(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106192019 A
(43)申请公布日 2016.12.07

(21)申请号 201610368543.8

(22)申请日 2016.05.30

(30)优先权数据
62/167,892 2015.05.29 US
62/218,906 2015.09.15 US

(71)申请人 分子克隆研究室有限公司
地址 美国加利福尼亚州

(72)发明人 郑建萍 史长平 沈丹 N·唐

(74)专利代理机构 北京市铸成律师事务所 11313
代理人 郝文博 郗名悦

(51)Int.Cl.
C40B 50/06(2006.01)
C12Q 1/68(2006.01)

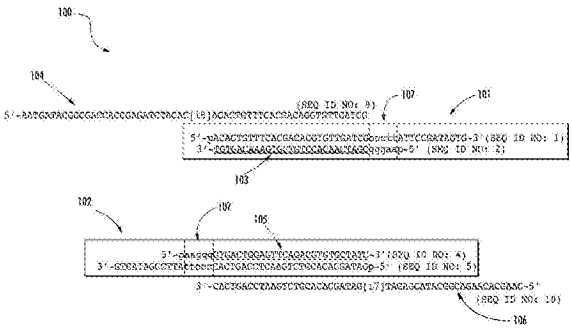
权利要求书3页 说明书24页
序列表11页 附图9页

(54)发明名称

用于制备测序文库的组合物和方法

(57)摘要

用于制备测序文库的组合物和方法。提供包括活化拓扑异构酶适配子(TOPO适配子)的组合物和使用所述活化TOPO适配子的方法来制备从样品多核苷酸(例如,DNA、RNA)得到的目标DNA双链体的文库以便精简化制备大量样品。这类文库可用于下一代测序(NGS)。



CN 106192019 A

1.一种用于制备目标DNA双链体的测序文库的方法,所述方法包括:

(a)提供具有第一端和第二端的多个所述目标DNA双链体;

(b)提供多个第一和第二线性拓扑异构酶(TOPO)活化测序适配子,其中所述第一TOPO活化测序适配子包括第一引物结合序列并且所述第二TOPO活化测序适配子包括第二引物结合序列,其中所述第一引物结合序列不同于所述第二引物结合序列,并且所述第一引物结合序列与第一寡核苷酸引物杂交,并且所述第二引物结合序列与第二寡核苷酸引物杂交;并且

(c)将所述第一和第二活化TOPO适配子共价连接至所述多个DNA双链体的所述第一和第二端以提供多个TOPO适配子-DNA双链体复合物,从而制备测序文库。

2.根据权利要求1所述的方法,其中所述目标DNA双链体在步骤(b)之前钝端化和去磷酸化。

3.一种用于制备目标DNA双链体的测序文库的方法,所述方法包括:

(a)提供具有第一端和第二端的多个所述目标DNA双链体;

(b)提供多个部分互补TOPO活化测序适配子,其中所述适配子包括包含所结合TOPO的双链体区域和至少一个单链区域,并且其中所述适配子包括使第一寡核苷酸引物杂交的第一寡核苷酸引物结合序列;

(c)将所述TOPO活化测序适配子共价连接至所述多个DNA双链体的所述第一和第二端以提供多个TOPO适配子-DNA双链体复合物;

(d)将所述第一寡核苷酸引物延伸以产生互补链,所述互补链包括使第二寡核苷酸引物杂交的第二引物结合位点;并且

(e)将所述第二寡核苷酸引物与所述第二引物结合位点杂交以延伸所述适配子-DNA双链体的第二链;从而制备测序文库。

4.根据权利要求3所述的方法,其中所述目标DNA双链体在步骤(b)之前钝端化和去磷酸化。

5.如权利要求3所述的方法,其中所述适配子包括单链发夹区域,其中所述发夹区域包括所述第一寡核苷酸引物结合序列和所述第二引物结合位点的反向互补序列。

6.如权利要求3所述的方法,其中所述适配子包括第一寡核苷酸和第二寡核苷酸,其中所述第一和第二寡核苷酸中的每一个的一部分彼此互补并且形成包括所结合TOPO的所述双链体区域,其中所述第一和第二寡核苷酸中的每一个的一部分是单链并且不彼此互补,其中所述第二寡核苷酸包括第一引物结合序列,并且其中所述第一寡核苷酸包括所述第二引物结合位点的反向互补序列。

7.如权利要求1至6中任一项所述的方法,其中步骤(c)包括:

(i)使所述TOPO活化适配子复合物的TOPO承载第一链与所述目标DNA双链体的第一链的所述5'端反应以共价连接所述活化TOPO活化适配子的所述第一链以形成所述TOPO适配子-目标DNA双链体复合物的第一链;并且

(ii)将所述TOPO活化适配子复合物的第二链连接至所述目标DNA双链体的第二链的所述3'端以形成所述TOPO适配子-目标DNA双链体复合物的第二链。

8.如权利要求7所述的方法,其还包括:使所述第一和第二寡核苷酸引物与所述第一和第二寡核苷酸引物结合序列杂交并且扩增所述TOPO适配子-目标DNA复合物。

9.如权利要求1至6中任一项所述的方法，其中步骤(c)在少于10分钟内执行。

10.如权利要求1至6中任一项所述的方法，其还包括制备所述活化TOPO适配子。

11.如权利要求1至6中任一项所述的方法，其还包括从生物样品获得DNA或RNA。

12.如权利要求11所述的方法，其中所述生物样品是生物流体样品或组织样品。

13.如权利要求1至6中任一项所述的方法，其中所述目标DNA双链体是无细胞DNA分子。

14.如权利要求1至6中任一项所述的方法，其中所述目标DNA双链体是基因组DNA的片段部分。

15.如权利要求1至6中任一项所述的方法，其中所述DNA双链体是从生物样品的细胞RNA转录的cDNA片段。

16.一种对多核苷酸进行测序的方法，所述方法包括：

(a)根据权利要求1-6中任一项所述来制备目标DNA双链体的测序文库；并且

(b)对TOPO适配子-目标DNA双链体复合物的所述文库进行测序。

17.如权利要求16所述的方法，其中所述测序包括经由选自合成测序、焦磷酸测序和连接测序的下一代测序(NGS)方法来大规模并行测序。

18.如权利要求17所述的方法，其中所述NGS是对标签或单一分子进行测序。

19.一种试剂盒，其包括：

(i)多个TOPO活化测序适配子；

(ii)一个或多个扩增引物；

(iii)一个或多个测序引物；

(iv)一个或多个反应缓冲液；和

(v)制备DNA测序文库的说明书。

20.如权利要求19所述的试剂盒，其中所述TOPO活化测序适配子是互补线性TOPO适配子。

21.如权利要求19所述的试剂盒，其中所述TOPO活化测序适配子是部分互补TOPO测序适配子，其包括TOPO活化DNA双链体和一个或多个单链部分。

22.如权利要求21所述的试剂盒，其中所述TOPO测序适配子包括单链发夹区域。

23.如权利要求22所述的试剂盒，其中所述适配子的所述发夹区域包括尿嘧啶残基，并且所述试剂盒还包括尿嘧啶DNA去糖基化酶(UDG)。

24.如权利要求21所述的试剂盒，其中所述TOPO测序适配子包括第一寡核苷酸和第二寡核苷酸，其中所述第一和第二寡核苷酸中的每一个的一部分彼此互补并且形成所述双链体区域，并且其中所述第一和第二寡核苷酸中的每一个的一部分是单链并且不彼此互补。

25.如权利要求19-24中任一项所述的试剂盒，其还包括连接酶。

26.如权利要求1所述的方法，其中所述第一线性TOPO活化测序适配子包括互补序列SEQ ID NO:2和3；并且其中所述第二线性TOPO活化测序适配子包括互补序列SEQ ID NO:4和6。

27.如权利要求5所述的方法，其中所述发夹拓扑异构酶活化(TOPO)测序适配子包括SEQ ID NO:20的寡核苷酸序列。

28.如权利要求6所述的方法，其中所述适配子包括以下序列的第一和第二寡核苷酸：

(a)分别为SEQ ID NO:30和31；或

(b)分别为SEQ ID NO:32和31。

29.如权利要求19所述的试剂盒,其中所述多个TOPO活化测序适配子包括:

(a)一组第一和第二线性适配子,其中所述第一线性TOPO活化测序适配子包括互补序列SEQ ID NO:2和3;并且其中所述第二线性TOPO活化测序适配子包括互补序列SEQ ID NO:4和6;

(b)发夹TOPO活化测序适配子,其各自包括SEQ ID NO:20;

(c)TOPO活化测序适配子,其包括分别为序列SEQ ID NO:30和31的第一和第二寡核苷酸;和/或

(d)TOPO活化测序适配子,其包括分别为序列SEQ ID NO:32和31的第一和第二寡核苷酸。

用于制备测序文库的组合物和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求2015年5月29日提交的美国临时申请号62/167,892,和2015年9月15日提交的美国临时申请号62/218,906的权益,所述申请全部通过引用并入本文。

[0003] 序列表

[0004] 包括SEQ ID NO:1-35的序列表附加于此。提供于序列表中的每个序列出于所有目的全部以引用方式并入本文。

技术领域

[0005] 提供用于制备DNA文库以通过下一代测序方法来测序的组合物和方法。提供测序适配子以改进文库DNA的产率并且加快文库制备过程。

背景技术

[0006] 下一代测序(NGS)方法的出现产生大量核苷酸序列信息,所述信息可用于提供与遗传信息有关的序列分析。NGS测序的基础是文库构建,即将要分析的目标多核苷酸,RNA和/或DNA制备成与待使用的测序系统相容的形式。

[0007] 制备测序文库的过程可显著影响测序数据的质量和输出。制备用于NGS的DNA文库的当前方法是耗费时间的、倾向于显著样品损失,并且导致所测序的遗传物质的低覆盖范围。

[0008] 存在对于制备多核苷酸DNA文库的改进方法的需要。可用于所需应用,例如,基因组测序、目标测序、RNA测序等,并且可适用于使用不同NGS平台来分析目标多核苷酸的方法将提供额外优势。

发明内容

[0009] 提供包括活化拓扑异构酶适配子(TOPO适配子)的组合物和使用活化TOPO适配子的方法来制备从样品多核苷酸(例如,DNA、RNA)得到的目标DNA双链体的文库以便精简化制备大量样品供下一代测序(NGS)。另外,此高通量方法可自动化以进一步缩减提供大量样品的遗传序列分析的时间和成本。

[0010] 使用活化TOPO适配子有利于在制备文库期间减少和/或防止形成适配子二聚体。另外,与使用标准寡核苷酸适配子的仅连接酶依赖性连接,例如,Illumina的TruSeq Nano DNA Library Prep所耗费时间和所获得的文库产物的产率相比,使用活化TOPO适配子加快样品分析过程,并且极大地改进文库DNA产物的产率。

[0011] 所提供的组合物和方法可适用于在医学、无创诊断例如出生前诊断、农业和环境监测和需要测序遗传物质的其他生物样品测试应用领域中分析样品。

[0012] 在一方面,提供用于制备目标DNA双链体的测序文库的方法。在一个实施方案中,所述方法包括(a)提供具有第一端和第二端的多个目标DNA双链体;(b)任选地,将目标DNA双链体钝端化并且去磷酸化;(c)提供多个第一和第二线性拓扑异构酶(TOPO)活化测序适

配子，其中第一活化TOPO适配子包括第一引物结合序列并且第二适配子包括第二引物结合序列，其中第一引物结合序列不同于第二引物结合序列，并且第一引物结合序列与第一寡核苷酸引物杂交，并且第二引物结合序列与第二寡核苷酸引物杂交；并且(d)将第一和第二活化TOPO适配子共价连接至多个DNA双链体的第一和第二端以提供多个TOPO适配子-DNA双链体复合物，从而制备测序文库。在一些实施方案中，第一线性TOPO活化测序适配子各自包括互补序列SEQ ID NO:2和3；并且第二线性TOPO活化测序适配子各自包括互补序列SEQ ID NO:4和6。

[0013] 在另一个实施方案中，用于制备目标DNA双链体的测序文库的方法包括：(a)提供具有第一端和第二端的多个目标DNA双链体；(b)任选地将目标DNA双链体钝端化并且去磷酸化；(c)提供多个部分互补TOPO活化测序适配子，其中所述适配子包括包含所结合TOPO的双链体区域和至少一个单链区域，并且其中所述适配子包括使第一寡核苷酸引物杂交的第一寡核苷酸引物结合序列；(d)将TOPO活化测序适配子共价连接至多个DNA双链体的第一和第二端以提供多个TOPO适配子-DNA双链体复合物；(e)将第一寡核苷酸引物延伸以产生互补链，所述互补链包括使第二寡核苷酸引物杂交的第二引物结合位点；并且(f)将第二寡核苷酸引物与第二引物结合位点杂交以延伸适配子-DNA双链体的第二链，从而制备测序文库。目标DNA双链体可任选地钝端化，例如使用T4聚合酶和/或Klenow DNA聚合酶。目标DNA双链体可任选地去磷酸化，例如使用碱性磷酸酶，例如，小牛肠磷酸酶、虾碱性磷酸酶、南极磷酸酶。

[0014] 在一些实施方案中，部分互补适配子包括单链发夹区域，其中发夹区域包括第一寡核苷酸引物结合序列和第二引物结合位点的反向互补序列。在一个实施方案中，发夹拓扑异构酶(TOPO)活化测序适配子各自包括SEQ ID NO:20的寡核苷酸序列。

[0015] 在一些实施方案中，部分互补适配子包括第一寡核苷酸和第二寡核苷酸，其中第一和第二寡核苷酸中的每一个的一部分彼此互补并且形成包括所结合TOPO的双链体区域，其中第一和第二寡核苷酸中的每一个的一部分是单链并且不彼此互补，其中第二寡核苷酸包括第一引物结合序列，并且其中第一寡核苷酸包括第二引物结合位点的反向互补序列。在一个实施方案中，适配子包括分别为SEQ ID NO:30和31的第一和第二寡核苷酸。在另一个实施方案中，适配子包括分别为SEQ ID NO:32和31的第一和第二寡核苷酸。

[0016] 在一些实施方案中，本文公开的用于制备目标DNA双链体的测序文库的方法的步骤(d)包括：(i)使活化TOPO适配子复合物的TOPO承载第一链与目标DNA双链体的第一链的5'端反应以使活化TOPO适配子的第一链共价连接至TOPO适配子-目标DNA双链体复合物的第一链；并且(ii)使活化TOPO适配子复合物的第二链连接至目标DNA双链体的第二链的3'端。

[0017] 在其他实施方案中，本文公开的用于制备目标DNA双链体的测序文库的方法还包括：使第一和第二寡核苷酸引物与第一和第二寡核苷酸引物结合序列杂交并且扩增TOPO适配子-多核苷酸复合物。

[0018] 在其他实施方案中，在本文公开的方法中将目标DNA双链体的端，例如，DNA的端修复和去磷酸化片段共价连接至拓扑异构酶活化测序适配子的步骤在少于10分钟内执行。

[0019] 在其他实施方案中，本文公开的用于制备目标DNA双链体的测序文库的方法还包括制备活化TOPO适配子。

[0020] 在其他实施方案中，本文公开的用于制备目标DNA双链体的测序文库的方法还包括从生物样品获得DNA或RNA。在一些实施方案中，生物样品是生物流体样品或组织样品。

[0021] 在其他实施方案中，制备测序文库的方法的目标DNA双链体是基因组DNA的片段部分。在其他实施方案中，目标DNA双链体是从生物样品的细胞RNA转录的cDNA的片段。

[0022] 在另一方面，提供多核苷酸的下一代测序(NGS)的方法。在一些实施方案中，所述方法包括：(a)制备如本文公开的目标DNA双链体的测序文库；并且(b)将TOPO适配子-目标DNA双链体复合物的文库下一代测序。在一些实施方案中，测序方法选自：合成测序、焦磷酸测序和连接测序。在其他实施方案中，测序方法是标签或单一分子的NGS。

[0023] 在另一方面，提供试剂盒。在一些实施方案中，试剂盒包括：(i)多个TOPO活化测序适配子；(ii)一个或多个扩增引物；(iii)一个或多个测序引物；(iv)一个或多个反应缓冲液；和(v)制备DNA测序文库的说明书。在一些实施方案中，试剂盒的TOPO活化适配子是互补线性TOPO适配子。在一些其他实施方案中，试剂盒的TOPO活化适配子是部分互补的例如发夹或Y形适配子。在其他实施方案中，试剂盒还包括尿嘧啶DNA糖基化酶(UDG)。在其他实施方案中，试剂盒还包括连接酶。

附图说明

[0024] 图1示出两个互补线性前适配子(101)和(102)，和通用寡核苷酸引物(104)和(106)的示例性集合(100)。参考实施例2。

[0025] 图2示出示例性部分互补发夹前适配子(200)，和通用寡核苷酸引物(203)和(205)。

[0026] 图3示出使用一组两个互补线性前适配子来制备TOPO适配子-目标DNA双链体-TOPO适配子复合物的示例性过程(300)。参考实施例1。

[0027] 图4示出使用活化线性TOPO适配子来制备目标DNA双链体的文库供NGS测序的示例性过程(400)。

[0028] 图5示出使用活化发夹TOPO适配子来制备目标DNA双链体的文库供NGS测序的示例性过程(500)。

[0029] 图6A-D示出如与使用标准寡核苷酸适配子(6A)和(6B)的仅连接酶依赖性连接相比，使用TOPO活化与目标DNA双链体(6C)和(6D)的拓扑异构酶依赖性连接来制备的样品DNA测序文库的示例性电泳图。

[0030] 图7示出示例性部分互补“单一指标”前适配子(700)，和通用寡核苷酸引物(703)和(705)。

[0031] 图8示出示例性部分互补“双重指标”前适配子(800)，和通用寡核苷酸引物(803)和(805)。

[0032] 图9示出使用活化部分互补TOPO适配子来制备目标DNA双链体的文库供NGS测序的示例性过程(900)。

[0033] 以引用的方式并入

[0034] 本文中提及的所有专利、专利公开案和其它出版物，包括在这些参考文献中公开的所有序列，以引用的方式并入本文，并入程度就如同每个个别的出版物、专利或专利公开案被具体和单独地说明以引用的方式并入一样。然而，引用任何文件不应理解为承认它是

相对于本发明的先前技术。

具体实施方式

[0035] 提供包括活化拓扑异构酶(TOPO)适配子的组合物和使用活化TOPO适配子的方法来制备从样品多核苷酸(例如,DNA、RNA)得到的目标DNA双链体的文库。这些文库可用于精简化制备大量样品用于诸如但不限于多核苷酸测序的应用,例如下一代测序(NGS)。使用TOPO适配子来制备文库改进文库DNA的产率、加快文库制备以及进而样品分析的过程,并且最大限度地减少或防止形成适配子二聚体。

[0036] 除非在本文中另外定义,否则本文所使用的所有技术和科学术语均具有与本发明所属领域的普通技术人员通常所理解的相同的意思。Singleton等人DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY,第2版,John Wiley和Sons,New York(1994),以及Hale和Marham,THE HARPER COLLINS DICTIONARY OF BIOLOGY,Harper Perennial,NY(1991)为本领域技术人员提供用于本发明中的许多术语的一般词典。虽然在本发明的实践或测试中可以使用类似或等同于本文所描述的那些方法和材料的任何方法和材料,但描述优选方法和材料。关于本领域的定义和术语,从业者可具体参考Sambrook等人1989和Ausubel FM等人1993。应了解本发明不限于所描述的具体方法、方案和试剂,因为这些方法、方案和试剂可变化。

[0037] 数字范围包括定义所述范围的数字。术语约在本文中用于意指数值的加上或减去百分之十(10%)。例如"约100"是指介于90和110之间的任何数字。

[0038] 除非另有指明,否则分别地核酸从左至右以5'至3'方向书写;氨基酸序列从左至右以氨基至羧基方向书写。

[0039] 术语"双链体"在本文中是指在两个多核苷酸序列之间存在的互补区域。术语"双链体区域"在涉及TOPO适配子使用时,在本文中是指在两个寡核苷酸或单一寡核苷酸的两个部分之间存在的序列互补区域,涵盖拓扑异构酶的识别序列。

[0040] 术语"目标DNA双链体"在本文中是指从样品多核苷酸得到的双链DNA分子,所述多核苷酸是DNA,例如,基因组或无细胞DNA,和/或RNA。

[0041] 术语"前适配子"在本文中是指双链体寡核苷酸(例如,DNA)底物,相容位点特异性拓扑异构酶,例如,拓扑异构酶I可裂解所述底物并且拓扑异构酶在裂解点处共价连接至所述底物以产生活化TOPO适配子。

[0042] 术语"活化TOPO适配子"在本文中是指包括具有在第一端的3'末端处或附近共价结合的单一拓扑异构酶的双链体寡核苷酸区域的多核苷酸结构。

[0043] 术语"第一端"和"第二端"在涉及核酸分子使用时,在本文中是指线性核酸分子的端。

[0044] 术语"单链突出物"或"突出物"在本文中用于意指延伸超过ds核酸分子的互补链的末端的双链(ds)核酸分子的链。术语"5'突出物"或"5'突出序列"在本文中用于意指在5'方向上延伸超过ds核酸分子的互补链的3'末端的ds核酸分子的链。术语"3'突出物"或"3'突出序列"在本文中用于意指在3'方向上延伸超过ds核酸分子的互补链的5'末端的ds核酸分子的链。

[0045] 术语"易分割的链"在本文中是指包括拓扑异构酶("TOPO")的位点特异性识别序

列的双链体寡核苷酸适配子的链，所述拓扑异构酶例如拓扑异构酶I，例如，痘苗病毒拓扑异构酶I，和其重组形式。

[0046] 术语“文库”在本文中是指在其5’端处共享共同序列并且在其3’端处共享共同序列的一些或多个模板分子，即，目标DNA双链体。使用术语“文库”来意指一些或多个模板分子不应理解为暗示构成文库的模板从特定来源得到，或“文库”具有特定组成。举例来说，使用术语“文库”不应理解为暗示文库内的个别模板必须具有不同核苷酸序列或模板必须在序列和/或来源方面相关联。

[0047] 术语“下一代测序(NGS)”在本文中是指允许大规模并行测序克隆扩增和单一核酸分子的测序方法，在此期间来自单一样品或来自多个不同样品的多个，例如，数百万个核酸片段一齐测序。NGS的非限制实例包括合成测序、连接测序、实时测序和纳米孔测序。

[0048] 术语“生物测定”在本文中是指包括样品核酸例如DNA的NGS测序的多步骤测定。多步骤生物测定可包括样品收集、样品分级、核酸净化步骤和制备测序文库的必需核酸修饰步骤中的一个或多个。

[0049] 术语“端修复DNA”在本文中是指在体外经受酶促反应以使5’和/或3’突出物钝端化的DNA。钝端可通过使用聚合酶在5’至3’方向上填充链的缺失碱基，并且使用外切核酸酶来移除3’突出物而获得。举例来说，T4聚合酶和/或Klenow DNA聚合酶可用于DNA端修复。

[0050] 术语“测序文库”在本文中是指经过加工供例如使用大规模并行方法例如NGS来测序的DNA。DNA可任选地经扩增以获得可通过NGS来测序的经过加工的DNA的多个拷贝的群体。

[0051] 术语“适配子”在本文中是指连接至双链DNA分子的两个链的核酸。适配子可由彼此碱基配对，即，互补的两个不同寡核苷酸分子组成。或者，适配子可由包括一个或多个互补区域，和一个或多个非互补区域的单一寡核苷酸组成。

[0052] 如本文使用的术语“碱基对”或“bp”是指双链DNA分子中的腺嘌呤(A)与胸腺嘧啶(T)，或胞嘧啶(C)与鸟嘌呤(G)的合作关系(即，氢键配对)。在一些实施方案中，碱基对可包括与尿嘧啶(U)配对的A，例如，在DNA/RNA双链体中。

[0053] 术语“互补”在本文中是指在单一多核苷酸链的双链体区域中或在两个多核苷酸链之间、在成对核苷酸之间经由碱基配对的序列互补性的广泛概念。已知腺嘌呤核苷酸能够与作为胸腺嘧啶或尿嘧啶的核苷酸形成特定氢键(“碱基配对”)。类似地，已知胞嘧啶核苷酸能够与鸟嘌呤核苷酸碱基配对。

[0054] 术语“基本上互补”在本文中是指在适配子的单一多核苷酸链的双链体区域中或两个多核苷酸链之间的序列互补性，其中互补性小于100％但是大于90％，并且在适配子共价连接至目标DNA双链体的条件下保持双链体区域的稳定性。

[0055] 术语“净化”在本文中是指分子在包含所述分子的样品的至少90重量％，或至少95重量％，或至少98重量％的浓度下存在于样品中。

[0056] 术语“分离”在本文中是指与例如在其天然环境下通常与它缔合的至少一种其他分子分离的核酸分子。分离的核酸分子包括通常例如通过染色体表达来表达核酸分子的细胞中所含的核酸分子，但核酸分子存在于染色体外或存在于不同于其天然染色体位置的染色体位置。

[0057] 术语“核苷酸”在本文中是指由糖部分(戊糖)、磷酸和含氮的杂环碱基组成的DNA

或RNA的单体单元。碱基经由糖苷碳(戊糖的1′碳)连接至糖部分并且碱基与糖的此组合是核苷。当核苷含有键合至戊糖的3’或5’位置的磷酸基团时,它被称为核苷酸。聚合性可操作连接核苷酸的序列通常在本文中称为“碱基序列”、“核苷酸序列”或核酸或多核苷酸“链”并且在本文中通过其从左至右方向是在5’末端至3’末端的常规方向上的化学式来表示,所述5’末端和3’末端分别是指聚合物序列的“5’”和“3’”端处的末端5’磷酸基团和末端3’羟基。

[0058] 术语“寡核苷酸”、“多核苷酸”和“核酸”在本文中是指包括两个或更多个,优选地超过三个脱氧核糖核苷酸和/或核糖核苷酸的分子。它的确切大小取决于许多因素,进而取决于寡核苷酸的最终功能或用途。寡核苷酸可合成或通过克隆或从天然(例如,基因组)来源得到。如本文使用,术语“多核苷酸”是指由以链形式共价键合的核苷酸单体组成的聚合物分子。DNA(脱氧核糖核酸)和RNA(核糖核酸)是多核苷酸的实例。

[0059] 术语“引物”在本文中是指寡核苷酸,不论天然地发生或合成产生,其在安置于诱导与核酸链互补的引物延伸产物的合成的条件下时能够充当核酸合成的启始点,例如,在四种不同核苷酸三磷酸和聚合酶例如耐热酶存在下,在适当缓冲液(“缓冲液”包括pH、离子强度、辅助因子等)中并且在合适温度下。为了获得扩增中的最大效率,引物优选地是单链,但是可替代地是双链。如果是双链,引物首先经处理以将其链分离,然后用于制备延伸产物。优选地,引物是寡脱氧核苷酸。引物必须足够长以在聚合酶例如耐热聚合酶的存在下起动延伸产物的合成。引物的确切长度取决于许多因素,包括温度、引物来源和使用方法。举例来说,取决于目标序列的复杂性,寡核苷酸引物通常含有15-25个核苷酸,但是它可含有更多或更少的核苷酸。短引物分子总体上需要更冷温度以与模板形成足够稳定的杂交复合物。

[0060] 如本文使用,术语“指标”在涉及核苷酸序列使用时意指可不同于其他指标并且不同于样品内包含的多核苷酸内的其他核苷酸序列的独特核苷酸序列。核苷酸指标可为随机或专门设计核苷酸序列。指标可具有任何所需序列长度,只要它具有足够长度以在群体中的多个指标内和/或在所分析或询问的多个多核苷酸内成为独特核苷酸序列。本公开的核苷酸指标适用于例如连接至目标多核苷酸以将特定种类加标签或做记号以便识别群体内的加标签种类的所有成员。因此,指标适用作为“条码”,其中相同分子种类的不同成员可含有相同指标并且其中不同多核苷酸的群体内的不同种类可具有不同指标。举例来说,指标序列可在测序文库制备期间并入多核苷酸例如目标DNA中以便对于从不同来源制备的汇集文库进行多重测序。

[0061] 术语“合成”在本文中是指以模板依赖性方式制造新的多核苷酸链或延长现有多核苷酸(即,DNA或RNA)的任何体外方法。根据本发明的合成可包括扩增,其借助于聚合酶来增加多核苷酸模板序列的拷贝数目。多核苷酸合成(例如,扩增)导致将核苷酸并入多核苷酸中(例如,从引物延伸),从而形成与多核苷酸模板互补的新的多核苷酸分子。所形成的多核苷酸分子和其模板可用作合成额外多核苷酸分子的模板。如本文使用,“DNA合成”包括但不限于聚合酶链反应(PCR),并且可包括使用经过标记的核苷酸,例如,用于探针和寡核苷酸引物,或用于多核苷酸测序。

[0062] 术语“模板DNA分子”在本文中是指例如在引物延伸反应中通过DNA聚合酶来合成互补核酸链的核酸链。

[0063] 术语“模板依赖性方式”在本文中是指涉及引物分子的模板依赖性延伸的过程(例

如，通过DNA聚合酶的DNA合成)。术语“模板依赖性方式”通常是指RNA或DNA的多核苷酸合成，其中多核苷酸的新合成链的序列与模板互补，即，由互补碱基配对的熟知规则来决定(参见，例如，Watson，J.D.等人，于Molecular Biology of the Gene，第4版，W.A.Benjamin，Inc.，Menlo Park，Calif.(1987)中)。

[0064] 在关于拓扑异构酶使用时，术语“在…处或附近”意指拓扑异构酶共价结合至双链体核酸分子例如TOPO适配子的一个链，以使得它可将它所结合的链的末端连接至含有自由5’末端羟基的第二核酸分子。总体上，由于共价结合至端的一个末端，拓扑异构酶在端“处或附近”。举例来说，当拓扑异构酶是IB型拓扑异构酶诸如痘苗拓扑异构酶时，拓扑异构酶结合于双链体核酸分子的端的3’末端处。然而，具有共价结合至端的末端的拓扑异构酶的端还可在互补链中含有单链突出物序列，由此延伸超过拓扑异构酶所结合的末端。这种配置是在ds核酸分子的端附近的拓扑异构酶的实例。

[0065] 本文所提供的标题不是对本发明的各个方面或实施方案的限制，这些方面或实施方案可以通过参考整个说明书来获得。因此，即将在下文定义的术语通过参考说明书全文而得到更充分地定义。

[0066] NGS文库构建的基础是将来自生物样品的核酸，即样品RNA或DNA制备成与待使用的测序系统相容的形式。本文提供的TOPO适配子可用于根据NGS平台的需求来制备测序文库，例如，由Illumina，Life Technologies、Roche和Pacific Biosciences开发的平台。

[0067] 通常，在制备RNA或DNA供NGS分析中的核心步骤是：(i)分裂样品多核苷酸且/或将样品多核苷酸设定尺寸至所需长度，(ii)如果样品是单链，将片段样品多核苷酸转化成双链DNA，(iii)将寡核苷酸适配子连接至目标片段的端，(iv)扩增适配子-片段复合物，和(iv)量化用于测序的最终文库产物。

[0068] 在制备用于NGS的测序文库的当前方法中，将寡核苷酸适配子连接至目标片段的端的步骤(iii)通常要求已转化成双链DNA的样品多核苷酸经历(a)双链体链的端修复，(b)5’端的磷酸化，(c)3’端的A-加尾以促进连接至测序适配子，(f)将适配子仅连接酶依赖性连接至片段DNA双链体，和任选地(g)有限扩增以富集具有连接至两端的适配子的产物，即，适配子-目标DNA双链体-适配子产物。当前方法在提供DNA文库产物的可接受产率方面是耗费时间的和效率低的。另外，通过将适配子连接至模板核酸片段的端来制备核酸片段文库的另一个主要缺点是形成适配子二聚体。适配子二聚体通过将两个适配子直接彼此连接而形成以致于其不含有模板核酸片段作为插入物。这些分子是不合需要的，因为在任何扩增步骤期间，例如在通用扩增反应期间，适配子二聚体与核酸片段文库并列地扩增。由于适配子二聚体总体上小于文库中所包含的片段，因此其以更快速率来扩增并积聚。这样由于在扩增反应中耗尽组件，诸如例如dNTP和引物而限制文库片段的扩增，从而降低扩增反应的效率。另一个更严重问题是在将这些扩增片段测序时，其不给出有用序列信息，因为其不含有插入物。在集群阵列的情况下，不具有目标DNA序列的集群的大量群体是不合需要的，因为从有限大小的芯片获得的实际序列数据的密度较低。因此，测序效率可显著降低。因此，制备具有低水平适配子二聚体的文库在多核苷酸的测序中是非常有利的，尤其在这些过程是高通量过程时。

[0069] 如本文提供的TOPO适配子组合物和使用TOPO适配子来制备测序文库的方法需要5’端的去磷酸化，排除3’端的A-加尾，并且利用适配子的拓扑异构酶催化连接。或者，所述

方法利用连接酶与拓扑异构酶的组合来催化将适配子连接至目标DNA。此方法将适配子连接步骤通常耗费的时间从约4至0.5小时的范围显著减少至约10分钟，或更少。文库产物的产率还大于广泛使用方法，例如，Illumina的TruSeq文库制备的产率。另外，用于根据所提供的方法来制备DNA文库的活化TOPO适配子最大限度地减少或防止适配子二聚体的形成，因为连接活化TOPO适配子需要存在去磷酸化受体DNA，第二活化TOPO适配子不能提供上述受体DNA，因为它具有磷酸化5'端。

[0070] 文库的形成方法是通过将活化TOPO适配子共价连接至多个目标DNA双链体的每端，即，5'和3'端以形成TOPO适配子-目标DNA双链体-TOPO适配子复合物。每个TOPO适配子包括至少一个寡核苷酸引物结合序列，其用来使寡核苷酸引物杂交以启始引物延伸反应，执行所述反应以产生与每个拓扑异构酶适配子-样品DNA复合物的至少一个链互补的延伸产物。可任选地经受有限循环扩增的所得引物延伸产物共同地提供样品/目标核酸的文库。

[0071] 前TOPO适配子

[0072] 在一些实施方案中，活化TOPO适配子从包括双链体寡核苷酸(例如，DNA)区域的前适配子得到，所述区域包括相容位点特异性拓扑异构酶可裂解的序列，例如C/TCCTTN(SEQ ID NO:12和13)，并且拓扑异构酶在裂解点处共价连接至所述区域以产生本文描述的活化TOPO测序适配子。在一些实施方案中，双链体寡核苷酸区域通过将两个分离互补寡核苷酸粘接来形成。在其他实施方案中，寡核苷酸区域由单一寡核苷酸的两个互补部分提供。

[0073] 前适配子的双链体区域的第一链，或易分割的链包括指定拓扑异构酶的识别序列的序列。拓扑异构酶，例如，痘苗拓扑异构酶的识别序列包括第一链中的共有五嘧啶元件5'-(C/T)CCTT⇓N(SEQ ID NO:12和13)(或相关序列)。在一些实施方案中，易分割的键位于距适配子的双链体区域的第一链的3'端至少2bp处。举例来说，识别序列可位于距双链体适配子DNA的3'端2-12bp范围内。在一些实施方案中，通过拓扑异构酶I来形成可裂解复合物需要裂解位点上游的六个双链体核苷酸和下游的两个核苷酸。在一些实施方案中，识别序列还包括五嘧啶元件上游至少1bp的序列，即，识别序列的5'端。识别序列的实例包括但不限于+6/-6双链体GCCCTTATTCCC(SEQ ID NO:14)、+8/-4双链体TCGCCCTTATTC(SEQ ID NO:15)、+10/-2双链体TGTCGCCCTTAT(SEQ ID NO:16)、+11/-2双链体GTGTCGCCCTTA(SEQ ID NO:17)和+10/-12GATTCCCCTTATTCCGATAGTG(顶部链)(SEQ ID NO:18)。

[0074] 在一些实施方案中，前适配子包括至少在指定拓扑异构酶识别序列的序列上互补的两个分离寡核苷酸，第一和第二寡核苷酸链。在一些实施方案中，第一链，或易分割的链具有至少2个、至少5个、至少10个、至少15个或至少20个核苷酸的3'突出物。在一些实施方案中，需要一组两个前适配子，第一和第二前适配子来提供第一和第二活化TOPO适配子，如本文中别处描述。图1示出一组两个互补线性前适配子(101)和(102)，其分别包括用于寡核苷酸引物(104)的第一通用引物结合序列(103)，和用于寡核苷酸引物(106)的第二通用引物结合序列(105)，和包括拓扑异构酶识别序列(107)的双链体区域。图1中示出的示例性前适配子包括具有3'突出物的易分割的链。3'突出物的一部分为前适配子的拓扑异构酶活化所需要，如本文中别处描述。在一些实施方案中，前适配子还包括用于结合测序引物的一个或多个序列。

[0075] 在其他实施方案中，前适配子是单一寡核苷酸，其包括指定拓扑异构酶识别序列的至少一个互补双链体区域，和至少一个非互补区域。在一些实施方案中，寡核苷酸前适配

子的双链体区域的第一链，易分割的链包括3'突出物。3'突出物是至少2个、至少5个、至少10个、至少15个或至少20个核苷酸。在一些实施方案中，需要一组两个前适配子，第一和第二前适配子来提供第一和第二活化TOPO适配子，如本文中别处描述。

[0076] 图2示出示例性部分互补发夹前适配子(200)，其由单一寡核苷酸形成，所述寡核苷酸包括其双链体区域内的拓扑异构酶识别序列(201)、可与第一寡核苷酸引物(203)杂交的第一通用引物结合序列(202)，和第二序列(204)，其反向互补序列产生可与第二寡核苷酸引物(205)杂交的第二寡核苷酸引物结合序列。在一个实施方案中，部分互补寡核苷酸前适配子包括SEQ ID NO:20的单一寡核苷酸。单一寡核苷酸前适配子包括两个部分，所述部分互补并且指定拓扑异构酶-I识别序列CCCTTN(SEQ ID NO:12)，其中N是A。单一寡核苷酸前适配子还包括SEQ ID NO:21的第一通用引物结合序列，和第二序列(SEQ ID NO:22)，其反向互补序列产生第二寡核苷酸引物结合序列(SEQ ID NO:23)。SEQ ID NO:24和25的寡核苷酸引物分别在初始引物延伸反应和延伸产物的随后扩增过程中结合至SEQ ID NO:21和23。第一和第二寡核苷酸引物结合序列通过可裂解尿嘧啶碱基来分离。如图2示出的寡核苷酸引物内的[i5]和[i7]插入物指示其中指标序列可引入寡核苷酸引物中的位置。

[0077] 图7示出示例性部分互补"Y形"前适配子(700)，其由第一和第二寡核苷酸形成，所述寡核苷酸包括双链体区域内的拓扑异构酶识别序列(701)，第二寡核苷酸的非互补(非双链体)区域中的可与第一寡核苷酸引物(703)杂交的第一引物结合序列(702)，和第一寡核苷酸(704)的非互补区域中的第二序列，其反向互补序列产生可与第二寡核苷酸引物(705)杂交的第二寡核苷酸引物结合序列。在图7描绘的一个实施方案中，部分互补寡核苷酸前适配子包括具有分别在SEQ ID NO:30和31中描绘的序列的第一和第二寡核苷酸。前适配子的两个寡核苷酸包括以下部分，所述部分互补并且指定拓扑异构酶-I识别序列CCCTTN(SEQ ID NO:12)，其中N是A。前适配子还包括在SEQ ID NO:33中描绘的第一引物结合序列，和第二序列(SEQ ID NO:27)，其反向互补序列产生第二寡核苷酸引物结合序列(SEQ ID NO:34)。在SEQ ID NO:29中描绘的第一引物在初始引物延伸反应中结合至SEQ ID NO:32，并且具有分别在SEQ ID NO:29和27中描绘的序列的第一和第二引物用于延伸产物的随后扩增。如图7示出的第二寡核苷酸内的[i7]插入物指示其中指标序列可引入扩增产物中的位置。

[0078] 图8示出示例性部分互补"Y形"前适配子(800)，其由第一和第二寡核苷酸形成，所述寡核苷酸包括双链体区域内的拓扑异构酶识别序列(801)，第二寡核苷酸的非互补(非双链体)区域中的可与第一寡核苷酸引物(803)杂交的第一引物结合序列(802)，和第一寡核苷酸(804)的非互补区域中的第二序列，其反向互补序列产生可与第二寡核苷酸引物(805)杂交的第二寡核苷酸引物结合序列。在图8描绘的一个实施方案中，部分互补寡核苷酸前适配子包括具有分别在SEQ ID NO:32和31中描绘的序列的第一和第二寡核苷酸。前适配子的两个寡核苷酸包括以下部分，所述部分互补并且指定拓扑异构酶-I识别序列CCCTTN(SEQ ID NO:12)，其中N是A。前适配子还包括在SEQ ID NO:33中描绘的第一引物结合序列，和第二序列(SEQ ID NO:27)，其反向互补序列产生第二寡核苷酸引物结合序列(SEQ ID NO:35)。在SEQ ID NO:29中描绘的第一引物在初始引物延伸反应中结合至SEQ ID NO:32，并且具有分别在SEQ ID NO:29和28中描绘的序列的第一和第二引物用于延伸产物的随后扩增。分别如图8示出的第一和第二寡核苷酸内的[I5]和[i7]插入物指示其中指标序列可引入扩增产物中的位置。

[0079] 在一些实施方案中，前适配子还包括用于结合测序引物的一个或多个序列。

[0080] 在一些实施方案中，第一和/或第二寡核苷酸引物的一部分需要粘接至由适配子指定的引物结合序列。在一些实施方案中，结合至适配子序列的寡核苷酸引物的一部分的长度是在5与50个之间、10与40个之间，或20与30个之间的核苷酸。在其他实施方案中，结合至适配子序列的寡核苷酸引物的一部分的长度是约10、20、25、30、35、40、45或50个核苷酸中的任何一个，或至少约10、20、30、35、40、45或50个核苷酸中的任何一个。

[0081] 总体上，适配子的双链体寡核苷酸区域尽可能短而不损失功能是有利的。在这种情况下，‘功能’是指在本领域技术人员熟知的用于拓扑异构酶和连接酶催化核酸连接反应的标准反应条件下(例如，在16℃至25℃范围内的温度下在适合于酶的缓冲液中孵育)，双链区域形成稳定双链体，以使得形成适配子的两个链在适配子连接至目标DNA双链体期间保持部分粘接。在通常用于引物延伸或PCR反应的粘接步骤的条件下，适配子的双链体区域并非绝对必需稳定。总体上，优选双链体区域具有20个或更少、15个或更少或10个或更少碱基对长度以便减少此效应。通过包含与标准Watson-Crick碱基对相比展现更强的碱基配对的非天然核苷酸，可增加双链区域的稳定性，并且由此其长度潜在地减少。优选地，但是并非绝对必需地，适配子的两个链在双链体区域中100％互补。应认识到可在双链区域中耐受一个或多个核苷酸错配，只要两个链能够在标准连接条件下形成稳定双链体。用于本发明的适配子总体上包括双链体区域，其形成适配子的‘可连接’端，即，在连接反应中接合至目标DNA双链体的端。在一些实施方案中，适配子的可连接端可为钝的。在其他实施方案中，适配子的可连接端可包括一个或多个核苷酸的短5’或3’突出物，其可存在以有助于/促进连接。在其他实施方案中，适配子包括单一核苷酸突出物。举例来说，单一核苷酸突出物可为胸苷。适配子的可连接端处的5’末端核苷酸应磷酸化以便能够实现与目标多核苷酸上的3’羟基的磷酸二酯键合。

[0082] 寡核苷酸适配子可由常规脱氧核苷酸或核苷酸单元或修饰合成寡核甘酸结构组成。应进一步了解本发明不限于包含常规脱氧核苷酸或核苷酸单元的寡核苷酸适配子组成。寡核苷酸的修饰可在碱基、糖、链的端，或骨架的磷酸基团处产生。碱基或糖的改变必须加以设计以避免干扰对于必要的寡核苷酸碱基配对相互作用起关键作用的氢键合。分子的端和骨架的修饰总体上更容易实现并且这些位点提供以上论述官能团的便利连接点。用于本发明的寡核苷酸的化学修饰磷酸骨架包括甲基膦酸酯、磷酸三酯、硫代磷酸酯和氨基磷酸酯(关于综述，参见Goodchild(1990)Bioconjugate Chemistry 1:165-187)。用于本发明的适当磷酸骨架修饰的选择由适配子或适配子-目标核酸拓扑异构酶连接产物的预定用途来控制。注意事项包括所需化学和生物稳定性和亲脂性质。具体修饰磷酸基团的优势在本领域中是熟知的并且已经详细地综述(参见Goodchild(1990)，同上文。

[0083] 前TOPO适配子的活化

[0084] 前适配子的活化总体上通过在合适条件下将前适配子与位点特异性拓扑异构酶一起孵育来发生，所述条件导致酶在裂解位点处裂解前适配子的双链体DNA，并且共价连接至其中的裂解位点的3’端，从而形成活化TOPO适配子。除了这些要求以外，对于双链体适配子区域中的两个寡核苷酸的数目或组成(即，核苷酸序列)没有限制，除了其必须选择以使得两个寡核苷酸或单一寡核苷酸的寡核苷酸部分粘接并且在拓扑异构酶连接至前适配子期间保持粘接以外。两个寡核苷酸或单一寡核苷酸的寡核苷酸部分的长度和核苷酸组成可

选择以便于避免可能由于将不合需要的内切核酸酶位点并入连接子中所导致的不利效应。示例性活化条件在本领域中为已知的并且可发现于美国专利号5,766,891中，所述专利的内容通过引用并入本文。

[0085] 拓扑异构酶是经由DNA链的断裂和重新接合来修饰DNA的拓扑状态的一种酶(Shuman等人，美国专利号5,766,891，以引用方式并入本文)。拓扑异构酶分类为裂解双链核酸分子的单一链的I型，包括IA型和IB型拓扑异构酶，和裂解核酸分子的两个链的II型拓扑异构酶(旋转酶)。如本文公开，I型和II型拓扑异构酶，以及其催化域和突变体形式，可用于产生本文提供的活化TOPO适配子。II型拓扑异构酶总体上不用于产生重组核酸分子或克隆程序，而IB型拓扑异构酶用于各种程序。

[0086] 在一些实施方案中，活化TOPO适配子通过使IA型拓扑异构酶与前适配子反应来产生。在其他实施方案中，活化TOPO适配子通过使IB型拓扑异构酶与前适配子反应来产生。在其他实施方案中，活化TOPO适配子通过使II型拓扑异构酶与前适配子反应来产生。

[0087] IA型和IB拓扑异构酶裂解ds核酸分子的一个链。通过IA型拓扑异构酶来裂解ds核酸分子在裂解位点处产生5'磷酸和3'羟基，并且IA型拓扑异构酶共价结合至裂解链的5'末端。相比之下，通过IB型拓扑异构酶来裂解ds核酸分子在裂解位点处产生3'磷酸和5'羟基，并且IB型拓扑异构酶共价结合至裂解链的3'末端。IA型拓扑异构酶包括，例如，大肠杆菌拓扑异构酶I和拓扑异构酶III、真核拓扑异构酶II，和太古菌反向旋转酶(参见Berger，Biochim.Biophys.Acta1400:3-18,1998，以引用方式并入本文)。

[0088] IB型拓扑异构酶包括存在于所有真核细胞中的细胞核I型拓扑异构酶和由痘苗和其他细胞痘病毒编码的酶(参见Cheng等人Cell92:841-850,1998，以引用方式并入本文)。真核IB型拓扑异构酶由在酵母、果蝇和哺乳动物细胞包括人细胞中表达的酶来例示(参见Caron和Wang,Adv.Pharmacol.29B:271-297,1994;Gupta等人Biochim.Biophys.Acta 1262:1-14,1995，其分别以引用方式并入本文;还参见Berger,1998，同上文)。病毒IB型拓扑异构酶由通过脊椎动物痘病毒(痘苗、肖普纤维瘤病毒、ORF病毒、鸡痘病毒和触染性软疣病毒)，和昆虫痘病毒(桑灯蛾昆虫痘病毒)产生的酶来例示(参见Shuman，Biochim.Biophys.Acta 1400:321-337,1998;Petersen等人Virology230:197-206,1997;Shuman和Moss,Proc.Natl.Acad.Sci.,USA84:7478-7482,1987;Shuman和Prescott(1990)J Biol Chem265(29):17826-36;Shuman,J.Biol.Chem.269:32678-32684,1994;美国专利号5,766,891;PCT/US95/16099;PCT/US98/12372，其分别以引用方式并入本文;还参见Cheng等人,1998，同上文)。

[0089] II型拓扑异构酶包括例如细菌旋转酶、细菌DNA拓扑异构酶IV、真核DNA拓扑异构酶II，和T偶列噬菌体编码DNA拓扑异构酶(Roca和Wang,Cell 71:833-840,1992;Wang,J.Biol.Chem.266:6659-6662,1991，其分别以引用方式并入本文;Berger1998，同上文)。如同IB型拓扑异构酶，II型拓扑异构酶具有裂解和连接活性。另外，如同IB型拓扑异构酶，底物ds核酸分子可经制备以使得II型拓扑异构酶可在裂解位点处形成与一个链的共价键合。举例来说，小牛胸腺II型拓扑异构酶可裂解含有定位于距5'端三个核苷酸处的5'凹陷拓扑异构酶识别位点的底物ds核酸分子，导致位于裂解位点5'处的三个核酸分子解离并且拓扑异构酶共价结合至ds核酸分子的5'末端(Andersen等人(1991)J Biol Chem 266:9203-9210)。此外，在这类负载II型拓扑异构酶的ds核酸分子与含有3'羟基的第二核酸分子接触

后，II型拓扑异构酶可将序列连接在一起，然后从重组核酸分子中释放。因此，II型拓扑异构酶还可用于执行本发明方法。

[0090] 痘苗IB型拓扑异构酶的突变分析揭示可突变而不影响拓扑异构酶的活性的大量氨基酸残基，并且已经识别为活性所需要的若干氨基酸(Shuman，1998，同上文)。考虑到痘苗拓扑异构酶催化域与其他IB型拓扑异构酶之间共享的较高同源性，和痘苗拓扑异构酶的详细突变分析，应认识到IB型拓扑异构酶和具有各种氨基酸突变的IB型拓扑异构酶的分离催化域可用于本文提供的方法中，并且因而被认为是用于本发明用途的拓扑异构酶。

[0091] 在一方面，携带共有拓扑异构酶识别序列，例如，(T/C)CCTTN(SEQ ID NO:12和13)的前适配子使用痘苗病毒拓扑异构酶I(TOPO-I)来活化。图3示出其中线性前TOPO适配子(301)和(302)通过TOPO-I(303)来活化的过程(300)的实例。参照图3，TOPO-I，例如，痘苗TOPO-I，在由碱基配对垂直线示出的特定五嘧啶序列例如5'-CCCTT-3'(SEQ ID NO:11)处结合并裂解适配子双链体DNA。裂解通过转酯反应来发生，其中CCCTT↓N磷酸二酯受到活性位点酪氨酸(例如，Tyr-274)攻击，导致形成DNA-(3'-磷酸酪氨酸)蛋白质加合物(304)和释放3'突出物(305)。裂解可使用较小的含有CCCTT的寡核苷酸来发生，只要在易分割的磷酸的上游存在至少六个核苷酸和下游的两个核苷酸(Shuman(1991)J Biol Chem 266:11372-79)。在DNA样品双链体(306)的异源受体DNA 5'端存在下，共价结合拓扑异构酶催化各种DNA链转移反应。它可在最初裂解的相同键上重新连接含有CCCTT的链(如在超螺旋DNA的松弛期间所发生)或它可将链连接至异源受体DNA 5'端，从而产生重组核酸分子，即，TOPO适配子-目标DNA双链体-TOPO适配子复合物(307)。当含有CCCTT的底物在易分割的键3'处含有不超过十五个碱基对时，发生不可逆的或"自杀"裂解，因为较短离去链(305)从蛋白质-DNA复合物中解离。自杀中间体可将切割CCCTT链转移至具有自由5'-OH的DNA受体，从而产生分子间连接产物。为了在形成活化TOPO适配子期间确保痘苗拓扑异构酶共价连接至受体双链体DNA(并且防止裂解链的重新连接)，前适配子的双链体链中的第二核苷酸的5'端(图3示出)在其5'端处磷酸化，从而将反应朝向裂解产物驱动。一旦痘苗拓扑异构酶共价连接至适配子并且离去基团从前适配子中分离，反应实际上是定量和不可逆的直到提供受体DNA为止(即，具有5'-OH基团的双链体DNA)。

[0092] 另外，痘苗拓扑异构酶I活化DNA中间体可重新连接至RNA分子的5'-OH端，从而允许快速形成DNA-RNA共价加合物(参见WO98/56943)。因此，除了将活化TOPO适配子连接至双链体DNA以外，在一些实施方案中，本文提供的活化TOPO适配子可适用于将适配子偶合至具有自由5'-OH部分的RNA分子。

[0093] 活化拓扑异构酶适配子序列和前体前TOPO适配子不包括功能序列，所述功能序列通常包含于在宿主生物体中复制样品DNA分子的分子克隆方法中所使用的载体中发现的适配子序列中。"克隆"在本文中是指涉及复制一个分子以产生具有相同DNA分子的细胞群体的方法。分子克隆总体上使用来自两种不同生物体的DNA序列：作为将要克隆的DNA的来源的物种，和充当用于复制重组DNA的活宿主的物种。因此，本文提供的活化topo适配子能够实现样品DNA的与载体无关的操作，例如，测序。在某些实施方案中，不包括于本文提供的适配子序列中的功能序列包括启动子序列、增强子序列、转录起始序列、多聚腺苷酸化信号、内含子序列、翻译启始序列、表位标签序列、整合促进因子序列、mRNA稳定性调控序列、限制性内切酶识别/裂解序列、合成多克隆位点序列，和细胞定位编码序列。

[0094] 使用活化TOPO适配子来制备测序文库

[0095] 通常，将TOPO适配子连接至目标双链体DNA分子为目标DNA提供用于引物，即适配子特异性引物的粘接位点，以便执行所连接目标DNA双链体的扩增和/或测序。因此，TOPO适配子中的寡核苷酸的全部或任何部分可充当用于如本文描述方法中所用一个或多个适配子特异性引物的粘接位点。

[0096] 用于高通量DNA测序的若干新方法(Nature.437,376-380(2005);Science.309,5741,1728-1732(2005))依赖于通用扩增反应，其中DNA样品随机分裂，然后处理以使得不同片段的端都含有相同DNA序列。具有通用端的片段可在单一反应中用一对扩增引物来扩增。

[0097] 与所有扩增技术，例如，在Illumina测序中所使用的固相桥连扩增一样，需要使用正反向扩增引物，所述引物包括'模板特异性'核苷酸序列，所述序列能够在扩增反应的粘接步骤的条件下粘接至将要扩增的模板中的序列，或其互补序列。与引物在扩增反应条件下粘接的模板中的序列在本文中称为'引物结合'序列。WO 98/44151和WO 00/18957所描述的方法的某些实施方案利用'通用'引物来扩增包括可变模板部分的模板，因此需要通过共同或'通用'引物结合序列来扩增侧接5'和3'。'通用'正反向引物包括能够粘接至模板构建体中的'通用'引物结合序列的序列。可变模板部分或'目标'本身可为已知、未知或部分已知序列。此方法具有以下优势：不需要针对将要扩增的每个目标序列来设计特定引物对；相同引物可用于扩增不同模板，只要每个模板通过将相同通用引物结合序列添加至其5'和3'端来修饰。因此，可变目标序列可为所关注的任何DNA片段。类似方法可用于使用一对通用正反向引物来扩增模板(具有已知端的目标)的混合物，诸如多个目标核酸分子或目标核酸分子文库(例如，基因组DNA片段)，只要混合物中的每个模板分子通过添加相同通用引物结合序列来修饰。

[0098] PCR扩增的这些'通用引物'方法是有利的，因为其使得相同或不同、已知或未知序列的多个模板分子能够在单一扩增反应中扩增，所述反应可在承载一对'通用'引物的固体载体上执行。否则，同时扩增不同序列的模板的混合物将需要多个引物对，每个对与混合物中的每个独特模板互补。对于例如通过NGS方法来测序的模板的复杂混合物来说，为每个个别模板产生多个引物对并非可行选项。

[0099] 单一适配子或两个不同适配子可连接至目标DNA双链体。

[0100] 在一些实施方案中，两个不同活化TOPO适配子连接至目标DNA双链体的端。两个不同适配子分别由两个寡核苷酸形成，所述寡核苷酸在其整个长度上基本上彼此互补以提供包括拓扑异构酶识别序列和至少一个寡核苷酸引物结合序列的线性互补适配子。两个不同活化线性适配子中的第一个与第二个活化线性适配子的不同之处在于寡核苷酸引物结合序列。举例来说，参照图1，在一个实施方案中，第一适配子包括优选地与第一寡核苷酸引物，例如，SEQ ID NO:9完全互补的第一寡核苷酸引物结合序列例如SEQ ID NO:7；并且第二适配子包括优选地与第二寡核苷酸引物，例如，SEQ ID NO:10完全互补的第二寡核苷酸引物结合序列例如SEQ ID NO:8。粘接至第一适配子上的第一引物结合序列的第一引物的延伸提供包括与目标DNA双链体的顶部链互补的序列的延伸产物，并且粘接至第二适配子上的第二引物结合序列的第二引物的延伸提供包括与目标DNA双链体的底部链互补的序列的延伸产物。

[0101] 图4示出使用一组活化互补线性适配子(401)和(402)例如图1示出的适配子来制备测序文库的过程(400)。参照图4,第一(401)和第二(402)活化线性TOPO适配子连接至目标DNA双链体(403)的第一和第二端。如本文中别处描述,第一活化适配子包括第一引物结合序列,并且第二活化适配子包括第二引物结合序列。两个引物结合序列彼此不同。第一和第二活化适配子包括结合至3'端的TOPO酶。活化TOPO适配子在DNA连接酶例如T4DNA连接酶或T7DNA连接酶存在下与钝端化和去磷酸化目标DNA双链体(403)一起孵育,所述连接酶与结合至每个适配子的3'端的TOPO酶的活性组合,将适配子共价连接至目标DNA双链体,从而提供不同TOPO适配子-目标DNA-TOPO适配子复合物(404)的文库。将目标DNA的5'端连接至适配子通过适配子的拓扑异构酶来介导,并且目标DNA的3'端的连接通过连接酶来介导。任选地,在一些实施方案中,第一寡核苷酸引物(405)粘接至第一适配子上的第一结合引物序列的至少一部分;并且第二寡核苷酸引物(406)粘接至第二适配子上的第二结合引物序列的至少一部分。每个引物可通过添加核苷酸来延伸,并且延伸产物例如通过PCR来扩增(407),并测序。

[0102] 在一些实施方案中,单一TOPO适配子连接至目标DNA双链体的两端。如本文中别处描述,单一TOPO适配子由包含5'和3'末端区域的寡核苷酸形成,所述末端区域包括双链体干区域和形成单链环的非互补区域,例如,图2示出的发夹适配子。部分互补适配子的双链体干区域包括拓扑异构酶识别序列,和至少一个寡核苷酸引物结合序列,例如,扩增引物结合序列,至少一个引物粘接至所述结合序列并且延伸以提供与每个TOPO适配子-目标DNA双链体构建体的至少一个链互补的延伸产物。在一些实施方案中,第一引物结合序列和第二引物序列通过尿嘧啶来分离,所述尿嘧啶可移除以打开环部分并且使得它可用于扩增例如PCR。

[0103] 图5示出使用TOPO活化部分互补发夹适配子来制备测序文库的示例性过程(500)。举例来说,如图5示出,在图2示出的前适配子的TOPO活化之后,将包括拓扑异构酶识别序列,例如,CCCTT(SEQ ID NO:11),和结合至寡核苷酸的3'端的拓扑异构酶的单一钝端化部分互补活化发夹TOPO适配子寡核苷酸适配子(501),例如,SEQ ID NO:26,连接至目标DNA双链体(502)的每一端。将适配子连接至DNA双链体的端通过将活化发夹适配子在DNA连接酶,例如,T4DNA连接酶或T7DNA连接酶存在下与钝端化和去磷酸化目标DNA双链体(502)一起孵育来获得,所述连接酶与结合至每个适配子的3'端的TOPO酶的活性组合,将适配子共价连接至目标DNA双链体,从而提供TOPO适配子-目标DNA双链体-TOPO适配子复合物(503)的文库。形成部分互补TOPO适配子的单一寡核苷酸包括寡核苷酸例如SEQ ID NO:20,所述寡核苷酸将第一引物结合序列例如SEQ ID NO:21与第二引物序列例如SEQ ID NO:22分离,并且可包括识别位点例如尿嘧啶碱基,所述识别位点用于DNA糖基化酶例如尿嘧啶DNA糖基化酶(UDG),和DNA内切核酸酶例如核酸内切酶VIII。在一些实施方案中,环未例如通过UDG和DNA内切核酸酶来裂解。未裂解环TOPO适配子-目标DNA双链体-TOPO适配子复合物(503)可经受滚动扩增并且使用例如经由Pacific Biosciences NGS平台的实时测序来测序。或者,在一些实施方案中,尿嘧啶碱基通过UDG和内切核酸酶的混合物对DNA进行酶促消化来移除以打开部分互补TOPO适配子的环部分并且使得它可用于延伸和/或扩增(504)。裂解环TOPO适配子-目标DNA复合物可例如使用其他NGS平台包括由Roche、Life Technologies和ABI开发的平台来测序。

[0104] 在一个实施方案中，并且参照图2和5，将第一寡核苷酸引物(505)例如SEQ ID NO:24粘接至第一寡核苷酸引物结合序列(506)例如SEQ ID NO:21，并且延伸以提供延伸产物，所述延伸产物包括第一通用适配子序列(505)例如SEQ ID NO:24，和第二引物结合序列(507)例如SEQ ID NO:23。随后，第二寡核苷酸引物(508)例如SEQ ID NO:25可粘接至第二引物结合序列(507)例如SEQ ID NO:23，并且延伸以提供包括第一和第二通用适配子的序列的产物(509)。因此，产生TOPO适配子-目标DNA-TOPO适配子复合物的文库。任选地，可执行延伸产物的扩增以提供目标DNA(510)的多个拷贝。

[0105] 图9示出使用TOPO活化部分互补“Y形”适配子来制备测序文库的示例性过程(900)。举例来说，如图9示出，在图7或8示出的前适配子的TOPO活化之后，将包括拓扑异构酶识别序列，例如，CCCTT(SEQ ID NO:11)，和结合至寡核苷酸的3’端的拓扑异构酶的单一钝端化部分互补活化Y形TOPO适配子寡核苷酸适配子(901)连接至目标DNA双链体(902)的每一端。将适配子连接至DNA双链体的端通过将活化适配子在DNA连接酶，例如，T4DNA连接酶或T7DNA连接酶存在下与钝端化和去磷酸化目标DNA双链体(902)一起孵育来获得，所述连接酶与结合至每个适配子的3’端的TOPO酶的活性组合，将适配子共价连接至目标DNA双链体，从而提供TOPO适配子-目标DNA双链体-TOPO适配子复合物(903)的文库。

[0106] 在一个实施方案中，并且参照图7、8和9，第一寡核苷酸引物(905)粘接至第一寡核苷酸引物结合序列(906)，并且延伸以提供包括第一通用适配子序列(905)和第二引物结合序列(907)例如SEQ ID NO:23的延伸产物。随后，第二寡核苷酸引物(908)可粘接至第二引物结合序列(907)并且延伸以提供包括第一和第二通用适配子的序列的产物(909)。因此，产生TOPO适配子-目标DNA-TOPO适配子复合物的文库。任选地，可执行延伸产物的扩增以提供目标DNA(910)的多个拷贝。

[0107] 本领域技术人员了解，通常将设计用于适配子介导的PCR扩增和/或测序的适配子特异性引物的寡核苷酸序列以便在用于进行适配子介导的PCR扩增和/或测序的条件下与在双链体连接子中使用的任何特定第二或第三寡核苷酸杂交。

[0108] NGS测序

[0109] 本文提供的TOPO适配子可用于构建供测序基因组DNA和RNA的NGS文库，所述DNA和RNA可为细胞或无细胞的。在长达13年的人类基因组计划完成之后，NGS顺应对于更廉价和更快速测序方法的需求而产生。自从第一个人类基因组序列完成以来，对于更廉价和更快速测序方法的需求极大地增加。此需求推动下一代测序(NGS)的研发。NGS平台执行大规模并行测序，在此期间来自单一样品或多个不同样品的数百万个DNA片段一齐测序。大规模并行测序技术促进高通量测序，允许整个基因组在少于一天内测序。虽然每个NGS平台在如何完成测序的方式上是唯一的，但是这些平台共有类似基本方法，包括模板制备、测序和成像和数据分析(Metzker ML(2010)Sequencing technologies—the next generation.Nat Rev Genet11:31-46)。

[0110] 本文提供的TOPO适配子可用于从DNA样品制备文库以便测序整个基因组、基因组内的目标区域(例如，外显子组测序)、ChIP-seq实验，或PCR扩增子遵循相同一般工作流程。利用活化TOPO适配子来进行NGS分析的文库制备工作流程的实施方案总体上需要：(i)任选分裂目标多核苷酸，(ii)5’端的端修复和去磷酸化，(iii)连接活化TOPO适配子，和(iv)任选地，有限PCR扩增以富集具有连接至两端的TOPO适配子的产物(图4、5和9)。

[0111] 包括可采用本文所述方法的基因组核酸的样品可为生物样品诸如组织样品、生物流体样品或细胞样品和其加工部分。生物流体样品包括作为非限制实例的血液、血浆、血清、汗、眼泪、痰、尿液、耳朵流体、淋巴、间质液、唾液、脑脊液、残迹、骨髓悬浮液、阴道流体、经子宫颈灌洗、大脑流体、腹水、乳、呼吸、肠和生殖泌尿道的分泌物、羊水和白细胞单采样品。在一些实施方案中，来源样品是通过无创程序可容易地获得的样品，例如，血液、血浆、血清、汗、眼泪、痰、尿液、耳朵流体和唾液。优选地，生物样品是外周血样或血浆和血清部分。在其他实施方案中，生物样品是拭子或涂片、活检样本或细胞培养物。在另一个实施方案中，样品是两种或更多种生物样品的混合物，例如，包括生物流体样品、组织样品和细胞培养物样品中的两种或两种以上的生物样品。如本文使用，术语“血液”、“血浆”和“血清”明确包括其组分或加工部分。类似地，当样品从活检、拭子、涂片等获得时，“样品”明确包括从活检、拭子、涂片等得到的加工部分或份。

[0112] 在一些实施方案中，样品可从来源获得，包括但不限于，来自不同个体、相同或不同个体的不同发育阶段、不同患病个体(例如，患有癌症或怀疑患有遗传病症的个体)、正常个体的样品，在个体中的不同疾病阶段获得的样品，从经受疾病的不同治疗的个体获得的样品，来自经受不同环境因子的个体，或具有病变的易患病体质的个体、暴露于病原体诸如传染病因子(例如，HIV)的个体，和作为供体细胞、组织和/或器官的接受者的个体的样品。在一些实施方案中，样品是包括从相同或不同受试者得到的不同来源样品的混合物的样品。举例来说，样品可包括从两个或更多个个体得到的细胞的混合物，如经常在犯罪现场发现。在一个实施方案中，样品是从怀孕女性，例如怀孕妇女获得的母体样品。在此情况下，样品可使用本文描述的方法分析以提供潜在胎儿病症的产前诊断。除非另外规定，否则母体样品包括胎儿和母体DNA，例如，cfDNA的混合物。在一些实施方案中，母体样品是生物流体样品，例如，血样。在其他实施方案中，母体样品是净化cfDNA样品。

[0113] 样品可为未加工生物样品，例如，全血样。来源样品可为部分加工生物样品，例如，已经分级以提供大致上无细胞血浆部分的血液样品。来源样品可为含有净化核酸的生物样品，例如，从基本上无细胞血浆样品得到的净化cfDNA的样品。样品的加工可包括冷冻样品，例如，组织活检样品，固定样品例如福尔马林固定，和包埋样品，例如，石蜡包埋。样品的部分加工包括样品分级，例如，从血液样品获得血浆部分，和分析在常规临床工作期间、在临床试验和/或科学研究情形下收集的样品所需要的其他加工步骤。额外加工步骤可包括分离并净化样品核酸的步骤。净化样品的进一步加工包括例如在测序准备过程中对样品核酸进行必要修饰的步骤。优选地，样品是未经加工或部分加工样品。

[0114] 样品也可从体外培养组织、细胞或含有多核苷酸的其他来源获得。培养样品可从包括但不限于以下的来源获得：保持于不同培养基和条件(例如，pH、压力或温度)下的培养物(例如，组织或细胞)、保持不同时间长度的培养物(例如，组织或细胞)、用不同因子或试剂(例如，候选药物或调节剂)处理的培养物(例如，组织或细胞)，或不同类型组织或细胞的培养物。

[0115] 生物样品可从各种受试者获得，包括但不限于，人和其他生物体，包括哺乳动物、植物或来自受试者的细胞，或微生物(例如，细菌、真菌)。

[0116] 可如本文描述来分析的样品多核苷酸包括基因组细胞DNA、无细胞DNA(cfDNA)、线粒体DNA、RNA和cDNA。针对一些NGS测序平台来制备测序文库要求多核苷酸具有特定范围的

片段大小，并且要求将较大多核苷酸，例如，细胞基因组DNA加以分裂。通过机械手段来分裂多核苷酸分子在C-O、P-O和C-C处裂解DNA骨架产生具有断裂C-O、P-O和/C-C键的钝端和3'-和5'-突出物的异质混合物(Alnemri和Litwack,J Biol Chem 265:17323-17333[1990]；Richards和Boyer,J Mol Biol 11:327-340[1965])，其需要修复以便随后进行制备DNA供测序所需要的酶促反应，例如，连接测序适配子。因此，可需要分裂多核苷酸，例如，细胞基因组DNA。或者，对于使用cfDNA样品来产生测序文库来说，不需要分裂以<300碱基片段的形式存在的cfDNA。

[0117] 一旦起始DNA或cDNA已经分裂，将片段钝化，即端修复和5'去磷酸化。不同于制备测序文库供NGS的大多数方法，本文提供的TOPO适配子不需要样品DNA的3'端加A尾。随后，TOPO适配子共价连接至样品DNA，即，目标DNA双链体，如上所述。存储在TOPO中的能量释放以接合样品DNA的5'OH基团，并且样品DNA的第二链的5'端连接至适配子的第二链的5'-磷酸基团。将TOPO适配子共价连接至样品多核苷酸，例如，DNA可在几分钟内执行。反应耗费少于15分钟、少于10分钟或少于5分钟。典型地，只需要5分钟来将样品DNA共价连接至TOPO适配子。连接步骤之后，TOPO适配子-目标DNA双链体-TOPO适配子复合物可扩增以富集连接产物。在一些实施方案中，扩增步骤包括粘接包括指标序列的扩增引物。在其他实施方案中，扩增步骤通过添加用于固相扩增的与引物互补的序列来完成TOPO适配子。在其他实施方案中，PCR扩增步骤包括添加用于固相扩增的与引物互补的序列，和指标序列。在其他实施方案中，带不同条码的TOPO适配子可与不同DNA样品一起使用。

[0118] 适配子特异性引物内的指标序列可用于将指标引入扩增目标DNA双链体中。因此，每个适配子特异性引物含有识别其相应目标多核苷酸的独特指标。举例来说，从一个样品得到的目标DNA可不同于从不同样品得到的目标DNA。因此，对应于不同样品的多个不同目标多核苷酸的多个适配子特异性引物可用来扩增多个不同目标多核苷酸并且导致将唯一识别的指标并入由此产生的多个扩增子内的每一个所得扩增子种类。

[0119] 指标可为可区别于其他指标的独特核苷酸序列。它也可通过目标多核苷酸内的序列或位置而可区别于多个多核苷酸内的其他核苷酸序列。核苷酸指标可为随机或专门设计的核苷酸序列。指标可为任何所需序列长度，只要它具有足够长度以在群体中的多个指标内和/或在所分析或询问的多个多核苷酸内成为独特核苷酸序列。在一些实施方案中，指标是约6个至约30个或约8个至约30个范围内的核苷酸的多核苷酸或多核苷酸内的区域。指标可为例如约10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个或更长核苷酸中的任何一个。举例来说，指标可为约35、40、45或50个或更长核苷酸中的任何一个。

[0120] 图1示出被设计用来将指标并入目标多核苷酸中的示例性适配子特异性引物。扩增使用适配子特异性正向引物，即由在图1中示出，并且包括指标[i5]的引物(104)来例示的正向引物，和适配子特异性反向引物，即，由如在图1中示出，并且包括如其中示出的指标[i7]的引物(106)来例示的反向引物将指标并入扩增目标DNA双链体中。

[0121] 不包括将样品DNA加dA-尾的步骤，并且使用拓扑异构酶来将TOPO适配子共价连接至样品DNA显著减少制备测序文库所需要的时间。在一些实施方案中，不包括PCR扩增时间，根据本文提供的方法从DNA样品制备测序文库所需要的时间是约1.5小时、约2小时或约1.5至约2小时，显著少于使用只采用连接酶来连接至目标DNA双链体的标准寡核苷酸适配子来

制备测序文库所耗费的时间。举例来说，不包括PCR扩增所耗费的时间，Illumina TruSeq Nano DNA文库制备耗费约3.5小时。

[0122] 可用于验证来源样品的完整性的测序方法包括下一代测序技术，允许多个样品个别地测序(即，单重测序)或在单一测序运作中作为指标目标DNA分子的汇集样品(即，多重测序)，并且产生DNA序列的多达几亿次读取。目标核酸，和指标目标核酸的序列可使用下一代测序技术(NGS)来确定，其中克隆扩增DNA模板或单一DNA分子相应地以大规模并行方式来测序(例如Voelkerding等人Clin Chem55:641-658[2008];Metzker M,Nature Rev 11:31-46[2010]所描述)。NGS技术有时细分为第一、第二和第三代测序(Pareek和Smoczynski,J Appl Genetics 52:413-435[2011])。除了高通量序列信息以外，NGS提供定量信息，因为每个序列读取可为可计数的“序列标签”，代表个别克隆DNA模板或单一DNA分子。NGS的测序技术包括但不限于焦磷酸测序、使用可逆染料终止子的合成测序、通过寡核苷酸探针连接来测序和离子半导体测序。

[0123] 涉及下一代测序技术中的一般可适用于所有当前技术的主要步骤是文库选择/构建、制备文库供测序和大规模并行测序。

[0124] 市售并且可用于测序TOPO适配子-目标DNA复合物的一些测序技术在以下描述。

[0125] 在一个实施方案中，本方法可应用于使用454测序的生物测定(http://www.454.com/)(例如，如Margulies,M.等人Nature 437:376-380[2005]所描述)。454的总体方法是基于焦磷酸测序的。测序制备开始于使用具有适配子序列的PCR引物或通过连接来产生的在任一端具有适配子的DNA(例如，扩增子或雾化基因组/宏基因组DNA)的长度;这些DNA固定至悬浮于油包水乳液中的微小珠粒(理想地，一个珠粒具有一个DNA片段)。然后，执行乳液PCR步骤以产生每个DNA片段的多个拷贝，由此产生一组珠粒，其中每个珠粒含有相同DNA片段的许多克隆拷贝。然后，将填充有微孔场的光纤芯片，被称为PicoTiterPlate，用乳液冲刷，从而使单一珠粒落入每个孔中。孔还以用于测序过程的一组酶(例如，DNA聚合酶、ATP硫酸化酶和荧光素酶)填充。此时，随着添加触发焦磷酸释放的碱基，可开始合成测序，所述焦磷酸释放产生闪光，记录这些闪光以推断在添加每个碱基类型(A、C、G、T)时每个孔中的DNA片段的序列。

[0126] 在另一个实施方案中，本方法可应用于使用Illumina测序器的生物测定。Illumina生产最广泛使用的平台家族。此技术在2006年引入(www.illumina.com)并且迅速被许多研究人员接受，因为更大量数据可以更成本有效方式产生。Illumina测序是合成测序方法，其在以下两个主要方面不同于454:(1)它使用具有附接寡核苷酸场的流槽，代替含有具有珠粒的个别微孔的芯片，并且(2)它不涉及焦磷酸测序，而是实际上涉及可逆染料终止子。染料终止方法类似于“传统”桑格测序。然而，它与桑格的不同之处在于染料终止子是可逆的，因此其在每个成像循环之后移除以让路给下一个可逆染料终止核苷酸。测序制备开始于冲刷流槽的在任一端具有特异性适配子的DNA的长度，所述流槽用与片段端杂交的特异性寡核苷酸填充。然后复制每个片段以产生相同片段的集群。然后，将可逆的染料终止子核苷酸冲刷流槽并且给予附接时间;过量核苷酸被冲刷掉，将流槽成像，并且将终止子逆转以使得过程可重复并且可继续在随后循环中添加核苷酸。

[0127] 在另一个实施方案中，本方法可应用于使用Applied Biosystems SOLiD过程的生物测定(http://solid.appliedbiosystems.com)。SOLiD过程开始于与由454使用的PCR步

骤类似的乳液PCR步骤，但是测序本身完全不同于以前描述的系统。测序涉及多回合、交错、二碱基并入系统。DNA连接酶用于并入，使得它成为“连接测序”方法，与以前提到的“合成测序”方法相对。Mardis(Mardis ER.,Next-generation DNA sequencing methods,Annu Rev Genomics Hum Genet2008;9:387-402)提供涉及使用此系统的复杂测序和解码过程的全面概述。

[0128] 在另一个实施方案中，本方法可应用于使用Ion Torrent系统的生物测定(http://www.iontorrent.com/)。Ion Torrent系统以与454相似的方式开始，使用含有DNA片段附接的珠粒的微孔板。然而，它在检测碱基并入的方式上不同于所有其他系统。当碱基添加至生长DNA链时，质子释放，从而轻微改变周围pH。对于pH敏感的微量检测器与板上的孔相关联，所述板本身是半导体芯片，并且当这些变化发生时，所述检测器进行记录。当不同碱基(A、C、G、T)依序冲刷过时，记录添加，从而允许推断来自每个孔的序列。

[0129] 在另一个实施方案中，本方法可应用于使用PacBio单一分子、实时测序方法的生物测定(http://www.pacificbiosciences.com/)。PacBio测序系统不涉及扩增步骤，使它有别于其他主要下一代测序系统。测序在含有许多零模式波导(ZMW)检测器的芯片上执行。DNA聚合酶附接至ZMW检测器并且当合成DNA链时，荧光标记的染料示踪核苷酸并入得以即时成像。PacBio的RS II C2XL当前提供最大的读取长度(平均约4,600个碱基)和每次运作最高数目的读取(约47,000)。典型“成对端”方法不用于PacBio，因为读取通常足够长以致于可经由CCS来覆盖片段多次而无需独立地从每个端测序。使用PacBio的多重测序不涉及独立读取，而是实际上遵循标准“内联”条码模型。

[0130] 在另一个实施方案中，本方法可应用于使用纳米孔测序的生物测定(例如，如Soni G V and Meller A.,Clin Chem 53:1996-2001[2007]所描述)。纳米孔测序DNA分析技术在工业上由许多公司开发，包括Oxford Nanopore Technologies(Oxford,United Kingdom)、Roche和Illumina。纳米孔测序是单一分子测序技术，其中当DNA的单一分子穿过纳米孔时直接对其测序。纳米孔测序是直接核苷酸询问测序的实例，其中当链穿过检测器时，测序过程直接检测核酸链的碱基。纳米孔是大约1纳米直径的小孔，将纳米孔浸没于导电流体中并且越过其施加电位(电压)导致由于离子传导穿过纳米孔而产生轻微电流。流动的电流的量对于纳米孔的大小和形状是敏感的。当DNA分子穿过纳米孔时，DNA分子上的每个核苷酸不同程度地阻碍纳米孔，不同程度地改变经过纳米孔的电流的幅值。因此，当DNA分子穿过纳米孔时电流的此变化代表DNA序列的读取。直接核苷酸询问测序的另一个实例是Halcyon的测序。

[0131] 实施例

[0132] 本发明在以下实施例中进一步解释描述，所述实施例不希望以任何方式限制要求保护的本发明范围。附图意图被看作是本发明的说明书和描述的完整部分。叙述的所有参考文献针对其中描述的所有内容特定地以引用方式并入本文。提供以下实施例来说明，但并非限制所要求保护的发明。

[0133] 实施例1

[0134] 制备拓扑异构酶活化测序适配子

[0135] 活化拓扑异构酶适配子(TOPO适配子)通过将合成寡核苷酸(SEQ ID NO:1和2)，和(SEQ ID NO:4和5)杂交来制备。适配子组的第一适配子通过将第一寡核苷酸

ACACTGTTTCACGACAGGTGTTGATCCCTTATTCCGATAGTG(SEQ ID NO:1)与第二寡核苷酸AAGGGCGATCAACACCTGTCGTGAAACAGTGT(SEQ ID NO:2)杂交来制备。适配子组的第二适配子通过将第三寡核苷酸AAGGGGTGACTGGAGTTCAGACGTGTGCTATC(SEQ ID NO:4)与第四寡核苷酸GATAGCACACGTCTGAACTCCAGTCACCCCTTATTCCGATAGTG(SEQ ID NO:5)杂交来制备。寡核苷酸的杂交为每个适配子提供单一拓扑异构酶识别序列/位点CCCTT(SEQ ID NO:11)。寡核苷酸(10μM)在10mM Tris-HCl(pH 7.5)、160mM NaCl中杂交并且在以下循环的热循环仪中扩增:98℃5分钟、85℃5分钟、72℃5分钟、65℃5分钟、55℃5分钟、42℃5分钟、37℃15分钟、20℃30分钟,然后保持在20℃。

[0136] 拓扑异构酶通过将以下物质混合至200μl的最终反应体积来结合:50μl粘接寡核苷酸、50μl的1x浓度下的5x TOPO结合缓冲液:20mM Tris-醋酸盐、pH 7.9@25℃、50mM醋酸钾、10mM醋酸镁、100μg/ml BSA、2.3mM ATP、100μg痘苗拓扑异构酶I(31.25pmol/μg;500单位/pmol)、46单位T4多核苷酸激酶(MCLAB产品号T4PK)和H_2O。在37℃下孵育90分钟之后,活化拓扑异构酶适配子通过HPLC来分离。

[0137] 如图3示出,适配子的活化产生具有连接至3'端的拓扑异构酶的钝端-双链适配子。活化第一适配子的顶部和底部链分别具有序列SEQ ID NO:3和2,并且活化第二适配子的顶部和底部链分别具有序列SEQ ID NO:4和6。

[0138] 实施例2

[0139] 使用拓扑异构酶活化适配子来制备测序文库

[0140] 为了证明在制备测序文库供大规模并行测序中使用TOPO适配子的优势,相等量的片段样品DNA用于根据本公开所提供的基于拓扑异构酶的方法来制备测序文库,并且与使用Illumina的仅基于连接酶的方法来制备文库的Illumina方法相比并行地制备测序文库。

[0141] DNA样品制备

[0142] 使用Covaris M220聚焦超声发生器将Lambda DNA或人基因组DNA剪切成大部分350bp的片段。DNA样品(10-500ng)在20℃下在含有10μl的5X端修复缓冲液(1x浓度:20mM Tris-醋酸酯,在25℃下pH 7.9,50mM醋酸钾,10mM醋酸镁,100μg/ml牛血清清蛋白(BSA))的50μl反应混合物中端修复15分钟,所述反应混合物含有0.2-0.6mM dNTP,3μl端修复酶(T4DNA聚合酶,MCLAB产品号T4DP,9单位)。端修复酶在75℃下热灭活10分钟。DNA的5'磷酸基团通过在50℃下添加3μl(30单位)碱性磷酸酶(小牛肠磷酸酶(CIP))10分钟而去磷酸化。

[0143] 端修复去磷酸化样品DNA通过将53μl反应混合物与80μl MCMag™净化珠粒(MCLAB)(在在此实施例中描述的条件下结合至带负电荷DNA的磁性珠粒)一起孵育来净化。DNA样品与珠粒一起孵育5分钟,并且珠粒结合DNA用乙醇冲刷两次。允许珠粒在室温下干燥。然后,在洗脱缓冲液(0.1X TE缓冲液)中,DNA从珠粒洗脱5分钟。1X TE缓冲液是10mM Tris-HCl,pH 7.5,1mM EDTA。

[0144] 样品DNA测序文库的基于拓扑异构酶的制备

[0145] 十五微升的净化端修复和去磷酸化样品DNA与25μl反应混合物中的根据实施例1制备的3μl活化拓扑异构酶适配子、2μl的T7连接酶、1μl ATP、2.5μl 10X连接缓冲液一起孵育。将反应混合物在室温下孵育5分钟。适配子-DNA-适配子复合物通过将25μl复合物与30μl结合缓冲液和MCMag™净化珠粒(MCLAB)一起孵育5分钟来净化。样品用80%乙醇冲刷两次,然后在25μl洗脱缓冲液中洗脱。适配子-样品复合物的数量和质量在生化分析仪上验证。

[0146] 有限循环扩增

[0147] 将二十三微升的净化适配子-DNA样品复合物与25μl的PCR预混液(KAPA文库扩增试剂盒或2X MCAmp™文库扩增预混液(MCLAB;产品号LIBA-50或LIBAP-50)以及含有正向引物AATGATACGGCGACCACCGAGATCTACACACACTGTTTCACGACAGGTGTTGATCG(SEQ ID NO:9)和

[0148] 反向引物

[0149] CAAGCAGAAGACGGCATACGAGATGATAGCACACGTCTGAACTCCAGTCAC(SEQ ID NO:10)的2μl的15μM PCR引物混合物混合。图1示出SEQ ID NO:9的PCR引物与第一适配子的第二链上的第一引物结合序列(SEQ ID NO:7)互补,并且SEQ ID NO:10的PCR引物与第二适配子的第二链上的第二引物结合序列(SEQ ID NO:8)互补。将DNA样品扩增6-10个循环。

[0150] 扩增产物(35μl)结合至净化珠粒(42μl;在25μl的0.1X TE缓冲液中冲刷并洗脱的MCMag™净化珠粒。PCR产物通过琼脂糖凝胶电泳和/或使用生化分析仪来分析。

[0151] 样品DNA测序文库的仅基于连接酶的制备

[0152] 将一百纳克的净化端修复和3'腺苷酸化样品DNA与Illumina的TruSeq适配子一起孵育,并且根据Illumina的方案使用连接酶(由Illumina的TruSeq Nano DNA LT文库制备试剂盒提供)连接。连接执行10分钟。相比之下,如上所述制备的100ng净化端修复和去磷酸化样品DNA与2微升净化TOPO适配子和40单位连接酶一起孵育,并且允许连接进行5分钟。使用本公开提供的基于拓扑异构酶的方法所获得的连接产物,即,适配子-样品DNA复合物的数量大于使用Illumina的仅基于连接酶的方法所获得的连接产物的数量。图6中示出的示例性数据示出与根据Illumina的仅基于连接酶的方法所获得的数量相比,基于拓扑异构酶的方法产生多3倍的适配子-样品复合物产物。两种文库产物使用Illumina的MiSeq测序器来测序。对于两种文库,获得样品DNA的100%覆盖度。

[0153] 另外,基于TOPO的方法可在制备Illumina文库所耗费的约一半时间内执行。基于TOPO的文库的制备时间是120分钟,而在使用商用文库制备时的制备时间在155与210分钟之间。

[0154] 因此,与广泛用于制备文库供大规模并行测序的金标准仅基于连接酶方法相比,测序文库的基于拓扑异构酶的制备提供更有效方法和文库产物的更大产率。

[0155] 实施例3

[0156] DNA样品制备

[0157] 使用Covaris M220聚焦超声发生器将Lambda DNA或人基因组DNA剪切成大部分350bp的片段。DNA样品(10-500ng)在50℃下在含有10X端修复缓冲液和5μl碱性磷酸酶(虾碱性磷酸酶(5单位)或南极磷酸酶(25单位))的70μl反应混合物中去磷酸化25分钟。将碱性磷酸酶在75℃下热灭活10分钟。去磷酸化样品DNA在72℃下通过添加端修复缓冲液、dNTP s和2.5μl(10单位)端修复酶(Pfu DNA聚合酶和/或KOD DNA聚合酶)的8μl混合物来端修复5分钟。

[0158] 样品DNA测序文库的基于拓扑异构酶的制备

[0159] 去磷酸化和端修复样品DNA与100μl反应混合物中的根据实施例1制备的2μl活化拓扑异构酶适配子、2μl连接酶(诸如T4 DNA连接酶或T7DNA连接酶)、1μl ATP、10X连接缓冲液一起孵育。反应混合物在16℃下孵育15分钟。适配子-DNA-适配子复合物通过将100μl复合物与160μl稀释MCMag净化珠粒(MCLAB)一起孵育5分钟来选择大小。上清液通过与30μl

MCMag净化珠粒(MCLAB)一起孵育5分钟来净化。样品用80％乙醇冲刷两次，然后在25μl洗脱缓冲液中洗脱。适配子-样品复合物的数量和质量在生化分析仪上验证。

[0160] 有限循环扩增

[0161] 将二十三微升净化适配子-DNA样品复合物与25ul的PCR预混液(诸如KAPA文库扩增试剂盒或2X MCAmp文库扩增预混液(MCLAB))和含有单一指标适配子的正向引物

[0162] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA(SEQ ID NO:27)

[0163] 反向引物

[0164] CAAGCAGAAGACGGCATACGAGAT(SEQ ID NO:29)的2μl的15μM PCR引物混合物混合。将DNA样品扩增6-10个循环。

[0165] 扩增产物(35μl)结合至净化珠粒(42μl;MCMag™DNA

[0166] 文库净化珠粒如上所述在25μl的0.1XTE缓冲液中冲刷并洗脱。PCR产物通过琼脂糖凝胶电泳和/或使用生化分析仪来分析。

[0167] 应该理解本文所描述的实施例和实施方案仅出于说明目的，并且将建议本领域技术人员根据它们进行各种修改或变化，并且它们被包括在本申请的精神和范围以及随附权利要求书的范围之内。出于所有目的，本文所引用的所有出版物、专利和专利申请特此以引用的方式整体并入。

[0168] 不含序列表的文本

[0169] SEQ ID NO:1-第一线性前TOPO适配子的寡核苷酸1(顶部链)

[0170] ACACTGTTTCACGACAGGTGTTGATCCCTTATTCCGATAGTG

[0171] SEQ ID NO:2-第一线性前TOPO适配子的寡核苷酸2(底部链)

[0172] AAGGGCGATCAACACCTGTCGTGAAACAGTGT

[0173] SEQ ID NO:3-第一线性活化TOPO适配子的寡核苷酸1(顶部链)

[0174] ACACTGTTTCACGACAGGTGTTGATCCCTT

[0175] SEQ ID NO:4-第二线性前TOPO适配子的寡核苷酸1(顶部链)

[0176] AAGGGGTGACTGGAGTTCAGACGTGTGCTATC

[0177] SEQ ID NO:5-第二线性前TOPO适配子的寡核苷酸2(底部链)

[0178] GATAGCACACGTCTGAACTCCAGTCACCCCTTATTCCGATAGTG

[0179] SEQ ID NO:6-第二线性活化TOPO适配子的寡核苷酸1(底部链)

[0180] GATAGCACACGTCTGAACTCCAGTCACCCCTT

[0181] SEQ ID NO:7-第一引物结合序列

[0182] CGATCAACACCTGTCGTGAAACAGTGT

[0183] SEQ ID NO:8-第二引物结合序列

[0184] GTGACTGGAGTTCAGACGTGTGCTATC

[0185] SEQ ID NO:9-第一寡核苷酸引物

[0186] AATGATACGGCGACCACCGAGATCTACACACACTGTTTCACGACAGGTGTTGATCG

[0187] SEQ ID NO:10-第二寡核苷酸引物

[0188] CAAGCAGAAGAGGGCATACGAGATGATAGCCACAGTCTGAACTCCAGTCAC

[0189] SEQ ID NO:11-图1和2的适配子中的Topo识别序列

[0190] CCCTT

[0191] SEQ ID NO:12-拓扑异构酶I的识别序列-共有序列加下划线-N是任何核苷酸

[0192] CCCTTN

[0193] SEQ ID NO:13-拓扑异构酶I的识别序列-共有序列加下划线-N是任何核苷酸

[0194] TCCTTN

[0195] SEQ ID NO:14-拓扑异构酶I的识别序列-共有序列加下划线

[0196] GCCCTTATTCCC

[0197] SEQ ID NO:15-拓扑异构酶I的识别序列-共有序列加下划线

[0198] TCGCCCTTATTC

[0199] SEQ ID NO:16-拓扑异构酶I的识别序列-共有序列加下划线

[0200] TGTCGCCCTTAT

[0201] SEQ ID NO:17-拓扑异构酶I的识别序列-共有序列加下划线

[0202] GTGTCGCCCTTA

[0203] SEQ ID NO:18-拓扑异构酶I的识别序列-共有序列加下划线

[0204] GATTCCCCTTATTCCGATAGTG

[0205] SEQ ID NO:19-拓扑异构酶I的识别序列-共有序列加下划线

[0206] AAGGGCGATC

[0207] SEQ ID NO:20-发夹前TOPO适配子的寡核苷酸

[0208] AAGGGCGATCAAGAGTTCAGACGTGTGCTATCUACACTGTTTCACGACAGGTGTTCATCGCCCTTATTCCGATAGTG

[0209] SEQ ID NO:21-发夹适配子中的第一引物结合序列

[0210] CTATCGTGCAGACTTGAG

[0211] SEQ ID NO:22发夹环中的第二引物序列

[0212] ACACTGTTTCACGACAGGTG

[0213] SEQ ID NO:23-发夹适配子中的第二引物结合序列通过延伸作为SEQ ID NO:22的反向互补序列的第一引物来产生

[0214] TCAAGACCTGTACTGAAACAGTGT

[0215] SEQ ID NO:24-发夹适配子的第一寡核苷酸引物

[0216] CAAGCAGAAGACGGCATACGAGATGATGATAGCACACGTCTGAACTCTTGA

[0217] SEQ ID NO:25-发夹适配子的第二寡核苷酸引物

[0218] AATGATACGACCACCGAGTCTACACACACTGTTTCACGACAGGTGTTGA

[0219] SEQ ID NO:26-活化发夹适配子的序列

[0220] AAGGGCGATCAAGAGTTCAGACGTGTGCTATCUACACTGTTTCACGACAGGTGTTCATCGCCCTT

[0221] SEQ ID NO:27-单一指标适配子的正向引物

[0222] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA

[0223] SEQ ID NO:28-双重指标适配子的正向引物

[0224] AATGATACGGCGACCACCGAGATCTACAC

[0225] SEQ ID NO:29-单一指标或双重指标适配子的反向引物

[0226] CAAGCAGAAGACGGCATACGAGAT

[0227] SEQ ID NO:30-单一指标适配子的前适配子顶部链

[0228] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCGCGGCCGATGCCCGGAGGCCGATAGTG

[0229] SEQ ID NO:31-单一或双重指标适配子的前适配子底部链

[0230] GTTCGTCTTCTGCCGTATGCTCTACACTGACCTCAAGTCTG CACACGAGAAGGCTACGGGAA

[0231] SEQ ID NO:32-双重指标适配子的前适配子顶部链

[0232] AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATGCCCTTATTCCGATAGTG

[0233] SEQ ID NO:33-单一或双重指标适配子的第一引物结合序列

[0234] GTTCGTCTTCTGCCGTATGCTCTA

[0235] SEQ ID NO:34-单一指标适配子的第二结合序列的互补序列

[0236] TCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

[0237] SEQ ID NO:35-双重指标适配子的第二结合序列的互补序列

[0238] GTGTAGATCTCGGTGGTCGCCGTATCATT

序列表

<110>分子克隆研究室有限公司

<120> 用于制备测序文库的组合物和方法

<130> 05185.001CN1

<140>
<141>

<150> 62/218,906
<151> 2015-09-15

<150> 62/167,892
<151> 2015-05-29

<160> 35

<170> PatentIn 3.5 版

[0001] <210> 1
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 1
acactgtttc acgacaggtg ttgatcccct attccgatag tg 42

<210> 2
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 2
aagggcgatc aacacctgtc gtgaaacagt gt 32

<210> 3
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 3
acactgtttc acgacaggtg ttgatccctt 30

<210> 4
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 4
[0002] aaggggtgac tggagttcag acgtgtgcta tc 32

<210> 5
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 5
gatagcacac gtctgaactc cagtcacccc ttattccgat agtg 44

<210> 6
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 6
gatagcacac gtctgaactc cagtcacccc tt 32

<210> 7
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 7
cgatcaacac ctgtcgtgaa acagtgt 27

<210> 8
<211> 27
<212> DNA
<213> 人工序列

[0003] <220>
<223> 人工序列的描述：合成引物

<400> 8
gtgactggag ttcagacgtg tgctatc 27

<210> 9
<211> 56
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 9
aatgatacgg cgaccaccga gatctacaca cactgtttca cgacaggtgt tgatcg 56

<210> 10
<211> 51
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 10
caagcagaag agggcatacg agatgatagc cacagtctga actccagtca c 51

<210> 11
<211> 5
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 11
ccctt 5

[0004] <210> 12
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (6)..(6)
<223> a、c、t、g、未知或其它

<400> 12
cccttn 6

<210> 13
<211> 6
<212> DNA
<213> 人工序列

〈220〉
〈223〉 人工序列的描述：合成寡核苷酸

〈220〉
〈221〉 修饰的碱基
〈222〉 (6)..(6)
〈223〉 a、c、t、g、未知或其它

〈400〉 13
tccttn 6

〈210〉 14
〈211〉 12
〈212〉 DNA
〈213〉 人工序列

〈220〉
〈223〉 人工序列的描述：合成寡核苷酸

[0005]
〈400〉 14
gcccttattc cc 12

〈210〉 15
〈211〉 12
〈212〉 DNA
〈213〉 人工序列

〈220〉
〈223〉 人工序列的描述：合成寡核苷酸

〈400〉 15
tcgcccttat tc 12

〈210〉 16
〈211〉 12
〈212〉 DNA
〈213〉 人工序列

〈220〉
〈223〉 人工序列的描述：合成寡核苷酸

<400> 16
tgtcgccctt at 12

<210> 17
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 17
gtgtcgccct ta 12

<210> 18
<211> 22
<212> DNA
<213> 人工序列

[0006] <220>
<223> 人工序列的描述：合成寡核苷酸

<400> 18
gattcccctt attccgatag tg 22

<210> 19
<211> 10
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 19
aagggcgatc 10

<210> 20
<211> 77
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<223> 组合 DNA/RNA 分子的描述：合成寡核苷酸

<400> 20
aagggcgatc aagagttcag acgtgtgcta tcuacactgt ttcacgacag gtgttcatcg 60

cccttattcc gatagtg 77

<210> 21
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0007]

<400> 21
ctatcgtgca gacttgag 18

<210> 22
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 22
acactgtttc acgacaggtg 20

<210> 23
<211> 24
<212> DNA
<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 23
tcaagacctg tactgaaaca gtgt 24

<210> 24
<211> 51
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 24
caagcagaag acggcatacg agatgatgat agcacacgtc tgaactcttg a 51

<210> 25
<211> 49
<212> DNA
[0008] <213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 25
aatgatacga ccaccgagtc tacacacact gtttcacgac aggtgttga 49

<210> 26
<211> 65
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<223> 组合 DNA/RNA 分子的描述：合成寡核苷酸

<400> 26
aagggcgatc aagagttcag acgtgtgcta tcuacactgt ttcacgacag gtgttcatcg 60

ccctt 65

<210> 27
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 27
aatgatacgg cgaccaccga gatctacact ctttccctac acga 44

<210> 28
<211> 29
<212> DNA
<213> 人工序列

<220>
[0009] <223> 人工序列的描述：合成引物

<400> 28
aatgatacgg cgaccaccga gatctacac 29

<210> 29
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 29
caagcagaag acggcatacg agat 24

<210> 30
<211> 74
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 30
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgcgcg gccgatgccc 60

ggaggccgat agtg 74

<210> 31
<211> 62
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 31
gttcgtcttc tgccgtatgc tctacactga cctcaagtct gcacacgaga aggctacggg 60

aa 62

[0010]

<210> 32
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 32
aatgatacgg cgaccaccga gatctacaca cactctttcc ctacacgacg ctcttccgat 60

gcccttattc cgatagtg 78

<210> 33
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 33
gttcgtcttc tgccgtatgc tcta 24

<210> 34
<211> 44
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0011]
<400> 34
tcgtgtaggg aaagagtgta gatctcggtg gtcgccgtat catt 44

<210> 35
<211> 29
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 35
gtgtagatct cggtggtcgc cgtatcatt 29

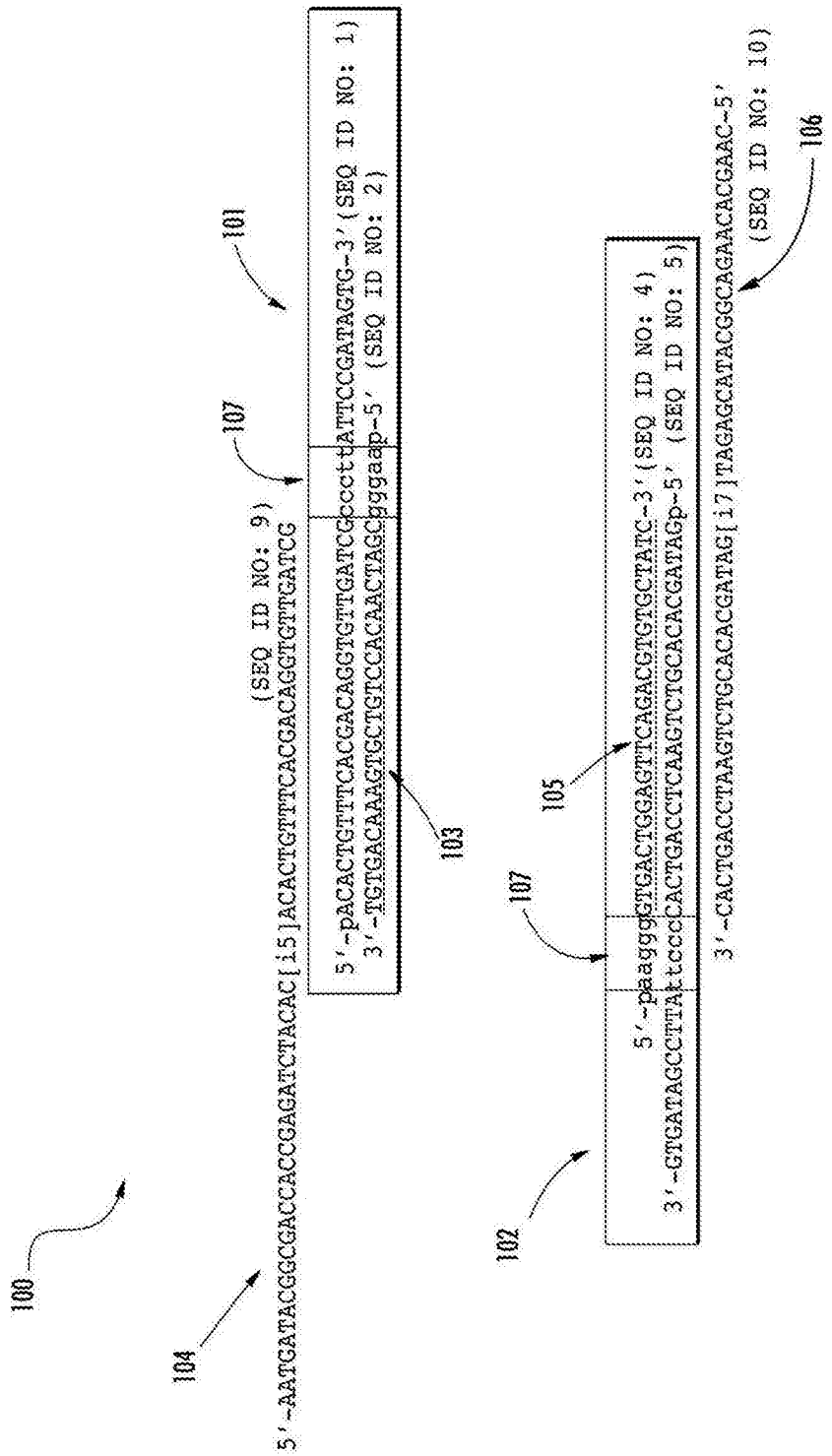

100
101
102
103
104
105
106
107
5′-AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTGTTTCACGACAGGTGTTGATCG
(SEQ ID NO: 9)
5′-pACACTGTTTCACGACAGGTGTTGATCGccttATTCCGATAGTG-3′(SEQ ID NO: 1)
3′-TGTGACAAAGTGCTGTCCACAACTAGCgggaap-5′ (SEQ ID NO: 2)
5′-paagggGTGACTGGAGTTCAGACGTGTGCTATC-3′(SEQ ID NO: 4)
3′-GTGATAGCCTTATtcccCACTGACCTCAAGTCTGCACACGATAGp-5′ (SEQ ID NO: 5)
3′-CACTGACCTAAGTCTGCACACGATAG[i7]TAGAGCATACGGCAGAACACGAAC-5′
(SEQ ID NO: 10)

图1

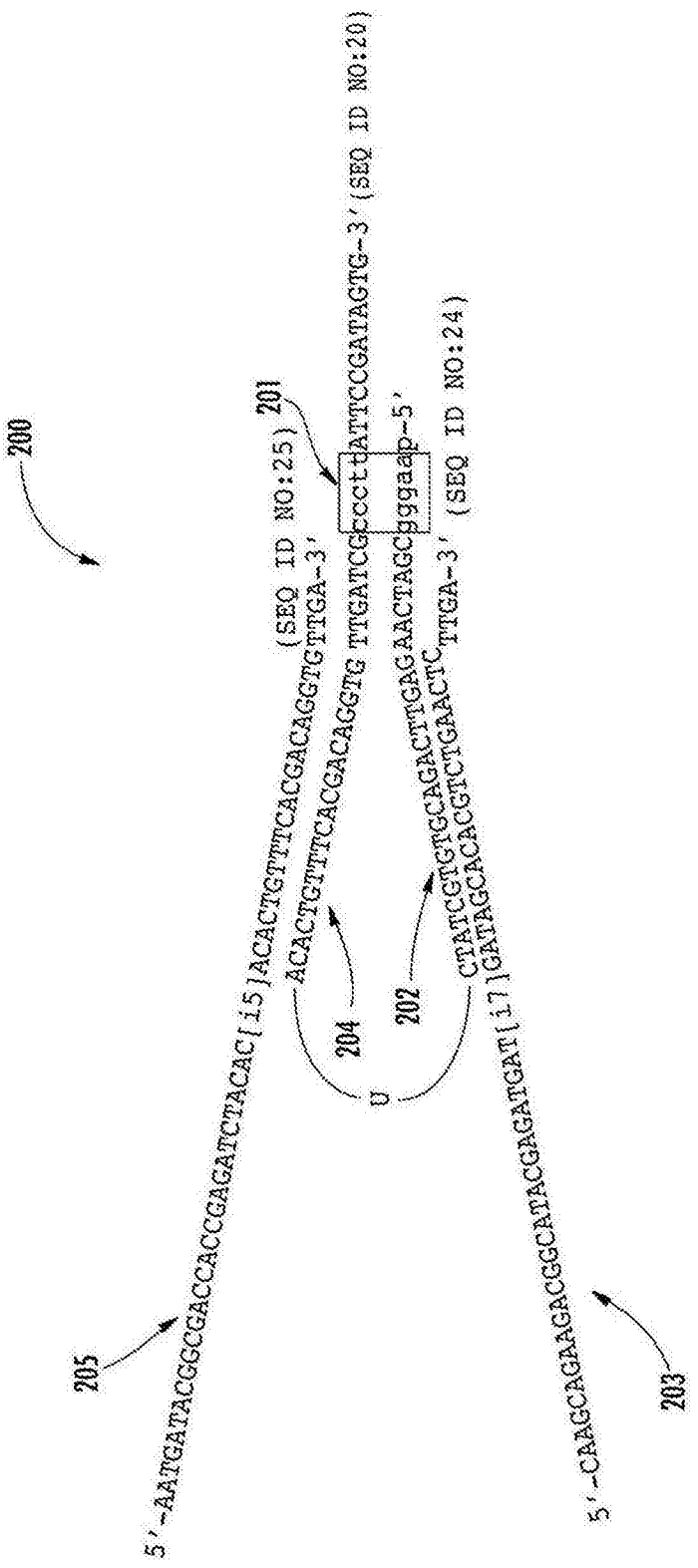

200
201
(SEQ ID NO:25)
ACACTGTTTCACGACAGGTGTTTGA-3'
TTGATCGccttATTCCGATAGTG-3'(SEQ ID NO:20)
204
U
202
CTATCGTGTGCAGACTTGAGAACTAGCgggaap-5'
CTTGAACTC TTGA-3' (SEQ ID NO:24)
205
5'-AATGATACGGCGACCACCGAGATCTACAC[15]ACACTGTTTCACGACAGGTGTTTGA-3'
5'-CAAGCAGAAGACGGCATACGAGATGAT[17]GATAGCACACCGTCTG
203

图2

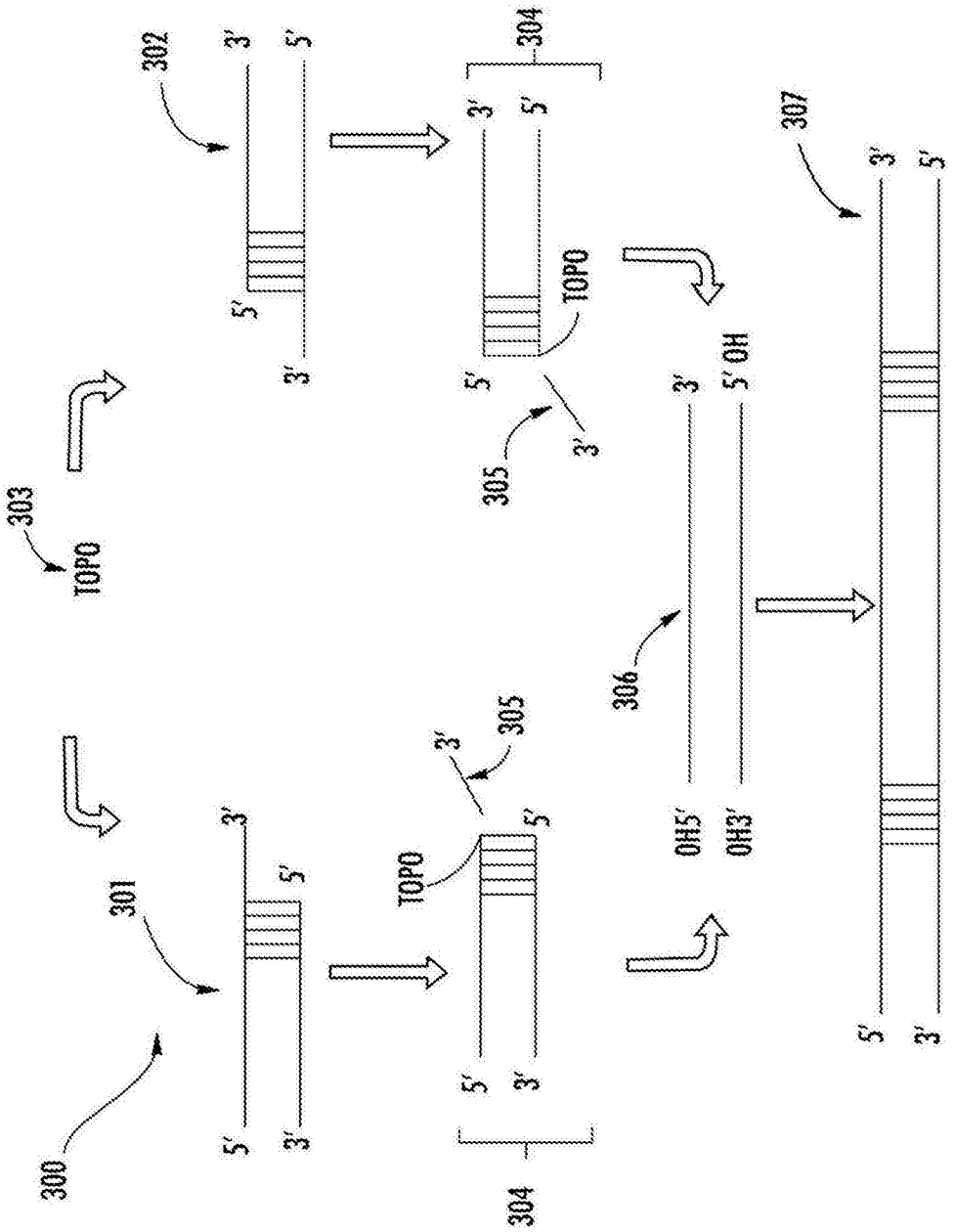

300
301
3'
5'
5'
3'
303
TOPO
302
3'
5'
5'
3'
304
3'
5'
TOPO
5'
3'
305
3'
5'
TOPO
5'
3'
304
305
3'
306
3'
5' OH
OH5'
OH3'
307
3'
5'
5'
3'

图3

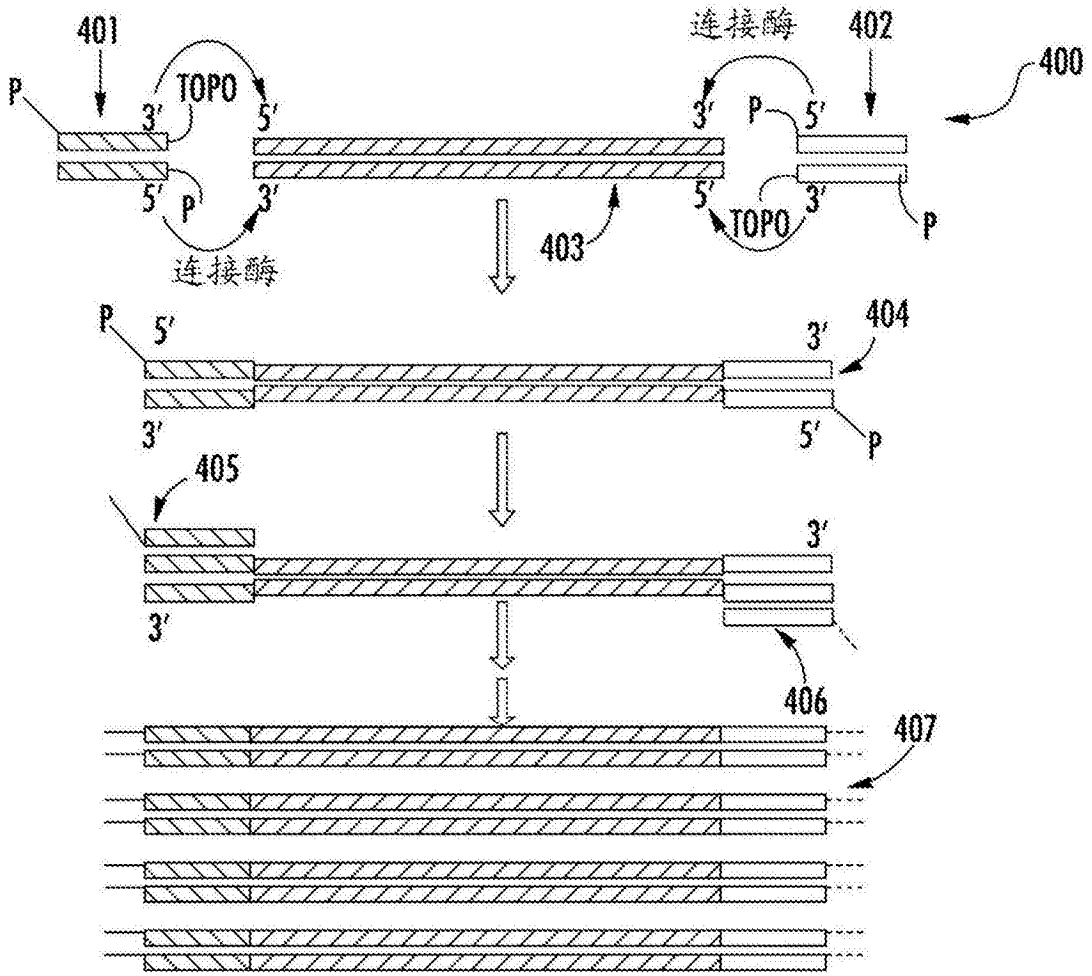


图4

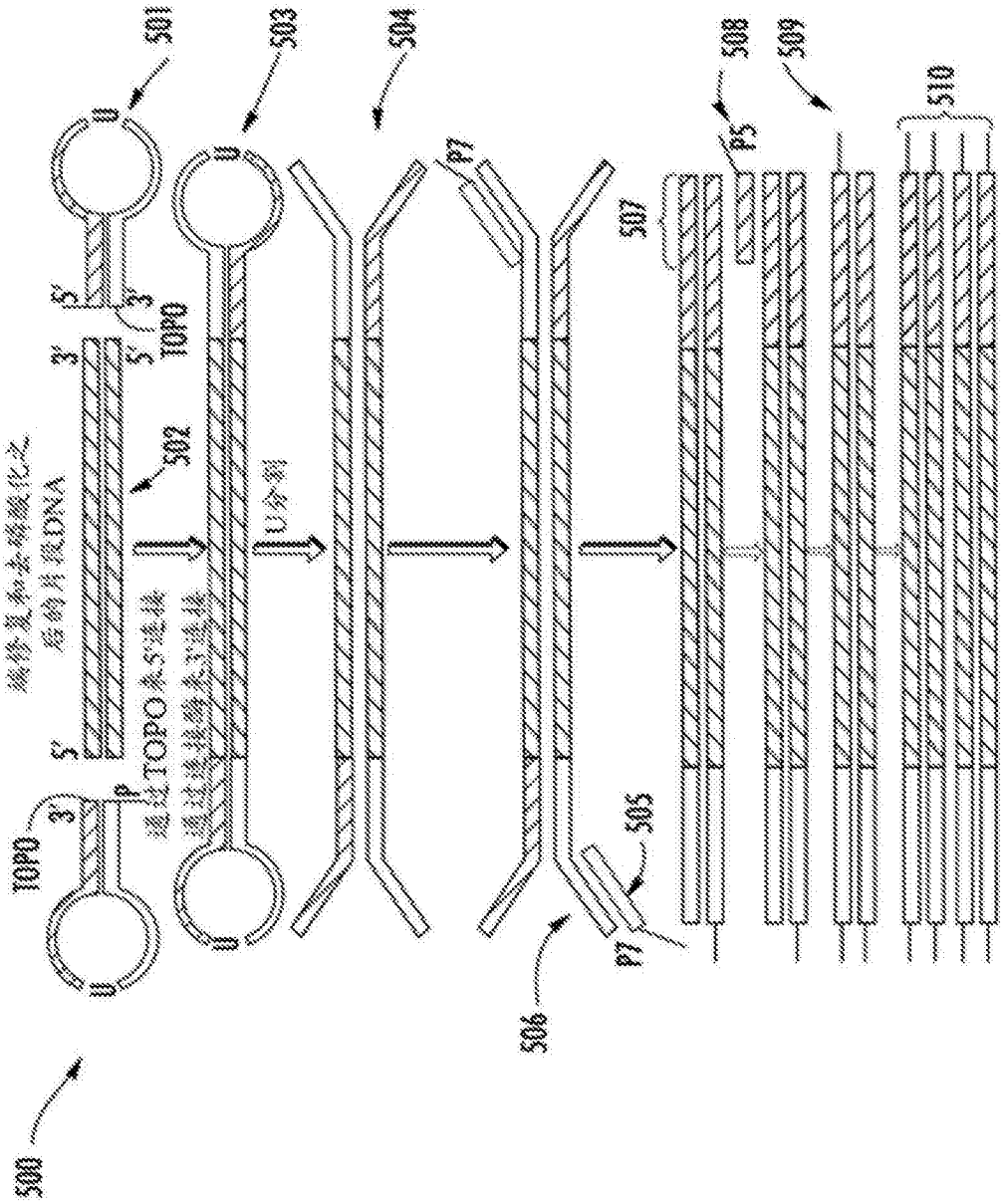
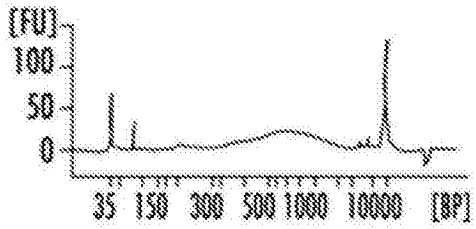

501
503
504
508
509
510
P5
P7
507
TOPO
5'
3'
502
P
TOPO
3'
5'
505
P7
506
500

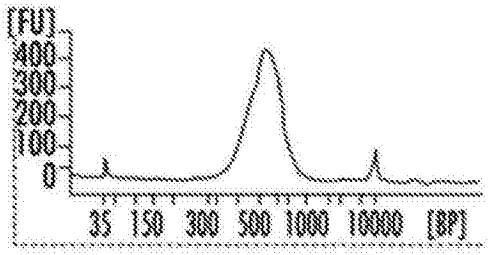
图5

用ILLUMINIA TRUSEQ NANO试剂盒制备的100 ng 350 BP LAMBDA文库



连接之后 (nM)
0.15

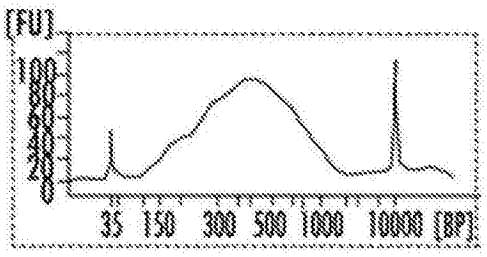
图 6A



扩增之后 (nM)
32.6

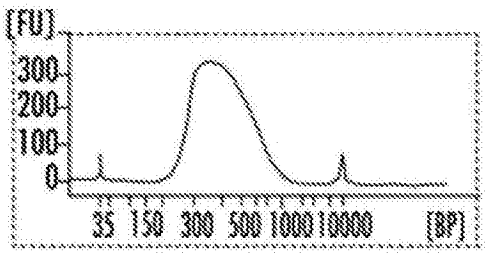
图 6B

用MCLAB TOPO连接文库制备试剂盒制备的100 ng 350 BP LAMBDA文库



连接之后 (nM)
0.5

图 6C



扩增之后 (nM)
50.1

图 6D

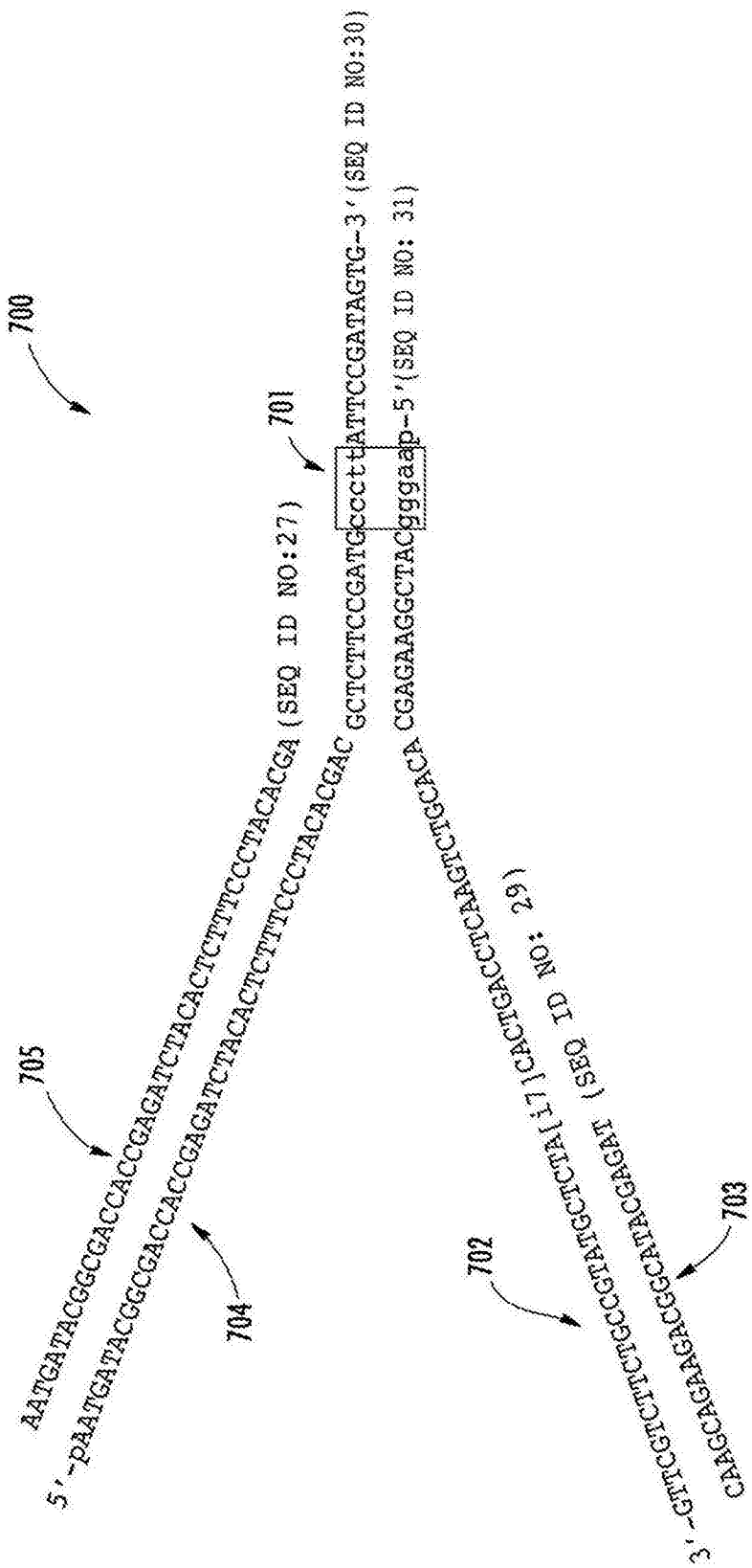

700
701
705
704
702
703
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA (SEQ ID NO:27)
5'-PAATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
GCTCTTCCGATGccctttATTCCGATAGTG-3'(SEQ ID NO:30)
CGAGAAGGCTACgggaaap-5'(SEQ ID NO: 31)
3'-GMTCGTCTTCCTMCGCCGTATGCTCTAT/1/CACTGACCTCAGTCTGCACA (SEQ ID NO: 29)
CAGCAGAAGGACGCATACGAGAT

图7

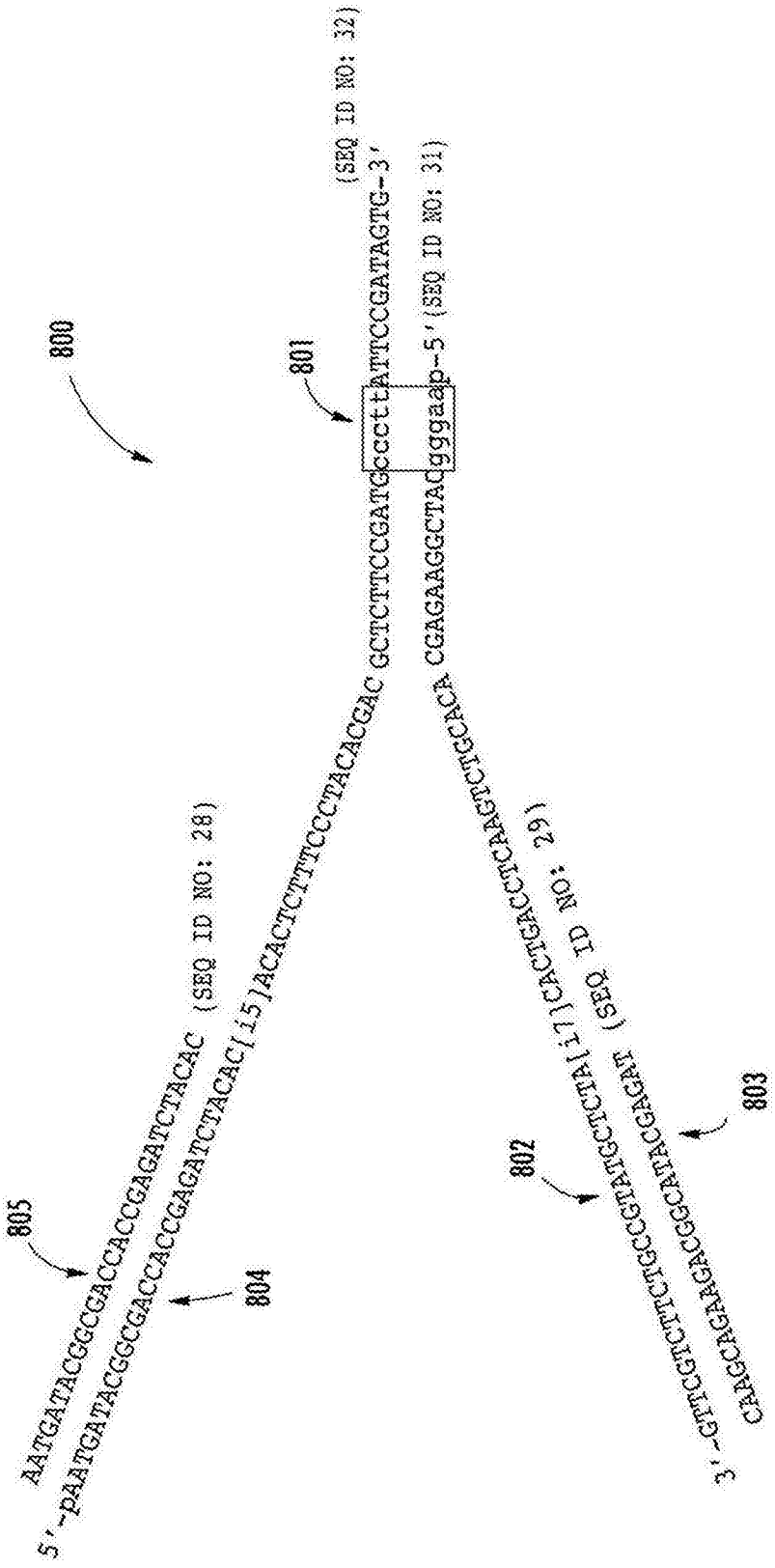

800
801
(SEQ ID NO: 32)
GCTCTTCCGATCccttATTCCGATAGTG-3'
5'-PAATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGAC
(SEQ ID NO: 28)
AATGATACGGCGACCACCGAGATCTACAC
CGAGAAGGCTACgggaap-5'(SEQ ID NO: 31)
802
803
804
805
3'-GTTCGTCTTCTTCGCCGTATGTCTCTAT[i7]CACTGACCTCAAGTCTGCACA
(SEQ ID NO: 29)
CAAGCAGAAGACGGCATACGAGAT

图8

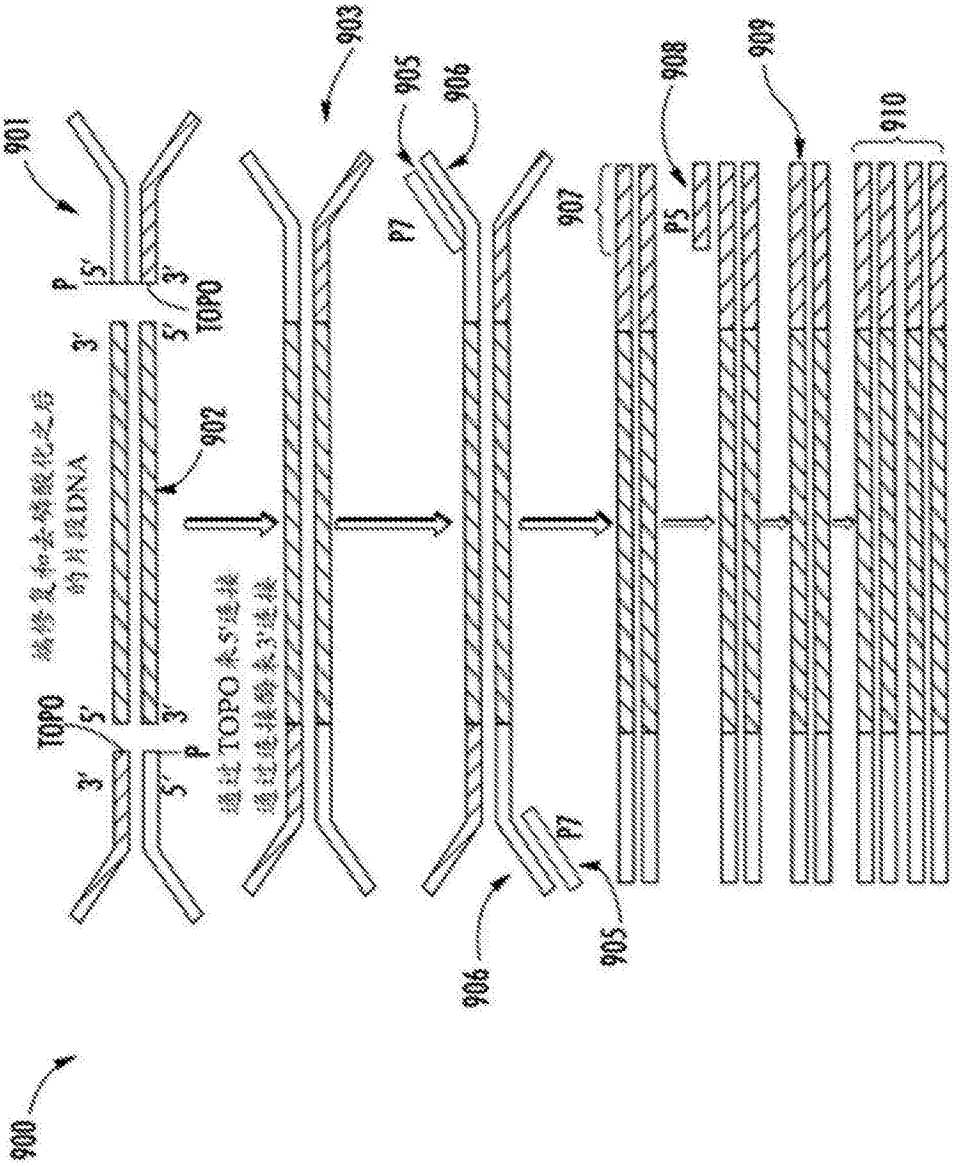

901
903
905
906
908
909
910
907
P7
P5
TOPO
902
P
5'
3'
通过TOPO末端连接
通过连接酶末端连接
900

图9