



(12)发明专利

(10)授权公告号 CN 105492607 B

(45)授权公告日 2019.07.02

(21)申请号 201480047858.1

(72)发明人 本杰明·辛德森

(22)申请日 2014.06.26

克里斯多佛·辛德森

(65)同一申请的已公布的文献号

迈克尔·史诺-莱文 凯文·奈斯

申请公布号 CN 105492607 A

米尔娜·雅罗什

(43)申请公布日 2016.04.13

唐纳德·马斯克利耶

(30)优先权数据

瑟奇·萨克森诺夫 兰登·梅里尔

61/840,403 2013.06.27 US

安德鲁·普莱斯 保罗·哈登伯尔

61/844,804 2013.07.10 US

媛·李

13/966,150 2013.08.13 US

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262

PCT/US2013/054797 2013.08.13 US

代理人 郑霞

61/896,060 2013.10.26 US

(51)Int.Cl.

61/909,974 2013.11.27 US

C12N 15/10(2006.01)

61/937,344 2014.02.07 US

C12Q 1/6806(2018.01)

61/940,318 2014.02.14 US

C12Q 1/6869(2018.01)

61/991,018 2014.05.09 US

(85)PCT国际申请进入国家阶段日

(56)对比文件

2016.02.29

WO 2010115154 A1, 2010.10.07,

(86)PCT国际申请的申请数据

WO 2012048341 A1, 2012.04.12,

PCT/US2014/044398 2014.06.26

US 20120316074 A1, 2012.12.13,

(87)PCT国际申请的公布数据

US 20120309002 A1, 2012.12.06,

W02014/210353 EN 2014.12.31

王兴春 等.高通量测序技术及其应用.《中

(73)专利权人 10X基因组学有限公司

国生物工程杂志》.2012,第32卷(第1期),第109-114页.

地址 美国加利福尼亚州

审查员 熊壮壮

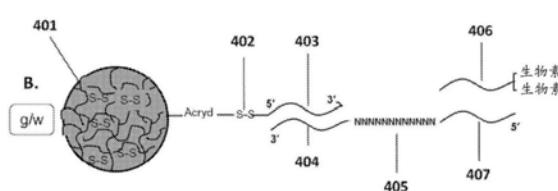
(54)发明名称

权利要求书5页 说明书93页 附图54页

用于样品处理的组合物和方法

第一分区释放至所述第二分区内。释放的条形码分子附接至所述第二分区内的样品材料的组分或其片段。所述第一分区可包含至少1,000,000个与之关联的条形码分子，所述条形码分子具有相同的条形码序列。

(57)摘要



用于样品处理，尤其是用于测序应用的方法和组合物。用于附接至含有条形码的大量寡核苷酸的珠子的多样性文库的珠子组合物。方法包括

B 制备条形码化珠子的文库的方法以及使珠子与样品组合的方法，例如通过使用微流体装置。第一分区包含与之关联的多个核酸条形码分子，且所述核酸条形码分子可包含相同的核酸条形码序列。所述第一分区可与样品材料的组分共同分配至第二分区，然后可将所述条形码分子从所述

CN 105492607 B

1. 一种对样品材料进行条形码化的方法,其包括:
 - (a) 提供珠子,其包含与之关联的多个核酸条形码分子,所述多个核酸条形码分子包含共同条形码序列;
 - (b) 将所述珠子与样品材料的组分共同分配至分区;
 - (c) 将所述多个核酸条形码分子中的核酸条形码分子从所述珠子释放至所述分区;以及
 - (d) 将所述核酸条形码分子附接至所述分区内的样品材料的一种或多种组分或其片段。
2. 根据权利要求1所述的方法,其中所述珠子包含至少1,000个与之关联的核酸条形码分子,所述至少1,000个核酸条形码分子具有共同条形码序列。
3. 根据权利要求1所述的方法,其中所述珠子包含至少10,000个与之关联的核酸条形码分子,所述至少10,000个核酸条形码分子具有共同条形码序列。
4. 根据权利要求1所述的方法,其中所述珠子包含至少100,000个与之关联的核酸条形码分子,所述至少100,000个核酸条形码分子具有共同条形码序列。
5. 根据权利要求1所述的方法,其中所述珠子包含至少1,000,000个与之关联的核酸条形码分子,所述至少1,000,000个核酸条形码分子具有共同条形码序列。
6. 根据权利要求1所述的方法,其中所述多个核酸条形码分子可释放地偶联至该珠子。
7. 根据权利要求1所述的方法,其中所述珠子包括凝胶珠子。
8. 根据权利要求1所述的方法,其中所述分区包含小滴。
9. 根据权利要求6所述的方法,其中所述(b)包括使包含珠子的第一水性流体与包含样品材料的组分的第二水性流体在不混溶流体内的小滴中组合。
10. 根据权利要求1所述的方法,其中(c)包括降解所述珠子。
11. 根据权利要求6所述的方法,其中(c)包括降解所述珠子,以将所述核酸条形码分子释放至所述分区。
12. 根据权利要求6所述的方法,其中(c)包括切割所述核酸条形码分子与所述珠子之间的化学联接。
13. 根据权利要求6所述的方法,其中所述珠子的交联和所述珠子与所述多个核酸条形码分子之间的联接中的至少一种包括二硫键,并且其中(c)包括使所述珠子暴露于还原剂。
14. 根据权利要求13所述的方法,其中所述还原剂是二硫苏糖醇(DTT)或三(2-羧基乙基)膦(TCEP)。
15. 根据权利要求1所述的方法,其中所述样品材料的组分包含一个或多个模板核酸分子,并且其中(d)包括将所述核酸条形码分子附接至所述一个或多个模板核酸分子的一个或多个片段。
16. 根据权利要求1所述的方法,其中所述样品材料的组分包含一个或多个模板核酸分子,所述多个核酸条形码分子还包含与所述一个或多个模板核酸分子的至少一部分互补的引物序列,并且(d)包括延伸所述核酸条形码分子以复制所述一个或多个模板核酸分子的至少一部分。
17. 根据权利要求1所述的方法,其中所述样品材料的组分包含单细胞的内含物。
18. 根据权利要求17所述的方法,其中所述单细胞包括癌细胞。

19. 根据权利要求17所述的方法,其中所述单细胞包括细菌细胞。
20. 根据权利要求19所述的方法,其中所述细菌细胞是从人的微生物组样品中分离的。
21. 根据权利要求1所述的方法,其中所述分区包含单个珠子。
22. 根据权利要求1所述的方法,其中:
 - (a) 包括提供包含多个不同核酸条形码分子的多个珠子,其中所述多个珠子中的每一个均包含多个核酸条形码分子,该核酸条形码分子包含与之关联的共同条形码序列;
 - (b) 包括将所述多个珠子与样品材料的组分共同分配至多个分区;
 - (c) 包括将所述多个不同核酸条形码分子中的不同核酸条形码分子从所述多个珠子释放至所述多个分区;以及
 - (d) 包括将所述多个不同核酸条形码分子附接至所述多个分区内的所述样品材料的组分或其片段。
23. 根据权利要求22所述的方法,其中所述多个不同的核酸条形码分子包括至少1,000个不同的核酸条形码分子。
24. 根据权利要求22所述的方法,其中所述多个不同的核酸条形码分子包括至少10,000个不同的核酸条形码分子。
25. 根据权利要求22所述的方法,其中所述多个不同的核酸条形码分子包括至少100,000个不同的核酸条形码分子。
26. 根据权利要求22所述的方法,其中所述多个不同的核酸条形码分子包括至少500,000个不同的核酸条形码分子。
27. 根据权利要求22所述的方法,其中所述多个分区包含多个子集,且所述多个子集的至少一个包含多个核酸条形码分子,所述核酸条形码分子包含共同条形码序列。
28. 根据权利要求27所述的方法,其中至少1%的所述多个分区包含多个核酸条形码分子,所述核酸条形码分子包含共同条形码序列。
29. 根据权利要求27所述的方法,其中至少2%的所述多个分区包含多个核酸条形码分子,所述核酸条形码分子包含共同条形码序列。
30. 根据权利要求27所述的方法,其中至少5%的所述多个分区包含多个核酸条形码分子,所述核酸条形码分子包含共同条形码序列。
31. 根据权利要求27所述的方法,其中所述样品材料的组分的片段包括一个或多个模板核酸分子的一个或多个片段,并且所述方法还包括对所述一个或多个模板核酸分子的一个或多个片段进行测序和至少部分地基于附接至其上的核酸条形码分子来表征所述模板核酸序列的一个或多个片段。
32. 根据权利要求31所述的方法,其中表征一个或多个模板核酸分子的一个或多个片段包括使所述一个或多个模板核酸分子中的单独模板核酸分子的片段定位到所述一个或多个模板核酸分子的单独模板核酸分子或所述单独模板核酸分子所来源的基因组。
33. 根据权利要求31所述的方法,其中表征所述模板核酸分子的一个或多个片段包括至少鉴定所述不同的核酸条形码分子中的单独核酸条形码分子,以及鉴定附接至该单独核酸条形码分子的一个或多个模板核酸分子的一个或多个片段中的单独片段的分子。
34. 根据权利要求22所述的方法,其中至少50%的所述多个分区恰好包含一个珠子。
35. 根据权利要求22所述的方法,其中至少70%的所述多个分区恰好包含一个珠子。

36. 根据权利要求22所述的方法,其中至少90%的所述多个分区恰好包含一个珠子。
37. 一种对样品材料进行条形码化的方法,其包括:
- (a) 提供包含多个不同核酸条形码分子的多个珠子,其中所述多个珠子中的每一个均包含多个核酸条形码分子,该核酸条形码分子具有与之关联的共同条形码序列;
 - (b) 将所述多个珠子与样品材料的组分共同分配至多个分区;
 - (c) 将所述多个不同核酸条形码分子中的不同核酸条形码分子从所述多个珠子释放至所述多个分区;以及
 - (d) 使所述不同核酸条形码分子附接至所述多个分区内的样品材料的组分。
38. 一种对样品材料进行条形码化的方法,其包括:
- (a) 提供可激活的核酸条形码序列,其中所述可激活的核酸条形码序列可释放地附接于珠子并包含条形码序列;
 - (b) 将所述可激活的核酸条形码序列与样品材料的一种或多种组分一起分配至小滴;
 - (c) 激活所述可激活的核酸条形码序列,以在所述小滴中产生活性核酸条形码序列;以及
 - (d) 将所述活性核酸条形码序列附接至所述样品材料的一种或多种组分。
39. 根据权利要求38所述的方法,其中(c)包括将该可激活的核酸条形码序列从所述小滴内的珠子中释放。
40. 根据权利要求38所述的方法,其中(c)包括从所述可激活的核酸条形码序列中去除可去除的保护基团。
41. 一种包含分区和珠子的组合物,其中
所述分区包含一种或多种样品材料的组分;以及
所述珠子包含在所述分区,并具有可释放地与之关联的多个核酸条形码分子,所述
多个核酸条形码分子包含共同的条形码序列。
42. 根据权利要求41所述的组合物,其中所述分区包含乳液中的水性小滴。
43. 根据权利要求41所述的组合物,其中所述多个核酸条形码分子与所述珠子通过可
切割的联接可释放地关联。
44. 根据权利要求43所述的组合物,其中所述可切割的联接包括选自下组的联接:化学
可切割的联接、光可切割的联接和热可切割的联接。
45. 根据权利要求43所述的组合物,其中所述可切割的联接包括二硫键。
46. 根据权利要求41所述的组合物,其中所述珠子包含可降解的珠子。
47. 根据权利要求46所述的组合物,其中所述可降解的珠子选自下组:光可降解的珠
子、化学可降解的珠子和热可降解的珠子。
48. 根据权利要求46所述的组合物,其中所述可降解的珠子包含化学可切割的交联。
49. 根据权利要求48所述的组合物,其中所述化学可切割的交联包括二硫键交联。
50. 根据权利要求41所述的组合物,其中所述一种或多种样品材料的组分包括核酸或
其片段。
51. 根据权利要求50所述的组合物,其中所述核酸包括基因组脱氧核糖核酸。
52. 根据权利要求50所述的组合物,其中所述核酸包括具有1kb至100kb长度的核酸片
段。

53. 根据权利要求50所述的组合物，其中所述核酸包括具有5kb至50kb长度的核酸片段。

54. 根据权利要求50所述的组合物，其中所述核酸包括具有10kb至30kb长度的核酸片段。

55. 根据权利要求41所述的组合物，其包含：

多个分区；和

多个不同的珠子，所述多个不同的珠子中的每一个均布置在所述多个分区中的单独一个内并且包含可释放地与之关联的多个核酸条形码分子，与所述多个不同珠子中的每一个关联的所述多个核酸条形码分子包含共同条形码序列，并且与不同的珠子关联的多个核酸条形码分子包含不同的条形码序列。

56. 根据权利要求55所述的组合物，其中所述多个不同的珠子包括至少1,000个不同的珠子。

57. 根据权利要求55所述的组合物，其中所述多个不同的珠子包括至少10,000个不同的珠子。

58. 根据权利要求55所述的组合物，其中所述多个不同的珠子包括至少100,000个不同的珠子。

59. 根据权利要求55所述的组合物，其中所述多个不同的珠子包括至少500,000个不同的珠子。

60. 一种对样品材料进行条形码化的方法，其包括：

使包含核酸的样品与条形码化珠子的文库组合，以形成混合物，其中每个所述条形码化珠子包含多个核酸条形码分子，所述多个核酸条形码分子包含共同条形码序列；

将该混合物分配至多个分区，其中所述分区的至少一个子集包含至多一个条形码化珠子；以及

在所述分区，将包含共同条形码序列的所述多个核酸条形码分子中的核酸条形码分子从所述条形码化珠子释放。

61. 根据权利要求60所述的方法，其中所述条形码化珠子能够被还原剂溶解。

62. 根据权利要求61所述的方法，其中所述条形码化珠子包含二硫键。

63. 根据权利要求60所述的方法，其中所述包含核酸的样品包含基因组脱氧核糖核酸。

64. 根据权利要求63所述的方法，其中所述基因组脱氧核糖核酸在与所述条形码化珠子组合之前被片段化。

65. 根据权利要求60所述的方法，其还包括使用微流体装置使所述包含核酸的样品与所述条形码化珠子的文库组合。

66. 根据权利要求60所述的方法，其还包括使用微流体装置将所述混合物分配至所述多个分区。

67. 根据权利要求60所述的方法，其中所述分区是油包水乳液内的水性小滴。

68. 根据权利要求67所述的方法，其还包括使用微流体装置将所述混合物分配至所述多个水性小滴。

69. 根据权利要求68所述的方法，其中所述微流体装置是小滴发生器。

70. 根据权利要求67所述的方法，其中所述微流体装置包括第一输入通道和第二输入

通道,它们在与输出通道流体连接的接合处汇合。

71. 根据权利要求70所述的方法,其中所述方法还包括将所述包含核酸的样品引入所述第一输入通道,以及将所述条形码化珠子的文库引入所述第二输入通道,从而在所述输出通道中生成所述包含核酸的样品和所述条形码化珠子的文库的混合物。

72. 根据权利要求71所述的方法,其中所述输出通道与第三输入通道在接合处流体连接。

73. 根据权利要求72所述的方法,其还包括将油引入所述第三输入通道,使得形成油包水乳液内的水性小滴,单独的小滴平均包含至多十个条形码化珠子。

74. 根据权利要求73所述的方法,其还包括将油引入第四输入通道,其中所述第四输入通道与所述第三输入通道和所述输出通道在接合处相交。

75. 根据权利要求70所述的方法,其还包括将还原剂引入所述第一或第二输入通道。

76. 根据权利要求70所述的方法,其中所述微流体装置还包括第三输入通道,该第三输入通道与所述第一输入通道、所述第二输入通道或所述第一通道与所述第二通道的接合处相交,并且该方法还包括将还原剂引入所述第三输入通道。

77. 根据权利要求70所述的方法,其中所述第一输入通道和所述第二输入通道彼此之间形成基本上垂直的角度。

78. 根据权利要求60所述的方法,其中所述核酸条形码分子通过还原剂的作用从所述条形码化珠子释放。

79. 根据权利要求60所述的方法,其中所述条形码化珠子包含与二硫键交联的基质,并且所述核酸条形码分子通过溶解该条形码化珠子的还原剂的作用从该条形码化珠子释放。

80. 根据权利要求60所述的方法,其还包括通过加热所述分区将所述核酸条形码分子从该条形码化珠子释放。

81. 根据权利要求60所述的方法,其中所述包含共同条形码序列的多个核酸条形码分子用已知的序列预合成。

82. 根据权利要求60所述的方法,其中所述包含共同条形码序列的多个核酸条形码分子还包括多个随机Nmer。

83. 根据权利要求82所述的方法,其还包括使所述多个随机Nmer与所述包含核酸的样品杂交。

84. 根据权利要求82所述的方法,其还包括在所述分区内进行核酸扩增反应。

用于样品处理的组合物和方法

[0001] 交叉引用

[0002] 本申请要求2013年6月27日提交的美国临时专利申请号61/840,403、2013年7月10日提交的美国临时专利申请号61/844,804、2013年8月13日提交的美国专利申请号13/966,150、2013年8月13日提交的PCT国际专利申请号PCT/US13/54797、2013年10月26日提交的美国临时专利申请号61/896,060、2013年11月27日提交的美国临时专利申请号61/909,974、2014年2月7日提交的美国临时专利申请号61/937,344、2014年2月14日提交的美国临时专利申请号61/940,318和2014年5月9日提交的美国临时专利申请号61/991,018的权益，所述申请为了所有目的通过引用以其全文并入本文。此外，本申请与2012年8月14日提交的美国临时专利申请号61/683,192相关，该申请为了所有目的通过引用以其全文并入本文。

背景技术

[0003] 基因组测序可以用来获得各种各样的生物医学背景的信息，包括诊断学、预后、生物技术和法医生物学。测序可包括基本方法，其包括Maxam-Gilbert测序和链终止法或从头测序法（包括鸟枪法测序和桥式PCR），或新一代方法，其包括聚合酶克隆测序、454焦磷酸测序、Illumina测序、SOLiD测序、Ion Torrent半导体测序、HeliScope单分子测序、**SMRT®**测序，等等。对于大多数测序应用，样品如核酸样品在引入测序仪之前先进行处理。例如，可通过扩增或通过附接独特标识符处理样品。独特标识符通常用来标识特定样品的来源。

发明内容

[0004] 本公开内容概括地提供了用于生成具有共价附接的多核苷酸的珠子的方法、组合物、装置和试剂盒。这样的珠子可用于任何合适的应用。

[0005] 本公开内容的一方面提供了对样品材料进行条形码化的方法。可提供第一分区，其包含与之关联的多个核酸条形码分子，并且所述核酸条形码分子可包含相同的核酸条形码序列。可将第一分区与样品材料的组分共同分配至第二分区，随后可将条形码分子从第一分区释放至第二分区。释放的条形码分子可附接至第二分区内的样品材料的一种或多种组分或其片段。在一些情况下，第一分区可包含与之关联的具有相同条形码序列的至少1,000个条形码分子、至少10,000个条形码分子、至少100,000个条形码分子或至少1,000,000个条形码分子。而且，在一些实例中，第一分区可以是珠子、微胶囊或小滴。在一些情况下，第一分区可包含珠子（例如，凝胶珠子）并且条形码分子可与珠子可释放地偶联。而且，第二分区可包含小滴和/或可包含不多于一个第一分区。

[0006] 在一些情况下，将第一分区与样品材料的组分共同分配至第二分区可包括使包含珠子的第一水性流体与包含样品组分的第二水性流体在不混溶流体内的小滴中组合。而且，可通过降解第一分区将条形码分子从第一分区中释放。在第一分区是珠子的情况下，可通过降解珠子和/或切割条形码分子与珠子之间的化学联接在第二分区中释放条形码分子。在一些情况下，珠子的交联和珠子与条形码分子之间的联接中的至少一个可包括二硫键。在一些情况下，可通过使珠子暴露于还原剂（例如，二硫苏糖醇（DTT）和三（2-羧基乙基）

膦(TCEP)将条形码分子从珠子释放。

[0007] 样品材料可包含一个或多个模板核酸分子，并且条形码分子可附接至该模板核酸分子的一个或多个片段。在一些情况下，条形码分子可包含与模板核酸分子的至少一部分互补的引物序列，并且条形码分子可通过延伸条形码分子以复制模板核酸分子的至少一部分而附接至模板核酸分子或其片段。而且，样品材料可包含单细胞(例如，癌细胞或细菌细胞(例如，从人的微生物组样品中分离的细菌细胞))的内含物。

[0008] 此外，可提供包含多个不同的核酸条形码序列的多个第一分区。每一个第一分区均可包含多个至少1000个核酸条形码分子，所述核酸条形码分子具有与之关联的相同的核酸条形码序列。可将第一分区与样品材料的组分共同分配至多个第二分区。随后可将来自第一分区的核酸条形码分子释放至第二分区。释放的核酸条形码分子随后可附接至第二分区内的样品材料的组分或其片段。在一些情况下，所述多个不同的核酸条形码序列可包含至少约1,000个不同的条形码序列、至少约10,000个不同的条形码序列、至少约100,000个不同的条形码序列或至少约500,000个不同的条形码序列。此外，在一些实例中，第二分区的子集可包含相同的核酸条形码序列。例如，至少约1%、至少约2%或至少约5%的第二分区可包含相同的核酸条形码序列。另外，在一些情况下，至少50%的第二分区、至少70%的第二分区或至少90%的第二分区可含有不多于一个第一分区。在一些情况下，至少50%的第二分区、至少70%的第二分区或至少90%的第二分区可含有恰好一个第一分区。

[0009] 样品材料的组分的片段可包括一个或多个模板核酸序列的一个或多个片段。可对模板核酸序列的片段进行测序，并至少部分地基于附接至其上的核酸条形码序列进行表征。在一些情况下，可通过将模板核酸序列中的单独模板核酸序列的片段定位(map)到模板核酸序列的单独模板核酸序列或该单独模板核酸序列所来源的基因组，来表征模板核酸序列的片段。在一些情况下，可通过至少鉴定不同的核酸条形码序列中的单独核酸条形码序列，以及鉴定附接至该单独核酸条形码序列的模板核酸序列的片段中的单独片段的序列，来表征模板核酸序列的片段。

[0010] 本公开内容的另一方面提供了对样品材料进行条形码化的方法。可提供多个第一分区，其包含多个不同的核酸条形码序列。每一个第一分区均可包含多个核酸条形码分子，所述核酸条形码分子具有与之关联的相同的核酸条形码序列。可将第一分区与样品材料的组分共同分配至多个第二分区。可将条形码分子从第一分区释放至第二分区。释放的条形码分子随后可附接至第二分区内的样品材料的组分。

[0011] 本公开内容的另一方面提供了对样品材料进行条形码化的方法。可提供可激活的核酸条形码序列，并将其与样品材料的一种或多种组分分配至第一分区。可激活所述可激活的核酸条形码序列，以在第一分区中产生活性核酸条形码序列。活性核酸条形码序列可附接至样品材料的一种或多种组分。在一些情况下，可通过将可激活的核酸条形码序列从第一分区内的第二分区中释放来激活所述可激活的核酸条形码序列。在一些情况下，可通过从可激活的核酸条形码序列中去除可去除的保护基团来激活所述可激活的核酸条形码序列。

[0012] 本公开内容的另一方面提供了一种组合物，其包含含有一种或多种样品组分的第一分区和包含在第一分区内的第二分区。第二分区可具有可释放地与之关联的多个寡核苷酸，并且所述寡核苷酸可包含共同的条形码序列。在一些情况下，第一分区可包含乳液中的

水性小滴,和/或第二分区可包含微胶囊或珠子。在一些情况下,第二分区可包含可降解的珠子,其可以是光可降解的珠子、化学可降解的珠子和/或热可降解的珠子。可降解的珠子可包含化学可切割的交联,例如二硫键交联。而且,在一些情况下,所述寡核苷酸可通过可切割的联接与第二分区可释放地关联。可切割的联接可包括,例如,化学可切割的联接、光可切割的联接和/或热可切割的联接。在一些情况下,可切割的联接是二硫键。此外,样品组分可包括,例如,核酸(例如,基因组核酸,如基因组DNA)或其片段。该核酸可包括核酸片段,该核酸片段可以具有约1kb至约100kb的长度、约5kb至约50kb的长度或约10kb至约30kb的长度。

[0013] 在一些情况下,所述组合物包含多个第一分区和多个不同的第二分区。每一个不同的第二分区均可被布置在单独的第一分区,并可包含可释放地与之关联的多个寡核苷酸。与每一个第二分区关联的寡核苷酸可包含共同的条形码序列,并且与不同的第二分区关联的寡核苷酸可包含不同的条形码序列。在一些情况下,不同的第二分区可包含至少1,000个不同的第二分区、至少10,000个不同的第二分区、至少100,000个不同的第二分区或至少500,000个不同的第二分区。

[0014] 本公开内容的另一方面提供了一种方法,其包括使核酸的样品与条形码化珠子的文库组合以形成混合物。可将该混合物分配至多个分区,使得分区的至少一个子集包含至多一个条形码化珠子。在所述分区,可将条形码从条形码化珠子释放。在一些情况下,条形码可用已知的序列预合成和/或可包含多个随机N-mer。随机N-mer可与核酸的样品杂交,以便在分区内进行例如核酸扩增反应。在一些情况下,条形码化珠子可能能够被还原剂溶解并可包含二硫键。而且,在一些情况下,样品核酸可以是基因组DNA,其在与条形码化珠子组合之前可以片段化或可以不片段化。在一些情况下,可通过还原剂的作用将条形码从条形码化珠子释放。在一些情况下,条形码化珠子可包含与二硫键交联的基质,并且可通过溶解条形码化珠子的还原剂的作用将条形码从该条形码化珠子释放。在一些情况下,可通过加热分区将条形码从条形码化珠子释放。

[0015] 在一些情况下,核酸的样品可与条形码化珠子的文库组合,和/或可使用微流体装置将两者的混合物分配至多个分区。在一些实例中,分区可以是油包水乳液内的水性小滴。将混合物分配至油包水乳液内的水性小滴可使用微流体装置来完成。

[0016] 微流体装置可以是小滴发生器,并且在一些情况下,可包括第一输入通道和第二输入通道,它们在与输出通道流体连接的接合处汇合。可将核酸的样品引入第一输入通道,而可将条形码化珠子的文库引入第二输入通道,以在输出通道中生成样品核酸和条形码化珠子的文库的混合物。在一些情况下,也可以将还原剂引入第一输入通道和第二输入通道中的任一个或两个中。而且,第一输入通道和第二输入通道彼此之间可形成基本上垂直的角度。

[0017] 在一些情况下,输出通道可与第三输入通道在接合处流体连接。可将油引入第三输入通道,使得形成包含条形码化珠子的油包水乳液内的水性小滴。例如,该小滴可包含平均至多十个条形码化珠子、至多七个条形码化珠子、至多五个条形码化珠子、至多三个条形码化珠子、至多两个条形码化珠子或至多一个条形码化珠子。而且,微流体装置可包括与第三输入通道和输出通道在接合处相交的第四输入通道。在一些情况下,也可向第四输入通道提供油。在一些情况下,微流体装置可包括与第一输入通道、第二输入通道或第一输入通

道与第二输入通道的接合处相交的另一输入通道。在一些情况下，可将还原剂引入所述另一输入通道。

[0018] 本公开内容的另一方面提供了包含珠子的组合物，该珠子与包含相同条形码序列和可变结构域的多个寡核苷酸共价连接。在一些情况下，该寡核苷酸还可包含引物结合位点和/或通用引物。此外，该相同条形码序列可以是约6个核苷酸至约20个核苷酸的长度。而且，该寡核苷酸可与珠子通过二硫键共价连接，和/或该珠子可包含胱胺或修饰的胱胺。在一些情况下，该珠子可能能够被还原剂基本上溶解。此外，在一些情况下，该珠子可包含至少约1,000,000个包含相同条形码序列的寡核苷酸。在一些情况下，至少约30%的寡核苷酸可包含具有不同序列的可变结构域。在一些情况下，该可变结构域可以是随机N-mer。在一些情况下，珠子可与寡核苷酸通过可切割的联接共价连接，例如化学可切割的联接、光可切割的联接和热可切割的联接。

[0019] 本公开内容的另一方面提供了包含珠子的组合物，该珠子可包含多个多于1,000,000个寡核苷酸，其中每一个寡核苷酸均包含恒定区和可变区。该珠子可能能够被还原剂基本上溶解。在一些情况下，每一个寡核苷酸可包含相同的恒定区。在一些情况下，至少25%的寡核苷酸可具有相同的恒定区。在一些情况下，该恒定区可以是条形码序列。在一些情况下，至少约25%的寡核苷酸可具有包含不同序列的可变区。本公开内容的另一方面提供了包含至少约1,000,000个珠子的文库，其中每一个珠子包含多个多于1,000,000个包含恒定区和可变区的寡核苷酸。在一些情况下，至少约25%的珠子包含具有不同核苷酸序列的寡核苷酸。

[0020] 本公开内容的另一方面提供了包含多个珠子的组合物，其中每一个珠子均包含多个可释放地偶联至其上的寡核苷酸。与单独珠子关联的寡核苷酸可包含共同的条形码结构域和可变结构域。共同的条形码结构域可在两个或更多个珠子之间不同。在一些情况下，所述珠子可包含至少约10,000个与不同珠子偶联的不同的条形码结构域。在一些情况下，每一个珠子均可包含至少约1,000,000个可释放地偶联至其上的寡核苷酸。

[0021] 本公开内容的另一方面提供了生成功能化珠子的方法。多个聚合物或单体可与一个或多个寡核苷酸混合。可交联所述聚合物或单体，使得在聚合物或单体之间形成二硫键，从而形成硬化的珠子。而且，可导致在寡核苷酸与聚合物或单体之间形成共价联接。在一些情况下，所述聚合物或单体可包含丙烯酰胺。在一些情况下，可交联所述聚合物或单体以形成硬化的珠子，并可导致在寡核苷酸与聚合物或单体之间同时或顺序地形成共价联接。而且，在一些情况下，所述寡核苷酸可包含可连接至acrydite部分的引物（例如，通用引物、测序引物）。

[0022] 此外，一个或多个附加寡核苷酸可附接至所述寡核苷酸。所述附加寡核苷酸可以是条形码序列，并因此当附接至寡核苷酸时可形成条形码化珠子。在一些情况下，该条形码序列可以是约6个核苷酸至约20个核苷酸的长度。

[0023] 在一些情况下，功能化的珠子可与多个第一附加寡核苷酸组合以产生混合物。可将该混合物分配至多个分区，使得平均每个分区包含不多于一个第一附加寡核苷酸。在一些情况下，所述分区可以是油包水乳液内的水性小滴和/或可由微流体装置生成。在一些情况下，所述分区可通过批量乳化过程生成。而且，第一附加寡核苷酸可在所述分区内扩增，以产生包含扩增的第一寡核苷酸的珠子。在一些情况下，在扩增过程中可使用捕获引物，并

且该捕获引物可附接至捕获部分,例如,生物素、链霉亲和素或谷胱甘肽-S-转移酶(GST)。扩增之后,可将分区的内含物合并到共同的容器中。可从分区的内含物中分离出包含扩增的第一寡核苷酸的珠子。在一些情况下,可使探针与扩增的第一寡核苷酸杂交。该探针可包含捕获部分。

[0024] 此外,一个或多个第二附加寡核苷酸可附接至所述扩增的第一寡核苷酸。在一些情况下,第二附加寡核苷酸可包含随机N-mer序列和/或伪随机N-mer序列。在一些情况下,第二附加寡核苷酸可包含引物结合位点,该引物结合位点可包含通用序列部分。在一些情况下,该引物结合位点可包含含有尿嘧啶的核苷酸。而且,该通用序列部分可与测序装置相容和/或可包含含有尿嘧啶的核苷酸的子段。

[0025] 本公开内容的另一方面提供了制备条形码文库的方法。可提供多个单独的第一珠子群体,并且包含第一条形码序列段的第一寡核苷酸可附接至所述单独的第一珠子群体,使得每一个单独的第一珠子群体均包含附接至其上的不同的第一条形码序列段。然后可合并所述单独的珠子群体以提供第一合并的珠子群体。然后可将所述第一合并的珠子群体分离成多个第二珠子群体。包含第二条形码序列段的第二寡核苷酸可附接至已经附接至所述第二珠子群体的第一寡核苷酸,使得每一个单独的第二珠子群体包含不同的第二条形码序列段。然后可合并所述单独的第二珠子群体以提供包含条形码文库的第二合并的珠子群体。

[0026] 在一些情况下,第一条形码序列段和第二条形码序列段可独立地选自第一组条形码序列段。另外,第一条形码序列段和第二条形码序列段可独立地包含至少4个核苷酸的长度、至少6个核苷酸的长度或至少10个核苷酸的长度。在一些情况下,第一条形码序列段和第二条形码序列段可独立地包含约4个核苷酸的长度至约20个核苷酸的长度。而且,在一些情况下,第一珠子群体可包含至少100个不同的第一条形码序列段或至少1,000个不同的第一条形码序列段。此外,在一些情况下,至少1,000,000个第一寡核苷酸分子可附接至每一个单独的第一珠子群体中的每一个珠子。在一些情况下,第二珠子群体可包含至少100个不同的第二条形码序列段或至少1,000个不同的第二条形码序列段。在一些情况下,至少1,000,000个第二寡核苷酸分子可附接至每一个第二珠子群体中的每一个珠子。

[0027] 另外,在一些情况下,第一寡核苷酸和第二寡核苷酸中的至少一个可包含功能序列,例如引物序列、引物退火序列、附接序列和测序引物序列。在一些情况下,第一寡核苷酸和第二寡核苷酸中的至少一个可包含序列段,该序列段包含含有尿嘧啶的核苷酸和非天然核苷酸中的一种或多种。

[0028] 在一些情况下,可通过提供夹板序列将第一寡核苷酸附接至单独的第一珠子群体,该夹板序列部分地互补于第一寡核苷酸的至少一部分,并且部分地互补于附接至单独的第一珠子群体的寡核苷酸的至少一部分。在一些情况下,第一寡核苷酸可附接至单独的第一珠子群体,使得其可释放地附接至单独的第一珠子群体。例如,第一寡核苷酸可通过可切割的联接附接至单独的第一珠子群体。在一些情况下,第一寡核苷酸可直接或间接地附接至单独的第一珠子群体。

[0029] 此外,在一些情况下,第二寡核苷酸可通过连接而附接至第一寡核苷酸。在一些情况下,可通过提供夹板序列将第二寡核苷酸附接至第一寡核苷酸,该夹板序列部分地互补于第一寡核苷酸的至少一部分,并且部分地互补于第二寡核苷酸的至少一部分。在一些情

况下,夹板序列当与第一寡核苷酸杂交时可提供第一突出端序列,并且第二条形码序列段可包含与第一突出端序列互补的第二突出端序列。在一些情况下,第一突出端序列和第二突出端序列可为约2个核苷酸的长度至约6个核苷酸的长度。此外,在一些情况下,第一突出端序列可包含多个不同的突出端序列,而第二寡核苷酸可包含与多个不同的第一突出端序列互补的多个不同的第二突出端序列。

[0030] 而且,单独的第一珠子群体可包含可降解的珠子,例如,化学可降解的珠子、光可降解的珠子和/或热可降解的珠子。在一些情况下,单独的第一珠子群体可包含珠子,所述珠子可包含化学可还原的交联剂,例如包含二硫键的化学可还原的交联剂。

[0031] 在一些情况下,第三寡核苷酸可附接至已经附接至第一寡核苷酸的第二寡核苷酸。第三寡核苷酸可包含功能序列,该功能序列可以是引物序列(例如,通用引物序列、靶向引物序列或随机序列)和/或可以是随机N-mer序列。在第三寡核苷酸包含随机N-mer序列的情况下,该随机N-mer序列可为约5个核苷酸的长度至约25个核苷酸的长度。

[0032] 本公开内容的另一方面提供了制备条形码文库的方法。可提供包含多个不同的第一珠子群体的第一合并的珠子群体,其中每一个不同的第一珠子群体包含附接至其上的不同的第一寡核苷酸。每一个不同的第一寡核苷酸可包含不同的第一条形码序列段。可以将所述第一合并的珠子群体分离成多个第二珠子群体。包含第二条形码序列段的第二寡核苷酸可附接至已经附接至第二珠子群体的第一寡核苷酸,其中每一个第二珠子群体包含不同的第二条形码序列段。可合并第二珠子群体以提供包含条形码文库的第二合并的珠子群体。

[0033] 在一些情况下,第一寡核苷酸可以可释放地附接至第一合并的珠子群体中的珠子。在一些情况下,第一寡核苷酸可通过可切割的联接附接至第一合并的珠子群体中的珠子。在一些情况下,第一合并的珠子群体中的珠子可各自包含至少1,000,000个附接至其上的第一寡核苷酸。在一些情况下,第一合并的珠子群体可包含至少10个不同的第一珠子群体、至少100个不同的第一珠子群体或至少500个不同的第一珠子群体。

[0034] 本公开内容的另一方面提供了包含多个不同的寡核苷酸的条形码文库。每一个不同的寡核苷酸可包含选自第一组条形码序列段的第一条形码序列段;选自第二组条形码序列段的第二条形码序列段;和连接第一条形码序列段与第二条形码序列段的连接序列。该连接序列可以是约2个核苷酸的长度至约6个核苷酸的长度,并且可选自一组连接序列。在一些情况下,这组连接序列包括约2个不同的连接序列至约50个不同的连接序列。在一些情况下,第一组条形码序列段和第二组条形码序列段是相同的。

[0035] 本公开内容的另一方面提供了扩增模板核酸序列的方法。可将模板核酸序列和包含多个可释放地附接的寡核苷酸的珠子共同分配至分区。该寡核苷酸可包含与模板核酸序列的一个或多个区域互补的引物序列,并可包含共同序列。该引物序列可与模板核酸序列退火,并且可延伸该引物序列以产生模板核酸序列的至少一部分的一个或多个第一拷贝,其中所述一个或多个第一拷贝包含该引物序列和该共同序列。

[0036] 在一些情况下,引物序列可包括可变引物序列(例如,随机N-mer)和/或可包括靶向引物序列。在一些情况下,所述分区可包含乳液中的小滴。在引物序列与模板核酸序列退火之前,可将寡核苷酸从珠子上释放至分区。在一些实例中,分区中可提供聚合酶(例如,外切核酸酶缺陷的聚合酶)。而且,引物序列的延伸可包括使用链置换聚合酶(例如,基本上不

具有外切核酸酶活性的热稳定的链置换聚合酶) 延伸引物序列。此外, 寡核苷酸可以是外切核酸酶抗性的。例如, 寡核苷酸可包含一个或多个硫代磷酸酯联接。在一些情况下, 硫代磷酸酯联接可包括在寡核苷酸中的末端核苷酸间联接处的硫代磷酸酯联接。

[0037] 此外, 一个或多个可变引物序列可与第一拷贝退火并延伸, 以由第一拷贝产生一个或多个第二拷贝, 使得第二拷贝包含所述一个或多个可变引物序列和共同序列。在一些情况下, 第二拷贝可包含与第一拷贝中的单独第一拷贝的至少一部分互补的序列和与所述一个或多个可变引物序列中的单独可变序列互补的序列。在一些情况下, 第二拷贝可在退火条件下优先形成发夹分子。而且, 在一些情况下, 所述寡核苷酸可包含在可变引物序列的延伸过程中未拷贝的序列段。未拷贝的序列段可包含, 例如, 一个或多个含尿嘧啶的核苷酸。此外, 可重复所述方法的任何步骤以产生扩增的核酸。

[0038] 本公开内容的另一方面提供了扩增多个不同核酸的方法。可将不同的核酸分配至单独的第一分区, 其中每一个第一分区包含具有可释放地与之关联的多个寡核苷酸的第二分区。与给定的第二分区关联的多个寡核苷酸可包含可变引物序列和条形码序列, 与不同的第二分区关联的寡核苷酸包含不同的条形码序列。与所述多个第二分区关联的寡核苷酸可释放至第一分区。第一分区中的可变引物序列可释放至第一分区内的核酸并延伸, 以在第一分区产生核酸的至少一部分的一个或多个拷贝, 使得所述拷贝包含释放至第一分区的寡核苷酸和关联的条形码序列。在一些情况下, 第一分区可包含乳液中的小滴而第二分区可包含珠子。在一些情况下, 每一个珠子可包含多于100,000个与之关联的寡核苷酸或多于1,000,000个与之关联的寡核苷酸。在一些情况下, 第二分区可包含至少1,000个不同的条形码序列、至少10,000个不同的条形码序列或至少100,000个不同的条形码序列。

[0039] 本公开内容的另一方面提供了全基因组扩增方法。随机引物可与基因组核酸杂交。随机引物可附接至通用核酸序列和核酸条形码序列, 其中该通用核酸序列可包含一个或多个含尿嘧啶的核苷酸。随机引物可延伸以形成扩增的产物, 并且该扩增的产物可暴露于适合使扩增产物经历形成部分发夹分子的分子内杂交反应的条件。在一些情况下, 随机引物可以是随机N-mer序列。在一些情况下, 通用核酸序列可包含至少10个不含尿嘧啶的核苷酸的区段。而且, 所述方法可在寡核苷酸阻断剂的存在下进行。该寡核苷酸阻断剂可能能够与通用核酸序列的至少一部分杂交和/或可包含C3间隔区(/3SpC3/)、双脱氧-C(/3ddC/)或3' 磷酸。

[0040] 本公开内容的另一方面提供了扩增核酸的方法。基因组组分可被片段化成多个第一片段。可将第一片段与多个寡核苷酸共同分配至多个分区。每一个分区中的寡核苷酸可包含引物序列和共同序列。每一个分区中的引物序列可与每一个分区内的第一片段的多个不同区域退火, 并且引物序列沿着第一片段延伸以在每一个分区产生扩增的第一片段。在一些情况下, 分区内扩增的第一片段可包含基因组组分的至少1X的覆盖范围、基因组组分的至少2X的覆盖范围或基因组组分的至少10X的覆盖范围。在一些情况下, 基因组组分可包括染色体。在一些情况下, 基因组组分可包括生物体的整个基因组。

[0041] 本公开内容的另一方面提供了表征核酸段的方法。可将核酸段与包含多个寡核苷酸的珠子共同分配至分区, 所述多个寡核苷酸包含共同的核酸条形码序列。所述寡核苷酸可附接至该核酸段的片段或该核酸段的部分的拷贝, 使得所述共同的核酸条形码序列附接至该核酸段的片段或该核酸段的部分的拷贝。可对核酸段的片段或核酸段的部分的拷贝和

附接的共同的核酸条形码序列进行测序，并且可至少部分基于它们与共同的核酸条形码序列的附接，将核酸段的片段或核酸段的拷贝表征为在该核酸段内连接。例如，可将核酸段和珠子共同分配至乳液中的小滴，或可共同分配至微胶囊。在一些情况下，核酸段的片段可包括核酸段的重叠片段。在一些情况下，核酸段的片段可包含核酸段的大于2X的覆盖范围或核酸段的大于10X的覆盖范围。

[0042] 而且，在一些情况下，所述寡核苷酸可以可释放地附接至珠子。例如，寡核苷酸可以是当向珠子施加刺激（例如，热刺激、光刺激、化学刺激等）时可从珠子上释放的。在一些情况下，刺激的施加可导致寡核苷酸与珠子之间的联接切割和/或可导致珠子的降解，使得寡核苷酸从珠子释放。而且，珠子可包含至少约10,000个附接至其上的寡核苷酸、至少约100,000个附接至其上的寡核苷酸、至少约1,000,000个附接至其上的寡核苷酸、至少约10,000,000个附接至其上的寡核苷酸或至少约100,000,000个附接至其上的寡核苷酸。此外，在一些情况下，寡核苷酸可包含一个或多个功能序列，例如，引物序列、引物退火序列或固定序列。在一些情况下，可经由合成测序方法对核酸段的片段或核酸段的部分的拷贝和附接的共同的核酸条形码序列进行测序。

[0043] 此外，在一些情况下，所述寡核苷酸可包含能够与核酸段的一部分或其互补序列退火的引物序列。可延伸该引物序列以复制核酸段的至少一部分或其互补序列，以产生核酸段的一部分或其互补序列的拷贝，其包含共同的核酸条形码序列。在一些情况下，寡核苷酸可至少包含第一测序引物序列。

[0044] 在一些情况下，可将多个核酸段与多个不同的珠子共同分配至多个单独的分区，使得所述单独分区的多个不同分区中的每一个分区含有单个珠子。每一个珠子可包含含有附接至其上的共同条形码序列的多个寡核苷酸，其中不同的珠子包含多个不同的条形码序列。每个分区中的条形码序列可附接至单独分区内的核酸段的片段或核酸段的部分的拷贝。随后可从单独的分区中合并所述片段或拷贝，并且可对该片段或拷贝和任何关联的条形码序列进行测序，以提供经测序的片段或经测序的拷贝。可部分基于包含共同条形码序列的经测序的片段或经测序的拷贝，将经测序的片段或经测序的拷贝表征为来源于共同的核酸段。在一些情况下，核酸段可包含基因组的至少一部分的片段。在这样的情况下，可从经测序的片段或经测序的拷贝组装序列，以提供基因组的至少一部分的毗连序列。从经测序的片段或经测序的拷贝组装序列可部分基于经测序的片段或经测序的拷贝的每一个核苷酸序列和包含共同条形码序列的经测序的片段或经测序的拷贝。而且，在一些情况下，可部分基于核酸段的片段或核酸段的部分的拷贝和包含共同条形码序列的经测序的片段或经测序的拷贝的每一个核苷酸序列，来表征核酸段的片段或核酸段的部分的拷贝。

[0045] 在一些情况下，不同的珠子可包含至少1,000个不同的条形码序列、至少10,000个不同的条形码序列或至少1,000,000个不同的条形码序列。在一些情况下，单独的分区中的两个或更多个分区可包含含有相同的条形码序列的珠子。在一些情况下，至少1%的单独的分区包含具有相同的条形码序列的珠子。

[0046] 本公开内容的另一方面提供了表征靶核酸的方法。可将靶核酸的第一片段分配至多个小滴，其中每个小滴包含具有多个附接至其上的寡核苷酸的珠子。附接至给定珠子的寡核苷酸可包含共同的条形码序列。共同的条形码序列可附接至第一片段的第二片段，并可合并所述小滴。可对第二片段和附接的条形码序列进行测序，并且可至少部分地基于包

含共同的条形码序列的第二片段，将第二片段定位到第一片段中的一个或多个。

[0047] 本公开内容的另一方面提供了核酸测序方法。可提供多个靶核酸序列，并且可将其分离至多个单独的分区。所述单独的分区中的每一个分区可包含一个或多个靶核酸序列和包含多个附接至其上的寡核苷酸的珠子。附接至给定珠子的寡核苷酸可包含共同的条形码序列。所述寡核苷酸可附接至分区内的一个或多个靶核酸序列的片段或一个或多个靶核酸序列的部分的拷贝，从而将所述共同的条形码序列附接至所述一个或多个靶核酸序列的片段或所述一个或多个靶核酸序列的部分的拷贝。可将单独的分区合并，并可对所述一个或多个靶核酸序列的片段或所述一个或多个靶核酸序列的部分的拷贝和附接的条形码序列进行测序，以提供条形码化片段序列或条形码化拷贝序列。在一些情况下，可部分基于条形码化片段序列或条形码化拷贝序列的条形码部分，将所述条形码化片段序列或条形码化拷贝序列组装成一个或多个毗连的核酸序列。

[0048] 本公开内容的另一方面提供了表征核酸段的方法。可将核酸段与包含多个寡核苷酸的珠子共同分配至第一小滴，所述多个寡核苷酸包含共同的核酸条形码序列。所述寡核苷酸可附接至该核酸段的片段或该核酸段的部分的拷贝，从而将所述共同的核酸条形码序列附接至该核酸段的片段或该核酸段的部分的拷贝。可对该核酸段的片段或该核酸段的部分的拷贝和附接的共同的核酸条形码序列进行测序，以提供多个条形码化片段序列或条形码化拷贝序列。可至少部分地基于所述共同的核酸条形码序列，将所述条形码化片段序列或条形码化拷贝序列组装成一个或多个毗连的核酸序列。在一些情况下，可部分基于条形码化片段序列或条形码化拷贝序列的非条形码部分的核酸序列，组装所述条形码化片段序列或条形码化拷贝序列。

[0049] 本公开内容的另一方面提供了核酸测序方法。可提供多个靶核酸序列，并且将所述靶核酸序列分离至多个单独的分区。所述单独的分区中的每一个分区可包含一个或多个靶核酸序列和多个寡核苷酸。给定分区中的寡核苷酸可包含共同的条形码序列，并且所述多个单独的分区可包含至少10,000个不同的条形码序列。每一个分区中的共同的条形码序列可附接至该分区内的所述一个或多个靶核酸序列的片段或所述一个或多个靶核酸序列的部分的拷贝。可将单独的分区合并，并可对所述一个或多个靶核酸序列的片段或所述一个或多个靶核酸序列的部分的拷贝和附接的条形码序列进行测序。在一些情况下，所述单独的分区可包含至少100,000个不同的条形码序列。

[0050] 援引并入

[0051] 在本说明书中提到的所有出版物、专利和专利申请均为所有目的通过引用以其整体并入本文，其引用程度如同具体地且单独地指出每个单独的出版物、专利或专利申请通过引用而并入。

附图说明

[0052] 图1A是制备条形码化珠子的流程图。

[0053] 图1B是处理用于测序的样品的流程图。

[0054] 图2是制备珠子的流程图。

[0055] 图3A是通过有限稀释向珠子上添加条形码的流程图。

[0056] 图3B是向附接至珠子的寡核苷酸上添加附加序列的流程图。

- [0057] 图4是将序列附接至珠子的图示(A-N)。“g/w”意指水包凝胶;“g/w/o”意指油包水包凝胶;
- [0058] 图5提供了附接至寡核苷酸的凝胶珠子的图示(5A),用于制备乳液中的凝胶珠子(GEM)的微流体芯片的图像(5B),以及GEM的图像(5C、D、E)。
- [0059] 图6提供了具有附接的寡核苷酸的珠子的亮视野图像(A、C、E)和荧光图像(B、D、F)。
- [0060] 图7提供了附接至DNA的珠子的荧光图像(A-C)。
- [0061] 图8提供了富含条形码的珠子群体的图像(A-F)。
- [0062] 图9提供了通过加热溶解珠子的图像(A-D)。
- [0063] 图10A提供了功能化珠子的示意图。图10提供了用还原剂溶解的珠子的图像(B-G)。
- [0064] 图11A提供了功能化珠子的示意图。图11提供了当使用不同条件制备珠子时条形码寡核苷酸和引物二聚体对的存在的图形描述(B-D)。
- [0065] 图12是附接至珠子的内含物的图形描述。
- [0066] 图13A是说明使用分区将条形码添加至珠子的流程图。
- [0067] 图13B是说明将附加序列添加至珠子的流程图。
- [0068] 图13C是说明在微孔板中使用组合方法制备条形码化珠子的图示。
- [0069] 图14是含有通用序列(R1、P5)的寡核苷酸和含尿嘧啶的核苷酸的图示(A-C)。
- [0070] 图15是在用于测序(PHASE)过程的部分发夹扩增中所用的步骤的图示(A-G)。
- [0071] 图16A是在引物的通用部分中包括含尿嘧啶的核苷酸的图形描述。
- [0072] 图16B是通过在反应混合物中包括acyNTP来控制扩增产物长度的图形描述。
- [0073] 图17是添加阻断寡核苷酸来降低起始位点偏倚的图形描述。
- [0074] 图18是数字处理器及其相关组件的流程图。
- [0075] 图19是为Illumina测序仪提供示例序列的表格。
- [0076] 图20是提供用来标记珠子的示例捕获部分浓度列表的表格。
- [0077] 图21是提供使用包含含有胸腺嘧啶的核苷酸的引物获得的测序度量列表的表格。
- [0078] 图22是提供使用包含含有尿嘧啶的核苷酸的引物获得的测序度量列表的表格。
- [0079] 图23是说明使用基于连接的示例组合方法来制备条形码化珠子的图示(A-D)。
- [0080] 图24是说明在基于连接的组合方法中示例性地使用间隔区碱基来制备条形码化珠子的图示(A-B)。
- [0081] 图25是说明使用基于连接的示例组合方法来制备条形码化珠子的图示(A-C)。
- [0082] 图26是说明在用来制备条形码化珠子的基于连接的示例组合方法中所用示例核酸的图示。
- [0083] 图27是说明用来制备条形码化珠子的基于连接的示例组合方法的图示(A-B)。
- [0084] 图28是适合于链特异性扩增的示例靶向条形码构建体的图示(A-B)。
- [0085] 图29是可聚合生成珠子的示例单体和交联剂的结构描述(A-C)。
- [0086] 图30是可用于生成珠子的示例方法的结构描述(A-C)。
- [0087] 图31是包含可用于将物质附接至珠子的官能团的示例珠子的图示。
- [0088] 图32提供了可在聚合反应过程中使用的示例引发剂的结构描述。

[0089] 图33A是条形码引物的图示。图33B-E是对应于实施例16中描述的示例扩增反应实验的数据的图形描述。

[0090] 图34是示例发夹构建体的图示(A-C)。

[0091] 图35是用于对珠子进行功能化的示例方法的图示(A-B)。

[0092] 图36是在实施例17中描述的凝胶电泳实验过程中获得的凝胶照片。

[0093] 图37A是在实施例18中描述的寡核苷酸的图示。图37B是在实施例18中描述的凝胶电泳实验过程中获得的凝胶照片。图37C是在实施例18中描述的荧光显微术实验过程中获得的珠子的显微照片。

[0094] 图38提供了示例性核酸条形码化和扩增过程的示意图。

[0095] 图39提供了本文所述的方法对于核酸测序和组装的示例性应用的示意图。

[0096] 图40显示了如本文所述的核酸的条形码化和扩增之后备选的处理步骤的实例。

具体实施方式

I. 概述

[0098] 本公开内容提供了在通过向样品组分的子集受控递送试剂来处理样品材料、随后部分地利用所递送的试剂来分析那些样品组分中有用的方法、系统和组合物。在许多情况下，该方法和组合物用于样品处理，尤其是通常用于核酸分析应用，特别是核酸测序应用。本公开内容包括包含多样性的组的试剂的珠子组合物，如附接至大量含有条形码序列的寡核苷酸的多样性珠子文库，及其制备和使用方法。

[0099] 制备珠子的方法通常包括，例如，将珠子前体(如单体或聚合物)、引物和交联剂在水溶液中混合，将所述水溶液与油相混合，有时使用微流体装置或小滴发生器，以及使油包水小滴形成。在一些情况下，可在小滴形成之前或之后添加催化剂，如加速剂和/或引发剂。在一些情况下，可通过施加能量，例如通过施加热或光(例如，紫外线)来实现引发。可以在小滴中发生聚合反应以产生珠子，在一些情况下该珠子与寡核苷酸(例如，引物)的一个或多个拷贝共价连接。可采用多种方法将附加序列连接至功能化的珠子。在一些情况下，将功能化的珠子与模板寡核苷酸(例如，含有条形码)组合并分配，使得平均一个或少数几个模板寡核苷酸与功能化珠子占据相同的分区。尽管分区可以是多种不同类型分区(例如，孔、微孔、管、小瓶、微胶囊等)中的任一种，但在优选的方面，分区可以是乳液内的小滴(例如，水性小滴)。可通过反应如引物延伸反应、连接反应或其他方法将寡核苷酸(例如，条形码)序列附接至分区内的珠子。例如，在一些情况下，将用引物功能化的珠子与包含引物结合位点的模板条形码寡核苷酸组合，使得引物能在珠子上延伸。在多轮扩增之后，使单个条形码序列的拷贝与附接至珠子的多个引物相附接。条形码序列附接至珠子后，可将乳液破坏，并且可将条形码化的珠子(或与另一种类型的扩增产物连接的珠子)与不具有扩增的条形码的珠子分离。然后，可采用例如引物延伸法或其他扩增反应，将附加序列如随机序列(例如，随机N-mer)或靶向序列添加至与珠子结合的条形码序列。该过程可产生条形码化珠子的多样性的大文库。

[0100] 图1A示出了用于产生条形码化珠子的示例方法。首先，可将凝胶前体(例如，线性聚合物和/或单体)、交联剂和引物在水溶液中混合，101。下一步，在微流体装置中，然后可将水溶液与油相混合，102。油相与水溶液的混合可导致油包水小滴的形成，103。在油包水

小滴内,发生凝胶前体的聚合,从而形成包含引物的多个拷贝的珠子,104。含引物的珠子产生之后,可将乳液破坏,105,并回收珠子。可通过例如洗涤和引入到任何合适的溶剂(例如,水性溶剂和非水性溶剂)将回收的珠子与未反应的组分分离。在一些情况下,然后可在另一乳液的小滴中将含引物的珠子与模板条形码序列组合(例如,通过有限稀释法),使得每个小滴包含平均至少一个珠子和平均一个或更少的模板条形码序列分子。可采用附接至珠子的引物,克隆扩增模板条形码序列,从而导致与模板互补的条形码序列的多个拷贝附接至珠子,106。然后,可将条形码化的珠子合并到含条形码或不含条形码的珠子群体中,107。然后,可通过例如富集步骤分离条形码化的珠子。条形码分子还可具有用于在后续处理中利用的附加功能序列组分。例如,可将引物序列引入到包含条形码序列段的相同寡核苷酸中,以使得能够使用含有条形码的寡核苷酸作为用于复制相同核酸的延伸引物,或作为用于后续测序或扩增反应的引物位点。在一个实例中,然后可通过引物延伸或其他扩增反应将随机N-mer序列添加至条形码化的珠子,108,由此可获得条形码化珠子的多样性文库,110,其中这类N-mer序列可提供通用引物序列。同样地,功能序列可包括用于将含有条形码的序列固定到表面上(例如,用于测序应用)的固定序列。为方便讨论,在下文中描述了若干特定的功能序列,诸如P5、P7、R1、R2、样品索引、随机Nmer等,和这些序列的部分序列,以及上述任一种的互补序列。但是,应当理解,这些说明是为了讨论的目的,并且含有条形码的寡核苷酸内包含的多种功能序列中的任一种均可替代这些特定的序列,包括但不限于,不同的附接序列、不同的测序引物区域、不同的n-mer区域(靶向的和随机的)以及具有不同功能(例如,二级结构形成,例如,发夹或其他结构)的序列、探针序列(例如,允许探询寡核苷酸的存在与否或允许沉降(pull down)所产生的扩增子),或多种其他功能序列中的任一种。

[0101] 本公开内容还包括用于核酸分析、特别是用于测序应用的样品制备方法。样品制备通常可包括,例如,从来源获得包含样品核酸的样品,任选地进一步处理样品,使样品核酸与条形码化的珠子混合,以及形成含有包含样品核酸和条形码化珠子的流体小滴的乳液。例如,可借助于微流体装置和/或通过任何合适的乳化方法产生小滴。流体小滴也可包含这样的试剂,该试剂能够溶解、降解或以其他方式破坏条形码化的珠子,和/或破坏与所附接的序列的联接,由此将所附接的条形码序列从珠子释放。可通过降解珠子、使寡核苷酸与珠子分离(如通过切割反应)或两者的组合来释放条形码序列。例如,通过扩增(例如,经由本文所述的扩增方法)流体小滴中的样品核酸,游离的条形码序列可附接至样品核酸。然后可破坏包含流体小滴的乳液,并且如果需要,然后可使用例如其他扩增方法将附加序列(例如,有助于特定测序方法的序列、附加条形码序列等)添加至条形码化的样品核酸。然后可在条形码化的、扩增的样品核酸上进行测序,并采用一种或多种测序算法来解释测序数据。如本文所用的,样品核酸可包括多种核酸中的任一种,包括例如,DNA和RNA,特别包括例如,基因组DNA、cDNA、mRNA、总RNA以及由mRNA或总RNA转录物产生的cDNA。

[0102] 图1B示出了用于对样品核酸进行条形码化并随后测序的示例方法。首先,可从来源获得包含核酸的样品,111,并且例如,可如本文所述获得一组条形码化的珠子,112。珠子优选地与含有一种或多种条形码序列以及引物如随机N-mer或其他引物的寡核苷酸连接。优选地,条形码序列可从条形码化的珠子释放,例如通过条形码与珠子之间的联接的切割,或通过基础珠子的降解来释放条形码,或两者的组合。例如,在某些优选的方面,可用试剂如还原剂降解或溶解条形码化的珠子以释放条形码序列。在该实例中,将包含核酸的样品

113、条形码化的珠子114以及例如还原剂116混合并进行分配。举例而言，这样的分区可包括将组分引入小滴生成系统，诸如微流体装置115。在微流体装置115的帮助下，可形成油包水乳液117，其中乳液含有水性小滴，该水性小滴含有样品核酸、还原剂和条形码化的珠子，117。还原剂可溶解或降解条形码化的珠子，由此从小滴内的珠子释放带有条形码和随机N-mer的寡核苷酸，118。然后随机N-mer可引发样品核酸的不同区域，从而在扩增后产生样品的扩增拷贝，其中每个拷贝均用条形码序列标记，119。优选地，每个小滴含有一组含有相同条形码序列和不同随机N-mer序列的寡核苷酸。随后，将乳液破坏，120，并可通过例如扩增方法(例如，PCR)加入附加序列(例如，有助于特定测序方法的序列、附加条形码等)，122。然后可进行测序，123，并采用算法解释测序数据，124。例如，测序算法通常能够进行条形码的分析，以比对测序读取结果和/或鉴定特定序列读取结果所属的样品。

[0103] 本公开内容的方法和组合物可与任何合适的数字处理器一起使用。数字处理器可被编程，例如，用以操作装置的任何组件和/或执行本文所述的方法。在一些实施方案中，可在与小滴发生器通信的数字处理器的帮助下执行珠子形成。数字处理器可控制小滴形成的速度或控制产生的小滴的总数。在一些实施方案中，可在微流体装置和与该微流体装置通信的数字处理器的帮助下完成条形码序列与样品核酸的附接。在一些情况下，数字处理器可控制向微流体装置的通道提供的样品和/或珠子的量，通道内材料的流速，以及包含条形码序列和样品核酸的小滴的产生速率。

[0104] 本公开内容的方法和组合物可用于多种不同的分子生物学应用，包括但不限于核酸测序、蛋白质测序、核酸定量、测序优化、检测基因表达、定量基因表达、表观遗传学应用以及基因组的或表达的标记物的单细胞分析。此外，本公开内容的方法和组合物具有许多医学应用，包括多种遗传和非遗传疾病和病症(包括癌症)的鉴别、检测、诊断、治疗、分期或风险预测。

[0105] II. 珠子或颗粒

[0106] 本公开内容的方法、组合物、装置和试剂盒可与任何合适的珠子或颗粒(包括凝胶珠子和其他类型珠子)一起使用。珠子可充当将依照本文所述的方法被递送的试剂的载体。特别地，这些珠子可提供表面(试剂可释放地附接至其上)或体积(试剂被夹带在其中或以其他方式可释放地分配)。然后可依照期望的方法(例如，在试剂向离散分区中的受控递送中)递送这些试剂。许多不同的试剂或试剂类型可以与珠子关联，其中可希望向分区递送这类试剂。这类试剂的非限制性实例包括，例如，酶、多肽、抗体或抗体片段、标记试剂(例如，染料、荧光团、生色团等)、核酸、多核苷酸、寡核苷酸，以及上述两种或更多种的任意组合。在一些情况下，珠子可提供表面，在该表面上合成或附接寡核苷酸序列。包括寡核苷酸、条形码序列、引物、交联剂等在内的多种实体可以与珠子的外表面关联。在多孔珠子的情况下，实体可与珠子的外表面和内表面均关联。实体可直接附接至珠子的表面(例如，通过共价键、离子键、范德华相互作用等)、可附接至其他已附接至珠子表面的寡核苷酸序列(例如，衔接子或引物)、可在珠子的整个内部扩散和/或可在分区(例如，流体小滴)中与珠子组合。在优选的实施方案中，寡核苷酸共价地附接至珠子的聚合物基质内的位点，因此其存在于珠子的内部和外部。在一些情况下，实体如细胞或核酸被封装在珠子内。包括扩增试剂(例如，PCR试剂、引物)在内的其他实体还可在整个珠子中扩散或在珠子的内部(例如，通过孔，共价附接至聚合物基质)化学地连接。

[0107] 珠子可用来定位实体或样品。在一些实施方案中，实体(例如，寡核苷酸、条形码序列、引物、交联剂、衔接子等)可以与珠子的外表面和/或内表面关联。在一些情况下，实体可位于整个珠子上。在一些情况下，实体可与珠子的整个表面关联或与珠子表面的至少一半关联。

[0108] 珠子可作为支持物，在其上合成寡核苷酸序列。在一些实施方案中，寡核苷酸的合成可包括连接步骤。在一些情况下，寡核苷酸的合成可包括将两个较小的寡核苷酸连接在一起。在一些情况下，通过附接至珠子的引物，引物延伸或其他扩增反应可用于在珠子上合成寡核苷酸。在这类情况下，附接至珠子的引物可以与也含有模板核苷酸序列的寡核苷酸的引物结合位点杂交。然后可通过引物延伸反应或其他扩增反应将引物延伸，并且与模板寡核苷酸互补的寡核苷酸可由此附接至珠子。在一些情况下，与珠子关联的一组相同的寡核苷酸可以与一组多样性的寡核苷酸相连接，使得每个相同的寡核苷酸附接至这组多样性的寡核苷酸的不同成员。在其他情况下，与珠子关联的一组多样性的寡核苷酸可以与一组相同的寡核苷酸相连接。

[0109] 珠子的特征

[0110] 本公开内容的方法、组合物、装置和试剂盒可与任何合适的珠子一起使用。在一些实施方案中，珠子可以是多孔的、无孔的、固体的、半固体的、半流体的或流体的。在一些实施方案中，珠子可以是可溶解的、可破坏的或可降解的。在一些情况下，珠子可以不是可降解的。在一些实施方案中，珠子可以是凝胶珠子。凝胶珠子可以是水凝胶珠子。凝胶珠子可以由分子前体如聚合物或单体物质形成。半固体珠子可以是脂质体珠子。固体珠子可包含金属，包括氧化铁、金和银。在一些情况下，珠子是二氧化硅珠子。在一些情况下，珠子是刚性的。在一些情况下，珠子可以是柔性的。

[0111] 在一些实施方案中，珠子可含有分子前体(例如，单体或聚合物)，该分子前体可通过前体的聚合形成聚合物网络。在一些情况下，前体可以是已经聚合的、能够经历进一步聚合(例如，通过化学交联)的物质。在一些情况下，前体包含丙烯酰胺或甲基丙烯酰胺单体、寡聚物或聚合物中的一种或多种。在一些情况下，珠子可包含预聚物，该预聚物是能够进一步聚合的寡聚物。例如，可采用预聚物制备聚氨酯珠子。在一些情况下，珠子可含有可进一步聚合在一起的单独的聚合物。在一些情况下，可通过不同前体的聚合产生珠子，使得其包含混合的聚合物、共聚物和/或嵌段共聚物。

[0112] 珠子可包含天然的和/或合成的材料，包括天然的和合成的聚合物。天然聚合物的实例包括蛋白质和糖，诸如脱氧核糖核酸、橡胶、纤维素、淀粉(例如，直链淀粉、支链淀粉)、蛋白质、酶、多糖、丝、聚羟基烷酸酯、壳聚糖、葡聚糖、胶原、角叉菜胶、卵叶车前子、阿拉伯树胶、琼脂、明胶、虫胶、梧桐胶、黄原胶、玉米糖胶、瓜尔胶、刺梧桐胶(gum karaya)、琼脂糖、藻酸、藻酸盐或其天然聚合物。合成聚合物的实例包括丙烯酸、尼龙、硅酮、氨纶(spandex)、粘胶人造丝、聚羧酸、聚醋酸乙烯酯、聚丙烯酰胺、聚丙烯酸酯、聚乙二醇、聚氨酯、聚乳酸、二氧化硅、聚苯乙烯、聚丙烯腈、聚丁二烯、聚碳酸酯、聚乙烯、聚对苯二甲酸乙二醇酯、聚(氯三氟乙烯)、聚(环氧乙烷)、聚(对苯二甲酸乙二醇酯)、聚乙烯、聚异丁烯、聚(甲基丙烯酸甲酯)、聚(甲醛)、聚甲醛、聚丙烯、聚苯乙烯、聚(四氟乙烯)、聚(醋酸乙烯酯)、聚(乙烯醇)、聚(氯乙烯)、聚(偏二氯乙烯)、聚(偏二氟乙烯)、聚(氟乙烯)及其组合(例如，共聚物)。珠子也可由除聚合物之外的材料形成，该材料包括脂质、胶束、陶瓷、玻璃陶瓷、材

料复合物、金属、其他无机材料以及其他材料。

[0113] 在一些情况下,化学交联剂可以是在单体聚合过程中用于交联单体的前体和/或可用于用物质将珠子功能化。在一些情况下,聚合物可与交联剂物质或其他类型的单体进一步聚合,以产生进一步聚合的网络。化学交联剂的非限制性实例包括胱胺、戊二醛、辛二亚氨酸二甲酯、N-羟基琥珀酰亚胺交联剂BS3、甲醛、碳二亚胺(EDC)、SMCC、Sulfo-SMCC、乙烯基硅烷(vinylsilane)、N,N'二烯丙基酒石二酰胺(DATD)、N,N'-二(丙烯酰)胱胺(BAC)或其同系物。在一些情况下,在本公开内容中使用的交联剂含有胱胺。

[0114] 根据所使用的特定交联剂,交联可以是永久的或可逆的。可逆交联可使得聚合物线性化,或在合适的条件下离解。在一些情况下,可逆的交联也可使得结合至珠子表面的材料能够可逆附接。在一些情况下,交联剂可以形成二硫键。在一些情况下,形成二硫键的化学交联剂可以是胱胺或修饰的胱胺。在一些实施方案中,二硫键可在分子前体单元(例如,单体、寡聚物或线性聚合物)之间形成。在一些实施方案中,二硫键可在分子前体单元(例如,单体、寡聚物或线性聚合物)或引入到珠子和寡核苷酸中的前体之间形成。

[0115] 例如,胱胺(包括修饰的胱胺)是包含二硫键的有机试剂,其可用作珠子的单独单体或聚合前体之间的交联剂。聚丙烯酰胺可在胱胺或包含胱胺的物质(例如,修饰的胱胺)的存在下聚合,以产生包含二硫键的聚丙烯酰胺凝胶珠子(例如,含有可化学还原的交联剂的可化学降解的珠子)。二硫键可允许珠子在暴露于还原剂时被降解(或溶解)。

[0116] 在至少一个可替代的实例中,一种线性多糖聚合物—壳聚糖,可通过亲水链与戊二醛交联以形成珠子。壳聚糖聚合物的交联可通过由热、压力、pH变化和/或辐射引发的化学反应实现。

[0117] 在一些实施方案中,珠子可在聚合物前体(例如,单体、寡聚物、线性聚合物)、寡核苷酸、引物和其他实体之间包含共价键或离子键。在一些情况下,该共价键包含碳-碳键或硫醚键。

[0118] 在一些情况下,珠子可包含acrydite部分,其在某些方面可用于将一种或多种物质(例如,条形码序列、引物、其他寡核苷酸)附接至珠子。在一些情况下,acrydite部分可以指由acrydite与一种或多种物质反应(例如,在聚合反应期间acrydite与其他单体和交联剂的反应)产生的acrydite类似物。可以修饰acrydite部分以与待附接的物质如寡核苷酸(例如,条形码序列、引物、其他寡核苷酸)形成化学键。例如,acrydite部分可以用能够形成二硫键的巯基基团修饰,或者可以用已经包含二硫键的基团修饰。巯基或二硫键(通过二硫键交换)可用作待附接的物质的锚定点,或者acrydite部分的另一部分可用于附接。在一些情况下,附接是可逆的,使得当二硫键断裂时(例如,在还原剂的存在下),试剂从珠子释放。在其他情况下,acrydite部分包含可用于附接的反应性羟基基团。

[0119] 为附接其他物质(例如,核酸)对珠子的功能化可通过多种不同的途径实现,包括聚合物内化学基团的活化、聚合物结构中活性或可激活的官能团的引入、或在珠子产生的预聚物或单体阶段的附接。

[0120] 例如,在一些实例中,聚合以形成珠子的前体(例如,单体、交联剂)可包含acrydite部分,使得当珠子产生时,该珠子还包含acrydite部分。通常,acrydite部分附接至寡核苷酸序列,如期望引入珠子中的引物(例如,用于扩增靶核酸和/或对靶核酸条形码序列、结合序列等进行测序中的一种或多种的引物)。在一些情况下,该引物包含P5序列。例

如,丙烯酰胺前体(例如,交联剂、单体)可包含acrydite部分,使得当其聚合以形成珠子时,该珠子还包含acrydite部分。

[0121] 在一些情况下,诸如单体和交联剂的前体可包含例如单个寡核苷酸(例如,如引物或其他序列)或其他物质。图29A描绘了示例单体,其包含acrydite部分和通过二硫键与acrydite部分连接的单个P5序列。在一些情况下,诸如单体和交联剂的前体可包含多个寡核苷酸、其他序列或其他物质。图29B描绘了包含多个acrydite部分的示例单体,每个acrydite部分均通过二硫键与P5引物连接。此外,图29C描绘了包含多个acrydite部分的示例交联剂,每个acrydite部分均通过二硫键与P5物质连接。在每个前体中包含多个acrydite部分或其他连接体物质可改善所连接的物质(例如,寡核苷酸)向由前体产生的珠子中的负载,因为每个前体均包含待负载的物质的多个拷贝。

[0122] 在一些情况下,包含反应性官能团或能够被激活从而使其变为反应性的官能团的前体可与其他前体聚合,以形成包含激活的或可激活的官能团的凝胶珠子。然后,官能团可用于使附加物质(例如,二硫键连接体、引物、其他寡核苷酸等)附接至凝胶珠子。例如,如图31所示,一些包含羧酸(COOH)基团的前体可与其他前体共聚,以形成还包含COOH官能团的凝胶珠子。在一些情况下,丙烯酸(包含游离COOH基团的物质)、丙烯酰胺和二(丙烯酰)胱胺可以共聚在一起以生成包含游离羟基的凝胶珠子。该凝胶珠子的COOH基团可被激活(例如,如图31所示,通过1-乙基-3-(3-二甲基氨基丙基)碳二亚胺(EDC)和N-羟基琥珀酰亚胺(NHS)或4-(4,6-二甲氧基-1,3,5-三嗪-2-基)-4-二甲基吗啉鎓氯化物(DMTMM),使其是反应性的(例如,在EDC/NHS或DMTMM用于激活的情况下,对胺官能团是反应性的)。然后,激活的COOH基团可以与包含将要连接至珠子的部分的合适物质(例如,在羧酸基团被激活为对胺官能团具有反应性的情况下,包含胺官能团的物质)反应。

[0123] 包含通过二硫键与P5引物连接的胺基团(例如, $\text{H}_2\text{N}-\text{C}_6-\text{S}-\text{S}-\text{C}_6-\text{P}_5$)的示例物质在图31中示出。可用EDC/NHS或DMTMM激活凝胶珠子的COOH官能团,以在一个或多个COOH位点处产生胺反应性物质。然后,如图31所示,物质 $\text{H}_2\text{N}-\text{C}_6-\text{S}-\text{S}-\text{C}_6-\text{P}_5$ 部分的胺基团可与激活的羧酸反应,使得该部分和所附接的P5寡核苷酸变得共价连接至珠子。可将未反应的COOH物质转化为其他物质,使其被阻断。

[0124] 可通过使一些二硫键还原为游离的巯基,将在其聚合物网络中包含二硫键的珠子用附加物质功能化。例如,可通过还原剂(例如,DTT、TCEP等)的作用,在不溶解珠子的情况下使二硫键还原,以产生游离的巯基基团。然后,珠子的游离巯基可以与物质的游离巯基或者与包含另一个二硫键的物质反应(例如,通过巯基-二硫键交换),使得该物质可连接至珠子(例如,通过产生的二硫键)。然而,在一些情况下,珠子的游离巯基可以与任何其他合适的基团反应。例如,珠子的游离巯基可以与包含acrydite部分的物质反应。珠子的游离巯基基团可通过迈克尔加成化学与acrydite反应,使得包含acrydite的物质连接至珠子。在一些情况下,可以通过包含巯基封端剂例如N-乙基马来酰亚胺和碘乙酸盐来防止不受控制的反应。

[0125] 可以控制珠子内的二硫键的活化,使得只有少量的二硫键被激活。例如,可通过控制用于产生游离巯基基团的还原剂的浓度和/或控制用于在珠子聚合中形成二硫键的试剂的浓度来施加控制。在一些情况下,可将低浓度(例如,还原剂分子:凝胶珠子的比例小于约10000、100000、1000000、10000000、100000000、1000000000或10000000000)

的还原剂用于还原。控制被还原成游离巯基的二硫键的数目对于在功能化期间确保珠子结构完整性可能是有用的。在一些情况下，光活性剂如荧光染料可通过珠子的游离巯基偶联至珠子，并用于定量珠子中存在的游离巯基的数目和/或追踪珠子。

[0126] 用于对包含二硫键的凝胶珠子进行功能化的示例方案在图35A中示出。如图所示，可采用例如任何本文所述的方法产生包含二硫键的珠子3501(例如，凝胶珠子)。当还原剂3502(例如，DTT、TCEP或本文所述的任何其他还原剂)在不适合于珠子降解的浓度下作用时，一些凝胶珠子3501的二硫键可被还原为游离的巯基，以产生包含游离巯基的珠子3503。一旦去除还原剂(例如，通过洗涤)3504，珠子3503可以与acrydite-S-S-物质部分3505反应，该部分3505包含通过二硫键与acrydite连接的待负载的物质(例如，示出的是P5寡核苷酸，但是该物质可以是另一种类型的多核苷酸，诸如，例如，包含P5、条形码、R1和随机N-mer的寡核苷酸)。部分3505可通过迈克尔加成化学与凝胶珠子3503偶联，以产生包含部分3505的珠子3506。然后，可通过对去除不想要的(例如，未附接的)物质纯化所产生的珠子3506。

[0127] 用于对包含二硫键的凝胶珠子进行功能化的另一个示例方案在图35B中示出。如图所示，可采用例如任何本文所述的方法产生包含二硫键的珠子3501(例如，凝胶珠子)。当还原剂3502(例如，DTT、TCEP或本文所述的任何其他还原剂)在不适合于珠子降解的浓度下作用时，一些凝胶珠子3501的二硫键可被还原为游离的巯基，以产生包含游离巯基的珠子3503。一旦去除还原剂(例如，通过洗涤)3504，珠子3503可以与2,2'-二硫吡啶3507反应以产生通过二硫键与吡啶部分连接的凝胶珠子3509。作为2,2'-二硫吡啶的替代物，可使用其他相似的物质，如4,4'-二硫吡啶或5,5'-二硫双-(2-硝基苯甲酸)(例如，DTNB或Ellman试剂)。2,2'-二硫吡啶3507可通过二硫键交换与凝胶珠子3503偶联，以产生包含通过二硫键与珠子3509连接的吡啶部分的珠子3509。然后，可将凝胶珠子3509与未反应的物质分离(例如，通过洗涤)。

[0128] 然后，纯化的凝胶珠子3509可与部分3508反应，该部分3508包含将要偶联至凝胶珠子的感兴趣的物质(例如，如图所示的P5寡核苷酸)和游离的巯基基团。在一些情况下，部分3508可由包含二硫键的另一物质产生，使得当二硫键被还原(例如，通过还原剂如DTT、TCEP等的作用)时，获得具有游离巯基基团的部分3508。部分3508可参与与珠子3509的吡啶基团的巯基-二硫键交换，以产生包含部分3508的凝胶珠子3510。吡啶基团通常是很好的离去基团，其可允许与部分3508的游离巯基的有效巯基-二硫键交换。然后，可通过对去除不想要的物质纯化(例如，通过洗涤)所产生的珠子3510。

[0129] 在一些情况下，在凝胶珠子形成之后向凝胶珠子添加部分可以是有利的。例如，在凝胶珠子形成后物质的添加可避免在聚合期间可能发生的链转移终止过程中的物质损失。此外，较小的前体(例如，不包含侧链基团和连接的部分的单体或交联剂)可用于聚合，并可最低限度地阻碍由粘滞效应引起的生长链末端。在一些情况下，在凝胶珠子合成之后的功能化可使待负载的物质(例如，寡核苷酸)在可能的损伤剂(例如，自由基)和/或化学环境中的暴露最小化。在一些情况下，所产生的凝胶可具有上临界溶解温度(UCST)，UCST可允许珠子的温度驱使的溶胀和瓦解。这种功能性可在随后用物质对珠子的功能化期间帮助物质(例如，引物，P5引物)渗透到珠子中。产生后的功能化还可用于控制珠子中物质的负载比，使得例如负载比的可变性最小化。另外，物质负载可以在分批工艺中进行，使得多个珠子可在单批中用该物质功能化。

[0130] 在一些情况下,与前体连接的acrydite部分、与前体连接的另一物质或前体本身包含不稳定的键,诸如,例如,化学、热或光敏感的键,例如,二硫键、UV敏感的键等。一旦包含不稳定键的acrydite部分或其他部分被引入珠子中,该珠子也可包含该不稳定键。例如,不稳定的键可以在将物质(例如,条形码、引物等)可逆地连接(共价连接)至珠子时是有用的。在一些情况下,热不稳定的键可包括基于核酸杂交的附接(例如,当寡核苷酸与附接至珠子的互补序列杂交时),使得杂合体的热解链从珠子或微胶囊释放寡核苷酸,例如,含有条形码的序列。此外,向凝胶珠子添加多种类型的不稳定键可导致能够响应于不同刺激的珠子的产生。每种类型的不稳定键可以对相关的刺激(例如,化学刺激、光、温度等)敏感,使得可通过施加合适的刺激来控制通过每种不稳定键附接至珠子的物质的释放。这种功能性对物质从凝胶珠子上的受控释放是有用的。在一些情况下,例如,通过如上文所述的凝胶珠子的活化官能团,可在凝胶珠子形成后将包含不稳定键的另一物质连接至凝胶珠子。应当理解,可释放地、可切割地或可逆地附接至本文所述的珠子的条形码包括通过条形码分子与珠子之间的联接的切割而被释放或可释放的条形码,或通过基础珠子自身的降解而被释放的条形码,从而使条形码能够被其他试剂接近或可接近,或包括这两者。通常,如本文所述的可释放的条形码通常可被称为可激活的,因为其一旦被释放即可用于反应。因此,例如,可通过将条形码从珠子(或本文所述的其他合适类型的分区)释放来激活可激活的条形码。应当理解,在描述的方法和系统的背景下也想到其他可激活的配置。特别地,可提供可释放地附接至珠子或以其他方式布置在分区中的试剂(带有关联的可激活的基团),使得一旦递送至期望的一组试剂(例如,通过共同分配),可激活的基团可以与期望的试剂反应。这类可激活的基团包括笼罩基团、可去除的阻断或保护基团,例如,光不稳定基团、热不稳定基团,或可化学去除的基团。

[0131] 除热可切割的键、二硫键和UV敏感的键之外,可与前体或珠子偶联的不稳定键的其他非限制性实例还包括酯联接(例如,可用酸、碱或羟胺切割的)、邻二醇联接(例如,可通过高碘酸钠切割的)、Diels-Alder联接(例如,可通过热切割的)、砜联接(例如,可通过碱切割的)、甲硅烷基醚联接(例如,可通过酸切割的)、糖苷联接(例如,可通过淀粉酶切割的)、肽联接(例如,可通过蛋白酶切割的)或磷酸二酯联接(例如,可通过核酸酶(DNA酶)切割的)。

[0132] 珠子可以与不同数目的acrydite部分连接。例如,珠子可包含约1、10、100、1000、10000、100000、1000000、10000000、100000000或1000000000个连接至珠子的acrydite部分。在其他实例中,珠子可包含至少1、10、100、1000、10000、100000、1000000、10000000或100000000个连接至珠子的acrydite部分。例如,珠子可包含约1、10、100、1000、10000、100000、1000000、10000000、100000000或1000000000个共价连接至珠子(如通过acrydite部分)的寡核苷酸。在其他实例中,珠子可包含至少1、10、100、1000、10000、100000、1000000、10000000、100000000或1000000000个共价连接至珠子(如通过acrydite部分)的寡核苷酸。

[0133] 还可在珠子产生期间(例如,在前体聚合期间)将不参与聚合的物质封装在珠子中。这类物质可以进入聚合反应混合物,使得在珠子形成时所产生的珠子包含该物质。在一些情况下,这类物质可在凝胶珠子形成后添加至凝胶珠子。这类物质可包括,例如,寡核苷酸、包括本文所述的那些的核酸扩增反应所需的物质(例如,引物、聚合酶、dNTP、辅因子(例

如,离子型辅因子))、酶反应所需的物质(例如,酶、辅因子、底物)或核酸修饰反应如聚合、连接或消化所需的物质。这类物质的捕获可通过在前体聚合期间产生的聚合物网络密度、凝胶珠子内离子电荷的控制(例如,通过与聚合的物质连接的离子型物质)或通过其他物质的释放来控制。封装的物质可在珠子降解时和/或通过施加能够使物质从珠子释放的刺激而从珠子释放。

[0134] 珠子可以具有均匀的大小或不均匀的大小。在一些情况下,珠子的直径可以为约1 μm 、5 μm 、10 μm 、20 μm 、30 μm 、40 μm 、45 μm 、50 μm 、60 μm 、65 μm 、70 μm 、75 μm 、80 μm 、90 μm 、100 μm 、250 μm 、500 μm 或1mm。在一些情况下,珠子可以具有至少约1 μm 、5 μm 、10 μm 、20 μm 、30 μm 、40 μm 、45 μm 、50 μm 、60 μm 、65 μm 、70 μm 、75 μm 、80 μm 、90 μm 、100 μm 、250 μm 、500 μm 、1mm或更大的直径。在一些情况下,珠子可以具有小于约1 μm 、5 μm 、10 μm 、20 μm 、30 μm 、40 μm 、45 μm 、50 μm 、60 μm 、65 μm 、70 μm 、75 μm 、80 μm 、90 μm 、100 μm 、250 μm 、500 μm 或1mm的直径。在一些情况下,珠子可以具有在约40–75 μm 、30–75 μm 、20–75 μm 、40–85 μm 、40–95 μm 、20–100 μm 、10–100 μm 、1–100 μm 、20–250 μm 或20–500 μm 的范围内的直径。

[0135] 在某些优选的方面,珠子以具有相对单分散的大小分布的珠子群体提供。应当理解,在一些期望在分区提供相对一致量的试剂的应用中,维持相对一致的珠子特征如大小有助于总体的一致性。特别地,本文所述的珠子可以具有这样的大小分布,该大小分布具有小于50%、小于40%、小于30%、小于20%的珠子横截面尺寸的变化系数,并且在一些情况下,该变化系数小于15%、小于10%或者甚至小于5%。

[0136] 珠子可以具有规则形状或不规则形状。珠子形状的实例包括球形、非球形、椭圆形、长方形、无定形的、圆形、圆柱形和其同系物。

[0137] 可降解的珠子

[0138] 除了上述珠子与相关联的分子(例如,含有条形码的寡核苷酸)之间的可切割联接之外或作为其替代,珠子可以是自发地或在暴露于一种或多种刺激(例如,温度变化、pH变化、暴露于特定化学物质或化学相、暴露于光、还原剂等)时可降解的、可破坏的或可溶解的。在一些情况下,珠子可以是可溶解的,使得当暴露于特定化学物质或环境变化(诸如,例如温度或pH)时,珠子的材料组分溶解。例如,凝胶珠子可以在升高的温度和/或碱性条件下降解或溶解。在一些情况下,珠子可以是可热降解的,使得当珠子暴露于合适的温度变化(例如,热)时,珠子降解。与物质(例如,核酸物质)结合的珠子的降解或溶解可导致物质从珠子上的释放。

[0139] 可降解的珠子可包含一种或多种具有不稳定键的物质,使得当珠子/物质暴露于合适的刺激时,该键被破坏并且珠子降解。该不稳定键可以是化学键(例如,共价键、离子键),或者可以是另一种类型的物理相互作用(例如,范德华相互作用、偶极-偶极相互作用等)。在一些情况下,用于产生珠子的交联剂可包含不稳定键。当暴露于合适的条件时,不稳定键被破坏并且珠子被降解。例如,聚丙烯酰胺凝胶珠子可包含胱胺交联剂。当珠子暴露于还原剂时,胱胺的二硫键被破坏并且珠子被降解。

[0140] 可降解的珠子对于在向珠子施加合适的刺激时更快地从珠子释放所附接的物质(例如,寡核苷酸、条形码序列)可以是有用的。例如,对于结合至多孔珠子内表面的物质或在封装的物质的情况下,当珠子降解时,该物质可以具有更高的迁移性和对溶液中其他物质的易接近性。在一些情况下,还可通过可降解的连接体(例如,二硫键连接体)将物质附接

至可降解的珠子。可降解的连接体可以与可降解的珠子响应于相同的刺激，或者两种可降解的物质可以响应于不同的刺激。例如，可通过二硫键将条形码序列附接至包含胱胺的聚丙烯酰胺珠子。当条形码化的珠子暴露于还原剂时，珠子降解，并且在条形码序列与珠子之间的二硫键以及珠子中胱胺的二硫键被破坏时条形码序列得到释放。

[0141] 可将可降解的珠子引入到分区如乳液的小滴或孔中，使得当施加合适的刺激时，分区内的珠子降解并且任何关联的物质均释放到小滴中。游离的物质可以与其他物质相互作用。例如，可将包含胱胺且与条形码序列通过二硫键连接的聚丙烯酰胺珠子与还原剂在油包水乳液的小滴内混合。在小滴内，该还原剂破坏多个二硫键，导致珠子降解和条形码序列向小滴的水性内环境中的释放。在另一个实例中，将包含结合珠子的条形码序列的小滴在碱性溶液中加热也可导致珠子降解和所附接的条形码序列向小滴的水性内环境中的释放。

[0142] 从上述公开内容将会理解，尽管称为珠子的降解，但在许多上面提到的情况下，该降解可以指结合的或所夹带的物质从珠子上的离解，伴随和不伴随物理珠子自身的结构降解。例如，可通过例如由化学环境改变导致的渗透压差，使所夹带的物质从珠子释放。举例而言，通常可能在珠子自身没有结构降解的情况下发生由渗透压差引起的珠子孔大小的改变。在一些情况下，由珠子的渗透性溶胀导致的孔大小的增加可使得珠子中所夹带的物质能够释放。在其他情况下，由于孔大小收缩，珠子的渗透性萎缩可使珠子更好地保持所夹带的物质。

[0143] 应当理解，当提供可降解的珠子时，可能期望避免在所需的时间之前将这类珠子暴露于导致这类降解的一种或多种刺激，以便避免珠子过早降解以及由这种降解所导致的问题，包括例如流动特性差、结块和聚集。举例而言，当珠子包含可还原的交联基团如二硫键基团时，将期望避免使这类珠子与还原剂（例如，DTT或其他二硫键切割试剂）接触。在这类情况下，对本文所述的珠子的处理在一些情况下将是无还原剂如DTT的。因为还原剂常常在商用的酶制剂中提供，所以通常期望在处理本文所述的珠子时提供无还原剂（或无DTT）的酶制剂。这类酶的实例包括，例如，聚合酶制剂、连接酶制剂以及许多其他可用于处理本文所述的珠子的酶制剂。所谓“无还原剂的”或“无DTT的”制剂是指对于在降解珠子中使用的这类材料，该制剂将具有小于其下限的1/10、小于1/50和甚至小于1/100。例如，对于DTT，无还原剂的试剂通常将含有少于0.01mM、0.005mM、0.001mM的DTT、0.0005mM DTT或者甚至少于0.0001mM的DTT或更少。在许多情况下，DTT的量将是检测不到的。

[0144] 用于降解珠子的方法

[0145] 在一些情况下，刺激可用于触发珠子的降解，这可导致内含物从珠子的释放。通常，刺激可导致珠子结构的降解，如共价键或其他类型的物理相互作用的降解。这些刺激在诱导珠子降解和/或释放其内含物方面可能是有用的。如下面更详细描述的，可使用的刺激的实例包括化学刺激、热刺激、光刺激及其任意组合。

[0146] 许多化学触发物可用于触发珠子的降解。这些化学变化的实例可包括但不限于珠子内组分完整性的pH介导的变化、通过交联键的切割发生的珠子组分的降解以及珠子组分的解聚。

[0147] 在一些实施方案中，珠子可以由包含可降解化学交联剂如BAC或胱胺的材料形成。这类可降解交联剂的降解可通过多种机制实现。在一些实例中，可以使珠子与可诱导氧化、

还原或其他化学变化的化学降解剂接触。例如，化学降解剂可以是还原剂，如二硫苏糖醇 (DTT)。还原剂的其他实例可包括β-巯基乙醇、(2S)-2-氨基-1,4-二巯基丁烷 (二硫丁胺或 DTBA)、三 (2-羧乙基) 脍 (TCEP) 或其组合。还原剂可以降解在形成珠子的凝胶前体之间形成的二硫键，因此使珠子降解。在其他情况下，溶液 pH 的变化，如 pH 增加，可触发珠子的降解。在其他情况下，暴露于水溶液如水可触发水解降解，因此使珠子降解。

[0148] 当施加热刺激时，也可诱导珠子释放其内含物。温度的变化可导致珠子的多种变化。例如，热可导致固体珠子液化。热的变化可导致珠子的熔化，使得珠子的一部分降解。在其他情况下，热可增加珠子组分的内部压力，使得珠子破裂或爆炸。热还可作用于用作构建珠子的材料的热敏性聚合物。

[0149] 本公开内容的方法、组合物、装置和试剂盒可以与任何合适的试剂一起使用以降解珠子。在一些实施方案中，温度或 pH 的变化可用于降解珠子内的热敏感的或 pH 敏感的键。在一些实施方案中，化学降解剂可用于通过氧化、还原或其他化学变化降解珠子内的化学键。例如，化学降解剂可以是还原剂，如 DTT，其中 DTT 可以降解在交联剂与凝胶前体之间形成的二硫键，从而使珠子降解。在一些实施方案中，可添加还原剂以降解珠子，这可导致或可不导致珠子释放其内含物。还原剂的实例可包括二硫苏糖醇 (DTT)、β-巯基乙醇、(2S)-2-氨基-1,4-二巯基丁烷 (二硫丁胺或 DTBA)、三 (2-羧乙基) 脍 (TCEP) 或其组合。还原剂可以以 0.1mM、0.5mM、1mM、5mM 或 10mM 存在。还原剂可以以高于 0.1mM、0.5mM、1mM、5mM、10mM 或更高的浓度存在。还原剂可以以低于 0.1mM、0.5mM、1mM、5mM 或 10mM 存在。

[0150] 降解步骤的时机

[0151] 可以降解珠子以释放附接至和包含在珠子内的内含物。该降解步骤可以在样品与珠子组合的同时发生。该降解步骤可以在样品与珠子在可在微流体装置中形成的流体小滴内组合时同时发生。该降解步骤可以在样品与珠子在可在微流体装置中形成的流体小滴内组合之后发生。应当理解，在许多应用中可以不发生降解步骤。

[0152] 还原剂可以与样品组合，然后与珠子组合。在一些情况下，可将还原剂与样品同时引入微流体装置。在一些情况下，可在引入样品之后将还原剂引入微流体装置。在一些情况下，可使样品与还原剂在微流体装置中混合，然后与微流体装置中的凝胶珠子接触。在一些实施方案中，可使样品与还原剂预混合，然后将其加入装置中并与凝胶珠子接触。

[0153] 可降解的珠子可以在施加在合适的刺激时立即降解。在其他情况下，珠子的降解可随时间发生。例如，珠子可以在施加合适的刺激时立即降解或在约 0、0.01、0.1、0.5、1、1.5、2.0、2.5、3.0、3.5、4.0、4.5、5.0、5.5、6.0、6.5、7.0、7.5、8.0、8.5、9.0、9.5、10.0、11、12、13、14、15 或 20 分钟之内降解。在其他实例中，珠子可以在施加合适的刺激时立即降解或在至多约 0、0.01、0.1、0.5、1、1.5、2.0、2.5、3.0、3.5、4.0、4.5、5.0、5.5、6.0、6.5、7.0、7.5、8.0、8.5、9.0、9.5、10.0、11、12、13、14、15 或 20 分钟之内降解。

[0154] 珠子还可以相对于与样品组合在不同的时间降解。例如，珠子可以与样品组合并随后在稍后的时间点降解。样品与珠子组合与随后的珠子降解之间的时间可以为约 0.0001、0.001、0.01、1、10、30、60、300、600、1800、3600、18000、36000、86400、172800、432000 或 864000 秒。样品与珠子组合与随后的珠子降解之间的时间可以超过约 0.0001、0.001、0.01、1、10、30、60、300、600、1800、3600、18000、36000、86400、172800、432000、864000 秒或更长。样品与珠子组合与随后的珠子降解之间的时间可以少于约 0.0001、

0.001、0.01、1、10、30、60、300、600、1800、3600、18000、36000、86400、172800、432000或864000秒。

[0155] 制备用寡核苷酸预功能化的珠子

[0156] 可采用多种方法产生本文所述的珠子。在一些情况下，珠子可以由含有分子前体(例如，线性聚合物、单体、交联剂)的液体形成。然后使该液体经历聚合反应，并由此硬化或胶凝成为珠子(或凝胶珠子)。该液体还可含有在聚合过程中引入珠子中的实体如寡核苷酸。该引入可通过与珠子共价或非共价缔合而实现。例如，在一些情况下，可在形成期间使寡核苷酸夹带在珠子中。或者，它们可以在形成过程中或形成之后偶联至珠子或珠子框架。通常，寡核苷酸与acrydite部分连接，该acrydite部分在聚合过程中变得与珠子交联。在一些情况下，寡核苷酸通过二硫键附接至acrydite部分。结果，形成了包含珠子-acrydite-S-S-寡核苷酸联接的组合物。图4A是用acrydite连接的引物功能化的珠子的示例性图示。

[0157] 在一个示例性过程中，可通过将多种聚合物和/或单体与一种或多种寡核苷酸例如一种或多种包含引物(例如，通用引物、测序引物)的寡核苷酸混合产生功能化的珠子。该聚合物和/或单体可包含丙烯酰胺并且可以交联，使得在聚合物和/或单体之间形成二硫键，从而导致形成硬化的珠子。该寡核苷酸可在硬化的珠子形成过程中(例如，同时地)共价连接至所述多种聚合物和/或单体，或者可在硬化的珠子形成之后(例如，顺序地)共价连接至所述多种聚合物和/或单体。在一些情况下，该寡核苷酸可通过acrydite部分连接至珠子。

[0158] 在多数情况下，一个群体的珠子均用相同的寡核苷酸如通用引物或引物结合位点预功能化。在一些情况下，珠子群体中的珠子用多种不同的寡核苷酸预功能化。这些寡核苷酸可任选地包含多种不同功能序列中的任何序列，例如，用于在随后的珠子处理或应用中使用。功能序列可包括，例如，引物序列(诸如靶向引物序列、通用引物序列，例如，足够短从而能够与样品核酸上的许多不同位置杂交并引发从所述位置延伸的引物序列，或随机引物序列)、附接或固定序列、连接序列、发夹序列、标记序列(例如，条形码或样品索引序列)或多种其他核苷酸序列中的任何序列。

[0159] 举例而言，在一些情况下，通用引物(例如，P5或其他合适的引物)可用作每个珠子上的引物，以将附加内含物(例如，条形码、随机N-mer、其他功能序列)附接至珠子。在一些情况下，通用引物(例如，P5)还可与测序装置相兼容，并且可稍后使得期望的链能够附接至测序装置内的流动池。例如，这类附接或固定序列可提供约束在测序装置中流动池表面的寡核苷酸的互补序列，以使得序列能够固定到该表面上以供测序。或者，这类附接序列可额外地在附接至珠子的寡核苷酸序列内提供或添加至其上。在一些情况下，珠子和其附接的物质可提供为与后续分析过程如测序装置或系统相兼容。在一些情况下，可将超过一种引物附接至珠子，并且超过一种引物可含有通用序列，以便例如使得寡核苷酸以及任何偶联至该序列的附加序列能够在不同的顺序或平行处理步骤中差异处理，例如，第一引物用于靶序列的扩增，而第二引物用于扩增产物的测序。例如，在一些情况下，附接至珠子的寡核苷酸将包含用于进行第一扩增或复制过程(例如，沿着靶核酸序列使引物延伸)的第一引物序列，以便产生扩增的条形码化靶序列。通过还在寡核苷酸内包含测序引物，所得的扩增靶序列将包含这样的引物，并且容易地转移至测序系统中。例如，在一些情况下，例如，当希望采用例如Illumina测序系统对扩增的靶标进行测序时，还可将R1引物或引物结合位点附接

至珠子。

[0160] 引入珠子中的实体可包括具有如上所述的多种功能序列中的任何序列的寡核苷酸。例如，这些寡核苷酸可包括以下的任一个或多个：P5、R1和R2序列、不可切割的5' acrydite-P5、可切割的5' acrydite-SS-P5、R1c、测序引物、读取引物、通用引物、P5_U、通用读取引物和/或任意这些引物的结合位点。在一些情况下，引物可含有一种或多种修饰的核苷酸、核苷酸类似物或核苷酸模拟物。例如，在一些情况下，寡核苷酸可包括肽核酸(PNA)、锁核酸(LNA)核苷酸，等等。在一些情况下，这些寡核苷酸可另外或备选地包括可被不同地处理的核苷酸或类似物，以便允许在其应用的不同步骤中差异处理。例如，在一些情况下，一种或多种功能序列可包括不受特定的聚合酶处理的核苷酸或类似物，因此在使用该酶的过程步骤中未被拷贝。例如，在一些情况下，寡核苷酸的一种或多种功能序列组分将包括例如，含有尿嘧啶的核苷酸、含有非天然碱基的核苷酸、阻断寡核苷酸、阻断的3'端、3' ddCTP。图19提供了其他的实例。应当理解，根据特定的应用，任何这些实体的序列均可作为引物或引物结合位点起作用。

[0161] 聚合可自发地发生。在一些情况下，聚合可由引发剂和/或加速剂、电磁辐射、温度变化(例如，添加或去除热)、pH变化、其他方法及其组合来引发。引发剂可以指能够通过激活(例如，通过产生自由基)聚合反应中所用的一种或多种前体而引发聚合反应的物质。加速剂可以指能够加快聚合反应发生的速率的物质。在一些情况下，加速剂可以加快随后用于激活单体(例如，通过产生自由基)的引发剂的活化(例如，通过产生自由基)，并因此引发聚合反应。在一些情况下，引发剂更快的活化可导致更快的聚合速率。然而，在一些情况下，加速还可通过非化学手段实现，诸如热(例如，添加和去除热)手段、不同类型的辐射手段(例如，可见光、紫外线等)或任何其他合适的手段。为创建含有分子前体(其然后可聚合形成硬化的珠子)的小滴，可采用乳液技术。例如，可将分子前体添加至水溶液。然后可以用油将该水溶液乳化(例如，通过搅拌、微流体小滴发生器或其他方法)。然后，该分子前体可在乳化的小滴中聚合以形成珠子。

[0162] 例如，可通过任何合适的方法制备乳液，该方法包括本领域公知的方法，如批量振荡、批量搅拌、流动聚焦和微筛分(见例如，Weizmann等人，Nature Methods, 2006, 3 (7) : 545-550; Weitz等人，美国公开号2012/0211084)。在一些情况下，可使用微流体装置制备乳液。在一些情况下，可使用油包水乳液。这些乳液中可引入氟表面活性剂，如具有含PEG的化合物如bis krytox peg(BKP)的Krytox FSH。在一些情况下，可使用水包油乳液。在一些情况下，可形成多分散乳液。在一些情况下，可形成单分散乳液。在一些情况下，单分散乳液可在微流体流动聚焦装置中形成。(Gartecki等人，Applied Physics Letters, 2004, 85 (13) : 2649-2651)。

[0163] 在至少一个实例中，用于制备珠子的微流体装置可含有通道区段，该通道区段在组合两条或更多条不混溶流体流(如含有分子前体和油的水溶液)的单个交叉点处相交。在单个交叉点处组合两条不混溶的流体可导致形成流体小滴。所形成的流体小滴的大小可取决于流体流进入流体交叉点的流速、两种流体的性质以及微流体通道的大小。在离开流体交叉点的流体小滴形成后引发聚合可导致由流体小滴形成硬化的珠子。用于产生小滴，以便同时用于珠子形成和用于将珠子分配至如本文其他各处讨论的离散小滴的微流体装置、通道网络和系统的实例在例如于2014年4月4日提交的美国临时专利申请号61/977,804中

描述,其为了所有目的通过引用以全文并入本文。

[0164] 为了操控单独的分子前体、寡聚物或聚合物何时开始聚合以形成硬化的珠子,可在珠子形成过程中的不同时间点添加引发剂和/或加速剂。加速剂可以是这样一种试剂,其可引发聚合过程(例如,在一些情况下,通过聚合引发剂的活化)并因此可缩短珠子硬化的时间。在一些情况下,单种加速剂或多种加速剂可用于聚合。仔细调整加速对实现合适的聚合反应可能是重要的。例如,如果加速过快,重量和过度的链转移事件可能导致较差的凝胶结构和任何期望的物质的低负载。如果加速过慢,由于聚合物缠结和高粘度,高分子量聚合物可产生受限制的活化位点(例如,自由基)。高粘度可阻碍用于珠子负载的物质的扩散,从而导致物质的低负载或无负载。例如,可通过选择合适的加速剂、合适的加速剂组合或通过选择合适的加速剂以及任何能够调节加速剂作用的刺激(例如,热、电磁辐射(例如,光、紫外线)、另一种化学物质等)来实现加速剂作用的调整。引发剂作用的调整也可以类似的方式实现。

[0165] 加速剂可以是水溶性的、油溶性的,或者可以既是水溶性的又是油溶性的。例如,加速剂可以是四甲基乙二胺(TMEDA或TEMED)、二甲基乙二胺、N,N,N',N'-四甲基甲烷二胺、N,N'-二吗啉基甲烷或N,N,N',N'-四(2-羟基丙基)乙二胺。例如,引发剂可以是过硫酸铵(APS)、钙离子或如图32所示的任何化合物(I-IX)。图32所示的化合物(I-IX)可充当水溶性的基于偶氮的引发剂。基于偶氮的引发剂可以在TEMED和APS不存在时使用,并可以充当基于热的引发剂。基于热的引发剂可通过热激活物质(例如,通过产生自由基),因此引发剂作用的速率可通过温度和/或引发剂的浓度来调整。聚合加速剂或引发剂可包含官能团,该官能团包括膦酸酯、磺酸酯、羧酸酯、羟基、白蛋白结合部分、N-乙烯基基团以及磷脂。聚合加速剂或引发剂可以是低分子量单体化合物。加速剂或引发剂可以a)在小滴产生前加入油中,b)在小滴产生后加入管线中,c)在小滴产生后加入出口储器中,或d)其组合。

[0166] 聚合还可通过电磁辐射引发。某些类型的单体、寡聚物或聚合物可含有光敏性质。因此,可通过将这类单体、寡聚物或聚合物暴露于紫外线、可见光、与敏化剂组合的紫外线、与敏化剂组合的可见光或其组合来引发聚合。敏化剂的实例可以是核黄素。

[0167] 珠子完全聚合或硬化的时间可根据珠子的大小、是否可添加加速剂、何时可添加加速剂、引发剂的类型、何时可施加电磁辐射、溶液的温度、聚合物组成、聚合物浓度以及其他相关参数而变化。例如,聚合可以在约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20分钟后完成。聚合可以在超过约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20分钟或更长时间后完成。聚合可以在少于约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20分钟内完成。

[0168] 可通过连续相交换从乳液(例如,凝胶-水-油)中回收珠子。可向乳液(例如,凝胶-水-油)中添加过量的水性流体,并且硬化的珠子可经历沉降,其中可以使珠子聚集并且可去除含有过量油的上清液。可以重复进行添加过量水性流体、随后沉降并去除过量油的这一过程,直至珠子相对于连续相油悬浮在给定纯度的水性缓冲液中。水性缓冲液的纯度可以为约80%、90%、95%、96%、97%、98%或99%(v/v)。水性缓冲液的纯度可以超过约80%、90%、95%、96%、97%、98%或99%或更高(v/v)。水性缓冲液的纯度可以低于约80%、90%、95%、96%、97%、98%或99%(v/v)。沉降步骤可以重复约2、3、4或5次。沉降步骤可以重复超过约2、3、4或5次或更多次。沉降步骤可以重复少于约2、3、4或5次。在一些情

况下,沉降和上清液的去除也可以去除未反应的起始材料。

[0169] 小滴发生器的实例可包括单一流动聚焦器、平行流动聚焦器和微筛分膜,如 Nanomi B.V. 使用的那些,及其他。优选地,微流体装置用于产生小滴。

[0170] 图30示出了用于产生凝胶珠子的基于乳液的示例方案,该凝胶珠子用通过二硫键与P5引物连接的acrydite部分预功能化。如图30A所示,将丙烯酰胺、二(丙烯酰)胱胺、acrydite-S-S-P5部分和过硫酸铵组合到乳液的小滴中。可将TEMED添加至乳液油相,并且其可扩散至小滴中以引发聚合反应。如图30A所示,TEMED作用于过硫酸铵导致SO₄⁻自由基的产生,该自由基然后可通过在碳-碳双键的一个碳处产生自由基来激活丙烯酰胺的碳-碳双键。

[0171] 如图30B所示,活化的丙烯酰胺可与未活化的丙烯酰胺反应(再次在其碳-碳双键处),从而开始聚合。所产生的每个产物均可通过自由基的形成而再次被激活,从而导致聚合物延伸(propagation)。此外,二(丙烯酰)胱胺交联剂和acrydite-S-S-P5部分均包含可与活化的物质反应的碳-碳双键,并且产物本身随后可变为活化的。将二(丙烯酰)胱胺交联剂包含在聚合反应中可导致如图30C所示产生的聚合物链的交联。因此,如图30C所示,可以产生包含连接至聚合物骨架的acrydite-S-S-P5部分的水凝胶聚合物网络。聚合反应可持续直至其终止。当反应终止时,可使用连续相交换或其他合适的方法破坏乳液,并获得包含与acrydite-S-S-P5部分偶联的交联水凝胶的凝胶珠子(在图30A中图示)。条形码和随机N-mer(介绍)

[0172] 某些应用,例如多核苷酸测序,可依赖于独特标识符(“条形码”)来鉴定序列以及例如由经测序的片段组装更大的序列。因此,在测序前将条形码添加至多核苷酸片段可能是期望的。在核酸应用的情况下,这类条形码通常由附接至样品序列的核苷酸的相对较短的序列组成,其中条形码序列是已知的或可通过其位置或序列元件鉴定的。在一些情况下,独特标识符对样品索引化可以是有用的。然而,在一些情况下,条形码在其他的情况下也可以是有用的。例如,条形码可用于在整个处理过程中追踪样品(例如,样品在实验室中的位置、样品在多个反应容器中的位置等);提供制造信息;追踪条形码随时间(例如,从条形码制造到使用)和在现场的表现;追踪条形码批次在现场随时间的表现;在测序期间提供产物信息以及可能当与产物关联的条形码在测序期间被读取时触发自动化方案(例如,自动化方案在计算机的帮助下开始并执行);追踪并修复有问题的条形码序列或产物批次;充当涉及条形码的反应中的分子触发器,以及其组合。在特别优选的方面,并如上文所提到的,如本文所述的条形码序列可用于提供两个确定的离散核酸序列之间的连接信息。该连接信息可包括,例如,与共同的样品、共同的反应容器(例如,孔或分区)或甚至共同的起始核酸分子的连接。特别地,通过将共同条形码附接至特定样品组分或给定反应体积内样品组分的子集,可以把带有该条形码的所得序列归于该反应体积。反过来,当根据其来源样品、其随后暴露于的处理步骤或在单个分子的基础上将样品分配至该反应体积时,可以更好地将所得序列鉴定为是来源于该反应体积的。

[0173] 可以从不同的形式产生条形码,该形式包括批量合成的多核苷酸条形码,随机合成的条形码序列,基于微阵列的条形码合成,天然核苷酸,与N-mer、随机N-mer、伪随机N-mer部分互补的互补序列,或其组合。条形码的合成在本文中以及在例如于2014年2月7日提交的美国专利申请号14/175,973中描述,其全部公开内容为了所有目的通过引用以全文并

入本文。

[0174] 如上文所述,充当独特标识符的引入条形码序列段的寡核苷酸还可包含附加序列段。这类附加序列段可包括功能序列,诸如引物序列、引物退火位点序列、固定序列或对后续处理有用的其他识别或结合序列,例如用于附接有含条形码的寡核苷酸的样品测序的测序引物或引物结合位点。此外,如本文所用的,提及特定功能序列被包含在含有条形码的序列内还涵盖包含任何这类序列的互补序列,使得互补复制将产生特定的所描述的序列。

[0175] 在一些实例中,条形码或部分条形码可由获自或适用于寡核苷酸阵列如微阵列或珠子阵列的寡核苷酸产生。在这类情况下,可切割微阵列的寡核苷酸(例如,使用将寡核苷酸锚定在阵列上的可切割的联接或部分(诸如光可切割的、化学可切割的或可以其他方式切割的联接)),使得游离的寡核苷酸能充当条形码或部分条形码。在一些情况下,条形码或部分条形码从具有已知序列的阵列获得。例如,已知序列(包括从阵列获得的那些序列)的使用可以有利于避免与未知序列的条形码相关的测序错误。微阵列可以提供至少约10,000,000、至少约1,000,000、至少约900,000、至少约800,000、至少约700,000、至少约600,000、至少约500,000、至少约400,000、至少约300,000、至少约200,000、至少约100,000、至少约50,000、至少约10,000、至少约1,000、至少约100或至少约10种可用作条形码或部分条形码的不同序列。

[0176] 本文提供的珠子可以附接至可作为独特标识符(例如,条形码)的寡核苷酸序列。通常,本文提供的珠子群体含有条形码的多样性文库,其中每个珠子均附接至单一条形码序列的多个拷贝。在一些情况下,条形码序列是预先合成的和/或用已知序列设计的。在一些情况下,文库中的每个珠子均附接至独特的条形码序列。在一些情况下,多个珠子将附接有相同的条形码序列。例如,在一些情况下,文库中约1%、2%、3%、4%、5%、10%、20%、25%、30%、50%、75%、80%、90%、95%或100%的珠子附接至与附接至该文库中不同珠子的条形码序列相同的条形码序列。有时,约1%、2%、3%、4%、5%、10%、20%、25%或30%的珠子附接至相同的条形码序列。

[0177] 条形码序列的长度可以是任何合适的长度,这取决于应用。在一些情况下,条形码序列的长度可以为约2到约500个核苷酸、约2到约100个核苷酸、约2到约50个核苷酸、约2到约20个核苷酸、约6到约20个核苷酸或约4到16个核苷酸。在一些情况下,条形码序列的长度为约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、85、90、95、100、150、200、250、300、400或500个核苷酸。在一些情况下,条形码序列的长度大于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、85、90、95、100、150、200、250、300、400、500、750、1000、5000或10000个核苷酸。在一些情况下,条形码序列的长度小于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、85、90、95、100、150、200、250、300、400、500、750或1000个核苷酸。

[0178] 可将条形码负载到珠子中,使得一种或多种条形码被引入到特定的珠子中。在一些情况下,每个珠子可含有同一组条形码。在其他情况下,每个珠子可含有不同组的条形码。在其他情况下,每个珠子可包含一组相同的条形码。在其他情况下,每个珠子可包含一组不同的条形码。

[0179] 本文提供的珠子可以附接至寡核苷酸序列,该寡核苷酸是能够在下游过程中引发

样品(例如,基因组样品)的随机的、伪随机的或靶向的N-mer。在一些情况下,在附接至单个珠子或珠子群体的寡核苷酸上将存在相同的n-mer序列。这可能是靶向引发方法的情况,例如,当选择引物以靶向较大靶序列内的某些序列区段时。在其他情况下,本文珠子群体内的每个珠子均附接至大量多样化的N-mer序列,从而使这些引物针对模板分子的采样多样化等,藉此这类随机n-mer序列将针对样品核酸的不同部分随机引发。

[0180] N-mer的长度可以变化。在一些情况下,N-mer(例如,随机N-mer、伪随机N-mer或靶向N-mer)的长度可以在约2到约100个核苷酸之间、约2到约50个核苷酸之间、约2到约20个核苷酸之间、约5到约25个核苷酸之间或约5到约15个核苷酸之间。在一些情况下,N-mer(例如,随机N-mer、伪随机N-mer或靶向N-mer)的长度可以为约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、85、90、95、100、150、200、250、300、400或500个核苷酸。在一些情况下,N-mer(例如,随机N-mer、伪随机N-mer或靶向N-mer)的长度可以大于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、85、90、95、100、150、200、250、300、400、500、750、1000、5000或10000个核苷酸。在一些情况下,N-mer(例如,随机N-mer、伪随机N-mer或靶向N-mer)的长度可以小于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、85、90、95、100、150、200、250、300、400、500、750或1000个核苷酸。

[0181] 可以对N-mer(包括随机N-mer)进行工程化以用于引发特定的样品类型。例如,可以针对不同类型的样品核酸或样品核酸的不同区域产生不同长度的N-mer,使得每个N-mer长度与每个不同类型的样品核酸或样品核酸的每个不同区域相对应。例如,可以针对来源于一个物种的基因组(例如,人类基因组)的样品核酸产生一种长度的N-mer,并且可以针对来源于另一物种(例如,酵母基因组)的样品核酸产生另一种长度的N-mer。在另一个实例中,可以针对包含基因组的特定序列区域的样品核酸产生一种长度的N-mer,并且可以针对包含基因组的另一序列区域的样品核酸产生另一种长度的N-mer。此外,除N-mer长度之外或作为其替代,还可对N-mer的碱基组成(例如,N-mer的GC含量)进行工程化,以与样品核酸的特定类型或区域相对应。例如,特定类型的样品核酸或样品核酸的特定区域中的碱基含量可以变化,因此具有不同碱基含量的N-mer可以用于引发不同样品类型的核酸或样品核酸的不同区域。

[0182] 可以用针对特定样品类型或特定样品序列区域工程化的N-mer产生在本文其他各处描述的珠子群体。在一些情况下,可以产生在N-mer长度和含量方面混合的珠子群体(例如,包含针对一种样品类型或序列区域工程化的N-mer的珠子和包含针对另一样品类型或序列区域工程化的另一种N-mer的珠子的混合物)。在一些情况下,可以产生珠子群体,其中一个或多个珠子可包含针对多种样品类型或序列区域工程化的N-mer的混合群体。

[0183] 如上所述,在一些情况下,随机的或靶向的N-mer均可包含核苷酸类似物、模拟物或非天然核苷酸,以便提供在后续处理步骤中具有改善的性能的引物。例如,在一些情况下,可能期望提供在经历热循环时(例如,在扩增期间)具有不同解链/退火谱的N-mer引物,以便提高n-mer序列的相对引发效率。在一些情况下,可将核苷酸类似物或非天然核苷酸掺入到N-mer引物序列中,以便改变该引物序列相比于包含天然核苷酸的相应引物的解链温度谱。在某些情况下,引物序列,如本文所述的N-mer序列,可在该序列内的一个或多个位置处包含修饰的核苷酸或核苷酸类似物,例如,LNA碱基,以便在与模板序列杂交时为引物提

供提高的温度稳定性,以及提供通常增强的双链体稳定性。在一些情况下,在引物合成中使用LNA核苷酸代替A或T碱基,以便用更紧密结合的LNA类似物替换那些较弱结合的碱基。通过提供增强的杂交引物序列,可采用这类引物产生更高效率的扩增过程,同时能够在不同的温度模式中操作。

[0184] 还可为上述寡核苷酸提供其他修饰。例如,在一些情况下,寡核苷酸可以具有受保护的末端或其他区域,以便防止或减少寡核苷酸例如通过任何存在的核酸外切酶活性的任何降解。在一个实例中,寡核苷酸可以在寡核苷酸序列内的一个或多个位置(例如,邻近或接近3'和/或5'末端位置)处具有一种或多种硫代磷酸核苷酸类似物。这些硫代磷酸核苷酸通常在寡核苷酸内的核苷酸间联接中提供硫基团以代替非连接氧,从而减少或消除核酸酶对寡核苷酸的活性,包括,例如,3'-5'和/或5'-3'核酸外切酶。通常,硫代磷酸类似物可用于为包含它们的寡核苷酸提供外切和/或内切核酸酶抗性,包括针对例如寡核苷酸的3'-5'和/或5'-3'核酸外切酶消化提供保护。因此,在一些方面,这些一个或多个硫代磷酸酯联接将处于寡核苷酸3'或5'末端的最后5到10个核苷酸间联接的一个或多个中,并且优选地包括最后的3'或5'末端核苷酸间联接和倒数第二个5'末端核苷酸间联接中的一个或多个,以便针对3'-5'或5'-3'核酸外切酶活性提供保护。还可在寡核苷酸内的其他位置提供硫代磷酸酯联接。除了对包含条形码序列(和任何关联的功能序列)的寡核苷酸提供这类保护之外,上述修饰在本文所述的阻断序列的情况下也是有用的,例如,在阻断序列中(例如,邻近或接近3'和/或5'末端位置以及可能寡核苷酸内的其他位置)引入硫代磷酸类似物。

[0185] 将内含物附接至预功能化的珠子

[0186] 可将多种内含物附接至本文所述的珠子,包括用寡核苷酸功能化的珠子。通常,附接寡核苷酸,特别是具有所需序列(例如,条形码、随机N-mer)的寡核苷酸。在本文提供的许多方法中,通过引物延伸反应将寡核苷酸附接至珠子。用引物预功能化的珠子可以与寡核苷酸模板接触。然后可进行扩增反应,使得引物得到延伸,从而使寡核苷酸模板的互补序列的拷贝附接至引物。其他附接方法如连接反应也是可能的。

[0187] 在一些情况下,具有不同序列(或相同序列)的寡核苷酸在单独的步骤中附接至珠子。例如,在一些情况下,具有独特序列的条形码附接至珠子,使得每个珠子之上均具有第一条形码序列的多个拷贝。在第二步中,可用第二序列将珠子进一步功能化。第一和第二系列的组合可充当附接至珠子的独特条形码或独特标识符。可继续该过程以添加起条形码序列作用的附加序列(在一些情况下,向每个珠子连续添加超过1、2、3、4、5、6、7、8、9或10个条形码序列)。还可进一步用随机N-mer将珠子功能化,该随机N-mer可例如充当下游全基因组扩增反应的随机引物。

[0188] 在一些情况下,在用某些寡核苷酸序列(例如,条形码序列)功能化后,可将珠子合并,然后与大量的随机N-mer接触,这些N-mer随后附接至珠子。在一些情况下,特别是当在随机N-mer附接之前合并珠子时,每个珠子均具有一个与之附接的条形码序列(经常为多个拷贝),但具有许多不同的与之附接的随机N-mer序列。图4提供了一种示例方法的分步图示,该方法是用于将寡核苷酸如条形码或N-mer附接至珠子的示例有限稀释法。

[0189] 有限稀释可用于将寡核苷酸附接至珠子,使得珠子平均附接至不超过一个独特寡核苷酸序列如条形码。通常,在该过程中珠子已经用某种寡核苷酸如引物功能化。例如,可将用引物(例如,诸如通用引物)功能化的珠子和多种模板寡核苷酸混合(通常以高的珠子:

模板寡核苷酸比),以产生珠子和模板寡核苷酸的混合物。然后,可通过例如批量乳化过程(平板内的乳液)或通过微流体装置,例如,微流体小滴发生器,将混合物分配到多个分区(例如,油包水乳液内的水性小滴)中。在一些情况下,可将混合物分配到多个分区中,使得平均每个分区包含不超过一个模板寡核苷酸。

[0190] 此外,模板寡核苷酸可通过附接至珠子的引物在分区扩增(例如,经由引物延伸反应)。扩增可导致产生包含扩增的模板寡核苷酸的珠子。扩增后,可将分区的内含物合并到共同的容器(例如,管、孔等)中。然后,可通过任何合适的方法,包括,例如,离心和磁性分离,在有或没有本文其他各处所述的捕获部分的帮助下,将包含扩增的模板寡核苷酸的珠子与分区的其他内含物(包括不包含扩增的模板寡核苷酸的珠子)分离。

[0191] 可将包含扩增的模板寡核苷酸的珠子与附加模板寡核苷酸组合,以产生包含珠子和附加模板寡核苷酸的批量混合物(bulk mixture)。附加模板寡核苷酸可包含与珠子上扩增的模板寡核苷酸至少部分互补的序列,使得附加模板寡核苷酸与扩增的模板寡核苷酸杂交。然后,可在扩增反应中通过杂交的附加模板寡核苷酸将扩增的模板寡核苷酸延伸,使得附加模板寡核苷酸的互补序列附接至扩增的模板寡核苷酸。对于任意所需数目的、将被添加至珠子的附加寡核苷酸,可以重复进行附加模板寡核苷酸与扩增的寡核苷酸结合、随后在扩增反应中使扩增的寡核苷酸延伸的循环。

[0192] 附接至扩增的模板寡核苷酸的寡核苷酸可包含,例如,随机N-mer序列、伪随机N-mer序列或引物结合位点(例如,通用序列部分,诸如与测序装置兼容的通用引物部分)中的一种或多种。如本文其他各处所述,任何这些序列或附接至珠子的任何其他序列可至少包含含有尿嘧啶的核苷酸的片段。

[0193] 用于将条形码序列和随机N-mer附接至珠子的有限稀释法的实例在图4中示出。如图4A所示,珠子401(例如,二硫键交联的聚丙烯酰胺凝胶珠子)用第一引物403预功能化。例如,第一引物403可使用结合至珠子401表面的acrydite部分通过二硫键402偶联至珠子。然而,在一些情况下,第一引物403可通过acrydite部分偶联至珠子,而无需二硫键402。第一引物403可以是用于引发将要附接至珠子的寡核苷酸的模板序列的通用引物,并且/或者可以是用于对包含第一引物403的寡核苷酸进行测序的引物结合位点(例如,P5)。

[0194] 然后,如图4B所示,可将第一引物403功能化的珠子401在水溶液中与模板寡核苷酸(例如,包含第一引物结合位点404(例如,P5c)、模板条形码序列405以及模板引物结合位点407(例如,R1c)的寡核苷酸)以及核酸扩增所需的试剂(例如,dNTP、聚合酶、辅因子等)混合。该水性混合物还可包含与捕获部分(例如,生物素)连接的捕获引物406(例如,有时称为读取引物),该捕获引物与模板寡核苷酸的模板引物结合位点407在序列上相同。

[0195] 然后,在水/油乳液中将水性混合物乳化,以在连续油相中产生水性小滴(例如,包含一个或多个珠子401、模板寡核苷酸、核酸扩增所需的试剂以及(需要时)任何捕获引物406的小滴)。通常,每个小滴平均包含至多一个模板寡核苷酸。如图4B和4C所示,小滴的第一轮热循环导致模板寡核苷酸在引物结合位点404处被第一引物403和第二引物403的延伸引发,使得与模板寡核苷酸序列互补的寡核苷酸在第一引物403处附接至珠子。互补的寡核苷酸包含第一引物403、条形码序列408(例如,与模板条形码序列405互补)以及与模板引物结合位点407和捕获引物406均互补的捕获引物结合位点415。在互补寡核苷酸的测序期间,捕获引物结合位点415还可用作读取引物结合位点(例如,R1)。

[0196] 如图4D所示，捕获引物406可在下一轮热循环期间与捕获引物结合位点415结合。然后，如图4E所示，在其5'端包含捕获部分(例如，生物素)的捕获引物406可被延伸以产生附加模板寡核苷酸(例如，包含序列404、405和406)。热循环可持续所需次数的循环(例如，至少约1、5、10、15、20、25、30、35、40、45、50个或更多个循环)，直至珠子401的所有第一引物403位点均与条形码序列408和捕获引物结合位点415连接。因为每个小滴通常在开始时包含一个或零个模板寡核苷酸，所以每个小滴将通常包含附接至与模板寡核苷酸互补的序列的多个拷贝的珠子或没有附接与模板寡核苷酸互补的序列的拷贝的珠子。如图4E所示，在热循环结束时，附接至珠子的寡核苷酸产物与同样包含捕获部分(例如，生物素)的模板寡核苷酸杂交。

[0197] 然后可通过任何合适的手段破坏乳液，并且可将释放的珠子合并到共同的容器中。如图4F和图4G所示，使用与能够和捕获引物406的捕获部分结合的部分(例如，链霉亲和素)连接的捕获珠子(或其他装置，包括本文所述的捕获装置)409，可通过捕获珠子与捕获部分的相互作用从阴性珠子(例如，不包含序列403、408和415的珠子)中富集阳性珠子(例如，包含序列403、408和415的珠子)。在使用捕获珠子的情况下，珠子可以是磁性的，使得可以使用磁体进行富集。作为替代，可以使用离心进行富集。如图4H所示，在富集阳性珠子时，可通过热或化学手段，包括本文所述的化学手段，使包含捕获部分并连接至捕获珠子的杂交的模板寡核苷酸从与珠子结合的寡核苷酸变性。然后，可通过附接至变性的寡核苷酸的捕获珠子将变性的寡核苷酸(例如，包含序列404、405和406的寡核苷酸)与阳性珠子分离。如图4H所示，获得了包含序列403、408和415的珠子。作为捕获珠子的替代，还可通过在分区中包含例如能够与珠子或偶联至珠子的物质结合的光学活性染料，经由流式细胞术从阳性珠子中分选阳性珠子。

[0198] 如图4I所示，在批量水性流体中，然后可将包含序列403、408和415的珠子与模板随机序列(例如，随机N-mer)413组合，该模板随机序列均连接至与捕获引物结合位点415互补的序列412。如图4J所示，捕获引物结合位点415可在加热时在序列412处引发包含模板随机序列413的寡核苷酸。引发后，捕获引物结合位点415可被延伸(例如，通过聚合酶)以将捕获引物结合位点415与随机序列414连接，该随机序列414与模板随机序列413互补。可采用热或化学手段，包括本文所述的化学手段，将包含模板随机序列413和序列412的寡核苷酸从珠子变性。例如，可以采用离心和洗涤珠子将珠子与变性的寡核苷酸分离。如图4K、4L和4M所示，在去除变性的寡核苷酸后，获得了包含条形码序列408和随机序列414的珠子。因为随机序列414的附接批量完成，所以包含独特条形码序列408的多个拷贝的每个珠子还包含不同的随机序列414。

[0199] 为了从珠子释放与珠子结合的寡核苷酸，可以使用本文其他各处所述的刺激，例如，还原剂。如图4N所示，包含二硫键和通过二硫键与寡核苷酸的联接的珠子与还原剂的接触使珠子和二硫键均降解，从而从珠子释放寡核苷酸。与还原剂的接触可在例如另一分区(例如，另一乳液的小滴)中完成，使得当寡核苷酸从珠子释放时，每个小滴通常包含游离的寡核苷酸，该游离的寡核苷酸全都包含相同的条形码序列408，以及不同的随机序列414。通过充当随机引物的随机序列414，可使用游离的寡核苷酸对同样在该分区中的样品核酸的不同区域进行条形码化。扩增或连接方案(包括本文所述的那些)可用于完成条形码与样品核酸的附接。

[0200] 通过有限稀释,每个分区(例如,小滴)可平均含有至多一个寡核苷酸。在给定序列-珠子稀释下的分布频率服从泊松分布。因此,在一些情况下,约6%、10%、18%、20%、30%、36%、40%或50%的小滴或分区可包含一个或更少的寡核苷酸序列。在一些情况下,超过约6%、10%、18%、20%、30%、36%、40%、50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更多的小滴可包含一个或更少的寡核苷酸序列。在其他情况下,少于约6%、10%、18%、20%、30%、36%、40%或50%的小滴可包含一个或更少的寡核苷酸序列。

[0201] 在一些情况下,可在随机N-mer序列添加之前重复有限稀释步骤,以便增加具有条形码拷贝的阳性珠子的数目。例如,可以准备有限稀释,使得所需分数(例如,1/10到1/3)的乳液小滴包含用于扩增的模板。可通过模板的扩增产生阳性珠子(如图4所示),使得阳性珠子通常不再包含用于扩增的引物(例如,所有的P5引物均已延伸)。然后可破坏乳液小滴,并随后在有限稀释下用新鲜的模板再乳化,以用于第二轮扩增。因为其引发位点将已经被占据,所以在第一轮扩增中产生的阳性珠子通常将不参与进一步的扩增。扩增随后再乳化的过程可重复合适的步骤次数,直至获得所需分数的阳性珠子。

[0202] 在一些情况下,可以回收在有限稀释功能化后在分选期间获得的阴性珠子并进一步处理以产生另外的阳性珠子。例如,可在回收后将阴性珠子分布到平板(例如,384孔板)的孔中,使得每个孔通常包含1个珠子。在一些情况下,可借助于流式细胞术(例如,在分选期间流式细胞仪引导每个阴性珠子进入孔中——一种示例流式细胞仪为BD FACS Jazz)或通过分布装置(例如,自动分布装置)完成分布。每个孔还可包含模板条形码序列,并且除了每个孔分配每个珠子而不是流体小滴之外,重复图4中示出的过程。因为每个孔均包含模板和珠子,所以每个孔均可产生阳性珠子。然后可将珠子从每个孔合并,并且可如本文其他各处所述批量添加附加序列(例如,随机N-mer序列)。

[0203] 可以以预期或预测的条形码/待条形码化的珠子的比例,将条形码负载到珠子中。在一些情况下,负载条形码,使得每个珠子负载约0.0001、0.001、0.1、1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000、20000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000、10000000000、50000000000或100000000000个条形码的比例。在一些情况下,负载条形码,使得每个珠子负载超过0.0001、0.001、0.1、1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000、20000、50000、100000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000、200000000、300000000、400000000、500000000、600000000、700000000、800000000、900000000、1000000000、2000000000、3000000000、4000000000、5000000000、6000000000、7000000000、8000000000、9000000000、10000000000个或更多个条形码的比例。在一些情况下,负载条形码,使得每个珠子负载小于约0.0001、0.0002、0.0003、0.0004、0.0005、0.0006、0.0007、0.0008、0.0009、0.001、0.002、0.003、0.004、0.005、0.006、0.007、0.008、0.009、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、

5000、10000、20000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000、10000000000、50000000000或10000000000个条形码的比例。

[0204] 珠子，包括本文所述的那些珠子(例如，基本可溶解的珠子，在一些情况下，基本可被还原剂溶解的)，可与多个寡核苷酸共价或非共价地连接，其中所述寡核苷酸的至少一个子集包含恒定区或结构域(例如，条形码序列、条形码结构域、共同的条形码结构域或在该子集的寡核苷酸之间恒定的其他序列)和可变区或结构域(例如，随机序列、随机N-mer或在该子集的寡核苷酸之间可变的其他序列)。在一些情况下，如本文其他各处所述，寡核苷酸可以可释放地偶联至珠子。寡核苷酸可通过任何合适的联接(包括本文其他各处所述的共价和非共价联接的类型)共价或非共价地连接至珠子。在一些情况下，寡核苷酸可通过可切割的联接例如化学可切割的联接(例如，二硫键)、光可切割的联接或热可切割的联接共价连接至珠子。珠子可包含超过约或至少约1、10、50、100、500、1000、5000、10000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000、10000000000、50000000000或100000000000个寡核苷酸，该寡核苷酸包含恒定区或结构域以及可变区或结构域。

[0205] 在一些情况下，所述寡核苷酸可各自包含相同的恒定区或结构域(例如，相同的条形码序列、相同的条形码结构域、共同的结构域等)。在一些情况下，所述寡核苷酸可各自包含具有不同序列的可变结构域。在一些情况下，包含相同恒定区(或共同的结构域)的寡核苷酸的百分比可以至少为约0.01%、0.1%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。在一些情况下，包含具有不同序列的可变区的寡核苷酸的百分比可以至少为约0.01%、0.1%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。在一些情况下，多个珠子中包含具有不同核苷酸序列的寡核苷酸(包括包含可变和恒定区或结构域的那些)的珠子的百分比至少为约0.01%、0.1%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。在一些情况下，所述寡核苷酸还可包含一个或多个附加序列，例如，引物结合位点(例如，测序引物结合位点)、通用引物序列(例如，预期将与具有特定长度的任意核酸片段上的一个或多个基因座杂交并将其引发的引物序列，基于具有该长度的序列内存在这类基因座的可能性)或任何其他期望的序列，包括本文其他各处所述的附加序列的类型。

[0206] 如本文其他各处所述，可产生多个珠子以形成例如珠子文库(例如，条形码化珠子文库)。在一些情况下，在所述多个珠子中的至少一个子集的单独珠子之间，共同结构域(例如，共同的条形码结构域)或区域的序列可以不同。例如，在多个珠子中的约2个或更多个、10个或更多个、50或更多个、100个或更多个、500个或更多个、1000个或更多个、5000个或更多个、10000个或更多个、50000个或更多个、100000个或更多个、500000个或更多个、1000000个或更多个、5000000个或更多个、10000000个或更多个、50000000个或更多个、100000000个或更多个、500000000个或更多个、1000000000个或更多个、5000000000个或更多个或者10000000000个或更多个珠子之间，所述多个珠子中的单独珠子之间的共同结构域或区域的序列可以不同。在一些情

况下，多个珠子中的每个珠子可包含不同的共同结构域或区域。在一些情况下，多个珠子中包含不同的共同结构域或区域的单独珠子的百分比可以至少为约0.01%、0.1%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%。在一些情况下，多个珠子可包含至少约2、10、50、100、500、1000、5000、10000、50000、100000、500000、1000000、5000000、10000000、50000000个或更多个不同的共同结构域，该结构域偶联至所述多个珠子中的不同珠子。

[0207] 作为有限稀释的替代(例如，通过乳液的小滴)，可采用其他分配方法将寡核苷酸附接至珠子。如图13A所示，可使用平板的孔。可在平板的孔中将包含引物(例如，通过acrydite和任选的二硫键连接至珠子的P5引物)的珠子与模板寡核苷酸(例如，包含条形码序列的模板寡核苷酸)以及扩增试剂组合。每个孔可包含独特模板条形码序列的一个或多个拷贝和一个或多个珠子。平板的热循环通过模板寡核苷酸与引物的杂交而使引物延伸，使得珠子包含具有与寡核苷酸模板互补的序列的寡核苷酸。热循环可持续期望次数的循环(例如，至少约1、2、5、10、15、20、25、30、35、40、45、50个或更多个循环)，直至所有的引物均已经被延伸。

[0208] 热循环完成时，可将珠子合并到共同的容器中，洗涤(例如，通过离心、磁性分离等)，将互补链变性，再次洗涤，并且如果需要然后使其经历额外轮次的批量处理。例如，可以采用上文针对有限稀释所述的并且如图13B和图4I-M所示的引物延伸方法，将随机N-mer序列添加至与珠子结合的寡核苷酸。

[0209] 作为有限稀释的另一个替代方法，如图13C所示，可使用包括在多孔板中分配的组合过程来产生具有寡核苷酸序列的珠子。在这样的方法中，孔可含有预合成的寡核苷酸，如寡核苷酸模板。可将珠子(例如，具有预引入的寡核苷酸如引物的珠子)分到多孔板的单个孔中。例如，可将含有P5寡核苷酸的珠子的混合物分到多孔板(例如，384个孔)的单个孔中，其中每个孔均含有独特的寡核苷酸模板(例如，包含第一部分条形码模板或条形码模板的寡核苷酸)。例如，可使用寡核苷酸模板作为模板和附接至珠子的引物作为引物，在单个孔内进行引物延伸反应。随后，可以将所有的孔合并在一起并可以去除未反应的产物。

[0210] 可以将与扩增产物附接的珠子的混合物再分到第二多孔板(例如，384孔板)的孔中，其中第二多孔板的每个孔均包含另一种寡核苷酸序列(例如，包括第二部分条形码序列和/或随机N-mer)。在一些情况下，该寡核苷酸序列可以附接(例如，通过杂交)至阻断寡核苷酸。在第二多孔板的孔内，可进行反应如单链连接反应，以向每个珠子添加附加序列(例如，通过如在第一步中附接至珠子的引物延伸产物与第二步的孔中的寡核苷酸连接)。在一些情况下，在第一步中连接至珠子的部分条形码序列与第二步中的第二部分条形码序列相连接，以产生包含完整条形码序列的珠子。在一些情况下，包含完整条形码序列的珠子还包含随机序列(例如，随机N-mer)和/或阻断寡核苷酸。在一些情况下，进行PCR反应或引物延伸反应以将附加序列附接至珠子。可将来自孔的珠子合并在一起，并去除未反应的产物。在一些情况下，使用另外的多孔板来重复该过程。该过程可重复超过1、2、3、4、5、6、7、8、9、10、15、20、50、100、500、1000、5000或10000次。

[0211] 在一些组合方法中，可使用连接方法在珠子(例如，本文其他各处所述的可降解的珠子)上组装包含条形码序列的寡核苷酸序列。例如，可以提供单独的珠子群体，含有条形码的寡核苷酸将与之附接。这些群体可包含用于附接核苷酸的锚定组分(或联接)，诸如可

激活的化学基团(亚磷酰胺、acrydite部分或其他热、光或化学可激活的基团)、可切割的连接、先前附接的寡核苷酸分子(含有条形码的寡核苷酸可与之连接、杂交或以其他方式附接)、DNA结合蛋白、用于静电附接的带电荷基团,或多种其他附接机制中的任一种。

[0212] 第一寡核苷酸或包含第一条形码序列段的寡核苷酸区段附接至单独的群体,其中不同的群体包含与其附接的不同条形码序列段。每个单独群体中的每个珠子可附接于至少2、10、50、100、500、1000、5000、10000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000个或更多个第一寡核苷酸分子或寡核苷酸区段分子。第一寡核苷酸或寡核苷酸区段可以可释放地附接至单独的群体。在一些情况下,第一寡核苷酸或寡核苷酸区段可直接附接至单独群体中的相应珠子,或可间接附接(例如,如上所述,通过偶联至珠子的锚定组分)至单独群体中的相应珠子。

[0213] 在一些情况下,第一寡核苷酸可在夹板(夹板的一个实例在图23A中以2306示出)的帮助下附接至单独的群体。如本文所用的,夹板通常指双链的核酸,其中核酸的一条链包含将与一个或多个接收寡核苷酸附接的寡核苷酸,并且其中核酸的另一条链包含具有以下序列的寡核苷酸,该序列与将要附接的寡核苷酸的至少一部分部分互补并且与一个或多个接收寡核苷酸的至少一部分部分互补。在一些情况下,寡核苷酸可以通过如图23A所示的突出端序列与接收寡核苷酸的至少一部分部分互补。如本文其他各处所述,突出端序列可具有任何合适的长度。

[0214] 例如,夹板可被配置为使其包含与以下寡核苷酸杂交的第一寡核苷酸或寡核苷酸区段,该寡核苷酸包含与第一寡核苷酸或寡核苷酸区段的至少一部分部分互补的序列以及与附接至单独群体的寡核苷酸的至少一部分部分互补的序列(例如,突出端序列)。夹板可通过其互补序列与附接至单独群体的寡核苷酸杂交。一旦杂交,然后可通过任何合适的附接机制,例如,连接反应,使夹板的第一寡核苷酸或寡核苷酸区段与附接至单独群体的寡核苷酸附接。

[0215] 在第一寡核苷酸或寡核苷酸区段与单独群体附接之后,然后合并单独群体以创建混合的合并群体,然后将其分离成混合的、合并的群体的多个单独群体。然后将包含第二条形码序列段的第二寡核苷酸或区段附接至每个单独混合、合并的群体中珠子上的第一寡核苷酸,使得不同的混合合并的珠子群体附接有不同的第二条形码序列段。混合、合并的群体的单独群体中的每个珠子可附接至至少2、10、50、100、500、1000、5000、10000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000个或更多个第二寡核苷酸分子或寡核苷酸区段分子。

[0216] 在一些情况下,第二寡核苷酸可在夹板的帮助下附接至第一寡核苷酸。例如,用于在产生混合合并的群体之前将第一寡核苷酸或寡核苷酸区段附接至单独群体的夹板还可包含与第二寡核苷酸的至少一部分部分互补的序列(例如,突出端序列)。夹板可通过互补序列与第二寡核苷酸杂交。一旦杂交,然后可通过任何合适的附接机制,例如,连接反应,使第二寡核苷酸附接至第一寡核苷酸。然后,可通过进一步的处理将与第一和第二寡核苷酸都互补的夹板链变性(或去除)。或者,可提供包含第二寡核苷酸的单独夹板,以将第二寡核苷酸附接至第一寡核苷酸,其附接方式类似于以上对于在夹板的帮助下将第一寡核苷酸与附接至单独群体的寡核苷酸进行附接所述的方式。另外,在一些情况下,可通过如本文其他各处所述的连接序列将第一寡核苷酸的第一条形码区段和第二寡核苷酸的第二条形码区

段接合。

[0217] 然后可将混合的、合并的群体的单独群体合并,且所得的合并的珠子群体包括条形码序列的多样性群体,或者由不同第一条形码序列的数目和不同第二条形码序列的数目的乘积所表示的条形码文库。例如,当第一和第二寡核苷酸包括例如全部256个4-mer条形码序列段时,完整的条形码文库可包含65,536种不同的8碱基条形码序列。

[0218] 条形码序列段可独立地选自一组条形码序列段,或者第一或第二条形码序列段可各自选自不同组的条形码序列段。此外,条形码序列段在长度上可单独地且独立地包含2到20个核苷酸、优选约4到约20个核苷酸、更优选约4到约16个核苷酸,或约4到约10个核苷酸。在一些情况下,条形码序列段在长度上可单独地且独立地包含至少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。特别地,条形码序列段可包含2-mer、3-mer、4-mer、5-mer、6-mer、7-mer、8-mer、9-mer、10-mer、11-mer、12-mer、13-mer、14-mer、15-mer、16-mer、17-mer、18-mer、19-mer、20-mer或更长的序列段。

[0219] 此外,包含在第一和第二寡核苷酸序列或序列段内的条形码序列段通常将表示至少10个不同的条形码序列段、至少50个不同的条形码序列段、至少100个不同的条形码序列段、至少500个不同的条形码序列段、至少1,000个不同的条形码序列段、至少约2,000个不同的条形码序列段、至少约4,000个不同的条形码序列段、至少约5,000个不同的条形码序列段、至少约10,000个不同的条形码序列段、至少50,000个不同的条形码序列段、至少100,000条形码序列段、至少500,000个条形码序列段、至少1,000,000个条形码序列段或更多。依照上述过程,这些不同的寡核苷酸可在第一或第二寡核苷酸添加步骤中分配到相似或相同数目的单独的珠子群体中,例如,至少10、100、500、1000、2000、4000、5000、10000、50000、100000、500000、1000000个等不同的条形码序列段被分别添加到至少10、100、500、1000、2000、4000、5000、10000、50000、100000、1000000个等单独的珠子群体。

[0220] 因此,在文库中表示的所得条形码文库在多样性上可以在至少约100个不同的条形码序列段到至少约1,000,000、2,000,000、5,000,000、10,000,000、100,000,000个或更多个不同的如本文其他各处所述的条形码序列段的范围内。

[0221] 如上所述,第一和第二寡核苷酸序列或序列段中的任一个或两个,或者随后添加的寡核苷酸(例如,第三寡核苷酸添加至第二寡核苷酸、第四寡核苷酸添加至第三寡核苷酸等),可包含附加序列,例如,完全或部分的功能序列(例如,引物序列(例如,通用引物序列、靶向引物序列、随机引物序列)、引物退火序列、附接序列、测序引物序列、随机N-mer等),以用于随后的处理。在许多情况下,这些序列在单独的群体、群体子集中的珠子之间和/或在总群体中的所有珠子之间将是共同的。在一些情况下,功能序列在不同的珠子亚群、不同的珠子或甚至附接至单个珠子的不同分子之间可以是可变的。此外,如本文其他各处所述,第一和第二寡核苷酸序列或序列段中的任一个或两个可包含序列段,该序列段包含一个或多个含尿嘧啶的核苷酸和非天然的核苷酸。另外,尽管描述为包含条形码序列的寡核苷酸,但应当理解,这样提到的包括由两个、三个或更多离散的条形码序列段组成的寡核苷酸,这些离散的条形码序列段在该寡核苷酸内被一个或多个碱基分隔开,例如,第一条形码区段与第二条形码区段在其所在的寡核苷酸中被1、2、3、4、5、6、10个或更多个碱基分隔开。优选地,条形码序列段在其所在的寡核苷酸序列中将彼此相邻或彼此位于6个碱基、4个碱基、3个碱基或2个碱基内。总之,无论是在寡核苷酸序列内为连续的,还是被一个或多个碱基分

隔开,给定寡核苷酸内的全部这些条形码序列段在本文中都被称为条形码序列、条形码序列段或条形码结构域。

[0222] 图23示出了用于产生具有以下序列的珠子的示例组合方法,该序列包含条形码序列以及特定类型的功能序列。尽管为了说明的目的就某些特定的序列段进行了描述,但应当理解,多种不同的配置可并入含有附接至本文所述的珠子的寡核苷酸的条形码中,包括多种不同的功能序列类型、引物类型(例如,对不同测序系统特异的)等等。如图23A所示,可产生珠子2301并使其与第一寡核苷酸组分共价连接(例如,通过acrydite部分或其他物质),该第一寡核苷酸组分将被用作锚定组分和/或功能序列或部分功能序列,例如,部分P5序列2302。在平板(例如,384孔板)的每个孔中,包含剩余的P5序列和独特的第一部分条形码序列(在寡核苷酸2303中以碱基“DDDDDD”表示)的寡核苷酸2303可与寡核苷酸2304杂交,该寡核苷酸2304包含寡核苷酸2303的互补序列和突出于寡核苷酸2303的每一端的其他碱基。可由此产生杂交的产物(“夹板”)2306。夹板的每个突出端可用阻断部分阻断(在图23A中用“X”示出),以防止副产物形成。阻断部分的非限制性实例包括3' 反向dT、双脱氧胞苷(ddC)和3' C3间隔区。因此,在描述的实例中,可产生不同的夹板,每个夹板具有独特的第一部分条形码序列或其互补序列,例如,如所述的,384个不同的夹板。

[0223] 如图23B所示,可将珠子2301加入平板的每个孔中,并且每个孔中的夹板2306可通过寡核苷酸2304的一个突出端与珠子2301的相应锚定序列(例如,部分P5序列2302)杂交。寡核苷酸2304的突出端在杂交部分P5序列2302中的有限稳定性可允许夹板2306的动态采样,这可帮助确保随后寡核苷酸2303与部分P5序列2302的连接是有效的。连接的酶(例如,连接酶)可将部分P5序列2302与寡核苷酸2303连接。连接酶的一个实例将是T4DNA连接酶。连接之后,可将产物合并,并洗涤珠子以去除未连接的寡核苷酸。

[0224] 如图23C所示,然后可将经洗涤的产物再分配到另一平板(例如,384孔板)的孔中,其中该平板的每个孔包含寡核苷酸2305,该寡核苷酸2305具有独特的第二部分条形码序列(在寡核苷酸2305中以“DDDDDD”表示)和与寡核苷酸2304的剩余突出端互补的相邻短序列(例如,与第二部分条形码序列相邻且在寡核苷酸2305的末端的“CC”)。寡核苷酸2305还可包含附加序列,如R1序列和随机N-mer(在寡核苷酸2305中以“NNNNNNNNNN”表示)。在一些情况下,寡核苷酸2305可包含含尿嘧啶的核苷酸。在一些情况下,寡核苷酸2305的任何含胸腺嘧啶的核苷酸可被含尿嘧啶的核苷酸替代。在一些情况下,为了提高包含第二部分条形码序列(例如,序列2305)的寡核苷酸与第一部分条形码序列(例如,序列2303)的连接效率,可提供例如与寡核苷酸2305的全部或一部分互补的双链体链,该链与寡核苷酸2305的一些部分或全部杂交,同时留下可用于与夹板2304杂交的突出端碱基。如上文所述,可提供在其3' 和5' 端中的一端或两端被阻断的夹板2304和/或双链体链,以防止由夹板和双链体链中的一个或两个形成副产物或在其之间形成副产物。在优选的方面,双链体链可与寡核苷酸2305的全部或部分互补。例如,当寡核苷酸2305包含随机n-mer时,可提供不与寡核苷酸的该部分杂交的双链体链。

[0225] 如图23C所示,通过相邻的短序列,寡核苷酸2305可与寡核苷酸2304杂交。此外,突出端在杂交寡核苷酸2305的短互补序列中的有限稳定性可允许寡核苷酸2305的动态采样,这可帮助确保随后寡核苷酸2305与寡核苷酸2303的连接是有效的。然后,连接的酶(例如,连接酶)可将寡核苷酸2305与寡核苷酸2303连接。通过寡核苷酸2305的第一部分条形码序

列与寡核苷酸2303的第二部分条形码序列的接合,寡核苷酸2305与寡核苷酸2303的连接可导致完整的条形码序列的产生。如图23D所示,然后可以合并产物,可将寡核苷酸2304从产物变性,并且然后可洗掉未结合的寡核苷酸。洗涤之后,可获得条形码化珠子的多样性文库,其中每个珠子均与例如包含P5序列、完整条形码序列、R1序列和随机N-mer的寡核苷酸结合。在此实例中,可以获得147,456个独特的条形码序列(例如,384个独特的第一部分条形码序列x384个独特的第二部分条形码序列)。

[0226] 在一些情况下,如上所述包含帮助寡核苷酸连接的突出端碱基可导致全部产物都在给定位置具有相同的碱基,包括如图24A所示在条形码序列的部分之间。根据所采用的特定测序方法,在测序读取结果之间在给定序列位置处没有或只有有限的碱基多样性可导致测序运行失败。因此,在一些方面,可提供在不同的夹板之间具有一定的可变性(就寡核苷酸的全部测序部分内的碱基身份或位置而言)的突出端碱基。例如,在第一个实例中,可将一个或多个间隔区碱基2401(例如,图24B中2401处的“1”“2”)添加到一些用于在珠子上合成较大寡核苷酸的寡核苷酸中,使得寡核苷酸产物彼此之间在长度上略微不同,并由此将突出端碱基定位在不同序列的不同位置处。还可将互补的间隔区碱基添加到序列组分连接所需的夹板中。如图24B所示,产物之间寡核苷酸长度的轻微差别可导致给定读取位置处的碱基多样性。

[0227] 在图25中所示的另一实例中,包含随机碱基突出端的夹板可用于在与夹板突出端互补的读取位置处引入碱基多样性。例如,双链夹板2501可以在一条链上包含随机碱基(例如,图25A中的“NN”)突出端2503和确定的碱基(例如,图25A中的“CTCT”)突出端2506,并且在另一条链上包含第一部分条形码序列(例如,图25A中的“DDDDDD”)。使用如以上针对图23中示出的实例所述的类似连接方案,可利用确定的突出端2506通过杂交来捕获序列2502(其可如图23所示附接至珠子),以用于随后与夹板2501的上方链(如图25A所示)的连接。尽管突出端2506以四个碱基的确定的序列突出端示出,但应当理解,该序列可以更长,以便提高在第一连接步骤中的杂交和连接效率。这样确定的碱基突出端2506在长度上可包括4、6、8、10个或更多个与部分P5序列2502互补的碱基。此外,随机碱基突出端2503可用于捕获最终所需序列的剩余组分(例如,序列2504)。序列2504可包含第二部分条形码序列(图25C的序列2504中的“DDDDDD”)、在一端的随机碱基突出端2503的互补序列2505(例如,图25C中2505处的“NN”)和在其另一端的随机N-mer2507(例如,图25C的序列2504中的“NNNNNNNNNN”)。

[0228] 由于随机碱基突出端2503中的碱基随机性,在互补序列2505处掺入到连接产物中的碱基可以是不同的,使得产物在互补序列2505的读取位置包含各种碱基。应当理解,在优选的方面,将要与第一部分条形码序列连接的第二部分条形码序列部分通常将包括这类第二部分条形码序列的群体,该群体包括随机突出端序列的所有互补序列,例如,给定的部分条形码序列将存在例如16种不同的突出端部分,以便将相同的第二部分条形码序列添加到呈现出多种突出端序列的给定孔中的每个珠子。虽然对于图25中的随机突出端2503和互补序列2505仅示出了两个碱基,但该实例并非意在限制。在突出端中可以使用任何合适数目的随机碱基。此外,尽管描述为随机突出端序列,但在一些情况下,这些突出端序列可以选自突出端序列的子集。例如,在一些情况下,突出端将选自突出端序列的子集,该子集包括比该突出端长度的所有可能的突出端序列更少的突出端序列,这可以不止一个突出端序

列，并且在一些情况下，超过2个、超过4个、超过10个、超过20个、超过50个或甚至更多的突出端序列。

[0229] 在另一实例中，一组夹板—其中每个夹板均具有选自给定长度的一组突出端序列（例如，一组至少2、4、10、20个或者更多个突出端序列）的确定的突出端—可用于在与夹板突出端互补的读取位置处引入碱基多样性。此外，因为这些突出端用于将第二部分条形码序列与第一条形码序列连接，所以在第二部分条形码序列的群体中呈现出所有可能的突出端互补序列是期望的。同样，在许多情况下，保持不同突出端序列的数目较低将是优选的，例如，低于50个、低于20个、低于10个或低于5个不同的突出端序列。在许多情况下，条形码文库中不同连接序列的数目将在2到4096个不同的连接序列之间，优选的文库具有约2到约50个不同的连接序列。同样，通常将期望保留这些具有相对较短长度的突出端序列，以便避免向最终的序列读取结果中引入不相关的碱基。这样，这些突出端序列通常将被设计用于向总的寡核苷酸构建体中引入不超过10个、不超过9个、不超过8个、不超过7个、不超过6个、不超过5个、不超过4个并且在一些情况下3个或更少的核苷酸。在一些情况下，突出端序列的长度可以是约1到约10个核苷酸长、约2到约8个核苷酸长、约2到约6个核苷酸长或约2到约4个核苷酸长。通常，该组中的每个夹板可包含与该组中的其他夹板具有不同序列的突出端，使得突出端的每个位置处的碱基与该组中其他夹板的相同碱基位置处的碱基不同。图26示出了一组夹板的示例。该组包含夹板2601（包含突出端“AC”2602）、夹板2603（包含突出端“CT”2604）、夹板2605（包含突出端“GA”2606）和夹板2607（包含突出端“TG”2608）。每个夹板还可包含突出端2609（例如，每个夹板中的“CTCT”）和第一部分条形码序列（“DDDDDD”）。如图26所示，每个夹板在其独特突出端（例如，夹板2601中的突出端2602、夹板2603中的突出端2604、夹板2605中的突出端2606以及夹板2607中的突出端2608）的每个位置处可包含不同的碱基，使得没有夹板突出端在相同的碱基位置包含相同的碱基。因为每个夹板在其独特突出端的每个位置处包含不同的碱基，所以与另一夹板所产生的产物相比，由每个夹板产生的产物在每个互补位置也可具有不同的碱基。因此，可以实现这些位置处的碱基多样性。

[0230] 这样的产物可以如下生成：将所需序列的第一组分（例如，图25中包含第一部分条形码序列的序列2502；第一组分也可以附接至珠子）与每个夹板所共同的突出端（例如，图26中的突出端2609）杂交；将该序列的第一组分与夹板连接；将所需序列的第二部分（例如，与图25中包含第二部分条形码序列的序列2504相似的序列，不同之处在于该序列在位置2505处包含与独特突出端序列互补的碱基，而不是随机碱基）与夹板的独特突出端杂交；以及将所需序列的第二组分与夹板连接。然后可如上文所述，使夹板的未连接部分（例如，如图26所示的包含突出端的下方序列）变性、对产物进行洗涤，等等，以获得最终产物。应当理解，并且如上文所述，这些突出端序列可在不同的部分条形码序列（或条形码序列段）之间提供1、2、3、4、5、6个或更多个碱基，使得其在条形码序列段之间提供具有上述特征的连接序列。这样的连接序列可具有不同的长度，例如，约2到约10个核苷酸长、约2到约8个核苷酸长、约2到约6个核苷酸长、约2到约5个核苷酸长或约2到约4个核苷酸长。

[0231] 使用图26中示出的这组夹板的示例流程在图27中示出。对于该组中的每个夹板，包含独特突出端序列的夹板链（例如，图26所示的夹板的下方链）可在每一个或多个平板的每个孔中提供。在图27中，对于总共8个平板，为四个夹板类型中的每一个提供包含独特突出

端序列的夹板链的两个96孔板。在八个平板中,两个平板(2601a,2601b)与图26中包含独特突出端序列(“AC”)的夹板2601的下方链对应,两个平板(2603a,2603b)与图26中包含独特突出端序列(“CT”)的夹板2603的下方链对应,两个平板(2605a,2605b)与图26中包含独特突出端序列(“GA”)的夹板2605的下方链对应,两个平板(2607a,2607b)与图26中包含独特突出端序列(“TG”)的夹板2607的下方链对应。每个96孔板(2601a、2601b、2603a、2603b、2605a、2605b、2607a和2607b)中的寡核苷酸可被转移到另一组96孔板2702中,其中每个平板转移到其自己的单独平板中(也是总共八个平板),并且每个平板的每个孔转移到下一个平板中其对应的孔中。

[0232] 可在一个或多个平板中提供包含独特的第一部分条形码序列(例如,图26中示出的夹板的上方链)和第一部分P5序列的夹板链。在图27中,在两个96孔板2708a和2708b中提供这样的夹板链,其中两个平板的每个孔均包含具有独特的第一部分条形码序列的寡核苷酸,两个平板中总共有192个独特的第一部分条形码序列。可将平板2708a的每个孔添加到四个平板2702中其对应的孔中,并且可将平板2708b的每个孔添加到另外四个平板2702中其对应的孔中。因此,每个孔中的两条夹板链可以杂交以产生完整的夹板。在夹板产生后,图27中两个96孔板2702的每个孔均包含配置为图26中的夹板2601、夹板2603、夹板2605或夹板2607的夹板和独特的第一部分条形码序列,总共有192个独特的第一部分条形码序列。

[0233] 然后,可向平板2702的每个孔添加包含第二部分P5序列(例如,与图25中的序列2502相似或等价)的珠子2709。每个孔中的夹板可通过每个夹板的共同突出端序列2609与第二部分P5序列杂交。然后,连接的酶(例如,连接酶)可将第二部分P5序列与包含剩余第一部分P5序列和第一部分条形码序列的夹板链连接。由此产生了包含珠子的第一产物,该珠子与包含P5序列和第一部分条形码序列的序列相连接,且仍然与包含突出端序列的夹板链杂交。连接之后,然后可将来自每个平板的孔的第一产物分别合并以产生平板池2703。与每个双平板组(例如,每组对应于特定的夹板配置)对应的平板池2703也可以单独合并,以产生第一产物池2704,使得每个第一产物池2704包含由仅包含一个独特突出端序列的夹板产生的产物。在图27中,产生了四个第一产物池2704,每一个与该实例中所用的四种夹板类型之一相对应。可洗涤每个平板池2703中的产物以去除未结合的寡核苷酸,可洗涤每个第一产物池2704中的产物以去除未结合的寡核苷酸,或者洗涤可在两个合并步骤中进行。在一些情况下,可以绕过平板池2703,每个双平板组的内含物直接进入第一产物池2704中。

[0234] 下一步,如图27所示,可将每个第一产物池2704等分到两个96孔板2705的每个孔中,总共8个平板(例如,每个产物池2704两个平板)。单独地,可在96孔板2706中提供包含独特的第二部分寡核苷酸、与四种独特突出端序列之一互补的末端序列以及待添加的任何其他序列(例如,附加测序引物位点、随机N-mer等)的寡核苷酸。例如,这样的寡核苷酸可包含与图25中的序列2504相似的序列,不同之处在于该序列在位置2505处包含与独特突出端序列互补的碱基,而不是随机碱基。例如,对于图26中所示的夹板2601,在位置2505处的碱基将是“TG”,其与夹板2601的独特突出端2602(“AC”)互补。在平板2706中,双平板的组可各自包含寡核苷酸,该寡核苷酸包含与四种独特突出端序列之一互补的序列,总共8个平板和四个平板组在图27中示出。平板2706可被配置为使得每个孔均包含独特的第二部分条形码序列,8个平板中总共有768个独特的第二部分条形码序列。

[0235] 如图27所示,根据进入平板2705的平板中的第一产物的合适的独特突出端序列,

平板2706的每个平板可以与平板2705的相应平板配对。可将来自平板2706的平板的每个孔中的寡核苷酸添加至来自平板2705的其相应平板中的相应孔，使得每个孔均包含来自合适的第一产物池2704的第一产物的等份以及来自平板2706的包含独特的第二条形码序列和任何其他序列(例如，随机N-mer)的寡核苷酸。在平板2705的每个孔中，每种第一产物的独特突出端序列可通过与独特突出端序列互补的寡核苷酸的碱基与包含第二部分条形码序列的寡核苷酸杂交。然后，连接的酶(例如，连接酶)可将寡核苷酸与第一产物连接。连接时，通过第一产物的第一部分条形码序列与第二产物的第二部分条形码序列的接合来产生包含完整条形码序列的第二产物。可移取获自平板2705的第二产物除并将其存放到共同的第二产物池2707中。然后可使包含突出端的夹板链(如图26所示)在产物池2707中变性，并且洗涤产物以获得最终产物。可获得总共147,456个独特的条形码序列(例如，192个第一部分条形码序列 \times 768个第二部分条形码序列)，其在与连接过程中所用的独特突出端序列互补的碱基位置中具有碱基多样性。

[0236] 上述关于夹板组的实例，以及组合合成所用的平板数目和类型都并非意在限制。一组夹板可包含任何合适数目的夹板。此外，根据例如所需的独特条形码序列的数目、用于产生条形码序列的碱基的数目等，每组夹板可被设计为具有合适的第一部分条形码序列多样性。

[0237] 使用组合平板方法，可以产生具有高多样性的条形码化珠子的文库。例如，如果使用两个384孔板，每个平板均具有预存放在每个孔中的包含部分条形码序列的寡核苷酸，则可能可以产生 384×384 或147,456个独特的条形码序列。本文所示的组合示例并非意在限制，因为可以使用任何合适的平板组合。例如，尽管在一些情况下，在每个组合步骤中所添加的条形码序列段可选自相同组的条形码序列段。然而，在许多情况下，在每个组合步骤中所添加的条形码序列段可选自部分或完全不同组的寡核苷酸序列。例如，在一些情况下，第一寡核苷酸片段可包括来自第一组条形码序列(例如，4-mer序列)的条形码序列，而第二寡核苷酸序列可包括来自部分或完全不同组的条形码序列段(例如，4-mer序列、6-mer序列、8-mer序列等，或者甚至混合长度的序列，例如，当第二寡核苷酸区段选自包含具有不同长度和序列的条形码序列的一组寡核苷酸时)的条形码序列，以在所产生的条形码中产生多参数可变性，例如，序列和长度。

[0238] 关于上述实例，例如，组合方法的每一步所用的平板(和条形码)的类型和数目不必相同。例如，384孔板可用于第一步而96孔板可用于第二步，总共产生36,864个独特的条形码序列。此外，在每个组合步骤中添加的完整条形码序列的碱基数目不需要相同。例如，在第一组合步骤中，可添加12碱基条形码序列的4个碱基，其余8个碱基在第二组合步骤中添加。此外，用于产生完整条形码序列的组合步骤的数目也可以变化。在一些情况下，使用约2、3、4、5、6、7、8、9或10个组合步骤。

[0239] 可在多孔板中通过标准技术和试剂进行引物延伸反应和连接反应。例如，聚合物聚乙二醇(PEG)可在单链连接反应过程中以约15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%或75%的浓度存在。在一些情况下，PEG可在连接反应过程中以超过约6%、10%、18%、20%、30%、36%、40%、50%或更高的浓度存在。在一些情况下，PEG可在第二平板中的连接反应过程中以低于约6%、10%、18%、20%、30%、36%、40%或50%的浓度存在。

[0240] 本文所提供的方法可降低连接反应中的核苷酸偏倚。当第一孔板中的第一延伸反应可以运行至完成时,可出现更好的结果。对于第二孔板中的单链连接步骤,当仅使用一种类型的寡核苷酸序列时,可以不存在竞争。用于将内含物附接至珠子的孔中分区方法可以避免错误形成具有8N末端的衔接子,特别是当第一孔板中的第一延伸反应运行至完成时。

[0241] 孔中分区过程的可能的变动可包括通过提供具有简并碱基的第二寡核苷酸序列来用PCR替代单链连接步骤,修饰第一寡核苷酸序列以使其比第二寡核苷酸序列更长,和/或在单链连接步骤之后在单独的批量反应中添加随机N-mer序列,因为这可以节约合成成本并可以减少N-mer序列偏倚。

[0242] 在一些情况下,过程的后续序列可用于将条形码序列附接至珠子。该条形码序列可以与合适的PCR试剂和多个珠子在水性流体中混合。该水性流体可在不混溶流体如油内被乳化,以形成乳液。该乳液可产生含有条形码序列、珠子和PCR试剂的单独的流体小滴。可将单独的流体小滴暴露于热循环条件,其中多轮温度循环使得条形码序列能够引发和延伸。可通过本公开内容中其他各处描述的连续相交换将含有流体小滴的乳液破坏。所得的悬浮在水溶液中的条形码化珠子可通过磁性分离或其他分选方法来分选,从而在水性流体中获得一批纯化的条形码化珠子。

[0243] 在一些情况下,过程的后续序列可用于将N-mer序列附加至珠子。N-mer序列可以与合适的PCR试剂和多个合并的条形码化珠子在水性流体中混合。可加热该水性流体以使得N-mer序列能够杂交和延伸。额外的加热可允许去除互补链。

[0244] PCR试剂可包括任何合适的PCR试剂。在一些情况下,在引物延伸或其他扩增反应过程中,dUTP可以替代dTTP,使得寡核苷酸产物包含含有尿嘧啶的核苷酸,而不是含有胸腺嘧啶的核苷酸。通用序列的含尿嘧啶的部分稍后可以与不接受或处理含尿嘧啶的模板的聚合酶一起使用,以减少不想要的扩增产物。

[0245] 扩增试剂可包括通用引物、通用引物结合位点、测序引物、测序引物结合位点、通用读取引物、通用读取结合位点或者与测序装置(例如,Illumina测序仪、Ion Torrent测序仪等)兼容的其他引物。扩增试剂可包括P5、不可切割的5' acrydite-P5、可切割的5' acrydite-SS-P5、R1c、生物素R1c、测序引物、读取引物、P5_Universal、P5_U、52-BioR1-rc、随机N-mer序列、通用读取引物等。在一些情况下,引物可包含修饰的核苷酸、锁核酸(LNA)、LNA核苷酸、含尿嘧啶的核苷酸、含非天然碱基的核苷酸、阻断寡核苷酸、阻断的3'末端和3' ddCTP。图19提供了其他的实例。

[0246] 如本文所述,在一些情况下,分配包含条形码的寡核苷酸,使得每个珠子平均被分配有少于一个独特寡核苷酸序列、少于两个独特寡核苷酸序列、少于三个独特寡核苷酸序列、少于四个独特寡核苷酸序列、少于五个独特寡核苷酸序列或者少于十个独特寡核苷酸序列。因此,在一些情况下,一部分珠子不含有寡核苷酸模板,并因此不可能含有扩增的寡核苷酸。因此,将包含寡核苷酸的珠子与不包含寡核苷酸的珠子分离可能是期望的。在一些情况下,这可以使用捕获部分实现。

[0247] 在一些实施方案中,捕获部分可以与分离方法如磁性分离一起使用,以将包含条形码的珠子与可能不包含条形码的珠子分离。这样,在一些情况下,扩增试剂可包含附接至引物或探针的捕获部分。捕获部分可允许从未标记的珠子中分选标记的珠子,以确认引物和下游扩增产物附接至珠子。示例性的捕获部分包括生物素、链霉亲和素、谷胱甘肽-S-转

移酶(GST)、cMyc、HA等。该捕获部分可以是或包括荧光标记物或磁性标记物。该捕获部分可包含多个分子的捕获部分,例如,多个分子的生物素、链霉亲和素等。在一些情况下,扩增反应可以利用附接至捕获部分(如本文其他各处所述)的捕获引物,使得该引物与扩增产物杂交,并且该捕获部分在扩增反应的额外循环过程中整合到其他扩增的寡核苷酸中。在其他情况下,在扩增反应完成后,包含捕获部分的探针可以与扩增的寡核苷酸杂交,使得该捕获部分与扩增的寡核苷酸相关联。

[0248] 捕获部分可以是结合对的成员,使得在分离期间该捕获部分可以与其结合对结合。例如,可产生包含寡核苷酸的珠子,该寡核苷酸包含为结合对的成员的捕获部分(例如,生物素)。可将珠子与包含结合对的另一成员(例如,链霉亲和素)的捕获珠子混合,使得两个结合对成员在所得的混合物中结合。然后,可采用任何合适的手段,包括,例如离心和磁性分离(例如,包括捕获珠子是磁珠的情况),将珠子-捕获珠子复合物与混合物的其他组分分离。

[0249] 在所述的许多情况下,单独的珠子通常具有与之附接的寡核苷酸,该寡核苷酸具有共同的总条形码序列段。如本文所述,当珠子包含具有共同条形码序列的寡核苷酸时,这通常是指在偶联至给定珠子的寡核苷酸中,显著百分比(例如,超过70%、超过80%、超过90%、超过95%或甚至超过99%)的具有或大于给定长度(例如,包括最终寡核苷酸的预期全长或长度,并且不包括未反应的锚定序列或部分条形码序列)的寡核苷酸包含相同或同样的条形码序列段。该条形码序列段或结构域(再次,其可由被一个或多个碱基分开的两个或更多个序列段组成)可被包含在单个珠子内的其他共同或可变序列或结构域之间。同样如所述的,珠子的总群体将包括具有大量不同条形码序列段的珠子。然而,在许多情况下,在给定珠子群体内的许多单独的珠子可包含相同的条形码序列段。特别地,具有1000、10,000、1,000,000、10,000,000个或更多个不同序列的条形码序列文库可在具有超过100,000、1,000,000、10,000,000、100,000,000、10亿、100亿、1000亿或更多个离散珠子的珠子群体中呈现,使得相同的条形码序列在给定的珠子群体或亚群内多次呈现。例如,相同的条形码序列可在给定的分析中存在于两个或更多个珠子(10个或更多个珠子、100个或更多个珠子等)上。

[0250] 可以添加具有相应的联接如链霉亲和素的捕获装置如磁珠以结合捕获部分,例如,生物素。然后,附接的磁珠可使得能够通过例如磁性分选来分离条形码化的珠子。除链霉亲和素外,磁珠还可包覆有其他连接实体(包括镍-IMAC)以使得能够分离His-标记的融合蛋白,包覆有二氧化钛以使得能够分离磷酸化的肽,或者包覆有胺反应性NHS-酯基团以固定蛋白质或其他配体以供分离。

[0251] 在一些实施方案中,捕获部分可附接至引物、内部序列、扩增产物内的特定序列、条形码序列、通用序列或互补序列。可通过PCR扩增或连接来附接捕获部分。在添加至珠子群体之前,捕获部分可包含附接至特定靶标如引物的通用标签,如生物素。在其他情况下,捕获部分可包含识别可独立地添加至珠子群体的特定序列或蛋白质或抗体的特定标签。在一些实施方案中,捕获部分可以预连接至分选珠子,如磁珠。在一些情况下,捕获部分可以是荧光标记物,其可以使得能够通过荧光活化细胞分选(FACS)来分选。

[0252] 在一些情况下,核酸标记物(例如,荧光标记物)可用于鉴定含有寡核苷酸的流体小滴、乳液或珠子。然后,可进行标记的小滴或珠子的分选(例如,通过流式细胞术),以便分

离附接至扩增的寡核苷酸的珠子。示例性的染色剂包括嵌入染料、小沟结合剂、大沟结合剂、外部结合剂和双嵌入剂。这类染料的具体实例包括SYBR绿、SYBR蓝、DAPI、碘化丙啶、SYBR金、溴化乙锭、碘化丙啶、咪唑(例如、Hoechst 33258、Hoechst 33342、Hoechst 34580和DAPI)、7-AAD、SYTOX Blue、SYTOX Green、SYTOX Orange、POPO-1、POPO-3、YOYO-1、YOYO-3、TOTO-1、TOTO-3、J0J0-1、LOLO-1、BOBO-1、BOBO-3、PO-PRO-1、PO-PRO-3、BO-PRO-1、BO-PRO-3、T0-PRO-1、T0-PRO-3、T0-PRO-5、J0-PRO-1、L0-PRO-1、YO-PRO-1、Y0-PRO-3、PicoGreen、OliGreen、RiboGreen、EvaGreen、SYBR Green、SYBR Green II、SYBR DX、SYTO-40、-41、-42、-43、-44、-45(蓝色)、SYTO-13、-16、-24、-21、-23、-12、-11、-20、-22、-15、-14、-25(绿色)、SYTO-81、-80、-82、-83、-84、-85(橙色)、SYTO-64、-17、-59、-61、-62、-60和-63(红色)。

[0253] 多功能珠子

[0254] 珠子可以与多种物质(包括非核酸物质)连接,使得它们是多功能的。例如,珠子可以与多种类型的包含条形码序列和N-mer(例如,如下文所述的随机N-mer或靶向N-mer)的寡核苷酸连接。每种类型的寡核苷酸在其条形码序列、其N-mer或该寡核苷酸的任何其他序列上可以不同。此外,每个珠子可以与包含条形码序列和N-mer的寡核苷酸连接,并且也可以与能够阻断包含条形码序列和N-mer的寡核苷酸的阻断寡核苷酸连接。寡核苷酸阻断剂和包含条形码序列和N-mer的寡核苷酸的负载可以以不同的比例完成,以便获得寡核苷酸阻断剂与包含条形码序列和N-mer的寡核苷酸的所需化学计量。通常,可将多种物质以不同的比例负载到珠子上,以便获得珠子上物质的所需化学计量。

[0255] 此外,珠子还可以与一种或多种不同类型的多功能寡核苷酸连接。例如,多功能寡核苷酸可以能够起到以下两项或更多项的作用:引物、用于连接的工具、寡核苷酸阻断剂、能够杂交检测的寡核苷酸、报告寡核苷酸、寡核苷酸探针、功能寡核苷酸、富集引物、靶向引物、非特异性引物和荧光探针。起荧光探针作用的寡核苷酸可用于,例如,珠子检测或表征(例如,珠子数目的定量、附接至珠子的物质(例如,引物、连接体等)的定量、珠子大小/拓扑学的确定、珠子孔隙率的确定等)。

[0256] 也可附接或偶联至珠子的物质的其他非限制性实例包括全细胞、染色体、多核苷酸、有机分子、蛋白质、多肽、碳水化合物、糖类、糖、脂质、酶、限制性内切酶、连接酶、聚合酶、条形码、衔接子、小分子、抗体、抗体片段、荧光团、脱氧核苷酸三磷酸(dNTP)、双脱氧核苷酸三磷酸(ddNTP)、缓冲液、酸性溶液、碱性溶液、温度敏感的酶、pH值敏感的酶、光敏感的酶、金属、金属离子、氯化镁、氯化钠、锰、水性缓冲液、温和缓冲液、离子缓冲液、抑制剂、糖类、油、盐、离子、去污剂、离子型去污剂、非离子型去污剂、寡核苷酸、核苷酸、DNA、RNA、肽多核苷酸、互补DNA(cDNA)、双链DNA(dsDNA)、单链DNA(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、scRNA、微RNA、dsRNA、核酶、核糖开关(riboswitch)及病毒RNA、完整或部分的锁核酸(LNA)、锁核酸核苷酸、任何其他类型的核酸类似物、蛋白酶、核酸酶、蛋白酶抑制剂、核酸酶抑制剂、螯合剂、还原剂、氧化剂、探针、生色团、染料、有机物、乳化剂、表面活性剂、稳定剂、聚合物、水、小分子、药物、放射性分子、防腐剂、抗生素、适体及其组合。可通过任何合适的方法将附加寡核苷酸物质和其他类型的物质均偶联至珠子,该方法包括共价和非共价的手段(例如,离子键、范德华相互作用、疏水相互作用、封装、物质向珠子内的扩散等)。在一

些情况下,附加物质可以是用于包含珠子上另一种类型物质的反应的反应物。例如,偶联至珠子的附加物质可以是适用于扩增反应的反应物,该扩增反应包含同样附接至珠子的寡核苷酸。

[0257] 在一些情况下,珠子可包含一个或多个捕获配体,每个捕获配体均能够捕获特定类型的样品组分,包括可能包含核酸的组分。例如,珠子可以包含能够从样品中捕获细胞的捕获配体。例如,该捕获配体可以是针对特定细胞表面上独特的和/或过表达的物质的抗体、抗体片段、受体、蛋白质、肽、小分子或任何其他物质。通过与细胞靶标的相互作用,可以从样品中捕获特定细胞类型,使其保持与珠子结合。与细胞结合的珠子可以进入如本文其他各处所述的分区,以将从细胞中获得的核酸进行条形码化。在一些情况下,从样品中捕获细胞可以在分区中发生。例如,分区中可包含裂解剂,以便从细胞中释放核酸。可采用本文所述的任何方法对释放的核酸进行条形码化并处理。

[0258] III. 条形码文库

[0259] 珠子可含有一个或多个附接的条形码序列。附接至单个珠子的条形码序列可以是相同或不同的。在一些情况下,每个珠子可以附接至约1、5、10、50、100、500、1000、5000、10000、20000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000、10000000000、50000000000或100000000000个相同的条形码序列。在一些情况下,每个珠子可以附接至约1、5、10、50、100、500、1000、5000、10000、20000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000或100000000000个不同的条形码序列。在一些情况下,每个珠子可以附接至至少约1、5、10、50、100、500、1000、5000、10000、20000、50000、100000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000、200000000、300000000、400000000、500000000、600000000、700000000、800000000、900000000、1000000000、2000000000、3000000000、4000000000、5000000000、6000000000、7000000000、8000000000、9000000000、10000000000个或更多个相同的条形码序列。在一些情况下,每个珠子可以附接至至少约1、5、10、50、100、500、1000、5000、10000、20000、50000、100000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000、200000000、300000000、400000000、500000000、600000000、700000000、800000000、900000000、1000000000个或更多个不同的条形码序列。在一些情况下,每个珠子可以附接至少于约1、2、3、4、5、6、7、8、9、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、20000、

30000、40000、50000、60000、70000、80000、90000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000或100000000000个或更多个相同的条形码序列。在一些情况下，每个珠子可以附接至少于约1、2、3、4、5、6、7、8、9、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、20000、30000、40000、50000、60000、70000、80000、90000、100000、500000、1000000、5000000、10000000、50000000或100000000000个或更多个不同的条形码序列。

[0260] 单个条形码文库可包含一个或多个条形码化的珠子。在一些情况下，单个条形码文库可以包含约1、5、10、50、100、500、1000、5000、10000、20000、50000、100000、500000、1000000、5000000、10000000、50000000、100000000、500000000、1000000000、5000000000或100000000000个单独的条形码化珠子。在一些情况下，每个文库可包含至少约1、5、10、50、100、500、1000、5000、10000、20000、50000、100000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000、200000000、300000000、400000000、500000000、600000000、700000000、800000000、900000000、1000000000、2000000000、3000000000、4000000000、5000000000、6000000000、7000000000、8000000000、9000000000、10000000000、20000000000、30000000000、40000000000、50000000000、60000000000、70000000000、80000000000、90000000000、100000000000个或更多个单独的条形码化珠子。在一些情况下，每个文库可包含少于约1、2、3、4、5、6、7、8、9、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、90000、100000、500000、1000000、5000000、10000000、50000000或100000000000个单独的条形码化珠子。文库内的条形码化珠子可具有相同的序列或不同的序列。

[0261] 在一些实施方案中，每个珠子可具有独特的条形码序列。但是，条形码文库内的具有独特条形码序列的珠子的数目可以受到组合限度的限制。例如，使用四种不同的核苷酸，如果条形码的长度为12个核苷酸，那么独特构建体的数目可限制在 $4^{12}=16777216$ 个独特构建体。由于条形码文库可包含多于1677216个珠子，因此可能存在一些具有多个拷贝的相同条形码的文库。在一些实施方案中，给定文库内多个拷贝的相同条形码的百分比可以是1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、40%或50%。在一些情况下，给定文库内多个拷贝的相同条形码的百分比可以大于1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%或50%。

[0262] 在一些实施方案中，每个珠子可包含一个独特的条形码序列，但包含多个不同的随机N-mer。在一些情况下，每个珠子可具有一个或多个不同的随机N-mer。另外，条形码文库内的具有不同随机N-mer的珠子的数目可以受到组合限度的限制。例如，使用四种不同的核苷酸，如果N-mer序列的长度为12个核苷酸，那么不同构建体的数目可限制在 $4^{12}=$

16777216个不同的构建体。由于条形码文库可包含多于16777216个珠子,因此可能存在一些具有多个拷贝的相同N-mer序列的文库。在一些实施方案中,给定文库内多个拷贝的相同N-mer序列的百分比可以是1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、40%或50%。在一些情况下,给定文库内多个拷贝的相同N-mer序列的百分比可以大于1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、40%、50%或更多。在一些情况下,给定文库内多个拷贝的相同N-mer序列的百分比可以小于1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%或50%。

[0263] 在一些实施方案中,对于每个珠子内的每个引物,条形码内的独特标识符序列可以是不同的。在一些情况下,对于每个珠子内的每个引物,条形码序列内的独特标识符序列可以是相同的。

[0264] IV. 将条形码化的珠子与样品组合

[0265] 样品的类型

[0266] 本公开内容的方法、组合物、装置和试剂盒可与任何合适的样品或物质一起使用。例如,样品(例如,样品材料、样品材料的组分、样品材料的片段等)或物质可以是样品处理中所用的任何物质,诸如试剂或分析物。示例性的样品可包括以下的一个或多个:全细胞、染色体、多核苷酸、有机分子、蛋白质、核酸、多肽、碳水化合物、糖类、糖、脂质、酶、限制性内切酶、连接酶、聚合酶、条形码(例如,包括条形码序列、核酸条形码序列、条形码分子)、衔接子、小分子、抗体、荧光团、脱氧核苷酸三磷酸(dNTP)、双脱氧核苷酸三磷酸(ddNTP)、缓冲液、酸性溶液、碱性溶液、温度敏感的酶、pH值敏感的酶、光敏感的酶、金属、金属离子、氯化镁、氯化钠、锰、水性缓冲液、温和缓冲液、离子缓冲液、抑制剂、油、盐、离子、去污剂、离子型去污剂、非离子型去污剂、寡核苷酸、模板核酸分子(例如,模板寡核苷酸、模板核酸序列)、核酸片段、模板核酸片段(例如,片段化过程中由模板核酸的片段化产生的模板核酸的片段、由核酸扩增反应产生的模板核酸的片段)、核苷酸、DNA、RNA、肽多核苷酸、互补DNA(cDNA)、双链DNA(dsDNA)、单链DNA(ssDNA)、质粒DNA、粘粒DNA、染色体DNA、基因组DNA(gDNA)、病毒DNA、细菌DNA、mtDNA(线粒体DNA)、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scaRNA、微RNA、dsRNA、核酶、核酸开关及病毒RNA、蛋白酶、完整或部分的锁核酸、锁核酸核苷酸、核酸酶、蛋白酶抑制剂、核酸酶抑制剂、螯合剂、还原剂、氧化剂、探针、生色团、染料、有机物、乳化剂、表面活性剂、稳定剂、聚合物、水、药物、放射性分子、防腐剂、抗生素、适体等等。总之,所使用的样品将根据特定的处理需求而不同。

[0267] 样品可来自于人类和非人类来源。在一些情况下,样品来自于哺乳动物、非人类哺乳动物、啮齿动物、两栖动物、爬行动物、狗、猫、牛、马、山羊、绵羊、鸡、鸟、鼠、兔、昆虫、蛞蝓、微生物、细菌、寄生虫或鱼。样品可以来自于多种细胞,包括但不限于:真核细胞、原核细胞、真菌细胞、心脏细胞、肺细胞、肾细胞、肝细胞、胰腺细胞、生殖细胞、干细胞、诱导多能干细胞、胃肠细胞、血细胞、癌细胞、细菌细胞、从人类微生物组样品分离的细菌细胞,等等。在一些情况下,样品可包含细胞的内含物,诸如,例如,单个细胞的内含物或多个细胞的内含物。本文所述的方法和系统的单细胞应用的实例在与本申请同日提交的美国临时专利申请号_____ (代理人案号43487-728.101) 中提出。样品还可以是无细胞的,如循环核酸(例如,DNA、RNA)。

[0268] 样品可以是天然存在的或合成的。样品可以获自任何合适的位置，包括生物体、全细胞、细胞制品以及来自任何生物体、组织、细胞或环境的无细胞组合物。样品可以从环境的活检物、抽吸物、福尔马林固定包埋的组织、空气、农业样品、土壤样品、石油样品、水样品或灰尘样品中获得。在一些情况下，样品可以从体液获得，体液可包括血液、尿液、粪便、血清、淋巴液、唾液、粘膜分泌物、汗液、中枢神经系统液体、阴道液或精液。样品还可以从制造的产品如化妆品、食品、个人护理用品等中获得。样品可以是实验操作的产物，该实验操作包括重组克隆、多核苷酸扩增、聚合酶链反应(PCR)扩增、纯化方法(如基因组DNA或RNA的纯化)和合成反应。

[0269] 将条形码附接至样品的方法

[0270] 可通过将两个核酸片段通过酶的作用接合在一起，来将条形码(或其他寡核苷酸，例如，随机N-mer)附接至样品。这可以通过引物延伸、聚合酶链反应(PCR)、使用聚合酶的另一种类型的反应或通过使用连接酶的连接来实现。当使用连接方法将样品附接至条形码时，在连接步骤之前，该样品可以被或可以不被片段化。在一些情况下，将寡核苷酸(例如，条形码、随机N-mer)附接至样品，同时寡核苷酸仍附接至珠子。在一些情况下，在寡核苷酸从珠子释放(例如，通过包含条形码的寡核苷酸从珠子上切割和/或通过珠子的降解)之后，将寡核苷酸(例如，条形码、随机N-mer)附接至样品。

[0271] 寡核苷酸可包含一个或多个随机N-mer序列。一批独特的随机N-mer序列可引发DNA区段的随机部分，由此扩增样品(例如，全基因组)。所得产物可以是代表整个样品(例如，基因组)的一批条形码化的片段。

[0272] 样品在连接至条形码化的珠子之前可以被或可以不被片段化。DNA片段化可包括将DNA链分离或破坏成小片或区段。可采用多种方法对DNA进行片段化，包括限制性消化或产生剪切力的多种方法。限制性消化可利用限制性内切酶，以通过对两条链的平端切割或通过不均匀切割以产生粘端而在DNA序列中制造有意的切口。剪切力介导的DNA链破坏的实例可以包括超声处理、声剪切、针剪切、移液或雾化。超声处理是一种类型的流体动力学剪切，其使DNA序列暴露于短期的剪切力，这可产生约700bp的片段大小。声剪切向碗形换能器内的DNA样品施加高频声能。针剪切通过使DNA穿过小直径的针来产生剪切力，以将DNA物理地撕裂成较小的区段。雾化力可通过使DNA通过喷雾器单元的小孔而产生，在该单元中从离开该单元的微细雾沫中收集所得的DNA片段。

[0273] 在一些情况下，使用连接反应将寡核苷酸与样品连接。该连接可包括通过催化磷酸二酯键的形成将两个核酸区段如条形码序列和样品接合在一起。连接反应可包括DNA连接酶，诸如大肠杆菌DNA连接酶、T4DNA连接酶、哺乳动物连接酶(诸如DNA连接酶I、DNA连接酶III、DNA连接酶IV)、热稳定连接酶等。T4DNA连接酶可以连接含有DNA、寡核苷酸、RNA和RNA-DNA杂合体的区段。连接反应可以不包括DNA连接酶，而是采用替代物如拓扑异构酶。为了将样品与条形码序列连接，采用高浓度的DNA连接酶且包含PEG可实现快速连接。为了选择连接反应的有利温度，可以考虑DNA连接酶的最适温度(其可以是37°C)以及待连接的DNA的解链温度(可以变化)。可将样品和条形码化的珠子悬浮在缓冲液中以使可能影响连接的离子作用最小化。

[0274] 尽管上文就条形码序列与样品核酸组分的连接或直接附接进行了描述，但如本文所用的，条形码与样品核酸的附接还包括条形码序列与样品的互补序列或该互补序列的拷

贝或互补序列的附接,例如,如本文其他各处更详细描述的,当条形码与用于复制样品核酸的引物序列相关联时。特别地,当在使用样品核酸(或样品核酸的复制物)作为模板的引物延伸反应中使用包含引物序列的条形码时,所得的延伸产物,无论是样品核酸的互补序列还是样品核酸的复制物,都将被称为具有与之附接的条形码序列。

[0275] 在一些情况下,将样品与条形码化的珠子组合(手动地或在微流体装置的帮助下),并且将组合的样品和珠子例如在微流体装置中分配。分区可以是油包水乳液内的水性小滴。当样品与条形码化的珠子组合时,在每个流体小滴中平均可存在少于两种目标分析物。在一些实施方案中,每个流体小滴平均可出现少于三种目标分析物。在一些情况下,每个流体小滴平均可出现超过两种目标分析物。在其他情况下,每个流体小滴平均可出现超过三种目标分析物。在一些情况下,在同一流体小滴中可出现相同目标分析物的一条或多条链。在一些情况下,在流体小滴内存在少于1、2、3、4、5、6、7、8、9、10、50、100、1000、5000、10000或100000种目标分析物。在一些情况下,在流体小滴内存在多于1、2、3、4、5、6、7、8、9、10、50、100、1000、5000、10000或100000种目标分析物。通常,本文所述分区的特征在于具有极小的体积。例如,在基于小滴的分区的情况下,小滴可以具有小于1000pL、小于900pL、小于800pL、小于700pL、小于600pL、小于500pL、小于400pL、小于300pL、小于200pL、小于100pL、小于50pL、小于20pL、小于10pL或者甚至小于1pL的总体积。当与珠子共同分配时,应当理解,分区内的样品流体体积可以小于上述体积的90%、小于80%、小于70%、小于60%、小于50%、小于40%、小于30%、小于20%或者甚至小于上述体积的10%。

[0276] 当样品与条形码化的珠子组合时,在每个流体小滴中平均可存在少于一个珠子。在一些实施方案中,每个流体小滴中平均可存在少于两个珠子。在一些实施方案中,每个流体小滴平均可存在少于三个珠子。在一些情况下,每个流体小滴中平均可存在超过一个珠子。在其他情况下,每个流体小滴中平均可存在超过两个珠子。在其他情况下,每个流体小滴平均可存在超过三个珠子。在一些实施方案中,可采用有限稀释技术实现每个流体小滴少于一个条形码化珠子的比例。这里,条形码化的珠子可在与样品混合前稀释、在与样品混合期间稀释或在与样品混合后稀释。

[0277] 被分配的不同条形码或不同组条形码的数目(例如,不同组条形码,每个不同组偶联至不同的珠子)可以根据例如待分配的特定条形码和/或应用而不同。不同组条形码可以是,例如,相同条形码在各组之间有差异的成组相同条形码。或者不同组条形码可以是,例如,每一组就其包含的条形码而言有差异的成组不同条形码。在一些情况下,通过将不同的条形码附接至不同的珠子(例如,凝胶珠子)而分配不同的条形码。在一些情况下,通过将每个不同的组放置在不同的分区中而分配不同组的条形码。在一些情况下,分区可包含一个或多个不同的条形码组。例如,每个不同组的条形码可偶联至不同的珠子(例如,凝胶珠子)。可将每个不同的珠子分配到流体小滴中,使得每个不同组的条形码被分配到不同的流体小滴中。例如,可以分配约1、5、10、50、100、1000、10000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000、100000000个或更多个不同的条形码或不同组的条形码。在一些实例中,可以分配至少约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,

000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000、100000000个或更多个不同的条形码或不同组的条形码。在一些实例中,可以分配少于约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个不同的条形码或不同组的条形码。在一些实例中,可以分配约1-5、5-10、10-50、50-100、100-1000、1000-10000、10000-100000、100000-1000000、1000000-10000000或10000-100000000个不同的条形码或不同组的条形码。

[0278] 条形码可以以特定密度进行分配。例如,可以分配条形码以使得每个分区含有约1、5、10、50、100、1000、10000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个条形码/分区。可以分配条形码以使得每个分区含有至少约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000、100000000个或更多个条形码/分区。可以分配条形码以使得每个分区含有少于约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个条形码/分区。可以分配条形码以使得每个分区含有约1-5、5-10、10-50、50-100、100-1000、1000-10000、10000-100000、100000-1000000、1000000-10000000或10000-100000000个条形码/分区。在一些情况下,分配的条形码可以偶联至一个或多个珠子,例如,凝胶珠子。在一些情况下,该分区是流体小滴。

[0279] 可以分配条形码以使得相同的条形码以特定密度分配。例如,可以分配相同的条形码以使得每个分区含有约1、5、10、50、100、1000、10000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个相同的条形码/分区。可以分配条形码以使得每个分区含有至少约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个或更多个相同的条形码/分区。可以分配条形码以使得每个分区含有少于约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个或更多个相同的条形码/分区。可以分配条形码以使得每个分区含有少于约1、5、10、50、100、1000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个或更多个相同的条形码/分区。

100000000个相同的条形码/分区。可以分配条形码以使得每个分区含有约1-5、5-10、10-50、50-100、100-1000、1000-10000、10000-100000、100000-1000000、10000-10000000、10000-10000000或10000-100000000个相同的条形码/分区。在一些情况下,分配的相同条形码可以偶联至珠子,例如,凝胶珠子。在一些情况下,该分区是流体小滴。

[0280] 可以分配条形码以使得不同的条形码以特定密度分配。例如,可以分配不同的条形码以使得每个分区含有约1、5、10、50、100、1000、10000、20,000、30,000、40,000、50,000、60,000、70,000、80,000、90,000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个不同的条形码/分区。可以分配条形码以使得每个分区含有至少约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个或更多个不同的条形码/分区。可以分配条形码以使得每个分区含有少于约1、5、10、50、100、1000、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1000000、2000000、3000000、4000000、5000000、6000000、7000000、8000000、9000000、10000000、20000000、50000000或100000000个不同的条形码/分区。可以分配条形码以使得每个分区含有约1-5、5-10、10-50、50-100、100-1000、1000-10000、10000-100000、100000-1000000、10000-10000000或10000-100000000个不同的条形码/分区。在一些情况下,分配的不同条形码可以偶联至珠子,例如,凝胶珠子。在一些情况下,该分区是流体小滴。

[0281] 用于分配条形码或不同组的条形码的分区的数目可以变化,例如,取决于应用和/或待分配的不同条形码或不同组的条形码的数目。例如,用于分配条形码或不同组的条形码的分区的数目可以是约5、10、50、100、250、500、750、1000、1500、2000、2500、5000、7500或10,000、20000、30000、40000、50000、60000、70000、80000、90000、100,000、200000、300000、400000、500000、600000、700000、800000、900000、1,000,000、2000000、3000000、4000000、5000000、10000000、20000000或更多。用于分配条形码或不同组的条形码的分区的数目可以是至少约5、10、50、100、250、500、750、1000、1500、2000、2500、5000、7500、10,000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、2000000、3000000、4000000、5000000、10000000、20000000或更多。用于分配条形码或不同组的条形码的分区的数目可以少于约5、10、50、100、250、500、750、1000、1500、2000、2500、5000、7500、10,000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200000、300000、400000、500000、600000、700000、800000、900000、1000000、2000000、3000000、4000000、5000000、10000000、20000000或20000000。用于分配条形码的分区的数目可以是约5-10000000、5-5000000、5-1,000,000、10-10,000、10-5,000、10-1,000、1,000-6,000、1,000-5,000、1,000-4,000、1,000-3,000或1,000-2,000。在一些情况下,该分区可以是流体小滴。

[0282] 如上所述,可以分配不同的条形码或不同组的条形码(例如,每组包含多个相同的条形码或不同的条形码),以使得每个分区通常包含不同的条形码或不同的条形码组。在一

些情况下,每个分区可以包含不同组的相同条形码,如相同组的偶联至珠子(例如凝胶珠子)的条形码。当分配不同组的相同条形码时,每个分区中相同条形码的数目可以变化。例如,约100,000个或更多个不同组的相同条形码(例如,附接至珠子的一组相同的条形码)可以在约100,000个或更多个不同的分区中进行分配,使得每个分区包含不同组的相同条形码(例如,每个分区包含偶联至不同组的相同条形码的珠子)。在每个分区中,每组条形码的相同条形码的数目可以是约1,000,000个或更多相同的条形码(例如,每个分区包含偶联至一个或多个珠子的1,000,000个或更多相同的条形码)。在一些情况下,不同组条形码的数目可以等于或基本等于分区的数目,或者可以小于分区的数目。不同条形码或不同条形码组的任何合适的数目、每个分区中条形码的数目和分区的数目可以组合起来。因此,应当理解,可以用任何上述的每个分区的条形码密度以及任何上述的分区数目提供任何上述不同数目的条形码。

[0283] 微流体装置和小滴

[0284] 在一些情况下,本公开内容提供了用于制备珠子和用于将珠子(或其他类型的分区)与样品组合的装置,例如,用于共同分配样品组分和珠子。这样的装置可以是微流体装置(例如,小滴发生器)。该装置可以由任何合适的材料形成。在一些实例中,装置可由选自下组的材料形成:熔融二氧化硅、钠钙玻璃、硼硅酸盐玻璃、聚(甲基丙烯酸甲酯)PMMA、PDMS、蓝宝石、硅、锗、环烯烃共聚物、聚乙烯、聚丙烯、聚丙烯酸酯、聚碳酸酯、塑料、热固性塑料、水凝胶、热塑性塑料、纸、弹性体及其组合。

[0285] 可以以装置包含流体流动通道的方式形成装置。可以使用任何合适的通道。在一些情况下,装置包含一个或多个流体输入通道(例如,入口通道)和一个或多个流体出口通道。在一些实施方案中,流体通道的内径可以为约10 μm 、20 μm 、30 μm 、40 μm 、50 μm 、60 μm 、65 μm 、70 μm 、75 μm 、80 μm 、85 μm 、90 μm 、100 μm 、125 μm 或150 μm 。在一些实施方案中,流体通道的内径可以大于10 μm 、20 μm 、30 μm 、40 μm 、50 μm 、60 μm 、65 μm 、70 μm 、75 μm 、80 μm 、85 μm 、90 μm 、100 μm 、125 μm 、150 μm 或更大。在一些实施方案中,流体通道的内径可以小于约10 μm 、20 μm 、30 μm 、40 μm 、50 μm 、60 μm 、65 μm 、70 μm 、75 μm 、80 μm 、85 μm 、90 μm 、100 μm 、125 μm 或150 μm 。流体通道内的体积流速可以本领域已知的任何流速。

[0286] 如本文其他各处所述,可以使用微流体装置通过形成包含一种或多种凝胶前体、一种或多种交联剂、任选的引发剂和任选的水性表面活性剂的流体小滴而形成珠子。该流体小滴可以被不混溶的连续流体如油包围,该流体还可以包含表面活性剂和/或加速剂。

[0287] 在一些实施方案中,可以使用微流体装置通过形成同时包含珠子和样品的流体小滴(或其他类型的第二分区,包括本文所述的任何合适类型的分区)而使珠子(例如,条形码化的珠子或其他类型的第一分区,包括本文所述的任何合适类型的分区)与样品(例如,核酸样品)组合。该流体小滴可以具有被油相包围的水性核心,例如,油包水乳液内的水性小滴。该流体小滴可含有一个或多个条形码化的珠子、样品、扩增试剂和还原剂。在一些情况下,该流体小滴可包含以下的一种或多种:水、无核酸酶的水、乙腈、珠子、凝胶珠子、聚合物前体、聚合物单体、聚丙烯酰胺单体、丙烯酰胺单体、可降解的交联剂、不可降解的交联剂、二硫键、acrydite部分、PCR试剂、引物、聚合酶、条形码、多核苷酸、寡核苷酸、核苷酸、DNA、RNA、肽多核苷酸、互补DNA(cDNA)、双链DNA(dsDNA)、单链DNA(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RNA、dsRNA、探针、染料、有机物、乳化剂、表面活性剂、稳定剂、聚合物、适体、还原剂、引发剂、生物素标记物、荧光团、缓冲液、酸性溶液、碱性溶液、光敏感的酶、pH敏感的酶、水性缓冲液、油、盐、去污剂、离子型去污剂、非离子型去污剂，等等。总之，该流体小滴的组成将根据特定的处理需求而改变。

[0288] 流体小滴可以具有均匀的大小或不均匀的大小。在一些情况下，流体小滴的直径可以为约1μm、5μm、10μm、20μm、30μm、40μm、45μm、50μm、60μm、65μm、70μm、75μm、80μm、90μm、100μm、250μm、500μm或1mm。在一些情况下，流体小滴可以具有至少约1μm、5μm、10μm、20μm、30μm、40μm、45μm、50μm、60μm、65μm、70μm、75μm、80μm、90μm、100μm、250μm、500μm、1mm或更大的直径。在一些情况下，流体小滴可以具有小于约1μm、5μm、10μm、20μm、30μm、40μm、45μm、50μm、60μm、65μm、70μm、75μm、80μm、90μm、100μm、250μm、500μm或1mm的直径。在一些情况下，流体小滴可以具有在约40–75μm、30–75μm、20–75μm、40–85μm、40–95μm、20–100μm、10–100μm、1–100μm、20–250μm或20–500μm的范围内的直径。

[0289] 在一些实施方案中，所述装置可包含两个或更多个流体输入通道的一个或多个交叉点。例如，该交叉点可以是流体交叉点。该流体交叉点可包含两个或更多个流体输入通道和一个或多个流体出口通道。在一些情况下，该流体交叉点可包含两个流体输入通道和两个流体出口通道。在其他情况下，该流体交叉点可包含三个流体输入通道和一个流体出口通道。在一些情况下，该流体交叉点可以在形成该交叉点的两个或更多个流体通道之间形成基本垂直的角度。

[0290] 在一些情况下，微流体装置可包含在接合处与出口通道流体连接的第一和第二输入通道。在一些情况下，例如，出口通道可以与第三输入通道在接合处流体连接。在一些情况下，可包含第四输入通道并且其可以与第三输入通道和出口通道在接合处相交。在一些情况下，微流体装置可包含第一、第二和第三输入通道，其中第三输入通道与第一输入通道、第二输入通道或者第一输入通道与第二输入通道的接合处相交。

[0291] 如本文其他各处所述，所述微流体装置可用于从液体产生凝胶珠子。例如，在一些实施方案中，在流体输入通道内包含一种或多种凝胶前体、一种或多种交联剂和任选的引发剂、任选的水性表面活性剂以及任选的醇的水性流体可以进入流体交叉点。在第二流体输入通道内，任选地具有表面活性剂和加速剂的油可以进入相同的流体交叉点。水性和油组分可在流体交叉点处混合，导致在连续油相中形成水性流体小滴。离开流体交叉点的流体小滴内的凝胶前体可以聚合形成珠子。

[0292] 如本文其他各处所述，所述微流体装置（例如，小滴发生器）可用于将样品与珠子（例如，条形码化珠子的文库）以及（如果需要）能够降解珠子的试剂（例如，如果珠子以二硫键连接，则是还原剂）组合。在一些实施方案中，可向与第一流体交叉点（例如，第一流体接合处）流体连接的第一流体输入通道提供样品（例如，核酸样品）。可以向同样与第一流体交叉点流体连接的第二流体输入通道提供预形成的珠子（例如，条形码化的珠子、可降解的珠子），其中第一流体输入通道与第二流体输入通道在该第一流体交叉点交汇。样品和珠子可以在第一流体交叉点混合以形成混合物（例如，水性混合物）。在一些情况下，可向第三流体输入通道提供还原剂，该第三流体输入通道同样与第一流体交叉点流体连接，并且与第一和第二流体输入通道在第一流体交叉点交汇。然后，还原剂可以与珠子和样品在第一流体交叉点混合。在其他情况下，可在进入微流体装置之前将还原剂与样品和/或珠子预混合，

使得通过第一流体输入通道向微流体装置提供样品和/或通过第二流体输入通道向微流体装置提供珠子。在其他情况下，可以不添加还原剂。

[0293] 在一些实施方案中，样品和珠子混合物可以通过与第一流体交叉点（并因此，与构成第一流体交叉点的任何流体通道）流体连接的第一出口通道离开第一流体交叉点。可以向与第一出口通道流体连接的第二流体交叉点（例如，第二流体接合处）提供混合物。在一些情况下，油（或其他合适的不混溶的）流体可以从与第二流体交叉点（并因此，与构成该交叉点的任何流体通道）流体连接且在第二流体交叉点与第一出口通道交汇的一个或多个单独的流体输入通道进入第二流体交叉点。在一些情况下，可以在与第二流体交叉点（并因此，与第一出口通道）流体连接且在第二流体交叉点与第一出口通道以及彼此交汇的一个或两个单独的流体输入通道中提供油（或其他合适的不混溶的流体）。两种组分—油以及样品与珠子的混合物—可以在第二流体交叉点混合。该混合将样品与珠子混合物分配至多个流体小滴（例如，油包水乳液内的水性小滴），其中形成的小滴的至少一个子集封装条形码化的珠子（例如，凝胶珠子）。形成的流体小滴可在油内被运送通过从第二流体交叉点离开的第二流体出口通道。在一些情况下，从第二流体交叉点离开第二出口通道的流体小滴可被分配到孔中以供进一步处理（例如，热循环）。

[0294] 在许多情况下，控制所得的小滴（或第二分区）相对于珠子（或第一分区）的占有率将是期望的。这种控制在例如于2014年4月4日提交的美国临时专利申请号61/977,804中描述，其全部公开内容为了所有目的通过引用以全文并入本文。通常，将形成小滴（或第二分区），使得至少50%、60%、70%、80%、90%或更多的小滴（或第二分区）含有不超过一个珠子（或第一分区）。另外，或备选地，将形成小滴（或第二分区），使得至少50%、60%、70%、80%、90%或更多的小滴（或第二分区）包含恰好一个珠子（或第一分区）。在一些情况下，所得的小滴（或第二分区）平均可各自包含至多约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个珠子（或第一分区）。在一些情况下，所得的小滴（或第二分区）平均可各自包含至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个珠子（或第一分区）。

[0295] 在一些实施方案中，可以在混合物进入微流体装置中之前将样品与包含条形码和任何其他试剂（例如，样品扩增所需的试剂、还原剂等）的珠子（例如，可降解的珠子）预混合以产生水性反应混合物。在水性混合物进入流体装置时，该混合物可从第一流体输入通道流动并进入流体交叉点。在一些情况下，油相可以从同样与流体交叉点流体连接的第二流体输入通道（例如，与第一流体输入通道垂直或基本垂直的流体通道）进入流体交叉点。该水性混合物和油可以在流体交叉点混合，使得乳液（例如，凝胶-水-油乳液）形成。该乳液可包含在连续油相中的多个流体小滴（例如，包含水性反应混合物的小滴）。在一些情况下，每个流体小滴可包含单个珠子（例如，附接至一组相同的条形码的凝胶珠子）、样品的等份以及任何其他试剂（例如，还原剂、样品扩增所需的试剂等）的等份。然而，在一些情况下，流体小滴可包含多个珠子。在小滴形成时，小滴可通过连续油相被运送通过离开流体交叉点的流体出口通道。离开出口通道的流体小滴可被分配到孔中以供进一步处理（例如，热循环）。

[0296] 在可在进入微流体装置之前将还原剂添加至样品或者可在第一流体交叉点添加还原剂的情况下，在第二流体交叉点形成的流体小滴可含有还原剂。在这种情况下，当小滴穿过离开第二流体交叉点的出口通道行进时，还原剂可降解或溶解流体小滴内含有的珠

子。

[0297] 在一些实施方案中，微流体装置可含有平行的三个离散的流体交叉点。流体小滴可以在这三个流体交叉点的任一个处形成。样品和珠子可以在这三个流体交叉点的任一个内混合。还原剂可以在这三个流体交叉点的任一个处添加。油可以在这三个流体交叉点的任一个处添加。

[0298] 本公开内容的方法、组合物、装置和试剂盒可与任何合适的油一起使用。在一些实施方案中，油可用于产生乳液。该油可以包括氟化油、硅油、矿物油、植物油及其组合。

[0299] 在一些实施方案中，微流体装置内的水性流体也可含有醇。例如，醇可以是甘油、乙醇、甲醇、异丙醇、戊醇、乙烷、丙烷、丁烷、戊烷、己烷及其组合。该醇可以以约5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20% (v/v) 存在于水性流体内。在一些情况下，该醇可以以至少约5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%或更高 (v/v) 的浓度存在于水性流体内。在一些情况下，该醇可以以小于约5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%或20% (v/v) 存在于水性流体内。

[0300] 在一些实施方案中，所述油也可含有表面活性剂以稳定乳液。例如，表面活性剂可以是含氟表面活性剂、Krytox润滑剂、Krytox FSH、工程化的流体、HFE-7500、硅酮化合物、含PEG的硅化合物，如bis krytox peg (BKP)。该表面活性剂可以以约0.1%、0.5%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、5%或10% (w/w) 存在。在一些情况下，该表面活性剂可以以至少约0.1%、0.5%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、5%、10% (w/w) 或更高的浓度存在。在一些情况下，该表面活性剂可以以小于约0.1%、0.5%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、5%或10% (w/w) 存在。

[0301] 在一些实施方案中，可向油中添加加速剂和/或引发剂。例如，加速剂可以是四甲基乙二胺 (TMEDA或TEMED)。在一些情况下，引发剂可以是过硫酸铵或钙离子。该加速剂可以以约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%或2% (v/v) 存在。在一些情况下，该加速剂可以以至少约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%或2% (v/v) 或更高的浓度存在。在一些情况下，该加速剂可以以小于约0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%或2% (v/v) 存在。

[0302] V. 扩增

[0303] DNA扩增是一种用于创建DNA小区段或长区段的多个拷贝的方法。本公开内容的方法、组合物、装置和试剂盒可采用DNA扩增来将一个或多个期望的寡核苷酸序列(如条形码序列或随机N-mer序列)附接至单独的珠子。DNA扩增还可用于利用随机N-mer序列沿感兴趣的样品如基因组DNA引发和延伸，以便产生样品序列的片段并将与引物相关联的珠子偶联至该片段。

[0304] 例如，可通过将模板核酸序列和包含多个附接的寡核苷酸(例如，可释放地附接的寡核苷酸)的珠子共同分配到分区(例如，乳液的小滴、微胶囊或任何其他合适的分区类型，

包括本文其他各处描述的合适的分区类型)中而扩增核酸序列。附接的寡核苷酸可包含与模板样品核酸序列的一个或多个区域互补的引物序列(例如,可变引物序列,如随机N-mer,或者靶向引物序列,如靶向N-mer),并且此外还可包含共同序列(例如,如条形码序列)。该引物序列可以与模板核酸序列退火并延伸(例如,在引物延伸反应或任何其他合适的核酸扩增反应中),以产生模板核酸的至少一部分的一个或多个第一拷贝,使得该一个或多个第一拷贝包含引物序列和共同序列。在包含引物序列的寡核苷酸可释放地附接至珠子的情况下,可在引物序列与模板核酸序列退火之前从珠子释放寡核苷酸。此外,通常,该引物序列可通过同样在该分区中提供的聚合酶(例如,如本文其他各处所述的链置换聚合酶、如本文其他各处所述的核酸外切酶缺陷的聚合酶或任何其他类型的合适的聚合酶,包括本文其他各处所述的聚合物类型)延伸。此外,可释放地附接至珠子的寡核苷酸可以是核酸外切酶抗性的,并因此可以包含一个或多个如本文其他各处所述的硫代磷酸酯联接。在一些情况下,所述一个或多个硫代磷酸酯联接可包含在寡核苷酸的末端核苷酸间联接处的硫代磷酸酯联接。

[0305] 在一些情况下,在产生一个或多个第一拷贝后,可使引物序列与一个或多个第一拷贝退火,并且再次延伸引物序列以产生一个或多个第二拷贝。所述一个或多个第二拷贝可包含引物序列、共同序列,并且还可以包含与所述一个或多个第一拷贝的单个拷贝的至少一部分互补的序列,和/或与可变引物序列互补的序列。上述步骤可以重复所需的循环数以产生扩增的核酸。

[0306] 上述寡核苷酸可包含在延伸反应(如产生上述一个或多个第一或第二拷贝的延伸反应)期间没有被拷贝的序列段。如本文其他各处所述,这样的序列段可包含一个或多个含尿嘧啶的核苷酸,并且还可导致在退火条件下形成发夹(或部分发夹)分子的扩增子的产生。

[0307] 在另一实例中,可通过将不同的核酸分配到各自包含第二分区(例如,珠子,包括本文其他各处所述的珠子类型)的单独的第一分区(例如,乳液中的小滴)中而扩增多种不同的核酸。第二分区可以与多个寡核苷酸可释放地关联。第二分区可包含任何合适数目的寡核苷酸(例如,每个本文所述的分区有超过1,000个寡核苷酸、超过10,000个寡核苷酸、超过100,000个寡核苷酸、超过1,000,000个寡核苷酸、超过10,000,000个寡核苷酸或任何其他数目的寡核苷酸)。此外,第二分区可包含任何合适数目的不同条形码序列(例如,至少1,000个不同的条形码序列、至少10,000个不同的条形码序列、至少100,000个不同的条形码序列、至少1,000,000个不同的条形码序列、至少10,000,000个不同的条形码序列或任何其他数目的本文其他各处所述的不同条形码序列)。

[0308] 此外,与给定第二分区相关联的多个寡核苷酸可包含引物序列(例如,可变引物序列、靶向引物序列)和共同序列(例如,条形码序列)。而且,与不同第二分区相关联的多个寡核苷酸可包含不同的条形码序列。与多个第二分区相关联的寡核苷酸可被释放到第一分区中。释放之后,第一分区内的引物序列可以与第一分区内的核酸退火,然后可以延伸引物序列以在第一分区产生该核酸的至少一部分的一个或多个拷贝。通常,所述一个或多个拷贝可包含释放到第一分区中的条形码序列。

[0309] 在小滴内的扩增和样品索引化

[0310] 可以对流体小滴内的内含物进行核酸(例如,DNA)扩增。如本文所述,流体小滴可

含有附接至珠子的寡核苷酸。流体小滴还可包含样品。流体小滴还可包含适合于扩增反应的试剂,该试剂可包括Kapa HiFi Uracil Plus、修饰的核苷酸、天然核苷酸、含尿嘧啶的核苷酸、dTTP、dUTP、dCTP、dGTP、dATP、DNA聚合酶、Taq聚合酶、突变校正聚合酶、9degrees North、修饰的(NEB)、exo (-)、exo (-) Pfu、Deep Vent exo (-)、Vent exo (-)以及无环核苷酸(acyNTPS)。

[0311] 附接至流体小滴内的珠子的寡核苷酸可用于扩增样品核酸,使得该寡核苷酸变为附接至样品核酸。该样品核酸可包含几乎任何寻求分析的核酸,包括,例如,全基因组、外显子组、扩增子、靶向的基因组区段(例如,基因或基因家族)、细胞核酸、循环核酸等,并且如上所述,可包括DNA(包括gDNA、cDNA、mtDNA等)、RNA(例如,mRNA、rRNA、总RNA等)。这类用于条形码化的核酸的制备通常可通过易于获得的方法来实现,例如,富集或沉降方法、分离方法、扩增方法等。为了扩增期望的样品,如gDNA,流体小滴内的寡核苷酸的随机N-mer序列可用于引发期望的靶序列并作为靶序列的互补序列延伸。在一些情况下,可在引发前,如本文其他各处所述将寡核苷酸从小滴中的珠子释放。对于这些引发和延伸过程,可以使用任何合适的DNA扩增方法,包括聚合酶链反应(PCR)、数字PCR、逆转录PCR、多重PCR、巢式PCR、重叠-延伸PCR、定量PCR、多重置换扩增(MDA)或连接酶链反应(LCR)。在一些情况下,可以进行流体小滴内的扩增,直到可产生一定量的包含条形码的样品核酸。在一些情况下,扩增可以进行约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个循环。在一些情况下,扩增可以进行超过约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多的循环。在一些情况下,扩增可以进行少于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个循环。

[0312] 如本文所述的示例性扩增和条形码化过程在图38中示意性地示出。如图所示,包含条形码序列的寡核苷酸与样品核酸3804一起被共同分配到例如乳液中的小滴3802中。如图A所示,如本文其他各处所述,可在与样品核酸3804共同分配的珠子3806上提供寡核苷酸3808,该寡核苷酸优选地是可从珠子3806释放的。除了一个或多个功能序列,例如,序列3810、3814和3816之外,寡核苷酸3808还包含条形码序列3812。例如,显示寡核苷酸3808包含条形码序列3812,以及可以起到给定测序系统的附接或固定序列作用的序列3810,例如,用于在Illumina Hiseq或Miseq系统的流动池中附接的P5序列。如图所示,该寡核苷酸还包含引物序列3816,该引物序列3816可包含用于引发样品核酸3804的部分的复制的随机或靶向N-mer。寡核苷酸3808还包含序列3814,该序列3814可提供用于在测序系统中引发聚合酶介导的、模板引导的合成测序反应的测序引发区,如“read1”或R1引发区。在许多情况下,条形码序列3812、固定序列3810和R1序列3814将是附接至给定珠子的所有寡核苷酸所共同的。对于随机N-mer引物,引物序列3816可以变化,或者可以是用于某些靶向应用的给定珠子上的寡核苷酸所共同的。

[0313] 基于引物序列3816的存在,所述寡核苷酸能够如图B所示引发样品核酸,这允许利用聚合酶和其他同样与珠子3806和样品核酸3804共同分配的延伸试剂来延伸寡核苷酸3808和3808a。如本文其他各处所述,这些聚合酶可包括热稳定的聚合酶,例如,当期望有分区内的双链样品核酸的初始变性时。或者,样品核酸的变性可以在分配之前,使得单链靶核酸被存放到分区中,从而允许在需要时使用非热稳定的聚合酶,例如,Klenow、phi29、Pol 1等。如图C所示,在对于随机N-mer引物,将与样品核酸3804的多个不同区域退火的寡核苷酸

的延伸后,创建核酸的多个重叠互补序列或片段,例如,片段3818和3820。尽管包含与样品核酸的部分互补的序列部分,例如,序列3822和3824,但这些构建体在本文中通常被称为包含样品核酸3804片段,具有附接的条形码序列。在一些情况下,可能期望人工地限制所产生的复制片段的大小,以便从第一扩增步骤维持可处理的片段大小。在一些情况下,如上所述,这可以通过机械手段实现,例如,采用片段化系统如Covaris系统,或者可以通过引入随机延伸终止剂(例如,以低浓度)以防止过长片段的形成而实现。

[0314] 然后,这些片段可进行序列分析,或者其可以如图D所示在该过程中进一步扩增。例如,同样从珠子3806释放的附加寡核苷酸,例如寡核苷酸3808b,可引发片段3818和3820。这针对片段3818示出。特别地,再次,基于寡核苷酸3808b中随机N-mer引物3816b(其在许多情况下将不同于给定分区中的其他随机N-mer,例如,引物序列3816)的存在,该寡核苷酸与片段3818退火,并被延伸以创建包含序列3828的片段3818的至少一部分的互补序列3826,其包含样品核酸序列的一部分的复制物。寡核苷酸3808b的延伸持续直到其已经复制通过片段3818的寡核苷酸部分3808。如上文其他各处所述,以及如图D中所示,该寡核苷酸可被配置用于在期望的点促进聚合酶引起的复制的停止,例如,在复制通过包含在片段3818内的寡核苷酸3808的序列3816和3814之后。如本文所述,这可以通过不同的方法实现,包括例如引入不能被所使用的聚合酶处理的不同核苷酸和/或核苷酸类似物。例如,这可以包括在序列区域3812内包含含尿嘧啶的核苷酸,以使非尿嘧啶耐受的聚合酶停止该区域的复制。结果,创建了片段3826,其在一端包含全长寡核苷酸3808b,包括条形码序列3812、附接序列3810、R1引物区域3814以及随机n-mer序列3816b。在该序列的另一端将包含第一寡核苷酸3808的随机n-mer的互补序列3816',以及R1序列的全部或一部分的互补序列,示为序列3814'。然后,R1序列3814和其互补序列3814'能够杂交在一起以形成部分发夹结构3828。应当理解,因为不同寡核苷酸之间随机n-mer不同,所以预计这些序列和其互补序列将不会参与发夹形成,例如,随机N-mer序列3816的互补序列—序列3816',预计将不会与随机n-mer序列3816b互补。对于其他应用将不是这样,例如,靶向引物,其中N-mer可以在给定分区内的寡核苷酸之间是共同的。

[0315] 通过形成这些部分发夹结构,将允许从进一步复制中去除样品序列的第一水平复制物,例如,防止拷贝的反复复制。这种部分发夹结构还提供对所创建的片段(例如,片段3826)的后续处理有用的结构。

[0316] 在条形码附接至样品后,可进行另外的扩增步骤(例如,PCR)以在测序前扩增条形码化的片段,以及任选地向那些条形码化的片段添加附加功能序列,例如,与测序装置(例如,Illumina MiSeq)兼容的附加引物结合位点(例如,Read2序列引物、Index引物),和任选的一个或多个附加条形码序列(例如,见图14C),以及其他功能序列,例如,附加固定序列或其互补序列,例如P7序列。在一些情况下,附加条形码序列可充当样品索引,其中原始条形码和样品索引允许进行多重测序(例如,同时的分子标记和样品鉴定)。原始条形码可在测序期间使用,以比对与该条形码关联的核酸分子所对应的序列读取结果(例如,通过条形码鉴定的)。由每个不同样品产生的、测序仪就绪产物可包含不同的样品索引。因此,该样品索引可在测序期间使用,以用于鉴定特定序列读取结果所属的样品,并且可以实现多重化。

[0317] 在一些情况下,可在使用或不使用分区或产生或不产生附加分区的情况下,在原始条形码添加至样品核酸之后将样品索引添加至样品核酸。在一些情况下,样品索引批量

加入。在一些情况下，样品索引向样品核酸的添加可在条形码向样品核酸的添加之前发生。在一些情况下，样品索引向样品核酸的添加可以与样品索引向样品核酸的添加同时或平行发生。

[0318] 在一些情况下，可在向样品核酸添加条形码序列之后将样品索引物添加至样品核酸。例如，如本文其他各处所述，可以使用扩增方法将条形码序列和其他序列（例如，P5、R1等）附接至样品核酸。在一些情况下，例如，随机扩增方案，如用于测序的部分发夹扩增（PHASE—如本文其他各处所述），可帮助将条形码序列和其他序列附接至样品核酸。在一个实例中，使用多个引物随机引发和扩增样品核酸，这些引物各自包含不同的随机N-mer、测序仪附接或固定位点（例如，P5）、条形码序列（例如，相同的条形码序列）和测序引物结合位点（例如，R1）。测序仪引物结合位点、条形码序列和/或测序引物结合位点中的任一种可包含含尿嘧啶的核苷酸。该引物也可包含在引物的一个或多个序列处与引物杂交的寡核苷酸阻断剂，以确保样品核酸的引发仅通过随机N-mer发生。示例引物的图形表示如下（寡核苷酸阻断剂未示出）：

[0319] P5-条形码-R1-随机NMer

[0320] 样品核酸的随机引发和多轮扩增可产生扩增子，该扩增子包含在一端与测序仪附接或固定位点（例如，P5）连接的样品核酸的一部分、条形码、测序引物结合位点（例如，R1）以及随机N-mer。在其另一端，该样品核酸的部分可以连接至与测序引物结合位点互补或部分互补的区域（例如，R1c或R1c（部分的））。示例序列（呈线性构型）的图形表示如下：

[0321] P5-条形码-R1-随机Nmer-插入物-R1c（部分的）

[0322] 其中“插入物”对应于在扩增期间复制的样品核酸的部分。测序引物结合位点（例如，R1）和在复制的样品核酸部分（插入物）的相对端的其部分互补序列（例如，R1c（部分的））可以分子内杂交，以形成如本文其他各处所述的部分发夹结构。

[0323] 创建样品核酸的条形码化片段之后，并且如上所述，可希望进一步扩增这些片段，以及将附加功能序列附接至扩增的、条形码化的片段。可采用任何合适的扩增过程进行该扩增，包括，例如，PCR、LCR、线性扩增等。通常，可采用针对所创建的片段中的已知末端序列引发的靶向引物（例如，针对图38中附接序列3810和序列3814'中的一个或两个引发）来启动这种扩增。进一步通过在这些引物中引入附加功能序列，例如，附加附接序列如P7，附加测序引物，例如读取2或R2引发序列，以及任选的样品索引化序列，可进一步配置所扩增的条形码化片段。

[0324] 举例而言，在产生部分发夹扩增子之后，可通过将该部分发夹扩增子接触与发夹的双链体部分互补的引物（例如，3814'），以便破坏发夹和引发沿发夹结构的延伸，从而破坏该部分发夹扩增子的分子内杂交。在许多情况下，将期望为这些引物提供比发夹结构更强的杂交亲和性，以便优先地破坏该发夹。这样，在至少一个实例中，引物包含锁核酸（LNA）或锁核酸核苷酸。LNA包括这样的核苷酸，该核苷酸中的核糖核酸碱基包含将核苷酸核糖部分的2'-氧与4'-碳连接的分子桥。LNA通常具有较高的解链温度和较低的杂交能。因此，LNA可以通过与部分发夹扩增子的任意杂交的序列结合来与部分发夹扩增子的分子内杂交有利地竞争。破坏的扩增子随后通过包含LNA的引物和其他引物的扩增可产生包含将要添加至该序列的任何附加序列（包括样品索引）的线性产物。

[0325] 对于上述的示例部分发夹P5-条形码-R1-随机Nmer-插入物-R1c（部分的）构型，部

分发夹可以与包含LNA以及与R1c(部分的)互补的序列的引物接触(例如,见图14C)。该引物还可包含将要添加到构建体的任何附加序列的互补序列。例如,该附加序列(例如,R2(部分的))可以是这样一种序列,当其与R1c(部分的)偶联时,其产生另外的测序引物结合位点(例如,R2)。引物与部分发夹的杂交可以破坏部分发夹的分子内杂交并使构建体线性化。杂交可以发生,例如,使得引物通过其互补序列与R1c(部分的)杂交(例如,见图14C)。引物的延伸可以产生包含与一种序列连接的引物的构建体,该序列与线性化的部分发夹扩增子互补。示例构建体的图示如下:

[0326] P5c-条形码,c-R1c-随机Nmer,c-插入物,c-R1(部分的)-R2(部分的),c其中P5c对应于P5的互补序列、条形码,c对应于条形码的互补序列、随机Nmer,c对应于随机N-mer的互补序列、插入物,c对应于插入物的互补序列、并且R1(部分的)-R2(部分的),c对应于R2的互补序列。

[0327] 在使用第二引物(例如,P5,在P5c处杂交)的另一轮扩增时,可以产生包含部分发夹扩增子序列和与该引物互补的序列的线性构建体。示例构型的图形表示如下:

[0328] P5-条形码-R1-随机Nmer-插入物-R1c(部分的)-R2(部分的)或者P5-条形码-R1-随机Nmer-插入物-R2

[0329] 其中R1c(部分的)和R2(部分的)的组合序列可对应于附加测序引物结合位点(例如,R2)。

[0330] 可采用额外轮次的这种扩增将附加序列添加至构建体,可进行无论多少个期望的附加序列/扩增轮次。对于上述的示例P5-条形码-R1-随机Nmer-插入物-R2构建体,包含与R2互补的序列(例如,R2c)、样品索引序列的互补序列(例如,SIc,样品条形码)和附加测序仪引物结合位点序列的互补序列(P7c)的引物可以在R2处通过该引物的R2c与构建体杂交(例如,见图14C)。引物的延伸可以产生包含与互补于构建体的序列连接的引物的构建体。示例构型的图形表示如下:

[0331] P5c-条形码,c-R1c-随机Nmer,c-插入物,c-R2,c-SIc-P7c

[0332] 在使用第二引物(例如,P5,在P5c处杂交)的另一轮扩增时,可以产生包含构建体序列和与该引物互补的序列的测序仪就绪构建体。这种测序仪就绪构建体的示例构型的图形表示如下:

[0333] P5-条形码-R1-随机Nmer-插入物-R2-样品索引-P7。作为替代,起始引物可包含条形码序列、P7和R2(而不是P5和R1)。示例引物的图形表示如下:

[0334] P7-条形码-R2-随机Nmer

[0335] 采用如上述的类似扩增方案(例如,使用包含LNA的引物的扩增、额外轮次的扩增,等),可以对包含样品核酸的一部分的插入物进行测序,可将P5、R1和样品索引添加至引物以产生测序仪就绪产物。示例产物的图形表示如下:

[0336] P7-条形码-R2-随机Nmer-插入物-R1-样品索引-P5

[0337] 在其他情况下,可在向样品核酸添加条形码序列的同时将样品索引添加至样品核酸。例如,用于产生条形码化的样品核酸的引物可同时包含条形码序列和样品索引,使得当条形码与样品核酸偶联时,样品索引同时得到偶联。样品索引可以被放置在引物序列中的任何位置。在一些情况下,引物可以是能够通过随机扩增如PHASE扩增产生条形码化的样品核酸的引物。这类引物的实例的图形表示包括:

[0338] P5-条形码-R1-样品索引-随机Nmer

[0339] P5-条形码-样品索引-R1-随机Nmer

[0340] P5-样品索引-条形码-R1-随机Nmer

[0341] 当采用各自的引物随机引发样品核酸并且样品核酸在分区中扩增时,可以产生包含条形码序列和样品索引序列的部分发夹扩增子。由上述引物产生的这类部分发夹扩增子的实例的图形表示(以线性形式示出)分别包括:

[0342] P5-条形码-R1-样品索引-随机Nmer-插入物-R1c(部分的)

[0343] P5-条形码-样品索引-R1-随机Nmer-插入物-R1c(部分的)

[0344] P5-样品索引-条形码-R1-随机Nmer-插入物-R1c(部分的)

[0345] R1c(部分的)可以与R1中的其互补序列发生分子内杂交以形成部分发夹扩增子。

[0346] 举例而言,在一些情况下,部分发夹扩增子产生后,可将附加序列(例如,如R2和P7序列的功能序列)添加至(例如,批量地)部分发夹扩增子。以与本文其他各处所述的扩增方法类似的方式,例如,通过针对上述部分发夹的5'端(例如,R1c序列)引发,包含这些附加功能序列的引物可用于引发部分发夹分子的复制。在许多情况下,将期望提供高亲和性引物序列,例如,以超过发夹结构的再杂交,以便提供更好的引发和复制。在这类情况下,更紧密结合的引物序列(例如,在其序列中包含一种或多种更高亲和性的核苷酸类似物,如LNA等)可用于破坏部分发夹扩增子并向该扩增子添加附加序列。例如,对于上述实例,引物可包含LNA、与R1c(部分的)互补的序列以及包含R2(部分的)的互补序列的序列,使得当该引物延伸并且所得产物进一步经由P5引物扩增时,R1c(部分的)和R2(部分的)接合以产生R2。由上述引物产生的这类构建体的实例的图形表示分别包括:

[0347] P5-条形码-R1-样品索引-随机Nmer-插入物-R2

[0348] P5-条形码-样品索引-R1-随机Nmer-插入物-R2

[0349] P5-样品索引-条形码-R1-随机Nmer-插入物-R2

[0350] 如上所述,可以利用额外轮次的扩增循环将附加序列添加至构建体。例如,引物可包含与R2互补的序列以及包含P7的互补序列的序列,使得当该引物延伸并且所得产物进一步经由P5引物扩增时,P7与R2连接并产生测序仪就绪构建体。由上述引物产生的这类测序仪就绪构建体的实例的图形表示分别包括:

[0351] P5-条形码-R1-样品索引-随机Nmer-插入物-R2-P7

[0352] P5-条形码-样品索引-R1-随机Nmer-插入物-R2-P7

[0353] P5-样品索引-条形码-R1-随机Nmer-插入物-R2-P7

[0354] 将条形码和样品索引组合为能够扩增样品核酸的区域的引物(例如,通过PHASE扩增)可使得样品索引化的平行化成为可能。引物组可用于对来自不同样品的核酸进行索引化。每组引物均可与获自特定样品的核酸分子相关联,并且包括含有多样性的条形码序列和共同样品索引序列的引物。

[0355] 在一些情况下,可期望将附加序列段附接至本文所述的部分发夹分子的5'端,以便不仅向如上所述的样品核酸的扩增片段提供额外的功能性,而且确保对那些分子的列有效的后续处理,例如,扩增和/或测序。例如,当部分发夹分子经受延伸反应条件时,通过经由5'末端处的延伸来引发其自身的“补平”反应,可能易于进行部分发夹结构的补平。因此,可以创建更加难以扩增(由于其双链体部分的更高稳定性)的完整发夹结构。在这类情况

下,可期望优先地附接不与相对端序列互补的附加序列段,以便防止完整发夹结构的形成。在一个示例性过程中,上述用于扩增部分发夹结构的LNA引物可以具有附加突出序列,包括,例如,上述的R2互补序列,以及与其他功能序列组分可能互补的序列,例如,附接序列如P7、样品索引序列等。使部分发夹和引物经历以上对于部分发夹的扩增所述的延伸反应也将导致部分发夹沿LNA引物上突出序列的延伸。延伸的序列可以仅包含非互补的序列,或者可以包含如上所述的附加功能序列或其互补序列,使得该延伸反应导致那些功能序列附接至部分发夹结构的5'末端。

[0356] 在备选的方面,附加序列段可以与部分发夹结构的5'端连接,其中这类序列段不与发夹结构的非重叠部分互补。前面所述的那些在图40中图示。如图A所示,当经历引物延伸反应时,部分发夹结构可充当其自身的引物,并且其5'序列得到延伸(如由虚线箭头示出的),直到其形成完整的或几乎完整的发夹结构,例如,没有或几乎没有突出端序列。该完整的发夹结构将具有大得多的双链体稳定性,由此可能消极地影响破坏发夹结构以引发其复制的能力,即使当采用更高亲和性的引物(例如,含LNA的引物/探针)时。

[0357] 为了使该可能性最小化,如路径B和C所示,将单独的序列段4006添加到发夹结构的5'端,以向部分发夹提供非互补的尾序列4008,以便防止产生完整或几乎完整的发夹结构。如图所示,这可以通过一些不同的方式实现。例如,在路径B中所示的第一过程中,侵入式探针4010可用于破坏该发夹结构并与序列段4012杂交。例如,通过使用更高亲和性的核苷酸类似物如LNA等,这类侵入式探针可以具有比固有的部分发夹结构更高亲和力的结合。特别地,与序列段4012杂交的侵入序列4010的部分可以在其序列内包含LNA,其方式等同于本文所述的在随后扩增中使用的LNA引物序列所使用的方式。

[0358] 部分发夹(和序列段4012)的5'部分的延伸如路径B中的虚线箭头所示,然后将序列4006附于部分发夹结构的5'末端以提供结构4008。或者,序列4006可以与部分发夹结构4002(或序列段4012)的5'端连接。如路径C所示,这通过使用夹板序列4014实现,该夹板序列与序列4006部分互补并与序列4012部分互补,以便使序列4006保持与序列段4012相邻以便连接。应当理解,夹板序列4014可再次利用更高亲和性的侵入式探针,如探针4010,来破坏发夹结构并与序列段4012杂交。特别地,再次,旨在与序列段4012杂交的夹板序列4014的部分可以在其序列内具有一种或多种LNA核苷酸类似物,以便优先破坏部分发夹结构4002,并使得序列4006能够与其5'端连接。

[0359] 在一些情况下,微流体装置(例如,微流体芯片)可能在样品索引化的平行化方面是有用的。这样的装置可包含平行模块,每个模块均能够通过同时包含条形码序列和样品索引的引物将条形码序列和样品索引添加至样品的核酸分子。每个平行模块可包含含有不同样品索引的引物组,使得在每个模块中处理的样品与不同的样品索引和条形码组相关联。例如,具有8个模块的微流体装置可以能够对8种不同的样品进行样品索引化。在通过将序列附接至样品核酸而进行条形码化和样品索引化后,可以利用通过例如连续扩增批量添加附加序列(例如,R2、P7、其他条形码序列)来产生如本文其他各处所述的测序仪就绪产物。

[0360] 在一些情况下,样品索引化可以在条形码化期间实现,而无需在用于将条形码附接至样品核酸的引物中包含单独的样品索引序列。在此类情况下,例如,条形码序列还可充当样品索引。具有同时起条形码序列和样品索引作用的序列的测序仪就绪构建体的示例性

配置如下：

[0361] P5-BSI-R1-随机Nmer-插入物-R2-P7

[0362] 其中“BSI”是同时起条形码序列和样品索引作用的序列。

[0363] 测序仪就绪产物可以包含条形码序列，该条形码序列可以用来比对序列读取结果并提供样品核酸的序列。可例如采用PHASE扩增并随后如本文其他各处所述进行批量扩增来生成测序仪就绪产物。此外，条形码序列可属于特定的一组已知条形码序列。这组条形码序列可与特定的样品关联，使得可通过所读取的条形码序列来实现特定测序读取结果所源自的样品的鉴定。每个样品均可与一组已知条形码序列关联，其中每个条形码序列组包含不与和其他样品关联的其他条形码组中的条形码序列重叠的条形码序列。因此，条形码序列的独特性以及其在不同条形码序列组之间的独特性可以用于多重化。

[0364] 例如，测序读取结果可包含条形码序列“GAGCCG”。条形码序列“GAGCCG”可以是与样品A关联的一组已知条形码序列中的条形码序列。没有在与另一样品关联的一组已知条形码序列中发现该序列。因为序列“GAGCCG”对于与样品A关联的这组条形码序列是独特的，所以在读取序列“GAGCCG”时，可以确定该序列读取结果与样品A相关联。此外，另一测序读取结果可包含条形码序列“AGCAGA”。条形码序列“AGCAGA”可以是与样品B关联的一组已知条形码序列中的条形码序列。没有在与另一样品关联的一组已知条形码序列中发现该序列。因为“AGCAGA”对于与样品B关联的这组条形码序列是独特的，所以在读取序列“AGCAGA”时，可以确定该序列读取结果与样品B相关联。

[0365] 在另一个实例中，可将样品索引序列嵌入在一个或多个扩增反应中使用的引物的随机序列中，以将条形码附接至样品核酸。例如，引物可包含条形码序列和随机序列，该序列可用于随机引发样品核酸并将条形码序列附接至样品核酸。在一些情况下，该随机序列可以是伪随机序列，使得该随机序列的特定碱基在所有引物之间是保守的。该保守碱基模式可以用作样品索引，使得从特定样品获得的所有测序仪就绪产物均在随机序列区中包含保守的碱基模式。每个样品均可以与不同的保守碱基模式相关联，并因此可以实现多重化。在一些情况下，该模式是伪随机序列的毗连序列区（例如，“NNNATACNNN”），或在其他情况下，该模式是伪随机序列的非毗连序列区（例如，“NCNGNNAANN”），其中“N”对应于随机碱基。此外，在伪随机序列中任何合适数目的碱基可以以任何模式保守，并且本文所述的实例并非意在限制。具有同时起条形码序列和样品索引作用的序列的测序仪就绪构建体的示例配置如下：

[0366] P5-条形码-R1-NQNQNNQQNN-插入物-R2-P7

[0367] 其中“Q”是随机区域中的保守碱基。

[0368] 例如，测序仪就绪产物可包含10-mer伪随机序列“NCNGNNAANN”，其中对于由样品A生成的全部测序仪就绪产物，所述伪随机序列的第二个碱基（“C”）、第四个碱基（“G”）、第七个碱基（“A”）和第八个碱基（“A”）是保守的。测序读取结果可在随机序列区中包括这样的保守碱基模式。在阅读该保守碱基模式时，可确定该序列读取结果与样品A相关联，因为“NCNGNNAANN”保守碱基模式与样品A相关联。而且，测序仪就绪产物可包含10-mer伪随机序列“NNGCNGNGNN”，其中对于由样品B生成的全部测序仪就绪产物，所述伪随机序列的第三个碱基（“G”）、第四个碱基（“C”）、第六个碱基（“G”）和第八个碱基（“G”）是保守的。测序读取结果可在随机序列区中包括这样的保守碱基模式。在阅读该保守碱基模式时，可确定该序列

读取结果与样品B相关联,因为“NNGCNGNGNN”保守碱基模式与样品B相关联。

[0369] 在其他情况下,可在将条形码序列添加至样品核酸之前将样品索引添加至样品核酸。例如,可批量预扩增样品核酸,使得所得扩增子在条形码化之前附接至样品索引序列。例如,可采用包含样品索引序列的引物扩增样品,使得该样品索引序列可附接至样品核酸。在一些情况下,所述引物可以是随机引物(例如,包括随机N-mer)并且扩增可以是随机的。随后可使用任何合适的方法(包括本文所述的条形码化方法)对产生的包含样品索引的扩增子进行条形码化。

[0370] 可将样品核酸分子与上述引物组合至分区(例如,乳液的小滴)。在一些情况下,每一个分区可包含多个样品核酸分子(例如,较大核酸的较小碎片)。在一些情况下,每一个分区中存在独特样品核酸分子的不多于一个拷贝。在一些情况下,每一个分区通常可包含含有相同条形码序列和样品引发序列(例如,可变随机N-mer、靶向N-mer)的引物,其中分区之间的条形码序列通常不同。在这样的情况下,每一个分区(因此分区中的样品核酸)可与独特的条形码序列关联,并且所述独特的条形码序列可用于确定分区中生成的条形码化的样品核酸的序列。

[0371] 在一些情况下,生成条形码化的样品核酸时,该条形码化的样品核酸可从它们各自的分区中释放、合并并经历批量扩增方案,以添加全部下游测序仪就绪产物共同的附加序列(例如,附加测序引物结合位点、附加测序仪引物结合位点、附加条形码序列、样品索引序列)。在分区为乳液的小滴的情况下,可破乳并合并条形码化的样品核酸。可例如使用本文所述的系列扩增方法向释放的、条形码化的样品核酸中批量添加样品索引。在批量添加样品索引的情况下,由相同的样品生成的每一个测序仪就绪产物将包含相同的样品索引,该样品索引可用于鉴定生成测序仪就绪产物的读取结果的样品。在条形码化期间添加样品索引的情况下,用于条形码化的每一个引物可包含相同的样品索引序列,使得由相同样品生成的每一个测序仪就绪产物将包含相同的样品索引序列。

[0372] 对于每一个样品,可使用针对每一个样品不同的样品索引,重复进行样品核酸的分配以生成条形码化的(或条形码化的且样品索引化的)样品核酸以及随后向条形码化的样品核酸添加附加序列(例如,包括样品索引)。在一些情况下,微流体小滴发生器可用于分配样品核酸。在一些情况下,微流体芯片可包含多个小滴发生器,使得可在每一个小滴发生器处理不同的样品,从而允许平行的样品索引化。经由每一个不同的样品索引,可实现测序期间的多重化。

[0373] 一旦生成测序仪就绪寡核苷酸,随后可将测序仪就绪寡核苷酸提供给测序装置以供测序。因此,例如,向测序装置提供的整个序列可包含一个或多个与测序装置相兼容的衔接子(例如,P5、P7)、一个或多个条形码序列、一个或多个引物结合位点(例如,读取1(R1)序列引物,读取2(R2)测序引物、索引引物)、N-mer序列、通用序列、感兴趣的序列及其组合。条形码序列可位于序列的任一端。在一些情况下,条形码序列可位于P5与读取1序列引物结合位点之间。在其他情况下,条形码序列可位于P7与读取2序列引物结合位点之间。在一些情况下,第二条形码序列可位于P7与读取2序列引物结合位点之间。可在测序装置中使用索引序列引物结合位点以确定条形码序列。

[0374] 有待向测序仪装置提供的序列的各种组分(例如,衔接子、条形码序列、样品索引序列、样品序列、引物结合位点等)的配置可根据例如期望的特定配置和/或序列的各种组

分的添加顺序而变化。可以使用用于测序的任何合适的配置并且可以以任何合适的顺序向寡核苷酸添加任何序列。可在样品核酸的条形码化之前、期间和之后向该样品核酸添加附加序列。例如,可在条形码化期间向样品核酸添加P5序列,并可在该样品核酸的条形码化之后以批量扩增方式添加P7。或者,可在条形码化期间向样品核酸添加P7序列,并可在该样品核酸的条形码化之后以批量扩增方式添加P5序列。本文实例所示的示例性配置并非旨在限制。而且,经由扩增向寡核苷酸添加序列组分也并不意味着限制。还可使用其他方法,例如,连接。此外,本文所述的衔接子、条形码序列、样品索引序列、引物结合位点、测序仪就绪产物等并不意味着限制。可使用本文所述的方法,针对任何合适类型的测序平台(例如,Illumina测序、Life Technologies Ion Torrent、Pacific Biosciences SMRT、Roche 454测序、Life Technologies SOLiD测序等),产生本文所述的任何类型的寡核苷酸,包括测序仪就绪产物。

[0375] 可使用本文所述方法采用适合于特定测序平台的任何衔接子序列生成测序仪就绪寡核苷酸。例如,可使用本文所述方法生成包含一个或多个条形码序列以及在Life Technologies Ion Torrent测序中有用的P1和A衔接子序列的测序仪就绪寡核苷酸。在一个实例中,可生成包含经由二硫键与P1序列连接的acrydite部分的珠子(例如,凝胶珠子)。可生成包含P1序列、条形码序列和随机N-mer序列的条形码构建体。该条形码构建体可进入扩增反应(例如,在分区如流体小滴中)以对样品核酸进行条形码化。条形码化的扩增子随后可经历进一步的批量扩增,以添加A序列和任何其他期望的序列,诸如样品索引。或者,P1和A序列可互换,使得在样品条形码化期间添加A并且批量添加P1。完整序列随后可进入Ion Torrent测序仪。可以以类似方式添加用于其他测序平台的其他衔接子序列(例如,用于Life Technologies SOLiD测序的P1衔接子序列、用于Roche 454的A和B衔接子序列等)。

[0376] 尽管本文描述为生成部分发夹分子,以及在一些情况下,防止形成完整发夹,但在一些情况下可能希望提供包含本文所述的条形码序列的完整发夹片段。特别地,这样的完整发夹分子可以进一步经历常规样品制备步骤,该步骤是通过在常规测序工作流程中处理单个发夹分子的3' 和5' 端,如同双链双链体分子的一个末端。特别地,通过使用常规的连接步骤,可容易地将合适的衔接子序列附接至发夹分子的3' 和5' 端两个末端,其附接方式与它们附接至双链体分子的3' 和5' 末端的方式相同。例如,在基于Illumina的测序过程的情况下,可使用标准的Illumina方案将包括P5和P7衔接子的标准Y衔接子以及R1和R2引物序列附接至发夹的一端,如同它是双链体分子的一个末端。

[0377] 减少不期望的扩增产物的方法(用于测序的部分发夹扩增(PHASE))

[0378] 可使用随机N-mer序列来随机引发样品,诸如基因组DNA(gDNA)。在一些实施方案中,随机N-mer可包括引物。在一些情况下,随机N-mer可引发样品。在一些情况下,随机N-mer可引发基因组DNA。在一些情况下,随机N-mer可引发DNA片段。

[0379] 此外,随机N-mer序列还可附接至另一寡核苷酸。该寡核苷酸可以是通用序列,和/或可含有一个或多个可与测序装置相兼容的引物读取序列(例如,读取1引物位点、读取2引物位点、索引引物位点)、一个或多个条形码序列和一个或多个可与测序装置相兼容的衔接子区段(例如P5、P7)。或者,该寡核苷酸可能不包含这些中的任何一个,并且可能包含另一序列。

[0380] 经由随后的扩增方法,用随机N-mer引发样品核酸可用于将与随机N-mer连接的寡

核苷酸序列(例如,包含条形码序列的寡核苷酸序列)附接至样品核酸,包括待测序的样品核酸。由于例如不期望的扩增产物的产生,利用随机引物引发样品可能引入显著的序列读取错误。

[0381] 为了减少不期望的扩增产物,寡核苷酸序列的至少一个子段可分别用dUTP或含尿嘧啶的核苷酸置换,以替代dTTP或含胸腺嘧啶的核苷酸。在一些情况下,置换可以是完全的(例如,所有含胸腺嘧啶的核苷酸均被置换为含尿嘧啶的核苷酸),或者可以是部分的,使得寡核苷酸的含胸腺嘧啶的核苷酸的一部分被置换为含尿嘧啶的核苷酸。在一些情况下,除了与随机N-mer序列相邻的最后约10至约20个、最后约10至30个、最后约10至40个或最后约5至40个核苷酸,寡核苷酸序列的全部含胸腺嘧啶的核苷酸都被置换为dUTP或含尿嘧啶的核苷酸。此外,可使用不接受或处理含尿嘧啶的模板的聚合酶来扩增样品核酸。在这一情况下,约10至约20个核苷酸的不含尿嘧啶的部分可得到扩增,而剩余的含有dUTP的部分或含尿嘧啶的核苷酸可能不被扩增。在一些情况下,寡核苷酸序列中包含dUTP或含尿嘧啶的核苷酸的部分可能与N-mer序列相邻。在一些情况下,寡核苷酸序列中包含dUTP或含尿嘧啶的核苷酸的部分可与条形码序列相邻。根据寡核苷酸序列的配置,寡核苷酸序列的任何部分(包括衔接子区段、条形码或读取引物序列)可包含dUTP或含尿嘧啶的核苷酸(例如,取代含胸腺嘧啶的核苷酸)。

[0382] 而且,可使用寡核苷酸中含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的置换的数目和定位,来例如调整采用以下所述的扩增方法获得的部分发夹产物的大小和/或调整聚合酶与含尿嘧啶的引物序列的结合。此外,还可在反应混合物内,例如在分区内,以期望的浓度提供游离的含尿嘧啶的核苷酸例如UTP或其类似物,以介导聚合酶/尿嘧啶-引物结合动力学。在一些情况下,较小的部分发夹产物可产生更准确的测序结果。因此,寡核苷酸可包含至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30个或更多个含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的置换,这取决于例如由寡核苷酸生成的部分发夹产物的期望长度。

[0383] 采用与寡核苷酸序列(例如,包含上述含尿嘧啶的核苷酸的寡核苷酸序列)连接的随机N-mer随机引发样品核酸时(图15A),第一轮的扩增(例如,使用不接受或处理作为模板的含尿嘧啶的核苷酸的聚合酶)可导致寡核苷酸序列附接至样品核酸的互补序列,图15B和图15C。采用包含随机N-mer的寡核苷酸序列的另一拷贝引发(经由随机N-mer)和进一步扩增所述扩增产物时(图15D),可生成这样的扩增产物,其包含寡核苷酸序列、样品核酸序列的一部分以及在与寡核苷酸序列相对的扩增产物的一端的部分互补(例如,与不包含含尿嘧啶的核苷酸的寡核苷酸序列的一部分互补)寡核苷酸序列。该部分互补寡核苷酸序列与该寡核苷酸序列可杂交以形成部分发夹,该部分发夹在一些情况下可不再参与核酸扩增。由于包含含尿嘧啶的核苷酸的原始寡核苷酸序列的一部分没有被拷贝,因此可生成部分发夹。扩增可持续所需数目的循环(例如,5、10、15、20、25、30、35、40、45或50个循环),直到包含随机N-mer的所有寡核苷酸序列耗尽(图15E-G)。

[0384] 在一些实施方案中,为了确保仅用随机N-mer而非部分附接的寡核苷酸序列引发样品核酸(例如,基因组DNA(gDNA)),可经由阻断寡核苷酸(例如,图15中的黑色哑铃)的杂交来阻断寡核苷酸序列。阻断寡核苷酸(在本文其他地方也称为寡核苷酸阻断剂)可与寡核苷酸序列的任何部分(包括条形码序列,读取引物位点序列,寡核苷酸的含尿嘧啶部分的全

部或一部分,或寡核苷酸的全部或任何其他部分,或其中的其他序列)杂交。阻断寡核苷酸可以是DNA或RNA。在一些情况下,阻断寡核苷酸可包含含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的置换。在一些情况下,阻断寡核苷酸的全部含胸腺嘧啶的核苷酸可被置换为含尿嘧啶的核苷酸。在一些情况下,阻断寡核苷酸的含胸腺嘧啶的核苷酸的一部分可被置换为含尿嘧啶的核苷酸。在一些情况下,阻断寡核苷酸可包含锁核酸(LNA)、LNA核苷酸、桥核酸(BNA)和/或BNA核苷酸。而且,阻断寡核苷酸可以具有对于阻断剂功能性所必需的任何合适的长度。阻断寡核苷酸可以具有适合于阻断寡核苷酸的一部分的长度,或可以具有与旨在阻断的寡核苷酸相同或基本相同的长度。阻断寡核苷酸可确保仅随机N-mer而非寡核苷酸序列的其他部分与样品核酸(例如,基因组DNA)结合。

[0385] 阻断寡核苷酸与寡核苷酸的化学计量比(例如,阻断寡核苷酸:寡核苷酸)可以变化。例如,阻断寡核苷酸:寡核苷酸的化学计量比可为约0.01、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、1.00、1.05、1.10、1.15、1.20、1.25、1.30、1.35、1.40、1.45、1.50、1.55、1.60、1.65、1.70、1.75、1.80、1.85、1.90、1.95、2.00、2.10、2.20、2.30、2.40、2.50、2.60、2.70、2.80、2.90、3.00、3.50、4.00、4.50、5.0、5.5、6.0、6.5、7.0、7.5、8.0、8.5、9.0、10.0、20、30、40、50、100或更大。在一些情况下,阻断寡核苷酸:寡核苷酸的化学计量比可为至少约0.01、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、1.00、1.05、1.10、1.15、1.20、1.25、1.30、1.35、1.40、1.45、1.50、1.55、1.60、1.65、1.70、1.75、1.80、1.85、1.90、1.95、2.00、2.10、2.20、2.30、2.40、2.50、2.60、2.70、2.80、2.90、3.00、3.50、4.00、4.50、5.0、5.5、6.0、6.5、7.0、7.5、8.0、8.5、9.0、10.0、20、30、40、50、100或更大。在一些情况下,阻断寡核苷酸:寡核苷酸的化学计量比可为至多约0.01、0.05、0.10、0.15、0.20、0.25、0.30、0.35、0.40、0.45、0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90、0.95、1.00、1.05、1.10、1.15、1.20、1.25、1.30、1.35、1.40、1.45、1.50、1.55、1.60、1.65、1.70、1.75、1.80、1.85、1.90、1.95、2.00、2.10、2.20、2.30、2.40、2.50、2.60、2.70、2.80、2.90、3.00、3.50、4.00、4.50、5.0、5.5、6.0、6.5、7.0、7.5、8.0、8.5、9.0、10.0、20、30、40、50或100。

[0386] 而且,阻断剂部分引入(例如,经由在阻断寡核苷酸的3'或5'端的双脱氧核苷酸(ddNTP)、ddCTP、ddATP、ddGTP、ddTTP等)到阻断寡核苷酸中和/或在阻断寡核苷酸中包含含尿嘧啶的核苷酸(例如,取代全部或部分的含胸腺嘧啶的核苷酸)可防止被阻断的寡核苷酸序列的阻断部分与样品核酸优先结合。阻断剂部分的另外的实例包括3'磷酸、阻断的3'端、3' ddCTP、C3间隔区(/3SpC3/)、双脱氧-C(/3ddC/)。可通过RNA酶、RNaseH、反义DNA寡核苷酸和/或碱性磷酸酶从寡核苷酸序列切割阻断寡核苷酸。

[0387] 在一些情况下,可采用阻断寡核苷酸来阻断寡核苷酸序列,使得该寡核苷酸序列包含阻断的5'端,包含阻断的3'端,可被完全阻断(例如,除了其随机N-mer序列之外可被完全阻断),或可在另一位置(例如,寡核苷酸的部分序列,其不同于寡核苷酸序列的随机N-mer)处被阻断。在一些情况下,寡核苷酸序列可包含多个阻断剂,使得寡核苷酸的多个位点被阻断。在一些情况下,寡核苷酸序列可包含阻断的3'端和含尿嘧啶的核苷酸两者。在一些情况下,包含含尿嘧啶的核苷酸和阻断的3'端的寡核苷酸序列可与N-mer序列相邻。在一些情况下,寡核苷酸序列可包含阻断的3'端。在一些情况下,寡核苷酸序列可包含含尿嘧啶的

核苷酸。在一些情况下，寡核苷酸序列可包含阻断的5' 端和含尿嘧啶的核苷酸两者。

[0388] 在一些情况下，包含含尿嘧啶的核苷酸和阻断的3' 端的寡核苷酸序列可与N-mer序列相邻。在一些情况下，包含含尿嘧啶的核苷酸和阻断的3' 端的寡核苷酸序列可与条形码序列相邻。在一些情况下，寡核苷酸序列可包含阻断的3' 端。在一些情况下，寡核苷酸序列可包含含尿嘧啶的核苷酸。在一些情况下，寡核苷酸序列可包含阻断的3' 端和含尿嘧啶的核苷酸两者。阻断寡核苷酸的添加可防止与可能不希望进行扩增的部分通用序列优先结合。

[0389] 在一些情况下，适合于经由其随机N-mer而引发样品核酸的寡核苷酸还可包含阻断序列，该阻断序列可起到与阻断寡核苷酸相同的作用。例如，寡核苷酸可以以发夹配置布置，该发夹配置具有可起到与阻断寡核苷酸相同作用的阻断序列。包含随机N-mer、R1c序列、P5序列、条形码序列和R1序列的示例性寡核苷酸可如下配置：

[0390] 5' -随机Nmer-R1c-P5-条形码-R1-3'

[0391] 寡核苷酸的R1序列和R1c序列可杂交以生成具有包含P5和条形码序列的发夹环的发夹。R1c序列可起到与阻断寡核苷酸相同的作用，使得用寡核苷酸引发样品核酸仅通过寡核苷酸的随机N-mer而发生。在一些情况下，如果需要，一个或多个切割位点(例如，限制位点、切割位点、脱碱基位点等)可包含在布置成具有阻断序列的发夹(包含寡核苷酸的发夹环)的寡核苷酸中，以分离寡核苷酸下游的序列组分。分离可例如经由酶反应、氧化还原、辐射(例如，紫外线)、加热或其他合适的手段而发生。

[0392] 图14B中示出了与随机N-mer连接的含尿嘧啶的核苷酸取代的示例性寡核苷酸序列。特别地，约8N-12N长度的随机引物(例如，随机N-mer)1404可与寡核苷酸序列连接。可使用随机N-mer来随机引发样品核酸并从样品核酸延伸，该样品核酸例如是基因组DNA(gDNA)。寡核苷酸序列包含：(1)用于与测序装置如流通池相兼容的序列(例如，Illumina的P5(1401)和读取1引物位点(1402))，以及(2)条形码(BC) — 1403(例如，6-12个碱基的序列)。此外，寡核苷酸序列的读取1引物位点1402可与包含含尿嘧啶的核苷酸和阻断剂部分的阻断寡核苷酸在其3' 端(例如，3' ddCTP，以“X”表示)杂交。阻断寡核苷酸可用于促进仅用随机N-mer序列引发样品核酸，并防止寡核苷酸序列与样品核酸中与读取1引物位点1402互补的部分优先结合。任选地，为了进一步限制产物长度，小百分比的终止核苷酸(例如，0.1-2%的无环核苷酸(acyNTP))(图16B)可包含在寡核苷酸序列中以减少不期望的扩增产物。

[0393] 在图15中示出了用于将含尿嘧啶的核苷酸取代的、包含随机N-mer的寡核苷酸序列附接至样品核酸(例如，基因组DNA(gDNA))的部分发夹扩增的实例。首先，在变性温度(例如，98°C，持续2分钟)下实现样品核酸的初始变性，然后在引发温度(例如，在4°C下30秒)下用随机N-mer序列引发样品核酸的随机部分，图15A。使寡核苷酸序列与阻断寡核苷酸(图15中的黑色哑铃)杂交，以确保仅随机N-mer而非寡核苷酸序列的另一部分引发样品核酸。随后，随着温度升至更高温度(例如，以0.1°C/秒升至45°C(保持1秒))可进行序列延伸(例如，经由不接受或处理作为模板的含尿嘧啶的核苷酸的聚合酶)(图15A)。随后可在升高的温度(例如，20秒时在70°C)下继续延伸，继续置换上游链并产生第一阶段的冗余(图15B)。随后可在变性温度(例如，98°C持续30秒)下发生扩增产物的变性，以释放样品核酸和扩增产物用于另外的引发。

[0394] 第一循环之后，扩增产物具有包含寡核苷酸序列的单一的5' 尾(图15C)。重复上述

提到的那些步骤以用寡核苷酸序列经由其随机N-mer引发扩增产物和样品核酸。黑色序列表示添加的5'尾的部分(在循环1中添加的),其包含含尿嘧啶的核苷酸,并因此将不在扩增产物的引发和扩增时被拷贝(图15D)。第二轮扩增后,可生成5'标记的产物及3'和5'标记的产物两者(图15E)。3'和5'标记的产物包含在一端的完整的寡核苷酸序列、样品核酸序列以及在寡核苷酸另一端的与寡核苷酸序列部分互补(例如,与不含有含尿嘧啶的核苷酸的寡核苷酸序列区互补)的序列。寡核苷酸序列可与其部分互补序列杂交以生成部分发夹结构(图15F)。扩增可继续重复所需数目的循环(例如,高达20次),直到所有寡核苷酸序列耗尽(图15G)。

[0395] 部分发夹的形成可防止生成拷贝的拷贝,并反而可促使仅产生原始模板的拷贝,从而减少潜在的扩增偏倚和其他伪迹。部分发夹的形成可促使期望产物的分离并可减少拷贝的产生。

[0396] 不读取尿嘧啶的聚合酶的形成部分发夹的期望性质可包括外切核酸酶缺陷的聚合酶(例如,具有低外切核酸酶活性、基本不具有外切核酸酶活性、不具有外切核酸酶活性)、链置换能力(例如,热稳定的链置换聚合酶)、在温度<50°C下的残余活性以及对于含尿嘧啶的核苷酸与含胸腺嘧啶的核苷酸的区别。这样的聚合酶的实例可包括9degrees North、修饰的(NEB)、exo minus Pfu、Deep Vent exo minus、Vent exo minus及其同系物。而且,具有低外切核酸酶活性的聚合酶可以是这样的聚合酶,其外切核酸酶活性小于具有正常外切核酸酶活性的热稳定聚合酶(例如,Taq聚合酶)的外切核酸酶活性的90%、小于80%、小于70%、小于60%、小于50%、小于40%、小于30%、小于20%、小于10%、小于5%或0%。在一些情况下,用于部分发夹扩增的聚合酶可以是能够链置换的。在一些情况下,限制扩增序列的长度可减少不期望的扩增产物,其中较长长度的产物可包括不期望的上游部分如条形码序列。可通过包含终止核苷酸来限制扩增产物的长度。终止核苷酸的实例可包括无环核苷酸(acyNTP)。终止核苷酸可以以扩增产物长度的约0%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、2.1%、2.2%、2.3%、2.4%、2.5%或更大的百分比而存在。在一些情况下,终止核苷酸可以以大于扩增产物长度的约0%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、2.1%、2.2%、2.3%、2.4%、2.5%或更大的百分比而存在。在一些情况下,终止核苷酸可以以小于扩增产物长度的约0%、0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、1.1%、1.2%、1.3%、1.4%、1.5%、1.6%、1.7%、1.8%、1.9%、2%、2.1%、2.2%、2.3%、2.4%、2.5%或更大的百分比而存在。

[0397] 也可通过在开始PHASE扩增之前预扩增样品核酸来控制扩增产物的长度。例如,随机N-mer可用于样品核酸的预扩增。随机N-mer可用于引发样品核酸,之后使用合适的热循环条件延伸引物。除了随机引发样品核酸外还可通过热循环条件(例如,热循环数、所用温度、循环时间、总运行时间等)来控制产物长度。在一些情况下,可获得小于原始样品核酸的预扩增产物。预扩增期间生成的扩增产物可随后进入PHASE扩增并如上所述进行条形码化。

[0398] 如图17所示,阻断寡核苷酸的加入可将起始位点偏倚减少50%。向通用序列中引入含尿嘧啶的核苷酸而不是含胸腺嘧啶的核苷酸以及使用不接受或处理含尿嘧啶的模板的聚合酶可显著减少测序错误,如图21和图22所示。例如,Q40错误可从约0.002减少至约

0.001,未定位的末端的分数可从约0.996减少至约0.03,插入物大小的中值可从约399减少至约310,IQR插入物大小可从约413减少至约209,并且零覆盖范围分数可从约0.9242减少至约0.0093。

[0399] 为了生成部分发夹物质,还设想了不包括用含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的扩增方案。在一些情况下,可使用不能被聚合酶识别或拷贝的其他物质(例如,甲基化的碱基、脱碱基位点、与大体积侧基连接的碱基等)代替含尿嘧啶的核苷酸,以生成部分发夹扩增子。在一些情况下,可生成完整的发夹扩增子并在合成后进行处理以生成部分发夹物质。在一些情况下,可生成完整的发夹扩增子并随后去除部分以生成部分发夹物质。例如,如图34A所示,当包含随机N-mer的寡核苷酸引物不包含含尿嘧啶的核苷酸和/或能够接收或处理含尿嘧啶的模板的聚合酶用于扩增时,可经由图15所示的扩增方案生成完整的发夹扩增子3401。生成完整的发夹扩增子3401时,该完整的发夹扩增子可在一个或多个合适的位点进行酶学(例如,经由限制性内切酶或其他位点特异性酶如切口酶)或化学切割3403,以生成部分发夹物质3402。

[0400] 在一些情况下,可生成完整的发夹扩增子并且可将部分添加至该完整的发夹扩增予以生成部分发夹物质。例如,包含与随机N-mer偶联的测序引物结合位点(例如,R1)而不包含含尿嘧啶的核苷酸的引物可用于扩增样品核酸,并经由图15所示的扩增方案生成完整的发夹扩增子(例如,完整的发夹—图34B中的3404,其包含测序引物结合位点(例如,R1)、拷贝的样品核酸以及与测序引物结合位点(例如,R1c)杂交的测序引物结合位点的互补序列)。生成完整的发夹扩增子3404时,该完整的发夹扩增子可具有例如经由连接3406而添加的附加序列(例如,包含P5序列和条形码序列的序列)3405。

[0401] 在一些情况下,可以共价修饰用于生成完整的发夹扩增子的引物(例如,包含随机N-mer的寡核苷酸),以经由例如连接体(例如,不包含核酸的连接体或包含不参与扩增的核酸的连接体)包含附加序列。在一些情况下,该连接体可以是聚乙二醇或碳基连接体。因此,由引物生成(例如,经由图15所示的扩增方案)的完整的发夹扩增子也可经由该连接体与附加序列共价连接。附接的序列随后可与完整的发夹扩增子连接以生成部分发夹物质。图34C示出了经由连接体3407包含附加序列3408的完整的发夹扩增子3409的一个实例。完整的发夹生成后,附加序列3408可与完整的发夹扩增子3409连接,使得可生成包含附加序列3408的部分发夹物质(3410)。

[0402] 靶向N-mer和靶向扩增

[0403] 除了随机扩增方案之外,还设想了包含靶向引发序列(例如,靶向N-mer)的条形码构建体(例如,包含条形码序列和用于引发样品核酸的N-mer的寡核苷酸)和靶向扩增方案。靶向扩增方案可能在例如经由测序方法检测特定基因或感兴趣的序列方面是有用的,可能在检测特定类型的核酸方面是有用的,可能在检测包含序列的核酸的特定链方面是有用的,以及上述的组合。通常,靶向扩增方案依赖靶向引物来完成特定核酸序列的扩增。在一些实例中,通过使用针对感兴趣的特定基因序列或感兴趣的特定基因序列上游的特定序列的引物,PCR方法可用于靶向扩增,使得感兴趣的特定基因序列在PCR期间得到扩增。

[0404] 也可以修改上述PHASE扩增反应,使得样品核酸的靶扩增得以实现。如上所述,包含靶向引发序列(例如,靶向N-mer)而非随机序列(例如,随机N-mer)的条形码构建体可用于在PHASE扩增期间引发特异性序列。该特异性序列例如可以是感兴趣的特定基因序列,使

得扩增子的生成指示该序列的存在。或者，该特异性序列可以是感兴趣的特定基因序列上游的已知序列。可以生成这样的构建体，并且如果需要，使用本文所述的任何方法将该构建体与珠子偶联，该方法包括图4所示的有限稀释方案和本文其他各处所述的组合平板方案。

[0405] 例如，如前文关于图4所述，可生成包含引物403(例如，P5)、条形码序列408和读取引物结合位点(例如R1)415的构建体(参见图4A-4H)。如图4I所示，可经由包含与读取引物结合位点415互补的序列412的引物将附加序列413添加(任选地批量添加)至所述构建体。序列413可充当靶向序列(例如，靶向N-mer)，使得靶向序列对应于感兴趣的特定靶序列。如本文其他各处所述，所述构建体还可包含寡核苷酸阻断剂，以便确保仅靶向序列而非该构建体的其他序列部分引发样品核酸。在完成的构建体进入与样品核酸的PHASE反应时，例如，靶向构建体可引发样品核酸(例如，在期望的序列位点)，并且可启动扩增反应以由如上所述的样品核酸生成部分发夹。在一些情况下，使用靶向N-mer引物和随机N-mer引物的组合来生成部分发夹扩增子。在一些情况下，靶向扩增在控制针对特定靶标的扩增期间生成的部分发夹扩增子的大小(例如，序列长度)方面可能是有用的。

[0406] 在一些情况下，多个包含条形码序列和靶向N-mer的构建体可与珠子(例如，凝胶珠子)偶联。在一些情况下，所述多个构建体可包含相同的条形码序列和/或相同的靶向N-mer序列。在一些情况下，靶向N-mer序列可在所述多个构建体的单独的构建体之间变化，使得样品核酸上的多个靶序列可经由各种靶向N-mer而引发。如上所述，可将珠子与样品核酸一起分配(例如，在流体小滴中)，降解每一个分区中的珠子以将偶联的构建体释放至分区，并且经由该构建体的靶向N-mer扩增样品核酸。可采用本文所述的任何方法来实现生成的扩增子的后处理(例如，添加附加序列(例如，P7、R2)、添加样品索引等)，该方法包括批量扩增方法(例如，批量PCR)和批量连接。

[0407] 在分区中，包含条形码序列和靶向N-mer的构建体可与珠子偶联，可在溶液中游离(例如，在流体小滴的水性内部游离)，或两者兼具。而且，分区可包含靶向构建体(例如，包含靶向N-mer序列的构建体)和非靶向构建体(例如，包含随机N-mer序列的构建体)两者。靶向和非靶向构建体中的每一个均可与珠子偶联，两者之一可与珠子偶联，且任一构建体还可在分区内的溶液中。

[0408] 在一个分区中存在每一种类型的构建体的情况下，样品核酸的靶向和非靶向扩增均可发生。例如，关于PHASE扩增反应，可使用靶向条形码构建体来初始引发和延伸样品核酸。通常，这些步骤对应于上文关于图15A-C所述的PHASE扩增的第一个循环，不同之处在于靶向构建体用于初始引发。随后可采用包含随机N-mer的条形码构建体来引发延伸产物，使得生成部分发夹，这些步骤对应于上文关于图15D-F所述的PHASE的第二个循环。扩增可持续另外数轮(例如，图15G)直到完成所需轮数的循环。可采用本文所述的任何方法来实现生成的部分发夹扩增子的后处理(例如，添加附加序列(例如，P7、R2)、添加样品索引等)，该方法包括批量扩增方法(例如，批量PCR)和批量连接。

[0409] 而且，可生成靶向条形码构建体，使得构建体的靶向N-mer针对除DNA之外的核酸物质，例如，RNA物质。在一些情况下，靶向条形码构建体的靶向N-mer可针对特定的RNA序列，例如对应于转录的基因的序列或信使RNA(mRNA)转录物上的其他序列。在一些情况下，对由RNA(例如，mRNA)生成的条形码化的产物的测序可帮助确定由RNA转录的基因的表达水平。在一些情况下，靶向N-mer可以是能够与聚腺嘌呤(聚A(poly-A)序列)杂交的聚胸腺嘧

啶(例如,聚T(poly-T)序列)序列,其可见于例如mRNA转录物的3'端。在采用包含聚T序列的靶向条形码构建体经由条形码构建体的聚T序列与mRNA的聚A序列杂交引发mRNA时,靶向条形码构建体可经由逆转录反应延伸,以生成包含该条形码构建体的互补DNA(cDNA)产物。在一些情况下,包含聚T靶向N-mer的靶向条形码构建体还可包含如本文其他各处所述的寡核苷酸阻断剂,使得仅聚T序列与RNA杂交。

[0410] 针对RNA物质的靶向条形码构建体也可能在经由例如PHASE扩增反应生成部分发夹扩增子方面是有用的。例如,包含聚T序列的靶向条形码构建体可与mRNA经由其聚A序列杂交。靶向条形码构建体可经由逆转录反应(例如,经由逆转录酶的作用)延伸,使得生成包含条形码构建体的cDNA。这些步骤可对应于上文关于图15A-C所述的PHASE扩增的第一个循环,不同之处在于逆转录用于生成延伸产物。逆转录(例如,第一个PHASE循环)之后,包含随机N-mer的条形码构建体可引发延伸产物,使得如上文关于图15D-F所述生成部分发夹。扩增可持续另外数轮(例如,图15G)直到完成所需轮数的循环。

[0411] 在一些情况下,多个包含条形码序列和靶向N-mer(其包含聚T序列)的靶向构建体可与珠子(例如,凝胶珠子)偶联。在一些情况下,所述多个构建体可包含相同的条形码序列。可将珠子与包含RNA的样品核酸一起分配(例如,在流体小滴中),降解每一个分区中的珠子以将偶联的构建体释放至分区,并且经由构建体的靶向N-mer捕获样品RNA。分区还可包含条形码构建体(例如,具有与靶向构建体相同的条形码序列),该条形码构建体包含随机N-mer。在第一个扩增循环中,可经由每一个分区内的逆转录发生靶向构建体的延伸,以生成包含靶向构建体的延伸产物。随后可采用包含随机N-mer的条形码构建体来引发每一个分区中的延伸产物,以生成部分发夹扩增子,如上文关于图15-A-G所述。可采用本文所述的任何方法来实现生成的扩增子的后处理(例如,添加附加序列(例如,P7、R2)、添加样品索引等),该方法包括批量扩增方法(例如,批量PCR)和批量连接。

[0412] 在一些情况下,样品中RNA的逆转录也可在不使用靶向条形码构建体的情况下使用。例如,包含RNA的样品核酸可首先经历使用其他类型的逆转录引物的逆转录反应,使得由RNA生成cDNA。生成的cDNA随后可经历如本文所述的靶向或非靶向扩增。例如,包含RNA的样品核酸可经历逆转录反应,使得由RNA生成cDNA。该cDNA随后可进入PHASE扩增反应(使用如上文关于图15A-G所述的具有随机N-mer的条形码构建体),以生成包含构建体的条形码序列的部分发夹扩增子。可采用本文所述的任何方法来实现生成的部分发夹扩增子的后处理(例如,添加附加序列(例如,P7、R2)、添加样品索引等),该方法包括批量扩增方法(例如,批量PCR)和批量连接。

[0413] 也可生成针对核酸的特定链上的特定序列(例如,基因序列)的靶向条形码构建体,使得保留针对每一条链生成的测序仪就绪产物的链型信息。例如,样品核酸可包括双链核酸(例如,双链DNA),使得核酸的每一条链包含一个或多个不同的靶基因序列。由于每一条链的相反的5'到3'方向性和/或碱基组成,因此互补DNA链可包含不同的基因序列。对于每一条链(基于链的5'到3'方向性)可基于条形码构建体的靶向N-mer和配置来生成靶向条形码构建体。图28A中示出了针对双链样品核酸的正向和反向链的靶向条形码构建体的示例组。

[0414] 图28A中示出了靶向条形码构建体的示例组2801和2802,其各自靶向双链样品核酸的正向(2801)链和反向(2802)链中的任一个。组2801包含靶向条形码构建体2803和

2804,其包含P5序列、条形码序列以及第一靶序列(2803)或第二靶序列(2804)中的任一个的靶向N-mer。组2802包含靶向条形码构建体2805和2806,其包含P5序列、条形码序列以及第一靶序列(2805)和第二靶序列(2806)中的任一个的靶向N-mer。每一个构建体还可包含条形码与靶向N-mer之间的任何附加序列(由图28A所示的每一个构建体中的箭头表示)。

[0415] 组2801中的条形码构建体被配置为引发它们各自在双链样品核酸的正向链上的靶序列。组2802的条形码构建体被配置为引发它们各自在双链样品核酸的反向链上的靶序列。如图所示,每一个组中的靶向条形码构建体以相反方向性进行配置,该相反方向性对应于双链样品核酸的正向链和反向链的相反方向性。每一个条形码构建体可引发其各自样品核酸链上各自的靶序列,以经由扩增反应如本文所述的任何扩增反应生成条形码化的扩增子。

[0416] 可使用本文所述的扩增方法将附加序列添加至条形码化的扩增子,该方法包括批量扩增、批量连接或其组合。图28B中示出了可用于将样品索引和P7序列添加至由图28A中的靶向条形码构建体生成的扩增子的示例引物组。引物组2808对应于靶向条形码构建体组2801(例如,靶向条形码构建体2803对应于引物2811,靶向条形码构建体2804对应于引物2812),而引物组2808对应于靶向条形码构建体组2801(例如,靶向条形码构建体2505对应于引物2809,靶向条形码构建体2806对应于引物2810)。每一个引物均可引发其各自链上其各自的靶序列,并且以与本文别处所述的批量扩增方法类似的方式启动批量扩增(例如,批量PCR)以生成包含P7和样品索引序列的测序仪就绪构建体。基于每一个测序仪就绪构建体的各种组分(例如,P5、条形码、靶向N-mer、样品插入物等)的配置和方向性,可确定/保留生成测序仪就绪产物的链。

[0417] 可生成针对双链核酸的正向链和反向链两者的条形码构建体(例如,靶向条形码构建体)的文库。例如,可使用本文所述方法生成包含靶向条形码构建体的珠子(例如,凝胶珠子)的两个文库,使得一个文库包含针对样品核酸的正向链的靶向条形码构建体,而另一个文库包含针对样品核酸的反向链的靶向条形码构建体。在一些情况下,每一个文库可包含珠子,每一个珠子包含相同的靶向N-mer。在一些情况下,每一个文库可包含两组或更多组珠子,且组中的每一个珠子包含相同的靶向N-mer(例如,靶向特定基因的靶向N-mer),而不同的组包含不同的靶向N-mer。在一些情况下,这两个文库可以合并,使得生成正向链和反向链珠子的文库。

[0418] 例如,文库可包含两种类型的正向链珠子和两种类型的反向链珠子,总共四种类型的珠子。文库中的每一个珠子可包含独特的条形码序列。一种类型的正向链珠子和一种类型的反向链珠子可包含对应于靶序列(例如,靶基因序列)的靶向N-mer。例如,一种类型的正向链珠子可包含如图28A中2803所示的靶向条形码构建体,而一种类型的反向链珠子可包含如图28A中2805所示的靶向条形码构建体。类似地,第二种类型的正向链珠子可包含如图28A中2804所示的靶向条形码构建体,而一种类型的反向链珠子可包含如图28A中2806所示的靶向条形码构建体。

[0419] 可将包含正向链和反向链珠子(例如,凝胶珠子)的条形码文库分配至如本文别处所述的条形码样品核酸,其中每一个珠子包含独特的条形码序列。例如,可将上述两种类型的正向链和两种类型的反向链珠子的混合文库与样品核酸(例如,基因组DNA)以及任何其他所需的试剂(例如,扩增样品核酸所必需的试剂、还原剂)一起分配。分区可以是,例如,流

体小滴,如乳液的小滴。通常,每一个分区可包含与包含独特条形码序列和靶向N-mer的靶向条形码构建体偶联的珠子(例如,正向链珠子或反向链珠子)。但在一些情况下,所述分区中的一个或多个可包含相同类型或不同类型的多个珠子。可将靶向条形码构建体从分区中的珠子释放(例如,经由珠子的降解—例如,在珠子是包含二硫键的凝胶珠子的情况下经由还原剂),并使靶向条形码构建体引发其各自的样品核酸链(例如,正向链或反向链)上的靶序列。

[0420] 第一产物链合成可经由杂交的靶向条形码构建体的延伸,经由例如样品核酸的线性扩增在每一个分区中发生。例如,采用靶向条形码构建体进行的附加轮次的样品核酸的线性扩增可用于生成第一产物链的附加拷贝。随后可将第一产物链从分区中移出(例如,在分区是乳液的小滴的情况下,可破乳以释放第一产物)并合并。可洗涤第一产物以去除靶向条形码构建体和任何其他的废产物。在一些情况下,可完成任选的双链消化,以消化样品核酸并将其从第一产物链中去除。

[0421] 接着,可使第一产物链经历批量扩增,以将附加序列(例如,P7、样品索引等)添加至第一产物链,从而导致生成第二产物链。批量扩增反应混合物可包含多个引物,所述多个引物中对应于一种珠子类型(因此,靶向条形码构建体的类型)的每一个引物用于生成第一产物链。对于上述包含两种类型的正向链珠子和两种类型的反向链珠子的示例文库,图28B中以2809、2810、2811和2812示出的引物可用于将附加的样品索引和P7序列添加至第一产物链,该第一产物链分别由靶向条形码构建体2803、2804、2805和2806经由批量扩增生成。随后可洗涤第二产物链以将引物从反应混合物中去除。新鲜的引物(例如,对于上述实例,包含P5和P7的引物)随后可增加另外一轮或多轮扩增(例如,经由PCR)以生成最终的测序仪就绪产物。因此,最终产物可包含原始靶向条形码构建体、扩增的样品核酸的链以及添加至第一产物链的附加序列(例如,P7、样品索引)。

[0422] 本文所述方法在全基因组扩增中可能是有用的。在全基因组扩增的一些实施方案中,随机引物(例如,随机N-mer序列)可与基因组核酸杂交。该随机引物可以是较大寡核苷酸的组分,该较大寡核苷酸还可包含通用核酸序列(包括本文所述的任何类型的通用核酸序列)和核酸条形码序列。在一些情况下,该通用核酸序列可包含一个或多个含尿嘧啶的核苷酸。而且,在一些情况下,该通用核酸序列可包含具有至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个不包含尿嘧啶的核苷酸的区段。该随机引物可延伸(例如,在引物延伸反应中或任何其他合适类型的核酸扩增反应中)以形成扩增产物。

[0423] 如本文别处所述,扩增产物可经历分子内杂交反应以形成发夹分子,例如,部分发夹分子。在一些情况下,全基因组扩增可在可能包含或可能不包含阻断剂部分(例如,C3间隔区(/3SpC3/)、双脱氧-C(/3ddC/)、3'磷酸或本文别处所述的任何其他类型的阻断剂部分)的寡核苷酸阻断剂(本文别处也称为阻断寡核苷酸)的存在下发生。此外,寡核苷酸阻断剂可以能够与通用核酸序列的至少一部分或包含随机引物的寡核苷酸的任何其他部分杂交。

[0424] 在全基因组扩增的一些实施方案中,可将基因组组分(例如,染色体、基因组核酸如基因组DNA、生物体的整个基因组或本文所述的任何其他类型的基因组组分)片段化成多个第一片段。可将第一片段与多个寡核苷酸共同分配至多个分区。每一个分区中的寡核苷酸可包含引物序列(包括本文别处所述的引物序列的类型)和共同序列(例如,条形码序

列)。每一个分区中的引物序列随后可与每一个分区内的第一片段的多个不同区域退火。该引物序列随后可沿着第一片段延伸以在所述多个分区中的每一个分区内产生扩增的第一片段。所述分区内的扩增的第一片段可包含基因组组分的任何合适的覆盖范围(如本文别处所述)。在一些情况下,所述分区内的扩增的第一片段可包含基因组组分的至少1X覆盖范围、至少2X覆盖范围、至少5X覆盖范围、至少10X覆盖范围、至少20X覆盖范围、至少40X覆盖范围或更大的覆盖范围。

[0425] VII. 数字处理器

[0426] 本公开内容的方法、组合物、装置和试剂盒可与任何合适的处理器、数字处理器或计算机一起使用。可将数字处理器编程,例如,以操作装置的任何部件和/或执行本文所述的方法。该数字处理器可以能够通过计算机网络例如因特网发送或接收电子信号和/或与远程计算机通信。一个或多个外围设备如屏幕显示器、打印机、存储器、数据存储和/或电子显示适配器可以与该数字处理器通信。一个或多个输入设备如键盘、鼠标或控制杆可与该数字处理器通信。该数字处理器还可与探测器通信,以使得探测器在期望的或其他预定的时间点或在由从预处理单元或其他装置接收的反馈所确定的时间点执行测量。

[0427] 图18中示出了示例控制组件的概念示意图。计算机充当控制组件的中央集线器(hub)。该计算机与显示器、一个或多个输入设备(例如,鼠标、键盘、照相机等)和任选的打印机通信。该控制组件经由其计算机与以下的一个或多个装置通信:任选的样品预处理单元、一个或多个样品处理单元(诸如测序仪、热循环仪或微流体装置)和任选的探测器。该控制组件可经由例如以太网连接来联网。用户可使用输入设备向计算机提供输入(例如,对于期望的一组核酸扩增反应所必需的参数或微流体装置的流速)。该输入由计算机解释以生成指令。该计算机将这样的指令传递至任选的样品预处理单元,所述一个或多个样品处理单元和/或任选的探测器以供执行。

[0428] 而且,在任选的样品预处理单元、一个或多个样品处理单元和/或任选的探测器的运行过程中,每一个装置可将信号回传至计算机。这样的信号可由计算机解释并使用,以确定任何装置是否需要进一步的指令。该计算机也可调节样品预处理单元,使得样品的组分适当地混合并以期望的或以其他方式预定的速率进料至样品处理单元(诸如微流体装置)。

[0429] 计算机还可与探测器通信,以使得该探测器在期望的或以其他方式预定的时间点或在由从预处理单元或样品处理单元接收的反馈所确定的时间点执行测量。该探测器还可将测量过程中获得的原始数据回传至计算机,以供进一步分析和解释。

[0430] 分析可经由显示器和/或由打印机生成的打印输出以对终端用户有用的格式进行汇总。用于控制样品预处理单元、样品处理单元和/或探测器的指令或程序,通过执行本文所述的任何方法而获得的数据,或经分析和/或解释的数据,可经由网络(例如可以是因特网)传输至一个或多个远程计算机或从一个或多个远程计算机接收。

[0431] 在一些实施方案中,珠子形成方法可在与小滴发生器通信的数字处理器的帮助下执行。该数字处理器可控制小滴形成的速度或控制生成的小滴总数。在一些实施方案中,将样品附接至条形码化珠子的方法可在与微流体装置通信的数字处理器的帮助下执行。具体地,该数字处理器可控制向输入通道中注入的样品和/或珠子的体积量,并且还可控制该通道内的流速。在一些实施方案中,附接寡核苷酸、引物等的方法可在与热循环仪或其他可编程的加热元件通信的数字处理器的帮助下执行。具体地,该数字处理器可在连接或扩增过

程中控制循环的时间和温度。在一些实施方案中，样品测序方法可在与测序装置通信的数字处理器的帮助下执行。

[0432] VIII. 试剂盒

[0433] 在一些情况下，本公开内容提供了一种试剂盒，其包含微流体装置、多个条形码化的珠子以及关于使用该微流体装置和将条形码化的珠子与客户样品组合以创建含有条形码化珠子和客户样品的流体小滴的说明。如贯穿本公开内容所详细说明的，任何合适的样品均可并入该流体小滴。如贯穿本公开内容所述，珠子可被设计为可降解的或不可降解的。在这种情况下，该试剂盒可包含或可不包含用于珠子降解的还原剂。

[0434] 在一些情况下，本公开内容提供了一种试剂盒，其包含多个条形码化的珠子、合适的扩增试剂(例如，任选地包括聚合酶、核苷三磷酸或其类似物、引物序列、缓冲液等中的一种或多种)以及关于将条形码化珠子与客户样品组合的说明。如贯穿本公开内容所详细说明的，可使用任何合适的样品。如贯穿本公开内容所详细说明的，所述扩增试剂可包括将不接受或处理含尿嘧啶的模板的聚合酶。本公开内容的试剂盒还可提供用来形成乳液的试剂，包括油和表面活性剂。

[0435] IX. 应用

[0436] 样品材料的条形码化

[0437] 本文所述的方法、组合物和系统对于将条形码、尤其是核酸条形码序列附接至样品材料和这些样品材料的组分是特别有用的。通常，这通过将样品材料组分分配至具有共同分配的多个条形码(该条形码随后附接至相同分区内的样品组分)的单独的分区或反应体积中来完成。

[0438] 在示例性过程中，提供包含多个寡核苷酸(例如，核酸条形码分子)的第一分区，所述多个寡核苷酸各自包含共同的核酸条形码序列。第一分区可包含多种便携式分区中的任何分区，例如，珠子(例如，可降解的珠子、凝胶珠子)、小滴(例如，乳液中的水性小滴)、微胶囊等，寡核苷酸与该分区可释放地附接、可释放地偶联或可释放地关联。而且，第一分区中可包含任何合适数目的寡核苷酸，包括本文别处所述的每一分区的寡核苷酸数。例如，寡核苷酸可经由可切割的联接如化学可切割的联接(例如，二硫键或本文所述的任何其他类型的化学可切割的联接)、光可切割的联接和/或热可切割的联接与第一分区可释放地附接、可释放地偶联或可释放地关联。在一些情况下，所述第一分区可以是珠子，并且该珠子可以是可降解的珠子(例如，光可降解的珠子、化学可降解的珠子、热可降解的珠子或本文别处所述的任何其他类型的可降解的珠子)。而且，该珠子可包含如本文别处所述的化学可切割的交联(例如，二硫键交联)。

[0439] 随后将第一分区与样品材料、样品材料组分、样品材料的片段或样品材料组分的片段共同分配至第二分区。样品材料(或其组分或片段)可以是任何合适的样品类型，包括本文别处所述的示例样品类型。在样品材料或样品材料的组分包含一个或多个核酸片段的情况下，所述一个或多个核酸片段可具有任何合适的长度，包括例如本文别处所述的核酸片段的长度。所述第二分区可包括多种分区中的任何分区，包括例如孔、微孔、纳米孔、管或容器，或者优选情况下的小滴(例如，乳液中的水性小滴)或微胶囊，在该分区中第一分区可被共同分配。在一些情况下，第一分区可在第一水性流体中提供，并且样品材料、样品材料组分或样品材料组分的片段可在第二水性流体中提供。在共同分配过程中，第一水性流体

和第二水性流体可在不混溶流体内的小滴内组合。在一些情况下，第二分区可包含不多于一个第一分区。在其他情况下，第二分区可包含不多于一个、两个、三个、四个、五个、六个、七个、八个、九个或十个第一分区。在其他情况下，第二分区可包含至少一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个第一分区。

[0440] 一旦共同分配，包含条形码序列的寡核苷酸可从第一分区中释放至第二分区（例如，经由第一分区的降解、切割寡核苷酸与第一分区之间的化学联接或任何其他合适类型的释放，包括本文别处所述的释放类型），并附接至与其共同分配的样品组分。在一些情况下，第一分区可包含珠子，并且珠子的交联可包括二硫键。此外，或者作为备选方案，寡核苷酸可经由二硫键与珠子连接。在任一情况下，可通过将第一分区暴露于还原剂（例如，DTT、TCEP或本文别处所述的任何其他示例性的还原剂）来使寡核苷酸从第一分区中释放。

[0441] 如本文别处所述，将条形码附接至样品组分包括例如通过连接、杂交或其他关联将条形码寡核苷酸直接附接至样品材料。此外，在许多情况下，例如，在核酸样品材料（例如，模板核酸序列、模板核酸分子）、其组分或片段的条形码化中，这样的附接可另外包括使用含条形码的寡核苷酸，该寡核苷酸还包含引发序列。该引发序列可与核酸样品材料的至少一部分互补，并可沿着核酸样品材料延伸以产生此类样品材料的互补序列以及这些序列或其互补序列的至少部分扩增产物。

[0442] 在另一个示例性过程中，可提供包含多个不同的核酸条形码序列的多个第一分区。每一个第一分区可包含多个核酸条形码分子，该核酸条形码分子具有与之关联的相同的核酸条形码序列。任何合适数目的核酸条形码分子可与每一个第一分区关联，包括本文别处所述的每一分区的核酸条形码分子的数目。第一分区可包含任何合适数目的不同核酸条形码序列，包括例如至少约2、10、100、500、1000、5000、10000、50000、100000、500000、1000000、5000000、10000000、50000000或1000000000个或更多个不同的核酸条形码序列。

[0443] 在一些情况下，所述多个第一分区可包含多个不同的第一分区，其中每一个不同的第一分区包含多个可释放地附接、可释放地偶联或可释放地关联的包含共同条形码序列的寡核苷酸，且与每一个不同的第一分区关联的寡核苷酸包含不同的条形码序列。不同的第一分区的数目可以是，例如，至少约2、10、100、500、1000、5000、10000、50000、100000、500000、1000000、5000000、10000000、50000000或1000000000个或更多个不同的第一分区。

[0444] 可将第一分区与样品材料、样品材料的片段、样品材料的组分或样品材料的组分的片段共同分配至多个第二分区。在一些情况下，第二分区的子集可包含相同的核酸条形码序列。例如，至少约1%、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%或更多的第二分区可包含相同的核酸条形码序列。而且，每个第二分区的第一分区的分布还可根据例如本文别处所述的占有率为变化。在多个第一分区包含多个不同的第一分区的情况下，可将每一个不同的第一分区安置在单独的第二分区内。

[0445] 共同分配之后，可将与第一分区关联的核酸条形码分子释放至多个第二分区。所释放的核酸条形码分子随后可附接至第二分区内的样品材料、样品材料组分、样品材料的片段或样品材料组分的片段。在条形码化的核酸物质（例如，条形码化的样品核酸、条形码化的模板核酸、一个或多个模板核酸序列的条形码化的片段等）的情况下，可如本文别处所述对条形码化的核酸物质进行测序。

[0446] 在另一个示例性过程中,可提供可激活的核酸条形码序列,并且可将它与一个或多个样品材料、样品材料的组分、样品材料的片段或样品材料的组分的片段分配至第一分区。在第一分区中,可将可激活的核酸条形码序列激活,以产生活性核酸条形码序列。该活性核酸条形码序列随后可附接至一个或多个样品材料、样品材料的组分、样品材料的片段或样品材料的组分的片段。

[0447] 在一些情况下,所述可激活的核酸条形码序列可与第二分区偶联,该第二分区也被分配在具有可激活的核酸条形码序列的第一分区中。如本文别处所述,可通过将可激活的核酸条形码序列从关联的分区(例如,珠子)中释放,来激活可激活的核酸条形码序列。因此,在可激活的核酸条形码序列与被分配在第一分区(例如,流体小滴)中的第二分区(例如,珠子)关联的情况下,可通过将可激活的核酸条形码序列从其所关联的第二分区中释放,来激活所述可激活的核酸条形码序列。此外,或者作为备选方案,也可通过从可激活的核酸条形码序列中去除可去除的阻断基团或保护基团,来激活可激活的条形码。

[0448] 在另一个示例性过程中,核酸样品可与条形码化的珠子(包括本文别处所述的珠子类型)的文库组合以形成混合物。在一些情况下,珠子的条形码除了条形码序列之外还可各自包含一个或多个附加序列,例如,通用序列和/或功能序列(例如,如本文别处所述的随机N-mer或靶向N-mer)。可将该混合物分配至多个分区,其中分区的至少一个子集包含至多一个条形码化的珠子。在所述分区中,可使用任何合适的途径(包括本文所述的释放类型)使条形码从珠子释放。条形码化的珠子的文库可经由任何合适的途径生成,包括使用本文别处所述的方法和组合物。在一些情况下,如本文别处所述,核酸样品可在微流体装置的帮助下与条形码化珠子的文库和/或所得到的经分配的混合物组合。在释放的条形码也包含引物序列(例如,如本文别处所述的靶向N-mer或随机N-mer)的情况下,条形码的引物序列可与样品核酸杂交,并且如果需要,扩增反应可在分区中完成。

[0449] 多核苷酸测序

[0450] 通常,本文提供的方法和组合物可用于制备寡核苷酸片段以用于下游应用如测序。特别是,这些方法、组合物和系统在测序文库的制备中是有用的。测序可以通过任何可用的技术进行。例如,测序可以通过经典Sanger测序方法进行。测序方法还可以包括:高通量测序、焦磷酸测序、连接测序、合成测序、杂交测序、RNA-Seq(Illumina)、数字基因表达(Helicos)、下一代测序、单分子合成测序(SMSS)((Helicos)、大规模平行测序、克隆单分子阵列(Solexa)、鸟枪法测序、Maxim-Gilbert测序、引物步移法以及本领域中已知的任何其他测序方法。

[0451] 例如,可通过提供多个靶核酸序列和将靶核酸序列分离至多个单独的分区来对多个靶核酸序列进行测序。每一个单独的分区可包含一个或多个靶核酸序列和多个寡核苷酸。单独的分区可包含任何合适数目的不同的条形码序列(例如,至少1,000个不同的条形码序列、至少10,000个不同的条形码序列、至少100,000个不同的条形码序列、至少1,000,000个不同的条形码序列、至少10,000,000个不同的条形码序列或如本文别处所述的任何其他数目的不同的条形码序列)。而且,给定分区中的寡核苷酸可包含共同的条形码序列。给定分区中的寡核苷酸和关联的共同的条形码序列可附接至该给定分区内的一个或多个靶核酸的片段或靶核酸序列的部分的拷贝。附接之后,可随后合并单独的分区。然后可对靶核酸的片段或靶核酸的部分的拷贝以及附接的条形码序列进行测序。

[0452] 在另一个实例中,可通过提供靶核酸序列和将它们分离至多个单独的分区来对多个靶核酸序列进行测序。所述多个单独的分区中的每一个分区可包含一个或多个靶核酸序列和具有多个附接的寡核苷酸的珠子。附接至给定珠子的寡核苷酸可包含共同的条形码序列。与珠子关联的寡核苷酸可附接至给定分区内的靶核酸序列的片段或靶核酸序列的部分的拷贝,以使得给定分区的片段或拷贝也附接至与珠子关联的共同的条形码序列。将寡核苷酸附接至靶核酸序列的片段或靶核酸序列的部分的拷贝之后,可随后合并单独的分区。然后可对靶核酸序列的片段或靶核酸序列的部分的拷贝以及任何附接的条形码序列进行测序(例如,使用任何合适的测序方法,包括本文别处所述的那些方法),以提供条形码化的片段序列或条形码化的拷贝序列。部分地基于条形码化的片段序列或条形码化的拷贝序列的条形码部分,条形码化的片段序列或条形码化的拷贝序列可组装成一个或多个毗连的核酸序列。

[0453] 在一些情况下,对不同数目的条形码化的寡核苷酸进行测序。例如,在一些情况下,对约30%-90%的条形码化的寡核苷酸进行测序。在一些情况下,对约35%-85%、40%-80%、45%-75%、55%-65%或50%-60%的条形码化的寡核苷酸进行测序。在一些情况下,对至少约30%、40%、50%、60%、70%、80%或90%的条形码化的寡核苷酸进行测序。在一些情况下,对小于约30%、40%、50%、60%、70%、80%或90%的条形码化的寡核苷酸进行测序。

[0454] 在一些情况下,从片段组装序列,从而为可能比单独的序列读取结果更长的原始靶多核苷酸的毗连区提供序列信息。单独的序列读取结果可为约10-50、50-100、100-200、200-300、300-400个或更多个核苷酸的长度。序列组装方法的实例包括在与本申请同日提交的美国临时专利申请号_____ (代理人案号43487-729.101) 中阐述的那些。

[0455] 条形码的身份可用于整理来自单独片段的序列读取结果以及区别单元型。例如,当在流体小滴内组合单独的样品片段与条形码化的珠子时,亲本多核苷酸片段可分离至不同的小滴。随着流体小滴和小滴内珠子的数目的增加,来自母本和父本单元型两者的片段包含在与相同珠子关联的相同流体小滴内的可能性可能变得极小。因此,可组装并整理来自与相同珠子关联的相同流体小滴中的片段的序列读取结果。

[0456] 在至少一个实例中,本公开内容提供了核酸测序方法、系统、组合物和这些的组合,其可用于在序列组装和读长等价两方面提供种种益处,而这样做具有非常高的通量并减少了样品制备时间和成本。

[0457] 通常,本文所述的测序方法提供了遗传序列的片段的局部标记或条形码化。通过标记来源于较大遗传序列内相同位置的片段,可利用标签或条形码的存在来告知如上所提到的组装过程。此外,可使用本文所述的方法由单一的长核酸分子生成较短的片段并对该较短的片段进行条形码化。对这些较短片段的测序和组装提供了长读等价序列,但并不需要低通量较长读长测序技术。

[0458] 图39提供了示例测序方法的示意图。如图所示,将可包含例如染色体或其他大核酸分子的第一遗传组分3902片段化成一组大的第一核酸片段,例如,包括片段3904和3906。大遗传组分的片段可以是非重叠的或重叠的,并且在一些情况下可包含多倍重叠片段,以便提供较大组分的序列的高置信度组装。在一些情况下,较大遗传组分的片段提供较大组分的1X、2X、5X、10X、20X、40X覆盖范围或更大的覆盖范围。

[0459] 然后处理第一片段3904中的一个或多个,以分别提供第一片段的第二片段的重叠组,例如,第二片段组3908和3910。该处理还为第二片段提供条形码序列,该条形码序列对于来源于特定第一片段的每一个第二片段都是相同的。如图所示,第二片段组3908的条形码序列表示为“1”,而片段组3910的条形码序列表示为“2”。可使用条形码的多样性文库来对大量不同的片段组区别地进行条形码化。然而,来自不同第一片段的各个第二片段组不必用不同的条形码序列进行条形码化。事实上,在许多情况下,可同时处理多个不同的第一片段,以使其包含相同的条形码序列。本文别处详细描述了多样性的条形码文库。

[0460] 随后可将来自例如片段组3908和3910的条形码化的片段合并用于测序。一旦测序,可至少部分地基于所包含的条形码,以及任选地和优选地部分基于片段自身的序列,而将序列读取结果3912归属于其各自的片段组,例如,如在聚集的读取结果3914和3916中所示的。随后组装每一个片段组的归属的序列读取结果以提供第一片段的组装的序列,例如,片段序列3918和3920,其反过来可组装成较大遗传组分的序列3922。

[0461] 根据前文,将大遗传组分如长核酸片段(例如长度为1、10、20、40、50、75、100、1000kb或更多kb)、染色体片段或全染色体或整个基因组的一部分(例如,基因组DNA)片段化成较小的第一片段。通常,这些片段可为约1000至约100000个碱基之间的任何长度。在某些优选方面,片段将为约1kb至约100kb,或约5kb至约50kb,或约10kb至约30kb,并且在一些情况下,为约15kb至约25kb。这些较大遗传组分的片段化可通过多种方便可得的方法中的任何方法来进行,该方法包括可商购获得的基于剪切的片段化系统(例如,Covaris片段化系统)、大小靶向片段化系统(例如Blue Pippin(Sage Sciences))、酶片段化方法(例如,使用限制性内切核酸酶)等等。如上所述,较大遗传组分的第一片段可包括重叠或非重叠的第一片段。尽管本文描述为在分配之前进行片段化,但应当理解片段化可任选地和/或另外地在例如一个或多个扩增步骤之后的过程中较晚地进行,以产生用于测序应用的所需大小的片段。

[0462] 在优选的方面,第一片段由较大遗传组分或其部分的多个拷贝生成,使得产生重叠的第一片段。在优选的方面,重叠的片段将构成潜在较大遗传组分或其部分的大于1X覆盖范围、大于2X覆盖范围、大于5X覆盖范围、大于10X覆盖范围、大于20X覆盖范围、大于40X覆盖范围或甚至更大的覆盖范围。随后将第一片段隔离至不同的反应体积。在一些情况下,可分离第一片段以使得反应体积含有一个或较少的第一片段。这通常通过以在溶液中有限稀释的方式提供片段来实现,以使得将溶液分配至不同的反应体积导致多于一个片段被放置于给定反应体积的概率非常低。然而,在多数情况下,给定反应体积可包含多个不同的第一片段,并且可能甚至具有2、5、10、100、1000个或甚至高达10,000个或更多个在给定反应体积中的不同的第一片段。此外,通常基于对该起始材料中的核酸浓度的理解,通过对第一片段所源自的溶液的适当稀释来得到在单独反应体积内期望范围的片段数目。

[0463] 反应体积可包含各种不同类型的容器或分区中的任何一种。例如,反应体积可包括常规反应容器,诸如试管、反应孔、微孔、纳米孔,或者反应体积可包括较小的常规反应体积,诸如稳定乳液,例如油包水乳液体系内的小滴。在优选的方面,由于小滴具有极高的多重化能力,例如,允许使用单一容器内成千上万、数百万、数千万个或甚至更多个离散小滴/反应体积,因此优选小滴作为反应体积。在每一个反应体积内,随后使其中包含的片段经历处理,该处理衍生每一个第一片段的重叠第二片段组,并且还为这些第二片段提供附接的

条形码序列。应当理解，在优选的方面，将第一片段分配至还含有一个或多个微胶囊或珠子的小滴中，所述微胶囊或珠子包含用于生成第二片段和对第二片段进行条形码化的条形码文库的成员。

[0464] 在优选的方面，这些第二片段的生成通过引入包含条形码序列的引物序列以及能够与部分第一片段杂交并沿着第一片段延伸的引物序列来进行，以提供包含条形码序列的第二片段。这些引物可包含靶向引物序列例如以衍生与第一片段的特定部分重叠的片段，或者它们可包含通用引发序列，例如，将引发第一片段的多个不同区域以产生大且多样性组的第二片段（其跨越第一片段并提供多倍重叠覆盖范围）的随机引物。这些延伸的引物序列可用作第二片段，或者它们可被进一步复制或扩增。例如，针对延伸的序列反复引发（例如，使用含相同引物的条形码化的寡核苷酸）。在某些优选的方面，第二组片段的生成产生了第一片段的部分的部分发夹复制物，如本文别处所述，这些复制物各自包含条形码序列，例如，用于如本文所述的PHASE扩增。如本文别处所述，通常期望形成部分发夹以防止重新引发复制的链，例如，制备拷贝的拷贝。同样地，相对于与扩增产物退火的引物，部分发夹通常优先由扩增产物在退火期间形成，例如，发夹将具有高于引物产物对的T_m。

[0465] 通常选择具有适合于后续测序的长度的第二片段。对于短读测序技术，这样的片段将通常为约50个碱基至约1000个碱基的可测序长度、约50个碱基至约900个碱基的可测序长度、约50个碱基至约800个碱基的可测序长度、约50个碱基至约700个碱基的可测序长度、约50个碱基至约600个碱基的可测序长度、约50个碱基至约500个碱基的可测序长度、约50个碱基至约400个碱基的可测序长度、约50个碱基至约300个碱基的可测序长度、约50个碱基至约250个碱基的可测序长度、约50个碱基至约200个碱基的可测序长度、或约50个碱基至约100个碱基的可测序长度（包括条形码序列段和经历测序过程的功能序列）。

[0466] 一旦重叠，则生成条形码化的第二片段组，它们可合并用于后续处理和最终的测序。例如，在一些情况下，如本文别处所述，随后可使条形码化的片段经历附加扩增，例如，PCR扩增。同样地，这些片段可另外地或同时被提供有用于鉴定条形码化的片段的集合所源自的样品的样品索引序列，以及提供用于在测序过程中使用的附加功能序列。

[0467] 此外，还可任选地进行清理步骤，例如，以从其他杂质中纯化核酸组分，按大小选择用于测序的片段组等。这样的清理步骤可包括对SPRI珠子（诸如可从Beckman Coulter, Inc. 获得的 Ampure® 珠子）进行纯化和/或大小选择。在一些情况下，当片段与SPRI珠子关联时，多个过程步骤可在一集成过程中执行，例如，如在Fisher等人，Genome Biol. 2011;12(1):R1 (E-pub 2011年1月4日) 中所述，该文献为了所有目的通过引用以全文并入本文。

[0468] 如先前所述，在许多情况下，使用短读测序技术来提供第二片段组的序列信息。因此，在优选的方面，第二片段组通常将包含片段，该片段在包含条形码序列时将在所用测序系统的读长内。例如，对于 Illumina HiSeq® 测序，当进行配对末端测序时，这样的片段可通常为约100个碱基至约200个碱基的长度。在一些情况下，当通过测序过程仅获得片段的末端部分时可对较长的第二片段进行测序。

[0469] 如上文关于图39所述，不同第二片段的序列读取结果随后归属于其各自的起始核酸区段，这部分地基于特定条形码序列的存在，以及在一些情况下部分地基于片段的实际

序列,即,片段序列的非条形码部分。应当理解,尽管基于短序列数据,但可推断出共享相同条形码的两个序列很可能源于相同的较长第一片段序列,尤其是在这样的序列能够以其他方式组装成毗连序列段的情况下,例如,使用带有共同条形码的其他重叠序列。一旦第一片段被组装,则第一片段可组装成较大的序列段,例如,完整长度的遗传组分。

[0470] 在一个示例性过程中,一个或多个模板核酸序列的一个或多个片段可使用本文所述方法进行条形码化。所述一个或多个片段中的片段可至少部分地基于附接至其上的核酸条形码序列进行表征。片段的表征还可包括将片段定位到其各自的模板核酸序列或模板核酸序列所源自的基因组。而且,表征还可包括鉴定单独的核酸条形码序列以及附接至其上的模板核酸序列的片段的序列。

[0471] 在一些情况下,本文所述的测序方法在表征核酸段或靶核酸中可能是有用的。在一些示例性方法中,可通过将核酸区段和包含多个寡核苷酸(其包含共同的核酸条形码序列)的珠子(例如,包括本文所述的任何合适类型的珠子)共同分配至分区(包括本文所述的任何合适类型的分区,例如,小滴)中来表征核酸区段。所述寡核苷酸可被可释放地附接至如本文别处所述的珠子(例如,向珠子施加刺激时可从珠子释放,例如,热刺激、光刺激和化学刺激),和/或可包含一个或多个功能序列(例如,引物序列、引物退火序列、固定序列、本文别处所述的任何其他合适的功能序列等)和/或如本文别处所述的一个或多个测序引物序列。而且,可将任何合适数目的寡核苷酸附接至珠子,包括本文别处所述的附接至珠子的多个数目的寡核苷酸。

[0472] 在所述分区内,寡核苷酸可附接至核酸区段的片段或核酸区段的部分的拷贝,以使得该片段或拷贝也附接至共同的核酸条形码序列。该片段可以是核酸区段的重叠片段,并且可提供例如核酸区段的大于2X覆盖范围、大于5X覆盖范围、大于10X覆盖范围、大于20X覆盖范围、大于40X覆盖范围或甚至更大的覆盖范围。在一些情况下,寡核苷酸可包含能够与核酸区段的一部分或其互补序列退火的引物序列。在一些情况下,可通过延伸寡核苷酸的引物序列来附接寡核苷酸,以复制核酸区段的至少一部分或其互补序列,从而产生核酸区段的至少一部分的拷贝,该拷贝包含寡核苷酸并因此包含共同的核酸条形码序列。

[0473] 将寡核苷酸附接至核酸区段的片段或核酸区段的部分的拷贝之后,可经由任何合适的测序方法(包括本文所述的任何类型的测序方法)对核酸区段的片段或核酸区段的部分的拷贝以及附接的寡核苷酸(包括寡核苷酸的条形码序列)进行测序,以提供多个条形码化的片段序列或条形码化的拷贝序列。测序之后,核酸区段的片段或核酸区段的部分的拷贝可至少部分地基于它们附接至共同的核酸条形码序列而被表征为在核酸段内连接。应当理解,这样的表征可包括被表征为连接的和毗连的序列,以及可能在相同的片段内连接的序列,而非毗连的序列。而且,测序过程中生成的条形码化的片段序列或条形码化的拷贝序列可至少部分地基于共同核酸条形码序列和/或条形码化的片段序列或条形码化的拷贝序列的非条形码部分而组装成一个或多个毗连的核酸序列。

[0474] 在一些情况下,可将多个核酸区段(例如,如本文别处所述的基因组的至少一部分的片段)与多个不同的珠子共同分配至多个单独的分区中,使得单独分区中的多个不同分区中的每一个分区均含有单个珠子。所述多个不同珠子可包含多个不同的条形码序列(例如,至少1,000个不同的条形码序列、至少10,000个不同的条形码序列、至少100,000个不同的条形码序列、至少1,000,000个不同的条形码序列或如本文别处所述的任何其他数目的

不同的条形码序列)。在一些情况下,所述多个单独分区中的两个或更多个、三个或更多个、四个或更多个、五个或更多个、六个或更多个、七个或更多个可包含含有相同条形码序列的珠子。在一些情况下,至少0.01%、0.1%、1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或99%的单独分区可包含具有相同条形码序列的珠子。而且,每一个珠子可包含多个附接的寡核苷酸,该寡核苷酸包含共同的核酸条形码序列。

[0475] 共同分配之后,可将条形码序列附接至每一个分区中核酸区段的片段或核酸区段的部分的拷贝。随后可将核酸区段的片段或核酸区段的部分的拷贝从单独的分区中合并。合并后,可对核酸区段的片段或核酸区段的部分的拷贝以及任何关联的条形码序列进行测序(例如,使用任何合适的测序方法,包括本文所述的那些方法),以提供经测序的片段或经测序的拷贝。经测序的片段或经测序的拷贝可至少部分地基于包含共同的条形码序列的经测序的片段或经测序的拷贝而被表征为来源于共同的核酸区段。而且,可组装从经测序的片段或经测序的拷贝获得的序列,以提供经测序的片段或经测序的拷贝所源自的序列的毗连序列(例如,基因组的至少一部分)。可至少部分地基于经测序的片段的每一个核苷酸序列和经测序的片段的共同条形码序列完成从经测序的片段或经测序的拷贝进行的序列组装。

[0476] 在另一示例方法中,可通过将靶核酸的片段分至配多个小滴中来表征靶核酸。每一个小滴可包含附接至包含共同的条形码序列的多个寡核苷酸的珠子。共同的条形码序列可附接至小滴中靶核酸的片段的片段。然后可将小滴合并,并使用任何合适的测序方法(包括本文所述的测序方法)对合并的小滴的片段和关联的条形码序列进行测序。测序之后,可至少部分地基于包含共同的条形码序列的靶核酸的片段的片段将靶核酸的片段的片段定位到靶核酸的片段。

[0477] 本文所述的方法、组合物和系统在测序中的应用通常可适用于多种不同的测序技术中的任何技术,包括NGS测序技术如Illumina MiSeq、HiSeq和X10测序系统,以及可从Life Technologies, Inc.获得的测序系统如Ion Torrent测序系统线。尽管关于条形码序列进行了讨论,但应当理解经测序的条形码序列可能不包括被包含在内的整个条形码序列,例如,考虑到测序错误。同样地,当提及将两个条形码序列表征为相同的条形码序列时,应当理解这可能是基于对条形码序列的很大一部分的识别,例如变化了少于5、4、3、2个或甚至单个碱基。

[0478] 从少量细胞进行测序

[0479] 本文提供的方法还可以用来以使得能够获得细胞特异性信息的方式制备细胞内含有的多核苷酸。该方法使得能够从极小的样品如包含约10–100个细胞的样品检测遗传变异。在一些情况下,在本文描述的方法中可以使用约1、5、10、20、30、40、50、60、70、80、90或100个细胞。在一些情况下,在本文描述的方法中可以使用至少约1、5、10、20、30、40、50、60、70、80、90或100个细胞。在其他情况下,在本文描述的方法中可以使用至多约5、10、20、30、40、50、60、70、80、90或100个细胞。

[0480] 在一个实例中,方法可包括分配细胞样品(或粗细胞提取物)以使得在分区例如微流体小滴内存在至多一个细胞(或一个细胞的提取物),并且将其与例如以上所述的条形码寡核苷酸共同分配。处理然后包括裂解细胞,对细胞内包含的多核苷酸进行片段化,将片段

化的多核苷酸附接至条形码化的珠子，合并该条形码化的珠子，并对所得的条形码化的核酸片段进行测序。

[0481] 如本文别处所述，条形码和其他试剂可以封装在珠子（例如，凝胶珠子）内、包覆于其上、与之关联或分散于其内。该珠子可以与样品（例如，细胞）负载同时负载到流体小滴内，使得每个细胞与不同的珠子接触。该技术可以用来将独特条形码附接至从每个细胞获得的寡核苷酸。然后可以将所得的标记的寡核苷酸合并且测序，并可以使用条形码追踪该寡核苷酸的起源。例如，可以确定具有相同条形码的寡核苷酸起源于相同的细胞，同时可以确定具有不同条形码的寡核苷酸起源于不同的细胞。

[0482] 本文描述的方法可以用来检测可以指示存在疾病如癌症的特定基因突变。例如，检测结肠组织样品的BRAF基因中V600突变的存在可以指示结肠癌的存在。在其他情况下，预后应用可包括检测可充当发生特定疾病的提高的风险因素的一个或多个特定基因中的突变。例如，检测乳腺组织样品中BRCA1突变的存在可指示比没有该突变的个人更高水平的发生乳腺癌的风险。在一些实例中，本公开内容提供了鉴别两个不同癌基因（例如，KRAS和EGFR）中的突变的方法。如果相同细胞包含具有两个突变的基因，这可以指示癌症的更具侵袭性的形式。与此相反，如果突变位于两个不同的细胞中，这可以指示该癌症可能是更加良性的或不到晚期。

[0483] 基因表达的分析

[0484] 本公开内容的方法可适用于处理样品以用于检测基因表达的变化。样品可包含细胞、mRNA或由mRNA反转录的cDNA。样品可以是包含来自数个不同细胞或组织的提取物的合并样品，或者包含来自单个细胞或组织的提取物的样品。

[0485] 可将细胞直接置于流体小滴中并裂解。裂解后，可使用本公开内容的方法对细胞的寡核苷酸进行片段化和条形码化以供测序。寡核苷酸也可以在将其引入在本公开内容的方法中使用的流体小滴中之前从细胞提取。mRNA的逆转录可在本文所述的流体小滴中，或在这样的流体小滴的外部进行。cDNA测序可提供随时间或在暴露于特定条件后特定细胞中的特定转录物的丰度的指示。

[0486] 从细胞或蛋白质分配多核苷酸

[0487] 在一个实例中，本公开内容提供的组合物、方法、装置和试剂盒可用来将细胞或蛋白质封装在流体小滴内。在一个实例中，可以将单个细胞或多个细胞（例如，2、10、50、100、1000、10000、25000、50000、10000、50000、1000000个或更多个细胞）与流体小滴内的裂解缓冲液一起负载至珠子之上、之内或内部，并温育指定的一段时间。该珠子可以是多孔的，以允许洗涤珠子的内含物以及将试剂引入珠子中，同时将一个或多个细胞的多核苷酸（例如，染色体）保持在流体小滴内。然后可根据本公开内容中提供的或本领域已知的任何方法，处理封装的一个或多个细胞的多核苷酸（例如，染色体）。该方法也可应用于任何其他的细胞组分，如蛋白质。

[0488] 表观遗传应用

[0489] 本公开内容的组合物、方法、装置和试剂盒在表观遗传应用中可能中是有用的。例如，DNA甲基化可以是表观遗传的指标，包括单核苷酸多态性（SNP）。因此，可对包含核酸的样品进行处理，以便确定在测序过程中被甲基化的碱基。在一些情况下，可将待条形码化的包含核酸的样品分成两个等份。样品的一个等份可用亚硫酸氢盐处理，以便将未甲基化的

含胞嘧啶的核苷酸转化为含尿嘧啶的核苷酸。在一些情况下，亚硫酸氢盐处理可在样品分配之前发生或可在样品分配之后发生。然后可分配每一个等份(如果还没有分配)，在分区中进行条形码化，并如本文所述批量添加附加序列以生成测序仪就绪产物。针对每一个等份(例如，亚硫酸氢盐处理的样品与未处理的样品)获得的测序数据的比较可用于确定样品核酸中哪些碱基被甲基化。

[0490] 在一些情况下，分开的样品的一个等份可用甲基化敏感性限制酶(MSRE)进行处理。甲基化特异性酶可处理样品核酸，使得样品核酸在甲基化位点切割。样品等份的处理可在样品分配之前发生或可在样品分配之后发生，并可分配每一个等份以用于生成条形码化的测序仪就绪产物。针对每一个等份(例如，MSRE处理的样品与未处理的样品)获得的测序数据的比较可用于确定样品核酸中哪些碱基被甲基化。

[0491] 低输入DNA应用

[0492] 本文所述的组合物和方法在低多核苷酸输入应用的分析和测序中可能是有用的。本文所述的方法如PHASE可帮助在低多核苷酸输入应用中获得良好的数据质量和/或帮助滤掉扩增错误。这些低输入DNA应用包括用于对在不相关的或相关性较低的核酸的混合物(其中感兴趣的序列仅为少数组分)中的感兴趣的特定核酸序列进行测序和鉴定的样品分析，以能够对存在于不同核酸的聚集体中的多个不同核酸进行单独测序和鉴定，以及其中输入DNA的绝对量非常低的分析。具体的实例包括对来自组织样品或来自循环细胞的体细胞突变的测序和鉴定，其中样品的绝大部分将来源于正常的健康细胞，而小部分可来源于肿瘤或其他癌细胞。其他的实例包括在例如微生物组分析应用中对多个单独群体组分的表征，其中在庞大而多样化的微生物元素群体中单个群体成员的贡献可能不能以其他方式容易地鉴定。在进一步的实例中，能够单独地测序并鉴定来自不同染色体例如母本和父本染色体的相同区域的不同链允许鉴定每一个染色体上的独特变体。本文所述的组合物、方法和系统的低多核苷酸输入应用的其他实例在与本申请同日提交的美国临时专利申请号_____ (代理人案号43487-727.101) 中进行了阐述。

[0493] 经对本领域现状所面临的问题的讨论，本文所述的方法和系统的优势更加清晰。在分析样品材料例如细胞或组织样品的遗传组成时，多数测序技术依赖于样品中靶核酸的广泛扩增，以便产生足以用于测序过程的材料。遗憾的是，在这些扩增过程中，多数本发明的材料将优先压倒性多于样品中以较低水平存在的部分。例如，在来自样品的遗传材料由95%的正常组织DNA和5%的来自肿瘤细胞的DNA组成的情况下，典型的扩增过程例如基于PCR的扩增将快速扩增多数本发明的材料(少数本发明的材料排除在外)。此外，因为这些扩增反应通常在合并的环境中进行，因此在该过程中，就特定的染色体、多核苷酸或生物体而言，扩增序列的起源通常不会保留下。

[0494] 相比之下，本文所述的方法和系统将单独的或小数目的核酸分配至这些核酸组分可进行初始扩增的单独的反应体积例如小滴中。在该初始扩增中，独特标识符可与在那些单独的反应体积中的组分偶联。在整个测序过程(包括后续的扩增过程，例如PCR扩增)中，不同组分的单独的分配扩增以及独特标识符例如条形码序列的应用允许保留每一个样品组分的贡献以及其起源的属性。

[0495] 如本文和整个公开内容所用的术语“约”通常指可以比特定使用的情况内的规定数值大15%或小15%的范围。例如，“约10”将包括8.5至11.5的范围。

[0496] 可以理解,本公开内容提供了本文所述的任何组合物、文库、方法、装置和试剂盒用于特定用途或目的(包括本文所述的各种应用、用途和目的)的应用。例如,本公开内容提供了本文所述的组合物、方法、文库、装置和试剂盒在分配物质、分配寡核苷酸、物质从分区中的刺激选择性释放、在分区中进行反应(例如,连接和扩增反应)、进行核酸合成反应、对核酸进行条形码化、为测序准备多核苷酸、多核苷酸测序、多核苷酸取相(参见,例如,与本申请同日提交的美国临时专利申请号_____ , (代理人案号43487-726.101))、从少量细胞对多核苷酸进行测序、分析基因表达、从细胞分配多核苷酸、突变检测、神经系统病症诊断、糖尿病诊断、胎儿非整倍性诊断、癌症突变检测和法医学、疾病检测、医学诊断、低输入核酸应用如循环肿瘤细胞(CTC)测序、其组合中以及在本文所述的任何其他应用、方法、过程或使用中的用途。

[0497] 本文提供的任何浓度值作为混合浓度值提供,不考虑任何原位转化、修饰、反应、螯合等。而且,在适当的情况下,本文所述方法(例如,测序方法、条形码化方法、扩增方法、靶向扩增方法、分析条形码化的样品的方法等)的灵敏度和/或特异性可变化。例如,本文所述方法可具有大于50%、70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.5%的特异性,和/或大于50%、70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或99.5%的灵敏度。

[0498] X. 实施例

[0499] 实施例1:用Acrydite引物功能化的凝胶珠子的产生

[0500] 根据图2所示的方法产生凝胶珠子。在不含核酸酶的水中,制备1mL下列浓度的储备溶液:丙烯酰胺前体(化合物A)=40% (v/v) 储备溶液、交联剂(双-丙烯酰胺-化合物B)=3.19mg/mL (在乙腈:水为50:50的混合物中)、引发剂(化合物C)=20mg/mL 以及二硫键acrydite引物(化合物D)=1mM。由这些储备溶液通过混合下列体积来制备1mL的水性凝胶珠子(GB)工作溶液:不含核酸酶的水=648μL、化合物A=150μL、化合物B=100μL、化合物C=100μL 和化合物D=2μL。每日制备化合物A和B的储备溶液以及GB工作溶液。

[0501] 凝胶珠子(GB)工作溶液201是含有交联剂、BAC以及具有约0.1至约100μm浓度的二硫键修饰的acrydite寡核苷酸的聚合物前体溶液的水性流体。第二流体202是含有表面活性剂(Krytox FSH 1.8% w/w HFE7500)的氟化油。将加速剂四甲基乙二胺(TEMED) a) 在小滴生成之前添加至油中(203), b) 在小滴生成之后添加至管线中(205), 和/或c) 在小滴生成之后添加至出口储器中(206), 以得到最终浓度1% (v/v)。每日制备新鲜的TEMED。通过将水相和油相流体送至小滴发生器来生成凝胶珠子(204)。小滴生成之后立即引发聚合并且聚合持续至出口孔。15-20分钟之后认为胶凝完成。胶凝之后,通过在HFE 7500中洗涤而使生成的凝胶珠子经历连续相交换(207), 以去除过量的油并将珠子重悬于水溶液中。在一些情况下,所得珠子可能以团聚体的形式存在。采用涡流将凝胶珠子的团聚体分离成单个凝胶珠子。在显微镜下对凝胶珠子进行可视化。

[0502] 实施例2:通过有限稀释产生条形码化的凝胶珠子

[0503] 根据图3A和图4所示的方法通过有限稀释产生功能化的凝胶珠子。将具有acrydite寡核苷酸(有或没有二硫键修饰)的凝胶珠子301、401与含条形码的模板序列302在有限稀释下混合。将包含生物素标记的读取引物406的PCR试剂303与凝胶珠子和模板序

列混合,304。将珠子、条形码模板和PCR试剂通过振摇/搅拌、流动聚焦或微筛分乳化为凝胶-水-油乳液(305),优选使得至多一个条形码模板存在于乳液内的分区(例如,小滴)中。将乳液暴露于一个或多个热循环(306)。第一热循环引入互补的条形码序列408,并将其固定在凝胶珠子上。

[0504] 进行持续的热循环以克隆扩增凝胶珠子各个部分的条形码,并将5'生物素标记的引物引入互补链中以用于含有条形码序列的珠子与不含条形码序列的珠子的下游分选。通过添加全氟癸醇,去除油,用HFE-7500洗涤,添加水性缓冲液,离心,去除上清液,去除不期望的产物(例如引物二聚体、起始材料、脱氧核苷酸三磷酸(dNTP)、酶等)以及将可降解的凝胶珠子回收至水性悬浮液中来破乳(307)。将功能化的凝胶珠子重悬于高盐缓冲液中(308)。将链霉亲和素标记的磁珠添加至重悬液中,随后将重悬液温育以允许与附接至生物素化的条形码的凝胶珠子结合(308、410)。然后使用磁性装置将未附接至条形码的珠子与阳性条形码化的凝胶珠子分离(308)。向凝胶珠子施加变性条件309(例如,热或化学变性剂),以便从条形码化的珠子中分离生物素化的互补链。随后从溶液中去除磁珠;并将所得的部分功能化的条形码化的珠子的溶液合并,以供进一步处理。

[0505] 实施例3:条形码化的珠子的进一步功能化

[0506] 如图3B所示,如下所述将来自实施例2的条形码化的凝胶珠子311进一步功能化。珠子与另外的模板寡核苷酸310(诸如含随机N-mer序列的寡核苷酸413,如图4所示)和PCR试剂312组合(313),并且经受能够使模板寡核苷酸与附接至凝胶珠子的读取引物杂交的条件。进行延伸反应,使得条形码链延伸(314),从而引入模板寡核苷酸的互补序列。将所得到的功能化的凝胶珠子重悬于水性缓冲液中(315),并使其暴露于加热条件以去除互补链(316),并将其放置于水性储存缓冲液中(317)。

[0507] 实施例4:珠子功能化的分步描述

[0508] 图4提供了用条形码和随机N-mer对凝胶珠子进行功能化的示例过程的分步描述。如图4A所示,该过程始于凝胶珠子401,其附接至通用引物如P5引物(或其互补序列)403。珠子可经由二硫键402与引物连接。在水性溶液中提供凝胶珠子(g/w)。通过使用有限稀释和分配,使独特的条形码序列模板405与珠子组合,以使得至多一个独特条形码序列占据与凝胶珠子相同的分区。通常,该分区为凝胶/水/油(g/w/o)乳液内的水性小滴。如图4B所示,条形码序列模板405被包含在较大的核苷酸链中,该核苷酸链包含与通用引物403互补的序列404以及与生物素标记的读取引物406在序列上相同的序列407。

[0509] 如图4C所示,随后进行扩增反应,以将条形码模板405的互补序列408引入到附接至珠子的链上。该扩增反应还导致与序列407互补的序列415的引入。附加的扩增循环导致生物素标记的读取引物406与序列415杂交(图4D),随后生物素标记的读取引物延伸(图4E)。然后可破乳,并且随后可将凝胶珠子合并入凝胶/水的共同溶液中。

[0510] 在凝胶/水溶液中,磁性捕获珠子409随后用于捕获附接至凝胶珠子的生物素化的核酸,该凝胶珠子随后从仅含有原始引物的珠子中分离出(图4F和图4G)。然后将生物素化的链从附接至凝胶珠子的链中去除(图4H)。随后可将随机N-mer序列414附接至已附接至凝胶珠子的链。对于每一个凝胶珠子,相同的条形码序列408附接至整个凝胶珠子中的每一个引物;然后用随机N-mer序列414对每一个条形码序列进行功能化,以使得多个不同的随机N-mer序列附接至每一个珠子。对于这一过程,与序列412(与序列415互补)连接的随机N-

mer模板序列413被引入到含有合并的珠子的溶液中(图4I)。使溶液经受能够将模板与附接至珠子的链杂交的条件,并将序列415延伸以使之包含随机N-mer 414(图4J)。然后将完全功能化的珠子(图4K)与样品核酸和还原剂(例如,1mM浓度的二硫苏糖醇(DTT))组合,并将其分配在凝胶/水/油乳液的小滴内(图4L)。可采用微流体装置进行该组合步骤(图5A)。随后在每一个分区(例如,小滴)内例如通过还原剂的作用将凝胶珠子降解,并且使条形码化的序列从小滴释放(图4M和图4N)。在条形码化的序列内的随机N-mer可充当扩增样品核酸的引物。

[0511] 实施例5:使用微流体芯片组合乳液中的凝胶-珠子(GEM)与样品

[0512] 可使用图5所示的双交叉(double-cross)微流体装置使功能化的凝胶珠子与样品组合。将可降解的凝胶珠子在含有约7%甘油的流体流中引入至流体输入端501。将感兴趣的实验样品在为水相的流体流中引入至流体输入端502。将约1mM浓度的还原剂二硫苏糖醇(DTT)在含有约7%甘油的流体流中引入至流体输入端503。流体输入端501、502和503在微流体交叉接合处504混合,并进入第二微流体交叉接合处506。第二微流体交叉接合处可用于产生含凝胶珠子的乳化(w/o)小滴。使用流体输入端505来引入具有2%(w/w)bis krytox peg(BKP)的油。将从第二微流体交叉接合处离开的单独的小滴507添加至微板孔(图5C)中,以用于进一步的下游应用。图5D是在不存在DTT的情况下生成的小滴的图像(并因此含有凝胶珠子)。图5E是采用DTT生成的小滴的图像,该DTT使内部凝胶珠子降解。

[0513] 实施例6:阳性凝胶珠子的荧光鉴定

[0514] 图6示出了含有已采用荧光标记物标记的已扩增核酸的凝胶珠子的图像。首先使用有限稀释进行凝胶珠子的功能化,使得仅一部分凝胶珠子用条形码进行功能化。将悬浮于bis krytox peg(BKP)乳液中的凝胶珠子在PCR热循环后但在洗涤之前以4X放大倍数成像。亮视野图像,图6A,显示了全部的乳液生成的小滴,而荧光图像,图6B,显示了仅阳性功能化的凝胶珠子。生成许多指示空小滴的非荧光小滴,该小滴不含凝胶珠子和/或寡核苷酸。通过多次重悬以及在HFE-7500中洗涤来洗掉空小滴。图6C和图6D显示了破乳和进一步的洗涤步骤后的阳性凝胶珠子富集。亮视野图像(4X)图6C和(10X)图6E显示了全部凝胶珠子。荧光图像(4X)图6D和(10X)图6F显示了来自SYBR染色的30%阳性珠子。30%阳性珠子的结果与来自gDNA输入的预测值相匹配。

[0515] 图7显示了含有单链(ss)DNA、双链(ds)DNA和变性ssDNA的凝胶珠子的图像。从在以下步骤所拍摄的荧光图像明显地看出用1X EvaGreen染色的凝胶珠子在dsDNA的存在下更加明亮,步骤1:制备(ssDNA)(图7A),步骤2:延伸(dsDNA)(图7B),以及步骤3:变性(ssDNA)(图7C)。荧光图像显示珠子在延伸之后变得较亮并且在变性之后变得较暗。

[0516] 实施例7:使用链霉亲和素包覆的磁珠富集阳性凝胶珠子

[0517] 图8中示出了使用链霉亲和素包覆的磁珠富集阳性凝胶珠子。图8A(亮视野)和图8B(荧光)提供了添加磁珠后24小时的SYBR染色的凝胶珠子的图像。磁性包覆的阳性凝胶珠子由于SYBR染色而更亮。在40mg/mL磁珠浓度下在分选之前(图8C)和分选之后(图8D)的亮视野图像显示了阳性凝胶珠子富集,其中包覆的珠子在光学上更亮。在60mg/mL磁珠浓度下在分选之前(图8E)和分选之后(图8F)的亮视野图像显示了阳性凝胶珠子富集,其中包覆的珠子在光学上更亮。在每一个磁珠工作浓度下,单个凝胶珠子由约100-1000个磁珠包覆。

[0518] 实施例8:凝胶珠子的溶解

[0519] 从图9中明显看出,在碱性溶液中加热凝胶珠子使凝胶珠子降解。将凝胶珠子在95 °C下在碱性溶液中加热,并以5分钟的加热时间间隔进行监测:t=0min(图9A),t=5min(图9B),t=10min(图9C),t=15min(图9D)。15分钟后,凝胶珠子完全降解。凝胶珠子的大小在降解时超过两倍。图10示出了使用三(2-羧基乙基)膦(TCEP)溶解凝胶珠子,三(2-羧基乙基)膦是有效且不可逆的二硫键还原剂。将功能化的凝胶珠子(图10A)放置于具有1mM TCEP的碱性溶液(pH=8)中,并以2分钟的时间间隔进行监测:t=0min(图10B),t=2min(图10C),t=4min(图10D),t=6min(图10E),t=8min(图10F),t=10min(图10G)。经约6分钟至约10分钟,功能化的凝胶珠子完全降解。

[0520] 实施例9:溶解凝胶珠子(GB)之后对内含物的分析

[0521] 图11和图12中提供了对附接至凝胶珠子的内含物的分析。用条形码或条形码互补序列(N12C)以及8个核苷酸长度的随机N-mer(8mer)1102对凝胶珠子进行功能化,1101。通过使用含有R1C和随机N-mer1102的模板构建体进行引物延伸反应来附接随机N-mer。整个寡核苷酸链(包括条形码和随机N-mer)的长度为82bp,1101。随机N-mer和R1C的链长度为42个碱基对(bp),1102。在65°C下使用KAPA HIFI RM Master Mix在高引物浓度(10μm)下进行延伸反应1小时。增加降解凝胶珠子步骤之前的洗涤步骤的数目导致样品内引物二聚体的量减少。当不进行洗涤时(1103)可观察到42bp产物1106和80bp产物1107两者。三次洗涤后,引物二聚体的水平相对于无洗涤实验降低(1104)。六次洗涤后(1105),观察到80bp产物1107,但未观察到引物二聚体。

[0522] 还可使用用于延伸步骤的六个不同温度(65°C、67°C、69°C、71°C、73°C、75°C,图11C)进行六次洗涤实验。在这一具体实施例中,使用高引物浓度(10μm)并且延伸步骤持续一小时。似乎67°C是针对优化80bp产物的水平和最小化42bp产物1109的数目两者的最优温度。

[0523] 选择温度67°C用于后续的变性研究。互补链的热变性(其中将样品六次加热至95°C并洗涤以去除互补链)显示,在变性之前有84bp峰1202,在变性之后有降低的峰1201。在1203处示出了从步骤1测得的对照值。

[0524] 实施例10:通过在孔中分配产生条形码化的凝胶珠子

[0525] 根据图13A和13B所示的方法,通过在孔中分配而产生功能化的珠子。图13A中概述了第一功能化步骤,图13B中概述了第二功能化步骤。图13C中概述了示例性多重衔接子产生方案并在实施例11中对该方案进行了描述。如图13A所示,将功能化的珠子1301(例如具有acrydite寡聚物和引物(例如,5'-AAUGAUACGGCGACCACCGAGA-3')的珠子)、具有条形码序列的模板1302(例如,5'-XXXXXXTCTCGGTGGTCGCCGTATCATT-3')和合适的PCR试剂1303混合在一起(1304/1305),并分至多孔板的384个孔中。每一个孔包含独特条形码序列的多个拷贝和多个珠子。伴有延伸反应的热循环1306在每一个单独的孔中进行,以形成具有附接的条形码的珠子。将全部的孔合并在一起并批量清理(1307/1308)。

[0526] 为了添加随机N-mer,将部分功能化的珠子1310、模板随机N-mer寡核苷酸1309和合适的PCR试剂1311混合在一起(1312),并使功能化的珠子1310经历延伸反应1313以将与随机N-mer模板互补的随机N-mer序列添加至珠子中。热循环之后,整体清理珠子(1314-1316)。

[0527] 实施例11:组合平板技术

[0528] 如图13C所示,将附接至引物(例如,P5寡聚体,即5'-AAUGAUACGGCGACCACCGAGA-3')1318的珠子1317分配至多孔板(诸如5X-1 384-孔板1319)的孔中,其中模板1321的多个拷贝包含独特的模板部分条形码序列(例如,5'-XXXXXXTCTCGGTGGTCGCCGTATCATT-3)。进行延伸反应(例如,引物1318经由模板1321的延伸)以在每个孔中生成包含延伸产物(例如,包含引物1318的寡核苷酸以及与模板部分条形码序列互补的部分条形码序列)的珠子-P5-[5X-1]1320。将珠子从孔中移出并合并在一起,并批量进行清理步骤。

[0529] 随后将合并的混合物重新分至第二多孔板1322(诸如具有5X-2的384-孔板)的孔中,其中每一个孔还包含含有第二独特部分条形码序列和随机N-mer(例如,5'P-YYYYYYCG CACACUCUUUCCCUACACGACGCUCUCCGAUC UNNNNNNNN-阻断)的寡核苷酸。该寡核苷酸可具有附接(例如,经由杂交)的阻断寡核苷酸(例如,“阻断”)。在和珠子结合的延伸产物与包含第二部分条形码序列和随机N-mer的寡核苷酸之间进行单链连接反应1324。连接反应之后,生成包含完整条形码序列(例如,XXXXXXXXYYYYY)和随机N-mer的珠子(例如,珠子-P5-[5X-1][5X-2]R1[8N-阻断剂])(1323)。该珠子还包含阻断寡核苷酸。随后将全部的孔合并在一起,切割阻断基团,并批量清理珠子产物。获得了包含大量多样性的条形码序列的珠子。

[0530] 实施例12:用于测序的部分发夹扩增(PHASE)反应

[0531] 用于测序的部分发夹扩增(PHASE)反应是这样一种技术,其可用于根据图14和图15中概述的方法,通过形成部分发夹结构来减少不期望的扩增产物。具体地,用通用序列部分1401/1402/1403标记的约为8N-12N长度的随机引物1404可用于随机地引发并从核酸如基因组DNA(gDNA)延伸。该通用序列包含:(1)用于与测序装置如流通池相兼容的序列(例如,Illumina的P5(1401)和读取1引物位点(1402))和(2)条形码(BC)(1403)(例如,6个碱基的序列)。为了减少这种长通用序列部分的不期望的后果,将含尿嘧啶的核苷酸全部替换为含胸腺嘧啶的核苷酸(除了通用序列部分的最后10-20个核苷酸),并将不接受或处理含尿嘧啶的模板的聚合酶用于核酸的扩增,从而导致关键测序度量的显著改善(图16A、图21和图22)。此外,包含含有尿嘧啶的核苷酸和阻断的3'端(例如,3' ddCTP)的阻断寡核苷酸用于促进通过随机N-mer序列引发核酸并防止与核酸中与读取1引物位点1402互补的部分优先结合。此外,通过包含小百分比的终止核苷酸(例如,0.1-2%的无环核苷酸(acyNTP))进一步限制产物长度(图16B),以减少不期望的扩增产物。

[0532] 本文提供了用于防止扩增不期望的产物的部分发夹形成的实例。首先,在98°C下实现初始变性2分钟,然后通过用作引物的随机N-mer序列在4°C下引发基因组DNA序列的随机部分30秒(图15A)。随后,随着温度以0.1°C/秒升至45°C(保持1秒)发生序列延伸(图15A)。延伸在升高的温度(20秒时在70°C)下继续,从而继续置换上游链并产生第一阶段的冗余(图15B)。变性在98°C下发生30秒,以释放用于附加引发的基因组DNA。第一循环之后,扩增产物具有单一的5'尾(图15C)。例如通过在4°C下开始第2个循环和使用随机N-mer序列重复上述提到的那些步骤高达20次,以再次引发基因组DNA,其中黑色序列指示添加的5'尾(在第1个循环中添加的)中不能被拷贝的部分(图15D)。变性在98°C下发生,以再次将基因组DNA和扩增产物从第一循环中释放,以用于附加引发。在第二轮热循环后,5'标记的产物以及3'和5'标记的产物均存在(图15E)。由3'和5'标记的产物形成部分发夹结构,从而防止不期望的产物的扩增(图15F)。基因组DNA序列的新的随机引发再次在4°C下开始(图15G)。

[0533] 实施例13:通过扩增添加附加序列

[0534] 为了完成测序仪就绪文库,完成附加扩增(例如,聚合酶链反应(PCR)步骤)以添加附加序列(图14C)。为了竞争去除(out-compete)发夹,使用含引物的锁核酸(LNA)或锁核酸核苷酸。此外,在先前步骤中采用包含含尿嘧啶的核苷酸的情况下,则将不能区别含模板尿嘧啶的核苷酸的聚合酶用于该步骤。图17中所示的结果显示阻断寡核苷酸减少了起始位点偏倚,如通过在Illumina MiSeq测序仪上测序所测得的。在该情况下的核酸模板为酵母gDNA。

[0535] 实施例14:数字处理器

[0536] 图18中示出了示例控制组件1801的概念示意图。计算机1802充当控制组件1801的中央集线器。计算机1802与显示器1803、一个或多个输入设备(例如,鼠标、键盘、照相机等)1804和任选的打印机1805通信。控制组件1801经由其计算机1802与以下的一个或多个装置通信:任选的样品预处理单元1806、一个或多个样品处理单元(诸如测序仪、热循环仪或微流体装置)1807和任选的探测器1808。控制组件可例如经由以太网连接来联网。用户可使用输入设备1804向计算机1802提供输入(例如,对于期望的一组核酸扩增反应所必需的参数或微流体装置的流速)。该输入由计算机1802解释以生成指令。计算机1802将这样的指令传递至任选的样品预处理单元1806、一个或多个样品处理单元1807和/或任选的探测器1808以供执行。而且,在任选的样品预处理单元1806、一个或多个样品处理单元1807和/或任选的探测器1808的操作过程中,每一个装置均可将信号回传至计算机1802。这样的信号可由计算机1802解释并使用,以确定任何装置是否需要进一步的指令。计算机1802也可调节样品预处理单元1806,使得样品的组分适当地混合并以期望的或其他预定的速率进料至样品处理单元(诸如微流体装置)1807。计算机1802还可与探测器1808通信,使得探测器在期望的或其他预定的时间点或在由从预处理单元1806或样品处理单元1807接收的反馈所确定的时间点执行测量。探测器1808还可将测量过程中获得的原始数据回传至计算机1802,以供进一步分析和解释。分析可经由显示器1803和/或由打印机1805生成的打印输出以对终端用户有用的格式进行汇总。用于控制样品预处理单元1806、样品处理单元1807和/或探测器1808的指令或程序;通过执行本文所述的任何方法而获得的数据;或经分析和/或解释的数据,可经由网络1810(例如可以是因特网)传输至一个或多个远程计算机1809或从一个或多个远程计算机1809进行接收。

[0537] 实施例15:经由连接的组合技术

[0538] 如图23A所示,生成珠子2301并且它与部分P5序列2302共价连接(例如,经由acrydite部分)。在4个96孔板的每个孔的50L中,分别包含剩余的P5序列和独特的部分条形码序列(在寡核苷酸2303中用碱基“DDDDDD”表示)的寡核苷酸2303与寡核苷酸2304杂交,该寡核苷酸2304包含寡核苷酸2303的反向互补序列和悬于寡核苷酸2303的每个末端之上的附加碱基。生成了夹板2306。用3' C3间隔区、3' 反向dT或双脱氧-C(ddC)阻断每一个突出端(在图23中用“X”表示),以防止副产物形成。

[0539] 如图23B所示,分别将夹板2306添加至4个96深孔板中,其中每一个孔包含2mL珠子2301以及包含独特的部分条形码序列的夹板。在每一个孔中,夹板2306经由寡核苷酸2304的相应突出端与珠子2301的部分P5序列2302杂交。杂交后,在16°C下,部分P5序列2302与寡核苷酸2303(其通常已经5' 磷酸化)经由连接酶例如T4连接酶的作用连接1小时。连接后,合并产物并洗涤珠子以去除未连接的寡核苷酸。

[0540] 如图23C所示,随后将洗涤后的产物重新分布至4个新的96孔板的孔中,其中该板的每一个孔包含2mL的珠子2301和寡核苷酸2305,该寡核苷酸2305具有独特的部分条形码序列(在寡核苷酸2305中用“DDDDDD”表示)以及与寡核苷酸2304的剩余突出端互补的相邻短序列(例如,与部分条形码序列相邻并且在寡核苷酸2305末端的“CC”)。寡核苷酸2305还包含随机N-mer(在寡核苷酸2305中用“NNNNNNNNNN”表示)。寡核苷酸2305经由相邻的短序列通过寡核苷酸2304的剩余的突出端与寡核苷酸2304杂交。随后在16℃下,寡核苷酸2305与寡核苷酸2303经由连接酶的作用连接1小时。寡核苷酸2305与寡核苷酸2303的连接导致完整条形码序列的生成。如图23D所示,随后合并产物,使寡核苷酸2304从产物中变性,并随后洗掉未结合的寡核苷酸。洗涤之后,获得了条形码化的珠子的多样性文库,其中与寡核苷酸结合的每一个珠子包含P5序列、完整条形码序列和随机N-mer。生成的文库包含约147,000个不同的条形码序列。

[0541] 实施例16:在条形码引物中用含尿嘧啶的核苷酸置换含胸腺嘧啶的核苷酸

[0542] 如图33A所示,使用适合于PHASE扩增的两个条形码引物3301和3302来扩增从酵母基因组中获得的样品核酸。PHASE扩增之后,添加附加序列(例如,经由批量PCR)以生成测序仪就绪产物。条形码引物3301(在图33A中也显示为U.2)和3302(在图33A中也显示为U.1)包含相同的序列(除了条形码引物3301在位置3306处包含附加的含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的置换)。针对每一个条形码引物运行扩增实验组,其中每组对应于与各自的条形码引物以不同的化学计量比混合的特定阻断寡核苷酸。对于条形码引物3302,进行对应于标准阻断寡核苷酸3303、包含桥核酸(BNA)的完整阻断寡核苷酸3304(在图33A中也显示为BNA阻断剂)或完整阻断寡核苷酸3305的扩增实验组。阻断寡核苷酸3303和3305在全部含胸腺嘧啶的核苷酸位置和ddC阻断的末端处包含含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的置换。在每组中,阻断寡核苷酸:条形码引物化学计量比为0、0.4、0.8或1.2。对于条形码引物3301,在阻断寡核苷酸:条形码引物化学计量比为0.8时测试每种类型的阻断寡核苷酸3303、3304和3305。

[0543] 图33B中示出了PHASE扩增产物的大小结果。如图所示,与阻断寡核苷酸3303偶联的条形码引物3302(例如,其包含额外的含尿嘧啶的核苷酸取代含胸腺嘧啶的核苷酸的置换)通常在所测试的化学计量比中产生最小的扩增产物。条形码引物3302对于阻断寡核苷酸3304和3305的结果发生变化,其大小通常大于对于阻断寡核苷酸3303的结果。对于条形码引物3301,在所测试的阻断寡核苷酸中扩增产物大小也通常大于针对与阻断寡核苷酸3303偶联的条形码引物3301所获得的结果。图33C中示出了测序仪就绪产物的大小结果。

[0544] 图33D中示出了从扩增产物中获得的关键测序度量。如图所示,对于由条形码引物3302生成的扩增产物的测序运行,未定位的读取结果的分数(图33D中的I图)通常较低。例如,对于由条形码引物3302和阻断寡核苷酸3303以0.8的阻断寡核苷酸:条形码引物化学计量比生成的扩增产物,未定位的读取结果的分数为约7-8%,而使用条形码引物3301在相同条件下获得的结果为约17-18%。而且,对于条形码引物3302,Q40错误率(图33D中的II图)也较低。例如,由条形码引物3302和阻断寡核苷酸3303以0.8的阻断寡核苷酸:条形码引物化学计量比生成的扩增产物的Q40错误率为约0.105%,而使用条形码引物3301在相同条件下获得的结果为约0.142%。图33E中示出了在测序过程中确定的读取1起始位点(III图)和读取2起始位点(IV图)相对熵。

[0545] 实施例17:凝胶珠子经由二硫键交换的合成功能化

[0546] 根据本文所述的一种或多种方法生成包含二硫键的凝胶珠子。然后使凝胶珠子与TCEP以多个TCEP分子与凝胶珠子之比(TCEP:GB)反应。测试的比值为0、25亿和100亿。TCEP充当还原剂以在凝胶珠子内生成游离的巯基。还原后,洗涤凝胶珠子一次以去除TCEP。接着,使生成的凝胶珠子的游离巯基与acrydite-S-S-P5物质(例如,图35A中的3505)反应,以使acrydite-S-S-P5经由如图35A所示的Michael加成化学与凝胶珠子连接。测试acrydite-S-S-P5与每一种类型的活化的凝胶珠子的不同比值(例如,用于在凝胶珠子上生成游离巯基的TCEP:GB比)。测试的acrydite-S-S-P5物质与活化的凝胶珠子之比(P5:GB)为5000万、5亿和50亿。

[0547] 合成后,将来自每一个反应的凝胶珠子洗涤,并在反应混合物中用DTT处理以降解凝胶珠子并释放任何结合的acrydite-S-S-P5物质。使每一个反应混合物的等份进入凝胶泳道,并使游离的寡核苷酸经历如图36所示的凝胶电泳(例如,图36中的泳道3-11)。还使50皮摩尔acrydite-S-S-P5标准(例如,图36中的泳道1)与25个碱基对的序列梯(例如,图36中的泳道2)一起运行。对应于加载的acrydite-S-S-P5的条带在泳道5和8中生成(图36中的箭头所指示的)。泳道5对应于以25亿的TCEP:GB比处理的凝胶珠子以及与acrydite-S-S-P5以50亿的P5:GB比反应的TCEP处理的凝胶珠子。泳道8对应于以100亿的TCEP:GB比处理的凝胶珠子以及与acrydite-S-S-P5以50亿的P5:GB比反应的TCEP处理的凝胶珠子。

[0548] 实施例18:凝胶珠子经由二硫键交换的合成功能化

[0549] 根据本文所述的一种或多种方法生成包含二硫键的凝胶珠子。随后使凝胶珠子与在0.1M磷酸盐缓冲液中的TCEP以4 μ g TCEP/100,000个凝胶珠子的浓度反应。TCEP可充当还原剂,以生成具有游离巯基的凝胶珠子。还原后,洗涤凝胶珠子一次以将凝胶珠子与TCEP分离。接着,使凝胶珠子的游离巯基与在2,2'-二硫吡啶的饱和溶液(\sim 0.2mM)中的2,2'-二硫吡啶(例如,图35B中的3507)反应,以使吡啶基团与凝胶珠子经由如图35B中所示的二硫键交换化学而连接。合成后,洗涤凝胶珠子三次,以去除过量的2,2'-二硫吡啶。

[0550] 然后,使洗涤后的凝胶珠子在一端与包含完整构建体条形码(FCBC-例如,包含P5、条形码序列、R1和随机N-mer的寡核苷酸)序列的寡核苷酸3702反应,而在其另一端与游离巯基反应。两个反应以两个不同的FCBC分子与凝胶珠子之比(例如,FCBC:GB)完成,并使反应过夜进行。测试的FCBC:GB比为4亿和16亿。最初供应的寡核苷酸3702,其游离巯基在二硫键中受保护,如图37A中的3701所示。为了生成如寡核苷酸3702中的游离巯基,将寡核苷酸3701采用在1X Tris-EDTA缓冲液(TE)中的0.1M DTT处理30分钟。在Sephadex(NAP-5)柱上的盐交换用于在还原之后去除DTT并纯化寡核苷酸3702。对于每一个反应,随后使纯化的寡核苷酸3702与凝胶珠子的二硫吡啶物质通过巯基二硫键交换(例如,参见图35B)反应,以生成包含寡核苷酸3702的凝胶珠子。反应后,通过洗涤珠子三次来纯化凝胶珠子。

[0551] 为了比较的目的,还通过如本文别处所述的单体聚合来生成包含二硫键和FCBC序列的凝胶珠子。FCBC与包含能够参与与丙烯酰胺和双(丙烯酰)胱胺的聚合的acrydite物质的单体连接,以生成凝胶珠子。FCBC序列与凝胶珠子经由acrydite部分连接。

[0552] 合成后,将来自每一个反应的凝胶珠子洗涤,并在反应混合物中用DTT处理以降解凝胶珠子并释放任何结合的寡核苷酸3702。经由聚合而合成的包含FCBC序列的凝胶珠子也在反应混合物中用DTT进行处理。使每一个反应混合物的等份进入凝胶泳道,并使游离的寡

核苷酸经历如图37B所示的凝胶电泳。如图37B中示出的凝胶照片所示，泳道1对应于50个碱基对的分子量标记(ladder)；泳道2对应于经由二硫键交换化学以4亿的FCBC:GB比功能化的凝胶珠子；泳道3对应于经由二硫键交换化学以16亿的FCBC:GB比功能化的凝胶珠子；泳道4对应于经由acrydite物质的聚合而生成的功能化的凝胶珠子。对于以两种FCBC:GB比生成的功能化的凝胶珠子生成了对应于加载的寡核苷酸的条带，并且所述条带位于与针对经由acrydite物质的聚合而生成的功能化的凝胶珠子所生成的条带相似的位置。

[0553] 合成后，还将来自每一个反应的凝胶珠子洗涤并用SYBR Gold荧光染色剂染色。经由聚合而合成的包含FCBC序列的凝胶珠子也用SYBR Gold染色。SYBR Gold可通过嵌入任何结合的寡核苷酸来对功能化的珠子染色。染色后，将珠子合并，并使用荧光显微术成像，如在图37C中示出的显微照片所示的。图37C中较亮的珠子(3704)对应于在珠子的聚合过程中功能化的珠子，而暗淡的珠子(仍显示SYBR Gold信号)(3705)对应于在凝胶珠子生成之后用二硫键交换化学功能化的珠子。经由二硫键交换的寡核苷酸的加载量为采用在凝胶珠子聚合过程中将珠子功能化所实现的加载量的约30%。

[0554] 根据上文应当理解，虽然已示出和描述了特定实施方式，但可以对其进行各种修改，并且这些修改是本发明所设想到的。本说明书中提供的具体实例并非意图限制本发明。虽然已参照上述说明书对本发明进行了描述，但此处的优选实施方案的说明和图示并不意味着以限制性的意义来解释。此外，应当理解，本发明的所有方面均不限于本文阐述的具体描述、构造或相对比例，这些取决于各种条件和变量。对本发明实施方案的形式和细节的各种修改对本领域技术人员来说将是显而易见的。因此可以预期，本发明也应涵盖任何这样的修改、变化和等同物。意图以所附权利要求限定本发明的范围，由此涵盖这些权利要求的范围内的方法和结构及其等同物。

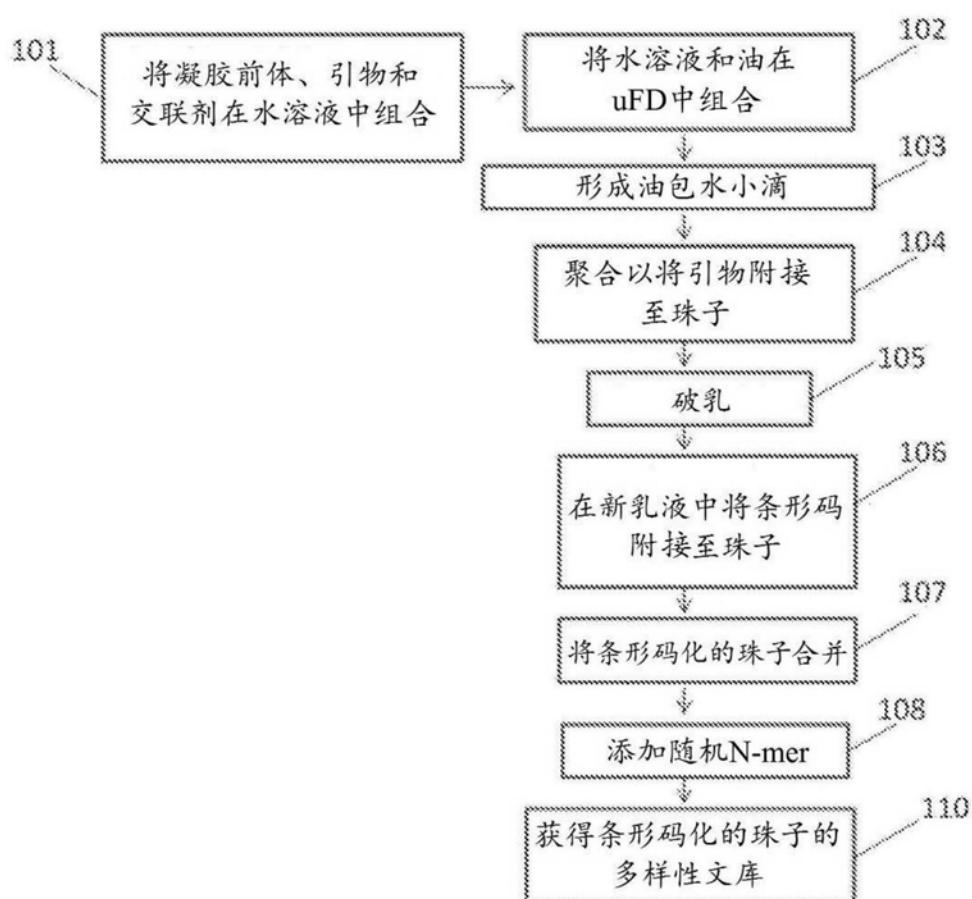


图1A

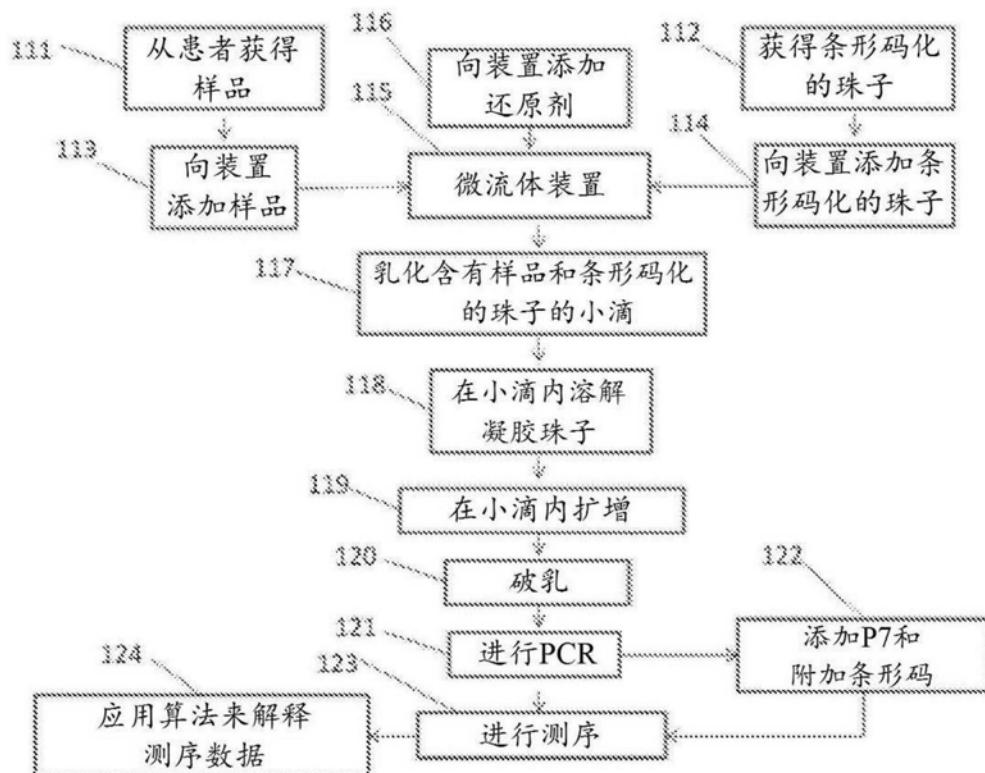


图1B

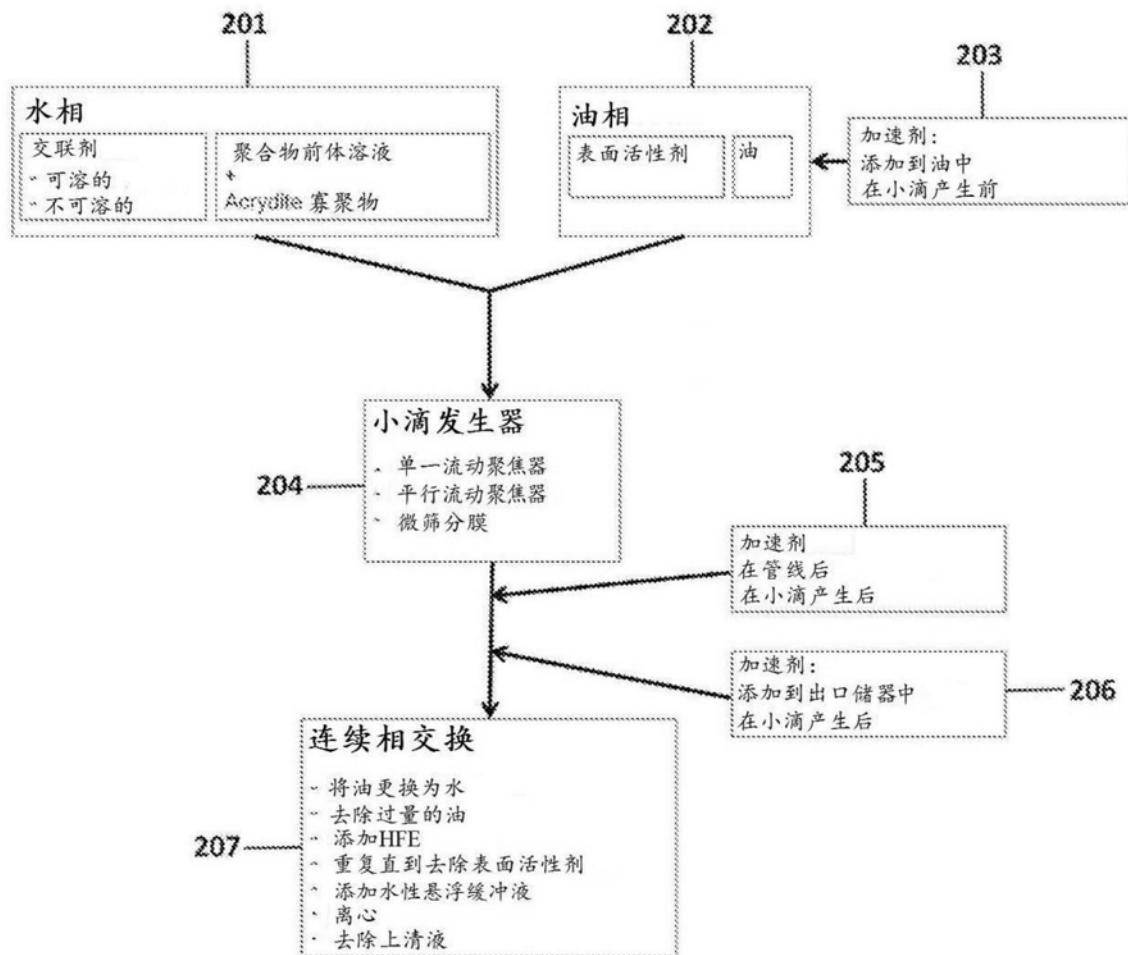


图2

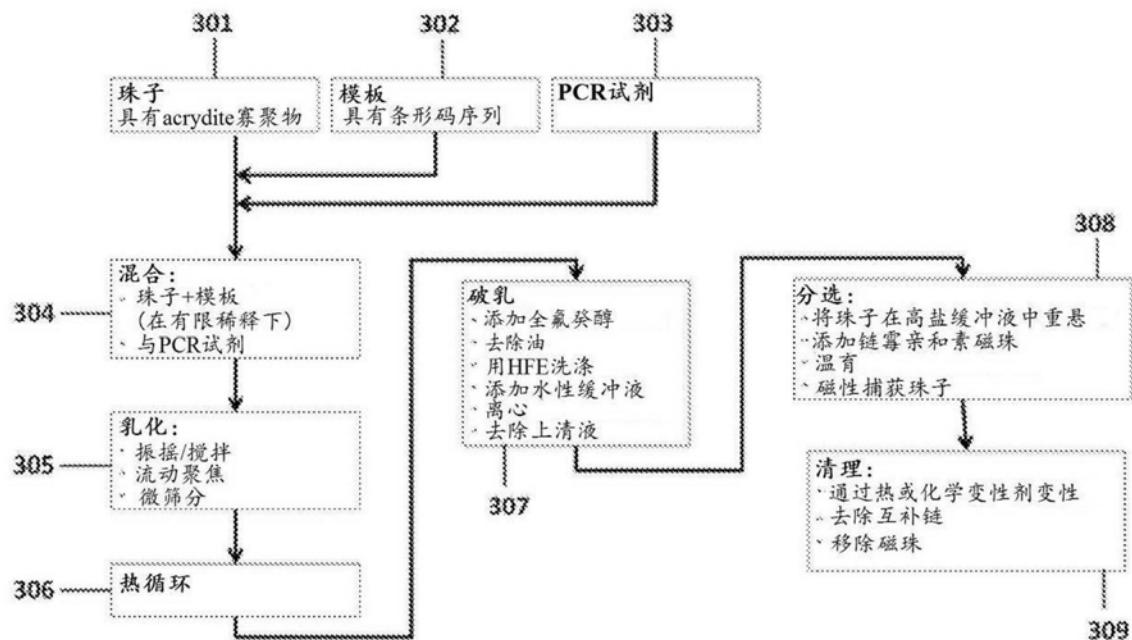


图3A

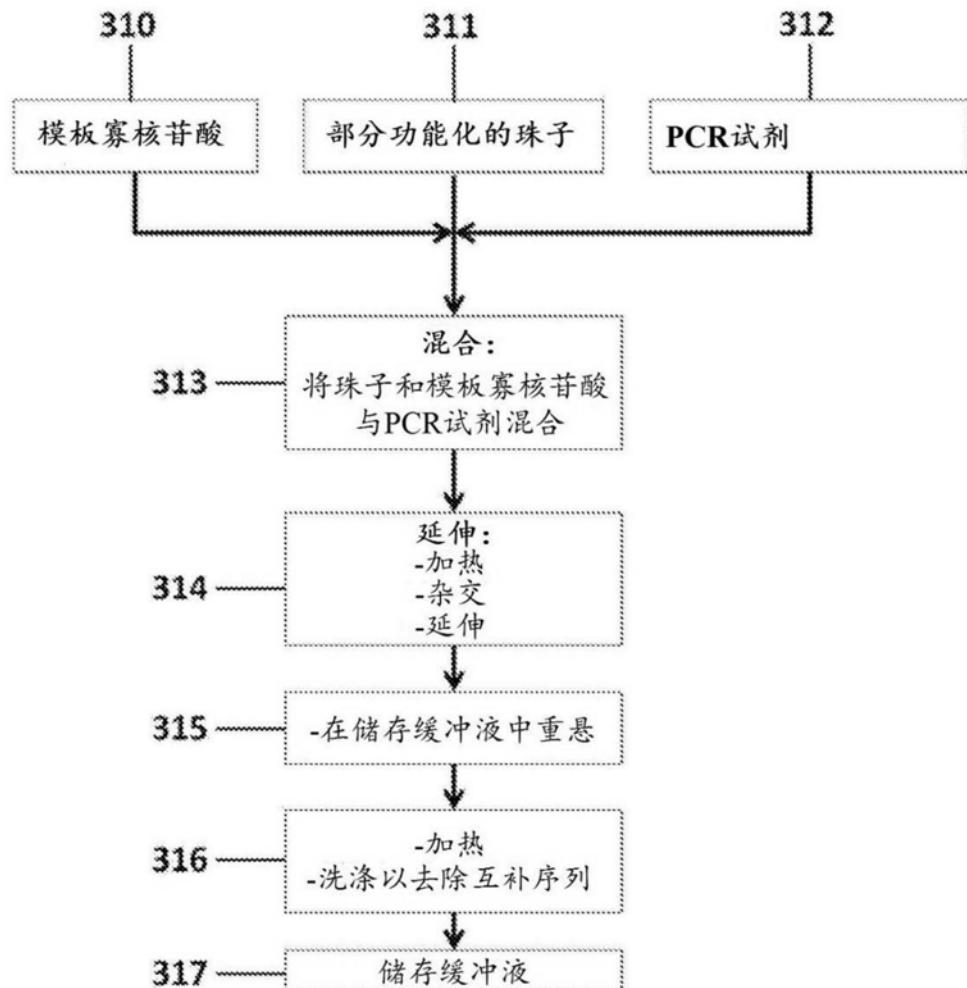
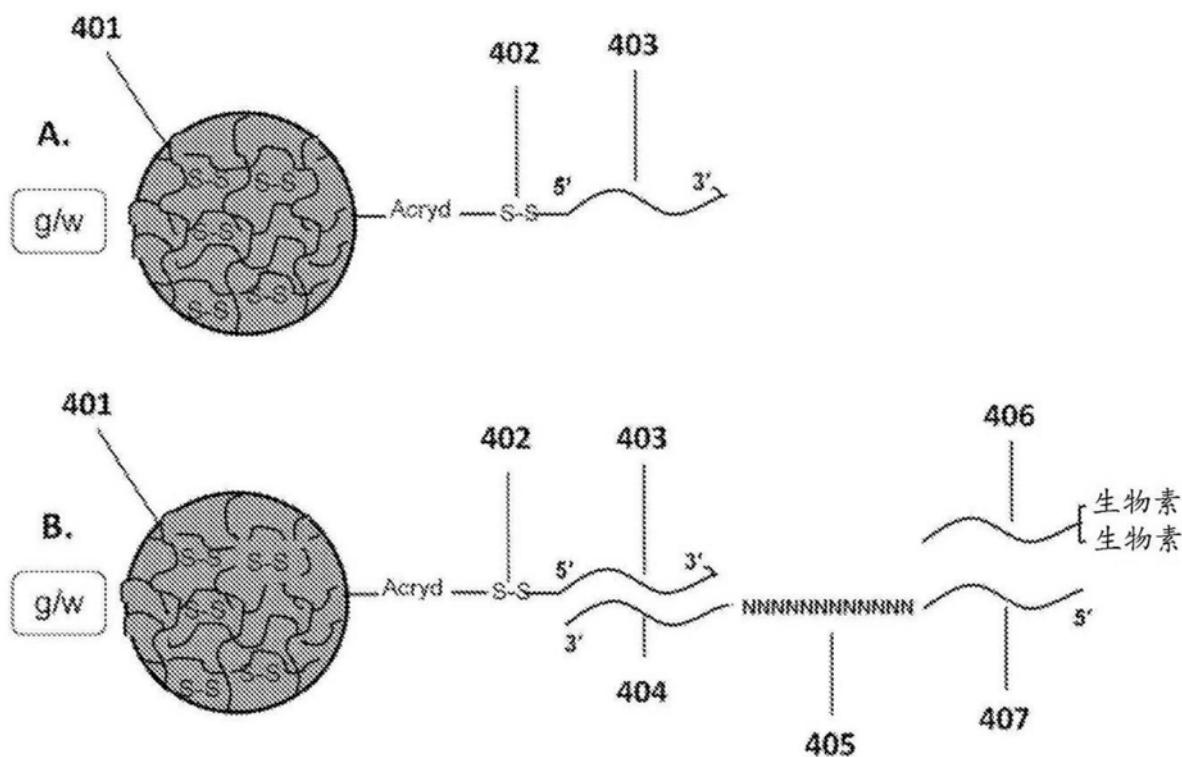
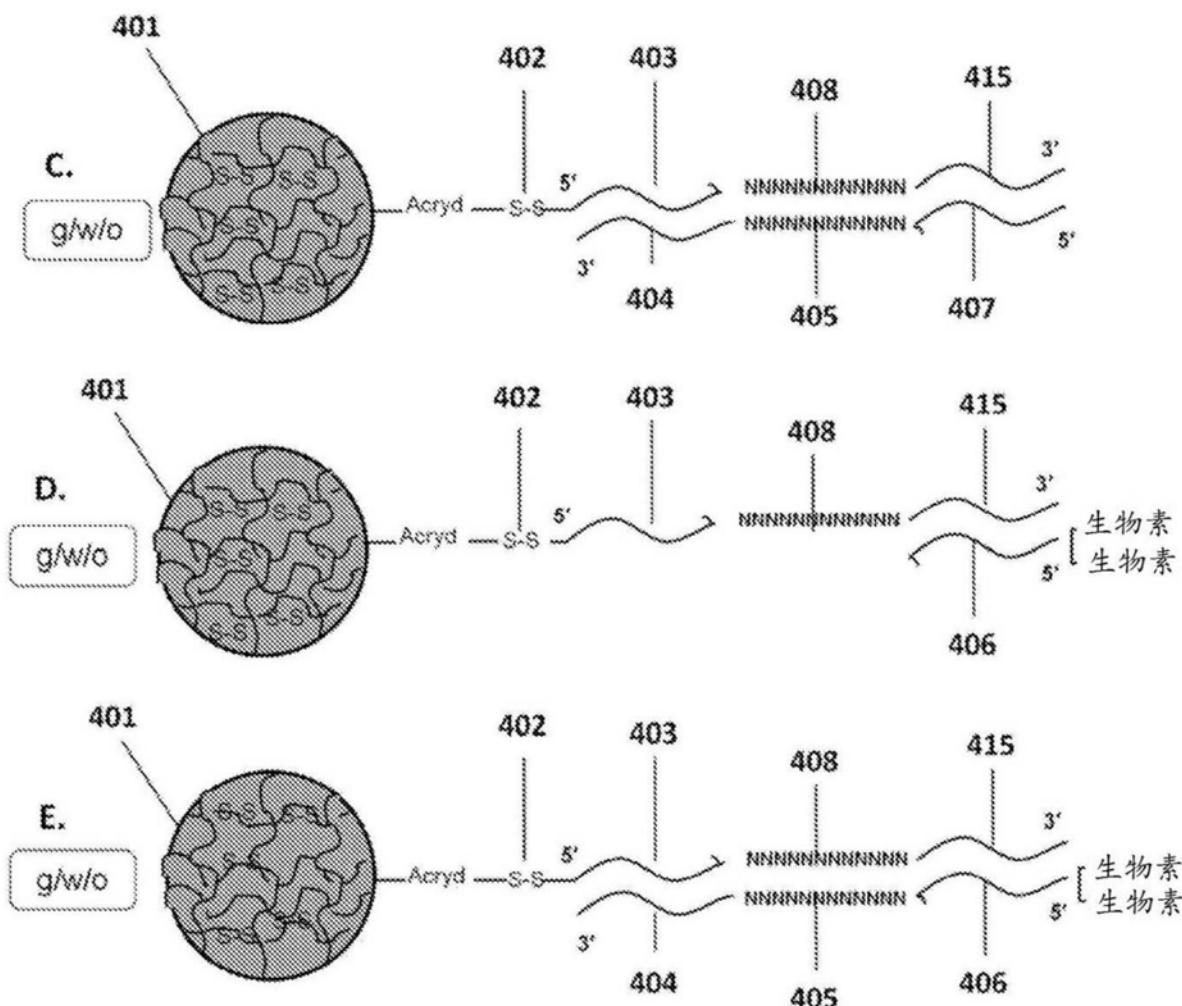
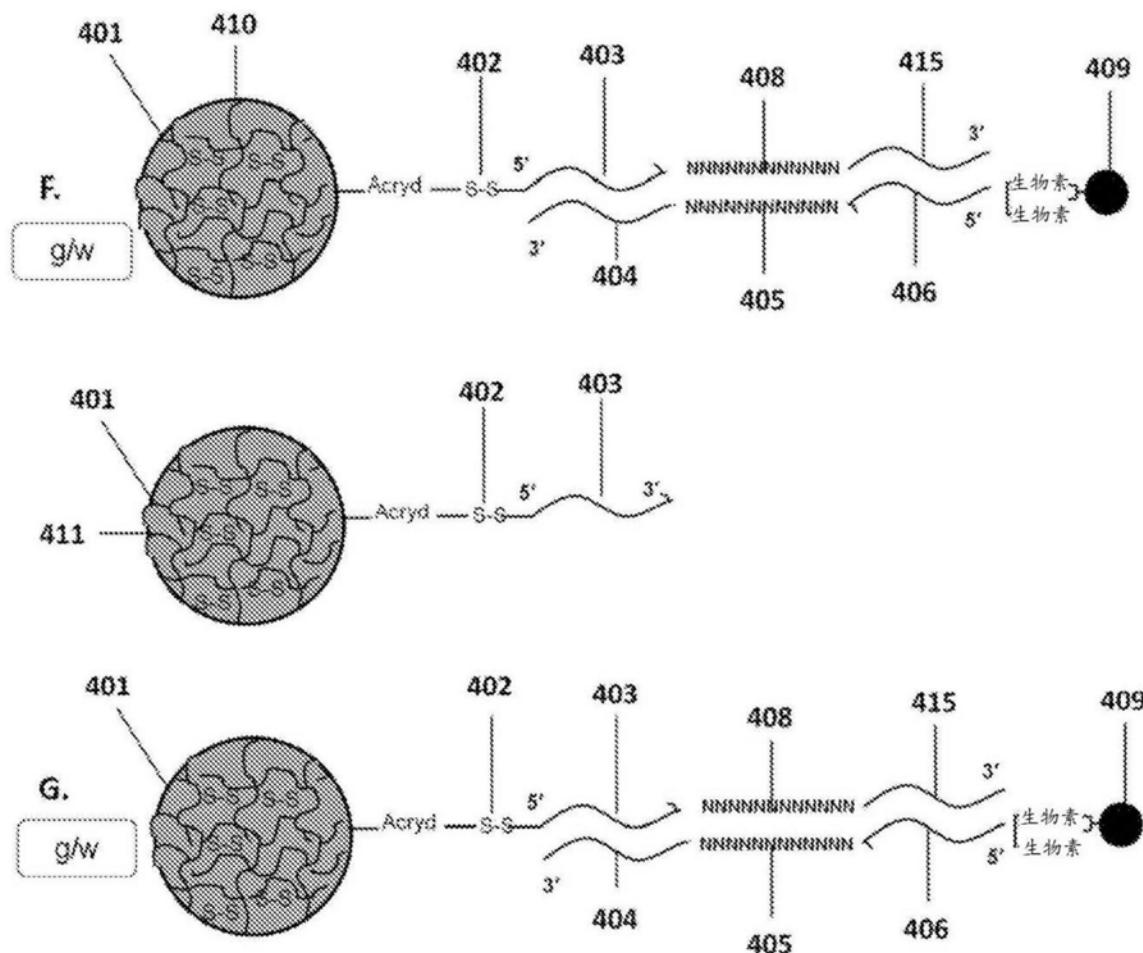
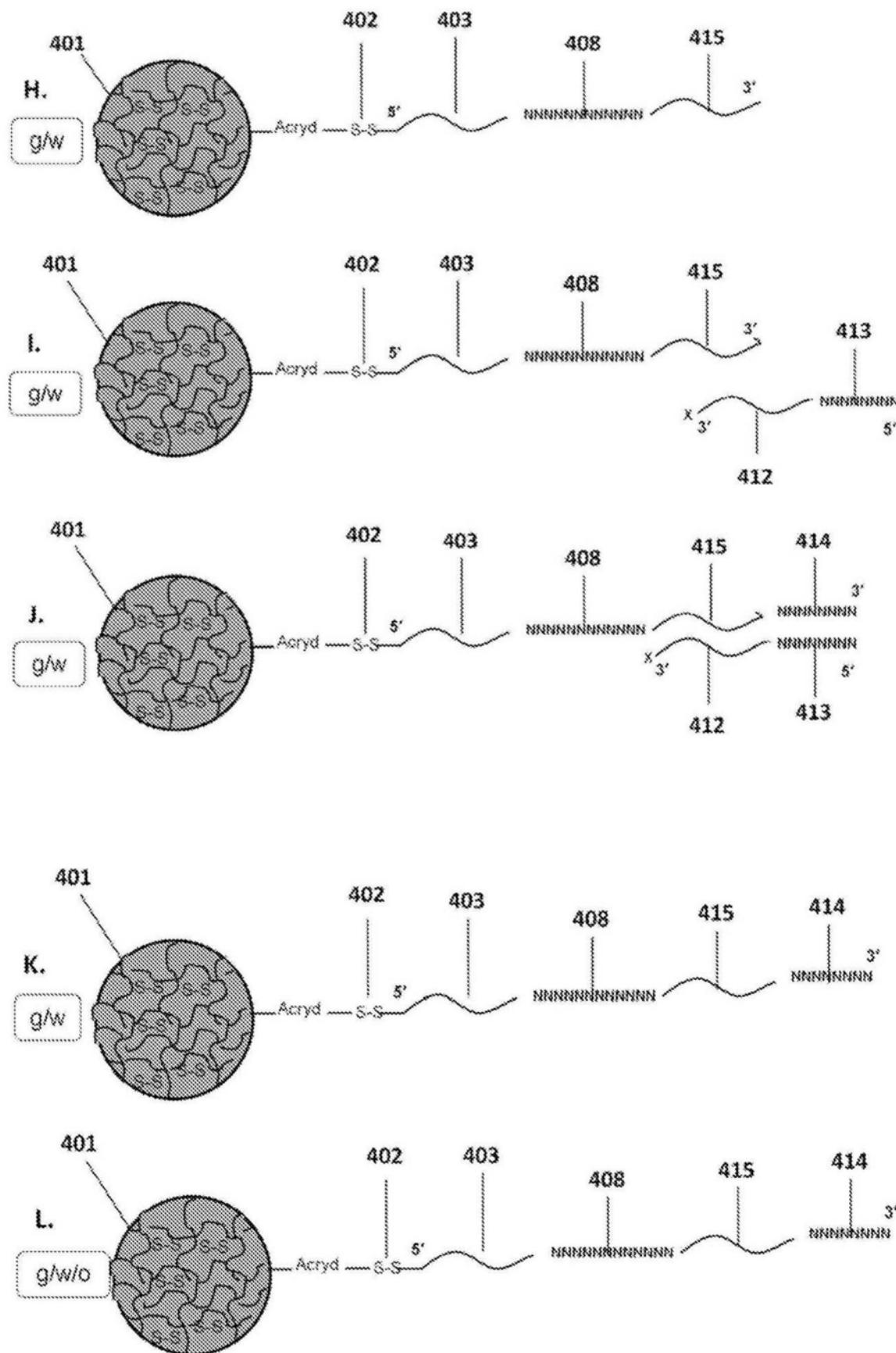


图3B









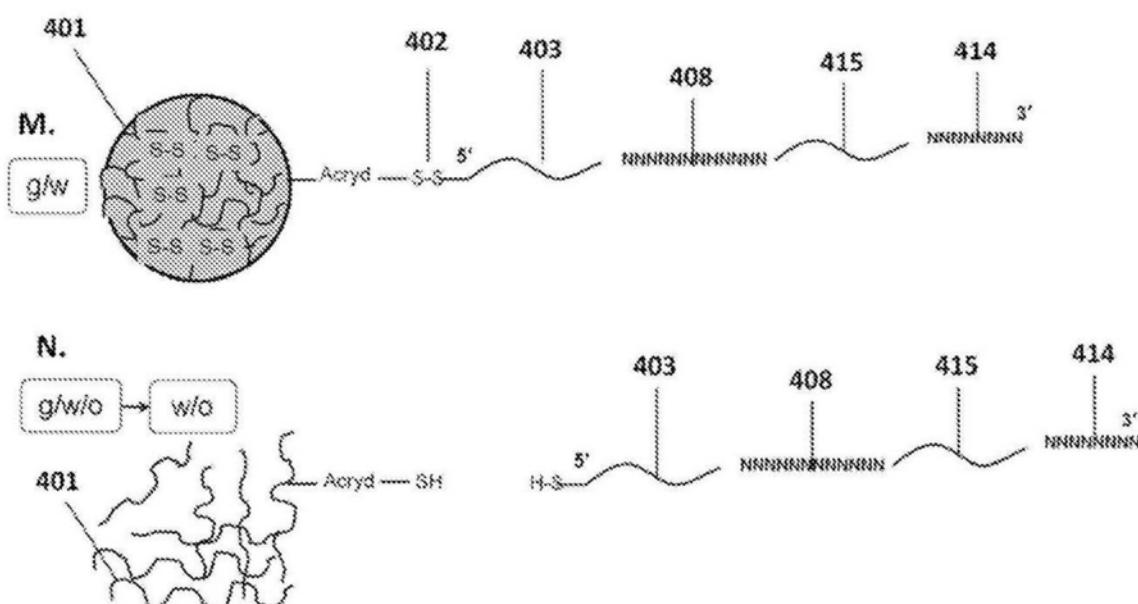


图4

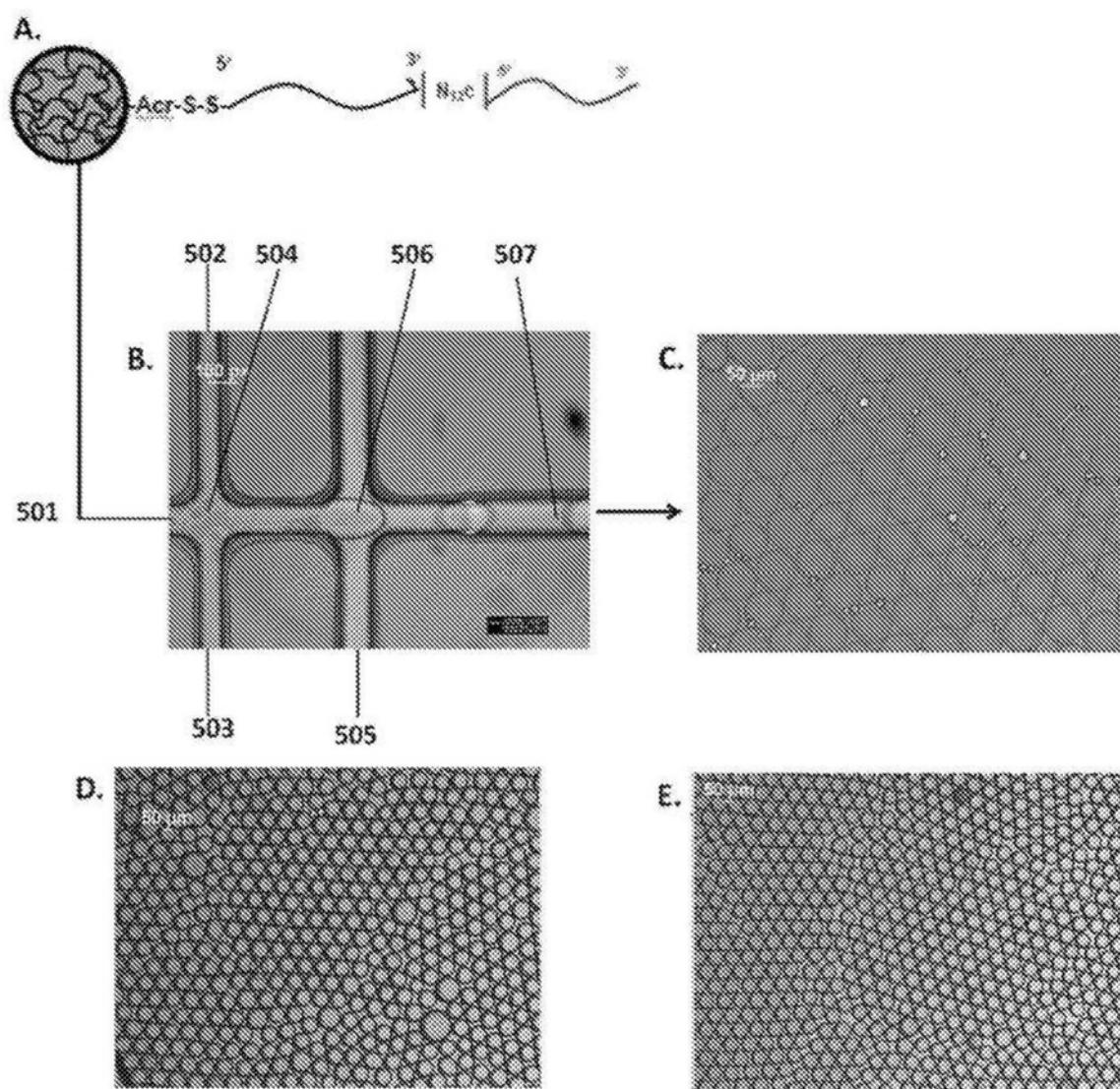


图5

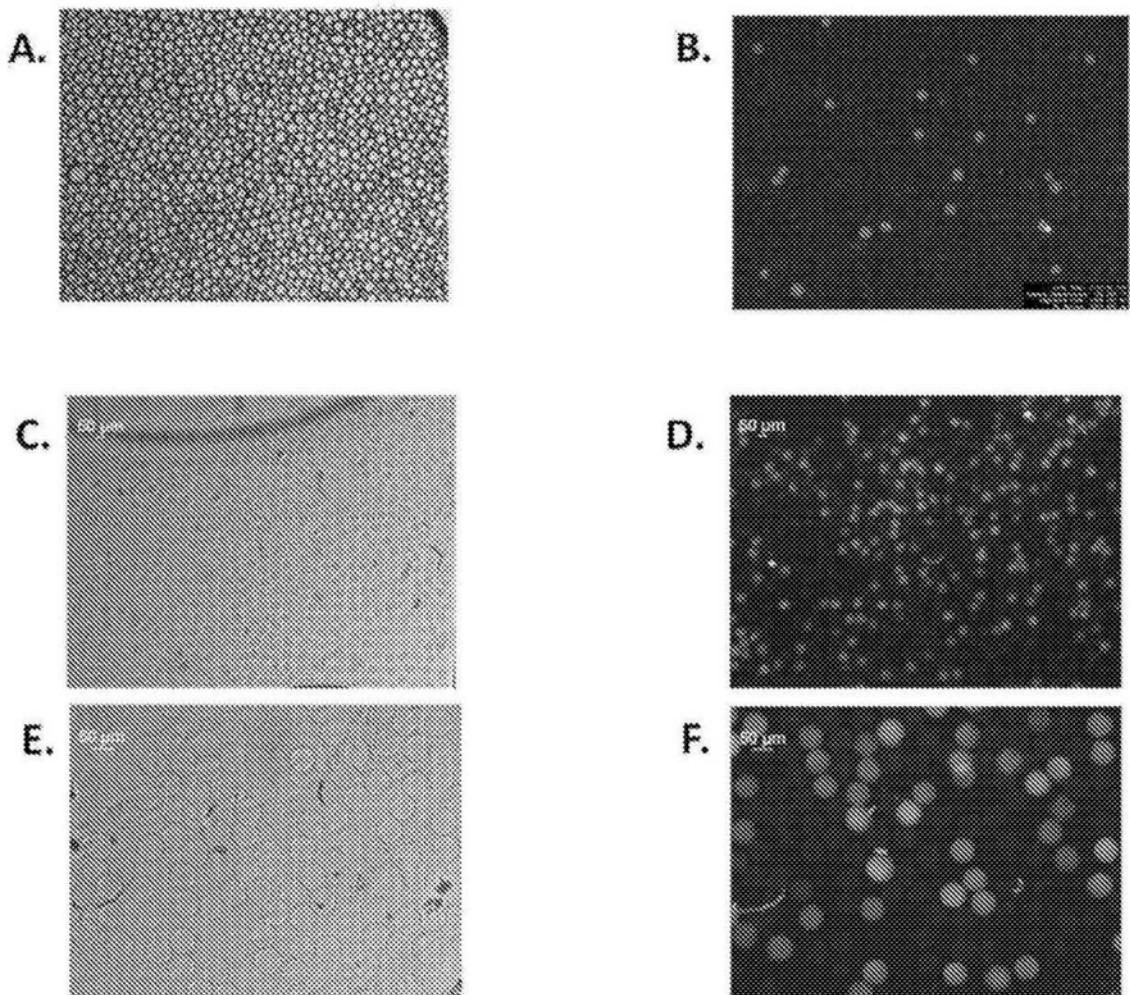


图6

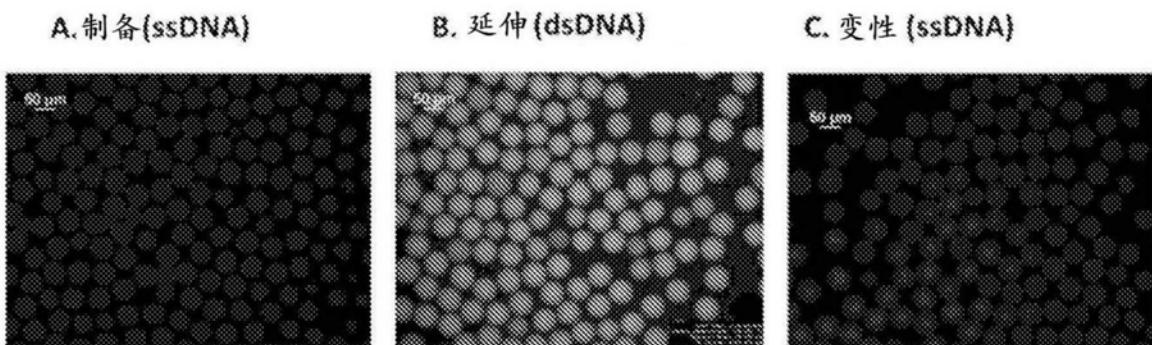


图7

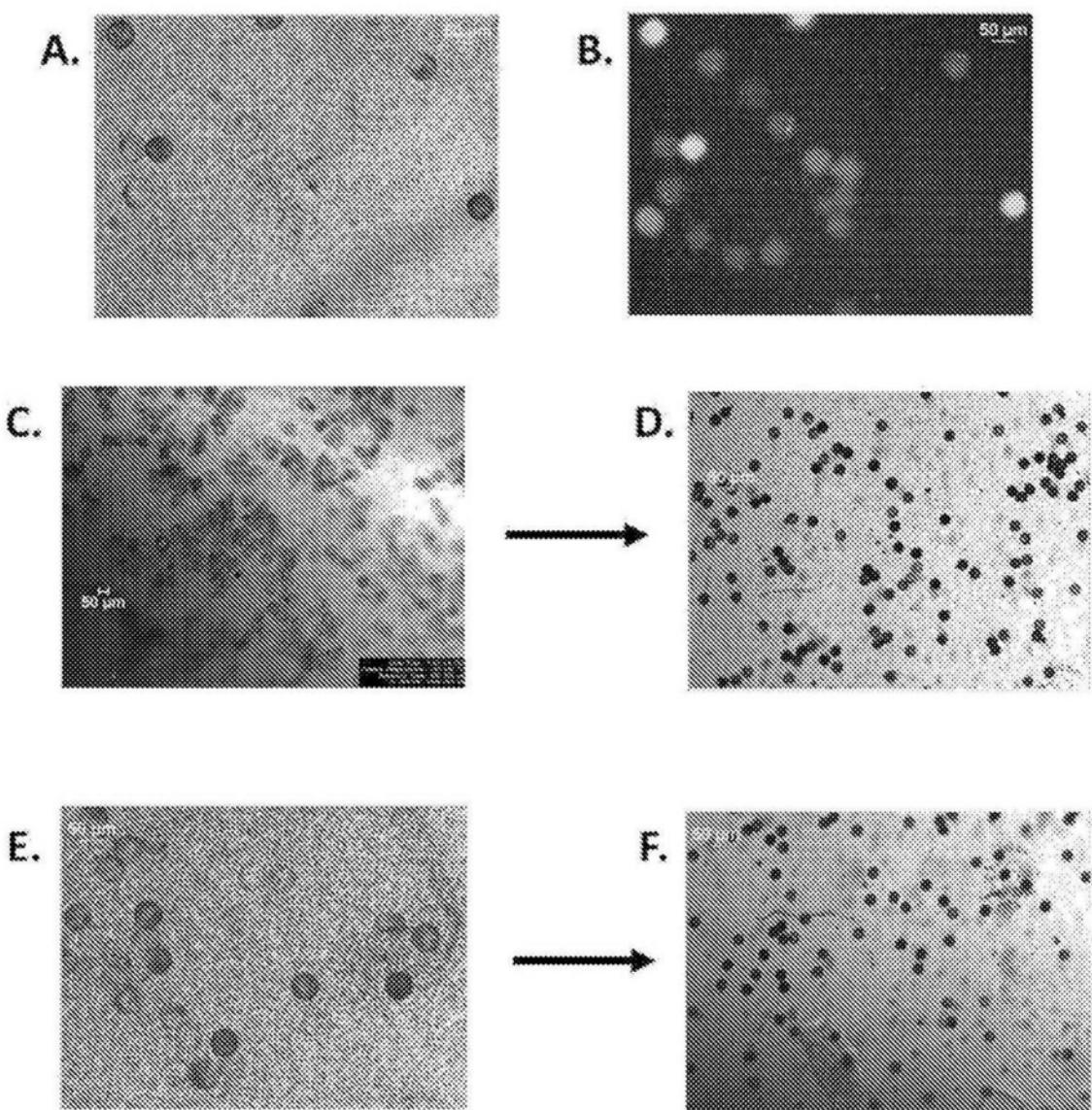


图8

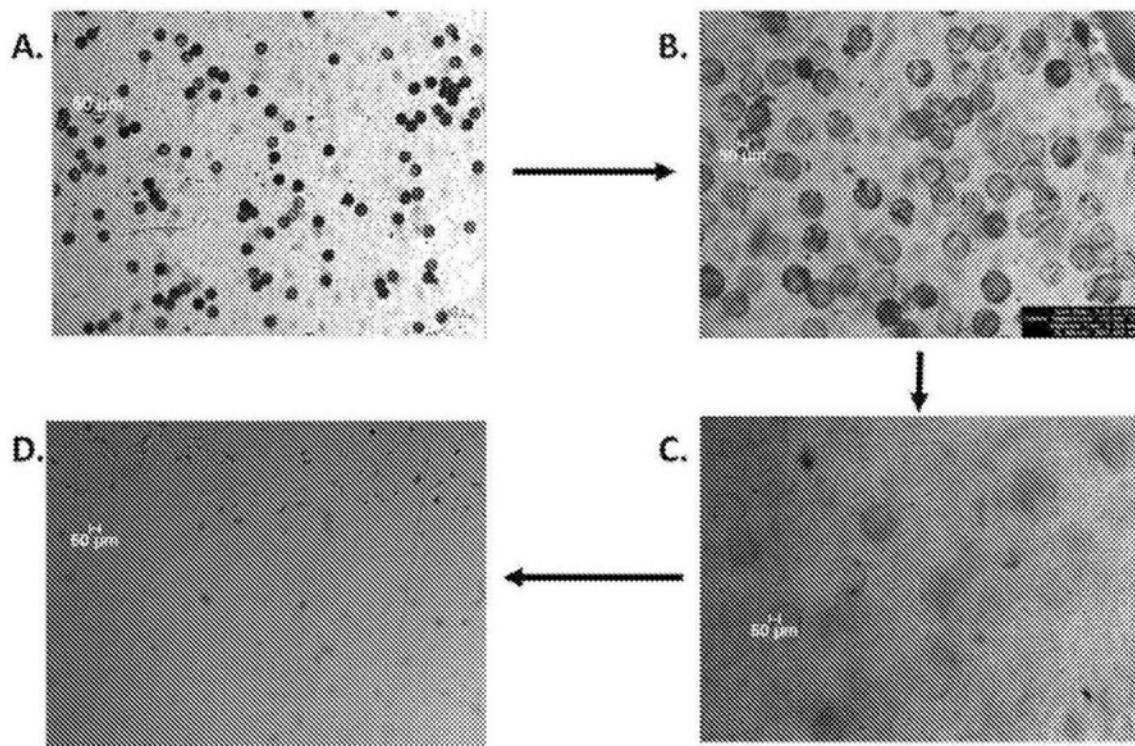


图9

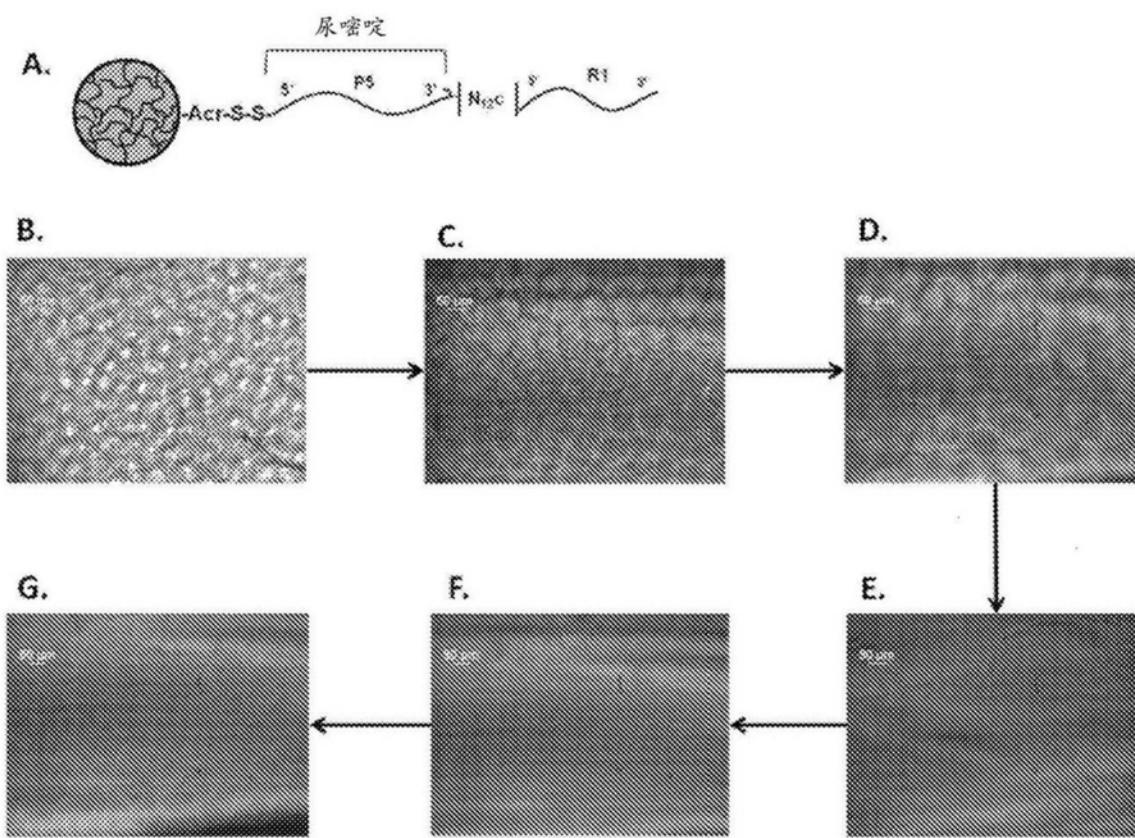
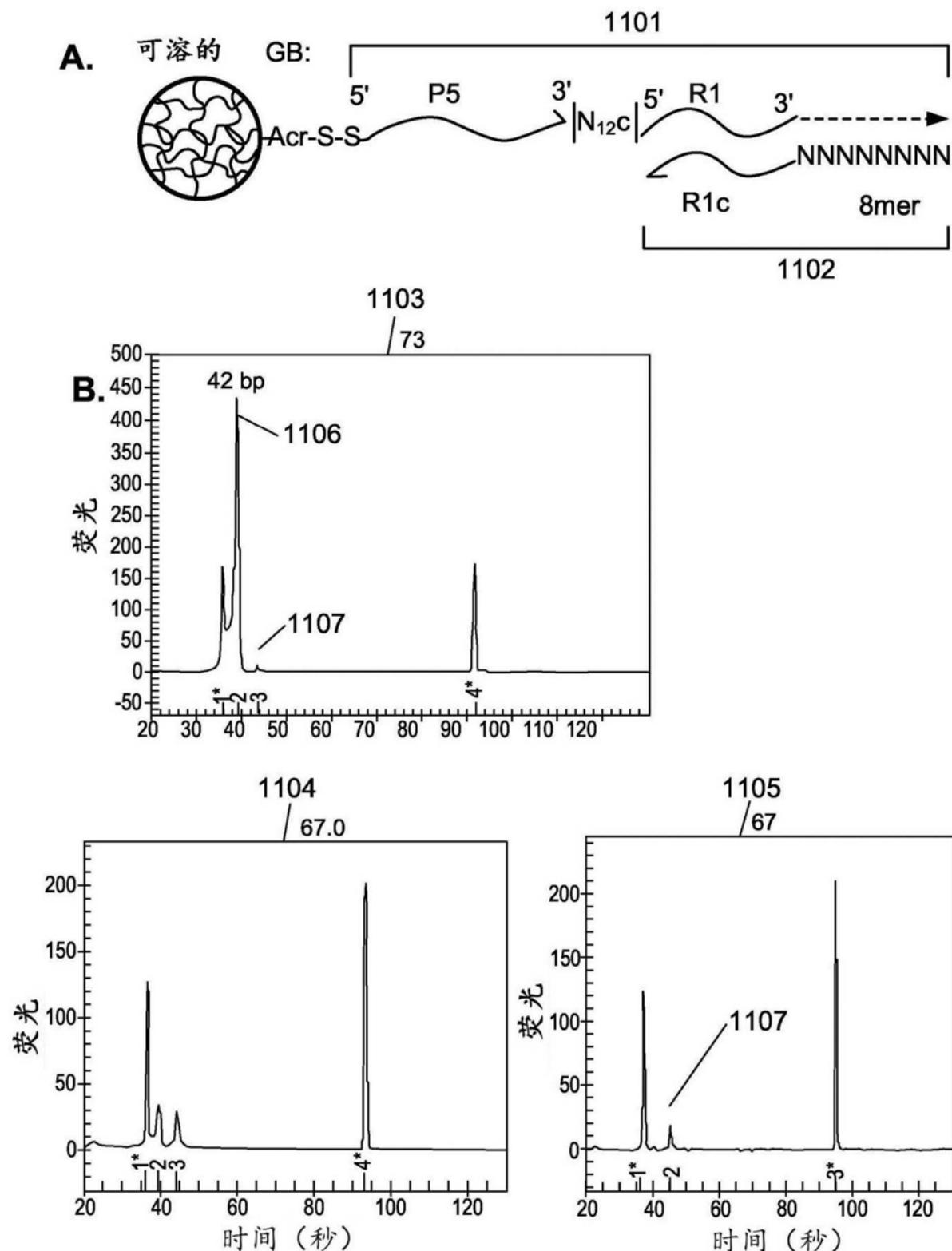


图10



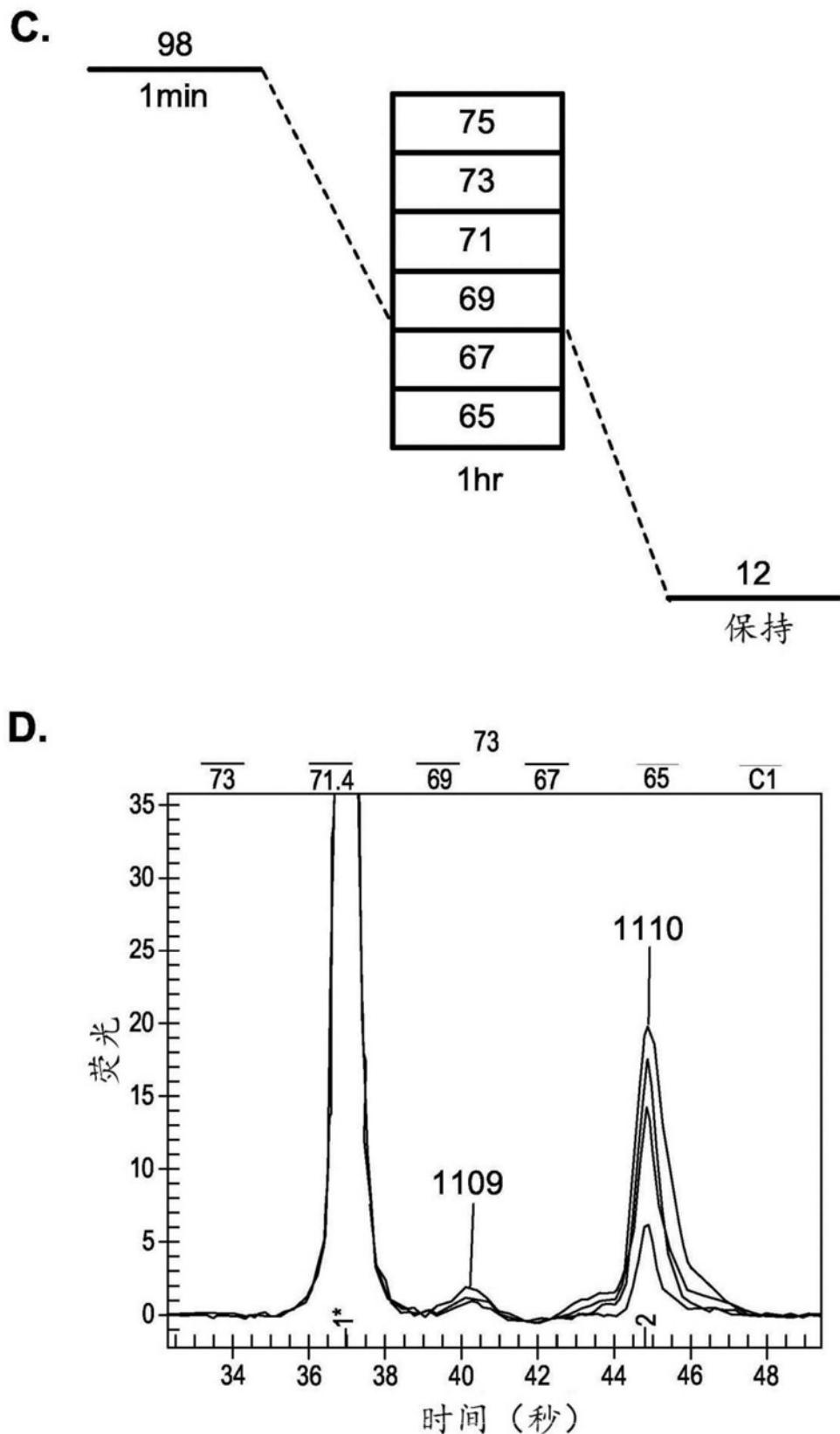


图11

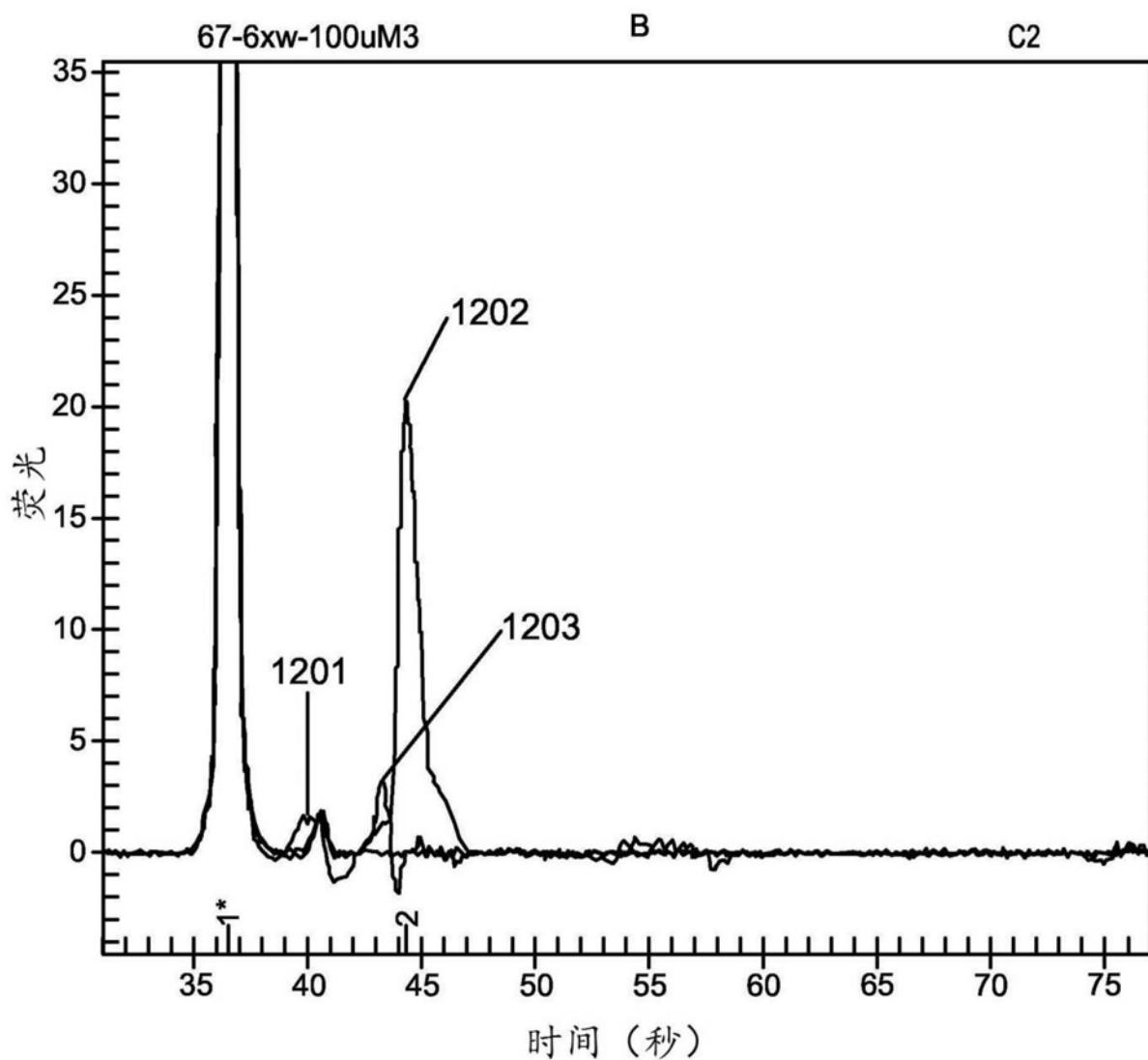


图12

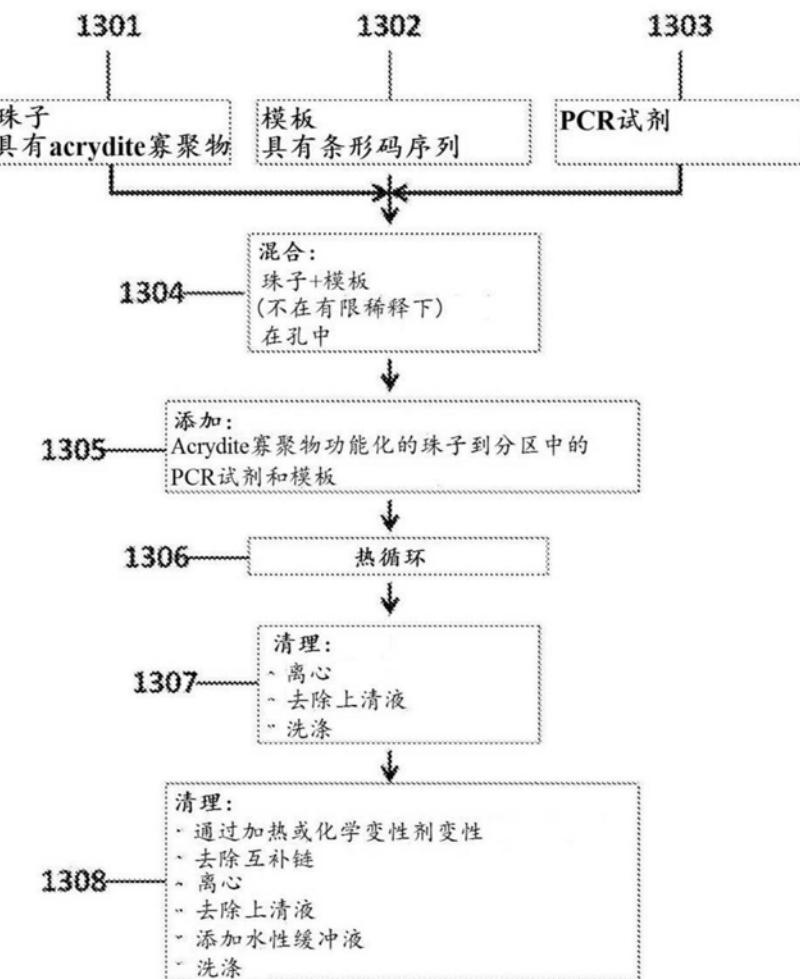


图13A

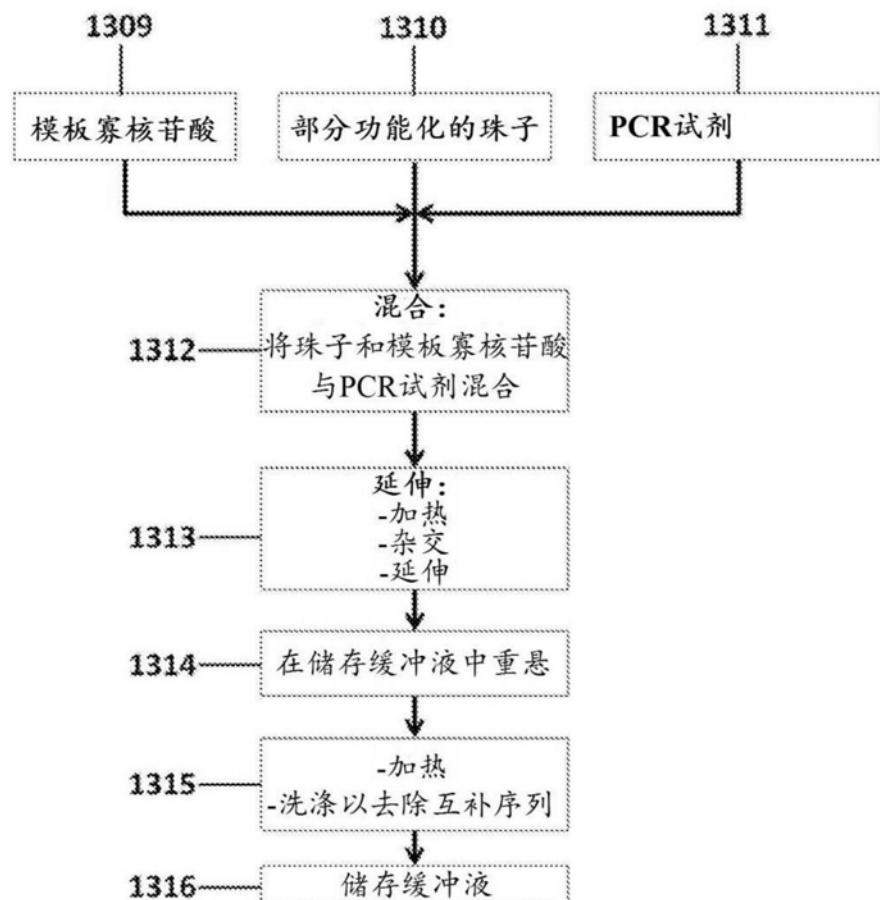


图13B

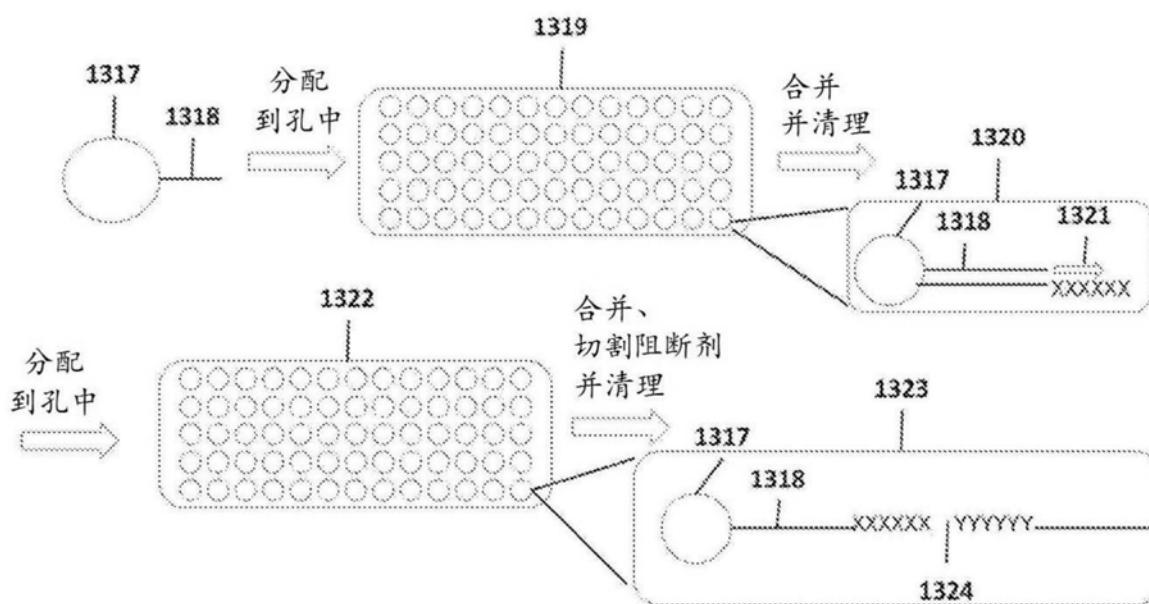


图13C

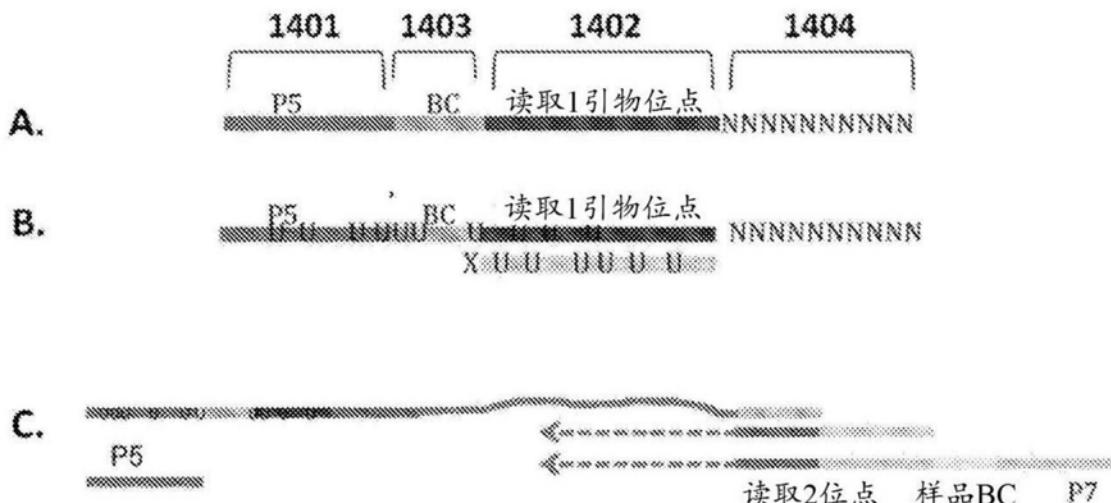


图14

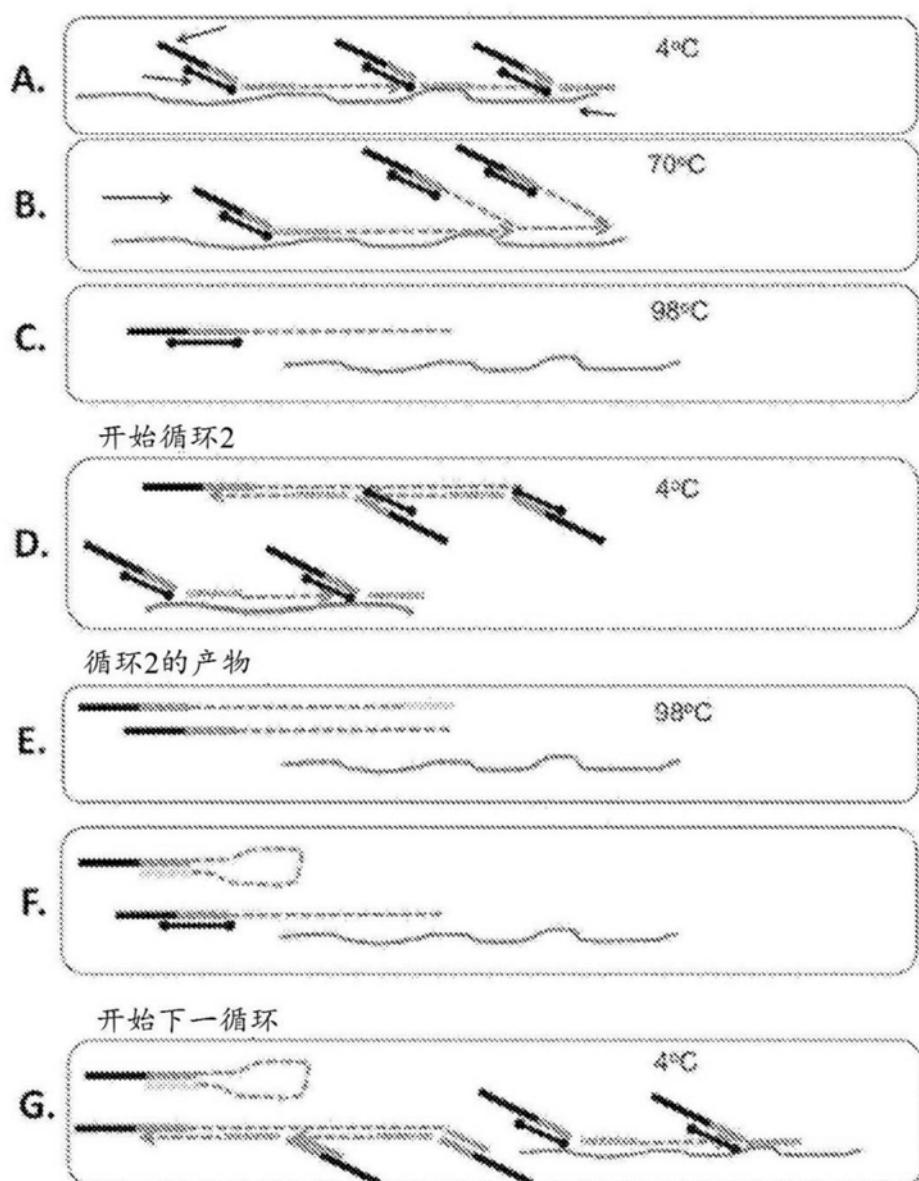


图15

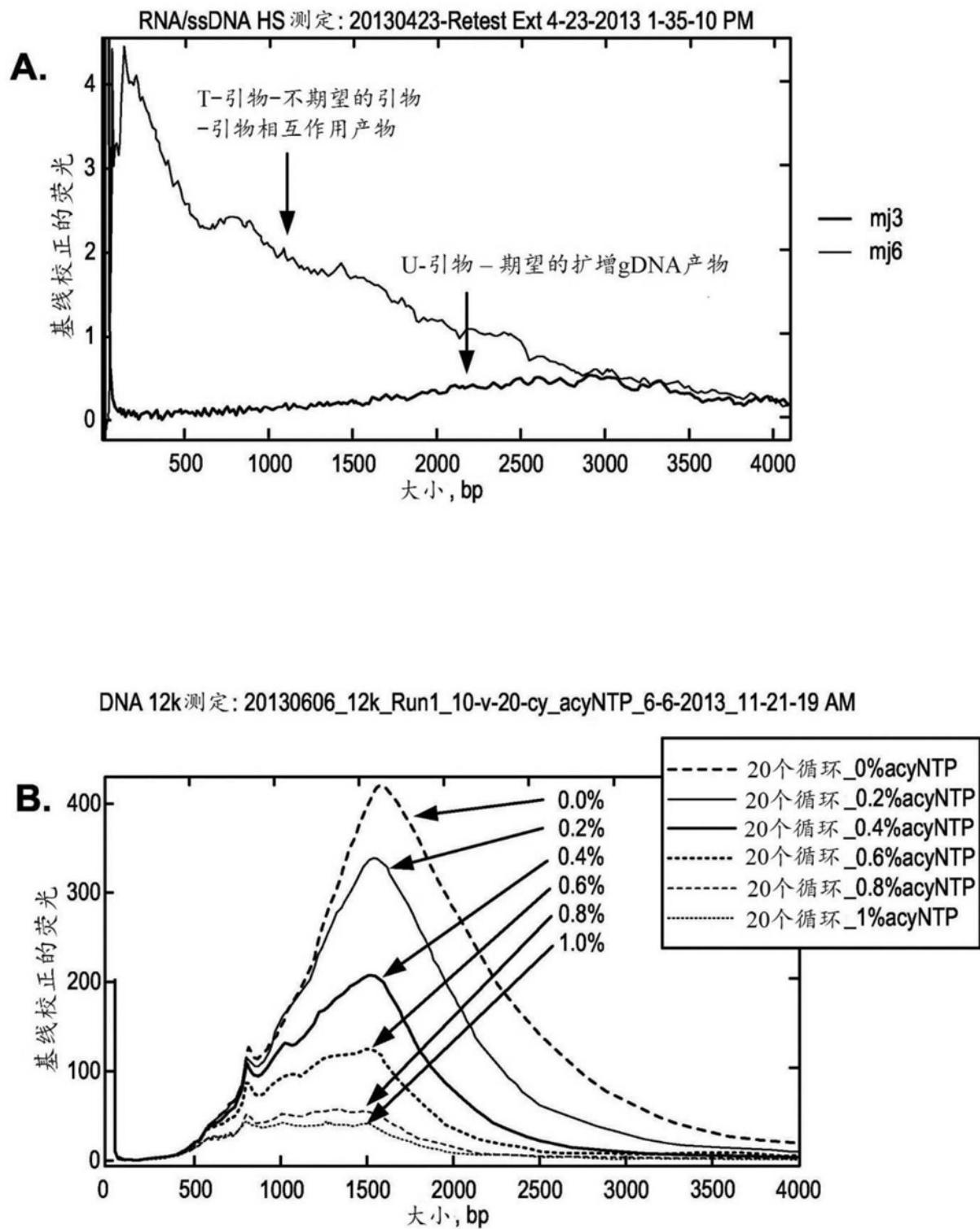


图16

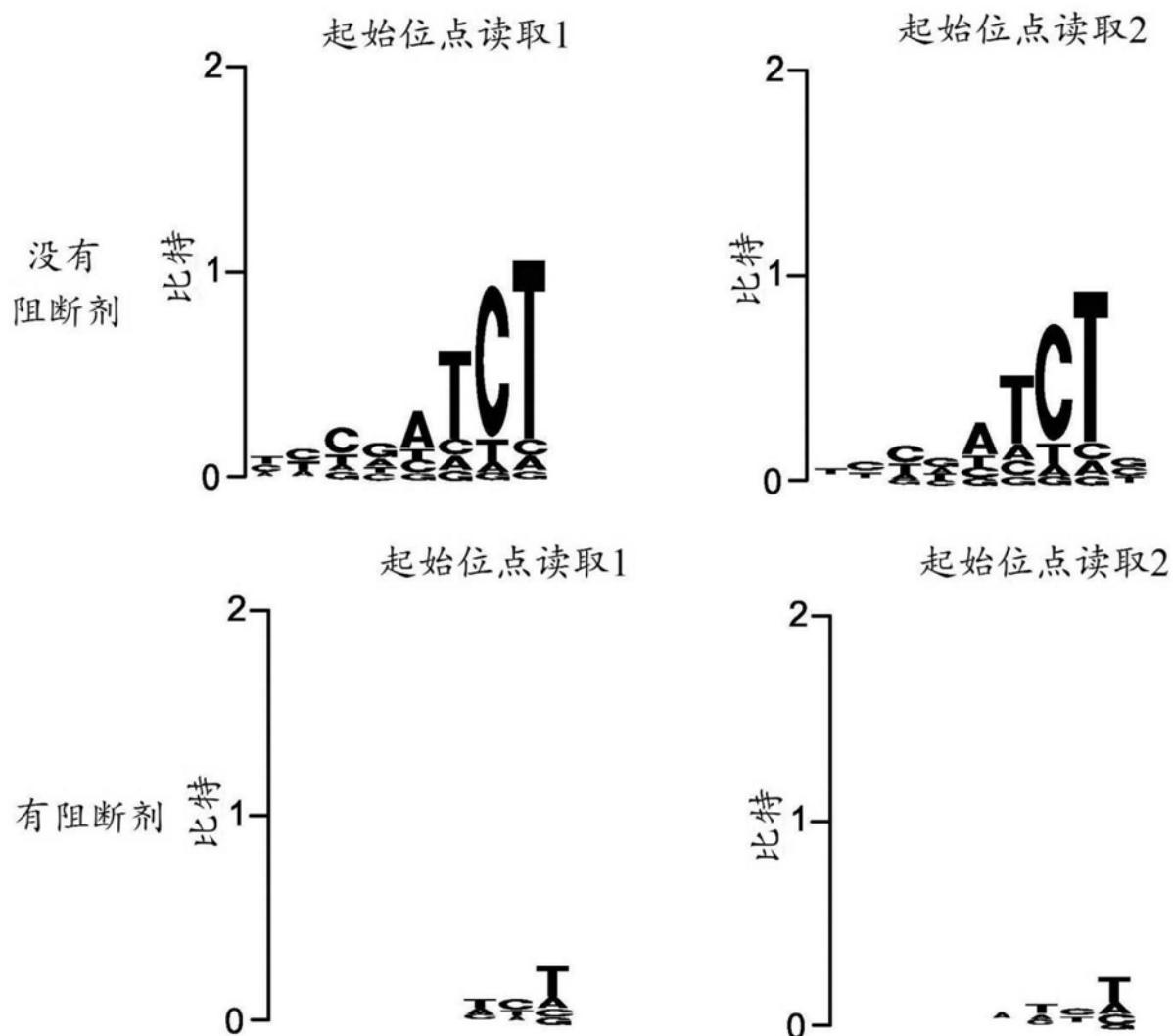


图17

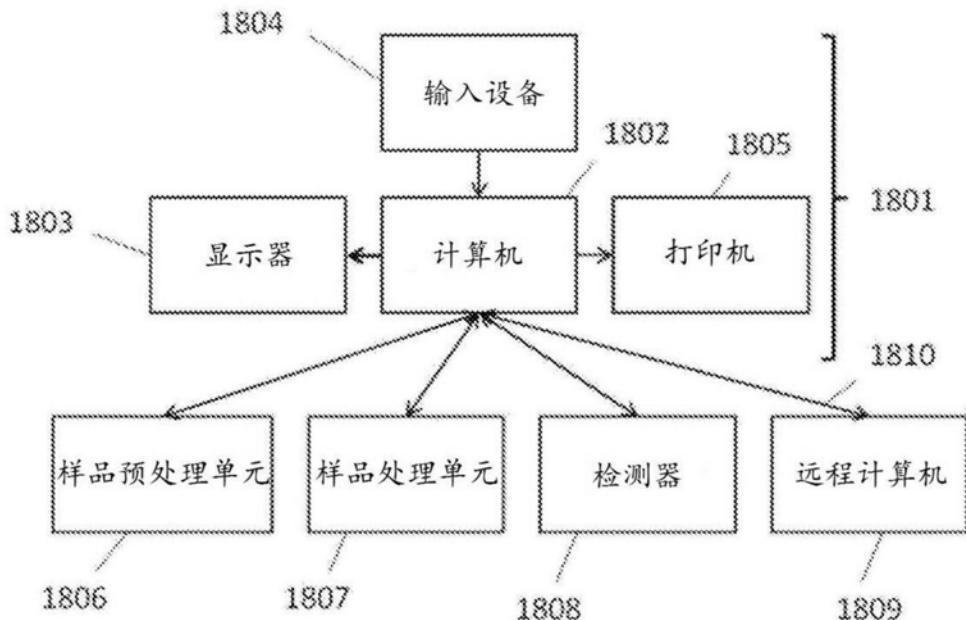


图18

功能	名称	序列 5'-3'	修饰
用于实际文库产生的引物	PS_M	/S'Acryd/TTTGCG-3'/ AAG GAA ACG GCG ACC ACC GAG A	"5' Acrydite", "ThioMCS-3'", 3' 脱氧尿苷, HPLC 纯化)
用于分选的反向引物	S2-Bio-R1-re	/5'-Bio/ GCG TCG TGT AGG GAA AGA GTG T	"5'双生物素", HPLC 纯化)
有限稀释模板	R1-条形码- PS_ 模板	GCG TCG TGT AGG GAA AGA GTG TNN NNN NNN NSN NGT GTA GAT CTC GGT GGT CGC CGT ATC ATT	手工混合, 具有天然核苷酸
用于加入N-mer随机引物的模板	R1-部分-Nmer-re	NNN NNN NNA GAT CGG AAG AGC GTC GTG TAG GGA AAG AGT GT	手工混合, 具有天然核苷酸, 没有3'阻断剂

图19

凝胶珠子等份	最终体积 (uL)	磁珠的浓度 mg/mL	温育时间 (hr)	选择性沉积	回收 (Y/N)
1	100	100	24	是	否
2	100	80	24	是	否
3	100	60	24	是	是
4	100	40	24	是	是
5	100	20	24	否	否
6	100	10	24	否	否

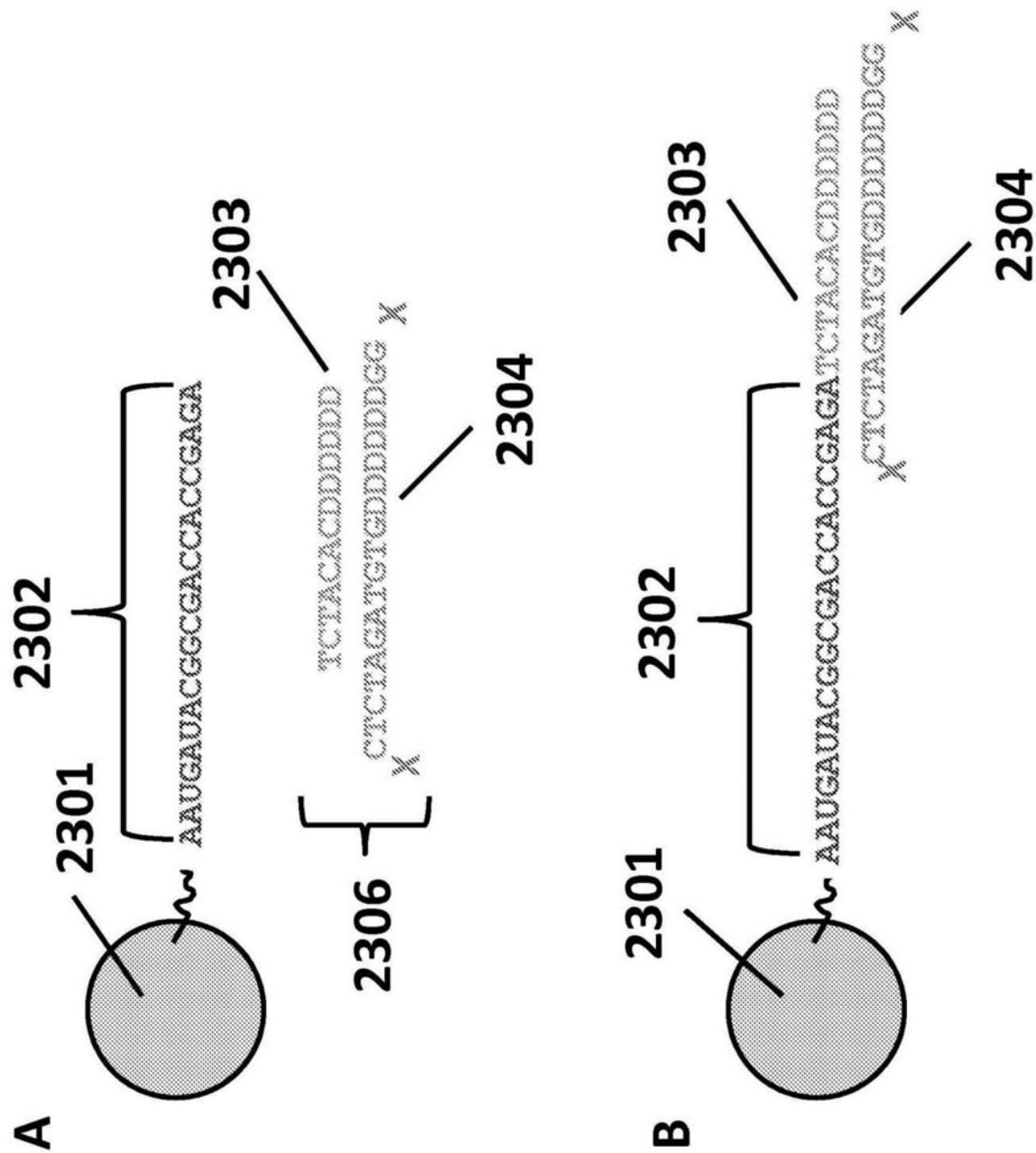
图20

度量	说明	值	目标值
Q40 错误	全部q40位置的错误率	0.002	<=0.02
未定位的分数	两端均未定位的读取结果的分数	0.996	<=0.04
插入物大小中值	插入物大小中值	399	[200:400]
IQR插入物大小	插入物大小的四分位距	413	<=250
零覆盖分数	零覆盖的基因组的分数	0.9242	<=0.01

图21

度量	说明	值	目标值
Q40 错误	全部q40位置的错误率	0.001	<=0.02
未定位的分数	两端均未定位的读取结果的分数	0.03	<=0.04
插入物大小中值	插入物大小中值	310	[200:400]
IQR 插入物大小	插入物大小的四分位距	209	<=250
零覆盖分数	零覆盖的基因组的分数	0.0093	<=0.01

图22



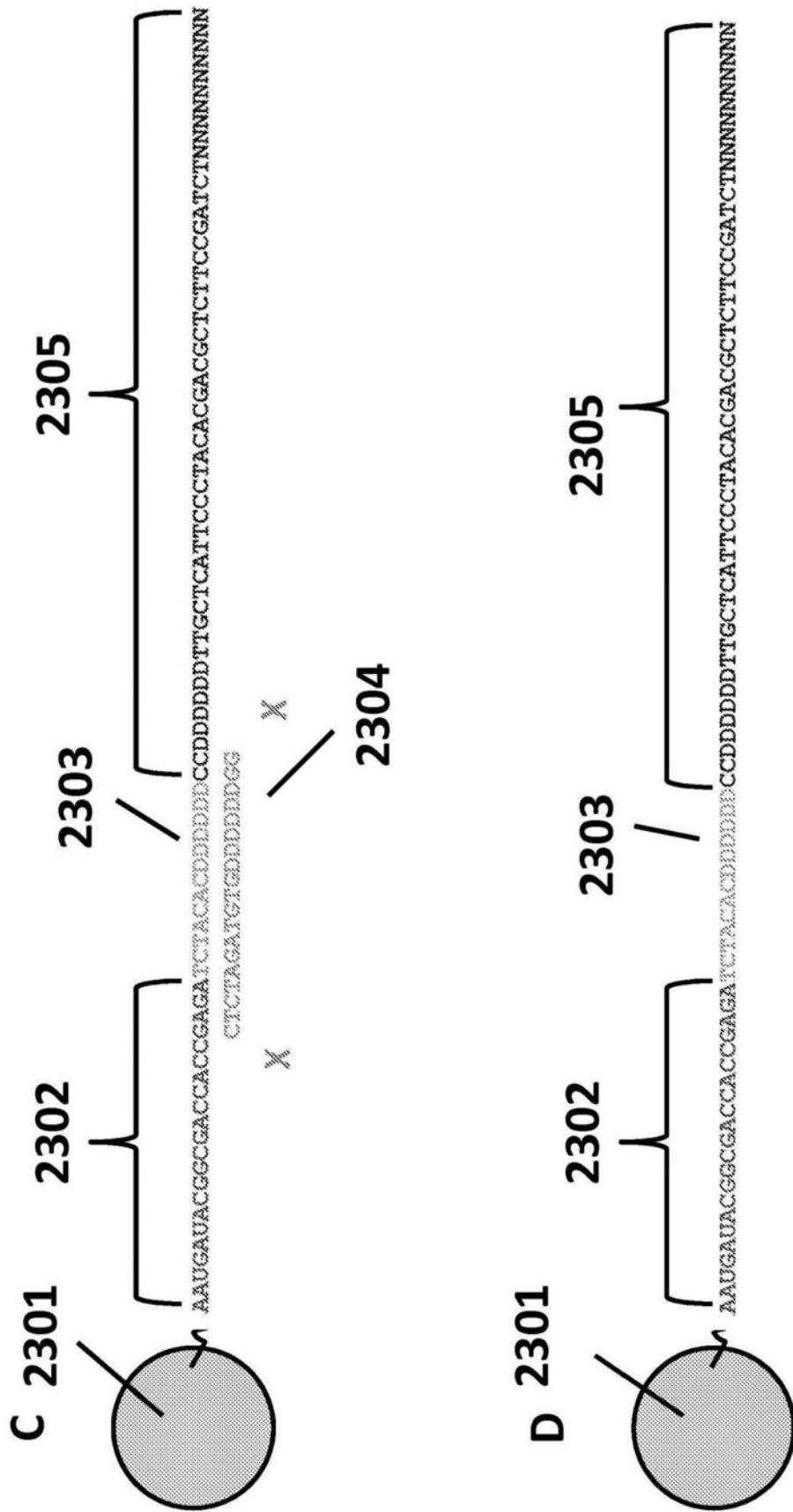


图23

A

2

2401

图24

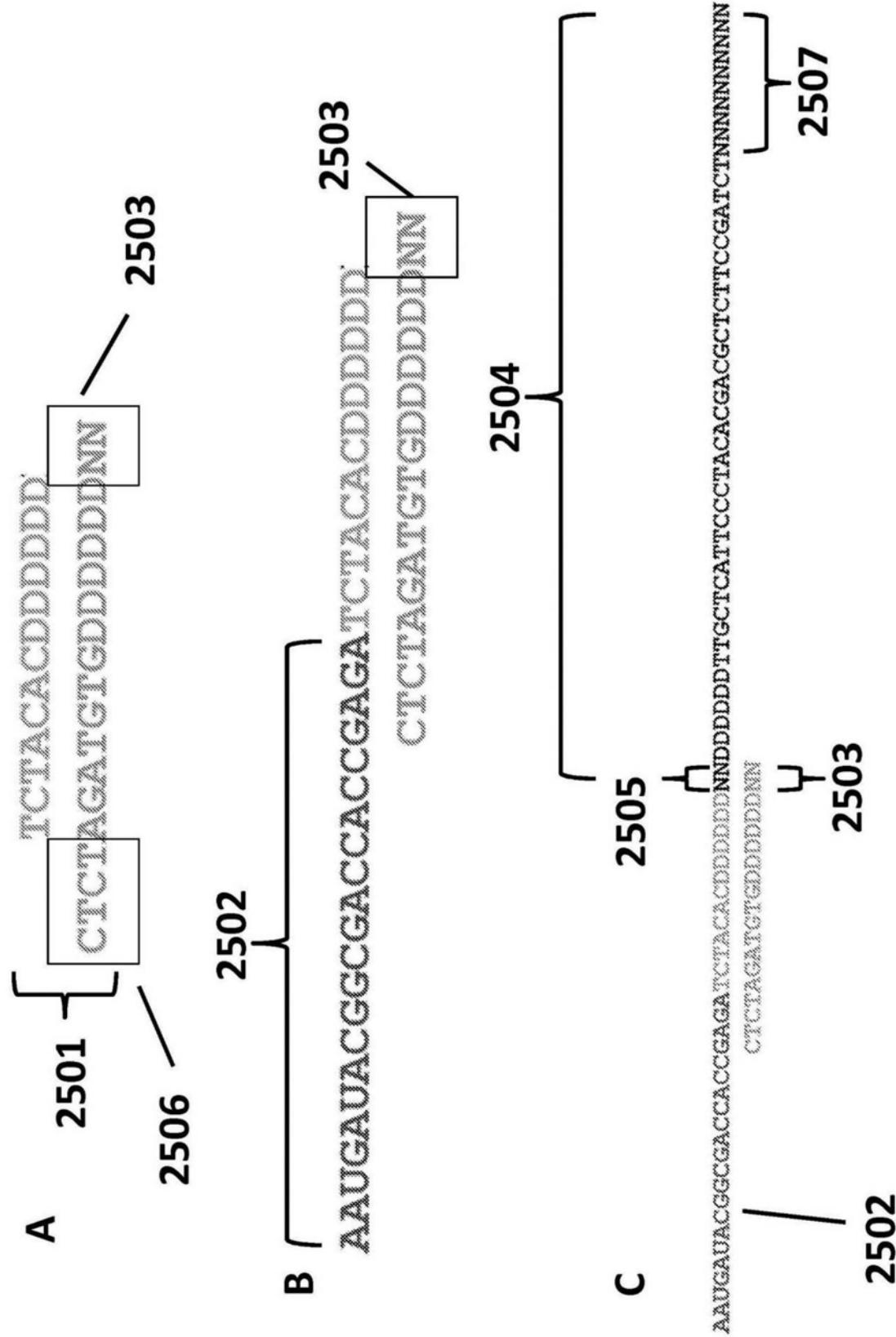


图25

2601

TCTACACDDDDDD
CTCTAGATGTGDDDDDDAC

2609**2602****2603**

TCTACACDDDDDD
CTCTAGATGTGDDDDDDCT

2609**2604****2605**

TCTACACDDDDDD
CTCTAGATGTGDDDDDDGA

2609**2606****2607**

TCTACACDDDDDD
CTCTAGATGTGDDDDDDTG

2609**2608**

图26

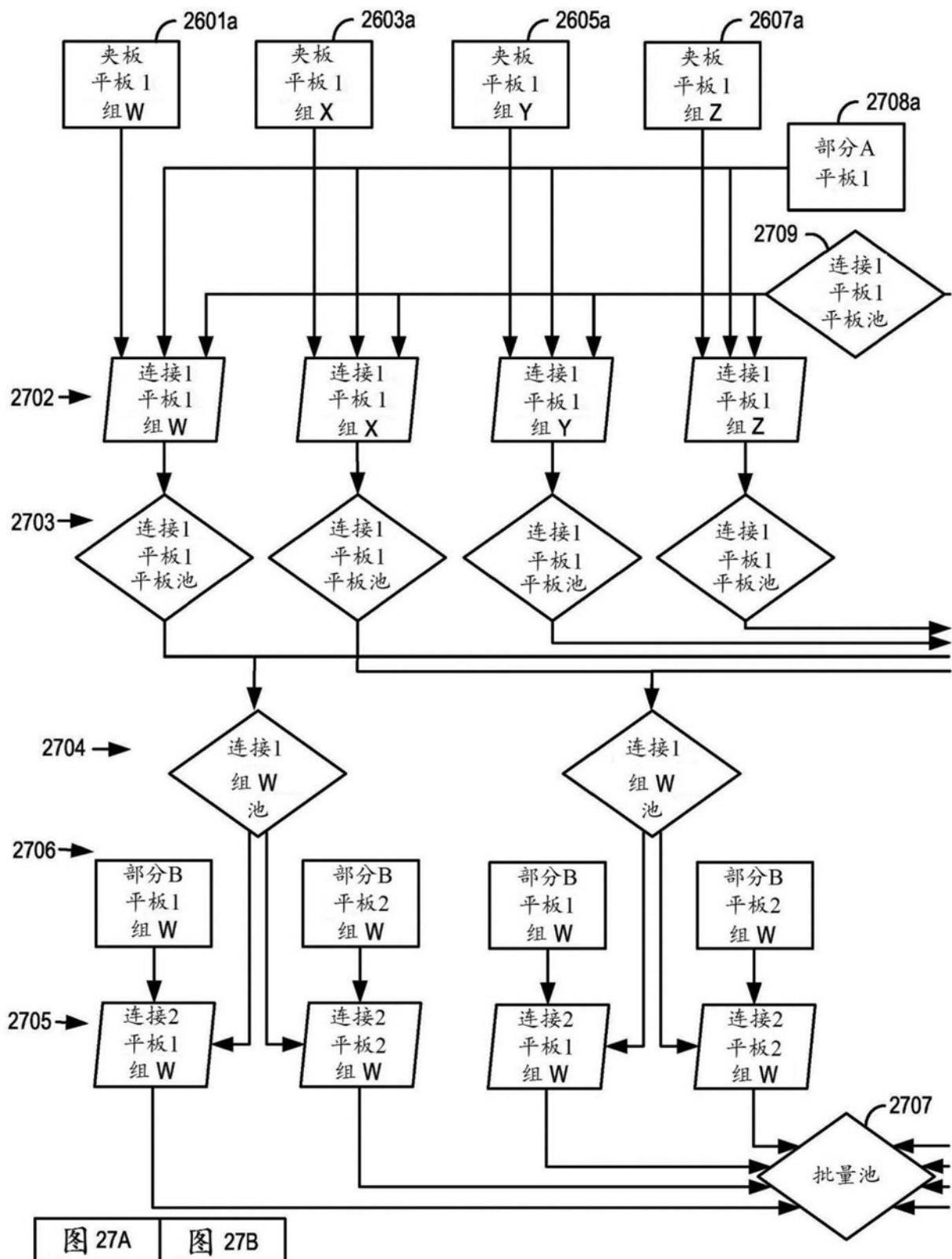


图27A

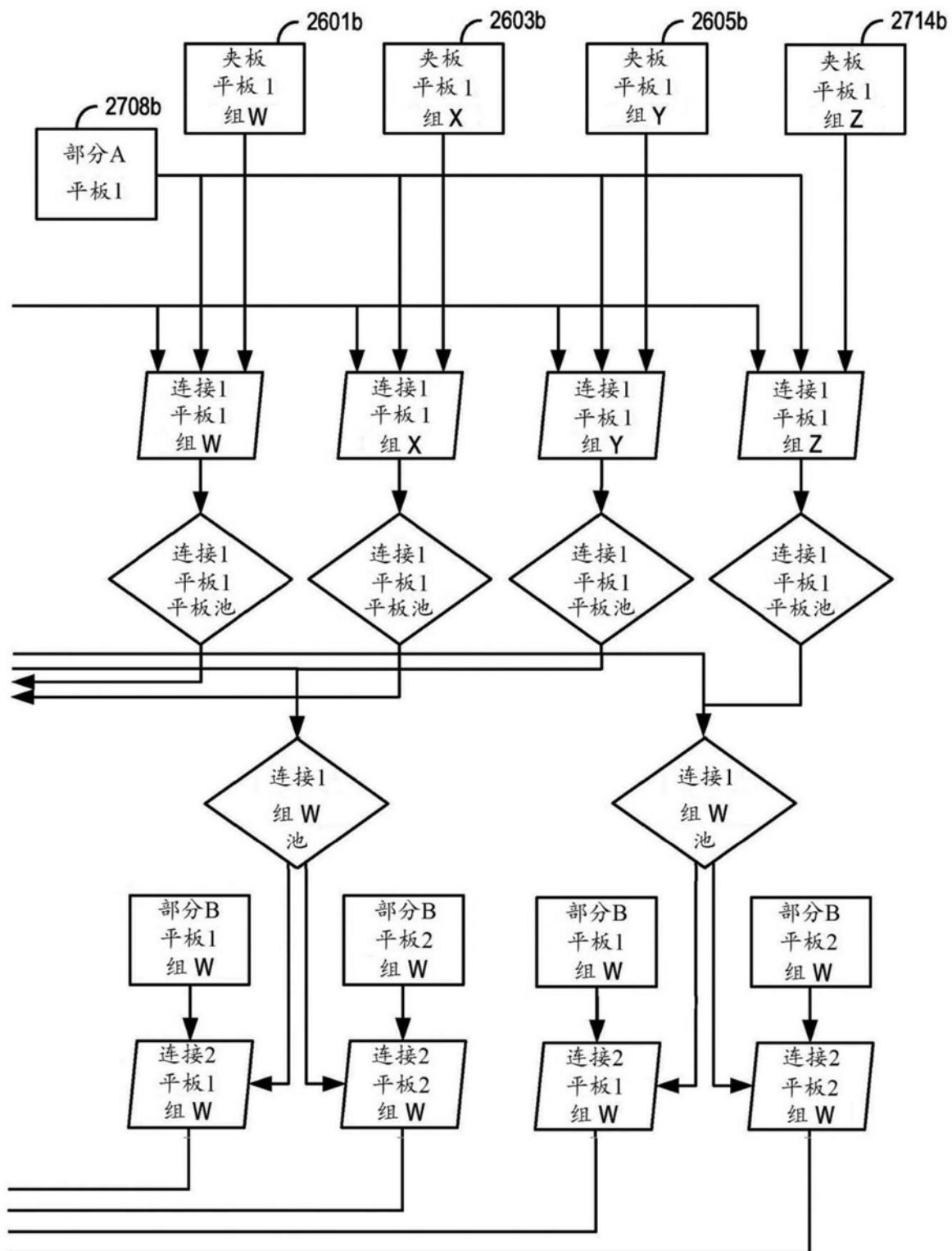


图27B

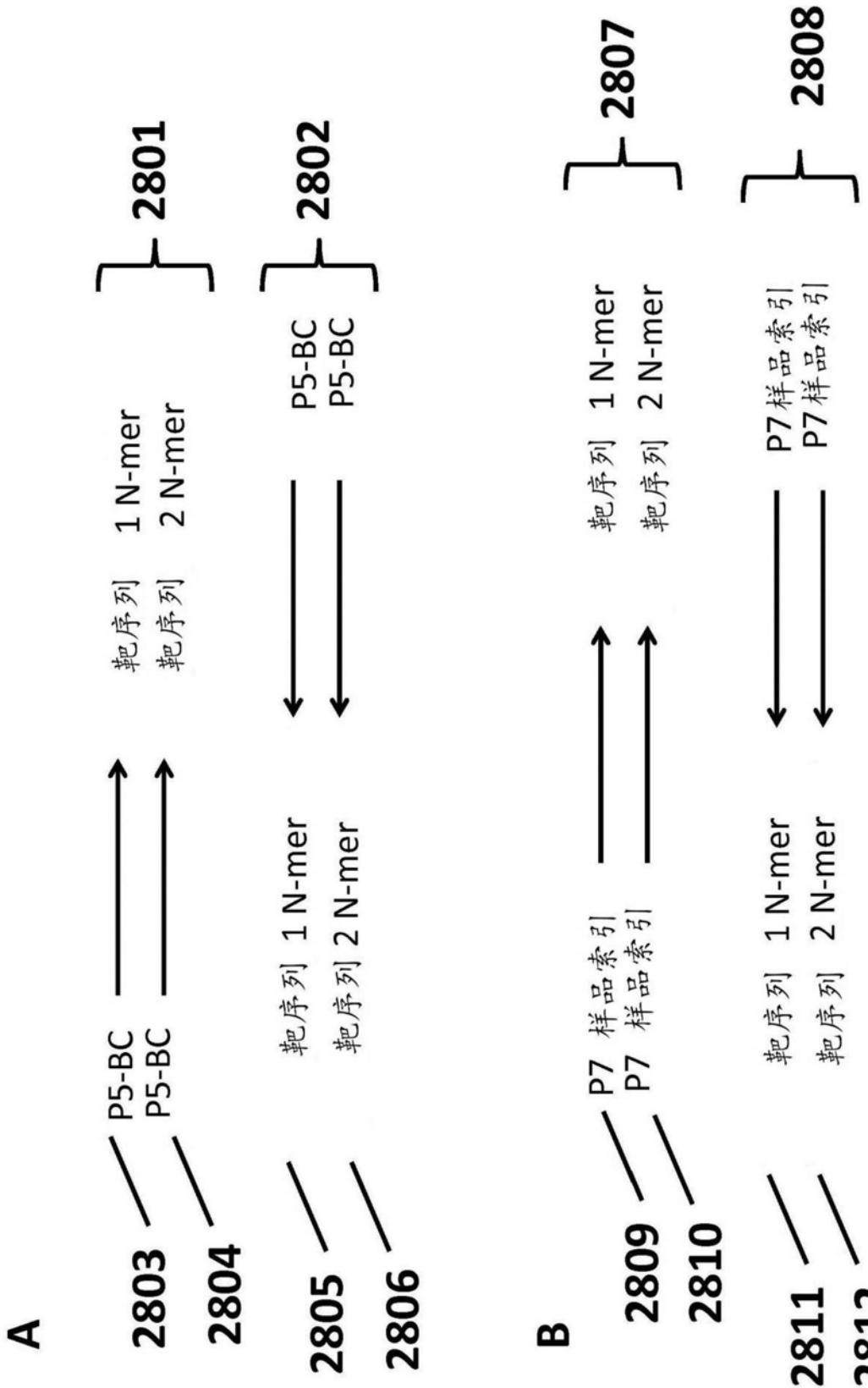
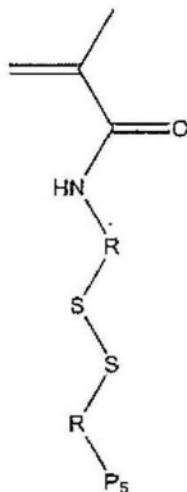
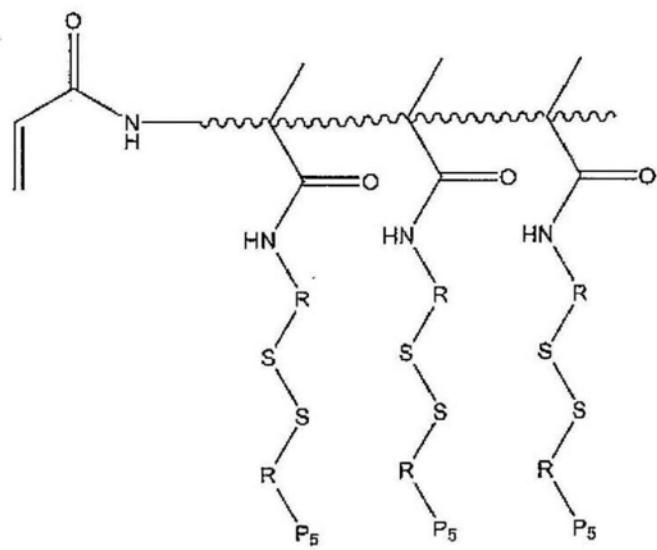


图28

A P_5 单体B

多 p5 单体

C

多 p5 交联剂

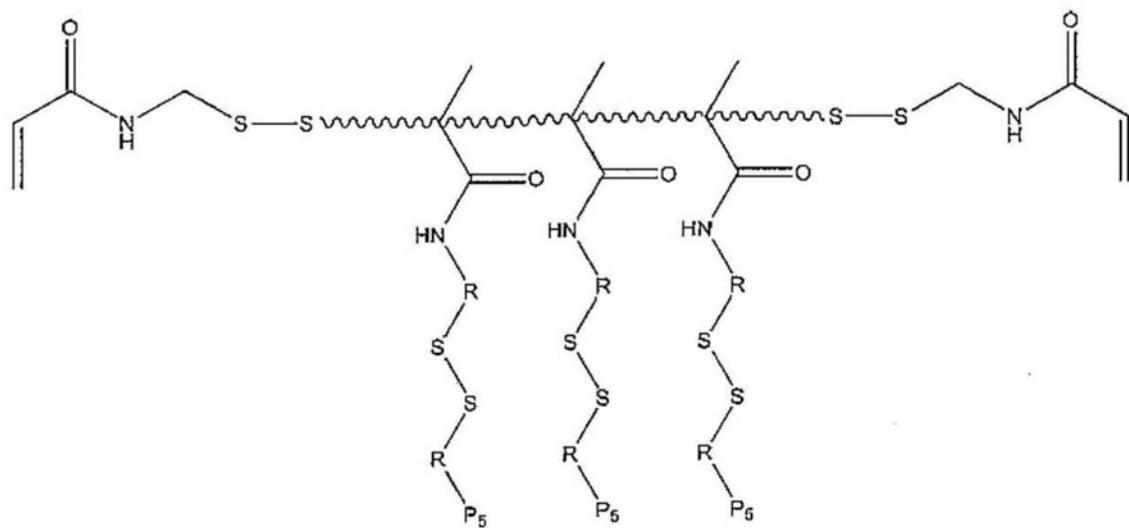
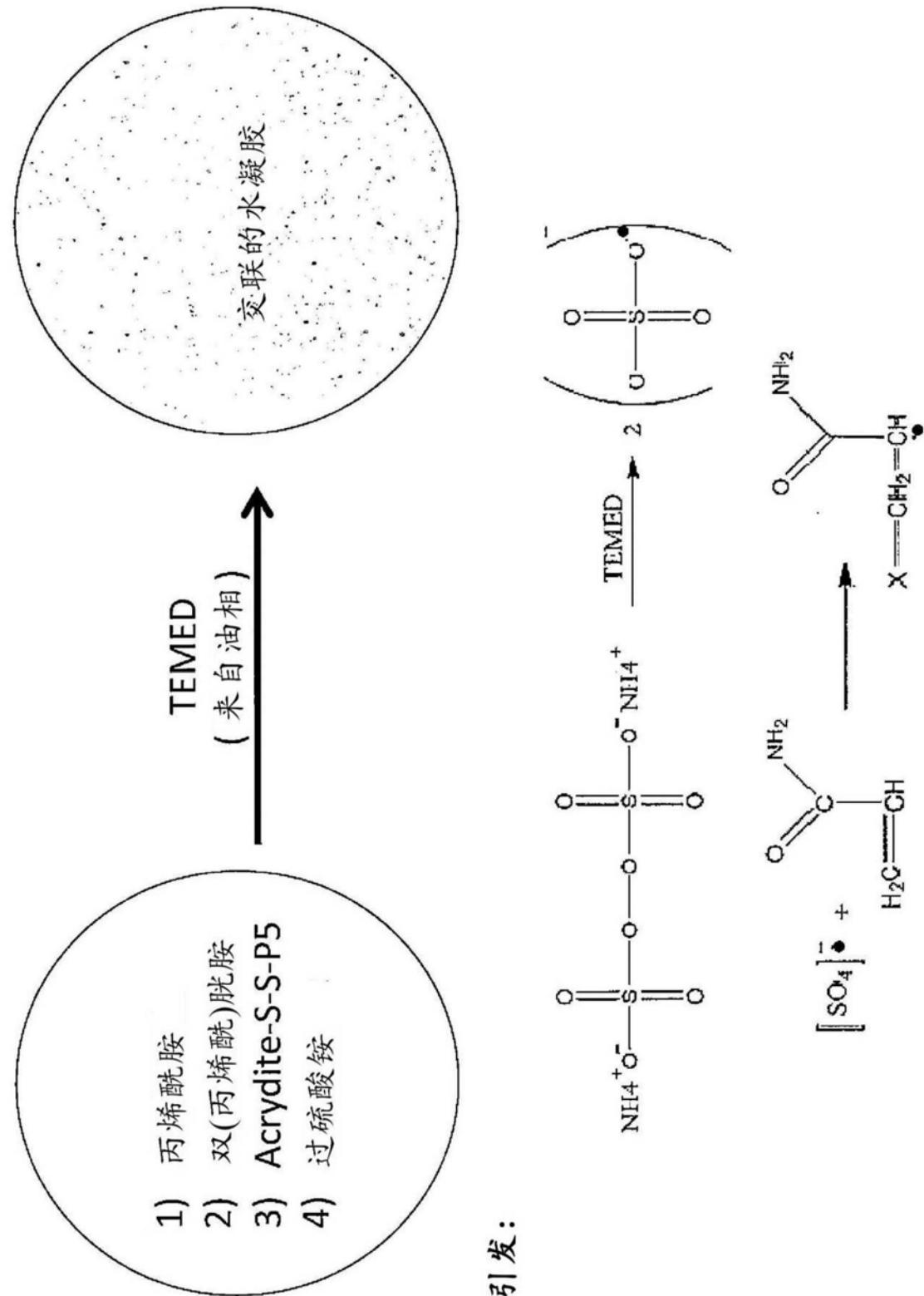


图29



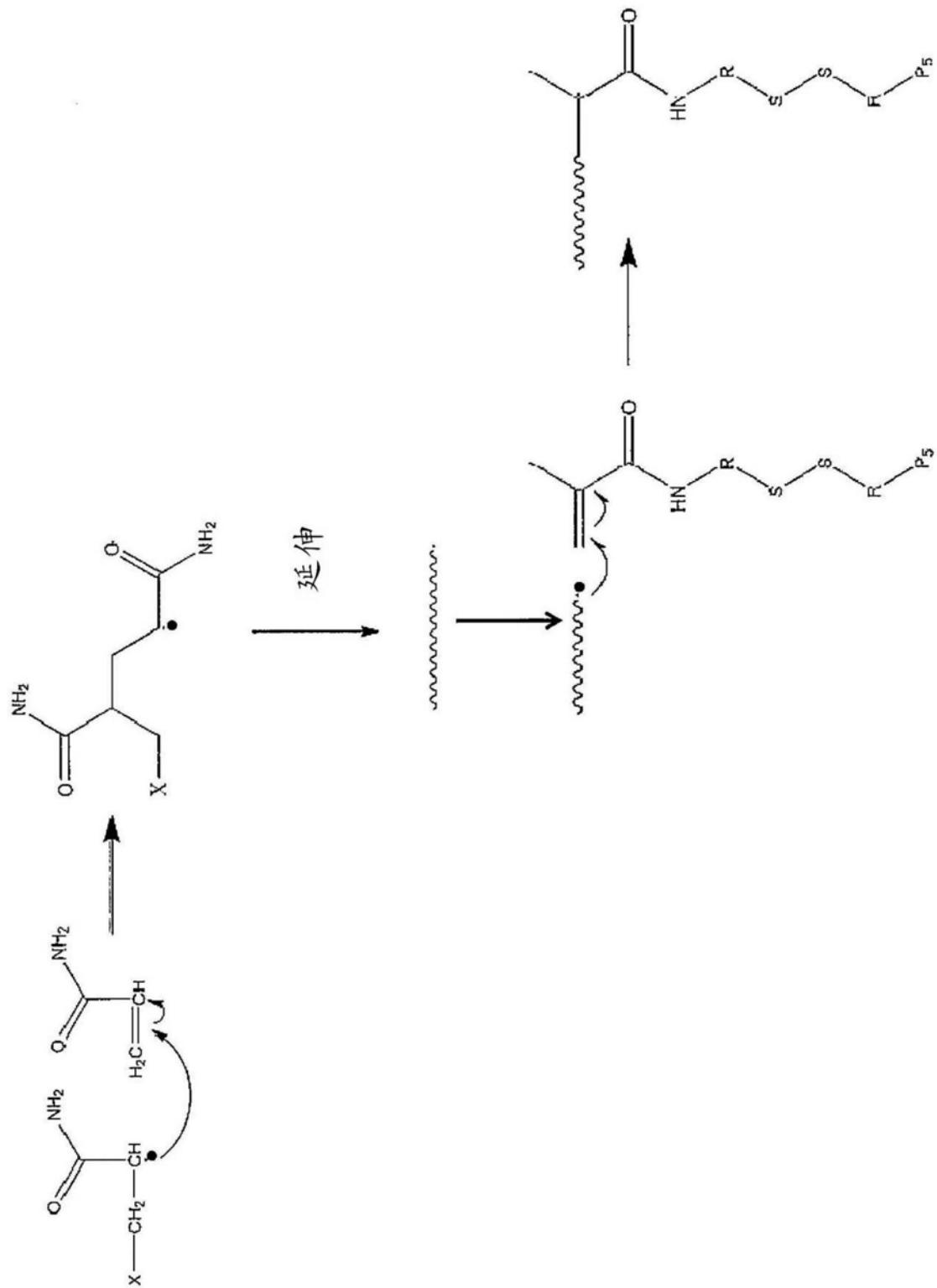


图30B

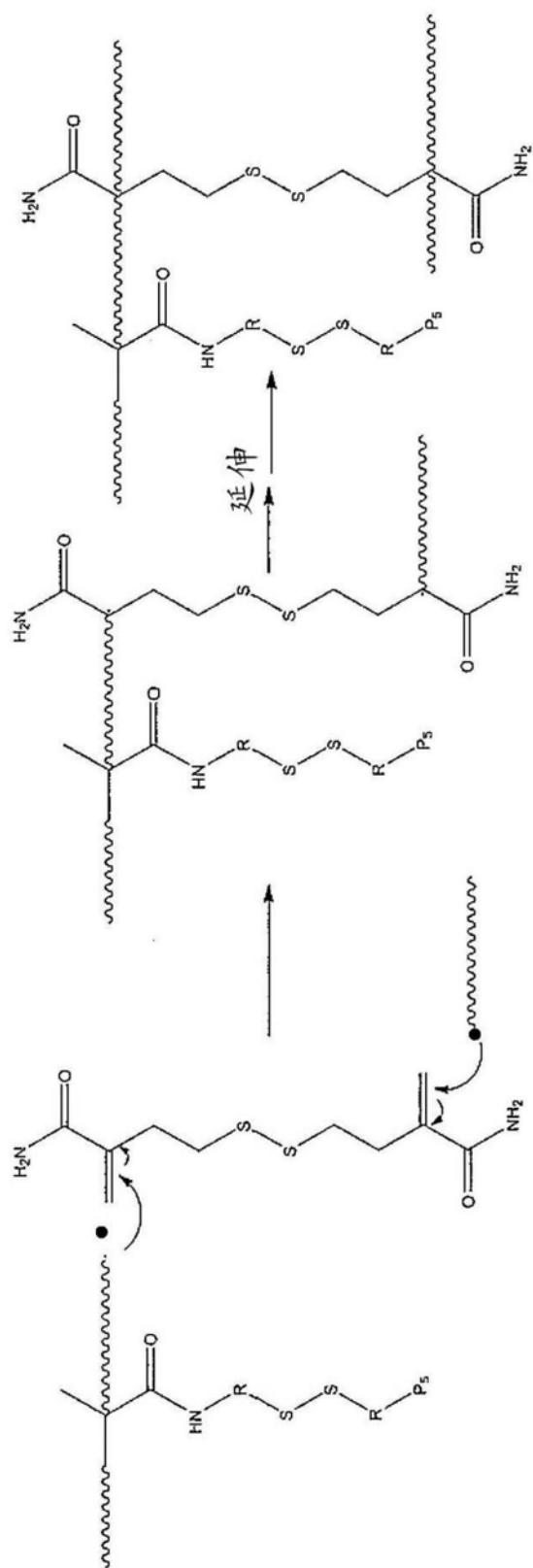


图30C

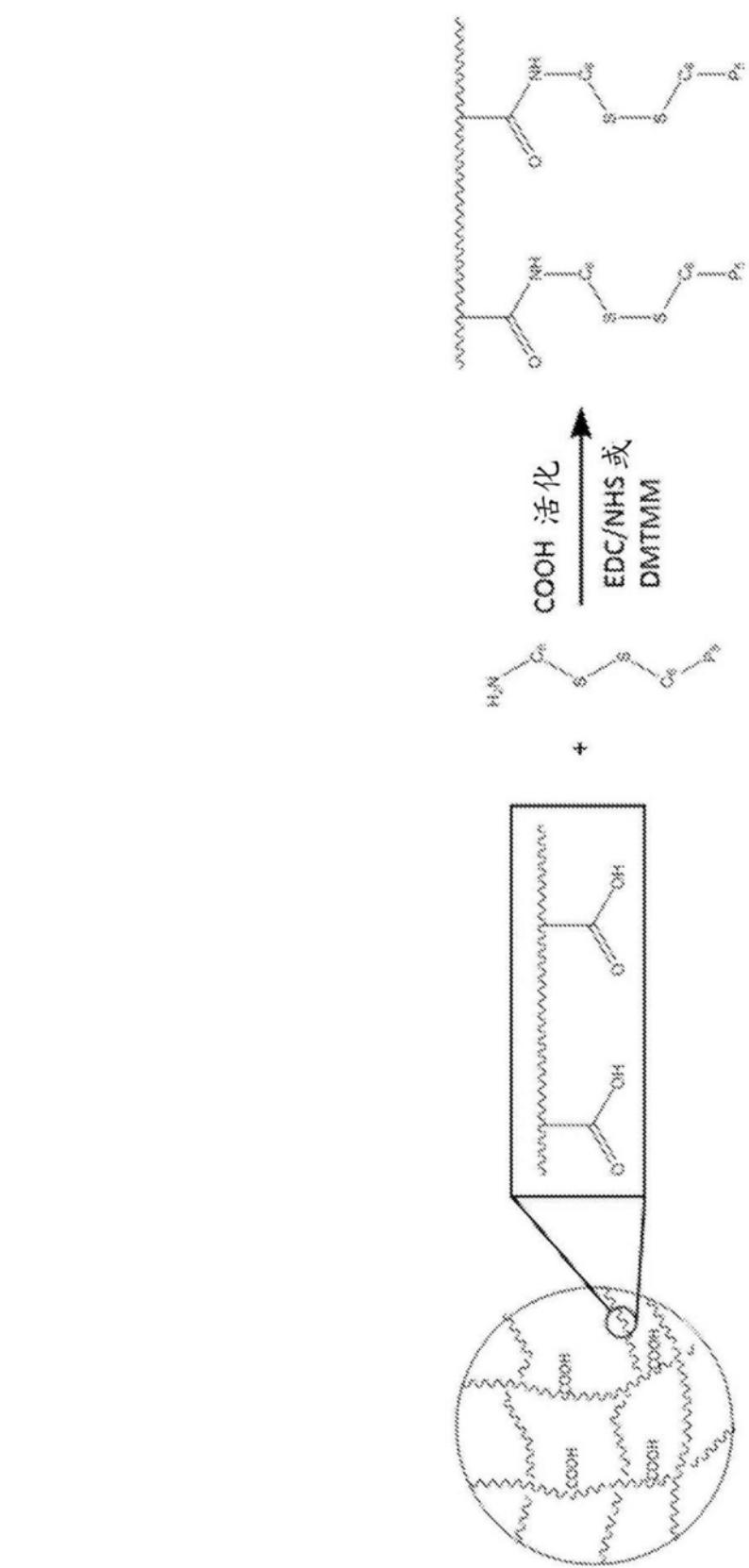


图31

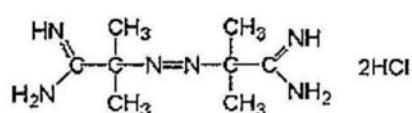
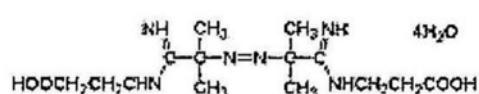
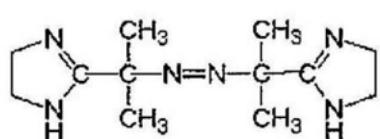
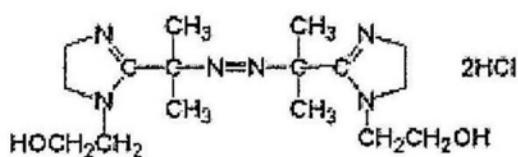
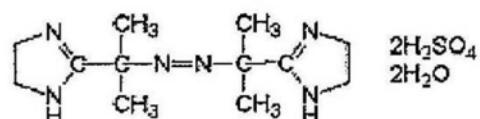
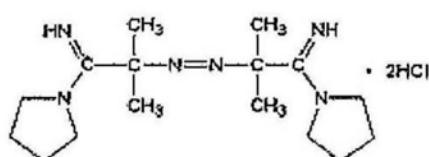
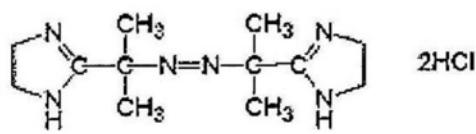
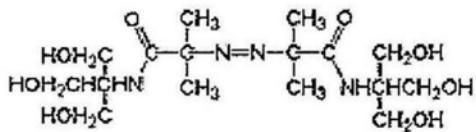
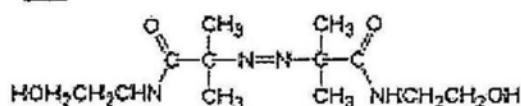
IIIIIIIVVVIVIIVIIIIX

图32

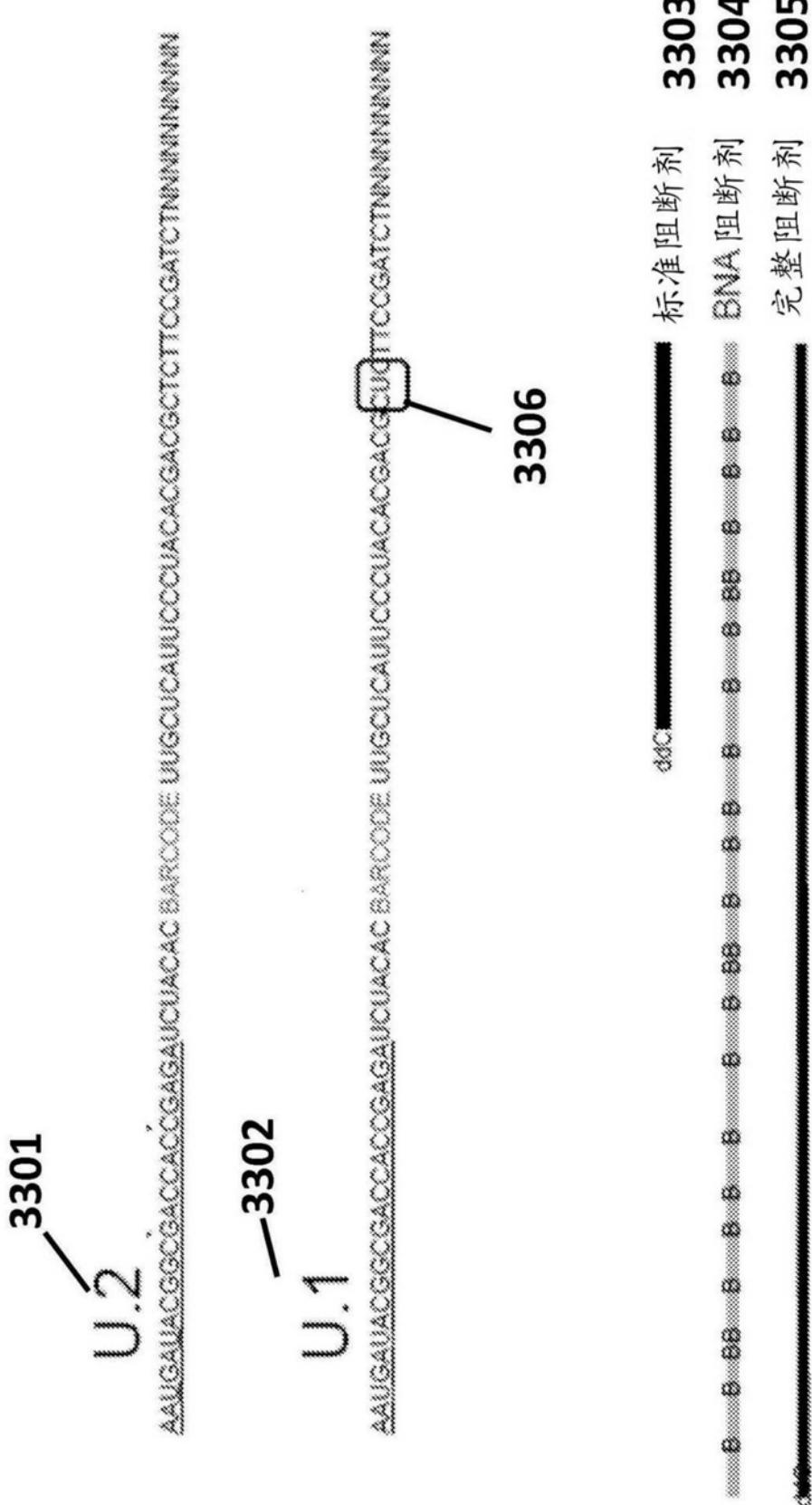


图33A

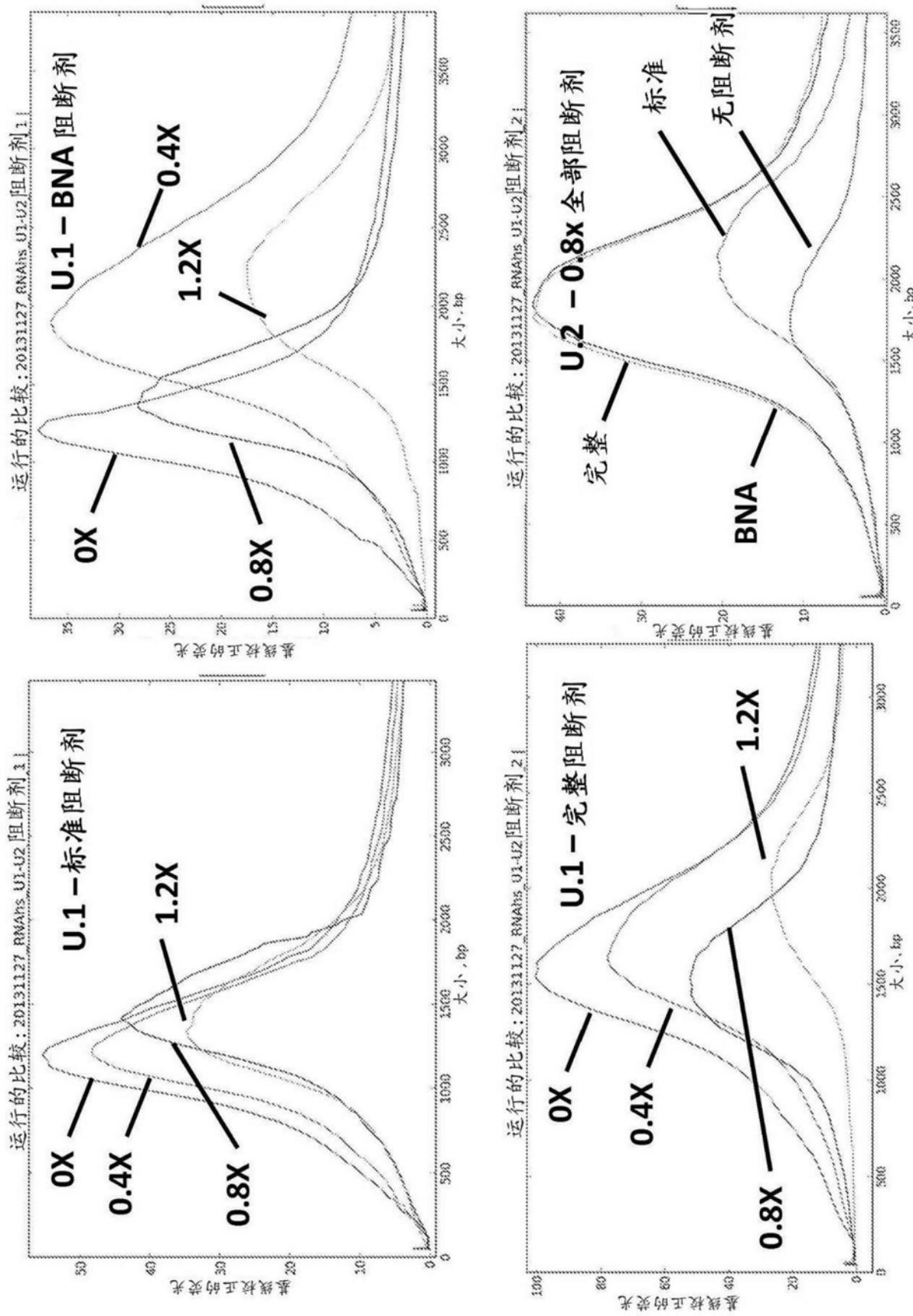


图33B

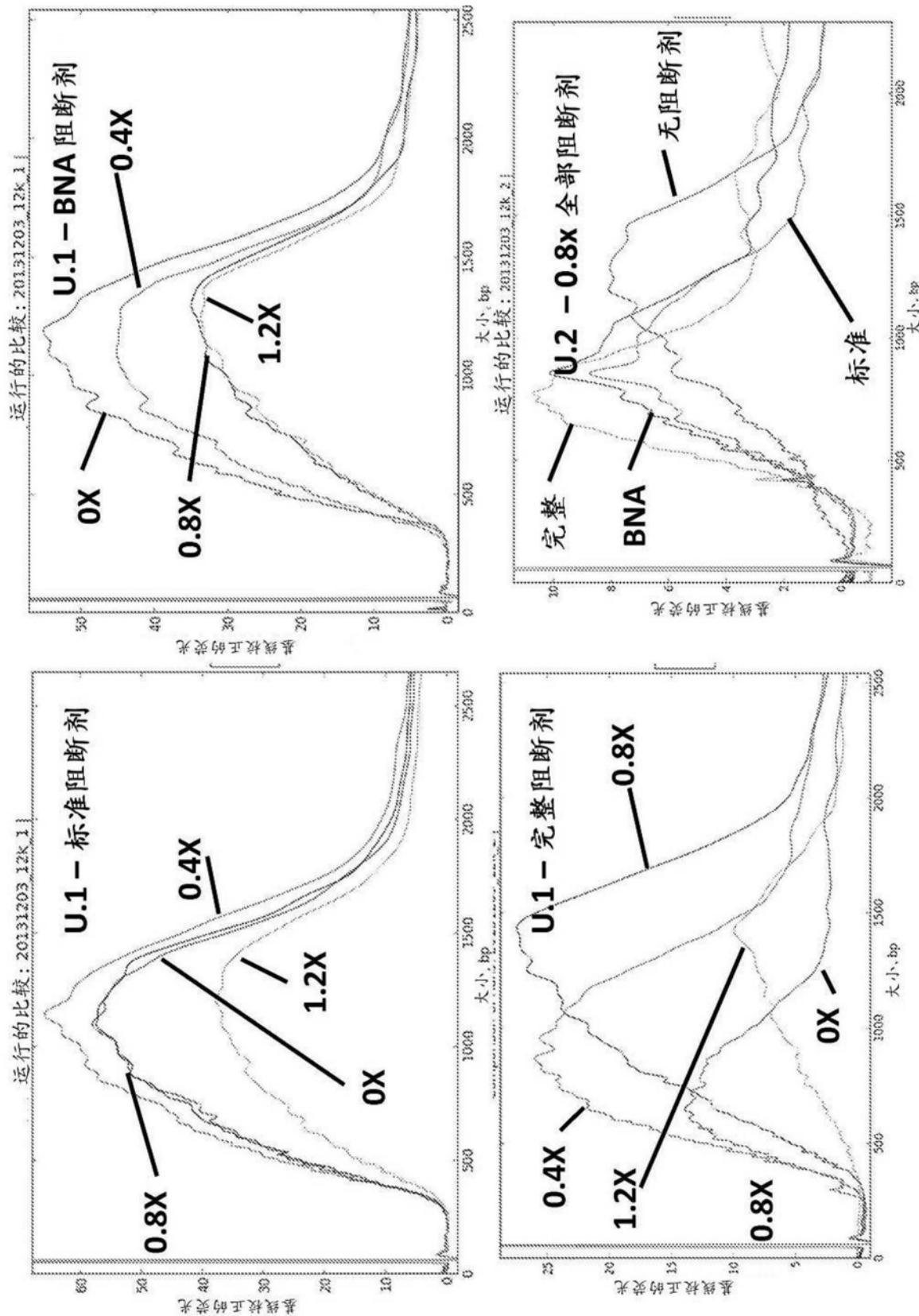
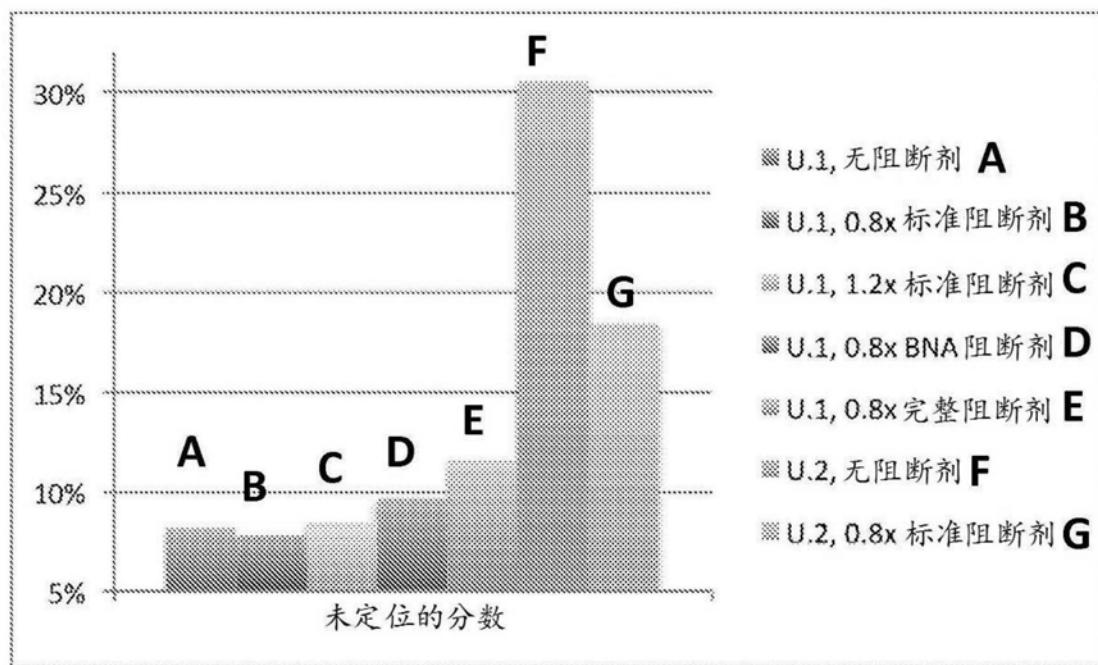


图33C

I



II

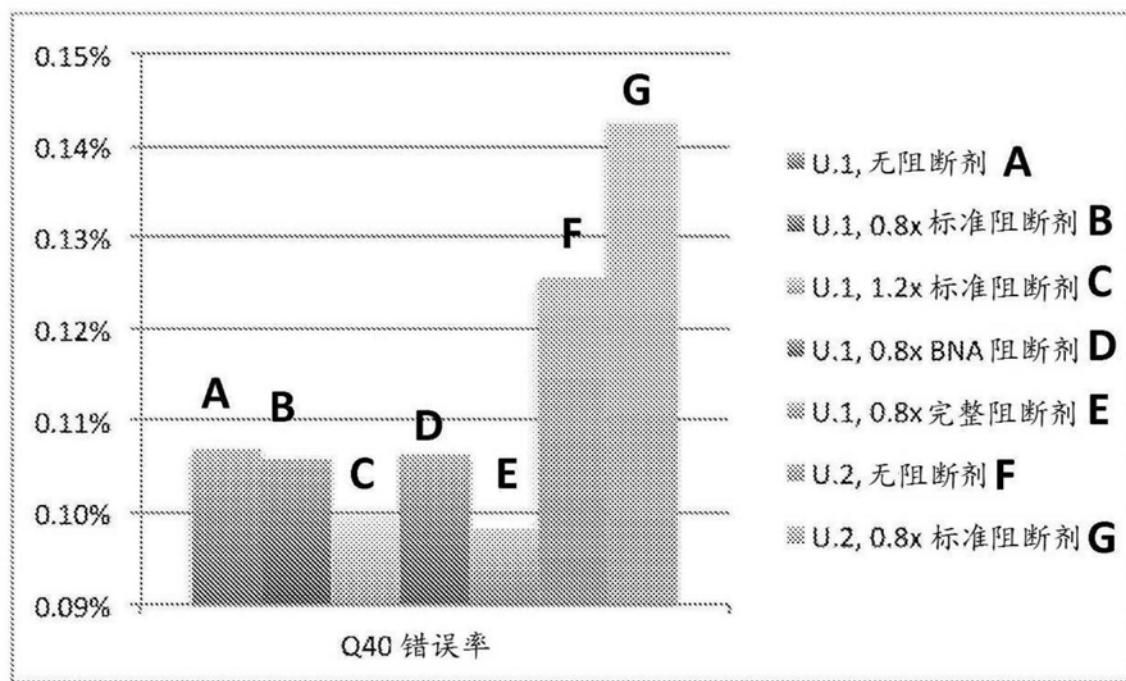


图33D

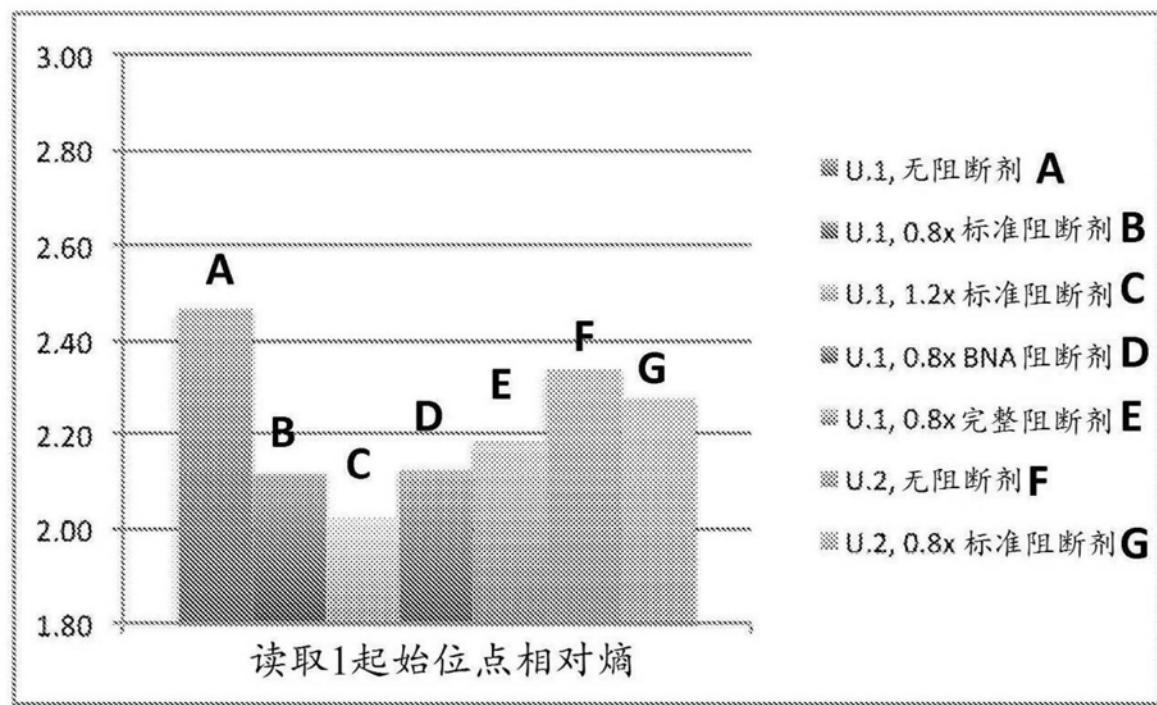
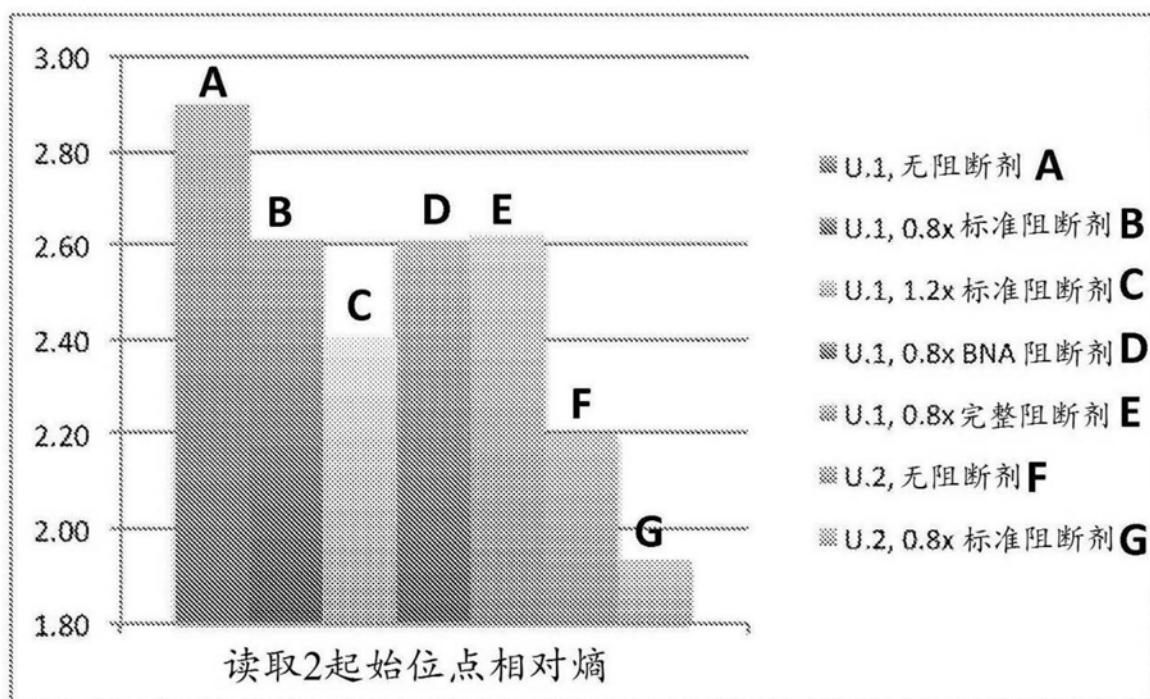
III**IV**

图33E

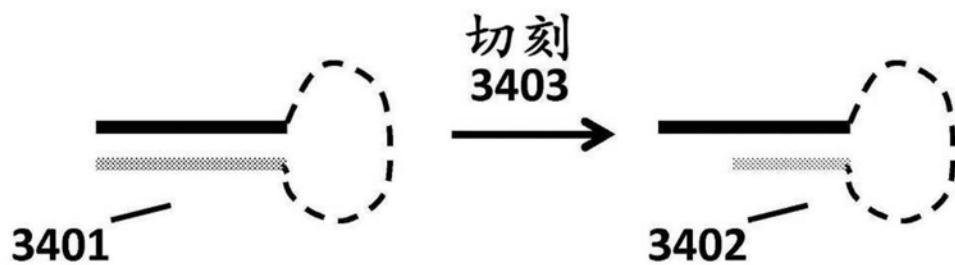
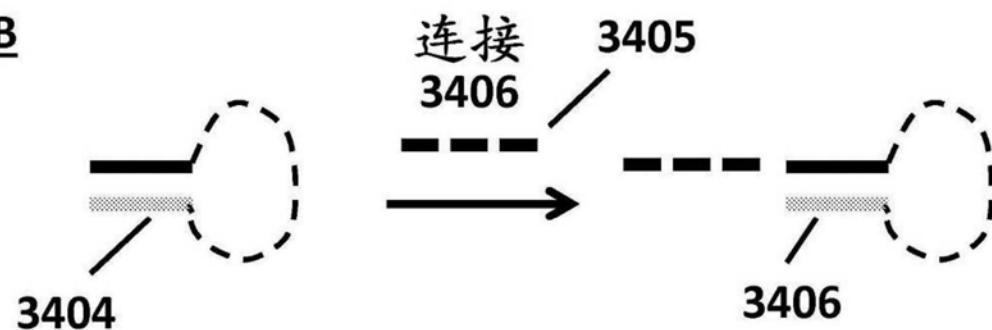
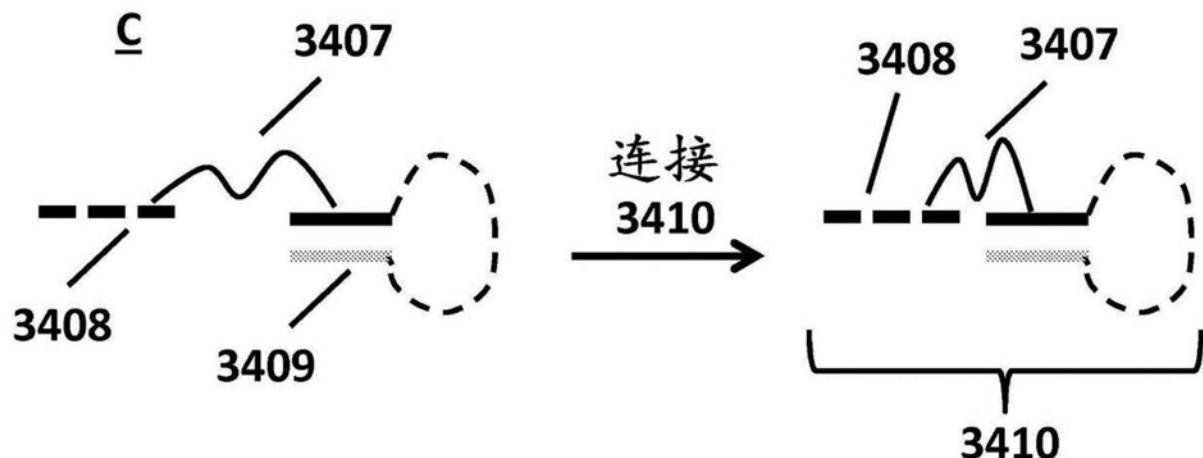
A**B****C**

图34

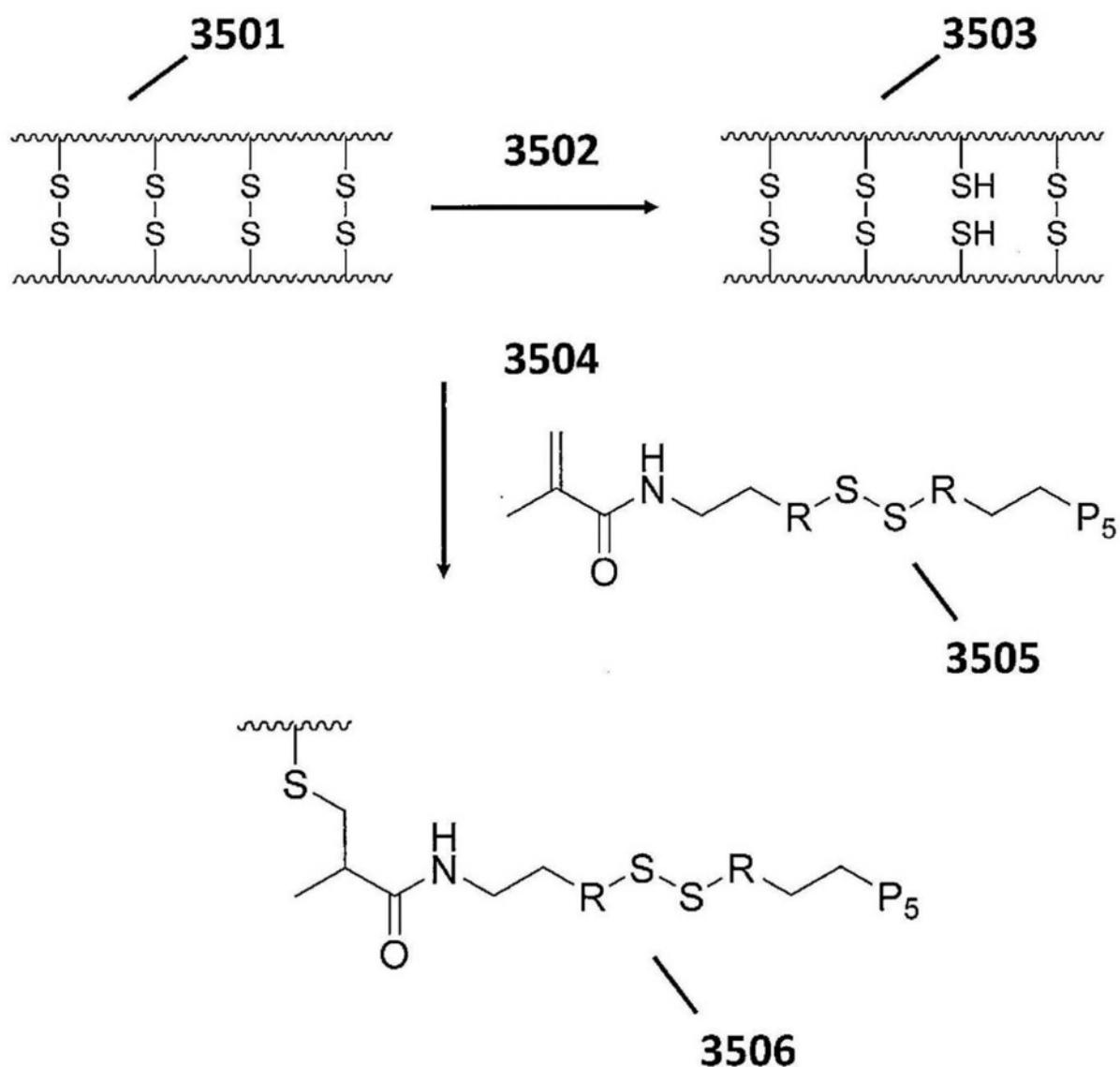


图35A

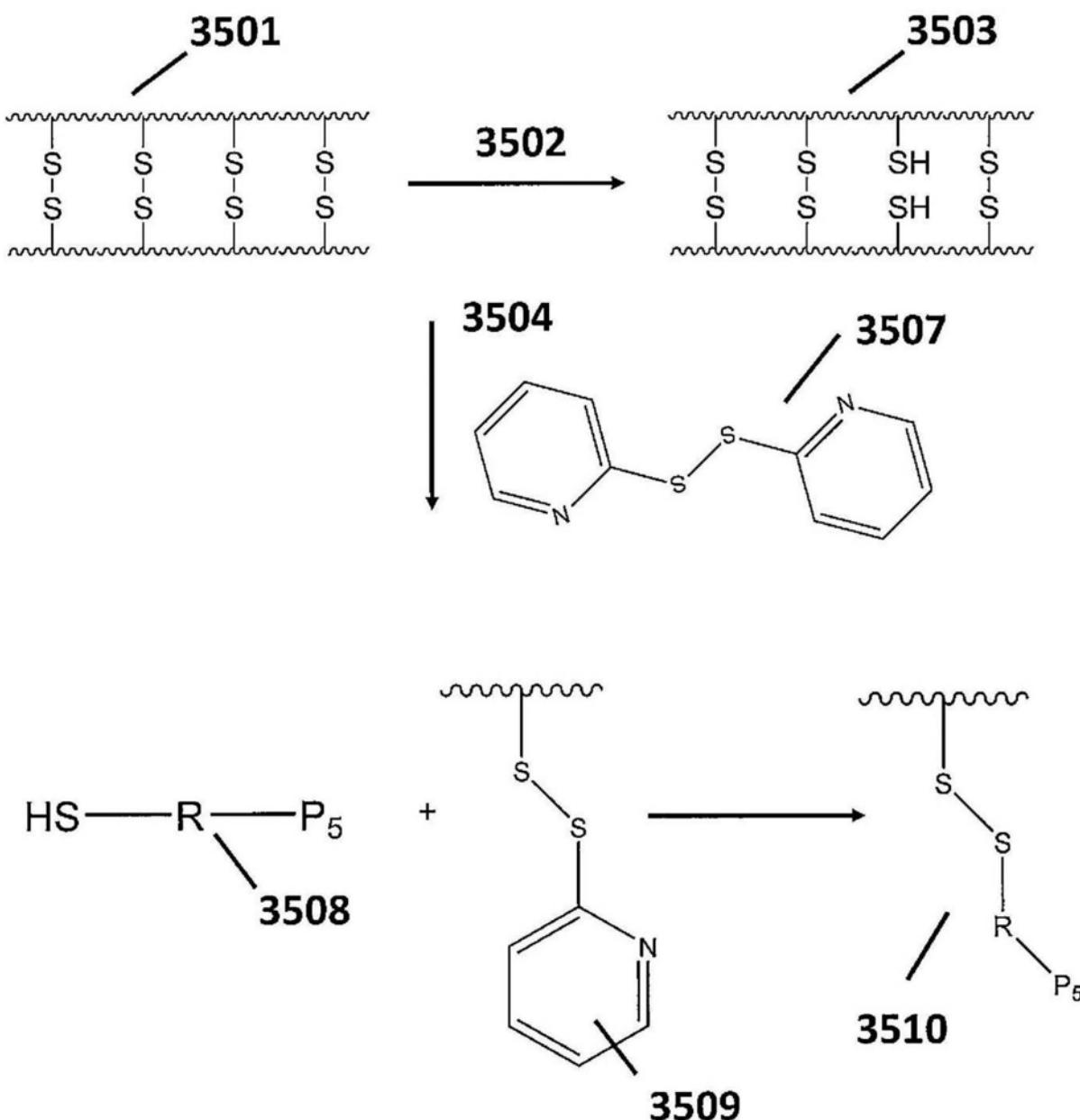
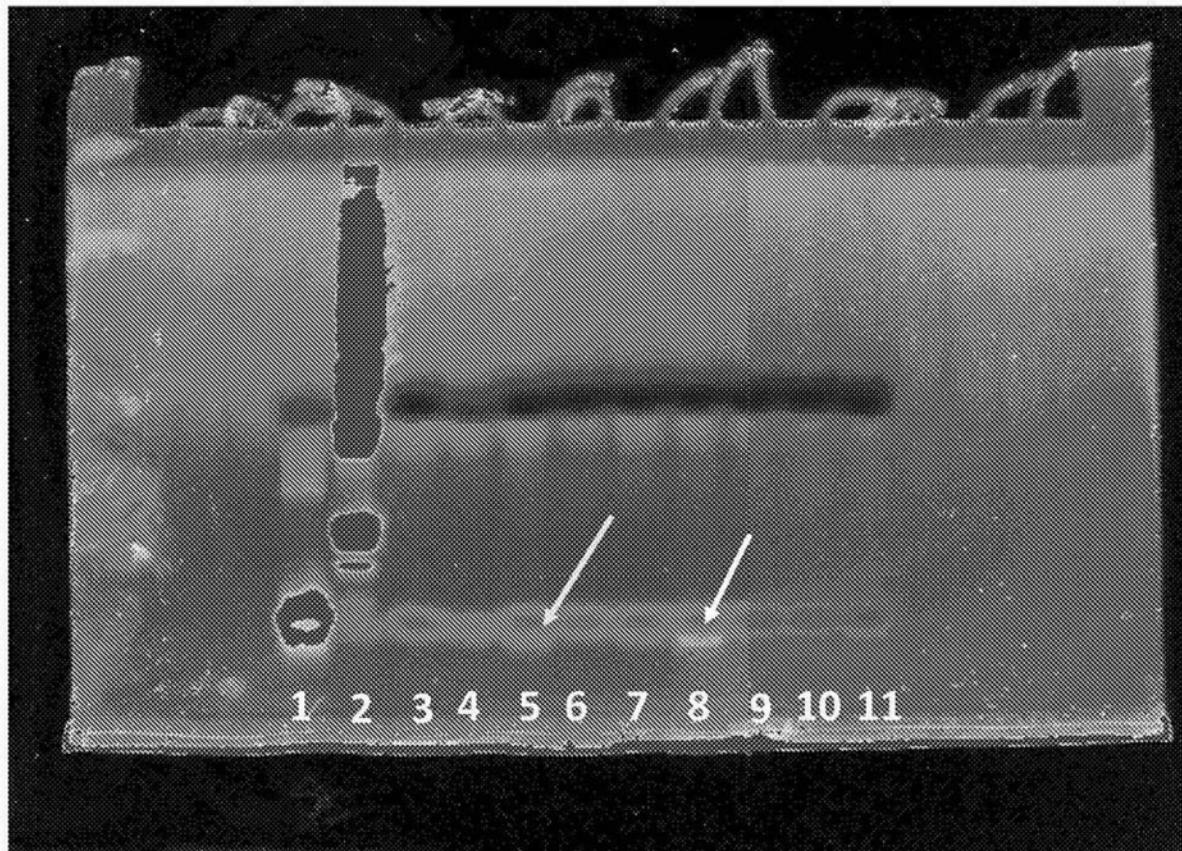


图35B



1: 50 皮摩尔 P5-acrydite
2: 25 bp 分子量标记
3-5: TCEP:GB = 25亿
6-8: TCEP:GB = 100亿

9-11: TCEP:GB = 0
3,6,9: P5:GB = 5千万
4,7,10: P5:GB = 5亿
5,8,11: P5:GB = 50亿

图36

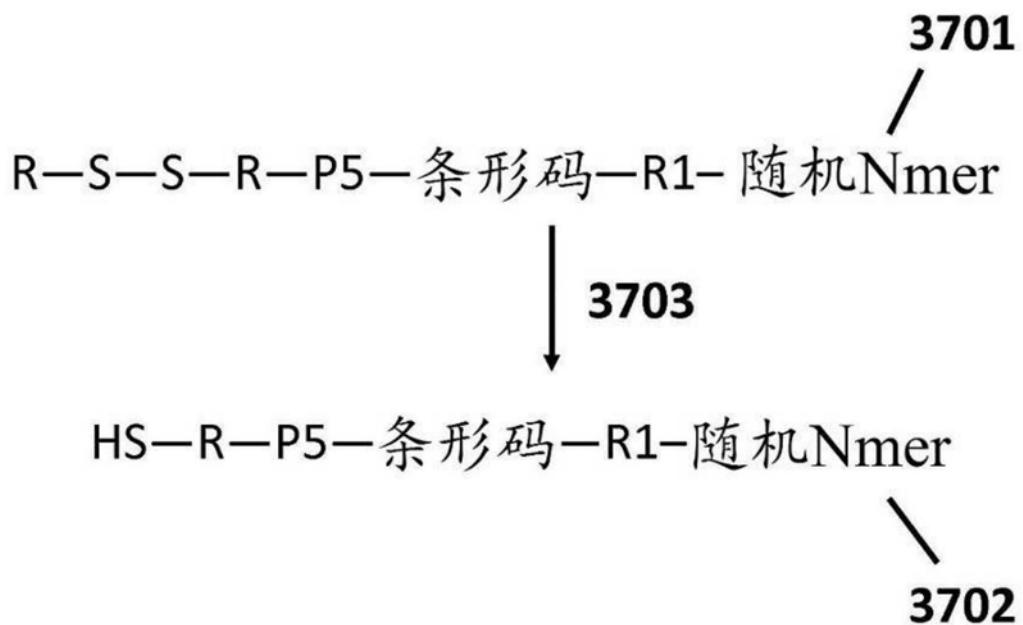
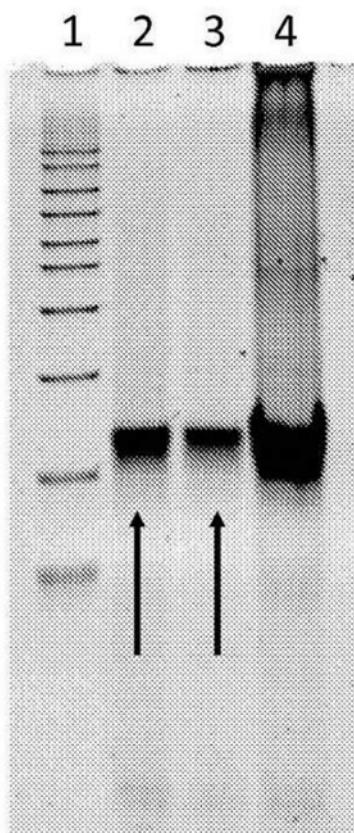
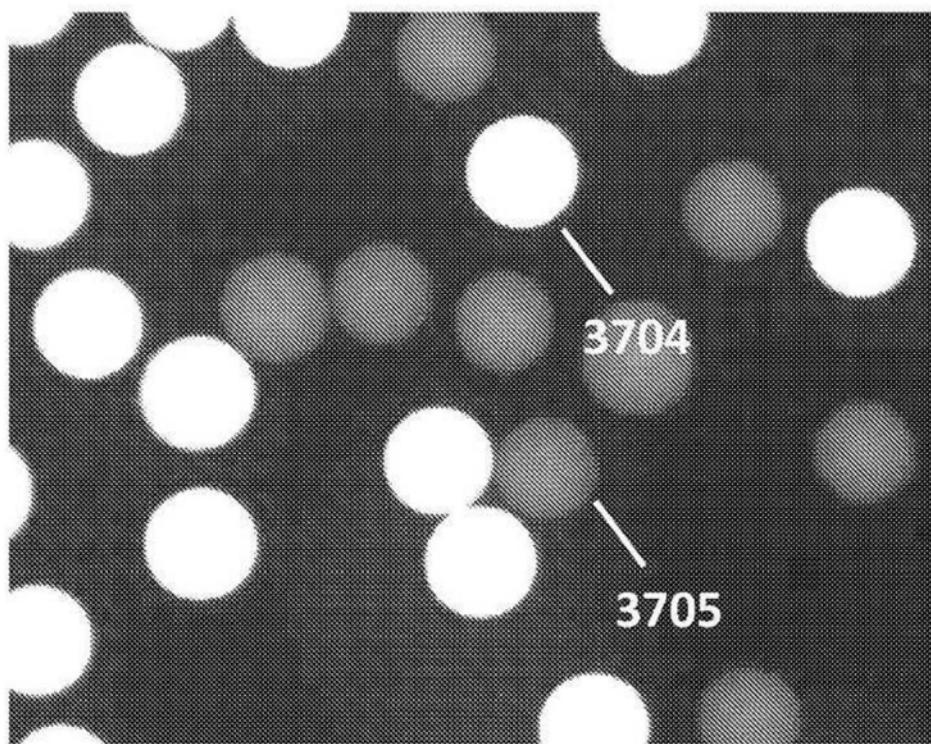


图37A



- 1: 50 bp 分子量标记
- 2: FCBC:GB = 4亿
- 3: FCBC:GB = 16亿
- 4: Acrydite 聚合化学

图37B



明亮的珠子：Acrydite聚合化学
(在凝胶珠子产生期间功能化)
暗淡的珠子：巯基-二硫键聚合化学
(在凝胶珠子产生后功能化)

图37C

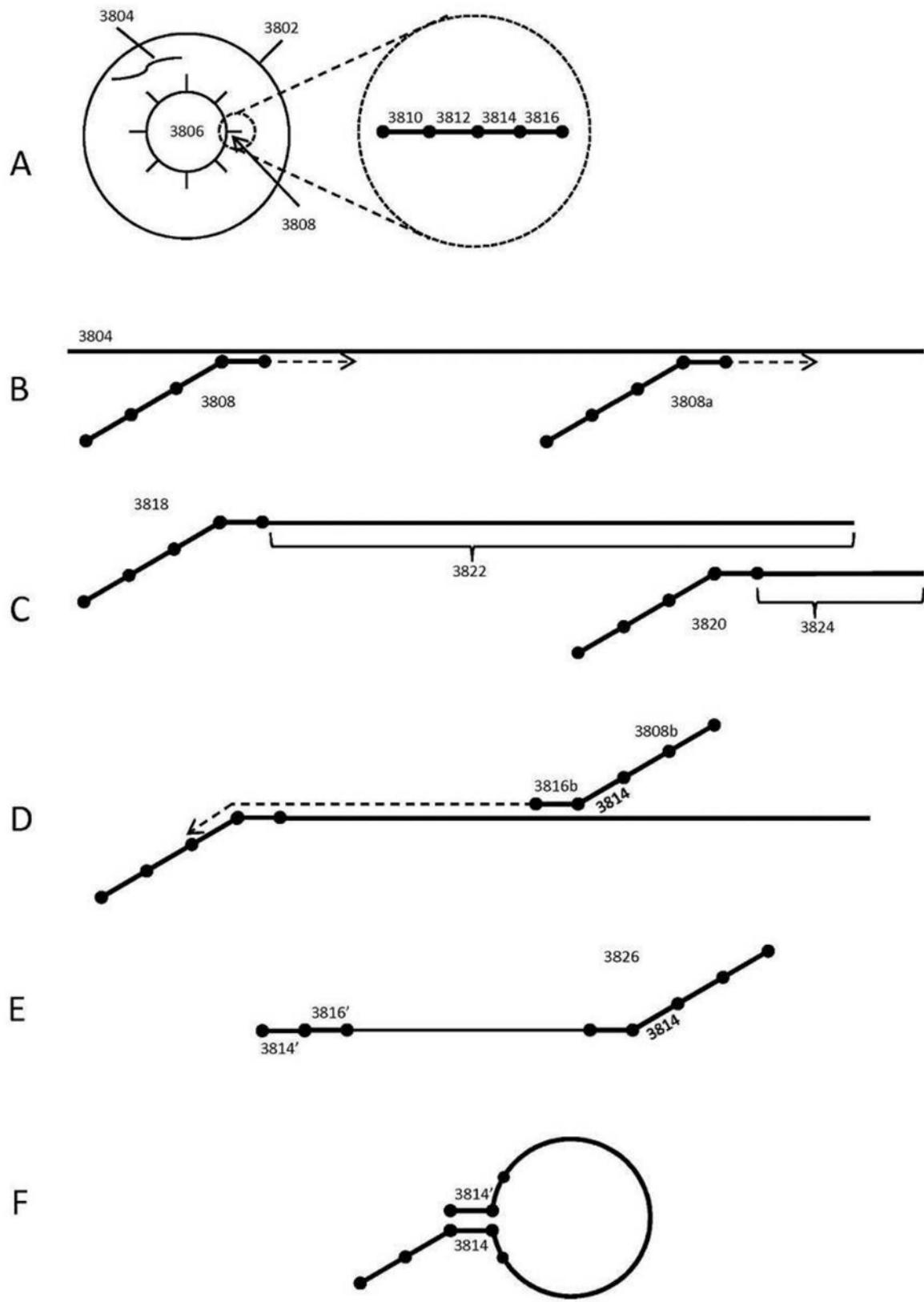


图38

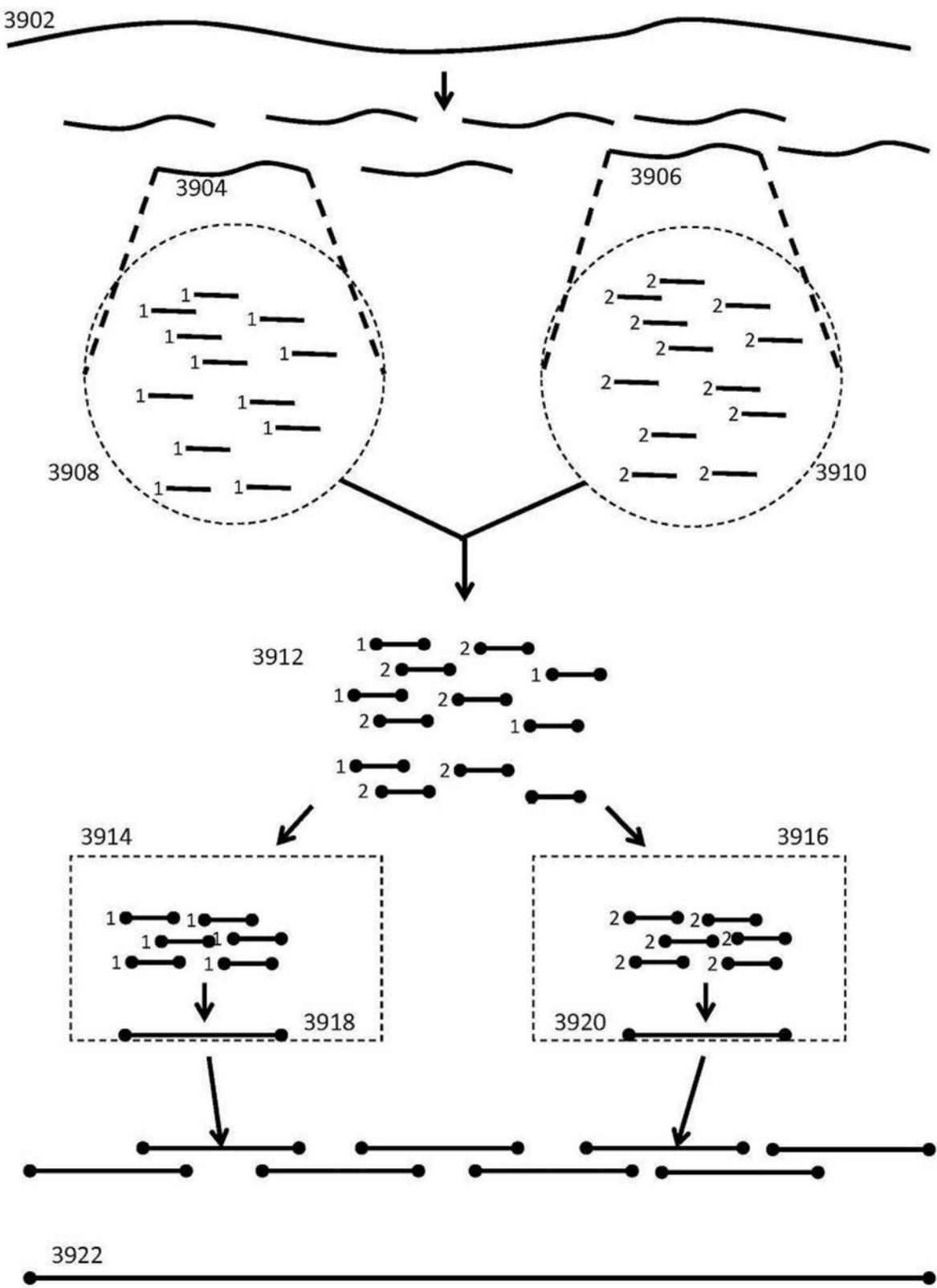


图39

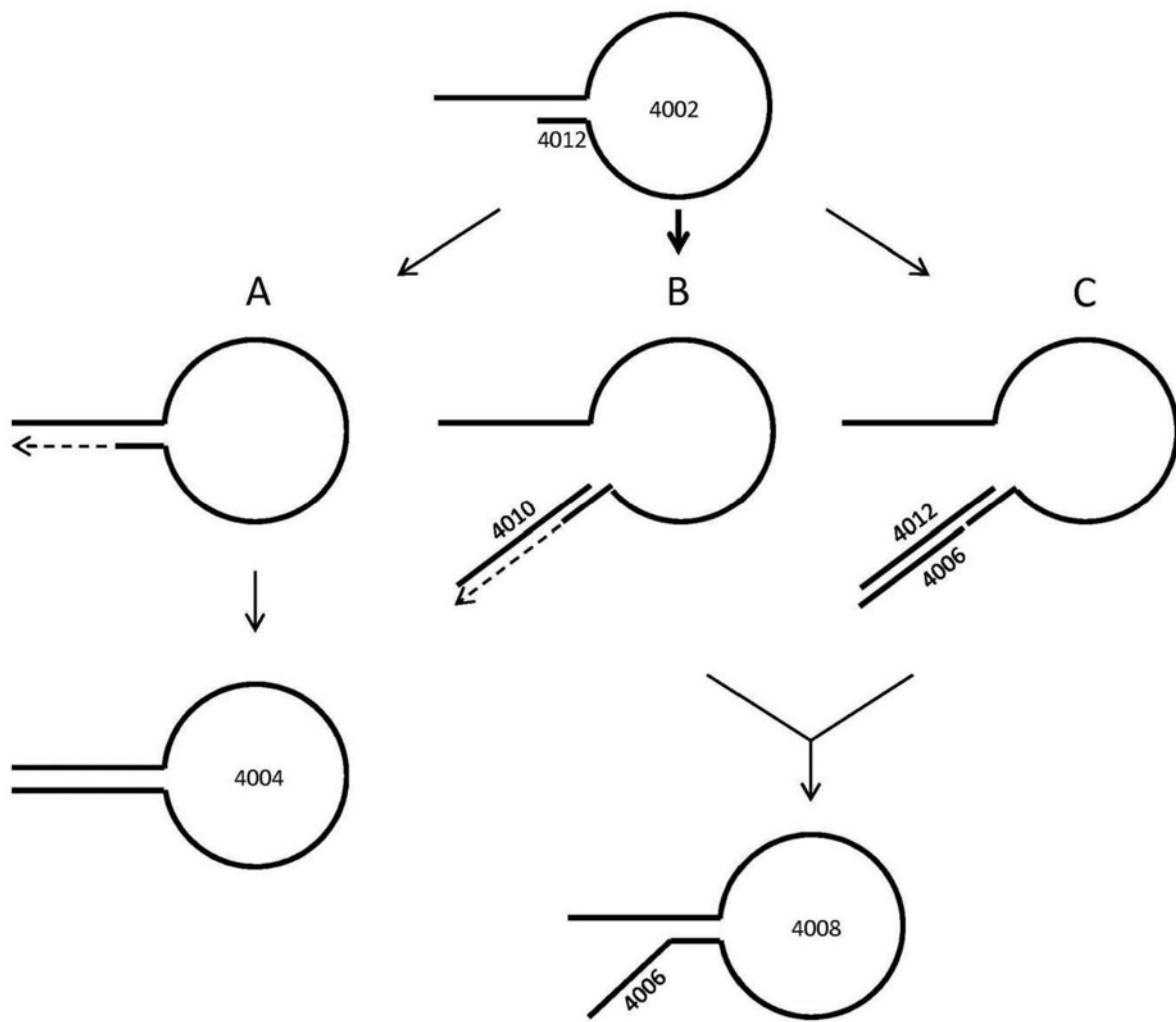


图40