(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104372080 B
(45)授权公告日 2018.03.30

(21)申请号 201410584764.X

(22)申请日 2009.11.18

(65)同一申请的已公布的文献号
申请公布号 CN 104372080 A

(43)申请公布日 2015.02.25

(30)优先权数据
61/115,704 2008.11.18 US

(62)分案原申请数据
200980154567.1 2009.11.18

(73)专利权人 博纳基因技术有限公司
地址 美国宾夕法尼亚州

(72)发明人 肖明 曹涵
帕里克史特·A·德施潘德
迈克尔·博伊斯-加西诺

(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
代理人 杨青 穆德骏

(51)Int.Cl.
C12Q 1/68(2006.01)

(56)对比文件

Chan et al.DNA Mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags.《genome research》.2004,第14卷(第6期),第1137-1146页.

Ming Xiao et al.Rapid DNA mapping by fluorescent single molecule detecion.《Nucleic Acids Research》.2007,第35卷(第3期),第1-12页.

审查员 张蕊

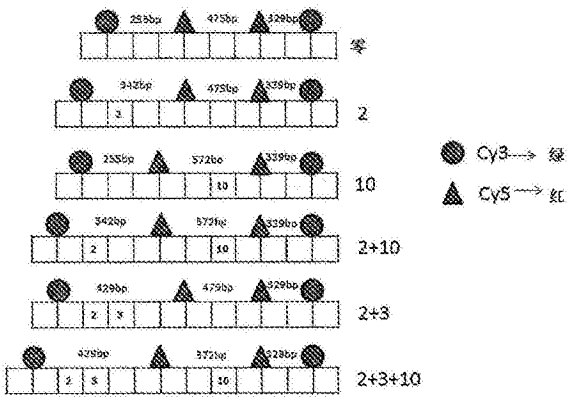
权利要求书2页 说明书13页 附图19页

(54)发明名称

多核苷酸作图和测序

(57)摘要

本发明涉及多核苷酸作图和测序。本发明提供了获得关于生物聚合物样本的结构信息的方法。所述方法包括标记生物聚合物如DNA或RNA的部分、在一些例子中对所述生物聚合物进行线性化、以及确定标记物之间的距离。然后使用者可比较不同样本的标记物间距离以定性地比较不同样本并分析指定样本的在两侧带有标记物的区域中核苷酸的添加或删除。所述方法还可对生物聚合物进行测序。



CN 104372080 B

1.一种获得关于DNA样本的序列信息的方法，其包含：

使用第一序列特异性切口核酸内切酶使第一双链DNA样本产生切口；

使所述第一双链DNA样本与一种或多种标记的核苷酸、聚合酶和连接酶接触，并将所述标记的核苷酸掺入至所述第一双链DNA样本中，在由所述切口核酸内切酶产生的两个或更多个切口位点处；

对包含至少两个染料标记的核苷酸的所述第一双链DNA样本的部分进行线性化；和

确定第一双链DNA样本的所述部分内的、其中的至少两个彼此隔开30bp至1000bp的两个或更多个标记的核苷酸的相对位置。

2.权利要求1的方法，其中所述两个或更多个标记的核苷酸中的至少一个包含终止核苷酸。

3.权利要求1的方法，其还包含确定所述第一双链DNA样本在主双链DNA样本内的相对位置，所述第一双链DNA样本来源于所述主双链DNA样本。

4.权利要求3的方法，其中所述第一双链DNA样本是通过消化从所述主双链DNA样本中获得的。

5.权利要求1至4任一项的方法，其还包含：使用序列特异性切口核酸内切酶使第二双链DNA样本产生切口；将一个或多个标记的核苷酸掺入至所述第二双链DNA样本中，在由所述切口核酸内切酶产生的两个或更多个切口位点处；对包含至少两个标记的核苷酸的所述第二双链DNA样本的一部分进行线性化；以及确定第二双链DNA样本的所述部分内的两个或更多个标记的核苷酸的相对位置。

6.权利要求1至4任一项的方法，其还包含将所述两个或更多个标记的核苷酸的相对位置，与在接触过相同切口核酸内切酶的第二双链DNA样本上的、相同的标记的核苷酸的位置进行比较。

7.权利要求1至4任一项的方法，其中确定两个或更多个标记的核苷酸的相对位置包括：确定位于第一双链DNA的线性化部分内的两个或更多个标记的核苷酸之间的距离。

8.权利要求1至4任一项的方法，其中将标记的核苷酸掺入在切口位点处包括：使切口位点与三种或更多种携带荧光标记物的核苷酸接触，所述核苷酸选自携带荧光标记物A的第一核苷酸、携带荧光标记物B的第二核苷酸、携带荧光标记物C的第三核苷酸以及携带荧光标记物D的第四核苷酸。

9.权利要求8的方法，其还包括对所述第一双链DNA样本进行照射以确定一个或多个携带荧光标记物的核苷酸的存在或相对位置。

10.权利要求9的方法，其还包括通过将在照射下可见的荧光团的顺序与一个或多个所述荧光团的对应核苷酸联系起来，以确定所述第一双链DNA样本的至少一部分的序列。

11.权利要求1至4任一项的方法，其还包括使用第二序列特异性切口核酸内切酶使所述第一双链DNA样本产生切口，从而引入与位于由第一序列特异性切口核酸内切酶产生的两个或更多个切口位点处的核苷酸相邻的切口位点。

12.权利要求1至4任一项的方法，其还包括基于所确定的、标记的核苷酸的相对位置，获得DNA样本的序列信息。

13.权利要求1至4任一项的方法，其中将标记的核苷酸掺入包括：使切口位点与两种或更多种标记的核苷酸接触，所述核苷酸选自携带标记物A的第一核苷酸、携带标记物B的第

二核苷酸、携带标记物C的第三核苷酸以及携带标记物D的第四核苷酸。

14.权利要求13的方法，其中标记物A、标记物B、标记物C和标记物D中的至少两种、三种或四种是不同的。

15.权利要求13的方法，其中使切口位点与三种或更多种标记的核苷酸接触，并且其中标记物A、标记物B、标记物C和标记物D中的至少三种是不同的。

16.权利要求13的方法，其中使切口位点与四种或更多种标记的核苷酸接触，并且其中标记物A、标记物B、标记物C和标记物D中的至少四种是不同的。

17.权利要求1至4以及14至16任一项的方法，其中标记的核苷酸是染料标记的核苷酸。

多核苷酸作图和测序

[0001] 本申请为国际申请日2009年11月18日、国际申请号PCT/US2009/064996于2011年7月13日进入中国国家阶段、申请号200980154567.1、发明名称“多核苷酸作图和测序”的分案申请。

[0002] 相关申请

[0003] 本申请要求2008年11月18日提交的美国申请No.61/115,704的优先权,通过参考将其全部内容纳入本文以用于任何和所有的目的。

技术领域

[0004] 本公开的发明涉及核酸测序领域和分子成像领域。本公开的发明还涉及纳米技术领域。

背景技术

[0005] 随着分子生物技术的进展,对于以不断增加的分辨率和精确度来分析越来越小的样本有了更多的兴趣。这其中的一部分受到种群异质性可经常掩盖样本的重要特征这一认识的驱动。有限的样本体积也是一些应用的考虑因素。

[0006] 虽然现有的技术在理论上能够从身体小样本(physically small sample)中提取重要信息,但是这类技术的有效性受限于它们在这样的小样本上分辨结构特征的能力。因此,在本技术领域中对于能够基于单分子或其它身体小样本获得基因组信息的方法和相关装置有需求。如果这类方法能够改进到超过目前技术所达到的1000bp(1kb)的精确度的话,那么这类方法的价值将会提高。

发明内容

[0007] 为了应对所述挑战,要求保护的本发明首先提供了用于分析一个或多个外显子的存在或相对位置的方法,所述方法包含分别使用第一和第二标记物来标记生物聚合物上的第一和第二位置,使得所述第一和第二标记物位于包括至少一个常外显子(constant exon)的所述生物聚合物的第一区域的两侧;以及对所述生物聚合物进行线性化,并将所述第一和第二标记物之间的距离与生物聚合物的所述第一区域中可变外显子(alternative exon)的存在、不存在、或相对位置联系起来。

[0008] 在第二个方面,本发明提供了获得关于DNA样本的结构信息的方法,其包含使用序列特异性切口核酸内切酶(nicking endonuclease)使第一个双链DNA样本产生切口;使一个或多个染料标记的核苷酸掺入通过所述切口核酸内切酶所产生的两个或更多个切口位点处;对包括至少两个染料标记的核苷酸的所述第一个双链DNA样本的一部分进行线性化;以及记录两个或更多个标记性染料标记的核苷酸的相对位置。

[0009] 还提供了获得关于核酸生物聚合物的序列信息的方法,其包含使具有第一结合序列的第一荧光标记序列特异性探针与单链核酸生物聚合物结合;使所述单链核酸生物聚合物与携带第一荧光标记物的第一终止核苷酸(terminator nucleotide)、与携带第二荧光

标记物的第二终止核苷酸、与携带第三荧光标记物的第三终止核苷酸、以及与携带第四荧光标记物的第四终止核苷酸接触；以及对所述核酸生物聚合物进行线性化和照射以确定与所述第一标记序列特异性探针相邻的所述第一终止核苷酸、第二终止核苷酸、第三终止核苷酸、第四终止核苷酸、或其任何组合的存在或相对位置。

[0010] 本发明还提供了获得关于核酸生物聚合物的结构信息的方法，其包含使双链生物聚合物与切口核酸内切酶接触以产生第一切口位点；使所述第一切口位点与携带荧光标记物A的第一终止核苷酸、与携带荧光标记物B的第二终止核苷酸、与携带荧光标记物C的第三终止核苷酸、以及与携带荧光标记物D的第四终止核苷酸接触；以及对所述双链生物聚合物进行线性化和照射以确定所述第一终止核苷酸、第二终止核苷酸、第三终止核苷酸、第四终止核苷酸、或其任何组合的相对位置。

[0011] 进一步提供了用于进行多重杂交(multiplex hybridization)的试剂盒，其包含各具不同颜色的多个杂交探针；以及将所述多个杂交探针中的至少两个应用在核酸样本上并线性化和成像至少一个杂交核酸的说明书。

附图说明

[0012] 当与附图结合起来阅读时，可对发明概要以及下面的详细描述有进一步的理解。为了对本发明进行图示说明，在附图中显示了本发明的示例性实施方案；然而，本发明不限于所公开的具体方法、组合物、以及装置。另外，所述附图不一定要按比例来绘制。在附图中：

[0013] 图1A图示说明了对于Nt.BstNBI切口核酸内切酶的作图统计数据(mapping statistics)，其显示100bp的光学分辨率显著提高了图谱的精确度和覆盖度；

[0014] 图1B图示说明了对于Nt.BspQI切口核酸内切酶的独特作图统计数据，其显示100bp的光学分辨率对于图谱的精确度和覆盖度影响甚微；

[0015] 图1C图示说明了与相对粗略的图谱(16kb)相比，相对精细的图谱(1.5kb)对于结构变异具有更好的检测力；

[0016] 图2A描绘了MAPT的基因结构；

[0017] 图2B列出了存在于MAPT基因中的每个外显子的每个外显子(可变外显子显示为阴影)尺寸。

[0018] 图2C图示说明了应用于RNA外显子剪接时的超分辨率成像的条形码或作图方案；

[0019] 图2D图示说明了多重条形码方案；

[0020] 图3图示说明了用于测序的起始材料；

[0021] 图4描绘了测序反应的第一个循环；

[0022] 图5描绘了开始于图4的第二个测序循环；

[0023] 图6显示了多重测序方案显著增加了产量；

[0024] 图7A描绘了用于证明SHRIMP分辨率的741bp PCR产物模型系统；

[0025] 图7B图示说明了标记DNA分子在玻璃表面上被线性化后的成像结果，表明三个(3)Cy3染料分子相隔30nm和60nm，这与三个(3)Cy3探针之间94bp和172bp的距离具有良好的一致性。

[0026] 图8A描绘了用于证明SHRIMP和SHREC分辨率的741bp PCR产物模型系统；

[0027] 图8B图示说明了标记DNA分子在玻璃表面上被线性化后的成像结果—Cy3-Cy5对之间的距离为37±5nm(预期为32nm)和91±5nm(预期为87nm),以及Cy3-Cy3对之间的距离为56±3nm(预期为58nm)(图4),显示了极好的一致性;

[0028] 图9描绘了要求保护的本方法查明关于遗传物质的结构信息的范例性、非限制性实施方案;

[0029] 图10描绘了要求保护的本方法查明关于遗传物质的结构信息的第二个范例性、非限制性实施方案;

[0030] 图11描绘了要求保护的本方法的一个非限制性实施方案;和

[0031] 图12描绘了要求保护的本方法的又一个非限制性实施方案。

具体实施方式

[0032] 通过与形成本公开的一部分的附图和实施例相结合参考下列详细描述,可更容易地理解本发明。要理解这一发明不限于在本文中所描述和/或显示的具体装置、方法、应用、条件、或参数,而且本文中使用的术语是为了仅通过实例来描述具体实施方案的目的,且不旨在限制所要求保护的本发明。此外,当在包括随附权利要求书的本说明书中使用时,除非在上下文中另有明确指示,否则不含具体数量的名词包含其复数形式,而且对于特定数值的引用至少包括该特定值。当在本文中使用时,术语"多个"是指大于一。当表达了值的范围时,另一个实施方案包括从一个特定值和/或至另一个特定值。类似地,当通过使用先行词"约"来将值表达为近似值时,将理解为所述特定值形成了另一个实施方案。所有的范围都是包含其中一切的和可以组合的。

[0033] 要理解,为了清晰起见在本文不同实施方案的内容中所描述的本发明的某些特征,还可以在单个实施方案中被组合提供。相反地,为了简洁起见在单个实施方案的内容中所描述的本发明的各种特征,还可被分别地或在任何子组合中提供。而且,引用陈述范围的值包括该范围内的每一个和所有值。

[0034] 在第一个实施方案中,本发明提供了分析一个或多个外显子的存在或甚至相对位置的方法。这些方法适宜包括分别使用第一和第二标记物来标记生物聚合物样本上的第一和第二位置,使得所述第一和第二标记物位于包括至少一个常外显子的所述生物聚合物样本第一区域的两侧。然后使用者再将所述第一和第二标记物之间的距离与生物聚合物的所述第一区域中可变外显子(即不出现在每个mRNA中的外显子)的存在或不存在(或相对位置)联系起来。(所述生物聚合物是与mRNA互补的合适的DNA;这样的DNA可由本技术领域中的普通专业人员容易地合成。)

[0035] 在一些实施方案中,所述第一和第二标记物是相同的荧光团。范围广泛的荧光团适用于本发明,其中包括荧光团的Cy-家族。其它荧光团对于本技术领域的专业人员来说将是已知的;荧光团的列表可见于例如http://info.med.yale.edu/genetics/ward/tavi/FISHdyes2.html。所述标记物可以具有相同的荧光团,但也可以具有不同的荧光团。

[0036] 使用者将所述第一和第二标记物之间的距离与一个或多个可变外显子的存在、不存在、或二者(或甚至所述外显子的相对位置)适当联系起来,这包含将存在于生物聚合物样本上的第一和第二标记物之间的距离与位于已知不含可变外显子的生物聚合物第一区域两侧的标记物之间的距离进行比较。这适宜通过对生物聚合物中包含荧光标记物的区域

进行线性化来完成。生物聚合物的线性化在美国专利申请No.10/484,293(于2009年11月9日授权)中进行了详细讨论,通过参考将其全部内容纳入本文以用于所有目的。

[0037] 所述建立联系适宜包括将第一和第二标记物之间的距离与在第一和第二位置之间缺乏可变外显子的生物聚合物上的第一和第二位置之间的距离进行比较。

[0038] 通过例如图2C来对此进行说明,所述图在"零"处描绘了不具有可变外显子的生物聚合物。在该图中的实施方案"2"描绘了具有可变外显子"2"的生物聚合物,所述可变外显子可通过观察到该外显子导致Cy3和Cy5染料之间的间隔距离增加(342bp)来检测,所述Cy3和Cy5染料在所述图顶端所示的无可变外显子生物聚合物中仅分开了255bp。

[0039] 图2B是显示存在于MAPT基因中的每个外显子(可变外显子通过阴影块来显示)尺寸的表。图2A概括地图示说明了在MAPT基因中可能出现的各种剪接排列。如该图中所示,外显子2、3、以及10被认为是"可变"外显子,并可以-或可以不-存在于MAPT mRNA中。

[0040] 使用者还可以使用另外的标记物(使用例如标记的核苷酸)来适当地标记生物聚合物上的第三、第四、或甚至更多的位置。这类另外的标记物可包括与所述第一和第二标记物相同的荧光团,或可包括与第一和第二标记物上的那些不同的荧光团。然后使用者可将第三标记物和第一标记物、第三标记物和第二标记物之间的距离、或二者与位于第三标记物和第一标记物之间、第三标记物和第二标记物之间、或二者的可变外显子的存在或不存在联系起来。所述关联还可提供一个或多个标记物的相对位置。

[0041] 图2也对此进行了显示。在标记了"2+10"的实施方案中,生物聚合物包括可变外显子2和10,所述外显子位于所述生物聚合物上的第一和第二以及第二和第三标记物(从左至右读取)之间。然后使用者可通过将"2+10"实施方案上标记物之间的距离与在该图顶端所示的"零"实施方案上的标记物之间距离进行比较,来确定这些外显子的存在(或相对位置)。

[0042] 除了通过标记物之间的距离来收集关于所研究的生物聚合物的结构信息外,使用者还可基于两个或更多个探针的相对顺序来获得结构信息,这可通过携带不同颜色荧光团的探针而变得更为便利。例如,如果使用了三个探针(红、黄、以及绿),那么所述探针以红-黄-绿的顺序来结合的序列与所述探针以黄-红-绿的顺序来结合的序列在结构上是不同的。因此,使用者可以通过观察探针在样本上结合/排列的相对顺序以及探针之间的相对距离这二者来收集关于所述样本的信息。

[0043] 回到上述非限制性实施例,对两种样本进行比较的使用者能够-通过说明探针的相对顺序以及探针之间的距离-来确定两种样本在如下方面的差异:(1)某些核苷酸序列显现(appeal)的顺序(通过在不同样本上探针的不同顺序来证实)和(2)在给定样本中例如拷贝变异(copy variation)的数量(通过某些探针在一个样本上比另一个上间隔得更远来证实)。

[0044] 所述标记物适宜彼此隔开约30bp至约1000bp,但更适当约30bp。如在本文的其它地方中所述,许多技术(例如SHRIMP、FIONA、SHREC、或本技术领域的普通专业人员已知的其它技术)能够分辨彼此分开仅数百或甚至数十碱基对左右的小距离的标记物。

[0045] 在另一方面,本发明提供了获得关于DNA样本的结构信息的方法。这些方法适宜包括使用序列特异性切口核酸内切酶使第一个双链DNA样本产生切口。这样的"切口酶"是本技术领域中已知的,且可获自例如New England Biolabs(www.neb.com)。

[0046] 所述方法适宜包括将一个或多个染料标记的核苷酸掺入在由所述切口核酸内切酶产生的两个或更多个切口位点处。根据所述核酸内切酶以及所分析的样本，切口作用(nickage)可沿着所述样本的长度产生一个、两个、或多个切口位点。标记的核苷酸可通过聚合酶被适当地掺入生物聚合物中。在适当的实施方案中，标记的核苷酸是抵消聚合酶作用且不促进链进一步延长的终止核苷酸。根据使用者的需要，所述核苷酸可携带相同的荧光团标记物或不同的标记物。

[0047] 所述方法还适宜包括对包含至少两个染料标记的核苷酸的所述第一个双链DNA样本的一部分进行线性化。一旦标记DNA被线性化，则使用者就可记录或以其它方式来说明两个或更多个标记性染料标记的核苷酸的位置，以用于进一步的分析。

[0048] 一种这样的分析包括将两个或更多个染料标记的核苷酸的相对位置与所述第一个双链DNA样本的一个或多个结构特征联系起来。这可能需要-如图9中所示-确定已知位于目标区域两侧的两个标记物之间的距离，所述目标区域例如在一些个体中已知含有某个突变或拷贝数变异的区域。通过对所述样本上的标记物间距离与对照样本上的标记物间距离(或取自另一个体或个体组(individuals)的另一个样本上的标记物间距离)进行比较，使用者可确定是否所分析的受试者可能具有(或不具有)特定的突变。

[0049] 在一些实施方案中，来源于存在于生物聚合物样本上标记物的相对位置的“条形码”提供了关于主双链DNA样本内的第一个双链DNA样本的相对位置的信息，所述第一个双链DNA样本来源于所述主双链DNA样本。术语“条形码”是指代表样本结构特征的一组信号(例如来自彼此隔开的荧光标记物的信号)(例如两个标记物之间的距离可以与标记物之间区域中基因的额外拷贝的存在联系起来)。所述“条形码”还可用于鉴别特定的样本，在所述特定样本中来自位于样本上的标记物的信号组对于该样本来说是特有的或可将该样本与所研究的其它样本区别开。

[0050] 例如，使用者可确定取自“母本”样本的第一样本上的一部分条形码与取自“母本”样本的第二样本上的条形码重叠，因此表明所述“母本”样本包含第一和第二样本所共有的区域。可对这样的“母本”样本进行消化以产生更小的寡核苷酸，然后所述更小的寡核苷酸自身可通过本文中所述的各种方法进行分析，然后通过对所述更小的寡核苷酸“制作条形码”，使用者可接着确定所述寡核苷酸在“母本”样本中的相对位置。

[0051] 这显示于图13中，其描绘(作图显示)了如下步骤：消化DNA的母本样本，将条形码放置在由消化所产生的产物上，以及将具有相应条形码的产物适当排列-适宜通过计算方法进行，从而将所述母本以及用于母本的有效条形码拼在一起。以这种方法，然后使用者可将母本上的条形码与例如受试者的生理状况联系起来。这在已知用于消化母本的限制性内切酶可分离可能含有拷贝数变异、外显子、或其它突变的基因组区域的情况下可以进行；所述复制数变异、外显子、或其它突变可通过将位于目标区域上的两个标记物之间的距离与位于已知缺乏(或具有)目标突变或外显子的“对照”或“标准”上的两个标记物之间的距离进行比较来检测。

[0052] 作为非限制性实施例，使用者可通过本文中所述的方法将标记物的条形码放置在“母本样本”的消化产物上，然后使用条形码通过计算机来重新装配那些产物以重新形成“母本”。然后使用者可将“母本”的条形码与其它已知样本进行比较以确定所述母本的一种或多种特征，例如拷贝数变异、外显子的添加或缺失等。以这种方法，使用者可通过有效地

将所有消化产物和它们的条形码放置在它们在“母本”内的适当前后关系中来进行“母本”样本的定性评估。

[0053] 所述方法可适宜包括：使用序列特异性切口核酸内切酶使第二个双链DNA样本产生切口，将一个或多个染料标记的核苷酸掺入由所述切口核酸内切酶所产生的两个或更多个切口位点处，对包含至少两个染料标记的核苷酸的所述第二个双链DNA样本的一部分进行线性化，并登记(例如记录或注释)两个或更多个标记性染料标记的核苷酸的相对位置。

[0054] 标记物的这些相对位置-即条形码，可(如前所述)用于确定在主双链DNA样本中第一和第二双链DNA样本之间的关系，所述第一和第二个双链DNA样本都来源于所述主双链DNA样本。

[0055] 在一些实施方案中，使用者将两个或更多个染料标记的核苷酸的相对位置与接触过相同切口核酸内切酶的第二个双链DNA样本上的相同染料标记的核苷酸的位置进行比较。以这种方法，使用者可比较取自不同来源的不同样本上的“条形码”。这使得如图10中所示多个样本之间能够定性比较。在该图中，样本来自受试者A、B、以及C，且根据要求保护的本方法进行了处理。如所示，受试者C的样本缺乏与受试者A和B的样本相结合的标记物，这提示受试者C的DNA缺乏该特定区域。然后使用者可将这一缺失区域与受试者C的生理特征联系起来，或可将受试者C的结果与另外的其它受试者的结果进行比较以鉴别遗失了该DNA区域的个体所共有的那些特征。

[0056] 还提供了获得关于核酸生物聚合物的序列信息的方法。这些方法适宜包括将具有第一结合序列的第一荧光标记序列特异性探针与单链核酸生物聚合物结合。这在例如图11中显示。然后使用者将所述单链核酸生物聚合物与携带荧光标记物A的第一终止核苷酸(例如携带Cy5的腺嘌呤)、与携带荧光标记物B的第二终止核苷酸(例如携带Alexa 405的胞嘧啶)、与携带荧光标记物C的第三终止核苷酸、以及与携带荧光标记物D的第四终止核苷酸接触。然后使用者对所述核酸生物聚合物进行照射以确定与所述第一标记序列特异性探针相邻的所述第一终止核苷酸、第二终止核苷酸、第三终止核苷酸、第四终止核苷酸、或其任何组合的存在(或相对位置)。

[0057] 所述第一探针的结合序列适宜在4至6个核苷酸之间。在一些实施方案中，核苷酸的荧光标记物具有不同的激发波长。在其它情况下，两个或更多个标记物共享一个激发波长。标记的核苷酸的激发波长与标记的序列特异性探针的激发波长可以是相同-或不同的。

[0058] 所述方法还适宜包括将分别具有第二、第三、第四、以及第五结合序列的至少四个荧光标记探针与所述单链核酸生物聚合物接触。所述第二结合序列是通过除去位于第一结合序列5’端的碱基并将第一替代碱基添加至第一结合序列的3’端来适当地构成的。

[0059] 类似地，所述第三结合序列是通过除去位于第一结合序列5’端的碱基并将第二替代碱基添加至第一结合序列的3’端来构成的。所述第四结合序列是通过除去位于第一结合序列5’端的碱基并将第三替代碱基添加至第一结合序列的3’端来适当地构成的，且第五结合序列是通过除去位于第一结合序列5’端的碱基并将第四替代碱基添加至第一结合序列的3’端来构成的。这些探针适宜携带彼此不同的荧光团，且可具有与第一探针不同的荧光团。

[0060] 作为非限制性实例，所述第一探针可包含序列5’-CTAGC-3’。在第二个探测循环中，位于所述探针5’端的C被除去，然后T成为探针的5’端，且探针的3’端如下：5’-TAGCA-

3';5'-TAGCT-3';5'-TAGCG-3';5'-TAGCC-3'。然后将这些标记探针与生物聚合物接触,且通过使用适当的激发波长对探针进行照射,使用者可确定新探针的位置并因此获得关于所研究的生物聚合物的序列信息。虽然在这一实例中所示的结合序列的长度为5bp,但是结合序列适宜1至100bp的长度,但更适宜4bp至6bp的长度。

[0061] 所述方法还适宜包括对核酸生物聚合物进行照射以确定与第二标记序列特异性探针相邻的第一终止核苷酸、第二终止核苷酸、第三终止核苷酸、第四终止核苷酸、或其任何组合的存在(或相对位置)。

[0062] 图11是所述方法的一个非限制性实施方案。如该图中所示,使用者可将具有不同结合序列的第一和第二探针与生物聚合物样本结合。然后使用者可在使相邻所结合的探针只有单个核苷酸与单链DNA结合的条件下,使样本与标记的核苷酸接触。这产生了展示出两个标记物的给定探针-核苷酸对,所述标记物可以-如图中所示-彼此不同。然后使用者可按需照射样本以可视化或以其它方式来定位所述探针-核苷酸对。探针与核苷酸可以通过连接酶来连接。在一些实施方案中,在探针与核苷酸之间可以有间隙(gap)(1+bps),所述间隙可以通过聚合酶和提供核苷酸来填充,所述核苷酸自身可被标记。还可使用连接酶来与探针连接,其中间隙通过标记的核苷酸来"填入"。可以使用非荧光探针。

[0063] 使用者可以在完成探针结合的第一个循环以及随后的标记的核苷酸结合之后,开始使用探针进行第二个循环,第二个循环的探针顾及了在第一个循环中获悉的序列信息。例如,第一个探针可具有AAGG序列,且相邻该探针结合的标记的核苷酸是T。在下一个循环中,使用者可利用这一信息并使用具有序列AGGT的探针,以便获得如上文中所述的更多的序列信息。

[0064] 在另一方面,本发明提供了获得关于核酸生物聚合物的结构信息的方法。这些方法适宜包括(a)将双链生物聚合物与切口核酸内切酶接触以产生至少两个切口位点;(b)将所述至少两个切口位点与携带荧光标记物A(例如Cy3)的第一核苷酸接触;(c)除去过量的第一核苷酸;(d)照射所述双链生物聚合物以确定第一核苷酸的存在或相对位置;(e)将所述至少两个切口位点与携带荧光标记物B(例如Cy5)的第二核苷酸接触和(f)除去过量的第二核苷酸。使用者适当地照射双链生物聚合物以确定第二核苷酸的存在或相对位置。

[0065] 使用者适宜将至少两个切口位点与携带荧光标记物C(例如Alexa405)的第三核苷酸接触;除去过量的第三核苷酸;(j)照射双链生物聚合物以确定第三核苷酸的存在或相对位置。所述方法还包括(k)将所述至少两个切口位点与携带荧光标记物D的第四核苷酸接触,(l)除去过量的第四核苷酸;以及(m)照射双链生物聚合物以确定第一核苷酸的存在或相对位置。

[0066] 以这种方法,所述切口酶"打开"了所述双链样本,使得与切口酶结合的位置相邻的核苷酸可用。然后使用者引入所述第一标记的核苷酸(例如胞嘧啶),并对所述生物聚合物进行分析以确定是否该核苷酸可能已结合以及在哪里结合。然后使用其它核苷酸(鸟嘌呤、酪氨酸、腺嘌呤核苷)来重复这一操作,在引入了所述其它核苷酸中的每一个之后,使用者(通过照射)来分析每个新引入核苷酸的结合。

[0067] 然后可重复前面的步骤(确定为(b)至(m)),使得使用者能够通过添加每个相继的标记的核苷酸来获得更多的序列信息。

[0068] 所述照射还适宜确立一个或多个标记的核苷酸的相对位置。具有两个或更多个标

记物的样本的至少一部分适宜线性化以用于这一分析。然后使用者确定位于双链生物聚合物的线性化部分内的两个或更多个标记的核苷酸之间的距离。然后可使用这些距离来获得用于所分析样本的条形码。

[0069] 在一些变化形式中，使用者可产生与位于第一个切口位点处的终止核苷酸相邻的第二切口位点。使用者适宜将第二切口位点与携带荧光标记物A的第一核苷酸、与携带荧光标记物B的第二核苷酸、与携带荧光标记物C的第三核苷酸、以及与携带荧光标记物D的第四核苷酸接触，并照射双链生物聚合物以确定掺入第二切口位点掺入的标记的核苷酸。

[0070] 这显示于图12中。如该图中所示，两个切口酶分子与双链DNA样本结合并在它们的末端产生了切口位点，通过该图中的加框N来显示。然后使用者依次引入标记的核苷酸。如该图中所示，腺嘌呤核苷被首先引入并与位于左侧探针对面的DNA链上的T结合。因为右侧探针的对面有腺嘌呤核苷，所以标记腺嘌呤核苷没有结合在该位点，且“X”表示在引入第一个标记碱基后那里没有发生结合。引入其它的切口酶和标记碱基，使用者能够通过相继添加标记碱基接着照射所标记的样本，对生物聚合物靶进行测序。然后从所述方法收集的序列信息可用于设计与特定序列结合的探针，然后可使用所述探针对指定样本“制条形码”，以用于进一步表征，例如将第一样本上两个或更多个标记探针之间的相对距离与不同或对照样本上相应的标记探针的距离进行比较。

[0071] 本发明还提供了用于进行多重杂交的试剂盒。这些试剂盒首先适宜包括多个杂交探针。每个探针适宜具有不同颜色或对不同的激发波长作出反应。所述试剂盒还适宜包括对于将这些杂交探针中的至少两个应用于核酸样本上、线性化所标记的样本、以及使至少一个杂交核酸成像的说明书。在一些实施方案中，使用者使两个或更多个杂交探针成像以确定两个探针之间的距离或两个探针之间的相对位置。

[0072] 根据某些条件，使用者可使用标记的核苷酸来填充(populate)相邻切口位点之间的整个生物聚合物区域。这适宜在所述切口位点彼此比较靠近时完成。在照射下，具有至少一些标记的核苷酸的生物聚合物区域比较亮；缺乏标记的核苷酸的区域比较暗。然而，使用者仍然可以从亮和暗区二者中收集信息。

[0073] 所谓亮区提供序列信息，因为使用者可以使用与位于所述区域内的各种标记的核苷酸相应的激发波长来照射该区域。在其它实施方案中，使用者可以通过确定位于暗区两侧的亮区(或甚至核苷酸)之间的距离来评估是否所述暗区-凭借其尺寸-包含拷贝数变异、外显子、或其它目标结构特征。因此，结构信息可从亮和暗区二者来收集。

[0074] 在一些实施方案中，使用者可选择使用的切口酶具有与生物聚合物样本上具有特别关注的区域互补的结合序列。以这种方法，使用者可有效地获得仅针对被认为具有最大兴趣或重要性的那个区域(或多个区域)的序列信息。

[0075] 使用者还可通过将在照射下可见的荧光团的顺序与一个或多个所述荧光团与之相对应的核苷酸联系起来，以适当地确定生物聚合物样本中至少一部分的序列。

[0076] 其它的公开

[0077] 成像技术

[0078] 几种技术将荧光成像中的光学分辨率提高了至少一个数量级。这些成像技术在单分子DNA和RNA分析上的应用大大促进了上文中所讨论的应用。

[0079] 一种这样的技术被称为具有单纳米精确荧光成像(FIONA)，包括通过将分布函数

与从荧光团收集的光相拟合来对单个有机荧光团进行定位。这一分布的核心是能够以1.5nm的精确度来定位。FIONA已用于研究分子发动机的移位或测量小距离。

[0080] 这一技术的扩展包括使用光漂白的单分子高分辨率成像(SHRIMP)，其能够以约10nm的分辨率来分辨具有相同颜色的相邻荧光团。FIONA已被扩展至两个颜色，开发了被称作单分子高分辨率共定位(SHREC)的方法。使用者可以，例如，对一起接近到10nm的Cy3和Cy5染料进行共定位，所述染料可连接在短DNA的末端。还可使用多色随机光学重建显微术(STORM)方法，其允许报告分子与激活分子的组合配对。对这些探针的稀疏子集(sparse subset)进行反复、颜色特异性的激活可允许具有纳米精确度的定位。

[0081] 基因组作图方法

[0082] 结构变异在人类健康和常见疾病方面起着非常重要的作用。这些变异被定义为长于1kb。但是尽管它们很重要，大多数用于检测拷贝数变异(CNVs)的基因组层面方法(genome-wide approach)是间接的，要根据样本和对照之间信号强度的差异来预测变异区域。因此这类方法提供了有限的定量信号和位置信息，且无法检测平衡事件如倒位和易位。例如，包括SNP阵列、寡核苷酸比较基因组杂交(CGH)阵列、以及BAC CGH阵列在内的基于微阵列的平台是用于发现结构变异的主要技术。这些平台灵敏度、特异性、以及探针密度不一致经常导致相冲突的结果，即使对于完全相同的样本也是如此。这一定性测量需要通过低通量检测方法如PCR和FISH来进一步证实。

[0083] 光学作图

[0084] 上文中所述的单分子技术非常适合于研究结构变异。然而，由于所述作图的光学性质，它们在分辨比约～1kbp更靠近的基序的能力上是有限的。明显更高的作图效率可通过分辨相隔小于100bp的特征来实现。进而，这大大提高了我们鉴别天然的、长基因组DNA分子中的结构变异的能力。

[0085] 合适的作图方案是基于对由切口核酸内切酶所产生的位点进行标记。具有五个碱基识别序列的切口核酸内切酶将平均产生跨越整个基因组的1kb物理图谱。基于计算机(in silico)全基因组作图，这类切口位点的大部分都落在彼此的1000bp内，这一距离无法使用传统的光学装置来分辨。这降低了图谱分辨率并使图谱的装配更加困难。

[0086] 一个例子是两种可商购的切口核酸内切酶的识别序列(基序)，具有5个碱基至7个碱基识别位点。设计出了运算法则来绘制针对人类参考基因组的所有切口位点的图谱。

[0087] 在酶Nt.BstNBI(5碱基基序GACTC)的例子中，跨越整个人类基因组有2.1×10^6个位点，这表明在切口之间平均为1.5kb。对于酶Nt.BspQI(7碱基基序GCTCTTC)来说，有平均以15kbp隔开的2.2×10^5个切口位点。原则上，使用所述5碱基基序的切口位点可以使用传统的光学装置(～1kbp)来分辨，但是计算机分析显示差不多有一半切口位点落在了彼此1kbp以内，这使得它们彼此难以区分。使用所述7碱基基序，人们可以分辨数量更大的位点。如下面所讨论的，这导致了在DNA片段的唯一性作图中的挑战。

[0088] 提高DNA作图中的分辨率

[0089] 计算机作图用于确定基于目前可获得的切口酶和我们现有的光学检测系统可唯一作图的DNA片段的百分比。

[0090] 图1A显示了切口核酸内切酶Nt.BstNBI(5碱基基序)的结果。对于1000bp的光学分辨率来说，仅有约12%的片段可使用8个切口位点被唯一鉴别。另一方面，达到100bp的分辨

率后,97%以上的片段是唯一的。紧密聚集的切口位点装载了更多序列信息,且它们的分布是唯一的。此外,在仅有8个切口位点的情况下,人们仅需要12kb的片段(平均起来)就能够实现片段针对参考基因组的唯一性作图。

[0091] 酶Nt.BspQI(7碱基基序)的切口图谱(图1B)显示,将分辨率提高至100bp,人们获益极小,因为落入彼此1kbp之内的Nt.BspQI切口位点较少。使用这种酶,需要平均8个连续Nt.BstQI切口位点来唯一地鉴别DNA片段,但是片段的平均尺寸为约120kb。由于在可使用现有方法合理提取出来的DNA长度内缺乏连续的切口位点,所以有相当多的基因组区域(~30%)不能作图。

[0092] 在不受任何单个理论制约的情况下,可鉴别出要求保护的本发明的一些优势。第一,在分辨间隔紧密的切口位点时,可获得多得多的关于DNA片段的信息。将片段针对基因组唯一作图的能力大大提高了。

[0093] 第二,随着分辨率提高,人们可分辨比目前使用光学方法可能分辨的小得多的结构变异。最后,提高分辨率还有助于我们鉴别大级别的结构变异。

[0094] 附图上的其它背景

[0095] 图1C中显示了具有150kbp插入物的片段的实例。成功地绘制所述片段的图谱(并因此对基因组内插入物的位置进行鉴别),可使用与插入物相邻的8个切口位点的连续组(contiguous set)。以有限的光学分辨率,这需要大的(>300kbp)基因组片段。这些难以使用标准的DNA提取规程来生成。相反,以100bp的分辨率,人们可以使用仅略大于所述插入物的片段,利用密集切口位点分布来唯一地绘制所述片段的图谱。

[0096] 对于可变转录组(alternative transcriptome)高通量数字作图(profiling)的需求

[0097] 可极大地受益于作图能力提高的另一种核酸分析是RNA的可变剪接。在RNA前体的剪接过程中,内含子被除去,且外显子被连接在一起形成成熟RNA。通过可变剪接这一过程,单个的初级转录物产生了不同的成熟RNA。这导致了具有多样性且甚至拮抗功能的蛋白质同工型的产生。最近的研究显示大蛋白质组学的(proteomic)复杂性和多样性是以有限数量的基因来实现的。在人类基因组中,~75%的人类基因显示了可变剪接。虽然人类基因组含有25,000个基因,但是它能通过可变剪接产生几十万种不同类型的蛋白质。

[0098] 许多基因的可变剪接变体对于细胞生物学的所有主要方面包括细胞周期调控、凋亡等具有决定性影响。已发现异常的剪接与包括癌症在内的各种疾病相关,且最近的研究显示mRNA在癌变组织中比在正常组织中更频繁地被可变剪接。其它例子包括由于包含和包含(inclusion and inclusion)异常的外显子所致的全长跨膜传导调节(CFTR)基因的显著降低,这引起非典型形式的囊性纤维化。另一个例子是微管相关蛋白τ(MAPT基因)。MAPT对于微管的聚合和稳定性以及神经元中的轴突运输来说是必需的。τ外显子10的异常剪接导致了神经退行性疾病—痴呆FTDP-17的发展。

[0099] 已开发了许多技术来量化RNA剪接变体。第一,寡核苷酸微阵列和光纤阵列已被用于整体检测基因剪接变体。然而,因为在阵列技术中每次查询一个全RNA转录物的小片段,所以每次仅可检测一个剪接事件(每次两个外显子)。因此,难以量化在一个特定剪接变体中包含或排除了多少外显子。而且,非特异性杂交可产生许多需要进一步证实的假阳性。

[0100] 第二,实时PCR可以通过每次量化一个外显子连接来获得剪接信息,但它受到严格

的反应条件、低通量、以及高成本的限制。第三，所谓的下一代测序技术已用于数字基因表达图谱中，且可用于绘制可变剪接变体的图谱。然而，它们主要基于短序列读取，且就全长RNA样本来说它们具有与微阵列相同的限制。

[0101] 现有的聚焦于转录组的技术所共有的劣势是没有一个能够监测可变剪接的外显子的组合，因为它们发生于个体转录物内。在现有的方法下，难以对外显子的排除进行证实，这可导致某些外显子的假排除。

[0102] 尽管可变剪接对于哺乳动物生物学来说极为重要，但是目前破译这一问题的解决方案面临着挑战。的确，由于缺乏强大的方法来量化RNA剪接变体，所以对于在整个发育阶段中如何调节和协调可变剪接知之甚少。

[0103] 提高分辨率以超过传统的光学限制

[0104] 作为提高分辨率可获得的优势的一个例子，人们可思考用于微管相关蛋白τ(MAPT)基因的光学条形码方法，所述基因对于微管的聚合和稳定性以及神经元中的轴突运输来说是必需的。τ外显子10的异常剪接导致了神经退行性疾病如痴呆FTDP-17的发展。

[0105] 一个示例性RNA条形码方案显示于图2中。MAPT转录物中的三个外显子(2、3、以及10)可经历可变剪接，外显子2和外显子3总是剪接在一起。因此，通过可变剪接可产生六个不同的MAPT转录物。所述MAPT基因结构显示于图2A中。

[0106] 显示了所有六种可能的可变剪接同工型(零、2、10、2+10等)，且每个外显子的长度显示于图2B中。传统的光学分辨率不能辨别与不同外显子相连的标记物。如果外显子的位置能够被分辨的话，那么所测得的标记物之间的距离将以与读取条形码类似的方法来鉴别每个剪接变体。

[0107] 为了形成这个实施例中的条形码，可设计四个外显子特异性寡核苷酸探针以分别与外显子1(Cy3-绿)、外显子7(Cy5-红)、外显子11(Cy5-红)、外显子13(Cy3-绿)进行特异性杂交，如图2C中以绿和红箭头所示。标记物之间的距离可用于鉴别存在哪个变体且颜色序列(即绿-红-红-绿)显示全标记转录物的存在。而且，本公开的条形码方案易于多重进行。

[0108] 例如，如果使用相同的两个颜色(例如绿和红)与四种不同的探针来标记一个不同的基因，则可设计出与MAPT基因不同的用于这一特定基因的颜色序列。因此颜色序列可用于定义特定基因，且该颜色序列的标记物之间的距离确定了该特定基因的个体剪接变体。在这个二色、四探针方法中，有$2^4=16$个不同的颜色序列以无限的能力来同时查询16个不同基因的剪接变体。如果使用8种不同探针的4种颜色，则可同时调查$4^6=65536$个不同基因，这超过了整个人类转录组(图2C)。

[0109] 这一方法与目前用于查询RNA剪接的表达作图技术相比有三个重要优势：(i)通过同时绘制单个转录物内外显子的分布图谱，人们可以确定同一个转录物内多个可变剪接外显子之间的关系。(ii)所述条形码方案的数字特性意味着人们不仅可量化个体剪接变体，人们还可通过将所有剪接变体加在一起来量化全部的基因表达。(iii)所述条形码方案将提供最大的多重检测能力。实现这些优势需要所具有的分辨率远超过传统光学方法的成像技术。

[0110] 对于低成本和高通量的全基因组测序的需求

[0111] 人类基因组计划(HGP)的成功很大程度上归功于凭借平行化、自动化、微型化、更好的化学和信息学的Sanger测序法的持续发展。作为人类基因组计划的主力，Sanger测序

法已支配了DNA测序领域将近三十年，且它的800Q20碱基读取长度具有重要的意义。

[0112] 这些新兴的测序技术可基于检测方法被分为两类：通过总体检测(ensemble detection)或通过单分子检测的测序。因为在总体检测中需要多个DNA拷贝，所以在该过程期间遗传信息如单倍体型和RNA剪接模式被遗失。虽然通过单分子检测的测序可能能够回收单倍体型信息，但是目前的单分子测序法(例如Helicos tSMS)的读取长度是50bp或以下，这远远短于两个SNP之间1kbp的平均距离。因此，如同使用前辈Sanger测序法一样，关键性遗传信息如单倍体型和RNA剪接模式仍然难以使用这些“下一代”测序技术来获得。本发明尤其实现了超过10kb的DNA测序长度。

[0113] 通过杂交来进行测序是利用基于微阵列的杂交分析来确定核酸分子序列的公知方法。通常，使用在微阵列上构建的具有已知序列(<100聚体)的短寡核苷酸来捕获(即杂交)并查询靶分子。所述微阵列分析产生了在靶分子中至少发现一次的杂交寡核苷酸所有序列的列表。然而，所述列表不显示杂交寡核苷酸序列的位置，所述列表也不提供寡核苷酸可存在于靶分子上的次数。然而，本发明获得了这类信息。

[0114] 图3显示了用于测序的起始材料。一组在5’端以不同颜色荧光团标记的5聚体(即长度为5个核苷酸)寡核苷酸；以不同颜色荧光团标记的4个核苷酸终止子；线性化单链DNA分子、或具有部分ssDNA间隙的双链DNA分子的阵列。

[0115] 图4描述了示例性测序反应的第一个循环。在第一个循环后，通过STORM成像技术沿着线性化DNA分子记录并定位每个杂交和掺入事件。然后洗掉探针。在下一个循环中，另有4种5聚体探针AGTCA、AGTCT、AGTCG、以及AGTCT被引入并杂交在与先前探针相同的位置上，因为它们享有与先前探针相同的序列。然后聚合酶掺入核苷酸终止子(图5)。

[0116] 将这一过程改变为多重进行的(使用不同颜色的标记物)并产生在一个循环期间读取的大量序列(图6)。还开发了运算法则来确定5聚体探针连续性添加的优选次序。本文中使用的超级成像技术包括SHRIMP、SHREC、STORM。

[0117] 实施例

[0118] 已开发了单分子高分辨率共定位(SHREC)和使用光漂白的单分子高分辨率成像(SHRImP)方法来测量比瑞利极限(Rayleigh limit)(对于可见激发来说≈250nm)更靠近的两个荧光团之间的距离。

[0119] 对所述两个技术进行组合，在定位方法学的能力上增加了另一个方面，且通过使用各具多个成员的几个不同颜色荧光团能够有潜力分辨几十个距离。为了将此应用在DNA上，可将双链DNA延展在聚丙烯酸和聚丙烯胺包被的表面上，使得所述DNA相对变直。为了测试SHRIMP，使用生物素来制造DNA构建物，然后在与32nm、58nm、以及90nm的Cy3之间距离相对应的位置475bp、172bp、以及94bp处添加三个Cy-3(图7B)。

[0120] 更多的细节提供在图7A中。一个PCR引物使用Cy3在5’端标记，且另一个引物在5’端被磷酸化。在PCR反应后，Cy3的5’端保护该链不被λ核酸外切酶消化，这产生了单链DNA分子。一旦产生了所述单链DNA分子，就可进行引物延伸反应以在每个特定序列位置处引入荧光染料。在这一情况下，5’端具有Cy3的两个短寡核苷酸分别在距一端94bp和256bp处被杂交。5’端具有生物素的另一个短寡核苷酸在单链模板的3’端被杂交。通过聚合酶进行延伸后，单链模板被转化为双链DNA分子，且两个Cy3染料分子被引入在特定位点处。

[0121] 测得距离为27nm、61nm、以及95nm，与所预期的距离有极佳的一致性。为了同时测

试SHRIMP和SHREC，将Cy5置于零位置处，且两个Cy3在94bp位置和172bp位置处，其中使用双视成像系统(dual-view imaging system)来测量它们的位置。Cy3-Cy5对之间的距离为37±5nm(预期为32nm)和91±5nm(预期为87nm)，且Cy3-Cy3对之间的距离为56±3nm(预期为58nm)(图8)。一致性极佳。

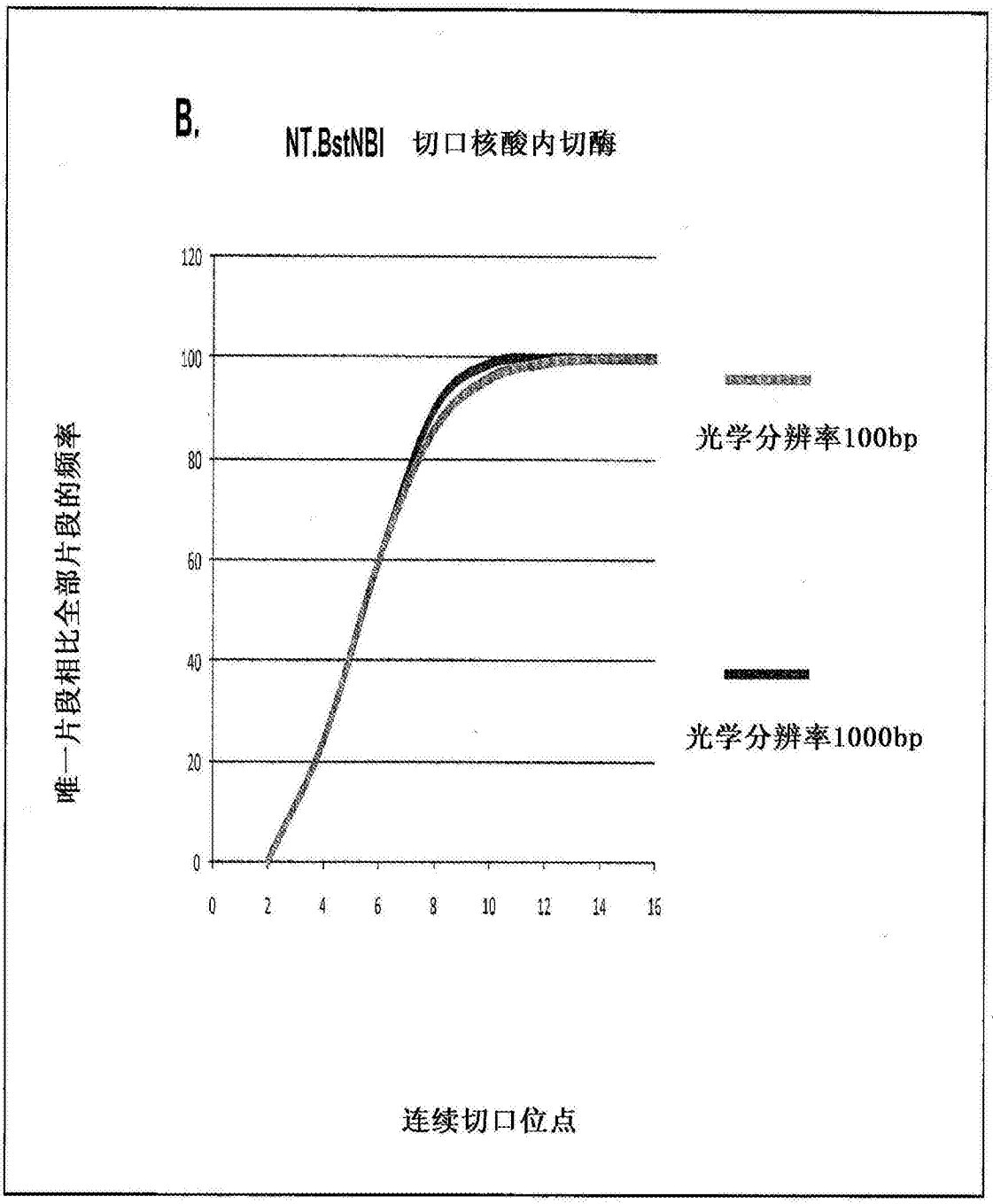

B.
NT.BstNBI 切口核酸内切酶
唯一片段相比全部片段的频率
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
16
光学分辨率100bp
光学分辨率1000bp
连续切口位点

图1A

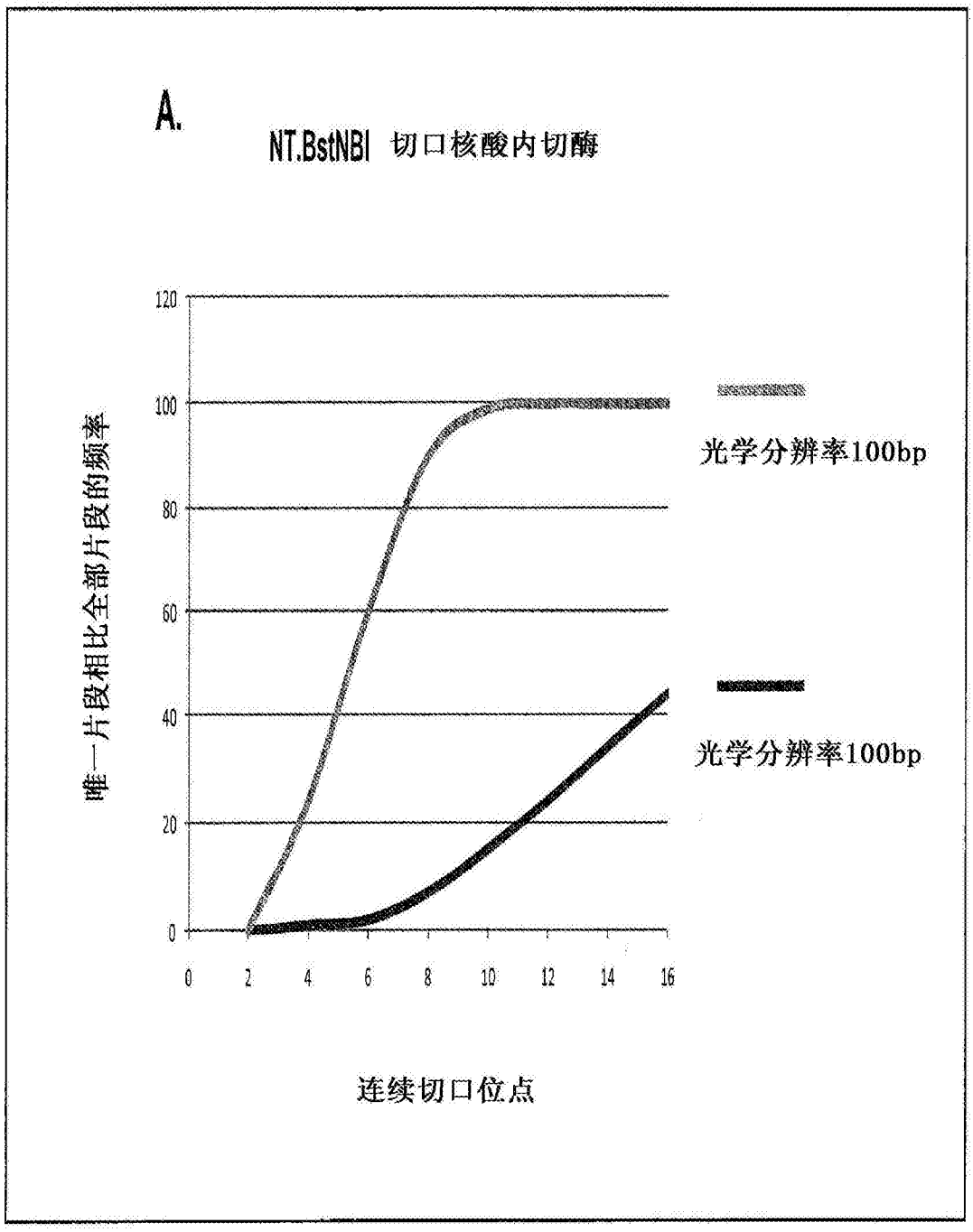

A.
NT.BstNBI 切口核酸内切酶
唯一片段相比全部片段的频率
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
16
连续切口位点
光学分辨率100bp
光学分辨率100bp

图1B

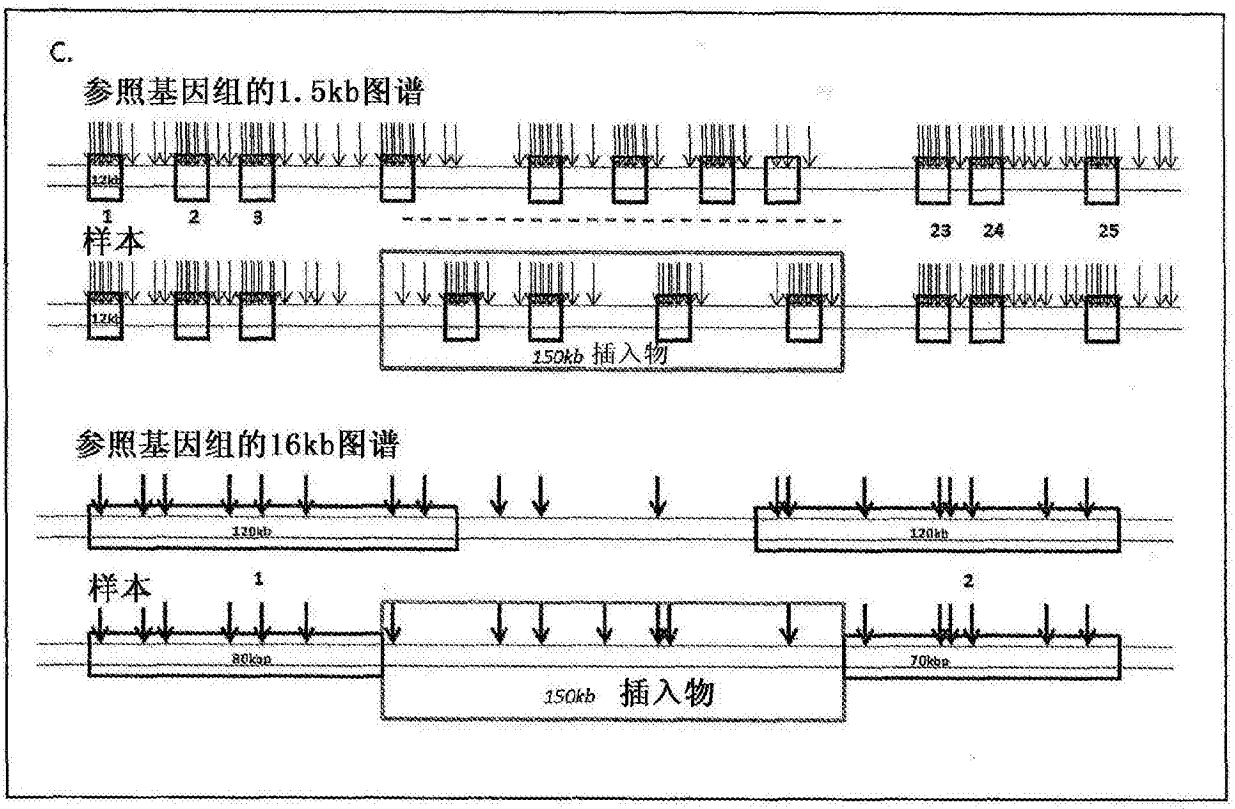
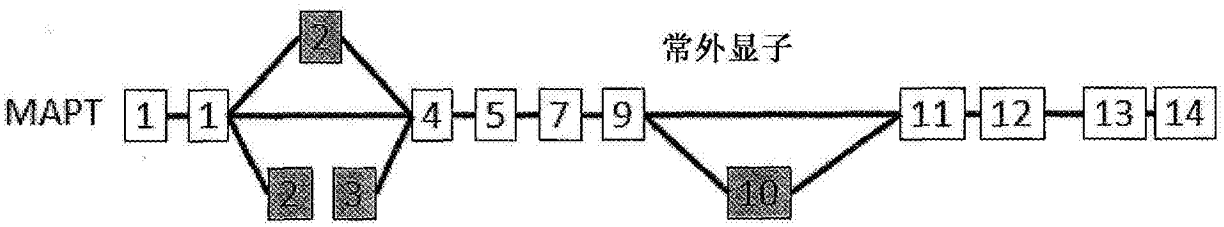

C.
参照基因组的1.5kb图谱
1
2
3
23
24
25
样本
150kb 插入物
参照基因组的16kb图谱
120kb
120kb
样本
1
2
80kbp
70kbp
150kb 插入物

图1C





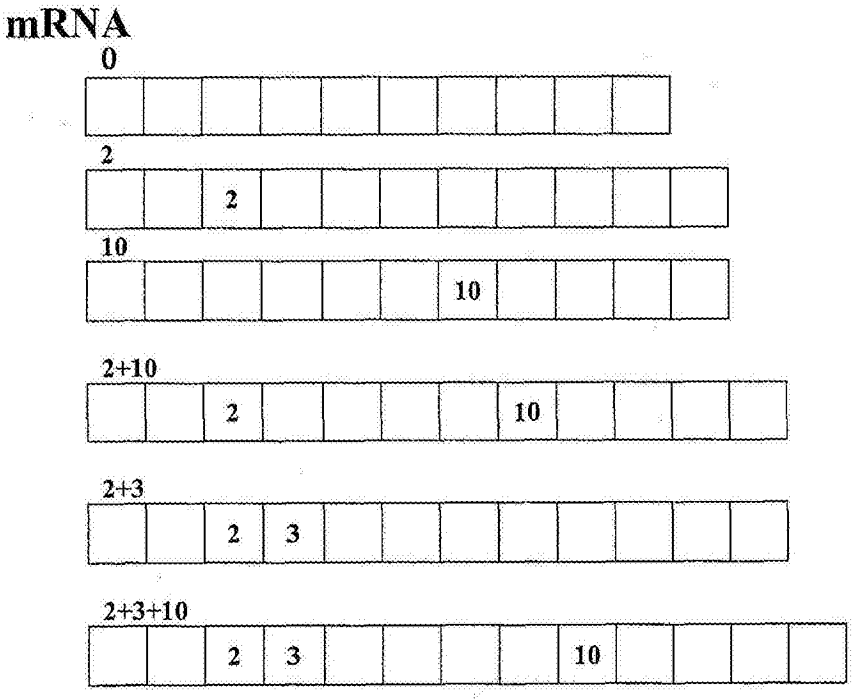


图2A

外显子1	外显子1 133bp	外显子2 87bp	外显子3 87bp	外显子4 66bp	外显子6 56bp	外显子7 127bp	外显子9 266bp	外显子10 93bp	外显子11 87bp	外显子12 113bp	外显子13 216bp	外显子14 87bp

图2B

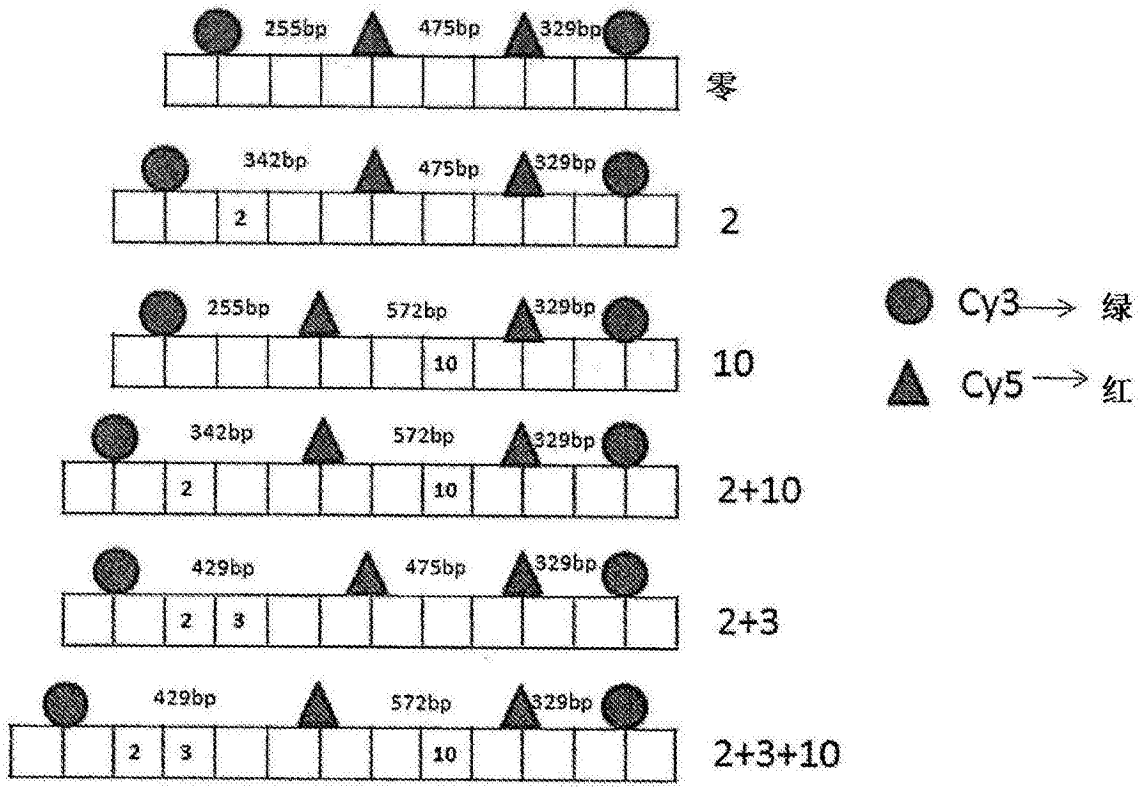
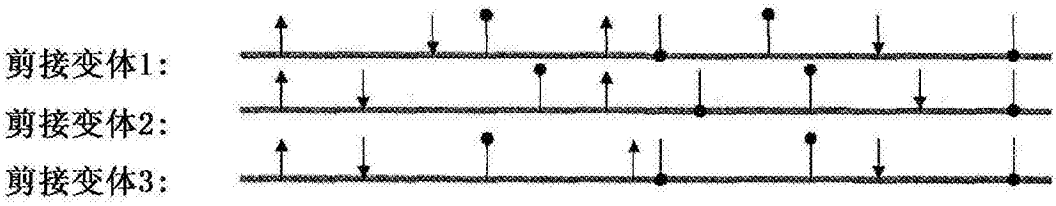
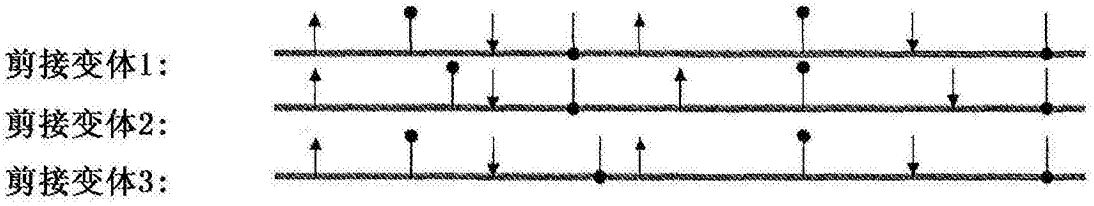

255bp
475bp
329bp
零
342bp
475bp
329bp
2
2
255bp
572bp
329bp
10
10
Cy3 → 绿
Cy5 → 红
342bp
572bp
329bp
2
10
2+10
429bp
475bp
329bp
2
3
2+3
429bp
572bp
329bp
2
3
10
2+3+10

图2C

D 四色八探针可将65536个基因的基因表达和剪接模式制作条形码



具有绿-蓝-黄-红-绿-蓝-黄-红条形码的基因2，
颜色之间的不同距离确定了基因内的每个剪接变体



基因3



图2D

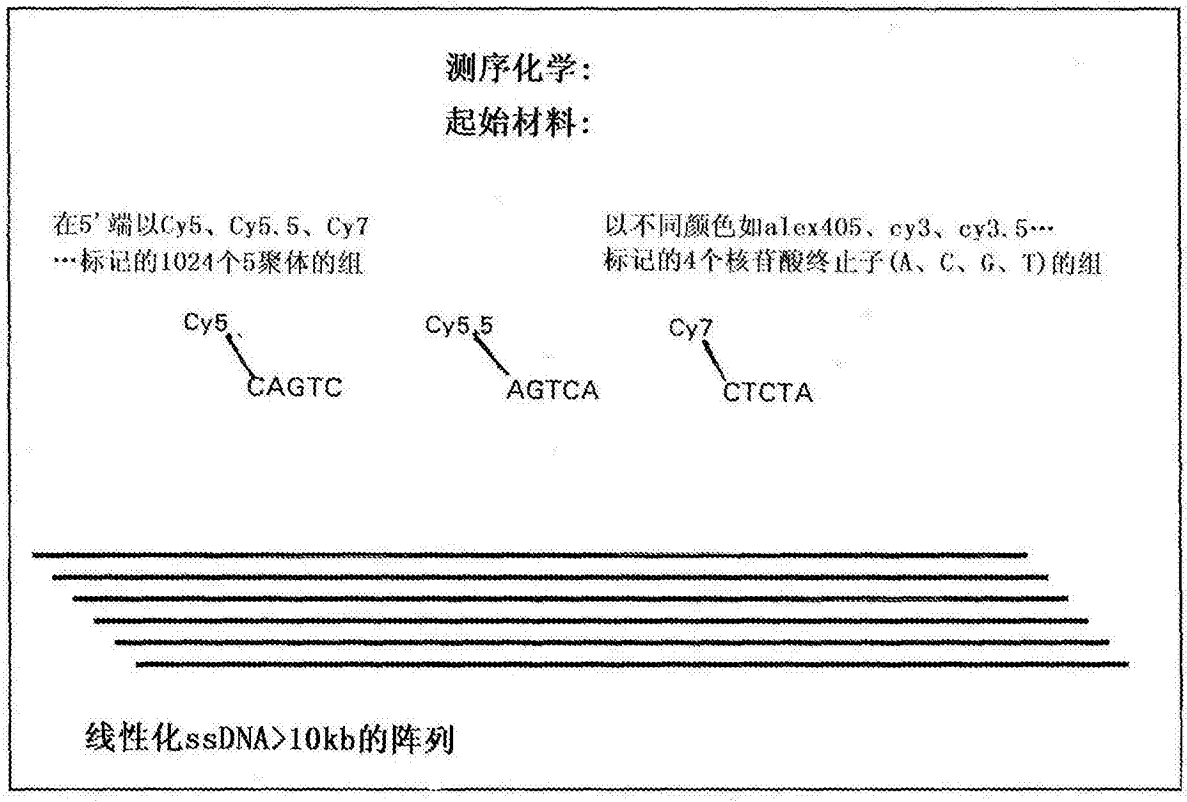

测序化学:
起始材料:
在5'端以Cy5、Cy5.5、Cy7
…标记的1024个5聚体的组
以不同颜色如alex405、cy3、cy3.5…
标记的4个核苷酸终止子(A、C、G、T)的组
Cy5
CAGTC
Cy5.5
AGTCA
Cy7
CTCTA
线性化ssDNA>10kb的阵列

图3

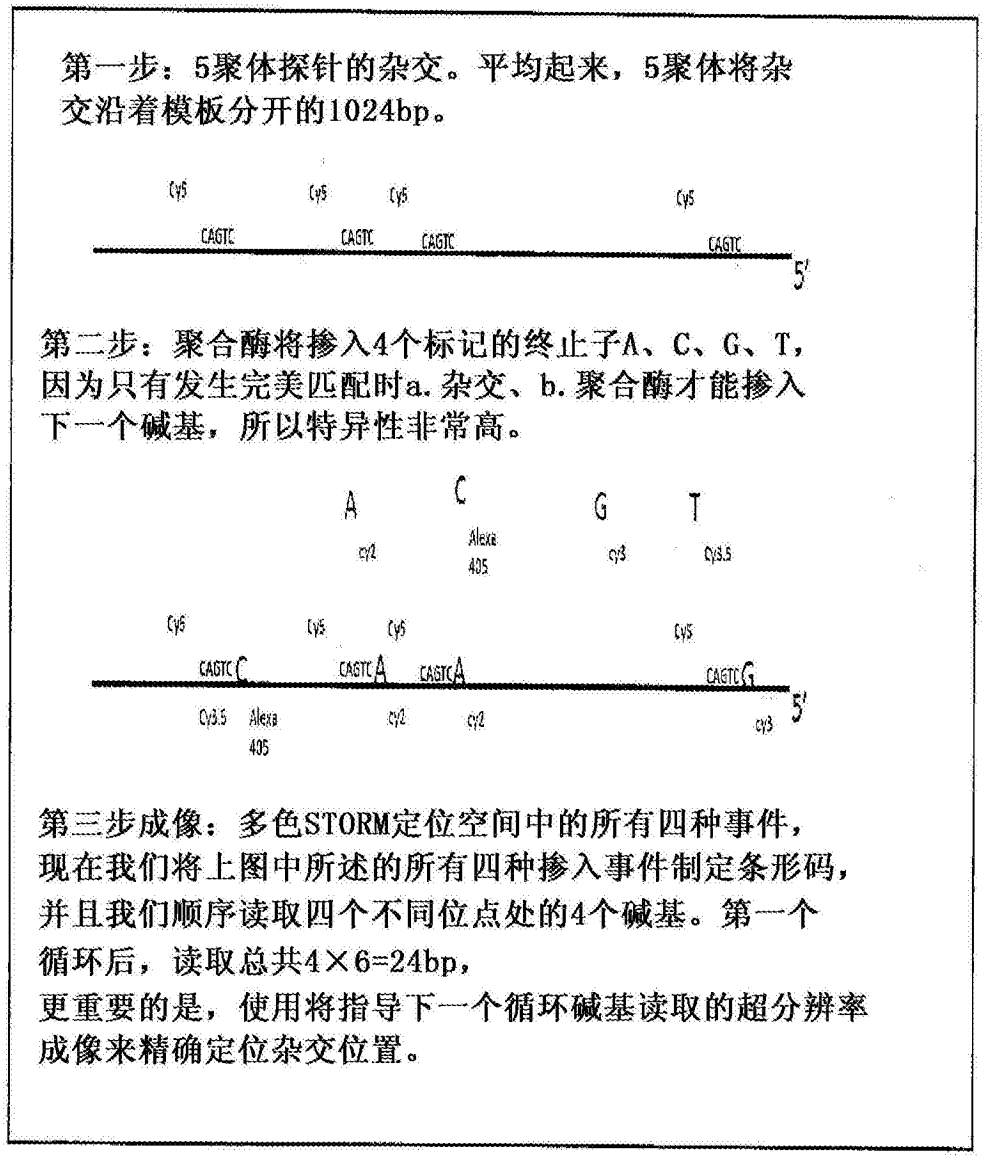
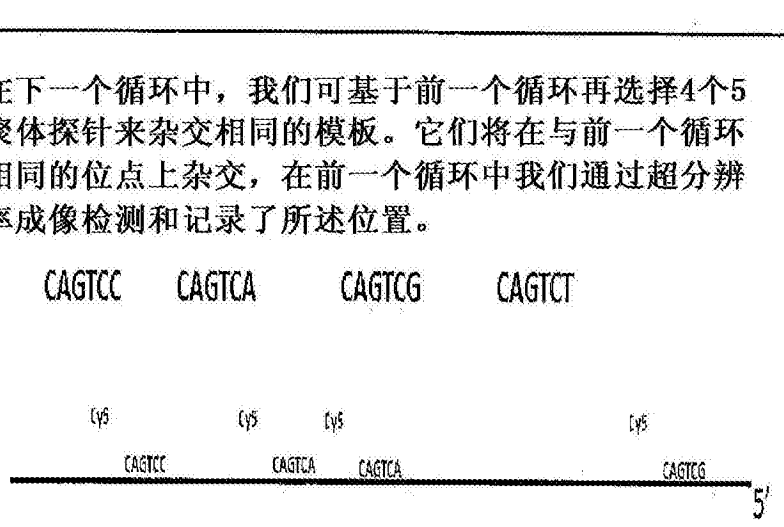
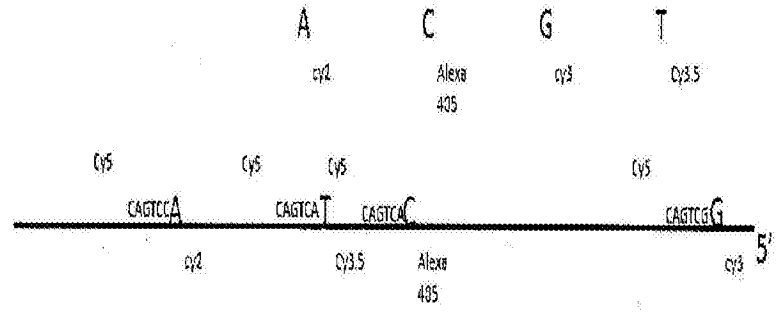

第一步：5聚体探针的杂交。平均起来，5聚体将杂交沿着模板分开的1024bp。
Cy5
CAGTC
5'
第二步：聚合酶将掺入4个标记的终止子A、C、G、T，因为只有发生完美匹配时a.杂交、b.聚合酶才能掺入下一个碱基，所以特异性非常高。
A
C
G
T
cy2
Alexa 405
cy3
Cy3.5
第三步成像：多色STORM定位空间中的所有四种事件，现在我们将上图中所述的所有四种掺入事件制定条形码，并且我们顺序读取四个不同位点处的4个碱基。第一个循环后，读取总共4×6=24bp，
更重要的是，使用将指导下一个循环碱基读取的超分辨率成像来精确定位杂交位置。

图4

在下一个循环中，我们可基于前一个循环再选择4个5聚体探针来杂交相同的模板。它们将在与前一个循环相同的位点上杂交，在前一个循环中我们通过超分辨率成像检测和记录了所述位置。



重复4个标记终止子A、C、G、T的掺入步骤。

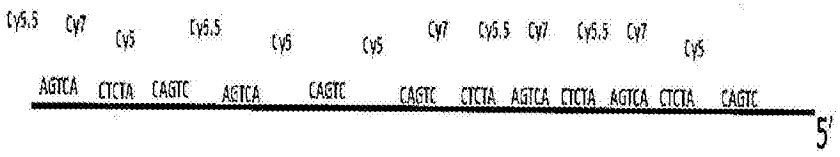


两个循环后，我们读取了7×4=28个碱基。这可被容易地扩展至5-10个循环，其中共有4×10=40个碱基。

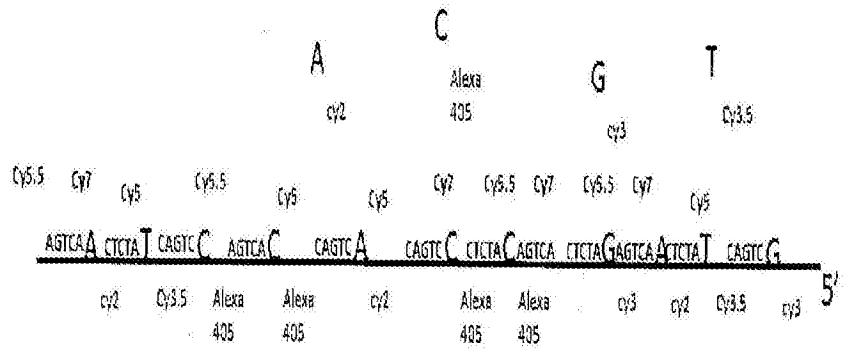
图5

为了增加复杂性和产量，我们可以用不同颜色的探针进行多重杂交。

在下面的实施例中，我们以四种不同颜色的杂交探针开始。



假设有12个杂交事件，在掺入后，我们读取了12个碱基，且总计12×5=60bp将被读取。



在这一情形中，人们需要对挑选用于杂交的探针和顺序的聚合酶掺入进行平衡以使多重性的能力最大化。

图6

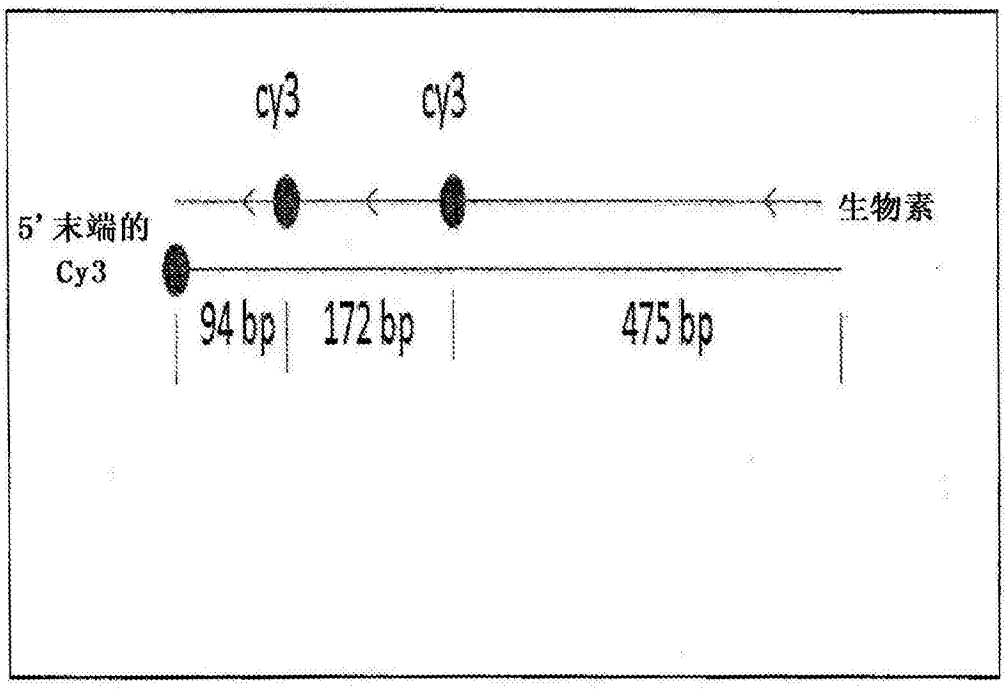


图7A

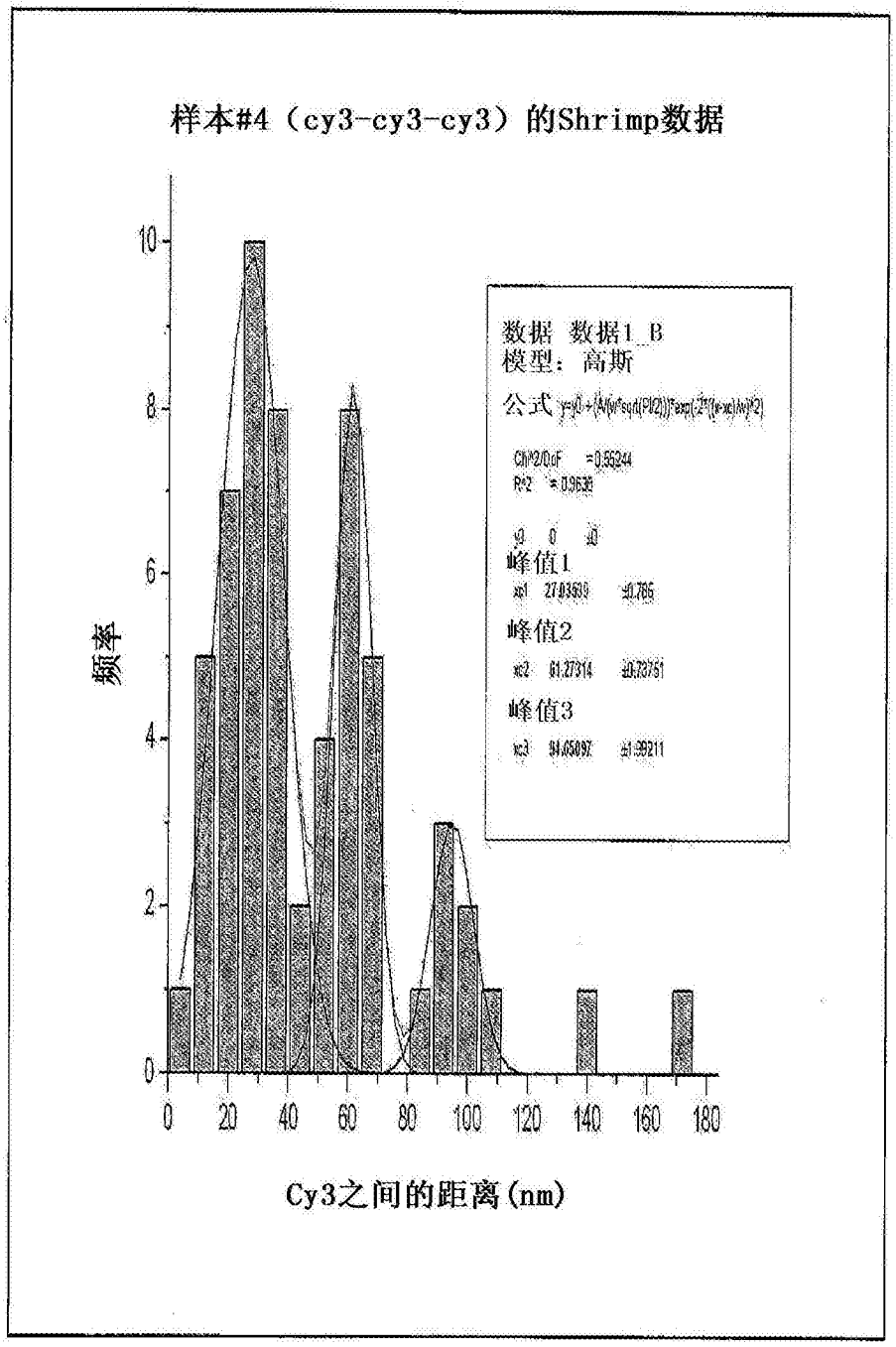

样本#4（cy3-cy3-cy3）的Shrimp数据
频率
Cy3之间的距离(nm)
数据 数据1_B
模型：高斯
公式
峰值1
峰值2
峰值3
0
2
4
6
8
10
0
20
40
60
80
100
120
140
160
180

图7B

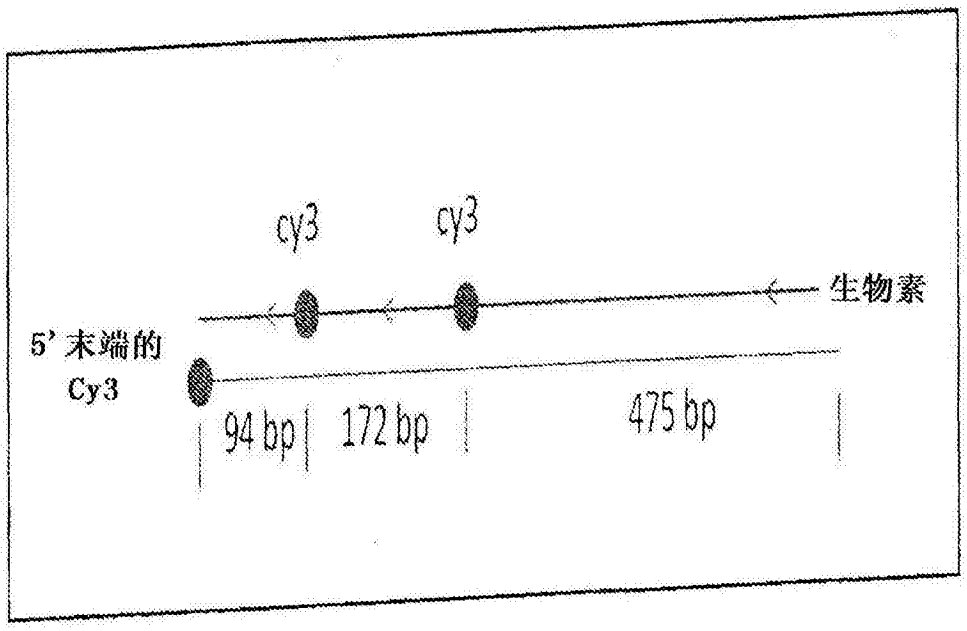


图8A

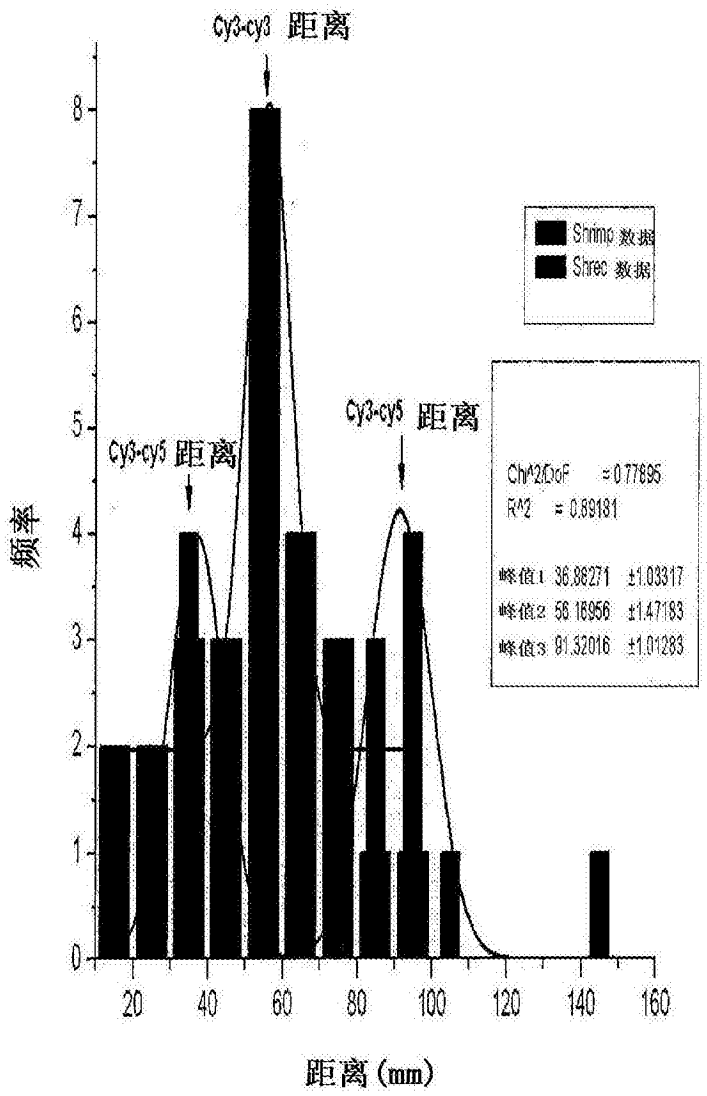
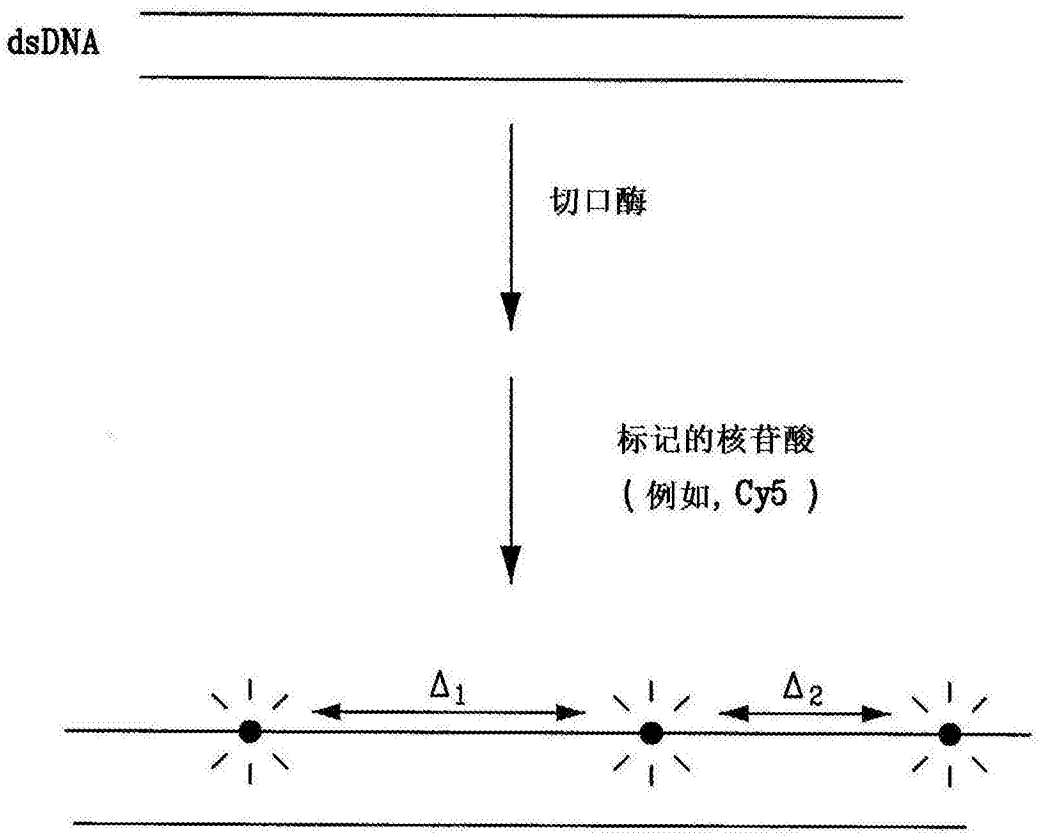

Cy3-cy3 距离
Cy3-cy5 距离
Cy3-cy5 距离
Shrimp 数据
Shrec 数据
Chi^2/DoF = 0.77895
R^2 = 0.89181
峰值1 36.88271 ±1.03317
峰值2 58.16956 ±1.47183
峰值3 91.32016 ±1.01283
频率
距离(mm)
0
1
2
3
4
5
6
7
8
20
40
60
80
100
120
140
160

图8B



Δ_1 和 Δ_2 可提供关于存在于dsDNA样本中的外显子或拷贝数变异的信息

图9

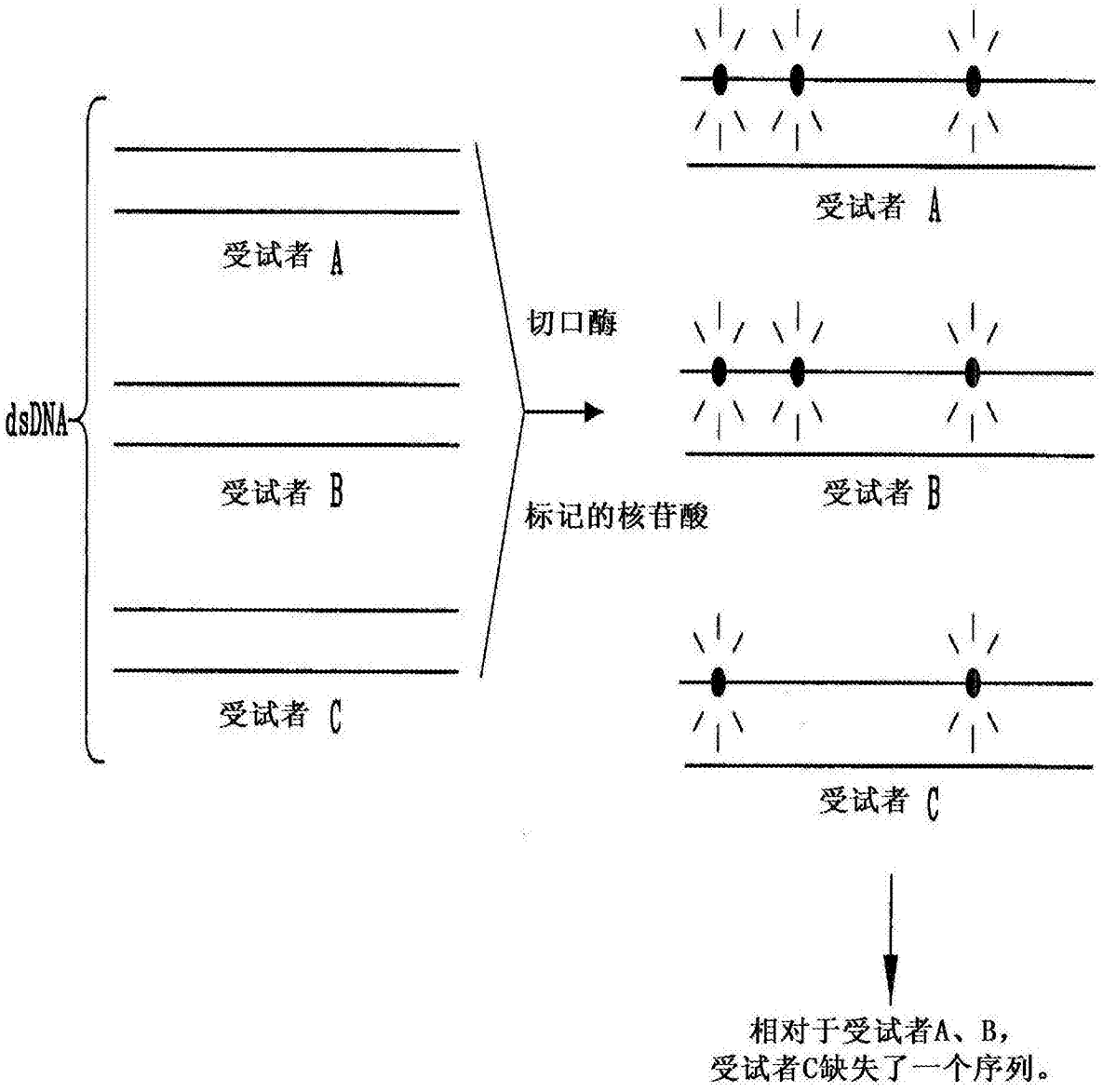

dsDNA
受试者 A
受试者 B
受试者 C
切口酶
标记的核苷酸
受试者 A
受试者 B
受试者 C
相对于受试者A、B，
受试者C缺失了一个序列。

图10

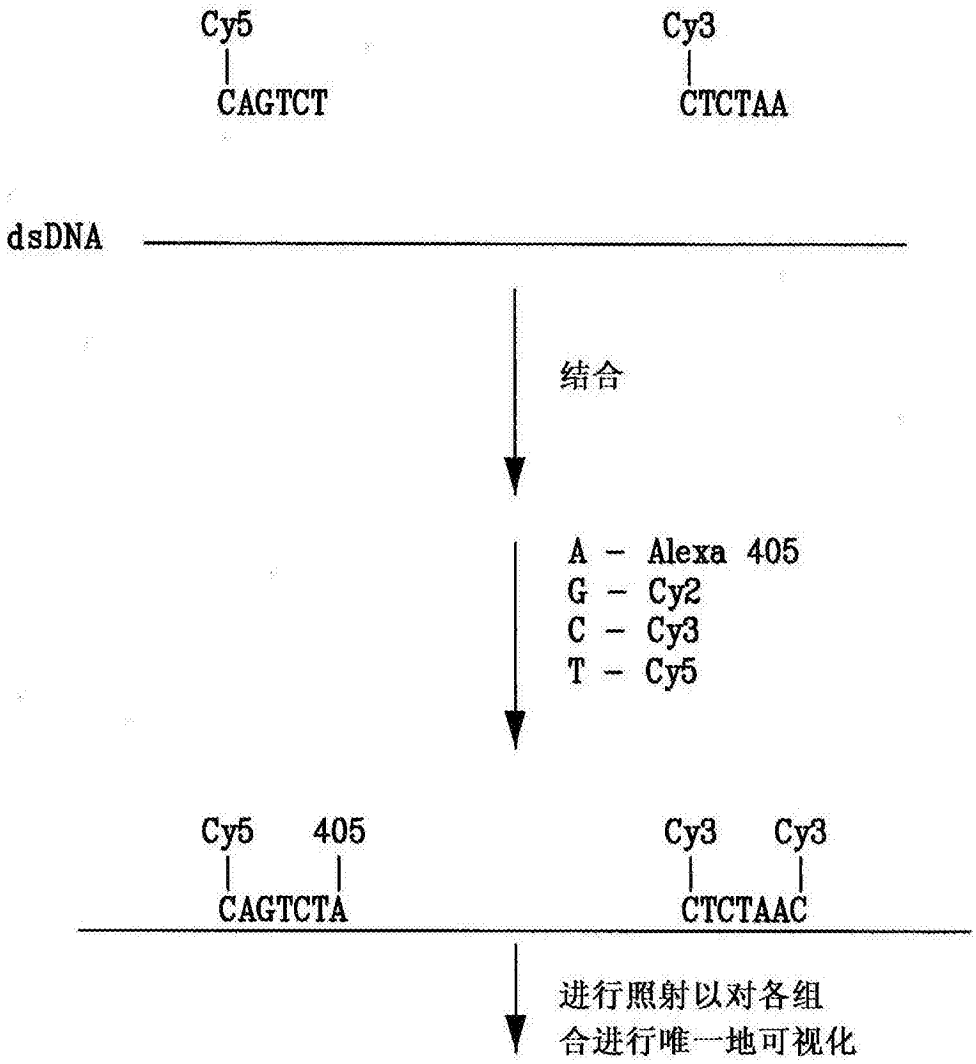

Cy5
CAGTCT
Cy3
CTCTAA
dsDNA
结合
A – Alexa 405
G – Cy2
C – Cy3
T – Cy5
Cy5 405
CAGTCTA
Cy3 Cy3
CTCTAAC
进行照射以对各组
合进行唯一地可视化

图11

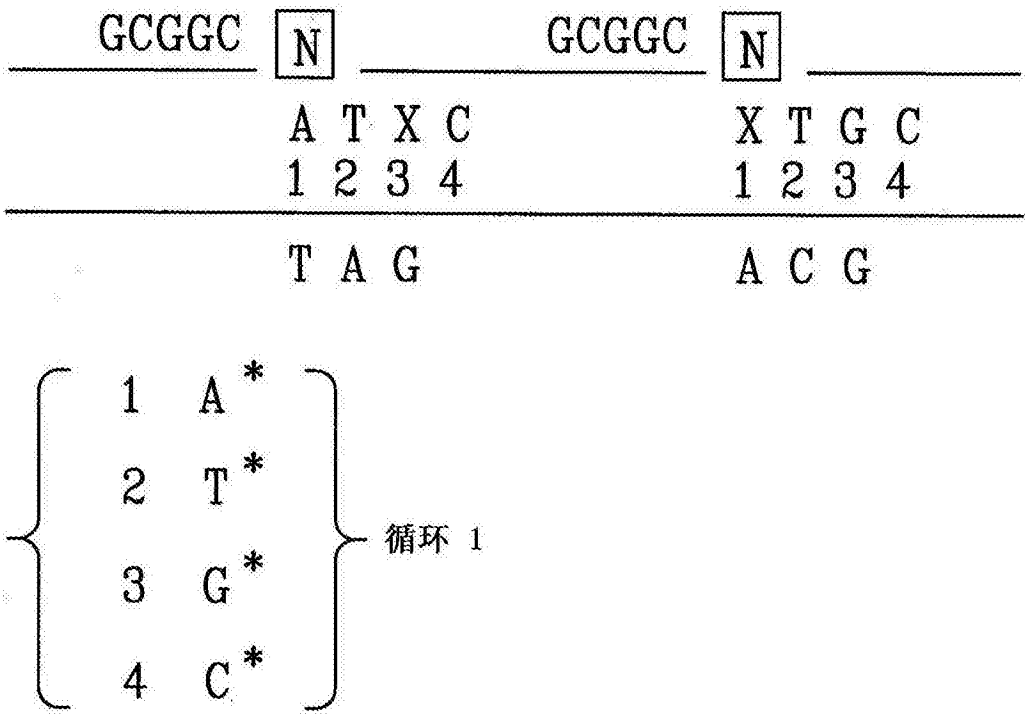


图12

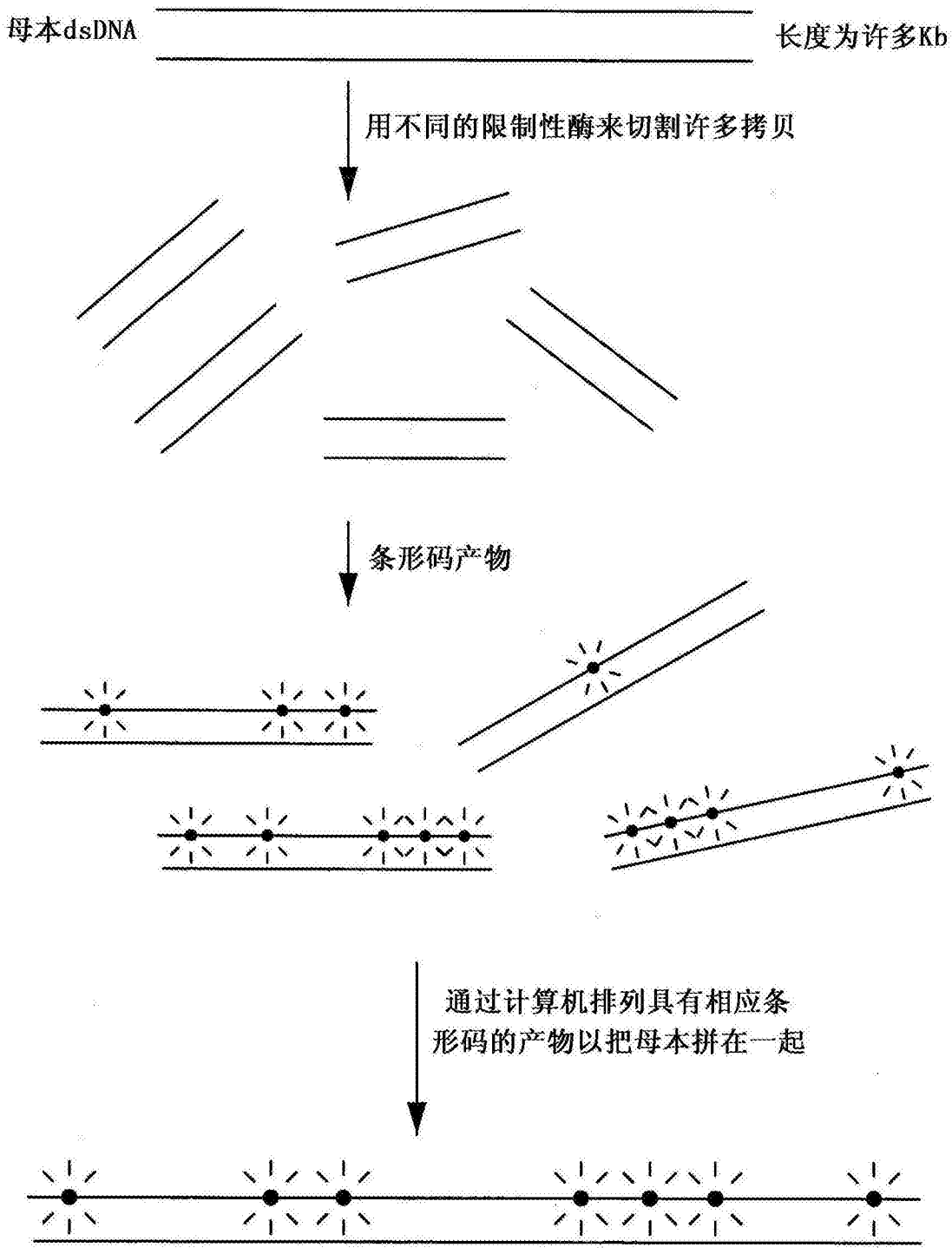

母本dsDNA
长度为许多Kb
用不同的限制性酶来切割许多拷贝
条形码产物
通过计算机排列具有相应条
形码的产物以把母本拼在一起

图13