(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 11 月 26 日 (26.11.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/232635 A1

(51) 国际专利分类号:
C12Q 1/6806 (2018.01) *C40B 50/06* (2006.01)

(21) 国际申请号: PCT/CN2019/087824

(22) 国际申请日: 2019 年 5 月 21 日 (21.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 深圳华大智造科技有限公司 (**MGI TECH CO., LTD.**) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。

(72) 发明人: 杨林 (**YANG, Lin**); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 张艳艳 (**ZHANG, Yanyan**); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 王其伟 (**WANG, Qiwei**); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 卢佳 (**LU, Jia**); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 陈芳 (**CHEN, Fang**); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。 蒋慧 (**JIANG, Hui**); 中国广东省深圳市盐田区北山工业区综合楼及11栋2楼, Guangdong 518083 (CN)。

(74) 代理人: 北京清亦华知识产权代理事务所 (普通合伙) (**TSINGYIHUA INTELLECTUAL PROPERTY LLC**); 中国北京市海淀区清华园清华大学照澜院商业楼301室, Beijing 100084 (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, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,

(54) **Title:** METHOD AND SYSTEM FOR CONSTRUCTING SEQUENCING LIBRARY ON THE BASIS OF METHYLATED DNA TARGET REGION, AND USE THEREOF

(54) 发明名称: 基于甲基化DNA目标区域构建测序文库及系统和应用

(57) **Abstract:** Disclosed are a method and a system for constructing a sequencing library on the basis of a methylated DNA target region, and a use thereof. The method comprises: obtaining a DNA sample that has been subjected to bisulfite conversion and that comprises a universal sequence; conducting amplification with first specific primers and first universal primers, the first specific primers being located upstream of the target region, and the first universal primers at least partly matching or overlapping the universal sequence; and conducting amplification with second specific primers, second universal primers and labeled primers to obtain the sequencing library; wherein the second specific primers are located downstream of the first specific primers and upstream of the target region, the second universal primers overlap at least part of the sequence of the second specific primers, and the labeled primers overlap part of the sequence of the first universal primer; or, the second specific primers are located downstream of the target region, the second universal primers overlap at least part of the sequence of the first specific primers, and the labelled primers overlap part of the sequence of the second specific primers.

(57) 摘要: 提供了基于甲基化DNA目标区域构建测序文库的方法及系统和应用。该方法包括: 获得经重亚硫酸氢盐转化的带有通用序列的DNA样本; 利用第一特异性引物和第一通用引物扩增获得, 第一特异性引物位于目标区域的上游, 第一通用引物与通用序列至少部分匹配或重叠; 利用第二特异性引物、第二通用引物和标签引物扩增得到测序文库; 其中第二特异性引物位于第一特异性引物的下游和目标区域的上游, 第二通用引物与第二特异性引物的至少部分序列重叠, 标签引物与第一通用引物的部分序列重叠; 或者第二特异性引物位于目标区域的下游, 第二通用引物和第一特异性引物的至少部分序列重叠, 标签引物与第二特异性引物的部分序列重叠。

WO 2020/232635 A1

[见续页]

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)。

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))。

基于甲基化 DNA 目标区域构建测序文库及系统和应用

优先权信息

无。

技术领域

本发明涉及基因测序领域，具体涉及一种基于甲基化 DNA 目标区域构建测序文库的方法及系统和应用。

背景技术

DNA 甲基化是一种表观调控修饰，它在不改变碱基序列的情况下，参与调控蛋白质合成的多少。对人类来说，DNA 甲基化是一种非常奇妙的化学修饰，亲人的关怀，机体的衰老、抽烟、酗酒甚至肥胖，都会被甲基化如实地记录到基因组上。基因组就像是一个日记本，甲基化作为文字，记录下人体的经历。DNA 甲基化是重要的表观遗传学标记信息，获得全基因组范围内所有 C 位点的甲基化水平数据，对于表观遗传学的时空特异性研究具有重要意义。以新一代高通量测序平台为基础，进行全基因组 DNA 甲基化水平图谱绘制，特定物种的高精确度甲基化修饰模式的分析，必将在表观基因组学研究中具有里程碑式的意义，并为细胞分化、组织发育等基础机制研究，以及动植物育种、人类健康与疾病研究奠定基础。

全基因组甲基化测序 WGBS（Whole Genome Bisulfite Sequencing），即全基因组亚硫酸氢盐测序，是研究生物甲基化的最常用手段，它可以覆盖所有甲基化位点，能够获得更加全面的甲基化图谱。但其在高通量测序中遇到了很多挑战：1、亚硫酸氢盐处理会对 DNA 单链化并造成严重的损伤；2、亚硫酸氢盐处理后的未甲基化 C 碱基会转变成 U 碱基，整个基因组的 GC 含量发生极端变化，造成后续扩增产生极大的偏好性；3、建库需要微克级别的起始 DNA，对于微量 DNA 很难有很有效的建库方法。对于临床检测和某些特定的研究来讲，全基因组甲基化测序操作复杂并且成本还过于昂贵，而采用靶向甲基化测序技术可以有效解决这些问题。

靶向甲基化测序技术可以分为以探针捕获和以多重 PCR 为基础的测序技术，对于探针捕获，其要求的起始量高，对于一些微量样本如血浆游离 DNA，很难进行捕获，并且探针捕获探针的设计和操作流程也过于复杂，检测周期长，成本高；而基于 DNA 重亚硫酸盐处理后的多重 PCR 起始要求量低，操作简单，灵敏度高，但该技术还需要进一步改进。

发明内容

本发明旨在至少在一定程度上解决相关技术中的技术问题之一。为此，本发明的一个目的在于提出一种基于甲基化 DNA 目标区域构建测序文库的方法及系统和应用。通过本发明提供的方法对甲基化 DNA 样本的目标区域进行建库，在建库过程中只针对甲基化 DNA 样

本的一条链进行扩增，建库。通过设计特异性引物和通用引物进行扩增，得到目标产物，可以有效解决引物二聚体的问题。同时利用多种特异性引物对同一甲基化 DNA 模板的目标区域进行扩增，可以保证扩增的特异性。

本发明的发明人在研究过程中注意到，基于 DNA 重亚硫酸盐处理后的多重 PCR，操作简单，灵敏度高，但技术要求高。先前有过报道，使用微滴技术进行单分子 BS-PCR，可以同时检测九千个左右的靶标，但是起始量较高，需要 2 μg DNA。2015 年，Lu Wen 等研究人员巧妙地利用 CpG 岛的特征序列作为引物结合位点，开发了基于 PCR 技术的 MCTA-seq，可以同时检测大量 CpG 岛区域的甲基化信号，该技术极其灵敏，能够对 7.5 pg 的 gDNA 进行检测，不过，MCTA-seq 更像是一种固定的 CGI Panel，作为靶向测序平台，灵活性稍显不足。因此开发一个起始量要求低，灵活性强的靶向甲基化技术是未来靶向甲基化的发展方向。

发明人通过研究发现：如何有效的进行超多重靶标扩增是主要瓶颈，即使是开展上万重的基因组扩增子测序也是非常具有挑战性的工作，更别说针对 Bisulfite 转化后序列的多重甲基化 PCR，主要是由于在 PCR 过程中形成了严重的引物二聚体。在重亚硫酸盐处理后的 DNA 进行多重 PCR 的过程中，DNA 经过重亚硫酸盐处理后，未甲基化的胞嘧啶转换为尿嘧啶，基因组上大部分胞嘧啶都是未甲基化的，因此大部分序列的碱基由以前 A/T/C/G 四种组成变为 A/T/G 组成。在常规的 PCR 中，一条引物是针对正链设计，一条是针对互补的链设计，因此用于 PCR 的一条链是富含 ATG 的序列，另一条链是富含 ATC 的序列，这种“天然互补”的引物序列很容易形成引物二聚体。当引物对数增加时，引物二聚体的形成也急剧增加，在多重 PCR 过程中，过多的引物由于引物二聚体的产生而消耗殆尽，造成多重 PCR 的失败，因此要解决多重亚硫酸盐多重 PCR 问题就先得解决引物容易形成引物二聚体的问题。

针对引物二聚体的问题，我们创造性的发明了单方向的引物扩增方法，只针对 DNA 模板两条链中其中一条链设计特异性引物，所有的特异性引物都只含有 ATG 或者 ATC，这些引物互相之间很难形成引物二聚体。通过这些单方向的特异性引物和一些固定的通用引物进行扩增，得到目标产物，可以有效地解决引物二聚体的问题。

具体而言，本发明提供了如下技术方案：

根据本发明的第一方面，本发明提供了一种基于甲基化 DNA 目标区域构建测序文库的方法，包括：（1）基于所述甲基化 DNA 样本，在所述甲基化 DNA 样本的至少一端连接通用序列，用重亚硫酸氢盐处理 DNA 样本，以便获得经转化的带有通用序列的 DNA 样本；（2）利用第一特异性引物和第一通用引物对所述经转化的带有通用序列的 DNA 样本进行第一扩增，以便获得第一扩增产物；其中，所述第一特异性引物位于所述目标区域的上游，所述第一通用引物与所述通用序列至少部分匹配或重叠；所述通用序列位于目标区域下游；（3）利用第二特异性引物、第二通用引物和标签引物对所述第一扩增产物进行第二扩增，以便获得第二扩增产物，得到测序文库；其中所述第二特异性引物位于所述第一特异性引物的下游和所述目标区域的上游，所述第二通用引物与所述第二特异性引物的至少部分序

列重叠，所述标签引物中含有标签序列，所述标签引物与所述第一通用引物的部分序列重叠；或者其中所述第二特异性引物位于所述目标区域的下游，所述第二通用引物和所述第一特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第二特异性引物的部分序列重叠。

本发明提供的基于甲基化 DNA 目标区域构建测序文库的方法，其是针对甲基化 DNA 模板的一条链设计特异性引物来得到对目标区域富集，建库的目的。首先，在甲基化 DNA 模板的至少一端引入通用序列，进行重亚硫酸盐处理，或者先进行重亚硫酸盐处理，然后再引入通用序列也可。即首先获得经转化的带有通用序列的 DNA 样本。然后只针对该 DNA 样本的一条链设计引物。即通过第一特异性引物和第一通用引物来针对该 DNA 样本的一条链进行扩增，第一特异性引物能够和该 DNA 样本的一条链匹配，第一通用引物能够和通用序列匹配，从而实现特异扩增。而且由于所用到的 DNA 模板是经过重亚硫酸盐转化后的样本，所以所设计的第一特异性引物是富含碱基 A、T、G 或者碱基 A、T、C 的序列，相互之间不会形成二聚体。而第一通用引物上含有 A、T、C、G 四种碱基，也不会和第一特异性引物形成引物二聚体，因此可以完全避免引物二聚体的形成。

同时，为了保证引物扩增的特异性，在第一特异性引物的下游和目标区域的上游或者目标区域的下游再设计一条第二特异性引物，利用第二特异性引物、第二通用引物和标签引物，针对第一扩增产物进行第二扩增，获得第二扩增产物，得到所需要的测序文库。

根据本发明的实施例，以上所述的基于甲基化 DNA 目标区域构建测序文库的方法可以进一步包括如下技术特征：

在本发明的一些实施例中，步骤（3）中，所述第二特异性引物的 5’端与所述第二通用引物的 3’端的至少部分序列重叠，所述标签引物的 3’端与所述第一通用引物的 5’端的部分序列重叠。第二特异性引物的 5’端序列能够与第二通用引物的 3’端的至少部分序列重叠，3’端序列能够和 DNA 模板上位于第一特异性引物下游和目标区域上游的模板区域进行匹配，从而可以基于第一扩增产物，实现对于目标区域的特异性扩增。

在本发明的一些实施例中，步骤（3）中，所述第二特异性引物的 5’端与所述标签引物 3’端的至少部分序列重叠，所述第二通用引物的 3’端与所述第一特异性引物的 5’端的部分序列重叠。第二特异性引物的 5’端序列与标签引物 3’端的至少部分序列重叠，其 3’端序列能够和 DNA 模板上位于目标区域下游的模板区域进行匹配，从而可以实现对于目标区域的特异性扩增。

在本发明的一些实施例中，标签引物中含有标签序列，这些标签序列可以是一些测序平台常用的用于区分不同样本的标签序列，方便同时用于多个混合样本测序，根据实施例，这些标签序列的长度可以为8~12bp，例如可以为10bp，8bp等。

在本发明的一些实施例中，步骤（1）进一步包括：（1-a）利用重亚硫酸氢盐对所述甲基化 DNA 样本进行处理，以便获得经转化的 DNA 样本；（1-b）利用 DNA 聚合酶和带有第一测序序列的随机引物，对所述经转化的 DNA 样本进行复制，以便获得所述经转化的带有通用序列的 DNA 样本，所述随机引物的 3’端为随机碱基序列，所述随机引物 5’端为通

用序列。

在本发明的一些实施例中，所述随机碱基序列为 6~12 个，所述随机碱基为 A、T、C 或者 G。

在本发明的一些实施例中，所述随机碱基序列为 6~12 个，所述随机碱基为 A、T 或者 C。

在本发明的一些实施例中，所述通用序列为测序接头序列或固定序列。

在本发明的一些实施例中，所述测序接头序列或者所述固定序列中胞嘧啶为甲基化修饰的胞嘧啶。

在本发明的一些实施例中，步骤（1）进一步包括：（1-1）对所述甲基化 DNA 样本进行末端修复加 A，以便获得修复的 DNA 样本；（1-2）将所述修复的 DNA 样本的至少一端与通用序列连接，以便获得带有通用序列的 DNA 样本；（1-3）利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

在本发明的一些实施例中，所述通用序列选自下列中的至少一种：测序接头序列或者经修饰的测序接头序列。

在本发明的一些实施例中，所述经修饰的测序接头序列为一条链胞嘧啶进行甲基化修饰，一条链胞嘧啶不进行甲基化修饰、一条链 3’端碱基经非羟基修饰的测序接头序列、带有固定序列和随机序列的测序接头序列、或者一条链 3’端碱基经非羟基修饰的带有固定序列和随机序列的测序接头序列。

在本发明的一些实施例中，所述随机序列为分子标签序列。通过大量不同的分子标签序列可以对原始的 DNA 模板个数进行计数，通过后续对分子标记序列的统计来追溯原始模板的个数和矫正测序或者 PCR 过程中产生的错误，从而可以实现对于 DNA 模板的精确检测和定量研究。

在本发明的一些实施例中，步骤（1）进一步包括：①利用转座酶对所述 DNA 样本进行打断和转座处理，以便获得带有通用序列的 DNA 样本，所述转座酶中包埋有通用序列；②利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

在本发明的一些实施例中，所述通用序列为转座酶效应序列或者带有测序接头的转座酶效应序列，优选为转座酶效应序列，所述转座酶可以为 Tn5、MuA 或其他具有类似功能的转座酶，优选为 Tn5 转座酶。

在本发明的一些实施例中，转座酶效应序列中的胞嘧啶为甲基化修饰的胞嘧啶。未甲基化胞嘧啶转化为鸟嘧啶不是一个 100%的过程，可能转化也可能不转化，那么在后续用通用引物扩增的会增加不确定性。而甲基化修饰的胞嘧啶在后续的亚硫酸盐处理的条件下胞嘧啶不会转化为尿嘧啶，保持序列信息不变。因此为了测序更加精准，转座酶效应序列中的胞嘧啶可以进行甲基化修饰。当然，胞嘧啶也可以不进行甲基化修饰处理。

在本发明的一些实施例中，所述甲基化 DNA 样本为基因组 DNA、片段化的基因组 DNA、或者游离 DNA。

根据本发明的第二方面，本发明提供了一种基于甲基化 DNA 目标区域构建测序文库的系统，包括：通用转化模块，所述通用转化模块基于所述甲基化 DNA 样本，构建在所述甲基化 DNA 样本的至少一端连接有通用序列，且经重亚硫酸氢盐处理的 DNA 样本，以便获得经转化的带有通用序列的 DNA 样本；第一扩增模块，所述第一扩增模块和所述通用转化模块相连，所述第一扩增模块利用第一特异性引物和第一通用引物对所述经转化的带有通用序列的 DNA 样本进行第一扩增，以便获得第一扩增产物；其中，所述第一特异性引物位于所述目标区域的上游，所述第一通用引物与所述通用序列至少部分匹配或重叠；第二扩增模块，所述第二扩增模块和所述第一扩增模块相连，所述第二扩增模块利用第二特异性引物、第二通用引物和标签引物对所述第一扩增产物进行第二扩增，获得第二扩增产物，得到测序文库；其中所述第二特异性引物位于所述第一特异性引物的下游和所述目标区域的上游，所述第二通用引物与所述第二特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第一通用引物的部分序列重叠；或者其中所述第二特异性引物位于所述目标区域的下游，所述第二通用引物和所述第一特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第二特异性引物的部分序列重叠。

根据本发明的实施例，以上所述基于甲基化 DNA 目标区域构建测序文库的系统可以进一步包括如下技术特征：

在本发明的一些实施例中，以上系统中，所述第二扩增模块中所述第二特异性引物的 5’端与所述第二通用引物的 3’端的至少部分序列重叠，所述标签引物的 3’端与所述第一通用引物的 5’端的部分序列重叠。

在本发明的一些实施例中，以上系统中，所述第二扩增模块中所述第二特异性引物的 5’端与所述标签引物 3’端的至少部分序列重叠，所述第二通用引物的 3’端与所述第一特异性引物的 5’端的部分序列重叠。

在本发明的一些实施例中，以上系统中，所述标签序列的长度为 8~12bp。

在本发明的一些实施例中，所述通用转化模块进一步包括：转化单元，所述转化单元利用重亚硫酸氢盐对所述甲基化 DNA 样本进行处理，以便获得经转化的 DNA 样本；扩增单元，所述扩增单元与所述转化单元相连，所述扩增单元利用 DNA 聚合酶和第一测序引物，对所述经转化的 DNA 样本进行复制，以便获得所述经转化的带有通用序列的 DNA 样本，所述第一测序引物的 3’端为随机碱基，所述第一测序引物的 5’端为通用序列。

在本发明的一些实施例中，以上系统中，所述随机碱基为 6~12 个，所述随机碱基为 A、T、C 或者 G。

在本发明的一些实施例中，以上系统中所述随机碱基为 6~12 个，所述随机碱基为 A、T 或者 C。

在本发明的一些实施例中，以上系统中，所述通用序列为测序接头序列或固定序列。

在本发明的一些实施例中，以上系统中，所述测序接头序列或者所述固定序列中胞嘧啶为甲基化修饰的胞嘧啶。

在本发明的一些实施例中，所述通用转化模块进一步包括：修复单元，所述修复单元用于对所述甲基化 DNA 样本进行末端修复加 A，以便获得修复的 DNA 样本；连接单元，所述连接单元与所述修复单元相连，所述连接单元用于将所述修复的 DNA 样本的至少一端与通用序列连接，以便获得带有通用序列的 DNA 样本；转化单元，所述转化单元与所述连接单元相连，所述转化单元利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

在本发明的一些实施例中，所述通用转化模块中，所述通用序列选自下列中的至少一种：测序接头序列或者经修饰的测序接头序列。

在本发明的一些实施例中，所述通用转化模块中，所述经修饰的测序接头序列为一条链胞嘧啶进行甲基化修饰，一条链胞嘧啶不进行甲基化修饰、一条链 3’端碱基经非羟基修饰的测序接头序列、带有固定序列和随机序列的测序接头序列、或者一条链 3’端碱基经非羟基修饰的带有固定序列和随机序列的测序接头序列。

在本发明的一些实施例中，所述通用转化模块中，所述随机序列为分子标签序列。通过大量不同的分子标记序列可以对原始的 DNA 模板个数进行计数，通过后续对分子标记序列的统计来追溯原始模板的个数和矫正测序或者 PCR 过程中产生的错误，从而可以实现对于 DNA 模板的精确检测和定量研究。

在本发明的一些实施例中，所述通用转化模块进一步包括：转座单元，所述转座单元利用转座酶对所述 DNA 样本进行转座处理，以便获得带有通用序列的 DNA 样本，所述转座酶中包埋有通用序列；转化单元，所述转化单元与所述转座单元相连，所述转化单元利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

在本发明的一些实施例中，以上转座单元中，所述通用序列为转座酶效应序列或者带有测序接头的转座酶效应序列，优选为转座酶效应序列。

在本发明的一些实施例中，以上转座单元中，所述转座酶效应序列中的胞嘧啶为甲基化修饰的胞嘧啶。

在本发明的一些实施例中，所述甲基化 DNA 样本为基因组 DNA、片段化的基因组 DNA 或者游离 DNA。

上述对本发明任一实施例中的基于甲基化 DNA 目标区域构建测序文库的方法的优点和技术特征的描述，同样适用本发明中上述任一实施例中基于甲基化 DNA 目标区域构建测序文库的系统，在此不再赘述。

根据本发明的第三方面，本发明提供了一种对甲基化 DNA 样本进行测序的方法，包括：

基于所述甲基化 DNA 样本，根据本发明第一方面任一实施例所述的方法或者利用本发明第二方面任一实施例所述的系统构建得到测序文库；对所述测序文库进行高通量测序，以便获得测序结果。

在本发明的一些实施例中，利用测序平台对所述测序文库进行高通量测序，所述测序平台选自 MGISEQ、Illumina、Proton 中的至少一种。

根据本发明的第四方面，本发明提供了一种确定甲基化DNA样本的甲基化状态的方法，包括：

基于所述甲基化DNA样本，根据本发明第一方面任一实施例所述的方法或者利用本发明第二方面任一实施例所述的系统构建得到测序文库；对所述测序文库进行高通量测序，以便获得测序结果；将所述测序结果与参考基因组进行比对，以便确定所述甲基化DNA样本的甲基化状态。

在本发明的一些实施例中，所述参考基因组为人类基因组hg19或炎黄基因组。

根据本发明的第五方面，本发明提供了一种试剂盒，包括：通用序列，标签引物，第一通用引物，第二通用引物和甲基化常规检测试剂；其中所述标签引物中含有标签序列，所述第一通用引物与所述通用序列的至少部分匹配或重叠，所述第一通用引物为SEQ ID NO:1，所述第二通用引物为SEQ ID NO:22。所述甲基化常规检测试剂例如可以为重亚硫酸盐检测试剂或者相应的试剂盒等。

根据本发明的实施例，以上所述的试剂盒进一步包括如下附加技术特征：

在本发明的一些实施例中，所述标签引物为SEQ ID NO:23所示。

在本发明的一些实施例中，所述试剂盒进一步包括：第一特异性引物和第二特异性引物，所述第一特异性引物包括SEQ ID NO:1~SEQ ID NO:10所示序列，所述第二特异性引物包括SEQ ID NO:11~SEQ ID NO:20所示序列。

在本发明的一些实施例中，所述试剂盒利用本发明第一方面所述的方法基于甲基化DNA目标区域构建测序文库。

附图说明

本发明的上述和/或附加的方面和优点从结合下面附图对实施例的描述中将变得明显和容易理解，其中：

图1是根据本发明的一个实施例提供的随机引物建库流程图。

图2是根据本发明的一个实施例提供的接头连接建库流程图。

图3是根据本发明的一个实施例提供的转座子建库流程图。

图4是根据本发明的一个实施例提供的不同接头序列的示意图。

图5是根据本发明的一个实施例提供的测序文库质检图。

图6是根据本发明的一个实施例提供的各个扩增子的测序深度结果图。

图7是根据本发明的一个实施例提供的测序文库质检图。

图8是根据本发明的一个实施例提供的各个扩增子的测序深度结果图。

图9是根据本发明的实施例提供的基于甲基化DNA目标区域构建测序文库的系统的结构示意图。

图10是根据本发明的实施例提供的一种通用转化模块的结构示意图。

图11是根据本发明的实施例提供的一种通用转化模块的结构示意图。

图12是根据本发明的实施例提供的一种通用转化模块的结构示意图。

具体实施方式

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本发明，而不能理解为对本发明的限制。

为了对于本申请有更为直观的理解，下面对本申请中存在的术语进行解释和说明。本领域技术人员需要理解的是，这些解释和说明仅为了理解更为方便，不应看做是对本申请保护范围的限制。本文中，如无特殊说明，当提到两个核酸序列相连时，是指通过 3'-5'磷酸二酯键相连。本文中如无特别说明，当表示碱基时，碱基 N 或者 n 代表任意碱基 A、T、C 或者 G。

本文中，术语“上游”、“下游”指的是按照核苷酸5'-3'的排列顺序，两个或者多个核酸序列相比，位于上游的核酸序列相比较位于下游的核酸序列，其识别或者匹配区域更靠近模板序列的5'端。当然，由于不同核酸序列的长度可能不一样，其识别或者匹配的区域的长度可能也会不一样。当表示A核酸序列位于B核酸序列的下游时，只需要A核酸序列的3'端识别或者结合位点相比较于B核酸序列的3'端的识别或者结合位点，更靠近模板序列的3'端即可。

本文中，当表示两个核酸序列之间“匹配”时，是指两个核酸序列的碱基之间发生互补配对。当表示两个核酸序列至少部分序列重叠时，是指两个核酸序列至少有一段相同的核酸序列。

本文中，无论是“重亚硫酸氢盐”、“亚硫酸盐”或者“亚硫酸氢盐”处理均指使DNA中的胞嘧啶脱氨转变为尿嘧啶的试剂或者过程。所以无论是基于重亚硫酸氢盐处理、亚硫酸盐处理、或者亚硫酸氢盐处理，均包含在本发明的保护范围之内。

为了解决在对甲基化 DNA 进行扩增的过程中，多对甲基化特异性引物之间的引物二聚体问题，本发明创造性的发明了单方向的引物扩增方法， 即只针对 DNA 模板的一条链进行引物设计，由此所设计的特异性引物都只含有 A、T、G 或者 A、T、C，相互之间很难形成引物二聚体。同时，为了保证引物扩增的特异性，在第二轮 PCR 扩增的过程中，针对第一轮扩增的产物，在其上设计特异性引物进行扩增，进一步保证扩增的特异性。由此所制备的测序文库满足测序的要求。

详细地说，基因组DNA（gDNA）通过Tn5转座子转座，打断的gDNA或者游离DNA（cfDNA）分子上通过接头连接或者是DNA随机复制在原始DNA上引入一段通用序列，对引入通用序列的DNA进行重亚硫酸盐处理（BS处理），得到重亚硫酸盐转化后的DNA序列（原始DNA未甲基化修饰的胞嘧啶(C)被转化为尿嘧啶（U））。根据所引入的通用序列上设计通用引物，在转化后DNA序列的目标区域上游设计特异性引物，特异性引物只针对DNA模板上的一条链进行设计，通过通用引物和特异性引物进行PCR扩增，得到PCR产物。同时为增加扩增的特异性，在上述特异性引物的下游设计巢式引物或者在目标区域的下游设计特异性引物，巢式引物或者特异性引物都只针对DNA模板上的一条链进行设计，通过巢式

引物或者下游特异性引物和通用引物对第一步PCR的产物进行第二步扩增，最终得到针对重亚硫酸盐处理后的模板的PCR扩增产物（BS-PCR）。

在本发明的一个方面，本发明提供了一种基于甲基化 DNA 目标区域构建测序文库的方法，包括：（1）基于所述甲基化 DNA 样本，构建在所述甲基化 DNA 样本的至少一端连接有通用序列，且经重亚硫酸氢盐处理的 DNA 样本，以便获得经转化的带有通用序列的 DNA 样本；（2）利用第一特异性引物和第一通用引物对所述经转化的带有通用序列的 DNA 样本进行第一扩增，以便获得第一扩增产物；其中，所述第一特异性引物位于所述目标区域的上游，所述第一通用引物与所述通用序列至少部分重叠或匹配；所述通用序列位于目标区域的下游；（3）利用第二特异性引物、第二通用引物和标签引物对所述第一扩增产物进行第二扩增，获得第二扩增产物，得到测序文库；其中所述第二特异性引物位于所述第一特异性引物的下游和所述目标区域的上游，所述第二通用引物与所述第二特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第一通用引物的部分序列重叠；或者其中所述第二特异性引物位于所述目标区域的下游，所述第二通用引物和所述第一特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第二特异性引物的部分序列重叠。

其中，在获得经转化的带有通用序列的DNA样本的过程中，根据通用序列和重亚硫酸盐处理的先后，根据需要，可以采取不同的方式：

在本发明的至少一些实施方式中，通过如下方法引入通用序列：

1. gDNA、打断的gDNA或cfDNA，先用重亚硫酸盐对DNA分子进行处理，然后用第一测序引物，即3’端带有6-12个随机N碱基（A/T/C/G组成的兼并碱基）或者6-12个随机H碱基（A/T/C组成的兼并碱基），5’端带有部分、全部测序接头序列或者固定序列（其中，序列中的胞嘧啶优先采用甲基化修饰的胞嘧啶）的引物和DNA聚合酶对模板进行复制，得到5’端带有通用序列的重亚硫酸盐处理后的DNA模板（如图1所示）。其中，可用的测序接头序列包括但不限于MGI平台的测序接头也包括illumina和proton平台的测序接头序列。在至少一些实施例中，可用的DNA聚合酶可以是常规的rTaq、Fusion、也可以是Bst或者phi29等。

在本发明的至少一些实施方式中，通过如下方法引入通用序列：

打断的gDNA或cfDNA通过末端修复加A，然后加上特定的接头序列，序列可以是部分、全部测序接头序列或者经修饰的测序接头序列，这些经修饰的测序接头序列可以是一条链3端碱基经非羟基修饰的带有固定序列的测序接头序列，或者是带有固定序列的测序接头序列，或者是一条链3’端碱基经非羟基修饰带有固定序列的测序接头序列，如图4所示标号1、标号2、标号3和标号4所示。其中纯化后，采用亚硫酸盐对加上通用序列的产物进行处理得到转化后的DNA模板（图2）。

在本发明的另一些实施方式中，通过如下方法引入通用序列：

Tn5转座酶包埋一段接头序列，该接头可以是Tn5转座酶本身的起效应的19bp特定序列，也可以使起效应的序列+其他序列（如测序接头序列）的组合，优先采用19bp特定序列，19bp特定序列中的胞嘧啶优先采用甲基化修饰的胞嘧啶，通过Tn5转座对gDNA进行转座加上特

定的接头，纯化后，采用重亚硫酸盐对加上特定接头的产物进行处理得到转化后的DNA模板（如图3所示）。

在获得上述经转化的带有通用序列的DNA样本之后，通过单方向特异性引物进行PCR扩增得到测序文库，扩增方式可以采用以下的任意一种：

在本发明的至少一些实施方式中，通过如下方法进行PCR扩增得到测序文库：

通过第一特异性引物和第一通用引物对亚硫酸盐处理后的DNA进行第一步PCR扩增。第一通用引物3’端序列和上述引入的通用序列部分或者全部互补或者重叠。例如，第一通用序列的5’端是测序接头序列的部分或者全部（优先部分序列）。第一特异性引物序列的结合位点位于需要扩增的目标区域的上游，其设计是针对重亚硫酸盐处理后的DNA模板序列；得到的产物经过纯化后，再通过第二特异性引物（在以下实施例中也被称为巢式引物）、第二通用引物、标签引物进行第二部PCR扩增。其中，在第二步PCR的第一个循环，第二特异性引物和标签引物先进行PCR，后续循环通过第二特异性引物、第二通用引物和标签引物一起进行多轮PCR。第二特异性引物5’端与第二通用引物的3’端的部分或者全部序列重叠，第二特异性引物的3’端为特异性序列，特异性序列设计在第一特异性引物和目标区域之间；第二通用引物可以为测序通用接头的部分或者全部序列，3’端和第二特异性引物的5’端部分或者全部序列相同；标签引物3’端和第一通用引物的5’端部分或者全部序列相同，中间有8-12bp的固定标签序列（每个平台用于区分样本混样的标签序列），用于后续多样本混合测序（附图1A，附图2A，附图3A）。

在本发明的另一些实施方式中，通过如下方法进行PCR扩增获得测序文库。

通过第一特异性引物（在下述实施例中也被称为上游特异性引物）和第一通用引物对亚硫酸盐处理后的DNA进行第一步PCR扩增。第一通用引物3’端序列和上述引入的通用序列部分或者全部互补或者重叠（此处的通用序列优先采用除测序接头序列之外的固定序列），第一特异性引物3’端的特异性序列设计需要扩增的目标区域的上游，其设计是针对重亚硫酸盐处理后的DNA模板序列，5’端为测序接头序列的部分或者全部序列（优先部分序列）。得到的产物经过纯化后，再通过第二特异性引物（相应地，在以下实施例中也可以称为下游特异性引物）、第二通用引物、标签引物进行第二步PCR扩增。其中，在第二步PCR的第一个循环，第二特异性引物和第二通用引物先进行PCR扩增，后续循环第二特异性引物、第二通用引物和标签引物一起进行多轮PCR；下游特异性引物5’端与标签引物的3’端的部分或者全部序列重叠，第二特异性引物的3’端为特异性序列，特异性序列设计在目标区域的下游；第二通用引物可以为测序接头序列的部分或者全部序列，其3’端和第一特异性引物的5’端部分或者全部序列重叠；标签引物3’端和第二特异性引物的5’端部分或者全部序列相同，中间有8-12bp的固定标签序列（每个平台用于区分样本混样的标签序列），用于后续多样本混合测序（附图1B，附图2B，附图3B）。

*根据*本发明的另一方面，本发明提供了一种基于甲基化 DNA 目标区域构建测序文库的系统，如图 9 所示，包括通用转化模块、第一扩增模块和第二扩增模块，各模块依次相连。其中，所述通用转化模块基于所述甲基化 DNA 样本，构建在所述甲基化 DNA 样本的至少

一端连接有通用序列，且经重亚硫酸氢盐处理的DNA样本，以便获得经转化的带有通用序列的DNA样本。所述第一扩增模块利用第一特异性引物和第一通用引物对所述经转化的带有通用序列的DNA样本进行第一扩增，以便获得第一扩增产物，其中所述第一特异性引物位于所述目标区域的上游，所述第一通用引物与所述通用序列至少部分匹配或重叠。所述第二扩增模块利用第二特异性引物、第二通用引物和标签引物对所述第一扩增产物进行第二扩增，获得第二扩增产物，得到测序文库；其中所述第二特异性引物，所述通用引物和所述标签引物如（i）或（ii）所示：（i）所述第二特异性引物位于所述第一特异性引物的下游和所述目标区域的上游，所述第二通用引物与所述第二特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第一通用引物的部分序列重叠；（ii）所述第二特异性引物位于所述目标区域的下游，所述第二通用引物和所述第一特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第二特异性引物的部分序列重叠。

在本发明的至少一些实施方式中，所述通用转化模块如图10所示，包括转化单元和与转化单元相连的扩增单元。其中，所述转化单元利用重亚硫酸氢盐对甲基化DNA样本进行处理，以便获得经转化的DNA样本。所述扩增单元利用DNA聚合酶和第一测序引物，对所述经转化的DNA样本进行复制，以便获得所述经转化的带有通用序列的DNA样本，所述第一测序引物的3’端为随机碱基，所述第一测序引物的5’端为通用序列。

在本发明的至少一些实施方式中，所述通用转化模块如图11所示，包括修复单元、连接单元和转化单元，各单元依次相连。所述修复单元用于对所述甲基化DNA样本进行末端修复加A，以便获得修复的DNA样本。所述连接单元用于将所述修复的DNA样本的至少一端与通用序列连接，以便获得带有通用序列的DNA样本。所述转化单元利用重亚硫酸氢盐对所述带有通用序列的DNA样本进行处理，以便获得所述经转化的带有通用序列的DNA样本。

在本发明的至少一些实施方式中，所述通用转化模块如图12所示，包括转座单元和与转座单元相连的转化单元。所述转座单元利用转座酶对所述DNA样本进行打断和转座处理，以便获得带有通用序列的DNA样本，所述转座酶中包埋有通用序列。所述转化单元利用重亚硫酸氢盐对所述带有通用序列的DNA样本进行处理，以便获得所述经转化的带有通用序列的DNA样本。

下面将结合实施例对本发明的方案进行解释。本领域技术人员将会理解，下面的实施例仅用于说明本发明，而不应视为限定本发明的范围。实施例中未注明具体技术或条件的，按照本领域内的文献所描述的技术或条件或者按照产品说明书进行。所用试剂或仪器未注明生产厂商者，均为可以通过市购获得的常规产品。

实施例1：甲基化多重PCR建库测序

实验设计：用100ng炎黄基因组DNA进行重亚硫酸盐处理，然后按照发明的步骤对DNA靶向甲基化文库制备，文库到MGISEQ-2000测序仪上进行上机测序，测序类型PE100，然后进行数据分析，包括数据利用率、比对率、扩增子特异性、均一性等性能。

1、重亚硫酸盐处理

采用 EZ DNA Methylation-Gold Kit™(美国 ZYMO 公司 货号 D5005)试剂盒，将上述的 DNA 进行重亚硫酸盐共处理。

配制溶液：

制备 CT 转换试剂（CT Conversion Reagent）溶液：从试剂盒中取出 CT 转换试剂（固体混合物），分别加入 900μL 的水、50μL 的 M-溶解缓冲液（M-Dissolving Buffer）和 300μL 的 M-稀释缓冲液（M-Dilution Buffer），室温下溶解并且震荡 10 分钟或在摇床上摇动 10 分钟。

M-洗涤缓冲液的制备：向 M-洗涤缓冲液中添加 24mL 100%的乙醇，备用。

具体步骤如下：

（1）在 PCR 管中加入 130μL 的 CT 转换试剂溶液和上述 DNA，轻弹或移液器吹悬混合样品。

然后将样品管放到 PCR 仪上按以下步骤操作:

98℃下持续 5 分钟， 64℃下持续 2.5 小时。

完成上述操作后，立刻进行下一步操作。

（2）将 Zymo-Spin IC™ Column 放入收集管（Collection Tube）中，并加入 600μL 的 M-结合缓冲液（M-Binding Buffer）。

然后将重亚硫酸盐处理的样品加入到含 M-结合缓冲液的 Zymo-Spin IC™ Column 中，盖上盖子颠倒混匀。

全速(>10,000 x g)离心 30 秒，弃收集管中的收集液。向柱中加入 100μL 的 M-洗涤缓冲液，全速(>10,000 x g)离心 30 秒，弃收集管中的液体。

向柱中添加 200μL 的 M-Desulphonation Buffer，室温放置 15min，全速(>10,000 x g)离心 30s，弃收集管中的液体。

向柱中添加 200μL 的 M-洗涤缓冲液，全速(>10,000 x g)离心 30s，弃收集管中的液体，并再重复此步骤 1 次。

将 Zymo-Spin IC™ Column 置于新的 1.5mL EP 管中，加入 40μL 的 M-洗脱缓冲液 r 到柱基质中，室温放置 2min，全速(>10,000 x g)离心洗脱目的片段 DNA。

2、DNA复制

（1）在PCR管中按照以下反应体系对重亚硫酸盐处理后的DNA进行DNA复制

上一步连接后的 DNA	38μL
5 × Bst buffer	5μL
随机引物（10μM）	5μL
BST 酶（美国 NEB 公司,货号 M0538）	1 μL

dNTP mix(10mM)	1μL
总体积	50 μL

其中随机引物序列（即本文中所提到的第一测序引物）：CGCTTGGCCTCCGACTTNNNNNNNN（SEQ ID NO:24），其中N为A/T/C/G四种碱基组成的随机序列。

（2）对上述反应体系放置到PCR仪上，65度，反应10分钟，

（3）反应完后用1.5×AMPure磁珠进行纯化（Beckman 公司AMPure XP，货号A63881），最后将纯化产物溶于22μl洗脱缓冲液。

3、第一轮PCR

（1）在PCR管中按照以下反应体系配置PCR体系

上一步处理后的 DNA	20μL
2 ×KAPA2G Fast ReadyMix（美国 kapa 公司，KK5102）	25μL
第一特异性引物池（10μM）	2.5μL
第一通用引物（10μM）	2.5μL
总体积	50 μL

（2）PCR反应条件为

94℃	1min	
94℃	30 s	15 个循环
58℃	2min	
72℃	30 s	
72℃	5min	
12℃	保持	

（3）反应完后用1.5× AMPure磁珠进行纯化，最后将纯化产物溶于22μl洗脱缓冲液。

3、第二轮PCR

（1）在PCR管中按照以下反应体系配置PCR体系，其中，巢氏引物池如下表4所示，标签引物如下表5所示。

上一步处理后的 DNA	17.5μL
2 × KAPA2G Fast ReadyMix	25μL
第二特异性引物池（10μM）	2.5μL
第二通用引物（10μM）	2.5μL
标签引物（10μM）	2.5
总体积	50 μL

（2）PCR反应条件

94℃	1min	
94℃	30 s	20 个循环
58℃	2min	
72℃	30 s	
72℃	5min	
12℃	保持	

（3）反应完后用1.0× AMPure磁珠进行纯化，最后将纯化产物溶于22μl洗脱缓冲液。

4、文库检测：

使用 Bioanalyzer 分析系统(Agilent, Santa Clara, USA)检测文库插入片段的大小及含量，其结果如图 5 所示；

5、上机测序

将得到的文库进行高通量测序，测序平台 MGISEQ-2000，测序类型 PE100，测序后数据经过比对后统计各项基本参数，包括下机数据、可用数据、比对率、GC 含量等，其结果如下表 1 所示。其中各扩增子的深度如图 6 所示。其中，图 6 中横坐标代表不同的 CpG 位点。

表1 测序检测结果

编号	下机数据	接头过滤比例	比对率	特异性	0.1X 均一性
样本 1	136227	1.3%	89.6%	78.6%	90%
样本 2	115298	1.0%	88.5%	77.5%	90%
样本 3	114045	0.9%	88.1%	78.7%	90%

表1中，样本1~样本3分别代表同一个样本做了三次重复，比对率是指比对到基因组上的比例，特异性是指目标区域的reads占总测序总reads的比例，均一性是指目标区域深度大于目标区域平均深度0.1倍个数占总目标区域数的比例。

从表1可以看出，各样本的接头过滤比例在1%上下，结合图5所示的文库质检结果，说明所形成的引物二聚体极少，比对率均在88-89%，特异性77-79%，性能良好。且各个扩增子的深度均一性良好。

实施例2：甲基化多重PCR建库测序

实验设计：采用打断到200-300bp的炎黄基因组DNA，然后按照本发明所提供的方法对DNA靶向甲基化文库制备，文库到MGISEQ-2000测序仪上进行上机测序，测序类型PE100，然后进行数据分析，包括数据利用率、比对率、扩增子特异性、均一性等性能。

1、末端修复

（1）将上一步获得的DNA片段按照下表在1.5mL的离心管中配制末端修复反应体系：

DNA 片段	30 μL
H_2O	45 μL
10×多核苷酸激酶缓冲液（eznymatic 公司,货号 Y9040L）	10 μL
dNTPs（每种成分均为 10mM）（eznymatic 公司,货号 N2010L）	4 μL
T4 DNA 聚合酶（eznymatic 公司,货号 P7080L）	5 μL
Klenow 片段（eznymatic 公司,货号 P7060L）	1 μL
T4 多核苷酸激酶（eznymatic 公司，货号 Y9040L）	5 μL
总体积	100 μL

（2） 将上述反应体系置于 20℃的 Thermomixer (Eppendorf)上，进行反应 30 min。反应完后用 AMPure 磁珠进行纯化，最后将纯化产物溶于 34μl 洗脱缓冲液。上述试剂均采用 enzymatic 公司的试剂。

2、末端添加碱基A：

（1）将上一步得到的 DNA 按下表在 1.5 mL 的离心管中配制添加碱基 A 的反应体系：

DNA	32 μL
10× Klenow 缓冲液 （eznymatic 公司，货号 P7010-HC-L）	5 μL
dATP(稀释为 1mM，eznymatic 公司，货号 N2010-A-L)	10 μL
Klenow (3’-5’ exo-，eznymatic 公司，货号 P7010-HC-L)	3 μL
总体积	50 μL

（2）将上述反应体系置于 37℃的 Thermomixer(Eppendorf)上，进行反应 30 min。反应完后用 AMPure 磁珠进行纯化，最后将纯化产物溶于 20μl 洗脱缓冲液。

2、连接甲基化接头1：

（1）将上一步得到的 DNA 按下表配制甲基化接头（有时也称为“甲基化标签接头”）的连接反应体系：

DNA	18 μL
2×Rapid 连接缓冲液	25 μL
甲基化标签接头*	4 μL
T4 DNA 连接酶 (Rapid，enzymatic，L603-HC-L)	3 μL
总体积	50 μL

*甲基化接头序列为：

接头 1：5’ /5Phos/AGTCGGAGGCCAAGCGGT（SEQ ID NO:25）

接头 2：5’ ACATGGCTACGATCCGACTddT（SEQ ID NO:26）

接头 1 序列中的 C 均进行了甲基化修饰保护，接头 2 中的序列可进行或不进行甲基化修饰保护，接头 2 中的 3 端的最后一个碱基进行阻断修饰防止和模板进行连接，即进行了双脱氧修饰。

（2）将上述反应体系置于 20℃的 Thermomixer (Eppendorf)上，进行反应 15 min，获得连接产物。反应完后用 AMPure 磁珠进行纯化，最后将纯化产物溶于 22μl 洗脱缓冲液。

3、亚硫酸盐处理

采用试剂盒 EZ DNA Methylation-Gold Kit™(ZYMO 公司）将上述连接好的 DNA 进行重亚硫酸盐共处理。

（1）准备试剂：

制备 CT 转换试剂（CT Conversion Reagent）溶液：从试剂盒中取出 CT 转换试剂（固体混合物），分别加入 900μL 的水、50μL 的 M-溶解缓冲液（M-Dissolving Buffer）和 300μL 的 M-稀释缓冲液（M-Dilution Buffer），室温下溶解并且震荡 10 分钟或在摇床上摇动 10 分钟。

M-洗涤缓冲液的制备：向 M-洗涤缓冲液中添加 24mL 100%的乙醇，备用。

（2）在 PCR 管中加入 130μL 的 CT 转换试剂溶液和上述连接好的 DNA，轻弹或移液器吹悬混合样品。

然后将样品管放到 PCR 仪上按以下步骤操作：

98℃下持续 5 分钟，64℃下持续 2.5 小时。

完成上述操作后，立刻进行下一步操作或者在 4℃下存储(最多 20 小时)备用。

（3） 将 Zymo-Spin IC™ Column 放入收集管（Collection Tube）中，并加入 600μL 的 M-结合缓冲液（M-Binding Buffer）。

然后将上述重亚硫酸盐处理的样品加入到含 M-结合缓冲液的 Zymo-Spin IC™ Column 中，盖上盖子颠倒混匀。

全速(>10,000 x g)离心 30 秒，弃收集管中的收集液。

向柱中加入 100μL 的 M-洗涤缓冲液，全速(>10,000 x g)离心 30 秒，弃收集管中的液体。

向柱中添加 200μL 的 M-Desulphonation Buffer，室温放置 15min，全速(>10,000 x g)离心 30s，弃收集管中的液体。

向柱中添加 200μL 的 M-洗涤缓冲液，全速(>10,000 x g)离心 30s，弃收集管中的液体，并再重复此步骤 1 次。

将 Zymo-Spin IC™ Column 置于新的 1.5mL EP 管中，加入 18μL 的 M-洗脱缓冲液 r 到柱基质中，室温放置 2min，全速(>10,000 x g)离心洗脱目的片段 DNA。

4、第一轮PCR

（1）在PCR管中按照以下反应体系配置PCR体系，其中上游特异性引物池包含的引物如下表3所示，第一通用引物如下表5所示。

上一步处理后的 DNA	20μL
2 ×KAPA2G Fast ReadyMix	25μL
第一特异性引物池（10μM）	2.5μL
第一通用引物（10μM）	2.5μL
总体积	50 μL

（2）PCR反应条件

94℃	1min	
94℃	30 s	15 个循环
58℃	2min	
72℃	30 s	
72℃	5min	
12℃	保持	

反应完后用1.5× AMPure磁珠进行纯化，最后将纯化产物溶于22μl洗脱缓冲液。

5、第二轮PCR

（1）在PCR管中按照以下反应体系配置PCR体系，其中巢氏引物池中所包含的引物如下表4所示，第二通用引物和标签引物如下表5所示。

上一步处理后的 DNA	17.5μL
2 × KAPA2G Fast ReadyMix	25μL
第二特异性引物池（10μM）	2.5μL
第二通用引物（10μM）	2.5μL
标签引物（10μM）	2.5
总体积	50 μL

（2）PCR反应条件

94℃	1min	
94℃	30 s	20 个循环
58℃	2min	
72℃	30 s	
72℃	5min	
12℃	保持	

反应完后用1.0× AMPure磁珠进行纯化，最后将纯化产物溶于22μl洗脱缓冲液。

6、文库检测：

使用 Bioanalyzer 分析系统(Agilent, Santa Clara, USA)检测文库插入片段的大小及含量，

其结果如图 7 所示。

7、上机测序

将得到的文库进行高通量测序，测序平台采用华大智造MGISEQ-2000，测序类型PE100，测序后数据经过比对后统计各项基本参数，包括下机数据、可用数据、比对率、特异性和均一性等，其结果如表 2 所示。各扩增子的测序深度如图 8 所示。

表2 测序结果

编号	原始数据	接头过滤比例	比对率	特异性	均一性
样本 1	112792	0.8%	84.3%	89.3%	100%
样本 2	131590	1.1%	85.6%	90.8%	100%
样本 3	120311	0.9%	86.1%	90.7%	100%

表2中，样本1~样本3分别代表同一个样本做了三次重复，比对率是指比对到基因组上的比例，特异性是指目标区域的reads占总测序总reads的比例，均一性是指目标区域深度大于目标区域平均深度0.1倍个数占总目标区域数的比例。

从表2、图7以及图8的结果可以看出，利用本发明所提供的扩增方法接头过滤比例在1%上下，引物二聚体少，比对率在84-86%之间，特异性在89-90%之间，性能良好，而且各扩增子之间的覆盖深度均一性良好。

表3：第一特异性引物池

目标 CpG 位点	编号	序列
cg21646186	第一特异性引物 01	GGAGGYSTAGYGATTTTAG（SEQ ID NO:1）
cg19426625	第一特异性引物 02	GGGAGAATTTTGAAAATGAAATATATTTTT（SEQ ID NO:2）
cg00960700	第一特异性引物 03	TTTTYGTTTTTYGTTTTYGTTTTT（SEQ ID NO:3）
cg06310157	第一特异性引物 04	TTTTTGAATTYGAGGTATYGGTT（SEQ ID NO:4）
cg15025536	第一特异性引物 05	TTTTAATTTAGAATTTATTATTATTTGAAGTTTTA（SEQ ID NO:5）
cg12743416	第一特异性引物 06	ATTTGGATYGTATTTTTAAGATATTTAATTATTAA（SEQ ID NO:6）
cg07382129	第一特异性引物 07	TGTGTTTYTATAAAGGTTAGGAGTTT（SEQ ID NO:7）
cg24084681	第一特异性引物 08	GGGTGGTTGATTTATGTAYGG（SEQ ID

		NO:8）
cg06837426	第一特异性引物 09	AGATTGTGYGGTAGTAAGTTTTT（SEQ ID NO:9）
cg00648301	第一特异性引物 10	GTTTGTTTGYGYGTTTG（SEQ ID NO:10）

第一特异性引物池有上述引物等摩尔混合而成，Y碱基为C/T的兼并碱基

表4：巢式引物池

目标 CpG 位点	编号	序列
cg21646186	第二特异性引物 01	ACATGGCTACGATCCGACTTGGAGTTTYGGGGYGYGTG（SEQ ID NO:11）
cg19426625	第二特异性引物 02	ACATGGCTACGATCCGACTTTTTTTGATATTGAAAATGTAATTGGTTTTT（SEQ ID NO:12）
cg00960700	第二特异性引物 03	ACATGGCTACGATCCGACTTGGTYTYGGTTGGYGTTTT（SEQ ID NO:13）
cg06310157	第二特异性引物 04	ACATGGCTACGATCCGACTTGGAGTATTTTATTTTTGTTGTTTATTATTATTTTT（SEQ ID NO:14）
cg15025536	第二特异性引物 05	ACATGGCTACGATCCGACTTGTTGAAGTGAGAATGTGATTATTAATTTTT（SEQ ID NO:15）
cg12743416	第二特异性引物 06	ACATGGCTACGATCCGACTTGTGTGTGTGTGTGTATTTATATATTTATATAAAA（SEQ ID NO:16）
cg07382129	第二特异性引物 07	ACATGGCTACGATCCGACTTTTAGAATTGAGATTAGAGAGGTAAGTAATG（SEQ ID NO:17）
cg24084681	第二特异性引物 08	ACATGGCTACGATCCGACTTGTTAAGTTGAAAAGTTGAATTTGTTTTT（SEQ ID NO:18）
cg06837426	第二特异性引物 09	ACATGGCTACGATCCGACTTYGGGTTGTTTTTGTATTTATTGTTG（SEQ ID NO:19）
cg00648301	第二特异性引物 10	ACATGGCTACGATCCGACTTGTATTTYGGTAATTTYGAGGTTG（SEQ ID NO:20）

第二特异性引物池有上述引物等摩尔混合而成，Y碱基为C/T的兼并碱基

表5：通用引物

编号	序列
第一通用引物	CGCTTGGCCTCCGACTT（SEQ ID NO:21）

第二通用引物	/5Phos/#GAACGACATGGCTACGATCCGACTT （SEQ ID NO:22）
标签引物	TGTGAGCCAAGGAGTTNNNNNNNNNNTTGTCTTCCTAA GACCGCTTGGCCTCCGACTT（SEQ ID NO:23）

其中，N碱基为MGI测序平台上的barcode序列。

在本发明的描述中，术语“第一”、“第二”等仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本发明的描述中，“多个”的含义是至少两个，例如两个，三个等，除非另有明确具体的限定。

在本发明中，除非另有明确的规定和限定，术语“相连”、“连接”、“固定”等术语应做广义理解，例如，可以是固定连接，也可以是可拆卸连接，或成一体；可以是机械连接，也可以是电连接或彼此可通讯；可以是直接相连，也可以通过中间媒介间接相连，可以是两个元件内部的连通或两个元件的相互作用关系，除非另有明确的限定。对于本领域的普通技术人员而言，可以根据具体情况理解上述术语在本发明中的具体含义。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、 “示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

尽管上面已经示出和描述了本发明的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本发明的限制，本领域的普通技术人员在本发明的范围内可以对上述实施例进行变化、修改、替换和变型。

权利要求书

1、一种基于甲基化 DNA 目标区域构建测序文库的方法，其特征在于，包括：

（1）基于所述甲基化 DNA 样本，在所述甲基化 DNA 样本的至少一端连接有通用序列，重亚硫酸氢盐处理 DNA 样本，以便获得经转化的带有通用序列的 DNA 样本；

（2）利用第一特异性引物和第一通用引物对所述经转化的带有通用序列的 DNA 样本进行第一扩增，以便获得第一扩增产物；

其中，所述第一特异性引物位于所述目标区域的上游，所述第一通用引物与所述通用序列至少部分匹配或重叠；

（3）利用第二特异性引物、第二通用引物和标签引物对所述第一扩增产物进行第二扩增，获得第二扩增产物，得到测序文库；

其中所述第二特异性引物、所述第二通用引物和所述标签引物如（i）或（ii）所示：

（i）所述第二特异性引物位于所述第一特异性引物的下游和所述目标区域的上游，所述第二通用引物与所述第二特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第一通用引物的部分序列重叠；

（ii）所述第二特异性引物位于所述目标区域的下游，所述第二通用引物和所述第一特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第二特异性引物的部分序列重叠。

2、根据权利要求 1 所述的方法，其特征在于，步骤（3）中所述第二特异性引物的 5’端与所述第二通用引物的 3’端至少部分序列重叠，所述标签引物的 3’端与所述第一通用引物的 5’端的部分序列重叠。

3、根据权利要求 1 所述的方法，其特征在于，步骤（3）中所述第二特异性引物的 5’端与所述标签引物的 3’端的至少部分序列重叠，所述第二通用引物的 3’端与所述第一特异性引物的 5’端的部分序列重叠。

4、根据权利要求 1 所述的方法，其特征在于，所述标签序列的长度为 8~12bp。

5、根据权利要求 1 所述的方法，其特征在于，步骤（1）进一步包括：

（1-a）利用重亚硫酸氢盐对所述甲基化 DNA 样本进行处理，以便获得经转化的 DNA 样本；

（1-b）利用 DNA 聚合酶和第一测序引物，对所述经转化的 DNA 样本进行复制，以便获得所述经转化的带有通用序列的 DNA 样本，所述第一测序引物的 3’端为随机碱基，所述第一测序引物的 5’端为通用序列。

6、根据权利要求 5 所述的方法，其特征在于，所述随机碱基为 6~12 个，所述随机碱基为 A、T、C 或者 G；

任选地，所述随机碱基为 6~12 个，所述随机碱基为 A、T 或者 C。

7、根据权利要求 5 所述的方法，其特征在于，所述通用序列为测序接头序列或固定序列；

任选地，所述测序接头序列或者所述固定序列中胞嘧啶为甲基化修饰的胞嘧啶。

8、根据权利要求 1 所述的方法，其特征在于，步骤（1）进一步包括：

（1-1）对所述甲基化 DNA 样本进行末端修复加 A，以便获得修复的 DNA 样本；

（1-2）将所述修复的 DNA 样本的至少一端与通用序列连接，以便获得带有通用序列的 DNA 样本；

（1-3）利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

9、根据权利要求 8 所述的方法，其特征在于，所述通用序列选自下列中的至少一种：

测序接头序列或者经修饰的测序接头序列；

任选地，所述经修饰的测序接头序列为一条链胞嘧啶进行甲基化修饰，一条链胞嘧啶不进行甲基化修饰、一条链 3’端碱基经非羟基修饰的测序接头序列、带有固定序列和随机序列的测序接头序列、或者一条链 3’端碱基经非羟基修饰的带有固定序列和随机序列的测序接头序列；

任选地，所述随机序列为分子标签序列。

10、根据权利要求 1 所述的方法，其特征在于，步骤（1）进一步包括：

①利用转座酶对所述 DNA 样本进行打断和转座处理，以便获得带有通用序列的 DNA 样本，所述转座酶中包埋有通用序列；

②利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

11、根据权利要求 10 所述的方法，其特征在于，所述通用序列为转座酶效应序列或者带有测序接头的 Tn5 转座酶效应序列，优选为转座酶效应序列；

优选地，转座酶效应序列中的胞嘧啶为甲基化修饰的胞嘧啶。

12、根据权利要求 1 所述的方法，其特征在于，所述甲基化 DNA 样本为基因组 DNA、片段化的基因组 DNA、或者游离 DNA。

13、一种基于甲基化 DNA 目标区域构建测序文库的系统，其特征在于，包括：

通用转化模块，所述通用转化模块基于所述甲基化 DNA 样本，在所述甲基化 DNA 样本的至少一端连接有通用序列，经重亚硫酸氢盐处理 DNA 样本，以便获得经转化的带有通用序列的 DNA 样本；

第一扩增模块，所述第一扩增模块和所述通用转化模块相连，所述第一扩增模块利用第一特异性引物和第一通用引物对所述经转化的带有通用序列的 DNA 样本进行第一扩增，以便获得第一扩增产物，其中所述第一特异性引物位于所述目标区域的上游，所述第一通用引物与所述通用序列至少部分匹配或重叠；

第二扩增模块，所述第二扩增模块和所述第一扩增模块相连，所述第二扩增模块利用第二特异性引物、第二通用引物和标签引物对所述第一扩增产物进行第二扩增，获得第二扩增产物，得到测序文库；

其中所述第二特异性引物，所述通用引物和所述标签引物如（i）或（ii）所示：

（i）所述第二特异性引物位于所述第一特异性引物的下游和所述目标区域的上游，所述第二通用引物与所述第二特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第一通用引物的部分序列重叠；

（ii）所述第二特异性引物位于所述目标区域的下游，所述第二通用引物和所述第一特异性引物的至少部分序列重叠，所述标签引物中含有标签序列，所述标签引物与所述第二特异性引物的部分序列重叠。

14、根据权利要求 13 所述的系统，其特征在于，所述第二扩增模块中所述第二特异性引物的 5’端与所述第二通用引物的 3’端的至少部分序列重叠，标签引物的 3’端与所述第一通用引物的 5’端的部分序列重叠。

15、根据权利要求 13 所述的系统，其特征在于，所述第二扩增模块中所述第二特异性引物的 5’端与所述标签引物的 3’段的至少部分序列重叠，所述第二通用引物的 3’端与所述第一特异性引物的 5’端的部分序列重叠。

16、根据权利要求 13 所述的系统，其特征在于，所述标签序列的长度为 8~12bp。

17、根据权利要求 13 所述的系统，其特征在于，所述通用转化模块进一步包括：

转化单元，所述转化单元利用重亚硫酸氢盐对所述甲基化 DNA 样本进行处理，以便获得经转化的 DNA 样本；

扩增单元，所述扩增单元与所述转化单元相连，所述扩增单元利用 DNA 聚合酶和第一测序引物，对所述经转化的 DNA 样本进行复制，以便获得所述经转化的带有通用序列的 DNA 样本，所述第一测序引物的 3’端为随机碱基，所述第一测序引物的 5’端为通用序列。

18、根据权利要求 17 所述的系统，其特征在于，所述随机碱基为 6~12 个，所述随机碱基为 A、T、C 或者 G；

任选地，所述随机碱基为 6~12 个，所述随机碱基为 A、T 或者 C。

19、根据权利要求 17 所述的系统，其特征在于，所述通用序列为测序接头序列或固定序列；

任选地，所述测序接头序列或者所述固定序列中胞嘧啶为甲基化修饰的胞嘧啶。

20、根据权利要求 13 所述的系统，其特征在于，所述通用转化模块进一步包括：

修复单元，所述修复单元用于对所述甲基化 DNA 样本进行末端修复加 A，以便获得修复的 DNA 样本；

连接单元，所述连接单元与所述修复单元相连，所述连接单元用于将所述修复的 DNA 样本的至少一端与通用序列连接，以便获得带有通用序列的 DNA 样本；

转化单元，所述转化单元与所述连接单元相连，所述转化单元利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

21、根据权利要求 20 所述的系统，其特征在于，所述通用序列选自下列中的至少一种：

测序接头序列或者经修饰的测序接头序列；

任选地，所述经修饰的测序接头序列为一条链胞嘧啶进行甲基化修饰，一条链胞嘧啶不进行甲基化修饰、一条链 3’端碱基经非羟基修饰的测序接头序列、带有固定序列和随机

序列的测序接头序列、或者一条链 3’端碱基经非羟基修饰的带有固定序列和随机序列的测序接头序列；

任选地，所述随机序列为分子标签序列。

22、根据权利要求 13 所述的系统，其特征在于，所述通用转化模块进一步包括：

转座单元，所述转座单元利用转座酶对所述 DNA 样本进行打断和转座处理，以便获得带有通用序列的 DNA 样本，所述转座酶中包埋有通用序列；

转化单元，所述转化单元与所述转座单元相连，所述转化单元利用重亚硫酸氢盐对所述带有通用序列的 DNA 样本进行处理，以便获得所述经转化的带有通用序列的 DNA 样本。

23、根据权利要求 22 所述的系统，其特征在于，所述通用序列为转座酶效应序列或者带有测序接头的转座酶效应序列，优选为转座酶效应序列；

优选地，所述转座酶效应序列中的胞嘧啶为甲基化修饰的胞嘧啶。

24、根据权利要求 13 所述的系统，其特征在于，所述甲基化 DNA 样本为基因组 DNA、片段化的基因组 DNA 或者游离 DNA。

25、一种对甲基化 DNA 样本进行测序的方法，其特征在于，包括：

基于所述甲基化 DNA 样本，根据权利要求 1~12 任一项所述的方法或者利用权利要求 13~24 任一项所述的系统构建得到测序文库；

对所述测序文库进行高通量测序，以便获得测序结果。

26、根据权利要求 25 所述的方法，其特征在于，利用测序平台对所述测序文库进行高通量测序，所述测序平台选自 MGISEQ、Illumina、Proton 中的至少一种。

27、一种确定甲基化 DNA 样本的甲基化状态的方法，其特征在于，包括：

基于所述甲基化 DNA 样本，根据权利要求 1~12 任一项所述的方法或者利用权利要求 13~24 任一项所述的系统构建得到测序文库；

对所述测序文库进行高通量测序，以便获得测序结果；

将所述测序结果与参考基因组进行比对，以便确定所述甲基化 DNA 样本的甲基化状态。

28、根据权利要求 27 所述的方法，其特征在于，所述参考基因组为人类基因组 hg19 或者炎黄基因组。

29、一种试剂盒，其特征在于，包括：通用序列，标签引物，第一通用引物，第二通用引物和甲基化检测试剂；

其中所述标签引物中含有标签序列，所述第一通用引物与所述通用序列的至少部分匹配或重叠，所述第一通用引物为 SEQ ID NO:21，所述第二通用引物为 SEQ ID NO:22。

30、根据权利要求 29 所述的试剂盒，其特征在于，所述标签引物为 SEQ ID NO:23 所示。

31、根据权利要求 29 或 30 所述的试剂盒，其特征在于，进一步包括：第一特异性引物和第二特异性引物，所述第一特异性引物包括 SEQ ID NO:1~SEQ ID NO:10 所示序列，所述第二特异性引物包括 SEQ ID NO:11~SEQ ID NO:20 所示序列。

32、根据权利要求 29 所述的试剂盒，其特征在于，所述试剂盒利用权利要求 1~12 中

所述的方法基于甲基化 DNA 目标区域构建测序文库。

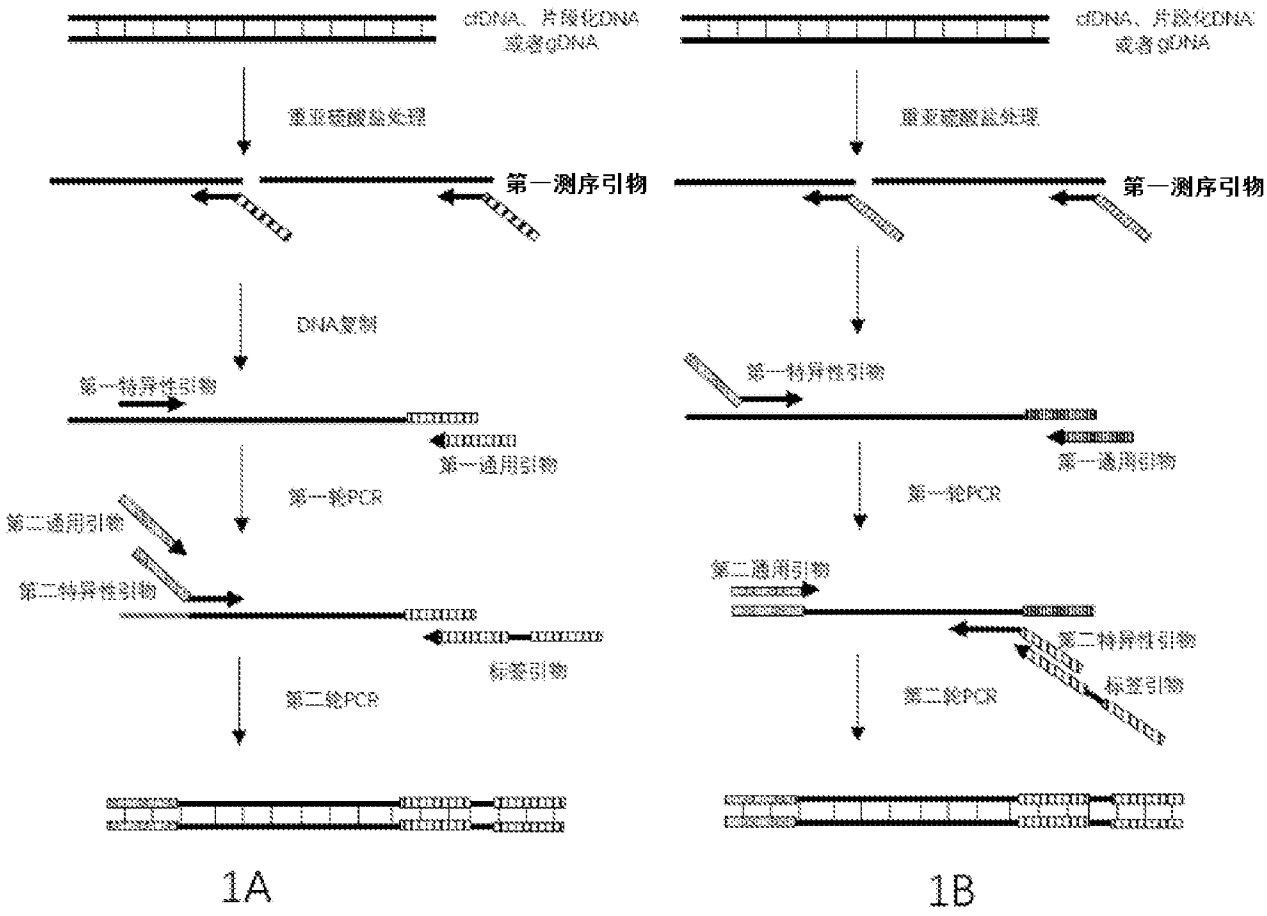

cDNA、片段化DNA
或者gDNA
重亚硫酸盐处理
第一测序引物
DNA复制
第一特异性引物
第一通用引物
第一轮PCR
第二通用引物
第二特异性引物
标签引物
第二轮PCR
1A
cDNA、片段化DNA
或者gDNA
重亚硫酸盐处理
第一测序引物
第一特异性引物
第一通用引物
第一轮PCR
第二通用引物
第二特异性引物
标签引物
第二轮PCR
1B

图 1

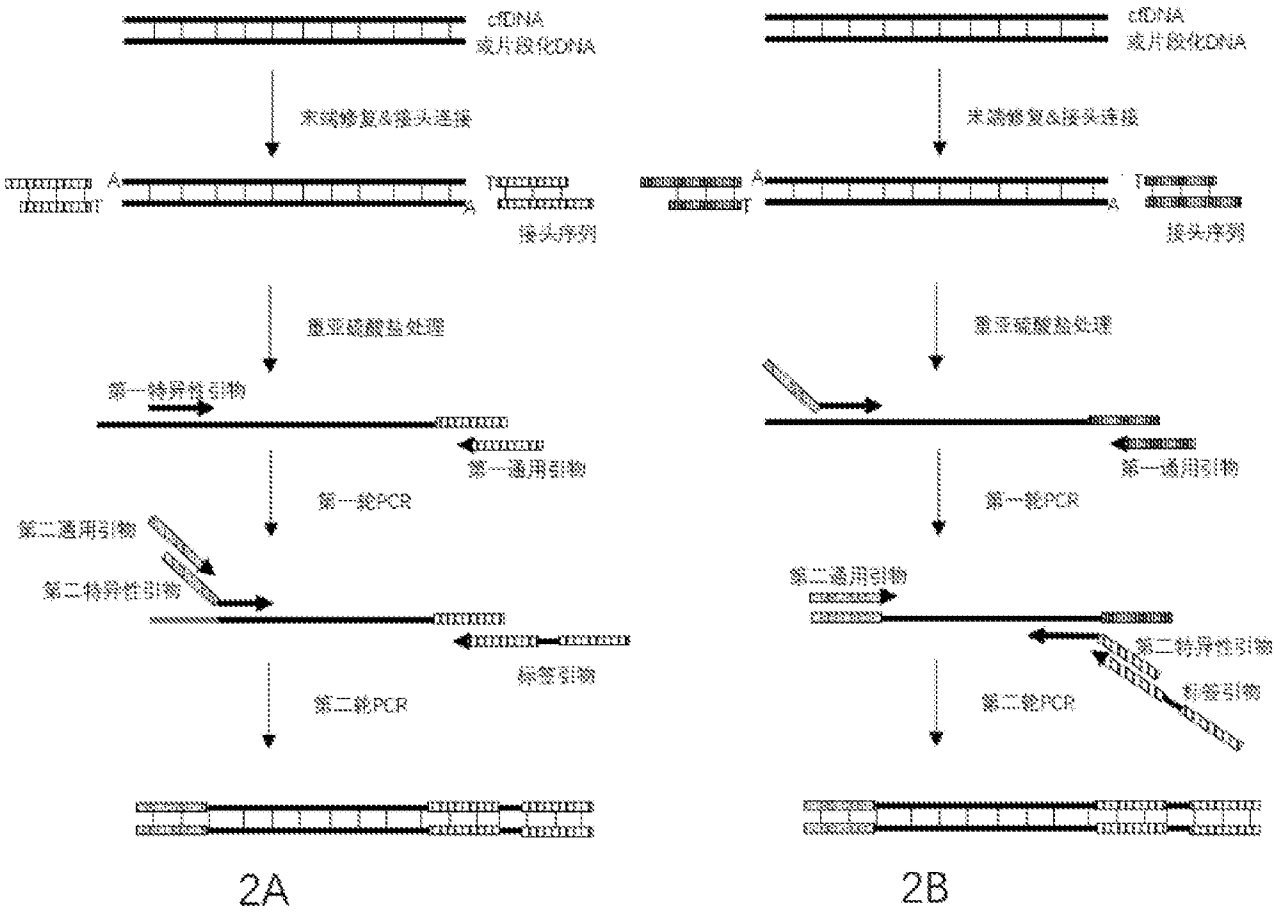

cfDNA
或片段化DNA
末端修复&接头连接
接头序列
重亚硫酸盐处理
第一特异性引物
第一通用引物
第一轮PCR
第二通用引物
第二特异性引物
标签引物
第二轮PCR
2A
cfDNA
或片段化DNA
末端修复&接头连接
接头序列
重亚硫酸盐处理
第一通用引物
第一轮PCR
第二通用引物
第二特异性引物
标签引物
第二轮PCR
2B

图 2

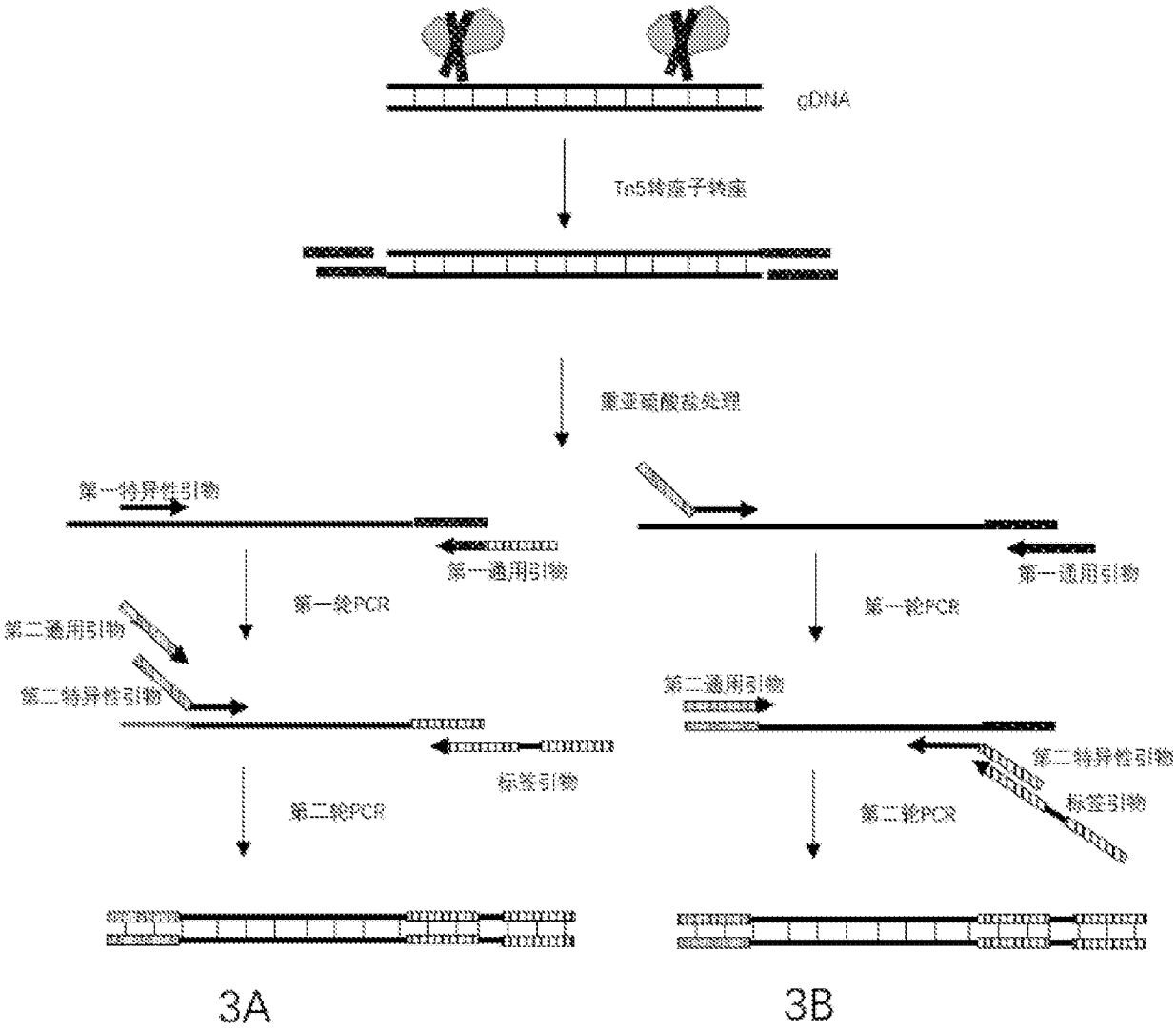

gDNA
Tn5转座子转座
变性破膜处理
第一特异性引物
第一通用引物
第一轮PCR
第二通用引物
第二特异性引物
标签引物
第二轮PCR
第一通用引物
第一轮PCR
第二通用引物
第二特异性引物
标签引物
第二轮PCR
3A
3B

图 3

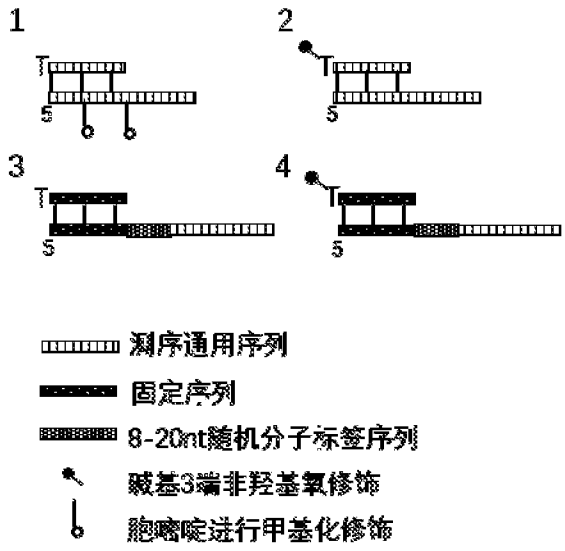

1
2
3
4
5
测序通用序列
固定序列
8-20nt随机分子标签序列
碱基3端非羟基氧修饰
胞嘧啶进行甲基化修饰

图 4

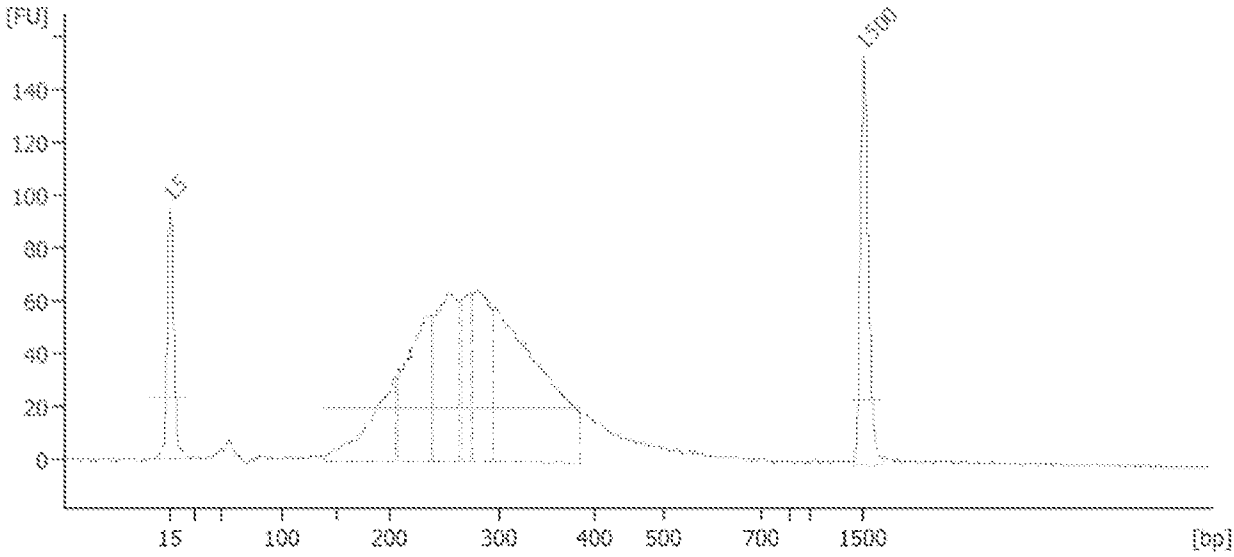

[FU]
140
120
100
80
60
40
20
0
LS
LM
15
100
200
300
400
500
700
1500
[bp]

图 5

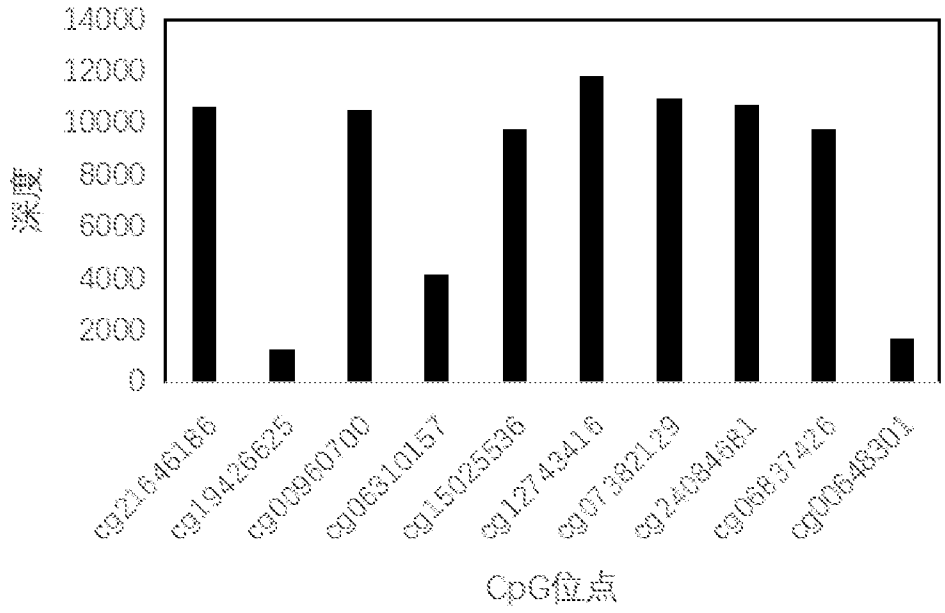

14000
12000
10000
8000
6000
4000
2000
0
甲基化水平
cg22646166
cg19426625
cg00960700
cg06310157
cg15025536
cg12743416
cg07382129
cg24084681
cg06837426
cg00648301
CpG位点

图 6

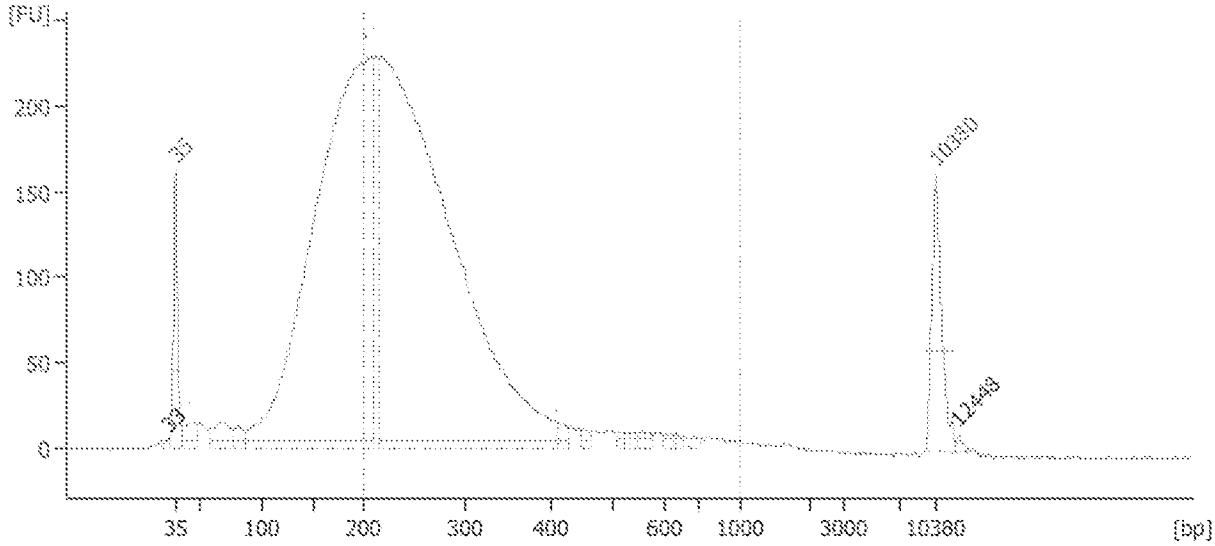

[FU]
200
150
100
50
0
35
100
200
300
400
600
1000
3000
10380
[bp]
35
10380
12448

图 7

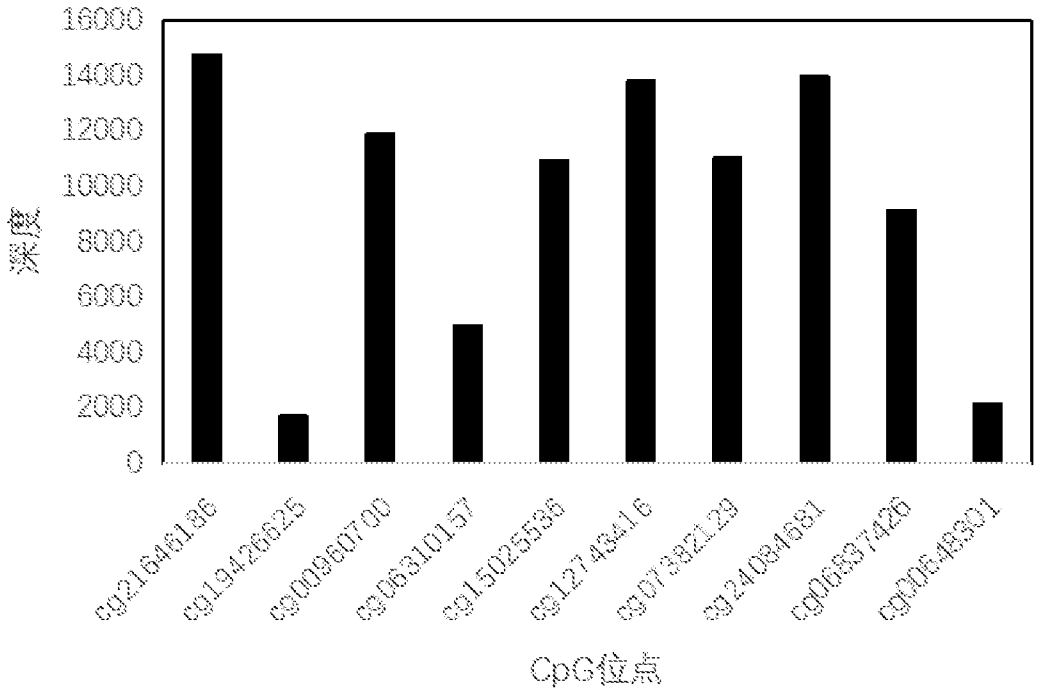

16000
14000
12000
10000
8000
6000
4000
2000
0
深度
cg21646136
cg19426625
cg00960700
cg06310157
cg15025536
cg12743416
cg07382129
cg24084681
cg06837426
cg20648301
CpG位点

图 8



图 9

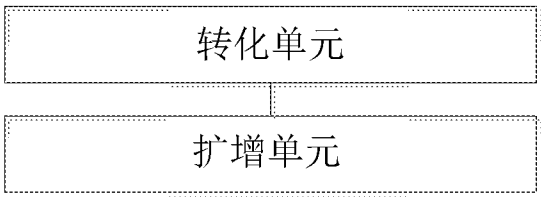


图 10



图 11

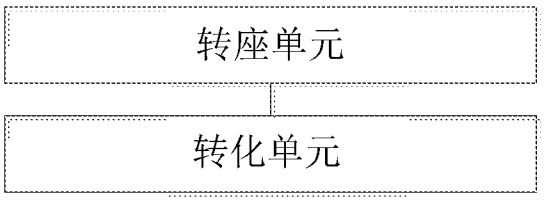


图 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/087824

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/6806(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, C12N

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, DWPI, SIPOABS, CNKI, NCBI, ISI Web of Science: 甲基化, 文库, 测序, 亚硫酸, 扩增, 聚合酶链, 重叠, methylat+, library, sequenc+, sulfite, sulphite, bisulfite, bisulphite, amplif+, polymerase chain reaction, PCR, overlap; GenBank, 中国专利生物序列检索系统: SEQ ID NO: 21, 22 of the present application

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 109666720 A (BEIJING ZHONGKE GENETICS AND REPRODUCTIVE MEDICINE RESEARCH INSTITUTE CO., LTD.) 23 April 2019 (2019-04-23) entire document, in particular embodiments	1-32
A	CN 107541791 A (BEIJING INSTITUTE OF GENOMICS, CHINESE ACADEMY OF SCIENCES) 05 January 2018 (2018-01-05) entire document, in particular embodiments	1-32

☐ Further documents are listed in the continuation of Box C. ☑ 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
"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	Date of mailing of the international search report
29 November 2019	**18 February 2020**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/087824

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed:
      ☑ in the form of an Annex C/ST.25 text file.
      ☐ on paper or in the form of an image file.
   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/087824

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109666720 A	23 April 2019	None	
CN 107541791 A	05 January 2018	None	

国际检索报告

国际申请号
PCT/CN2019/087824

A. 主题的分类

C12Q 1/6806(2018.01)i; C40B 50/06(2006.01)i

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

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C12Q, C40B, C12N

包含在检索领域中的除最低限度文献以外的检索文献

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））

CNABS、DWPI、SIPOABS、CNKI、NCBI、ISI Web of Science: 甲基化，文库，测序，亚硫酸，扩增，聚合酶链，重叠，methylat+, library, sequenc+, sulfite, sulphite, bisulfite, bisulphite, amplif+, polymerase chain reaction, PCR, overlap; GenBank、中国专利生物序列检索系统:本申请的SEQ ID NO：21和22

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 109666720 A (北京中科遗传与生殖医学研究院有限责任公司) 2019年 4月 23日 (2019-04-23) 全文，尤其是实施例	1-32
A	CN 107541791 A (中国科学院北京基因组研究所) 2018年 1月 5日 (2018-01-05) 全文，尤其是实施例	1-32

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“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

“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

“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

“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

“&” 同族专利的文件

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 29日	2020年 2月 18日
ISA/CN的名称和邮寄地址 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 传真号 (86-10)62019451	受权官员 吴永庆 电话号码 +86-10-62089161

国际检索报告

国际申请号
PCT/CN2019/087824

第I栏　核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列表进行的：

a. ☑ 作为国际申请的一部分提交的：
   - ☑ 附件C/ST.25文本文件形式
   - ☐ 纸件或图形文件形式

b. ☐ 根据细则13之三.1(a)仅为国际检索目的以附件C/ST.25文本文件形式与国际申请同时提交的：

c. ☐ 仅为国际检索目的在国际申请日之后提交的：
   - ☐ 附件C/ST.25文本文件形式（细则13之三.1(a)）
   - ☐ 纸件或图形文件形式（细则13之三.1(b)和行政规程第713段）

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列表的信息相同或未超出申请时提交的申请中的信息范围（如适用）的所需声明。

3. 补充意见：

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/087824

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 109666720 A	2019年 4月 23日	无	
CN 107541791 A	2018年 1月 5日	无	