

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020年4月16日 (16.04.2020)



(10) 国际公布号
WO 2020/073748 A1

(51) 国际专利分类号: *C12Q 1/68* (2018.01) *C40B 50/06* (2006.01) IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(21) 国际申请号: PCT/CN2019/102651

(22) 国际申请日: 2019年8月26日 (26.08.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权: 201811185409.X 2018年10月11日 (11.10.2018) CN

(71) 申请人: 北京优乐复生科技有限责任公司 (BEIJING EULER TECHNOLOGY LIMITED COMPANY) [CN/CN]; 中国北京市昌平区中关村生命科学园生命园路8号11号楼4层, Beijing 102206 (CN)。

(72) 发明人: 张翼 (ZHANG, Yi); 中国北京市昌平区中关村生命科学园生命园路8号11号楼4层, Beijing 102206 (CN)。 陈琼 (CHEN, Qiong); 中国北京市昌平区中关村生命科学园生命园路8号11号楼4层, Beijing 102206 (CN)。

(74) 代理人: 北京安杰律师事务所 (ANJIE LAW FIRM); 中国北京市朝阳区东方东路19号亮马桥外交办公大楼D1座19层王颖, Beijing 100600 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,

本国际公布:
— 包括国际检索报告(条约第21条(3))。
— 包括说明书序列表部分(细则5.2(a))。



WO 2020/073748 A1

(54) Title: METHOD FOR CONSTRUCTING SEQUENCING LIBRARY

(54) 发明名称: 一种测序文库的构建方法

(57) Abstract: Disclosed are a method and kit for constructing a second generation high-throughput sequencing library, wherein same can improve the utilization efficiency of the nucleic acid template and simplify the construction process of the sequencing library to make the sequencing results more accurate and the coverage more uniform.

(57) 摘要: 一种用于第二代高通量测序文库构建的方法和试剂盒, 可提高核酸模板的利用效率, 简化测序文库的构建流程, 使得测序结果更加准确和覆盖度更加均一。

一种测序文库的构建方法

技术领域

本发明涉及用于第二代高通量测序文库构建的方法和试剂盒，更具体地，本发明涉及3'末端悬随机碱基的测序接头用于高通量测序文库构建的方法和试剂盒。

背景技术

相对于第一代测序技术，第二代测序技术测序速度更快，通量更高，符合目前科技发展对测序的需求。目前，第二代测序技术的平台主要包括Illumina公司的HiSeq、Miseq、Nextseq、Novaseq以及Life Technologies公司的SOLID system、PGM、Proton等。第二代测序技术的技术思路是边合成边测序，即根据新合成的不同碱基带来的信号变化确定DNA序列，比如，Illumina测序平台是检测光信号，Life测序平台是检测酸碱变化引起的电流变化。第二代测序技术是迄今为止发展最为成熟、使用最为广泛的DNA高通量测序手段，在基因组大规模测序和基因诊断治疗中功不可没，其在临床方面的应用也将越来越广泛。

循环DNA又称为游离DNA，是血液或在细胞外存在的DNA。游离DNA的主要来源是凋亡细胞或骨髓细胞，这些细胞释放的DNA再经体内核酸酶切割后，产生了长度约为166 bp的小片段DNA(Y. M. Dennis Lo et al. Science Translational Medicine. 2010.10:61ra91)。游离DNA在体内处于一个动态平衡状态，所以，游离DNA可以作为健康评估的一个重要参数。肿瘤发生、器官移植等变化都会导致外周血游离DNA的性质发生改变，这些性质包括游离DNA的长度、碱基信息、表观修饰等；所以，游离DNA可以作为疾病的早期诊断、监测和预后评估的一种无创检测的重要标志物。

目前，以游离DNA作为分子标记开展的无创产前诊断临床应用已经获得了全方面认可，多个国家已经全面推进该技术应用。除了碱基信息，游离DNA的长度信息也是一种非常重要的分子标记。有研究发现，不同组织或不同状态细胞的核小体、转录因子或DNA结合蛋白会与DNA的不同区域结合，最终导致游离DNA长度和测序覆盖度发生变化，根据这些差异可以追溯这些游离DNA的来源，这将给癌症早诊、器官移植、监控等领域带来新的曙光(Matthew W. Snyder et al. Cell 2016.1:57-68)。另一个方面，利用高通

量测序方法对肿瘤甲基化的研究发现,利用甲基化测序分析肿瘤与正常组织的DNA甲基化差异信号后,可以通过此差异实现癌症的早期诊断,再结合不同组织特异的甲基化信号,还可以对肿瘤的具体位置进行定位,这对于癌症早筛后诊治具有重大意义(Kun Sun et al. 2015.5:5503-12; ShichengGuo et al.2017.3:635-642)。

利用第二代高通量测序技术对游离DNA进行甲基化测序前,需要先构建甲基化测序文库。目前,第二代高通量甲基化测序文库的传统构建流程(参见图4)是先进行预文库构建,包括末端补平,5'末端磷酸化,3'末端悬A和接头连接步骤;在预文库构建完成后,再进行亚硫酸氢盐处理,亚硫酸氢盐处理会导致大量DNA损伤,最终可以进行测序的模板占原始模板的比例不到10%(Masahiko Shiraishi et al. 2004.10:409-415)。甲基化测序文库的构建流程需要,1) 每一步都需要纯化,操作繁琐;2) 补平步骤会人为引入核苷酸,改变真实的甲基化状态;3) 大量DNA模板在亚硫酸氢盐处理时被破坏,并在PCR扩增后丢失。

目前,Swift甲基化测序文库构建方法较传统方法能更高效的建库(CN104395480,参见图5),构建流程是先进行亚硫酸氢盐处理,再进行文库构建,包括3'末端加尾和接头连接,再进行延伸反应,在得到双链脱氧多核苷酸的基础上,再进行另一端测序接头的连接,因为延伸反应时的DNA模板含有大量的dUTP,加上亚硫酸氢盐处理对DNA模板的损伤,只发生一轮延伸反应,得到完整双链脱氧多核苷酸的效率低,可用于另一端测序接头连接的模板少,最终可以进行测序的模板少。而在基因诊断领域一直需要开发出更优、更高效的建库方法,以提高模板的利用效率。

发明内容

鉴于目前基于亚硫酸氢盐处理的DNA甲基化测序文库构建过程中所遇到的问题,本发明提供了一种对单链脱氧多核苷酸进行接头连接的方法;并进而提供了一种第二代高通量测序文库构建方法。本发明的方法不但适用于正常DNA,还适用于FFPE样本、古DNA、亚硫酸氢盐处理后的DNA样本等损伤严重的样本。

本发明涉及一种对脱氧多核苷酸底物进行建库的方法,所述方法包括如下步骤:

(1) 将所述脱氧多核苷酸单链底物与如下物质混合以形成第一混合物: a) 选自dGTP、dCTP、dATP和dTTP中的一种脱氧核苷酸;b)末端脱氧核苷酸转移酶和DNA连接酶;c) 控尾组分,其中该控尾组分是由5至20个核苷酸长度的多核苷酸同聚物和X

区, 以及与 X 区互补的连接子多核苷酸组成的部分双链核苷酸分子; 其中多核苷酸同聚物与 a) 中的脱氧核苷酸互补;

(2) 孵育所述的第一混合物, 脱氧多核苷酸单链底物的 3' 端与溶液中的脱氧核苷酸发生加尾反应, 添加了多核苷酸同聚尾的底物 3' 端与控尾组分的连接子连接, 得到加尾后的底物;

(3) 在步骤(2)的反应体系中, 添加 DNA 聚合酶、包含 dGTP、dCTP、dATP 和 dTTP 的脱氧核苷酸和线性扩增引物, 以形成第二混合物;

(4) 孵育所述第二混合物, 以步骤(2)得到的加尾后的底物为模板进行第一次线性延伸反应, 合成底物的互补链, 再解链, 线性扩增引物与底物互补后, 再次进行后续的线性延伸反应, 其中线性延伸反应的次数不低于 3 次;

(5) 使步骤(4)的产物解链;

(6) 在步骤(5)的溶液中添加 5' 测序接头和 DNA 连接酶形成第三混合物;

(7) 孵育所述的第三混合物, 5' 测序接头与底物互补链结合, 制备得到 DNA 文库。

在一个具体的实施方式, 在步骤(4)的第一次线性延伸反应中, 所用的引物是控尾分子 (也就是控尾区和 X 区所组成的单链), 后续的延伸反应中, 所用的引物为步骤(3)中添加的线性扩增引物。

在一个具体的实施方式, 在步骤(4)延伸反应开始之前, 降解控尾组分中的多核苷酸同聚物和 X 区片段, 以步骤(3)中添加的线性扩增引物针对底物进行线性延伸反应。

在一个具体的实施方式, 步骤(3)中添加的线性扩增引物对底物进行竞争性结合, 从而该添加的线性扩增引物针对底物进行线性延伸反应。

本发明进一步涉及一种试剂盒, 该试剂盒能对脱氧多核苷酸底物进行建库, 其包含:

组分一: 包含选自 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸、末端脱氧核苷酸转移酶、DNA 连接酶和控尾组分, 其中所述的控尾组分由 5 至 20 个核苷酸长度的多核苷酸同聚物和 X 区, 以及与 X 区互补的连接子多核苷酸组成的部分双链核苷酸分子, 该多核苷酸同聚物与所述的选自 dGTP、dCTP、dATP 和 dGTP 中的一种脱氧核苷酸互补;

组分二: 包含 DNA 聚合酶、包含 dGTP、dCTP、dATP 和 dTTP 的脱氧核苷酸, 和线性扩增引物;

组分三：包含 5'测序接头和 DNA 连接酶。

本发明通过设计线性扩增，可有效提高原始单链多核苷酸底物的互补链的数量。通过设计一条链的3'末端悬若干个随机碱基的5'测序接头，可以非常高效的在多核苷酸底物的互补链的3'末端添加5'测序接头。进一步，将多核苷酸底物变性为单链，在底物加尾后与连接子连接、之后完成多核苷酸底物的线性扩增得到互补链、互补链3'端连接5'测序接头后，PCR富集，得到可进行下一代测序的文库。

此外，由于底物被甲基化，得到了含U碱基的脱氧多核苷酸底物模板，由于U碱基可能导致线性扩增的中止，得到的互补链长短不一，但是本发明通过一条链的3'末端悬若干个随机碱基的5'测序接头与互补链单链多核苷酸底物的连接，大大提高了互补链的利用率。依据本发明，本建库流程可对低至2ng人类培养细胞来源的基因组DNA构建全基因组甲基化测序文库，并得到高效的测序结果。

附图说明

图 1：本发明方法 DNA 文库构建流程图

图 2：控尾组分的结构

图 3：悬 6N “5'测序接头” 的结构

图 4：传统方法 DNA 甲基化文库构建流程图

图 5：Swift 方法 DNA 甲基化文库构建流程图

发明详述

说明书中的 3'、3'端和 3'末端的含义相同，5'、5'端和 5'末端的含义相同，他们分别指核苷酸序列的 3'端或者 5'端。

多核苷酸底物

多核苷酸底物是需要进行加尾反应和文库构建的多核苷酸底物片段。在各种实施方案中，多核苷酸底物为单链或双链的 DNA。在另外的实施方案中，多核苷酸底物是经过化学处理的核苷酸序列，包括但不限于是经过亚硫酸氢盐处理的多核苷酸。

多核苷酸底物可以是天然来源的或合成的。天然来源是来自原核生物或真核生物，如人、小鼠、病毒、植物或细菌的多核苷酸序列。本发明的多核苷酸底物还可以是，FFPE 样本、古 DNA、亚硫酸氢盐处理后的 DNA 等损伤严重的样本。多核苷酸底物被加尾，

能用于涉及微阵列的测定并且产生用于下一代核酸测序的文库。加尾的多核苷酸底物还可以用于多核苷酸序列的有效克隆。

在一些实施方式中，多核苷酸底物是单链的或双链的并且包含 3'端游离羟基。在一些方面，多核苷酸底物是双链的并且包含平末端。在其它方面，双链多核苷酸底物包含 3'凹陷末端。多核苷酸底物的突出末端或凹陷末端的长度可以变化。在各个方面，多核苷酸底物的突出末端或凹陷末端的长度为 1、2、3、4、5、6、7、8、9、10 或更多个核苷酸。

在一些方面，多核苷酸底物的长度介于约 10 个与约 5000 个核苷酸之间，或介于约 40 个与约 2000 个核苷酸之间，或介于约 50 个与约 1000 个核苷酸之间，或介于约 100 个与约 500 个核苷酸之间。在另外的方面，多核苷酸底物的长度为至少 3 个到至多约 50、100 或 1000 个核苷酸。

DNA 去磷酸化

DNA 去磷酸化是指 DNA 5'端和 3'端的第一个氨基酸残基磷酸基团的除去；一般是采用碱性磷酸酶处理 DNA 来实现 5'端和 3'端残基的去磷酸化。

在一些实施方式中，本发明在对脱氧多核苷酸底物进行加尾之前，对脱氧多核苷酸底物进行去磷酸反应。

加尾反应

如本文所使用的术语“加尾”可与术语“受控加尾”互换。本发明所述方法的步骤(2)对脱氧多核苷酸底物进行加尾，所述方法用于以受控方式将所需数量的核苷酸添加至多核苷酸底物 3'端。通过举例并且非限制性地，通过添加控尾组分，使多核苷酸底物的尾部控制在一定的长度范围内(见图 1)。该控尾组分包含了 5 至 20 个核苷酸长度的多核苷酸同聚物，控尾组分与底物新添加的核苷酸同聚尾序列形成双链结构，因此降低了聚合过程的速率，使多核苷酸底物的尾部控制在一定的长度范围内(见图 1)。

加尾反应中所添加的核苷酸是选自 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸，例如可以是 dGTP 溶液，或 dCTP 溶液，或 dATP 溶液，或 dTTP 溶液。在一个具体实施方案中，采用含聚(dT)的核苷酸同聚序列的控尾组分控制 TdT 酶(末端脱氧核苷酸转移酶)在多核苷酸底物的 3'端添加聚(dA)尾(又称为，核苷酸(dA)同聚尾)。进一步，多核苷酸底物的同聚尾与控尾组分的连接子连接，形成底物的 3'端加尾区。

在一个具体实施方案中，控尾组分包含了 5-20 个、优选 5-13 个、进一步优选为 7-10 个、更优选为 7-9 个相同核苷酸的核苷酸同聚物序列。

在一个具体实施方案中，多核苷酸底物与控尾组分的摩尔浓度比范围为 1:1-1:100，优选 1:5-1:50。

在一个具体实施方案中，本发明所述方法的步骤(2)的加尾反应的 pH 范围为约 5.0 到约 9.0；多核苷酸底物与单核苷酸摩尔浓度比范围为 1:10-1:20000，优选 1:100-1:2000；孵育的时间在 1 分钟到 120 分钟，优选 0.5-60 分钟，0.5-30 分钟，1-20 分钟，1-15 分钟或 1-10 分钟；孵育的温度为 20°C-50°C，优选 25°C-45°C，更优选 25°C-37°C。

控尾组分

本发明通过添加控尾组分来控制多核苷酸底物的加尾长度和效率。控尾组分是由控尾区和 X 区，以及能够与 X 区互补的连接子序列（见图 2），控尾组分又称为“控尾接头”。控尾区和 X 区所组成的单链被称为“控尾分子”。

本发明的控尾区是一段 5-20 个核苷酸长度的多核苷酸同聚物。多核苷酸同聚物又称“poly 区”，是相同的核苷酸连接成的多核苷酸链。本发明的控尾区是 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸组成的核苷酸同聚物序列；优选是 dTTP 组成的聚(dT)的控尾区。优选地，本发明控尾区的多核苷酸同聚物的长度为 5-20 个核苷酸，优选 7-20、9-20 个核苷酸，进一步优选为 5-10 个核苷酸，7-10 个核苷酸，更优选为 7-9 个核苷酸。一定长度的 dGTP 或 dCTP 多核苷酸同聚物可以有效控制多核苷酸底物加尾在 20 个核苷酸左右。

“X 区序列”提供用于核酸片段的扩增或测序的引发序列，还可以包含用于区分不同的底物分子的标记序列，标记序列可以包含 4-16 个碱基，并且在一些方面用于下一代测序应用。在本发明的一些实施方式中，X 区序列可以是但不限于与 Illumina、Ion Torrent、Roche 454 或 SOLiD 测序平台相容的含下一代测序(NGS)接头序列。X 区序列可以是 DNA 序列、RNA 序列或者包含 DNA 和 RNA 的杂聚序列。

控尾组分中的连接子，只具有与 X 区序列互补的连接子被称为“短连接子”；除了与 X 区互补序列外还包括延伸引物结合区的连接子序列，被称为“长连接子”，如表 2 所示。

本发明使用了一种对脱氧多核苷酸底物进行加尾的方法，所述方法用于以受控方式将所需数量的核苷酸添加至多核苷酸底物 3' 端。通过举例并且非限制性地，通过添加

控尾组分, 该控尾组分包含了 5-20 个核苷酸长度的多核苷酸同聚物, 控尾组分与底物新添加的核苷酸同聚尾序列互补形成双链结构, 因此降低了聚合过程的速率, 使多核苷酸底物的尾部控制在一定的长度范围内 (见图 1)。

在本发明的一些实施方式中, 采用含聚(dT)的核苷酸同聚序列的控尾组分控制 TdT 酶在多核苷酸底物的 3' 端添加聚(dA)尾 (又称为核苷酸(dA)同聚尾)。进一步, 多核苷酸底物的聚(dA)尾与控尾组分的连接子连接, 形成底物的 3' 端加尾区。

在本发明的一些实施方式中, 控尾组分包含封闭基团。本文所使用的封闭基团是阻止通过酶进行延伸的部分。如果没有封闭基团, 酶能够通过添加核苷酸合成多核苷酸。封闭基团包括但不限于: 磷酸基团、碳三间臂、双脱氧核苷酸、核糖核苷酸、氨基以及反向脱氧胸苷。

在本发明的一些实施方式中, 控尾组分连接子的 5' 端有磷酸化修饰, 3' 端有封闭基团, 控尾区的 3' 有封闭基团。

线性扩增反应

线性扩增反应也称为“线性延伸反应”或“线性延伸”。

本发明在对脱氧多核苷酸底物进行加尾之后, 进一步以脱氧多核苷酸底物为模板进行线性扩增反应。本发明提供了一种对脱氧核苷酸底物进行线性扩增的方法, 所述方法用于以线性扩增的方式来增加脱氧多核苷酸底物的数量 (见图 1)。

在本发明的一些实施方式中, 在加尾反应之后, 加入线性扩增引物, 以加尾的底物为模板进行延伸反应, 延伸反应可以首先是通过控尾分子发生延伸反应。在本发明的一些实施方式中, 线性扩增引物通过竞争的方式使控尾分子与底物多核苷酸底物分离, 进而以脱氧多核苷酸底物为模板进行线性扩增反应。在一些具体实施方式中, 在步骤(4)延伸反应开始之前, 降解控尾组分中的多核苷酸同聚物和 X 区片段, 以步骤(3)中添加的线性扩增引物针对底物进行线性延伸反应。

在一个具体实施方式中, 该方法包括: 在加尾反应之后, 多核苷酸底物变性呈单链状态, 加入与底物 3' 连接子序列互补的线性扩增引物、DNA 聚合酶和脱氧核苷酸, 与多核苷酸底物进行反应; 在线性扩增引物的 3' 端发生核苷酸延伸反应, 合成底物的互补链, 得到双链脱氧多核苷酸, 双链脱氧多核苷酸通过变性使底物的互补链与底物分离, 底物再次与线性扩增引物、DNA 聚合酶和脱氧核苷酸发生延伸反应, 发生延伸反应的次数称为线性扩增循环数。在一些方面, 所述的 DNA 聚合酶可以高效扩增含 U 碱基的

脱氧多核苷酸底物模板。

在一个具体实施例方案中，线性扩增循环数不低于 3 次，优选不低于 4 次，优选为 4-50 次，进一步优选 4-20 次，更优选为 4-12 次，8-12 次。

5'测序接头的连接反应

本发明的 5'测序接头是具有部分双链结构的脱氧多核苷酸；“部分双链”指的是所述悬随机个数碱基的 5'测序接头包含单链部分和双链部分。

5'测序接头提供用于核酸片段的扩增或测序的引发序列，并且在一些方面用于下一代测序应用。

本发明通过添加悬随机碱基的 5'测序接头与线性扩增反应之后得到的底物的互补链发生连接反应，悬随机碱基的 5'测序接头的 3'末端包含一段随机碱基单链多核苷酸 (见图 3)，因此悬随机碱基的 5'测序接头是多分子结构的具有部分双链的多核苷酸，悬随机碱基的 5'测序接头又称“悬 N 5'测序接头”。例如悬 6 个随机碱基的 5'测序接头又称“悬 6N 5'测序接头”，N 代表一种脱氧核苷酸碱基，也就是 6 个 N 中每一个都是随机选自 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸碱基，“悬 6N 5'测序接头”是不同随机碱基连接后的混合分子。

本发明的 5'测序接头悬随机碱基的个数为 0-50 个，优选 2-30 个，进一步优选为 2-17，4-15，4-10，更优选 7-10，7-9 个。

在本发明的具体实施方式中，5'测序接头由两条多核苷酸链退火形成，不含随机碱基的多核苷酸链的 5'端有磷酸化修饰，3'端有封闭基团，含随机碱基的多核苷酸链的 3'端有封闭基团 (见图 3)。

在对单链脱氧多核苷酸进行接头连接的方法中，所述方法用于以悬随机碱基的 5'测序接头的随机碱基单链多核苷酸部分与线性扩增反应之后得到的底物的互补链的 3'端形成互补，得到除 5'测序接头双链部分之外的部分双链结构，在 DNA 连接酶的作用下，5'测序接头不含随机碱基的多核苷酸链的 5'端与多核苷酸底物的互补链的 3'端连接 (见图 1)；在线性扩增步骤及纯化步骤之后用于与 5'测序接头连接的底物互补链增多，连接效率提高。

在一个具体实施例方案中，多核苷酸底物与 5'测序接头的摩尔浓度比范围为 1:100-1:4000，优选 1:500-1:1000。

连接酶

可用于本发明方法的连接酶可以是 DNA 连接酶和 RNA 连接酶，包括但不限于 T4 DNA 连接酶、大肠杆菌 DNA 连接酶、T7 DNA 连接酶以及 T4 RNA 连接酶。

本发明的连接酶使控尾组分中的连接子与底物加尾的多核苷酸连接。在另一些实施方式中，本发明的连接酶使 5'测序接头与合成的底物的互补链单链脱氧多核苷酸连接。

分离步骤

在一些实施方案中，对本发明步骤(7)之后的多核苷酸产物进行纯化。多核苷酸产物的纯化通过本领域技术人员已知和理解的任何方法来进行。在本发明多核苷酸底物的纯化可以通过加入表面是羧基修饰的磁珠来进行。在其它的具体实施方式中，通过柱纯化和沉淀来进行多核苷酸底物的纯化。

具体实施方式

实施例中所用的序列参见如下表1-表6

表 1. 用于实施例的 DNA 多核苷酸

序号	序列 (方向 5'-3')	描述	序列序号
001	AGACGTGTGCTCTTCCGATCTTTTTTTTTT*T*T*T -C3 Spacer	具有 12b 聚 (dT) 控尾区的控尾分子	SEQ ID NO: 1
002	Phos- AGATCGGAAGAGCACACGTCTGAACTCCAGT*C*A*C- C3 Spacer	长连接子	SEQ ID NO: 2
003	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	线性扩增引物	SEQ ID NO: 3
004	GTGACTGGAGTTCAGACGTGT	qPCR 反向引物	SEQ ID NO: 4
005	CACTCTTCCCTACACGACG	qPCR 正向引物	SEQ ID NO: 5
006	Phos- NNNNNNNNAGATCGGAAGAGCACACGTCTGAACTCCA GT*C*A*C-C3 Spacer	带标签长连接子	SEQ ID NO: 6
007	CACTCTTCCCTACACGACGCTCTTCCGATC*T*N*N- C3 Spacer	3'末端含 2 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 7
008	CACTCTTCCCTACACGACGCTCTTCCGATCT*N*N*N- C3 Spacer	3'末端含 3 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 8
009	CACTCTTCCCTACACGACGCTCTTCCGATCTN*N*N*N -C3 Spacer	3'末端含 4 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 9
010	CACTCTTCCCTACACGACGCTCTTCCGATCTNN*N*N* N-C3 Spacer	3'末端含 5 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 10

011	ACACTCTTTCCTACACGACGCTCTTCCGATCTNNN*N* N*N-C3 Spacer	3'末端含 6 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 11
012	ACACTCTTTCCTACACGACGCTCTTCCGATCTNNNN*N* *N*N-C3 Spacer	3'末端含 7 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 12
013	ACACTCTTTCCTACACGACGCTCTTCCGATCTNNNNN* N*N*N-C3 Spacer	3'末端含 8 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 13
014	ACACTCTTTCCTACACGACGCTCTTCCGATCTNNNNNN* *N*N*N-C3 Spacer	3'末端含 9 个随机碱基的 5'测序接头正向引物	SEQ ID NO: 14
015	Phos- AGATCGGAAGAGCGTCGTGTAGGGAAAGAG*T*G*T-C3 Spacer	5'测序接头反向引物	SEQ ID NO: 15
016	A/5mC/A/5mC/T/5mC/TTT/5mC//5mC//5mC/TA/5mC/A/5mC/ GA/5mC/G/5mC/T/5mC/TT/5mC//5mC/GAT/5mC/T	传统甲基化测序接头正向 引物	SEQ ID NO: 16
017	Phos- GAT/5mC/GGAAGAG/5mC/A/5mC/A/5mC/GT/5mC/TGAA/5 mC/T/5mC//5mC/AGT/5mC/A/5mC/	传统甲基化测序接头反向 引物	SEQ ID NO: 17
018	AATGATACGGCGACCACCGAGATCTACACACAGTGCTA CACTCTTTCCTACACGACG	P5 PCR 标记引物 M501	SEQ ID NO: 18
019	AATGATACGGCGACCACCGAGATCTACACCGCATGTAA CACTCTTTCCTACACGACG	P5 PCR 标记引物 M502	SEQ ID NO: 19
020	AATGATACGGCGACCACCGAGATCTACACTATCGCAGA CACTCTTTCCTACACGACG	P5 PCR 标记引物 M503	SEQ ID NO: 20
021	AATGATACGGCGACCACCGAGATCTACACGTGCATACA CACTCTTTCCTACACGACG	P5 PCR 标记引物 M504	SEQ ID NO: 21
022	AATGATACGGCGACCACCGAGATCTACACAGCTAGTCA CACTCTTTCCTACACGACG	P5 PCR 标记引物 M505	SEQ ID NO: 22
023	AATGATACGGCGACCACCGAGATCTACACCTAGCTGAA CACTCTTTCCTACACGACG	P5 PCR 标记引物 M506	SEQ ID NO: 23
024	CAAGCAGAAGACGGCATAACGAGATGTCTTAGGGTGACT GGAGTTCAGACG	P7 PCR 标记引物 N726	SEQ ID NO: 24
025	CAAGCAGAAGACGGCATAACGAGATTAGCTGCAGTGACT GGAGTTCAGACG	P7 PCR 标记引物 N728	SEQ ID NO: 25
026	AATGATACGGCGACCACCGAGATCTACACTCTTTCCT ACACG*A*C	P5 PCR 引物	SEQ ID NO: 26
027	CAAGCAGAAGACGGCATAACGAGATTTGACTGTGACTGG AGTTCAGA*C*G	P7 PCR 标记引物 13	SEQ ID NO: 27
028	CAAGCAGAAGACGGCATAACGAGATTGACATGTGACTG GAGTTCAGA*C*G	P7 PCR 标记引物 14	SEQ ID NO: 28
029	AGACGTGTGCTCTTCCGATCNNNNNNNTTTTTTTTT* T*T*T-C3 Spacer	带标签具有 12b 聚 (dT) 控尾区的控尾分子	SEQ ID NO: 29

Phos: 磷酸; *: 硫代位点; C3 Spacer: 碳 3 间臂; 5mC: 5-甲基-胞嘧啶脱氧核苷酸;
N: dA、dT、dC 或 dG 核苷酸

表 2. 用于甲基化测序文库构建的悬聚 (dT) 控尾组分

多核苷酸序号	接头结构
002	长连接子 /5' Phos/AGATCGGAAGAGCACACGTCTGAACTCCAGT*C*A*C /3' C3 Spacer
001	控尾分子 /3' C3 Spacer/T*T*T*TTTTTTTTTCTAGCCTTCTCGTGTGCAGA 5' <div style="display: flex; justify-content: space-around; width: 100%;"> 控尾区 X区 </div>

Phos: 磷酸; C3 Spacer: 碳 3 间臂; *: 硫代位点

表 3. 用于实施例 1 的悬不同数量 N 的 5' 测序接头

多核苷酸序号	接头结构
012	悬 7 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCTNNNN*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
007	悬 2 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCT*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
008	悬 3 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCT*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
009	悬 4 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCTN*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
010	悬 5 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCTNN*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
011	悬 6 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCTNNN*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
013	悬 8 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCTNNNN*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/
014	悬 9 个 N 的 5' 测序接头 /3' C3 Spacer/5' ACACTCCTTCCCTACACGACGCTCTTCCGATCTNNNNN*N*N*N/3' C3 Spacer/
015	控尾分子 /3' C3 Spacer/T*G*T*GAGAAAGGGATGTGCTGCGAGAGGCTAGA/5' Phos/

Phos: 磷酸; C3 Spacer: 碳 3 间臂; *: 硫代位点; N: dA、dT、dC 或 dG 的核苷酸

表 4. 用于甲基化测序文库构建的带分子标签的悬聚 (dT) 控尾组分

多核苷酸序号	接头结构
006	带标签长连接子 /5' Phos/NNNNNNNNAGATGAGAGAGCAGCAGTCTGAACTCCAGT*C*A*C /3' C3 Spacer
029	带标签控尾分子 /3' C3 Spacer/T*T*T*TTTTTTTTTNNNNNNNNCTAGCCTTCTCGTGTGCAGA 5' <div style="display: flex; justify-content: space-around; width: 100%;"> 控尾区 X区 </div>

Phos: 磷酸; C3 Spacer: 碳 3 间臂; *: 硫代位点; N: dA、dT、dC 或 dG 的核苷酸

表 5. 用于甲基化测序文库构建的“传统甲基化测序接头”

多核苷酸序号	接头结构
016	5' A/5mC/A/5mC/T/5mC/TTT/5mC//5mC//5mC/TA/5mC/A/5mC/GA/5mC/G/5mC/T/5mC//5mC/GAT/5mC/T 3'
017	/5mC/A/5mC/TGA/5mC//5mC/T/5mC/AAAT/5mC/TG/5mC/A/5mC/A/5mC/CAGAAAG/5mC/TAG/5' Phos/

Phos: 磷酸; 5mC: 5-甲基-胞嘧啶脱氧核苷酸

实施例

实施例1. 悬不同数量随机碱基的5'测序接头对5'测序接头连接的影响

材料:

5 x退火缓冲液 (碧云天, 目录号D0251)

无酶水 (索莱宝, 目录号R1600-100)

λ-DNA (takara, 目录号3019)

亚硫酸氢盐处理试剂盒 (Zymo Research, 目录号D5005)

FastAP 温敏碱性磷酸酶 (ThermoFisher, EF0651, 1 U/μL)

10 x CutSmart缓冲液 (New England Biolabs, 目录号B7204S)

10 x绿色缓冲液 (Enzymatics, 目录号B0120, 20 mM Tris-醋酸盐、50 mM 乙酸钾、10 mM 乙酸镁, pH 7.9)

β-烟酰胺腺嘌呤二核苷酸 (New England Biolabs, 目录号B9007S, 50 mM)

dATP (Takara, 目录号4026, 100 mM)

TdT酶 (Enzymatics, 目录号P7070L, 20 U/μL)

大肠杆菌DNA连接酶 (Takara, 目录号2161, 60 U/μL)

EB缓冲液 (Qiagen, 目录号19086)

dNTP (takara, 目录号4030, 每种2.5 mM)

Phusion U 热启动DNA聚合酶 (ThermoFisher, 目录号F555L, 2 U/μL)

5 x Phusion HF 缓冲液 (ThermoFisher, 目录号F555L)

Beckman Ampure XP磁珠 (Beckman, 目录号A63882)

SB缓冲液: 20% PEG8000, 2.5M NaCl, 10mM Tris-盐酸, 1mM EDTA

2 x T4 DNA快速连接反应缓冲液 (Enzymatics, 目录号B1010)

T4 DNA快速连接酶 (Enzymatics, 目录号L6030-HC-L, 600 U/μl)

方法:

(1) 接头制备: 将多核苷酸对 (001/002, 007/015, 008/015, 009/015, 010/015, 011/015,

012/015, 013/015, 014/015)等摩尔量混合, 在1 x退火缓冲液中于95°C孵育2分钟, 然后缓慢冷却至室温, 得到控尾接头 (如表2所示)和3'末端悬不同数量随机碱基的5'测序接头 (如表3所示)。

- (2) 使用亚硫酸氢盐处理试剂盒对 40 ng λ -DNA (takara, 目录号 3019)进行亚硫酸氢盐处理。
- (3) 使用聚焦超声仪 (Covaris, 目录号 S220)将步骤(2)的 DNA 产物片段化至 300 bp, 待用。
- (4) 按照表 1-1 所示制备 DNA 去磷酸化反应混合液, 在 37°C温浴 30 分钟后, 95°C处理 5 分钟, 之后立即插入冰上并孵育 2 分钟后待用, 使底物保持单链状态。

表 1-1

DNA 去磷酸反应混合液	
DNA (来自步骤(3))	17 μ l
10 x CutSmart缓冲液	2 μ l
FastAP 温敏碱性磷酸酶	1 μ l
总体积	20 μ l

- (5) 对多核苷酸底物进行加尾: 按照表 1-2 所示制备加尾和连接反应混合液, 将反应混合液在 37°C温浴 30 分钟后, 95°C处理 5 分钟, 然后保持在 4°C。

表 1-2

加尾和连接反应混合液	
DNA (来自步骤(4))	20 μ l
无酶水	4 μ l
2.5 x绿色缓冲液加12.5% PEG 8000	8 μ l
2 mM dATP加2 mM β -烟酰胺腺嘌呤二核苷酸	2 μ l
10 μ M控尾接头 (表2所示的001/002)	2 μ l
1:33 稀释的TdT酶 (1 μ l TdT酶加32 μ l EB缓冲液)	2 μ l
1:6 稀释的大肠杆菌DNA连接酶 (1 μ l TdT酶加5 μ l EB缓冲液)	2 μ l
总体积	40 μ l

(6) 线性扩增：制备如表 1-3 所示的线性扩增的反应混合液，按照表 1-4 所示的 PCR 扩增程序运行 4 个线性扩增；使用 166 μl 1:6 稀释的 Beckman Ampure XP 磁珠(1 体积的 Beckman Ampure XP 磁珠加上 5 体积的 SB 缓冲液)和 280 μl 的 1.8:1 稀释的 SB 缓冲液(1.8 体积的 SB 缓冲液加上 1 体积的无酶水)纯化回收线性扩增产物，再加入 100 μl EB 缓冲液洗脱，将 100 μl 洗脱液分成 5 μl /份于 200 μl PCR 管中，分出 18 份用于下步反应。

表 1-3

线性扩增的反应混合液	
DNA (来自步骤(5))	40 μl
无酶水	29 μl
5 x Phusion HF缓冲液	20 μl
100 μM 线性扩增引物 (003)	2 μl
2.5 mM dNTP	8 μl
Phusion U 热启动DNA聚合酶	1 μl
总体积	100 μl

表 1-4

PCR反应条件		
温度	时间	循环数
95°C	3分钟	1
95°C	30秒	4
60°C	30秒	
68°C	1分钟	
68°C	5分钟	1
4	保存	1

(7) 将步骤(6)的 18 份 DNA 在 95°C 反应 5 分钟，立即置于冰上 2 分钟待用，使双链解链保持单链待用。

(8) 按照如表 1-5 所示制备来自步骤(7)的 DNA 与悬 2N-9N 的“5'测序接头”连接的反应

混合液，每种接头两个平行，在 25°C 温浴 15 分钟，使用 17 μl Beckman Ampure XP 磁珠回收连接后的 DNA，再加入 26 μl 无酶水洗脱，得到 PCR 扩增前文库。

表 1-5

“5’测序接头”连接反应混合液	
DNA (来自步骤(7))	5 μl
无酶水	2 μl
2 x T4 DNA快速连接反应缓冲液	10 μl
10 μM 悬n个N “5’测序接头” (表3)	2 μl
T4 DNA快速连接酶	1 μl
总体积	20 μl

(9) 使用文库定量试剂盒 (KAPA Biosystems, 目录号 KK4824) 以及 DNA 定量标准品和预混合引物试剂盒 (KAPA Biosystems, 目录号 KK4808), 加上 qPCR 正向引物/qPCR 反向引物 (005/004, 如表 1, 使用浓度同预混合引物试剂盒) 检测 PCR 扩增前文库的摩尔浓度, 计算时 PCR 扩增前文库片段大小为 320 bp。

实验结果: 如表 1-6 所示, 使用悬 2 个 N 的 “5’测序接头” 构建的 PCR 扩增前文库浓度最低, 为 0.000780 nM, “5’测序接头” 悬 N 的数量从 2 增加到 4 个, PCR 扩增前文库浓度从 0.000780 nM 增加到 0.0254 nM, 呈指数增加; 从 4 到 7, PCR 扩增前文库浓度增加趋势减弱, 使用悬 7N “5’测序接头” 构建的 PCR 扩增前文库浓度最高, 为 0.0653 nM; 使用悬 7 个 N、8 个 N 和 9 个 N 的 “5’测序接头” (见表 3) 构建的 PCR 扩增前文库浓度基本一致, 分别为 0.0653 nM、0.0627 nM 和 0.0646 nM。

结论: 悬 2-9 个 N 的 “5’测序接头” 均能有效与线性扩增后得到的核苷酸底物的互补链发生连接反应, 并对亚硫酸氢盐处理后的 DNA 进行甲基化文库的构建。

表 1-6

“5’测序接头”悬随 机碱基(N)个数	2	3	4	5	6	7	8	9
文库浓度 (nM)	0.000780	0.00503	0.0254	0.0404	0.0503	0.0653	0.0627	0.0646

实施例2. 不同接头浓度对5’测序接头连接的影响

方法:

- (1) 接头制备: 按照实施例1中接头制备方法制备控尾接头 (001/002, 如表2所示)以及3'末端悬6个N的“5'测序接头”(011/015, 如表3所示)
- (2) 使用亚硫酸氢盐处理试剂盒对 24 ng λ -DNA 进行亚硫酸氢盐处理。
- (3) 制备线性扩增产物: 按照实施例1中所描述的方法制备线性扩增产物, 线性扩增产物纯化回收时加入31.2 μ l EB缓冲液洗脱, 将31.2 μ l洗脱液分成2.6 μ l/份于200 μ l PCR管中, 分出10份用于下步反应。
- (4) 将步骤(3)的 DNA 在 95°C反应 5 分钟, 立即置于冰上 2 分钟待用。
- (5) 反应: 如下表 2-1 所示配制进行“5'测序接头”连接反应的混合液, 在 25°C温浴 15 分钟, 使用 17 μ l Beckman Ampure XP 磁珠回收连接后 DNA, 再加入 26 μ l 无酶水洗脱, 得到 PCR 扩增前文库。

表 2-1

“5'测序接头”连接反应混合液					
DNA底物: 5'测序接头比例	1:100	1:500	1:1000	1:2000	1:4000
DNA (来自步骤(4))	2.6 μ l	2.6 μ l	2.6 μ l	2.6 μ l	2.6 μ l
无酶水	6 μ l	4.4 μ l	2.4 μ l	3.2 μ l	-
2 x T4 DNA快速连接反应缓冲液	10 μ l	10 μ l	10 μ l	10 μ l	10 μ l
10 μ M悬6个N“5'测序接头”	0.4 μ l	2 μ l	4 μ l	-	-
25 μ M悬6个N“5'测序接头”	-	-	-	3.2 μ l	6.4 μ l
T4 DNA快速连接酶	1 μ l	1 μ l	1 μ l	1 μ l	1 μ l
总体积	20 μ l	20 μ l	20 μ l	20 μ l	20 μ l

- (6) 按照实施例 1 所描述的方法检测 PCR 扩增前文库的摩尔浓度。

实验结果: 如表 2-2 所示, DNA 底物与 5'测序接头比例为 1:100 时构建的 PCR 扩增前文库浓度最低, 为 0.0210 nM, 1:4000 时构建的 PCR 扩增前文库浓度最高, 为 0.0836 nM。

结论: DNA 底物与 5'测序接头比例为 1:100 到 1:4000 均能有效的对亚硫酸氢盐处理后的 DNA 实施甲基化文库的构建。

表 2-2

DNA 底物: 5'测序接头比例	1:100	1:500	1:1000	1:2000	1:4000
文库浓度 (nM)	0.0210	0.0499	0.0641	0.0693	0.0836

实施例 3. 不同线性扩增循环数对甲基化 PCR 扩增前文库构建的影响

- (1) 接头制备: 按照实施例1中接头制备方法制备控尾接头 (001/002, 如表2所示)以及3'末端悬7个N的“5'测序接头”(012/015, 如表3所示)。
- (2) 使用亚硫酸氢盐处理试剂盒对 20 ng λ -DNA 进行亚硫酸氢盐处理。
- (3) 使用聚焦超声仪 (Covaris, 目录号 S220)将步骤(2)的 DNA 产物片段化至 300 bp, 待用。
- (4) 按照表 3-1 所示制备 DNA 去磷酸化反应混合液, 在 37°C温浴 30 分钟后, 95°C处理 5 分钟, 之后立即插入冰上并孵育 2 分钟后待用。

表 3-1

DNA 去磷酸反应混合液	
DNA (来自步骤(3))	17 μ l
10 x CutSmart缓冲液	2 μ l
FastAP 温敏碱性磷酸酶	1 μ l
总体积	20 μ l

- (5) 按照表 3-2 所示制备加尾和连接反应混合液, 将反应混合液在 37°C混浴 30 分钟后, 95°C处理 5 分钟, 然后保持在 4°C, 待反应完成后, 将反应混合液平均分成 4 份, 每份 10 μ l, 待用。

表 3-2

加尾和连接反应混合液	
DNA (来自步骤(4))	20 μ l
无酶水	4 μ l
2.5 x绿色缓冲液加12.5% PEG 8000	8 μ l
2 mM dATP加2 mM β -烟酰胺腺嘌呤二核苷酸	2 μ l
10 μ M控尾接头 (001/002)	2 μ l

1:33 稀释的TdT酶 (1 μl TdT酶加32 μl EB缓冲液)	2 μl
1:6 稀释的大肠杆菌DNA连接酶 (1 μl TdT酶加5 μl EB缓冲液)	2 μl
总体积	40 μl

(6) 制备如表 3-3 所示的线性扩增的反应混合液,按照表 3-4 所示的 PCR 扩增程序运行,其中 95°C, 30 秒, 60°C, 30 秒, 68°C, 1 分钟的反应循环数分别使用 4、6、8 和 12;均使用 166 μl 1: 6 稀释的 Beckman Ampure XP 磁珠 (1 体积的 Beckman Ampure XP 磁珠加上 5 体积的 SB 缓冲液)和 280 μl 的 1.8: 1 稀释的 SB 缓冲液 (1.8 体积的 SB 缓冲液加上 1 体积的无酶水)纯化回收线性扩增产物,再加入 12.5 μl EB 缓冲液洗脱,四种线性扩增循环数的 12.5 μl 洗脱液都分成 5 μl /份于 200 μl PCR 管中,都分出 2 份用于下步反应。

表 3-3

线性扩增的反应混合液	
DNA (来自步骤(5))	10 μl
无酶水	59 μl
5 x Phusion HF缓冲液	20 μl
100 μM 线性扩增引物 (003)	2 μl
2.5 mM dNTP	8 μl
Phusion U 热启动DNA聚合酶	1 μl
总体积	100 μl

表 3-4

温度	时间	循环数
95°C	3分钟	1
95°C	30秒	4或者6或者8或者12
60°C	30秒	
68°C	1分钟	
68°C	5分钟	1
4°C	保存	1

- (7) 将步骤(6)制备得到的 DNA 在 95°C 反应 5 分钟，立即置于冰上 2 分钟待用。
- (8) 按照如表 3-5 所示制备四种“5'测序接头”连接的反应混合液，分别在 25°C 温浴 15 分钟，分别使用 17 μl Beckman Ampure XP 磁珠回收连接后 DNA，再加入 26 μl 无酶水洗脱，得到 PCR 扩增前文库。

表 3-5

“5'测序接头”连接反应混合液				
线性扩增循环数	4 个	6 个	8 个	12 个
DNA (来自步骤(7))	5 μl	5 μl	5 μl	5 μl
无酶水	3.2 μl	2.8 μl	2.4 μl	1.6 μl
2 x T4 DNA 快速连接反应缓冲液	10 μl	10 μl	10 μl	10 μl
25 μM 悬 7N “5'测序接头” (012/015)	0.8 μl	1.2 μl	1.6 μl	2.4 μl
T4 DNA 快速连接酶	1 μl	1 μl	1 μl	1 μl
总体积	20 μl	20 μl	20 μl	20 μl

- (9) 按照实施例 1 所描述的方法检测 PCR 扩增前文库的摩尔浓度。

实验结果：如表 3-6 所示，线性扩增循环数为 4、6、8 和 12，PCR 扩增前文库的摩尔浓度分别为 0.0553 nM、0.0947 nM、0.131 nM 和 0.199 nM，线性扩增循环数为 12 时构建的文库浓度最高。

表 3-6

线性扩增使用循环数	4 个	6 个	8 个	12 个
文库浓度 (nM)	0.0553	0.0947	0.131	0.199

结论：对加尾的脱氧多核苷酸底物进行线性扩增，线性扩增 4-12 个循环，然后连接悬 7N “5'测序接头”，均能得到有效的甲基化文库。

实施例 4. NGS 检测不同线性扩增循环数对甲基化测序文库构建的影响

材料：

2 x 高保真热启动 PCR 混合液 (KAPA Biosystems, 目录号 KK2602)

方法：

- (1) 接头制备：按照实施例 1 中所描述的方法制备带随机分子标签的控尾接头(006/029, 如表 4 所示)，以及 3'末端悬 7 个 N 的“5'测序接头” (012/015, 如表 3 所示)。

- (2) 甲基化测序文库制备：按照实施例 3 中所描述的方法构建甲基化测序文库至“5'测序接头”连接反应后；使用 17 μl Beckman Ampure XP 磁珠回收连接后 DNA，再加入 20 μl 无酶水洗脱；取 2 μl 洗脱液于 18 μl 无酶水中稀释 10 倍，得到 10 倍稀释液，再取 2 μl 10 倍稀释液于 18 μl 无酶水中稀释 10 倍，得到 100 倍稀释液，依次进行稀释得到 10000 倍稀释液，取 10000 倍稀释液 5.34 μl ，待用。
- (3) 如表 4-1 所示制备 PCR 扩增反应混合液，按照表 4-2 所示的 PCR 扩增程序运行；使用 80 μl Beckman Ampure XP 磁珠回收 PCR 产物，再加入 50 μl 无酶水洗脱，再使用 40 μl Beckman Ampure XP 磁珠回收 50 μl 洗脱产物，最后加入 25 μl EB 缓冲液洗脱，得到最终测序文库。

表 4-1

PCR 扩增反应混合液	
DNA (来自步骤(2))	5.34 μl
无酶水	40.66 μl
50 μM MP5 PCR 标签引物 (每个文库使用引物 018、019、020、021、022 和 023 中的一种，每种引物浓度为 50 μM)	2 μl (需单加)
25 μM MP7 PCR 标签引物 (每个文库使用引物 024 和 025 中的一种，每种引物浓度为 25 μM)	2 μl (需单加)
2 x 高保真热启动 PCR 混合液	50 μl
总体积	100 μl

表 4-2

PCR 扩增程序		
温度	时间	循环数
98°C	45 秒	1
98°C	15 秒	28
60°C	30 秒	
72°C	30 秒	
72°C	5 分钟	1
4°C	保存	1

- (4) 使用安捷伦 2100 生物分析仪 (Agilent Technologies, 目录号 G2939BA) 和安捷伦高灵

敏 DNA 试剂盒 (Agilent Technologies, 目录号 5067-4626)检测文库片段分布; 使用文库定量试剂盒 (KAPA Biosystems, 目录号 KK4808)以及 DNA 定量标准品和预混引物试剂盒 (KAPA Biosystems, 目录号 KK4808)检测文库摩尔浓度。

(5) 使用Illumina-NS500测序仪对步骤(3)得到的文库进行75PE模式测序, 利用软件Cutadapt (v1.12) 去除接头序列; 采用软件Bwa-Meth (v0.2.0) 对甲基化测序序列进行基因组比对; 利用软件包Sambamba (v0.5.4) 标记重复序列; 最后, 利用软件包bedtools (v2.25.0) 对测序深度进行统计。

实验结果如下表 4-3 所示, 线性扩增循环数为 4、6、8 和 12 时, 构建的甲基化测序文库的浓度分别为 32.22 nM, 54.90 nM, 134.24 nM 和 139.95 nM。高通量测序结果显示的测序数据量为 4 Mb 时, 建库效率分别为 30.35%, 48.36%, 70.36%和 92.63%, 变异系数分别为 0.455, 0.275, 0.999 和 0.637。

其中, 从 1000 个模板 λ -DNA 基因组起始建库, 通过测序, 最终在每个 λ -DNA 基因组位点上平均得到 100 个独一无二的测序片段, 即称为建库效率 10%。对每个 λ -DNA 基因组位点上得到的独一无二的测序片段数进行统计, 计算其平均值及标准差, 以标准差除以平均值, 得“变异系数”。变异系数代表建库方法的均一性, 越低则均一性越好。建库效率指代建库方法的有效性, 越高越好。

结论: 线性扩增循环数为 4、6、8 和 12 均能有效的完成线性扩增, 并对亚硫酸氢盐处理后的 DNA 构建甲基化测序文库; 线性扩增循环数为 12 时, 建库效率最高, 达 92.63%。

表 4-3

DNA 类型	λ -DNA	λ -DNA	λ -DNA	λ -DNA
DNA 投入量	2 ng	2 ng	2 ng	2 ng
线性扩增循环数	4	6	8	12
PCR 扩增时 DNA 投入量	步骤(2)记载的 20 μ l 连接纯化产物的 10000 倍稀释液取 5.34 μ l	步骤(2)记载的 20 μ l 连接纯化产物的 10000 倍稀释液取 5.34 μ l	步骤(2)记载的 20 μ l 连接纯化产物的 10000 倍稀释液取 5.34 μ l	步骤(2)记载的 20 μ l 连接纯化产物的 10000 倍稀释液取 5.34 μ l
文库浓度 (nM)	32.22	54.90	134.24	139.95
测序仪	Illumina-NS500	Illumina-NS500	Illumina-NS500	Illumina-NS500

测序模式	75PE	75PE	75PE	75PE
测序数据量	4 Mb	4 Mb	4 Mb	4 Mb
建库效率	30.35%	48.36%	70.36%	92.63%
变异系数	0.455	0.275	0.999	0.637

实施例 5

比较本发明方法与传统方法以及 Swift 方法在构建甲基化测序文库上效率的差异

材料:

10 x 末端修复缓冲液 (New England Biolabs, 目录号B6052S, 50 mM Tris-盐酸、10 mM 氯化镁、10 mM二硫苏糖醇、1 mM三磷酸腺苷、0.4 mM dATP、0.4 mM dCTP、0.4 mM dGTP、0.4 mM dTTP, pH 7.5)

T4 DNA聚合酶 (Enzymatics, 目录号P7080L, 3 U/ μ L)

T4多聚核苷酸激酶 (Enzymatics, 目录号Y9040L, 10 U/ μ L)

10 x dA加尾缓冲液 (New England Biolabs, 目录号B6059S, 10 mM Tris-盐酸、10 mM 氯化镁、50 mM氯化钠、1 mM二硫苏糖醇、0.2 mM dATP, pH 7.9)

Klenow大片段 (Exo-) (Enzymatics, 目录号P7010-LC-L, 10 U/ μ L)

(1) 利用本发明方法构建 λ -DNA 甲基化测序文库

(1-1) 接头制备: 按照实施例1中接头制备方法制备控尾接头 (001/002, 如表2所示)以及 3'末端悬7个N的“5'测序接头”(012/015, 如表3所示)。

(1-2) 使用亚硫酸氢盐处理试剂盒对 20 ng λ -DNA 进行亚硫酸氢盐处理。

(1-3) 使用聚焦超声仪 (Covaris, 目录号 S220)将步骤(1-2)DNA 产物片段化至 300 bp, 均分成两份 (即两个平行), 待用。

(1-4) 按照表 5-1 所示制备 DNA 去磷酸化反应混合液, 在 37°C温浴 30 分钟后, 95°C处理 5 分钟, 之后立即插入冰上并孵育 2 分钟后待用。

表 5-1

DNA 去磷酸反应混合液	
DNA (来自步骤(1-3))	17 μ l
10 x CutSmart缓冲液	2 μ l

FastAP 温敏碱性磷酸酶	1 μ l
总体积	20 μ l

(1-5) 按照表 5-2 所示制备加尾和连接反应混合液，将反应混合液在 37°C 混浴 30 分钟后，95°C 处理 5 分钟，然后保持在 4°C，待用。

表 5-2

加尾和连接反应混合液	
DNA (来自步骤(1-4))	20 μ l
无酶水	4 μ l
2.5 x 绿色缓冲液加 12.5% PEG 8000	8 μ l
2 mM dATP 加 2 mM β -烟酰胺腺嘌呤二核苷酸	2 μ l
10 μ M 控尾接头 (001/002)	2 μ l
1:33 稀释的 TdT 酶 (1 μ l TdT 酶加 32 μ l EB 缓冲液)	2 μ l
1:6 稀释的大肠杆菌 DNA 连接酶 (1 μ l TdT 酶加 5 μ l EB 缓冲液)	2 μ l
总体积	40 μ l

(1-6) 制备如表 5-3 所示的线性扩增的反应混合液，按照表 5-4 所示的 PCR 扩增程序运行，反应完成后使用 166 μ l 1:6 稀释的 Beckman Ampure XP 磁珠 (1 体积的 Beckman Ampure XP 磁珠加上 5 体积的 SB 缓冲液) 和 280 μ l 的 1.8:1 稀释的 SB 缓冲液 (1.8 体积的 SB 缓冲液加上 1 体积的无酶水) 纯化回收线性扩增产物，用 6.6 μ l EB 缓冲液洗脱。

表 5-3

线性扩增的反应混合液	
DNA (来自步骤(1-5))	40 μ l
无酶水	29 μ l
5 x Phusion HF 缓冲液	20 μ l
100 μ M 线性扩增引物 (003)	2 μ l
2.5 mM dNTP	8 μ l
Phusion U 热启动 DNA 聚合酶	1 μ l

总体积	100 μl
-----	-------------------

表 5-4

温度	时间	循环数
95°C	3分钟	1
95°C	30秒	12
60°C	30秒	
68°C	1分钟	
68°C	5分钟	1
4°C	保存	1

(1-7) 将步骤(1-6)的 DNA 洗脱液在 95°C 反应 5 分钟，立即置于冰上 2 分钟待用。

(1-8) 按照如表 5-5 所示制备“5'测序接头”连接的反应混合液，在 25°C 温浴 15 分钟，使用 17 μl Beckman Ampure XP 磁珠回收连接后 DNA，再加入 100 μl 无酶水洗脱；取 2 μl 洗脱液于 18 μl 无酶水中稀释 10 倍，得到 10 倍稀释液，再取 2 μl 10 倍稀释液于 18 μl 无酶水中稀释 10 倍，得到 100 倍稀释液，依次进行稀释得到 10000 倍稀释液，取 10000 倍稀释液 5.34 μl ，待用。

表 5-5

DNA (来自步骤(1-7))	6.6 μl
2 x T4 DNA 快速连接反应缓冲液	10 μl
25 μM 悬 7N “5'测序接头” (012/015)	2.4 μl
T4 DNA 快速连接酶	1 μl
总体积	20 μl

(1-9) 如表 5-6 所示制备 PCR 扩增反应混合液，按照表 5-7 所示的 PCR 扩增程序运行；使用 80 μl Beckman Ampure XP 磁珠回收 PCR 产物，再加入 50 μl 无酶水洗脱，再使用 40 μl Beckman Ampure XP 磁珠回收 50 μl 洗脱产物，最后加入 25 μl EB 缓冲液洗脱，得到最终测序文库。

表 5-6

PCR 扩增反应混合液

DNA (来自步骤(1-8))	5.34 μl
无酶水	40.66 μl
50 μM P5 PCR 标签引物 (023)	2 μl (需单加)
25 μM P7 PCR 标签引物 (每个文库使用引物 024 或 025, 每种引物的浓度为 25 μM)	2 μl (需单加)
2 x 高保真热启动 PCR 混合液	50 μl
总体积	100 μl

表 5-7

PCR 扩增程序		
温度	时间	循环数
98°C	45 秒	1
98°C	15 秒	28
60°C	30 秒	
72°C	30 秒	
72°C	5 分钟	1
4°C	保存	1

(1-10) 文库浓度检测方法同实施例 4.

(1-11) 使用 Illumina-Nova 测序仪对步骤(1-9)得到的文库进行 150PE 模式测序, 分析方法同实施例 4 中的步骤(5)。

(2) 利用传统方法构建 λ -DNA 甲基化测序文库 (建库流程示意图如图 4)

(2-1) 接头制备, 按照实施例 1 中接头制备方法制备“传统甲基化测序接头”(016/017, 如表 5 所示)。

(2-2) 取 20 ng λ -DNA, 使用聚焦超声仪将 DNA 片段化至 300 bp, 均分成两份 (即两个平行), 待用。

(2-3) 如表 5-8 所示配制末端修复反应混合液, 在 20°C 下反应 30 分钟, 然后加入 45 μl Beckman Ampure XP 磁珠回收修复后 DNA, 使用 26 μl 无酶水洗脱。

表 5-8

末端修复反应混合液

DNA (来自步骤(2-2))	24 μ l
10 x 末端修复缓冲液	3 μ l
T4 DNA聚合酶	1.5 μ l
T4多聚核苷酸激酶	1.5 μ l
总体积	30 μ l

(2-4) 如表 5-9 所示配制 dA 加尾反应混合液，在 37°C 下反应 30 分钟，然后加入 45 μ l Beckman Ampure XP 磁珠回收完成 dA 加尾的 DNA，使用 12 μ l 无酶水洗脱。

表 5-9

dA加尾反应混合液	
DNA (来自步骤(2-3))	26 μ l
10 x dA加尾反应缓冲液	3 μ l
Klenow大片段 (Exo-)	1 μ l
总体积	30 μ l

(2-5) 如表 5-10 所示配制接头连接反应混合液，在 25°C 下反应 15 分钟，然后加入 21 μ l Beckman Ampure XP 磁珠回收完成接头连接的 DNA，使用 20 μ l 无酶水洗脱。

表 5-10

接头连接反应混合液	
DNA (来自步骤(2-4))	12 μ l
2 x T4 DNA快速连接缓冲液	15 μ l
1.5 μ M “传统甲基化测序接头” (016/017)	2 μ l
T4 DNA快速连接酶	1 μ l
总体积	30 μ l

(2-6) 使用亚硫酸氢盐处理试剂盒对 DNA (来自步骤(2-5)) 进行亚硫酸氢盐处理，使用 100 μ l 无酶水洗脱；取 2 μ l 洗脱液按照步骤(1-8)所描述的稀释方法进行稀释，得到 10000 倍稀释液，取 10000 倍稀释液 5.34 μ l，待用。

(2-7) 如表 5-11 所示制备 PCR 扩增反应混合液，按照表 5-12 所示的 PCR 扩增程序运行；使用 40 μ l Beckman Ampure XP 磁珠回收 PCR 产物，再加入 50 μ l 无酶水洗脱，再使用 40 μ l Beckman Ampure XP 磁珠回收 50 μ l 洗脱产物，最后加入 25 μ l EB 缓冲液洗脱，得到最终测序文库。

表5-11

PCR扩增反应混合液	
DNA (来自步骤(2-6))	5.34 μ l
无酶水	17.66 μ l
2 x 高保真热启动甲基化PCR混合液	25 μ l
20 μ M P5 PCR引物 (026)	1 μ l
20 μ M P7 PCR标签引物 (每个文库使用引物027或028, 每种引物的浓度为20 μ M)	1 μ l
总体积	50 μ l

表5-12

PCR扩增程序		
温度	时间	循环数
98°C	45秒	1
98°C	15秒	28
62°C	30秒	
72°C	30秒	
72°C	2分钟	1
4°C	保存	1

(2-8) 文库浓度检测方法同实施例 4。

(2-9) 使用 Illumina-Nova 测序仪对步骤(2-7)得到的文库进行 150PE 模式测序, 分析方法同实施例 4 中的步骤(5)。

(3) 利用 Swift 方法构建 λ -DNA 甲基化测序文库 (建库流程示意图如图 5)

(3-1) 使用亚硫酸氢盐处理试剂盒对 20 ng λ -DNA 进行亚硫酸氢盐处理。

(3-2) 使用聚焦超声仪将步骤(3-1)产物片段化至 300 bp, 均分成两份 (即两个平行), 待用。

(3-3) 按照 Swift DNA 甲基化建库试剂盒 (Swift Biosciences, 目录号 30024)的建库步骤对来自步骤(3-2)的 DNA 进行甲基化文库的构建, 不同之处在文库构建至 5'测序接头连接完纯化后 (Index PCR 之前)用 100 μ l 无酶水洗脱; 取 2 μ l 洗脱液按照步骤 1-8 所描述的稀释方法进行稀释, 得到 10000 倍稀释液, 取 10000 倍稀释液 5.34 μ l, 补无酶水 14.66 μ l 至终体积为 20 μ l, 待用。

(3-4) 按照 Swift DNA 甲基化建库试剂盒 (Swift Biosciences, 目录号 30024)中的 Index PCR 步骤对来自步骤(3-3)的 DNA 进行 PCR 扩增, 使用的 Index 为 Index 试剂盒 (Swift

Biosciences, 目录号 36024)中的 I16 或 I19, 不同之处在 PCR 扩增循环数为 28。

(3-5) 按照步骤(2-7)所示的 PCR 产物纯化方法对来自步骤(3-4)的 PCR 产物进行纯化回收, 得到最终测序文库。

(3-6) 按照实施例 4 所描述的方法检测文库摩尔浓度。

(3-7) 使用 Illumina-Nova 测序仪对步骤(2-7)得到的文库进行 150PE 模式测序, 分析方法同实施例 4 中的步骤(5)。

实验结果如表 5-13 所示, 分别用本发明方法, Swift 方法和传统建库方法, 构建的甲基化测序文库的浓度分别为 18.684 nM, 1.641 nM 和 0.146 nM。测序数据量为 4 Mb 时, 建库效率分别为 56.1%, 22%和 3.92%, 变异系数分别为 0.463, 8.49, 和 3.73。

表 5-13

建库方法	本发明方法	Swift 方法	传统方法
DNA 类型	λDNA	λDNA	λDNA
建库起始 DNA 投入量	10 ng	10 ng	10 ng
PCR 实际投入量	步骤(1-8)记载的 100 μl 连接纯化产物的 10000 倍稀释液取 5.34 μl	步骤(2-6)记载的 100 μl 连接纯化产物的 10000 倍稀释液取 5.34 μl	步骤(3-3)记载的 100 μl 连接纯化产物的 10000 倍稀释液取 5.34 μl
亚硫酸氢盐处理时间点	建库前	建库前	建库后
线性扩增循环数	12	1	无
5'测序接头	悬 7 个随机碱基	悬 T	悬 T
文库浓度	18.684 nM	1.641 nM	0.146 nM
测序仪	Illumina-nova	Illumina-nova	Illumina-nova
测序模式	150 PE	150 PE	150 PE
测序数据量	4 Mb	4 Mb	4 Mb
建库效率	56.1%	22%	3.92%
变异系数	0.463	8.49	3.73

结论: 使用本发明方法能够高效的对少量基因组 DNA 进行甲基化测序文库构建, 建库效率和建库方法的均一性都远远优于 Swift 方法和传统方法。

权利要求书

1. 一种对脱氧多核苷酸底物进行建库的方法，所述方法包括如下步骤：
 - (1) 将所述脱氧多核苷酸单链底物与如下物质混合以形成第一混合物：a) 选自 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸；b) 末端脱氧核苷酸转移酶和 DNA 连接酶；c) 控尾组分，其中该控尾组分是由 5 至 20 个核苷酸长度的多核苷酸同聚物和 X 区，以及与 X 区互补的连接子多核苷酸组成的部分双链核苷酸分子；其中多核苷酸同聚物与 a) 中的脱氧核苷酸互补；
 - (2) 孵育所述的第一混合物，脱氧多核苷酸单链底物的 3' 端与溶液中的脱氧核苷酸发生加尾反应，添加了多核苷酸同聚尾的底物 3' 端与控尾组分的连接子连接，得到加尾后的底物；
 - (3) 在步骤(2)的反应体系中，添加 DNA 聚合酶、包含 dGTP、dCTP、dATP 和 dTTP 的脱氧核苷酸和线性扩增引物，以形成第二混合物；
 - (4) 孵育所述第二混合物，以步骤(2)得到的加尾后的底物为模板进行第一次线性延伸反应，合成底物的互补链，再解链，线性扩增引物与底物互补后，再次进行后续的线性延伸反应，其中线性延伸反应的次数不低于 3 次；
 - (5) 使步骤(4)的产物解链；
 - (6) 在步骤(5)的溶液中添加 5' 测序接头和 DNA 连接酶形成第三混合物；
 - (7) 孵育所述的第三混合物，5' 测序接头与底物互补链结合，制备得到 DNA 文库。
2. 如权利要求 1 所述的方法，其中 a) 中的脱氧多核苷酸是 dATP。
3. 前述任一项权利要求所述的方法，其中步骤(1)中的控尾组分包含 5 至 13 个核苷酸长度的多核苷酸同聚物。
4. 前述任一项权利要求所述的方法，其中步骤(1)中的控尾组分包含 7 至 10 个核苷酸长度的多核苷酸同聚物。
5. 前述任一项权利要求所述的方法，其中步骤(2)在温度 20°C-50°C 下进行，优选 25°C-45°C，更优选 25°C-37°C。
6. 前述任一项权利要求所述的方法，其中步骤(3)的线性扩增引物与加尾后的底物的 3' 末端互补。

7. 权利要求 6 所述的方法，其中步骤(3)的线性扩增引物与控尾组分的连接子互补。

8. 前述任一项权利要求所述的方法，其中在步骤(4)的第一次线性延伸反应中，所用的引物是控尾分子。

9. 前述权利要求 1-7 任一项所述的方法，其中在步骤(4)的第一次线性延伸反应中，所用的引物是步骤(3)中添加的线性扩增引物。

10. 权利要求 9 所述的方法，其中在步骤(4)的第一次线性延伸反应中，控尾组分中与底物互补的序列被降解或被步骤(3)中添加的线性扩增引物所竞争。

11. 前述任一项权利要求所述的方法，其中步骤(4)的线性延伸反应的次数不低于 4 次，优选为 4-50 次，进一步优选 4-20 次，更优选为 4-12 次。

12. 前述任一项权利要求所述的方法，其中步骤(6)和(7)的 5'测序接头的一条链的 3'末端悬了 0-50 个随机碱基的单链多核苷酸。

13. 权利要求 12 所述的方法，其中步骤(6)和(7)的 5'测序接头的一条链的 3'末端悬了 0-30 个随机碱基的单链多核苷酸。

14. 权利要求 13 所述的方法，其中 5'测序接头的一条链的 3'末端悬了 2-30 个随机碱基的单链多核苷酸，优选 2-17 个，更优选 4-15 个，更优选 7-10 个。

15. 前述任一项权利要求所述的方法，其特征在于控尾组分的多核苷酸同聚物和连接子包含 3'封闭基团。

16. 前述任一项权利要求所述的方法，其特征在于 5'测序接头包含 3'封闭基团。

17. 前述任一项权利要求所述的方法，封闭基团选自以下组分的一种或几种：核糖核苷酸、碳三间臂、磷酸、双脱氧核苷酸、氨基基团和倒置脱氧胸腺嘧啶核苷。

18. 前述任一项权利要求所述的方法，控尾组分中的连接子多核苷酸包含 5'端磷酸和 3'端封闭基团。

19. 前述任一项权利要求所述的方法，5'测序接头中与含有随机碱基链的互补链上包含 5'端磷酸。

20. 前述任一项权利要求所述的方法，其中，脱氧多核苷酸底物是去磷酸化的多核苷酸序列。

21. 前述任一项权利要求所述的方法，其中在步骤(4)之后，从所述的第二混合物中分离双链及单链脱氧多核苷酸，解链后得到单链脱氧多核苷酸用于后续步骤。

22. 前述任一项权利要求所述的方法，其中在步骤(7)之后，对制备得到的 DNA 进行

PCR 扩增。

23. 权利要求 22 所述的方法，其中在步骤(7)之后，对制备得到的 DNA 进行纯化，纯化后再进行 PCR 扩增。

24. 一种试剂盒，其包含：

组分一：包含选自 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸、末端脱氧核苷酸转移酶、DNA 连接酶和控尾组分，其中所述的控尾组分由 5 至 20 个核苷酸长度的多核苷酸同聚物和 X 区，以及与 X 区互补的连接子多核苷酸组成的部分双链核苷酸分子，该多核苷酸同聚物与所述的选自 dGTP、dCTP、dATP 和 dGTP 中的一种脱氧核苷酸互补；

组分二：包含 DNA 聚合酶、包含 dGTP、dCTP、dATP 和 dTTP 的脱氧核苷酸，和线性扩增引物；

组分三：包含 5'测序接头和 DNA 连接酶。

25. 权利要求 24 所述的试剂盒，其中组分一中选自 dGTP、dCTP、dATP 和 dTTP 中的一种脱氧核苷酸是 dATP。

26. 权利要求 24 或 25 所述的试剂盒，其中的控尾组分包含 5 至 13 个，7 至 10 个，更优选 7 至 9 个核苷酸长度的多核苷酸同聚物。

27. 权利要求 24-26 任一项所述的试剂盒，其中所述的线性扩增引物与控尾组分的 3' 端互补，优选与控尾组分的连接子互补。

28. 权利要求 24-27 任一项所述的试剂盒，其中所述的 5'测序接头的一条链的 3' 末端悬了 0-50 个随机碱基的单链多核苷酸，优选 2-30，优选 2-17 个，更优选 4-15 个，更优选 7-10 个随机碱基的单链多核苷酸。

29. 权利要求 24-28 任一项所述的试剂盒在对脱氧多核苷酸底物进行建库中的应用。

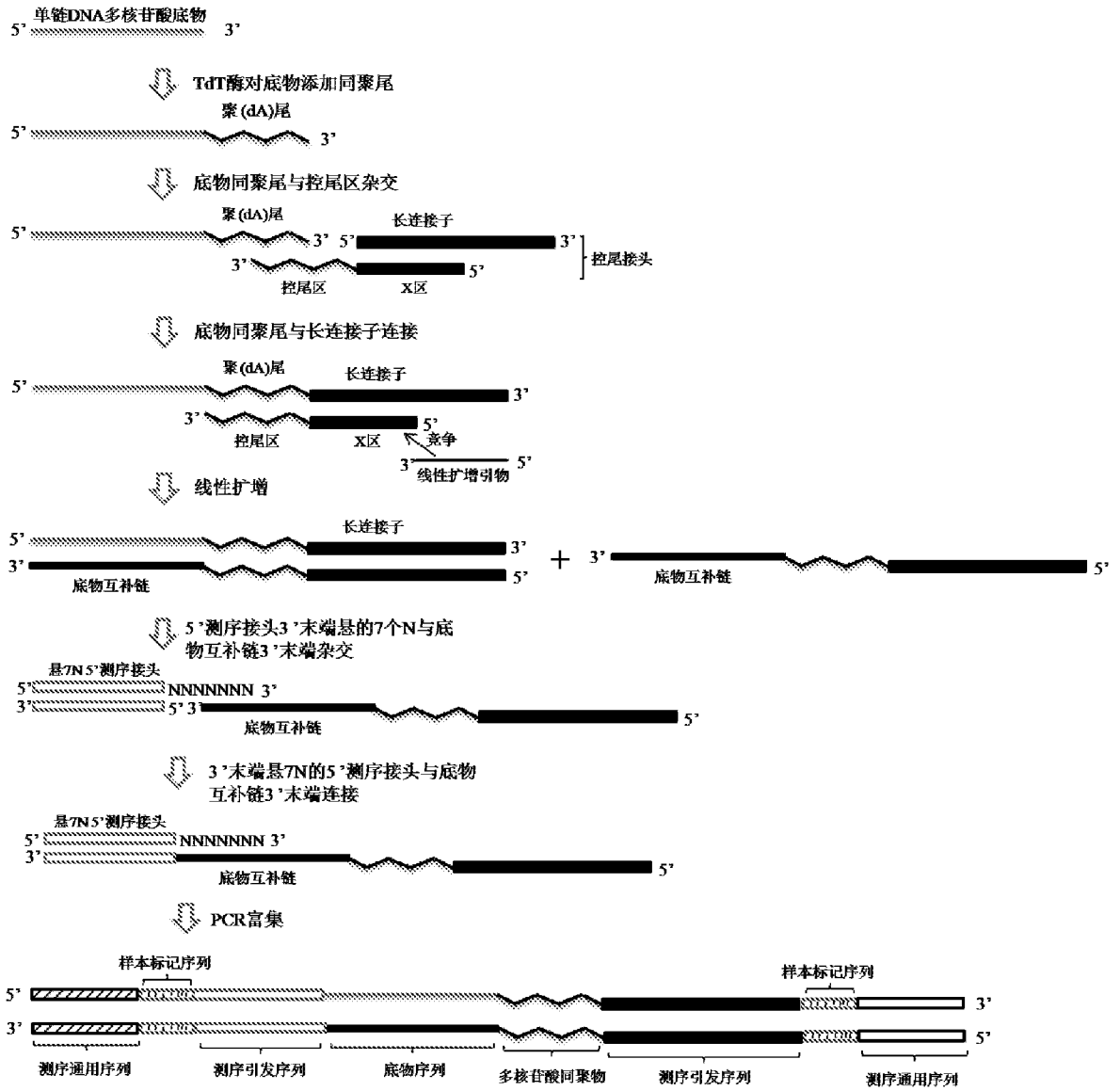


图 1

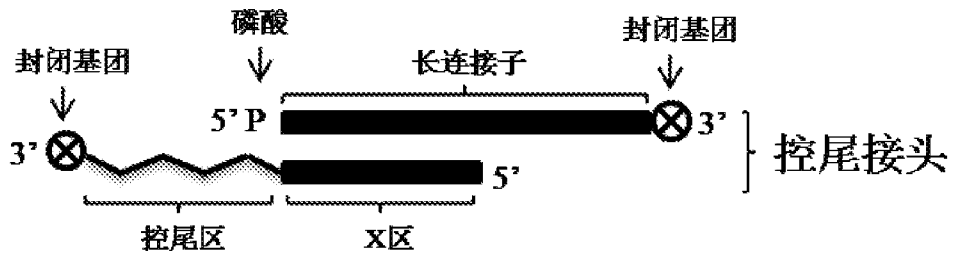


图2

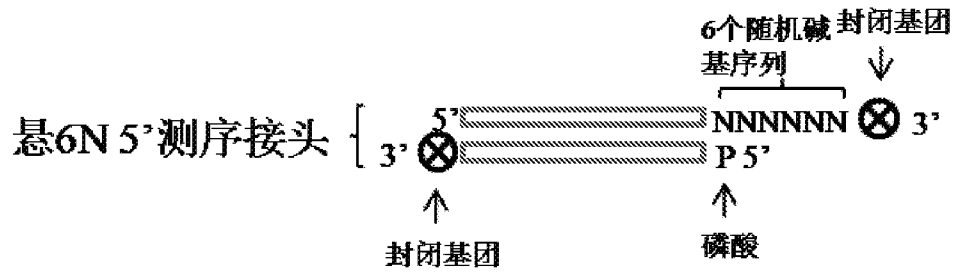


图3

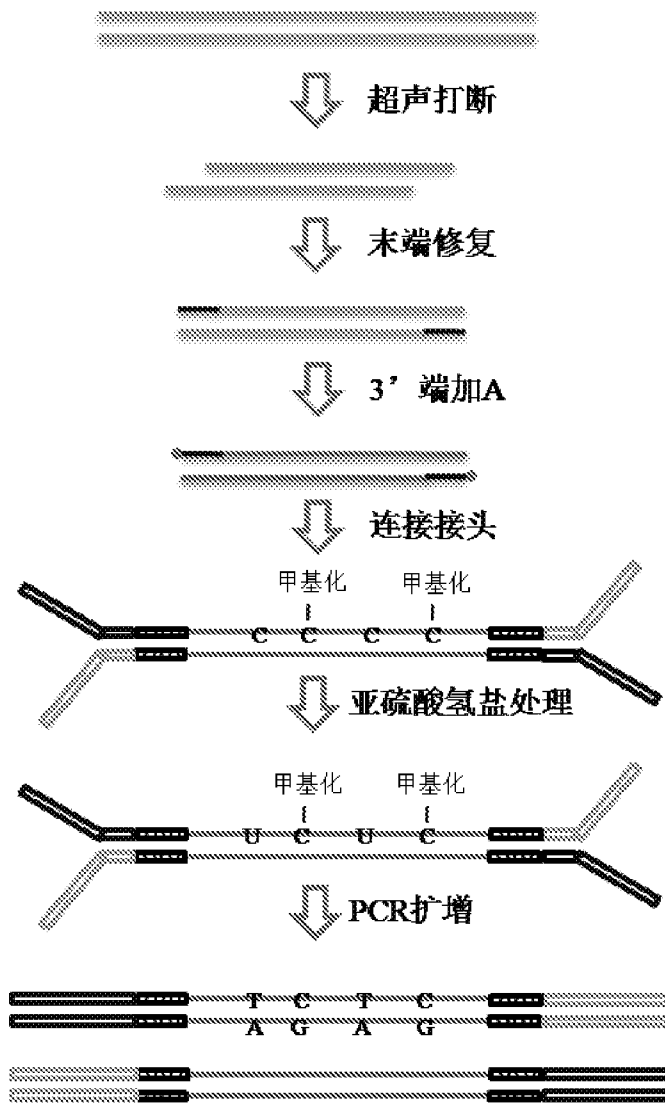


图4

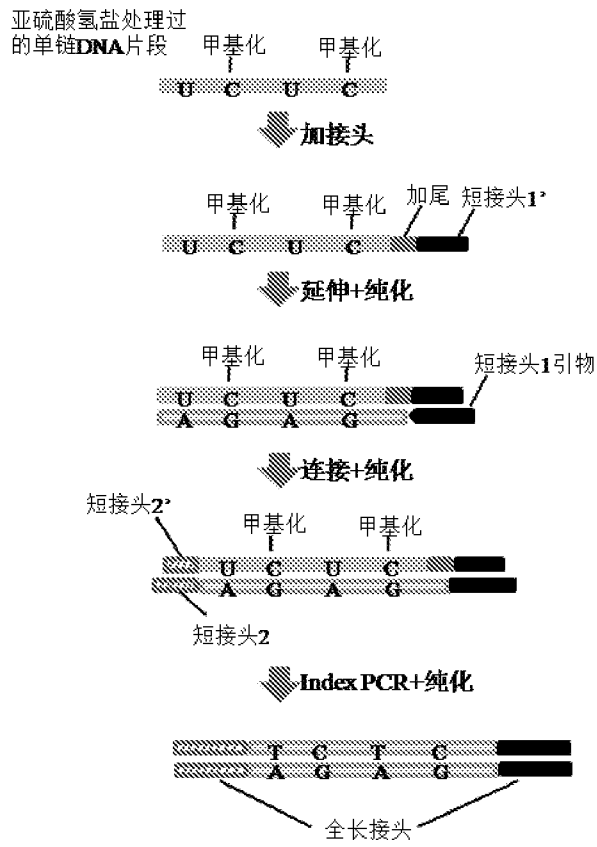


图5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102651

A. CLASSIFICATION OF SUBJECT MATTER C12Q 1/68(2018.01)i; C40B 50/06(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12Q; C40B Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNABS, VEN, CNKI, CNTXT, USTXT, GOOGLE, PUBMED, 文库, 二代测序, 末端脱氧核苷酸转移酶, DNA连接酶, DNA聚合酶, 线性扩增, 加尾, 去磷酸化, 控尾, library, next-generation sequencing, NGS, terminal deoxynucleotidyl transferase, TDT, DNA ligase, DNA polymerase, linear amplification, tailing, dephosphorylation		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104395480 A (SWIFT BIOSCIENCES INC.) 04 March 2015 (2015-03-04) description, paragraphs [0051]-[0053], [0113], [0150] and [0217]-[0224]	1-29
A	CN 108456713 A (TIANJIN NOVOGENE BIOINFORMATICS TECHNOLOGY CO., LTD.) 28 August 2018 (2018-08-28) claims 1-9	1-29
A	US 2010221785 A1 (HUMAN GENETIC SIGNATURES PTY LTD.) 02 September 2010 (2010-09-02) entire document	1-29
A	CN 108004301 A (GENOSABER BIOTECHNOLOGY NANTONG CO., LTD.; SHANGHAI GENO BIOTECHNOLOGY CO., LTD.) 08 May 2018 (2018-05-08) entire document	1-29
A	CN 105525357 A (BGI SHENZHEN CO., LTD.) 27 April 2016 (2016-04-27) entire document	1-29
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 08 November 2019		Date of mailing of the international search report 04 December 2019
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/ CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China		Authorized officer
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/102651

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105442054 A (HANGZHOU GUKUN BIOTECHNOLOGY CO., LTD.) 30 March 2016 (2016-03-30) entire document	1-29
A	CN 104313699 A (TIANJIN NOVOGENE BIOINFORMATICS TECHNOLOGY CO., LTD.) 28 January 2015 (2015-01-28) entire document	1-29
A	Van DIJK, E. L. "Library Preparation Methods for Next-generation Sequencing: Tone down the Bias" <i>CELL RESEARCH</i> , Vol. 322, 15 January 2014 (2014-01-15), pp. 12-20	1-29

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/102651

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	104395480	A	04 March 2015	CN	108300765	A	20 July 2018
				HK	1206395	A1	08 January 2016
				CA	2866625	A1	19 September 2013
				AU	2013232131	B2	20 September 2018
				US	9896709	B2	20 February 2018
				JP	2018130124	A	23 August 2018
				LT	2825672	T	10 June 2019
				EP	2825672	B1	13 February 2019
				SG	11201405669 X	A	30 October 2014
				CN	104395480	B	30 January 2018
				JP	6509723	B2	08 May 2019
				JP	2015510766	A	13 April 2015
				US	2018223321	A1	09 August 2018
				US	2015087027	A1	26 March 2015
				AU	2013232131	A1	18 September 2014
				EP	2825672	A4	18 November 2015
				WO	2013138536	A1	19 September 2013
				EP	2825672	A1	21 January 2015
<hr/>							
CN	108456713	A	28 August 2018	None			
<hr/>							
US	2010221785	A1	02 September 2010	EP	1883707	A4	05 August 2009
				US	8431347	B2	30 April 2013
				CA	2609218	A1	30 November 2006
				CN	101203618	A	18 June 2008
				WO	2006125267	A1	30 November 2006
				CN	101203618	B	13 March 2013
				EP	1883707	B1	09 April 2014
				CA	2609218	C	11 October 2016
				AU	2006251866	A1	30 November 2006
				JP	2008541705	A	27 November 2008
				EP	1883707	A1	06 February 2008
				AU	2006251866	B2	29 November 2007
<hr/>							
CN	108004301	A	08 May 2018	WO	2019114146	A1	20 June 2019
<hr/>							
CN	105525357	A	27 April 2016	HK	1223660	A1	04 August 2017
				CN	105525357	B	21 August 2018
<hr/>							
CN	105442054	A	30 March 2016	CN	105442054	B	03 April 2018
<hr/>							
CN	104313699	A	28 January 2015	None			
<hr/>							

<p>A. 主题的分类</p> <p>C12Q 1/68(2018.01)i; C40B 50/06(2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																																					
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>C12Q; C40B</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNABS, VEN, CNKI, CNTXT, USTXT, GOOGLE, PUBMED, 文库, 二代测序, 末端脱氧核苷酸转移酶, DNA连接酶, DNA聚合酶, 线性扩增, 加尾, 去磷酸化, 控尾, library, next-generation sequencing, NGS, terminal deoxynucleotidyl transferase, TDT, DNA ligase, DNA polymerase, linear amplification, tailing, dephosphorylation.</p>																																					
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 104395480 A (斯威夫特生物科学公司) 2015年 3月 4日 (2015 - 03 - 04) 说明书[0051]- [0053]、[0113]、[150]、[0217]- [0224]段</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>CN 108456713 A (天津诺禾致源生物信息科技有限公司) 2018年 8月 28日 (2018 - 08 - 28) 权利要求1-9</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>US 2010221785 A1 (HUMAN GENETIC SIGNATURES PTY LTD.) 2010年 9月 2日 (2010 - 09 - 02) 全文</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>CN 108004301 A (格诺思博生物科技南通有限公司 上海格诺生物科技有限公司) 2018年 5月 8日 (2018 - 05 - 08) 全文</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>CN 105525357 A (深圳华大基因股份有限公司) 2016年 4月 27日 (2016 - 04 - 27) 全文</td> <td>1-29</td> </tr> <tr> <td>A</td> <td>CN 105442054 A (杭州谷坤生物技术有限公司) 2016年 3月 30日 (2016 - 03 - 30) 全文</td> <td>1-29</td> </tr> </tbody> </table> <p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <table border="0"> <tr> <td>* 引用文件的具体类型:</td> <td>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</td> </tr> <tr> <td>“A” 认为不特别相关的表示了现有技术一般状态的文件</td> <td>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</td> </tr> <tr> <td>“D” 申请人在国际申请中引证的文件</td> <td>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</td> </tr> <tr> <td>“E” 在国际申请日的当天或之后公布的在先申请或专利</td> <td>“&” 同族专利的文件</td> </tr> <tr> <td>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</td> <td></td> </tr> <tr> <td>“O” 涉及口头公开、使用、展览或其他方式公开的文件</td> <td></td> </tr> <tr> <td>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</td> <td></td> </tr> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 104395480 A (斯威夫特生物科学公司) 2015年 3月 4日 (2015 - 03 - 04) 说明书[0051]- [0053]、[0113]、[150]、[0217]- [0224]段	1-29	A	CN 108456713 A (天津诺禾致源生物信息科技有限公司) 2018年 8月 28日 (2018 - 08 - 28) 权利要求1-9	1-29	A	US 2010221785 A1 (HUMAN GENETIC SIGNATURES PTY LTD.) 2010年 9月 2日 (2010 - 09 - 02) 全文	1-29	A	CN 108004301 A (格诺思博生物科技南通有限公司 上海格诺生物科技有限公司) 2018年 5月 8日 (2018 - 05 - 08) 全文	1-29	A	CN 105525357 A (深圳华大基因股份有限公司) 2016年 4月 27日 (2016 - 04 - 27) 全文	1-29	A	CN 105442054 A (杭州谷坤生物技术有限公司) 2016年 3月 30日 (2016 - 03 - 30) 全文	1-29	* 引用文件的具体类型:	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件	“A” 认为不特别相关的表示了现有技术一般状态的文件	“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性	“D” 申请人在国际申请中引证的文件	“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性	“E” 在国际申请日的当天或之后公布的在先申请或专利	“&” 同族专利的文件	“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)		“O” 涉及口头公开、使用、展览或其他方式公开的文件		“P” 公布日先于国际申请日但迟于所要求的优先权日的文件	
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																																			
A	CN 104395480 A (斯威夫特生物科学公司) 2015年 3月 4日 (2015 - 03 - 04) 说明书[0051]- [0053]、[0113]、[150]、[0217]- [0224]段	1-29																																			
A	CN 108456713 A (天津诺禾致源生物信息科技有限公司) 2018年 8月 28日 (2018 - 08 - 28) 权利要求1-9	1-29																																			
A	US 2010221785 A1 (HUMAN GENETIC SIGNATURES PTY LTD.) 2010年 9月 2日 (2010 - 09 - 02) 全文	1-29																																			
A	CN 108004301 A (格诺思博生物科技南通有限公司 上海格诺生物科技有限公司) 2018年 5月 8日 (2018 - 05 - 08) 全文	1-29																																			
A	CN 105525357 A (深圳华大基因股份有限公司) 2016年 4月 27日 (2016 - 04 - 27) 全文	1-29																																			
A	CN 105442054 A (杭州谷坤生物技术有限公司) 2016年 3月 30日 (2016 - 03 - 30) 全文	1-29																																			
* 引用文件的具体类型:	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件																																				
“A” 认为不特别相关的表示了现有技术一般状态的文件	“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性																																				
“D” 申请人在国际申请中引证的文件	“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性																																				
“E” 在国际申请日的当天或之后公布的在先申请或专利	“&” 同族专利的文件																																				
“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)																																					
“O” 涉及口头公开、使用、展览或其他方式公开的文件																																					
“P” 公布日先于国际申请日但迟于所要求的优先权日的文件																																					
国际检索实际完成的日期	国际检索报告邮寄日期																																				
2019年 11月 8日	2019年 12月 4日																																				
ISA/CN的名称和邮寄地址	授权官员																																				
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	彭海航																																				
传真号 (86-10)62019451	电话号码 86-(010)-53961949																																				

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 104313699 A (天津诺禾致源生物信息科技有限公司) 2015年 1月 28日 (2015 - 01 - 28) 全文	1-29
A	van DIJK, E. L. "Library preparation methods for next-generation sequencing:Tone down the bias" CELL RESEARCH, 第322卷, 2014年 1月 15日 (2014 - 01 - 15), 第12-20页	1-29

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/102651

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	104395480	A	2015年 3月 4日	CN 108300765 A	2018年 7月 20日
				HK 1206395 A1	2016年 1月 8日
				CA 2866625 A1	2013年 9月 19日
				AU 2013232131 B2	2018年 9月 20日
				US 9896709 B2	2018年 2月 20日
				JP 2018130124 A	2018年 8月 23日
				LT 2825672 T	2019年 6月 10日
				EP 2825672 B1	2019年 2月 13日
				SG 11201405669X A	2014年 10月 30日
				CN 104395480 B	2018年 1月 30日
				JP 6509723 B2	2019年 5月 8日
				JP 2015510766 A	2015年 4月 13日
				US 2018223321 A1	2018年 8月 9日
				US 2015087027 A1	2015年 3月 26日
				AU 2013232131 A1	2014年 9月 18日
				EP 2825672 A4	2015年 11月 18日
				WO 2013138536 A1	2013年 9月 19日
				EP 2825672 A1	2015年 1月 21日

CN	108456713	A	2018年 8月 28日	无	

US	2010221785	A1	2010年 9月 2日	EP 1883707 A4	2009年 8月 5日
				US 8431347 B2	2013年 4月 30日
				CA 2609218 A1	2006年 11月 30日
				CN 101203618 A	2008年 6月 18日
				WO 2006125267 A1	2006年 11月 30日
				CN 101203618 B	2013年 3月 13日
				EP 1883707 B1	2014年 4月 9日
				CA 2609218 C	2016年 10月 11日
				AU 2006251866 A1	2006年 11月 30日
				JP 2008541705 A	2008年 11月 27日
				EP 1883707 A1	2008年 2月 6日
				AU 2006251866 B2	2007年 11月 29日

CN	108004301	A	2018年 5月 8日	WO 2019114146 A1	2019年 6月 20日

CN	105525357	A	2016年 4月 27日	HK 1223660 A1	2017年 8月 4日
				CN 105525357 B	2018年 8月 21日

CN	105442054	A	2016年 3月 30日	CN 105442054 B	2018年 4月 3日

CN	104313699	A	2015年 1月 28日	无	
