



(12) 发明专利申请

(10) 申请公布号 CN 105102697 A

(43) 申请公布日 2015. 11. 25

(21) 申请号 201480020238. 9

(22) 申请日 2014. 02. 07

(30) 优先权数据

61/762, 435 2013. 02. 08 US

61/800, 223 2013. 03. 15 US

61/840, 403 2013. 06. 27 US

61/844, 804 2013. 07. 10 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

代理人 郑霞

(51) Int. Cl.

C40B 50/06(2006. 01)

C40B 20/00(2006. 01)

(85) PCT国际申请进入国家阶段日

2015. 10. 08

(86) PCT国际申请的申请数据

PCT/US2014/015427 2014. 02. 07

(87) PCT国际申请的公布数据

W02014/124338 EN 2014. 08. 14

(71) 申请人 10X 基因组学有限公司

地址 美国加利福尼亚州

(72) 发明人 本杰明·辛德森 米尔娜·雅罗什

保罗·哈登伯尔

迈克尔·史诺-莱文 凯文·奈斯

瑟奇·萨克森诺夫

权利要求书4页 说明书43页

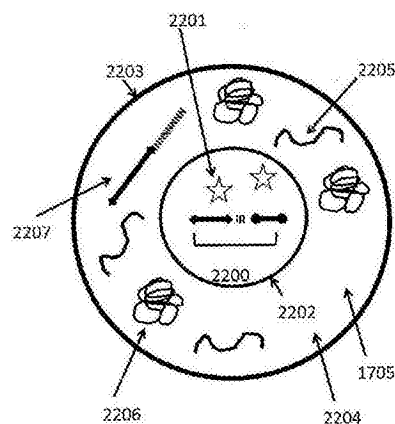
序列表15页 附图32页

(54) 发明名称

多核苷酸条形码生成

(57) 摘要

本公开内容提供了用于多核苷酸处理的组合物、方法、系统和装置。这样的多核苷酸处理可以用于多种应用,包括多核苷酸测序。在一些情况下,本公开内容提供了用于生成多核苷酸条形码文库以及用于将这样的多核苷酸附接至靶多核苷酸上的方法。



1. 一种包含至少约 1,000 个珠子的文库,其中所述至少约 1,000 个珠子中的每个珠子包含不同的条形码序列。

2. 如权利要求 1 所述的文库,其中所述不同的条形码序列包括在包含固定序列和 / 或针对测序引物的退火序列的多核苷酸中。

3. 如权利要求 1 所述的文库,其中所述不同的条形码序列为至少约 5 个核苷酸的长度。

4. 如权利要求 1 所述的文库,其中所述不同的条形码序列为至少约 10 个核苷酸的长度。

5. 如权利要求 1 所述的文库,其中所述不同的条形码序列是随机多核苷酸序列。

6. 如权利要求 1 所述的文库,其中所述不同的条形码序列是组合地生成的。

7. 如权利要求 1 所述的文库,其中所述 1,000 个珠子中的每一个包含所述不同的条形码序列的多个拷贝。

8. 如权利要求 1 所述的文库,其中所述 1,000 个珠子中的每一个包含所述不同的条形码序列的至少约 100,000 个拷贝。

9. 如权利要求 1 所述的文库,其中所述 1,000 个珠子中的每一个包含所述不同的条形码序列的至少约 1,000,000 个拷贝。

10. 如权利要求 1 所述的文库,其中所述 1,000 个珠子中的每一个包含所述不同的条形码序列的至少约 10,000,000 个拷贝。

11. 如权利要求 1 所述的文库,其进一步包含两个或更多个包含相同的条形码序列的珠子。

12. 如权利要求 1 所述的文库,其中所述 1,000 个珠子中的至少两个珠子包含相同的条形码序列。

13. 如权利要求 1 所述的文库,其中所述至少约 1,000 个珠子包含至少约 10,000 个珠子。

14. 如权利要求 1 所述的文库,其中所述至少约 1,000 个珠子包含至少约 100,000 个珠子。

15. 如权利要求 1 所述的文库,其中所述文库包含至少约 1,000 个不同的条形码序列。

16. 如权利要求 1 所述的文库,其中所述文库包含至少约 10,000 个不同的条形码序列。

17. 如权利要求 1 所述的文库,其中所述文库包含至少约 100,000 个不同的条形码序列。

18. 如权利要求 1 所述的文库,其中所述文库包含至少约 1,000,000 个不同的条形码序列。

19. 如权利要求 1 所述的文库,其中所述文库包含至少约 2,500,000 个不同的条形码序列。

20. 如权利要求 1 所述的文库,其中所述文库包含至少约 5,000,000 个不同的条形码序列。

21. 如权利要求 1 所述的文库,其中所述文库包含至少约 10,000,000 个不同的条形码序列。

22. 如权利要求 1 所述的文库,其中所述文库包含至少约 25,000,000 个不同的条形码序列。

23. 如权利要求 1 所述的文库,其中所述文库包含至少约 50,000,000 个不同的条形码序列。

24. 如权利要求 1 所述的文库,其中所述文库包含至少约 100,000,000 个不同的条形码序列。

25. 如权利要求 1 所述的文库,其中所述至少约 1,000 个珠子跨多个分区分布。

26. 如权利要求 25 所述的文库,其中所述分区是乳液的小滴。

27. 如权利要求 1 或 8-10 中任一项所述的文库,其中所述 1,000 个珠子中的每一个珠子包括在不同的分区中。

28. 如权利要求 27 所述的文库,其中所述不同的分区是乳液的小滴。

29. 如权利要求 1 所述的文库,其中所述 1,000 个珠子中的两个或更多个珠子包括在不同的分区中。

30. 如权利要求 29 所述的文库,其中所述不同的分区是乳液的小滴。

31. 如权利要求 1 所述的文库,其中所述 1,000 个珠子是水凝胶珠子。

32. 一种合成包含条形码序列的多核苷酸的文库的方法,所述方法包括:

- a. 合成多个包含条形码序列的多核苷酸;
- b. 将所述多核苷酸分隔至多个分区内,从而生成分区的多核苷酸;
- c. 扩增所述分区的多核苷酸,从而生成扩增的多核苷酸;以及
- d. 分离包含扩增的多核苷酸的分区。

33. 如权利要求 32 所述的方法,其中所述合成包括在偶联反应中包括腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶的混合物。

34. 如权利要求 32 所述的方法,其中所述分隔包括进行有限稀释,从而生成稀释的多核苷酸。

35. 如权利要求 34 所述的方法,其中所述分隔进一步包括对所述稀释的多核苷酸进行分区。

36. 如权利要求 32 所述的方法,其中所述分区中的每一个平均包含约 1 个包含条形码序列的多核苷酸。

37. 如权利要求 32 所述的方法,其中所述分区中的每一个平均包含约 0.5 个包含条形码序列的多核苷酸。

38. 如权利要求 32 所述的方法,其中所述分区中的每一个平均包含约 0.1 个包含条形码序列的多核苷酸。

39. 如权利要求 32 所述的方法,其中所述扩增通过选自聚合酶链反应、不对称聚合酶链反应、乳液 PCR(ePCR)、包括使用珠子的 ePCR、包括使用水凝胶的 ePCR、基于退火和成环的多重扩增循环(MALBAC)、单引物等温扩增及其组合的方法进行。

40. 如权利要求 32 所述的方法,其中所述扩增是使用 RNA 引物进行的。

41. 如权利要求 40 所述的方法,其进一步包括使所述扩增的多核苷酸暴露于 RNA 酶 H。

42. 如权利要求 32 所述的方法,其中所述包含条形码序列的多核苷酸中的每一个是 MALBAC 引物。

43. 如权利要求 32 所述的方法,其中所述分离是通过流动辅助的分选进行的。

44. 如权利要求 32 所述的方法,其中由多核苷酸形成发夹结构,该多核苷酸选自所述

包含条形码序列的多核苷酸和所述扩增的多核苷酸。

45. 如权利要求 44 所述的方法,其进一步包括在未退火的区域内切割所述发夹结构。

46. 如权利要求 32 所述的方法,其中选自所述包含条形码序列的多核苷酸、所述分区的多核苷酸和所述扩增的多核苷酸的多核苷酸附接至珠子上。

47. 如权利要求 32 所述的方法,其进一步包括使所述扩增的多核苷酸与部分互补序列退火。

48. 如权利要求 47 所述的方法,其中所述部分互补序列包含条形码序列。

49. 如权利要求 32 所述的方法,其进一步包括将所述扩增的多核苷酸中的至少一个附接至靶序列。

50. 如权利要求 49 所述的方法,其中所述靶序列被片段化。

51. 如权利要求 50 所述的方法,其中所述靶序列通过选自机械剪切和酶处理的方法被片段化。

52. 如权利要求 51 所述的方法,其中所述机械剪切由超声诱导。

53. 如权利要求 51 所述的方法,其中所述酶选自限制酶、片段化酶和转座酶。

54. 如权利要求 49 所述的方法,其中所述附接通过选自连接和扩增的方法进行。

55. 如权利要求 54 所述的方法,其中所述扩增是用 MALBAC 引物进行的 MALBAC 扩增,从而生成 MALBAC 扩增产物。

56. 如权利要求 55 所述的方法,其中所述 MALBAC 引物包含所述扩增的多核苷酸。

57. 如权利要求 55 所述的方法,其中所述 MALBAC 引物包含不是所述扩增的多核苷酸的多核苷酸。

58. 如权利要求 57 所述的方法,其进一步包括将所述 MALBAC 扩增产物附接至所述扩增的多核苷酸。

59. 如权利要求 32 所述的方法,其中所述分区选自小滴、胶囊和孔。

60. 如权利要求 32 所述的方法,其中所述文库包含至少约 1,000 个不同的条形码序列。

61. 如权利要求 32 所述的方法,其中所述分区包含相同的包含条形码序列的多核苷酸的多个拷贝。

62. 如权利要求 32 所述的方法,其中所述包含条形码序列的多核苷酸进一步包含选自固定序列、针对测序引物的退火序列和对于与靶多核苷酸的连接相容的序列的序列。

63. 如权利要求 32 所述的方法,其中所述文库包含至少约 100,000 个不同的条形码序列。

64. 如权利要求 32 所述的方法,其中所述文库包含至少约 500,000 个不同的条形码序列。

65. 如权利要求 32 所述的方法,其中所述文库包含至少约 1,000,000 个不同的条形码序列。

66. 如权利要求 32 所述的方法,其中所述文库包含至少约 2,500,000 个不同的条形码序列。

67. 如权利要求 32 所述的方法,其中所述文库包含至少约 5,000,000 个不同的条形码序列。

68. 如权利要求 32 所述的方法,其中所述文库包含至少约 10,000,000 个不同的条形码

序列。

69. 如权利要求 32 所述的方法,其中所述文库包含至少约 25,000,000 个不同的条形码序列。

70. 如权利要求 32 所述的方法,其中所述文库包含至少约 50,000,000 个不同的条形码序列。

71. 如权利要求 32 所述的方法,其中所述文库包含至少约 100,000,000 个不同的条形码序列。

72. 如权利要求 1-31 中任一项所述的文库或如权利要求 32-71 中任一项所述的方法在对物质进行分区、对寡核苷酸进行分区、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、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、循环肿瘤细胞(CTC)测序、多核苷酸取相、从少量细胞对多核苷酸进行测序、分析基因表达、从细胞对多核苷酸进行分区或其组合中的用途。

多核苷酸条形码生成

交叉引用

[0001] 本申请要求 2013 年 2 月 8 日提交的美国临时专利申请号 61/762, 435、2013 年 3 月 15 日提交的美国临时专利申请号 61/800, 223、2013 年 6 月 27 日提交的美国临时专利申请号 61/840, 403 和 2013 年 7 月 10 日提交的美国临时专利申请号 61/844, 804 的权益, 所述申请为所有目的通过引用以其整体并入本文。

背景技术

[0002] 多核苷酸条形码在包括下一代测序技术在内的大量应用中具有实用性。这样的条形码通常含有独特的标识符序列, 后者在以充分的多样性和规模生产时可能是极其昂贵的。合成单多核苷酸条形码的成本是合成过程中每个碱基的成本和多核苷酸长度的函数。因此, 合成多个各自具有不同序列的条形码的成本等于每个碱基的成本乘以每个分子的碱基数, 再乘以该多个条形码内的分子数。当前, 合成 DNA 序列要花费每个碱基约 0.10 美元。对于包含上万个至数百万个条形码的条形码文库, 该成本太过高昂。因此, 非常需要改善的生成条形码文库的方法。

发明内容

[0003] 本公开内容提供了用于生成多核苷酸条形码的方法、组合物、系统和试剂盒以及这类多核苷酸条形码的用途。这类多核苷酸条形码可以用于任何合适的应用。

[0004] 本公开内容的一个方面提供了包含一个或多个多核苷酸的文库, 每个多核苷酸包含条形码序列, 其中该多核苷酸布置在一个或多个分区内, 并且其中该文库包含至少约 1,000 个不同的条形码序列。

[0005] 在一些情况下, 所述条形码序列是至少约 5 个核苷酸的长度。另外, 该条形码序列可以是随机的多核苷酸序列。

[0006] 而且, 所述分区可以平均包含约 1 个多核苷酸、约 0.5 个多核苷酸或约 0.1 个多核苷酸。所述分区可以是小滴、胶囊、孔或珠子。

[0007] 另外, 所述文库可以包含至少约 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、至少约 500,000 个不同的条形码序列、至少约 1,000,000 个不同的条形码序列、至少约 2,500,000 个不同的条形码序列、至少约 5,000,000 个不同的条形码序列、至少约 10,000,000 个、至少约 25,000,000 个、至少约 50,000,000 个或至少约 100,000,000 个不同的条形码序列。

[0008] 在一些情况下, 所述分区可以包含相同多核苷酸的多个拷贝。

[0009] 另外, 所述多核苷酸中的每一个可以包含选自固定序列、针对测序引物的退火序列和对于与靶多核苷酸的连接相容的序列的序列。

[0010] 在一些情况下, 所述多核苷酸中的每一个是 MALBAC 引物。

[0011] 本公开内容的另一方面提供了合成包含条形码序列的多核苷酸的文库的方法, 该方法包括: a) 合成多个包含条形码序列的多核苷酸; b) 将该多核苷酸分隔至多个分区内,

从而生成分区的多核苷酸;c) 扩增该分区的多核苷酸,从而生成扩增的多核苷酸;以及 d) 分离包含该扩增的多核苷酸的分区。在一些情况下,所述合成包括在偶联反应中包括腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶的混合物。

[0012] 而且,所述分隔可以包括进行有限稀释,从而生成稀释的多核苷酸。在一些情况下,所述分隔进一步包括对所述稀释的多核苷酸进行分区。

[0013] 另外,所述扩增可以通过选自聚合酶链反应、不对称聚合酶链反应、乳液 PCR(ePCR)、包括使用珠子的 ePCR、包括使用水凝胶的 ePCR、基于退火和成环的多重扩增循环(MALBAC)、单引物等温扩增及其组合的方法来进行。在一些情况下,所述扩增使用 RNA 引物来进行,并且可以包括将扩增的多核苷酸暴露于 RNA 酶 H。

[0014] 在一些情况下,所述包含条形码序列的多核苷酸中的每一个是 MALBAC 引物。

[0015] 在一些情况下,所述分离可以通过流动辅助的分选来进行。

[0016] 另外,可以由多核苷酸形成发夹结构,该多核苷酸选自包含条形码序列的多核苷酸和扩增的多核苷酸。在一些情况下,方法可以进一步包括在未退火的区域内切割发夹结构。

[0017] 而且,选自所述包含条形码序列的多核苷酸、所述分区的多核苷酸和所述扩增的多核苷酸的多核苷酸可以附接至珠子上。

[0018] 所述方法可以进一步包括使扩增的多核苷酸与部分互补序列退火。所述部分互补序列可以包含条形码序列。

[0019] 所述方法可以进一步包括将扩增的多核苷酸中的至少一个附接至靶序列。该靶序列可以被片段化。在一些情况下,该靶序列通过选自机械剪切和酶处理的方法被片段化。该机械剪切可以由超声诱导。在一些情况下,所述酶选自限制酶、片段化酶和转座酶。另外,所述附接可以通过选自连接和扩增的方法进行。

[0020] 在一些情况下,所述扩增是用 MALBAC 引物进行的 MALBAC 扩增,从而生成 MALBAC 扩增产物。在一些情况下,所述 MALBAC 引物包含扩增的多核苷酸。在一些情况下,所述 MALBAC 引物包含不是所述扩增的多核苷酸的多核苷酸。在这样的情况下,所述方法可以进一步包括将 MALBAC 扩增产物附接至扩增的多核苷酸。

[0021] 另外,每个分区可以平均包含约 1 个包含条形码序列的多核苷酸、0.5 个包含条形码序列的多核苷酸或 0.1 个包含条形码序列的多核苷酸。而且,所述分区可以选自小滴、胶囊和孔。

[0022] 在一些情况下,所述文库包含至少约 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、至少约 500,000 个不同的条形码序列、至少约 1,000,000 个不同的条形码序列、至少约 2,500,000 个不同的条形码序列、至少约 5,000,000 个不同的条形码序列、至少约 10,000,000 个、至少约 25,000,000 个、至少约 50,000,000 个或至少约 100,000,000 个不同的条形码序列。

[0023] 在一些情况下,所述分区包含相同的包含条形码序列的多核苷酸的多个拷贝。

[0024] 而且,所述包含条形码序列的多核苷酸可以包含选自固定序列、针对测序引物的退火序列和对于与靶多核苷酸的连接相容的序列的序列。

[0025] 本公开内容的另一方面提供了一种包含至少约 1,000 个珠子的文库,其中所述至少约 1,000 个珠子中的每个珠子包含不同的条形码序列。在一些情况下,所述不同的条形

码序列可以包括在包含固定序列和 / 或针对测序引物的退火序列的多核苷酸中。在一些情况下,所述不同的条形码序列可以是至少约 5 个核苷酸或至少约 10 个核苷酸的长度。在一些情况下,所述不同的条形码序列可以是随机多核苷酸序列或者可以组合地生成。

[0026] 而且,所述 1,000 个珠子中的每一个可以包含不同条形码序列的多个拷贝。例如,所述 1,000 个珠子中的每一个可以包含不同条形码序列的至少约 100,000 个、至少约 1,000,000 个或至少约 10,000,000 个拷贝。在一些情况下,所述文库可以进一步包含两个或更多个包含相同条形码序列的珠子。在一些情况下,所述 1,000 个珠子中的至少两个珠子可以包含相同的条形码序列。此外,所述至少约 1,000 个珠子包含至少约 10,000 个或至少约 100,000 个珠子。

[0027] 另外,所述文库可以包含至少约 1,000 个、至少约 10,000 个、至少约 100,000 个、至少约 1,000,000 个、至少约 2,500,000 个、至少约 5,000,000 个、至少约 10,000,000 个、至少约 25,000,000 个、至少约 50,000,000 个或至少约 100,000,000 个不同的条形码序列。

[0028] 在一些情况下,所述至少约 1,000 个珠子可以跨多个分区分布。在一些情况下,所述分区可以是乳液的小滴。在一些情况下,所述 1,000 个珠子中的每一个珠子可以包括在不同的分区中。在一些情况下,所述不同的分区可以是乳液的小滴。在一些情况下,所述 1,000 个珠子中的两个或更多个珠子可以包括在不同的分区中。在一些情况下,所述不同的分区可以是乳液的小滴。在一些情况下,所述 1,000 个珠子可以是水凝胶珠子。

[0029] 本公开内容的另一方面提供了本文描述的文库、组合物、方法、装置或试剂盒在对物质进行分区、对寡核苷酸进行分区、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、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、循环肿瘤细胞(CTC)测序、多核苷酸取相、从少量细胞对多核苷酸进行测序、分析基因表达、从细胞对多核苷酸进行分区或其组合中的用途。

援引并入

[0030] 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而并入本文,其程度如同具体地和单独地指出各个单独的出版物、专利或专利申请通过引用而并入。

附图说明

[0031] 本公开内容的方法、组合物、系统和装置的新特征在所附的权利要求中具体阐述。通过参考以下对其中利用了本发明方法、组合物、系统和装置的原理的说明性实施方案加以阐述的详细描述和附图,将会获得对本公开内容的特征和优点的更好的理解,附图中:

[0032] 图 1 示意性地描绘了示例性叉状衔接子。

[0033] 图 2 示意性地描绘了条形码区域的示例性布置。

[0034] 图 3 描绘了连接至靶多核苷酸的相对末端的两个叉状衔接子的示例性序列。

[0035] 图 4 是用于产生实施例 1 中描述的叉状衔接子的示意性示例方法。

[0036] 图 5 是实施例 2 中描述的在胶囊内的胶囊的示意性示例。

[0037] 图 6 是实施例 3 中描述的在胶囊内的胶囊的示意性示例。

- [0038] 图 7 是可根据实施例 4 的方法产生的产物（或中间体）的示意性示例。
- [0039] 图 8a-c 描绘了实施例 4 中描述的示例性序列。
- [0040] 图 9a-j 描绘了实施例 5 中描述的示例性序列。
- [0041] 图 10a-e 描绘了实施例 6 中描述的示例性序列。
- [0042] 图 11a-d 示意性地描绘了实施例 7 中描述的方法和结构。
- [0043] 图 12 示意性地描绘了通过示例性流动聚焦方法产生胶囊。
- [0044] 图 13 示意性地描绘了通过示例性流动聚焦方法产生在胶囊内的胶囊。
- [0045] 图 14a-e 示意性地描绘了实施例 8 中描述的方法和结构。
- [0046] 图 15a-e 示意性地描绘了实施例 9 中描述的方法和结构。
- [0047] 图 16 示意性地描绘了实施例 10 中描述的方法和结构。
- [0048] 图 17 示意性地描绘了实施例 11 中描述的在胶囊内的胶囊。
- [0049] 图 18 示意性地描绘了实施例 12 中描述的在胶囊内的胶囊。
- [0050] 图 19a-e 描绘了实施例 13 中描述的示例性序列。图 19f 描绘了实施例 13 中描述的示例性方法和结构。
- [0051] 图 20 示意性地描绘了实施例 14 中描述的在胶囊内的胶囊。
- [0052] 图 21a-c 示意性地描绘了实施例 15 中描述的方法和结构。
- [0053] 图 22 示意性地描绘了实施例 16 中描述的在胶囊内的胶囊。
- [0054] 图 23 示意性地描绘了实施例 17 中描述的在胶囊内的胶囊。

具体实施方式

[0055] 虽然本文已经显示并描述了本发明的多种实施方案，但是对本领域技术人员来说显而易见的是，这些实施方案仅通过举例的方式来提供。本领域技术人员在不背离本发明的情况下可以想到许多变化、改变和替换。应当理解，可以采用本文所述的本发明实施方案的各种替代方案。

[0056] 本公开内容提供了用于生成多核苷酸条形码的方法、组合物、系统和试剂盒以及这类多核苷酸条形码的用途。这类多核苷酸条形码可以用于任何合适的应用。在一些情况下，本公开内容中提供的多核苷酸条形码可以在下一代测序反应中使用。下一代测序反应包括全基因组的测序、特定序列如单核苷酸多态性 (SNP) 和其他突变的检测、核酸（例如脱氧核糖核酸）插入的检测和核酸缺失的检测。

[0057] 除非另有说明，否则本文描述的方法、组合物、系统和试剂盒的应用可以结合有机化学、聚合物技术、微流体、分子生物学、重组技术、细胞生物学、生物化学和免疫学的任何常规技术。这样的常规技术包括孔和微孔构建、胶囊生成、乳液生成、点印、微流体装置构建、聚合物化学、限制性消化、连接、克隆、多核苷酸测序和多核苷酸序列装配。合适的技术的具体的、非限制性的说明在本公开内容通篇描述。然而，也可以使用等同的程序。对某些技术的描述可见于标准实验室手册，如 Genome Analysis: A Laboratory Manual Series (第 I-IV 卷), Using Antibodies: A Laboratory Manual, Cells: A Laboratory Manual, PCR Primer: A Laboratory Manual, 和 Molecular Cloning: A Laboratory Manual (均来自 Cold Spring Harbor Laboratory Press), 和 “Oligonucleotide Synthesis: A Practical Approach” 1984, IRL Press London, 全部为了所有目的而通过引

用以其整体并入本文。

I. 定义

[0058] 本文所用的术语仅仅是为了描述具体实施方案的目的,而并非旨在限制。

[0059] 如本文使用的,单数形式“一个”、“一种”和“该”也意图包括复数形式,除非上下文明确指出不是这样。此外,就术语“包括”、“含有”、“具有”、“有”、“带有”、“如”或其变化形式在说明书和 / 或权利要求中的使用来说,这些术语并非限制性的,且旨在以类似于术语“包含”的方式包含在内。

[0060] 如本文所用的术语“约”通常是指在特定应用的背景中比指定数值大或小 15% 的范围。例如,约 10 将包括从 8.5 到 11.5 的范围。

[0061] 如本文所用的术语“条形码”通常是指可以附接至分析物以传递关于该分析物的信息的标记。例如,条形码可以是附接至在特定分区内包含的靶多核苷酸的片段上的多核苷酸序列。该条形码然后可以用靶多核苷酸的片段进行测序。多个序列上相同条形码的存在可以提供关于该序列的起源的信息。例如,条形码可以指示序列来自于特定分区和 / 或基因组的邻近区域。这尤其可以用于在测序前合并数个分区时的序列装配。

[0062] 如本文使用的,术语“bp”通常是指“碱基对”的缩写。

[0063] 如本文使用的,术语“微孔”通常是指容积小于 1mL 的孔。根据应用,微孔可以制成多种容积。例如,微孔可以制成适合容纳本文描述的任何分区体积的大小。

[0064] 如本文使用的,术语“分区”可以是动词或名词。当用作动词(例如,“进行分区”)时,该术语通常是指物质或样品(例如,多核苷酸)在可用来将一个级分(或子部分)与另一个隔离开的容器之间的分级(例如,细分)。使用名词“分区”提及这样的容器。例如,可以使用微流体、稀释、分配等进行分区。分区可以是,例如,孔、微孔、孔洞、小滴(例如,乳液中的小滴)、乳液的连续相、试管、斑点、胶囊、珠子、稀溶液中的珠子的表面,或任何其他用于将样品的一个级分与另一个级分离开的合适的容器。分区也可以包含另一个分区。

[0065] 如本文使用的,术语“多核苷酸”或“核酸”通常是指包含多个核苷酸的分子。示例性的多核苷酸包括脱氧核糖核酸、核糖核酸及其合成类似物,包括肽核酸。

[0066] 如本文使用的,术语“物质”通常是指能够与本公开内容的方法、组合物、系统、装置和试剂盒一起使用的任何物质。物质的实例包括试剂、分析物、细胞、染色体、标记分子或成组分子、条形码和包含任何这样的物质的任何样品。可以使用任何合适的物质,如在本公开内容中的其他部分更加充分讨论的。

II. 多核苷酸条形码化

[0067] 某些应用,例如多核苷酸测序,可以依赖于独特的标识符(“条形码”)来鉴定序列的起源,以及,例如,从经测序的片段装配较大的序列。因此,可能希望在测序前向多核苷酸片段上添加条形码。条形码可以是多种不同的形式,包括多核苷酸条形码。根据具体应用,条形码可以以可逆或不可逆的方式附接至多核苷酸片段上。另外,条形码可以允许在测序过程中对单个多核苷酸片段进行鉴别和 / 或量化。

[0068] 可以将条形码加载至分区中以使得向特定分区中引入一个或多个条形码。在一些情况下,每个分区可以含有不同组的条形码。这可以通过将条形码直接分配至分区中,或者通过将条形码置于在另一分区内包含的分区中来完成。

[0069] 可以以预期的或预测的每个待条形码化的物质(例如,多核苷酸片段、多核苷酸

链、细胞等)上的条形码的比例,将条形码加载到分区内。在一些情况下,将条形码加载到分区内,以使得每个物质加载约 0.0001、0.001、0.1、1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000 或 200000 个条形码。在一些情况下,将条形码加载到分区内,以使得每个物质加载多于约 0.0001、0.001、0.1、1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000 或 200000 个条形码。在一些情况下,将条形码加载到分区内,以使得每个物质加载少于约 0.0001、0.001、0.1、1、2、3、4、5、6、7、8、9、10、20、50、100、500、1000、5000、10000 或 200000 个条形码。

[0070] 当每个多核苷酸片段存在多于一个条形码时,这样的条形码可以是相同条形码的拷贝,或者可以是不同的条形码。例如,可将附接过程设计成将多个相同的条形码附接至单个多核苷酸片段上,或将多个不同的条形码附接至该多核苷酸片段上。

[0071] 本文所提供的方法可以包括向分区中加载对于条形码与多核苷酸片段的附接而言所必需的试剂。在连接反应的情况下,可以将包括限制酶、连接酶、缓冲液、衔接子、条形码等的试剂加载到分区中。在通过扩增进行条形码化的情况下,可以将包括引物、DNA 聚合酶、dNTP、缓冲液、条形码等的试剂加载到分区中。在转座子介导的条形码化(例如 NEXTERA)的情况下,可以将包括转座体(即,转座酶和转座子末端复合物)、缓冲液等的试剂加载到分区中。在 MALBAC 介导的条形码化的情况下,可以将包括 MALBAC 引物、缓冲液等的试剂加载到分区中。如本公开内容通篇所描述的,这些试剂可以直接地或经由另一分区加载到分区内。

[0072] 条形码可以利用粘端或平端连接至多核苷酸片段。条形码化的多核苷酸片段也可以通过用包含条形码的引物扩增多核苷酸片段而生成。在一些情况下,多核苷酸片段的 MALBAC 扩增可以用来生成条形码化的多核苷酸片段。用于 MALBAC 的引物可以包含或者可以不包含条形码。在 MALBAC 引物不包含条形码的情况下,可以通过其他扩增方法如 PCR 将条形码添加至 MALBAC 扩增产物上。条形码化的多核苷酸片段也可以使用转座子介导的方法生成。对于本公开内容讨论的任何其他物质,根据测定或过程的需要,这些模块可以包含在相同或不同的分区内。

[0073] 在一些情况下,条形码可以从设计用于以模块形式装配的较小组分组合地装配。例如,1A、1B 和 1C 这三个模块可以组合地装配以产生条形码 1ABC。这样的组合装配可以显著地降低合成多个条形码的成本。例如,由 3 个 A 模块、3 个 B 模块和 3 个 C 模块组成的组合系统可以仅从 9 个模块生成 $3 \times 3 \times 3 = 27$ 个可能的条形码序列。

[0074] 在一些情况下,如本公开内容的其他部分进一步描述的,可以通过混合两种寡核苷酸并使其杂交以产生退火的或部分退火的寡核苷酸(例如,叉状衔接子)而组合地装配条形码。这些条形码可以包含一种或多种核苷酸的突出端,以利于与待条形码化的多核苷酸片段的连接。在一些情况下,反义链的 5' 端可以被磷酸化以确保双链连接。使用这种方法,可以通过例如混合寡核苷酸 A 和 B、A 和 C、A 和 D、B 和 C、B 和 D 等来装配不同的模块。如本公开内容的其他部分更详细描述,退火的寡核苷酸也可以合成为具有发夹环的单分子,该发夹环可以在连接至待条形码化的多核苷酸后切割。

[0075] 如本公开内容的其他部分更详细描述,多核苷酸彼此之间的附接可以依赖于杂交相容的突出端。例如,A 与 T 之间的杂交通常用来确保片段之间的连接相容性。在一些情况下,A 突出端可以通过用酶如 Taq 聚合酶处理来产生。在一些情况下,可以使用限制酶

产生具有单碱基 3' 突出端的切割产物,该突出端可以是例如 A 或 T。留下单碱基 3' 突出端的限制酶的实例包括 MnII、HphI、Hpy188I、HpyAV、HpyCH4III、MboII、BciVI、BmrI、AhdI 和 XcmI。在其他情况下,可以用限制酶生成不同的突出端(例如,5' 突出端、大于单碱基的突出端)。可以用来生成突出端的限制酶包括 BfuCI、Taq^oI、BbvI、BccI、BceAI、BcoDI、BsmAI 和 BsmFI。

III. 分区的条形码文库的生成

[0076] 在一些情况下,本公开内容提供了生成分区的条形码文库的方法以及按照这样的方法产生的文库。在一些情况下,本文提供的方法结合 DNA 序列的随机合成、分隔成分区、分隔的序列的扩增,以及扩增的分隔的序列的分离以提供包含在分区内的条形码的文库。

a. 多核苷酸条形码的随机合成

[0077] 在一些情况下,本文描述的方法利用多核苷酸合成的随机方法,包括 DNA 合成的随机方法。在随机 DNA 合成过程中,可以将 A、C、G 和 / 或 T 的任意组合添加到偶联步骤中,以使得该偶联步骤中每种类型的碱基偶联至产物的子集。如果 A、C、G 和 T 以相等的浓度存在,则大约四分之一的产物将掺入每种碱基。连续的偶联步骤,以及偶联反应的随机性质,使得能够生成 4^n 个可能的序列,其中 n 是多核苷酸中碱基的数目。例如,长度为 6 的随机多核苷酸的文库可具有 $4^6 = 4,096$ 个成员的多样性,而长度为 10 的文库将具有 1,048,576 个成员的多样性。因此,能够生成非常大且复杂的文库。这些随机序列可以充当条形码。

[0078] 任何合适的合成碱基也可以与本发明一起使用。在一些情况下,可以改变每个偶联步骤中包括的碱基以便合成优选的产物。例如,在每个偶联步骤中存在的碱基数目可以是 1、2、3、4、5、6、7、8、9、10 或更多。在一些情况下,在每个偶联步骤中存在的碱基数目可以是至少 1、2、3、4、5、6、7、8、9、10 或更多。在一些情况下,在每个偶联步骤中存在的碱基数目可以少于 2、3、4、5、6、7、8、9 或 10。

[0079] 也可以改变单个碱基的浓度以便合成优选的产物。例如,任何碱基可以以另一碱基的浓度的约 0.1、0.5、1、5 或 10 倍的浓度存在。在一些情况下,任何碱基可以以另一碱基的浓度的至少约 0.1、0.5、1、5 或 10 倍的浓度存在。在一些情况下,任何碱基可以以另一碱基的浓度的低于约 0.1、0.5、1、5 或 10 倍的浓度存在。

[0080] 根据应用,随机多核苷酸序列的长度可以是任何合适的长度。在一些情况下,随机多核苷酸序列的长度可以是 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。在一些情况下,随机多核苷酸序列的长度可以是至少 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。在一些情况下,随机多核苷酸序列的长度可以少于 4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个核苷酸。

[0081] 在一些情况下,所述文库由成员的数目来定义。在一些情况下,文库可以包含约 256、1024、4096、16384、65536、262144、1048576、4194304、16777216、67108864、268435456、1073741824、4294967296、17179869184、68719476736、 2.74878×10^{11} 或 1.09951×10^{12} 个成员。在一些情况下,文库可以包含至少约 256、1024、4096、16384、65536、262144、1048576、4194304、16777216、67108864、268435456、1073741824、4294967296、17179869184、68719476736、 2.74878×10^{11} 或 1.09951×10^{12} 个成员。在一些情况下,文库可以包含少于约 256、1024、4096、16384、65536、262144、1048576、4194304、16777216、67108864、268435456、1073741824、4294967296、17179869184、68719476736、 2.74878×10^{11} 或 1.09951×10^{12} 个成

员。在一些情况下,该文库是条形码文库。在一些情况下,条形码文库可以包含至少约 1000、10000、100000、1000000、2500000、5000000、10000000、25000000、50000000 或 100000000 个不同的条形码序列。

[0082] 随机条形码文库还可以包含其他多核苷酸序列。在一些情况下,这些其他多核苷酸序列在性质上是非随机的,并且包括例如引物结合位点、用于生成叉状衔接子的退火位点、固定序列,和能够与靶多核苷酸序列退火,因而能够对多核苷酸序列进行条形码化的区域。

b. 多核苷酸向分区内的分隔

[0083] 在包含随机条形码序列的多核苷酸合成后,将多核苷酸分区到单独的区室中以生成包含条形码序列的分区的多核苷酸的文库。可以使用任何合适的分隔方法和任何合适的分区或分区内的分区。

[0084] 在一些情况下,通过稀释包含随机条形码序列的多核苷酸的混合物以使得特定体积的该稀释液平均含有少于一个多核苷酸来进行分区。然后将该特定体积的稀释液转移至分区。在任何多个分区中,每个分区因此有可能具有一个或零个多核苷酸分子。

[0085] 在一些情况下,可以进行稀释以使得每个分区包含约 0.001、0.01、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2 个或更多个分子。在一些情况下,可以进行稀释以使得每个分区包含至少约 0.001、0.01、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2 个或更多个分子。在一些情况下,可以进行稀释以使得每个分区包含少于约 0.001、0.01、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1 或 2 个分子。

[0086] 在一些情况下,约 10%、20%、30%、40%、50%、60%、70%、80% 或 90% 的分区包含指定数目的分子。在一些情况下,至少约 10%、20%、30%、40%、50%、60%、70%、80% 或 90% 的分区包含指定数目的分子。在一些情况下,少于约 10%、20%、30%、40%、50%、60%、70%、80% 或 90% 的分区包含指定数目的分子。

[0087] 在一些情况下,约 10%、20%、30%、40%、50%、60%、70%、80% 或 90% 的分区包含一个或更少的多核苷酸。在一些情况下,至少约 10%、20%、30%、40%、50%、60%、70%、80% 或 90% 的分区包含一个或更少的多核苷酸。在一些情况下,少于约 10%、20%、30%、40%、50%、60%、70%、80% 或 90% 的分区包含一个或更少的多核苷酸。

[0088] 在一些情况下,分区是孔、微孔、孔洞、小滴(例如,乳液中的小滴)、乳液的连续相、试管、斑点、胶囊、珠子的表面,或任何其他用于将样品的一个级分与另一个级分离的合适的容器。在分区包括珠子的情况下,用于扩增的引物可以衔接至该珠子。分区在本公开内容的其他部分更详细地描述。

c. 分区的多核苷酸的扩增

[0089] 然后扩增如上所述分区的多核苷酸,以便生成对于靶多核苷酸序列的条形码化而言足够的材料。可以使用任何合适的扩增方法,包括聚合酶链反应(PCR)、连接酶链反应(LCR)、解旋酶依赖性扩增、指数后线性 PCR(LATE-PCR)、不对称扩增、数字 PCR、简并寡核苷酸引物 PCR(DOP-PCR)、引物延伸预扩增 PCR(PEP-PCR)、连接介导的 PCR、滚环扩增、多重置换扩增(MDA)和单引物等温扩增(SPIA)、乳液 PCR(ePCR)、包括使用珠子的 ePCR、包括使用水凝胶的 ePCR、基于退火和成环的多重扩增循环(MALBAC)及其组合。MALBAC 方法描述于例如 Zong 等人, *Science*, 338(6114), 1622-1626(2012), 其通过引用以其整体并入本文。

[0090] 在一些情况下,例如,生成单链产物的扩增方法(例如,不对称扩增、SPIA 和 LATE-PCR)可能是优选的。在一些情况下,生成双链产物的扩增方法(例如,标准 PCR)可能是优选的。在一些情况下,扩增方法将指数扩增分区的多核苷酸。在一些情况下,扩增方法将线性扩增分区的多核苷酸。在一些情况下,扩增方法将首先指数扩增然后线性扩增多核苷酸。而且,单一类型的扩增可以用来扩增多核苷酸,或者可以用不同类型的扩增的顺序步骤来完成扩增。例如,ePCR 可以与其他轮次的 ePCR 组合,或者可以与不同类型的扩增组合。

[0091] 进行扩增直到产生适量的包含条形码的多核苷酸。在一些情况下,可以进行扩增 10、15、20、25、30、35、40、45、50、55、60 个或更多个循环。在一些情况下,可以进行扩增至少 10、15、20、25、30、35、40、45、50、55、60 个或更多个循环。在一些情况下,可以进行扩增少于 10、15、20、25、30、35、40、45、50、55 或 60 个循环。

[0092] 在一些情况下,可以进行扩增直到在每个分区中产生特定量的多核苷酸产物。在一些情况下,进行扩增直到多核苷酸产物的量为约 10,000,000,000、5,000,000,000、1,000,000,000、500,000,000、100,000,000、50,000,000、10,000,000、5,000,000、1,000,000、500,000、400,000、300,000、200,000 或 100,000 个分子。在一些情况下,进行扩增直到多核苷酸产物的量为至少约 100,000、200,000、300,000、400,000、500,000、1,000,000、5,000,000、10,000,000、50,000,000、100,000,000、500,000,000、1,000,000,000、5,000,000,000 或 10,000,000,000 个分子。在一些情况下,进行扩增直到多核苷酸产物的量少于约 10,000,000,000、5,000,000,000、1,000,000,000、500,000,000、100,000,000、50,000,000、10,000,000、5,000,000、1,000,000、500,000、400,000、300,000、200,000 或 100,000 个分子。

d. 包含扩增的序列的分区的分离

[0093] 如上所述,在一些情况下,对包含条形码的多核苷酸进行分区,以使得每个分区含有平均少于一个多核苷酸序列。因此,在一些情况下,分区的一部分将不含有多核苷酸,因此不能含有扩增的多核苷酸。因此,可能希望将包含多核苷酸的分区与不包含多核苷酸的分区分隔开。

[0094] 在一种情况下,使用能够鉴别包含多核苷酸的分区的基于流动的分选方法,将包含多核苷酸的分区与不包含多核苷酸的分区分隔开。在一些情况下,可以使用多核苷酸的存在指示物,以便区分包含多核苷酸的分区与不包含多核苷酸的分区。

[0095] 在一些情况下,可以使用核酸染料来鉴别包含多核苷酸的分区。示例性的染料包括嵌入染料、小沟结合剂、大沟结合剂、外部结合物和双嵌入剂。这类染料的具体实例包括 SYBR 绿、SYBR 蓝、DAPI、碘化丙锭、SYBR 金、溴化乙锭、吖啶、原黄素、吖啶橙、吖啶黄素、荧光香豆素 (fluorcoumanin)、玫瑰树碱、道诺霉素、氯喹、偏端霉素 D、色霉素、二胺乙基苯菲啶 (homidium)、光辉霉素、钆聚吡啶、氨基霉素、菲啶和吖啶、溴化乙锭、碘化丙锭、碘化己锭、二氢乙锭、乙锭同型二聚体 -1 和 -2、单叠氮乙锭、ACMA、吖啶、咪唑(例如, Hoechst 33258、Hoechst 33342、Hoechst 34580 和 DAPI)、吖啶橙(也能够嵌入)、7-AAD、放线菌素 D、LDS751、羟芪巴脒、SYTOX 蓝、SYTOX 绿、SYTOX 橙、POPO-1、POPO-3、YOYO-1、YOYO-3、TOTO-1、TOTO-3、JOJO-1、LOLO-1、BOBO-1、BOBO-3、PO-PRO-1、PO-PRO-3、BO-PRO-1、BO-PRO-3、TO-PRO-1、TO-PRO-3、TO-PRO-5、JO-PRO-1、LO-PRO-1、YO-PRO-1、

YO-PRO-3、PicoGreen、OliGreen、RiboGreen、SYBR 金、SYBR 绿 I、SYBR 绿 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)。

[0096] 在一些情况下,可以使用分离方法如颗粒的磁性分离或沉降。这样的方法可以包括,例如,将待扩增的多核苷酸、对应于所述待扩增的多核苷酸的引物和/或扩增的多核苷酸产物附接至珠子上的步骤。在一些情况下,待扩增的多核苷酸、对应于所述待扩增的多核苷酸的引物和/或多核苷酸产物向珠子上的附接可以经由对光不稳定的连接体例如 PC 氨基 C6 来进行。在使用对光不稳定的连接体的情况下,可以利用光来从珠子上释放连接的多核苷酸。该珠子可以是,例如,磁珠或胶乳珠子。珠子然后可允许通过例如磁性分选或沉降来分离。胶乳颗粒的沉降可以例如通过在比胶乳更稠的液体如甘油中离心来进行。在一些情况下,可以使用密度梯度离心。

[0097] 珠子可以具有均匀的大小或不均匀的大小。在一些情况下,珠子的直径可以是约 0.001 μm 、0.01 μm 、0.05 μm 、0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 或 1mm。珠子可以具有至少约 0.001 μm 、0.01 μm 、0.05 μm 、0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 或 1mm 的直径。在一些情况下,珠子可以具有小于约 0.001 μm 、0.01 μm 、0.05 μm 、0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 或 1mm 的直径。在一些情况下,珠子可以具有约 0.001 μm 至 1mm、0.01 μm 至 900 μm 、0.1 μm 至 600 μm 、100 μm 至 200 μm 、100 μm 至 300 μm 、100 μm 至 400 μm 、100 μm 至 500 μm 、100 μm 至 600 μm 、20 μm 至 50 μm 、150 μm 至 200 μm 、150 μm 至 300 μm 或 150 μm 至 400 μm 的直径。

[0098] 在一些情况下,可以利用包含多核苷酸的分区与不包含多核苷酸的分区之间的差异电荷,例如通过对分区进行电泳或介电电泳来分离包含多核苷酸的分区。

[0099] 在一些情况下,基于渗透压的差异,可以利用分区的选择性膨胀或收缩来鉴别包含多核苷酸的分区。在一些情况下,包含多核苷酸的分区可以通过流动分级分离、溶剂萃取、差别解链(例如,使用核酸探针)或冷冻来分离。

[0100] 包含多核苷酸的分区的分离提供了具有显著多样性的分区的多核苷酸条形码的文库,同时仅导致一次批量合成的费用。

IV. 包含条形码的衔接子的生成

[0101] 在本公开内容中描述的条形码可以具有多种结构。在一些情况下,本公开内容的条形码是衔接子的一部分。通常,“衔接子”是用来使条形码能够附接到靶多核苷酸上的结构。衔接子可以包含,例如,条形码,对于与靶多核苷酸的连接相容的多核苷酸序列,和功能序列如引物结合位点和固定区域。

[0102] 在一些情况下,衔接子是叉状衔接子。在图 1 中以图形方式示出了叉状衔接子的一个实例。参照图 1,在靶多核苷酸 105 的相对侧上示出了两个拷贝的叉状衔接子结构 106。每个叉状衔接子包含第一固定区域 101、第二固定区域 102、第一测序引物区域 103、第二测序引物区域 104 和彼此退火的一对部分互补的区域(在 103 和 104 内)。测序引物区域或

固定区域可以用来将条形码化的多核苷酸固定到例如珠子的表面上。测序引物区域可以用作例如测序引物的退火位点。在一些情况下,可以设计突出端以便能够与靶序列相匹配。在图 1 中,这对退火的多核苷酸 103 和 104 具有 3'-T 突出端,该突出端与靶多核苷酸 105 上的 3'-A 突出端相匹配。条形码可以包括在叉状衔接子的任何合适的部分中。在包含条形码的叉状衔接子与靶序列 105 附接后,测序引物区域 103 和 104 能够用来对靶多核苷酸进行测序。叉状衔接子结构的另一个实例包括在 Illumina™文库制品中使用的叉状衔接子和可从 New England Biolabs™获得的用于 Illumina 的 NEBNext® Multiplex Oligos。非叉状衔接子的实例包括 Merriman 等人, Electrophoresis, 33 (23) 3397-3417 (2012) 中描述的那些,其通过引用以其整体并入本文。

[0103] 图 2 示出了条形码区域在图 1 所示的叉状衔接子内的放置的三个示意性实例。在一个实例中,条形码 205 (BC1) 置于第一固定区域 201 内或第一固定区域 201 与第一测序引物区域 203 之间。在另一个实例中,条形码 206 (BC2) 置于第一测序引物区域 203 内或与之邻近。在又一个实例中,条形码 207 (BC3) 置于第二固定区域 202 内或第二固定区域 202 与第二测序引物区域 204 之间。尽管图 2 在靶序列的两端都示出了条形码,但这不是必需的,因为每个靶序列只有一个条形码对于某些应用是足够的。然而,如本公开内容的其他部分所述,也可以使用每个靶序列超过一个条形码。

[0104] 图 3 提供了连接至靶多核苷酸 (NNN) 的相对端的两个叉状衔接子的示例性序列 (SEQ ID NO:1 和 SEQ ID NO:22),并在序列水平上显示了每个叉状衔接子的条形码区域(粗体,核苷酸 30-37、71-77、81-87 和 122-129)。在图 3 中,核苷酸 1-29 代表第一叉状衔接子的固定区域,核苷酸 38-70 代表第一叉状衔接子的测序引物区域,核苷酸 78-80 (NNN) 代表任意长度的靶多核苷酸,核苷酸 88-120 代表第二叉状衔接子的测序引物区域,而核苷酸 129-153 代表第二叉状衔接子的固定区域。

V. 分区

a. 分区的一般特性

[0105] 如本公开内容通篇所述,本公开内容的某些方法、组合物、系统、装置和试剂盒可以利用某些物质向单独的分区内的细分(分区)。分区可以是,例如,孔、微孔、孔洞、小滴(例如,乳液中的小滴)、乳液的连续相、试管、斑点、胶囊、珠子的表面,或任何其他用于隔离样品或物质的一个级分的合适的容器。分区可以用来容纳用于进一步处理的物质。例如,如果物质是多核苷酸分析物,则进一步的处理可以包括切割、连接和/或用作试剂的物质进行条形码化。可以用任何数目的装置、系统或容器来容纳、支持或包含分区。在一些情况下,可以用微孔板来容纳、支持或包含分区。可以使用任何合适的微孔板,例如,具有 96、384 或 1536 个孔的微孔板。

[0106] 每个分区还可以含有任何其他合适的分区或包容在任何其他合适的分区中。例如,孔、微孔、孔洞、珠子的表面或管可包含小滴(例如,乳液中的小滴)、乳液中的连续相、斑点、胶囊或任何其他合适的分区。小滴可以包含胶囊、珠子或另一小滴。胶囊可以包含小滴、珠子或另一胶囊。这些描述仅是说明性的,还想到所有合适的组合和复数。例如,任何合适的分区可以包含多个相同或不同的分区。在一个实例中,孔或微孔包含多个小滴和多个胶囊。在另一个实例中,胶囊包含多个胶囊和多个小滴。想像到分区的所有组合。表 1 显示了可以彼此组合的分区的非限制性实例。

表 1. 可以彼此组合的分区的实例。

	孔	斑点	小滴	胶囊
孔	• 孔内的孔	• 孔内的斑点	• 孔内的小滴	• 孔内的胶囊
斑点	• 孔内的斑点	• 斑点内的斑点	• 斑点内的小滴	• 斑点内的胶囊
小滴	• 孔内的小滴	• 斑点内的小滴	• 小滴内的小滴	• 胶囊内的小滴 • 小滴内的胶囊
胶囊	• 孔内的胶囊	• 斑点内的胶囊 • 胶囊内的斑点	• 小滴内的胶囊 • 胶囊内的小滴	• 胶囊内的胶囊
珠子的表面	• 孔内的珠子	• 珠子上的斑点 • 斑点内的珠子	• 小滴内的珠子	• 胶囊内的珠子

[0107] 本文描述的任何分区可以包含多个分区。例如,分区可以包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000 或 50000 个分区。分区可以包含至少 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000 或 50000 个分区。在一些情况下,分区可以包含少于 3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000 或 50000 个分区。在一些情况下,每个分区可以包含 2-50、2-20、2-10 或 2-5 个分区。

[0108] 分区可以包含任何合适的物质或物质的混合物。例如,在一些情况下,分区可以包含试剂、分析物、样品、细胞及其组合。包含其他分区的分区可以在相同的分区中包含某些物质以及在不同的分区中包含某些物质。根据特定过程的需要,物质可以分布在任何合适的分区之间。例如,表 1 中的任何分区可以含有至少一种第一物质,且表 1 中的任何分区可以含有至少一种第二物质。在一些情况下,第一物质可以是试剂,而第二物质可以是分析物。

[0109] 在一些情况下,物质是从细胞中分离的多核苷酸。例如,在一些情况下,利用任何合适的方法(例如可商购的试剂盒)从细胞中分离多核苷酸(例如,基因组 DNA、RNA 等)。可以对多核苷酸进行量化。然后将量化的多核苷酸分配到如本文所述的多个分区中。根据量化和测定的需要,多核苷酸的分区可以在预定的覆盖量下进行。在一些情况下,所有或大多数(例如,至少 50%、60%、70%、80%、90%或 95%)分区不包含重叠的多核苷酸,使得在多个分区间形成非重叠片段的单独的混合物。分区的多核苷酸然后可以按照本领域已知的或本公开内容描述的任何合适的方法来处理。例如,可以对分区的多核苷酸进行片段化、扩增、条形码化等。

[0110] 物质可以使用多种方法进行分区。例如,可以将物质稀释并在多个分区间进行分配。可以对包含物质的介质进行终末稀释,以使得分区的数目超过物质的数目。也可以在

形成乳液或胶囊之前,或在将物质点印至基底上之前使用稀释。物质的数目与分区的数目之比可以是约 0.1、0.5、1、2、4、8、10、20、50、100 或 1000。物质的数目与分区的数目之比可以是至少约 0.1、0.5、1、2、4、8、10、20、50、100 或 1000。物质的数目与分区的数目之比可以小于约 0.1、0.5、1、2、4、8、10、20、50、100 或 1000。物质的数目与分区的数目之比可以为约 0.1-10、0.5-10、1-10、2-10、10-100、100-1000 或更大。

[0111] 也可以使用压电小滴生成(例如,Bransky 等人,Lab on a Chip, 2009, 9, 516-520)或表面声波(例如,Demirci 和 Montesano, Lab on a Chip, 2007, 7, 1139-1145)进行分区。

[0112] 使用的分区的数目可以根据应用而不同。例如,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。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、100,000、200000、300000、400000、500000、600000、700000、800000、900000、1,000,000、2000000、3000000、4000000、5000000、10000000、20000000 或更多。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、100,000、200000、300000、400000、500000、600000、700000、800000、900000、1,000,000、2000000、3000000、4000000、5000000、10000000、20000000。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。

[0113] 被分区的不同条形码或不同组条形码的数目可以根据例如待分区的特定条形码和/或应用而不同。不同组条形码可以是,例如,相同条形码在各组中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。例如,可以对约 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、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000 个或更多个不同的条形码或不同组的条形码进行分区。在一些实例中,可以对至少约 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、30000000、40000000、50000000、60000000、70000000、80000000、90000000、100000000 个或更多个不同的条形码或不同组的条形码进行分区。在一些实例中,可以对约 1-5、5-10、10-50、50-100、100-1000、1000-10000、10000-100000、100000-1000000、10000-1000000、10000-10000000 或 10000-100000000 个条形码进行分

区。

[0114] 条形码可以以特定密度进行分区。例如,可以对条形码进行分区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。可以对条形码进行分区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。可以对条形码进行分区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 个条形码 / 分区。可以对条形码进行分区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、10000-1000000、10000-10000000 或 10000-100000000 个条形码 / 分区。

[0115] 可以对条形码进行分区以使得相同的条形码以特定密度分区。例如,可以对相同的条形码进行分区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。可以对条形码进行分区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 个或更多个相同的条形码 / 分区。可以对条形码进行分区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 个相同的条形码 / 分区。可以对条形码进行分区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、10000-1000000、10000-10000000 或 10000-100000000 个相同的条形码 / 分区。

[0116] 可以对条形码进行分区以使得不同的条形码以特定密度分区。例如,可以对不同的条形码进行分区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 1000000000 个不同的条形码 / 分区。可以对条形码进行分区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 个或更多个不同的条形码 / 分区。可以对条形码进行分区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。可以对条形码进行分区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、100000-1000000、10000-1000000、10000-10000000 或 10000-100000000 个不同的条形码 / 分区。

[0117] 用于对条形码进行分区的分区的数目可以变化,例如,取决于应用和 / 或待分区的不同条形码的数目。例如,用于对条形码进行分区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。用于对条形码进行分区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。用于对条形码进行分区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 或 20000000。用于对条形码进行分区的分区的数目可以是约 5-10000000、5-50000000、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。

[0118] 如上所述,可以对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)进行分区,以使得每个分区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。当对不同组的相同条形码进行分区时,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可以在约 100,000 个或更多个不同的分区中进行分区,使得每个分区包含不同组的相同条形码。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。不同条形码或不同条形码组的任何合适的数目(包括本文其他部分描述的待分区的不同条形码或不同条形码组的数目)、每个分区中条形码的数目(包括本文其他部分描述的每个分区中条形码的数目)和分区的数目(包括本文其他部分描述的分区的数目)可以组合起来以生成被分区的条形码的多样性文库,其中每个分区具有大数目的条形码。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。

[0119] 分区的容积可以根据应用而不同。例如,本公开内容描述的任何分区(例如孔、斑点、小滴(例如,在乳液中)和胶囊)的容积可以是约 1000 μ l、900 μ l、800 μ l、700 μ l、600 μ l、500 μ l、400 μ l、300 μ l、200 μ l、100 μ l、50 μ l、25 μ l、10 μ l、5 μ l、1 μ l、900nL、800nL、700nL、600nL、500nL、400nL、300nL、200nL、100nL、50nL、25nL、10nL、5nL、2.5nL、1nL、900pL、800pL、700pL、600pL、500pL、400pL、300pL、200pL、100pL、50pL、25pL、10pL、5pL、1pL、900fL、800fL、700fL、600fL、500fL、400fL、300fL、200fL、100fL、50fL、25fL、10fL、5fL、1fL 或 0.5fL。分区的容积可以是至少约 1000 μ l、900 μ l、800 μ l、700 μ l、600 μ l、500 μ l、400 μ l、300 μ l、200 μ l、100 μ l、50 μ l、25 μ l、10 μ l、5 μ l、1 μ l、900nL、800nL、700nL、600nL、500nL、400nL、300nL、200nL、100nL、50nL、25nL、10nL、5nL、5nL、2.5nL、1nL、900pL、800pL、700pL、600pL、500pL、400pL、300pL、200pL、100pL、50pL、25pL、10pL、5pL、1pL、900fL、800fL、700fL、600fL、500fL、400fL、300fL、200fL、100fL、50fL、25fL、10fL、5fL、1fL 或 0.5fL。分区的容积可以小于约 1000 μ l、900 μ l、800 μ l、700 μ l、600 μ l、500 μ l、400 μ l、300 μ l、200 μ l、100 μ l、50 μ l、25 μ l、10 μ l、5 μ l、1 μ l、900nL、800nL、700nL、600nL、500nL、400nL、300nL、200nL、100nL、50nL、25nL、10nL、5nL、5nL、2.5nL、1nL、900pL、800pL、700pL、600pL、500pL、400pL、300pL、200pL、100pL、50pL、25pL、10pL、5pL、1pL、900fL、800fL、700fL、600fL、500fL、400fL、300fL、200fL、100fL、50fL、25fL、10fL、5fL、1fL 或 0.5fL。分区的容积可以是约 0.5fL–5pL、10pL–10nL、10nL–10 μ l、10 μ l–100 μ l 或 100 μ l–1mL。

[0120] 在不同分区中的流体体积可存在可变性。更具体地,在一组分区之间,不同分区的容积可以变化至少(或至多)+或-1%、2%、3%、4%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、200%、300%、400%、500%或1000%。例如,孔(或其他分区)可以包含为第二孔(或其他分区)内的流体体积的至多80%的流体体积。

[0121] 特定物质也可以靶向至特定分区。例如,在一些情况下,捕获试剂(例如,寡核苷酸探针)可以固定在或置于分区内以捕获特定物质(例如,多核苷酸)。例如,捕获寡核苷酸可以固定在珠子的表面上,以捕获包含具有互补序列的寡核苷酸的物质。

[0122] 物质也可以以特定密度进行分区。例如,可以对物质进行分区,以使得每个分区含有约1、5、10、50、100、1000、10000、100000或1000000个物质/分区。可以对物质进行分区,以使得每个分区含有至少约1、5、10、50、100、1000、10000、100000、1000000个或更多个物质/分区。可以对物质进行分区,以使得每个分区含有少于约1、5、10、50、100、1000、10000、100000或1000000个物质/分区。可以对物质进行分区,以使得每个分区含有约1–5、5–10、10–50、50–100、100–1000、1000–10000、10000–100000或100000–1000000个物质/分区。

[0123] 可以对物质进行分区,以使得至少一个分区包含在该分区内独特的物质。这可能适用于约1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%或更多的分区。这可能适用于至少约1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%或更多的分区。这可能适用于少于约1%、5%、10%、20%、30%、40%、50%、60%、70%、80%或90%的分区。

b. 孔作为分区

[0124] 在一些情况下,使用孔作为分区。孔可以是微孔。孔可包含介质,该介质包含一个或多个物质。物质可以以多种配置包含在孔内。在一个实例中,将物质直接分配到孔中。直

接分配到孔中的物质可以覆盖有层,该层例如是可溶解的、可熔化的或可渗透的。该层可以是,例如,油、蜡、膜等。该层可以在将另外的物质引入孔中之前或之后溶解或熔化。可以在任何时间点,例如在添加任何物质之后,用密封层对孔进行密封。

[0125] 在一个实例中,将用于样品处理的试剂直接分配到孔中,并用可溶解的、可熔化的或可渗透的层覆盖。将包含待处理的分析物的样品引入到层的顶部。该层溶解或熔化,或者分析物(或试剂)扩散通过该层。将孔密封,并在用于分析物处理的合适的条件下温育。随后可回收处理的分析物。

[0126] 在一些情况下,孔包含其他分区。孔可包含任何合适的分区,包括,例如,另外的孔、斑点、小滴(例如,乳液中的小滴)、胶囊、珠子等。每个分区可以作为单个分区或多个分区存在,并且每个分区可包含相同的物质或不同的物质。

[0127] 在一个实例中,孔包含胶囊,该胶囊包含用于样品处理的试剂。可使用液体介质将胶囊加载至孔中,或在不使用液体介质(例如,基本上干燥)的情况下将胶囊加载至孔中。如本公开内容的其他地方所述的,胶囊可含有一个或多个胶囊,或其他分区。可将包含待处理的分析物的样品引入到孔中。可将孔密封,并可施加刺激以使胶囊的内容物释放至孔中,从而导致试剂与待处理的分析物之间的接触。可将孔在用于处理分析物的合适的条件下温育。随后可回收处理的分析物。尽管该实例描述了其中试剂在胶囊中并且分析物在孔中的实施方案,但相反的配置——即,试剂在孔中并且分析物在胶囊中——也是可能的。

[0128] 在另一个实例中,孔包含乳液,且乳液的小滴包含含有用于样品处理的试剂的胶囊。包含待处理的分析物的样品包含在乳液的小滴内。将孔密封,并施加刺激以使胶囊的内容物释放至小滴中,从而导致试剂与待处理的分析物之间的接触。将孔在用于处理分析物的合适的条件下温育。随后可回收处理的分析物。尽管该实例描述了其中试剂在胶囊中并且分析物在小滴中的实施方案,但相反的配置——即,试剂在小滴中并且分析物在胶囊中——也是可能的。

[0129] 可将孔布置成阵列,例如,微孔阵列。根据单个孔的尺寸和基底的大小,孔阵列可包含一定范围的孔密度。在一些情况下,孔密度可以为 10 个孔/cm²、50 个孔/cm²、100 个孔/cm²、500 个孔/cm²、1000 个孔/cm²、5000 个孔/cm²、10000 个孔/cm²、50000 个孔/cm²或 100000 个孔/cm²。在一些情况下,孔密度可以为至少 10 个孔/cm²、50 个孔/cm²、100 个孔/cm²、500 个孔/cm²、1000 个孔/cm²、5000 个孔/cm²、10000 个孔/cm²、50000 个孔/cm²或 100000 个孔/cm²。在一些情况下,孔密度可以小于 10 个孔/cm²、50 个孔/cm²、100 个孔/cm²、500 个孔/cm²、1000 个孔/cm²、5000 个孔/cm²、10000 个孔/cm²、50000 个孔/cm²或 100000 个孔/cm²。c. 斑点作为分区

[0130] 在一些情况下,使用斑点作为分区。斑点可以例如通过将物质分配到表面上而制得。物质可以以多种配置包含在斑点内。在一个实例中,通过将物质包含在用于形成斑点的介质中而将物质直接分配到斑点中。直接分配到斑点上的物质可以覆盖有层,该层例如是可溶解的、可熔化的或可渗透的。该层可以是,例如,油、蜡、膜等。该层可以在将另外的物质引入斑点之前或之后溶解或熔化。可以在任何时间点,例如在添加任何物质之后,通过覆盖物来密封斑点。

[0131] 在一个实例中,将用于样品处理的试剂直接分配到斑点,例如载玻片上,并用可溶解的、可熔化的或可渗透的层覆盖。将包含待处理的分析物的样品引入到层的顶部。该

层溶解或熔化,或者分析物(或试剂)扩散通过该层。将斑点密封,并在用于分析物处理的合适的条件下温育。随后可回收处理的分析物。

[0132] 如本公开内容的其他地方(例如,表1)所述,也可将斑点布置在孔内。在一些情况下,可将多个斑点布置在孔内,使得每个斑点的内容物不会混合。这样的配置可能是有用的,例如,当期望阻止物质彼此接触时。在一些情况下,孔可包含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个或更多个斑点。在一些情况下,孔可包含至少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个或更多个斑点。在一些情况下,孔可包含少于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或30个斑点。在一些情况下,孔可包含2-4、2-6、2-8、4-6、4-8、5-10或4-12个斑点。在将物质(例如,包含分析物的介质)加入到孔中时,斑点中的物质可以混合。此外,使用分开的斑点来容纳不同的物质(或物质的组合)也可用于防止用来将斑点置于孔内的装置的交叉污染。

[0133] 在一些情况下,斑点包含其他分区。斑点可包含任何合适的分区,包括,例如,另外的斑点、小滴(例如,乳液中的小滴)、胶囊、珠子等。每个分区可以作为单个分区或多个分区存在,并且每个分区可包含相同的物质或不同的物质。

[0134] 在一个实例中,斑点包含胶囊,该胶囊包含用于样品处理的试剂。如本公开内容的其他地方所述,胶囊可含有一个或多个胶囊,或其他分区。将包含待处理的分析物的样品引入到斑点中。将斑点密封,并施加刺激以使胶囊的内容物释放至斑点中,从而导致试剂与待处理的分析物之间的接触。将斑点在用于处理分析物的合适的条件下温育。随后可回收处理的分析物。尽管该实例描述了其中试剂在胶囊中并且分析物在斑点中的实施方案,但相反的配置——即,试剂在斑点中并且分析物在胶囊中——也是可能的。

[0135] 在另一个实例中,斑点包含乳液,且乳液的小滴包含含有用于样品处理的试剂的胶囊。包含待处理的分析物的样品包含在乳液的小滴内。将斑点密封,并施加刺激以使胶囊的内容物释放至小滴中,从而导致试剂与待处理的分析物之间的接触。将斑点在用于处理分析物的合适的条件下温育。随后可回收处理的分析物。尽管该实例描述了其中试剂在胶囊中并且分析物在小滴中的实施方案,但相反的配置——即,试剂在小滴中并且分析物在胶囊中——也是可能的。

[0136] 斑点可以具有均匀的大小或不均匀的大小。在一些情况下,斑点的直径可以为约0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 、1mm、2mm、5mm或1cm。斑点可以具有至少约0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 、1mm、1mm、2mm、5mm或1cm的直径。在一些情况下,斑点可以具有小于约0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 、1mm、1mm、2mm、5mm或1cm的直径。在一些情况下,斑点可以具有约0.1 μm 至1cm、100 μm 至1mm、100 μm 至500 μm 、100 μm 至600 μm 、150 μm 至300 μm 、或150 μm 至400 μm 的直径。

[0137] 可以将斑点布置成阵列,例如斑点阵列。根据单个斑点的尺寸和基底的大小,斑点阵列可包含一定范围的斑点密度。在一些情况下,斑点密度可以为10个斑点/ cm^2 、50个斑点/ cm^2 、100个斑点/ cm^2 、500个斑点/ cm^2 、1000个斑点/ cm^2 、5000个斑点/ cm^2 、10000个斑点

/cm²、50000 个斑点 /cm²或 100000 个斑点 /cm²。在一些情况下,斑点密度可以为至少 10 个斑点 /cm²、50 个斑点 /cm²、100 个斑点 /cm²、500 个斑点 /cm²、1000 个斑点 /cm²、5000 个斑点 /cm²、10000 个斑点 /cm²、50000 个斑点 /cm²或 100000 个斑点 /cm²。在一些情况下,斑点密度可以小于 10 个斑点 /cm²、50 个斑点 /cm²、100 个斑点 /cm²、500 个斑点 /cm²、1000 个斑点 /cm²、5000 个斑点 /cm²、10000 个斑点 /cm²、50000 个斑点 /cm²或 100000 个斑点 /cm²。d. 乳液作为分区

[0138] 在一些情况下,使用乳液中的小滴作为分区。例如,可通过任何合适的方法,包括本领域中已知的方法制备乳液。(参见,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,可使用碳氟化合物包水乳液。这些乳液可以掺入含氟表面活性剂,诸如含聚乙二醇 (PEG) 的寡聚全氟聚醚 (PFPE)。(Holtz 等人, *Lab on a Chip*, 2008, 8(10):1632-1639)。在一些情况下,可以在微流体流动聚焦装置中形成单分散乳液。(Garstecki 等人, *Applied Physics Letters*, 2004, 85(13):2649-2651)。

[0139] 物质可容纳在乳液中的小滴内,该乳液包含,例如,形成小滴的第一相(例如,油或水)和第二(连续)相(例如,水或油)。乳液可以是单乳液,例如,油包水或水包油乳液。乳液可以是双乳液,例如水包油包水或油包水包油乳液。更高阶的乳液也是可能的。该乳液可容纳在任何合适的容器,包括本公开内容中描述的任何合适的分区中。

[0140] 在一些情况下,乳液中的小滴包含其他分区。乳液中的小滴可包含任何合适的分区,包括,例如,另外的小滴(例如,乳液中的小滴)、胶囊、珠子等。每个分区可作为单个分区或多个分区存在,并且每个分区可包含相同的物质或不同的物质。

[0141] 在一个实例中,乳液中的小滴包含胶囊,该胶囊包含用于样品处理的试剂。如本公开内容的其他地方所述的,胶囊可含有一个或多个胶囊,或其他分区。包含待处理的分析物的样品包含在小滴内。施加刺激以使胶囊的内容物释放至小滴中,从而导致试剂与待处理的分析物之间的接触。将小滴在用于处理分析物的合适的条件下温育。随后可回收处理的分析物。尽管该实例描述了其中试剂在胶囊中并且分析物在小滴中的实施方案,但相反的配置——即,试剂在小滴中并且分析物在胶囊中——也是可能的。

[0142] 乳液中的小滴可以具有均匀的大小或不均匀的大小。在一些情况下,乳液中的小滴的直径可以为约 0.001 μm、0.01 μm、0.05 μm、0.1 μm、0.5 μm、1 μm、5 μm、10 μm、50 μm、100 μm、150 μm、200 μm、300 μm、400 μm、500 μm、600 μm、700 μm、800 μm、900 μm 或 1mm。小滴可以具有至少约 0.001 μm、0.01 μm、0.05 μm、0.1 μm、0.5 μm、1 μm、5 μm、10 μm、50 μm、100 μm、150 μm、200 μm、300 μm、400 μm、500 μm、600 μm、700 μm、800 μm、900 μm 或 1mm 的直径。在一些情况下,小滴可以具有小于约 0.001 μm、0.01 μm、0.05 μm、0.1 μm、0.5 μm、1 μm、5 μm、10 μm、50 μm、100 μm、150 μm、200 μm、300 μm、400 μm、500 μm、600 μm、700 μm、800 μm、900 μm 或 1mm 的直径。在一些情况下,小滴可以具有约 0.001 μm 至 1mm、0.01 μm 至 900 μm、0.1 μm 至 600 μm、100 μm 至 200 μm、100 μm 至 300 μm、100 μm 至 400 μm、100 μm 至 500 μm、100 μm 至 600 μm、150 μm 至 200 μm、150 μm 至 300 μm、或 150 μm 至 400 μm 的直径。

[0143] 乳液中的小滴也可具有特定的密度。在一些情况下,小滴不如水性流体(例如,水)稠密;在一些情况下,小滴比水性流体稠密。在一些情况下,小滴不如非水性流体

(例如,油)稠密;在一些情况下,小滴比非水性流体稠密。小滴可以具有约 $0.05\text{g}/\text{cm}^3$ 、 $0.1\text{g}/\text{cm}^3$ 、 $0.2\text{g}/\text{cm}^3$ 、 $0.3\text{g}/\text{cm}^3$ 、 $0.4\text{g}/\text{cm}^3$ 、 $0.5\text{g}/\text{cm}^3$ 、 $0.6\text{g}/\text{cm}^3$ 、 $0.7\text{g}/\text{cm}^3$ 、 $0.8\text{g}/\text{cm}^3$ 、 $0.81\text{g}/\text{cm}^3$ 、 $0.82\text{g}/\text{cm}^3$ 、 $0.83\text{g}/\text{cm}^3$ 、 $0.84\text{g}/\text{cm}^3$ 、 $0.85\text{g}/\text{cm}^3$ 、 $0.86\text{g}/\text{cm}^3$ 、 $0.87\text{g}/\text{cm}^3$ 、 $0.88\text{g}/\text{cm}^3$ 、 $0.89\text{g}/\text{cm}^3$ 、 $0.90\text{g}/\text{cm}^3$ 、 $0.91\text{g}/\text{cm}^3$ 、 $0.92\text{g}/\text{cm}^3$ 、 $0.93\text{g}/\text{cm}^3$ 、 $0.94\text{g}/\text{cm}^3$ 、 $0.95\text{g}/\text{cm}^3$ 、 $0.96\text{g}/\text{cm}^3$ 、 $0.97\text{g}/\text{cm}^3$ 、 $0.98\text{g}/\text{cm}^3$ 、 $0.99\text{g}/\text{cm}^3$ 、 $1.00\text{g}/\text{cm}^3$ 、 $1.05\text{g}/\text{cm}^3$ 、 $1.1\text{g}/\text{cm}^3$ 、 $1.2\text{g}/\text{cm}^3$ 、 $1.3\text{g}/\text{cm}^3$ 、 $1.4\text{g}/\text{cm}^3$ 、 $1.5\text{g}/\text{cm}^3$ 、 $1.6\text{g}/\text{cm}^3$ 、 $1.7\text{g}/\text{cm}^3$ 、 $1.8\text{g}/\text{cm}^3$ 、 $1.9\text{g}/\text{cm}^3$ 、 $2.0\text{g}/\text{cm}^3$ 、 $2.1\text{g}/\text{cm}^3$ 、 $2.2\text{g}/\text{cm}^3$ 、 $2.3\text{g}/\text{cm}^3$ 、 $2.4\text{g}/\text{cm}^3$ 或 $2.5\text{g}/\text{cm}^3$ 的密度。小滴可以具有至少约 $0.05\text{g}/\text{cm}^3$ 、 $0.1\text{g}/\text{cm}^3$ 、 $0.2\text{g}/\text{cm}^3$ 、 $0.3\text{g}/\text{cm}^3$ 、 $0.4\text{g}/\text{cm}^3$ 、 $0.5\text{g}/\text{cm}^3$ 、 $0.6\text{g}/\text{cm}^3$ 、 $0.7\text{g}/\text{cm}^3$ 、 $0.8\text{g}/\text{cm}^3$ 、 $0.81\text{g}/\text{cm}^3$ 、 $0.82\text{g}/\text{cm}^3$ 、 $0.83\text{g}/\text{cm}^3$ 、 $0.84\text{g}/\text{cm}^3$ 、 $0.85\text{g}/\text{cm}^3$ 、 $0.86\text{g}/\text{cm}^3$ 、 $0.87\text{g}/\text{cm}^3$ 、 $0.88\text{g}/\text{cm}^3$ 、 $0.89\text{g}/\text{cm}^3$ 、 $0.90\text{g}/\text{cm}^3$ 、 $0.91\text{g}/\text{cm}^3$ 、 $0.92\text{g}/\text{cm}^3$ 、 $0.93\text{g}/\text{cm}^3$ 、 $0.94\text{g}/\text{cm}^3$ 、 $0.95\text{g}/\text{cm}^3$ 、 $0.96\text{g}/\text{cm}^3$ 、 $0.97\text{g}/\text{cm}^3$ 、 $0.98\text{g}/\text{cm}^3$ 、 $0.99\text{g}/\text{cm}^3$ 、 $1.00\text{g}/\text{cm}^3$ 、 $1.05\text{g}/\text{cm}^3$ 、 $1.1\text{g}/\text{cm}^3$ 、 $1.2\text{g}/\text{cm}^3$ 、 $1.3\text{g}/\text{cm}^3$ 、 $1.4\text{g}/\text{cm}^3$ 、 $1.5\text{g}/\text{cm}^3$ 、 $1.6\text{g}/\text{cm}^3$ 、 $1.7\text{g}/\text{cm}^3$ 、 $1.8\text{g}/\text{cm}^3$ 、 $1.9\text{g}/\text{cm}^3$ 、 $2.0\text{g}/\text{cm}^3$ 、 $2.1\text{g}/\text{cm}^3$ 、 $2.2\text{g}/\text{cm}^3$ 、 $2.3\text{g}/\text{cm}^3$ 、 $2.4\text{g}/\text{cm}^3$ 或 $2.5\text{g}/\text{cm}^3$ 的密度。在其他情况下,小滴密度可以为最大约 $0.7\text{g}/\text{cm}^3$ 、 $0.8\text{g}/\text{cm}^3$ 、 $0.81\text{g}/\text{cm}^3$ 、 $0.82\text{g}/\text{cm}^3$ 、 $0.83\text{g}/\text{cm}^3$ 、 $0.84\text{g}/\text{cm}^3$ 、 $0.85\text{g}/\text{cm}^3$ 、 $0.86\text{g}/\text{cm}^3$ 、 $0.87\text{g}/\text{cm}^3$ 、 $0.88\text{g}/\text{cm}^3$ 、 $0.89\text{g}/\text{cm}^3$ 、 $0.90\text{g}/\text{cm}^3$ 、 $0.91\text{g}/\text{cm}^3$ 、 $0.92\text{g}/\text{cm}^3$ 、 $0.93\text{g}/\text{cm}^3$ 、 $0.94\text{g}/\text{cm}^3$ 、 $0.95\text{g}/\text{cm}^3$ 、 $0.96\text{g}/\text{cm}^3$ 、 $0.97\text{g}/\text{cm}^3$ 、 $0.98\text{g}/\text{cm}^3$ 、 $0.99\text{g}/\text{cm}^3$ 、 $1.00\text{g}/\text{cm}^3$ 、 $1.05\text{g}/\text{cm}^3$ 、 $1.1\text{g}/\text{cm}^3$ 、 $1.2\text{g}/\text{cm}^3$ 、 $1.3\text{g}/\text{cm}^3$ 、 $1.4\text{g}/\text{cm}^3$ 、 $1.5\text{g}/\text{cm}^3$ 、 $1.6\text{g}/\text{cm}^3$ 、 $1.7\text{g}/\text{cm}^3$ 、 $1.8\text{g}/\text{cm}^3$ 、 $1.9\text{g}/\text{cm}^3$ 、 $2.0\text{g}/\text{cm}^3$ 、 $2.1\text{g}/\text{cm}^3$ 、 $2.2\text{g}/\text{cm}^3$ 、 $2.3\text{g}/\text{cm}^3$ 、 $2.4\text{g}/\text{cm}^3$ 或 $2.5\text{g}/\text{cm}^3$ 。这些密度可反映任何特定流体(例如,水性流体、水、油等)中胶囊的密度。

e. 胶囊作为分区

[0144] 在一些情况下,使用胶囊作为分区。可通过任何合适的方法制备胶囊,包括本领域中已知的方法,包括乳化聚合(Weitz 等人(美国公开号 2012/0211084))、采用聚合电解质的逐层装配、凝聚、内相分离和流动聚焦。任何合适的物质可容纳在胶囊内。该胶囊可容纳在任何合适的容器中,包括本公开内容中描述的任何合适的分区中。

[0145] 在一些情况下,胶囊包含其他分区。胶囊可包含任何合适的分区,包括,例如,另外的胶囊、乳液中的小滴、珠子等。每个分区可以作为单个分区或多个分区存在,并且每个分区可包含相同的物质或不同的物质。

[0146] 在一个实例中,外层胶囊包含内层胶囊。内层胶囊包含用于样品处理的试剂。将分析物包封在内层胶囊与外层胶囊之间的介质中。施加刺激以使内层胶囊的内容物释放至外层胶囊中,从而导致试剂与待处理的分析物之间的接触。将外层胶囊在用于处理分析物的合适的条件下温育。随后可回收处理的分析物。尽管该实例描述了其中试剂在内层胶囊中并且分析物在内层胶囊与外层胶囊之间的介质中的实施方案,但相反的配置——即,试剂在内层胶囊与外层胶囊之间的介质中并且分析物在内层胶囊中——也是可能的。

[0147] 胶囊可预先形成并通过注射以试剂来填充。例如,在 Abate 等人(Proc. Natl. Acad. Sci. U. S. A., 2010, 107(45), 19163-19166)和 Weitz 等人(美国公开号 2012/0132288)中描述的皮量注射(picoinjection)方法可用于将试剂引入到本文所述的胶囊的内部。通常,皮量注射将在胶囊壳硬化之前进行,例如通过在胶囊壳形成之前,将物质注射入胶囊前体如乳液的小滴的内部。

[0148] 胶囊可以具有均匀的大小或不均匀的大小。在一些情况下,胶囊的直径可以为约 $0.001\text{ }\mu\text{m}$ 、 $0.01\text{ }\mu\text{m}$ 、 $0.05\text{ }\mu\text{m}$ 、 $0.1\text{ }\mu\text{m}$ 、 $0.5\text{ }\mu\text{m}$ 、 $1\text{ }\mu\text{m}$ 、 $5\text{ }\mu\text{m}$ 、 $10\text{ }\mu\text{m}$ 、 $50\text{ }\mu\text{m}$ 、 $100\text{ }\mu\text{m}$ 、 $150\text{ }\mu\text{m}$ 、

200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 或 1mm。胶囊可以具有至少约 0.001 μm 、0.01 μm 、0.05 μm 、0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 或 1mm 的直径。在一些情况下,胶囊可以具有小于约 0.001 μm 、0.01 μm 、0.05 μm 、0.1 μm 、0.5 μm 、1 μm 、5 μm 、10 μm 、50 μm 、100 μm 、150 μm 、200 μm 、300 μm 、400 μm 、500 μm 、600 μm 、700 μm 、800 μm 、900 μm 或 1mm 的直径。在一些情况下,胶囊可以具有约 0.001 μm 至 1mm、0.01 μm 至 900 μm 、0.1 μm 至 600 μm 、100 μm 至 200 μm 、100 μm 至 300 μm 、100 μm 至 400 μm 、100 μm 至 500 μm 、100 μm 至 600 μm 、150 μm 至 200 μm 、150 μm 至 300 μm 、或 150 μm 至 400 μm 的直径。

[0149] 胶囊还可具有特定的密度。在一些情况下,胶囊不如水性流体(例如,水)稠密;在一些情况下,胶囊比水性流体稠密。在一些情况下,胶囊不如非水性流体(例如,油)稠密;在一些情况下,胶囊比非水性流体稠密。胶囊可以具有约 0.05g/cm³、0.1g/cm³、0.2g/cm³、0.3g/cm³、0.4g/cm³、0.5g/cm³、0.6g/cm³、0.7g/cm³、0.8g/cm³、0.81g/cm³、0.82g/cm³、0.83g/cm³、0.84g/cm³、0.85g/cm³、0.86g/cm³、0.87g/cm³、0.88g/cm³、0.89g/cm³、0.90g/cm³、0.91g/cm³、0.92g/cm³、0.93g/cm³、0.94g/cm³、0.95g/cm³、0.96g/cm³、0.97g/cm³、0.98g/cm³、0.99g/cm³、1.00g/cm³、1.05g/cm³、1.1g/cm³、1.2g/cm³、1.3g/cm³、1.4g/cm³、1.5g/cm³、1.6g/cm³、1.7g/cm³、1.8g/cm³、1.9g/cm³、2.0g/cm³、2.1g/cm³、2.2g/cm³、2.3g/cm³、2.4g/cm³ 或 2.5g/cm³ 的密度。胶囊可以具有至少约 0.05g/cm³、0.1g/cm³、0.2g/cm³、0.3g/cm³、0.4g/cm³、0.5g/cm³、0.6g/cm³、0.7g/cm³、0.8g/cm³、0.81g/cm³、0.82g/cm³、0.83g/cm³、0.84g/cm³、0.85g/cm³、0.86g/cm³、0.87g/cm³、0.88g/cm³、0.89g/cm³、0.90g/cm³、0.91g/cm³、0.92g/cm³、0.93g/cm³、0.94g/cm³、0.95g/cm³、0.96g/cm³、0.97g/cm³、0.98g/cm³、0.99g/cm³、1.00g/cm³、1.05g/cm³、1.1g/cm³、1.2g/cm³、1.3g/cm³、1.4g/cm³、1.5g/cm³、1.6g/cm³、1.7g/cm³、1.8g/cm³、1.9g/cm³、2.0g/cm³、2.1g/cm³、2.2g/cm³、2.3g/cm³、2.4g/cm³ 或 2.5g/cm³ 的密度。在其他情况下,胶囊密度可以为最大约 0.7g/cm³、0.8g/cm³、0.81g/cm³、0.82g/cm³、0.83g/cm³、0.84g/cm³、0.85g/cm³、0.86g/cm³、0.87g/cm³、0.88g/cm³、0.89g/cm³、0.90g/cm³、0.91g/cm³、0.92g/cm³、0.93g/cm³、0.94g/cm³、0.95g/cm³、0.96g/cm³、0.97g/cm³、0.98g/cm³、0.99g/cm³、1.00g/cm³、1.05g/cm³、1.1g/cm³、1.2g/cm³、1.3g/cm³、1.4g/cm³、1.5g/cm³、1.6g/cm³、1.7g/cm³、1.8g/cm³、1.9g/cm³、2.0g/cm³、2.1g/cm³、2.2g/cm³、2.3g/cm³、2.4g/cm³ 或 2.5g/cm³。这些密度可反映任何特定流体(例如,水性流体、水、油等)中胶囊的密度。

1. 通过流动聚焦产生胶囊

[0150] 在一些情况下,胶囊可通过流动聚焦产生。流动聚焦是这样的一种方法,通过该方法与第二流体不混溶的第一流体流入第二流体中。参照图 12,包含单体、交联剂、引发剂和水性表面活性剂的第一(例如,水性)流体 1201 流入到包含表面活性剂和促进剂的第二(例如,油)流体 1202 中。在微流体装置的 T 型接合点 1203 处进入第二流体后,第一流体的小滴脱离第一流体流,并且由于第一流体中的单体、交联剂和引发剂与第二流体中的促进剂混合,胶囊壳开始形成 1204。由此,形成胶囊。随着胶囊向下游行进,壳由于更多地暴露于促进剂而变得更厚。也可以利用改变试剂的浓度来改变胶囊壳的厚度和渗透性。

[0151] 可通过例如,在第一流体中包含物质来包封物质,或其他分区如小滴。将物质包含在第二流体中可使物质嵌入胶囊的壳中。当然,根据特定样品处理方法的需要,也可以将相

反转——即，第一相可以是油相而第二相可以是水相。

2. 通过流动聚焦产生在胶囊内的胶囊

[0152] 在一些情况下，可通过流动聚焦产生在胶囊内的胶囊。参照图 13，包含胶囊、单体、交联剂、引发剂和水性表面活性剂的第一（例如，水性）流体 1301 流入到包含表面活性剂和促进剂的第二（油）流体 1302 中。在微流体装置的 T 型接合点 1303 处进入第二流体后，第一流体的小滴脱离第一流体流，并且由于第一流体中的单体、交联剂和引发剂与第二流体中的促进剂混合，第二胶囊壳开始在胶囊周围形成 1304。由此，形成在胶囊内的胶囊。随着胶囊向下游行进，壳由于更多地暴露于促进剂而变得更厚。也可以利用改变试剂的浓度来改变第二胶囊壳的厚度和渗透性。

[0153] 可通过例如，在第一流体中包含物质来包封物质。将物质包含在第二流体中可使物质嵌入胶囊的第二壳中。当然，根据特定样品处理方法的需要，也可以将相反转——即，第一相可以是油相而第二相可以是水相。

3. 胶囊的批量产生

[0154] 在一些情况下，可使用胶囊前体，如乳液中的小滴，批量产生胶囊。可通过任何合适的方法，例如，通过产生具有包含单体、交联剂、引发剂和表面活性剂的小滴的乳液而形成胶囊前体。然后可将促进剂添加至介质中，从而导致胶囊形成。对于流动聚焦的方法，可通过改变反应物的浓度和暴露于促进剂的时间来改变壳的厚度。然后可将胶囊洗涤并回收。对于本文所述的任何方法，可将物质，包括其他分区，包封在胶囊内，或者如果合适，包封在壳内。

[0155] 在另一个实例中，在乳液产生过程期间，可将乳液的小滴暴露于存在于出口孔中的促进剂。例如，可通过任何合适的方法，如图 12 中示出的流动聚焦方法形成胶囊前体。可将促进剂包含在位于 T 型接合点的出口处的介质（例如，位于图 12 的水平通道的极右端处的介质）中，而非将促进剂包含在第二流体 1202 中。当乳液小滴（即，胶囊前体）离开通道时，其接触包含促进剂的介质（即，出口介质）。如果胶囊前体具有小于出口介质的密度的密度，则胶囊前体将上升通过介质，从而确保对流和扩散暴露于促进剂并减小在通道出口处聚合的可能性。

VI. 物质

[0156] 本公开内容的方法、组合物、系统、装置和试剂盒可与任何合适的物质一起使用。物质可以是，例如，在样品处理中使用的任何物质，如试剂或分析物。示例性的物质包括全细胞、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、限制酶、连接酶、聚合酶、条形码、衔接子、小分子、抗体、荧光团、脱氧核苷三磷酸（dNTP）、双脱氧核苷三磷酸（ddNTP）、缓冲液、酸性溶液、碱性溶液、对温度敏感的酶、对 pH 敏感的酶、对光敏感的酶、金属、金属离子、氯化镁、氯化钠、锰、水性缓冲液、温和的缓冲液、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、scaRNA、微 RNA、dsRNA、核酶、核糖开关和病毒 RNA、锁核酸（LNA）的全部或一部分、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、适体等。总之,所使用的物质将根据特定的样品处理需要而变化。

[0157] 在一些情况下,分区包含具有相似属性的一组物质(例如,一组酶、一组矿物质、一组寡核苷酸、不同条形码的混合物、相同条形码的混合物)。在其他情况下,分区包含物质的不均匀混合物。在一些情况下,物质的不均匀混合物包含对于进行特定反应所必需的所有组分。在一些情况下,这样的混合物包含除了对于进行反应所必需的1、2、3、4、5种或更多种组分外,对于进行该反应所必需的所有组分。在一些情况下,这样的附加组分包含在不同的分区内,或在分区之内或周围的溶液内。

[0158] 物质可以是天然存在的或合成的。物质可存在于使用本领域已知的任何方法获得的样品中。在一些情况下,样品可在分析其中的分析物之前进行处理。

[0159] 物质可以从任何合适的位置获得,包括从生物体、全细胞、细胞制品和来自任何生物体、组织、细胞或环境的无细胞组合物获得。物质可从环境样品、活检物、抽出物、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、水样品或灰尘样品获得。在一些情况下,物质可从体液获得,该体液可以包括血液、尿液、粪便、血清、淋巴、唾液、粘膜分泌物、汗液、中枢神经系统液、阴道液或精液。物质也可从生产的产品如化妆品、食品、个人护理产品等获得。物质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。

[0160] 在一些情况下,物质可根据质量来量化。物质可以以约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000ng、1 μ g、5 μ g、10 μ g、15 μ g或20 μ g的质量提供。物质可以以至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000ng、1 μ g、5 μ g、10 μ g、15 μ g或20 μ g的质量提供。物质可以以少于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000ng、1 μ g、5 μ g、10 μ g、15 μ g或20 μ g的质量提供。物质可以以范围为约1-10、10-50、50-100、100-200、200-1000、1000-10000ng、1-5 μ g或1-20 μ g的质量提供。如本公开内容的其他地方所述,如果物质为多核苷酸,则可使用扩增来增加多核苷酸的量。

[0161] 也可将多核苷酸量化为“基因组当量”。基因组当量是等同于从中获得靶多核苷酸的生物体的一个单倍体基因组的多核苷酸的量。例如,单个二倍体细胞含有两个基因组当量的DNA。多核苷酸可以以范围为约1-10、10-50、50-100、100-1000、1000-10000、10000-100000或100000-1000000个基因组当量的量提供。多核苷酸可以以至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、8500、9000、9500、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200000、300000、400000、500000、600000、700000、800000、900000或1000000个基因组当量的量提供。多核苷酸可以以少于约2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、50、100、500、1000、1500、2000、2500、3000、3500、4000、4500、5000、5500、6000、6500、7000、7500、8000、

8500、9000、9500、10000、20000、30000、40000、50000、60000、70000、80000、90000、100000、200000、300000、400000、500000、600000、700000、800000、900000 或 1000000 个基因组当量的量提供。

[0162] 多核苷酸还可以根据所提供的序列覆盖的量来量化。序列覆盖的量是指代表重构序列中给定核苷酸的读取值的平均数。通常,区域被测序的次数越多,所获得的序列信息就越准确。多核苷酸可以以提供约 0.1X-10X、10X-50X、50X-100X、100X-200X 或 200X-500X 的序列覆盖范围的量提供。多核苷酸可以以提供至少约 0.1X、0.2X、0.3X、0.4X、0.5X、0.6X、0.7X、0.8X、0.9X、1.0X、5X、10X、25X、50X、100X、125X、150X、175X 或 200X 的序列覆盖的量提供。多核苷酸可以以提供小于约 0.2X、0.3X、0.4X、0.5X、0.6X、0.7X、0.8X、0.9X、1.0X、5X、10X、25X、50X、100X、125X、150X、175X 或 200X 的序列覆盖的量提供。

[0163] 在一些情况下,物质在特定步骤之前或之后引入到分区中。例如,可在将细胞样品分区分到分区中之后,将裂解缓冲试剂引入到分区中。在一些情况下,顺序引入试剂和/或包含试剂的分区,使得不同的反应或操作在不同的步骤发生。还可在穿插有反应或操作步骤的步骤中加载试剂(或包含试剂的分区)。例如,可将包含用于分子(例如,核酸)片段化的试剂的胶囊加载到孔中,接着进行片段化步骤,该步骤之后可以是包含用于连接条形码(或其他独特标识符,例如,抗体)的试剂的胶囊的加载以及随后的条形码与片段化的分子的连接。

VII. 分析物和其他物质的处理

[0164] 在一些情况下,本公开内容的方法、组合物、系统、装置和试剂盒可用于处理含有物质例如分析物的样品。可进行任何合适的处理。

a. 靶多核苷酸的片段化

[0165] 在一些情况下,本公开内容的方法、组合物、系统、装置和试剂盒可用于多核苷酸片段化。多核苷酸的片段化被用作包括多核苷酸测序在内的多种方法中的步骤。通常以长度描述的多核苷酸片段的大小(根据每个片段中核苷酸的线性数目来量化)可根据靶多核苷酸的来源、用于片段化的方法和期望的应用而变化。可使用单个片段化步骤或多个片段化步骤。

[0166] 使用本文所述的方法产生的片段可以为约 1-10、10-20、20-50、50-100、50-200、100-200、200-300、300-400、400-500、500-1000、1000-5000、5000-10000、10000-100000、100000-250000 或 250000-500000 个核苷酸的长度。使用本文所述的方法产生的片段可以为至少约 10、20、100、200、300、400、500、1000、5000、10000、100000、250000、500000 个或更多个核苷酸的长度。使用本文所述的方法产生的片段可以为小于约 10、20、100、200、300、400、500、1000、5000、10000、100000、250000、500000 个核苷酸的长度。

[0167] 使用本文所述的方法产生的片段可以具有约 1-10、10-20、20-50、50-100、50-200、100-200、200-300、300-400、400-500、500-1000、1000-5000、5000-10000、10000-100000、100000-250000 或 250000-500000 个核苷酸的平均长度或中值长度。使用本文所述的方法产生的片段可以具有至少约 10、20、100、200、300、400、500、1000、5000、10000、100000、250000、500000 个或更多个核苷酸的平均长度或中值长度。使用本文所述的方法产生的片段可以具有小于约 10、20、100、200、300、400、500、1000、5000、10000、100000、250000、500000 个核苷酸的平均长度或中值长度。

[0168] 许多片段化方法是本领域中已知的。例如,可通过物理、机械或酶方法进行片段化。物理片段化可包括将靶多核苷酸暴露于热或紫外线。机械破坏可用于将靶多核苷酸机械地剪切成所需范围的片段。机械剪切可通过本领域中已知的许多方法来完成,包括靶多核苷酸的重复移液、超声处理(例如,使用超声波)、空腔化和雾化。还可使用酶方法对靶多核苷酸进行片段化。在一些情况下,可使用酶如使用限制酶进行酶消化。

[0169] 尽管在本公开内容的上文或任何其他地方参照“靶”多核苷酸对前述段落和本公开内容的一些段落中所述的片段化方法进行了描述,但这并不意味着限制。本文所述或本领域中已知的任何片段化方法均可应用于本发明中使用的任何多核苷酸。在一些情况下,该多核苷酸可以是靶多核苷酸,如基因组。在其他情况下,该多核苷酸可以是希望将其进一步片段化的靶多核苷酸的片段。在另外的其他情况下,进一步片段化的片段仍可以再进行进一步片段化。任何合适的多核苷酸可根据本文所述的方法进行片段化。

[0170] 限制酶可用于进行靶多核苷酸的特异性或非特异性片段化。本公开内容的方法可使用一种或多种类型的限制酶,其通常被描述为 I 型酶、II 型酶和 / 或 III 型酶。II 型和 III 型酶通常是可商购的并且是本领域公知的。II 型和 III 型酶识别双链多核苷酸序列内的核苷酸碱基对的特定序列(“识别序列”或“识别位点”)。在结合并识别这些序列后,II 型和 III 型酶切割该多核苷酸序列。在一些情况下,切割将产生具有被称为“粘端”的突出单链 DNA 部分的多核苷酸片段。在其他情况下,切割将不会产生具有突出端的片段,从而形成“平端”。本公开内容的方法可以包括使用产生粘端或平端的限制酶。

[0171] 限制酶可识别靶多核苷酸中的多个识别位点。一些限制酶(“精确切割物”)只识别单个识别位点(例如,GAATTC)。其他限制酶更混杂,并识别多于一个识别位点,或多个识别位点。一些酶在识别位点内的单个位置切割,而其他酶可能在多个位置切割。一些酶在识别位点内的相同位置切割,而其他酶在可变的位置切割。

[0172] 本公开内容提供了选择一种或多种限制酶来产生所需长度的片段的方法。多核苷酸片段化可用计算机(in silico)模拟,并且可对片段化进行优化以获得在特定大小范围内的多核苷酸片段的最大数目或分数,同时使得在不期望大小范围内的片段的数目或分数减至最小。可应用优化算法来选择两种或更多种酶的组合以产生具有所需片段数量分布的所需片段大小。

[0173] 多核苷酸可同时或顺序地暴露于两种或更多种限制酶。这可以通过,例如,向分区添加多于一种限制酶,或通过向分区添加一种限制酶、进行消化、灭活该限制酶(例如,通过热处理)并随后加入第二种限制酶来完成。任何合适的限制酶可在本文提出的方法中单独或组合使用。

[0174] 在一些情况下,物质是一种为“稀有切割物(rare-cutter)”的限制酶。如本文所用的术语“稀有切割物”通常是指具有仅稀有地在基因组中出现的识别位点的酶。通过用限制酶切割假设的随机基因组而产生的限制性片段的大小可以近似于 4^N , 其中 N 是在酶的识别位点中的核苷酸的数目。例如,具有由 7 个核苷酸组成的识别位点的酶将每 4^7 bp 切割一次基因组,产生约 16,384bp 的片段。稀有切割物酶通常具有包含 6 个或者更多个核苷酸的识别位点。例如,稀有切割物酶可具有包含或由 6、7、8、9、10、11、12、13、14 或 15 个核苷酸组成的识别位点。稀有切割物酶的实例包括 NotI (GCGGCCGC)、XmaIII (CGGCCG)、SstII (CCGCGG)、SalI (GTCGAC)、NruI (TCGCGA)、NheI (GCTAGC)、Nb. BbvCI (CCTCAGC)、BbvCI (CCTCAGC)、

AscI (GGCGCGCC)、AseI (GCGATCGC)、FseI (GGCCGGCC)、PacI (TTAATTAA)、PmeI (GTTTAAAC)、SbfI (CCTGCAGG)、SgrAI (CRCCGGYG)、SwaI (ATTTAAAT)、BspQI (GCTCTTC)、SapI (GCTCTTC)、SfiI (GGCCNNNNNGGCC)、CspCI (CAANNNNNGTGG)、AbsI (CCTCGAGG)、CciNI (GCGGCCGC)、FspAI (RTGCGCAY)、MauBI (CGCGCGCG)、MreI (CGCCGGCG)、MssI (GTTTAAAC)、PalAI (GGCGCGCC)、RgaI (GCGATCGC)、RigI (GGCCGGCC)、SdaI (CCTGCAGG)、SfaI (GCGATCGC)、SgfI (GCGATCGC)、SgrDI (CGTCGACG)、SgsI (GGCGCGCC)、SmiI (ATTTAAAT)、SrfI (GCCCCGGG)、Sse2321 (CGCCGGCG)、Sse83871 (CCTGCAGG)、LguI (GCTCTTC)、PciSI (GCTCTTC)、AarI (CACCTGC)、AjuI (GAANNNNNNNTGG)、AloI (GAACNNNNNTCC)、BarI (GAAGNNNNNTAC)、PpiI (GAACNNNNNTC)、PsrI (GAACNNNNNTAC) 和其他稀有切割物酶。

[0175] 在一些情况下,多核苷酸可以同时片段化和条形码化。例如,可使用转座酶(例如,NEXTERA)对多核苷酸进行片段化并向该多核苷酸添加条形码。

VIII. 刺激响应性

[0176] 在一些情况下,刺激可以用于触发物质从分区的释放。通常,刺激可导致对分区结构如孔壁、斑点的组分、小滴(例如,乳液中的小滴)的稳定性或胶囊壳的破坏。这些刺激在诱导分区释放其内容物方面是特别有用的。由于分区可包含在另一分区内,并且各个分区可对不同的刺激有响应(或无响应),因此刺激响应性可用于将一个分区(例如,对刺激有响应的分区)的内容物释放到另一个分区(例如,对该刺激无响应或对该刺激有较低响应的分区)中。

[0177] 在一些情况下,可以通过施加溶解内层胶囊的刺激将内层胶囊的内容物释放到外层胶囊的内容物中,从而产生含有混合样品的胶囊。当然,该实施方案仅仅是说明性的,并且刺激响应性可以用于将任何合适的分区的内容物释放到任何其他合适的分区、介质或容器(关于分区内的分区的更具体的实例参见,例如,表1)中。

[0178] 可以使用的刺激的实例包括化学刺激、本体变化(bulk change)、生物刺激、光、热刺激、磁刺激、向孔中添加介质及其任意组合,如下文更充分描述的。(参见,例如,Esser-Kahn 等人,(2011)Macromolecules44:5539-5553;Wang 等人,(2009)ChemPhysChem 10:2405-2409)。

a. 化学刺激和本体变化

[0179] 许多化学触发物可以用来触发分区的破裂(例如,Plunkett 等人,Biomacromolecules, 2005, 6:632-637)。这些化学变化的实例可以包括但不限于pH介导的分区组分完整性的变化,通过交联键的化学裂解引起的分区组分的崩解,以及触发的分区组分的解聚。本体变化也可用于触发分区的破裂。

[0180] 溶液pH的变化,如pH的下降,可通过许多不同的机制触发分区的破裂。酸的加入可通过多种机制引起分区的一部分的降解或分解。质子的加入可以分解分区组分中的聚合物的交联,破坏分区组分中的离子键或氢键,或在分区组分中产生纳米孔以允许内部内容物渗漏至外部。pH的变化也可以使乳液去稳定,从而导致小滴的内容物的释放。

[0181] 在一些实例中,分区由包含酸可降解的化学交联剂如缩酮的材料生产。pH的降低,特别是降至低于5的pH,可诱使缩酮转化成酮和两个醇并促进分区的破裂。在其他实例中,分区可以由包含一种或多种对pH敏感的聚合电解质的材料生产。pH的降低可破坏这类分区的离子键或氢键相互作用,或在其中产生纳米孔。在一些情况下,由包含聚合电解质的材

料制成的分区包含当 pH 变化时扩张和收缩的带电荷的、基于凝胶的芯。

[0182] 构成分区的交联材料的破坏可以通过多种机制来完成。在一些实例中,分区可以与诱导氧化、还原或其他化学变化的各种化学品接触。在一些情况下,可以使用还原剂,如 β -巯基乙醇,使得分区的二硫键断裂。另外,可以加入酶,以裂解形成分区的材料中的肽键,从而导致分区的完整性的丧失。

[0183] 解聚也可以用于破坏分区。可加入化学触发物,以促进保护头基的移除。例如,触发物可导致聚合物内的碳酸酯或氨基甲酸酯的头基的移除,而这又引起解聚和物质从分区内部的释放。

[0184] 在又一个实例中,化学触发物可包括渗透触发物,从而溶液中的离子或溶质浓度的变化诱发用于形成分区的材料的膨胀。膨胀可引起内压的积聚,使得分区破裂以释放其内容物。膨胀也可以引起材料的孔径的增大,从而使分区内包含的物质扩散出,反之亦然。

[0185] 也可以通过本体或物理变化,如压力诱发的破裂、熔化或孔隙率的变化,使分区释放其内容物。

b. 生物刺激

[0186] 也可利用生物刺激来触发分区的破裂。通常,生物触发物类似于化学触发物,但许多实例使用生物分子或在生命系统中常见的分子,如酶、肽、糖类、脂肪酸、核酸等。例如,分区可以由包含具有肽交联的聚合物的材料制成,该肽交联对由特异性蛋白酶引起的裂解是敏感的。更具体地,一个实例可以包括由包含 GFLGK 肽交联的材料制成的分区。在加入生物触发物如蛋白酶组织蛋白酶 B 时,壳壁的肽交联被裂解并释放胶囊的内容物。在其他情况下,蛋白酶可以被热活化。在另一实例中,分区包含含有纤维素的组分。水解酶壳聚糖的加入用作对纤维素键的裂解、包含壳聚糖的分区的组分的解聚以及其内部内容物的释放的生物触发物。

c. 热刺激

[0187] 当施加热刺激时,也可以诱使分区释放其内容物。温度的变化可引起分区的多种变化。热的变化可以导致分区的熔化以使得分区的一部分崩解,或乳液的破裂。在其他情况下,热可以增加分区的内部组分的内压,使得分区破裂或爆破。在另一些情况下,热可以将分区转变成皱缩的脱水状态。热也可以作用于用作构建分区的材料的热敏感性聚合酶。

[0188] 在一个实例中,分区由包含热敏感性水凝胶的材料制成。在施加热,如高于 35°C 的温度时,水凝胶材料收缩。材料的突然收缩增大压力并破坏分区。

[0189] 在一些情况下,用来产生分区的材料可以包含二嵌段聚合物或具有不同热敏性的两种聚合物的混合物。一种聚合物可能在施加热后特别可能收缩,而另一种则是更加热稳定的。当向这样的壳壁施加热时,热敏性聚合物可以收缩,而另一种则保持不变,从而引起孔隙的形成。在另一些情况下,用来产生分区的材料可以包含磁性纳米颗粒。暴露于磁场可导致热的生成,从而导致分区的破裂。

d. 磁刺激

[0190] 在用来产生分区的材料中包含磁性纳米颗粒可以允许触发分区的破裂,如上所述,以及使得能够将这些分区引导至其他分区(例如,将胶囊引导至阵列中的孔)。在一个实例中,将 Fe_3O_4 纳米颗粒引入用来产生分区的材料中在振荡磁场刺激的存在下触发破裂。

e. 电和光刺激

[0191] 分区也可由于电刺激而破裂。类似于前一节中描述的磁性颗粒,电敏感性颗粒可以允许分区的触发破裂以及其他功能,如在电场中的对齐或氧化还原反应。在一个实例中,由包含电敏感性材料的材料制成的分区在电场中对齐,使得可以控制内部试剂的释放。在其他实例中,电场可以在可增大孔隙率的分区内诱发氧化还原反应。

[0192] 光刺激也可以用于破坏分区。许多光触发物是可能的,并可包括使用多种分子(如纳米颗粒和能够吸收特定范围波长的光子的生色团)的系统。例如,金属氧化物涂层可以用来产生某些分区。对涂覆有 SiO₂/TiO₂ 的分区的紫外线照射可导致分区壁的崩解。在又一实例中,可以在用来产生分区的材料中引入光开关材料(photo switchable materials),如偶氮苯基团。在施加紫外线或可见光时,诸如这些的化学物质在光子吸收时经历可逆的顺式至反式异构化。在该方面,光开关的引入导致分区的一部分的崩解,或分区的一部分的孔隙率的增大。

f. 刺激的施加

[0193] 本公开内容的装置、方法、组合物、系统和试剂盒可以与提供这样的触发物或刺激的任何设备或装置组合使用。例如,如果刺激是热刺激,则装置可以与加热的或热控制的板组合使用,其允许对孔的加热并可诱发胶囊的破裂。对于热刺激可以使用诸多热传递方法中的任何方法,包括但不限于通过辐射传热、对流传热或传导传热施加热。在其他情况下,如果刺激是生物酶,则该酶可以注入到装置中,使得它沉积至各个孔中。在另一方面,如果刺激是磁场或电场,则装置可与磁板或电板组合使用。

IX. 应用

a. 多核苷酸测序

[0194] 通常,本文提供的方法和组合物可用于制备多核苷酸片段以用于下游应用如测序。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、合成测序、单分子测序、纳米孔测序、连接测序、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。

[0195] 在一些情况下,对不同数目的片段进行测序。例如,在一些情况下,对约 30% -90% 的片段进行测序。在一些情况下,对约 35% -85%、40% -80%、45% -75%、50% -70%、55% -65% 或 50% -60% 的片段进行测序。在一些情况下,对至少约 30%、40%、50%、60%、70%、80% 或 90% 的片段进行测序。在一些情况下,对少于约 30%、40%、50%、60%、70%、80% 或 90% 的片段进行测序。

[0196] 在一些情况下,装配来自片段的序列以提供关于比单个序列读取值更长的原始靶多核苷酸的连续区域的序列信息。单个序列读取值可以是约 10-50、50-100、100-200、200-300、300-400 个或更多个核苷酸的长度。

[0197] 条形码标签的身份可用来排列来自单个片段的序列以及区分单元型。例如,在单个片段的分区过程中,可以将亲本多核苷酸片段分隔至不同的分区中。随着分区数目的增加,来自母本和父本单元型两者的片段包含在同一分区内的可能性变得小到可以忽略不计。因此,可以装配并排列来自同一分区中的片段的序列读取值。

b. 多核苷酸取相

[0198] 本公开内容还提供了用于以可使得能够生成取相或连接信息的方式制备多核苷酸片段的方法和组合物。这样的信息可以允许检测序列中的连接的遗传变异,包括被长多核苷酸段隔开的遗传变异(例如,SNP、突变、插入缺失(indel)、拷贝数变异、颠换、易位、倒位等)。术语“插入缺失”是指导致共定位的插入和缺失以及核苷酸的净增加或损失的突变。“微插入缺失(microindel)”是导致1-50个核苷酸的净增加或损失的插入缺失。这些变异可能以顺式或反式关系存在。在顺式关系中,两个或更多个遗传变异存在于同一多核苷酸或链中。在反式关系中,两个或更多个遗传变异存在于多个多核苷酸分子或链上。

[0199] 本文提供的方法可以用来确定多核苷酸取相。例如,可以对多核苷酸样品(例如,横跨给定基因座或多个基因座的多核苷酸)进行分区以使得每个分区存在多核苷酸的至多一个分子。然后可以对多核苷酸进行片段化、条形码化和测序。可以检查序列的遗传变异。检测到用两个不同条形码标记的同一序列中的遗传变异可以指示这两个遗传变异源自DNA的两条单独的链,反映出反式关系。相反地,检测到用相同条形码标记的两个不同的遗传变异可以指示这两个遗传变异来自DNA的同一条链,反映出顺式关系。

[0200] 相位信息对多核苷酸片段的表征可能至关重要,尤其是当多核苷酸片段源自患有或疑似患有特定疾病或病症(例如,遗传性隐性疾病,如囊性纤维化、癌症等)或处于其风险中的受试者时。该信息可以能够区分以下可能性:(1)两个遗传变异位于DNA的同一条链上的同一基因内,以及(2)两个遗传变异位于同一基因内但位于DNA的单独的链上。可能性(1)可以指示该基因的一个拷贝是正常的并且该个体没有该疾病,而可能性(2)可以指示该个体具有或将发展成该疾病,尤其是当这两个遗传变异在存在于同一基因拷贝内时损害基因的功能时。类似地,相位信息还可以能够区分以下可能性:(1)两个遗传变异,各自位于DNA的同一条链上的不同基因内,以及(2)两个遗传变异,各自位于不同基因内但位于DNA的单独的链上。

c. 从少量细胞对多核苷酸进行测序

[0201] 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(例如,SNP、突变、插入缺失、拷贝数变异、颠换、易位、倒位等)。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。

[0202] 在一个实例中,方法包括对细胞样品(或粗细胞提取物)进行分区以使得每个分区存在至多一个细胞(或一个细胞的提取物),裂解细胞,通过本文所述的任何方法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,合并,并测序。

[0203] 如本文其他部分所述,条形码和其他试剂可以包含在分区(例如,胶囊)内。这些胶囊可以在细胞加载之前、之后或同时加载到另一分区(例如,孔)内,使得每个细胞与不同的胶囊接触。该技术可以用来将独特条形码附接至从每个细胞获得的多核苷酸。然后将所得的标记的多核苷酸合并并且测序,并可以使用条形码追踪多核苷酸的起源。例如,可以确定具有相同条形码的多核苷酸起源于相同细胞,同时可以确定具有不同条形码的多核苷酸起源于不同细胞。

[0204] 本文描述的方法可以用于检测致癌突变在整个癌性肿瘤细胞群体中的分布。例如,一些肿瘤细胞可以具有癌基因(例如,HER2、BRAF、EGFR、KRAS)在两个等位基因中的突变或扩增(纯合的),另外一些肿瘤细胞可以具有在一个等位基因中的突变(杂合的),而再一些肿瘤细胞可能不具有突变(野生型)。本文所述的方法可以用来检测这些差异,并且还可以量化纯合、杂合和野生型细胞的相对数目。例如,这样的信息可用于对特定癌症进行分期和/或用于随时间监测癌症的进展及其治疗。

[0205] 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,这可以指示该癌症是更加良性的或不到晚期。

d. 基因表达的分析

[0206] 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。

[0207] 可将细胞直接置于分区(例如,微孔)中并裂解。裂解后,可使用本发明的方法对细胞的多核苷酸进行片段化和条形码化以供测序。多核苷酸也可以在将其引入在本发明的方法中使用的分区中之前从细胞提取。mRNA 的逆转录可在本文所述的分区中,或在这样的分区的外部进行。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。

[0208] 上面提出的方法相比于当前的多核苷酸处理方法提供了几个优势。第一,操作者之间的可变性大大降低。第二,该方法可在具有低成本并可容易地制造的微流体装置中进行。第三,靶多核苷酸的受控片段化使使用者能制备具有确定和合适长度的多核苷酸片段。这有助于对多核苷酸进行分区,并且由于过大片段的存在,还减少了序列信息丢失的程度。所述方法和系统还提供了保持所处理的多核苷酸的完整性的简易工作流程。此外,限制酶的使用使得使用者能够产生 DNA 突出端(“粘端”),其可针对与衔接子和/或条形码的相容性而设计。

e. 从细胞对多核苷酸如染色体的分区

[0209] 在一个实例中,本公开内容提供的方法、组合物、系统、装置和试剂盒可用于从细胞对多核苷酸(包括全部染色体)进行分区。在一个实例中,将单个细胞或多个细胞(例如,2、10、50、100、1000、10000、25000、50000、100000、500000、1000000 个或更多细胞)加载至具有裂解缓冲液和蛋白酶 K 的容器中,并温育指定的一段时间。多个细胞的使用将使得能够进行多核苷酸取相,例如,通过将每个待分析的多核苷酸在其自己的分区中进行分区。

[0210] 温育后,例如通过将细胞裂解物流动聚焦至胶囊中,对细胞裂解物进行分区。如果将要进行取相,则进行流动聚焦,使得每个胶囊仅包含单个分析物(例如,单个染色体),或仅包含任何特定染色体的单个拷贝(例如,第一染色体的一个拷贝和第二染色体的一个拷贝)。在一些情况下,可将多个染色体包封在相同的胶囊内,只要这些染色体不是相同的染色体。在温和流动下进行包封,以使多核苷酸的剪切减至最小。胶囊可以是多孔的,以允许洗涤胶囊的内容物以及将试剂引入胶囊中,同时将多核苷酸(例如,染色体)保持在胶囊

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,处理包封的多核苷酸(例如,染色体)。胶囊壳保护包封的多核苷酸(例如,染色体)免遭剪切和进一步的降解。当然,也可将该方法应用于任何其他细胞组分。

[0211] 如上所述,胶囊壳可用于保护多核苷酸免遭剪切。然而,胶囊也可用作分区,以能够进行多核苷酸或其他分析物的区室化剪切。例如,在一些情况下,多核苷酸可包封在胶囊内,随后经历超声剪切或任何其他合适的剪切。胶囊壳可配置为在剪切下保持完整,而包封的多核苷酸可受到剪切,但仍将保留在胶囊内。在一些情况下,水凝胶小滴可用于达到相同的目的。

f. 癌症突变检测和法医学

[0212] 本文所述的通过分区中基于扩增的条形码化方案的条形码化方法可用于由降解的样品如福尔马林固定、石蜡包埋的 (FFPE) 组织切片产生条形码文库。本文所述的方法可以能够鉴定分区内的源自相同初始分子的所有扩增子。确实,通过采用分区条形码化,可以保留关于独特起始多核苷酸的信息。由于可以区分来自不同的原始分子的扩增子,因此这样的鉴定可有助于文库复杂性的确定。此外,本文所述的方法可允许评估独特的覆盖,这可有助于确定变体判定 (calling) 灵敏度。这些优势在癌症突变检测和法医学中可能是特别有用的。

g. 低输入 DNA 应用 (循环肿瘤细胞 (CTC) 测序)

[0213] 本文所述的条形码化方法可用于低多核苷酸输入应用,例如,循环肿瘤细胞 (CTC) 的核酸的测序。例如,在分区内进行的本文所述的 MALBAC 方法可有助于在低多核苷酸输入应用中获得良好的数据质量和 / 或有助于过滤掉扩增错误。

X. 试剂盒

[0214] 在一些情况下,本公开内容提供了包含用于产生分区的试剂的试剂盒。该试剂盒可包含用于产生分区和在分区内的分区的任何合适的试剂和说明。

[0215] 在一个实例中,试剂盒包含用于在乳液中的小滴内产生胶囊的试剂。例如,试剂盒可包含用于产生胶囊的试剂、用于产生乳液的试剂和关于将胶囊引入到乳液的小滴中的说明。如本公开内容通篇所说明的,可将任何合适的物质掺入小滴和 / 或胶囊中。本公开内容的试剂盒还可提供这些物质中的任何物质,如预分区的包含条形码的多核苷酸。相似地,如本公开内容通篇所述,可将胶囊设计成在施加刺激时释放其内容物至乳液的小滴中。

[0216] 在另一个实例中,试剂盒包含用于产生在胶囊内的胶囊的试剂。例如,试剂盒可包含用于产生内层胶囊的试剂、用于产生外层胶囊的试剂和关于产生在胶囊内的胶囊的说明。如本公开内容通篇所说明的,可将任何合适的物质掺入到内层和 / 或外层胶囊中。本公开内容的试剂盒还可提供这些物质中的任何物质,如预分区的包含条形码的多核苷酸。相似地,如本公开内容通篇所述,可将内层胶囊设计成在施加刺激时释放其内容物至外层胶囊中。

XI. 装置

[0217] 在一些情况下,本公开内容提供了包含用于处理分析物的分区的装置。装置可以是如本公开内容的其他地方所述的微孔阵列或微斑点阵列 (microspot array)。装置可以以包含任何合适的分区的形式形成。在一些情况下,装置包含多个孔,或多个斑点。当然,装置中的任何分区也可容纳其他分区,如胶囊、乳液中的小滴等。

[0218] 装置可由任何合适的材料形成。在一些实例中,装置由选自熔融二氧化硅、钠钙玻璃、硼硅酸盐玻璃、聚(甲基丙烯酸甲酯)、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、塑料及其组合的材料形成。

[0219] 在一些情况下,装置包含供流体向分区内和在分区间流动的通道。可使用任何合适的通道。装置可包含流体入口和流体出口。入口和出口可附接至液体处理装置,以将物质引入装置中。可在引入任何物质之前或之后对装置进行密封。

[0220] 可在装置的不同部分中使用亲水性和/或疏水性的材料。例如,在一些情况下,本公开内容的装置包含分区,该分区具有包含亲水性材料的内表面。在一些情况下,分区的外表面包含疏水性材料。在一些情况下,流体流动路径用亲水性或疏水性材料包覆。

[0221] 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、装置和试剂盒在对物质进行分区、对寡核苷酸进行分区、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、核苷酸取相、从少量细胞对多核苷酸进行测序、分析基因表达、从细胞对多核苷酸进行分区、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)测序或其组合中的用途。

实施例

实施例 1:通过不对称 PCR 和添加部分互补通用序列产生包含条形码序列的叉状衔接子的文库

[0222] 本实施例提供了用于制备包含与下一代测序技术(例如,ILLUMINA)兼容的条形码序列的叉状衔接子的方法。在该实施例中,条形码置于位置 207 中,如图 2 所示。

[0223] 参照图 4,合成包含第一固定区域 402、条形码区域 403 和第一测序引物区域 404 的单链衔接子-条形码多核苷酸序列 401。条形码区域 403 是通过在每一偶联步骤中包含等摩尔浓度的 A、G、T 和 C 而合成的七个核苷酸的随机序列。

[0224] 合成后,将单链衔接子-条形码多核苷酸 401 稀释到油包水乳液中的水性小滴内,使得每个小滴平均包含 0.1 个多核苷酸。该小滴还包含用于通过不对称 PCR 扩增单链衔接子-条形码多核苷酸 401 的试剂(例如,聚合酶、引物、dNTP、缓冲液、盐)和 DNA 嵌入染料(例如,溴化乙锭)。反向引物以多于正向引物的量存在,或反之亦然,使得不对称扩增能够进行。扩增该多核苷酸,并且该反应继续进行通过指数期扩增 410(产生双链产物 405)和线性期扩增 411(产生单链产物 406)。

[0225] 将小滴在荧光辅助细胞分选仪(FACS)412 上分选以收集包含扩增的多核苷酸的小滴。将部分互补通用序列 407 添加至分区以产生部分退火的叉状结构 413。部分互补通用序列 407 包含第二固定区域 408 和第二测序引物区域 409,后者包含与待测序的多核苷酸靶标上的 A 突出端(未示出)相容的 T 突出端。

实施例 2:用片段化酶(Fragmentase)进行的片段化和条形码化

[0226] 如实施例 1 中所述或通过本公开内容中描述的任何其他方法合成、分区、扩增并分选包含第一固定区域 402、条形码区域 403 和第一测序引物区域 404 的单链衔接子-条形码多核苷酸序列(例如,图 4:401)。在包含单链衔接子-条形码多核苷酸序列的小滴上进

行界面聚合,以产生多个包含单链衔接子-条形码多核苷酸序列 406 的文库的胶囊,其中文库中的每一个(或大多数)序列在其各自的条形码区域 403 的序列上不同。因此,产生了包封的单链衔接子-条形码多核苷酸的文库。

[0227] 制备了两种混合物。混合物 Z1 包含靶多核苷酸(即,待片段化和条形码化的多核苷酸)、片段化酶(例如,NEBNEXT DSDNA FRAGMENTASE)和部分互补通用序列(例如,图 4:407)。第二混合物 Z2 包含如上所述产生的包封的单链衔接子-条形码多核苷酸的文库,以及足以活化片段化酶的浓度的氯化镁。混合物 Z1、Z2,或者 Z1 和 Z2 两者还包含 T4 聚合酶、Taq 聚合酶和热稳定连接酶。

[0228] 将混合物 Z1 和 Z2 合并,并且根据本公开内容的其他地方所述的方法,如流动聚焦,形成在胶囊内的胶囊。图 5 示出了根据上述方法产生的在胶囊内的胶囊。外层胶囊 501 包含内层胶囊 502 和介质 504。内层胶囊 502 是包封的单链衔接子-条形码多核苷酸的文库中的一员。因此,内层胶囊 502 包含单链衔接子-条形码多核苷酸 503 的多个拷贝,后者可用于将相同的条形码附接至分区如外层胶囊 501 内的多核苷酸。

[0229] 介质 504 包含上述混合物 Z1 和混合物 Z2 的内容物。更具体而言,介质 504 包含靶多核苷酸 505、部分互补通用序列 506 以及包含片段化酶、T4 聚合酶、Taq 聚合酶、热稳定连接酶、氯化镁和合适的缓冲液的酶混合物 507。

[0230] 在产生胶囊内的胶囊并将胶囊内的胶囊暴露于合适的条件后,酶处理靶多核苷酸。更具体而言,片段化酶对靶多核苷酸进行片段化并且 T4 聚合酶使得片段化的靶多核苷酸的末端成为平端。然后热灭活片段化酶和 T4 聚合酶,并使用刺激使内层胶囊 502 破裂,释放其内容物至外层胶囊 501 中。Taq 聚合酶向片段化的、平端化的靶多核苷酸上添加 3'-A 突出端。单链衔接子-条形码多核苷酸 503 与部分互补通用序列 506 杂交,从而形成具有与在片段化的靶多核苷酸上的 3'-A 突出端相容的 3'-T 突出端的叉状衔接子。热稳定连接酶将叉状衔接子连接至片段化的靶多核苷酸,从而产生条形码化的靶多核苷酸。然后使外层胶囊 501 破裂,将来自所有外层胶囊的样品合并,并对靶多核苷酸进行测序。随后在测序前,根据需要可进行另外的准备步骤(例如,本体扩增(bulk amplification)、大小选择等)。

[0231] 在一些情况下,混合物 Z1 包含部分互补通用序列 506 的多个版本,其中每个版本具有其自身的样品特异性条形码。

[0232] 此外,尽管上述实施例使用了热稳定连接酶来将包含条形码序列的叉状衔接子附接至靶多核苷酸,但是也可使用 PCR 来完成该步骤,如本公开内容的其他地方所述。

实施例 3:通过超声处理进行的片段化和条形码化

[0233] 如实施例 2 中所述或通过本公开内容中描述的任何其他合适的方法产生包封的单链衔接子-条形码多核苷酸的文库。将靶多核苷酸(即,待片段化的多核苷酸)分区到胶囊中。将包含靶多核苷酸的胶囊配置成能抵抗超声应力。使包含靶多核苷酸的胶囊暴露于超声应力(例如,COVARIS 聚焦超声发生器(COVARIS Focused-Ultrasonicator)),并对靶多核苷酸进行片段化,从而产生片段化的靶多核苷酸胶囊。

[0234] 制备混合物 Z1,其包含包封的单链衔接子-条形码多核苷酸(例如,图 4:406)的文库、片段化的靶多核苷酸胶囊、部分互补通用序列(例如,图 4:407)、酶混合物(T4 聚合酶、Taq 聚合酶和热稳定连接酶)和合适的缓冲液。根据本公开内容的其他地方所述的方

法,如流动聚焦,产生在胶囊内的胶囊。

[0235] 图 6 示出了根据上述方法产生的在胶囊内的胶囊。外层胶囊 601 包含多个内层胶囊 602 和 605 以及介质 604。内层胶囊 602 和 605 分别包括包含单链衔接子-条形码多核苷酸 603 的胶囊和包含片段化的靶多核苷酸 606 的胶囊。内层胶囊 602 包含单链衔接子-条形码多核苷酸 603 的多个拷贝,后者可用于将相同的条形码附接至分区内的多核苷酸,诸如包含在内层胶囊 605 内的片段化的多核苷酸 606。

[0236] 介质 604 含有上述混合物 Z1 的内容物。更具体而言,介质 604 包含部分互补通用序列 607、酶混合物 (T4 聚合酶、Taq 聚合酶和热稳定连接酶) 608 和合适的缓冲液。

[0237] 使包含片段化的靶多核苷酸 606 的内层胶囊 605 暴露于刺激以使其破裂并释放其内容物至外层胶囊 601 的内容物中。T4 聚合酶使得片段化的靶多核苷酸的末端成为平端; Taq 聚合酶向片段化的、平端化的靶多核苷酸上添加 3' -A 突出端。然后热灭活 T4 聚合酶和 Taq 聚合酶,并且施加刺激以将内层胶囊 602 的内容物释放至外层胶囊 601 中。单链衔接子-条形码多核苷酸 603 与部分互补通用序列 607 杂交,从而形成具有与片段化的靶多核苷酸上的 3' -A 突出端相容的 3' -T 突出端的叉状衔接子。热稳定连接酶将叉状衔接子连接至片段化的靶多核苷酸,从而产生条形码化的靶多核苷酸。然后使外层胶囊 601 破裂,将来自所有外层胶囊的样品合并,并对靶多核苷酸进行测序。

[0238] 如实施例 2 中所述,在一些情况下,Z1 可包含部分互补通用序列 607 的多个版本。此外,尽管该实施例示出了通过使用热稳定连接酶对靶多核苷酸进行加条形码,但也可以使用 PCR 来完成该步骤。

实施例 4:通过单引物等温扩增 (SPIA) 和限制消化产生叉状衔接子

[0239] 本实施例示出了通过 SPIA 和限制消化合成叉状衔接子。图 7 提供了可根据本实施例的方法产生的产物 (或中间体) 的实例。参照图 7,示出了发夹衔接子 701 (SEQ ID NO:2),其可用作如本公开内容的其他地方所述的叉状衔接子的前体。在该实施例中,使用 SPIA 将发夹衔接子合成为单链扩增产物。发夹衔接子 701 包含双链区域 702、用于 AT 连接的 3' -T 突出端 703 和可被限制酶切割的区域 704 (即,在位置 33 与 34 之间)。发夹衔接子可包含条形码区域和功能区域,诸如固定区域和用于测序引物退火的区域。

[0240] 对衔接子的切割 (例如,在位置 33 与 34 之间) 产生了图 8a 中示出的叉状衔接子 (SEQ ID NO:3-4)。通过引入与待切割的区域互补的寡核苷酸序列并将退火的衔接子暴露于限制酶来切割衔接子。图 8a 中示出的叉状衔接子区域与靶多核苷酸的连接产生了图 8b 中示出的结构 (SEQ ID NO:5-6)。参照图 8,图 8b 中的序列的加下划线部分包含具有 3' -A 突出端的靶多核苷酸,该 3' -A 突出端对于与图 8a 中示出的叉状衔接子的连接相容。

[0241] 然后通过聚合酶链反应扩增图 8b 中所示的序列 (SEQ ID NO:5-6) 以产生图 8c 中所示的 SEQ ID NO:7 (SEQ ID NO:5 的扩增产物) 和 SEQ ID NO:8 (SEQ ID NO:6 的扩增产物)。在图 8c 中,SEQ ID NO:7 代表 SEQ ID NO:5 的扩增产物,其将第一固定序列 (加下划线的 5' 部分) 和第二固定序列 (加下划线的 3' 部分) 添加至 SEQ ID NO:5。SEQ ID NO:8 代表 SEQ ID NO:6 的扩增产物,其用不同的序列 (加下划线的 3' 部分和加下划线的 5' 部分) 替换 SEQ ID NO:6 的未杂交部分。此外,SEQ ID NO:8 在多核苷酸的 5' 未杂交区域内包含六个核苷酸的条形码 (TAGTGC;加粗的)。因此扩增产物包含条形码化的靶多核苷酸序列 (由 111 表示)、固定序列和条形码。

实施例 5:通过单引物等温扩增 (SPIA) 和限制消化产生另外的叉状衔接子

[0242] 本实施例示出了通过 SPIA 和限制消化合成如图 9a 中所示的叉状衔接子 (SEQ ID NO:9-10), 其中 N 代表 A、T、G 或 C。图 9b 显示了单链形式的叉状衔接子 (SEQ ID NO:11), 其中该单链形式能够形成发夹结构。在星号指示的位置处切割发夹结构产生图 9a 中所示的叉状衔接子。

[0243] 用于 SPIA 的模板将是图 9c 中所示的序列 (SEQ ID NO:12)。在图 9c 中,“R”代表 RNA 区域。图 9d 示出了由图 9c 中的序列形成的发夹结构。用聚合酶处理图 9d 中的序列 (SEQ ID NO:12) 以将核苷酸添加至 3' 末端, 从而产生图 9e 中所示的序列 (SEQ ID NO:13)。然后用 RNA 酶 H 处理图 9e 中的序列 (SEQ ID NO:13), 该 RNA 酶 H 降解与 DNA 杂交的 RNA, 从而产生图 9f 中的序列 (SEQ ID NO:14)。

[0244] 然后在 SEQ ID NO:14 上进行链置换 SPIA。链置换扩增中的引物具有 RRRRRRRRRRRR (即, R_{13}) 的形式。该引物为 RNA 引物, 其比 SEQ ID NO:14 (即, N_{12}) (图 9f) 的未杂交的 3' 末端长一个碱基。更具体而言, 如图 9f 中所示, SEQ ID NO:14 的 3' 末端含有十二个 N 核苷酸。该 RNA 引物含有 13 个核苷酸。该 RNA 引物的核苷酸 2-13 与 SEQ ID NO:14 的十二个未杂交的 N 核苷酸互补。该 RNA 引物的核苷酸 1 与第一杂交的碱基 (从 3' 至 5') (在这种情况下为 T) 互补。该 RNA 引物置换 A 并产生图 9g 中所示的双链延伸产物 (SEQ ID NO:15-16)。由于仅存在一条引物, 因此该反应产生单链产物的多个拷贝。用 RNA 酶 H 处理单链扩增产物以产生图 9h 中所示的单链扩增产物 (SEQ ID NO:17)。图 9i 示出了该序列的 5' -3' 形式 (SEQ ID NO:17)。图 9j 示出了该序列的发夹形式 (SEQ ID NO:17)。

[0245] 然后将图 9j 中示出的发夹衔接子连接至具有 3' -A 突出端的片段化的多核苷酸。通过添加与该区域互补的寡核苷酸并用限制酶切割, 在图 9j 中由曲线分开的 A 与 C 残基之间切割发夹。这产生叉状衔接子。然后如实施例 4 中所述进行 PCR 扩增, 以将固定区域和条形码添加到附接至靶多核苷酸的叉状衔接子上。

实施例 6:通过指数 PCR 和杂交产生包含条形码的叉状衔接子

[0246] 本实施例示出了通过杂交产生包含条形码的叉状衔接子。图 10a 示出了图 8a 中提供的示例性叉状衔接子。如实施例 4 中所述, 可将该衔接子连接至靶多核苷酸, 随后可进行扩增反应以添加另外的功能序列, 包括条形码。然而, 也可在将叉状衔接子附接至靶多核苷酸之前, 将条形码 (和其他功能序列) 直接掺入到叉状衔接子中。例如, 图 10b 示出了添加有第一固定区域 (加下划线的) 和七个核苷酸的条形码区域 (加粗 / 加下划线的; “N”) 的图 10a 的叉状衔接子。

[0247] 图 10b 的条形码化的叉状衔接子如下产生: 首先合成作为单链的 SEQ ID NO:18。如本公开内容通篇所述, 使用 A、G、T 和 C 的等摩尔混合物产生条形码区域中的多样性。如实施例 1 中所述进行基于小滴的 PCR。然而, 使用一条 DNA 引物和一条 RNA 引物来扩增小滴中的 SEQ ID NO:18。如实施例 1 中所述, 在嵌入染料的存在下进行扩增, 并且分离包含扩增的 SEQ ID NO:18 的小滴。图 10c 示出了双链扩增产物。SEQ ID NO:19 的加下划线部分是来源于 RNA 引物的 RNA 链。然后用 RNA 酶 H 处理图 10c 中所示的序列, 该 RNA 酶 H 消化加下划线的 RNA 区域, 产生图 10d 中所示的构建体。为产生叉状构建体, 将部分互补通用序列 (SEQ ID NO:21) 添加到图 10d 中所示的构建体中, 从而产生图 10e 中所示的产物。采用该方法的优点在于, 相比于由 SPIA 提供的多核苷酸的线性扩增, 其使用了由指数 PCR 提供的

显著更高水平的多核苷酸扩增。

实施例 7:双索引方法

[0248] 本实施例示出了合成用于双索引读取的条形码的方法。双索引读取是使用附接至每条链的条形码对双链片段的两条链的读取。图 11 示出了用于双索引方法的条形码的合成实例,以及在胶囊中的胶囊的配置中使用条形码的实例。

[0249] 如图 11a 中所示,合成包含第一固定区域 1102、第一条形码区域 1103 和第一测序引物区域 1104 的第一单链衔接子-条形码多核苷酸序列 1101。平行地,如图 11b 中所示,合成包含第二固定区域 1132、第二条形码区域 1133 和第二测序引物区域 1134 的第二单链衔接子-条形码多核苷酸序列 1131。在一些情况下,条形码区域 1103 和 1133 具有相同的序列。在其他情况下,条形码区域 1103 和 1133 具有不同的序列或部分不同的序列。

[0250] 合成后,平行地将单链衔接子-条形码多核苷酸 1101(图 11a)和 1131(图 11b)稀释到油包水乳液中的水性小滴中。该小滴还包含用于通过不对称 PCR(例如,聚合酶、引物、dNTP、缓冲液、盐)和 DNA 嵌入染料(例如,溴化乙锭)分别扩增单链衔接子-条形码多核苷酸 1101(图 11a)和 1131(图 11b)的试剂。反向引物以多于正向引物的量存在,或反之亦然,从而使得不对称扩增能够进行。对多核苷酸 1101(图 11a)和 1131(图 11b)进行扩增,并且该反应继续进行分别通过指数期扩增 1110(产生双链产物 1105(图 11a)和 1135(图 11b))以及线性期扩增 1111(产生单链产物 1106(图 11a)和 1136(图 11b))。

[0251] 将小滴在荧光辅助细胞分选仪(FACS)1112 上进行分选以收集包含扩增的多核苷酸的小滴。

[0252] 然后分别在包含单链衔接子-条形码多核苷酸序列 1106 和 1136 的小滴上进行界面聚合,以产生两种类型的胶囊 1120(图 11a)和 1150(图 11b),每一种胶囊分别包含单链衔接子-条形码多核苷酸序列 1106 或 1136 中的一个。

[0253] 制备两种混合物。混合物 Z1 包含靶多核苷酸(即,待片段化并加条形码的多核苷酸)1170 和片段化酶(例如,NEBNext DSDNA FRAGMENTASE)。第二混合物 Z2 包含如上所述产生的胶囊 1120 和 1180,以及足以活化片段化酶的浓度的氯化镁。混合物 Z1、Z2,或者 Z1 和 Z2 两者还包含 T4 聚合酶、Taq 聚合酶和热稳定连接酶。

[0254] 将混合物 Z1 和 Z2 合并,并且根据本公开内容的其他地方所述的方法,如流动聚焦,形成在胶囊内的胶囊。图 11c 示出了根据上述方法产生的在胶囊内的胶囊。外层胶囊 1160 包含胶囊 1120 和 1150 以及介质 1190。因此,胶囊 1120 和 1150 各自分别包含单链衔接子-条形码多核苷酸 1106 和 1136 的多个拷贝,并可用于将条形码 1103 和 1133 附接至分区内的多核苷酸,诸如在外层胶囊 1160 的介质 1190 中的靶多核苷酸 1170。

[0255] 介质 1190 含有上述混合物 Z1 和混合物 Z2 的内容物。更具体而言,介质 1190 包含靶多核苷酸 1170 和包含片段化酶、T4 聚合酶、Taq 聚合酶、热稳定连接酶、氯化镁和合适的缓冲液的酶混合物 1180。

[0256] 在产生胶囊内的胶囊并使胶囊内的胶囊暴露于合适的条件后,酶处理靶多核苷酸。更具体而言,片段化酶对靶多核苷酸进行片段化并且 T4 聚合酶使片段化的靶多核苷酸的末端成为平端。然后热灭活片段化酶和 T4 聚合酶,并使用刺激使胶囊 1120 和 1150 破裂,从而释放其内容物至外层胶囊 1160 的介质 1190 中。Taq 聚合酶向片段化的、平端化的靶多核苷酸上添加 3'-A 突出端。单链衔接子-条形码多核苷酸 1106 与单链衔接子-条形

码多核苷酸 1136 杂交,从而形成包含条形码区域 1103 和 1133 的叉状衔接子,该叉状衔接子具有与在片段化的靶多核苷酸上的 3' -A 突出端(未示出)相容的 3' -T 突出端。热稳定连接酶将叉状衔接子连接到片段化的靶多核苷酸,从而产生条形码化的靶多核苷酸。然后使外层胶囊 1160 破裂,将来自所有外层胶囊的样品合并,并对靶多核苷酸进行测序。

[0257] 此外,尽管上述实施例使用热稳定连接酶将包含条形码序列的叉状衔接子连接到靶多核苷酸,但如本公开内容的其他地方所述,也可使用 PCR 来完成该步骤。

实施例 8:通过珠子乳液 PCR 和添加部分互补通用序列产生包含条形码序列的叉状衔接子

[0258] 如图 14a 所示,合成包含第一固定区域 1402、条形码区域 1403 和第一测序引物区域 1404 的单链衔接子-条形码序列 1401。合成后,将单链衔接子-条形码序列 1401 稀释到油包水乳液中的水性小滴中,使得每个小滴平均包含 1 个多核苷酸。该小滴还包含通过光不稳定连接体连接至 RNA 引物 1406 的一个或多个拷贝的第一珠子 1405,该 RNA 引物 1406 与包含在第一测序引物区域 1404 中的序列互补;与包含在第一固定区域 1402 中的序列(未示出)互补的 DNA 引物;以及对于单链衔接子-条形码序列 1401 的扩增所必需的试剂(例如,聚合酶、dNTP、缓冲液、盐)。对该多核苷酸进行扩增 1407,产生均附接至第一珠子 1405 上的双链产物 1408,以形成结构 1420 并且在溶液中(未示出)。

[0259] 然后破乳,并将乳液组分合并以形成产物混合物。如图 14b 所示,然后将释放的珠子用合适的介质洗涤 1409(通过离心)数次,用氢氧化钠(NaOH)处理 1410 以使附接至第一珠子 1405 上的双链产物变性,随后进一步进行洗涤 1411。在对结构 1420 进行变性 1410 和洗涤 1411 后,所得结构 1430 包含单链衔接子-条形码序列 1401 的单链互补序列 1412,该单链互补序列 1412 包含互补固定区域 1413、互补条形码区域 1414 和互补测序引物区域 1415。如所示,互补测序引物区域 1415 包含 RNA 引物 1406。然后将结构 1430 重悬浮于合适的介质中。

[0260] 接着,如图 14c 中所示,将包含与互补固定区域 1413 互补的 DNA 多核苷酸 1417 的一个或多个拷贝的第二珠子 1416 添加至介质中。通过互补 DNA 多核苷酸 1417 和单链互补序列 1412 的互补固定区域 1413,第二珠子 1416 与单链互补序列 1412 结合。该单链互补序列现在在一个末端与第一珠子 1405 结合,而在其另一末端与第二珠子 1416 结合,以形成结构 1440。

[0261] 如图 14d 所示,然后使用甘油梯度对结构 1440 进行离心 1418 以将结构 1440 与不包含在结构 1440 中的结构 1430 分离。在第二珠子 1416 为磁性的情况下,可使用磁性分离作为替代手段。然后用 NaOH 处理 1419 产物以使来自第二珠子 1416 的单链互补序列 1412 变性,从而导致结构 1430 再生。然后使结构 1430 经历数轮洗涤(通过离心),以去除第二珠子 1416。附接至结构 1430 的单链互补序列 1412 代表单链条形码衔接子。

[0262] 如图 14e 所示,可使用单链互补序列 1412 产生叉状衔接子。为产生叉状衔接子 1450,然后用光使单链互补序列 1412 从结构 1430 中释放 1424,随后与通用互补序列 1426 组合 1425,或先与通用互补序列 1426 组合 1425 随后从结构 1430 中释放 1424。为产生可连接的末端,使用 RNA 酶 H 来消化单链互补序列 1412 的 RNA 引物 1406,并使用 IIs 型限制酶在通用互补序列 1426 上产生单碱基 T 突出端。该 T 突出端与在待测序的多核苷酸靶标上的 A 突出端(未示出)相容。

实施例 9 :通过珠子乳液 PCR 和添加部分互补通用序列产生包含条形码序列的叉状衔接子

[0263] 如图 15a 所示,合成包含第一固定区域 1502、条形码区域 1503 和第一测序引物区域 1504 的单链衔接子 - 条形码序列 1501。合成后,将单链衔接子 - 条形码序列 1501 稀释到油包水乳液中的水性小滴中,使得每个小滴平均包含 1 个多核苷酸。该小滴还包含通过光不稳定连接体连接至 RNA 引物 1506 的一个或多个拷贝的第一珠子 1505,该 RNA 引物 1506 与包含在第一固定区域 1502 中的序列互补;与包含在第一测序引物区域 1504 中的序列(未示出)互补的 DNA 引物;以及对于单链衔接子 - 条形码序列 1501 的扩增所必需的试剂(例如,聚合酶、dNTP、缓冲液、盐)。对该多核苷酸进行扩增 1507,产生均附接至第一珠子 1505 上的双链产物 1508,以形成结构 1520 并且在溶液中(未示出)。

[0264] 然后破乳,并将乳液组分合并以形成产物混合物。如图 15b 所示,然后将释放的珠子用合适的介质洗涤 1509(通过离心)数次,用氢氧化钠(NaOH)处理 1510 以使附接至第一珠子 1505 上的双链产物变性,随后进一步洗涤 1511。在对结构 1520 进行变性 1510 和洗涤 1511 后,所得结构 1430 包含单链衔接子 - 条形码序列 1501 的单链互补序列 1512,该单链互补序列 1512 包含互补固定区域 1513、互补条形码区域 1514 和互补测序引物区域 1515。如所示,互补测序引物区域 1515 包含 RNA 引物 1506。然后将结构 1530 重悬浮于合适的介质中。

[0265] 接着,如图 15c 中所示,将包含与互补测序引物区域 1515 互补的 DNA 多核苷酸 1517 的一个或多个拷贝的第二珠子 1516 添加至介质中。通过互补 DNA 多核苷酸 1517 和单链互补序列 1512 的互补测序引物区域 1515,第二珠子 1416 与单链互补序列 1512 结合。该单链互补序列现在在一个末端与第一珠子 1505 结合,而在其另一末端与第二珠子 1516 结合,以形成结构 1540。

[0266] 如图 15d 所示,然后使用甘油梯度对结构 1540 进行离心 1518 以将结构 1540 与不包含在结构 1540 中的结构 1530 分离。在第二珠子 1516 为磁性的情况下,可使用磁性分离作为替代手段。然后用 NaOH 处理 1519 产物以使来自第二珠子 1516 的单链互补序列 1512 变性,从而导致结构 1530 再生。然后使结构 1530 经历数轮洗涤(通过离心),以去除第二珠子 1516。附接至结构 1530 的单链互补序列 1512 代表单链条形码衔接子。

[0267] 如图 15e 所示,可使用单链互补序列 1512 产生叉状衔接子。为产生叉状衔接子 1550,然后任选地用光使单链互补序列 1512 从结构 1530 中释放 1524,随后与通用互补序列 1526 组合 1525。为产生可连接的末端,使用 IIs 型限制酶在通用互补序列 1526 上产生单碱基 T 突出端。该 T 突出端与在待测序的多核苷酸靶标上的 A 突出端(未示出)相容。

实施例 10 :通过珠子乳液 PCR 产生叉状衔接子模板条形码序列以及由其衍生的衔接子

[0268] 如图 16 所示,根据实施例 8、实施例 9 中所述的方法,或本文所述的任何其他方法,产生包含与磁珠(1601)结合的单链衔接子 - 条形码序列 1602 的结构 1600。接着,通过本文所述的方法,例如,界面聚合,将结构 1600 分区到胶囊(或可替代地,另一种乳液)1620 中。胶囊 1620 还包含用于通过不对称 PCR 扩增单链衔接子 - 条形码序列 1602 的试剂(例如,聚合酶、引物、dNTP、缓冲液、盐)。反向引物以多于正向引物的量存在,或反之亦然,从而使得不对称扩增能够进行。对单链衔接子 - 条形码序列 1602 进行扩增 1603,并且该反应继续进行通过线性期扩增 1604,这产生与单链条形码衔接子 - 模板 1602 互补的单链衍

接子产物 1605。在此时,胶囊 1620 既包含溶液中的单链衔接子 1605 又包含与磁珠 (1601) 结合的单链衔接子-条形码序列 1602。然后通过磁性分离 1606 将胶囊 1620 与不包含珠子 (并因此不包含模板 1602 和单链衔接子 1605) 的胶囊分离。可以使胶囊 1620 破裂并且如实施例 9 中所述产生叉状衔接子。

实施例 11:用珠子乳液 PCR 加条形码以及用片段化酶进行片段化

[0269] 如图 17 所示,根据实施例 8、实施例 9 中所述的方法,或本文所述的任何其他方法,产生包含与磁珠 (1701) 结合的单链衔接子-条形码序列 1702 的结构 1700。在包含结构 1700 的小滴上进行界面聚合,以产生包含通过光不稳定连接体附接至珠子 1701 上的单链衔接子-条形码序列 1702 的胶囊 1704。

[0270] 制备两种混合物。混合物 Z1 包含靶多核苷酸 (即,待片段化并加条形码的多核苷酸)、片段化酶 (例如,NEBNext DSDNA FRAGMENTASE) 和部分互补通用序列。第二混合物 Z2 包含如上所述产生的胶囊 1704,和足以活化片段化酶的浓度的氯化镁。混合物 Z1、Z2,或者 Z1 和 Z2 两者还包含 T4 聚合酶、Taq 聚合酶和热稳定连接酶。

[0271] 将混合物 Z1 和 Z2 合并,并且根据本公开内容的其他地方所述的方法,如流动聚焦,形成在胶囊内的胶囊。图 17 示出了根据上述方法产生的在胶囊内的胶囊。外层胶囊 1703 包含内层胶囊 1704 和介质 1705。内层胶囊 1704 是包封的、与珠子结合的单链条形码衔接子的文库中的一员。因此,内层胶囊 1704 包含结构 1700 的多个拷贝,结构 1700 可用于产生游离的单链衔接子-条形码序列 1702 并将相同的条形码衔接子附接至分区如外层胶囊 1703 内的多核苷酸。

[0272] 介质 1705 含有上述混合物 Z1 和混合物 Z2 的内容物。更具体而言,介质 1705 包含靶多核苷酸 1706、部分互补通用序列 1707 以及包含片段化酶、T4 聚合酶、Taq 聚合酶、热稳定连接酶、氯化镁和合适的缓冲液的酶混合物 1708。

[0273] 在产生胶囊内的胶囊并使胶囊内的胶囊暴露于合适的条件后,酶处理靶多核苷酸。更具体而言,片段化酶对靶多核苷酸进行片段化并且 T4 聚合酶使得片段化的靶多核苷酸的末端成为平端。然后热灭活片段化酶和 T4 聚合酶,并使用刺激使内层胶囊 1704 破裂,释放其内容物至外层胶囊 1703 中。Taq 聚合酶向片段化的、平端化的靶多核苷酸上添加 3'-A 突出端。单链衔接子-条形码序列 1702 与部分互补通用序列 1707 杂交,并通过光从珠子上释放,从而形成具有与在片段化的靶多核苷酸上的 3'-A 突出端相容的 3'-T 突出端的叉状衔接子。热稳定连接酶将叉状衔接子连接到片段化的靶多核苷酸,从而产生条形码化的靶多核苷酸。然后使外层胶囊 1703 破裂,将来自所有外层胶囊的样品合并,并对靶多核苷酸进行测序。在测序前,根据需要可进行另外的准备步骤 (例如,本体扩增、大小选择等)。

[0274] 在一些情况下,Z1 可包含部分互补通用序列 1707 的多个版本。此外,尽管该实施例示出了通过使用热稳定连接酶对靶多核苷酸加条形码,但也可以使用 PCR 完成该步骤。

实施例 12:用珠子乳液 PCR 加条形码以及通过超声处理进行片段化

[0275] 如图 18 所示,根据实施例 8、实施例 9 中所述的方法,或本文所述的任何其他方法,产生包含与磁珠 (1801) 结合的单链衔接子-条形码序列 1802 的结构 1800。在包含结构 1800 的小滴上进行界面聚合,以产生包含通过光不稳定连接体附接至珠子 1801 上的单链衔接子-条形码序列 1802 的胶囊 1803。将靶多核苷酸 (即,待片段化的多核苷酸) 分区到

胶囊 1804 中。将包含靶多核苷酸的胶囊 1804 配置成能抵抗超声应力。使包含靶多核苷酸的胶囊 1804 暴露于超声应力（例如，COVARIS 聚焦超声发生器），并对靶多核苷酸进行片段化，从而产生片段化的靶多核苷酸胶囊。

[0276] 制备混合物 Z1，其包含胶囊 1803、片段化的靶多核苷酸胶囊 1804、部分互补通用序列 1805、酶混合物（T4 聚合酶、Taq 聚合酶和热稳定连接酶）1806 和合适的缓冲液。根据本公开内容的其他地方所述的方法，如流动聚焦，产生在胶囊内的胶囊。

[0277] 图 18 示出了根据上述方法产生的在胶囊内的胶囊。外层胶囊 1807 包含胶囊 1803 和 1804 以及介质 1808。内层胶囊 1803 和 1804 分别包括包含结构 1800 的胶囊和包含片段化的靶多核苷酸 1809 的胶囊。内层胶囊 1803 包含结构 1800 的多个拷贝，结构 1800 可用于产生游离的单链条形码衔接子 1802 并将相同的条形码衔接子附接至分区内的多核苷酸，诸如包含在内层胶囊 1804 内的片段化的多核苷酸 1809。

[0278] 介质 1808 含有上述混合物 Z1 的内容物。更具体而言，介质 1808 包含部分互补通用序列 1805、酶混合物（T4 聚合酶、Taq 聚合酶和热稳定连接酶）1806 和合适的缓冲液。

[0279] 使包含片段化的靶多核苷酸 1809 的内层胶囊 1804 暴露于刺激，以使其破裂并释放其内容物至外层胶囊 1807 的内容物中。T4 聚合酶使得片段化的靶多核苷酸的末端成为平端；Taq 聚合酶向片段化的、平端化的靶多核苷酸上添加 3'-A 突出端。然后热灭活 T4 聚合酶和 Taq 聚合酶，并且施加刺激以将内层胶囊 1803 的内容物释放至外层胶囊 1807 中。单链衔接子 - 条形码序列 1802 与部分互补通用序列 1805 杂交，并且采用光使衔接子从珠子上释放，形成具有与在片段化的靶多核苷酸上的 3'-A 突出端相容的 3'-T 突出端的叉状衔接子。热稳定连接酶将该叉状衔接子连接至片段化的靶多核苷酸，从而产生条形码化的靶多核苷酸。然后使外层胶囊 1807 破裂，将来自所有外层胶囊的样品合并，并对靶多核苷酸进行测序。

[0280] 在一些情况下，Z1 可包含部分互补通用序列 1807 的多个版本。此外，尽管该实施例示出了通过使用热稳定连接酶对靶多核苷酸加条形码，但也可以使用 PCR 完成该步骤。

实施例 13：用基于退火和成环的多重扩增（MALBAC）进行的条形码化

[0281] 如图 19a 所示，制备包含 SEQ ID No. 22 的引物。该引物包含条形码区域（表示为“条形码”）、引物测序区域（命名为“引发序列”）和可能包含 A、T、C 或 G 的任何组合的八个核苷酸的可变区域（表示为“NNNNNNNN”）。将图 19 中所示的引物与靶多核苷酸（由图 19 中的环表示）连同具有链置换活性的聚合酶（例如，Vent、exo+DeepVent、exo - DeepVent）一起组合到分区（例如，胶囊、乳液的小滴等）内。在一些情况下，使用非链置换聚合酶（例如，Taq、PfuUltra）。然后对该分区进行 MALBAC 扩增。合适的 MALBAC 循环条件是已知的，并在例如，Zong 等人，Science, 338(6114), 1622-1626 (2012) 中描述，该文献通过引用以其全文并入本文。

[0282] 产生如图 19b 中所示的 SEQ ID No. 23 的环状 MALBAC 产物。环状 MALBAC 产物包含图 19a 中所示的原始引物、将在环中条形码化定向的靶多核苷酸以及与原始引物序列互补并杂交的区域。该分区被破坏并且回收内容物。在一些情况下，产生多个分区。这些分区被全体破坏，回收每一个分区的内容物，随后将其合并。

[0283] 接着，用限制酶（例如，BfuCl 或相似的酶）处理图 19b 中示出的所产生的 MALBAC 产物以在 MALBAC 产物上产生 4- 碱基对的突出端（在这种情况下，为以斜体显示的 GATC）。

该结构由 SEQ ID No. 24 表示,并在图 19c 中示出。图 19d 中显示的 SEQ ID No. 25 的叉状衔接子包含与在 MALBAC 产物上产生的突出端互补的突出端(在这种情况下,为以斜体显示的 CTAG)。叉状衔接子与图 19c 中的 MALBAC 产物混合,并且这些互补区域杂交。使用热稳定连接酶将叉状衔接子和 MALBAC 产物连接在一起,以形成图 19e 中 SEQ ID No. 26 的所需结构。可使用另外的扩增方法(例如,PCR)将另外的区域(例如,固定区域、另外的条形码等)添加至叉状衔接子。

[0284] 在一些情况下,可能期望其他碱基对突出端(例如,1 个碱基对的突出端-10 个碱基对的突出端)。需要时,用于产生这些突出端的限制酶可用作替代物,包括本文中所述的那些。在一个实例中,使用 Taq[®] I 在 MALBAC 产物上产生两个碱基对的突出端。

[0285] 作为备选,可设计图 19a 中所示的引物,使得 RNA 引物序列置于条形码区域的 5',以便使用 RNA 酶产生突出端。如图 19f 中所示,MALBAC 产物 1900 包含置于条形码区域 1902 的 5'的 RNA 引物序列 1901。MALBAC 产物 1900 还包含测序引物区域 1903、靶多核苷酸 1904、互补测序引物区域 1905、互补条形码区域 1906 和与 RNA 引物序列 1901 互补的区域 1907。用 RNA 酶 H 1908 处理 MALBAC 产物 1900,并且消化 RNA 引物区域序列 1901 以在 MALBAC 产物 1900 上产生 2-6 碱基对的突出端 1909 以生成结构 1920。然后将通用互补区域 1910 添加至包含与结构 1910 上的突出端互补的区域的区域 1910。然后通用互补区域 1910 与结构 1920 杂交,并使用热稳定连接酶将通用互补区域 1910 与结构 1920 连接。

实施例 14:用基于退火和成环的多重扩增(MALBAC)进行的条形码化

[0286] 如图 20 所示,使用例如界面聚合或本文所述的任何其他方法将包含条形码区域的模板 2000 与对于 PCR 所必需的试剂 2001 组合到胶囊 2002 内。使用 PCR 由模板 2000 产生 MALBAC 引物。接着,将胶囊 2000 包封到还包含混合物 2004 的外层胶囊 2003 中,该混合物 2004 包含待条形码化的靶多核苷酸 2005 和对于 MALBAC 扩增所必需的试剂 2006(例如,DeepVent 聚合酶、dNTP、缓冲液)。在胶囊 2002 恰当地暴露于被设计用于破坏胶囊 2002 的刺激时,胶囊 2002 被破坏,胶囊 2002 的内容物与混合物 2004 的内容物混合。靶多核苷酸 2005 的 MALBAC 扩增开始,产生与图 19f 中示为 1900 的所述产物相似的 MALBAC 产物。

[0287] 然后用合适的刺激破坏外层胶囊 2003,并回收内容物。然后用合适的限制酶处理 MALBAC 产物,并以如实施例 13 中所述的方式偶联到叉状衔接子上。然后根据需要进行另外的下游准备步骤(例如,本体扩增、大小选择等)。

实施例 15:用基于退火和成环的多重扩增(MALBAC)进行的条形码化

[0288] 如图 21a 所示,制备 MALBAC 引物 2100。MALBAC 引物 2100 包含序列引发区域 2101 和 8-核苷酸的可变区域 1902。引物 2100 与靶多核苷酸 2103 连同具有链置换活性的聚合酶(例如,Vent、exo+DeepVent、exo-DeepVent)一起组合到分区(例如,胶囊、乳液等)内。在一些情况下,使用非链置换聚合酶(例如,Taq、PfuUltra)。然后对该分区进行 MALBAC 扩增 2104。

[0289] 产生环状 MALBAC 产物 2110,其包含测序引发区域 2101、靶多核苷酸 2103 和互补序列引发区域 2105。然后使在图 21b 中以线性形式 2120 显示的 MALBAC 产物 2110 与包含测序引物区域 2106、条形码区域 2107 和固定区域 2108 的另一条引物 2130 接触。使用不对称数字 PCR 产生引物 2130。通过使用 PCR 单循环,用引物产生包含引物 2130 并因此包含条形码区域 2107 的双链产物 2140。

[0290] 然后可使双链产物 2140 变性,随后与图 21c 中所示的另一引物 2150 接触。引物 2150 包含条形码区域 2109、测序引物区域 2111 和固定区域 2112。在引物 2113 和 2114 的存在下,另外数轮 PCR 可将条形码区域 2109 添加至附接至条形码区域 2107 的靶多核苷酸的末端。随后根据需要进行另外的下游准备步骤(例如,本体扩增、大小选择等)。

实施例 16:用基于退火和成环的多重扩增(MALBAC)进行的条形码化

[0291] 如图 22 所示,使用例如界面聚合或本文所述的任何其他方法将包含条形码区域的引物模板 2200 与对于 PCR 所必需的试剂 2201 组合到胶囊 2202 内。然后使用 PCR 由模板 2200 产生引物。接着,将胶囊 2200 包封到还包含混合物 2204 的外层胶囊 2003 中,该混合物 2004 包含待条形码化的靶多核苷酸 2205、对于 MALBAC 扩增所必需的试剂 2206(例如,DeepVent 聚合酶、dNTP、缓冲液)和不包含条形码的 MALBAC 引物 2207(与实施例 15 中所述的 MALBAC 引物 2100 相似)。靶多核苷酸 2205 的 MALBAC 扩增开始,产生与图 21a 中示为 2110 的所述产物相似的 MALBAC 产物。随后在胶囊 2202 恰当地暴露于被设计用于破坏胶囊 2202 的刺激时,胶囊 2202 被破坏,胶囊 2202 的内容物与混合物 2004 的内容物混合。使用由模板 2200 产生的引物开始 PCR 单循环,产生与实施例 15 中所述的产物相似的条形码化的产物。

[0292] 然后用合适的刺激破坏外层胶囊 2203,并回收内容物。然后根据需要进行另外的下游准备步骤(例如,本体扩增、大小选择、附加条形码的添加等)。

实施例 17:用转座酶和 Tagmentation 进行的条形码化

[0293] 如图 23 所示,如实施例 1 中所述,或通过本公开内容中描述的任何其他方法,合成、分区、扩增并分选单链衔接子-条形码多核苷酸序列 2300。在包含单链衔接子-条形码多核苷酸序列的小滴上进行界面聚合,以产生胶囊 2301。

[0294] 制备两种混合物。混合物 Z1 包含靶多核苷酸 2302(即,待片段化并加条形码的多核苷酸)、转座酶 2303 和部分互补通用序列 2304。第二混合物 Z2 包含如上所述产生的胶囊 2301,和如本文其他地方所述的对于 PCR 所必需的试剂 2305。

[0295] 将混合物 Z1 和 Z2 合并,并且根据本公开内容的其他地方所述的方法,如流动聚焦,形成在胶囊内的胶囊。图 23 示出了根据上述方法产生的在胶囊内的胶囊。外层胶囊 2306 包含胶囊 2301 和介质 2307。胶囊 2301 是包封的单链衔接子-条形码多核苷酸的文库中的一员。因此,胶囊 2301 包含单链衔接子-条形码多核苷酸 2300 的多个拷贝,后者可用于将相同条形码附接至分区如外层胶囊 2306 内的多核苷酸。

[0296] 介质 2307 含有上述混合物 Z1 和混合物 Z2 的内容物。更具体而言,介质 2307 包含靶多核苷酸 2302、部分互补通用序列 2304 和对于 PCR 所必需的试剂 2305,包括热启动 Taq。

[0297] 在产生在胶囊内的胶囊并使在胶囊内的胶囊暴露于合适的条件后,转座酶处理靶多核苷酸。更具体而言,转座酶通过 tagmentation 对靶多核苷酸进行片段化,并用共同引发序列对其进行标记。然后将标记的靶多核苷酸加热以补平由转座酶产生的靶核苷酸中的任何缺口。然后热灭活转座酶,并使用刺激以使内层胶囊 2301 破裂,释放其内容物至外层胶囊 2306 中。通过加热外层胶囊 2306 至 95℃活化热启动 Taq。该反应继续进行有限循环 PCR 以将单链衔接子-条形码多核苷酸序列 2300 添加至靶多核苷酸 2302。然后使外层胶囊 2306 破裂,并对靶多核苷酸进行测序。

[0298] 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施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。

[0001]

<110> 10X 基因组学有限公司

<120> 多核苷酸条形码生成

<130> 43487-706, 601

<140> PCT/US2014/015427

<141> 2014-02-07

<150> 61/844, 804

<151> 2013-07-10

<150> 61/840, 403

<151> 2013-06-27

<150> 61/800, 223

<151> 2013-03-15

<150> 61/762, 435

<151> 2013-02-08

<160> 37

<170> PatentIn version 3.5

<210> 1

<211> 77

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成寡核苷酸

<400> 1

aatgatacgg cgaccaccga gatctacact agatcgcaca ctctttccct acaacgacgt 60

cttccgatct gatctaa 77

<210> 2

<211> 66

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述: 合成寡核苷酸

<400> 2

[0002]

gatcggaaga gcacacgtct gaactccagt cacacactct ticcetacac gacgctcttc	60
cgatct	66
<210> 3	
<211> 33	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<400> 3	
acactctttc cctacacgac gctcttccga tet	33
<210> 4	
<211> 33	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<400> 4	
gatcggaaga gcacacgtct gaactccagt cac	33
<210> 5	
<211> 67	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<220>	
<221> misc feature	
<222> (33)..(34)	
<223> 这些位置的核苷酸被未公开的条形码化的靶多核苷酸序列隔开	
<400> 5	
acactctttc cctacacgac gctcttccga tetagatcgg aagagcacac gtctgaactc	60
cagtcac	67
<210> 6	
<211> 67	
<212> DNA	

[0003]

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> misc_feature

<222> (33)..(34)

<223> 这些位置的核苷酸被未公开的条形码化的靶多核苷酸序列隔开

<400> 6

acaactctttc cctacacgac gctcttccga tctagatcgg aagagcacac gtctgaactc 60

cagtcac

67

<210> 7

<211> 122

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成多核苷酸

<220>

<221> misc_feature

<222> (58)..(59)

<223> 这些位置的核苷酸被未公开的条形码化的靶多核苷酸序列隔开

<400> 7

aatgatacgg cgaccacega gatciacact ctctccctac aegacgcict ccgaactag 60

atcggaagag cacacgtctg aactccagtc acatcagat ctcgtaigcc gtcttctgt 120

tg

122

<210> 8

<211> 122

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成多核苷酸

<220>

<221> misc_feature

<222> (64)..(65)

<223> 这些位置的核苷酸被未公开的条形码化的靶多核苷酸序列隔开

[0004]

<400> 8
caagcagaag acggcatacg agatcgtgat gtgactggag ttcagaecgtg tgcctcttcg 60
atctagatcg gaagagcgtc gtgtagggaa agagtgtaga tctcggtggt cgcctatca 120
tt 122

<210> 9
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述: 合成寡核苷酸

<220>
<221> 修饰的_碱基
<222> (34).. (41)
<223> a、c、t 或 g

<400> 9
acactctttc cctacacgac gctcttcgga tetnnnnnnn nt 42

<210> 10
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述: 合成寡核苷酸

<220>
<221> 修饰的_碱基
<222> (1).. (8)
<223> a、c、t 或 g

<400> 10
nnnnnnnag atcggaagag cacacgtctg aactecagtc ac 42

<210> 11
<211> 84
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述: 合成寡核苷酸

[0005]

<220>	
<221> 修饰的_碱基	
<222> (1)..(8)	
<223> a、c、t 或 g	
<220>	
<221> 修饰的_碱基	
<222> (76)..(83)	
<223> a、c、t 或 g	
<400> 11	
nnnnnnnnag atcgaagag cacacgtctg aactccagtc acacaetctt tccctacacg	60
acgetcttcc gatctnnnnn nnnt	84
<210> 12	
<211> 88	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述: 合成寡核苷酸	
<220>	
<223> 组合的 DNA/RNA 分子的描述: 合成寡核苷酸	
<220>	
<221> 修饰的_碱基	
<222> (1)..(12)	
<223> a、c、u 或 g	
<220>	
<221> 修饰的_碱基	
<222> (14)..(21)	
<223> a、c、t 或 g	
<400> 12	
nnnnnnnnnn nnannnnnnn nagatcggaa gagcgtcgtg tagggaaaga gtgtgtgact	60
ggagttcaga cgtgtgctct tccgatct	88
<210> 13	
<211> 109	
<212> DNA	
<213> 人工序列	
<220>	

[0006]

<223> 人工序列的描述：合成寡核苷酸

<220>

<223> 组合的 DNA/RNA 分子的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (1).. (12)

<223> a、c、u 或 g

<220>

<221> 修饰的_碱基

<222> (14).. (21)

<223> a、c、t 或 g

<220>

<221> 修饰的_碱基

<222> (89).. (96)

<223> a、c、t 或 g

<220>

<221> 修饰的_碱基

<222> (98).. (109)

<223> a、c、t 或 g

<400> 13

nnnnnnnnnn nnannnnnnn nagatcggaa gagcgtcgtg tagggaaaga gtgtgtgact 60

ggagttcaga cgtgtgetct tccgatctnn nnnnnntnnn nnnnnnnnnn 109

<210> 14

<211> 97

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (2).. (9)

<223> a、c、t 或 g

<220>

<221> 修饰的_碱基

<222> (77).. (84)

<223> a、c、t 或 g

<220>

[0007]

<221> 修饰的_碱基

<222> (86).. (97)

<223> a、c、t 或 g

<400> 14

annnnnnnna gatcggaaga gcgtcgtgta gggaaagagt gtgtgactgg agttcagacg 60

tgtgctcttc cgatctnnnn nnnntnnnnn nnnnnnn 97

<210> 15

<211> 96

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (2).. (9)

<223> a、c、t 或 g

<220>

<221> 修饰的_碱基

<222> (76).. (83)

<223> a、c、t 或 g

<220>

<221> 修饰的_碱基

<222> (85).. (96)

<223> a、c、t 或 g

<400> 15

annnnnnmng atcggaagag cacacgtctg aaetecagtc acacaetctt teectacaag 60

acgtctcttc gatctnnnn nnnntnnnnn nnnnnnn 96

<210> 16

<211> 96

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<223> 组合的 DNA/RNA 分子的描述：合成寡核苷酸

[0008]

<220>
<221> 修饰的_碱基
<222> (1)..(13)
<223> a、c、u 或 g

<220>
<221> 修饰的_碱基
<222> (14)..(21)
<223> a、c、t 或 g

<220>
<221> 修饰的_碱基
<222> (88)..(95)
<223> a、c、t 或 g

<400> 16
nnnnnnnnnn nnnnnnnnnn nagatcggaa gagcgtcgig tagggaaaga gtgtgtgact 60

ggagttcaga cgtgtgctct tccgatennn nnnmnt 96

<210> 17
<211> 83
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述: 合成寡核苷酸

<220>
<221> 修饰的_碱基
<222> (1)..(8)
<223> a、c、t 或 g

<220>
<221> 修饰的_碱基
<222> (75)..(82)
<223> a、c、t 或 g

<400> 17
nnnnnnnnnag atcggaagag cgtcgtgtag ggaaagagtg tgtgactgga gttcagacgt 60

gtgctcttcc gatennnnnnn nnt 83

<210> 18
<211> 70
<212> DNA
<213> 人工序列

<220>

[0009]

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (30).. (36)

<223> a、c、t 或 g

<400> 18

aatgatacgg cgaccaccga gatctacacn nnnnnnacac tctttcccta cagcagctc 60

tccgatctt 70

<210> 19

<211> 70

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (35).. (41)

<223> a、c、t 或 g

<400> 19

aagatcggaa gacgctcgtg tagggaaaga gtgtnnnnnn ngtagatc tcggtggtc 60

ccgtatcatt 70

<210> 20

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (21).. (27)

<223> a、c、t 或 g

<400> 20

gtcgtgtagg gaaagagtgt nnnnnnngtg tagatctcgg tggtcgccgt atcatt 56

[0010]

<210> 21	
<211> 34	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<400> 21	
agatcggaag agcacacgtc tgaactccag tcac	34
<210> 22	
<211> 73	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<400> 22	
ttagatcaga tcggaagagc acacgtctga actccagtc ctaaggcgaa tctcgtatgc	60
cgtcttctgc ttg	73
<210> 23	
<211> 36	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<220>	
<221> misc_feature	
<222> (18).. (19)	
<223> 这些位置的核苷酸被未公开的条形码和引物测序区域隔开	
<400> 23	
acgacgctct tccgatctag atcggaagag cgtcgt	36
<210> 24	
<211> 6	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	

[0011]

<220>	
<221> misc_feature	
<222> (5)..(6)	
<223> 这些位置的核苷酸被未公开的条形码和引物测序区域隔开	
<400> 24	
gatcta	6
<210> 25	
<211> 33	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<400> 25	
gateggaaga gcacacgtct gaactccagt cac	33
<210> 26	
<211> 67	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<220>	
<221> misc feature	
<222> (33)..(34)	
<223> 这些位置的核苷酸被未公开的条形码和引物测序区域隔开	
<400> 26	
acactctttc cctacacgac gctcttcga tctagatcgg aagagcacac gtctgaacte	60
cagtcac	67
<210> 27	
<211> 13	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工序列的描述：合成寡核苷酸	
<220>	

[0012]

⟨221⟩ 修饰的_碱基	
⟨222⟩ (5)..(9)	
⟨223⟩ a、c、t、g、未知的或其他	
⟨400⟩ 27	
ggccnnnnng gcc	13
⟨210⟩ 28	
⟨211⟩ 12	
⟨212⟩ DNA	
⟨213⟩ 人工序列	
⟨220⟩	
⟨223⟩ 人工序列的描述：合成寡核苷酸	
⟨220⟩	
⟨221⟩ 修饰的_碱基	
⟨222⟩ (4)..(8)	
⟨223⟩ a、c、t、g、未知的或其他	
⟨400⟩ 28	
caannnnngt gg	12
⟨210⟩ 29	
⟨211⟩ 14	
⟨212⟩ DNA	
⟨213⟩ 人工序列	
⟨220⟩	
⟨223⟩ 人工序列的描述：合成寡核苷酸	
⟨220⟩	
⟨221⟩ 修饰的_碱基	
⟨222⟩ (4)..(10)	
⟨223⟩ a、c、t、g、未知的或其他	
⟨400⟩ 29	
gaannnnnnn ttgg	14
⟨210⟩ 30	
⟨211⟩ 13	
⟨212⟩ DNA	
⟨213⟩ 人工序列	
⟨220⟩	
⟨223⟩ 人工序列的描述：合成寡核苷酸	

[0013]

<220> <221> 修饰的_碱基 <222> (5)..(10) <223> a、c、t、g、未知的或其他 <400> 30 gaacnnnnnn tcc	13
<210> 31 <211> 13 <212> DNA <213> 人工序列 <220> <223> 人工序列的描述：合成寡核苷酸	
<220> <221> 修饰的_碱基 <222> (5)..(10) <223> a、c、t、g、未知的或其他 <400> 31 gaagnnnnnn tac	13
<210> 32 <211> 12 <212> DNA <213> 人工序列 <220> <223> 人工序列的描述：合成寡核苷酸	
<220> <221> 修饰的_碱基 <222> (5)..(9) <223> a、c、t、g、未知的或其他 <400> 32 gaacnnnnnc tc	12
<210> 33 <211> 13 <212> DNA <213> 人工序列 <220>	

[0014]

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的_碱基

<222> (5).. (10)

<223> a、c、t、g、未知的或其他

<400> 33

gaacnnnnnn tac

13

<210> 34

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述：合成肽

<400> 34

Gly Phe Leu Gly Lys

1 5

<210> 35

<211> 153

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成多核苷酸

<220>

<221> 修饰的_碱基

<222> (78).. (80)

<223> a、c、t、g、未知的或其他

<400> 35

aatgatacgg cgaccaccga gatctacact agatcgacac ctcttttcct acacgacgt 60

cttcgatct gatctaannt ttagatcaga tcggaagagc acacgtcga aciccagtca 120

ctaaggcgaa tctcgtatgc cgtttttctg ttg 153

<210> 36

<211> 26

<212> DNA

<213> 人工序列

[0015]

<220>

<223> 人工序列的描述：合成寡核苷酸

<220>

<221> misc_feature

<222> (18).. (19)

<223> 这些位置的核苷酸被未公开的条形码和引物测序区域隔开

<220>

<221> 修饰的_碱基

<222> (19).. (26)

<223> a、c、t 或 g

<400> 36

acgacgctet tccgatctnn nnnnnn

26

<210> 37

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 37

acaactcttcc cctacacgac gctcttcc

28

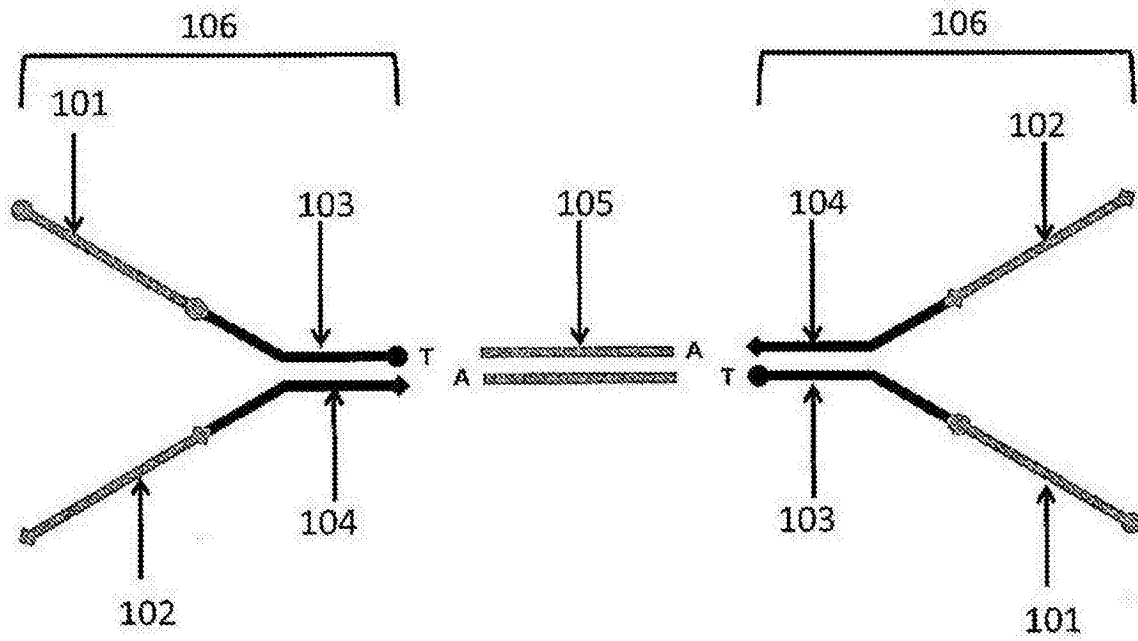


图 1

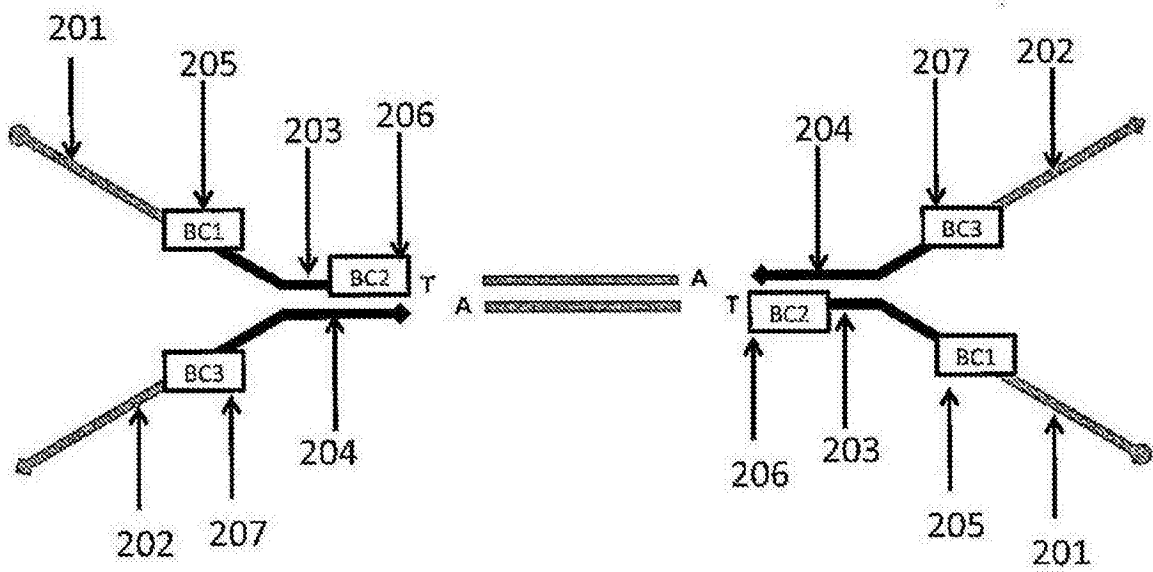


图 2

5' -
AATGATACGG CGACCACCGA GATCTACACT 30
AGATCGCACA CTCTTTCCCT ACACGACGCT 60
CTTCCGATCT **GATCTAA** SEQ ID NO: 1 77

-NNN- 80

TTAGATCAGA TCGGAAGAGC ACACGTCTGA 110
ACTCCAGTCA **CTAAGGCGAA** TCTCGTATGC 140
CGTCTTCTGC TTG 153
-3' SEQ ID NO: 22

图 3

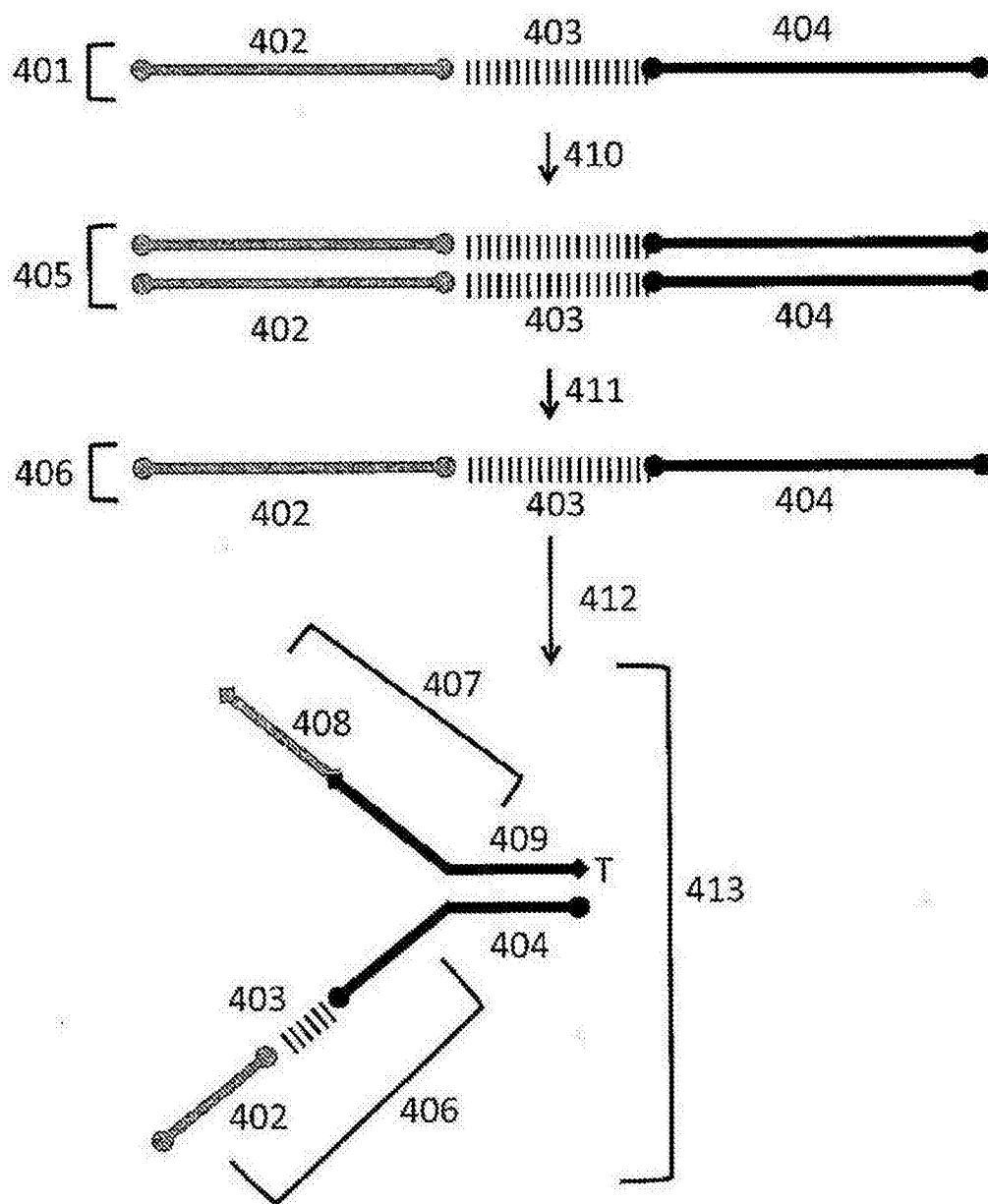


图 4

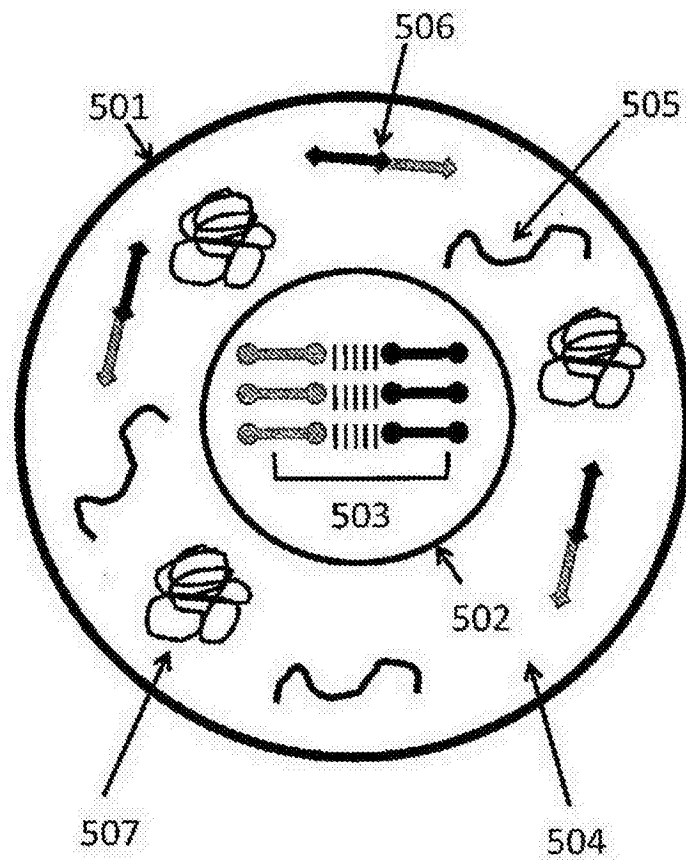


图 5

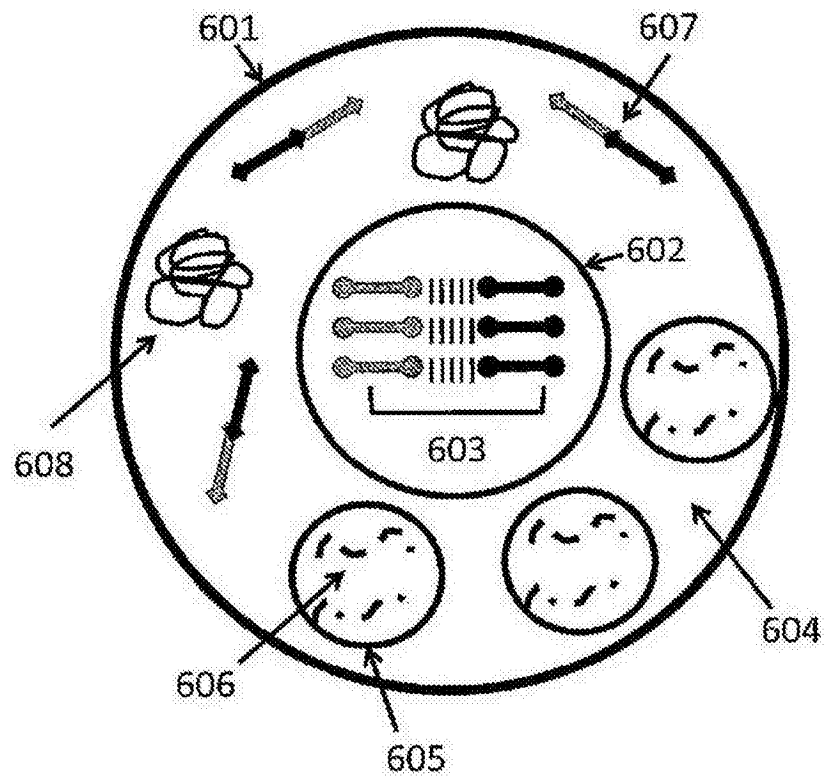


图 6

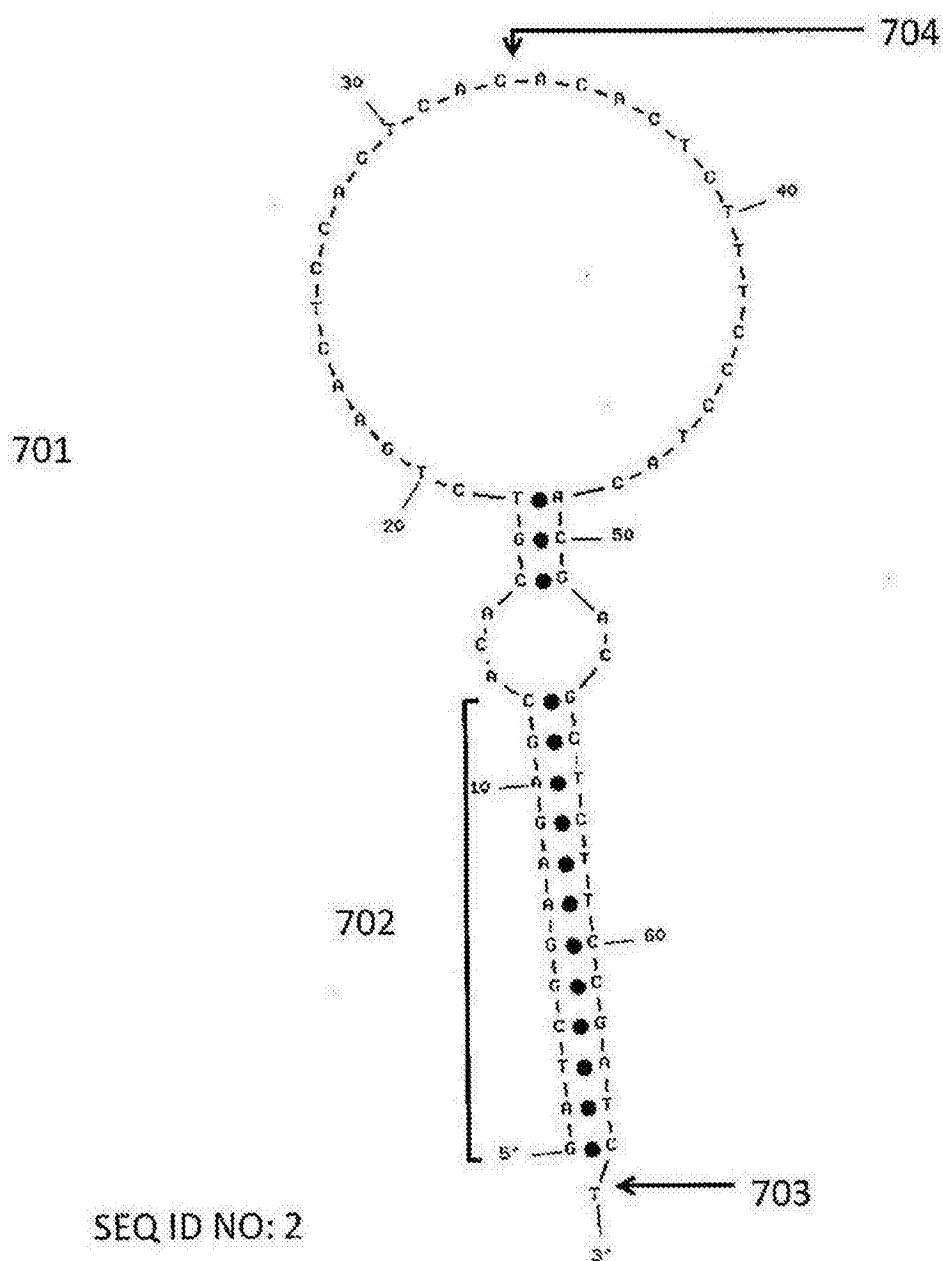


图 7

SEQ ID NO: 3
 5'-ACACTCTTTCCCTACACGAC
 GCTCTTCCGATCT
 CGAGAAGGCTAG
 3'-CACTGACCTCAAGTCTGCACA
 SEQ ID NO: 4

图 8a

SEQ ID NO: 5

5'-ACACTCTTTCCCTACACGAC GCTCTTCCGATCT111AGATCGGAAGAGC ACACGTCTGAACTCCAGTCAC
 CGAGAAGGCTAG111TCTAGCCTTCTCG
 3'-CACTGACCTCAAGTCTGCACA CAGCACATCCCTTCTCACA

SEQ ID NO: 6

图 8b

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT111AGATCGGAAGAGCA
 CACGTCTGAACTCCAGTCACATCAGGATCTCGTATGCCGTCTTCTGCTTG (SEQ ID NO: 7)

3'-
 TTACTATGCCGCTGCTGGCTCTAGATGTGAGAAAGGGATGTGCTGCCAGAAGGCTAG111TCTAGCCTTCTC
 GTGTGCAGACTTCAGGTCAGTGTAGTGCTAGAGCATACGGCAGAAGACGAAC (SEQ ID NO: 8)

图 8c

SEQ ID NO: 9

5'-ACACTCTTTCCCTACACGAC

GCTCTTCCGATCTNNNNNNNNNT

CGAGAAGGCTAGANNNNNNNN

3'-CACTGACCTCAAGTCTGCACA

SEQ ID NO: 10

图 9a

5'-

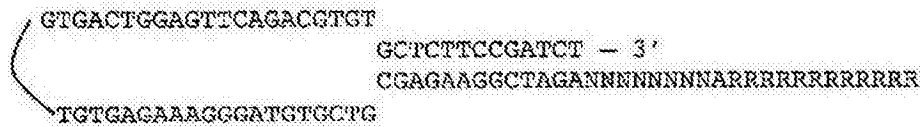
NNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC*ACA
 CTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNT
 (SEQ ID NO: 11)

图 9b

5' -
 RRRRRRRRRRRRRANNNNNNNNAGATCGGAAGAGCGTCGTGTAGGG
 AAAGAGTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
 (SEQ ID NO: 12)

图 9c

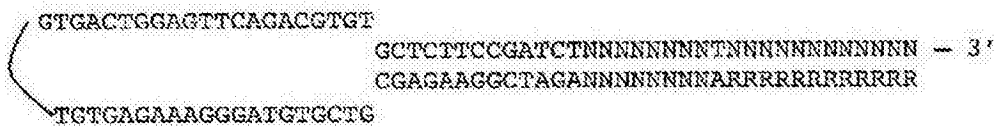
SEQ ID NO: 12



GTGACTGGAGTTCAGACGTGT
 GCTCTTCCGATCT - 3'
 CGAGAAGGCTAGANNNNNNNNARRRRRRRRRRR
 TGTGAGAAAGGGATGTGCTG

图 9d

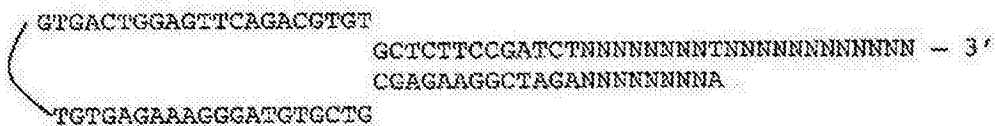
SEQ ID NO: 13



GTGACTGGAGTTCAGACGTGT
 GCTCTTCCGATCTNNNNNNNNNTNNNNNNNNNNNN - 3'
 CGAGAAGGCTAGANNNNNNNNARRRRRRRRRRR
 TGTGAGAAAGGGATGTGCTG

图 9e

SEQ ID NO: 14



GTGACTGGAGTTCAGACGTGT
 GCTCTTCCGATCTNNNNNNNNNTNNNNNNNNNNNN - 3'
 CGAGAAGGCTAGANNNNNNNNA
 TGTGAGAAAGGGATGTGCTG

图 9f

5' ANNNNNNNNGATCGGAAGAGCACACGTCTGAACTCCAGTCACACACTCTTCCCTACACGAGCTCTTCCG
 3' TNNNNNNNNCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTGTGTGAGAAAGGGATGTGCTGCGAGAAGGC

ATCTNNNNNNNNNTNNNNNNNNNNNN-3' (SEQ ID NO: 15)
 TAGANNNNNNNNNRRRRRRRRRRRR-5' (SEQ ID NO: 16)

图 9g

3' TNNNNNNNNCTAGCCTTCTCGTGTCAGACTTGAGGTCAGTGTGTGAGAAAGGGATGTGCTGCCAGAAGGC
TAGANNNNNNNN-5' (SEQ ID NO: 17)

图 9h

5' NNNNNNNNAGATCCGAAGAGCGTCGTGTAGGAAAGAGTGTGTCACTGGACTTCAGACGTGTGCTCTTCCG
ATCNNNNNNNNNT-3' (SEQ ID NO: 17)

图 9i

SEQ ID NO: 17

ACACTCTTTCCCTACACGAC

GCTCTTCCGATCTNNNNNNNNNT - 3'

CGAGAAGGCTAGANNNNNNNNN

CACTGACCTCAAGTCTGCACA

图 9j

SEQ ID NO: 3

5' -ACACTCTTTCCCTACACGAC

GCTCTTCCGATCT

CGAGAAGGCTAG

3' -CACTGACCTCAAGTCTGCACA

SEQ ID NO: 4

图 10a

SEQ ID NO: 18

5' -ATTGATACGGCGACCAACGAGATCTACACNNNNNNNACACTCTTTCCCTACACGAC

GCTCTTCCGATCT

CGAGAAGGCTAG

3' -CACTGACCTCAAGTCTGCACA

SEQ ID NO: 4

图 10b

SEQ ID NO: 18

5' - AATGATACGGCGACCACCGAGATCTACACNNNNNNNACACTCTTCCCTACACGACGCTCTTCCGATCTT - 3'
 3' - TTACTATGCCGCTGGTGGCTCTAGATGTGNNNNNNNTGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGAA - 5'

SEQ ID NO: 19

图 10c

SEQ ID NO: 18

5' AATGATACGGCGACCACCGAGATCTACACNNNNNNNACACTCTTCCCTACACGACGCTCTTCCGATCTT-3'
 3' TTACTATGCCGCTGGTGGCTCTAGATGTGNNNNNNNTGTGAGAAAGGGATGTGCTG-5'

SEQ ID NO: 20

图 10d

SEQ ID NO: 18

5' AATGATACGGCGACCACCGAGATCTACACNNNNNNNACACTCTTCCCTACACGAC GCTCTTCCGATCTT-3'
 3' TTACTATGCCGCTGGTGGCTCTAGATGTGNNNNNNNTGTGAGAAAGGGATGTGCTG CGAGAAGGCTAGA

SEQ ID NO: 20

3' -CACTGACCTCAAGTCTGCACA

SEQ ID NO: 21

图 10e

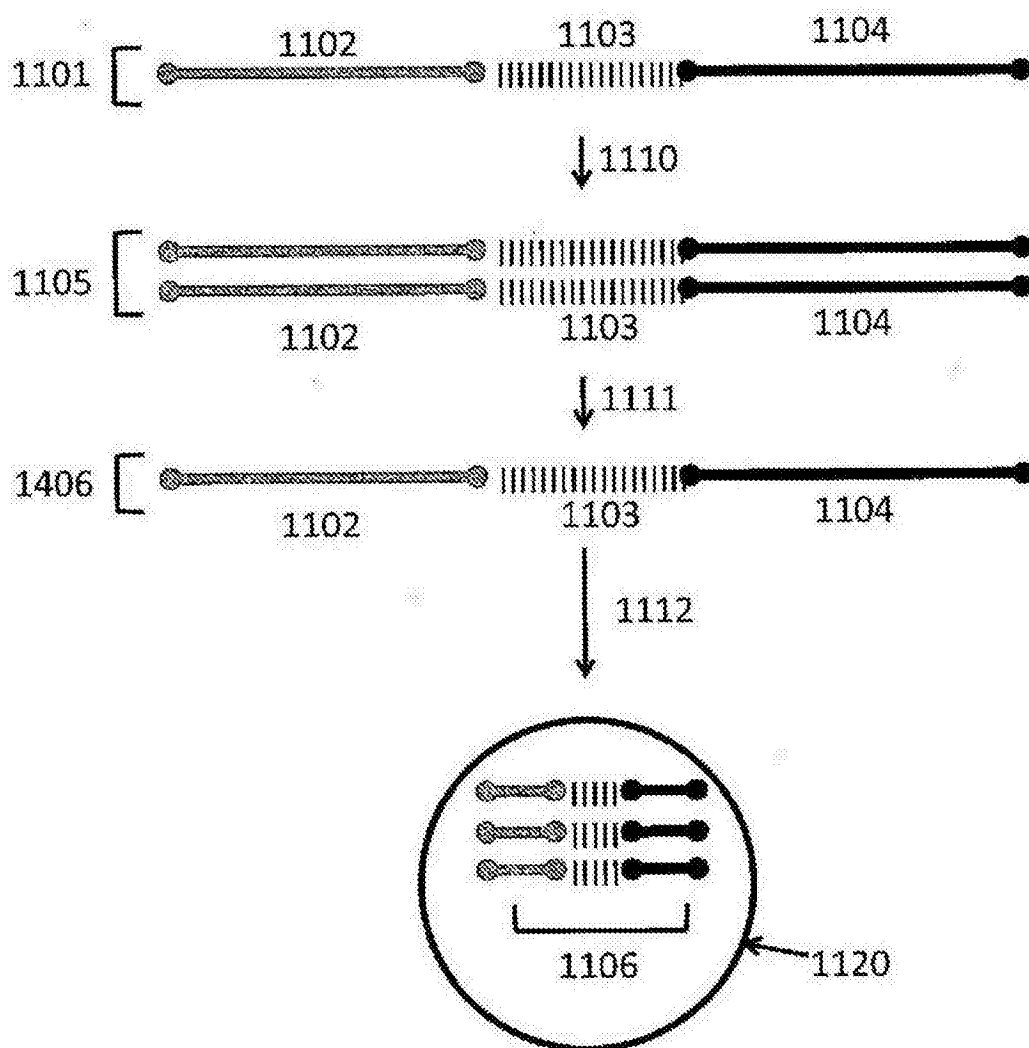


图 11a

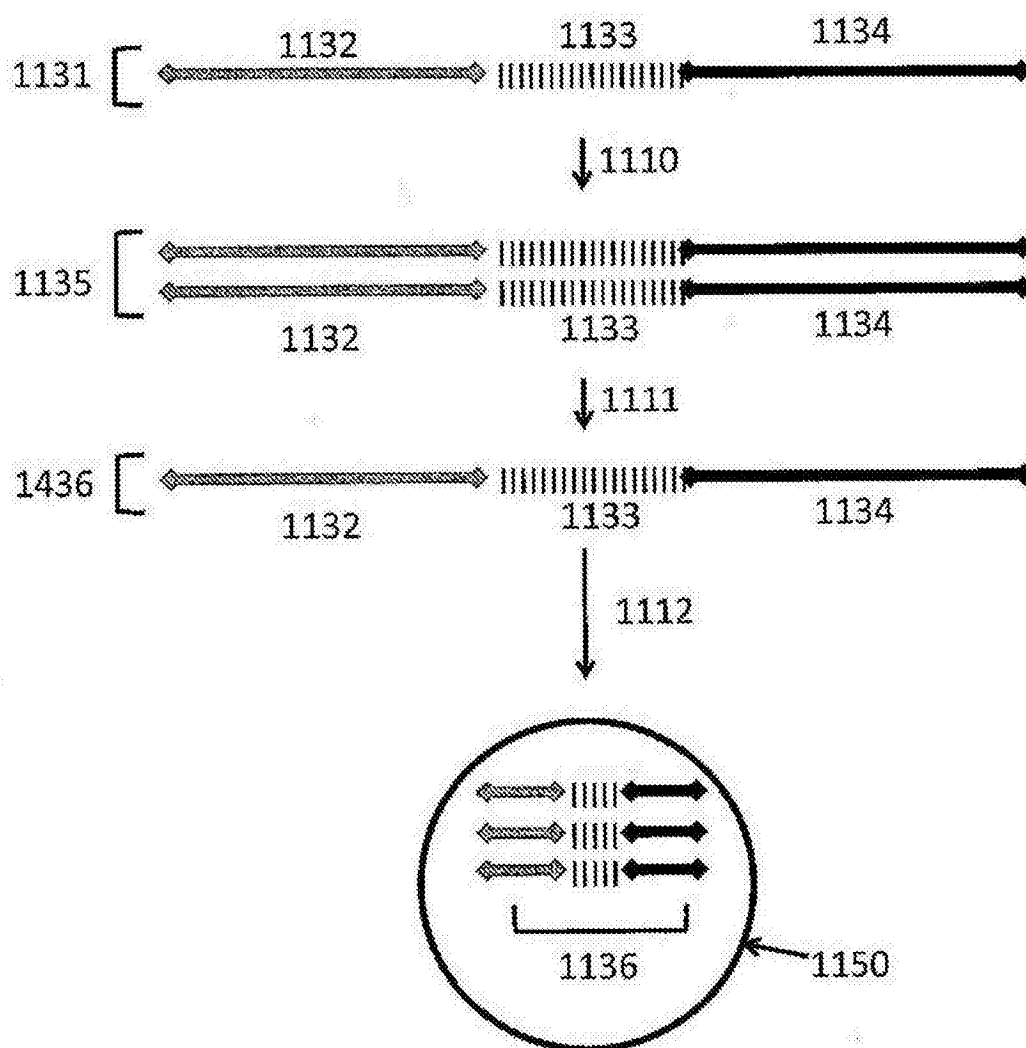


图 11b

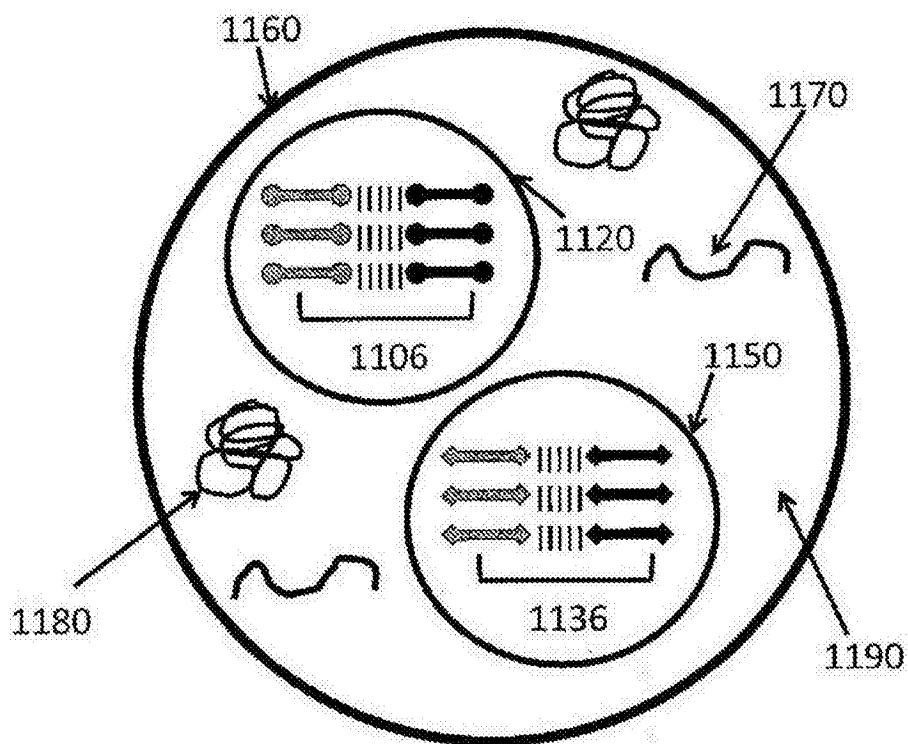


图 11c

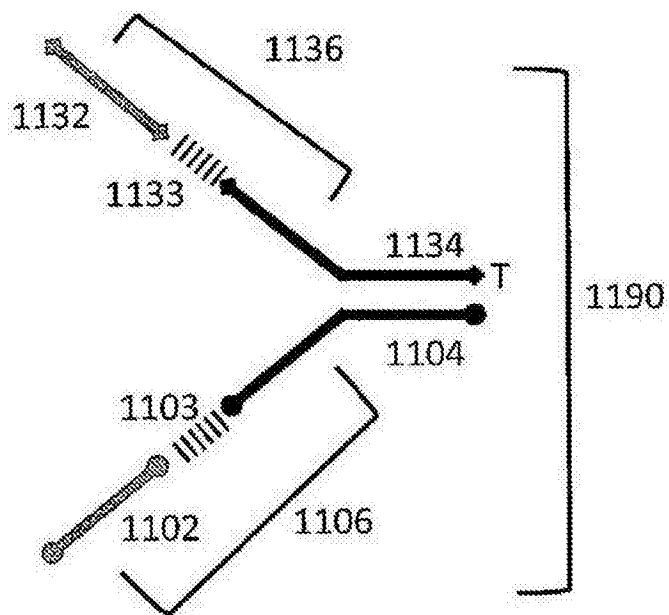


图 11d

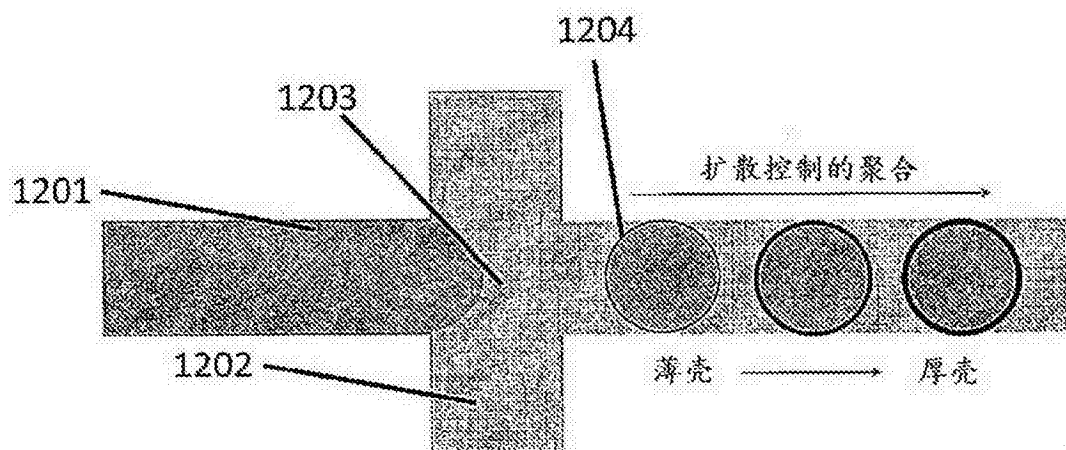


图 12

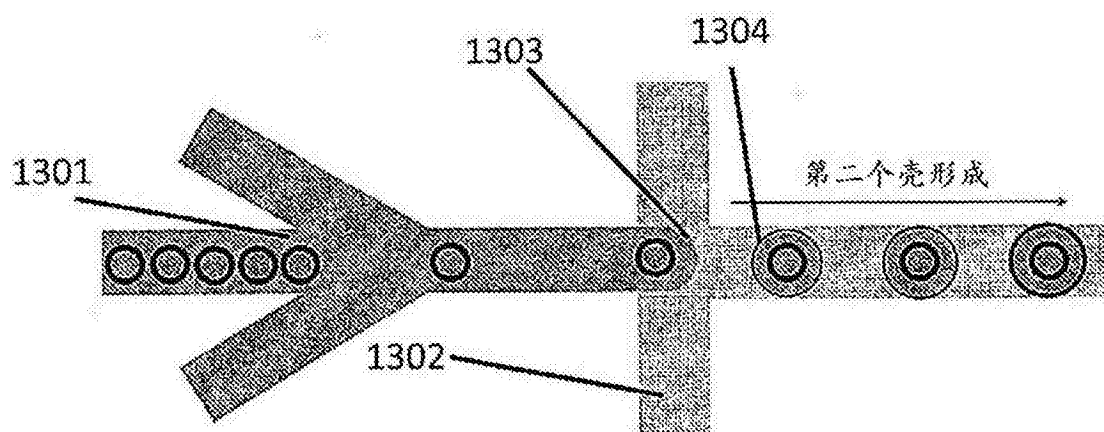


图 13

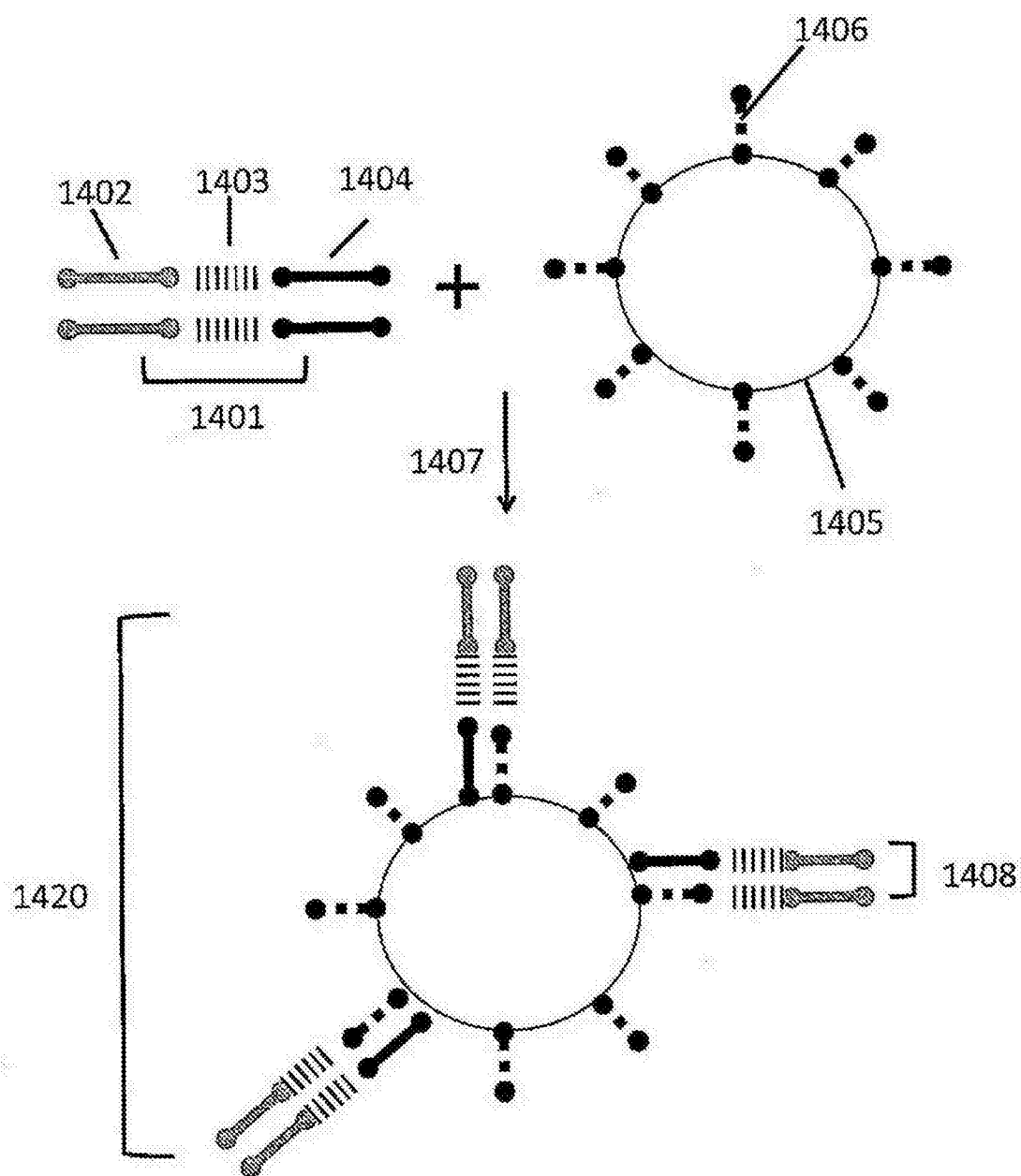


图 14a

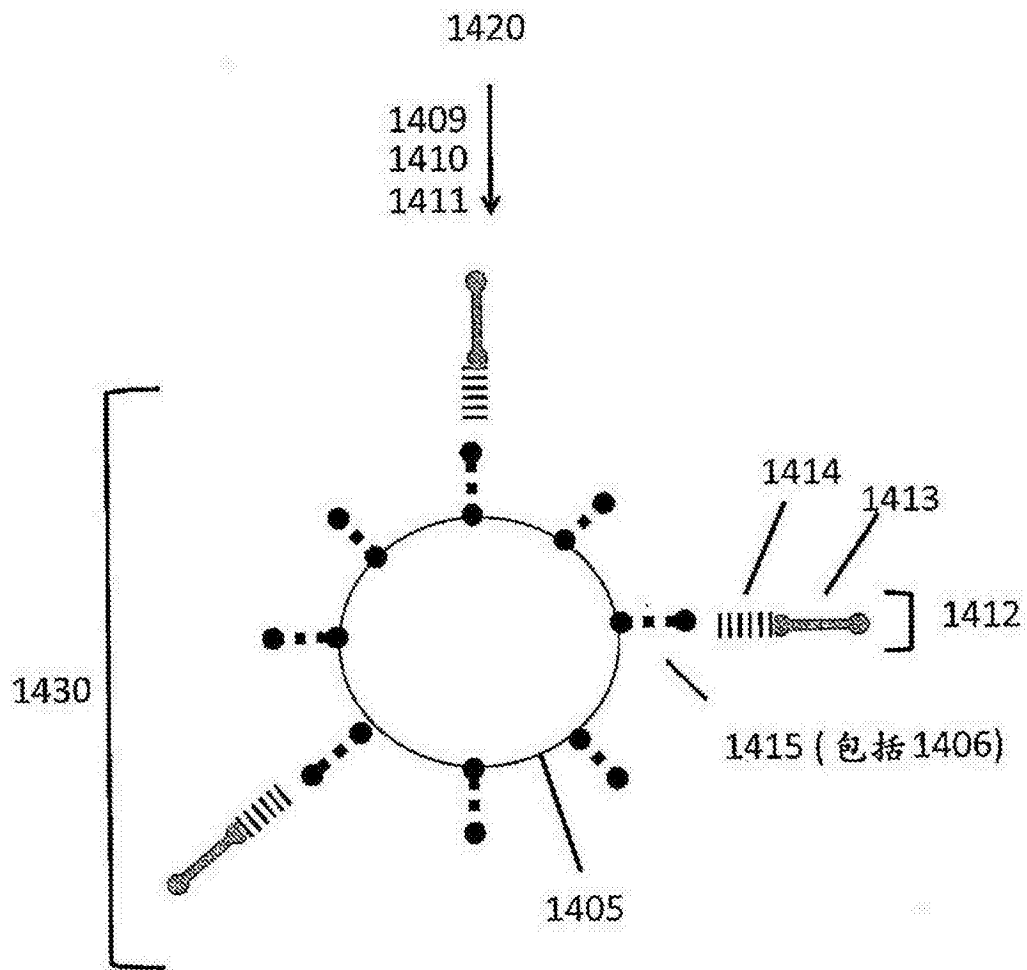


图 14b

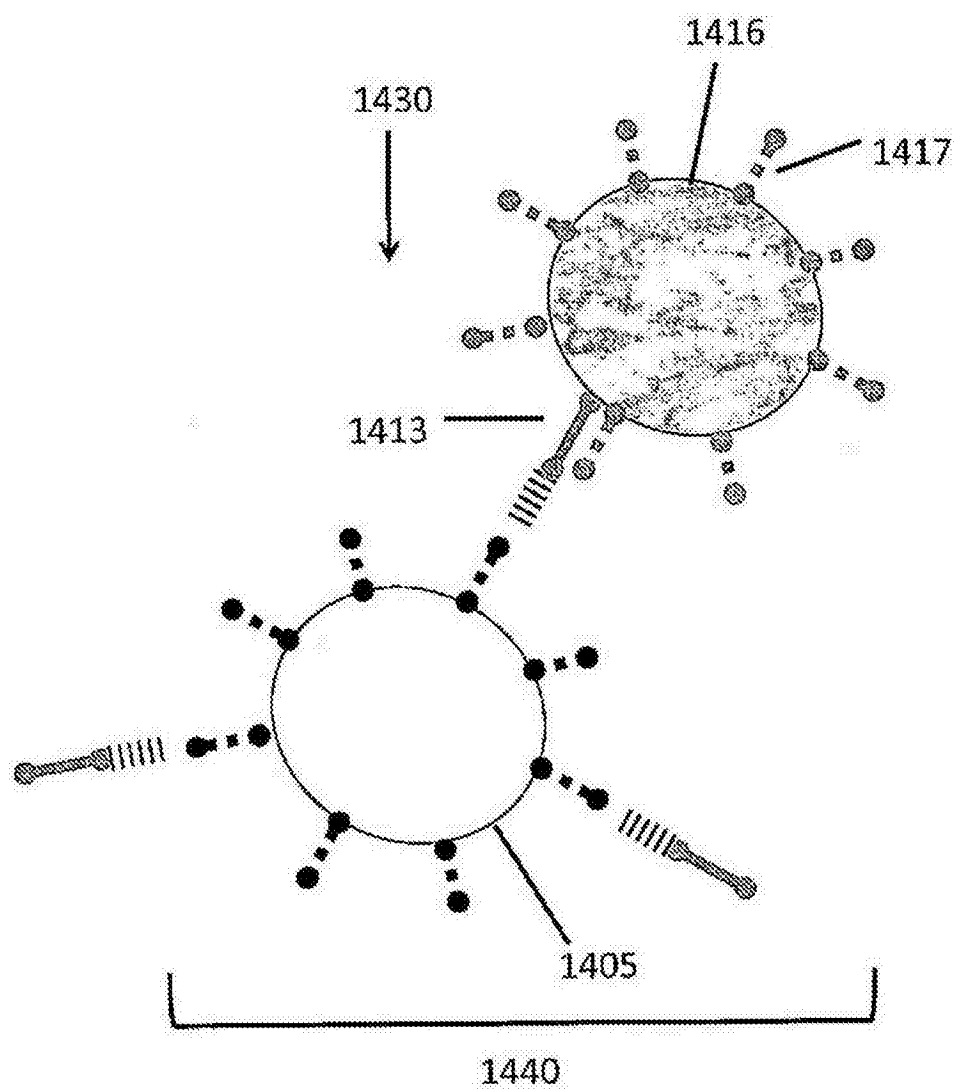


图 14c

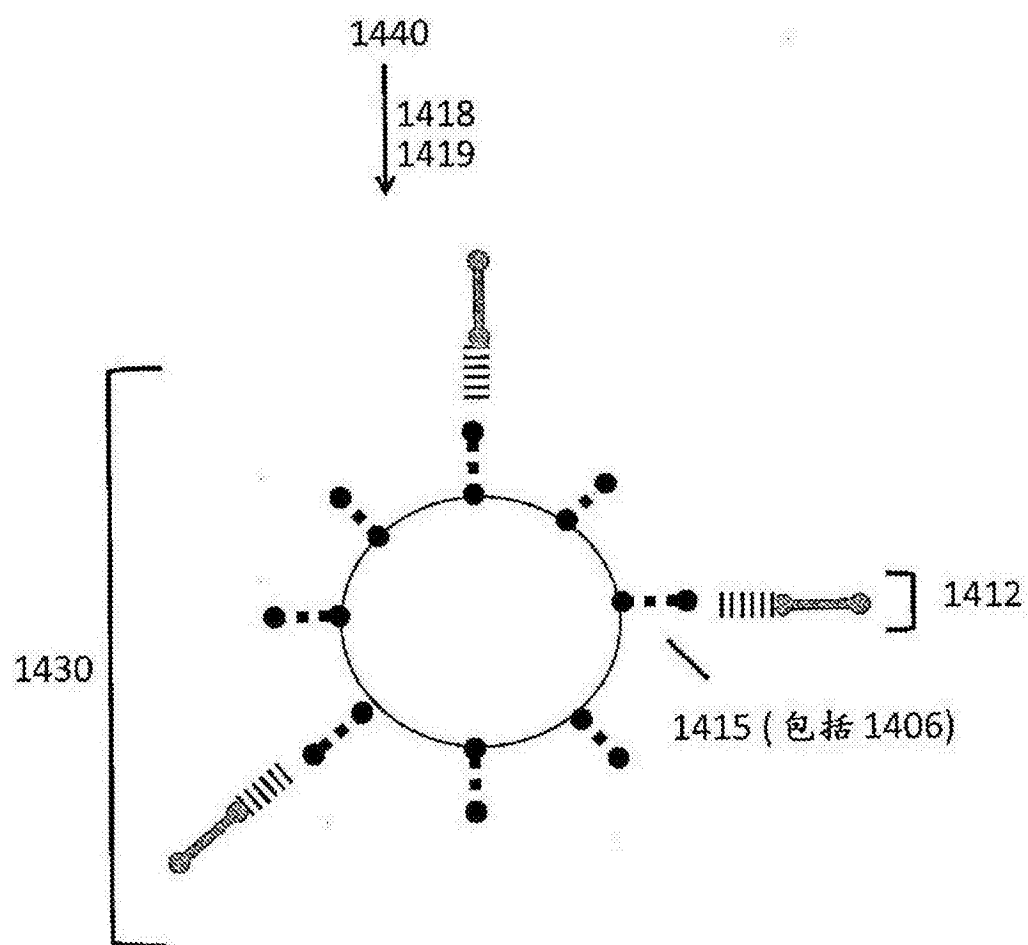


图 14d

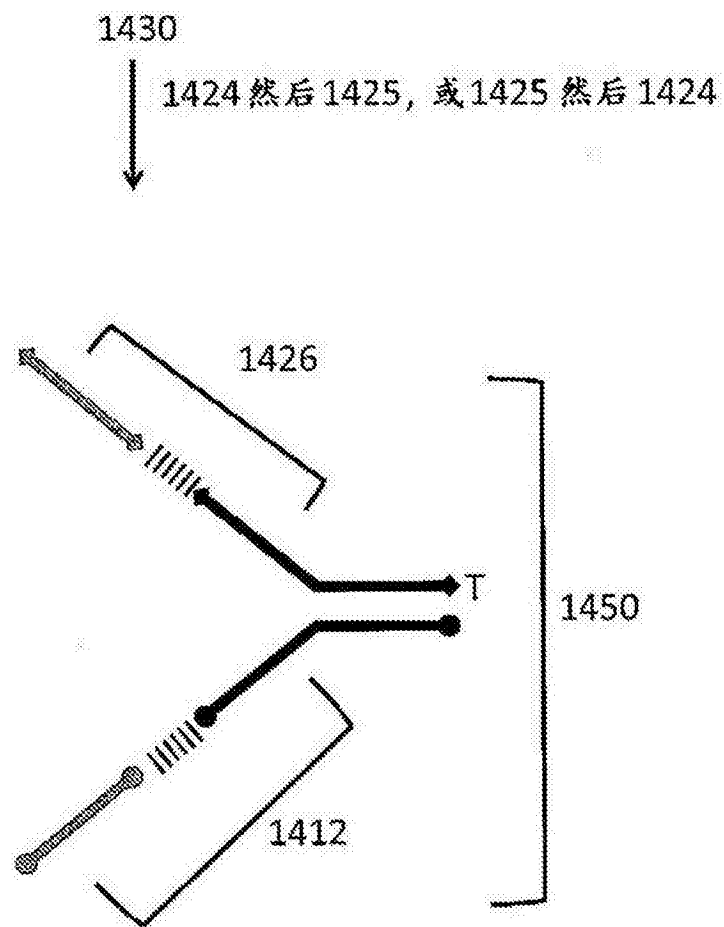


图 14e

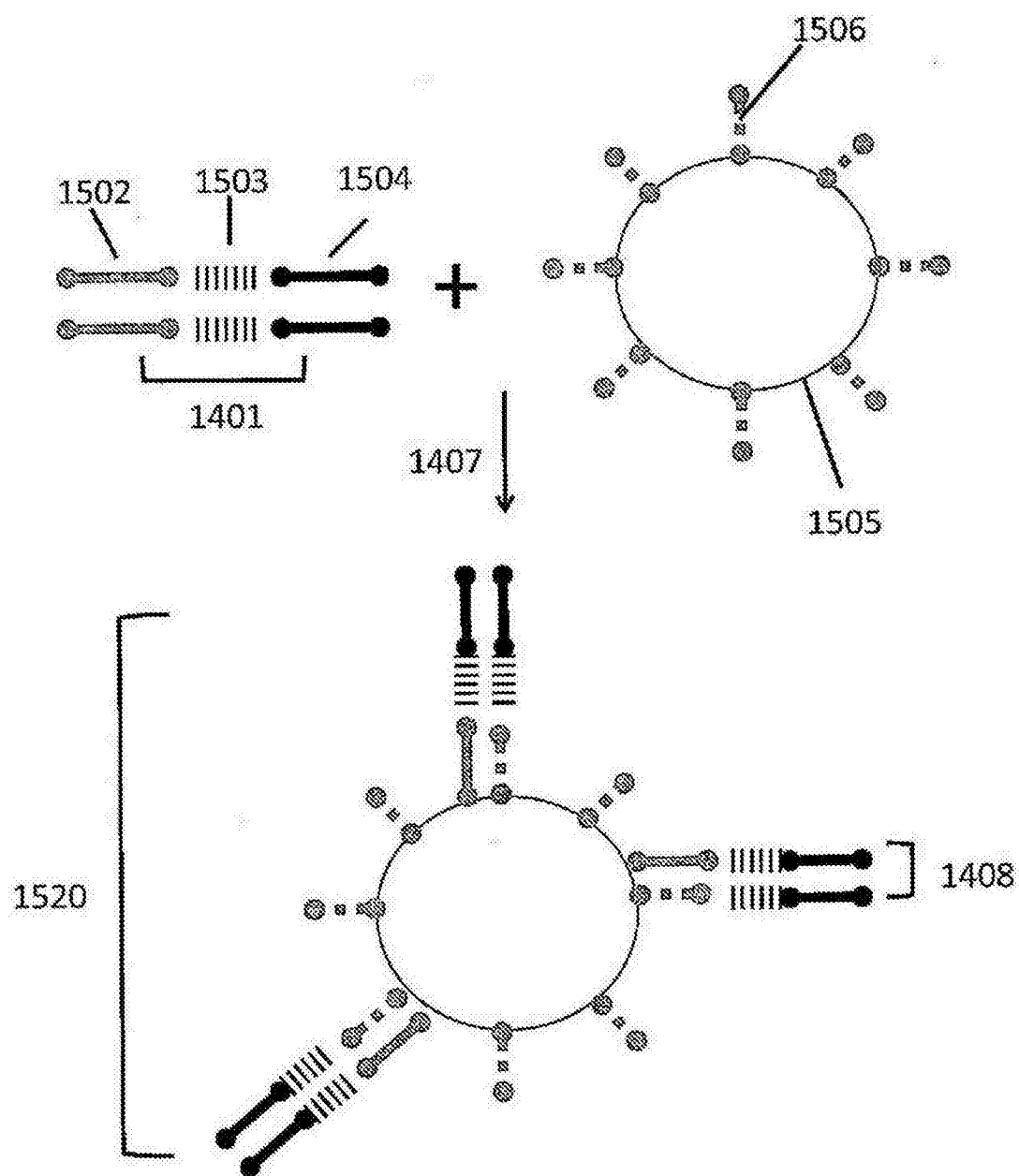


图 15a

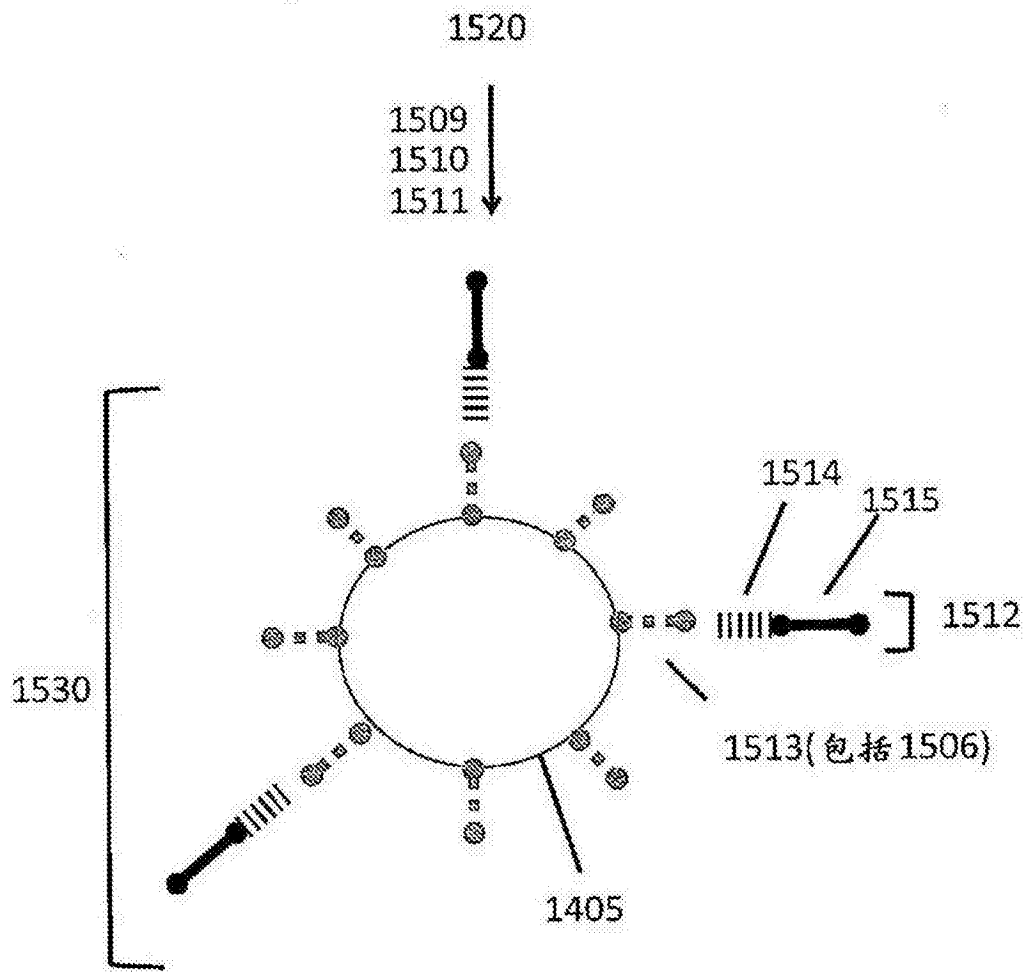


图 15b

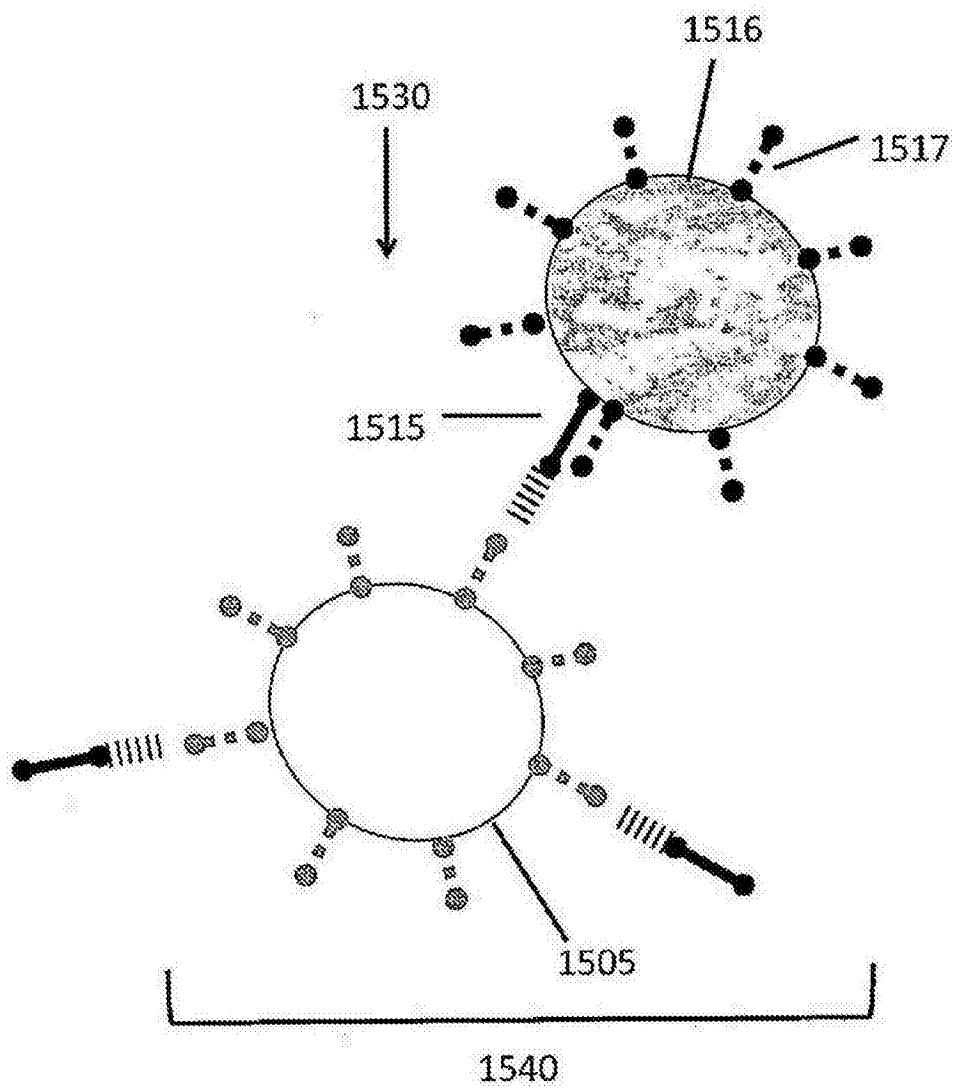


图 15c

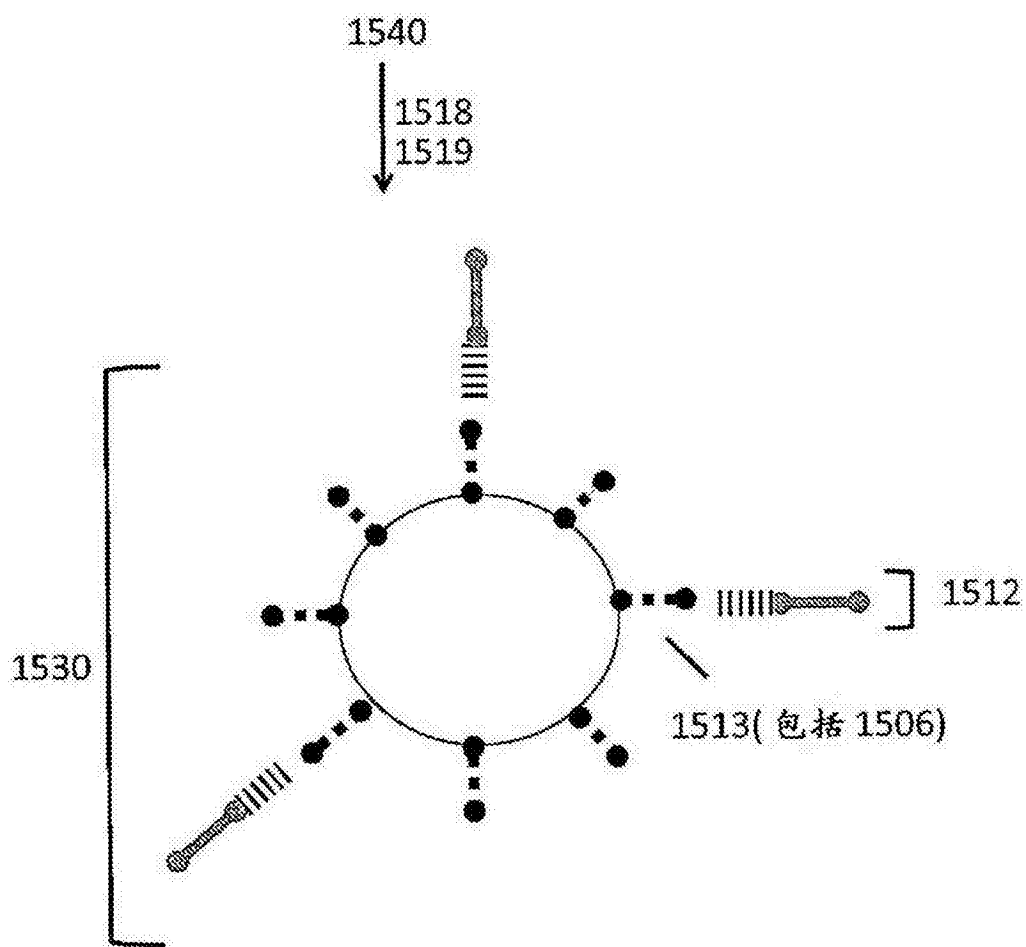


图 15d

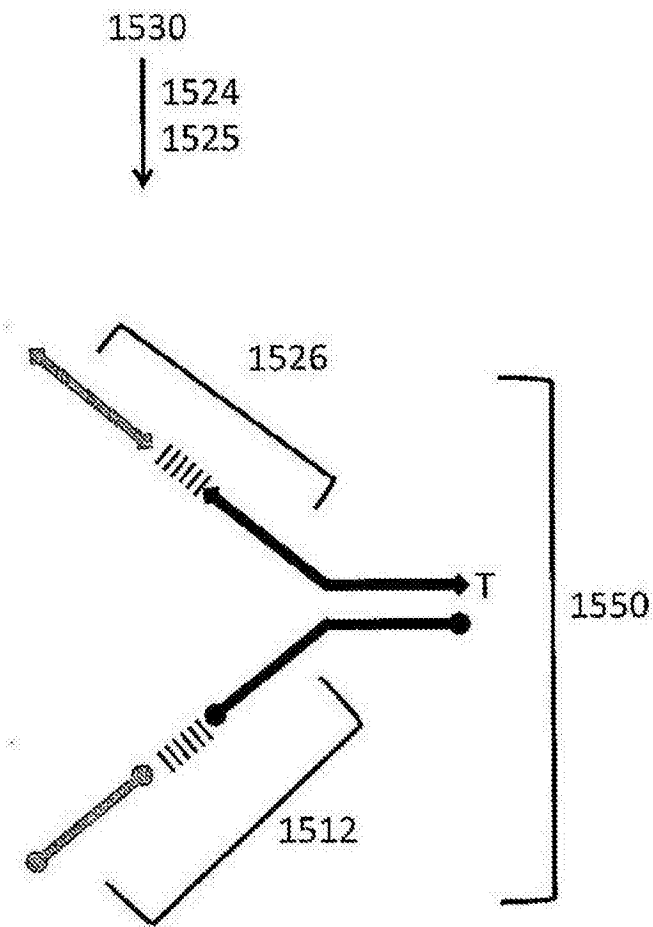


图 15e

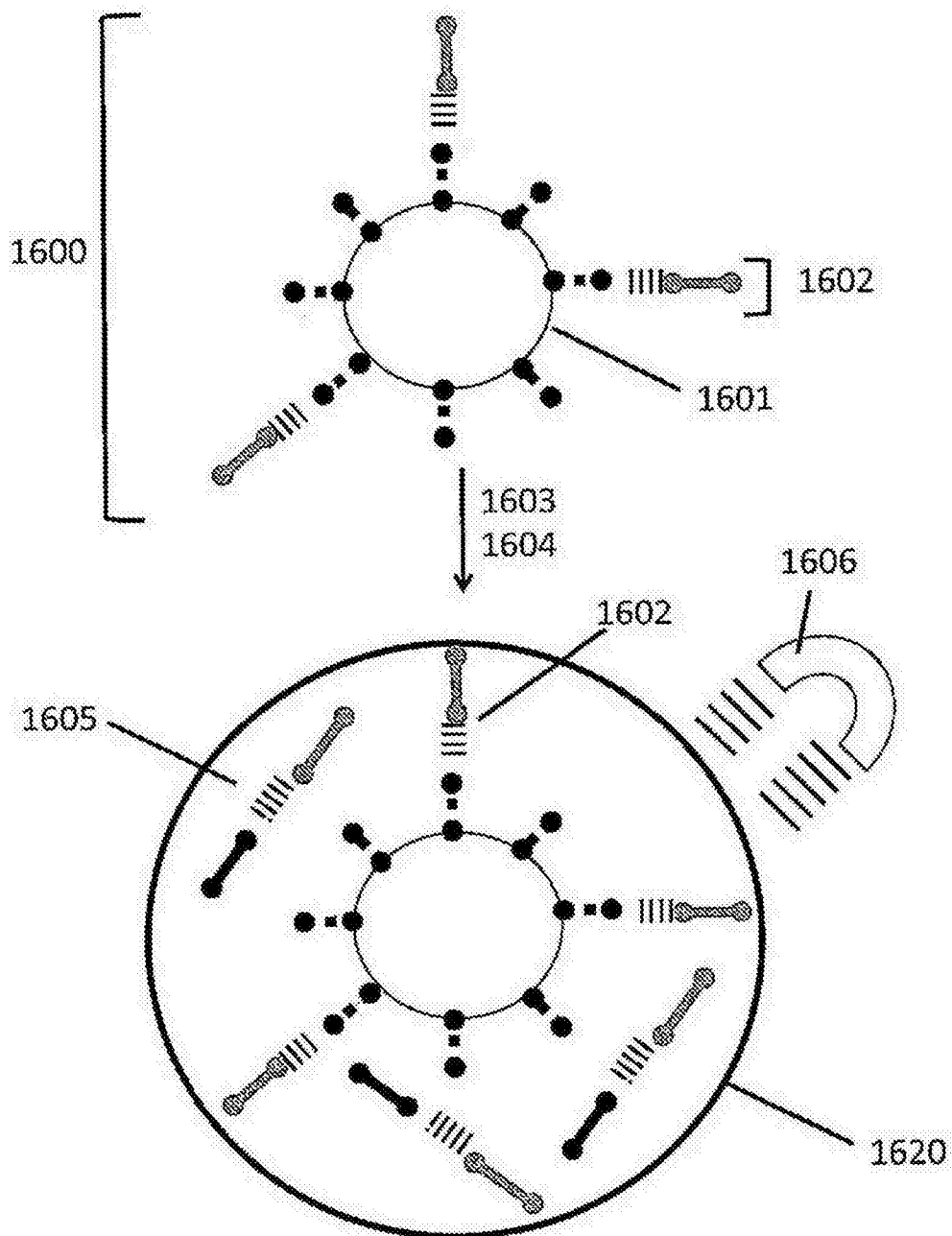


图 16

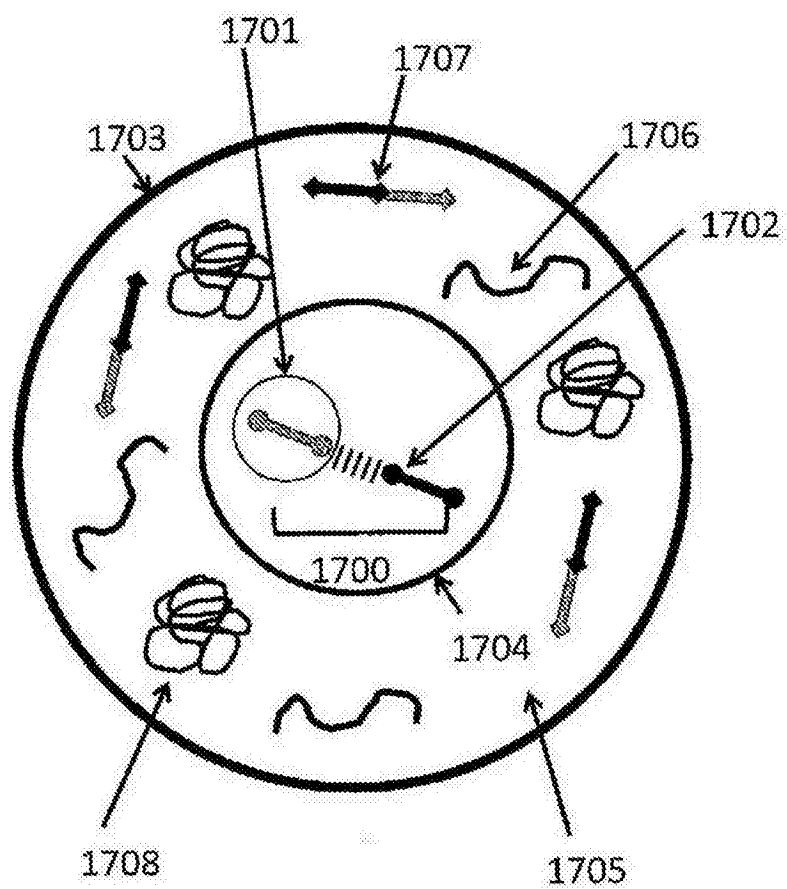


图 17

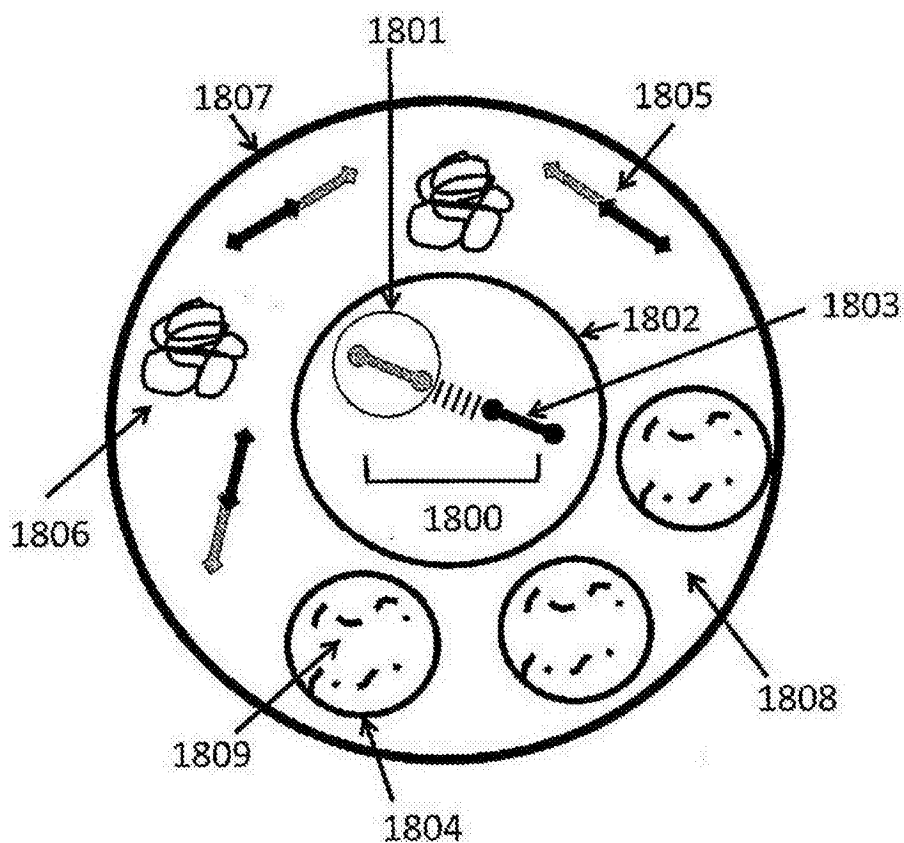


图 18

5'ACGACGCTCTTCCGATCT[条形码][引发序列]NNNNNNNN-3'
SEQ ID NO: 22

图 19a

5'ACGACGCTCTTCCGATCT[条形码][引发序列]
3'TGCTGCGAGAAGGCTAGA[条形码][引发序列]
SEQ ID NO: 23

图 19b

5'GATCT[条形码][引发序列]
3' A[条形码][引发序列]
SEQ ID NO: 24

图 19c

5'-ACACTCTTTCCCTACACGAC
 GCTCTTCC
 CGAGAAGGCTAG
 3'-CACTGACCTCAAGTCTGCACA
 SEQ ID NO: 25

图 19d

5'-ACACTCTTTCCCTACACGAC
 GCTCTTCCGATCT[条形码][引发序列]
 CGAGAAGGCTAGA[条形码][引发序列]
 3'-CACTGACCTCAAGTCTGCACA
 SEQ ID NO: 26

图 19e

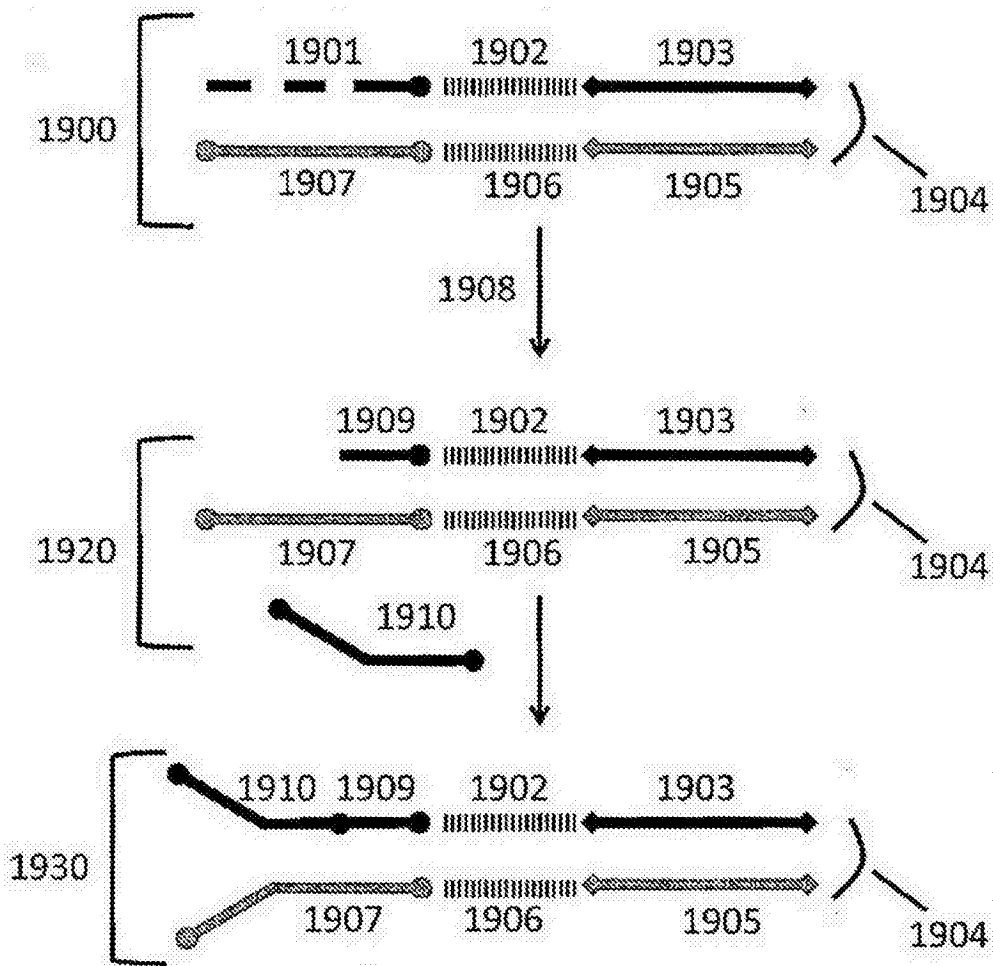


图 19f

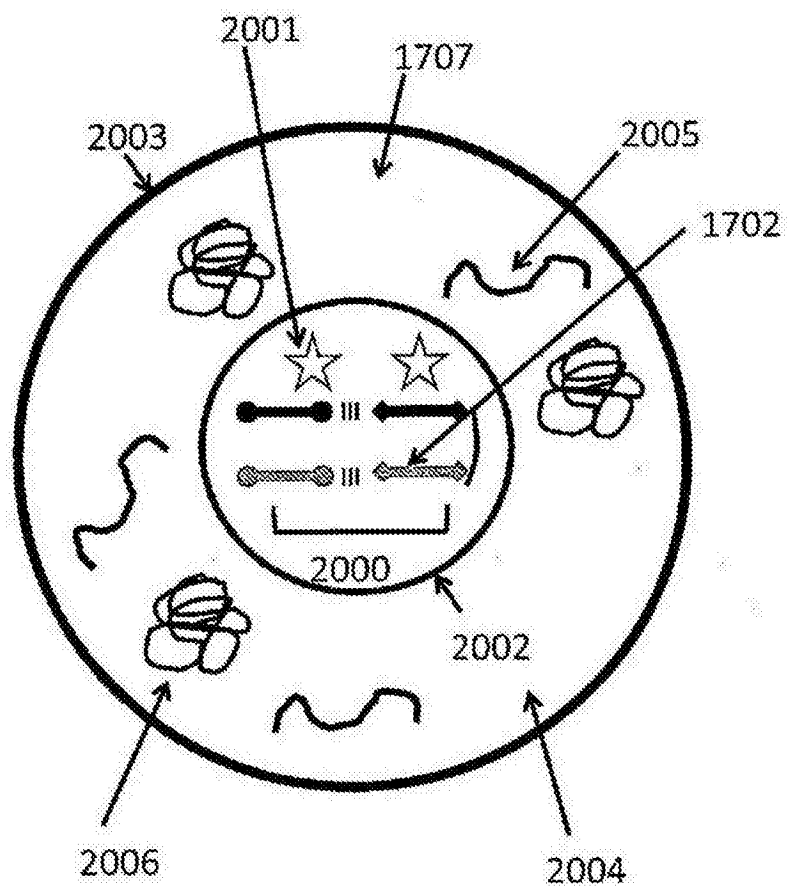


图 20

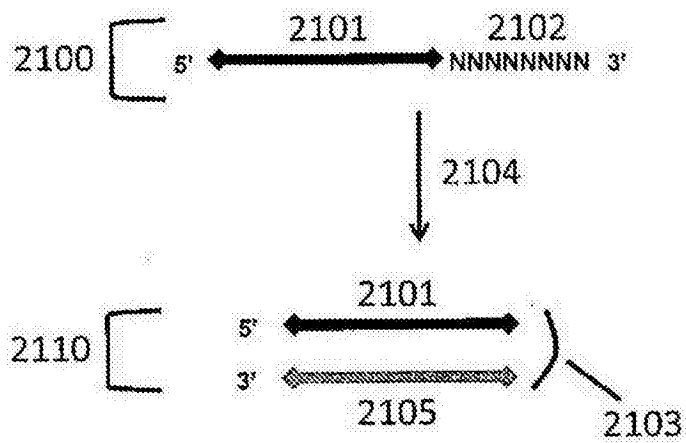


图 21a

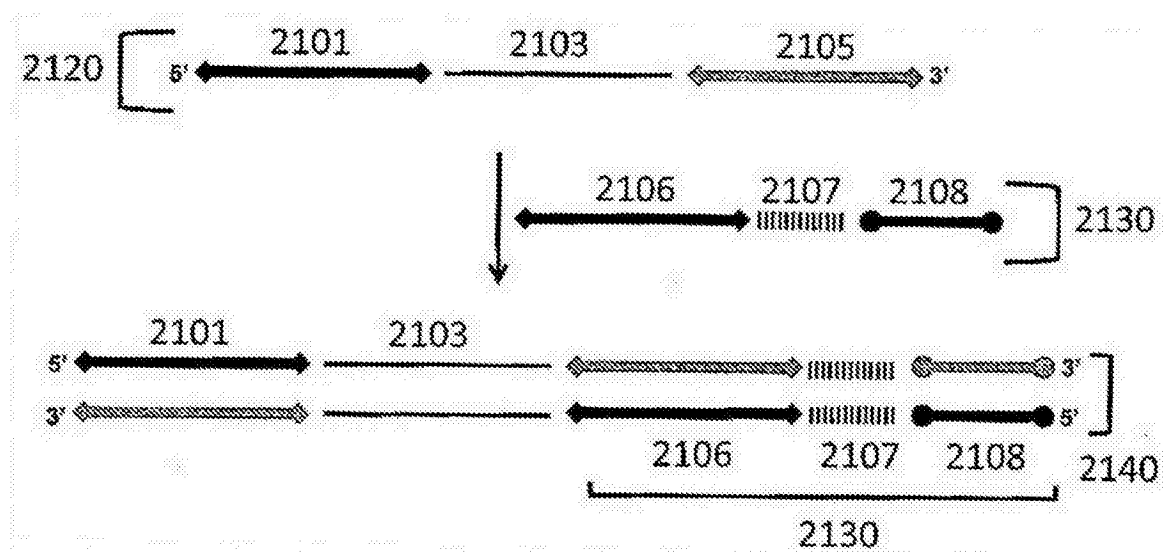


图 21b

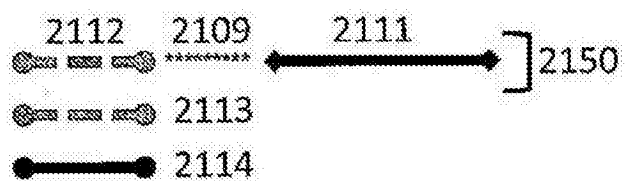


图 21c

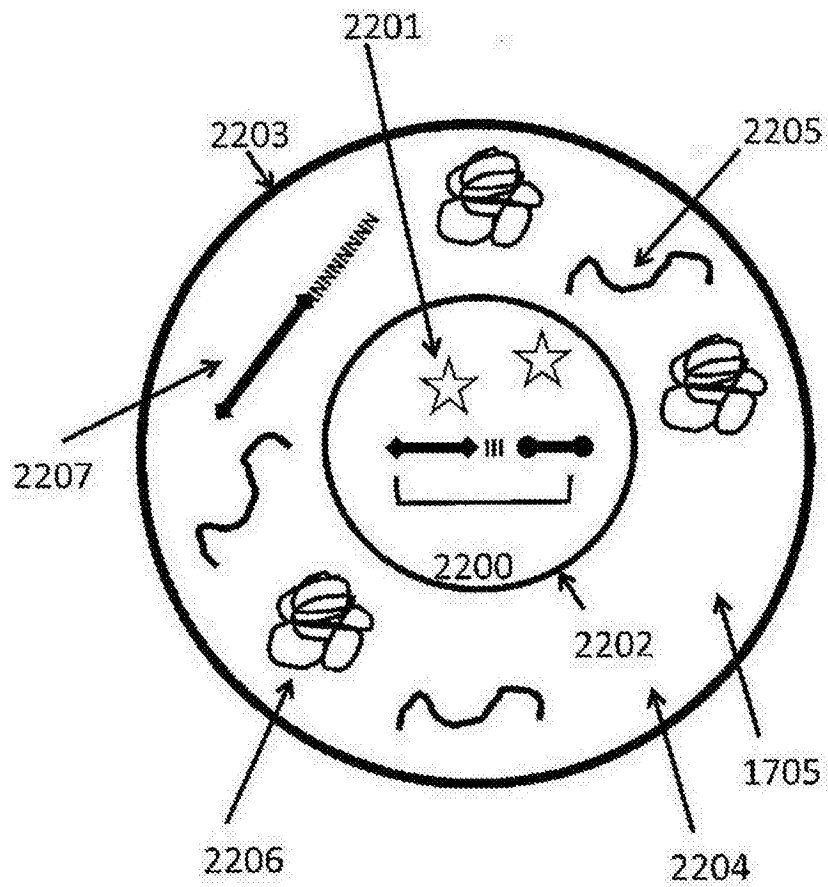


图 22

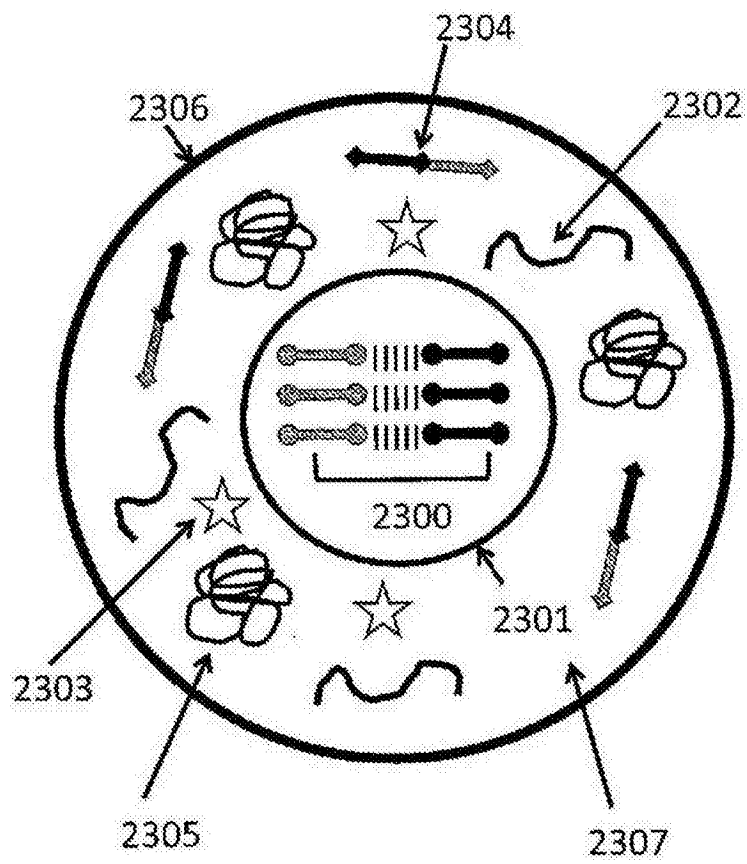


图 23