(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104736722 A
(43)申请公布日 2015.06.24

(21)申请号 201380038915.5
(22)申请日 2013.05.21
(30)优先权数据
61/649,836 2012.05.21 US
61/654,389 2012.06.01 US
61/716,378 2012.10.19 US
61/749,871 2013.01.07 US
61/763,441 2013.02.11 US
61/763,424 2013.02.11 US
(85)PCT国际申请进入国家阶段日
2015.01.21
(86)PCT国际申请的申请数据
PCT/US2013/042106 2013.05.21
(87)PCT国际申请的公布数据
WO2013/177220 EN 2013.11.28
(71)申请人 斯克利普斯研究所
地址 美国加利福尼亚州
(72)发明人 史蒂文·罗伯特·黑德
菲利普·T·欧道克哈尼安
丹尼尔·R·萨洛蒙
(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 王思琪 郑霞
(51)Int.Cl.
C12Q 1/68(2006.01)
C12P 19/34(2006.01)
G06F 19/18(2006.01)

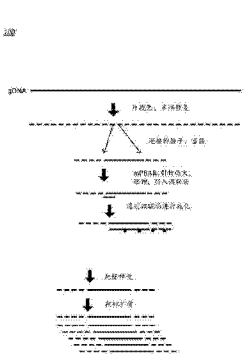
权利要求书6页 说明书53页 附图16页

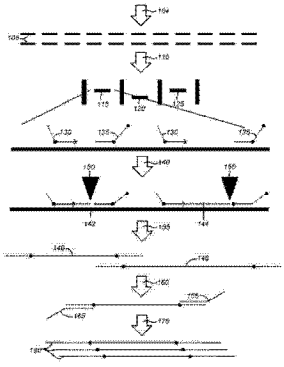
(54)发明名称

样品制备方法

(57)摘要

本发明提供了方法、组合物和用于该方法的试剂盒，其可改善核酸分析技术，并可允许更可靠且准确的靶向、多重、高通量测序。该方法、组合物和试剂盒可用于对核酸的靶基因座进行测序。本文公开的方法、组合物和试剂盒可用于辅助的从头靶向测序。本文公开的方法、组合物和试剂盒也可用于文库标记以供从头测序和定相。





CN 104736722 A

1. 一种方法，其包括：

a. 对核酸的核酸片段进行空间分离；

b. 产生一个或多个扩增子，其中该扩增子通过以下步骤产生：

i. 使引物和探针与核酸片段的共同链杂交；

ii. 进行引物延伸反应；

iii. 将引物延伸反应的产物与探针连接以形成扩增子；

c. 将一个或多个扩增子与标识符相关联；以及

d. 获得扩增子的序列。

2. 根据权利要求 1 所述的方法，进一步包括扩增所述扩增子。

3. 根据权利要求 2 所述的方法，其中所述扩增是线性的。

4. 根据权利要求 3 所述的方法，其中所述扩增通过滚环扩增而进行。

5. 根据权利要求 2 所述的方法，其中将至少一个扩增子与至少一个不同的扩增子连接。

6. 根据权利要求 2 所述的方法，其中所述扩增是非线性的。

7. 根据权利要求 1 所述的方法，其中所述标识符包含分子条形码。

8. 根据权利要求 1 所述的方法，其中所述标识符包含核酸序列。

9. 根据权利要求 1 所述的方法，其中所述标识符包含非 A、T、C 或 G 的核酸。

10. 根据权利要求 1 所述的方法，其中所述标识符位于扩增子的 5’端。

11. 根据权利要求 1 所述的方法，其中所述标识符位于扩增子的 3’端。

12. 根据权利要求 1 所述的方法，其中通过扩增子的扩增将标识符与扩增子相关联。

13. 根据权利要求 1 所述的方法，其中所述核酸选自 DNA、RNA、cDNA 和基因组 DNA。

14. 根据权利要求 1 所述的方法，其中核酸的片段化通过选自以下的方法进行：声处理、酶消化、热、暴露于紫外线、反复移液和雾化。

15. 根据权利要求 1 所述的方法，其中在分区中进行核酸片段的空间分离。

16. 根据权利要求 1 所述的方法，其中通过将核酸拴系于固体或半固体支持物上来进行核酸片段的空间分离。

17. 根据权利要求 16 所述的方法，其中所述核酸片段与拴系于固体或半固体支持物上的引物杂交，其中所述引物包含标识符。

18. 根据权利要求 16 所述的方法，其中所述核酸片段与拴系于固体或半固体支持物上的探针杂交，其中所述探针包含标识符。

19. 根据权利要求 17 所述的方法，其中对所述固体或半固体支持物进行编址。

20. 根据权利要求 1 所述的方法，其中所述探针包含标识符。

21. 根据权利要求 1 所述的方法，其中所述探针包含一个或多个衔接子序列。

22. 根据权利要求 1 所述的方法，其中所述探针与靶核酸杂交。

23. 根据权利要求 1 所述的方法，其中所述探针包含简并序列。

24. 根据权利要求 1 所述的方法，其中所述探针包含合成的核苷酸。

25. 根据权利要求 1 所述的方法，其中所述探针包含引物。

26. 根据权利要求 1 所述的方法，其中所述引物包含标识符。

27. 根据权利要求 1 所述的方法，其中所述引物包含一个或多个衔接子序列。

28. 根据权利要求 1 所述的方法，其中所述引物与靶核酸杂交。

29. 根据权利要求 1 所述的方法，其中所述引物包含简并序列。

30. 根据权利要求 1 所述的方法，其中所述引物包含合成的核苷酸。

31. 根据权利要求 1 所述的方法，其中所述扩增子与独特的标识符相关联。

32. 根据权利要求 15 所述的方法，其中所述标识符代表单独的分区。

33. 根据权利要求 1 所述的方法，其中扩增子序列的获得通过大规模平行测序而进行。

34. 根据权利要求 1 所述的方法，其中使用计算装置从包含标识符的序列读取值产生全部或部分核酸片段的共有序列。

35. 根据权利要求 1 所述的方法，其中使用计算装置从包含标识符的序列读取值产生全部或部分核酸的共有序列。

36. 根据权利要求 34 和 35 所述的方法，其中使用计算装置产生共有序列，而不将共有序列与参照进行比较。

37. 根据权利要求 34 和 35 所述的方法，其中所述共有序列具有至少 1X 的覆盖深度。

38. 根据权利要求 34 和 35 所述的方法，其中所述共有序列具有至少 5X 的覆盖深度。

39. 根据权利要求 34 和 35 所述的方法，其中所述共有序列具有至少 10X 的覆盖深度。

40. 根据权利要求 34 和 35 所述的方法，其中所述共有序列具有至少 50X 的覆盖深度。

41. 根据权利要求 1 所述的方法，其中该方法对于多个核酸片段是多重化的。

42. 根据权利要求 35 所述的方法，其中该方法对于至少 2 个核酸片段是多重化的。

43. 根据权利要求 35 所述的方法，其中该方法对于至少 10 个核酸片段是多重化的。

44. 根据权利要求 35 所述的方法，其中该方法对于至少 100 个核酸片段是多重化的。

45. 根据权利要求 35 所述的方法，其中该方法对于至少 10000 个核酸片段是多重化的。

46. 根据权利要求 35 所述的方法，其中该方法对于至少 100000 个核酸片段是多重化的。

47. 根据权利要求 35 所述的方法，其中该方法对于至少 1000000 个核酸片段是多重化的。

48. 根据权利要求 1 所述的方法，其中所述引物延伸反应的产物为至少 100 个核苷酸。

49. 根据权利要求 1 所述的方法，其中所述引物延伸反应的产物为至少 1000 个核苷酸。

50. 根据权利要求 1 所述的方法，其中所述引物延伸反应的产物为至少 10000 个核苷酸。

51. 一种方法，其包括传送通过权利要求 1 的方法产生的测序数据。

52. 一种方法，其包括接收通过权利要求 1 的方法产生的测序数据。

53. 一种方法，其包括存储通过权利要求 1 的方法产生的测序数据。

54. 一种方法，其包括比较或分析通过权利要求 1 的方法产生的测序数据。

55. 一种方法，其包括传送与通过权利要求 1 的方法产生的测序数据相关的报告。

56. 一种方法，其包括接收与通过权利要求 1 的方法产生的测序数据相关的报告。

57. 一种方法，其包括存储与通过权利要求 1 的方法产生的测序数据相关的报告。

58. 一种方法，其包括比较或分析与通过权利要求 1 的方法产生的测序数据相关的报告。

59. 根据权利要求 1 所述的方法，进一步包括 :使用包括非暂时性计算机可读介质的计

算装置将测序数据转变为与测序数据相关的报告。

60. 根据权利要求 1 所述的方法，其中所述引物或探针对于核酸片段的一个或多个区域具有特异性。

61. 根据权利要求 1 所述的方法，其中所述引物或探针与核酸片段的一个或多个区域至少 50％互补。

62. 根据权利要求 1 所述的方法，其中所述引物或探针与核酸片段的一个或多个区域至少 75％互补。

63. 根据权利要求 1 所述的方法，其中所述引物或探针与核酸片段的一个或多个区域至少 90％互补。

64. 根据权利要求 1 所述的方法，其中将一个或多个扩增子连接起来以形成连续的序列。

65. 根据权利要求 1 所述的方法，其中进行引物延伸反应包括加入链置换聚合酶。

66. 根据权利要求 1 所述的方法，进一步包括 ：进行引物延伸反应以形成引物延伸产物，其中该引物延伸产物包含亲和偶联物，并且其中该引物延伸产物包含靶序列。

67. 根据权利要求 1 所述的方法，进一步包括 ：进行引物延伸反应以形成引物延伸产物，并使用亲和偶联物进行引物延伸产物的亲和纯化。

68. 根据权利要求 66 或 67 所述的方法，其中所述亲和偶联物是生物素。

69. 根据权利要求 68 所述的方法，其中使用链霉亲和素进行引物延伸产物的亲和纯化。

70. 一种方法，其包括 ：

a. 获得核酸，其中该核酸包含靶序列 ；

b. 使 TELA 引物和 TELA 探针与核酸的共同链杂交 ；

c. 进行引物延伸反应 ；

d. 将引物延伸反应的产物与 TELA 探针连接以形成包含靶序列的连接产物 ；以及

e. 对靶序列进行测序。

71. 根据权利要求 70 所述的方法，其中所述靶序列为连接产物的至少 30％。

72. 一种方法，其包括传送通过权利要求 70 的方法产生的测序数据。

73. 一种方法，其包括接收通过权利要求 70 的方法产生的测序数据。

74. 一种方法，其包括存储通过权利要求 70 的方法产生的测序数据。

75. 一种方法，其包括比较或分析通过权利要求 70 的方法产生的测序数据。

76. 一种方法，其包括传送与通过权利要求 70 的方法产生的测序数据相关的报告。

77. 一种方法，其包括接收与通过权利要求 70 的方法产生的测序数据相关的报告。

78. 一种方法，其包括存储与通过权利要求 70 的方法产生的测序数据相关的报告。

79. 一种方法，其包括比较或分析与通过权利要求 70 的方法产生的测序数据相关的报告。

80. 根据权利要求 70 所述的方法，进一步包括 ：使用包括非暂时性计算机可读介质的计算装置将测序数据转变为与测序数据相关的报告。

81. 根据权利要求 70 所述的方法，其中将一个或多个连接产物连接起来以形成连续的序列。

82. 一种方法，其包括：

a. 获得核酸文库；

b. 将衔接子序列与该核酸文库的一个或多个核酸连接；

c. 使引物与衔接子序列杂交，其中该引物包含间隔区和基因座特异性区域；

d. 进行引物延伸反应以形成引物延伸产物，其中该引物延伸产物包含亲和偶联物，并且其中该引物延伸产物包含靶序列；

e. 使用该亲和偶联物进行引物延伸产物的亲和纯化。

83. 根据权利要求 82 所述的方法，其中所述核酸文库为片段化的基因组 DNA。

84. 根据权利要求 82 所述的方法，其中所述核酸文库为表达的序列。

85. 根据权利要求 82 所述的方法，其中所述核酸文库经表观遗传学分选。

86. 根据权利要求 82 所述的方法，其中对引物延伸产物进行测序。

87. 根据权利要求 82 所述的方法，其中所述核酸文库包含至少 2 个核酸片段。

88. 根据权利要求 82 所述的方法，其中所述核酸文库包含至少 10 个核酸片段。

89. 根据权利要求 82 所述的方法，其中所述核酸文库包含至少 100 个核酸片段。

90. 根据权利要求 82 所述的方法，其中所述核酸文库包含至少 10,000 个核酸片段。

91. 根据权利要求 82 所述的方法，其中所述核酸文库包含至少 100,000 个核酸片段。

92. 根据权利要求 82 所述的方法，其中所述核酸文库包含至少 1,000,000 个核酸片段。

93. 根据权利要求 82 所述的方法，其中所述核酸文库已进行片段化。

94. 根据权利要求 93 所述的方法，其中所述片段化通过选自以下的方法进行：声处理、酶消化、热、暴露于紫外线、反复移液和雾化。

95. 根据权利要求 82 所述的方法，进一步包括对核酸进行扩增以产生核酸文库。

96. 根据权利要求 82 所述的方法，进一步包括对衔接子连接的核酸文库进行扩增。

97. 根据权利要求 95 或 96 所述的方法，其中所述扩增是线性的。

98. 根据权利要求 97 所述的方法，其中所述扩增通过滚环扩增而进行。

99. 根据权利要求 96 所述的方法，其中所述扩增是非线性的。

100. 根据权利要求 82 所述的方法，其中所述引物延伸产物包含标识符。

101. 根据权利要求 82 所述的方法，其中所述引物延伸产物包含分子条形码。

102. 根据权利要求 82 所述的方法，其中所述引物延伸产物包含核酸序列。

103. 根据权利要求 82 所述的方法，其中所述引物延伸产物包含非 A、T、C 或 G 的核酸。

104. 根据权利要求 82 所述的方法，其中所述衔接子位于核酸文库的核酸的 5’端。

105. 根据权利要求 82 所述的方法，其中所述引物延伸产物位于核酸文库的核酸的 3’端。

106. 根据权利要求 82 所述的方法，其中所述核酸选自 DNA、RNA、cDNA 和基因组 DNA。

107. 根据权利要求 82 所述的方法，其中所述间隔区为简并或随机序列。

108. 根据权利要求 82 所述的方法，其中所述间隔区包含至少 1 个核苷酸。

109. 根据权利要求 82 所述的方法，其中所述间隔区包含至少 10 个核苷酸。

110. 根据权利要求 82 所述的方法，其中所述间隔区包含至少 100 个核苷酸。

111. 根据权利要求 82 所述的方法，其中所述间隔区包含分子条形码。

112. 根据权利要求 82 所述的方法，其中所述间隔区包含核酸序列。

113. 根据权利要求 82 所述的方法，其中所述间隔区包含非 A、T、C 或 G 的核酸。

114. 根据权利要求 82 所述的方法，其中所述间隔区包含酶的靶序列。

115. 根据权利要求 82 所述的方法，其中所述基因座特异性区域与核酸的基因座至少 50％互补。

116. 根据权利要求 82 所述的方法，其中所述基因座特异性区域与核酸的基因座至少 70％互补。

117. 根据权利要求 82 所述的方法，其中所述基因座特异性区域与核酸的基因座至少 80％互补。

118. 根据权利要求 82 所述的方法，其中所述基因座特异性区域与核酸的基因座至少 90％互补。

119. 根据权利要求 82 所述的方法，其中所述基因座特异性区域与核酸的基因座至少 99％互补。

120. 根据权利要求 82 所述的方法，其中所述基因座特异性区域结合靶序列上游的核酸的基因座。

121. 根据权利要求 82 所述的方法，其中所述亲和偶联物是生物素。

122. 根据权利要求 82 所述的方法，其中使用链霉亲和素进行引物延伸产物的亲和纯化。

123. 根据权利要求 82 所述的方法，进一步包括对引物延伸产物进行测序。

124. 根据权利要求 100 所述的方法，其中所述标识符代表产生核酸文库的样品的来源。

125. 根据权利要求 123 所述的方法，其中通过进行大规模平行测序来获得引物延伸产物的测序。

126. 根据权利要求 100 所述的方法，其中使用计算装置从包含标识符的序列读取值产生全部或部分靶序列的共有序列。

127. 根据权利要求 100 所述的方法，其中使用计算装置从包含标识符的序列读取值产生全部或部分靶序列的共有序列。

128. 根据权利要求 100 所述的方法，其中使用计算装置产生共有序列，而不将共有序列与参照进行比较。

129. 根据权利要求 126-128 所述的方法，其中所述共有序列具有至少 1X 的覆盖深度。

130. 根据权利要求 126-128 所述的方法，其中所述共有序列具有至少 5X 的覆盖深度。

131. 根据权利要求 126-128 所述的方法，其中所述共有序列具有至少 10X 的覆盖深度。

132. 根据权利要求 126-128 所述的方法，其中所述共有序列具有至少 50X 的覆盖深度。

133. 根据权利要求 82 所述的方法，其中该方法对于多个样品或靶序列是多重化的。

134. 根据权利要求 82 所述的方法，其中该方法对于至少 2 个样品或靶序列是多重化的。

135. 根据权利要求 82 所述的方法，其中该方法对于至少 10 个样品或靶序列是多重化的。

136. 根据权利要求 82 所述的方法，其中该方法对于至少 100 个样品或靶序列是多重化的。

137. 根据权利要求 82 所述的方法，其中该方法对于至少 10000 个样品或靶序列是多重化的。

138. 根据权利要求 82 所述的方法，其中该方法对于至少 100000 个样品或靶序列是多重化的。

139. 根据权利要求 82 所述的方法，其中该方法对于至少 1000000 个样品或靶序列是多重化的。

140. 根据权利要求 82 所述的方法，其中所述引物延伸反应的产物为至少 100 个核苷酸。

141. 根据权利要求 82 所述的方法，其中所述引物延伸反应的产物为至少 1000 个核苷酸。

142. 根据权利要求 82 所述的方法，其中所述引物延伸反应的产物为至少 10000 个核苷酸。

143. 一种方法，其包括传送通过权利要求 82 的方法产生的测序数据。

144. 一种方法，其包括接收通过权利要求 82 的方法产生的测序数据。

145. 一种方法，其包括存储通过权利要求 82 的方法产生的测序数据。

146. 一种方法，其包括比较或分析通过权利要求 82 的方法产生的测序数据。

147. 一种方法，其包括传送与通过权利要求 82 的方法产生的测序数据相关的报告。

148. 一种方法，其包括接收与通过权利要求 82 的方法产生的测序数据相关的报告。

149. 一种方法，其包括存储与通过权利要求 82 的方法产生的测序数据相关的报告。

150. 一种方法，其包括比较或分析与通过权利要求 82 的方法产生的测序数据相关的报告。

151. 根据权利要求 82 所述的方法，进一步包括 ：使用包括非暂时性计算机可读介质的计算装置将测序数据转变为与测序数据相关的报告。

152. 根据权利要求 82 所述的方法，其中将一个或多个引物延伸产物连接起来以形成连续的序列。

样品制备方法

相关申请的交叉引用

[0001] 本申请是以下申请的非临时申请并要求以下申请的权益：于 2013 年 2 月 11 日提交的美国临时专利申请号 61/763,441（代理人案号 44013-703.101）、于 2013 年 2 月 11 日提交的美国临时专利申请号 61/763,424（代理人案号 44013-701.105）、于 2013 年 1 月 7 日提交的美国临时专利申请号 61/749,871（代理人案号 44013-701.104）、于 2012 年 10 月 19 日提交的美国临时专利申请号 61/716,378（代理人案号 44013-701.103）、于 2012 年 6 月 1 日提交的美国临时专利申请号 61/654,389（代理人案号 44013-701.102）以及于 2012 年 5 月 21 日提交的美国临时专利申请号 61/649,836（代理人案号 44013-701.101），以上各申请的全部内容通过引用并入本文。

背景技术

[0002] 若干生物应用涉及核酸测序，包括下一代测序。下一代测序可放大克隆错误。此外，下一代测序的数据分析可能要求使用参照基因组。需要靶向基因组的特定区域以供测序分析的方法。

[0003] 当前的下一代测序平台在针对较长序列的测序仪读取值比对和基因组组装方面存在问题。重复序列、同源序列和可变序列的区域并没有被可靠地定位。这些短读取测序仪需要与参照进行比对的策略，且这一比对步骤可大幅度地增加针对较长的读取长度获得可靠的测序结果所需的计算步骤，以及偏差。在本领域中需要改进的方法和系统，其用于靶向目标序列并准备这些序列以用于较长读取长度的测序反应。

发明内容

[0004] 在一些实施方案中，本发明提供了一种方法，其包括：对核酸的核酸片段进行空间分离；产生一个或多个扩增子，其中该扩增子通过以下步骤产生：使引物和探针与核酸片段的共同链杂交，进行引物延伸反应，将引物延伸反应的产物与探针连接以形成扩增子；将一个或多个扩增子与标识符相关联；以及获得扩增子的序列。

[0005] 在一些实施方案中，可对扩增子进行扩增。扩增子的扩增可通过线性扩增、非线性扩增或滚环扩增而进行。

[0006] 在一些实施方案中，将至少一个扩增子与至少一个不同的扩增子连接。在一些情况下，标识符可包含分子条形码，核酸序列，非 A、T、C 或 G 的核酸。在一些情况下，标识符位于扩增子的 5’端。在一些情况下，标识符位于扩增子的 3’端。在一些情况下，通过扩增子的扩增将标识符与扩增子相关联。

[0007] 在一些实施方案中，核酸选自 DNA、RNA、cDNA 和基因组 DNA。在一些情况下，通过选自以下的方法对核酸进行片段化：声处理、酶消化、热、暴露于紫外线、反复移液和雾化。

[0008] 在一些实施方案中，在分区中进行核酸片段的空间分离。在一些情况下，通过将核酸拴系于固体或半固体支持物上来进行核酸片段的空间分离。在一些情况下，使核酸片段与拴系于固体或半固体支持物上的引物杂交，其中该引物包含标识符。在一些情况下，使核

酸片段与拴系于固体或半固体支持物上的探针杂交，其中该探针包含标识符。在一些情况下，对固体或半固体支持物进行编址 (address)。

[0009] 在一些实施方案中，探针包含标识符。在一些情况下，探针包含一个或多个衔接子序列。在一些情况下，探针与靶核酸杂交。在一些情况下，探针包含简并序列。在一些情况下，探针包含合成的核苷酸。在一些情况下，探针包含引物。

[0010] 在一些实施方案中，引物包含标识符。在一些情况下，引物包含一个或多个衔接子序列。在一些情况下，引物与靶核酸杂交。在一些情况下，引物包含简并序列。在一些情况下，引物包含合成的核苷酸。在一些情况下，将扩增子与独特的标识符相关联。在一些情况下，标识符代表单独的分区。

[0011] 在一些实施方案中，扩增子的测序通过大规模平行测序而进行。在一些情况下，使用计算装置从包含标识符的序列读取值产生全部或部分核酸片段的共有序列。

[0012] 在一些实施方案中，使用计算装置从包含标识符的序列读取值产生全部或部分核酸的共有序列。在一些情况下，使用计算装置产生共有序列，而不将共有序列与参照进行比较。

[0013] 在一些实施方案中，共有序列具有至少 1X、5X、10X 或 50X 的覆盖深度。

[0014] 在一些实施方案中，本发明的方法提供了对于多个核酸片段的多重化 (multiplexed) 分析。在一些情况下，该方法对于至少 2 个核酸片段、至少 10 个核酸片段、至少 100 个核酸片段、至少 10000 个核酸片段、至少 100000 个核酸片段或至少 1000000 个核酸片段是多重化的。

[0015] 在一些实施方案中，引物延伸反应的产物为至少 100 个核苷酸、至少 1000 个核苷酸或至少 10000 个核苷酸。

[0016] 在一些实施方案中，本发明提供了传送产生的测序数据、接收产生的测序数据、存储产生的测序数据，包括比较或分析产生的测序数据，传送与产生的测序数据相关的报告，接收与产生的测序数据相关的报告，存储与产生的测序数据相关的报告，存储与产生的测序数据相关的报告，比较或分析与通过本发明的方法产生的测序数据相关的报告。在一些情况下，本发明提供了使用包括非暂时性计算机可读介质的计算装置将测序数据转变为与测序数据相关的报告。

[0017] 在一些实施方案中，引物或探针对于核酸片段的一个或多个区域具有特异性。在一些情况下，引物或探针与核酸片段的一个或多个区域至少 50％互补。在一些情况下，引物或探针与核酸片段的一个或多个区域至少 75％互补。在一些情况下，引物或探针与核酸片段的一个或多个区域至少 90％互补。

[0018] 在一些实施方案中，将一个或多个扩增子连接起来以形成连续的序列。在一些情况下，本发明提供了进行引物延伸反应，所述反应包括加入链置换聚合酶。在一些情况下，本发明提供了进行引物延伸反应以形成引物延伸产物，其中该引物延伸产物包含亲和偶联物，并且其中该引物延伸产物包含靶序列。

[0019] 在一些实施方案中，本发明提供了进行引物延伸反应以形成引物延伸产物，并使用亲和偶联物进行引物延伸产物的亲和纯化。在一些情况下，该亲和偶联物是生物素。在一些情况下，使用链霉亲和素进行引物延伸产物的亲和纯化。

[0020] 在一些实施方案中，本发明提供了一种方法，其包括 :获得核酸，其中该核酸包含

靶序列；使 TELA 引物和 TELA 探针与核酸的共同链杂交；进行引物延伸反应；将引物延伸反应的产物与 TELA 探针连接以形成包含靶序列的连接产物；以及对靶序列进行测序。

[0021] 在一些实施方案中，靶序列为连接产物的至少 30％。在一些情况下，将一个或多个连接产物连接起来以形成连续的序列。

[0022] 在一些实施方案中，本发明提供了一种方法，其包括：获得核酸文库；将衔接子序列与该核酸文库的一个或多个核酸连接；使引物与衔接子序列杂交，其中该引物包含间隔区和基因座特异性区域；进行引物延伸反应以形成引物延伸产物，其中该引物延伸产物包含亲和偶联物，并且其中该引物延伸产物包含靶序列；使用亲和偶联物进行引物延伸产物的亲和纯化。

[0023] 在一些实施方案中，核酸文库为片段化的 gDNA。在一些情况下，核酸文库为表达的序列。在一些情况下，核酸文库经表观遗传学分选 (sort)。在一些实施方案中，对引物延伸产物进行测序。

[0024] 在一些情况下，核酸文库包含至少 2 个、至少 10 个、至少 100 个、至少 10000 个、至少 100000 个或至少 1000000 个核酸片段。

[0025] 在一些情况下，通过选自以下的方法对核酸进行片段化：声处理、酶消化、热、暴露于紫外线、反复移液和雾化。权利要求 1 的方法进一步包括对核酸进行扩增以产生核酸文库。

[0026] 在一些实施方案中，本发明提供了包括对衔接子连接的核酸文库进行扩增的方法。在一些情况下，扩增是线性的。在一些情况下，扩增通过滚环扩增而进行。在一些情况下，扩增是非线性的。在一些情况下，引物延伸产物包含标识符。

[0027] 在一些实施方案中，引物延伸产物包含分子条形码。在一些情况下，引物延伸产物包含核酸序列。在一些情况下，引物延伸产物包含非 A、T、C 或 G 的核酸。在一些情况下，衔接子位于核酸文库的核酸的 5’端。在一些情况下，引物延伸产物位于核酸文库的核酸的 3’端。在一些情况下，核酸选自 DNA、RNA、cDNA 和基因组 DNA。

[0028] 在一些实施方案中，本发明提供了其中间隔区为简并或随机序列的方法。在一些情况下，间隔区包含至少 1 个核苷酸、至少 10 个核苷酸或 100 个核苷酸。

[0029] 在一些情况下，间隔区包含分子条形码。在一些情况下，间隔区包含核酸序列。在一些情况下，间隔区包含非 A、T、C 或 G 的核酸。在一些情况下，间隔区包含酶的靶序列。

[0030] 在一些实施方案中，基因座特异性区域与核酸的基因座至少 50％互补。在一些情况下，基因座特异性区域与核酸的基因座至少 70％互补。在一些情况下，基因座特异性区域与核酸的基因座至少 80％互补。在一些情况下，基因座特异性区域与核酸的基因座至少 90％互补。在一些情况下，基因座特异性区域与核酸的基因座至少 99％互补。在一些情况下，基因座特异性区域结合靶序列上游的核酸的基因座。

[0031] 在一些实施方案中，亲和偶联物是生物素。在一些情况下，使用链霉亲和素进行引物延伸产物的亲和纯化。

[0032] 在一些实施方案中，本发明提供了对引物延伸产物进行测序。在一些情况下，标识符代表产生核酸文库的样品的来源。

[0033] 在一些情况下，通过进行大规模平行测序来获得引物延伸反应的序列。在一些情况下，使用计算装置从包含标识符的序列读取值产生全部或部分靶序列的共有序列。

[0034] 在一些实施方案中，共有序列具有至少 1X、5X、10X 或 50X 的覆盖深度。

[0035] 在一些实施方案中，本发明提供了对于多个样品或靶序列为多重化的方法。在一些情况下，该方法对于至少 2 个、10 个、100 个、10000 个、100000 个、1000000 个、1000000 个样品或靶序列是多重化的。

[0036] 在一些实施方案中，引物延伸反应的产物为至少 100 个核苷酸、至少 1000 个核苷酸或至少 10000 个核苷酸。

[0037] 在一些实施方案中，将一个或多个引物延伸产物连接起来以形成连续的序列。

援引并入

[0038] 本说明书中提及的所有出版物、专利和专利申请均通过引用并入本文，其程度如同具体地且个别地指明每个单独的出版物、专利或专利申请而并入。

附图说明

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

[0040] 图 1A 示出了涉及 mPEAR 的工作流程的示意性图示。

[0041] 图 1B 示出了涉及靶向测序的工作流程的示意性图示。

[0042] 图 2A 示出了供体引物 (D) 的实例。

[0043] 图 2B 示出了受体探针 (A) 的实例。

[0044] 图 3A 示出了普通 TELA 或 mPEAR 延伸反应的图示。

[0045] 图 3B 示出了 TELA 或 mPEAR 标记的产物的亲和纯化反应的图示。

[0046] 图 3C 示出了 TELA 或 mPEAR 标记的产物从固定的表面上释放的图示。

[0047] 图 4 示出了 RAPEL 方法的图示。

[0048] 图 5 示出了测序仪平台衔接子与靶向测序产物一起使用的图示。

[0049] 图 6 示出了侧翼为衔接子序列和分子标记或条形码的多个靶序列。

[0050] 图 7 示出了 mPEAR 或 TELA 引物和靶序列。

[0051] 图 8 示出了用于靶序列的从头测序的衔接子序列和条形码的附接。

[0052] 图 9 示出了从 mPEAR、TELA、RAPEL 或本发明的方法和组合物的其他产物产生的测序的示例性读取结构。

[0053] 图 10 示出了在 mPEAR、TELA、RAPEL 或本发明的方法和组合物的其他产物内的多个衔接子序列、条形码和引物位点。

[0054] 图 11 示出了用于传送通过本发明的组合物和方法产生的数据的计算机可读存储介质平台和示例性的装置。

[0055] 图 12 示出了使用滚环扩增组装多个连续序列靶标的本发明方法和组合物的示意性图示。

[0056] 图 13 示出了靶序列的物理分离以及基于条形码序列从头组装序列的示意图。

[0057] 图 14 示出了序列的实例，包括条形码和衔接子序列。

具体实施方式

[0058] 本发明提供了方法、组合物和用于该方法的试剂盒，其可改善核酸分析技术，并可允许更可靠且准确的靶向、多重、高通量测序。该方法、组合物和试剂盒可用于对核酸的靶基因座进行测序。本文公开的方法、组合物和试剂盒可用于辅助的从头靶向测序。本文公开的方法、组合物和试剂盒也可用于文库 DNA/RNA 标记，以用于真正的从头测序和定相 (phasing)。

I. 定义

[0059] 如本文所述的“亲和偶联物”提供了利用两个分子之间的特异性相互作用来进行靶分子的纯化。对靶分子具有亲和性的亲和配体可附接于不溶性支持物并充当捕获靶分子的诱饵。靶分子可以共价或非共价地连接至与亲和配体相互作用或结合的偶联物分子。亲和配体可以是结合靶标而不结合溶液中的其他分子的任何分子。

[0060] “扩增的核酸”或“扩增的多核苷酸”可以是通过在体外进行的任何核酸扩增或复制方法使其量与其起始量相比增加至至少两倍的任何核酸或多核苷酸分子。例如，扩增的核酸可从聚合酶链反应 (PCR) 获得，在一些情况下聚合酶链反应可以以指数方式 (例如，2^n) 扩增 DNA。扩增的核酸也可从线性扩增获得。扩增的核酸可通过引物延长而获得。Identify。

[0061] “扩增产物”可指从扩增反应获得的产物。

[0062] “扩增子”可以是多核苷酸或核酸，其为天然或人工扩增或复制事件的来源和 / 或产物。

[0063] 术语“生物样品”或“样品”通常是指从生物实体中分离的样品或部分。生物样品可表现出整体的性质，且实例包括但不限于体液、分离的肿瘤样本、培养的细胞、及其任意组合。生物样品可来自于一个或多个个体。一个或多个生物样品可来自于同一个个体。一个非限制性实例为，一个样品来自于个体的血液而第二个样品来自于个体的肿瘤活检。生物样品的实例可包括但不限于血液、血清、血浆、鼻拭取物或鼻咽洗液、唾液、尿液、胃液、脊髓液、泪液、粪便、粘液、汗液、耳垢、油、腺体分泌物、脑脊髓液、组织、精液、阴道液、间质液 (包括来源于肿瘤组织的间质液)、眼部液体、脊髓液、咽喉拭取物、呼气、毛发、指甲、皮肤、活检、胎盘液、羊水、脐带血、emphatic fluid、腔液、痰、脓、微生物群 (micropiota)、胎便、乳汁和 / 或其他分泌物。样品可包括鼻咽洗液。受试者的组织样品的实例可包括但不限于结缔组织、肌肉组织、神经组织、上皮组织、软骨、癌性或肿瘤样品或骨骼。样品可从人或动物提供。样品可从哺乳动物、脊椎动物如鼠类、猿猴、人、家畜、竞技动物 (sport animal) 或宠物提供。样品可从活的或死的受试者中采集。样品可从受试者中新鲜采集或者可能已经经过了某种形式的前处理、储存或运输。

[0064] “体液”通常可描述来源于受试者身体的液体或分泌物。在一些情况下，体液可以是混合在一起的多于一种类型的体液的混合物。体液的一些非限制性实例可以是 :血液、尿液、骨髓、脊髓液、胸膜液、淋巴液、羊水、腹水、痰或其组合。

[0065] “互补的”或“互补性”可指通过碱基配对而相关的核酸分子。互补的核苷酸通常为 A 与 T(或 A 与 U)，或者 C 与 G。两个单链 RNA 或 DNA 分子在以下情况下被称为基本上互补的 :一条链的核苷酸最佳地对准并具有适当的核苷酸插入或缺失，以至少约 90％至约 95％的互补性，并更优选以约 98％至约 100％的互补性，且甚至更优选以 100％的互补性配对。或者，当 RNA 或 DNA 链在选择性杂交条件下与其互补链杂交时存在基本互补性。选择

性杂交条件包括但不限于严格的杂交条件。杂交温度通常比解链温度（T_m）低至少约 2℃至约 6℃。

[0066] “条形码”或“分子条形码”可以是用于标记的物质。条形码可标记分子，例如核酸或多肽。用于标记的物质可与信息相关联。条形码可被称为序列标识符（即，基于序列的条形码或序列索引）。条形码可以是特定的核苷酸序列。条形码可用作标识符。条形码可以是不同大小的分子或相同分子的不同终点。条形码可包括分子内的特定序列和不同的终止序列 (ending sequence)。例如，从相同引物扩增且具有 25 个核苷酸位置的分子不同于经扩增且具有 27 个核苷酸位置的分子。在 27mer 序列中的额外位置可被认为是条形码。条形码可引入多核苷酸中。可通过许多方法将条形码引入多核苷酸中。用于引入条形码的一些非限制性方法可包括分子生物学方法。用于引入条形码的分子生物学方法的一些非限制性实例为通过引物（例如，加尾的引物延长）、探针（即，连接至探针的延长）或连接（即，已知序列与分子的连接）。

[0067] 条形码可引入多核苷酸的任何区域中。该区域可以是已知的。该区域可以是未知的。条形码可添加至沿多核苷酸的任何位置。条形码可添加至多核苷酸的 5’端。条形码可添加至多核苷酸的 3’端。条形码可添加至多核苷酸的 5’端和 3’端之间。条形码可与一个或多个其他的已知序列一起添加。一个非限制性实例是条形码与序列衔接子一起添加。

[0068] 条形码可与信息相关联。可与条形码相关联的信息类型的一些非限制性实例包括：样品的来源；样品的取向；在其中处理样品的区域或容器；相邻的多核苷酸；或其任意组合。

[0069] 在一些情况下，条形码可由序列的组合制成（不同于组合条形码化），并可用于鉴定样品或基因组坐标，以及不同的模板分子或单链（由其获得分子标记和链的拷贝）。在一些情况下，可将样品标识符、基因组坐标和每个生物分子的特异性标记一起扩增。

[0070] 条形码可在样品合并之前加入。当测定所合并的样品的序列时，可将条形码与多核苷酸的其余部分一起进行测序。条形码可用于将已测序的片段与样品的来源相关联。

[0071] 条形码也可用于鉴定样品的取向。一个或多个条形码可一起使用。两个或更多个条形码可彼此相邻、彼此不相邻或其任意组合。

[0072] 条形码可用于组合标记。

[0073] “组合标记”可为这样一种方法，通过该方法使用两个或更多个条形码进行标记。这两个或更多个条形码可标记多核苷酸。每个单独的条形码本身均可与信息相关联。条形码组合在一起可与信息相关联。在一些情况下，条形码的组合可一起使用，以确定在随机扩增的分子中扩增发生于原始样品模板而非该模板的合成拷贝。

[0074] “简并”可指由随机碱基组成的核酸或核酸区域。当指核酸序列（例如，“简并引物”或“随机引物”或“简并探针”或“随机探针”）时，术语“简并”和“随机”可互换使用。简并区可以是可变长度的。简并区可包含整个核酸的某部分（例如，半简并引物）。简并区可包含整个核酸（例如，“简并引物”）。简并核酸混合物或半简并核酸混合物可包含碱基对的每种可能的组合、少于碱基对的每种可能的组合，或碱基对的一些组合、碱基对的几个组合，或单一的碱基对组合。简并引物混合物或半简并引物混合物可包含

[0075] “双链”可指已通过互补碱基配对而退火的两条多核苷酸链。

[0076] “已知的寡核苷酸序列”或“已知的寡核苷酸”或“已知的序列”可指已知的多核苷

酸序列。已知的寡核苷酸序列可对应于已设计的寡核苷酸，例如，用于下一代测序平台（例如，Illumina，454）的通用引物、探针、衔接子、标签、引物、分子条形码序列、标识符。已知的序列可包含引物的一部分。已知的寡核苷酸序列可实际上不为特定的使用者所知，但可推定（constructively）获知，例如，通过存储为可由计算机访问的数据。已知的序列也可以是实际上未知的商业秘密或对于一个或多个使用者而言的秘密，但对于已设计用户使用的实验、试剂盒、设备或软件的特定组件的实体可为已知的。

[0077] “文库”可指核酸的集合。文库可包含一个或多个靶片段。在一些情况下，靶片段可以是扩增的核酸。在其他情况下，靶片段可以是未扩增的核酸。文库可包含具有一个或多个已知的寡核苷酸序列的核酸，该寡核苷酸序列添加到 3’端、5’端或添加到 3’和 5’端两端。可制备文库以使得片段可包含鉴定文库来源的已知的寡核苷酸序列（例如，鉴定患者或 DNA 来源的分子鉴定条形码）。在一些情况下，可以合并两个或更多个文库以创建文库池（library pool）。文库也可使用其他试剂盒和技术例如转座子（transpon）介导的标记或如本领域已知的“片段化”来生成。试剂盒可以是市售的，例如 Illumina Nextera 试剂盒。

[0078] “基因座特异性的”可指对应于核酸分子中的位置（例如，染色体或基因组内的位置）的一个或多个基因座。在一些情况下，基因座可与基因型相关联。在一些情况下，基因座可直接从样品中分离并富集，例如，基于杂交和 / 或其他基于序列的技术，或者它们可在序列检测之前使用样品作为模板进行选择性扩增。在一些情况下，可根据个体之间的 DNA 水平变化，基于对特定染色体的特异性、基于所选基因座的 CG 含量和 / 或所需的扩增条件或本领域技术人员在阅读本发明公开内容后将会明了的其他特征来选择基因座。基因座也可指特定的基因组坐标或基因组中的位置，正如该基因组的参照序列所指示的。

[0079] “长核酸”可指长于 1、2、3、4、5、6、7、8、9 或 10 千碱基的多核苷酸。

[0080] 术语“解链温度”或“T_m”通常指使双链核酸分子群体半解离为单链的温度。用于计算核酸的 T_m 的方程式是本领域公知的。给出 T_m 值的简单估计值的一个方程式如下：$T_m = 81.5+16.6(\log10[Na^+])0.41(\%\ [G+C])-675/n-1.0m$。当核酸在具有 0.5M 或更小的阳离子浓度的水溶液中时，(G+C) 含量为 30％至 70％，n 为碱基的数目，并且 m 为碱基对错配的百分比（参见，例如，Sambrook J 等人，Molecular Cloning,A Laboratory Manual, 第 3 版，Cold Spring Harbor Laboratory Press(2001)）。其他参考文献可包括更复杂的计算，这些计算在 T_m 的计算中还考虑结构以及序列特征。

[0081] “核苷酸”可指碱基 - 糖 - 磷酸组合。核苷酸是核酸序列（例如，DNA 和 RNA）的单体单元。术语核苷酸包括核糖核苷三磷酸 ATP、UTP、CTG、GTP 以及脱氧核糖核苷三磷酸例如 dATP、dCTP、dITP、dUTP、dGTP、dTTP 或其衍生物。此类衍生物可包括，例如，[αS]dATP、7- 脱氮 -dGTP 和 7- 脱氮 -dATP 及核苷酸衍生物，其赋予含有它们的核酸分子核酸酶抗性。如本文所用的术语核苷酸还指双脱氧核糖核苷三磷酸（ddNTP）及其衍生物。双脱氧核糖核苷三磷酸的说明性实例包括但不限于 ddATP、ddCTP、ddGTP、ddITP 和 ddTTP。核苷酸可以是未标记的或通过公知技术进行可检测的标记的。荧光标记及其与寡核苷酸的附接描述于多篇综述中，包括 Haugland,Handbook of Fluorescent Probes and Research Chemicals, 第 9 版，Molecular Probes,Inc.,Eugene Oreg.(2002)；Keller 和 Manak,DNA Probes, 第 2 版，Stockton Press,New York(1993)；Eckstein, 编 著，Oligonucleotides and

Analogues:A Practical Approach,IRL Press,Oxford(1991) ;Wetmur,Critical Reviews in Biochemistry and Molecular Biology,26:227-259(1991) ;等等。适用于本发明的其他方法公开于下述作为例子的参考文献中 :Fung 等人,美国专利号 4,757,141 ;Hobbs,Jr 等人,美国专利号 5,151,507 ;Cruickshank,美国专利号 5,091,519 ;Menchen 等人,美国专利号 5,188,934 ;Begot 等人,美国专利号 5,366,860 ;Lee 等人,美国专利号 5,847,162 ;Khanna 等人,美国专利号 4,318,846 ;Lee 等人,美国专利号 5,800,996 ;Lee 等人,美国专利号 5,066,580 ;Mathies 等人,美国专利号 5,688,648 ;等等。也可采用量子点进行标记,如在下列专利和专利出版物中所公开的 :美国专利号 6,322,901、6,576,291、6,423,551、6,251,303、6,319,426、6,426,513、6,444,143、5,990,479、6,207,392、2002/0045045 和 2003/0017264。可检测的标记包括,例如,放射性同位素、荧光标记、化学发光标记、生物发光标记和酶标记。核苷酸的荧光标记可包括但不限于荧光素、5- 羧基荧光素 (FAM)、2’7’- 二甲氧基 -4’5- 二氯 -6- 羧基荧光素 (JOE)、若丹明 (rhodamine)、6- 羧基若丹明 (R6G)、N,N,N’N’- 四甲基 -6- 羧基若丹明 (TAMRA)、6- 羧基 -X- 若丹明 (ROX)、4-(4’二甲基氨基苯基偶氮基) 苯甲酸 (DABCYL)、瀑布蓝 (Cascade Blue)、俄勒冈绿、德克萨斯红、花菁和 5-(2’- 氨基乙基) 氨基萘 -1- 磺酸 (EDANS)。荧光标记的核苷酸的具体实例包括可从加利福尼亚州福斯特市的 Perkin Elmer 获得的 [R6G]dUTP、[TAMRA]dUTP、[R110]dCTP、[R6G]dCTP、[TAMRA]dCTP、[JOE]ddATP、[R6G]ddATP、[FAM]ddCTP、[R110]ddCTP、[TAMRA]ddGTP、[ROX]ddTTP、[dR6G]ddATP、[dR110]ddCTP、[dTAMRA]ddGTP 和 [dROX]ddTTP,可从伊利诺伊州阿灵顿高地 (Arlington Heights,Ill.) 的 Amersham 获得的 FluoroLink 脱氧核苷酸 (FluoroLink DeoxyNucleotide)、FluoroLink Cy3-dCTP、FluoroLink Cy5-dCTP、FluoroLink Fluor X-dCTP、FluoroLink Cy3-dUTP 和 FluoroLink Cy5-dUTP,可从印第安纳州印第安纳波利斯 (Indianapolis,Ind.) 的 Boehringer Mannheim 获得的荧光素 -15-dATP(Fluorescein-15-dATP)、荧光素 -12-dUTP(Fluorescein-12-dUTP)、四甲基 - 若丹明 -6-dUTP、IR770-9-dATP、荧光素 -12-ddUTP(Fluorescein-12-ddUTP)、荧光素 -12-UTP(Fluorescein-12-UTP) 和荧光素 -15-2’-dATP(Fluorescein-15-2’-dATP),以及可从俄勒冈州尤金 (Eugene,Oreg.) 的 Molecular Probes 获得的 Chromosomee 标记的核苷酸 (Chromosomee Labeled Nucleotide)、BODIPY-FL-14-UTP、BODIPY-FL-4-UTP、BODIPY-TMR-14-UTP、BODIPY-TMR-14-dUTP、BODIPY-TR-14-UTP、BODIPY-TR-14-dUTP、瀑布蓝 -7-UTP(Cascade Blue-7-UTP)、瀑布蓝 -7-dUTP(Cascade Blue-7-dUTP)、荧光素 -12-UTP(fluorescein-12-UTP)、荧光素 -12-dUTP(fluorescein-12-dUTP)、俄勒冈绿 488-5-dUTP(Oregon Green 488-5-dUTP)、若丹明绿 -5-UTP(Rhodamine Green-5-UTP)、若丹明绿 -5-dUTP(Rhodamine Green-5-dUTP)、四甲基若丹明 -6-UTP、四甲基若丹明 -6-dUTP、德克萨斯红 -5-UTP(Texas Red-5-UTP)、德克萨斯红 -5-dUTP(Texas Red-5-dUTP) 和德克萨斯红 -12-dUTP(Texas Red-12-dUTP)。核苷酸也可通过化学修饰来标记或标志。化学修饰的单核苷酸可以是,例如生物素 -dNTP。生物素化的 dNTP 的一些非限制性实例可包括生物素 -dATP(例如,bio-N6-ddATP、生物素 -14-dATP)、生物素 -dCTP(例如,生物素 -11-dCTP、生物素 -14-dCTP) 和生物素 -dUTP(例如,生物素 -11-dUTP、生物素 -16-dUTP、生物素 -20-dUTP)。

[0082] “聚合酶”可指使用另一条链作为模板将单个核苷酸连接在一起成为一条链的酶。

[0083] “聚合酶链反应”或“PCR”可指用于（甚至在过量的非特异性 DNA 的存在下）体外复制所选定 DNA 的特定片段的技术。向选定的 DNA 添加引物，此时引物使用核苷酸，以及通常使用 Taq 聚合酶等来引发选定 DNA 的复制。通过使温度循环，选定的 DNA 反复变性并复制。选定 DNA 的单个拷贝，即使与其他、随机的 DNA 混合，也可得到扩增，以获得成千、数百万或数十亿个复制物。聚合酶链反应可用于检测和测量非常少量的 DNA 以及产生定制的 DNA 片段。

[0084] 术语“多核苷酸”可包括但不限于各种 DNA、RNA 分子，其衍生物或组合。这些可包括诸如以下的种类：dNTP、ddNTP、DNA、RNA、肽核酸、cDNA、dsDNA、ssDNA、质粒 DNA、粘粒 DNA、染色体 DNA、基因组 DNA、病毒 DNA、细菌 DNA、mtDNA（线粒体 DNA）、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scaRNA、微小 RNA(microRNA)、dsRNA、核酶、核糖开关和病毒 RNA。

[0085] “引物”通常指一种寡核苷酸，其用于，例如，引发核苷酸延伸、连接和 / 或合成，例如在聚合酶链反应的合成步骤中或在用于某些测序反应的引物延伸技术中。引物也可用于杂交技术中，作为提供基因座与捕获寡核苷酸的互补性以供检测特定核酸区域的手段。

[0086] “引物延伸产物”可指由引物延伸反应产生的产物，该反应使用邻近的多核苷酸作为模板，以及与所述邻近序列互补或部分互补的引物。

[0087] “测序”、“序列确定”等通常指可用于确定核酸中核苷酸碱基顺序的任何及所有的生物化学方法。

[0088] “链霉亲和素”可指可与生物素结合的蛋白质或肽，并且可包括：天然蛋清抗生物素蛋白 (egg-white avidin)、重组抗生物素蛋白、抗生物素蛋白的去糖基化形式、细菌链霉亲和素、重组链霉亲和素、截短型链霉亲和素和 / 或其任何衍生物。

[0089] “受试者”通常指目前存活的生物体或曾经存活的生物体或带有可复制的基因组的实体。本发明的方法、试剂盒和 / 或组合物可施用于一个或多个单细胞或多细胞受试者，包括但不限于：微生物，如细菌和酵母；昆虫，包括但不限于苍蝇、甲虫和蜜蜂；植物，包括但不限于玉米、小麦、海藻或藻类；以及动物，包括但不限于：人、实验室动物（如小鼠、大鼠、猴子和黑猩猩）、家畜（如狗和猫）、农畜（如牛、马、猪、绵羊、山羊）以及野生动物（如熊、大熊猫、狮、虎、豹、大象、斑马、长颈鹿、大猩猩、海豚和鲸）。本发明的方法也可施用于病菌或感染原 (infectious agent)，例如病毒或病毒颗粒，或者已被一种或多种病毒感染的一个或多个细胞。

[0090] “支持物”可以是固体、半固体、珠子、表面。支持物可在溶液中移动或者可以是固定的。

[0091] 独特的标识符可意指分子条形码，也可为混合物中核酸例如 dUTP 的百分比。

II. 多核苷酸

[0092] 可对多核苷酸分子进行处理。例如，可通过化学、物理和 / 或酶的作用来处理多核苷酸。核酸可从受试者或生物样本中获得。在一些实施方案中，核酸是 DNA。DNA 可以是基因组来源的，或为产自受试者 RNA 的 cDNA 文库，或无细胞的 DNA。

[0093] 在一些情况下，多核苷酸可代表生物体或受试者的整个遗传互补体。多核苷酸可以是来自真核生物的基因组 DNA 分子，其可包括内含子和外显子序列（编码序列）以及非编码的调节序列（如启动子和增强子序列）。在一些情况下（例如，DNA），分子可包含基因组 DNA 的多核苷酸序列的子集，诸如，例如，特定的染色体或染色体的片段。多核苷酸可以

是单链或双链的 RNA 或者 RNA 和 DNA 的组合。有时，主要的多核苷酸分子的序列可能是未知的。在一些实施方案中，多核苷酸分子是人基因组 DNA 分子。在一些实施方案中，多核苷酸不是基因组的，而可来自于线粒体、叶绿体、质粒、细菌和 / 或病毒。在一些情况下，多核苷酸分子是来自于已被病毒感染的生物体的染色体或基因组 DNA 分子；在一些情况下，病毒感染可能已引起 DNA 的改变或插入。

[0094] 可对 DNA 分子进行化学处理。在一些情况下，可在任何的片段化过程之前或之后，以及在连接衔接子序列之前或之后处理多核苷酸分子。

[0095] 在许多情况下，例如通过机械剪切或酶消化对多核苷酸的片段化产生具有平端及 3’和 5’突出端的异质混合物的片段。在一些情况下，该方法可提供使用本领域已知的方法或试剂盒（例如，Lucigen DNA 终止子末端修复试剂盒 (Lucigen DNA terminator End Repair Kit)）对片段末端的修复，以产生设计用于插入至例如克隆载体的平端位点的末端。在一些情况下，本发明的方法提供了经测序的核酸群体的平端片段末端。此外，在一些情况下，也可对平端片段进行磷酸化或去磷酸化以便于连接。可通过酶处理，例如使用激酶（例如，T4 多核苷酸激酶）引入磷酸部分，或者可使用碱性磷酸酶对磷酸部分进行去磷酸化。

[0096] 在其他情况下，例如通过某些类型的 DNA 聚合酶（如 Taq 聚合酶或 Klenow exo- 聚合酶）的活性制备具有单个突出的核苷酸的多核苷酸序列，该聚合酶具有非模板依赖性末端转移酶活性，该活性将单个脱氧核苷酸，例如脱氧腺苷 (A) 添加至例如 PCR 产物的 3’端。这类酶可用于将单个核苷酸‘A’添加至靶多核苷酸双链体的每条链的平端 3’末端。因此，可通过与 Taq 或 Klenow exo- 聚合酶的反应将‘A’添加至靶多核苷酸双链体的每条末端修复的双链体链的 3’末端，而衔接子多核苷酸构建体可以是 T- 构建体，其具有存在于衔接子构建体的每个双链体区的 3’末端上的兼容的‘T’突出端。这种末端修饰还防止衔接子和靶标自连接使得存在向形成合并的连接的衔接子 - 靶序列的偏差。

[0097] 多核苷酸可来自于多种来源，包括任何含有遗传物质的物种。在一些情况下，样品可来源于人、哺乳动物、非人类哺乳动物、猿、猴、黑猩猩、爬行动物、两栖动物、鸟类、昆虫或各种无脊椎动物来源。样品也可来源于微生物，该微生物可包括但不限于单细胞生物体或多细胞生物体，细菌、寄生虫、真菌、原生生物、藻类、幼虫、线虫、蠕虫、病毒及其任意组合。

[0098] 此外，样品可从多种组织和组织类型中提取。多核苷酸可以是来源于胚胎的（例如，采自妊娠受试者的流体），或者可来源于受试者自身的组织。多核苷酸也可以作为无细胞的多核苷酸而被发现，或处于不包含在细胞内的状态。多核苷酸可从，例如，体液或组织中提取。

[0099] 在收集含有多核苷酸的组织或体液后，可对样品进行处理。例如，可对核酸进行片段化、纯化、部分纯化和 / 或与来自不同来源的不同的多核苷酸混合，或其任意组合。起始材料核酸可包含已知来源或未知来源的 DNA，或其组合。起始材料核酸可包含已知来源的 DNA，并与已知序列的 DNA 混合。在一些情况下，已知序列的 DNA 可作为对照或样品参照。

[0100] 多核苷酸样品可通过本文的任何方法进行处理。样品可使用本领域已知的多种技术进行分离和提取。多核苷酸的分离和纯化可使用任何手段来完成，该手段包括但不限于使用由诸如 Sigma Aldrich、Life Technologies、Promega、Affymetrix、IBI 等公司提供的商业试剂盒和方案。试剂盒和方案也可以是非市售的。在一些情况下，多核苷酸如 DNA 可

使用市售的试剂盒（例如Qiagen Qiamp® Circulating Nucleic Acid Kit方案、Qiagen Qubit™dsDNA HS Assay 试剂盒方案、Agilent™DNA 1000 试剂盒或 TruSeq™Sequencing Library Preparation ;Low-Throughput(LT) 方案）来分离、提取和制备。多核苷酸样品可通过使用 Ficoll 试剂如 Ficoll-Paque PLUS GE Healthcare Life Sciences 从诸如血液的体液中纯化。

[0101] 如本文所述，多个多核苷酸序列，例如来自基因组的多核苷酸序列，可在后续步骤之前首先进行片段化。在长度方面描述的多核苷酸片段的大小可根据靶多核苷酸的来源、用于片段化的方法和所需的应用而不同。在一些情况下，可使用一个或多个片段化步骤。例如，可使用 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 个或更多个片段化步骤。

[0102] 片段的长度可为约 1-10、10-20、20-50、50-100、50-200、100-200、200-300、300-400、400-500、500-1000、1000-5000、5000-10000、10000-100000、100000-250000 或 250000-500000 个核苷酸。片段的长度可为至少约 10、20、100、200、300、400、500、1000、5000、10000、100000、250000、500000 个或更多个核苷酸。片段的长度可为少于约 10、20、100、200、300、400、500、1000、5000、10000、100000、250000、500000 个核苷酸。

[0103] 多种片段化方法在本文中描述并且是本领域已知的。例如，片段化可通过物理、机械或酶法进行。物理片段化可包括将靶多核苷酸暴露于热或紫外 (UV) 线。机械破碎可用于将靶多核苷酸机械剪切成所需范围的片段。机械剪切可通过包括靶多核苷酸的反复移液、声处理和雾化在内的本领域已知的多种方法来完成。靶多核苷酸也可使用酶法进行片段化。在一些情况下，可使用酶例如使用限制酶进行酶消化。

[0104] 限制酶可用于进行靶多核苷酸的特异性或非特异性片段化。该方法可使用一种或多种类型的限制酶，通常描述为 I 型酶、II 型酶和 / 或 III 型酶。II 型酶和 III 型酶通常可商购获得并且是本领域公知的。II 型酶和 III 型酶识别双链多核苷酸序列内的核苷酸的特定序列（“识别序列”或“识别位点”）。一旦结合并识别这些序列，II 型酶和 III 型酶即切割多核苷酸序列。在一些情况下，切割将产生具有突出的单链 DNA 的一部分（称为“粘端”）的多核苷酸片段。在其他情况下，切割将不会产生具有突出端的片段，而是形成“平端”。该方法可包括使用产生粘端或平端的限制酶。

[0105] 限制酶可识别靶多核苷酸中的多个识别位点。一些限制酶（“精确的切割酶”）仅识别单个识别位点（例如，GAATTC）。其他的限制酶更不加区别 (promiscuous)，识别多于一个识别位点或者多个识别位点。一些酶在识别位点中的单一位置进行切割，而其他酶可在多个位置进行切割。一些酶在识别位点中的相同位置进行切割，而其他酶在可变的位置进行切割。

[0106] 多核苷酸可同时或顺序暴露于两种或更多种限制酶。这可通过以下步骤来完成：例如，将多于一种限制酶添加至分区，或者将一种限制酶添加至分区，进行消化，灭活该限制酶（例如，通过热处理），然后添加第二种限制酶。

[0107] 在一些实施方案中，本发明可使用靶核酸的稀释和空间分离。在一些情况下，核酸的长片段在进行空间分离之前进行稀释。稀释可通过本领域已知的任何方法，例如通过加入稀释剂如水或合适的缓冲液来完成。稀释的示例性方法涉及在稀释前确定核酸的浓度，并计算要加入多少稀释剂才能使稀释的样品可被分配成含有亚基因组量的 DNA 的量（即，使一个样品包含少于一个完整的基因组）。在另一种示例性方法中，可计算使得样品可以以

每个分区含有约 1、2、3、5、10、20、50、80、100、150、200、400、500、1000、1500、5,000、10,000、100,000、1,000,000、10,000,000、100,000,000、1,000,000,000 个核酸片段的方式进行分配的稀释度。在另一种示例性方法中，进行稀释以便于样品的分配，从而可以分离基因组的一个拷贝的约 1%、2%、5%、10%、15%、20%、30%、50%、70%、80%、95%或 100%。

[0108] 空间分离可通过本领域中已知的许多方法如移液、微量移液或微流体技术来完成。分区可由本领域中任何已知的方法制得，该方法包括板（例如 96-孔）、微流体室、微滴或在固体表面如硅片或珠子或半固体表面上的简单空间分离。油和 / 或乳液可用于空间分离。

[0109] 在一个示例性方法中，进行稀释和空间分离，以使得两个分区包含来自每个亲本染色体的 DNA 的相同基因座的概率较低，或者使得来自相同的基因组基因座的多个片段将极为罕见。

III. 靶向测序

[0110] 靶向测序可包括检测复杂变异、避免克隆错误以及在计算上较不繁琐的分析的能力（例如，从头测序）。存在数个靶向测序的实施方案。一些实例可包括使用序列特异性引物 (PELA) 的引物延伸、靶向随机引物延伸连接和标记（靶向 RAPELLing)。靶向 RAPELL 可包括靶向步骤，其使用本文公开的 RAPELLing 方法。其他的靶向测序方法可包括高度多重 PCR，其可使用生物素 dUTP 进行大范围或小范围 PCR。靶向测序的另一个实施方案为靶向延伸连接和扩增 (TELA)。靶向测序的另一个实施方案可为多重引物延伸和亲和反应 (mPEAR)。靶向测序方法可包括环化滚环扩增 (CRCA)。在一些情况下，“靶向测序”是指分离和扩增生物相关基因组位置以供 DNA 测序的任何方法。在一些情况下，对基因组的保守或功能元件进行扩增，该元件与通过 DNA 测序进行的分析相关。在一些情况下，这可包括表观遗传学信息，例如核酸的甲基化，例如甲基化的 DNA。

A. 靶向引物延伸

[0111] 可通过使用包含可与已知序列杂交的区域的引物来靶向引物延伸。在一些实施方案中，已知序列在靶基因座内。在一些实施方案中，已知序列在靶基因座外。

[0112] 可将引物设计为覆盖式 (tiled) 的。引物覆盖策略可通过使用多个未配对的或配对的引物以使每个引物都可产生扩增子来完成。可设计引物以使得多个引物可产生多个可“覆盖”基因座的扩增子。覆盖可意指扩增子可彼此重叠。可设计引物以使得产生的扩增子基本上覆盖整个靶基因座 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、50、100、200、300、400、500、600、700、800、900、1000、2000、5000、10,000、50,000、100,000、500,000、1,000,000 次或更多次。在一些情况下，可通过扩增较长靶区域的区域来覆盖基因座。在一些实施方案中，该基因座是靶基因座。引物可分别与围绕并包括靶基因座的区域杂交。可设计最远的上游引物以使其能够与起始于靶基因座上游约 30、40、50、60、70、80、90、100、120、130、140、160、180、200、250、300、350、400 个或更多个核苷酸的区域杂交。在一些实施方案中，最远的上游引物的起始区域位于靶基因座上游约 100 至约 200 个核苷酸。可设计最远的下游引物以使其能够与起始于靶基因座下游约 30、40、50、60、70、80、90、100、120、130、140、160、180、200、250、300、350、400 个或更多个核苷酸的区域杂交。在一些实施方案中，最远的下游引物的起始区域位于靶基因座下游约 100 至约 200 个核苷酸。

[0113] 在一些情况下，可将引物设计成与参照基因组互补。所使用的参照基因组可以是标准参照或种族或群体特异性参照，例如可在每个多态性位置处包括主要的等位基因的参照。在 SNP 的情况下，四种核苷酸的简并或 N 种或每种可能的核苷酸均可用于引物中该位置的合成。参照设计可包括分配“窗口 (window)”，该窗口可包含针对靶位置的 100bp 序列。在一些情况下，窗口可为至少 1、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、250000 或 500000 个碱基对。在一些情况下，窗口可为至多 1、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、10000、250000 或 500000 个碱基对。可将覆盖的窗口分配给靶基因座，以使得可从每个窗口选出基于长度 TM 和特异性的表现最佳的探针。可调整窗口以对应来自所定义的窗口的任何偏差。在一些情况下，窗口可以是重叠的。

[0114] 在一些实施方案中，重叠的扩增子可覆盖基本上所有的靶基因座。在一些实施方案中，重叠的扩增子可覆盖基本上所有的靶基因座和靶区域的侧翼区域。用于覆盖靶基因座的扩增子的数目可为约 2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、150、200、250、300、350、400、500、600、700、800、900、1000、1500、2000、2500、3000、4000、5000、6000、7000、8000、9000、10,000 个或更多个。用于覆盖区域的扩增子的数目可依赖于该区域的长度、该区域内的序列变异性、该区域内的串联重复的数目、该区域中 CG 碱基的百分比或该区域内可影响测序质量的其他序列或结构变异。通常，可优选具有较多数目的扩增子覆盖区域，以提高测序的深度和准确性。

[0115] 在一些情况下，窗口或覆盖的设计使得特定序列基序如 CCG 多核苷酸可使测序化学法产生错误。读取值中的这些多核苷酸序列的位置可能影响测序质量。在一些情况下，如果序列基序位于固体表面附近，则与如果其位于读取值的末端相比，它们可能对读取的质量具有差别性影响。在一些情况下，调整这些基序在读取值中的位置产生共有且随机的错误分布，这可使用冗余读取值处理进行校正。这是在使用相同的读取结构的基于 PCR 的方法中无法获得的一种属性。PCR 方法也可引入最不繁琐的序列组合以避免测序期间聚合酶的打滑 (stuttering)。这样的结果可能导致在集群扩增中的 CGG 序列解读，而非被认为是系统误差的正确 CCG 构型。

[0116] 覆盖式策略也可改变靶标每个位置处的读取值中的错误分布 (例如，通过具有不同测序起始位点的覆盖式扩增子)。例如，诸如“CCG”的序列可能更难以精确地进行通过式测序 (sequence through)。然而，在一些情况下，出现 CCG 序列的读取值的位置可直接影响化学法通过该读取值进行测序的能力。因此，覆盖式策略可调整读取值中 CCG 位置的位置，这可允许提高化学法准确地通过 CCG 序列基序进行读取的可能性。

[0117] 在一些实施方案中，扩增的区域可包括靶区域。扩增子的长度可为约 50、60、70、80、90、100、110、120、140、150、160、170、180、190、200、210、220、230、240、250、270、290、300、320、340、360、380、400、450、500、550、600 个或更多个核苷酸。在一些实施方案中，每个扩增子通常为约 100 个核苷酸至约 200 个核苷酸。

[0118] 覆盖式扩增子可包含相互重叠的核苷酸。重叠可为约 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、45、50、55、60、65、70、75、80、85、90、95、100、110、120、130、140、150、160、170、

180、190、200、225、250、275、300 个或更多个核苷酸。

[0119] 可设计引物以使其得到优化。用于优化引物的方法的一些非限制性实例为：改变长度或序列以达到最佳 Tm；基于引物对靶杂交位置的特异性；避免彼此发生二聚化和 / 或避免共同的多态性。在一些实施方案中，可设计引物以避免具有共同的已知多态性的区域，即，在该区域中次要等位基因频率高于 1%、2%、3%、4%、5%、6%、7%、8%、9%、10%或更高。引物可包含可通过聚合酶延伸的 3’OH 基团。

[0120] 引物可通过使用核苷酸进行标记。引物可在 3’端、5’端和 / 或在中间进行标记。引物可包含标记的核苷酸（例如，与生物素或荧光部分偶联的核苷酸）。用于标记核苷酸的方法已在本文中进行了描述。引物可包含 5’核苷酸尾标记。引物可包含核苷酸尾和 / 或标记的核苷酸。5’核苷酸尾可包含已知的序列。添加至 5’核苷酸尾中的已知序列可使扩增子可用于下游反应。已知序列可包含衔接子、分子条形码和 / 或其他已知序列。

[0121] 衔接子或测序衔接子或衔接子序列可包含与反应中使用的测序平台的衔接子相对应的序列。例如，可将衔接子添加至 5’端或其附近。在一些实施方案中，5’衔接子被称为 A 衔接子。衔接子序列在本文中进一步描述。分子条形码可对应于样品来源、延伸的方向和 / 或靶区域。

[0122] 引物可与核酸杂交并可发生延伸。可使用链置换热稳定性聚合酶进行延伸。模板核酸的拷贝可在延伸过程中通过聚合酶产生。聚合酶也可用于置换与模板杂交的任何核酸。扩增子产生过程可被重复或循环。扩增子产生过程可包括：热变性、引物退火和引物延伸。

[0123] 可设计引物以使得仅模板的正向链被复制。在其他实施方案中，可设计引物以使得仅模板的反向链被复制。在其他实施方案中，可设计引物以使得模板的正向链和反向链均被复制。引物可包含含有两个或更多个核苷酸的分子条形码，该分子条形码可与有关取向或正被复制的模板链的信息相关联。

[0124] 正向和反向反应可一起进行。正向和反向反应可单独进行。当正向和反向反应单独进行时，它们可在后续步骤中组合。从单独的正向和反向反应产生的扩增子可在文库产生前混合。由单独的正向和反向反应产生的文库可在测序之前合并。正向和反向反应可分别进行测序并且数据可在计算机中 (in silico) 组合。

[0125] 正向和反向链扩增子可用于错误校正。正向和反向链扩增子可用于精修定位 (mapping)。正向和反向链扩增子可用于距离分析。

[0126] 可对扩增的产物或扩增子进行大小处理以减小或控制总长度。在一些情况下，大小处理可为片段化，在其他情况下，大小处理可以是停止延伸。大小处理可产生对于测序而言可能是最佳的大小的扩增子。扩增子的大小处理可通过本领域中已知的任何酶或物理手段而进行。

[0127] 酶片段化可通过使用可切割（例如，水解）核酸键的酶来进行。可切割核酸的酶的一些非限制性实例包括：水解酶、核酸酶、核糖核酸酶、脱氧核糖核酸酶、磷酸酯酶、拓扑异构酶、内切核酸酶、限制酶、II 型限制性内切核酸酶或 I 型限制性核酸酶。

[0128] 在一些情况下，核酸的片段化可通过物理或机械力来进行。对核酸进行物理片段化的一些非限制性方法包括：声处理、雾化或水力剪切 (hydroshearing)。

[0129] 扩增子的大小处理也可在延伸期间在 3’端进行。在一些情况下，大小处理通过

产生随机 3’端来完成。3’大小处理可通过使用缺少 3’OH 基团的核苷酸、生物素 -ddNTP、dUTP 后使用 UDG/APE1、甲基 C 或其他经修饰的核苷酸来进行。此类核苷酸的掺入可使延伸停止或终止。此类核苷酸的掺入频率可通过滴定此类核苷酸的量来改变。在一些实施方案中，使用生物素化的 ddNTP 终止的扩增子可通过使用链霉亲和素珠纯化而被进一步分离。

[0130] 可将核苷酸添加至扩增子的 3’端；这可被称为 3’核苷酸尾。3’核苷酸尾可包含已知的序列。已知的序列可使扩增子可用于下游反应。3’核苷酸尾可包含衔接子、分子条形码和 / 或用于扩增的已知序列。衔接子可包含与可与测序平台一起使用的衔接子相对应的序列。在一些实施方案中，该 3’衔接子可被称为 B- 衔接子。分子条形码可对应于样品来源、延伸的方向和 / 或靶区域。第二个已知序列可包含已知的核酸。用于添加 3’核苷酸尾的方法可根据用于大小处理的方法而不同。在一些情况下，3’尾可为进一步的核苷酸的延伸，从而使一个拷贝和 5’标记的分子区别于另一拷贝和 5’标记的分子。

[0131] 在 3’OH 的存在下（例如，如果使用酶或物理片段化作为大小处理的方法），3’核苷酸尾可通过连接或通过另外的引物延伸步骤来添加。3’核苷酸尾可通过连接来添加，步骤包括：具有随机突出端的双链构建体的末端修复和连接。3’核苷酸尾可通过引物在单链模板上的延伸和延长来添加。

[0132] 如果将终止核苷酸用于大小处理，由于缺乏 3’OH，分子的 3’端可能无法用于引物延伸或连接。在 3’OH 缺失的情况下，3’核苷酸尾可通过引物杂交来添加，该引物在 3’端上包含随机核苷酸序列且在 5’端上包含 3’核苷酸尾的互补序列（例如 5’- 已知序列 - 随机序列…-3’）。引物的随机序列可与扩增子杂交。随机序列可由 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 个或更多个核苷酸组成。包含随机序列的引物的混合物可包括所有可能的序列的混合物。在一个非限制性实例中，如果随机序列包含 5 个核苷酸，则可能存在 A、T、C 和 G 碱基的 4^5 或 1024 种可能的组合。在另一个非限制性实例中，如果随机序列包含 6 个核苷酸，则可能存在 A、T、C 和 G 碱基的 4^6 或 4,096 种组合，并且引物混合物可包括所有的或基本上所有的可能组合的混合物。

[0133] 在一些实施方案中，扩增子已通过引入结合链霉亲和素的生物素 -ddNTP 而纯化或捕获。链置换聚合酶可用于延伸 5-3’。在生物素化的模板末端的随机序列可延伸并置换所有其他随机关联的引物，从而成为模板扩增子上唯一延伸的核酸。在一个非限制性实例中，复合物可包含 5’A 衔接子 - 基因座特异性引物 - 靶标 - 终止 ddNTP- 生物素 3’；反向链可包含 5’-B 衔接子 - 随机序列 - 靶标 - 基因座特异性序列 -A 衔接子 -3’。该复合物可以是双链的。该复合物可进一步通过洗涤链霉亲和素珠并移除上清液而分离。该复合物的一条链可通过热变性从链霉亲和素珠上释放，而另一条链可通过生物素部分保持与链霉亲和素共价结合。

[0134] 已处理的扩增子可通过线性扩增或 PCR 进行扩增。为了扩增核酸，可使用与 A 和 B 衔接子序列互补的引物进行 2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30 次或更多次 PCR 循环。在示例性实施方案中，进行约 9 至约 12 次 PCR 循环。

[0135] 扩增步骤可在 5’和 3’端掺入更多已知的核苷酸（例如，通过使用在 5’端含有已知序列的 PCR 引物）。分子条形码可在该步骤中添加。在一些实施方案中，在 5’和 / 或 3’端的已知核苷酸序列可不包含可能是下一代测序反应所必需的全长测序衔接子。在这些情

况下，全长测序衔接子可在 PCR 步骤中引入。在一些情况下，例如对于 ILMN 测序，衔接子序列的大部分 3’端可能是相同的。在一些情况下，可能需要添加一个额外的碱基或几个额外的碱基，以确保文库分子的同一条链上的 A 和 B 衔接子序列的方向性。

[0136] 在示例性实施方案中，所得的测序仪就绪文库 (sequencer-ready libarary) 可由以下形式的双链分子组成 :5’ -A 衔接子 - 合成引物 - 扩增子 -B 衔接子 -3’。在一些实施方案中，所得的测序仪就绪文库可包含 1、2 或 3 个或更多个分子条形码。A 衔接子和 B 衔接子可对应于与测序平台一起使用的衔接子序列 ;衔接子序列在本文中进行了讨论。任何测序方法均可用于分析测序仪就绪文库 ;测序方法的汇总可见于本文中。从这样的测序反应中获得的数据可通过本领域中已知的任何方法进行存储、传送和 / 或分析 ;数据存储和传送方法可见于本文中。

[0137] 测序读取值可通过数据分析来进行分析。数据可使用软件来分析。软件可从读取值中修剪掉 (trim) 衔接子。如果引入了任选的分子条形码则可对样品进行鉴定。可去除重复的读取值。可通过读取开始时的已知合成序列来鉴别读取值的基因组坐标。与相同的已知基因组坐标相对应的读取值可被分箱 (binned) 在一起，并可产生共有序列。共有序列可在不使用参照基因组的情况下产生。可将共有序列与参照基因组进行比较。可从分析中去除不形成共有序列的读取值。靶核酸的每个邻近区段可被认为是单一的“靶标”，而与该靶标对应的所有引物被认为是该靶标的引物组。无论靶标大小如何，这都可产生等于靶标全长的计算机读取长度。

[0138] 在确定共有序列后，可查询靶区域的已知单元型。单元型可以是多态性集合的成员。单元型数据可以是关于样品的单元型的信息。可鉴定已知引起或不引起疾病的匹配单元型。对于与已知的单元型不匹配的共有序列，可使用从头测序来确定新的单元型、单倍群和 / 或结构变异。

[0139] 靶向测序的优点可包括，数据分析可能不需要将每个读取值与参照基因组进行比对 (即，每个读取开始时的合成序列可识别基因组位置，而该读取的剩余部分可以是从头的或没有参照的)。测序变体可更可靠地与基因组变体区分开来。例如，覆盖具有变异序列的同一基因组区域的几个扩增子可能指示该变体为基因组的。引物组可针对每个邻近的靶标进行分箱，从而允许组装。可检测复杂变异。读取值的随机 3’端可避免克隆错误 (即，可显示具有低频率变异例如体细胞突变的不同的模板)。靶向测序方法的其他优点包括 : ddNTP 终止可提供可能涉及较少清理 (clean-up) 的随机片段化。嵌合文库分子中的合成序列可用于鉴定样品和基因组坐标。在一些情况下，它们可从读取值和组装分析中移除，使得仅来自样品的序列用于辅助从头组装。读取值的随机 3’端可确保被测序的分子不是“克隆的”，从而可使错误大幅度减少并使检测体细胞变异的灵敏度更大。重复的读取值可在计算机中去除，而传统的基于 PCR 的靶向测序不能去除重复的读取值。覆盖式探针设计可允许冗余采样和共有读取长度。这可避免靶标由于在一个引物位点下的 SNP 或新的生物学而逃脱。引物延伸期间生物素的引入可简化清理。已分离的分子的随机引发可允许在不进行连接或另外的清理或末端修复的情况下引入 B- 衔接子。每个读取开始时的合成序列可通过避免参照基因组定位而大大降低计算负担。长的共有读取值可允许检测复杂变异。靶标特异性合成引物的线性消耗可降低每个样品的成本并增加每生产批次的体积。

B. 高度多重 PCR

[0140] 靶向测序方法可包括使用或不使用生物素的高度多重 PCR。可从基因组 DNA 样品中分离出靶区域。靶区域可包含多个靶基因。靶基因已在本文中公开。可使用 PCR 产物的生物素化捕获来分离靶区域。PCR 产物的大小范围可为 10-1000、100-10,000、100-20,000、1,000-20,000、2,000-15,000、10,000-15,000、10,000-20,000、10,000-100,000 或 10-200,000 个核苷酸的长度。

[0141] 可将引物设计为位于一个或多个目标区域（即靶区域或靶基因座）的侧翼。一个或多个目标区域可以是基因。引物可与靶序列杂交。可进行扩增反应。扩增反应可为 PCR。PCR 可以是大范围 PCR。扩增反应可以是高度多重化的。扩增反应可以是低拷贝或低循环。一个或多个基因座的 PCR 可以是多重化的。PCR 可为高度多重 PCR。高度多重化可以是多于 10、20、30、40、50、60、70、80、90、100、120、140、160、180、200、250、300、400、500、600、700、800、900、1000、1500、2000、3000、4000、5000、6000、7000、8000、9000 个或更多个靶基因座。基因座的扩增可独立地进行。可将一个或多个扩增产物合并。

[0142] 可掺入生物素偶联的 dNTP。生物素偶联的 dNTP 可在扩增期间掺入。可使用 dNTP 混合物进行扩增，该 dNTP 混合物包含约 0.1%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、12%、15%、20%、25%或更多的生物素偶联的 dNTP。在一些实施方案中，生物素偶联的 dNTP 可以是 dUTP。

[0143] PCR 扩增产物可在凝胶上电泳。该凝胶可以是琼脂糖凝胶。该凝胶可以是低熔点琼脂糖凝胶。可将分子量标准梯 (marker ladder) 加入邻近的孔中。可切下包含 PCR 产物的凝胶部分。在一些实施方案中，将要切下的部分将由基于引物设计所预期的长度来指导。在一些实施方案中，与 10-15K 范围的 PCR 产物相对应的凝胶部分为将要切下的凝胶部分。

[0144] 可从琼脂糖中纯化扩增产物。可将凝胶切下的琼脂糖溶解。凝胶切下的琼脂糖可溶解于溶解缓冲液中。可进一步分离纯化或半纯化的多核苷酸。该多核苷酸可与柱结合。该柱可以是基因组 DNA 结合柱。可洗涤该柱。可从柱上洗脱多核苷酸。可对洗脱的多核苷酸进行片段化。片段化方法已在本文中公开。可将多核苷酸片段化成范围为约 50-1000、100-1000、150-700、200-660、100-800 或 10-1500 个核苷酸的片段范围。

[0145] 可分离靶扩增产物。靶扩增产物可基于大小选择来分离。可通过凝胶电泳 / 凝胶纯化、大小排阻柱 / 柱清理和 / 或针对大小选择而优化的固相可逆固定 (SPIR) 珠子，基于大小来分离靶扩增产物。

[0146] 靶纯化产物可通过亲和纯化（例如，如果已掺入生物素 dNTP，则使用链霉亲和素）来分离。可将包含生物素的多核苷酸暴露于链霉亲和素。可在多核苷酸的片段化之前或之后将其暴露于链霉亲和素。链霉亲和素可包含链霉亲和素涂覆的珠子。可对链霉亲和素 - 生物素 - 片段复合物进行纯化或半纯化。可洗涤链霉亲和素 - 生物素 - 片段复合物。

[0147] 可对多核苷酸片段进行处理。在本文中公开了多核苷酸处理方法。可对多核苷酸片段进行末端修复。在本文中公开了末端修复的方法。可对多核苷酸片段进行衔接子加尾。在本文中公开了衔接子加尾的方法。可对多核苷酸片段进行扩增。在本文中公开了扩增的方法。可在链霉亲和素的存在下扩增多核苷酸片段。

[0148] 扩增产物可以是测序仪就绪的，并且可包括序列文库。可进一步处理扩增产物。本文中公开了处理多核苷酸的方法。扩增产物可在琼脂糖凝胶上电泳。可在测序之前将扩增产物从凝胶上切下并纯化。

[0149] 本文中公开了对序列文库进行测序的方法。可从对文库的测序中产生数据。本文中公开了存储和传送已从测序反应中产生的数据的方法。可对该数据进行分析或处理。本文中已公开了分析或处理数据的方法。可使用经分析的数据。本文中已公开了使用经分析的测序数据的方法。

C. 靶向延长和连接衔接子 (TELA)

[0150] 可使用可供选择的方法（本文中称为 TELA）对靶分子进行标记或条形码化。如图 10 中所示，TELA 方法通常在涉及引物杂交的多步骤过程中提供条形码的附接，该引物包含：基因座特异性序列、通用衔接子序列和条形码序列。具有这种构造的引物在本文中称为 TELA 引物。然后进行引物延伸，随后将产物连接，以形成连续的序列。在另一个实施方案中，与条形码序列邻接的通用衔接子序列可与包含其他通用衔接子序列的多核苷酸序列杂交，并一起扩增。在其他情况下，与条形码序列邻接的通用衔接子序列可与包含可能适于高通量测序平台或其他应用（例如基因座特异性引物与随机加尾的引物的连接）的其他衔接子序列的多核苷酸序列杂交。

D. TELA 引物的设计和退火

[0151] 在一些实施方案中，靶多核苷酸首先与 TELA 引物杂交，该 TELA 引物由基因座特异性序列或第二和第三探针结构域（其可为已知的、部分已知的或未知的序列）以及通用衔接子序列和条形码，或第一和第 4 探针结构域形成。TELA 引物与靶分子的杂交形成引物 - 靶标构建体，接着使用该引物 - 靶标构建体进行初始引物延伸反应，其中如图 7 所示形成与每个单独的衔接子 - 靶标构建体的模板链互补的延伸产物。接着将所得引物延伸产物连接并可进行扩增，以共同提供如图 6 中所示的标记的或条形码化的模板多核苷酸的文库。术语文库是指在其 3’ 和 5’ 端包含已知的共同序列的靶片段的集合，并且也可称为 3’ 和 5’ 修饰的文库。在一些情况下，基因座特异性引物和随机引物可以以相似的策略使用，由此正向或反向基因座特异性引物与相应的（即正向 / 反向）随机引物组合使用。在一些情况下，使用基因座特异性引物和随机引物的组合可产生重叠的扩增子。在一些情况下，这可产生用于测序的更长的靶标长度。

[0152] 在一些情况下，针对所有样品的通用衔接子序列或双链体的一条或两条链可携带标签序列，以条形码化或追踪样品的身份。在一些情况下，不包括条形码。

[0153] 如图 7 中所示，TELA 引物序列的一个重要特征是该序列的一部分可能不与靶序列 710 完全退火。通常，该序列的这一部分包含单独的引物位点 730，该引物位点 730 与互补于目标基因座的序列 700 相邻。TELA 序列通常通过两条部分互补的多核苷酸链的退火而形成，以便在两条链退火时提供至少一个双链区和至少一个非互补单链区。在一些情况下，基因座特异性区域可连接至随机间隔区序列 720。在一些情况下，间隔区序列的长度可以少于 20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900 或 1000 个核苷酸。在其他情况下，间隔区序列的长度可多于 20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900 或 1000 个核苷酸。

[0154] 通常，衔接子的双链区包含多个连续的核苷酸，其由两条部分互补的多核苷酸链的退火而形成，通常位于基因座特异性序列的互补序列之间。如本文所提供的，双链的通常是指已退火的两条链，而非指代任何特定的结构 DNA 特征。此外，双链区也可指代基因座特异性序列（当该序列互补于靶多核苷酸中的序列时）。

[0155] 通常，在不损失功能的情况下，可将如在 mPEAR 引物中的 TELA 引物的基因座特异性区域设计成尽可能短。在该上下文中，“功能”是指双链区在用于酶催化的核酸连接反应的标准反应条件（例如，在 4℃至 60℃范围的温度下在适于酶的引物延伸缓冲液中进行温育）下形成稳定的双链体，从而使形成衔接子的两条链在引物于靶分子上延伸期间仍保持部分退火的能力。

[0156] 将相同的衔接子连接到每个靶多核苷酸的两个末端。每个衔接子－靶标构建体中的靶序列的侧翼将为由引物的双链区衍生的互补序列。双链区越长，并因此在衔接子－靶标构建体中由其衍生的互补序列越长，则在用于引物延伸和 / 或 PCR 的退火条件下 TELA 引物－靶标构建体在这些内部自互补区域中能够回折并与其自身进行碱基配对的可能性越大。因此，通常双链区的长度可为少于 100、90、80、70、60、50、40、30、20、10 个核苷酸。在一些情况下，双链区的长度可为多于 100、90、80、70、60、50、40、30、20、10 个核苷酸，以降低这种影响。可通过包含表现出比标准 Watson-Crick 碱基对更强的碱基配对的非天然核苷酸来提高双链区的稳定性，并因此潜在地减小双链区的长度。

[0157] TELA 引物在基因座特异性序列中的互补百分比可不同。在一些情况下，其在双链区中可为 100％互补。在其他情况下，其可为大于 1％、10％、20％、30％、40％、50％、60％、70％、80％、90％互补。在其他情况下，其可为小于 1％、10％、20％、30％、40％、50％、60％、70％、80％、90％互补。在双链区内可容忍一个或多个核苷酸错配，条件是两条链能够在标准连接条件下形成稳定的双链体。

[0158] 用于该方法中的通用衔接子的序列通常可包括形成衔接子的“可连接”末端（即在连接反应中与靶多核苷酸连接的末端）的双链区。衔接子的可连接末端可以是平端，或者在其他实施方案中，可存在一个或多个核苷酸的短 5’或 3’突出端，以便于 / 促进连接。对处于衔接子的可连接末端的 5’末端核苷酸进行磷酸化，以使得能够与靶多核苷酸上的 3’羟基形成磷酸二酯键连接。

[0159] 另一个特征可包括通用衔接子序列的区域，在该区域中形成衔接子的两条多核苷酸链表现出一定程度的非互补性，使得这两条链在用于引物延伸或 PCR 反应的标准退火条件下不能彼此完全退火。在一些情况下，这一区域可在用于酶催化连接反应的标准反应条件下经受退火，条件是两条链在扩增反应中在退火条件下恢复到单链形式。

[0160] 通常，可将区域设计为防止以如本文所述的多种方式退火。

[0161] 就长度来说，可能不退火的区域可根据功能来确定，该功能例如是，需要为引物延伸、PCR 和 / 或测序提供适合引物结合的序列。通常，这样的区域的长度可延伸任何数目的核苷酸。在许多情况下，优选最小化衔接子的总长度，例如，以便在连接步骤之后从衔接子－靶标构建体中分离未结合的衔接子。因此，通常优选的是未退火区域的长度应少于 20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900 或 1000 个核苷酸。在其他情况下，未退火区域的长度应多于 20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900 或 1000 个核苷酸。在一些情况下，非互补区可使引物的 5’端不稳定。在一些情况下，最小化该引物的长度还可使尾序列可部分地与模板 DNA 的另一区域结合的可能性最小化。

[0162] TELA 引物的实际核苷酸序列可以是任何合适的序列和长度。TELA 引物可由使用者选择，以使所期望的序列元件最终包括在来源于衔接子的模板的文库的共同序列中，例

如，以便为特定的通用扩增引物和 / 或测序引物组提供结合位点。可包括另外的序列元件，例如，以便为最终将在文库中模板分子的测序中使用的测序引物或由模板文库的扩增（例如在测序应用中在固体支持物上）获得的产物提供结合位点。

[0163] TELA 引物序列可包含 DNA 的两条链，但也可包含可能适合的任何核苷酸或核苷酸衍生物。可替代的核苷酸可包括通过磷酸二酯和非磷酸二酯骨架键的混合物连接的天然和非天然核苷酸（例如，一种或多种核糖核苷酸）的混合物。可包括其他非核苷酸修饰，例如，生物素部分、封闭基团和用于附接的捕获部分，如生物素化的核苷酸。

[0164] 进一步地，衔接子序列还可包含外切核酸酶抗性修饰，如硫代磷酸酯键。这样的修饰减少了存在于文库中的衔接子二聚体的数目，因为两个衔接子不能在没有除去它们的非互补性突出端的情况下连接。衔接子可在与靶标的连接反应之前采用外切核酸酶进行处理，以确保链的突出端在连接过程中不能被去除。以这种方式处理衔接子减少了连接步骤中衔接子二聚体的形成。

E. 引物延伸和连接

[0165] 在使 TELA 引物退火之后，可使用如本文所述（参见第 II-F 部分）的任何合适的聚合酶进行引物延伸反应。此外，可使用如本文所述的连接方法将一个或多个产物连接在一起，以形成连续的序列。

[0166] 此外，如图 6 所示，可将一个或多个连续的引物延伸 - 连接产物进一步组装成更长的片段以供下游分析。在一些情况下，引物中的通用衔接子位点可用作杂交位点以将一个或多个产物退火在一起。使用 PCR 或进一步的引物延伸反应可将多个产物组装成较长的连续链。

[0167] 可将合并的连接的多核苷酸序列与未连接的衔接子多核苷酸构建体从连接反应的任何组分如酶、缓冲液、盐等中纯化。合适的纯化方法是本领域已知的，并采用标准方法 (Sambrook 和 Russell, Molecular Cloning, A Laboratory Manual, 第三版）。

[0168] 此外，在连接第一引物延伸产物后，可将衔接子附接至连接产物的侧翼区。包含衔接子区域和通用引发区域且互补或部分互补于连接产物中的通用衔接子位点的衔接子引物可用于产生具有附接的衔接子的多核苷酸。使用 TELA 引物的通用方案（即杂交，随后进行引物延伸和连接）以及 TELA 引物的通用设计可能适用于衔接子引物的应用。

[0169] 在一些情况下，衔接子序列对于下游应用（例如，如本文所述的测序）可能是有用的。

F. 扩增技术

[0170] 许多扩增方法和技术是本领域已知的。任何合适的方法可以在本发明的方法中使用，以便增加多核苷酸的数量或量，同时保持原样品或连接产物的序列信息的初始含量。可使用一种或多种扩增方法并将其以一种或多种组合使用。

[0171] 扩增方法的实例可包括但不限于：聚合酶链反应 (PCR)（美国专利号 4,683,195 和 4,683,202 ;PCR Technology:Principles and Applications for DNA Amplification, H.A.Erlich 编著，Freeman Press, NY, N.Y., 1992）、连接酶链反应 (LCR)(Wu 和 Wallace, Genomics4:560, 1989 ;Landegren 等人，Science 241:1077, 1988）、链置换扩增 (SDA)（美国专利号 5,270,184 和 5,422,252）、转录介导的扩增 (TMA)（美国专利号 5,399,491）、连锁线性扩增 (linked linear amplification, LLA)（美国专利号 6,027,923)

等，自动维持序列复制 (Guatelli 等人，Proc.Nat.Acad.Sci.USA,87,1874(1990)；和 WO90/06995)、靶多核苷酸序列的选择性扩增（美国专利号 6,410,276)、共有序列引发的聚合酶链反应 (CP-PCR)（美国专利号 4,437,975)、任意引发的聚合酶链反应 (AP-PCR)（美国专利号 5,413,909、5,861,245) 以及基于核酸的序列扩增 (NASBA)。（参见，美国专利号 5,409,818、5,554,517 和 6,063,603，各专利均通过引用并入本文）。其他可使用的扩增方法包括：Qβ 复制酶（描述于 PCT 专利申请号 PCT/US87/00880 中）；等温扩增方法，例如 SDA（描述于 Walker 等人，Nucleic Acids Res.20(7):1691-6(1992) 中）；以及滚环扩增（描述于美国专利号 5,648,245 中）。其他可使用的扩增方法描述于美国专利号 5,242,794、5,494,810、4,988,617 和美国序列号 09/854,317 以及美国公开号 20030143599 中，各自均通过引用并入本文。在一些方面，通过多重基因座特异性 PCR 对 DNA 进行扩增。在一个优选的方面，使用衔接子 - 连接和单引物 PCR 对 DNA 进行扩增。其他可用的扩增方法，如平衡 PCR(Makrigiorgos 等人，Nature Biotech,20:936-9(2002)) 和等温扩增方法，如基于核酸序列的扩增 (NASBA)，和自动维持序列复制 (Guatelli 等人，PNAS USA 87:1874(1990))。基于这些方法，本领域技术人员能够容易地设计任何合适的待扩增区域中的引物。

G. 扩增产物和条件

[0172] 通常，任何合适的扩增产物和用于生成产物的条件可在本发明的方法中使用。可使用适于不同扩增技术和序列的不同的扩增长度、循环次数、杂交、退火和延伸条件。

i. 扩增长度

[0173] 通常，扩增产物的长度可为任意长度并且包含可能在列举序列中有用的任意序列。通常，扩增的多核苷酸可为至少约 5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、75kb 或 100kb。通常，扩增的多核苷酸可为至多约 5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、75kb 或 100kb。

ii. 扩增条件

[0174] 通常，对于选择性或通用扩增可使用任何合适的扩增条件。在一些情况下，扩增可以是线性的。在一些情况下，扩增可以是对数式的。由于本发明的方法列举了一个或多个可被扩增的序列，因此在各步骤中控制扩增以控制样品之间的变异性可能是合适的。

[0175] 例如，在一些情况下，可在选择性或通用扩增步骤中使用有限数目的扩增循环。这可能特别适合于如下的选择性扩增：其中，在使用了多个基因座或条形码的多重条件下，针对不同的基因座或条形码的不同引物组可能表现不同。不同引物组中的引物在其与模板杂交的能力上可能有所不同，从而产生引物组之间在扩增效率上的差异。基于引物和样品 DNA 的序列环境 (context)、缓冲液条件和其他条件，针对给定基因座的每组引物可能表现不同。用于多重分析系统的通用 DNA 扩增通常可引入较少的偏差和可变性。

[0176] 为了使一个或多个基因座或条形码之间的扩增变异最小化，例如，可使用线性扩增方法，继以对数通用扩增来进行扩增。在一些情况下，将循环数限制在 1-50 个循环，使得扩增为线性或接近线性的。在一些情况下，用于线性扩增的扩增循环可为至少约 1、2、3、4、5、6、7、8、9、10、20、30、40 或 50 个循环。在一些情况下，用于线性扩增的扩增循环可为至多

约 1、2、3、4、5、6、7、8、9、10、20、30、40 或 50 个循环。在一些情况下，在从连接产物对序列进行线性选择性扩增之后，可如本文所述进行对数通用扩增步骤。其中可对多个基因座或条形码扩增产物使用共同引物组的通用扩增可进一步降低扩增变异性，同时产生增加的样品量。

[0177] 在其他情况下，可在线性扩增之前使用对数扩增。在一些情况下，用于对数扩增的扩增循环可为至少约 1、2、3、4、5、6、7、8、9、10、20、30、40 或 50 个循环。在一些情况下，用于对数扩增的扩增循环可为至多约 1、2、3、4、5、6、7、8、9、10、20、30、40 或 50 个循环。

[0178] 通常，任何合适数目的引物组可用于扩增。在一些情况下，扩增引物组可以约等于所检测的基因座的数目。在一些情况下，引物组可为至少约 1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175、200、300、400、500、600、700、800、900 或 1000 个引物组。在一些情况下，引物组可为至多约 1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、125、150、175、200、300、400、500、600、700、800、900 或 1000 个引物组。

IV. 测序方法

[0179] 许多序列测定方法与本发明的系统和方法兼容。用于序列测定的示例性方法包括但不限于：基于杂交的方法，例如在 Drmanac 的美国专利号 6,864,052、6,309,824 和 6,401,267 以及 Drmanac 等人的美国专利公开 2005/0191656 中所公开的，其均通过引用而并入；合成测序方法，例如，Nyren 等人的美国专利号 7,648,824、7,459,311 和 6,210,891，Balasubramanian 的美国专利号 7,232,656 和 6,833,246，Quake 的美国专利号 6,911,345，Li 等人，Proc. Natl. Acad. Sci., 100:414-419(2003)；焦磷酸测序，如在 Ronaghi 等人的美国专利号 7,648,824、7,459,311、6,828,100 和 6,210,891 中所描述的；以及基于连接的序列测定方法，例如，Drmanac 等人的美国专利申请号 20100105052，以及 Church 等人的美国专利申请号 20070207482 和 20090018024。

[0180] 可使用以本质上平行的方式测定许多（通常为数千至数十亿个）核酸序列的方法来确定序列信息，其中优选地使用高通量连续过程平行地读出许多序列。这样的方法包括但不限于：焦磷酸测序（例如，由 454Life Sciences, Inc., Branford, Conn. 商品化的）；连接测序（例如，Life Technology, Inc., Carlsbad, Calif. 在 SOLiD™技术中商品化的）；使用修饰的核苷酸的合成测序（例如，由 Illumina, Inc., San Diego, Calif. 在 TruSeq™和 HiSeq ™技术中以及由 Helicos Biosciences Corporation, Cambridge, Mass. 在 HeliScope™中以及由 Pacific Biosciences of California, Inc., Menlo Park, Calif. 在 PacBio RS 中商品化的）；通过离子检测技术的测序（Ion Torrent, Inc., South San Francisco, Calif.）；DNA 纳米球测序（Complete Genomics, Inc., Mountain View, Calif.）；基于纳米孔的测序技术（例如，由 Oxford Nanopore Technologies, LTD, Oxford, UK 开发的）；以及类似的高度平行测序方法。

V. 试剂盒

[0181] 试剂盒可用于使用利用 mPEAR、Rappel、靶向的 Rappel 或 TELA 方法的方法来制备标记的多核苷酸的文库。

[0182] 试剂盒可至少包含提供的 mPEAR、TELA 引物、通用衔接子或其组合（如本文中所定义的），加上提供的至少一种扩增引物，该扩增引物能够与衔接子引物退火并引发延伸产物的合成，当使用衔接子时，该延伸产物会包括与该衔接子连接的任何靶序列。

[0183] 在一些情况下，包含在试剂盒中的衔接子序列的特征如本文其他部分针对本发明的其他方面所述。扩增引物的结构和性质是本领域技术人员所熟知的。与包含在试剂盒中的衔接子一起使用的适当核苷酸序列的合适引物可使用标准的自动化核酸合成设备和本领域中常规使用的试剂容易地制备。试剂盒可包括提供的一种类型的引物或单独提供的两种不同引物（或甚至混合物），例如，适合于在溶液相中和/或在合适的固体支持物上（即固相扩增）对以衔接子序列修饰的模板进行 PCR 或等温扩增的一对扩增引物。试剂盒可包含用于连接至目标样品的双链衔接子，加上至少两种携带不同的标签序列的不同的扩增引物，其中该标签序列不与衔接子杂交。该试剂盒可用于扩增至少两个不同的样品，其中每个样品使用一种标记的引物进行扩增，然后在单独的扩增反应后合并。

[0184] 在试剂盒中，衔接子和/或引物可以以可当即使用的形式提供，或者更优选地，作为在使用前需要稀释的浓缩物提供，或者甚至以在使用前需要重建的冻干或干燥的形式提供。如有需要，试剂盒可进一步包含提供的用于稀释或重建引物的适当的稀释剂。任选地，试剂盒可进一步包含提供的用于进行 PCR 或等温扩增的试剂、缓冲液、酶、dNTP 等。这些试剂的合适（但非限制性）的实例如在所附实施例的“材料与方法”部分中所述。可任选地在试剂盒中提供的其他组分包括适于对使用错配的衔接子和引物制备的模板进行测序的“通用”测序引物。

D. 通过引物延伸对靶向文库分子的捕获 (mPEAR)

[0185] mPEAR 是可用于靶向测序的样品制备技术。其可涉及如图 3 中所示的若干步骤。

i. 片段化

[0186] 可在后续步骤之前首先对多核苷酸样品进行片段化。在本文中描述了片段化方法。在长度方面描述的多核苷酸片段的大小可根据靶多核苷酸的来源、用于片段化的方法和所需的应用而不同。在一些情况下，可使用一个或多个片段化步骤。例如，可使用 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 个或更多个片段化步骤。

ii. 多核苷酸链末端修复

[0187] 例如通过机械剪切或酶消化对核酸的片段化可产生具有平端和 3’- 和 5’- 突出端的异质混合物的片段。在一些实施方案中，可使用本领域已知的方法或试剂盒（即 Lucigen DNA 终止子末端修复试剂盒）修复或处理 DNA 片段，以产生设计用于插入至例如克隆载体的平端位点的末端。可对核酸群体的平端的片段末端进行测序。此外，在一些情况下，也可对平端片段进行磷酸化。可通过酶处理，例如使用激酶（即虾碱性激酶）来引入磷酸部分。可使用磷酸酶对平端片段进行去磷酸化。可使用核酸酶对粘端片段进行修剪。可通过本领域已知的任何方法向平端添加突出端。

[0188] 例如通过某些类型的 DNA 聚合酶（如 Taq 聚合酶或 Klenow exo- 聚合酶）的活性制备具有单个突出的核苷酸的多核苷酸序列，该聚合酶具有非模板依赖性末端转移酶活性，该活性将单个脱氧核苷酸，例如脱氧腺苷 (A) 添加至例如 PCR 产物的 3’ 端。这类酶可用于将单个核苷酸‘A’添加至靶多核苷酸双链体的每条链的平端 3’ 末端。因此，可通过使用 Taq 或 Klenow exo- 聚合酶的反应将‘A’添加至靶多核苷酸双链体的每条末端修复的双链体链的 3’ 末端，而衔接子多核苷酸构建体可以是 T- 构建体，其具有存在于衔接子构建体的每个双链体区的 3’ 末端上的兼容的‘T’突出端。这种末端修饰还防止衔接子和靶标自连接使得存在向形成合并的连接的衔接子 - 靶序列的偏差。

iii. 条形码化

[0189] 可使用标识符序列或分子条形码，360。这些序列提供可在下游应用如测序中鉴定的特定靶分子的来源的特征性标记物。通常，独特的标识符是用于标记靶分子的已知序列的条形码寡核苷酸。mPEAR 方法可包括通过酶反应如连接反应将寡核苷酸条形码附接至核酸靶分子。例如，连接酶可将 DNA 条形码共价附接至片段化的 DNA。

[0190] 用于添加分子条形码的另一示例性方法可包括使用包含用于扩增反应（例如，PCR 或线性扩增等）的条形码序列的寡核苷酸引物。

[0191] 通常，如本文所述，标识符可以是与探针组中的第一或第二探针邻接的寡核苷酸条形码序列。然而，在一些情况下，可使用不同的标识符。标识符，与条形码序列一样，可以是独特的或非独特的。例如，在一些情况下，独特的标识符可以是杂交探针。在一个实例中，杂交探针可包含寡核苷酸序列和另外的组分如荧光元件（即纳米颗粒、纳米探针、量子点等）。在一些情况下，也可将一个或多个荧光元件描述为条形码。例如，可将不同波长或颜色的荧光元件以独特的或非独特的图案或序列排列。在其他情况下，标识符是染料，在这种情况下，附接可包括将染料嵌入到分析物分子（如嵌入到 DNA 或 RNA）中，或者与标记有染料的探针结合。在其他情况下，标识符可以是核酸寡核苷酸，在这种情况下，附接至多核苷酸序列可包括寡核苷酸与序列之间的连接反应或通过 PCR 的引入。在其他情况下，所述反应可包括加入金属同位素，其中第一或第二探针用同位素进行标记。

[0192] 可将独特的标识符（例如，寡核苷酸条形码、探针等）以多种方式附接至多核苷酸序列。条形码可包含不同的长度。在一些情况下，它们可包含约 1、2、3、4、5、6、7、8、9、10、20、50、100、500 或 1,000 个核苷酸的长度。在一些情况下，分子条形码可为少于约 1、2、3、4、5、6、7、8、9、10、20、50、100、500 或 1000 个核苷酸的长度。在一些情况下，可将多个条形码连接至多核苷酸。在一些情况下，可将约 1、2、3、4、5、6、7、8、9、10、20、50、100、500 或 1000 个条形码连接至单个多核苷酸。在一些情况下，多核苷酸可连接有少于约 1、2、3、4、5、6、7、8、9、10、20、50、100、500 或 1000 个条形码。

iv. 已知序列的附接

[0193] 可将已知序列附接至片段化的多核苷酸样品的末端。已知序列可包含分子条形码、用于测序的衔接子或任何其他序列如通用引物序列。通用引物序列可包括，例如，可与引物杂交的已知序列，例如用于 PCR 扩增反应的已知序列。在一些实施方案中，可使用连接反应连接将已知序列共价附加至片段。连接方法可采用连接酶如 DNA 连接酶以附加多核苷酸链（例如片段和已知序列）的末端，从而形成共价键。5’- 磷酸部分可有利于与靶标 3’-OH 的连接。附加可意指先前未共价连接的多核苷酸链的共价连接。在一些实施方案中，附加可涉及在两条多核苷酸链之间形成磷酸二酯键，但也可使用其他共价连接方式（例如非磷酸二酯骨架连接）。已知序列可引入可有利于分离的标志物（例如生物素化的核苷酸，或附接至可通过抗体纯化方法分离的部分的核苷酸）。

[0194] 可将带有邻接的已知序列的片段进行扩增和 / 或纯化。纯化可以基于多核苷酸的大小。纯化可以基于分离技术，例如基于生物素 / 链霉亲和素或基于抗体的分离技术。纯化可意指从连接反应的一种或多种组分如酶、缓冲液、盐等中分离出多核苷酸。合适的纯化方法是本领域已知的，并采用标准方法。可对具有邻接的已知序列的片段进行扩增。扩增可通过包括 PCR 和 / 或线性扩增在内的本领域中已知的任何手段来完成。在扩增步骤中，

可使用具有尾的引物，该尾具有已知的序列，并且可以添加，例如，分子条形码和 / 或衔接子序列。

[0195] 在一些情况下，可由各种序列的组合来形成独特的条形码。在一些情况下，非独特的条形码可与另外的序列（例如探针序列、探针序列的部分或与探针连接的另外的序列）连接，以形成独特的条形码序列。例如，独特的序列在单独使用或与非独特的条形码序列组合使用时，可在探针序列的起始（开始）和末端（终止）部分形成。序列的组合（即探针序列和非独特的条形码序列）可提供独特的标识序列。例如，在一些情况下，可将条形码设计为具有通用结构，5’ XXXXYYYY，其中 X 是可变长度区域，其互补于选自包括但不限于第一和 / 或第二探针 / 杂交序列、衔接子序列、通用引发序列或连接序列的序列中的一个序列。Y 可选自具有可变长度的非独特的条形码序列。在一些情况下，Y 序列可以是样品中所有探针组所共有的。在其他情况下，Y 序列可以是一个基因座或多个基因座例如整个染色体或与特定疾病或基因型相关联的基因座所独特的。在一些情况下，定义为 X 或 Y 的核苷酸的长度或数目可为约 1-20、20-50、50-75、75-100、100-150、150-200、200-300、300-400 或 400-500 个核苷酸。在一些情况下，定义为 X 或 Y 的核苷酸的长度或数目可为至少约 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、75、100、125、150、175、200、300、400 或 500 个核苷酸。在一些情况下，定义为 X 或 Y 的核苷酸的长度或数目可为至多约 1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、75、100、125、150、175、200、300、400 或 500 个核苷酸。

[0196] 在可替代的构造中，条形码可通过至少约 1、2、3、4、5、6、7、8、9 或 10 个序列的组合而形成。条形码可通过至多约 1、2、3、4、5、6、7、8、9 或 10 个非独特的序列的组合而形成。

v. mPEAR 引物

[0197] mPEAR 引物可与已附加至片段的已知序列中的已知序列通用引发位点退火。示例性的 mPEAR 引物图示于图 3 中，350。mPEAR 引物可包含：通用扩增序列、靶基因座特异性序列、分子条形码、间隔区序列和 / 或其他已知序列。通用扩增序列可与附接至片段的通用衔接子序列杂交。间隔区序列可具有可变的长度并且可包含简并核苷酸、已知的核苷酸或其任意组合。基因座特异性序列可与靶基因座 300 杂交，或者可与目标区域 310 的直接上游或下游的区域杂交。可将 mPEAR 引物设计为与目标区域的上游或下游的位点退火。在一些实施方案中，这可允许提高的特异性，例如可避免假基因和具有相似的序列同源性的基因家族，从而减少假阳性。

[0198] 可使用多个 mPEAR 引物来靶向相同或不同的目标区域。可使用两个或更多个靶向相同的目标区域的 mPEAR 引物。可将两个或更多个 mPEAR 引物设计成使其覆盖目标区域。可将两个 mPEAR 引物设计成使其靶向相同的区域，但来自相反的方向。使用多个 mPEAR 引物可允许同时对若干目标区域进行分析。

[0199] mPEAR 引物可与已附加或连接至已知序列的靶片段 DNA 杂交。在一些实施方案中，进行结合的两个分离的区域增加了 mPEAR 引物的特异性。

[0200] 在一些实施方案中，mPEAR 引物的通用 5’ 端可用于改善下游测序。例如，通用 5’ 端可以使朝向 DNA 文库片段的末端的合成寡核苷酸稳定化。通用 5’ 端可提高测序仪的效率。通用 5’端可保持序列，例如锚定序列，朝向读取的起始部分。在下游应用如测序中，在一些情况下，这可允许测序仪适当地定位靶序列，而不浪费测序仪容量。

[0201] 在一些实施方案中，可使用封闭多核苷酸，例如封闭寡核苷酸或封闭寡聚体

(blocking oligos)。在一些实施方案中，引物可以在5'端结合并覆盖通用序列，从而任选地减少或消除对于另外的封闭寡核苷酸的需要。

[0202] 由简并核苷酸组成的间隔区序列可以合成为邻近通用序列。简并核苷酸的数目可以是可变的。简并核苷酸可允许DNA文库的开始和终止位置在DNA测序过程中具有一定的灵活性。这可允许在基因座特异性引发位点的设计中的灵活性。可变序列起始位点的存在可有助于避免测序步骤中的系统误差，并可允许在冗余DNA文库片段的整个读取值中的随机错误分布。

[0203] 可将基因座特异性引发位点设计为识别位于实际靶序列的上游的DNA序列。在不损失功能的情况下可将基因座特异性区域设计成尽可能短。在上下中，'功能'是指双链区在用于酶催化的核酸引物延伸反应的标准反应条件（例如，在4℃至60℃范围的温度下在适于酶的退火缓冲液中进行温育）下形成稳定的双链体，从而使形成衔接子的两条链在引物延伸至靶分子期间仍保持部分退火的能力。

[0204] 可将mPEAR引物区域设计成防止以各种方式自退火。在一些情况下，mPEAR引物可采取导致通用引发位点或基因座特异性位点可比其中另一个更长的形式。在这样的情况下，在其中一条链上存在单链区，或者选择序列以使两条链不杂交，从而形成连续的单链引物。在一些情况下，可设计序列以使它们以'气泡(bubble)'的构型退火，其中mPEAR引物构建体的两个末端均能够彼此杂交并形成双链体，但中心区域不能用于双链体。链中构成中心区域的部分在相同的两条链的其他部分退火而形成一个或多个双链区的条件下并不退火。在一些情况下，较长的mPEAR引物的长度可能与mPEAR能够与自身进行碱基配对的可能性相关。因此，在一些实施方案中，可将长度缩短，以减少这种影响。在一些实施方案中，也可通过包含比标准Watson-Crick碱基对表现出更强的碱基配对的非天然核苷酸来提高稳定性。

[0205] mPEAR引物的实际核苷酸序列可以是任何合适的序列，并可由使用者选择，以使所期望的序列元件最终包含于来源于引物的模板的文库的共同序列中，例如，以为特定通用扩增引物和/或测序引物组提供结合位点。可包含另外的序列元件，例如，以为最终将在文库中模板分子的测序中使用的测序引物或由模板文库的扩增（例如在测序应用中在固体支持物上）获得的产物提供结合位点。

[0206] 通常，mPEAR序列可包含DNA，但也可包含可能合适的任何核苷酸或核苷酸衍生物。可供选择的核苷酸可包括通过磷酸二酯和非磷酸二酯骨架键的混合物连接的天然和非天然核苷酸（例如，一种或多种核糖核苷酸）的混合物。可包含其他非核苷酸修饰，例如，生物素部分、封闭基团和用于附接的捕获部分，如生物素化的核苷酸。

vi.mPEAR延长

[0207] 可将杂交的mPEAR引物延伸以扩增靶片段的全部或一部分。如本文所述，可使用任何合适的聚合酶330进行引物延伸反应。引物延伸反应是本领域公知的，并且可包含用于反应的任何合适的试剂。聚合酶的选择可以基于不同的标准，包括引物延伸的长度、酶的保真度、速度、周转率等。在一些情况下，Klenow或Klenow片段可能适合于引物延伸反应。

[0208] 引物延伸反应370的条件可以变化，并且可使用循环、时间和温度的任意组合来进行一个或多个反应。反应条件通常可基于引物设计的各种参数（包括解链温度、预测的二聚体-二聚体形成、平均延伸长度等）而变化。

[0209] 在一些实施方案中，可在经标志的核苷酸的存在下进行延伸反应。在一些实施方案中，反应混合物中的核苷酸的一些部分采用亲和偶联物如生物素 340 进行标志。延伸可通过使用例如聚合酶的延长反应而发生。延伸反应可产生靶片段的互补物。在一些实施方案中，靶片段的互补物采用生物素化的核苷酸进行标志。所得引物延伸产物可包含模板多核苷酸的文库。在一些情况下，引物本身可被生物素化。在一些实施方案中，可使用链霉亲和素固定化的表面从混合物中的其他片段中分离出模板多核苷酸的文库，其中使用用来结合亲和偶联物的试剂以纯化靶分子，如图 3B 中所示。在一些情况下，固定化的表面可包含链霉亲和素以对偶联的生物素进行亲和纯化。在一些情况下，固定化的表面可以是链霉亲和素涂覆的珠子，其可通过使用磁体 380 进行纯化。在一些情况下，与本领域当前的技术相比，使用亲和偶联物的引物延伸提供了更高效或更具特异性的探针捕获。在一些情况下，当前的方法依赖于通过合成序列与来源于患者的序列之间的氢键产生的亲和力。合成序列通过使用参照基因组进行设计。在一些情况下，来源于患者的序列可与合成序列不同，并且在一些情况下，这限制了结合效率。

vii. 文库

[0210] 可对模板多核苷酸的文库进行扩增（例如通过 PCR 或线性扩增反应）。扩增可使用能在 5’和 3’端与已知序列杂交的引物进行。

[0211] 扩增反应的内容物通常是本领域已知的，并且可包括扩增反应所需的适合的底物（例如 dNTP）、酶（例如 DNA 聚合酶）和缓冲液组分。PCR 扩增反应可能需要通常表示为“正向”和“反向”引物（引物寡核苷酸）的两种扩增引物，其能够与待扩增的多核苷酸序列的一部分在扩增反应的每个循环中的引物退火步骤中所经受的条件下特异性地退火。在某些实施方案中，正向和反向引物可以是相同的。在线性扩增中，可能需要一种引物。

[0212] 在扩增步骤中，可使用针对所有样品的简并或通用引物或具有靶多核苷酸特异性序列（即基因座特异性序列）的正向引物进行扩增。

[0213] 扩增方法的实例包括但不限于：聚合酶链反应 (PCR)（美国专利号 4,683,195 和 4,683,202；PCR Technology:Principles and Applications for DNA Amplification,H. A.Erlich 编著，Freeman Press,NY,N.Y.,1992）、连接酶链反应 (LCR)(Wu 和 Wallace,Genomics4:560,1989；Landegren 等人，Science 241:1077,1988）、链置换扩增 (SDA)（美国专利号 5,270,184 和 5,422,252）、转录介导的扩增 (TMA)（美国专利号 5,399,491）、连锁线性扩增 (LLA)（美国专利号 6,027,923) 以及类似方法，自动维持序列复制 (Guatelli 等人，Proc.Nat.Acad.Sci.USA,87,1874(1990)；和 WO90/06995）、靶多核苷酸序列的选择性扩增（美国专利号 6,410,276）、共有序列引物聚合酶链反应 (CP-PCR)（美国专利号 4,437,975）、任意引物聚合酶链反应 (AP-PCR)（美国专利号 5,413,909、5,861,245) 以及基于核酸的序列扩增 (NASBA)。（参见，美国专利号 5,409,818、5,554,517 和 6,063,603，各专利均通过引用并入本文）。其他可使用的扩增方法包括：Qβ 复制酶（描述于 PCT 专利申请号 PCT/US87/00880 中）；等温扩增方法，例如 SDA（描述于 Walker 等人，Nucleic Acids Res.20(7):1691-6(1992) 中）；以及滚环扩增（描述于美国专利号 5,648,245 中）。其他可使用的扩增方法描述于美国专利号 5,242,794、5,494,810、4,988,617 和美国序列号 09/854,317 以及美国公开号 20030143599 中，各自均通过引用并入本文。在一些方面，通过多重基因座特异性 PCR 对 DNA 进行扩增。在一些方面，使用衔接

子-连接和单引物PCR对DNA进行扩增。其他可用的扩增方法，如平衡PCR(Makrigiorgos等人，Nature Biotech,20:936-9(2002))和等温扩增方法，如基于核酸序列的扩增(NASBA)，和自动维持序列复制(Guatelli等人，PNAS USA 87:1874(1990))。基于这些方法，本领域技术人员能够容易地设计出位于目标基因座的5'和3'侧的任何适当区域中的引物。这样的引物可用于扩增任意长度的DNA，只要它在其序列中包含目标基因座。

[0214] 可使用多于两种的扩增引物进行扩增反应。为防止mPEAR二聚体的扩增，可对扩增引物进行修饰，以使之包含在整个引物延伸产物上杂交并向靶分子模板(或连接至其3'端的dNTP)内杂交的核苷酸。可对第一扩增引物进行修饰和处理，以有助于防止链的外切核酸酶消化。通用的第一扩增引物可扩增所有的样品，而非单独地修饰和处理每个标记的引物。标记的引物可作为样品特异性第三引物而引入扩增反应中，但并不需要进行特别的修饰和处理来降低外切核酸酶消化。携带标签的第三扩增引物可包含与第一扩增引物的至少一部分相同的序列，以使其可用于扩增由第一扩增引物的延伸而产生的双链体。

[0215] 可在靶分子模板的一条或两条链上进行引物延伸。引物延伸和随后的扩增可运行穿过DNA文库分子的末端。在DNA的两条链上均使用酶促引物延伸可能是有利的。靶向同一序列的两个反应可增加特异性，并可降低失败率。

[0216] 扩增引物可为不同的长度。在巢式PCR的情况下，可将三种或更多种扩增引物设计为比用于扩增先前的扩增子的引物更长，因此所添加的核苷酸的长度是完全可控的，并且在需要时可为数百个核苷酸。

[0217] 正向和反向引物可具有足够的长度，以与整个通用衔接子序列和靶序列的至少一个碱基(或添加至靶链上作为3'-突出端的核苷酸dNTP)杂交。正向和反向引物还可包含可延伸超出衔接子构建体的区域。在一些实施方案中，扩增引物可为至少10、20、30、40、50、60、70、80、90、100、150、200、300、400或500个碱基的长度。在其他实施方案中，扩增引物可为至多10、20、30、40、50、60、70、80、90、100、150、200、300、400或500个碱基的长度。正向和反向引物可具有显著不同的长度。在一些实施方案中，第一引物可为20-40个碱基，而第二引物可为40-100个碱基的长度。可选择正向和反向引物的衔接子-靶标特异性部分的核苷酸序列，以实现与待扩增的衔接子-靶标序列在扩增反应的退火步骤的条件下的特异性杂交，同时最小化与存在的任何其他靶序列的非特异性杂交。

[0218] 扩增引物通常为单链多核苷酸结构。它们可包含天然和非天然碱基以及天然和非天然骨架键的混合物。

[0219] 引物可包含对各平台上的捕获具有特异性的序列。在一些情况下，可引入序列，以允许与各种高通量平台试剂盒，例如由Illumina提供的试剂盒中的已知序列杂交。将用于平行加样的杂交序列引入至用于测序的表面平台上是本领域已知的。

[0220] 引物可包含非核苷酸化学修饰，例如硫代磷酸酯，以增加外切核酸酶抗性，另一方面其条件是该修饰不阻碍引物功能。例如，修饰可以促进引物与固体支持物(例如生物素部分)的附接。某些修饰本身可改善分子作为引物的功能，或者可提供一些其他有用的功能，例如提供切割位点，以使引物(或由其衍生的延伸的多核苷酸链)能够被切割。

[0221] 可以对合并或未合并的样品进行扩增。标签可以是扩增引物的一部分。在一些实施方案中，可以在合并之前独立地扩增每个样品。可以对合并的核酸样品进行处理以供测序。

[0222] 扩增步骤可用于产生大量较高质量的样品。扩增步骤可以用来向靶多核苷酸序列引入另外的条形码或衔接子序列。

viii. 样品捕获

[0223] 可以捕获合并或未合并的样品，以准备测序。可以作为单一捕获的靶标的阵列进行测序。可以将扩增产物附接在平面表面上，或在珠子的集合体上。珠子的集合体可以分离成乳液，在该乳液的每个“分区”中含有单个珠子。在每个“分区”仅一个模板的浓度下，仅有一个模板在每个珠子上得到扩增。在一些实施方案中，mPEAR 靶向方法可以与基因组 RAPELLing 方法结合使用。

[0224] 用于捕获扩增产物的任何化学手段可以是适合的。在一些实施方案中，以下方式对于捕获来说是合适的：单点共价附接至固体支持物上引物的 5’ 端处或附近，使引物的模板特异性部分可自由地与其同源模板退火，并且 3’羟基基团可自由地进行引物延伸。本领域中已知的任何合适的共价附接手段可以用于此目的。所选择的附接化学将取决于固体支持物的性质，以及应用到其上的任何衍生物或官能化。引物本身可包含可能是非核苷酸化学修饰的部分以促进连接。在一个具体的实施方案中，引物可在 5’ 端包含含硫的亲核体，如硫代磷酸酯 (phosphorothioate) 或硫代磷酸酯 (thiophosphate)。

[0225] 在其他实施方案中，可以通过生物素 - 链霉亲和素或链霉亲和素衍生物的相互作用而实现捕获。包含生物素的扩增产物可以与带有链霉亲和素或链霉亲和素衍生物的表面进行温育，从而使产物得以固定化，如图 3B 所示。本领域已知的纯化方法可用于回收 (retrieve) 扩增产物，例如通过使用磁体和 / 或链霉亲和素涂覆的珠子。在一些实施方案中，可以使用额外的洗涤步骤。如图 3C 所示，可以通过用过量浓度的生物素或生物素相关化合物冲洗来洗脱扩增产物。本领域已知的用于多核苷酸的生物素 - 链霉亲和素亲和纯化的方法 (参见 US5405746、US5500356 和 US5759778) 通过引用以其整体并入本文。

ix. 测序与数据分析

[0226] 可以通过本领域已知的任何方法对 mPEAR 样品进行测序，在本文中公开了若干非限制性实例。测序可以生成数据。如本文所公开的，可以将数据存储、处理和传送。

[0227] 如图 9 所示的捕获的分子的读取值结构可以在数据分析的速度和质量上提供显著的优点。由于基因座特异性引物识别位点可以被设计为与靶序列的上游杂交，该基因座特异性位点可以用于鉴定读取值的基因组位置。通常可将此称为“锚式读取值比对”，其可以大大减少数据处理和统计分析的量。在一个非限制性的实例中，并非采用短序列读取值以及在计算机上将其与整个参照基因组进行比对，而是读取值可被自动分箱或与其正确的基因组位置相关联。这可以大大缩短数据分析的时间，提高准确度，并降低这类分析所需的计算能力。

[0228] 此外，mPEAR 方法可以允许多核苷酸靶标的真正的从头测序。因此，通过用已知序列“锚定”读取值，就可以不使用参照基因组而组装延伸产物。从而，mPEAR 允许对于与参照基因组极大不同的基因组基因座的更大灵敏度。临床相关的较长插入、较大缺失以及重复能够以更高的灵敏度和准确度得到检测。病毒插入位点和 / 或多核苷酸可动因子可以被检测和定位，并且该过程可以多重化以具有更高的效率。

[0229] 一个或多个条形码的任选的添加可允许，例如，样品对序列读取值的分箱。条形码化可用于多种应用，包括单独的多核苷酸分子的示踪，如图 13 中所示。在一些实施方案中，

DNA 分子读取值可以与样品相关联。在一些实施方案中，mPEAR 条形码化可用于对信息进行定相 (phase)，其中单个分子可被鉴定为父本或母本遗传自单个个体。

[0230] E. 单末端衔接子文库与滚环扩增

[0231] 滚环扩增可与靶向方法一起使用。滚环扩增可用来产生线性扩增反应。

i. 片段化

[0232] 如图 12 所示，在后续步骤之前，可以首先对多核苷酸样品 1200 进行片段化，1210。片段化方法已在本文中描述。在长度方面描述的多核苷酸片段的大小可根据靶多核苷酸的来源、用于片段化的方法和所需的应用而不同。在一些情况下，可使用一个或多个片段化步骤。例如，可使用 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15 个或更多个片段化步骤。

[0233] 在一些情况下，可以将核酸片段化为至少 10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、5000 个碱基对长度的大小。在某些情况下，可以将核酸片段化为最多 10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、5000 个碱基对长度的大小。

[0234] 多种片段化方法在本文中描述并且是本领域已知的。例如，片段化可通过物理、机械或酶法进行。物理片段化可包括将靶多核苷酸暴露于热或紫外线。机械破碎可用于将靶多核苷酸机械剪切成所需范围的片段。机械剪切可通过包括靶多核苷酸的反复移液、声处理和雾化在内的本领域已知的多种方法来完成。靶多核苷酸也可使用酶法进行片段化。在一些情况下，可使用酶例如使用限制酶进行酶消化。

[0235] 限制酶可用于进行靶多核苷酸的特异性或非特异性片段化。本发明的方法可使用一种或多种类型的限制酶，通常描述为 I 型酶、II 型酶和 / 或 III 型酶。II 型酶和 III 型酶通常可商购获得并且是本领域公知的。II 型酶和 III 型酶识别双链多核苷酸序列内的核苷酸碱基对的特定序列（“识别序列”或“识别位点”）。一旦结合并识别这些序列，II 型酶和 III 型酶即切割多核苷酸序列。在一些情况下，切割将产生具有突出的单链 DNA 的一部分（称为“粘端”）的多核苷酸片段。在其他情况下，切割将不会产生具有突出端的片段，而是形成“平端”。本发明的方法可包括使用产生粘端或平端的限制酶。

[0236] 限制酶可识别靶多核苷酸中的多个识别位点。一些限制酶（“精确的切割酶”）仅识别单个识别位点（例如，GAATTC）。其他的限制酶更不加区分，识别多于一个识别位点或者多个识别位点。一些酶在识别位点中的单一位置进行切割，而其他酶可在多个位置进行切割。一些酶在识别位点中的相同位置进行切割，而其他酶在可变的位置进行切割。

ii. 核酸链末端修复

[0237] 在许多情况下，例如通过机械剪切或酶消化对核酸的片段化产生具有平端和 3’- 和 5’- 突出端的异质混合物的片段。在一些情况下，本发明的组合物和方法提供了使用本领域中已知的方法或试剂盒（即 Lucigen DNA 终止子末端修复试剂盒）对片段末端的修复，以产生设计用于插入至例如克隆载体的平端位点的末端。在一些情况下，本发明的组合物和方法提供了经测序的 DNA 群体的平端片段末端。此外，在一些情况下，也可对平端片段进行磷酸化。可通过酶处理，例如使用激酶（即虾碱性激酶）引入磷酸部分。

[0238] 在其他情况下，例如通过某些类型的 DNA 聚合酶（如 Taq 聚合酶或 Klenow exo- 聚合酶）的活性制备具有单个突出的核苷酸的多核苷酸序列，该聚合酶具有非模板依赖性末端转移酶活性，该活性将单个脱氧核苷酸，例如脱氧腺苷 (A) 添加至例如 PCR 产物的 3’端。

这类酶可用于将单个核苷酸‘A’添加至靶多核苷酸双链体的每条链的平端 3’末端。因此，可通过与 Taq 或 Klenow exo- 聚合酶的反应将‘A’添加至靶多核苷酸双链体的每条末端修复的双链体链的 3’末端，而衔接子多核苷酸构建体可以是 T- 构建体，其具有存在于衔接子构建体的每个双链体区的 3’末端上的兼容的‘T’突出端。这种末端修饰还防止衔接子和靶标自连接使得存在向形成合并的连接的衔接子 - 靶序列的偏差。

[0239] 在一些情况下，可以使用 Nextera 试剂盒，例如 Illumina/Epicentre 提供的 Nextera 试剂盒，其使用 tn5 转座酶同时进行双链 DNA 的片段化和衔接子与片段末端的连接。例如，在含有 0.25μl 转座酶和 4μl 5×HMW 的 Nextera 反应缓冲液（包含 Illumina 兼容的衔接子）的 20μl 反应中，在 55℃下将扩增的 cDNA “标记片段化 (tagmentated)”5 分钟。为了将转座酶从 DNA 上剥离，接着将 35μl PB 加入标记片段化反应混合物中，并且使用 88μl SPRI XP 珠子（样品与珠子之比为 1 :1.6）纯化标记片段化的 DNA。用于该方法的试剂可在 Nextera DNA 样品试剂盒 (Epicentre/Illumina) 中获得。也可使用可替代的试剂盒，例如 Roche FLX 和 Titanium 测序系统提供的试剂盒。

[0240] 在一些情况下，可以不进行 cDNA 的片段化。而是在 RNA 分子反转录成 cDNA 之前，可以使用任何合适的方法，包括本文所述的以及 Hashimony 等人，Hashimshony, 2012 所描述的适用技术，将其片段化。

[0241] 在一些情况下，使用琼脂糖凝胶方法如 SizeSelect™凝胶 (Life Technologies) 或 Pippin Prep™试剂盒或珠子如 AMPure XP(Beckman Coulter) 对片段化的 DNA 进行大小选择。在其他实施方案中，对片段化的 DNA 进行末端修复或多核苷酸加尾，以备文库制备的后续步骤使用。

[0242] 例如通过机械剪切或酶消化对多核苷酸的片段化可产生具有平端和 3’- 和 5’- 突出端的异质混合物的片段，1230。在某些情况下，可对片段末端进行修复或处理，1220。这类用于多核苷酸链末端修复的方法已经在本文中描述。

iii. 已知序列的附接

[0243] 已知序列可以附接至片段化的多核苷酸样品的末端。已知序列的附接方法（例如连接）已经在本文中描述。

[0244] 已知序列 1235 可以包含分子条形码、一个或多个衔接子或任何其他序列，例如通用引物序列。在一些实施方案中，已知序列包含两个衔接子 :A 衔接子 1240 和 B 衔接子 1245，它们可以以“反向的”方式就位。“反向的”方式可以意味着 A 衔接子的 5’端可连接至 B 衔接子的 5’端。在一些实施方案中，可以在 A 和 B 衔接子之间工程构建限制酶或其他核酸酶位点。

iv. 连接

[0245] 附接至已知序列的片段化的多核苷酸可以环化。可以通过连接完成环化，1250。在一些实施方案中，将片段的 5’端连接到已知序列的 3’端。在其他实施方案中，将片段的 3’端连接到已知序列的 5’端。

v. 扩增

[0246] 环化的多核苷酸可以通过滚环扩增过程进行扩增。在这样的过程中，引物 1255 可与环化的多核苷酸 1260 杂交。聚合酶可以延长和拷贝环化的模板，即滚环扩增。聚合酶可以拷贝环化的模板 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、

50、60、70、80、90、100、200、300、400、500 次或更多次。这可能会产生包含一个或多个连续线性拷贝的拷贝，1265。在一些实施方案中，线性多核苷酸中可以存在 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、40、50、60、70、80、90、100、200、300、400、500 个拷贝。这些拷贝可以被限制酶或核酸酶切割，以产生用于测序的文库。在一些实施方案中，可以利用 PCR 扩增方法将全长衔接子序列添加到 5’和 3’端，由此产生用于测序的文库。本文公开了使用 PCR 引物将已知序列添加到扩增产物上的方法。

[0247] 在一些情况下，可以使用覆盖式基因座特异性引物扩增含有目标序列的环状分子。引物可被设计成位于靶标的上游并覆盖靶标的全长。引物也可被设计为处于两个方向上。双链 DNA “环”可以变性并与基因座特异性引物结合（双向）。当发生滚环扩增时，扩增产物可以从模板分子中置换。过量引物可以与生长链结合并在交替方向上扩增延伸的拷贝以形成具有重复序列的长双链 DNA 分子。该重复序列可包含位于靶序列侧翼的 A 和 B 衔接子序列。其现在可以作为 PCR 模板，用于在全长测序衔接子中进行扩增。在其他情况下，当分子的末端连接时，衔接子间的限制位点成为活性的。

vi. 测序和数据分析

[0248] 滚环扩增样品可以通过本领域已知的任何方法进行测序；本文公开了若干非限制性实例。测序可以产生数据。该数据可以如本文所公开的那样存储、处理和传送。

II. 用于从头测序和标记的 DNA 标记

A. 随机引物延伸、连接和标记 (RAPELL)

[0249] 随机引物延伸、连接和标记 (RAPELL) 法可用于使用短读取系统获取核苷酸长片段的序列信息。RAPELL 法可以包括：扩增前的多核苷酸分离、纯化、稀释和空间分离，及分子标记的引入以及最终测序。图 1 描述了一种根据本发明的示例性方法。图 1A 示出了使用短读取系统，使用如本文所述的 mPEAR 系统获取核苷酸的长片段的序列信息的过程 100。该过程 100 包括获取高分子量（超过 5 千碱基）的长核酸 105 的样品。图 1B 示出了用于进一步处理高分子量核酸的过程。将长核酸 105 以亚基因组的量稀释并空间分离 110 至若干分区（例如，115、120、125）。各个分区 115、120、125 可包含长核酸 105。在各个分区中，长核酸片段 105 与含有引物 130 和衔接子 135 的聚合酶 / 连接酶混合物接触。引物 130 可以在 3’端包含随机序列以允许沿长核酸片段 105 的随机结合，并包含已知序列的区域（以圆圈和虚线表示）和 3’帽（以菱形表示）。聚合酶反应 140 沿模板核酸的多个随机区域延伸引物 130，直到延长产物（例如，142、144）到达下游衔接子 135。连接酶 150 将延长产物连接 155 到下游衔接子上并创建扩增子文库（例如，146、148）。可以加入 160 第二组引物（例如，165、170）用于聚合酶链反应 (PCR)175，从而产生适合于测序的文库 180。

i. 样品获得

[0250] RAPELL 法可以使用长分子长度的多核苷酸。样品可以来自核酸文库，如 cDNA 文库。样品可以来自基因组 DNA。核酸还可以从一个或多个受试者中分离。在一些示例性方法中，核酸是高分子量的脱氧核酸 (DNA)。高分子量可以指，例如约大于 0.5、1、3、4、5、10、15、20、30、35、40、50、60、70、80、90、100、200 或更多的千碱基。核酸样品可以包含大小范围约为 1-50kb、5-50kb、5-100kb、20-90kb、50-100kb、5-200kb 的核酸片段。核酸可以主要从细胞成分中纯化。核酸样品可以是染色体形式。在一些情况下，多核苷酸可以片段化成较小的大小。用于多核苷酸片段化的方法在本文中已经公开。在一些情况下，可以利用本领

域已知的方法（例如，物理剪切或酶消化）将染色体核酸片段化成诸如约 1、3、4、5、10、15、20、30、35、40、50、60、70、80、90、100、200kb 的大小，或 0-50kb、5-50kb、5-100kb、20-90kb、50-100kb、5-200kb 范围的大小。

ii. 稀释和空间分离

[0251] 本发明使用样品核酸的稀释和空间分离。在一些情况下，核酸的长片段在空间分离之前进行稀释。稀释可以通过本领域已知的任何方法，例如通过添加稀释剂如水或适当的缓冲液来完成。稀释的示例性方法包括在稀释前确定核酸的浓度，并计算要加入多少稀释剂才能使稀释的样品可以被分配成含有亚基因组量的 DNA 的量（即，使一个样品包含小于一个完整的基因组）。在另一个示例性方法中，可以计算使得样品可以以每个分区包含 约 1、2、3、5、10、20、50、80、100、150、200、400、500、1000、1500、5,000、10,000、100,000、1,000,000、10,000,000、100,000,000、1,000,000,000 个核酸片段的方式进行分配的稀释度。在另一个示例性方法中，进行稀释以便于样品的分配，从而可以分离基因组的一个拷贝的约 1%、2%、5%、10%、15%、20%、30%、50%、70%、80%、95%或 100%。

[0252] 空间分离可以通过本领域已知的多种方法如移液、微量移液或微流体技术来完成。分区可以由本领域已知的方法制成，该方法包括板（例如，96- 孔）、微流体室、微滴或固体表面如硅芯片或珠上的简单空间分离。

[0253] 在一个示例性方法中，进行稀释和空间分离，以使得两个分区包含来自每个亲本染色体的 DNA 的相同基因座的概率较低，或者来自相同的基因组基因座的多个片段将极为罕见。

iii. 引物延伸、连接和扩增

[0254] 在一个分区中，稀释的多核苷酸可与合成寡核苷酸、天然 dNTP、聚合酶（或聚合酶片段）、连接酶和足以进行引物延伸和连接的相关缓冲液的混合物接触。合成寡核苷酸的所述混合物包含供体引物和受体探针。

a. 供体引物

[0255] 本发明可利用供体引物来产生稀释的模板的互补区域。图 22A 是说明了示例性供体引物 200 的图示。供体引物可以包括：衔接子序列 205、分子标记（即条形码）210 和随机引物 215 的区域。供体引物可由包含 DNA 核苷酸、RNA 核苷酸或其任意组合的核苷酸组成。

[0256] 衔接子序列 205 可以位于供体引物 200 的 5’端附近。衔接子序列可以具有约 4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、75、80、85、90、95、100、125、130、150、200 个或更多个核苷酸的长度。在一些情况下，衔接子序列 205 可以被设计成具有这样的序列，使得二级结构可以形成如发夹或茎环。为了促进二级结构的形成和释放，可以加入一个或多个尿嘧啶碱基。在一些情况下可以使用能够形成二级结构的衔接子，以减少供体引物与 RAPELL 产物的结合。可以设计二级结构以使其能够通过切割而选择性地除去。发夹的选择性切割可以通过使用酶如核酸酶例如 Drosha 来实现。

[0257] 分子条形码 210 可被设计为指定哪个分区中发生反应。因此，分子条形码的数目可以等于用于该反应的分区数。在一个非限制性实例中，如果将核苷酸片段分离到 96 孔板分区中，则可以使用 96 种不同的供体引物，各自具有不同的分子标记。该分子标记可以是 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 个或更多个核苷酸的长度。

[0258] 供体引物的3'端可包含随机序列，该随机序列可以充当随机引物。随机引物可以是能够与模板杂交并引发反应的核酸的短区段。随机区域可以包含6、7、8、9、10个或更多个核苷酸。在一个非限制性实例中，该区域可以是6个核苷酸长（即六聚体），因此引物的混合物可以包含碱基的每一种可能的组合（总共4^6= 4096种可能的组合）。在另一个非限制性实例中，该区域可以是8个核苷酸长（即八聚体），而供体引物的混合物可以包含碱基的每一种可能的组合（总共4^8= 65563种可能的组合）。

b. 受体探针

[0259] 本发明可以使用能够与模板结合的受体探针240。图2B是说明了示例性受体探针240的图示。受体探针可以包含：衔接子序列230、分子条形码225和随机引物220的区域。受体探针可以由包含DNA核苷酸、RNA核苷酸或其任意组合的核苷酸制成。受体探针240可以在最3'的核苷酸上包含3'帽235，以阻止从受体探针240延伸或延长。

[0260] 受体探针240的5'端包含随机引物230。随机引物230可以是由碱基的每一个可能的组合组成的核酸的短区段。随机区域可以包含6、7、8、9、10个或更多个核苷酸。在一个非限制性实例中，该区域可以是6个核苷酸长（即六聚体），因此引物的混合物可以包含碱基的每一种可能的组合（总共4^6= 4096种可能的组合）。在另一个非限制性实例中，该区域可以是8个核苷酸长（即八聚体），而受体探针的混合物可以包含碱基的每一种可能的组合（总共4^8= 65563种可能的组合）。

[0261] 分子条形码210可被设计为指定哪个分区中发生反应。因此，各个受体探针300所需要的分子条形码310的数目可以等于用于该反应的分区数。在一个非限制性实例中，如果将核苷酸片段分离到96孔板分区中，则可以使用96种不同的供体引物，各自具有不同的分子标记。该分子标记可以是1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个核苷酸的长度。具有相同分子标记的衔接子探针240和供体引物200可以配对并在同一个分区中使用。在一些实施方案中，衔接子探针300，240和供体引物200将会配对，使得它们的分子标记是不同的并且在同一个分区中使用。

[0262] 衔接子序列230可以位于衔接子探针300的3'端附近。衔接子序列可以具有约4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、45、50、55、60、75、80、85、90、95、100、125、130、150、200个或更多个核苷酸的长度。在一些情况下，衔接子序列230可以被设计成具有这样的序列，使得二级结构可以形成如发夹或茎环。为了促进二级结构的形成和释放，可以加入一个或多个尿嘧啶碱基。在一些情况下可以使用能够形成二级结构的衔接子，以减少衔接子探针240与RAPELL产物的结合。可以设计二级结构以使其能够通过切割而选择性地除去。发夹的选择性切割可以通过使用酶如Drosha来实现。

iv. 结合和延长

[0263] 图4示出了供体引物405和受体探针410的结合和延长400的示例性过程。受体探针410和供体引物405的随机引物区域可以沿模板与区域结合。由于供体引物405和受体探针410的混合物可以具有，例如，六聚体或八聚体碱基的每种组合，因此受体探针和供体引物可以在整个模板415中以能够产生统计学上随机的RAPELL产物430的方式退火。一旦供体引物405和受体探针410已经退火，那么非链置换聚合酶420就可以延伸供体引物405直到其到达受体探针410。非链置换聚合酶可用于充分延伸。一旦延伸已经到达受体

探针 410，那么连接酶 425 就可用于连接延伸产物和受体探针 410 以生成 RAPELL 产物 430。结合和延长方法可以在同一个分区中重复，以产生多个 RAPELL 产物 430。在一些情况下，多个 RAPELL 产物可以包含模板片段的数百、数千或数百万个短拷贝。所得 RAPELL 产物 430 可以是可变长度的并且可以通过解链从模板上释放下来。

[0264] 在一些实例中，可以使用预扩增步骤。例如通过将衔接子与长 DNA 片段的末端连接，空间分离该片段，然后进行大范围 PCR。然后如本文所述使用随机 RAPELL 引物。也可以通过使用先前描述的多重 PCR 预扩增进行靶向大范围 PCR。

v. RAPELL 产物

[0265] 本文所公开的方法可以产生 RAPELL 产物 430。RAPELL 产物可以包含：供体引物 405 和受体探针 410。RAPELL 产物可以在 5’和 3’端上具有衔接子区域。在一些情况下，例如，如果已在 5’或 3’端上产生二级结构，则可以加工 RAPELL 产物。二级结构可通过切割来选择性地去除。二级结构如发夹的选择性切割可以通过使用酶如 Drosha 来完成。

vi. 扩增

[0266] 图 5 示出了如何能够通过聚合酶链反应 (PCR) 扩增 RAPELL 产物 530 的示例性方法 500。为了扩增 RAPELL 产物 530，可以使用与衔接子序列互补（例如通过供体引物 505 和受体探针 510 整合至 RAPELL 产物中）的引物（例如，测序仪引物 A 515 和测序仪引物 B520)。在一些情况下，可以扩增 RAPELL 产物 530，以产生用于测序文库 525 的多个 RAPELL 产物。多个 RAPELL 产物 530 可以在使用的同一反应室中进行扩增。在一些实施方案中，可以将存在于独立的分区中的 RAPELL 产物 530（见图 1）合并，而不是在各个单独的分区中运行。

[0267] 测序仪引物（例如，测序仪引物 A 515 和测序仪引物 B 520）可以包含：能够与衔接子序列（例如，505 和 510）结合的第一区域和适合于特定测序平台衔接子序列（如 Illumina 序列）的区域。可以优化或调整扩增反应的循环条件，以产生期望的片段大小或对于测序仪性能任选的片段大小范围。大小选择和定量可用于实现最佳的测序仪性能。

vii. 测序和数据分析

[0268] 每个拷贝可以包含对反应的单个分区具有特异性的标识符。每个测序仪读取值可以被锚定到单个分区并因此可以拴系到较长的模板片段上。长读取长度通过组合来自相同分子条形码（即同一分区）的重叠序列而实现。

III. 测序

[0269] 可以使用众多的序列测定方法。示例性的序列测定方法包括但不限于：基于杂交的方法，例如在通过引用而并入的 Drmanac 的美国专利号 6,864,052、6,309,824 和 6,401,267 以及 Drmanac 等人的美国专利公开 2005/0191656 中所公开的方法；合成测序方法，例如 Nyren 等人的美国专利号 7,648,824、7,459,311 和 6,210,891，Balasubramanian 的美国专利号 7,232,656 和 6,833,246，Quake 的美国专利号 6,911,345，Li 等人，Proc. Natl. Acad. Sci.，100:414-419(2003)；如 Ronaghi 等人的美国专利号 7,648,824、7,459,311、6,828,100 和 6,210,891 中所述的焦磷酸测序；以及基于连接的测序方法，例如，Drmanac 等人的美国专利申请号 20100105052 和 Church 等人的美国专利申请号 20070207482 和 20090018024。

[0270] 序列测定也可使用以本质上平行方式确定许多（通常为数千至数十亿个）核酸

序列的方法来进行，其中优选地使用高通量连续过程平行地读出许多序列。这样的方法包括但不限于：焦磷酸测序（例如，由 454Life Sciences, Inc., Branford, Conn. 商品化的）；连接测序（例如，Life Technology, Inc., Carlsbad, Calif. 在 SOLiD™技术中商品化的）；使用修饰的核苷酸的合成测序（例如，由 Illumina, Inc., San Diego, Calif. 在 TruSeq™和 HiSeq ™技术中以及由 Helicos Biosciences Corporation, Cambridge, Mass. 在 HeliScope™中以及由 Pacific Biosciences of California, Inc., Menlo Park, Calif. 在 PacBio RS 中商品化的）；通过离子检测技术的测序（Ion Torrent, Inc., South San Francisco, Calif.）；DNA 纳米球测序（Complete Genomics, Inc., Mountain View, Calif.）；基于纳米孔的测序技术（例如，由 Oxford Nanopore Technologies, LTD, Oxford, UK 开发的）；以及类似的高度平行测序方法。

[0271] 一些测序方法需要衔接子序列，有时被称为“衔接子”或“序列衔接子”。衔接子序列可以是平台特异性的。衔接子可以包含锚。衔接子可以包含测序序列。衔接子可以包含扩增序列。在一些实施方案中，衔接子序列可包含锚、测序序列以及扩增序列。衔接子序列可被添加到 5’端。衔接子序列可被添加到 3’端。衔接子序列可以既被添加到 3’端又被添加到 5’端。衔接子序列可以有助于测序。

IV. 用于数据传送和存储的系统

[0272] 本发明的另一个方面提供了一种被配置成实施本发明的方法的系统。该系统可以包括被编程为执行本文所述方法的计算机服务器（“服务器”）。图 11 示出了一种适于使用户能够存储、分析和处理序列信息的系统。该系统包括被编程为执行本文中所述的示例性方法的中央计算机服务器。该服务器包括：中央处理器（CPU，也被称为“处理器”），其可以是单核处理器、多核处理器或用于平行处理的多个处理器。该服务器还包括存储器（例如，随机存取存储器、只读存储器、闪速存储器）；电子存储单元（例如硬盘）；用于与一个或多个其他系统进行通信的通信接口（例如，网络适配器）；和外围设备，可包括高速缓冲存储器、其他存储器、数据存储和 / 或电子显示适配器。存储器、存储单元、接口和外围设备可以通过通信总线（实线）如主板与处理器通信。存储单元可以是用于存储数据的数据存储单元。服务器在通信接口的帮助下可操作地连接到计算机网络（“网络”）。该网络可以是因特网、内联网和 / 或外联网，与因特网通信的内联网和 / 或外联网，电信或数据网络。在某些情况下，网络在服务器的帮助下可以实现对等网络，它可以使得连接到服务器的装置能够作为客户端或服务器运行。在一些实施方案中，计算资源可以被配置成云服务模型。

[0273] 存储单元可以存储文件，如序列数据、样品数据、分子条形码、软件或者与本发明相关联的数据的任何方面。数据存储单元可以与数据耦合，该数据可以利用分子条形码中含有的样品来源或其他信息对样品序列进行分箱。

[0274] 服务器可以通过网络与一个或多个远程计算机系统进行通信。所述一个或多个远程计算机系统可以是，例如，个人计算机、膝上型计算机、平板电脑、电话机、智能电话机或个人数字助理。例如，远程计算机系统可以用于将患者数据传送到照护者。该数据或硬件或系统例如可以被加密或修改（例如，以符合 HIPPA 规则和标准）。

[0275] 在一些情况下，该系统包括一台服务器。在其他情况下，该系统包括通过内联网、外联网和 / 或因特网彼此通信的多个服务器。

[0276] 服务器可以适于存储样品信息，例如，样品来源、日期、取向、序列、统计数据或者

可能相关的任何其他信息。这类信息可以存储在存储单元或服务器上，并且这类数据可以通过网络传送。

[0277] 本文所述的方法可以通过存储在服务器的电子存储位置上（例如存储器或电子存储单元中）的机器（或计算机处理器）可执行代码（或软件）来执行。在使用过程中，该代码可以由处理器来执行。在一些情况下，该代码可以从存储单元中取回并存储在存储器中以备处理器访问。在一些情况下，可以不包括电子存储单元，而机器可执行指令被存储在存储器中。或者，该代码可以在第二计算机系统中执行。

[0278] 本文所提供的系统和方法的各个方面，如服务器，可体现在编程中。该技术的各个方面可被认为是“产品”或“制品”，其形式通常为携载或体现在某种类型的机器可读介质中的机器（或处理器）可执行代码和 / 或关联数据。机器可执行代码可以存储在诸如存储器（例如，只读存储器、随机存取存储器、闪速存储器）或硬盘的电子存储单元中。“存储”型介质可包括计算机、处理器等或其关联模块的任何或所有的有形存储器，诸如可在任何时间为软件程序提供非暂时性存储的各种半导体存储器、磁带驱动器、磁盘驱动器等。软件的全部或部分可不时通过因特网或各种其他电信网络通信。这样的通信可例如实现软件从一台计算机或处理器向另一计算机或处理器中的加载，举例而言，从管理服务器或主计算机向应用服务器的计算机平台中的加载。因此，可承载软件元件的另一类介质包括光波、电波和电磁波，诸如跨本地设备之间的物理接口、通过有线网络和陆线光网络以及通过各种空中链路使用的光波、电波和电磁波。携载此类波的物理元件，诸如有线链路或无线链路、光链路等，亦可被认为是承载软件的介质。如本文中所使用的，除非限于非暂时性、有形“存储”介质，否则诸如计算机或机器“可读介质”等术语是指参与向处理器提供用于执行的指令的任何介质。

[0279] 因此，机器可读介质如计算机可执行代码可采取多种形式，包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质可包括例如光盘或磁盘，如任何计算机等中的任何存储装置等，这些可用于实施该系统。有形传输介质包括：同轴电缆、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) 数据通信期间生成的那些信号。计算机可读介质的常见形式因此包括，例如：软盘、柔性盘、硬盘、磁带、任何其他磁介质、CD-ROM、DVD、DVD-ROM、任何其他光介质、打孔卡片、纸带、任何其他具有孔图案的物理存储介质、RAM、ROM、PROM 和 EPROM、FLASH-EPROM、任何其他存储器芯片或盒式存储器、传送数据或指令的载波、传送此类载波的线缆或链路，或者可由计算机从中读取编程代码和 / 或数据的任何其他介质。这些形式的计算机可读介质中的许多介质可涉及将一个或多个序列的一个或多个指令携载至处理器以供执行。

[0280] 测序结果可以在用户界面如图形用户界面的帮助下呈现给用户。

V. 患者照护

[0281] 靶向测序技术可以在研究和临床设置中用于受试者的测序。在一个示例性实施方案中，对一种或多种已知的致病基因的测序可以在临床设置中进行，以用于携带者的筛查。可以建立临床试验组和潜在的包括基因组中的可注释位置的靶向测序组。该技术还可以用于拾取可动因子，如病毒插入位点，或可用于鉴定 SNP、突变、等位基团变异或遗传修饰。

[0282] 本文所述的发明可以提供数据，该数据可以被医疗专业人员或照护者使用以作出

给予照护的决定。在一些实施方案中，对一种或多种等位基因、基因变体、SNP 或其他基因组修饰的鉴定可将包括人类在内的动物鉴别为疾病携带者。作为疾病携带者的患者状态可以指导对患者或受试者的患者照护、生育决定、预后、防治、监测、诊断和 / 或治疗。

[0283] 诊断可以包括确定患者的状况。诊断可以在一个时间点进行或者持续地进行。例如，基于对基因组 DNA 中的病毒插入点的鉴定，可以诊断患者感染了病毒。基于遗传序列的存在或缺失，可以诊断患者患有某种病症。

[0284] 患者可以被鉴定为常染色体隐性突变或等位基因或染色体变体的携带者。作为携带者的患者的状态可以影响患者的避孕选择、患者对预防性照护的选择等。在一些情况下，携带者检测可以对个体进行。生育和 / 或避孕决定可以基于个体和或夫妇的携带者状态来作出。在一些情况下，对胎儿样品的测定（例如，产前检查）可以影响对母亲或胎儿的治疗决定。

[0285] 预后可以包括确定患者疾病的后果、恢复的机会或疾病将会如何进展。例如，鉴定染色体异常可以提供预后可能基于的信息。例如，染色体变体的存在或缺失可预测癌症存活率。

[0286] 监测可以包括对患者的一系列测试，以确定疾病的存在或监测疾病的进展。例如，已经被病毒感染的个体可以进行一系列监测，以确定抗病毒治疗是否预防该个体的进一步感染。

[0287] 防治可以包括确定治疗疗法。例如，患者的治疗疗法可以部分或全部基于某些基因组元件的存在或缺失来确定。例如，在个体的细胞色素 P450 基因内的某些基因组变体的存在或缺失可以影响药物代谢率并因此可影响对个体的治疗类型。

[0288] 可以使用本文所述的方法进行临床试验。在一些情况下，可以制定一个或多个方案以符合临床实验室改进修正案 (CLIA) 或食品药品管理局 (FDA) 规定。

VI. 临床或实验室研究

[0289] 方法、试剂盒和 / 或组合物可以在临床或实验室研究设置中使用，以研究疾病的遗传基础，例如，鉴定可能导致疾病、疾病的可能性或其他状况的新的遗传变异。方法可以用于研究病毒感染、疫苗的有效性。方法、试剂盒和 / 或组合物可以用于研究在传统上难以准确测序的基因组区域，诸如基因组内的高度多态性或可变性区域。

VII. 试剂盒

[0290] 用于实施本发明方法和测定的试剂任选地以试剂盒的形式提供以便于用户应用这些试验。这样的试剂盒通常还包括用于进行受试者测定的说明书，并且可任选地包括在其中进行反应的流体容器，例如小池、多孔板、微流体装置等。

[0291] 本发明的这些试剂盒试剂可以在供用户测量的小瓶中提供，或者可以在简单地组合起来以产生适当的反应混合物的预测量的小瓶或安瓿中提供。该试剂可以以液态和 / 或冻干的形式提供，并可任选地包括用于试剂的稀释和 / 或再水化的合适的缓冲溶液。通常，所有的试剂和说明书都共同包装在一个备用的盒子、袋或类似物中。

VIII. 靶基因

[0292] 本文提供的方法可以用于靶向于疾病相关基因的全部或一部分。例如，含有与一种或多种下列疾病相关的基因或基因区段的核酸可以使用公开的方法从样品中进行处理：ABCC8 相关的高胰岛素血症、全色盲、黑尿病、α -1 抗胰蛋白酶缺乏症、α - 甘露糖苷病、

Andermann 综合征、ARSACS、天冬氨酰葡糖胺尿症、共济失调伴维生素 E 缺乏、共济失调性毛细血管扩张症、常染色体隐性多囊肾病、BBS1 相关的 Bardet-Biedl 综合征、BBS10 相关的 Bardet-Biedl 综合征、生物素酶缺乏症、Bloom 综合征、卡纳万病、肉碱软脂酰转移酶 IA 缺乏症、肉碱软脂酰转移酶 II 缺乏症、软骨 - 毛发发育不全、无脉络膜、I 型瓜氨酸血症、CLN3 相关的神经元蜡样脂褐质沉积症、CLN5 相关的神经元腊样脂褐质沉积症、科恩综合征、Ia 型糖基化先天性疾病、Ib 型糖基化先天性疾病、先天性 Finnish 肾病、Costeff 视神经萎缩综合征、囊性纤维化、胱氨酸病、D- 双功能蛋白质缺乏症、因子 V 莱顿血栓形成倾向、因子 XI 缺乏症、家族性自主神经功能异常、家族性地中海热、C 型范科尼贫血、脆性 X 染色体综合征、半乳糖血症、戈谢病、GJB2 相关的 DFNB1 非综合征性听力损失和耳聋、葡萄糖 -6- 磷酸脱氢酶缺乏症、1 型戊二酸血症、Ia 型糖原贮积病、Ib 型糖原贮积病、III 型糖原贮积病、V 型糖原贮积病、GRACILE 综合征、Hb β 链相关的血红蛋白病（包括 β 地中海贫血和镰状细胞病）、遗传性果糖不耐受症、遗传性胸腺嘧啶 - 尿嘧啶尿、赫利茨大疱性结合性表皮松解症 (LAMA3 相关的）、赫利茨大疱性结合性表皮松解症 (LAMB3 相关）、赫利茨大疱性结合性表皮松解症 (LAMC2 相关）、氨基己糖苷酶 A 缺乏症（包括泰 - 萨克斯病）、HFE- 相关的遗传性血色病、由胱硫醚 β - 合酶缺乏引起的高胱氨酸尿、赫尔利综合征、低磷酸酯酶症（常染色体隐性）、包涵体肌病 2、异戊酸血症、朱伯特综合征 2、克拉伯病、2D 型肢带型肌营养不良症、2E 型肢带型肌营养不良症、硫辛酰胺脱氢酶缺乏症、长链 3- 羟酰辅酶 A 脱氢酶缺乏症、1B 型枫糖尿症、中等链长脂酰 CoA 脱氢酶缺乏症、巨脑性脑白质病 (Megalencephalic Leukoencephalopathy) 伴皮质下囊肿、异染性脑白质营养不良症、MTHFR 缺乏症、黏脂质累积病 IV、肌肉 - 眼 - 脑疾病、NEB 相关的线状体肌病、C 型尼曼 - 匹克病、尼曼 - 皮克病 (SMPD1- 相关的）、Nijmegen 破损综合征、Northern 癫痫、Pendred 综合征、PEX1 相关的 Zellweger 综合征谱、苯丙氨酸羟化酶缺乏症、1 型多腺性自身免疫性综合征、庞皮病、PPT1 相关的神经元蜡样质脂褐质沉积症、原发性肉碱缺乏症、1 型原发性高草酸尿症、2 型原发性高草酸尿症、PROP1- 相关的合并垂体激素缺乏症、凝血酶原血栓形成倾向、假性胆碱脂酶缺乏症、致密性成骨不全症、1 型肢根点状软骨发育不良、Salla 病、Segawa 综合征、短链酰基辅酶 A 脱氢酶缺乏症、Sjogren-Larsson 综合征、Smith-Lemli-Opitz 综合征、脊髓性肌萎缩、激素耐药型肾病综合征、硫酸盐转运蛋白相关的骨软骨发育不良、TPP1 相关的神经元蜡样脂褐质沉积症、I 型酪氨酸血症、1F 型乌谢尔综合征、3 型乌谢尔综合征、极长链酰基辅酶 A 脱氢酶缺乏症 (Wilson 病）和 X 染色体连锁的青年性视网膜劈裂症。

[0293] 该疾病可以是癌症。癌症可以是，例如，肿瘤，白血病如急性白血病、急性 T 细胞白血病、急性淋巴细胞性白血病、急性髓样白血病、成髓细胞性白血病、前髓细胞性白血病、粒单核细胞性白血病、单核细胞性白血病、红白血病、慢性白血病、慢性髓细胞性（粒细胞性）白血病或慢性淋巴细胞性白血病，真性红细胞增多症，淋巴瘤如霍奇金淋巴瘤、滤泡性淋巴瘤或非霍奇金淋巴瘤，多发性骨髓瘤，瓦尔登斯特伦巨球蛋白血症，重链病，实体瘤，肉瘤，癌瘤例如纤维肉瘤、粘液肉瘤、脂肪肉瘤、软骨肉瘤、成骨性肉瘤、淋巴管肉瘤、间皮瘤、Ewing 瘤、平滑肌肉瘤、横纹肌肉瘤、结肠癌、结直肠癌、胰腺癌、乳腺癌、卵巢癌、包括去势抗性前列腺癌在内的前列腺癌、鳞状细胞癌、基底细胞癌、腺癌、汗腺癌、皮脂腺癌、乳头状癌、乳头状腺癌、囊腺癌、髓样癌、支气管原癌、肾细胞癌、肝细胞瘤、胆管癌、绒毛膜癌、精原细胞瘤、胚胎性癌、肾母细胞瘤 (Wilms' tumor)、宫颈癌、子宫癌、睾丸肿瘤、肺癌、小细胞肺癌、

膀胱癌、上皮癌、神经胶质瘤、颅咽管瘤、室管膜瘤、松果体瘤、血管母细胞瘤、听神经瘤、少突神经胶质瘤、脑膜瘤、黑素瘤、神经母细胞瘤、视网膜母细胞瘤、子宫内膜癌、非小细胞肺癌、头颈癌或肾癌。

[0294] 该疾病可以是自身免疫性疾病。该自身免疫性疾病可以是针对个体自身组织或其共分离或表现并由其引起的疾病或病症，或由其导致的状况。自身免疫性疾病或病症的实例包括但不限于：急性播散性脑脊髓炎 (ADEM)、关节炎、急性坏死性出血性白质脑炎、阿狄森病、丙种球蛋白缺乏血症、斑秃、淀粉样变性、强直性脊柱炎、抗 GBM/ 抗 TBM 肾炎、抗磷脂综合征 (APS)、自身免疫性血管性水肿、自身免疫性再生障碍性贫血、自身免疫性自主神经功能异常、自身免疫性肝炎、自身免疫性高脂血症、自身免疫性免疫缺陷、自身免疫性内耳病 (AIED)、自身免疫性心肌炎、自身免疫性胰腺炎、自身免疫性视网膜病、自身免疫性血小板减少性紫癜 (ATP)、自身免疫性甲状腺病、自身免疫性荨麻疹、轴突和神经元神经病、巴洛病、白塞氏病、大疱性类天疱疮、心肌病、Castleman 病、乳糜泻、Chagas 病、慢性疲劳综合征、慢性炎性脱髓鞘性多发性神经病 (CIDP)、慢性复发性多病灶性骨髓炎 (CRMO)、-Strauss 综合征、瘢痕性类天疱疮 / 良性黏膜类天疱疮、克罗恩病、Cogans 综合征、冷凝集素病、先天性心脏传导阻滞、柯萨奇病毒性心肌炎、CREST 病、混合性冷沉球蛋白血症、神经病、疱疹样皮炎、皮肌炎、德维克病（视神经脊髓炎）、盘状狼疮、Dressler 综合征、子宫内膜异位症、嗜酸细胞增多性食管炎、嗜酸性筋膜炎、结节性红斑、实验性过敏性脑脊髓炎、Evans 综合征、纤维肌痛、纤维化肺泡炎、巨细胞动脉炎（颞动脉炎）、肾小球肾炎、肺出血肾炎综合征、肉芽肿病伴多血管炎 (GPA)、Graves 疾病、Guillain-Barre 综合征、桥本氏脑炎、桥本氏甲状腺炎、溶血性贫血、Henoch-Schonlein 紫癜、妊娠疱疹、低丙球蛋白血症、特发性血小板减少性紫癜 (ITP)、IgA 肾病、IgG4 相关的硬化性疾病、免疫调节性脂蛋白、包涵体肌炎、胰岛素依赖性糖尿病 (1 型)、间质性膀胱炎、少年关节炎、I 型糖尿病、川崎综合征、Lambert-Eaton 综合征、白细胞碎裂性血管炎、扁平苔藓、硬化性苔藓、木样结膜炎、线状 IgA 病 (LAD)、狼疮 (SLE)、莱姆病、Meniere 病、显微镜下多血管炎、混合性结缔组织病 (MCTD)、Mooren 溃疡、Mucha-Habermann 病、多发性硬化、重症肌无力、肌炎、发作性睡病、视神经脊髓炎 (Devic 病)、中性粒细胞减少症、眼瘢痕性类天疱疮、视神经炎、复发性风湿病、PANDAS（链球菌相关的小儿自身免疫性神经精神性障碍）、副肿瘤性小脑变性、发作性睡眠性血红蛋白尿 (PNH)、Parry Romberg 综合征、Parsonnage-Turner 综合征、睫状体扁平部炎（周边葡萄膜炎）、天疱疮、周围神经病、静脉周围脑脊髓炎、恶性贫血、POEMS 综合征、结节性多发性动脉炎、I、II 或 III 型自身免疫性多腺体综合征、风湿性多肌痛、多肌炎、心肌梗塞后综合征、心包切开术后综合征、黄体酮皮炎、原发性胆汁性肝硬化、原发性硬化性胆管炎、银屑病、银屑病性关节炎、特发性肺纤维化、坏疽性脓皮病、纯红细胞再生障碍、雷诺现象、反射性交感神经营养障碍、Reiter 综合征、复发性多软骨炎、不宁腿综合征、腹膜后纤维化、风湿热、类风湿性关节炎、结节病、Schmidt 综合征、巩膜炎、硬皮病、干燥综合征、精子与睾丸自身免疫、僵人综合征、亚急性细菌性心内膜炎 (SBE)、Susac 综合征、交感性眼炎、Takayasu 动脉炎、颞动脉炎 / 巨细胞动脉炎、血小板减少性紫癜 (TTP)、Tolosa-Hunt 综合征、横贯性脊髓炎、溃疡性结肠炎、未分化结缔组织病 (UCTD)、葡萄膜炎、血管炎、水疱大疱性皮肤疾病 (Vesiculobullous dermatosis)、白癜风和韦格纳肉芽肿病。

[0295] 本文提供的方法可用于制备含有与癌症相关的基因或基因部分的样品。例如，本

文提供的方法可以制备用于对与癌症相关的体细胞突变进行测序的样品。与癌症相关的体细胞突变可见于例如由 Wellcome Trust Sanger Institute 维护的 COSMIC 数据库中，其通过引用并入本文。

X. 预见性实施例

实施例 1 ：针对 mPEAR 生成的文库的样品方案

[0296] 获得输入多核苷酸。一般为至少 50ng 的高分子量基因组 DNA(gDNA)。对 gDNA 进行片段化。

[0297] 片段化 ：

[0298] 多种片段化方法是合适的。可以使用剪切片段化 (例如，Covaris)。平均片段大小可以是约 100 个、200 个、300 个、400 个或更多个核苷酸，但可以根据所使用测序仪平台而不同。对于第三代测序技术，片段大小可以更大。

[0299] 将 DNA 重悬在 75μL 1x Tris-EDTA(TE) 缓冲液中。将重悬混合物加至玻璃 Covaris 管中。可以使用下列设置 ：占空比 10％，强度 5，循环 / 猝发 200，时间 120 秒。

末端修复和衔接子连接 ：

[0300] 可以产生平端。对于 Illumina 文库生成，可以进行 A- 加尾。对于其他测序平台，A- 加尾可以是任选的。

[0301] 末端修复

[0302] 首先将连接酶缓冲液加至珠子。制备主混合物。

[0303] 在 0.5mL 低 DNA 结合管中制备下列反应混合物 ：

[0304] H_2O	75μL
[0305] 含有 10mM ATP 的 T4 DNA 连接酶缓冲液	10μL
[0306] 10mM dNTP 混合物	4μL
[0307] T4 DNA 聚合酶	5μL
[0308] Klenow 片段 5U/μL	1μL
[0309] T4 多核苷酸激酶	5μL
[0310] 共计 ：	100μL

[0311] 将珠子重悬于末端修复混合物中。将样品在 20℃的加热块中温育 30 分钟。置于磁性颗粒集合器 (MPC) 上并去除上清液。使用 MPC，用 200μL 无 tRNA 的 1X SSC 缓冲液洗涤固定化的文库 3 次。在每次洗涤之间充分混合。

A- 加尾混合物 ：

[0312] 首先将缓冲液加至珠子或制备主混合物。在 0.5mL 低 DNA 结合管中制备下列反应混合物 ：

[0313] H_2O	32μL
[0314] 10X 缓冲液 B011	5μL
[0315] 1mM dATP	10μL
[0316] Klenow Exo-	3μL
[0317] 共计 ：	50μL

[0318] 将珠子重悬于 A 加尾混合物中。将样品在 37℃的加热块中温育 30 分钟。置于 MPC 上并去除上清液。使用 MPC，将固定化的文库用 200μL 含 tRNA 的 1X SSC 洗涤 3 次，并用无

tRNA 的 1X SSC(1X ＝ 0.150M 氯化钠,0.015M 柠檬酸钠）缓冲液洗涤 1 次。在每次洗涤之间充分混合。

连接：

[0319] 将 2X 连接缓冲液加至珠子。重悬珠子,使得最终衔接子浓度为 0.3 μM。用退火溶液以 1:10 稀释原液。不同的条形码衔接子可用于以上的 10 个、15 个和 20 个循环。

[0320] dH_2O 18 μL

[0321] 2X 快速连接缓冲液 (Enzymatics B101L) 25 μL

[0322] TruSeq 条形码化衔接子 (1:10 稀释 bc10,11,12) 3 μL

[0323] 共计： 45 μL

[0324] 加入 5 μL 的 T4DNA 连接酶（快速）并通过上下移液而混合。在混合下,将样品在 20℃下温育 15 分钟。使用磁体,将珠子用 200 μL 1X SSC(含有 10ng/μL tRNA) 洗涤 3 次,并用 200 μL 1X SSC(无 tRNA) 洗涤 1 次。将珠子重悬于 23 μL H_2O 中。转移至薄壁 PCR 管中,用另外的 23 μL 水冲洗该管。

PCR 富集：

[0325] 这一步骤可以是任选的。

[0326] 在 500 μL 薄壁 PCR 管中制备下列 PCR 反应混合物：

[0327] 重悬有珠子的 H_2O 23 μL

[0328] 2x HiFi KAPA 主混合物 50 μL

[0329] Truseq PCR 引物 1 2 μL

[0330] Truseq PCR 引物 2 2 μL

[0331] 含核酸酶的水,来自以上 23 μL

[0332] 运行下列 PCR 方案 10 个和 15 个循环：

[0333] 98℃下 45 秒

[0334] 10 个循环后取出 50 μL,然后再进行 10 个循环。

[0335] 98℃,15 秒

[0336] 60℃,30 秒

[0337] 72℃,30 秒

[0338] 72℃下 1 分钟

[0339] 在 4℃下保持

[0340] 纯化 DNA(例如,使用 Zymo(25)PCR 纯化试剂盒）。在 30 μL 的 dH_2O 中洗脱。在 4％琼脂糖凝胶上纯化 10 个循环和 20 个循环的产物。

实施例 2：针对 mPEAR 的示例性衔接子设计

[0341] 衔接子序列可以由在 5'端的通用序列、关于取向信息的 3 个核苷酸的序列和在 3'端的 4-6 个核苷酸的分子条形码组成。引入分子条形码可允许进行多重靶标富集和测序。DNA 样品可以单独进行片段化、末端修复和衔接子连接。由于各个样品可具有条形码,所以 mPEAR 步骤可以是多重化的,从而通过消除文库生成后的单独样品处理步骤而极大降低了成本并增加了通量。使用基因组 RAPELLing 进行的 mPEAR 文库生成适合于所有测序平台。在从链霉亲和素珠上洗脱后通过低循环 PCR 反应引入测序仪平台特异性衔接子。较短的衔接子序列允许文库分子的更高的连接效率和更精确的大小选择。图 14 示出了合适序列的

实例。

实施例 3 :mPEAR

[0342] 通过对条形码化样品进行 12 个循环的 PCR 来制备文库。将 30 μL DNA 样品转移至 0.5mL LoBind 管中。加入 1nmol 的各种 3’ 封闭的衔接子封阻剂寡核苷酸 (各 10 μL, 100 μM,通用封阻剂 1、通用封阻剂 2、封阻剂 1 和封阻剂 2) 并干燥 (例如,真空离心蒸发浓缩 (speedvac))。

[0343] 重悬 :

[0344] 5 μL 10X 标准 Taq 缓冲液 (60%)5X 生物素 -dNTP

[0345] 10 μL 10X 60%生物素 -dNTP 6.25 μL dA、dC、dG(20mM)

[0346] 1 μL 60 引物混合物 (25 μM) 15 μL bio-dUTP(5mM)

[0347] 32 μL 不含核酸酶的水 2.5 μL dT(20mM)

[0348] 1 μL 100mM $MgCl_2$溶液 13.75 μL nH_2O

[0349] 1 μL Taq(5U/ μL) 1 μL

[0350] 共计 : 50 μL

[0351] 加热至 98℃ 2 分钟。或者在热循环仪上采用斜降缓慢冷却至 47℃,或者快速冷却至 47℃,然后在 47℃保持 4 分钟,继而快速斜升至 72℃ 10 分钟,然后加入 1 μL 的 0.5M EDTA 以猝灭。置于冰上。

通用封阻剂 1 :

GATCGGAAGAGCACACGTCTGAACTCCAGTCAC555555ATCTCGTATGCCGTCTTCTGCTTGX

通用封阻剂 2 :

CAAGCAGAAGACGGCATACGAGAT555555GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCX

文库固定化

[0352] 用使用 H_2O 填充至 10mL 的 B&W-100 μL,1M Tris pH 7.5,20 μL 0.5M EDTA,4mL 5M NaCl 洗涤链霉亲和素珠子 (例如, Dynal M280 链霉亲和素珠子)。将 25 μL 珠子转移至新的 0.5mL LoBind 管中。用 B&W 缓冲液 (含有 10ng/ μL tRNA) 洗涤链霉亲和素磁珠 3 次,每次洗涤使用 200 μL 缓冲液 (移液混合)。最后一次洗涤可以是 5 分钟。使用 MPC 使该珠子沉淀并除去缓冲液。加入 50 μL 的 B&W 缓冲液 (无 tRNA)。将 50 μL 的 DNA 转移至 50 μL 的 Dynal M-280 链霉亲和素珠中,或者反过来亦可,可以使用另外 20 μL B&W 缓冲液来冲洗管。充分混合并在室温下温育 1 小时,每 15 分钟混合一次。使用磁体,用 200 μL 的 B&W 缓冲液洗涤固定化的文库 3 次。用 200 μL 的 1X SSC 洗涤 2 次。使用最后一次洗涤将珠子转移至 PCR 管中。除去最后一次洗涤剩余的所有 1X SSC 缓冲液。现在珠子应该以扩增就绪的状态处于 PCR 管中。

[0353] 通过 PCR 进行富集—15 和 25 个循环

[0354] 向珠子加入 : 98℃,45 秒 - 保持

[0355] 50 μL 2X HiFi KAPA 就绪混合物 98℃,15 秒

[0356] 2 μL PCR 1 60℃,30 秒

[0357] 2 μL PCR 2 72℃,30 秒

[0358] 46 μL nH_2O 72℃,60 秒 - 保持

[0359] 100 μL 4℃ - 保持

[0360] 清洗 DNA(例如,使用 Zymo-25 柱)。用 30μL 洗脱并通过真空离心蒸发浓缩进行干燥。在 20μL 中重悬。加载至 2%琼脂糖凝胶上并同分子量标准梯(ladder)一起电泳。切下 320-400 个核苷酸的区域。切下的 DNA 可以进行处理并测序。

数据分析:

[0361] 在确定标准数据质量指标后,可以利用分子条形码对序列读取值进行分箱或关联。

锚式读取值比对:

[0362] 命中的测序仪读取值将含有 mPEAR 引物-退火位点。该退火位点用作"锚"以将读取值定位在基因组中。该"锚"的下游序列可以当作从头序列。这相对于鸟枪式外显子组测序方法具有显著优势,该鸟枪式外显子组测序方法需要使用与参考基因组相类似的读取值。更长的插入和缺失通过从头测序而检测到,但被再测序遗漏,这是因为它们与参考基因组有太大差异从而不能被大部分比对方法检测到。位于已知 mPEAR 退火位点侧翼的易位、病毒插入位点和其他可动因子也通过锚式读取值比对而检测。

脱离-配对测序:

[0363] mPEAR 允许脱离配对测序。两条链用不同的 mPEAR 退火位点来靶向。该退火位点可以在计算机中配对。读取值之间的重叠提高了序列质量,并且 mPEAR 退火位点之间的距离可用于确定拷贝数或染色体重排。在一些情况下,两个 mPEAR 退火位点将包含于单个读取值中。这允许确定长重复序列,例如预示脆性 X 染色体综合征的串联三核苷酸重复。更长的靶序列可能需要多锚比对。

实施例 3:基因分型诊断学

[0364] 对人类 DNA 样品(例如,怀疑为癌性的人皮肤细胞的群体)准备 TELA 反应。可以利用去污剂和热来裂解细胞,并且通过氯仿/乙醇提取来沉淀出二倍体 DNA 的约 15,000 个拷贝。DNA 的重悬可以收集为具有单倍体 DNA 的约 10,000 个拷贝。可向 DNA 样品应用包含 100 个 TELA 引物组的文库。该文库内的引物组可以包含已知与皮肤癌相关的各种癌基因和肿瘤抑制基因的基因座特异性序列。各个引物还可以包含条形码序列。引物延伸反应在如本文所述的类似反应条件下用 Klenow 片段聚合酶进行。使用通用引发位点(通过 TELA 引物)和简并引物的后续 PCR 扩增可使条形码化 DNA 的产量增加至 10ng。

[0365] 可使用多重测序策略以足够的覆盖范围(例如,500)对样品进行测序。单个 DNA 链的条形码化可允许从单个链获得测序信息,而非作为整个 DNA 样品的平均值而获得。基于所测序的 DNA 链和所分配的条形码的数目,得到了 SNP 定相/单元型分型信息,并可以解析 DNA 的许多重复区域。另外,通过丢弃相对于单元型随机出现的突变可以获得准确度的大幅提升,因为这些突变可能是测序错误。SNP 定相/单元型分型信息提供了可能存在于细胞中的皮肤癌类型的遗传学线索。可以为医师生成报告以供随访和评估。此外,这些报告可以通过因特网进行电子提交及访问。序列数据的分析可以在受试者所处位置之外的地点进行。生成该报告并将其提交至用户/医师所在位置。如图 11 所示,医师经由能够连接因特网的计算机来访问反映疑似癌症分析的报告。

实施例 4:RAPELL 方案

[0366] 加入 gDNA(例如,人脑 DNA)、NEB4 缓冲液、nH_2O 并在 96℃变性 10min。

[0367] 1μL 人脑 DNA(HBD)50ng

[0368] 2μL RAPELL 1(10μM)

[0369] 2μL RAPELL 2(10μM)

[0370] 1.2μL 10mM dNTP

[0371] 4μL 10X NEB4 缓冲液

[0372] 17.8μL nH_2O

[0373] 2μL T4DNA 聚合酶（无链置换）

[0374] 2μL DNA 连接酶 (600U/μL)

[0375] 3.6μL 10mM ATP

[0376] 35.6μL 反应

[0377] 在 20℃下温育 0.5 小时。在 96℃下变性 2 分钟。加入 2μL T4DNA 聚合酶。加入 2μL DNA 连接酶。在 20℃下温育 0.5 小时。在加热变性与加入酶之间循环 5 次，第 2 次循环加入 2μL 10X NEB4 缓冲液。在 75℃下温育 20 分钟以灭活该酶。

[0378] 使用来自以上的 53.6μL。加入下列试剂：

[0379] 6μL UDG

[0380] 6μL APE

[0381] 共计：65.6μL

[0382] 在 37℃下温育 1 小时。清洗 DNA（例如，用 5 倍体积的 DNA 结合缓冲液 (320μL) 进行 Zymo-5 清理）。用 20μL 洗脱液进行洗脱。

[0383] 在 15 个和 25 个循环之间的 PCR 循环

[0384] 来自以上的 20μL DNA 98℃，45 秒

[0385] 2μL PCR1(25μM) 98℃，15 秒

[0386] 2μL PCR2(25μM) 60℃，30 秒

[0387] 50μL 2X HiFi KAPA 混合物 72℃，30 秒

[0388] 26μL nH_2O 72℃，60 秒

[0389] 100μL 4℃ - 保持

[0390] 清洗 DNA（例如，Zymo-25 清理）。用 30μL 洗脱液进行洗脱并进行干燥（例如，利用真空离心蒸发浓缩）。在 2％琼脂糖凝胶上电泳。切下 DNA 以供进一步处理、测序和数据分析。

实施例 5：用于 RAPELLing 的示例性寡核苷酸

[0391] 图 18 公开了可用于进行本发明方法的示例性寡核苷酸。

实施例 6：

[0392] 用 mPEAR 生成的文库获得长片段测序读取值的样品方法。从来源获得多核苷酸。对多核苷酸进行片段化、末端修复，并进行由 mPEAR 介导的扩增靶文库生成。该靶文库然后经历 RAPELLing 方法，其中将靶文库片段化为长多核苷酸片段并进行分配、拷贝、扩增和测序。然后组装共有序列。

实施例 7：

[0393] 1. 引物设计：覆盖式 (tiling) 策略可用于靶向基因座的全面覆盖。从靶基因座上游 100-200 个核苷酸开始并穿过基因座继续在基因座后延伸 100-200 个核苷酸。靶标可以分解成 100-200 个核苷酸窗口，其中选择表现最佳的一个或多个引物。引物的长度可以

不同。可对引物进行优化以得到最优 Tm。可将引物设计成对靶位置具有最大特异性。引物可设计为避免二聚化。引物的 5’末端可以用对应于在反应中使用的测序仪平台的序列进行加尾。这种序列可以称为“A 衔接子”。该引物可以称为“A 衔接子加尾的引物”。引物可具有能够被聚合酶延伸的 3’OH 基团。

[0394] 2.A 衔接子的添加。A 衔接子加尾的引物可以与基因组 DNA(“gDNA”) 杂交。热稳定的链置换聚合酶可用于延伸。延伸可以在 5’-3’方向上发生。模板 DNA 的拷贝可以在置换已经与模板杂交的任何 DNA 链的同时通过聚合酶制备。该过程可以通过热变性、引物退火和引物延伸来重复或循环。只有正向链以这种方式进行拷贝。

[0395] 3. 片段化 :扩增产物的片段化可以通过酶促或物理手段完成以产生随机的 3’端。用于产生随机 3’端的优选方法是通过在引物延伸反应中使用低比例的生物素 ddNTP/dNTP 来掺入生物素化的 ddNTP。该 ddNTP 可随机地终止延伸的分子。其他片段化方法可包括诸如声处理的剪切或酶促片段化。通过使用生物素化的 ddNTP,可以通过链霉亲和素珠纯化来分离生物素化的分子。

[0396] 4.B 衔接子的添加。可添加 B 衔接子用于测序反应。如果使用酶促或物理片段化,B 衔接子可以通过连接或另外的引物延伸步骤来添加。当终止 ddNTP 在前一步骤中用于片段生成时,分子的 3’末端由于缺乏 3’OH 而无法进行引物延伸或连接。在这种情况下,B 衔接子可以通过将随机引物与 B 衔接子尾在分离的分子的 5’末端上或附近杂交来添加。在一个非限制性实例中,随机引物可以包含 :5’-B 衔接子 -NNNN…-3’,其中“NNNN”代表一个或多个核苷酸的随机核酸序列。引物的随机区段可以沿捕获的分子进行杂交。链置换聚合酶可以用于延伸。延伸可以在 5’-3’方向上发生。在生物素化的模板的末端处的随机引物可以延伸并置换所有其他引物。延伸和置换后,它可以是在生物素化的模板上剩余的唯一分子。双链复合物可以包含含有 5’A 衔接子 - 基因座特异性引物 - 靶标 - 终止 ddNTP- 生物素 3’的链。反义链可含有 5’-B 衔接子 - 随机序列 - 靶标 - 基因座特异性序列 -A 衔接子 -3’。该复合物可通过洗涤链霉亲和素珠子和除去上清液而再次分离。

[0397] 5. 释放 :既具有 A 衔接子又具有 B 衔接子的复合物可以通过热变性从珠子上释放。使用与 A 和 B 衔接子序列互补的引物进行 9-12 个循环的 PCR。如果它们尚未被引入,可以在 PCR 步骤中通过使用加尾的引物来引入全长测序衔接子。

[0398] 任选地,用于在测序过程中进行样品多重化的分子条形码可以在初始 A 衔接子引物延伸时、B 衔接子引物延伸时和 / 或最终 PCR 扩增期间添加。

[0399] 6. 测序仪就绪的文库 :得到的测序仪就绪的文库可以由下列形式的双链分子组成 :5’-A 衔接子 - 合成引物 - 靶标 - 靶标随机末端 -b 衔接子 -3’。可以对该文库进行测序并且可以存储和 / 或传送数据用于分析。

[0400] 7. 数据分析 :在数据分析期间,可以从读取值中修剪掉衔接子。样品可以基于任意任选的条形码来鉴定。可以除去重复的读取值。基因组坐标可以根据读取开始时已知的合成序列来鉴别。与相同的已知基因组坐标相对应的读取值可以分箱在一起。共有序列可以在不使用参考基因组的情况下生成。不形成共有序列的读取值可以作为脱靶读取值从分析中除去。靶 DNA 的各个连续区段可以被认为是单个靶标,并且与该靶标相对应的所有引物可以被认为是该靶标的引物组。这可以产生计算机读取长度,该长度可等于该靶标的全长。无论靶标大小如何都可以进行计算机读取长度的生成。确定共有序列之后,可以查询

靶区域的单元型。对于与已知单元型不匹配的共有序列，可以使用从头测序来确定新的单元型和 / 或结构变异。

[0401] 8. 应用 ：该技术可以用于已知致病基因的临床测序，诸如携带者检测。可以建立其他试验组，以及潜在的包括基因组中的可注释位置的靶向测序组。另外，由于测序可以从已知进行到未知，所以该技术可用于鉴定病毒插入位点。

实施例 8 ：

[0402] 1. 引物设计 ：可以设计多个引物以与选定靶区域之内或附近的特定靶序列杂交。各个引物的间隔是可以改变的。可变间隔可以导致沿选定靶区域的整个长度的高水平覆盖。

[0403] 2. A 衔接子的添加。单向引物还可以包含对将要使用的测序平台具有特异性的 5’序列（例如，在 5’末端上的“A”衔接子序列）。

[0404] 3. 延长 ：引物混合物与靶 DNA 杂交，并使用加有低浓度 dUTP 的 dNTP 混合物由聚合酶进行延伸。

[0405] 4. 大小控制 / 片段化 ：可以切割延伸的引物以生成一组嵌套的单链产物。切割方法的一个非限制性实例是使用尿嘧啶 DNA 糖基化酶（“UDG”）和 / 或人除嘌呤 / 除嘧啶核酸内切酶 I(APE I) 来生成一组嵌套的单链产物。这些产物可以在 5’末端被 A 衔接子序列锚定。一种可替代的片段化策略可以使用加有低浓度甲基 -dCTP 的 dNTP 混合物，随后使用 4 碱基切割酶进行限制酶消化，该切割酶将不会切割已掺入甲基 -C 的位点。

[0406] 5. B 衔接子的添加。可以使用具有 5’- 随机碱基突出端、含有 B 测序衔接子的双链构建体，通过连接或引物延伸将 B 测序衔接子添加到 3’末端。添加 B 衔接子后，文库可以使用 A 和 B 特异性 PCR 引物进行 PCR 扩增。如果已添加的 B 衔接子不是全长的，则可以通过引物尾添加全长衔接子。

[0407] 任选地，用于测序过程中的样品多重化的分子条形码可以在初始 A 衔接子引物延伸时、B 衔接子引物延伸时和 / 或最终 PCR 扩增期间加入。

[0408] 6. 测序仪就绪的文库 ：得到的测序仪就绪的文库可以由下列形式的双链分子组成 ：5’-A 衔接子 - 合成引物 - 靶标 - 靶标随机末端 -b 衔接子 -3’。可以对该文库进行测序并且可以存储和 / 或传送数据用于分析。

[0409] 7. 数据分析 ：在数据分析期间，可以从读取值中修剪掉衔接子。样品可以基于任意任选的条形码来鉴定。可以除去重复的读取值。基因组坐标可以根据读取开始时已知的合成序列来鉴别。与相同的已知基因组坐标相对应的读取值可以分箱在一起。共有序列可以在不使用参考基因组的情况下生成。不形成共有序列的读取值可以作为脱靶读取值从分析中除去。靶 DNA 的各个连续区段可以被认为是单个靶标，而与该靶标相对应的所有引物可以被认为是该靶标的引物组。这可以产生计算机读取长度，该长度可等于该靶标的全长。无论靶标大小如何都可以进行计算机读取长度的生成。确定共有序列之后，可以查询靶区域的单元型。对于与已知单元型不匹配的共有序列，可以使用从头测序来确定新的单元型或结构变异。

[0410] 8. 应用 ：该技术可在临床测序设置中使用。一个非限制性的实例是检测已知与疾病相关的等位基因，例如，携带者检测。可以建立其他试验组，以及潜在的包括基因组中的可注释位置的靶向测序组。另外，由于测序可以从已知进行到未知，所以该技术可用于鉴定

病毒插入位点。由于这种靶向方法在提供序列比对位置的同时还允许读取值的从头组装，并且由于靶标是通过引物延伸进行测序而不是基于纯化或通过杂交的识别，因此它可以富集含有其他基于杂交的靶标捕获方法难以保留的插入、缺失和/或其他遗传异常的测序文库。在一个非限制性的实例中，这种方法可能特别适合对人类组织相容性抗原(HLA)区域进行测序。

[0411] 这种方法还描述了用于靶向、半靶向和包括全基因组定相的全基因组扩增程序的有用的方法。在全基因组定相的情况下，该策略将开始于在多重反应中的高度稀释的、亚基因组量的DNA。该DNA可以是相对较大的片段(例如，10-40kb)。在该方案的这种重复中，在各个反应中生成的大多数序列源自受试者基因组的单个拷贝，因此大多数检测到的变异可能是纯合的。

实施例9：

[0412] 1.文库制备：文库可以使用平台特异性文库制备方法或试剂盒来制备。该方法或试剂盒可以是市售的，并且可以生成测序仪就绪的文库。平台特异性文库制备方法可将已知的序列添加至核酸分子的末端；该已知序列可以被称为衔接子序列。任选地，文库制备方法可以引入一种或多种分子条形码。

[0413] 2.靶向：来自测序仪就绪文库的DNA分子可以使用一种或多种引物(即mPEAR引物)的合并物进行选择(即，靶向)。mPEAR引物可以与靶文库分子或片段杂交。杂交的mPEAR引物可以使用聚合物来延伸。mPEAR引物可以包含通用或共同的5'末端、间隔区序列和靶标或基因座特异性序列。通用或共同末端可以与来自先前文库产生步骤的通用衔接子序列杂交。这可以用于将向DNA文库片段的末端的合成的寡核苷酸稳定化。向文库片段的末端的稳定可以在不浪费测序仪容量的情况下允许测序仪读取值适当地定位靶序列。间隔区序列可以包含可变数目的简并核苷酸。简并核苷酸允许DNA测序过程中的DNA文库起始和停止位置的长度灵活性。具有可变序列起始位点可减少测序步骤中的系统性错误。具有可变序列起始位点可允许在冗余DNA文库片段的读取值中的随机化错误分布。最后，基因座特异性结合位点位于mPEAR引物的3'末端附近。基因座特异性引发位点被设计成识别实际靶序列上游的DNA序列。一种或多种mPEAR引物可以串联使用，以靶向一个或多个目标区域。靶向该靶标的上游序列允许提高特异性，这是因为可以避免具有类似序列同源性的假基因和基因家族，从而减少数据中的假阳性。任选地，第二mPEAR引物可以设计成与相反链结合。任选地，mPEAR引物可以引入分子条形码。

[0414] 2.延伸：mPEAR引物的3'末端可用于引物延伸。可以使用聚合酶延伸分子。该延伸可以在5'-3'方向上发生。可以引入生物素化的dNTP(即天然的和生物素化的dNTP的混合物可以在延伸反应中使用)。任选地，mPEAR引物延伸可以在两条链上都发生。mPEAR引物延伸可以运行通过DNA文库分子的末端。在DNA两条链上发生的mPEAR引物延伸的任选使用可能是有优势的；有两个反应针对相同的序列可以提高特异性和/或可以减少失败(例如，如果mPEAR引物之一例如由于不能杂交而失败)。

[0415] 3.分离：新合成的、生物素化的DNA文库/捕获分子杂合体可以与链霉亲和素(例如涂覆在珠子上的链霉亲和素)一起温育。生物素化的靶DNA分子可以通过磁珠纯化进行分离。可以用合适的缓冲液进行一次或多次洗涤。任选地，DNA文库分子可以从捕获分子中洗脱出来。或者，可以在磁珠仍处于溶液中时进行扩增反应(即，在无洗脱步骤的情况下)。

扩增可以使用适当的引物通过 PCR 进行。在 PCR 扩增期间，可以引入全长测序平台特异性衔接子序列。得到的扩增后的分子可以是测序仪就绪的，或者可以在测序前通过本领域中已知的任何手段进一步纯化。

[0416] 4. 测序：可以对文库进行测序，并且可以存储和 / 或传送数据以供分析。

[0417] 5. 数据分析：在数据分析期间，可以从读取值中修剪掉衔接子。样品可以基于任意任选的条形码来鉴别。可以除去重复的读取值。基因组坐标可以根据读取开始时已知的合成序列来鉴别。与相同的已知基因组坐标相对应的读取值可以分箱在一起。共有序列可以在不使用参考基因组的情况下生成。不形成共有序列的读取值可以作为脱靶读取值从分析中去除。捕获的分子的读取结构在数据分析的速度和质量上可具有显著优势。由于基因座特异性引物识别位点位于靶 DNA 序列的上游，所以该基因座特异性位点用于鉴定读取值的基因组位置。参考基因组并不是绝对必需的；通过在读取值中“种入 (seeding)”已知序列，剩下的可以在不使用参考基因组的情况下进行组装。可以完成 DNA 靶标的真正的从头测序。从头测序可允许与参考基因组非常不同的基因组基因座的较大灵敏度。以较大的灵敏度实现临床相关的较长插入、缺失、重复和破坏基因功能的潜在病毒插入位点或可动因子的检测。对于测量质量来说，参考分数或与参考的相似性都不是绝对必需的。

实施例 10：

[0418] 1. 引物设计：引物可以设计成在靶序列或基因座的上游。第一引物可以包含：基因座特异性序列和 5’ 序列，其中 5’ 序列可以包含所正在使用的测序平台的第一衔接子序列的全部或部分（即，TELA 引物）。第二引物（即“探针”）可以包含：由 8 个核苷酸组成的随机序列和 3’ 序列，其中 3’ 序列包含所正在使用的测序平台的第二衔接子序列的全部或部分（即，TELA 探针）。

[0419] 2. 杂交：TELA 引物可以与特异性序列杂交。特异性序列可以在目标基因座附近。在一些实施方案中，特异性序列可以只是在目标基因座外部。TELA 探针可以与整个基因组中的随机序列杂交。TELA 探针杂交的间隔可以通过调节 TELA 探针浓度来调整。TELA 引物和 TELA 探针可以与相同模板链杂交。

[0420] 3. 延长：非置换聚合酶在其到达 DNA 模板上的第二引物之前可以用于延伸第一引物。DNA 连接酶可将 TELA 引物与 TELA 探针连接起来（即，连结或连接）。获得的产物可由侧翼为衔接子尾的 DNA 模板的单链拷贝组成。

[0421] 4. 扩增：可利用 PCR 扩增该产物。在该产物不包含完整（即，全长）衔接子序列的情况下，加尾引物扩增可引入第一和第二测序仪衔接子（即，有时被称为 A 衔接子和 B 衔接子）的剩余物。扩增可以产生测序仪就绪的文库。任选地，可以将该产物纯化或进一步处理成测序仪就绪的。

[0422] 任选地，可以同 TELA 引物、TELA 探针一起和 / 或在最终 PCR 加尾扩增的过程中加入一种或多种分子条形码。

[0423] 5. 测序：可以对文库进行测序，并且可以存储和 / 或传送数据以供分析。

[0424] 6. 数据分析：可以分析来自测序的数据。这种方法的优势包括在每次读取开始时由合成序列确定测序仪读取值的基因组位置的能力。另外，由于测序仪读取值的 3’ 末端是随机产生的（即，通过随机结合的 TELA 探针而产生的），所以可以避免或减少克隆错误。如果一个变体是从具有不同 3’ 末端的多个测序仪读取值中检测出来的，则这可能是真实的或

基因组的变体，而非与测序仪或读取变体。如果错误在扩增过程中发生，则其可能只在具有相同的 3’末端的读取值中出现。

实施例 11：

[0425] 测序文库按照标准文库制备方法来制备。两个样品用 Truseq 条形码 #5 和 #6 进行条形码化。

[0426] 通过以下步骤制备样品 A：使用 Covaris 剪切法进行片段化（即，剪切）。对片段化的样品进行：末端修复、A- 加尾和衔接子连接。继而，磁珠纯化两次——索引 5，将所有制备的样品加入方案中。

[0427] 样品 B：在 covaris 上片段化，末端修复，A- 加尾和衔接子连接，然后磁珠纯化 2 次，继而 PCR 循环该材料 12 次，然后磁珠纯化，之后在 2％琼脂糖凝胶上进一步纯化，将约 350-420bp 的条带切割、切下，并用琼脂糖溶解缓冲液和 zymo-25 柱进行分离——索引 6，约 200ng 输入。

[0428] 将所有 DNA 样品转移至 0.5mL 的 LoBind 管 (Eppendorf) 中。加入 1 纳摩尔的各种 3’封闭的衔接子封阻剂寡核苷酸（各 10μL，100μM，通用封阻剂 1、通用封阻剂 2、封阻剂 1 和封阻剂 2)。通过真空离心蒸发浓缩进行干燥并重悬。

[0429] 5μL 10X 标准 Taq 缓冲液 (60％)5X 生物素 -dNTP

[0430] 10μL 10X 60％生物素 -dNTP 6.25μL dA、dC、dG(20mM)

[0431] 1μL 60 引物混合物 (25μM) 15μl bio-dUTP(5mM)

[0432] 32μL 不含核酸酶的水 2.5μL dT(20mM)

[0433] 1μL 100mM 的 $MgCl_2$ 溶液 13.75μL nH_2O

[0434] 1μL Taq(5U/μL)

[0435] 共计： 50μL

[0436] 加热至 98℃ 2min。在热循环仪上采用斜降缓慢冷却至 47℃，然后至 72℃ 10 分钟，然后加入 1μL 的 0.5M EDTA 以猝灭，并置于冰上。

[0437] 文库的固定化

[0438] 洗涤 Dynal M280 链酶亲和素珠 (B&W-100μL，1M Tris pH 7.5，20μL 0.5M EDTA，4mL 5M NaCl，用 nH_2O 填充至 10mL)。将 25μL 珠子转移至新的 0.5mL LoBind 管中。用 B&W 缓冲液（含有 10ng/μL tRNA) 洗涤链霉亲和素磁珠 3 次，每次洗涤 200μL（移液混合）。最后一次洗涤等待 5 分钟后再取出。（使用 MPC 使珠子沉淀并除去缓冲液）。加入 50μL 的 B&W 缓冲液（无 tRNA)。将 50μL 的 DNA 转移至 50μL 的 Dynal M-280 链霉亲和素珠中，或者反过来亦可（可以使用额外 20μL B&W 缓冲液冲洗管）。充分混合并室温温育 1 小时，每 15 分钟混合一次。使用磁体，将固定化的文库用 200μL B&W 缓冲液洗涤 3 次并用 200μL 1X 柠檬酸钠盐水 (SSC) 洗涤 2 次。利用最后一次洗涤将珠子转移至 PCR 管。除去最后一次洗涤的所有剩余的 1X SSC 缓冲液。现在珠子应该以扩增就绪的状态处于 PCR 管中。

[0439] 通过 PCR 进行富集：在 15 个和 25 个循环之间（在 15 个循环时取出 50μL 并且使剩余物结束于 25 个循环）。循环

[0440] 向珠子加入： 98℃，45 秒 - 保持

[0441] 50μL 2X HiFi KAPA 就绪混合物 (KAPA Biosystems)98℃，15 秒

[0442] 2μL PCR 1 60℃，30 秒

[0443] 2μL PCR 2 72℃,30 秒

[0444] 46μL nH_2O 72℃,60 秒 - 保持

[0445] 100μL 4℃ - 保持

[0446] 清洗 DNA(例如,Zymo-25 柱),用 30μL 洗脱液进行洗脱,通过真空离心蒸发浓缩进行干燥,重悬 20μL,加载到 2%琼脂糖凝胶上。切下 320-400 个核苷酸的条带。将材料循环 25 次产生足以进行测序的 DNA 产物。扩增后的 DNA 在 2%琼脂糖上进行纯化,并将约 350-420 个核苷酸的条带切下、切出,并用琼脂糖溶解缓冲液和 zymo-25 柱进行分离。

实施例 12:

[0447] 制备文库:(12 个循环的 PCR)。两个样品,条形码 5,735ng 和条形码 6,684ng。将 30μL DNA 样品转移至 0.5mL LoBind 管。加入 1nmol 的各种 3' 封闭的衔接子封阻剂寡核苷酸(各 10μL,100μM,通用封阻剂 1、通用封阻剂 2、封阻剂 1 和封阻剂 2)。通过真空离心蒸发浓缩进行干燥。

[0448] 重悬:

[0449] 5μL 10X 标准 Taq 缓冲液 (60%)5X 生物素 -dNTP

[0450] 10μL 10X 60%生物素 -dNTP 6.25μL dA、dC、dG(20mM)

[0451] 1μL 60 引物混合物 (25μM) 15μl bio-dUTP(5mM)

[0452] 32μL 无核酸酶的水 2.5μL dT(20mM)

[0453] 1μL 100 mM 的 $MgCl_2$ 溶液 13.75μL nH_2O

[0454] 1μL Taq(5U/μL)

[0455] 共计: 50μL

[0456] 加热至 98℃ 2min。或者在热循环仪上利用斜降缓慢冷却至 47℃或者快速冷却至 47℃,然后在 47℃保持 4 分钟,继而快速斜升至 72℃ 10 分钟,然后加入 1μL 的 0.5M EDTA 以猝灭。置于冰上。

[0457] 文库固定化

[0458] 洗涤 Dynal M280 链霉亲和素珠:

[0459] (B&W-100μL,1M Tris pH7.5,20μL 0.5M EDTA,4mL 5M NaCl…用 n- 水加至 10mL)。将 25μL 珠子转移至新的 0.5mL LoBind 微量离心管中。用 B&W 缓冲液(含有 10ng/μL tRNA)洗涤链霉亲和素磁珠 3 次,每次洗涤 200μL(移液混合)。最后一次洗涤等待 5 分钟后再取出。使用 MPC 使珠子沉淀并除去缓冲液。加入 50μL 的 B&W 缓冲液(无 tRNA)。将 50μL 的 DNA 转移至 50μL 的 Dynal M-280 链霉亲和素珠中,或者反过来亦可。可以使用额外 20μL B&W 缓冲液冲洗管。充分混合并室温温育 1 小时,每 15 分钟混合一次。使用磁体,将固定化的文库用 200μL B&W 缓冲液洗涤 3 次并用 200μL 1XSSC 洗涤 2 次。使用最后一次洗涤将珠子转移至 PCR 管。除去最后一次洗涤的所有剩余 1X SSC 缓冲液。现在珠子应该以扩增就绪的状态处于 PCR 管中。

[0460] 通过 PCR 进行富集—15 个和 25 个循环

[0461] 向珠子加入: 98℃,45 秒 - 保持

[0462] 50μL 2X HiFi KAPA 就绪混合物 98℃,15 秒

[0463] 2μL PCR 1 60℃,30 秒

[0464] 2μL PCR 2 72℃,30 秒

[0465] 46 μL nH_2O 72℃,60 秒 - 保持

[0466] 100 μL 4℃ - 保持

[0467] 清洗 DNA(例如,Zymo-25 柱),用 30 μL 洗脱,通过真空离心蒸发浓缩进行干燥,重悬 20 μL,加载到 2% 琼脂糖凝胶上。切下 320-400 个核苷酸的区域并凝胶提取 DNA。DNA 可进一步处理并测序。

[0468] 通用封阻剂 1 :

GATCGGAAGAGCACACGTCTGAACTCCAGTCAC555555ATCTCGTATGCCGTCTTCTGCTTGX

[0469] 通用封阻剂 2 :

CAAGCAGAAGACGGCATACGAGAT555555GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCX

[0470] 该方法可用于人类 CCLS 基因的约 10kb 区域。该方法还可与高度多重化的短 PCR 反应一起进行。

[0471] 实施例 13

[0472] PCR 反应可以进行 2 次。1 μL 约 1.2 μg 的 gDNA(例如人脑 DNA)。2 μL 的 60 引物混合物 (25 μM 原液) 可与以下成分一起加入 :

[0473] 25 μL 2X KAPA HiFi

[0474] 22 μL nH_2O

[0475] 共计 :50 μL

[0476] PCR—10 个循环

[0477] 98℃,45 秒

[0478] 98℃,15 秒

[0479] 47℃,30 秒

[0480] 72℃,30 秒

[0481] 72℃,1 分钟

[0482] 4℃ - 保持

[0483] 清洗 DNA(例如,Zymo-25 柱),用 30 μL 洗脱并干燥 (例如,使用真空离心蒸发浓缩)。将 DNA 加载到 2% 琼脂糖上,2 个孔,凝胶纯化 300-600 个核苷酸的片段。从凝胶上洗脱,使用例如 Zymo-25 进行分离,用 30 μL 水洗脱,通过真空离心蒸发浓缩进行干燥。

[0484] 片段化 :在 50 μL 1X TE(含 2ng/ μL tRNA) 中重悬所有 DNA。加至玻璃 Covaris 管中。采用以下设置 :占空比 10%,强度 5,循环 / 猝发 200,时间 120 秒以剪切 DNA。

[0485] 50 μL 片段化的 DNA,20 μL 末端修复混合物 (8 μL 水,7 μL 10X KAPA 末端修复,5 μL KAPA 酶)。在 20℃温育 30 分钟。清理,将 120 μL AmpureXP 珠子加至 70 μL = 190 μL。充分混合,温育 10 分钟让 DNA 结合。将管移至磁体上,去除液体,3 分钟。用 200 μL 80% 乙醇洗涤珠子,等待 30 秒,除去,重复,共洗涤 2 次。使珠子干燥 10min。

[0486] 加入来自以上的珠子。使珠子再水合 3 分钟。50 μL A- 加尾主混合物 (42 μL 水,5 μL 10X KAPA A- 加尾,3 μL KAPA A- 加尾酶)。充分混合并在 30℃温育 30 分钟。通过添加 90 μL 20% PEG8000/2.5M NaCl 溶液进行清洗。总体积可以是 140 μL,通过移液充分混合。温育 10min 让 DNA 结合。将管移至磁体上,去除液体。用 200 μL 80% 乙醇洗涤珠子,等待 30 秒,除去,重复,共洗涤 2 次。使珠子干燥 5 分钟。连接衔接子。加入来自以上的珠子,使珠子再水合 3min。

[0487] 45μL连接主混合物(30μL水,10μL 5X KAPA Lig.,5μL T4DNA连接酶)。

[0488] 3μL衔接子(标准的1:10稀释液)(反应中的最终衔接子浓度=0.3μM)。

[0489] 充分混合并在20℃温育15分钟。洗涤2次。加入50μL 20%PEG8000/2.5M NaCl溶液。总体积100μL,通过移液充分混合。温育10分钟让DNA结合。将管移至磁体上,去除液体。用200μL 80%乙醇洗涤珠子,等待30秒,除去,重复,洗涤2次。使珠子干燥5分钟。在50μL水中重悬珠子,等待5分钟。加入50μL 20% PEG8000/2.5M NaCl溶液。温育10min让DNA结合。将管移至磁体上,去除液体。用200μL 80%乙醇洗涤珠子,等待30秒,除去,重复,共洗涤2次。使珠子干燥5分钟。在23μL Tris pH 8中重悬珠子,使其再水合3分钟。从洗脱液中收集DNA并转移至PCR管。通过PCR进行富集。在500μL薄壁PCR管中制备下列PCR反应混合物。

[0490] 使用来自以上的DNA 23μL

[0491] 2x HiFi KAPA主混合物 50μL

[0492] Truseq PCR引物1 2μL

[0493] Truseq PCR引物2 2μL

[0494] 含核酸酶的水 23μL

[0495] 运行下列PCR方案10个循环:

[0496] 98℃ 45秒

[0497] 10个循环的:

[0498] 98℃,15秒

[0499] 60℃,30秒

[0500] 72℃,30秒

[0501] 72℃ 1分钟

[0502] 4℃保持

[0503] 纯化DNA(例如,使用Zymo(25)PCR纯化试剂盒)并用30μL dH_2O洗脱。每个PCR 2个孔,在4%琼脂糖凝胶上对材料进行10个循环和20个循环的纯化。

[0504] 虽然本文已经表明和描述了本发明的优选实施方案,但本领域技术人员将会明白,这些实施方式仅以实例的方式提供。在不背离本发明的情况下,本领域技术人员将会想到许多变更、变化和替换。应当理解,本文描述的本发明实施方案的许多替代方案可用于实施本发明。下面的权利要求书旨在限定本发明的范围并且因此涵盖这些权利要求范围内的方法和结构及其等同物。

[0505] 虽然本文已经表明和描述了本发明的优选实施方案,但本领域技术人员将会明白,这些实施方式仅以实例的方式提供。在不背离本发明的情况下,本领域技术人员将会想到许多变更、变化和替换。应当理解,本文描述的本发明实施方案的许多替代方案可用于实施本发明。下面的权利要求书旨在限定本发明的范围并且因此涵盖这些权利要求范围内的方法和结构及其等同物。

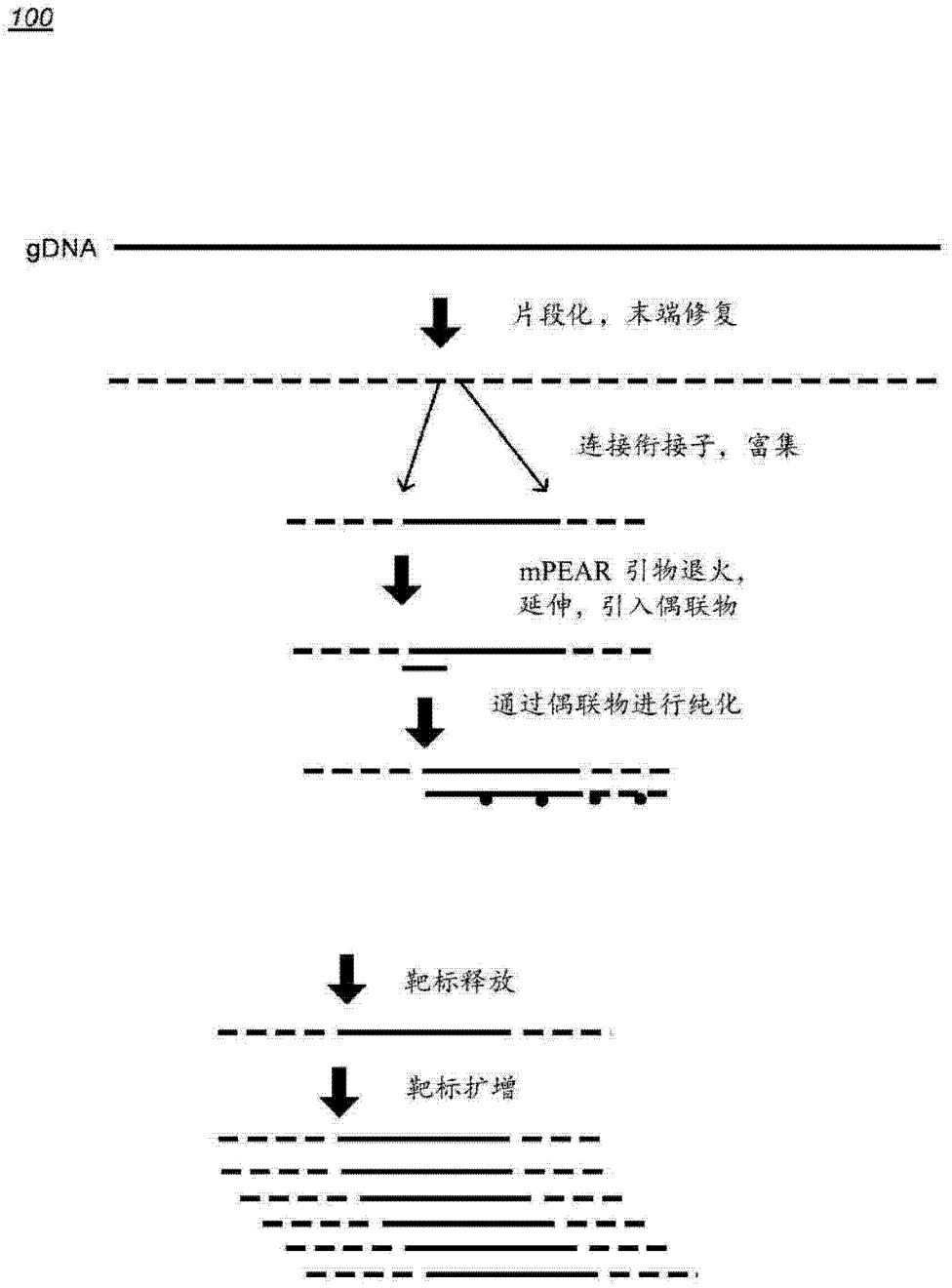

100
gDNA
片段化，末端修复
连接衔接子，富集
mPEAR 引物退火，
延伸，引入偶联物
通过偶联物进行纯化
靶标释放
靶标扩增

图 1A

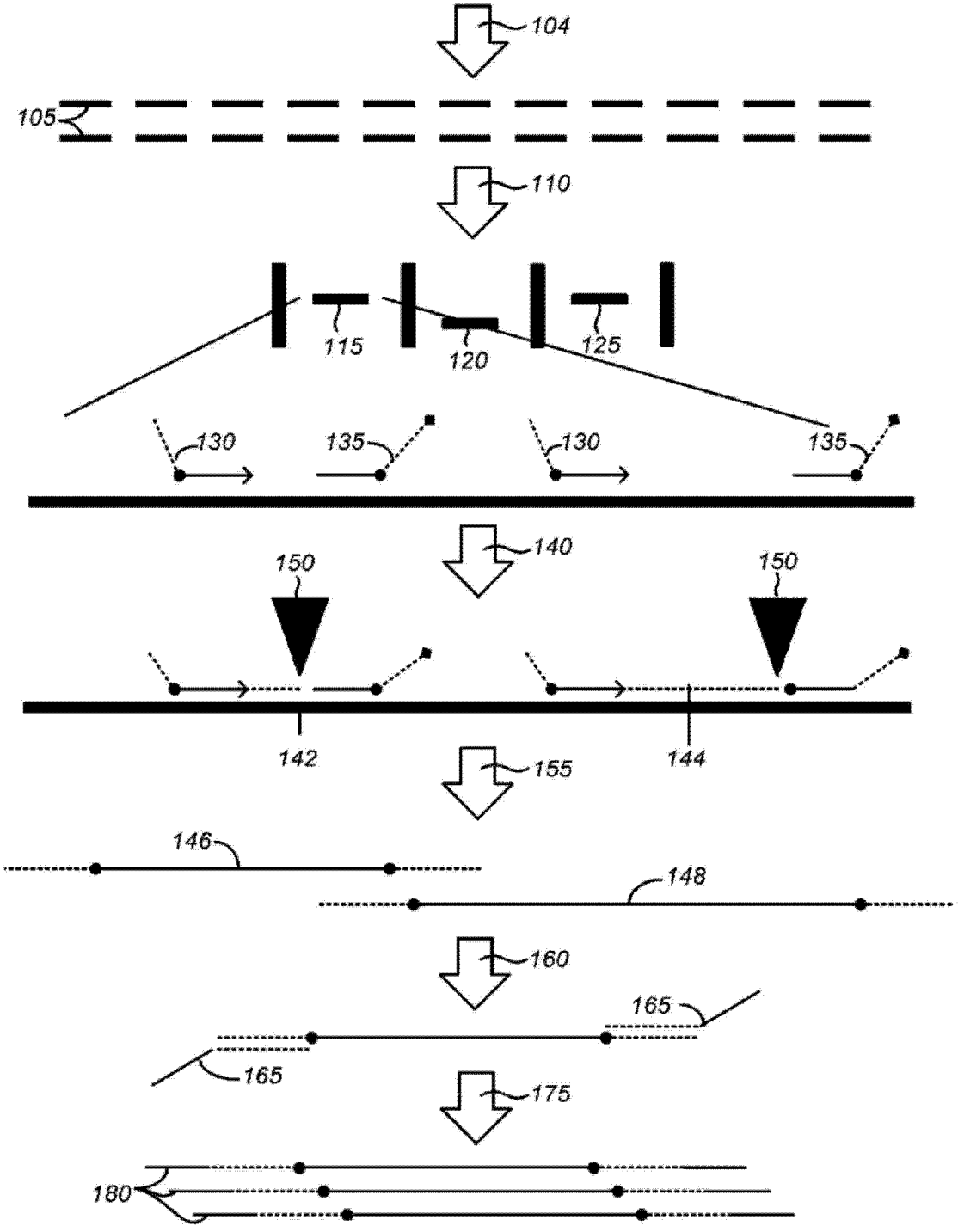

104
105
110
115
120
125
130
135
130
135
140
150
150
142
144
155
146
148
160
165
165
175
180

图 1B

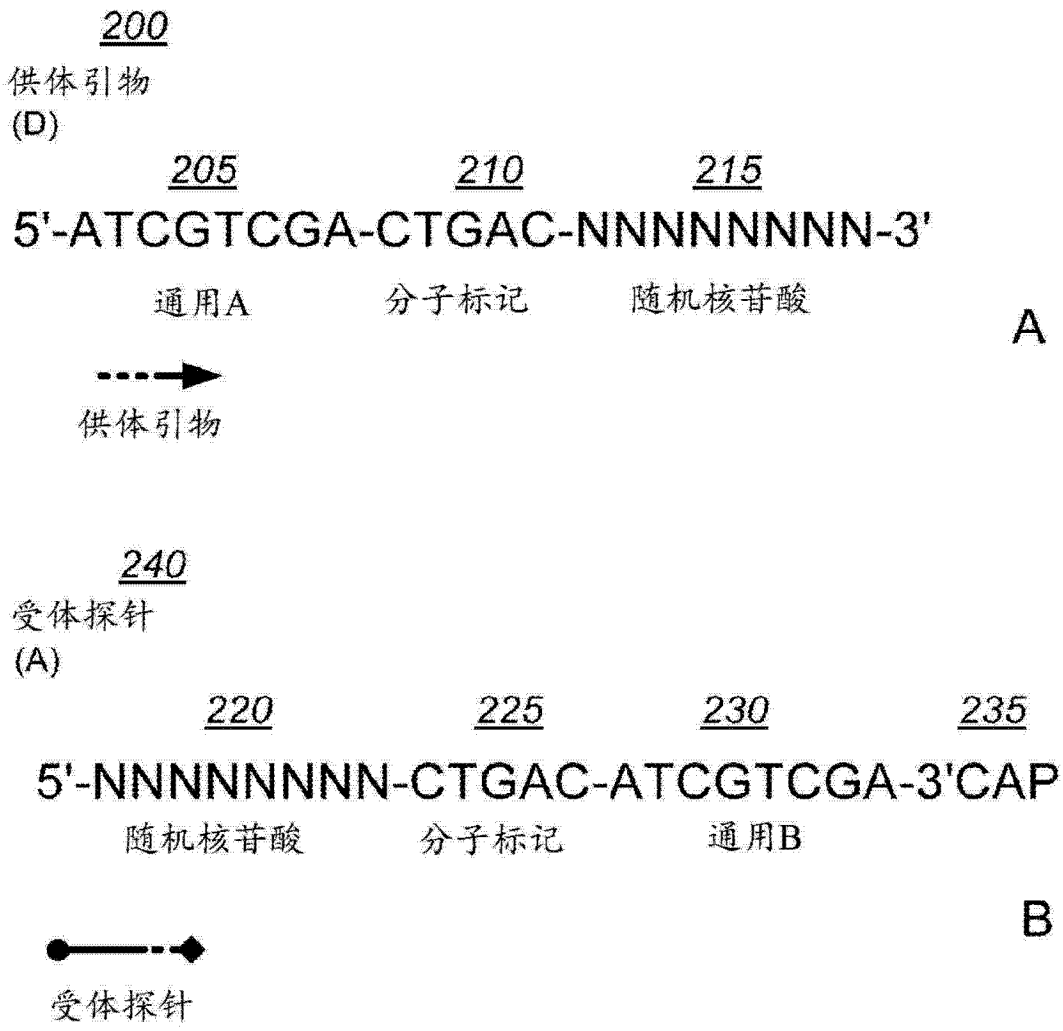

200
供体引物
(D)
205
210
215
5'-ATCGTCGA-CTGAC-NNNNNNNN-3'
通用A
分子标记
随机核苷酸
A
供体引物
240
受体探针
(A)
220
225
230
235
5'-NNNNNNNN-CTGAC-ATCGTCGA-3'CAP
随机核苷酸
分子标记
通用B
B
受体探针

图 2

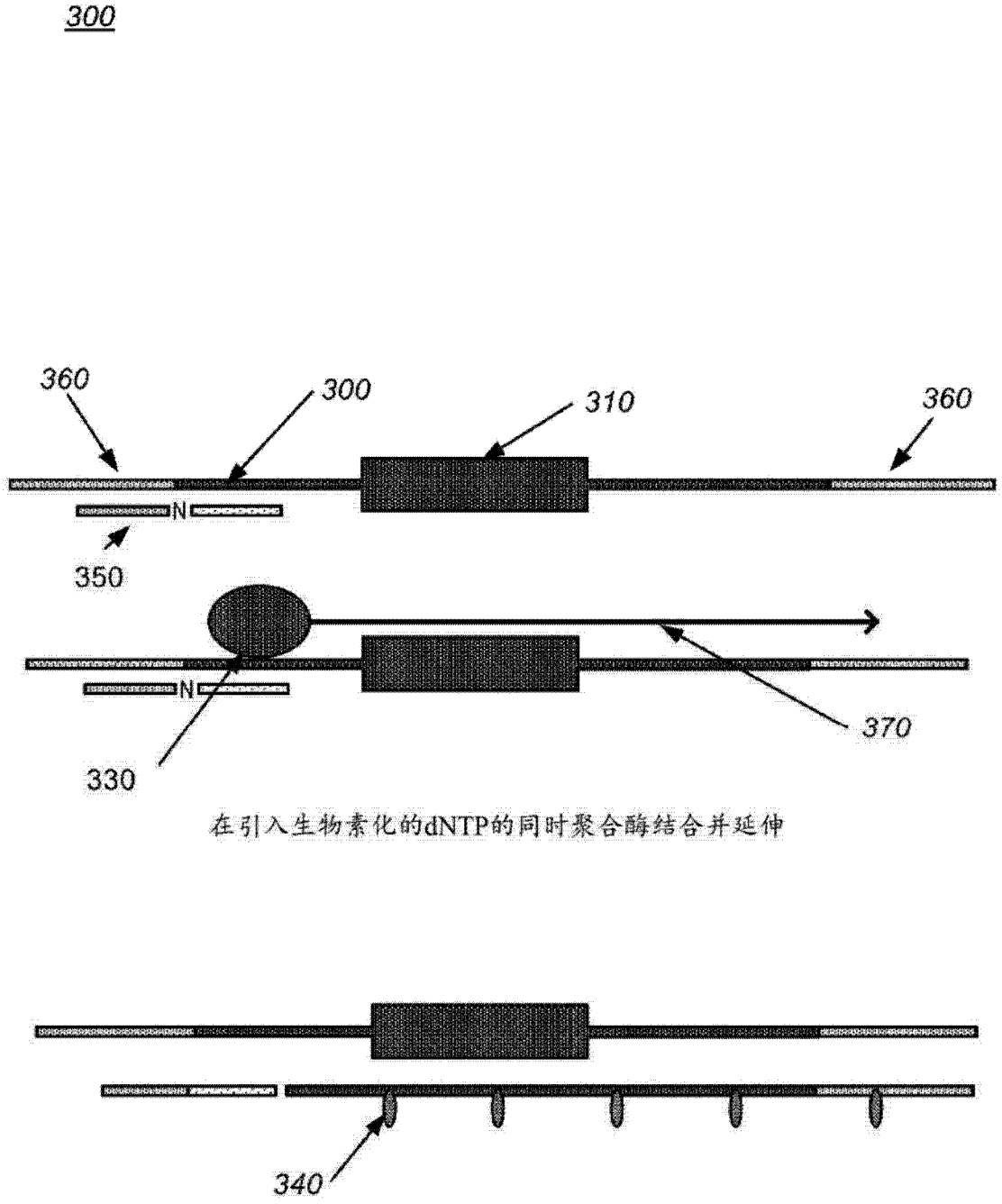

300
360
300
310
360
N
350
N
370
330
在引入生物素化的dNTP的同时聚合酶结合并延伸
340

图 3A

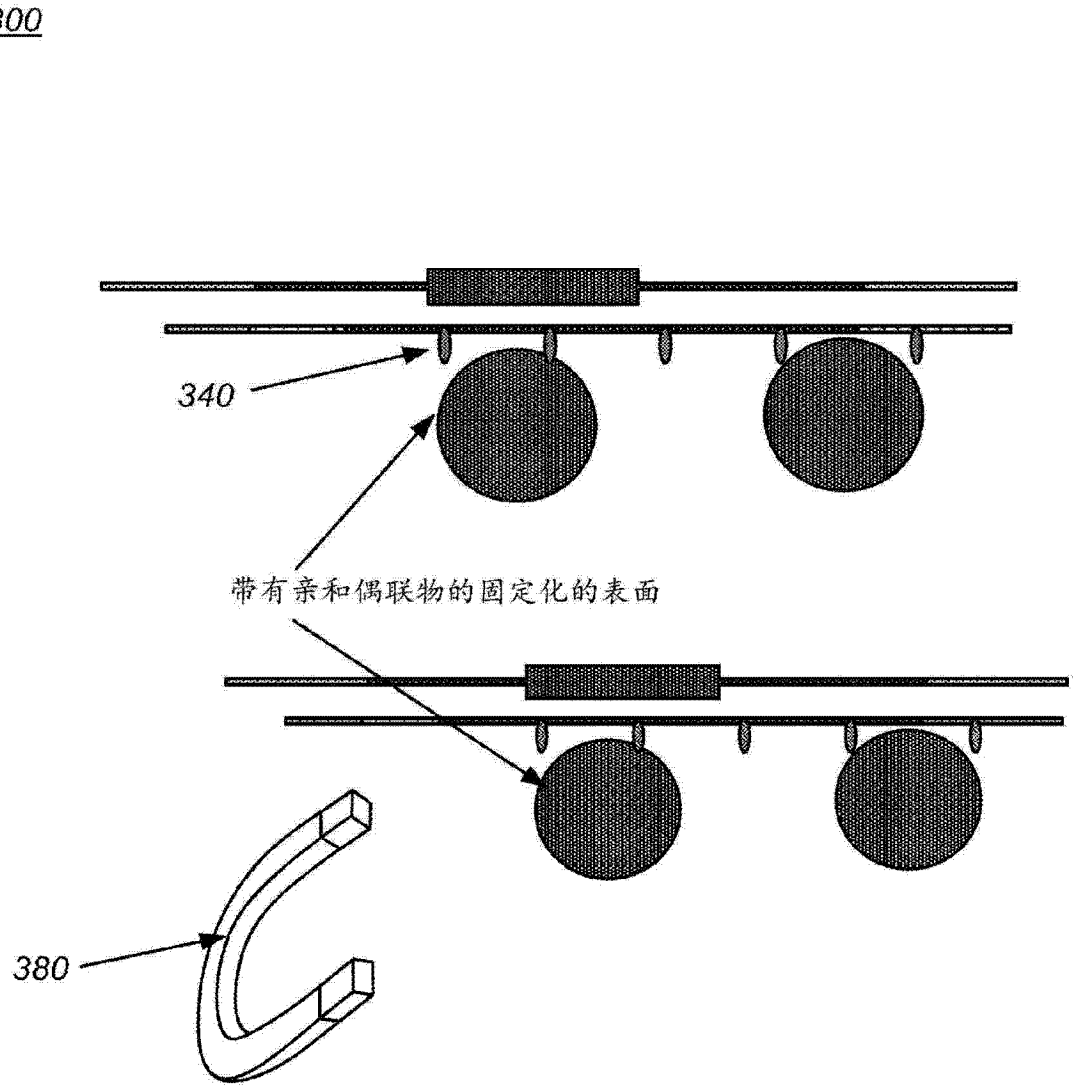


图 3B

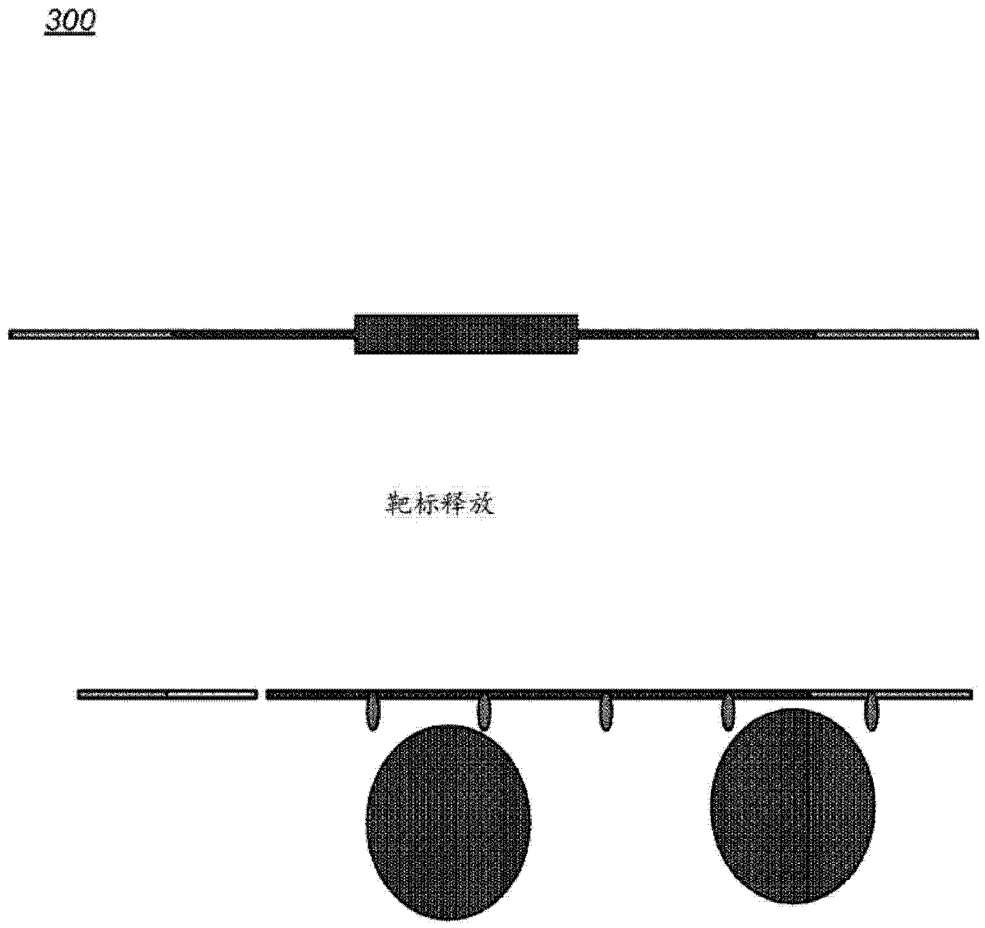


图 3C

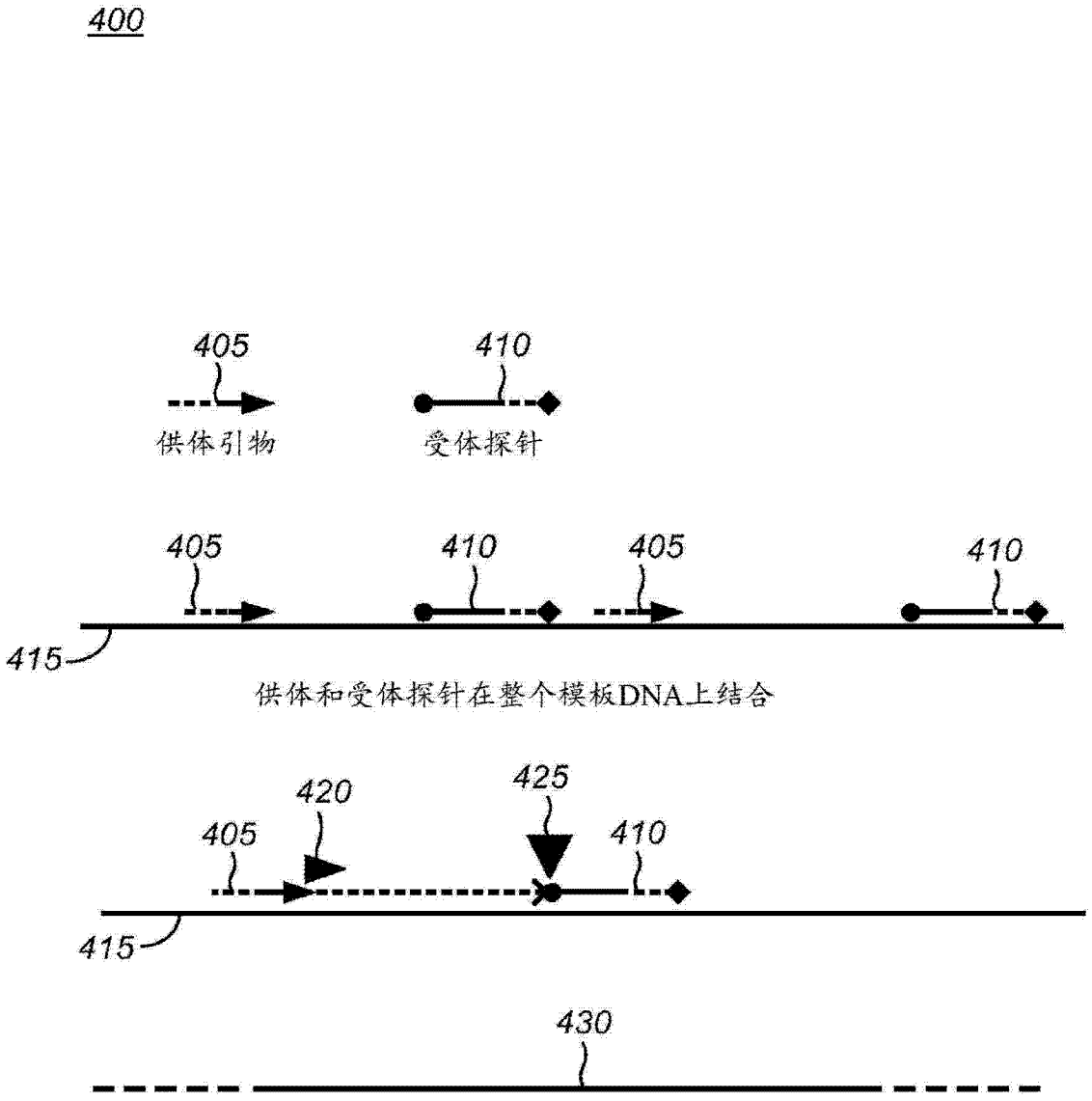

400
405
410
供体引物
受体探针
405
410
405
410
415
供体和受体探针在整个模板DNA上结合
420
425
405
410
415
430
RAPELL产物具有通用衔接子互补物和分子标记

图 4

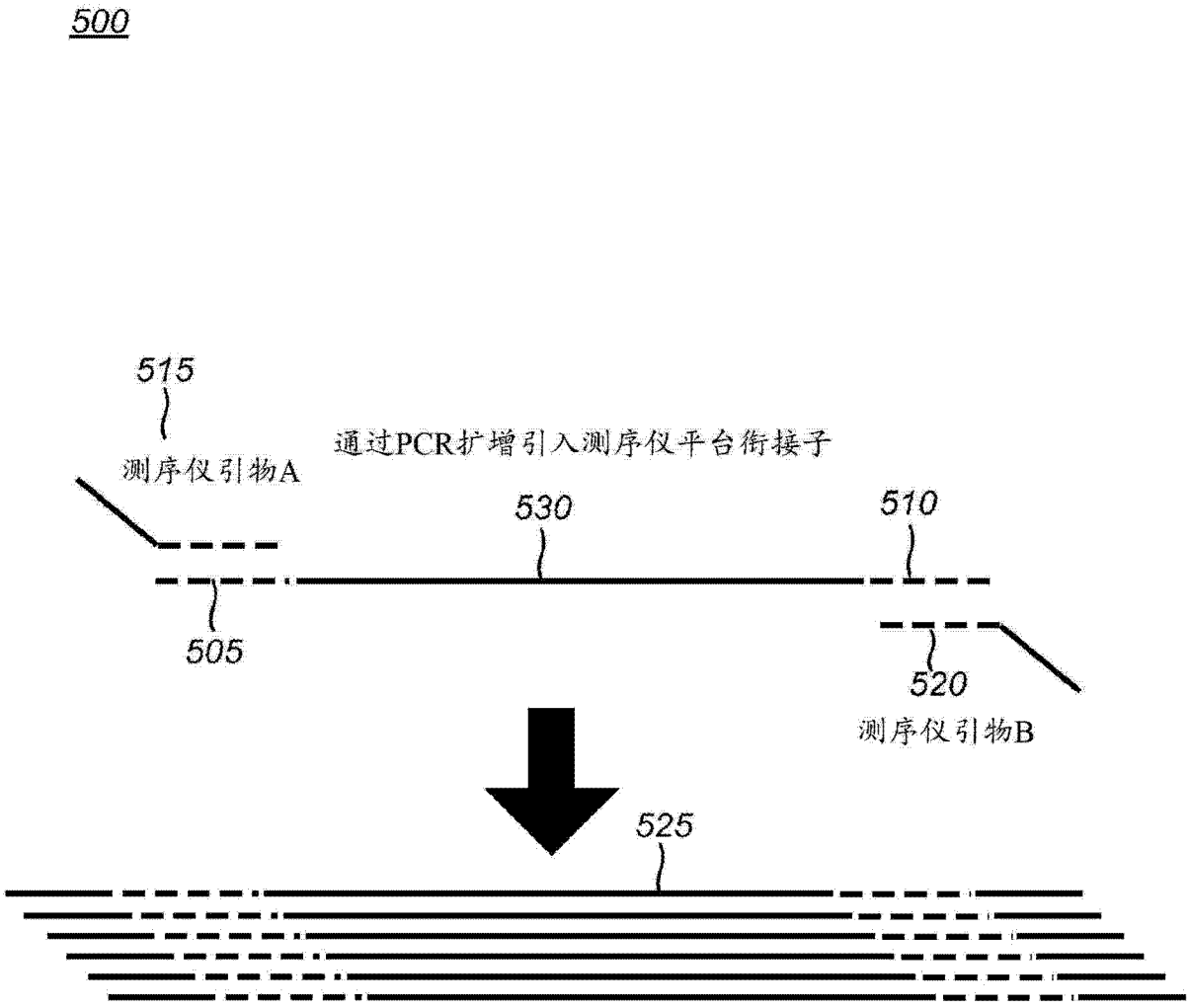

500
515
测序仪引物A
通过PCR扩增引入测序仪平台衔接子
530
510
505
520
测序仪引物B
525

图 5

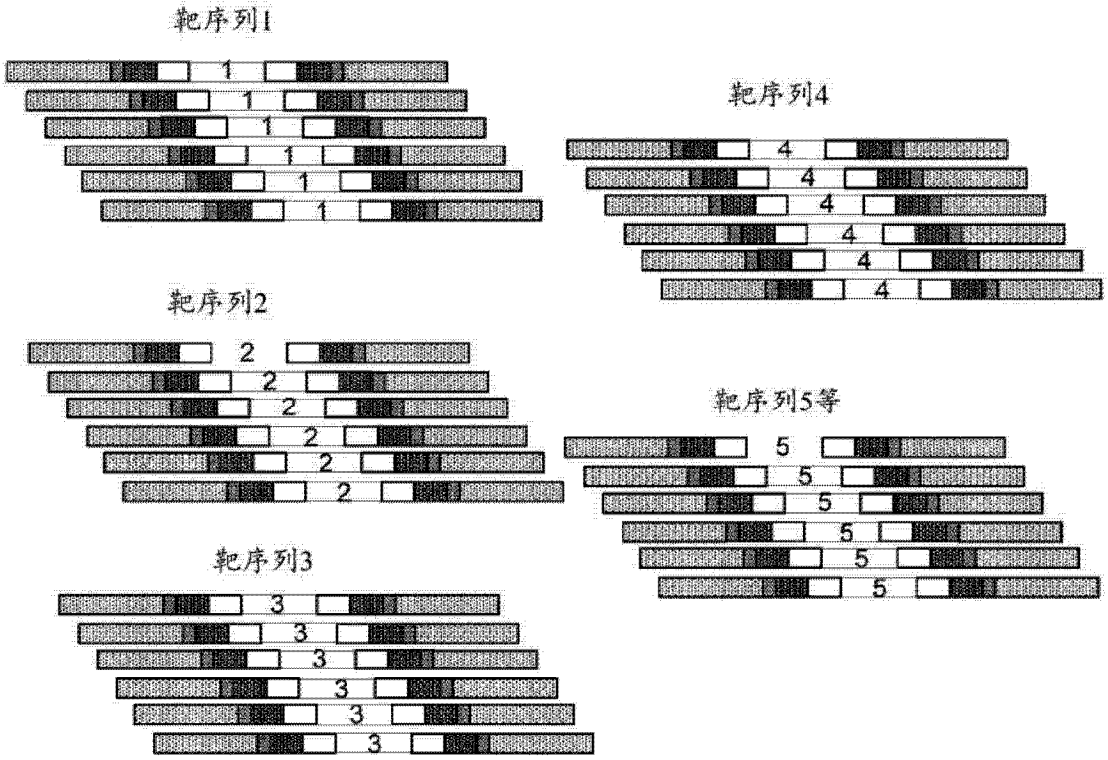

靶序列1
1
1
1
1
1
1
靶序列4
4
4
4
4
4
4
靶序列2
2
2
2
2
2
2
靶序列5等
5
5
5
5
5
5
靶序列3
3
3
3
3
3
3

图 6

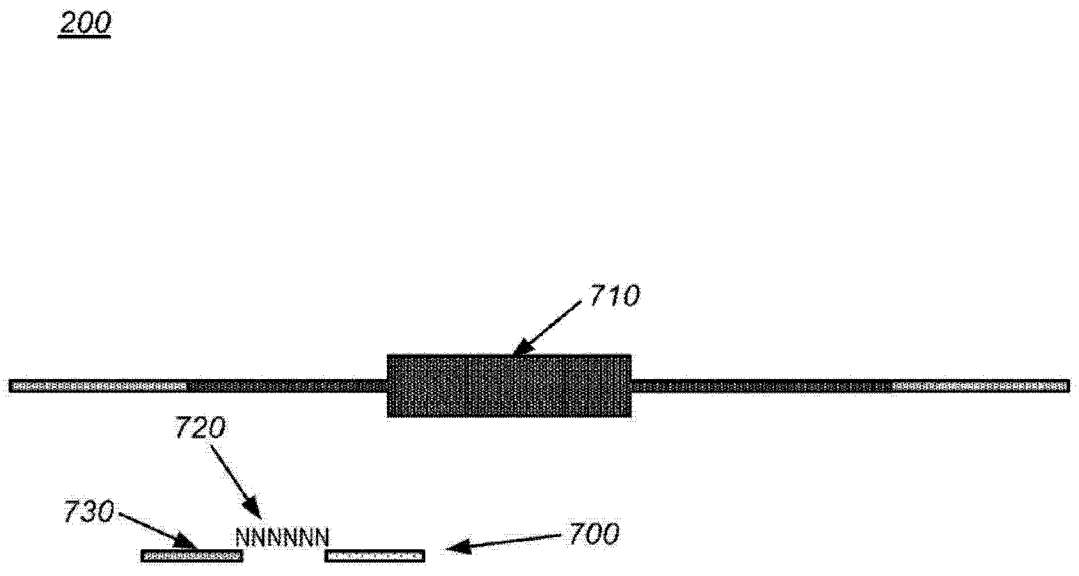
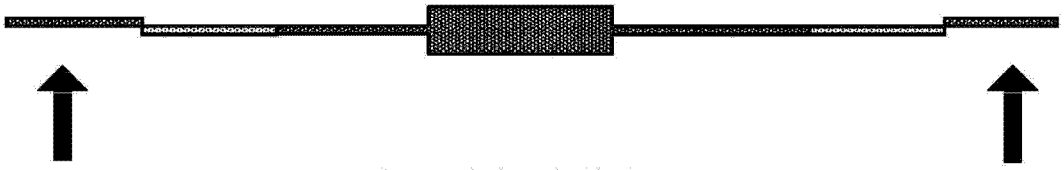
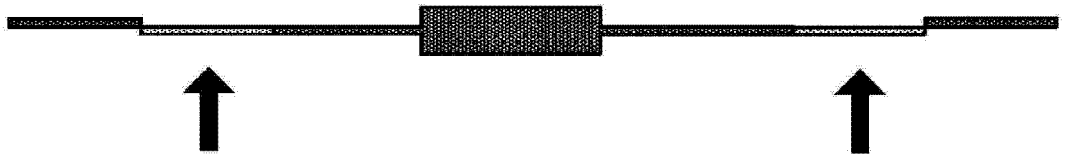

200
710
720
730
NNNNNN
700

图 7

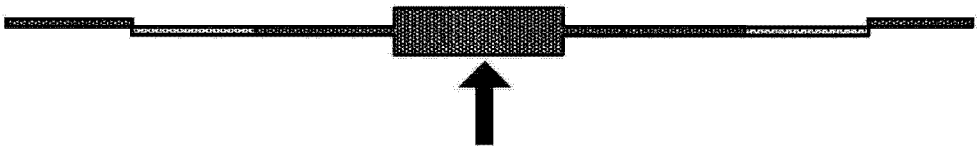


测序仪特异性衔接子



针对样品ID的分子条形码

用于“种入”读取位置的基因座特异性结合位点



靶标可被“从头”测序

图 8

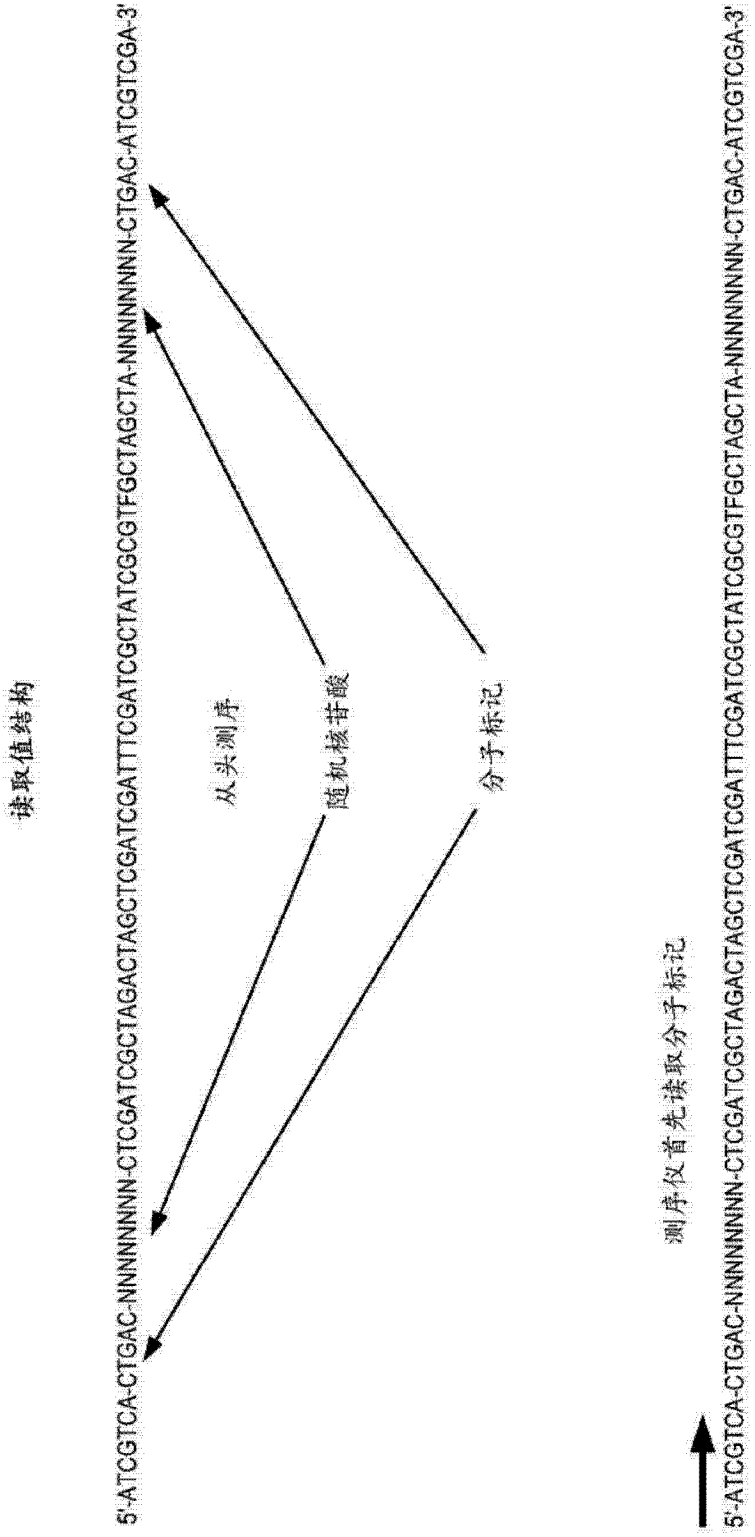

读取值结构
5'-ATCGTCA-CTGAC-NNNNNNNN-CTCGATCGCTAGACTAGCTCGATCGATTTCGATCGCTATCGCGTFGCTAGCTA-NNNNNNNN-CTGAC-ATCGTCGA-3'
从头测序
随机核苷酸
分子标记
测序仪首先读取分子标记
5'-ATCGTCA-CTGAC-NNNNNNNN-CTCGATCGCTAGACTAGCTCGATCGATTTCGATCGCTATCGCGTFGCTAGCTA-NNNNNNNN-CTGAC-ATCGTCGA-3'

图 9

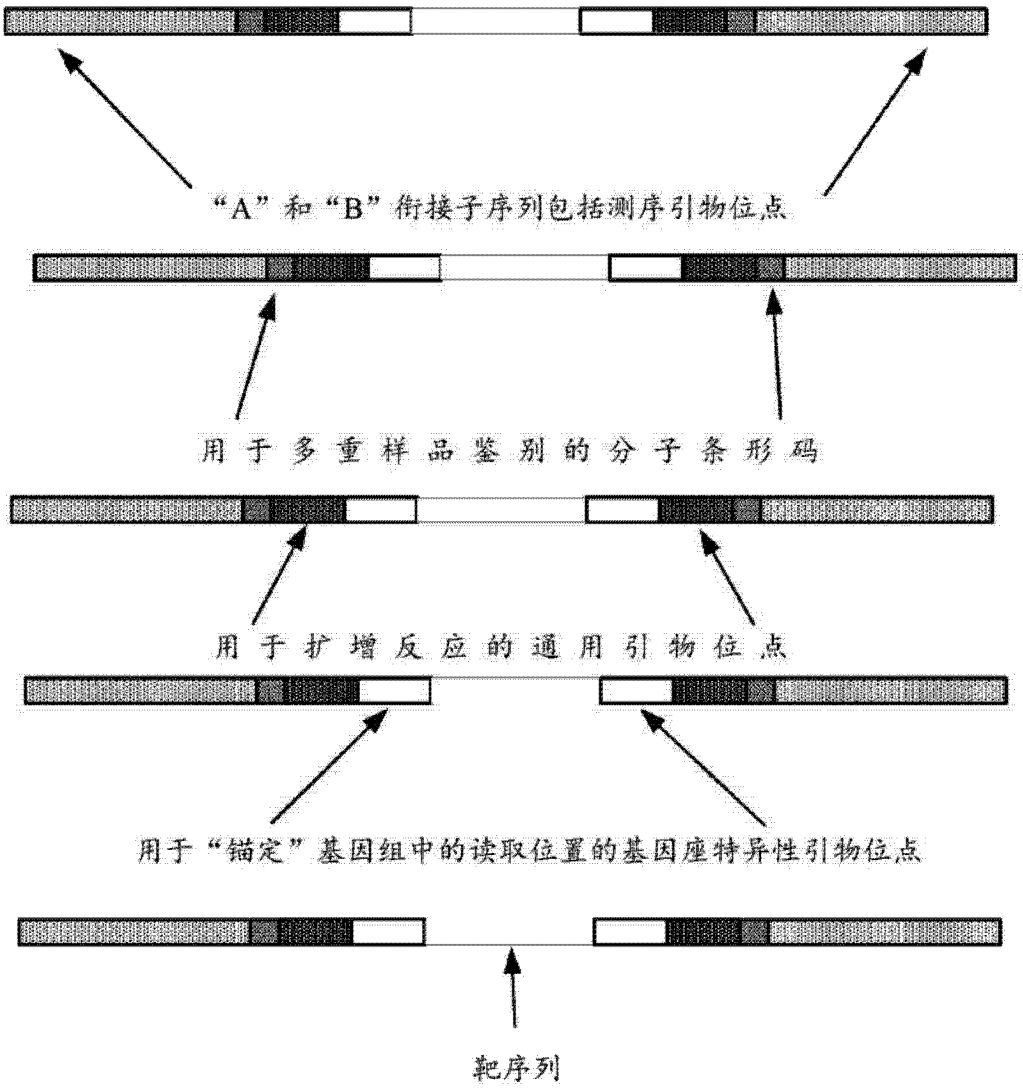

“A”和“B”衔接子序列包括测序引物位点
用于多重样品鉴别的分子条形码
用于扩增反应的通用引物位点
用于“锚定”基因组中的读取位置的基因座特异性引物位点
靶序列

图 10

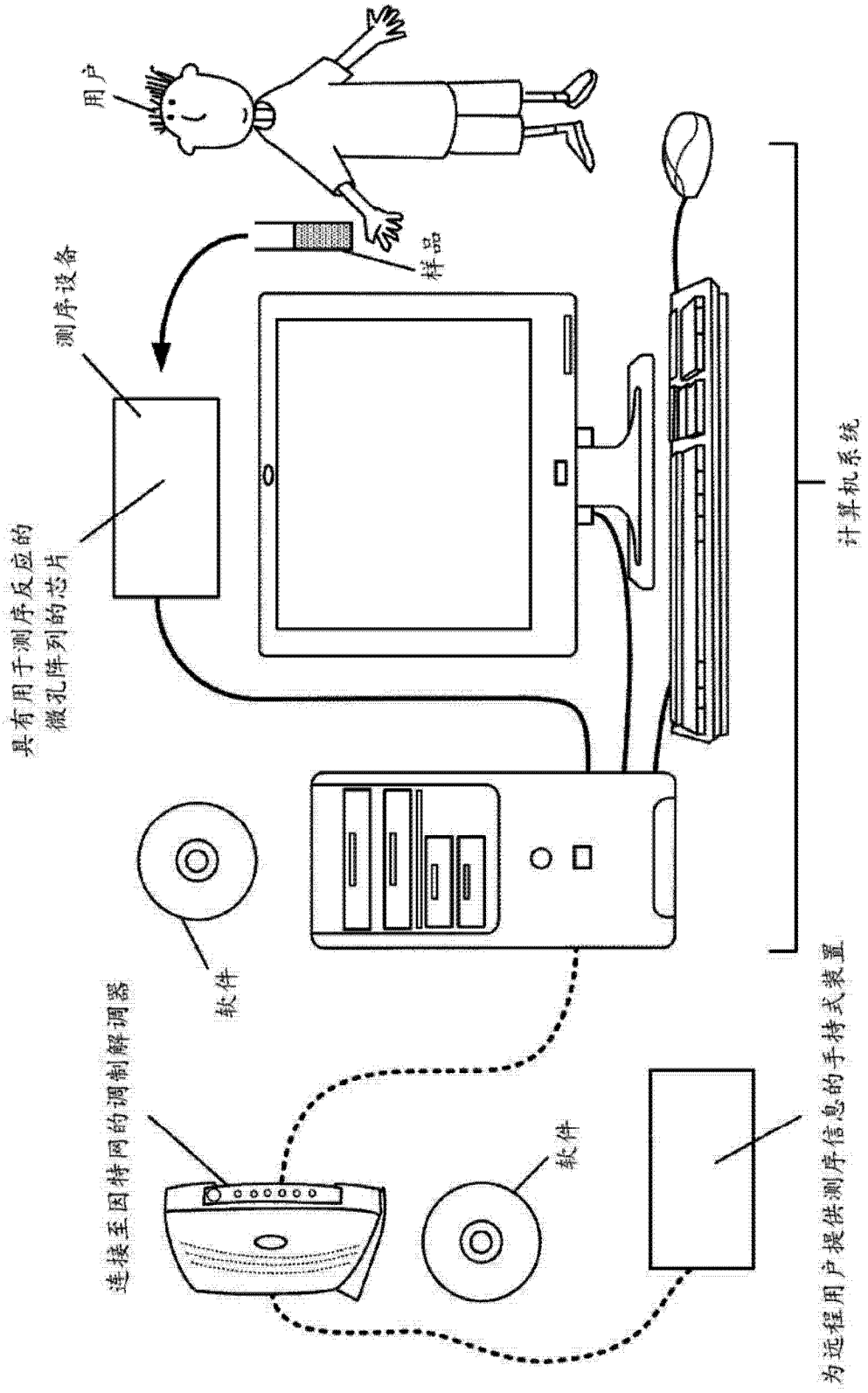

用户
样品
测序设备
具有用于测序反应的
微孔阵列的芯片
计算机系统
软件
连接至因特网的调制解调器
软件
为远程用户提供测序信息的手持式装置

图 11

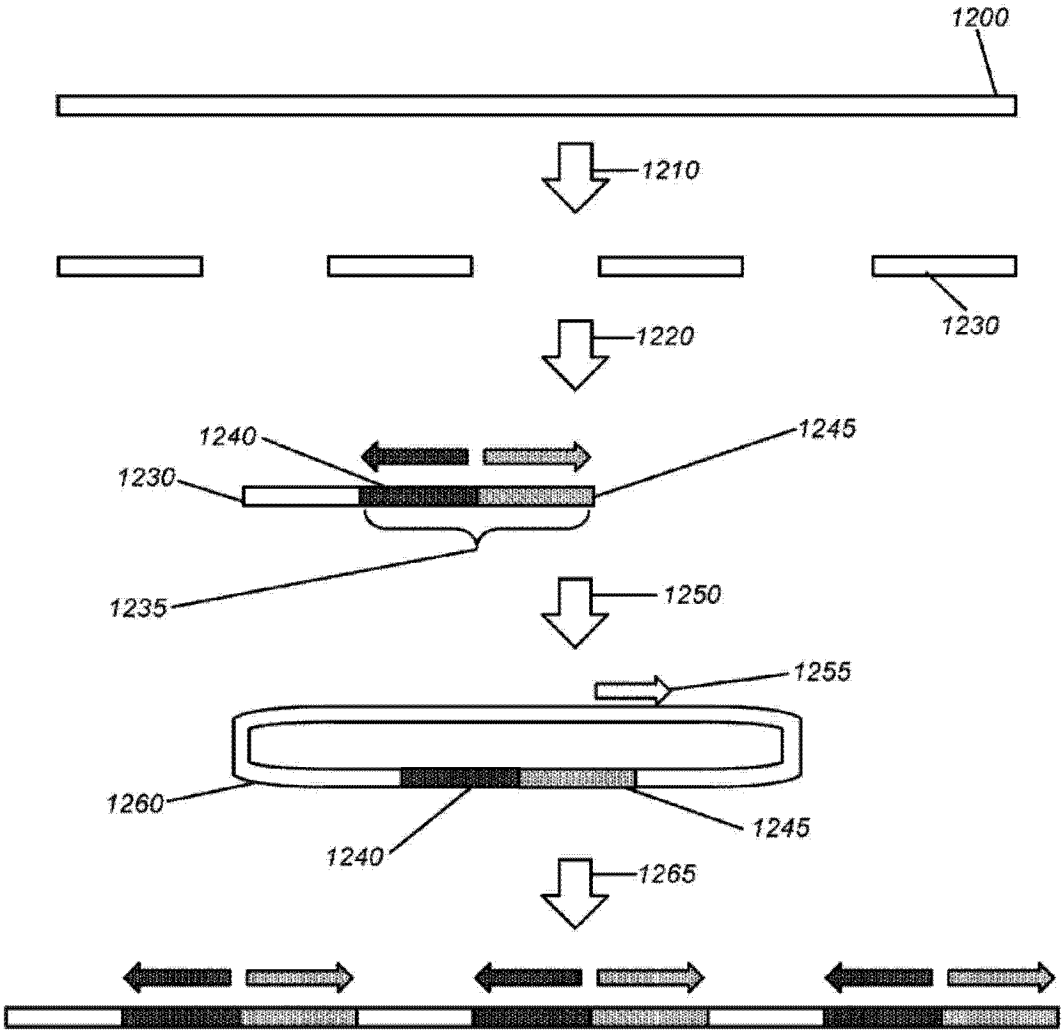

1200
1210
1230
1220
1240
1245
1230
1235
1250
1255
1260
1240
1245
1265

图 12

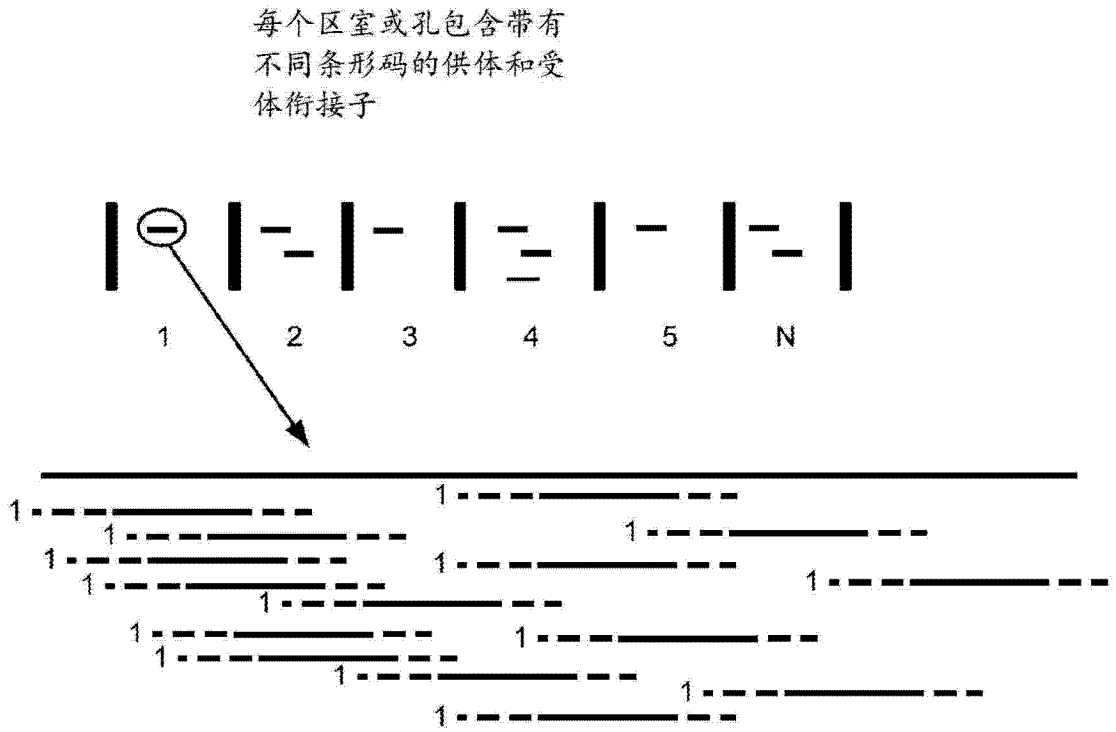

每个区室或孔包含带有
不同条形码的供体和受
体衔接子
1
2
3
4
5
N
基于分子标记（条形码），在计算机上从头组装短序列读取值

图 13

Truseq标准构建体 – 多重（红色=条形码）

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-XXX-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA-XXX-TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG

NNNNNNNATCTCGTATGCCGTCTTCTGCTTG

NNNNNNNTAGAGCATACGGCAGAAGACGAAC

定制三种赛核苷酸：Rapel 1和2以及PCR 1（我们已拥有PCR 2）。

Rapel-1:

条形码13

UUT-TCCCTACACGACGCTCTTCCGATCT-AGTCAA-N8-3'

-UT-AGGGATGTGCTGCGAGAAGGCTAGA-5'

Rapel-2:

条形码17

P-N8-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-GTAGAG-

ATCTCGTATGCCGTCTTCTGCTTG-amino-3'

PCR1

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

PCR2

CAAGCAGAAGACGGCATACGAGAT

IDT定制序列

Rapel 1: AGATCGGAAGAGCGTCGTGTAGGGAT/ideoxyU//ideoxyU//TTCCCTACACGACGCTCTTCCGATCTAGTC

AANNNNNNNN

Rapel 2: /5Phos/NNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCACGTAGAGATCTCGTATGCCGTCTTCTGCTTG

PCR 1: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

图 14