

对多核苷酸进行测序的方法

发明领域

5 本发明涉及对多核苷酸进行测序的方法，其中通过相同的发光信号检测不同核苷酸的依次掺入，从而实现多核苷酸序列的测定。

发明背景

1977 年，桑格发明了双脱氧末端终止测序法，成为第一代测序技术的代表。2001 年，依托第一代测序技术，完成了人类基因组草图。桑格测序法具有实验操作简单、
10 结果直观准确和实验周期短等特点，在对检测结果时效性要求较高的临床基因突变检测以及基因分型等领域有着广泛的应用。然而，桑格测序法通量小、成本高，限制了其在大规模基因测序中的应用。

为克服桑格测序法的缺点，第二代测序技术应运而生。与第一代测序技术相比，第二代测序技术具有通量大、成本低、自动化程度高等优点，适合于大规模测序。目前已开发的第二代测序技术主要涉及边连接边测序（sequencing by ligation, SBL）技术和边合成边测序（sequencing by synthesis, SBS）技术。这些测序技术的典型实例包括 Roche 454 测序法、Applied Biosystems 公司开发的 SOLiD 测序法、Complete Genomics 自主开发的联合探针锚定连接法（cPAL）和华大基因开发的联合探针锚定合成法（cPAS）、Illumina 公司和 Solexa technology 公司合作开发的 Illumina 测序法等等。测序检测方法主要有电化学法、光信号检测法等，其中，较为主流的检测方法为光信号检测。为了实现 4 种碱基（A、T/U、C 和 G）的鉴别和区分，需要使用 4 种荧光染料来分别标记 4 种碱基。目前也有报道使用 2 种荧光染料标记 4 种碱基，通过 2 种荧光染料的不同组合来实现 4 种碱基的鉴别和区分。而 Roche 454 测序法运用自发荧光原理，将 dNTP 合成到待测序列产生的焦磷酸转变为 ATP，利用生成的
20 ATP 和荧光素酶共同氧化荧光素产生荧光，通过检测荧光信号的有无及强弱，区分 4 种碱基及合成碱基的个数。第二代测序技术由于硬件要求，仪器普遍较为庞大，不利于携带及搬运。

目前测序技术已发展至第三代，其克服了第二代测序技术仪器庞大的缺点，如 Oxford Nanopore 的测序仪因为其测序原理的不同使其测序仪体积大大缩小，甚至可以
30 被携带至太空进行测序实验。但是目前第三代测序技术错误率较高，限制了其大规模推广。

以 Illumina、Complete Genomics 和华大基因开发的测序仪为例，用 4 种荧光染料分别标记 4 种碱基，通过激光激发，采集不同的荧光信号，以区分不同碱基。参见例如 Sara Goodwin, John D. McPherson and W. Richard McCombie, Coming of age: ten years of next-generation sequencing technologies. *Nature reviews*, 2016, 17: 333-351c。
35

Illumina 公司开发的 NextSeq 测序系统和 Mini-Seq 测序系统，以及华大基因 BGISEQ-50 测序系统均使用 2 种荧光染料标记 4 种碱基，通过 2 种荧光染料的不同组合来实现 4 种碱基的鉴别和区分。如，通过用第一荧光染料标记碱基 A，用第二荧光染料标记碱基 G，用第一和第二荧光染料同时标记碱基 C，且不对碱基 T 进行标记，从而区分四种碱基。参见例如美国专利 US 9453258 B2。
40

在 Roche 454 测序法中，依次分别通入一种脱氧核糖核苷酸（dNTP），若该 dNTP 能与待测序列配对，则在 dNTP 合成后释放焦磷酸，焦磷酸与测序反应体系中的 ATP 硫酸化酶反应生成 ATP，生成的 ATP 再与体系中的荧光素酶共同氧化荧光素发出荧光，荧光信号被检测器捕捉，经计算机分析转换为测序结果。参见例如 Martin Kircher and Janet Kelso. High-throughput DNA sequencing – concepts and limitations.
45

Bioessays, 2010, 32: 524-536。

Ion torrent 测序系统与 Roche 454 测序法类似,依次分别通入一种脱氧核糖核苷酸(dNTP),若该dNTP能与待测序列配对,则在dNTP合成后释放氢离子,产生的氢离子改变反应体系的pH值,集成在测序芯片上的电器元件将pH值变化转变为电信号传输至计算机,经计算机分析转换为测序结果。参见例如 Sara Goodwin, John D. McPherson and W. Richard McCombie, Coming of age: ten years of next-generation sequencing technologies. *Nature reviews*, 2016, 17: 333-351。

这些技术存在以下缺陷:

1. 利用4种荧光染料标记4种碱基,为了区分不同的荧光信号,测序设备至少配备2种单色激发光源和2个相机,这导致测序装置的制造成本昂贵且体积巨大。

2. 利用2种荧光染料标记4种碱基,相比于利用4种荧光染料标记,虽然降低了设备制造成本,缩小了设备体积,但是实验证明,由于该方案中其中一种dNTP同时标记了两种荧光,激光同时激发两种荧光发光,随着测序长度增加,模板状态变差(目前存在的二代测序技术,无论何种原理,都存在随着测序读长增加,测序质量变差的情况),导致标记的两种荧光不能被平衡激发(其中一种荧光发光强度明显高于另一种)使这种融合荧光的dNTP信号趋向于和单种荧光标记的dNTP信号揉合,导致无法区分不同的dNTP,因此其测序质量明显低于用4种荧光染料标记的检测方法。

3. 不论是利用4种或者2种荧光染料标记4种碱基的检测方法,不同荧光之间均可能存在信号互相干扰,影响测序质量。

4. Roche 测序法与 Ion torrent 测序法,虽然不需要激发光源与相机等设备,但是其使用的脱氧核糖核苷酸为天然状态,当遇到待测序列具有重复的碱基排布时,如5'-ATTG-3',与碱基排布为5'-ATG-3'的序列相比,只能通过信号强弱进行区分(理论上序列5'-ATTG-3'信号值约为5'-ATG-3'的3倍),这种判别方法受测序条件干扰较大,也不易控制,尤其在读长较长时,很难将两者区分开。

因此,本领域仍然存在对成本更低、效果更好的测序方法的需要。

发明内容

本发明涉及对多核苷酸进行测序的方法,其中通过相同的发光信号检测不同核苷酸的依次掺入,从而实现多核苷酸序列的测定。

在一个方面,本发明涉及用于确定靶多核苷酸的序列的方法,其包括:

(a)提供靶多核苷酸,

(b)使所述靶多核苷酸与引物接触,以使所述引物杂交至所述靶多核苷酸,从而形成靶多核苷酸与引物的部分双链体,

(c)在允许聚合酶进行核苷酸聚合反应的条件下,使所述部分双链体与聚合酶和核苷酸接触,以使得所述核苷酸掺入到所述引物上,

其中所述核苷酸选自以下的一种或多种:第一核苷酸、第二核苷酸、第三核苷酸和第四核苷酸,其中所述第一核苷酸包含经第一标记物标记的第一核苷酸和任选未经标记的第一核苷酸,所述第二核苷酸包含经第二标记物标记的第二核苷酸和任选未经标记的第二核苷酸,所述第三核苷酸选自:(1)经第一标记物标记的第三核苷酸和经第二标记物标记的第三核苷酸,或(2)经第一标记物和第二标记物同时标记的第三核苷酸,所述第四核苷酸包含未经标记的第四核苷酸,

其中所述核苷酸各自的核糖或脱氧核糖部分包含通过2'或3'氧原子附接的保护基团,

(d)检测步骤(c)的所述部分双链体上第一标记物的存在,

(e)检测步骤(c)的所述部分双链体上第二标记物的存在,

(f)任选地去除在步骤(c)的所述部分双链体中掺入的核苷酸上的保护基团和标记物,

(g)任选地重复步骤(c)-(f)一次或多次,从而获得所述靶多核苷酸的序列信息,其中所述第一标记物和所述第二标记物的存在通过相同的发光信号来检测。

在具体的实施方案中,所述第一标记物是发光标记物。

在具体的实施方案中,步骤(d)包括使步骤(c)的所述部分双链体与特异性结合所述第一标记物的经发光标记物标记的配体接触,随后检测所述部分双链体上所述发光标记物的存在。

在具体的实施方案中,在去除步骤(c)的所述部分双链体中掺入的核苷酸上的保护基团和标记物时一起去除所述配体。

在具体的实施方案中,步骤(e)包括使步骤(c)的所述部分双链体与特异性结合所述第二标记物的经发光标记物标记的配体接触,随后检测所述部分双链体上所述发光标记物的存在。

在具体的实施方案中,步骤(e)在步骤(d)之后进行。

在具体的实施方案中,所述发光标记物是相同的发光标记物。

在具体的实施方案中,所述发光标记物是荧光标记物,例如是荧光团,例如选自香豆素、AlexaFluor、Bodipy、荧光素、四甲基罗丹明、Cy5、Cy3、得克萨斯红及其衍生物。

在具体的实施方案中,所述第一核苷酸中所述经第一标记物标记的第一核苷酸和所述未经标记的第一核苷酸的比率为 4:1 至 3:2。

在具体的实施方案中,所述第二核苷酸中所述经第二标记物标记的第二核苷酸和所述未经标记的第二核苷酸的比率为 4:1 至 3:2。

在其他方面,本发明还涉及用于对多核苷酸进行测序的试剂盒,其包含:(a)选自以下的一种或多种核苷酸:第一核苷酸、第二核苷酸、第三核苷酸和第四核苷酸,其中所述第一核苷酸包含经第一标记物标记的第一核苷酸和任选未经标记的第一核苷酸,所述第二核苷酸包含经第二标记物标记的第二核苷酸和任选未经标记的第二核苷酸,所述第三核苷酸选自:(1)经第一标记物标记的第三核苷酸和经第二标记物标记的第三核苷酸,或(2)经第一标记物和第二标记物同时标记的第三核苷酸,所述第四核苷酸包含未经标记的第四核苷酸;和(b)它们的包装材料,其中所述核苷酸各自的核糖或脱氧核糖部分包含通过 2'或 3'氧原子附接的保护基团。

在具体的实施方案中,所述第一标记物是发光标记物。

在具体的实施方案中,试剂盒还包含特异性结合所述第一标记物的经发光标记物标记的配体。

在具体的实施方案中,试剂盒还包含特异性结合所述第二标记物的经发光标记物标记的配体。

在具体的实施方案中,所述发光标记物是相同的发光标记物。

在具体的实施方案中,其中所述发光标记物是荧光标记物,例如是荧光团,例如选自香豆素、AlexaFluor、Bodipy、荧光素、四甲基罗丹明、Cy5、Cy3、得克萨斯红及其衍生物。

在具体的实施方案中,试剂盒还包含酶和适合所述酶起作用的缓冲液。

附图说明

图 1 显示了实施例 1 中对大肠杆菌条形码序列进行测序的第 1 位碱基信号提取图。

图 2 显示了实施例 1 中对大肠杆菌条形码序列进行测序的第 10 位碱基信号提取图。

图 3 显示了实施例 2 中对大肠杆菌条形码序列进行测序的第 1 位碱基信号提取图。

图 4 显示了实施例 2 中对大肠杆菌条形码序列进行测序的第 50 位碱基信号提取图。

图 5 显示了实施例 1 中未添加不带标记核苷酸实验的第一位碱基的信号提取图。

图 6 显示了实施例 2 中未添加不带标记核苷酸实验的第一位碱基的信号提取图。

发明详述

除非另外定义，否则本文使用的所有技术和科学术语具有与本发明所属领域的普通技术人员通常理解的含义相同的含义。本文提及的所有专利、申请和其他出版物均通过引用整体并入本文。如果本文中提出的定义与通过引用并入本文的专利、申请和其他出版物中所述的定义相抵触或不一致，则以本文所述的定义为准。

如本文所用，术语“多核苷酸”是指脱氧核糖核酸(DNA)、核糖核酸(RNA)或其类似物。多核苷酸可以是单链的、双链的或含有单链和双链序列两者。多核苷酸分子可以来源于双链 DNA(dsDNA)形式(例如，基因组 DNA、PCR 和扩增产物等)，或者可以来源于单链形式的 DNA(ssDNA)或 RNA 并且其可以转化为 dsDNA 形式，并且反之亦然。多核苷酸分子的准确序列可以是已知的或未知的。以下是多核苷酸的示例性实例：基因或基因片段(例如，探针、引物、EST 或 SAGE 标签)、基因组 DNA、基因组 DNA 片段、外显子、内含子、信使 RNA(mRNA)、转运 RNA、核糖体 RNA、核糖酶、cDNA、重组多核苷酸、合成多核苷酸、分枝多核苷酸、质粒、载体、任何序列的分离的 DNA、任何序列的分离的 RNA、任何上述序列的核酸探针、引物或扩增拷贝。

多核苷酸可以包括核苷酸或核苷酸类似物。核苷酸通常含有糖(如核糖或脱氧核糖)、碱基和至少一个磷酸基。核苷酸可以是无碱基的(即，缺少碱基)。核苷酸包括脱氧核糖核苷酸、修饰的脱氧核糖核苷酸、核糖核苷酸、修饰的核糖核苷酸、肽核苷酸、修饰的肽核苷酸、修饰磷酸盐糖主链核苷酸及其混合物。核苷酸的实例包括(例如)腺苷一磷酸(AMP)、腺苷二磷酸(ADP)、腺苷三磷酸(ATP)、胸苷一磷酸(TMP)、胸苷二磷酸(TDP)、胸苷三磷酸(TTP)、胞苷酸(CMP)、胞苷二磷酸(CDP)、胞苷三磷酸(CTP)、鸟苷一磷酸(GMP)、鸟苷二磷酸(GDP)、鸟苷三磷酸(GTP)、尿苷一磷酸(UMP)、尿苷二磷酸(UDP)、尿苷三磷酸(UTP)、脱氧腺苷酸(dAMP)、脱氧腺苷二磷酸(dADP)、脱氧腺苷三磷酸(dATP)、脱氧胸腺嘧啶核苷一磷酸(dTMP)、脱氧胸腺嘧啶核苷二磷酸(dTDP)、脱氧胸苷三磷酸(dTTP)、去氧胞二磷(dCDP)、脱氧胞苷三磷酸(dCTP)、脱氧鸟苷一磷酸(dGMP)、脱氧鸟苷二磷酸(dGDP)、脱氧鸟苷三磷酸(dGTP)、脱氧尿苷一磷酸(dUMP)、脱氧尿苷二磷酸(dUDP)和脱氧尿苷三磷酸(dUTP)。还可以在本文所述的方法中使用包含修饰的碱基的核苷酸类似物。无论是具有天然主链还是类似结构，可以包含在多核苷酸中的示例性修饰的碱基包括(例如)肌苷、黄嘌呤(xanthine)、次黄嘌呤(hypoxanthine)、异胞嘧啶、异鸟嘌呤、2-氨基嘌呤、5-甲基胞嘧啶、5-羟甲基胞嘧啶、2-氨基腺嘌呤、6-甲基腺嘌呤、6-甲基鸟嘌呤、2-丙基鸟嘌呤、2-丙基腺嘌呤、2-硫脲嘧啶、2-硫胸腺嘧啶、2-硫胞嘧啶、15-卤代脲嘧啶、15-卤代胞嘧啶、5-丙炔基尿嘧啶、5-丙炔基胞嘧啶、6-偶氮尿嘧啶、6-偶氮胞嘧啶、6-偶氮胸腺嘧啶、5-尿嘧啶、4-硫尿嘧啶、8-卤代腺嘌呤或鸟嘌呤、8-氨基腺嘌呤或鸟嘌呤、8-硫腺嘌呤或鸟嘌呤、8-硫烷基腺嘌呤或鸟嘌呤、8-羟基腺嘌呤或鸟嘌呤、5-卤素取代的尿嘧啶或胞嘧啶、7-甲基鸟嘌呤、7-甲基腺嘌呤、8-氮杂鸟嘌呤、8-氮杂腺嘌呤、7-去氮鸟嘌呤、7-去氮腺嘌呤、3-去氮鸟嘌呤、3-去氮腺嘌呤等。如本领域中已知的，某些核苷酸类似物不能引入多核苷酸，例如，核苷酸类似物，如腺苷 5'-磷酸硫酸。

通常而言，核苷酸包括核苷酸 A、C、G、T 或 U。如本文所用，术语“核苷酸 A”是指含有腺嘌呤(A)或其修饰物或类似物的核苷酸，例如 ATP、dATP。“核苷酸 G”是指含有鸟嘌呤(G)或其修饰物或类似物的核苷酸，例如 GTP、dGTP。“核苷酸 C”是指含有胞嘧啶(C)或其修饰物或类似物的核苷酸，例如 CTP、dCTP。“核苷酸 T”是指含

有胸腺嘧啶(T)或其修饰物或类似物的核苷酸,例如 TTP、dTTP。“核苷酸 U”是指含有尿嘧啶(U)或其修饰物或类似物的核苷酸,例如 UTP、dUTP。

核苷酸的标记

本发明涉及用不同的标记物单独或组合地标记核苷酸,从而使得可以区分不同的核苷酸,其中所述不同的标记物可以通过相同的发光信号检测。

在具体的实施方案中,通过将不同的标记物与各自的用可产生相同的发光信号的发光标记物标记的配体特异性结合,来实现通过相同的发光信号检测不同的标记物。在优选的实施方案中,所述可产生相同的发光信号的发光标记物是相同的发光标记物。

如本文所用,用于标记核苷酸的所述标记物和与其特异性结合的配体可以是任何能够彼此特异性结合的分子,其结合对在本文中称为抗配体对。抗配体对的成员之间的结合可以是非共价的。抗配体对还不必限于成对的单分子。例如,单个配体可以通过两种或更多种抗配体的协同作用结合。抗配体对的成员之间的结合导致结合复合物的形成,有时称为配体/抗配体复合物或简单地作为配体/抗配体。示例性的抗配体对包括但不限于:(a)与相应抗体或其结合部分或片段组合的半抗原或抗原性化合物,例如地高辛-地高辛抗体, N3G-N3G 抗体, FITC-FITC 抗体;(b)核酸适配体和蛋白质;(c)非免疫结合对(例如生物素-抗生物素蛋白、生物素-链霉亲和素、生物素-中性抗生物素蛋白);(d)激素-激素结合蛋白;(e)受体-受体激动剂或拮抗剂;(f)凝集素-碳水化合物;(g)酶-酶辅因子;(h)酶-酶抑制剂;和(i)能够形成核酸双链体的互补的寡核苷酸或多核苷酸对。

在另一个具体的实施方案中,所述不同的标记物之一可以是发光标记物,从而可以直接检测。而其他的标记物则仍然通过与各自的用可产生相同的发光信号的发光标记物标记的配体特异性结合来检测。在优选的实施方案中,与所述不同的标记物相关的发光标记物是相同的发光标记物。

如本文所用,术语“发光标记物”是指在被合适的激发波长激发时,能够以特定的发射波长发射荧光的任何物质。这样的发光标记物可以是例如荧光团,例如选自香豆素、AlexaFluor、Bodipy、荧光素、四甲基罗丹明、吩噻嗪、吖啶、Cy5、Cy3、AF532、得克萨斯红及其衍生物。

多核苷酸的测序

本发明的用不同的标记物单独或组合地标记的核苷酸可用于各种核酸测序方法。优选地,本发明的用不同的标记物单独或组合地标记的核苷酸适用于合成法测序。如本文所用的合成法测序是本领域熟知的各种合成法测序方法。基本地,合成法测序涉及首先将被测序的核酸分子与测序引物杂交,随后在聚合酶的存在下,以被测序的核酸分子为模板在测序引物的 3'端聚合如本文所述的经标记的核苷酸。聚合之后,通过检测所述标记来鉴定该经标记的核苷酸。在从经标记的核苷酸上除去标记(即如本文所述的化学发光标记物)之后,开始下一个聚合测序循环。

此外核酸测序方法还可以用本文所述的核苷酸进行公开于美国专利号 5302509 中的方法。

用于测定靶多核苷酸序列的方法可以这样进行:使靶多核苷酸序列变性,使靶多核苷酸分别与不同的核苷酸接触,以便形成所述靶核苷酸的互补体,并且检测所述核苷酸的掺入。所述方法利用了聚合,使得聚合酶通过掺入互补于所述靶的正确的核苷酸,以延伸所述互补链。所述聚合反应还需要特殊引物来启动聚合作用。

对每一轮反应来说,所述经标记的核苷酸的掺入是通过聚合酶进行的,并随后测定所述掺入事件。存在很多不同的聚合酶,并且对本领域普通技术人员来说容易确定最适合的聚合酶。优选的酶包括 DNA 聚合酶 I、Klenow 片段、DNA 聚合酶 III、T4 或 T7 DNA 聚合酶、Taq 聚合酶或 vent 聚合酶。还可以使用通过工程方法改造成

具有特定性质的聚合酶。

所述测序方法优选对排列在固体支持物上的靶多核苷酸进行。可以通过接头分子将多个靶多核苷酸固定在所述固体支持物上，或者可以连接在诸如微球体的颗粒上，所述颗粒还可以连接在固体支持材料上。

5 可以通过多种方法将所述多核苷酸连接在所述固体支持物上，包括使用生物素-链亲和素相互作用。用于将多核苷酸固定在固体支持物上的方法为本领域所公知，并且包括石板印刷技术以及将每一种多核苷酸点样在固体支持物的特定位置上。合适的固体支持物为本领域所公知，并且包括玻璃载玻片和珠、陶瓷和硅表面和塑料材料。所述支持物通常是平面，尽管也可以使用微珠(微球体)，并且还可以通过已知方法将
10 后者连接在其他固体支持物上。所述微球体可以具有任何合适的大小，其直径通常为10-100 纳米。在优选实施方案中，将所述多核苷酸直接连接在平面上，优选连接在平的玻璃表面上。连接优选通过共价键的形式进行。所使用的阵列优选是单分子阵列，它包括位于独特的光学可分辨区域的多核苷酸，例如在国际申请号 WO00/06770 中所描述的。

15 进行聚合的必须条件对本领域技术人员来说是熟知的。为了进行所述聚合酶反应，通常首先必须使引物序列与所述靶多核苷酸退火，所述引物序列是由所述聚合酶识别的，并且起着所述互补链随后延伸的起始位点的作用。所述引物序列可以相对所述靶多核苷酸作为独立的成分添加。另外，所述引物和靶多核苷酸可以分别是一个单链分子的一部分，由所述引物部分与所述靶的一部分形成分子内双链体，即发卡环结构。
20 可以在所述分子的任何位点，将该结构固定在所述固体支持物上。进行所述聚合酶反应所必需的其他条件，对本领域技术人员来说是熟知的，这些条件包括温度、pH、缓冲液组成。

随后，使本发明的经标记的核苷酸与所述靶多核苷酸接触，以便能够进行聚合。所述核苷酸可以依次添加，即分别添加每一种类型的核苷酸(A, C, G 或 T/U)，或同时添加。
25

使所述聚合步骤进行足以掺入一个核苷酸的时间。

然后除去未掺入的核苷酸，例如，通过对所述阵列实施洗涤步骤，并且随后可以进行对掺入标记的检测。

检测可以通过常规方法进行。例如，检测荧光标记或信号的方式是本领域熟知的。
30 例如，可以通过检测荧光的波长的装置来实现。这样的装置是本领域熟知的。例如，这样的装置可以是共焦扫描显微镜，其用激光扫描固体支持物的表面，以便使直接结合被测序的核酸分子上的荧光团成像。另外，可以例如用灵敏的 2-D 探测器，如电荷偶连的探测器(CCD)观察所产生的每一种信号。还可以例如使用诸如扫描近场光学显微方法(SNOM)的其他技术。

35 在检测之后，可以用合适条件除去所述标记。

本发明的经标记的核苷酸的使用并不局限于 DNA 测序技术，还可用本发明的核苷酸实施包括多核苷酸合成，DNA 杂交分析，和单核苷酸多态性研究的其他形式。涉及到核苷酸和酶之间的相互作用的任何技术，都可以利用本发明的分子。例如，可以将所述分子用作逆转录酶或末端转移酶的底物。

40 在具体的实施方案中，本发明的经标记的核苷酸还具有 3'保护基团。在本发明的一些实施方案中，保护基团和标记物通常是 3'阻断的经标记的核苷酸上的两种不同的基团，但在另一些实施方案中，保护基团和标记物也可以是同一基团。

如本文所用，术语“保护基团”意指这样的基团，其阻止聚合酶（其将含有该基团的核苷酸掺入到正在合成的多核苷酸链上的）在将含有该基团的核苷酸掺入到正在合成的多核苷酸链上后继续催化另一核苷酸的掺入。这样的保护基团在本文中也称为 3'-OH 保护基团。包含这样的保护基团的核苷酸在本文中也称为 3'阻断的核苷酸。
45

保护基团可以是能够被添加到核苷酸上任何合适的基团，只要该保护基团能防止另外的核苷酸分子被加入至多核苷酸链中且同时在不破坏该多核苷酸链的情况下易于从核苷酸的糖部分除去。此外，经保护基团修饰的核苷酸需要耐受聚合酶或用于将该修饰的核苷酸掺入多核苷酸链内的其他适合的酶。因此，理想的保护基团表现出长期的稳定性，可被聚合酶高效地掺入，阻止核苷酸的二次掺入或进一步掺入，并且能够在不破坏多核苷酸结构的温和条件下优选地在水性条件下被除去。

现有技术已描述了多种符合上述描述的保护基团。例如，WO 91/06678 公开 3'-OH 保护基团包括酯和醚，-F，-NH₂，-OCH₃，-N₃，-OPO₃，-NHCOCH₃，2 硝基苯碳酸酯，2,4-次磺酰二硝基和四氢呋喃醚。Metzker 等人 (Nucleic Acids Research, 22(20):4259-4267, 1994) 公开了八种 3'-修饰的 2-脱氧核糖核苷 5'-三磷酸酯(3'-修饰的 dNTP)的合成和应用。WO2002/029003 描述了在聚合酶反应中使用烯丙基保护基团对 DNA 生长链上的 3'-OH 基团加帽。优选地，可以使用国际申请公开 WO2014139596 和 WO2004/018497 中报导的各种保护基团，包括例如 WO2014139596 的图 1A 中示例的那些保护基团和权利要求书中限定的那些 3'羟基保护基（即保护基团），和例如 WO2004/018497 的图 3 和 4 中示例的那些保护基团和权利要求书中限定的那些保护基团。上述参考文献均通过引用整体并入本文。

本领域技术人员将会理解如何将合适的保护基团连接在核糖环上，以便阻断与 3'-OH 的相互作用。所述保护基团可以直接连接在 3'位置上，或者可以连接在 2'位置上(所述保护基团具有足够的大小或电荷，以便阻断 3'位置上的相互作用)。另外，所述保护基团可以连接在 3'和 2'位置，并且可以被裂解，以便暴露出 3'-OH 基团。

在成功地将 3'阻断的核苷酸掺入核酸链后，测序方案需要除去保护基团以产生用于连续链合成的可用的 3'-OH 位点。如本文所用的可从经修饰的核苷酸上除去保护基团的试剂在很大程度上取决于所使用的保护基团。例如，从 3'羟基官能团除去酯保护基团通常通过碱水解来实现。除去保护基团的容易程度差异很大；通常，羰基碳上取代基的电负性越大，除去的容易度越大。例如，高电负性的三氟乙酸基团在甲醇中在 pH7 下能够从 3'羟基快速裂解 (Cramer 等人, 1963)，因此其在该 pH 下的聚合期间是不稳定的。苯氧基乙酸酯基团在少于 1 分钟内裂解，但是需要显著更高的 pH，例如用 NH-/甲醇实现 (Reese 和 Steward, 1968)。使用除碱水解以外的化学方法可以选择性地切割各种各样的羟基保护基团。通过用亲核试剂例如苯硫酚和硫代硫酸盐处理可迅速裂解 2,4-二硝基苯硫基 (Letsinger 等, 1964)。烯丙基醚通过用丙酮/水中的 Hg (II) 处理而裂解 (Gigg and Warren, 1968)。使用 Ag (I) 或 Hg (II) 在中性条件下除去四氢噻喃基醚 (Cohen and Steele, 1966; Cruse 等人, 1978)。光化学去阻断可以与可光化学裂解的保护基团一起使用。有几种保护基团可用于这种方法。使用邻硝基苄醚作为核糖核苷的 2'-羟基官能性的保护基团是已知且被证明的 (Ohtsuka 等, 1978)；其通过在 260nm 的照射进行除去。碳酸烷基邻硝基苄基碳酸酯保护基也通过在 pH7 的照射下被除去 (Cama and Christensen, 1978)。3'-OH 保护基团的酶解解阻断也是可能的。已经证明 T4 多核苷酸激酶可以将 3'-磷酸酯末端转化成 3'-羟基末端，然后可以用作 DNA 聚合酶 I 的引物 (Henner 等, 1983)。该 3'-磷酸酶活性用于除去含有磷酸酯作为保护基团的那些 dNTP 类似物的 3'保护基团。

可从 3'阻断的核苷酸上除去保护基团的其它试剂包括例如膦(例如三(羟甲基)膦 (THP))，其可以例如将含叠氮化物的 3'-OH 保护基团从核苷酸上除去（关于膦的此应用可参见例如 WO2014139596 中的记载，其全部内容通过引用并入本文）。可从 3'阻断的核苷酸上除去保护基团的其它试剂还包括例如 WO2004/018497 的说明书中第 114-116 页描述的用于除去作为 3'-OH 保护基团的 3'-烯丙基、3,4-二甲氧基苄氧基甲基或氟甲氧甲基的相应试剂。

在本发明的实施方案中，核苷酸的标记物优选在检测后与保护基团一起被除去。

在某些实施方案中，标记物可以被掺入保护基团，从而允许在将 3' 阻断的核苷酸掺入核酸链后其能够与保护基团一起被除去。

在其他实施方案中，标记物可以利用连接基团与保护基团分开地连接在核苷酸上。这样的标记物可以例如连接到核苷酸的嘌呤或嘧啶碱基上。在某些实施方案中，所用的连接基团是可裂解的。可裂解的连接基团的使用能确保所述标记可以在检测之后除去，这避免了与后续掺入的任何经标记的核苷酸的任何信号干扰。在另一些实施方案中，可以使用不可裂解的连接基团，因为在经标记的核苷酸掺入核酸链之后，不需要后续的核苷酸掺入，因此不需要将标记从核苷酸中除去。

在另外的实施方案中，标记物和/或连接基团可以具有足以发挥阻断其他核苷酸掺入到多核苷酸链上的大小或结构（也就是说，所述标记本身可用作保护基团）。所述阻断可能是由于空间位阻造成的，或者可能是由于大小、电荷和结构的组合造成的。

可裂解的连接基团为本领域所公知，并且可以采用常规化学方法，以便将连接基团连接在核苷酸碱基和标记物上。连接基团可以连接在核苷酸碱基的任何位置上，其前提是，仍然能进行 Watson-Crick 碱基配对。对于嘌呤碱基来说，如果所述连接基团是通过所述嘌呤或优选的脱氮嘌呤类似物的 7 号位置，通过 8-修饰的嘌呤，通过 N-6 修饰的腺嘌呤或 N-2 修饰的鸟嘌呤连接的话将是优选的。对于嘧啶来说，连接优选是通过胞嘧啶，胸腺嘧啶和尿嘧啶上的 5 号位置和胞苷上的 N-4 位置连接的。

使用术语“可裂解的连接基团”并非意味着需要除去整个连接基团（例如，从核苷酸碱基中除去）。当标记物与碱基相连接时，核苷裂解位点可位于连接基团上的位置，该位置能够确保在裂解后一部分的连接基团仍与所述核苷酸碱基保持连接。

合适的连接基团包括但不限于二硫连接基团，酸不稳定性连接基团（包括二烷基氧基苄基连接基团，Sieber 连接基团，吡啶连接基团，叔丁基 Sieber 连接基团），亲电可裂解的连接基团，亲核可裂解的连接基团，光可裂解的连接基团，在还原条件、氧化条件下裂解的连接基团，保险栓(safety-catch) 连接基团，以及通过消除机制进行裂解的连接基团。合适的连接基团可以用标准化学保护基团改良，正如在以下文献中所披露的：Greene & Wuts, *Protective Groups in Organic Synthesis*, John Wiley & Sons。Guillier 等披露了用于固相合成的其他合适的可裂解的连接基团 (Chem.Rev.100: 2092-2157, 2000)。

所述连接基团可以通过任何合适的方法裂解，包括接触酸，碱，亲核试剂，亲电试剂，自由基，金属，还原或氧化试剂，光照，温度，酶等，下文将示例性描述各种可裂解的连接基团的合适裂解方式。通常地，所述可裂解的连接基团可以在与所述保护基团相同的条件下裂解，以使得仅需要一次处理即可除去所述标记物和保护基团。

亲电裂解的连接基团典型地被质子所裂解，并包括对酸敏感的裂解。合适的亲电裂解的连接基团包括修饰的苄基系统，诸如三苯甲基、对炔氧基苄基酯和对炔氧基苄基酰胺。其他适合的连接基团包括叔丁氧羰基(Boc)基团和缩醛系统。为制备合适的连接分子，还可以考虑在硫缩醛或其他含硫保护基的裂解中使用诸如镍、银或汞的亲硫金属。亲核裂解的连接基团包括在水中不稳定的基团（即，能够在碱性 pH 值下简单地裂解），例如酯类，以及对非水性亲核试剂不稳定的基团。氟离子可用于裂解诸如三异丙基硅烷(TIPS)或叔丁基二甲基硅烷(TBDMS)的基团中的硅氧键。可光解的连接基团在糖化学中被广泛使用。优选地，激活裂解所需的光不影响修饰的核苷酸中的其他组分。例如，如果使用荧光团作为标记，优选地，该荧光团吸收与裂解所述连接分子所需的光不同波长的光。适合的连接基团包括那些基于 O-硝基苄基化合物和硝基藜芦基化合物的连接基团。也可以使用基于安息香化学的连接基团 (Lee 等人, J.Org.Chem.64:3454-3460, 1999)。已知多种对还原裂解敏感的连接基团。使用基于钨催化剂的催化氢化已用于裂解苄基和苄氧羰基基团。二硫键还原也为本领域所知。

基于氧化的方法为本领域所公知。这些方法包括对炔氧基苄基的氧化以及硫和硒连接基团的氧化。使用碘溶液(aqueous iodine)来使二硫化物和其他基于硫或硒的连接基团裂解也在本发明的范围内。安全拉手型连接基团(safety-catchlinker)为那些在两步中裂解的连接基团。在优选的系统中, 第一步是反应性亲核中心的产生, 随后的第二步涉及分子内环化, 这导致裂解。例如, 可以用肼或光化学方法处理乙酰丙酸酯连接来释放活性的胺, 然后所述胺被环化以使分子中其他位置的酯裂解(Burgess 等人, J.Org.Chem.62:5165-5168,1997)。也可以使用消除反应裂解连接基团。可以使用诸如苄基氧羰基和氰基乙基的基团的碱催化的消除以及烯丙基系统的钯催化的还原消除。

在某些实施方案中, 连接基团可包含间隔单元。连接基团的长度并不重要, 只要所述标记物与核苷酸保持足够的距离, 以免干扰核苷酸与酶之间的相互作用。

在某些实施方案中, 连接基团可由与 3'-OH 保护基团类似的官能团组成。这会使得仅需要单一处理就除去标记物和保护基团。特别优选的连接基团是可通过腈裂解的含叠氮化物的连接基团。

如本文所用的可从经修饰的核苷酸上除去标记物的试剂在很大程度上取决于所使用的标记物。例如, 在标记物掺入保护基团的情况下, 使用上文所述的除去保护基团的试剂除去标记物。或者, 在标记物通过可裂解的连接基团连接至核苷酸的碱基时, 使用如上文所述的裂解连接基团的试剂除去标记物。在优选的实施方案中, 使用相同的试剂来从经修饰的核苷酸上除去标记物和保护基团, 例如在连接基团由与 3'-OH 保护基团类似的官能团组成的情况下。

本发明的示例性实施方案

在一个具体的实施方案中, 本发明涉及用于确定靶多核苷酸的序列的方法, 其包括:

(a)提供靶多核苷酸,

(b)使所述靶多核苷酸与引物接触, 以使所述引物杂交至所述靶多核苷酸, 从而形成靶多核苷酸与引物的部分双链体,

(c)在允许聚合酶进行核苷酸聚合反应的条件下, 使所述部分双链体与聚合酶和核苷酸接触, 以使得所述核苷酸掺入到所述引物上,

其中所述核苷酸选自以下的一种或多种: 第一核苷酸、第二核苷酸、第三核苷酸和第四核苷酸, 其中所述第一核苷酸包含经第一标记物标记的第一核苷酸, 所述第二核苷酸包含经第二标记物标记的第二核苷酸, 所述第三核苷酸选自: (1) 经第一标记物标记的第三核苷酸和经第二标记物标记的第三核苷酸, 或 (2) 经第一标记物和第二标记物同时标记的第三核苷酸, 所述第四核苷酸包含未经标记的第四核苷酸,

其中所述核苷酸各自的核糖或脱氧核糖部分包含通过 2'或 3'氧原子附接的保护基团,

其中所述第一标记物是发光标记物,

(d)检测步骤(c)的所述部分双链体上所述发光标记物的存在,

(e)随后使步骤(c)的所述部分双链体与特异性结合所述第二标记物的经发光标记物标记的配体接触, 随后检测所述部分双链体上所述发光标记物的存在,

(f)任选地去除在步骤(c)的所述部分双链体中掺入的核苷酸上的保护基团和标记物,

(g)任选地重复步骤(c)-(f)一次或多次, 从而获得所述靶多核苷酸的序列信息,

其中所述发光标记物是相同的发光标记物。

在另一个具体的实施方案中, 本发明涉及用于确定靶多核苷酸的序列的方法, 其包括:

(a)提供靶多核苷酸,

(b)使所述靶多核苷酸与引物接触,以使所述引物杂交至所述靶多核苷酸,从而形成靶多核苷酸与引物的部分双链体,

(c)在允许聚合酶进行核苷酸聚合反应的条件下,使所述部分双链体与聚合酶和核苷酸接触,以使得所述核苷酸掺入到所述引物上,

5 其中所述核苷酸选自以下的一种或多种:第一核苷酸、第二核苷酸、第三核苷酸和第四核苷酸,其中所述第一核苷酸包含经第一标记物标记的第一核苷酸,所述第二核苷酸包含经第二标记物标记的第二核苷酸,所述第三核苷酸选自:(1)经第一标记物标记的第三核苷酸和经第二标记物标记的第三核苷酸,或(2)经第一标记物

10 和第二标记物同时标记的第三核苷酸,所述第四核苷酸包含未经标记的第四核苷酸,其中所述核苷酸各自的核糖或脱氧核糖部分包含通过 2'或 3'氧原子附接的保护基团,

(d)使步骤(c)的所述部分双链体与特异性结合所述第一标记物的经发光标记物标记的配体接触,随后检测所述部分双链体上所述发光标记物的存在,

随后从所述部分双链体上除去所述配体,

15 (e)使步骤(c)的所述部分双链体与特异性结合所述第二标记物的经发光标记物标记的配体接触,随后检测所述部分双链体上所述发光标记物的存在,

(f)任选地去除在步骤(c)的所述部分双链体中掺入的核苷酸上的保护基团和标记物,

(g)任选地重复步骤(c)-(f)一次或多次,从而获得所述靶多核苷酸的序列信息,

20 其中所述发光标记物是相同的发光标记物。

本发明的改进的实施方案

在作出本发明的过程中,本发明人还发现,通过添加一部分不带标记的核苷酸,可以控制单一标记的核苷酸产生的信号值,有利于不同核苷酸的区分及后续的数据分析并显著改善测序效果。

25 因此,在具体的实施方案中,除了经第一标记物标记的第一核苷酸之外,所述第一核苷酸还可以包含未经标记的第一核苷酸。除了经第二标记物标记的第二核苷酸之外,所述第二核苷酸还可以包含未经标记的第二核苷酸。

在具体的实施方案中,所述第一核苷酸中所述经第一标记物标记的第一核苷酸和所述未经标记的第一核苷酸的比率为 4:1 至 3:2。在具体的实施方案中,所述第二核苷酸中所述经第二标记物标记的第二核苷酸和所述未经标记的第二核苷酸的比率为 4:1 至 3:2。

本发明的有益技术效果

35 本发明仅基于单一激发荧光检测进行测序,相比于利用 4 种或者 2 种荧光染料标记 4 种核苷酸的检测方法,该测序方法仅需要单一激发光源与单一相机,可以缩小测序设备体积,降低测序设备制造成本。

本发明在测序过程仅产生一种荧光,可以避免因标记不同荧光染料而造成的不同荧光信号间干扰。与 2 种荧光染料标记检测相比,也避免了双色荧光与单色荧光的相互干扰。

40 与 Roche 测序法和 Ion torrent 测序法相比,本发明使用的核苷酸的 3'端羟基被修饰阻断,在测序过程中,每次反应只能合成一个脱氧核糖核苷酸,不会出现利用天然状态的脱氧核糖核苷酸测序过程中,遇到具有重复碱基的序列,一次反应合成多个脱氧核糖核苷酸的情况。因此本发明有助于提高测序的准确性。

实施例

45 实施例 1:

方法概述

(1)提供连接于支持物上的待测序的核酸分子，或者将待测序的核酸分子连接于支持物上；

(2)添加用于起始核苷酸聚合反应的引物，使引物退火至待测序的核酸分子上，引物作为起始的生长的核酸链，与待测序的核酸分子一起形成连接于支持物上的双链体；

(3)添加用于进行核苷酸聚合反应的聚合酶，以及四种核苷酸，从而形成含有溶液相和固相的反应体系；其中，四种核苷酸分别为核苷酸 A、(T/U)、C 和 G 的衍生物，并且具有碱基互补配对能力；所述四种化合物的核糖或脱氧核糖的 3'位置处的羟基(-OH)被保护基团保护；并且，第一核苷酸(如核苷酸 A)连接有第一分子标记(如生物素、N3G 等小分子)，第二核苷酸(如核苷酸 T)连接有第二分子标记(如地高辛、FITC 等)，第三核苷酸(如核苷酸 C)部分连接有第一分子标记和第二分子标记，第四核苷酸(如核苷酸 G)没有连接分子标记。为了利于不同核苷酸的区分及后续的数据分析，控制单一标记的核苷酸产生的信号值，添加部分对应的不带标记的核苷酸，如 A-cold 和 T-cold。带标记的核苷酸 A 与 A-cold 的比例范围为 4:1 到 3:2；带标记的核苷酸 T 与 T-cold 的比例范围为 4:1 到 3:2。4 类核苷酸在反应液中的终浓度在 0.5-5 μ M 之间。

(4)在允许聚合酶进行核苷酸聚合反应的条件下，以 150-350 μ l/min 的速度加入聚合反应液 150-200 μ l，反应温度 40-60 $^{\circ}$ C，反应时间 1-2min，从而将所述四种核苷酸中的一种掺入生长的核酸链的 3'端；

(5)用洗脱试剂(PBS 或 TBS) 300-400 μ l，以 150-350 μ l/min 的速度移除前一步骤的反应体系的溶液相，保留连接于支持物上的双链体。以 150-350 μ l/min 的速度加入与第一分子标记(生物素、N3G 等)特异性结合的配体(如 SA、N3G 抗体等) 150-200 μ l，该配体被荧光基团标记(如 AF532、CY3 等)，在 30-55 $^{\circ}$ C 条件下孵育 1-5min。后用洗脱试剂(PBS 或 TBS) 300-400 μ l，以 150-350 μ l/min 的速度洗脱游离的带荧光标记的配体，在拍照缓冲液中，以 50-1000ms 曝光条件检测发出的荧光信号。

(6)用洗脱试剂(PBS 或 TBS) 300-400 μ l，以 150-350 μ l/min 的速度置换前述的拍照缓冲液，后以 150-350 μ l/min 的速度加入与第二分子标记(地高辛、FITC 等)特异性结合的配体(地高辛抗体、FITC 抗体等) 150-200 μ l，该配体被荧光基团标记(如 AF532、CY3 等)，在 30-55 $^{\circ}$ C 条件下孵育 1-5min。后用洗脱试剂(PBS 或 TBS) 300-400 μ l，以 150-350 μ l/min 的速度洗脱游离的带荧光标记的抗体，在拍照缓冲液中，以 10-200ms 曝光条件检测发出的荧光信号。

(7)检测完毕后，以 150-200 μ l/min 的速度通入切除缓冲液 300-400 μ l，50-60 $^{\circ}$ C 孵育 1-2min，同时去除连接在脱氧核糖核苷酸类似物上的小分子标记物和 3'位置处的羟基(-OH)保护基团。

(8)重复步骤(3)-(7)。

(9)采集的信号经软件分析转化为序列信息。

大肠杆菌条形码序列的测定与分析

以生物素和地高辛标记核苷酸，以链霉亲和素和地高辛抗体作为其对应的配体。

1. 实验材料

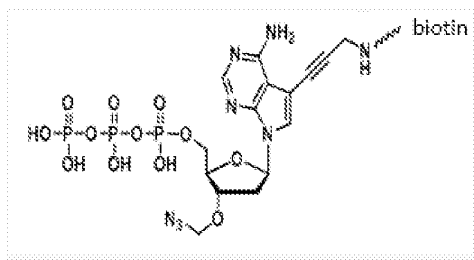
1). 大肠杆菌

2). BGISEQ-500 高通量测序试剂盒(SE100)

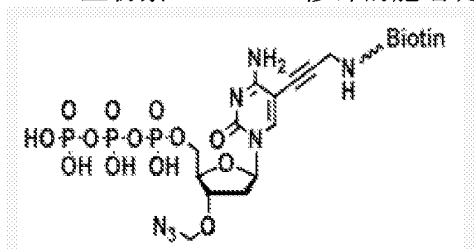
MGIEasy™ DNA 文库制备试剂盒

3). 脱氧核糖核苷酸类似物及聚合反应混合液

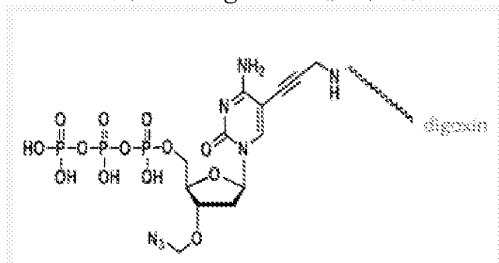
(1)生物素(biotin)修饰的腺嘌呤脱氧核糖核苷酸类似物



(2) 生物素 (biotin) 修饰的胞嘧啶脱氧核糖核苷酸类似物

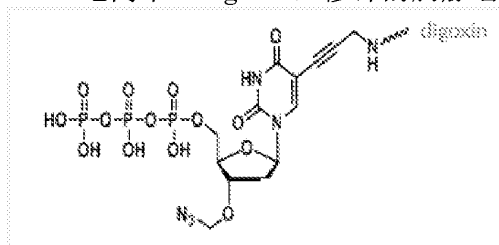


(3) 地高辛 (digoxin) 修饰的胞嘧啶脱氧核糖核苷酸类似物

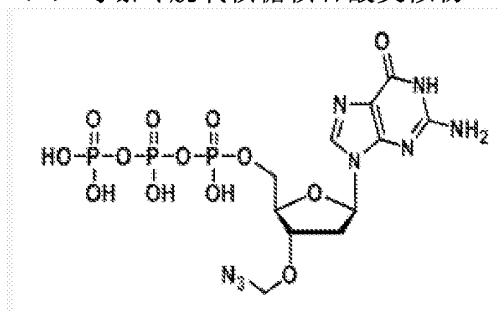


5

(4) 地高辛 (digoxin) 修饰的胸腺嘧啶脱氧核糖核苷酸类似物



(5) 鸟嘌呤脱氧核糖核苷酸类似物



10

脱氧核糖核苷酸类似物的混合反应液 1:

第一组: A-生物素+A-cold(A-生物素 : A-cold 为 4:1, A-生物素 + A-cold = 1 μ M)第二组: C-生物素+C-地高辛 (C-生物素 : C-地高辛为 2:1, C-生物素 + C-地高辛 = 2 μ M)

第三组: T-地高辛+T-cold (T-地高辛 : T-cold 为 4:1, T-地高辛 + T-cold = 1 μ M)

第四组: G-cold (1 μ M)

四组核苷酸类似物配按上述浓度及比例制成混合液。

5 脱氧核糖核苷酸类似物的混合反应液 2:

第一组: A-生物素 (1 μ M)

第二组: C-生物素+C-地高辛 (C-生物素 : C-地高辛为 2:1, C-生物素 + C-地高辛 = 2 μ M)

第三组: T-地高辛 (1 μ M)

10 第四组: G-cold (1 μ M)

四组核苷酸类似物配按上述浓度及比例制成混合液。

4). 磷酸盐缓冲液 (PBS) (生工生物)

该试剂即为抗体配体缓冲液也为洗脱试剂

15 5). 2 μ g/ml 链霉亲和素标记的 CY3 荧光(试剂厂商: Thermo Fisher scientific; 试剂货号: 434315); 2 μ g/ml 地高辛抗体标记的 CY3 荧光 (试剂厂商: Jackson ImmunoResearch; 试剂货号: 200-162-156)。

上述荧光标记抗体均用 PBS 配制。

2. 实验步骤

1) 参照如下文献提取大肠杆菌基因组 DNA。

20 So A, Pel J, Rajan S, Marziali A. Efficient genomic DNA extraction for low target concentration bacterial cultures using SCODA DNA extraction technology. Cold Spring Harb Protoc. 2010 (10): pdb. prot5506.

2) 参照 MGIEasy™ DNA 文库制备试剂盒及说明书, 制备环状单链 DNA。制备好的单链环状 DNA 已经标记有条形码序列。

25 3) 参照 BGISEQ-500 高通量测序试剂盒 (SE100) 说明书, 将环状单链 DNA 通过滚环复制, 制备成 DNA 纳米球。后继续参照 BGISEQ-500 高通量测序试剂盒 (SE100) 说明书, 将制备好的 DNA 纳米球装载到测序芯片上。

4) 向装载好 DNA 纳米球的芯片通入磷酸盐缓冲液 (生工), 通液体积 300 μ l, 通液速度 200 μ l/min。

30 5) 参照 BGISEQ-500 高通量测序试剂盒 (SE100) 说明书制备测序反应液, 将其中的脱氧核糖核苷酸替换为上述实验材料中的 4 组脱氧核苷酸类似物 1 或脱氧核苷酸类似物 2, 浓度参考实验材料。将新制备的测序反应液通入芯片, 通液体积 300 μ l, 通液速度 200 μ l/min。55 $^{\circ}$ C 孵育 1min。后通入磷酸盐缓冲液 (生工), 通液体积 300 μ l, 通液速度 200 μ l/min。

35 6) 向测序芯片通入链霉亲和素标记的 CY3 荧光 (2 μ g/ml, Thermo Fisher), 通液体积 150 μ l, 通液速度 150 μ l/min, 使带有荧光标记的链霉亲和素和生物素结合。35 $^{\circ}$ C 孵育 3min。后通入磷酸盐缓冲液 (生工), 通液体积 300 μ l, 通液速度 200 μ l/min, 去除游离的链霉亲和素标记的 CY3 荧光。

40 7) 向测序芯片通入信号采集缓冲液 (BGISEQ-500 高通量测序试剂盒 (SE100) 中已有), 通液体积 300 μ l, 通液速度 200 μ l/min, 后由激光器激发结合在待测序列上的荧光 (曝光时间 100ms) 并记录信号。

8) 向测序芯片通入磷酸盐缓冲液 (生工), 通液体积 300 μ l, 通液速度 200 μ l/min。后通入地高辛抗体标记的 CY3 荧光 (2 μ g/ml, Jackson ImmunoResearch), 通液体积 150 μ l, 通液速度 150 μ l/min, 35 $^{\circ}$ C 孵育 5min。后通入磷酸盐缓冲液 (生工), 通液体积 300 μ l, 通液速度 200 μ l/min, 去除游离的地高辛抗体标记的 CY3 荧光。

45 9) 向测序芯片通入信号采集缓冲液, 通液体积 300 μ l, 通液速度 200 μ l/min,

后由激光器激发结合在待测序列上的荧光（曝光时间 20ms）并记录信号。

10) 通入切除反应液（BGISEQ-500 高通量测序试剂盒（SE100）中已有），通液体积 300 μ l，通液速度 200 μ l/min，57℃孵育 1min。

11) 循环重复步骤 4-9。

5 12) 每次反应循环记录的荧光信号信息经分析软件转换为脱氧核糖核苷酸信息。

13) 共进行 10 个测序反应循环（测条形码），根据 500 平台自带的软件对所有读长进行条形码的拆分，计算每个条形码的拆分率。

3. 实验结果

10 经条形码序列分析软件分析，条形码拆分效率为 82%。

图 1 为待测条形码序列第一位碱基的信号提取图片，从图上可知 4 种脱氧核糖核苷酸根据检测规则被区分为 4 个信号团。左下角的为 G 碱基的信号团；水平的信号臂为 A 碱基信号团；竖直的信号臂为 T 碱基信号团；位于 AT 信号臂对角线的为 C 碱基信号团。

15 图 2 为待测条形码序列第十位碱基的信号提取图，信号臂的区分同第一位碱基的信号提取图。

图 5 为未添加不带标记核苷酸实验的第一位碱基的信号提取图，信号臂的区分同前述添加不带标记核苷酸实验的信号提取图。

20 另用上述相同的实验方法进行 50 个测序反应循环，经过分析显示映射率（mapping rate）为 70%，错误率为 2%。

实施例 2:

方法概述

(1) 提供连接于支持物上的待测序的核酸分子，或者将待测序的核酸分子连接于支持物上；

25 (2) 添加用于起始核苷酸聚合反应的引物，使引物退火至待测序的核酸分子上，引物作为起始的生长的核酸链，与待测序的核酸分子一起形成连接于支持物上的双链体；

30 (3) 添加用于进行核苷酸聚合反应的聚合酶，以及四种核苷酸，从而形成含有溶液相和固相的反应体系；其中，四种核苷酸分别为核苷酸 A、(T/U)、C 和 G 的衍生物，并且具有碱基互补配对能力；所述四种核苷酸的核糖或脱氧核糖的 3' 位置处的羟基（-OH）被保护基团保护；并且，第一核苷酸（如核苷酸 A）连接有第一分子标记（任意可被激发荧光，如 AF532、CY3 等），第二核苷酸（如核苷酸 T）连接有第二分子标记（如生物素、地高辛等小分子），第三核苷酸（如核苷酸 C）部分连接有第一分子标记和第二分子标记，第四核苷酸（如核苷酸 G）没有连接分子标记。为了利于不同核苷酸的区分及后续的数据分析，控制单一标记的核苷酸产生的信号值，添加部分对应的不带标记的核苷酸，如 A-cold 和 T-cold。带标记的核苷酸 A 与 A-cold 的比例范围为 4:1 到 3:2；带标记的核苷酸 T 与 T-cold 的比例范围为 4:1 到 3:2。4 类脱氧核糖核苷酸类似物在反应液中的终浓度在 0.5-5 μ M 之间。

40 (4) 在允许聚合酶进行核苷酸聚合反应的条件下，以 150-350 μ l/min 的速度加入聚合反应液 150-200 μ l，反应温度 40-60℃，反应时间 1-2min，从而将所述四种核苷酸中的一种并入生长的核酸链的 3' 端；

(5) 用洗脱试剂（PBS 或 TBS）300-400 μ l，以 150-350 μ l/min 的速度移除前一步骤的反应体系的溶液相，保留连接于支持物上的双链体。在拍照缓冲液中，以 50-1000ms 曝光条件检测发出的荧光信号。

45 (6) 用洗脱试剂（PBS 或 TBS）300-400 μ l，以 150-350 μ l/min 的速度置换前述的拍照缓冲液，后以 150-350 μ l/min 的速度加入与第二分子标记（biotin、digoxin 等）

特异性结合的配体 (SA、digoxin 抗体等) 150-200 μ l, 该物质被荧光基团标记 (与第一分子标记相同的荧光), 在 30-55 $^{\circ}$ C 条件下孵育 1-5min。后用洗脱试剂 (PBS 或 TBS) 300-400 μ l, 以 150-350 μ l/min 的速度洗脱游离的带荧光标记的抗体, 在拍照缓冲液中, 以 10-200ms 曝光条件检测发出的荧光信号。

- 5 (7) 检测完毕后, 以 150-200 μ l/min 的速度通入切除缓冲液 300-400 μ l, 50-60 $^{\circ}$ C 孵育 1-2min, 同时去除连接在脱氧核糖核苷酸类似物上的小分子标记物和 3'位置处的羟基 (-OH) 保护基团。

(8) 重复步骤 (3) - (7)。

(9) 采集的信号经软件分析转化为序列信息。

10

大肠杆菌 SE50 的测定与分析

1. 实验材料

1) 大肠杆菌

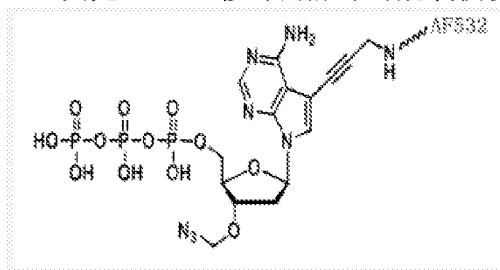
2) BGISEQ-500 高通量测序试剂盒 (SE100)

15

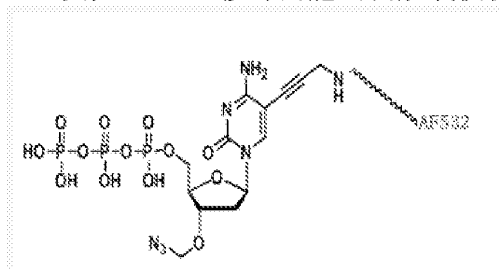
MGIEasy™ DNA 文库制备试剂盒

3) 脱氧核糖核苷酸类似物及聚合反应混合液

(1) 荧光 AF532 修饰的腺嘌呤脱氧核糖核苷酸类似物

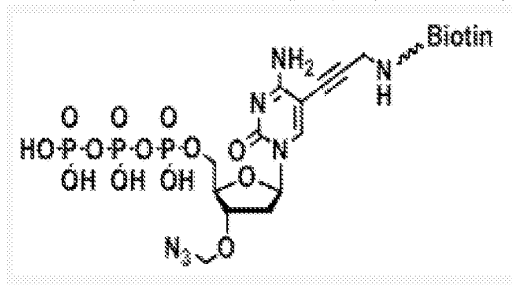


(2) 荧光 AF532 修饰的胞嘧啶脱氧核糖核苷酸类似物

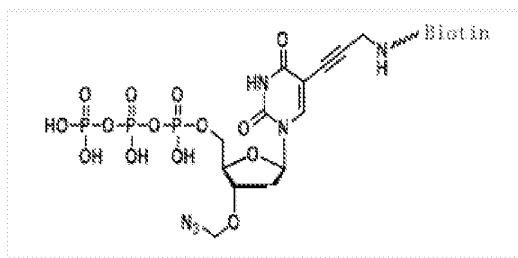


20

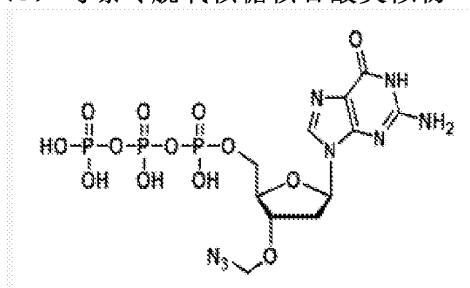
(3) 生物素 (biotin) 修饰的胞嘧啶脱氧核糖核苷酸类似物



(4) 生物素 (biotin) 修饰的胸腺嘧啶脱氧核糖核苷酸类似物



(5) 鸟嘌呤脱氧核糖核苷酸类似物



脱氧核糖核苷酸类似物的混合反应液 1:

- 5 第一组: A-AF532+A-cold (A-生物素: A-cold 为 4:1, A-生物素+ A-cold = 1 μ M)
第二组: C-生物素+C-AF532 (C-生物素: C-AF532 为 2:1, C-生物素+ C-AF532 = 2 μ M)

第三组: T-生物素+T-cold (T-生物素: T-cold 为 4:1, T-生物素 + T-cold = 1 μ M)

第四组: G-cold (1 μ M)

10 四组核苷酸类似物配按上述浓度及比例制成混合液。

脱氧核糖核苷酸类似物的混合反应液 2:

第一组: A-AF532 (1 μ M)

第二组: C-生物素+C-AF532 (C-生物素: C-AF532 为 2:1, C-生物素+ C-AF532 = 2 μ M)

15 第三组: T-生物素 (1 μ M)

第四组: G-cold (1 μ M)

四组核苷酸类似物配按上述浓度及比例制成混合液。

4) 磷酸盐缓冲液 (PBS) (生工生物)

该试剂即为抗体配体缓冲液也为洗脱试剂

20 5) 2 μ g/ml 链霉亲和素标记的 AF532 荧光(试剂厂商: Thermo Fisher scientific; 试剂货号: 434315);

上述荧光标记抗体均用 PBS 配制。

2. 实验步骤

1). 参照以下文献提取大肠杆菌基因组 DNA。

25 So A, Pel J, Rajan S, Marziali A. Efficient genomic DNA extraction for low target concentration bacterial cultures using SCODA DNA extraction technology. Cold Spring Harb Protoc. 2010 (10): pdb. prot5506.

2) 参照 MGIEasy™ DNA 文库制备试剂盒及说明书, 制备环状单链 DNA。制备好的单链环状 DNA 已经标记有条形码序列。

30 3) 参照 BGISEQ-500 高通量测序试剂盒 (SE100) 说明书, 将环状单链 DNA 通过滚环复制, 制备成 DNA 纳米球。后继续参照 BGISEQ-500 高通量测序试剂盒 (SE100) 说明书, 将制备好的 DNA 纳米球装载到测序芯片上。

4) 向装载好 DNA 纳米球的芯片通入磷酸盐缓冲液 (生工), 通液体积 300 μ l,

通液速度 200ul/min。

- 5 5) 参照 BGISEQ-500 高通量测序试剂盒 (SE100) 说明书制备测序反应液, 将其中的脱氧核糖核苷酸替换为上述实验材料中的 4 组脱氧核苷酸类似物 1 或脱氧核苷酸类似物 2, 浓度参考实验材料。将新制备的测序反应液通入芯片, 通液体积 300μl, 通液速度 200ul/min。55℃ 孵育 1min。后通入磷酸盐缓冲液 (生工), 通液体积 300μl, 通液速度 200ul/min。

6) 向测序芯片通入信号采集缓冲液 (BGISEQ-500 高通量测序试剂盒 (SE100) 中已有), 通液体积 300μl, 通液速度 200ul/min, 后由激光器激发结合在待测序列上的荧光 (曝光时间 100ms) 并记录信号。

- 10 7) 向测序芯片通入磷酸盐缓冲液 (生工), 通液体积 300μl, 通液速度 200ul/min。后通入链霉亲和素标记的 AF532 荧光 (2μg/ml, Thermo Fisher scientific), 通液体积 150μl, 通液速度 150ul/min, 35℃ 孵育 5min。后通入磷酸盐缓冲液 (生工), 通液体积 300μl, 通液速度 200ul/min, 去除游离的链霉亲和素标记的 AF532 荧光。

- 15 8) 向测序芯片通入信号采集缓冲液, 通液体积 300μl, 通液速度 200ul/min, 后由激光器激发结合在待测序列上的荧光 (曝光时间 20ms) 并记录信号。

9) 通入切除反应液 (BGISEQ-500 高通量测序试剂盒 (SE100) 中已有), 通液体积 300μl, 通液速度 200ul/min, 57℃ 孵育 1min。

10) 循环重复步骤 4-8。

- 20 11) 每次反应循环记录的荧光信号信息经分析软件转换为脱氧核糖核苷酸信息。

12) 共进行 50 个测序反应循环 (测条形码), 根据 500 平台自带的软件对所有读长进行条形码的拆分, 计算每个条形码的拆分率。

3. 实验结果

经条形码序列分析软件分析, 条形码拆分效率为 83.6%。

- 25 根据 50 个测序反应循环, 经过分析显示映射率 (mapping rate) 为 67%, 错误率为 2%。

- 30 图 3 为待测序列第一位碱基的信号提取图片, 从图上可知 4 种脱氧核糖核苷酸根据检测规则被区分为 4 个信号团。左下角的为 G 碱基的信号团; 水平的信号臂为 A 碱基信号团; 竖直的信号臂为 T 碱基信号团; 位于 AT 信号臂对角线的为 C 碱基信号团。

图 4 为待测条形码序列第 50 位碱基的信号提取图, 信号臂的区分同第一位碱基的信号提取图。

图 6 为未添加不带标记核苷酸实验的第一位碱基的信号提取图, 信号臂的区分同前述添加不带标记核苷酸实验的信号提取图。

35

权 利 要 求

1. 一种用于确定靶多核苷酸的序列的方法，其包括

(a)提供靶多核苷酸，

(b)使所述靶多核苷酸与引物接触，以使所述引物杂交至所述靶多核苷酸，从而形成靶多核苷酸与引物的部分双链体，

(c)在允许聚合酶进行核苷酸聚合反应的条件下，使所述部分双链体与聚合酶和核苷酸接触，以使得所述核苷酸掺入到所述引物上，

其中所述核苷酸选自以下的一种或多种：第一核苷酸、第二核苷酸、第三核苷酸和第四核苷酸，其中所述第一核苷酸包含经第一标记物标记的第一核苷酸和任选未经标记的第一核苷酸，所述第二核苷酸包含经第二标记物标记的第二核苷酸和任选未经标记的第二核苷酸，所述第三核苷酸选自：(1)经第一标记物标记的第三核苷酸和经第二标记物标记的第三核苷酸，或(2)经第一标记物和第二标记物同时标记的第三核苷酸；所述第四核苷酸包含未经标记的第四核苷酸；

其中所述核苷酸各自的核糖或脱氧核糖部分包含通过 2'或 3'氧原子附接的保护基团，

(d)检测步骤(c)的所述部分双链体上第一标记物的存在，

(e)检测步骤(c)的所述部分双链体上第二标记物的存在，

(f)任选地去除在步骤(c)的所述部分双链体中掺入的核苷酸上的保护基团和标记物，

(g)任选地重复步骤(c)-(f)一次或多次，从而获得所述靶多核苷酸的序列信息，

其中所述第一标记物和所述第二标记物的存在通过相同的发光信号来检测。

2. 权利要求 1 的方法，其中所述第一标记物是发光标记物。

3. 权利要求 1 的方法，其中步骤(d)包括使步骤(c)的所述部分双链体与特异性结合所述第一标记物的经发光标记物标记的配体接触，随后检测所述部分双链体上所述发光标记物的存在，

任选地，在去除步骤(c)的所述部分双链体中掺入的核苷酸上的保护基团和标记物

时一起去除所述配体。

4. 权利要求 1-3 中任一项的方法，其中步骤(e)包括使步骤(c)的所述部分双链体与特异性结合所述第二标记物的经发光标记物标记的配体接触，随后检测所述部分双链体上所述发光标记物的存在，

例如，步骤(e)在步骤(d)之后进行。

5. 权利要求 2-4 中任一项的方法，其中所述发光标记物是相同的发光标记物。

6. 权利要求 2-5 中任一项的方法，其中所述发光标记物是荧光标记物，例如是荧光团，例如选自香豆素、AlexaFluor、Bodipy、荧光素、四甲基罗丹明、吩噻嗪、吡啶、Cy5、Cy3、AF532、得克萨斯红及其衍生物。

7. 权利要求 1-6 中任一项的方法，其中所述第一核苷酸中所述经第一标记物标记的第一核苷酸和所述未经标记的第一核苷酸的比率为 4:1 至 3:2。

8. 权利要求 1-7 中任一项的方法，其中所述第二核苷酸中所述经第二标记物标记的第二核苷酸和所述未经标记的第二核苷酸的比率为 4:1 至 3:2。

9. 一种用于对多核苷酸进行测序的试剂盒，其包含：(a)选自以下的一种或多种核苷酸：第一核苷酸、第二核苷酸、第三核苷酸和第四核苷酸，其中所述第一核苷酸包含经第一标记物标记的第一核苷酸和任选未经标记的第一核苷酸，所述第二核苷酸包含经第二标记物标记的第二核苷酸和任选未经标记的第二核苷酸，所述第三核苷酸选自：(1)经第一标记物标记的第三核苷酸和经第二标记物标记的第三核苷酸，或(2)经第一标记物和第二标记物同时标记的第三核苷酸，所述第四核苷酸包含未经标记的第四核苷酸；和(b)它们的包装材料，

其中所述核苷酸各自的核糖或脱氧核糖部分包含通过 2'或 3'氧原子附接的保护基团。

10. 权利要求 9 的试剂盒，其中所述第一标记物是发光标记物。

11. 权利要求 9 的试剂盒，其还包含特异性结合所述第一标记物的经发光标记物标记的配体。

12. 权利要求 9-11 中任一项的试剂盒，其还包含特异性结合所述第二标记物的经发光标记物标记的配体。

13. 权利要求 9-12 中任一项的试剂盒，所述发光标记物是相同的发光标记物。

14. 权利要求 9-13 中任一项的试剂盒，其中所述发光标记物是荧光标记物，例如是荧光团，例如选自香豆素、AlexaFluor、Bodipy、荧光素、四甲基罗丹明、吩噻嗪、吡啶、Cy5、Cy3、AF532、得克萨斯红及其衍生物

15. 权利要求 9-14 中任一项的试剂盒，其还包含酶和适合所述酶起作用的缓冲液。