



(12) 发明专利申请

(10) 申请公布号 CN 119487205 A

(43) 申请公布日 2025. 02. 18

(21) 申请号 202380038013.5

(22) 申请日 2023.03.03

(30) 优先权数据

63/316790 2022.03.04 US

17/725042 2022.04.20 US

(85) PCT国际申请进入国家阶段日

2024.11.01

(86) PCT国际申请的申请数据

PCT/US2023/063736 2023.03.03

(87) PCT国际申请的公布数据

W02023/168444 EN 2023.09.07

(71) 申请人 元素生物科学公司

地址 美国加利福尼亚州

(72) 发明人 W·莱特 R·凯利 J·赵

(74) 专利代理机构 中国专利代理(香港)有限公司 72001

专利代理师 童春媛 彭昶

(51) Int.Cl.

C12Q 1/6806 (2006.01)

C12Q 1/6869 (2006.01)

权利要求书17页 说明书106页

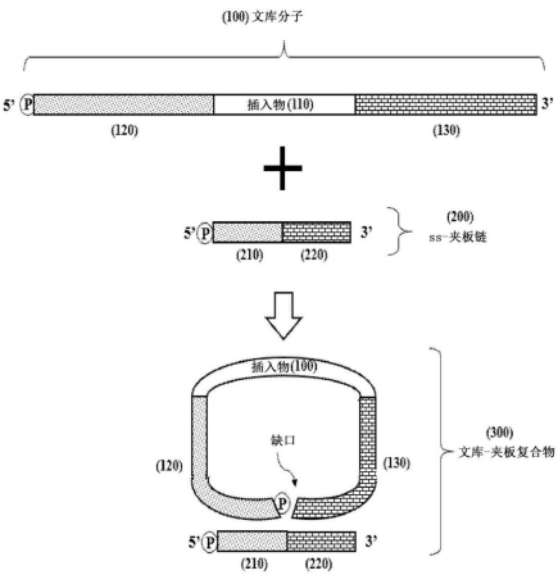
序列表(电子公布) 附图34页

(54) 发明名称

单链夹板链和使用方法

(57) 摘要

本公开提供了包含核酸单链夹板链的组合物,包括试剂盒,以及采用所述单链夹板链的方法。所述单链夹板链可以与线性文库分子的部分杂交以形成具有缺口的环化的文库-夹板复合物,其中所述缺口可以被连接以形成共价闭环状分子,所述共价闭环状分子可以经历下游扩增和测序工作流程。



1. 一种文库-夹板复合物(300),其包含:

(i) 单链核酸文库分子(100),其包括目的序列(110),所述目的序列在一侧上侧接至少第一左通用衔接子序列(120)且在另一侧上侧接至少第一右通用衔接子序列(130);以及

(ii) 单链夹板链(200),其包括第一区域(210),所述第一区域具有针对所述文库分子的所述第一左通用衔接子序列(120)的通用结合序列,并且所述单链夹板链进一步包含第二区域(220),所述第二区域具有针对所述文库分子的所述第一右通用衔接子序列(130)的通用结合序列,由此使所述文库分子成环以产生文库-夹板复合物(300),所述文库-夹板复合物在所述文库分子的5'端与所述文库分子的3'端之间具有缺口。

2. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含:第二左通用衔接子序列(140)。

3. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含:第二右通用衔接子序列(150)。

4. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含:第一左样品索引序列(160)。

5. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含:第一右样品索引序列(170)。

6. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含:第一左唯一识别序列(180)。

7. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含:第一右唯一识别序列(190)。

8. 根据权利要求1所述的文库-夹板复合物(300),其中所述核酸文库分子(100)进一步包含以下任意一者或者两者或更多者的任意组合:

(i) 第二左通用衔接子序列(140);

(ii) 第二右通用衔接子序列(150);

(iii) 第一左索引序列(160);

(iv) 第一右索引序列(170);

(v) 第一左唯一识别序列(180);和/或

(vi) 第一右唯一识别序列(190)。

9. 根据权利要求8所述的文库-夹板复合物(300),其中所述第一左通用衔接子序列(120)和/或所述第二左通用衔接子序列(140)包括:

(i) 针对正向测序引物的通用结合序列;

(ii) 针对反向测序引物的通用结合序列;

(iii) 针对第一表面引物的通用结合序列;

(iv) 针对第二表面引物的通用结合序列;

(v) 针对正向扩增引物的通用结合序列;

(vi) 针对反向扩增引物的通用结合序列;和/或

(vii) 针对压实寡核苷酸的通用结合序列。

10. 根据权利要求8所述的文库-夹板复合物(300),其中所述第一右通用衔接子序列(130)和/或所述第二右通用衔接子序列(150)包括:

- (i) 针对正向测序引物的通用结合序列;
- (ii) 针对反向测序引物的通用结合序列;
- (iii) 针对第一表面引物的通用结合序列;
- (iv) 针对第二表面引物的通用结合序列;
- (v) 针对正向扩增引物的通用结合序列;
- (vi) 针对反向扩增引物的通用结合序列;和/或
- (vii) 针对压实寡核苷酸的通用结合序列。

11. 一种用于产生多个共价闭合环状文库分子(400)的方法,其包括:

a) 提供多个单链核酸文库分子(100),其中所述多个单链核酸文库分子中的各个文库分子包括目的序列(110),所述目的序列在一侧上侧接至少第一左通用衔接子序列(120),并且在另一侧上侧接至少第一右通用衔接子序列(130);

b) 提供多个单链夹板链(200),其中所述多个单链夹板链中的各个单链夹板链(200)包含第一区域(210)和第二区域(220),所述第一区域能够与各个文库分子的所述至少第一左通用衔接子序列(120)杂交,所述第二区域能够与所述各个文库分子的所述至少第一右通用衔接子序列(130)杂交;

c) 使所述多个单链夹板链(200)与多个单链核酸文库分子(100)杂交,使得所述单链夹板链(210)中的一个的所述第一区域退火至所述文库分子的所述至少第一左通用衔接子序列(120),并且使得所述单链夹板链(220)的所述第二区域退火至所述文库分子的至少第一右通用序列(130),由此使各个文库分子成环以形成多个文库-夹板复合物(300),所述文库-夹板复合物在所述文库分子的末端5'与3'端之间具有缺口,其中所述缺口是可酶促连接的;以及

d) 连接所述多个文库-夹板复合物(300)中的所述缺口,由此产生多个共价闭合环状文库分子(400)。

12. 根据权利要求11所述的方法,其进一步包括:(e) 在适用于使各个共价闭合环状文库分子(400)与各个固定的表面引物杂交的条件下,将所述多个共价闭合环状文库分子(400)分配到支持物上,所述支持物具有固定在所述支持物上的多个表面引物,由此将所述多个共价闭合环状文库分子(400)固定到所述支持物。

13. 根据权利要求12所述的方法,其进一步包括:(f) 在适合于使用多个表面引物作为固定的扩增引物并使用所述多个共价闭合环状文库分子(400)作为模板分子在所述支持物上进行滚环扩增反应的条件下,使多个固定的共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,由此产生多个固定的核酸多联体分子。

14. 根据权利要求13所述的方法,其进一步包括:对所述多个固定的核酸多联体分子进行测序。

15. 根据权利要求14所述的方法,其中所述测序包括:

a) 使多个固定的多联体分子与(i) 多个测序聚合酶和(ii) 多个可溶性测序引物接触,其中所述接触在适合于形成各自包含结合至核酸双链体的测序聚合酶的多个复合聚合酶的条件下进行,其中所述核酸双链体包含与可溶性测序引物杂交的多联体分子;

b) 在适用于将至少一个核苷酸与复合测序聚合酶结合的条件下,使多个复合测序聚合酶与多个核苷酸接触,其中所述多个核苷酸包括至少一个用荧光团标记并且在糖3'位置处

具有可移除链终止部分的核苷酸类似物；

c) 将至少一个核苷酸掺入到杂交的测序引物的3'端中,由此产生多个新生的延伸的测序引物;以及

d) 检测经掺入的核苷酸并且识别所述经掺入的核苷酸的核碱基。

16. 根据权利要求14所述的方法,其中所述测序包括:

a) 使所述多个固定的多联体分子与(i) 多个测序聚合酶和(ii) 多个可溶性测序引物接触,其中所述接触在适合于形成各自包含结合至核酸双链体的测序聚合酶的多个第一复合聚合酶的条件下进行,其中所述核酸双链体包含与可溶性测序引物杂交的多联体分子;

b) 在适用于将多价分子的互补核苷酸单元与所述多个第一复合聚合酶中的至少两个进行结合以形成多个多价-复合聚合酶的条件下,使多个复合测序聚合酶与多个可检测地标记的多价分子接触,由此形成多个多价-复合聚合酶,并且所述条件抑制所述互补核苷酸单元掺入到所述多个多价-复合聚合酶的测序引物中,其中多个多价分子中的各个多价分子包含附接至多个核苷酸臂的核心,并且每个核苷酸臂附接至核苷酸单元;

c) 检测所述多个多价-复合聚合酶;以及

d) 识别结合至所述多个多价-复合聚合酶中的所述多个第一复合聚合酶的所述互补核苷酸单元的核碱基,由此确定核酸模板的序列。

17. 根据权利要求16所述的方法,其进一步包括:

e) 解离所述多个多价-复合聚合酶并移除多个第一测序聚合酶及其结合的多价分子,并且保留多个核酸双链体;

f) 使步骤(e)的多个保留的核酸双链体与多个第二测序聚合酶接触,其中所述接触在适用于将所述多个第二测序聚合酶与所述多个保留的核酸双链体结合的条件下进行,由此形成各自包含结合至保留的核酸双链体的第二测序聚合酶的多个第二复合聚合酶;以及

g) 使所述多个第二复合聚合酶与多个核苷酸接触,其中所述接触在适用于将来自所述多个核苷酸的互补核苷酸与步骤(f)的所述第二复合聚合酶中的至少两个进行结合的条件下进行,由此形成多个核苷酸-复合聚合酶,并且所述条件适用于促进结合的互补核苷酸掺入到所述核苷酸-复合聚合酶的测序引物中。

18. 根据权利要求17所述的方法,其进一步包括:

h) 检测被掺入到所述核苷酸-复合聚合酶的所述测序引物中的所述互补核苷酸。

19. 根据权利要求17所述的方法,其进一步包括:

h) 检测被掺入到所述核苷酸-复合聚合酶的所述测序引物中的所述互补核苷酸;以及

i) 识别被掺入到所述核苷酸-复合聚合酶的所述测序引物中的所述互补核苷酸的核碱基。

20. 一种形成至少一种亲和力复合物的方法,其包括:

a) 将第一通用核酸引物、第一DNA聚合酶和第一多价分子与根据权利要求16所述的多联体分子的第一部分结合,由此形成第一结合复合物,其中所述第一多价分子的第一核苷酸单元与所述第一DNA聚合酶结合;以及

b) 将第二通用核酸引物、第二DNA聚合酶以及所述第一多价分子与同一多联体模板分子的第二部分结合,由此形成第二结合复合物,其中所述第一多价分子的第二核苷酸单元与所述第二DNA聚合酶结合,其中包括同一多价分子的所述第一结合复合物和所述第二结

合复合物形成亲和力复合物,其中所述第一多价分子包含附接至多个核苷酸臂的核心并且每个核苷酸臂附接至核苷酸单元,并且其中所述多联体分子包含目的序列(110)的两个或更多个串联重复序列以及结合所述第一通用核酸引物和所述第二通用核酸引物的通用引物结合位点。

21.一种通过形成至少一种亲和力复合物进行测序的方法,其包括:

a) 将第一通用核酸引物、第一DNA聚合酶和第一多价分子与根据权利要求16所述的多联体分子的第一部分结合,由此形成第一结合复合物,其中所述第一多价分子的第一核苷酸单元与所述第一DNA聚合酶结合;

b) 将第二通用核酸引物、第二DNA聚合酶以及所述第一多价分子与同一多联体模板分子的第二部分结合,由此形成第二结合复合物,其中所述第一多价分子的第二核苷酸单元与所述第二DNA聚合酶结合,其中包括同一多价分子的所述第一结合复合物和所述第二结合复合物形成亲和力复合物,其中所述第一多价分子包含附接至多个核苷酸臂的核心并且每个核苷酸臂附接至核苷酸单元,其中所述多联体分子包含目的序列(110)的两个或更多个串联重复序列以及结合所述第一通用核酸引物和所述第二通用核酸引物的通用引物结合位点,并且其中所述接触在适合于抑制所述第一结合复合物和所述第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化掺入的条件下进行;

c) 检测所述同一多联体模板分子上的所述第一结合复合物和所述第二结合复合物,以及

d) 识别所述第一结合复合物中的所述第一核苷酸单元,由此确定所述多联体模板分子的所述第一部分的序列,并且识别所述第二结合复合物中的所述第二核苷酸单元,由此确定所述多联体模板分子的所述第二部分的序列。

22.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含:第二左通用衔接子序列(140)。

23.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含:第二右通用衔接子序列(150)。

24.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含:第一左样品索引序列(160)。

25.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含:第一右样品索引序列(170)。

26.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含:第一左唯一识别序列(180)。

27.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含:第一右唯一识别序列(190)。

28.根据权利要求11所述的方法,其中所述核酸文库分子(100)进一步包含以下任意一者或者两者或更多者的任意组合:

(vii) 第二左通用衔接子序列(140);

(viii) 第二右通用衔接子序列(150);

(ix) 第一左索引序列(160);

(x) 第一右索引序列(170);

(xi) 第一左唯一识别序列(180);和/或

(xii) 第一右唯一识别序列(190)。

29. 根据权利要求28所述的方法, 其中所述第一左通用衔接子序列(120) 和/或所述第二左通用衔接子序列(140) 包括:

(viii) 针对正向测序引物的通用结合序列;

(ix) 针对反向测序引物的通用结合序列;

(x) 针对第一表面引物的通用结合序列;

(xi) 针对第二表面引物的通用结合序列;

(xii) 针对正向扩增引物的通用结合序列;

(xiii) 针对反向扩增引物的通用结合序列;和/或

(xiv) 针对压实寡核苷酸的通用结合序列。

30. 根据权利要求28所述的方法, 其中所述第一右通用衔接子序列(130) 和/或所述第二右通用衔接子序列(150) 包括:

(viii) 针对正向测序引物的通用结合序列;

(ix) 针对反向测序引物的通用结合序列;

(x) 针对第一表面引物的通用结合序列;

(xi) 针对第二表面引物的通用结合序列;

(xii) 针对正向扩增引物的通用结合序列;

(xiii) 针对反向扩增引物的通用结合序列;和/或

(xiv) 针对压实寡核苷酸的通用结合序列。

31. 一种用于对核酸分子进行测序的方法, 其包括:

a) 提供固定在支持物上的随机位置处的多个核酸模板分子, 其中各个固定的模板分子包含插入序列区域和一个样品索引, 其中每个样品索引包含与识别插入序列的样品源的通用样品索引序列接合的3聚体随机序列, 其中不同的固定的模板分子具有不同的3聚体随机序列和相同的通用样品索引序列, 并且其中所述固定的模板分子具有不同的插入序列;

b) 使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂对多个固定的模板分子的所述3聚体随机序列进行三个循环的聚合酶介导的测序反应, 其中所述核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记, 其中三个循环的测序包括对从结合至所述固定的模板分子的所述可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像, 由此确定所述多个固定的模板分子中的各个模板分子中的所述3聚体随机序列的序列, 并且其中在第一测序循环、第二测序循环和第三测序循环中的每一者中, 在所述多个固定的模板分子之间, 对A、G、C和T/U核碱基的平衡多样性进行检测和成像; 以及

c) 使用在步骤(b) 中获得的图像来生成所述多个固定的模板分子的位置的图, 其中所述插入区域的序列不用于生成所述图。

32. 根据权利要求31所述的方法, 其中步骤(b) 的所述平衡多样性是在所述第一测序循环、所述第二测序循环和所述第三测序循环中的每一者中检测和成像的所述核碱基A、G、C和T/U中的每一者的25-75%。

33. 根据权利要求31所述的方法, 其进一步包括:

d) 对所述多个固定的模板分子的所述通用样品索引序列进行测序;
e) 对所述多个固定的模板分子的所述插入序列区域进行测序;以及
f) 将在步骤(b)中获得的给定模板分子的所述插入序列与在步骤(a)中获得的来自同一给定模板分子的所述通用样品索引序列进行指定,由此识别所述给定插入序列的所述样品源。

34. 根据权利要求31所述的方法,其中所述多个核酸模板分子进一步包含第二样品索引,所述第二样品索引包含识别所述插入序列的所述样品源的第二通用样品索引序列,并且所述第二样品索引缺少随机序列。

35. 根据权利要求34所述的方法,其进一步包括:

g) 对所述多个固定的模板分子的所述3聚体随机序列进行测序,以获得在所述第一测序循环、所述第二测序循环和所述第三测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性,从而生成所述多个固定的模板分子的位置的图;

h) 对所述多个固定的模板分子的所述第一通用样品索引序列进行测序;

i) 对所述多个固定的模板分子的所述第二通用样品索引序列进行测序;

j) 对所述多个固定的模板分子的所述插入序列区域进行测序;以及

k) 将在步骤(c)中获得的给定模板分子的所述插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的所述第一通用样品索引序列和所述第二通用样品索引序列进行指定,由此识别所述给定插入序列的所述样品源。

36. 一种用于对核酸分子进行测序的方法,其包括:

a) 提供固定和支持物上的随机位置处的多个核酸模板分子,其中各个固定的模板分子包含插入序列区域和一个样品索引,其中每个样品索引包含与识别插入序列的样品源的通用样品索引序列接合的3聚体随机序列,其中所述通用样品索引序列包含3-20个核苷酸,其中不同的固定的模板分子具有不同的3聚体随机序列和相同的通用样品索引序列,并且其中所述固定的模板分子具有不同的插入序列;

b) 使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂对所述多个固定的模板分子的所述3聚体随机序列和所述通用样品索引序列的第一碱基位置进行四个循环的聚合酶介导的测序反应,其中所述核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中四个循环的测序包括对从结合至所述固定的模板分子的所述可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定所述多个固定的模板分子中的各个模板分子中的所述通用样品索引序列的所述第一碱基位置和所述3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中,在所述多个固定的模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及

c) 使用在步骤(b)中获得的所述四个循环的聚合酶介导的测序反应的图像来生成所述多个固定的模板分子的位置的图,其中所述插入区域的序列不用于生成所述图。

37. 根据权利要求36所述的方法,其中步骤(b)的所述平衡多样性是在所述第一测序循环、所述第二测序循环、所述第三测序循环和所述第四测序循环中的每一者中检测和成像的所述核碱基A、G、C和T/U中的每一者的25-75%。

38. 根据权利要求36所述的方法,其进一步包括:

a) 对所述多个固定的模板分子的所述通用样品索引序列的其余碱基位置进行测序；
b) 对所述多个固定的模板分子的所述插入序列区域进行测序；以及
c) 将在步骤(b)中获得的给定模板分子的所述插入序列与在步骤(a)中获得的来自同一给定模板分子的所述通用样品索引序列进行指定，由此识别所述给定插入序列的所述样品源。

39. 根据权利要求36所述的方法，其中所述多个核酸模板分子进一步包含第二样品索引，所述第二样品索引包含识别所述插入序列的所述样品源的第二通用样品索引序列，并且所述第二样品索引缺少随机序列。

40. 根据权利要求39所述的方法，其进一步包括：

a) 对所述多个固定的模板分子的所述3聚体随机序列和所述通用样品索引序列的所述第一碱基位置进行测序，以获得在所述第一测序循环、所述第二测序循环、所述第三测序循环和所述第四测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性，从而生成所述多个固定的模板分子的位置的图；

b) 对所述多个固定的模板分子的所述第一通用样品索引序列的其余碱基位置进行测序；

c) 对所述多个固定的模板分子的所述第二通用样品索引序列进行测序；

d) 对所述多个固定的模板分子的所述插入序列区域进行测序；以及

e) 将在步骤(c)中获得的给定模板分子的所述插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的所述第一通用样品索引序列和所述第二通用样品索引序列进行指定，由此识别所述给定插入序列的所述样品源。

41. 根据权利要求30至40中任一项所述的方法，其中所述支持物包括玻璃或塑料基底。

42. 根据权利要求30至40中任一项所述的方法，其中所述支持物被配置在流通池通道、流通池或毛细管腔上。

43. 根据权利要求30至40中任一项所述的方法，其中所述支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。

44. 根据权利要求43所述的方法，其中所述至少一个亲水性聚合物涂层包含选自以下项组成的组的分子：聚乙二醇(PEG)、聚(乙烯醇)(PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮)(PVP)、聚(丙烯酸)(PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺)(PNIPAM)、聚(甲基丙烯酸甲酯)(PMA)、聚(2-羟乙基甲基丙烯酸酯)(HEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(POEGMA)、聚谷氨酸(PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。

45. 根据权利要求43所述的方法，其中所述至少一个亲水性聚合物涂层包含具有至少四个支链的支链亲水性聚合物分子。

46. 根据权利要求43所述的方法，其中所述至少一个亲水性聚合物涂层包含具有至少1000道尔顿的分子量的聚合物分子。

47. 根据权利要求30至40中任一项所述的方法，其中所述固定的模板分子包含具有所述一个样品索引和所述插入序列的串联重复序列的多个固定的多联体分子。

48. 根据权利要求30至40中任一项所述的方法，其中所述固定的模板分子包含多个不同的成簇模板分子，所述成簇模板分子具有所述插入序列的一个拷贝和所述一个样品索引的一个拷贝，其中所述成簇模板分子经由桥式扩增产生。

49. 根据权利要求30至40中任一项所述的方法,其中位于所述支持物上的随机位置处的所述固定的核酸模板分子的密度为 10^4 - 10^8 个/ mm^2 。

50. 根据权利要求30至40中任一项所述的方法,其中所述插入序列的所述样品源是基因组DNA、双链cDNA或细胞游离循环DNA。

51. 根据权利要求30至40中任一项所述的方法,其中所述可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。

52. 根据权利要求30至40中任一项所述的方法,其中所述可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。

53. 根据权利要求30至40中任一项所述的方法,其中所述可检测地标记的核苷酸试剂包含多价分子,每个多价分子包含(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中所述核心附接至所述多个核苷酸臂,其中所述间隔基附接至所述接头,其中所述接头附接至所述核苷酸单元。

54. 根据权利要求30至40中任一项所述的方法,其中结合至所述固定的模板分子的所述可检测地标记的核苷酸试剂包含与测序引物杂交以形成双链体的各个固定的模板分子,并且所述双链体结合至聚合酶以形成复合聚合酶,并且所述复合聚合酶结合至可检测地标记的核苷酸试剂。

55. 根据权利要求54所述的方法,其中在适用于将可检测地标记的核苷酸试剂与所述复合聚合酶结合并将所述可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,所述复合聚合酶结合至所述可检测地标记的核苷酸试剂,其中所述可检测地标记的核苷酸试剂包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团。

56. 根据权利要求54所述的方法,其中在适用于将可检测地标记的核苷酸试剂与所述复合聚合酶结合并将所述可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,所述复合聚合酶结合至所述可检测地标记的核苷酸试剂,其中所述可检测地标记的核苷酸试剂包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团。

57. 根据权利要求54所述的方法,其中在适用于将可检测地标记的核苷酸试剂与所述复合聚合酶结合的条件下,所述复合聚合酶结合至所述可检测地标记的核苷酸试剂,并且所述条件适用于抑制核苷酸掺入,其中所述可检测地标记的核苷酸试剂包含多价分子,所述多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中所述核心附接至所述多个核苷酸臂,其中所述间隔基附接至所述接头,其中所述接头附接至所述核苷酸单元。

58. 根据权利要求54所述的方法,其中所述固定的模板分子包含固定的多联体分子,所述固定的多联体分子与多个测序引物杂交以在同一多联体分子上形成至少第一双链体和第二双链体,其中所述第一和双链体结合至第一聚合酶并且所述第二双链体结合至第二聚合酶,以形成第一复合聚合酶和第二复合聚合酶,并且其中所述方法包括:

a) 使多个多价分子与同一多联体模板分子上的所述第一复合聚合酶和所述第二复合

聚合酶接触,其中各个多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中所述核心附接至所述多个核苷酸臂,并且所述间隔基附接至所述接头,并且所述接头附接至所述核苷酸单元,其中所述接触在适合于将来自所述多个多价分子中的单个多价分子与所述第一复合聚合酶和所述第二复合聚合酶结合的条件下进行,其中所述单个多价分子的第一核苷酸单元结合至包括与所述多联体模板分子的第一部分杂交的所述第一测序引物的所述第一复合聚合酶,由此形成第一结合复合物,并且其中所述单个多价分子的第二核苷酸单元结合至包括与所述多联体模板分子的第二部分杂交的第二测序引物的所述第二复合聚合酶,由此形成第二结合复合物,并且

- 其中结合至同一多价分子的所述第一结合复合物和所述第二结合复合物形成亲和复合物,

- 其中所述接触在适合于抑制所述第一结合复合物和所述第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化的掺入的条件下进行;

- b) 检测所述同一多联体模板分子上的所述第一结合复合物和所述第二结合复合物;

- c) 对从所述可检测地标记的多价分子发出的所述光学颜色信号进行成像,所述可检测地标记的多价分子在所述同一多联体模板分子上形成所述第一结合复合物和所述第二结合复合物;以及

- d) 识别所述第一结合复合物中的所述第一核苷酸单元,由此确定所述多联体模板分子的所述第一部分的序列,并且识别所述第二结合复合物中的所述第二核苷酸单元,由此确定所述多联体模板分子的所述第二部分的序列。

59. 一种用于对核酸分子进行测序的方法,其包括:

- a) 提供第一多个文库分子,所述多个文库分子中的每个分子包含(i)源自第一样品源的插入序列区域,(ii)第一样品索引,其具有与第一通用样品索引序列接合的3聚体随机序列,以及(iii)具有第二通用样品索引序列的第二样品索引,所述第二样品索引缺少随机序列,其中所述第一通用样品索引序列和所述第二通用样品索引序列的组合唯一地识别所述插入序列的第一样品源,其中不同的第一文库分子具有不同的3聚体随机序列并且具有不同的插入序列;

- b) 提供第二多个文库分子,所述多个文库分子中的每个分子包含(i)源自第二样品源的插入序列区域,(ii)第三样品索引,其具有与第三通用样品索引序列接合的3聚体随机序列,以及(iii)具有第四通用样品索引序列的第四样品索引,所述第四样品索引缺少随机序列,其中所述第三通用样品索引序列和所述第四通用样品索引序列的组合唯一地识别所述插入序列的第二样品源,其中不同的第二文库分子具有不同的3聚体随机序列并且具有不同的插入序列;

- c) 汇集所述第一多个文库分子和所述第二多个文库分子;

- d) 将经汇集的文库分子分配到支持物上并进行扩增反应以产生固定至所述支持物的随机位置处的多个克隆扩增的模板分子;

- e) 使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂,对所述第一样品索引和所述第三样品索引的所述3聚体随机序列进行三个循环的聚合酶介导的测序反应,其中所述核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可

检测颜色标记,其中三个循环的测序包括对从结合至固定的扩增模板分子的所述可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定所述多个固定的模板分子中的各个模板分子中的所述3聚体随机序列的序列,并且其中在所述第一测序循环、所述第二测序循环和所述第三测序循环的每一者中,在所述多个固定的扩增模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及

f) 使用在步骤(e)中获得的图像生成所述多个固定的模板分子的位置的图,其中所述插入区域的序列不用于生成所述图。

60. 根据权利要求1所述的方法,其中步骤(e)的所述平衡多样性是在所述第一测序循环、所述第二测序循环和所述第三测序循环中的每一者中检测和成像的所述核碱基A、G、C和T/U中的每一者的25-75%。

61. 根据权利要求59所述的方法,其进一步包括:

g) 对所述多个固定的模板分子的所述第一通用样品索引序列进行测序;

h) 对所述多个固定的模板分子的所述第二通用样品索引序列进行测序;

i) 对源自所述第一文库分子的所述多个固定的模板分子的所述插入序列区域进行测序;以及

j) 将在步骤(c)中获得的给定模板分子的所述插入序列与来自同一给定模板分子的所述第一通用样品索引序列和所述第二通用样品索引序列进行指定,由此识别所述给定插入序列的所述第一样品源。

62. 根据权利要求59所述的方法,其进一步包括:

g) 对所述多个固定的模板分子的所述第三通用样品索引序列进行测序;

h) 对所述多个固定的模板分子的所述第四通用样品索引序列进行测序;

i) 对源自所述第二文库分子的所述多个固定的模板分子的所述插入序列区域进行测序;以及

j) 将在步骤(c)中获得的给定模板分子的所述插入序列与来自同一给定模板分子的所述第三通用样品索引序列和所述第四通用样品索引序列进行指定,由此识别所述给定插入序列的第二样品源。

63. 一种用于对核酸分子进行测序的方法,其包括:

a) 提供第一多个文库分子,所述多个文库分子中的每个分子包含(i)源自第一样品源的插入序列区域,(ii)第一样品索引,其具有与第一通用样品索引序列接合的3聚体随机序列,以及(iii)具有第二通用样品索引序列的第二样品索引,所述第二样品索引缺少随机序列,其中所述第一通用样品索引序列和所述第二通用样品索引序列的组合唯一地识别所述插入序列的第一样品源,其中所述第一通用样品索引序列包含3-20个核苷酸,其中不同的第一文库分子具有不同的3聚体随机序列并且具有不同的插入序列;

b) 提供第二多个文库分子,所述多个文库分子中的每个分子包含(i)源自第二样品源的插入序列区域,(ii)第三样品索引,其具有与第三通用样品索引序列接合的3聚体随机序列,以及(iii)具有第四通用样品索引序列的第四样品索引,所述第四样品索引缺少随机序列,其中所述第三通用样品索引序列和所述第四通用样品索引序列的组合唯一地识别所述插入序列的第二样品源,其中所述第三通用样品索引序列包含3-20个核苷酸,其中不同的第二文库分子具有不同的3聚体随机序列并且具有不同的插入序列;

- c) 汇集所述第一多个文库分子和所述第二多个文库分子;
- d) 将经汇集的文库分子分配到支持物上并进行扩增反应以产生固定至所述支持物的随机位置处的多个克隆扩增的模板分子;
- e) 使用包含不同类型的核碱基A、G、C和T/U的混合物的多个可检测地标记的核苷酸试剂对所述第一样品序列和所述第三样品索引的所述3聚体随机序列进行四个循环的聚合酶介导的测序反应,并对所述第一通用样品索引序列和所述第三通用样品索引序列的第一碱基位置进行测序,其中所述核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的扩增模板分子的所述可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定所述多个固定的模板分子中的各个模板分子中的所述3聚体随机序列的序列,并且其中在所述第一测序循环、所述第二测序循环、所述第三测序循环和所述第四测序循环中的每一者中,在所述多个固定的扩增模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及
- f) 使用在步骤(e)中获得的图像生成所述多个固定的模板分子的位置的图,其中所述插入区域的序列不用于生成所述图。

64. 根据权利要求63所述的方法,其中步骤(e)的所述平衡多样性是在所述第一测序循环、所述第二测序循环、所述第三测序循环和所述第四测序循环中的每一者中检测和成像的所述核碱基A、G、C和T/U中的每一者的25-75%。

65. 根据权利要求63所述的方法,其进一步包括:

- g) 对所述多个固定的模板分子的所述第一通用样品索引序列的其余碱基位置进行测序;
- h) 对所述多个固定的模板分子的所述第二通用样品索引序列进行测序;
- i) 对源自所述第一文库分子的所述多个固定的模板分子的所述插入序列区域进行测序;以及
- j) 将在步骤(c)中获得的给定模板分子的所述插入序列与来自同一给定模板分子的所述第一通用样品索引序列和所述第二通用样品索引序列进行指定,由此识别所述给定插入序列的所述第一样品源。

66. 根据权利要求59所述的方法,其进一步包括:

- g) 对所述多个固定的模板分子的所述第三通用样品索引序列的其余碱基位置进行测序;
- h) 对所述多个固定的模板分子的所述第四通用样品索引序列进行测序;
- i) 对源自所述第二文库分子的所述多个固定的模板分子的所述插入序列区域进行测序;以及
- j) 将在步骤(c)中获得的给定模板分子的所述插入序列与来自同一给定模板分子的所述第三通用样品索引序列和所述第四通用样品索引序列进行指定,由此识别所述给定插入序列的第二样品源。

67. 根据权利要求59至66中任一项所述的方法,其中所述支持物包括玻璃或塑料基底。

68. 根据权利要求59至66中任一项所述的方法,其中所述支持物被配置在流通池通道、流通池或毛细管腔上。

69. 根据权利要求59至66中任一项所述的方法,其中所述支持物用具有不大于45度的

水接触角的至少一个亲水性聚合物涂层钝化。

70. 根据权利要求69所述的方法, 其中所述至少一个亲水性聚合物涂层包含选自以下项组成的组的分子: 聚乙二醇 (PEG)、聚(乙烯醇) (PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮) (PVP)、聚(丙烯酸) (PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺) (PNIPAM)、聚(甲基丙烯酸甲酯) (PMA)、聚(2-羟乙基甲基丙烯酸酯) (HEMA)、聚(低聚(乙二醇) 甲基醚甲基丙烯酸酯) (POEGMA)、聚谷氨酸 (PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。

71. 根据权利要求69所述的方法, 其中所述至少一个亲水性聚合物涂层包含具有至少四个支链的支链亲水性聚合物分子。

72. 根据权利要求69所述的方法, 其中所述至少一个亲水性聚合物涂层包含具有至少1000道尔顿的分子量的聚合物分子。

73. 根据权利要求59至66中任一项所述的方法, 其中所述固定的模板分子包含具有所述一个样品索引和所述插入序列的串联重复序列的多个固定的多联体分子。

74. 根据权利要求59至66中任一项所述的方法, 其中所述固定的模板分子包含多个不同的成簇模板分子, 所述成簇模板分子具有所述插入序列的一个拷贝和所述一个样品索引的一个拷贝, 其中所述成簇模板分子经由桥式扩增产生。

75. 根据权利要求59至66中任一项所述的方法, 其中位于所述支持物上的随机位置处的所述固定的核酸模板分子的密度为 10^4 - 10^8 个/ mm^2 。

76. 根据权利要求59至66中任一项所述的方法, 其中所述插入序列的所述样品源是基因组DNA、双链cDNA或细胞游离循环DNA。

77. 根据权利要求59至66中任一项所述的方法, 其中所述可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。

78. 根据权利要求59至66中任一项所述的方法, 其中所述可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。

79. 根据权利要求59至66中任一项所述的方法, 其中所述可检测地标记的核苷酸试剂包含多价分子, 每个多价分子包含(1) 核心、(2) 多个核苷酸臂和(3) 至少一个荧光团, 其中各个核苷酸臂包含(i) 核心附接部分、(ii) 包含PEG部分的间隔基、(iii) 接头和(iv) 核苷酸单元, 其中所述核心附接至所述多个核苷酸臂, 其中所述间隔基附接至所述接头, 其中所述接头附接至所述核苷酸单元。

80. 根据权利要求59至66中任一项所述的方法, 其中结合至所述固定的模板分子的所述可检测地标记的核苷酸试剂包含与测序引物杂交以形成双链体的各个固定的模板分子, 并且所述双链体结合至聚合酶以形成复合聚合酶, 并且所述复合聚合酶结合至可检测地标记的核苷酸试剂。

81. 根据权利要求80所述的方法, 其中在适用于将可检测地标记的核苷酸试剂与所述复合聚合酶结合并将所述可检测地标记的核苷酸掺入到杂交的测序引物中的条件下, 所述复合聚合酶结合至所述可检测地标记的核苷酸试剂, 其中所述可检测地标记的核苷酸试剂包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团。

82. 根据权利要求80所述的方法, 其中在适用于将可检测地标记的核苷酸试剂与所述复合聚合酶结合并将所述可检测地标记的核苷酸掺入到杂交的测序引物中的条件下, 所述

复合聚合酶结合至所述可检测地标记的核苷酸试剂,其中所述可检测地标记的核苷酸试剂包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团。

83. 根据权利要求80所述的方法,其中在适用于将可检测地标记的核苷酸试剂与所述复合聚合酶结合的条件下,所述复合聚合酶结合至所述可检测地标记的核苷酸试剂,并且所述条件适用于抑制核苷酸掺入,其中所述可检测地标记的核苷酸试剂包含多价分子,所述多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分,(ii)包含PEG部分的间隔基,(iii)接头和(iv)核苷酸单元,其中所述核心附接至所述多个核苷酸臂,其中所述间隔基附接至所述接头,其中所述接头附接至所述核苷酸单元。

84. 根据权利要求80所述的方法,其中所述固定的模板分子包含固定的多联体分子,所述固定的多联体分子与多个测序引物杂交以在同一多联体分子上形成至少第一双链体和第二双链体,其中所述第一和双链体结合至第一聚合酶并且所述第二双链体结合至第二聚合酶,以形成第一复合聚合酶和第二复合聚合酶,并且其中所述方法包括:

a) 使多个多价分子与同一多联体模板分子上的所述第一复合聚合酶和所述第二复合聚合酶接触,其中各个多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中所述核心附接至所述多个核苷酸臂,并且所述间隔基附接至所述接头,并且所述接头附接至所述核苷酸单元,其中所述接触在适合于将来自所述多个多价分子中的单个多价分子与所述第一复合聚合酶和所述第二复合聚合酶结合的条件下进行,其中所述单个多价分子的第一核苷酸单元结合至包括与所述多联体模板分子的第一部分杂交的所述第一测序引物的所述第一复合聚合酶,由此形成第一结合复合物,并且其中所述单个多价分子的第二核苷酸单元结合至包括与所述多联体模板分子的第二部分杂交的第二测序引物的所述第二复合聚合酶,由此形成第二结合复合物,并且

- 其中结合至同一多价分子的所述第一结合复合物和所述第二结合复合物形成亲和力和复合物,

- 其中所述接触在适合于抑制所述第一结合复合物和所述第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化的掺入的条件下进行;

b) 检测所述同一多联体模板分子上的所述第一结合复合物和所述第二结合复合物;

c) 对从所述可检测地标记的多价分子发出的所述光学颜色信号进行成像,所述可检测地标记的多价分子在所述同一多联体模板分子上形成所述第一结合复合物和所述第二结合复合物;以及

d) 识别所述第一结合复合物中的所述第一核苷酸单元,由此确定所述多联体模板分子的所述第一部分的序列,并且识别所述第二结合复合物中的所述第二核苷酸单元,由此确定所述多联体模板分子的所述第二部分的序列。

85. 一种用于产生多个文库-夹板复合物(300)的方法,其包括:

a) 提供多个单链核酸文库分子(100),其中所述多个单链核酸文库中的各个文库分子包含以5'至3'顺序排列的区域:(i)第一左通用衔接子序列(120),其具有针对第一表面引物的结合序列;(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列;

(iii) 目的序列(110); (iv) 第二右通用衔接子序列(150), 其具有针对第二测序引物的结合序列; 以及 (v) 第一右通用衔接子序列(130), 其具有针对第二表面引物的结合序列;

b) 提供多个单链夹板链(200), 其中所述多个单链夹板链中的各个单链夹板链(200) 包含以5'至3'顺序排列的区域: (i) 具有通用结合序列的第一区域(210), 其与线性单链文库分子(100)的所述第一左通用衔接子序列(120)杂交, 以及(ii)具有通用结合序列的第二区域(220), 其与所述线性单链文库分子(100)的所述第一右通用衔接子序列(130)杂交;

c) 在适合于使所述单链夹板链(200)的所述第一区域(210)与所述线性单链文库分子(100)的所述第一左通用衔接子序列(120)杂交并且适合于使所述单链夹板链(200)的所述第二区域(220)与所述线性单链文库分子(100)的所述第一右通用衔接子序列(130)杂交的条件下, 使所述多个单链夹板链(200)与所述多个单链核酸文库分子(100)杂交, 由此使各个文库分子成环以形成多个文库-夹板复合物(300), 所述多个文库-夹板复合物在所述文库分子的末端5'与3'端之间具有缺口, 其中所述缺口是可酶促连接的。

86. 根据权利要求85所述的方法, 其中所述多个单链核酸文库分子(100)进一步包含第一左索引序列(160)和/或第一右索引序列(170)。

87. 根据权利要求85所述的方法, 其中所述多个单链核酸文库分子(100)进一步包含第一左唯一识别序列(180)和/或第一右唯一识别序列(190)。

88. 根据权利要求85所述的方法, 其进一步包括: d) 连接各个文库-夹板复合物(300)中的所述缺口, 由此产生多个共价闭合环状文库分子(400)。

89. 根据权利要求88所述的方法, 其进一步包括: e) 使所述多个共价闭合环状文库分子(400)与至少一种核酸外切酶接触, 以移除所述多个单链夹板链(200)并保留所述多个共价闭合环状文库分子(400)。

90. 根据权利要求89所述的方法, 其中所述多个共价闭合环状文库分子中的各个共价闭合环状文库分子(400)包含具有针对第二表面引物的结合序列的第一右通用衔接子序列(130), 并且其中所述方法进一步包括: f) 在适用于使各个共价闭合环状文库分子(400)与各个固定的第二表面引物杂交的条件下, 将所述多个共价闭合环状文库分子(400)分配到具有固定在支持物上的多个所述第二表面引物的所述支持物上, 由此将所述多个共价闭合环状文库分子(400)固定到所述支持物。

91. 根据权利要求90所述的方法, 其进一步包括: g) 在适合于使用多个第二表面引物作为固定的扩增引物并使用所述多个共价闭合环状文库分子(400)作为模板分子在所述支持物上进行滚环扩增反应的条件下, 使多个固定的共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触, 由此产生多个固定的核酸多联体分子。

92. 根据权利要求91所述的方法, 其中所述多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。

93. 根据权利要求90所述的方法, 其中所述支持物上的多个固定的第二表面捕获引物位于所述支持物上的预定或随机位置处。

94. 根据权利要求90所述的方法, 其中所述支持物上的多个固定的第二表面捕获引物彼此流体连通, 以允许试剂(例如, 酶、核苷酸、二价阳离子等)的溶液流动到所述支持物上, 使得多个固定的第二表面引物可以基本上同时以大规模并行方式与所述试剂反应。

95. 根据权利要求91所述的方法, 其中所述支持物上的所述多个固定的核酸多联体分

子的密度为 10^4 - 10^8 个/ mm^2 。

96. 根据权利要求91所述的方法,其进一步包括:对多个固定的多联体进行测序,其中所述测序包括:

a) 使多个固定的多联体分子与 (i) 多个测序聚合酶和 (ii) 多个可溶性测序引物接触,其中所述接触在适合于形成各自包含结合至核酸双链体的测序聚合酶的多个复合聚合酶的条件下进行,其中所述核酸双链体包含与可溶性测序引物杂交的多联体分子;

b) 在适用于将至少一个核苷酸与复合测序聚合酶结合的条件下,使多个复合测序聚合酶与多个核苷酸接触,其中所述多个核苷酸包括至少一个用荧光团标记并且在糖3'位置处具有可移除链终止部分的核苷酸类似物;

c) 将至少一个核苷酸掺入到杂交的测序引物的3'端中,由此产生多个新生的延伸的测序引物;以及

d) 检测经掺入的核苷酸并且识别所述经掺入的核苷酸的核碱基。

97. 根据权利要求96所述的方法,其中所述多个核苷酸在3'糖基处包含可移除链终止部分,其中所述可移除链终止部分包含烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基团、叠氮基基团、O-叠氮基甲基基团、胺基团、酰胺基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团,并且其中所述可移除链终止部分可用化学化合物裂解以在所述糖基团上产生可延伸的3'OH部分。

98. 根据权利要求96所述的方法,其中所述多个核苷酸包括选自dATP、dGTP、dCTP、dTTP和dUTP组成的组的一个类型的核苷酸。

99. 根据权利要求96所述的方法,其中所述多个核苷酸包括选自dATP、dGTP、dCTP、dTTP和/或dUTP组成的组的两个或更多个类型的核苷酸的任意组合的混合物。

100. 根据权利要求91所述的方法,其进一步包括:对多个固定的多联体进行测序,其中所述测序包括:

a) 使多个固定的多联体分子与 (i) 多个测序聚合酶和 (ii) 多个可溶性通用测序引物接触,其中所述接触在适合于形成各自包含结合至核酸双链体的测序聚合酶的多个第一复合聚合酶的条件下进行,其中所述核酸双链体包含与可溶性测序引物杂交的多联体分子;

b) 在适用于将多价分子的互补核苷酸单元与所述多个第一复合聚合酶中的至少两个进行结合以形成多个多价-复合聚合酶的条件下,使多个复合测序聚合酶与多个可检测地标记的多价分子接触,由此形成多个多价-复合聚合酶,并且所述条件抑制所述互补核苷酸单元掺入到所述多个多价-复合聚合酶的测序引物中,其中多个多价分子中的各个多价分子包含附接至多个核苷酸臂的核心,并且每个核苷酸臂附接至核苷酸单元;

c) 检测所述多个多价-复合聚合酶;以及

d) 识别结合至所述多个多价-复合聚合酶中的所述多个第一复合聚合酶的所述互补核苷酸单元的核碱基,由此确定核酸模板的序列。

101. 根据权利要求100所述的方法,其进一步包括:

e) 解离所述多个多价-复合聚合酶并移除多个第一测序聚合酶及其结合的多价分子,并且保留多个核酸双链体;

f) 使步骤(e)的多个保留的核酸双链体与多个第二测序聚合酶接触,其中所述接触在

适用于将所述多个第二测序聚合酶与所述多个保留的核酸双链体结合的条件下进行,由此形成各自包含结合至保留的核酸双链体的第二测序聚合酶的多个第二复合聚合酶;

g) 使多个第二复合聚合酶与多个核苷酸接触,所述多个核苷酸包括至少一个在糖3'位置处具有可移除链终止部分的核苷酸类似物,其中所述接触在适用于将来自多个核苷酸的互补核苷酸与步骤(f)的所述第二复合聚合酶中的至少两个结合的条件下进行,由此形成多个核苷酸-复合聚合酶,并且所述条件适用于促进结合的互补核苷酸掺入到所述核苷酸-复合聚合酶的所述测序引物中。

102. 根据权利要求101所述的方法,其进一步包括:

h) 检测被掺入到所述核苷酸-复合聚合酶的所述测序引物中的所述互补核苷酸。

103. 根据权利要求101所述的方法,其进一步包括:

h) 检测被掺入到所述核苷酸-复合聚合酶的所述测序引物中的所述互补核苷酸;以及

i) 识别被掺入到所述核苷酸-复合聚合酶的所述测序引物中的所述互补核苷酸的核碱基。

104. 一种通过形成至少一种亲和力复合物进行测序的方法,其包括:

a) 将第一通用测序引物、第一测序聚合酶和第一可检测地标记的多价分子与权利要求16的多联体分子的第一部分结合,由此形成第一结合复合物,其中第一多价分子的第一核苷酸单元与所述第一测序聚合酶结合;

b) 将第二通用测序引物、第二测序聚合酶和所述第一可检测地标记的多价分子与同一多联体分子的第二部分结合,由此形成第二结合复合物,其中所述第一多价分子的第二核苷酸单元与所述第二测序聚合酶结合,其中包括同一多价分子的所述第一结合复合物和所述第二结合复合物形成亲和力复合物,其中所述第一可检测地标记的多价分子包含附接至多个核苷酸臂的核心,并且每个核苷酸臂附接至核苷酸单元,其中所述多联体分子包含目的序列(110)的两个或更多个串联重复序列以及结合所述第一通用测序引物和所述第二通用测序引物的通用引物结合位点,并且其中所述接触在适合于抑制所述第一结合复合物和所述第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化掺入的条件下进行;以及

c) 检测所述同一多联体分子上的所述第一结合复合物和所述第二结合复合物。

105. 根据权利要求100所述的方法,其中附接至所述各个多价分子的所述核心的所述多个核苷酸臂具有相同类型的核苷酸单元,并且其中所述类型的核苷酸单元选自由以下项组成的组:dATP、dGTP、dCTP、dTTP和dUTP。

106. 根据权利要求100所述的方法,其中所述多个多价分子包括两个或更多个类型的多价分子的任意组合的混合物,每个类型具有选自由以下项组成的组的核苷酸单元:dATP、dGTP、dCTP、dTTP和/或dUTP。

107. 根据权利要求101所述的方法,其中所述多个核苷酸在3'糖基处包含可移除链终止部分,其中所述可移除链终止部分包含烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基团、叠氮基基团、O-叠氮基甲基基团、胺基团、酰胺基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团,并且其中所述可移除链终止部分可用化学化合物裂解以在所述糖基团上产生可延伸的3'OH部分。

108. 根据权利要求101所述的方法, 其中所述多个核苷酸包括选自由dATP、dGTP、dCTP、dTTP和dUTP组成的组的一个类型的核苷酸。

109. 根据权利要求101所述的方法, 其中所述多个核苷酸包括选自由dATP、dGTP、dCTP、dTTP和/或dUTP组成的组的两个或更多个类型的核苷酸的任意组合的混合物。

110. 根据权利要求90所述的方法, 其中所述支持物包括玻璃或塑料基底。

111. 根据权利要求90所述的方法, 其中所述支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。

112. 根据权利要求111所述的方法, 其中所述至少一个亲水性聚合物涂层包含选自由以下项组成的组的分子: 聚乙二醇 (PEG)、聚(乙烯醇) (PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮) (PVP)、聚(丙烯酸) (PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺) (PNIPAM)、聚(甲基丙烯酸甲酯) (PMA)、聚(2-羟乙基甲基丙烯酸酯) (HEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯) (POEGMA)、聚谷氨酸 (PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。

单链夹板链和使用方法

[0001] 相关申请

[0002] 本申请要求2022年3月4日提交的美国临时专利申请号63/316,790和2022年4月20日提交的美国专利申请号17/725,042的优先权并要求其权益,这些申请中每一件的内容通过引用整体并入本文。

[0003] 通过引用并入序列表

[0004] 电子序列表的内容(ELEM-002_001W0_SequenceListing_ST26.xml;大小:68,293字节;以及创建日期:2023年3月3日)通过引用以其整体并入本文。

技术领域

[0005] 本公开提供了包含核酸单链夹板链的组合物以及使用单链夹板链的方法。单链夹板链可以与文库分子的部分杂交以形成具有缺口的文库-夹板复合物,其中缺口可以被连接以形成共价闭合环状分子,该共价闭合环状分子经历下游扩增和测序工作流程。

背景技术

[0006] 本公开涉及从共价闭合环状分子文库制备多核苷酸并对其进行测序。二代测序技术的改进极大地提高了测序速度和数据输出,由此导致当前测序平台的高样品通量。有效制备具有靶序列的闭合环状文库分子对于下游扩增和测序工作流程非常重要。提高测序通量的另一个方面是在文库制备过程中向DNA片段添加唯一索引序列,这允许在每次测序运行期间汇集大量文库并同时进行测序。因此,需要用于产生和测序含有靶序列和唯一索引序列的环状文库分子的替代方法,其与下游二代测序技术兼容。本文提供了满足这种需要的组合物、方法和试剂盒。

发明内容

[0007] 本公开提供了文库-夹板复合物(300),其包含:(i)单链核酸文库分子(100),其包括目的序列(110),该目的序列在一侧上侧接至少第一左通用衔接子序列(120),并且在另一侧上侧接至少第一右通用衔接子序列(130);以及(ii)单链夹板链(200),其包括第一区域(210),该第一区域具有针对文库分子的第一左通用衔接子序列(120)的通用结合序列,并且该单链夹板链进一步包含第二区域(220),该第二区域具有针对文库分子的第一右通用衔接子序列(130)的通用结合序列,由此使文库分子成环以产生文库-夹板复合物(300),该文库-夹板复合物在文库分子的5'端与文库分子的3'端之间具有缺口。

[0008] 在一些实施例中,核酸文库分子(100)进一步包含:第二左通用衔接子序列(140)。在一些实施例中,核酸文库分子(100)进一步包含:第二右通用衔接子序列(150)。在一些实施例中,核酸文库分子(100)进一步包含:第一左样品索引序列(160)。在一些实施例中,核酸文库分子(100)进一步包含:第一右样品索引序列(170)。在一些实施例中,核酸文库分子(100)进一步包含:第一左唯一识别序列(180)。在一些实施例中,核酸文库分子(100)进一步包含:第一右唯一识别序列(190)。

[0009] 在一些实施例中,核酸文库分子(100)进一步包含以下任意一者或者两者或更多者的任意组合:(i)第二左通用衔接子序列(140);(ii)第二右通用衔接子序列(150);(iii)第一左索引序列(160);(iv)第一右索引序列(170);(v)第一左唯一识别序列(180);和/或(vi)第一右唯一识别序列(190)。

[0010] 在一些实施例中,第一左通用衔接子序列(120)和/或第二左通用衔接子序列(140)包括:(i)针对正向测序引物的通用结合序列;(ii)针对反向测序引物的通用结合序列;(iii)针对第一表面引物的通用结合序列;(iv)针对第二表面引物的通用结合序列;(v)针对正向扩增引物的通用结合序列;(vi)针对反向扩增引物的通用结合序列;和/或(vii)针对压实寡核苷酸的通用结合序列。

[0011] 在一些实施例中,第一右通用衔接子序列(130)和/或第二右通用衔接子序列(150)包括:(i)针对正向测序引物的通用结合序列;(ii)针对反向测序引物的通用结合序列;(iii)针对第一表面引物的通用结合序列;(iv)针对第二表面引物的通用结合序列;(v)针对正向扩增引物的通用结合序列;(vi)针对反向扩增引物的通用结合序列;和/或(vii)针对压实寡核苷酸的通用结合序列。

[0012] 本公开提供了一种用于产生多个共价闭合环状文库分子(400)的方法,其包括:a)提供多个单链核酸文库分子(100),其中多个单链核酸文库分子中的各个文库分子包括目的序列(110),该目的序列在一侧上侧接至少第一左通用衔接子序列(120),并且在另一侧上侧接至少第一右通用衔接子序列(130);b)提供多个单链夹板链(200),其中该多个单链夹板链中的各个单链夹板链(200)包括第一区域(210)和第二区域(220),该第一区域能够与各个文库分子的至少第一左通用衔接子序列(120)杂交,该第二区域能够与各个文库分子的至少第一右通用衔接子序列(130)杂交;c)使多个单链夹板链(200)与多个单链核酸文库分子(100)杂交,使得单链夹板链(210)中的一个的第一区域退火至文库分子的至少第一左通用衔接子序列(120),并且使得单链夹板链(220)的第二区域退火至文库分子的至少第一右通用序列(130),由此使各个文库分子成环以形成多个文库-夹板复合物(300),该文库-夹板复合物在文库分子的末端5'与3'端之间具有缺口,其中该缺口是可酶促连接的;以及d)连接多个文库-夹板复合物(300)中的缺口,由此产生多个共价闭合环状文库分子(400)。

[0013] 在一些实施例中,该方法进一步包括:(e)在适用于使各个共价闭合环状文库分子(400)与各个固定的表面引物杂交的条件下,将多个共价闭合环状文库分子(400)分配到支持物上,该支持物具有固定在支持物上的多个表面引物,由此将多个共价闭合环状文库分子(400)固定到支持物。

[0014] 在一些实施例中,该方法进一步包括:(f)在适合于使用多个表面引物作为固定的扩增引物并使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个固定的共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,由此产生多个固定的核酸多联体分子。

[0015] 在一些实施例中,该方法进一步包括:对多个固定的核酸多联体分子进行测序。在一些实施例中,测序包括:a)使多个固定的多联体分子与(i)多个测序聚合酶和(ii)多个可溶性测序引物接触,其中该接触在适合于形成各自包含结合至核酸双链体的测序聚合酶的多个复合聚合酶的条件下进行,其中核酸双链体包含与可溶性测序引物杂交的多联体分

子;b)在适用于将至少一个核苷酸与复合测序聚合酶结合的条件下,使多个复合测序聚合酶与多个核苷酸接触,其中该多个核苷酸包括至少一个用荧光团标记并在糖3'位置处具有可移除链终止部分的核苷酸类似物;c)将至少一个核苷酸掺入到杂交的测序引物的3'端中,由此产生多个新生的延伸的测序引物;以及d)检测掺入的核苷酸并识别掺入的核苷酸的核碱基。

[0016] 在一些实施例中,测序包括:a)使多个固定的多联体分子与(i)多个测序聚合酶和(ii)多个可溶性测序引物接触,其中该接触在适合于形成各自包含结合至核酸双链体的测序聚合酶的多个第一复合聚合酶的条件下进行,其中该核酸双链体包含与可溶性测序引物杂交的多联体分子;b)在适用于将多价分子的互补核苷酸单元与多个第一复合聚合酶中的至少两个结合以形成多个多价-复合聚合酶的条件下,使多个复合测序聚合酶与多个可检测地标记的多价分子接触,由此形成多个多价-复合聚合酶,并且该条件抑制互补核苷酸单元掺入到多个多价-复合聚合酶的测序引物中,其中多个多价分子中的各个多价分子包含附接至多个核苷酸臂的核心,并且每个核苷酸臂附接至核苷酸单元;c)检测多个多价-复合聚合酶;以及d)识别结合至多个多价-复合聚合酶中的多个第一复合聚合酶的互补核苷酸单元的核碱基,由此确定核酸模板的序列。

[0017] 在一些实施例中,测序进一步包括:e)解离多个多价-复合聚合酶并移除多个第一测序聚合酶及其结合的多价分子,并保留多个核酸双链体;f)使步骤(e)的多个保留的核酸双链体与多个第二测序聚合酶接触,其中该接触在适用于将多个第二测序聚合酶与多个保留的核酸双链体结合的条件下进行,由此形成各自包含结合至保留的核酸双链体的第二测序聚合酶的多个第二复合聚合酶;g)使多个第二复合聚合酶与多个核苷酸接触,其中该接触在适用于使多个核苷酸中的互补核苷酸与步骤(f)的第二复合聚合酶中的至少两个结合的条件下进行,由此形成多个核苷酸-复合聚合酶,并且该条件适用于促进结合的互补核苷酸掺入到核苷酸-复合聚合酶的测序引物中。

[0018] 在一些实施例中,测序进一步包括:h)检测掺入到核苷酸-复合聚合酶的测序引物中的互补核苷酸。在一些实施例中,测序进一步包括:h)检测掺入到核苷酸-复合聚合酶的测序引物中的互补核苷酸;以及i)识别掺入到核苷酸-复合聚合酶的测序引物中的互补核苷酸的核碱基。

[0019] 本公开提供了一种用于形成至少一种亲和力复合物的方法,其包括:a)将第一通用核酸引物、第一DNA聚合酶和第一多价分子与根据权利要求16的多联体分子的第一部分结合,由此形成第一结合复合物,其中第一多价分子的第一核苷酸单元与第一DNA聚合酶结合;以及b)将第二通用核酸引物、第二DNA聚合酶和第一多价分子与同一多联体模板分子的第二部分结合,由此形成第二结合复合物,其中第一多价分子的第二核苷酸单元与第二DNA聚合酶结合,其中包括同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,其中第一多价分子包含附接至多个核苷酸臂的核心并且每个核苷酸臂附接至核苷酸单元,并且其中多联体分子包含目的序列(110)的两个或更多个串联重复序列以及结合第一通用核酸引物和第二通用核酸引物的通用引物结合位点。

[0020] 本公开提供了一种通过形成至少一种亲和力复合物进行测序的方法,其包括:a)将第一通用核酸引物、第一DNA聚合酶和第一多价分子与权利要求16的多联体分子的第一部分结合,由此形成第一结合复合物,其中第一多价分子的第一核苷酸单元与第一DNA聚合

酶结合;b)将第二通用核酸引物、第二DNA聚合酶以及第一多价分子与同一多联体模板分子的第二部分结合,由此形成第二结合复合物,其中第一多价分子的第二核苷酸单元与第二DNA聚合酶结合,其中包括同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,其中第一多价分子包含附接至多个核苷酸臂的核心并且每个核苷酸臂附接至核苷酸单元,其中多联体分子包含目的序列(110)的两个或更多个串联重复序列以及结合第一通用核酸引物和第二通用核酸引物的通用引物结合位点,并且其中该接触在适合于抑制第一结合复合物和第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化掺入的条件下进行;c)检测同一多联体模板分子上的第一结合复合物和第二结合复合物,以及d)识别第一结合复合物中的第一核苷酸单元,由此确定多联体模板分子的第一部分的序列,并且识别第二结合复合物中的第二核苷酸单元,由此确定多联体模板分子的第二部分的序列。

[0021] 在一些实施例中,核酸文库分子(100)进一步包含:第二左通用衔接子序列(140)。在一些实施例中,核酸文库分子(100)进一步包含:第二右通用衔接子序列(150)。在一些实施例中,核酸文库分子(100)进一步包含:第一左样品索引序列(160)。在一些实施例中,核酸文库分子(100)进一步包含:第一右样品索引序列(170)。在一些实施例中,核酸文库分子(100)进一步包含:第一左唯一识别序列(180)。在一些实施例中,核酸文库分子(100)进一步包含:第一右唯一识别序列(190)。

[0022] 在一些实施例中,核酸文库分子(100)进一步包含以下任意一者或者两者或更多者的任意组合:(i)第二左通用衔接子序列(140);(ii)第二右通用衔接子序列(150);(iii)第一左索引序列(160);(iv)第一右索引序列(170);(v)第一左唯一识别序列(180);和/或(vi)第一右唯一识别序列(190)。

[0023] 在一些实施例中,第一左通用衔接子序列(120)和/或第二左通用衔接子序列(140)包括:(i)针对正向测序引物的通用结合序列;(ii)针对反向测序引物的通用结合序列;(iii)针对第一表面引物的通用结合序列;(iv)针对第二表面引物的通用结合序列;(v)针对正向扩增引物的通用结合序列;(vi)针对反向扩增引物的通用结合序列;和/或(vii)针对压实寡核苷酸的通用结合序列。

[0024] 在一些实施例中,第一右通用衔接子序列(130)和/或第二右通用衔接子序列(150)包括:(i)针对正向测序引物的通用结合序列;(ii)针对反向测序引物的通用结合序列;(iii)针对第一表面引物的通用结合序列;(iv)针对第二表面引物的通用结合序列;(v)针对正向扩增引物的通用结合序列;(vi)针对反向扩增引物的通用结合序列;和/或(vii)针对压实寡核苷酸的通用结合序列。

[0025] 本公开提供了一种用于对核酸分子进行测序的方法,其包括:a)提供固定在支持物上的随机位置处的多个核酸模板分子,其中各个固定的模板分子包含插入序列区域和一个样品索引,其中每个样品索引包含与识别插入序列的样品源的通用样品索引序列接合的3聚体随机序列,其中不同的固定的模板分子具有不同的3聚体随机序列和相同的通用样品索引序列,并且其中固定的模板分子具有不同的插入序列;b)使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂对多个固定的模板分子的3聚体随机序列进行三个循环的聚合酶介导的测序反应,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的模板分

子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环和第三测序循环中的每一者中,在多个固定的模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及c) 使用在步骤(b)中获得的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。在一些实施例中,步骤b)的平衡多样性是在第一测序循环、第二测序循环和第三测序循环的每一者中检测和成像的核碱基A、G、C和T/U中的每一者的25-75%。

[0026] 在一些实施例中,该方法进一步包括:d)对多个固定的模板分子的通用样品索引序列进行测序;e)对多个固定的模板分子的插入序列区域进行测序;以及f)将在步骤(b)中获得的给定模板分子的插入序列与在步骤(a)中获得的来自同一给定模板分子的通用样品索引序列进行指定,由此识别给定插入序列的样品源。

[0027] 在一些实施例中,多个核酸模板分子进一步包含第二样品索引,该第二样品索引包含识别插入序列的样品源的第二通用样品索引序列,并且第二样品索引缺少随机序列。

[0028] 在一些实施例中,该方法进一步包括:g)对多个固定的模板分子的3聚体随机序列进行测序,以获得在第一测序循环、第二测序循环和第三测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性,从而生成多个固定的模板分子的位置的图;h)对多个固定的模板分子的第一通用样品索引序列进行测序;i)对多个固定的模板分子的第二通用样品索引序列进行测序;j)对多个固定的模板分子的插入序列区域进行测序;以及k)将在步骤(c)中获得的给定模板分子的插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定的插入序列的样品源。

[0029] 本公开提供了一种用于对核酸分子进行测序的方法,其包括:a)提供固定在支持物上的随机位置处的多个核酸模板分子,其中各个固定的模板分子包含插入序列区域和一个样品索引,其中每个样品索引包含与识别插入序列的样品源的通用样品索引序列接合的3聚体随机序列,其中通用样品索引序列包含3-20个核苷酸,其中不同的固定的模板分子具有不同的3聚体随机序列和相同的通用样品索引序列,并且其中固定的模板分子具有不同的插入序列;b)使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂对多个固定的模板分子的3聚体随机序列和通用样品索引序列的第一碱基位置进行四个循环的聚合酶介导的测序反应,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中四个循环的测序包括对从结合至固定的模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的通用样品索引序列的第一碱基位置和3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中,在多个固定的模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;;以及c)使用步骤(b)中获得的四个循环的聚合酶介导的测序反应的图像生成多个固定的模板分子的位置图,其中插入区域的序列不用于生成图。在一些实施例中,步骤(b)的平衡多样性是在第一测序循环、第二测序循环、第三测序循环和第四测序循环的每一者中检测和成像的核碱基A、G、C和T/U的25-75%。

[0030] 在一些实施例中,该方法进一步包括:a)对多个固定的模板分子的通用样品索引

序列的其余碱基位置进行测序;b)对多个固定的模板分子的插入序列区域进行测序;以及c)将在步骤(b)中获得的给定模板分子的插入序列与在步骤(a)中获得的来自同一给定模板分子的通用样品索引序列进行指定,由此识别给定插入序列的样品源。在一些实施例中,多个核酸模板分子进一步包含第二样品索引,该第二样品索引包含识别插入序列的样品源的第二通用样品索引序列,并且第二样品索引缺少随机序列。

[0031] 在一些实施例中,该方法进一步包括:a)对多个固定的模板分子的3聚体随机序列和通用样品索引序列的第一碱基位置进行测序,以获得在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性,从而生成多个固定的模板分子的位置的图;b)对多个固定的模板分子的第一通用样品索引序列的其余碱基位置进行测序;c)对多个固定的模板分子的第二通用样品索引序列进行测序;d)对多个固定的模板分子的插入序列区域进行测序;以及e)将在步骤(c)中获得的给定模板分子的插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定的插入序列的样品源。

[0032] 在一些实施例中,支持物包括玻璃或塑料基底。在一些实施例中,支持物被配置在流通池通道、流通池或毛细管腔上。在一些实施例中,支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。

[0033] 在一些实施例中,至少一个亲水性聚合物涂层包含选自以下项组成的组的分子:聚乙二醇(PEG)、聚(乙烯醇)(PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮)(PVP)、聚(丙烯酸)(PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺)(PNIPAM)、聚(甲基丙烯酸甲酯)(PMA)、聚(2-羟乙基甲基丙烯酸酯)(HEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(POEGMA)、聚谷氨酸(PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少四个支链的支链亲水性聚合物分子。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少1000道尔顿的分子量的聚合物分子。

[0034] 在一些实施例中,固定的模板分子包含具有一个样品索引和插入序列的串联重复序列的多个固定的多联体分子。在一些实施例中,固定的模板分子包含多个不同的成簇模板分子,该成簇模板分子具有插入序列的一个拷贝和一个样品索引的一个拷贝,其中成簇模板分子经由桥式扩增产生。

[0035] 在一些实施例中,位于支持物上的随机位置处的固定的核酸模板分子的密度为 10^4 - 10^8 个/ mm^2 。在一些实施例中,插入序列的样品源是基因组DNA、双链cDNA或细胞游离循环DNA。

[0036] 在一些实施例中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。在一些实施例中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。

[0037] 在一些实施例中,可检测地标记的核苷酸试剂包含多价分子,每个多价分子包含(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0038] 在一些实施例中,结合至固定的模板分子的可检测地标记的核苷酸试剂包含与测

序引物杂交以形成双链体的各个固定的模板分子,并且双链体结合至聚合酶以形成复合聚合酶,并且复合聚合酶结合至可检测地标记的核苷酸试剂。

[0039] 在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团。

[0040] 在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,并且该条件适用于抑制核苷酸掺入,其中可检测地标记的核苷酸试剂包含多价分子,该多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0041] 在一些实施例中,固定的模板分子包括固定的多联体分子,该多联体分子与多个测序引物杂交以在同一多联体分子上形成至少第一双链体和第二双链体,其中第一和双链体结合至第一聚合酶结合,并且第二双链体结合至第二聚合酶以形成第一复合聚合酶和第二复合聚合酶,并且该方法包括:a)使多个多价分子与同一多联体模板分子上的第一复合聚合酶和第二复合聚合酶接触,其中各个多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包括(i)核心附接部分,(ii)包含PEG部分的间隔基,(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,并且间隔基附接至接头,并且接头附接至核苷酸单元,其中接触在适用于将多个多价分子中的单个多价分子与第一复合聚合酶和第二复合聚合酶结合的条件下进行,其中单个多价分子的第一核苷酸单元与第一复合聚合酶结合,该第一复合酶包括与多联体模板分子的第一部分杂交的第一测序引物,由此形成第一结合复合物,并且其中单个多价分子的第二核苷酸单元与第二复合聚合酶结合,该第二复合聚合酶包括与多联体模板的第二部分杂交的第二测序引物,由此形成第二结合复合物,并且其中结合至同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,其中接触在适用于抑制第一结合复合物和第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化的掺入的条件下进行;b)检测同一多联体模板分子上的第一结合复合物和第二结合复合物;c)对从可检测地标记的多价分子发出的光学颜色信号进行成像,可检测地标记的多价分子在同一多联体模板分子上形成第一结合复合物和第二结合复合物;以及d)识别第一结合复合物中的第一核苷酸单元,由此确定多联体模板分子的第一部分的序列,并且识别第二结合复合物中的第二核苷酸单元,由此确定多联体模板分子的第二部分的序列。

[0042] 本公开提供了一种用于对核酸分子进行测序的方法,其包括:a)提供第一多个文库分子,该多个文库分子中的每个分子包含(i)源自第一样品源的插入序列区域,(ii)第一样品索引,其具有与第一通用样品索引序列接合的3聚体随机序列,以及(iii)具有第二通用样品索引序列的第二样品索引,该第二样品索引缺少随机序列,其中第一通用样品索引

序列和第二通用样品索引序列的组合唯一地识别插入序列的第一样品源,其中不同的第一文库分子具有不同的3聚体随机序列并且具有不同的插入序列;b)提供第二多个文库分子,该多个文库分子中的每个分子包含(i)源自第二样品源的插入序列区域,(ii)第三样品索引,其具有与第三通用样品索引序列接合的3聚体随机序列,以及(iii)具有第四通用样品索引序列的第四样品索引,该第四样品索引缺少随机序列,其中第三通用样品索引序列和第四通用样品索引序列的组合唯一地识别插入序列的第二样品源,其中不同的第二文库分子具有不同的3聚体随机序列并且具有不同的插入序列;c)汇集第一多个文库分子和第二多个文库分子;d)将经汇集的文库分子分配到支持物上并进行扩增反应以产生固定至支持物的随机位置处的多个克隆扩增的模板分子;(e)使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂,对第一样品索引和第三样品索引的3聚体随机序列进行三个循环的聚合酶介导的测序反应,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的扩增模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环和第三测序循环的每一者中,在多个固定的扩增模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及f)使用在步骤(e)中获得的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。在一些实施例中,步骤(e)的平衡多样性是在第一测序循环、第二测序循环和第三测序循环的每一者中检测和成像的核碱基A、G、C和T/U中的每一者的25-75%。

[0043] 在一些实施例中,该方法进一步包括:g)对多个固定的模板分子的第一通用样品索引序列进行测序;h)对多个固定的模板分子的第二通用样品索引序列进行测序;i)对源自第一文库分子的多个固定的模板分子的插入序列区域进行测序;以及j)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定插入序列的第一样品源。

[0044] 在一些实施例中,该方法进一步包括:g)对多个固定的模板分子的第三通用样品索引序列进行测序;h)对多个固定的模板分子的第四通用样品索引序列进行测序;i)对源自第二文库分子的多个固定的模板分子的插入序列区域进行测序;以及j)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第三通用样品索引序列和第四通用样品索引序列进行指定,由此识别给定插入序列的第二样品源。

[0045] 本公开提供了一种用于对核酸分子进行测序的方法,其包括:a)提供第一多个文库分子,该多个文库分子中的每个分子包含(i)源自第一样品源的插入序列区域,(ii)第一样品索引,其具有与第一通用样品索引序列接合的3聚体随机序列,以及(iii)具有第二通用样品索引序列的第二样品索引,该第二样品索引缺少随机序列,其中第一通用样品索引序列和第二通用样品索引序列的组合唯一地识别插入序列的第一样品源,其中第一通用样品索引序列包含3-20个核苷酸,其中不同的第一文库分子具有不同的3聚体随机序列并且具有不同的插入序列;b)提供第二多个文库分子,该多个文库分子中的每个分子包含(i)源自第二样品源的插入序列区域,(ii)第三样品索引,其具有与第三通用样品索引序列接合的3聚体随机序列,以及(iii)具有第四通用样品索引序列的第四样品索引,该第四样品索引缺少随机序列,其中第三通用样品索引序列和第四通用样品索引序列的组合唯一地识别

插入序列的第二样品源,其中第三通用样品索引序列包含3-20个核苷酸,其中不同的第二文库分子具有不同的3聚体随机序列并且具有不同的插入序列;c)汇集第一多个文库分子和第二多个文库分子;d)将经汇集的文库分子分配到支持物上并进行扩增反应以产生固定至支持物的随机位置处的多个克隆扩增的模板分子;e)使用包含不同类型的核碱基A、G、C和T/U的混合物的多个可检测地标记的核苷酸试剂对第一样品序列和第三样品索引的3聚体随机序列进行四个循环的聚合酶介导的测序反应,并对第一通用样品索引序列和第三通用样品索引序列的第一碱基位置进行测序,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的扩增模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中,在多个固定的扩增模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及f)使用在步骤(e)中获得的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。在一些实施例中,步骤(e)的平衡多样性是在第一测序循环、第二测序循环、第三测序循环和第四测序循环的每一者中检测和成像的核碱基A、G、C和T/U中的每一者的25-75%。

[0046] 在一些实施例中,该方法进一步包括:g)对多个固定的模板分子的第一通用样品索引序列的其余碱基位置进行测序;h)对多个固定的模板分子的第二通用样品索引序列进行测序;i)对源自第一文库分子的多个固定的模板分子的插入序列区域进行测序;以及j)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定插入序列的第一样品源。

[0047] 在一些实施例中,该方法进一步包括:g)对多个固定的模板分子的第三通用样品索引序列的其余碱基位置进行测序;h)对多个固定的模板分子的第四通用样品索引序列进行测序;i)对源自第二文库分子的多个固定的模板分子的插入序列区域进行测序;以及j)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第三通用样品索引序列和第四通用样品索引序列进行指定,由此识别给定插入序列的第二样品源。

[0048] 在一些实施例中,该方法进一步包括:a)对多个固定的模板分子的通用样品索引序列的其余碱基位置进行测序;b)对多个固定的模板分子的插入序列区域进行测序;以及c)将在步骤(b)中获得的给定模板分子的插入序列与在步骤(a)中获得的来自同一给定模板分子的通用样品索引序列进行指定,由此识别给定插入序列的样品源。在一些实施例中,多个核酸模板分子进一步包含第二样品索引,该第二样品索引包含识别插入序列的样品源的第二通用样品索引序列,并且第二样品索引缺少随机序列。

[0049] 在一些实施例中,该方法进一步包括:a)对多个固定的模板分子的3聚体随机序列和通用样品索引序列的第一碱基位置进行测序,以获得在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性,从而生成多个固定的模板分子的位置的图;b)对多个固定的模板分子的第一通用样品索引序列的其余碱基位置进行测序;c)对多个固定的模板分子的第二通用样品索引序列进行测序;d)对多个固定的模板分子的插入序列区域进行测序;以及e)将在步骤(c)中获得的给定模板分子的插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定的插入序列的样品源。

[0050] 在一些实施例中,支持物包括玻璃或塑料基底。在一些实施例中,支持物被配置在流通池通道、流通池或毛细管腔上。在一些实施例中,支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。

[0051] 在一些实施例中,至少一个亲水性聚合物涂层包含选自由以下项组成的组的分子:聚乙二醇(PEG)、聚(乙烯醇)(PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮)(PVP)、聚(丙烯酸)(PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺)(PNIPAM)、聚(甲基丙烯酸甲酯)(PMA)、聚(2-羟乙基甲基丙烯酸酯)(HEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(POEGMA)、聚谷氨酸(PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少四个支链的支链亲水性聚合物分子。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少1000道尔顿的分子量的聚合物分子。

[0052] 在一些实施例中,固定的模板分子包含具有一个样品索引和插入序列的串联重复序列的多个固定的多联体分子。在一些实施例中,固定的模板分子包含多个不同的成簇模板分子,该成簇模板分子具有插入序列的一个拷贝和一个样品索引的一个拷贝,其中成簇模板分子经由桥式扩增产生。

[0053] 在一些实施例中,位于支持物上的随机位置处的固定的核酸模板分子的密度为 10^4 - 10^8 个/ mm^2 。在一些实施例中,插入序列的样品源是基因组DNA、双链cDNA或细胞游离循环DNA。

[0054] 在一些实施例中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。在一些实施例中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。

[0055] 在一些实施例中,可检测地标记的核苷酸试剂包含多价分子,每个多价分子包含(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0056] 在一些实施例中,结合至固定的模板分子的可检测地标记的核苷酸试剂包含与测序引物杂交以形成双链体的各个固定的模板分子,并且双链体结合至聚合酶以形成复合聚合酶,并且复合聚合酶结合至可检测地标记的核苷酸试剂。

[0057] 在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团。

[0058] 在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,并且该条件适用于抑制核苷酸掺入,其中可检测地标记的核苷酸试剂包含多价分子,该多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间

隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0059] 在一些实施例中,固定的模板分子包括固定的多联体分子,该多联体分子与多个测序引物杂交以在同一多联体分子上形成至少第一双链体和第二双链体,其中第一和双链体结合至第一聚合酶结合,并且第二双链体结合至第二聚合酶以形成第一复合聚合酶和第二复合聚合酶,并且该方法包括:a)使多个多价分子与同一多联体模板分子上的第一复合聚合酶和第二复合聚合酶接触,其中各个多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包括(i)核心附接部分,(ii)包含PEG部分的间隔基,(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,并且间隔基附接至接头,并且接头附接至核苷酸单元,其中接触在适用于将多个多价分子中的单个多价分子与第一复合聚合酶和第二复合聚合酶结合的条件下进行,其中单个多价分子的第一核苷酸单元与第一复合聚合酶结合,该第一复合酶包括与多联体模板分子的第一部分杂交的第一测序引物,由此形成第一结合复合物,并且其中单个多价分子的第二核苷酸单元与第二复合聚合酶结合,该第二复合聚合酶包括与多联体模板的第二部分杂交的第二测序引物,由此形成第二结合复合物,并且其中结合至同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,其中接触在适用于抑制第一结合复合物和第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化的掺入的条件下进行;b)检测同一多联体模板分子上的第一结合复合物和第二结合复合物;c)对从可检测地标记的多价分子发出的光学颜色信号进行成像,可检测地标记的多价分子在同一多联体模板分子上形成第一结合复合物和第二结合复合物;以及d)识别第一结合复合物中的第一核苷酸单元,由此确定多联体模板分子的第一部分的序列,并且识别第二结合复合物中的第二核苷酸单元,由此确定多联体模板分子的第二部分的序列。

[0060] 本公开提供了用于产生多个文库-夹板复合物(300)的方法,其包括:a)提供多个单链核酸文库分子(100),其中多个单链核酸文库分子中的各个文库分子包含以5'至3'顺序排列的区域:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列;(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列;(iii)目的序列(110);(iv)第二右通用衔接子序列(150),其具有针对第二测序引物的结合序列;(v)第一右通用衔接子序列(130),其具有针对第一表面引物的结合序列;b)提供多个单链夹板链(200),其中多个单链夹板链中的各个单链夹板链(200)包含以5'至3'顺序排列的区域:(i)具有通用结合序列的第一区域(210),其与线性单链文库分子(100)的第一左通用衔接子序列(120)杂交,以及(ii)具有通用结合序列的第二区域(220),其与线性单链文库分子(100)的第一右通用衔接子序列(130)杂交;c)在适合于使单链夹板链(200)的第一区域(210)与线性单链文库分子(100)的第一左通用衔接子序列(120)杂交并且适合于使单链夹板链(200)的第二区域(220)与线性单链文库分子(100)的第一右通用衔接子序列(130)杂交的条件下,使多个单链夹板链(200)与多个单链核酸文库分子(100)杂交,由此使各个文库分子成环以形成多个文库-夹板复合物(300),该多个文库-夹板复合物在文库分子的末端5'与3'端之间具有缺口,其中该缺口是可酶促连接的。

[0061] 在一些实施例中,多个单链核酸文库分子(100)进一步包含第一左索引序列(160)和/或第一右索引序列(170)。在一些实施例中,多个单链核酸文库分子(100)进一步包含第

一左唯一识别序列(180)和/或第一右唯一识别序列(190)。

[0062] 在一些实施例中,该方法进一步包括:d)连接各个文库-夹板复合物(300)中的缺口,由此产生多个共价闭合的环状文库分子(400)。在一些实施例中,该方法进一步包括:e)使多个共价闭合环状文库分子(400)与至少一种核酸外切酶接触,以移除多个单链夹板链(200)并保留多个共价闭合环状文库分子(400)。

[0063] 在一些实施例中,多个共价闭合环状文库分子中的各个共价闭合环状文库分子(400)包含具有针对第一表面引物的结合序列的第一右通用衔接子序列(130),并且该方法进一步包括:f)在适用于使各个共价闭合环状文库分子(400)与各个固定的第二表面引物杂交的条件下,将多个共价闭合环状文库分子(400)分配到具有固定在支持物上的多个第一表面引物的支持物上,由此将多个共价闭合环状文库分子(400)固定到支持物。

[0064] 在一些实施例中,该方法进一步包括:g)在适合于使用多个第一表面引物作为固定的扩增引物并使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个固定的共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,由此产生多个固定的核酸多联体分子。在一些实施例中,多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。

[0065] 在一些实施例中,支持物上的多个固定的第一表面捕获引物位于支持物上的预定或随机位置处。在一些实施例中,支持物上的多个固定的第二表面捕获引物彼此流体连通,以允许试剂(例如,酶、核苷酸、二价阳离子等)的溶液流动到支持物上,使得多个固定的第一表面引物可以基本上同时以大规模并行方式与试剂反应。在一些实施例中,支持物上的多个固定的核酸多联体分子的密度为 10^4 - 10^8 个/ mm^2 。

[0066] 在一些实施例中,该方法进一步包括:a)对多个固定的多联体进行测序,其中测序包括:b)使多个固定的多联体分子与(i)多个测序聚合酶和(ii)多个可溶性测序引物接触,其中该接触在适合形成多个复合聚合酶的条件下进行,每个复合聚合酶包含结合至核酸双链体的测序聚合酶,其中该核酸双链体包含与可溶性测序引物杂交的多联体分子;c)在适用于将至少一个核苷酸与复合测序聚合酶结合的条件下,使多个复合测序聚合酶与多个核苷酸接触,其中多个核苷酸包括至少一个用荧光团标记并在糖3'位置处具有可移除链终止部分的核苷酸类似物;d)将至少一个核苷酸掺入杂交的测序引物的3'端,由此产生多个新生延伸测序引物;以及e)检测掺入的核苷酸并识别掺入的核苷酸的核碱基。

[0067] 在一些实施例中,多个核苷酸在3'糖基处包含可移除链终止部分,其中可移除链终止部分包含烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基团、叠氮基基团、0-叠氮基甲基基团、胺基团、酰胺基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团,并且其中可移除链终止部分可用化学化合物裂解以在糖基团上产生可延伸的3'OH部分。在一些实施例中,多个核苷酸包括选自dATP、dGTP、dCTP、dTTP和dUTP的一种类型的核苷酸。在一些实施例中,多个核苷酸包括选自由dATP、dGTP、dCTP、dTTP和/或dUTP组成的组的两个或更多个类型的核苷酸的任意组合的混合物。

[0068] 在一些实施例中,该方法进一步包括:对多个固定的多联体分子进行测序,其中测序包括:a)使多个固定的多联体分子与(i)多个测序聚合酶和(ii)多个可溶性通用测序引物接触,其中该接触在适合形成多个第一复合聚合酶的条件下进行,每个第一复合聚合酶

包含结合至核酸双链体的测序聚合酶,其中该核酸双链体包含与可溶性测序引物杂交的多联体分子;b)在适用于将多价分子的互补核苷酸单元与多个第一复合聚合酶中的至少两个结合以形成多个多价-复合聚合酶的条件下,使多个复合测序聚合酶与多个可检测地标记的多价分子接触,由此形成多个多价-复合聚合酶,并且该条件抑制互补核苷酸单元掺入到多个多价-复合聚合酶的测序引物中,其中多个多价分子中的各个多价分子包含附接至多个核苷酸臂的核心,并且每个核苷酸臂附接至核苷酸单元;c)检测多个多价-复合聚合酶;以及d)识别结合至多个多价-复合聚合酶中的多个第一复合聚合酶的互补核苷酸单元的核碱基,由此确定核酸模板的序列。

[0069] 在一些实施例中,该方法进一步包括:e)解离多个多价-复合聚合酶并移除多个第一测序聚合酶及其结合的多价分子,并保留多个核酸双链体;f)使步骤(e)的多个保留的核酸双链体与多个第二测序聚合酶接触,其中该接触在适用于将多个第二测序聚合酶与多个保留的核酸双链体结合的条件下进行,由此形成各自包含结合至保留的核酸双链体的第二测序聚合酶的多个第二复合聚合酶;g)使多个第二复合聚合酶与多个核苷酸接触,该多个核苷酸包括至少一个在糖3'位置处具有可移除链终止部分的核苷酸类似物,其中该接触在适用于将来自多个核苷酸的互补核苷酸与步骤(f)的第二复合聚合酶中的至少两个结合的条件下进行,由此形成多个核苷酸-复合聚合酶,并且该条件适用于促进结合的互补核苷酸掺入到核苷酸-复合聚合酶的测序引物中。

[0070] 在一些实施例中,该方法进一步包括:h)检测掺入到核苷酸-复合聚合酶的测序引物中的互补核苷酸。在一些实施例中,该方法进一步包括:h)检测掺入到核苷酸-复合聚合酶的测序引物中的互补核苷酸;以及i)识别掺入到核苷酸-复合聚合酶的测序引物中的互补核苷酸的核碱基。

[0071] 本公开提供了一种通过形成至少一种亲和力复合物进行测序的方法,其包括:a)将第一通用测序引物、第一测序聚合酶和第一可检测地标记的多价分子与权利要求16的多联体分子的第一部分结合,由此形成第一结合复合物,其中第一多价分子的第一核苷酸单元与第一测序聚合酶结合;b)将第二通用测序引物、第二测序聚合酶和第一可检测地标记的多价分子与同一多联体分子的第二部分结合,由此形成第二结合复合物,其中第一多价分子的第二核苷酸单元与第二测序聚合酶结合,其中包括同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,其中第一可检测地标记的多价分子包含附接至多个核苷酸臂的核心,并且每个核苷酸臂附接至核苷酸单元,其中多联体分子包含目的序列(110)的两个或更多个串联重复序列以及结合第一通用测序引物和第二通用测序引物的通用引物结合位点,并且其中该接触在适合于抑制第一结合复合物和第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化掺入的条件下进行;以及c)检测同一多联体分子上的第一结合复合物和第二结合复合物。

[0072] 在一些实施例中,附接至各个多价分子的核心的多个核苷酸臂具有相同类型的核苷酸单元,并且其中该核苷酸单元的类型选自由以下项组成的组:dATP、dGTP、dCTP、dTTP和dUTP。在一些实施例中,多个多价分子包括两个或更多个类型的多价分子的任意组合的混合物,每个类型具有选自由以下项组成的组的核苷酸单元:dATP、dGTP、dCTP、dTTP和/或dUTP。

[0073] 在一些实施例中,多个核苷酸在3'糖基处包含可移除链终止部分,其中可移除链

终止部分包含烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基团、叠氮基基团、O-叠氮基甲基基团、胺基团、酰胺基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团,并且其中可移除链终止部分可用化学化合物裂解以在糖基团上产生可延伸的3'OH部分。在一些实施例中,多个核苷酸包括选自dATP、dGTP、dCTP、dTTP和dUTP的一种类型的核苷酸。在一些实施例中,多个核苷酸包括选自dATP、dGTP、dCTP、dTTP和/或dUTP组成的组的两个或更多个类型的核苷酸的任意组合的混合物。

[0074] 在一些实施例中,支持物包括玻璃或塑料基底。在一些实施例中,支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。在一些实施例中,至少一个亲水性聚合物涂层包含选自由以下项组成的组的分子:聚乙二醇(PEG)、聚(乙烯醇)(PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮)(PVP)、聚(丙烯酸)(PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺)(PNIPAM)、聚(甲基丙烯酸甲酯)(PMA)、聚(2-羟乙基甲基丙烯酸酯)(HEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(POEGMA)、聚谷氨酸(PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。

附图说明

[0075] 本发明的新颖特征在所附权利要求中具体阐述。通过参考阐述例示性实施例的以下详细说明,将更好地理解本发明的特征和优点,在例示性实施例中利用了本发明的原理,并且在其附图中:

[0076] 图1是示出示例性线性单链文库分子(100)与单链夹板分子/链(200)杂交,由此使文库分子成环,以形成具有缺口的文库-夹板复合物(300)的示意图。文库分子(100)包括目的序列(110),该目的序列在一侧上侧接第一左通用衔接子序列(120),并且在另一侧上侧接第一右通用衔接子序列(130)。单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。

[0077] 图2是示出示例性线性单链文库分子(100)与单链夹板分子/链(200)杂交,由此使文库分子成环,以形成具有缺口的文库-夹板复合物(300)的示意图。文库分子(100)包括目的序列(110),该目的序列在一侧上侧接第一左通用衔接子序列(120)和第二左通用衔接子序列(140),并且在另一侧上侧接第二右通用衔接子序列(150)和第一右通用衔接子序列(130)。单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。

[0078] 图3是示出示例性线性单链文库分子(100)与单链夹板分子/链(200)杂交,由此使文库分子成环,以形成具有缺口的文库-夹板复合物(300)的示意图。示例性文库分子(100)包含:第一左通用衔接子序列(120);任选的第一左唯一识别序列(180);第一左索引序列(160);第二左通用衔接子序列(140);目的序列(110);第二右通用衔接子序列(150);第一右索引序列(170);和第一右通用衔接子序列(130)。单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。

[0079] 图4是示出示例性线性单链文库分子(100)与单链夹板分子/链(200)杂交,由此使文库分子成环,以形成具有缺口的文库-夹板复合物(300)的示意图。示例性文库分子(100)

包含:第一左通用衔接子序列(120);第一左索引序列(160);第二左通用衔接子序列(140);目的序列(110);第二右通用衔接子序列(150);第一右索引序列(170);任选的第一右唯一识别序列(190);和第一右通用衔接子序列(130)。单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。

[0080] 图5是示出示例性线性单链文库分子(100)与单链夹板分子/链(200)杂交,由此使文库分子成环,以形成具有缺口的文库-夹板复合物(300)的示意图。文库分子(100)包含:第一左通用衔接子序列(120);第一左接合衔接子序列(125);第一左索引序列(160);第二左接合衔接子序列(165);第二左通用衔接子序列(140);第三左接合衔接子序列(145);目的序列(110);第三右接合衔接子序列(155);第二右通用衔接子序列(150);第二右接合衔接子序列(175);第一右索引序列(170);第一右接合衔接子序列(135);任选的第一右唯一识别序列(190);和第一右通用衔接子序列(130)。单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。为了简单起见,文库-夹板复合物(300)没有示出任何接合衔接子。技术人员将认识到,文库-夹板复合物(300)可包括文库分子(100)中存在的接合衔接子中的任意一者或者两者或更多者的任意组合。

[0081] 图6示出了示例性共价闭合环状文库分子的三个示意图,每个共价闭合环状文库分子与单链夹板链(200)杂交。顶部示意图(A)示出了共价闭合环状文库分子(400),其具有目的序列(110)、第一右通用衔接子序列(130)和第一左通用衔接子序列(120)。中间示意图(B)示出了共价闭合环状文库分子(400),其具有目的序列(110)、第二右通用衔接子序列(150)、第一右通用衔接子序列(130)、第一左通用衔接子序列(120)和第二左通用衔接子序列(140)。底部示意图(C)示出了共价闭合环状文库分子(400),其具有目的序列(110)、第二右通用衔接子序列(150)、第一右索引序列(170)、第一右通用衔接子序列(130)、第一左通用衔接子序列(120)、任选的第一左唯一识别序列(180)、第一左索引序列(160)和第二左通用衔接子序列(140)。

[0082] 图7是示出示例性文库-夹板复合物,经历连接反应以闭合缺口,以形成共价闭合环状文库分子(400)的示意图,该文库分子与单链夹板链(200)杂交,其中单链夹板链(200)用作扩增引物进行滚环扩增反应。虚线代表新生的延伸产物。

[0083] 图8示出了示例性单链夹板分子/链(200)的核苷酸序列。示例性单链夹板链包括第一区域(210;SEQ ID NO:193),第二区域(220;SEQ ID NO:194)。

[0084] 图9A-9F是表1(共6张),其列出具有短随机序列(例如,NNN)的示例性第一左索引序列(160)和第一右索引序列(170)的序列。

[0085] 图10是示例性Y形衔接子的示意图,每个衔接子具有双链退火区域,其包括平端末端,或者末端可以具有5'或3'悬垂区域。图10还示出了具有不同长度的错配部分的示例性Y形衔接子。

[0086] 图11是示例性Y形衔接子的示意图,该Y形衔接子具有包含第一测序引物的结合序列和第一表面引物的结合序列的第一链。图11还示出了具有第二链的示例性Y形衔接子,该第二链包含第二测序引物的结合序列和第二表面引物的结合序列。第一链任选地包括第一样品索引序列。第二链任选地包括第二样品索引序列。

[0087] 图12示例性Y形衔接子的示意图的示意图,其具有包含第一测序引物的结合序列和第一表面引物的结合序列的第一链。图12还示出了具有第二链的示例性Y形衔接子,该第二链包含第二测序引物的结合序列和第二表面引物的结合序列。第一链任选地包括具有短随机序列(例如,NNN)的第一样品索引序列。第二链任选地包括具有短随机序列(例如,NNN)的第二样品索引序列。

[0088] 图13是示出示例性文库制备工作流程的示意图,其中核酸片段在两端处连接至Y形衔接子。Y形衔接子包括第一链,该第一链包含第一测序引物的结合序列、第一样品索引序列和第一表面引物的结合序列。Y形衔接子包括第二链,该第二链包含第二测序引物的结合序列、第二样品索引序列和第二表面引物的结合序列。

[0089] 图14是示出示例性文库制备工作流程的示意图,其中核酸片段在两端处连接至Y形衔接子。Y形衔接子包括第一链,该第一链包含第一测序引物的结合序列、具有短随机序列(例如,NNN)的第一样品索引序列和第一表面引物的结合序列。Y形衔接子包括第二链,该第二链包含第二测序引物的结合序列、第二样品索引序列和第二表面引物的结合序列。

[0090] 图15是示出示例性文库制备工作流程的示意图,其中核酸片段在两端处连接至Y形衔接子以产生衔接子-插入物-衔接子分子。Y形衔接子包括第一链,该第一链包含第一测序引物的结合序列。Y形衔接子包括第二链,该第二链包含第二测序引物的结合序列。使用携带第一表面引物的结合序列的第一有尾PCR引物对衔接子-插入物-衔接子分子经历第一引物延伸反应,以产生第一延伸产物。使用携带第二表面引物的结合序列的第二有尾PCR引物对第一延伸产物经历第二引物延伸反应,以产生第二延伸产物。

[0091] 图16是示出示例性文库制备工作流程的示意图,其中核酸片段在两端处连接至Y形衔接子以产生衔接子-插入物-衔接子分子。Y形衔接子包括第一链,该第一链包含第一测序引物的结合序列。Y形衔接子包括第二链,该第二链包含第二测序引物的结合序列。使用携带第一表面引物的结合序列、第一样品索引序列和第一表面引物的结合序列的第一有尾PCR引物对衔接子-插入物-衔接子分子经历第一引物延伸反应,以产生第一延伸产物。使用携带第二表面引物的结合序列、第二样品索引序列和第二表面引物的结合序列的第二有尾PCR引物对第一延伸产物经历第二引物延伸反应,以产生第二延伸产物。

[0092] 图17是示出示例性文库制备工作流程的示意图,其中核酸片段在两端处连接至Y形衔接子以产生衔接子-插入物-衔接子分子。Y形衔接子包括第一链,该第一链包含第一测序引物的结合序列。Y形衔接子包括第二链,该第二链包含第二测序引物的结合序列。使用携带第一表面引物的结合序列的第一有尾PCR引物、具有短随机序列(例如,NNN)的第一样品索引序列和第一表面引物的结合序列,使衔接子-插入物-衔接子分子经历第一引物延伸反应,以产生第一延伸产物。使用携带第二表面引物的结合序列、第二样品索引序列和第二表面引物的结合序列的第二有尾PCR引物对第一延伸产物经历第二引物延伸反应,以产生第二延伸产物。

[0093] 图18是示例性低结合支持物的示意图,其包括玻璃基底和共价或非共价粘附至玻璃的交替的亲水涂层,并且其进一步包含充当寡核苷酸引物(例如,捕获寡核苷酸)的附着位点的化学反应性官能团。在替代性实施例,支持物可由任何材料(诸如玻璃、塑料或聚合物材料)制成。

[0094] 图19是多价分子的各种示例性构型的示意图。左(I类):具有“星爆(starburst)”

或“旋梯(helter-skelter)”构型的多价分子的示意图。中心(II类):具有树枝状高分子构型的多价分子的示意图。右(III类):通过使链霉亲和素与4臂或8臂PEG-NHS与生物素和dNTP反应而形成的多个多价分子的示意图。核苷酸单元被指定为‘N’,生物素被指定为‘B’,并且链霉亲和素被指定为‘SA’。

[0095] 图20是包含附接至多个核苷酸臂的通用核心的示例性多价分子的示意图。

[0096] 图21是包含附接至多个核苷酸臂的树枝状核心的示例性多价分子的示意图。

[0097] 图22示出了包含附接至多个核苷酸臂的核心的示例性多价分子的示意图,其中核苷酸臂包含生物素、间隔基、接头和核苷酸单元。

[0098] 图23是包含核心附接部分、间隔基、接头和核苷酸单元的示例性核苷酸臂的示意图。

[0099] 图24示出了示例性间隔基的化学结构(顶部)以及各种示例性接头的化学结构,包括11-原子接头、16-原子接头、23-原子接头和N3接头(底部)。

[0100] 图25示出了各种示例性接头的化学结构,包括接头1-9。

[0101] 图26示出了接合/附接至核苷酸单元的各种示例性接头的化学结构。

[0102] 图27示出了接合/附接至核苷酸单元的各种示例性接头的化学结构。

[0103] 图28示出了接合/附接至核苷酸单元的各种示例性接头的化学结构。

[0104] 图29示出了接合/附接至核苷酸单元的各种示例性接头的化学结构。

[0105] 图30示出了示例性生物素化核苷酸臂的化学结构。在该实例中,核苷酸单元经由在嘧啶碱基的5位或嘌呤碱基的7位处的炔丙胺附接物与接头连接。

[0106] 图31是示出包括通用右样品索引和3聚体随机序列(NNN)的右样品索引序列(170)的核苷酸碱基多样性的图。该图示出,A和T碱基识别的3聚体随机序列(NNN)的核苷酸多样性约为30%,并且C和G碱基识别的核苷酸多样性约为20%。

[0107] 图32是示出缺少3聚体随机序列(NNN)的左样品索引序列(160)的核苷酸碱基多样性的图。该图示出A和T碱基识别的核苷酸多样性约为40%,C碱基识别的核苷酸多样性约为15%,并且G碱基识别的核苷酸多样性约为5%。

具体实施方式

[0108] 在整个本申请中,引用了各种出版物、专利和/或专利申请。出版物、专利和/或专利申请的公开特此通过引用以其整体并入本申请中,以便更全面地描述本公开所属领域的现状。

[0109] 定义:

[0110] 本文提供的标题不是对本公开的各个方面的限制,这些方面可通过参考作为整体的说明书来理解。

[0111] 除非另有定义,否则本文所用的技术和科学术语具有本领域普通技术人员通常理解的含义,除非另有定义。一般而言,与本文所述的分子生物学、核酸化学、蛋白质化学、遗传学、微生物学、转基因细胞生产和杂交的技术相关的术语是本领域中众所周知和通常使用的那些。本文所述的技术和程序通常根据本领域中众所周知的常规方法以及如整个本说明书中引用和讨论的各种一般性和更具体的参考文献中所述来执行。例如,参见Sambrook等人,Molecular Cloning:A Laboratory Manual(第三版,Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, N.Y. 2000)。另请参见Ausubel等人, Current Protocols in Molecular Biology, Greene Publishing Associates (1992)。与本文所述的实验室程序和技术结合利用的命名法是本领域中众所周知和通常使用的那些。

[0112] 除非本文上下文另外要求, 否则单数术语应包括复数, 并且复数术语应包括单数。单数形式“一”、“一个”和“该”以及任何词语的单数使用包括复数所指对象, 除非明确且毫不含糊地限于一个所指对象。

[0113] 应理解, 替代性术语 (例如, “或”) 的使用被视为意指替代方案中的一者或两者或它们的任意组合。

[0114] 本文所用的术语“和/或”应被视为意指指定特征或组分中的每一个与或不与另一个一起的具体公开。例如, 如本文中诸如“A和/或B”的短语中所使用的术语“和/或”旨在包括: “A和B”; “A或B”; “A” (仅A); 和“B” (仅B)。类似地, 如本文中诸如“A、B和/或C”中所使用的术语“和/或”旨在涵盖以下方面中的每个方面: “A、B和C”; “A、B或C”; “A或C”; “A或B”; “B或C”; “A和B”; “B和C”; “A和C”; “A” (仅A); “B” (仅B); 和“C” (仅C)。

[0115] 如本文和所附权利要求中所用, 如本文所用的术语“包含”、“包括”、“具有”和“含有”以及它们的语法变体旨在是非限制性的, 使得列表中的一个项目或多个项目不排除可被取代或添加到列出的项目的其他项目。应当理解, 每当在本文中用语言“包含/包括”描述方面时, 也以其他方式提供以“由...组成”和/或“基本上由...组成”描述的类型方面。

[0116] 如本文所用, 术语“约”和“大约”是指如由本领域普通技术人员所确定的在针对特定值或组合物的可接受误差范围内的值或组合物, 这将部分地取决于如何测量或确定值或组合物, 即, 测量系统的限制。例如, 根据本领域中的实践, “约”或“大约”可意指在一个或超过一个标准偏差内。替代性地, “约”或“大约”可意指至多10% (即, $\pm 10\%$) 或更大的范围, 这取决于测量系统的限制。例如, 约5mg可包括4.5mg至5.5mg之间的任何数量。此外, 特别是对于生物学系统或过程, 这些术语可意指至多一个数量级或至多5倍的值。当在本公开中提供特定值或组合物时, 除非另有说明, 否则“约”或“大约”的含义应被假定是在针对该特定值或组合物的可接受误差范围内。此外, 在提供值的范围和/或子范围的情况下, 所述范围和/或子范围可以包括所述范围和/或子范围的端点。

[0117] 本文使用的术语“肽”、“多肽”和“蛋白质”以及其他相关术语可互换使用, 并且是指氨基酸的聚合物并且不限于任何特定长度。多肽可包括天然和非天然氨基酸。多肽包括重组或化学合成的形式。多肽还包括尚未经历翻译后修饰的前体分子, 诸如蛋白水解切割、由于核糖体跳跃导致的切割、羟基化、甲基化、脂化、乙酰化、SUMO化、泛素化、糖基化、磷酸化和/或二硫键形成。这些术语涵盖蛋白质序列的天然和人工蛋白质、蛋白质片段和多肽类似物 (诸如突变蛋白、变体、嵌合蛋白和融合蛋白) 以及翻译后或以其他方式共价或非共价修饰的蛋白质。

[0118] 术语“细胞生物样品”是指单个细胞、多个细胞、组织、器官、生物体或任何这些细胞生物样品的切片。细胞生物样品可以从生物体中提取 (例如, 活检), 或者从在液体中或培养皿中生长的细胞培养物获得。细胞生物样品包含新鲜的、冷冻的、新鲜冷冻的或存档的样品 (例如, 福尔马林固定石蜡包埋的; FFPE)。细胞生物样品可以包埋在蜡、树脂、环氧树脂或琼脂中。可以将细胞生物样品固定在例如丙酮、乙醇、甲醇、甲醛、多聚甲醛-Triton或戊二醛中的任意一者或者两者或更多者的任意组合中。细胞生物样品可以是切片的或未切片

的。细胞生物样品可以染色的、脱色的或不染色的。

[0119] 可以使用本领域技术人员已知的多种技术中的任何技术从细胞或细胞生物样品中提取目的核酸。例如,典型DNA提取程序包括:(i)收集要从中提取DNA的细胞样品或组织样品,(ii)破坏细胞膜(即,细胞溶解)以释放DNA和其他细胞质组分,(iii)用浓盐溶液处理经溶解的样品以使蛋白质、脂质和RNA沉淀,之后进行离心以分离出沉淀的蛋白质、脂质和RNA,以及(iv)从上清液纯化DNA以移除去污剂、蛋白质、盐或在细胞膜溶解期间所使用的其他试剂。多种合适的商业核酸提取和纯化试剂盒符合本文的公开。实例包括但不限于:来自Qiagen(Germantown,MD)的QIAamp试剂盒(用于从人类样品分离基因组DNA)和DNAeasy试剂盒(用于从动物或植物样品分离基因组DNA),或来自Promega(Madison,WI)的Maxwell®和ReliaPrep™系列试剂盒。

[0120] 如本文所用,术语“聚合酶”及其变体包括包含结合核苷酸(或核苷)的结构域的酶,其中聚合酶可形成具有模板核酸和互补核苷酸的复合物。聚合酶可具有一个或多个活性,包括但不限于:碱基类似物检测活性、DNA聚合活性、逆转录酶活性、DNA结合、链置换活性以及核苷酸结合和识别。聚合酶可以为可催化核苷酸(包括其类似物)聚合成核酸链的任何酶。通常但不一定,此类核苷酸聚合可以模板依赖性方式发生。通常,聚合酶包含在其处可发生核苷酸结合和/或核苷酸聚合的催化作用的一个或多个活性位点。在一些实施例中,聚合酶包括其他酶活性,诸如例如3'至5'核酸外切酶活性或5'至3'核酸外切酶活性。在一些实施例中,聚合酶具有链置换活性。聚合酶可包括但不限于:天然存在的聚合酶以及它们的任何亚基和截短物、突变聚合酶、变体聚合酶、重组、融合或以其他方式工程化的聚合酶、经化学修饰的聚合酶、合成分子或组装体、以及保留催化核苷酸聚合的能力的任何类似物、衍生物或它们的片段(例如,催化活性片段)。聚合酶包括催化失活聚合酶、催化活性聚合酶、逆转录酶和包含核苷酸结合结构域的其他酶。在一些实施例中,聚合酶可从细胞被分离或者使用重组DNA技术或化学合成方法来生成。在一些实施例中,聚合酶可在原核生物、真核生物、病毒或噬菌体生物体中被表达。在一些实施例中,聚合酶可以为经翻译后修饰的蛋白质或其片段。聚合酶可源自原核生物、真核生物、病毒或噬菌体。聚合酶包括DNA引导的DNA聚合酶和RNA引导的DNA聚合酶。

[0121] 术语“链置换”是指聚合酶局部地分离双链核酸链并且以基于模板的方式合成新链的能力。链置换聚合酶从模板链置换互补链并且催化新链合成。链置换聚合酶包括嗜中温和嗜热聚合酶。链置换聚合酶包括野生型酶和变体(包括核酸外切酶突变体、突变体版本、嵌合酶和经截短的酶)。链置换聚合酶的实例包括:phi29 DNA聚合酶、Bst DNA聚合酶的大片段、Bsu DNA聚合酶的大片段(exo-)、Bca DNA聚合酶(exo-)、大肠杆菌(E.coli)DNA聚合酶的Klenow片段、T5聚合酶、M-MuLV逆转录酶、HIV病毒逆转录酶、Deep Vent DNA聚合酶和KOD DNA聚合酶。phi29 DNA聚合酶可以为野生型phi29 DNA聚合酶(例如,来自Expedeon的MagniPhi)、或变体EquiPhi29 DNA聚合酶(例如,来自赛默飞世尔科技公司)、或嵌合QualiPhi DNA聚合酶(例如,来自4basebio)。

[0122] 本文所用的术语“核酸”、“多核苷酸”和“寡核苷酸”以及其他相关术语可互换地使用,并且是指核苷酸的聚合物且不限于任何特定长度。核酸包括重组和化学合成形式。核酸可被分离。核酸包括DNA分子(例如,cDNA或基因组DNA)、RNA分子(例如,mRNA)、使用核苷酸类似物生成的DNA或RNA的类似物(例如,肽核酸和非天然存在的核苷酸类似物)以及含有

DNA和RNA的嵌合形式。核酸可以是单链或双链的。核酸包含核苷酸的聚合物,其中核苷酸包括天然或非天然碱基和/或糖。核酸包含天然存在的核苷间键,例如磷酸二酯键。核酸包含非天然核苷间键,包含硫代磷酸酯、硫醇代磷酸酯或肽核酸(PNA)键。在一些实施例中,核酸包含一个类型的多核苷酸或者两个或更多个不同类型的多核苷酸的混合物。

[0123] 如本文所用,术语“可操作地链接”和“可操作地接合”或相关术语是指组分的并置。经并置的组分可共价链接在一起。例如,两个核酸组分可以酶促方式连接在一起,其中将该两个组分连接在一起的键包括磷酸二酯键。第一和第二核酸组分可链接在一起,其中第一核酸组分可对第二核酸组分赋予功能。例如,引物结合序列与目的序列之间的链接形成具有可与引物结合的部分的核酸文库分子。在另一实例中,转基因(例如,编码多肽的核酸或目的核酸序列)可连接至载体,其中该链接允许载体中含有的转基因序列进行表达或起作用。在一些实施例中,转基因可操作地链接至影响转基因表达的宿主细胞调控序列(例如,启动子序列)。在一些实施例中,载体包含至少一个宿主细胞调控序列,包括启动子序列、增强子、转录和/或翻译起始序列、转录和/或翻译终止序列、多肽分泌信号序列等。在一些实施例中,宿主细胞调控序列控制转基因的水平、定时和/或位置的表达。

[0124] 术语“链接的”、“接合的”、“附接的”、“附加的”以及它们的变体包含化合物或分子的任意组合之间任何类型的融合、键合、粘附或缔合,其具有足够的稳定性以承受在特定程序中的使用。该程序可以包括但不限于:核苷酸结合;核苷酸掺入;去封闭(例如,移除链终止部分);洗涤;移除;流动;检测;成像和/或识别。例如,此类键可包括:共价键、离子键、氢键、偶极-偶极键、亲水键、疏水键或亲和键、涉及范德华力的键或缔合、机械键等。在一些实施例中,此类键在分子内出现,例如将单链或双链线性核酸分子的端部链接在一起以形成环状分子。在一些实施例中,这种键可以发生在不同分子的组合之间,或分子与非分子之间,包括但不限于:核酸分子与固体表面之间的键;蛋白质和可检测报告部分之间的键;核苷酸和可检测报告部分之间的键;等。键的一些实例可以在例如Hermanson, G., “Bioconjugate Techniques”, 第二版(2008); Aslam, M., Dent, A., “Bioconjugation: Protein Coupling Techniques for the Biomedical Sciences”, London: Macmillan (1998); Aslam, M., Dent, A., “Bioconjugation: Protein Coupling Techniques for the Biomedical Sciences”, London: Macmillan (1998)。

[0125] 本文所用的术语“引物”和相关术语是指能够与DNA和/或RNA多核苷酸模板杂交以形成双链体分子的寡核苷酸。引物可以沿着其整个长度是单链的或者具有单链和双链部分。引物包括天然核苷酸和/或核苷酸类似物。引物可以为重组核酸分子。引物可具有任何长度,但通常范围为4至50个核苷酸。典型的引物包括5'端和3'端。引物的3'端可包括充当聚合酶催化的引物延伸反应中的核苷酸聚合起始位点的3'OH部分。替代性地,引物的3'端可缺少3'OH部分或可包括抑制聚合酶催化的反应中的核苷酸聚合的末端3'阻断基团。沿引物的长度的任何一个核苷酸或超过一个核苷酸可用可检测报告基因部分来标记。引物可在溶液中(例如,可溶性引物)或可被固定至支持物(例如,捕获引物)。

[0126] 术语“模板核酸”、“模板多核苷酸”、“靶核酸”、“靶多核苷酸”、“模板链”和其他变体是指在本文描述的任何扩增和/或测序方法中用作基础核酸分子的核酸链。模板核酸可以为单链或双链的,或者模板核酸可具有单链或双链部分。模板核酸可从天然存在的来源、重组形式获得或经化学合成以包括任何类型的核酸类似物。模板核酸可以是线性的、多联

体的、环状的或其他形式。

[0127] 术语“衔接子”和相关术语是指可以可操作地链接(附加)至靶多核苷酸的寡核苷酸,其中衔接子对共接合衔接子-靶分子赋予功能。衔接子包括DNA、RNA、嵌合DNA/RNA或它们的类似物。衔接子可包括至少一个核糖核苷残基。衔接子可以为单链、双链的或具有单链和/或双链部分。衔接子可被配置为线性的、茎环的、发夹或Y形形式。衔接子可以为任何长度,包括4个至100个核苷酸或更长。衔接子可具有平端、悬垂端或两者的组合。悬垂端包括5'悬垂和3'悬垂端。单链衔接子的5'端或双链衔接子的一个链可具有5'磷酸酯基团或缺少5'磷酸酯基团。衔接子可包括未与靶多核苷酸杂交的5'尾(例如,有尾衔接子),或者衔接子可以为无尾的。衔接子的至少一部分包含已知的和预定的序列。衔接子可包括与引物(诸如扩增引物、测序引物或捕获引物(例如,可溶性或经固定的捕获引物))的至少一部分互补的序列。衔接子可包括随机序列或简并序列。衔接子可包括至少一个肌苷残基。衔接子可包括至少一个硫代磷酸酯、硫醇代磷酸酯和/或磷酸胺键。衔接子可包括至少一种条形码序列,其可用于在多重测定中区分来自不同样品源的多核苷酸(例如,插入序列)。衔接子可包括至少一种唯一识别序列(例如,分子标识),其可用于唯一地识别衔接子所附加的核酸分子。在一些实施例中,唯一识别序列包含2-12个或更多个具有已知序列的核苷酸。例如,唯一识别序列包括已知的随机序列,其中每个位置处的核苷酸随机选自具有碱基A、G、C、T或U的核苷酸。衔接子可包括至少一种限制酶识别序列,包括选自I型、II型、III型、IV型、Hs型或IIB型组成的组中的任意一者或者两者或更多者的任意组合。

[0128] 术语“通用序列”和相关术语是指核酸分子中两个或更多个多核苷酸分子共有的序列。例如,具有通用序列的衔接子可以可操作地接合至多个多核苷酸,使得共接合的分子群体携带相同的通用衔接子序列。通用衔接子序列的实例包括扩增引物序列、测序引物序列或捕获引物序列(例如,可溶性或固定的捕获引物)。

[0129] 当用于指代核酸分子时,术语“杂交(hybridize)”或“杂交(hybridizing)”或“杂交(hybridization)”或其他相关术语是指两个不同核酸之间进行氢键结合以形成双链体核酸。杂交也包括单个核酸分子的两个不同区域之间进行氢键结合以形成具有双链体区的自杂交分子。杂交可包括Watson-Crick或Hoogsteen结合以形成双链体双链核酸或核酸分子内的双链区域。双链核酸或单个核酸的该两个不同区可以为完全地互补的或部分地互补的。互补核酸链不需要跨它们的整个长度彼此杂交。互补碱基配对可以为标准A-T或C-G碱基配对或者可以为碱基配对相互作用的其他形式。双链体核酸可包括错配的经碱基配对的核苷酸。

[0130] 当用于指代核酸时,术语“延伸(extend)”、“延伸(extending)”、“延伸(extension)”和其他变体是指一个或多个核苷酸到核酸分子中的掺入。核苷酸掺入包括一个或多个核苷酸到核酸链的末端3'OH端中的聚合,导致核酸链的延伸。可用天然核苷酸和/或核苷酸类似物来进行核苷酸掺入。通常,但不一定,核苷酸掺入以模板依赖性方式发生。可使用延伸核酸分子的任何合适的方法,包括由DNA聚合酶或RNA聚合酶催化的引物延伸。

[0131] 术语“核苷酸”和相关术语是指包含芳香族碱基、五碳糖(例如,核糖或脱氧核糖)和至少一个磷酸酯基团的分子。正则或非正则核苷酸与该术语的使用一致。在一些实施例中,核苷酸包含单磷酸盐、二磷酸盐或三磷酸盐,或相应的磷酸盐类似物。术语“核苷”是指包括芳香族碱基和糖的分子。核苷酸和核苷可以为未经标记的或用可检测报道基因部分标

记的。

[0132] 核苷酸(和核苷)通常包含杂环碱基,包括经取代或未经取代的含氮母体杂芳环,其通常存在于核酸中,包括天然存在的、经取代的、经修饰的或工程化的变体或它们的类似物。核苷酸(或核苷)的碱基能够与适当的互补碱基形成Watson-Crick和/或Hoogsteen氢键。示例性的碱包括但不限于:嘌呤类和嘧啶类,诸如:2-氨基嘌呤、2,6-二氨基嘌呤、腺嘌呤(A)、乙烯腺嘌呤、 N^6 - Δ^2 -异戊烯基腺嘌呤(6iA)、 N^6 - Δ^2 -异戊烯基-2-甲硫腺嘌呤(2ms6iA)、 N^6 -甲基腺嘌呤、鸟嘌呤(G)、异鸟嘌呤、 N^2 -二甲基鸟嘌呤(dmG)、7-甲基鸟嘌呤(7mG)、2-硫代嘧啶、6-硫鸟嘌呤(6sG)、次黄嘌呤和 O^6 -甲基鸟嘌呤;7-脱氮嘌呤,诸如7-脱氮腺嘌呤(7-脱氮-A)和7-脱氮鸟嘌呤(7-脱氮-G);嘧啶类,诸如胞嘧啶(C)、5-丙炔基胞嘧啶、异胞嘧啶、胸腺嘧啶(T)、4-硫胸腺嘧啶(4sT)、5,6-二氢胸腺嘧啶、 O^4 -甲基胸腺嘧啶、尿嘧啶(U)、4-硫尿嘧啶(4sU)和5,6-二氢尿嘧啶(二氢尿嘧啶;D);吡咯类,诸如硝基吡咯和4-甲基吡咯;吡咯类,诸如硝基吡咯;水粉草素;肌苷类;羟甲基胞嘧啶类;5-甲基胞嘧啶类;碱基(Y);以及甲基化、糖基化和酰化的碱基部分;等。另外的示例性碱基可见于Fasman,1989,在“Practical Handbook of Biochemistry and Molecular Biology”中,第385至394页,CRC Press,Boca Raton,Fla。

[0133] 核苷酸(和核苷)通常包含糖部分,诸如碳环部分(Ferraro和Gotor 2000Chem.Rev.100:4319-48),无环部分(Martinez等人,1999Nucleic Acids Research 27:1271-1274;Martinez等人,1997Bioorganic&Medicinal Chemistry Letters第7卷:3013-3016)和其他糖部分(Joeng等人,1993J.Med.Chem.36:2627-2638;Kim,等人,1993J.Med.Chem.36:30-7;Eschenmosser 1999Science 284:2118-2124;和美国专利号5,558,991)。糖部分包含:核糖基;2'-脱氧核糖基;3'-脱氧核糖基;2',3'-二脱氧核糖基;2',3'-二脱氢二脱氧核糖基;2'-烷氧基核糖基;2'-叠氮核糖基;2'-氨基核糖基;2'-氟核糖基;2'-巯基羧基;2'-烷硫基核糖基;3'-烷氧基核糖基;3'-叠氮核糖基;3'-氨基核糖基;3'-氟核糖基;3'-巯基羧基;3'-烷硫基核糖基碳环;无环的或其他修饰的糖。

[0134] 在一些实施例中,核苷酸包含具有一个、两个或三个磷原子的链,其中链通常经由酯键或磷酸胺键附接至糖部分的5'碳。在一些实施例中,核苷酸为具有磷链的类似物,其中磷原子用居间的O、S、NH、亚甲基或亚乙基而链接在一起。在一些实施例中,链中的磷原子包括经取代的侧基,包括O、S或 BH_3 。在一些实施例中,链包括用类似物取代的磷酸酯基团,包括磷酸胺、硫代磷酸酯、二硫代磷酸酯和O-甲基亚磷酸胺基团。

[0135] 术语“滚环扩增”通常是指采用环化核酸模板分子的扩增方法,该模板分子含有靶标目的序列、扩增引物结合序列和任选的一个或多个衔接子序列,诸如测序引物结合序列和/或样品索引序列。滚环扩增反应可以在等温扩增条件下进行,并且包括成环的核酸模板分子、扩增引物、链置换聚合酶和多个核苷酸,以生成含有环状模板分子的串联重复序列的多联体以及原始成环的核酸模板分子中存在的任何衔接子序列。多联体可以自塌陷形成核酸纳米球。通过在环状模板分子中包括一对反向重复序列,或者通过用一个或多个压实寡核苷酸进行滚环扩增反应,可以进一步压缩纳米球的形状和尺寸。使用滚环扩增生用于测序工作流程的克隆扩增子的优点中的一个是可以同时对纳米球中靶标序列的重复拷贝进行测序以增加信号强度。在一些实施例中,滚环扩增反应可以在多个具有至少四个连续鸟嘌呤的压实存在下进行。滚环扩增反应生成多联体,其包含针对压实寡核苷酸的通用结

合序列的重复拷贝。至少一种压实寡核苷酸可以形成鸟嘌呤四联体并与针对压实寡核苷酸的通用结合序列杂交,并且所得的多联体可以折叠形成分子内G-四链体结构。多联体可自塌陷以形成紧凑的纳米球。纳米球中鸟嘌呤四联体和G-四链体的形成可以增加纳米球的稳定性,以保持其紧凑的尺寸和形状,这样可以承受用于进行本文描述的任何测序工作流程的试剂的重复流动。

[0136] 当用于指代核酸时,术语“扩增(amplify、amplifying、amplification)”和其他相关术语包括产生原始多核苷酸模板分子的多个拷贝,其中拷贝包含与模板序列互补的序列,和/或拷贝包含与模板序列相同的序列。在一些实施例中,拷贝包含与模板序列基本相同的序列,和/或与模板序列互补的序列基本相同的序列。

[0137] 术语“报道基因部分(reporter moiety)”、“报道基因部分(reporter moieties)”或相关术语是指生成或导致生成可检测信号的化合物。报道基因部分有时被称为“标记”。可使用任何合适的报道基因部分,包括发光、光致发光、电致发光、生物发光、化学发光、荧光、磷光、发色团、放射性同位素、电化学、质谱、Raman、半抗原、亲和标签、原子或酶。报道基因部分生成由化学或物理变化(例如,热、光、电、pH、盐浓度、酶活性或邻近事件)引起的可检测信号。邻近事件包括彼此接近、或彼此缔合、或彼此结合的两个报道基因部分。本领域技术人员众所周知对报道基因部分进行选择,使得每一者在可与其他报道基因部分相区别的波长处吸收激发辐射和/或发出荧光,以允许监测相同反应中或不同反应中不同报道基因部分的存在。可选择具有在光谱上有区别的发射曲线或具有最小重叠光谱发射曲线的两个或更多个不同报道基因部分。报道基因部分可链接(例如,可操作地链接)至核苷酸、核苷、核酸、酶(例如,聚合酶或逆转录酶)或支持物(例如,表面)。

[0138] 报道基因部分(或标记)包括荧光标记或荧光团。可以用作荧光标记或荧光团的示例性荧光部分包括但不限于,荧光素和荧光素衍生物,诸如羧基荧光素、四氯荧光素、六氯荧光素、羧基萘酚荧光素、异硫氰酸荧光素、NHS-荧光素、碘乙酰胺基荧光素、马来酰亚胺荧光素、SAMSA-荧光素、氨基硫脲荧光素、碳酸胍甲硫乙酰胺基荧光素,罗丹明和罗丹明衍生物,诸如TRITC、TMR、丽丝胺罗丹明、德克萨斯红、罗丹明B、罗丹明6G、罗丹明10、NHS-罗丹明、TMR-碘乙酰胺、丽丝胺罗丹明B磺酰氯、丽丝胺罗丹明B磺酰肼、德克萨斯红磺酰氯、德克萨斯红磺酰肼,香豆素和香豆素衍生物,诸如AMCA、AMCA-NHS、AMCA-磺基-NHS、AMCA-HPDP、DCIA、AMCE-酰肼, BODIPY和衍生物诸如BODIPY FL C3-SE、BODIPY 530/550C3、BODIPY 530/550C3-SE、BODIPY 530/550C3酰肼、BODIPY 493/503C3酰肼、BODIPY FL C3酰肼、BODIPY FLIA、BODIPY 530/551IA、Br-BODIPY 493/503、Cascade Blue和衍生物,诸如Cascade Blue乙酰三氮化物、Cascade Blue尸胺、Cascade Blue乙二胺、Cascade Blue酰肼, Lucifer Yellow和衍生物,诸如Lucifer Yellow碘乙酰胺、Lucifer Yellow CH,花青和衍生物,诸如吲哚鎓基花青染料、苯并吲哚鎓基花青染料、吡啶鎓基花青染料、噻唑鎓基花青染料、喹啉鎓基花青染料、咪唑鎓基花青染料、Cy 3、Cy5、镧系元素螯合物和衍生物,诸如BCPDA、TBP、TMT、BHHCT、BCOT、铕螯合物、铽螯合物, Alexa Fluor染料、DyLight染料、Atto染料、LightCycler Red染料、CAL Fluor染料、JOE和其衍生物、Oregon Green染料、WellRED染料、IRD染料、藻红蛋白和藻胆素染料、孔雀石绿、二苯基乙烯、DEG染料、NR染料、近红外染料和本领域已知的其他染料,诸如Haugland, Molecular Probes Handbook, (Eugene, Oreg.) 第6版; Lakowicz, Principles of Fluorescence Spectroscopy, 第2版, Plenum Press New

York(1999)或Hermanson,Bioconjugate Techniques,第2版中所描述的那些,或其衍生物;或它们的任意组合。花青染料可以磺化或非磺化形式存在,并且由被两个氮原子之间的聚甲炔桥分开的两个假吡啶、苯并吡啶鎓、吡啶鎓、噻唑鎓和/或喹啉鎓基团组成。可商购获得的花青荧光团包括例如: Cy3(其可包含1-[6-(2,5-二氧化吡咯烷-1-基氧基)-6-氧代己基]-2-(3-{1-[6-(2,5-二氧化吡咯烷-1-基氧基)-6-氧代己基]-3,3-二甲基-1,3-二氢-2H-吡啶-2-亚基}丙-1-烯-1-基)-3,3-二甲基-3H-吡啶鎓或1-[6-(2,5-二氧化吡咯烷-1-基氧基)-6-氧代己基]-2-(3-{1-[6-(2,5-二氧化吡咯烷-1-基氧基)-6-氧代己基]-3,3-二甲基-5-磺基-1,3-二氢-2H-吡啶-2-亚基}丙-1-烯-1-基)-3,3-二甲基-3H-吡啶鎓-5-磺酸酯)、Cy5(其可包含1-(6-((2,5-二氧化吡咯烷-1-基)氧基)-6-氧代己基)-2-((1E,3E)-5-((E)-1-(6-((2,5-二氧化吡咯烷-1-基)氧基)-6-氧代己基)-3,3-二甲基-5-吡啶-2-亚基)五-1,3-二烯-1-基)-3,3-二甲基-3H-吡啶-1-鎓或1-(6-((2,5-二氧化吡咯烷-1-基)氧基)-6-氧代己基)-2-((1E,3E)-5-((E)-1-(6-((2,5-二氧化吡咯烷-1-基)氧基)-6-氧代己基)-3,3-二甲基-5-磺基吡啶-2-亚基)五-1,3-二烯-1-基)-3,3-二甲基-3H-吡啶-1-鎓-5-磺酸酯)和Cy7(其可包含1-(5-羧基戊基)-2-[(1E,3E,5E,7Z)-7-(1-乙基-1,3-二氢-2H-吡啶-2-亚基)七-1,3,5-三烯-1-基]-3H-吡啶鎓或1-(5-羧基戊基)-2-[(1E,3E,5E,7Z)-7-(1-乙基-5-磺基-1,3-二氢-2H-吡啶-2-亚基)七-1,3,5-三烯-1-基]-3H-吡啶鎓-5-磺酸酯),其中“Cy”代表‘花青’,并且第一个数字识别两个假吡啶基团之间的碳原子的数量。Cy2是一种噻唑衍生物而不是假吡啶,并且苯并衍生化的Cy3.5、Cy5.5和Cy7.5是该规则的例外。

[0139] 在一些实施例中,报道基因部分可以为FRET对,使得可在单个激发和成像步骤下执行多个分类。如本文所用,FRET可包括激发交换(Forster)转移或电子交换(Dexter)转移。

[0140] 如本文所用,术语“支持物”是指被设计用于生物学分子或生物学样品的沉积以进行测定和/或分析的底物。要沉积到支持物上的生物学分子的示例包括核酸(例如,DNA、RNA)、多肽、糖类、脂质、单个细胞或多个细胞。生物学样品的实例包括但不限于:唾液、痰、粘液、血液、血浆、血清、尿液、粪便、汗液、泪液和来自组织或器官的液体。

[0141] 在一些实施例中,支持物为固体、半固体或两者的组合。在一些实施例中,支持物为多孔的、半多孔的、无孔的或孔隙率的任意组合。在一些实施例中,支持物可以为基本上平面的、凹面的、凸面的或它们的任意组合。在一些实施例中,支持物可以为柱形的,例如包含毛细管或毛细管的内表面。

[0142] 在一些实施例中,支持物的表面可以为基本上光滑的。在一些实施例中,支持物可具有规则或不规则的纹理,包括凸块、经蚀刻、孔、三维支架或它们的任意组合。

[0143] 在一些实施例中,支持物包含具有任何形状的微珠,包括球形、半球形、柱形、桶形、环形、盘形、杆状、锥形、三角形、立方体、多边形、管状或线状。

[0144] 支持物可由任何材料制成,包括但不限于:玻璃、熔融二氧化硅、硅、聚合物(例如,聚苯乙烯(PS)、大孔聚苯乙烯(MPPS)、聚甲基丙烯酸甲酯(PMMA)、聚碳酸酯(PC)、聚丙烯(PP)、聚乙烯(PE)、高密度聚乙烯(HDPE)、环烯烃聚合物(COP)、环烯烃共聚物(COC)、聚对苯二甲酸乙二醇酯(PET))或它们的任意组合。设想了玻璃和塑料基板两者的各种组合物。

[0145] 本公开提供了固定至支持物的多个(例如,两个或更多个)核酸模板分子。在一些

实施例中,固定的多个核酸模板分子具有相同的序列或具有不同的序列。在一些实施例中,多个核酸模板分子中的各个核酸模板分子被固定至支持物上的不同位点。在一些实施例中,多个核酸模板中的两个或更多个各个核酸模板分子被固定至支持物上的位点。

[0146] 术语“阵列”是指包含位于支持物上预定位置处的多个位点以形成位点阵列的支持物。位点可以为离散的并且被间隙区分开。在一些实施例中,支持物上预定的位点可以一维排列成行或列,或者以二维排列成行和列。在一些实施例中,该多个预定的位点在支持物上以有组织的方式排列。在一些实施例中,该多个预先确定的位点以任何有组织的图案排列,包括直线、六边形图案、网格图案、具有反射对称性的图案、具有旋转对称性的图案等等。不同位点对之间的间距可相同或可不同。在一些实施例中,支持物包含至少 10^2 个位点、至少 10^3 个位点、至少 10^4 个位点、至少 10^5 个位点、至少 10^6 个位点、至少 10^7 个位点、至少 10^8 个位点、至少 10^9 个位点、至少 10^{10} 个位点、至少 10^{11} 个位点、至少 10^{12} 个位点、至少 10^{13} 个位点、至少 10^{14} 个位点、至少 10^{15} 个位点或更多个位点,其中位点位于支持物上预定的位置处。在一些实施例中,支持物上的多个预定位点(例如, 10^2 - 10^{15} 个位点或更多)固定有核酸模板分子以形成核酸模板阵列。在一些实施例中,通过与固定的表面捕获引物杂交而将核酸模板分子固定在多个预定位点处,或者将核酸模板分子共价附接至表面捕获引物。在一些实施例中,核酸模板分子固定在多个预定位点,例如固定在 10^2 - 10^{15} 个位点或更多。在一些实施例中,固定的核酸模板分子被克隆扩增以在多个预定位点处生成固定的核酸簇。在一些实施例中,各个固定的核酸簇包含线性簇,或包含单链或双链多联体。

[0147] 在一些实施例中,包含位于支持物上的随机位置处的多个位点的支持物在本文中被认为在其上具有随机定位的位点的支持物。支持物上随机地定位的位点的位置不是预先确定的。该多个随机地定位的位点在支持物上以无序和/或不可预测的方式排列。在一些实施例中,支持物包含至少 10^2 个位点、至少 10^3 个位点、至少 10^4 个位点、至少 10^5 个位点、至少 10^6 个位点、至少 10^7 个位点、至少 10^8 个位点、至少 10^9 个位点、至少 10^{10} 个位点、至少 10^{11} 个位点、至少 10^{12} 个位点、至少 10^{13} 个位点、至少 10^{14} 个位点、至少 10^{15} 个位点或更多个位点,其中位点在支持物上是随机定位的。在一些实施例中,支持物上的多个随机定位的位点(例如, 10^2 - 10^{15} 个位点或更多)固定有核酸模板分子。在一些实施例中,通过与固定的表面捕获引物杂交将核酸模板分子固定在多个随机定位的位点,或者将核酸模板分子共价附接至表面捕获引物。在一些实施例中,核酸模板被固定在多个随机地定位的位点处,例如被固定在 10^2 - 10^{15} 个位点或更多个位点处。在一些实施例中,固定的核酸模板被克隆扩增以在多个随机定位的位点处生成固定的核酸簇。在一些实施例中,各个固定的核酸簇包含线性簇,或包含单链或双链多联体。

[0148] 在一些实施例中,支持物上的多个固定的表面捕获引物(例如,位于支持物上的预定或随机位置处)彼此流体连通,以允许试剂(例如,核酸模板分子、可溶性引物、酶、核苷酸、二价阳离子、缓冲液等)的溶液流动到支持物上,使得支持物上多个固定的表面捕获引物可以基本上同时以大规模并行方式与试剂反应。在一些实施例中,所述多个固定的表面捕获引物的流体连通可用于在该多个固定的表面捕获引物上基本上同时进行核酸扩增反应(例如,RCA、MDA、PCR和桥式扩增)。

[0149] 在一些实施例中,支持物上的多个固定的核酸簇彼此流体连通,以允许试剂(例如,酶、核苷酸、二价阳离子等)的溶液流动到支持物上,使得支持物上多个固定的核酸簇可

以基本上同时以大规模并行方式与试剂反应。在一些实施例中,多个固定的核酸簇的流体连通可用于在多个固定的核酸簇上基本上同时进行核苷酸结合测定和/或进行核苷酸聚合反应(例如,引物延伸或测序),并且任选地进行大规模并行测序的检测和成像。

[0150] 在一些实施例中,术语“固定的”和相关术语是指通过共价键或非共价相互作用附接至支持物、或附接至支持物上的涂层、或埋入由支持物上涂层形成的基质内的核酸分子,其中核酸分子包括表面捕获引物、核酸模板分子和捕获引物的延伸产物。捕获引物的延伸产物包括核酸多联体(例如,核酸簇)。核酸分子可以固定在支持物上的预定或随机位置。核酸分子可以固定在支持物上钝化的涂层之上或之内的预定或随机位置。

[0151] 在一些实施例中,术语“固定的”和相关术语是指通过共价键或非共价相互作用附接至支持物、或附接至支持物上的涂层、或埋入由支持物上涂层形成的基底内的酶(例如,聚合酶)。酶可以固定在支持物上的预定或随机位置。酶可以固定在支持物上钝化的涂层之上或之内的预定或随机位置。

[0152] 在一些实施例中,一种或更多种核酸模板分子固定在支持物上,例如固定在支持物上的位点处。在一些实施例中,一种或更多种核酸模板分子被克隆扩增。在一些实施例中,一种或更多种核酸模板分子从支持物上克隆扩增下来(例如,在溶液中),然后沉积到支持物上并固定在支持物上。在一些实施例中,一种或更多种核酸模板分子的克隆扩增反应在支持物上进行,导致固定在支持物上。在一些实施例中,使用核酸扩增反应对一种或更多种核酸模板分子进行克隆扩增(例如,在溶液中或在支持物上),包括以下任何一种或任意组合:聚合酶链式反应(PCR)、多重置换扩增(MDA)、转录介导扩增(TMA)、基于核酸序列的扩增(NASBA)、链置换扩增(SDA)、实时SDA、桥式扩增、等温桥式扩增、滚环扩增(RCA)、环对环扩增、解旋酶依赖性扩增、重组酶依赖性扩增和/或单链结合(SSB)蛋白依赖性扩增。

[0153] 术语“表面引物”和相关术语是指固定至支持物并包含可与核酸模板分子的至少一部分杂交的序列的单链寡核苷酸。表面捕获引物可用于经由杂交将模板分子固定到支持物。表面捕获引物可以在流动、洗涤、抽吸以及温度、pH、盐、化学品和/或酶条件的变化期间,以抵抗引物移除的方式固定至支持物。通常,但不一定,表面捕获引物的5'端可以固定到支持物或支持物上的涂层(或包埋入支持物上的涂层中)。替代性地,表面捕获引物的内部或3'端可以固定到支持物。

[0154] 表面捕获引物的序列可以沿着其长度与核酸模板分子的至少一部分完全或部分互补。支持物可包括多个具有相同序列或具有两个或更多个不同序列的固定表面捕获引物。表面捕获引物可以是任何长度,例如4-50个核苷酸,或50-100个核苷酸,或100-150个核苷酸,或更长的长度。

[0155] 表面捕获引物可以具有末端3'核苷酸,该末端3'核苷酸具有糖3'OH部分,其可延伸用于核苷酸聚合(例如,聚合酶催化的聚合)。表面捕获引物可以具有末端3'核苷酸,该末端3'核苷酸具有与抑制核苷酸聚合的链终止部分连接的3'糖位置。可使用去阻断剂移除(例如,去阻断)3'链终止部分,以将3'端转化为可延伸3'OH端。链终止部分的实例包括:烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基基团、胺基基团、酰胺基基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团。叠氮型链终止部分包括叠氮基、叠氮基和叠氮基甲基基团。去阻断剂的实例包括膦化合物,诸如三(2-羧乙基)膦(TCEP)和双磺基三苯基膦(BS-TPP),用于链终止基

团叠氮基、叠氮基和叠氮甲基基团。去阻断剂的实例包括四(三苯基膦)钯(0) ($\text{Pd}(\text{PPh}_3)_4$)、用吡啶或用2,3-二氯-5,6-二氰基-1,4-苯醌(DDQ),用于链终止基团烷基、烯基、炔基和烯丙基。去阻断剂的实例包括Pd/C,用于链终止基团芳基和苄基。去阻断剂的实例包括膦、 β -巯基乙醇或二硫苏糖醇(DTT),用于链终止基团胺、酰胺、酮、异氰酸酯、磷酸酯、硫代和二硫。去阻断剂的实例包括MeOH中的碳酸钾(K_2CO_3)、吡啶中的三乙胺和乙酸(AcOH)中的Zn,用于碳酸酯链终止基团。去阻断剂的实例包括四丁基氟化铵、吡啶-HF、与氟化铵和三乙胺三氢氟化物,用于链终止基团脲和甲硅烷基。

[0156] 术语“测序”和相关术语指用于从核酸分子获得核苷酸序列信息的方法,通常通过确定核酸分子内的至少一些核苷酸(包括它们的核碱基组分)的身份。在一些实施例中,核酸分子的给定区域的序列信息包括识别被测序的区域内的每个核苷酸。在一些实施例中,测序信息仅确定区域中的一些核苷酸,而一些核苷酸的身份仍未确定或不正确地确定。可以使用任何合适的用于测序的方法。在示例性实施例中,测序可以包括无标记或基于离子的测序方法。在一些实施例中,测序可以包括标记的或含有染料的核苷酸或基于荧光的核苷酸测序方法。在一些实施例中,测序可以包括基于聚合酶克隆的测序或桥式测序方法。在一些实施例中,测序采用聚合酶和多价分子来生成至少一种亲和力复合物,其中各个多价分子包含多个拴系至核心的核苷酸单元。在一些实施例中,测序采用聚合酶和游离核苷酸来进行边合成边测序。在一些实施例中,测序采用连接酶和多个序列特异性寡核苷酸来进行边连接边测序。

[0157] 具体实施方式

[0158] 简介:使用单链夹板链进行环化

[0159] 本公开提供了包含核酸单链夹板链的组合物,包括试剂盒,以及采用所述单链夹板链的方法。

[0160] 单链夹板链(200)可用于一锅多酶反应以环化线性文库分子(例如,参见图1-图5)。单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。单链夹板链的两个区域(例如,(210)和(220))被设计成提高线性核酸分子成环的效率。单链夹板链和使用它们的方法具有有用低至0.25pmol文库分子提供高效环化的优点。

[0161] 本文描述的方法可以手工进行或用于自动化,因为退火和多酶反应可以通过组合一些酶促反应(例如,磷酸化和连接)并通过添加后续酶(例如,核酸外切酶)在单个反应容器(一锅)中进行,无需中间的酒精沉淀或有机萃取。

[0162] 单链夹板链

[0163] 本公开提供了核酸单链夹板链(200),其包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。单链夹板链的两个区域(例如,(210)和(220))被设计成与具有目的序列(110)的单链线性文库分子(100)末端的通用衔接子序列杂交。例如,单链夹板链的第一区域(210)与文库分子的一端杂交,并且单链夹板链的第二区域(220)与文库分子的另一端杂交,由此使文库分子成环,以产生文库-夹板复合物(300),其包括缺口(例如,参见图1-图5)。可以酶促连接缺口以产生共价闭环环状分子(400),其中文库分子的末端共价接合。

[0164] 线性核酸文库分子的末端序列分别包含至少第一通用衔接子序列和第二通用衔接

接子序列。在一些实施例中,线性文库分子的第一通用衔接子序列和第二通用衔接子序列分别包含针对固定在支持物上的第一捕获引物和第二捕获引物的结合序列。

[0165] 在一些实施例中,单链夹板链的第一区域(210)包括第一通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列,或针对压实寡核苷酸的通用结合序列。

[0166] 在一些实施例中,单链夹板链的第二区域(220)包括第二通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列,或针对压实寡核苷酸的通用结合序列。

[0167] 在一些实施例中,单链夹板链(200)在其5'和/或3'端包含一个或多个硫代磷酸酯键以赋予核酸外切酶抗性。在一些实施例中,单链夹板链(200)在内部位置包含一个或多个硫代磷酸酯键以赋予核酸内切酶抗性。在一些实施例中,单链夹板链(200)在其5'和/或3'端或在内部位置包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,单链夹板链(200)的5'端为磷酸化或非磷酸化的。在一些实施例中,单链夹板链(200)的3'端包含末端3'OH基团或末端3'阻断基团。

[0168] 在一些实施例中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列

[0169] 5'-ACCCTGAAAGTACGTGCATTACATG-3'(SEQ ID NO:193)(例如,图7A-图7C)。

[0170] 在一些实施例中,文库分子中的第一左通用衔接子序列(120)包括序列5'-CATGTAATGCACGTACTTTCAGGGT-3'(SEQ ID NO:199)。

[0171] 在一些实施例中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括以下序列

[0172] 5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)(例如,图7A-图7C)。

[0173] 在一些实施例中,文库分子中的第一右通用衔接子序列(130)包括序列5'-AGTCGTCGACGCTCACCTGATC-3'(SEQ ID NO:200)。

[0174] 在一些实施例中,单链夹板链(200)包括序列

[0175] 5'-ACCCTGAAAGTACGTGCATTACATGGATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:195)(例如,图7A-图7C)。

[0176] 在一些实施例中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列

[0177] 5'-TCGGTGGTCGCCGTATCATT-3'(SEQ ID NO:196)。

[0178] 在一些实施例中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括以下序列

[0179] 5'-CAAGCAGAAGACGGCATA CGA-3'(SEQ ID NO:197)。

[0180] 在一些实施例中,单链夹板链(200)包括序列5'-TCGGTGGTCGCCGTATCATTCAAGCAG AAGACGGCATA CGA-3'(SEQ ID NO:198)。

[0181] 文库-夹板复合物

[0182] 本公开提供了文库-夹板复合物(300),其包含:(i)单链核酸文库分子(100),其包

括目的序列(110),该目的序列在一侧上侧接至少第一左通用衔接子序列(120),并且在另一侧上侧接至少第一右通用衔接子序列(130);以及(ii)单链夹板链(200)包括与线性单链文库分子的一端上的序列杂交的第一区域(210)和与线性单链文库分子的另一端上的序列杂交的第二区域(220)。在文库-夹板复合物(300)中,单链夹板链的第一区域(210)与文库分子的至少第一左通用衔接子序列(120)杂交,并且单链夹板链的第二区域(220)与文库分子的至少第一右通用衔接子序列(130)杂交,由此使文库分子成环以产生文库-夹板复合物(300)(例如参见图1-图5)。

[0183] 在文库-夹板复合物(300)中,单链夹板链的第一区域(210)包含第一通用衔接子序列,其可以与线性核酸文库分子一端的第一通用结合序列杂交(例如,参见图1-图5)。在一些实施例中,单链夹板链的第一区域(210)包括第一通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列、针对压实寡核苷酸的通用结合序列。在一些实施例中,单链夹板链(200)的长度可为20-150个核苷酸、或长度为60-100个核苷酸、长度为70-90个核苷酸或长度为60-80个核苷酸。在一些实施例中,单链夹板链(200)在5'和/或3'端包含一个或多个硫代磷酸酯键以赋予核酸外切酶抗性。在一些实施例中,单链夹板链(200)在内部位置包含一个或多个硫代磷酸酯键以赋予核酸内切酶抗性。在一些实施例中,单链夹板链(200)在5'和/或3'端或在内部位置包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,单链夹板链(200)的5'端为磷酸化的或缺少磷酸酯基团。在一些实施例中,单链夹板链(200)的3'端包括末端3'OH基团或末端3'阻断基团。

[0184] 单链夹板链的第二区域(220)包含第二通用衔接子序列,其可以与线性核酸文库分子另一端的第二通用结合序列杂交(例如,参见图1-图5)。在一些实施例中,单链夹板链的第二区域(220)包括第二通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列,或针对压实寡核苷酸的通用结合序列。

[0185] 在文库-夹板复合物(300)中,单链夹板链的第一区域(210)与文库分子的至少第一左通用衔接子序列(120)杂交,并且单链夹板链的第二区域(220)与文库分子的至少第一右通用衔接子序列(130)杂交,由此使文库分子成环以产生文库-夹板复合物(300)。文库-夹板复合物(300)包含文库分子的5'端与文库分子的3'端之间的缺口(例如,参见图1-图5)。在一些实施例中,缺口是可酶促连接的。

[0186] 在文库-夹板复合物(300)中,单链夹板链(210)的第一区域可以与双链核酸文库分子的有义链或反义链杂交。在文库-夹板复合物(300)中,单链夹板链(220)的第二区域可以与双链核酸文库分子的有义链或反义链杂交。双链核酸文库分子可以变性以产生单链的有义和反义文库链。单链文库分子(100)可以是有义链或反义链。

[0187] 在文库-夹板复合物(300)中,单链夹板链的第一区域(210)不与目的序列(110)杂交,并且单链夹板链的第二区域(220)不与目的序列(110)杂交。

[0188] 在一些实施例中,在文库-夹板复合物(300)中,单链文库分子(100)的5'端被磷酸化或缺少磷酸酯基团。在一些实施例中,单链文库分子的3'端包括末端3'OH基团或末端3'阻断基团。

[0189] 在一些实施例中,核酸文库分子(100)进一步包含第二左通用衔接子序列(140)。

在一些实施例中,核酸文库分子(100)进一步包含第二右通用衔接子序列(150)。示例性文库-夹板复合物(300)示出于图1-图5中。在一些实施例中,核酸文库分子(100)可进一步包含额外的左通用衔接子序列和/或右通用衔接子序列。

[0190] 在一些实施例中,核酸文库分子(100)进一步包含第一左索引序列(160)。在一些实施例中,核酸文库分子(100)进一步包含第一右索引序列(170)。第一右索引序列(170)的长度可以是3-20个核苷酸。在一些实施例中,第一左索引序列(160)包括样品索引序列。第一左索引序列(160)的长度可以是3-20个核苷酸。在一些实施例中,第一右索引序列(170)包括另一个样品索引序列。左样品索引序列和右样品索引序列(例如,(160)和(170))的序列可以彼此相同或不同。样品索引序列可用于区分多重测定中从不同样品源获得的目的序列。示例性文库-夹板复合物(300)示出于图1-图5中。

[0191] 示例性第一左索引序列(160)和第一右索引序列(170)的列表在图9A-图9F的表1中提供。第一右索引序列(170)可以包括短随机序列(例如,NNN)或缺少短随机序列。短随机序列的长度可以是3-20个核苷酸。

[0192] 在一些实施例中,核酸文库分子(100)进一步包含位于本文所述的任何通用衔接子序列之间的至少一个接合衔接子序列(例如,参见图5)。例如,第一左接合衔接子序列(125)可位于第一左通用衔接子序列(120)与第一左索引序列(160)之间。第二左接合衔接子序列(165)可位于第一左索引序列(160)与第二左通用衔接子序列(140)之间。第三接合衔接子序列(145)可位于第二左通用衔接子序列(140)与目的序列(110)之间。第一右接合衔接子序列(135)可位于第一右通用序列(130)与第一右索引序列(170)之间。第二右接合衔接子序列(175)可位于第一右索引序列(170)与第二右通用衔接子序列(150)之间。第三右接合衔接子序列(155)可位于第二右通用衔接子序列(150)与目的序列(110)之间。任何接合衔接子序列包含任何序列并且长度可为3-60个核苷酸。任何接合衔接子序列包含通用序列或唯一序列。任何接合衔接子序列包括具有3-20个核苷酸的随机序列(例如,NNN)。任何接合衔接子序列包含扩增引物、测序引物或压实寡核苷酸的结合序列。任何接合衔接子序列包含固定的表面引物(例如,捕获引物)的结合序列。任何接合衔接子序列均包含样品索引序列。任何接合衔接子序列均包含唯一识别序列。任何接合衔接子序列,特别是接合衔接子序列(145)包括Tn5转座子端序列5'-AGATGTGTATAAGAGACAG-3'(SEQ ID NO:209)。任何接合衔接子序列,特别是接合衔接子序列(155)包括Tn5转座子端序列5'-CTGTCTCTTATACACATCT-3'(SEQ ID NO:210)。可以在适合形成转座子突触复合物的条件下,经由转座酶介导的反应将Tn5转座子端序列引入文库分子(100),该反应包括使双链输入DNA(例如,基因组DNA)与Tn-5型转座酶和双链寡核苷酸接触,该双链寡核苷酸包含与通用衔接子序列或样品索引序列链接的Tn转座子端序列(SEQ ID NO:209)。在双链寡核苷酸中,Tn转座子端序列(SEQ ID NO:209)可位于相对于通用衔接子序列或样品索引序列的5'或3'。

[0193] 通过使用一个或两个索引序列(例如,左索引序列和/或右索引序列)准备具有样品索引的文库来启动多重工作流程。可使用第一左索引序列(160)和/或第一右索引序列(170),使用从不同来源分离的输入核酸,来制备单独的具有样品索引的文库。可以将具有样品索引的文库汇集在一起以产生多重文库混合物,并且可以对汇集的文库进行成环、扩增和/或测序。插入区域的序列连同第一左索引序列(160)和/或第一右索引序列(170)可用

于识别输入核酸的来源。在一些实施例中,可以将任意数量的具有样品索引的文库汇集在一起,例如可以汇集2-10、或10-50、或50-100、或100-200、或多于200个具有样品索引的文库。示例性的核酸来源包括天然存在的、重组的或化学合成的来源。示例性的核酸来源包括单个细胞、多个细胞、组织、生物流体、环境样品或整个生物体。示例性的核酸来源包括新鲜的、冷冻的、新鲜冷冻的或存档的来源(例如,福尔马林固定石蜡包埋的;FFPE)。技术人员将认识到核酸可以从许多其他来源分离。核酸文库分子可以以单链或双链形式制备。

[0194] 在一些实施例中,核酸文库分子(100)进一步包含:任选的第一左唯一识别序列(180)。在一些实施例中,核酸文库分子(100)进一步包含任选的第一右唯一识别序列(190)。在一些实施例中,第一左唯一识别序列(180)和第一右唯一识别序列(190)各自包含一段序列,该段序列用于在其他目的分子序列的群体中,唯一地识别其上附加有唯一衔接子的各个目的序列(例如,插入序列)。在一些实施例中,第一左唯一识别序列(180)和/或第一右唯一识别序列(190)可用于分子标记。在一些实施例中,唯一识别序列(180)和/或(190)包含2-12个或更多个具有已知序列的核苷酸。例如,唯一识别序列包括已知的随机序列,其中每个位置处的核苷酸随机选自具有碱基A、G、C、T或U的核苷酸。唯一识别序列(180)和/或(190)可以是用于分子标记程序。示例性文库-夹板复合物(300)示出于图1-图5中。

[0195] 在一些实施例中,核酸文库分子(100)包含以下任意一者或者两者或更多者的任意组合:第一左通用衔接子序列(120);第二左通用衔接子序列(140);第一左索引序列(160);第一左唯一识别序列(180);第一右通用衔接子序列(130);第二右通用衔接子序列(150);第一右索引序列(170);和/或第一右唯一识别序列(190)。图1-图5示出了一些示例性文库-夹板复合物(300)。

[0196] 在一些实施例中,第一左通用衔接子序列(120)和/或第二左通用衔接子序列(140)包括:针对正向测序引物或反向测序引物的通用结合序列;针对第一表面引物或第二表面引物的通用结合序列;针对正向扩增引物或反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。在一些实施例中,核酸文库分子(100)可进一步包含额外的左通用衔接子序列。

[0197] 在一些实施例中,第一右通用衔接子序列(130)和/或第二右通用衔接子序列(150)包括:针对正向测序引物或反向测序引物的通用结合序列;针对第一表面引物或第二表面引物的通用结合序列;针对正向扩增引物或反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。在一些实施例中,核酸文库分子(100)可进一步包含额外的右通用衔接子序列。

[0198] 在一些实施例中,示例性文库-夹板复合物(300)包含:(a)单链核酸文库分子(100);以及(b)单链夹板链(200)。

[0199] 在示例性文库-夹板复合物(300)中,单链核酸文库分子(100)包含以5'至3'顺序排列的组分:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列;(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列;(iii)目的序列(110);(iv)第二右通用衔接子序列(150),其具有针对第二测序引物的结合序列;以及(v)第一右通用衔接子序列(130),具有针对第一表面引物的结合序列。

[0200] 在示例性的文库-夹板复合物(300)中,单链夹板链(200)包括以5'至3'顺序排列的组分:第一区域(210);以及第二区域(220)。

[0201] 在示例性的文库-夹板复合物(300)中,单链夹板链(200)的部分与文库分子(100)的部分杂交,由此使文库分子成环以产生文库-夹板复合物(300),使得单链夹板链的第一区域(210)与针对第一表面引物的结合序列(120)杂交,并且单链夹板链的第二区域(220)与针对第二表面引物的结合序列(130)杂交。文库-夹板复合物(300)包含文库分子的5'端与文库分子的3'端之间的缺口,其中该缺口是可酶促连接的。示例性文库-夹板复合物(300)示出于图1-图5中。

[0202] 在示例性文库-夹板复合物(300)中,单链夹板链的第一区域(210)不与目的序列(110)杂交,并且单链夹板链的第二区域(220)不与目的序列(110)杂交。

[0203] 在一些实施例中,本文描述的任何文库-夹板复合物(300)包含多个文库-夹板复合物(300),其中多个文库-夹板复合物中的各个文库-夹板复合物的目的序列(110)包含相同的目的序列或不同的目的序列。

[0204] 在一些实施例中,文库分子中的第一左通用衔接子序列(120)包括序列5'-CATGTAATGCACGTACTTTCAGGGT-3'(SEQ ID NO:199)。

[0205] 在一些实施例中,文库分子中的第一左通用衔接子序列(120)包括序列5'-AATGATACGGCGACCACCGA-3'(SEQ ID NO:203)。

[0206] 在一些实施例中,文库分子中的第二左通用衔接子序列(140)包括序列5'-CGTGC TGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:201)。

[0207] 在一些实施例中,文库分子中的第二左通用衔接子序列(140)包括序列5'-ACACT CTTTCCCTACACGACGCTCTTCCGATCT-3'(SEQ ID NO:205)。

[0208] 在一些实施例中,文库分子中的第二左通用衔接子序列(140)包括序列5'-TCGTC GGCAGCGTCAGATGTGTATAAGAGACAG-3'(SEQ ID NO:206)。

[0209] 在一些实施例中,文库分子中的第二右通用衔接子序列(150)包括序列5'-ATGTC GGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:202)。

[0210] 在一些实施例中,文库分子中的第二右通用衔接子序列(150)包括序列5'-AGATC GGAAGAGCACACGTCTGAACTCCAGTCAC-3'(SEQ ID NO:207)。

[0211] 在一些实施例中,文库分子中的第二右通用衔接子序列(150)包括序列5'-CTGTC TCTTATACACATCTCCGAGCCACGAGAC-3'(SEQ ID NO:208)。

[0212] 在一些实施例中,文库分子中的第一右通用衔接子序列(130)包括序列5'-AGTCGTCGAGCCTCACCTGATC-3'(SEQ ID NO:200)。

[0213] 在一些实施例中,文库分子中的第一右通用衔接子序列(130)包括序列5'-TCGTATGCCGTCTTCTGCTTG-3'(SEQ ID NO:204)。

[0214] 在一些实施例中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列5'-ACCCTGAAAGTACGTGCATTACATG-3'(SEQ ID NO:193)。

[0215] 在一些实施例中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括以下序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)。

[0216] 在一些实施例中,单链夹板链(200)包括序列5'-ACCCTGAAAGTACGTGCATTACATGGA TCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:195)。例如参见图7A-图7C。

[0217] 在一些实施例中,在该方法中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列5'-TCGGTGGTCGCCGTATCATT-3'(SEQ ID NO:196)。

[0218] 在一些实施例中,在该方法中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括序列5'-CAAGCAGAAGACGGCATACGA-3'(SEQ ID NO:197)。

[0219] 在一些实施例中,在该方法中,单链夹板链(200)包括序列5'-TCGGTGGTCGCCGTATCATTCAAGCAGAAGACGGCATACGA-3'(SEQ ID NO:198)。

[0220] 本公开提供了包含多个本文所述的任何文库-夹板复合物(300)的反应混合物。在一些实施例中,反应混合物包含多个本文所述的任何文库-夹板复合物(300)以及T4多核苷酸激酶。在一些实施例中,反应混合物包含多个本文所述的任何文库-夹板复合物(300)以及连接酶。在一些实施例中,反应混合物包含多个本文所述的任何文库-夹板复合物(300)以及T4多核苷酸激酶和连接酶。在一些实施例中,连接酶包含T7 DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。

[0221] 共价闭合环状分子

[0222] 本公开提供共价闭合环状文库分子(400),其包含:目的序列(110)、至少第一左通用衔接子序列(120)和至少第一右通用衔接子序列(130)。示例性的共价闭合环状文库分子示出于图6A-图6C中。在一些实施例中,共价闭合环状文库分子(400)进一步包含第二左通用衔接子序列(140)。在一些实施例中,共价闭合环状分子(400)进一步包含第二右通用衔接子序列(150)。在一些实施例中,共价闭合环状分子(400)进一步包含额外的左和/或右通用衔接子序列。

[0223] 在一些实施例中,共价闭合环状分子(400)进一步包含第一左索引序列(160)和/或第一右索引序列(170)。第一左索引序列(160)的长度可以是3-20个核苷酸。第一右索引序列(170)的长度可以是3-20个核苷酸。样品索引序列可用于区分多重测定中从不同样品源获得的目的序列。示例性第一左索引序列(160)和第一右索引序列(170)的列表在图9A-图9F的表1中提供。第一右索引序列(170)和/或第一左索引序列(160)可以包括短随机序列(例如,NNN)或缺少短随机序列。短随机序列的长度可以是3-20个核苷酸。

[0224] 通过使用一个或两个索引序列(例如,左索引序列和/或右索引序列)准备具有样品索引的文库来启动多重工作流程。可使用第一左索引序列(160)和/或第一右索引序列(170),使用从不同来源分离的输入核酸,来制备单独的具有样品索引的文库。可以将具有样品索引的文库汇集在一起以产生多重文库混合物,并且可以对汇集的文库进行成环、扩增和/或测序。插入区域的序列连同第一左索引序列(160)和/或第一右索引序列(170)可用于识别输入核酸的来源。在一些实施例中,可以将任意数量的具有样品索引的文库汇集在一起,例如可以汇集2-10、或10-50、或50-100、或100-200、或多于200个具有样品索引的文库。示例性的核酸来源包括天然存在的、重组的或化学合成的来源。示例性的核酸来源包括单个细胞、多个细胞、组织、生物流体、环境样品或整个生物体。示例性的核酸来源包括新鲜的、冷冻的、新鲜冷冻的或存档的来源(例如,福尔马林固定石蜡包埋的;FFPE)。技术人员将认识到核酸可以从许多其他来源分离。核酸文库分子可以以单链或双链形式制备。

[0225] 在一些实施例中,共价闭合环状分子(400)进一步包含任选的第一左唯一识别序

列(180)和/或任选的第一右唯一识别序列(190)。在一些实施例中,第一左唯一识别序列(180)和第一右唯一识别序列(190)各自包含一段序列,该段序列用于在其他目的分子序列(例如,分子标记)的群体中,唯一地识别其上附加有唯一衔接子的各个目的序列(例如,插入序列)。

[0226] 在一些实施例中,共价闭合环状分子(400)包括任意一者或者两者或更多者的任意组合:第一左通用衔接子序列(120);第二左通用衔接子序列(140);第一左索引序列(160);第一左唯一识别序列(180);第一右通用衔接子序列(130);第二右通用衔接子序列(150);第一右索引序列(170);和/或第一右唯一识别序列(190)。在一些实施例中,第一左索引序列(160)包括样品索引序列。在一些实施例中,第一右索引序列(170)包括另一个样品索引序列。右样品索引序列和左样品索引序列(例如,(170)和(160))的序列可以彼此相同或不同。样品索引序列可用于区分多重测定中从不同样品源获得的目的序列。在一些实施例中,第一左唯一识别序列(180)和第一右唯一识别序列(190)各自包含一段序列,该段序列用于在其他目的分子序列的群体中,唯一地识别其上附加有唯一衔接子的各个目的序列(例如,插入序列)。在一些实施例中,第一左唯一识别序列(180)和/或第一右唯一识别序列(190)可用于分子标记。在一些实施例中,唯一识别序列包含2-12个或更多个具有已知序列的核苷酸。例如,唯一识别序列(180)和/或(190)包括已知的随机序列,其中唯一识别序列中每个位置处的核苷酸随机选自具有核碱基A、G、C、T或U。唯一识别序列(180)和/或(190)可用于分子标记程序。示例性共价闭合环状文库分子示出于图6A-图6C中。

[0227] 在一些实施例中,在共价闭合环状分子(400)中,第一左通用衔接子序列(120)和/或第二左通用衔接子序列(140)包含:针对正向测序引物或反向测序引物的通用结合序列;针对第一表面引物或第二表面引物的通用结合序列;针对正向扩增引物或反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。在一些实施例中,共价闭合环状分子(400)可进一步包含额外的左通用衔接子序列。

[0228] 在一些实施例中,在共价闭合环状分子(400)中,第一右通用衔接子序列(130)和/或第二右通用衔接子序列(150)包括:针对正向测序引物或反向测序引物的通用结合序列;针对第一表面引物或第二表面引物的通用结合序列;针对正向扩增引物或反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。在一些实施例中,共价闭合环状分子(400)可进一步包含额外的右通用衔接子序列。

[0229] 在一些实施例中,共价闭合环状分子(400)进一步包含至少一个接合衔接子序列,其位于本文所述的任何通用衔接子序列之间(例如,参见图5)。例如,第一左接合衔接子序列(125)可位于第一左通用衔接子序列(120)与第一左索引序列(160)之间。第二左接合衔接子序列(165)可位于第一左索引序列(160)与第二左通用衔接子序列(140)之间。第三接合衔接子序列(145)可位于第二左通用衔接子序列(140)与目的序列(110)之间。第一右接合衔接子序列(135)可位于第一右通用序列(130)与第一右索引序列(170)之间。第二右接合衔接子序列(175)可位于第一右索引序列(170)与第二右通用衔接子序列(150)之间。第三右接合衔接子序列(155)可位于第二右通用衔接子序列(150)与目的序列(110)之间。任何接合衔接子序列包含任何序列并且长度可为3-60个核苷酸。任何接合衔接子序列包含通用序列或唯一序列。任何接合衔接子序列包括具有3-20个核苷酸的随机序列(例如,NNN)。任何接合衔接子序列包含扩增引物、测序引物或压实寡核苷酸的结合序列。任何接合衔接

子序列包含固定的表面引物(例如,捕获引物)的结合序列。任何接合衔接子序列均包含样品索引序列。任何接合衔接子序列均包含唯一识别序列。任何接合衔接子序列,特别是接合衔接子序列(145)包括Tn5转座子端序列5'-AGATGTGTATAAGAGACAG-3'(SEQ ID NO:209)。任何接合衔接子序列,特别是接合衔接子序列(155)包括Tn5转座子端序列5'-CTGTCTCTTATACACATCT-3'(SEQ ID NO:210)。可以在适合形成转座子突触复合物的条件下,经由转座酶介导的反应将Tn5转座子端序列引入文库分子(100),该反应包括使双链输入DNA(例如,基因组DNA)与Tn-5型转座酶和双链寡核苷酸接触,该双链寡核苷酸包含与通用衔接子序列或样品索引序列链接的Tn转座子端序列(SEQ ID NO:209)。在双链寡核苷酸中,Tn转座子端序列(SEQ ID NO:209)可位于相对于通用衔接子序列或样品索引序列的5'或3'。

[0230] 在一些实施例中,示例性共价闭合环状分子(400)包含:(i)第一左通用衔接子序列(120),其具有针对第二表面引物(120)的结合序列;(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列;(iii)目的序列(110);(iv)第二右通用衔接子序列(150),其具有针对第二测序引物的结合序列;(v)第一右通用衔接子序列(130),其具有针对第一表面引物的结合序列(130);以及(vi)任选地,共价闭合环状分子(400)与单链夹板链(200)杂交(例如,参见图6A-图6C)。

[0231] 在示例性共价闭合环状分子(400)中,单链夹板链的第一区域和第二区域(分别为210和220)不与目的序列(110)杂交。

[0232] 在一些实施例中,本文所述的任何共价闭合环状分子(400)包含多个共价闭合环状分子(400),其中多个共价闭合环状分子中的各个共价闭合环状分子(400)的目的序列(110)包含相同的目的序列或不同的目的序列。

[0233] 在一些实施例中,共价闭合环状分子中的第一左通用衔接子序列(120)包括序列5'-CATGTAATGCACGTACTTTCAGGGT-3'(SEQ ID NO:199)。

[0234] 在一些实施例中,共价闭合环状分子中的第一左通用衔接子序列(120)包括序列5'-AATGATACGGCGACCACCGA-3'(SEQ ID NO:203)。

[0235] 在一些实施例中,共价闭合环状分子中的第二左通用衔接子序列(140)包括序列5'-CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:201)。

[0236] 在一些实施例中,共价闭合环状分子中的第二左通用衔接子序列(140)包括序列5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'(SEQ ID NO:205)。

[0237] 在一些实施例中,共价闭合环状分子中的第二左通用衔接子序列(140)包括序列5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'(SEQ ID NO:206)。

[0238] 在一些实施例中,共价闭合环状分子中的第二右通用衔接子序列(150)包括序列5'-ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:202)。

[0239] 在一些实施例中,共价闭合环状分子中的第二右通用衔接子序列(150)包括序列5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'(SEQ ID NO:207)。

[0240] 在一些实施例中,共价闭合环状分子中的第二右通用衔接子序列(150)包括序列5'-CTGTCTCTTATACACATCTCCGAGCCACGAGAC-3'(SEQ ID NO:208)。

[0241] 在一些实施例中,共价闭合环状分子中的第一右通用衔接子序列(130)包括序列5'-AGTCGTCGCAGCCTCACCTGATC-3'(SEQ ID NO:200)。

[0242] 在一些实施例中,共价闭合环状分子中的第一右通用衔接子序列(130)包括序列5'-TCGTATGCCGTCTTCTGCTTG-3'(SEQ ID NO:204)。

[0243] 在一些实施例中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列5'-ACCCTGAAAGTACGTGCATTACATG-3'(SEQ ID NO:193)。

[0244] 在一些实施例中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括以下序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)。

[0245] 在一些实施例中,单链夹板链(200)包括序列5'-ACCCTGAAAGTACGTGCATTACATGGA TCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:195)。例如参见图7A-图7C。

[0246] 在一些实施例中,在该方法中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列5'-TCGGTGGTCGCCGTATCATT-3'(SEQ ID NO:196)。

[0247] 在一些实施例中,在该方法中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括序列5'-CAAGCAGAAGACGGCATACGA-3'(SEQ ID NO:197)。

[0248] 在一些实施例中,在该方法中,单链夹板链(200)包括序列5'-TCGGTGGTCGCCGTAT CATTCAAGCAGAAGACGGCATACGA-3'(SEQ ID NO:198)。

[0249] 本公开提供了一种反应混合物,其包含多个本文所述的任何共价闭合环状分子(400)和至少一种核酸外切酶。在一些实施例中,核酸外切酶包含核酸外切酶I、不耐热核酸外切酶I和/或T7核酸外切酶中的任意一者或者两者或更多者的任意组合。

[0250] 包括单链夹板链的试剂盒

[0251] 本公开提供了至少一种用于使线性核酸文库分子成环以形成多个文库-夹板复合物(300)的试剂盒,每个文库-夹板复合物具有缺口。在一些实施例中,该试剂盒可用于使具有目的序列(110)的单链核酸文库分子成环,该目的序列在一侧上侧接至少第一左通用衔接子序列(120),并且在另一侧上侧接至少第一右通用衔接子序列(130)。在一些实施例中,成环的文库分子可转化为共价闭合环状分子,其可经历滚环扩增(RCA)反应以生成核酸多联体。多联体可以固定到支持物。固定的多联体可以经历大规模平行测序反应。

[0252] 本公开提供了包含核酸单链夹板链(200)的试剂盒,每个核酸单链夹板链包括第一区域(210)和第二区域(220)。第一区域(210)包含第一通用衔接子序列,其可以与线性核酸文库分子一端的第一通用结合序列杂交。第二区域(220)包含第二通用衔接子序列,其可以与线性核酸文库分子另一端的第二通用结合序列杂交。在一些实施例中,单链夹板链的第一区域(210)包括第一通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列,或针对压实寡核苷酸的通用结合序列。在一些实施例中,单链夹板链的第二区域(220)包括第二通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列,或针对压实寡核苷酸的通用结合序列。在一些实施例中,单链夹板链(200)的长度可为20-150个核苷酸、或长度为60-100个核苷酸、长度为70-

90个核苷酸或长度为60-80个核苷酸。在一些实施例中,单链夹板链(200)在5'和/或3'端包含一个或多个硫代磷酸酯键以赋予核酸外切酶抗性。在一些实施例中,单链夹板链(200)在内部位置包含一个或多个硫代磷酸酯键以赋予核酸内切酶抗性。在一些实施例中,单链夹板链(200)在5'和/或3'端或在内部位置包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,单链夹板链(200)的5'端为磷酸化或非磷酸化的。在一些实施例中,单链夹板链(200)的3'端包含末端3'OH基团或末端3'阻断基团。

[0253] 在一些实施例中,试剂盒包括单链夹板链(200),其具有包括序列5'-ACCCTGAAAGTACGTGCATTACATG-3'(SEQ ID NO:193)的第一区域(210)。

[0254] 在一些实施例中,试剂盒包括单链夹板链(200),其具有包括序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)的第二区域(220)。

[0255] 在一些实施例中,试剂盒包括单链夹板链(200),其包括序列5'-ACCCTGAAAGTACGTGCATTACATGGATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:195)。

[0256] 在一些实施例中,试剂盒进一步包含具有左通用衔接子序列(120)的衔接子,用于制备多个文库分子,其中文库分子将包括可以结合表面引物的序列5'-CATGTAATGCACGTACTTTTCAGGGT-3'(SEQ ID NO:199)。在一些实施例中,具有左通用衔接子序列(120)的衔接子还包括样品索引序列(160)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。例如参见图12A、图12B和图12C。

[0257] 在一些实施例中,试剂盒进一步包含具有右通用衔接子序列(130)的衔接子,用于制备多个文库分子,其中文库分子将包括可以结合表面引物的序列5'-AGTCGTCGCAGCCTCACCTGATC-3'(SEQ ID NO:200)。在一些实施例中,具有右通用衔接子序列(130)的衔接子还包括样品索引序列(170)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。例如参见图12A、图12B和图12C。

[0258] 在一些实施例中,试剂盒进一步包含具有左通用衔接子序列(140)的衔接子,用于制备多个文库分子,其中文库分子将包括以下序列5'-CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:201),其可以结合测序引物。在一些实施例中,具有左通用衔接子序列(140)的衔接子还包括样品索引序列(160)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。例如参见图12A、图12B和图12C。

[0259] 在一些实施例中,试剂盒进一步包含具有右通用衔接子序列(150)的衔接子,用于制备多个文库分子,其中文库分子将包括以下序列5'-ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:202),其可以结合测序引物。在一些实施例中,具有右通用衔接子序列(150)的衔接子还包括样品索引序列(170)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。例如参见图12A、图12B和图12C。

[0260] 在一些实施例中,试剂盒包括单链夹板链(200),其具有包括序列5'-TCGGTGGTCGCCGTATCATT-3'(SEQ ID NO:196)的第一区域(210)。

[0261] 在一些实施例中,试剂盒包括单链夹板链(200),其具有包括序列5'-CAAGCAGAAGACGGCATAACGA-3'(SEQ ID NO:197)的第二区域(220)。

[0262] 在一些实施例中,试剂盒包括单链夹板链(200),其包括序列5'-TCGGTGGTCGCCGTATCATTCAAGCAGAAGACGGCATAACGA-3'(SEQ ID NO:198)。

[0263] 在一些实施例中,试剂盒进一步包含具有左通用衔接子序列(120)的衔接子,用于

制备多个文库分子,其中文库分子将包括可以结合表面引物的序列5'-AATGATACGGCGACCACCGA-3'(SEQ ID NO:203)。在一些实施例中,具有左通用衔接子序列(120)的衔接子还包括样品索引序列(160)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。

[0264] 在一些实施例中,试剂盒进一步包含具有右通用衔接子序列(130)的衔接子,用于制备多个文库分子,其中文库分子将包括可以结合表面引物的序列5'-TCGTATGCCGTCTTCTGCTTG-3'(SEQ ID NO:204)。在一些实施例中,具有右通用衔接子序列(130)的衔接子还包括样品索引序列(170)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。

[0265] 在一些实施例中,试剂盒进一步包括具有左通用衔接子序列(140)的衔接子,用于制备多个文库分子,其中文库分子将包括以下序列5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'(SEQ ID NO:205),其可以结合测序引物。在一些实施例中,具有左通用衔接子序列(140)的衔接子还包括样品索引序列(160)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。

[0266] 在一些实施例中,试剂盒进一步包括具有右通用衔接子序列(150)的衔接子,用于制备多个文库分子,其中文库分子将包括以下序列5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'(SEQ ID NO:207),其可以结合测序引物。在一些实施例中,具有右通用衔接子序列(150)的衔接子还包括样品索引序列(170)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。

[0267] 在一些实施例中,试剂盒进一步包括具有左通用衔接子序列(140)的衔接子,用于制备多个文库分子,其中文库分子将包括以下序列5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'(SEQ ID NO:206),其可以结合测序引物。在一些实施例中,具有左通用衔接子序列(140)的衔接子还包括样品索引序列(160)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。

[0268] 在一些实施例中,试剂盒进一步包括具有右通用衔接子序列(150)的衔接子,用于制备多个文库分子,其中文库分子将包括以下序列5'-CTGTCTCTTATACACATCTCCGAGCCCAGACAGAC-3'(SEQ ID NO:208),其可以结合测序引物。在一些实施例中,具有右通用衔接子序列(150)的衔接子还包括样品索引序列(170)。衔接子可以是单链衔接子(例如,PCR引物)、双链衔接子、气泡衔接子或Y形衔接子。

[0269] 在一些实施例中,试剂盒进一步包括多个样品索引,其包括多个第一左索引(160)和/或多个第一右索引序列(170)。第一左索引序列(160)的长度可以是3-20个核苷酸。第一右索引序列(170)的长度可以是3-20个核苷酸。在一些实施例中,试剂盒可包括容纳单个第一左索引(160)或单个第一右索引(170)的单独的容器或多孔板(例如,96孔板)的孔。在一些实施例中,试剂盒可包括容纳一对单个第一左索引(160)和单个第一右索引(170)的单独的容器或多孔板的孔。在一些实施例中,试剂盒包括位于多孔板(例如,96孔板)中的第一左索引(160)和/或多个第一右索引序列(170)。示例性第一左索引序列(160)和第一右索引序列(170)的列表在图9A-图9F的表1中提供。第一左索引序列(160)可以包括短随机序列(例如,NNN)或缺少短随机序列。第一右索引序列(170)可以包括短随机序列(例如,NNN)或缺少短随机序列。短随机序列的长度可以是3-20个核苷酸。

[0270] 在一些实施例中,试剂盒包括多个有尾PCR引物,该多个有尾PCR引物包括第一左索引序列(160)。有尾PCR引物可以包括:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列;(ii)第一左索引序列(160);(iii)第二左通用衔接子序列(140),其具有针对第一测序引物或第二测序引物的结合序列。第一左索引序列(160)还可以包括随机序列(例如,NNN)。参见例如图17。

[0271] 在一些实施例中,试剂盒包括多个有尾PCR引物,该多个有尾PCR引物包括第一右索引序列(170)。有尾PCR引物可以包括:(i)第一右通用衔接子序列(130),其具有针对第一表面引物的结合序列;(ii)第一右索引序列(170);以及(iii)第二右通用衔接子序列(150),其具有针对第一测序引物或第二测序引物的结合序列。第一右索引序列(170)还可以包括随机序列(例如,NNN)。参见例如图17。

[0272] 在一些实施例中,试剂盒包括核酸单链夹板链(200)并且进一步包括T4多核苷酸激酶。在一些实施例中,试剂盒进一步包括连接酶,其中连接酶包含T7 DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。在一些实施例中,试剂盒进一步包括至少一种核酸内切酶,其包括核酸外切酶I、不耐热核酸外切酶I和/或T7核酸外切酶中的任意一者或者两者或更多者的任意组合。

[0273] 在一些实施例中,试剂盒包括至少一种用于使多个单链夹板链(200)和多个核酸文库分子(100)杂交的缓冲液。在一些实施例中,试剂盒包括一种用于在单个反应容器中进行多种酶促反应的缓冲液,包括(i)使单链夹板链(200)的5'端磷酸化,(ii)连接文库-夹板复合物(300)中的缺口,和/或(iii)核酸外切酶消化来自共价闭合环状分子(400)单链夹板链(200)。替代性地,试剂盒包括两种或更多种单独的缓冲液,其中第一缓冲液可用于进行磷酸化反应,第二缓冲液可用于进行连接反应,并且第三缓冲液可用于进行外切核酸酶消化反应。

[0274] 在一些实施例中,试剂盒包括一个或多个容器,其含有本文所述的单链夹板链中的任一个(200)。该试剂盒可进一步包括一个或多个容器,其含有T4多核苷酸激酶、至少一种连接酶和/或至少一种核酸外切酶。该试剂盒可以包括任意组合中的这些组分的任一者,并且可以包括在单个容器中,或者可以包含在单独的容器中,或者它们的任意组合。

[0275] 该试剂盒可以包括使用该试剂盒进行反应来使线性核酸文库分子成环以形成多个文库-夹板复合物(300)的说明书,每个文库-夹板复合物都具有缺口。该试剂盒可以包括使用该试剂盒通过连接文库-夹板复合物(300)中的缺口来产生共价闭合环状文库分子(400)的说明书。

[0276] 包括用于文库制备的衔接子和有尾PCR引物的试剂盒

[0277] 本公开提供了一种文库制备试剂盒,用于通过将片段化的靶核酸(例如,插入区域(110))的两端连接至Y形衔接子以产生多个衔接子-插入物-衔接子构建体来制备多个线性文库分子,并且通过使用第一有尾PCR引物和第二有尾PCR引物进行引物延伸或PCR反应,将额外的衔接子序列附加至衔接子-插入物-衔接子构建体(例如,图17)。可以使用本文所述的任何单链夹板分子(例如,图3和图4)使所得线性文库分子成环以形成共价闭合环状分子。

[0278] 在一些实施例中,文库制备试剂盒包括Y形衔接子。Y形衔接子包含第一核酸链和第二核酸链,其中两条链的一部分彼此完全互补并退火在一起,并且两条链的另一部分彼

此不互补且错配。在一些实施例中, Y形衔接子的连接端包含形成平端或悬垂端(例如, 5'或3'悬垂端)的退火部分。在一些实施例中, 在文库制备试剂盒中, Y形衔接子的第一链包含具有序列5'-TGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:211)的寡核苷酸。在一些实施例中, 在文库制备试剂盒中, Y形衔接子的第二链包含具有以下序列的寡核苷酸: 5'-CGTGCTGGATTGGCTCACCAGACACCTTCC GACAT-3'(SEQ ID NO:212)

[0279] 在一些实施例中, 文库制备试剂盒包含第一有尾PCR引物, 该第一有尾PCR引物包括序列5'-GATCAGGTGAGGCTGCGACGACTNNNNNNNNNNN AGTTGACAAGCGGTAGCCTGCACACCTTCCGACAT-3'(SEQ ID NO:213), 其中12碱基N序列代表右样品索引序列(170)。在一些实施例中, 12碱基N序列中的前九个'N'包括右样品索引的通用样品索引序列。在一些实施例中, 12碱基N序列中的最后三个"NNN"(下划线)包含3聚体随机序列, 其被设计为在每个测序循环中提供核苷酸多样性和颜色平衡。

[0280] 在一些实施例中, 文库制备试剂盒包含第二有尾PCR引物, 该第二有尾PCR引物包括序列5'-CATGTAATGCACGTACTTTTCAGGGTNNNNNNNNNN

[0281] CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:214), 其中10碱基N序列代表左样品索引序列(160)。在一些实施例中, 10碱基N序列包括左样品索引的通用样品索引序列。

[0282] 形成多个文库-夹板复合物的方法

[0283] 本公开提供了用于形成多个文库-夹板复合物(300)的方法, 包括: (a) 提供多个单链核酸文库分子(100), 其中多个单链核酸文库分子中的各个文库分子包括目的序列(110), 该目的序列在一侧上侧接至少第一左通用衔接子序列(120), 并且在另一侧上侧接至少第一右通用衔接子序列(130)(例如, 参见图1-图5)。

[0284] 用于形成多个文库-夹板复合体(300)的方法进一步包括步骤(b): 提供多个单链夹板链(200), 其中多个单链夹板链中的各个单链夹板链(200)包括第一区域(210)和第二区域(220), 该第一区域能够与各个文库分子的至少第一左通用衔接子序列(120)杂交, 该第二区域能够与各个文库分子的至少第一右通用衔接子序列(130)杂交。示例性单链夹板链(200)示出于图1-图5中。在一些实施例中, 单链夹板链(200)的长度可为20-150个核苷酸、或长度为60-100个核苷酸、长度为70-90个核苷酸或长度为60-80个核苷酸。

[0285] 形成多个文库-夹板复合物(300)的方法进一步包括步骤(c): 使多个单链夹板链(200)与多个单链核酸文库分子(100)杂交。杂交是在适用于使各个文库分子与各个单链夹板链杂交的条件下进行的, 使得单链夹板链(210)的一个的第一区域退火至文库分子的至少第一左通用衔接子序列(120), 并且使得单链夹板链(220)的第二区域退火至文库分子的至少第一右通用序列(130), 由此使各个文库分子成环以形成多个库-夹板复合物(300)。在一些实施例中, 文库-夹板复合物(300)包括文库分子的末端5'与3'端之间的缺口(例如, 图1-图5)。在一些实施例中, 缺口是可酶促连接的。

[0286] 在一些实施例中, 在形成多个文库-夹板复合物(300)的方法中, 单链夹板链(210)的第一区域包括第一通用衔接子序列, 该第一通用衔接子序列可与线性核酸文库分子(120)一端的第一通用结合序列杂交。在一些实施例中, 单链夹板链的第一区域(210)包括第一通用衔接子序列, 其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结

合序列,或针对压实寡核苷酸的通用结合序列。

[0287] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,单链夹板链的第二区域(220)包含第二通用衔接子序列,其可以在线性核酸文库分子的另一端与第二通用结合序列杂交。在一些实施例中,单链夹板链的第二区域(220)包括第二通用衔接子序列,其包括针对正向测序引物或反向测序引物的通用结合序列、针对第一表面引物或第二表面引物的通用结合序列、针对正向扩增引物或反向扩增引物的通用结合序列,或针对压实寡核苷酸的通用结合序列。

[0288] 在一些实施例中,在形成多个文库-夹板复合物(300)的方法中,单链夹板链(200)在其5'和/或3'端包含一个或多个硫代磷酸酯键以赋予核酸外切酶抗性。在一些实施例中,单链夹板链(200)在内部位置包含一个或多个硫代磷酸酯键以赋予核酸内切酶抗性。在一些实施例中,单链夹板链(200)在其5'和/或3'端或在内部位置包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,单链夹板链(200)的5'端为磷酸化或非磷酸化的。在一些实施例中,单链夹板链(200)的3'端包含末端3'OH基团或末端3'阻断基团。

[0289] 在一些实施例中,在形成多个文库-夹板复合物(300)的方法中,单链夹板链(210)的第一区域可以与双链核酸文库分子的有义链或反义链杂交。在文库-夹板复合物(300)中,单链夹板链(220)的第二区域可以与双链核酸文库分子的有义链或反义链杂交。双链核酸文库分子可以变性以产生单链的有义和反义文库链。可以使双链核酸文库分子变性以产生步骤(a)的单链核酸文库分子(100)。

[0290] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,单链夹板链的第一区域和第二区域(分别为210和220)不与目的序列(110)杂交。

[0291] 在一些实施例中,在形成多个文库-夹板复合物(300)的方法中,核酸文库分子(100)进一步包含第二左通用衔接子序列(140)。在一些实施例中,核酸文库分子(100)进一步包含第二右通用衔接子序列(150)。在一些实施例中,核酸文库分子(100)可进一步包含额外的左通用衔接子序列和/或右通用衔接子序列。

[0292] 在一些实施例中,在形成多个文库-夹板复合物(300)的方法中,核酸文库分子(100)进一步包含第一左索引序列(160)和/或第一右索引序列(170)。在一些实施例中,第一左索引序列(160)包括样品索引序列。第一左索引序列(160)的长度可以是3-20个核苷酸。在一些实施例中,第一右索引序列(170)包括另一个样品索引序列。第一右索引序列(170)的长度可以是3-20个核苷酸。左样品索引序列和右样品索引序列(例如,(160)和(170))的序列可以彼此相同或不同。样品索引序列可用于区分多重测定中从不同样品源获得的目的序列。示例性第一左索引序列(160)和第一右索引序列(170)的列表在图9A-图9F的表1中提供。第一左索引序列(160)可以包括短随机序列(例如,NNN)或缺少短随机序列。第一右索引序列(170)可以包括短随机序列(例如,NNN)或缺少短随机序列。短随机序列(例如,NNN)的长度可以是3-20个核苷酸。

[0293] 通过使用一个或两个索引序列(例如,左索引序列和/或右索引序列)准备具有样品索引的文库来启动多重工作流程。可使用第一左索引序列(160)和/或第一右索引序列(170),使用从不同来源分离的输入核酸,来制备单独的具有样品索引的文库。可以将具有样品索引的文库汇集在一起以产生多重文库混合物,并且可以对汇集的文库进行成环、扩增和/或测序。插入区域的序列连同第一左索引序列(160)和/或第一右索引序列(170)可用

于识别输入核酸的来源。在一些实施例中,可以将任意数量的具有样品索引的文库汇集在一起,例如可以汇集2-10、或10-50、或50-100、或100-200、或多于200个具有样品索引的文库。示例性的核酸来源包括天然存在的、重组的或化学合成的来源。示例性的核酸来源包括单个细胞、多个细胞、组织、生物流体、环境样品或整个生物体。示例性的核酸来源包括新鲜的、冷冻的、新鲜冷冻的或存档的来源(例如,福尔马林固定石蜡包埋的;FFPE)。技术人员将认识到核酸可以从许多其他来源分离。核酸文库分子可以以单链或双链形式制备。

[0294] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,核酸文库分子(100)进一步包含:任选的第一左唯一识别序列(180)和/或任选的第一右唯一识别序列(190)。在一些实施例中,第一左唯一识别序列(180)和第一右唯一识别序列(190)各自包含一段序列,该段序列用于在其他目的分子序列的群体中,唯一地识别其上附加有唯一衔接子的各个目的序列(例如,插入序列)。在一些实施例中,第一左唯一识别序列(180)和/或第一右唯一识别序列(190)可用于分子标记。在一些实施例中,唯一识别序列包含2-12个或更多个具有已知序列的核苷酸。例如,唯一识别序列包括已知的随机序列,其中每个位置处的核苷酸随机选自具有碱基A、G、C、T或U的核苷酸。唯一识别序列(180)和/或(190)可以是用于分子标记程序。

[0295] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,核酸文库分子(100)包含以下任意一者或者两者或更多者以任何顺序的任意组合:第一左通用衔接子序列(120);第二左通用衔接子序列(140);第一左索引序列(160);第一左唯一识别序列(180);第一右通用衔接子序列(130);第二右通用衔接子序列(150);第一右索引序列(170);和/或第一右唯一识别序列(190)。

[0296] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,第一左通用衔接子序列(120)和/或第二左通用衔接子序列(140)包括:针对正向测序引物或反向测序引物的通用结合序列;针对第一表面引物或第二表面引物的通用结合序列;针对正向扩增引物或反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。在一些实施例中,核酸文库分子(100)可进一步包含额外的左通用衔接子序列。

[0297] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,第一右通用衔接子序列(130)和/或第二右通用衔接子序列(150)包括:针对正向测序引物或反向测序引物的通用结合序列;针对第一表面引物或第二表面引物的通用结合序列;针对正向扩增引物或反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。在一些实施例中,核酸文库分子(100)可进一步包含额外的右通用衔接子序列。

[0298] 在一些实施例中,在用于形成多个文库-夹板复合物(300)的方法中,核酸文库分子(100)进一步包含位于本文所述的任何通用衔接子序列之间的至少一个接合衔接子序列(例如,参见图5)。例如,第一左接合衔接子序列(125)可位于第一左通用衔接子序列(120)与第一左索引序列(160)之间。第二左接合衔接子序列(165)可位于第一左索引序列(160)与第二左通用衔接子序列(140)之间。第三接合衔接子序列(145)可位于第二左通用衔接子序列(140)与目的序列(110)之间。第一右接合衔接子序列(135)可位于第一右通用序列(130)与第一右索引序列(170)之间。第二右接合衔接子序列(175)可位于第一右索引序列(170)与第二右通用衔接子序列(150)之间。第三右接合衔接子序列(155)可位于第二右通用衔接子序列(150)与目的序列(110)之间。任何接合衔接子序列包含任何序列并且长度可

为3-60个核苷酸。任何接合衔接子序列包含通用序列或唯一序列。任何接合衔接子序列包括具有3-20个核苷酸的随机序列(例如,NNN)。任何接合衔接子序列包含扩增引物、测序引物或压实寡核苷酸的结合序列。任何接合衔接子序列包含固定的表面引物(例如,捕获引物)的结合序列。任何接合衔接子序列均包含样品索引序列。任何接合衔接子序列均包含唯一识别序列。任何接合衔接子序列,特别是接合衔接子序列(145)包括Tn5转座子端序列5'-AGATGTGTATAAGAGACAG-3'(SEQ ID NO:209)。任何接合衔接子序列,特别是接合衔接子序列(155)包括Tn5转座子端序列5'-CTGTCTCTTATACACATCT-3'(SEQ ID NO:210)。可以在适合形成转座子突触复合物的条件下,经由转座酶介导的反应将Tn5转座子端序列引入文库分子(100),该反应包括使双链输入DNA(例如,基因组DNA)与Tn-5型转座酶和双链寡核苷酸接触,该双链寡核苷酸包含与通用衔接子序列或样品索引序列链接的Tn转座子端序列(SEQ ID NO:209)。在双链寡核苷酸中,Tn转座子端序列(SEQ ID NO:209)可位于相对于通用衔接子序列或样品索引序列的5'或3'。

[0299] 本公开提供了用于形成多个文库-夹板复合物(300)的方法,其包括:(a)提供多个单链夹板链(200),其中各个单链夹板链(200)包括以5'至3'顺序排列的区域:(i)具有通用结合序列的第一区域(210),其与线性单链文库分子(例如120)一端上的序列杂交,以及(ii)具有通用结合序列的第二区域(220),其与线性单链文库分子(例如,130)另一端上的序列杂交。用于形成多个文库-夹板复合物(300)的方法进一步包括步骤(b):使多个单链夹板链(200)与多个单链核酸文库分子(100)杂交,其中各个文库分子包含以5'至3'顺序排列的区:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列;(ii)第二左通用衔接子序列(140),其具有针对第一测序引物或第二测序引物的结合序列;(iii)目的序列(110);(iv)第二右通用衔接子序列(150),其具有针对第二测序引物或第一测序引物的结合序列;以及(v)具有第一表面引物的结合序列的第一右通用衔接子序列(130),其中杂交在适合于单链夹板链(200)与文库分子(100)杂交的条件下进行,由此使文库分子成环以产生文库-夹板复合物(300),使得单链夹板链的第一区域(210)与针对第二表面引物(120)的结合序列杂交,并且使得单链夹板链的第二区域(220)与针对第一表面引物(130)的结合序列杂交,其中文库-夹板复合物(300)包含在文库分子的末端5'端与3'端之间的缺口,并且其中缺口是可酶促连接的(例如,参见图2)。在一些实施例中,多个单链核酸文库分子(100)进一步包含第一左索引序列(160)和/或第一右索引序列(170)(例如,参见图3和图4)。在一些实施例中,在多个给定文库-夹板复合物(300)中,第一左索引(160)和第一右索引(170)的序列彼此相同或不同。第一左索引序列(160)的长度可以是3-20个核苷酸。第一右索引序列(170)的长度可以是3-20个核苷酸。第一左索引序列(160)和/或第一右索引序列(170)可以包括短随机序列(例如,NNN)。短随机序列的长度可以是3-20个核苷酸。示例性第一左索引序列(160)和第一右索引序列(170)的列表在图9A-图9F的表1中提供。在一些实施例中,多个单链核酸文库分子(100)进一步包含可用于分子标记的第一左唯一识别序列(180)和/或第一右唯一识别序列(190)(例如,参见图3和图4)。

[0300] 在一些实施例中,在该方法中,文库分子中的第一左通用衔接子序列(120)包括序列5'-CATGTAATGCACGTACTTTCAGGGT-3'(SEQ ID NO:199)。

[0301] 在一些实施例中,在该方法中,文库分子中的第一左通用衔接子序列(120)包括序列5'-AATGATACG GCGACCACCGA-3'(SEQ ID NO:203)。

[0302] 在一些实施例中,在该方法中,文库分子中的第二左通用衔接子序列(140)包括序列5'-CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:201)。

[0303] 在一些实施例中,在该方法中,文库分子中的第二左通用衔接子序列(140)包括序列5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'(SEQ ID NO:205)。

[0304] 在一些实施例中,在该方法中,文库分子中的第二左通用衔接子序列(140)包括序列5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'(SEQ ID NO:206)。

[0305] 在一些实施例中,在该方法中,文库分子中的第二右通用衔接子序列(150)包括序列5'-ATGTCGGAAGGTGTGCAGGCTACCGCTTGCTCAACT-3'(SEQ ID NO:202)。

[0306] 在一些实施例中,在该方法中,文库分子中的第二右通用衔接子序列(150)包括序列5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'(SEQ ID NO:207)。

[0307] 在一些实施例中,在该方法中,文库分子中的第二右通用衔接子序列(150)包括序列5'-CTGTCTCTTATACACATCTCCGAGCCCACGAGAC-3'(SEQ ID NO:208)。

[0308] 在一些实施例中,在该方法中,文库分子中的第一右通用衔接子序列(130)包括序列5'-AGTCGTCGCAGCCTCACCTGATC-3'(SEQ ID NO:200)。

[0309] 在一些实施例中,在该方法中,文库分子中的第一右通用衔接子序列(130)包括序列5'-TCGTATGCCGTCTTCTGCTTG-3'(SEQ ID NO:204)。

[0310] 在一些实施例中,在该方法中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列5'-ACCCTGAAAGTACGTGCATTACATG-3'(SEQ ID NO:193)。

[0311] 在一些实施例中,在该方法中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)。

[0312] 在一些实施例中,在该方法中,单链夹板链(200)包括序列5'-ACCCTGAAAGTACGTGCATTACATGGATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:195)。例如参见图7A-图7C。

[0313] 在一些实施例中,在该方法中,单链夹板链(210)的第一区域包括文库分子的第一左通用衔接子序列(120)的通用结合序列,其中第一区域(210)包括以下序列5'-TCGGTGGTCGCCGTATCATT-3'(SEQ ID NO:196)。

[0314] 在一些实施例中,在该方法中,单链夹板链(220)的第二区域包括文库分子的第一右通用衔接子序列(130)的通用结合序列,其中第二区域(220)包括序列5'-CAAGCAGAAGACGGCATAACGA-3'(SEQ ID NO:197)。

[0315] 在一些实施例中,在该方法中,单链夹板链(200)包括序列5'-TCGGTGGTCGCCGTATCATTCAAGCAGAAGACGGCATAACGA-3'(SEQ ID NO:198)。

[0316] 在一些实施例中,用于形成本文所述的多个文库-夹板复合物(300)的任何方法可进一步包括至少一种酶促反应,包括磷酸化反应、连接反应和/或核酸外切酶反应。酶促反应可以顺序进行或基本上同时进行。酶促反应可以在单个反应容器中进行。替代性地,可以在第一反应容器中进行第一酶促反应,然后转移至进行第二酶促反应的第二反应容器,然后转移至进行第三酶促反应的第三反应容器。

[0317] 在一些实施例中,用于形成本文所述的多个文库-夹板复合物(300)的任何方法进一步包括进行单独且连续的磷酸化和连接反应,该磷酸化和连接反应在单独的反应容器中

进行。在一些实施例中,用于形成多个文库-夹板复合物(300)的方法进一步包括步骤(c1):在适合于使多个单链夹板链(200)和/或多个单链核酸文库分子(100)的5'端磷酸化的条件下,在第一反应容器中,使多个单链夹板链(200)和多个单链核酸文库分子(100)与T4多核苷酸激酶接触;并将磷酸化反应转移至第二反应容器。在一些实施例中,用于形成多个文库-夹板复合物(300)的方法进一步包括步骤(d1):在第二反应容器中,在适用于酶促连接缺口的条件下,使多个磷酸化的单链夹板链(200)和多个磷酸化单链核酸文库分子(100)与连接酶接触,由此产生各自与单链夹板链(200)杂交的多个共价闭环状文库分子(400)。在一些实施例中,连接酶包含T7 DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。

[0318] 在一些实施例中,用于形成本文所述的多个文库-夹板复合物(300)的任何方法进一步包括进行连续的磷酸化和连接反应,该反应在同一反应容器中连续进行。在一些实施例中,用于形成多个文库-夹板复合物(300)的方法进一步包括步骤(c2):在适合于使多个单链夹板链(200)和多个单链核酸文库分子(100)的5'端磷酸化的条件下,在第一反应容器中,使多个单链夹板链(200)和多个单链核酸文库分子(100)与T4多核苷酸激酶接触。在一些实施例中,用于形成多个文库-夹板复合物(300)的方法进一步包括步骤(d2):在同一第一反应容器中,在适用于酶促连接缺口的条件下,使磷酸化的单链夹板链(200)和磷酸化的单链核酸文库分子(100)与连接酶接触,由此产生各自与单链夹板链(200)杂交的多个共价闭环状文库分子(400)。在一些实施例中,连接酶包含T7 DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。

[0319] 在一些实施例中,用于形成本文所述的多个文库-夹板复合物(300)的任何方法进一步包括进行基本上同时的磷酸化和连接反应,该反应在同一反应容器中一起进行。在一些实施例中,用于形成多个文库-夹板复合物(300)的方法进一步包括步骤(c3):在适合使多个单链夹板链(200)和多个单链核酸文库分子(100)的5'端磷酸化的条件下,并且该条件适用于酶促连接缺口,在第一反应容器中,使多个单链夹板链(200)和多个单链核酸文库分子(100)与(i) T4多核苷酸激酶和(ii)连接酶接触,由此产生各自与单链夹板链(200)杂交的多个共价闭环状文库分子(400)。在一些实施例中,连接酶包含T7 DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。

[0320] 在一些实施例中,用于形成本文所述的多个文库-夹板复合物(300)的任何方法进一步包括以下任选的步骤:从多个共价闭环状文库分子(400)酶促移除多个单链夹板链(200),其包括步骤:使多个共价闭环状文库分子(400)与至少一种核酸外切酶接触,以移除多个单链夹板链(200)并保留多个共价闭环状文库分子(400)。在一些实施例中,核酸外切酶反应可以在用于进行磷酸化和/或连接反应的相同反应缓冲液中进行,或者在不同的反应缓冲液中进行。在一些实施例中,可以在第一反应容器中进行磷酸化反应(步骤c1,参见上文)并在第二反应容器中进行连接反应(步骤d1,参见上文)之后,在第三反应容器中进行核酸外切酶反应。在一些实施例中,可以在第一反应容器中进行磷酸化反应(步骤c2,参见上文)并在第一反应容器中进行顺序的连接反应(步骤d2,参见上文)之后,在第一反应容器中进行核酸外切酶反应。在一些实施例中,可以在第一反应容器中进行基本上同时的磷酸化和连接反应(步骤c3,参见上文)之后,在第一反应容器中进行核酸外切酶反应。在一些实施例中,至少一种核酸外切酶包含核酸外切酶I、不耐热核酸外切酶I和/或T7核酸外切酶中的两种或更多种的任意组合。

[0321] 滚环扩增方法

[0322] 本公开提供了对共价闭合环状文库分子(400)进行滚环扩增反应的方法。滚环扩增反应可以在磷酸化和连接反应之后进行,或者在连接反应之后进行。在一些实施例中,滚环扩增反应可以在与单链夹板链(200)杂交的共价闭合环状文库分子(400)上进行。在一些实施例中,例如滚环扩增反应可在核酸外切酶反应后进行,此时共价闭合环状文库分子(400)不再与单链夹板链(200)杂交。在一些实施例中,共价闭合环状文库分子(400)可以分配在支持物上,然后经历滚环扩增反应。在一些实施例中,共价闭合环状文库分子(400)可以在溶液中经历滚环扩增反应,然后分配到支持物上。在一些实施例中,滚环扩增反应可以采用保留的单链夹板链(200)作为扩增引物,或者可以移除单链夹板链(200)(例如,经由核酸外切酶消化)并用可溶性扩增引物替换。

[0323] 支持物上滚环扩增

[0324] 在一些实施例中,对缺少杂交的单链夹板链(200)的多个共价闭合环状文库分子进行滚环扩增反应的方法,并且其中多个共价闭合环状文库分子中的各个共价闭合环状文库分子(400)包含第一表面引物的通用结合序列,该方法包括步骤(a):在适用于使各个共价闭合环状文库分子(400)与各个固定的第一表面引物杂交的条件下,将多个共价闭合环状文库分子(400)分配到支持物上,该支持物上固定有多个第一表面引物,由此固定多个共价闭合环状文库分子(400)。

[0325] 在一些实施例中,在进行滚环扩增反应的方法中,固定和支持物上的多个第一表面引物包括序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)。

[0326] 在一些实施例中,在进行滚环扩增反应的方法中,固定和支持物上的多个第一表面引物包括序列5'-CAAGCAGAAGACGGCATACGA-3'(SEQ ID NO:197)。

[0327] 单个第一表面引物可以与具有第一表面引物的通用结合序列的共价闭合环状文库分子(400)杂交。

[0328] 在一些实施例中,进行滚环扩增反应的方法进一步包括步骤(b):在合适于使用多个第一表面引物作为固定的扩增引物,并使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个固定的共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,由此产生多个固定于第一表面引物的核酸多联体分子。在一些实施例中,多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP中的两种或更多种的任意组合。在一些实施例中,各个固定的多联体共价接合至各个第一表面引物。在一些实施例中,多个共价闭合环状文库分子中的各个共价闭合环状文库分子(400)包含第一表面引物和第二表面引物(例如,分别为(130)和(120))的通用结合序列,使得滚环扩增反应产生具有针对第一表面引物和第二表面引物的通用结合序列的多个串联拷贝的多联体分子。在一些实施例中,支持物进一步包含多个第二表面引物。在一些实施例中,固定的第二表面引物用于将多联体分子的至少一部分固定至支持物。在一些实施例中,固定的第二表面引物具有不可延伸的3'端并且不能用于扩增。在一些实施例中,可以对固定的多联体经历测序反应。

[0329] 在一些实施例中,固定和支持物上的多个第二表面引物包括序列5'-CATGTAATGCACGTACTTTTCAGGGT-3'(SEQ ID NO:199或其互补序列)。

[0330] 在一些实施例中,固定和支持物上的多个第二表面引物包括序列5'-

AATGATACGGCGACCACCGA-3' (SEQ ID NO:203或其互补序列)。

[0331] 单个第二表面引物可以与具有第二表面引物的通用结合序列的多联体分子的一部分杂交。在一些实施例中,固定的第二表面引物用于将多联体分子的至少一部分固定至支持物。在一些实施例中,固定的第二表面引物具有不可延伸的3'端并且不能用于扩增。在一些实施例中,可以对固定的多联体经历测序反应。

[0332] 在一些实施例中,在进行滚环扩增反应的方法中,多个共价闭合环状文库分子(400)可以分配到一种支持物上,该支持物涂覆有一种或更多种化合物,以在支持物上产生钝化层(例如,图18)。在一些实施例中,钝化层形成多孔或半多孔层。在一些实施例中,一种或多种类型的表面引物、多联体模板分子和/或聚合酶可以附接至钝化层,以固定至支持物。在一些实施例中,支持物包含低非特异性结合表面,其能够改善支持物上的核酸杂交和扩增性能。一般而言,支持物可包含一层或更多层共价或非共价附接的低结合化学修饰层,例如硅烷层、聚合物膜,以及可使用的一种或更多种共价或非共价附接的寡核苷酸将多个核酸多联体分子固定至支持物。在一些实施例中,支持物可包含官能化聚合物涂层,其经由支持物上的化学基团、接枝至官能化聚合物涂层的引物、以及引物和官能化的聚合物涂层上的水溶性保护层,共价结合至支持物的至少一部分。在一些实施例中,官能化聚合物涂层包含聚(N-(5-叠氨基乙酰胺基戊基)丙烯酰胺-共-丙烯酰胺(PAZAM)。在一些实施例中,支持物包含表面涂层,该表面涂层具有至少一层亲水性聚合物涂层和至少一层的多个寡核苷酸。亲水性聚合物涂层可包含聚乙二醇(PEG)。亲水性聚合物涂层可包含具有至少4个支链的支化PEG。在一些实施例中,低非特异性结合涂层具有可以水接触角测量的亲水性程度,其中水接触角不超过45度。在一些实施例中,固定至支持物或固定至支持物上的涂层的共价闭合环状文库分子(400)的密度为约 10^2 - 10^6 个/ mm^2 ,或约 10^6 - 10^9 个/ mm^2 ,或约 10^9 - 10^{12} 个/ mm^2 。在一些实施例中,多个共价闭合环状文库分子(400)在支持物(或支持物上的涂层)上的预定位点处,固定至支持物或固定至支持物上的涂层;或固定至涂层在支持物(或支持物上的涂层)上的随机位点处。

[0333] 在一些实施例中,在进行滚环扩增反应的方法中,步骤(a)的分配(例如,将多个共价闭合环状文库分子(400)分配到支持物上)可以在高效杂交缓冲液的存在下进行,该高效杂交缓冲液包含:(i)第一极性非质子溶剂,其具有不大于40的介电常数并且具有4-9的极性指数;(ii)第二极性非质子溶剂,其具有不大于115的介电常数并且以有效使双链核酸变性的量存在于杂交缓冲液制剂中;(iii)pH缓冲系统,其将杂交缓冲液制剂的pH维持在约4-8的范围内;以及(iv)足以增强或促进分子拥挤的量的拥挤剂。在一些实施例中,高效杂交缓冲液包含:(i)第一极性非质子溶剂,其包含占杂交缓冲液体积25-50%的乙腈;(ii)第二极性非质子溶剂,其包含占杂交缓冲液体积5-10%的甲酰胺;(iii)pH缓冲系统,其包含2-(N-吗啉代)乙磺酸(MES),pH为5-6.5;以及(iv)拥挤剂,其包含占杂交缓冲液体积5-35%的聚乙二醇(PEG)。在一些实施例中,高效率杂交缓冲液进一步包含甜菜碱。使用可溶性扩增引物进行溶液内滚环扩增

[0334] 在一些实施例中,对缺少杂交的单链夹板链(200)的多个共价闭合环状文库分子(400)进行滚环扩增反应的方法,其中多个共价闭合环状文库分子中的各个共价闭合环状文库分子(400)包括针对正向扩增引物的通用结合序列和针对第一表面引物的通用结合序列,该方法包括:(a)在溶液中杂交多个共价闭合环状文库分子和多个可溶性正向扩增引

物;以及(b)在适合于使用多个正向扩增引物,并使用多个共价闭合环状文库分子(400)作为模板分子,在溶液中进行滚环扩增反应的条件下,通过使多个共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,来进行第一滚环扩增反应,由此产生多个具有仍与其共价闭合环状文库分子(400)杂交的部分的核酸多联体分子。在一些实施例中,用于进行滚环扩增反应的方法进一步包括步骤(c):在适用于使多联体分子的至少一部分与多个固定的第一表面引物杂交的条件下,将多个多联体分子分配到其上固定有多个第一表面引物的支持物上,由此固定多个多联体分子。多个固定的多联体分子仍与其共价闭合环状文库分子(400)杂交。在一些实施例中,用于进行滚环扩增反应的方法进一步包括步骤(d):在适合于使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行第二滚环扩增反应的条件下,使固定的多个多联体分子与多个链置换聚合酶和多个核苷酸接触,由此延伸多个固定的核酸多联体分子。在一些实施例中,第一滚环扩增反应和/或第二滚环扩增反应可以用包括dATP、dGTP、dCTP、dTTP和/或dUTP中的两个或更多个的任意组合的多个核苷酸进行。在一些实施例中,各个固定的多联体与各个第一表面引物杂交。在一些实施例中,多共价闭合环状文库分子个中的各个共价闭合环状文库分子(400)包含第一表面引物和第二表面引物(例如,分别为(130)和(120))的通用结合序列,使得溶液内滚环扩增反应产生具有针对第一表面引物和第二表面引物的通用结合序列的多个串联拷贝的多联体分子。在一些实施例中,支持物进一步包含多个第二表面引物。在一些实施例中,固定的第二表面引物用于将多联体分子的至少一部分固定至支持物。在一些实施例中,固定的第二表面引物具有不可延伸的3'端并且不能用于扩增。在一些实施例中,可以对固定的多联体经历测序反应。

[0335] 在一些实施例中,在进行滚环扩增反应的方法中,固定在支持物上的多个第一表面引物包括序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)。单个第一表面引物可以与具有第一表面引物的通用结合序列的共价闭合环状文库分子(400)杂交。

[0336] 在一些实施例中,在进行滚环扩增反应的方法中,固定在支持物上的多个第一表面引物包括序列5'-CAAGCAGAAGACGGCATACGA-3'(SEQ ID NO:197)。单个第一表面引物可以与具有第一表面引物的通用结合序列的共价闭合环状文库分子(400)杂交。

[0337] 在一些实施例中,固定在支持物上的多个第二表面引物包括序列5'-CATGTAATGCACGTA CTTTCAGGGT-3'(SEQ ID NO:199或其互补序列)。单个第二表面引物可以与具有第二表面引物的通用结合序列的多联体分子的一部分杂交。

[0338] 在一些实施例中,固定在支持物上的多个第二表面引物包括序列5'-AATGATACGGCGACCACCGA-3'(SEQ ID NO:203或其互补序列)。单个第二表面引物可以与具有第二表面引物的通用结合序列的多联体分子的一部分杂交。

[0339] 在一些实施例中,固定的第二表面引物用于将多联体分子的至少一部分固定至支持物。在一些实施例中,固定的第二表面引物具有不可延伸的3'端并且不能用于扩增。在一些实施例中,可以对固定的多联体经历测序反应。

[0340] 在一些实施例中,可将步骤(c)的多个多联体分子可以分配到支持物上,该支持物上涂覆有一种或更多种化合物,以在支持物上产生钝化层(例如,图18)。在一些实施例中,钝化层形成多孔或半多孔层。在一些实施例中,一种或多种类型的表面引物、多联体模板分子和/或聚合酶可以附接至钝化层,以固定至支持物。在一些实施例中,支持物包含低非特

异性结合表面,其能够改善支持物上的核酸杂交和扩增性能。一般而言,支持物可包含一层或更多层共价或非共价附接的低结合化学修饰层,例如硅烷层、聚合物膜,以及可使用的一种或更多种共价或非共价附接的寡核苷酸将多个核酸多联体分子固定至支持物。在一些实施例中,支持物可包含官能化聚合物涂层,其经由支持物上的化学基团、接枝至官能化聚合物涂层的引物、以及引物和官能化的聚合物涂层上的水溶性保护层,共价结合至支持物的至少一部分。在一些实施例中,官能化聚合物涂层包含聚(N-(5-叠氮基乙酰胺基戊基)丙烯酰胺-共-丙烯酰胺(PAZAM)。在一些实施例中,支持物包含表面涂层,该表面涂层具有至少一层亲水性聚合物涂层和至少一层的多个寡核苷酸。亲水性聚合物涂层可包含聚乙二醇(PEG)。亲水性聚合物涂层可包含具有至少4个支链的支化PEG。在一些实施例中,低非特异性结合涂层具有可以水接触角测量的亲水性程度,其中水接触角不超过45度。在一些实施例中,固定至支持物或固定至支持物上的涂层的多联体分子的密度为约 10^2 - $10^6/\text{mm}^2$ 、或约 10^6 - $10^9/\text{mm}^2$ 、或约 10^9 - $10^{12}/\text{mm}^2$ 。在一些实施例中,多个多联体分子在支持物(或支持物上的涂层)上的预定位点,固定至支持物或固定至支持物上的涂层;或在支持物(或支持物上的涂层)上的随机位点,固定至支持物上的涂层。

[0341] 在一些实施例中,步骤(c)的分配可以在高效杂交缓冲液存在下进行,所述高效杂交缓冲液包含:(i)第一极性非质子溶剂,其具有不大于40的介电常数并且具有4-9的极性指数;(ii)第二极性非质子溶剂,其具有不大于115的介电常数并且以有效使双链核酸变性的量存在于杂交缓冲液制剂中;(iii)pH缓冲系统,其将杂交缓冲液制剂的pH维持在约4-8的范围内;以及(iv)足以增强或促进分子拥挤的量的拥挤剂。在一些实施例中,高效杂交缓冲液包含:(i)第一极性非质子溶剂,其包含占杂交缓冲液体积25-50%的乙腈;(ii)第二极性非质子溶剂,其包含占杂交缓冲液体积5-10%的甲酰胺;(iii)pH缓冲系统,其包含2-(N-吗啉代)乙磺酸(MES),pH为5-6.5;以及(iv)拥挤剂,其包含占杂交缓冲液体积5-35%的聚乙二醇(PEG)。在一些实施例中,高效率杂交缓冲液进一步包含甜菜碱。使用单链夹板链进行溶液内滚环扩增

[0342] 在一些实施例中,用于对与单链夹板链(200)杂交的多个共价闭合环状文库分子进行滚环扩增反应的方法,其中多个共价闭合环状文库分子中的各个共价闭合环状文库分子(400)包含针对第一表面引物的通用结合序列,该方法包括(a):在适用于将单链夹板链(200)作为扩增引物,进行第一滚环扩增反应的条件下,在溶液中使与单链夹板链(200)杂交的多个共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,由此产生仍与其共价闭合环状文库分子(400)杂交的多个多联体分子(例如,参见图6A-图6C)。

[0343] 在一些实施例中,用于进行滚环扩增反应的方法进一步包括步骤(b):在适用于使多联体的至少一部分与多个固定的第一表面引物杂交的条件下,将与其共价闭合环状文库分子(400)杂交的多个多联体分子分配到其上固定有多个第一表面引物的支持物上,由此固定多个多联体分子。多个固定的多联体分子仍与其共价闭合环状文库分子(400)杂交。

[0344] 在一些实施例中,用于进行滚环扩增反应的方法进一步包括步骤(c):在适合于使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行第二滚环扩增反应的条件下,使多个固定的多联体分子与多个链置换聚合酶和多个核苷酸接触,由此延伸多个固定的核酸多联体分子。

[0345] 在一些实施例中,第一滚环扩增反应和/或第二滚环扩增反应可以用包括dATP、

dGTP、dCTP、dTTP和/或dUTP中的两个或更多的任意组合的多个核苷酸进行。在一些实施例中,各个固定的多联体与各个第一表面引物杂交。在一些实施例中,多共价闭环文库分子个中的各个共价闭环文库分子(400)包含第一表面引物和第二表面引物(例如,分别为(130)和(120))的通用结合序列,使得溶液内滚环扩增反应产生具有针对第一表面引物和第二表面引物的通用结合序列的多个串联拷贝的多联体分子。在一些实施例中,支持物进一步包含多个第二表面引物。在一些实施例中,固定的第二表面引物用于将多联体分子的至少一部分固定至支持物。在一些实施例中,固定的第二表面引物具有不可延伸的3'端并且不能用于扩增。在一些实施例中,可以对固定的多联体经历测序反应。

[0346] 在一些实施例中,在进行滚环扩增反应的方法中,固定在支持物上的多个第一表面引物包括序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:194)。单个第一表面引物可以与具有第一表面引物的通用结合序列的共价闭环文库分子(400)杂交。

[0347] 在一些实施例中,在进行滚环扩增反应的方法中,固定在支持物上的多个第一表面引物包括序列5'-CAAGCAGAAGACGGCATACTGA-3'(SEQ ID NO:197)。单个第一表面引物可以与具有第一表面引物的通用结合序列的共价闭环文库分子(400)杂交。

[0348] 在一些实施例中,固定在支持物上的多个第二表面引物包括序列5'-CATGTAATGCACGTACTTTCAGGGT-3'(SEQ ID NO:199或其互补序列)。单个第二表面引物可以与具有第二表面引物的通用结合序列的多联体分子的一部分杂交。

[0349] 在一些实施例中,固定在支持物上的多个第二表面引物包括序列5'-AATGATACGGCGACCACTG-3'(SEQ ID NO:203或其互补序列)。单个第二表面引物可以与具有第二表面引物的通用结合序列的多联体分子的一部分杂交。

[0350] 在一些实施例中,固定的第二表面引物用于将多联体分子的至少一部分固定至支持物。在一些实施例中,固定的第二表面引物具有不可延伸的3'端并且不能用于扩增。在一些实施例中,可以对固定的多联体经历测序反应。

[0351] 在一些实施例中,可以将步骤(b)的多个多联体分子分配到支持物上,该支持物涂覆有一种或更多种化合物,以在支持物上产生钝化层(例如,图18)。在一些实施例中,钝化层形成多孔或半多孔层。在一些实施例中,表面引物、多联体模板分子和/或聚合酶可以附接至钝化层,以固定至支持物。在一些实施例中,支持物包含低非特异性结合表面,其能够改善支持物上的核酸杂交和扩增性能。一般而言,支持物可包含一层或更多层共价或非共价附接的低结合化学修饰层,例如硅烷层、聚合物膜,以及可使用的一种或更多种共价或非共价附接的寡核苷酸将多个核酸多联体分子固定至支持物。在一些实施例中,支持物可包含官能化聚合物涂层,其经由支持物上的化学基团、接枝至官能化聚合物涂层的引物、以及引物和官能化的聚合物涂层上的水溶性保护涂层,共价结合至支持物的至少一部分。在一些实施例中,官能化聚合物涂层包含聚(N-(5-叠氮基乙酰胺基戊基)丙烯酰胺-共-丙烯酰胺(PAZAM))。在一些实施例中,支持物包含表面涂层,该表面涂层具有至少一层亲水性聚合物涂层和至少一层的多个寡核苷酸。亲水性聚合物涂层可包含聚乙二醇(PEG)。亲水性聚合物涂层可包含具有至少4个支链的支化PEG。在一些实施例中,低非特异性结合涂层具有可以水接触角测量的亲水性程度,其中水接触角不超过45度。在一些实施例中,固定至支持物或固定至支持物上的涂层的多联体分子的密度为约 10^2 - 10^6 /mm²、或约 10^6 - 10^9 /mm²、或约 10^9 - 10^{12} /mm²。在一些实施例中,多个多联体分子在支持物(或支持物上的涂层)上的预定位

点,固定至支持物或固定至支持物上的涂层;或在支持物(或支持物上的涂层)上的随机位点,固定至支持物上的涂层。

[0352] 在一些实施例中,步骤(b)的分配可以在高效杂交缓冲液存在下进行,该高效杂交缓冲液包含:(i)第一极性非质子溶剂,其具有不大于40的介电常数并且具有4-9的极性指数;(ii)第二极性非质子溶剂,其具有不大于115的介电常数并且以有效使双链核酸变性的量存在于杂交缓冲液制剂中;(iii)pH缓冲系统,其将杂交缓冲液制剂的pH维持在约4-8的范围内;以及(iv)足以增强或促进分子拥挤的量的拥挤剂。在一些实施例中,高效杂交缓冲液包含:(i)第一极性非质子溶剂,其包含占杂交缓冲液体积25-50%的乙腈;(ii)第二极性非质子溶剂,其包含占杂交缓冲液体积5-10%的甲酰胺;(iii)pH缓冲系统,其包含2-(N-吗啉代)乙磺酸(MES),pH为5-6.5;以及(iv)拥挤剂,其包含占杂交缓冲液体积5-35%的聚乙二醇(PEG)。在一些实施例中,高效率杂交缓冲液进一步包含甜菜碱。

[0353] 用于测序的方法

[0354] 本公开提供了对本文所述的任何固定的多联体分子进行测序的方法。本文描述的进行滚环扩增反应的任何方法都可以用于产生固定于支持物的多个多联体分子,并且可以对固定的多联体分子经历测序反应。在一些实施例中,测序反应采用可检测地标记的核苷酸类似物。在一些实施例中,测序反应采用两阶段测序反应,包括结合可检测地标记的多价分子,并掺入核苷酸类似物。在一些实施例中,测序反应采用未标记的核苷酸类似物。术语多联体分子和模板分子可互换使用。

[0355] 在一些实施例中,本文描述的任何滚环扩增反应(例如,在支持物上或在溶液中进行的RCA)可用于生成固定的多联体,每个多联体含有目的序列和存在于共价闭合环状文库分子(400)中的任何衔接子序列的串联重复单元。例如,串联重复单元包含:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列,(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列,(iii)目的序列(110),(iv)第二右通用衔接子序列(150),其具有针对第二测序引物的结合序列,(v)第一右通用衔接子序列(130),其具有针对第一表面引物的结合序列,和(vii)第一左索引序列(160)和/或第一右索引序列(170)(例如参见图3-图4)。在一些实施例中,串联重复单元进一步包含第一左唯一识别序列(180)和/或第一右唯一识别序列(190)。

[0356] 固定的多联体可以自行塌陷成紧凑的核酸纳米球。在RCA反应期间包括一个或多个压实寡核苷酸可进一步压实纳米球的尺寸和/或形状。给定多联体中串联重复单元数量的增加,增加了多联体周围用于与多个测序引物(例如,具有通用序列的测序引物)杂交的位点数量的位点的数量,该测序引物用作聚合酶催化的测序反应的多个起始位点。当测序反应采用经可检测地标记的核苷酸和/或经可检测地标记的多价分子(例如,具有核苷酸单元)时,由参与沿多联体的并行测序反应的核苷酸或核苷酸单元发出的信号产生对于每个多联体而言增加的信号强度。给定多联体的多个部分可被同时测序。此外,多个结合复合物可沿特定多联体分子形成,每个结合复合物包含结合至多价分子的测序聚合酶,其中该多个结合复合物保持稳定而不离解,导致增加的存留时间,这增加了信号强度并且减少了成像时间。

[0357] 用于使用核苷酸类似物进行测序的方法

[0358] 本公开提供了用于对本文所述的任何固定的多联体分子进行测序的方法,该方法

包括步骤(a):使测序聚合酶与(i)核酸多联体分子和(ii)核酸测序引物接触,其中接触在适合于将测序聚合酶结合至杂交至核酸引物的核酸多联体分子的条件下进行,其中杂交至核苷酸引物的核酸多联体分子形成核酸双链体。在一些实施例中,测序聚合酶包括可以结合并掺入核苷酸类似物的重组突变体测序聚合酶。在一些实施例中,测序引物包含3'可延伸端。

[0359] 在一些实施例中,用于对多联体分子进行测序的方法中,测序引物包含3'可延伸端或3'不可延伸端。在一些实施例中,多个核酸多联体分子包含扩增的模板分子(例如,克隆扩增的模板分子)。在一些实施例中,多个核酸多联体分子包含靶标目的序列的一个拷贝。在一些实施例中,多个核酸分子包含靶标目的序列的两个或更多个串联拷贝(例如,多联体)。在一些实施例中,多个核酸多联体分子中的核酸多联体分子包含相同的靶标目的序列或不同的靶标目的序列。在一些实施例中,多个核酸多联体分子和/或多个核酸引物为在溶液中的或固定至支持物的。在一些实施例中,当多个核酸多联体分子和/或多个核酸引物固定至支持物时,与第一测序聚合酶的结合生成多个固定的第一复合聚合酶。在一些实施例中,多个核酸多联体分子和/或核酸引物被固定至支持物上的 10^2 - 10^{15} 个不同位点。在一些实施例中,多个多联体分子和核酸引物与多个第一测序聚合酶的结合生成固定于支持物上的 10^2 - 10^{15} 个不同位点的多个第一复合聚合酶。在一些实施例中,支持物上的多个固定的第一复合聚合酶被固定于支持物上的预定或随机位点。在一些实施例中,多个固定的第一复合聚合酶彼此流体连通以允许试剂(例如,包括测序聚合酶的酶、多价分子、核苷酸和/或二价阳离子)的溶液流动到支持物上,使得支持物上的多个固定的复合聚合酶以大规模并行方式与试剂溶液反应。

[0360] 在一些实施例中,用于测序的方法进一步包括步骤(b):在适用于将至少一个核苷酸与测序聚合酶(该测序聚合酶结合至核酸双链体)结合的条件下,并且适用于核苷酸的聚合酶催化掺入的条件下,使测序聚合酶与多个核苷酸接触。在一些实施例中,在至少一种包含镁和/或锰的催化阳离子存在下,使测序聚合酶与多个核苷酸接触。在一些实施例中,多个核苷酸包括至少一个在糖2'或3'位置处具有链终止部分的核苷酸类似物。在一些实施例中,链终止部分从糖2'或3'位置可移除,以将链终止部分转化为OH或H基团。在一些实施例中,多个核苷酸包括至少一个缺失链终止部分的核苷酸。在一些实施例中,至少一个核苷酸用可检测报告部分(例如荧光团)标记。

[0361] 在一些实施例中,用于测序的方法进一步包括步骤(c):在适用于掺入至少一种核苷酸的条件下,将至少一种核苷酸掺入到可延伸引物的3'端中。在一些实施例中,适用于核苷酸结合聚合酶和掺入核苷酸的条件可以相同或不同。在一些实施例中,适用于掺入核苷酸的条件包含加入至少一种包含镁和/或锰的催化阳离子。在一些实施例中,至少一种核苷酸结合测序聚合酶并掺入可延伸引物的3'端。在一些实施例中,步骤(c)中将核苷酸掺入引物的3'端中包括引物延伸反应。

[0362] 在一些实施例中,用于测序的方法进一步包括步骤(d):将步骤(b)和(c)中将至少一个核苷酸掺入可延伸引物的3'端中重复至少一次。在一些实施例中,多个核苷酸包括用可检测报告部分标记的多个核苷酸。可检测报告基因部分包括荧光团。在一些实施例中,荧光团附接至核苷酸碱基。在一些实施例中,荧光团用接头附接至核苷酸碱基,该接头可从碱基裂解/移除。在一些实施例中,多个核苷酸中的至少一个未用可检测报告基因部分标记。

在一些实施例中, 附接至核苷酸的特定可检测报道基因部分 (例如, 荧光团) 可对应于核苷酸碱基 (例如, dATP、dGTP、dCTP、dTTP 或 dUTP) 以允许对核苷酸碱基的检测和识别。在一些实施例中, 该方法进一步包括: 在步骤 (c) 和/或 (d) 处检测该至少一个掺入的核苷酸。在一些实施例中, 该方法进一步包括: 在步骤 (c) 和/或 (d) 处识别该至少一个掺入的核苷酸。在一些实施例中, 可以通过检测和识别结合测序聚合酶的核苷酸来确定核酸多联体分子的序列, 由此确定多联体分子的序列。在一些实施例中, 核酸多联体分子的序列可以通过检测和识别掺入引物 3' 端的核苷酸来确定, 由此确定多联体分子的序列。

[0363] 在一些实施例中, 在用于测序的方法中, 结合至核酸双链体的多个测序聚合酶包含多个复合聚合酶, 其具有至少第一复合聚合酶和第二复合聚合酶, 其中 (a) 第一复合聚合酶包含结合至第一核酸双链体的第一测序聚合酶, 其中该第一核酸双链体包含与第一核酸引物杂交的第一核酸模板序列, (b) 第二复合聚合酶包含结合至第二核酸双链体的第二测序聚合酶, 其中该第二核酸双链体包含与第二核酸引物杂交的第二核酸模板序列, (c) 第一核酸模板序列和第二核酸模板序列包含相同或不同的序列, (d) 第一核酸多联体和第二核酸多联体被克隆扩增, (e) 第一引物和第二引物包含可延伸的 3' 端或不可延伸的 3' 端, 并且 (f) 多个复合聚合酶固定于支持物。在一些实施例中, 固定于支持物的多个复合聚合酶的密度为约 10^2 - 10^{15} 个复合聚合酶/ mm^2 。

[0364] 用于对核酸进行测序的两阶段方法

[0365] 本公开提供了用于对本文所述的任何固定的多联体分子进行测序的两阶段方法。在一些实施例中, 第一阶段通常包括将多价分子与复合聚合酶结合以形成多价-复合聚合酶, 并检测多价-复合聚合酶。

[0366] 在一些实施例中, 第一阶段包括步骤 (a): 使多个第一测序聚合酶与 (i) 多个核酸多联体分子和 (ii) 多个核酸引物接触, 其中该接触在适用于将多个第一测序聚合酶与多个核酸多联体分子和多个核酸引物结合的条件下进行, 由此形成多个第一复合聚合酶, 每个第一复合聚合酶包含结合至核酸双链体的第一测序聚合酶, 其中核酸双链体包含与核酸引物杂交的核酸多联体分子。在一些实施例中, 第一聚合酶包含重组突变体测序聚合酶。在一些实施例中, 测序引物包含 3' 可延伸端。

[0367] 在一些实施例中, 用于对多联体分子进行测序的方法中, 测序引物包含 3' 可延伸端或 3' 不可延伸端。在一些实施例中, 多个核酸多联体分子包含扩增的模板分子 (例如, 克隆扩增的模板分子)。在一些实施例中, 多个核酸多联体分子包含靶标目的序列的一个拷贝。在一些实施例中, 多个核酸分子包含靶标目的序列的两个或更多个串联拷贝 (例如, 多联体)。在一些实施例中, 多个核酸多联体分子中的核酸多联体分子包含相同的靶标目的序列或不同的靶标目的序列。在一些实施例中, 多个核酸多联体分子和/或多个核酸引物为在溶液中的或固定至支持物的。在一些实施例中, 当多个核酸多联体分子和/或多个核酸引物固定至支持物时, 与第一测序聚合酶的结合生成多个固定的第一复合聚合酶。在一些实施例中, 多个核酸多联体分子和/或核酸引物被固定至支持物上的 10^2 - 10^{15} 个不同位点。在一些实施例中, 多个多联体分子和核酸引物与多个第一测序聚合酶的结合生成固定于支持物上的 10^2 - 10^{15} 个不同位点的多个第一复合聚合酶。在一些实施例中, 支持物上的多个固定的第一复合聚合酶被固定于支持物上的预定或随机位点。在一些实施例中, 多个固定的第一复合聚合酶彼此流体连通以允许试剂 (例如, 包括测序聚合酶的酶、多价分子、核苷酸和/或

二价阳离子)的溶液流动到支持物上,使得支持物上的多个固定的复合聚合酶以大规模并行方式与试剂溶液反应。

[0368] 在一些实施例中,用于测序的方法进一步包括步骤(b):使多个第一复合聚合酶与多个多价分子接触以形成多个多价-复合聚合酶(例如,结合复合物)。在一些实施例中,多个多价分子中的各个多价分子包含衔接至多个核苷酸臂的核心,并且每个核苷酸臂衔接至核苷酸(例如,核苷酸单元)(例如,图19-图23)。在一些实施例中,在适用于将多价分子的互补核苷酸单元与该多个第一复合聚合酶中的至少两个结合,由此形成多个多价复合聚合酶的条件下进行步骤(b)的接触。在一些实施例中,该条件适用于抑制互补核苷酸单元的聚合酶催化的掺入多个多价-复合聚合酶的引物中。在一些实施例中,该多个多价分子包含至少一个具有多个核苷酸臂(例如,图19-图23)的多价分子,每个核苷酸臂衔接有核苷酸类似物(例如,核苷酸类似物单元),其中核苷酸类似物包括位于糖2'和/或3'位置的链终止部分。在一些实施例中,该多个多价分子包括包含多个核苷酸臂的至少一个多价分子,每个核苷酸臂衔接有缺失链终止部分的核苷酸单元。在一些实施例中,该多个多价分子中的至少一个多价分子用可检测报道基因部分标记。在一些实施例中,可检测报道基因部分包含荧光团。在一些实施例中,在至少一种包含镱、钡和/或钙的非催化阳离子存在下,进行步骤(b)的接触。

[0369] 在一些实施例中,用于测序的方法进一步包括步骤(c):检测多个多价-复合聚合酶。在一些实施例中,检测包括检测结合至复合聚合酶的多价分子,其中多价分子的互补核苷酸单元结合至引物,但互补核苷酸单元的掺入被抑制。在一些实施例中,多价分子用可检测报道基因部分标记以允许检测。在一些实施例中,标记的多价分子包含衔接至多价分子的核心、接头和/或核苷酸单元的荧光团。

[0370] 在一些实施例中,测序方法进一步包括步骤(d):识别结合至多个第一复合聚合酶的互补核苷酸单元的核碱基,由此确定多联体分子的序列。在一些实施例中,多价分子用对应于衔接至核苷酸臂的特定核苷酸单元的可检测报道基因部分进行标记,以允许识别结合至多个第一复合聚合酶的互补核苷酸单元(例如,核苷酸碱基腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶)。

[0371] 在一些实施例中,两阶段测序方法的第二阶段通常包括核苷酸掺入。在一些实施例中,用于测序的方法进一步包括步骤(e):解离多个多价-复合聚合酶并移除多个第一测序聚合酶及其结合的多价分子,并保留多个核酸双链体。

[0372] 在一些实施例中,用于测序的方法进一步包括步骤(f):使步骤(e)的多个保留的核酸双链体与多个第二测序聚合酶接触,其中该接触在适用于将多个第二测序聚合酶与多个保留的核酸双链体结合的条件下进行,由此形成各自包含结合至保留的核酸双链体的第二测序聚合酶的多个第二复合聚合酶。在一些实施例中,第二测序聚合酶包含重组突变体测序聚合酶。

[0373] 在一些实施例中,步骤(a)的多个第一测序聚合酶具有与步骤(f)的多个第二测序聚合酶的氨基酸序列100%相同的氨基酸序列。在一些实施例中,步骤(a)的多个第一测序聚合酶具有与步骤(f)的多个第二测序聚合酶的氨基酸序列不同的氨基酸序列。

[0374] 在一些实施例中,用于测序的方法进一步包括步骤(g):使多个第二复合聚合酶与多个核苷酸接触,其中该接触在适用于将来自多个核苷酸的互补核苷酸与复合聚合酶的至

少两个结合的条件下进行,由此形成多个核苷酸-复合聚合酶。在一些实施例中,步骤(g)的接触在适用于促进结合的互补核苷酸经聚合酶催化掺入核苷酸复合-聚合酶的引物中条件下进行,由此形成多个核苷酸-复合聚合酶。在一些实施例中,步骤(g)中将核苷酸掺入引物的3'端中包括引物延伸反应。在一些实施例中,步骤(g)的接触在至少一种包含镁和/或锰的催化阳离子存在下进行。在一些实施例中,该多个核苷酸包括天然核苷酸(例如,非类似物核苷酸)或核苷酸类似物。在一些实施例中,该多个核苷酸包含可移除或不可移除2'和/或3'链终止部分。在一些实施例中,该多个核苷酸包括多个未标记的或用可检测的报道部分标记的核苷酸。可检测报道基因部分包括荧光团。在一些实施例中,荧光团附接至核苷酸碱基。在一些实施例中,荧光团用接头附接至核苷酸碱基,该接头可从碱基裂解/移除或不可从碱基移除。在一些实施例中,多个核苷酸中的至少一个未用可检测报道基因部分标记。在一些实施例中,附接至核苷酸的特定可检测报道基因部分(例如,荧光团)可对应于核苷酸碱基(例如,dATP、dGTP、dCTP、dTTP或dUTP)以允许对核苷酸碱基的检测和识别。

[0375] 在一些实施例中,用于测序的方法进一步包括步骤(h):检测掺入到核苷酸-复合聚合酶的引物中的互补核苷酸。在一些实施例中,该多个核苷酸用可检测报道部分标记以允许检测。在一些实施例中,在用于对多联体分子进行测序的方法中,省略步骤(h)的检测。

[0376] 在一些实施例中,用于测序的方法进一步包括步骤(i):识别掺入核苷酸-复合聚合酶的引物中的互补核苷酸的碱基。在一些实施例中,步骤(i)中识别掺入的互补核苷酸可用于确认结合至步骤(d)中的多个第一复合聚合酶的多价分子的互补核苷酸的身份。在一些实施例中,步骤(i)的识别可用于确定核酸多联体分子的序列。在一些实施例中,在用于对多联体分子进行测序的方法中,省略了步骤(i)的识别。

[0377] 在一些实施例中,用于测序的方法进一步包括步骤(j):当通过使多个第二复合聚合酶与包含至少一个具有2'和/或3'链终止部分的核苷酸的多个核苷酸接触来进行步骤(g)时,从掺入的核苷酸中移除链终止部分。

[0378] 在一些实施例中,用于测序的方法进一步包括步骤(k):重复步骤(a)-(g)和(j)至少一次。在一些实施例中,用于测序的方法进一步包括步骤(k):重复步骤(a)-(j)至少一次。在一些实施例中,核酸多联体分子的序列可以通过在步骤(c)和(d)中检测和识别结合测序聚合酶但不掺入引物的3'端的多价分子来确定。在一些实施例中,核酸多联体分子的序列可以通过检测和识别在步骤(h)和(i)中掺入引物3'端的核苷酸来确定(或确认)。

[0379] 在一些实施例中,在用于对核酸分子进行测序的任何方法中,使多个第一复合聚合酶与多个多价分子结合,形成至少一个亲和力复合物,该方法包括以下步骤:(a)使第一核酸引物、第一测序聚合酶和第一多价分子与多联体模板分子的第一部分结合,由此形成第一结合复合物,其中第一多价分子的第一核苷酸单元与第一测序聚合酶结合;以及(b)将第二核酸引物、第二测序聚合酶和第一多价分子与同一多联体模板分子的第二部分结合,由此形成第二结合复合物,其中第一多价分子的第二核苷酸单元与第二测序聚合酶结合,其中包括同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物。在一些实施例中,第一测序聚合酶包含本文所述的任何野生型或突变型聚合酶。在一些实施例中,第二测序聚合酶包含本文所述的任何野生型或突变型聚合酶。多联体模板分子包含目的序列和至少一个通用测序引物结合位点的串联重复序列。第一核酸引物和第二核酸引物可沿多联体模板分子与测序引物结合位点结合。示例性多价分子示出于图19-图23中。

[0380] 在一些实施例中,在用于对核酸分子进行测序的任何方法中,其中该方法包括将多个第一复合聚合酶与多个多价分子结合,以形成至少一个亲和力复合物,该方法包括以下步骤:(a)使多个测序聚合酶和多个核酸引物与多联体核酸多联体分子的不同部分接触,以在同一多联体模板分子上形成至少第一和第二复合聚合酶;(b)在适用于将来自多个多价分子的单个多价分子与第一复合聚合酶和第二复合聚合酶结合的条件下,使多个可检测地标记的多价分子与同一多联体模板分子上的至少第一复合聚合酶和第二复合聚合酶接触,其中单个多价分子的至少第一核苷酸单元结合至第一复合聚合酶,该第一复合聚合酶包括与多联体模板分子的第一部分杂交的第一引物,由此形成第一结合复合物(例如,第一三元复合物),并且其中单个多价分子的至少第二核苷酸单元结合至第二复合聚合酶,该第二复合聚合酶包括与多联体模板分子的第二部分杂交的第二引物,由此形成第二结合复合物(例如,第二三元复合物),其中接触在适用于抑制结合的第一核苷酸单元和第二核苷酸单元经聚合酶催化掺入第一结合复合物和第二结合复合物中的条件下进行,并且其中与同一多价分子结合的第一结合复合物和第二结合复合物形成亲和力复合物;以及(c)检测同一多联体模板分子上的第一结合复合物和第二结合复合物,以及(d)识别第一结合复合物中的第一核苷酸单元,由此确定多联体模板分子的第一部分的序列,并且识别第二结合复合物中的第二核苷酸单元,由此确定多联体模板分子的第二部分的序列。在一些实施例中,多个测序聚合酶包含本文所述的任何野生型或突变型测序聚合酶。多联体模板分子包含目的序列和至少一个通用测序引物结合位点的串联重复序列。多个核酸引物可沿多联体模板分子与测序引物结合位点结合。示例性多价分子示出于图19-图23中。

[0381] 结合测序

[0382] 本公开提供了用于对本文所述的任何固定的多联体分子进行测序的方法,其中测序方法包括采用未标记的链终止核苷酸的结合测序(SBB)程序。在一些实施例中,结合测序(SBB)方法包括步骤(a)在使三元复合物稳定的条件下,依次使引发的模板核酸与至少两种单独的混合物接触,其中所述至少两种单独的混合物各自包括聚合酶和核苷酸,由此顺序接触导致引发的模板核酸在三元复合物稳定条件下与模板中第一、第二和第三碱基类型同源的碱基类型的核苷酸接触;(b)检查至少两种单独的混合物以确定是否形成三元复合物;以及(c)识别引发的模板核酸分子的下一个正确核苷酸,其中如果在步骤(b)中检测到三元复合物,则下一个正确核苷酸被识别为第一碱基类型、第二碱基类型或第三碱基类型的同源物,并且其中如果步骤(b)中不存在三元复合物,将下一个正确的核苷酸推算为第四碱基类型的核苷酸同源物;(d)在步骤(b)之后将下一个正确的核苷酸添加到引发的模板核酸的引物中,由此产生延伸的引物;以及(e)对包含延伸引物的引发的模板核酸重复步骤(a)至(d)至少一次。示例性的结合测序方法描述于美国专利第10,246,744号和第10,731,141号(其中这两个专利的内容通过引用整体并入本文中)。

[0383] 多重工作流程

[0384] 本公开提供了多重工作流程,其通常包括使用单链夹板链(200)和携带一个或两个索引序列(例如,左(160)和/或右(170)样品索引的序列)的具有样品索引的核酸文库制备单独的具有样品索引的共价闭合环状文库分子的群体。第一左索引序列(160)和第一右索引序列(170)是已知序列。可以从不同来源分离的输入核酸制备单独的具有样品索引的文库,其中样品索引序列用于区分不同来源。汇集步骤可以在产生具有样品索引的共价闭

合环状文库分子之后或产生具有样品索引的文库-夹板复合物之前进行。汇集的分子可以经历下游多重扩增和/或多重测序反应。

[0385] 用于多重分析的传统汇集工作流程

[0386] 在一些实施例中,一种用于制备从多个样品源分离的目的序列的多重混合物的方法,其包括:(a)提供单链核酸文库分子(100)的两个或更多个群体,文库分子(100)的每个群体包含在单独区室中,其中给定群体中的核酸文库分子包含:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列,(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列,(iii)从样品源分离的目的序列(110),(iv)第二右通用衔接子序列(150),其具有针对第二测序引物的结合序列,(v)第一右通用衔接子序列(130),其具有针对第一表面引物的结合序列的,和(vii)第一左索引序列(160)和/或第一右索引序列(170),其中左(160)和/或右(170)索引单独地或组合地识别从中分离出目的序列(110)的样品源(例如,参见图3和图4)。在一些实施例中,单链核酸文库分子(100)进一步包含第一左唯一识别序列(180)和/或第一右唯一识别序列(190)。

[0387] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(b):提供多个单链夹板链(200),其中各个单链夹板链(200)包括第一区域(210)和第二区域(220),其中第一区域(210)可以与单链文库分子的第一左通用衔接子序列(120)杂交,并且其中第二区域(220)可以与单链文库分子的第一右通用衔接子序列(130)杂交。

[0388] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(c):在单独的区室中,使单链核酸文库分子(100)的群体与配额的多个单链夹板链(200)接触,其中该接触是在适用于使单链夹板链的第一区域(210)与单链文库分子的第一左通用衔接子序列(120)杂交的条件下进行,并且该条件适合于使单链夹板链的第二区域(220)与单链文库分子的第一右通用衔接子序列(130)杂交,由此使文库分子成环以产生在文库分子的5'端与文库分子的3'端之间具有缺口的文库-夹板复合物的群体(300)(例如,参见图1-图5)。在一些实施例中,缺口是可酶促连接的。

[0389] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(d):在单独的区室中,在适用于酶促连接缺口的条件下,使文库-夹板复合物(300)的群体与连接酶接触,由此产生各自与单链夹板链(200)杂交的共价闭合环状文库分子(400)群体(例如,参见图6A-图6C)。

[0390] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(e):将来自单独区室的共价闭合环状文库分子(400)的群体汇集在一起,以生成共价闭合环状文库分子(400)的多重混合物,其包含从多个样品源分离的目的序列的多重混合物。

[0391] 在一些实施例中,目的序列可以从两个或更多个不同的样品源(例如,2-10个、或10-50个、或50-100个、或100-250个、或多于250个不同的样品源)分离。示例性的核酸来源包括天然存在的、重组的或化学合成的来源。示例性的核酸来源包括单个细胞、多个细胞、组织、生物流体、环境样品或整个生物体。示例性的核酸来源包括新鲜的、冷冻的、新鲜冷冻的或存档的来源(例如,福尔马林固定石蜡包埋的;FFPE)。技术人员将认识到核酸可以从许多其他来源分离。给定群体中的目的序列具有相同或不同的序列。

[0392] 在一些实施例中,步骤(a)的单链核酸文库分子(100)的群体的数量可以是2-10个、或10-50个、或50-100个、或100-250个、或多于250个不同的单链核酸文库分子(100)群体。在一些实施例中,任何数量的不同群体的共价闭合环状文库分子(400)可以在步骤(e)中汇集在一起,例如2-10个、或10-50个、或50-100个、或100-200个、或超过200个不同群体的共价闭合环状文库分子(400)可以汇集在一起。

[0393] 技术人员将认识到,步骤(a)中可以使用任何数量的单独的区室(例如,2-10个、或10-50个、或50-100个、或100-250个、或更多个单独的区室)(例如多孔板,诸如例如96孔板)。

[0394] 在一些实施例中,与步骤(d)或(e)的共价闭合环状文库分子(400)杂交的单链夹板链(200)的3'端包含可延伸的3'OH端,其可以充当引物延伸反应(例如滚环扩增反应)的起始点。

[0395] 在一些实施例中,在步骤(d)或(e)中,与单链夹板链(200)杂交的共价闭合环状文库分子(400)群体可以任选地与至少一种核酸外切酶反应,以移除多个单链夹板链(200)并保留多个共价闭合的环状文库分子(400)。在一些实施例中,至少一种核酸外切酶包含核酸外切酶I、不耐热核酸外切酶I和/或T7核酸外切酶。

[0396] 在一些实施例中,步骤(a)的单链核酸文库分子(100)进一步包含以下任意一者或者两者或更多者的任意组合:针对正向扩增引物的通用结合序列;针对反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。

[0397] 用于多重分析的非传统汇集工作流程

[0398] 在一些实施例中,一种用于制备从多个样品源分离的目的序列的多重混合物的方法,其包括:(a)提供单链核酸文库分子(100)的两个或更多个群体,每个文库分子(100)群体,每个单链核酸文库分子群体包含在单独的区室中,其中给定群体的核酸文库分子包含:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列,(ii)第二左通用衔接子序列(140),其具有针对第一测序引物的结合序列,(iii)从样品源分离的目的序列(110),(iv)第二右通用衔接子序列(150),其具有针对第二测序引物的结合序列,(v)第一右通用衔接子序列(130),其具有针对第一表面引物的结合序列,和(vii)第一左索引序列(160)和/或第一右索引序列(170),其中左索引(160)和/或右(170)索引序列单独地或组合地识别从中分离出目的序列(110)的样品源;(例如,参见图3和图4)。在一些实施例中,单链核酸文库分子(100)进一步包含第一左唯一识别序列(180)和/或第一右唯一识别序列(190)。

[0399] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(b):将单链核酸文库分子(100)的两个或更多个群体汇集到同一个区室。

[0400] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(c):向步骤(b)的隔室添加多个单链夹板链(200),其中各个单链夹板链(200)包括第一区域(210)和第二区域(220),其中第一区域(210)可以与单链文库分子的第一左通用衔接子序列(120)杂交,并且其中第二区域(220)可以与单链文库分子的第一右通用衔接子序列(130)杂交。

[0401] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(d):孵育单链核酸文库分子(100)的群体和多个单链夹板链(200),在适合于

将单链夹板链的第一区域(210)与单链文库分子的第一左通用衔接子序列(120)杂交的条件下,并且该条件适合于将单链夹板链的第二区域(220)与单链文库分子的第一右通用衔接子序列(130)杂交,由此使文库分子成环以产生文库-夹板复合物(300)的群体,其具有文库分子的5'端与文库分子的3'端之间的缺口(例如,参见图1-图5)。在一些实施例中,缺口是可酶促连接的。

[0402] 在一些实施例中,用于制备从多个样品源分离的目的序列的多重混合物的方法进一步包括步骤(e):在步骤(d)的区室中,在适用于酶促连接缺口的条件下,使文库-夹板复合物(300)群体与连接酶接触,由此产生各自与单链夹板链(200)杂交的共价闭环状文库分子(400)的群体,由此产生共价闭环状文库分子(400)的多重混合物,其包含从多个样品源分离的目的序列的多重混合物(例如,参见图6A-图6C)。

[0403] 在一些实施例中,目的序列可以从两个或更多个不同的样品源(例如,2-10个、或10-50个、或50-100个、或100-250个、或多于250个不同的样品源)分离。示例性的核酸来源包括天然存在的、重组的或化学合成的来源。示例性的核酸来源包括单个细胞、多个细胞、组织、生物流体、环境样品或整个生物体。示例性的核酸来源包括新鲜的、冷冻的、新鲜冷冻的或存档的来源(例如,福尔马林固定石蜡包埋的;FFPE)。技术人员将认识到核酸可以从许多其他来源分离。给定群体中的目的序列具有相同或不同的序列。

[0404] 在一些实施例中,步骤(a)的单链核酸文库分子(100)的群体的数量可以是2-10个、或10-50个、或50-100个、或100-250个、或多于250个不同的单链核酸文库分子(100)群体。在一些实施例中,任何数量的不同群体的单链核酸文库分子可以在步骤(b)中汇集在一起,例如2-10个、或10-50个、或50-100个、或100-200个、或超过200个不同群体的共价闭环状文库分子(400)可以汇集在一起。

[0405] 技术人员将认识到,步骤(a)中可以使用任何数量的单独的区室(例如,2-10个、或10-50个、或50-100个、或100-250个、或更多个单独的区室)(例如多孔板,诸如例如96孔板)。

[0406] 在一些实施例中,与步骤(e)的共价闭环状文库分子(400)杂交的单链夹板链(200)的3'端包含可延伸的3'OH端,其可以充当引物延伸反应(例如滚环扩增反应)的起始点。

[0407] 在一些实施例中,在步骤(e)中,与单链夹板链(200)杂交的共价闭环状文库分子(400)群体可以任选地与至少一种核酸外切酶反应,以移除多个单链夹板链(200)并保留多个共价闭环的环状文库分子(400)。在一些实施例中,至少一种核酸外切酶包含核酸外切酶I、不耐热核酸外切酶I和/或T7核酸外切酶。

[0408] 在一些实施例中,步骤(a)的单链核酸文库分子(100)进一步包含以下任意一者或者两者或更多者的任意组合:针对正向扩增引物的通用结合序列;针对反向扩增引物的通用结合序列;和/或针对压实寡核苷酸的通用结合序列。

[0409] 支持物上多重滚环扩增反应

[0410] 在一些实施例中,本文所述的共价闭环状文库分子(400)的任何多价混合物可以分配到支持物上,其中该方法包括:在适用于使各个共价闭环状文库分子(400)与各个固定的第一表面引物杂交的条件下,将多个共价闭环状文库分子(400)分配到其上固定有多个第一表面引物的支持物上,由此固定多个共价闭环状文库分子(400);以及在适合

于使用多个第一表面引物作为固定的扩增引物,并使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个固定的共价闭合环状文库分子(400)与多个链置换聚合酶和多个核苷酸接触,由此产生多个固定的核酸多联体分子,其中多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。在一些实施例中,共价闭合环状文库分子的群体仍与单链夹板链(200)杂交并分配到支持物上。在一些实施例中,已在适合于移除多个单链夹板链(200)并保留共价闭合环状文库分子(400)群体的条件下,使用至少一种核酸外切酶,从共价闭合环状文库分子的群体(400)移除单链夹板链(200)。

[0411] 使用可溶性扩增引物并固定在支持物上的溶液内多重滚环扩增反应

[0412] 在一些实施例中,本文所述的共价闭合环状文库分子(400)的任何多价混合物可经历溶液内滚环扩增反应,该方法包括:在适用于移除多个单链夹板链(200)并保留共价闭合环状文库分子(400)群体的条件下,使与单链夹板链(200)杂交的共价闭合环状文库分子(400)群体与至少一种核酸外切酶接触;以及在适合进行滚环扩增反应的条件下,使保留的共价闭合环状文库分子(400)群体与多个可溶性扩增引物、多个链置换聚合酶和多个核苷酸接触,由此产生多个核酸多联体分子,其中多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。在一些实施例中,共价闭合环状文库分子(400)包含针对正向扩增引物的通用结合序列和/或针对反向扩增引物的通用结合序列。在一些实施例中,可溶性扩增引物可以与针对正向扩增引物的通用结合序列或针对反向扩增引物的通用结合序列杂交。在一些实施例中,至少一种核酸外切酶包含核酸外切酶I、不耐热核酸外切酶I和/或T7核酸外切酶。

[0413] 在一些实施例中,可以将多个核酸多联体分子分配到支持物上,该方法包括:在适用于使多联体的至少一部分与多个固定的第一表面引物的至少一部分杂交的条件下,将多个多联体分子分配到其上固定有多个第一表面引物的支持物上,由此固定多个多联体分子;以及在适合于使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行滚环扩增反应的条件下,使固定的多个多联体分子(其与共价闭合环状文库分子(400)杂交)与多个链置换聚合酶和多个核苷酸接触,由此延伸多个固定的核酸多联体分子,其中多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。

[0414] 使用单链夹板链并固定在支持物上的溶液内多重滚环扩增反应

[0415] 在一些实施例中,本文所述的任何共价闭合环状文库分子(400)的多价混合物可经历溶液内滚环扩增反应,该方法包括:在适合进行滚环扩增反应的条件下,通过将单链夹板链(200)杂交的共价闭合环状文库分子(400)群体与多个链置换聚合酶和多个核苷酸接触,来进行滚环扩增反应,由此产生多个多联体分子,其中多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。在一些实施例中,单链夹板链(200)的3'端是可延伸的。

[0416] 在一些实施例中,多个多联体分子可以分配到支持物上,该方法包括:在适用于使多联体的至少一部分与多个固定的第一表面引物的至少一部分杂交的条件下,将与共价闭合环状文库分子(400)杂交的多个多联体分子分配到其上固定有多个第一表面引物的支持物上,由此固定与共价闭合环状文库分子(400)杂交的多个多联体分子;以及在适合于使用多个共价闭合环状文库分子(400)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个固定的多联体分子与多个链置换聚合酶和多个核苷酸接触,由此延伸多个固定的核酸多联体分子,其中多个核苷酸包括dATP、dGTP、dCTP、dTTP和/或dUTP。

[0417] 固定的多联体分子的多重测序

[0418] 本文所述的任何多重固定的多联体分子均可经历测序反应,其一般包括:对多个固定的核酸多联体分子的插入区域(110)和第一左索引序列(160)和/或第一右索引序列(170)进行测序,其中插入区域(110)的序列连同第一左索引序列(160)和/或第一右索引序列(170)可用于识别给定的固定的多联体分子的目的序列的样品源。固定的核酸多联体分子包含插入区域(110)和一个或两个样品索引序列(例如,第一左索引序列(160)和/或第一右索引序列(170))。插入区域(110)、第一左索引序列(160)和/或第一右索引序列(170)可以以任何顺序测序。在一些实施例中,首先可以对插入区域(110)进行测序,第二可以对第一左索引序列(160)进行测序,并且如果第一右索引序列(170)存在于固定的核酸多联体分子中,则第三可以对它进行测序。在一些实施例中,首先可以对第一左索引序列(160)进行测序,第二可以对插入区域(110)进行测序,并且如果第一右索引序列(170)存在于固定的核酸多联体分子中,则第三可以对它进行测序。在一些实施例中,首先可以对第一右索引序列(170)进行测序,如果第一左索引序列(160)存在于固定的核酸多联体分子中,则其次可以对第一左索引序列进行测序,并且可以对插入区域(110)进行测序。

[0419] 在一些实施例中,第一左索引(160)和第一右索引(170)的测序读数可以与它们各自的已知序列比对并指定比对分数,其中比对分数指示该索引的已知索引序列和测序读数之间的相似性。第一左索引(160)和/或第一右索引(170)的比对分数以及相关插入区域(110)的序列读数可用于确定插入区域的样品源。

[0420] 使用核苷酸类似物进行多重测序

[0421] 在一些实施例中,可对本文所述的任何多重固定的多联体分子进行测序,测序方法包括:(a)在适用于将可检测地标记的核苷酸类似物掺入测序引物的3'端中的条件下,将多个固定的核酸多联体分子与多个测序引物、多个测序聚合酶和多个可检测地标记的核苷酸类似物结合,该可检测地标记的核苷酸类似物各自包含2'或3'链终止部分;(b)检测掺入的可检测地标记的核苷酸类似物;以及(c)识别掺入的可检测地标记的核苷酸类似物的核碱基。上文详细描述了使用核苷酸类似物的测序方法。

[0422] 使用两阶段测序方法的多重测序

[0423] 在一些实施例中,本文所述的任何多重固定的多联体分子可以使用两阶段测序方法进行测序,其中第一阶段包括:(a)在适用于将各个多联体分子与测序引物、第一测序聚合酶和可检测地标记的多价分子结合(例如,形成多个多价-复合聚合酶)的条件下,将多个固定的核酸多联体分子与多个测序引物、多个第一测序聚合酶和多个可检测地标记的多价分子结合,其中合适的条件抑制可检测地标记的多价分子的核苷酸单元的聚合酶催化的掺入;(b)检测结合的可检测地标记的多价分子;以及(c)识别可检测地标记的多价分子的核碱基。在一些实施例中,各个可检测地标记的多价分子包含衔接至多个核苷酸臂的核心,并且每个核苷酸臂衔接至核苷酸(例如,核苷酸单元)(例如,参见图19-图23)。在一些实施例中,该多个多价分子中的至少一个多价分子用可检测报道基因部分标记。多价分子的任何部分都可以被标记,包括核心、核苷酸臂或核碱基。在一些实施例中,可检测报道基因部分包含荧光团。在一些实施例中,步骤(a)的结合在至少一种包含镨、钕和/或钙的非催化阳离子存在下进行。在一些实施例中,可检测地标记的多价分子与多个固定的核酸多联体分子的结合形成至少一种如上所述的亲和力复合物。

[0424] 在一些实施例中,两阶段测序方法的第二阶段包括:(d)从多个固定的核酸多联体

分子中移除多个第一测序聚合酶和多个结合的多价分子,并保留各自杂交测序引物的固定的多联体分子(核酸双链体);以及(e)在适用于将核苷酸掺入测序引物的3'端中的条件下,使保留的固定的多联体分子和杂交的测序引物与多个第二聚合酶和多个核苷酸接触。上文详细描述了采用两阶段方法的测序方法。

[0425] 用于改进碱基识别的样品索引

[0426] 通常,需要制备待分配到支持物(例如,涂覆的流通池)上的核酸文库,其中文库分子被转化为模板分子,模板分子以高密度固定到支持物,以用于大规模并行测序。对于以高密度固定在支持物上随机位置的模板分子,在测序运行期间解析高密度荧光图像以实现准确碱基识别的考验变得具有挑战性。

[0427] 固定的模板分子群体的核苷酸多样性是指,每个测序循环中存在的核苷酸A、G、C和T的相对比例。最佳的高多样性文库通常包括目的序列(插入)区域,其具有在测序运行的每个循环中代表的所有四种核苷酸碱基(例如,A、G、C和T/U)的比例大致相等。低多样性文库通常包括具有高比例的某些核苷酸和低比例的其他核苷酸的目的序列(插入)区域。为了克服低多样性文库的问题,通常将从PhiX噬菌体制备的少量高多样性文库与目的文库混合(例如,PhiX内标文库),并在同一流通池上一起测序。虽然PhiX文库内标文库提供了核苷酸多样性,但它也占据了流通池上的空间,由此取代了携带目的序列的靶标文库,并减少了可从靶标文库获得的测序数据量(例如,降低了测序通量)。克服低多样性文库问题的另一种方法是,制备具有至少一个被设计为颜色平衡的样品索引序列的靶标文库分子。然而,可能需要设计大量样品索引集,例如用于16重、24重、96重或更大重级的一组单索引样品序列或成对索引样品序列。对于其中所有样品索引序列都是颜色平衡的大样品索引集(例如,参见图32),将样品索引序列设计为单个或成对样品索引是具有挑战性的。

[0428] 克服低多样性文库分子测序的挑战(例如,在支持物上以高密度)的替换方法是,制备具有至少一个样品索引序列的文库,该样品索引序列包括直接链接至通用样品索引序列的短随机序列(例如,3聚体随机序列,NNN),其中短随机序列提供核苷酸多样性和颜色平衡。图9A-图9F中表1提供了具有短随机序列(例如,NNN)和通用样品索引序列的示例性右索引序列(170)的列表。在具有样品索引的文库分子群体中,样品索引的短随机序列(例如,(170))提供高核苷酸多样性,其包括几乎相等比例的所有四种核苷酸(例如,A、G、C、T和/或U),其将在测序运行的每个循环中表示(参见图31)。短随机序列的高核苷酸多样性也在测序运行的每个循环期间提供颜色平衡。设计样品索引(例如,(160)和/或(170))中包括短随机序列(例如,NNN)的优点是,在文库分子的低重群体(例如,2重或4重)中,识别两个或四个不同样品的通用样品索引序列不需要表现出核苷酸多样性(例如,参见图32)。另外,短随机序列(例如,NNN)的核苷酸多样性可以消除包括PhiX内标文库的需要,或者允许使用减少量的PhiX内标文库来分配到流通池上并测序。

[0429] 目标文库分子可以包括单个样品索引序列(例如,样品索引(170)),其包括短随机序列(例如,NNN)和通用样品索引序列。在一些实施例中,仅来自单个样品索引序列(例如,(170))的测序数据用于聚合酶克隆作图和模板配准,因为短随机序列(例如,NNN)提供了足够的核苷酸多样性和颜色平衡。来自通用样品索引序列的测序数据可用于在多重分析中区分从不同样品来源获得的目的序列。

[0430] 靶标文库分子可进一步包括第二样品索引序列(例如,双样品索引),其包含第二

通用样品索引序列(例如,(160))。在一些实施例中,仅来自单个样品索引序列(例如,(170))的测序数据用于聚合酶克隆作图和/或模板配准,因为短随机序列提供了足够的核苷酸多样性和颜色平衡。来自第一通用样品索引序列(170)和第二通用样品索引序列(160)的测序数据可以用作双样品索引以在多重测定中区分从不同样品源获得的目的序列。在一些实施例中,第二样品索引序列(例如,(160))可以包括或不包括第二短随机序列(例如,NNN)。

[0431] 对目的序列区和样品索引区域进行测序的顺序也可用于改善对低多样性文库分子进行测序的挑战。例如,可以在对目的序列区域进行测序之前首先对样品索引区进行测序,并且可以将样品索引序列与目的序列区相关联。例如,可以首先对样品索引区进行测序,包括对短随机序列(例如,NNN)进行测序,并任选地对通用样品索引的至少一部分进行测序,然后对目的序列区域进行测序。在具有样品索引的文库分子群体中,短随机序列(例如,NNN)提供了文库分子的目的序列区可能无法提供的核苷酸多样性。样品索引的序列为聚合酶克隆作图和模板配准提供了改进的核苷酸多样性和颜色平衡。

[0432] 另外,当首先对样品索引区域进行测序时,测序后的样品索引区域的长度相对较短(例如,长度小于30个核苷酸),使得测序后的样品索引区域的产物去杂交更加完全。更温和的去杂交条件可用于移除测序样品索引区域的大部分或全部产物,这降低了来自与模板分子保持杂交的任何测序产物的残留信号水平。相比之下,目的序列区域通常比样品索引区域长得多(例如,长度超过100个核苷酸)。当目的序列区域在样品索引区域之前测序时,测序的目的序列区域的产物必须经历更严格的去杂交条件,以移除任何杂交至模板分子的剩余产物,这可能会损坏模板分子。

[0433] 本公开提供了核酸文库分子(100),其各自包含至少一个样品索引序列,其可用于在多重测定中区分从不同样品源获得的目的序列,其中至少一个样品索引序列包括链接至通用样品索引序列的短随机序列(例如,NNN)。在一些实施例中,左样品索引(160)包括链接至通用左样品索引序列的短随机序列(例如,NNN),且/或右样品索引(170)包含链接至右通用样品索引序列的短随机序列(例如,NNN)。至少一个样品索引序列可以包括用于改善碱基识别的序列多样性。至少一个样品索引序列可用于提高碱基识别的准确性。

[0434] 在一些实施例中,短随机序列(例如,NNN)位于通用样品索引序列(例如,(170)和/或(160))的上游,使得在测序运行期间,随机序列部分在通用样品索引序列之前被测序。在一些实施例中,短随机序列位于通用样品索引序列的下游,使得在测序运行期间,随机部分在通用样品索引序列之后被测序。

[0435] 在一些实施例中,在随机序列中,给定位置处的每个碱基“N”独立地选自A、G、C、T或U。在一些实施例中,随机序列缺乏具有2或3个相同核碱基的连续重复序列,例如AA、TT、CC、GG、UU、AAA、TTT、CCC、GGG或UUU。在一些实施例中,在文库分子群体中,通用样品索引序列(例如,(170)和/或(160))包括具有高多样性序列的短随机序列,该高多样性序列包括大致相等比例的所有四种核苷酸(例如,A、G、C、T和/或U),其将在测序运行的每个循环中表示。

[0436] 在一些实施例中,短随机序列(例如,NNN)包含3-20个核苷酸、或3-10个核苷酸、或3-8个核苷酸、或3-6个核苷酸、或3-5个核苷酸、或3-4个核苷酸。

[0437] 在一些实施例中,短随机序列(例如,NNN)包括但不限于:AGC、AGT、GAC、GAT、CAT、

CAG、TAG、TAC。技术人员将认识到,可以制备更多的随机序列(例如,64种可能的组合),其中随机序列中给定位置处的每个碱基“N”独立地选自A、G、C、T或U。

[0438] 在一些实施例中,通用样品索引序列包含5-20个核苷酸、或7-18个核苷酸、或9-16个核苷酸。

[0439] 在一些实施例中,右样品索引序列的群体中的各个右样品索引序列(例如,(170))包含通用样品索引序列和短随机序列(例如,NNN)。在一些实施例中,右样品索引序列的群体中的短随机序列的所有四个核苷酸碱基(例如,A、G、C和T/U)占总体碱基组成的约25%或约20-30%,以在对短随机序列(例如,NNN)进行测序期间,在每个测序循环中提供核苷酸多样性。

[0440] 在一些实施例中,在右样品索引序列的群体中,短随机序列中任何给定位置处的腺嘌呤(A)的比例为约20-30%或约15-35%或约10-40%。在一些实施例中,在右样品索引序列的群体中,短随机序列中任何给定位置处的鸟嘌呤(G)的比例为约20-30%或约15-35%或约10-40%。在一些实施例中,在右样品索引序列的群体中,短随机序列中任何给定位置处的胞嘧啶(C)的比例为约20-30%或约15-35%或约10-40%。在一些实施例中,在右样品索引序列的群体中,短随机序列中任何给定位置处的胸腺嘧啶(T)或尿嘧啶(U)的比例为约20-30%或约15-35%或约10-40%。

[0441] 在一些实施例中,在右样品索引序列的群体中,短随机序列中任何给定位置处的腺嘌呤(A)和胸腺嘧啶(T)的比例或腺嘌呤(A)和尿嘧啶(U)的比例为约10-65%。在一些实施例中,在右样品索引序列的群体中,短随机序列中任何给定位置处的鸟嘌呤(G)和胞嘧啶(C)的比例为约10-65%。

[0442] 在一些实施例中,在右样品索引序列的群体中,短随机序列的序列多样性确保,在至少对短随机序列(例如,NNN)进行测序期间,不出现少于四个不同核苷酸碱基的测序循环。

[0443] 包括直接链接至通用样品索引序列的短随机序列NNN的示例性样品索引序列包括但不限于:NNNGTAGGAGCC(SEQ ID NO:97);NNNCCGCTGCTA(SEQ ID NO:98);NNNAACAACAAG(SEQ ID NO:99);NNNGGTGGTCTA(SEQ ID NO:100);NNNTTGGCCAAC(SEQ ID NO:101);NNNCAGGAGTGC(SEQ ID NO:105);以及NNNATCACACTA(SEQ ID NO:106)。技术人员将认识到,通用样品索引可以是任何长度并且具有可用于在多重测定中区分从不同样品源获得的目的序列的任何序列。在给定样品索引的群体中,例如NNNGTAGGAGCC(SEQ ID NO:97),该群体含有个样品索引分子的混合物,该各个样品索引分子各自携带相同的通用样品索引序列(例如,GTAGGAGCC)和不同的短随机序列(例如,NNN),其中给定样品索引的群体中可能存在最多64种不同的短随机序列。

[0444] 包括短随机序列NNN的示例性样品索引序列在图9A-图9F的表1中列出。在一些实施例中,至少一种文库分子包含右样品索引,该右样品索引包含直接链接至右通用样品索引序列的短随机序列(例如,NNN)。在一些实施例中,至少一种文库分子进一步包含左样品索引,其包含不同于右通用样品索引序列的左通用样品索引序列。

[0445] 在一些实施例中,随机序列(例如,NNN)提供比率平衡的核碱基腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和/或尿嘧啶(参见图31)。在一些实施例中,在具有样品索引的文库分子群体中,随机序列(例如,NNN)与通用样品索引序列的至少一部分一起,提供了比率平衡的核

碱基腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和/或尿嘧啶,在测序运行的每个循环中表示。

[0446] 在一些实施例中,测序反应包括使用聚合酶和以对应于核碱基的不同荧光团标记的核苷酸(例如,核苷酸类似物)。在一些实施例中,使用标记的核苷酸对随机序列(例如,NNN)进行测序,提供了比率平衡的荧光颜色,其对应于测序运行的每个循环中的核碱基腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和/或尿嘧啶。在一些实施例中,使用标记的核苷酸对随机序列(例如,NNN)和通用样品索引序列的至少一部分进行测序,提供了比率平衡的荧光颜色,其对应于核碱基腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和/或尿嘧啶(例如,参见图31)。标记的核苷酸在测序反应期间发出荧光信号。在一些实施例中,测序反应在具有检测器的测序装置上进行,所述检测器捕获来自固定的模板分子上的测序反应的荧光图像。测序装置可被配置为将检测器捕获的荧光成像数据转发至计算机系统,该计算机系统经编程以确定流通池上固定的模板分子的位置(例如,作图)。计算机系统可以仅基于随机序列(例如,NNN),或基于随机序列(例如,NNN)和通用样品索引序列的至少一部分的荧光成像数据,生成固定模板分子的位置的图。因此,用于对随机序列(例如,NNN)和任选的通用样品索引序列的一部分进行测序的少量测序循环,可用于生成固定模板分子的位置的图。计算机系统可以被配置为仅提取随机序列(例如,NNN)或随机序列(例如,NNN)和通用样品索引序列的至少一部分的荧光颜色和强度。计算机系统可以被配置为使用给定固定的模板分子的位置以及与给定模板分子相关的荧光颜色和强度(其在对随机序列进行测序时建立),用于在对插入区域(110)进行测序时进行碱基识别。计算机系统可以被配置为在对随机序列(例如,NNN)和通用样品索引序列以及插入区域(110)进行测序的同时检测滞后和超前。在一些实施例中,随机序列(例如,NNN)在每个测序循环提供的荧光颜色的平衡比率,可以提高从检测器捕获的荧光图像处理的数据的质量,并且可以反过来提高由计算机系统确定固定模板分子在流通池上的位置以及颜色和强度的能力,所有这些都提高测序后插入区域(110)碱基识别的准确性和质量分数。

[0447] 在一些实施例中,测序反应包括使用聚合酶和以不同荧光团标记的多价分子,该不同荧光团对应于核苷酸单元的核碱基(例如,腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶),所述核苷酸单元附接至给定多价分子中核苷酸臂。在一些实施例中,各个多价分子的核心附接至对应核苷酸单元(例如,腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶)的荧光团,所述核苷酸单元附接至给定多价分子中的核苷酸臂(例如,参见图19-图23)。在一些实施例中,多价分子的至少一个核苷酸臂包含连附接至荧光团的接头和/或核苷酸碱基,并且其中附接至给定接头或核苷酸碱基的荧光团对应于核苷酸臂的核苷酸碱基(例如,腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶或尿嘧啶)。在一些实施例中,使用标记的多价分子对随机序列(例如,NNN)进行测序,提供了对应于测序运行的每个循环中的核碱基腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和/或尿嘧啶的荧光颜色的平衡比例。在一些实施例中,使用标记的多价分子对随机序列(例如,NNN)和通用样品索引序列的至少一部分进行测序,提供了对应于核碱基腺嘌呤、胞嘧啶、鸟嘌呤、胸腺嘧啶和/或尿嘧啶的荧光颜色的平衡比例(例如,参见图31)。标记的多价分子在测序反应期间发出荧光信号。在一些实施例中,测序反应在具有检测器的测序装置上进行,所述检测器捕获来自固定的模板分子上的测序反应的荧光图像。测序装置可被配置为将检测器捕获的荧光成像数据转发至计算机系统,该计算机系统经编程以确定流通池上固定的模板分子(聚合酶克隆)的位置(例如,作图)。计算机系统可以仅基于随

机序列(例如,NNN),或基于随机序列(例如,NNN)和通用样品索引序列的至少一部分的荧光成像数据,生成固定模板分子的位置的图。因此,用于对随机序列(例如,NNN)和任选的通用样品索引序列的一部分进行测序的少量测序循环,可用于生成固定模板分子的位置的图。计算机系统可以被配置为仅提取随机序列(例如,NNN)或随机序列(例如,NNN)和通用样品索引序列的荧光颜色和强度。计算机系统可以被配置为使用给定固定的模板分子的位置以及与给定模板分子相关的荧光颜色和强度(其在对随机序列进行测序时建立),用于在对插入区域(110)进行测序时进行碱基识别。计算机系统可以被配置为在对随机序列(例如,NNN)和通用样品索引序列以及插入区域(110)进行测序的同时检测滞后和超前。在一些实施例中,随机序列(例如,NNN)在每个测序循环提供的荧光颜色的平衡比率,可以提高从检测器捕获的荧光图像处理的数据的质量,并且可以反过来提高由计算机系统确定固定模板分子在流通池上的位置以及颜色和强度的能力,所有这些都可以提高测序后插入区域(110)碱基识别的准确性和质量分数。

[0448] 第一实施例:对样品索引序列测序的顺序

[0449] 在一些实施例中,测序的顺序包括:(1)对右样品索引(170)进行测序,其中右样品索引包含第一随机序列(例如,NNN)和右通用样品索引序列;(2)对左样品索引(160)进行测序;以及(3)对插入区域(110)进行测序。在一些实施例中,左样品索引(160)包含左通用样品索引序列。在一些实施例中,左样品索引(160)包含第二随机序列(例如,NNN)和左通用样品索引序列。在一些实施例中,对包括第一随机序列(例如,NNN)和右通用样品索引序列的右样品索引区域(170)进行测序,可以提供足够的核苷酸多样性,使得可以省略对左样品索引(160)的测序。

[0450] 在一些实施例中,用于对固定至支持物的模板分子进行测序的方法,其中各个模板分子包含:(i)针对第二表面引物的通用结合序列(120),(ii)左样品索引序列(160),其具有通用左样品索引序列,(iii)针对正向测序引物的通用结合序列(140),(iv)目的序列(110),(v)针对反向测序引物的通用结合序列(150),(vi)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),和(vii)针对第一表面引物的通用结合序列(130)(例如,参见图3和图4),其中该方法包括步骤(a):将模板分子与第一多个可溶性测序引物杂交,该第一多个可溶性测序引物与针对反向测序引物的通用结合序列(150)杂交,并对包括短随机序列(例如,NNN)的右样品索引序列(170)和右通用样品索引序列进行测序,由此产生与固定的模板分子杂交的第一多个样品索引延伸产物,其中第一多个样品索引延伸产物与右样品索引序列互补,该右样品索引序列具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN)。

[0451] 在一些实施例中,用于测序的方法进一步包括步骤(b):移除第一多个样品索引延伸产物并保留固定的模板分子。

[0452] 在一些实施例中,用于测序的方法进一步包括步骤(c):将保留的固定的模板分子与第二多个可溶性测序引物杂交,所述第二多个可溶性测序引物与针对第二表面引物的通用结合序列(120)杂交,并对左样品索引序列(160)进行测序,由此产生与固定模板分子杂交的第二多个样品索引延伸产物,其中第二多个样品索引延伸产物与具有左通用样品索引序列的左样品索引序列(160)互补。

[0453] 在一些实施例中,用于测序的方法进一步包括步骤(d):移除第二多个样品索引延

伸产物并保留固定的模板分子。

[0454] 在一些实施例中,用于测序的方法进一步包括步骤(e):将保留的固定的模板分子与第三多个可溶性测序引物杂交,该第三多个可溶性测序引物与针对正向测序引物的通用结合序列(140)杂交,并对插入区域(110)进行测序,由此产生与固定的模板分子杂交的多个插入延伸产物,其中该多个插入延伸产物与目的序列(110)互补。

[0455] 在一些实施例中,用于测序的方法进一步包括步骤(f1):将(i)插入区域(110)的序列指定到(ii)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),由此将插入区域识别为从第一来源获得。

[0456] 在一些实施例中,用于测序的方法进一步包括步骤(f2):将(i)插入区域(110)的序列指定到(ii)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),以及(iii)左样品索引序列(160),由此将插入区域识别为从第一来源获得。

[0457] 在一些实施例中,可以使用变性试剂,在诸如例如50-90℃的促进核酸变性的温度,移除步骤(b)和(d)的多个测序延伸产物,该变性试剂包含具有或不具有甲酰胺的SSC(例如,柠檬酸钠盐)缓冲液。

[0458] 在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何测序方法,其使用测序聚合酶和可检测地标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何两阶段测序方法,其采用测序聚合酶、可检测地标记的多价分子和核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何结合测序方法。

[0459] 在一些实施例中,固定至支持物的多个模板分子的密度为约 10^2 - 10^{15} 个/ mm^2 。在一些实施例中,多个模板分子固定在支持物上的随机位置。在一些实施例中,多个模板分子以预定图案固定在支持物上。

[0460] 第二实施例:对样品索引序列测序的顺序

[0461] 在一些实施例中,测序的顺序包括:(1)对右样品索引(170)进行测序,其中右索引包含随机序列(例如,NNN)和通用样品索引序列;(2)对插入区域(110)进行测序;以及(3)对左样品索引(160)进行测序。在一些实施例中,左样品索引(160)包含左通用样品索引序列。在一些实施例中,左样品索引(160)包含第二随机序列(例如,NNN)和左通用样品索引序列。在一些实施例中,对包括第一随机序列(例如,NNN)和右通用样品索引序列的右样品索引区域(170)进行测序,可以提供足够的核苷酸多样性,使得可以省略对左样品索引(160)的测序。

[0462] 在一些实施例中,用于对固定至支持物的模板分子进行测序的方法,其中各个模板分子包含:(i)针对第二表面引物的通用结合序列(120),(ii)左样品索引序列(160),其具有左通用样品索引序列,(iii)针对正向测序引物的通用结合序列(140),(iv)目的序列(110),(v)针对反向测序引物的通用结合序列(150),(vi)右样品索引序列(170),其具有直接链接至右通用样品索引序列的短随机序列(例如,NNN),和(vii)针对第一表面引物的通用结合序列(130)(例如,参见图3和图4),其中该方法包括步骤(a):将模板分子与第一多个可溶性测序引物杂交,该第一多个可溶性测序引物与针对反向测序引物的通用结合序列(150)杂交,并对包括短随机序列(例如,NNN)的右样品索引序列(170)和右通用样品索引序

列进行测序,由此产生与固定的模板分子杂交的第一多个样品索引延伸产物,其中第一多个样品索引延伸产物与右样品索引序列(170)互补,该右样品索引序列具有直接链接至右通用样品索引序列的短随机序列(例如,NNN)。

[0463] 在一些实施例中,用于测序的方法进一步包括步骤(b):移除第一多个样品索引延伸产物并保留固定的模板分子。

[0464] 在一些实施例中,用于测序的方法进一步包括步骤(c):将保留的固定的模板分子与第二多个可溶性测序引物杂交,所述第二多个可溶性测序引物与针对正向测序引物的通用结合序列(140)杂交,并对插入区域(110)进行测序,由此产生与固定模板分子杂交的第二多个插入延伸产物,其中第二多个插入延伸产物与目的序列(110)互补。

[0465] 在一些实施例中,用于测序的方法进一步包括步骤(d):移除多个插入延伸产物并保留固定的模板分子。

[0466] 在一些实施例中,用于测序的方法进一步包括步骤(e):将保留的固定的模板分子与第三多个可溶性测序引物杂交,该第三多个可溶性测序引物与针对第二表面引物的通用结合序列(120)杂交,并对左样品索引序列(160)进行测序,由此产生与固定的模板分子杂交的第二多个样品索引延伸产物,其中该第二多个样品索引延伸产物与具有左通用样品索引序列的左样品索引序列(160)互补。

[0467] 在一些实施例中,用于测序的方法进一步包括步骤步骤(f1):将(i)插入区域(110)的序列指定到(ii)右样品索引序列(170),其具有直接链接至右通用样品索引序列的短随机序列(例如,NNN),由此将插入区识别为从第一来源获得。

[0468] 在一些实施例中,用于测序的方法进一步包括步骤(f2):将(i)插入区域(110)的序列指定到(ii)右样品索引序列(170),其具有直接链接至右通用样品索引序列的短随机序列(例如,NNN),以及(iii)左样品索引序列(160),由此将插入区域识别为从第一来源获得。

[0469] 在一些实施例中,可以使用变性试剂,在诸如例如50-90℃的促进核酸变性的温度,移除步骤(b)和(d)的多个测序延伸产物,该变性试剂包含具有或不具有甲酰胺的SSC(例如,柠檬酸钠盐)缓冲液。

[0470] 在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何测序方法,其使用测序聚合酶和可检测地标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何两阶段测序方法,其采用测序聚合酶、可检测地标记的多价分子和核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何结合测序方法。

[0471] 在一些实施例中,固定至支持物的多个模板分子的密度为约 10^2 - 10^{15} 个/ mm^2 。在一些实施例中,多个模板分子固定在支持物上的随机位置。在一些实施例中,多个模板分子以预定图案固定在支持物上。

[0472] 第三实施例:对样品索引序列测序的顺序

[0473] 在一些实施例中,测序的顺序包括:(1)对插入区域(110)进行测序;(2)对右样品索引(170)进行测序,其中右索引包含随机序列(例如,NNN)和通用样品索引序列;以及(3)对左样品索引(160)进行测序。在一些实施例中,左样品索引(160)包含左通用样品索引序列。在一些实施例中,左样品索引(160)包含第二随机序列(例如,NNN)和左通用样品索引序

列。在一些实施例中,对包括第一随机序列(例如,NNN)和右通用样品索引序列的右样品索引区域(170)进行测序,可以提供足够的核苷酸多样性,使得可以省略对左样品索引(160)的测序。

[0474] 在一些实施例中,用于对固定至支持物的模板分子进行测序的方法,其中各个模板分子包含:(i)针对第二表面引物的通用结合序列(120),(ii)左样品索引序列(160),其具有左通用样品索引序列,(iii)针对正向测序引物的通用结合序列(140),(iv)目的序列(110),(v)针对反向测序引物的通用结合序列(150),(vi)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),和(vii)针对第一表面引物的通用结合序列(130)(例如,参见图3和图4),其中该方法包括步骤(a):将模板分子与第一个可溶性测序引物杂交,该第一个可溶性测序引物与针对正向测序引物的通用结合序列(140)杂交,并对插入区域(110)进行测序,由此产生与固定的模板分子杂交的多个插入延伸产物,其中多个插入延伸产物与目的序列(110)互补。

[0475] 在一些实施例中,用于测序的方法进一步包括步骤(b):移除多个插入延伸产物并保留固定的模板分子。

[0476] 在一些实施例中,用于测序的方法进一步包括步骤(c):将模板分子与第二个可溶性测序引物杂交,该第二个可溶性测序引物与针对反向测序引物的通用结合序列(150)杂交,并对包括短随机序列(例如,NNN)的右样品索引序列(170)和右通用样品索引序列进行测序,由此产生与固定的模板分子杂交的第一个样品索引延伸产物,其中第一个样品索引延伸产物与右样品索引序列互补,该右样品索引序列具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN)。

[0477] 在一些实施例中,用于测序的方法进一步包括步骤(d):移除第一个样品索引延伸产物并保留固定的模板分子。

[0478] 在一些实施例中,用于测序的方法进一步包括步骤(e):将保留的固定的模板分子与第三个可溶性测序引物杂交,该第三个可溶性测序引物与针对第二表面引物的通用结合序列(120)杂交,并对左样品索引序列(160)进行测序,由此产生与固定的模板分子杂交的第二个样品索引延伸产物,其中该第二个样品索引延伸产物与具有左通用样品索引序列的左样品索引序列(160)互补。

[0479] 在一些实施例中,用于测序的方法进一步包括步骤(f1):将(i)插入区域(110)的序列指定到(ii)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),由此将插入区域识别为从第一来源获得。

[0480] 在一些实施例中,用于测序的方法进一步包括步骤(f2):将(i)插入区域(110)的序列指定到(ii)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),以及(iii)左样品索引序列(160),由此将插入区域识别为从第一来源获得。

[0481] 在一些实施例中,可以使用变性试剂,在诸如例如50-90℃的促进核酸变性的温度,移除步骤(b)和(d)的多个测序延伸产物,该变性试剂包含具有或不具有甲酰胺的SSC(例如,柠檬酸钠盐)缓冲液。

[0482] 在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何测序方法,其使用测序聚合酶和可检测地标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)

的测序包括进行本文所述的任何两阶段测序方法,其采用测序聚合酶、可检测地标记的多价分子和核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)的测序包括进行本文所述的任何结合测序方法。

[0483] 在一些实施例中,固定至支持物的多个模板分子的密度为约 10^2 - 10^{15} 个/ mm^2 。在一些实施例中,多个模板分子固定在支持物上的随机位置。在一些实施例中,多个模板分子以预定图案固定在支持物上。

[0484] 第四实施例:测序的顺序

[0485] 在一些实施例中,测序的顺序包括:(1)对插入区域(110)的前3-5个碱基进行测序;(2)对右样品索引(170)进行测序,其中右索引包含任选的随机序列(例如,NNN)和通用样品索引序列;以及(3)对左样品索引(160)进行测序。在一些实施例中,左样品索引(160)包含左通用样品索引序列。在一些实施例中,左样品索引(160)包含第二随机序列(例如,NNN)和左通用样品索引序列。在一些实施例中,对插入区域(110)的前3-5个碱基进行测序可以提供足够的序列多样性,使得右样品索引(170)和左样品索引(160)不包括短随机序列(例如,NNN)。在一些实施例中,对包括第一随机序列(例如,NNN)和右通用样品索引序列的右样品索引区域(170)进行测序,可以提供足够的核苷酸多样性,使得可以省略对左样品索引(160)的测序。

[0486] 在一些实施例中,用于对固定至支持物的模板分子进行测序的方法,其中各个模板分子包含:(i)针对第二表面引物的通用结合序列(120),(ii)左样品索引序列(160),其具有左通用样品索引序列,(iii)针对正向测序引物的通用结合序列(140),(iv)目的序列(110),(v)针对反向测序引物的通用结合序列(150),(vi)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),和(vii)针对第一表面引物的通用结合序列(130)(例如,参见图3和图4),其中该方法包括步骤(a):将模板分子与第一多个可溶性测序引物杂交,该第一多个可溶性测序引物与针对正向测序引物的通用结合序列(140)杂交,并对插入区域(110)的前3-5个碱基进行测序,由此产生与固定的模板分子杂交的多个插入延伸产物,其中多个插入延伸产物与目的序列(110)互补。插入区域(110)的前3-5个碱基的序列可以为聚合酶克隆作图和模板配准提供足够的序列多样性和颜色平衡。

[0487] 在一些实施例中,用于测序的方法进一步包括步骤(b):移除多个插入延伸产物并保留固定的模板分子。

[0488] 在一些实施例中,用于测序的方法进一步包括步骤(c):将模板分子与第二多个可溶性测序引物杂交,该第二多个可溶性测序引物与针对反向测序引物的通用结合序列(150)杂交,并对包括若存在的短随机序列(例如,NNN)的右样品索引序列(170)和右通用样品索引序列进行测序,由此产生与固定的模板分子杂交的第一多个样品索引延伸产物,其中第一多个样品索引延伸产物与右样品索引序列(170)互补。

[0489] 在一些实施例中,用于测序的方法进一步包括步骤(d):移除第一多个样品索引延伸产物并保留固定的模板分子。

[0490] 在一些实施例中,用于测序的方法进一步包括步骤(e):将保留的固定的模板分子与第三多个可溶性测序引物杂交,该第三多个可溶性测序引物与针对第二表面引物的通用结合序列(120)杂交,并对左样品索引序列(160)进行测序,由此产生与固定的模板分子杂交的第二多个样品索引延伸产物,其中该第二多个样品索引延伸产物与具有左通用样品索

引序列的左样品索引序列(160)互补。

[0491] 在一些实施例中,用于测序的方法进一步包括步骤(f):移除第二多个样品索引延伸产物并保留固定的模板分子。

[0492] 在一些实施例中,用于测序的方法进一步包括步骤(g):将模板分子与第四多个可溶性测序引物杂交,所述第四多个可溶性测序引物与针对正向测序引物的通用结合序列(140)杂交,并对插入区域(110)的全长进行测区,由此产生与固定的模板分子杂交的多个全长插入延伸产物,其中多个全长插入延伸产物与目的序列(110)互补。

[0493] 在一些实施例中,用于测序的方法进一步包括步骤(h1):将(i)插入区域(110)的全长序列指定到(ii)右样品索引序列(170),由此将插入区域识别为从第一来源获得。

[0494] 在一些实施例中,用于测序的方法进一步包括步骤(h2):将(i)插入区域(110)的全长序列指定到(ii)右样品索引序列(170)和(iii)左样品索引序列(160),由此将插入区域识别为从第一来源获得。

[0495] 在一些实施例中,可以使用变性试剂进行,在诸如例如50-90℃的促进核酸变性的温度,移除步骤(b)、(d)和(f)的多个测序延伸产物,所述变性试剂包含具有或不具有甲酰胺的SSC(例如,柠檬酸钠盐)缓冲液。

[0496] 在一些实施例中,步骤(a)、(c)、(e)和(g)的测序包括进行本文所述的任何测序方法,其使用测序聚合酶和可检测地标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)、(e)和(g)的测序包括进行本文所述的任何两阶段测序方法,其采用测序聚合酶、可检测地标记的多价分子和核苷酸类似物。在一些实施例中,步骤(a)、(c)、(e)和(g)的测序包括进行本文所述的任何结合测序方法。

[0497] 在一些实施例中,固定至支持物的多个模板分子的密度为约 10^2 - 10^{15} 个/ mm^2 。在一些实施例中,多个模板分子固定在支持物上的随机位置。在一些实施例中,多个模板分子以预定图案固定在支持物上。

[0498] 第五实施例:测序的顺序

[0499] 在一些实施例中,测序的顺序包括:(1)对固定的模板分子的插入区域(110)的前3-5个碱基进行测序(例如,正向测序);(2)对右样品索引(170)进行测序,其中右索引包含第一随机序列(例如,NNN)和右通用样品索引序列;(3)对左样品索引(160)进行测序;(4)进行成对轮转反应,使固定的模板分子被与模板分子互补的固定的链取代;以及(5)对固定互补链的插入区域(110)的全长进行测序(例如,反向测序)。在一些实施例中,文库分子群体的插入区域(110)的前3-5个碱基的序列可以提供足够的序列多样性以改善碱基识别准确性。在一些实施例中,左样品索引(160)包含左通用样品索引序列。在一些实施例中,左样品索引(160)包含第二随机序列(例如,NNN)和左通用样品索引序列。在一些实施例中,对包括第一随机序列(例如,NNN)和右通用样品索引序列的右样品索引区域(170)进行测序,可以提供足够的核苷酸多样性,使得可以省略对左样品索引(160)的测序。

[0500] 在一些实施例中,用于对固定至支持物的模板分子进行测序的方法,其中各个模板分子共价链接至固定的缺少尿嘧啶碱基的捕获引物,并且各个模板分子包含随机分配的尿嘧啶碱基,并且各个模板分子包含:(i)针对第二表面引物的通用结合序列(120),(ii)左样品索引序列(160),其具有左通用样品索引序列,(iii)针对正向测序引物的通用结合序列(140),(iv)目的序列(110),(v)针对反向测序引物的通用结合序列(150),(vi)右样品索

引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),和(vii)针对第一表面引物的通用结合序列(130)(例如,参见图3和图4),其中该方法包括步骤(a):将模板分子与第一多个可溶性测序引物(例如,正向测序引物)杂交,该第一多个可溶性测序引物与针对正向测序引物的通用结合序列(140)杂交,并对插入区域(110)的前3-5个碱基进行测序,由此产生与固定的模板分子杂交的多个插入延伸产物,其中多个插入延伸产物与目的序列(110)互补。

[0501] 在一些实施例中,用于测序的方法进一步包括步骤(b):移除多个插入延伸产物并保留固定的模板分子。

[0502] 在一些实施例中,用于测序的方法进一步包括步骤(c):将模板分子与第二多个可溶性测序引物杂交,该第二多个可溶性测序引物与针对反向测序引物的通用结合序列(150)杂交,并对包括短随机序列(例如,NNN)的右样品索引序列(170)和右通用样品索引序列进行测序,由此产生与固定的模板分子杂交的第一多个样品索引延伸产物,其中第一多个样品索引延伸产物与右样品索引序列互补,该右样品索引序列具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN)。

[0503] 在一些实施例中,用于测序的方法进一步包括步骤(d):移除第一多个样品索引延伸产物并保留固定的模板分子。

[0504] 在一些实施例中,用于测序的方法进一步包括步骤(e):将保留的固定的模板分子与第三多个可溶性测序引物杂交,该第三多个可溶性测序引物与针对第二表面引物的通用结合序列(120)杂交,并对左样品索引序列(160)进行测序,由此产生与固定的模板分子杂交的第二多个样品索引延伸产物,其中该第二多个样品索引延伸产物与具有左通用样品索引序列的左样品索引序列(160)互补。

[0505] 在一些实施例中,用于测序的方法进一步包括步骤(f):通过使用链置换聚合酶和多个核苷酸,进行引物延伸反应,来替换与固定的模板分子杂交的第二多个样品索引延伸产物,以产生与固定的模板分子(包括固定的捕获引物)杂交的延伸产物。

[0506] 在一些实施例中,用于测序的方法进一步包括步骤(g):通过在固定的模板分子中的尿嘧啶位点处生成脱碱基位点,并在脱碱基位点处生成间隙,来移除固定的模板分子,由此产生含有间隙的模板分子,同时保留在步骤(f)中生成的延伸产物,其中各个延伸产物通过与固定的捕获引物杂交而被保留。在一些实施例中,通过进行步骤(g)和(h)来实现成对轮转。

[0507] 在一些实施例中,用于测序的方法进一步包括步骤(h):将保留的延伸产物与第四多个可溶性测序引物(例如,反向测序引物)杂交,该第四多个可溶性测序引物与针对反向测序引物的通用结合序列(150)杂交,并且对插入区域(110)进行测序(例如,对插入区域(110)的至少一部分或全长进行测序)。

[0508] 在一些实施例中,用于测序的方法进一步包括步骤步骤(i1):将(i)插入区域(110)的序列指定到(ii)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),由此将插入区域识别为从第一来源获得。

[0509] 在一些实施例中,用于测序的方法进一步包括步骤(i2):将(i)插入区域(110)的序列指定到(ii)右样品索引序列,其具有直接链接至右通用样品索引序列(170)的短随机序列(例如,NNN),以及(iii)左样品索引序列(160),由此将插入区域识别为从第一来源获

得。

[0510] 在一些实施例中,可以使用变性试剂,在诸如例如50-90℃的促进核酸变性的温度,移除步骤(b)和(d)的多个测序延伸产物,该变性试剂包含具有或不具有甲酰胺的SSC(例如,柠檬酸钠盐)缓冲液。

[0511] 在一些实施例中,步骤(a)、(c)、(e)和(h)的测序包括进行本文所述的任何测序方法,其使用测序聚合酶和可检测地标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)、(e)和(h)的测序包括进行本文所述的任何两阶段测序方法,其采用测序聚合酶、可检测地标记的多价分子和核苷酸类似物。在一些实施例中,步骤(a)、(c)、(e)和(h)的测序包括进行本文所述的任何结合测序方法。

[0512] 在一些实施例中,固定至支持物的多个模板分子的密度为约 10^2 - 10^{15} 个/ mm^2 。在一些实施例中,多个模板分子固定在支持物上的随机位置。在一些实施例中,多个模板分子以预定图案固定在支持物上。

[0513] 对3聚体随机序列进行测序以生成聚合酶克隆图

[0514] 本公开提供了用于对核酸进行测序的方法,其包括:(a)提供固定在支持物上(例如,固定在随机或预定位置处)的多个核酸模板分子,其中各个固定的模板分子包含插入序列区域和一个样品索引,其中每个样品索引包含与识别插入序列的样品源的通用样品索引序列接合的3聚体随机序列,其中不同的固定的模板分子具有不同的3聚体随机序列和相同的通用样品索引序列,并且其中固定的模板分子具有不同的插入序列;(b)使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂对多个固定的模板分子的3聚体随机序列进行三个循环的聚合酶介导的测序反应,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环和第三测序循环中的每一者中,在多个固定的模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及(c)使用在步骤(b)中获得的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。

[0515] 在一些实施例中,在用于对核酸进行测序的方法中,步骤(b)的平衡多样性是在第一测序循环、第二测序循环和第三测序循环中的每一者中检测和成像的核碱基A、G、C和T/U中的每一者的约5-85%、或约5-60%、或约10-50%、或约15-55%、或约25-75%。

[0516] 在一些实施例中,在用于对核酸进行测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的通用样品索引序列进行测序;(b)对多个固定的模板分子的插入序列区域进行测序;以及(c)将在步骤(b)中获得的给定模板分子的插入序列与在步骤(a)中获得的来自同一给定模板分子的通用样品索引序列进行指定,由此识别给定插入序列的样品源。

[0517] 在一些实施例中,在用于对核酸进行测序的方法中,多个核酸模板分子进一步包含第二样品索引,该第二样品索引包含识别插入序列的样品源的第二通用样品索引序列,并且第二样品索引缺少随机序列。

[0518] 在一些实施例中,在用于对核酸进行测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的3聚体随机序列进行测序,以获得在第一测序循环、第二测序循环和第

三测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性,从而生成多个固定的模板分子的位置的图;(b)对多个固定的模板分子的第一通用样品索引序列进行测序;(c)对多个固定的模板分子的第二通用样品索引序列进行测序;(d)对多个固定的模板分子的插入序列区域进行测序;以及(e)将在步骤(c)中获得的给定模板分子的插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定插入序列的样品源。

[0519] 本公开提供了用于对核酸进行测序的方法,其包括:(a)提供固定在支持物上(例如,固定在随机或预定位置处)的多个核酸模板分子,其中各个固定模板分子包含插入序列区域和一个样品索引,其中每个样品索引包含与通用样品索引序列接合的3聚体随机序列,该通用样品索引序列识别插入序列的样品源,其中通用样品索引序列包含3-20个核苷酸,其中不同的固定的模板分子具有不同的3聚体随机序列和相同的通用样品索引序列,并且其中固定的模板分子具有不同的插入序列;(b)使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂对多个固定的模板分子的3聚体随机序列和通用样品索引序列的第一碱基位置进行四个循环的聚合酶介导的测序反应,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中四个循环的测序包括对从结合至固定的模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的通用样品索引序列的第一碱基位置和3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中,在多个固定的模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及(c)使用在步骤(b)中获得的四个循环的聚合酶介导的测序反应的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。

[0520] 在一些实施例中,在用于对核酸进行测序的方法中,步骤(b)的平衡多样性是在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者检测和成像的核碱基A、G、C和T/U中的每一者的约5-85%、或约5-60%、或约10-50%、或约15-55%、或约25-75%。

[0521] 在一些实施例中,在对核酸进行测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的通用样品索引序列的其余碱基位置进行测序;(b)对多个固定的模板分子的插入序列区域进行测序;以及(c)将在步骤(b)中获得的给定模板分子的插入序列与在步骤(a)中获得的来自同一给定模板分子的通用样品索引序列进行指定,由此识别给定插入序列的样品源。

[0522] 在一些实施例中,在用于对核酸进行测序的方法中,多个核酸模板分子进一步包含:第二样品索引,该第二样品索引包含识别插入序列的样品源的第二通用样品索引序列,并且第二样品索引缺少随机序列。

[0523] 在一些实施例中,在用于对核酸进行测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的3聚体随机序列和通用样品索引序列的第一碱基位置进行测序,以获得在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中检测和成像的A、G、C和T/U核碱基的平衡多样性,从而生成多个固定的模板分子的位置的图;(b)对多个固定的模板分子的第一通用样品索引序列的其余碱基位置进行测序;(c)对多个固定的模板分子的第二通用样品索引序列进行测序;(d)对多个固定的模板分子的插入序列区域

进行测序;以及(e)将在步骤(c)中获得的给定模板分子的插入序列与在步骤(a)和(b)中获得的来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定插入序列的样品源。

[0524] 在一些实施例中,在任何用于对核酸进行测序的方法中,支持物包括玻璃或塑料基底。在一些实施例中,支持物被配置在流通池通道、流通池或毛细管腔上。在一些实施例中,支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。在一些实施例中,至少一个亲水性聚合物涂层包含选自以下项组成的组的分子:聚乙二醇(PEG)、聚(乙烯醇)(PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮)(PVP)、聚(丙烯酸)(PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺)(PNIPAM)、聚(甲基丙烯酸甲酯)(PMA)、聚(2-羟乙基甲基丙烯酸酯)(PHEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(POEGMA)、聚谷氨酸(PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少四个支链的支链亲水性聚合物分子。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少1000道尔顿的分子量的聚合物分子。

[0525] 在一些实施例中,在用于对核酸进行测序的任何方法中,固定的模板分子包含多个具有插入序列和所述一个样品索引的串联重复序列的固定的多联体分子。在一些实施例中,固定的模板分子包含多个不同的成簇模板分子,该成簇模板分子具有插入序列的一个拷贝和一个样品索引的一个拷贝,其中成簇模板分子经由桥式扩增产生。在一些实施例中,位于支持物上随机或预定位置处的固定的核酸模板分子的密度为 10^4 - 10^8 个/ mm^2 。在一些实施例中,插入序列的样品源是基因组DNA、双链cDNA或细胞游离循环DNA。

[0526] 在一些实施例中,在用于对核酸进行测序的任何方法中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。在一些实施例中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。在一些实施例中,可检测地标记的核苷酸试剂包含多价分子,每个多价分子包含(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0527] 在一些实施例中,在用于对核酸进行测序的任何方法中,结合至步骤(b)中固定的模板分子的可检测地标记的核苷酸试剂包含,与测序引物杂交以形成双链体的各个固定的模板分子,并且该双链体结合至聚合酶形成复合聚合酶,并且该复合聚合酶结合至可检测地标记的核苷酸试剂。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,并且该条件适用于抑制核苷酸掺入,其中可检测

地标记的核苷酸试剂包含多价分子,该多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0528] 在一些实施例中,在用于对核酸进行测序的任何方法中,固定的模板分子包括固定的多联体分子,该多联体分子与多个测序引物杂交以在同一多联体分子上形成至少第一双链体和第二双链体,其中第一和双链体结合至第一聚合酶,并且第二双链体结合至第二聚合酶,以形成第一复合聚合酶和第二复合聚合酶,并且其中该方法包括:(a)使多个多价分子与同一多联体模板分子上的第一复合聚合酶和第二复合聚合酶接触,其中各个多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,并且间隔基附接至接头,并且接头附接至核苷酸单元,其中接触在适用于将多个多价分子中的单个多价分子与第一复合聚合酶和第二复合聚合酶结合的条件下进行,其中单个多价分子的第一核苷酸单元与第一复合聚合酶结合,该第一复合酶包括与多联体模板分子的第一部分杂交的第一测序引物,由此形成第一结合复合物,并且其中单个多价分子的第二核苷酸单元与第二复合聚合酶结合,该第二复合聚合酶包括与多联体模板分子的第二部分杂交的第二测序引物,由此形成第二结合复合物,其中结合至同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,并且其中接触在适用于抑制的第一结合复合物和第二结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化的掺入的条件下进行;(b)检测同一多联体模板分子上的第一结合复合物和第二结合复合物;(c)对从可检测地标记的多价分子发出的光学颜色信号进行成像,可检测地标记的多价分子在同一多联体模板分子上形成第一结合复合物和第二结合复合物;以及(d)识别第一结合复合物中的第一核苷酸单元,由此确定多联体模板分子的第一部分的序列,并且识别第二结合复合物中的第二核苷酸单元,由此确定多联体模板分子的第二部分的序列。

[0529] 对3聚体随机序列进行多重测序以生成聚合酶克隆图

[0530] 本公开提供了对核酸进行多重测序的方法,其包括:(a)提供第一多个文库分子,该多个文库分子中的每个分子包含(i)源自第一样品源的插入序列区域,(ii)第一样品索引,其具有与第一通用样品索引序列接合的3聚体随机序列,以及(iii)具有第二通用样品索引序列的第二样品索引,该第二样品索引缺少随机序列,其中第一通用样品索引序列和第二通用样品索引序列的组合唯一地识别插入序列的第一样品源,其中不同的第一文库分子具有不同的3聚体随机序列并且具有不同的插入序列;(b)提供第二多个文库分子,该多个文库分子中的每个分子包含(i)源自第二样品源的插入序列区域,(ii)第三样品索引,其具有与第三通用样品索引序列接合的3聚体随机序列,以及(iii)具有第四通用样品索引序列的第四样品索引,该第四样品索引缺少随机序列,其中第三通用样品索引序列和第四通用样品索引序列的组合唯一地识别插入序列的第二样品源,其中不同的第二文库分子具有不同的3聚体随机序列并且具有不同的插入序列;(c)汇集第一多个文库分子和第二多个文库分子;(d)将经汇集的文库分子分配到支持物上并进行扩增反应以产生固定至支持物的随机位置处的多个克隆扩增的模板分子(例如,固定在随机或预定位置处);(e)使用包含不同类型的核碱基A、G、C和T/U的混合物的多种可检测地标记的核苷酸试剂,对第一样品索引

和第三样品索引的3聚体随机序列进行三个循环的聚合酶介导的测序反应,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的扩增模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环和第三测序循环的每一者中,在多个固定的扩增模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及(f)使用在步骤(e)中获得的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。

[0531] 在一些实施例中,在用于对核酸进行多重测序的方法中,步骤(e)的平衡多样性是在第一测序循环、第二测序循环和第三测序循环中的每一中检测和成像的核碱基A、G、C和T/U中的每一者的约5-85%、或约5-60%、或约10-50%、或约15-55%、或约25-75%。

[0532] 在一些实施例中,在用于对核酸进行多重测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的第一通用样品索引序列进行测序;(b)对多个固定的模板分子的第二通用样品索引序列进行测序;(c)对源自第一文库分子的多个固定的模板分子的插入序列区域进行测序;以及(d)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定插入序列的第一样品源。

[0533] 在一些实施例中,在用于对核酸进行多重测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的第三通用样品索引序列进行测序;(b)对多个固定的模板分子的第四通用样品索引序列进行测序;(c)对源自第二文库分子的多个固定的模板分子的插入序列区域进行测序;以及(d)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第三通用样品索引序列和第四通用样品索引序列进行指定,由此识别给定插入序列的第二样品源。

[0534] 本公开提供了对核酸进行多重测序的方法,其包括:(a)提供第一多个文库分子,该多个文库分子中的每个分子包含(i)源自第一样品源的插入序列区域,(ii)第一样品索引,其具有与第一通用样品索引序列接合的3聚体随机序列,以及(iii)具有第二通用样品索引序列的第二样品索引,该第二样品索引缺少随机序列,其中第一通用样品索引序列和第二通用样品索引序列的组合唯一地识别插入序列的第一样品源,其中第一通用样品索引序列包含3-20个核苷酸,其中不同的第一文库分子具有不同的3聚体随机序列并且具有不同的插入序列;(b)提供第二多个文库分子,该多个文库分子中的每个分子包含(i)源自第二样品源的插入序列区域,(ii)第三样品索引,其具有与第三通用样品索引序列接合的3聚体随机序列,以及(iii)具有第四通用样品索引序列的第四样品索引,该第四样品索引缺少随机序列,其中第三通用样品索引序列和第四通用样品索引序列的组合唯一地识别插入序列的第二样品源,其中第三通用样品索引序列包含3-20个核苷酸,其中不同的第二文库分子具有不同的3聚体随机序列并且具有不同的插入序列;(c)汇集第一多个文库分子和第二多个文库分子;(d)将经汇集的文库分子分配到支持物上并进行扩增反应以产生固定至支持物的随机位置处的多个克隆扩增的模板分子(例如,固定在随机或预定位置处);(e)使用包含不同类型的核碱基A、G、C和T/U的混合物的多个可检测地标记的核苷酸试剂对第一样品序列和第三样品索引的3聚体随机序列进行四个循环的聚合酶介导的测序反应,并对第

一通用样品索引序列和第三通用样品索引序列的第一碱基位置进行测序,其中核苷酸试剂包含与每种不同类型的核碱基相对应的不同的可检测颜色标记,其中三个循环的测序包括对从结合至固定的扩增模板分子的可检测地标记的核苷酸试剂发出的光学颜色信号进行检测和成像,由此确定多个固定的模板分子中的各个模板分子中的3聚体随机序列的序列,并且其中在第一测序循环、第二测序循环、第三测序循环和第四测序循环中的每一者中,在多个固定的扩增模板分子之间,对A、G、C和T/U核碱基的平衡多样性进行检测和成像;以及(f)使用在步骤(e)中获得的图像来生成多个固定的模板分子的位置的图,其中插入区域的序列不用于生成图。

[0535] 在一些实施例中,在用于对核酸进行多重测序的方法中,步骤(e)的平衡多样性是在第一测序循环、第二测序循环、第三测序循环和第四个测序循环的每一者中检测到并成像的核碱基A、G、C和T/U中的每一者的约5-85%、或约5-60%、或约10-50%、或约15-55%、或约25-75%。

[0536] 在一些实施例中,在用于对核酸进行多重测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的第一通用样品索引序列的其余碱基位置进行测序;(b)对多个固定的模板分子的第二通用样品索引序列进行测序;(c)对源自第一文库分子的多个固定的模板分子的插入序列区域进行测序;以及(d)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第一通用样品索引序列和第二通用样品索引序列进行指定,由此识别给定插入序列的第一样品源。

[0537] 在一些实施例中,在用于对核酸进行多重测序的方法中,该方法进一步包括:(a)对多个固定的模板分子的第三通用样品索引序列的其余碱基位置进行测序;(b)对多个固定的模板分子的第四通用样品索引序列进行测序;(c)对源自第二文库分子的多个固定的模板分子的插入序列区域进行测序;以及(d)将在步骤(c)中获得的给定模板分子的插入序列与来自同一给定模板分子的第三通用样品索引序列和第四通用样品索引序列进行指定,由此识别给定插入序列的第二样品源。

[0538] 在一些实施例中,在用于对核酸进行多重测序的任何方法中,支持物包括玻璃或塑料基底。在一些实施例中,支持物被配置在流通池通道、流通池或毛细管腔上。在一些实施例中,支持物用具有不大于45度的水接触角的至少一个亲水性聚合物涂层钝化。在一些实施例中,至少一个亲水性聚合物涂层包含选自以下项组成的组的分子:聚乙二醇(PEG)、聚(乙烯醇)(PVA)、聚(乙烯基吡啶)、聚(乙烯基吡咯烷酮)(PVP)、聚(丙烯酸)(PAA)、聚丙烯酰胺、聚(N-异丙基丙烯酰胺)(PNIPAM)、聚(甲基丙烯酸甲酯)(PMA)、聚(2-羟乙基甲基丙烯酸酯)(PHEMA)、聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(POEGMA)、聚谷氨酸(PGA)、聚赖氨酸、聚葡萄糖苷、链霉亲和素和葡聚糖。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少四个支链的支链亲水性聚合物分子。在一些实施例中,至少一个亲水性聚合物涂层包含具有至少1000道尔顿的分子量的聚合物分子。

[0539] 在一些实施例中,在用于对核酸进行多重测序的任何方法中,固定的模板分子包含具有一个样品索引和插入序列的串联重复序列的多个固定的多联体分子。在一些实施例中,固定的模板分子包含多个不同的成簇模板分子,该成簇模板分子具有插入序列的一个拷贝和一个样品索引的一个拷贝,其中成簇模板分子经由桥式扩增产生。在一些实施例中,支持物上固定的核酸模板分子(例如,固定在随机或预定位置处)的密度是 10^4 - 10^8 个/ mm^2 。

在一些实施例中,插入序列的样品源是基因组DNA、双链cDNA或细胞游离循环DNA。

[0540] 在一些实施例中,在用于对核酸进行多重测序的任何方法中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。在一些实施例中,可检测地标记的核苷酸试剂包含各自包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团的核苷酸。在一些实施例中,可检测地标记的核苷酸试剂包含多价分子,每个多价分子包含(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0541] 在一些实施例中,在用于对核酸进行多重测序的任何方法中,在结合至步骤(e)中固定的模板分子的可检测地标记的核苷酸试剂包含各个固定的模板分子,其与测序引物杂交以形成双链体,并且该双链体结合至聚合酶以形成复合聚合酶,并且该复合聚合酶结合至可检测地标记的核苷酸试剂。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、五碳糖部分、1-10个磷酸酯基团和荧光团。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合并将可检测地标记的核苷酸掺入到杂交的测序引物中的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,其中可检测地标记的核苷酸试剂包含芳香族核碱基、在3'碳糖位置处具有链终止基团的五碳糖部分、1-10个磷酸酯基团和荧光团。在一些实施例中,在适用于将可检测地标记的核苷酸试剂与复合聚合酶结合的条件下,复合聚合酶结合至可检测地标记的核苷酸试剂,并且该条件适用于抑制核苷酸掺入,其中可检测地标记的核苷酸试剂包含多价分子,该多价分子包括(1)核心、(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分、(ii)包含PEG部分的间隔基、(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,其中间隔基附接至接头,其中接头附接至核苷酸单元。

[0542] 在一些实施例中,在用于对核酸进行多重测序的任何方法中,固定的模板分子包括固定的多联体分子,其与多个测序引物杂交以在同一多联体分子上形成至少第一双链体和第二双链体,其中第一和双链体结合至第一聚合酶,并且第二双链体结合至第二聚合酶,以形成第一复合聚合酶和第二复合聚合酶,并且其中该方法包括:(a)使多个多价分子与同一多联体模板分子上的第一复合聚合酶和第二复合聚合酶接触,其中各个多价分子包括(1)核心,(2)多个核苷酸臂和(3)至少一个荧光团,其中各个核苷酸臂包含(i)核心附接部分,(ii)包含PEG部分的间隔基,(iii)接头和(iv)核苷酸单元,其中核心附接至多个核苷酸臂,并且间隔基附接至接头,并且接头附接至核苷酸单元,其中接触在适用于将多个多价分子中的单个多价分子与第一复合聚合酶和第二复合聚合酶结合的条件下进行,其中单个多价分子的第一核苷酸单元与第一复合聚合酶结合,该第一复合酶包括与多联体模板分子的第一部分杂交的第一测序引物,由此形成第一结合复合物,并且其中单个多价分子的第二核苷酸单元与第二复合聚合酶结合,该第二复合聚合酶包括与多联体模板的第二部分杂交的第二测序引物,由此形成第二结合复合物,其中结合至同一多价分子的第一结合复合物和第二结合复合物形成亲和力复合物,并且其中接触在适用于抑制第一结合复合物和第二

结合复合物中结合的第一核苷酸单元和第二核苷酸单元的聚合酶催化的掺入的条件下进行; (b) 检测同一多联体模板分子上的第一结合复合物和第二结合复合物; (c) 对从可检测地标记的多价分子发出的光学颜色信号进行成像, 可检测地标记的多价分子在同一多联体模板分子上形成第一结合复合物和第二结合复合物; 以及 (d) 识别第一结合复合物中的第一核苷酸单元, 由此确定多联体模板分子的第一部分的序列, 并且识别第二结合复合物中的第二核苷酸单元, 由此确定多联体模板分子的第二部分的序列。

[0543] 文库分子

[0544] 本公开提供了本文所述的核酸文库分子 (100) 中的任一个, 以及使用核酸文库分子 (100) 制备文库-夹板复合物 (300) 和共价闭合环状文库分子 (400) 的方法。核酸文库分子包含DNA、RNA、cDNA或嵌合DNA/RNA。核酸文库分子可以为单链或双链的, 或者可包括单链或双链部分。核酸文库分子可以是线性的、多联体的、共价闭合环状的、哑铃形、发夹形或其他形式。

[0545] 本文所述的核酸文库分子通常是指各自包含共价接合至至少一个通用衔接子序列 (例如, (120) 和 (130)) 的目的序列 (例如, 插入片段 (110)) 的核酸分子群体。群体中的各个文库分子可以包括额外的通用衔接子序列, 并且可以进一步包含至少一种索引序列或唯一识别序列。群体中的各个文库分子可以具有与群体中的其他文库分子相同或不同的目的序列。

[0546] 核酸文库分子的插入区域包含从包括生物学样品 (例如, 新鲜或活样品), 诸如单个细胞、多个细胞、或组织的任何来源提取的目的序列。插入区域可从健康或疾病细胞或组织分离。插入区域可以从存档样品获得, 诸如新鲜冷冻石蜡包埋 (FFPE) 样品, 或从针活检、循环肿瘤细胞、细胞游离循环DNA (例如, 来自肿瘤细胞或胎儿) 获得。通常用裂解缓冲液来处理细胞或组织以释放其DNA和RNA, 并且将期望的核酸与不期望的大分子 (诸如蛋白质) 分离。

[0547] 核酸文库分子的插入区域可以任何形式被分离, 包括克隆或扩增的染色体的、基因组的 (例如, 全基因组的)、细胞器的 (例如, 线粒体的、叶绿体或核糖体的)、重组分子。核酸文库分子的插入区域可以是甲基化的或非甲基化的。

[0548] 插入区域可从任何生物体被分离, 包括病毒、真菌、原核生物或真核生物。插入区域可从任何生物体被分离, 包括人、猿、类人猿、犬、猫科动物、牛、马、鼠、猪、山羊、狼、蛙、鱼、植物、昆虫或细菌。插入区域可从空气、水、土壤或食物中携带的生物体被分离。

[0549] 插入区域可从任何生物学流体被分离, 包括血液、尿液、血清、淋巴液、肿瘤、唾液、肛门分泌物、阴道分泌物、羊膜样品、汗液、精液、环境样品或培养物样品。插入区域可从任何器官被分离, 包括头、颈、脑、乳房、卵巢、子宫颈、结肠、直肠、子宫内膜、胆囊、肠、膀胱、前列腺、睾丸、肝、肺、肾、食道、胰腺、甲状腺、垂体、胸腺、皮肤、心脏、喉或其他器官。

[0550] 可使用重组核酸技术来制备插入区域, 包括但不限于载体克隆、转基因宿主细胞制备、宿主细胞培养和/或PCR扩增的任意组合。

[0551] 插入区域可以在一端或两端附加至至少一种通用衔接子序列以形成重组核酸文库分子。可使用天然核苷酸, 通过化学合成程序来制备通用衔接子序列, 该天然核苷酸具有或不具有核苷酸类似物或修饰的核苷酸键, 来赋予某些性质, 包括对酶消化的抗性或增加的热稳定性。抑制核酸酶消化的核苷酸类似物和经修饰的核苷酸键的实例包括: 硫代磷酸

酯、2'-O-甲基RNA、反向dT和2'3'双脱氧-dT。包括锁核酸(LNA)的插入区域具有增加的热稳定性。

[0552] 插入区域可以为经片段化或未经片段化的形式。片段化的插入区域可以通过机械力、酶促或化学片段化方法获得。可使用产生具有重叠序列或非重叠序列的片段的群体的程序来生成片段化的插入区。

[0553] 机械片段化通常生成经随机地片段化的核酸分子。机械片段化方法包括机械剪切,诸如流体剪切、恒定剪切和脉动剪切。机械片段化方法也包括机械应力,包括超声处理、雾化和声空化。

[0554] 可在适合于生成经随机地或经非随机地片段化的核酸分子的条件下进行酶促片段化程序。例如,可以进行限制性核酸内切酶消化完全以产生非随机片段化的核酸分子。替代性地,可进行部分或不完全限制酶消化以生成经随机地片段化的核酸分子。使用限制性核酸内切酶的酶促片段化,包括选自由I型、II型、IIs型、IIB型、III型或IV型限制性酶组成的组的任意一种或者两种或更多种限制性酶的任意组合。酶促片段化包括用稀切限制酶消化核酸,该稀切限制酶包含Not I、Asc I、Bae I、AspC I、Pac I、Fse I、Sap I、Sfi I或Psr I。酶促片段化包括使用缺口限制性核酸内切酶、核酸内切酶和/或核酸外切酶的任意组合。酶促片段化可以通过进行缺口平移反应来实现。

[0555] 可使用序列特异性引物用PCR来生成插入区域的片段,该序列特异性引物与基因组DNA样品中的靶区杂交以生成具有已知片段长度和序列的插入区域。

[0556] 使用CRISPR/Cas9的靶向基因组片段化方法可用于生成经片段化的插入区域。

[0557] 也可使用基于转座酶的标签化(tagmentation)方法使用NEXTERA(来自Epicentre)来生成插入部分的片段。

[0558] 插入区域可以是单链或双链的。双链插入区域域的末端可以是平端的,或具有5'悬垂端或3'悬垂端,或其任意组合。可使插入区域的一个或两个端部经历酶促加尾反应,以通过采用末端转移酶反应来生成非模板多A尾。插入区域的末端可以兼容接合到至少一种通用衔接子序列。

[0559] 插入区域可以是任何长度,例如插入区域的长度可以是约50-250、或约250-500、或约500-750、或约750-1000个碱基或碱基对。

[0560] 可以对包含插入区域的片段经历大小选择的过程,或者不对片段经历大小选择。例如,可以通过凝胶电泳和凝胶切片提取来选择片段的大小。可以使用固相粘附/固定方法来选择片段的大小,该方法通常采用涂有化学官能团的微顺磁珠,该化学官能团在有或没有聚乙二醇或聚亚烷基二醇的情况下在某些离子强度条件下与核酸相互作用。市售的固相粘附珠包括来自Beckman Coulter的SPRI(固相可逆固定)微珠(AMPUR XP顺磁珠,目录号B23318)、MAGNA PURE磁性玻璃颗粒(Roche Diagnostics,目录号03003990001)、来自Promega的MAGNASIL顺磁珠(目录号MD1360)、来自Precision System Science的MAGTRATION顺磁珠和系统(目录号A1120和A1060)、来自Omega Bio-Tek的MAG-BIND(目录号M1378-01)、来自Millapore的MAGPREP二氧化硅(目录号101193)、来自Bangs Laboratories的SNARE DNA纯化系统(目录号BP691、BP692和BP693)以及来自Perkin Elmer的CHEMAGEN M-PVA珠(目录号CMG-200)。

[0561] 插入区域可以使用连接酶和/或引物延伸反应在一端或两端接合至至少一个通用

衔接子序列。插入区域和通用衔接子之间的共价链接可以通过DNA或RNA连接酶实现。可以连接双链DNA分子的示例性DNA连接酶包括T4 DNA连接酶和T7 DNA连接酶。可以通过使用有尾引物的PCR将通用衔接子序列附加至插入序列,该有尾引物具有携带通用衔接子序列的5'区域和与插入序列的一部分互补的3'区域。可以通过PCR,使用具有携带第三通用衔接子序列的5'区域和与第一通用衔接子序列或第二通用衔接子序列的一部分互补的3'区域的有尾引物,将通用衔接子序列附加至一侧上或两侧上侧接第一衔接子序列和第二衔接子序列的插入序列。

[0562] 在一些实施例中,可以通过采用连接反应和引物延伸反应来产生文库分子(100)。文库分子可通过将双链插入区域(110)的第一端接合至具有第二左通用衔接子序列(140)的第一双链衔接子,并将双链插入区域(110)接合至具有第二右通用衔接子序列(150)的第二双链衔接子,其中使用DNA连接酶进行接合以产生双链重组分子。在一些实施例中,第一双链衔接子进一步包含左样品索引序列(160)。在一些实施例中,第二双链衔接子进一步包含右样品索引序列(170)。

[0563] 在一些实施例中,可以通过采用连接反应和引物延伸反应来产生文库分子(100)。文库分子可以通过将双链插入区域(110)的第一端和第二端接合至双链Y形衔接子(例如,分叉衔接子)来产生,该双链Y形衔接子具有退火的部分和错配的另一部分。Y形衔接子可包括形成平端或悬垂端(例如,5'或3'悬垂端)的退火部分(例如,参见图10A-图10D)。Y形衔接子的退火部分和/或错配部分可包括第二左通用衔接子序列(140)的至少一部分(或其互补序列)和第二右通用衔接子序列(150)(或其互补序列)的至少一部分(例如,图10A-图10D、图11A-图11D和图12A-图12C)。在一些实施例中,Y形衔接子的退火部分和/或错配部分可进一步包括左样品索引序列(160)。在一些实施例中,Y形衔接子的退火部分和/或错配部分可进一步包括右样品索引序列(170)(例如,图11A-图11D和图12A-图12C)。可以使用DNA连接酶将双链插入区域(110)接合至双链Y形衔接子以产生双链重组分子。

[0564] 在一些实施例中,可以使双链重组分子经历变性条件以产生单链重组分子。通过使用有尾引物(例如,有尾PCR引物)进行引物延伸反应,将单链重组分子与第一有尾引物接触/杂交,该第一有尾引物具有携带第一右通用衔接子序列(130)的5'区域、包含右样品索引序列(170)的内部区域以及与第二右通用衔接子序列(150)的至少一部分互补的3'区域。该方法进一步包括进行引物延伸反应以产生第一双链有尾分子,其包含:第一右通用衔接子序列(130);第一右样品索引(170);第二右通用衔接子序列(150);插入区域(110);和第二左通用衔接子序列(140)。参见例如图17。

[0565] 通过第一双链有尾分子与第二有尾引物接触/杂交,可以将另一通用衔接子序列附加至第一双链有尾分子,该第二有尾引物具有携带第一左通用衔接子序列(120)的5'区域、具有左样品索引序列(160)内部区域和与第二左通用衔接子序列(140)的至少一部分互补的3'区域。该方法进一步包括进行引物延伸反应以产生第二双链有尾分子,其包含第一左通用衔接子序列(120)、第一左样品索引(160)、第二左通用衔接子序列(140)、插入区域(110)、第二右通用衔接子序列(150)、第一右样品索引(170)和第一右通用衔接子序列(130)。参见例如图17。在一些实施例中,使用连接和引物延伸反应产生文库分子,不进行PCR扩增以减少扩增偏差。

[0566] 使用Y形衔接子制备文库

[0567] 本公开提供了使用Y形衔接子制备核酸文库分子(100)的方法。制备的核酸文库分子可以与本文所述的任何单链夹板链杂交,以产生本文所述的任何文库-夹板复合物(300)和共价闭合环状文库分子(400)。共价闭合环状文库分子可以分配在支持物上并经历滚环扩增反应以产生多个可以测序的固定的多联体。替代性地,可以将制备的核酸文库分子固定至具有多个表面引物的支持物,并且可以经由桥式扩增来扩增固定的文库分子(例如,固定的克隆扩增的文库分子),并且可以对扩增的产物进行测序。

[0568] 文库制备方法包括步骤(a):提供多个双链核酸Y形衔接子,其中各个双链衔接子包含杂交在一起的第一寡核苷酸链和第二寡核苷酸链,其中第一寡核苷酸链和第二寡核苷酸链均包含互补区域和错配区域,由此形成具有双链退火部分和具有两条单链的错配部分的Y形衔接子(例如,图10A-图10D、11A-图11D和图12A-图12D;顶链是第一个寡核苷酸,以及底链是第二个寡核苷酸)。

[0569] 在一些实施例中,形成Y形衔接子的第一寡核苷酸和/或第二寡核苷酸链的5'端可以被磷酸化。在一些实施例中,第一寡核苷酸链包含第一通用衔接子序列,该第一通用衔接子序列包含针对第一测序引物(1st SeqP; (150))的结合序列或其互补序列。在一些实施例中,第二寡核苷酸链包含第二通用衔接子序列,该第二通用衔接子序列包含针对第二测序引物(2nd SeqP; (140))的结合序列或其互补序列。参见例如图10A-图10D、图11A-图11D和图12A-图12C。在一些实施例中,在Y形衔接子的群体中,形成Y形衔接子的第一寡核苷酸链中的每一个都具有相同的序列。在一些实施例中,在Y形衔接子的群体中,形成Y形衔接子的第二寡核苷酸链中的每一个都具有相同的序列。

[0570] 在一些实施例中,Y形衔接子的双链退火区域包括至少4个连续碱基配对的核苷酸。在一些实施例中,双链退火区域包括可以经由酶促连接反应接合至具有目的序列的核酸片段的末端。在一些实施例中,双链退火区域包括平端末端,或者末端可以具有5'或3'悬垂区域(例如,参见图10A-图10C)。在一些实施例中,错配部分的第一寡核苷酸链和第二寡核苷酸链可以是相同的长度或不同的长度(例如,参见图10D)。

[0571] 在一些实施例中,Y形衔接子的双链退火区域的第一链包含针对第一测序引物(例如,反向测序引物或正向测序引物)(1st SeqP; (150))的结合序列的至少一部分或其互补序列。在一些实施例中,Y形衔接子的双链退火区域的第二链包含针对第二测序引物(例如,正向测序引物或反向测序引物)(2nd SeqP; (140))的结合序列的至少一部分或其互补序列。在一些实施例中,针对第一链的第一测序引物(1st SeqP; (150))的结合序列包括序列5'-TGTCGGAAGGTGTGCAGGCTACCGCTTGCTCAACT-3'(SEQ ID NO:211)。在一些实施例中,针对第二链的第二测序引物(2nd SeqP; (140))的结合序列包括序列5'-CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:212)。

[0572] 在一些实施例中,错配区域的第一链包括针对第一测序引物(例如,反向测序引物或正向测序引物)(1st SeqP; (150))的结合序列的至少一部分或其互补序列。在一些实施例中,错配区域的第一链进一步包含针对第一表面引物(1st SurfP; (130))的结合序列的至少一部分或其互补序列(例如,图11A-图11D)。在一些实施例中,错配区域的第一链包含通用衔接子序列,该通用衔接子序列包括第一样品索引序列(1st index; (170))(例如,图11A和图11D)。

[0573] 在一些实施例中,错配区域的第二链包括针对第二测序引物(例如,正向测序引物

或反向测序引物) (2nd SeqP; (140)) 的结合序列的至少一部分或其互补序列。在一些实施例中, 错配区域的第二链进一步包含针对第二表面引物 (2nd SurfP; (120)) 的结合序列的至少一部分或其互补序列 (例如, 图11A-图11D)。在一些实施例中, 错配区域的第二链进一步包含通用衔接子序列, 该通用衔接子序列包含第二样品索引序列 (2nd index; (160)) (例如, 图11C和图11D)。

[0574] 在一些实施例中, 仅错配区域的第一链包含通用衔接子, 其包含第一样品索引序列 (例如, 图11B)。在一些实施例中, 仅错配区域的第二链包含通用衔接子, 其包含第二样品索引序列 (例如, 图11C)。在一些实施例中, 错配区域的第一链包含第一样品索引序列, 并且错配区域的第二链包含第二样品索引序列 (例如, 图11D)。

[0575] 该文库制备方法进一步包括步骤 (b): 提供多个核酸片段。多个核酸片段可以具有相同的序列或具有不同的序列。多个核酸片段包含双链核酸分子。在一些实施例中, 可以通过将来自至少一个多核苷酸样品的靶多核苷酸 (例如, 基因组DNA) 片段化来产生多个核酸片段。可以使用机械力、酶促 (例如, 限制性核酸内切酶) 或化学片段化方法来片段化靶多核苷酸。替代性地, 多个核酸片段包含使用逆转录酶和DNA聚合酶从RNA产生的双链cDNA。在另一实施例中, 可以使用模板多核苷酸和一对PCR引物经由PCR产生多个核酸片段。在另一实施例中, 靶多核苷酸可以与酶混合物反应, 例如产生单链缺口的酶和催化双链切割的另一种酶。示例性的酶混合物为FRAGMENTASE (例如, 来自New England Biolabs)。在又一个实施例中, 多个核酸片段包含尚未经历片段化程序的循环细胞游离DNA (例如, 双链cfDNA)。细胞游离DNA的长度可为50-200bp。核酸片段可以为尺寸选择的或缺少尺寸选择。技术人员将认识到, 可以使用这些方法中的任一种来产生核酸片段, 并且片段的长度可以为50-1000bp或大于1000bp。核酸片段可以包括具有5'悬垂端和/或3'悬垂端的片段的异质混合物。

[0576] 该文库制备方法进一步包括步骤 (c): 在适合产生具有平端5'磷酸化端的核酸片段的条件下, 使步骤 (b) 的多个核酸片段与多种酶接触。在一些实施例中, 多种酶产生在其3'端处具有非模板A尾的平端片段。多种酶包括两种或更多种可以催化核酸末端修复、磷酸化和/或A加尾的酶。末端修复酶包括DNA聚合酶 (例如, T4 DNA聚合酶) 和Klenow片段。5'端磷酸化酶包括T4多核苷酸激酶。A加尾酶包括Taq聚合酶 (例如, 非校对聚合酶) 和dATP。在一些实施例中, 片段化、末端修复、磷酸化和A加尾可以使用酶的混合物在一锅反应中进行。

[0577] 文库制备方法进一步包括步骤 (d): 在适合于将Y形衔接子连接至各个核酸片段的两端, 以产生多个衔接子-插入物-衔接子文库分子的条件下, 使步骤 (a) 的多个双链核酸Y形衔接子和步骤 (c) 的多个核酸片段接触, 其中衔接子-插入物-衔接子文库分子中的衔接子包含双链退火区域和错配区域 (例如, 图13-图17)。在一些实施例中, 连接酶包括T4 DNA连接酶。

[0578] 文库制备方法进一步包括步骤 (e): 通过在适合于使各个第一有尾PCR引物的至少一部分与各个衔接子-插入物-衔接子文库分子的错配区域杂交的条件下, 并且条件适合于进行第一引物延伸反应, 通过使步骤 (d) 的多个衔接子-插入物-衔接子文库分子与第一多个有尾PCR引物接触来进行第一引物延伸反应。(例如, 图15-图17)。在一些实施例中, 各个第一有尾PCR引物包含 (i) 5'区域, 其具有能够与第一表面引物 (1st SurfP; (130)) 杂交的通用衔接子序列, 其中第一有尾PCR引物的5'区域不与衔接子-插入物-衔接子文库分子杂交, 以及 (ii) 3'区域, 其具有包含针对第一测序引物 (1st SeqP; (150)) 的结合序列的通用

衔接子序列或其互补序列,其中有尾引物的3'区域退火至衔接子-插入物-衔接子文库分子的连接的Y形衔接子的错配区域的第一寡核苷酸链(例如,图15)。在一些实施例中,第一多个有尾PCR引物进一步包含(iii)通用衔接子序列,其包括第一样品索引序列(1st index; (170)) (例如,图16)。第一个样品索引序列(1st index; (170))可用于区分多重测定中从不同样品源获得的目的序列(例如,插入序列)。在一些实施例中,第一多个有尾PCR引物具有相同的序列(例如,单个通用第一引物种类)。在一些实施例中,第一引物延伸反应可以使用多个聚合酶、多个核苷酸和与各个衔接子-插入物-衔接子文库分子的一部分杂交的第一多个有尾PCR引物进行。第一引物延伸产物包含:针对第一表面引物(1st SurfP; (130))的结合序列;任选的第一样品索引序列(1st index; (170))的结合序列;针对第一测序引物(1st SeqP; (150))的结合序列;插入序列; (110)和针对第二测序引物(2nd SeqP; (140))的结合序列(例如,图15-图17)。

[0579] 在一些实施例中,步骤(e)的第一引物延伸反应可以使用第一多个有尾PCR引物进行,该第一多个有尾PCR引物包括(i)5'区域,其具有能够与第一表面引物(1st SurfP; (130))杂交的通用衔接子序列,(ii)第一样品索引序列(1st index; (170)),以及(iii)3'区域,其具有包括针对第一测序引物(1st SeqP; (150))的结合序列的通用衔接子序列或其互补序列。参见例如图16和图17。第一多个有尾PCR引物中的有尾PCR引物各自包含第一样品索引序列(1st index; (170)),该第一样品索引序列包括直接连接到通用样品索引序列的短随机序列(例如,NNN)(图17),或通用样品索引序列但缺少短随机序列(例如,NNN)(图16)。通用样品索引序列可用于区分多重测定中从不同样品源获得的目的序列(例如,插入序列)。在短随机序列中,给定位置处的每个碱基“N”独立地选自A、G、C、T或U。短随机序列在测序运行的每个循环中提供核苷酸多样性和颜色平衡。图9A-图9F中表1提供了具有短随机序列(例如,NNN)和通用样品索引序列的示例性第一样品序列的列表。在样品索引文库分子的群体中,样品索引(例如,1st index; (170))的短随机序列提供了高核苷酸多样性,其包括将在测序运行的每个循环中表示的所有四种核苷酸(例如,A、G、C、T和/或U)的比例大致相等。短随机序列的高核苷酸多样性也在测序运行的每个循环期间提供颜色平衡。在第一多个有尾PCR引物中,其中各个引物包括第一样品索引序列(1st index; (170))具有短随机序列(例如,NNN),例如NNGTAGGAGCC (SEQ ID NO:97),第一多个有尾PCR引物含有引物分子的混合物,该引物分子各自携带同一通用样品索引序列(例如,GTAGGAGCC)和不同的短随机序列(例如,NNN),其中第一多个有尾PCR引物中可以存在多达64个不同的短随机序列。在一些实施例中,第一有尾引物包括序列5'-GATCAGGTGAGGCTGCGACGACTNNNNNNNNNN

[0580] AGTTGACAAGCGGTAGCCTGCACACCTCCGACAT-3' (SEQ ID NO:213),其中12碱基N序列代表右样品索引序列(170)。在一些实施例中,12碱基N序列中的前九个“N”包括右样品索引的通用样品索引序列。在一些实施例中,12碱基N序列中的最后三个“NNN”(下划线)包含3聚体随机序列,其被设计为在每个测序循环中提供核苷酸多样性和颜色平衡。

[0581] 文库制备方法进一步包括步骤(f):在适合于使各个第二有尾PCR引物的至少一部分与各个第一引物延伸产物的一部分杂交的条件下,并且条件适合于进行第二引物延伸反应,通过使步骤(e)的第一引物延伸产物与第二多个有尾PCR引物接触来进行第二引物延伸反应。(例如,图15-图17)。在一些实施例中,各个第二有尾PCR引物包含(i)5'区域,其具有能够与第二表面引物(2nd SurfP; (120))杂交的通用衔接子序列,其中第二有尾PCR引物的

5'区域不与第一引物延伸产物杂交,以及(ii)3'区域,其具有包含针对第二测序引物(2nd SeqP; (140))的结合序列的通用衔接子序列或其互补序列,其中3'端退火至针对第一引物延伸产物的第二测序引物(2nd SeqP; (140))的结合序列(例如,图15)。在一些实施例中,第二多个有尾PCR引物进一步包含(iii)包含第二样品索引序列的通用衔接子序列(例如,图16和图17)。第二样品索引序列(2nd index; (160))可用于区分多重测定中从不同样品源获得的目的序列(例如,插入序列)。在一些实施例中,第二多个有尾PCR引物具有相同的序列(例如,单个通用第二引物种类)。在一些实施例中,第二引物延伸反应可以使用多个聚合酶、多个核苷酸和与各个第一引物延伸产物的一部分杂交的第二多个有尾PCR引物进行。在一些实施例中,第二引物延伸产物包含以下组分:针对第二表面引物(2nd SurfP; (120))的结合序列;任选的第二样品索引序列;针对第二测序引物(2nd SeqP; (140))的结合序列;插入序列(110);针对第一测序引物(1st SeqP; (150))的结合序列;任选的第一样品索引序列;以及针对第一表面引物(1st SurfP; (130))的结合序列(例如,图15-图17)。在一些实施例中,第二有尾引物包括序列5'-CATGTAATGCACGTACTTTCAGGGTNNNNNNNNN CGTGCTGGATTG GCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:214),其中10碱基N序列代表左样品索引序列(160)。在一些实施例中,10碱基N序列包括左样品索引的通用样品索引序列。

[0582] 文库制备方法进一步包括步骤(g):在支持物上产生多个固定的模板分子,其中固定的模板分子包含步骤(f)的第二引物延伸产物的组分,其包含:针对第二表面引物(2nd SurfP; (120))的结合序列;任选的第二样品索引序列;针对第二测序引物(2nd SeqP; (140))的结合序列;插入序列(110);针对第一测序引物(1st SeqP; (150))的结合序列;任选的第一样品索引序列;以及针对第一表面引物(1st SurfP; (130))的结合序列。

[0583] 在一些实施例中,可以使用桥式扩增反应来产生多个固定的模板分子(例如,成簇模板分子或克隆扩增模板分子)。步骤(f)的第二引物延伸产物为线性文库分子,其可以分配到具有多个两种类型的固定的表面引物的支持物上,其中一种类型的表面引物捕获线性文库分子。捕获的线性分子可以经历桥式扩增反应,该反应采用第一类型和第二类型的固定的表面引物来产生彼此紧密接近的第一模板分子和第二模板分子。第一模板分子和第二模板分子彼此互补并且包含步骤(f)的第二引物延伸产物的组分。第一模板分子和第二模板分子包括克隆扩增的模板分子。第一模板分子和第二模板分子包含目的序列和至少一个通用测序引物结合位点。第一模板分子和第二模板分子包含以下成分:针对第二表面引物(2nd SurfP; (120))的结合序列(或其互补序列);任选的第二样品索引序列;针对第二测序引物(2nd SeqP; (140))的结合序列(或其互补序列);插入序列(110);针对第一测序引物(1st SeqP; (150))的结合序列(或其互补序列);任选的第一样品索引序列;以及针对第一表面引物(1st SurfP; (130))的结合序列(或其互补序列)。

[0584] 在一些实施例中,可以通过使用本文所述的任何单链夹板链成环步骤(f)的第二引物延伸产物,并进行本文所述的用于形成文库-夹板复合物(300)的任何方法来产生多个固定的模板分子,以及本文所述的用于形成共价闭合环状分子(400)的任何方法。共价闭合环状分子(400)可经历本文所述的任何滚环扩增(RCA)反应,包括支持物上RCA、溶液内RCA、或使用单链夹板链引发的溶液内RCA滚环扩增反应。滚环扩增反应可用于产生固定至支持物的多个多联体。固定的多联体包含步骤(f)的第二引物延伸产物的组分的多个串联拷贝,其包括:针对第二表面引物(2nd SurfP; (120))的结合序列(或其互补序列);任选的第二样

品索引序列;针对第二测序引物(2nd SeqP; (140))的结合序列(或其互补序列);插入序列(110);针对第一测序引物(1st SeqP; (150))的结合序列(或其互补序列);任选的第一样品索引序列;以及针对第一表面引物(1st SurfP; (130))的结合序列(或其互补序列)。各个固定的多联体包含克隆扩增的模板分子。固定的多联体可以自行塌陷成紧凑的核酸纳米球。在任何RCA反应期间包括一个或多个压实寡核苷酸可进一步压缩纳米球的尺寸和/或形状。给定多联体中串联重复单元数量的增加,增加了多联体周围用于与多个测序引物(例如,具有通用序列的测序引物)杂交的位点数量的位点的数量,该测序引物用作聚合酶催化的测序反应的多个起始位点。

[0585] 文库制备方法进一步包括步骤(h):对来自步骤(g)的多个固定的模板分子的至少一部分进行测序。测序反应包括本文所述的任何测序方法,包括采用可检测地标记的核苷酸类似物的测序反应,或采用包括结合可检测地标记的多价分子并掺入核苷酸类似物的两阶段测序反应的测序反应,或采用非可检测地标记的核苷酸类似物的测序反应(例如,结合测序)。

[0586] 可以根据本文所述的任何工作流程以任何顺序对固定的模板分子的不同组分进行测序。技术人员将认识到,固定的模板分子的不同组分可以按照本文未描述的其他顺序进行测序。

[0587] 在一些实施例中,测序顺序包括:(1)对第一样品索引序列(170)进行测序(如果存在,使用短随机序列NNN);(2)对第二样品索引序列(160)进行测序;以及(3)对部分或全长插入区域(110)进行测序。

[0588] 在一些实施例中,测序顺序包括:(1)对第一样品索引序列(170)进行测序(如果存在,使用短随机序列NNN);(2)对部分或全长插入区域(110)进行测序;以及(3)对第二样品索引序列进行测序(160)。

[0589] 在一些实施例中,测序顺序包括:(1)对部分或全长插入区域(110)进行测序;(2)对第一样品索引序列(170)进行测序(如果存在,使用短随机序列NNN);以及(3)对第二样品索引序列进行测序(160)。

[0590] 在一些实施例中,测序顺序包括:(1)对插入区域(110)的3-5个碱基进行测序;(2)对第一样品索引序列(170)进行测序(如果存在,使用短随机序列NNN);(3)对第二样品索引序列(160)进行测序;以及(4)对全长插入区域(110)进行测序。

[0591] 在一些实施例中,测序顺序包括:(1)对插入区域(110)的3-5个碱基进行测序;(2)对第一样品索引序列(170)进行测序(如果存在,使用短随机序列NNN);(3)对第二样品索引序列(160)进行测序;(4)进行成对反应(pairwise reaction),使固定的模板分子被与模板分子互补的固定的链所取代;以及(5)对插入区域(110)的全长进行测序。

[0592] 在对多个固定的模板分子中的第一样品索引的短随机序列(例如,NNN)进行测序时,短随机序列提供高核苷酸多样性,其中包括将在测序运行的每个循环中表示的所有四种核苷酸(例如,A、G、C、T和/或U)的大约相等比例。短随机序列的高核苷酸多样性也在测序运行的每个循环期间提供颜色平衡。在一些实施例中,仅来自第一样品索引序列的短随机序列的测序数据可用于聚合酶克隆作图和模板配准,因为短随机序列提供了足够的核苷酸多样性和颜色平衡。来自通用样品索引序列的测序数据可用于在多重分析中区分从不同样品来源获得的目的序列。

[0593] 该文库制备方法进一步包括步骤(i):将插入区域的序列指定到第一样品索引序列,由此将插入区域识别为从第一来源获得。在一些实施例中,用于测序的方法进一步包括:将插入区域的序列指定到第一样品索引序列和第二样品索引序列,由此将插入区域识别为从第一来源获得。

[0594] 具有低非特异性结合涂层的支持物

[0595] 本公开提供了使用其上固定有多个表面引物的支持物来制备本文所述的任何固定的多联体的组合物和方法。在一些实施例中,用低非特异性结合涂层使支持物钝化(例如,图18)。本文所述的表面涂层表现出与通常用于核酸捕获、扩增和测序工作流程的试剂(诸如染料、核苷酸、酶和核酸引物)的非常低非特异性结合。表面涂层表现出与常规表面涂层相比的低背景荧光信号或高对比度噪声(CNR)比。

[0596] 一般而言,支持物包含基底(或支持结构)、一层或更多层共价或非共价附接的低结合化学改性层,例如硅烷层、聚合物膜,以及一层或更多层共价或非共价附接的引物序列,其可用于将单链靶核酸栓系至支持物表面。在一些实施例中,表面的配方,例如一层或多层的化学组成、用于将一层或多层与支持物表面和/或彼此交联的偶联化学,以及层的总数目,可以变化,使得蛋白质、核酸分子以及其他杂交和扩增反应组分与支持物表面的非特异性结合相对于可比较的单层被最小化或减少。通常,表面的配方可以变化,使得支持物表面上的非特异性杂交相对于可比较的单层被最小化或减少。表面的配方可以变化,使得支持物表面上的非特异性扩增相对于可比较的单层被最小化或减少。表面的配方可以变化,使得支持物表面上的特定扩增速率和/或产量最大化。在本文所公开的一些情况下,适用于检测的扩增水平在不超过2个、3个、4个、5个、6个、7个、8个、9个、10个、15个、20个、25个、30个或超过30个扩增循环中被实现。

[0597] 包含一个或多个化学改性层(例如低非特异性结合聚合物层)的基底或支持结构可以是独立的或集成到另一结构或组件中。例如,在一些实施例中,基底或支持结构可包括集成或组装的微流体流通池内的一个或多个表面。基底或支持结构可包括微板形式内的一个或多个表面,例如微板中的孔的底表面。如上所述,在一些优选实施例中,基底或支持结构包括毛细管的内表面(诸如内腔表面)。在替代的优选实施例中,基底或支持结构包括蚀刻到平面芯片中的毛细管的内表面(诸如内腔表面)。

[0598] 用于将第一化学改性层接枝至表面的附接化学通常取决于制造表面的材料和层的化学性质两者。在一些实施例中,第一层可共价附接至表面。在一些实施例中,第一层可以非共价附接,例如通过表面与第一层的分子组分之间的非共价相互作用(诸如静电相互作用、氢键或范德华相互作用)吸附至表面。在任一情况下,可以在附接或沉积第一层之前处理基底表面。本领域技术人员已知的多种表面制备技术中的任一者可用于清洁或处理表面。例如,可使用Piranha溶液(硫酸(H_2SO_4)和过氧化氢(H_2O_2)的混合物)对玻璃或硅表面进行酸洗,在KOH和NaOH中进行碱处理,和/或使用氧等离子体处理方法进行清洁。

[0599] 硅烷化学构成了一种非限制性方法,用于共价修饰玻璃或硅表面上的硅烷醇基团以附接更多反应性官能团(例如,胺或羧基),然后可用于将接头分子(例如,各种长度的直链碳氢化合物分子,诸如C6、C12、C18碳氢化合物或直链聚乙二醇(PEG)分子)或层分子(例如,支化PEG分子或其他聚合物)偶联到表面。可用于产生任何公开的低结合表面的合适的硅烷的实例包括但不限于:(3-氨基丙基)三甲氧基硅烷(APTMS)、(3-氨基丙基)三乙氧基

硅烷 (APTES)、PEG-硅烷 (例如, 包含1K、2K、5K、10K、20K等的分子量)、氨基-PEG硅烷 (即, 包含游离氨基官能团)、马来酰亚胺-PEG硅烷、生物素-PEG硅烷等。

[0600] 本领域技术人员已知的多种分子中的任意者, 包括但不限于: 氨基酸、肽、核苷酸、寡核苷酸、其他单体或聚合物、或其组合, 可用于在表面上产生一种或更多种化学修饰的分子层, 其中可以改变所使用的组分的选择以改变表面的一种或更多种性质, 例如官能团和/或拴系的寡核苷酸引物的表面密度、表面的亲水性/疏水性或这三种表面的三维性质 (即“厚度”)。可用于在任何所公开的表面中产生一层或多层低非特异性结合材料的优选聚合物的实例包括但不限于: 各种分子量和分支结构的聚乙二醇 (PEG)、链霉亲和素、聚丙烯酰胺、聚酯、葡聚糖、聚赖氨酸和聚赖氨酸共聚物, 或其任意组合。可用于将材料的一个或多个层 (例如聚合物层) 接枝至表面和/或使层彼此交联的缀合化学的实例包括但不限于: 生物素-链霉亲和素相互作用 (或它们的变体)、his标签-Ni/NTA缀合化学、甲氧基醚缀合化学、羧酸酯缀合化学、胺缀合化学、NHS酯、马来酰亚胺、硫醇、环氧树脂、叠氮化物、酰肼、炔烃、异氰酸酯和硅烷。

[0601] 低非特异性结合表面涂层可以均匀地施加在基底上。或者, 表面涂层可被图案化, 使得化学改性层被限制于基材的一个或多个离散区域。例如, 可以使用光刻技术对表面进行图案化, 以在表面上产生化学改性区域的有序阵列或随机图案。替代地或组合地, 可以使用例如接触印刷和/或喷墨印刷技术对基底表面进行图案化。在一些实施例中, 经化学修饰的区域的有序阵列或随机图案可包括至少1个、5个、10个、20个、30个、40个、50个、60个、70个、80个、90个、100个、200个、300个、400个、500个、600个、700个、800个、900个、1000个、2000个、3000个、4000个、5000个、6000个、7000个、8000个、9000个或10,000个或更多个离散区域。

[0602] 为了实现低非特异性结合表面, 亲水性聚合物可以非特异性吸附或共价接枝至表面。通常, 利用聚 (乙二醇) (PEG, 也称为聚环氧乙烷 (PEO) 或聚氧乙烯) 或具有不同分子量和端基团的其他亲水性聚合物来进行钝化, 这些聚合物使用例如硅烷化学链接到表面。远离表面的端基团可包括但不限于: 生物素、甲氧基醚、羧酸酯、胺、NHS酯、马来酰亚胺和双硅烷。在一些实施例中, 亲水性聚合物 (例如, 直链聚合物、支化聚合物或多支化聚合物) 的两个或更多个层可被沉积到表面上。在一些实施例中, 两个或更多个层可以彼此共价偶联或内部交联以改善所得表面的稳定性。在一些实施例中, 具有不同碱基序列和碱基修饰的寡核苷酸引物 (或其他生物分子, 例如酶或抗体) 可以各种表面密度拴系至所得表面层。在一些实施例中, 例如, 可以改变表面官能团密度和寡核苷酸浓度两者以达成特定的引物密度范围。另外, 可以通过用带有相同官能团的其他分子稀释寡核苷酸来控制引物密度。例如, 胺标记的寡核苷酸可以在与NHS-酯涂覆的表面反应中用胺标记的聚乙二醇稀释, 以降低最终的引物密度。杂交区域与表面附接官能团之间具有不同长度的接头的引物也可用于控制表面密度。合适的接头的实例包括: 在引物的5'端处的聚T和聚A链 (例如, 0至20个碱基)、PEG接头 (例如, 3至20个单体单元) 和碳链 (例如, C6、C12、C18等)。为了测量引物密度, 可将经以荧光方式标记的引物拴系至表面, 并且随后将荧光读段与针对已知浓度的染料溶液的荧光读段进行比较。

[0603] 为了缩放引物表面密度并向亲水性或两性表面添加额外的维度, 已经开发了包含PEG和其他亲水性聚合物的多层涂层的表面。通过使用包括但不限于下述聚合物/共聚物材

料的亲水性和两性表面分层方法,可以显著增加表面上的引物负载密度。传统PEG涂层方法使用单层引物沉积,该方法已通常被报告用于单分子应用,但对于核酸扩增应用来说不产生高拷贝数。如本文所述,可使用传统交联方法与任何相容的聚合物或单体亚基来实现“分层”,使得可循序地构建包括两个或更多个高度交联的层的表面。合适的聚合物的实例包括但不限于:链霉亲和素、聚丙烯酰胺、聚酯、葡聚糖、聚赖氨酸、以及聚赖氨酸和PEG的共聚物。在一些实施例中,不同的层可通过多种缀合反应中的任一者彼此附接,该多种缀合反应包括但不限于:生物素-链霉亲和素结合、叠氮化物-炔烃点击反应、胺-NHS酯反应、硫醇-马来酰亚胺反应、以及带正电荷的聚合物与带负电荷的聚合物之间的离子相互作用。在一些实施例中,高引物密度材料可在溶液中被构造并且随后在多个步骤中被分层放置到表面上。

[0604] 如所指出的,本公开的低非特异性结合涂层表现出蛋白质、核酸和用于固相核酸扩增的杂交和/或扩增制剂的其他组分的减少的非特异性结合。可以定性方式或以定量方式来评估给定支持物表面所表现出的非特异性结合的程度。例如,在一些实施例中,在标准化设置条件下,将表面暴露于荧光染料(例如,花青染料诸如Cy3或Cy5等,荧光素、香豆素、罗丹明等或本文公开的其他染料)、荧光标记的核苷酸、荧光标记的寡核苷酸和/或荧光标记的蛋白质(例如聚合酶),然后进行指定的冲洗方案和荧光成像,可以用作定性工具,用于比较包含不同表面配方的支持物上的非特异性结合。在一些实施例中,在标准化设置条件下,将表面暴露于荧光染料、荧光标记的核苷酸、荧光标记的寡核苷酸和/或荧光标记的蛋白质(例如聚合酶),然后进行指定的冲洗方案和荧光成像,可以用作定量工具,用于比较包含不同表面配方的支持物上的非特异性结合——前提是已注意确保,荧光成像是在荧光信号与支持表面上的荧光团的数量(例如,在荧光团的信号饱和和/或自猝灭不成问题的条件下)线性相关(或以可预测的方式相关)的条件下进行的,并使用合适的校准标准。在一些实施例中,本领域技术人员已知的其他技术(例如,放射性同位素标记和计数方法)可用于定量评估本公开的不同支持物表面配方所表现出的非特异性结合的程度。

[0605] 本文所公开的一些表面表现出至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、50、75、100、或大于100、或本文中的范围所涵盖的任何中间值的荧光团(诸如Cy3)的特异性与非特异性结合的比率。本文所公开的一些表面表现出至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、50、75、100、或大于100、或本文中的范围所涵盖的任何中间值的荧光团(诸如Cy3)的特异性与非特异性荧光的比率。

[0606] 如所指出的,在一些实施例中,所公开的低结合支持物表现出的非特异性结合程度,可以使用标准化的实验方案进行评估:使表面与标记蛋白质(例如,牛血清白蛋白(BSA)、链霉亲和素、DNA聚合酶、逆转录酶、解旋酶、单链结合蛋白(SSB)等,或其任意组合)、标记的核苷酸、标记的寡核苷酸等接触,在孵育和漂洗的标准化设置条件下,随后检测保留在表面上的标记的量,并将由此产生的信号与适当的校准标准进行比较。在一些实施例中,标记可包括荧光标记。在一些实施例中,标记可包括放射性同位素。在一些实施例中,标记可包括本领域技术人员已知的任何其他可检测标记。在一些实施例中,因此可以从每单位面积非特异性结合的蛋白质分子(或其他分子)的数量方面来评估由给定支持物表面配方所表现出的非特异性结合程度。在一些实施例中,本公开的低结合支持物可以表现出非特

异性蛋白质结合(或其他指定分子(例如,花青染料诸如Cy3或Cy5等、荧光素、香豆素、罗丹明等或本文公开的其他染料)的非特异性结合)为:少于0.001个分子/ μm^2 、少于0.01个分子/ μm^2 、少于0.1个分子/ μm^2 、少于0.25个分子/ μm^2 、少于0.5个分子/ μm^2 、少于1个分子/ μm^2 、少于10个分子/ μm^2 、少于100分子/ μm^2 、或少于1,000个分子/ μm^2 。本领域技术人员将认识到,本公开的给定支持表面可以表现出落在该范围内的任何位置的非特异性结合,例如少于86个分子/ μm^2 。例如,本文公开的一些修饰表面在与磷酸盐缓冲盐水(PBS)缓冲液中的1 μM Cy3标记的链霉亲和素(GE Amersham)溶液接触15分钟,然后用去离子水漂洗3次后,表现出小于0.5个分子/ μm^2 的非特异性蛋白质结合。本文公开的一些改性表面表现出Cy3染料分子的非特异性结合小于0.25个分子/ μm^2 。在独立的非特异性结合测定中,1 μM 标记的Cy3 SA (ThermoFisher)、1 μM Cy5 SA染料(ThermoFisher)、10 μM 氨基烯丙基-dUTP-ATTO-647N (Jena Biosciences)、10 μM 氨基烯丙基-dUTP-ATTO-Rho11 (Jena Biosciences)、10 μM 氨基烯丙基-dUTP-ATTO-Rho11 (Jena Biosciences)、10 μM 7-丙炔氨基-7-脱氮-dGTP-Cy5 (Jena Biosciences)和10 μM 7-丙炔氨基-7-脱氮-dGTP-Cy3 (Jena Biosciences)在384孔板格式中在低结合基底上于37 $^{\circ}\text{C}$ 孵育15分钟。每个孔用50 μl 去离子不含RNA酶/DNA酶的水冲洗2-3次、并用pH 7.4的25mMACES缓冲液冲洗2-3次。使用如由制造商指定的Cy3、AF555或Cy5滤光片组(根据所执行的染料测试)以800的PMT增益设定和50 μm 至100 μm 的分辨率在GE Typhoon仪器上对384孔板进行成像。为了获得更高分辨率的成像,在带有全内反射荧光(TIRF)物镜(100X,1.5NA,Olympus)、CCD相机(例如,Olympus EM-CCD单色相机、Olympus XM-10单色相机或Olympus DP80彩色和单色相机)、照明源(例如,Olympus 100W Hg灯、Olympus 75W Xe灯或Olympus U-HGLGPS荧光光源)的Olympus IX83显微镜(Olympus Corp.,Center Valley,PA)上收集图像,并且激发波长为532nm或635nm。二向色镜购自Semrock (IDEX Health&Science,LLC,Rochester,New York),例如405、488、532或633nm二向色反射器/分束器,并且带通滤光片选择为532LP或645LP,符合适当的激发波长。本文公开的一些改性表面表现出小于0.25个分子/ μm^2 的染料分子的非特异性结合。

[0607] 在一些实施例中,本文所公开的一些表面表现出荧光团(诸如Cy3)的特异性与非特异性结合的比率为至少2:3、4:5、6:7、8:9、10:11、12:13、14:15、16:17、18:19、20:25、30:35、40:50、75:100、或大于100、或本文中的范围所涵盖的任何中间值。在一些实施例中,本文所公开的一些表面表现出荧光团(诸如Cy3)的特异性与非特异性荧光信号的比率为至少2:3、4:5、6:7、8:9、10:11、12:13、14:15、16:17、18:19、20:25、30:35、40:50、75:100、或大于100、或本文中的范围所涵盖的任何中间值。

[0608] 符合本文公开的低背景表面可表现出至少4:1、5:1、6:1、7:1、8:1、9:1、10:1、15:1、20:1、30:1、40:1、50:1、或超过每个非特异性地吸附的分子50个附接的特定染料分子的特异性染料附接(例如,Cy3附接)与非特异性染料吸附(例如,Cy3染料吸附)比率。类似地,当经历激发能量时,符合本文公开的已附接有荧光团(例如,Cy3)的低背景表面可表现出,相对于非特异性吸附的染料荧光信号至少4:1、5:1、6:1、7:1、8:1、9:1、10:1、15:1、20:1、30:1、40:1、50:1或超过50:1的特定荧光信号(例如,从附接至表面的经Cy3标记的寡核苷酸产生)比率。

[0609] 在一些实施例中,可例如通过测量水接触角来评估所公开的支持物表面的亲水性的程度(或与水性溶液的“可湿性”),其中将一小滴水放置在表面上,并且使用例如光学张

力计来测量其与表面的接触角。在一些实施例中,可确定静态接触角。在一些实施例中,可确定前进或后退接触角。在一些实施例中,本文所公开的亲水性低结合支持物表面的水接触角的范围可以为约0度至约30度。在一些实施例中,本文所公开的亲水性低结合支持物表面的水接触角可不超过50度、40度、30度、25度、20度、18度、16度、14度、12度、10度、8度、6度、4度、2度或1度。在许多情况下,接触角不超过40度。本领域技术人员将认识到,本公开的给定亲水性低结合支持物表面可表现出具有在该范围内的任何位置的值的水接触角。

[0610] 在一些实施例中,本文所公开的亲水性表面有助于减少针对生物测定的洗涤时间,这通常归功于生物分子与低结合表面的减少的非特异性结合。在一些实施例中,可在小于60秒、50秒、40秒、30秒、20秒、15秒、10秒或小于10秒内执行足够的洗涤步骤。例如,在一些实施例中,可以在少于30秒内执行足够的清洗步骤。

[0611] 本公开的低结合表面对于长时间暴露于溶剂和高温、或对于溶剂暴露或温度变化的重复循环的稳定性或耐久性表现出显著改善。例如,在一些实施例中,可以通过荧光标记表面上的官能团或表面上的拴系的生物分子(例如,寡核苷酸引物),并在长时间暴露于溶剂和高温,或重复暴露于溶剂或温度变化的循环之前、期间和之后监测荧光信号,来测试所公开的表面的稳定性。在一些实施例中,用于评估表面的质量的荧光中的变化的程度,可以为历经暴露于溶剂和/或高温1分钟、2分钟、3分钟、4分钟、5分钟、10分钟、20分钟、30分钟、40分钟、50分钟、60分钟、2小时、3小时、4小时、5小时、6小时、7小时、8小时、9小时、10小时、15小时、20小时、25小时、30小时、35小时、40小时、45小时、50小时或100小时的时间段后,小于1%、2%、3%、4%、5%、10%、15%、20%或25%(或历经这些时间段后所测量的这些百分比的任意组合)。在一些实施例中,用于评估表面的质量的荧光中的变化的程度,可以为历经重复暴露于溶剂变化和/或温度变化5个循环、10个循环、20个循环、30个循环、40个循环、50个循环、60个循环、70个循环、80个循环、90个循环、100个循环、200个循环、300个循环、400个循环、500个循环、600个循环、700个循环、800个循环、900个循环或1,000个循环后,小于1%、2%、3%、4%、5%、10%、15%、20%或25%(或历经该循环的范围后所测量的这些百分比的任意组合)。

[0612] 在一些实施例中,本文所公开的表面可表现出特定信号与非特定信号或其他背景的高比率。例如,当用于核酸扩增时,一些表面可表现出比表面的相邻无种群区域(unpopulated region)的信号大的扩增信号,倍数为至少4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、75倍、100倍或大于100倍。类似地,一些表面表现出比表面的相邻经扩增核酸群体区域的信号大的扩增信号,倍数为至少4倍、5倍、6倍、7倍、8倍、9倍、10倍、15倍、20倍、30倍、40倍、50倍、75倍、100倍或大于100倍。

[0613] 在一些实施例中,当用于核酸杂交或扩增应用以产生杂交或克隆扩增的核酸分子(例如,已经用荧光团直接或间接标记的)簇时,所公开的低背景表面的荧光图像表现出对比-噪声比(CNR)为至少10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200、210、220、230、240、250或大于250。

[0614] 一种或多种类型的引物(例如,捕获引物)可以附接或拴系至支持物表面。在一些实施例中,一种或更多种类型的衔接子或引物可包含间隔基序列、用于与衔接子连接的靶文库核酸序列杂交的衔接子序列、正向扩增引物、反向扩增引物、测序引物和/或分子条形码序列,或其任意组合。在一些实施例中,1个引物或衔接子序列可以拴系至表面的至少一

层。在一些实施例中,至少2、3、4、5、6、7、8、9、10或多于10个不同的引物或衔接子序列可拴系至表面的至少一层。

[0615] 在一些实施例中,拴系的衔接子和/或引物序列的长度范围可为约10个核苷酸至约100个核苷酸。在一些实施例中,拴系的衔接子和/或引物序列长度可为至少10、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、或至少100个核苷酸。在一些实施例中,拴系的衔接子和/或引物序列长度可为至多100、至多90、至多80、至多70、至多60、至多50、至多40、至多30、至多20、或至多10个核苷酸。本段落中描述的任何下限值和上限值可以组合以形成包括在本公开内的范围,例如,在一些实施例中,拴系的衔接子和/或引物序列的长度可以在约20个核苷酸至约80个核苷酸的范围内。本领域技术人员将认识到,拴系的衔接子和/或引物序列的长度可以具有该范围内的任何值,例如约24个核苷酸。

[0616] 在一些实施例中,本公开的低结合支持物表面上的引物的最终表面密度可以在约100个引物分子/ μm^2 至约100,000个引物分子/ μm^2 的范围内。在一些实施例中,本公开的低结合支持物表面上的引物的最终表面密度可以在约100,000个引物分子/ μm^2 至约 10^{15} 个引物分子/ μm^2 的范围内。在一些实施例中,引物的表面密度可为至少1,000、至少10,000、至少100,000或至少 10^{15} 个引物分子/ μm^2 。在一些实施例中,引物的表面密度可为至多10,000、至多100,000、至多1,000,000或至多 10^{15} 个引物分子/ μm^2 。本段落中描述的任何下限值和上限值可以组合以形成包括在本公开内的范围,例如,在一些实施例中,引物的表面密度可以在从约10,000个分子/ μm^2 至约 10^{15} 个分子/ μm^2 的范围内。本领域技术人员将认识到,引物分子的表面密度可以具有该范围内的任何值,例如约455,000个分子/ μm^2 。在一些实施例中,最初与支持物表面上的衔接子或引物序列杂交的目标文库核酸序列的表面密度可以小于或等于所指示的拴系的引物的表面密度。在一些实施例中,与支持物表面上的衔接子或引物序列杂交的克隆扩增的靶文库核酸序列的表面密度可以跨越与所指示的拴系的引物的表面密度相同的范围。

[0617] 上面列出的局部密度不排除整个表面上密度的变化,使得表面可以包含具有例如500,000/ μm^2 的寡核苷酸密度的区域,同时也包含具有显著不同的局部密度的至少第二区域。

[0618] 低非特异性结合涂层包括多层表面涂层的一层或多层,其可以包含支化聚合物或者可以是直链的。合适的支链聚合物的实例包括但不限于:支链PEG、支链聚(乙烯醇)(支链PVA)、支链聚(乙烯基吡啶)、支链聚(乙烯基吡咯烷酮)(支链PVP)、支链聚(丙烯酸)(支链PAA)、支链聚丙烯酰胺、支链聚(N-异丙基丙烯酰胺)(支链PNIPAM)、支链聚(甲基丙烯酸甲酯)(支链PMA)、支链聚(2-羟乙基甲基丙烯酸酯)(支链PHEMA)、支链聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(支链POEGMA)、支链聚谷氨酸(支链PGA)、支链聚赖氨酸、支链聚葡萄糖苷和葡聚糖。

[0619] 在一些实施例中,用于产生本文公开的多层表面中的任意的一个或多个层的支化聚合物可包含:至少4个支链、至少5个支链、至少6个支链、至少7个支链、至少8个支链、至少9个支链、至少10个支链、至少12个支链、至少14个支链、至少16个支链、至少18个支链、至少20个支链、至少22个支链、至少24个支链、至少26个支链、至少28个支链、至少30个支链、至少32个支链、至少34个支链、至少36个支链、至少38个支链或至少40个支链。

[0620] 用于形成本文公开的多层表面中的任一者的一个或多个层的线性、支链或多支链

聚合物可具有至少500、至少1,000、至少2,000、至少3,000、至少4,000、至少5,000、至少10,000、至少15,000、至少20,000、至少25,000、至少30,000、至少35,000、至少40,000、至少45,000或至少50,000道尔顿的分子量。

[0621] 在一些实施例中,例如,其中多层表面的至少一层包含支化聚合物,正在沉积的层的支化聚合物分子与前一层的分子之间的共价键的数量可以在每个分子约一个共价键至约32个共价键的范围内。在一些实施例中,新层的支链聚合物分子与前一层的分子之间的共价键的数量可以为每个分子至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少12、至少14、至少16、至少18、至少20、至少22、至少24、至少26、至少28、至少30或至少32个共价键。

[0622] 材料层与表面偶联后剩余的任何反应性官能团可任选地通过使用高收率偶联化学偶联小的惰性分子来阻断。例如,在使用胺偶联化学将新材料层附接至前一材料层的情况下,任何残留的胺基可随后通过与小氨基酸(诸如甘氨酸)偶联而被乙酰化或失活。

[0623] 沉积在表面上的低非特异性结合材料(例如,亲水性聚合物材料)的层的数量的范围可以为1至约10个。在一些实施例中,层的数量为至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9或至少10个。在一些实施例中,层的数量可以为最多10、最多9、最多8、最多7、最多6、最多5、最多4、最多3、最多2或最多1个。本段中所述的下限和上限值中的任意者可进行组合以形成包括在本公开内的范围,例如,在一些实施例中,层的数量的范围可以为约2至约4个。在一些实施例中,所有层可包含相同材料。在一些实施例中,每个层可包含不同材料。在一些实施例中,多个层可包含多种材料。在一些实施例中,至少一个层可包含支化聚合物。在一些实施例中,所有层可包含支化聚合物。

[0624] 在一些情况下,可使用极性质子溶剂、极性或极性非质子溶剂、非极性溶剂或它们的任意组合来将低非特异性结合材料的一个或多个层沉积在底物表面上和/或缀合至底物表面。在一些实施例中,用于层沉积和/或偶联的溶剂可包括:醇(例如,甲醇、乙醇、丙醇等)、另一种有机溶剂(例如,乙腈、二甲基亚砜(DMSO)、二甲基甲酰胺(DMF)等)、水、水性缓冲溶液(例如,磷酸盐缓冲液、磷酸盐缓冲盐水、3-(N-吗啉代)丙烷磺酸(MOPS)等)或它们的任意组合。在一些实施例中,所使用的溶剂混合物的有机组分可占总量的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、70%、75%、80%、85%、90%、95%、98%或99%,其中余量由水或水性缓冲溶液组成。在一些实施例中,所使用的溶剂混合物的水性组分可占总量的至少1%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、70%、75%、80%、85%、90%、95%、98%或99%,其中余量由有机溶剂组成。所使用的溶剂混合物的pH可小于6、约6、6.5、7、7.5、8、8.5、9或大于pH 9。

[0625] 荧光成像可以使用本领域技术人员已知的多种荧光团、荧光成像技术和荧光成像仪器中的任一种来进行。可以使用的合适的荧光染料的实例(例如,通过与核苷酸、寡核苷酸或蛋白质缀合)包括但不限于:荧光素、罗丹明、香豆素、花青及其衍生物,包括花青衍生物花青染料-3(Cy3)、花青染料-5(Cy5)、花青染料-7(Cy7)等。可以使用的荧光成像技术的实例包括但不限于荧光显微镜成像、荧光共焦成像、双光子荧光等。可以使用的荧光成像仪器的实例包括但不限于:配备有图像传感器或相机的荧光显微镜、共焦荧光显微镜、双光子荧光显微镜或包括适当选择的光源、透镜、镜子、棱镜、二向色反射器、光圈和图像传感器或相机的定制仪器。配备用于获取所公开的低结合支持物表面和其上杂交的模板核酸序列的

克隆扩增集落(克隆)的图像的荧光显微镜的非限制性实例是Olympus IX83倒置荧光显微镜,配备)20x、0.75NA、532nm光源,针对532nm长通激发和Cy3荧光发射滤光片优化的带通和二向色镜滤光片组、Semrock 532nm二向色反射器和相机(Andor sCMOS,Zyla 4.2),其中调节激发光强度以避免信号饱和。通常,在采集图像时,可以将支持物表面浸入缓冲液(例如,25mMACES、pH 7.4缓冲液)中。

[0626] 在一些情况下,使用所公开的反应试剂和低非特异性结合支持物的核酸杂交和/或扩增反应的性能可以使用荧光成像技术来评估,其中图像的对比度噪声比(CNR)提供评估扩增特异性和支持物上非特异性结合的关键指标。CNR通常定义为: $CNR = (\text{信号} - \text{背景}) / \text{噪声}$ 。背景项通常被认为是针对指定的目的区域(ROI)中的特定特征(衍射极限点(diffraction limited spot),DLS)周围的间隙区所测量的信号。虽然信噪比(SNR)通常被认为是总体信号质量的基准,但可显示,在需要快速图像捕获的应用中,相对于作为针对信号质量的基准的SNR,改善的CNR可提供显著优点(例如,对于其而言必须使循环时间最小化的测序应用),如下文实例中所示。本公开的表面也在共同未决的国际申请序列号PCT/US2019/061556中提供,该申请的全部内容通过引用并入本文。

[0627] 在大多数基于集成的测序方法中,背景项通常被测量为与“间质”区域相关的信号。除了“间质”背景($B_{\text{间质}}$)之外,扩增DNA集落占据的区域内存在“质内”背景($B_{\text{质内}}$)。这两个背景信号的组合决定了可实现的CNR,并随后直接影响光学仪器要求、架构成本、试剂成本、运行时间、成本/基因组,并最终影响基于循环阵列的测序应用的准确性和数据质量。 $B_{\text{间质}}$ 背景信号有多种来源;一些实例包括:来自可消耗流通池的自发荧光、生成虚假荧光信号的检测分子的非特异性吸附(其可能遮蔽来自ROI的信号)、非特异性DNA扩增产物(例如,从引物二聚体生成的那些)的存在。在典型的下一代测序(NGS)应用中,当前视场(FOV)中的该背景信号随时间推移被平均和减去。单个DNA菌落(即FOV中的(S) - $B_{\text{间质}}$)产生的信号生成可分类的可辨别特征。在某些情况下,质内背景($B_{\text{质内}}$)可能会产生混杂的荧光信号,该信号并非特定于目的靶标,而是存在于同一ROI中,由此使其更难以平均和减去。

[0628] 在本公开内容的低结合基底上实施核酸扩增可以通过减少非特异性结合来降低 $B_{\text{间质}}$ 背景信号,可以导致特异性核酸扩增的改善,并且可以导致非特异性扩增的减少,这会影晌间质和质内区域产生的背景信号。在一些情况下,所公开的低结合支持物表面,任选地与所公开的杂交缓冲液试剂组合使用,可以导致CNR比使用常规支持物和杂交、扩增和/或测序方案实现的那些提高2、5、10、100或1000倍。尽管本文在使用荧光成像作为读出或检测模式的背景下进行了描述,但是相同的原理也适合于将所公开的低非特异性结合支持物和核酸杂交和扩增试剂用于其他检测模式,包括光学和非光学检测模式。

[0629] 所公开的低结合支持物,任选地与所公开的杂交和/或扩增方案组合使用,产生固相反应,其表现出:(i)蛋白质和其他反应组分的可忽略的非特异性结合(由此最小化基地背景),(ii)可忽略的非特异性核酸扩增产物,以及(iii)提供可调节的核酸扩增反应。

[0630] 在一些实施例,当用于核酸杂交或扩增应用以产生经杂交或克隆扩增的核酸分子(例如,已直接地或间接地用荧光团标记)的群落时,所公开的低背景表面的荧光图像表现出至少10、20、30、40、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、20、210、220、230、240、250或大于250的对比度噪声比(CNR)。

[0631] 在一些实施例,当样品核酸分子或其互补序列用花青染料3(Cy3)荧光团标记

时,并且当使用具有 $20\times 0.75\text{NA}$ 物镜、 532nm 光源、针对 532nm 激发和Cy3荧光发射优化的带通和二向色镜滤光片组以及相机(例如,Andor sCMOS、Zyla 4.2)的倒置荧光显微镜(例如,Olympus IX83),在非信号饱和条件下,同时表面浸入缓冲液(例如, 25mM ACES、 $\text{pH } 7.4$ 缓冲液)中时,表面的荧光图像表现出至少20的对比度噪声比(CNR)。

[0632] 杂交缓冲液

[0633] 本公开提供了使用杂交缓冲液(例如高效杂交缓冲液)将本文所述的多个共价闭合环状文库分子(400)分配到具有多个固定的第一表面引物的支持物上的组合物和方法,使得各个共价闭合环状文库分子(400)与各个固定的第三表面引物杂交,由此固定多个共价闭合环状文库分子(400)。

[0634] 杂交缓冲液(例如,高效杂交缓冲液)包含第一极性非质子溶剂和第二极性非质子溶剂、 pH 缓冲系统和拥挤剂。本文所述的杂交组合物中包括的极性溶剂是包含一种或更多种特征在于存在永久偶极矩的分子的溶剂或溶剂系统,即具有电荷密度空间不均匀分配的分子。极性溶剂的特征可在于20、25、30、35、40、45、50、55、60的介电常数或在于前述值中的值或具有前述值中的任何值的值的范围。本文所述的极性溶剂可包含极性非质子溶剂。本文所述的极性非质子溶剂可进一步在分子中不含可电离的氢。此外,在本发明所公开的组合物的上下文中,极性溶剂或极性非质子溶剂可优选地用强极化官能团(诸如腈、羰基、硫醇、内酯、砜、亚硫酸盐和碳酸酯基团)取代,使得下面的溶剂分子具有偶极矩。极性溶剂和极性非质子溶剂可以脂族和芳香族或环状两种形式存在。在一些实施例中,极性溶剂为乙腈。

[0635] 本文所述的极性或非极性非质子溶剂可具有与乙腈相同或接近的介电常数。极性或非极性非质子溶剂的介电常数可在约20至60、约25至55、约25至50、约25至45、约25至40、约30至50、约30至45、或约30至40的范围内。极性或非极性非质子溶剂的介电常数可大于20、25、30、35或40。极性或非极性非质子溶剂的介电常数可低于30、40、45、50、55或60。极性或非极性非质子溶剂的介电常数可以为约35、36、37、38或39。

[0636] 本文所述的极性或非极性非质子溶剂可具有与乙腈相同或接近的极性指数。极性或非极性非质子溶剂的极性指数可在约2-9、2-8、2-7、2-6、3-9、3-8、3-7、3-6、4-9、4-8、4-7或4-6的范围内。极性或非极性非质子溶剂的极性指数可大于约2、3、4、4.5、5、5.5或6。极性或非极性非质子溶剂的极性指数可低于约4.5、5、5.5、6、6.5、7、7.5、8、8.5、9或10。极性或非极性非质子溶剂的极性指数可以为约5.5、5.6、5.7或5.8。

[0637] 极性或非极性非质子溶剂的一些实例包括但不限于:乙腈、二甲基甲酰胺(DMF)、二甲基亚砜(DMSO)、乙酰苯胺、*N*-乙酰基吡咯烷酮、4-氨基吡啶、苯甲酰胺、苯并咪唑、1,2,3-苯并三唑、二氧化丁二烯、2,3-碳酸丁烯酯、 γ -丁内酯、己内酯(ϵ)、氯马来酸酐、2-氯环己酮、氯代碳酸乙烯酯、氯硝基甲烷、柠康酸酐、巴豆酰内酯、5-氰基-2-硫尿嘧啶、环丙基腈、硫酸二甲酯、二甲基砜、1,3-二甲基-5-四唑、1,5-二甲基四唑、1,2-二硝基苯、2,4-二硝基甲苯、二苯砜、1,2-二硝基苯、2,4-二硝基甲苯、二苯砜、 ϵ -己内酰胺、乙烷磺酰氯、乙基次膦酸乙酯、*N*-乙基四唑、碳酸乙烯酯、三硫代碳酸乙烯酯、硫酸乙二醇酯、亚硫酸乙二醇酯、糠醛、2-呋喃甲酰腈、2-咪唑、靛红、异噁唑、丙二腈、4-甲氧基苯甲腈、1-甲氧基-2-硝基苯、 α -溴代特窗酸甲酯(methyl α bromo tetronate)、1-甲基咪唑、*N*-甲基咪唑、3-甲基异噁唑、*N*-甲基吗啉-*N*-氧化物、甲基苯基砜、*N*-甲基吡咯烷酮、甲基环丁砜、甲基-4-甲苯磺酸

盐、3-硝基苯胺、硝基苯并咪唑、2-硝基呋喃、1-亚硝基-2-吡咯烷酮、2-硝基噻吩、2-噁唑烷酮、9,10-菲醌、N-苯基悉尼酮、邻苯二甲酸酐、吡啶甲腈(2-氰基吡啶)、1,3-丙磺酸内酯、 β -丙内酯、碳酸丙烯酯、4H-吡喃-4-硫酮、4H-吡喃-4-酮(γ -吡喃酮)、哒嗪、2-吡咯烷酮、糖精、琥珀腈、磺胺、环丁砜、2,2,6,6-四氯环己酮、四氢噻喃氧化物、四亚甲基砜(环丁砜)、噻唑、2-硫脲嘧啶、3,3,3-三氯丙烯、1,1,2-三氯丙烯、1,2,3-三氯丙烯、三亚甲基硫化物-二氧化物和亚硫酸三亚甲基酯。

[0638] 极性溶剂或极性非质子溶剂的量以有效使双链核酸变性的量存在。在一些实施例中,按制剂的总体积计,极性或极性非质子溶剂的量为大于约以体积计10%。按制剂的总体积计,极性或极性非质子溶剂的量为约或超过约以体积计5%、10%、15%、20%、25%、30%、35%、40%、50%、60%、70%、80%、90%或更高。按制剂的总体积计,极性或极性非质子溶剂的量为低于约以体积计15%、20%、25%、30%、35%、40%、50%、60%、70%、80%、90%或更高。在一些实施例中,极性或极性非质子溶剂的量按配制物的总体积计在约以体积计10%至90%的范围内。在一些实施例中,极性或极性非质子溶剂的量按配制物的总体积计在约以体积计25%至75%的范围内。在一些实施例中,按制剂的总体积计,极性或极性非质子溶剂的量为在约以体积计10%至95%、10%至85%、20%至90%、20%至80%、20%至75%、或30%至60%的范围内。

[0639] 在一些实施例中,所公开的杂交缓冲液制剂可以包括添加有机溶剂。合适的溶剂的实例包括但不限于:乙腈、乙醇、DMF和甲醇、或它们的处于不同百分比(通常>5%)的任意组合。在一些实施例中,杂交缓冲液中包括的有机溶剂的百分比(以体积计)可以在约1%至约20%的范围内。在一些实施例中,有机溶剂的体积百分比可以为至少1%、至少2%、至少3%、至少4%、至少5%、至少6%、至少7%、至少8%、至少9%、至少10%、至少15%、或至少20%。在一些实施例中,有机溶剂的体积百分比可以为最多20%、最多15%、最多10%、最多9%、最多8%、最多7%、最多6%、最多5%、最多4%、最多3%、最多2%、最至多1%。本段中所述的下限和上限值中的任一者可进行组合以形成包括在本公开内的范围,例如,有机溶剂的体积百分比的范围可以为约4%至约15%。本领域技术人员将认识到,有机溶剂的体积百分比可具有在该范围内的任何值,例如,约7.5%。

[0640] 杂交速率的改进:在一些实施例中,使用本文公开的优化的缓冲液制剂(任选地,与低非特异性结合表面组合使用)产生是常规杂交方案的缓冲液制剂的约2倍至约20倍范围内的更快的相对杂交速率。在一些实施例中,相对杂交速率可以是常规杂交方案的相对杂交速率的至少2倍、至少3倍、至少4倍、至少5倍、至少6倍、至少7倍、至少8倍、至少9倍、至少10倍、至少12倍、至少14倍、至少16倍、至少18倍或至少20倍。

[0641] 杂交效率(或产量)的提高是与互补序列杂交的固体表面上的可用拴系的衔接子序列、引物序列或寡核苷酸序列整体的百分比的量度。在一些实施例中,与常规杂交方案的缓冲液制剂相比,使用本文公开的优化缓冲液制剂(任选地,与低非特异性结合表面组合使用)产生改进的杂交效率。在一些实施例中,在上文指定的杂交反应时间中的任一者可实现的杂交效率优于80%、85%、90%、95%、98%或99%。

[0642] 杂交特异性的改进是对拴系的衔接子序列、引物序列或寡核苷酸序列整体仅与完全互补的序列正确杂交的能力的一种衡量。在一些实施例中,与常规杂交方案的缓冲液制剂相比,使用本文公开的优化的缓冲液制剂(任选地,与低非特异性结合表面组合使用)产

生改进的杂交特异性。在一些实施例中,可实现的杂交特异性优于10个杂交事件中1个碱基错配、100个杂交事件中1个碱基错配、1,000个杂交事件中1个碱基错配、或10,000个杂交事件中1个碱基错配。

[0643] 术语“拥挤剂”和相关术语是指改变溶液中其他分子的性质的化合物。拥挤剂通常具有高分子量和/或庞大的结构。溶液中的拥挤剂可增加溶液中其他分子的浓度。拥挤剂可减少溶液中对于其他分子而言可用的溶剂的体积,这可产生分子拥挤环境。溶液中的拥挤剂可生成对于溶液中的分子而言的拥挤环境。拥挤剂可改变反应的速率或平衡常数。拥挤剂的实例包括:聚乙二醇(例如,PEG)、聚蔗糖、葡聚糖、糖原、聚乙烯醇、三嵌段聚合物(例如,Pluronic)、聚苯乙烯、聚乙烯吡咯烷酮(PVP)、羟丙基甲基纤维素(HPMC)、羟乙基甲基纤维素(HEMC)、羟丁基甲基纤维素、羟丙基纤维素、甲基纤维素和羟甲基纤维素。在一些实施例中,拥挤剂包含直链或支化PEG。在一些实施例中,拥挤剂包括PEG 400、PEG 1500、PEG 2000、PEG 3400、PEG 3350、PEG 4000、PEG 6000、或PEG 8000。在一些实施例中,按溶液的体积计,溶液可包括约1%、2%、3%、5%、10%、15%、20%、25%、30%、35%、40%、50%、60%或更高百分比的至少一种拥挤剂。在一些实施例中,溶液可用于核酸扩增,包括滚环扩增和/或多重置换扩增反应。

[0644] 组合物中合适量的拥挤剂允许、增强或促进分子拥挤。按制剂的总体积计,拥挤剂的量为约或多于约以体积计1%、2%、3%、5%、10%、15%、20%、25%、30%、35%、40%、50%、60%或更高。在一些情况下,按制剂的总体积计,分子拥挤剂的量大于以体积计5%。按制剂的总体积计,拥挤剂的量为低于约以体积计3%、5%、10%、12.5%、15%、20%、25%、30%、35%、40%、50%、60%、70%、80%、90%或更高。在一些情况下,按制剂的总体积计,分子拥挤剂的量可以小于以体积计30%。在一些实施例中,极性或极性非质子溶剂的量按配制物的总体积计在约以体积计25%至75%的范围内。在一些实施例中,按制剂的总体积计,极性或极性非质子溶剂的量在约以体积计1%至40%、1%至35%、2%至50%、2%至40%、2%至35%、2%至30%、2%至25%、2%至20%、2%至10%、5%至50%、5%至40%、5%至35%、5%至30%、5%至25%、5%至20%的范围内。在一些情况下,按制剂的总体积计,分子拥挤剂的量可在约以体积计5%至约20%的范围内。在一些实施例中,按制剂的总体积计,拥挤剂的量在约以体积计1%至30%的范围内。

[0645] 在一些实施例中,所公开的杂交缓冲液制剂可以包括添加分子拥挤剂或体积排阻剂。分子拥挤或体积排阻剂通常为大分子(例如蛋白质),其当以高浓度被添加到溶液时可通过减少对于其他分子而言可用的溶剂的体积来改变溶液中的其他分子的性质。在一些实施例中,杂交缓冲液制剂中包括的分子拥挤剂或体积排阻剂的体积百分比可以在约1%至约50%的范围内。在一些实施例中,分子拥挤或体积排阻剂的体积百分比可以为至少1%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、或至少50%。在一些实施例中,分子拥挤或体积排阻剂的体积百分比可以为最多50%、最多45%、最多40%、最多35%、最多30%、最多25%、最多20%、最多15%、最多10%、最多5%、或最多1%。本段中所述的下限和上限值中的任一者可进行组合以形成包括在本公开内的范围,例如,分子拥挤或体积排阻剂的体积百分比的范围可以为约5%至约35%。本领域技术人员将认识到,分子拥挤或体积排阻剂的体积百分比可具有在该范围内的任何值,例如,约12.5%。

[0646] 本文描述的杂交缓冲液包括将组合物的pH维持在适合杂交过程的范围内的pH缓冲系统。pH缓冲液系统可包括选自以下项组成的组的一种或多种缓冲剂:Tris、HEPES、TAPS、Tricine、Bicine、Bis-Tris、NaOH、KOH、TES、EPPS、MES和MOPS。pH缓冲系统可进一步包括溶剂。优选的pH缓冲系统包括:MOPS、MES、TAPS、与甲醇、乙腈、乙醇、异丙醇、丁醇、叔丁醇组合的磷酸盐缓冲剂、DMF、DMSO、或它们的任意组合。

[0647] 杂交缓冲液包括有效维持制剂的pH在适合杂交的范围内的一定量的pH缓冲系统。在一些实施例中,pH可以为至少3、至少4、至少5、至少6、至少7、至少8、至少9、或至少10。在一些实施例中,pH可以为最多10、最多9、最多8、最多7、最多6、最多5、最多4、或最多3。本段中所述的下限和上限值中的任一者可进行组合以形成包括在本公开内的范围,例如,杂交缓冲液的pH的范围可以为约4至约8。本领域技术人员将认识到杂交缓冲液的pH可具有在该范围内的任何值,例如,约pH 7.8。在一些情况下,pH范围为约3至约10。在一些实施例中,所公开的杂交缓冲液制剂可包括在约pH 3至pH 10的范围内pH的调节,其中优选的缓冲范围为5至9。

[0648] 本文描述的杂交缓冲液包括用于控制核酸解链温度的添加剂(例如,极性非质子溶剂),其可以根据组合物中使用的其他试剂而变化。按制剂的总体积计,用于控制核酸的解链温度的添加剂的量约为或超过约以体积计1%、2%、3%、5%、10%、15%、20%、25%、30%、35%、40%、50%、60%或更高。在一些情况下,按制剂的总体积计,用于控制核酸的解链温度的添加剂的量大于约以体积计2%。在一些情况下,按制剂的总体积计,用于控制核酸的解链温度的添加剂的量大于以体积计5%。在一些情况下,按制剂的总体积计,用于控制核酸的解链温度的添加剂的量低于约以体积计3%、5%、10%、12.5%、15%、20%、25%、30%、35%、40%、50%、60%、70%、80%、90%或更高。在一些实施例中,按制剂的总体积计,用于控制核酸的解链温度的添加剂的量在约以体积计1%至40%、1%至35%、2%至50%、2%至40%、2%至35%、2%至30%、2%至25%、2%至20%、2%至10%、5%至50%、5%至40%、5%至35%、5%至30%、5%至25%、5%至20%的范围内。在一些实施例中,按制剂的总体积计,用于控制核酸的解链温度的添加剂的量在约以体积计2%至20%的范围内。在一些情况下,用于控制核酸的解链温度的添加剂的量在按配制物的总体积计约5体积%至10体积%的范围内。

[0649] 在一些实施例中,所公开的杂交缓冲液制剂可以包括添加改变核酸双链体解链温度的添加剂。可用于改变核酸解链温度的合适添加剂的实例包括但不限于甲酰胺。在一些实施例中,杂交缓冲液制剂中包括的解链温度添加剂的体积百分比可以在约1%至约50%的范围内。在一些实施例中,解链温度添加剂的体积百分比可以为至少1%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、或至少50%。在一些实施例中,解链温度添加剂的体积百分比可以为最多50%、最多45%、最多40%、最多35%、最多30%、最多25%、最多20%、最多15%、最多10%、最多5%、或最多1%。本段中所述的下限和上限值中的任一者可进行组合以形成包括在本公开内的范围,例如,解链温度添加剂的体积百分比的范围可以为约10%至约25%。本领域技术人员将认识到,解链温度添加剂的体积百分比可具有在该范围内的任何值,例如,约22.5%。

[0650] 在一些实施例中,本文所述的杂交缓冲液包括影响DNA水合的添加剂:在一些实施例中,所公开的杂交缓冲液制剂可包括添加影响核酸水合的添加剂。实例包括但不限于:甜

菜碱、脲、甘氨酸甜菜碱、或它们的任意组合。在一些实施例中,包括在杂交缓冲液制剂中的水合添加剂的体积百分比的范围可以为约1%至约50%。在一些实施例中,水合添加剂的体积百分比可以为至少1%、至少5%、至少10%、至少15%、至少20%、至少25%、至少30%、至少35%、至少40%、至少45%、或至少50%。在一些实施例中,水合添加剂的体积百分比可以为最多50%、最多45%、最多40%、最多35%、最多30%、最多25%、最多20%、最多15%、最多10%、最多5%、或最多1%。本段中所述的下限和上限值中的任一者可进行组合以形成包括在本公开内的范围,例如,水合添加剂的体积百分比的范围可以为约1%至约30%。本领域技术人员将认识到,解链温度添加剂的体积百分比可具有在该范围内的任何值,例如,约6.5%。

[0651] 测序聚合酶

[0652] 本公开内容提供了用于对核酸分子进行测序的方法,其中本文所述的任何测序方法采用至少一种类型的测序聚合酶和多个核苷酸,或者采用至少一种类型的测序聚合酶和多个核苷酸以及多个多价分子。在一些实施例中,测序聚合酶能够掺入与多联体模板分子中的核苷酸相对的互补核苷酸。在一些实施例中,测序聚合酶能够结合与多联体模板分子中的核苷酸相对的多价分子的互补核苷酸单元。在一些实施例中,多种测序聚合酶包括重组突变体聚合酶。

[0653] 用于以核苷酸和/或多价分子测序的合适聚合酶的实例包括但不限于:Klenow DNA聚合酶;水栖栖热菌DNA聚合酶I(Taq聚合酶);KlenTaq聚合酶;Candidatus altiarthaeales古菌;Candidatus Hadarchaeum Yellowstoneense;Hadesarchaea古菌;广古菌(Euryarchaeota)古菌;热原体古菌;嗜热球菌聚合酶,诸如Thermococcus litoralis、噬菌体T7 DNA聚合酶;人类 α 、 δ 和 ϵ DNA聚合酶;噬菌体聚合酶,诸如T4、RB69和phi29噬菌体DNA聚合酶;激烈火球菌DNA聚合酶(Pfu聚合酶);枯草芽孢杆菌DNA聚合酶III;大肠杆菌DNA聚合酶III α 和 ϵ ;9度N聚合酶;逆转录酶,诸如HIV M型或O型逆转录酶;禽成髓细胞瘤病毒逆转录酶;莫洛尼鼠白血病病毒(MMLV)逆转录酶;或端粒酶。DNA聚合酶的另外的非限制性实例包括来自各种古生菌属(诸如,气热菌属、古生球菌属、除硫球菌属、火棒菌属、火球菌属、火叶菌属、热网菌属、葡萄嗜热菌属、斯特氏菌属、硫化叶菌属、热球菌属和火山鬃菌属等等或它们的变体)的那些,包括如本领域中已知的此类聚合酶,诸如9度N、VENT、DEEP VENT、THERMINATOR、Pfu、KOD、Pfx、Tgo和RB69聚合酶。

[0654] 核苷酸

[0655] 本公开提供了用于对核酸分子进行测序的方法,其中本文所述的任何测序方法均采用至少一个核苷酸。核苷酸包括碱基、糖和至少一个磷酸酯基团。在一些实施例中,多个核苷酸中的至少一个核苷酸包含芳香族碱基、五碳糖(例如,核糖或脱氧核糖)以及一个或多个磷酸基团(例如,1至10个磷酸基团)。该多个核苷酸可包括选自以下项组成的组的至少一个类型的核苷酸:dATP、dGTP、dCTP、dTTP和dUTP。多个核苷酸可包括选自dATP、dGTP、dCTP、dTTP和/或dUTP组成的组的两个或更多个类型的核苷酸的任意组合的混合物。在一些实施例中,多个核苷酸中的至少一个核苷酸不是核苷酸类似物。在一些实施例中,多个核苷酸中的至少一个核苷酸包括核苷酸类似物。

[0656] 在一些实施例中,在本文所述用于对核酸分子进行测序的任何方法中,多个核苷酸中的至少一个核苷酸包含一个、两个或三个磷原子的链,其中该链通常经由酯键或磷酸

胺键附接至该糖部分的5'碳。在一些实施例中,多个核苷酸中的至少一个核苷酸为具有磷链的类似物,其中磷原子用居间的O、S、NH、亚甲基或亚乙基链接在一起。在一些实施例中,链中的磷原子包括经取代的侧基,包括O、S或BH₃。在一些实施例中,链包括用类似物取代的磷酸酯基团,包括磷酰胺、硫代磷酸酯、二硫代磷酸酯和O-甲基亚磷酰胺基团。

[0657] 在一些实施例中,在本文所述用于对核酸分子进行测序的任何方法中,多个核苷酸中的至少一个核苷酸包括终止子核苷酸类似物,其在糖2'位置处、在糖3'位置处、或在糖2'和3'位置处具有链终止部分(例如,阻断部分)。在一些实施例中,在引物延伸反应期间,链终止部分可抑制后续核苷酸单元或游离核苷酸在新生链中的聚合酶催化的掺入。在一些实施例中,链终止部分附接至3'糖羟基位置,其中糖包括核糖或脱氧核糖部分。在一些实施例中,链终止部分可从3'糖羟基位置移除/裂解以生成具有3'OH糖基团的核苷酸,其可在聚合酶催化的核苷酸的掺入反应中用后续核苷酸延伸。在一些实施例中,链终止部分包括烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基基团、胺基基团、酰胺基基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团。在一些实施例中,链终止部分可从核苷酸裂解/移除,例如通过用化学剂、pH变化、光或热使链终止部分发生反应。在一些实施例中,链终止部分烷基、烯基、炔基和烯丙基可用四(三苯基膦)钯(0) (Pd(PPh₃)₄)、用哌啶或用2,3-二氯-5,6-二氰基-1,4-苯醌(DDQ)裂解。在一些实施例中,链终止部分芳基和苄基可用H₂ Pd/C裂解。在一些实施例中,链终止部分胺、酰胺、酮、异氰酸酯、磷酸酯、硫代、二硫可用磷化氢或用硫醇基团裂解,该硫醇基团包括β-巯基乙醇或二硫苏糖醇(DTT)。在一些实施例中,链终止部分碳酸酯可用MeOH中的碳酸钾(K₂CO₃)、用吡啶中的三乙胺或用乙酸(AcOH)中的Zn裂解。在一些实施例中,链终止部分脲和甲硅烷基可用四丁基氟化铵、吡啶-HF、用氟化铵或用三乙胺三氢氟化物裂解。

[0658] 在一些实施例中,在本文所述用于对核酸分子进行测序的任何方法中,多个核苷酸中的至少一个核苷酸包括终止子核苷酸类似物,其在糖2'位置处、在糖3'位置处、或在糖2'和3'位置处具有链终止部分(例如,阻断部分)。在一些实施例中,链终止部分包括叠氮、叠氮基或叠氮基甲基基团。在一些实施例中,链终止部分包括3'-O-叠氮基或3'-O-叠氮基甲基基团。在一些实施例中,链终止部分叠氮、叠氮基和叠氮基甲基基团可用膦化合物裂解/移除。在一些实施例中,膦化合物包括衍生的三烷基膦部分或衍生的三芳基膦部分。在一些实施例中,膦化合物包括:三(2-羧乙基)膦(TCEP)、或双磺基三苯基膦(BS-TPP)、或三(羟丙基)膦(THPP)。在一些实施例中,裂解剂包括4-二甲基氨基吡啶(4-DMAP)。

[0659] 在一些实施例中,在本文所述用于对核酸分子进行测序的任何方法中,核苷酸包含链终止部分,其选自由以下项组成的组:3'-脱氧核苷酸、2',3'-双脱氧核苷酸、3'-甲基、3'-叠氮基、3'-叠氮基甲基、3'-O-叠氮基烷基、3'-O-乙炔基、3'-O-氨基烷基、3'-O-氟烷基、3'-氟甲基、3'-二氟甲基、3'-三氟甲基、3'-磺酰基、3'-丙二酰基、3'-氨基、3'-O-氨基、3'-巯基、3'-氨基甲基、3'-乙基、3'-丁基、3'-叔丁基、3'-苄基甲氧基羰基、3'-叔-丁氧基羰基、3'-O-烷基羰基、3'-硫代磷酸酯和3-O-苄基或其衍生物。

[0660] 在一些实施例中,在本文所述用于对核酸分子进行测序的任何方法中,多个核苷酸包括用可检测的报道部分标记的多个核苷酸。可检测报道基因部分包括荧光团。在一些实施例中,荧光团附接至核苷酸碱基。在一些实施例中,荧光团用接头附接至核苷酸碱基,该接头可从碱基裂解/移除。在一些实施例中,多个核苷酸中的至少一个未用可检测报道基

因部分标记。在一些实施例中, 附接至核苷酸的特定可检测报道基因部分 (例如, 荧光团) 可对应于核苷酸碱基 (例如, dATP、dGTP、dCTP、dTTP 或 dUTP) 以允许对核苷酸碱基的检测和识别。

[0661] 在一些实施例中, 在本文所述用于对核酸分子进行测序的任何方法中, 核苷酸碱基上的可裂解接头包含可裂解部分, 其包含烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基团、胺基团、酰胺基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团。在一些实施例中, 通过用化学剂、pH 变化、光或热使可裂解部分发生反应, 碱基上的可裂解接头可从碱基裂解/移除。在一些实施例中, 可裂解部分烷基、烯基、炔基和烯丙基可用四 (三苯基膦) 钯 (0) ($\text{Pd}(\text{PPh}_3)_4$)、用吡啶或用 2,3-二氯-5,6-二氰基-1,4-苯醌 (DDQ) 裂解。在一些实施例中, 可裂解部分芳基和苄基可用 H_2 Pd/C 裂解。在一些实施例中, 可裂解部分胺、酰胺、酮、异氰酸酯、磷酸酯、硫代、二硫可用磷化氢或用硫醇基团裂解, 该硫醇基团包括 β -巯基乙醇或二硫苏糖醇 (DTT)。在一些实施例中, 可裂解部分碳酸酯可用 MeOH 中的碳酸钾 (K_2CO_3)、用吡啶中的三乙胺或用乙酸 (AcOH) 中的 Zn 裂解。在一些实施例中, 可裂解部分脲和甲硅烷基可用四丁基氟化铵、吡啶-HF、用氟化铵或用三乙胺三氢氟化物裂解。

[0662] 在一些实施例中, 在本文所述用于对核酸分子进行测序的任何方法中, 核苷酸碱基上的可裂解接头包含可裂解部分, 其包括叠氮、叠氮基或叠氮基甲基。在一些实施例中, 可裂解部分叠氮、叠氮基和叠氮基甲基基团可用膦化合物裂解/移除。在一些实施例中, 膦化合物包括衍生的三烷基膦部分或衍生的三芳基膦部分。在一些实施例中, 膦化合物包括: 三 (2-羧乙基) 膦 (TCEP)、或双磺基三苯基膦 (BS-TPP)、或三 (羟丙基) 膦 (THPP)。在一些实施例中, 裂解剂包括 4-二甲基氨基吡啶 (4-DMAP)。

[0663] 在一些实施例中, 在本文所述用于对核酸分子进行测序的任何方法中, 链终止部分 (例如, 在糖 2' 和/或糖 3' 位置处) 和核苷酸碱基上的可裂解接头具有相同或不同的可裂解部分。在一些实施例中, 链接至碱基的链终止部分 (例如, 在糖 2' 和/或糖 3' 位置处) 和可检测报道基因部分可用相同化学剂化学裂解/移除。在一些实施例中, 链接至碱基的链终止部分 (例如, 在糖 2' 和/或糖 3' 位置处) 和可检测报道基因部分可用不同化学剂化学裂解/移除。

[0664] 多价分子

[0665] 本公开提供了用于对核酸分子进行测序的方法, 其中本文描述的任何用于测序的方法采用至少一个多价分子。在一些实施例中, 多价分子包含附接至核心并具有任何构型的多个核苷酸臂, 包括星爆型、旋梯或瓶刷构型 (例如, 图 19-图 23)。多价分子包含: (1) 核心; 以及 (2) 多个核苷酸臂, 其包含: (i) 核心附接部分, (ii) 包含 PEG 部分的间隔基, (iii) 接头, 和 (iv) 核苷酸单元, 其中所述核心附接至所述多个核苷酸臂, 其中所述间隔基附接至所述接头, 其中所述接头附接至所述核苷酸单元。在一些实施例中, 核苷酸单元包含碱基、糖和至少一个磷酸酯基团, 并且接头通过碱基附接至核苷酸单元。在一些实施例中, 接头包含脂族链或低聚乙二醇链, 其中两个接头链均具有 2 至 6 个亚基。在一些实施例中, 接头也包括芳香族部分。图 15 中示出了示例性核苷酸臂。示例性多价分子示出于图 19-图 23 中。示例性间隔基示出于图 24 (顶部) 中且示例性接头示出于图 24 (底部) 和图 25 中。附接至接头的示例性核苷酸示出于图 26-图 29 中。图 30 中示出了示例性生物素化核苷酸臂。

[0666] 在一些实施例中,多价分子包含附接至多个核苷酸臂的核心,并且其中多个核苷酸臂具有选自dATP、dGTP、dCTP、dTTP和dUTP组成的组的相同类型的核苷酸单元。

[0667] 在一些实施例中,多价分子包含附接至多个核苷酸臂的核心,其中每个臂包括核苷酸单元。在一些方面,该多核苷酸单元包括芳香族碱基、五碳糖(例如,核糖或脱氧核糖)以及一个或多个磷酸酯基团(例如,1至10个磷酸酯基团)。该多个多价分子可包括具有选自以下项组成的组的一个类型的核苷酸单元的一个类型的多价分子:dATP、dGTP、dCTP、dTTP和dUTP。多个多价分子可包括两个或更多个类型的多价分子的任意组合的混合物,其中混合物中的各个多价分子包括选自以下项组成的组的核苷酸单元:dATP、dGTP、dCTP、dTTP和/或dUTP。

[0668] 在一些实施例中,核苷酸单元包含一个、两个或三个磷原子的链,其中该链通常经由酯键或磷酸胺键附接至糖部分的5'碳。在一些实施例中,至少一个核苷酸单元为具有磷链的核苷酸类似物,其中磷原子用居间的O、S、NH、亚甲基或亚乙基链接在一起。在一些实施例中,链中的磷原子包括经取代的侧基,包括O、S或BH₃。在一些实施例中,链包括用类似物取代的磷酸酯基团,包括磷酸胺、硫代磷酸酯、二硫代磷酸酯和O-甲基亚磷酸胺基团。

[0669] 在一些实施例中,多价分子包含附接至多个核苷酸臂的核心,并且其中各个核苷酸臂包含核苷酸单元,该核苷酸单元是在糖2'位置处、在糖3'位置、或糖2'和3'位置处具有链终止部分(例如,阻断部分)的核苷酸类似物。在一些实施例中,核苷酸单元包含在糖2'位置处、在糖3'位置处或在糖2'和3'位置处的链终止部分(例如,阻断部分)。在一些实施例中,在引物延伸反应期间,链终止部分可抑制后续核苷酸单元或游离核苷酸在新生链中的聚合酶催化的掺入。在一些实施例中,链终止部分附接至3'糖羟基位置,其中糖包括核糖或脱氧核糖部分。在一些实施例中,链终止部分可从3'糖羟基位置移除/裂解以生成具有3'OH糖基团的核苷酸,其可在聚合酶催化的核苷酸的掺入反应中用后续核苷酸延伸。在一些实施例中,链终止部分包括烷基基团、烯基基团、炔基基团、烯丙基基团、芳基基团、苄基基团、叠氮基基团、胺基基团、酰胺基基团、酮基基团、异氰酸酯基团、磷酸酯基团、硫代基团、二硫基团、碳酸酯基团、脲基团或甲硅烷基基团。在一些实施例中,链终止部分可从核苷酸单元裂解/移除,例如通过用化学剂、pH变化、光或热使链终止部分发生反应。在一些实施例中,链终止部分烷基、烯基、炔基和烯丙基可用四(三苯基膦)钯(0) (Pd(PPh₃)₄)、用哌啶或用2,3-二氯-5,6-二氰基-1,4-苯醌(DDQ)裂解。在一些实施例中,链终止部分芳基和苄基可用H₂Pd/C裂解。在一些实施例中,链终止部分胺、酰胺、酮、异氰酸酯、磷酸酯、硫代、二硫可用磷化氢或用硫醇基团裂解,该硫醇基团包括β-巯基乙醇或二硫苏糖醇(DTT)。在一些实施例中,链终止部分碳酸酯可用MeOH中的碳酸钾(K₂CO₃)、用吡啶中的三乙胺或用乙酸(AcOH)中的Zn裂解。在一些实施例中,链终止部分脲和甲硅烷基可用四丁基氟化铵、吡啶-HF、用氟化铵或用三乙胺三氢氟化物裂解。

[0670] 在一些实施例中,核苷酸单元包含在糖2'位置处、在糖3'位置处或在糖2'和3'位置处的链终止部分(例如,阻断部分)。在一些实施例中,链终止部分包括叠氮、叠氮基或叠氮基甲基基团。在一些实施例中,链终止部分包括3'-O-叠氮基或3'-O-叠氮基甲基基团。在一些实施例中,链终止部分叠氮、叠氮基和叠氮基甲基基团可用膦化合物裂解/移除。在一些实施例中,膦化合物包括衍生的三烷基膦部分或衍生的三芳基膦部分。在一些实施例中,膦化合物包括:三(2-羧乙基)膦(TCEP)、或双磺基三苯基膦(BS-TPP)、或三(羟丙基)膦

(THPP)。在一些实施例中,裂解剂包括4-二甲基氨基吡啶(4-DMAP)。

[0671] 在一些实施例中,所述核苷酸单元包含选链终止部分,其自由以下项组成的组:3'-脱氧核苷酸、2',3'-二脱氧核苷酸、3'-甲基、3'-叠氮基、3'-叠氮基甲基、3'-0-叠氮烷基、3'-0-乙炔基、3'-0-氨基烷基、3'-0-氟烷基、3'-氟甲基、3'-二氟甲基、3'-三氟甲基、3'-磺酰基、3'-丙二酰基、3'-氨基、3'-0-氨基、3'-巯基、3'-氨基甲基、3'-乙基、3'-丁基、3'-叔丁基、3'-苄基甲氧基羰基、3'-叔丁氧基羰基、3'-0-烷基羟氨基、3'-硫代磷酸酯和3'-0-苄基或其衍生物。

[0672] 在一些实施例中,多价分子包含附接至多个核苷酸臂的核心,其中核苷酸臂包含间隔基、接头和核苷酸单元,并且其中核心、接头和/或核苷酸单元用可检测的报道部分标记。在一些实施例中,可检测报道基因部分包含荧光团。在一些实施例中,附接至多价分子的特定可检测报道基因部分(例如,荧光团)可对应于核苷酸单元的碱基(例如,dATP、dGTP、dCTP、dTTP或dUTP),以允许对核苷酸碱基的检测和识别。

[0673] 在一些实施例中,多价分子的至少一个核苷酸臂具有附接至可检测的报道部分的核苷酸单元。在一些实施例中,可检测报道部分附接至核苷酸碱基。在一些实施例中,可检测报道基因部分包含荧光团。在一些实施例中,附接至多价分子的特定可检测报道基因部分(例如,荧光团)可对应于核苷酸单元的碱基(例如,dATP、dGTP、dCTP、dTTP或dUTP),以允许对核苷酸碱基的检测和识别。

[0674] 在一些实施例中,多价分子的核心包含亲和素样或链霉亲和素样部分,并且核心附接部分包含生物素。在一些实施例中,核心包含链霉亲和素型或亲和素型部分,其包括亲和素蛋白,以及可以与至少一个生物素部分结合的亲和素的任何衍生物、类似物和其他非天然形式。其他形式的亲和素部分包括天然和重组亲和素和链霉亲和素以及衍生分子,例如非糖基化亲和素和截短的链霉亲和素。例如,亲和素部分包括亲和素的去糖基化形式、由链霉菌属(例如,阿维丁链霉菌)产生的细菌链霉亲和素,以及衍生形式,例如N-酰基亲和素,例如N-乙酰基、N-邻苯二甲酰基和N-琥珀酰基亲和素,以及市售产品EXTRAVIDIN、CAPTAVIDIN、NEUTRAVIDIN和NEUTRALITE AVIDIN。

[0675] 在一些实施例中,本文所述用于对核酸分子进行测序的任何方法可包括形成结合复合物,其中结合复合物包含(i)聚合酶、与引物形成双链体的核酸多联体分子和核苷酸,或结合复合物包含(ii)聚合酶、与引物形成双链体的核酸多联体分子和多价分子的核苷酸单元。在一些实施例中,结合复合物的存留时间大于约0.1秒、0.2秒、0.3秒、0.4秒、0.5秒、0.6秒、0.7秒、0.8秒、0.9秒或1秒。结合复合物具有大于约0.1秒-0.25秒、或约0.25秒-0.5秒、或约0.5秒-0.75秒、或约0.75秒-1秒、或约1秒-2秒、或约2秒-3秒、或约3秒-4秒、或约4秒-5秒的存留时间,和/或其中该方法在或可在处于或高于15℃、处于或高于20℃、处于或高于25℃、处于或高于35℃、处于或高于37℃、处于或高于42℃、处于或高于55℃、处于或高于60℃、或处于或高于72℃、或处于或高于80℃或在由前述中的任一者定义的范围内的温度被执行。结合复合物(例如,三元复合物)在经历导致聚合酶、模板分子、引物和/或核苷酸单元或核苷酸中的任一者之间的相互作用的解离的条件之前保持稳定。例如,离解条件包括:使结合复合物与去污剂、EDTA和/或水中的任一者或它们的任意组合接触。在一些实施例中,本公开提供了所述方法,其中结合复合物沉积在、附接至或杂交至在检测步骤中表现出大于20的对比度噪声比的表面。在一些实施例中,本公开提供了所述方法,其中在当核苷

酸或核苷酸单元与模板核酸的下一个碱基互补时使结合复合物稳定,并且当核苷酸或核苷酸单元与模板核酸的下一个碱基不互补时使结合复合物不稳定的条件下进行接触。

[0676] 实例

[0677] 以下实例旨在是例示性的并可用于进一步理解本公开的实施例,并且不应被解释为以任何方式限制本教导的范围。

[0678] 实例1:线性核酸文库

[0679] 使用酶FRAGMENTASE ULTRA(来自New England Biolabs)对基因组DNA进行片段化。测试了一系列基因组DNA的输入量,包括1ng/uL至1000ng/uL。在最低范围内,仅测试了1ng的基因组DNA。片段化反应包括7uL的FRAGMENTASE反应缓冲液、2uL的FRAGMENTASE酶、26uL基因组DNA。片段化反应在37℃下孵育15分钟,并在65℃下加热灭活30分钟。平均片段化的DNA约为250bp。片段化的DNA通过酶促末端修复和A加尾反应进行处理。

[0680] 通过衔接子连接和使用有尾引物的PCR来制备线性文库分子。所得文库分子包含以5'至3'顺序排列的区域:(i)第一左通用衔接子序列(120),其具有针对第二表面引物的结合序列;(ii)左样品索引序列(160);(iii)第二左通用衔接子序列(140),其具有针对测序引物的结合序列;(iv)目的序列(110);(v)第二右通用衔接子序列(150),其具有针对测序引物的结合序列;(vi)包括短随机序列的右样品索引序列(例如,NNN);以及(v)第一右通用衔接子序列(130),具有针对第一表面引物的结合序列。例如见图3和图4。

[0681] 实例2:文库成环

[0682] 通过将0.5pmol的线性文库分子与具有以下序列的单链夹板链(200)杂交,使来自实例1的线性文库分子成环5'-ACCCTGAAAGTACGTGCATTACATGGATCAGGTGAGGC TGCGACGACT-3'(SEQ ID NO:195)(参见图8)。杂交反应包括:双链体缓冲液(100mM乙酸钾,30mM HEPES pH 7.5)中的30uL的线性文库分子(16.7nM)、14uL单链夹板链(77nM)。杂交反应在热循环仪中进行:95℃5分钟,37℃5分钟,保持37℃。

[0683] 成环分子(例如,含有缺口)与连接酶反应。连接反应包括:44uL成环分子、1uL T7 DNA连接酶(3,000,000单位/mL)和5uL T4 DNA连接酶反应缓冲液(10X浓度)。连接反应在热循环仪中37℃孵育10分钟,65℃孵育10分钟,并保持在4℃。所得共价闭合环状分子仍与单链夹板链杂交(例如,参见图6A-图6C)。

[0684] 实例3:单链夹板链的降解

[0685] 通过添加1uL不耐热核酸外切酶I(20,000单位/mL)和1uL T7核酸外切酶(10,000单位/mL)降解单链夹板链。降解反应在热循环仪中进行:37℃10分钟,80℃2分钟,并保持在4℃。

[0686] 使用SPRIselect珠子、80%乙醇和磁力板将降解反应物净化两次。

[0687] 通过定量PCR或单链Qubit对共价闭合环状文库分子的产量进行定量。对于用于制备线性文库的约100ng基因组DNA的输入,测得共价闭合环状分子的产量约为50-100fmol(数据未示出)。少至12fmol的共价闭合环状分子已成功用于支持物上的滚环扩增,以产生用于测序的固定的多联体模板分子(数据未示出)。

[0688] 实例4:滚环扩增和测序

[0689] 将来自实例3的共价闭合环状文库分子分配到在高效杂交缓冲液存在下、用低非特异性结合涂层钝化的支持物上,并经历支持物上滚环扩增以产生固定的多联体。

[0690] 实例5:使用多价分子和核苷酸进行测序

[0691] 对多联体经历递归两阶段测序反应,在第一阶段使用荧光标记的多价分子,在第二阶段使用未标记的核苷酸类似物(例如,3'链终止子阻断基团)。

[0692] 两阶段测序反应在其上固定有多个多联体模板分子(例如,固定的聚合酶克隆)的流通池上进行。

[0693] 通过将多个可溶性测序引物与固定至流通池的多联体模板分子杂交,以形成固定的引物-多联体双链体,来进行第一阶段测序反应。将多个第一测序聚合酶流到流通池上(例如,接触固定的引物-多联体双链体)并在适用于将测序聚合酶与双链体结合以形成复合聚合酶的条件下孵育。在包括非催化阳离子(例如,锶、钡和/或钙)的缓冲液存在下,使荧光标记的多价分子的混合物(例如,以约20-100nM的不同浓度)流到流通池上,并在适合将多价分子的互补核苷酸单元与复合聚合酶结合以形成亲和力复合物的条件下孵育,而无需聚合酶催化掺入核苷酸单元。荧光标记的多价分子在其核心处被标记。对复合聚合酶进行洗涤。获得了保持结合至复合聚合酶的经以荧光方式标记的多价分子的图像。通过用包含去污剂的缓冲液洗涤,移除第一测序聚合酶和多价分子,同时保留与固定的多联体杂交的测序引物(保留的双链体)。

[0694] 第一阶段测序反应适用于在多联体模板分子(例如,聚合酶克隆)上形成多个亲和力复合物。例如,第一阶段测序反应包括:(a)将第一核酸引物、第一聚合酶和第一多价分子结合至多联体模板分子的第一部分,由此形成第一结合复合物,其中第一多价分子的第一核苷酸单元与第一聚合酶结合;以及(b)将第二核酸引物、第二聚合酶和第一多价分子结合至同一多联体模板分子的第二部分,由此形成第二结合复合物,其中第一多价分子的第二核苷酸单元结合至第二聚合酶,其中包括同一多价分子的第一结合复合物和第二结合复合物形成第一亲和力复合物。

[0695] 通过使保留的双链体与多个第二测序聚合酶接触以形成复合聚合酶来进行第二阶段测序反应。在包括催化阳离子(例如,镁和/或锰)的缓冲液存在下,将未标记的核苷酸类似物(例如,3'-O-甲基叠氮基核苷酸)(例如,以约1-5uM的不同浓度)的混合物添加到复合聚合酶中,并在适合于将互补核苷酸与复合聚合酶结合并促进聚合酶催化的核苷酸掺入以生成新生的延伸测序引物的条件下孵育。对复合聚合酶进行洗涤。没有获得图像。使掺入的未标记核苷酸类似物与裂解剂反应,移除3'-O-甲基叠氮基基团并生成可延伸的3'OH基团。

[0696] 在替代的第二阶段测序反应中,在包括催化阳离子(例如,镁和/或锰)的缓冲液存在下,将荧光标记的核苷酸类似物(例如,3'-O-甲基叠氮基核苷酸)(例如,约1-5uM)的混合物添加到复合聚合酶中,并在适合于使互补核苷酸与复合聚合酶结合并促进聚合酶催化的核苷酸掺入以生成新生的延伸测序引物的条件下孵育。对复合聚合酶进行洗涤。获得了作为复合聚合酶的一部分的经掺入的以荧光方式标记的核苷酸类似物的图像。使经掺入的经以荧光方式标记的核苷酸类似物与裂解试剂发生反应,该裂解试剂移除3'-O-甲基叠氮基基团并且生成可延伸3'OH基团。

[0697] 通过用包括去污剂的缓冲液进行洗涤来移除第二测序聚合酶,同时保留与多联体杂交的新生的延伸的测序引物(保留的双链体)。通过执行第一阶段和第二阶段测序反应的多个循环来进行重复测序反应,以生成延伸的正向测序引物链。

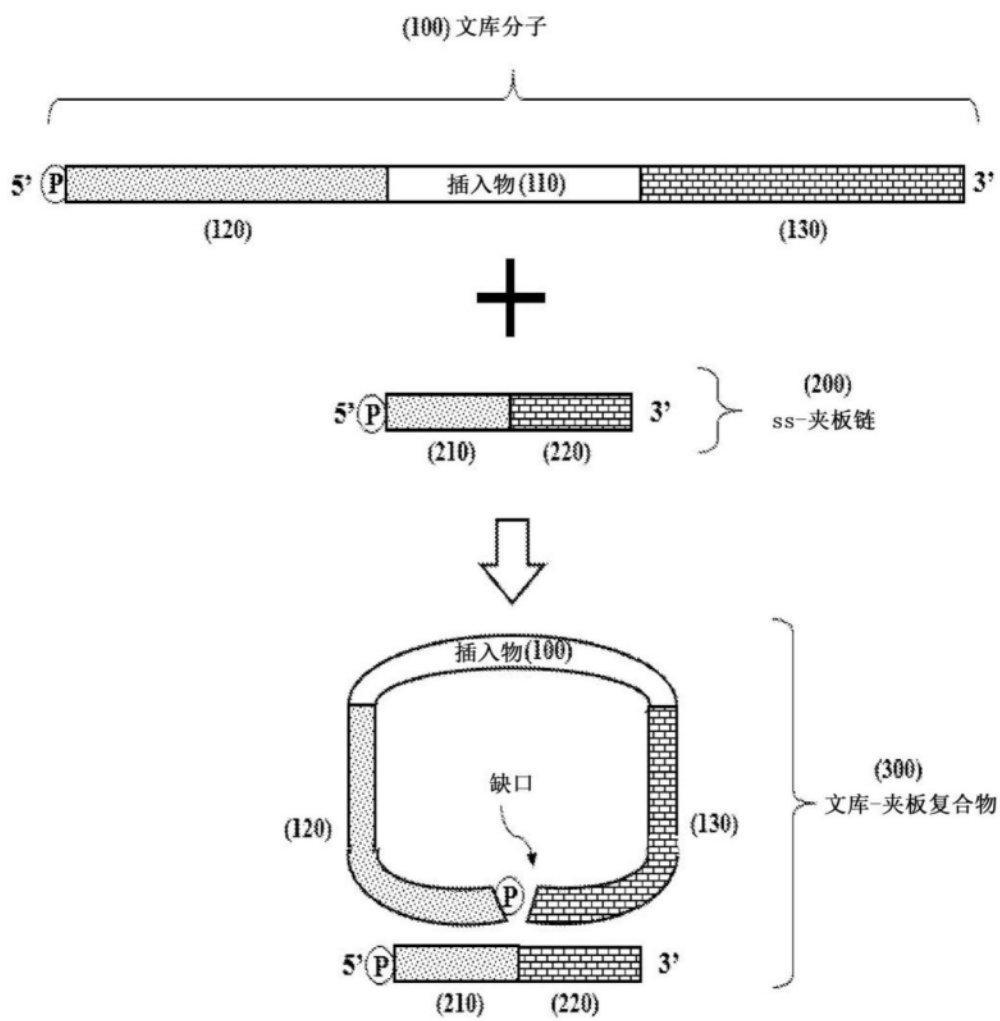


图1

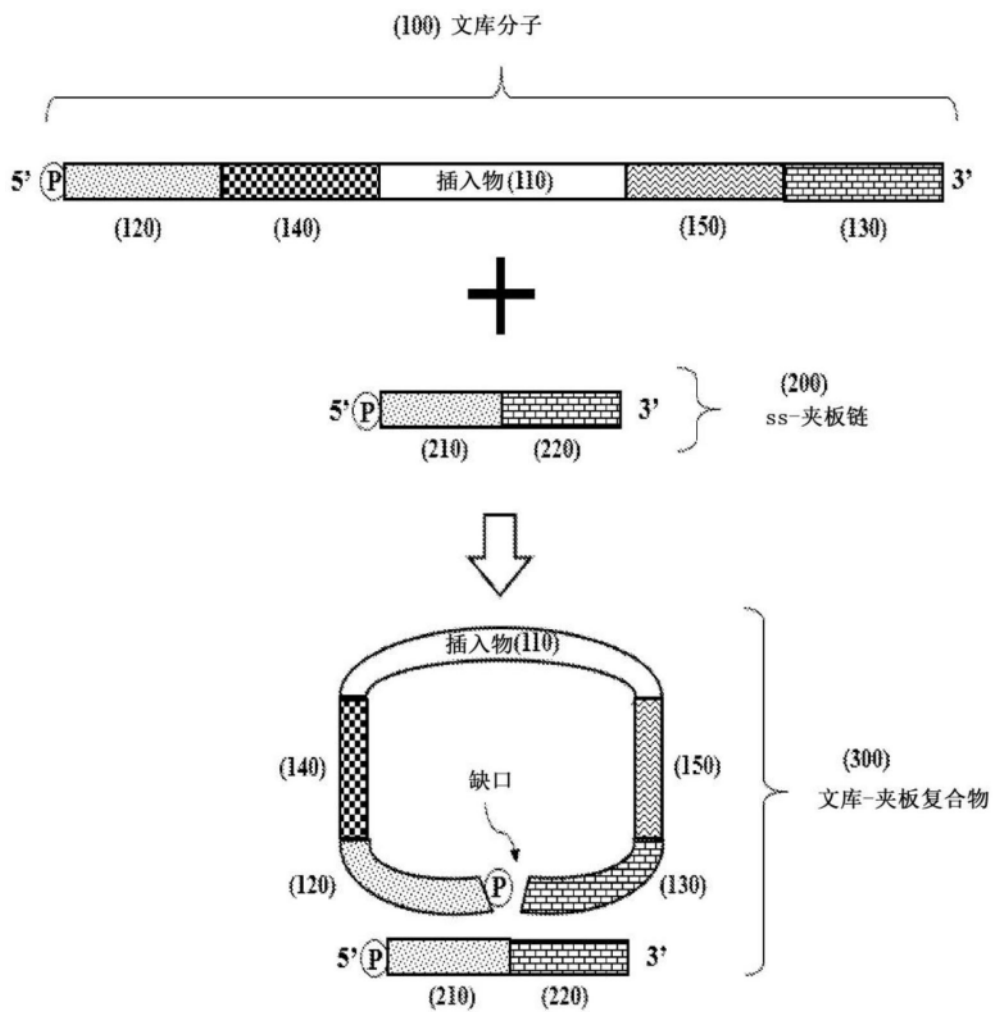


图2

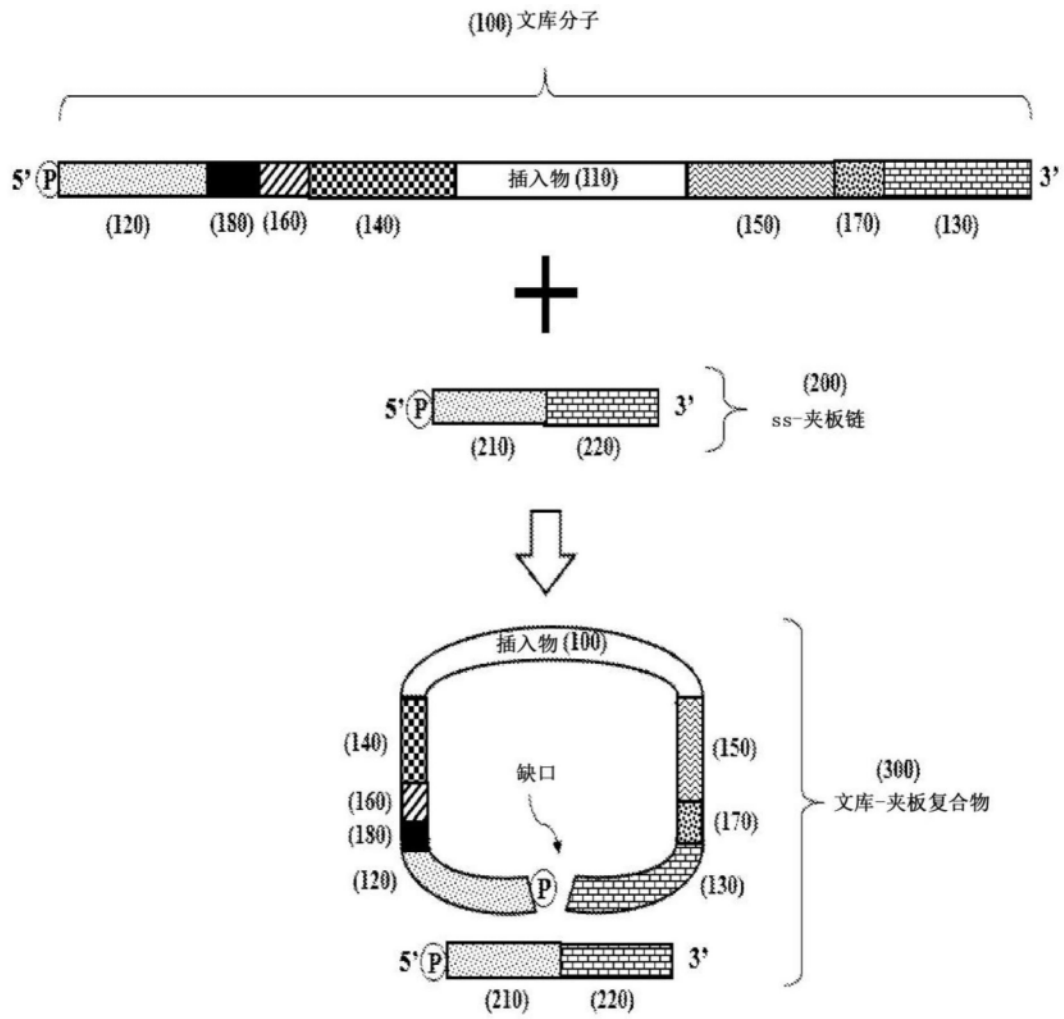


图3

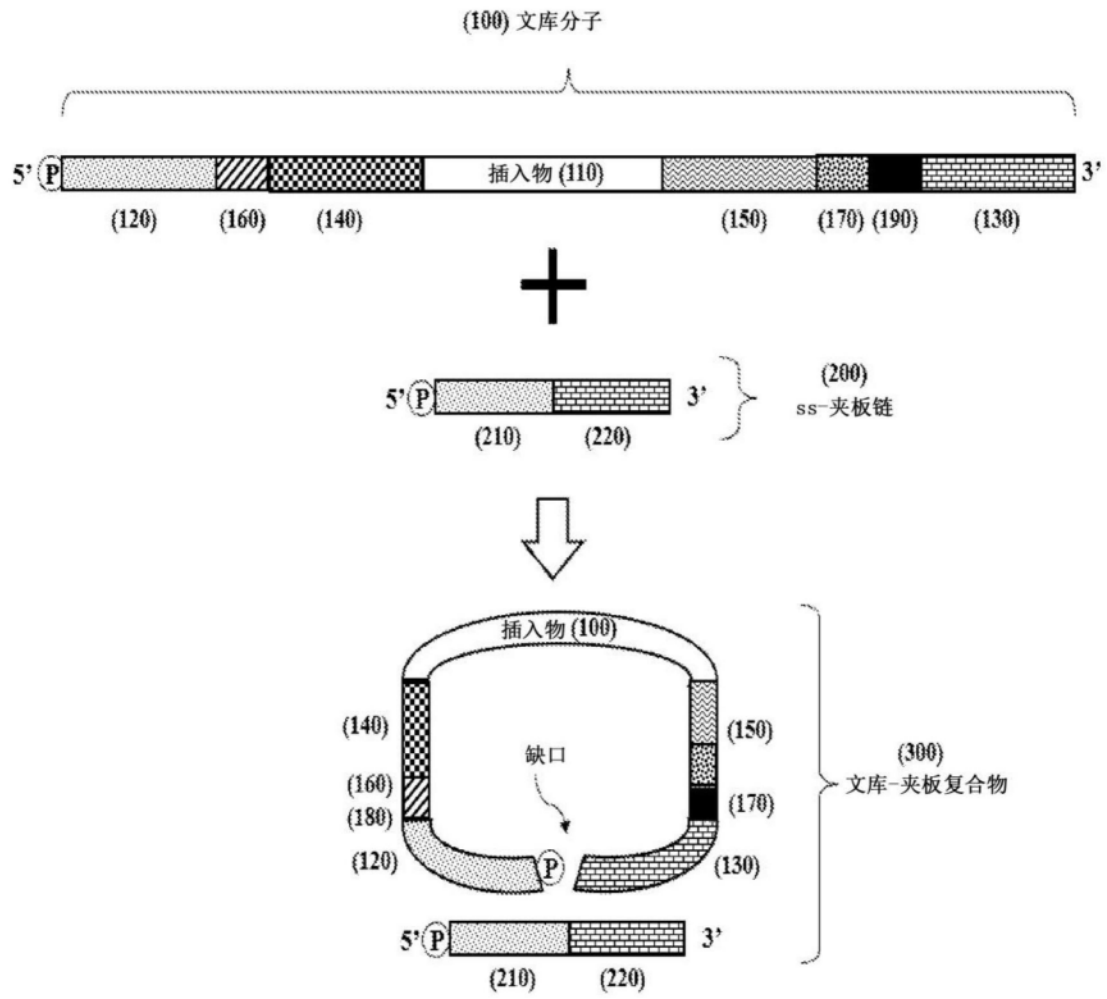


图4

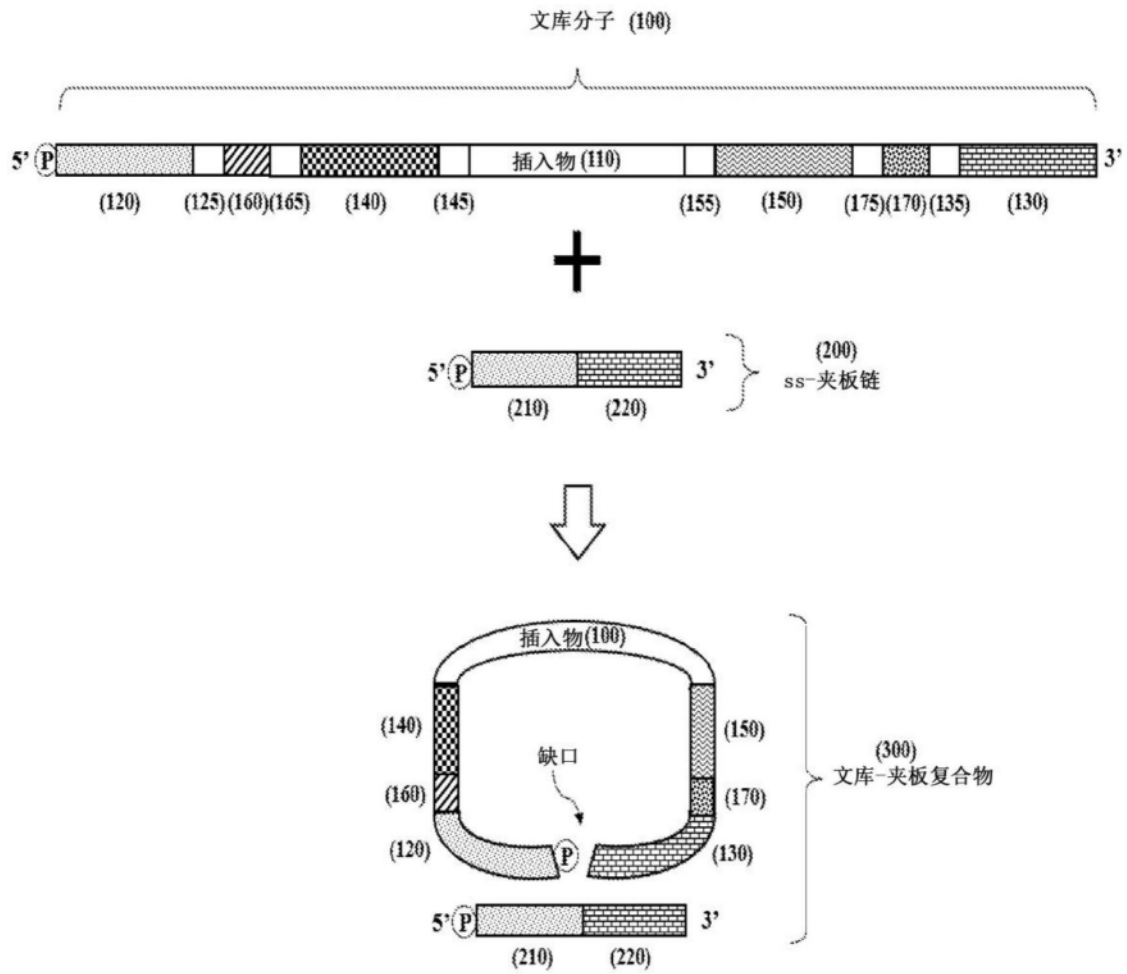


图5

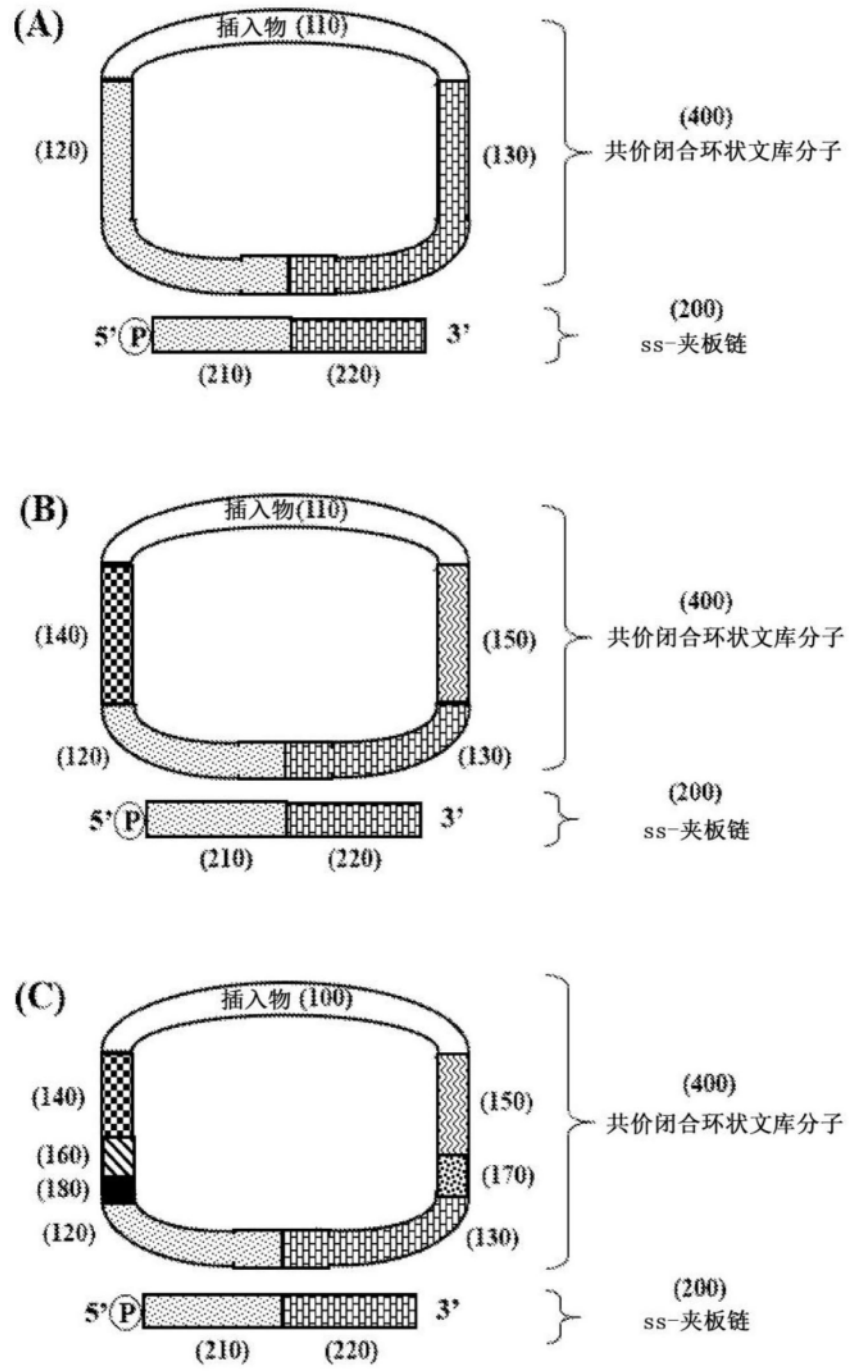


图6

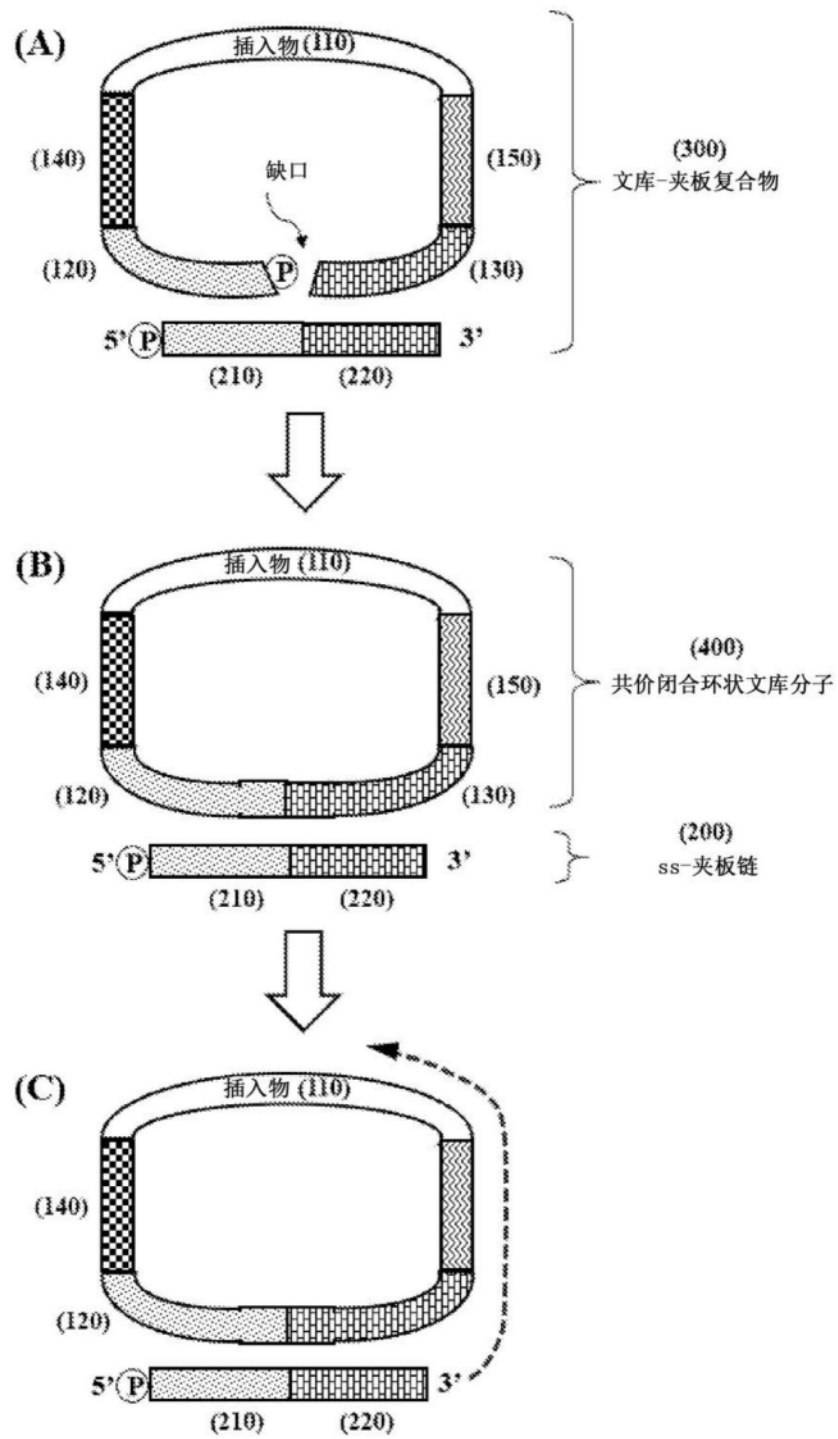


图7



图8

表1:

第一左索引序列(160)		
索引名称	序列(方向: 5' 到3')	SEQ ID NO:
007	CCACTCATT	1
008	AACTAACAC	2
037	TTATGGAGA	3
022	GAGTTCGCT	4
002	TTGTTGTTC	5
035	CCATATTCT	6
001	AATCATGCG	7
014	GGACAGCTC	8
004	CCAACACGA	9
034	TGTAAGCGG	10
012	ATAACTACC	11
019	CGCACTGTG	12
020	TTACACTCA	13
045	ATTGACCTG	14
011	CACCAGTGT	15
038	GGTGCTCCT	16
033	ATCCGCGTC	17
026	CGGTCCTAA	18
046	TGGTATGCC	19
036	AAGATAGAG	20
015	CCGGTATCG	21
042	GGTCCGAGA	22
023	CGTGATAGC	23
081	TATGTGGCC	24
024	GTACCAGAG	25
016	TACATCAGA	26
028	AGGCGACCT	27
039	CAGCACAAG	28
031	CTCTGTCCA	29
013	TCTTGTGAT	30
043	CCAGGAGTC	31
025	ACCGTTCGT	32
044	GACACTTAT	33

图9A

表1（续）：

第一左索引序列(160)		
索引名称	序列（方向：5' 到3'）	SEQ ID NO:
005	TTGGCTGGA	34
017	ATTGCGTAT	35
080	CTCGAAGAT	36
094	TGAACGGAT	37
049	AATACCGGT	38
058	GGCCTTCAA	39
067	TGAGAACAA	40
079	AGATCTCTA	41
054	GGCAACGCA	42
086	CGAATGACA	43
051	CAAGTGCAG	44
059	CAAGCAACT	45
065	TTCAGGTCT	46
064	CAGTAGGTA	47
060	ACGAAGGCT	48
053	TTCTATTGG	49
062	GTGTCTCAC	50
077	GCTGAATGT	51
071	GAAGATGTT	52
088	ATTCGTTGA	53
095	TCGCATAGT	54
057	TCATGCTTG	55
050	TGGAGTATA	56
055	AAGGTGATT	57
063	AGAGGAAGG	58
069	GATCGCGAA	59
056	CTTCGACAC	60
066	GCTATTATG	61
082	GACTGATTA	62
085	ACGGCCAAC	63
075	CTTATCGCG	64
061	TCTCTCTGC	65
073	CAGCGTCC	66

图9B

表1（续）：

第一左索引序列(160)		
索引名称	序列（方向：5' 到3'）	SEQ ID NO:
092	GATTCACGG	67
074	TCCTCGCAA	68
072	ACCACAATT	69
097	CGTTGAATT	70
106	CAGTTGACC	71
111	CCACGAACG	72
112	AGCACCTAG	73
113	TAGGCCGTC	74
114	GTTCTGTAA	75
115	CACTGACGT	76
116	TCCAATACC	77
117	ATATTGGTG	78
119	CCGACCACT	79
120	GAACAACGA	80
121	ATTGTAACC	81
123	TCCACCTGT	82
124	CTACATGAA	83
127	CCAATGTTG	84
132	GTTACCAAC	85
133	CAATTGTGA	86
134	AGGCGTAAG	87
135	ACAGCCGTT	88
136	CACCAACCG	89
137	TCTAGCCTC	90
138	GTGTTAGGA	91
139	CGCGAGACT	92
140	TCAACTTAG	93
141	AAGAGCGCC	94
143	CGACCTTGA	95
144	TACGTACTC	96

图9C

表1 (续) :

第一右索引序列(170)		
索引名称	序列 (方向: 5' 到3')	SEQ ID NO:
001	NNNGTAGGAGCC	97
003	NNNCCGCTGCTA	98
004	NNNAACAACAAG	99
006	NNNGGTGGTCTA	100
008	NNNTTGGCCAAC	101
011	NNNTCCATCGTT	102
013	NNNATTCTAGAG	103
014	NNNGGATGCAAT	104
015	NNNCAGGAGTGC	105
016	NNNATCACACTA	106
018	NNNGCGGATGAT	107
019	NNNCGAATCTGG	108
021	NNNATGCGTCCT	109
022	NNNGGTCTGAGC	110
023	NNNGCCTCATAG	111
024	NNNCTATACCTC	112
025	NNNAAGATTGGA	113
026	NNNTGCCAGCAT	114
027	NNNACTAGCTCA	115
028	NNNTAAGTACTG	116
029	NNNCGTGCAACG	117
030	NNNGTCCACGGA	118
034	NNNAGGTCCGTG	119
035	NNNCTCGTTAGT	120
036	NNNACACAATCC	121
037	NNNTCTTCGCAA	122
038	NNNGGTCACCTG	123
039	NNNCCGATATAT	124
040	NNNACAGGTCGC	125
042	NNNGTGCGGTTC	126
043	NNNCACGCACAT	127
044	NNNTGGATCACA	128
045	NNNATATTCGGT	129

图9D

表1 (续) :

第一右索引序列(170)		
索引名称	序列 (方向: 5' 到3')	SEQ ID NO:
047	NNNCCAGAGATG	130
048	NNNTACCGTTGA	131
049	NNNATTGATACC	132
051	NNNCCAAGACCG	133
052	NNNAAGCCGGAC	134
054	NNNGGATCTCCG	135
055	NNNCCTGGAGAA	136
056	NNNATCCTCATC	137
057	NNNTATGTCTCT	138
059	NNNACGTAGAGT	139
060	NNNGTACTTCAA	140
061	NNNCATTCCAGC	141
062	NNNTGGCGTGAG	142
063	NNNTCAACGTCT	143
064	NNNAACGAAGTC	144
065	NNNCTTAATCGA	145
066	NNNGCGGCCTTA	146
068	NNNTTCCGAACG	147
071	NNNCGCACCGAA	148
073	NNNTAGTGACCA	149
074	NNNGTTATGCTG	150
075	NNNCTCCGCTAT	151
076	NNNTGAGATGGA	152
078	NNNGATAGATGT	153
079	NNNCCTCATGTC	154
081	NNNGACTTAACC	155
082	NNNCAAGGCGTT	156
083	NNNAGGAAGCCG	157
084	NNNTCACTCAAG	158
085	NNNTCCACTAGC	159
087	NNNCTGTCATGA	160
088	NNNAGCGTTCAG	161
089	NNNGCATAGGCA	162

图9E

表1（续）：

第一右索引序列(170)		
索引名称	序列（方向：5' 到3'）	SEQ ID NO:
091	NNNTATCAACGC	163
092	NNNACAGTGTA	164
096	NNNTGCGGATAC	165
097	NNNGCTATTACT	166
106	NNNGCTCAATAA	167
111	NNNCAGAGCTCC	168
112	NNNACGGCACGA	169
113	NNNTTCCTTGGC	170
114	NNNGGTTGGATG	171
115	NNNCCACACCAT	172
116	NNNTACATATCG	173
117	NNNATTGCCATA	174
119	NNNCCATCTTGC	175
120	NNNTGGAATGCT	176
121	NNNAATTCTGCT	177
123	NNNCACCGACTC	178
124	NNNACAATCAGC	179
127	NNNCCGCGAACA	180
132	NNNGAGCATCTA	181
133	NNNCTATTAGAC	182
134	NNNTGTGGCAGT	183
135	NNNCAACCGCAG	184
136	NNNACCTATTCT	185
137	NNNTGTAGGCCA	186
138	NNNGTGGTTATC	187
139	NNNCACCAATGG	188
140	NNNACAACCGAT	189
141	NNNTGTCTGTAA	190
143	NNNTCCGCACTC	191
144	NNNGAGAGCAGA	192

图9F

现有技术

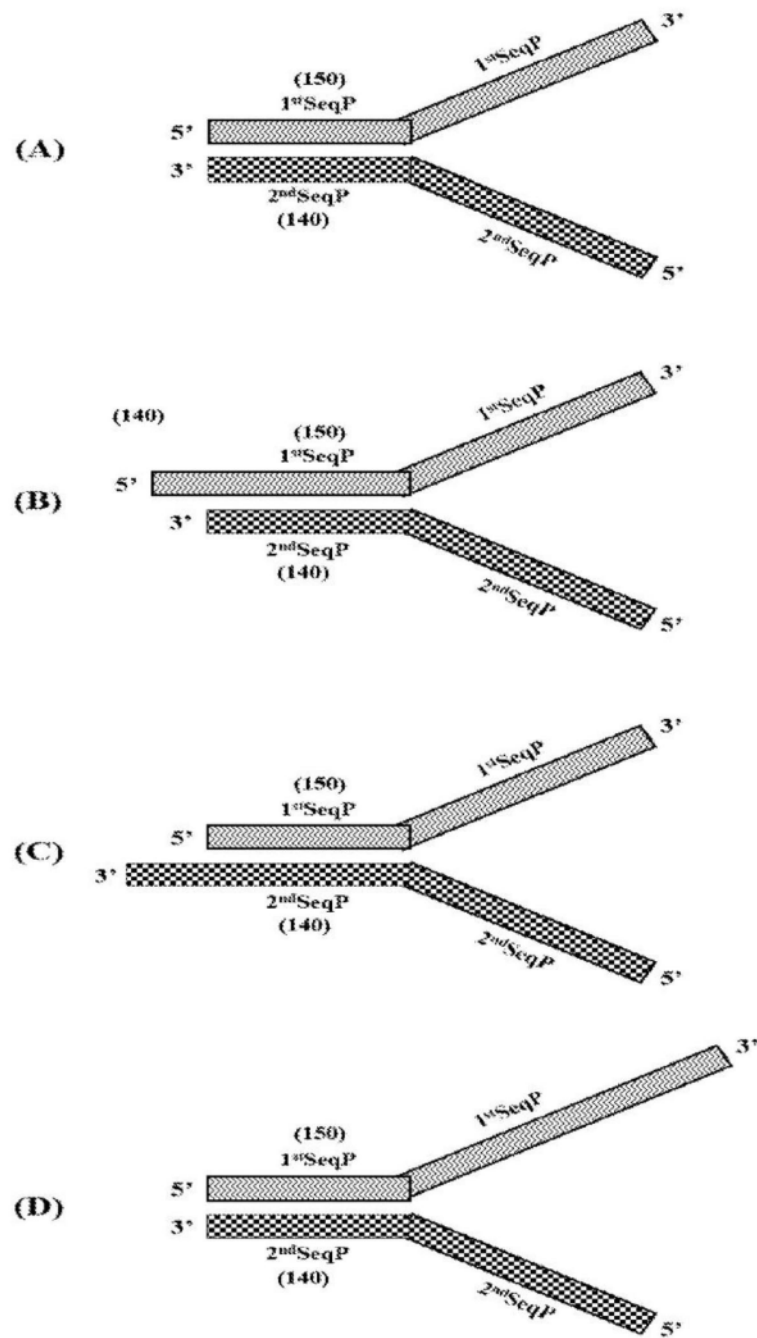


图10

现有技术

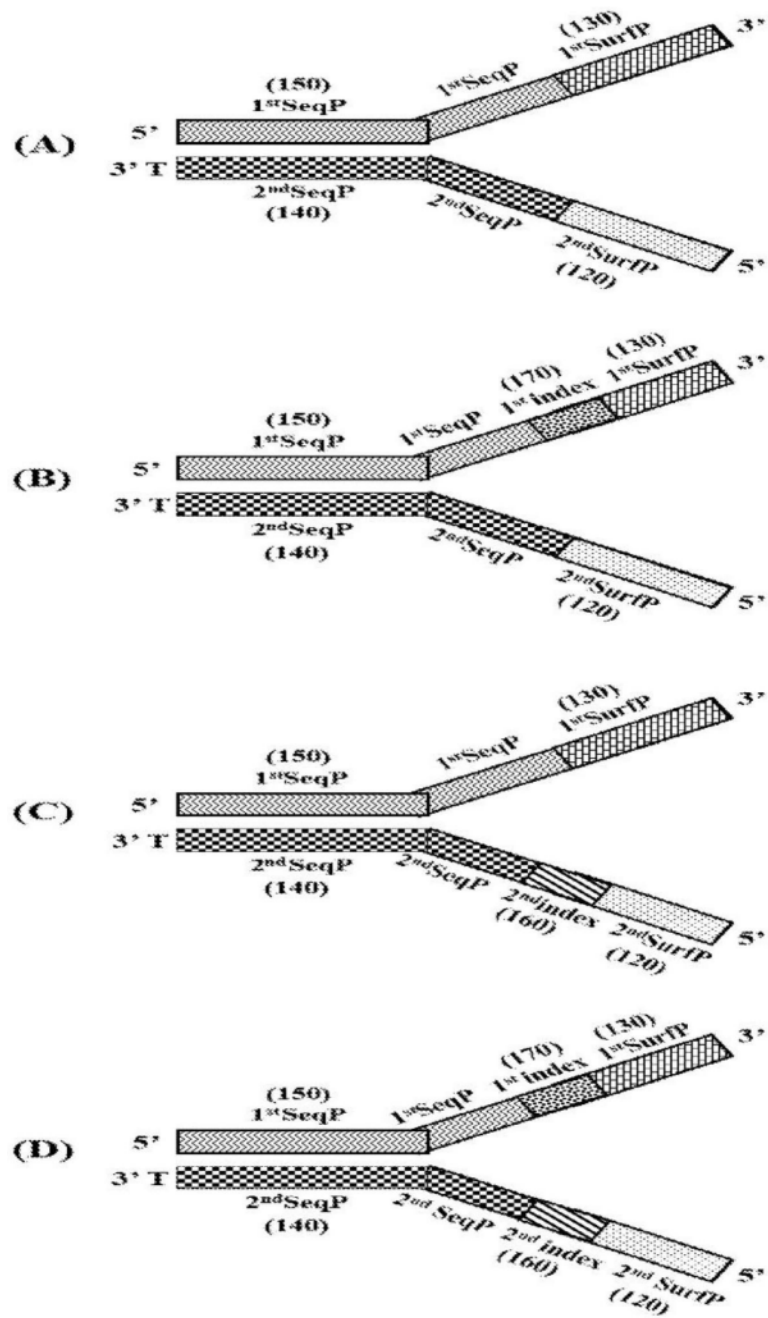


图11

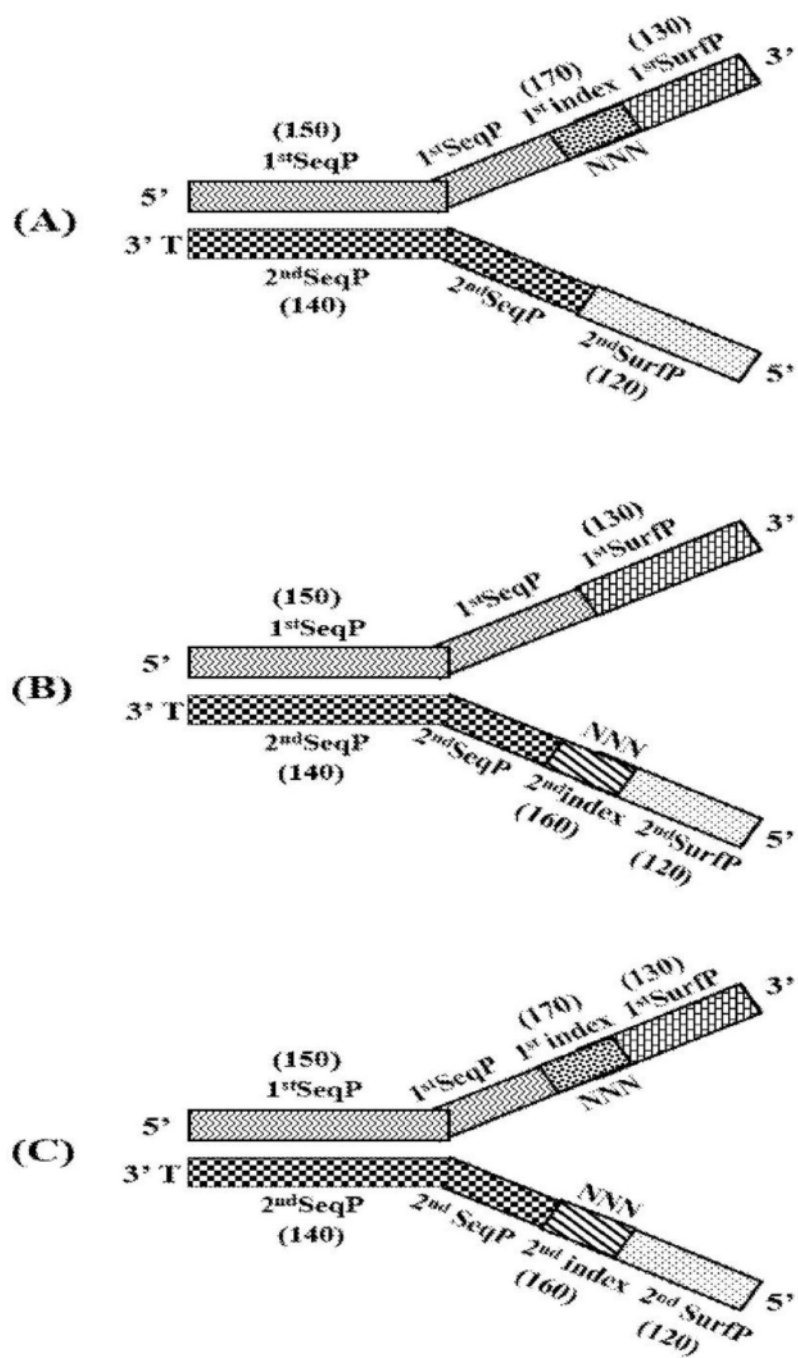


图12

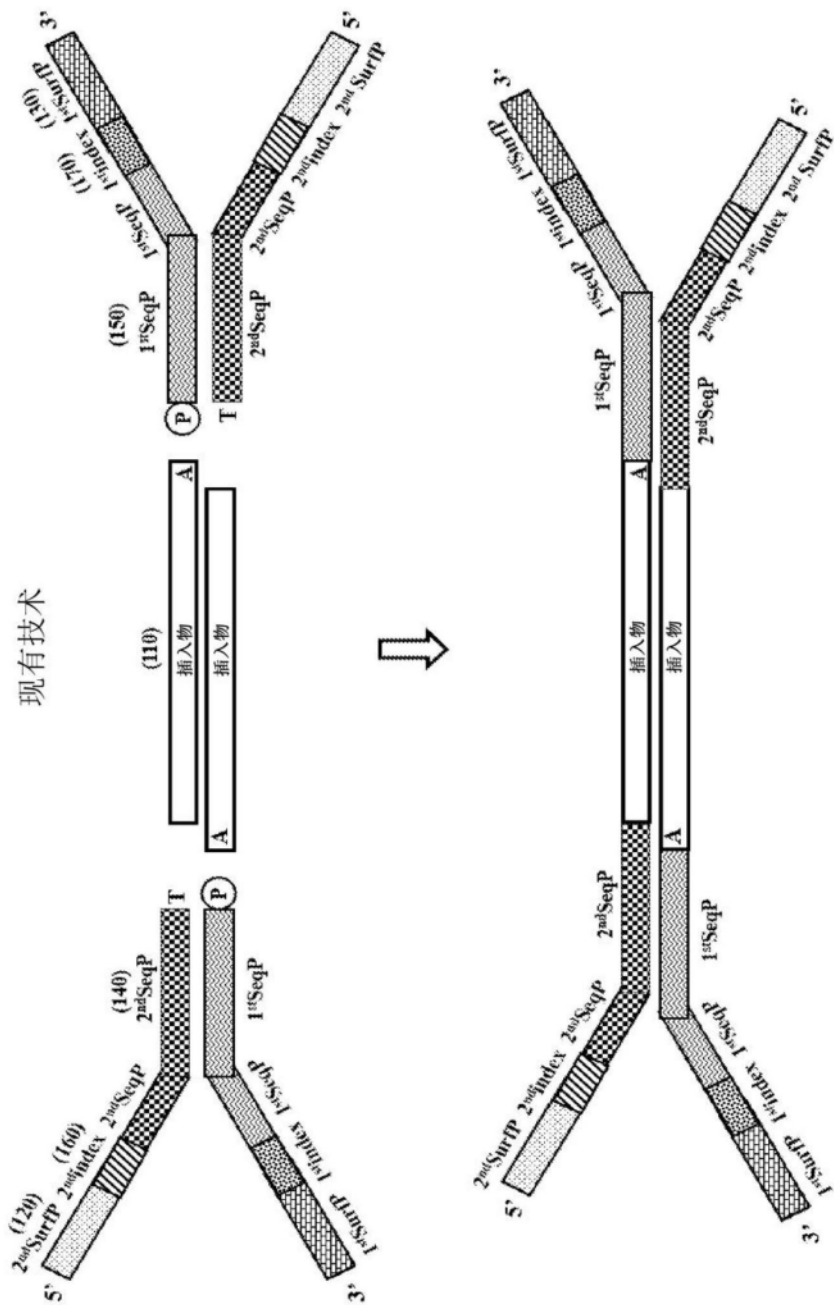


图13

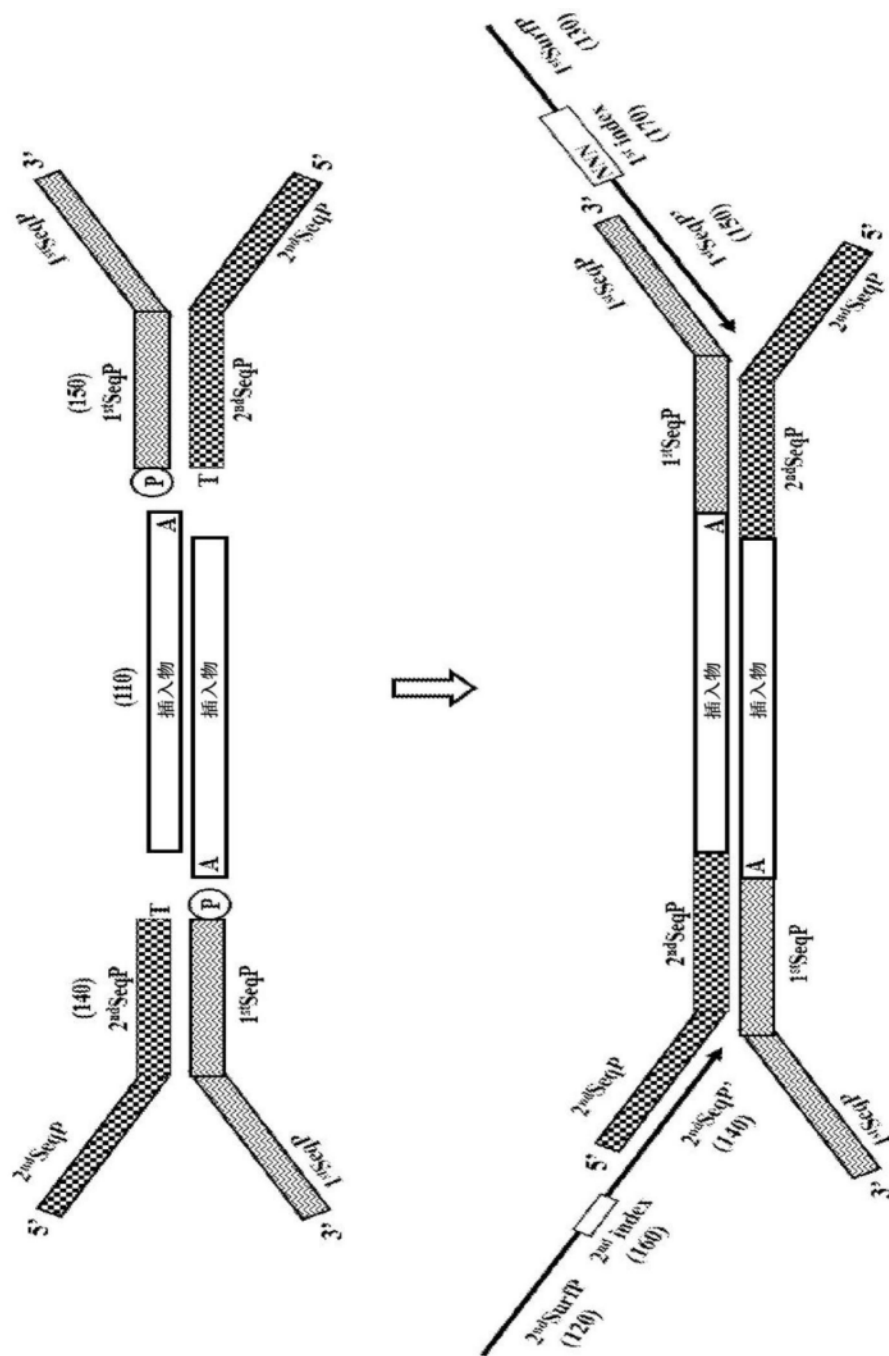


图17

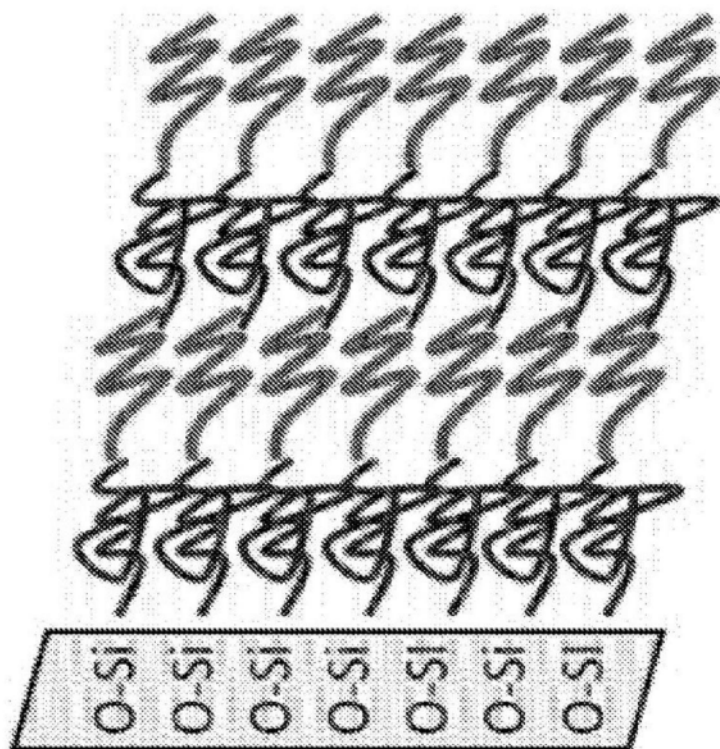


图18

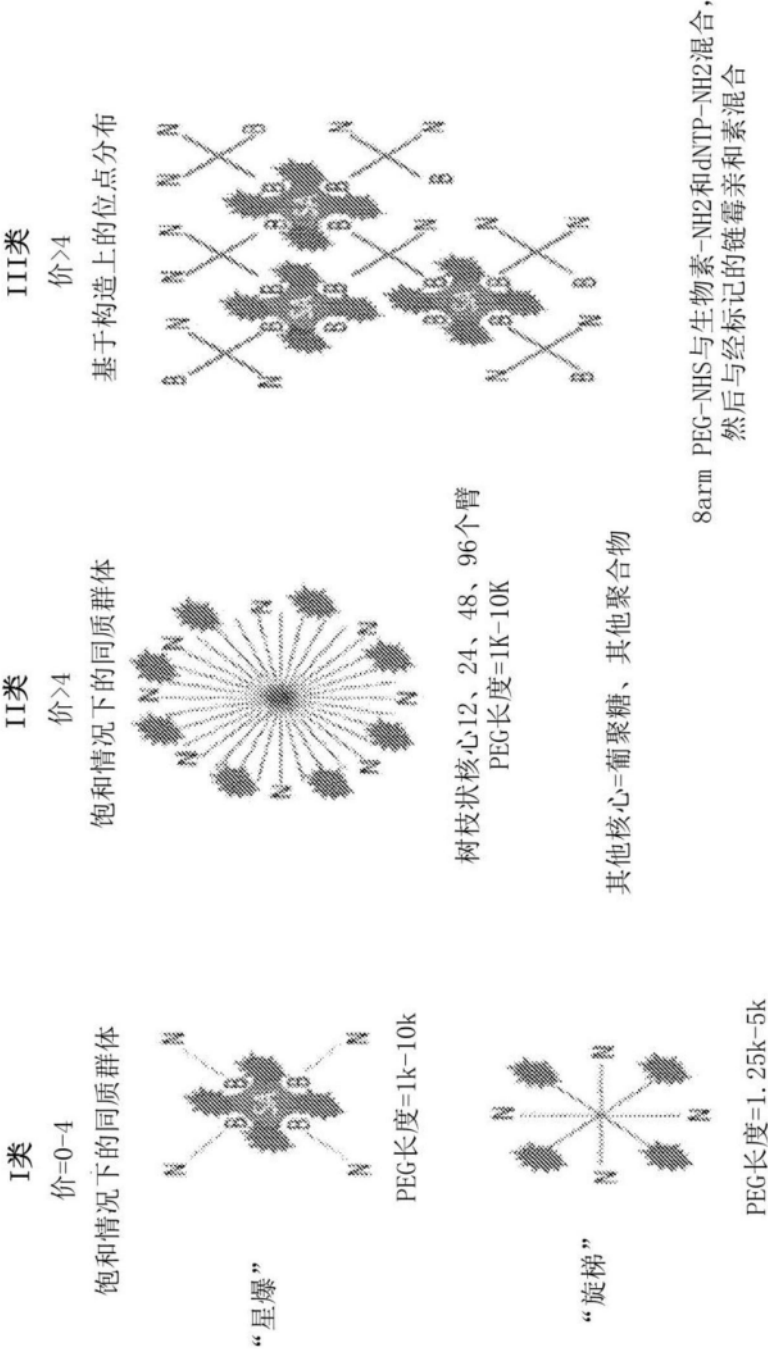


图19

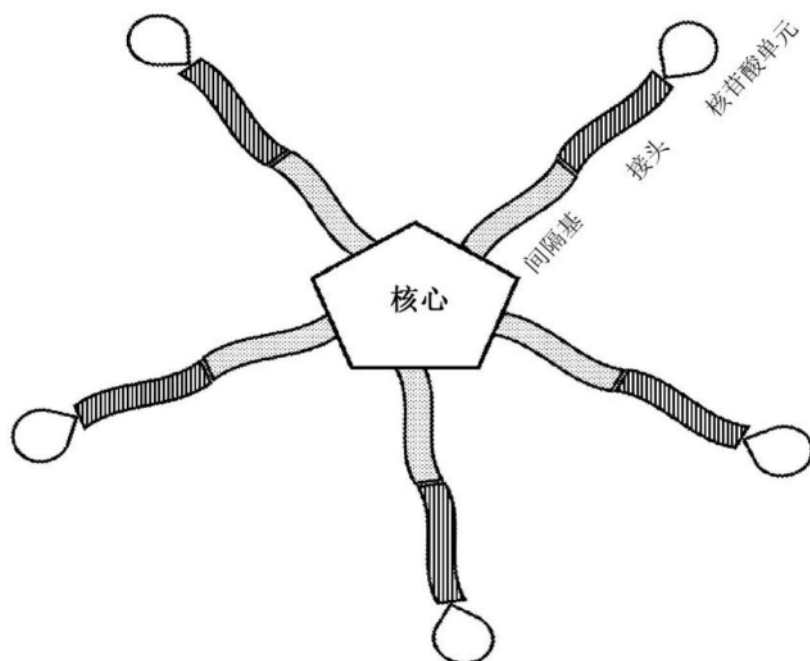


图20

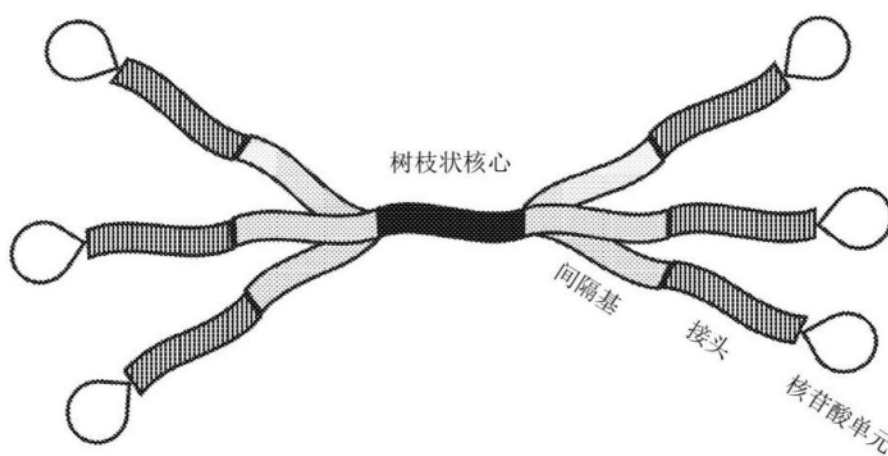


图21

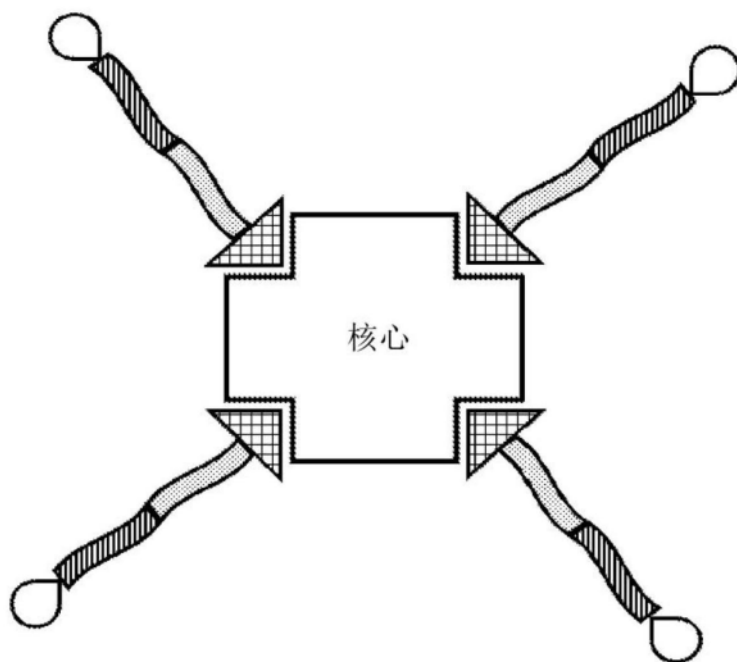


图22

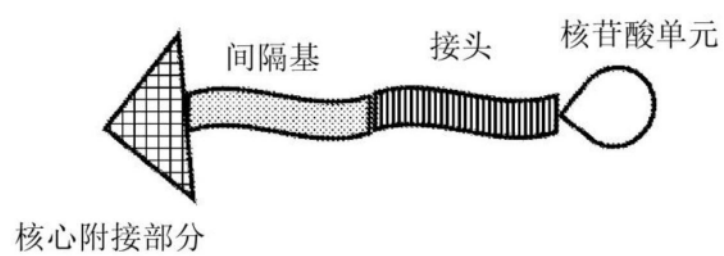
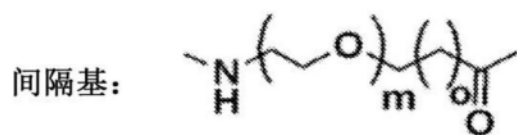


图23

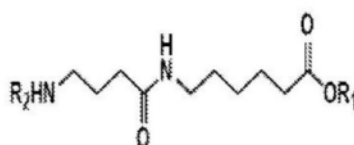


接头:

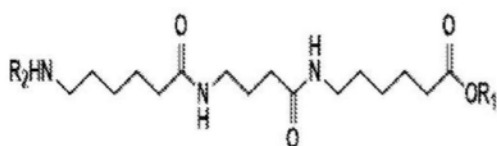
11 原子接头:



16 原子接头:



23 原子接头:



N3 接头:

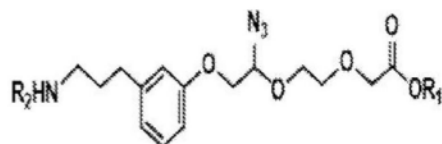


图24

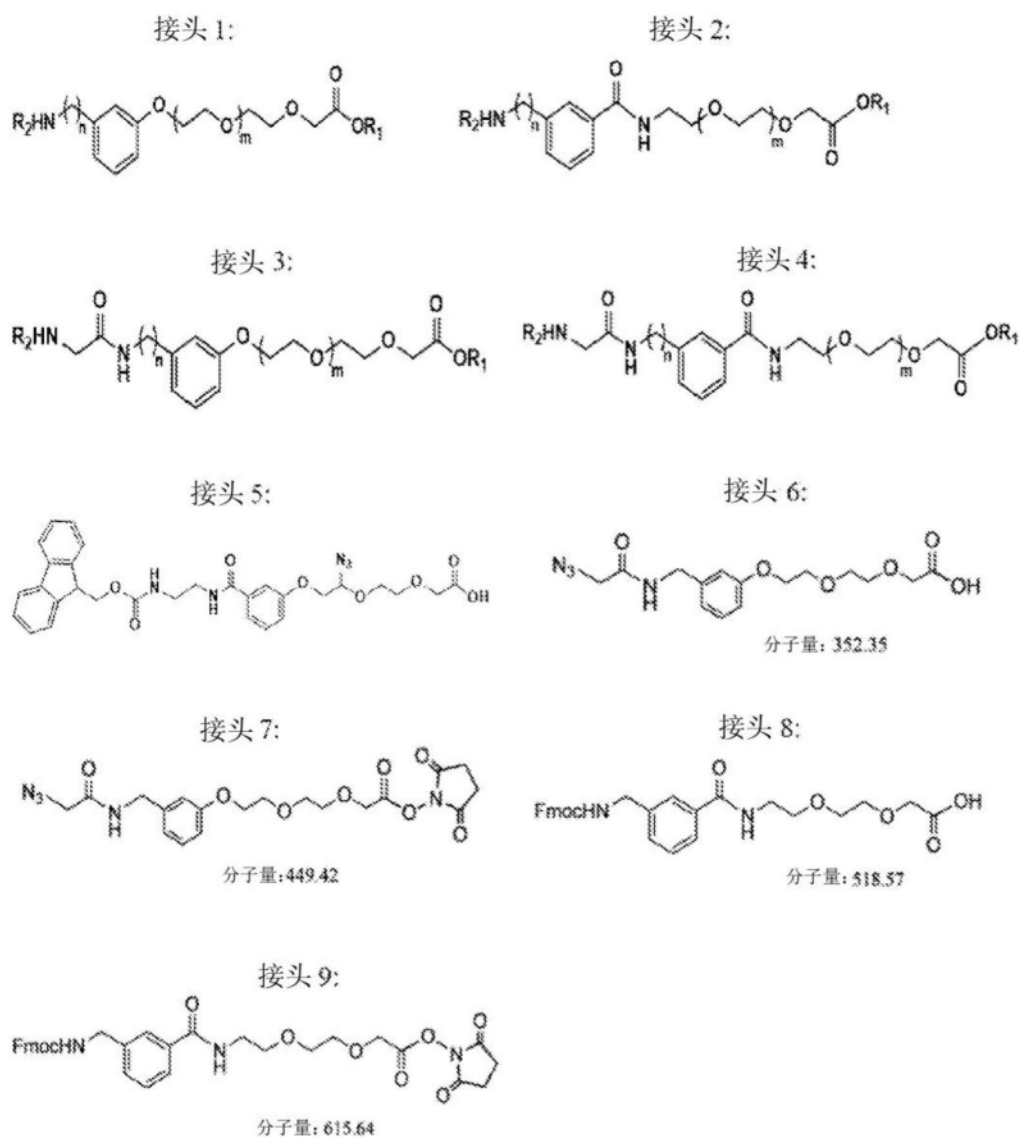


图25

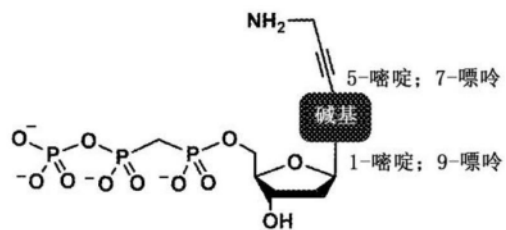
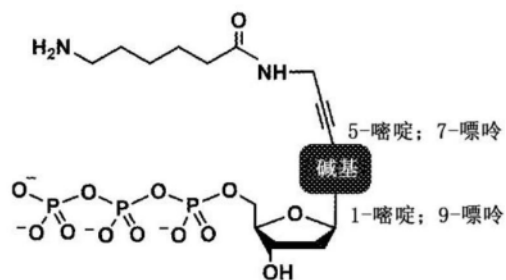
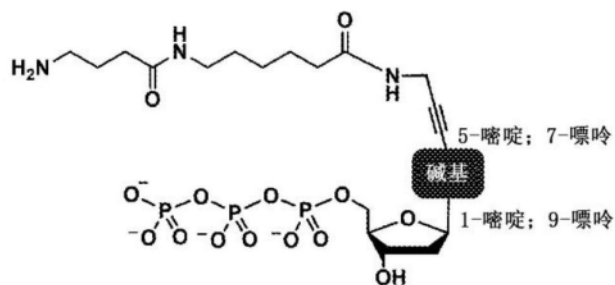
dNTP-PA-NH₂:**dNTP-PA-11 原子接头 -NH₂:****dNTP-PA-16 原子接头 -NH₂:**

图26

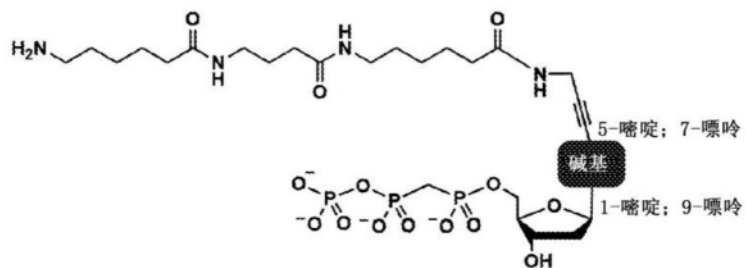
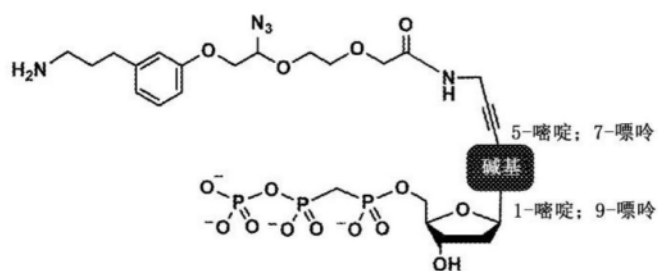
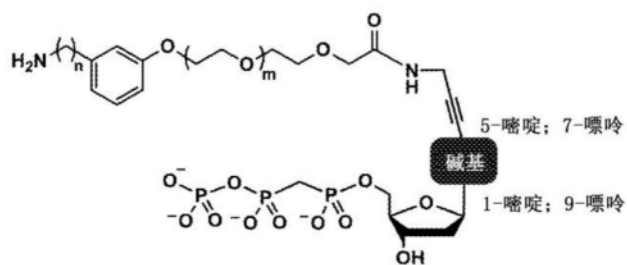
dNTP-PA-23 原子接头 -NH₂:**dNTP-PA-N3 接头 -NH₂:****dNTP-PA-接头 1-NH₂:**

图27

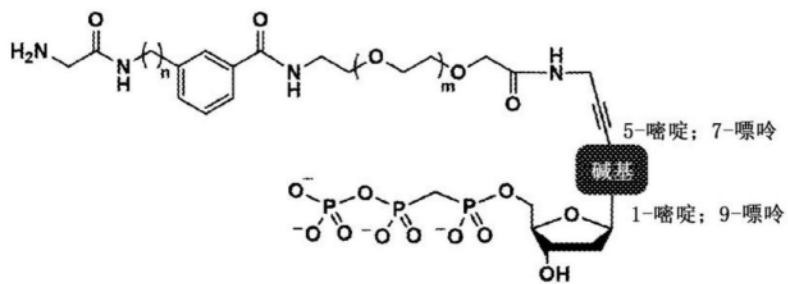
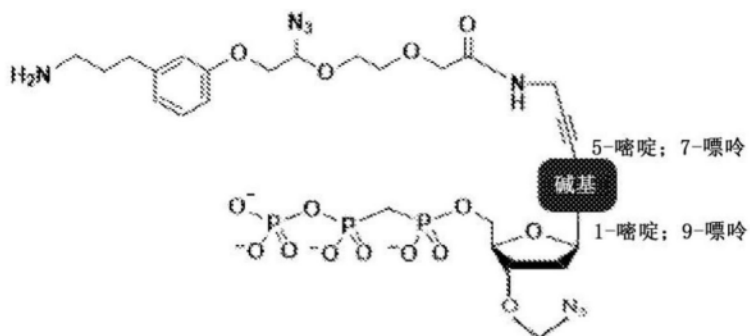
dNTP-PA-接头 4-NH₂:dNTP-PA-N3接头 -NH₂:

图29

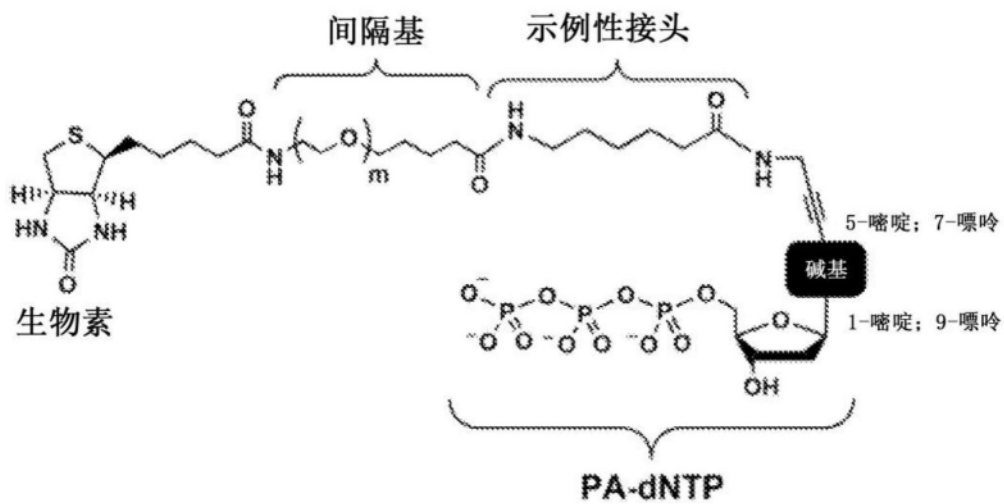


图30

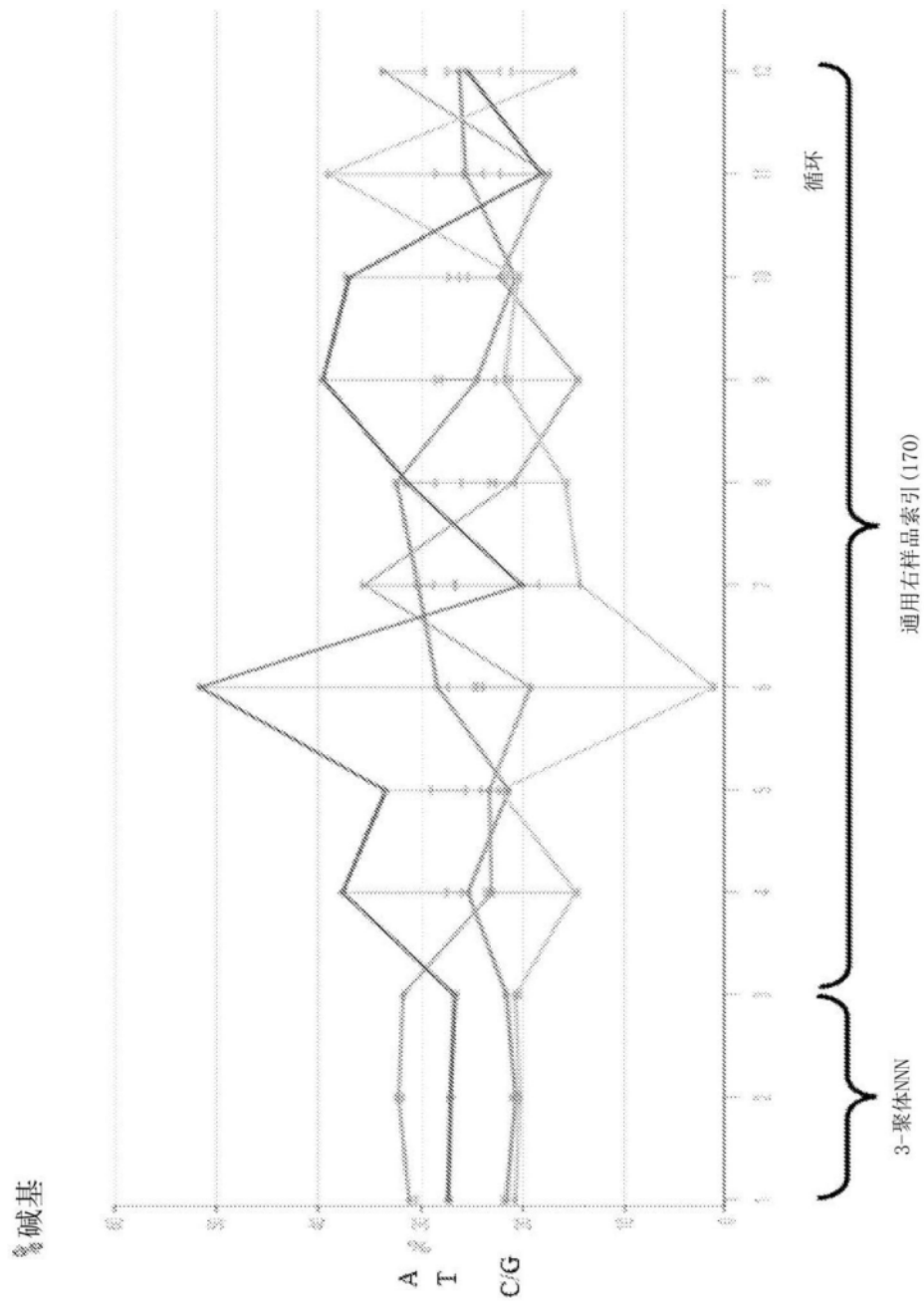


图31

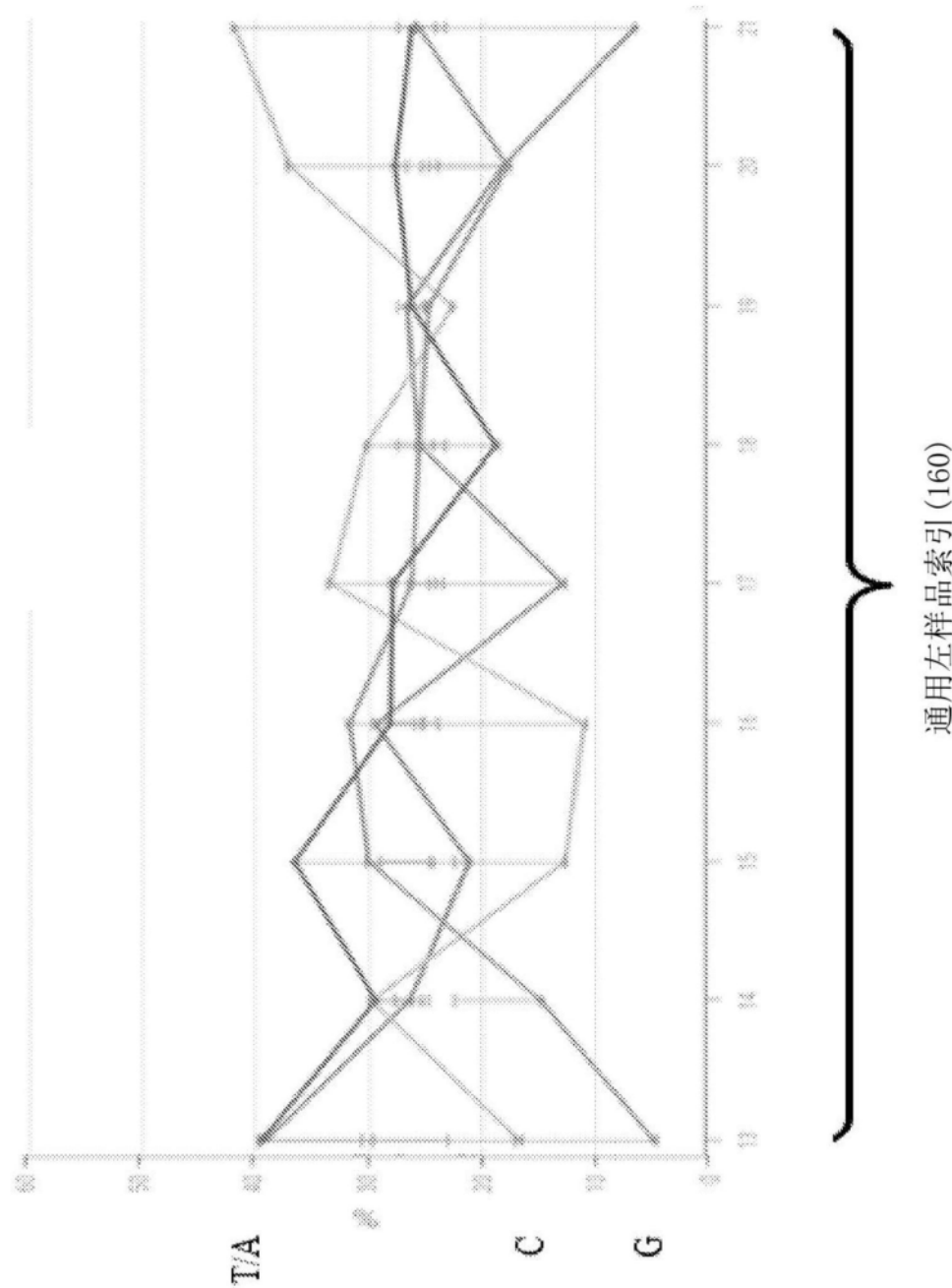


图32