(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119923478 A
(43)申请公布日 2025.05.02

(21)申请号 202380063405.7

(22)申请日 2023.07.05

(30)优先权数据
63/358,491 2022.07.05 US
63/508,833 2023.06.16 US

(85)PCT国际申请进入国家阶段日
2025.03.03

(86)PCT国际申请的申请数据
PCT/US2023/069662 2023.07.05

(87)PCT国际申请的公布数据
WO2024/011145 EN 2024.01.11

(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 J·赵 S·莱维 X·齐

(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 权陆军 彭昶

(51)Int.Cl.
C12Q 1/6855(2006.01)

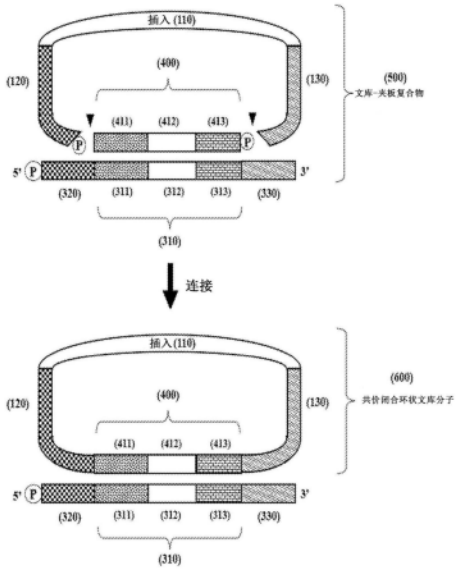
权利要求书2页 说明书104页
序列表（电子公布） 附图26页

(54)发明名称

使用双链夹板衔接子的无PCR文库制备以及使用方法

(57)摘要

本公开提供了包含核酸双链夹板衔接子的组合物,包括试剂盒,以及采用所述双链夹板衔接子的方法,例如无PCR的工作流程。所述双链夹板衔接子(200)可用于一锅多酶反应以将一个或多个新衔接子序列引入文库分子中。所述双链夹板衔接子(200)包含第一夹板链(长夹板链(300))和第二夹板链(短夹板链(400)),其中所述第一夹板链和所述第二夹板链杂交在一起以形成具有双链区和两个侧接单链区的所述双链夹板衔接子(200)。所述第二夹板链(400)携带待引入的新衔接子序列,诸如例如通用结合序列和/或索引序列。



CN 119923478 A

1.一种用于形成多个文库-夹板复合物(500)的方法,所述方法包括:

a)提供多个双链夹板衔接子(200),其中所述多个双链夹板衔接子中的各个双链夹板衔接子(200)包含与第二夹板链(400)杂交的第一夹板链(300),其中所述双链夹板衔接子包括双链区和两个侧接单链区,其中所述第一夹板链包含第一区(320)、内部区(310)和第二区(330),并且其中所述第一夹板链的所述内部区(310)与所述第二夹板链(400)杂交;以及

b)使所述多个双链夹板衔接子与多个单链核酸文库分子(100)杂交,其中各个文库分子包含目的序列(110),所述目的序列在第一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列并在第二侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列,

从而使所述多个文库分子成环以形成各自具有两个缺口的多个文库-夹板复合物(500)。

2.根据权利要求1所述的方法,其进一步包括:(c)使所述多个文库-夹板复合物(500)与连接酶接触以生成多个共价闭合环状文库分子(600)。

3.根据权利要求1或2所述的方法,其中所述杂交在适合于使所述第一夹板链的所述第一区(320)与所述文库分子的针对正向测序引物结合位点(120)的所述通用衔接子序列杂交的条件下进行。

4.根据权利要求3所述的方法,其中所述条件适合于使所述第一夹板链的所述第二区(330)与所述文库分子的针对反向测序引物结合位点(130)的所述通用衔接子序列杂交。

5.根据权利要求1至4中任一项所述的方法,其中所述第一夹板链(300)的所述内部区(310)包括至少三个子区。

6.根据权利要求5所述的方法,其中所述至少三个子区包括子区(311)、子区(312)和子区(313)。

7.根据权利要求6所述的方法,其中所述子区(311)包含针对表面捕获引物结合位点的通用衔接子序列、针对表面钉扎引物结合位点的通用衔接子序列、样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。

8.根据权利要求6或7所述的方法,其中所述子区(312)包含针对表面捕获引物结合位点的通用衔接子序列、针对表面钉扎引物结合位点的通用衔接子序列、样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。

9.根据权利要求6至8中任一项所述的方法,其中所述子区(313)包含针对表面捕获引物结合位点的通用衔接子序列、针对表面钉扎引物结合位点的通用衔接子序列、样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。

10.根据权利要求6至9中任一项所述的方法,其中所述子区(311)、(312)或(313)包含样品索引序列,并且其中所述样品索引序列包括:

- 缺少短随机序列(NNN)的样品索引序列;
- 样品索引序列和短随机序列(NNN);
- 在两侧上侧接可转换为脱碱基碱基的核苷酸碱基的样品索引;
- 至少一个可转化为脱碱基碱基的核苷酸碱基;
- 至少一种脱氧肌苷;
- 18碳间隔基;和/或

• 18碳间隔基和至少一种脱氧肌苷。

11.根据权利要求6至10中任一项所述的方法,其进一步包括:

i)在适合于使各个共价闭合环状文库分子(600)与各个固定的表面捕获引物杂交的条件下,将所述多个共价闭合环状文库分子(600)分布到支持物上,所述支持物具有固定到所述支持物的多个表面捕获引物,从而将所述多个共价闭合环状文库分子(600)固定到所述支持物;

12.根据权利要求11所述的方法,其中所述支持物进一步包含固定到所述支持物的多个表面钉扎引物。

13.根据权利要求11所述的方法,其进一步包括:

ii)在适合于使用所述多个表面捕获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(600)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,使多个经固定的共价闭合环状文库分子(600)与多个链置换聚合酶和多个核苷酸接触,

从而生成固定到所述表面捕获引物的多个核酸多联体分子。

14.根据权利要求13所述的方法,其进一步包括:

iii)对固定到所述表面捕获引物的所述多个核酸多联体分子进行测序,其中所述测序包括(i)对所述样品索引进行测序以及(ii)对所述目的序列(110)进行测序。

15.根据权利要求13所述的方法,其进一步包括:

iv)对固定到所述表面捕获引物的所述多个核酸多联体分子进行测序,其中所述测序包括(A)对一个或多个短随机序列NNN进行测序,(B)对一个或多个样品索引进行测序,以及(C)对所述目的序列(110)进行测序。

16.根据权利要求7至15中任一项所述的方法,其中所述第一夹板链(300)的所述内部区(310)包含一个样品索引。

17.根据权利要求7至15中任一项所述的方法,其中所述第一夹板链(300)的所述内部区(310)包含一个样品索引和短随机序列(NNN)。

18.根据权利要求1至17中任一项所述的方法,其中所述文库分子(100)包含选自表1的一个或多个核苷酸序列。

19.根据权利要求1至18中任一项所述的方法,其中所述第一夹板链(300)包含选自表2的一个或多个核苷酸序列。

20.根据权利要求1至19中任一项所述的方法,其中所述第二夹板链(400)包含选自表3的一个或多个核苷酸序列。

使用双链夹板衔接子的无PCR文库制备以及使用方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2022年7月5日提交的美国临时专利申请号63/358,491和2023年6月16日提交的美国临时专利申请号63/508,833的优先权和权益,这些申请各自的内容通过引用以其整体并入本文。

[0003] 电子序列表的引用

[0004] 电子序列表的内容(ELEM-013_001WO_SeqListing_ST26.xml;大小50,242个字节;以及创建日期:2023年6月30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双链夹板衔接子的组合物以及用于使用双链夹板衔接子制备核酸文库的方法。双链夹板衔接子可以与文库分子部分杂交以形成具有缺口的文库-夹板复合物,其中缺口可以连接以形成共价闭合环状分子,其可以进行下游扩增和测序工作流程。

背景技术

[0006] 从传统的Sanger式测序方法到下一代测序方法的转变已经降低了测序的成本,但是下一代测序方法仍然存在明显的局限性。在一个方面,可用的测序平台生成的测序读段虽然数量众多,但相对较短,并且可能需要通过计算重新组装成完整的目的序列。可用的组装方法可能是缓慢的、费力的、昂贵的、计算要求高的和/或不适合于类似个体(例如,病毒)的群体。对于复杂基因组的测序尤其如此。组装具有挑战性,部分是由于与短读段的组装相关的不断膨胀的测序数据集。此类数据集会给计算机集群带来很大的压力。例如,从头组装可能需要将测序读段(或由其衍生的k聚体)同时存储在随机存取存储器(RAM)中。对于大型数据集,这一要求并非微不足道。此外,即使当组装是可能的时,关键的单倍型信息通常也无法恢复。事实上,可用的技术的固有局限性阻碍了对克服现状测序技术的缺点的改进。因此,需要改进的测序方法和相关的组装技术,该改进的测序方法和相关的组装技术减少获得准确序列所需的时间和/或计算要求。

发明内容

[0007] 本公开提供了一种用于形成多个文库-夹板复合物(500)的方法,该方法包括:(a)提供多个双链夹板衔接子(200),其中多个双链夹板衔接子中的各个双链夹板衔接子(200)包含与第二夹板链(400)杂交的第一夹板链(300),其中双链夹板衔接子包含双链区和两个侧接单链区,其中第一夹板链包含第一区(320)、内部区(310)和第二区(330),并且其中第一夹板链的内部区(310)与第二夹板链(400)杂交;以及使多个双链夹板衔接子与多个单链核酸文库分子(100)杂交,其中各个文库分子包含目的序列(110),该目的序列在第一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列并在第二侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列,从而使多个文库分子成环以形成各自具有两个缺

口的多个文库-夹板复合物(500)。

[0008] 在一些实施例中,该方法进一步包括:(c)使多个文库-夹板复合物(500)与连接酶接触以生成多个共价闭合环状文库分子(600)。

[0009] 在一些实施例中,杂交在适合于使第一夹板链的第一区(320)与文库分子的针对正向测序引物结合位点(120)的通用衔接子序列杂交的条件下进行。在一些实施例中,该条件适合于使第一夹板链的第二区(330)与文库分子的针对反向测序引物结合位点(130)的通用衔接子序列杂交。

[0010] 在一些实施例中,第一夹板链(300)的内部区(310)包括至少三个子区。在一些实施例中,至少三个子区包括子区(311)、子区(312)和子区(313)。在一些实施例中,子区(311)包含针对表面捕获引物结合位点的通用衔接子序列、针对表面钉扎引物结合位点的通用衔接子序列、样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。在一些实施例中,子区(312)包含针对表面捕获引物结合位点的通用衔接子序列、针对表面钉扎引物结合位点的通用衔接子序列、样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。在一些实施例中,子区(313)包含针对表面捕获引物结合位点的通用衔接子序列、针对表面钉扎引物结合位点的通用衔接子序列、样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。

[0011] 在一些实施例中,子区(311)、(312)或(313)包含样品索引序列,并且其中样品索引序列包括:缺乏短随机序列(NNN)的样品索引序列;样品索引序列和短随机序列(NNN);在两侧上侧接可转换为脱碱基碱基的核苷酸碱基的样品索引;至少一个可转化为脱碱基碱基的核苷酸碱基;至少一种脱氧肌苷;18碳间隔基;和/或18碳间隔基和至少一种脱氧肌苷。

[0012] 在一些实施例中,该方法进一步包括:在适合于使各个共价闭合环状文库分子(600)与各个固定的表面捕获引物杂交的条件下,将多个共价闭合环状文库分子(600)分布到支持物上,该支持物具有固定到支持物的多个表面捕获引物,从而将多个共价闭合环状文库分子(600)固定到支持物。在一些实施例中,支持物进一步包含固定到支持物的多个表面钉扎引物。

[0013] 在一些实施例中,该方法进一步包括:在适合于使用多个表面捕获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(600)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个经固定的共价闭合环状文库分子(600)与多个链置换聚合酶和多个核苷酸接触,从而生成固定到表面捕获引物的多个核酸多联体分子。

[0014] 在一些实施例中,该方法进一步包括:iii)对固定到表面捕获引物的多个核酸多联体分子进行测序,其中该测序包括(i)对样品索引进行测序以及(ii)对目的序列(110)进行测序。

[0015] 在一些实施例中,该方法进一步包括:iv)对固定到表面捕获引物的多个核酸多联体分子进行测序,其中该测序包括(A)对一个或多个短随机序列NNN进行测序,(B)对一个或多个样品索引进行测序,以及(C)对目的序列(110)进行测序。

[0016] 在一些实施例中,第一夹板链(300)的内部区(310)包含一个样品索引。

[0017] 在一些实施例中,第一夹板链(300)的内部区(310)包含一个样品索引和短随机序列(NNN)。

[0018] 在一些实施例中,文库分子(100)包含选自表1的一个或多个核苷酸序列。

[0019] 在一些实施例中,第一夹板链(300)包含选自表2的一个或多个核苷酸序列。

[0020] 在一些实施例中,第二夹板链(400)包含选自表3的一个或多个核苷酸序列。

附图说明

[0021] 在所附权利要求书中对本公开的特征进行了具体阐述。通过参考阐述了说明性实施例的以下详细说明,将获得对本公开的特征和优点的更好理解,在所述实施例中利用了本公开的原理,并且在其附图中:

[0022] 图1是示出示例性线性核酸文库分子(100)的示意图,该线性核酸文库分子包含在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列的插入区(110)(例如,目的序列),并且插入区(110)在另一侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列。

[0023] 图2是示例性双链夹板衔接子(200)的示意图,该双链夹板衔接子包含与第二夹板链(短链(400))杂交的第一夹板链(长链(300))。

[0024] 图3是示出示例性文库成环工作流程的示意图,其包括使线性单链文库分子(100)与双链夹板衔接子(200)杂交,从而使文库分子成环以形成具有两个缺口的文库-夹板复合物(500)。

[0025] 图4是示出示例性连接反应的示意图,其包括对文库-夹板复合物(500)中的缺口进行酶促连接反应,从而闭合缺口以形成与第一夹板链(300)杂交的共价闭合环状文库分子(600)。

[0026] 图5是示出与扩增引物杂交的示例性共价闭合环状文库分子(600)的示意图。虚线代表新生的延伸产物。

[0027] 图6是示出第一夹板链(300;长夹板链)的核酸序列的几个实施例的示意图,第一夹板链包含:外部第一区(320);外部第二区(330);以及三个内部子区,该三个内部子区包括子区(311)、子区(312)和子区(313)。该示意图还示出了第二夹板链(400;短夹板链)的示例性核酸序列,第二夹板链包含:三个子区,该三个子区包括:子区(411);子区(412);和子区(413)。

[0028] 图7是示出第一夹板链(300;长夹板链)的示例性核酸序列的示意图,第一夹板链包含:外部第一区(320);外部第二区(330);以及两个内部子区,包括子区(311)和子区(313)。该示意图还示出了第二夹板链(400;短夹板链)的示例性核酸序列,第二夹板链包含:三个子区,该三个子区包括:子区(411);子区(412);和子区(413)。子区(412)处的环表示由于缺乏第一夹板链(300)的子区(312)而形成的环。

[0029] 图8A至图8D是示出各自携带针对正向测序引物结合位点(120)的通用衔接子序列的双链核酸衔接子的各种实施例的一组示意图。图8A是携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的双链衔接子的示意图。图8B是携带针对正向测序引物结合位点(120)的通用衔接子序列的截短序列的双链衔接子的示意图。图8C是具有5'突出端的双链衔接子的示意图,其中衔接子链中的一者携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列,而另一衔接子链携带针对正向测序引物结合位点(120)的通用衔接子序列的截短序列。图8D是具有3'突出端的双链衔接子的示意图,其中衔接子链中的一者携带针对正向测序引物结合位点(120)的通用衔接子序列的截短序列,而

另一衔接子链携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列。

[0030] 图9A至图9D是示出各自携带针对反向测序引物结合位点(130)的通用衔接子序列的双链核酸衔接子的各种实施例的一组示意图。图9A是携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的双链衔接子的示意图。图9B是携带针对反向测序引物结合位点(130)的通用衔接子序列的截短序列的双链衔接子的示意图。图9C是具有5'突出端的双链衔接子的示意图,其中衔接子链中的一者携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列,而另一衔接子链携带针对反向测序引物结合位点(130)的通用衔接子序列的截短序列。图9D是具有3'突出端的双链衔接子的示意图,其中衔接子链中的一者携带针对反向测序引物结合位点(130)的通用衔接子序列的截短序列,而另一衔接子链携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列。

[0031] 图10A至图10D是示出Y形衔接子的各种实施例的一组示意图,每个Y形衔接子包含杂交在一起并具有双链退火区和错配部分的两个寡核苷酸。图10A是Y形衔接子的示意图,该Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸和携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。图10B是Y形衔接子的示意图,该Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的截短序列的第一寡核苷酸和携带针对反向测序引物结合位点(130)的通用衔接子序列的截短序列的第二寡核苷酸。图10C是Y形衔接子的示意图,该Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸和携带针对反向测序引物结合位点(130)的通用衔接子序列的截短序列的第二寡核苷酸。图10D是Y形衔接子的示意图,该Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的截短序列的第一寡核苷酸和携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。

[0032] 图11A至图11B是示出转座体的实施例的一组示意图。图11A是两个转座体的示意图,其中第一个转座体(顶部)包含与双链多核苷酸结合的转座酶,该双链多核苷酸包含转座子末端序列和针对正向测序引物结合位点(120)的通用衔接子序列。转座子末端序列特异性结合转座酶。第二转座体(底部)包含与双链多核苷酸结合的转座酶,该双链多核苷酸包含转座子末端序列和针对反向测序引物结合位点(130)的通用衔接子序列。转座子末端序列特异性结合转座酶。图11B是包含与第一双链多核苷酸结合的转座酶的示例性转座体的示意图,第一双链多核苷酸包含转座子末端序列和针对正向测序引物结合位点(120)的通用衔接子序列,并且转座酶与包含转座子末端序列和针对反向测序引物结合位点(130)的通用衔接子序列的第二双链多核苷酸结合。转座子末端序列特异性结合转座酶。

[0033] 图12是示出包含与Y形衔接子结合的转座酶的转座体的实施例的示意图。每个Y形衔接子包含杂交在一起并具有双链退火区和错配部分的两个寡核苷酸。

[0034] 图13是示出生成双链线性核酸文库分子的示例性衔接子连接工作流程的示意图。双链核酸片段在一侧酶促连接至双链衔接子,该双链衔接子携带针对正向测序引物结合位点(120)的通用衔接子序列。双链核酸片段在另一侧上酶促连接至双链衔接子,该双链衔接子携带针对反向测序引物结合位点(130)的通用衔接子序列。

[0035] 图14是示出生成双链线性核酸文库分子的示例性衔接子连接工作流程的示意图。双链核酸片段在一侧酶促连接至第一Y形衔接子,第一Y形衔接子包含携带针对正向测序引

物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。双链核酸片段在另一侧上酶促连接至第二Y形衔接子,第二Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。

[0036] 图15是示出生成双链线性核酸文库分子的示例性衔接子连接工作流程的示意图。双链核酸片段在一侧酶促连接至第一Y形衔接子,第一Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。双链核酸片段在另一侧上酶促连接至第二Y形衔接子,第二Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。使用具有与错配部分中的一者(例如,(130))杂交的区域的引物对所得双链线性核酸文库分子进行引物延伸,并且引物还在另一端携带样品索引序列和通用表面引物结合位点(SurfPBS)。使用杂交的引物作为模板进行引物延伸反应,以生成延伸的文库分子,其包含:正向测序引物结合位点(120);插入序列(110);反向测序引物结合位点(130);样品索引序列;以及通用表面引物结合位点(SurfPBS)。

[0037] 图16是示出生成双链线性核酸文库分子的示例性衔接子连接工作流程的示意图。双链核酸片段在一侧酶促连接至第一Y形衔接子,第一Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。双链核酸片段在另一侧上酶促连接至第二Y形衔接子,第二Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。所得双链线性核酸文库分子与具有3'突出端和平端的线性双链衔接子杂交。3'突出端包含可与具有反向测序引物结合位点(130)的错配部分杂交的序列,其生成具有缺口(实心三角形)的部分双链区。可以连接缺口,并且可以去除未连接的链。

[0038] 图17是使用多个转座体的示例性标签片段化(tagmentation)工作流程的示意图。使输入的双链DNA与多个转座体接触,其中各个转座体包含与第一双链多核苷酸结合的转座酶,第一双链多核苷酸包含转座子末端序列和针对正向测序引物结合位点(120)的通用衔接子序列,并且转座酶与包含转座子末端序列和针对反向测序引物结合位点(130)的通用衔接子序列的第二双链多核苷酸结合。

[0039] 图18是示例性文库成环工作流程的示意图,其包括使线性单链文库分子(100)与第一夹板链(200)杂交,从而使文库分子成环以形成在文库分子的末端之间具有间隙的文库单夹板复合物(700)。线性核酸文库分子(100)包含在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列的插入区(110)(例如,目的序列),并且插入区(110)在另一侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列。第一夹板链(200)包含与线性单链文库分子(100)一端上的针对正向测序引物结合位点(120)的通用衔接子序列杂交的第一区(320),并且第一夹板链包含与线性单链文库分子另一端上的针对反向测序引物结

合位点(130)的通用衔接子序列杂交的第二区(330)。用聚合酶催化的填充反应和酶促连接反应闭合该间隙,以生成共价闭合环状分子。

[0040] 图19是示出包括短3聚体随机序列(NNN)的样品索引序列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%。

[0041] 图20是示出缺乏短3聚体随机序列(NNN)的样品索引序列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%。

[0042] 图21是示例性低结合支持物的示意图,其包含玻璃基底和共价或非共价粘附至玻璃的亲水涂层的交替层,并且其进一步包含用作寡核苷酸引物(例如,捕获寡核苷酸和环化寡核苷酸)的附接位点的化学反应性官能团。

[0043] 图22是多价分子的各种示例性构型的示意图。左:具有星暴(starburst)或helter-skelter构型的多价分子的示意图。中心:具有树枝状高分子构型的多价分子的示意图。右: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’。

[0044] 图23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。

[0045] 图24是包含附接至多个核苷酸臂的树枝状高分子核心的示例性多价分子的示意图。

[0046] 图25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。

[0047] 图26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。

[0048] 图27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(包括11-原子接头、16-原子接头、23-原子接头和N3接头)的化学结构(底部)。

[0049] 图28示出了各种示例性接头(包括接头1-9)的化学结构。

[0050] 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。

[0051] 图30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。

[0052] 图31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。

[0053] 图32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。

具体实施方式

[0054] 定义:

[0055] 本文所提供的标题不是对本公开的各个方面的限制,这些方面可以通过整体参考本说明书来理解。

[0056] 除非另有定义,否则本文所使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。一般而言,与本文所描述的分子生物学、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术语。

本文所描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描述来执行。例如,参见Sambrook等人,MolecularCloning:ALaboratoryManual(第三版,Cold SpringHarborLaboratoryPress,Cold Spring Harbor,N.Y.2000)。另请参见Ausubel等人,Current Protocols in Molecular Biology,GreenePublishingAssociates(1992)。与本文所描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命名法。

[0057] 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。除非明确且肯定地限于一个所指对象,否则单数形式“一个/一种(a/an)”和“所述(the)”和任何词的单数用途包括多个所指对象。

[0058] 应当理解,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。

[0059] 本文所使用的术语“和/或”应被视为意指指定特征或组分中的每一者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)。

[0060] 如本文和所附权利要求中所使用的,如本文所使用的术语“包括(comprising)”、“包含(including)”、“具有(having)”和“含有(containing)”以及其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所列项目中的其他项目。应当理解,每当在本文中用语言“包括”描述方面时,也提供了以“由……组成”和/或“基本上由……组成”描述的其他类似方面。

[0061] 如本文所使用的,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。可替代地,“约”或“大约”可意指至多10%(即,±10%)或更大的范围,这取决于测量系统的限制。例如,约5mg可包括介于4.5mg与5.5mg之间的任何数值。此外,特别是对于生物系统或过程,这些术语可以表示高达一个数量级或高达值的5倍。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,这些范围和/或子范围可以包括这些范围和/或子范围的端点。

[0062] 本文所使用的术语“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。多肽可以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。

[0063] 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中获得。细胞生物样品包括新鲜的、冷冻的、新鲜冷冻的或存档的样品(例如,福尔马林固定的石蜡包埋的;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。

[0064] 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)收集要从其提取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以去除去污剂、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。

[0065] 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,这种核苷酸聚合可以以模板依赖性方式发生。通常,聚合酶包括在其处可发生核苷酸结合和/或核苷酸聚合的催化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。聚合酶可以包括但不限于:天然存在的聚合酶和其任何亚基和截短物、突变体聚合酶、变体聚合酶、重组、融合或以其他方式经工程化的聚合酶、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、衍生物或其片段(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,聚合酶可从细胞中分离或者使用重组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。聚合酶可以来源于原核生物、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。

[0066] 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(包括核酸外切酶减突变体、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但不限于phi29 DNA聚合酶、Bst DNA聚合酶大片段、Bsu DNA聚合酶大片段(exo-)、Bca DNA聚合酶(exo-)、大肠杆菌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Thermo Fisher Scientific™)、或嵌合QualiPhi™DNA聚合酶(例如,来自4basebio™)。

[0067] 本文所使用的术语“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的形式。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。核酸包括核苷酸的聚合物,其中核苷酸可包含天然或非天然碱基和/或糖。核酸包含天然存在的核苷间键,例如且不限于磷酸二酯键。核酸可包含非天然核苷间键,包括硫代磷酸酯、含硫磷酸酯和/或肽核酸(PNA)键。在一些实施例中,核酸包含一种类型的多核苷酸或者两种或更多种不同类型的多核苷酸的混合物。

[0068] 如本文所使用的术语“可操作地连接”和“可操作地接合”或相关术语是指组分的并置。经并置的组分可以共价连接在一起。例如,两种核酸组分可以以酶促方式连接在一起,其中将该两种组分接合在一起的键包括磷酸二酯键。第一核酸组分和第二核酸组分可以连接在一起,其中第一核酸组分可以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键形成具有可以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,其中该连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,转基因可操作地连接至影响转基因的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,载体包含至少一种宿主细胞调控序列,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。

[0069] 术语“连接的”、“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,去除链终止部分);洗涤;去除;流动;检测;成像和/或识别。这种键可包括例如共价键、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、机械键合等。在一些实施例中,这种键在分子内出现,例如将单链或双链线性核酸分子的端部连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;蛋白质与可检测报告基因部分之间的连接;核苷酸与可检测报告基因部分之间的连接;等等。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)中找到。

[0070] 如本文所使用的术语“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。引物可以包括天然核苷酸和/或核苷酸类似物。引物可以为重组核酸分子。引物可以具有任何长度,但通常范围为4至50个核苷酸。典型的引物包含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。可替代地,引物的3’端可缺乏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以用可检测报告基因部分来标记。引物可以在溶液中(例如,可溶性引物)或可以固定到支持物(例如,捕获引物)。

[0071] 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。模板核酸可以从天然存在的来源、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。

[0072] 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之间进行氢键合以形成具有双链体区的自杂交分子。杂交可包括Watson-Crick或Hoogstein结合以形成双链体双链核酸或核酸分子内的双链区。双链核酸或单个核酸的两个不同区可以是完全互补的或部分互补的。互补核酸链不需要跨其整个长度彼此杂交。互补碱基配对可以是标准A-T或C-G碱基配对或者可以是碱基配对相互作用的其他形式。双链体核酸可以包括错配的经碱基配对的核苷酸。

[0073] 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。可以用天然核苷酸和/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。可以使用延伸核酸分子的任何合适的方法,这些方法包括由DNA聚合酶或RNA聚合酶催化的引物延伸。

[0074] 术语“核苷酸”和相关术语是指包括芳香族碱基、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。

[0075] 核苷酸(和核苷)通常包含杂环碱基,包括通常在核酸中发现的经取代或未经取代的含氮母体杂芳香族环,包括天然存在的、经取代的、经修饰的或经工程化的变体或其类似物。核苷酸(或核苷)的碱基能够与适当的互补碱基形成沃森-克里克和/或胡斯坦氢键。示例性的碱基包括但不限于嘌呤和嘧啶,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-甲基硫代腺嘌呤(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-硫代鸟嘌呤(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);吲哚类,诸如硝基吲哚和4-甲基吲哚;吡咯类,诸如硝基吡咯;水粉蕈素(nebularine);肌苷;羟甲基胞嘧啶;5-甲基胞嘧啶;碱基(Y);以及经甲基化、糖基化和酰化的碱基部分;等等。另外的示例性碱基可见于Fasman,1989,在“Practical Handbook ofBiochemistry and Molecular Biology”中,第385至394页,CRC Press,Boca Raton,Fla。

[0076] 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)。糖部分包括: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'-巯基核糖基;2'-烷基硫代核糖基;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'-巯基核糖基;3'-烷基硫代核糖基碳环;无环的或其他修饰的糖。

[0077] 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,其中该链通常经由酯键或磷酰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,其中磷原子通过居间的O、S、NH、亚甲基或亚乙基连接在一起。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,这些类似物包括磷酰胺、硫代磷酸酯、二硫代磷酸酯和O-甲基亚磷酰胺基团。

[0078] 如本文所使用,"核苷酸单元"或"核苷酸部分"是指包含碱基、糖和至少一个磷酸酯基团的核苷酸(例如,dATP、dTTP、dGTP、dCTP或dUTP)或其类似物。核苷酸单元可以附接至本文所述的测序反应中使用的多价分子。一般而言,附接至同一多价分子的所有核苷酸单元将具有相同的同一性(例如,全部为A、全部为T、全部为C或全部为G),尽管本领域技术人员将理解,可能存在其中多价分子包含不同同一性的核苷酸单元将是有利的情况。

[0079] 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。使用滚环扩增生成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具有至少四个连续鸟嘌呤(例如,4、5、6、7、8、9、10、11、12或超过12个鸟嘌呤)的多个压实寡核苷酸的存在下进行。滚环扩增反应可生成多联体,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。不希望受理论束缚,假设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测序工作流程中任一者的试剂的重复流动。

[0080] 当用于指代核酸时,术语"扩增(amplify)"、"扩增(amplifying)"、"扩增(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。

[0081] 术语“报告基因部分(reportermoiety)”、“报告基因部分(reportermoieties)”或相关术语是指生成或导致生成可检测信号的化合物。报告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。报告基因部分生成由化学或物理变化(例如,热、光、电、pH、盐浓度、酶促活性或邻近事件)引起的可检测信号。邻近事件包含两个报告基因部分相互靠近、相互缔合或相互结合。本领域技术人员众所周知对报告基因部分进行选择,使得每个报告基因部分在与其他报告基因部分可区分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告基因部分的存在。可以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告基因部分。报告基因部分可连接(例如,可操作地连接)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)。

[0082] 报告基因部分(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、碳酸肼甲硫乙酰氨基荧光素,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BODIPYFLC3-SE、BODIPY 530/550C3、BODIPY 530/550C3-SE、BODIPY 530/550C3酰肼、BODIPY 493/503C3酰肼、BODIPY FL C3酰肼、BODIPY FL IA、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,LuciferYellow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、Cy3、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,AlexaFluor染料、DyLight染料、Atto染料、LightCycler Red染料、CALFlour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 ofFluorescence 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,或其任意组合。花青染料可以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吲哚、苯并吲哚鎓、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(其可以包括1-[6-(2,5-二氧代吡咯烷-1-基氧基)-6-氧代己基]-2-(3-{1-[6-(2,5-二氧代吡咯烷-1-基氧基)-6-氧代己基]-3,3-二甲基-1,3-二氢-2H-吲哚-2-亚基}丙-1-烯-1-基)-3,3-二甲基-3H-吲哚鎓或1-[6-(2,5-二氧代吡咯烷-1-基氧基)-6-氧代己基]-2-(3-{1-[6-(2,5-二氧代吡咯烷-1-基氧基)-6-氧代己基]-3,3-二甲基-5-磺基-1,3-二氢-2H-吲哚-2-亚基}丙-1-烯-1-基)-3,3-二甲基-3H-吲哚鎓-5-磺酸酯)、Cy5(其可以包括1-(6-((2,5-二氧代吡咯烷-1-基)氧基)-6-氧代己基)-2-((1E,3E)-5-((E)-1-(6-

((2,5-二氧代吡咯烷-1-基)氧基)-6-氧代己基)-3,3-二甲基-5-吲哚啉-2-亚基)五-1,3-二烯-1-基)-3,3-二甲基-3H-吲哚-1-鎓或1-(6-((2,5-二氧代吡咯烷-1-基)氧基)-6-氧代己基)-2-((1E,3E)-5-((E)-1-(6-((2,5-二氧代吡咯烷-1-基)氧基)-6-氧代己基)-3,3-二甲基-5-磺基吲哚啉-2-亚基)五-1,3-二烯-1-基)-3,3-二甲基-3H-吲哚-1-鎓-5-磺酸酯)和Cy7(其可以包括1-(5-羧基戊基)-2-[(1E,3E,5E,7Z)-7-(1-乙基-1,3-二氢-2H-吲哚-2-亚基)七-1,3,5-三烯-1-基]-3H-吲哚鎓或1-(5-羧基戊基)-2-[(1E,3E,5E,7Z)-7-(1-乙基-5-磺基-1,3-二氢-2H-吲哚-2-亚基)七-1,3,5-三烯-1-基]-3H-吲哚鎓-5-磺酸酯),其中"Cy"代表'花青',并且第一个数字表示两个假吲哚基团之间的碳原子的数量。Cy2是一种噁唑衍生物而不是假吲哚,并且苯并衍生的Cy3.5、Cy5.5和Cy7.5是此规则的例外。

[0083] 在一些实施例中,报告基因部分可以为FRET对,例如,使得可在单个激发和成像步骤下进行多个分类。如本文所使用的,FRET可以包括激发交换(Forster)转移或电子交换(Dexter)转移。

[0084] 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。待沉积到支持物上的生物学分子的实例包括核酸(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。

[0085] 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、凸面的或其任意组合。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。

[0086] 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,包括例如凸块、经蚀刻、孔、三维支架或它们的任意组合。

[0087] 在一些实施例中,支持物包含具有任何形状的珠,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。

[0088] 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。

[0089] 在一些方面,本公开提供了固定到支持物的多个(例如,两个或更多个)核酸模板分子。在一些实施例中,经固定的多个核酸模板分子具有相同的序列。在一些实施例中,经固定的多个核酸模板分子具有不同的序列。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到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单独的核酸模板分子被固定到支持物上的位点。

[0090] 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。这些位点可以是离散的并且被间质区分开。在一些实施例中,支持物上预定的位点可以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,多个预定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。不同位点对之间的间距可以相同或可以不同。在一些实施例中,支持物包括至少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，其中所述位点位于支持物上预先确定的位置处。在一些实施例中，支持物上的多个预定位点(例如，10^2至10^{15}个位点或更多，例如，10^2个位点、10^3个位点、10^4个位点、10^5个位点、10^6个位点、10^7个位点、10^8个位点、10^9个位点、10^{10}个位点、10^{11}个位点、10^{12}个位点、10^{13}个位点、10^{14}个位点、10^{15}个位点或更多)用核酸模板分子固定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处，例如固定在10^2至10^{15}个位点或更多(例如，10^2至10^{15}个位点或更多，例如，10^2个位点、10^3个位点、10^4个位点、10^5个位点、10^6个位点、10^7个位点、10^8个位点、10^9个位点、10^{10}个位点、10^{11}个位点、10^{12}个位点、10^{13}个位点、10^{14}个位点、10^{15}个位点或更多)处。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。

[0091] 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称为在其上具有随机定位的位点的支持物。在此类实施例中，支持物上随机定位的位点的位置不是预定的。因此，多个随机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，支持物包括至少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，其中所述位点随机地定位在支持物上。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至10^{15}个位点或更多，例如，10^2至10^{15}个位点或更多，例如，10^2个位点、10^3个位点、10^4个位点、10^5个位点、10^6个位点、10^7个位点、10^8个位点、10^9个位点、10^{10}个位点、10^{11}个位点、10^{12}个位点、10^{13}个位点、10^{14}个位点、10^{15}个位点或更多)用核酸模板分子固定。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，核酸模板固定在多个随机定位的位点处，例如固定在10^2至10^{15}个位点或更多，例如，10^2至10^{15}个位点或更多，例如，10^2个位点、10^3个位点、10^4个位点、10^5个位点、10^6个位点、10^7个位点、10^8个位点、10^9个位点、10^{10}个位点、10^{11}个位点、10^{12}个位点、10^{13}个位点、10^{14}个位点、10^{15}个位点或更多处。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。

[0092] 在一些实施例中，支持物上的多个经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，使得支持物上多个经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，多个经固定的表面捕获引物的流体连通可用于在该多个经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)。

[0093] 在一些实施例中，支持物上的多个经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,多个经固定的核酸簇的流体连通可用于在多个经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。

[0094] 在一些实施例中,术语“经固定的”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可包括核酸多联体(例如,核酸簇)。核酸分子可以固定在支持物上的预定随机位置处。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。

[0095] 在一些实施例中,术语“经固定的”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、或埋入由支持物上的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。

[0096] 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),并且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),该核酸扩增反应包括以下任一者或任意组合: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。

[0097] 术语“表面引物”和相关术语是指固定到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到支持物。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)。可替代地,表面捕获引物的内部或3’端可以固定到支持物。

[0098] 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至50个核苷酸,或50至100个核苷酸,或100至150个核苷酸,或更长的长度。

[0099] 表面捕获引物可以具有末端3’核苷酸,该末端3’核苷酸具有可延伸用于核苷酸聚合(例如,聚合酶催化的聚合)的糖3’OH部分。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。可使用去阻断剂去除(例如,去阻断)3’链终止部分,以将3’端转化为可延伸的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、缩醛基团或甲硅烷基基团。叠氮型链终止部分包括叠氮基团、叠氮基基团和叠氮基甲基基团。去封闭

剂的实例包括膦化合物,例如三(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团、叠氮基基团和叠氮基甲基基团。去封闭剂的实例包括四(三苯基膦)钯(0)($Pd(PPh_3)_4$)与哌啶,或与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。

[0100] 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测序方法。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群落(polony)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。

[0101] 在一些方面,本公开提供了各种试剂以及采用这些试剂进行核酸变性(去杂交)和测序的方法。各种试剂可以包括至少一种pH缓冲剂。本文列出了示例性非限制性pH缓冲剂的全名。

[0102] 术语“Tris”是指pH缓冲剂三(羟甲基)-氨基甲烷。术语“Tris-HCl”是指pH缓冲剂三(羟甲基)-氨基甲烷盐酸盐。术语“Tris-乙酸盐”是指包含三(羟甲基)-氨基甲烷的乙酸盐的pH缓冲剂。

[0103] 术语“Tricine”是指pH缓冲剂N-[三(羟甲基)甲基]甘氨酸。

[0104] 术语“Bicine”是指pH缓冲剂N,N-双(2-羟乙基)甘氨酸。

[0105] 术语“Bis-Tris丙烷”是指pH缓冲剂1,3双[三(羟甲基)甲基氨基]丙烷

[0106] 术语“HEPES”是指pH缓冲剂4-(2-羟乙基)-1-哌嗪乙烷磺酸。

[0107] 术语“MES”是指pH缓冲剂2-(N-吗啉代)乙烷磺酸)。

[0108] 术语“MOPS”是指pH缓冲剂3-(N-吗啉代)丙烷磺酸。

[0109] 术语“MOPSO”是指pH缓冲剂3-(N-吗啉代)-2-羟基丙烷磺酸。

[0110] 术语“BES”是指pH缓冲剂N,N-双(2-羟乙基)-2-氨基乙烷磺酸。

[0111] 术语“TES”是指pH缓冲剂2-[(2-羟基-1,1双(羟甲基)乙基)氨基]乙烷磺酸)。

[0112] 术语“CAPS”是指pH缓冲剂3-(环己基氨基)-1-丙烷磺酸。

[0113] 术语“TAPS”是指pH缓冲剂N-[三(羟甲基)甲基]-3-氨基丙烷磺酸。

[0114] 术语“TAPSO”是指pH缓冲剂N-[三(羟甲基)甲基]-3-氨基-2-羟基丙烷磺酸。

[0115] 术语“ACES”是指pH缓冲剂N-(2-乙酰胺基)-2-氨基乙烷磺酸。

[0116] 术语“PIPES”是指pH缓冲剂哌嗪-1,4-双(2-乙烷磺酸)。

[0117] 术语“乙醇胺”是指也称为2-氨基乙醇的pH缓冲剂。

[0118] 简介:双链夹板衔接子

[0119] 在一些方面,本公开提供了包含核酸双链夹板衔接子的组合物,包括试剂盒,以及采用双链夹板衔接子的方法。

[0120] 所述双链夹板衔接子(200)可用于一锅多酶反应以将一个或多个新衔接子序列引入文库分子中。在一些实施例中,双链夹板衔接子(200)包含第一夹板链(长夹板链(300))和第二夹板链(短夹板链(400))。在某些实施例中,第一夹板链和第二夹板链杂交在一起以形成具有双链区和两个侧接单链区的双链夹板衔接子(200)(例如,参见图2和3)。在一些实施例中,第一夹板链(长夹板链(300))和第二夹板链(短夹板链(400))沿它们的整个长度彼此杂交。在一些实施例中,第一夹板链(长夹板链(300))和第二夹板链(短夹板链(400))包含未杂交在一起的部分。第二夹板链(400)可以携带待引入的新衔接子序列,诸如例如新通用结合序列和/或新索引序列。第一夹板链可包含第一区(320)、内部区(310)和第二区(330)。第一夹板链的内部区(310)可以与第二夹板链(400)杂交。在一些实施例中,双链夹板式衔接子的两个侧接单链区(例如,(320)和(330))被设计成与具有目的序列(110)的单链线性文库分子(100)的端部处的通用衔接子序列杂交。例如,第一夹板链的第一区(320)可以与文库分子的一端(120)杂交,并且第一夹板链的第二区(330)可以与文库分子的另一端(130)杂交,从而使文库分子成环以生成包括两个缺口的文库-夹板复合物(500)(例如,参见图3)。

[0121] 在一些实施例中,线性核酸文库分子(100)包含在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列的插入区(110)(例如,目的序列),并且插入区(110)在另一侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列。在一些实施例中,双链夹板衔接子(200)包含与第二夹板链(短链(400))杂交的第一夹板链(长链(300))。第一夹板链可以包含与在线性单链文库分子(100)一端上的针对正向测序引物结合位点(120)的通用衔接子序列杂交的第一区(320),并且第一夹板链可以包含与在线性单链文库分子另一端上的针对反向测序引物结合位点(130)的通用衔接子序列杂交的第二区(330)。第一夹板链的内部区(310)可以与第二夹板链(400)杂交。缺口可被酶促连接以生成共价闭合环状分子(600),其中第二夹板链(400)在两端处共价接合至文库分子,从而将新的衔接子序列引入文库分子中(例如,参见图4)。连接反应可以将来自第二夹板链(400)的序列接合至文库分子(100)的端部。

[0122] 在一些实施例中,第二夹板链(400)的子区中的任一者可设计成与扩增引物杂交。在一些实施例中,扩增引物具有可延伸的3’端,其可用于引发引物延伸反应,例如,如图5所示。在一些实施例中,扩增引物可以是可溶性引物或固定到支持物(例如,表面捕获引物)。在一些实施例中,扩增引物可用于进行滚环扩增反应以生成与共价闭合环状文库分子(600)互补的核酸多联体分子。

[0123] 因此,本文所描述的双链夹板衔接子和方法可用于将任何线性文库分子转化为共价闭合环状分子。第二夹板链(400)可包括至少一种新的通用衔接子序列(例如,新的表面引物序列),从而使得共价闭合环状分子(600)能够结合至具有固定在其上的多个表面捕获引物的支持物。第二夹板链(400)中的新的通用衔接子序列允许在扩增和测序工作流程中使用共价闭合环状分子(600)。

[0124] 本文描述的方法还提供了采用连接反应而不是间隙填充反应来引入新衔接子序列的优点。连接反应可用低至0.25pmol的文库分子产生高效的环化。

[0125] 本文描述的方法可以手动进行或易于自动化进行,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,无需居间的酒精沉淀或有机萃取。

[0126] 双链夹板衔接子

[0127] 在一些方面,本公开提供了核酸双链夹板衔接子(200),其包含:(i)第一夹板链(长夹板链(300)),该第一夹板链与(ii)第二夹板链(短夹板链(400))杂交(例如,参见图2和3)。第一夹板链可包含第一区(320)、内部区(310)和第二区(330)。第一夹板链的内部区(310)可以与第二夹板链(400)杂交以形成具有双链区和两个侧接单链区的双链夹板衔接子(200)。双链夹板衔接子(200)的两个侧接单链区可以被设计成与线性核酸文库分子(100)的末端序列杂交。线性核酸文库分子的末端序列可以分别包含第一(120)和第二(130)序列通用衔接子序列。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正向(120)和反向(130)测序引物结合位点的结合序列。

[0128] 如图2所示,在一些实施例中,第一夹板链包含与在线性单链文库分子的一端上的序列杂交的第一区(320)和与在线性单链文库分子的另一端上的序列杂交的第二区(330)。第一夹板链的内部区(310)可以与第二夹板链(400)杂交。第一夹板链(300)的内部区(310)可包括至少三个子区,包括子区(311)、(312)和(313)。第二夹板链(400)可包括至少三个子区,包括子区(411)、(412)和(413)。例如,子区(311)与子区(411)杂交。在另一实例中,子区(312)与子区(412)杂交。在另一实例中,子区(313)与子区(413)杂交。第二夹板链(400)的子区可以以任何组合、以任何顺序包含以下任一者或多者:针对表面捕获引物的通用引物结合序列、针对表面钉扎引物的通用引物结合序列、样品索引序列、短随机序列和/或唯一分子索引(UMI)序列。

[0129] 在一些实施例中,第一夹板链的第一区(320)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处的第一通用结合序列杂交(例如,参见图2)。第一夹板链的第二区(330)可包含第二通用衔接子序列,其可与线性核酸文库分子另一端处的第二通用结合序列杂交(例如,参见图2)。在一些实施例中,第一夹板链的第一区(320)包含第一通用衔接子序列,其包括针对正向或反向测序引物的通用结合序列。在一些实施例中,第一夹板链的第二区(330)包含第二通用衔接子序列,其包括针对正向或反向测序引物的通用结合序列。在一些实施例中,第一夹板链(300)的5'端为磷酸化的或非磷酸化的。在一些实施例中,第一夹板链(300)的3'端包含末端3'OH基团或末端3'阻断基团。

[0130] 在一些实施例中,第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区(例如,参见图2和3)。第一子区(411)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第二子区(412)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第三子区(413)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、

17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。在一些实施例中,第二夹板链(400)被设计成表现出与文库分子(100)的插入序列(110)的杂交减少或无杂交。

[0131] 第二夹板链(400)的子区中的序列以3'至5'定向的示例性排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含样品索引序列和任选的短随机序列NNN]-[(413)包含用于结合表面捕获引物的通用序列]。

[0132] 第二夹板链(400)的子区中的序列以3'至5'定向的示例性排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含用于结合表面捕获引物的通用序列]-[(413)包含样品索引序列和任选的短随机序列NNN]。

[0133] 第二夹板链(400)的子区中的序列以3'至5'定向的示例性排列包括:[(411)包含样品索引序列和任选的短随机序列NNN]-[(412)包含用于结合表面钉扎引物的通用序列]-[(413)包含用于结合表面捕获引物的通用序列]。

[0134] 在一些实施例中,第二夹板链(400)包含携带第二样品索引序列和任选的短随机序列NNN的附加子区。例如,附加子区可以位于子区(411)与(412)之间,或者位于子区(412)与(413)之间。

[0135] 在一些实施例中,第二夹板链(400)的长度可以为20至100(例如,约20、25、30、35、40、45、50、60、70、80、90或100)个核苷酸。在一些实施例中,第二夹板链(400)的长度可以为30至80(例如,约30、35、40、45、50、60、70或80)个核苷酸,长度为80个核苷酸。长度为40至6060(例如,约40、45、50或60)个核苷酸。在一些实施例中,第二夹板链(400)的5'端为磷酸化的;可替代地,第二夹板链(400)的5'端为非磷酸化的。在一些实施例中,第二夹板链(400)的3'端包含末端3'OH基团;可替代地,第二夹板链(400)的3'端包含末端3'阻断基团。在一些实施例中,第二夹板链(400)在5'端和/或3'端处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,第二夹板链(400)在内部位置处包含一个或多个硫代磷酸酯键例如以赋予核酸内切酶抗性。在一些实施例中,第二夹板链(400)在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0136] 在一些实施例中,第一夹板链(300)包含内部区(310),该内部区包括至少三个子区,包括第四子区(311)、第五子区(312)和第六子区(313)。第四子区(311)可以与第二夹板链(400)的第一子区(411)杂交。第四子区(311)可以与第二夹板链(400)的第一子区(411)完全或部分互补。第五子区(312)可以与第二夹板链(400)的第二子区(412)杂交。第五子区(312)可以与第二夹板链(400)的第二子区(412)完全或部分互补。第六子区(313)可以与第二夹板链(400)的第三子区(413)杂交。第六子区(313)可以与第二夹板链(400)的第三子区(413)完全或部分互补。第四子区、第五子区和第六子区不与目的序列、表面捕获引物或表面钉扎引物杂交(或至少表现出非常少的杂交)。

[0137] 在一些实施例中,第一夹板链(300)的子区中的一者包含索引或随机序列,例如样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。例如,子区(311)、(312)或(313)包含样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索

引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。

[0138] 如图6所示,在一些实施例中,子区(312)包含以下中的任一者或者两者或更多者的任意组合:样品索引序列(用"S"表示);随机序列(用"N"表示);至少一个可以转化为脱碱基碱基的核苷酸(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基(例如,18碳间隔基)。图6描绘了第二夹板链(400;短夹板链)的替代示例性核酸序列,第二夹板链包含:三个子区,该三个子区包括:子区(411);子区(412);和子区(413)。

[0139] 在一些实施例中,包含索引或随机序列的第一夹板链子区包含以下中的任一者或者两者或更多者的任意组合:样品索引序列(用"S"表示);随机序列(用"N"表示);至少一个可以转化为脱碱基碱基的核苷酸(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。对于非限制性实例,参见图6。

[0140] 在一些实施例中,双链夹板衔接子(200)包含与第二夹板链(400)部分地杂交的第一夹板链(300),其中第一夹板链(300)包括通用长夹板链。在一些实施例中,通用长夹板链包括至少一个仅与第二夹板链(400)部分地杂交的子区。例如但不限于,通用长夹板链(300)包含携带至少一个核苷酸的子区,该核苷酸可以转化为脱碱基碱基(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。示例性的通用长夹板链(300)如图6所示。

[0141] 在一些实施例中,如图7中所描绘,第一夹板链缺乏子区(312)。图7还示出了第二夹板链(400;短夹板链)的示例性核酸序列,第二夹板链包含:三个子区,该三个子区包括:子区(411);子区(412);和子区(413)。在某些实施例中,当第一夹板链(300)与第二夹板链(400)杂交时,第二夹板链(400)的子区(412)环出,因为第一夹板链(300)缺乏子区(312)。

[0142] 在一些实施例中,第一夹板链(300)缺乏子区(311)、(312)或(313),使得通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体具有环出的第二夹板链的一部分。例如但不限于,第一夹板链(300)缺乏子区(312),并且在通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体中,第二夹板链(400)的子区(412)环出(例如,参见图7)。在一些实施例中,缺乏子区的第一夹板链(300)是通用长夹板链(300)的实例。

[0143] 在一些实施例中,第一夹板链(300)的长度可为50至150个核苷酸、或长度为60至100个核苷酸、或长度为70至90个核苷酸。在一些实施例中,第一夹板链(300)在5'端和/或3'端处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,第一夹板

链(300)在内部位置处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,第一夹板链(300)在5'端和/或3'端处包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,第一夹板链(300)在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0144] 下表1至6列出了文库分子(100)中的通用衔接子序列、第一夹板链(300)中的序列、第二夹板链(400)中的序列、经固定的表面引物的序列和测序引物的序列的各种实施例。

[0145] 在一些实施例中,表1中列出的文库分子(100)中的任何通用衔接子序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在一些实施例中,截短的长度可以为1、2、3、4、5、6、7、8、9、10、11或12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0146] 在一些实施例中,测序引物包含与下表1中列出的序列中的任一者互补的序列。

[0147] 在一些实施例中,测序引物包含与下表1中列出的序列中的任一者互补的序列,其中测序引物在5'端和/或3'端处被截短,并且其中截短可以为1至12个核苷酸。在一些实施例中,截短的长度可以为1、2、3、4、5、6、7、8、9、10、11或12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0148] 在一些实施例中,第一夹板链(300)和/或第二夹板链(400)中的任何序列包括与下表1至6中列出的任何序列互补的序列。

[0149] 在一些实施例中,表2至3中列出的第一夹板链(300)和/或第二夹板链(400)中的任何序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在一些实施例中,截短的长度可以为1、2、3、4、5、6、7、8、9、10、11或12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0150] 表1:文库分子(100)中的通用序列

[0151]

SEQ ID NO:	文库分子(100)中的序列:
1	正向测序引物结合位点(120): 5'- CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT -3'
2	正向测序引物结合位点(120): 5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCT -3'
3	正向测序引物结合位点(120): 5'- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG -3'
4	反向测序引物结合位点(130): 5'- ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT -3'
5	反向测序引物结合位点(130): 5'- AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -3'
6	反向测序引物结合位点(130): 5'- CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -3'
43	反向测序引物结合位点(130): 5'- ATGTCGGAAGGTGTGCAGGCTACCGCTTGT -3'
44	反向测序引物结合位点(130): 5'- ATGTCGGAAGGTGTGCAGGCTACCG -3'

[0152] 表2:第一夹板链(300):长夹板链中的序列

[0153]

SEQ ID NO:	第一夹板链(300)：长夹板链中的序列：
7	长夹板链(300)的第一区(320)： 5'- ATGTCGGAAGGTGTCTGGTGAGCCAATCCAGCACG -3'
8	长夹板链(300)的第一区(320)： 5'- GTGAGCCAATCCAGCACG -3'
9	长夹板链(300)的第一区(320)： 5' – AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -3'
10	长夹板链(300)的第一区(320)： 5'- CTGTCTCTTATACACATCTGACGCTGCCGACGA – 3'
11	长夹板链(300)的第二区(330)： 5'- ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT -3'
12	长夹板链(300)的第二区(330)： 5'- AGTTGACAAGCGGTAGCC -3'
13	长夹板链(300)的第二区(330)： 5'- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT -3'

[0154]

SEQ ID NO:	第一夹板链(300)：长夹板链中的序列：
14	长夹板链(300)的第二区(330)： 5'- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG -3'
15	在子区(311)中： (311)是与短夹板链(400)中的(411)中的序列完全或部分互补的序列。
16	在子区(312)中： (312)是与短夹板链(400)中的(412)中的序列完全或部分互补的序列。
17	在子区(313)中： (313)是与短夹板链(400)中的(413)中的序列完全或部分互补的序列。
N/A	在一些实施例中，第一夹板链(300)缺乏子区(312)。

[0155] 表3:第二夹板链(400):短夹板链中的序列

[0156]

SEQ ID NO:	第二夹板链(400)：短夹板链中的序列：
18	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'-CTAGTCCACTCCGACGCTGCTGA -5'
19	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- TAATGTACCTAGTCCACTCCGA -5'
20	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- TAATGTACCTAGTCCACTCCGACGCTGCTGA -5'
21	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- CTCCGACGCTGCTGA -5'
45	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- TAATGTACCTAGTCCACTCCGACGC -5'
46	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- CGTACTAATGTACCTAGTCCACTCCGA -5'
47	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- TGGGACTTTCATGCACGTAATGTAC -5'
22	短夹板链(400)的表面*捕获*引物结合位点 在子区：(411)、(412)或(413)中： 3'- GTTCGTCTTCTGCCGTATGCTCTA -5'
23	短夹板链(400)的表面*钉扎*引物结合位点 在子区：(411)、(412)或(413)中： 3'- TGGGACTTTCATGCACGTAATGTAC -5'
48	短夹板链(400)的表面钉扎引物结合位点 在子区：(411)、(412)或(413)中： 3'- TAATGTAC -5'
24	短夹板链(400)的表面*钉扎*引物结合位点

[0157]

SEQ ID NO:	第二夹板链(400)：短夹板链中的序列：
	在子区：(411)、(412)或(413)中： 3'- TGGGACTTTCATGCACG -5'
49	短夹板链(400)的表面钉扎引物结合位点 在子区：(411)、(412)或(413)中： 3'- TGGGACTTTCATGCACGATTGC -5'
25	短夹板链(400)的表面*钉扎*引物结合位点 在子区：(411)、(412)或(413)中： 3'- TGGGACTTTCATGCACGTAATGTACCTAGTCCA -5'
26	短夹板链(400)的表面*钉扎*引物结合位点 在子区：(411)、(412)或(413)中： 3'- CTAGAGCCACCAGCGGCATAGTAA -5'
27	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- GTAGGAGCCNNN -3'
28	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- CCGCTGCTANNN -3'
29	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- AACAACAAGNNN -3'
30	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- GGTGGTCTANNN -3'
31	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- TTGGCCAACNNN -3'
32	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- CAGGAGTGCNNN -3'
33	短夹板链(400)的示例性样品索引 在子区：(411)、(412)或(413)中： 5'- ATCACACTANNN -3'

[0158] 表4:经固定的表面引物

[0159]

SEQ:	经固定的表面引物：
34	表面*捕获*引物： 5'-GATCAGGTGAGGCTGCGACGACT -3'
35	表面*捕获*引物： 5'- ATTACATGGATCAGGTGAGGCT -3'
36	表面*捕获*引物： 5'- GAGGCTGCGACGACT -3'
50	表面捕获引物： 5'- ATTACATGGATCAGGTGAGGCTGCG -3'
37	表面*捕获*引物： 5'- CAAGCAGAAGACGGCATACGA-3'
38	表面*捕获*引物：

[0160]

SEQ：	经固定的表面引物：
	5'-CAAGCAGAAGACGGCATACGAGAT-3'
39	表面*钉扎*引物： 5'- CATGTAATGCACGTACTTTCAGGGT -3'
40	表面*钉扎*引物： 5'- ACCTGATCCATGTAATGCACGTACTTTCAGGGT -3'
41	表面*钉扎*引物： 5'- AATGATACGGCGACCACCGA-3'
42	表面*钉扎*引物： 5'- AATGATACGGCGACCACCGAGATC-3'

[0161] 表5：

[0162]

SEQ ID NO:	表 5：用于测序插入(110)的测序引物
51	FWD 测序引物： 5'- CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT -3'
52	REV 测序引物： 5'- AGTTGACAAGCGGTAGCCTGCACACCTTCCGACAT -3'
53	REV 测序引物： 5'- ACAAGCGGTAGCCTGCACACCTTCCGACAT -3'

[0163] 表6：

[0164]

SEQ ID NO:	表 6： 用于子区(411)、(412)或(413)中的测序索引的测序引物
54	FWD 测序引物： 5'- AGCCTCACCTGATCCATGTAAT -3'
55	FWD 测序引物： 5'- AGCCTCACCTGATCCATGTAATCATGC -3'
56	FWD 测序引物： 5'- AGTCGTCGCAGCCTCACCTGATC -3'
57	FWD 测序引物： 5'- AGTCGTCGCAGCCTCACCTGATCCATGTAAT -3'
58	FWD 测序引物： 5'- AGTCGTCGCAGCCTCACCTGATCCATGTAATCGTGA -3'
59	FWD 测序引物： 5'- AGTCGTCGCAGCCTCACCTGATCCATGTAATCATGC -3'
60	FWD 测序引物： 5' – ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT -3'
61	REV 测序引物： 5'- ACCCTGAAAGTACGTGCATTACATG -3'
62	REV 测序引物： 5'- ACCCTGAAAGTACGTGC -3'
63	REV 测序引物： 5'- ACCCTGAAAGTACGTGCTAACG -3'
64	REV 测序引物： 5'- ACCCTGAAAGTACGTGCATTACATGGATCAGGT -3'

[0165] 文库-夹板复合物

[0166] 在一些方面,本公开提供了一种文库-夹板复合物(500),其包含:(i)单链核酸文库分子(100),其包含在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列的插入区(110)(例如,目的序列),并且插入区(110)在另一侧上侧接针对反向测序引物结合

位点(130)的通用衔接子序列;以及(ii)双链夹板衔接子(200),其包含与第二夹板链(短夹板链(400))杂交的第一夹板链(长夹板链(300)),其中第一夹板链包含第一区(320)、内部区(310)和第二区(330)。在某些实施例中,第一夹板链的内部区(310)与第二夹板链(400)杂交以形成具有双链区和两个侧接单链区的双链夹板衔接子(200)。

[0167] 在一些实施例中,单链核酸文库分子(100)缺乏针对表面捕获引物结合位点的通用衔接子序列。在一些实施例中,单链核酸文库分子(100)缺乏针对表面钉扎引物结合位点的通用衔接子序列。在一些实施例中,单链核酸文库分子(100)缺乏针对样品索引序列的通用衔接子序列。在一些实施例中,单链核酸文库分子(100)缺乏唯一分子索引(UMI)序列。

[0168] 在文库-夹板复合物(500)中,第一夹板链的第一区(320)可以与文库分子的针对正向测序引物结合位点(120)的通用衔接子序列杂交,并且第一夹板链的第二区(330)可以与文库分子的针对反向测序引物结合位点(130)的通用衔接子序列杂交,从而使文库分子成环以生成文库-夹板复合物(500)(例如,参见图3和4)。

[0169] 在文库-夹板复合物(500)中,第二夹板链(400)可以使文库分子(100)接近至少针对表面钉扎引物结合位点的通用衔接子序列、针对表面捕获引物结合位点的通用衔接子序列和样品索引序列。在一些实施例中,第二夹板链(400)使文库分子(100)接近包括短随机序列(例如,NNN)和/或唯一分子索引(UMI)序列的其他序列。

[0170] 文库-夹板复合物(500)可以包含在文库分子的5'端与第二夹板链的3'端之间的第一缺口。在某些实施例中,文库-夹板复合物(500)还包含在第二夹板链的5'端与文库分子的3'端之间的第二缺口(例如,参见图3和4)。在一些实施例中,第一和第二缺口是可酶促连接的。连接反应会将来自第二夹板链(400)的序列接合至文库分子(100)的端部。

[0171] 在文库-夹板复合物(500)中,第一夹板链的第一区(320)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(500)中,第一夹板链的第二区(330)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。

[0172] 在文库-夹板复合物(500)中,第一夹板链的第一区(320)不与目的序列(110)杂交,并且第一夹板链的第二区(330)不与目的序列(110)杂交。

[0173] 在文库-夹板复合物(500)中,第二夹板链(400)不与目的序列(110)杂交,并且第一夹板链的内部区(310)不与目的序列(110)杂交。

[0174] 在一些实施例中,在文库-夹板复合物(500)中,第一夹板链(300)的内部区(310)与第二夹板链(400)杂交。第一夹板链(300)的内部区(310)可包括至少三个子区,包括子区(311)、(312)和(313)。第二夹板链(400)可包括至少三个子区,包括子区(411)、(412)和(413)。例如,子区(311)与子区(411)杂交。在另一实例中,子区(312)与子区(412)杂交。在另一实例中,子区(313)与子区(413)杂交。

[0175] 在一些实施例中,本文描述的任何文库-夹板复合物(500)包含多个文库-夹板复合物(500),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。

[0176] 在一些实施例中,在文库-夹板复合物(500)中,单链文库分子(100)的5'端为磷酸化的。在一些实施例中,在文库-夹板复合物(500)中,单链文库分子(100)的5'端缺乏磷酸酯基团。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。

[0177] 在一些实施例中,在文库-夹板复合物(500)中,第一夹板链(300)的长度可以为50至150(例如,约50、60、70、80、90、100、110、120、130、140或150)个核苷酸。在一些实施例中,在文库-夹板复合物(500)中,第一夹板链(300)的长度可以为60至100(例如,约60、70、80、90或100)个核苷酸。在一些实施例中,在文库-夹板复合物(500)中,第一夹板链(300)的长度可以为70至90(例如,约70、75、80、85或90)个核苷酸。在一些实施例中,第一夹板链(300)在5'端和/或3'端处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,第一夹板链(300)在内部位置处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,第一夹板链(300)在5'端和/或3'端处包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,第一夹板链(300)在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,第一夹板链(300)的5'端为磷酸化的。在一些实施例中,第一夹板链(300)的5'端缺乏磷酸酯基团。在一些实施例中,第一夹板链(300)的3'端包含末端3'OH基团或末端3'阻断基团。

[0178] 在一些实施例中,在文库-夹板复合物(500)中,第一夹板链(300)包含内部区(310),该内部区包括至少三个子区,包括第四子区(311)、第五子区(312)和第六子区(313)。第四子区(311)可以与第二夹板链(400)的第一子区(411)杂交。第四子区(311)可以与第二夹板链(400)的第一子区(411)完全或部分互补。第五子区(312)可以与第二夹板链(400)的第二子区(412)杂交。第五子区(312)可以与第二夹板链(400)的第二子区(412)完全或部分互补。第六子区(313)可以与第二夹板链(400)的第三子区(413)杂交。第六子区(313)可以与第二夹板链(400)的第三子区(413)完全或部分互补。第四子区、第五子区和第六子区不与目的序列、表面捕获引物或表面钉扎引物杂交(例如,至少表现出非常少的杂交)。

[0179] 在一些实施例中,第一夹板链(300)的子区中的一者包含索引或随机序列,例如样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。例如但不限于,子区(311)、(312)或(313)包含样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。

[0180] 在一些实施例中,包含索引或随机序列的第一夹板链子区包含以下中的任一者或者两者或更多者的任意组合:样品索引序列(用"S"表示);随机序列(用"N"表示);至少一个可以转化为脱碱基碱基的核苷酸(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。

[0181] 在一些实施例中,双链夹板衔接子(200)包含与第二夹板链(400)部分地杂交的第一夹板链(300),其中第一夹板链(300)包括通用长夹板链。在一些实施例中,通用长夹板链包括至少一个仅与第二夹板链(400)部分地杂交的子区。例如,通用长夹板链(300)可以包

含携带至少一个核苷酸的子区,该核苷酸可以转化为脱碱基碱基(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用“I”表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。示例性的通用长夹板链(300)如图6所示。

[0182] 在一些实施例中,在文库-夹板复合物(500)中,第一夹板链(300)缺乏子区(311)、(312)或(313),使得通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体具有环出的第二夹板链的一部分。例如,第一夹板链(300)缺乏子区(312),并且在通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体中,第二夹板链(400)的子区(412)环出(例如,参见图7)。在一些实施例中,缺乏子区的第一夹板链(300)是通用长夹板链(300)的实例。

[0183] 在一些实施例中,在文库-夹板复合物(500)中,第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区(例如,参见图3和4)。第一子区(411)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第二子区(412)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第三子区(413)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。第二夹板链(400)被设计成表现出与文库分子(100)的插入序列(110)的杂交减少或无杂交。

[0184] 在一些实施例中,在文库-夹板复合物(500)中,第二夹板链(400)的子区中的序列以3’至5’定向的排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含样品索引序列和任选的短随机序列NNN]-[(413)包含用于结合表面捕获引物的通用序列]。

[0185] 在一些实施例中,在文库-夹板复合物(500)中,第二夹板链(400)的子区中的序列以3’至5’定向的排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含用于结合表面捕获引物的通用序列]-[(413)包含样品索引序列和任选的短随机序列NNN]。

[0186] 在一些实施例中,在文库-夹板复合物(500)中,第二夹板链(400)的子区中的序列以3’至5’定向的排列包括:[(411)包含样品索引序列和任选的短随机序列NNN]-[(412)包含用于结合表面钉扎引物的通用序列]-[(413)包含用于结合表面捕获引物的通用序列]。

[0187] 在一些实施例中,在文库-夹板复合物(500)中,第二夹板链(400)包含携带第二样品索引序列和任选的短随机序列NNN的附加子区。例如,但不限于,附加子区可以位于子区(411)与(412)之间,或者位于子区(412)与(413)之间。

[0188] 在一些实施例中,在文库-夹板复合物(500)中,第二夹板链(400)的长度可以为20至100(例如,约20、25、30、35、40、45、50、60、70、80、90或100)个核苷酸。在一些实施例中,第

二夹板链(400)的长度可以为30至80(例如,约30、35、40、45、50、60、70或80)个核苷酸,长度为80个核苷酸。长度为40至6060(例如,约40、45、50或60)个核苷酸。在一些实施例中,第二夹板链(400)的5'端为磷酸化的;可替代地,第二夹板链(400)的5'端为非磷酸化的。在一些实施例中,第二夹板链(400)的3'端包含末端3'OH基团;可替代地,第二夹板链(400)的3'端包含末端3'阻断基团。在一些实施例中,第二夹板链(400)在5'端和/或3'端处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,第二夹板链(400)在内部位置处包含一个或多个硫代磷酸酯键例如以赋予核酸内切酶抗性。在一些实施例中,第二夹板链(400)在5'端和/或3'端处包含一个或多个2'-0-甲基胞嘧啶碱基。在一些实施例中,第二夹板链(400)在内部位置处包含一个或多个2'-0-甲基胞嘧啶碱基。

[0189] 在文库-夹板复合物(500)中,第二夹板链(400)使文库分子(100)接近至少针对表面钉扎引物结合位点的通用衔接子序列、针对表面捕获引物结合位点的通用衔接子序列和样品索引序列。在一些实施例中,第二夹板链(400)使文库分子(100)接近包括短随机序列(例如,NNN)和/或唯一分子索引(UMI)序列的其他序列。

[0190] 上表1至4列出了形成文库-夹板复合物(500)的各种分子的序列的各种实施例,各种分子包括文库分子(100)、第一夹板链(300)和第二夹板链(400)。在一些实施例中,第一夹板链(300)和/或第二夹板链(400)中的序列可以是与表2至3中列出的序列互补的序列。在一些实施例中,表2至3中列出的第一夹板链(300)和/或第二夹板链(400)中的任何序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。在一些实施例中,在文库分子中,表1中列出的正向测序引物结合位点(120)的序列和/或反向测序引物结合位点(130)的序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0191] 在一些方面,本公开提供了包含多个本文所述的任何文库-夹板复合物(500)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(500)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(500),并且缺乏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(500)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(500)以及T4多核苷酸激酶和连接酶。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(500)和连接酶,并且缺乏T4多核苷酸激酶。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。

[0192] 共价闭合环状分子

[0193] 在一些方面,本公开提供了一种共价闭合环状文库分子(600),其包含:目的序列(110)、正向测序引物结合位点(120)、反向测序引物结合位点(130),并且序列形成第二夹板链(400)。在一些实施例中,共价闭合环状文库分子(600)包含在一侧上侧接正向测序引物结合位点(120)且在另一侧上侧接反向测序引物结合位点(130)的目的序列(110),其中正向测序引物结合位点(120)和反向测序引物结合位点(130)共价接合至第二夹板链(400)。示例性共价闭合环状文库分子(600)示出于图4和5中。

[0194] 在一些实施例中,本文所述的任何共价闭合环状文库分子(600)进一步包含多个共价闭合环状文库分子(600),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(600)的目的序列(110)包含相同的目的序列。可替代地,在一些实施例中,目的序列(110)包含不同的目的序列。

[0195] 在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区(例如,参见图4和5)。第一子区(411)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第二子区(412)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第三子区(413)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基,其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基,其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标签)。在一些实施例中,唯一分子索引(UMI)包含随机序列。第二夹板链(400)被设计成表现出与文库分子(100)的插入序列(110)的杂交减少或无杂交。

[0196] 在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)的子区中的序列以3'至5'定向的排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含样品索引序列和任选的短随机序列NNN]-[(413)包含用于结合表面捕获引物的通用序列]。

[0197] 在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)的子区中的序列以3'至5'定向的排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含用于结合表面捕获引物的通用序列]-[(413)包含样品索引序列和任选的短随机序列NNN]。

[0198] 在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)的子区中的序列以3'至5'定向的排列包括:[(411)包含样品索引序列和任选的短随机序列NNN]-[(412)包含用于结合表面钉扎引物的通用序列]-[(413)包含用于结合表面捕获引物的通用序列]。

[0199] 在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)包含携带第二样品索引序列和任选的短随机序列NNN的附加子区。例如,附加子区可以位于子区(411)与(412)之间,或者位于子区(412)与(413)之间。

[0200] 在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)的长度可以为20至100个核苷酸、或长度为30至80个核苷酸、或长度为40至60个核苷酸。在一些实施例中,第二夹板链(400)在5'端和/或3'端处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,第二夹板链(400)在内部位置处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,第二夹板链(400)在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0201] 在一些实施例中,共价闭合环状文库分子(600)与第一夹板链(300)杂交或第一夹板链(300)不存在。在一些实施例中,第一夹板链(300)的长度可以为50至150(例如,约50、60、70、80、90、100、110、120、130、140或150)个核苷酸。在一些实施例中,第一夹板链(300)的长度可以为60至100(例如,约60、70、80、90或100)个核苷酸。在一些实施例中,第一夹板

链(300)的长度可以为70至90(例如,约70、75、80、85或90)个核苷酸。在一些实施例中,第一夹板链(300)在5'端和/或3'端处包含一个或多个硫代磷酸酯键例如以赋予核酸外切酶抗性。在一些实施例中,第一夹板链(300)在内部位置处包含一个或多个硫代磷酸酯键例如以赋予核酸内切酶抗性。在一些实施例中,第一夹板链(300)在5'端和/或3'端处包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,第一夹板链(300)在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,第一夹板链(300)的5'端为磷酸化的;可替代地,第一夹板链(300)的5'端缺乏磷酸酯基团。在一些实施例中,第一夹板链(300)的3'端包括末端3'OH基团;可替代地,第一夹板链(300)的3'端包括末端3'阻断基团。

[0202] 在一些实施例中,在共价闭合环状文库分子(600)中,第一夹板链(300)包含内部区(310),该内部区包括至少三个子区,包括第四子区(311)、第五子区(312)和第六子区(313)。第四子区(311)可以与第二夹板链(400)的第一子区(411)杂交。第四子区(311)可以与第二夹板链(400)的第一子区(411)完全或部分互补。第五子区(312)可以与第二夹板链(400)的第二子区(412)杂交。第五子区(312)可以与第二夹板链(400)的第二子区(412)完全或部分互补。第六子区(313)可以与第二夹板链(400)的第三子区(413)杂交。第六子区(313)可以与第二夹板链(400)的第三子区(413)完全或部分互补。第四子区、第五子区和第六子区不与目的序列、表面捕获引物或表面钉扎引物杂交(例如,表现出非常少的杂交)。

[0203] 在一些实施例中,在共价闭合环状文库分子(600)中,第一夹板链(300)的子区中的一者包含索引或随机序列,例如样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。例如,子区(311)、(312)或(313)包含样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基,其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基,其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标签)。在一些实施例中,唯一分子索引(UMI)包含随机序列。

[0204] 在一些实施例中,在共价闭合环状文库分子(600)中,子区(312)包含以下中的任一者或者两者或更多者的任意组合:样品索引序列(用"S"表示);随机序列(用"N"表示);至少一个可以转化为脱碱基碱基的核苷酸(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。

[0205] 在一些实施例中,当共价闭合环状文库分子(600)与第一夹板链(300)杂交时,第一夹板链与来自第二夹板链(400)的序列部分地杂交,其中第一夹板链(300)包括通用长夹板链。在一些实施例中,通用长夹板链可包含至少一个仅部分地与来自第二夹板链(400)的序列杂交的子区。例如,通用长夹板链(300)可以包含携带至少一个核苷酸的子区,该核苷酸可以转化为脱碱基碱基(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。示例性的通用长夹板链(300)如图6所示。

[0206] 在一些实施例中,在共价闭合环状文库分子(600)中,第一夹板链(300)缺乏子区(311)、(312)或(313),使得通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体具有环出的第二夹板链的一部分。例如,第一夹板链(300)可以缺乏子区(312),并且在通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体中,第二夹板链(400)的子区(412)环出(例如,参见图7)。在一些实施例中,缺乏子区的第一夹板链(300)是通用长夹板链(300)的实例。

[0207] 在一些实施例中,示例性共价闭合环状分子(600)包含:(i)目的序列(110);(ii)具有针对正向测序引物(140)的结合序列的通用衔接子序列;(iii)具有针对反向测序引物(150)的结合序列的通用衔接子序列;(iv)具有针对表面钉扎引物(子区411)的结合序列的通用衔接子序列;(v)具有针对表面捕获引物(子区413)的结合序列的通用衔接子序列;以及(vi)样品索引、短随机序列(NNN)和/或唯一分子索引(UMI)(子区412),其中共价闭合环状分子(600)任选地与第一夹板链(300)杂交。

[0208] 在一些实施例中,示例性共价闭合环状分子(600)包含:(i)目的序列(110);(ii)具有针对正向测序引物(140)的结合序列的通用衔接子序列;(iii)具有针对反向测序引物(150)的结合序列的通用衔接子序列;(iv)具有针对表面钉扎引物(子区411)的结合序列的通用衔接子序列;(v)具有针对表面捕获引物(子区412)的结合序列的通用衔接子序列;以及(vi)样品索引、短随机序列(NNN)和/或唯一分子索引(UMI)(子区413),其中共价闭合环状分子(600)任选地与第一夹板链(300)杂交。

[0209] 在一些实施例中,示例性共价闭合环状分子(600)包含:(i)目的序列(110);(ii)具有针对正向测序引物(140)的结合序列的通用衔接子序列;(iii)具有针对反向测序引物(150)的结合序列的通用衔接子序列;(iv)具有针对表面钉扎引物(子区412)的结合序列的通用衔接子序列;(v)具有针对表面捕获引物(子区413)的结合序列的通用衔接子序列;以及(vi)样品索引、短随机序列(NNN)和/或唯一分子索引(UMI)(子区411),其中共价闭合环状分子(600)任选地与第一夹板链(300)杂交。

[0210] 在一些实施例中,本文所述的任何共价闭合环状分子(600)包含多个共价闭合环状分子(600),其中多个共价闭合环状分子中的各个共价闭合环状分子(600)的目的序列(110)包含相同的目的序列。在一些实施例中,多个共价闭合环状分子中的各个共价闭合环状分子(600)的目的序列(110)包含不同的目的序列。

[0211] 上表1至4列出了形成共价闭合环状文库分子(600)的序列的各种实施例,包括文库分子(100)、第一夹板链(300)和第二夹板链(400)。在一些实施例中,第一夹板链(300)和/或第二夹板链(400)中的序列包括与表2至3中列出的序列互补的序列。在一些实施例中,表2至3中列出的第一夹板链(300)和/或第二夹板链(400)中的任何序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。在一些实施例中,在文库分子中,表1中列出的正向测序引物结合位点(120)的序列和/或反向测序引物结合位点(130)的序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0212] 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(600)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核酸外切酶中的任一者或者两者或更多者的任意组合。

[0213] 通过使用携带至少一个样品索引序列的双链夹板衔接子制备具有样品索引的共价闭合环状文库分子(600),可以实现多重工作流程。样品索引序列可用于使用从不同来源分离的输入核酸来制备单独批次的具有样品索引的共价闭合环状文库分子(600)。可以将具有样品索引的共价闭合环状文库分子(600)汇集在一起以生成多重共价闭合环状文库分子(600)混合物,并且可以对汇集的共价闭合环状文库分子(600)进行扩增和/或测序。插入区的序列连同样品索引序列可用于识别输入核酸的来源。在一些实施例中,可以将任何数量的共价闭合环状文库分子(600)汇集在一起,例如可以汇集2至10、或10至50、或50至100、或100至200、或超过200个批次的具有样品索引的共价闭合环状文库分子(600)。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,福尔马林固定的石蜡包埋的;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。

[0214] 包含双链夹板衔接子的试剂盒

[0215] 本公开内容提供了用于将一个或多个新衔接子序列引入线性核酸文库分子中的试剂盒。在一些实施例中,试剂盒可用于使单链核酸文库分子成环,该单链核酸文库分子具有在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列并且在另一侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列的目的序列(110)。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。多联体可以固定于大规模并行测序的支持物。

[0216] 本公开提供了包括核酸双链夹板衔接子(200)的试剂盒,该核酸双链夹板衔接子包含:(i)第一夹板链(长夹板链(300)),以及(ii)第二夹板链(短夹板链(400))。第一夹板链包含第一区(320)、内部区(310)和第二区(330)。第一夹板链的内部区(310)可以与第二夹板链(400)杂交以形成具有双链区和两个侧接单链区的双链夹板衔接子(200)。双链夹板衔接子(200)的两个侧接单链区被设计成与线性核酸文库分子(100)的末端序列杂交。线性核酸文库分子的末端序列分别包含第一(120)和第二(130)序列通用衔接子序列。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正向(120)和反向(130)测序引物结合位点的结合序列。

[0217] 第一夹板链的内部区(310)的至少一部分可以与第二夹板链(400)杂交以形成具有双链区和两个侧接单链区的双链夹板衔接子(200)。第二夹板链(400)包括至少一个新的衔接子序列。文库-夹板复合物(500)的形成使文库分子(100)接近至少一个新的通用衔接子序列,包括例如针对表面钉扎引物结合位点的通用序列、针对表面捕获引物结合位点的通用序列、样品索引序列、短随机序列(例如,NNN)和/或唯一分子索引(UMI)序列。示例性的双链夹板衔接子示出于图2和3中。该试剂盒可包括容器,该容器含有杂交或非杂交形式的第一夹板链(300)和第二夹板链(400)。该试剂盒可包括第一容器和第二容器,该第一容器含有第一夹板链(300)并且该第二容器含有第二夹板链(400)。

[0218] 在一些实施例中,在该试剂盒中,第一夹板链的第一区(320)包含第一通用衔接子序列,其可以与在线性核酸文库分子一端处的第一通用结合序列杂交(例如,参见图2和3)。第一夹板链的第二区(330)包含第二通用衔接子序列,其可以与在线性核酸文库分子另一

端处的第二通用结合序列杂交(例如,参见图2和3)。在一些实施例中,第一夹板链的第一区(320)包含第一通用衔接子序列,其包括针对正向或反向测序引物的通用结合序列。在一些实施例中,第一夹板链的第二区(330)包含第二通用衔接子序列,其包括针对正向或反向测序引物的通用结合序列。在一些实施例中,第一夹板链(300)的5'端为磷酸化的;可替代地,第一夹板链(300)的5'端为非磷酸化的。在一些实施例中,第一夹板链(300)的3'端包含末端3'OH基团;可替代地,第一夹板链(300)的5'端为末端3'阻断基团。

[0219] 在一些实施例中,在该试剂盒中,第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区(例如,参见图2和3)。第一子区(411)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第二子区(412)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(NNN)和/或唯一分子索引(UMI)。第三子区(413)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。第二夹板链(400)被设计成表现出与文库分子(100)的插入序列(110)的杂交减少或无杂交。

[0220] 第二夹板链(400)的子区中的序列以3'至5'定向的示例性排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含样品索引序列和任选的短随机序列NNN]-[(413)包含用于结合表面捕获引物的通用序列]。

[0221] 第二夹板链(400)的子区中的序列以3'至5'定向的示例性排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含用于结合表面捕获引物的通用序列]-[(413)包含样品索引序列和任选的短随机序列NNN]。

[0222] 第二夹板链(400)的子区中的序列以3'至5'定向的示例性排列包括:[(411)包含样品索引序列和任选的短随机序列NNN]-[(412)包含用于结合表面钉扎引物的通用序列]-[(413)包含用于结合表面捕获引物的通用序列]。

[0223] 在一些实施例中,在该试剂盒中,第二夹板链(400)的长度可以为20至100(例如,约20、25、30、35、40、45、50、60、70、80、90或100)个核苷酸。在一些实施例中,第二夹板链(400)的长度可以为30至80(例如,约30、35、40、45、50、60、70或80)个核苷酸,长度为80个核苷酸。长度为40至6060(例如,约40、45、50或60)个核苷酸。在一些实施例中,第二夹板链(400)的5'端为磷酸化的;可替代地,第二夹板链(400)的5'端为非磷酸化的。在一些实施例中,第二夹板链(400)的3'端包含末端3'OH基团;可替代地,第二夹板链(400)的3'端包含末端3'阻断基团。在一些实施例中,第二夹板链(400)在5'端和/或3'端处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,第二夹板链(400)在内部位置处包含一个或多个硫代磷酸酯键例如以赋予核酸内切酶抗性。在一些实施例中,第二夹板链(400)在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0224] 在一些实施例中,在该试剂盒中,第一夹板链(300)包含内部区(310),该内部区包

括至少三个子区,包括第四子区(311)、第五子区(312)和第六子区(313)。第四子区(311)可以与第二夹板链(400)的第一子区(411)杂交。第四子区(311)可以与第二夹板链(400)的第一子区(411)完全或部分互补。第五子区(312)可以与第二夹板链(400)的第二子区(412)杂交。第五子区(312)可以与第二夹板链(400)的第二子区(412)完全或部分互补。第六子区(313)可以与第二夹板链(400)的第三子区(413)杂交。第六子区(313)可以与第二夹板链(400)的第三子区(413)完全或部分互补。第四子区、第五子区和第六子区不与目的序列、表面捕获引物或表面钉扎引物杂交(或至少表现出非常少的杂交)。

[0225] 在一些实施例中,在该试剂盒中,第一夹板链(300)的子区中的一者包含索引或随机序列,例如样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。例如,子区(311)、(312)或(313)包含样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。

[0226] 在一些实施例中,在该试剂盒中,包含索引或随机序列的第一夹板链子区包含以下中的任一者或者两者或更多者的任意组合:样品索引序列(用“S”表示);随机序列(用“N”表示);至少一个可以转化为脱碱基碱基的核苷酸(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用“I”表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。

[0227] 在一些实施例中,在该试剂盒中,双链夹板衔接子(200)包含与第二夹板链(400)部分地杂交的第一夹板链(300),其中第一夹板链(300)包括通用长夹板链。在一些实施例中,通用长夹板链包括至少一个仅与第二夹板链(400)部分地杂交的子区。例如但不限于,通用长夹板链(300)包含携带至少一个核苷酸的子区,该核苷酸可以转化为脱碱基碱基(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用“I”表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。示例性的通用长夹板链(300)如图6所示。

[0228] 在一些实施例中,在该试剂盒中,第一夹板链(300)缺乏子区(311)、(312)或(313),使得通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体具有环出的第二夹板链的一部分。例如但不限于,第一夹板链(300)缺乏子区(312),并且在通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体中,第二夹板链(400)的子区(412)环出(例如,参见图7)。在一些实施例中,缺乏子区的第一夹板链(300)是通用长夹板链(300)的实例。

[0229] 在一些实施例中,在该试剂盒中,第一夹板链(300)的长度可以为50至150个核苷酸、或长度为60至100个核苷酸、或长度为70至90个核苷酸。在一些实施例中,第一夹板链

(300)在5'端和/或3'端处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,第一夹板链(300)在内部位置处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,第一夹板链(300)在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0230] 上表1至4列出了可以包括在该试剂盒中的文库分子(100)、第一夹板链(300)、第二夹板链(400)和经固定的表面引物中的通用衔接子序列的各种实施例。在一些实施例中,在本文所述的任何试剂盒中,第一夹板链(300)和/或第二夹板链(400)中的序列包括与表2至3中列出的序列互补的序列。在一些实施例中,在本文所述的任何试剂盒中,表2至3中列出的第一夹板链(300)和/或第二夹板链(400)中的任何序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0231] 在一些实施例中,试剂盒包含核酸双链夹板衔接子(200)并且进一步包含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,任选地其中连接酶包括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核酸外切酶中的任一者或者两者或更多者的任意组合。

[0232] 在一些实施例中,试剂盒包含至少一种用于使多个双链夹板衔接子(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重酶促反应的缓冲液,包括以下的任何组合:(i)使第一和/或第二夹板链(例如,(300)和/或(400))的5'端磷酸化,(ii)连接文库-夹板复合物(500)中的缺口,和/或(iii)来自共价闭合环状分子(600)的第一夹板链(300)的核酸外切酶消化。可替代地,试剂盒包括两种或更多种单独的缓冲液,例如,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核酸外切酶消化反应。

[0233] 在一些实施例中,试剂盒包含一个或多个容器,其含有本文所述的任何双链夹板衔接子(200)或本文所述的第一夹板链(300)和第二夹板链(400)中的任一者。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。该试剂盒可以任何组合包含这些组分中的任何者,并且可以包含在单个容器中,或者可以包含在单独的容器中,或者它们的任何组合。

[0234] 试剂盒可包括使用试剂盒例如进行反应以将一种或多种新的衔接子序列引入线性核酸文库分子中的说明书。

[0235] 形成多个文库-夹板复合物的方法

[0236] 在一些方面中,本公开提供了用于形成多个文库-夹板复合物(500)的方法,该方法包括:(a)提供多个双链夹板衔接子(200),其中多个双链夹板衔接子中的各个双链夹板衔接子(200)包含与第二夹板链(400)杂交的第一夹板链(300),其中双链夹板衔接子包括双链区和两个侧接单链区,其中第一夹板链包含第一区(320)、内部区(310)和第二区(330),并且其中第一夹板链的内部区(310)与第二夹板链(400)杂交。示例性的双链夹板衔接子(200)示出于图2和3中。在一些实施例中,第一夹板链(300)的内部区(310)的一部分不与第二夹板链(400)的一部分杂交。

[0237] 在一些实施例中,用于形成多个文库-夹板复合物(500)的方法进一步包括步骤

(b):使多个双链夹板衔接子与多个单链核酸文库分子(100)杂交,其中各个文库分子包含目的序列(110),该目的序列在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列并且在另一侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列(例如,图3)。杂交可以在适合于使第一夹板链的第一区(320)与文库分子的部分(120)杂交的条件下进行,并且该条件适合于使第一夹板链的第二区(330)与文库分子的部分(130)杂交,从而使多个文库分子成环以形成多个文库-夹板复合物(500)。

[0238] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链的第一区(320)包含可与在线性核酸文库分子一端处的针对正向测序引物结合位点(120)的通用衔接子序列杂交的序列。

[0239] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链的第二区(330)包含可与在线性核酸文库分子另一端处的针对反向测序引物结合位点(130)的通用衔接子序列杂交的序列。

[0240] 在一些实施例中,第二夹板链(400)的5'端为磷酸化的或缺乏磷酸酯基团。在一些实施例中,第二夹板链(400)的3'端包括末端3'OH基团或末端3'阻断基团。

[0241] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链的第一区(320)与文库分子的针对正向测序引物结合位点(120)的通用衔接子序列杂交,并且第一夹板链的第二区(330)与文库分子的针对反向测序引物结合位点(130)的通用衔接子序列杂交,从而使文库分子成环以生成文库-夹板复合物(500)。文库-夹板复合物(500)可以包含在文库分子的5'端与第二夹板链的3'端之间的第一缺口(例如,图3和4)。文库-夹板复合物(500)还可以包含在第二夹板链的5'端与文库分子的3'端之间的第二缺口(例如,图3和4)。在一些实施例中,第一和第二缺口是可酶促连接的。连接反应会将来自第二夹板链(400)的序列接合至文库分子(100)的端部。

[0242] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链的第一区(320)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(500)中,第一夹板链的第二区(330)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。

[0243] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)不与目的序列(110)杂交,并且第一夹板链的内部区(310)不与目的序列(110)杂交。

[0244] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链的第一区(320)不与目的序列(110)杂交,并且第一夹板链的第二区(330)不与目的序列(110)杂交。

[0245] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,单链文库分子(100)的5'端为磷酸化的;可替代地,单链文库分子(100)的5'端缺乏磷酸酯基团。在一些实施例中,单链文库分子的3'端包括末端3'OH基团;可替代地,单链文库分子的3'端包括末端3'阻断基团。

[0246] 在一些实施例中,单链核酸文库分子(100)缺乏针对表面捕获引物结合位点的通用衔接子序列。在一些实施例中,单链核酸文库分子(100)缺乏针对表面钉扎引物结合位点的通用衔接子序列。在一些实施例中,单链核酸文库分子(100)缺乏针对样品索引序列的通用衔接子序列。在一些实施例中,单链核酸文库分子(100)缺乏唯一分子索引(UMI)序列。

[0247] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区(例如,参见图2和3)。第一子区(411)包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列和/或唯一分子索引(UMI)。第二子区(412)可包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列和/或唯一分子索引(UMI)。第三子区(413)可以包含针对经固定的表面钉扎引物的通用结合序列、经固定的表面捕获引物、至少一个样品索引序列、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。第二夹板链(400)被设计成表现出与文库分子(100)的插入序列(110)的杂交减少或无杂交。

[0248] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)的子区中的序列以3'至5'定向的排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含样品索引序列和任选的短随机序列NNN]-[(413)包含用于结合表面捕获引物的通用序列]。

[0249] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)的子区中的序列以3'至5'定向的排列包括:[(411)包含用于结合表面钉扎引物的通用序列]-[(412)包含用于结合表面捕获引物的通用序列]-[(413)包含样品索引序列和任选的短随机序列NNN]。

[0250] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)的子区中的序列以3'至5'定向的排列包括:[(411)包含样品索引序列和任选的短随机序列NNN]-[(412)包含用于结合表面钉扎引物的通用序列]-[(413)包含用于结合表面捕获引物的通用序列]。

[0251] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)包含携带第二样品索引序列和任选的短随机序列NNN的附加子区。例如,附加子区可以位于子区(411)与(412)之间,或者位于子区(412)与(413)之间。

[0252] 在一些实施例中,在形成多个文库-夹板复合物(500)的方法中,第二夹板链(400)的长度可以为20至100(例如,约20、25、30、35、40、45、50、60、70、80、90或100)个核苷酸。在一些实施例中,第二夹板链(400)的长度可以为30至80(例如,约30、35、40、45、50、60、70或80)个核苷酸,长度为80个核苷酸。长度为40至6060(例如,约40、45、50或60)个核苷酸。在一些实施例中,第二夹板链(400)的5'端为磷酸化的;可替代地,第二夹板链(400)的5'端为非磷酸化的。在一些实施例中,第二夹板链(400)的3'端包含末端3'OH基团;可替代地,第二夹板链(400)的3'端包含末端3'阻断基团。在一些实施例中,第二夹板链(400)在5'端和/或3'端处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,第二夹板链(400)在内部位置处包含一个或多个硫代磷酸酯键例如以赋予核酸内切酶抗性。在一些实施例中,第二夹板链(400)在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0253] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链(300)的长度可以为50至150个核苷酸、或长度为60至100个核苷酸、或长度为70至90个核苷酸。在一些实施例中,第一夹板链(300)在5'端和/或3'端处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,第一夹板链(300)在内部位置处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,第一夹板链(300)在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。在一些实施例中,第一夹板链(300)的5'端为磷酸化的或缺乏磷酸酯基团。在一些实施例中,第一夹板链(300)的3'端包含末端3'OH基团或末端3'阻断基团。

[0254] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链(300)包含内部区(310),该内部区包括至少三个子区,包括第四子区(311)、第五子区(312)和第六子区(313)。第四子区(311)可以与第二夹板链(400)的第一子区(411)杂交。第四子区(311)可以与第二夹板链(400)的第一子区(411)完全或部分互补。第五子区(312)可以与第二夹板链(400)的第二子区(412)杂交。第五子区(312)可以与第二夹板链(400)的第二子区(412)完全或部分互补。第六子区(313)可以与第二夹板链(400)的第三子区(413)杂交。第六子区(313)可以与第二夹板链(400)的第三子区(413)完全或部分互补。第四子区、第五子区和第六子区不与目的序列、表面捕获引物或表面钉扎引物杂交(或至少表现出非常少的杂交)。

[0255] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,第一夹板链(300)包含子区(312),该子区包含样品索引、短随机序列(例如,NNN)和/或唯一分子索引(UMI)。在一些实施例中,样品索引包含5至20个碱基(例如,约5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。在一些实施例中,唯一分子索引(UMI)包含3至20个碱基(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个碱基),其可用于唯一地识别衔接子所附加的核酸分子(例如,分子标记)。在一些实施例中,唯一分子索引(UMI)包含随机序列。

[0256] 在一些实施例中,在用于形成多个文库-夹板复合物(500)的方法中,子区(312)包括以下中的任一者或者两者或更多者的任意组合:样品索引序列(用"S"表示);随机序列(用"N"表示);至少一个可以转化为脱碱基碱基的核苷酸(用实心黑色矩形表示)(例如,尿苷、8-氧代-7,8-二氢鸟嘌呤(例如,8oxoG)或脱氧肌苷);脱氧肌苷(用"I"表示);和/或间隔基。在一些实施例中,间隔基包括18碳间隔基(例如,包括六乙二醇间隔基)、多个C3间隔基亚磷酰胺、或包括三亚甲基二醇间隔基的间隔基9。在一些实施例中,间隔基包括聚乙二醇间隔基,例如,PEG2、PEG3或PEG4间隔基。

[0257] 在一些实施例中,在形成多个文库-夹板复合物(500)的方法中,第一夹板链(300)缺乏子区(311)、(312)或(313),使得通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体具有环出的第二夹板链的一部分。例如,第一夹板链(300)缺乏子区(312),并且在通过第一夹板链(300)与第二夹板链(400)之间的杂交形成的双链体中,第二夹板链(400)的子区(412)环出(例如,参见图7)。

[0258] 上表1至4列出了形成文库-夹板复合物(500)的各种分子的序列的各种实施例,各种分子包括文库分子(100)、第一夹板链(300)和第二夹板链(400)。在一些实施例中,在用于形成本文所述的多个文库-夹板复合物(500)的任何方法中,第一夹板链(300)和/或第二

夹板链(400)中的序列包括与表2至3中列出的序列互补的序列。在一些实施例中,在用于形成本文所述的多个文库-夹板复合物(500)的任何方法中,表2至3中列出的第一夹板链(300)和/或第二夹板链(400)中的任何序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在一些实施例中,截短的长度可以为1、2、3、4、5、6、7、8、9、10、11或12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。在一些实施例中,在文库分子中,表1中列出的正向测序引物结合位点(120)的序列和/或反向测序引物结合位点(130)的序列可以在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在一些实施例中,截短的长度可以为1、2、3、4、5、6、7、8、9、10、11或12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0259] 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(500)的任何方法可以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。

[0260] 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(500)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,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(500)的方法进一步包括步骤(c1):在适合于使多个双链夹板衔接子(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双链夹板衔接子(200)和多个单链核酸文库分子(100)与T4多核苷酸激酶接触;以及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(500)的方法进一步包括步骤(d1):在适合于酶促连接第一缺口和第二缺口的条件下,在第二反应容器中使多个磷酸化的双链夹板衔接子(200)和多个磷酸化的单链核酸文库分子(100)与连接酶接触,从而生成多个共价闭合环状文库分子(600),每个共价闭合环状文库分子与第一夹板链(300)杂交。在一些实施例中,连接酶包含T7DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。用于生成共价闭合环状文库分子(600)的示例性方法示出于图4中。酶促连接反应可以将来自第二夹板链(400)的序列接合至文库分子(100)的端部。双链衔接子(200)可以将新的通用衔接子序列和样品索引序列接合至文库分子(100)的两端,而不需要进行引物延伸或PCR反应。因此,采用双链衔接子(200)的无PCR工作流程可用于生成具有下游工作流程(诸如扩增和测序)所需的衔接子序列的共价环化文库分子(600)。

[0261] 如图4所示,在一些实施例中,在共价闭合环状文库分子(600)中,第二夹板链(400)的子区共价接合至通用正向测序引物结合位点(120)和通用反向测序引物结合位点(130)。在一些实施例中,第一夹板链(300)具有可延伸的3'端,其可用于引发引物延伸反应。在一些实施例中,第一夹板链(300)可用作扩增引物以进行滚环扩增反应以生成与共价闭合环状文库分子(600)互补的核酸多联体分子。

[0262] 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(500)的任何方法进一步包括进行连续的磷酸化和连接反应,磷酸化和连接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(500)的方法进一步包括步骤(c2):在适合于使多个双链夹板衔接子(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双链夹板衔接子(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(500)的方法进一步包括步骤(d2):在适合于酶促连接第一缺口和第二缺口的条件下,在同一第一反应容器中使磷酸化的双链夹板衔接子(200)和磷酸化的单链核酸文库分子(100)与连接酶接触,从而生成多个共价闭合环状文库分子(600),每个共价闭合环状文库分子与第一夹板链(300)杂交。在一些实施例中,连接酶包括T7 DNA连接酶、T3连接酶、T4连接酶或Taq连接酶。用于生成共价闭合环状文库分子(600)的示例性方法示出于图4中。

[0263] 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(500)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,磷酸化和连接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(500)的方法进一步包括步骤(c3):在适合于使多个双链夹板衔接子(200)和多个单链核酸文库分子(100)的5'端磷酸化的条件下,并且该条件适合于酶促连接第一缺口和第二缺口,在第一反应容器中使多个双链夹板衔接子(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,从而生成多个共价闭合环状文库分子(600),每个共价闭合环状文库分子与第一夹板链(300)杂交。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。用于生成共价闭合环状文库分子(600)的示例性方法示出于图4中。

[0264] 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(500)的任何方法进一步包括以下任选步骤:从多个共价闭合环状文库分子(600)酶促移除多个第一夹板链(300),其包括:使多个共价闭合环状文库分子(600)与至少一种核酸外切酶接触,以移除多个第一夹板链(300)并保留多个共价闭合环状文库分子(600)。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(c1)并在第二反应容器中进行连接反应(d1)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(c2)并在第一反应容器中进行顺序的连接反应(d2)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(c3)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。

[0265] 通过使用携带至少一个样品索引序列的双链夹板衔接子制备具有样品索引的共价闭合环状文库分子(600),可以实现多重工作流程。样品索引序列可用于使用从不同来源分离的输入核酸来制备单独批次的具有样品索引的共价闭合环状文库分子(600)。可以将具有样品索引的共价闭合环状文库分子(600)汇集在一起以生成多重共价闭合环状文库分子(600)混合物,并且可以对汇集的共价闭合环状文库分子(600)进行扩增和/或测序。插入区的序列连同样品索引序列可用于识别输入核酸的来源。在一些实施例中,可以将任何数量的共价闭合环状文库分子(600)汇集在一起,例如可以汇集2至10、或10至50、或50至100、或100至200、或超过200个批次的具有样品索引的共价闭合环状文库分子(600)。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,福尔马林固定的石蜡包埋的;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。

[0266] 滚环扩增方法

[0267] 本公开提供了对共价闭合环状文库分子(600)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,滚环扩增反应可在核酸外切酶反应后进行,此时共价闭合环状文库分子(600)不再与第一夹板链(300)杂交。在一些实施例中,滚环扩增反应可以在与第一夹板链(300)杂交的共价闭合环状文库分子(600)上进行。在一些实施例中,共价闭合环状文库分子(600)可以分布在支持物上,并且然后进行滚环扩增反应。在一些实施例中,共价闭合环状文库分子(600)可以在溶液中进行滚环扩增反应,并且然后分布在支持物上。在一些实施例中,滚环扩增反应可以采用保留的第一夹板链(300)作为扩增引物(例如,图4),或者可以去除第一夹板链(300)(例如,经由核酸外切酶消化)并用可溶性扩增引物替换(例如,图5)。

[0268] 支持物上滚环扩增

[0269] 在一些实施例中,用于对缺乏杂交的第一夹板链(300)的多个共价闭合环状文库分子(600)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(600)包含针对表面捕获引物的通用结合序列,该方法包括步骤(a):在适合于使各个共价闭合环状文库分子(600)与各个经固定的表面捕获引物杂交的条件下,将多个共价闭合环状文库分子(600)分布到支持物上,该支持物上固定有多个表面捕获引物,从而将多个共价闭合环状文库分子(600)固定到支持物。

[0270] 在一些实施例中,多个表面捕获引物包含表4中列出的序列中的任一者或其互补序列。

[0271] 各个表面捕获引物可以与具有针对表面捕获引物的通用结合序列的共价闭合环状文库分子(600)杂交。

[0272] 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):在适合于使用多个表面捕获引物作为经固定的扩增引物并使用多个共价闭合环状文库分子(600)作为模板分子,在支持物上进行滚环扩增反应的条件下,使多个经固定的共价闭合环状文库分子(600)与多个链置换聚合酶和多个核苷酸接触,从而生成多个固定到表面捕获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中的两者或更多者的任意组合。在一些实施例中,各个经固定的多联体共价接合至各个表面捕获引物。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(600)包含针对表面捕获引物和表面钉扎引物的通用结合序列,使得滚环扩增反应生成具有由共价闭合环状文库分子(600)携带的序列的多个串联拷贝的多联体分子,该多联体分子包含针对表面捕获引物和表面钉扎引物的通用结合序列。在一些实施例中,支持物进一步包含多个表面钉扎引物。在一些实施例中,经固定的表面钉扎引物用于将多联体分子的至少一部分钉扎到支持物上。在一些实施例中,经固定的表面钉扎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,多个表面钉扎引物包含表4中列出的序列中的任一者或其互补序列。在一些实施例中,可以使固定的多联体经历测序反应。

[0273] 各个表面钉扎引物可以与具有针对表面钉扎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,经固定的表面钉扎引物用于将多联体分子的至少一部分钉扎到支持物上。在一些实施例中,经固定的表面钉扎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,可以使固定的多联体经历测序反应。

[0274] 在一些实施例中,在用于进行滚环扩增反应的方法中,可以将多个共价闭合环状文库分子(600)分布到支持物上,该支持物涂覆有一种或多种化合物,以在支持物上产生钝化层(例如,图21)。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,以固定到支持物。在一些实施例中,支持物包含低非特异性结合表面,其使得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到支持物。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作为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,固定到支持物或固定到支持物上的涂层的共价闭合环状文库分子(600)的密度为每mm^2约10^2至10^6(例如,约10^2、10^3、10^4、10^5或10^6)。在一些实施例中,固定到支持物或固定到支持物上的涂层的共价闭合环状文库分子(600)的密度为每mm^2约10^6至10^9(例如,约10^6、10^7、10^8或10^9)。在一些实施例中,固定到支持物或固定到支持物上的涂层的共价闭合环状文库分子(600)的密度为每mm^2约10^9至10^{12}(例如,约10^9、10^{10}、10^{11}或10^{12})。在一些实施例中,多个共价闭合环状文库分子(600)在支持物(或支持物上的涂层)上的预定位点处固定到支持物或固定到支持物上的涂层。在一些实施例中,多个共价闭合环状文库分子(600)在支持物(或支持物上的涂层)上的随机位点处固定到支持物或固定到支持物上的涂层。

[0275] 使用可溶性扩增引物进行溶液内滚环扩增

[0276] 在一些实施例中,用于对缺乏杂交的第一夹板链(300)的多个共价闭合环状文库分子(600)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(600)包含针对正向扩增引物的通用结合序列和针对表面捕获引物的通用结合序列,该方法包括:(a)在溶液中使多个共价闭合环状文库分子与多个可溶性正向扩增引物杂交;以及(b)在适合于使用多个正向扩增引物和多个共价闭合环状文库分子(600)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(600)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,从而生成多个核酸多联体分子,该多个核酸多联体分子具有仍与其共价闭合环状文库分子(600)杂交的部分。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):在适合于使多联体分子的至少一部分与多个经固定的表面捕获引物杂交的条件下,将多个多联体分子分布到其上固定有多个表面捕获引物的支持物上,从而固定多个多联体分子。多个经固定的多联体分子仍与其共价闭合环状文库分子(600)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(600)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中的两者或更多者的任意组合的多个核苷酸进行。在一些实施例中,各个经固定的多联体与各个表面捕获引物杂交。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(600)包含针对表面捕获引物和表面钉扎引物的通用结合序列,使得溶液中滚环扩增反应生成具有由共价闭合环状文库分子(600)携带的序列的多个串联拷贝的多联体分子,该多联体分子包含针对表面捕获引物和表面钉扎引物的通用结合序列。在一些实施例中,支持物进一步包含多个表面钉扎引物。在一些实施例中,经固定的表面钉扎引物用于将多联体分子的至少一部分钉扎到支持物上。在一些实施例中,经固定的表面钉扎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,可以使固定的多联体经历测序反应。在一些实施例中,可溶性正向扩增引物具有与表面捕获引物相同的序列。

[0277] 在一些实施例中,多个表面捕获引物包含表4中列出的序列中的任一者或其互补序列。在一些实施例中,多个表面钉扎引物包含表4中列出的序列中的任一者或其互补序列。

[0278] 在一些实施例中,在用于进行滚环扩增反应的方法中,固定在支持物上的多个表面捕获引物包括序列5'-GATCAGGTGAGGCTGCGACGACT-3'(SEQ ID NO:34)(或其互补序列)。

[0279] 在一些实施例中,在用于进行滚环扩增反应的方法中,固定在支持物上的多个表面捕获引物包括序列5'-ATTACATGGATCAGGTGAGGCT-3'(SEQ ID NO:35)(或其互补序列)。

[0280] 在一些实施例中,经固定的表面钉扎引物用于将多联体分子的至少一部分钉扎到支持物上。在一些实施例中,经固定的表面钉扎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,可以使固定的多联体经历测序反应。

[0281] 在一些实施例中,可将步骤(c)的多个多联体分子可以分布到支持物上,该支持物上涂覆有一种或更多种化合物,以在支持物例如玻璃上产生钝化层(参见,例如,图21)。在替代性实施例中,支持物可由任何材料(诸如玻璃、塑料或聚合物材料)制成。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,以固定到支持物。在一些实施例中,支持物包含低非特异性结合表面,其使得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到支持物。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个支链至少4个支链(例如,4、5、6、7、8、9、10或更多个支链)的支化PEG。在一些实施例中,低非特异性结合涂层具有可以作为水接触角测量的亲水性程度,其中水接触角不超过45度(例如,不超过5度、不超过10度、不超过15度、不超过20度、不超过25度、不超过30度、不超过35度、不超过40度或不超过45度)。在一些实施例中,固定到支持物或固定到支持物上的涂层的多联体分子的密度为每mm^2约10^2至10^6

(例如,约10^2、10^3、10^4、10^5或10^6)。在一些实施例中,固定到支持物或固定到支持物上的涂层的多联体分子的密度为每mm^2约10^6至10^9(例如,约10^6、10^7、10^8或10^9)。在一些实施例中,固定到支持物或固定到支持物上的涂层的多联体分子的密度为每mm^2约10^9至10^{12}(例如,约10^9、10^{10}、10^{11}或10^{12})。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)上的预定位点处固定到支持物或固定到支持物上的涂层。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)上的随机位点处固定到支持物上的涂层。

[0282] 使用第一夹板链进行溶液内滚扩增

[0283] 在一些实施例中,用于对与第一夹板链(300)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(600)包含针对表面捕获引物的通用结合序列,该方法包括(a):在适合于使用第一夹板链(300)作为扩增引物进行第一滚环扩增反应的条件下,在溶液中使与第一夹板链(300)杂交的多个共价闭合环状文库分子(600)与多个链置换聚合酶和多个核苷酸接触,从而生成仍与其共价闭合环状文库分子(600)杂交的多个多联体分子(例如,参见图4)。

[0284] 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):在适合于使多联体的至少一部分与多个经固定的表面捕获引物杂交的条件下,将与其共价闭合环状文库分子(600)杂交的多个多联体分子分布到其上固定有多个表面捕获引物的支持物上,从而固定多个多联体分子。多个经固定的多联体分子仍与其共价闭合环状文库分子(600)杂交。

[0285] 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(600)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,使多个经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,从而延伸多个经固定的核酸多联体分子。

[0286] 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中的两者或更多者的任意组合的多个核苷酸进行。在一些实施例中,各个经固定的多联体与各个表面捕获引物杂交。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(600)包含针对表面捕获引物和表面钉扎引物的通用结合序列,使得溶液中滚环扩增反应生成具有由共价闭合环状文库分子(600)携带的序列的多个串联拷贝的多联体分子,该多联体分子包含针对表面捕获引物和表面钉扎引物的通用结合序列。在一些实施例中,支持物进一步包含多个表面钉扎引物。在一些实施例中,经固定的表面钉扎引物用于将多联体分子的至少一部分钉扎到支持物上。在一些实施例中,经固定的表面钉扎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,可以使固定的多联体经历测序反应。

[0287] 在一些实施例中,多个表面捕获引物包含表4中列出的序列中的任一者或其互补序列。在一些实施例中,多个表面钉扎引物包含表4中列出的序列中的任一者或其互补序列。

[0288] 在一些实施例中,经固定的表面钉扎引物用于将多联体分子的至少一部分钉扎到支持物上。在一些实施例中,经固定的表面钉扎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,可以使固定的多联体经历测序反应。

[0289] 在一些实施例中,可以将步骤(b)的多个多联体分子分布到支持物上,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21)。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,以固定到支持物。在一些实施例中,支持物包含低非特异性结合表面,其使得能够改善支持物上的核酸杂交和扩增性能。在一些实施例中,支持物可包含一层或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到支持物。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

[0290] 至少4个支链(例如,4、5、6、7、8、9、10或更多个支链)的支化PEG。在一些实施例中,低非特异性结合涂层具有可以作为水接触角测量的亲水性程度,其中水接触角不超过45度(例如,不超过5度、不超过10度、不超过15度、不超过20度、不超过25度、不超过30度、不超过35度、不超过40度或不超过45度)。在一些实施例中,固定到支持物或固定到支持物上的涂层的多联体分子的密度为每mm^2约10^2至10^6(例如,约10^2、10^3、10^4、10^5或10^6)。在一些实施例中,固定到支持物或固定到支持物上的涂层的多联体分子的密度为每mm^2约10^6至10^9(例如,约10^6、10^7、10^8或10^9)。在一些实施例中,固定到支持物或固定到支持物上的涂层的多联体分子的密度为每mm^2约10^9至10^{12}(例如,约10^9、10^{10}、10^{11}或10^{12})。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)上的预定位点处固定到支持物或固定到支持物上的涂层。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)上的随机位点处固定到支持物上的涂层。

[0291] 压实寡核苷酸

[0292] 在一些方面,本公开提供了用于在存在或不存在多个压实寡核苷酸的情况下进行滚环扩增的组合物和方法。压实寡核苷酸是单链的并且可以包括5'区、任选的内部区和3'区。压实寡核苷酸的5'区和3'区可与多联体杂交,以将多联体的远端部分拉到一起,引起多联体的压实以形成纳米结构。例如但不限于,压实寡核苷酸的5'区可以被设计成与多联体分子的第一部分(例如,第一通用衔接子序列)杂交,并且压实寡核苷酸的3'区可以被设计成与多联体分子的第二部分(例如,第二通用衔接子序列)杂交。与在不存在压实寡核苷酸的情况下生成的多联体相比,在滚环扩增期间包含压实寡核苷酸可促进具有更紧密尺寸和形状的纳米结构的形成。核酸纳米结构的紧凑且稳定的特征例如通过增加信号强度来改善测序准确性,并且它们在多个测序循环期间保持其形状和大小。在一些实施例中,压实寡核苷酸的3'端被设计成阻断引物延伸。

[0293] 在一些实施例中,压实寡核苷酸是包含DNA、RNA或DNA和RNA的组合的单链寡核苷酸。压实寡核苷酸可以为任何长度,包括20至150个核苷酸,例如20、30、40、50、60、70、80、90、100、110、120、130、140或150个核苷酸,或其间的任何范围。在一些实施例中,压实寡核苷酸的长度可以为30至100个核苷酸。在一些实施例中,压实寡核苷酸的长度可以为40至80个核苷酸。

[0294] 在一些实施例中,压实寡核苷酸包含5'区和3'区,以及任选的5'区与3'区之间的居间区。居间区可以为任何长度,例如,长度为约2至20个核苷酸,例如2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个核苷酸,或其间的任何范围。在一些实施例中,居间区包括具有连续相同碱基(例如,AAA、GGG、CCC、TTT或UUU)的均聚物。在一些实施例中,居间区包括非均聚物序列。

[0295] 在一些实施例中,压实寡核苷酸的5'区可以沿其长度与多联体分子的第一部分完全互补或部分互补。例如但不限于,压实寡核苷酸的5'区包含可与表1至3中列出的通用衔接子序列杂交的序列,包括表面捕获引物结合位点、表面钉扎引物结合位点、正向测序引物结合位点或反向测序引物结合位点。

[0296] 在一些实施例中,压实寡核苷酸的3'区可以沿其长度与多联体分子的第二部分完全互补或部分互补。例如但不限于,压实寡核苷酸的3'区包含可以与表1至3中列出的通用衔接子序列杂交的序列,包括表面捕获引物结合位点、表面钉扎引物结合位点、正向测序引物结合位点或反向测序引物结合位点。

[0297] 压实寡核苷酸的5'区可与多联体分子的第一通用序列部分杂交。压实寡核苷酸的3'区可与多联体分子的第二通用序列部分杂交。压实寡核苷酸的5'区和3'区可与多联体杂交,以将多联体的远端部分拉到一起,引起多联体的压实以形成纳米结构。

[0298] 在一些实施例中,压实寡核苷酸的5'区可以具有与3'区相同的序列。在一些实施例中,压实寡核苷酸的5'区可以具有与3'区不同的序列。在一些实施例中,压实寡核苷酸的3'区可以具有作为5'区的反向序列的序列。

[0299] 在一些实施例中,压实寡核苷酸在其5'端或3'端处包含一个或多个经修饰的碱基或键以赋予某些功能。在一些实施例中,压实寡核苷酸在其5'端和/或3'端处包含至少一个硫代磷酸酯键,例如以赋予核酸外切酶抗性。在一些实施例中,在3'端处或附近的至少一个核苷酸包含赋予核酸外切酶抗性的2'氟碱基。在一些实施例中,压实寡核苷酸的3'端包含至少一个2'-O-甲基RNA碱基,其阻断聚合酶催化的延伸。在一些实施例中,压实寡核苷酸在其3'端包含3'反向dT,其阻断聚合酶催化的延伸。在一些实施例中,压实寡核苷酸包含阻断聚合酶催化的延伸的3'磷酸化。在一些实施例中,压实寡核苷酸的内部区包含至少一种锁核酸(LNA),其增加通过将压实寡核苷酸与多联体分子杂交形成的双链体的热稳定性。

[0300] 压实寡核苷酸可包含具有连续鸟嘌呤的至少一个区。例如,压实寡核苷酸可包含具有2、3、4、5、6或更多个连续鸟嘌呤的至少一个区。在一些实施例中,压实寡核苷酸包含可形成鸟嘌呤四联体结构的四个连续鸟嘌呤。例如可经由Hoogsteen氢键合使鸟嘌呤四联体结构稳定。可通过中心阳离子(例如,钾、钠、锂、铷或铯)使鸟嘌呤四联体结构稳定。

[0301] 可在存在具有至少四个连续鸟嘌呤的多个压实寡核苷酸的情况下进行滚环扩增反应。所得的多联体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结构。在纳米结构中形成鸟嘌呤四联体和G-四联体可以增加纳米结构的稳定性,以保持其紧凑的大小和形状,其可以承受pH、温度和/或试剂的重复流动的变化。

[0302] 用于测序的方法

[0303] 本公开内容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。术语多联体分子和模板分子在本文中可互换使用。

[0304] 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)可用于生成经固定的多联体,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(600)中的任何衔接子序列的串联重复单元。

[0305] 在一些实施例中,示例性串联重复单元包含:(i)具有针对表面捕获引物(子区413)的结合序列的通用衔接子序列;(ii)具有针对反向测序引物(130)的结合序列的通用衔接子序列;(iii)目的序列(110);(iv)具有针对正向测序引物(120)的结合序列的通用衔接子序列;(v)具有针对表面钉扎引物(子区411)的结合序列的通用衔接子序列;以及(vi)样品索引、短随机序列(NNN)和/或唯一分子索引(UMI)(子区412)。

[0306] 在一些实施例中,示例性串联重复单元包含:(i)具有针对表面捕获引物(子区412)的结合序列的通用衔接子序列;(ii)样品索引、短随机序列(NNN)和/或唯一分子索引(UMI)(子区413);(iii)具有针对反向测序引物(130)的结合序列的通用衔接子序列;(iv)目的序列(110);(v)具有针对正向测序引物(120)的结合序列的通用衔接子序列;以及(vi)具有针对表面钉扎引物(子区411)的结合序列的通用衔接子序列。

[0307] 在一些实施例中,示例性串联重复单元包含:(i)具有针对表面捕获引物(子区413)的结合序列的通用衔接子序列;(ii)具有针对反向测序引物(150)的结合序列的通用衔接子序列;(iii)目的序列(110);(iv)具有针对正向测序引物(140)的结合序列的通用衔接子序列;(v)样品索引、短随机序列(NNN)和/或唯一分子索引(UMI)(子区412);以及(vi)具有针对表面钉扎引物(子区411)的结合序列的通用衔接子序列。

[0308] 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目的增加增加了沿多联体用于与多个测序引物(例如,具有通用序列的测序引物)杂交的位点的数目,多个测序引物充当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可以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,可以沿特定多联体分子形成多个结合复合物,每个结合复合物包含与多价分子结合的测序聚合酶,其中多个结合复合物保持稳定而不解离,导致持续时间增加,这增加了信号强度并减少了成像时间。

[0309] 使用工程化聚合酶的测序方法

[0310] 在一些方面,本公开提供了多联体模板分子,其可以使用采用标记或未标记的链终止核苷酸的任何核酸测序方法进行测序,其中链终止核苷酸包括3'-O-叠氮基团(或3'-O-甲基叠氮基团)或糖3'位置上的任何其他类型的大块阻断基团。在一些实施例中,可以使用标记的多价分子和未标记的链终止核苷酸使用亲和力测序方法(SBA)对多联体模板分子进行测序。在一些实施例中,可以使用合成测序(SBS)方法对多联体模板分子进行测序,该方法采用标记的链终止核苷酸。在一些实施例中,可以使用采用未标记的链终止核苷酸的结合测序方法(SBB)对多联体模板分子进行测序。在一些实施例中,可以使用经磷酸链标记

的核苷酸对多联体模板分子进行测序。

[0311] 用于使用核苷酸类似物进行测序的方法

[0312] 在一些方面,本公开提供了用于测序的方法,其包括步骤(a):使测序聚合酶与(i)核酸多联体分子和(ii)核酸引物接触,其中接触在适合使测序聚合酶与核酸多联体分子(该核酸多联体分子与核酸引物杂交)结合的条件下进行,其中与核酸引物杂交的核酸多联体分子形成核酸双链体。在一些实施例中,测序聚合酶包含重组突变体测序聚合酶。在一些实施例中,引物包含3'可延伸端。

[0313] 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,多个核苷酸包括至少一个缺乏链终止部分的核苷酸。

[0314] 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。

[0315] 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(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基因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,该接头可从碱基切割/去除。在一些实施例中,多个核苷酸中的至少一者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告基因部分(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的3'端的核苷酸来确定,从而确定多联体分子的序列。

[0316] 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,多个复合聚合酶的密度为每mm^2约10^2

至10^{15}个(例如,10^{2}至10^{15}个或更多,例如,10^{2}、10^{3}、10^{4}、10^{5}、10^{6}、10^{7}、10^{8}、10^{9}、10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、10^{15}个)固定到支持物的复合聚合酶。

[0317] 用于对核酸进行测序的两阶段方法

[0318] 在一些方面,本公开提供了一种用于对核酸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。

[0319] 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与核酸双链体结合的第一测序聚合酶,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。

[0320] 在一些实施例中,在用于对多联体分子进行测序的方法中,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目的序列或不同的靶目的序列。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到支持物的。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到支持物时,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/或核酸引物被固定到支持物上的10^{2}至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多个第一复合聚合酶,其固定到支持物上的10^{2}至10^{15}个不同位点(例如,10^{2}至10^{15}个位点或更多,例如,10^{2}个位点、10^{3}个位点、10^{4}个位点、10^{5}个位点、10^{6}个位点、10^{7}个位点、10^{8}个位点、10^{9}个位点、10^{10}个位点、10^{11}个位点、10^{12}个位点、10^{13}个位点、10^{14}个位点、10^{15}个位点)。在一些实施例中,支持物上的多个经固定的第一复合聚合酶被固定到支持物上的预定或随机位点。在一些实施例中,多个经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经固定的复合聚合酶以大规模并行方式与试剂溶液反应。

[0321] 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(例如,核苷酸单元)(例如,图22至25)。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者结合,从而形成多个多价-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。在一些方面,多个多价分子包括至少一个具有多个核苷酸臂的多价分子(例如,图22至25),每个核苷酸臂附接有核苷酸类似物(例如,核苷酸类似物单元),其中核苷酸类似物包括在糖2'和/或3'位置处的链终止部分。在一些实施例中,多个多价分子包括至少一个多价分子,该多价分子包含多个核苷酸臂,每个核苷酸臂附接有缺乏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告基因部分标记。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,可检测报告基因部分包括荧光团。在一些实施例中,在至少一种包含锶、钡和/或钙的非催化阳离子存在下,进行步骤(b)的接触。

[0322] 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告基因部分标记以允许检测。在一些实施例中,经标记的多价分子包含附接至多价分子的核心、接头和/或核苷酸单元的荧光团。

[0323] 在一些实施例中,用于测序的方法进一步包括步骤(d):识别与多个第一复合聚合酶结合的互补核苷酸单元的碱基,由此确定多联体分子的序列。在一些实施例中,多价分子用对应于附接至核苷酸臂的特定核苷酸单元的可检测报告基因部分进行标记,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)。

[0324] 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。

[0325] 在一些实施例中,用于测序的方法进一步包括步骤(f):将步骤(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使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,由此形成多个第二复合聚合酶,每个第二复合聚合酶包含与核酸双链体结合的第二测序聚合酶。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。

[0326] 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。

[0327] 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,步骤(g)的接触在镁和/或锰的存在下进行。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去除或不可去除的2'和/或3'链终止部分。在一些实施例中,多个核苷酸包括用可检测报告基因部分标记的多个核苷酸。可检测报告基因部分可以包括荧光团。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,该接头可从碱基切割/去除或不可从碱基去除。在一些实施例中,附接至核苷酸的特定可检测报告基因部分(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,多个核苷酸中的至少一者未用可检测报告基因部分标记。在一些实施例中,该多个核苷酸未用可检测

报告基因部分标记。

[0328] 在一些实施例中,用于测序的方法进一步包括步骤(h):当核苷酸用可检测报告基团部分标记时,步骤(h)包括检测掺入核苷酸-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告基因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,当核苷酸未被标记时,则省略检测步骤。

[0329] 在一些实施例中,用于测序的方法进一步包括步骤(i):当核苷酸用可检测报告基因部分标记时,步骤(i)包括识别掺入核苷酸-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,当核苷酸未被标记时,则省略识别步骤。

[0330] 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括至少一个具有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,从掺入的核苷酸中去除链终止部分。

[0331] 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)至(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检测和识别在步骤(c)和(d)中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)中掺入引物的3'端的核苷酸来确定(或确认)。

[0332] 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,多个第一复合聚合酶与多个多价分子的结合形成至少一个亲和力复合物,该方法包括以下步骤:(a)将第一核酸引物、第一测序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,从而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结合至第一测序聚合酶;以及(b)将第二核酸引物、第二测序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,从而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测序聚合酶,其中包含相同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。示例性的多价分子示出于图22至25。

[0333] 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复合聚合酶和第二复合聚合酶;(b)在适合于将来自多个多价分子中的单个多价分子结合至第一复合聚合酶和第二复合聚合酶的条件下,使多个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与第一复合聚合酶结合,该第一复合聚合酶包含与多联体模板分子的第一部分杂交的第一引物,从而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与第二复合聚合酶结合,该第二复合聚合酶包含与多联体模板分子的第二部分杂交的第二引物,从而形成第二结合复合物(例

如,第二三元复合物),其中接触在适合于抑制聚合酶催化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掺入的条件下进行,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合力复合物;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,从而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,从而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。多个核酸引物可以沿多联体模板分子与测序引物结合位点结合。示例性的多价分子示出于图22至25。

[0334] 结合测序

[0335] 在一些方面,本公开提供了用于对本文所述的任何经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)方法包括以下步骤:(a)在三元复合物稳定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,其中至少两种单独的混合物各自包含聚合酶和核苷酸,其中依次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碱基类型、第二碱基类型和第三碱基类型碱基类型的核苷酸同源物接触;(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,并且其中基于步骤(b)中三元复合物的不存在,将下一个正确核苷酸推算为第四碱基类型的核苷酸同源物;(d)在步骤(b)之后将下一个正确核苷酸添加到引发的模板核酸的引物中,从而产生延伸的引物;以及(e)对包含延伸的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)。用于使用磷酸链标记的核苷酸进行测序的方法

[0336] 本公开提供了用于使用与未经固定的模板分子结合的经固定的测序聚合酶进行测序的方法,其中用经磷酸酯链标记的核苷酸进行测序反应。在一些实施例中,共价闭合环状文库分子(600)可以充当未经固定的模板分子。在一些实施例中,测序方法包括步骤(a):提供其上固定有多种测序聚合酶的支持物。在一些实施例中,测序聚合酶包括持续性DNA聚合酶。在一些实施例中,测序聚合酶包括野生型或突变型DNA聚合酶,包括例如但不限于Phi29 DNA聚合酶。在一些实施例中,支持物包括多个单独的隔室,并且测序聚合酶固定到隔室的底部。在一些实施例中,单独的隔室包括光可以穿透的二氧化硅底部。在一些实施例中,单独的隔室包括配置有包括金属包覆膜(例如,铝包覆膜)中的孔的纳米光子限制结构的二氧化硅底部。在一些实施例中,金属包覆中的孔的小孔径是例如大约70nm。在一些实施例中,纳米光子限制结构的高度是大约100nm。在一些实施例中,纳米光子限制结构包括零模式波导(ZMW)。在一些实施例中,纳米光子限制结构含有液体。

[0337] 在一些实施例中,测序方法进一步包括步骤(b):在适合于各个经固定的测序聚合酶与单链环状模板分子结合和适合于各个测序引物与各个单链环状模板分子杂交的条件下,使多个经固定的测序聚合酶与多个单链环状核酸模板分子(例如,共价闭合环状文库分子(600))和多个寡核苷酸测序引物接触,从而生成多个聚合酶/模板/引物复合物。在一些

实施例中,单独的测序引物与单链环状模板分子上的通用测序引物结合位点杂交。

[0338] 在一些实施例中,测序方法进一步包括步骤(c):使多个聚合酶/模板/引物复合物与多个经磷酸酯链标记的核苷酸接触,每个经磷酸酯链标记的核苷酸包括芳香族碱基、五碳糖(例如,核糖或脱氧核糖)和包括3至20个磷酸酯基团(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个磷酸酯基团)的磷酸酯链,其中末端磷酸酯基团与可检测报告基因部分(例如,荧光团)连接。第一磷酸酯基团、第二磷酸酯基团和第三磷酸酯基团可以被称为α、β和γ磷酸酯基团。在一些实施例中,与末端磷酸酯基团附接的特定可检测报告基因部分对应于核苷酸碱基(例如,dATP、dGTP、dCTP、dTTP或dUTP),以允许检测和识别核碱基。在一些实施例中,在适于聚合酶催化的核苷酸掺入的条件下,使多种聚合酶/模板/引物复合物与多个经磷酸酯链标记的核苷酸接触。在一些实施例中,测序聚合酶能够与经互补磷酸酯链标记的核苷酸结合,并掺入与模板分子中的核苷酸相对的互补核苷酸。在一些实施例中,聚合酶催化的核苷酸掺入反应在α磷酸酯基团与β磷酸酯基团之间切割,从而释放与荧光团连接的多磷酸酯链。

[0339] 在一些实施例中,测序方法进一步包括步骤(d):检测由经磷酸酯链标记的核苷酸发出的荧光信号,所述经磷酸酯链标记的核苷酸由测序聚合酶结合并且掺入到测序引物的末端中。在一些实施例中,步骤(d)进一步包括鉴定由测序聚合酶结合并掺入到测序引物的末端中的经磷酸酯链标记的核苷酸。

[0340] 在一些实施例中,测序方法进一步包括步骤(d):重复步骤(c)至(d)至少一次。在一些实施例中,采用经磷酸酯链标记的核苷酸的测序方法可以根据美国专利第7,170,050号;第7,302,146号;和/或第7,405,281号中所述的方法进行。

[0341] 用于多重分析的传统汇集工作流程

[0342] 在一个实施例中,一种用于制备从多个样品来源分离的目的序列的多重混合物的方法包括:(a)提供单链核酸文库分子的两个或更多个群体(100),文库分子(100)的每个群体包含在单独的隔室中,其中给定群体中的核酸文库分子包含(i)目的序列(110);(ii)具有针对正向测序引物(120)的结合序列的通用衔接子序列;以及(iii)具有针对反向测序引物(130)的结合序列的通用衔接子序列(例如,图1)。

[0343] 在一些实施例中,用于制备从多个样品来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):提供多个双链夹板衔接子(200),其中各个双链夹板衔接子(200)包含与第二夹板链(400)杂交的第一夹板链(300)(例如,图2)。在一些实施例中,第二夹板链(400)包含至少一个样品索引序列和任选的短随机序列NNN。

[0344] 在一些实施例中,用于制备从多个样品来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):在单独的隔室中,使单链核酸文库分子(100)的群体与配额的多个双链夹板衔接子(200)接触,其中该接触在适合于使第一夹板链(300)的部分与文库分子(100)的部分杂交的条件下进行,从而使文库分子成环以生成文库-夹板复合物(500)的群体,使得各个第一夹板链的区(320)与各个文库分子(100)的针对正向测序引物结合位点(120)的通用衔接子序列杂交,并且各个第一夹板链的区(330)与各个文库分子(100)的针对反向测序引物结合位点(130)的通用衔接子序列杂交,其中文库-夹板复合物(500)中的每一者包含在文库分子的5'端与第二夹板链(300)的3'端之间的第一缺口,其中文库-夹板复合物(500)中的每一者包含在第二夹板链(300)的5'端与文库分子(100)的3'端之间第二

缺口,并且其中第一缺口和第二缺口是可酶促连接的(例如,参见图3和4)。

[0345] 在一些实施例中,用于制备从多个样品来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):在适合于酶促连接第一缺口和第二缺口的条件下,在单独的隔室中使文库-夹板复合物(500)的群体与连接酶接触,从而生成各自与第一夹板链(300)杂交的共价闭合环状文库分子(600)的群体(例如,参见图4)。

[0346] 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(600)的群体汇集在一起,以生成共价闭合环状文库分子(600)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。

[0347] 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来源(例如,2至10、10至50、50至100、100至250、其间的任何范围或超过250个不同的样品来源)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,福尔马林固定的石蜡包埋的;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。

[0348] 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至10、10至50、50至100、100至250、其间的任何范围,或者超过250个单链核酸文库分子(100)的不同群体。在一些实施例中,可以在步骤(e)中将任意数量的共价闭合环状文库分子(600)的不同群体汇集在一起,例如2至10、10至50、50至100、100至200、其间的任意范围,或者可以将超过200个共价闭合环状文库分子(600)的不同群体汇集在一起。

[0349] 本领域技术人员将认识到,在步骤(a)中可以使用任何数量的单独隔室(例如,2至10、10至50、50至100、100至250、其间的任何范围或更多个单独隔室)(例如,多孔板,诸如例如96孔板)。

[0350] 在一些实施例中,与步骤(d)或(e)的共价闭合环状文库分子(600)杂交的第一夹板链(300)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。

[0351] 在一些实施例中,在步骤(d)或(e)中,与第一夹板链(300)杂交的共价闭合环状文库分子(600)群体可以任选地与至少一种核酸外切酶反应,以移除多个第一夹板链(300)并保留多个共价闭合环状文库分子(600)。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。

[0352] 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。

[0353] 用于改进碱基识别的样品索引

[0354] 通常,需要制备待分布到支持物(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。

[0355] 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。在一些实施例中,最佳的高多样性模板分子包括在测序运行的每个循环中呈现的所有四种核苷酸的比例大致相等的目的序列(插入)区。在一些实施例中,低多样性模板分子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。为了克服低多样性模板分子的问题,可以将少量从PhiX噬菌体制备的高多样性分子与目的模板分子(例如,PhiX掺入文库)混合并一起测序,例如在同一流通池上。虽然PhiX掺入文库提供了核苷酸多样性,但它也占据流通池上的空间,从而替换携带目的序列的模板分子并减少可从模板分子获得的测序数据量(例如,减少测序通量)。

[0356] 克服低多样性模板分子的问题的另一种方法是制备具有至少一个被设计成颜色平衡的样品索引序列的模板分子。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型样品索引集,将样品索引序列设计为单个或成对的样品索引可能具有挑战性(例如,参见图19和20)。

[0357] 如本文所述,克服对低多样性模板分子(例如,在支持物上以高密度)测序的挑战的替代方法是制备具有至少一个样品索引序列的模板分子,该样品索引序列包括与样品索引序列直接连接的短随机序列(例如,NNN),其中短随机序列(NNN)提供核苷酸多样性和颜色平衡。在一些实施例中,样品索引序列包含直接连接至样品索引序列的短随机序列(NNN)。在一些实施例中,短随机序列(NNN)位于样品索引序列的上游或下游。一些示例性的样品索引序列包括但不限于:NNNGTAGGAGCC;NNNCCGCTGCTA;NNNAACAACAAG;NNNGGTGGTCTA;NNNTTGGCCAAC;NNNCAGGAGTGC;和NNNATCACACTA(例如,参见表3)。

[0358] 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,例如但不限于NNNGTAGGAGCC,该群体含有各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。

[0359] 在具有样品索引的模板分子群体中,样品索引的短随机序列(NNN)可以提供高核苷酸多样性,其包括大约相等比例的所有四种核苷酸(例如,A、G、C、T和/或U),这四种核苷酸将在测序运行的每个循环中表示(参见图19和20)。短随机序列的高核苷酸多样性也可以在测序运行的每个循环期间提供颜色平衡。设计样品索引以包括短随机序列(例如,NNN)的优点是,在模板分子的低重群体(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19和20)。另外,短随机序列(例如,NNN)的核苷酸多样性可以消除包括PhiX掺入文库的需要,或者允许使用减少量的PhiX掺入文库来分布到流通池上并测序。

[0360] 模板分子可以包括包含短随机序列(NNN)的第一样品索引序列和缺乏短随机序列的第二样品索引序列。在一些实施例中,仅来自具有短随机序列(NNN)的样品索引序列的测序数据用于聚合酶群落作图和模板配准,因为短随机序列(例如,NNN)提供足够的核苷酸多样性和颜色平衡。两种类型的样品索引可用于在多重测定中区分从不同样品来源获得的目的序列。

[0361] 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模板分子进行测序的挑战。例如但不限于,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。例如但不限于,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,并且然后对目的序列区进行测序。在具有样品索引的模板分子的群体中,短随机序列(例如,NNN)可以提供可能不是由模板分子的目的序列区提供的核苷酸多样性。样品索引的序列可以为聚合酶群落作图和模板配准提供了改进的核苷酸多样性和颜色平衡。

[0362] 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可以相对较短(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)。在一些实施例中,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,以去除任何与模板分子杂交的剩余产物,这可能会损坏模板分子。

[0363] 在一些方面,本公开提供了模板分子,每个模板分子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来源获得的目的序列,其中至少一个样品索引序列包括连接至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。

[0364] 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。

[0365] 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,例如且不限于AA、TT、CC、GG、UU、AAA、TTT、CCC、GGG或UUU。在一些实施例中,在模版分子的群体中,通用样品索引序列包括具有高多样性序列的短随机序列,其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),这四种核苷酸将在测序运行的每个循环中表示。

[0366] 在一些实施例中,短随机序列(例如,NNN)包含3至20个核苷酸、3至10个核苷酸、3至8个核苷酸、3至6个核苷酸、3至5个核苷酸、或3至4个核苷酸、或其间的任何范围。

[0367] 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。

[0368] 在一些实施例中,通用样品索引序列包含5至20个核苷酸、7至18个核苷酸、或9至16个核苷酸、或其间的任何范围。

[0369] 在一些实施例中,样品索引的群体中的各个样品索引序列包括通用样品索引序列和短随机序列(例如,NNN)。在一些实施例中,样品索引序列的群体中的短随机序列具有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%至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。

[0370] 在一些实施例中,在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%至30%、约15%至35%、约10%至40%或其间的任何范围。在一些

实施例中,在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%至30%、约15%至35%、约10%至40%或其间的任何范围。在一些实施例中,在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%至30%、约15%至35%、约10%至40%或其间的任何范围。在一些实施例中,在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%至30%、约15%至35%、约10%至40%或其间的任何范围。

[0371] 在一些实施例中,在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%至65%。在一些实施例中,在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%至65%。

[0372] 在一些实施例中,在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。

[0373] 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19)。在一些实施例中,在具有样品索引的模版分子的群体中,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。

[0374] 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,使用经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19)。标记的核苷酸在测序反应期间发出荧光学信号。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,生成经固定的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,可用于生成经固定的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(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(110)碱基识别的准确性和质量分数。

[0375] 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),这些核苷酸单元附接至给定多价分子中的核苷酸臂。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,这些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22至25)。在一些实施例中,多价分子的至少一个核苷酸臂包含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,使用经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,使用经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19)。标记的多价分子在测序反应期间发出荧光学信号。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(聚合酶群落)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,可用于生成经固定的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(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(110)碱基识别的准确性和质量分数。

[0376] 第一实施例:对样品索引序列测序的顺序

[0377] 在一些实施例中,共价闭合环状文库分子(600)包含:(i)第一子区(411),其包含针对经固定的表面钉扎引物的通用结合序列;(ii)第二子区(412),其包含短随机序列(NNN)和样品索引序列;(iii)第三子区(413),其包含针对经固定的表面捕获引物的通用结合序列;(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列(例如,参见图4)。在一些实施例中,通过使多个共价闭合环状文库分子(600)与固定到支持物的多个表面捕获引物杂交,对多个共价闭合环状文库分子(600)进行滚环扩增,其中表面捕获引物引发扩增,从而生成多个经固定的多联体,每个多联体具有其同源共价闭合环状文库分子(600)的串联重复序列。在一些实施例中,滚环扩增在存在多个压实寡核苷酸的情况下进行,压实寡核苷酸可以在其针对表面钉扎引物的通用结合序列、针对表面捕获引物的通用结合序列、针对反向测序引物的通用结合序列和/或针对正向测序引物的通用结合序列处与多联体杂交。在一些实施例中,多个经固定的

捕获引物缺乏尿嘧啶碱基。在一些实施例中,滚环扩增反应包括多个核苷酸,包括dATP、dGTP、dCTP、dTTP和dUTP,以生成多个经固定的多联体,其中各个多联体分子包含随机分布的尿嘧啶碱基。在一些实施例中,对经固定的多联体的至少一部分进行测序。

[0378] 在一些实施例中,测序的顺序包括:(1)对多联体的短随机序列(NNN)和样品索引序列(412)进行测序;以及(2)对多联体的插入区(110)进行测序。在一些实施例中,测序的顺序进一步包括:(3)进行成对轮转反应,使得经固定的多联体分子被与该多联体分子互补的经固定的第二链替换;以及(4)对第二链上的插入区(110)进行测序。在一些实施例中,对短随机序列(NNN)和样品索引序列(412)进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。

[0379] 在一些实施例中,用于对固定到支持物的多联体分子进行测序的方法包括步骤(a):使多联体分子与与第三子区(413)的至少一部分杂交的第一多个可溶性测序引物杂交,并对短随机序列(例如,NNN)和第三子区(413)的通用样品索引序列进行测序,从而生成与经固定的多联体分子杂交的多个样品索引延伸产物,其中多个样品索引延伸产物与短随机序列(例如,NNN)和通用样品索引序列互补。

[0380] 在一些实施例中,用于测序的方法进一步包括步骤(b):去除第一多个样品索引延伸产物并保留经固定的多联体分子。

[0381] 在一些实施例中,用于测序的方法进一步包括步骤(c):将保留的经固定的多联体分子与第二多个可溶性测序引物杂交,第二多个可溶性测序引物与针对正向测序引物(120)的通用结合序列杂交,并对插入区(110)进行测序,从而生成与经固定的多联体分子杂交的多个插入序列延伸产物。

[0382] 在一些实施例中,用于测序的方法进一步包括步骤(d):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经固定的多联体分子杂交的多个插入序列延伸产物,以生成与经固定的多联体分子(包括经固定的捕获引物)杂交的第二链延伸产物。

[0383] 在一些实施例中,用于测序的方法进一步包括步骤(e):通过在经固定的多联体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,来去除经固定的多联体分子,从而生成含有间隙的多联体分子,同时保留在步骤(d)中生成的第二链延伸产物,其中各个第二链延伸产物通过与经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(d)和(e)来实现成对轮转。

[0384] 在一些实施例中,用于测序的方法进一步包括步骤(f):将保留的第二链延伸产物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反向测序引物(130)的通用结合序列杂交,并且对插入区(110)至少一部分进行测序。

[0385] 在一些实施例中,用于测序的方法进一步包括:将(i)插入区(110)的序列分配至(ii)样品索引序列,从而将插入区识别为从第一来源获得。在一些实施例中,可以在步骤(c)和/或(f)之后进行分配。

[0386] 在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用包含具有或不具有甲酰胺的SSC(例如,盐水-柠檬酸钠)缓冲液的变性试剂,在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用去杂交试剂在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离

液剂。

[0387] 在一些实施例中,步骤(a)、(c)和(f)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(f)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经标记的核苷酸类似物或未经标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(f)的测序包括进行任何结合测序方法或采用本文所述的经磷酸酯链标记的核苷酸的任何测序方法。

[0388] 在一些实施例中,固定到支持物的多个多联体分子的密度为每mm^2约10^2至10^{15}(例如,10^2、10^3、10^4、10^5、10^6、10^7、10^8、10^9、10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、10^{15})。在一些实施例中,多个多联体分子固定在支持物上的随机位置处。在一些实施例中,多个多联体分子以预定图案固定在支持物上。

[0389] 第二实施例:对样品索引序列测序的顺序

[0390] 在一些实施例中,共价闭合环状文库分子(600)包含:(i)第一子区(411),其包含针对经固定的表面钉扎引物的通用结合序列;(ii)第二子区(412),其包含短随机序列(NNN)和样品索引序列;(iii)第三子区(413),其包含针对经固定的表面捕获引物的通用结合序列;(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列(例如,参见图4)。在一些实施例中,通过使多个共价闭合环状文库分子(600)与固定到支持物的多个表面捕获引物杂交,对多个共价闭合环状文库分子(600)进行滚环扩增,其中表面捕获引物引发扩增,从而生成多个经固定的多联体,每个多联体具有其同源共价闭合环状文库分子(600)的串联重复序列。在一些实施例中,滚环扩增在存在多个压实寡核苷酸的情况下进行,压实寡核苷酸可以在其针对表面钉扎引物的通用结合序列、针对表面捕获引物的通用结合序列、针对反向测序引物的通用结合序列和/或针对正向测序引物的通用结合序列处与多联体杂交。在一些实施例中,多个经固定的捕获引物缺乏尿嘧啶碱基。在一些实施例中,滚环扩增反应包括多个核苷酸,包括dATP、dGTP、dCTP、dTTP和dUTP,以生成多个经固定的多联体,其中各个多联体分子包含随机分布的尿嘧啶碱基。在一些实施例中,对经固定的多联体的至少一部分进行测序。

[0391] 在一些实施方案中,测序的顺序包括:(1)对多联体的插入区(110)进行测序;以及(2)对多联体的短随机序列(NNN)和样品索引序列(412)进行测序。在一些实施例中,测序的顺序进一步包括:(3)进行成对轮转反应,使得经固定的多联体分子被与该多联体分子互补的经固定的第二链替换;以及(4)对第二链上的插入区(110)进行测序。在一些实施例中,对短随机序列(NNN)和样品索引序列(412)进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。

[0392] 在一些实施例中,用于对固定到支持物的多联体分子进行测序的方法包括步骤(a):使经固定的多联体分子与第一多个可溶性测序引物杂交,第一多个可溶性测序引物与针对正向测序引物(120)的通用结合序列杂交,并对插入区(110)进行测序,从而生成与经固定的多联体分子杂交的多个插入序列延伸产物。

[0393] 在一些实施例中,用于测序的方法进一步包括步骤(b):去除多个插入序列延伸产物并保留经固定的多联体分子。

[0394] 在一些实施例中,用于测序的方法进一步包括步骤(c):使保留的多联体分子与与

第三子区(413)的至少一部分杂交的第二多个可溶性测序引物杂交,并对短随机序列(例如,NNN)和第三子区(412)的通用样品索引序列进行测序,从而生成与经固定的多联体分子杂交的多个样品索引延伸产物,其中多个样品索引延伸产物与短随机序列(例如,NNN)和通用样品索引序列互补。

[0395] 在一些实施例中,用于测序的方法进一步包括步骤(d):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经固定的多联体分子杂交的多个样品索引延伸产物,以生成与经固定的多联体分子(包括经固定的捕获引物)杂交的第二链延伸产物。

[0396] 在一些实施例中,用于测序的方法进一步包括步骤(e):通过在经固定的多联体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,来去除经固定的多联体分子,从而生成含有间隙的多联体分子,同时保留在步骤(d)中生成的第二链延伸产物,其中各个第二链延伸产物通过与经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(d)和(e)来实现成对轮转。

[0397] 在一些实施例中,用于测序的方法进一步包括步骤(f):将保留的第二链延伸产物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反向测序引物(130)的通用结合序列杂交,并且对插入区(110)至少一部分进行测序。

[0398] 在一些实施例中,用于测序的方法进一步包括:将(i)插入区(110)的序列分配至(ii)样品索引序列,从而将插入区识别为从第一来源获得。在一些实施例中,可以在步骤(c)和/或(f)之后进行分配。

[0399] 在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用包含具有或不具有甲酰胺的SSC(例如,盐水-柠檬酸钠)缓冲液的变性试剂,在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用去杂交试剂在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离液剂。

[0400] 在一些实施例中,步骤(a)、(c)和(f)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(f)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经标记的核苷酸类似物或未经标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(f)的测序包括进行任何结合测序方法或采用本文所述的经磷酸酯链标记的核苷酸的任何测序方法。

[0401] 在一些实施例中,固定到支持物的多个多联体分子的密度为每mm^2约10^2至10^{15}(例如,10^2、10^3、10^4、10^5、10^6、10^7、10^8、10^9、10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、10^{15})。在一些实施例中,多个多联体分子固定在支持物上的随机位置处。在一些实施例中,多个多联体分子以预定图案固定在支持物上。

[0402] 第三实施例:测序的顺序

[0403] 在一些实施例中,共价闭合环状文库分子(600)包含:(i)第一子区(411),其包含针对经固定的表面钉扎引物的通用结合序列;(ii)第二子区(412),其包含短随机序列(NNN)和样品索引序列;(iii)第三子区(413),其包含针对经固定的表面捕获引物的通用结合序列;(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对

正向测序引物(120)的通用结合序列(例如,参见图4)。在一些实施例中,通过使多个共价闭合环状文库分子(600)与固定到支持物的多个表面捕获引物杂交,对多个共价闭合环状文库分子(600)进行滚环扩增,其中表面捕获引物引发扩增,从而生成多个经固定的多联体,每个多联体具有其同源共价闭合环状文库分子(600)的串联重复序列。在一些实施例中,滚环扩增在存在多个压实寡核苷酸的情况下进行,压实寡核苷酸可以在其针对表面钉扎引物的通用结合序列、针对表面捕获引物的通用结合序列、针对反向测序引物的通用结合序列和/或针对正向测序引物的通用结合序列处与多联体杂交。在一些实施例中,多个经固定的捕获引物缺乏尿嘧啶碱基。在一些实施例中,滚环扩增反应包括多个核苷酸,包括dATP、dGTP、dCTP、dTTP和dUTP,以生成多个经固定的多联体,其中各个多联体分子包含随机分布的尿嘧啶碱基。在一些实施例中,对经固定的多联体的至少一部分进行测序。

[0404] 在一些实施例中,测序的顺序包括:(1)对多联体的插入区(110)的前3至5个碱基进行测序;(2)对多联体的短随机序列(NNN)和样品索引序列(412)进行测序;以及(3)对多联体的插入区(110)的剩余部分进行测序。在一些实施例中,测序的顺序进一步包括:(4)进行成对轮转反应,使得经固定的多联体分子被与该多联体分子互补的经固定的第二链替换;以及(5)对第二链上的插入区(110)进行测序。在一些实施例中,对多联体的插入区(110)的前3至5个碱基进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。在一些实施例中,对短随机序列(NNN)和样品索引序列(412)进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。

[0405] 在一些实施例中,用于对固定到支持物的多联体分子进行测序的方法包括步骤(a):使多联体分子与第一多个可溶性测序引物杂交,第一多个可溶性测序引物与针对正向测序引物(120)的通用结合序列杂交,并对插入区(110)的前3至5个碱基进行测序,从而生成与经固定的多联体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(110)的前3至5个碱基的序列可以为聚合酶群落作图和多联体配准提供足够的序列多样性和颜色平衡。

[0406] 在一些实施例中,用于测序的方法进一步包括步骤(b):去除多个插入延伸产物并保留经固定的多联体分子。

[0407] 在一些实施例中,用于测序的方法进一步包括步骤(c):使保留的多联体分子与与第三子区(413)的至少一部分杂交的第二多个可溶性测序引物杂交,并对短随机序列(例如,NNN)和第三子区(412)的通用样品索引序列进行测序,从而生成与经固定的多联体分子杂交的多个样品索引延伸产物,其中多个样品索引延伸产物与短随机序列(例如,NNN)和通用样品索引序列互补。

[0408] 在一些实施例中,用于测序的方法进一步包括步骤(d):去除多个样品索引延伸产物并保留经固定的多联体分子。

[0409] 在一些实施例中,用于测序的方法进一步包括步骤(e):使保留的多联体分子与第三多个可溶性测序引物杂交,第三多个可溶性测序引物与针对正向测序引物(120)的通用结合序列杂交,并对插入区(110)的其余部分进行测序或者对插入区(110)的全长进行测序,从而生成与经固定的多联体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。在一些实施例中,在步骤(e)中,可以使用经标记的核苷酸和/或经标记的多价分子对插入区(110)的前3至5个碱基进行测序。在一些实施例中,在步骤(e)中,可

以使用未经标记的核苷酸和/或未经标记的多价分子对插入区(110)的前3至5个碱基进行测序(例如,暗测序)。在一些实施例中,在步骤(e)中,可以使用经标记的核苷酸和/或经标记的多价分子对插入区(110)的剩余部分进行测序。在一些实施例中,在步骤(e)中,可以使用经标记的核苷酸和/或经标记的多价分子对插入区(110)的全长进行测序。

[0410] 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经固定的多联体分子杂交的多个插入序列延伸产物,以生成与经固定的多联体分子(包括经固定的捕获引物)杂交的第二链延伸产物。

[0411] 在一些实施例中,用于测序的方法进一步包括步骤(g):通过在经固定的多联体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,来去除经固定的多联体分子,从而生成含有间隙的多联体分子,同时保留在步骤(f)中生成的第二链延伸产物,其中各个第二链延伸产物通过与经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(f)和(g)来实现成对轮转。

[0412] 在一些实施例中,用于测序的方法进一步包括步骤(h):将保留的第二链延伸产物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反向测序引物(130)的通用结合序列杂交,并且对插入区(110)至少一部分进行测序。

[0413] 在一些实施例中,用于测序的方法进一步包括:将(i)插入区(110)的序列分配至(ii)样品索引序列,从而将插入区识别为从第一来源获得。在一些实施例中,可以在步骤(c)、(e)和/或(h)之后进行分配。

[0414] 在一些实施例中,步骤(b)和(d)的多个测序延伸产物的去除可以使用包含具有或不具有甲酰胺的SSC(例如,盐水-柠檬酸钠)缓冲液的变性试剂,在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,步骤(b)和(d)的多个测序延伸产物的去除可以使用去杂交试剂在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离液剂。

[0415] 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以及经标记或未经标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(e)的测序包括进行任何结合测序方法或采用本文所述的经磷酸酯链标记的核苷酸的任何测序方法。

[0416] 在一些实施例中,固定到支持物的多个多联体分子的密度为每mm^2约10^2至10^{15}(例如,10^2、10^3、10^4、10^5、10^6、10^7、10^8、10^9、10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、10^{15})。在一些实施例中,多个多联体分子固定在支持物上的随机位置处。在一些实施例中,多个多联体分子以预定图案固定在支持物上。

[0417] 第四实施例:测序的顺序

[0418] 在一些实施例中,共价闭合环状文库分子(600)包含:(i)第一子区(411),其包含针对经固定的表面钉扎引物的通用结合序列;(ii)第二子区(412),其包含短随机序列(NNN)和样品索引序列;(iii)第三子区(413),其包含针对经固定的表面捕获引物的通用结合序列;(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对

正向测序引物(120)的通用结合序列(例如,参见图4)。在一些实施例中,通过使多个共价闭合环状文库分子(600)与固定到支持物的多个表面捕获引物杂交,对多个共价闭合环状文库分子(600)进行滚环扩增,其中表面捕获引物引发扩增,从而生成多个经固定的多联体,每个多联体具有其同源共价闭合环状文库分子(600)的串联重复序列。在一些实施例中,滚环扩增在存在多个压实寡核苷酸的情况下进行,压实寡核苷酸可以在其针对表面钉扎引物的通用结合序列、针对表面捕获引物的通用结合序列、针对反向测序引物的通用结合序列和/或针对正向测序引物的通用结合序列处与多联体杂交。在一些实施例中,多个经固定的捕获引物缺乏尿嘧啶碱基。在一些实施例中,滚环扩增反应包括多个核苷酸,包括dATP、dGTP、dCTP、dTTP和dUTP,以生成多个经固定的多联体,其中各个多联体分子包含随机分布的尿嘧啶碱基。在一些实施例中,对经固定的多联体的至少一部分进行测序。

[0419] 在一些实施例中,测序的顺序包括:(1)对多联体的插入区(110)的前3至5个碱基进行测序;以及(2)对多联体的短随机序列(NNN)和样品索引序列(412)进行测序。在一些实施例中,测序的顺序进一步包括:(3)进行成对轮转反应,使得经固定的多联体分子被与该多联体分子互补的经固定的第二链替换;以及(4)对第二链上的插入区(110)进行测序。在一些实施例中,对多联体的插入区(110)的前3至5个碱基进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。在一些实施例中,对短随机序列(NNN)和样品索引序列(412)进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。

[0420] 在一些实施例中,用于对固定到支持物的多联体分子进行测序的方法包括步骤(a):使多联体分子与第一多个可溶性测序引物杂交,第一多个可溶性测序引物与针对正向测序引物(120)的通用结合序列杂交,并对插入区(110)的前3至5个碱基进行测序,从而生成与经固定的多联体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(110)的前3至5个碱基的序列可以为聚合酶群落作图和多联体配准提供足够的序列多样性和颜色平衡。

[0421] 在一些实施例中,用于测序的方法进一步包括步骤(b):去除多个插入延伸产物并保留经固定的多联体分子。

[0422] 在一些实施例中,用于测序的方法进一步包括步骤(c):使保留的多联体分子与与第三子区(413)的至少一部分杂交的第二多个可溶性测序引物杂交,并对短随机序列(例如,NNN)和第三子区(412)的通用样品索引序列进行测序,从而生成与经固定的多联体分子杂交的多个样品索引延伸产物,其中多个样品索引延伸产物与短随机序列(例如,NNN)和通用样品索引序列互补。

[0423] 在一些实施例中,用于测序的方法进一步包括步骤(d):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经固定的多联体分子杂交的多个插入序列延伸产物,以生成与经固定的多联体分子(包括经固定的捕获引物)杂交的第二链延伸产物。

[0424] 在一些实施例中,用于测序的方法进一步包括步骤(e):通过在经固定的多联体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,来去除经固定的多联体分子,从而生成含有间隙的多联体分子,同时保留在步骤(d)中生成的第二链延伸产物,其中各个第二链延伸产物通过与经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(d)和(e)来实现成对轮转。

[0425] 在一些实施例中,用于测序的方法进一步包括步骤(f):将保留的第二链延伸产物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反向测序引物(130)的通用结合序列杂交,并且对插入区(110)至少一部分进行测序。

[0426] 在一些实施例中,用于测序的方法进一步包括:将(i)插入区(110)的序列分配至(ii)样品索引序列,从而将插入区识别为从第一来源获得。在一些实施例中,可以在步骤(c)和/或(f)之后进行分配。

[0427] 在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用包含具有或不具有甲酰胺的SSC(例如,盐水-柠檬酸钠)缓冲液的变性试剂,在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用去杂交试剂在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离液剂。

[0428] 在一些实施例中,步骤(a)、(c)和(f)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(f)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经标记的核苷酸类似物或未经标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(f)的测序包括进行任何结合测序方法或采用本文所述的经磷酸酯链标记的核苷酸的任何测序方法。

[0429] 在一些实施例中,固定到支持物的多个多联体分子的密度为每mm^2约10^2至10^{15}。在一些实施例中,多个多联体分子固定在支持物上的随机位置处。在一些实施例中,多个多联体分子以预定图案固定在支持物上。

[0430] 第五实施例:对样品索引序列测序的顺序

[0431] 在一些实施例中,共价闭合环状文库分子(600)包含:(i)第一子区(411),其包含针对经固定的表面钉扎引物的通用结合序列;(ii)第二子区(412),其包含针对经固定的表面捕获引物的通用结合序列;(iii)第三子区(413),其包含短随机序列(NNN)和样品索引序列;(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列(例如,参见图4)。

[0432] 在一些实施例中,通过使多个共价闭合环状文库分子(600)与固定到支持物的多个表面捕获引物杂交,对多个共价闭合环状文库分子(600)进行滚环扩增,其中表面捕获引物引发扩增,从而生成多个经固定的多联体,每个多联体具有其同源共价闭合环状文库分子(600)的串联重复序列。在一些实施例中,滚环扩增在存在多个压实寡核苷酸的情况下进行,压实寡核苷酸可以在其针对表面钉扎引物的通用结合序列、针对表面捕获引物的通用结合序列、针对反向测序引物的通用结合序列和/或针对正向测序引物的通用结合序列处与多联体杂交。在一些实施例中,多个经固定的捕获引物缺乏尿嘧啶碱基。在一些实施例中,滚环扩增反应包括多个核苷酸,包括dATP、dGTP、dCTP、dTTP和dUTP,以生成多个经固定的多联体,其中各个多联体分子包含随机分布的尿嘧啶碱基。在一些实施例中,对经固定的多联体的至少一部分进行测序。

[0433] 在一些实施例中,测序的顺序包括:(1)对多联体的短随机序列(NNN)和样品索引序列(413)进行测序;以及(2)对多联体的插入区(110)进行测序。在一些实施例中,测序的

顺序进一步包括:(3)进行成对轮转反应,使得经固定的多联体分子被与该多联体分子互补的经固定的第二链替换;以及(4)对第二链上的插入区(110)进行测序。在一些实施例中,对短随机序列(NNN)和样品索引序列(413)进行测序可以为聚合酶群落作图和多联体配准提供足够的核苷酸多样性和颜色平衡。

[0434] 在一些实施例中,用于对固定到支持物的多联体分子进行测序的方法包括步骤(a):使多联体分子与第一多个可溶性测序引物杂交,第一多个可溶性测序引物与针对反向测序引物(130)的通用结合序列的至少一部分杂交,并对短随机序列(例如,NNN)和第三子区(413)的通用样品索引序列进行测序,从而生成与经固定的多联体杂交的多个样品索引延伸产物分子,其中多个样品索引延伸产物与短随机序列(例如,NNN)和通用样品索引序列互补。

[0435] 在一些实施例中,用于测序的方法进一步包括步骤(b):去除第一多个样品索引延伸产物并保留经固定的多联体分子。

[0436] 在一些实施例中,用于测序的方法进一步包括步骤(c):将保留的经固定的多联体分子与第二多个可溶性测序引物杂交,第二多个可溶性测序引物与针对正向测序引物(120)的通用结合序列杂交,并对插入区(110)进行测序,从而生成与经固定的多联体分子杂交的多个插入序列延伸产物。

[0437] 在一些实施例中,用于测序的方法进一步包括步骤(d):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经固定的多联体分子杂交的多个插入序列延伸产物,以生成与经固定的多联体分子(包括经固定的捕获引物)杂交的第二链延伸产物。

[0438] 在一些实施例中,用于测序的方法进一步包括步骤(e):通过在经固定的多联体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,来去除经固定的多联体分子,从而生成含有间隙的多联体分子,同时保留在步骤(d)中生成的第二链延伸产物,其中各个第二链延伸产物通过与经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(d)和(e)来实现成对轮转。

[0439] 在一些实施例中,用于测序的方法进一步包括步骤(f):将保留的第二链延伸产物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反向测序引物(130)的通用结合序列杂交,并且对插入区(110)至少一部分进行测序。

[0440] 在一些实施例中,用于测序的方法进一步包括:将(i)插入区(110)的序列分配至(ii)样品索引序列,从而将插入区识别为从第一来源获得。在一些实施例中,可以在步骤(c)和/或(f)之后进行分配。

[0441] 在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用包含具有或不具有甲酰胺的SSC(例如,盐水-柠檬酸钠)缓冲液的变性试剂,在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,步骤(b)的多个测序延伸产物的去除可以使用去杂交试剂在促进核酸变性的温度(诸如例如50℃至90℃)下进行。在一些实施例中,去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离液剂。

[0442] 在一些实施例中,步骤(a)、(c)和(f)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(f)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

价分子和经标记的核苷酸类似物或未经标记的核苷酸类似物。在一些实施例中,步骤(a)、(c)和(f)的测序包括进行任何结合测序方法或采用本文所述的经磷酸酯链标记的核苷酸的任何测序方法。

[0443] 在一些实施例中,固定到支持物的多个多联体分子的密度为每mm^2约10^2至10^{15}(例如,10^2、10^3、10^4、10^5、10^6、10^7、10^8、10^9、10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、10^{15})。在一些实施例中,多个多联体分子固定在支持物上的随机位置处。在一些实施例中,多个多联体分子以预定图案固定在支持物上。

[0444] 去杂交试剂

[0445] 在一些方面,本公开提供了一种或多种可以促进任意两条核酸链之间的核酸变性的核酸去杂交试剂。在一些实施例中,去杂交试剂可以促进核酸模板分子与核酸延伸产物之间的核酸变性。在一些实施例中,去杂交试剂可以促进多联体分子与多个样品索引延伸产物之间的核酸变性,同时保留经固定的多联体分子。在一些实施例中,去杂交试剂可以促进多联体分子与多个插入延伸产物之间的核酸变性,同时保留经固定的多联体分子。在一些实施例中,去杂交试剂可以促进第二链分子与多个插入延伸产物之间的核酸变性,同时保留经固定的第二链分子。

[0446] 在一些方面,本公开提供一种或多种核酸去杂交试剂(例如,变性试剂),以及使用核酸去杂交试剂的方法,其中该方法包括测序工作流程的任何顺序。例如,测序工作流程的顺序可以包括上述那些,例如包括从经固定的多联体分子去除多个样品索引延伸产物、从经固定的多联体分子去除多个插入延伸产物和/或从经固定的第二链分子去除多个插入延伸产物的方法。

[0447] 在一些实施例中,去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离液剂。在一些实施例中,去杂交试剂的pH范围为约5至5.25,或pH范围为约5.25至5.5,或pH范围为约5.5至5.75,或pH范围为约5.75至6,或其间的任何范围。

[0448] 在一些实施例中,本文描述的任何去杂交试剂可以包括pH缓冲剂,其可以将试剂的pH维持在适合核酸杂交的范围内。pH缓冲剂可以包括以下中的任一者或者两者或更多者的任何组合:Tris、Tris-HCl、Tris-乙酸盐、Tricine、Bicine、Bis-Tris丙烷、HEPES、MES、MOPS、MOPSO、BES、TES、CAPS、TAPS、TAPSO、ACES、PIPES、乙醇胺(又名2-氨基甲醇;MEA)、柠檬酸盐化合物、柠檬酸盐混合物、NaOH和/或KOH。在一些实施例中,pH缓冲剂可以以约1至100mM、或约10至50mM、或约10至25mM、或其间的任何范围的浓度存在于本文所述的任何去杂交试剂中。在一些实施例中,可以将存在于本文所述的任何试剂中的pH缓冲剂的pH调节至约4至9.5的pH,或约5至9的pH,或约5至8的pH,或约5.5至7的pH,或其间的任何范围。

[0449] 在一些实施例中,本文所述的任何去杂交试剂可包括至少一种还原剂,该还原剂包括DTT(二硫苏糖醇)、2-β巯基乙醇、TCEP、(三(2-羧乙基)膦)、甲酰胺、DMSO(二甲亚砜)、连二亚硫酸钠($Na_2S_2O_4$)、谷胱甘肽、蛋氨酸、甜菜碱、三(3-羟丙基)膦(THPP)和/或N-乙酰半胱氨酸。去杂交试剂可包括浓度为约0.1至0.5M、或约0.5至1M、或约1至2M、或其间任何范围的还原剂。去杂交试剂可以包括浓度为约0.01至0.1mM、或约0.1至1mM、或约1至2.5mM、或约2.5至5mM、或约5至7.5mM、或约7.5至9mM、或约9至12mM、或约12至25mM、或约25至50mM,或其间的任何范围的还原剂。去杂交试剂可包括浓度为约1%至5%、或约5%至10%、或约10%

至20％、或约20％至30％、或约30％至40％、或约40％至50％,或其间的任何范围的还原剂。

[0450] 在一些实施例中,本文所述的任何去杂交试剂可包括至少一种单价盐,该单价盐包括NaCl、KCl、NH_2SO_4和/或谷氨酸钾。去杂交试剂可以包括浓度为约25至500mM、或约50至250mM、或约100至200mM、或约500mM至750mM、或约750mM-1M、或约1M-1.5M,或其间的任何范围的单价盐。

[0451] 在一些实施例中,本文所述的任何去杂交试剂可以包括增加分子拥挤的拥挤剂。在一些实施例中,拥挤剂包括聚乙二醇(PEG,例如,1至50K分子量)、葡聚糖、硫酸葡聚糖、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。基于杂交试剂的总体积,拥挤剂可以以约1％至10％、或约10％至25％、或约25％至50％、或更高的体积百分比存在于去杂交试剂中。

[0452] 在一些实施例中,本文所述的任何去杂交试剂可包括可破坏非共价键诸如氢键或范德华力的离液剂。在一些实施例中,离液剂包括SDS(十二烷基硫酸钠)、脲、硫脲、氯化胍、盐酸胍、硫氰酸胍、异硫氰酸胍、异硫磺酸胍、硫氰酸钾、氯化锂、碘化钠或高氯酸钠。去杂交试剂可包括浓度为约0.5至1M、0.1至1M、或约1至2M、或约2至3M、或约3至4M、或约4至5M或其间的任何范围的离液剂。

[0453] 在一些实施例中,去杂交试剂进一步包含(a)多个经固定的多联体分子和(b)多个与子区(411)、(412)或(413)杂交的可溶性测序引物。在一些实施例中,各个多联体分子包含串联重复多核苷酸单元,其中每个单元包含(i)第一子区(411);(ii)第二子区(412);(iii)第三子区(413);(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列。

[0454] 在一些实施例中,去杂交试剂进一步包含(a)多个经固定的多联体分子和(b)多个样品索引延伸产物,每个样品索引延伸产物包含子区(411)、(412)或(413)的短随机序列(例如,NNN)和通用样品索引序列。在一些实施例中,各个多联体分子包含串联重复多核苷酸单元,其中每个单元包含(i)第一子区(411);(ii)第二子区(412);(iii)第三子区(413);(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列。

[0455] 在一些实施例中,去杂交试剂进一步包含(a)多个经固定的多联体分子和(b)多个与针对正向测序引物(120)的通用结合序列杂交的可溶性测序引物。在一些实施例中,各个多联体分子包含串联重复多核苷酸单元,其中每个单元包含(i)第一子区(411);(ii)第二子区(412);(iii)第三子区(413);(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列。

[0456] 在一些实施例中,去杂交试剂进一步包含(a)多个经固定的多联体分子和(b)多个与针对反向测序引物(130)的通用结合序列杂交的可溶性测序引物。在一些实施例中,各个多联体分子包含串联重复多核苷酸单元,其中每个单元包含(i)第一子区(411);(ii)第二子区(412);(iii)第三子区(413);(iv)针对反向测序引物(130)的通用结合序列;(v)插入序列(110);以及(vi)针对正向测序引物(120)的通用结合序列。

[0457] 在一些实施例中,去杂交试剂进一步包含多种测序聚合酶和多个可检测地标记的多价分子。在一些实施例中,去杂交试剂进一步包含多种测序聚合酶和多个核苷酸类似物。在一些实施例中,多个核苷酸类似物是未标记的。在一些实施例中,多个核苷酸类似物被可

检测地标记。在一些实施例中,多个核苷酸类似物包含未标记的核苷酸类似物和可检测地标记的核苷酸类似物的混合物。

[0458] 在一些实施例中,多个经固定的多联体分子以每mm^2约10^2至10^{15}(例如,10^2、10^3、10^4、10^5、10^6、10^7、10^8、10^9、10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、10^{15})的密度固定到支持物。在一些实施例中,将多个经固定的多联体分子固定到用具有约45至50度的水接触角的亲水性聚合物涂层钝化的支持物。在一些实施例中,多个多联体分子固定在支持物上的随机位置处。在一些实施例中,多个多联体分子以预定图案固定在支持物上。

[0459] 包括去杂交试剂的试剂盒

[0460] 在一些方面,本公开提供了用于进行上述测序工作流程的任何顺序的试剂盒,例如包括从经固定的多联体分子去除多个样品索引延伸产物、从经固定的多联体分子去除多个插入延伸产物和/或从经固定的第二链分子去除多个插入延伸产物的方法。在一些实施例中,试剂盒包括去杂交试剂,该去杂交试剂包括pH缓冲剂、还原剂、单价盐和拥挤剂。在一些实施例中,去杂交试剂进一步包含离液剂。在一些实施例中,试剂盒进一步包括以下中的任一者或者两者或更多者的任意组合:用于对插入区进行测序的多个正向测序引物、用于对插入区进行测序的多个反向测序引物、与子区(411)、(412)或(413)杂交的多个测序引物、多个测序聚合酶、多个可检测地标记的核苷酸类似物、多个未标记的核苷酸类似物、多个可检测地标记的多价分子和/或可以从核苷酸类似物的2'糖位置或3'糖位置去除链终止部分(例如,阻断部分)的切割试剂。

[0461] 对3聚体随机序列进行测序以生成聚合酶克隆图

[0462] 在一些方面,本公开提供了用于对核酸进行测序的方法,该方法包括:(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,其中各个经固定的模板分子包含插入序列区和一个样品索引,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(例如,短随机序列NNN),该通用样品索引序列识别插入序列的样品来源,其中不同的经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,并且其中经固定的模板分子具有不同的插入序列;(b)使用多种可检测地标记的核苷酸试剂对多个经固定的模板分子的3聚体随机序列进行三个循环的聚合酶介导的测序反应,该核苷酸试剂包含不同类型的核碱基A、G、C和T/U的混合物,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,其中测序的三个循环包括对从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,该可检测地标记的核苷酸试剂结合至经固定的模板分子,从而确定多个经固定的模板分子的各个模板分子中的3聚体随机序列的序列,并且其中在多个经固定的模板分子中的第一测序循环、第二测序循环和第三测序循环的每一者中检测和成像A、G、C和T/U的核碱基的平衡多样性;以及(c)使用步骤(b)中获得的图像生成多个经固定的模板分子的位置图,其中插入区的序列不用于生成图。

[0463] 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%至85%、或约5%至60%、或约10%至50%、或约15%至55%、或约25%至75%、或其间的任何范围。

[0464] 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)对多个经固定的模板分子的通用样品索引序列进行测序;(b)对多个经固定的模板分子的插入

序列区进行测序;以及(c)将步骤(b)中获得的给定模板分子的插入序列与来自步骤(a)中获得的相同给定模板分子的通用样品索引序列进行分配,从而识别给定插入序列的样品来源。

[0465] 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来源的第二通用样品索引序列,并且第二样品索引缺乏短随机序列。

[0466] 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)对多个经固定的模板分子的3聚体随机序列进行测序以获得A、G、C和T/U的核碱基的平衡多样性,这些核碱基在第一测序循环、第二测序循环和第三测序循环中的每一者中检测并成像,以生成多个经固定的模板分子的位置图;(b)对多个经固定的模板分子的第一通用样品索引序列进行测序;(c)对多个经固定的模板分子的第二通用样品索引序列进行测序;(d)对多个经固定的模板分子的插入序列区进行测序;以及(e)将步骤(c)中获得的给定模板分子的插入序列与来自步骤(a)和(b)中获得的相同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分配,从而识别给定插入序列的样品来源。

[0467] 本公开提供了用于对核酸进行测序的方法,该方法包括:(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,其中各个经固定的模板分子包含插入序列区和一个样品索引,其中每个样品索引包含与识别插入序列的样品来源的通用样品索引序列连接的3聚体随机序列(例如,短随机序列NNN),其中通用样品索引序列包含3至20个核苷酸,其中不同的经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,并且其中经固定的模板分子具有不同的插入序列;(b)使用多种可检测地标记的核苷酸试剂对多个经固定的模板分子的3聚体随机序列和通用样品索引序列的第一个碱基位置进行四个循环的聚合酶介导的测序反应,该核苷酸试剂包含不同类型的核碱基A、G、C和T/U的混合物,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,其中测序的四个循环包括对从结合至经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,从而确定多个经固定的模板分子的各个模板分子中的3聚体随机序列的序列和通用样品索引序列的第一个碱基位置,并且其中在多个经固定的模板分子中的第一测序循环、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并成像A、G、C和T/U的核碱基的平衡多样性;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经固定的模板分子的位置图,其中插入区的序列不用于生成图。

[0468] 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%至85%、或约5%至60%、或约10%至50%、或约15%至55%、或约25%至75%、或其间的任何范围。

[0469] 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)对多个经固定的模板分子的通用样品索引序列的剩余碱基位置进行测序;(b)对多个经固定的模板分子的插入序列区进行测序;以及(c)将步骤(b)中获得的给定模板分子的插入序列与来自步骤(a)中获得的相同给定模板分子的通用样品索引序列进行分配,从而识别给定插入序列的样品来源。

[0470] 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来源的第二通用样品索引序列,并且第二样品索引缺乏随机序列。

[0471] 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)对多个经固定的模板分子的3聚体随机序列和通用样品索引序列的第一个碱基位置进行测序,以获得A、G、C和T/U的核碱基的平衡多样性,这些核碱基在第一测序循环、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并成像,以生成多个经固定的模板分子的位置图;(b)对多个经固定的模板分子的第一通用样品索引序列的剩余碱基位置进行测序;(c)对多个经固定的模板分子的第二通用样品索引序列进行测序;(d)对多个经固定的模板分子的插入序列区进行测序;以及(e)将步骤(c)中获得的给定模板分子的插入序列与来自步骤(a)和(b)中获得的相同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分配,从而识别给定插入序列的样品来源。

[0472] 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,支持物用具有不超过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各项组成的组的分子: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化亲水性聚合物分子。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。

[0473] 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,定位在支持物上的随机或预定位置处的固定的核酸模板分子的密度为每$mm^2$$10^4$至$10^8$,或其间的任何范围。在一些实施例中,插入序列的样品来源是基因组DNA、双链cDNA或细胞游离循环DNA。

[0474] 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至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至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。

[0475] 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至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至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。

[0476] 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,该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,从而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,从而确定多联体模板分子的第二部分的序列。

[0477] 对3聚体随机序列进行多重测序以生成聚合酶克隆图

[0478] 在一些方面,本发明提供了用于对核酸进行多重测序的方法,该方法包括:(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)衍生自第一样品来源的插入序列区,(ii)具有与第一通用样品索引序列接合的3聚体随机序列(例如,短随机序列NNN)的第一样品索引,以及(iii)具有缺乏随机序列的第二通用样品索引序列的第二样品索引,

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来源,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)衍生自第二样品来源的插入序列区,(ii)具有与第三通用样品索引序列接合的3聚体随机序列的第三样品索引,以及(iii)具有缺乏随机序列的第四通用样品索引序列的第四样品索引,其中第三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来源,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)将汇集的文库分子分布到支持物上并进行扩增反应以生成固定到支持物上(例如,固定在随机或预定位置处)的多个克隆扩增模板分子;(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,其中测序的三个循环包括对从与经固定的扩增模板分子结合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,从而确定多个经固定的模板分子中各个模板分子中的3聚体随机序列的序列,并且其中在多个经固定的扩增模板分子中的第一测序循环、第二测序循环和第三测序循环的每一者中检测和成像A、G、C和T/U的核碱基的平衡多样性;以及(f)使用在步骤(e)中获得的图像生成多个经固定的模板分子的位置图,其中插入区的序列不用于生成该图。

[0479] 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%至85%、或约5%至60%、或约10%至50%、或约15%至55%、或约25%至75%、或其间的任何范围。

[0480] 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)对多个经固定的模板分子的第一通用样品索引序列进行测序;(b)对多个经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经固定的模板分子的插入序列区进行测序;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相同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分配,从而识别给定插入序列的第一样品来源。

[0481] 在一些实施例中,在对核酸进行多重测序的方法中,该方法进一步包括:(a)对多个经固定的模板分子的第三通用样品索引序列进行测序;(b)对多个经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经固定的模板分子的插入序列区进行测序;以及(d)将步骤(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和第四通用样品索引序列分配,从而识别给定插入序列的第二样品来源。

[0482] 在一些方面,本发明提供了用于对核酸进行多重测序的方法,该方法包括:(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)衍生自第一样品来源的插入序列区,(ii)具有与第一通用样品索引序列接合的3聚体随机序列(例如,短随机序列NNN)的第一样品索引,以及(iii)具有缺乏随机序列的第二通用样品索引序列的第二样品索引,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来源,其中第一通用样品索引序列包含3至20个核苷酸(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或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)衍生自第二样品来源的插入序列区,(ii)具有与第三通用样品索引序列接合的3聚体随机序列的第三样品索引,以及(iii)具有缺乏随机序列的第四通用样品索引序列的第四样品索引,其中第三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来源,其中第三样品索引序列通用样品索引序列包含3至20个核苷酸(例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或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)将汇集的文库分子分布到支持物上并进行扩增反应以生成固定到支持物上(例如,固定在随机或预定位置处)的多个克隆扩增模板分子;(e)对第一样品索引和第三样品索引的3聚体随机序列进行四个循环的聚合酶介导的测序反应,并使用多种可检测地标记的核苷酸试剂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,该核苷酸试剂包括不同类型的核碱基A、G、C和T/U的混合物,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,其中测序的三个循环包括对从与经固定的扩增模板分子结合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,从而确定多个经固定的模板分子的各个模板分子中的3聚体随机序列的序列,并且其中在多个经固定的扩增模板分子中的第一测序循环、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A、G、C和T/U的核碱基的平衡多样性;以及(f)使用在步骤(e)中获得的图像生成多个经固定的模板分子的位置图,其中插入区的序列不用于生成该图。

[0483] 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%至85%、或约5%至60%、或约10%至50%、或约15%至55%、或约25%至75%、或其间的任何范围。

[0484] 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)对多个经固定的模板分子的第一通用样品索引序列的剩余碱基位置进行测序;(b)对多个经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经固定的模板分子的插入序列区进行测序;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相同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分配,从而识别给定插入序列的第一样品来源。

[0485] 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)对多个经固定的模板分子的第三通用样品索引序列的剩余碱基位置进行测序;(b)对多个经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经固定的模板分子的插入序列区进行测序;以及(d)将步骤(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和第四通用样品索引序列分配,从而识别给定插入序列的第二样品来源。

[0486] 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,支持物用具有不超过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各项组成的组的分子: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化亲水性聚合物分子。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。

[0487] 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)的密度为每$mm^2$$10^4$至$10^8$个。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。

[0488] 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至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至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。

[0489] 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至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至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。

[0490] 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,该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,从而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,从而确定多联体模板分子的第二部分的序列。

[0491] 文库分子

[0492] 在一些实施例中,本公开提供了核酸文库分子(100)、文库-夹板复合物(500)和共价闭合环状文库分子(600)及其制备方法。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。

[0493] 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(100))的核酸分子群体。在一些实施例中,群体中的各个文库分子进一步包含或缺乏额外的通用衔接子序列、至少一种样品索引序列和/或唯一分子索引(UMI)。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。

[0494] 在一些实施例中,核酸文库分子的插入区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。在一些实施例中,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。

[0495] 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。

[0496] 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。插入区可从空气、水、土壤或食物中携载的生物体被分离。

[0497] 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。

[0498] 插入区可以使用重组核酸技术来制备,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扩增的任何组合。

[0499] 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。

[0500] 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。

[0501] 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例如但不限于机械剪切,诸如流体剪切、恒定剪切和脉动剪切。机械片段化方法还可以包括机械应力,包括超声处理、雾化和声空化。在一些实施例中,聚焦声能可用于随机片段化核酸分子。市售装置(例如,Covaris™)可用于利用聚焦声能对核酸分子进行片段化。

[0502] 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。例如但不限于,可以进行限制性核酸内切酶酶消化至完成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可以包括选自由I型、II型、IIs型、IIB型、III型或IV型限制性酶组成的组的任一者或者两者或更多者的任意组合。酶促片段化可以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。酶促片段化可以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。

[0503] 在一些实施例中,酶促片段化可以通过使核酸与酶混合物反应来实现,该酶混合物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™)。

[0504] 可以使用与输入DNA样品中的靶区杂交的序列特异性引物用PCR生成插入区的片段,以生成具有已知片段长度和序列的插入区。

[0505] 使用CRISPR/Cas9的靶向片段化方法可用于从基因组DNA或其他输入DNA来源生成片段化插入区。

[0506] 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。

[0507] 在一些实施例中,可以使用基于Tn5转座酶的工作流程来片段化输入DNA,该工作流程采用包含双链DNA的转座子末端序列,该双链DNA具有可以结合转座酶以形成DNA-转座酶复合物(例如,转座体)的序列,其中复合物可以在体外标签化反应中将转座子末端序列

转置/插入到DNA中。转座子末端序列包含第一DNA链和第二DNA链。在一些实施例中,第一DNA链包含19个碱基转移末端序列5'AGATGTGTATAAGAGACAG 3'(SEQ ID NO:3)。在此类实施例中,第二DNA链包含19碱基非转移末端序列5'CTGTCTCTTATACACATCT 3'(SEQ ID NO:6),其可在其5'端被磷酸化。

[0508] 当需要使用基于Tn5转座酶的工作流程(例如,标签化工作流程)对输入DNA进行片段化并附加衔接子时,第一转座子DNA链进一步包含衔接子序列的第一链。在一些实施例中,第二转座子DNA链进一步包含衔接子序列的第二链。在一些实施例中,第一通用样品索引序列和第二衔接子序列沿其长度完全互补,从而形成线性双链转座子末端衔接子分子(图11)。在一些实施例中,如图11所示,转座子末端序列特异性结合转座酶。在一些实施例中,第一衔接子序列和第二衔接子序列沿其长度部分互补,从而形成Y形双链转座子末端衔接子分子(图12)。在Y形衔接子内,第一寡核苷酸可包含转座子末端序列和针对正向测序引物结合位点(120)的通用衔接子序列,并且第二寡核苷酸可包含转座子末端序列和针对反向测序引物结合位点(130)的通用衔接子序列。转座子末端序列可以特异性结合转座酶。在一些实施例中,Y形双链转座子末端衔接子分子可以是全长或短截Y形衔接子。

[0509] 在一些实施例中,多个Tn5转座酶和多个双链转座子末端衔接子分子可以在适合于将双链转座子末端衔接子分子结合/负载到转座酶上的条件下混合在一起以形成多个DNA-转座酶复合物(例如,转座体)(例如,图11和12)。

[0510] 在一些实施例中,输入DNA(例如,双链DNA)可以在适合于将转座子末端序列转座/插入到输入DNA中的随机位点的条件下与多个转座体混合,这将输入DNA片段化并将转移的末端序列共价附接至输入DNA的一条链的5'端,并且非转移的末端序列与转移的末端序列杂交,在互补的输入DNA链的3'端有一个间隙(例如,9个碱基间隙)。可以对间隙进行聚合酶催化的填充反应和酶促连接,以生成没有间隙并在两端携带衔接子的标签化双链DNA(图17)。转座子末端序列可以特异性结合转座酶。输入DNA和多个转座体的接触可以适合于将转座子末端序列和它们的通用衔接子序列插入输入DNA的随机位点,这生成输入DNA的双链片段,其具有共价接合至第一片段链的一端的针对正向测序引物结合位点(120)的通用衔接子序列和退火至所接合的针对正向测序引物结合位点(120)的通用衔接子序列的针对反向测序引物结合位点(130)的通用衔接子序列,以及针对正向测序引物结合位点(120)的通用衔接子序列与第二片段链之间的间隙(空心三角形)。转座子末端序列的插入还可以生成输入DNA的双链片段,其具有共价接合至第二片段链一端的针对反向测序引物结合位点(130)的通用衔接子序列和退火至所接合的针对反向测序引物结合位点(130)的通用衔接子序列的针对反向测序引物结合位点(130)的通用衔接子序列,以及针对反向测序引物结合位点(130)的通用衔接子序列与第一个片段链之间的间隙(空心三角形)。可以通过进行聚合酶催化的填充反应和酶促连接来闭合间隙,以生成包含第一DNA链和第二DNA链的双链DNA文库分子,每条DNA链包含在一侧上侧接针对正向测序引物结合位点(120)的通用衔接子序列的插入区(110)(例如,目的序列),并且插入区(110)在另一侧上侧接针对反向测序引物结合位点(130)的通用衔接子序列。

[0511] 基于转座酶的标签化工作流程可以使用美国专利号10,184,122、10,287,574、11,028,438和公开的美国申请号2019/0194737中描述的任何方法进行,所有这些文献的内容通过引用以其整体并入。

[0512] 在一些实施例中,插入区可以是单链或双链的。在一些实施例中,上述任何片段化方法可用于生成具有平端的一个或两个端、或具有5'突出端、或具有3'突出端、或其任何组合的DNA片段。

[0513] 在一些实施例中,片段化的核酸可以进行酶促反应以进行末端修复和/或A加尾。片段化的核酸的一端或两端可以进行酶促加尾反应,以通过采用末端转移酶反应生成非模板A尾。片段化的核酸可以在适合产生具有平端5'磷酸化末端的核酸片段的条件下与多种酶接触。在一些实施例中,多种酶生成在其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'端磷酸化酶可包含T4多核苷酸激酶。A加尾酶可以包括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。片段化的核酸的端部可以为相容的以连接至至少一个衔接子。

[0514] 插入区可以是任何长度,例如插入区的长度可以为约50至250、或约250至500、或约500至750、或约750至1000个碱基或碱基对,或其间的任何范围。

[0515] 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(AMPURXP™顺磁珠,目录号B23318)、MAGNAPURE™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Millipore™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)。

[0516] 将衔接子附加到片段化的核酸

[0517] 在一些方面,本公开提供了用于将一个或多个衔接子序列附加至片段化的核酸的试剂、试剂盒和方法。在一些实施例中,单独的片段化的核酸将与至少一个用于文库制备的通用衔接子序列共价接合。一般而言,核酸片段在两端共价接合至一个或多个通用衔接子以生成具有左衔接子-插入-右衔接子排列的线性文库分子。在一些实施例中,片段化核酸群体中的至少一个片段包含目的序列。文库分子群体中的单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插入区。在一些实施例中,约1至10ng、或约10至50ng、或约50至100ng、或其间的任何范围的输入片段化核酸可附加至一个或多个通用衔接子以生成线性文库。

[0518] 在一些实施例中,衔接子包含可以可操作地连接(附加)至靶多核苷酸的寡核苷酸,其中衔接子赋予共接合衔接子-靶分子功能。衔接子可以包括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。衔接子可以被配置为线性的(例如,图8和9)、茎环、发夹或Y形形式(例如,图10)。在一些实施例中,任何Y形衔接子(例如,如图10中所示)还包括在一个或两个错

配部分上的样品索引序列和任选的短随机序列(例如,NNN)。衔接子可以为任何长度,包括4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。突出端包括5'突出端和3'突出端。单链衔接子的5'端或双链衔接子的一个链可具有5'磷酸酯基团或缺乏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,加尾衔接子),或者衔接子可以为无尾的。衔接子的至少一部分可以包含已知的和预定的序列。衔接子可包括与扩增引物结合位点、正向测序引物结合位点、反向测序引物结合位点、表面捕获引物结合位点和/或表面钉扎引物结合位点的至少一部分相同或互补的序列。衔接子可包括至少一个样品索引序列,其可用于在多重测定中区分来自不同样品来源的多核苷酸(例如,插入序列)。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唯一分子索引(例如,UMI,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至12个或更多个具有已知序列的核苷酸。例如但不限于,唯一识别序列包含已知的随机序列,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,包括任何一种或者选自由I型、II型、III型、IV型、Hs型或IIB型组成的组的两者或更多者的任意组合。衔接子可包括至少一个硫代磷酸酯、硫醇代磷酸酯和/或磷酰胺键。

[0519] 在一些实施例中,衔接子包含通用序列,其是多核苷酸分子群体中的两个或更多个多核苷酸分子之间共有的核酸分子中的序列。例如但不限于,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但不限于通用扩增引物结合位点、通用正向测序引物结合位点、通用反向测序引物结合位点、通用表面捕获引物结合位点和通用表面钉扎引物结合位点。

[0520] 在一些实施例中,衔接子包含至少一个样品索引序列。样品索引可用于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线性文库。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(110)的序列连同一个或多个样品索引序列可用于识别输入核酸的来源。样品索引序列可用于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,样品索引包含3至10个核苷酸。在一些实施例中,样品索引包括短随机序列(例如,NNN),其可以位于样品索引序列的上游或下游。示例性样品索引列于上表3中。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,例如可以将2至10个、或10至50个、或50至100个、或100至200个、或其间的任何范围,或超过200个具有样品索引的文库汇集在一起。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,福尔马林固定的石蜡包埋的;FFPE)。技术人员将认识到核酸可以从许多其他来源分离。具有样品索引的线性文库分子可以以单链或双链形式制备。

[0521] 在一些实施例中,可以使用具有或不具有赋予某些性质(包括对酶消化的抗性或增加的热稳定性)的核苷酸类似物或经修饰的核苷酸键的天然核苷酸,使用化学合成程序来制备衔接子。抑制核酸酶消化的核苷酸类似物和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。

[0522] 在一些实施例中,衔接子包含具有杂交在一起的两条寡核苷酸链的双链线性衔接子(例如,图8和9)。在一些实施例中,双链线性衔接子的一端或两端可具有平端、5'突出端或3'突出端。在一些实施例中,双链线性衔接子的一端或两端可具有与核酸片段上的突出端互补的至少一个碱基突出(例如,A碱基突出)。在一些实施例中,双链线性衔接子的一端或两端包含磷酸化或非磷酸化的末端5'端。在一些实施例中,双链线性衔接子的一端或两端包含可通过聚合酶催化的引物延伸反应延伸的末端3'端,或末端3'端被阻断且不可延伸。在一些实施例中,双链线性衔接子的一条或两条链在5'端和/或3'端处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,双链线性衔接子的一条或两条链在内部位置处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,双链线性衔接子的一条或两条链在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0523] 在一些实施例中,双链线性衔接子包含至少一个衔接子序列。例如,双链线性衔接子包含针对正向测序引物结合位点(120)的通用序列的全长或部分长度序列(例如,图8)。在一些实施例中,任何双链核酸衔接子(例如,如图8中所示)还包括样品索引序列和任选的短随机序列(例如,NNN)。在一些实施例中,双链线性衔接子包含针对反向测序引物结合位点(130)的通用序列的全长或部分长度序列(例如,图9)。在一些实施例中,任何双链核酸衔接子(例如,如图9中所示)还包括样品索引序列和任选的短随机序列(例如,NNN)。在一些实施例中,双链线性衔接子进一步包含样品索引序列。在一些实施例中,双链线性衔接子进一步包含唯一分子索引(UMI)序列。在一些实施例中,双链线性衔接子缺乏待附加至核酸片段的额外序列。在一些实施例中,双链线性衔接子缺乏针对通用扩增引物结合位点、通用表面捕获引物结合位点和/或通用表面钉扎引物结合位点的序列。在一些实施例中,双链线性衔接子缺乏针对样品索引的序列。在一些实施例中,双链线性衔接子缺乏针对唯一分子索引(UMI)序列的序列。

[0524] 在一些实施例中,本文所述的任何双链线性衔接子包含上表1中列出的或与上表1中列出的序列互补的通用正向测序引物结合位点(120)或通用反向测序引物结合位点(130)。在一些实施例中,表1中列出的通用正向测序引物结合位点(120)或通用反向测序引物结合位点(130)中的任一者可在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0525] 在一些实施例中,衔接子包含具有两条寡核苷酸链的Y形衔接子,这两条寡核苷酸链杂交在一起以形成双链退火区和错配区(例如,图10)。在一些实施例中,Y形衔接子的退火部分可具有平端、5'突出端或3'突出端。在一些实施例中,Y形衔接子的退火部分可具有与核酸片段上的突出端互补的至少一个碱基突出(例如,A碱基突出)。在一些实施例中,Y形衔接子的退火部分包含磷酸化或非磷酸化的末端5'端。在一些实施例中,Y形衔接子的退火部分包含可通过聚合酶催化的引物延伸反应延伸的末端3'端,或末端3'端被阻断且不可延伸。在一些实施例中,退火部分和/或错配部分在5'端和/或3'端处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,退火部分和/或错配部分在内部位置处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,退火部分和/或错配部分在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0526] 在一些实施例中,Y形衔接子的退火部分和/或错配部分包含至少一个衔接子序

列。例如,Y形衔接子的第一寡核苷酸链包含针对正向测序引物结合位点(120)的通用序列的全长或部分长度序列。在一些实施例中,Y形衔接子的第二寡核苷酸链包含针对反向测序引物结合位点(130)的通用序列的全长或部分长度序列。图10示出了示例性Y形衔接子。

[0527] 在一些实施例中,多个Y形衔接子包括至少第一Y形衔接子和第二Y形衔接子。在一些实施例中,多个Y形衔接子中的Y形衔接子是相同的并且具有相同的第一寡核苷酸链和第二寡核苷酸链。在一些实施例中,多个Y形衔接子中的Y形衔接子是不同的并且具有不同的第一寡核苷酸链和第二寡核苷酸链。

[0528] 在一些实施例中,Y形衔接子是“短截”Y形衔接子,其中Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏待附加至核酸片段的额外序列,或短截Y形衔接子具有部分长度的通用衔接子序列。在一些实施例中,全长Y形衔接子或短截Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链进一步包含样品索引序列。在一些实施例中,样品索引序列位于第一寡核苷酸和/或第二寡核苷酸的错配区上。在一些实施例中,全长Y形衔接子或短截Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链进一步包含唯一分子索引(UMI)序列。在一些实施例中,唯一分子索引(UMI)序列位于第一寡核苷酸和/或第二寡核苷酸的错配区上。在一些实施例中,全长Y形衔接子或短截Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏针对通用扩增引物结合位点、通用表面捕获引物结合位点和/或通用表面钉扎引物结合位点的序列。在一些实施例中,全长Y形衔接子或短截Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏针对样品索引的序列。在一些实施例中,全长Y形衔接子或短截Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏针对唯一分子索引(UMI)序列的序列。使用较短的短截Y形衔接子的优点在于,与较长的Y形衔接子的连接效率相比,短截Y形衔接子的连接效率更高。

[0529] 在一些实施例中,Y形衔接子可连接至核酸片段的端部以生成具有附加至目的序列两端的衔接子序列的文库分子。在一些实施例中,可以通过使用加尾引物(例如,加尾PCR引物)进行引物延伸反应来附加额外的衔接子序列。例如,引物延伸反应包括将现在连接的Y形衔接子的错配部分与携带附加衔接子序列的加尾引物接触/杂交,并进行至少一个引物延伸反应(例如,图15)。在一些实施例中,第一Y形衔接子和第二Y形衔接子携带针对正向测序引物结合位点(120)的相同通用衔接子序列,并且携带针对反向测序引物结合位点(130)的相同通用衔接子序列。可以使用具有与错配部分中的一者(例如,(130))杂交的区的引物对所得双链线性核酸文库分子进行引物延伸,并且引物还可以在另一端处携带样品索引序列和通用表面引物结合位点(SurfPBS)。可以使用杂交的引物作为模板进行引物延伸反应,以生成延伸的文库分子,其包含:正向测序引物结合位点(120);插入序列(110);反向测序引物结合位点(130);样品索引序列;以及通用表面引物结合位点(SurfPBS)。在一些实施例中,延伸的文库分子可以使用具有与错配部分中的一者(例如,(120))杂交的区的第二引物进行另一引物延伸反应,并且引物还可以在另一端处携带通用表面引物结合位点(SurfPBS)和任选的样品索引序列。使用第二杂交引物作为模板进行引物延伸反应。

[0530] 如图16所示,在一些实施例中,第一Y形衔接子和第二Y形衔接子携带针对正向测序引物结合位点(120)的相同通用衔接子序列,并且携带针对反向测序引物结合位点(130)的相同通用衔接子序列。在一些实施例中,Y形衔接子还包括在一个或两个错配部分上的样品索引序列和任选的短随机序列(例如,NNN)。所得双链线性核酸文库分子可以与具有3’突

出端和平端的线性双链衔接子杂交。3'突出端包含可与具有反向测序引物结合位点(130)的错配部分杂交的序列,其生成具有缺口(实心三角形)的部分双链区。可以连接缺口,并且可以去除未连接的链。

[0531] 加尾引物可以携带额外衔接子序列中的任一者或任何组合,包括通用扩增引物结合位点、通用表面捕获引物结合位点、通用表面钉扎引物结合位点、样品索引序列和/或唯一分子索引(UMI)。

[0532] 在一些实施例中,第一加尾PCR引物可用于进行第一引物延伸反应并且第二加尾PCR引物可用于进行第二引物延伸以生成包含在两侧附加有至少一个额外衔接子的插入区的文库分子。在一些实施例中,第一加尾PCR引物和第二加尾PCR引物可用于进行多个PCR循环(例如,约5至20个PCR循环)以生成包含在两侧上附加有至少一个额外衔接子的插入区的文库分子。在一些实施例中,工作流程不包括使用第一加尾PCR引物或第二加尾PCR引物的PCR。

[0533] 在一些实施例中,作为Y形衔接子的一部分的本文所述的第一寡核苷酸链或第二寡核苷酸链中的任一者包含通用正向测序引物结合位点(120)(或其互补序列)或通用反向测序引物结合位点(130)(或其互补序列),其列于上表1中。在一些实施例中,在Y形衔接子的第一寡核苷酸链或第二寡核苷酸链中,表1中列出的通用正向测序引物结合位点(120)或通用反向测序引物结合位点(130)中的任一者可在5'端和/或3'端处被截短,其中截短可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0534] 在一些实施例中,衔接子包含具有一条寡核苷酸链的发夹衔接子,该寡核苷酸链具有双链的分子内退火区和单链非互补环部分。在一些实施例中,发夹衔接子的双链部分可具有平端、5'突出端或3'突出端。在一些实施例中,发夹衔接子的双链部分可具有与核酸片段上的突出端互补的至少一个碱基突出(例如,A碱基突出)。在一些实施例中,发夹衔接子的双链部分包含磷酸化或非磷酸化的末端5'端。在一些实施例中,发夹衔接子的双链部分包含可通过聚合酶催化的引物延伸反应延伸的末端3'端,或末端3'端被阻断且不可延伸。在一些实施例中,环部分包含至少一个可切割位点,包括核酸内切酶限制性酶位点、脱碱基位点和/或核糖核苷酸。在一些实施例中,脱碱基位点包含一个或多个尿苷、脱氧-8-氧代-鸟嘌呤三磷酸(d-8-oxoG);8-氧代-7,8-二氢鸟嘌呤(8oxoG)或脱氧肌苷。在一些实施例中,双链部分和/或环部分在5'端和/或3'端处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,双链部分和/或环部分在内部位置处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,双链部分和/或环部分在5'端和/或3'端处或在内部位置处包含一个或多个2'-O-甲基胞嘧啶碱基。

[0535] 在一些实施例中,发夹衔接子的双链部分和/或环部分包含至少一个衔接子序列。例如,发夹衔接子的双链部分的一条链包含针对正向测序引物结合位点(120)的通用序列的全长或部分长度序列。在一些实施例中,发夹衔接子的双链部分的另一条链包含针对反向测序引物结合位点(130)的通用序列的全长或部分长度序列。在一些实施例中,发夹衔接子的环部分的一个区包含针对正向测序引物结合位点(120)的通用序列的全长或部分长度序列。在一些实施例中,发夹衔接子的环部分的另一个区包含针对反向测序引物结合位点(130)的通用序列的全长或部分长度序列。在一些实施例中,发夹衔接子进一步包含样品索引序列。在一些实施例中,发夹衔接子进一步包含唯一分子索引(UMI)序列。在一些实施例

中,发夹衔接子缺乏待附加至核酸片段的额外序列。在一些实施例中,发夹衔接子缺乏针对通用扩增引物结合位点、通用表面捕获引物结合位点和/或通用表面钉扎引物结合位点的序列。在一些实施例中,发夹衔接子缺乏针对样品索引的序列。在一些实施例中,发夹衔接子缺乏针对唯一分子索引(UMI)序列的序列。

[0536] 在一些实施例中,作为发夹衔接子的一部分的本文所述的任何寡核苷酸链包含上表1中列出的通用正向测序引物结合位点(120)和通用反向测序引物结合位点(130)。在一些实施例中,作为发夹衔接子的一部分的本文所述的任何寡核苷酸链包括上表1中列出的通用正向测序引物结合位点(120)的互补序列和/或通用反向测序引物结合位点(130)的互补序列。在一些实施例中,在作为发夹的一部分的链的第一寡核苷酸链或第二寡核苷酸链中,表1中列出的通用正向测序引物结合位点(120)或通用反向测序引物结合位点(130)中的任一者可以在5'端和/或3'端处被截断,其中截断可以为1至12个核苷酸。在某些实施例中,5'端被截短。在某些实施例中,3'端被截短。

[0537] 各个核酸片段可以在一端或两端附加到至少一个通用衔接子序列,以形成具有一般排列左衔接子-插入片段-右衔接子的重组核酸线性文库分子。

[0538] 在一些实施例中,可以使用连接酶和/或引物延伸反应将核酸片段在一端或两端处与至少一个通用衔接子序列接合以生成线性文库分子。核酸片段与通用衔接子之间的共价连接可通过DNA连接酶或RNA连接酶实现。可以连接双链DNA分子的示例性非限制性DNA连接酶包括T4 DNA连接酶和T7 DNA连接酶。在一些实施例中,通用衔接子序列可使用具有携带通用衔接子序列的5'区和与核酸片段的一部分互补的3'区的加尾引物通过引物延伸或PCR附加至核酸片段。在一些实施例中,可以通过引物延伸或PCR,使用具有携带额外通用衔接子序列的5'区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'区的加尾引物,将通用衔接子序列附加至在一侧上或两侧上侧接第一通用衔接子序列和第二通用衔接子序列的插入序列。

[0539] 在一些实施例中,可以通过采用连接反应,并且然后使用双链夹板衔接子(200)生成文库-夹板复合物(500)来生成文库分子(100)。在一些实施例中,用于生成文库分子的方法包括:(a)提供包含目的序列的双链核酸片段(例如,双链插入区(110));以及(b)将双链插入区(110)的第一端接合至具有针对正向测序引物结合位点(120)的通用序列的第一双链线性衔接子,并将双链插入区(110)的第二端接合至具有针对反向测序引物结合位点(130)的通用序列的第二双链线性衔接子,其中使用DNA连接酶进行接合以生成双链重组文库分子(100)(例如,图13)。双链核酸片段可以在另一侧上酶促连接至双链衔接子,该双链衔接子携带针对反向测序引物结合位点(130)的通用衔接子序列。在一些实施例中,双链衔接子中的一者或两者还可以包含样品索引序列和任选的短随机序列(例如,NNN)。

[0540] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(c):使双链重组文库分子(100)变性以生成两个单链文库分子(100)。两个单链文库分子(100)可与本文所述的任何双链夹板衔接子(200)杂交以生成多个文库-夹板复合物(500),如下文步骤(d)中所述。

[0541] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(d):使单链文库分子(100)与本文所述的多个任何双链夹板衔接子(200)杂交。在一些实施例中,各个双链夹板衔接子(200)包含与第二夹板链(400)杂交的第一夹板链(300),其中双链夹板衔接子包含

双链区和两个侧接单链区,其中第一夹板链包含第一区(320)、内部区(310)和第二区(330),并且其中第一夹板链的内部区(310)与第二夹板链(400)杂交。在一些实施例中,第一夹板链的第一区(320)被设计成与文库分子(100)的针对正向测序引物结合位点(120)的通用序列杂交。在一些实施例中,第一夹板链的第二区(330)被设计成与文库分子(100)的针对反向测序引物结合位点(130)的通用序列杂交。在一些实施例中,步骤(c)的杂交在适合于使第一夹板链的第一区(320)与单链文库分子的至少一部分(120)杂交的条件下进行,并且该条件适合于使第一夹板链的第二区(330)与单链文库分子的至少一部分(130)杂交,从而使单链文库分子成环以形成各自具有两个缺口的多个文库-夹板复合物(500)(例如,图3)。

[0542] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(e):使多个文库-夹板复合物(500)与连接酶接触以生成多个共价闭合环状文库分子(600)(例如,图4)。

[0543] 在一些实施例中,在步骤(b)之后不进行引物延伸或PCR反应以将任何另外的衔接子序列接合至双链重组文库分子(100),另外的衔接子序列包括:通用扩增引物结合位点、通用表面捕获引物结合位点、通用表面钉扎引物结合位点、样品索引或唯一分子索引(UMI)序列。

[0544] 在一些实施例中,可以通过采用连接反应,并且然后使用双链夹板衔接子(200)生成文库-夹板复合物(500)来生成文库分子(100)。在一些实施例中,用于生成文库分子的方法包括:(a)提供包含目的序列的双链核酸片段(例如,双链插入区(110))。

[0545] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(b):将双链插入区(110)的第一端接合至第一Y形衔接子,并将双链插入区(110)的第二端接合至第二Y形衔接子,其中使用DNA连接酶进行接合以生成双链重组文库分子(100)。在一些实施例中,第一Y形衔接子和第二Y形衔接子是相同的,并且各自包含杂交在一起以形成双链退火区和错配区的两条寡核苷酸链。在一些实施例中,Y形衔接子的第一寡核苷酸链包含针对正向测序引物结合位点(120)的通用序列的全长或部分长度序列(例如,图14)。在一些实施例中,Y形衔接子的第二寡核苷酸链包含针对反向测序引物结合位点(130)的通用序列的全长或部分长度序列(例如,图14)。双链核酸片段可以在另一侧上酶促连接至第二Y形衔接子,第二Y形衔接子包含携带针对正向测序引物结合位点(120)的通用衔接子序列的全长序列的第一寡核苷酸,以及携带针对反向测序引物结合位点(130)的通用衔接子序列的全长序列的第二寡核苷酸。在一些实施例中,第一Y形衔接子和第二Y形衔接子携带针对正向测序引物结合位点(120)的相同通用衔接子序列,并且携带针对反向测序引物结合位点(130)的相同通用衔接子序列。在一些实施例中,Y形衔接子还包括在一个或两个错配部分上的样品索引序列和任选的短随机序列(例如,NNN)。在一些实施例中,Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏针对通用扩增引物结合位点、通用表面捕获引物结合位点和/或通用表面钉扎引物结合位点的序列。在一些实施例中,Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏针对样品索引的序列。在一些实施例中,Y形衔接子的第一寡核苷酸链和/或第二寡核苷酸链缺乏针对唯一分子索引(UMI)序列的序列。

[0546] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(c):使双链重组文库分子(100)变性以生成两个单链文库分子(100)。两个单链文库分子(100)可与本文所述的任何双链夹板衔接子(200)杂交以生成多个文库-夹板复合物(500),如下文步骤(d)中所

述。

[0547] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(d):使单链文库分子(100)与本文所述的多个任何双链夹板衔接子(200)杂交。在一些实施例中,各个双链夹板衔接子(200)包含与第二夹板链(400)杂交的第一夹板链(300),其中双链夹板衔接子包含双链区和两个侧接单链区,其中第一夹板链包含第一区(320)、内部区(310)和第二区(330),并且其中第一夹板链的内部区(310)与第二夹板链(400)杂交。在一些实施例中,第一夹板链的第一区(320)被设计成与文库分子(100)的针对正向测序引物结合位点(120)的通用序列杂交。在一些实施例中,第一夹板链的第二区(330)被设计成与文库分子(100)的针对反向测序引物结合位点(130)的通用序列杂交。在一些实施例中,步骤(c)的杂交在适合于使第一夹板链的第一区(320)与单链文库分子的至少一部分(120)杂交的条件下进行,并且该条件适合于使第一夹板链的第二区(330)与单链文库分子的至少一部分(130)杂交,从而使单链文库分子成环以形成各自具有两个缺口的多个文库-夹板复合物(500)(例如,图3)。

[0548] 在一些实施例中,用于生成文库分子的方法进一步包括步骤(e):使多个文库-夹板复合物(500)与连接酶接触以生成多个共价闭合环状文库分子(600)(例如,图4)。

[0549] 在一些实施例中,在步骤(b)之后不进行引物延伸或PCR反应以将任何另外的衔接子序列接合至双链重组文库分子(100),另外的衔接子序列包括:通用扩增引物结合位点、通用表面捕获引物结合位点、通用表面钉扎引物结合位点、样品索引或唯一分子索引(UMI)序列。

[0550] 具有低非特异性结合涂层的支持物

[0551] 在一些方面,本公开提供了用于测序的组合物和方法,其采用具有固定在其上的多个表面引物的支持物。在一些实施例中,用低非特异性结合涂层使支持物钝化。在一些实施例中,本文所描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)的非常低的非特异性结合。与常规表面涂层相比,该表面涂层可以表现出低背景荧光学信号或高对比度噪声(CNR)比。

[0552] 在一些实施例中,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,在不超过2、3、4、5、6、7、8、9、10、15、20、25、30或超过30个扩增循环中实现适于检测的扩增水平。

[0553] 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)。

[0554] 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种表面制备技术可以用于清洁或处理表面。例如,可以使用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。

[0555] 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具反应性的官能团(例如,胺或羧基基团),然后可以将其用于偶联接头分子(例如,各种长度的直链烃分子,诸如C6、C12、C18烃或线性聚乙二醇(PEG)分子)或层分子(例如,支化PEG分子或其他聚合物)至表面。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。

[0556] 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。

[0557] 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。可替代地或组合地,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,化学修饰区域的有序阵列或随机图案可包含至少1、5、10、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000、2000、3000、4000、5000、6000、7000、8000、9000、或10,000或更多个离散区域。

[0558] 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

表面的端基团可以包含但不限于: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。例如但不限于,经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引物5'端的poly-T和poly-A链(例如,0至20个碱基,例如,0、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个)、PEG接头(例如,3至20个单体单元,例如,约3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或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,并且随后将荧光读数与已知浓度的染料溶液的荧光读数进行比较。

[0559] 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。如本文所描述的,可以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",使得可以依序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。

[0560] 如上所述,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的——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。

[0561] 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)的特异性结合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)的特异性荧光与非特异性荧光的比率。

[0562] 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1μM Cy3标记的链霉亲和素(GEAmersham)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,1μM标记的Cy3 SA(ThermoFisher)、1μM Cy5 SA染料(ThermoFisher)、10μM氨基烯丙基-dUTP-ATTO-647N(Jena Biosciences)、10μM氨基烯丙基-dUTP-ATTO-Rho11(Jena Biosciences)、10μM氨基烯丙基-dUTP-ATTO-Rho11(JenaBiosciences)、10μM 7-丙炔氨基-7-脱氮-dGTP-Cy5(JenaBiosciences和10μM 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℃孵育15分钟。每个孔用50ul去离子不含RNA酶/DNA酶的水冲洗2至3次,并用pH 7.4的25mMACES缓冲液冲洗2至3次。使用如由制造商指定的Cy3、AF555或Cy5滤波器集合(根据所执行的染料测试)以800的PMT增益设定和50至100μm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,NewYork),例如405、488、532或633nm二向色反射器/分束器,并且选择带通滤波器作为与适当激发波长一致的532LP或645LP。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。

[0563] 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。

[0564] 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。

[0565] 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以表现出具有在此范围内的任何位置的值的水接触角。

[0566] 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、50、40、30、20、15、10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。

[0567] 本公开的低结合表面可以对长时间暴露于溶剂和升高的温度,或对溶剂暴露或温度变化的重复循环表现出稳定性或耐久性的显著改善。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,用于评定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,用于评定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)。

[0568] 在一些实施例中,本文公开的表面可以表现出特异性信号与非特异性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(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。

[0569] 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、20、210、220、230、240、250或大于250。

[0570] 一种或更多种类型的引物(例如,捕获寡核苷酸和/或环化寡核苷酸)可以附接或栓系至支持物表面。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或超过10个不同的引物或衔接子序列可拴系至表面的至少一层。

[0571] 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最多100、最多90、最多80、最多70、最多60、最多50、最多40、最多30、最多20、或最多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。

[0572] 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最多10,000、最多100,000、最多1,000,000或最多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。

[0573] 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。

[0574] 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。合适的支化聚合物的实例包括但不限于:支化PEG、支化聚(乙烯醇)(支化PVA)、支化聚(乙烯基吡啶)、支化聚(乙烯基吡咯烷酮)(支化PVP)、支化)、聚(丙烯酸)(支化PAA)、支化聚丙烯酰胺、支化聚(N-异丙基丙烯酰胺)(支化PNIPAM)、支化聚(甲基丙烯酸甲酯)(支化PMA)、支化聚(2-羟乙基甲基丙烯酸酯)(支化PHEMA)、支化聚(低聚(乙二醇)甲基醚甲基丙烯酸酯)(支化POEGMA)、支化聚谷氨酸(支化PGA)、支化聚赖氨酸、支化聚葡萄糖苷和葡聚糖。

[0575] 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。

[0576] 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、支化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。

[0577] 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。

[0578] 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。

[0579] 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。

[0580] 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。

[0581] 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等的定制仪器。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,Zyla4.2),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。

[0582] 在一些情况下,使用所公开的反应制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通常被定义为:CNR=(信号-背景)/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)周围的间质区所测量的信号。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。

[0583] 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($B_{间质}$)之外,扩增DNA集落占据的区域内存在“质内”背景($B_{质内}$)。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。$B_{间质}$背景信号来自多种来源;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)中的此背景信号随着时间的推移被平均和减去。单个DNA菌落(即,FOV中的$(S)-B_{间质}$)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($B_{质内}$)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。

[0584] 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$B_{间质}$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,这会影响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制剂组合使用,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制剂用于其他检测模式,包括光学和非光学检测模式。

[0585] 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。

[0586] 在一些实施例中,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)的聚合酶群落时,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)。

[0587] 在一些实施例中,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20×0.75NA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,Zyla4.2)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、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)。

[0588] 测序聚合酶

[0589] 在一些方面,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型聚合酶。

[0590] 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 altiarchaeales古菌;Candidatus HadarchaeumYellowstonense;Hadesarchaea古菌;广古菌(Euryarchaeota)古菌;热原体古菌;嗜热球菌聚合酶,诸如嗜热球菌、噬菌体T7 DNA聚合酶;人α、δ和εDNA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IIIα和ε;9度N聚合酶;逆转录酶,诸如HIV M型或O型逆转录酶;禽类髓母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、DEEPVENT、THERMINATOR、Pfu、KOD、Pfx、Tgo和RB69聚合酶。

[0591] 核苷酸

[0592] 在一些方面,本公开提供了用于使用核苷酸对核酸分子进行测序的方法,其中多个核苷酸中的至少一种核苷酸包含碱基、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酯基团(例如,1至10个磷酸酯基团)。多个核苷酸可以包括选自由以下各项组成的组的至少一种类型的核苷酸:dATP、dGTP、dCTP、dTTP和dUTP。多个核苷酸可包括选自由dATP、dGTP、dCTP、dTTP和/或dUTP组成的组的两种或更多种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。

[0593] 在一些实施例中,在本文所述的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,其中该链通常通过酯键或磷酰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、亚甲基或亚乙基连接在一起。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,这些类似物包括磷酰胺、硫代磷酸酯、二硫代磷酸酯和O-甲基亚磷酰胺基团。

[0594] 在一些实施例中,在本文所述的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,链终止部分可以在引物延伸反应期间抑制聚合酶催化的后续核苷酸单元或游离核苷酸在新生链中的掺入。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'糖羟基位置去除/切割以生成具有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、二硫化物基团、碳酸基团、脲基团、缩醛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切割/去除,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)($Pd(PPh_3)_4$)用哌啶或用2,3-二氯-5,6-二氰基-1,4-苯醌(DDQ)切割。在一些实施例中,链终止部分芳基和苄基可用H2 Pd/C切割。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切割,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(K_2CO_3)、用吡啶中的三乙胺或用乙酸(AcOH)中的Zn切割。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酸盐裂解。

[0595] 在一些实施例中,在本文所述的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切

割/去除。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,切割剂包括4-二甲基氨基吡啶(4-DMAP)。

[0596] 在一些实施例中,在本文所述的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,该链终止部分选自由以下各项组成的组: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'-劳基甲氧羰基、3'叔丁氧羰基、3'-O-烷基羟氨基、3'-硫代磷酸酯和3-O-苄基,或其衍生物。

[0597] 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,该接头可从碱基切割/去除。在一些实施例中,多个核苷酸中的至少一者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告基因部分(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。

[0598] 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,核苷酸碱基上的可切割接头包含可切割部分,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、缩醛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、光或热使可切割部分发生反应,碱基上的可切割接头可从碱基切割/去除。在一些实施例中,可切割部分烷基、烯基、炔基和烯丙基可用四(三苯基膦)-钯(0)($Pd(PPh_3)_4$)用哌啶或用2,3-二氯-5,6-二氰基-1,4-苯醌(DDQ)切割。在一些实施例中,可切割部分芳基和苄基可用H2 Pd/C切割。在一些实施例中,可切割部分胺、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切割,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,可切割部分碳酸酯可用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切割。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酸盐裂解。

[0599] 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基团。在一些实施例中,可切割部分叠氮、叠氮基和叠氮基甲基基团可用膦化合物切割/去除。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,切割剂包括4-二甲基氨基吡啶(4-DMAP)。

[0600] 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切割接头具有相同或不同的可切割部分。在一些实施例中,连接至碱基的链终止部分(例如,在糖2'和/或糖3'位置处)和可检测报告基因部分可用相同化学剂化学切割/去除。在一些实施例中,连接至碱基的链终止部分(例如,在糖2'和/或糖3'位置处)和可检测报告基因部分可用不同化学剂化学切割/去

除。

[0601] 多价分子

[0602] 在一些方面,本公开提供了用于采用多价分子对核酸分子进行测序的方法。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22)。在一些实施例中,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。示例性核苷酸臂在图26中示出。示例性的多价分子示出于图22至25。示例性间隔基在图27(顶部)中示出并且示例性接头在图27(底部)和图28中示出。附接至接头的示例性核苷酸在图29至31中示出。示例性生物素化核苷酸臂在图32中示出。

[0603] 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由dATP、dGTP、dCTP、dTTP和dUTP组成的组的相同类型的核苷酸单元。

[0604] 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)。多个多价分子可以包括具有选自由以下各项组成的组的一种类型的核苷酸单元的一种类型的多价分子: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由以下各项组成的组的核苷酸单元:dATP、dGTP、dCTP、dTTP和/或dUTP。

[0605] 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,其中该链通常通过酯键或磷酰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、亚甲基或亚乙基连接在一起。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,这些类似物包括磷酰胺、硫代磷酸酯、二硫代磷酸酯和O-甲基亚磷酰胺基团。

[0606] 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,链终止部分可以在引物延伸反应期间抑制聚合酶催化的后续核苷酸单元或游离核苷酸在新生链中的掺入。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'糖羟基位置去除/切割以生成具有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、二硫化物基团、碳酸基团、脲基团、缩醛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

单元切割/去除,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)($Pd(PPh_3)_4$)用哌啶或用2,3-二氯-5,6-二氰基-1,4-苯醌(DDQ)切割。在一些实施例中,链终止部分芳基和苄基可用H2 Pd/C切割。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切割,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(K_2CO_3)、用吡啶中的三乙胺或用乙酸(AcOH)中的Zn切割。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酸盐裂解。

[0607] 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切割/去除。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,切割剂包括4-二甲基氨基吡啶(4-DMAP)。

[0608] 在一些实施例中,核苷酸单元包含链终止部分,该链终止部分选自由以下各项组成的组: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'-芴基甲氧羰基、3'叔丁氧羰基、3'-O-烷基羟氨基、3'-硫代磷酸酯和3-O-苄基,或其衍生物。

[0609] 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报告基因部分标记。在一些实施例中,可检测报告基因部分包括荧光团。在一些实施例中,附接至多价分子的特定可检测报告基因部分(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。

[0610] 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,可检测报告基因部分附接至核苷酸碱基。在一些实施例中,可检测报告基因部分包括荧光团。在一些实施例中,附接至多价分子的特定可检测报告基因部分(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。

[0611] 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,核包括链霉亲和素型或亲和素型部分(包含亲和素蛋白),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、类似物和其它非天然形式。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,亲和素部分可以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™。

[0612] 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、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9或1秒。结合复合物具有大于约0.1至0.25秒、或约0.25至0.5秒、或约0.5至0.75秒、或约0.75至1秒、或约1至2秒、或约2至3秒、或约3至4秒、或约4至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)在经历引起聚合酶、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、附接至或杂交于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。

[0613] 实例

[0614] 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。

[0615] 实例1:将输入DNA片段化

[0616] 使用两种不同的方法将基因组DNA片段化。使用酶FRAGMENTASE ULTRA™(来自New England Biolabs™)对基因组DNA进行片段化。测试了一系列输入量的基因组DNA,包括1ng/μL至1000ng/μL。在最低范围内,仅测试了1ng的基因组DNA。片段化反应包括7μL的FRAGMENTASE™反应缓冲液、2μL的FRAGMENTASE™酶、26μL的基因组DNA。将片段化反应在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。

[0617] 还使用Covaris通过剪切将基因组DNA片段化,以实现约200、250、350和400bp的插入片段大小。用于制备线性文库分子的输入片段化核酸的量包括50ng、100ng、500ng和1ug。

[0618] 实例2:制备文库分子

[0619] 将片段化的双链DNA在两端连接至Y形衔接子。Y形衔接子各自包含第一寡核苷酸链和第二寡核苷酸链,该第一寡核苷酸链具有针对正向测序引物结合位点(120)的通用序列的全长或部分长度序列,该第二寡核苷酸链具有针对反向测序引物结合位点(130)的通用序列的全长或部分长度序列。针对正向测序引物结合位点(120)和针对反向测序引物结合位点(130)的通用序列的示例性通用序列列于表1中。将所得的双链文库分子(100)热变性,以生成多个单链文库分子(100)。示例性的单链文库分子(100)示出于图1和3中。

[0620] 实例3:制备双链夹板衔接子

[0621] 双链夹板衔接子包含与第二夹板链(400)杂交的第一夹板链(300)。示例性双链夹板衔接子示出于图2中。在室温下使用NaOH使双链夹板衔接子变性5分钟,并且然后在37℃下退火5分钟。可替代地,将双链夹板衔接子热变性,并且然后在37℃下退火5分钟。

[0622] 第一夹板链包含第一区(320)、内部区(310)和第二区(330)。第一夹板链的第一区(320)可以与单链文库分子(100)的针对正向测序引物结合位点(120)的通用序列杂交。第一夹板链的第二区(330)可以与单链文库分子(100)的针对反向测序引物结合位点(130)的通用序列杂交。第一夹板链的区(320)和(330)的示例性序列列于表2中。

[0623] 实例4:制备文库-夹板复合物

[0624] 在热循环仪装置中,在含有100mM乙酸钾和30mM HEPES(pH 7.5)的退火缓冲液中,将单链文库分子(100)退火为双链夹板衔接子(200)(例如,图3)。退火程序包括:95摄氏度5分钟、37摄氏度5分钟、并保持在37摄氏度。

[0625] 在文库-夹板复合物的第一实施例中,各个单链文库分子(100)包含插入区(110),该插入区包含人基因组序列、具有序列

[0626] 5'-CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:51)的针对正向测序引物结合位点(120)的通用序列,以及具有序列

[0627] 5'-ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:4)的针对反向测序引物结合位点(130)的通用序列。第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区。第一子区(411)包含具有序列3'-TGGGACTTTCATGCACG-5'(SEQ ID NO:24)的针对经固定的表面钉扎引物的通用结合序列。第二子区(412)包含短随机序列(NNN)和四个样品索引序列中的一者。第三子区(413)包含具有序列3'-TAATGTACCTAGTCCACTCCGA-5'(SEQ ID NO:19)的针对经固定的表面捕获引物的通用结合序列。表3中列出了短夹板链(400)中的其他示例性序列,包括表面捕获引物结合序列、表面钉扎引物结合序列和样品索引序列。

[0628] 在文库-夹板复合物的第二实施例中,各个单链文库分子(100)包含插入区(110),该插入区包含人基因组序列、具有序列

[0629] 5'-CGTGCTGGATTGGCTCACCAGACACCTTCCGACAT-3'(SEQ ID NO:1)的针对正向测序引物结合位点(120)的通用序列,以及具有序列

[0630] 5'-ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:4)的针对反向测序引物结合位点(130)的通用序列。第二夹板链(400)包含至少三个子区,包括第一子区、第二子区和第三子区。第一子区(411)包含具有序列3'-TGGGACTTTCATGCACGTAATGTAC-5'(SEQ ID NO:7)的针对经固定的表面钉扎引物的通用结合序列。第二子区(412)包含具有序列3'-CTAGTCCACTCCGACGCTGCTGA-5'(SEQ ID NO:18)的针对经固定的表面捕获引物的通用结合序列。第三子区(413)包含短随机序列(NNN)和四个样品索引序列中的一者。在该第二实施例中,在生成共价闭合环状分子(600)的连接反应之后,将子区(413)中的短随机序列(NNN)和样品索引序列接合至位于共价闭合环状文库分子(600)中的针对反向测序引物结合位点(130)的通用序列(例如,参见下面的实例5)。表3中列出了针对表面捕获引物结合序列、表面钉扎引物结合序列和样品索引序列的其他示例性序列。

[0631] 实例5:制备共价闭合环状文库分子

[0632] 通过向实例4的退火混合物中添加T7 DNA连接酶和T4多核苷酸激酶以及T4 DNA连接酶反应缓冲液,使退火混合物进行酶促连接和磷酸化反应。将酶混合物在热循环装置中孵育,加热盖设置为75摄氏度。热循环装置程序包括:37度10分钟,65度10分钟,并保持在4度。连接和磷酸化反应生成与第一夹板链(300)杂交的共价闭合环状文库分子(600)(例如,

图4)。

[0633] 通过向连接/磷酸化反应混合物添加T7核酸外切酶和不耐热核酸外切酶来进行酶促核酸外切酶消化。将核酸外切酶反应混合物在热循环装置中孵育,其程序设定为:37摄氏度10分钟,80摄氏度2分钟,并保持在4摄氏度。

[0634] 使用SPRI SELECT™珠(来自Beckman Coulter™)对核酸外切酶反应混合物进行多次循环净化。

[0635] 使用Qubit或qPCR对共价闭合环状文库分子(例如,单链分子)的清洁制备物的产率进行定量。

[0636] 实例6:滚环扩增和测序

[0637] 将来自实例5的共价闭合环状文库分子分布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,并进行支持物上滚环扩增以生成经固定的多联体。低非特异性结合涂层包括多个经固定的表面捕获引物和表面钉扎引物。表4列出了示例性的表面捕获和钉扎引物。

[0638] 实例7:使用多价分子和核苷酸进行测序

[0639] 在第一阶段中使用经荧光标记的多价分子和在第二阶段中使用未标记的核苷酸类似物(例如,3'链终止子阻断基团)对多联体进行递归两阶段测序反应。

[0640] 在其上固定有多个多联体模板分子(例如,经固定的聚合酶群落)的流通池上进行两阶段测序反应。

[0641] 通过使多个可溶性测序引物与固定到流通池的多联体模板分子杂交,以形成经固定的引物-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,接触经固定的引物-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,使经以荧光方式标记的多价分子的混合物(例如,以约20至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。将经以荧光方式标记的多价分子在其核心处标记。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(未示出)。通过用包含去污剂的缓冲液洗涤,去除第一测序聚合酶和多价分子,同时保留与经固定的多联体杂交的测序引物(保留的双链体)。

[0642] 第一阶段测序反应适合于在多联体模板分子(例如,聚合酶群落)上形成多个亲合力复合物。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,从而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结合至第一聚合酶;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,从而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,其中包含相同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。

[0643] 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至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。使掺入的未标记的核苷酸类似物与去除3'O-甲基叠氮基团并生成可延伸的3'OH基团的切割试剂反应。

[0644] 通过用包括去污剂的缓冲液进行洗涤来去除第二测序聚合酶,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进行重复测序反应,以产生延伸的正向测序引物链。

[0645] 对于由上述实例4所述的文库-夹板复合物的第一实施例生成的经固定的多联体,测序的顺序包括:(1)使用正向测序引物对插入区(110)进行测序,以生成多个正向插入延伸产物;(2)进行成对轮转反应;(3)使用反向测序引物对插入区(110)进行测序,以生成多个反向插入延伸产物;以及(4)使用具有序列5'-ACCCTGAAAGTACGTGC-3'的反向测序引物对短随机序列(NNN)和样品索引序列进行测序,以生成多个样品索引延伸产物。测序结果列于下表A中。

[0646] 表A:

[0647]

循环文库负载浓度	原始读数(百万)	过滤通过率	R1 Q30	R2 Q30	索引分配率
9 pM	883 M	96.2%	99.0%	97.7%	95%
9 pM	750 M	95.3%	94.8%	91.5%	94.4%

[0648] 对于由上述实例4中描述的文库-夹板复合物的第二实施例生成的经固定的多联体,测序的顺序包括:(1)使用具有序列

[0649] 5'-ATGTCGGAAGGTGTGCAGGCTACCGCTTGTCAACT-3'(SEQ ID NO:4)的正向测序引物对短随机序列(NNN)和样品索引序列进行测序,以生成多个样品索引延伸产物;(2)使用正向测序引物对插入区(110)进行测序,以生成多个正向插入延伸产物;(3)进行成对轮转反应;以及(4)使用反向测序引物对插入区(110)进行测序,以生成多个反向插入延伸产物。测序结果列于下表B中。

[0650] 表B:

[0651]

循环文库负载浓度	原始读数(百万)	通过过滤读数(百万)	过滤通过率	R1 Q30	R2 Q30
12 pM	1116 M	994.3	89.1%	94.4%	93.4%
12 pM	1126 M	991.6	88/0%	94.3%	92.9%

[0652] 通过引用结合

[0653] 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。

[0654] 等效形式

[0655] 本公开的一个或多个实施例的细节在上面的随附描述中阐述。尽管在本公开的实践或测试中可以使用与本文描述的那些类似或等同的任何方法和材料,但是现在描述优选的方法和材料。根据说明书和权利要求书,本公开的其他特征、目的和优点将变得显而易见。在说明书和所附权利要求中,除非上下文另有明确规定,否则单数形式包括复数指示物。除非另外定义,否则本文所用的所有技术和科学术语具有与本发明所属领域的普通技

术人员通常所理解相同的含义否则本文所用的所有技术和科学术语具有与本公开所属领域的普通技术人员通常所理解相同的含义。本说明书中引用的所有专利和出版物通过引用并入本文。

[0656] 前述描述仅出于说明的目的而呈现,并且不旨在将本公开限制为所公开的精确形式,而是由所附权利要求限定。

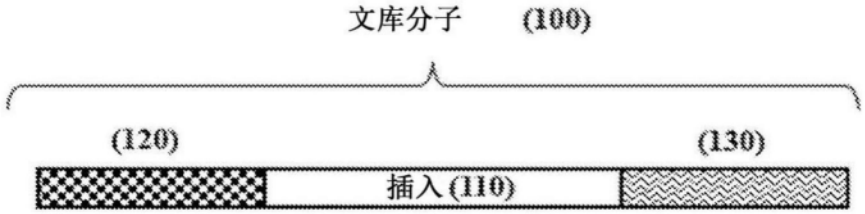

文库分子 (100)
(120)
(130)
插入(110)

图1

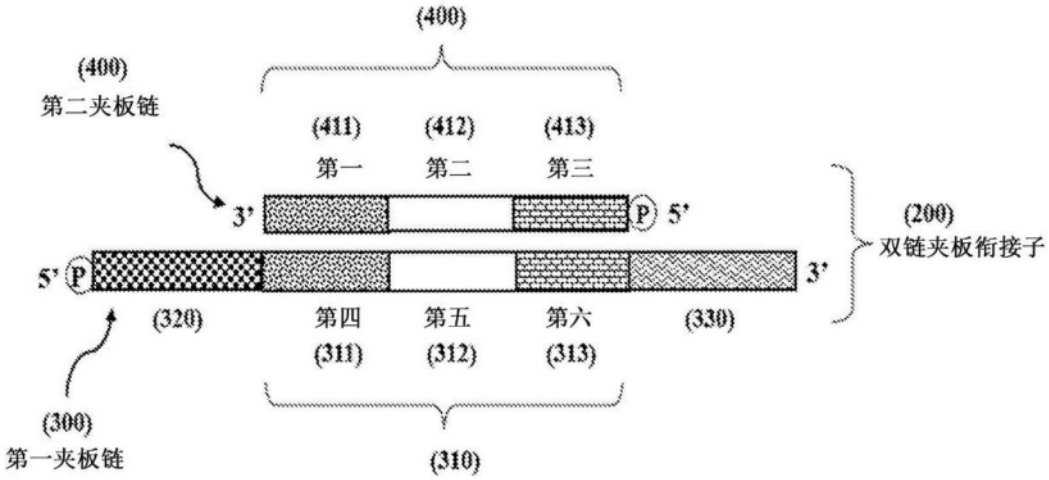

(400)
(400)
第二夹板链
(411)
(412)
(413)
第一
第二
第三
3'
P
5'
(200)
双链夹板衔接子
5'
P
3'
(320)
第四
第五
第六
(330)
(311)
(312)
(313)
(300)
第一夹板链
(310)

图2

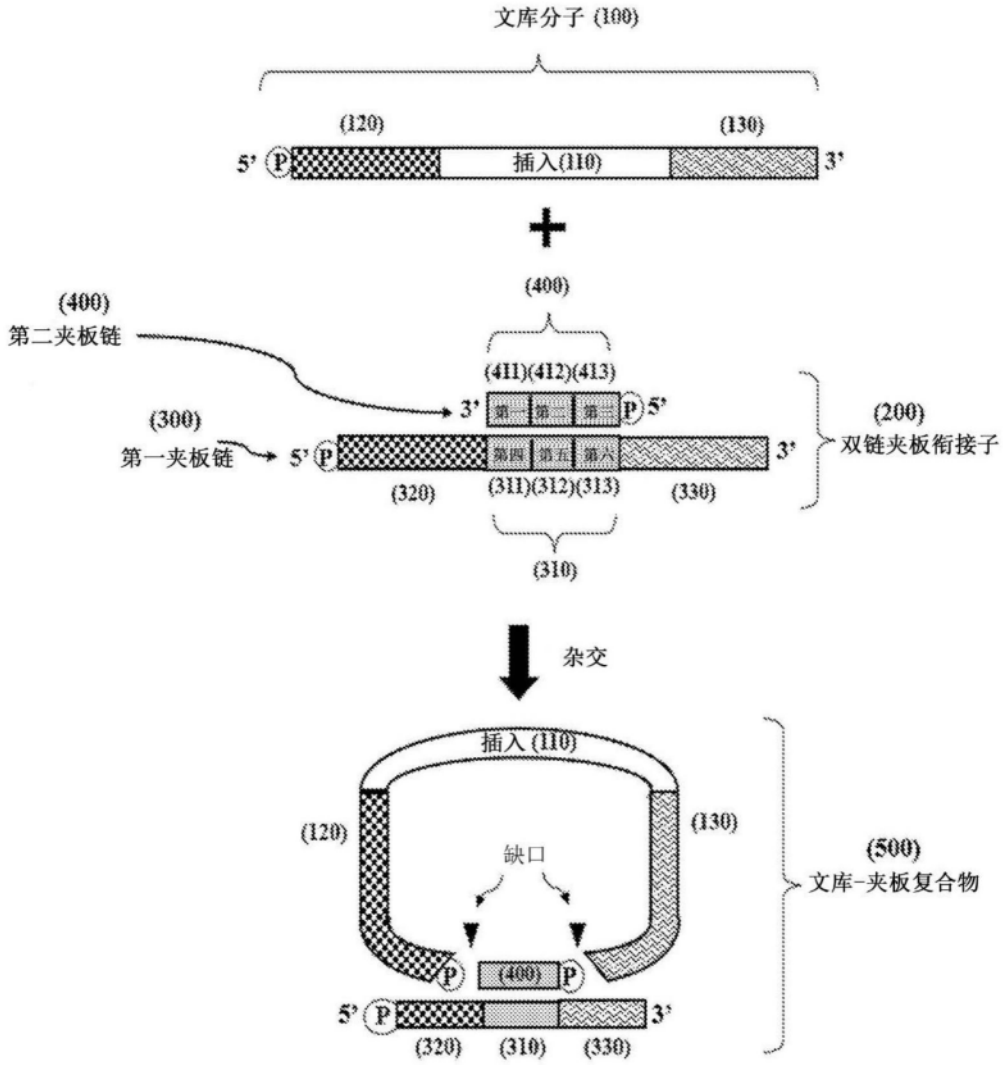

文库分子 (100)
(120)
(130)
5'
P
插入(110)
3'
(400)
第二夹板链
(400)
(411)(412)(413)
3'
第一
第二
第三
P
5'
(300)
第一夹板链
5'
P
第四
第五
第六
3'
(200)
双链夹板衔接子
(320)
(311)(312)(313)
(330)
(310)
杂交
插入 (110)
(120)
(130)
缺口
(500)
文库-夹板复合物
P
(400)
P
5'
P
3'
(320)
(310)
(330)

图3

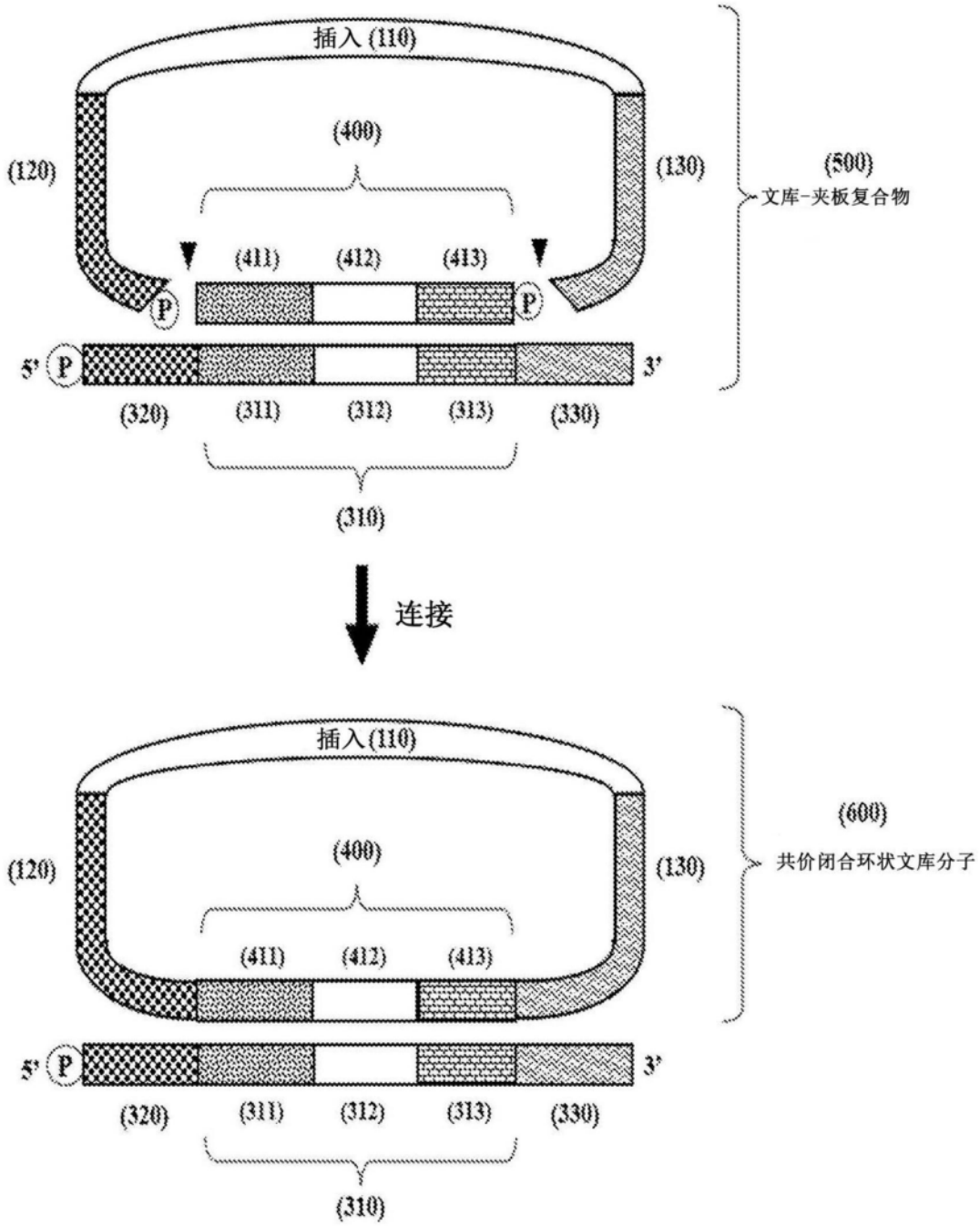

插入 (110)
(400)
(120)
(130)
(500)
文库-夹板复合物
(411)
(412)
(413)
P
P
5'
P
3'
(320)
(311)
(312)
(313)
(330)
(310)
连接
插入 (110)
(600)
共价闭合环状文库分子
(400)
(120)
(130)
(411)
(412)
(413)
5'
P
3'
(320)
(311)
(312)
(313)
(330)
(310)

图4

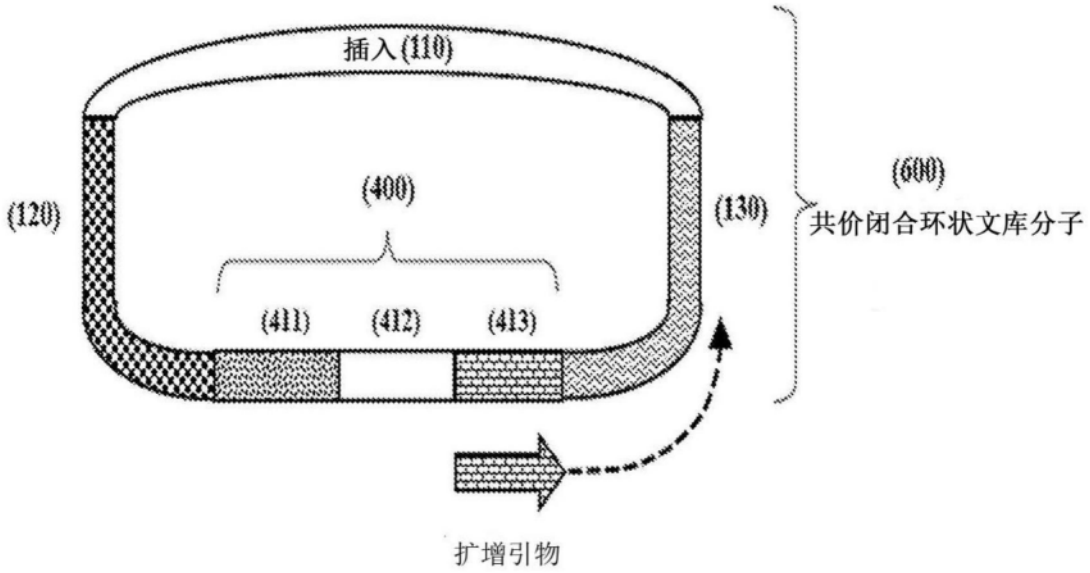


图5

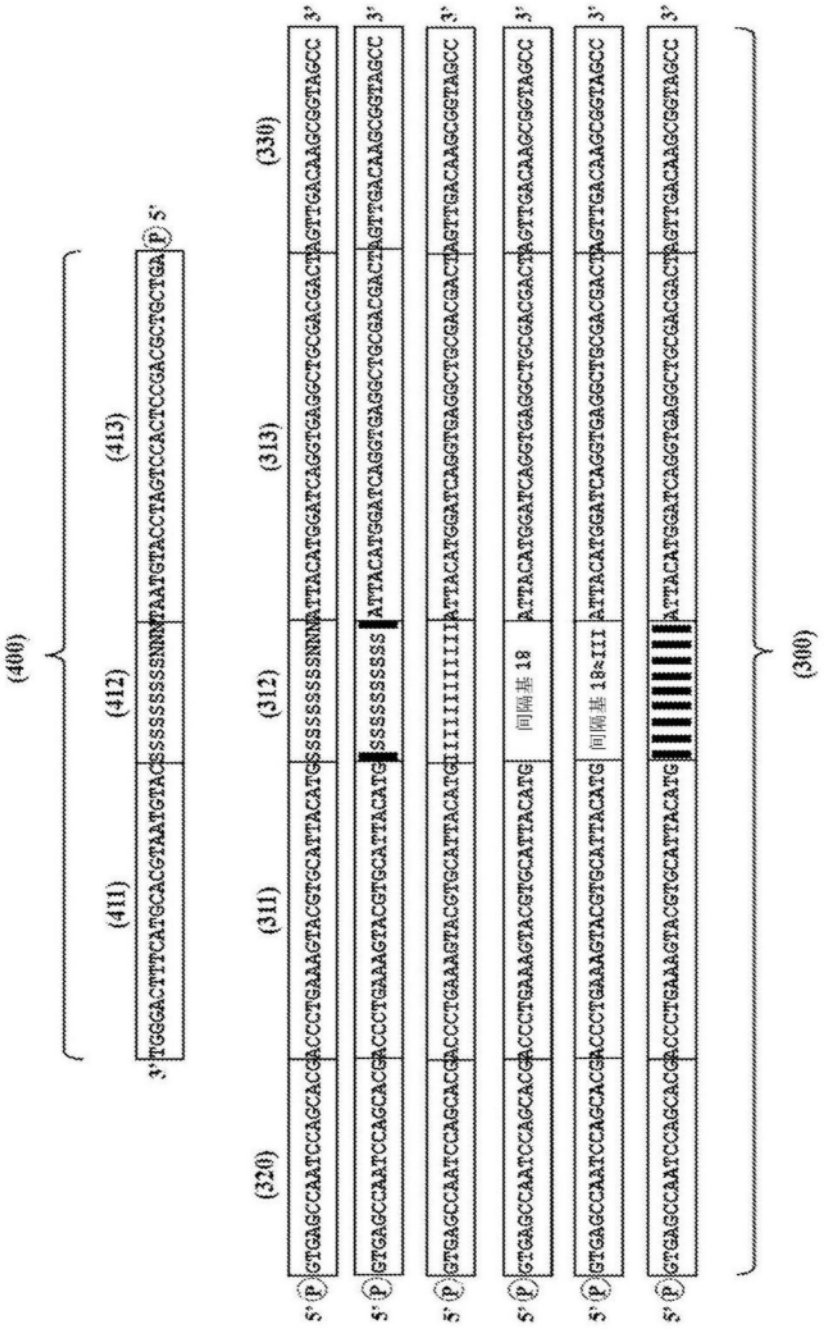

(400)
(411)
(412)
(413)
3' TGGGACTTTCATGCACGTAATGTAC SSSSSSSSSNNNTAATGTACCTAGTCCACTCCGACGCTGCTGA Ⓟ 5'
(320)
(311)
(312)
(313)
(330)
5' Ⓟ GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATG SSSSSSSSSNNNTTACATGGATCAGGTGAGGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
5' Ⓟ GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATG SSSSSSSSSS ATTACATGGATCAGGTGAGGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
5' Ⓟ GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATG IIIIIIIIIII ATTACATGGATCAGGTGAGGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
5' Ⓟ GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATG 间隔基 18 ATTACATGGATCAGGTGAGGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
5' Ⓟ GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATG 间隔基 18×III ATTACATGGATCAGGTGAGGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
5' Ⓟ GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATG ATTACATGGATCAGGTGAGGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
(300)

图6

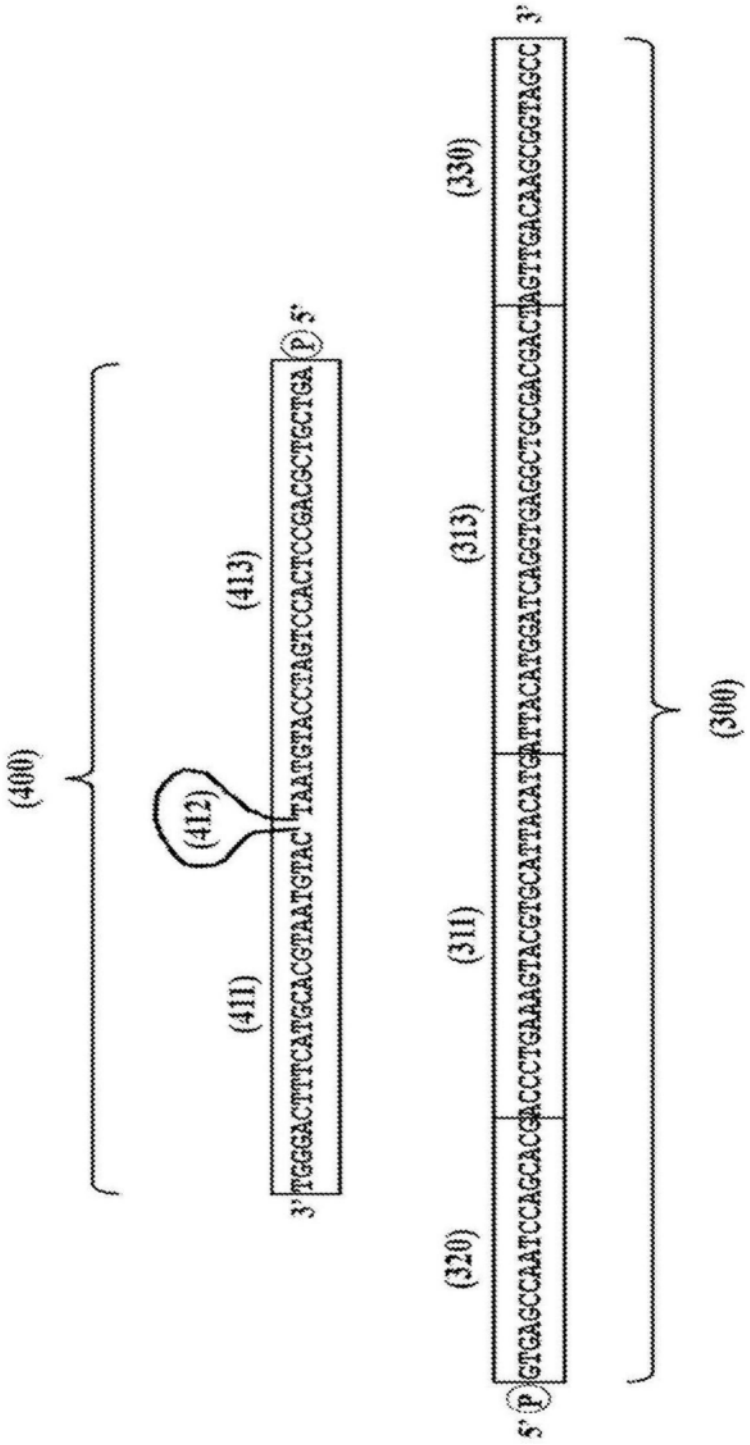

(400)
(412)
(411)
(413)
3'TGGGACTTTCATGCACGTAATGTAC TAATGTACCTAGTCCACTCCGACGCTGCTGA P 5'
(320)
(311)
(313)
(330)
5' P GTGAGCCAATCCAGCACGACCCTGAAAGTACGTGCATTACATGATTACATGGATCAGGTGACGCTGCGACGACTAGTTGACAAGCGGTAGCC 3'
(300)

图7

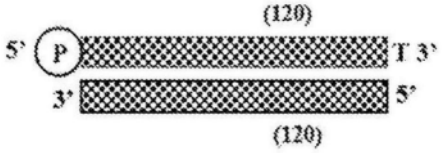

(120)
5'
P
T 3'
3'
5'
(120)

图8A

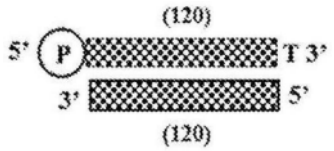


图8B

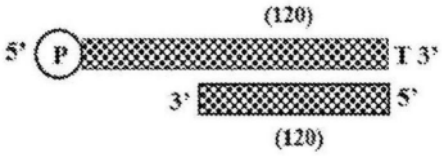


图8C

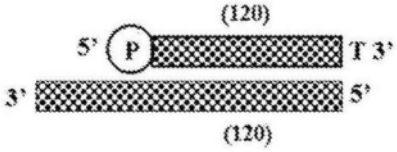


图8D

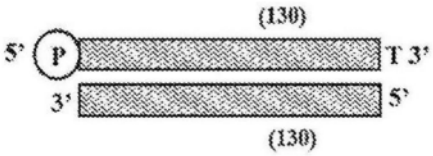


图9A

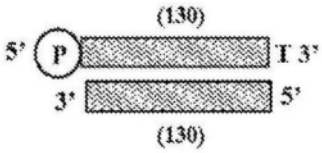


图9B

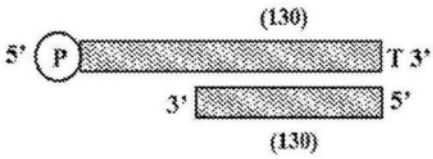


图9C

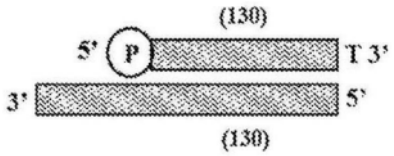


图9D

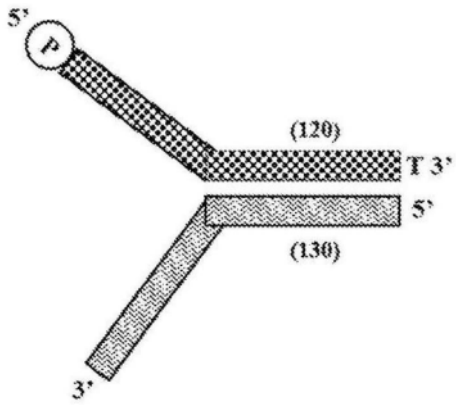


图10A

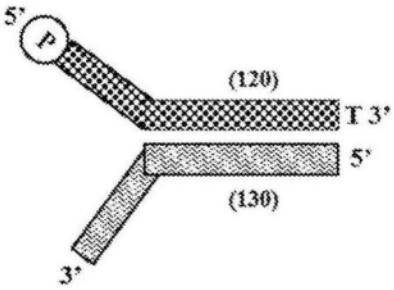


图10B

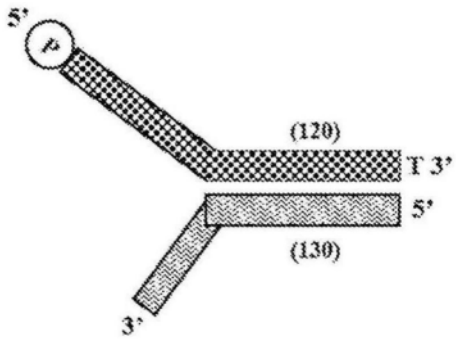


图10C

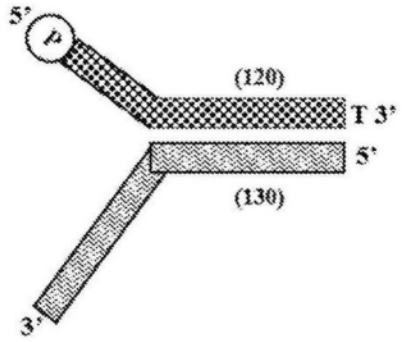


图10D

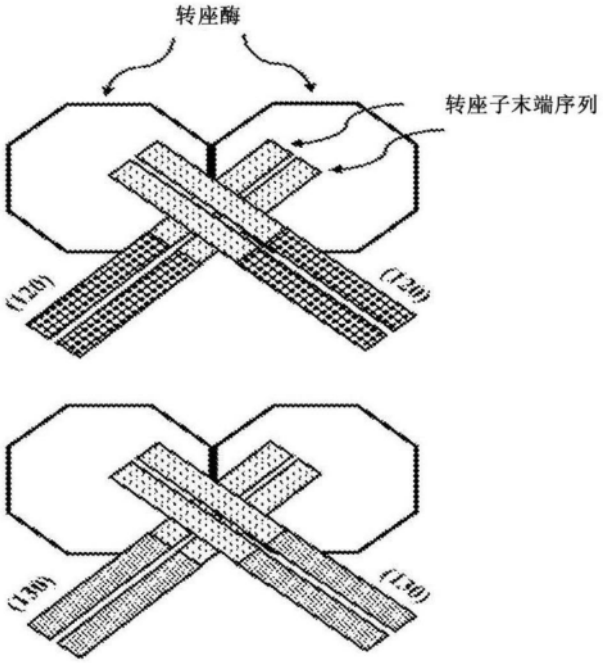


图11A

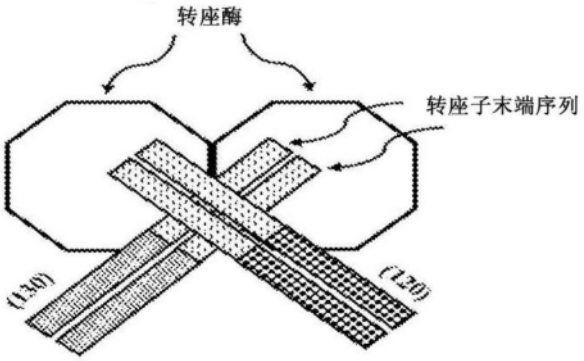


图11B

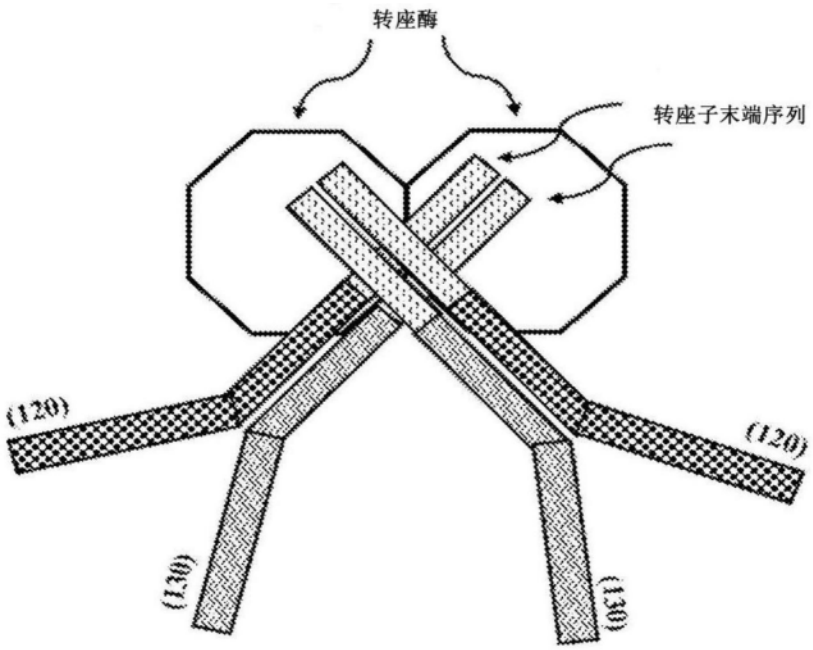

转座酶
转座子末端序列
(120)
(120)
(130)
(130)

图12

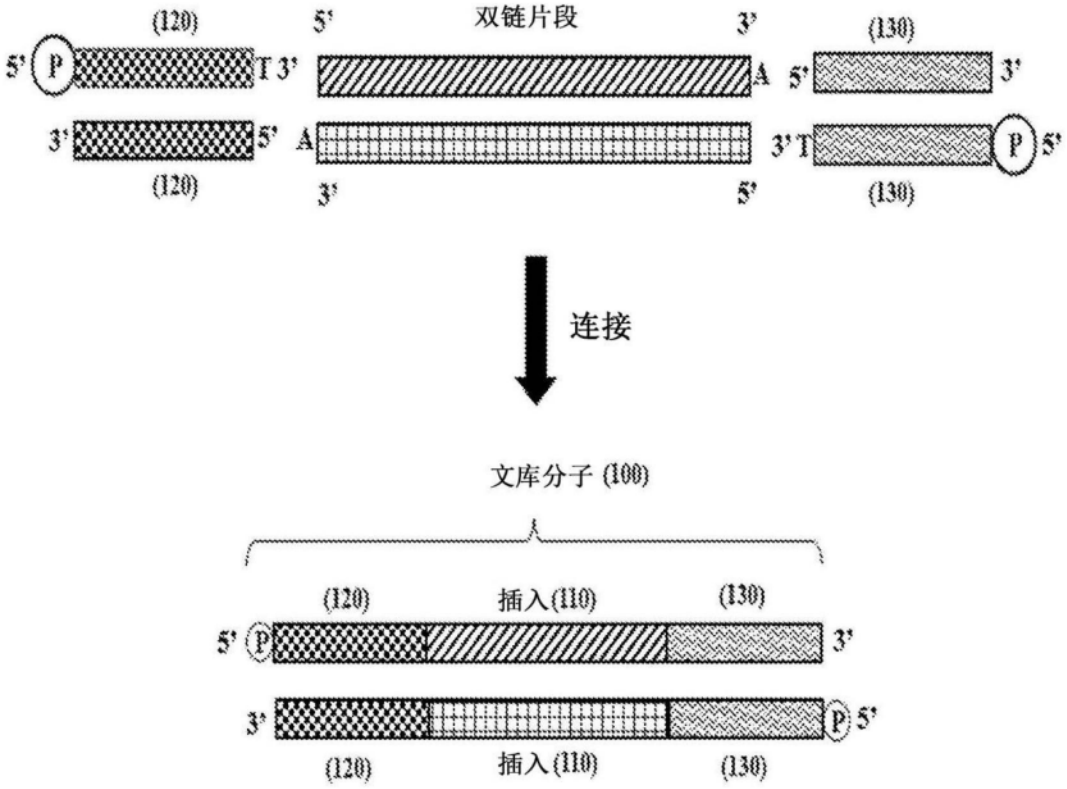

(120)
5'
双链片段
3'
(130)
5'
P
T 3'
A
5'
3'
3'
5'
A
3' T
P
5'
(120)
3'
5'
(130)
连接
文库分子 (100)
(120)
插入(110)
(130)
5'
P
3'
3'
P
5'
(120)
插入(110)
(130)

图13

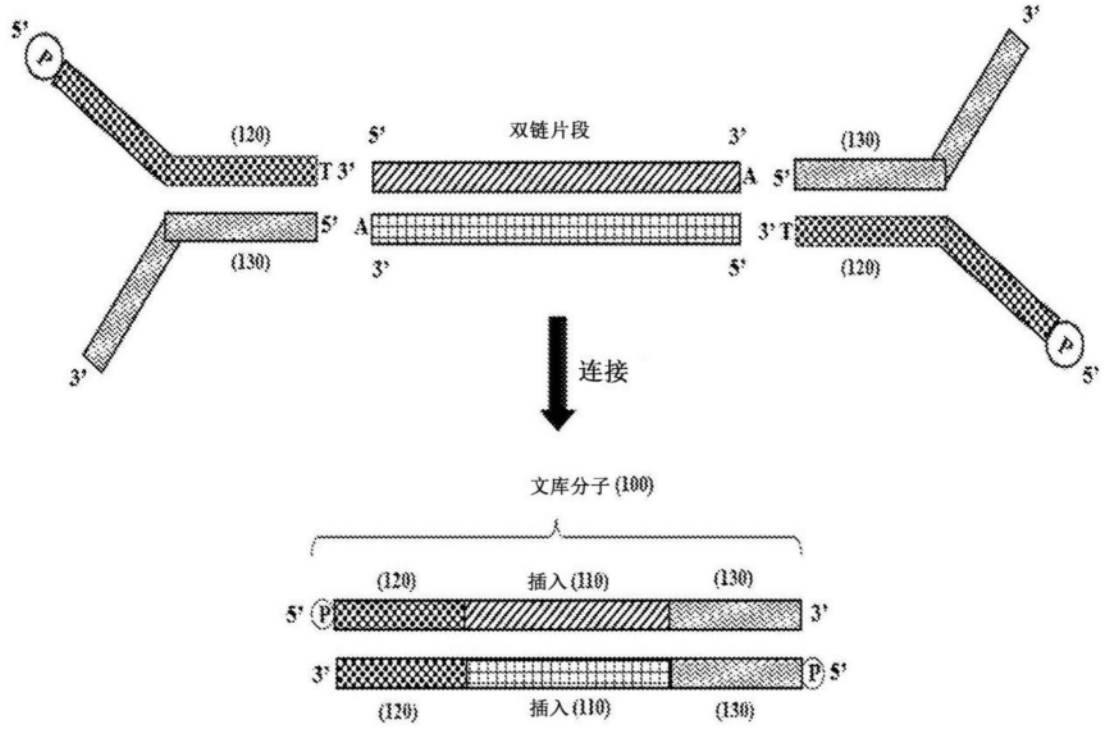

5'
P
(120)
T 3'
5'
双链片段
3'
A 5'
(130)
3'
5' A
(130)
3'
3' T
5'
(120)
P
5'
连接
文库分子 (100)
(120)
插入 (110)
(130)
5' P
3'
3'
P 5'
(120)
插入 (110)
(130)

图14

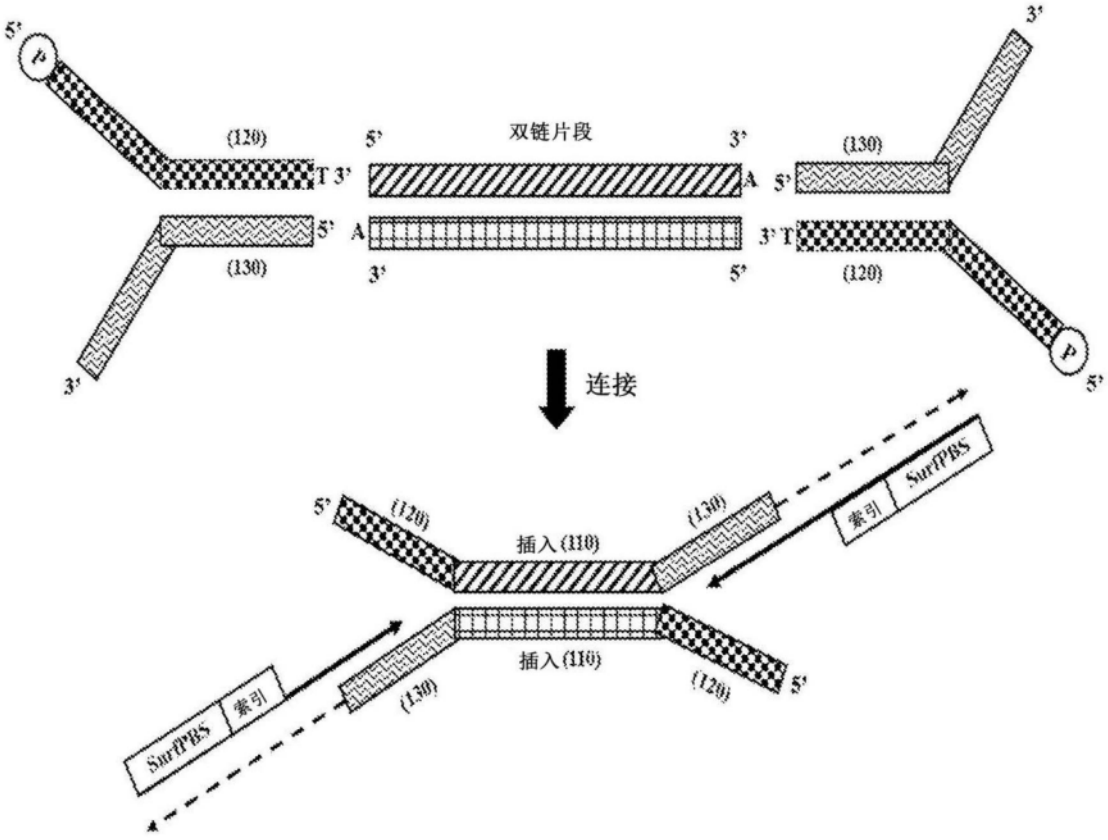

5'
P
(120)
T 3'
5'
双链片段
3'
A 5'
(130)
3'
5' A
(130)
3'
3' T
5'
(120)
P
5'
连接
5'
(120)
插入 (110)
(130)
索引
SurfPBS
SurfPBS
索引
(130)
插入 (110)
(120)
5'

图15

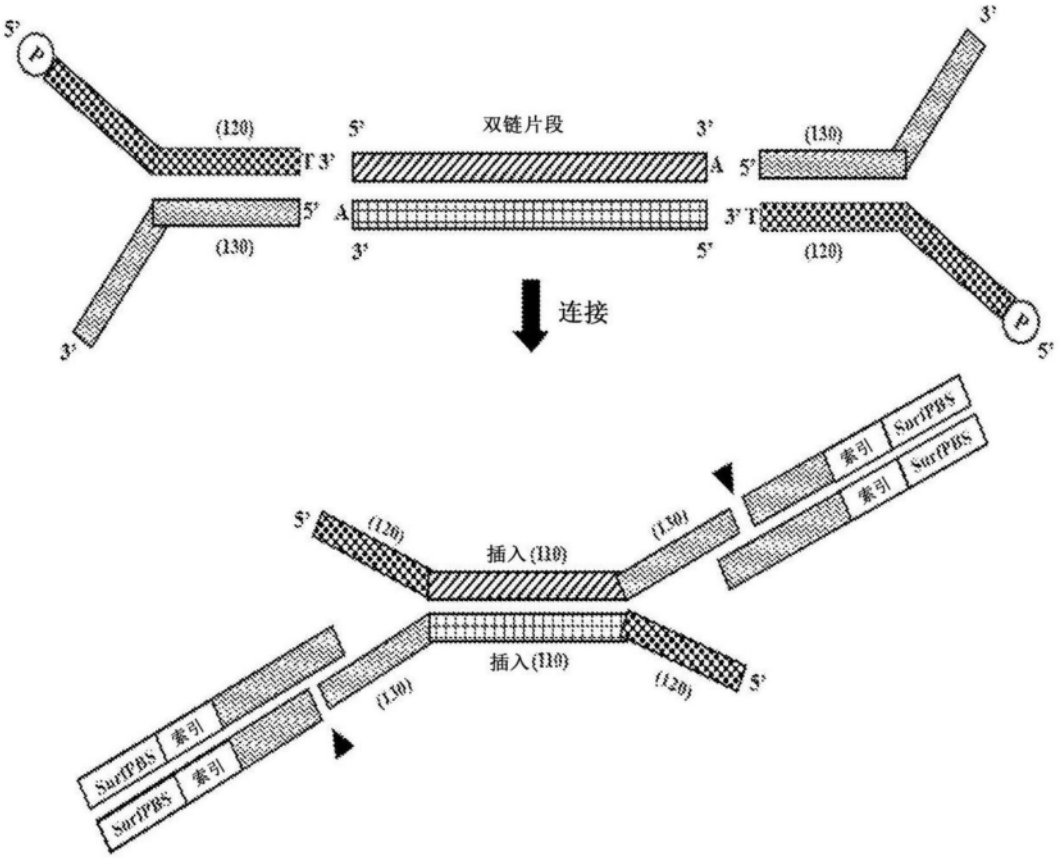

5'
P
(120)
5'
双链片段
3'
3'
(130)
T 3'
A
5'
5'
A
3' T
(130)
3'
5'
(120)
3'
连接
P
5'
5'
(120)
插入(110)
(130)
SurfPBS
索引
SurfPBS
索引
插入(110)
(130)
(120)
5'
SurfPBS
索引
SurfPBS
索引

图16

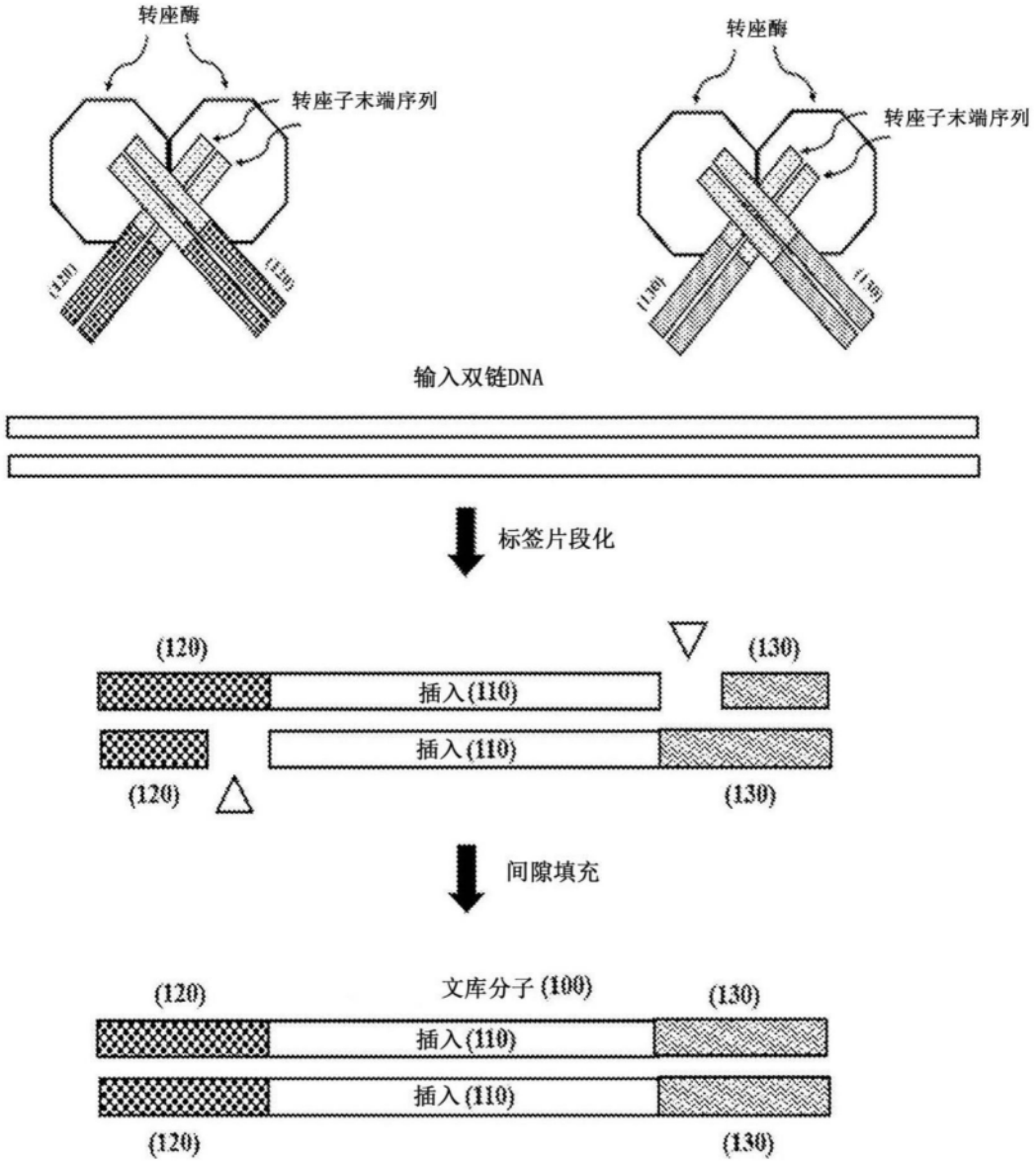

转座酶
转座子末端序列
转座酶
转座子末端序列
输入双链DNA
标签片段化
(120)
(130)
插入(110)
插入(110)
(120)
(130)
间隙填充
文库分子(100)
(120)
(130)
插入(110)
插入(110)
(120)
(130)

图17

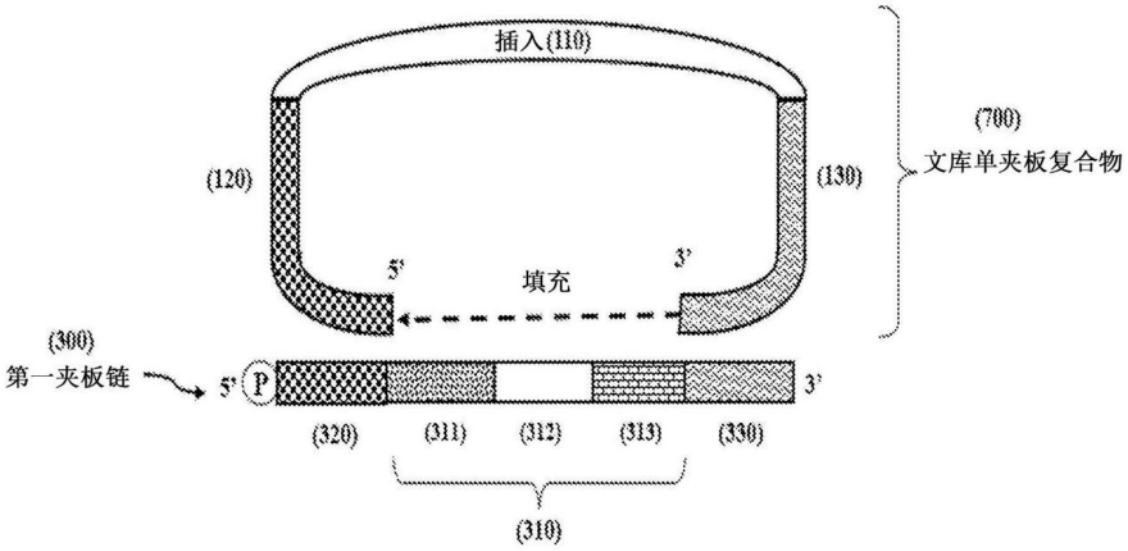

插入(110)
(120)
(130)
(700)
文库单夹板复合物
5'
填充
3'
(300)
第一夹板链
5'
P
3'
(320)
(311)
(312)
(313)
(330)
(310)

图18

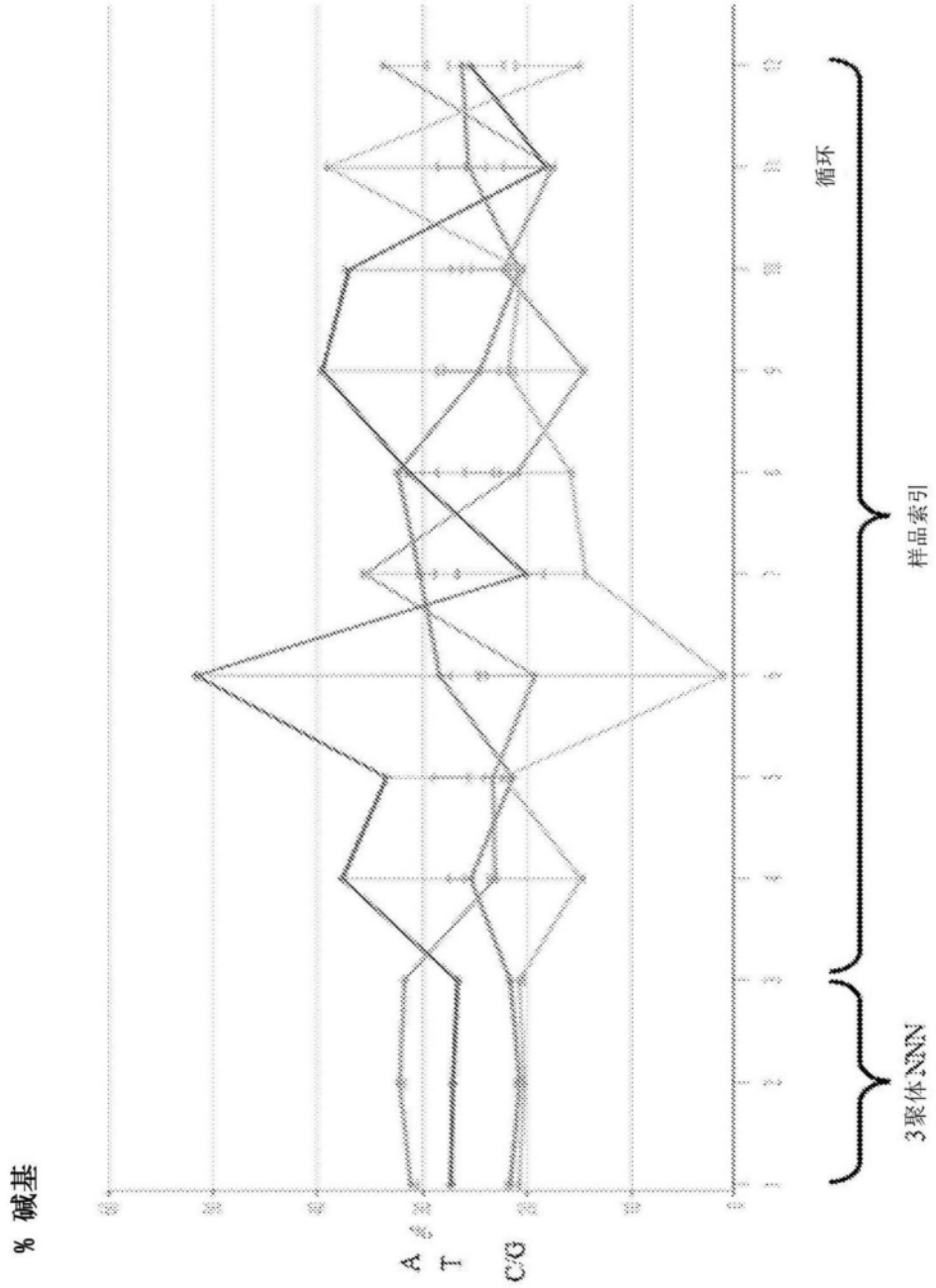


图19

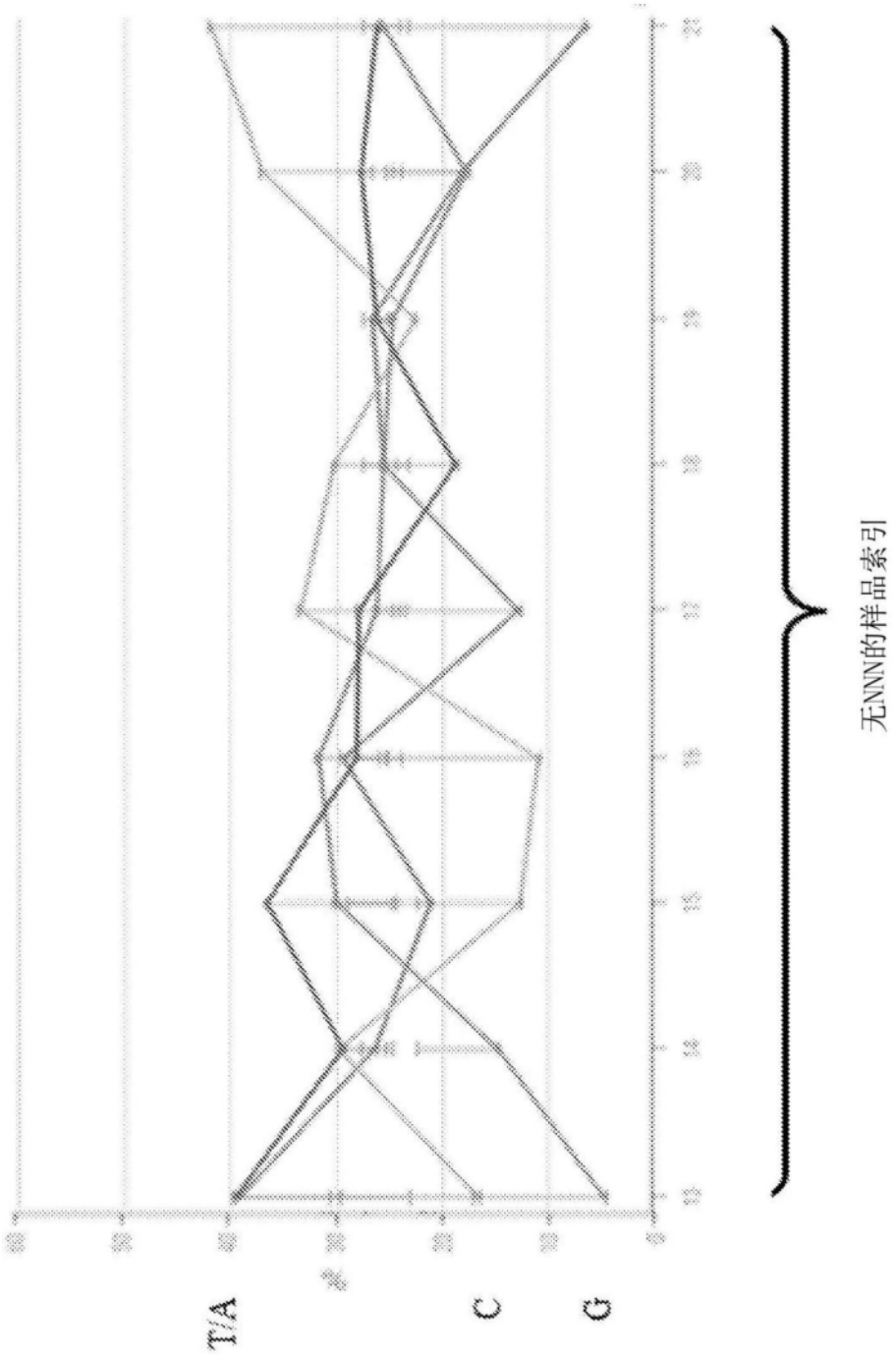


图20

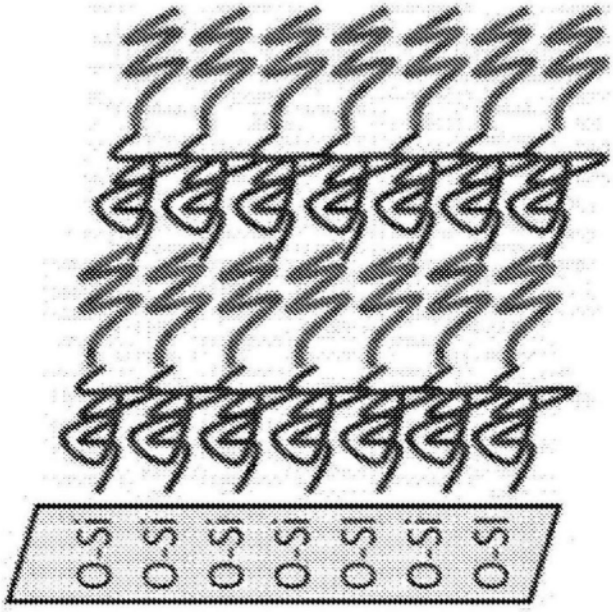


图21

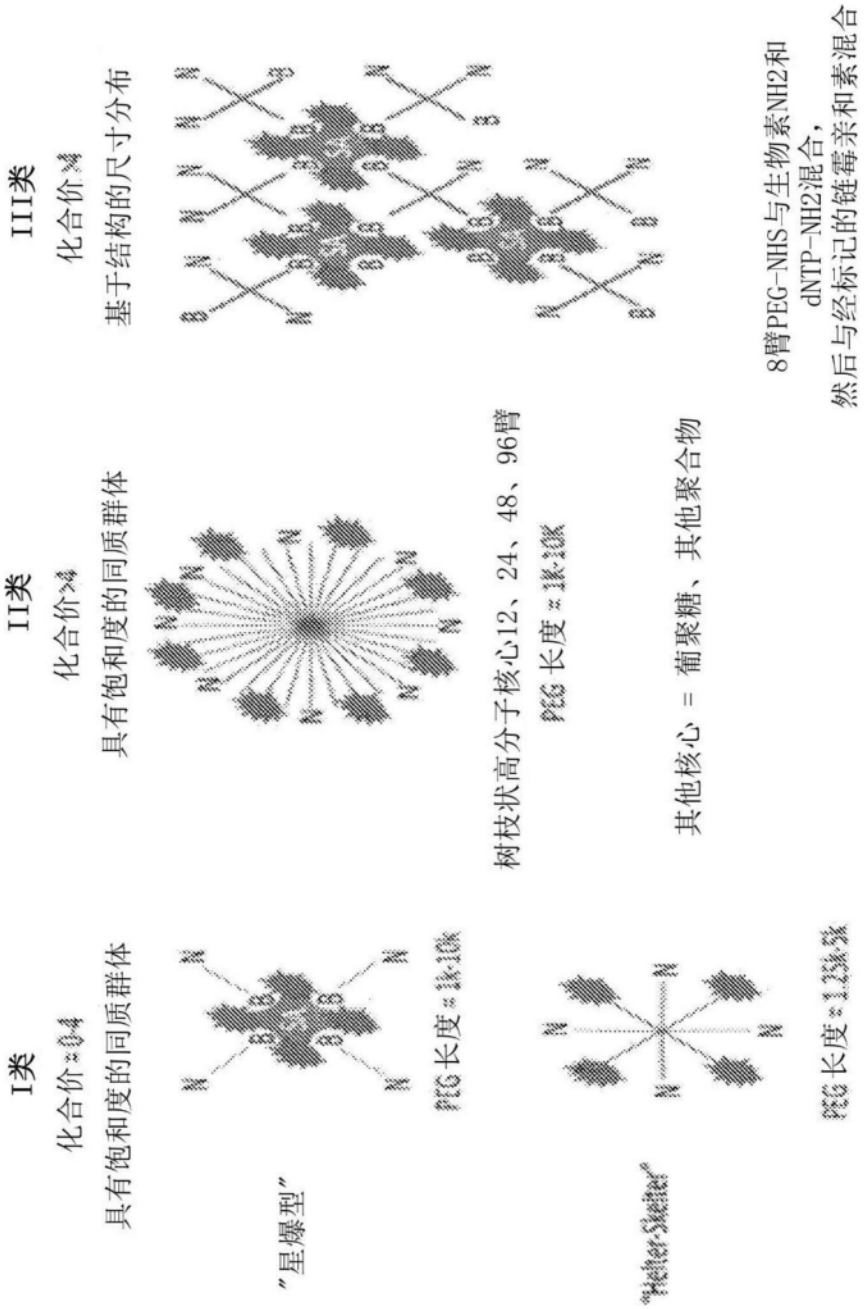

I类
化合价=0-4
具有饱和度的同质群体
"星爆型"
PEG长度=1k-10k
"Heiter-Skeiter"
PEG长度=1.25k-5k
II类
化合价>4
具有饱和度的同质群体
树枝状高分子核心12、24、48、96臂
PEG长度=1k-10K
其他核心 = 葡聚糖、其他聚合物
III类
化合价>4
基于结构的尺寸分布
8臂PEG-NHS与生物素NH2和
dNTP-NH2混合，
然后与经标记的链霉亲和素混合

图22

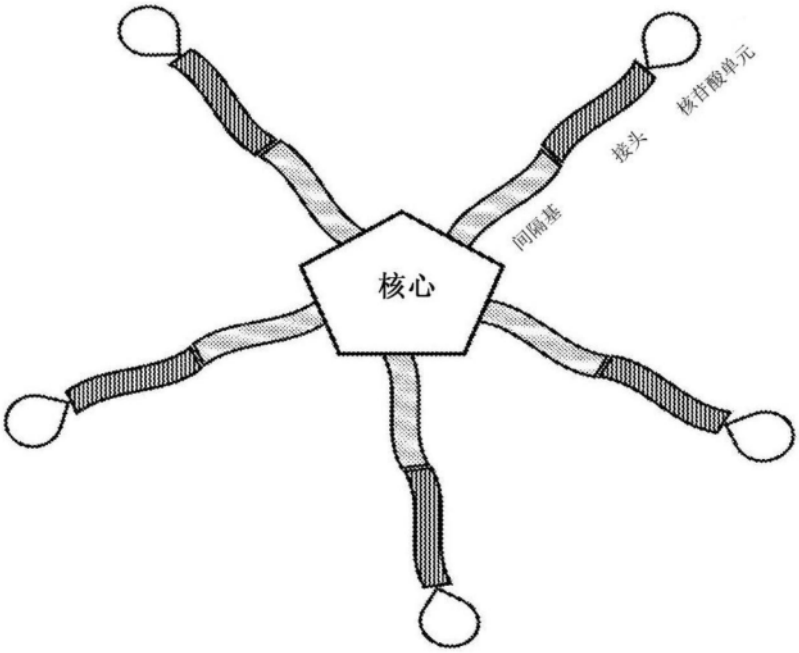


图23

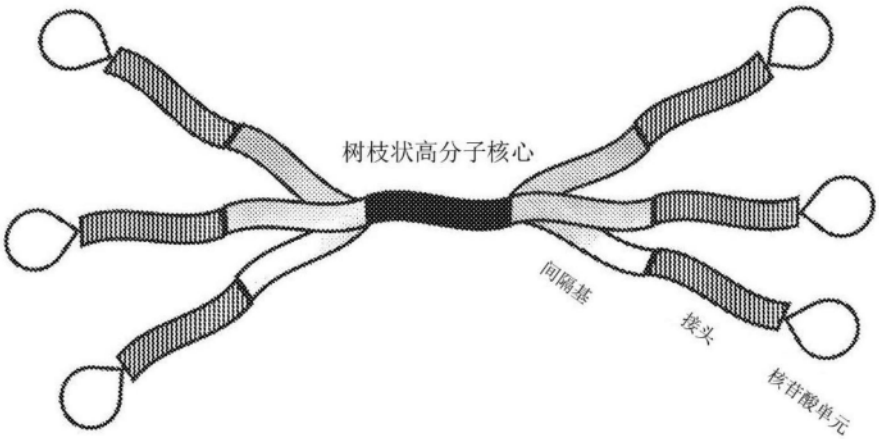


图24

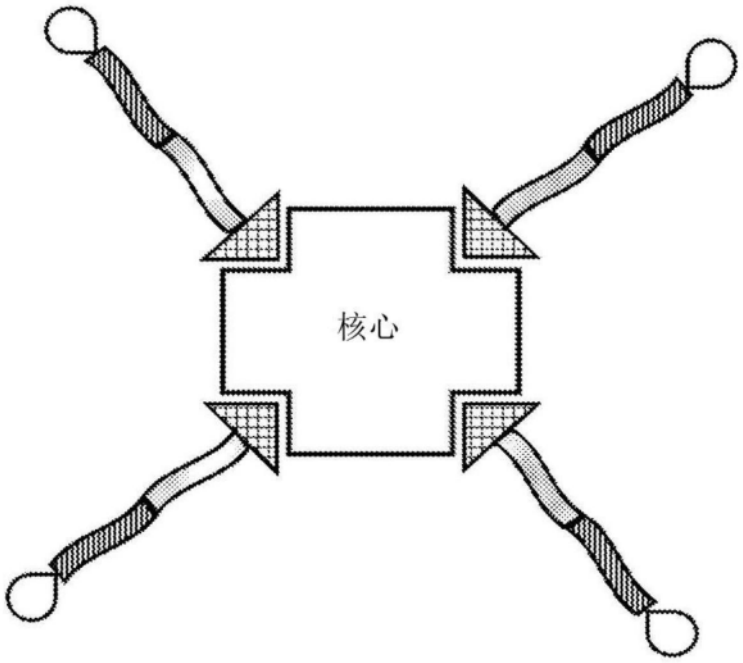


图25

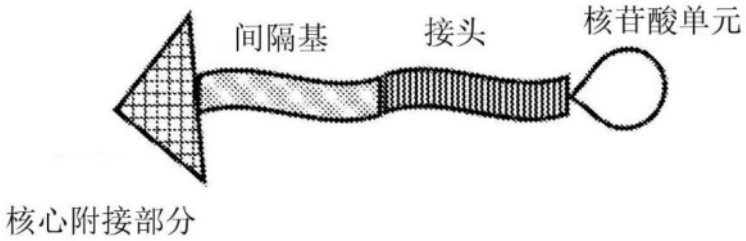
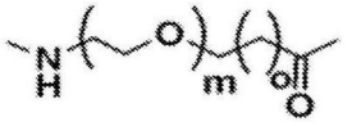


图26

间隔基：

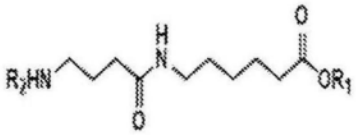


接头：

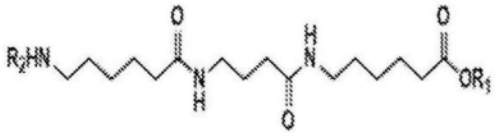
11个原子接头：



16个原子接头：



23个原子接头：



N3接头：

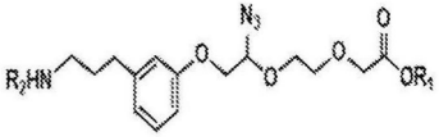
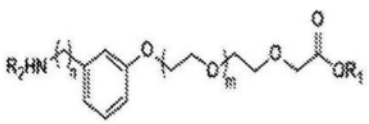
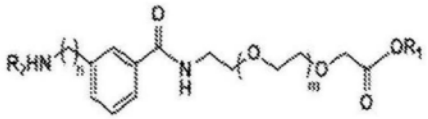


图27

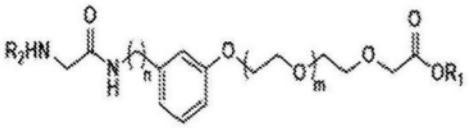
接头1:



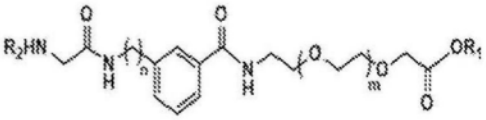
接头2:



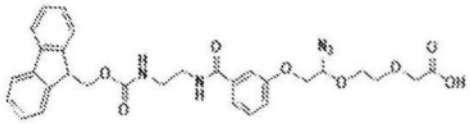
接头3:



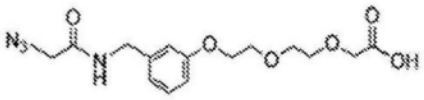
接头4:



接头5:

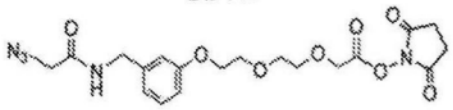


接头6:



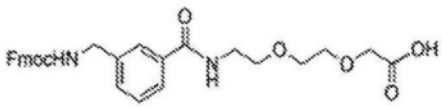
分子量：352.35

接头7:



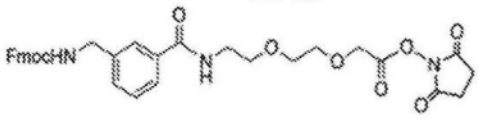
分子量：449.42

接头8:



分子量：518.57

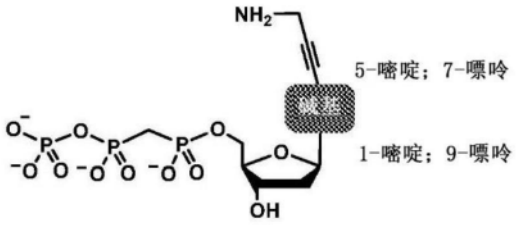
接头9:



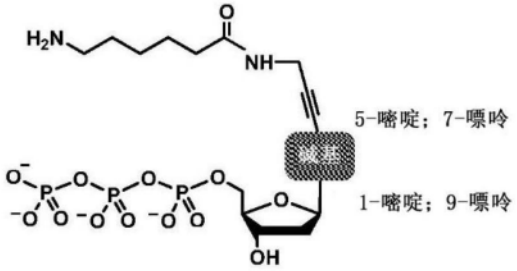
分子量：615.64

图28

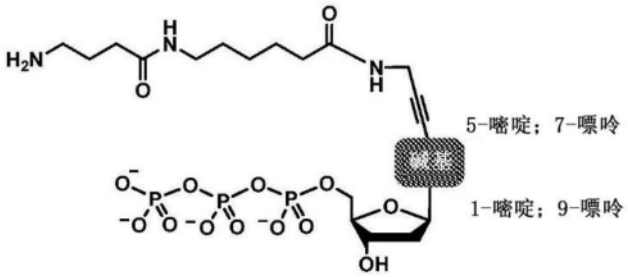
dNTP-PA-NH_2:



dNTP-PA-11个原子接头 -NH_2:



dNTP-PA-16 个原子接头 -NH_2:



dNTP-PA-23个原子接头 -NH_2:

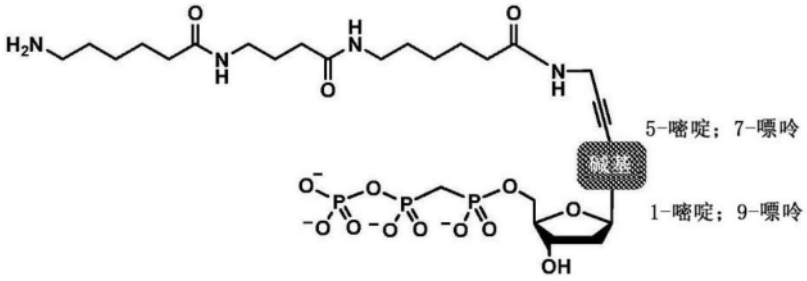
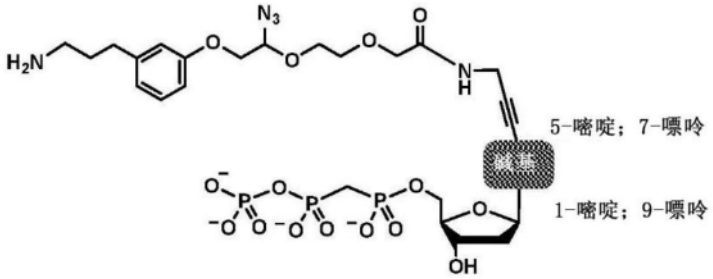
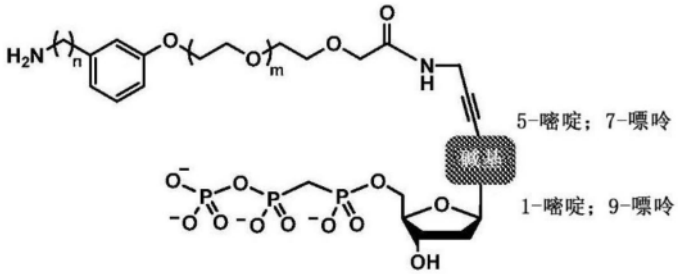


图29

dNTP-PA-N3接头 -NH_2:



dNTP-PA-接头1-NH_2:



dNTP-PA-接头2-NH_2:

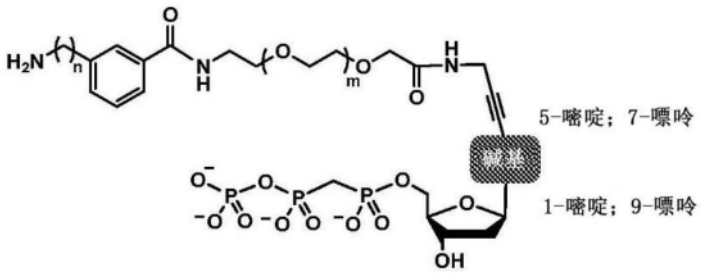
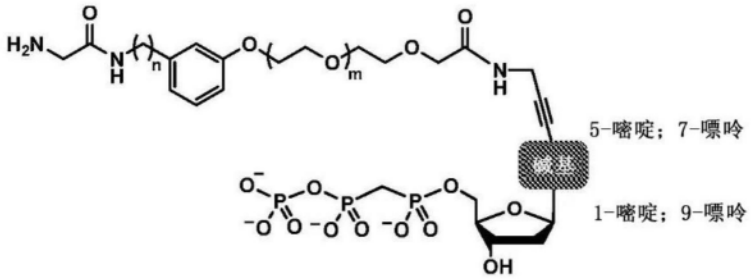
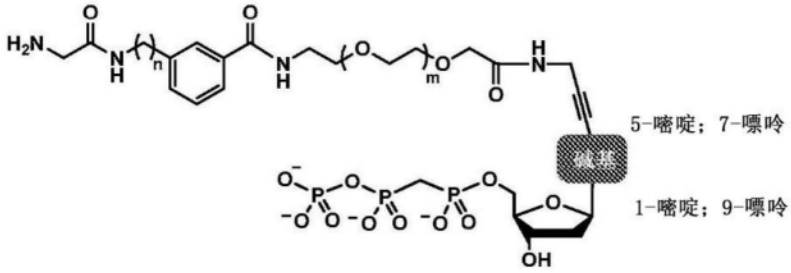


图30

dNTP-PA-接头 3-NH_2:



dNTP-PA-接头 4-NH_2:



dNTP-PA-N3 接头 -NH_2:

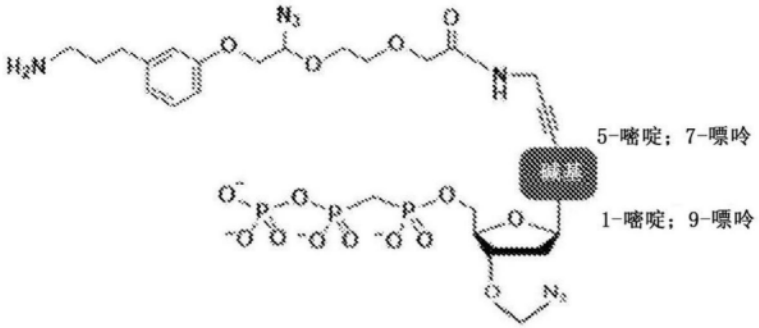


图31

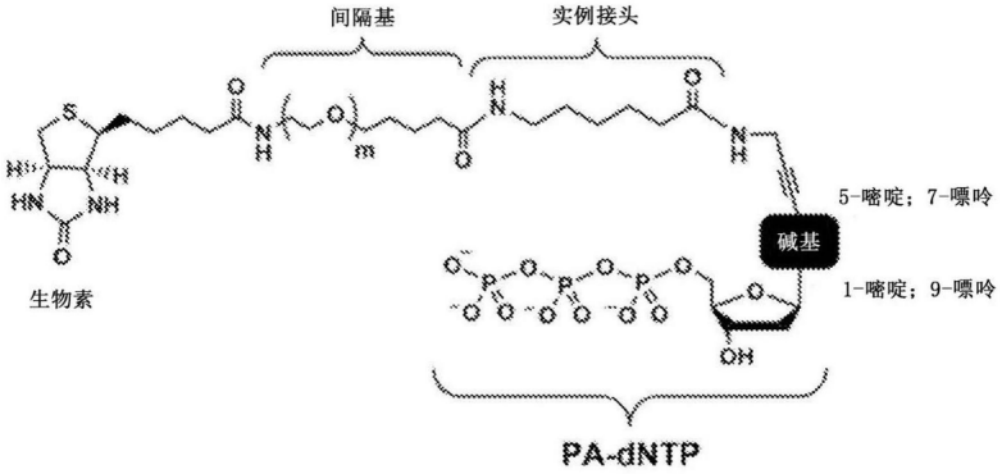


图32