和序列B。在一些实施方案中,该方法包括使反应混合物经历核酸扩增反应,其中反应混合物包含:(a)多种多联体,其中多种多联体中的个体多联体包含通过将具有5'端和3'端的个体多核苷酸环化而形成的不同接点;(b)包含序列A'的第一引物,其中第一引物通过序列A与序列A'之间的序列互补性与靶序列的序列A特异性杂交;(c)包含序列B的第二引物,其中第二引物通过序列B与B'之间的序列互补性与包含靶序列的互补序列的互补多核苷酸中存在的序列B'特异性杂交;以及(d)延伸第一引物和第二引物以产生扩增的多核苷酸的聚合酶;其中靶序列中序列A的5'端与序列B的3'端之间的距离为75nt或更小。

[0180] 在相关方面,本公开内容提供了在反应混合物中扩增包含靶序列的多种不同环状多核苷酸的方法,其中靶序列包含以5'至3'方向取向的序列A和序列B。在一些实施方案中,该方法包括使反应混合物经历核酸扩增反应,其中反应混合物包含:(a)多种环状多核苷酸,其中多种环状多核苷酸中的个体多核苷酸包含通过将具有5'端和3'端的个体多核苷酸环化而形成的不同接点;(b)包含序列A'的第一引物,其中第一引物通过序列A与序列A'之间的序列互补性与靶序列的序列A特异性杂交;(c)包含序列B的第二引物,其中第二引物通过序列B与B'之间的序列互补性与包含靶序列的互补序列的互补多核苷酸中存在的序列B'特异性杂交;以及(d)延伸第一引物和第二引物以产生扩增的多核苷酸的聚合酶;其中序列A和序列B是内源序列,并且靶序列中序列A的5'端与序列B的3'端之间的距离为75nt或更小。

[0181] 无论扩增环状多核苷酸还是多联体,这样的多核苷酸都可以来自任何合适的样品来源(直接地或间接地,如通过扩增)。本文描述了多种合适的样品来源、任选的提取方法、多核苷酸的类型与序列变体的类型,如关于本公开内容的多个方面中的任一方面所述。环状多核苷酸可以衍生自将非环状多核苷酸环化。本文中提供了环化过程(例如,使用和不使用衔接子寡核苷酸)、试剂(例如衔接子的类型、连接酶的使用)、反应条件(例如有利于自接合)、任选的额外加工(例如反应后纯化),和由此形成的接点的非限制性实例,如关于本公开内容的多个方面中的任一方面所述。多联体可以衍生自环状多核苷酸的扩增。有多种扩增多核苷酸(例如DNA和/或RNA)的方法可用,本文还描述了其非限制性的实例。在一些实施方案中,通过环状多核苷酸的滚环扩增产生多联体。

[0182] 图10示出了在单重复(除非为环状,否则通常不会被扩增)和包含多个拷贝的靶序列的多联体的情况下,第一和第二引物相对于靶序列的示例性排列。如关于本文所述的其他方面所记载,引物的这种排列可以称为“背对背”(B2B)或“倒置”引物。用B2B引物扩增有助于富集环状和/或串联模板。此外,这种取向结合相对较小的足迹(由一对引物跨越的总距离)允许在靶序列周围的更多种片段化事件的扩增,因为相比于出现在典型的扩增反应(彼此面对,跨越靶序列)中发现的引物排列中,接合出现在引物之间的可能性更小。在一些实施方案中,序列A的5'端与序列B的3'端之间的距离为约或小于约200、150、100、75、50、

40、30、25、20、15或更少的核苷酸。在一些实施方案中,序列A是序列B的互补序列。在一些实施方案中,针对多个不同靶序列的多对B2B引物用于同一反应中,以平行扩增多个不同的靶序列(例如,约或至少约10、50、100、150、200、250、300、400、500、1000、2500、5000、10000、15000或更多不同的靶序列)。引物可具有任何合适的长度,如本文其他地方所述。扩增可以包括在适当条件下的任何合适的扩增反应,如本文所述的扩增反应。在一些实施方案中,扩增是聚合酶链反应。

[0183] 在一些实施方案中,B2B引物包含至少两个序列元件,通过序列互补性与靶序列杂交的第一元件,和在第一元件进行杂交的第一杂交温度下的第一扩增阶段期间不与靶序列杂交的5'“尾部”(例如,由于尾部与紧邻第一元件所结合的位置3'侧的靶序列部分之间缺少序列互补性)。例如,第一引物包含在序列A'的5'侧的序列C,第二引物包含在序列B的5'侧的序列D,并且在第一杂交温度下的第一扩增阶段期间,序列C和序列D都不与多种多联体(或环状多核苷酸)杂交。在使用这种有尾引物的一些实施方案中,扩增可包括第一阶段和第二阶段;第一阶段包括在第一温度下的杂交步骤,在此期间第一和第二引物与多联体(或环状多核苷酸)杂交并引物延伸;第二阶段包括在高于第一温度的第二温度下的杂交步骤,在此期间第一和第二引物与包含延伸的第一或第二引物或其互补序列的扩增产物杂交,并引物延伸。在两个温度的每一个下的扩增循环数可以基于期望的产物调节。通常,第一温度将用于数目相对较少的循环,例如约或小于约15、10、9、8、7、6、5个或更少的循环。在较高温度下的循环数可以独立于第一温度下的循环数来选择,但通常会有相同或更多个循环,例如约或至少约5、6、7、8、9、10、15、20、25或更多个循环。较高的温度有利于引物延伸产物中引物的第一元件与尾元件之间的杂交,而不利引物中的仅第一元件与多联体内靶序列之间的杂交形成的较短片段。因此,可以使用两阶段扩增以减少可能对短扩增产物有利的程度,从而保持相对较高比例的具有两个或更多个靶序列拷贝的扩增产物。例如,在5个循环的第二温度下杂交和引物延伸(例如,至少5、6、7、8、9、10、15、20或更多个循环)后,反应混合物中至少5%(例如至少5%、6%、7%、8%、9%、10%、15%、20%、25%、30%或更多)的扩增多核苷酸包含两个或更多个靶序列拷贝。根据这种两阶段有尾B2B引物扩增过程的实施方案的说明如图11A-11D所示。图15A-15C提供了进一步实施方式的说明。

[0184] 在一些实施方案中,扩增在倾斜的条件下进行,以增加来自多联体扩增子的长度。例如,可以降低引物浓度,使得不是每个引发位点都会与引物杂交,从而延长PCR产物。类似地,在循环期间减少引物杂交时间也会减少引物杂交,从而增加平均PCR扩增子大小。此外,增加循环的温度和/或延伸时间可以类似地增加PCR扩增子的平均长度。可以使用这些技术的任何组合。

[0185] 在一些实施方案中,特别是在使用B2B引物进行扩增的情况下,处理扩增产物,以基于大小过滤所得的扩增子,从而减少和/或消除包含多联体混合物中的单体数目。这可以使用多种可用的技术完成,包括但不限于,从凝胶中切除片段和凝胶过滤(例如,以富集长度大于约300、400、500个或更多个核苷酸的片段);以及通过微调结合缓冲液浓度进行SPRI珠子(Agencourt AMPure XP)的大小选择。例如,在与DNA片段混合期间使用0.6x结合缓冲液可用于优先结合大于约500个碱基对(bp)的DNA片段。

[0186] 在一些实施方案中,第一引物包含在序列A'的5'侧的序列C,第二引物包含在序列B的5'侧的序列D,并且在第一杂交温度下的第一扩增阶段期间,序列C和序列D都不与多种

环状多核苷酸杂交。扩增可包括第一阶段和第二阶段；其中第一阶段包括在第一温度下的杂交步骤，在此期间第一和第二引物在引物延伸之前与环状多核苷酸或其扩增产物杂交；而第二阶段包括在高于第一温度的第二温度下的杂交步骤，在此期间第一和第二引物与包含延伸的第一或第二引物或其互补序列的扩增产物杂交。例如，第一温度可以选择为约或大于约序列A'的 $T_m$ 、序列B的 $T_m$ 或其平均值，或者大于1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C、10°C或高于其中一个 $T_m$ 的温度。在该实例中，第二温度可以选择为约或大于约组合序列(A' + C)的 $T_m$ 、组合序列(B + D)的 $T_m$ 或其平均值，或者大于1°C、2°C、3°C、4°C、5°C、6°C、7°C、8°C、9°C、10°C或高于其中一个 $T_m$ 的温度。术语“ $T_m$ ”也称为“解链温度”，并且通常表示由参考序列（实际上可能是较大的多核苷酸中的子序列）组成的寡核苷酸的50%与其互补序列杂交（或分离）的温度。通常， $T_m$ 随着长度的增加而增加，因此，序列A'的 $T_m$ 预计低于组合序列(A' + C)的 $T_m$ 。

[0187] 在一方面，本公开内容提供了一种用于检测序列变体的系统。在一些实施方案中，该系统包含(a)计算机，该计算机被配置用于接收关于对样品进行检测反应的用户请求；(b)扩增系统，该扩增系统响应于用户请求，对样品或其部分进行核酸扩增反应，其中扩增反应包括以下步骤：(i)使用连接酶将多种多核苷酸中的个体多核苷酸环化以形成多种环状多核苷酸，多种多核苷酸中的每种多核苷酸在连接前在5'端和3'端之间具有接点；(ii)将连接酶降解；以及(iii)在将连接酶降解后扩增环状多核苷酸，以产生扩增的多核苷酸；其中不在步骤(i)与(iii)之间纯化或分离多核苷酸；(c)测序系统，该测序系统产生由扩增系统扩增的多核苷酸的测序读取，鉴定测序读取和与参考序列之间的序列差异，并将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体；以及(d)报告生成器，该报告生成器向接收者发送报告，其中该报告包含序列变体的检测结果。在一些实施方案中，接收者是用户。图32示出了可用于本公开内容的方法的系统的非限制性实例。图29和30提供了示例性工作流程设计的说明性示意图。

[0188] 系统中使用的计算机可包含一个或多个处理器。处理器可以与一个或多个控制器、计算单元和/或计算机系统的其他单元相关联，或者根据需要植入固件中。如果在软件中实现，则例程可存储在任何计算机可读存储器中，诸如存储在RAM、ROM、闪速存储器、磁盘、激光盘或其他合适的存储介质中。同样地，该软件可经由任何已知的传送方法而传送至计算设备，所述传送方法例如包括通过诸如电话线、因特网、无线连接等通信信道，或者经由诸如计算机可读盘、闪存驱动器等可移动介质。各个步骤可实现为各个区块、操作、工具、模块和技术，而所述各个区块、操作、工具、模块和技术继而可在硬件、固件、软件或者硬件、固件和/或软件的任何组合中实现。当在硬件中实现时，所述区块、操作、技术等之中的一些或全部区块、操作、技术可例如在定制集成电路(IC)、专用集成电路(ASIC)、现场可编程逻辑阵列(FPGA)、可编程逻辑阵列(PLA)等中实现。可以在系统的实施方案中使用客户端-服务器关系数据库架构。客户端-服务器架构是网络架构，其中网络上的每个计算机或进程都是客户端或服务器。服务器计算机通常是用于管理磁盘驱动器(文件服务器)、打印机(打印服务器)或网络流量(网络服务器)的强力计算机。客户端计算机包括用户运行应用程序的PC(个人计算机)或工作站，以及如本文公开的示例输出设备。客户端计算机依赖服务器计算机来获取资源，例如文件、设备甚至处理能力。在一些实施方案中，服务器计算机处理所有数据库功能。客户端计算机可以具有处理所有前端数据管理的软件，并且还可以接收来

自用户的数据输入。

[0189] 可以将系统配置用于接收关于对样品进行检测反应的用户请求。用户请求可以是直接的或间接的。直接请求的实例包括通过输入设备(如键盘、鼠标或触摸屏)传输的请求。间接请求的实例包括通过通信介质的传输,如通过互联网(有线或无线)的传输。

[0190] 该系统可以进一步包括扩增系统,其响应于用户请求,对样品或其部分进行核酸扩增反应。多种扩增多核苷酸(例如DNA和/或RNA)的方法是可用的。扩增可以是线性的,指数式的,或在多阶段扩增过程中涉及线性和指数阶段。扩增方法可包括温度的改变,例如热变性步骤,或者可以是不需要热变性的等温过程。本文描述了合适的扩增过程的非限制性实例,如关于本公开内容的多个方面中的任一方面所述。在一些实施方案中,扩增包括滚环扩增(RCA)。用于扩增多核苷酸的多种系统是可用的,并且可以基于待进行的扩增反应的类型而变化。例如,对于包含温度变化循环的扩增方法,扩增系统可包括热循环仪。扩增系统可包括实时扩增和检测仪器,例如Applied Biosystems、Roche和Stratagene制造的系统。在一些实施方案中,扩增反应包括以下步骤:(i)将个体多核苷酸环化以形成多种环状多核苷酸,每种环状多核苷酸在5'端和3'端之间具有接点;以及(ii)扩增环状多核苷酸。样品、多核苷酸、引物、聚合酶和其他试剂可以是本文所述的那些,如关于多个方面中的任一方面所述。本文中提供了环化过程(例如,使用和不使用衔接子寡核苷酸)、试剂(例如衔接子的类型、连接酶的使用)、反应条件(例如有利于自接合)、任选的额外加工(例如反应后纯化),和由此形成的接点的非限制性实例,如关于本公开内容的多个方面中的任一方面所述。可以选择和/或设计系统来执行任何此类方法。

[0191] 系统可以进一步包括测序系统,该测序系统产生由扩增系统扩增的多核苷酸的测序读取,鉴定测序读取与参考序列之间的序列差异,并将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体。测序系统和扩增系统可以相同,或包括重叠的设备。例如,扩增系统和测序系统可以使用相同的热循环仪。多种用于系统的测序平台是可用的,并且可以基于所选择的测序方法来选择。本文描述了测序方法的实例。扩增和测序可涉及使用液体处理器。几种商业上可用的液体处理系统可以用来运行这些过程的自动化(实例请参见例如来自Perkin-Elmer、Beckman Coulter、Caliper Life Sciences、Tecan、Eppendorf、Apricot Design、Velocity 11的液体处理器)。多种自动测序机器是可商购的,并且包括由Life Technologies(SOLiD平台和基于pH值检测)、Roche(454平台)、Illumina(例如基于流动池的系统,如基因组分析仪装置)制造的测序仪。在2、3、4、5或更多个自动化装置之间(例如,在液体处理器和测序装置中的一个或多个之间)的传输可以是手动的或自动的。

[0192] 本文描述了用于相对于参考序列鉴定序列差异和判定序列变体的方法,如关于本公开内容的多个方面中的任一方面所述。测序系统通常将包括响应于测序数据的输入和所需参数的输入(例如,参考基因组的选择)而执行这些步骤的软件。本文描述了比对算法和实现这些算法的比对器的实例,包括但不限于Needleman-Wunsch算法(参见,例如,EMBOSS Needle比对器,可从[www.ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)获得,任选地具有默认设置)、BLAST算法(参见,例如,BLAST比对工具,可从[blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)获得,任选地具有默认设置)或者Smith-Waterman算法(参见,例如,EMBOSS Water比对器,可从[www.ebi.ac.uk/Tools/psa/emboss\\_water/](http://www.ebi.ac.uk/Tools/psa/emboss_water/)



nucleotide.html获得,任选地具有默认设置)。可以使用所选算法的任何合适参数来评估最佳比对,包括默认参数。这种比对算法可以形成测序系统的一部分。

[0193] 该系统可以进一步包括报告生成器,其向接收者发送报告,其中该报告包含序列变体的检测结果。报告可以实时生成,例如在测序读取期间或在分析测序数据时生成,随着过程的进展定期更新。附加地或替代地,可以在分析结束时生成报告。报告可以自动生成,如当测序系统完成判定所有序列变体的步骤时。在一些实施方案中,报告响应于来自用户的指令而生成。除了检测序列变体的结果之外,报告还可以包含基于一种或多种序列变体的分析。例如,在一种或多种序列变体与特定污染物或表型相关联的情况下,报告可以包含关于该关联的信息,如污染物或表型存在的可能性,存在的水平,以及任选的基于这些信息的建议(例如附加测试、监测或补救措施)。报告可以采用多种形式。可以设想,与本公开内容相关的数据可以通过这样的网络或连接(或任何其他合适的传送信息的方法,包括但不限于,邮寄实物报告,如打印稿)进行传送以供接收和/或由接收方审查。接收方可以是但不限于个人或电子系统(例如,一个或多个计算机,以及/或一个或多个服务器)。

[0194] 在一方面,本公开内容提供了一种包含代码的计算机可读介质,该代码在由一个或多个处理器执行时实施检测序列变体的方法。在一些实施方案中,所实施的方法包括:(a)接收关于对样品进行检测反应的客户请求;(b)响应于客户要求请求,对样品或其部分进行核酸扩增反应,其中扩增反应包括以下步骤:(i)使用连接酶将多种多核苷酸中的个体多核苷酸环化,以形成多种环状多核苷酸,其中多种多核苷酸中的每种多核苷酸在连接前具有5'端和3'端;(ii)将连接酶降解;以及(iii)在将连接酶降解后扩增环状多核苷酸,以产生扩增的多核苷酸;其中不在步骤(i)与(iii)之间纯化或分离多核苷酸;(c)进行包括以下步骤的测序分析:(i)产生在扩增反应中扩增的多核苷酸的测序读取;(ii)鉴定测序读取与参考序列之间的序列差异;以及(iii)将出现在具有不同接点的至少两种环状多核苷酸中的序列差异判定为序列变体;以及(d)生成包含序列变体的检测结果的报告。

[0195] 包含计算机可执行代码的机器可读介质可以采用许多形式,包括但不限于,有形存储介质、载波介质或物理传输介质。非易失性存储介质例如包括光盘或磁盘(诸如任何计算机中的任何存储设备)等,诸如可用于实现数据库的存储介质等。易失性存储介质包括动态存储器,例如这种计算机平台的主存储器。有形传输介质包括同轴线缆、铜线和光纤,包括构成计算机系统内的总线的导线。载波传输介质可以采取电信号或电磁信号或者声波或光波的形式,诸如在射频(RF)和红外(IR)数据通信期间所生成的电信号或电磁信号或者声波或光波。因此,计算机可读介质的常见形式包括,例如:软盘、柔性盘、硬盘、磁带、任何其他磁介质、CD-ROM、DVD或DVD-ROM、任何其他光学介质、穿孔卡片纸带、任何具有孔洞图案的其他物理存储介质、RAM、ROM、PROM和EPROM、FLASH-EPROM、任何其他存储器芯片或盒、载波传输数据或指令、传输此类载波的缆线或链路,或者任何可让计算机从中读取编程代码和/或数据的其他介质。这些计算机可读介质的形式中的许多形式可参与向处理器传送一个或多个序列的一个或多个指令以供执行。

[0196] 本计算机可执行代码可以在包含处理器的任何合适的设备上执行,该设备包括服务器、PC或移动设备如智能手机或平板计算机。任何控制器或计算机任选地包括显示器,该显示器可以是阴极射线管("CRT")显示器、平板显示器(例如,有源矩阵液晶显示器、液晶显示器等)或其他显示器。计算机电路通常放置在盒子中,该盒子包含许多集成电路芯片,例



如微处理器、存储器、接口电路等。该盒子还任选地包含硬盘驱动器、软盘驱动器、大容量可移动驱动器(如可写CD-ROM)以及其他常见外围元件。输入设备(例如键盘、鼠标或触摸敏感屏幕)任选地提供来自用户的输入。计算机可以包括用于接收用户指令的适当软件,该软件可以是用户输入成一组参数字段的形式,例如GUI的形式,或者是预编程指令的形式,例如,针对各种不同的特定操作预编程的指令形式。

[0197] 在本公开内容的各个方面中的任何方面的一些实施方案中,所述方法、组合物和系统具有治疗应用,例如表征患者样品和任选地诊断受试者的状况。治疗应用还可以包括基于本文所述方法的结果来告知患者可能最有反应的疗法的选择(也称为“治疗诊断学”),以及对有需要的受试者的实际治疗。特别地,本文公开的方法和组合物可用于诊断肿瘤的存在、进展和/或肿瘤转移,尤其是当分析的多核苷酸包含cfDNA、ctDNA、cfRNA或片段化肿瘤DNA或由它们组成时。在一些实施方案中,监测对受试者的治疗功效。例如,通过随时间监测ctDNA,ctDNA的减少可以用作有效治疗的指标,而ctDNA的增加可以促进选择不同的治疗或不同的剂量。其他用途包括评估移植受者的器官排斥(与移植供体基因组相对应的循环DNA的量的增加被用作移植排斥的早期指标),以及病原体感染如病毒或细菌感染的基因分型/分型(isotyping)。检测循环胎儿DNA中的序列变体可用于诊断胎儿的状况。

[0198] 如本文所用,“治疗”或“处理”或“缓解”或“减轻”可互换使用。这些术语是指获得有益或期望结果的方法,该有益或期望结果包括但不限于治疗益处和/或预防益处。治疗益处是指对治疗中的一种或多种疾病、状况或症状的任何治疗相关的改善或效果。对于预防益处,可将组合物施用于存在发展出特定疾病、状况或症状的风险的受试者,或施用于报告了疾病的一种或多种生理症状的受试者,即使疾病、状况或症状可能尚未表现出来。通常,预防益处包括降低治疗中的一种或多种疾病、状况或症状的发生率和/或减缓其恶化(例如在经治疗群体与未治疗群体之间,或在受试者的治疗状态与未治疗状态之间)。改善治疗结果可以包括诊断受试者的状况,以鉴定受试者是否将受益于用一种或多种治疗剂的治疗,或其他治疗干预(如手术)。在这种诊断应用中,相对于未根据本公开内容的方法进行诊断而分组的患者中的有效性,用一种或多种治疗剂的总体成功治疗率可以改善(例如,治疗功效的测量值改善至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、95%或更多)。

[0199] 术语“受试者”、“个体”和“患者”在本文中可互换使用,指脊椎动物,优选指哺乳动物,更优选指人。哺乳动物包括但不限于鼠、猿猴、人、农场动物、竞技动物和宠物。还包括体内获得或体外培养的生物实体的组织、细胞及其后代。

[0200] 术语“治疗剂(therapeutic agent)”、“治疗能力剂(therapeutic capable agent)”或“处理剂(treatment agent)”可互换使用,是指在向受试者施用后带来一些有益效果的分子或化合物。有益效果包括实现确诊;疾病、症状、病症或病理状况减轻;减少或预防疾病、症状、病症或状况的发作;以及通常抵抗疾病、症状、病症或病理状况。

[0201] 在本文所述的各种方法的一些实施方案中,样品来自受试者。受试者可以是任何生物体,其非限制性实例包括植物、动物、真菌、原生生物、原核生物、病毒、线粒体和叶绿体。样品多核苷酸可以从受试者分离,例如细胞样品、组织样品、体液样品或器官样品(或来自其中任何一种的细胞培养物),包括,例如,培养的细胞系、活检物、血液样品、脸颊拭子或含有细胞的流体样品(如唾液)。在一些情况下,样品不包含完整细胞,被处理以除去细胞,或在没有细胞提取步骤的情况下分离多核苷酸(例如分离无细胞多核苷酸,例如无细胞

DNA)。其他样品来源的实例包括血液、尿液、粪便、鼻孔、肺、肠、其他体液或排泄物,由此衍生的物质或其组合。受试者可以是动物,包括但不限于,牛、猪、小鼠、大鼠、鸡、猫、狗等,并且通常是哺乳动物,如人。在一些实施方案中,样品包含肿瘤细胞,例如来自受试者的肿瘤组织样品。在一些实施方案中,样品是血液样品或其一部分(例如血浆或血清)。血清和血浆可能是特别感兴趣的,因为与此类组织中恶性细胞死亡率较高相关的肿瘤DNA的相对富集。样品可以是新鲜样品,或是经历一种或多种储存过程的样品(例如石蜡包埋样品,特别是福尔马林固定、石蜡包埋的(FFPE)样品)。在一些实施方案中,来自单个个体的样品被分成多个单独的样品(例如,2、3、4、5、6、7、8、9、10或更多个单独的样品),该单独的样品独立地经历本公开内容的方法,如一式两份、一式三份、一式四份或一式更多份的分析。当样品来自受试者时,参考序列也可以来源于受试者,例如来自分析样品的共有序列或来自相同受试者的另一样品或组织的多核苷酸的序列。例如,可以分析血液样品的ctDNA突变,同时分析来自另一个样品(例如口腔或皮肤样品)的细胞DNA,以确定参考序列。

[0202] 可以根据任何合适的方法,在有或没有从样品中的细胞进行提取的情况下从样品中提取多核苷酸。有多种试剂盒可用于提取多核苷酸,其选择可取决于样品的类型或待分离的核酸的类型。本文提供了提取方法的实例,如关于本文公开的多个方面中的任一方面描述。在一个实例中,样品可以是血液样品,例如在EDTA管(例如BD Vacutainer)中收集的样品。可以通过离心(例如在1900xg,4℃下10分钟)将血浆与外周血细胞分离。以此种方式在6mL血液样品上进行的血浆分离通常将产生2.5至3mL血浆。根据制造商的方案,可以从血浆样品中提取循环无细胞DNA,如通过使用QIAmp循环核酸试剂盒(Qiagen)。然后可以定量DNA(例如,在具有高灵敏度DNA试剂盒(Agilent)的Agilent 2100生物分析仪上)。例如,来自健康人的这种血浆样品的循环DNA的产量可以是每毫升血浆1ng至10ng,在癌症患者样品中明显更高。

[0203] 多核苷酸还可以源自储存的样品,例如冷冻或存档的样品。用于存储样品的一种常用方法是将其福尔马林固定和石蜡包埋。然而,该过程还与核酸的降解有关。从FFPE样品处理和分析的多核苷酸可包括短多核苷酸,例如50-200个碱基对或更短的片段。从固定的石蜡包埋样品中纯化核酸的技术有很多,如在W02007133703中描述的方法,以及由Foss等人,Diagnostic Molecular Pathology, (1994) 3:148-155和Paska,C等人,Diagnostic Molecular Pathology, (2004) 13:234-240所描述的方法。可商购的试剂盒可用于纯化FFPE样品中的多核苷酸,如Ambion's Recoverall总核酸分离试剂盒。典型的方法是从通过用二甲苯或其他有机溶剂提取而除去组织中的石蜡的步骤开始,然后用热和蛋白酶如蛋白酶K处理,该蛋白酶切割组织和蛋白质并帮助从组织释放基因组物质。然后可以将释放的核酸捕获在膜上,或从溶液中沉淀,洗涤以除去杂质,并且对于mRNA分离的情况,有时增加DNase处理步骤以降解不需要的DNA。其他提取FFPE DNA的方法是可用的,并且可以用于本公开内容的方法中。

[0204] 在一些实施方案中,多种多核苷酸包含无细胞多核苷酸,如无细胞DNA(cfDNA)、无细胞RNA(cfRNA)、循环肿瘤DNA(ctDNA)或循环肿瘤RNA(ctRNA)。无细胞DNA在健康和患病个体中循环。无细胞RNA在健康和患病个体中循环。来自肿瘤的cfDNA(ctDNA)不限于任何特定的癌症类型,但似乎是不同恶性肿瘤中常见的发现。根据一些测量,血浆中的游离循环DNA浓度在对照受试者中为约14-18ng/ml,在肿瘤患者中为约180-318ng/ml。凋亡和坏死的细

胞死亡有助于体液中的无细胞循环DNA。例如,在前列腺癌患者和其他前列腺疾病如良性前列腺增生和前列腺炎的血浆中观察到循环DNA水平显著升高。此外,循环肿瘤DNA存在于原发性肿瘤所在器官的体液中。因此,乳腺癌检测可以在导管灌洗液中实现;结直肠癌检测在大便中实现;肺癌检测在痰中实现,并且前列腺癌检测在尿液或精液中检测。无细胞DNA可以从多种来源获得。一个常见的来源是受试者的血液样品。然而,cfDNA或其他片段化DNA可以来自多种其他来源。例如,尿液和粪便样品可以是包括ctDNA在内的cfDNA的来源。无细胞RNA可以从多种来源获得。

[0205] 在一些实施方案中,在没有提取步骤和/或没有纯化步骤的情况下使多核苷酸经历后续步骤(例如环化和扩增)。例如,可以在没有提取步骤的情况下处理流体样品以除去细胞,以产生纯化的液体样品和细胞样品,然后从纯化的流体样品中分离DNA。有多种分离多核苷酸的方法可用,例如通过沉淀,或与基底非特异性结合并随后洗涤该基底以释放结合的多核苷酸。在没有细胞提取步骤的情况下从样品中分离多核苷酸时,多核苷酸将主要是细胞外或“无细胞”多核苷酸。例如,无细胞多核苷酸可包括无细胞DNA(也称为“循环”DNA)。在一些实施方案中,循环DNA是来自肿瘤细胞的循环肿瘤DNA(ctDNA),例如来自体液或排泄物(例如血液样品)。无细胞多核苷酸可包括无细胞RNA(也称为“循环”RNA)。在一些实施方案中,循环RNA是来自肿瘤细胞的循环肿瘤RNA(ctRNA)。肿瘤经常显示出凋亡或坏死,使得肿瘤核酸通过各种机制,以不同形式和不同水平释放到体内,包括受试者的血流中。通常,ctDNA的大小可以在较高浓度的较小片段(通常长度为70至200个核苷酸)至较低浓度的高达数千碱基的大片段之间。

[0206] 在本文所述的各个方面中的任一方面的一些实施方案中,检测序列变体包括检测相对于参考序列的突变(例如罕见的体细胞突变)或无突变的背景中的突变,其中序列变体与疾病相关。一般而言,有统计、生物和/或功能证据表明与疾病或性状相关的序列变体被称为“因果遗传变体”。单一因果遗传变体可与超过一种疾病或性状有关。在一些实施方案中,因果遗传变体可以与孟德尔性状、非孟德尔性状或两者关联。因果遗传变体可以表现为多核苷酸的变异,如1、2、3、4、5、6、7、8、9、10、20、50或更多序列差异(例如,在包含因果遗传变体的多核苷酸与在相同的相对基因组位置缺少该因果遗传变体的多核苷酸之间)。因果遗传变体类型的非限制性实例包括单核苷酸多态性(SNP)、缺失/插入多态性(DIP)、拷贝数变体(CNV)、短串联重复(STR)、限制片段长度多态性(RFLP)、简单序列重复(SSR)、可变数目串联重复(VNTR)、随机扩增多态性DNA(RAPD)、扩增片段长度多态性(AFLP)、反转录转座子间扩增多态性(IRAP)、长短穿插元件(LINE/SINE)、长串联重复(LTR)、移动元件、反转录转座子微卫星扩增多态性、基于反转录转座子的插入多态性、序列特异性扩增多态性,以及可遗传表观遗传修饰(例如,DNA甲基化)。因果遗传变体也可以是一组密切相关的因果遗传变体。一些因果遗传变体可能作为RNA多核苷酸中的序列变异发挥作用。在该水平上,一些因果遗传变体还通过某种RNA多核苷酸是否存在来指示。此外,一些因果遗传变体导致蛋白质多肽的序列变异。已经报道了许多因果遗传变体。导致镰状细胞贫血的血红蛋白的Hb S变体是SNP因果遗传变体的一个实例。引起囊性纤维化的CFTR基因的 $\delta 508$ 突变是DIP因果遗传变体的一个实例。引起唐氏综合症的21三体是CNV因果遗传变体的一个实例。引起亨廷顿氏病的串联重复是STR因果遗传变体的一个实例。表1提供了因果遗传变体及与其相关的疾病的非限制性实例。在W02014015084中描述了因果遗传变体的其他非限制性实例。表2提供了

基因的其他实例,该基因中的突变与疾病相关并且其中的序列变体可以根据本公开内容的方法来检测。

表1

疾病	基因	变体名称
21-羟化酶缺乏症	CYP21A2	F306+t
21-羟化酶缺乏症	CYP21A2	F306+t
21-羟化酶缺乏症	CYP21A3	g.655A/C>G
21-羟化酶缺乏症	CYP21A4	g.655A/C>G
21-羟化酶缺乏症	CYP21A6	G110del8nt
21-羟化酶缺乏症	CYP21A5	G110del8nt
21-羟化酶缺乏症	CYP21A7	I172N, rs34607927
21-羟化酶缺乏症	CYP21A2	I236N
21-羟化酶缺乏症	CYP21A2	M239K, rs6476
21-羟化酶缺乏症	CYP21A2	P30L
21-羟化酶缺乏症	CYP21A2	P453S
21-羟化酶缺乏症	CYP21A2	Q318X
21-羟化酶缺乏症	CYP21A2	R356W
21-羟化酶缺乏症	CYP21A2	V237E, rs12530380
21-羟化酶缺乏症	CYP21A2	V281L, rs6471
ABCC8 相关的高胰岛素血症	ABCC8	3992-9G>A
ABCC8 相关的高胰岛素血症	ABCC8	delF1388
ABCC8 相关的高胰岛素血症	ABCC8	delF1388
ABCC8 相关的高胰岛素血症	ABCC8	V187D
软骨发育不全	FGFR3	G375C
软骨发育不全	FGFR3	G380R, rs28931614
软骨发育不全	CNGB3	c.1148delC
软骨发育不全	CNGB3	c.1148delC
软骨发育不全	CNGB3	c.819-826del8
软骨发育不全	CNGB3	c.819-826del8
软骨发育不全	CNGB3	c.886-896del11insT
软骨发育不全	CNGB3	c.886-896del11insT
软骨发育不全	CNGB3	c.991-3T>G
软骨发育不全	CNGB3	p.Arg403Gln
软骨发育不全	CNGB3	p.Glu336X
腺苷一磷酸脱氢酶 1	AMPD1	P48L
腺苷一磷酸脱氢酶 1	AMPD1	Q12X, rs17602729
胼胝体发育不全伴神经病变	SLC12A6	c.2436delG
胼胝体发育不全伴神经病变	SLC12A6	c.2436delG
尿黑酸尿	HGD	c.174delA
尿黑酸尿	HGD	c.174delA
尿黑酸尿	HGD	c.457_458insG
尿黑酸尿	HGD	c.457_458insG

疾病	基因	变体名称
尿黑酸尿	HGD	G161R
尿黑酸尿	HGD	G270R
尿黑酸尿	HGD	IVS1-1G>A
尿黑酸尿	HGD	IVS5+1G>A
尿黑酸尿	HGD	Met368Val
尿黑酸尿	HGD	P230S
尿黑酸尿	HGD	S47L
尿黑酸尿	HGD	V300G
$\alpha$ -1-抗胰蛋白酶缺乏症	SERPINA1	Arg101His, rs709932
$\alpha$ -1-抗胰蛋白酶缺乏症	SERPINA1	Glu264Val
$\alpha$ -1-抗胰蛋白酶缺乏症	SERPINA1	Glu342Lys, rs28929474
$\alpha$ -1-抗胰蛋白酶缺乏症	SERPINA1	Glu376Asp, rs1303
$\alpha$ -甘露糖苷贮积症	MAN2B1	IVS14+1G>C
$\alpha$ -甘露糖苷贮积症	MAN2B1	p.L809P
$\alpha$ -甘露糖苷贮积症	MAN2B1	p.R750W
$\alpha$ -肌聚糖病	SGCA	R77C, rs28933693
$\alpha$ -地中海贫血	HBA2	H19D
$\alpha$ -地中海贫血	HBA1	HbQ
$\alpha$ -地中海贫血	HBA1,HBA2	3.7 kb (I 型) 缺失 $\alpha$ -2
$\alpha$ -地中海贫血	HBA1,HBA2	3.7 kb (I 型) 缺失 $\alpha$ -2
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1086, --(SEA); 缺失约 20 kb, 包括 $\alpha$ -球蛋白基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1086, --(SEA); 缺失约 20 kb, 包括 $\alpha$ -球蛋白基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1087, --(MED-I); 缺失约 17.5 kb, 包括 $\alpha$ -球蛋白基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1087, --(MED-I); 缺失约 17.5 kb, 包括 $\alpha$ -球蛋白基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1088, -( $\alpha$ )20.5; 这种 20.5 kb 的缺失涉及 $\alpha$ 2 和 $\alpha$ 1 的 5'端; $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1088, -( $\alpha$ )20.5; 这种 20.5 kb 的缺失涉及 $\alpha$ 2 和 $\alpha$ 1 的 5'端; $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1094, --(FIL); 缺失 30-34 kb, 涉及 $\alpha$ 1、 $\alpha$ 2 和 zeta 基因 $\alpha$ -Thal-1

疾病	基因	变体名称
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1094, --(FIL); 缺失 30-34 kb, 涉 及 $\alpha 1$ 、 $\alpha 2$ 和 zeta 基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1095, --(THAI); 缺失 34-38 kb, 涉 及 $\alpha 1$ 、 $\alpha 2$ 和 $\zeta$ 基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1095, --(THAI); 缺失 34-38 kb, 涉 及 $\alpha 1$ 、 $\alpha 2$ 和 $\zeta$ 基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1097, --(MED-II); 缺失 26.5 kb 涉 及两个 $\alpha$ 和 $\zeta$ 基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 1097, --(MED-II); 缺失 26.5 kb 涉 及两个 $\alpha$ 和 $\zeta$ 基因 $\alpha$ -Thal-1
$\alpha$ -地中海贫血	HBA2	HbVar 数据库 id # 187
$\alpha$ -地中海贫血	HBA2	HbVar 数据库 id # 2598, IVS I-5 (G>A)
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 703
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 704
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 705, Hb Koya Dora
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 707, rs41412046
$\alpha$ -地中海贫血	HBA1	HbVar 数据库 id # 87
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 969, Poly A (A->G); AATAAA->AATGAA $\beta\beta$ +
$\alpha$ -地中海贫血	HBA1,HBA2	HbVar 数据库 id # 971, Poly A (A->G); AATAAA->AATAAG $\beta\beta$ +
$\alpha$ -地中海贫血	HBA2	M1T
$\alpha$ -地中海贫血	HBA1	W14X
血管紧张素 II 受体, 1 型	AGTR1	A1166C
载脂蛋白 E 基因分型	APOE	p.C112R, rs429358
载脂蛋白 E 基因分型	APOE	p.R158C, rs7412
精氨基琥珀酸尿	ASL	R385C
ARSACS	SACS	5254C>T
ARSACS	SACS	6594delT
ARSACS	SACS	6594delT
天冬氨酰基葡萄糖胺尿	AGA	c.199_200delGA
天冬氨酰基葡萄糖胺尿	AGA	c.199_200delGA
天冬氨酰基葡萄糖胺尿	AGA	C163S
共济失调伴维生素 E 缺乏	TTPA	744delA

疾病	基因	变体名称
共济失调伴维生素 E 缺乏	TTPA	744delA
共济失调-毛细血管扩张症	ATM	R35X
自身免疫性多内分泌腺病综合征 1 型	AIRE	c.1163_1164insA
自身免疫性多内分泌腺病综合征 1 型	AIRE	c.1163_1164insA
自身免疫性多内分泌腺病综合征 1 型	AIRE	c.769C>T
自身免疫性多内分泌腺病综合征 1 型	AIRE	c.967_979del
自身免疫性多内分泌腺病综合征 1 型	AIRE	c.967_979del
自身免疫性多内分泌腺病综合征 1 型	AIRE	Y85C
Bardet-Biedl 综合症	BBS1	M390R
Bardet-Biedl 综合症	BBS10	p.C91LfsX4
Bardet-Biedl 综合症	BBS10	p.C91LfsX4
贝斯特卵黄状黄斑营养不良	BEST1	c.G383C
$\beta$ -肌聚糖病	SGCB	S114F
$\beta$ -地中海贫血	HBB	-28 (A->G) $\beta\beta$ +
$\beta$ -地中海贫血	HBB	-29 (A->G) $\beta\beta$ +
$\beta$ -地中海贫血	HBB	-29A>G
$\beta$ -地中海贫血	HBB	-30 (T->A) $\beta\beta$ +
$\beta$ -地中海贫血	HBB	-87 (C->G) $\beta\beta$ +
$\beta$ -地中海贫血	HBB	-88C>T
$\beta$ -地中海贫血	HBB	CAP+1 (A->C) $\beta\beta$ +
$\beta$ -地中海贫血	HBB	密码子 15 (G->A); TGG(Trp)->TAG(终止密码子) $\beta\beta$ 0, rs34716011
$\beta$ -地中海贫血	HBB	密码子 15 (G->A); TGG(Trp)->TAG(终止密码子) $\beta$ 0, rs34716011
$\beta$ -地中海贫血	HBB	密码子 16 (-C); GGC(Gly)->GG- $\beta$ 0
$\beta$ -地中海贫血	HBB	密码子 16 (-C); GGC(Gly)->GG- $\beta$ 0
$\beta$ -地中海贫血	HBB	密码子 17 (A->T); AAG(Lys)->TAG(终止密码子) $\beta$ 0
$\beta$ -地中海贫血	HBB	密码子 24 (T->A); GGT(Gly)->GGA(Gly) $\beta$ +
$\beta$ -地中海贫血	HBB	密码子 39 (C->T); CAG(Gln)->TAG(终止密码子) $\beta$ 0

疾病	基因	变体名称
β-地中海贫血	HBB	密码子 5 (-CT); CCT(Pro)->C--β0
β-地中海贫血	HBB	密码子 5 (-CT); CCT(Pro)->C--β0
β-地中海贫血	HBB	密码子 6 (-A); GAG(Glu)->G-G β0
β-地中海贫血	HBB	密码子 6 (-A); GAG(Glu)->G-G β0
β-地中海贫血	HBB	密码子 8 (-AA); AAG(Lys)->--G β0
β-地中海贫血	HBB	密码子 8 (-AA); AAG(Lys)->--G β0
β-地中海贫血	HBB	密码子 41/42 (-TTCT); TTCTTT(Phe-Phe)->----TT β0
β-地中海贫血	HBB	密码子 41/42 (-TTCT); TTCTTT(Phe-Phe)->----TT β0
β-地中海贫血	HBB	密码子 71/72 (+A); TTT AGT(Phe Ser)->TTT A AGT; β0
β-地中海贫血	HBB	密码子 71/72 (+A); TTT AGT(Phe Ser)->TTT A AGT; β0
β-地中海贫血	HBB	密码子 8/9 (+G); AAG TCT(Lys;Ser)->AAG G TCT β0
β-地中海贫血	HBB	密码子 8/9 (+G); AAG TCT(Lys;Ser)->AAG G TCT β0
β-地中海贫血	HBB	HbVar 数据库 id # 889, IVS-II-654 (C->T); AAGGCAATA->AAG^GTA ATA β+(严重)
β-地中海贫血	HBB	HbVar 数据库 id # 890, IVS-II-705 (T->G); GATGTAAGA->GAG^GTA AGA β+
β-地中海贫血	HBB	HbVar 数据库 id # 891, IVS-II-745 (C->G); CAGCTACCAT->CAG^GT ACCAT β+
β-地中海贫血	HBB	HbVar 数据库 id # 979, 619 bp 缺失 β0
β-地中海贫血	HBB	619 bp 缺失 β0



疾病	基因	变体名称
β-地中海贫血	HBB	IVS-I-1 (G->A); AG^GTTGGT->AGATTGG T β0
β-地中海贫血	HBB	IVS-I-1 (G->T); AG^GTTGGT->AGTTTGGT β0
β-地中海贫血	HBB	IVS-I-110 (G->A) β+;突变是 受体剪接位点 AG^GC 的 5' 侧的 21 个核苷酸
β-地中海贫血	HBB	IVS-I-5 (G->C) β+(严重)
β-地中海贫血	HBB	IVS-II-1 (G->A); β0
β-地中海贫血	HBB	IVS-II-844 (C->G); β+
β-地中海贫血	HBB	IVS1+6T>C
β-地中海贫血	HBB	IVS11-849A>C
β-地中海贫血	HBB	IVS11-849A>G
生物素酰胺酶缺乏症	BTB	A171T, rs13073139
生物素酰胺酶缺乏症	BTB	D252G, rs28934601
生物素酰胺酶缺乏症	BTB	D444H, rs13078881
生物素酰胺酶缺乏症	BTB	F403V
生物素酰胺酶缺乏症	BTB	G98:d7i3
生物素酰胺酶缺乏症	BTB	G98:d7i3
生物素酰胺酶缺乏症	BTB	Q456H
生物素酰胺酶缺乏症	BTB	R157H
生物素酰胺酶缺乏症	BTB	R538C
Blau 综合征	NOD2	E383K
Blau 综合征	NOD2	L469F
Blau 综合征	NOD2	R334Q
Blau 综合征	NOD2	R334W
布卢姆综合征	BLM	2407insT
布卢姆综合征	BLM	2407insT
布卢姆综合征	BLM	736delATCTGAinsTAGATT C (2281del6/ins7)
布卢姆综合征	BLM	736delATCTGAinsTAGATT C (2281del6/ins7)
BRCA1 遗传性乳腺癌/卵巢癌	BRCA1	185delAG
BRCA1 遗传性乳腺癌/卵巢癌	BRCA1	185delAG
BRCA1 遗传性乳腺癌/卵巢癌	BRCA1	5382insC
BRCA1 遗传性乳腺癌/卵巢癌	BRCA1	5382insC
BRCA1 遗传性乳腺癌/卵巢癌	BRCA1	Tyr978X
BRCA2 遗传性乳腺癌/卵巢癌	BRCA2	6174delT
BRCA2 遗传性乳腺癌/卵巢癌	BRCA2	6174delT
BRCA2 遗传性乳腺癌/卵巢癌	BRCA2	8765delAG
BRCA2 遗传性乳腺癌/卵巢癌	BRCA2	8765delAG

疾病	基因	变体名称
卡纳万病	ASPA	A305E (914C>A), rs28940574
卡纳万病	ASPA	E285A (854A>C), rs28940279
卡纳万病	ASPA	IVS2-2A>G (433-2A>G)
卡纳万病	ASPA	Y231X (693C>A)
肉碱棕榈酰转移酶 IA 缺乏症	CPT1A	G710E
肉碱棕榈酰转移酶 IA 缺乏症	CPT1A	P479L
肉碱棕榈酰转移酶 II 缺乏症	CPT2	G549D
肉碱棕榈酰转移酶 II 缺乏症	CPT2	L178F 534 ins/25 bp del
肉碱棕榈酰转移酶 II 缺乏症	CPT2	L178F 534 ins/25 bp del
肉碱棕榈酰转移酶 II 缺乏症	CPT2	P227L
肉碱棕榈酰转移酶 II 缺乏症	CPT2	P50H
肉碱棕榈酰转移酶 II 缺乏症	CPT2	P604S
肉碱棕榈酰转移酶 II 缺乏症	CPT2	Q413fs
肉碱棕榈酰转移酶 II 缺乏症	CPT2	Q413fs
肉碱棕榈酰转移酶 II 缺乏症	CPT2	Q550R
肉碱棕榈酰转移酶 II 缺乏症	CPT2	R124X
肉碱棕榈酰转移酶 II 缺乏症	CPT2	R503C
肉碱棕榈酰转移酶 II 缺乏症	CPT2	R631C
肉碱棕榈酰转移酶 II 缺乏症	CPT2	S113L
肉碱棕榈酰转移酶 II 缺乏症	CPT2	s38fs
肉碱棕榈酰转移酶 II 缺乏症	CPT2	s38fs
肉碱棕榈酰转移酶 II 缺乏症	CPT2	Y628S, rs28936673
软骨-毛发发育不全	RMRP	g.262G>T
软骨-毛发发育不全	RMPR	g.70A>G
CFTR 相关的病症	CFTR	1811+1.6kbA->G
CFTR 相关的病症	CFTR	2183AA>G
CFTR 相关的病症	CFTR	2183AA>G
CFTR 相关的病症	CFTR	3849+10kbC>T
CFTR 相关的病症	CFTR	A455E
CFTR 相关的病症	CFTR	A559T
CFTR 相关的病症	CFTR	C524X
CFTR 相关的病症	CFTR	574delA, 574delA
CFTR 相关的病症	CFTR	574delA, 574delA
CFTR 相关的病症	CFTR	2108delA, 2108delA
CFTR 相关的病症	CFTR	2108delA, 2108delA
CFTR 相关的病症	CFTR	3171delC, 3171delC
CFTR 相关的病症	CFTR	3171delC, 3171delC
CFTR 相关的病症	CFTR	621+1G->T
CFTR 相关的病症	CFTR	2105-2117del13insAGAAA
CFTR 相关的病症	CFTR	2105-2117del13insAGAAA
CFTR 相关的病症	CFTR	711+1G->T

疾病	基因	变体名称
CFTR 相关的病症	CFTR	711+5G->A
CFTR 相关的病症	CFTR	712-1G->T
CFTR 相关的病症	CFTR	1288insTA, 1288insTA
CFTR 相关的病症	CFTR	1288insTA, 1288insTA
CFTR 相关的病症	CFTR	936delTA
CFTR 相关的病症	CFTR	936delTA
CFTR 相关的病症	CFTR	[δ]F311
CFTR 相关的病症	CFTR	[δ]F311
CFTR 相关的病症	CFTR	1078delT, 1078delT
CFTR 相关的病症	CFTR	1078delT, 1078delT
CFTR 相关的病症	CFTR	1161delC, 1161delC
CFTR 相关的病症	CFTR	1161delC, 1161delC
CFTR 相关的病症	CFTR	1609delCA, 1609delCA
CFTR 相关的病症	CFTR	1609delCA, 1609delCA
CFTR 相关的病症	CFTR	[δ]I507
CFTR 相关的病症	CFTR	[δ]I507
CFTR 相关的病症	CFTR	rs332, [δ]F508
CFTR 相关的病症	CFTR	rs332, [δ]F508
CFTR 相关的病症	CFTR	1677delTA, 1677delTA
CFTR 相关的病症	CFTR	1677delTA, 1677delTA
CFTR 相关的病症	CFTR	1717-1G->A
CFTR 相关的病症	CFTR	1812-1G->A
CFTR 相关的病症	CFTR	1898+1G->A
CFTR 相关的病症	CFTR	1898+1G->T
CFTR 相关的病症	CFTR	1898+5G->T
CFTR 相关的病症	CFTR	1949del84, 1949del84
CFTR 相关的病症	CFTR	1949del84, 1949del84
CFTR 相关的病症	CFTR	2043delG, 2043delG
CFTR 相关的病症	CFTR	2043delG, 2043delG
CFTR 相关的病症	CFTR	2055del9->A
CFTR 相关的病症	CFTR	2055del9->A
CFTR 相关的病症	CFTR	2143delT, 2143delT
CFTR 相关的病症	CFTR	2143delT, 2143delT
CFTR 相关的病症	CFTR	2184delA, 2184delA
CFTR 相关的病症	CFTR	2184delA, 2184delA
CFTR 相关的病症	CFTR	2184insA, 2184insA
CFTR 相关的病症	CFTR	2184insA, 2184insA
CFTR 相关的病症	CFTR	2307insA, 2307insA
CFTR 相关的病症	CFTR	2307insA, 2307insA
CFTR 相关的病症	CFTR	296+12T->C
CFTR 相关的病症	CFTR	2789+5G->A
CFTR 相关的病症	CFTR	2869insG, 2869insG
CFTR 相关的病症	CFTR	2869insG, 2869insG

疾病	基因	变体名称
CFTR 相关的病症	CFTR	3120G->A
CFTR 相关的病症	CFTR	3120+1G->A
CFTR 相关的病症	CFTR	3272-26A->G
CFTR 相关的病症	CFTR	3659delC, 3659delC
CFTR 相关的病症	CFTR	3659delC, 3659delC
CFTR 相关的病症	CFTR	3667del4, 3667del4
CFTR 相关的病症	CFTR	3667del4, 3667del4
CFTR 相关的病症	CFTR	3791delC, 3791delC
CFTR 相关的病症	CFTR	3791delC, 3791delC
CFTR 相关的病症	CFTR	3821delT, 3821delT
CFTR 相关的病症	CFTR	3821delT, 3821delT
CFTR 相关的病症	CFTR	3905insT, 3905insT
CFTR 相关的病症	CFTR	3905insT, 3905insT
CFTR 相关的病症	CFTR	4016insT, 4016insT
CFTR 相关的病症	CFTR	4016insT, 4016insT
CFTR 相关的病症	CFTR	394delTT, 394delTT
CFTR 相关的病症	CFTR	394delTT, 394delTT
CFTR 相关的病症	CFTR	405+1G->A
CFTR 相关的病症	CFTR	405+3A->C
CFTR 相关的病症	CFTR	444delA
CFTR 相关的病症	CFTR	444delA
CFTR 相关的病症	CFTR	3876delA, 3876delA
CFTR 相关的病症	CFTR	3876delA, 3876delA
CFTR 相关的病症	CFTR	457TAT->G
CFTR 相关的病症	CFTR	457TAT->G
CFTR 相关的病症	CFTR	3199del6, 3199del6
CFTR 相关的病症	CFTR	3199del6, 3199del6
CFTR 相关的病症	CFTR	406-1G->A
CFTR 相关的病症	CFTR	663delT, 663delT
CFTR 相关的病症	CFTR	663delT, 663delT
CFTR 相关的病症	CFTR	935delA, 935delA
CFTR 相关的病症	CFTR	935delA, 935delA
CFTR 相关的病症	CFTR	CFTR dele2,3 (21kb)
CFTR 相关的病症	CFTR	CFTR dele2,3 (21kb)
CFTR 相关的病症	CFTR	D1152H
CFTR 相关的病症	CFTR	E60X
CFTR 相关的病症	CFTR	E92X
CFTR 相关的病症	CFTR	F508C, rs1800093
CFTR 相关的病症	CFTR	G178R
CFTR 相关的病症	CFTR	G330X
CFTR 相关的病症	CFTR	G480C
CFTR 相关的病症	CFTR	G542X
CFTR 相关的病症	CFTR	G551D

疾病	基因	变体名称
CFTR 相关的病症	CFTR	G622D
CFTR 相关的病症	CFTR	G85E
CFTR 相关的病症	CFTR	G91R
CFTR 相关的病症	CFTR	I148T, rs35516286
CFTR 相关的病症	CFTR	I506V
CFTR 相关的病症	CFTR	IVS8-5T
CFTR 相关的病症	CFTR	IVS8-7T
CFTR 相关的病症	CFTR	IVS8-9T
CFTR 相关的病症	CFTR	K710X
CFTR 相关的病症	CFTR	L206W
CFTR 相关的病症	CFTR	M1101K, rs36210737
CFTR 相关的病症	CFTR	N1303K
CFTR 相关的病症	CFTR	P574H
CFTR 相关的病症	CFTR	Q1238X
CFTR 相关的病症	CFTR	Q359K/T360K_wt
CFTR 相关的病症	CFTR	Q493X
CFTR 相关的病症	CFTR	Q552X
CFTR 相关的病症	CFTR	Q890X
CFTR 相关的病症	CFTR	R1066C
CFTR 相关的病症	CFTR	R1070Q
CFTR 相关的病症	CFTR	R1158X
CFTR 相关的病症	CFTR	R1162X
CFTR 相关的病症	CFTR	R117C
CFTR 相关的病症	CFTR	R117H
CFTR 相关的病症	CFTR	R1283M
CFTR 相关的病症	CFTR	R334W
CFTR 相关的病症	CFTR	R347H
CFTR 相关的病症	CFTR	R347P
CFTR 相关的病症	CFTR	R352Q
CFTR 相关的病症	CFTR	R553X
CFTR 相关的病症	CFTR	R560T
CFTR 相关的病症	CFTR	R709X
CFTR 相关的病症	CFTR	R75X
CFTR 相关的病症	CFTR	R764X
CFTR 相关的病症	CFTR	S1196X
CFTR 相关的病症	CFTR	S1235R, rs34911792
CFTR 相关的病症	CFTR	S1251N
CFTR 相关的病症	CFTR	S1255X
CFTR 相关的病症	CFTR	S364P
CFTR 相关的病症	CFTR	S549I
CFTR 相关的病症	CFTR	S549N
CFTR 相关的病症	CFTR	S549R
CFTR 相关的病症	CFTR	S549R

疾病	基因	变体名称
CFTR 相关的病症	CFTR	T338I
CFTR 相关的病症	CFTR	V520F
CFTR 相关的病症	CFTR	W1089X
CFTR 相关的病症	CFTR	W1204X
CFTR 相关的病症	CFTR	W1204X
CFTR 相关的病症	CFTR	W1282X
CFTR 相关的病症	CFTR	Y1092X
CFTR 相关的病症	CFTR	Y122X
无脉络膜	CHM	c.1609+2dupT
无脉络膜	CHM	c.1609+2dupT
CLN3 相关的神经元蜡样质脂褐质沉积症	CLN3	c.461_677del
CLN3 相关的神经元蜡样质脂褐质沉积症	CLN3	c.461_677del
CLN3 相关的神经元蜡样质脂褐质沉积症	CLN3	c.791_1056del
CLN3 相关的神经元蜡样质脂褐质沉积症	CLN3	c.791_1056del
CLN5 相关的神经元蜡样质脂褐质沉积症	CLN5	c.1175_1176delAT
CLN5 相关的神经元蜡样质脂褐质沉积症	CLN5	c.1175_1176delAT
CLN5 相关的神经元蜡样质脂褐质沉积症	CLN5	c.225G>A
CLN8 相关的神经元蜡样质脂褐质沉积症	CLN8	c.70C>G
科恩综合征	VPS13B	c.3348_3349delCT
科恩综合征	VPS13B	c.3348_3349delCT
先天性白内障、面部变形和神经病变	CTDP1	IVS6+389C>T
先天性糖基化障碍 Ia	PMM2	p.F119L
先天性糖基化障碍 Ia	PMM2	p.R141H
先天性糖基化障碍 Ib	MPI	R295H, rs28928906
先天性芬兰型肾病	NPHS1	c.121_122del
先天性芬兰型肾病	NPHS1	c.121_122del
先天性芬兰型肾病	NPHS1	c.3325C>T
克罗恩病	NOD2	3020 ins C
克罗恩病	NOD2	3020 ins C
克罗恩病	NOD2	G908R, rs2066845
克罗恩病	NOD2	R702W, rs2066844
胱氨酸贮积症	CTNS	1035insC
胱氨酸贮积症	CTNS	1035insC
胱氨酸贮积症	CTNS	537del21

疾病	基因	变体名称
胱氨酸贮积症	CTNS	537del21
胱氨酸贮积症	CTNS	57kb 缺失
胱氨酸贮积症	CTNS	57kb 缺失
胱氨酸贮积症	CTNS	D205N
胱氨酸贮积症	CTNS	L158P
胱氨酸贮积症	CTNS	W138X
DFNA 9 (COCH)	COCH	P51S
糖尿病与听力损失	mtDNA	3234A>G
糖尿病与听力损失	mtDNA	3271T>C
糖尿病与听力损失	mtDNA	G8363A
糖尿病与听力损失	mtDNA	T14709C
早发性原发性张力障碍 (DYT1)	TOR1A	904_906delGAG
早发性原发性张力障碍 (DYT1)	TOR1A	904_906delGAG
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMB3	3024delT
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMB3	3024delT
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMB3	p.Q243X
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMB3	R144X
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMB3	R42X
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMB3	R635X
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMA3	R650X
交界性大疱性表皮松解症, Herlitz-Pearson 型	LAMC2	R95X
因子 V Leiden 血友病	F5	H1299R
因子 V Leiden 血友病	F5	R506Q, rs6025
因子 V R2 突变血栓性	F5	rs6027
因子 XI 缺乏	F11	E117X (576G>T)
因子 XI 缺乏	F11	F283L (1074T>C)
因子 XI 缺乏	F11	IVS14 +1G>A
因子 XI 缺乏	F11	IVS14del14
因子 XI 缺乏	F11	IVS14del14
因子 XIII 缺乏	F13A1	V34L, rs5985
家族性腺瘤性息肉病	APC	I1307K, rs1801155
家族性自主神经机能异常	IKBKAP	2507+6T>C
家族性自主神经机能异常	IKBKAP	P914L
家族性自主神经机能异常	IKBKAP	R696P
家族性高胆固醇血症 B 型	APOB	R3500Q, rs5742904

疾病	基因	变体名称
家族性高胆固醇血症 B 型	APOB	R3500W
家族性高胆固醇血症 B 型	APOB	R3531C, rs12713559
家族性地中海热	MEFV	A744S (2230G>T)
家族性地中海热	MEFV	del1692 (del2076_2078)
家族性地中海热	MEFV	del1692 (del2076_2078)
家族性地中海热	MEFV	E148Q (442 G>C), rs3743930
家族性地中海热	MEFV	E167D (501 G>C)
家族性地中海热	MEFV	F479L (1437 C>G)
家族性地中海热	MEFV	K695R (2084A>G)
家族性地中海热	MEFV	M680I (2040G>C)
家族性地中海热	MEFV	M694I (2082G>A), rs28940578
家族性地中海热	MEFV	M694V (2080A>G)
家族性地中海热	MEFV	P369S (1105 C>T), rs11466023
家族性地中海热	MEFV	R408Q (1223G>A), rs11466024
家族性地中海热	MEFV	R653H (1958G>A)
家族性地中海热	MEFV	R761H (2282G>A)
家族性地中海热	MEFV	T267I (800 C>T)
家族性地中海热	MEFV	V726A (2177T>C), rs28940579
FANCC 相关的范科尼贫血症	FANCC	322delG
FANCC 相关的范科尼贫血症	FANCC	322delG
FANCC 相关的范科尼贫血症	FANCC	IVS4+4A>T (711 +4A>T)
FANCC 相关的范科尼贫血症	FANCC	Q13X (37C>T)
FANCC 相关的范科尼贫血症	FANCC	R547X
FGFR1 相关的颅缝早闭	FGFR1	P252R
FGFR2 相关的颅缝早闭	FGFR2	P253R
FGFR2 相关的颅缝早闭	FGFR2	S252W
FGFR3 相关的颅缝早闭	FGFR3	A391E, rs28931615
FGFR3 相关的颅缝早闭	FGFR3	P250R, rs4647924
游离唾液酸储存障碍	SLC17A5	c.1007_1008delTA
游离唾液酸储存障碍	SLC17A5	c.1007_1008delTA
游离唾液酸储存障碍	SLC17A5	c.115C>T
额颞叶痴呆伴帕金森综合征-17	MAPT	IVS10+16
额颞叶痴呆伴帕金森综合征-17	MAPT	P301L
额颞叶痴呆伴帕金森综合征-17	MAPT	P301S
额颞叶痴呆伴帕金森综合征-17	MAPT	R406W
延胡索酸酶缺乏症	FH	c.1431_1433dupAAA
延胡索酸酶缺乏症	FH	c.1431_1433dupAAA
半乳糖血症	GALT	5.0 Kb 基因缺失
半乳糖血症	GALT	5.0 Kb 基因缺失
半乳糖血症	GALT	5'UTR-119del



疾病	基因	变体名称
半乳糖血症	GALT	5'UTR-119del
半乳糖血症	GALT	IVS2-2 A>G
半乳糖血症	GALT	K285N
半乳糖血症	GALT	L195P T>C
半乳糖血症	GALT	L218L
半乳糖血症	GALT	N314D, rs2070074
半乳糖血症	GALT	Phe171Ser
半乳糖血症	GALT	Q169K
半乳糖血症	GALT	Q188R
半乳糖血症	GALT	S135L
半乳糖血症	GALT	T138M C>T
半乳糖血症	GALT	X380R
半乳糖血症	GALT	Y209C A>G
戈谢病	GBA	1035insG
戈谢病	GBA	1035insG
戈谢病	GBA	84insG
戈谢病	GBA	84insG
戈谢病	GBA	D409H, rs1064651
戈谢病	GBA	D409V
戈谢病	GBA	IVS2(+1)G>A
戈谢病	GBA	L444P (1448T>C), rs35095275
戈谢病	GBA	N370S
戈谢病	GBA	R463C
戈谢病	GBA	R463H
戈谢病	GBA	R496H (1604G>A)
戈谢病	GBA	V394L
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	167delT
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	167delT
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	235delC
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	235delC
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	35delG
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	35delG
GJB2 相关的 DFNA 3 非综合征性听力丧失与耳聋	GJB2	IVS1+1G>A
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	101delAG

疾病	基因	变体名称
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	313del14
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	313del14
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	delE120
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	delE120
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	M34T, rs35887622
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	Q124X
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	R184P
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	V37I
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	W24X
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	W77R
GJB2 相关的 DFNB 1 非综合征性听力丧失和耳聋	GJB2	W77X
葡萄糖-6-磷酸脱氢酶缺乏症	G6PD	A335V
葡萄糖-6-磷酸脱氢酶缺乏症	G6PD	R459L
葡萄糖-6-磷酸脱氢酶缺乏症	G6PD	R459P
葡萄糖-6-磷酸脱氢酶缺乏症	G6PD	rs1050828, rs1050828
葡萄糖-6-磷酸脱氢酶缺乏症	G6PD	rs1050829, rs1050829
葡萄糖-6-磷酸脱氢酶缺乏症	G6PD	rs5030868, rs5030868
戊二酸血症 1 型	GCDH	A421V
戊二酸血症 1 型	GCDH	R402W
糖原贮积病 1a 型	G6PC	459insTA
糖原贮积病 1a 型	G6PC	459insTA
糖原贮积病 1a 型	G6PC	727G/T
糖原贮积病 1a 型	G6PC	del F327
糖原贮积病 1a 型	G6PC	del F327
糖原贮积病 1a 型	G6PC	G188R
糖原贮积病 1a 型	G6PC	G270V
糖原贮积病 1a 型	G6PC	Q242X
糖原贮积病 1a 型	G6PC	Q27fsdelC
糖原贮积病 1a 型	G6PC	Q27fsdelC
糖原贮积病 1a 型	G6PC	Q347X
糖原贮积病 1a 型	G6PC	R83C
糖原贮积病 1a 型	G6PC	R83H

疾病	基因	变体名称
糖原贮积病 1b 型	G6PT1	1211delCT
糖原贮积病 1b 型	G6PT1	A367T
糖原贮积病 1b 型	G6PT1	G339C
糖原贮积病 1b 型	G6PT1	G339D
糖原贮积病 1b 型	G6PT1	W118R
糖原贮积病 II 型	GAA	Arg854X
糖原贮积病 II 型	GAA	Asp645Glu, rs28940868
糖原贮积病 II 型	GAA	IVS1(-13t>g)
糖原贮积病 III 型	AGL	1484delT
糖原贮积病 III 型	AGL	1484delT
糖原贮积病 III 型	AGL	17delAG
糖原贮积病 III 型	AGL	17delAG
糖原贮积病 III 型	AGL	Q6X
糖原贮积病 V 型	PYGM	G204S
糖原贮积病 V 型	PYGM	K542T
糖原贮积病 V 型	PYGM	K542X
糖原贮积病 V 型	PYGM	R49X
GNE 相关的肌病	GNE	M712T, rs28937594
Gracile 综合征	BCS1L	c.232A>G, rs28937590
血红蛋白 S $\beta$ -地中海贫血	HBB	c.19G>A
血红蛋白 S $\beta$ -地中海贫血	HBB	c.20A>T
血红蛋白 S $\beta$ -地中海贫血	HBB	c.79G>A
血红蛋白 S $\beta$ -地中海贫血	HBB	Hb CS
血红蛋白 S $\beta$ -地中海贫血	HBB	Hb D
血红蛋白 S $\beta$ -地中海贫血	HBB	Hb O
遗传性果糖不耐症	ALDOB	A149P, rs1800546
遗传性果糖不耐症	ALDOB	A174D
遗传性果糖不耐症	ALDOB	$\delta$ 4E4
遗传性果糖不耐症	ALDOB	$\delta$ 4E4
遗传性果糖不耐症	ALDOB	N334K
遗传性果糖不耐症	ALDOB	Y203X
遗传性胰腺炎	PRSS1	A16V
遗传性胰腺炎	SPINK1	M1T
遗传性胰腺炎	PRSS1	N29I
遗传性胰腺炎	SPINK1	N34S, rs17107315
遗传性胰腺炎	PRSS1	R122C
遗传性胰腺炎	PRSS1	R122H
遗传性胸腺嘧啶-尿嘧啶尿症	DPYD	rs3918290
己糖胺酶 A 缺乏症	HEXA	1278insTATC
己糖胺酶 A 缺乏症	HEXA	1278insTATC
己糖胺酶 A 缺乏症	HEXA	G269S (805G>A)
己糖胺酶 A 缺乏症	HEXA	IVS12 +1G>C
己糖胺酶 A 缺乏症	HEXA	IVS7 +1G>A

疾病	基因	变体名称
己糖胺酶 A 缺乏症	HEXA	IVS9 +1G>A
己糖胺酶 A 缺乏症	HEXA	R178C
己糖胺酶 A 缺乏症	HEXA	R178H
己糖胺酶 A 缺乏症	HEXA	R247W (739C>T)
己糖胺酶 A 缺乏症	HEXA	R249W (745C>T)
HFE 相关的遗传性血色素沉着症	HFE	E168Q
HFE 相关的遗传性血色素沉着症	HFE	E168X
HFE 相关的遗传性血色素沉着症	HFE	HM971246, H63H
HFE 相关的遗传性血色素沉着症	HFE	P160delC
HFE 相关的遗传性血色素沉着症	HFE	P160delC
HFE 相关的遗传性血色素沉着症	HFE	Q127H, rs28934595
HFE 相关的遗传性血色素沉着症	HFE	Q283P
HFE 相关的遗传性血色素沉着症	HFE	rs1799945, rs1799945
HFE 相关的遗传性血色素沉着症	HFE	rs1800562, rs1800562
HFE 相关的遗传性血色素沉着症	HFE	rs1800730, rs1800730
HFE 相关的遗传性血色素沉着症	HFE	V53M
HFE 相关的遗传性血色素沉着症	HFE	V59M
HFE 相关的遗传性血色素沉着症	HFE	W169X
有汗性外胚层发育不良 2	GJB6	A88V, rs28937872
有汗性外胚层发育不良 2	GJB6	G11R
有汗性外胚层发育不良 2	GJB6	V37E
胱硫醚 $\beta$ 合酶缺乏引起的高胱氨酸尿症	CBS	G307S 919G->A
胱硫醚 $\beta$ 合酶缺乏引起的高胱氨酸尿症	CBS	I278T 833T->C, rs5742905
高钾性周期性麻痹 1 型	SCN4A	I693T
高钾性周期性麻痹 1 型	SCN4A	L689I
高钾性周期性麻痹 1 型	SCN4A	L689V
高钾性周期性麻痹 1 型	SCN4A	M1360V
高钾性周期性麻痹 1 型	SCN4A	M1592V
高钾性周期性麻痹 1 型	SCN4A	p.A1156T
高钾性周期性麻痹 1 型	SCN4A	p.M1370V
高钾性周期性麻痹 1 型	SCN4A	p.R1448C
高钾性周期性麻痹 1 型	SCN4A	p.T1313M
高钾性周期性麻痹 1 型	SCN4A	R675G
高钾性周期性麻痹 1 型	SCN4A	R675Q
高钾性周期性麻痹 1 型	SCN4A	R675W
高钾性周期性麻痹 1 型	SCN4A	T704M
高钾性周期性麻痹 1 型	SCN4A	V781I
高鸟氨酸血症-高氨血症-高瓜氨酸血症候群	SLC25A15	F188del
高鸟氨酸血症-高氨血症-高瓜氨酸血症候群	SLC25A15	F188del

疾病	基因	变体名称
高草酸尿, 原发性, 1 型	AGXT	33insC
高草酸尿, 原发性, 1 型	AGXT	33insC
高草酸尿, 原发性, 1 型	AGXT	F152I
高草酸尿, 原发性, 1 型	AGXT	G170R
高草酸尿, 原发性, 1 型	AGXT	I244T
高草酸尿, 原发性, 2 型	GRHPR	103delG
高草酸尿, 原发性, 2 型	GRHPR	103delG
软骨发育不良	FGFR3	Asn328Ile
软骨发育不良	FGFR3	I538V
软骨发育不良	FGFR3	K650M
软骨发育不良	FGFR3	K650N 1950G>T
软骨发育不良	FGFR3	K650Q
软骨发育不良	FGFR3	N540K 1620C>A
软骨发育不良	FGFR3	N540S
软骨发育不良	FGFR3	N540T
低钾型周期性麻痹 1 型	CACNA1S	R528G
低钾型周期性麻痹 1 型	CACNA1S	R528H
低钾型周期性麻痹 1 型	CACNA1S	rs28930068, rs28930068
低钾型周期性麻痹 1 型	CACNA1S	rs28930069, rs28930069
低钾型周期性麻痹 2 型	SCN4A	R669H
低钾型周期性麻痹 2 型	SCN4A	R672C
低钾型周期性麻痹 2 型	SCN4A	R672G
低钾型周期性麻痹 2 型	SCN4A	R672H
低钾型周期性麻痹 2 型	SCN4A	R672S
低磷酸酯酶症	ALPL	Asp361Val
低磷酸酯酶症	ALPL	c.1559delT
低磷酸酯酶症	ALPL	c.1559delT
低磷酸酯酶症	ALPL	E174K
低磷酸酯酶症	ALPL	G317D
低磷酸酯酶症	ALPL	Phe310Leu
异戊酸血症	IVD	A282V
异戊酸血症	IVD	rs28940889
克拉伯病	GALC	EX11-17DEL
克拉伯病	GALC	EX11-17DEL
克拉伯病	GALC	G270D
克拉伯病	GALC	rs1805078, rs1805078
克拉伯病	GALC	rs398607
Leber 遗传性视神经病变	mtDNA	14484T>C
Leber 遗传性视神经病变	mtDNA	15257G>A
Leber 遗传性视神经病变	mtDNA	G14459A
Leber 遗传性视神经病变	mtDNA	G3460A
Leber 遗传性视神经病变	mtDNA	m.11778G>A
Leber 遗传性视神经病变	mtDNA	m.13708G>A

疾病	基因	变体名称
Leber 遗传性视神经病变	mtDNA	m.15812G>A
Leber 遗传性视神经病变	mtDNA	m.3394T>C
Leber 遗传性视神经病变	mtDNA	m.4216T>C
Leber 遗传性视神经病变	mtDNA	m.4917A>G
Leigh 综合症, 法国-加拿大型	LRPPRC	A354V
LGMD2I	FKRP	L276I, rs28937900
长链 3-羟酰辅酶 A 脱氢酶缺乏症	HADHA	E474Q c.1528G>C
长链 3-羟酰辅酶 A 脱氢酶缺乏症	HADHA	Q342X 1132C>T
枫糖尿症 1A 型	BCKDHA	Y438N
枫糖尿症 1B 型	BCKDHB	E372X
枫糖尿症 1B 型	BCKDHB	G278S
枫糖尿症 1B 型	BCKDHB	R183P
McCune-Albright 综合征	GNAS	R201C
McCune-Albright 综合征	GNAS	R201G
McCune-Albright 综合征	GNAS	R201H
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	244insT
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	244insT
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	250C>T
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	583G>A
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	616C>T
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	617G>A
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	799G>A
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	K304E
中链酰基辅酶 A 脱氢酶缺乏症	ACADM	Y42H
巨脑性脑白质病伴皮质下囊肿	MLC1	135insC
巨脑性脑白质病伴皮质下囊肿	MLC1	135insC
MELAS	mtDNA	3243A>G
MELAS	mtDNA	3250T>C
MELAS	mtDNA	3252A>G
MELAS	mtDNA	A12770G
MELAS	mtDNA	C3256T
MELAS	mtDNA	G13513A
MELAS	mtDNA	T3291C
MELAS	mtDNA	T8356C
MELAS	mtDNA	T9957C
MERRF	mtDNA	8361G>A
MERRF	mtDNA	A8296G
MERRF	mtDNA	m.8344A>G
异染性脑白质营养不良	ARSA	c.459+1G>A
异染性脑白质营养不良	ARSA	p.P426L, rs28940893
异染性脑白质营养不良	ARSA	p.T274M
异染性脑白质营养不良	ARSA	P377L
线粒体心肌病	mtDNA	A3260T
线粒体心肌病	mtDNA	A4300G

疾病	基因	变体名称
线粒体心肌病	mtDNA	C3303T
线粒体心肌病	mtDNA	T9997C
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	5537insT
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	5537insT
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	8993T>C
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	8993T>G
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	C11777A
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	T10158C
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	T10191C
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	T8851C
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	T9176C
线粒体 DNA 相关的 Leigh 综合征和 NARP	mtDNA	T9176G
MTHFR 缺乏症	MTHFR	1298A>C
MTHFR 缺乏症	MTHFR	rs1801133, rs1801133
MTRNR1 相关的听力丧失和耳聋	mtDNA	1095T>C
MTRNR1 相关的听力丧失和耳聋	mtDNA	1494C>T
MTRNR1 相关的听力丧失和耳聋	mtDNA	1555A>G
MTRNR1 相关的听力丧失和耳聋	mtDNA	961T>G
MTRNR1 相关的听力丧失和耳聋	mtDNA	A7445G
MTTS1 相关的听力丧失和耳聋	mtDNA	7443A>G
MTTS1 相关的听力丧失和耳聋	mtDNA	7444G>A
MTTS1 相关的听力丧失和耳聋	mtDNA	7472insC
MTTS1 相关的听力丧失和耳聋	mtDNA	7472insC
MTTS1 相关的听力丧失和耳聋	mtDNA	7510T>C
MTTS1 相关的听力丧失和耳聋	mtDNA	7511T>C
MTTS1 相关的听力丧失和耳聋	mtDNA	7512T>C
粘脂质贮积症 IV	MCOLN1	delEx1 3 Ex7 (511>6944del)
粘脂质贮积症 IV	MCOLN1	delEx1 3 Ex7 (511>6944del)
粘脂质贮积症 IV	MCOLN1	IVS-2A>G
粘多糖贮积症 I 型	IDUA	c.46_57del
粘多糖贮积症 I 型	IDUA	c.46_57del
粘多糖贮积症 I 型	IDUA	p.A327P
粘多糖贮积症 I 型	IDUA	p.P533R
粘多糖贮积症 I 型	IDUA	Q70X

疾病	基因	变体名称
粘多糖贮积症 I 型	IDUA	W402X
粘多糖贮积症 IIIA 型	SGSH	p.R245H
粘多糖贮积症 IIIA 型	SGSH	p.R74C
粘多糖贮积症 IIIA 型	SGSH	p.S66W
粘多糖贮积症 VII 型	GUSB	p.D152N
多发性内分泌腺癌 2 型	RET	2047T>A
多发性内分泌腺癌 2 型	RET	2047T>A
多发性内分泌腺癌 2 型	RET	2047T>C
多发性内分泌腺癌 2 型	RET	2047T>G
多发性内分泌腺癌 2 型	RET	2048G>A
多发性内分泌腺癌 2 型	RET	A883F 2647 G>T
多发性内分泌腺癌 2 型	RET	Glu768Asp G>C
多发性内分泌腺癌 2 型	RET	M918T
肌肉-眼-脑疾病	POMGNT1	c.1539+1G>A
MYH 相关的息肉病	MUTYH	c.1376C>A
MYH 相关的息肉病	MUTYH	c.494A>G, rs34612342
MYH 相关的息肉病	GENE_SYMBOL _TBD	rs36053993, rs36053993
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	c.990delC
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	c.990delC
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	fsP330
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	fsP330
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	L302P
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	R496L
由于鞘磷脂酶缺乏引起的尼曼-皮克病	SMPD1	R608del
尼曼-皮克病 C1 型	NPC1	I1061T
Nijmegen 断裂综合症	NBN	657del5
Nijmegen 断裂综合症	NBN	657del5
Pallister-Hall 综合征	GLI3	2012delG
Pallister-Hall 综合征	GLI3	2012delG
Pallister-Hall 综合征	GLI3	2023delG
Pallister-Hall 综合征	GLI3	2023delG
Pendred 综合征	SLC26A4	1197delT
Pendred 综合征	SLC26A4	1197delT
Pendred 综合征	SLC26A4	E384G
Pendred 综合征	SLC26A4	IV58+1(G->A)



疾病	基因	变体名称
Pendred 综合征	SLC26A4	L236P
Pendred 综合征	SLC26A4	T416P
过氧化物酶体双功能酶缺乏症	HSD17B4	c.302+1G>C
过氧化物酶体双功能酶缺乏症	HSD17B4	c.303-1G>A
综合性精神发育障碍	NLGN4X	D396X
综合性精神发育障碍	NLGN4X	D396X
综合性精神发育障碍	NLGN4X	NLGN4X:1253delAG
综合性精神发育障碍	NLGN4X	NLGN4X:1253delAG
综合性精神发育障碍	NLGN3	R451C
苯丙氨酸羟化酶缺乏症	PAH	G272X
苯丙氨酸羟化酶缺乏症	PAH	I65T
苯丙氨酸羟化酶缺乏症	PAH	IVS12+1G>T
苯丙氨酸羟化酶缺乏症	PAH	L48S, rs5030841
苯丙氨酸羟化酶缺乏症	PAH	R158Q, rs5030843
苯丙氨酸羟化酶缺乏症	PAH	R252W, rs5030847
苯丙氨酸羟化酶缺乏症	PAH	R261Q, rs5030849
苯丙氨酸羟化酶缺乏症	PAH	R408Q, rs5030859
苯丙氨酸羟化酶缺乏症	PAH	R408W, rs5030858
苯丙氨酸羟化酶缺乏症	PAH	rs5030855
苯丙氨酸羟化酶缺乏症	PAH	rs5030861
苯丙氨酸羟化酶缺乏症	PAH	Y414C, rs5030860
纤溶酶原激活物抑制剂 I	SERPINE1	-844 G>A
纤溶酶原激活物抑制剂 I	SERPINE1	4G/5G
多囊肾病, 常染色体隐性	PKHD1	c.10412T>G
多囊肾病, 常染色体隐性	PKHD1	c.107C>T
多囊肾病, 常染色体隐性	PKHD1	c.1486C>T
多囊肾病, 常染色体隐性	PKHD1	c.5895dupA
多囊肾病, 常染色体隐性	PKHD1	c.5895dupA
多囊肾病, 常染色体隐性	PKHD1	c.9689delA
多囊肾病, 常染色体隐性	PKHD1	c.9689delA
PPT1 相关的神经元蜡样质脂褐质沉积症	PPT1	c.364A>T
PPT1 相关的神经元蜡样质脂褐质沉积症	PPT1	p.L10X
PPT1 相关的神经元蜡样质脂褐质沉积症	PPT1	p.R151X
PPT1 相关的神经元蜡样质脂褐质沉积症	PPT1	T75P
PROP1 相关的垂体激素缺乏症	PROP1	301-302delAG
PROP1 相关的垂体激素缺乏症	PROP1	301-302delAG
凝血酶原血栓形成倾向	F2	rs1799963
凝血酶原血栓形成倾向	F2	rs6025, rs6025
假性维生素 D 缺乏性佝偻病	CYP27B1	外显子 8 中 7 bp 重复

疾病	基因	变体名称
假性维生素 D 缺乏性佝偻病	CYP27B1	外显子 8 中 7 bp 重复
假性维生素 D 缺乏性佝偻病	CYP27B1	958delG
假性维生素 D 缺乏性佝偻病	CYP27B1	958delG
Rett 综合征	MECP2	806delG
Rett 综合征	MECP2	806delG
Rett 综合征	MECP2	A140V, rs28934908
Rett 综合征	MECP2	P152R
Rett 综合征	MECP2	P225R
Rett 综合征	MECP2	R106W, rs28934907
Rett 综合征	MECP2	R133C
Rett 综合征	MECP2	R168X
Rett 综合征	MECP2	R255X
Rett 综合征	MECP2	R270X
Rett 综合征	MECP2	R294X
Rett 综合征	MECP2	R306C, rs28935468
Rett 综合征	MECP2	S134C
Rett 综合征	MECP2	T158M, rs28934906
肢根点状软骨发育不良 1 型	PEX7	p.A218V
肢根点状软骨发育不良 1 型	PEX7	p.G217R
肢根点状软骨发育不良 1 型	PEX7	p.L292X, rs1805137
短链酰基辅酶 A 脱氢酶缺乏症	ACADS	c.511C>T, rs1800556
短链酰基辅酶 A 脱氢酶缺乏症	ACADS	c.625G>A
短链酰基辅酶 A 脱氢酶缺乏症	ACADS	R107C
Shwachman-Diamond 综合征	SBDS	183_184TA>CT
Shwachman-Diamond 综合征	SBDS	183_184TA>CT
Shwachman-Diamond 综合征	SBDS	258+2T>C
Sjogren-Larsson 综合征	ALDH3A2	c.943C>T
Smith-Lemli-Opitz 综合征	DHCR7	C380Y
Smith-Lemli-Opitz 综合征	DHCR7	IVS8-1G>C
Smith-Lemli-Opitz 综合征	DHCR7	L109P
Smith-Lemli-Opitz 综合征	DHCR7	L157P
Smith-Lemli-Opitz 综合征	DHCR7	R352Q
Smith-Lemli-Opitz 综合征	DHCR7	R352W
Smith-Lemli-Opitz 综合征	DHCR7	R404C
Smith-Lemli-Opitz 综合征	DHCR7	R446Q
Smith-Lemli-Opitz 综合征	DHCR7	T93M
Smith-Lemli-Opitz 综合征	DHCR7	V326L
Smith-Lemli-Opitz 综合征	DHCR7	W151X
Smith-Lemli-Opitz 综合征	DHCR7	W151X
痉挛性截瘫 13	HSPD1	V72I
硫酸盐转运蛋白相关的骨软骨发育不良	SLC26A2	340delV

疾病	基因	变体名称
硫酸盐转运蛋白相关的骨软骨发育不良	SLC26A2	340delV
硫酸盐转运蛋白相关的骨软骨发育不良	SLC26A2	c.837C>T
硫酸盐转运蛋白相关的骨软骨发育不良	SLC26A2	C653S
硫酸盐转运蛋白相关的骨软骨发育不良	SLC26A2	IVS1+2T>C
硫酸盐转运蛋白相关的骨软骨发育不良	SLC26A2	R178X
TFR2 相关的遗传性血色素沉着症	TFR2	AVAQ594-597del
TFR2 相关的遗传性血色素沉着症	TFR2	AVAQ594-597del
TFR2 相关的遗传性血色素沉着症	TFR2	E60X
TFR2 相关的遗传性血色素沉着症	TFR2	E60X
TFR2 相关的遗传性血色素沉着症	TFR2	M172K
TFR2 相关的遗传性血色素沉着症	TFR2	Y250X
致死性发育不全	FGFR3	G370C
致死性发育不全	FGFR3	K650E
致死性发育不全	FGFR3	R248C
致死性发育不全	FGFR3	S249C
致死性发育不全	FGFR3	S371C
致死性发育不全	FGFR3	X807C A>T
致死性发育不全	FGFR3	X807G
致死性发育不全	FGFR3	X807L
致死性发育不全	FGFR3	X807R
致死性发育不全	FGFR3	X807S
致死性发育不全	FGFR3	X807W
致死性发育不全	FGFR3	Y373C
TPP1 相关的神经元蜡样质脂褐质沉积症	TPP1	c.509-1G>A
TPP1 相关的神经元蜡样质-脂褐质沉积症	TPP1	c.509-1G>C
TPP1 相关的神经元蜡样质-脂褐质沉积症	TPP1	G284V
TPP1 相关的神经元蜡样质-脂褐质沉积症	TPP1	p.R208X
转甲状腺素蛋白淀粉样变性	TTR	c.148G>A
酪氨酸羟化酶缺乏的 DRD	TH	L205P
酪氨酸羟化酶缺乏的 DRD	TH	R202H
酪氨酸血症 I 型	FAH	E357X
酪氨酸血症 I 型	FAH	IVS12+5 G>A
酪氨酸血症 I 型	FAH	IVS7-6 T>G
酪氨酸血症 I 型	FAH	IVS8-1G>C

疾病	基因	变体名称
酪氨酸血症 I 型	FAH	p.W262X
酪氨酸血症 I 型	FAH	P261L
酪氨酸血症 I 型	FAH	Q64H
Wilson 病	ATP7B	1340del4
Wilson 病	ATP7B	3402delC
Wilson 病	ATP7B	3402delC
Wilson 病	ATP7B	H1069Q
Wilson 病	ATP7B	R778G
Wilson 病	ATP7B	W779X
Wilson 病	ATP7B	W779X
X 染色体连锁型青年性视网膜裂损症	RS1	E72K
X 染色体连锁型青年性视网膜裂损症	RS1	G109R
X 染色体连锁型青年性视网膜裂损症	RS1	G74V
Zellweger 综合征谱	PEX1	c.2097_2098insT
Zellweger 综合征谱	PEX1	c.2097_2098insT
Zellweger 综合征谱	PEX1	c.2916delA
Zellweger 综合征谱	PEX1	c.2916delA
Zellweger 综合征谱	PEX1	p.G843D

表2

疾病/障碍	基因
瘤形成	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR $\alpha$ ; PPAR $\gamma$ ; WT1 (维尔姆斯瘤); FGF 受体家族成员(5 个成员: 1, 2, 3, 4, 5); CDKN2a; APC; RB (视网膜母细胞瘤); MEN1; VHL; BRCA1; BRCA2; AR (雄激素受体); TSG101; IGF; IGF 受体; Igf1 (4 种变体); Igf2 (3 种变体); Igf 1 受体; Igf 2 受体; Bax; Bcl2; 胱天蛋白酶家族(9 个成员: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
老年性黄斑变性	Aber; Ccl2; Cc2; cp (血浆铜蓝蛋白); Timp3; 组织蛋白酶 D; Vldlr; Ccr2
精神分裂症	神经调节蛋白 1 (Nrg1); Erb4 (神经调节蛋白的受体); 复蛋白 1 (Cplx1); Tph1 色氨酸羟化酶; Tph2 色氨酸羟化酶 2; 神经元表面蛋白 1; GSK3; GSK3a;

疾病/障碍	基因
	GSK3b
三核苷酸重复疾病	HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's Dx); FXN/X25 (Friedrich 共济失调); ATX3 (Machado-Joseph's Dx); ATXN1 与 ATXN2 (脊髓小脑共济失调); DMPK (肌强直性营养不良); 萎缩蛋白-1 和 Atn1 (DRPLA Dx); CBP (Creb-BP –全球不稳定); VLDLR (Alzheimer's); Atxn7; Atxn10
脆性 X 综合征	FMR2; FXR1; FXR2; mGLUR5
分泌相关疾病	APH-1 ( $\alpha$ 和 $\beta$ ); 早老蛋白(Psen1); 呆蛋白(Ncstn); PEN-2
朊病毒相关疾病	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c)
药物成瘾	Prkce (醇); Drd2; Drd4; ABAT (醇); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (醇)
孤独症	Mecp2; BZRAP1; MDGA2; Sema5A; 神经元表面蛋白 1; 脆性 X (FMR2 (AFF2); FXR1; FXR2; Mglur5)
阿尔茨海默病	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; 簇蛋白; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1, 水通道蛋白 1); Uchl1; Uchl3; APP
炎症	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
帕金森病	x-突触核蛋白; DJ-1; LRRK2; 帕金蛋白; PINK1

[0207] 在一些实施方案中,方法进一步包括基于判定步骤来诊断受试者的步骤,例如诊断患有与检测到的因果遗传变体相关的疾病的受试者,或报告患者已患或将患此疾病的可能性。本文提供了疾病、相关基因和相关序列变体的实例。在一些实施方案中,通过报告生成器报告结果,如本文所述。

[0208] 在一些实施方案中,一种或多种因果遗传变体是与特定类型或阶段的癌症相关的序列变体,或是具有特定特征(例如转移潜能、耐药性、药物反应性)的癌症的序列变体。在一些实施方案中,本公开内容提供了确定预后的方法,如已知某些突变与患者结果相关的情况。例如,ctDNA已被证明是比传统癌抗原53(CA-53)和循环肿瘤细胞计数(参见,例如,

Dawson等人, *N Engl J Med* 368:1199 (2013)) 更好的乳腺癌预后生物标志物。此外,本公开内容的方法可用于治疗决策、指导和监测,以及癌症疗法的开发和临床试验。例如,可以通过将采用特定疗法治疗之前、期间和之后得到的患者ctDNA样品进行比较来监测治疗功效,该特定疗法是例如分子靶向疗法(单克隆药物)、化疗药物、放疗方案等或其组合。例如,可以监测ctDNA以确定某些突变在治疗后是否增加或减少,是否出现新突变等,与跟踪患者症状的监测方法相比,这可以让医生在更短的时间内改变治疗(例如,继续、停止或改变治疗)。在一些实施方案中,方法进一步包括基于判定步骤来诊断受试者的步骤,例如诊断具有与检测到的序列变体相关的特定阶段或类型的癌症的受试者,或报告患者已患或将患此疾病的可能性。

[0209] 例如,对于基于分子标志物(例如赫赛汀和her2/neu状态)的特异性靶向患者的疗法,对患者进行测试,以确定他们的肿瘤中是否存在某些突变,并且这些突变可以用于预测对治疗的响应或抵抗,并指导决定是否使用该疗法。因此,在治疗过程中检测和监测ctDNA对指导治疗选择可能非常有用。发现一些原发性(治疗前)或继发性(治疗后)的癌症突变是癌症对某些疗法产生抗性的原因(Misale等人, *Nature* 486 (7404):532 (2012))。

[0210] 与一种或多种癌症相关且可用于诊断、预后或治疗决策的多种序列变体是已知的。本公开内容的方法中可用的具有肿瘤学意义的合适靶序列包括但不限于TP53基因、ALK基因、KRAS基因、PIK3CA基因、BRAF基因、EGFR基因和KIT基因的改变。可以被特异性扩增和/或针对序列变体进行特异性分析的靶序列可以是癌症相关基因的全部或部分。在一些实施方案中,在TP53基因中鉴定了一种或多种序列变体。TP53是人类癌症中最常见的突变基因之一,例如,TP53突变可见于45%的卵巢癌、43%的大肠癌和42%的上呼吸消化道癌中(参见,例如,M.Olivier等人, *TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use*. Cold Spring Harb Perspect Biol. 2010年1月;2(1))。表征TP53的突变状态可有助于临床诊断,提供预后价值,并影响对癌症患者的治疗。例如,TP53突变可用作患有源自神经胶质细胞的大脑肿瘤的患者预后不良的预测因子,以及患有慢性淋巴细胞白血病的患者的疾病快速进展的预测因子(参见McLendon RE等人, *Cancer*. 2005年10月15日;104(8):1693-9; Dicker F等人, *Leukemia*. 2009年1月;23(1):117-24)。序列变异可以发生在基因的任何地方。因此,可以在此评估全部或部分TP53基因。即,如本文其他地方所述,当使用靶标特定组分(例如,靶标特异性引物)时,可以使用多个TP53特异性序列,例如以扩增和检测跨越基因的片段,而不仅是可用于选定靶标的一个或多个选定子序列(如突变“热点”)。或者,可以设计与一个或多个选定子序列的上游或下游(该核苷酸或核苷酸区域与一类受试者中突变率增加有关,也包括在术语“热点”中)杂交的靶标特异性引物。可以设计跨越该子序列的标准引物,并且/或者可以设计与该子序列的上游或下游杂交的B2B引物。

[0211] 在一些实施方案中,在ALK基因的全部或部分中鉴定一种或多种序列变体。据报道,多达7%的肺肿瘤中存在ALK融合,其中一些与EGFR酪氨酸激酶抑制剂(TKI)耐药性有关(参见,例如,Shaw等人, *J Clin Oncol*. 2009年9月10日;27(26):4247-4253)。截止至2013年,在具有ALK酪氨酸激酶抑制剂(TKI)继发性耐药性的患者中发现了跨越整个ALK酪氨酸激酶结构域的几个不同的点突变(Katayama R 2012 *Sci Transl Med*. 2012年2月8日;4(120))。因此,ALK基因中的突变检测可用于辅助癌症疗法决策。

[0212] 在一些实施方案中,在KRAS基因的全部或部分中鉴定一种或多种序列变体。据报

道,大约15-25%的肺腺癌患者和40%的结肠直肠癌患者携带肿瘤相关的KRAS突变(参见,例如,Neuman 2009,Pathol Res Pract.2009;205(12):858-62)。大多数突变位于KRAS基因的密码子12、13和61。这些突变激活KRAS信号传导通路,从而触发肿瘤细胞的生长和增殖。一些研究表明,具有携带KRAS突变的肿瘤的患者无法受益于单独或与化疗联合的抗EGFR抗体疗法(参见,例如,Amado等人,2008J Clin Oncol.2008年4月1日;26(10):1626-34, Bokemeyer等人,2009J Clin Oncol.2009年2月10日;27(5):663-71)。可以为了鉴定序列变异而靶向的一个特定的序列变异“热点”是基因的35位。KRAS序列变体的鉴定可用于治疗选择,如结肠直肠癌受试者的治疗选择。

[0213] 在一些实施方案中,在PIK3CA基因的全部或部分中鉴定一种或多种序列变体。PIK3CA的体细胞突变经常出现在各种类型的癌症中,例如,在10-30%的结肠直肠癌中(参见,例如,Samuels等人,2004Science.2004年4月23日;304(5670):554.)。这些突变最常位于外显子9(螺旋结构域)和外显子20(激酶结构域)内的两个“热点”区域内,其可以被特异性地靶向用于扩增和/或分析以供序列变体检测。位置3140也可以被特异性靶向。

[0214] 在一些实施方案中,在BRAF基因的全部或部分中鉴定一种或多种序列变体。据报道,在所有恶性黑色素瘤中有近50%都携带BRAF中的体细胞突变(参见,例如,Maldonado等人,J Natl Cancer Inst.2003年12月17日;95(24):1878-90)。BRAF突变可见于所有黑色素瘤亚型中,但最常见于源自没有慢性阳光诱导损伤的皮肤的黑素瘤中。在黑素瘤中最常见的BRAF突变是错义突变V600E,其将600位的缬氨酸用谷氨酰胺置换。BRAF V600E突变与BRAF抑制剂疗法的临床益处相关。BRAF突变的检测可用于黑素瘤治疗的选择和对靶向疗法的耐药性的研究。

[0215] 在一些实施方案中,在EGFR基因的全部或部分中鉴定一种或多种序列变体。EGFR突变通常与非小细胞肺癌相关(在美国约为10%,在东亚约为35%;参见,例如,Pao等人,Proc Natl Acad Sci USA.2004年9月7日;101(36):13306-11)。这些突变通常发生在EGFR外显子18-21内,并且通常是杂合的。这些突变中大约90%是外显子19缺失或外显子21L858R点突变。

[0216] 在一些实施方案中,在KIT基因的全部或部分中鉴定一种或多种序列变体。据报道,近85%的胃肠道间质瘤(GIST)携带KIT突变(参见,例如,Heinrich等人,2003J Clin Oncol.2003年12月1日;21(23):4342-9)。大多数KIT突变发生在近膜结构域(外显子11,70%)、细胞外二聚化基序(外显子9,10-15%)、酪氨酸激酶I(TKI)结构域(外显子13,1-3%)和酪氨酸激酶2(TK2)结构域和激活环(外显子17,1-3%)中。在靶向疗法伊马替尼后和患者对疗法产生耐药性后,通常会发现继发性KIT突变。

[0217] 可根据本文所述的方法来分析其全部或部分的序列变体的癌症相关基因的其他非限制性实例包括但不限于PTEN;ATM;ATR;EGFR;ERBB2;ERBB3;ERBB4;Notch1;Notch2;Notch3;Notch4;AKT;AKT2;AKT3;HIF HIF1a;HIF3a;Met;HRG;Bcl2;PPAR $\alpha$ ;PPAR $\gamma$ ;WT1(维尔姆斯瘤);FGF受体家族成员(5个成员:1、2、3、4、5);CDKN2a;APC;RB(视网膜母细胞瘤);MEN1;VHL;BRCA1;BRCA2;AR(雄激素受体);TSG101;IGF;IGF受体;Igfl(4种变体);Igf2(3种变体);Igf 1受体;Igf 2受体;Bax;Bcl2;胱天蛋白酶家族(9个成员:1、2、3、4、6、7、8、9、12);Kras;以及Apc。其他实例在本文其他地方提供。可以基于根据本文公开的方法判定一种或多种序列变体来诊断的癌症的实例包括但不限于棘皮瘤、腺泡细胞癌、听神经瘤、肢端

着色斑性黑素瘤、顶端螺旋瘤、急性嗜酸细胞白血病、急性淋巴母细胞性白血病、急性巨核母细胞性白血病、急性单核细胞性白血病、成熟的急性髓细胞白血病、急性髓样树突细胞白血病、急性髓样白血病、急性早幼粒细胞性白血病、黏质瘤、腺癌、腺样囊性癌、腺瘤、牙源性腺瘤样瘤、肾上腺皮质癌、成人T-细胞白血病、侵袭性NK-细胞白血病、AIDS相关癌症、AIDS相关淋巴瘤、软组织腺泡状肉瘤、成釉细胞纤维瘤、肛门癌、间变性大细胞淋巴瘤、甲状腺未分化癌、血管免疫母细胞性T细胞淋巴瘤、血管肌脂瘤、血管肉瘤、阑尾癌、星形细胞瘤、非典型性畸胎样横纹肌样瘤、基底细胞癌、基底细胞样癌、B-细胞白血病、B细胞淋巴瘤、Bellini管癌、胆道癌、膀胱癌、母细胞瘤、骨癌、骨瘤、脑干胶质瘤、脑瘤、乳腺癌、Brenner瘤、支气管瘤、细支气管肺泡癌、棕色瘤、伯基特淋巴瘤、原发灶不明的癌症、类癌瘤、癌、原位癌、阴茎癌、原发灶不明癌、癌肉瘤、Castleman病、中枢神经系统胚胎瘤、小脑星形细胞瘤、大脑星形细胞瘤、宫颈癌、胆管细胞癌、软骨瘤、软骨肉瘤、脊索瘤、绒毛膜癌、脉络丛乳头状瘤、慢性淋巴细胞白血病、慢性单核细胞性白血病、慢性髓性白血病、慢性骨髓增生性疾病、慢性嗜中性粒细胞白血病、明细胞肿瘤、结肠癌、结直肠癌、颅咽管瘤、皮肤T细胞淋巴瘤、Degos病、隆凸性皮肤纤维肉瘤、皮样囊肿、结缔组织增生性小圆细胞瘤、弥漫性大B细胞淋巴瘤、胚胎发育不良神经上皮瘤、胚胎癌、内胚窦瘤、子宫内膜癌、子宫内膜子宫癌、子宫内膜样肿瘤、肠病相关的T细胞淋巴瘤、室管膜母细胞瘤、室管膜瘤、上皮样肉瘤、红白血病、食管癌、鼻腔神经胶质瘤、尤因家族肿瘤、尤因家族肉瘤、尤因肉瘤、颅外生殖细胞瘤、性腺外生殖细胞瘤、肝外胆管癌、非乳腺性佩吉特病、输卵管癌、胎中胎、纤维瘤、纤维肉瘤、滤泡性淋巴瘤、滤泡性甲状腺癌、胆囊癌、胆囊癌、神经节胶质瘤、神经节瘤、胃癌、胃淋巴瘤、胃肠癌、胃肠类癌瘤、胃肠间质瘤、胃肠间质瘤、生殖细胞肿瘤、生殖细胞瘤、妊娠性绒毛膜癌、妊娠滋养细胞瘤、骨巨细胞瘤、多形性胶质母细胞瘤、胶质瘤、大脑胶质瘤病、血管球瘤、胰高血糖素瘤、成性腺细胞瘤、粒层细胞瘤、毛细胞白血病、毛细胞白血病、头颈癌、头颈癌、心脏癌、血管母细胞瘤、血管外皮细胞瘤、血管肉瘤、恶性血液肿瘤、肝细胞癌、肝脾T细胞淋巴瘤、遗传性乳腺癌-卵巢癌综合征、霍奇金淋巴瘤、霍奇金淋巴瘤、下咽癌、下丘脑胶质瘤、炎性乳腺癌、眼内黑素瘤、胰岛细胞癌、胰岛细胞瘤、幼年型粒单核细胞白血病、卡波西肉瘤、卡波西肉瘤、肾癌、Klatskin瘤、Krukenberg瘤、喉癌、喉癌、恶性雀斑样痣黑素瘤、白血病、白血病、唇癌和口腔癌、脂肪肉瘤、肺癌、黄体瘤、淋巴管瘤、淋巴管肉瘤、淋巴上皮瘤、淋巴样白血病、淋巴瘤、巨球蛋白血症、恶性纤维组织细胞瘤、恶性纤维组织细胞瘤、骨恶性纤维组织细胞瘤、恶性胶质瘤、恶性间皮瘤、恶性外周神经鞘瘤、恶性棒状瘤、恶性特利顿瘤、MALT淋巴瘤、套细胞淋巴瘤、肥大细胞淋巴瘤、纵膈生殖细胞瘤、纵膈肿瘤、甲状腺髓样癌、髓母细胞瘤、髓母细胞瘤、髓上皮瘤、黑素瘤、黑素瘤、脑膜瘤、Merkel细胞癌、间皮瘤、间皮瘤、原发灶隐匿转移性鳞状颈癌、转移性尿路上皮癌、混合Mullerian瘤、单核细胞性白血病、口腔癌、粘液瘤、多发性内分泌瘤综合征、多发性骨髓瘤、多发性骨髓瘤、蕈样肉芽肿、蕈样肉芽肿、骨髓增生异常疾病、骨髓增生异常综合征、髓样白血病、髓样肉瘤、骨髓增生性疾病、粘液瘤、鼻腔癌、鼻咽癌、鼻咽癌、赘生物、神经鞘瘤、神经母细胞瘤、神经母细胞瘤、神经纤维瘤、神经瘤、结节型黑素瘤、非霍奇金淋巴瘤、非霍奇金淋巴瘤、非黑素瘤皮肤癌、非小细胞肺癌、眼肿瘤、少突星形细胞瘤、少突神经胶质瘤、嗜酸粒细胞腺瘤、视神经鞘脑膜瘤、口癌、口癌、口咽癌、骨肉瘤、骨肉瘤、卵巢癌、卵巢癌、卵巢上皮癌、卵巢生殖细胞瘤、卵巢低恶性潜能肿瘤、乳腺Paget病、Pancoast瘤、胰腺癌、胰腺癌、乳头状甲状腺癌、乳头状瘤



病、副神经节瘤、副鼻窦癌、甲状旁腺癌、阴茎癌、血管周上皮样细胞瘤、咽癌、嗜铬细胞瘤、中度分化的松果体实质瘤、成松果体细胞瘤、垂体细胞瘤、垂体腺瘤、垂体瘤、浆细胞瘤、胸膜肺母细胞瘤、多胚瘤、前体T成淋巴细胞淋巴瘤、原发性中枢神经系统淋巴瘤、原发性渗出性淋巴瘤、原发性肝细胞癌、原发性肝癌、原发性腹膜癌、原始神经外胚层瘤、前列腺癌、腹膜假粘液瘤、直肠癌、肾细胞癌、涉及染色体15上NUT基因的呼吸道癌、视网膜母细胞瘤、横纹肌瘤、横纹肌肉瘤、Richter转化、骶尾部畸胎瘤、唾液腺癌、肉瘤、施万鞘瘤神经鞘瘤病(Schwannomatosis)、皮脂腺癌、继发性肿瘤、精原细胞瘤、浆液性肿瘤、Sertoli-Leydig细胞瘤、性索间质瘤、Sezary综合征、Signet细胞癌、皮肤癌、小蓝圆细胞瘤、小细胞癌、小细胞肺癌、小细胞淋巴瘤、小肠癌、软组织肉瘤、生长抑素瘤、煤烟疣、脊髓瘤、脊髓肿瘤、脾边缘区淋巴瘤、鳞状细胞癌、胃癌、表浅扩散性黑素瘤、幕上原始神经外胚层瘤、表面上皮-间质瘤、滑膜肉瘤、T细胞急性淋巴母细胞性白血病、T细胞大颗粒淋巴细胞白血病、T细胞白血病、T细胞淋巴瘤、T细胞幼淋巴细胞白血病、畸胎瘤、晚期淋巴瘤、睾丸癌、泡膜细胞瘤、喉癌、胸腺癌、胸腺瘤、甲状腺癌、肾盂和输尿管移行细胞癌、移行细胞癌、脐尿管癌、尿道癌、泌尿生殖系统肿瘤、子宫肉瘤、葡萄膜黑素瘤、阴道癌、Verner Morrison综合征、疣状癌、视通路胶质瘤、外阴癌、瓦尔登斯特伦巨球蛋白血症、Warthin瘤、维尔姆斯瘤,及其组合。与癌症相关的特定序列变体的非限制性实例在表3中提供。

表3

基因	突变 氨基酸 核苷酸	COSMIC 黑素瘤中的突变%	临床相关性
BRAF	V600E 1799T>A	50% (BRAF 突变中 80-90%)	增加对 BRAF 抑制剂的敏感性 增加对 MEK 抑制剂的敏感性
BRAF	V600E 1799_1800delTGinsAA	50% (BRAF 突变中 80-90%)	增加对 BRAF 抑制剂的敏感性 增加对 MEK 抑制剂的敏感性
BRAF	V600R 1798_1799delGTinsAG	50% (BRAF 突变中 <5%)	增加对 BRAF 抑制剂的敏感性 响应于 MEK 抑制剂
BRAF	V600M 1798G>A	50% (BRAF 突变中 <1%)	增加对 BRAF 抑制剂的敏感性 响应于 MEK 抑制剂
BRAF	V600K 1798_1799delGTinsAA	50% (BRAF 突变中 5%)	增加对 BRAF 抑制剂的敏感性 增加对 MEK 抑制剂的敏感性
BRAF	V600G 1799T>G	50% (BRAF 突变中 <1%)	增加对 BRAF 抑制剂的敏感性 响应于 MEK 抑制剂
BRAF	V600D 1799-800delTGinsAT	50% (BRAF 突变中 <5%)	增加对 BRAF 抑制剂的敏感性 响应于 MEK 抑制剂
BRAF	L597V 1789C>G	50% (BRAF 突变中 1%)	响应于 BRAF 抑制剂 响应于 MEK 抑制剂
BRAF	L597S 1789_1790delCTinsTC	50% (BRAF 突变中 <1%)	响应于 BRAF 抑制剂 响应于 MEK 抑制剂
BRAF	L597Q 1790T>A	50% (BRAF 突变中 <1%)	响应于 BRAF 抑制剂 响应于 MEK 抑制剂
BRAF	L597R 1790T>G	50% (BRAF 突变中 <1%)	响应于 BRAF 抑制剂 响应于 MEK 抑制剂
BRAF	D594N 1799_1780delTGinsGA	50% (BRAF 突变中 <1%)	不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
BRAF	D594H 1780C>G	50% (BRAF 突变中 <1%)	不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
BRAF	D594E 1782T>A	50% (BRAF 突变中 <1%)	不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂

基因	突变 氨基酸 核苷酸	COSMIC 黑色素瘤中的突变%	临床相关性
BRAF	D594E 1782T>G	50% (BRAF 突变中 <1%)	不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
BRAF	D594G 1781A>G	50% (BRAF 突变中 <1%)	不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
BRAF	D594V 1781A>T	50% (BRAF 突变中 <1%)	不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
NRAS	G12D 35G>A	13-25% (NRAS 突变中 4%)	细胞毒性化疗
NRAS	G13R 37G>C	13-25% (NRAS 突变中 2%)	细胞毒性化疗 不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
NRAS	G13D 38G>A	13-25% (NRAS 突变中 2%)	细胞毒性化疗 不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
NRAS	G13V 38G>T	13-25% (NRAS 突变中 2%)	细胞毒性化疗 不响应于 BRAF 抑制剂 可能响应于 MEK 抑制剂
NRAS	Q61K 181C>A	13-25% (NRAS 突变中 34%)	响应于 MEK 抑制剂
NRAS	Q61L 182A>T	13-25% (NRAS 突变中 8%)	响应于 MEK 抑制剂
NRAS	Q61R 182A>G	13-25% (NRAS 突变中 35%)	响应于 MEK 抑制剂
NRAS	Q61H 183A>C	13-25% (NRAS 突变中 2%)	响应于 MEK 抑制剂
NRAS	Q61H 183A>T	13-25% (NRAS 突变中 2%)	响应于 MEK 抑制剂
CTNNB1	S37F 110C>T	原发性葡萄膜黑色素瘤中 约 2-3%(CTNNB1 突变 中 46%)	
CNA11	Q209L 626A>T	原发性葡萄膜黑色素瘤中 34%(CNA11 突变中 92%)	
CNA11	Q209P 626A>C	原发性葡萄膜黑色素瘤中 34%(CNA11 突变中 1%)	
GNAQ	Q209L 626A>T	原发性葡萄膜黑色素瘤中 约 50% (CNAQ 突变中约 33%)	对 MEK 抑制剂敏感
GNAQ	Q209P 626A>C	原发性葡萄膜黑色素瘤中 约 50% (CNAQ 突变中约 64%)	对 MEK 抑制剂敏感 对 MEK 抑制剂敏感
GNAQ	Q209R 626A>G	原发性葡萄膜黑色素瘤中 约 50% (CNAQ 突变中约 2%)	
KIT	K624E	KIT 突变恶性黑色素瘤中	增加对 KIT 抑制剂的敏感性

基因	突变 氨基酸 核苷酸	COSMIC 黑素瘤中的突变%	临床相关性
	1924A>G	约 20%	
KIT	D816H 2446G>C	KIT 突变恶性黑素瘤中 约 5%	增加对 KIT 抑制剂的敏感性
KIT	L567P 1727T>C	KIT 突变恶性黑素瘤中 约 25%	增加对 KIT 抑制剂的敏感性
KIT	V559A 1676T>C	KIT 突变恶性黑素瘤中 约 20%	增加对 KIT 抑制剂的敏感性
KIT	V559D 1676T>A	KIT 突变恶性黑素瘤中 约 5%	增加对 KIT 抑制剂的敏感性
KIT	W557R 1669T>C	KIT 突变恶性黑素瘤中 约 10%	增加对 KIT 抑制剂的敏感性
KIT	W557R 1669T>A	KIT 突变恶性黑素瘤中 约 10%	增加对 KIT 抑制剂的敏感性

[0218] 此外,本文公开的方法和组合物可用于发现与一种或多种癌症类型、阶段或癌症特征相关的新的罕见突变。例如,根据本公开内容,可以使共享分析特征(例如特定疾病、癌症类型、癌症阶段等)的个体群体经历检测序列变体的方法,以鉴定序列变体或序列变体的类型(例如特定基因或基因部分的突变)。相比于没有该特征的个体,在共享该特征的一组个体中被鉴定为以统计学上显著更高的频率出现的变体序列可以被指定与该特征有一定程度的关联。如此鉴定的序列变体或序列变体类型随后可用于诊断或治疗被发现携带这些它们的个体。

[0219] 其他治疗应用包括非侵入性胎儿诊断。胎儿DNA可以在孕妇的血液中找到。本文所述的方法和组合物可用于鉴定循环胎儿DNA中的序列变体,因此可用于诊断胎儿中的一种或多种遗传疾病,如与一种或多种因果遗传变体相关的遗传疾病。本文描述了因果遗传变体的非限制性实例,包括三体性、囊性纤维化、镰状细胞贫血和Tay-Saks病。在该实施方案中,母亲可以提供对照样品和用于比较的血液样品。对照样品可以是任何合适的组织,并且通常将会被处理以提取细胞DNA,然后可以对该DNA进行测序以提供参考序列。随后可以将对应于胎儿基因组DNA的cfDNA序列鉴定为相对于母体参照的序列变体。父亲也可以提供参考样品,以帮助鉴定胎儿序列和序列变体。

[0220] 更进一步的治疗应用包括检测外源多核苷酸,如来自病原体(例如细菌、病毒、真菌和微生物)的外源多核苷酸,该信息可告知诊断和治疗选择。例如,一些HIV亚型与耐药性相关(参见,例如,hivdb.stanford.edu/pages/genotype-rx)。类似地,HCV分型、亚型分型和同种型突变也可以使用本公开内容的方法和组合物来实现。此外,在HPV亚型与宫颈癌风险相关的情况下,这种诊断可以进一步告知癌症风险的评估。可检测的病毒的进一步非限制性实例包括乙型肝炎病毒(HBV)、土拨鼠肝炎病毒、地松鼠(肝脱氧核糖核酸病毒科)肝炎病毒、鸭乙型肝炎病毒、苍鹭乙型肝炎病毒、疱疹病毒属单纯疱疹病毒(HSV)1型和2型、水痘-带状疱疹病毒、巨细胞病毒(CMV)、人巨细胞病毒(HCMV)、小鼠巨细胞病毒(MCMV)、豚鼠巨细胞病毒(GPCMV)、EB病毒(EBV)、人疱疹病毒6(HHV变体A和B)、人疱疹病毒7(HHV-7)、人疱疹病毒8(HHV-8)、卡波西氏肉瘤相关疱疹病毒(KSHV)、B病毒痘病毒痘苗病毒、天花病毒

(variola virus)、天花病毒 (smallpox virus)、猴痘病毒、牛痘病毒、骆驼痘病毒、鼠痘病毒、小鼠痘病毒、兔痘病毒、浣熊痘病毒、传染性软疣病毒、口疮病毒、副痘苗病毒、牛丘疹性口炎病毒、绵羊痘病毒、山羊痘病毒、结节性皮肤病病毒、禽痘病毒、金丝雀痘病毒、鸽痘病毒、麻雀痘病毒、粘液瘤病毒、野兔纤维瘤病毒、兔纤维瘤病毒、松鼠纤维瘤病毒、猪痘病毒、塔纳痘病毒、亚巴猴痘病毒、黄病毒登革病毒、丙型肝炎病毒 (HCV)、GB 肝炎病毒 (GBV-A、GBV-B 和 GBV-C)、西尼罗河病毒、黄热病病毒、圣路易斯脑炎病毒、日本脑炎病毒、波瓦生病毒、蜱传脑炎病毒、夸赛纳森林病病毒、披膜病毒、委内瑞拉马脑炎 (VEE) 病毒、基孔肯雅病毒、罗斯河病毒、马亚罗病毒、辛德毕斯病毒、风疹病毒、逆转录病毒人类免疫缺陷病毒 (HIV) 1 型和 2 型、人类 T 细胞白血病毒 (HTLV) 1 型、2 型和 5 型、小鼠乳腺肿瘤病毒 (MMTV)、劳氏肉瘤病毒 (RSV)、慢病毒、冠状病毒、严重急性呼吸道综合征 (SARS) 病毒、纤丝病毒埃博拉病毒、马尔堡病毒、偏肺病毒 (MPV) 如人类偏肺病毒 (HMPV)、弹状病毒狂犬病病毒、水疱性口膜炎病毒、布尼亚病毒、克里米亚-刚果出血热病毒、裂谷热病毒、拉克罗斯病毒、汉坦病毒、正粘病毒、流感病毒 (A、B 和 C 型)、副粘病毒、副流感病毒 (PIV 1、2 和 3 型)、呼吸道合胞病毒 (A 型和 B 型)、麻疹病毒、腮腺炎病毒、砂砾样病毒、淋巴细胞性脉络丛脑膜炎病毒、胡宁病毒、马丘坡病毒、瓜纳瑞托病毒、拉沙病毒、Ampari 病毒、弗莱克索尔病毒、伊派病毒、Mobala 病毒、Mopeia 病毒、拉丁美洲病毒、巴拉那病毒、皮秦德病毒、庞塔托鲁病毒 (PTV)、塔卡里伯病毒和太米阿米病毒。

[0221] 可以通过本公开内容的方法检测的细菌病原体的具体实例包括但不限于以下的任何一种或多种 (或任何组合): 鲍氏不动杆菌 (*Acinetobacter baumannii*)、放线杆菌属 (*Actinobacillus* sp.)、放线菌 (*Actinomycetes*)、放线菌属 (*Actinomyces* sp.) (如衣氏放线菌 (*Actinomyces israelii*) 和内氏放线菌 (*Actinomyces naeslundii*))、气单胞菌属 (*Aeromonas* sp.) (如嗜水气单胞菌 (*Aeromonas hydrophila*)、维氏气单胞菌温和生物变种 (*Aeromonas veronii* biovar *sobria*) (温和气单胞菌 (*Aeromonas sobria*)) 和豚鼠气单胞菌 (*Aeromonas caviae*))、嗜吞噬细胞无形体 (*Anaplasma phagocytophilum*)、木糖氧化产碱菌 (*Alcaligenes xylosoxidans*)、鲍氏不动杆菌、伴放线放线杆菌 (*Actinobacillus actinomycetemcomitans*)、芽孢杆菌属 (*Bacillus* sp.) (如炭疽芽孢杆菌 (*Bacillus anthracis*)、蜡状芽孢杆菌 (*Bacillus cereus*)、枯草芽孢杆菌 (*Bacillus subtilis*)、苏云金芽孢杆菌 (*Bacillus thuringiensis*) 和嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*))、拟杆菌属 (*Bacteroides* sp.) (如脆弱拟杆菌 (*Bacteroides fragilis*))、巴尔通氏体属 (*Bartonella* sp.) (如杆状巴尔通氏体 (*Bartonella bacilliformis*) 和汉氏巴尔通氏体 (*Bartonella henselae*))、双歧杆菌属 (*Bifidobacterium* sp.)、博德特氏菌属 (*Bordetella* sp.) (如百日咳博德特氏菌 (*Bordetella pertussis*)、副百日咳博德特氏菌 (*Bordetella parapertussis*) 和支气管炎博德特氏菌 (*Bordetella bronchiseptica*))、疏螺旋体属 (*Borrelia* sp.) (如回归热疏螺旋体 (*Borrelia recurrentis*) 和布氏疏螺旋体 (*Borrelia burgdorferi*))、布鲁氏菌属 (*Brucella* sp.) (如流产布鲁氏菌 (*Brucella abortus*)、狗布鲁氏菌 (*Brucella canis*)、马尔他布鲁氏菌 (*Brucella melitensis*) 和猪布鲁氏菌 (*Brucella suis*))、伯克霍尔德氏菌属 (*Burkholderia* sp.) (如类鼻疽伯克霍尔德氏菌 (*Burkholderia pseudomallei*) 和洋葱伯克霍尔德氏菌 (*Burkholderia cepacia*))、弯曲杆菌属 (*Campylobacter* sp.) (如空肠弯

曲杆菌空肠亚种 (*Campylobacter jejuni*)、大肠弯曲杆菌 (*Campylobacter coli*)、红嘴鸥弯曲杆菌 (*Campylobacter lari*) 和胚胎弯曲杆菌 (*Campylobacter fetus*)、二氧化碳嗜纤维菌属 (*Capnocytophaga* sp.)、人心杆菌 (*Cardiobacterium hominis*)、砂眼衣原体 (*Chlamydia trachomatis*)、肺炎嗜衣原体 (*Chlamydophila pneumoniae*)、鹦鹉热嗜衣原体 (*Chlamydophila psittaci*)、柠檬酸杆菌属 (*Citrobacter* sp.)、伯氏考克斯氏体 (*Coxiella burnetii*)、棒杆菌属 (*Corynebacterium* sp.) (如白喉棒杆菌 (*Corynebacterium diphtheriae*)、杰氏棒杆菌 (*Corynebacterium jeikeum*) 和棒杆菌 (*Corynebacterium*))、梭菌属 (*Clostridium* sp.) (如产气荚膜梭菌 (*Clostridium perfringens*)、艰难梭菌 (*Clostridium difficile*)、肉毒梭菌 (*Clostridium botulinum*) 和破伤风梭菌 (*Clostridium tetani*))、嗜蚀艾肯氏菌 (*Eikenella corrodens*)、肠杆菌属 (*Enterobacter* sp.) (如产气肠杆菌 (*Enterobacter aerogenes*)、成团肠杆菌 (*Enterobacter agglomerans*)、阴沟肠杆菌 (*Enterobacter cloacae*) 和大肠杆菌 (*Escherichia coli*), 包括机会性大肠杆菌, 如肠毒性大肠杆菌, 肠侵袭性大肠杆菌, 肠致病性大肠杆菌, 肠出血性大肠杆菌, 肠聚集性大肠杆菌和尿路致病大肠杆菌)、肠球菌属 (*Enterococcus* sp.) (如粪肠球菌 (*Enterococcus faecalis*) 和屎肠球菌 (*Enterococcus faecium*))、埃里希氏体属 (*Ehrlichia* sp.) (如恰菲埃里希氏体 (*Ehrlichia chafeensis*) 和犬埃里希氏体 (*Ehrlichia canis*))、猪红斑丹毒丝菌 (*Erysipelothrix rhusiopathiae*)、真杆菌属 (*Eubacterium* sp.)、土拉热弗朗西丝氏菌 (*Francisella tularensis*)、具核梭杆菌 (*Fusobacterium nucleatum*)、阴道加德纳氏菌 (*Gardnerella vaginalis*)、麻疹孪生球菌 (*Gemella morbillorum*)、嗜血菌属 (*Haemophilus* sp.) (如流感嗜血菌 (*Haemophilus influenzae*)、杜氏嗜血菌 (*Haemophilus ducreyi*)、流感嗜血菌埃及生物群 (*Haemophilus aegyptius*)、副流感嗜血菌 (*Haemophilus parainfluenzae*)、溶血嗜血菌 (*Haemophilus haemolyticus*) 和副溶血嗜血菌 (*Haemophilus parahaemolyticus*))、螺杆菌属 (*Helicobacter* sp.) (如幽门螺杆菌 (*Helicobacter pylori*)、同性恋螺杆菌 (*Helicobacter cinaedi*) 和芬纳尔螺杆菌 (*Helicobacter fennelliae*))、金氏金氏杆菌 (*Kingella kingii*)、克雷伯氏菌属 (*Klebsiella* sp.) (如肺炎克雷伯氏菌 (*Klebsiella pneumoniae*)、肉芽肿克雷伯氏菌 (*Klebsiella granulomatis*) 和产酸克雷伯氏菌 (*Klebsiella oxytoca*))、乳杆菌属 (*Lactobacillus* sp.)、单核细胞增生利斯特氏菌、问号钩端螺旋体 (*Leptospira interrogans*)、侵肺军团菌 (*Legionella pneumophila*)、问号钩端螺旋体、消化链球菌属 (*Peptostreptococcus* sp.)、粘膜莫拉氏菌 (*Moraxella catarrhalis*)、摩根氏菌属 (*Morganella* sp.)、动弯杆菌属 (*Mobiluncus* sp.)、微球菌属 (*Micrococcus* sp.)、分枝杆菌属 (*Mycobacterium* sp.) (如麻风分枝杆菌 (*Mycobacterium leprae*)、结核分枝杆菌 (*Mycobacterium tuberculosis*)、胞内分枝杆菌 (*Mycobacterium intracellulare*)、鸟分枝杆菌 (*Mycobacterium avium*)、牛分支杆菌 (*Mycobacterium bovis*) 和瘰癧分枝杆菌 (*Mycobacterium marinum*))、枝原体属 (*Mycoplasma* sp.) (如肺炎枝原体 (*Mycoplasma pneumoniae*)、人型枝原体 (*Mycoplasma hominis*) 和生殖道枝原体 (*Mycoplasma genitalium*))、诺卡氏菌属 (*Nocardia* sp.) (如星状诺卡氏菌 (*Nocardia asteroides*)、盖尔森基兴诺卡氏菌 (*Nocardia cyriacigeorgica*) 和巴西诺卡氏菌 (*Nocardia brasiliensis*))、奈瑟氏球菌属 (*Neisseria* sp.) (如淋病奈瑟氏球菌

(*Neisseria gonorrhoeae*) 和脑膜炎奈瑟氏球菌 (*Neisseria meningitidis*))、多杀巴斯德氏菌 (*Pasteurella multocida*)、类志贺邻单胞菌 (*Plesiomonas shigelloides*)、普雷沃氏菌属 (*Prevotella* sp.)、卟啉单胞菌属 (*Porphyromonas* sp.)、产黑素普雷沃氏菌 (*Prevotella melaminogenica*)、变形菌属 (*Proteus* sp.) (如普通变形菌 (*Proteus vulgaris*) 和奇异变形菌 (*Proteus mirabilis*))、普罗维登斯菌属 (*Providencia* sp.) (如产碱普罗维登斯菌 (*Providencia alcalifaciens*)、雷氏普罗维登斯菌 (*Providencia rettgeri*) 和斯氏普罗维登斯菌 (*Providencia stuartii*))、铜绿假单胞菌 (*Pseudomonas aeruginosa*)、痤疮丙酸杆菌 (*Propionibacterium acnes*)、马红球菌 (*Rhodococcus equi*)、立克次氏体属 (*Rickettsia* sp.) (如立氏立克次氏体 (*Rickettsia rickettsii*)、蜱立克次氏体 (*Rickettsia akari*) 和普氏立克次氏体 (*Rickettsia prowazekii*))、恙虫病东方体 (*Orientia tsutsugamushi*) (以前称:恙虫热立克次氏体 (*Rickettsia tsutsugamushi*)) 和斑疹伤寒立克次氏体 (*Rickettsia typhi*))、红球菌属 (*Rhodococcus* sp.)、粘质沙雷氏菌 (*Serratia marcescens*)、嗜麦芽糖寡养单胞菌 (*Stenotrophomonas maltophilia*)、沙门氏菌属 (*Salmonella* sp.) (如肠沙门氏菌 (*Salmonella enterica*)、伤寒沙门氏菌 (*Salmonella typhi*)、副伤寒沙门氏菌 (*Salmonella paratyphi*)、肠炎沙门氏菌 (*Salmonella enteritidis*)、猪霍乱沙门氏菌 (*Salmonella choleraesuis*) 和鼠伤寒沙门氏菌 (*Salmonella typhimurium*))、沙雷氏菌属 (*Serratia* sp.) (如粘质沙雷氏菌和液化沙雷氏菌 (*Serratia liquifaciens*))、志贺氏菌属 (*Shigella* sp.) (如痢疾志贺氏菌 (*Shigella dysenteriae*)、弗氏志贺氏菌 (*Shigella flexneri*)、鲍氏志贺氏菌 (*Shigella boydii*) 和索氏志贺氏菌 (*Shigella sonnei*))、葡萄球菌属 (*Staphylococcus* sp.) (如金黄色葡萄球菌 (*Staphylococcus aureus*)、表皮葡萄球菌 (*Staphylococcus epidermidis*)、溶血葡萄球菌 (*Staphylococcus hemolyticus*)、腐生葡萄球菌 (*Staphylococcus saprophyticus*))、链球菌属 (*Streptococcus* sp.) (如肺炎链球菌 (*Streptococcus pneumoniae*) (如氯霉素耐药血清型4的肺炎链球菌, 奇霉素耐药血清型6B的肺炎链球菌, 链霉素耐药血清型9V的肺炎链球菌, 红霉素耐药血清型14的肺炎链球菌, 奥普托欣耐药血清型14的肺炎链球菌, 利福平耐药血清型18C的肺炎链球菌, 四环素耐药血清型19F的肺炎链球菌, 青霉素耐药血清型19F的肺炎链球菌, 以及甲氧苄氨嘧啶耐药血清型23F的肺炎链球菌, 氯霉素耐药血清型4的肺炎链球菌, 奇霉素耐药血清型6B的肺炎链球菌, 链霉素耐药血清型9V的肺炎链球菌, 奥普托欣耐药血清型14肺炎链球菌, 利福平耐药血清型18C肺炎链球菌, 青霉素耐药血清型19F的肺炎链球菌, 或甲氧苄氨嘧啶耐药血清型23F的肺炎链球菌)、无乳链球菌 (*Streptococcus agalactiae*)、变异链球菌 (*Streptococcus mutans*)、酿脓链球菌 (*Streptococcus pyogenes*), A群链球菌, 酿脓链球菌, B群链球菌, 无乳链球菌, C群链球菌, 咽峡炎链球菌 (*Streptococcus anginosus*), 似马链球菌 (*Streptococcus equisimilis*), D群链球菌, 牛链球菌 (*Streptococcus bovis*), F群链球菌, 以及咽峡炎链球菌G群链球菌、减少螺菌 (*Spirillum minus*)、念珠状链杆菌 (*Streptobacillus moniliformis*)、密螺旋体属 (*Treponema* sp.) (如品他病密螺旋体 (*Treponema carateum*)、极细密螺旋体 (*Treponema petenue*)、苍白密螺旋体 (*Treponema pallidum*) 和 *Treponema endemicum*)、*Tropheryma whippelii*、解脲支原体 (*Ureaplasma urealyticum*)、韦荣氏球菌属 (*Veillonella* sp.)、*Vibrio* sp. (弧菌属) (如霍乱弧菌 (*Vibrio cholerae*)、副溶血弧菌 (*Vibrio*

parahemolyticus)、创伤弧菌 (*Vibrio vulnificus*)、副溶血弧菌 (*Vibrio parahaemolyticus*)、创伤弧菌、溶藻弧菌 (*Vibrio alginolyticus*)、模拟弧菌 (*Vibrio mimicus*)、霍利斯弧菌 (*Vibrio hollisae*)、河流弧菌 (*Vibrio fluvialis*)、麦奇尼科夫氏弧菌 (*Vibrio metchnikovii*)、海鱼弧菌 (*Vibrio damsela*) 和弗尼斯弧菌 (*Vibrio furnisii*)、耶尔森氏菌属 (*Yersinia* sp.) (如小肠结肠炎耶尔森氏菌 (*Yersinia enterocolitica*)、鼠疫耶尔森氏菌 (*Yersinia pestis*) 和假结核耶尔森氏菌 (*Yersinia pseudotuberculosis*)) 和嗜麦芽黄单胞菌 (*Xanthomonas maltophilia*)，及其他。

[0222] 在一些实施方案中，本公开内容的方法和组合物用于监测器官移植受体。通常将会发现，来自供体细胞的多核苷酸在来自受体细胞的多核苷酸的背景下循环。如果器官被较好地接受，供体循环DNA的水平通常是稳定的，而供体DNA的快速增加(例如，作为给定样品的频率)可以用作移植排斥的早期信号。在此阶段可以给予治疗以防止移植失败。对供体器官的排斥已被证明会导致血液中供体DNA增加；参见Snyder等人，PNAS 108 (15) :6629 (2011)。本公开内容提供了对该领域中的现有技术的显著灵敏度改进。在该实施方案中，受体对照样品(例如面颊拭子等)和供体对照样品可用于比较。受体样品可用于提供该参考序列，而对应于供体基因组的序列可被鉴定为相对于该参考序列的序列变体。监测可以包括在一段时间内从受体获得样品(例如血液样品)。早期样品(例如在最初几周内的样品)可用于建立供体cfDNA部分的基线。后续样品可以与基线进行比较。在一些实施方案中，供体cfDNA部分增加约或至少约10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、100%、250%、500%、1000%或更多可以作为受体正在排斥供体组织的指示。

[0223] 在一些实施方案中，提供了检测方法。

#### 实施例

[0224] 以下实施例是为了说明本发明的各种实施方案而给出，并不意味着以任何方式限制本发明。本实施例以及本文所述的方法目前是优选实施方案的代表，是示例性的，并不旨在作为对本发明范围的限制。本领域技术人员将想到在由权利要求的范围所限定的本发明的精神内所包含的变化和其他用途。

#### 实施例1：制备用于突变检测的串联重复测序文库

[0225] 从在12μL水或10mM Tris-HCl (pH 8.0) 中>10ng的约150bp DNA片段开始，添加2μL 10X CircLigase缓冲混合物，并将混合物加热至95℃持续2分钟，并在冰上冷却5分钟。向该混合物添加4μL 5M甜菜碱，1μL 50mM MnCl<sub>2</sub>和1μL CircLigase II。将反应在60℃下温育至少12小时。然后添加2μL的RCA引物混合物(各50nM，至最终浓度5nM)，并混合。将混合物加热至95℃持续2分钟，并冷却至42℃持续2小时。用Zymo寡核苷酸纯化试剂盒纯化环化连接产物。根据制造商的说明，将28μL水添加至22μL环化连接产物中，总体积为50μL。将其与100μL的寡核苷酸结合缓冲液和400μL乙醇混合。然后在>10,000×g下旋转30秒，弃去流过物。添加750μL DNA洗涤缓冲液，然后在>10,000×g下旋转30秒，弃去流过物，并以最高速度再旋转1分钟。将柱移至新的Eppendorf管中并用17μL水洗脱(最终洗脱体积为约15μL)。

[0226] 滚环扩增以约50μL的体积进行。向15μL洗脱样品中添加5μL 10X RepliPhi缓冲液(Epicenter)，1μL 25mM dNTP，2μL 100mM DTT，1μL 100U/μL RepliPhi Phi29和26μL水。将反应混合物在30℃下温育1小时。按照制造商对剩余洗涤步骤的说明，通过添加80μL Ampure珠子来纯化RCA产物。对于洗脱，添加22.5μL洗脱缓冲液，并将珠子在65℃温育5分



钟。短暂旋转后,将管返回磁体。

[0227] 将来自RCA反应的约20 $\mu$ L洗脱产物与25 $\mu$ L 2X Phusion主混合物,2.5 $\mu$ L DMSO和0.5 $\mu$ L 10 $\mu$ M的每种B2B引物混合物混合。扩增使用以下PCR程序:95 $^{\circ}$ C 1分钟,5个循环的延伸(95 $^{\circ}$ C 15秒,55 $^{\circ}$ C 15秒,72 $^{\circ}$ C 1分钟),13–18个循环的复制(95 $^{\circ}$ C 15秒,68 $^{\circ}$ C 15秒,72 $^{\circ}$ C 1分钟),和72 $^{\circ}$ C的7分钟最终延伸。通过运行E-凝胶来检查PCR产物大小。如果范围为100–500bp,则进行0.6X Ampure珠子纯化,以富集300–500bp,并取1–2ng用小RNA文库衔接子引物进行另一轮PCR。如果产品大小范围为>1000bp,使用1.6X Ampure珠子将产物纯化,取2–3ng用于Nextera XT扩增子文库制备,以通过0.6X Ampure珠子纯化来富集400–1000bp范围内的大小。

[0228] 为了对测序数据进行生物信息学,从MiSeq运行获得FASTQ文件。使用BWA将序列在FASTQ文件中与含有靶序列(例如KRAS和EGFR)的参考基因组序列比对。使用比对结果找到每个序列(两次读取)的重复单元的区域和长度及其参考位置。使用比对结果和每个序列的重复单元的信息找到所有基因座中的变体。合并两次读取的结果。计算变量的归一化频率和噪声水平。应用来自经确认变体的变体判定中的多个附加标准,包括qscore>30和p值<0.0001。符合这些标准的经确认变体被报告为真实变体(突变)。该过程可以通过计算机语言(例如python)自动完成。

实施例2:制备用于检测序列变体的串联重复测序文库

[0229] 12 $\mu$ L体积的具有150bp平均长度的10ng DNA片段被用于串联重复测序文库构建。DNA预先用T4多核苷酸激酶(New England Biolabs)处理,以在5'端添加磷酸基团并在3'端留下羟基基团。对于由DNA酶I或酶促片段化产生或从血清或血浆中提取的DNA片段,跳过末端处理步骤。将DNA与2 $\mu$ L 10X CircLigase缓冲液(Epicenter CL9021K)混合。将混合物加热至95 $^{\circ}$ C持续2分钟并在冰上冷却5分钟,然后添加4 $\mu$ L甜菜碱,1 $\mu$ L 50mM MnCl<sub>2</sub>和1 $\mu$ L CircLigase II(Epicenter CL9021K)。连接反应在60 $^{\circ}$ C下进行至少12小时。将200nM 1 $\mu$ L每种RCA引物混合物添加至(至最终浓度10nM)连接产物并混合,加热至96 $^{\circ}$ C持续1分钟,冷却至42 $^{\circ}$ C,并在42 $^{\circ}$ C温育2小时。

[0230] 用Zymo寡核苷酸纯化试剂盒(Zymo Research,D4060)纯化具有杂交RCA引物的CircLigation产物。为此,将21 $\mu$ L产物用28 $\mu$ L水和1 $\mu$ L载体RNA(Sigma-Aldrich,R5636,用1X TE缓冲液以200ng/ $\mu$ L稀释)稀释至50 $\mu$ L。将稀释样品与100 $\mu$ L寡核苷酸结合缓冲液和400 $\mu$ L 100%乙醇混合。将混合物加载到柱,并在>10,000 $\times$ g下离心30秒。弃去流过物。通过以>10,000 $\times$ g离心30秒,弃去流过物,并以最高速度再离心1分钟,用750 $\mu$ L DNA洗涤缓冲液洗涤该柱。将柱移至新的1.5mL Eppendorf管中,并用17 $\mu$ L洗脱缓冲液(10mM Tris-Cl pH值8.0,最终洗脱体积约15 $\mu$ L)洗脱DNA。

[0231] 将5 $\mu$ L 10X RepliPhi缓冲液,2 $\mu$ L 25mM dNTP,2 $\mu$ L 100mM DTT,1 $\mu$ L 100U/ $\mu$ L RepliPhi Phi29和25 $\mu$ L水(Epicenter,RH040210)添加到来自柱的15 $\mu$ L洗脱样品中,总反应体积为50 $\mu$ L。将反应混合物在30 $^{\circ}$ C下温育2小时。通过添加80 $\mu$ L AmpureXP珠(Beckman Coulter,A63881)将RCA产物纯化。按照制造商的说明进行洗涤步骤。在65 $^{\circ}$ C温育5分钟后,在22.5 $\mu$ L洗脱缓冲液中洗脱RCA产物。将管短暂离心,然后返回磁体。

[0232] 将来自RCA反应的约20 $\mu$ L洗脱产物与25 $\mu$ L 2X Phusion主混合物(New England Biolabs M0531S),2.5 $\mu$ L水,2.5 $\mu$ L DMSO和0.5 $\mu$ L B2B引物混合物(各10 $\mu$ M)混合。用以下热

循环程序进行扩增:95℃2分钟,5个循环的延伸(95℃30秒,55℃15秒,72℃1分钟),18个循环的复制(95℃15秒,68℃15秒,72℃1分钟),和72℃7分钟的最终延伸。通过电泳检测PCR产物大小。一旦通过电泳确认长PCR产物,将PCR产物与30μL Ampure珠子(0.6X体积)混合用于纯化,以富集>500bp的PCR产物。用Qubit 2.0定量平台(Invitrogen)将纯化的产物18个循环。将约1ng纯化的DNA用于Nextera XT扩增子文库制备(Illumina FC-131-1024)。通过采用0.6X Ampure珠子的纯化来富集插入大小>500bp的文库元件。

[0233] 使用2100生物分析仪的Agilent DNA高灵敏度试剂盒(Agilent Technologies Inc.,Santa Clara,CA)来分析扩增文库的浓度和大小分布。使用具有2-250bp MiSeq测序试剂盒的Illumina MiSeq进行测序。根据MiSeq手册,在测序运行中加载了12pM变性文库。

[0234] 在该程序的变化中,在文库制备中使用Illumina衔接子代替Nextera制备。为此,使用含有B2B引物的通用部分和Illumina衔接子序列(P5和P7;5' CAAGCAGAAGACGGCATACGA3'和5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT3')的一对引物,将约1ng类似纯化的DNA用于PCR扩增。使用Phusion主混合物,进行12个循环的复制步骤(95℃30秒,55℃15秒,72℃60秒)。该扩增步骤的目的是添加用于扩增子测序的Illumina衔接子。用0.6X Ampure珠子富集长度>500bp的扩增子。使用2100生物分析仪的Agilent DNA高灵敏度试剂盒(Agilent technologies Inc.,Santa Clara,CA)分析扩增子文库的浓度和大小分布。使用具有2x250bp MiSeq测序试剂盒的Illumina MiSeq进行测序。B2B引物的通用部分也用作测序引物序列,并且如果引物未包含在Illumina试剂盒中,则添加定制测序引物。在测序运行中加载了12pM变性文库。

[0235] 在一个实例分析中的靶区域覆盖如图33所示。下表6描述了靶区域的分析结果。

[0236] 表4提供了可用于本公开内容的方法的RCA引物的实例。表5提供了可用于本公开内容的方法的B2B引物的实例。

表4

基因名称	RCA 引物 A 名称	RCA 引物 A 序列	RCA 引物 B 名称	RCA 引物 B 序列
PIK3CA	PIK3CA-E1a	GCTTTGAGCTGT TCTTTGTCATT	PIK3CA-E1b	AAAGCAATTTCTACAC GAGATCC
PIK3CA	PIK3CA-E2a	TTTAATTGTGTG GAAGATCCAATC	PIK3CA-E2b	ATTAAACAGCATGCAT TGAAC TG
EGFR	EGFR-E1a	CTTTCTCACCTTC TGGGATCC	EGFR-E1b	AAATTCCCCTCGCTAT CAAG
EGFR	EGFR-E2a	CCATCACGTAGG CTTCCTG	EGFR-E2b	ATGGCCAGCGTGGACA AC
EGFR	EGFR-E3a	GACATAGTCCAG GAGGCAGC	EGFR-E3b	TGTCCGGGAACACAAA GAC
EGFR	EGFR-E4a	AAGCGACGGTCC TCCAAG	EGFR-E4b	TGGCAGCCAGGAACGT AC
EGFR	EGFR-E5a	AGTACGTTCTG GCTGCC	EGFR-E5b	AACACCGCAGCATGTC AAG
EGFR	EGFR-E6a	ATCCACTTGATA GGCACCTTG	EGFR-E6b	AAGTGGATGGCATTGG AATC
EGFR	EGFR-E7a	TCTCGCTGGCAG GGATTC	EGFR-E7b	CCTGGAGAAAGGAGA ACGC
EGFR	EGFR-E9a	AACTTTGGGCGA CTATCTGC	EGFR-E9b	AGTTCCGTGAGTTGAT CATCG
EGFR	EGFR-E10a	TTGGAGTCTGTA GGACTTGGC	EGFR-E10b	ACTTCTACCGTGCCCT GATG
EGFR	EGFR-E11a	CTGCTGTGGGAT GAGGTACTC	EGFR-E11b	CACAGCAGGGCTTCTT CAG
EGFR	EGFR-E12a	CATGGAATGCTT GTACCACATC	EGFR-E12b	CATGGGCAACTTCTCT GTTTC
BRAF	BRAF-E1a	CAGTTTGAACAG TTGTCTGGATC	BRAF-E1b	AAACTGATGGGACCCA CTCC
PTEN	PTEN-E1a	TGTTTCTGCTAAC GATCTCTTTG	PTEN-E1b	AGGAGATATCAAGAGG ATGGATTC
PTEN	PTEN-E2a	CAGGAAATCCCA TAGCAATAATG	PTEN-E2b	TCCTGCAGAAAGACTT GAAGG
PTEN	PTEN-E3a	GCTTTGAATCCA AAAACCTTAAAA C	PTEN-E3b	GGATTCAAAGCATAAA AACCATTAC
PTEN	PTEN-E4a	TACAGTACATTC ATACCTACCTCT GC	PTEN-E4b	TATGTTGTATAACTTA AACCCGATAGAC

PTEN	PTEN-E5a	AAAGGATATTGT GCAACTGTGG	PTEN-E5b	TTGAAGACCATAACCC ACCAC
PTEN	PTEN-E6a	CCATAGAAATCT AGGGCCTCTTG	PTEN-E6b	AAGTAAGGACCAGAG ACAAAAAGG
PTEN	PTEN-E7a	CCAGATGATTCT TTAACAGGTAGC	PTEN-E7b	GGATTATAGACCAGTG GCACTG
PTEN	PTEN-E9a	GAACCTTGCTTC CCGTCGTG	PTEN-E9b	CATGTACTTTGAGTTCC CTCAGC
PTEN	PTEN-E12a	TCTGGTCCTGGT ATGAAGAATG	PTEN-E12b	CAGGACCAGAGGAAA CCTCAG
PTEN	PTEN-E13a	GCTCTATACTGC AAATGCTATCG	PTEN-E13b	CGTGCAGATAATGACA AGGAATATC
PTEN	PTEN-E14a	TTGGAGAAAAGT ATCGGTTGG	PTEN-E14b	GGTCAGTTAAATTAAA CATTTTGTGG
PTEN	PTEN-E15a	TGGTGTTACAGA AGTTGAACTGC	PTEN-E15b	GATGTTAGTGACAATG AACCTGATC
KRAS	KRAS-E1a	AAGAGTGCCTTG ACGATACAGC	KRAS-E1b	TCTTGCCTACGCCACC AG
TP53	TP53-E1a	CCTGACTCAGAC TGACATTCTCC	TP53-E1b	CAGGCCCTTCTGTCTTG AAC
TP53	TP53-E2a	ATGTTCCGAGAG CTGAATGAG	TP53-E2b	GAACATCTCGAAGCGC TCAC
TP53	TP53-E3a	TTAAAGGACCAG ACCAGCTTTC	TP53-E3b	TTATGGTATAAGTTGG TGTTCTGAAG
TP53	TP53-E4a	CTTGGGACCTCT TATCAAGTGG	TP53-E4b	AGAGGTCCCAAGACTT AGTACCTG
TP53	TP53-E5a	AAGCAAGCAGGA CAAGAAGC	TP53-E5b	GCTTGCTTACCTCGCTT AGTG
TP53	TP53-E6a	GGGACGGAACAG CTTTGAG	TP53-E6b	TTCCGTCCCAGTAGAT TACCAC
TP53	TP53-E7a	CAACTACATGTG TAACAGTTCCTG C	TP53-E7b	CATGTAGTTGTAGTGG ATGGTGG
TP53	TP53-E8a	GTGGAGTATTTG GATGACAGAAAC	TP53-E8b	ATACTCCACACGCAAA TTTCC
TP53	TP53-E9a	TGCTCAGATAGC GATGGTGAG	TP53-E9b	CTATCTGAGCAGCGCT CATG
TP53	TP53-E10a	CTGTGCAGCTGT GGGTGTA	TP53-E10b	GCAGGTCTTGGCCAGT TG
TP53	TP53-E12a	AAGTCTGTGACT TGCACGGTC	TP53-E12b	TGTCCCAGAATGCAAG AAGC
TP53	TP53-E13a	CCTGTCATCTTCT GTCCCTTC	TP53-E13b	GATGACAGGGGCCAGG AG
TP53	TP53-E14a	AAGACCCAGGTC CAGATGAAG	TP53-E14b	TGGGTCTTCAGTGAAC CATTG
TP53	TP53-E15a	CTGCTCTTGTCTT TCAGACTTCC	TP53-E15b	GAGCAGAAAGTCAGTC CCATG
TP53	TP53-E16a	CTCTGAGTCAGG AAACATTTTCAG	TP53-E16b	GCTCGACGCTAGGATC TGAC

表5

基因名称	B2B 引物 A	引物 A 序列	B2B 引物 B	引物 B 序列
PIK3CA	PIK3CA-BX1a	G TTCAGAGTTCTACAG TCCGACGATCGCTTTG AGCTGTTCTTTGTCAAT	PIK3CA-BX1b	CCTTGGCACCCGAGAATT CCAAAAGCAATTTCTACA CGAGATCC
PIK3CA	PIK3CA-BX2a	G TTCAGAGTTCTACAG TCCGACGATCTTTAATT GTGTGGAAGATCCAAT C	PIK3CA-BX2b	CCTTGGCACCCGAGAATT CCAATTAAACAGCATGCA TTGAACTG
EGFR	EGFR-BX1a	G TTCAGAGTTCTACAG TCCGACGATCCTTTCTC ACCTTCTGGGATCC	EGFR-BX1b	CCTTGGCACCCGAGAATT CCAAAATTTCCCGTCGCTA TCAAG
EGFR	EGFR-BX2a	G TTCAGAGTTCTACAG TCCGACGATCCCATCA CGTAGGCTTCCTG	EGFR-BX2b	CCTTGGCACCCGAGAATT CCAATGGCCAGCGTGGAC AAC
EGFR	EGFR-BX3a	G TTCAGAGTTCTACAG TCCGACGATCGACATA GTCCAGGAGGCAGC	EGFR-BX3b	CCTTGGCACCCGAGAATT CCATGTCCGGGAACACAA AGAC
EGFR	EGFR-BX4a	G TTCAGAGTTCTACAG TCCGACGATCAAGCGA CGGTCCTCCAAG	EGFR-BX4b	CCTTGGCACCCGAGAATT CCATGGCAGCCAGGAAC GTAC
EGFR	EGFR-BX5a	G TTCAGAGTTCTACAG TCCGACGATCAGTACG TTCCTGGCTGCC	EGFR-BX5b	CCTTGGCACCCGAGAATT CCAAACACCGCAGCATGT CAAG
EGFR	EGFR-BX6a	G TTCAGAGTTCTACAG TCCGACGATCATCCAC TTGATAGGCACCTTG	EGFR-BX6b	CCTTGGCACCCGAGAATT CCAAAGTGGATGGCATTG GAATC
EGFR	EGFR-BX7a	G TTCAGAGTTCTACAG TCCGACGATCTCTCGC TGGCAGGGATTC	EGFR-BX7b	CCTTGGCACCCGAGAATT CCACCTGGAGAAAGGAG AACGC
EGFR	EGFR-BX9a	G TTCAGAGTTCTACAG TCCGACGATCAACTTT GGGCGACTATCTGC	EGFR-BX9b	CCTTGGCACCCGAGAATT CCAAGTTCCGTGAGTTGA TCATCG
EGFR	EGFR-BX10a	G TTCAGAGTTCTACAG TCCGACGATCTTGGAG TCTGTAGGACTTGGC	EGFR-BX10b	CCTTGGCACCCGAGAATT CCAACTTCTACCGTGCCC TGATG
EGFR	EGFR-BX11a	G TTCAGAGTTCTACAG TCCGACGATCCTGCTG TGGGATGAGGTACTC	EGFR-BX11b	CCTTGGCACCCGAGAATT CCACACAGCAGGGCTTCT TCAG
EGFR	EGFR-BX12a	G TTCAGAGTTCTACAG TCCGACGATCCATGGA ATGCTTGTACCACATC	EGFR-BX12b	CCTTGGCACCCGAGAATT CCACATGGGCAACTTCTC TGTTTC
BRAF	BRAF-BX1a	G TTCAGAGTTCTACAG TCCGACGATCCAGTTT GAACAGTTGTCTGGAT C	BRAF-BX1b	CCTTGGCACCCGAGAATT CCAAAAGTATGGGACCC ACTCC
PTEN	PTEN-BX1a	G TTCAGAGTTCTACAG TCCGACGATCTGTTTCT GCTAACGATCTCTTTG	PTEN-BX1b	CCTTGGCACCCGAGAATT CCAAGGAGATATCAAGA GGATGGATTC
PTEN	PTEN-BX2a	G TTCAGAGTTCTACAG TCCGACGATCCAGGAA ATCCCATAGCAATAAT G	PTEN-BX2b	CCTTGGCACCCGAGAATT CCATCCTGCAGAAAGACT TGAAGG

PTEN	PTEN-B X3a	G TTCAGAGTTCTACAG TCCGACGATCGCTTTG AATCCAAAAACCTTAA AAC	PTEN-B X3b	CCTTGGCACCCGAGAATT CCAGGATTCAAAGCATAA AAACCATTAC
PTEN	PTEN-B X4a	G TTCAGAGTTCTACAG TCCGACGATCTACAGT ACATTCTATACCTACCT CTGC	PTEN-B X4b	CCTTGGCACCCGAGAATT CCATATGTTGTATAACTT AAACCCGATAGAC
PTEN	PTEN-B X5a	G TTCAGAGTTCTACAG TCCGACGATCAAAGGA TATTGTGCAACTGTGG	PTEN-B X5b	CCTTGGCACCCGAGAATT CCATTGAAGACCATAACC CACCAC
PTEN	PTEN-B X6a	G TTCAGAGTTCTACAG TCCGACGATCCCATAG AAATCTAGGGCCTCTT G	PTEN-B X6b	CCTTGGCACCCGAGAATT CCAAAGTAAGGACCAGA GACAAAAAGG
PTEN	PTEN-B X7a	G TTCAGAGTTCTACAG TCCGACGATCCAGAT GATTCTTTAACAGGTA GC	PTEN-B X7b	CCTTGGCACCCGAGAATT CCAGGATTATAGACCAGT GGCACTG
PTEN	PTEN-B X9a	G TTCAGAGTTCTACAG TCCGACGATCGAACTT GTCTTCCCGTCGTG	PTEN-B X9b	CCTTGGCACCCGAGAATT CCACATGTACTTTGAGTT CCCTCAGC
PTEN	PTEN-B X12a	G TTCAGAGTTCTACAG TCCGACGATCTCTGGT CCTGGTATGAAGAATG	PTEN-B X12b	CCTTGGCACCCGAGAATT CCACAGGACCAGAGGAA ACCTCAG
PTEN	PTEN-B X13a	G TTCAGAGTTCTACAG TCCGACGATCGCTCTA TACTGCAAATGCTATC G	PTEN-B X13b	CCTTGGCACCCGAGAATT CCACGTGCAGATAATGAC AAGGAATATC
PTEN	PTEN-B X14a	G TTCAGAGTTCTACAG TCCGACGATCTTGGAG AAAAGTATCGGTTGG	PTEN-B X14b	CCTTGGCACCCGAGAATT CCAGGTCAGTTAAATTAA ACATTTTGTGG
PTEN	PTEN-B X15a	G TTCAGAGTTCTACAG TCCGACGATCTGGTGT TACAGAAAGTTGAACTG C	PTEN-B X15b	CCTTGGCACCCGAGAATT CCAGATGTTAGTGACAAT GAACCTGATC
KRAS	KRAS-B X1a	G TTCAGAGTTCTACAG TCCGACGATCAAGAGT GCCTTGACGATACAGC	KRAS-B X1b	CCTTGGCACCCGAGAATT CCATCTTGCCTACGCCAC CAG
TP53	TP53-BX 1a	G TTCAGAGTTCTACAG TCCGACGATCCCTGAC TCAGACTGACATTCTC C	TP53-BX 1b	CCTTGGCACCCGAGAATT CCACAGGCCCTTCTGTCT TGAAC
TP53	TP53-BX 2a	G TTCAGAGTTCTACAG TCCGACGATCATGTTT CGAGAGCTGAATGAG	TP53-BX 2b	CCTTGGCACCCGAGAATT CCAGAACATCTCGAAGCG CTCAC
TP53	TP53-BX 3a	G TTCAGAGTTCTACAG TCCGACGATCTTAAAG GACCAGACCAGCTTTC	TP53-BX 3b	CCTTGGCACCCGAGAATT CCATTATGGTATAAGTTG GTGTTCTGAAG
TP53	TP53-BX 4a	G TTCAGAGTTCTACAG TCCGACGATCCTTGGG ACCTCTTATCAAGTGG	TP53-BX 4b	CCTTGGCACCCGAGAATT CCAAGAGGTCCCAAGACT TAGTACCTG
TP53	TP53-BX 5a	G TTCAGAGTTCTACAG TCCGACGATCAAGCAA GCAGGACAAGAAGC	TP53-BX 5b	CCTTGGCACCCGAGAATT CCAGCTTGCTTACCTCGC TTAGTG

TP53	TP53-BX 6a	G TTCAGAGTTCTACAG TCCGACGATCGGGACG GAACAGCTTTGAG	TP53-BX 6b	CCTTGGCACCCGAGAATT CCATTCCGTCCCAGTAGA TTACCAC
TP53	TP53-BX 7a	G TTCAGAGTTCTACAG TCCGACGATCCAAC TA CATGTGTAACAGTTCC TGC	TP53-BX 7b	CCTTGGCACCCGAGAATT CCACATGTAGTTGTAGTG GATGGTGG
TP53	TP53-BX 8a	G TTCAGAGTTCTACAG TCCGACGATCGTGGAG TATTTGGATGACAGAA AC	TP53-BX 8b	CCTTGGCACCCGAGAATT CCAATACTCCACACGCAA ATTTC
TP53	TP53-BX 9a	G TTCAGAGTTCTACAG TCCGACGATCTGCTCA GATAGCGATGGTGAG	TP53-BX 9b	CCTTGGCACCCGAGAATT CCACTATCTGAGCAGCGC TCATG
TP53	TP53-BX 10a	G TTCAGAGTTCTACAG TCCGACGATCCTGTGC AGCTGTGGGTTGA	TP53-BX 10b	CCTTGGCACCCGAGAATT CCAGCAGGTCTTGGCCAG TTG
TP53	TP53-BX 12a	G TTCAGAGTTCTACAG TCCGACGATCAAGTCT GTGACTTGCACGGTC	TP53-BX 12b	CCTTGGCACCCGAGAATT CCATGTCCCAGAATGCAA GAAGC
TP53	TP53-BX 13a	G TTCAGAGTTCTACAG TCCGACGATCCCTGTC ATCTTCTGTCCCTTC	TP53-BX 13b	CCTTGGCACCCGAGAATT CCAGATGACAGGGGCCA GGAG
TP53	TP53-BX 14a	G TTCAGAGTTCTACAG TCCGACGATCAAGACC CAGGTCCAGATGAAG	TP53-BX 14b	CCTTGGCACCCGAGAATT CCATGGGTCTTCAGTGAA CCATTG
TP53	TP53-BX 15a	G TTCAGAGTTCTACAG TCCGACGATCCTGCTC TTGTCTTTCAGACTTCC	TP53-BX 15b	CCTTGGCACCCGAGAATT CCAGAGCAGAAAGTCAG TCCCATG
TP53	TP53-BX 16a	G TTCAGAGTTCTACAG TCCGACGATCCTCTGA GTCAGGAAACATTTTC AG	TP53-BX 16b	CCTTGGCACCCGAGAATT CCAGCTCGACGCTAGGAT CTGAC

表6

	结果
读取	1.5M
%靶碱基, 1x	97.8%
%中靶	63.4%
%重复	18.2%
平均覆盖	74.5x
覆盖的标准偏差	0.21

#### 实施例3: 用于测序文库构建的基因组DNA片段化

[0237] 使用NEBNext dsDNA片段化酶试剂盒(New England Biolabs)按照制造商的方案处理1 $\mu$ L基因组DNA。将温育时间延长至在37℃下45分钟。通过添加5 $\mu$ L的0.5M EDTA (pH 8.0) 终止片段化反应,并根据制造商的方案,通过添加2X体积的Ampure XP珠子(Beckman Coulter, A63881)进行纯化。在具有高灵敏度DNA试剂盒(Agilent)的生物分析仪上分析片段化的DNA。片段化DNA的大小范围通常为约100bp至约200bp,峰值为约150bp。

#### 实施例4: 文库制备程序

[0238] 在本实施例中,使用KAPA文库制备试剂盒(KK8230)进行说明。

[0239] 在涉及珠子纯化的步骤中,将AMPure XP珠子(cat#A63881)平衡至室温并充分重悬浮,之后与样品混合。在涡旋混合器上与样品充分混合后,将其在室温下温育15分钟,以使DNA与珠子结合。然后将珠子放在磁力架上,直到液体澄清。然后用200 $\mu$ L 80%的乙醇洗涤珠子两次,并在室温下干燥15分钟。

[0240] 对于进行末端修复反应,将高达50 $\mu$ L (2–10ng) 的无细胞DNA与20 $\mu$ L末端修复主混合物(8 $\mu$ L水,7 $\mu$ L 10X KAPA末端修复缓冲液和5 $\mu$ L KAPA末端修复酶混合物)混合,并在20 $^{\circ}$ C温育30分钟。然后将120 $\mu$ L AMPure XP珠子添加到70 $\mu$ L末端修复反应中。然后如上所述纯化样品。

[0241] 对于进行加A尾反应,将含有末端修复的DNA片段的干燥珠子与A尾主混合物(42 $\mu$ L水,5 $\mu$ L 10X KAPA A尾缓冲液和KAPA A尾酶)混合。将反应在30 $^{\circ}$ C下温育30分钟。添加90 $\mu$ L PEG溶液(20%PEG 8000,2.5M NaCl)后,根据上述珠子纯化方案洗涤混合物。对于平端连接反应,跳过该加A尾步骤。

[0242] 对于衔接子连接,使用具有以下序列(5'至3')的两个寡核苷酸来形成衔接子多核苷酸双链体:/5Phos/CCATTTCATTACCTCTTTCTCCGCACCCGACATAGAT\*T和/5Phos/ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGG\*T。将含有末端修复(用于平端连接)或a尾(用于基于连接体的连接)的干燥珠子与45 $\mu$ L连接主混合物(30 $\mu$ L水,10 $\mu$ L 5x KAPA连接缓冲液和5 $\mu$ L KAPA T4DNA连接酶),以及5 $\mu$ L水(用于平端连接)或5 $\mu$ L等摩尔的连接体寡核苷酸混合物(用于基于连接体的连接)混合。将珠子充分重悬浮,并在20 $^{\circ}$ C下温育15分钟。添加50 $\mu$ L PEG溶液(参见上文)后,根据上述珠子纯化方案洗涤混合物。

[0243] 使用Illustra Genomiphi V2DNA扩增试剂盒进行多重置换扩增(MDA)。将含有连接的片段链的干燥珠子重悬于9 $\mu$ L含有随机六聚体的缓冲液中,并在95 $^{\circ}$ C下加热3分钟,然后在冰上快速冷却。在添加1 $\mu$ L酶混合物后,将冷却的样品在30 $^{\circ}$ C下温育90分钟。然后在65 $^{\circ}$ C加热10分钟停止反应。在添加30 $\mu$ L PEG溶液(参见上文)后,根据上述纯化方案洗涤混合物,并将其重悬于200 $\mu$ L TE中(在65 $^{\circ}$ C温育5分钟)。如果需要,可以用定量PCR、数字小液滴PCR(ddPCR)将纯化的产物定量,或用于下一代测序(NGS)。

[0244] 在MDA后,使用Covaris S220在130 $\mu$ L总体积中将较长的连接片段链(例如>2kb)超声处理至约300bp。制造商的方案表明峰值功率为140W,占空系数为10%,每脉冲循环200次,处理时间为80秒。选择约300bp的片段长度,以增加保持完整的原始细胞游离DNA片段的几率。如果需要,可以使用标准文库制备方案将衔接子放在经超声处理的DNA片段上以供测序。从Illumina测序仪(HiSeq或MiSeq)上的成对末端测序运行返回了多种读取组合物。其中接点(自接合,或在连接步骤中包含衔接子的情况下的衔接子连接)在读取内部(非衔接子序列在5'和3'侧)的那些片段被用于将感兴趣的序列条形码化。

#### 实施例5:环化与扩增

[0245] 此处提供了根据本方法的环化和扩增程序的示例说明。该程序使用以下用品:PCR仪(例如MJ research PTC-200Peltier热循环仪);Circligase II,ssDNA连接酶Epicenter cat#CL9025K;外切核酸酶(例如ExoI,NEB Biolabs cat#M0293S;ExoIII,NEB Biolabs cat#M0206S);T4多核苷酸激酶(NEB Biolab cat#M0201S);全基因组扩增试剂盒(例如GE Healthcare,Illustra,Ready-To-Go,Genomiphi,V3 DNA扩增试剂盒);GlycoBlue(例如Ambion cat#AM9515);微型离心机(例如Eppendorf 5415D);DNA纯化珠(例如Agencourt,



Ampure XP, Beckman Coulter cat#A63881); 磁力架 (例如 The MagnaRack™ Invitrogen cat#CS15000); Qubit® 2.0 荧光计 (Invitrogen, cat#Q32866); 分子探针 dsDNA HS 测定试剂盒 (Life Technology cat#032854); 和生物分析仪 (Agilent 2100) 以及高灵敏度 DNA 试剂 (cat#5067-4626)。

[0246] 对于扩增缺乏 5' 端磷酸盐的 DNA 片段 (例如无细胞 DNA), 第一步是末端修复和单链形成。将 DNA 在 96℃ 变性 30 秒 (例如在 PCR 仪上)。通过将 40μL DNA 和 5μL 10X PNK 反应缓冲液合并来制备多核苷酸激酶 (PNK) 反应, 然后在 37℃ 温育 30 分钟。将 1mM ATP 和 PNK 酶添加至反应中, 并在 37℃ 下温育 45 分钟。通过将 DNA 沉淀和重悬浮来进行缓冲液交换。将来自 PNK 反应的 50μL DNA, 5μL 乙酸钠 (0.5M, pH 5.2), 1μL GlycoBlue, 1μL 寡核苷酸 (100ng/μL) 和 150μL 100% 乙醇合并。将混合物在 -80℃ 温育 30 分钟, 并以 16k rpm 离心 5 分钟, 以使 DNA 沉淀。用 500μL 70% 乙醇洗涤 DNA 沉淀物, 在室温下空气干燥 5 分钟, 并将 DNA 悬浮于 12μL 10mM Tris-Cl (pH 8.0) 中。

[0247] 然后通过连接使重悬的 DNA 环化。将 DNA 在 96℃ 变性 30 秒, 将样品在冰上冷却 2 分钟, 并添加连接酶混合物 (2μL 10X CircLigase 缓冲液, 4μL 5M 甜菜碱, 1μL 50mM MnCl<sub>2</sub>, 1μL CircLigase II)。连接反应在 PCR 仪上于 60℃ 温育 16 小时。未连接的多核苷酸通过外切核酸酶消化降解。为此, 将 DNA 在 80℃ 变性 45 秒, 并向每个管中添加 1μL 外切核酸酶混合物 (ExoI 20U/μL:ExoIII 100U/μL=1:2)。上下移液 5 次对其进行混合, 并短暂旋转。将消化混合物在 37℃ 下温育 45 分钟。用 30μL 水将体积调至 50μL, 并通过如上所述的沉淀和再悬浮进行进一步的缓冲液交换。

[0248] 对于进行全基因组扩增 (WGA), 首先将纯化的 DNA 在 65℃ 变性 5 分钟。将来自 GE WGA 试剂盒的 10μL 变性缓冲液添加至 10μL 纯化的 DNA 中。将 DNA 在冷块或冰上冷却 2 分钟。将 20μL DNA 添加至 Ready-To-Go GenomiPhi V3 饼 (WGA) 中。WGA 反应在 30℃ 下温育 1.5 小时, 然后在 65℃ 下加热灭活 10 分钟。

[0249] 使用 Ampure XP 磁珠 (1.6X) 来纯化样品。将珠子涡旋, 并取 80μL 等分到 1.5mL 管中。然后将 30μL 水, 20μL 扩增的 DNA 和 80μL 珠子合并, 并在室温下温育 3 分钟。将管置于磁力架上 2 分钟, 移取澄清溶液。将珠子用 80% 乙醇洗涤两次。添加 200μL 10mM Tris-Cl (pH 8.0) 洗脱 DNA。将 DNA 珠子混合物在 65℃ 温育 5 分钟。将管放回磁力架 2 分钟。将 195μL DNA 转移至新管。使用 1μL 通过 Qubit 定量。最后, 使用 Covaris S220 对 130μL WGA 产物进行超声处理, 以达到约 400bp 的大小。

#### 实施例 6: 连接效率和中靶率分析

[0250] 通过定量 PCR (qPCR) 来分析在如上述实施例中环化并经历全基因组应用的 cfDNA。样品靶标的 qPCR 扩增曲线结果 (使用 KRAS 引物) 示于图 18A 和 18B 中。如图 18A 所示, 输入 cfDNA 的 1/10 的 qPCR 扩增的平均 Ct (循环阈值) 为 31.75, 而相同样品的连接产物的 1/10 的平均 Ct 为 31.927, 表明约 88% 的高连接效率。连接效率可以为约或大于约 70%、80%、90%、95% 或更高, 如约 100%。在一些实例中, 除去未环化的线性 DNA, 使得所有 DNA 都可以从环状形式扩增。每个样品运行三次, 一式两份。如图 18B 所示, 10ng WGA 产物和参考基因组 DNA (gDNA) (12878, 10ng) 的扩增曲线几乎彼此重合。WGA 样品的平均 Ct 为 26.655, 而 gDNA 样品的平均 Ct 为 26.605, 表明超过 96% 的高中靶率。给定量的扩增 DNA 中的 KRAS 的数量与未扩增的 gDNA 的该数量相当, 表明扩增过程无偏差。每个样品测试三次, 一式两份。作为对比点, 还测

试了Lou等人(PNAS, 2013, 110 (49))提供的环化方案。使用缺少上述实施例的沉淀和纯化步骤的Lou方法,仅10-30%的线性输入DNA被转化为环状DNA。如此低的回收率对下游测序和变体检测提出了挑战。

#### 实施例7:通过ddPCR分析扩增的环化DNA

[0251] 使用小液滴数字PCR (ddPCR) 评估从环化多核苷酸产生的全基因组扩增产物中的等位基因频率保持和偏差。一般而言, ddPCR是指一种数字PCR测定法, 它通过计算包封在支持PCR扩增的离散的、体积确定的、油包水的小液滴分区中的核酸分子数目来测量绝对量(Hinson等人, 2011, Anal. Chem. 83:8604-8610; Pinheiro等人, 2012, Anal. Chem. 84:1003-1011)。单个ddPCR反应可以由每孔至少20,000个分区的小液滴组成。小液滴数字PCR可以使用执行数字PCR测定的任何平台进行, 该数字PCR测定通过计算包封在支持PCR扩增的离散的、体积确定的、油包水的小液滴分区中的核酸分子数目来测量绝对量。小液滴数字PCR的典型策略可总结如下: 将样品稀释并分成数千至数百万个独立的反应室(油包水小液滴), 使得每个反应室包含一个感兴趣的核酸分子拷贝或不含该拷贝检测到的含有靶标扩增子(即感兴趣的核酸分子)的“阳性”小液滴的数目与不含靶标扩增子(感兴趣的核酸分子)的“阴性”小液滴的数目相比, 可用于确定原始样品中存在的感兴趣的核酸分子的拷贝数。小液滴数字PCR系统的实例包括Bio-Rad的QX100™小液滴数字PCR系统, 其将含有核酸模板的样品分成20,000纳升大小的小液滴; 以及RainDance的RainDrop™数字PCR系统, 其将含有核酸模板的样品分成1,000,000到10,000,000皮升大小的小液滴。ddPCR方法的其他实例在W02013181276A1中提供。

[0252] 本实施例中, 将来自黑素瘤细胞系的BRAF V600E基因组DNA (gDNA) 以特定比例(0%、0.67%、2.0%、6.67%、20%或100%)与参考基因组DNA 12878混合, 并且片段化以产生与cfDNA中发现的片段大小相似的片段(本实施例中约150bp)。根据实施例2对混合的DNA样品(10ng)进行环化和扩增。针对BRAF V600E和野生型, 使40ng扩增的DNA经历ddPCR。观察到的突变等位基因频率在图19中图示并制表。如图所示, 在有扩增的情况下观察到的突变等位基因频率(图19表中间行)反映了输入突变等位基因频率(顶行), 还示出了来自100ng未扩增基因组DNA的ddPCR结果(底行)。ddPCR输出的等位基因频率被计算为含BRAF突变的小液滴的数目比上含突变体的液体和含野生型的小液滴的总和。具有扩增的DNA表示为空心圆, 无扩增的DNA表示为小的实心圆。除了0.67%的小偏差外, 两个数据集完全重叠。这表明保留了突变等位基因频率的真实呈现, 基本无偏差。

#### 实施例8: 背景上的序列变体的检测

[0253] 如实施例2所述, 将10ng超声处理的gDNA (150bp, 多基因多重参照DNA, Horizon) 环化并扩增, 然后进行超声处理。然后使片段化的DNA经历Rubicon测序文库构建。捕获测序后, 绘制距参考热点50bp内的变体。变体检测的结果如图20所示, 其中判定变体需要在两种不同的多核苷酸中进行检测, 该不同多核苷酸通过不同的接点来区分。七个预期的参考热点(KIT D816V、EGFR G719S、EGFR T790M、EGFR L858R、KRAS G13D、KRAS G12D、NRAS Q61K) 绘制在位置0处。还确认了两个其他变体, 如图20中的空心三角形和菱形所示。

[0254] 为了比较, 如上所述对gDNA进行超声处理, 但根据惯例, 使10ng经超声处理的gDNA直接经历Rubicon测序文库构建, 无需环化且未要求确认两种不同多核苷酸的序列变体。在捕获测序后, 再次绘制来自参考热点的50bp内的变体, 结果如图21所示。七个预期的参考热

点 (KIT D816V、EGFR G719S、EGFR T790M、EGFR L858R、KRAS G13D、KRAS G12D、NRAS Q61K) 绘制在位置0处。其他位置的变体不是预期的,并且很可能是由于测序错误所致。与生成图20所采用的方法的结果相比,图21的结果表明,标准测序方法具有更高的随机错误率,当等位基因频率低(如低于5%)时,可能会掩盖真实的突变信号。

[0255] 图16-17示出了在有和没有要求在两种不同的多核苷酸中进行检测的情况下通过测序方法所检测的灵敏度和背景噪声的单独分析结果。如图所示,验证要求大大降低了背景噪声并提高了灵敏度。

实施例9:GC组成与大小分布分析

[0256] 如实施例5所述将10ng经超声处理的gDNA (150bp,多基因多重参照DNA,Horizon) 环化并扩增,对其进行测序并用变体判定的双多核苷酸验证过滤器(左)进行分析。将序列数目与CG百分比范围制表并绘图,如图22所示。如最左边的图所示,除中心峰(对应于基础基因组的总GC含量)之外,根据实施例5制备的样品的序列在很大程度上类似于理论分布。相比之下,当采用Rubicon测序文库构建试剂盒使用相同量的gDNA直接构建测序文库而不进行扩增时,测序结果与理论分布的差异非常明显(参见中间图)。这种直接Rubicon测序的中心峰值高于理论分布。据Newman等人(2014;Nature Medicine, (20):548-54)报道,当使用32ng cfDNA时,cfDNA测序GC含量分布与理论分布相似。这在最右边的图中说明。

[0257] 对已经如实施例5所述环化、扩增和测序的cfDNA进行DNA大小分布评估。如图23所示,测序结果表明的片段长度分布的峰值在约150-180bp,这类似于cfDNA的典型分布模式。

实施例10:扩增均一性评估

[0258] 将根据实施例5环化和扩增的10种产物的qPCR结果与未扩增的参考DNA(来自12878细胞系的gDNA,Coriell Institute)进行比较。每个实时qPCR反应使用10ng基因组参考DNA或扩增产物,并通过扩增产物与基因组参照的相对定量来产生比率。如图24所示,每次PCR的比率在2倍变化内,表明扩增的DNA池中这些靶标的拷贝数与未扩增的参考DNA非常相似。设计并在之前验证了来自6个基因(BRAF、cKIT、EGFR、KRAS、NRAS、PI3KCA)的10对PCR引物。

实施例11:DNA片段的扩增产量的定量

[0259] 从四例患者(患者1-4)和一例健康对照分离cfDNA。将基因组DNA(gDNA,Multi-Gene Multiplex Horizon)超声处理至约150bp片段。将DNA环化并用随机引物扩增。表7示出了扩增反应的DNA输入量,以及扩增产生的DNA量。即使对最小的样品(0.4ng)也获得了显著的扩增,并且所有样品都被扩增至少600倍。

表7

样品类型	输入 (ng)	产量 (ng)
gDNA	10	6100
gDNA	4	6880
gDNA	2	5700
gDNA	1	5760
gDNA	0.5	3280
cfDNA健康对照	0.4	6240
cfDNA患者1	4	13800

cfDNA患者2	3	9480
cfDNA患者3	3	7840
cfDNA患者4	1	3180

#### 实施例12:检测癌症患者cfDNA的低频突变

[0260] 在步骤1中,将cfDNA环化。在室温下在PCR管中制备环状连接混合物。将4ng-10ng cfDNA以12μl的体积移液至PCR管。将DNA在96℃变性30秒,然后将PCR管在冰上冷却2分钟。将连接混合物(2μl 10X CircLigase缓冲液,4μl 5M甜菜碱,1μl 50mM MnCl<sub>2</sub>,1μl CircLigase II)添加到每个管,并在PCR仪上于60℃反应16个小时。

[0261] 在步骤2中,对连接反应进行处理,以除去未连接的线性DNA。将1μl外切核酸酶混合物(NEB M0206S,M0293S;ExoI 20u/μl;ExoIII 100u/μl=1:2)添加到每个管,混合,并在PCR仪中于37℃温育30分钟。

[0262] 在步骤3中,对连接反应进行纯化,用于缓冲液交换。用Oligo Clean&Concentrator(Zymo Research)纯化连接产物。在外切核酸酶处理后将结合混合物(30μl 10mM Tris,100μl寡聚结合缓冲液,400μl 100%乙醇)添加至连接反应,混合并短暂离心。加载Zymo-spin柱,并以大于10,000xg旋转30秒。用750μl DNA洗涤缓冲液洗涤该柱,并以14,000xg离心1分钟。通过以大于10,000xg离心30秒,将DNA用15μl 10mM Tris洗脱。

[0263] 在步骤4中,通过随机引发来扩增DNA。用Ready-To-Go Genomiphi V3DNA扩增试剂盒(GE Healthcare)进行全基因组扩增(WGA)。将10μl经纯化的连接物与10μl 2x变性缓冲液混合,在95℃下温育3分钟,然后在冰上冷却至4℃。将20μl变性的DNA添加到WGA预混合物中,将样品在30℃下温育1.5小时,然后在65℃下灭活10分钟。

[0264] 在步骤5中,使用Agencourt AMPure XP Purification(1.6X)(Beckman Coulter)来清洁扩增产物。将30μl 10mM Tris和80μl Ampure珠子添加到20μl WGA反应中。将混合物在室温下温育2分钟。将试管置于磁铁架上,并温育2分钟。将上清液取出丢弃。将样品用200μl乙醇(80%)洗涤两次,空气干燥5分钟,并用200μl 10mM Tris(pH 8.0)洗脱DNA。

[0265] 在步骤6中,将WGA DNA片段化。使用Covaris S220超声仪对130μl WGA产物进行超声处理,以获得大小约400bp的片段。Covaris S220的设置如下:峰值入射功率=140W,占空系数=10%,每脉冲循环=200次,处理时间=55秒。

[0266] 在步骤7中,通过qPCR定量样品。1/10的连接输入和连接产物用于qPCR反应,重复三次,以测量连接效率。将10ng片段化的WGA产物与10ng参考gDNA(12878细胞系)用于qPCR,以测量中靶率。反应包含5μl 2x主混合物(TaqMan Fast Universal PCR主混合物(2x), Applied Biosystems;Evagreen染料,Biotium),0.5μl引物(5μM),1.2μl H<sub>2</sub>O,10μl DNA。扩增按照以下程序进行:95℃2分钟;和40次循环[95℃,10秒;60℃,20秒]。

[0267] 在步骤8中,构建测序文库。使用KAPA Hyper制备试剂盒(KK8500)或具有标准PCR文库的KAPA文库制备试剂盒(KK8200),从500-1000ng经超声处理的扩增DNA制备测序文库。根据制造商的方案制备衔接子连接物(衔接子最终浓度为1uM)。将连接产物的衔接子连接洗涤液——30μl(0.3x)20%PEG 8000/2.5M NaCl溶液添加到100μl重悬的连接产物。将珠子与连接产物充分混合,并在室温下温育15分钟。然后在磁体上捕获珠子直至液体澄清。然后使用Ampure XP珠子使130μl上清液经历大小选择。将样品转移至新板,然后添加20μl Ampure XP珠子(0.5x)。现在连接产物被捕获在珠子中,并用200μl 80%乙醇洗涤两次。然

后将连接产物重新悬浮并洗脱在20μl EB缓冲液中。在大小选择和纯化后,将20μl连接产物添加至25μl 2x KAPA HiFi Hotstart预备混合物和5μl 10μM P5+P7引物(5' CAAGCAGAAGACGGCATACGA3', 5' ACACTCTTCCCTACACGACGCTCTCCGATCT3')中,以使用以下循环程序扩增该文库:98℃,45秒;5个循环的(98℃,15秒;60℃,30秒;72℃,30秒);72℃,60秒。将扩增的文库稀释20倍,然后加载到片段分析仪或生物分析仪(高灵敏度芯片)上进行定量。通过凝胶大小选择器(来自Sage Science的Blue Pippin prep)进行进一步的大小选择。

[0268] 在步骤9中,使用来自xGEN Pan-Cancer Panel v1.5,127908597 (IDT)的探针通过探针捕获富集来富集测序文库。在步骤10中,文库在HiSeq 2500中测序,平均深度为1000x。

[0269] 在步骤11中,分析测序数据,以作出变体判定。变体判定包括要求在两种不同的多核苷酸(例如通过不同的接点鉴定)上出现序列差异方可作为变体计数的步骤。检测到几种体细胞突变,并且它们在公共数据库(COSMIC(癌症体细胞突变目录))中也有报道。在鉴定的突变中,BRAF V600M在0.05%的等位基因频率下得到鉴定,表明该系统即使在输入较低时也具有高灵敏度。各种突变的检测结果,包括其在样品中的频率,如表8所示。

表8

癌症类型	突变基因	突变	样品中的频率
黑素瘤	FGFR2	V191I	0.20%
黑素瘤	NRAS	Q61K	0.20%
黑素瘤	CTNNB1	S37F	0.10%
黑素瘤	BRAF	V600E	0.10%
黑素瘤	BRAF	V600M	0.05%
乳腺癌	PIK3CA	H1407L	53%
乳腺癌	LIFR	S679L	0.60%
乳腺癌	KRAS	G12D	0.60%
乳腺癌	ATM	R2993*	0.30%
乳腺癌	ATRX	S618F	0.10%
前列腺	AR	T346A	6.40%
前列腺	SPOP	F133L	3.30%
前列腺	BRAF	V600M	1.10%
前列腺	NSD1	R1557C	1.00%
前列腺	SF3B1	K700E	1.60%
胰腺癌	KRAS	G12V	1.70%
胰腺癌	TP53	7578176C>T (剪接供体变体)	0.5%

#### 实施例13:来自FFPE样品的多重参考DNA的准确突变检测

[0270] 按照制造商的方案(Covaris truXTRAC™ FFPE DNA试剂盒),从样品Horizon FFPE-multiplex (HD200)中提取DNA。使用Covaris S220超声仪对130μl FFPE gDNA进行超声处理,以获得大小约150bp的片段(Covaris S220的设置:峰值入射功率=175W,占空系数=10%,每脉冲循环=200次,处理时间=430秒)。将11μl体积的50ng DNA在95℃变性30秒。添加10.5μl H<sub>2</sub>O,2.5μl连接酶缓冲液(NEB B0202S)和1μl T4多核苷酸激酶(NEB M0201S)。将反应在37℃下温育30分钟,使其磷酸化。

[0271] 对样品进行连接,然后用Oligo Clean&Concentrator (Zymo Research) 纯化。在外切核酸酶处理后将结合混合物 (30 $\mu$ l 10mM Tris, 100 $\mu$ l寡聚结合缓冲液, 400 $\mu$ l 100%乙醇) 添加到连接反应中,通过涡旋混合,并短暂旋转。将样品加载到Zymo-spin柱上,并以大于10,000xg旋转30秒。用750 $\mu$ l DNA洗涤缓冲液对柱进行洗涤,并以14,000xg离心1分钟。通过以大于10,000xg离心30秒,将DNA用15 $\mu$ l 10mM Tris洗脱。

[0272] 根据实施例13中的步骤5-11进一步处理和分析样品。结果总结在表9中。Horizon的多重突变标准DNA中的九个突变的呈现大致被该过程保留,而DNA的量增加了至少600倍。

表9

基因	突变	等位基因频率	测量的等位基因频率
BRAF	p.V600E	10.5%	8%
KIT	p.D816V	10%	9.4%
EGFR	p.L858R	3%	4%
EGFR	p.T790M	1%	1%
EGFR	p.G719S	24.5%	14%
KRAS	p.G13D	15%	9.1%
KRAS	p.G12D	6%	5.4%
NRAS	p.Q61K	12.5%	8.3%
PIK3CA	p.H1047R	17.5%	14%

#### 实施例14:检测来自癌症突变细胞系gDNA多聚体的低频突变

[0273] 本实施例中,将经超声处理的基因组DNA连接形成多聚体,然后使其经历扩增、片段化和分析。图25A和25B示出了该过程的实例。图31中提供了说明性工作流程。

[0274] 将来自含有BRAF V600E突变的黑素瘤细胞系SK--mel-28 (ATCC) 的gDNA与参考gDNA (12878Coriell Institute) 混合,以获得1%的BRAF V600E。如实施例14所述对DNA进行超声处理,以获得大小约150bp的片段。将11 $\mu$ l体积的100ng DNA在95℃下变性30秒。添加10.5 $\mu$ l H<sub>2</sub>O, 2.5 $\mu$ l连接酶缓冲液 (NEB B0202S) 和1 $\mu$ l T4多核苷酸激酶 (NEB M0201S), 然后在37℃温育30分钟,使DNA磷酸化。

[0275] 使用Oligo Clean&Concentrator (Zymo Research) 纯化样品。这包括在外切核酸酶处理后将结合混合物 (25 $\mu$ l 10mM Tris, 100 $\mu$ l寡核苷酸结合缓冲液, 400 $\mu$ l 100%乙醇) 添加到连接反应中。通过涡旋对其混合并短暂旋转。加载Zymo-spin柱,并以大于10,000xg旋转30秒,用750 $\mu$ l DNA洗涤缓冲液洗涤,并以14,000xg离心1分钟。通过以大于10,000xg离心30秒,将DNA用15 $\mu$ l 10mM Tris洗脱。

[0276] 为了连接,将4 $\mu$ l体积的6ng DNA与0.45 $\mu$ l 10x末端修复缓冲液 (Enzymatics), 0.05 $\mu$ l 25mM的dNTP, 0.5 $\mu$ l 10mM的ATP, 末端修复酶混合物 (Enzymatics) 和2000单位/ $\mu$ l的T4连接酶混合。反应在25℃下温育30分钟,然后在75℃下温育20分钟。

[0277] 用Ready-To-Go Genomiphi V3DNA扩增试剂盒 (GE Healthcare) 进行全基因组扩增。将8 $\mu$ l H<sub>2</sub>O和10 $\mu$ l经纯化的连接物与10 $\mu$ l 2x变性缓冲液混合。将DNA在95℃变性3分钟,然后在冰上冷却至4℃。将20 $\mu$ l变性的DNA添加至WGA预混合物中,并在30℃下温育1.5小时,然后在65℃下灭活10分钟。

[0278] 然后使用Agencourt AMPure XP Purification (1.6X) (Beckman Coulter) 来清洁

扩增反应。将30 $\mu$ l 10mM的Tris和80 $\mu$ l Ampure珠子添加到20 $\mu$ l的WGA反应中。将其在室温下温育2分钟。将管置于磁力架上,并温育2分钟。上清液取出并丢弃。将珠子用200 $\mu$ l乙醇(80%)洗涤两次,然后空气干燥5分钟。用200 $\mu$ l 10mMTris (pH 8.0)洗脱DNA。然后使用Covaris S220超声仪将130 $\mu$ l WGA产物片段化以获得大小约400bp的片段(Covaris S220的设置:峰值入射功率=140W,占空系数=10%,每脉冲循环=200次,处理时间=55秒)。

[0279] 使用BioRad Prime PCR ddPCR突变检测测定法通过ddPCR检测突变。突变检测ddPCR反应于室温下在PCR管中组装(80ng扩增DNA,探针的10 $\mu$ l 2x ddPCR supermix,1 $\mu$ l 20x靶(BRAF V600E,BioRad)引物(9 $\mu$ M)/探针(FAM;5 $\mu$ M),1 $\mu$ l 20x野生型引物(9 $\mu$ M)/探针(HEX;5 $\mu$ M),8 $\mu$ l DNA样品(50ng))。通过上下移液将反应混合5次,然后转移到小液滴生成器的盒中。使用QX200小液滴生成器生成小液滴,将其转移到96孔PCR板中,并使用以下PCR程序进行扩增:95 $^{\circ}$ C,10分钟;40个循环的[94 $^{\circ}$ C,30秒,55 $^{\circ}$ C 1分钟];98 $^{\circ}$ C,10分钟。将PCR反应板转移到QX200小液滴读取器中,以将结果定量。根据输入的DNA,BRAF V600E突变的预期频率为1%。通过该连接和扩增程序,大致维持该频率(根据ddPCR分析,该频率为1.41%),而DNA的量增加约200倍。

[0280] 图26和27示出了图25过程的示例性变化。

实施例15:在单一反应测定中检测罕见突变

[0281] 使用NGS检测循环无细胞DNA中的罕见突变的能力受到两个因素的限制:一是测序技术的准确性;二是测定的总转化率。本文描述了使用单一反应测定工作流程和准确判定变体的算法来灵敏检测靶区域中的罕见变体的方法。用于单一反应测定的示例性工作流程如图34A和34B所示。

[0282] 将线性单链多核苷酸通过末端结合而环化。在需要单链环的情况下,多核苷酸可以是最初分离的单链多核苷酸,或者可以对其进行处理以使多核苷酸成为单链(例如通过变性)。在本实施例中,使多核苷酸环化的方法涉及酶,例如使用连接酶(例如RNA连接酶或DNA连接酶)。反应条件由所选酶的制造商指定。将多核苷酸的末端接合以形成环状多核苷酸(自身直接接合或或一种或多种其他多核苷酸结合,例如包含两种靶多核苷酸的环状靶多核苷酸)产生具有接合序列的接点。

[0283] 在环化后,包括外切核酸酶步骤以在环化反应后消化任何未连接的核酸。即,封闭环不含游离的5'或3'末端,因此5'或3'外切核酸酶的引入不会消化封闭环,但会消化未连接的组分。在一些情况下,这特别适用于多路复用系统。

[0284] 在环化后,连接酶可以保持结合在连接的分子上。连接接点处连接酶的存在可阻止引物在聚合反应中延伸并降低扩增效率(如图34A所示)。通过降解来去除环化反应中使用的连接酶允许聚合酶通过接点延伸,并有效地将环化的靶分子作为多联体扩增。连接酶的降解包括用热不稳定蛋白酶如Qiagen蛋白酶处理,其可以通过在70 $^{\circ}$ C温育15分钟而热灭活。

[0285] 在除去连接酶后,可以通过用于全基因组扩增的随机引物或用于靶向扩增的特定引物来扩增环化的DNA分子。将扩增的多联体随机超声处理,以产生大小范围为约500bp-1000bp的片段,用于NGS文库的构建和测序。

[0286] 图35中示出了本实施例中的算法的改进。多联体序列中的接点(由参照中的起始和结束位置表征)可以通过与参考序列比对来鉴定,并且可以用作原始输入DNA片段的标

记。然而,在给定的靶区域中,接点类型(参照中的开始和结束的组合)的数目可能有限;当存在大量输入DNA片段时,来自两个独立的输入DNA片段的接点出现冲突的几率相对较高。因此,在许多应用中,“接点标记”可能不被视为唯一的标记,并且基于这种标记的错误校正和分子计数可能不是唯一的区别。

[0287] 通过将多联体粉碎点与多联体序列中的接点相结合,可以为输入分子创建更多不同的标记。如图35所示,使用超声处理或其他片段化方法(例如酶促切割)将可以在环连接和RCA后产生的长串多联体粉碎以形成更短的多联体。这导致许多具有不同结构的多联体:例如,不同数量的重复或在读取中相对于接点有不同的起始/结束位置——多联体粉碎点在两端。

[0288] 通过结合多联体粉碎点来创建不同的标记,将有可能基于由不同标记鉴定的读取家族所建立的共有序列来有效地执行额外的错误校正。使用各种投票(voting)方案从读取家族构建共有序列并进行变体判定,例如,只有当家族中的所有读取都报告同一变体时,才会判定变体。其次,不同的标记可用于帮助计算输入分子的数目和计算等位基因频率。

[0289] 用于识别多联体粉碎点的示例算法如下。1) 在多联体序列(读取)中鉴定重复长度、重复区域和接点,这可以通过自我比对,与参考序列比对或其他计算方法完成;2) 通过将读取与参考序列比对来确定读取内接点的位置;3) 通过从接点的参考起始或结束位置移位来计算多联体粉碎点(在两端)的位置,同时通过相邻接点的读取位置来确定要移动的碱基数。

实施例16:在有或没有连接酶去除的情况下,对连接酶处理cfDNA分子的qPCR分析

[0290] 通过qPCR来分析在各种条件下环化和纯化的连接酶处理的cfDNA样品。测试中包括三种纯化条件:1. 通过色谱柱纯化的环化连接的DNA;2. 用苯酚氯仿纯化的环状连接的DNA;3. 用蛋白酶K处理的环状连接的DNA以除去连接酶,随后通过色谱柱进行纯化。每种条件(1、2和3)均使用10ng cfDNA。将cfDNA在96℃变性30秒,在冰块上冷却2分钟,然后添加连接混合物(2μL 10X Circligase缓冲液,4μL 5M甜菜碱,1μL 50mM MnCl<sub>2</sub>,1μL Circligase II (Epicenter#CL9025K)。同时设置‘无连接酶对照’(4)和‘无DNA对照’(5)。(‘无连接酶对照’含有10ng cfDNA,其与2μL 10μL Circligase缓冲液,4μL 5M甜菜碱,1μL 50mM MnCl<sub>2</sub>混合但无连接酶。‘无DNA对照’含有所有连接试剂但无cfDNA。所有反应在60℃温育1小时。

[0291] 在连接后,通过色谱柱纯化‘无连接酶对照’(4),‘无DNA对照’(5)和条件1(通过色谱柱纯化的环化连接DNA)。将每个反应加载到经10mM Tris-Cl (pH 8.0) 预洗涤的micro bio-spin P-6柱上,并通过在1,000x (g) 下将该柱离心4分钟来洗脱。

[0292] 条件2(通过苯酚氯仿纯化的环化连接DNA)通过苯酚氯仿提取和乙醇沉淀来纯化。将180μL 10mM的Tris添加至来自外切核酸酶处理的20μL DNA,使体积为200μL,并使用200μL苯酚提取DNA。收集水层,并乙醇沉淀来回收DNA。乙醇共沉淀混合物(苯酚提取后的200μL DNA溶液,20μL醋酸钠(0.5M, pH 5.2),1μL GlycoBlue,1μL载体寡核苷酸(100ng/μL),600μL 100%乙醇)在-80℃温育保持30分钟,并以16k rpm离心5分钟以使DNA沉淀。用500μL 70%乙醇洗涤DNA沉淀物。将DNA沉淀物在室温下空气干燥5分钟,并用11μL 10mM的Tris-Cl (pH 8.0) 重悬浮。

[0293] 条件3(用蛋白酶K处理的环状连接DNA)首先在37℃用蛋白酶K处理30分钟,然后使用色谱柱纯化。将反应加载到经10mM Tris-Cl (pH 8.0) 预洗涤的micro bio-spin P-6柱



上,并通过在1,000x (g) 下将该柱离心4分钟来洗脱。

[0294] 然后通过定量PCR (qPCR) 分析来自每种条件 (1-5) 的产物。对每个样品设置3个重复qPCR反应,并计算平均Ct。如表10所示,无DNA对照 (5) 的qPCR扩增的平均Ct (循环阈值) 为40,无连接酶对照 (4) 产物的平均Ct为30.79,表明在无连接酶处理的条件下,输入DNA的回收率较高。条件1和2的平均Ct分别为34.80和35.18。与无连接酶对照 (4) 相比,条件1和2在连接酶处理后,可扩增DNA仍有显著损失,即使两种产物都被纯化以除去游离酶。其中在纯化前通过酶降解除去了连接酶的条件3的产物显示平均Ct为30.42,这与无连接酶对照 (4) 相当,表明连接酶的除去对于可扩增DNA的有效回收至关重要。

表10

样品	平均Ct
1.通过柱纯化的连接的DNA	34.80
2.通过苯酚-氯仿提取纯化的连接的DNA	35.18
3.用蛋白酶处理并通过柱纯化的连接的DNA	30.42
4.通过柱纯化的无连接酶对照DNA	30.79
5.无DNA对照	40.00

#### 实施例17:NGS文库复杂性的比较

[0295] 本实施例比较了由不同工作流程制备的NGS文库的复杂性。本实施例中比较的工作流程包括:

- 1.通过环化和在扩增前通过苯酚氯仿提取步骤进行纯化而制备的文库;
- 2.通过环化和在未纯化的情况下直接扩增而制备的文库;
- 3.通过环化和蛋白酶处理15分钟,然后在未纯化的情况下扩增而制备的文库;
- 4.通过环化和蛋白酶处理20分钟,然后在未纯化的情况下扩增而制备的文库;
- 5.通过环化和蛋白酶处理30分钟,然后在未纯化的情况下扩增而制备的文库;
- 6.通过环化和蛋白酶处理60分钟,然后在未纯化的情况下扩增而制备的文库。

[0296] 对于每种条件,使用12μL 20ng cfDNA作为文库构建的输入。将DNA样品在96℃变性30秒,并在冰块上冷却2分钟。在冷块上建立连接混合物 (12μL cfDNA, 2μL 10X Circligase缓冲液, 4μL 5M甜菜碱, 1μL 50mM MnCl<sub>2</sub>, 1μL Circligase II (Epicentre# CL9025K)) 的添加,并且连接在60℃下进行3小时。将连接DNA混合物在PCR仪上于80℃温育45秒,然后进行外切核酸酶处理。将1μL外切核酸酶混合物 (ExoI 20U/μL:ExoIII 100U/μL =1:2) 添加至每个管,并且反应在37℃下温育30分钟。

[0297] 对于工作流程1,连接产物是用苯酚氯仿提取并用盐和乙醇沉淀。将180μL 10mM的Tris添加至来自外切核酸酶处理的20μL DNA,使体积为200μL,并使用200μL苯酚提取DNA。收集水层,并通过乙醇沉淀回收DNA。将乙醇共沉淀混合物 (苯酚提取后的200μL DNA溶液, 20μL醋酸钠 (0.5M, pH 5.2), 1μL GlycoBlue, 1μL载体寡核苷酸 (100ng/μL), 600μL 100%乙醇) 在-80℃下温育30分钟,并以16k rpm离心5分钟以使DNA沉淀。将DNA沉淀物用500μL 70%乙醇洗涤。

[0298] 将DNA沉淀物在室温下空气干燥5分钟,并用11μL 10mM的Tris-Cl (pH 8.0) 重悬浮。为了进行全基因组扩增 (WGA), 首先将纯化的DNA在65℃变性5分钟。将来自GE WGA试剂盒的10μL变性缓冲液添加到10μL纯化的DNA。将DNA在冷块或冰上冷却2分钟。将20μL DNA添

加到Ready-To-Go GenomiPhi V3饼(WGA)。将WGA反应在30℃下温育1.5小时,然后在65℃下加热灭活10分钟。

[0299] 对于工作流程2,将0.12μL 0.5M EDTA和0.58μL 1M KCl添加到经外切核酸酶处理的连接产物中,并充分混合。将连接混合物在95℃变性2分钟,并在冰上冷却至4℃,然后添加Ready-To-Go GenomiPhi V3饼(WGA)。将WGA反应在30℃下温育4.5小时,然后在65℃下加热灭活10分钟。

[0300] 对于工作流程3-6,首先将经外切核酸酶处理的连接产物用蛋白酶进行处理,以除去环化酶II。向每个反应中添加1μL丝氨酸蛋白酶,并在55℃下的温育时间被定为15分钟至60分钟(条件3:15分钟;条件4:20分钟;条件5:30分钟;条件6:60分钟),然后在70℃下加热灭活15分钟,并向经处理的连接产物添加0.12μL 0.5M EDTA和0.58μL 1M KCl。然后将连接混合物在95℃变性2分钟,并在冰上冷却至4℃,然后添加Ready-To-Go GenomiPhi V3饼(WGA)。将WGA反应在30℃下温育4.5小时,然后在65℃下加热灭活10分钟。

[0301] 对于所有条件,WGA产物均使用AmpureXP磁珠进行珠纯化,并被超声处理至平均大小为800bp。然后将经超声处理的DNA样品用作使用KAPA文库制备试剂盒的标准测序文库构建的输入。然后通过Illumina HiSeq2500对文库进行测序,并通过计算每个文库中检测到的独特分子数目来评估文库复杂性。文库复杂性的比较如下图所示,独特分子的数目按相对基础进行衡量。工作流程1显示出最低的复杂性;工作流程2的文库复杂性略高于工作流程1,表明通过跳过纯化减少了分子损失;去除连接酶(工作流程3-6)显著增加了文库中独特分子的数目,蛋白酶处理时间越长,检测到的独特分子越多。在该实验中,在温育30分钟后,检测到的分子的增加开始趋于平稳。结果如图38所示。

#### 实施例18:变体判定

[0302] 本实施例评估了使用本文所述的方法的变体判定,其中在至少两种不同的剪切多核苷酸中出现的序列变体被鉴定为真正的序列变体。

[0303] 将来自9个细胞系的基因组DNA片段化至平均大小约150bp并混合以产生DNA混合物。DNA混合物覆盖了0.1%等位基因频率(AF)的8个癌症热点(表11)。将12μL 20ng的DNA混合物用作每个反应的文库构建的输入。将DNA样品在96℃变性30秒,并在冰块上冷却2分钟。在添加连接混合物(12μL cfDNA,2μL 10X Circligase缓冲液,4μL 5M甜菜碱,1μL 50mM MnCl<sub>2</sub>,1μL Circligase II(Epicenter#CL9025K))后,反应建立在冷块上,并在60℃下连接3小时。将连接DNA混合物在PCR仪上于80℃温育45秒,然后进行外切核酸酶处理。将1μL外切核酸酶混合物(ExoI 20U/μL:ExoIII100U/μL=1:2)添加至每个管中,并且反应在37℃下温育30分钟。通过蛋白酶处理除去Circligase II,并向反应中添加0.12μL 0.5M EDTA/0.58μL 1M KCl。然后将连接混合物在95℃变性2分钟,并在冰上冷却至4℃,然后添加Ready-To-Go GenomiPhi V3饼(全基因组扩增,WGA)。将WGA反应在30℃下温育4.5小时,然后在65℃下加热灭活10分钟。

[0304] 使用AmpureXP磁珠对WGA产物进行珠纯化,并超声处理至平均大小约800bp。然后将经超声处理的DNA样品用作使用KAPA文库制备试剂盒的标准测序文库构建的输入。如图39所示,超声处理或其他片段化方法(例如,酶促切割)可用于从由RCA产生的长串多联体产生更短的多联体。由此产生的多联体可具有多种结构。所形成的较短多联体可以具有不同的重复数目,或不同个拷贝的输入DNA序列,和/或在读取中相对于接点(例如,多联体粉碎

点)有不同起始/结束位置。在一些情况下,通过随机引发产生了具有多种结构的多联体(图40)。

[0305] 通过Illumina HiSeq2500对文库进行测序,并使测序读取经历序列分析。在序列分析中,多联体粉碎点可以与测序读取的接点序列相结合,以产生可以与测序读取相关的唯一标记。由唯一标记鉴定的读取家族所建立的共有序列可用于执行额外的错误校正。

[0306] 可以使用各种投票方案从读取家族建立共有序列并执行变体判定。例如,可以基于家族中的大多数读取报告相同变体来建立共有序列。序列差异可以在读取家族共有序列与参考序列之间鉴定,并且在一些情况下,当在至少两种不同的读取家族中出现相同的序列差异时,变体得到判定。唯一标记还可用于帮助计算输入分子的数目和计算等位基因频率。

[0307] 在用于鉴定多联粉碎点并执行变体判定的示例性算法中,

1) 在多联体序列(读取)中鉴定重复长度、重复区域和接点,例如,通过自我比对、与参考序列比对或其他计算方法。

2) 通过将读取与参考序列比对来确定读取内接点的位置。

3) 通过从接点的参考起始或结束点移动来计算多联体粉碎点(在两端)的位置,同时通过相邻接点的读取位置来确定要移动的碱基数。

4) 基于读取的粉碎点结合其“接点标记”(例如,唯一标记)将其分组,并创建读取的不同读取家族,每个读取家族具有唯一标记。

5) 如果读取家族中的大多数读取一致地投票该变体,即,所有读取都报告了该变体且有多联体确认,则该家族将被视为变体家族(图41A);否则该家族将不被视为变体家族(图41B)。

6) 变体家族的数目可用于确定由“接点标记”鉴定的输入分子是否可以视为“变体分子”,其中可以应用各种截断。例如,在图41C中,当在至少两种不同的读取家族中出现相同的序列差异(开始)时,变体得到判定。在一些情况下,“变体分子”的数目可用于确定是否应该判定变体,并且可用于计算等位基因频率。

[0308] 为了缺失使用本文所述的方法对变体检测的改进,将通过当序列差异出现在来自第一剪切多核苷酸的大多数测序读取和来自第二剪切多核苷酸的大多数测序读取中时将测序读取中所检测到的序列差异判定为变体而鉴定的变体数目与通过无此要求的方法鉴定的变体进行了比较。如表12中所示,需求序列变体出现在至少两种不同的剪切多核苷酸中使非特异性假阳性变体减少约23.39%,而没有将真正的阳性判定去除(0%的特异性真正变体被去除),表明特异性的显著改善,而不影响灵敏度。

表11. 混合细胞系DNA中的癌症热点的列表

基因	变体
PIK3CA	H1047R
KRAS	G12D
EGFR	L747-E749delA750P
EGFR	T790M
EGFR	L858R
NRAS	Q61R

BRAF	V600E
EGFR	G719S

表12. 使用本文所述方法的变体检测

	过 滤 前 的 总 变 体	过 滤 后 的 总 变 体	除 去 的 非 特 异 性 变 体 的 百 分 比	过 滤 前 检 测 到 的 掺 入 癌 症 热 点 的 数 目	过 滤 后 检 测 到 的 掺 入 癌 症 热 点 的 数 目	除 去 的 特 异 性 变 体 的 百 分 比
复制 1	1260	945	25.00%	6	6	0.00%
复制 2	1194	934	21.78%	7	7	0.00%
平均	1227	939.5	23.39%	6.5	6.5	0.00%

[0309] 尽管本文中已经示出并描述了本发明的优选实施方案,但对于本领域技术人员显而易见的是,这些实施方案仅以示例的方式提供。本领域技术人员在不脱离本公开内容的前提下现将想到多种变化、改变和替代。应当理解,本文中所述的本发明实施方案的各种替代方案均可用于实施本发明。旨在由以下权利要求来限定本发明的范围,并由此涵盖这些权利要求范围内的方法和结构及其等同物。

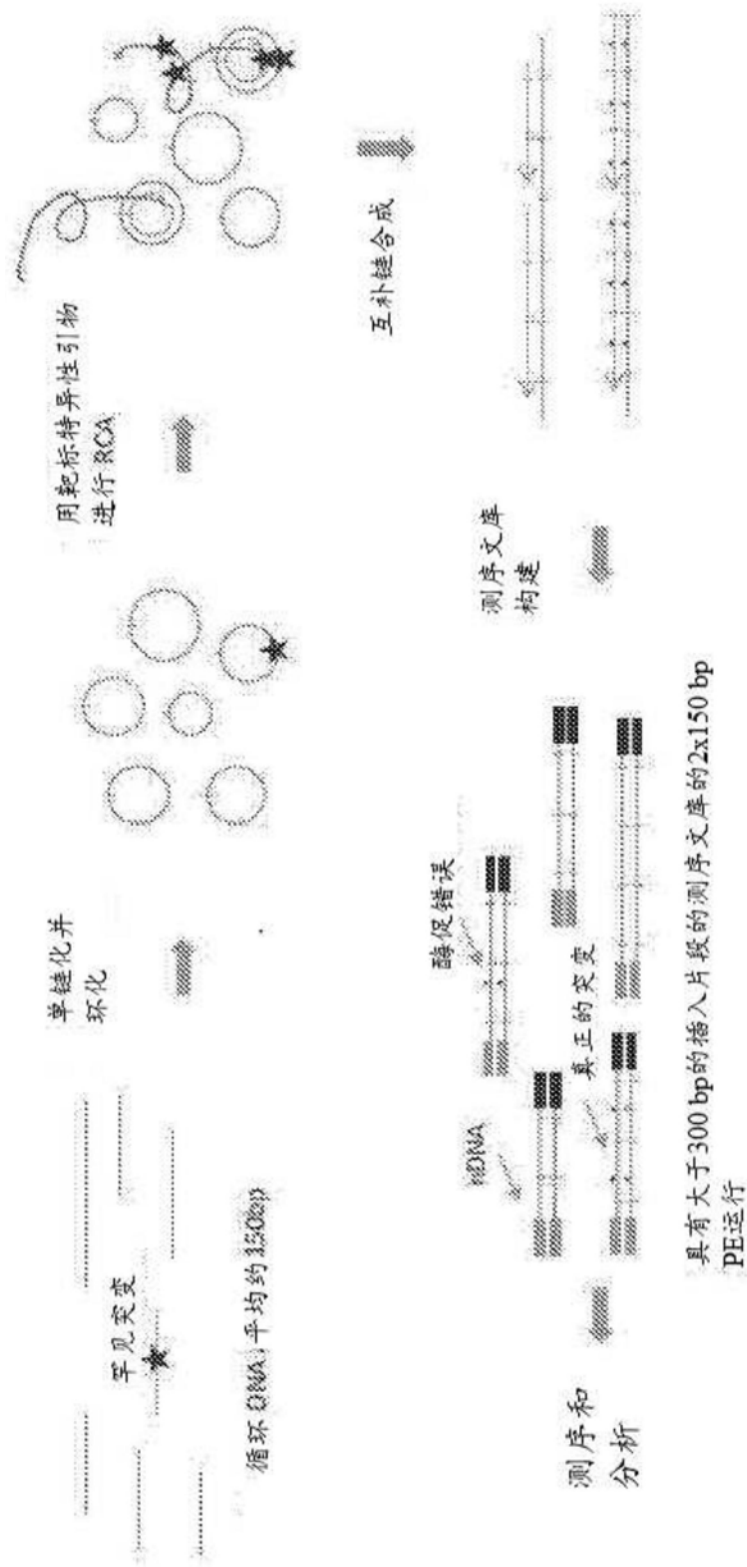


图1

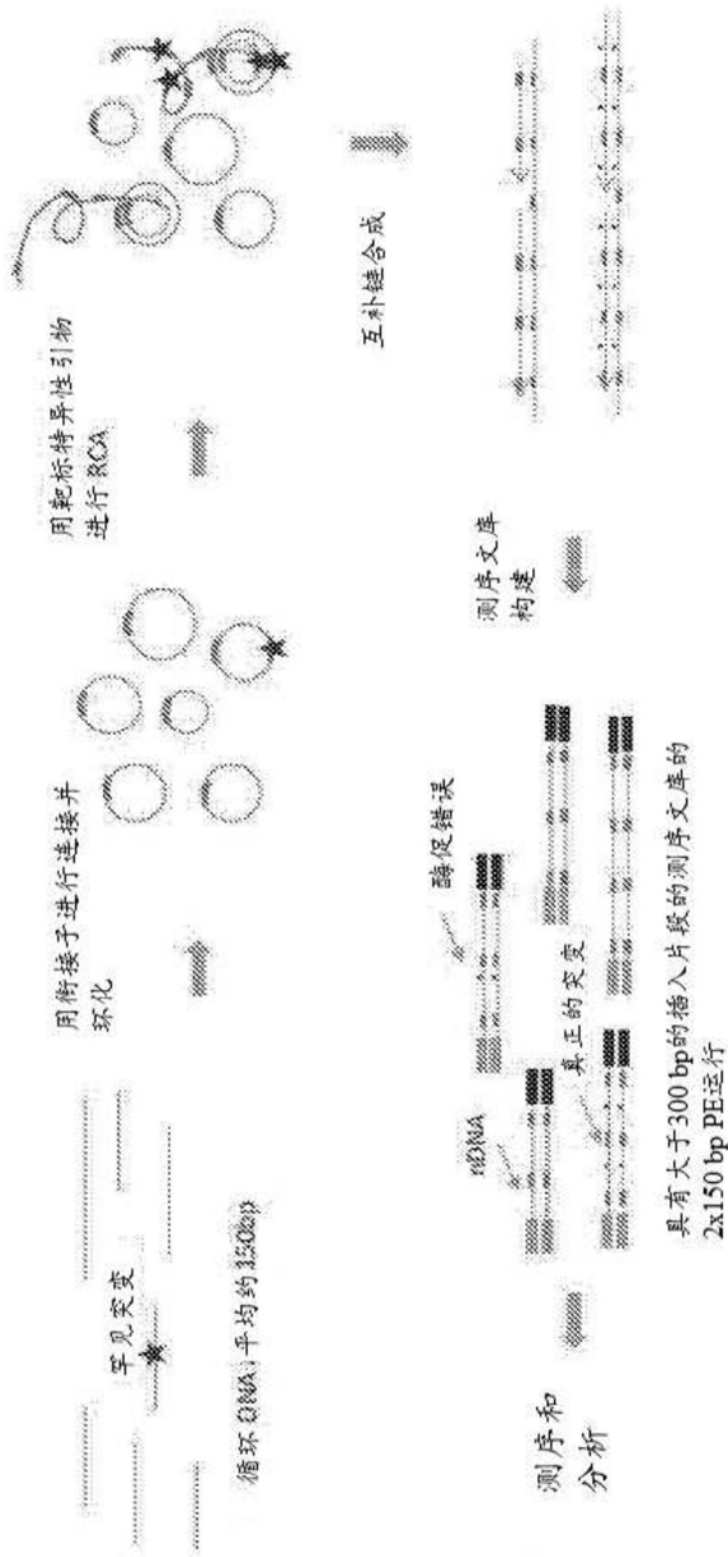


图2

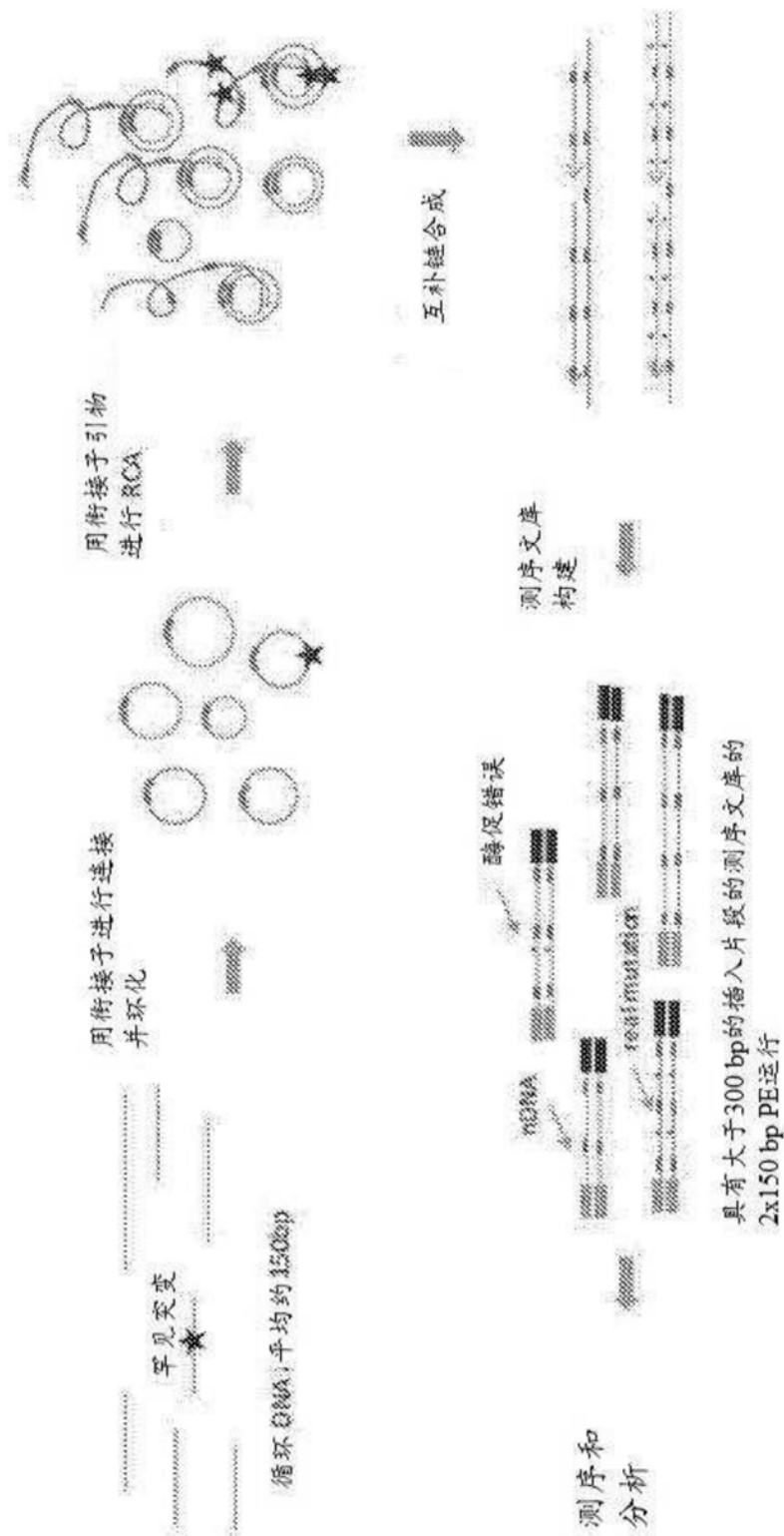


图3

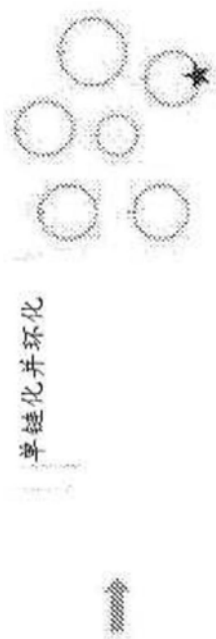


图4A



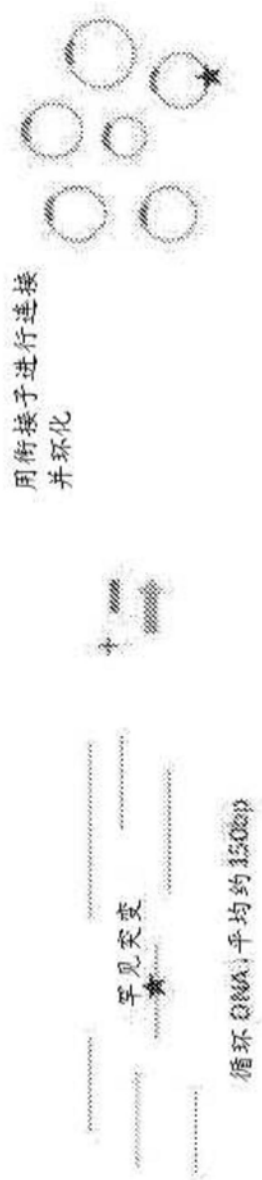


图4B



图4C

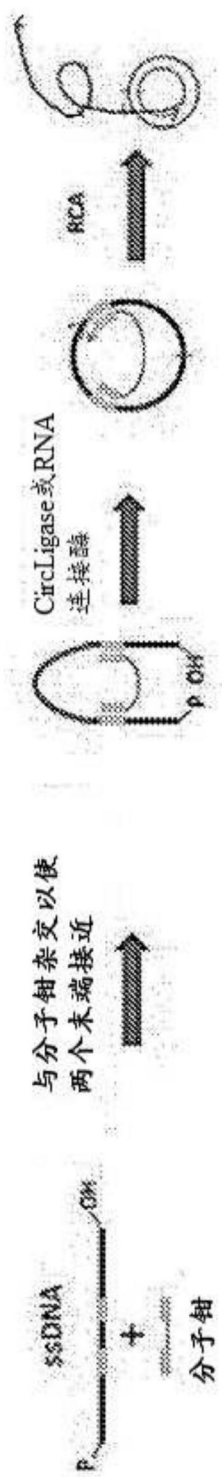


图5



图6A

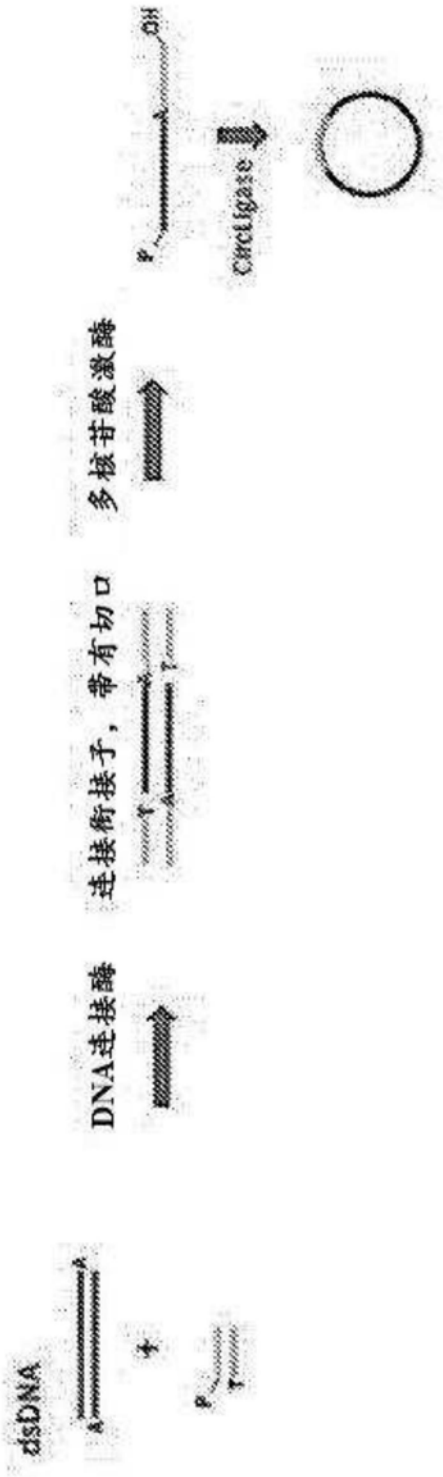


图6B

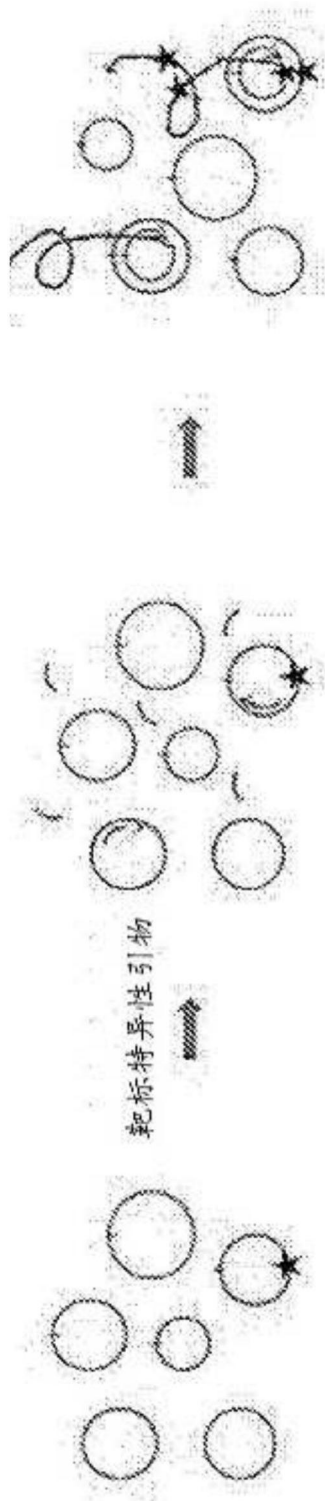


图7A

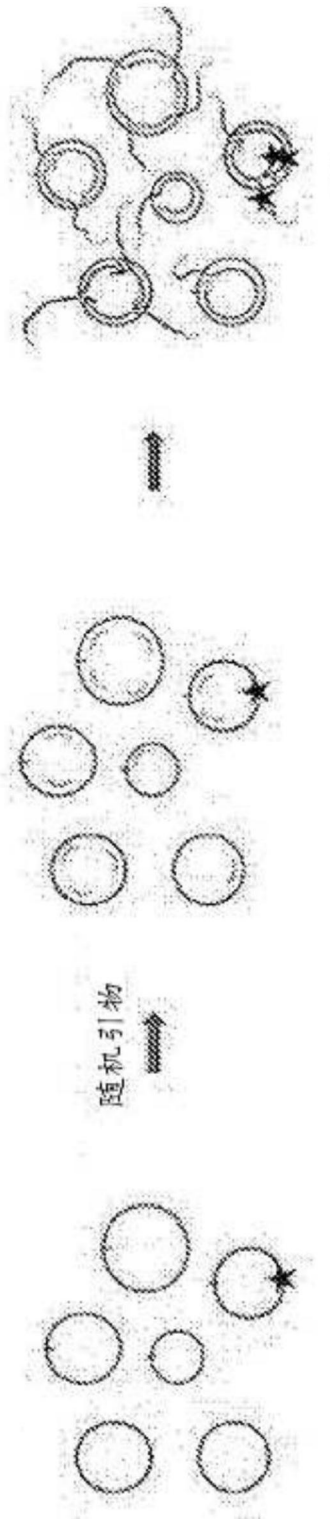


图7B

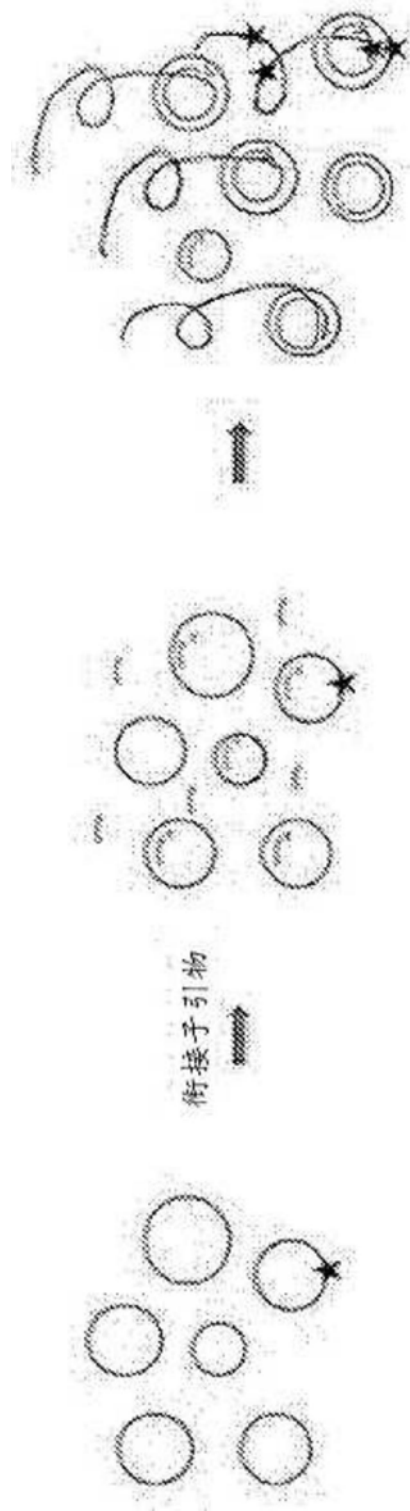


图7C



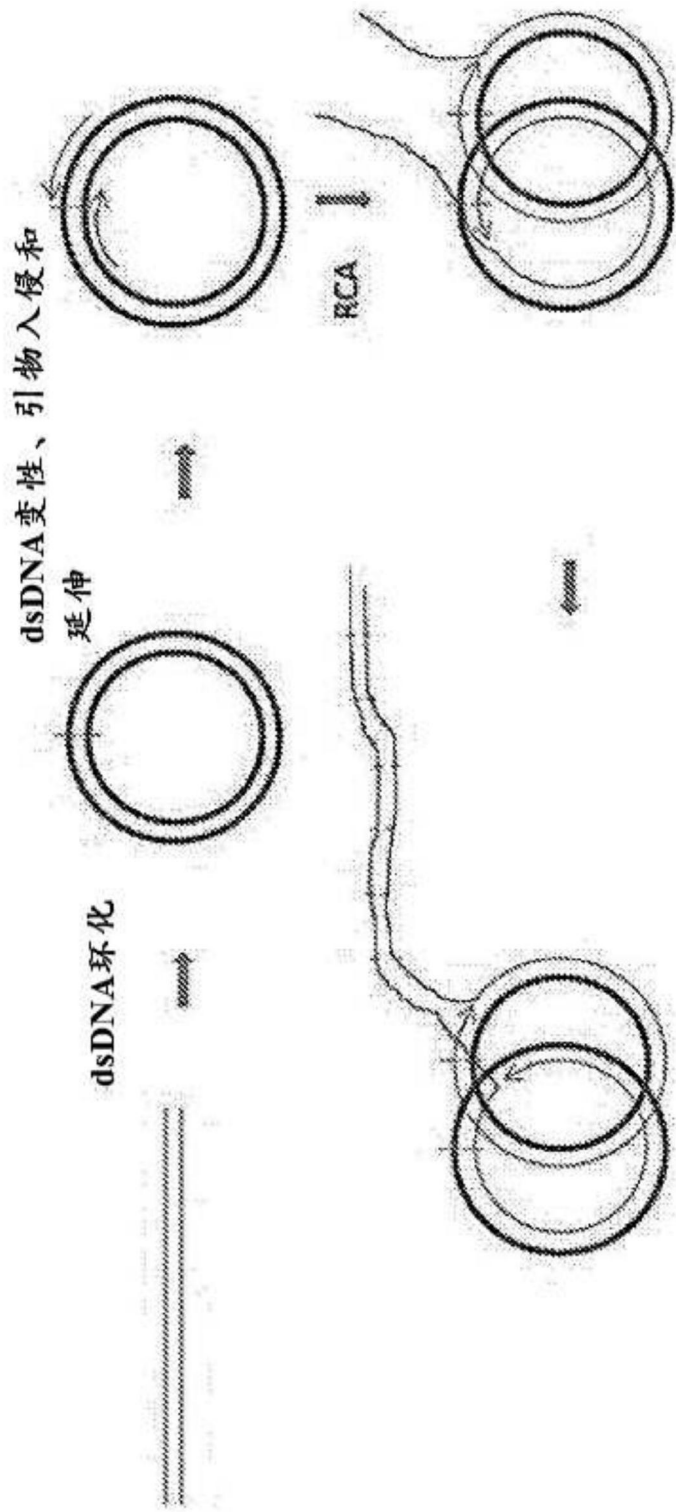


图8



图9A



图9B



图9C

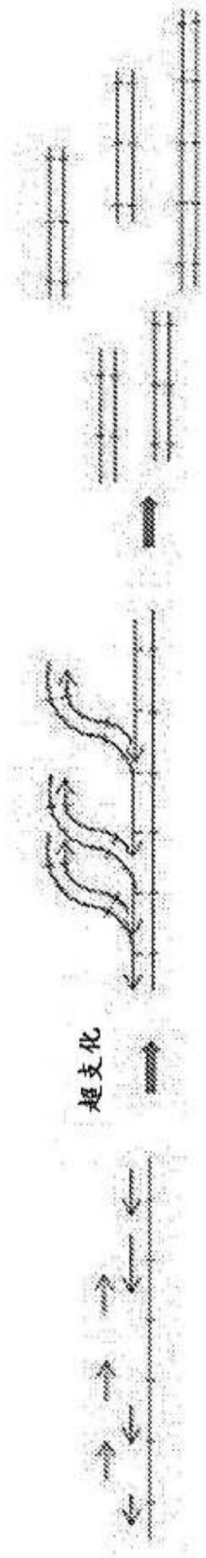


图9D

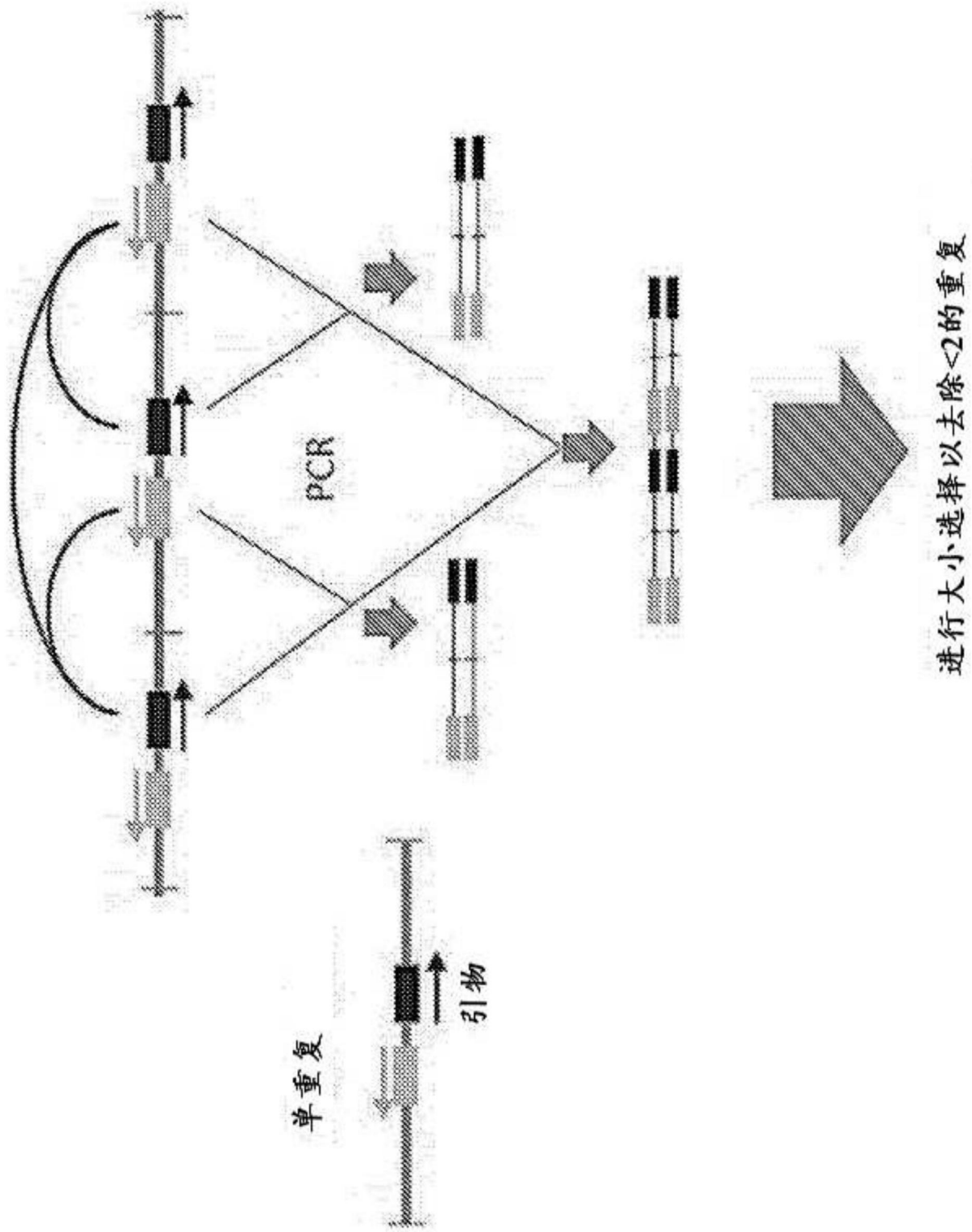


图10

图 11A

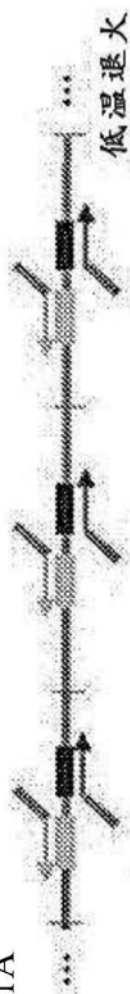


图 11B

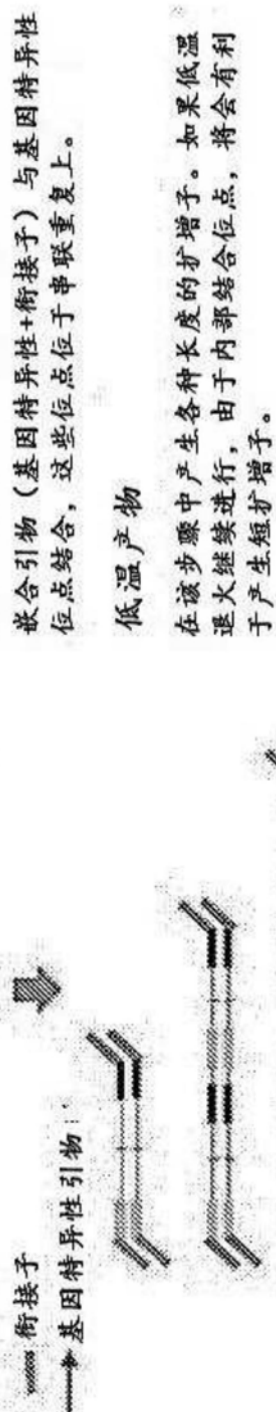


图 11C



图 11D



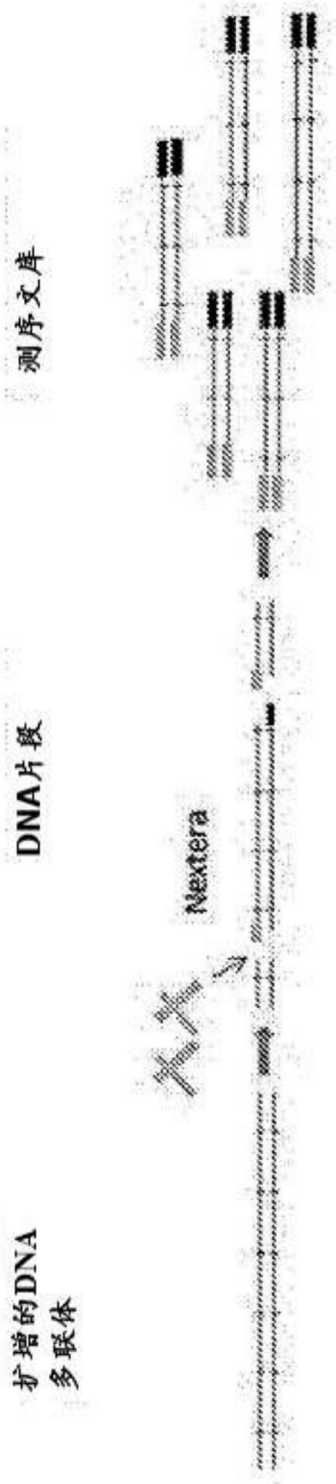


图12A





图12B

图 13A

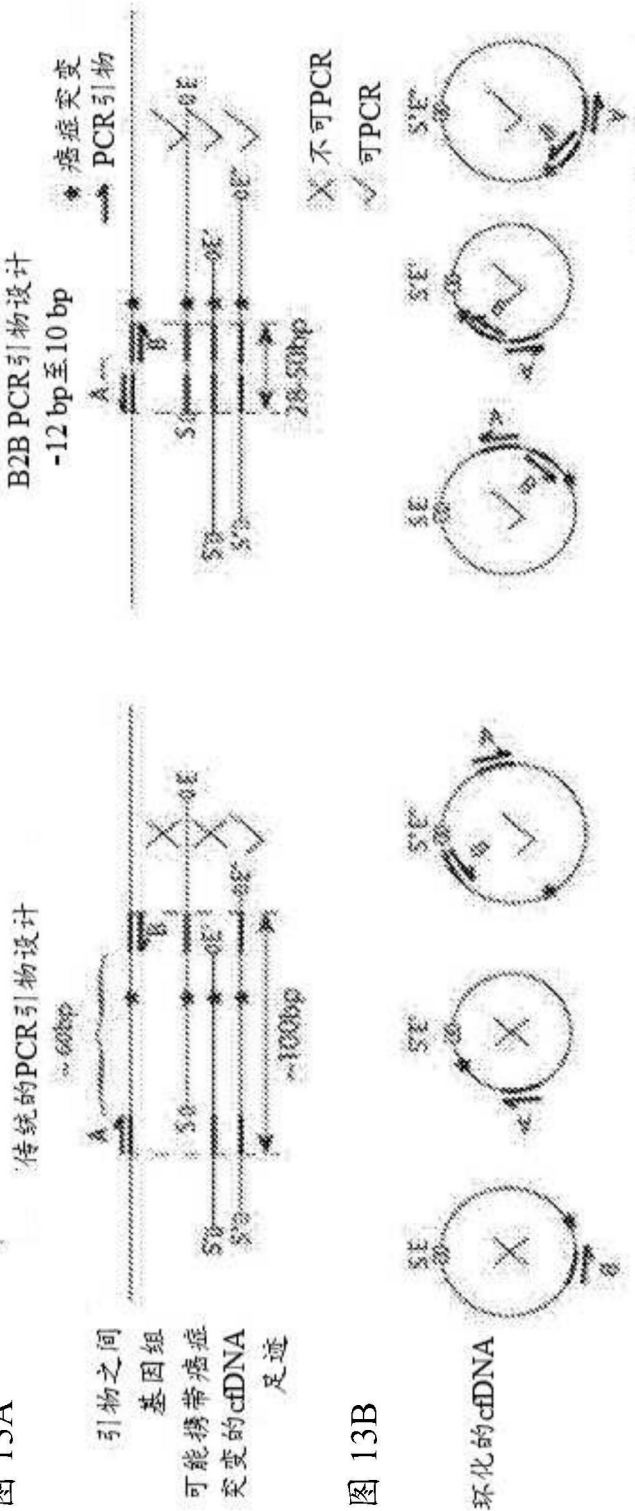


图 13B





图13C

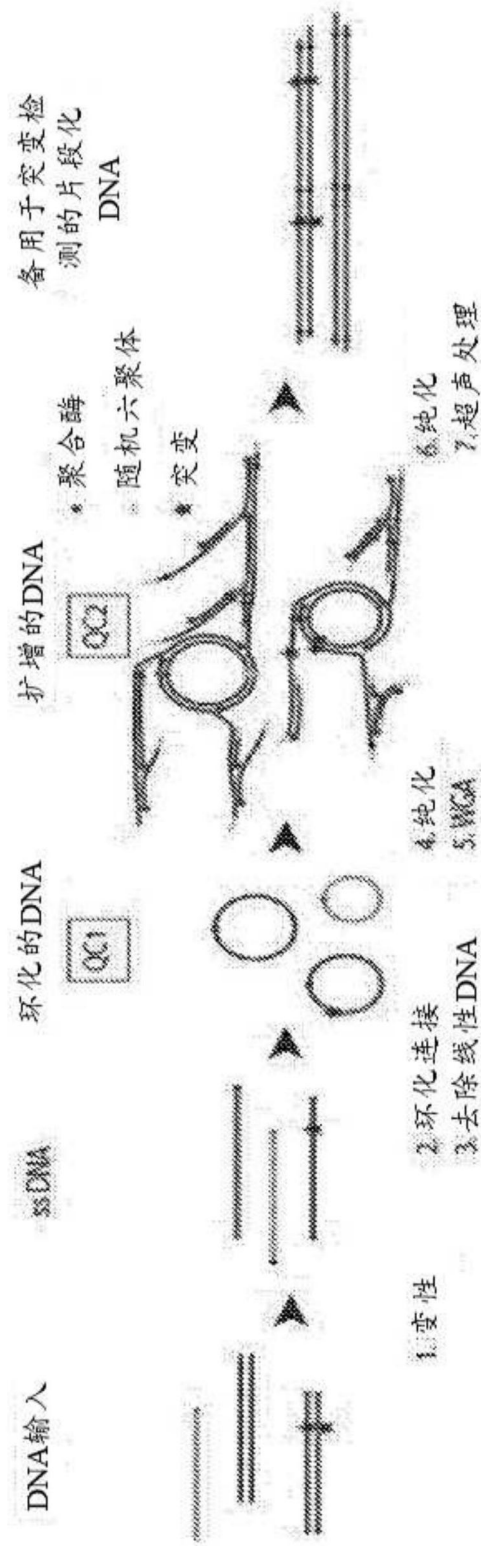
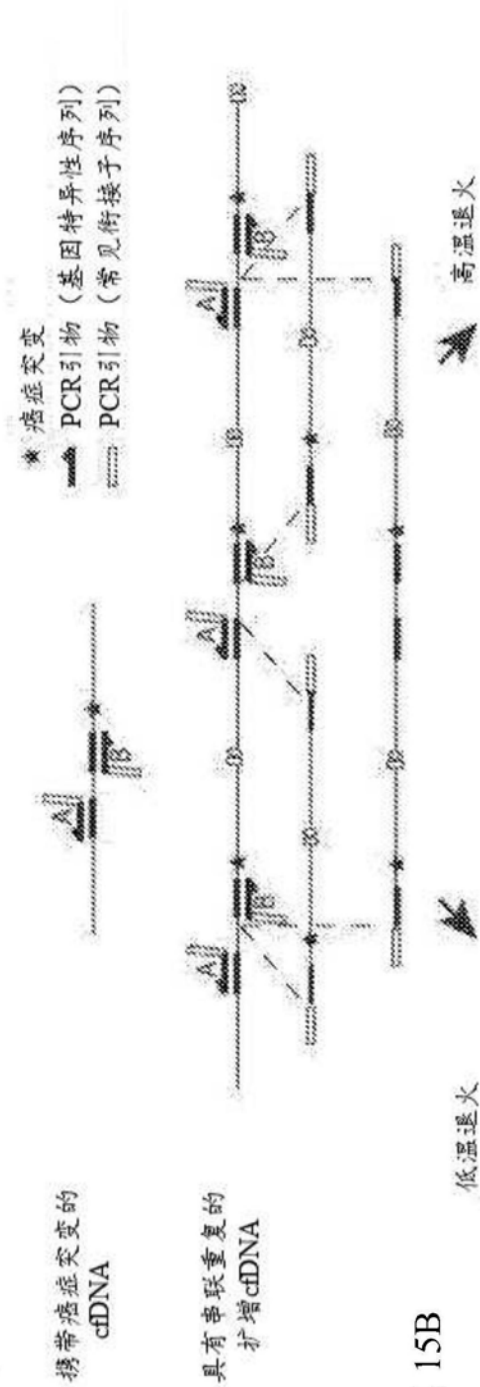


图14

图 15A



CYP2C19点突变（外显子5中的核苷酸681）

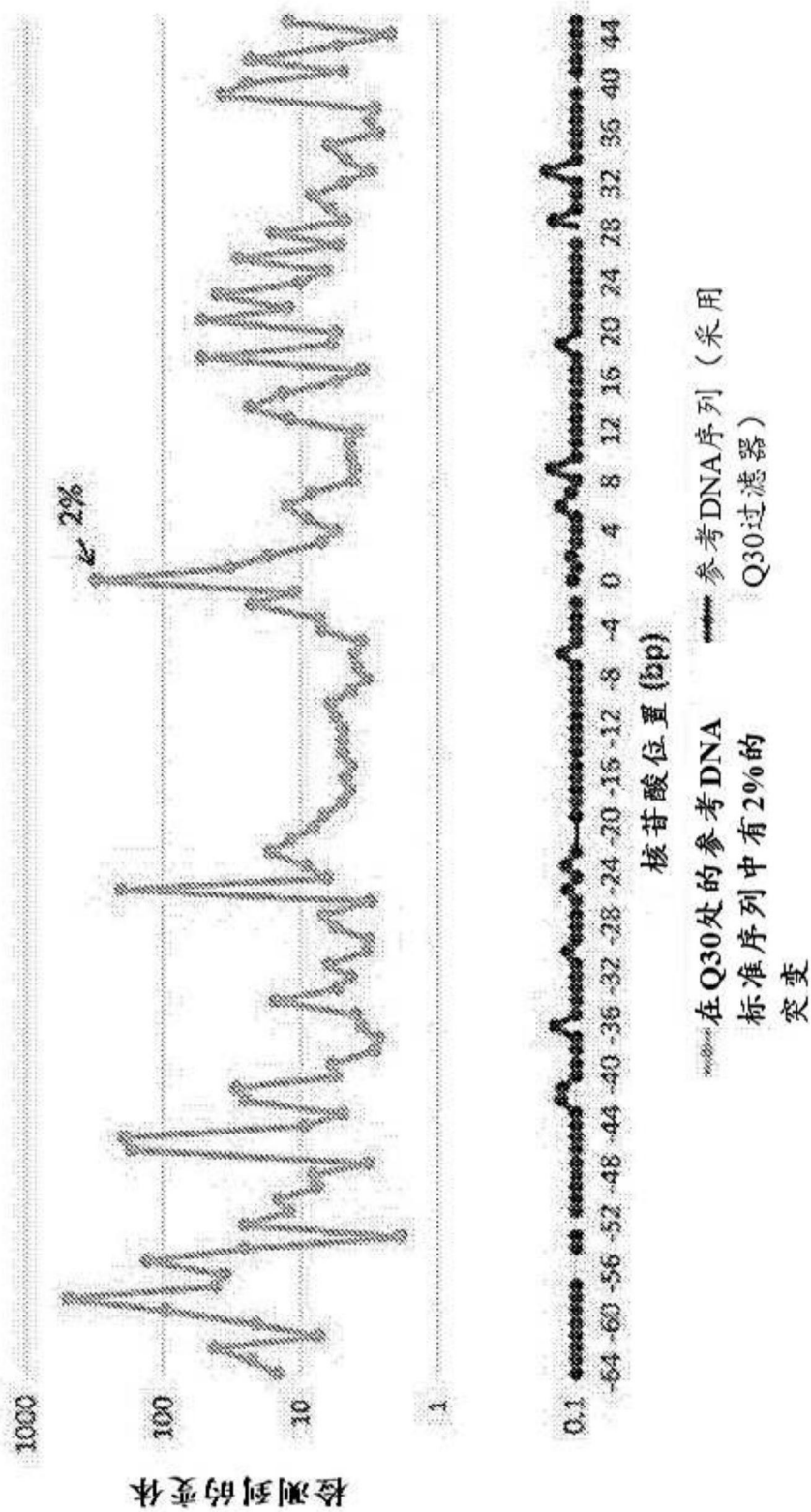


图16

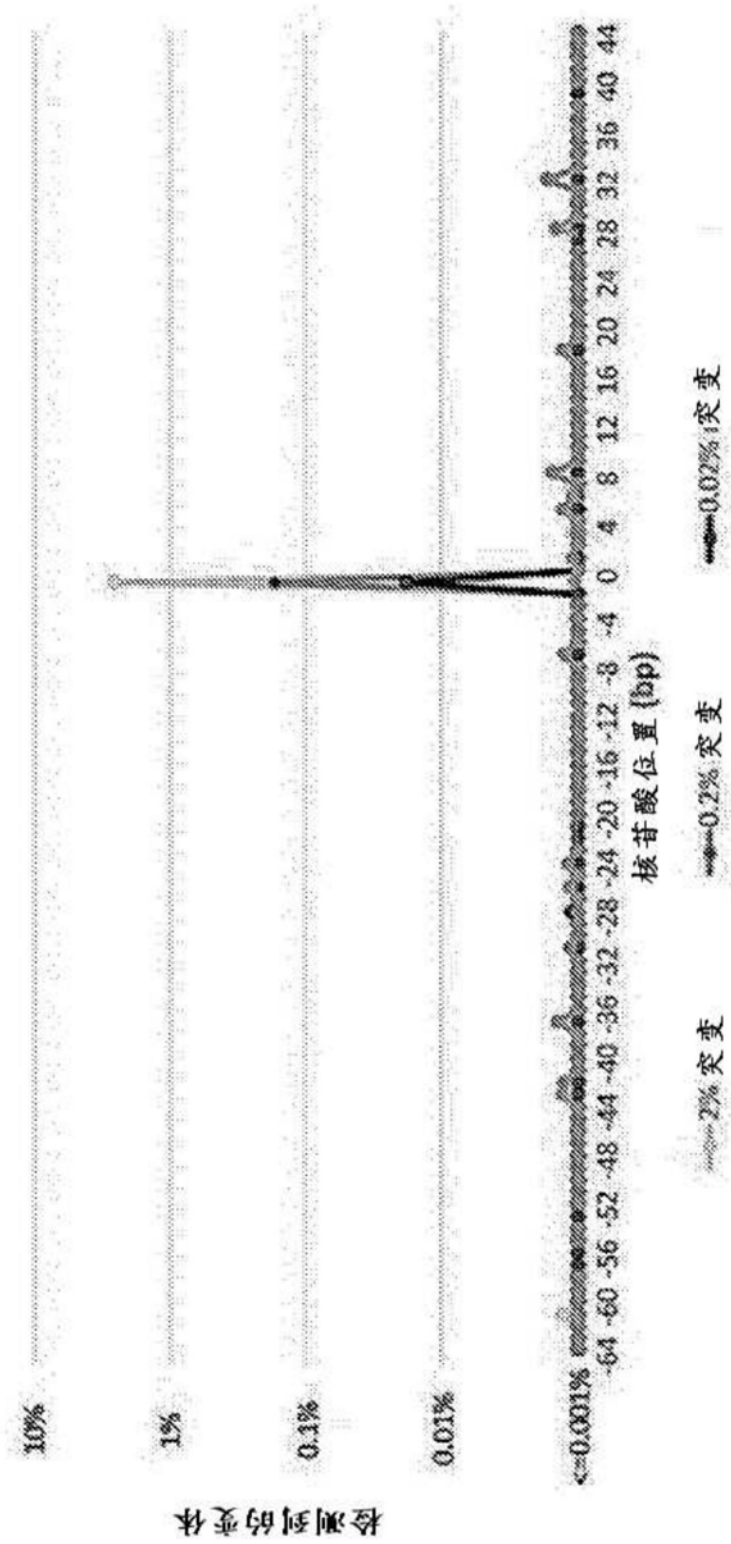


图17

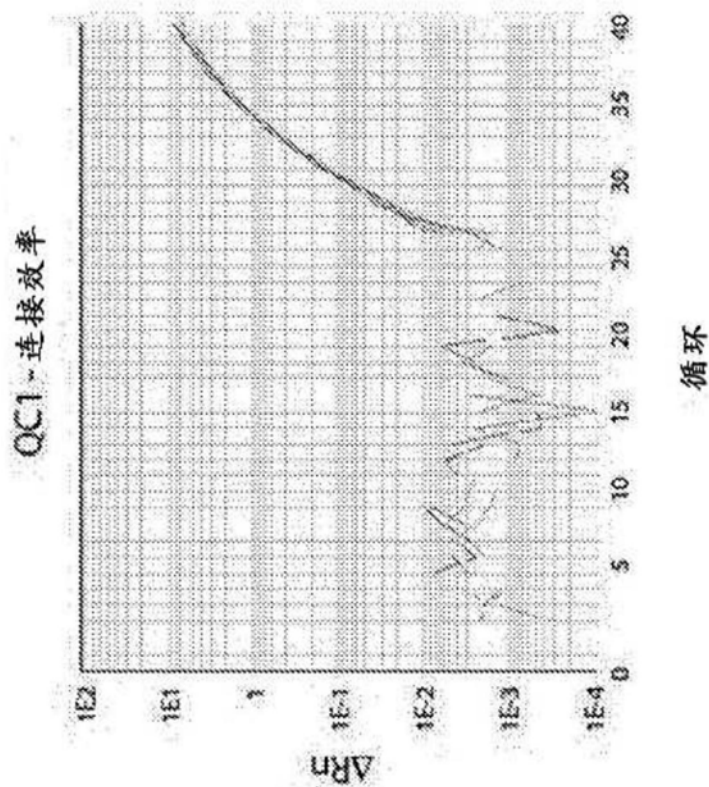


图18A

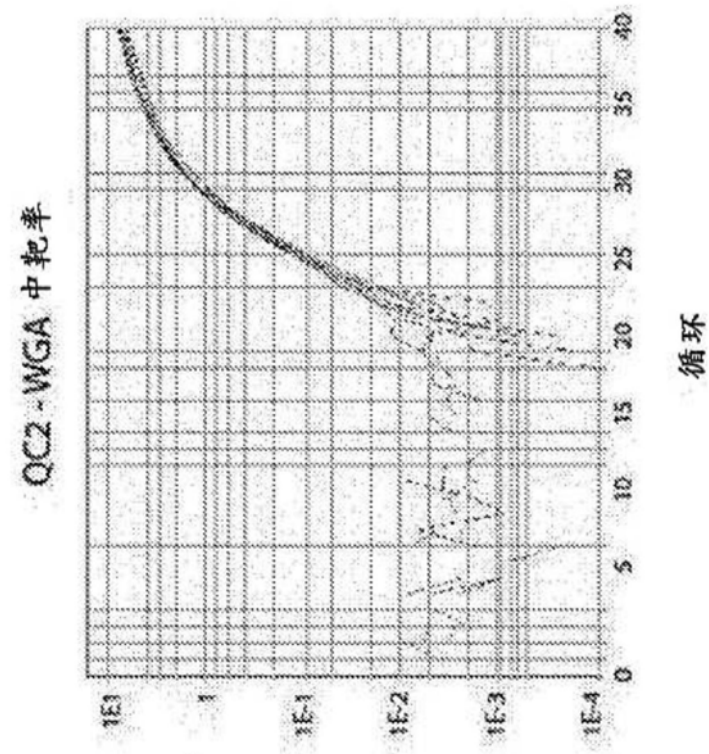


图18B



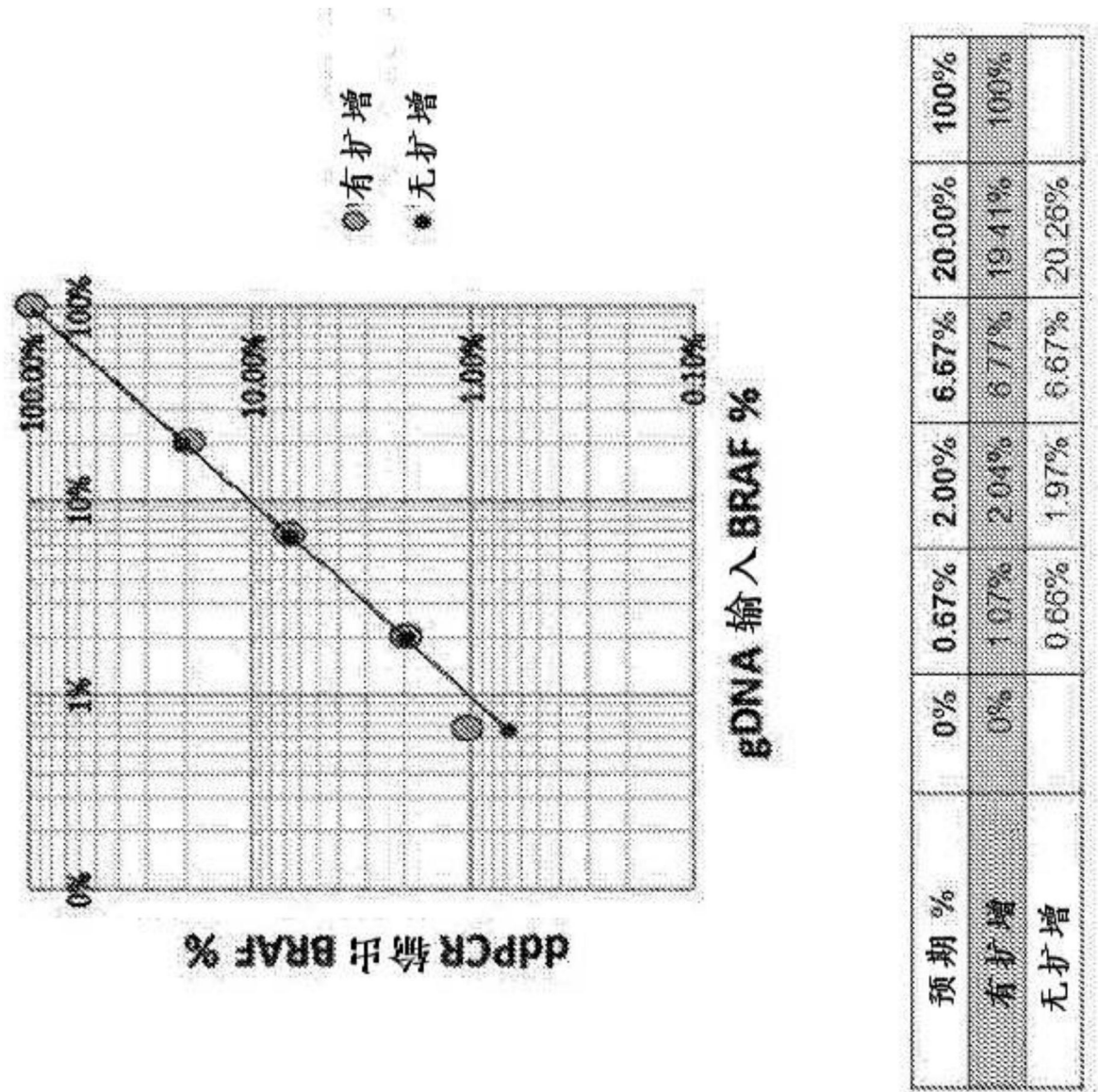


图19

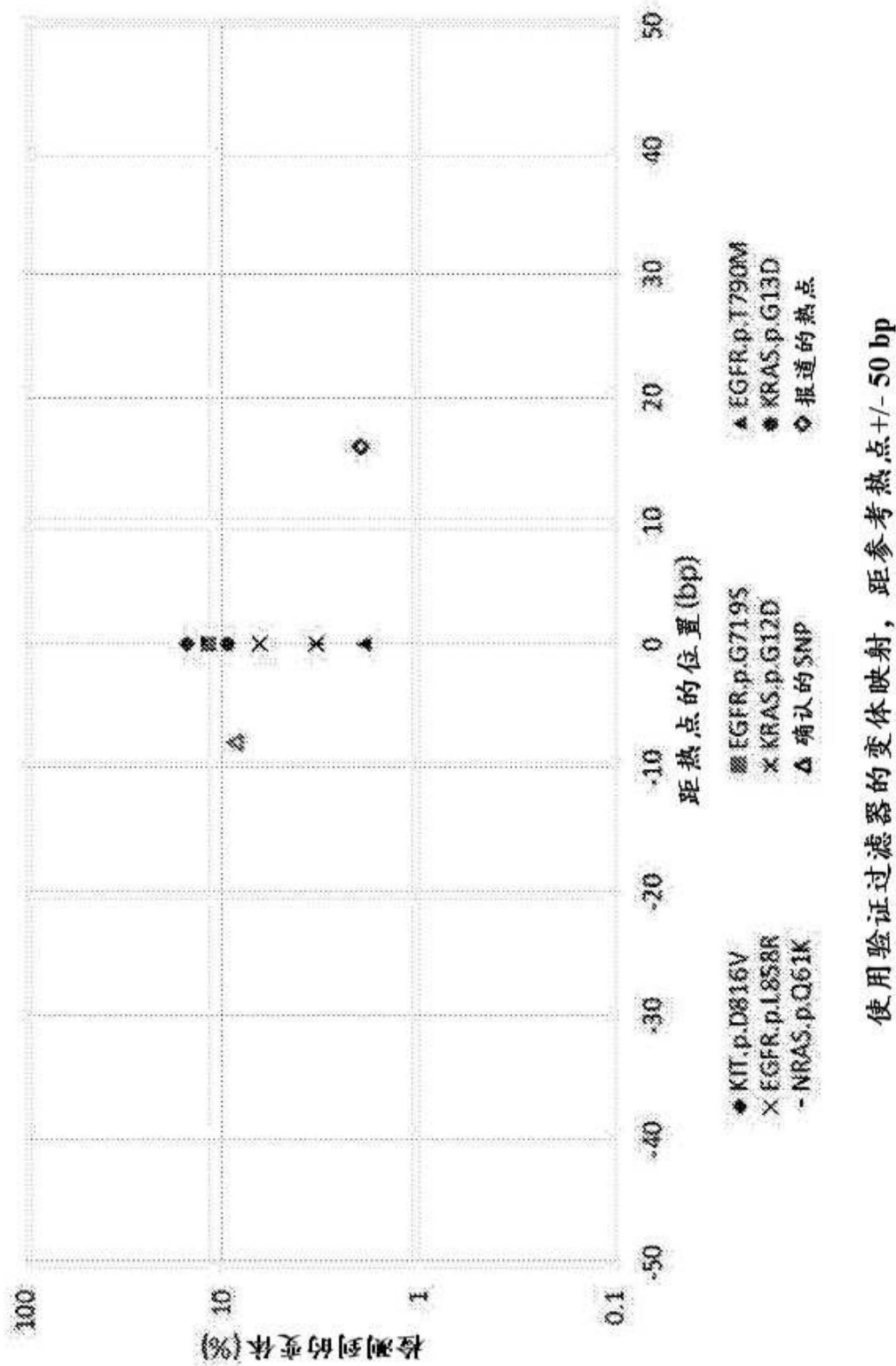


图20

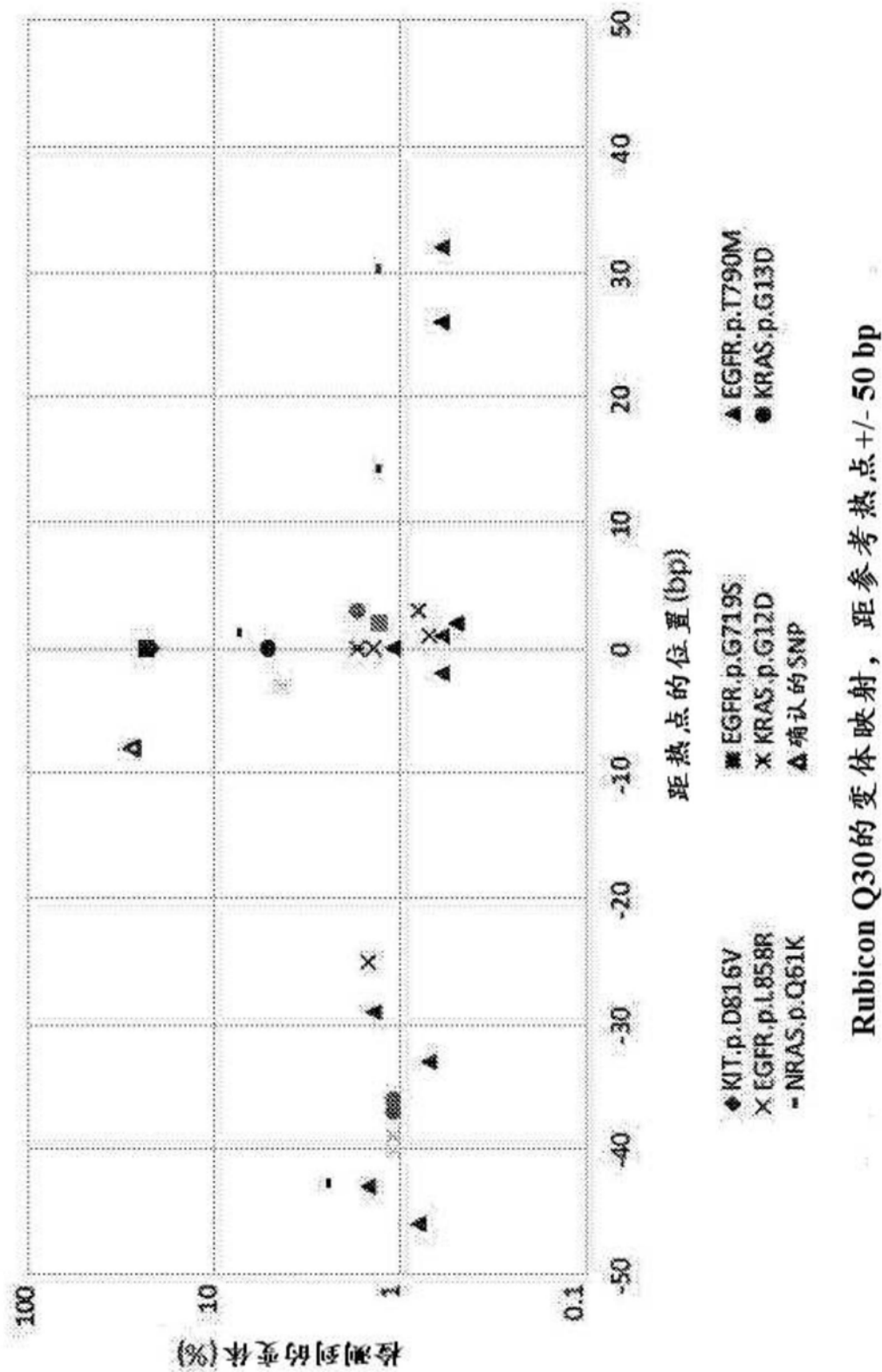


图21

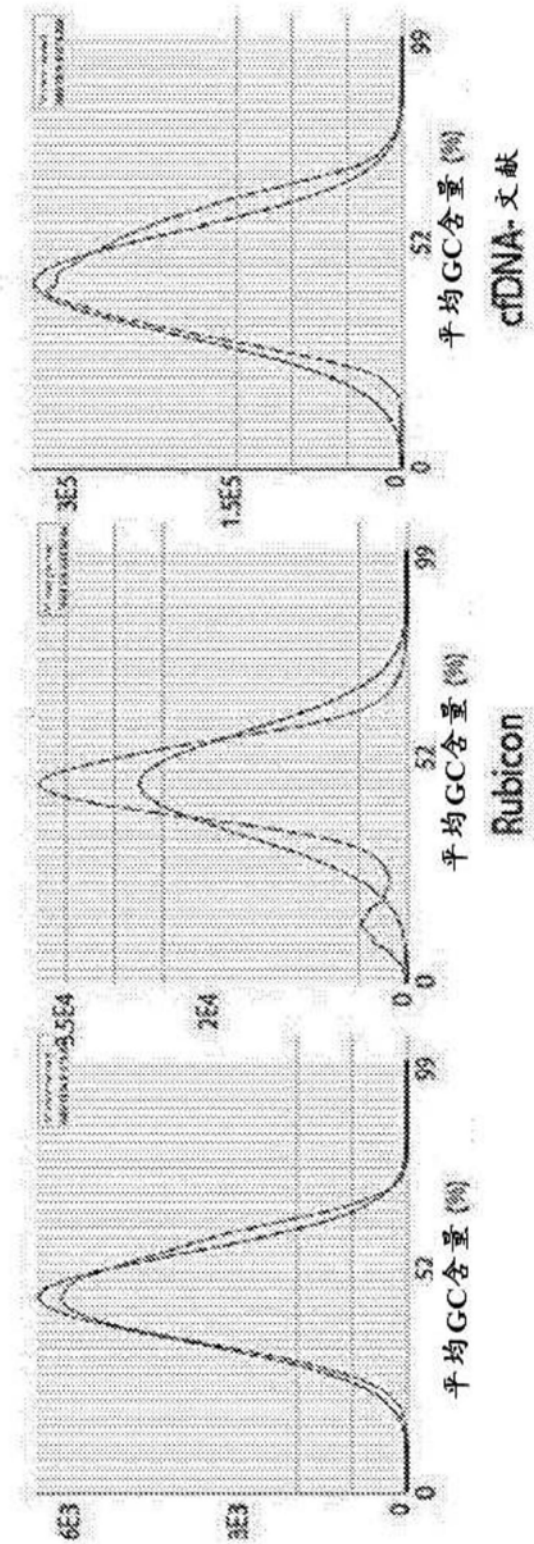


图22

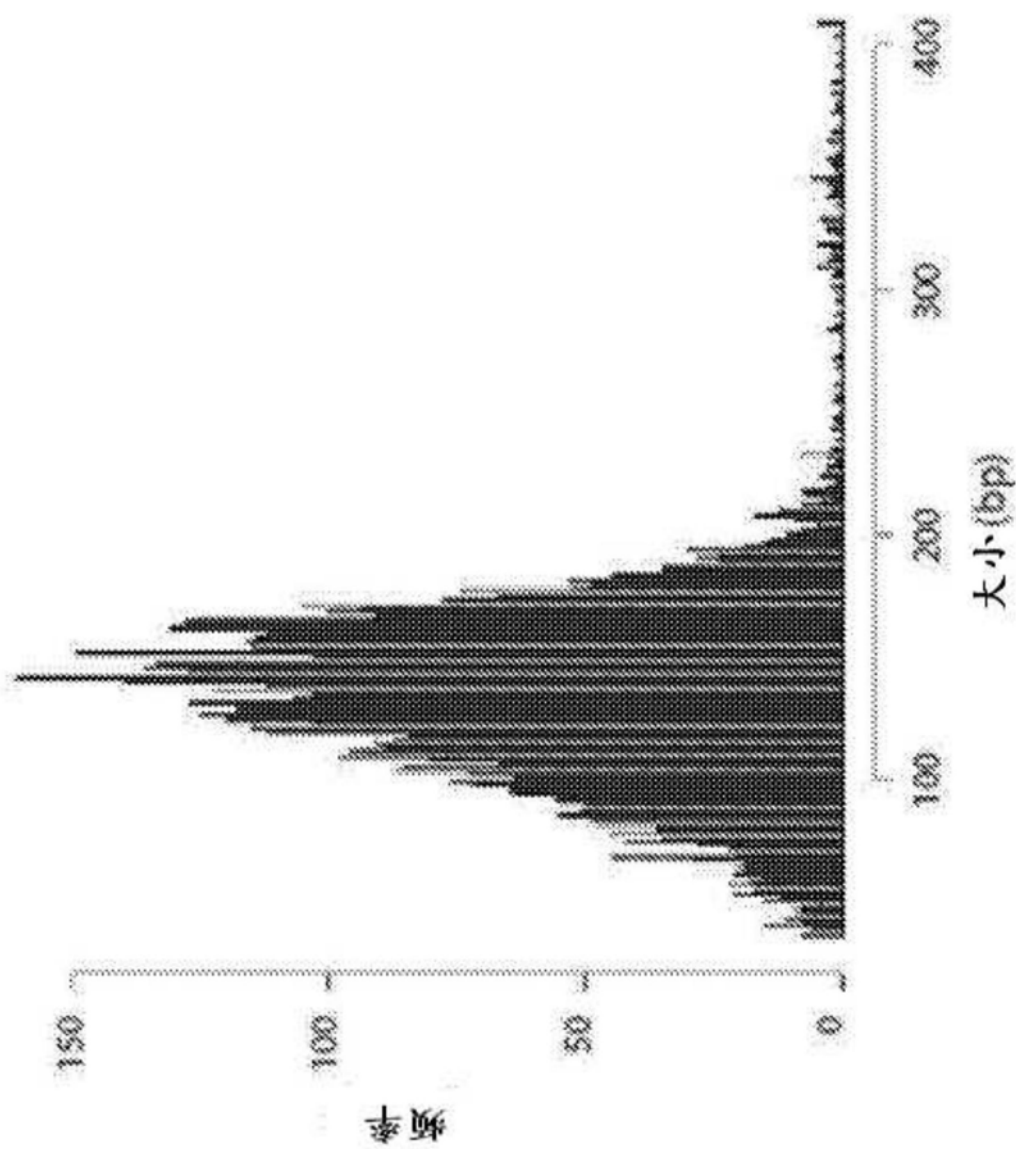


图23

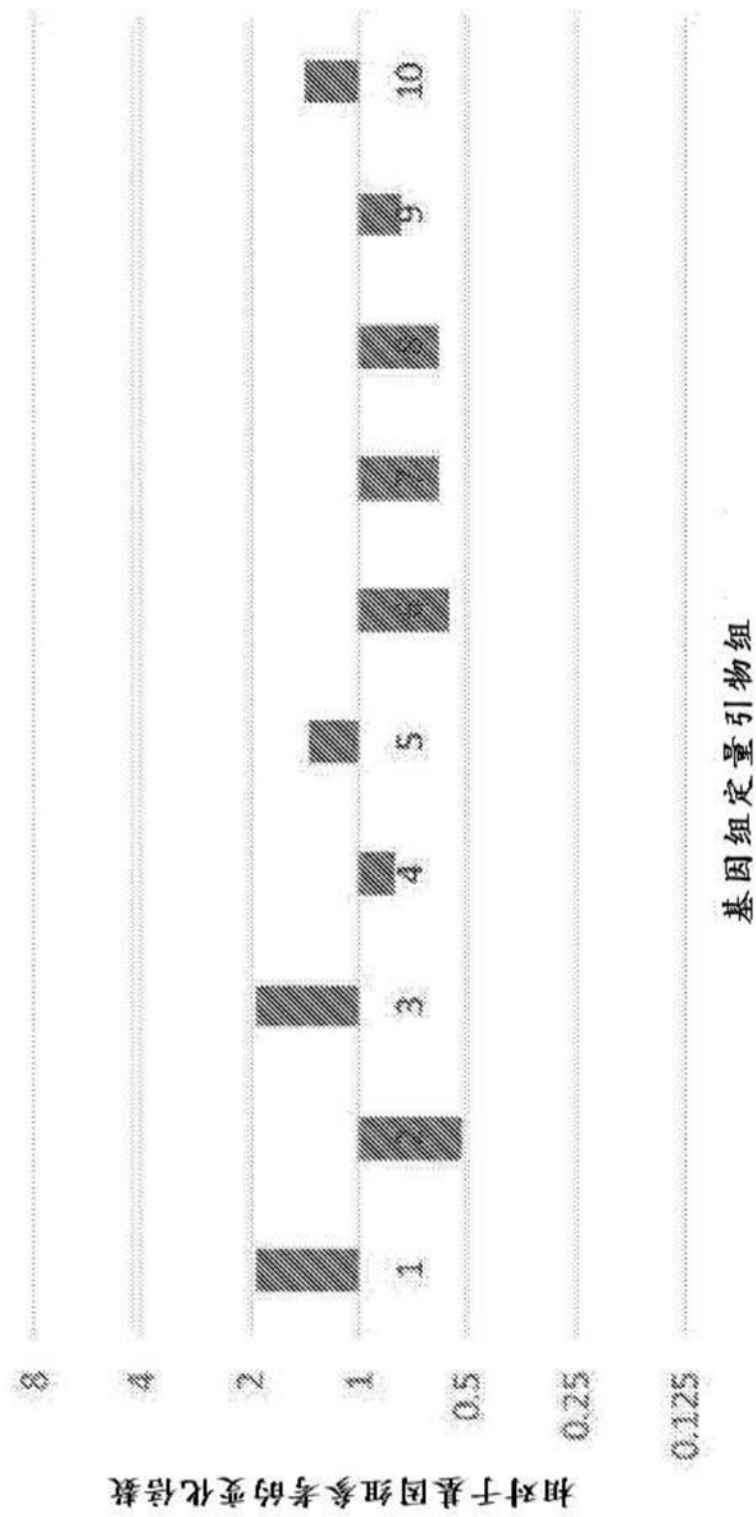


图24

图 25A  
平端连接 (自动标记)

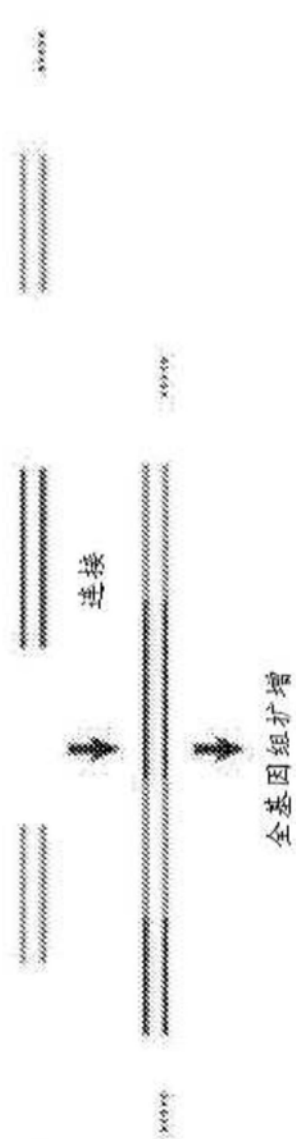
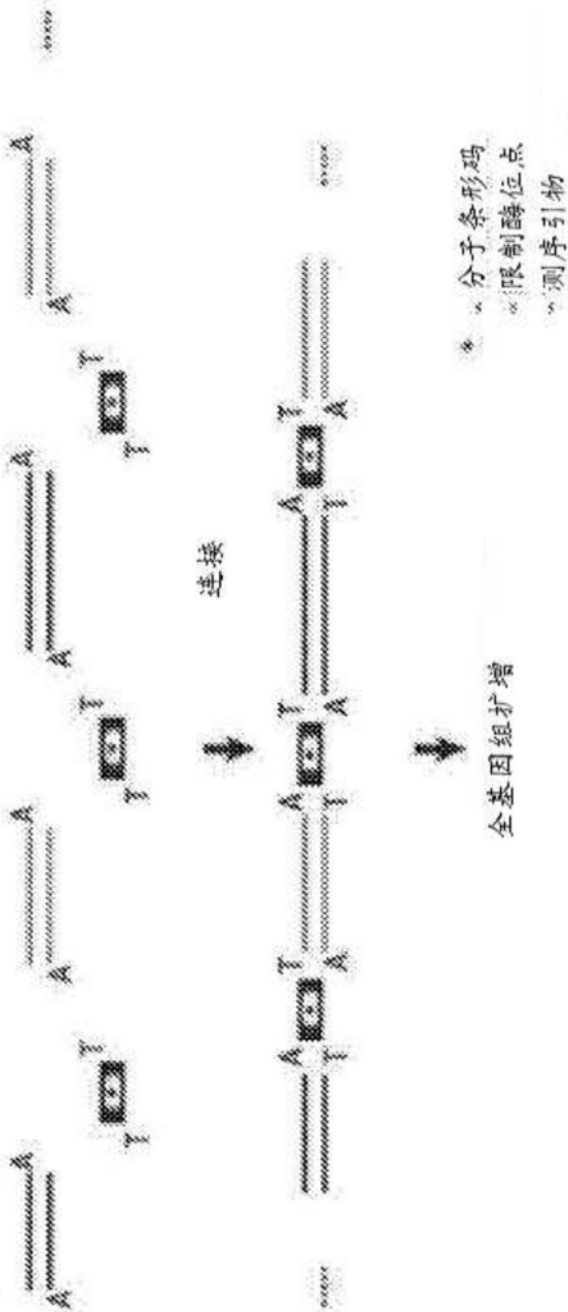


图 25B  
衔接子 (条形码) 连接



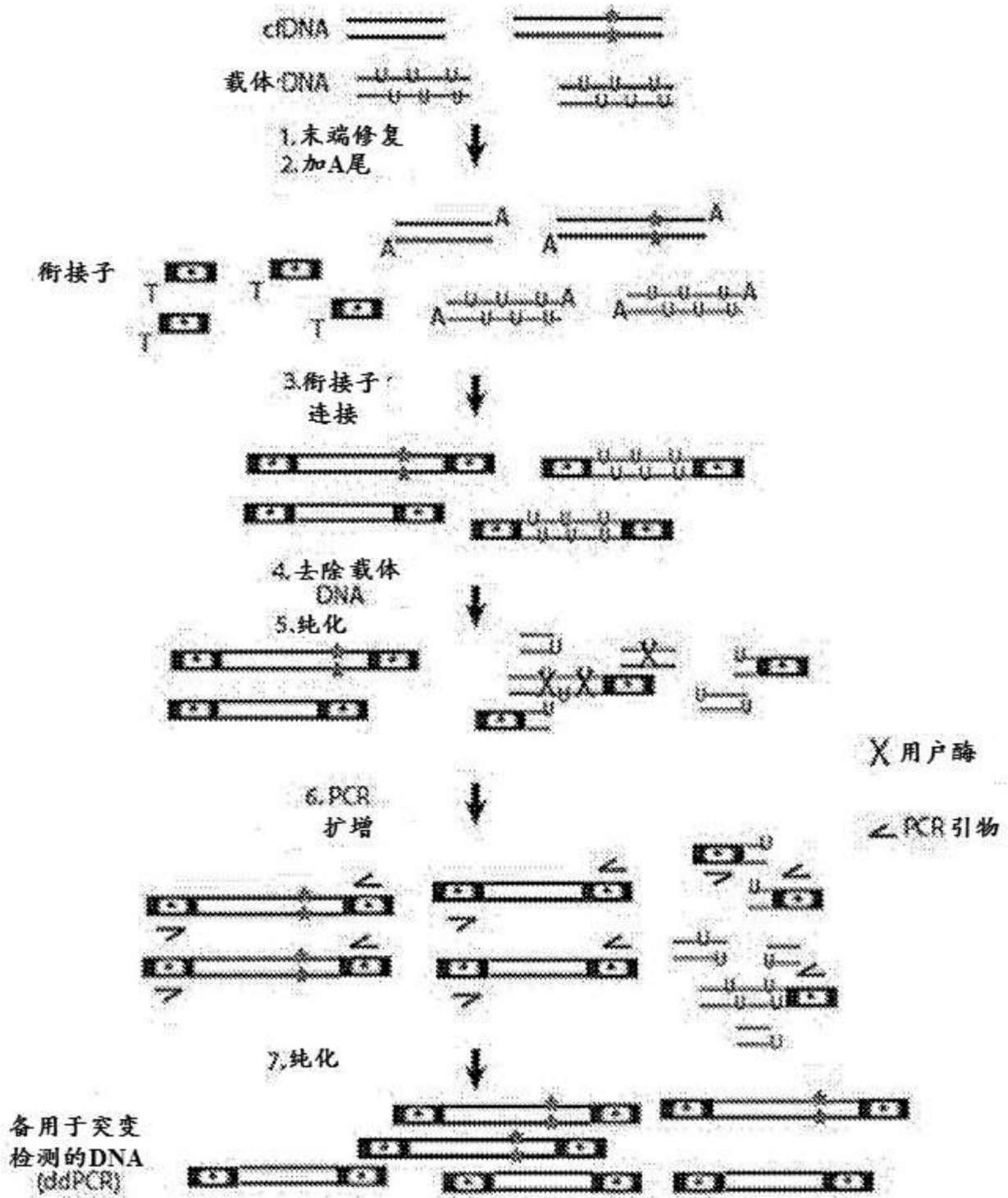


图26



图 27A

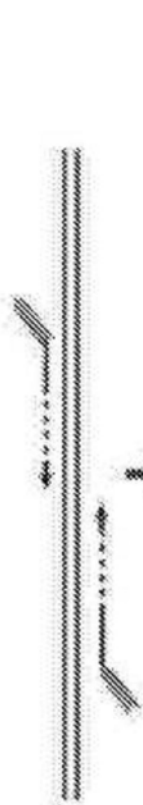


图 27B



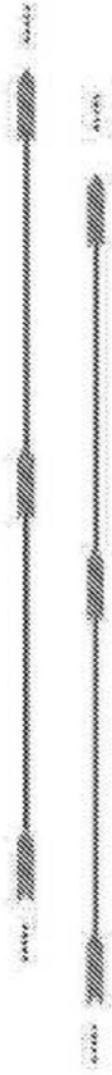
图 27C



图 27D



图 27E



高通量测序



图28A



图28B



图28C

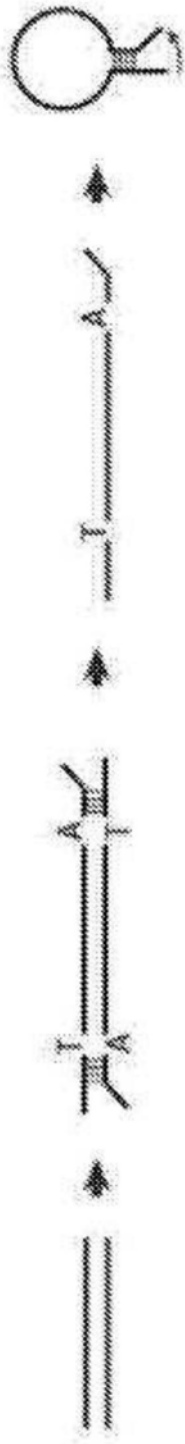


图28D

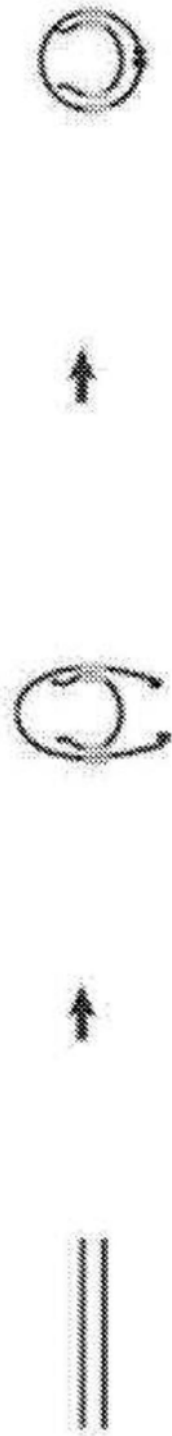


图28E

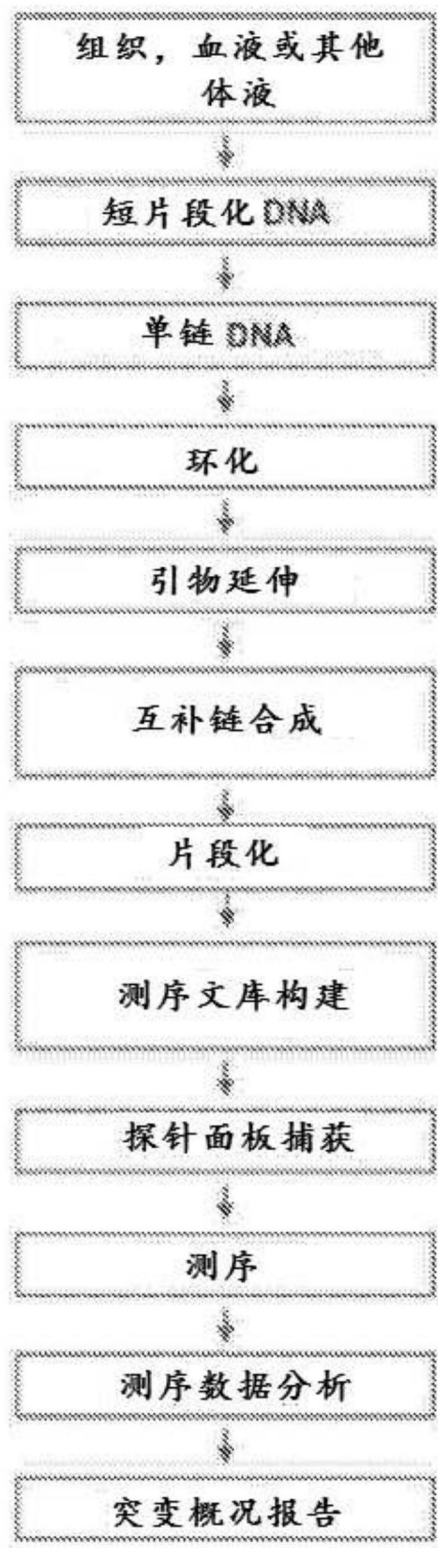


图29

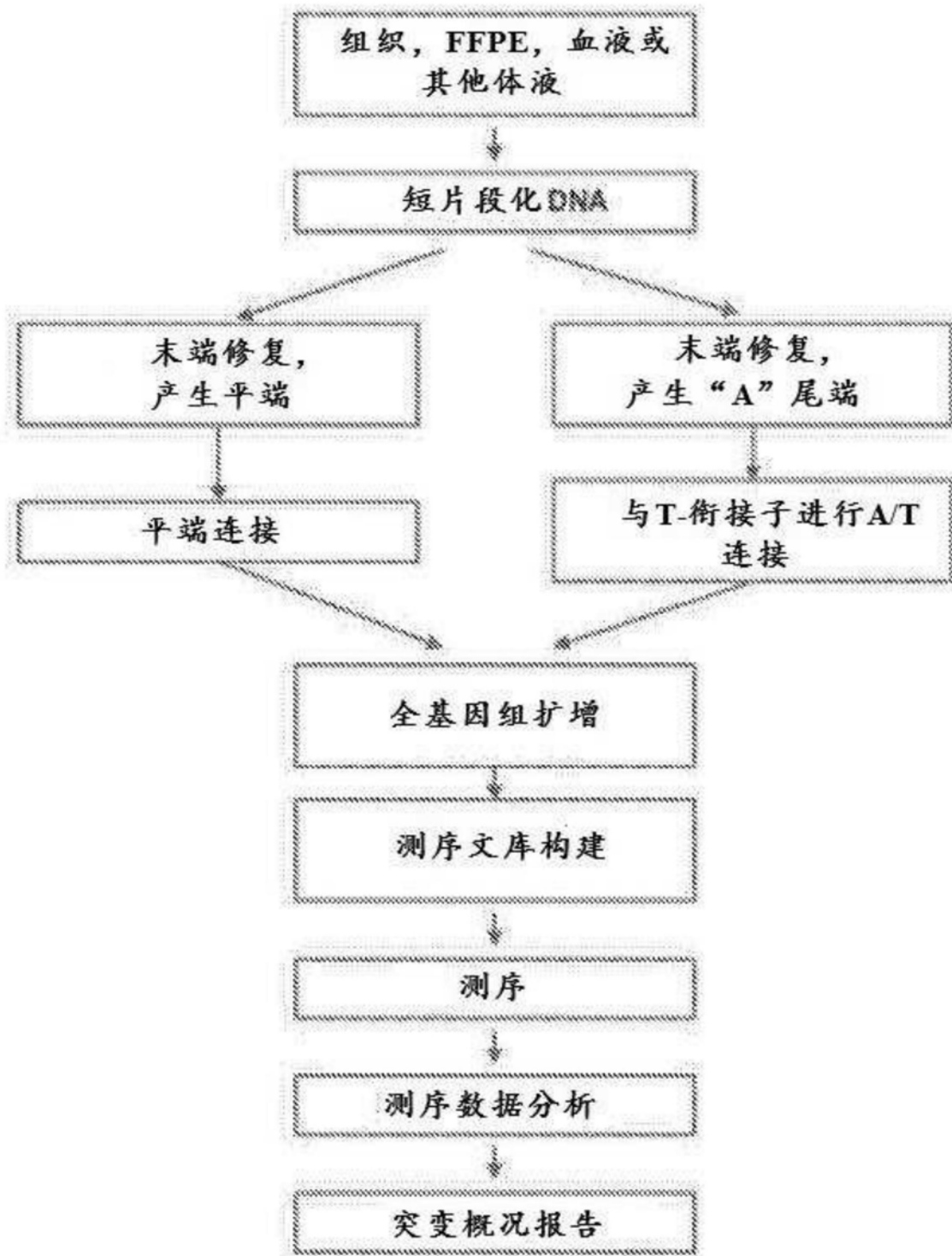


图30



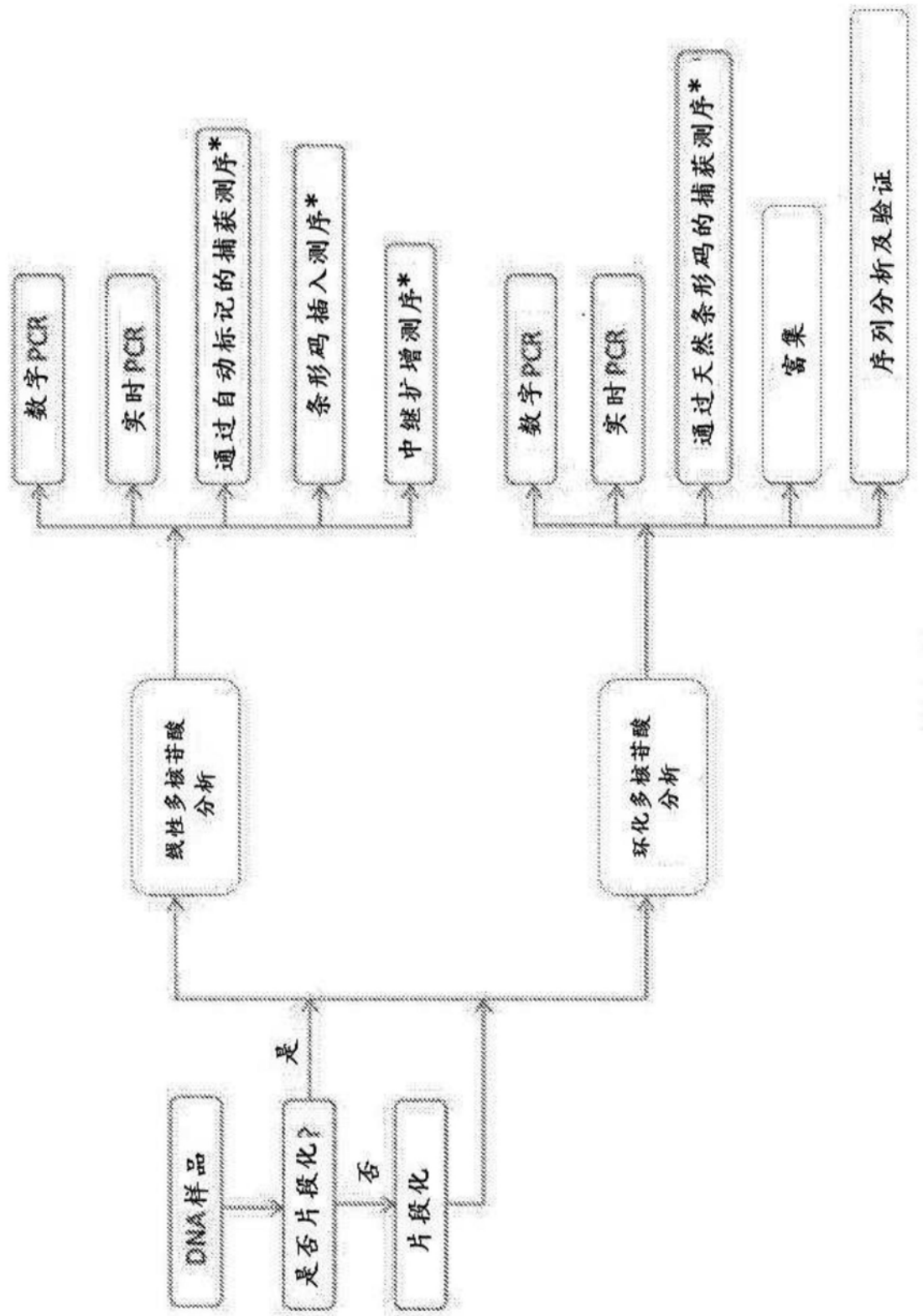


图31

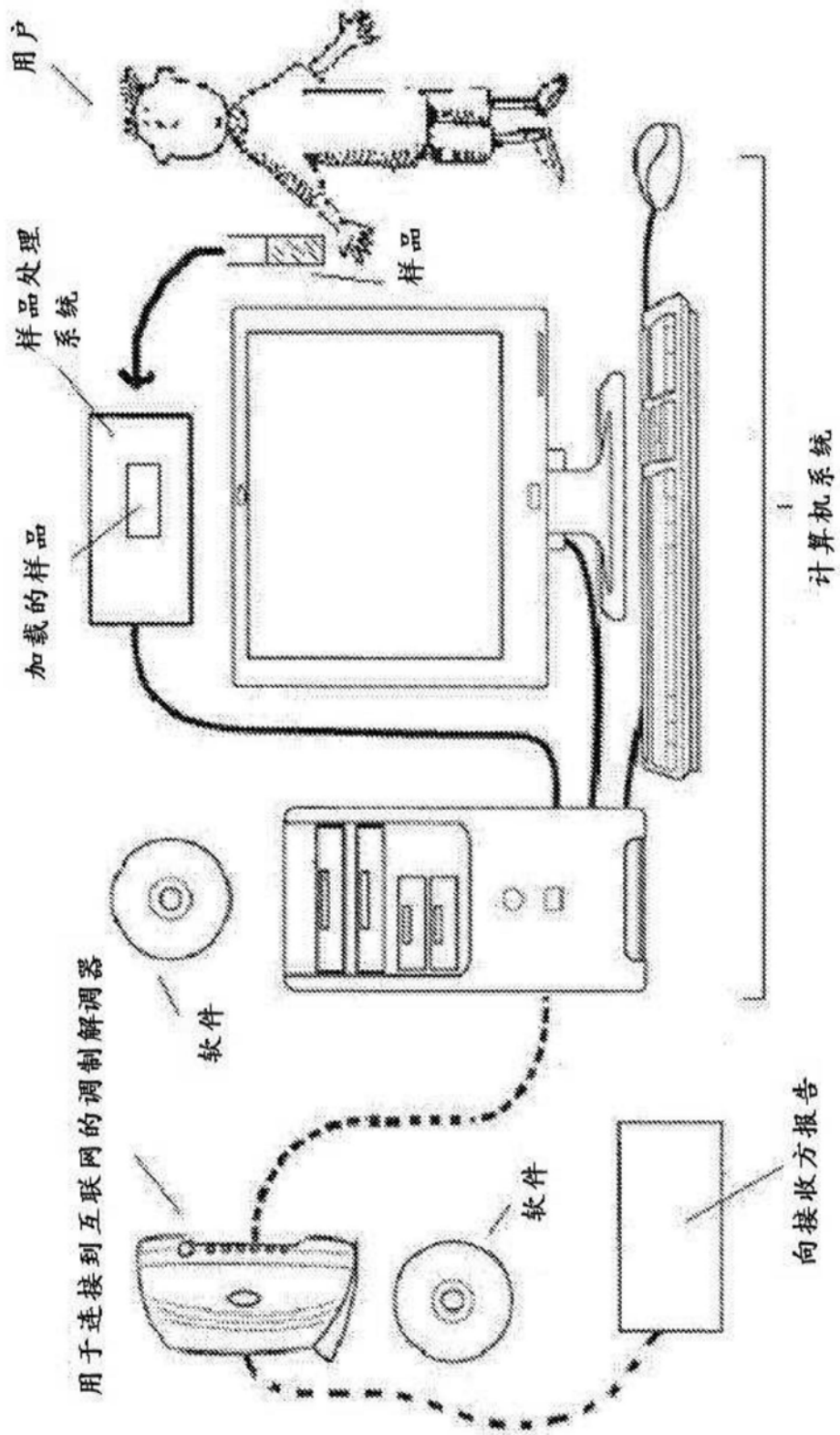


图32

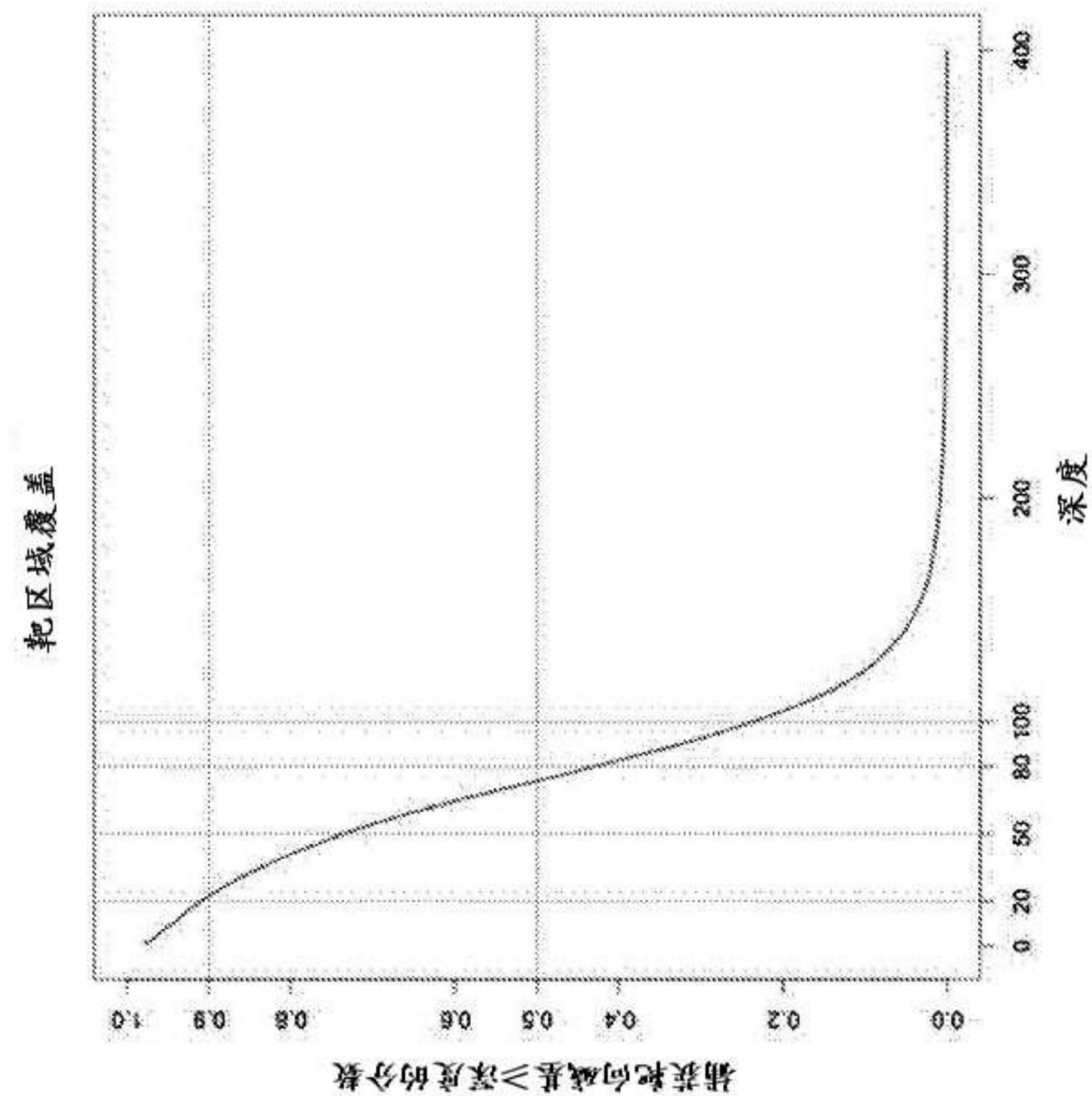


图33

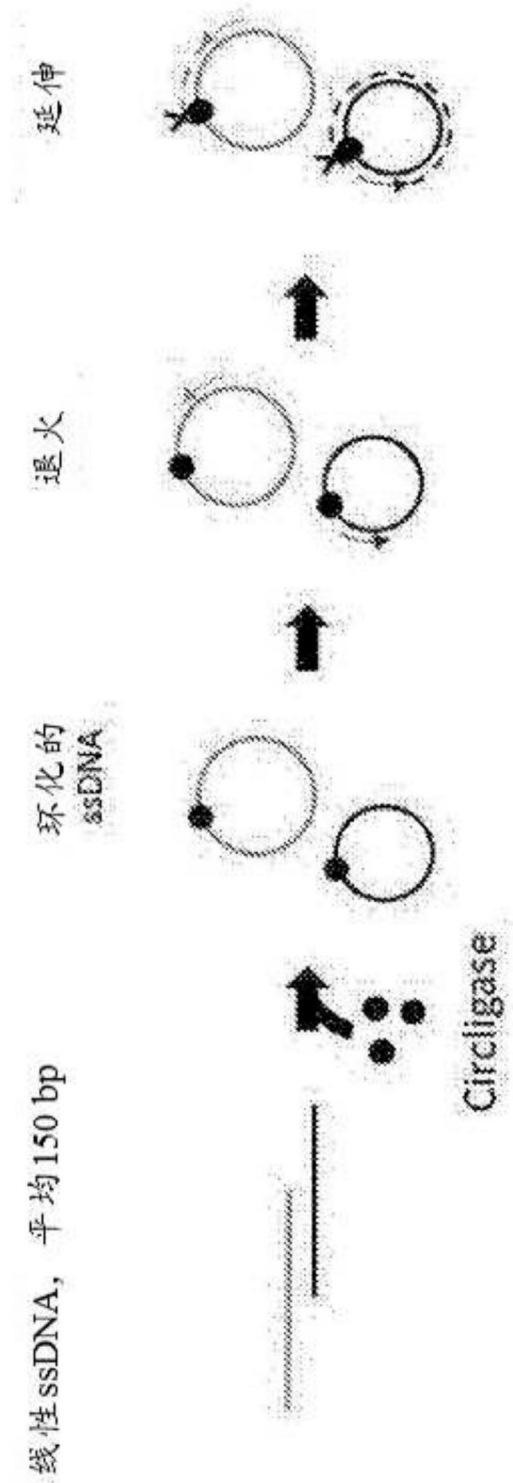


图34A

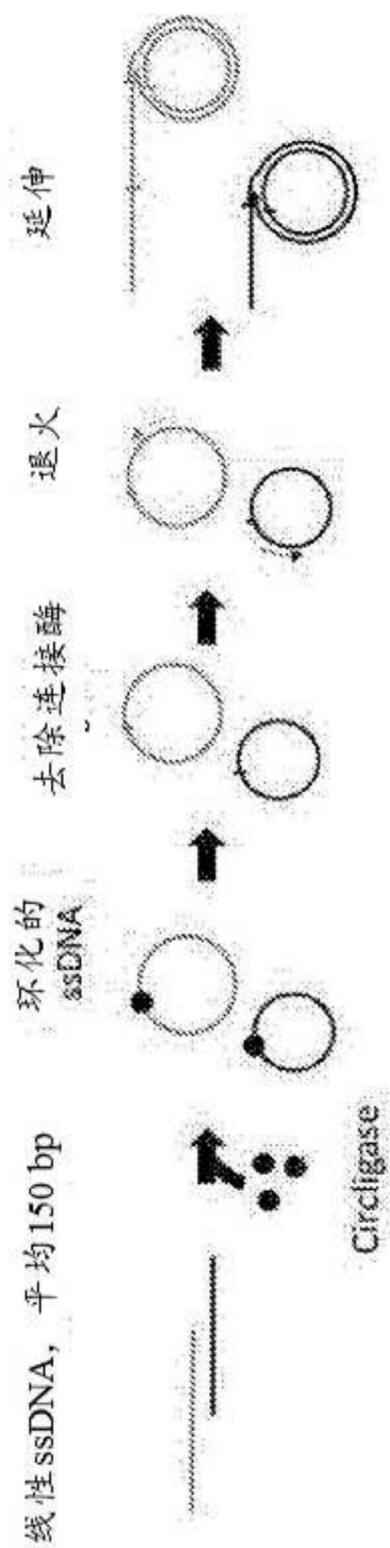


图34B

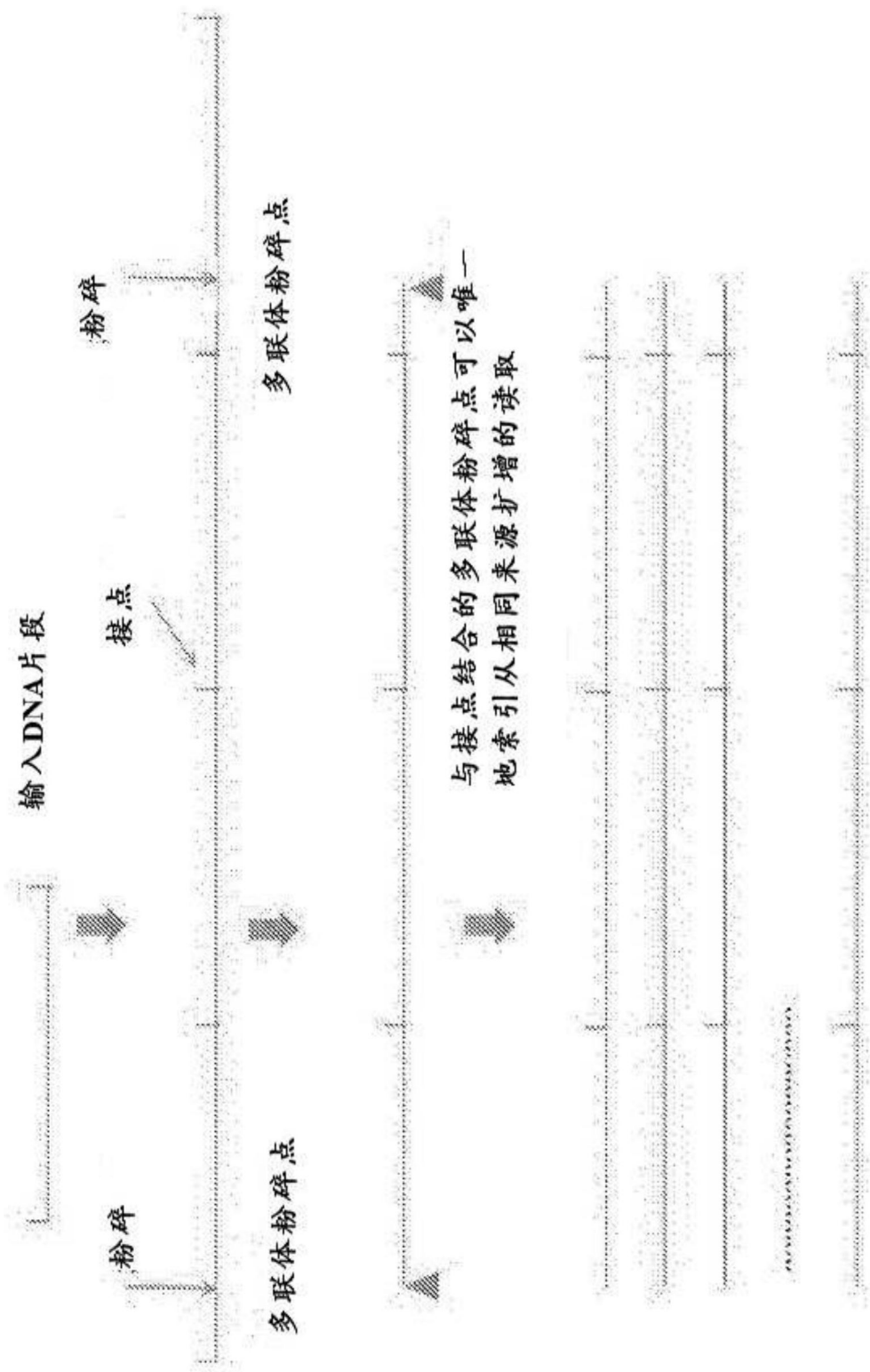


图35

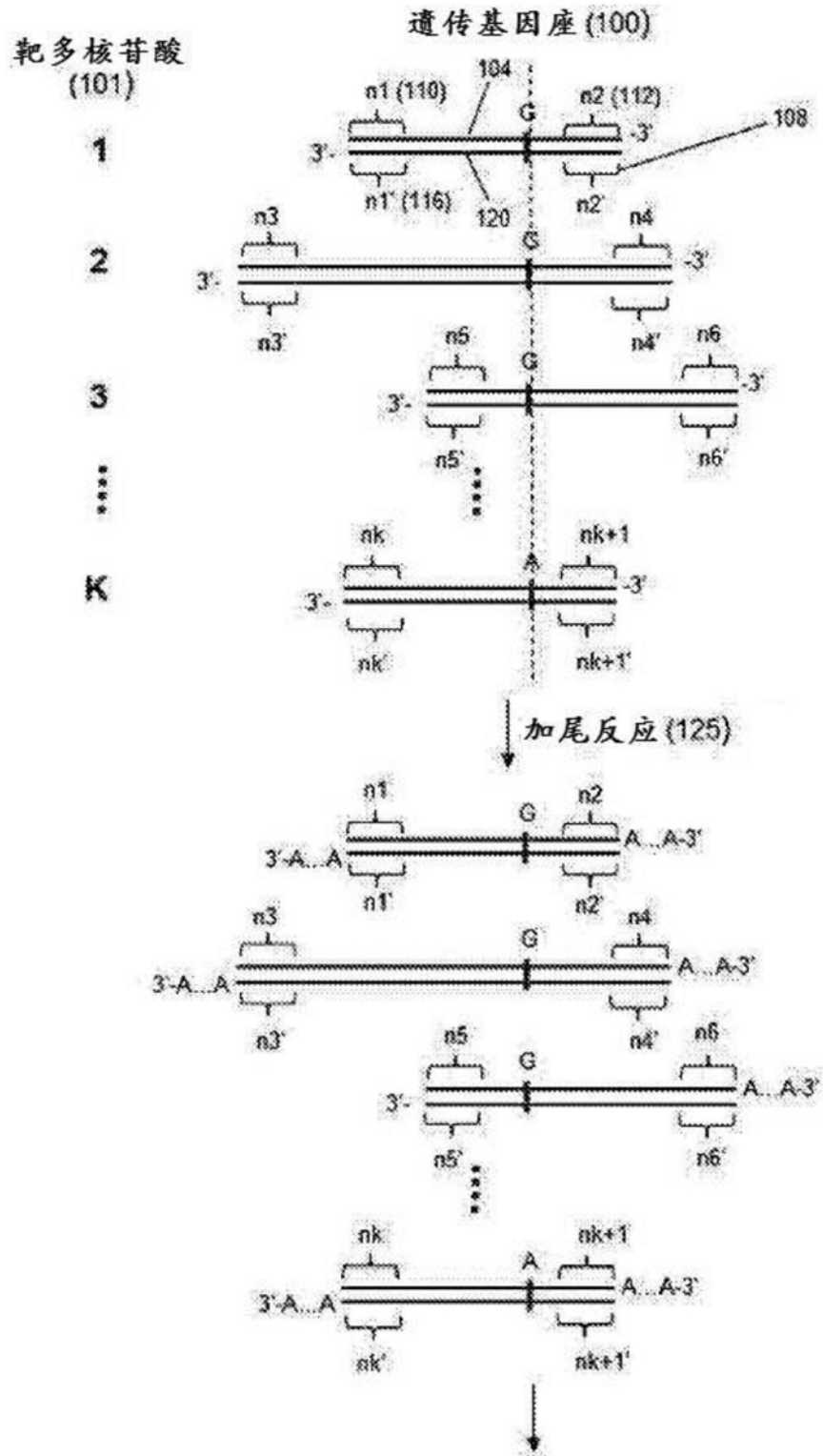


图36A

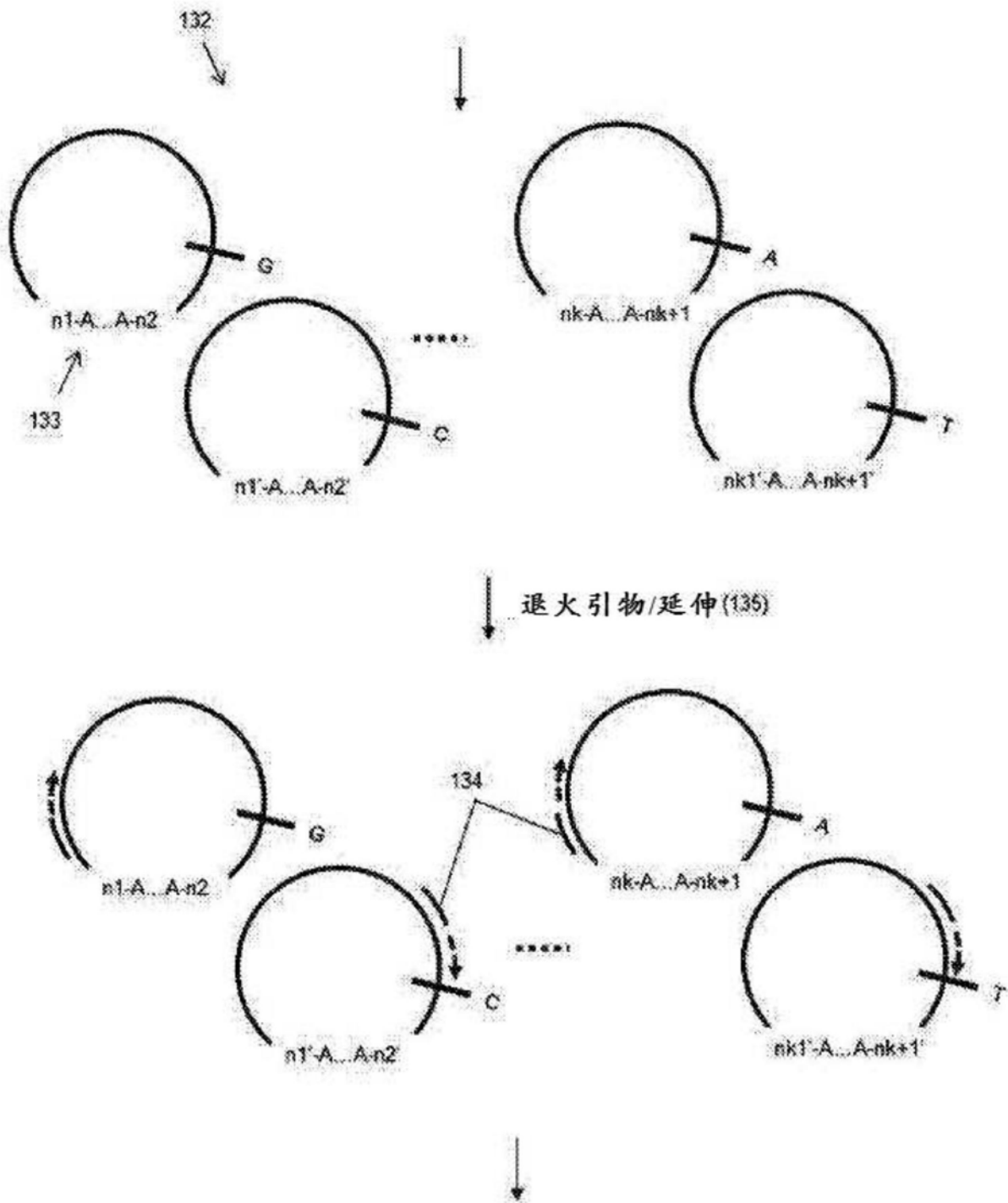


图36B



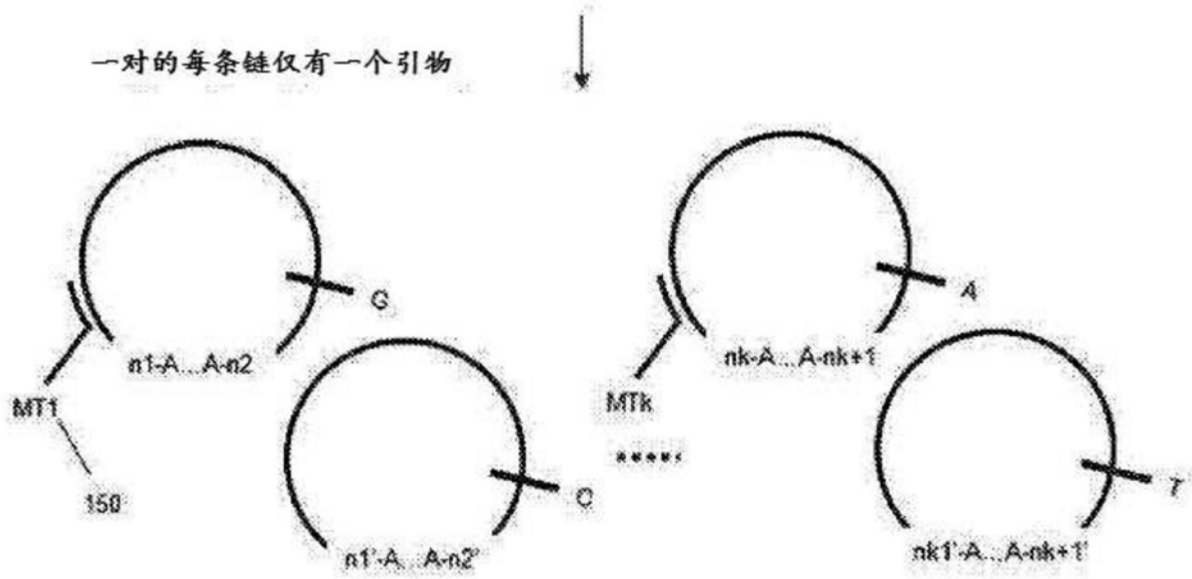


图36C

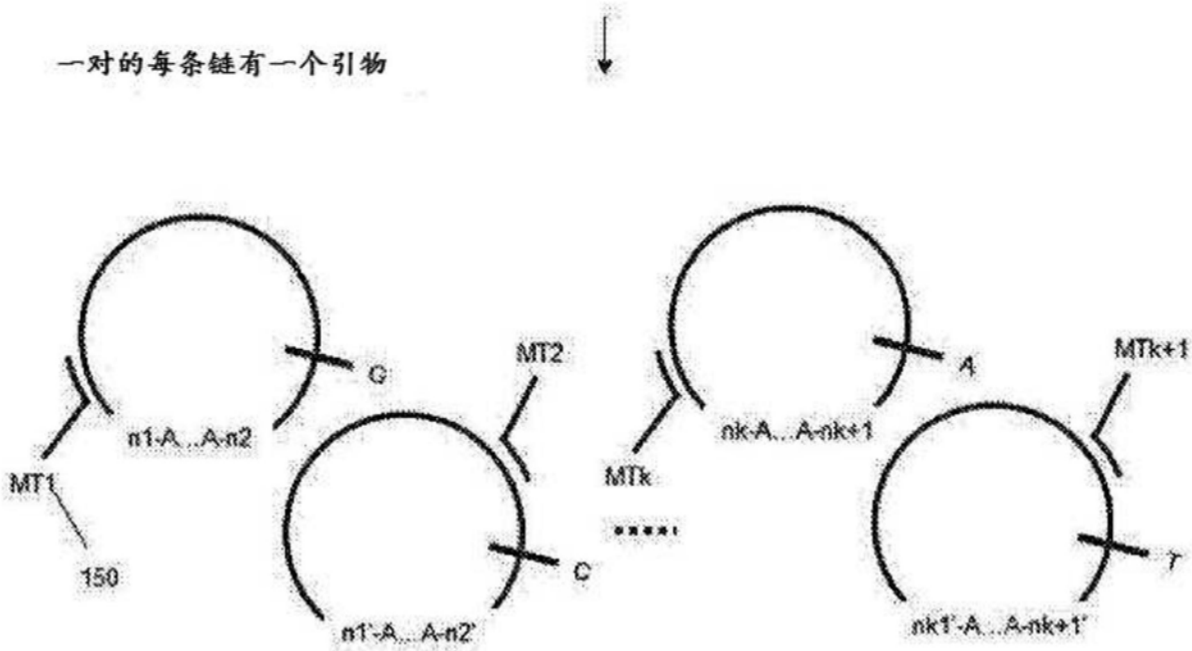


图36D

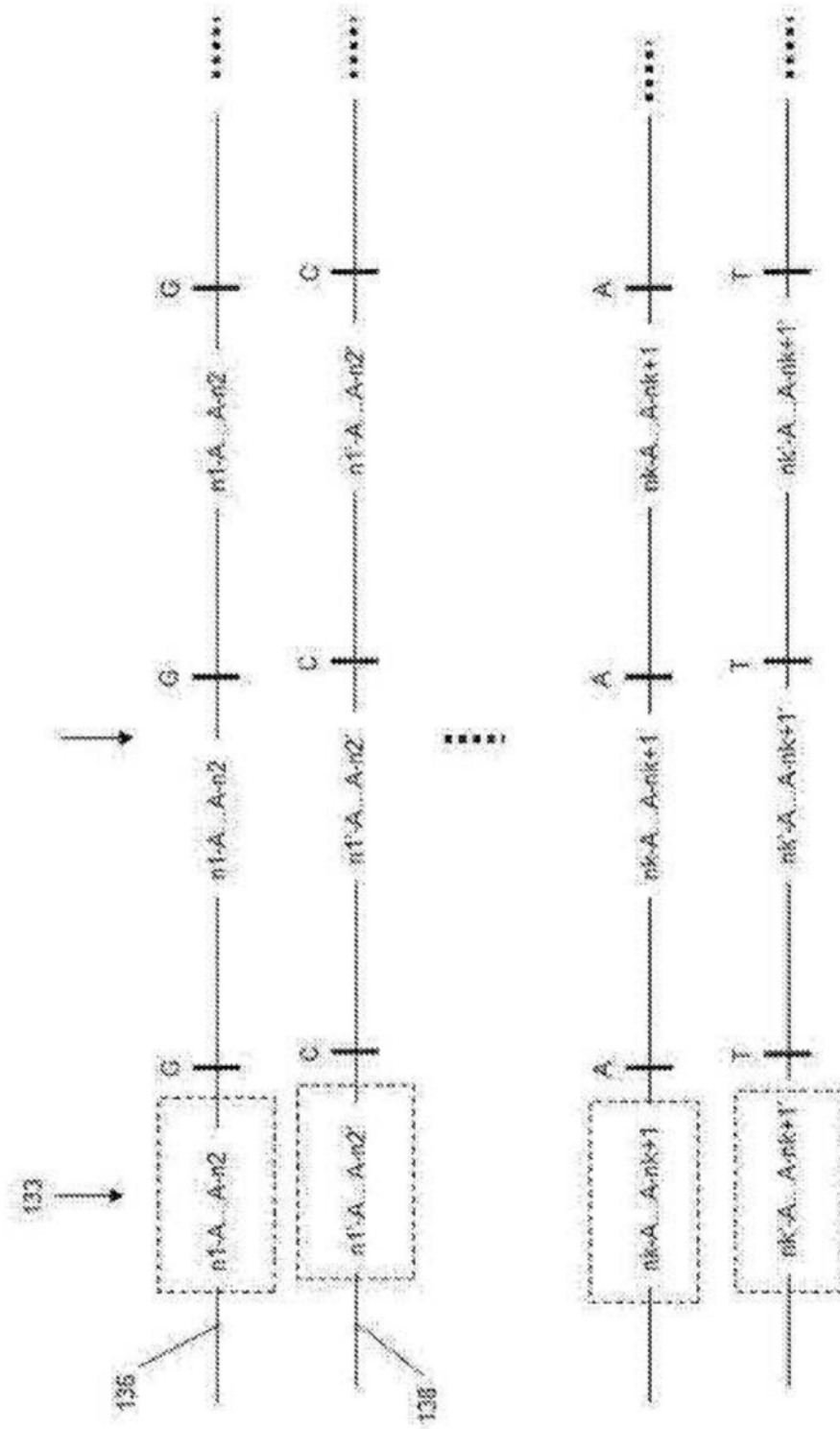


图36E

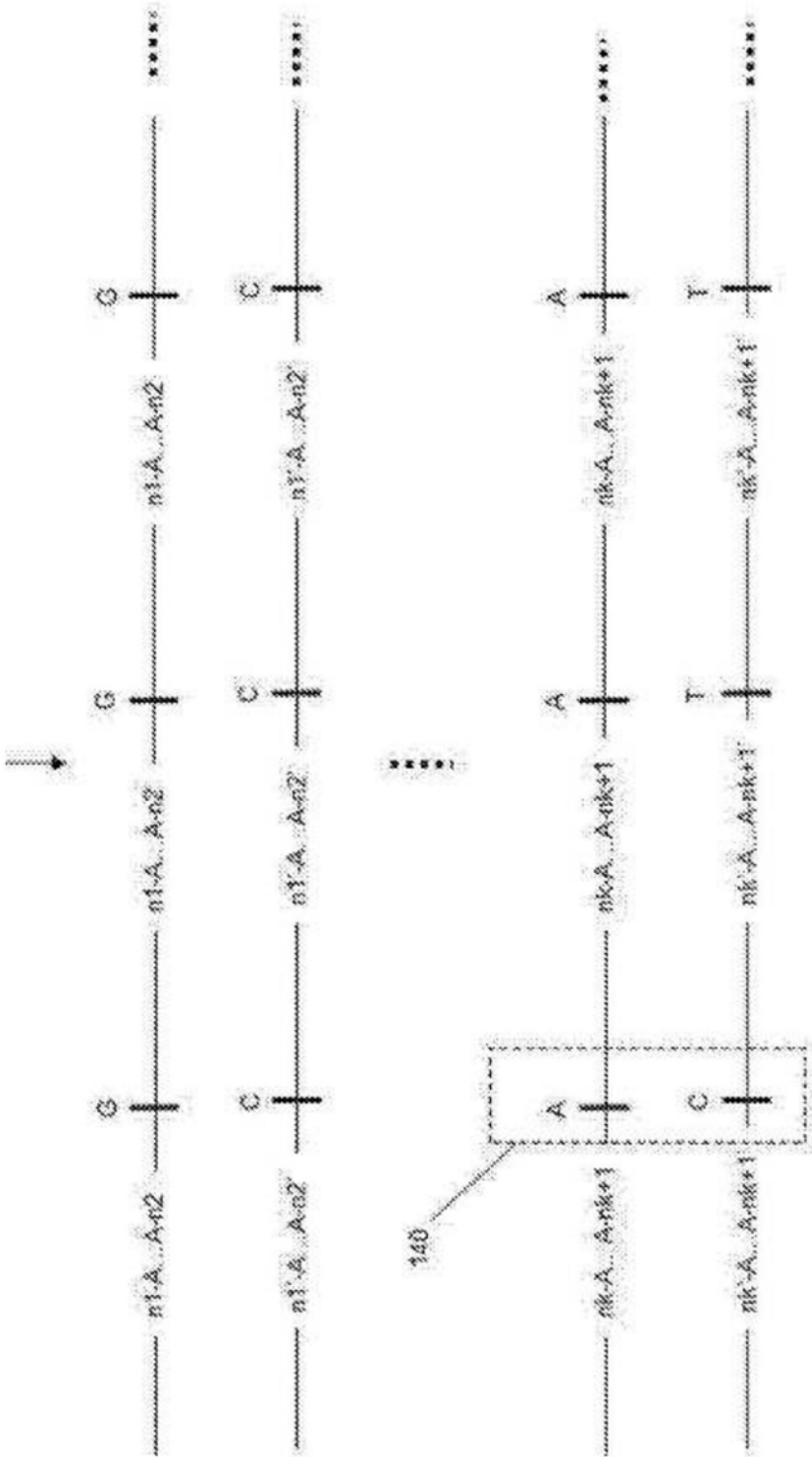


图36F

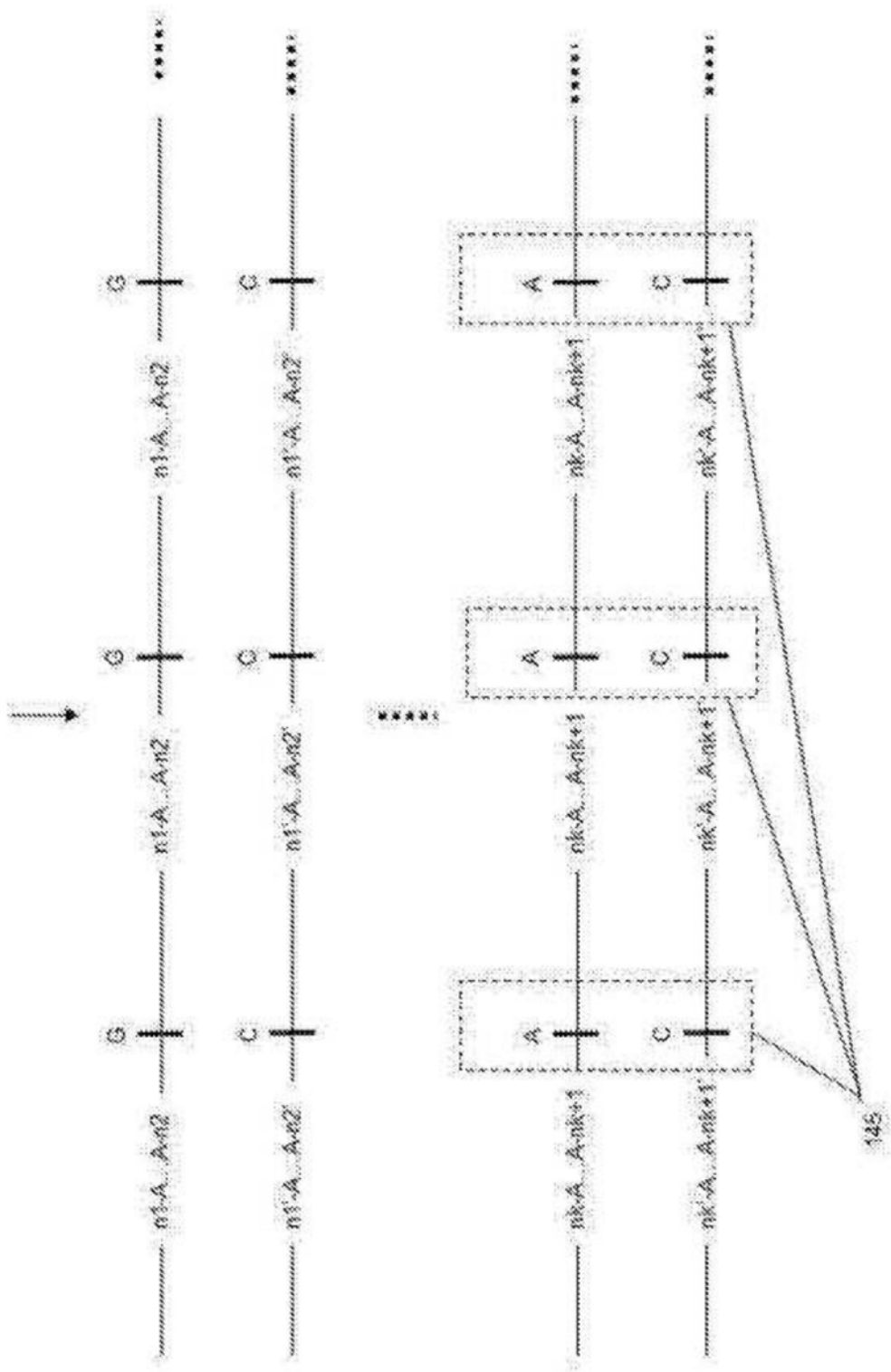


图36G

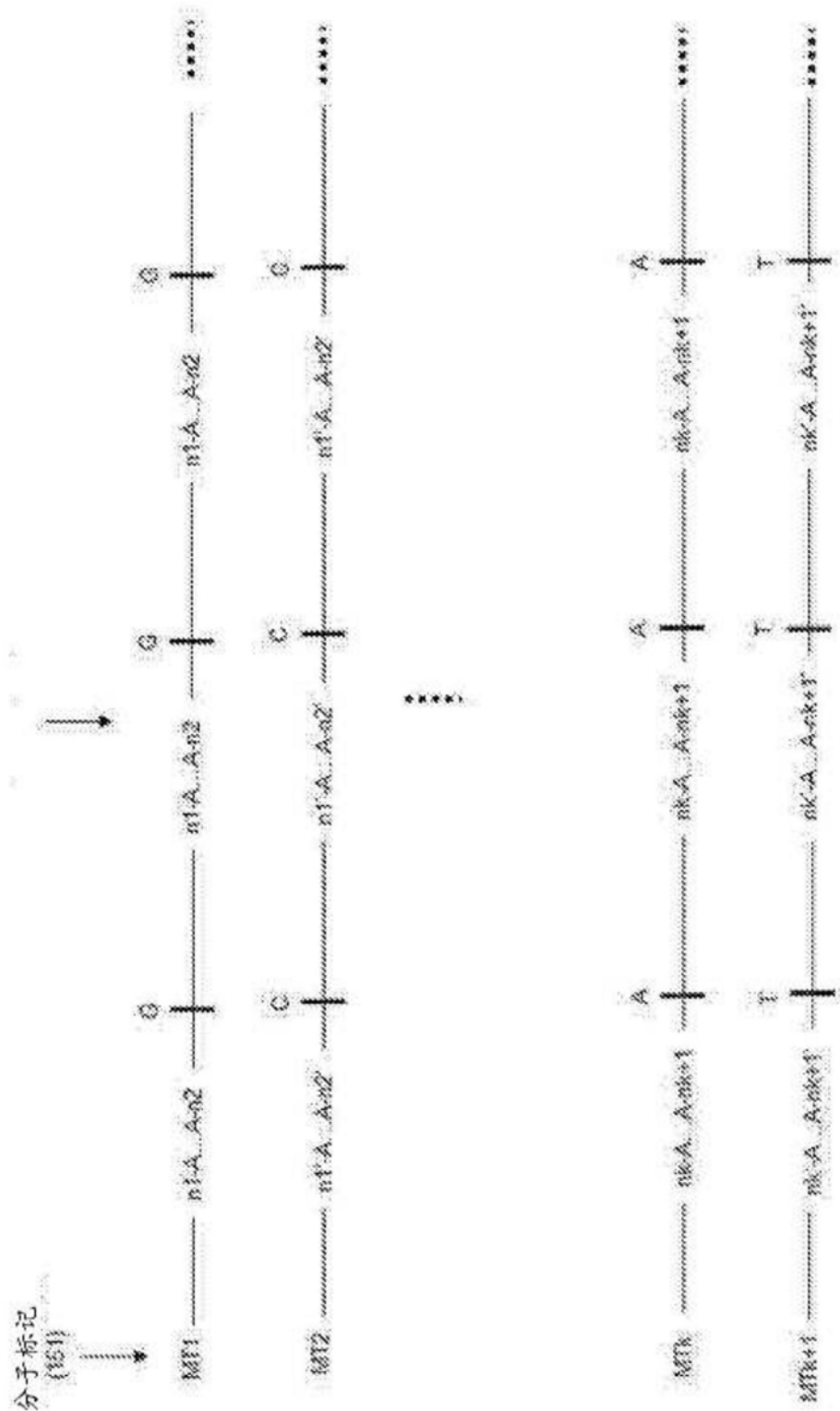


图36H

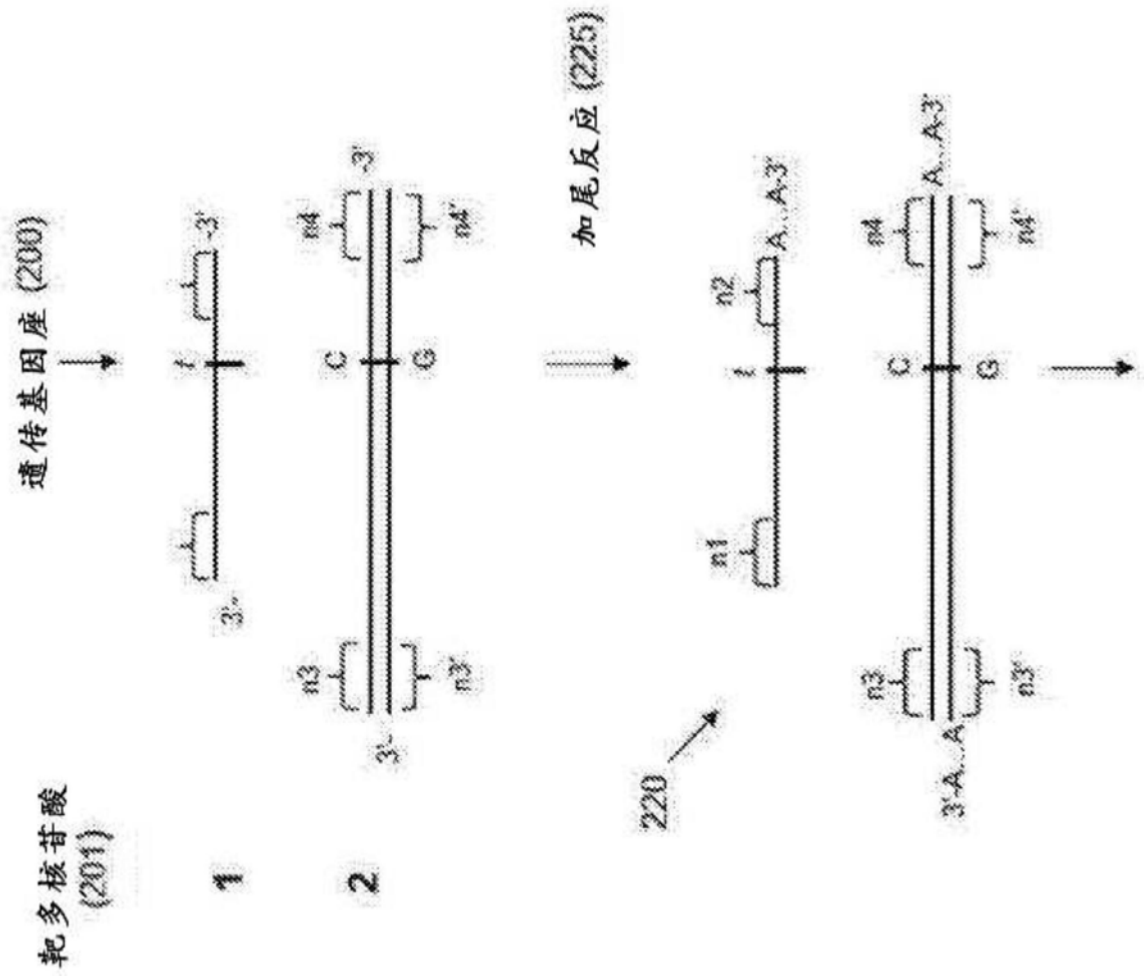


图37A

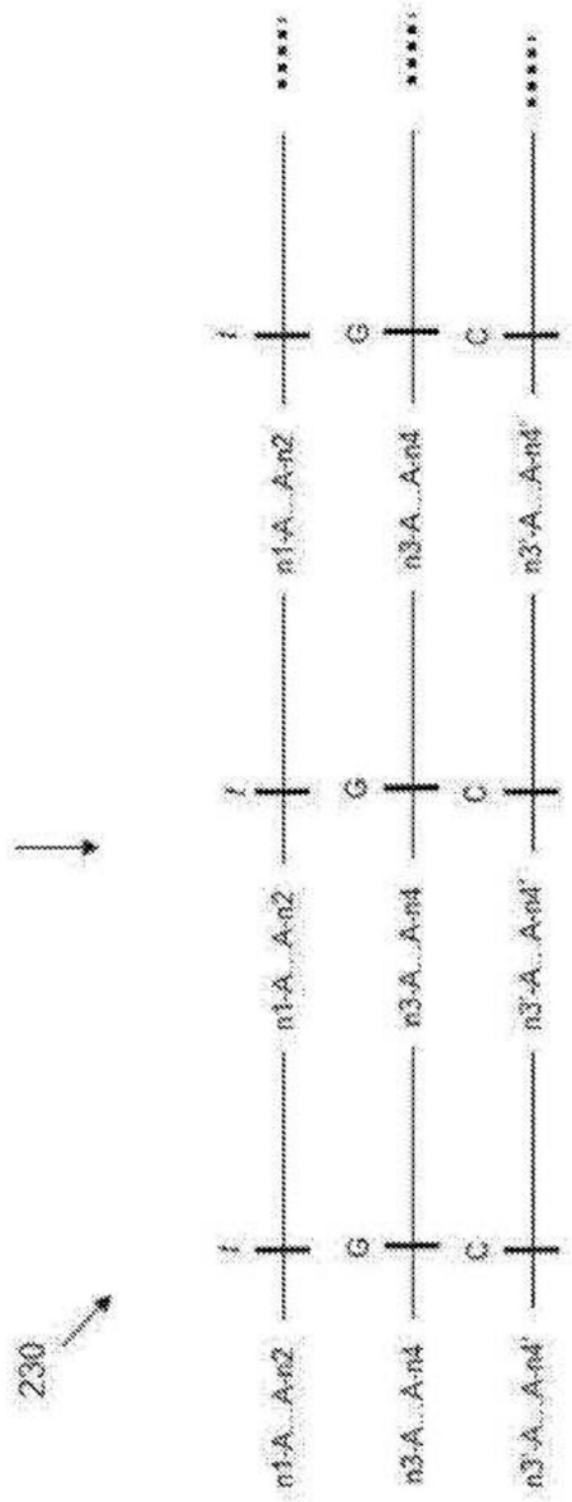


图37B

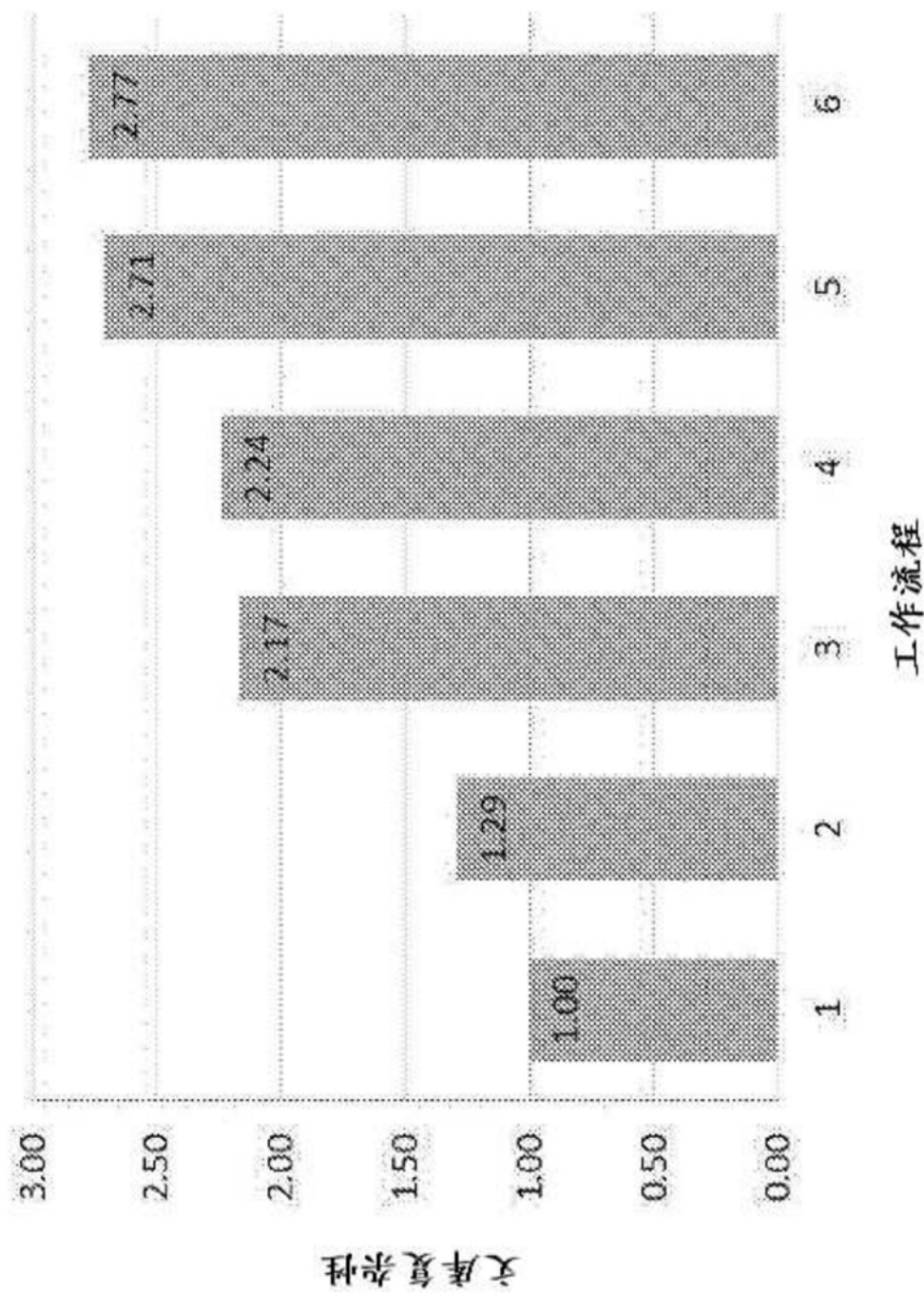


图38



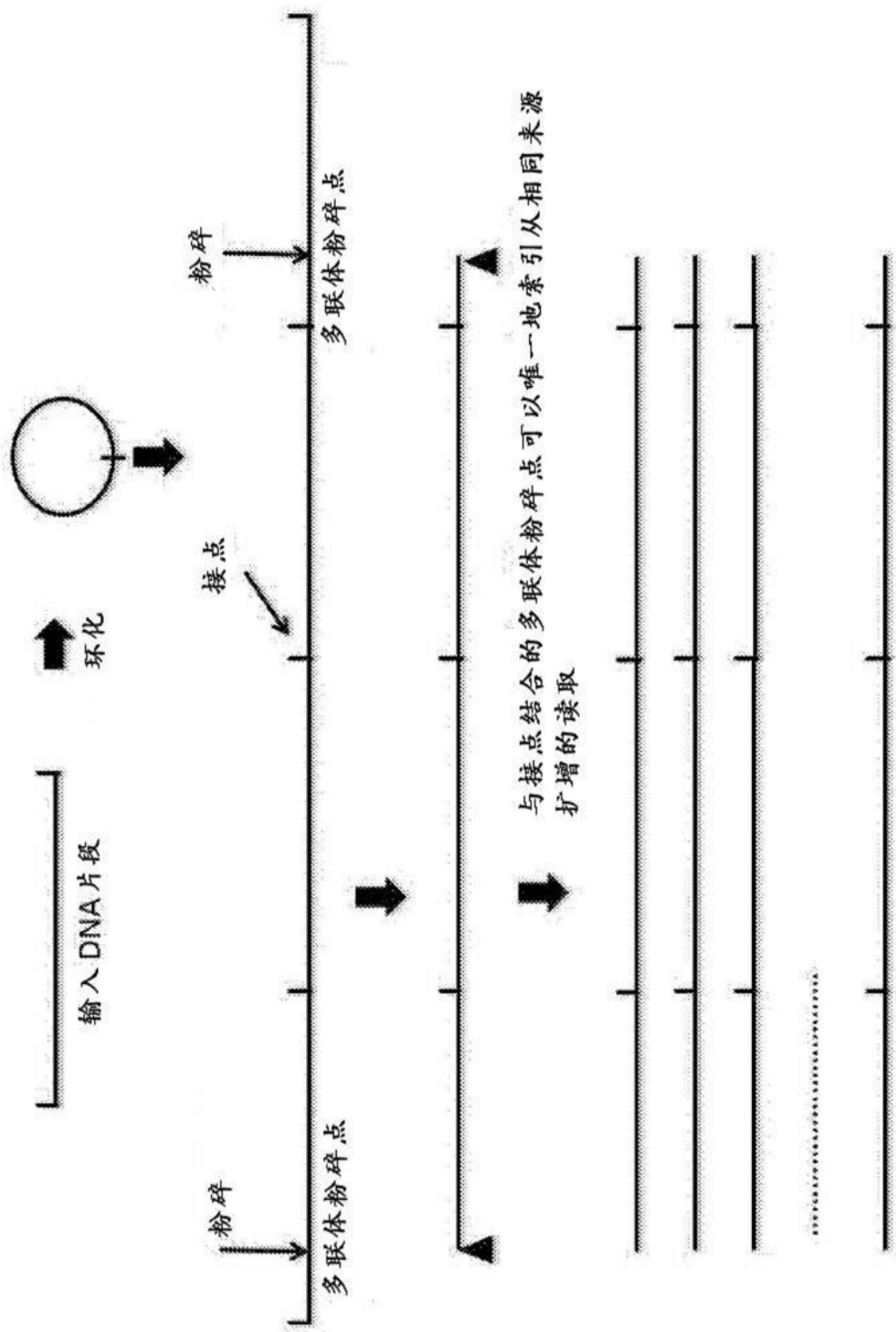


图39

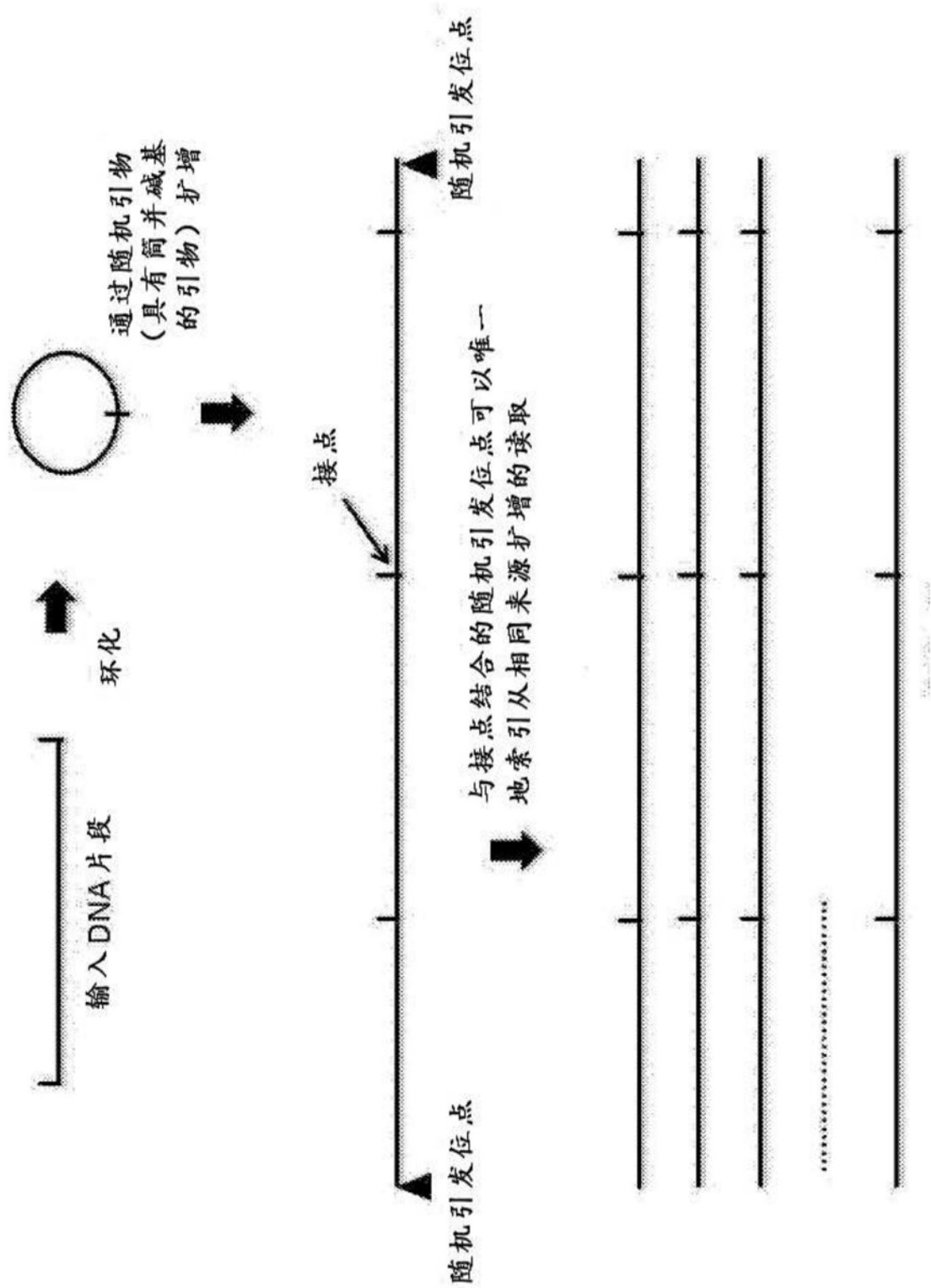


图40

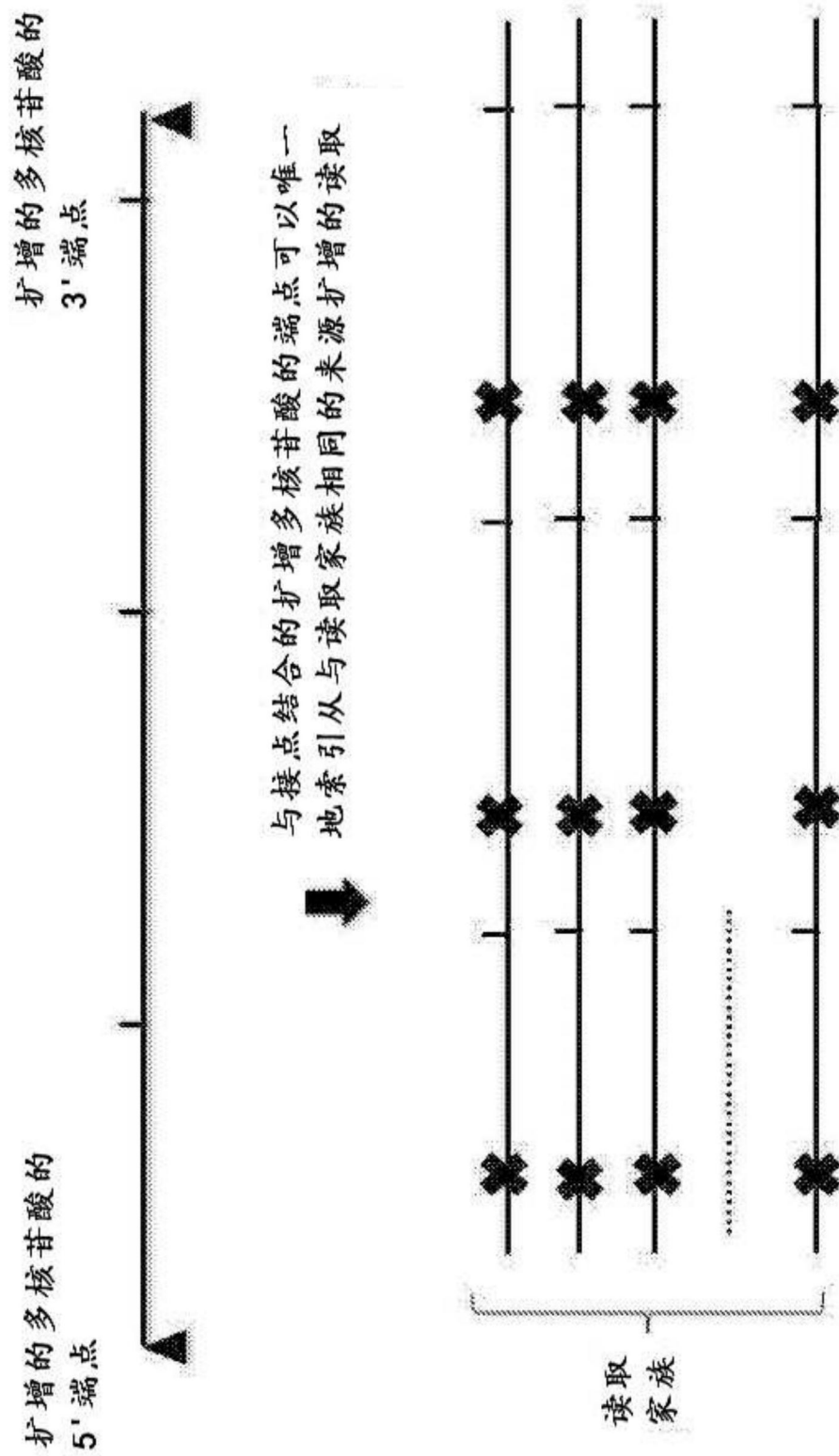


图41A

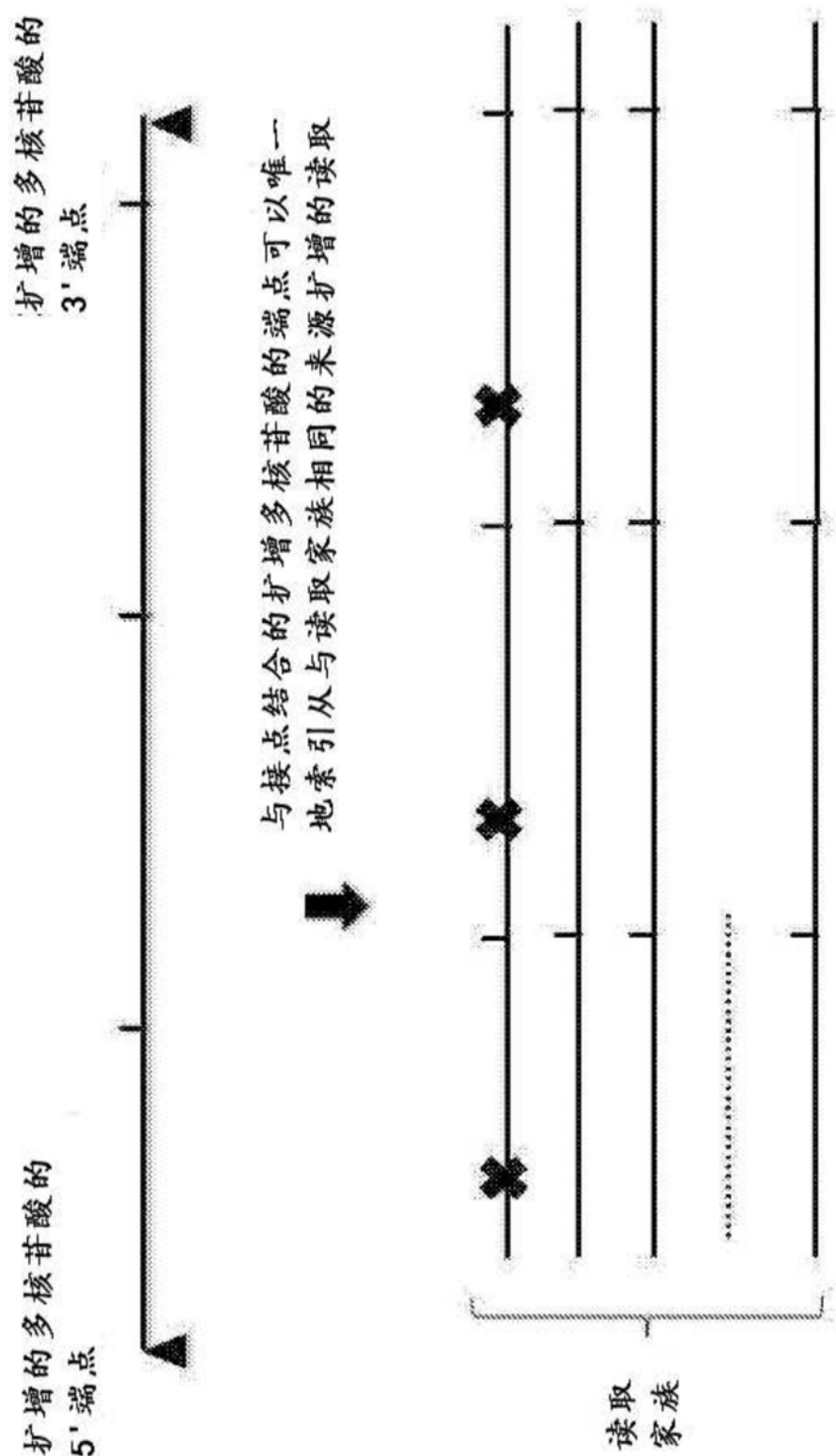


图41B

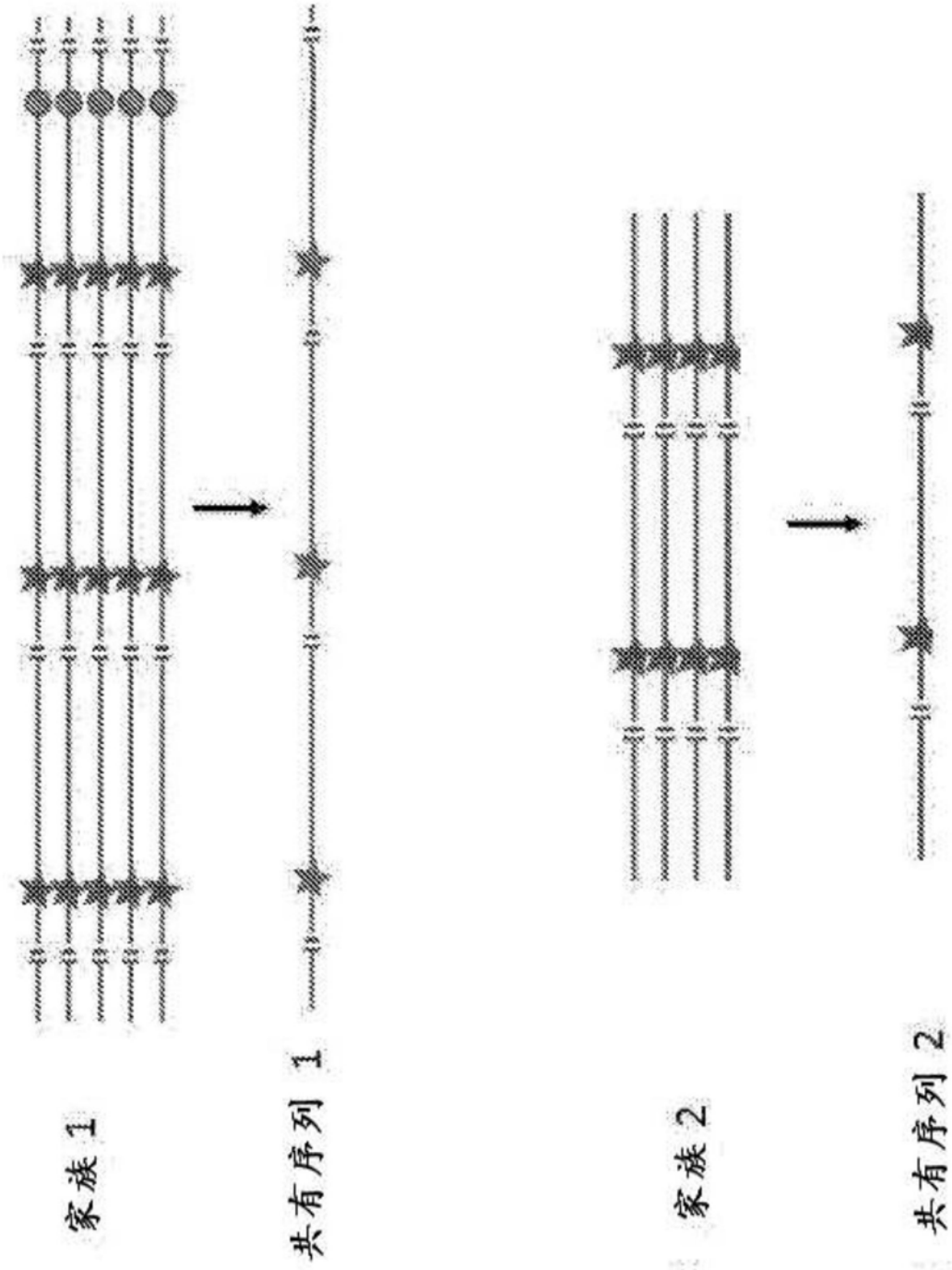


图41C