



# (12) 发明专利

(10) 授权公告号 CN 112243462 B

(45) 授权公告日 2025. 02. 25

(21) 申请号 201980038046.3

(22) 申请日 2019.06.05

(65) 同一申请的已公布的文献号  
申请公布号 CN 112243462 A

(43) 申请公布日 2021.01.19

(30) 优先权数据  
62/681,524 2018.06.06 US

(85) PCT国际申请进入国家阶段日  
2020.12.04

(86) PCT国际申请的申请数据  
PCT/US2019/035617 2019.06.05

(87) PCT国际申请的公布数据  
W02019/236726 EN 2019.12.12

(73) 专利权人 加利福尼亚大学董事会  
地址 美国加利福尼亚州

(72) 发明人 R·格林 J·卡普

(74) 专利代理机构 北京龙双利达知识产权代理有限公司 11329  
专利代理师 孙涛 肖鹏

(51) Int.Cl.  
C12Q 1/6806 (2006.01)

(56) 对比文件  
CN 103354840 A, 2013.10.16  
WO 2016058517 A1, 2016.04.21  
WO 2016156529 A1, 2016.10.06

审查员 周怡婷

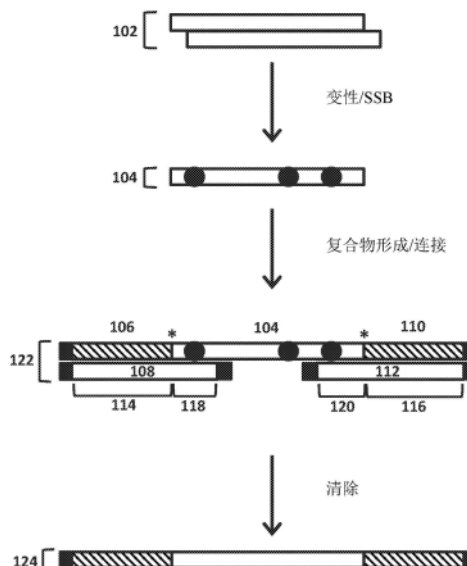
权利要求书3页 说明书16页  
序列表2页 附图6页

## (54) 发明名称

产生核酸文库的方法以及用于实践所述方法的组合物和试剂盒

## (57) 摘要

提供了产生核酸文库的方法。所述方法包含将单链核酸结合蛋白结合的单链核酸 (SSB结合 ssNA)、衔接子寡核苷酸和夹板寡核苷酸结合,以形成包含所述夹板寡核苷酸与所述SSB结合的 ssNA的末端区域并与所述衔接子寡核苷酸杂交的复合物。所述第一衔接子寡核苷酸的端部与所述SSB结合的 ssNA的第一末端区域的端部相邻,并且所述方法可以进一步包含共价连接相邻端部。还提供了例如在实践本公开的所述方法中有用的组合物和试剂盒。



1. 一种产生核酸文库的方法,所述方法包括:

使单链核酸(ssNA)与单链核酸结合蛋白(SSB)接触,以产生SSB结合的ssNA;

将以下组合:

所述SSB结合的ssNA,

第一衔接子寡核苷酸,以及

第一夹板寡核苷酸,所述第一夹板寡核苷酸包括SSB结合的ssNA杂交区域和第一衔接子寡核苷酸杂交区域,

以形成包括以下的复合物:所述第一夹板寡核苷酸通过所述SSB结合的ssNA杂交区域与所述SSB结合的ssNA的末端区域杂交;并且所述第一夹板寡核苷酸通过所述第一衔接子寡核苷酸杂交区域与所述第一衔接子寡核苷酸杂交,使得所述第一衔接子寡核苷酸的端部与所述SSB结合的ssNA的所述末端区域的端部相邻;以及

将以下组合:

所述SSB结合的ssNA,

第二衔接子寡核苷酸;以及

第二夹板寡核苷酸,所述第二夹板寡核苷酸包括SSB结合的ssNA杂交区域和第二衔接子寡核苷酸杂交区域,

其中形成的复合物进一步包括:所述第二夹板寡核苷酸通过所述SSB结合的ssNA杂交区域与所述SSB结合的ssNA的所述末端区域杂交的位置与和所述第一夹板寡核苷酸杂交的所述末端区域的位置相反;并且所述第二夹板寡核苷酸通过所述第二衔接子寡核苷酸杂交区域与所述第二衔接子寡核苷酸杂交,使得所述第二衔接子寡核苷酸的端部与所述SSB结合的ssNA的端部相邻的位置与和所述第一衔接子寡核苷酸相邻的所述端部的位置相反,其中所述第一衔接子寡核苷酸和所述第二衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部包括阻断修饰,以及

所述端部包括阻断修饰指端部不能够使用用于共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部的方法与任何其它寡核苷酸组分的端部连接;

所述方法进一步包括共价连接所述第一衔接子寡核苷酸和所述SSB结合的ssNA的相邻端部,以及

共价连接所述第二衔接子寡核苷酸和所述SSB结合的ssNA的相邻端部。

2. 根据权利要求1所述的方法,其中接触、组合和共价连接步骤的总持续时间小于或等于3小时。

3. 根据权利要求1所述的方法,其中所述复合物由大于或等于80%的所述ssNA与所述SSB接触而形成。

4. 根据权利要求1所述的方法,其中所述ssNA来自经降解核酸样品。

5. 根据权利要求4所述的方法,其中所述ssNA来自古老(ancient)核酸样品。

6. 根据权利要求1所述的方法,其中所述ssNA来自法医核酸样品。

7. 根据权利要求1所述的方法,其中所述ssNA为单链DNA(ssDNA)。

8. 根据权利要求7所述的方法,其中所述ssDNA衍生自双链DNA(dsDNA)。

9. 根据权利要求8所述的方法,其进一步包括在使所述ssDNA与SSB接触之前,通过使所述dsDNA变性来产生所述ssDNA。

10. 根据权利要求8所述的方法,其进一步包括在形成所述复合物之后,使所述ssDNA再杂交以产生dsDNA。

11. 根据权利要求10所述的方法,其进一步包括对所产生的dsDNA进行测序。

12. 根据权利要求11所述的方法,其中对所产生的dsDNA的子样品进行测序。

13. 根据权利要求1所述的方法,其中所述ssNA为单链RNA(ssRNA)。

14. 根据权利要求1所述的方法,其中所述共价连接包括化学连接或酶连接。

15. 根据权利要求1所述的方法,其中所述阻断修饰为连接阻断修饰。

16. 根据权利要求1所述的方法,其中所述阻断修饰选自由以下组成的组:在所述衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部不存在3' OH;以及在所述衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部处存在被修饰的3' OH。

17. 根据权利要求16所述的方法,其中所述阻断修饰是在所述衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部处存在被修饰的3' OH,并且其中所述阻断修饰选自由以下组成的组:氨基修饰剂、间隔子、双脱氧碱基、反向的双脱氧碱基和3' 磷酸。

18. 根据权利要求1所述的方法,其中所述SSB结合的ssNA杂交区域包括随机序列。

19. 根据权利要求1所述的方法,其中所述SSB结合的ssNA杂交区域包括通用碱基。

20. 根据权利要求1所述的方法,其中所述SSB结合的ssNA杂交区域的长度小于或等于10个核苷酸。

21. 根据权利要求1所述的方法,其中所述衔接子寡核苷酸包括用于PCR扩增的衔接子或其补体。

22. 根据权利要求1所述的方法,其中所述衔接子寡核苷酸包括部分或完全测序衔接子或其补体。

23. 根据权利要求1到22中任一项所述的方法,其进一步包括对所述ssNA的至少一部分进行测序。

24. 一种组合物,其包括:

单链核酸结合蛋白结合的单链核酸(SSB结合的ssNA);

第一衔接子寡核苷酸;

第一夹板寡核苷酸,所述第一夹板寡核苷酸包括SSB结合的ssNA杂交区域和第一衔接子寡核苷酸杂交区域;

第二衔接子寡核苷酸;以及

第二夹板寡核苷酸,所述第二夹板寡核苷酸包括SSB结合的ssNA杂交区域和第二衔接子寡核苷酸杂交区域,其中所述第一衔接子寡核苷酸和所述第二衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部包括阻断修饰,以及

所述端部包括阻断修饰指端部不能够使用用于共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部的方法与任何其它寡核苷酸组分的端部连接;

所述组合物进一步包括:

用于将衔接子寡核苷酸端部共价连接到所述SSB结合的ssNA的端部的试剂。

25. 根据权利要求24所述的组合物,其中所述ssNA来自经降解核酸样品。

26. 根据权利要求25所述的组合物,其中所述ssNA来自古老核酸样品。

27. 根据权利要求24所述的组合物,其中所述ssNA来自法医核酸样品。

28. 根据权利要求24所述的组合物,其中所述ssNA为单链DNA(ssDNA)。
29. 根据权利要求28所述的组合物,其中所述ssDNA衍生自双链DNA(dsDNA)。
30. 根据权利要求24所述的组合物,其中所述ssNA为单链RNA(ssRNA)。
31. 根据权利要求24所述的组合物,其中所述试剂为连接酶。
32. 一种试剂盒,其包括:  
单链核酸结合蛋白(SSB);  
第一衔接子寡核苷酸;  
第一夹板寡核苷酸,所述第一夹板寡核苷酸包括SSB结合的ssNA杂交区域和第一衔接子寡核苷酸杂交区域,  
说明书,所述说明书用于使用所述SSB、所述第一衔接子寡核苷酸和所述第一夹板寡核苷酸来产生核酸文库;  
第二衔接子寡核苷酸;以及  
第二夹板寡核苷酸,所述第二夹板寡核苷酸包括SSB结合的ssNA杂交区域和第二衔接子寡核苷酸杂交区域,其中所述第一衔接子寡核苷酸和所述第二衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部包括阻断修饰,以及  
所述端部包括阻断修饰指端部不能够使用用于共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部的方法与任何其它寡核苷酸组分的端部连接,以及  
其中所述说明书用于使用所述SSB、所述第一衔接子寡核苷酸、所述第一夹板寡核苷酸、所述第二衔接子寡核苷酸和所述第二夹板寡核苷酸来产生核酸文库;  
所述试剂盒进一步包括:  
用于将衔接子寡核苷酸的端部连接到单链核酸结合蛋白结合的单链核酸(SSB结合的ssNA)的端部的试剂。
33. 根据权利要求32所述的试剂盒,其中所述SSB为单链DNA结合蛋白。
34. 根据权利要求32所述的试剂盒,其中所述SSB为单链RNA结合蛋白。
35. 根据权利要求32所述的试剂盒,其中所述试剂为连接酶。
36. 根据权利要求32所述的试剂盒,其中所述阻断修饰为连接阻断修饰。
37. 根据权利要求32所述的试剂盒,其中所述阻断修饰选自自由以下组成的组:在所述衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部不存在3' OH;以及在所述衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部处存在被修饰的3' OH。
38. 根据权利要求32所述的试剂盒,其中所述阻断修饰是在所述衔接子寡核苷酸的与所述SSB结合的ssNA不相邻的端部处存在被修饰的3' OH,并且其中所述阻断修饰选自自由以下组成的组:氨基修饰剂、间隔子、双脱氧碱基、反向的双脱氧碱基和3'磷酸。
39. 根据权利要求32所述的试剂盒,其中所述SSB结合的ssNA杂交区域包括随机序列。
40. 根据权利要求32所述的试剂盒,其中所述SSB结合的ssNA杂交区域包括通用碱基。
41. 根据权利要求32至40中任一项所述的试剂盒,其中所述SSB结合的ssNA杂交区域的长度小于或等于10个核苷酸。

## 产生核酸文库的方法以及用于实践所述方法的组合物和试剂盒

[0001] 相关申请的交叉引用

[0002] 本申请要求于2018年6月6日提交的美国临时专利申请第62/681,524号的权益,所述申请通过引用以其整体并入本文。

### 背景技术

[0003] 核酸测序已成为遗传学研究的越来越重要的领域,用于诊断和其它应用中。通常,核酸测序由确定核酸,如RNA或DNA的片段的核苷酸顺序组成。通常会分析相对短的序列,并且可以在各种生物信息学方法中使用所得序列信息来将片段与参考序列进行比对,或者将片段在逻辑上整合在一起,以便可靠地确定衍生出所述片段的长得多的遗传物质的序列。已经开发了对特性片段的自动化的基于计算机的检查,并且最近已将其用于基因组图谱定位、个体之间的遗传变异分析、基因和其功能的鉴定等。

[0004] 用于高通量DNA测序的几种方法依赖于通用扩增反应,由此对DNA样品进行随机片段化,然后进行处理,从而使得不同片段的端部都含有相同的DNA序列。具有通用端部的片段可以在单个反应中用单个扩增引物对进行扩增。通用引发序列到待通过PCR扩增的靶的端部上的添加可以通过多种方法实现。例如,可以使用在其5'端处具有通用序列并且在其3'端处具有简并序列的通用引物从复杂靶序列或靶序列的复杂混合物中随机扩增片段。引物的简并3'部分在DNA上的随机位置处退火,并且可以扩展以生成在其5'端处具有通用序列的靶的副本。

[0005] 可替代地,含有通用引发序列的衔接子可以连接到靶序列的端部。一种或多种衔接子可以用于与靶序列的连接反应。与目前用于通过一个或多个用于通用扩增的衔接子序列的连接制备核酸测序文库的方法相关联的缺点是此类方法所需的时间和费用。

### 发明内容

[0006] 提供了产生核酸文库的方法。所述方法包含将单链核酸结合蛋白结合的单链核酸(SSB结合的ssNA)、衔接子寡核苷酸和夹板寡核苷酸结合,以形成包含所述夹板寡核苷酸与所述SSB结合的ssNA的末端区域并与所述衔接子寡核苷酸杂交的复合物。所述第一衔接子寡核苷酸的端部与所述SSB结合的ssNA的第一末端区域的端部相邻,并且所述方法可以进一步包含共价连接相邻端部。还提供了例如在实践本公开的所述方法中有用的组合物和试剂盒。

### 附图说明

[0007] 图1.根据本公开的一个实施例的产生核酸文库的方法的示意性说明。

[0008] 图2.本公开的示例性方法与用于降解DNA(上图)和现代(modern)DNA(下图)的其它方法的比较。

[0009] 图3.头发DNA长度分布。左图示出了通过圣克鲁斯方法(Santa Cruz method)

(SRL3) 由现代头发DNA产生的模板分子的所观察长度。如对毛干中的DNA所期望的,完整分子的长度通常很短。右图 (SRL4) 中示出了类似的长度分布。

[0010] 图4. SRL3和SRL4中的估计的文库复杂度(独特分子的数量)。

[0011] 图5. 文库复杂度比较。使用本公开的示例方法SS2.0、BEST和叉状衔接子连接制备了测序文库。在三次实验中,使用Preseq(左)或通过qPCR(右)根据几百万个读段对文库复杂度(文库中独特分子的数量)进行了估计。本公开的示例方法将比下一最佳方案SS2.0多2到3倍的提取DNA转化成测序文库。

## 具体实施方式

[0012] 提供了产生核酸文库的方法。所述方法包含将单链核酸结合蛋白结合的单链核酸(SSB结合的ssNA)、衔接子寡核苷酸和夹板寡核苷酸结合,以形成包含所述夹板寡核苷酸与所述SSB结合的ssNA的末端区域并与所述衔接子寡核苷酸杂交的复合物。所述第一衔接子寡核苷酸的端部与所述SSB结合的ssNA的第一末端区域的端部相邻,并且所述方法可以进一步包含共价连接相邻端部。还提供了例如在实践本公开的所述方法中有用的组合物和试剂盒。

[0013] 在更详细地描述本公开的方法、组合物和试剂盒之前,应当理解的是,方法、组合物和试剂盒不限于所描述的特定实施例,因为此类方法、组合物和试剂盒当然可以改变。还应当理解的是,因为方法、组合物和试剂盒的范围将仅由所附权利要求书限定,所以本文中所使用的术语仅是出于描述特定实施例的目的,而不旨在是限制性的。

[0014] 在提供取值范围的情况下,应当理解的是,在所述范围的上限与下限之间的每个插入值(到下限的十分之一单位,除非上下文清楚地另外指明)以及在所陈述范围内的任何其它所陈述的值或插入值均被涵盖在方法、组合物和试剂盒之内。这些更小范围的上限和下限可以独立地被包含在更小范围之内,并且也被涵盖在方法、组合物和试剂盒之内,受制于在所陈述范围内任何明确排除的限制。在所陈述的范围包含限制中的一者或两者的情况下,排除了所包含的限制中的任一者或两者的范围也被包含在方法、组合物和试剂盒之内。

[0015] 本文提供了某些范围,其中数值前面出现术语“约”。术语“约”在本文用于为其后面出现的精确数字以及接近或靠近所述术语后面的数字的数提供文字性支持。在确定数字是否接近或靠近具体叙述的数字时,接近或靠近的未叙述的数字可以为这样的数字,其在出现的上下文中,提供具体叙述的数字的基本等同形式。

[0016] 除非另外定义,否则本文使用的所有技术术语和科学术语均具有与所述方法、组合物和试剂盒所属领域的普通技术人员通常所理解的含义相同的含有。尽管与本文所述的方法、组合物和试剂盒类似或等同的任何方法、组合物和试剂盒也可以用于实践或测试所述方法、组合物和试剂盒,但现在描述代表性的说明性方法、组合物和试剂盒。

[0017] 本说明书中引用的所有出版物和专利均通过引用并入本文,如同每篇单独的出版物或专利被具体和单独地指出通过引用并入本文一样,并且通过引用并入本文以公开和描述与所引用的出版物相关的材料和/或方法。任何出版物的引用均为其在申请日之前的公开内容,并且不应被解释为承认本方法、组合物和试剂盒无权先于此类出版物,因为所提供的出版日期可能与实际出版日期不同,所述实际出版日期可能需要独立确认。

[0018] 应当指出的是,除非上下文另外清楚地指明,否则如本文中以及所附权利要求书

中所使用的,单数形式“一个/种(a/an)”和“所述(the)”包含复数指代物。应当进一步指出的是,权利要求书可以撰写为排除任何可选择的要素。因此,本声明旨在充当对于此类与权利要求要素的叙述相结合的排除性术语(如“单独地”、“仅”等)的使用或“否定型”限制的使用的先行基础。

[0019] 应当理解的是,为清除起见而在单独实施例的上下文中描述的方法、组合物和试剂盒的某些特征还可以以组合形式提供于单个实施例中。相反,为清除起见而在单个实施例的上下文中描述的方法、组合物和试剂盒的各种特征也可以分开提供或以任何合适的子组合形式提供。实施例的所有组合均被明确涵盖在本公开中,并且在本文中公开,如同每种组合被单独和明确地公开一样,只要这种组合包括可操作的过程和/或组合物。另外,描述这些变量的实施例中列出的所有子组合也被明确地涵盖在本方法、组合物和试剂盒中,并且在本文中公开,如同每个此类子组合在本文中单独和明确地公开一样。

[0020] 当阅读了本公开,对本领域技术人员显而易见,本文描述和阐释的每个单个实施例具有分开的组分和特征,其可以容易与任何其它数个实施例的特征分开或结合,而不背离本发明的精神或范围。任何叙述的方法可以以所叙述事件的顺序或以逻辑上可能的任何其它顺序进行。

[0021] 方法

[0022] 如上所概述的,本公开提供了产生核酸文库的方法。所述方法包含使单链核酸(ssNA)与单链核酸结合蛋白(SSB)接触,以产生SSB结合的ssNA。所述方法进一步包含将所述SSB结合的ssNA、第一衔接子寡核苷酸以及第一夹板寡核苷酸组合,所述第一夹板寡核苷酸包含SSB结合的ssNA杂交区域和第一衔接子寡核苷酸杂交区域。所述组合引起包含以下的复合物的形成:所述第一夹板寡核苷酸通过所述SSB结合的ssNA杂交区域与所述SSB结合的ssNA的末端区域杂交;并且所述第一夹板寡核苷酸通过所述第一衔接子寡核苷酸杂交区域与所述第一衔接子寡核苷酸杂交,使得所述第一衔接子寡核苷酸的端部与所述SSB结合的ssNA的第一末端区域的端部相邻。在一些实施例中,所述组合进一步包含将所述SSB结合的ssNA、第二衔接子寡核苷酸以及第二夹板寡核苷酸组合,所述第二夹板寡核苷酸包含SSB结合的ssNA杂交区域和第二衔接子寡核苷酸杂交区域,其中形成的复合物进一步包含:所述第二夹板寡核苷酸通过所述SSB结合的ssNA杂交区域与所述SSB结合的ssNA的所述末端区域杂交的位置与和所述第一夹板寡核苷酸杂交的所述末端区域的位置相反;并且所述第二夹板寡核苷酸通过所述第二衔接子寡核苷酸杂交区域与所述第二衔接子寡核苷酸杂交,使得所述第二衔接子寡核苷酸的端部与所述SSB结合的ssNA的端部相邻的位置与和所述第一衔接子寡核苷酸相邻的所述端部的位置相反。

[0023] 采用第二衔接子寡核苷酸和第二夹板寡核苷酸的示例实施例示意性地示出在图1中。在此实例中,ssNA通过使dsNA变性,由dsNA产生(例如,由dsDNA产生的ssDNA)。在图1的顶部示出了dsNA 102。在dsNA 102变性后,使所得ssNA与单链核酸结合蛋白(SSB)接触,以产生SSB结合的ssNA。图1中示出的是衍生自dsNA 102的链的SSB结合的ssNA 104。在此实例中,ssNA 104与和第一夹板寡核苷酸108杂交的第一衔接子寡核苷酸106以及和第二夹板寡核苷酸112杂交的第二衔接子寡核苷酸110组合。第一夹板寡核苷酸108与第一衔接子寡核苷酸106的杂交是通过第一夹板寡核苷酸108的第一衔接子寡核苷酸杂交区域114。第二夹板寡核苷酸112与第二衔接子寡核苷酸110的杂交是通过第二夹板寡核苷酸112的第二衔接

子寡核苷酸杂交区域116。第一夹板寡核苷酸108和第二夹板寡核苷酸112与SSB结合的NA 104、第一衔接子区域106和第二衔接子区域110的杂交形成复合物122。第一夹板寡核苷酸108与SSB结合的ssNA 104的5'末端区域的杂交是通过第一夹板寡核苷酸108的第一SSB结合的ssNA杂交区域118。第二夹板寡核苷酸112与SSB结合的ssNA 104的3'末端区域的杂交是通过第二夹板寡核苷酸112的第二SSB结合的ssNA杂交区域120。夹板寡核苷酸被设计成,使得当夹板寡核苷酸的SSB结合的ssNA杂交区域与SSB结合的ssNA的其相应末端区域杂交时,衔接子寡核苷酸的端部与SSB结合的ssNA的端部相邻。相邻端部的位置由星号指示。这些相邻端部可以共价连接(例如,通过酶连接)以产生经过衔接的ssNA(例如,图1中示出的经过衔接的ssNA 124),过经衔接的ssNA然后可以在由经过衔接的ssNA的衔接子部分中的一个或多个序列促进的所关注的下游应用(例如,PCR扩增、下一代测序和/或等)中使用。如图1中示出的,在相邻端部共价连接后,可以执行任选清除步骤,以将经过衔接的ssNA与一种或多种试剂或形成的复合物的组分分离,例如,用于共价连接的酶、夹板寡核苷酸、SSB和/或等。此类清除步骤的合适方法包含但不限于固相可逆固定化(SPRI-例如,使用磁珠)和核酸柱纯化。在图1中示出的实例中,在夹板寡核苷酸(黑色矩形)的每个端部处存在阻断修饰,并且在每个衔接子寡核苷酸的与SSB结合的ssNA(黑色矩形)不相邻的端部处进一步存在阻断修饰。阻断修饰防止寡核苷酸和ssNA与那些端部的连接。

[0024] 如上所概述的,与目前用于通过一个或多个衔接子序列的连接制备核酸测序文库的方法相关联的缺点包含此类方法所需的时间和费用。本公开的方法构成了对单链文库制备的现有技术方法的改进,如由Gansauge等人(2017)《核酸研究(Nucleic Acids Research)》45(10):e79描述的方法(被称为“ssDNA2.0”),其中令人惊讶地发现本方法更有效,需要更少时间并且降低成本。现在将进一步详细地描述本公开的方法的各方面。

[0025] 本方法包含使单链核酸(ssNA)与单链核酸结合蛋白(SSB)接触,以产生SSB结合的ssNA。“单链核酸”或“ssNA”意指其长度的70%或以上为单链的(即,分子间或分子内未杂交的)多核苷酸的集合。在一些实施例中,在多核苷酸的长度的75%或以上、80%或以上、85%或以上、90%或以上、95%或以上或99%或以上,ssNA为单链的。在某些方面,在多核苷酸的整个长度上,ssNA为单链的。

[0026] ssNA可以是所关注的任何核酸样品(或可以由任何核酸样品制备),包含但不限于与单个细胞、多个细胞(例如,培养的细胞)、组织、器官或生物体(例如,细菌、酵母等)分离的核酸样品。示例性样品类型包含但不限于血液、血清、唾液、痰、尿液、粪便、呕吐物、粘液、头发、指甲(例如,指甲、趾甲)、拭子(例如,脸颊拭子、咽拭子、阴道拭子)、活检组织(例如,穿刺活检、细针活检、细针抽吸活检)、细胞培养物、环境样品(例如,水、土壤、空气、表面、接触DNA)和宏基因组样品。在某些方面,核酸样品是从动物的单个细胞、细胞的集合、组织、器官和/或等分离的。在一些情况下,核酸样品包括无细胞核酸(例如,无细胞DNA(cfDNA)),如但不限于胎儿无细胞核酸(例如,无细胞胎儿DNA(cffDNA))或循环肿瘤核酸(例如,循环肿瘤DNA(ctDNA))。在一些实施例中,动物是哺乳动物(例如,来自人属的哺乳动物、啮齿动物(例如,小鼠或大鼠)、狗、猫、马、牛或所关注的任何其它哺乳动物)。在其它方面,核酸样品是从除哺乳动物之外的来源,如细菌、酵母、昆虫(例如,果蝇)、两栖动物(例如,青蛙(例如,非洲爪蟾))、病毒、植物、或任何其它非哺乳核酸样品来源分离/获得的。

[0027] 在一些实施例中,ssNA来自经降解核酸样品。如本文所用,“经降解核酸样品”为已

经通过酶促、物理、化学或其它过程片段化的DNA的样品。经降解核酸样本的实例为从骨头残骸、头发、来自血浆中的无细胞DNA或从土壤或水中回收的环境DNA中回收的DNA片段。在某些方面,当ssNA来自经降解核酸样品时,ssNA来自古老(ancient)核酸样品。“古老核酸样品”意指从生物残骸中回收的核酸片段。所关注的古老核酸样品的非限制性实例为从灭绝的生物体或动物,例如,灭绝的哺乳动物获得(例如,分离)的核酸样品。在某些方面,灭绝的哺乳动物来自人属。在一些实施例中,ssNA来自法医核酸样品。如本文所用,“法医核酸样品”为与犯罪调查有关(例如,在其过程期间获得)的核酸样品。

[0028] 在某些方面,ssNA来自肿瘤核酸样品(也就是说,从肿瘤中分离的核酸样品)。如本文所用,“肿瘤”是指所有赘生性细胞生长和增殖,无论是恶性的还是良性的,以及所有癌前和癌性细胞和组织。术语“癌症”和“癌性”是指或描述哺乳动物中通常表征为不受调节的细胞生长/增殖的生理状况。癌症的实例包含但不限于癌、淋巴瘤、母细胞瘤、肉瘤和白血病。此类癌症的更具体的实例包含鳞状细胞癌、小细胞肺癌、非小细胞肺癌、肺腺癌、肺鳞状细胞癌、腹膜癌、肝细胞癌、胃肠癌、胰腺癌、胶质母细胞瘤、宫颈癌、卵巢癌、肝癌、膀胱癌、肝癌、乳腺癌、结肠癌、结直肠癌、子宫内膜癌或子宫癌、唾液腺癌、肾癌、肝癌、前列腺癌、外阴癌、甲状腺癌、肝癌、各种类型的头颈癌等。

[0029] 在一些实施例中,ssNA来自无细胞核酸样品,例如,无细胞DNA、无细胞RNA或两者。在某些方面,无细胞核酸从选自由以下组成的组的体液样品获得:全血、血浆、血清、羊水、唾液、尿液、胸腔积液、支气管灌洗、支气管抽吸物、母乳、初乳、眼泪、精液、腹膜液、胸腔积液和粪便。在一些实施例中,无细胞核酸为无细胞胎儿DNA。在某些方面,无细胞核酸为循环肿瘤DNA。在一些实施例中,无细胞核酸包括传染剂DNA。在一些实施例中,无细胞核酸包括来自移植物的DNA。

[0030] 在某些方面,ssNA为单链脱氧核糖核酸(ssDNA)。所关注的ssDNA包含但不限于衍生自双链DNA(dsDNA)的ssDNA。例如,ssDNA可以衍生自变性(例如,热变性和/或化学变性)以产生ssDNA的双链DNA。在一些实施例中,方法包含在使ssDNA与SSB接触之前,通过使dsDNA变性来产生ssDNA。

[0031] 当ssNA是衍生自dsDNA样品的ssDNA时,方法可以进一步包含在形成复合物之后,使ssDNA再杂交(其现在在一个或两个端部处包含一个或多个衔接子(例如,测序衔接子))以产生dsDNA。如果期望,可以对所产生dsDNA进行测序。在一些实施例中,再杂交在足够严格的杂交条件下执行,以产生类似于ssDNA所衍生的原始dsDNA的dsDNA。足够严格的杂交条件可以包含所选杂交温度、所选盐浓度和/或被选择以产生类似于ssDNA所衍生的原始dsDNA的dsDNA的任何其它方便的杂交参数。此类所产生dsDNA的至少一个子集的一个或两个端部将类似于/复制原始dsDNA的端部(例如,突出端)。使用本公开的方法确定端部/突出端含量(例如,通过测序)可以提供关于ssDNA所衍生的核酸样品的多种有用信息。例如,知道突出端含量在分析例如,来自血浆或另一合适的来源的无细胞DNA(cfDNA)中是有价值的。已经示出,cfDNA衍生自各种来源,包含血细胞、孕妇的胎儿细胞、患有癌症的个体中的肿瘤细胞、来自器官移植受体的移植器官组织等。通过本公开方法的实施例提供的突出端含量可以用于对测序读段进行分类,例如,通过原始来源以用于诊断目的。

[0032] 此外,端部/突出端含量可以用于分析来自法医样品的经混合DNA。例如,来自精液、血液或所关注的其它来源的DNA可以具有对所述来源具有诊断性的端部特性,并且可以

基于此信息对DNA序列进行分区。

[0033] 此外,确定古老DNA样品(例如,来自灭绝的生物体、植物或动物的样品)中的突出端含量提供了用于表征此类样品以及样品所衍生的生物体、植物、动物等的信息。例如,古老DNA样品(例如,来自灭绝的哺乳动物的DNA样品)通常包含污染的DNA(例如,污染细菌DNA等)。在此类情况下,当这些类型的突出端与DNA的特定来源相关联时,可以基于检测到的突出端的类型对所关注的DNA序列与污染的DNA序列进行分区。

[0034] 在某些实施例中,本公开的方法可用于根据突出端的长度和类型确定DNA提取物(例如,古老DNA提取物)中碱基损伤的速率和位置。

[0035] 因此,在一些实施例中,提供了包含以下的方法:将SSB结合的dsDNA衍生的ssDNA与衔接子寡核苷酸和夹板寡核苷酸组合以形成复合物,所述复合物包含与如本文所述的衔接子寡核苷酸和夹板寡核苷酸杂交的SSB结合的dsDNA衍生的ssDNA;以及在复合物形成之后,将ssDNA杂交以产生类似于ssDNA所衍生的原始dsDNA的dsDNA(也就是,现在在一个或两个端部处包含一个或多个衔接子(例如,测序衔接子)的“经过衔接的”dsDNA)。此类方法可以进一步包含对经过衔接的dsDNA进行测序。在某些方面,测序是确定ssDNA所衍生的dsDNA的端部/突出端含量。在涉及测序的本公开的方法的任何实施例中,所述方法可以包含对经过衔接的ssDNA的子样品进行测序,以降低测序期间的复杂度。

[0036] 在一些实施例中,ssNA为单链核糖核酸(ssRNA)。所关注的RNA包含但不限于信使RNA(mRNA)、微小RNA(miRNA)、小干扰RNA(siRNA)、反式作用小干扰RNA(ta-siRNA)、天然小干扰RNA(nat-siRNA)、核糖体RNA(rRNA)、转移RNA(tRNA)、小核仁RNA(snoRNA)、小核RNA(snRNA)、长非编码RNA(lncRNA)、非编码RNA(ncRNA)、转移信使RNA(tmRNA)、前体信使RNA(pre-mRNA)、小Cajal体特异性RNA(scaRNA)、piwi相互作用RNA(piRNA)、内切核糖核酸酶制备的siRNA(esiRNA)、小时序RNA(stRNA)、信号识别RNA、端粒RNA、核酶、或这些RNA类型或亚型的任何组合。在一些实施例中,当ssNA为ssRNA时,ssRNA为mRNA。

[0037] 用于从所关注的来源分离、纯化和/或浓缩DNA和RNA的方法、试剂和试剂盒是本领域中已知的并且可商购获得。例如,用于从所关注的来源分离DNA的试剂盒包含Qiagen公司(Qiagen, Inc.) (马里兰州,德国城(Germantown, Md))的DNeasy®、RNeasy®、QIAamp®、QIAprep®和QIAquick®核酸分离/纯化试剂盒;生命技术公司(Life Technologies, Inc.) (加利福尼亚州卡尔斯巴德(Carlsbad, CA))的DNAzol®、ChargeSwitch®、Purelink®、GeneCatcher®核酸分离/纯化试剂盒;克隆技术实验室公司(Clontech Laboratories, Inc.) (加利福尼亚州,芒廷维尤(Mountain View, CA))的NucleoMag®、NucleoSpin®和NucleoBond®核酸分离/纯化试剂盒。在某些方面,核酸从固定的生物样品,例如福尔马林固定的石蜡包埋的(FFPE)组织分离。可以使用市售试剂盒来分离来自FFPE组织的基因组DNA:如Qiagen公司(马里兰州,德国城)的AllPrep® DNA/RNA FFPE试剂盒、生命技术公司(加利福尼亚州,卡尔斯巴德)的用于FFPE的RecoverAll®全核酸分离试剂盒以及克隆技术实验室公司(加利福尼亚州,芒廷维尤)的NucleoSpin®FFPE试剂盒。

[0038] 当核酸样品所获得(例如,分离)的生物体、植物、动物等灭绝时,用于回收此类核

酸的合适策略是已知的并且包含例如Green等人, (2010)《科学 (Science)》328 (5979) :710-722;Poinar等人, (2006)《科学》311 (5759) :392-394;Stiller等人, 2006,《美国国家科学院院刊 (Proc.Natl.Acad.Sci.)》103 (37) :13578-13584;Miller等人, (2008)《自然 (Nature)》456 (7220) :387-90;Rasmussen等人, (2010)《自然》463 (7282) :757-762;以及其它地方所描述的那些策略。

[0039] 如以上所概述的,本方法包含使ssNA与单链核酸结合蛋白 (SSB) 接触,以产生SSB结合的ssNA。SSB以协作方式与ssNA结合,并且与双链核酸 (dsNA) 不很好地结合。在结合ssDNA后,SSB使螺旋双链体不稳定。在实践本主题方法时可以采用的SSB包含原核SSB (例如,细菌或古细菌SSB) 和真核SSB。在实践本主题方法时可以采用的SSB的非限制性实例包含大肠杆菌 (E.coli) SSB、大肠杆菌RecA、极热稳定单链DNA结合蛋白 (ET SSB)、嗜热栖热菌 (*Thermus thermophilus*) (Tth) RecA、T4基因32蛋白、复制蛋白A (RPA-真核SSB) 等。ET SSB、Tth RecA、大肠杆菌RecA、T4基因32蛋白以及用于使用此类SSB制备SSB结合的ssNA的缓冲液和详细方案可从例如新英格兰生物实验室公司 (New England Biolabs, Inc.) (马萨诸塞州伊普斯维奇 (Ipswich, MA)) 获得。发明人已经确定,给定相等摩尔浓度输入,SSB的较大输入对于具有较高平均片段长度的ssNA是有益的。下面的“实验”章节提供了关于用于使ssNA与SSB接触以产生与SSB结合的ssNA的示例方法的详细指南。

[0040] 如以上所概述的,本主题方法包含将SSB结合的ssNA、衔接子寡核苷酸和夹板寡核苷酸组合以形成复合物,所述夹板寡核苷酸包含SSB结合的ssNA杂交区域和衔接子寡核苷酸杂交区域。如本文所用,“寡核苷酸”为核苷酸的5个到500个核苷酸,例如5个到100个核苷酸的单链多聚体。寡核苷酸可以是合成的或者可以以酶促方式制备,并且在一些实施例中,在长度上为5个到50个核苷酸。寡核苷酸可以含有核糖核苷酸单体 (即,可以为寡核糖核苷酸或“RNA寡核苷酸”)、脱氧核糖核苷酸单体 (即,可以为寡脱氧核糖核苷酸或“DNA寡核苷酸”) 或其组合。例如,寡核苷酸在长度上可以为10个到20个、20个到30个、30个到40个、40个到50个、50个到60个、60个到70个、70个到80个、80个到100个、100个到150个或150个到200个或多达500个核苷酸。

[0041] 本公开的“衔接子寡核苷酸”为包含衔接子或其部分的寡核苷酸。“衔接子”意指可用于一种或多种下游应用 (例如,经过衔接的ssNA或其衍生物的PCR扩增、经过衔接的ssNA或其衍生物的测序和/或等) 的核苷酸序列。在某些方面,存在于衔接子寡核苷酸中的衔接子或其部分为测序衔接子。“测序衔接子”意指一个或多个核酸结构域,其包含如下所关注的测序平台利用的核苷酸序列 (或其补体) 的至少一部分:由Illumina®提供的测序平台 (例如,HiSeq™、MiSeq™和/或Genome Analyzer™测序系统);Oxford Nanopore™技术 (例如,MinION™测序系统)、Ion Torrent™ (例如,Ion PGM™和/或Ion Proton™测序系统);太平洋生物科学 (Pacific Biosciences) (例如,Sequel或PacBio RS II测序系统);Life Technologies™ (例如,SOLiD™测序系统);罗氏 (Roche) (例如,454GS FLX+和/或GS Junior测序系统);或所关注的任何其它测序平台。

[0042] 在某些方面,测序衔接子是或包含选自以下的核酸结构域:与表面附接的测序平台寡核苷酸 (例如,Illumina®测序系统中与流动细胞的表面附接的P5或P7寡核苷酸) 特异性结合的结构域 (例如,“捕获位点”或“捕获序列”);测序引物结合结构域 (例如,

Illumina®平台的读段1或读段2引物可以结合的结构域);唯一标识物(例如,条形码或其它结构域,其唯一地标识寡核苷酸探针的3'区域、探针补体寡核苷酸或两者,和/或通过用具体条形码或“标签”标记来自给定样品的每个分子来唯一地标识被测序以实现样品多路复用的rRNA的样品来源);条形码测序引物结合结构域(与用于对条形码进行测序的引物结合的结构域);分子标识结构域(例如,分子索引标签,如4个、6个或其它数量的核苷酸的随机化标签),其用于唯一地标记所关注的分子,例如,以基于对唯一标签进行测序的实例的数量来确定表达水平;任何此类结构域的补体;或其任何组合。在某些方面,条形码结构域(例如,样品索引标签)和分子标识结构域(例如,分子索引标签)可以包含在相同的核酸中。

[0043] 当衔接子寡核苷酸包含测序衔接子之一或一部分时,可以使用多种方法添加一种或多种另外的测序衔接子和/或测序衔接子的剩余部分。例如,可以通过连接、逆转录、PCR扩增和/或等添加测序衔接子的另外和/或剩余部分。在PCR的情况下,可以采用包含第一扩增引物和第二扩增引物的扩增引物对,所述第一扩增引物包含3'杂交区域(例如,用于与衔接子寡核苷酸的衔接子区域杂交)以及包含测序衔接子的另外和/或剩余部分的5'区域,所述第二扩增引物包含3'杂交区域(例如,用于与添加到ssNA分子的相反端部的第二衔接子寡核苷酸的衔接子区域杂交)以及任选地包含测序衔接子的另外和/或剩余部分的5'区域。

[0044] 本公开的“夹板寡核苷酸”为包含SSB结合的ssNA杂交区域和衔接子寡核苷酸杂交区域的寡核苷酸。SSB结合的ssNA杂交区域为与SSB结合的ssNA的末端区域杂交的区域(核苷酸序列)。衔接子寡核苷酸杂交区域为与衔接子寡核苷酸的全部或部分杂交的区域(核苷酸序列)。夹板寡核苷酸被设计成用于同时与SSB结合的ssNA和衔接子寡核苷酸杂交,使得在复合物形成时,衔接子寡核苷酸的端部与SSB结合的ssNA的末端区域的端部相邻。

[0045] 夹板寡核苷酸的SSB结合的ssNA杂交区域可以具有任何合适的长度和序列。在一些实施例中,SSB结合的ssNA杂交区域的长度为10个核苷酸或更少。在某些方面,SSB结合的ssNA杂交区域在长度上为4个到20个核苷酸,例如,在长度上5个到15个、5个到10个、5个到9个、5个到8个或5个到7个(例如,6个或7个)核苷酸。在一些实施例中,SSB结合的ssNA杂交区域包含随机核苷酸序列(例如,由其组成),使得当采用具有各种随机SSB结合的ssNA杂交区域的多个异源夹板寡核苷酸时,集合能够充当用于SSB结合的ssNA的异源群体的夹板寡核苷酸,而无论SSB结合的ssNA的末端区域的序列如何。

[0046] 因此,在某些方面,方法包含通过将SSB结合的ssNA、衔接子寡核苷酸和多个异源夹板寡核苷酸组合来形成复合物,所述多个异源夹板寡核苷酸具有能够充当用于具有未确定序列的末端区域的SSB结合的ssNA的异源群体的夹板寡核苷酸的各种随机SSB结合的ssNA杂交区域。

[0047] 在一些实施例中,SSB结合的ssNA杂交区域包含被设计成与已知序列的SSB结合的ssNA末端区域杂交的已知序列。在某些方面,采用了具有已知序列的不同SSB结合的ssNA杂交区域的两个或更多个异源夹板寡核苷酸,所述已知序列被设计成与已知序列的相应SSB结合的ssNA末端区域杂交。其中SSB结合的ssNA杂交区域具有已知序列的实施例用于例如,当期望从具有已知序列的末端区域的SSB结合的ssNA的仅一个子集产生核酸文库时。因此,在某些方面,方法包含通过将SSB结合的ssNA、衔接子寡核苷酸和一个或多个异源夹板寡核苷酸组合来形成复合物,所述一个或多个异源夹板寡核苷酸具有能够充当用于具有已知序列的一个或多个末端区域的一个或多个SSB结合的ssNA的夹板寡核苷酸的已知序列的一个

或多个不同的SSB结合的ssNA杂交区域。

[0048] 在某些方面,SSB结合的ssNA杂交区域包含一个或多个通用碱基。如本文所用,“通用碱基”是能够与四个标准核苷酸碱基:A、C、G和T中的每一个不区别地碱基配对的碱基。可以掺入到SSB结合的ssNA杂交区域中的通用碱基包含但不限于2'-脱氧肌苷(dI、dInosine)和5-硝基吡啶。

[0049] SSB结合的ssNA、衔接子寡核苷酸和夹板寡核苷酸组合的方式可以变化。在一些实施例中,组合包含将包含夹板寡核苷酸通过衔接子寡核苷酸杂交区域与衔接子寡核苷酸杂交的复合物与SSB结合的ssNA组合。在其它方面中,组合包含将包含夹板寡核苷酸通过SSB结合的ssNA杂交区域与SSB结合的ssNA杂交的复合物与衔接子寡核苷酸组合。在仍其它方面,组合包含将SSB结合的ssNA、衔接子寡核苷酸和夹板寡核苷酸组合,其中在组合之前,三种组分中的组分均未与另一组分预复合(也就是,与其杂交)。

[0050] 组合在杂交条件下执行,使得复合物包含夹板寡核苷酸通过SSB结合的ssNA杂交区域与SSB结合的ssNA的末端区域杂交,并且夹板寡核苷酸通过衔接子寡核苷酸杂交区域与衔接子寡核苷酸杂交。特异性杂交是否发生由夹板寡核苷酸的相关(也就是,杂交)区域之间的互补程度、SSB结合的ssNA的末端区域和衔接子寡核苷酸及其长度、盐浓度和杂交发生的温度等因素确定,所述温度可以通过相关区域的溶解温度( $T_m$ )告知。溶解温度是指相关区域的一半保持杂交并且相关区域的一半解离成单链的温度。双链体的 $T_m$ 可以以实验方式确定或使用以下公式预测: $T_m = 81.5 + 16.6 (\log_{10} [Na^+]) + 0.41 (\text{分数G+C}) - (60/N)$ ,其中N为链长度并且 $[Na^+]$ 小于1M。参见Sambrook和Russell (2001;《分子克隆:实验室手册(Molecular Cloning: A Laboratory Manual)》,第3版,冷泉港出版社(Cold Spring Harbor Press),纽约冷泉港(Cold Spring Harbor N.Y.),第10章)。依赖于各种参数的其它更先进的模型也可以用于预测相关区域的 $T_m$ ,这取决于各种杂交条件。用于实现特异性核酸杂交的方法可以在例如,Tijssen,《生物化学和分子生物学中的实验室技术-与核酸探针的杂交(Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes)》,第I部分,第2章,“杂交原理和核酸探针测定策略概述(Overview of principles of hybridization and the strategy of nucleic acid probe assays)”,Elsevier(1993)中找到。

[0051] 如本文所用的术语“互补的”或“互补性”是指通过与靶核酸的区域的非共价键连接基配对的核苷酸序列,例如,与SSB结合的ssNA的末端区域杂交的SSB结合的ssNA杂交区域的核苷酸序列,以及与探针补体寡核苷酸杂交的衔接子寡核苷酸杂交区域的核苷酸序列。在经典的Watson-Crick碱基配对中,腺嘌呤(A)与胸腺嘧啶(T)形成碱基对,和鸟嘌呤(G)在DNA中与胞嘧啶(C)形成碱基对一样。在RNA中,胸腺嘧啶被尿嘧啶(U)替代。因此,A与T互补,并且G与C互补。在RNA中,A与U互补,并且反之亦然。通常,“互补的”或“互补性”是指至少部分互补的核苷酸序列。这些术语还可以涵盖完全互补的双链体,使得一条链中的每个核苷酸与对应位置中另一条链中的每个核苷酸互补。在某些情况下,核苷酸序列可以与靶标部分互补,其中并非所有核苷酸在所有对应位置都与靶核酸中的每个核苷酸互补。例如,SSB结合的ssNA杂交区域可以与SSB结合的ssNA的末端区域完美(即,100%)互补,或者SSB结合的ssNA杂交区域可以共享某种程度的不太完美的互补性(例如,70%、75%、85%、90%、95%、99%)。可以通过出于最佳比较的目的比对序列来确定两个核苷酸序列的同一

性百分比(例如,可以在第一序列的序列中引入空位以进行最佳比对)。然后比较对应位置处的核苷酸,并且两个序列之间的同一性百分比是序列共有的相同位置的数量的函数(即, $\% \text{同一性} = \text{相同位置的}\# / \text{位置的总}\# \times 100$ )。当一个序列中的位置被与另一个序列中的对应位置相同的核苷酸占据时,则所述位置处的分子是相同的。此数学算法的非限制性实例描述于Karlin等人,《美国国家科学院院刊》90:5873-5877(1993)中。此算法并入到NBLAST和XBLAST程序(版本2.0)中,如Altschul等人,《核酸研究(Nucleic Acids Res.)》25:389-3402(1997)中所描述的。当利用BLAST和Gapped BLAST程序时,可以使用相应程序(例如,NBLAST)的默认参数。在一个方面,用于序列比较的参数可以被设置为评分=100、字长=12,或者可以改变(例如,字长=5或字长=20)。

[0052] 复合物形成为使得衔接子寡核苷酸的端部与SSB结合的ssNA的末端区域的端部相邻。“与…相邻”意指衔接子寡核苷酸的端部处的末端核苷酸和SSB结合的ssNA的末端区域的末端核苷酸端部彼此足够接近,使得末端核苷酸可以共价连接,例如,通过化学连接、酶连接等。在一些实施例中,端部借助于衔接子寡核苷酸的端部处的末端核苷酸和SSB结合的ssNA的末端区域的末端核苷酸端部与夹板寡核苷酸的相邻核苷酸杂交,而彼此相邻。夹板寡核苷酸可以被设计成确保衔接子寡核苷酸的端部与SSB结合的ssNA的末端区域的端部相邻。此类夹板寡核苷酸的非限制性实例在本文的实验章节中提供。

[0053] 本文所述的方法中的任何方法可进一步包含共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部。共价连接可以包含连接相邻端部。连接相邻端部可以使用任何合适的方法执行。在某些方面,连接是通过化学连接。在其它方面,连接是通过酶连接。用于执行酶连接反应的合适的试剂(例如,连接酶和对应缓冲液等)和试剂盒是已知的且是可获得的,例如,可从新英格兰生物实验室公司(马萨诸塞州伊普斯维奇)获得的瞬时粘性端连接酶主混合物(Instant Sticky-end Ligase Master Mix)。可以采用的连接酶包含例如T4 DNA连接酶(例如,处于低浓度或高浓度)、T4 DNA连接酶、T7 DNA连接酶、大肠杆菌DNA连接酶、Electro Ligase®等。适合于执行连接反应的条件将根据所使用的连接酶的类型而改变。关于此类条件的信息易于获得。必要时,可以使用例如合适的激酶,如T4多核苷酸激酶(PNK),在衔接子寡核苷酸的5'端或SSB结合的ssNA处添加磷酸基团。此类激酶和用于使用此类激酶对5'端进行磷酸化的指南可从例如新英格兰生物实验室公司(马萨诸塞州伊普斯维奇)获得。

[0054] 在一些实施例中,所述夹板寡核苷酸、所述衔接子寡核苷酸或两者均包含阻断修饰。例如,夹板寡核苷酸的一个或两个端部可以包含阻断修饰,和/或衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部可以包含阻断修饰。“阻断修饰”意指端部不能够使用用于共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部的方法与任何其它寡核苷酸组分的端部连接。在某些方面,阻断修饰是连接阻断修饰。可以包含在夹板寡核苷酸的一个或两个端部和/或衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部处的阻断修饰的实例包含在衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部处不存在3'OH,以及在衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部处的3'OH不可及。其中端部具有不可及的3'OH的阻断修饰的非限制性实例包含:氨基修饰剂、间隔子、双脱氧碱基、反向的双脱氧碱基和3'磷酸等。

[0055] 在某些方面,夹板寡核苷酸、衔接子寡核苷酸或两者均包含一个或多个非天然核苷酸(其也可以被称为核苷酸类似物)。可以包含在夹板寡核苷酸、衔接子寡核苷酸或两者

中的非天然核苷酸的非限制性实例为LNA(锁核酸)、PNA(肽核酸)、FANA(2'-脱氧-2'-氟阿拉伯糖核苷酸)、GNA(乙二醇核酸)、TNA(苏糖核酸)、2'-O-Me RNA、2'-氟RNA、吗啉代核苷酸和其任意组合。

[0056] 共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部会产生经过衔接的ssNA,其中“经过衔接的”意指ssNA现在包含一个或多个衔接子序列或其子区域。经过衔接的ssNA可以在作为所关注的下游应用的输入使用之前被纯化。例如,可以使复合物变性(例如,热变性)以将经过衔接的ssNA与夹板寡核苷酸分离,经过衔接的ssNA可以从SSB和/或在接触和/或组合步骤(例如,通过固相可逆固定化(SPRI)、柱纯化和/或等)期间存在的任何其它组分或其组合中纯化。

[0057] 在一些实施例中,一个或多个衔接子序列或其子区域是一个或多个测序衔接子或其子区域,并且方法进一步包含对经过衔接的ssNA或其任何衍生物的至少一部分(例如,使用经过衔接的ssNA作为模板通过PCR扩增产生的扩增子)进行测序。测序可以在任何合适的测序平台上执行,包含高通量测序(HTS)(或“下一代测序(NGS)”)平台等。所关注的HTS/NGS测序平台包含但不限于Illumina®提供的测序平台(例如,HiSeq™、MiSeq™和/或Genome Analyzer™测序系统);Oxford Nanopore™技术(例如,MinION™、GridIONx5™、PromethION™或SmidgION™基于纳米孔的测序系统)、Ion Torrent™(例如,Ion PGM™和/或Ion Proton™测序系统);太平洋生物科学(例如,Sequel或PacBio RS II测序系统);Life Technologies™(例如,SOLiD测序系统);罗氏(例如,454GS FLX+和/或GS Junior测序系统);或所关注的任何其它测序平台。用于直接测序(例如,通过基于纳米孔的测序)或制备用于在特定平台上进行测序(例如,通过扩增,例如,固相扩增等)的相容性核酸分子、对相容性分子进行测序以及分析测序数据的详细方案可从所关注的测序平台的制造商获得。

[0058] 如上所概述的,本公开的方法构成了对单链文库制备的现有技术方法的改进,如由Gansauge等人(2017)《核酸研究》45(10):e79描述的方法(被称为“ssDNA2.0”),其中令人惊讶地发现本方法更有效,需要更少时间并且降低成本。在一些实施例中,当方法包含共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部时,组合和共价连接步骤的总持续时间为4小时或更少、3小时或更少、2小时或更少或1小时或更少。在某些实施例中,当方法包含共价连接衔接子寡核苷酸和SSB结合的ssNA的相邻端部时,接触、组合和共价连接步骤的总持续时间为4小时或更少、3小时或更少、2小时或更少或1小时或更少。在一些实施例中,方法的效率为使得复合物由70%或更多、75%或更多、80%或更多、85%或更多、90%或更多、95%或更多或99%或更多的ssNA在接触步骤期间与SSB接触而形成。

[0059] 组合物

[0060] 如上文所概述的,本公开还提供了组合物。组合物可用于多种应用中,包含例如实践本公开的方法中的任何方法,包含执行以上在本公开的“方法”章节中描述的步骤中的任何步骤中的一个或多个步骤。如此,组合物可以包含以上在本公开的方法章节中描述的呈任何组合形式的寡核苷酸(包含多个异源寡核苷酸/其集合)、ssNA、SSB、其它试剂等。

[0061] 在某些方面,提供了包含SSB结合的ssNA、第一衔接子寡核苷酸和第一夹板寡核苷酸的组合物,所述第一夹板寡核苷酸包含SSB结合的ssNA杂交区域和第一衔接子寡核苷酸杂交区域。此类组合物可以进一步包含第二衔接子寡核苷酸和第二夹板寡核苷酸,所述第二夹板寡核苷酸包含SSB结合的ssNA杂交区域和第二衔接子寡核苷酸杂交区域。

[0062] 在某些方面,提供了包含复合物的组合物,所述复合物包含夹板寡核苷酸通过衔接子寡核苷酸杂交区域与衔接子寡核苷酸杂交,以杂交的复合物的形式存在(例如,在不存在于SSB结合的ssNA的情况下)。在其它方面中,提供了包含复合物的组合物,所述复合物包含夹板寡核苷酸通过SSB结合的ssNA杂交区域与SSB结合的ssNA杂交,以杂交的复合物的形式存在。

[0063] ssNA可以为ssDNA。当ssNA为ssDNA时,ssDNA可以衍生自dsDNA。在一些实施例中,ssNA为ssRNA。在一些实施例中,ssNA来自经降解核酸样品。在某些方面,当ssNA来自经降解核酸样品时,ssNA来自古老核酸样品,如从灭绝的生物体或动物,例如灭绝的哺乳动物获得(例如,分离)的核酸样品。在某些方面,灭绝的哺乳动物来自人属。在一些实施例中,ssNA来自法医核酸样品。

[0064] 本公开的组合物可以进一步包含用于将衔接子寡核苷酸端部与SSB结合的ssNA的端部共价连接的试剂。在一些实施例中,试剂是化学连接试剂或酶连接试剂,例如,连接酶。

[0065] 本公开的组合物可以包含存在于容器中的一种或多种组分。合适的容器包含但不限于管、小瓶和板(例如,96孔或其它孔板)。

[0066] 在某些方面,组合物包含液体介质中的一种或多种组分。液体介质可以是水性液体介质,如水、缓冲溶液等。一种或多种添加剂,如盐(例如NaCl、MgCl<sub>2</sub>、KCl、MgSO<sub>4</sub>)、缓冲剂(Tris缓冲剂、N-(2-羟乙基)哌嗪-N'-(2-乙磺酸)(HEPES)、2-(N-吗啉代)乙磺酸(MES)、2-(N-吗啉代)乙磺酸钠盐(MES)、3-(N-吗啉代)丙磺酸(MOPS)、N-三[羟基甲基]甲基-3-氨基丙磺酸(TAPS)等)、增溶剂、洗涤剂(例如,非离子型洗涤剂,如吐温(Tween)-20等)、核酸酶抑制剂、甘油、螯合剂等可以存在于此类组合物中。

[0067] 试剂盒

[0068] 如上文所概述的,本公开提供了试剂盒。试剂盒可用于多种应用中,包含例如实践本公开的方法中的任何方法,包含执行以上在本公开的方法章节中描述的步骤中的任何步骤中的一个或多个步骤。如此,试剂盒可以包含以上在本公开的方法章节中描述的呈任何组合形式的寡核苷酸(包含多个异源寡核苷酸/其集合)、ssNA、SSB、其它试剂等。

[0069] 在一些实施例中,本公开的试剂盒包含:单链核酸结合蛋白(SSB,例如,单链DNA结合蛋白、单链RNA结合蛋白或两者);第一衔接子寡核苷酸;第一夹板寡核苷酸,所述第一夹板寡核苷酸包括SSB结合的ssNA杂交区域和第一衔接子寡核苷酸杂交区域;以及说明书,所述说明书用于使用SSB、第一衔接子寡核苷酸和第一夹板寡核苷酸来产生核酸文库。在某些方面,此试剂盒进一步包含第二衔接子寡核苷酸和第二夹板寡核苷酸,所述第二夹板寡核苷酸包含SSB结合的ssNA杂交区域和第二衔接子寡核苷酸杂交区域,其中说明书用于使用SSB、第一衔接子寡核苷酸、第一夹板寡核苷酸、第二衔接子寡核苷酸和第二夹板寡核苷酸来产生核酸文库。

[0070] 本公开的试剂盒可以进一步包含用于将衔接子寡核苷酸端部与SSB结合的ssNA的端部共价连接的试剂。在一些实施例中,试剂是化学连接试剂或酶连接试剂,例如,连接酶。

[0071] 在一些实施例中,夹板寡核苷酸、衔接子寡核苷酸或两者均包含阻断修饰。例如,夹板寡核苷酸的一个或两个端部可以包含阻断修饰,和/或衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部可以包含阻断修饰。在某些方面,阻断修饰是连接阻断修饰。可以包含在夹板寡核苷酸的一个或两个端部和/或衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部

处的阻断修饰的实例包含在衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部处不存在3' OH,以及在衔接子寡核苷酸的与SSB结合的ssNA不相邻的端部处的3' OH不可及。其中端部具有不可及的3' OH的阻断修饰的非限制性实例包含:氨基修饰剂、间隔子、双脱氧碱基、反向的双脱氧碱基和3'磷酸等。

[0072] 在一些实施例中,本公开的试剂盒中提供的一个或多个夹板寡核苷酸包含SSB结合的ssNA杂交区域,所述SSB结合的ssNA杂交区域包含随机核苷酸序列(例如,由其组成),使得当试剂盒包含具有各种随机SSB结合的ssNA杂交区域的多个异源夹板寡核苷酸时,集合能够充当用于SSB结合的ssNA的异源群体的夹板寡核苷酸,而无论所关注的SSB结合的ssNA的末端区域的序列如何。

[0073] 在某些方面,本公开的试剂盒中提供的夹板寡核苷酸包含包括一个或多个通用碱基的SSB结合的ssNA杂交区域。可以掺入到SSB结合的ssNA杂交区域中的通用碱基包含但不限于2'-脱氧肌苷(dI、dInosine)和5-硝基吡啶。

[0074] 在一些实施例中,本公开的试剂盒中提供的夹板寡核苷酸的SSB结合的ssNA杂交区域的长度为10个核苷酸或更少。在某些方面,SSB结合的ssNA杂交区域在长度上为4个到20个核苷酸,例如,在长度上5个到15个、5个到10个、5个到9个、5个到8个或5个到7个(例如,6个或7个)核苷酸。

[0075] 主题试剂盒的组分可以存在于单独容器中,或者多个组分可以存在于单个容器中。合适的容器包含单独的管(例如,小瓶)、板(例如,96孔板、384孔板等)的一个或多个孔等。

[0076] 用于使用SSB、一个或多个衔接子寡核苷酸和一个或多个夹板寡核苷酸来产生核酸文库的说明书可以记录在合适的记录介质上。例如,说明书可以印刷在如纸或塑料等基材上。因此,说明书可以以包装插页形式存在于试剂盒中、存在于试剂盒的容器或其组分的标签中(即与包装或次包装关联)等。在其它实施例中,说明书以存在于合适的计算机可读存储介质(例如,便携式闪存驱动器、DVD、CD-ROM,软盘等)上的电子存储数据文件形式存在。在又其它实施例中,试剂盒中不存在实际说明书,但是提供了用于从远程源,例如通过因特网,获得说明书的手段。此实施例的实例是包含网址的试剂盒,可以在所述网址中查看说明书和/或可以从所述网址下载说明书。与说明书一样,用于获得说明书的手段记录在合适的基材上。

[0077] 以下实例以说明性方式而不是以限制性方式提供。

[0078] 实验

[0079] 实例1

[0080] 本文公开了一种用于在一个反应中将衔接子与单链DNA快速、高效且靶向连接的方法。在以下实例中,将含有Illumina P7或P5衔接子序列的测序衔接子寡核苷酸与针对P7夹板的含有一连串3'N并且针对P5夹板含有一连串5'N的夹板寡核苷酸杂交。夹板为可以仅高效执行双链连接以将衔接子连接到靶单链DNA的连接酶创造了机会。不需要参与连接的所有寡核苷酸DNA端部均被阻止连接的寡核苷酸修饰(例如,氨基修饰)阻断。

[0081] 将单链结合蛋白(SSB)添加到测定中提高了反应的效率。使用靶DNA的浓度和长度来计算适当量的SSB,以实现最佳连接效率。SSB可以防止单链DNA重新退火,同时防止二级结构。

[0082] 此方法可以与Gansauge等人(2017)《核酸研究》45(10):e79中描述的被称为SS2.0的单链文库制备进行有利地比较。与SS2.0相比,本方法需要显著更小的时间,并且展示出DNA转化为可以测序的合适衔接子连接的DNA分子的效率显著提高。另外,与SS2.0相比,本方法降低了试剂成本。

[0083] 衔接子和夹板寡核苷酸被工程化为在不应参与适当衔接子连接的所有端部上执行连接阻断修饰。这包含阻塞P5衔接子的5'端、P7衔接子的3'端和夹板的所有端部。这些连接阻断修饰可以是氨基修饰剂、碳间隔子、双脱氧碱基或阻止连接酶接近3'端的3'羟基或5'端的5'磷酸的任何其它合适的修饰。寡核苷酸可以例如通过整合DNA技术(IDT)合成。示例寡核苷酸示出在以下表1中。

[0084] 表1--示例寡核苷酸

[0085]	P5 衔接子 (5' → 3') SEQ ID NO: 1	/5AmMC12/ACACTCTTTCCCTACACGACGCTCTTCCGATCT
	P5 夹板 (5' → 3') SEQ ID NO: 2	/5AmMC6/NNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/ 3AmMO/
	P7 衔接子 (5' → 3') SEQ ID NO: 3	/5Phos/AGATCGGAAGAGCACACGTCTGAACTCCAGTCA/3ddC/
	P7 夹板 (5' → 3') SEQ ID NO: 4	/5AmMC12/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNN /3AmMO/

[0086] /5AmMC12/=5'氨基修饰剂C12

[0087] /5AmMC6/=5'氨基修饰剂C6

[0088] /3AmMO/=3'氨基修饰剂

[0089] /5Phos/=5'磷酸

[0090] /3ddC/=3'双脱氧胞嘧啶

[0091] 在此实例中采用的方案中,将模板DNA与SSB组合并进行热变性。在变性之后,将反应物放置在冰上或处于4°C的PCR冷却器上。在冷却之后,向每个反应物中添加衔接子。然后,添加反应主混合物,之后进行混合。37°C下的温育允许连接立即开始,其中大多数连接发生在45分钟之前。可以利用确立的方法清除反应,并且下游应用,如扩增和测序保持不变。

[0092] 为了制备衔接子,将1X最终浓度的T4 RNA连接酶缓冲液(目录号B0216L)与P5衔接子组合至最终浓度为10uM,并且与P5夹板寡核苷酸组合至最终浓度为20uM。使用P7衔接子寡核苷酸和P7夹板寡核苷酸制备相似的单独混合物。通过加热到95°C持续10秒,并且然后以0.1°C/s的速率降斜到10°C对衔接子进行杂交。

[0093] 以下提供了示例方案。

[0094] 1. 样品输入(36uL)

[0095] a. 将剪切的DNA和ET SSB(目录号M2401S)组合,达到36uL的体积

[0096] i. 1uL的ET SSB可促进连接,而不会抑制所测试的大多数样品类型

[0097] ii. 用缓冲液EBT(10mM Tris-HCl,pH 8.0和0.05%吐温20)填充剩余体积

[0098] 2. 使样品变形

- [0099] a. 在具有预加热到95℃的盖的热循环仪中, 孵育样品, 持续3分钟。
- [0100] b. 将所变性样品立即放置在冰或PCR冷却器上, 持续30秒。
- [0101] 3. 添加2uL池化的衔接子混合物 (相等体积的P5和P7衔接子)
- [0102] a. 衔接子输入将取决于输入的摩尔浓度。衔接子与模板的6到10:1的摩尔比率是优选的。
- [0103] 4. 添加反应主混合物并通过移液彻底混合
- [0104] a. 8uL的T4 DNA连接酶缓冲液 (目录#M0202M)
- [0105] b. 32uL的50%PEG 8000 (目录号B0216L)
- [0106] c. 1uL的T4多核苷酸激酶-10,000U/mL (目录号M0201L)
- [0107] d. 1uL的T4 DNA连接酶-2,000,000U/mL (目录号M0202M)
- [0108] 5. 在37℃下温育, 持续多达60分钟
- [0109] a. 大多数连接发生在前15分钟, 但直到45分钟左右才达到平稳。
- [0110] 6. 清除反应
- [0111] a. 柱清除 (例如, 用于降解DNA) 或SPRI (例如, 用于现代样品)。
- [0112] 比较和优化测定
- [0113] 在清除之后, 对经过预扩增的文库的稀释液执行qPCR, 以确定连接效率。较低CT值指示相对于同一运行中具有较高CT值的另一样品的较高连接效率。其值的差异大约等于文库效率的两倍差异。还用指数PCR反应 (index PCR reaction) 对经过预扩增的文库的等分试样进行扩增。后加索引。用SPRI对文库进行清理并使其在Agilent TapeStation 2200系统上可视化, 以估计每个文库中衔接子伪像的比例。
- [0114] 单链结合蛋白
- [0115] 观察到NEB所提供的如ET SSB蛋白等单链结合蛋白 (SSB) 增强了单链连接的连接效率。给定相等摩尔浓度输入, 具有较高平均片段长度的样品需要更大SSB输入以实现峰连接。反应中DNA的摩尔浓度也会影响所需的SSB的量。在大量过量ET中, SSB具有抑制连接的潜力。
- [0116] 方案比较
- [0117] 将本方案的效率与NEB Ultra 2试剂盒 (dsDNA)、SS2.0 (Gansauge等人 (2017)《核酸研究》45(10):e79) 和以下中描述的平端单管 (BEST) 方法进行了比较: Caroe等人 (2017)《生态学和进化方法 (Methods in Ecology and Evolution)》9(2):410-419; 以及Mak等人 (2017)《GigaScience》6:1-13。比较结果是使用平均片段长度为约350bp的现代人类DNA和严重降解的平均片段长度为约35-40bp的古老野牛样品获得的。
- [0118] NEB Ultra 2试剂盒被识别为针对现代样品的高效文库制备方法, 而SS2.0被识别为针对降解样品的高效文库制备方法。BEST方案涉及使用T4 DNA聚合酶和T4 PNK的平端修复, 以对DNA (无拖尾) 进行平端化并且对5'端进行磷酸化。接下来, 使用T4 DNA连接酶将平端dsDNA衔接子连接到平端, 之后使用Bst 2.0 Warmstart聚合酶进行填充反应, 并使用SPRI珠或柱进行清除。
- [0119] 比较结果在图2中提供。在上图上从左到右为: 此实例中描述的方法 (星号)、ss2.0和BEST。在下图上从左到右为: 此实例中描述的方法 (星号)、NEB Ultra 2试剂盒、ss2.0和BEST。此实例中描述的方法展示出, 与ss2.0相比, 对古老样品的连接效率更高。对于现代样

品,此实例中描述的方法介于0.3个与0.5个qPCR循环之间,在NEB Ultra 2试剂盒之后。

#### [0120] 实例2-头发DNA

[0121] 在高温下使用标准蛋白酶K处理从头发中收集了DNA。使用6纳克的DNA作为用于制备文库的模板。遵循如上所述的方案。由衔接子连接的产物产生两个测序文库。一个使用了50 $\mu$ L总连接产物中的1 $\mu$ L (SRL3)。另一个使用了此产物中的2.5 $\mu$ L (SRL4)。均在Illumina MiSeq测序平台上对两个文库进行测序,以评估文库特性和复杂度(独特文库分子的数量)。

[0122] 在2 $\times$ 75配对端测序后,使用SeqPrep程序将彼此重叠的正向和反向读段对进行组合。这发生在原始DNA模板短到足够使得正向读段和反向读段覆盖相同序列中的一些序列时(在本文中被称为“合并的读段”)时。在合并之后,将合并和未合并的读段图谱定位到参考人类基因组序列中。图3中示出的是SRL3和SRL4文库两者的合并的且图谱定位的,合并的且未图谱定位的以及合并的且未图谱定位的读段的观察到的原始模板长度分布。注意,对于未合并的且未图谱定位的,不可能推断模板DNA的长度。

[0123] 使用Preseq软件程序来估计两个文库中唯独特文库分子的数量。此程序对观察到的重复分子的数量进行计数,以根据如此处产生的观察到的读段的大样品对核酸文库的复杂度进行建模。此程序示出了对被预测为在文库测序的各个深度处独特的观察到的读段的分数的估计。如图4中示出的,两个文库都被预测为具有超过250,000,000个独特分子的复杂度。由2.5 $\mu$ L的衔接子连接的模板制成的SRL4比SRL3具有更多独特分子。

#### [0124] 实例3-来自野牛骨骼的古老DNA

[0125] 使用从古老野牛骨骼中提取的DNA对从模板DNA分子转化为测序文库的效率进行了比较。使用四种不同方案,包含以上描述的方案),从相同量的来自相同提取物的DNA中生成了文库。

[0126] 使用两种方法测量了文库的复杂度:衔接子连接的产物的qPCR和直接测序。qPCR贵是一式三份进行的。图5中示出的两种方法展示了本文所述的方法在将DNA转化为测序文库时更高效。

[0127] 因此,前述内容仅展示了本公开的原理。应当理解的是,本领域技术人员能够设计各种布置,尽管未在本文明确描述或示出,但所述各种布置体现了本发明的原理并且包含在本发明的精神和范围内。另外,在此叙述的所有实例和条件性语言主要打算帮助读者理解诸位发明人所贡献的本发明的原理和概念以推动本领域发展,并且将被视为而不限于这些特别叙述的实例和条件。此外,在此叙述本发明的原理、方面和实施例以及其特定实例的所有陈述打算涵盖其结构和功能等效物两者。另外,预期此类等效物包含当前已知的等效物以及有朝一日开发的等效物两者,即不论结构而执行相同功能的发展的任何要素。因此,本发明的范围不旨在限于本文中示出和描述的示例性实施例。

- [0001] 序列表
- [0002] <110> 加利福尼亚大学董事会董事(The Regents of the University of  
[0003] California)
- [0004] R·格林
- [0005] J·卡普
- [0006] <120> 产生核酸文库的方法以及用于实践所述方法的组合物和试剂盒
- [0007] <130> UCSC-376W0
- [0008] <150> 62/681,524
- [0009] <151> 2018-06-06
- [0010] <160> 4
- [0011] <170> PatentIn版本3.5
- [0012] <210> 1
- [0013] <211> 33
- [0014] <212> DNA
- [0015] <213> 人工序列(Artificial Sequence)
- [0016] <220>
- [0017] <223> 合成序列
- [0018] <400> 1
- [0019] acactctttc cctacacgac gctcttccga tct 33
- [0020] <210> 2
- [0021] <211> 40
- [0022] <212> DNA
- [0023] <213> 人工序列(Artificial Sequence)
- [0024] <220>
- [0025] <223> 合成序列
- [0026] <220>
- [0027] <221> misc\_feature
- [0028] <222> (1) .. (7)
- [0029] <223> n为a、c、g或t
- [0030] <400> 2
- [0031] nnnnnnaga tcggaagagc gtcgtgtagg gaaagagtgt 40
- [0032] <210> 3
- [0033] <211> 34
- [0034] <212> DNA
- [0035] <213> 人工序列(Artificial Sequence)
- [0036] <220>
- [0037] <223> 合成序列
- [0038] <400> 3

- 
- [0039] agatcggaag agcacacgtc tgaactccag tcac 34  
[0040] <210> 4  
[0041] <211> 41  
[0042] <212> DNA  
[0043] <213> 人工序列(Artificial Sequence)  
[0044] <220>  
[0045] <223> 合成序列  
[0046] <220>  
[0047] <221> misc\_feature  
[0048] <222> (35) .. (41)  
[0049] <223> n为a、c、g或t  
[0050] <400> 4  
[0051] gtgactggag ttcagacgtg tgctcttccg atctnnnnn n 41

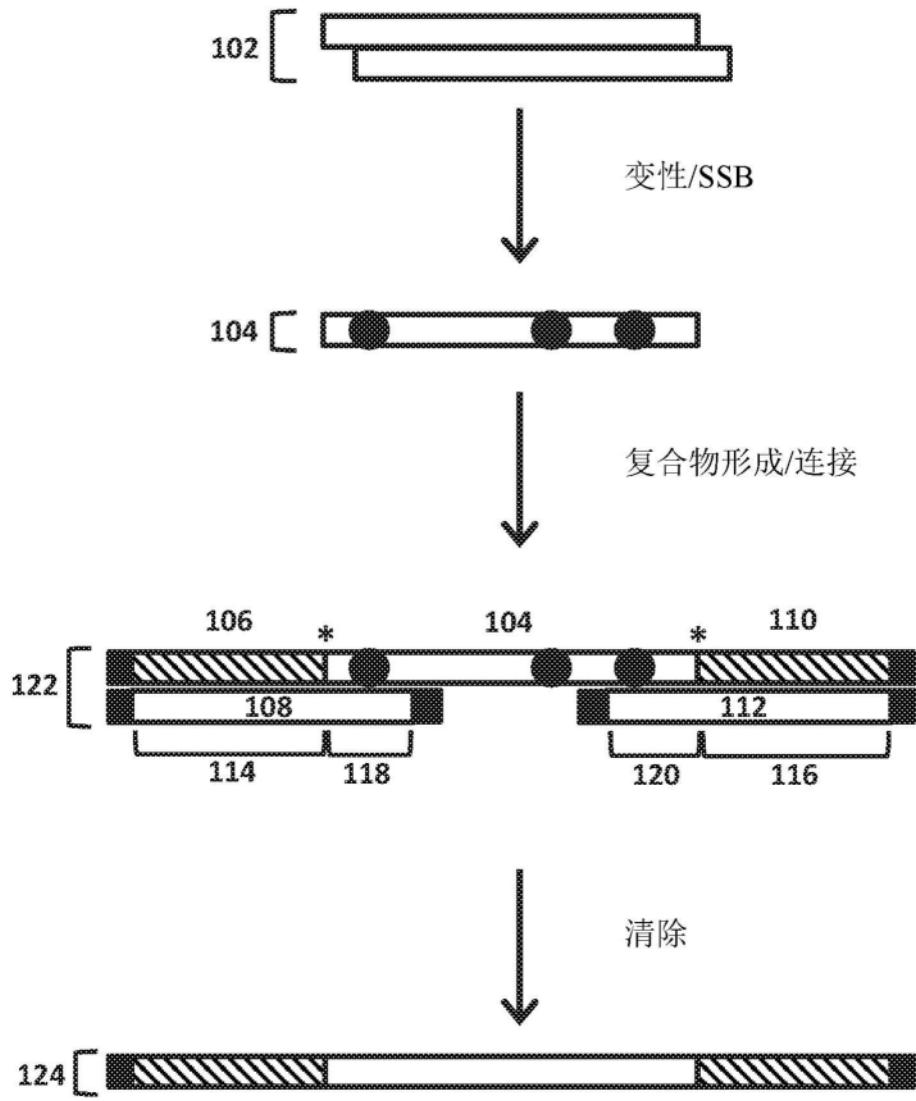


图1

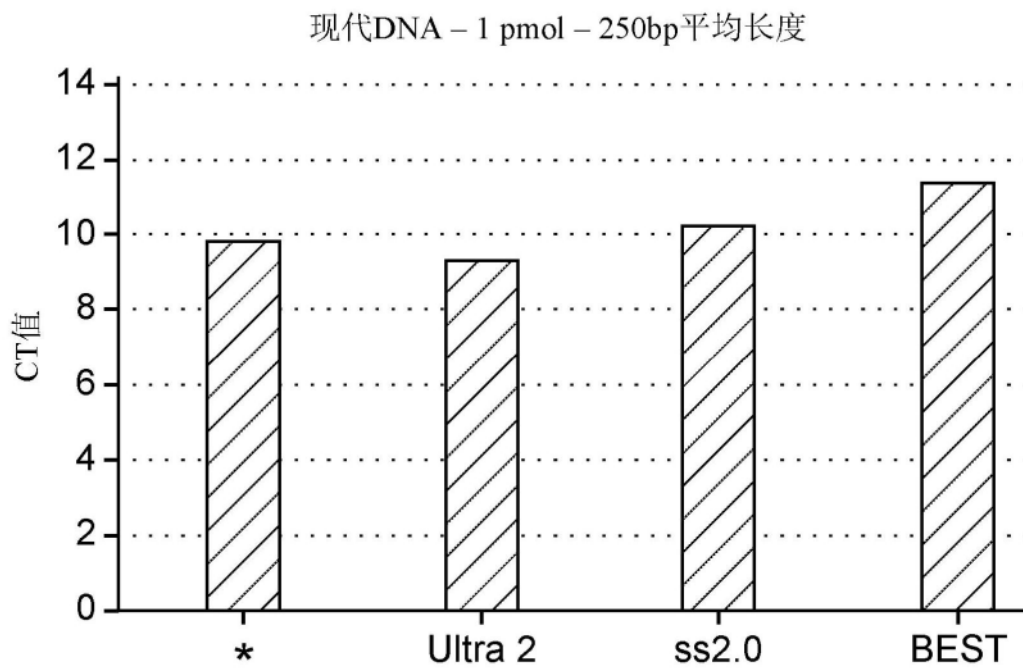
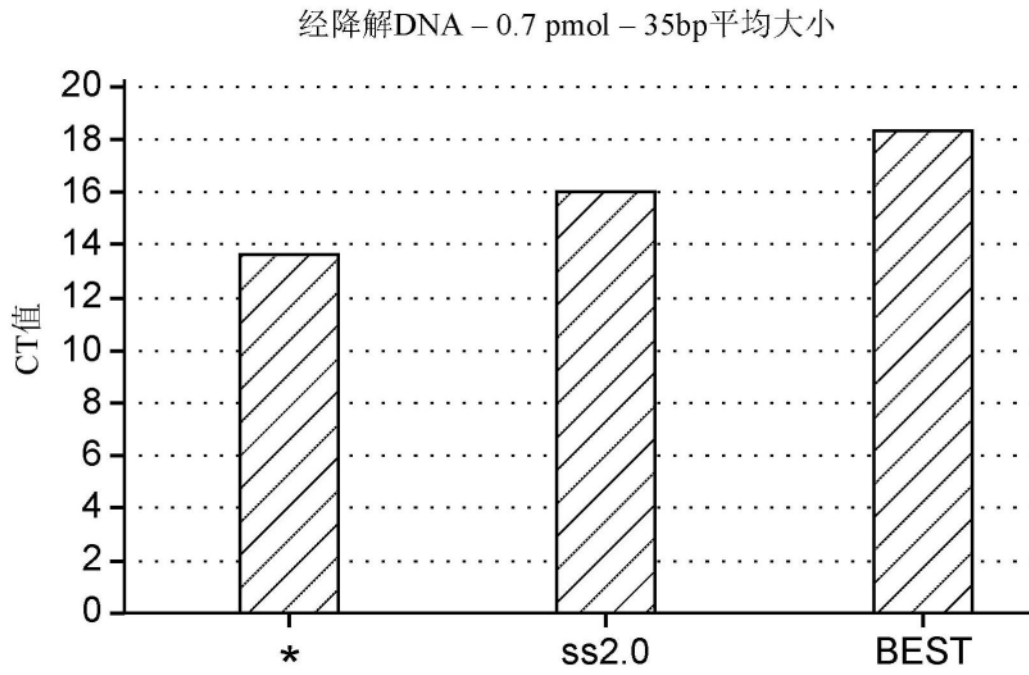


图2

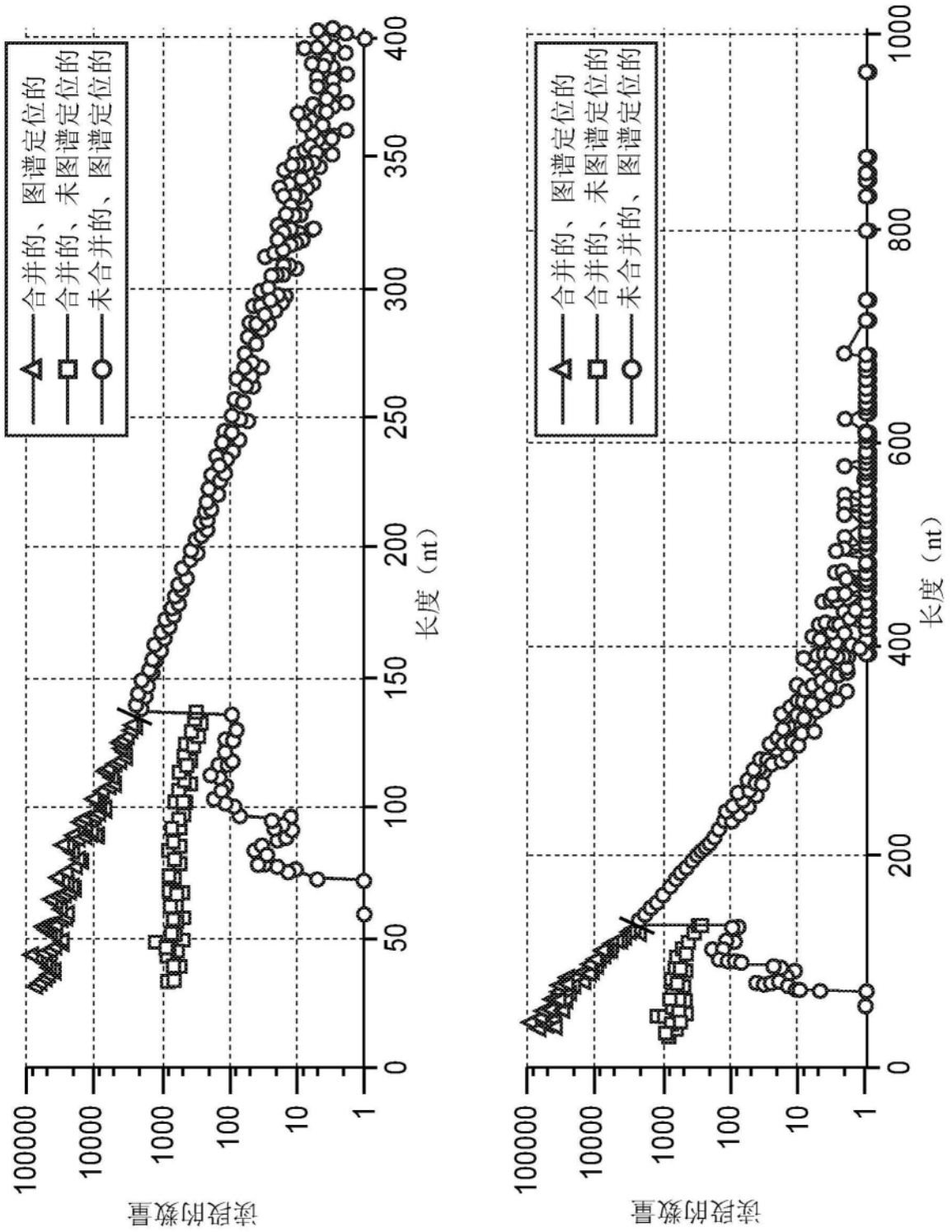


图3

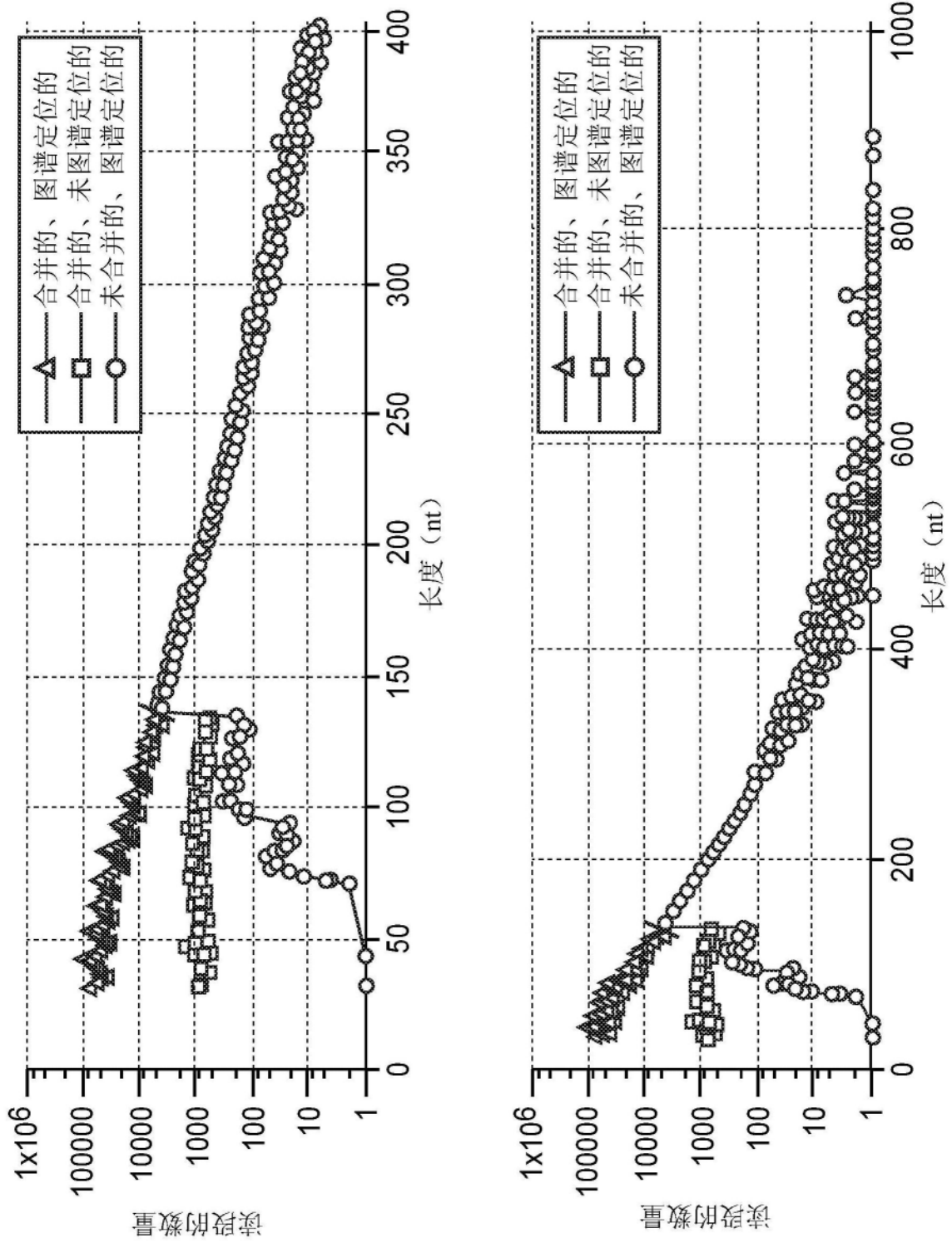


图3(续)

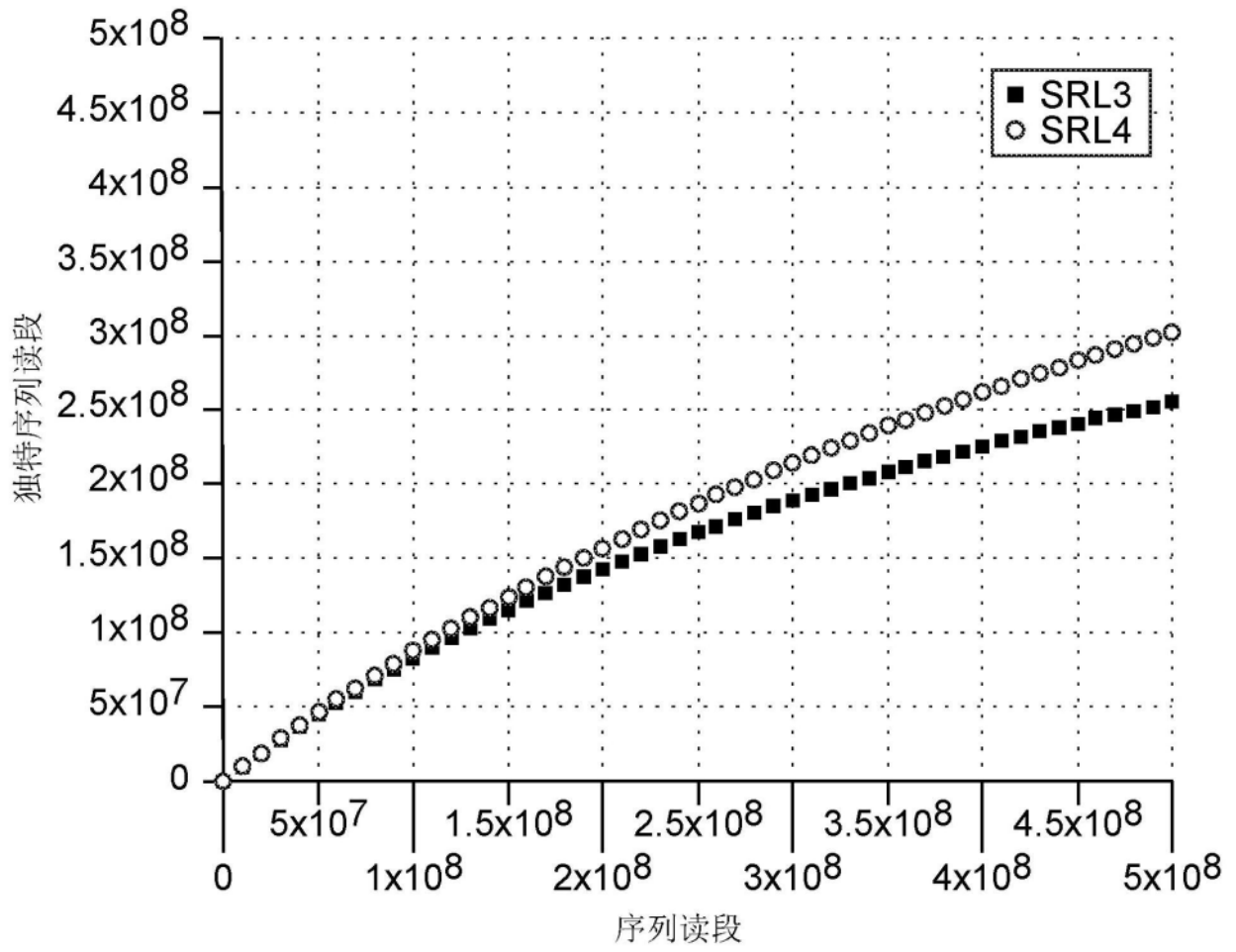


图4

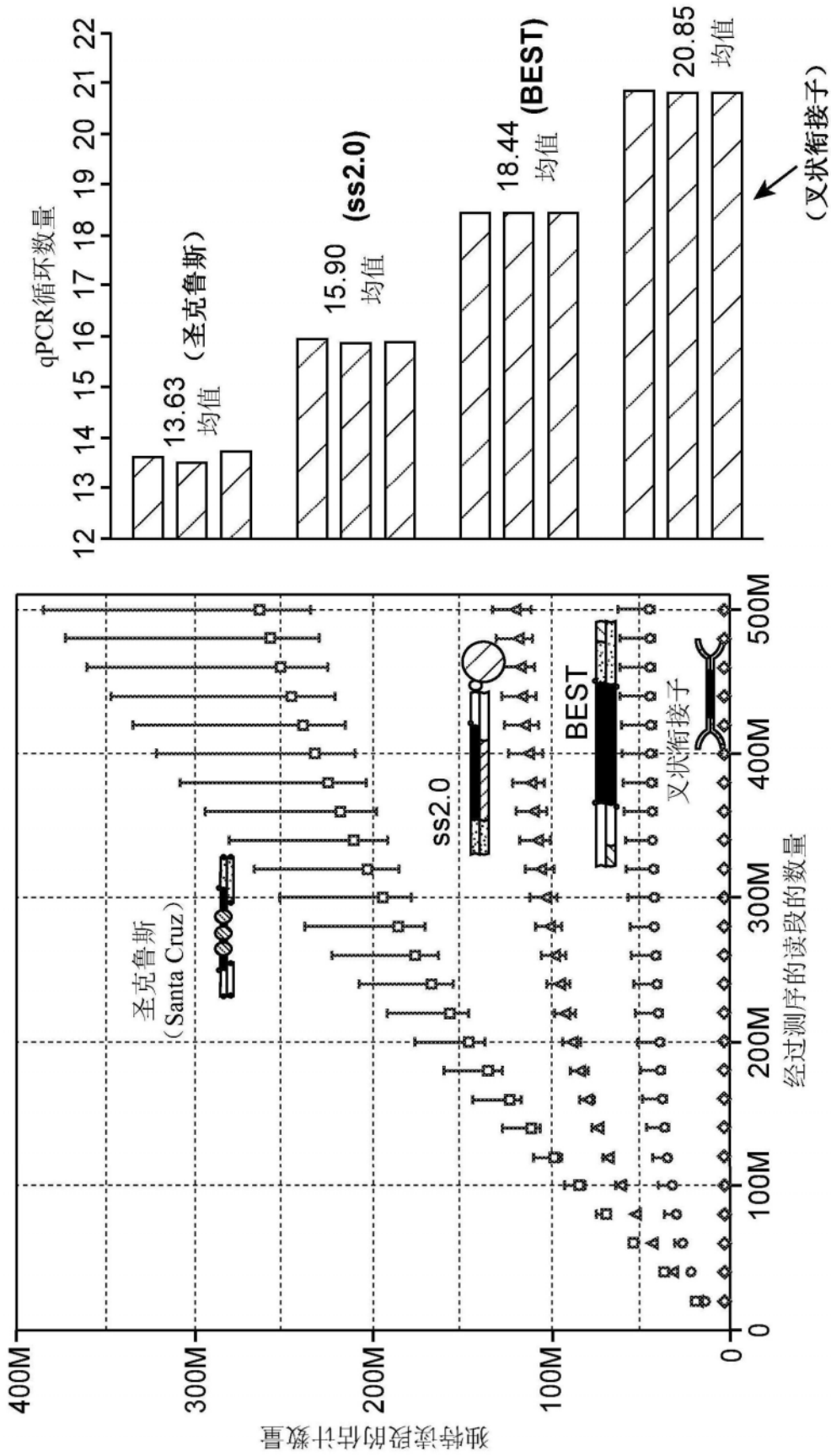


图5