



(12) 发明专利申请

(10) 申请公布号 CN 112430641 A

(43) 申请公布日 2021.03.02

(21) 申请号 202011083287.0

(51) Int.Cl.

(22) 申请日 2015.06.26

C12Q 1/6806 (2018.01)

(30) 优先权数据

C12N 15/10 (2006.01)

62/019,209 2014.06.30 US

C40B 50/06 (2006.01)

(62) 分案原申请数据

201580036230.6 2015.06.26

(71) 申请人 亿明达股份有限公司

地址 美国加利福尼亚州

(72) 发明人 F.J.斯蒂默斯 J.S.费希尔

K.L.冈德森 S.阿米尼

C.格勒克纳

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 涂滔

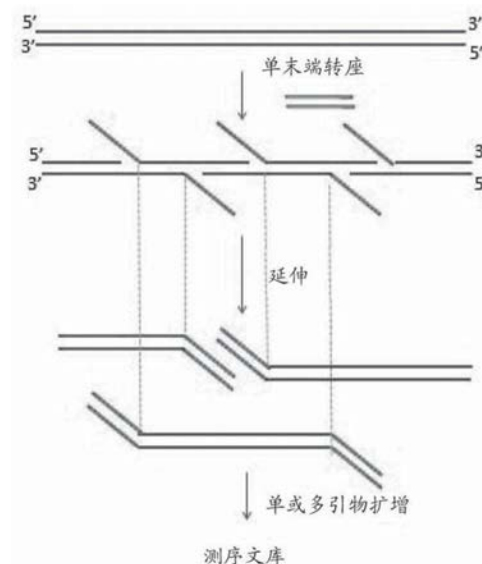
权利要求书1页 说明书31页 附图12页

(54) 发明名称

使用单侧转座的方法和组合物

(57) 摘要

本文提供的实施方案涉及用于下一代测序的方法和组合物。一些实施方案包括使用单侧转座从靶核酸制备模板文库，并且对该模板文库测序。



1. 从双链靶核酸制备测序文库的方法,其包括:

(a) 提供多个转座体,每个转座体单体包含转座酶和转座子核酸,其中所述转座体配置为仅对所述双链靶核酸的一条链产生切口;并且

(b) 使所述靶核酸与所述转座体接触,使得在所述靶核酸的多个位点处对所述靶核酸产生切口,并且单个转座子核酸附着到至少一个有切口的靶核酸以产生转座的核酸,从而获得经修饰的核酸文库用于测序。

2. 用于捕获靶DNA的邻接信息的方法,其包括:

(a) 提供多个转座体,每个转座体单体包含转座酶和转座子核酸,其中所述转座体配置为仅对所述双链靶核酸的一条链产生切口;

(b) 使所述靶DNA与所述转座体接触,使得在所述靶核酸的多个位点处对所述靶DNA产生切口;

(c) 向所述靶DNA序列中加入或插入一个或多个识别序列以产生经处理的靶DNA;

(d) 测序所述经处理的靶DNA;并且

(e) 通过鉴定所述靶DNA序列或具有共享特性的识别位点来捕获邻接信息。

3. 权利要求1或2的方法,其还包括在表面上捕获所述经修饰的核酸。

4. 权利要求1或2的方法,其中将在 (b) 中与所述靶核酸接触的所述转座体附着到表面,从而在所述表面上捕获所述经修饰的核酸。

5. 权利要求3或4的方法,其包括测序所述表面上的捕获的核酸。

6. 权利要求3-5中任一项的方法,其中从两个捕获的核酸获得的序列信息在所述靶核酸序列的线性表示中的接近性 (proximity) 指示在所述表面上捕获的核酸的接近性。

7. 权利要求6的方法,其中与较不接近 (in less close proximity) 的捕获的核酸相比,在所述表面上彼此更接近 (in closer proximity to one another) 的捕获的核酸包含在所述靶核酸序列的表示中更接近的序列。

8. 根据权利要求6和7中任一项的方法,其中所述靶核酸序列的表示包括单倍型表示。

9. 根据权利要求1-8中任一项的方法,其中所述转座体包含单侧转座酶活性。

10. 权利要求9的方法,其中所述转座体包含缺乏转座酶活性的单体亚单位。

使用单侧转座的方法和组合物

[0001] 本申请是基于申请日为2015年6月26日,优先权日为2014年6月30日,申请号为201580036230.6,发明名称为:“使用单侧转座的方法和组合物”的专利申请的分案申请。

[0002] 相关申请

[0003] 本申请要求于2014年6月30日提交的美国临时申请号62/019,209的优先权,其通过引用以其整体并入本文。

技术领域

[0004] 本文提供的实施方案涉及用于下一代测序的方法和组合物。一些实施方案包括使用单侧转座(one-sided transposition),也称为单侧转座(one sided transposition)从靶核酸制备模板文库,对模板文库测序,以及捕获邻接信息(contiguity information)。

[0005] 发明背景

[0006] 几种下一代测序技术可用于快速和经济测定基因组的整个序列。通常,在测序前从靶基因组DNA样品制备模板核酸文库。样品制备通常包括DNA片段化(fragmentation)步骤,其将较大的DNA链断裂成更适合下一代测序技术的较小的DNA片段。通常将衔接头(adaptor)附着至该DNA片段的末端,这可以通过DNA末端修复,然后是衔接头连接来完成,或最近通过使用转座体(transposome)系统来完成。转座体(其是转座酶和转座子核酸的复合物)的使用允许同时进行基因组片段化和片段的衔接头连接,从而简化文库制备。然而,基因组DNA的片段化可导致邻接,定相和单倍型方面就个别核酸分子而言的信息损失。因此,存在对备选文库制备方法的需要。

[0007] 发明概述

[0008] 在本文描述的一些实施方案中是用于单侧转座的方法。本申请的发明人惊奇地发现,通过进行单侧转座,双链靶DNA仅在一链上产生切口,并且即使在除去转座体后,此类转座后的靶DNA保持完整。因此,靶DNA的邻接在转座事件之后得到保持。在一些实施方案中,用于单侧转座的方法可以用于捕获邻接信息。在一些实施方案中,用于单侧转座的方法可用于制备测序文库。在一些实施方案中,用于单侧转座的方法可用于确定定相信息或单倍型信息。

[0009] 在一些实施方案中,转座体二聚体配置为仅对双链靶DNA的一条链产生切口,并将转座体单体的转座子的仅一条转移链转移至有切口的靶DNA。在一些实施方案中,转座体二聚体的一个单体亚单位不能转座,导致单侧转座。在一些实施方案中,转座体二聚体的一个转座酶可以通过结合转座子而形成转座体复合物,但不能对靶DNA产生切口。

[0010] 在一些实施方案中,转座子是功能性的,使得通过将转座子与转座酶接触形成转座体,并且可以将转座子序列转移到靶核酸。在一些实施方案中,转座子是非功能性的,使得通过使转座子与转座酶接触形成转座体,但转座子序列不能转移到靶核酸。在一些实施方案中,转移链的3'-末端包含不能对靶核酸的5'-末端进行亲核攻击的3'-末端核酸。在一些实施方案中,识别序列的3'-末端是封闭的。在一些实施方案中,封闭的识别序列的3'-末端包含3'-末端双脱氧核苷酸,胺基基团,烷基基团,芳基基团,硫醇基团,硫酸根基团,反向

核苷酸,叠氮基基团或生物素。

[0011] 在一些实施方案中,转座酶能够形成转座体,但不能对靶DNA产生切口。在一些实施方案中,转座酶包含一个或多个氨基酸修饰,使得其能够形成转座体,但不能对靶DNA产生切口。

[0012] 在一些实施方案中,转座体复合物以使得转座体不能有效形成二聚体的方式配置。在一些实施方案中,转座体复合物以使得转座体完全不能形成二聚体的方式配置。在一些实施方案中,转座体单体仅在双链DNA的一条链中形成切口,并将转座子的转移链转移至有切口的靶DNA。

[0013] 在一些实施方案中,通过利用对通过靶DNA的两条链的转座的差异抗性(resistance)进行单侧转座。在一些实施方案中,靶DNA的一条链包含修饰的碱基或修饰的磷酸二酯键,其对转座有抗性。将具有通过靶DNA的两条链的转座的差异抗性的靶DNA暴露于转座体导致单侧转座。在一些实施方案中,靶核酸是双链cDNA,其中cDNA的一条链包含修饰的碱基和/或修饰的磷酸二酯键,使得所述链对转座有抗性。在一些实施方案中,靶核酸是双链基因组DNA,其中一条链以使得所述链部分或完全对转座有抗性的方式修饰。

[0014] 在图14中示出示例性的单侧转座方案。在一些实施方案中,从单链核酸模板(实线)开始,合成互补链(虚线),其与原始模板链具有对转座的差异抗性。然后,使用即使正常的转座子复合物(例如活性和未封闭的),发生单侧转座。在一个实例中,新合成的链具有较高的转座抗性。在另一个实例中,原始模板是高度抗性的,并且合成的链允许转座为自身。在此实施方案中,较低抗性的链(less resistant strand)形成文库元件,其通过更具抗性的链保持邻接(which are held in contiguity by the more resistant strand)。

[0015] 申请人令人惊讶地发现,在进行单侧转座后,双链靶核酸保持完整而不丢失邻接信息,即使是在除去转座体的转座酶之后。在一些实施方案中,通过使用SDS,尿素,蛋白酶或热处理在转座后从转座的靶核酸中除去转座酶。因此,单侧转座可以有利于确定序列信息,邻接信息,定相信息和单倍型信息。邻接信息可以提供广泛的单倍型分辨能力。单倍型允许定相罕见等位基因和结构变体,如基因重排,基因复制。

[0016] 在一些实施方案中,可以将单侧转座与组合条形码化(combinatorial barcoding)相结合,其中第一组条形码通过单侧转座附着,而第二组条形码通过后续扩增附着。

[0017] 在一些实施方案中,在转座期间将第一组条形码引入靶核酸以产生包含第一组条形码的转座的靶核酸。合并转座的靶核酸以产生转座的靶核酸的第一合并物。将第二组条形码引入转座的靶核酸的第一合并物以产生包含第一和第二组条形码的靶核酸。第二组条形码可以通过随后的扩增,连接或另外的转座引入。在一些实施方案中,第一和第二组条形码是不同的。将包含第一和第二组条形码的靶核酸合并以产生转座的靶核酸的第二合并物。任选地,可以重复引入另外的条形码和合并的步骤以产生有条形码的(barcode)靶核酸文库。

[0018] 在一些实施方案中,单侧转座可用于确定来自单细胞的核酸的序列信息或邻接信息。核酸可以是单细胞的基因组核酸或从mRNA产生的cDNA。在一些实施方案中,可以将第一组条形码引入来自单细胞的核酸,其充当该单细胞的标识符(identifier)。在一些实施方案中,在将第一组条形码引入来自单细胞的核酸后,可以合并有条形码的核酸,并通过随后

的扩增,连接或另外的转座(在引入或不引入另外的条形码的情况下)进行进一步处理。

[0019] 本文提供的方法和组合物的一些实施方案包括从双链靶核酸制备测序文库的方法,包括:(a)提供多个转座体,每个转座体包含转座酶和转座子核酸,其中转座体配置为仅对靶核酸的一条链产生切割并将转座子转移至该靶核酸的一条链;并且(b)使靶核酸与转座体接触,使得在靶核酸的多个位点处对靶核酸产生切口,并且将转座子核酸附着至有切口的靶核酸,从而获得经修饰的核酸的文库用于测序。

[0020] 一些实施方案包括从双链靶核酸制备测序文库的方法,包括:(a)提供多个转座体,每个转座体包含转座酶和转座子核酸,其中转座体配置为仅对靶核酸的一条链产生切口并将转座子转移至靶核酸的一条链;(b)使靶核酸与转座体接触,使得在靶核酸的多个位点处对靶核酸产生切口,并且将转座子核酸附着至有切口的靶核酸;并且(c)将引物与转座子核酸杂交并延伸杂交的引物,从而获得修饰的核酸的文库用于测序。使用单侧转座的文库制备的示例性方案示于图13中。

[0021] 一些实施方案包括用于捕获靶DNA的邻接信息的方法。该方法包括(a)提供多个转座体,每个转座体单体包含转座酶和转座子核酸,其中转座体配置为仅对双链靶核酸的一条链产生切割;(b)使靶DNA与转座体接触,使得在靶核酸的多个位点处对靶DNA产生切口;(c)向靶DNA序列中加入或插入一个或多个识别序列以产生处理的靶DNA;(d)测序处理的靶DNA;并且(e)通过鉴定靶DNA序列或具有共享特性的识别序列来捕获邻接信息。

[0022] 一些实施方案包括捕获靶DNA的邻接信息的方法。该方法包括(a)提供多个转座体,每个转座体单体包含转座酶和包含识别序列的转座子核酸,其中转座体配置为仅对双链靶核酸的一条链产生切口;(b)将转座子核酸插入靶核酸的链中,包括:(i)使靶核酸与转座体接触,使得在多个位点处对靶核酸产生切口,并且将单个转座子核酸在产生切口位点的一侧上附着至有切口的链,和(ii)将附着的单个转座子核酸在产生切口位点的另一侧连接至有切口的链,从而获得经修饰的核酸;(c)扩增经修饰的核酸,从而获得包含插入的识别序列的多个核酸;(d)测序经处理的靶DNA;并且(e)通过鉴定靶DNA序列或具有共享特性的识别序列来捕获邻接信息。

[0023] 一些实施方案还包括在表面上捕获经修饰的核酸。

[0024] 在一些实施方案中,将在(b)中与靶核酸接触的转座体附着至表面,从而在表面上捕获经修饰的核酸。

[0025] 一些实施方案还包括在表面上测序捕获的核酸。

[0026] 在一些实施方案中,从两个捕获的核酸获得的序列信息在靶核酸序列的线性表示中的接近性(proximity)指示在表面上的捕获的核酸的接近性。

[0027] 在一些实施方案中,与较不接近(in less close proximity)的捕获的核酸相比,在表面上彼此更接近的捕获的核酸包含在靶核酸序列的表示中更接近的序列。

[0028] 在一些实施方案中,靶核酸序列的表示包含单倍型表示。在一些实施方案中,靶核酸序列的表示包括有序短读段(ordered short reads)。

[0029] 在一些实施方案中,转座酶包含单侧转座酶活性。

[0030] 在一些实施方案中,转座酶包含缺乏转座酶活性的单体亚单位。在一些实施方案中,转座酶包含共价连接的单体亚单位。在一些实施方案中,转座酶的四级结构是单体的。在一些实施方案中,转座酶缺乏形成二聚体的能力。

[0031] 在一些实施方案中,转座酶选自下组:Mu,Mu E392Q,Tn5,超活性Tn5(hyperactive Tn5)(Goryshin and Reznikoff,J.Biol.Chem.,273:7367(1998)),EZ-Tn5TM转座酶(Epicentre Biotechnologies,Madison,Wisconsin),Tn5的变体,RAG,Tn7,Tn10,Vibhar转座酶和Tn552.Tn5转座酶的变体,如具有氨基酸取代,插入,缺失和/或其它蛋白或肽的融合,公开于美国专利:5,925,545;5,965,443;7,083,980;7,608,434;和美国专利申请14/686,961。专利和专利申请通过引用以其整体并入本文。在一些实施方案中,相对于野生型蛋白质,Tn5转座酶在位置54,56,372,212,214,251和338处包含一个或多个取代,如美国专利申请14/686,961中所公开的。在一些实施方案中,Tn5野生型蛋白或其变体可以进一步包含融合多肽。在一些实施方案中,与转座酶融合的多肽结构域可以包含例如延伸因子Ts。本段中引用的每个参考文献通过引用以其整体并入本文。

[0032] 在一些实施方案中,封闭转座子核酸。在一些实施方案中,封闭转座子的转移链的3'-末端。在一些实施方案中,封闭的转座子核酸的3'末端选自下组:双脱氧基基团,间隔物基团,胺基基团,叠氮基基团,磷酸根基团,烷基基团,反向核苷酸和生物素基团。在一些实施方案中,可以通过对来自转座子序列的碱基的取代,添加或缺失改变转座子序列。

[0033] 在一些实施方案中,通过使转座酶与功能性转座子核酸和非功能性转座子核酸接触来制备多个转座体。在一些实施方案中,非功能性转座子包含封闭的转座子。在一些实施方案中,构成非功能性转座子核酸与功能性转座子核酸的转座子核酸的比率大于或等于1:1。在一些实施方案中,构成非功能性转座子核酸与功能性转座子核酸的转座子核酸的比率可以是1:2,1:3,1:5,1:10,1:20,1:30,1:40,1:50,1:75,1:100,2:1,3:1,4:1,5:1,6:1,7:1,8:1,9:1,10:1,20:1,30:1,40:1,50:1,60:1,70:1,80:1,90:1,或100:1。

[0034] 一些实施方案还包括扩增延伸的核酸。在一些实施方案中,用有尾扩增引物(tailed amplification primers)扩增延伸的核酸,该有尾扩增引物包含选自下组的序列:锚定位点,测序引物位点,扩增引物位点,和报告标签。

[0035] 一些实施方案还包括扩增捕获的核酸。在一些实施方案中,扩增捕获的核酸包括桥式扩增。

[0036] 在一些实施方案中,表面包含多个捕获探针。在一些实施方案中,捕获探针包含核酸。一些实施方案还包括将经修饰的核酸与捕获探针杂交。

[0037] 在一些实施方案中,经修饰的核酸和捕获探针各自包含亲和部分。在一些实施方案中,亲和部分可以是结合对的成员。在一些情况下,经修饰的核酸可以包含结合对的第一成员,而捕获探针可以包含结合对的第二成员。在一些情况下,捕获探针可以固定到固体表面,并且经修饰的核酸可以包含结合对的第一成员,并且捕获探针可以包含结合对的第二成员。在此类情况下,使结合对的第一和第二成员结合将修饰的核酸固定到固体表面。结合对的例子包括但不限于生物素-亲合素,生物素-链霉亲合素,生物素-中性亲合素(neutravidin),配体-受体,激素-受体,凝集素-糖蛋白和抗原-抗体。

[0038] 一些实施方案还包括使修饰的核酸的亲和部分与捕获探针的亲和部分结合。

[0039] 在一些实施方案中,转座子核酸包含选自下组的序列:锚定位点,条形码,测序引物位点,扩增引物位点,独特分子指标(unique molecular index)和报告标签。

[0040] 在一些实施方案中,至少一个转座体包含两个转座子核酸。

[0041] 在一些实施方案中,该两个转座子核酸具有不同的序列。

[0042] 在一些实施方案中,多个转座体包含至少两个不同的转座子核酸。

[0043] 在一些实施方案中,靶核酸选自DNA和RNA。在一些实施方案中,靶核酸选自基因组DNA和cDNA。在一些实施方案中,靶核酸是基因组DNA。

[0044] 在一些实施方案中,表面在选自下组的基底上:珠粒,载玻片,流动池,通道,浸渍片(dip-stick)和孔。

[0045] 在一些实施方案中,表面包含每 mm^2 至少约10,000个捕获的核酸。在一些实施方案中,表面包含每 mm^2 至少约100,000个捕获的核酸。在一些实施方案中,表面包含每 mm^2 至少约1,000,000,1,500,000,2,000,000,3,000,000,5,000,000,10,000,000,15,000,000,20,000,000,30,000,000,40,000,000,50,000,000,60,000,000,70,000,000,80,000,000,90,000,000,100,000,000,150,000,000,200,000,000,300,000,000,350,000,000,400,000,000,450,000,000,500,000,000,550,000,000,600,000,000,650,000,000,700,000,000,750,000,000,800,000,000,850,000,000,900,000,000,950,000,000,1000,000,000,1200,000,000,1300,000,000,1400,000,000,1500,000,000,1600,000,000,1700,000,000,1800,000,000,1900,000,000,2000,000,000,3000,000,000,4000,000,000,5000,000,000,6000,000,000,700,000,000,8,000,000,000,9,000,000,000,10,000,000,000或更多个捕获的核酸。

[0046] 一些实施方案包括通过前述方法的任一项制备的测序文库。

[0047] 本文提供的方法和组合物的一些实施方案包括从双链靶核酸制备具有条形码的测序文库,其包括:(a)提供多个转座体,每个转座体包含转座酶和包含条形码的转座子核酸,并且(b)将转座子核酸插入靶核酸的链中,包括:(i)使靶核酸与转座体接触,使得在多个位点处对靶核酸产生切口,并且将单个转座子核酸在产生切口位点的一侧附着至有切口的链,并且(ii)将附着的单个转座子核酸在产生切口位点的另一侧连接至有切口的链,从而获得修饰核酸的文库。

[0048] 一些实施方案还包括(c)在表面上捕获修饰的核酸。

[0049] 一些实施方案包括从双链靶核酸制备具有条形码的测序文库,其包括:(a)提供多个转座体,每个转座体单体包含转座酶和包含条形码的转座子核酸;和(b)将转座子核酸插入靶核酸的链,包括:(i)使靶核酸与转座体接触,使得在多个位点处对靶核酸产生切口,并且将单个转座子核酸在有切口的位点的一侧附着至有切口的链,并且(ii)将附着的单个转座子核酸在有切口的位点的另一侧连接至有切口的链,从而获得经修饰的核酸;(c)扩增经修饰的核酸,从而获得包含插入的条形码的多个核酸。

[0050] 一些实施方案还包括在表面上捕获修饰的核酸。

[0051] 在一些实施方案中,将在(b)中与靶核酸接触的转座体附着至表面,从而在该表面上捕获经修饰的核酸。

[0052] 一些实施方案还包括测序捕获的核酸。

[0053] 在一些实施方案中,从两个捕获的核酸获得的序列信息在靶核酸序列的线性表示中的接近性指示在表面上的捕获的核酸的接近性。

[0054] 在一些实施方案中,与较不接近的捕获的核酸相比,在表面上彼此更接近的捕获的核酸包含在靶核酸序列的表示中更接近的序列。

[0055] 在一些实施方案中,靶核酸序列的表示包含单倍型表示。

- [0056] 在一些实施方案中,至少一个转座子核酸的条形码是不同的。
- [0057] 在一些实施方案中,转座子核酸的条形码不相同。
- [0058] 一些实施方案还包括根据序列中共同条形码的存在来比对核酸序列以产生靶核酸的表示。
- [0059] 在一些实施方案中,转座酶包含单侧转座酶活性。
- [0060] 在一些实施方案中,转座酶包含缺乏转座酶活性的单体亚单位。
- [0061] 在一些实施方案中,转座酶包含共价连接的单体亚单位。在一些实施方案中,转座酶的四级结构是单体的。在一些实施方案中,转座酶缺乏形成二聚体的能力。
- [0062] 在一些实施方案中,转座酶选自下组:Mu,Mu E392Q,Tn5,超活性Tn5,EZ-Tn5TM,Tn5的变体,RAG,Tn7,Tn10,Tn552和Vibhar转座酶。
- [0063] 在一些实施方案中,转座子核酸是封闭的。
- [0064] 在一些实施方案中,封闭的转座子核酸的3'末端选自下组:双脱氧基基团,间隔物基团,胺基基团,叠氮基基团,烷基基团,芳基基团,反向核苷酸,硫代磷酸根基团(thiophosphate group)和生物素基团。
- [0065] 在一些实施方案中,通过使转座酶与非功能性转座子核酸和功能性转座子核酸接触来制备多个转座体。在一些实施方案中,非功能性转座子包含封闭的3'末端。在一些实施方案中,通过使转座酶与封闭的转座子核酸和非封闭的转座子核酸接触来制备多个转座体。在一些实施方案中,构成封闭的转座子核酸与非封闭的转座子核酸的转座子核酸的比率大于或等于1:1。在一些实施方案中,构成封闭的转座子核酸与非封闭的转座子核酸的转座子核酸的比率可以是1:2,1:3,1:5,1:10,1:20,1:30,1:40,1:50,1:75,1:100,2:1,3:1,4:1,5:1,6:1,7:1,8:1,9:1,10:1,20:1,30:1,40:1,50:1,60:1,70:1,80:1,90:1,或100:1。
- [0066] 一些实施方案还包括将扩增接头附着至靶核酸。在一些实施方案中,扩增接头包含选自下组的序列:锚定位点,测序引物位点,扩增引物位点和报告标签。
- [0067] 一些实施方案还包括扩增捕获的核酸。在一些实施方案中,扩增捕获的核酸包括桥式扩增。
- [0068] 在一些实施方案中,表面包含多个捕获探针。在一些实施方案中,捕获探针包括核酸。在一些实施方案中,捕获探针各自包含亲和部分。在一些实施方案中,亲和部分选自下组:生物素,亲合素,链霉亲合素和重组酶。
- [0069] 在一些实施方案中,转座子核酸包含选自下组的序列:锚定位点,测序引物位点,扩增引物位点,独特分子指标和报告标签。
- [0070] 在一些实施方案中,至少一个转座体包含两个转座子核酸。在一些实施方案中,该两个转座子核酸具有不同的序列
- [0071] 在一些实施方案中,多个转座子包含至少两个不同的转座子核酸。
- [0072] 在一些实施方案中,靶核酸选自基因组DNA的DNA片段和cDNA。在一些实施方案中,靶核酸选自基因组DNA和cDNA。在一些实施方案中,靶核酸是基因组DNA。
- [0073] 在一些实施方案中,表面在选自下组的基底上:珠粒,载玻片,流动池,通道,浸渍片和孔。
- [0074] 在一些实施方案中,表面包含每mm²至少约10,000个捕获的核酸。在一些实施方案中,表面包含每mm²至少约100,000个捕获的核酸。在一些实施方案中,表面包含每mm²至少约

1,000,000,1,500,000,2,000,000,3,000,000,5,000,000,10,000,000,15,000,000,20,000,000,30,000,000,40,000,000,50,000,000,60,000,000,70,000,000,80,000,000,90,000,000,100,000,000,150,000,000,200,000,000,300,000,000,350,000,000,400,000,000,450,000,000,500,000,000,550,000,000,600,000,000,650,000,000,700,000,000,750,000,000,800,000,000,850,000,000,900,000,000,950,000,000,1000,000,000,1200,000,000,1300,000,000,1400,000,000,1500,000,000,1600,000,000,1700,000,000,1800,000,000,1900,000,000,2000,000,000,3000,000,000,4000,000,000,5000,000,000,6000,000,000,700,000,000,8,000,000,000,9,000,000,000,10,000,000,000或更多个捕获的核酸。

[0075] 一些实施方案包括通过前述方法中的任一项制备的包含条形码的测序文库。

[0076] 在一些实施方案中,在用转座酶处理后或在随后的扩增后,可以将一个或多个识别序列加入或插入到有切口的靶核酸中。一个或多个识别序列可以包括但不限于在产生切口的位点处的条形码,引物或衔接头DNA序列,其将靶核酸片段标记(tag)为就相邻,区划(compartmental)或距离空间关系而言是独特的。

[0077] 在标记后,可以使用上述测序平台对鸟枪(shotgun)核酸分子进行测序,通过鉴定具有共享特性的识别序列来捕获邻接信息。在一些实施方案中,共享特性是相同或互补的条形码序列。例如,可以通过共享的条形码序列来鉴定相邻来源的读段序列;或可以基于来源于相同靶DNA区段的共享的区室(compartment)特异性条形码通过区室限定读段。在其它实施方案中,共享特性是共享的或约束的(constrained)物理位置,其可以由流动池上的一个或多个x,y坐标来指示。“约束”物理位置可以指接近,相同或几乎相同的物理位置,或者指一组两个或多个物理位置,其相对物理坐标与衍生核酸片段的靶核酸序列上的相对序列坐标相关。例如,在涉及长范围邻接的方法中,使用衔接头序列在测序流动池的表面上原位转移到伸展的HMW基因组DNA中,来通过鉴定衔接头序列、杂交的DNA片段,或其组合的约束的物理位置(即,固定物理连接的测序模板处的相对坐标),来获得距离空间关系。该方法可以用于捕获短程(short range),中程(mid range)和长程(long range)的邻接信息。

[0078] 在一些实施方案中,可以将单侧转座与组合条形码化组合。单侧转座元件的使用是会实现组合条形码化的,而不需要任何另外的机制以在处理期间将相关的文库元件保持在一起。图15中示出了将单侧转座与组合条形码化组合的示例性方案。

[0079] 本公开包括以下实施方案:

[0080] 实施方案1.从双链靶核酸制备测序文库的方法,其包括:

[0081] (a) 提供多个转座体,每个转座体单体包含转座酶和转座子核酸,其中所述转座体配置为仅对所述双链靶核酸的一条链产生切口;并且

[0082] (b) 使所述靶核酸与所述转座体接触,使得在所述靶核酸的多个位点处对所述靶核酸产生切口,并且单个转座子核酸附着到至少一个有切口的靶核酸以产生转座的核酸,从而获得经修饰的核酸文库用于测序。

[0083] 实施方案2.用于捕获靶DNA的邻接信息的方法,其包括:

[0084] (a) 提供多个转座体,每个转座体单体包含转座酶和转座子核酸,其中所述转座体配置为仅对所述双链靶核酸的一条链产生切口;

[0085] (b) 使所述靶DNA与所述转座体接触,使得在所述靶核酸的多个位点处对所述靶

DNA产生切口；

[0086] (c) 向所述靶DNA序列中加入或插入一个或多个识别序列以产生经处理的靶DNA；

[0087] (d) 测序所述经处理的靶DNA；并且

[0088] (e) 通过鉴定所述靶DNA序列或具有共享特性的识别位点来捕获邻接信息。

[0089] 实施方案3.实施方案1或2的方法，其还包括在表面上捕获所述经修饰的核酸。

[0090] 实施方案4.实施方案1或2的方法，其中将在(b)中与所述靶核酸接触的所述转座体附着到表面，从而在所述表面上捕获所述经修饰的核酸。

[0091] 实施方案5.实施方案3或4的方法，其包括测序所述表面上的捕获的核酸。

[0092] 实施方案6.实施方案3-5中任一项的方法，其中从两个捕获的核酸获得的序列信息在所述靶核酸序列的线性表示中的接近性(proximity)指示在所述表面上捕获的核酸的接近性。

[0093] 实施方案7.实施方案6的方法，其中与较不接近(in less close proximity)的捕获的核酸相比，在所述表面上彼此更接近(in closer proximity to one another)的捕获的核酸包含在所述靶核酸序列的表示中更接近的序列。

[0094] 实施方案8.根据实施方案6和7中任一项的方法，其中所述靶核酸序列的表示包括单倍型表示。

[0095] 实施方案9.根据实施方案1-8中任一项的方法，其中所述转座体包含单侧转座酶活性。

[0096] 实施方案10.实施方案9的方法，其中所述转座体包含缺乏转座酶活性的单体亚单位。

[0097] 实施方案11.实施方案9的方法，其中所述转座体包含共价连接的单体亚单位。

[0098] 实施方案12.实施方案9-11中任一项的方法，其中所述转座酶的四级结构是单体的。

[0099] 实施方案13.实施方案9的方法，其中所述转座酶缺乏形成二聚体的能力。

[0100] 实施方案14.实施方案1-13中任一项的方法，其中所述转座酶选自下组：Mu，MuE392Q，Tn5，超活性Tn5(hyperactive Tn5)，Tn5变体，Vibhar，RAG和Tn552。

[0101] 实施方案15.实施方案1-14中任一项的方法，其中所述一种或多种转座子核酸是非功能性的。

[0102] 实施方案16.实施方案15的方法，其中所述非功能性转座子核酸的3'末端选自下组：双脱氧基基团，间隔物基团，胺基，烷基基团，芳基基团，磷酸根基团，硫醇基基团，反向核苷酸，叠氮基基团，硫酸根基团和生物素基团。

[0103] 实施方案17.实施方案15的方法，其中通过使所述转座酶与功能性转座子核酸和非功能性转座子核酸接触来制备所述多个转座体。

[0104] 实施方案18.实施方案15的方法，其中构成非功能性转座子核酸与功能性转座子核酸的转座子核酸的比率大于或等于1:1。

[0105] 实施方案19.实施方案15的方法，其中构成非功能性转座子核酸与功能性转座子核酸的转座子核酸的比率大于或等于10:1。

[0106] 实施方案20.实施方案2-19中任一项的方法，其还包括：

[0107] (i) 向有切口的靶核酸提供DNA聚合酶；

- [0108] (ii) 使用互补链作为模板延伸所述靶核酸的3'；
- [0109] (iii) 任选地, 扩增所述延伸的核酸。
- [0110] 实施方案21. 实施方案20的方法, 其中用有尾扩增引物(tailed amplification primers) 扩增所述延伸的核酸, 所述有尾扩增引物包含选自下组的序列: 锚定位点, 测序引物位点, 扩增引物位点, 条形码和报告标签。
- [0111] 实施方案22. 实施方案3-21中任一项的方法, 其包括扩增所述捕获的核酸。
- [0112] 实施方案23. 实施方案22的方法, 其中所述扩增所述捕获的核酸包括桥式扩增。
- [0113] 实施方案24. 实施方案3-23中任一项的方法, 其中所述表面包含多个捕获探针。
- [0114] 实施方案25. 实施方案24的方法, 其中所述捕获探针包含核酸。
- [0115] 实施方案26. 实施方案25的方法, 其包括使所述经修饰的核酸与所述捕获探针杂交。
- [0116] 实施方案27. 实施方案24的方法, 其中所述经修饰的核酸和所述捕获探针各自包含亲和部分。
- [0117] 实施方案28. 实施方案27的方法, 其中所述亲和部分选自下组: 生物素, 亲合素和链霉亲合素。
- [0118] 实施方案29. 实施方案27的方法, 其包括将所述经修饰的核酸的亲和部分与所述捕获探针的亲和部分结合。
- [0119] 实施方案30. 实施方案1-29中任一项的方法, 其中所述转座子核酸包含选自下组的序列: 锚定位点, 条形码, 测序引物位点, 扩增引物位点和报告标签。
- [0120] 实施方案31. 实施方案1-30中任一项的方法, 其中至少一个转座体包含两个转座子核酸。
- [0121] 实施方案32. 实施方案31的方法, 其中所述两个转座子核酸具有不同的序列。
- [0122] 实施方案33. 实施方案1-32中任一项的方法, 其中所述多个转座体包含至少两个不同的转座子核酸。
- [0123] 实施方案34. 实施方案1-33中任一项的方法, 其中在使用所述转座体处理所述靶核酸之后, 通过SDS, 尿素, 热或蛋白酶处理从所述靶核酸中除去所述转座酶。
- [0124] 实施方案35. 实施方案1-34中任一项的方法, 其中所述靶核酸选自下组: 基因组DNA, 基因组DNA的片段和cDNA。
- [0125] 实施方案36. 实施方案35的方法, 其中所述靶核酸是基因组DNA。
- [0126] 实施方案37. 实施方案3-36中任一项的方法, 其中所述表面在选自下组的基底上: 珠粒, 载玻片, 流动池, 通道, 浸渍片和孔。
- [0127] 实施方案38. 实施方案3-37中任一项的方法, 其中所述表面包含每 mm^2 至少约10,000个捕获的核酸。
- [0128] 实施方案39. 实施方案3-38中任一项的方法, 其中所述表面包含每 mm^2 至少约100,000个捕获的核酸。
- [0129] 实施方案40. 实施方案3-39中任一项的方法, 其中所述表面每 mm^2 包含至少约1,000,000个捕获的核酸。
- [0130] 实施方案41. 通过实施方案1-40中任一项的方法制备的测序文库。
- [0131] 实施方案42. 从双链靶核酸制备具有条形码的测序文库的方法, 其包括:

[0132] (a) 提供多个转座体,每个转座体包含转座酶和包含识别序列的转座子核酸,其中所述转座体配置为仅对所述双链靶核酸的一条链产生切口;并且

[0133] (b) 将所述转座子核酸插入所述靶核酸的链中,包括:

[0134] (i) 使所述靶核酸与所述转座体接触,使得在多个位点处对所述靶核酸产生切口,并且将单个转座子核酸在有切口的位点的一侧上附着到有切口的链以产生转座的核酸,并且

[0135] (ii) 将所述附着的单个转座子核酸在所述有切口的位点的另一侧上连接至所述有切口的链,从而获得经修饰的核酸的文库用于测序。

[0136] 实施方案43.捕获靶DNA的邻接信息的方法,其包括:

[0137] (a) 提供多个转座体,每个转座体单体包含转座酶和包含识别序列的转座子核酸,其中所述转座体配置为仅对所述双链靶核酸的一条链产生切口;

[0138] (b) 将所述转座子核酸插入所述靶核酸的链中,包括:

[0139] (i) 使所述靶核酸与所述转座体接触,使得在多个位点处对所述靶核酸产生切口,并且将单个转座子核酸在有切口的位点的一侧上附着至有切口的链,并且

[0140] (ii) 将所述附着的单个转座子核酸在所述有切口的位点的另一侧上连接至所述有切口的链,从而获得经修饰的核酸;

[0141] (c) 扩增所述经修饰的核酸,从而获得包含插入的识别序列的多个核酸;

[0142] (d) 测序经处理的靶DNA;并且

[0143] (e) 通过鉴定靶DNA序列或具有共享特性的识别序列来捕获邻接信息。

[0144] 实施方案44.实施方案42或43的方法,其还包括(c)在表面上捕获所述经修饰的靶核酸。

[0145] 实施方案45.实施方案42或43的方法,其中将在(b)中与所述靶核酸接触的所述转座体附着至表面,从而在所述表面上捕获所述经修饰的核酸。

[0146] 实施方案46.实施方案44或45的方法,包括对所述捕获的核酸测序。

[0147] 实施方案47.实施方案42-46中任一项的方法,其中从两个捕获的核酸获得的序列信息在所述靶核酸序列的线性表示中的接近性指示在所述表面上捕获的核酸的接近性。

[0148] 实施方案48.实施方案47的方法,其中与较不接近的捕获的核酸相比,在所述表面上彼此更接近的捕获的核酸包含在所述靶核酸序列的表示中更接近的序列。

[0149] 实施方案49.实施方案47和48中任一项的方法,其中所述靶核酸序列的表示包括单倍型表示。

[0150] 实施方案50.实施方案42-49中任一项的方法,其中所述识别序列是条形码,且其中至少一个转座子核酸的条形码是不同的。

[0151] 实施方案51.实施方案42-50中任一项的方法,其中所述识别序列是条形码,且其中所述转座子核酸的条形码是不相同的。

[0152] 实施方案52.实施方案46-51中任一项的方法,其包括根据所述序列中的共同条形码的存在来比对所述核酸序列以产生所述靶核酸的表示。

[0153] 实施方案53.实施方案42-52中任一项的方法,其中所述转座酶包含单侧转座酶活性。

[0154] 实施方案54.实施方案53的方法,其中所述转座酶包含缺乏转座酶活性的单体亚

单位。

[0155] 实施方案55.实施方案53的方法,其中所述转座酶包含共价连接的单体亚单位。

[0156] 实施方案56.实施方案53-55中任一项的方法,其中所述转座酶的四级结构是单体的。

[0157] 实施方案57.实施方案53的方法,其中所述转座酶缺乏形成二聚体的能力。

[0158] 实施方案58.实施方案42-57中任一项的方法,其中所述转座酶选自下组:Mu,MuE392Q,Tn5,超活性Tn5,Tn5变体,Vibhar,RAG和Tn552。

[0159] 实施方案59.实施方案42-58中任一项的方法,其中一个或多个转座子核酸是非功能性的。

[0160] 实施方案60.实施方案59的方法,其中所述非功能性转座子核酸的3'末端选自下组:双脱氧基基团,间隔物基团,胺基,烷基基团,芳基基团,磷酸根基团,硫醇基基团,反向核苷酸,叠氮基基团,硫酸根基团和生物素基团。

[0161] 实施方案61.实施方案59的方法,其中通过使所述转座酶与非功能性转座子核酸和功能性转座子核酸接触来制备所述多个转座体。

[0162] 实施方案62.实施方案61的方法,其中构成非功能性转座子核酸与功能性转座子核酸的转座子核酸的比率大于或等于1:1。

[0163] 实施方案63.实施方案61的方法,其中构成非功能性转座子核酸与功能性转座子核酸的转座子核酸的比率大于或等于10:1。

[0164] 实施方案64.实施方案44-63中任一项的方法,其中(c)包括将扩增衔接头(adaptor)附着至所述靶核酸。

[0165] 实施方案65.实施方案64的方法,其中所述扩增衔接头包含选自下组的序列:锚定位点,测序引物位点,扩增引物位点,和报告标签。

[0166] 实施方案66.实施方案44-65中任一项的方法,其包括扩增捕获的核酸。

[0167] 实施方案67.实施方案66的方法,其中所述扩增捕获的核酸包括桥式扩增。

[0168] 实施方案68.实施方案44-67中任一项的方法,其中所述表面包含多个捕获探针。

[0169] 实施方案69.实施方案68的方法,其中所述捕获探针包含核酸。

[0170] 实施方案70.实施方案68的方法,其中所述捕获探针各自包含亲和部分。

[0171] 实施方案71.实施方案70的方法,其中所述亲和部分选自下组:生物素,亲合素,链霉亲合素和重组酶。

[0172] 实施方案72.实施方案42-71中任一项的方法,其中所述转座子核酸包含选自下组的序列:锚定位点,测序引物位点,扩增引物位点和报告标签。

[0173] 实施方案73.实施方案42-72中任一项的方法,其中至少一个转座体包含两个转座子核酸。

[0174] 实施方案74.实施方案73的方法,其中所述两个转座子核酸具有不同的序列。

[0175] 实施方案75.实施方案42-74中任一项的方法,其中所述多个转座体包含至少两个不同的转座子核酸。

[0176] 实施方案76.实施方案1-75中任一项的方法,其中所述多个转座体包含至少一个具有形成转座体的能力但缺乏转座能力的转座酶。

[0177] 实施方案77.实施方案42-76中任一项的方法,其中所述靶核酸选自下组:基因组

DNA, 基因组DNA的片段和cDNA。

[0178] 实施方案78. 实施方案77的方法, 其中所述靶核酸是基因组DNA。

[0179] 实施方案79. 实施方案44-78中任一项的方法, 其中所述表面在选自下组的基底上: 珠粒, 载玻片, 流动池, 通道, 浸渍片和孔。

[0180] 实施方案80. 实施方案44-79中任一项的方法, 其中所述表面包含每 mm^2 至少约10,000个捕获的核酸。

[0181] 实施方案81. 实施方案44-80中任一项的方法, 其中所述表面包含每 mm^2 至少约100,000个捕获的核酸。

[0182] 实施方案82. 实施方案44-81中任一项的方法, 其中所述表面包含每 mm^2 至少约1,000,000个捕获的核酸。

[0183] 实施方案83. 通过实施方案42-82中任一项的方法制备的包含条形码的测序文库。

[0184] 实施方案84. 实施方案41-83中任一项的方法, 其还包括在使所述转座体与所述靶核酸接触之后除去所述转座酶。

[0185] 实施方案85. 实施方案84的方法, 其中通过SDS, 尿素, 热或蛋白酶处理除去所述转座酶。

[0186] 实施方案86. 实施方案1-85中任一项的方法, 还包括组合条形码化 (combinatorial barcoding), 其中所述转座子包含第一组条形码;

[0187] 在转座期间将所述第一组条形码引入所述靶核酸以产生包含第一组条形码的转座的靶核酸;

[0188] 合并所述转座的靶核酸以产生转座的靶核酸的第一合并物;

[0189] 将第二组条形码引入到所述转座的靶核酸的第一合并物以产生包含第一和第二组条形码的靶核酸;

[0190] 将包含第一和第二组条形码的靶核酸合并以产生转座的靶核酸的第二合并物;

[0191] 任选地重复引入另外的条形码和合并的步骤以产生有条形码的 (barcoded) 靶核酸的文库。

[0192] 附图描述

[0193] 图1描述了其中使靶核酸与包含转座子核酸的转座体群体接触的示例性实施方案。

[0194] 图2描绘了示例性实施方案, 其中使包含不同转座子核酸的转座体群体与靶核酸接触, 并且在不同切口位点处将不同的转座子核酸附着于靶核酸的链。

[0195] 图3描述了其中通过线性扩增来扩增经修饰的靶核酸以获得某些扩增产物的示例性实施方案。

[0196] 图4描绘了示例性实施方案, 其中转座体包含二聚体转座酶, 并且转座子核酸包含两个转座子元件, 所述转座子元件包含嵌合元件 (ME), 其中ME之一在3'末端用双脱氧基团封闭。

[0197] 图5描述了示例性实施方案, 其中包含不同条形码的转座体群体接触靶核酸的; 转座子核酸附着于切口位点的一侧, 而将转座子核酸的另一非附着末端通过连接附着于切口位点的另一侧; 并且通过全基因组扩增 (WGA) 扩增经修饰的靶核酸。

[0198] 图6描述了不用转座子 (扩增子) 或用包含转座酶和 (1) 用3'生物素基团 (3' Bio) 封

闭的转座子核酸；(2) 用3'间隔物基团(3'间隔物)封闭的转座子核酸；或(3) 非封闭的转座子核酸(TDE1)的转座酶处理基因组DNA的结果。

[0199] 图7显示了名义倍数覆盖率和具有100bp插入频率的500bp读段和具有50bp插入频率的300bp读段的平均合成读取长度的图。

[0200] 图8A和8B显示了具有Y形接头转座子的单侧转座。

[0201] 图9显示加载有来自单侧转座反应的样品的琼脂糖凝胶的照片。

[0202] 图10显示加载有来自用转座子的n+1和n-1变体进行的转座反应运行的样品的琼脂糖凝胶的照片。

[0203] 图11显示用活性和无活性转座体的混合物进行的转座反应的图。

[0204] 图12显示对靶核酸产生切口并连接寡核苷酸接头的示例性方案。

[0205] 图13显示使用单侧转座进行的文库制备的示例性方案。

[0206] 图14显示出了通过利用对通过DNA的两条链的转座的差异抗性进行的单侧转座的示例性方案。

[0207] 图15显示出与组合条形码化偶联的单侧转座的示例性方案。使用单侧转座，单侧产物本身可以维持邻接，而不需要外部机制。将独特的但不能区分的分子(A,B和C)包含在一起。将它们随机分入到单独的反应中，其中加入模块条形码。尽管在每个步骤中分离的反应的数目少于分子的数目，但是通过反应的路径趋向于对于每个分子是独特的，导致每个分子的独特条形码组合。

[0208] 发明详述

[0209] 本文提供的实施方案涉及用于下一代测序的方法和组合物。一些实施方案包括使用单侧转座从靶核酸制备模板文库，并测序模板文库。在一些实施方案中，单侧转座包括用转座酶对双链核酸的链产生切口，以及在切口位点的一侧将转座子核酸附着到有切口的链。有利地，与双侧转座(例如Nextera™)相比，单侧转座不会使双链靶核酸片段化。因此，对于某些靶核酸，如基因组DNA，可以保留邻接，单倍型和/或定相信息。

[0210] 本文提供的方法和组合物的一些实施方案包括具有单侧转座酶活性的转座体，使用这种转座体制备测序文库，以及测序此类文库。在一些实施方案中，转座体可以包括具有单侧转座酶活性的转座酶。在一些实施方案中，转座体可以包括转座子核酸，其可以具有抑制转座子插入双链靶核酸的两条链中的封闭基团。转座酶还包括来自逆转录转座子(retrotransposon)的整合酶和逆转录病毒转座酶。示例性的转座酶包括但不限于Mu，Tn10，Tn5和超活性Tn5(Goryshin and Reznikoff, J.Biol.Chem., 273:7367 (1998))。与本文提供的一些方法和组合物一起使用的转座酶的实施方案包括公开于美国专利申请公开号2010/0120098的那些，其通过引用整体并入本文。转座酶和转座子元件的更多实施方案包括超活性Tn5转座酶和Tn5型转座酶元件(Goryshin和Reznikoff, J.Biol.Chem., 273:7367 (1998)，其通过引用整体并入本文)，MuA转座酶和包含R1和R2末端序列的Mu转座酶元件(Mizuuchi, Cell, 35:785, (1983) 和Savilahti等人, EMBO J., 14:4893, 15 (1995)，将每篇所述文献通过引用整体并入本文)。与超活性Tn5转座酶(例如，EZ-Tn5™转座酶，Epicentre Biotechnologies, Madison, Wisconsin)形成复合物的转座酶元件的实例阐述于WO 2012/061832; U.S. 2012/0208724, U.S. 2012/0208705和WO2014018423，其各自通过引用整体并入本文。用于本文提供的一些方法和组合物的转座酶和转座子序列的更多实施方案包括金黄

色葡萄球菌Tn552 (*Staphylococcus aureus* Tn552) (Colegio et al., *J. Bacteriol.*, 183: 2384-8 (2001); Kirby et al., *Mol. Microbiol.*, 43: 173-86 (2002)), Tyl (Devine & Boeke, *Nucleic Acids Res.*, 22: 3765-72 (1994) and WO 95/23875), 转座子Tn7 (Craig, *Science* 271: 1512 (1996); Craig, *Curr Top Microbiol Immunol.*, 204: 27-48 (1996)), Tn/0和IS10 (Kleckner et al., *Curr Top Microbiol Immunol.*, 204: 49-82 (1996)), Mariner转座酶 (Lampe et al., *EMBO J.*, 15: 5470-9, (1996)), Tel (Plasterk, *Curro Topics Microbiol. Immunol.*, 204: 125-43, (1996)), P元件 (Gloor, *Methods Mol. Biol.*, 260: 97-114, (2004)), Tn3 (Ichikawa & Ohtsubo, *J. Biol. Chem.* 265: 18829-32, (1990)), 细菌插入序列 (Ohtsubo & Sekine, *Curro Top. Microbiol. Immunol.* 204: 1-26, (1996)), 逆转录病毒 (Brown, et al., *Proc Natl Acad Sci USA*, 86: 2525-9, (1989)), 和酵母的反转录转座子 (Boeke & Corces, *Annu Rev Microbiol.* 43: 403-34, (1989))。更多实例包括IS5, Tn10, Tn903, IS911和转座酶家族酶的工程化形式 (Zhang et al., *PLoS Genet.* 5: e1000689. Epub 2009 Oct 16; and Wilson et al. *Microbiol. Methods* 71: 332-5 (2007))。更多实例包括MuA转座酶 (参见例如 Rasila TS, et al., (2012) *PLoS ONE* 7 (5): e37922. doi:10.1371/journal.pone.0037922)。可用于本文提供的方法和组合物的一些实施方案的转座酶的实例描述于 Leschziner, A.E., et al., (1998) *P.N.A.S.* 95: 7345-7350; 以及 Haapa S., et al., (1999) *N.A. Res.* 27: 2777-2784, 其各自通过引用整体并入。Tn5转座酶的变体, 如具有氨基酸取代, 插入, 缺失和/或与其它蛋白质或肽的融合的变体, 公开于美国专利: 5,925,545; 5,965,443; 7,083,980; 7,608,434; 和美国专利申请14/686,961。将这些专利和专利申请通过引用整体并入本文。在一些实施方案中, 相对于野生型蛋白质, Tn5转座酶在位置54, 56, 372, 212, 214, 251, 和338处包含一个或多个取代, 如美国专利申请14/686,961中所公开的。在一些实施方案中, Tn5野生型蛋白或其变体可以进一步包含融合多肽。在一些实施方案中, 与转座酶融合的多肽结构域可以包含例如伸长因子Ts。将本段中引用的每个参考文献通过引用整体并入本文。

[0211] 在一些实施方案中, 使双链靶核酸与多个转座体接触, 使得靶核酸的链产生切口, 并且转座子核酸在切口位点的一侧附着到有切口的靶核酸的链, 以获得经修饰的靶核酸。在一些实施方案中, 转座体接触溶液中的靶核酸。在该实施方案中, 经修饰的靶核酸可以在溶液中产生并随后捕获于表面上。或者, 可以发生在表面上转座体和靶核酸之间的接触。转座体或靶核酸可以在进行接触之前附着于表面。由转座体和表面上的靶核酸之间的接触产生的经修饰的靶核酸可以保持在表面上捕获, 或者经修饰的靶核酸可以从表面释放。

[0212] 在一些实施方案中, 对捕获的核酸测序。在一些实施方案中, 在靶核酸序列的线性表示中从两个捕获的核酸获得的序列信息的接近性指示所捕获的核酸在表面上的接近性。在一些实施方案中, 在表面上彼此更接近的捕获的核酸包括与在较不接近的捕获的核酸相比在靶核酸序列的表示中更接近的序列。在一些实施方案中, 靶核酸序列的表示包括单倍型或装配表示。

[0213] 本文提供的方法和组合物的一些实施方案还包括在靶核酸的测序片段的从头装配中使用单侧转座。在一些实施方案中, 将界标 (landmark) 插入靶核酸中并且可以用于靶核酸的测序片段的装配中以产生靶核酸序列的表示。在一些实施方案中, 重叠片段可以包括共同插入的界标。使用界标对于包含高度重复序列的靶核酸是特别有利的。此外, 在一些

实施方案中,不需要参考序列。

[0214] 在一些实施方案中,通过使靶核酸与具有单侧转座酶活性的转座体群体和包含不同条形码的转座子核酸接触,将界标插入到靶核酸中。在一些实施方案中,通过单侧转座,然后连接,将转座子核酸插入到靶核酸的单链中。在一些实施方案中,转座酶在靶核酸的链中产生切口,并且将转座子核酸在有切口的位点处附着到有切口的靶核酸的一条链,并且转座子核酸的另一端在有切口侧的另一侧连接到的有切口的靶核酸,从而获得在构成环的链中具有插入的经修饰的双链靶核酸。在一些实施方案中,可以扩增和测序经修饰的核酸。在一些实施方案中,经修饰的核酸可以附着到表面。在一些实施方案中,可以通过单链结合蛋白或结合单链环的蛋白质(如重组酶)进行附着。

[0215] 在一些实施方案中,在没有转座的情况下修饰靶核酸。在一些实施方案中,使用切口内切核酸酶,例如来自New England Biolabs,MA,USA的切口核酸内切酶,或限制性内切核酸酶可以使靶核酸随机化。示例性的限制性内切核酸酶包括但不限于EcoRI,EcoRII,BamHI,HindIII,TaqI,NotI。限制性内切核酸酶의 其它实例可以在New England Biolabs目录中找到。任选地,可以用具有3'或5'核酸外切酶活性的酶,例如用核酸外切酶I或核酸外切酶II或核酸外切酶III延伸缺口。寡核苷酸接头可以连接到靶核酸的有切口的末端。在一些实施方案中,寡核苷酸接头可以包括引物结合位点,如测序引物位点和扩增引物位点,另外的序列还可以包括切割位点,独特的分子指数,锚定位点,报道标签和条形码。因此,使靶核酸产生切口并连接一个或多个接头保持靶核酸完整,而没有片段化。使靶核酸产生切口并连接寡核苷酸接头的示例性方案显示于图12中。

[0216] 如本文中所用,“核酸”包括连接在一起的至少两个核苷酸单体。实例包括但不限于DNA,如基因组或cDNA;RNA,如mRNA,sRNA或rRNA;或DNA和RNA的杂合体。从以下实例和本文其它地方显而易见,核酸可以具有天然存在的核酸结构或非天然存在的核酸类似物结构。核酸可以含有磷酸二酯键;然而,在一些实施方案中,核酸可以具有其它类型的主链,包括例如磷酰胺,硫代磷酸酯,二硫代磷酸酯,0-甲基亚磷酰胺和肽核酸主链和连接。核酸可以具有正的骨架;非离子主链和非核糖基主链。核酸也可以包含一个或多个碳环糖。在本文中的方法或组合物中使用的核酸可以是单链或备选地如规定的双链。在一些实施方案中,核酸可以含有双链和单链序列的部分,例如,如通过叉式接头所证明的。核酸可以包含脱氧核糖核苷酸和核糖核苷酸的任何组合,以及碱基的任何组合,包括尿嘧啶,腺嘌呤,胸腺嘧啶,胞嘧啶,鸟嘌呤,肌苷,黄嘌呤,次黄嘌呤,异胞嘧啶,异鸟嘌呤和碱基类似物,如硝基吡咯(包括3-硝基吡咯)和硝基咪唑(包括5-硝基咪唑)等。在一些实施方案中,核酸可以包括至少一个混杂碱基。混杂碱基可以与多于一种不同类型的碱基进行碱基配对,并且例如当包括在寡核苷酸引物或插入物中时可以有用的,所述寡核苷酸引物或插入物用于在复杂核酸样品,如基因组DNA样品中随机杂交。混杂碱基的实例包括可与腺嘌呤,胸腺嘧啶或胞嘧啶配对的肌苷。其它实例包括次黄嘌呤,5-硝基咪唑,无环5-硝基咪唑,4-硝基吡唑,4-硝基咪唑和3-硝基吡咯。可以使用可与至少两种,三种,四种或更多类碱基进行碱基配对的混杂碱基。

[0217] 如本文中所用,“核苷酸序列”包括核酸聚合物中核苷酸单体的顺序和类型。核苷酸序列是核酸分子的特征并且可以以多种形式中的任一种表示,包括例如描述,图像,电子介质,符号系列,数字系列,字母系列,系列颜色等。信息可以例如以单核苷酸分辨率,以较

高分辨率(例如指示核苷酸亚单位的分子结构)或以较低分辨率(例如指示染色体区域,如单倍型区块)表示。一系列“A”,“T”,“G”和“C”字母是公知的序列,其表示可以在单核苷酸分辨率下与DNA分子的实际序列相关联的DNA。除了在系列中用“U”代替“T”外,类似的表示法用于RNA。

[0218] 当用于指核酸时,如本文所使用的,术语“不同的”是指核酸具有彼此不同的核苷酸序列。两个或更多个核酸可以具有沿其整体长度不同的核苷酸序列。或者,两个或更多个核酸可以具有沿其长度的主要部分不同的核苷酸序列。例如,两个或更多个核酸可以具有对于两个或更多个分子不同的靶核苷酸序列部分,同时还具有在两个或更多个分子上相同的通用序列部分。通用序列可以存在于核酸的末端,或者在待复制,检测或扩增的核酸区域的侧翼。

[0219] 如本文所使用的,“单倍型”包括个体从其亲本之一遗传的多于一个位点的一组等位基因。单倍型可以包括来自染色体的全部或部分的两个或更多个基因座。等位基因包括例如单核苷酸多态性(SNP),短串联重复(STR),基因序列,染色体插入,染色体缺失等。术语“定相等位基因”是指来自特定染色体或其部分的特定等位基因的分布。因此,两个等位基因的“相位”可以指一个或多个染色体上的两个或多个等位基因的相对位置的表征或表示。

[0220] 如本文所用,核酸中的“切口”是指双链核酸的区域,其中两条链中的仅一条含有切割的主链结构。因此,“产生切口”是指打破双链核酸区域内的仅一条核酸链的共价结构的动作。该区域通常仅是双链核酸的一部分。该部分可以包括例如至多5个碱基对,10个碱基对,25个碱基对,50个碱基对,100个碱基对,200个碱基对,300个碱基对,400个碱基对,500个碱基对,1000个碱基对。所述区域可以包括双链核酸的较大部分或较小部分。例如,上文例示的上限的备选或在上文例示的上限外,产生切口的核酸部分的下限可以任选地为至少500个碱基对,400个碱基对,300个碱基对,200个碱基对,100个碱基对,50个碱基对,25个碱基对,10个碱基对或更小。在上述范围中列出的值可以限定核酸区域群体的所有成员的最大或最小尺寸,或者备选可以指具有该区域的核酸群体的平均值。应当理解,双链核酸可以在两条链中产生切口,第一切口出现在第一区域中,而第二切口出现在第二区域中。通常,在核酸保持杂交的双链形式的条件下,可以维持两个有切口的区域的有效连接性。相反,在双链核酸的相同区域中切割两条链可导致在切割位点侧翼的核酸区域之间的有效连接性的损失。

[0221] 如本文所用,术语“表面”旨在表示与周围流体,如气态流体或液体流体直接接触的固体支持物或凝胶材料的部分或层。该表面可以与另一种材料接触,如气体,液体,凝胶,聚合物,有机聚合物,相似或不同材料的第二表面,金属或涂层(coat)。表面或其区域可以是基本平坦的。表面可以具有表面特征,如孔,凹坑(pit),通道,脊(ridge),凸起区域(raised region),桩(pegs),柱(post)等。在多孔基底的情况下,表面可以位于流体与基底接触的孔中。例如,表面可以存在于凝胶的孔中,其中附着的部分与进入孔的流体相互作用。因此,在表面“上”的部分可以位于多孔材料,如凝胶的孔中。

[0222] 如本文所用,术语“固体支持物”是指不溶于水性液体的刚性基底。基底可以是无孔的或多孔的。基底可以任选地能够吸收液体(例如由于多孔性),但是通常将具有足够的刚性,使得当吸收液体时基底基本上不会溶胀,并且当通过干燥除去液体时基底基本上不会收缩。无孔固体支持物通常对液体或气体是不可渗透的。示例性的固体支持物包括但不

限于玻璃和改性或功能化玻璃,塑料(包括丙烯酸类,聚苯乙烯,苯乙烯和其它材料的共聚物,聚丙烯,聚乙烯,聚丁烯,聚氨酯,TeflonTM,环烯烃,聚酰亚胺等),尼龙,陶瓷,树脂,Zeonor,二氧化硅(silica)或二氧化硅基材料,包括硅和改性硅,碳,金属,无机玻璃,光纤束和聚合物。对于一些实施方案,特别有用的固体支持物位于流动池装置内。

[0223] 如本文所用,术语“凝胶材料”旨在表示可渗透液体和气体的半刚性基底。通常,当液体被吸收时,凝胶材料可以膨胀,并且当通过干燥除去液体时凝胶材料可以收缩。示例性的凝胶包括但不限于具有胶体结构的那些,如琼脂糖;聚合物网状结构,如明胶;或交联聚合物结构,如聚丙烯酰胺,SFA(参见例如美国专利申请公开号2011/0059865A1,其通过引用并入本文)或PAZAM(参见例如美国专利申请公开号61/753,833,其通过引用并入本文)。

[0224] 如本文所用,术语“附着”旨在表示通过防止扩散分离的力连接。该术语可以包括天然共价或非共价的连接。例如,核酸可以通过创建键链的一个或多个共价键共价附着到表面。当两种物质之间(例如在核酸和表面之间)的至少一个键不是共价键时,发生非共价附着。非共价键的实例包括例如氢键,离子键,范德华力,疏水键等。

[0225] 如本文所使用,术语“邻接信息”是指基于共享信息的两个或更多个DNA片段之间的空间关系。信息的共享方面可以是就相邻,区划和距离空间关系而言。关于这些关系的信息继而促进从DNA片段衍生的序列读段的分层装配或定位(mapping)。这种邻接信息改善了这种装配或定位的效率和准确性,因为与常规鸟枪测序结合使用的传统装配或定位方法没有考虑单个序列读取的相对基因组起点或坐标,这是因为它们涉及衍生出单个序列读段的两个或更多个DNA片段之间的空间关系。因此,根据本文描述的实施方案,捕获邻接信息的方法可以通过用于确定相邻空间关系的短程邻接方法,用于确定区室空间关系的中程邻接方法或用于确定距离空间关系的长程邻接方法来实现。这些方法有助于DNA序列装配或定位的准确性和质量,并且可以用于任何测序方法,如上述的那些方法。

[0226] 在一些实施方案中,该步骤导致产生源自靶DNA序列的鸟枪核酸分子文库。在替代实施方案中,甚至可以通过如下所述的Y衔接头方法来实现片段化或插入。一个或多个转座酶分子可以是可溶性游离转座酶或可以与表面结合的识别序列相关。

[0227] 如本文所用,术语“条形码”是指对靶核酸序列而言是独特的且完全独立的核酸序列。通常,条形码可包括可用于鉴定一种或多种特定核酸的一个或多个核苷酸序列。条形码可以是人工序列,或可以是在转座期间产生的天然存在的序列,如在先前并置的DNA片段末端的相同侧翼基因组DNA序列(g-代码)。条形码可以包含至少约1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20或更多个连续核苷酸。在一些实施方案中,条形码包含至少约10,20,30,40,50,60,70,80,90,100或更多个连续核苷酸。在一些实施方案中,包含条形码的核酸群体中的条形码的至少一部分是不同的。在一些实施方案中,条形码的至少约10%,20%,30%,40%,50%,60%,70%,80%,90%,95%,99%是不同的。在多个此类实施方案中,所有条形码是不同的。包含条形码的核酸群体中不同条形码的多样性可以是随机产生的或非随机产生的。

[0228] 在一些实施方案中,转座子序列包含至少一个条形码。在一些实施方案中,如包含两个非连续转座子序列的转座体,第一转座子序列包含第一条形码,并且第二转座子序列包含第二条形码。在一些实施方案中,转座子序列包含条形码,其包含第一条形码序列和第二条形码序列。在一些前述实施方案中,第一条形码序列可以被鉴定或指定为与第二条形

码序列配对。例如,可以使用包括已知彼此配对的多个第一和第二条形码序列的参考表知道已知的第一条形码序列与已知的第二条形码序列配对。

[0229] 在另一实例中,第一条形码序列可以包括与第二条形码序列相同的序列。在另一个实例中,第一条形码序列可以包括第二条形码序列的反向互补物。在一些实施方案中,第一条形码序列和第二条形码序列不同。第一和第二条形码序列可以包括双码(bi-code)。

[0230] 在本文所述的组合物和方法的一些实施方案中,条形码用于模板核酸的制备。如将理解的,大量可用的条形码允许每个模板核酸分子包含独特的鉴定。在模板核酸混合物中每个分子的独特的鉴定可用于几种应用中。例如,可以将独特鉴定的分子应用于在具有多个染色体的样品中,基因组中,细胞中,细胞类型中,细胞疾病状态中,以及物种中,例如单倍型测序中,亲本等位基因区分中,在宏基因组测序中,以及基因组的样品测序中鉴定单个核酸分子。

[0231] 在一些实施方案中,可以在转座期间插入贯穿靶核酸的多个独特条形码。在一些实施方案中,每个条形码包括第一条形码序列和第二条形码序列,具有设置在其间的片段化位点。第一条形码序列和第二条形码序列可以鉴定或指定为彼此配对。配对可以是信息性的,使得第一条形码与第二条形码相关联。有利地,配对的条形码序列可用于装配来自模板核酸文库的测序数据。例如,鉴定包含第一条形码序列的第一模板核酸和包含与第一条形码序列配对的第二条形码序列的第二模板核酸指示第一和第二模板核酸代表靶核酸的序列表示中彼此相邻的序列。此类方法可用于从头靶核酸的序列表示,而不需要参照基因组。

[0232] 如本文所使用的,术语“至少一部分”和/或其语法等同物可以指整个量的任何分数。例如,“至少一部分”可以指整个量的至少约1%,2%,3%,4%,5%,6%,7%,8%,9%,10%,15%,20%,25%,30%,35%,40%,45%,50%,55%,60%,65%,70%,75%,80%,85%,90%,95%,99%,99.9%或100%。

[0233] 如本文所用,术语“约”是指 $\pm 10\%$ 。

[0234] 靶核酸

[0235] 本文中提供的方法和组合物的一些实施方案包括靶核酸。在一些实施方案中,靶核酸包括双链核酸。在一些实施方案中,靶核酸包括基因组DNA或cDNA。在一些实施方案中,使用线粒体或叶绿体DNA。在一些实施方案中,靶核酸包括RNA或其衍生物,如mRNA或cDNA。本文中所述的一些实施方案可利用以一个拷贝(即单个分子)存在或者备选以多个拷贝(即具有相同序列的核酸分子的整体)存在的单个靶核酸种类。其它实施方案可以利用多种不同的靶核酸种类(例如,具有多种不同核苷酸序列的核酸分子)。因此,多个靶核酸可以包括多个相同的靶核酸,其中的一些靶核酸相同的多个不同的靶核酸,或所有靶核酸不同的多个靶核酸。靶核酸可以从获自单个生物体的核酸分子或从包括多于一种生物体的来源获得的核酸分子群制备。靶核酸可以来自单个细胞;来自单个生物体的多个细胞,组织或体液;来自同一物种的几种生物体的细胞,组织或体液;或来自多个物种(如宏基因组(metagenomic)样品一样),如来自环境样品。核酸分子的来源包括但不限于细胞器,细胞,组织,器官或生物体。

[0236] 在一些实施方案中,使靶核酸与转座子接触,使得转座子核酸插入或附着到靶核酸上以提供经修饰的核酸。在一些实施方案中,可以进一步操作经修饰的核酸,例如延伸,

扩增和连接。

[0237] 转座体

[0238] 本文提供的方法和组合物的一些实施方案包括转座体。在一些实施方案中,转座体包括与一个或多个转座子核酸结合的转座酶。在一些实施方案中,转座体包含单侧转座酶活性,其包括对双链核酸的链产生切口,并在切口位点一侧将转座子核酸附着至有切口的链。

[0239] 在一些实施方案中,具有单侧转座酶活性的转座体包括包含某些类型的具有单侧转座酶活性的转座酶的转座体。在一些实施方案中,野生型转座酶具有单侧转座酶活性或被修饰为具有单侧转座酶活性。具有单侧转座酶活性或可被修饰为具有单侧转座酶活性的转座酶的实例包括Mu, Mu E392Q, Tn5, RAG, 超活性Tn5, Tn5变体, Vibhar和Tn552 (Leschziner, A.E., et al., (1998) P.N.A.S. 95:7345-7350; 和Haapa S., et al., (1999) N.A.Res. 27:2777-2784, 其各自通过引用整体并入)。本文列出了具有单侧转座酶活性或可被修饰为具有单侧转座酶活性的转座酶的更多实例。在一些实施方案中,具有单侧转座酶活性的转座体包含单个单体和转座子核酸。在一些实施方案中,转座酶可以被修饰以缺乏形成二聚体的能力。在一些实施方案中,具有单侧转座酶活性的转座体包含其中单体之一缺乏转座酶活性的二聚体。在一些实施方案中,二聚体的单体亚单位可以是共价连接的。

[0240] 在一些实施方案中,具有单侧转座酶活性的转座体包含封闭的转座子核酸。在一些实施方案中,封闭的转座子核酸被封闭而不能附着于有切口的双链核酸的链。封闭的转座子核酸可以包括在转座子核酸的3'端,抑制转座子核酸与另一个核酸附着的封闭基团。在一些实施方案中,封闭基团可以包括双脱氧基团,间隔物基团和生物素基团。在一些实施方案中,具有单侧转座酶活性的转座体群体可以通过使转座酶与封闭的转座子核酸和非封闭的转座子核酸接触来制备。非封闭的转座子核酸包括缺少封闭基团的转座子核酸。在一些实施方案中,获得包含转座体的群体,所述转座体包含转座酶二聚体,所述转座酶二聚体包含封闭的转座子核酸和非封闭的转座子核酸。在一些实施方案中,转座酶二聚体包含两个封闭的转座子核酸。在一些实施方案中,转座酶二聚体包含两个非封闭的转座子核酸。在一些实施方案中,群体中不同类型的二聚体的比例可以通过使转座酶与各种比率的封闭的转座子核酸与非封闭的转座子核酸接触来操作。在一些实施方案中,封闭的转座子核酸与非封闭的转座子核酸的比率大于或等于1:1, 5:1, 10:1, 50:1, 100:1, 200:1, 500:1, 1000:1, 或前述比率之间的任何范围。

[0241] 其它有用的转座子核酸是比标准转座子短或长的那些。例如,可以使在3'末端转移的链更短(例如通过除去一个或多个碱基)以抑制转移反应的发生。类似地,3'末端可以更长以导致这种抑制。

[0242] 用于单侧转座的方法也可用于调节转座后文库的插入物大小。这种方法的优点是插入物大小的长度可以通过活性和无活性复合物之间的比率来确定,在许多情况下,其比温育或浓缩转座体和核酸的时间更容易控制。

[0243] 可以制备具有活性转座体单体亚单位和无活性转座体亚单位的转座体二聚体,并在转座反应开始之前或之后加入核酸。例如,可以通过加入Mg²⁺开始反应。在特定实施方案中,可以形成包含三种转座体二聚体(即活性:活性二聚体,无活性:无活性二聚体,和活性:无活的二聚体)的混合物的群体。具有不同活性的群体可以通过改变被组合形成混合物的

活性和无活性转座体亚单位的比率来制备。可以选择该比率以影响当用混合物处理核酸样品时产生的平均插入物大小。可以通过使用转座体种类的混合物来实现插入物大小的类似控制,所述转座体种类仅具有活性:活性和无活性:无活性的两种转座体种类。如图11中的图所示,无活性:无活性种类能够结合靶DNA但不能转座靶,将起间隔物作用。换句话说,无活性:无活性二聚体竞争在其它情况下将结合活性:活性二聚体并转座的位点。掺入转座酶反应混合物中的无活性:无活性二聚体的量的常规滴定可用于控制由混合物产生的平均片段大小。这些方法具有若干优点,包括例如相对不依赖于时间的且可控的。常规的转座反应(例如来自Illumina, Inc. (San Diego, CA)的Nextera[®]样品制备方法)需要仔细控制反应时间段以实现产生所需平均尺寸的片段的转座反应。另一方面,单侧转座的当前方法可以如上所述进行,从而对时间较不敏感。更具体地,可以选择活性单体亚单位与无活性单体亚单位的比率以确定文库中片段的大小。

[0244] 在一些实施方案中,具有单侧转座酶活性的转座体可以附着于表面。转座体可以通过转座酶或通过转座子核酸附着。例如,转座酶可以共价或非共价地附着到表面。或者/另外,转座子核酸可以共价或非共价地附着到表面。可用于其制备和使用的附着,表面和相关方法在本文和美国专利申请系列号13/790,220中进一步详细阐述,其通过引用并入本文。

[0245] 在一些实施方案中,转座酶包括能够与包含转座子元件或转座酶元件的转座子核酸形成功能性复合物,并催化所述转座子序列插入或转座到靶核酸中以提供经修饰的核酸的酶。例如,在体外转座反应中,将转座子核酸插入靶DNA中以提供经修饰的DNA。在一些实施方案中,转座酶包括能够与包含转座子元件或转座酶元件的转座子核酸形成功能性复合物,并催化单侧转座到靶核酸中以提供经修饰的核酸的酶。

[0246] 在一些实施方案中,通过转座酶可以在靶核酸中的随机或基本上随机的位点插入或附着转座子核酸。转座酶还包括来自反转录转座子的整合酶和逆转录病毒转座酶。与本文中提供的一些方法和组合物一起使用的转座酶的实施方案包括公开于US 2010/0120098中公开的那些转座酶,其通过引用整体并入本文。转座酶和转座子元件的更多实施方案包括超活性Tn5转座酶和Tn5型转座酶元件(Goryshin and Reznikoff, J. Biol. Chem., 273: 7367 (1998), 其通过引用整体并入本文), MuA转座酶和包含R1和R2末端序列的Mu转座酶元件(Mizuuchi, Cell, 35: 785, (1983) and Savilahti, et al., EMBO J., 14: 4893, 15 (1995), 通过引用整体并入本文)。与超活性Tn5转座酶(例如, EZ-Tn5[™]转座酶, Epicentre Biotechnologies, Madison, Wisconsin)形成复合物的实例转座酶元件阐述于WO 2012/061832; U.S. 2012/0208724, U.S. 2012/0208705和WO 2014018423, 其各自通过引用整体并入本文。用于本文提供的一些方法和组合物的转座酶和转座子核酸的更多实施方案包括金黄色葡萄球菌(*Staphylococcus aureus*) Tn552 (Colegio et al., J. Bacteriol., 183: 2384-8 (2001); Kirby et al., Mol. Microbiol., 43: 173-86 (2002)), Tyl (Devine & Boeke, Nucleic Acids Res., 22: 3765-72 (1994) 和WO 95/23875), 转座子Tn7 (Craig, Science 271: 1512 (1996); Craig, Curr Top Microbiol Immunol., 204: 27-48 (1996)), Tn/O和IS10 (Kleckner et al., Curr Top Microbiol Immunol., 204: 49-82 (1996)), Mariner转座酶 (Lampe et al., EMBO J., 15: 5470-9, (1996)), Tel (Plasterk, Curr Topics Microbiol. Immunol., 204: 125-43, (1996)), P元件 (Gloor, Methods Mol. Biol., 260: 97-

114, (2004)), Mos-1转座酶(Richardson et al., EMBO Journal 25:1324-1334 (2006)), Tn3 (Ichikawa&Ohtsubo, J BioI.Chem.265:18829-32, (1990)), 细菌插入序列(Ohtsubo& Sekine, Curro Top.Microbiol.Immunol.204:1-26, (1996)), 逆转录病毒(Brown, et al., Proc Natl Acad Sci USA, 86:2525-9, (1989)) 和酵母的反转录转座子(Boeke&Corces, Annu Rev Microbiol.43:403-34, (1989))。更多实例包括IS5, Tn10, Tn903, IS911和转座酶家族酶的工程化改造形式(Zhang et al., PLoS Genet.5:e1000689. Epub2009Oct 16; and Wilson et al. Microbiol.Methods 71:332-5 (2007))。更多实例包括MuA转座酶(参见例如 Rasila TS, et al., (2012) PLoS ONE 7 (5):e37922. doi:10.1371/journal.pone.0037922)。本段中引用的每个参考文献通过引用整体并入本文。

[0247] 在一些实施方案中,转座子核酸包含双链核酸。转座子元件包括核酸分子或它的包括与转座酶或整合酶形成转座体的核苷酸序列的部分。在一些实施方案中,转座子元件能够在转座反应中与转座酶形成功能性复合物。本文提供了转座子元件的实例,并且包括19-bp外端(“OE”)转座子末端,内端(“IE”)转座子末端或由例如野生型或突变体Tn5转座酶识别的“马赛克末端(mosaic end)” (“ME”)转座子末端,或R1和R2转座子末端(参见例如US 2010/0120098,其通过引用整体并入本文)。转座子元件可以包括适于在体外转座反应中与转座酶或整合酶形成功能性复合物的任何核酸或核酸类似物。例如,转座子末端可以包含DNA, RNA, 修饰的碱基, 非天然碱基, 修饰的主链, 并且可以包含一条或两条链中的切口。

[0248] 在一些实施方案中,转座子核酸可以包括转座子元件和另外的序列。在一些实施方案中,另外的序列可以在转座反应中插入靶核酸中或附着至靶核酸中。另外的序列可以包括引物结合位点,如测序引物位点和扩增引物位点,另外的序列还可以包括切割位点,独特的分子指数,锚定位点,报告标签和条形码。

[0249] 在一些实施方案中,引物结合位点可以包括用于测序引物以在测序反应中与核酸退火的序列。在一些实施方案中,引物结合位点可包括针对引物的在扩增反应或其它延伸反应中与核酸退火的序列。

[0250] 在一些实施方案中,切割位点可以包括可被片段化的转座子核酸中的位点。例如,包含切割位点的转座子核酸可以插入靶核酸中,然后修饰的核酸可以通过在插入的切割位点处片段化。在一些实施方案中,切割位点包括限制性内切酶识别序列和/或限制酶切割位点。在一些实施方案中,切割位点可以包括核酸中的至少一种核糖核苷酸,其在其它情况下可以包含脱氧核糖核苷酸并且可以用RNA酶切割。可以使用能够选择性切割脱氧核糖核苷酸和核糖核苷酸之间的磷酸二酯键的化学切割剂,包括例如金属离子,如稀土金属离子(例如La³⁺, 特别是Tm³⁺、Yb³⁺或Lu³⁺, Fe (3) 或Cu (3)), 或暴露于升高的pH。在一些实施方案中,切割位点可以包括切口酶(即断裂双链核酸的特定区域中的一条链的切口内切核酸酶)的一个或多个识别序列。因此,片段化位点可以包括第一切口酶识别序列和任选地第二切口酶识别序列。第一和第二切口酶识别序列可以彼此相同或彼此不同。在一些实施方案中,切割位点可以包括一个或多个核苷酸类似物,其包含无碱基位点并且允许在某些化学剂,如多胺, N,N'-二甲基乙二胺(DMED)存在下在片段化位点处切割(参见例如,US 2010/0022403,其通过引用整体并入本文)。在一些实施方案中,无碱基位点可以通过修饰切割位点内的尿嘧啶核苷酸来创建,例如使用尿嘧啶DNA糖基化酶(UDG)酶。然后可以通过用内切核酸酶(例如Endo IV内切核酸酶, AP裂合酶, FPG糖基化酶/AP裂合酶, Endo VIII糖基化酶/AP裂合

酶),热或碱处理,在无碱基位点处切割包含无碱基位点的多核苷酸链。无碱基位点还可以在除脱氧尿苷之外的核苷酸类似物上产生,并通过用内切核酸酶,热或碱处理以类似的方式切割。例如,8-氧代-鸟嘌呤可以通过暴露于FPG糖基化酶而转化为无碱基位点。脱氧肌苷可以通过暴露于AlkA糖基化酶而转化为无碱基位点。然后,可以通常通过用合适的内切核酸酶,如Endo IV或AP裂合酶处理切割如此产生的无碱基位点(参见例如,U.S.2011/0014657,其通过引用整体并入本文)。在另一个实例中,切割位点可包括允许通过用高碘酸盐(例如高碘酸钠)处理切割的二醇连接。在另一个实例中,切割位点可以包括允许用化学还原剂,例如三(2-羧乙基)-磷酸盐(TCEP)切割的二硫化物基团。在一些实施方案中,切割位点可以包括可光切割的部分。光化学切割可以通过利用光能断裂共价键的多种方法中的任一种来进行。用于光化学切割的位点可以由核酸中的非核苷酸化学部分提供,如亚磷酰胺[4-(4,4'-二甲氧基三苯甲基氧基)丁酰胺基甲基)-1-(2-硝基苯基)-乙基]-2-氰乙基-(N,N-二异丙基)-亚磷酰胺(Glen Research,Sterling,Va.,USA,产品目录号10-4913-XX)。

[0251] 在一些实施方案中,转座子核酸可以包括锚定位点。在一些实施方案中,锚定位点可以包括可以特异性结合捕捉探针的序列。在一些实施方案中,锚定位点包含与包含核酸的捕捉探针互补和/或基本上互补的序列。在一些实施方案中,锚定位点可以包括配体或受体,其结合包含相应受体或配体的捕捉探针的。换言之,锚定位点和捕捉探针可以包含配体/受体对。在一些实施方案中,配体或受体可以通过修饰的核苷酸与转座子核酸的锚定位点缔合。配体和受体的实例包括分别能结合链霉亲和素或镍的生物素或多聚His。其它实例包括本领域已知的配体对及其受体,例如亲和素-生物素,链霉亲和素-生物素,以及生物素,链霉亲和素或亲和素的衍生物,包括但不限于2-亚氨基生物素(2-aminobiotin),脱硫生物素,NeutrAvidin(Molecular Probes,Eugene,Oreg.),CaptAvidin(Molecular Probes)等;结合蛋白/肽,包括麦芽糖-麦芽糖结合蛋白(MBP),钙-钙结合蛋白/肽(CBP);抗原抗体,包括表位标签,包括c-MYC,HA,VSV-G,HSV,V5和FLAG TagTM,及其相应的抗表位抗体;半抗原,例如二硝基苯和地高辛配基(digoxigenin)及其相应的抗体;适体及其相应靶标;多聚His标签(例如戊-His和己-His)及其结合配偶体,包括相应的固定化金属离子亲和层析(IMAC)材料和抗聚His抗体;荧光团和抗荧光团抗体;核酸链及其互补链等。

[0252] 在一些实施方案中,转座子核酸可以包括报告标签。有用的报告标签包括本领域已知的多种可识别标签,标记物或基团中的任一种。在某些实施方案中,报告标签可以发射信号。信号的实例包括荧光,化学发光,生物发光,磷光,放射性,量热或电化学发光的那些。示例性的报告标签包括荧光团,放射性同位素,色原,酶,包括表位标签的抗原,半导体纳米晶体如量子点,重金属,染料,磷光基团,化学发光基团,电化学检测部分,结合蛋白,磷光体(phosphors),稀土螯合物,过渡金属螯合物,近红外染料,电化学发光标记物和质谱仪相容的报告标签,如质量标签,电荷标签和同位素。可以与本文的方法和组合物一起使用的更多的报告标签包括光谱标记物,如荧光染料(例如异硫氰酸荧光素,德克萨斯红,罗丹明等);放射性标记(例如,³H、¹²⁵I、³⁵S、¹⁴C、³²P、³³P等);酶(例如,辣根过氧化物酶,碱性磷酸酶等);光谱比色标记物,如胶体金或有色玻璃或塑料(例如聚苯乙烯,聚丙烯,乳胶等);珠;磁性标签;电标签;热标签;和质量标签。

[0253] 在一些实施方案中,转座子核酸可以包括条形码。在一些实施方案中,转座子群体

可以包括此类转座子核酸,其包含相同条形码,一个或多个不同条形码,或其中每个转座子核酸可以包括不同的条形码。在一些实施方案中,插入到靶核酸中或附着至靶核酸的条形码可以用于鉴定靶核酸。在一些实施方案中,条形码可用于鉴定靶核酸中的插入事件。在一些实施方案中,转座体群中的每个转座体包括具有不同条形码的转座子核酸,所述条形码可用于鉴定靶核酸中的插入位点。在一些实施方案中,条形码可用于鉴定在切割位点处片段化后的插入位点,例如条形码跨越切割位点的地方。示例条形码及其制备和使用的方法阐述于国际公开号W0 2012/061832;US 2012/0208724,US 2012/0208705和PCT/US2013/031023,其各自通过引用整体并入本文。在一些实施方案中,插入靶核酸中的条形码可以用作片段化序列的随后比对中的界标,以获得靶核酸的序列表达。在一些实施方案中,包括共同条形码的片段可以被鉴定为具有重叠序列。

[0254] 在一些实施方案中,转座子核酸可以包括彼此连接的两个转座子元件。接头可以包括在插入物中,使得第一转座子元件与第二转座子元件连续。特别有用的插入物是形成“环状”复合物的插入物,如国际公开号W02012/061832;US 2012/0208724,US 2012/0208724,US 2012/0208705和PCT/US2013/031023中列出,其各自通过引用整体并入本文。在这种结构中,具有连续转座子元件的单个插入物与两个转座酶亚单位结合形成“环状”复合体。在一些实施方案中,转座子核酸可以包括封闭基团。

[0255] 基底 (substrate)

[0256] 本文提供的方法和组合物的一些实施方案包括使用具有表面的基底。可用的基质包括例如固体支持物和凝胶。在一些实施方案中,表面结合核酸。在一些实施方案中,表面包含通过Watson-Crick互补性将核酸结合到表面的多个捕获探针。在一些实施方案中,捕获探针结合锚定标签。在一些实施方案中,捕获探针和锚标签各自包含核酸。在一些实施方案中,捕获探针和锚标签包含彼此特异性结合的小分子基团,诸如本文提供的受体或配体,例如生物素,亲合素,HisD,镍,抗体和抗原。

[0257] 基底可以是二维或三维的,并且可以是平面表面(例如,玻璃载玻片)或可以是成形的。有用的材料包括玻璃(例如可控孔径玻璃(CPG)),石英,塑料(如聚苯乙烯(低交联和高交联聚苯乙烯),聚碳酸酯,聚丙烯和聚(甲基丙烯酸甲酯)),丙烯酸共聚物,聚酰胺,硅,金属(例如烷基硫醇衍生化的金(alkanethiolate-derivatized gold)),纤维素,尼龙,胶乳,葡聚糖,凝胶基质(例如,硅胶),聚丙烯醛(polyacrolein)或复合材料。合适的三维固体支持物包括例如球体,微粒,珠,膜,载玻片,板,微机械加工芯片(micro machined chip),管(例如毛细管),微孔,微流体装置,通道,过滤器或适合于锚定核酸或其它捕捉探针的任何其它结构。固体支持物可以包括平面微阵列或能够具有包括核酸群或引物或其它捕捉探针的区域的矩阵。实例包括核苷衍生化的CPG和聚苯乙烯载玻片;衍生化的磁性载玻片;用聚乙二醇接枝的聚苯乙烯等。

[0258] 制备和使用那些组合物的各种组合物和相关方法可用于将捕获探针例如核酸附着,锚定或固定到基底表面。附着可以通过直接或间接键合到表面来实现。键合可以通过共价连接进行(参见例如Joos et al. (1997) *Analytical Biochemistry*, 247:96-101; Oroskar et al. (1996) *Clin. Chem.*, 42:1547-1555; and Khandjian (1986) *Mol. Bio. Rep.*, 11:107-11,其各自通过引用整体并入本文)。优选的附着是核酸的末端核苷酸与整合在表面上的环氧化物的直接胺键合。键合也可以通过非共价键连接。例如,生物素-链霉抗生物

素蛋白 (Taylor et al. (1991) 1. Phys. D: Appl. Phys., 24:1443, 其通过引用整体并入本文) 和地高辛配基与抗地高辛配基 (Smith et al., Science, 253:1122 (1992), 其通过引用整体并入本文) 是用于将核酸锚定于表面的常用工具。核酸与表面的附着可以通过中间结构如珠, 颗粒或凝胶。通过凝胶将核酸附着到阵列以可购自 Illumina Inc. (San Diego, CA) 或记载于 US 2010/10111768; U.S. 2012/0270305; 和 W005/065814 (其各自通过引用整体并入本文) 的流动池例示。

[0259] 在一些实施方案中, 基底可具有连续或整体表面。因此, 核酸片段可以附着在空间随机位置处, 其中最近邻片段 (或从片段衍生的最近邻簇) 之间的距离将是可变的。所得到的阵列可以具有特征的可变或随机空间模式。在一些实施方案中, 本文中列出的方法中使用的基底可包括以重复模式存在的捕捉探针阵列。在一些此类实施方案中, 捕捉探针提供核酸可以附着的位置。在一些实施方案中, 重复模式是六角形模式, 直线模式, 网格模式, 具有反射对称性的模式, 具有旋转对称性的模式等。修饰的核酸附着的捕捉探针各自可具有是, 或小于约 1mm^2 、 $500\mu\text{m}^2$ 、 $100\mu\text{m}^2$ 、 $25\mu\text{m}^2$ 、 $10\mu\text{m}^2$ 、 $5\mu\text{m}^2$ 、 $1\mu\text{m}^2$ 、 500nm^2 、或 100nm^2 , 或由前述值中的任何两个定义的范围的面积。或者/另外, 每个特征可具有是, 或大于约 100nm^2 、 250nm^2 、 500nm^2 、 $1\mu\text{m}^2$ 、 $2.5\mu\text{m}^2$ 、 $5\mu\text{m}^2$ 、 $10\mu\text{m}^2$ 、 $100\mu\text{m}^2$ 、或 $500\mu\text{m}^2$, 或由上述任何两个值定义的面积。由阵列上的片段 (无论是模式化的还是空间随机的) 的扩增产生的核酸的簇或集落可以类似地具有在上述范围中或者选自上面列举的上限和下限之间的面积。

[0260] 在一些实施方案中, 表面上的特征 (例如核酸, 捕获探针或捕获的核酸) 的密度可以为至少 1000 个特征/ mm^2 , 10000 个特征/ mm^2 , 100000 个特征/ mm^2 , 1000000 个特征/ mm^2 , 或在上述值之间的任何范围。在一些实施方案中, 表面上的特征 (例如核酸, 捕获探针或捕获的核酸) 的密度可以是至少 1000 个特征/ μm^2 , 10000 个特征/ μm^2 , 100000 个特征/ μm^2 , 1000,000 个特征/ μm^2 , 2000,000 个特征/ μm^2 , 3000,000 个特征/ μm^2 , 4000,000 个特征/ μm^2 , 5000,000 个特征/ μm^2 , 6000,000 个特征/ μm^2 , 7000,000 个特征/ μm^2 , 8000,000 个特征/ μm^2 , 9000,000 个特征/ μm^2 , 10,000,000 个特征/ μm^2 , 20,000,000 个特征/ μm^2 , 50,000,000 个特征/ μm^2 , 100,000,000 个特征/ μm^2 , 或在上述值之间的任何范围。

[0261] 几种商品化的测序平台利用具有孔的基底, 所述孔在序列检测步骤期间提供了检测试剂 (例如, 可购自 454 Life Sciences (Roche 的子公司, Basel Switzerland) 的平台中的焦磷酸盐或可购自 Ion Torrent (Life Technologies 的子公司, Carlsbad California) 的平台中的质子) 扩散的屏障。

[0262] 本文中提供的一些实施方案包括扩增靶核酸, 经修饰的核酸或其片段的部分。可以使用本领域中已知的任何合适的扩增方法。在一些实施方案中, 在基底之中或之上扩增核酸片段。例如, 在一些实施方案中, 使用桥式扩增方法来扩增核酸片段, 如通过美国专利号 5,641,658; 美国专利公开号 2002/0055100; 美国专利号 7,115,400; 美国专利公开号 2004/0096853; 10 美国专利公开号 No. 2004/0002090; 美国专利公开号 2007/0128624; 和美国专利公开号 2008/0009420 的公开内容例示, 其各自以全文引用的方式并入本文中。

[0263] 桥式扩增方法允许将扩增产物固定在基底之中或之上, 以形成由固定的核酸分子的簇 (或“集落”) 组成的阵列。这种阵列上的每个簇或集落由多个相同的固定的多核苷酸链和多个相同的固定的互补多核苷酸链形成。如此形成的阵列在本文中 can 称为“聚簇阵列”。当由固定的多核苷酸链和固定的互补链的退火对形成时, 固相扩增反应的产物是所谓

的“桥式”结构,其中两条链优选通过共价附着固定在5'端的固体支持物上。桥式扩增方法是其中使用固定的核酸模板产生固定的扩增子的方法的实例。其它合适的方法也可用于从根据本文提供的方法产生的固定化核酸片段产生固定的扩增子。例如,可以通过固相PCR,固相MDA,固相RCA等形成一个或多个簇或集落,无论每对扩增引物中的一个或两个引物是否被固定化。

[0264] 应当理解,本文中所述或本领域通常已知的任何扩增方法可以与通用引物或靶物特异性引物一起使用以扩增固定的DNA片段。用于扩增的合适方法包括但不限于聚合酶链式反应(PCR),链置换扩增(SDA),转录介导的扩增(TMA)和基于核酸序列的扩增(NASBA),例如描述于美国专利号8,003,354,其全部内容通过引用并入本文。上述扩增方法可用于扩增一种或多种目的核酸。例如,PCR,多重PCR,SDA,TMA,NASBA等可以用于扩增固定的核酸片段。在一些实施方案中,特异性针对目标核酸的引物包括在扩增反应中。

[0265] 用于扩增核酸的其它合适方法可包括寡核苷酸延伸和连接,滚环扩增(RCA)(Lizardi et al.,Nat.Genet.19:225-232(1998),其通过引用整体并入本文)和寡核苷酸连接测定(OLA)(参见例如美国专利号7,582,420,5,185,243,5,679,524和5,573,907;EP 0320308;EP 0336731;EP 0439182;WO 90101069;WO 89/12696;和WO 89109835,其通过引用整体并入本文)。应当理解,可以设计这些扩增方法以扩增固定的核酸片段。例如,在一些实施方案中,扩增方法可以包括连接探针扩增或寡核苷酸连接测定(OLA)反应,其包含特异性针对目标核酸的引物。在一些实施方案中,扩增方法可包括引物延伸-连接反应,其含有特异性针对目标核酸的引物。作为可以特异性设计以扩增目标核酸的引物延伸和连接引物的非限制性实例,扩增可包括用于GoldenGate测定(Illumina,Inc.,San Diego,CA)或一种或多种列于美国专利号7,582,420和7,611,869的测定法的引物,如通过美国专利号7,582,420和7,611,869例示,其各自通过引用整体并入本文。

[0266] 等温扩增技术可以用于本公开的方法中。示例性的等温扩增方法包括但不限于多重置换扩增(MDA),例如由Dean et al.,Proc.Natl.Acad.Sci.USA99:5261-66(2002)例示,或等温链置换核酸扩增,如以美国专利号6,214,587例示,其各自通过引用整体并入本文。可用于本公开的其它非基于PCR的方法包括例如在Walker et al.,Molecular Methods for Virus Detection,Academic Press,Inc.,1995;美国专利号5,455,166和5,130,238,及Walker et al.,Nucl.Acids Res.20:1691-96(1992)中描述的链置换扩增(SDA)或超分支链置换扩增(hyperbranched strand displacement amplification),其描述于例如Lage et al.,Genome Research 13:294-307(2003)中,其各自通过引用整体并入本文。

[0267] 扩增反应,条件和组分的另外描述在美国专利号7,670,810中阐述,其通过引用整体并入本文。其它有用的等温扩增技术包括重组酶促进的扩增技术,如由TwistDx(Cambridge,UK)以TwistAmp™试剂盒商业销售的那些。重组酶促进的扩增试剂和反应条件的有用组分阐述于US 5,223,414和US7,399,590中,其各自通过引用整体并入本文。也可以使用依赖于解旋酶的扩增,例如,如Xu et al.EMBO Rep 5:795-800(2004),其通过引用以其整体并入本文。

[0268] 在一些实施方案中,可能期望需要进行再接种(re-seeding)步骤。例如,修饰的核酸片段可以在表面区域内的位置捕获,在扩增过程的一个或多个循环上复制,其原始片段

和/或其扩增子可以从该位置释放,释放的核酸可以在相同区域中的其它位置捕获,并且可以扩增新捕获的核酸。在一个具体的实例中,可以对接种在表面上的片段进行单一循环的桥式扩增,而不是在从表面释放时洗掉原始模板片段,模板片段可以在新位置在表面上再接种,所述新位置邻近于其最初接种的位置。后续轮次的桥式扩增将允许在原始种子位置和再接种位置两者的簇生长。使用此类方法,可以在表面的区域创建重复的集落以提供技术重复。对技术重复的序列的分析可以提供错误检查的益处。例如,仅在近端簇(被鉴定为技术重复)的子集中出现的观察到的序列变体可以被鉴定为扩增错误,而发生在被鉴定为特定片段的技术重复的所有簇中的序列变体更多可能是真正的变体。

[0269] 测序核酸

[0270] 本文中所述的方法的一些实施方案可以包括对源自靶核酸的片段进行测序的步骤。一个实例是合成测序(SBS)。在SBS中,监测核酸引物沿着核酸模板(例如,靶核酸或其扩增子的片段)延伸以确定模板中的核苷酸序列。引物可以与存在于如上所述的插入物中的引发位点杂交。基础化学过程可以是聚合(例如,通过聚合酶催化)。在特定的基于聚合酶的SBS实施方案中,将荧光标记的核苷酸以模板依赖性方式添加至引物中(从而延伸引物),使得添加至引物的核苷酸的顺序和类型的检测可用于确定模板。使用本文阐述的步骤,附着在阵列的不同位置处的多个不同核酸片段可以在由于其在阵列中的位置而能够区分针对不同模板发生的事件的条件下经受SBS技术。

[0271] 在一些实施方案中,流动池提供用于容纳通过本公开的方法产生并且经受SBS或其它检测技术的核酸片段阵列的方便的形式,所述检测技术牵涉循环中的试剂的重复递送。如本文所用,“流动池”包括具有表面的室,一种或多种流体试剂可以流过该表面。通常,流动池将具有入口开口和出口开口以促进流体的流动。可以容易地用于本公开的方法中的流动池和相关流体系统和检测平台的实例描述于,例如,Bentley等,Nature 456:53-59 (2008),WO 04/018497;US 7,057,026;WO 91/06678;WO 071123744;US 7,329,492;US 7,211,414;US 7,315,019;US 7,405,281,和US 2008/0108082中,其通过引用以其整体并入本文。在具体的实施方案中,凝胶存在于流动池的内表面上,并且凝胶提供基底,对所述基底附着本文中列出的一种或多种组合物和/或其中发生本文中列出的一种或多种方法步骤。

[0272] 在一些实施方案中,为了启动第一SBS循环,可以使一种或多种标记的核苷酸,DNA聚合酶等流入/通过容纳核酸片段阵列的流动池。可以检测阵列的那些位点,其中引物延伸(例如通过引物与位于附着于核酸片段的插入物上的引发位点的杂交)引起经标记的核苷酸被掺入。任选地,核苷酸可以进一步包括一旦核苷酸已经添加到引物就终止进一步引物延伸的可逆终止性质。例如,具有可逆终止剂部分的核苷酸类似物可以添加到引物中,使得随后的延伸不能发生,直到递送去封闭剂以去除该部分。因此,对于使用可逆终止的实施方案,可以将去封闭试剂递送到流动池(在检测发生之前或之后)。可以在各个递送步骤之间进行洗涤。然后可以重复该循环“n”次以将引物延伸n个核苷酸,从而检测长度“n”的序列。可以容易地适用于与由本公开的方法产生的阵列一起使用的示例性SBS程序,流体系统和检测平台描述于例如Bentley et al.,Nature 456:53-59 (2008),WO 04/018497;US 7,057,026;WO 91/06678;WO 071123744;US 7,329,492;US 7,211,414;US 7,315,019;US 7,405,281,及US 2008/0108082,其各自通过引用整体并入本文。

[0273] 在一些实施方案中,可以使用利用循环反应的其它测序程序,例如焦磷酸测序。由于特定核苷酸被掺入新生核酸链中,焦磷酸测序检测无机焦磷酸 (PPi) 的释放 (Ronaghi, et al., *Analytical Biochemistry* 242 (1), 84-9 (1996); Ronaghi, *Genome Res.* 11 (1), 3-11 (2001); Ronaghi et al. *Science* 281 (5375), 363 (1998); US 6,210,891; US 6,258,568 和 US 6,274,320, 其各自通过引用整体并入本文)。在焦磷酸测序中,释放的 PPi 可以通过由 ATP 硫酸化酶转化为三磷酸腺苷 (ATP) 来检测,并且可以通过萤光素酶产生的光子检测产生的 ATP 水平。因此,可以通过发光检测系统监测测序反应。用于基于荧光的检测系统的激发辐射源对于焦磷酸测序程序不是必需的。可用于将焦磷酸测序应用于本公开的方法的有用的流体系统,检测器和程序描述于例如 WO 2012058096, US 2005/0191698, US 7,595,883, 和 US 7,244,559, 其各自通过引用整体并入本文。连接测序反应也是有用的,包括例如 Shendure et al. *Science* 309:1728-1732 (2005); US 5,599,675; 和 US 5,750,341 中描述, 其各自通过引用整体并入本文。一些实施方案可以包括杂交测序程序,如例如描述于 Bains et al., *Journal of Theoretical Biology* 135 (3), 303-7 (1988); Drmanac et al., *Nature Biotechnology* 16, 54-58 (1998); Fodor et al., *Science* 251 (4995), 767-773 (1995); 及 WO 1989110977, 其各自通过引用整体并入本文。

[0274] 在一些实施方案中,连接测序和杂交测序程序,存在于阵列位点处的靶核酸片段 (或其扩增子) 经历寡核苷酸递送和检测的重复循环。用于本文或本文引用的参考文献中所述的 SBS 方法的流体系统可容易地适用于递送用于连接测序或杂交测序程序的试剂。通常,寡核苷酸是经荧光标记的,并且可以使用与关于本文中或本文引用的参考文献中的 SBS 程序所述的荧光检测器类似的荧光检测器来检测。

[0275] 一些实施方案可以利用涉及 DNA 聚合酶活性的实时监测的方法。例如,可以通过带有荧光团的聚合酶和 γ -磷酸标记的核苷酸之间的荧光共振能量转移 (FRET) 相互作用或者用 Zeromode 波导 (ZMW) 来检测核苷酸掺入。基于 FRET 的测序的技术和试剂描述于例如 Levene et al. *Science* 299, 682-686 (2003); Lundquist et al. *Opt. Lett.* 33, 1026-1028 (2008); 和 Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176-1181 (2008), 其公开内容通过引用整体并入本文。

[0276] 一些 SBS 实施方案包括检测在将核苷酸掺入延伸产物中时释放的质子。例如,基于检测释放的质子的测序可以使用可购自 Ion Torrent (Guilford, CT, Life Technologies 子公司) 的电检测器和相关技术或 US 2009/10026082AI; US 2009/10127589AI; US 2010/10137143; 或 US 2010/10282617 (其各自以全文引用的方式并入本文中) 中描述的测序方法和系统。

[0277] 在一些实施方案中,本方法的测序步骤可包括纳米孔测序技术,例如 Deamer & Akeson *Trends Biotechnol.* 18, 147-151 (2000); Deamer & Branton, *Acc. Chem. Res.* 35: 817-825 (2002); 和 Li et al., *Nat. Mater.* 2: 611-615 (2003) 中描述的, 其各自通过引用整体并入本文。在此类实施方案中,靶核酸片段通过纳米孔。纳米孔可以是合成孔或生物膜蛋白,例如 α -溶血素。当靶核酸通过纳米孔时,可以通过测量孔的电导率的波动来鉴定每个碱基。(美国专利号 7,001,792; Soni & Meller *Clin. Chem.* 53, 1996-2001 (2007); Healy, *Nanomed.* 2: 459-481 (2007); 及 Cockroft et al., *J. Am. Chem. Soc.* 130: 818-820 (2008), 其各自通过引用整体并入本文)。在一些实施方案中,单个纳米孔的位置类似于本文示例的阵

列上的位点或特征。纳米孔彼此的接近性可以与它们读取的片段序列的接近性相关,例如,以便于将那些片段装配成它们来源的较大序列。

[0278] 在一些实施方案中,本文中所述的测序步骤可以有利地以多重形式进行,使得同时操作多种不同的靶核酸。在具体实施方案中,可以在共同的反应容器中或在特定基底的表面上处理不同的靶核酸。这允许以多重方式方便地递送测序试剂,除去未反应的试剂和检测掺入事件。在使用表面结合的靶核酸或其片段的实施方案中,靶核酸或片段可以是阵列形式。在阵列形式中,靶核酸的片段通常可以以空间可区分的方式偶联到表面,例如使用本文所述的附着技术。阵列可以包括在每个位点(也称为特征)处的靶核酸片段的单拷贝,或者具有相同序列的多个拷贝可以存在于每个位点或特征。可以通过扩增方法如桥式扩增或乳液PCR产生多个拷贝。

[0279] 制备和测序核酸

[0280] 本文提供的组合物和方法的一些实施方案包括从靶核酸制备测序文库。一些实施方案还包括对制备的文库进行测序。在一些实施方案中,转座体群体,每个包含转座酶和转座子核酸的转座体与靶核酸接触。接触可以在基底中或其上,或者在溶液中进行。转座体可以包含单侧转座酶活性,使得靶核酸在多个位点产生切口,并且将单个转座子核酸在有切口的位点的一侧连接至有切口的链。在一些实施方案中,引物可以与每个附着的转座子核酸杂交并延伸以获得单链修饰的核酸群体。在一些实施方案中,可以扩增延伸的核酸。在一些实施方案中,可以将延伸和/或扩增的核酸,即修饰的核酸,捕获到表面用于测序。一些实施方案还包括对捕获的核酸测序。

[0281] 图1描绘了示例性实施方案,其中使靶核酸与包含转座子核酸的转座体群体接触。靶核酸在多个位点产生切口,并且将转座子核酸在切口位点的一侧附着到有切口的靶核酸的一条链上。引物与附着的转座子核酸杂交以提供延伸核酸群体。在一些实施方案中,可以扩增延伸的核酸。在一些实施方案中,延伸的核酸提供用于测序文库的模板。

[0282] 在一些实施方案中,具有单侧转座酶活性的转座体包含具有单侧转座酶活性的转座酶。在一些实施方案中,转座体包含封闭的转座子核酸。本文提供了可用于从靶核酸制备和测序文库的方法和组合物的转座体。在一些实施方案中,转座子核酸包含锚定位点,条形码,测序引物位点,扩增引物位点和/或报道标签。图2描绘了示例性实施方案,其中包含不同的转座子核酸的转座体的群体与靶核酸接触并且不同的转座子核酸在不同切口位点与靶核酸的链附着。在一些实施方案中,不同的转座子核酸可以包括不同的锚定位点,条形码,测序引物位点,扩增引物位点和/或报道标签。

[0283] 在一些实施方案中,将扩增的核酸扩增。在一些实施方案中,扩增是使用有尾扩增引物。有尾引物可以包括另外的末端序列,使得额外的序列包括在扩增产物中。在一些实施方案中,扩增引物可以包括锚定位点,测序引物位点,扩增引物位点和报道标签。图3描绘了其中通过线性扩增来扩增经修饰的靶核酸以获得某些扩增产物的示例性实施方案。

[0284] 图4描绘了示例性实施方案,其中转座体包含二聚体转座酶,并且转座子核酸包含两个包含嵌合元件(ME)的转座子元件,其中ME之一在3'末端用双脱氧基封闭。在一些实施方案中,转座子核酸包含两个转座子元件之间的可切割接头。可以切割转座子核酸,并且转座子核酸的非封闭片段可以在切口位点处附着到有切口的靶核酸的链上。

[0285] 在一些实施方案中,在表面上捕获修饰的核酸。在一些实施方案中,表面包含多个

捕获探针。在一些实施方案中,捕获探针包括核酸。在一些实施方案中,捕获探针与经修饰的核酸特异性杂交。在一些实施方案中,捕获探针包含结合经修饰的核酸的亲的部分和亲的部分。在一些实施方案中,例如通过桥式扩增来扩增捕获的核酸。在一些实施方案中,在表面上测序捕获核酸。

[0286] 本文提供的方法和组合物的一些实施方案还包括制备包含条形码的测序文库。一些实施方案还包括测序此类文库。在一些实施方案中,条形码提供适用于靶核酸的测序片段的比对的界标。在一些实施方案中,通过单侧转座和连接将转座子核酸插入靶核酸的单链中。扩增经修饰的靶核酸并测序片段。重叠片段可包括常规插入,其可用于测序片段的比对以产生靶核酸的序列表示。图5描绘了示例性实施方案,其中包含不同条形码的转座体群体接触靶核酸;转座子核酸附着于切口位点的一侧,而转座子核酸的另一非附着末端通过连接附着于切口位点的另一侧;并通过全基因组扩增(WGA)扩增经修饰的靶核酸。

[0287] 在一些实施方案中,使具有单侧转座酶活性的转座体群体与靶核酸接触。转座体包含插入靶核酸链中的转座子核酸。本文描述了可用于此类实施方案的转座体。在一些实施方案中,通过使靶核酸与转座体接触将转座子核酸插入双链靶核酸的单链中,使得靶核酸在多个位点产生切口,并且单个转座子核酸附着到在切口位点一侧的有切口的链,并将附着的单个转座子核酸连接到切口位点另一侧的有切口的链上。在一些实施方案中,连接酶可以包括非同源末端连接连接酶。在一些实施方案中,连接酶可包括连接酶IV。在一些实施方案中,扩增经修饰的核酸。在一些实施方案中,在表面上捕获修饰的核酸。在一些实施方案中,对经修饰的核酸测序。在一些实施方案中,根据重叠序列中常见条形码的存在,对修饰的核酸的序列进行比对。一些实施方案包括测序文库,其包含通过本文提供的方法制备的条形码。

[0288] 获得单倍型信息

[0289] 靶核酸,如基因组DNA可以包括多于单个单倍型。例如,人基因组DNA含有两组DNA分子,每组有母本和父本序列的不同组合。本文中提供的一些实施方案可用于从单个核酸分子的片段或其拷贝获得序列信息。

[0290] 在一些实施方案中,维持基底上某些片段的物理接近性。在一些实施方案中,与在线性靶核酸的序列中彼此较不接近的片段的序列相比,在线性靶核酸的序列中彼此更接近的片段的序列在表面上彼此具有更接近的物理接近性。可以通过多种方法保留某些片段的物理接近性。

[0291] 在一些实施方案中,单侧转座不使靶核酸片段化。在一些实施方案中,可以使靶核酸与具有单侧转座酶活性的转座体接触,以获得修饰的核酸。在一些实施方案中,修饰的核酸可以与表面接触。在一些实施方案中,转座子核酸包括锚定标签,使得经修饰的序列可捕获在包含捕获探针的表面上。在一些实施方案中,修饰的核酸可以在与表面接触时被片段化。在一些实施方案中,修饰的核酸可以在接近表面的位置处片段化。在一些实施方案中,修饰的核酸可以在表面上测序。

[0292] 在一些实施方案中,获得单倍型信息的方法包括比较针对表面上的近端位置确定的互补序列以鉴定序列错误。在一些实施方案中,表面上任何两个片段种类的相对接近性可以提供对从两个片段获得的序列信息的比对有用的信息。具体地,在表面上来源于任意两个给定片段的簇之间的距离可以与两个簇来自相同的靶多核苷酸分子的概率正相关,如

WO 2012/025250中更加详细描述,其各自以全文引用的方式并入本文中。

[0293] 作为实例,在一些实施方案中,衍生自在流动池表面捕获的长核酸分子的片段出现在穿过流动池表面的线中(例如,如果核酸在片段化或扩增之前被延伸出来)或在表面上的云(cloud)中。此外,然后可以产生固定的核酸的物理图谱。因此,物理图关联扩增固定的核酸后的簇的物理关系。具体地,物理图谱用于计算从任何两个簇获得的序列数据连接的概率,如在WO2012/025250的并入材料中描述,其各自以全文引用的方式并入本文中。

[0294] 在一些实施方案中,通过使表面成像以建立固定的核酸分子在整个表面上的位置来产生物理图谱。在一些实施方案中,通过将成像剂添加固体支持物并检测来自成像剂的信号来成像固定的核酸。在一些实施方案中,成像剂是可检测标记物。合适的可检测标记物包括但不限于质子,半抗原,放射性核素,酶,荧光标记物,化学发光标记物和/或生色剂。例如,在一些实施方案中,成像剂是嵌入染料或非插入DNA结合剂。可以使用如本领域中已知的任何合适的嵌入染料或非插入DNA结合剂,包括但不限于在US2012/0282617中列出的,其通过引用整体并入本文。

[0295] 在某些实施方案中,多个经修饰的核酸分子流到包含多个纳米通道的流动池上。如本文中所用,术语纳米通道是指可以长线性核酸分子被延伸通过的窄通道。在一些实施方案中,链的数目是,或不超过1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900或1000个单独的长核酸链,或由任何两个前述值限定的范围,其被延伸穿过每个纳米通道。在一些实施方案中,单独的纳米通道被物理屏障隔开,所述物理屏障防止靶核酸的单个长链与多个纳米通道相互作用。在一些实施方案中,固体支持物包含,或至少包含,10、50、100、200、500、1000、3000、5000、10000、30000、50000、80000或100000个纳米通道,或由任何两个的上述值限定的范围。

[0296] 在一些实施方案中,一旦核酸已经沿着通道延伸,则切割修饰的核酸。可以任选地扩增所得片段以沿着通道的表面形成簇。然后,可以例如通过这些通道之一的长度向下追踪簇实施邻近定位(contiguity mapping)。作为实例,具有1000个或更多个纳米通道的流动池可用于对具有短“定位”读出的生物体的基因组进行测序,所述纳米通道在纳米通道中具有定位的固定化片段化产物。在一些实施方案中,纳米通道中定位的固定化片段化产物可以用于解析单倍型。在一些实施方案中,纳米通道中定位的固定化片段化产物可用于解决定相问题。

[0297] 在一些实施方案中,使用单侧转座将人工DNA插入gDNA中。一个实例中,将人工DNA插入基因组DNA(或其它核酸)的重复区域中以使重复区域是独特的。可以例如通过测序技术(例如上文所述的那些)来分析重复区域,以计数重复的数目,或以定向相对于重复区域基因组DNA中的其它序列。在另一个实例中,通过单侧转座插入的人工DNA使得双链核酸的顶链和底链不同。因此,插入方法的产物可以例如通过测序技术(例如上述那些)来分析,以将一条链与另一条链区分开。这可以进一步允许在基因组DNA(或其它双链核酸)的重建中顶部和底部链的独立装配。

实施例

[0298] 实施例1:封闭的转座子核酸

[0299] 不用转座体 (Amplicon), 或包含转座酶和 (1) 用3'生物素基团 (3'Bio) 封闭的转座子核酸; (2) 用3'间隔物基团 (3'间隔物) 封闭的转座子核酸; 或 (3) 非封闭的转座子核酸 (TDE1) 的转座体处理靶DNA。图6描绘了结果, 其中用包含封闭的转座子核酸 (3'Bio, 和3'间隔物) 的转座体不发生转座。

[0300] 实施例2: 界标 (landmark) 插入和装配模型

[0301] 将包含12bp随机序列的界标插入靶DNA中。对DNA进行测序, 并且序列片段从头装配。图7显示了名义倍数覆盖率和具有100bp插入频率的500bp读段和具有50bp插入频率的300bp读段的平均合成读段长度的图。证明了可以用50X覆盖从头装配6-7kb。

[0302] 实施例3: 在具有和不具有甘油的情况下的单侧转座

[0303] 将靶DNA (pUC19) 与转座体温育, 并在1.2%凝胶上分离DNA产物。运行10个样品: 前5个样品不含甘油, 而后5个样品包含65%的甘油。每组5个样品建立为转座体的不同浓度的滴定。转座体由转座酶和非封闭的转座子核酸 (TDE1) 组成。图9中显示了染色凝胶的照片。凝胶还加载有未切割的pUC19 (仅pUC19), 线性化的pUC19 (EcoRI) 和单链切口的pUC19 (Nb.Bsr DI) 的对照。如对于加载有“无甘油”样品的凝胶泳道所示, TDE浓度的增加导致单侧转座 (即有切口的) 产物和双侧转座 (即线性) 产物的增加。相比之下, 在65%甘油存在下进行的反应显示当TDE1增加时单侧转座产物的量增加, 但是在TDE1浓度增加的情况下存在很少到没有 (little to no) 双侧转座产物增加。

[0304] 实施例4: 转移链的转座子核酸的长度的改变抑制转座

[0305] 该实施例证明通过减去1个核苷酸 (n-1) 或添加1个核苷酸 (n+1) 改变转座子的转移链的长度降低了转座效率。

[0306] 与3'n-1和n+1METS转座子形成转座体, 并在室温下与0, 1%, 5%, 50%, 90%, 99%, 或100%TDE1杂交过夜。然后, 使所得转座体在室温下与1kb扩增子反应过夜, 接着用SDS处理, 然后在TBE凝胶上分离。图10显示了加载有反应产物的TBE凝胶以及分子量梯和不具有转座酶的对照样品。令人惊讶的是, 即使在过夜温育下, 大多数靶DNA仍然存在于每个样品中, 表明n-1和n+1转座子对转座具有抑制效应。此外, 抑制效应与增加的n-1和n+1转座子的百分比相关。

[0307] 如本文使用的术语“包含”与“包括”, “含有”或“特征在于”同义, 并且是包括性的或开放式的, 并且不排除额外的未记载的要素或方法步骤。

[0308] 上述描述公开了本发明的几种方法和材料。本发明易于对方法和材料的修改, 以及制作方法和设备的改变。通过考虑本公开或本文公开的本发明的实践, 此类修改对于本领域技术人员将变得显而易见。因此, 不希望本发明限于本文公开的具体实施方案, 而是它涵盖了落入本发明的真实范围和精神内的所有修改和替代。

[0309] 本文引用的所有参考文献, 包括但不限于公开的和未公开的申请, 专利和文献参考文献通过引用以其整体并入本文, 并且由此成为本说明书的一部分。在通过引用并入的出版物和专利或专利申请与本说明书中包含的公开内容相矛盾的情况下, 本说明书旨在代替和/或优先于任何此类矛盾材料。

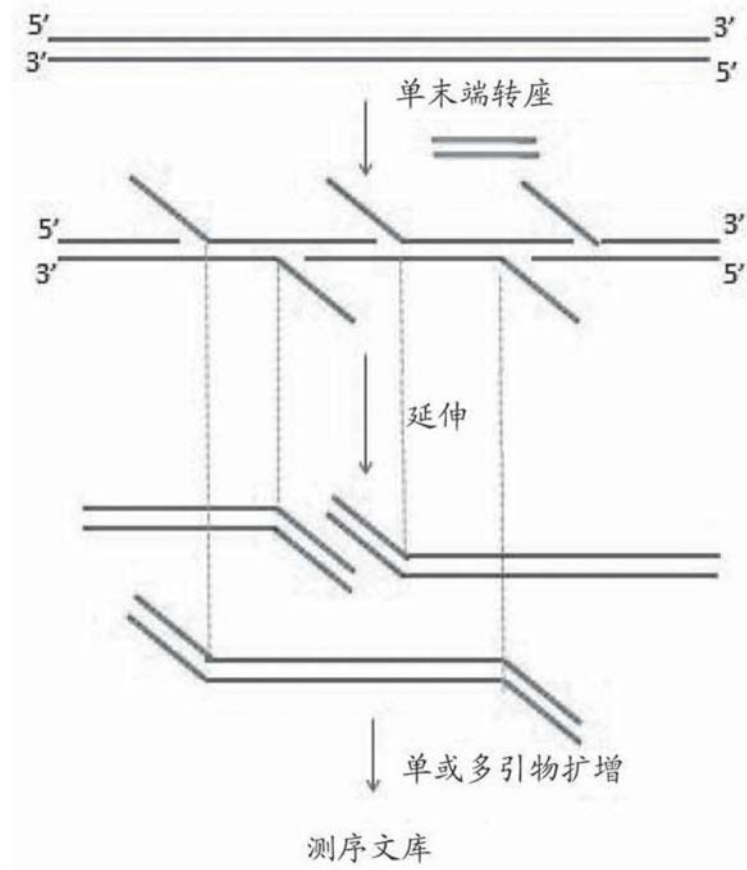


图1

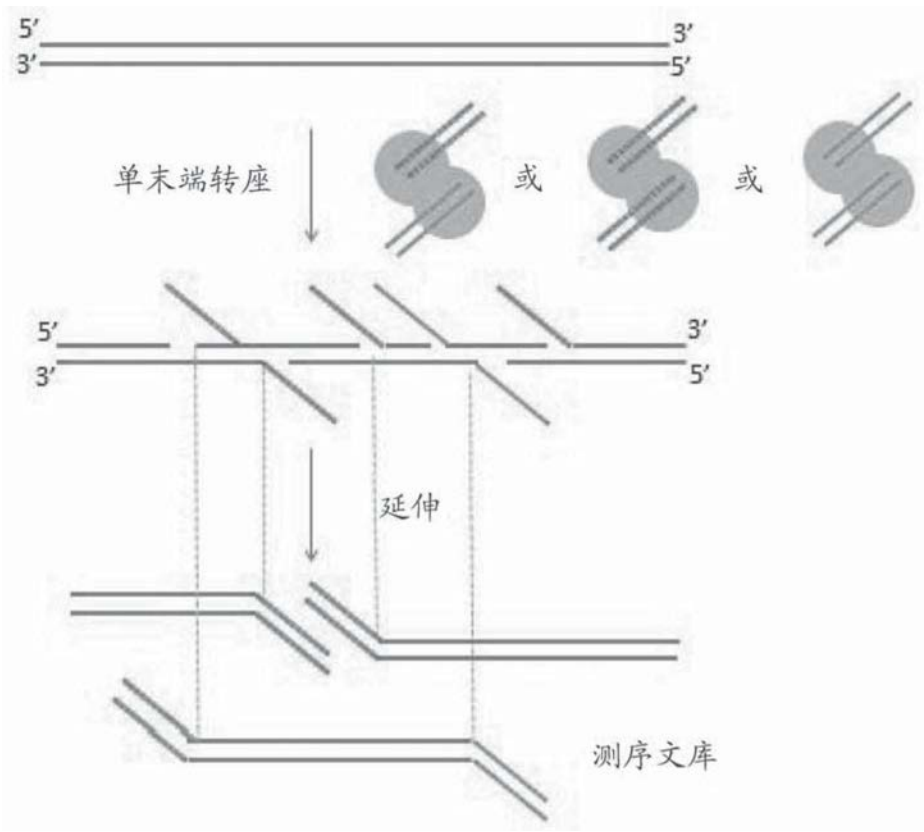


图2

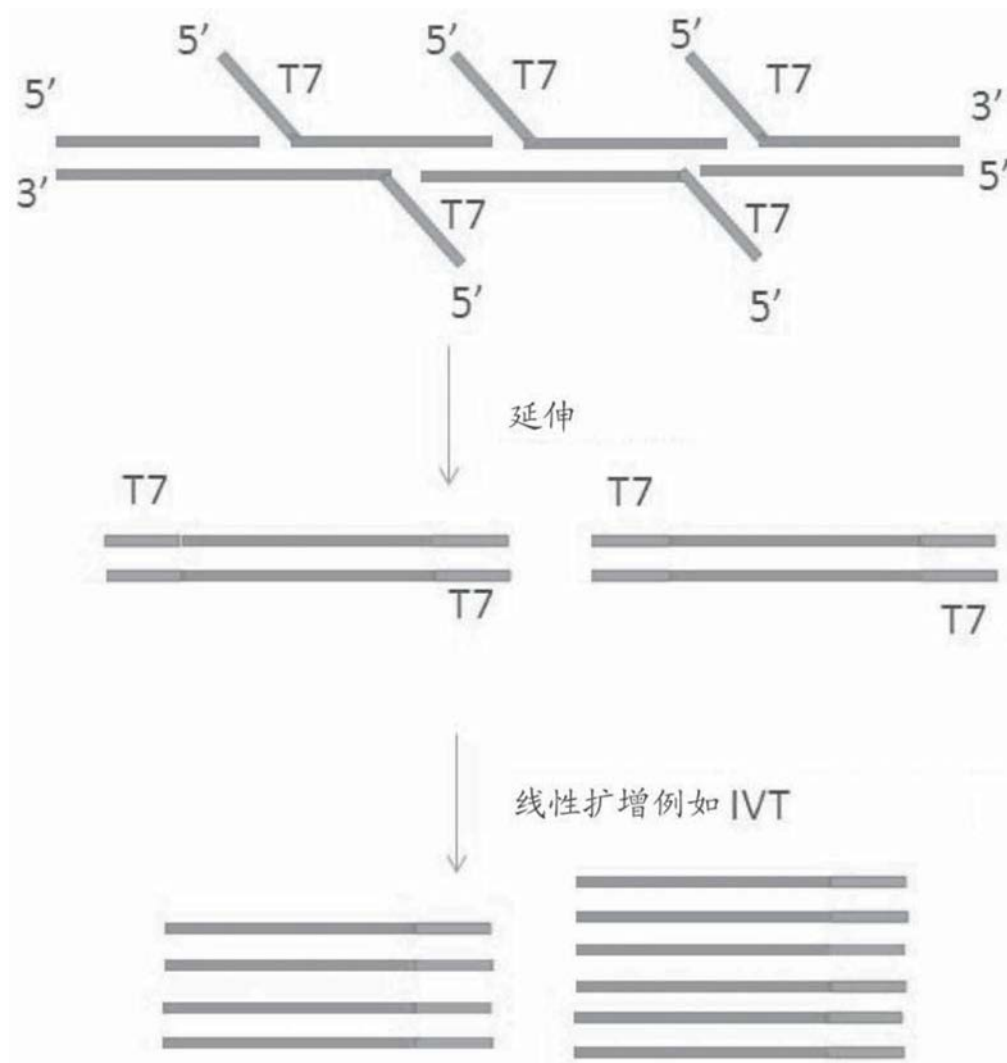


图3

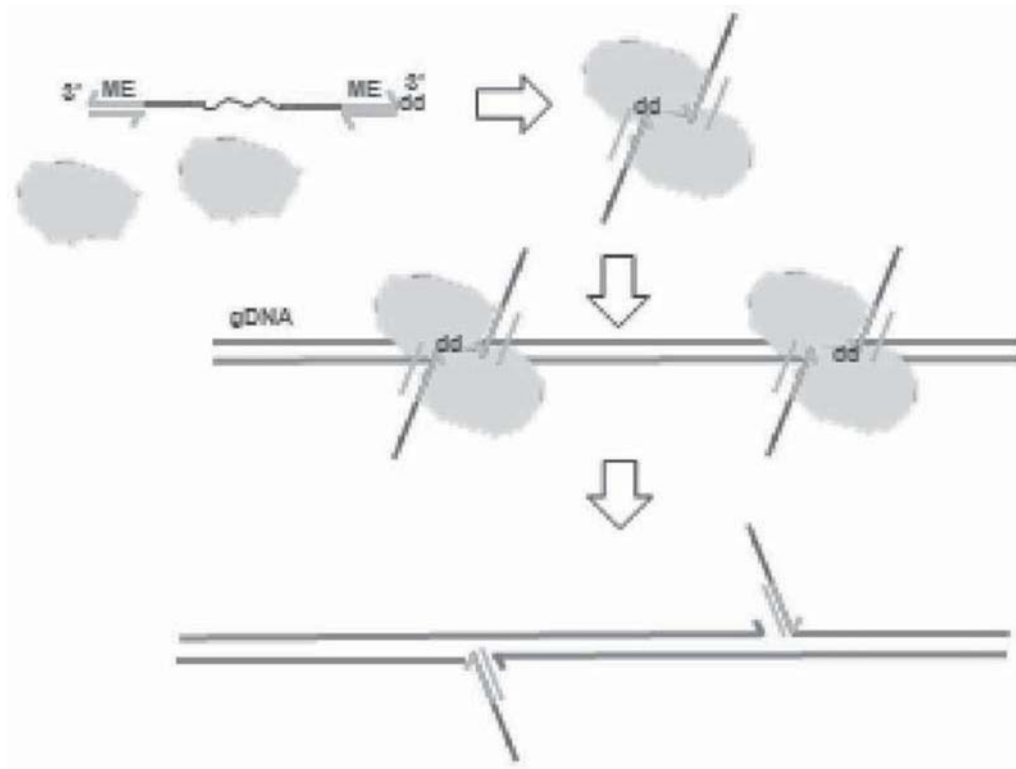


图4

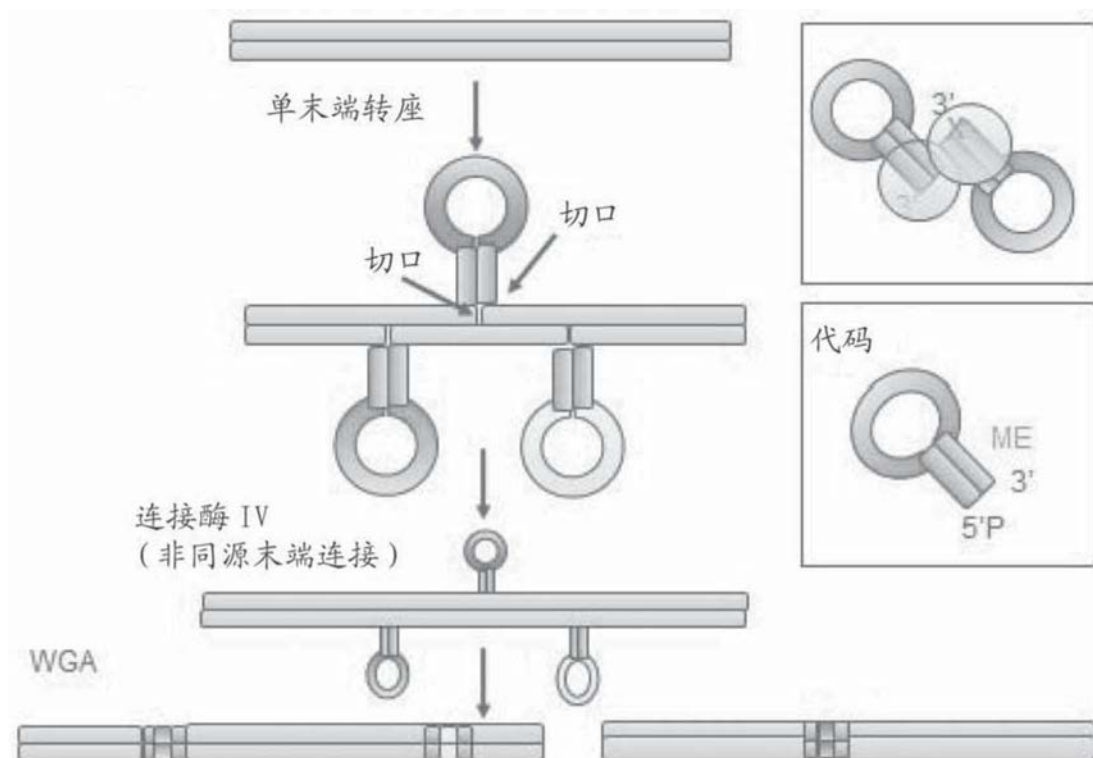


图5

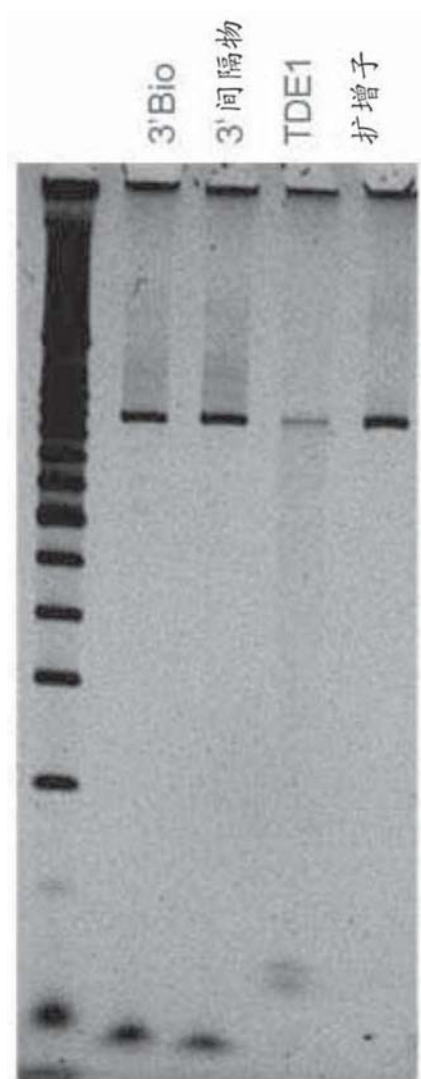


图6

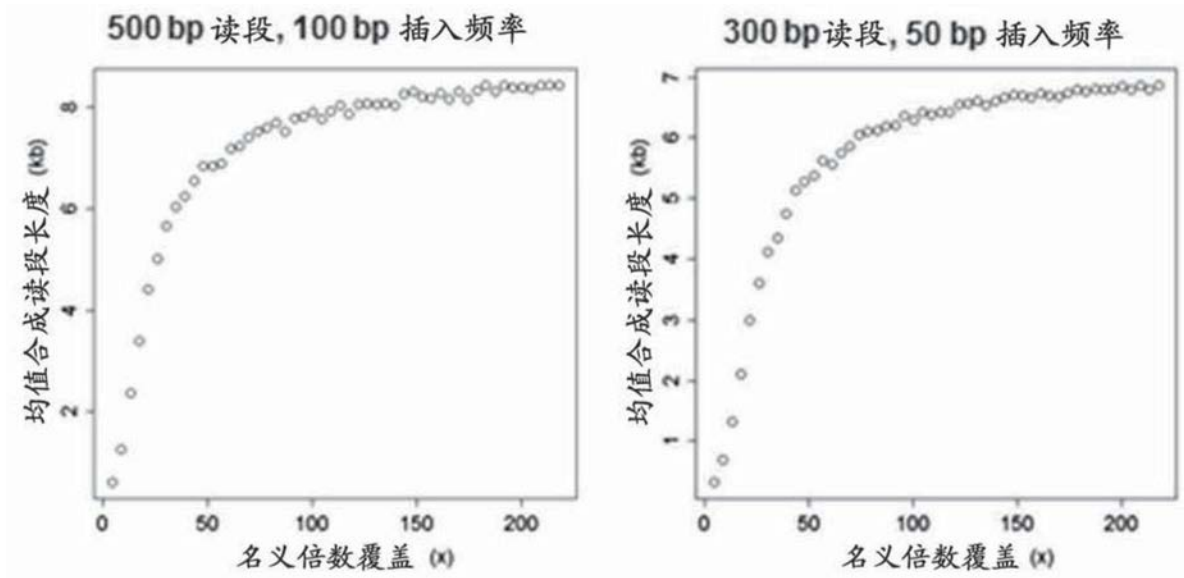


图7

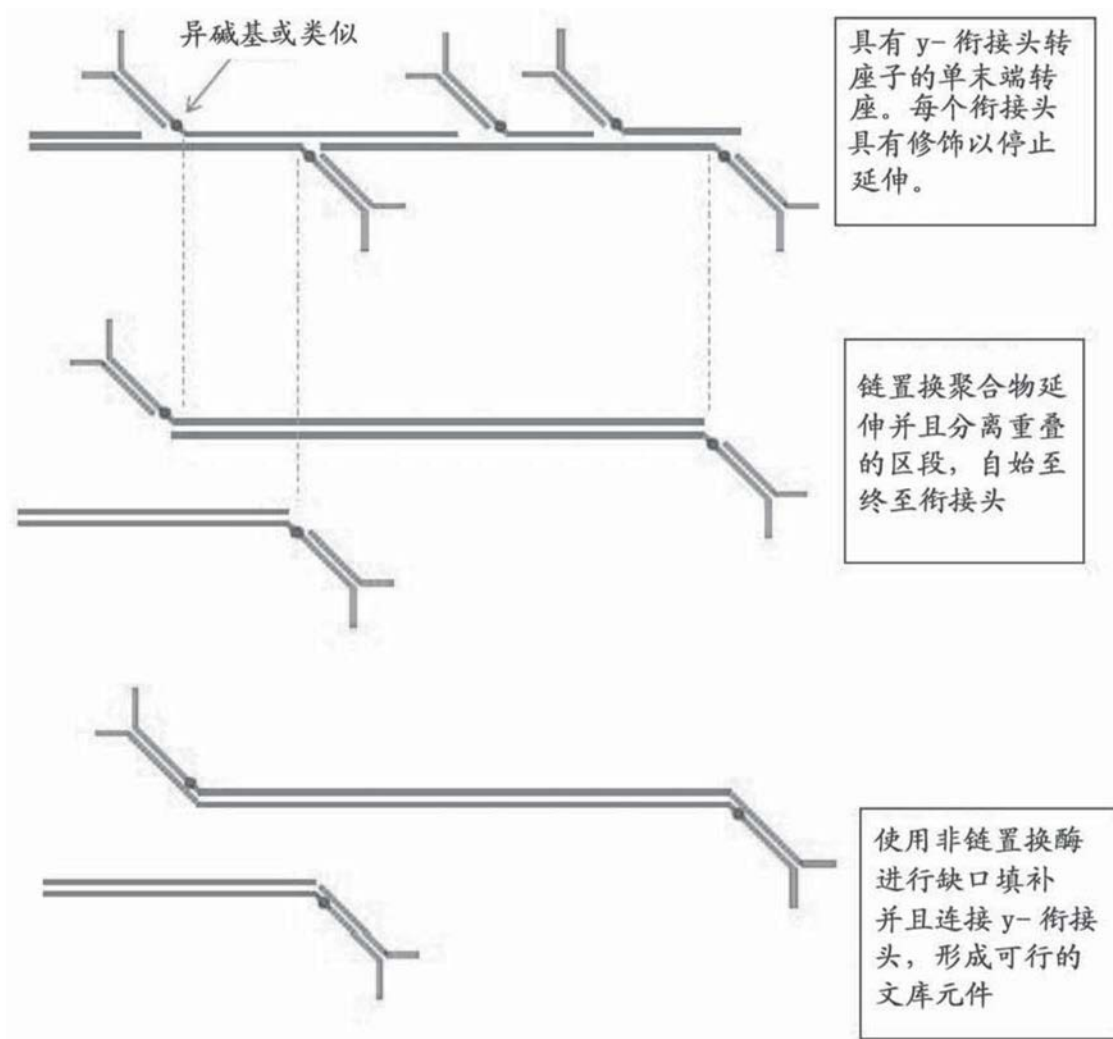


图8A

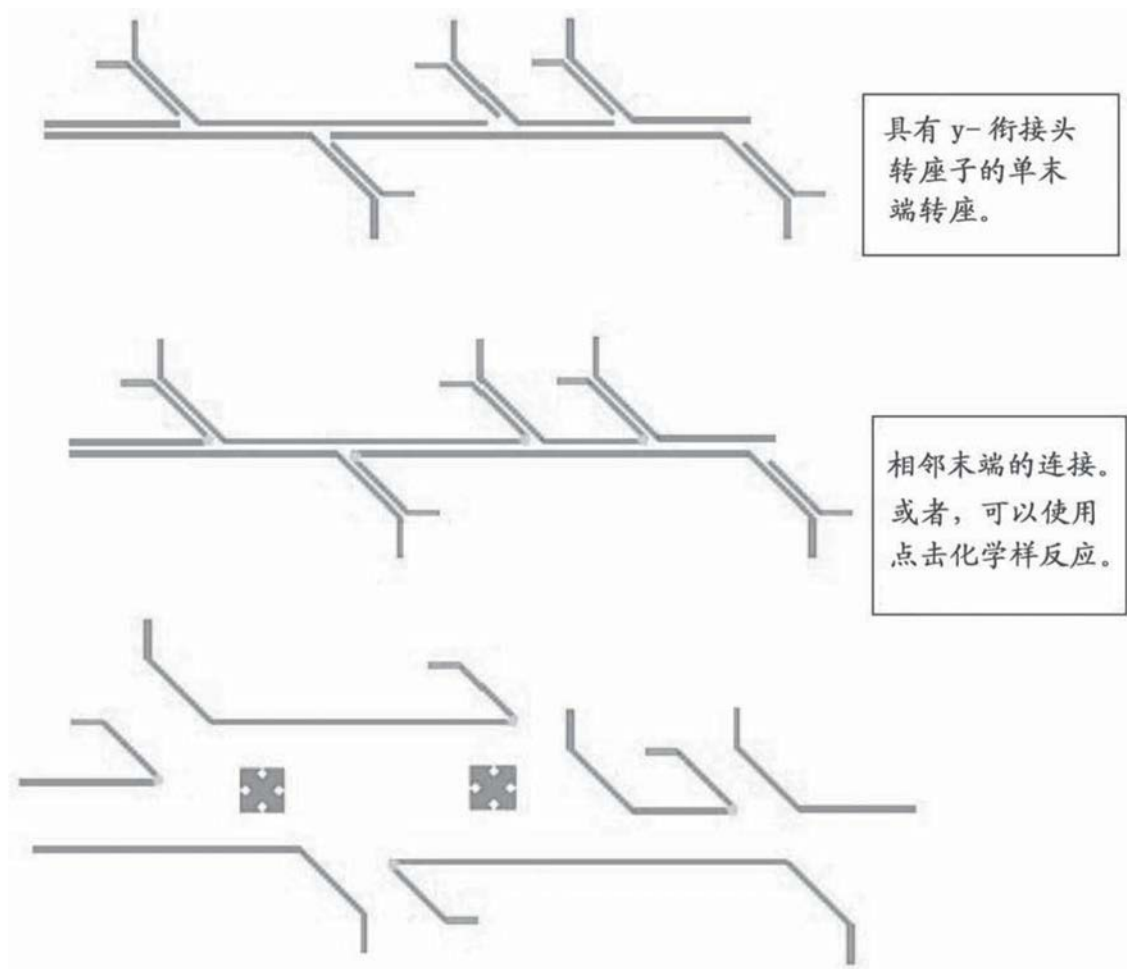


图8B

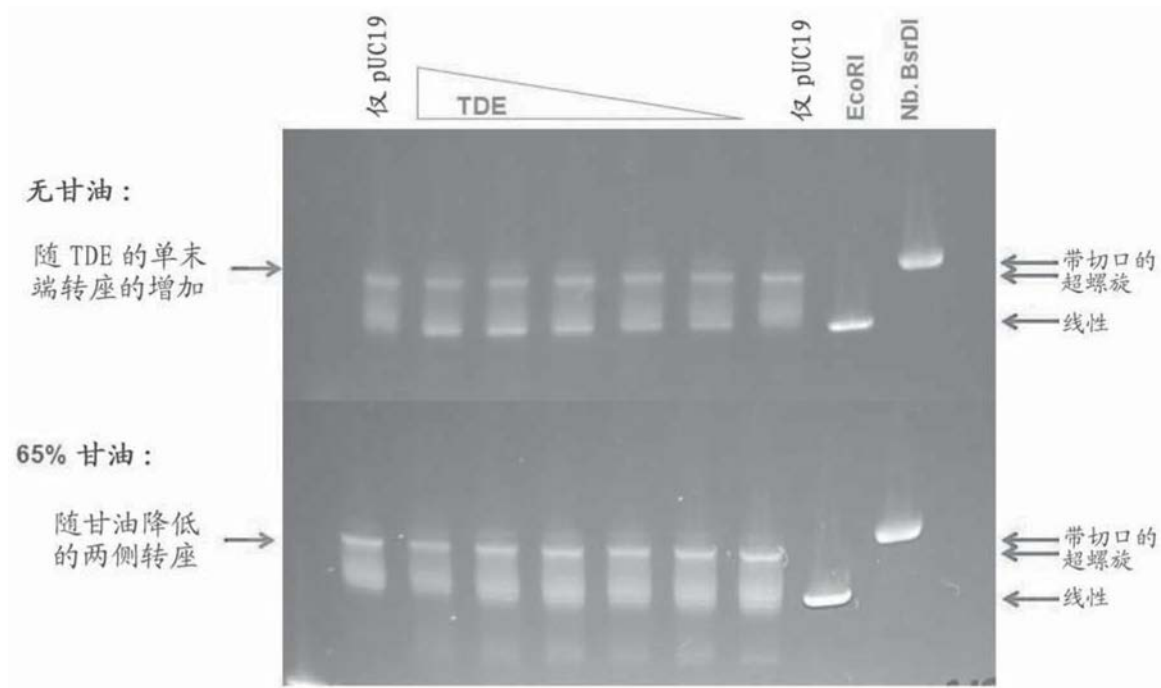


图9

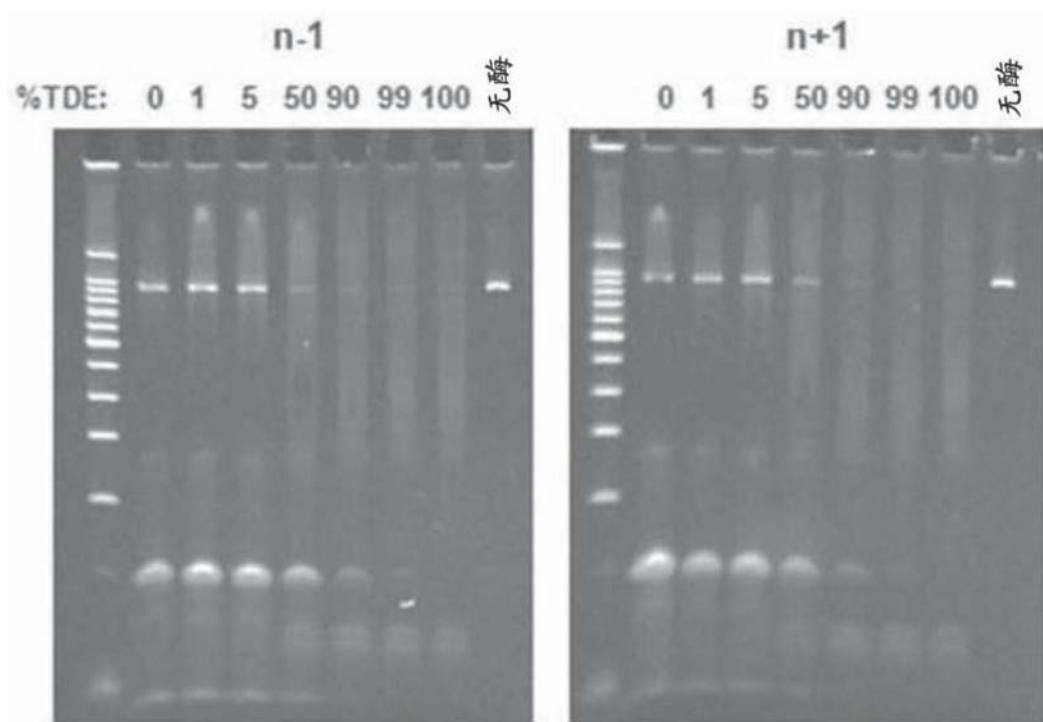


图10

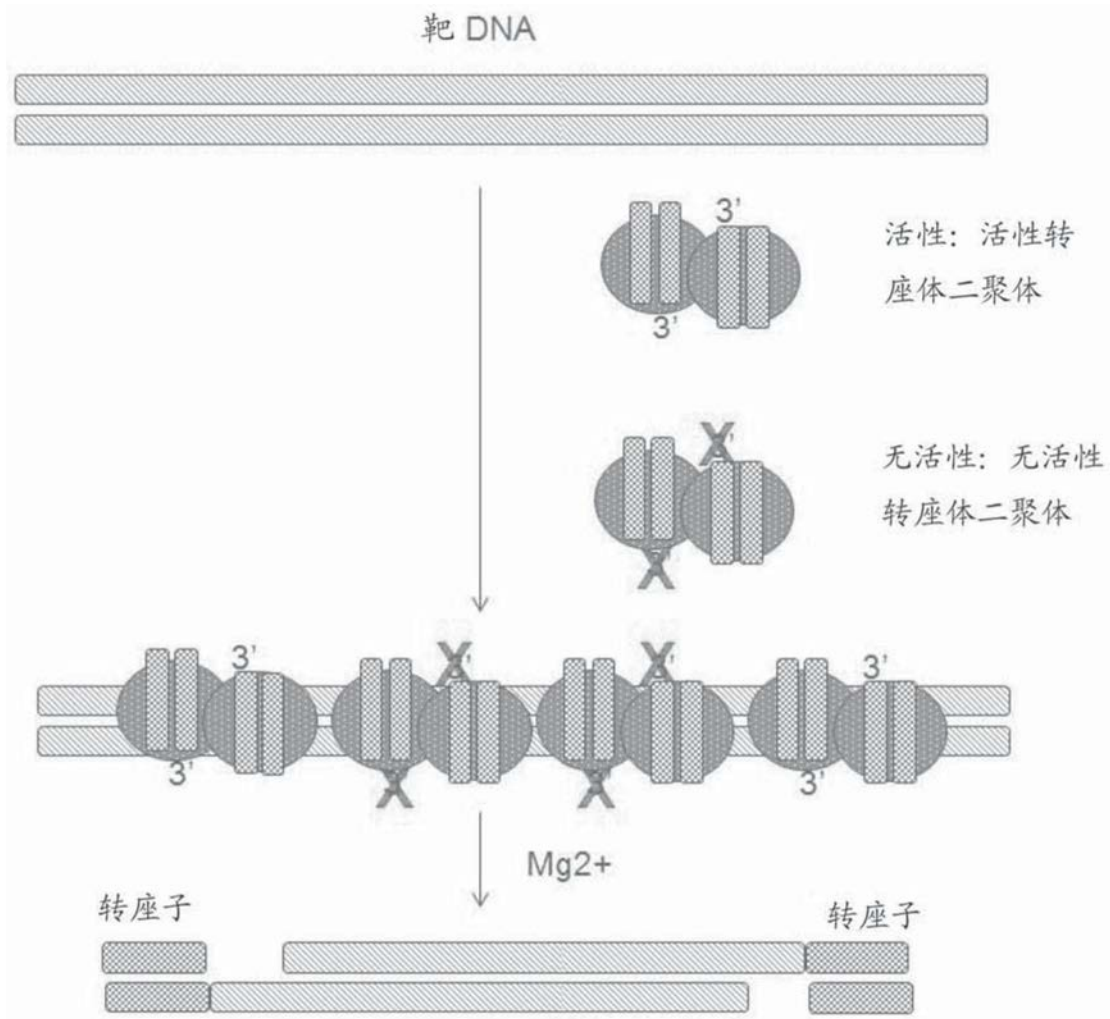


图11

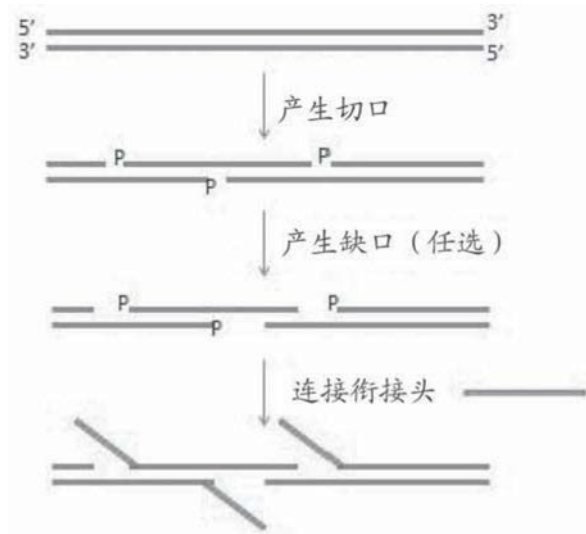


图12

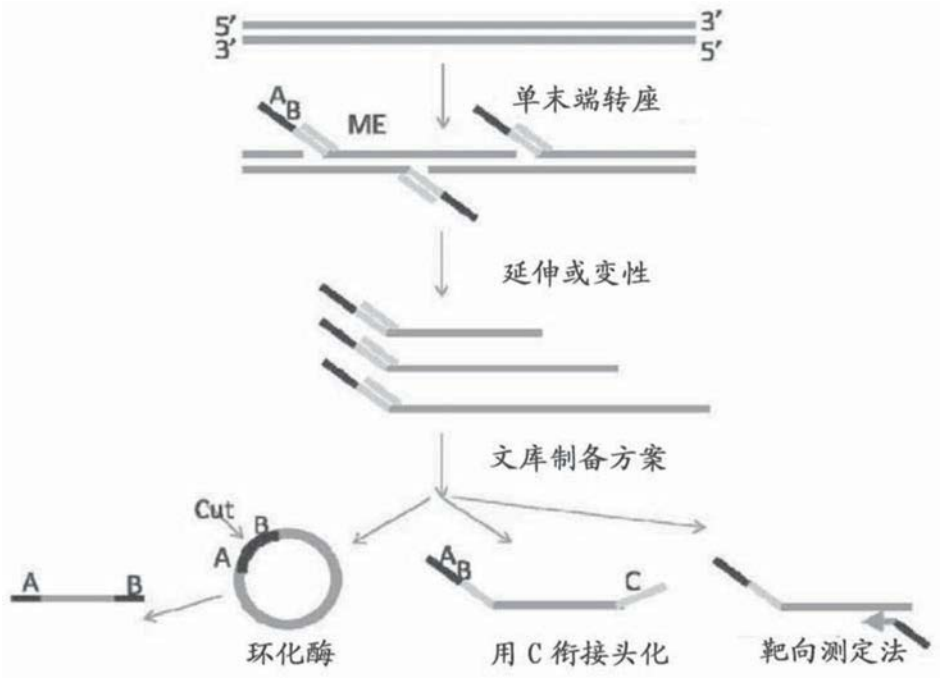


图13

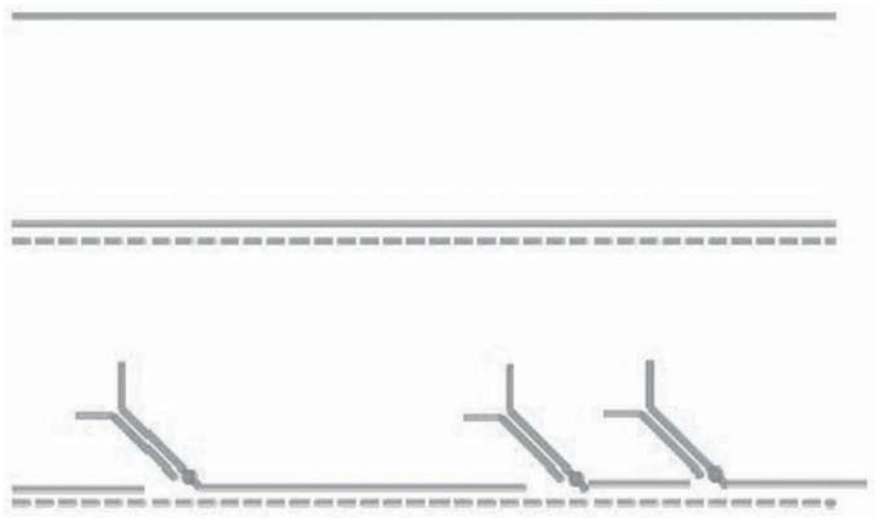


图14

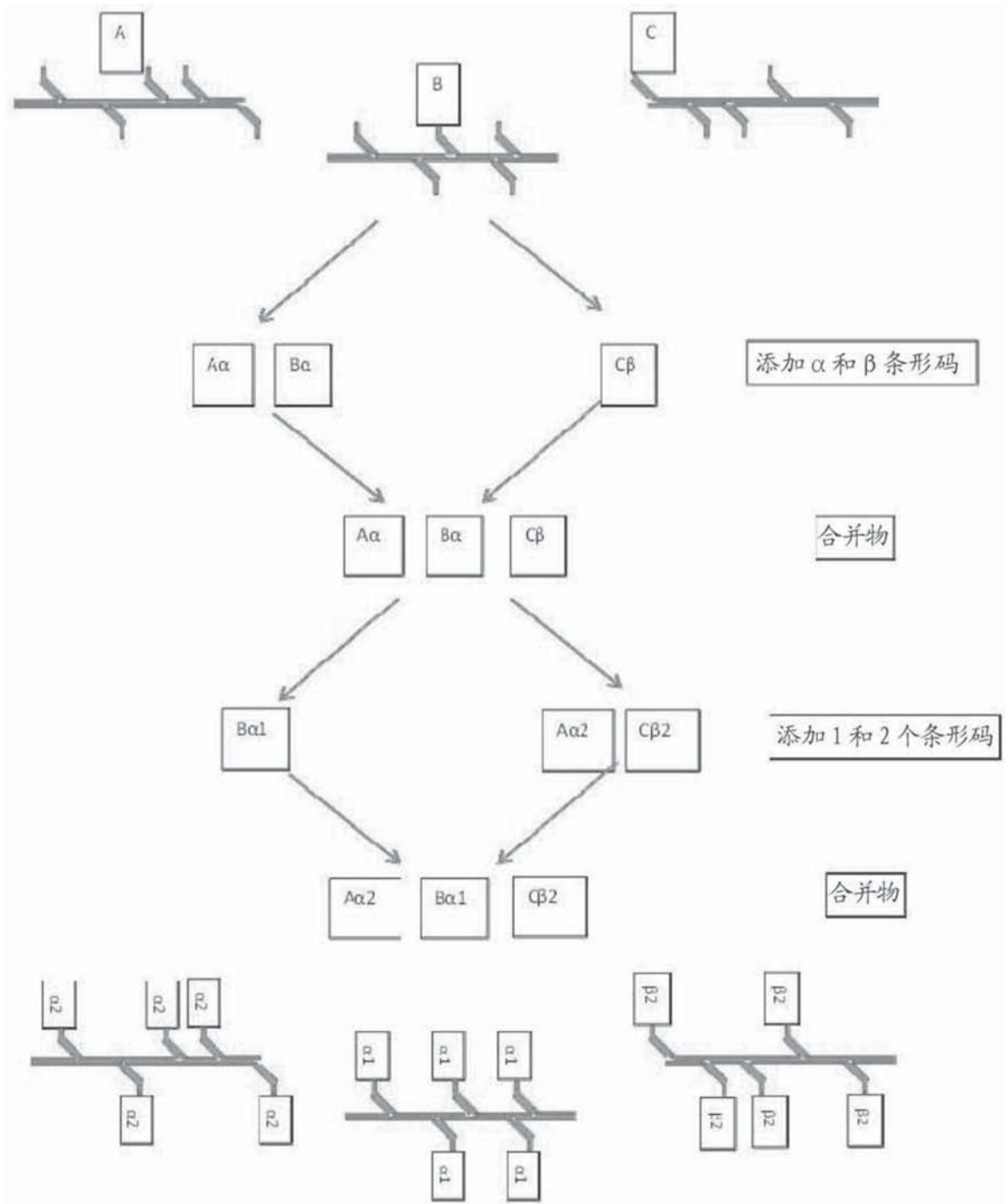


图15