(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
2012 年 2 月 16 日 (16.02.2012)

PCT

(10) 国际公布号
WO 2012/019320 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2010/001219

(22) 国际申请日: 2010 年 8 月 11 日 (11.08.2010)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人 (对除美国外的所有指定国): **中国科学院心理研究所 (INSTITUTE OF PSYCHOLOGY, CHINESE ACADEMY OF SCIENCES)** [CN/CN]; 中国北京市朝阳区大屯路甲 4 号, Beijing 100101 (CN)。 **深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。

(72) 发明人; 及

(75) 发明人/申请人 (仅对美国): **王燕 (WANG, Yan)** [CN/CN]; 中国北京市朝阳区大屯路甲 4 号, Beijing 100101 (CN)。 **叶明芝 (YE, Mingzhi)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。 **韩旭 (HAN, Xu)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。 **张秀清 (ZHANG, Xiuqing)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼, Guangdong 518083 (CN)。 **孙中生 (SUN, Zhongsheng)** [CN/CN]; 中国北京市朝阳区大屯路甲 4 号, Beijing 100101 (CN)。

(74) 代理人: **中国国际贸易促进委员会专利商标事务所 (CCPIT PATENT AND TRADEMARK LAW OFFICE)**; 中国北京市西城区阜成门外大街 2 号万通新世界广场 8 层, Beijing 100037 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV,

[见续页]

(54) Title: HIGH-THROUGHPUT SEQUENCING METHOD FOR METHYLATED DNA AND USE THEREOF

(54) 发明名称 : 一种甲基化 DNA 的高通量测序方法及其应用

(57) Abstract: The present invention provides a high-throughput sequencing method for methylated DNA, and the use thereof. Particularly, the present invention provides a high-throughput sequencing method for methylated DNA, which combines methylated DNA immunoprecipitation, removal of repetitive sequence, and bisulfite treatment. The size of sequencing library will be decreased, and the cost will be reduced by using the method disclosed in the present invention.

[见续页]

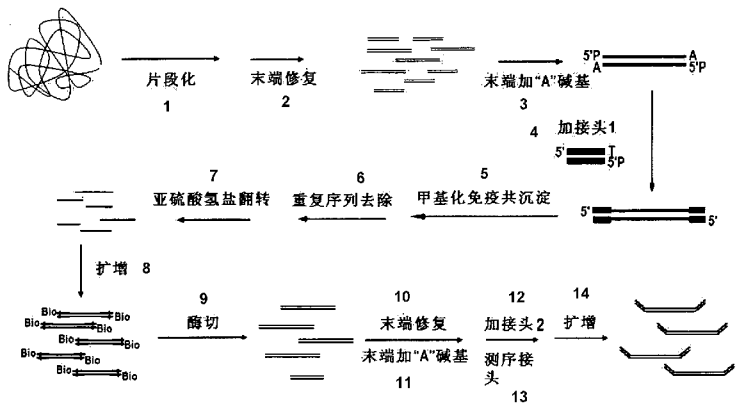


Fig.1

1 FRAGMENTATION
2 END REPAIRING
3 ADDING "A" BASE TO THE END
4 ADDING ADAPTER 1
5 DNA-METHYLATION IMMUNOPRECIPITATION
6 REPETITIVE SEQUENCES REMOVAL
7 BISULFITE CONVERSION
8 AMPLIFICATION
9 ENZYME DIGESTION
10 END REPAIRING
11 ADDING "A" BASE TO THE END
12 ADDING ADAPTER 2
13 SEQUENCING ADAPTER
14 AMPLIFICATION

WO 2012/019320 A1

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, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, 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, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, 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))。

— 包括经修改的权利要求及声明(条约第 19 条(1))。

(57) 摘要:

本发明提供了一种甲基化 DNA 的高通量测序方法及其应用。具体而言，本发明提供了将甲基化 DNA 免疫共沉淀，重复序列去除和亚硫酸氢盐处理三者结合的甲基化 DNA 的高通量测序方法。通过本发明公开的方法减小了测序文库大小，降低了测序成本。

一种甲基化DNA的高通量测序方法及其应用

技术领域

本发明涉及基因组学、生物技术领域，具体而言涉及将甲基化DNA免疫共沉淀、重复序列去除以及亚硫酸氢盐处理高通量测序技术结合，进行基因组功能区的胞嘧啶5’碳原子的甲基化状态进行精确测序的方法；进一步还提供了能够实施这种测序的装置，进而能够使测序成本降低、并且减少信息处理量，更高效地对甲基化DNA进行高通量测序。

背景技术

DNA甲基化与基因调控或疾病间的关系

在高等真核生物基因组中，DNA甲基化是在不改变DNA碱基种类与数量的前提下使得被修饰DNA的空间结构发生改变导致基因的沉默或过度表达，从而使生物体表型呈现出多样化。

例如，在正常细胞中常在CpG位点发生甲基化，但在启动子区域的CpG岛不发生甲基化。肿瘤细胞的整体DNA甲基化水平显著降低，基因丰度低的区域发生显著的去甲基化。这种DNA的低甲基化导致染色体的不稳定性与癌症的发生。如像睾丸特异性基因、黑素瘤相关基因和与增殖相关的基因在体细胞中被沉默，其启动子CpG岛被甲基化，而在相应的癌细胞中，这些启动子发生去甲基化，从而使得这些基因能够表达。另外，甲基化水平的降低促使了某些基因（如：与增殖相关的转录因子）的表达。在肿瘤发育过程中，DNA甲基化程度的降低还会使得损伤进一步的加剧，促进从良性扩增到恶性扩增的转变。

DNA甲基化对基因表达模式以及基因组稳定性均起着至关重要的作用。鉴于DNA甲基化在人类疾病发生、发展过程中的重要作用已经得到了世界上大多数研究者的普遍重视，已经成为当前研究的热点之一，并

且DNA甲基化修饰作用于全基因组水平，因此其检测技术的发展左右着对甲基化的研究与认识，进而在很大程度上影响着研究者们对于人类疾病特别是癌症相关疾病的研究。

DNA甲基化测序方法的现状

目前，现有的基于测序仪的DNA甲基化检测技术根据建序库方法的不同可分为直接亚硫酸氢盐测序、MeDIP测序、MBD测序、酶切-亚硫酸氢盐测序等几种方法。

● 亚硫酸氢盐（bisulfite）直接测序

亚硫酸盐直接测序法的主要步骤包括断裂DNA、DNA片段末端修复、连接甲基化测序接头、亚硫酸氢盐翻转、PCR扩增、测序和序列比较。具体而言是将断裂的DNA在经过末端修饰和3'端加“A”碱基后，直接连接到甲基化的测序接头（adapter）（接头上所有位点修饰成甲基化状态）上，在合适的反应条件下，针对单链DNA分子，使用亚硫酸氢盐脱去未甲基化胞嘧啶的氨基而使之转变成尿嘧啶，而甲基化的胞嘧啶保持不变，即进行亚硫酸氢盐翻转。然后进行PCR扩增使尿嘧啶全部转化成胸腺嘧啶。最后，对PCR产物进行测序，并且与未经处理的序列比较，判断CpG位点是否发生甲基化。

这种测序方法已经应用在拟南芥甲基化测序和人细胞系测序中，均得到了billion级的数据，测深分别达到20×和14×，即基因组碱基平均测序深度分别达到20倍和14倍。

该测序方法虽然解决了在全基因水平对DNA甲基化谱扫描的序列高通量问题，但是该方法产生了海量核苷酸序列，这也带来了新问题：首先是海量数据分析问题。尤其是对于高等哺乳动物这种庞大基因组的测序数据分析（20×的覆盖率，大约达到600亿碱基对的数量），测序后对海量数据的拼接和重新比对，工作量巨大，工作步骤复杂。其次是测

序成本问题。按此方法即便使用最新型的 3G 测序芯片，仅测序成本一项就很昂贵，所以无法适应大多数分子生物学实验室作为常用实验技术的要求。

● MeDIP 测序和 MBD 测序

由于在哺乳动物中甲基化一般发生在 CpG 的胞嘧啶 5 位碳原子上，所以可通过特异性结合甲基化 DNA 的蛋白 MBD 或 5'-甲基胞嘧啶抗体 MeDIP 富集高甲基化的 DNA 片段。并结合第二代高通量测序，对富集到的 DNA 片段进行测序。具体而言，MDB 方法分离甲基化 DNA 片段的方法被称为甲基化 CpG 免疫沉淀(MCIp)。MeDIP 是通过 5-甲基胞嘧啶抗体可用来进行免疫沉淀高特异性的富集甲基化 DNA 片段，5-甲基胞嘧啶抗体也可以结合非 CpG 位点的单个甲基化的胞嘧啶，因此比 MBD 有更高的特异性。这项技术被称为甲基化 DNA 免疫沉淀，结合新一代测序技术可以高通量的筛选异常甲基化的基因，此法避免了应用限制性酶在酶切位点上的局限性。

当采用 MeDIP 测序、MBD 测序时，要进行测序文库制备，基因组 DNA 经片段化之后，连接没有化学基团修饰的测序接头，连接好后先用 MBD 或者 5-甲基胞嘧啶抗体将含有甲基化胞嘧啶的 DNA 片段与未甲基化的 DNA 片段分离。甲基化的 DNA 片段纯化后不经亚硫酸氢盐翻转直接进行 PCR 和测序。

例如，David Serre 对 HCT116 结肠癌细胞系 DNA 进行 MBD 测序，测序结果显示将近一千九百万（占用芯片中两道）的数据可检测到所有已知的甲基化和一些未知甲基化区域，大大降低了测序费用。但是由于该技术没有在测序前进行亚硫酸氢盐的磺酸化处理，必需对甲基化位点进行鉴定，大大增加了后续的工作量。

● 酶切-亚硫酸氢盐测序

基于酶切方法的亚硫酸氢盐测序的目的是富集待测 DNA 片段，降低

测序 DNA 文库大小并降低测序费用。该方法能够成功地富集到一些 CpG 岛（所测数据的 8%比对到了不同的 CpG 岛）。而且该方法在一定程度上降低了测序 DNA 文库大小，并且利用亚硫酸氢盐翻转之后，不用再做后续的甲基化位点的鉴定工作。

例如，已知 Michael Zeschnigk 利用基于酶切方法的亚硫酸氢盐测序时利用 4 种内切酶对 CpG 富集区进行 DNA 片段的富集[Smiraglia DJ, Plass C. The study of aberrant methylation in cancer via restriction landmark genomic scanning. Oncogene 2002; 21: 5414-5426]。该方法的制作原理是 DNA 的片段化不是用超声断裂，而是利用多内切酶联合酶（MseI, Tsp 509I, NlaIII 和 Hpy CH4V）切方法断裂 DNA,MseI, Tsp509I,NlaIII 和 Hpy CH4V 酶切识别位点分别是 TTAA, AATT, CATG 和 TGCA。根据作者做的计算机预测，这四种酶的联合酶切在 DNA 片段大小，CpG 岛能被切割的个数等因素方面比其他酶联合酶切的优势大，因此选该四种酶联合酶切。酶切后选择 300bp-800bp 的片段进行纯化，连接甲基化的测序接头，经亚硫酸氢盐翻转 PCR 后进行测序。

但是该方法由于对基因组 DNA 的片段化是利用酶切完成的，酶切识别位点的固定性会导致片段大小的分布差异很大，小于 300bp 和大于 800bp 的 DNA 被舍弃将使部分基因组 DNA 不可能被测序到；另外，测序的读长只有 130bp、300bp-800bp 的片段不能被测通。因此，该方法使部分有生物学意义和功能的 DNA 甲基化不能被检测到。

无论是上述哪一种测序方法，都存在一个值得注意的问题，就是以上各种方法的测序均会致使大量无生物学功能的测序数据的产生。其原因在于在现有的建库和测序方法中，基因组由大量重复序列组成的异染色质区域数据占据测序数据的很大比例，这是因为在待测基因中含有高甲基化 CpG 的重复序列区（例如，丝粒和端粒含有重复序列尤其是高度重复序列，这些重复序列被认为参与染色体的结构组成，但是尚未发现

它直接参与基因的表达和调控）。而重复序列尤其是高度重复序列及其DNA 甲基化谱与目的基因表达关系的分析相对联系较弱[Herman JG 等人， Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A 1996; 93: 9821-9826]，因此，去除掉重复序列后只对功能区 DNA 的甲基化片段进行测序可大大节省测序成本。

重复序列去除技术

如何去除重复序列，目前已有研究人员对此进行了研究，例如将亚硫酸氢盐测序技术与高密度芯片结合，先通过芯片技术将待测甲基化DNA 的范围进行控制。已知 Agilent 和 NimbleGen 公司设计出的芯片探针位置集中于基因组的启动子和第一外显子以控制目标 DNA 的范围，这种方法可以去除如来自异染色质 DNA 片段的冗余。另外，还有以高通量测序技术为基础的单碱基多态性检测也同样用外显子捕获芯片来捕获外显子 DNA 片段以降低测序文库的大小来降低每个样本测序的检测成本。但是由于外显子捕获芯片的捕获 DNA 量有限，对后续的实验有影响，其去除冗余甲基化 DNA 的能力根本不足以满足在基因组水平的甲基化 DNA 的测序分析。

此外，COT-1 DNA 是一种荧光原位杂交和比较基因组杂交等杂交试验中重复序列的封闭的序列。发明人认为其可以作为去除重复序列的一个有力手段，已知 COT-1 DNA 富含高度和中度重复序列，是根据高、中度重复序列 DNA 变性后可以复性而单拷贝和低拷贝的 DNA 很难复性的原理制成。

COT-1 DNA 重复序列去除的常用方式如下所述：用生物素标记 COT-1 DNA，用亲和素包被磁珠，利用亲和素磁珠与生物素结合的原理，将标记有生物素的 COT-1 DNA 结合在包被有亲和素的磁珠上，得到生物素标记 COT-1 DNA－亲和素包被磁珠四者的复合体，将此复合体与可能存在

重复序列的待测 DNA 片段杂交，根据高、中度重复序列 DNA 变性后可以复性而单拷贝和低拷贝的 DNA 很难复性的原理，重复序列被杂交在标记生物素的 COT-1 DNA 上，得到含有重复序列－生物素标记 COT-1 DNA－亲和素包被磁珠五者的磁珠复合体，分离磁珠复合体并舍弃，同时回收经过磁珠处理的待测 DNA，回收的 DNA 即是去除了重复序列的 DNA 片段。

利用 COT-1 DNA 去除重复序列的特点在于位于功能区的甲基化 DNA 片段（启动子，外显子，部分内含子等区域）不会被捕获移除，而位于异染色质区的高度和中度甲基化重复序列被移除。这种重复序列去除方式能够满足基因组水平的甲基化 DNA 的测序分析的要求。

发明内容

发明人面对现有技术的缺陷，设想通过前期处理，去除冗余的甲基化序列，然后再进行测序，以此减少后期处理的信息量并降低测序成本。经过多方面的选择和调查，发明人首次构思了下述技术路线：

（1）先通过 MeDIP 或 MBD 技术从待测 DNA 中富集甲基化 DNA 片段；

（2）将（1）中富集得到甲基化 DNA 片段通 COT-1 DNA 重复序列去除技术处理，得到去除冗余序列的仅含有功能区 DNA 的甲基化片段的 DNA 文库；

（3）然后对（2）中的文库进行亚硫酸氢盐测序的高通量测序。

在该技术路线中首次将 COT-1 DNA 重复序列去除技术应用到高通量测序当中。

但是在实际实施时，遇到了很多技术问题。最关键的就是接头的匹配衔接问题。

目前，甲基化 DNA 的亚硫酸氢盐测序利用的甲基化的接头，由于该接头的所有胞嘧啶位点都是甲基化的，DNA 片段在连接甲基化的测序接头后经亚硫酸氢盐翻转之后，接头的序列不发生改变，PCR 之后仍和测序引物匹配。

但是将MeDIP和亚硫酸氢盐测序法联合应用的话，会出现下述几种情况：(1)先亚硫酸氢盐测序后MeDIP处理，在构建亚硫酸氢盐测序文库时，在亚硫酸氢盐磺酸化处理前引入甲基化修饰的接头的话，甲基化的接头的引入会造成后续的MeDIP甲基化免疫共沉淀出现“假阳性”的甲基化DNA富集。(2)先进行MeDIP处理然后进行亚硫酸氢盐测序，并在MeDIP处理前不加接头，双链DNA经MeDIP和亚硫酸氢盐之后变成单链DNA，并且序列发生改变，常规建库所用的测序接头与测序引物无法使用。(3) 先进行MeDIP处理然后进行亚硫酸氢盐测序，如果在MeDIP前连接没有甲基化的接头，亚硫酸氢盐翻转会致使接头序列发生改变而和测序引物不匹配。

为实现以上两种甲基化DNA富集技术和甲基化DNA的亚硫酸氢盐测序技术结合，发明人通过合理设计辅助接头及引物，合理安排该接头的连接和去除与各技术的使用顺序解决了上述难题。由此首次提供了甲基化DNA免疫共沉淀与重复序列去除的亚硫酸氢盐处理的高通量测序方法（MeDIP-repetitive elements removal-bisulfite，简称MRERB技术），这种方法既满足了对待测DNA片段检测的要求又降低了每个样本测序文库的大小和测序量，减少了测序后期处理信息量进而节约了检测费用，实现了低成本对功能区DNA甲基化片段进行甲基化精细图谱的绘制的目标。

具体而言，本发明提供了下述内容：

1. 一种甲基化DNA的高通量测序方法，包括建库步骤和测序步骤，

其中建库步骤是指获得待检测的甲基化DNA文库的步骤，所述建库步骤包括：

A)基因组DNA的片段化及双链DNA片段末端修复；

B)将A）得到的双链DNA与辅助接头连接，所述的辅助接头是指设计成能与修复后的双链DNA末端部分相连接的双链DNA序列，它可以使

经磺酸化处理的单链 DNA 经 PCR 后变成双链 DNA。辅助接头与修复的双链 DNA 连接的一端为连接末端，另一端为非连接末端；

C)将 B）的产物进行甲基化免疫共沉淀；

D)去除 C）的产物中的中度，高度重复序列；

E)将 D）得到的产物进行亚硫酸氢盐翻转处理；

F)将 E）得到的单链 DNA 通过引物（根据辅助接头序列设计）进行 PCR 扩增获得双链 DNA， G）酶切去除辅助接头；

其中测序步骤是指对前述建库步骤获得的文库进行序列测定，包括步骤：

H）对 G）中获得的双链 DNA 进行末端修复，连接测序接头；

J）对 H）中的产物进行 DNA 测序。

2.项目 1 的测序方法：

其中，步骤 B）中辅助接头选自下述 a-h 中的至少一种：

a 为不含酶切位点，非连接端为一条链突出结构，连接端为平末端；

b 为含酶切位点，非连接端为一条链突出结构，连接端为平末端；

c 为不含酶切位点，非连接端为树杈结构，连接端为平末端；

d 为含酶切位点，非连接端为树杈结构，连接端为平末端；

e 为不含酶切位点，非连接端为一条链突出结构，连接端为粘末端；

f 为含酶切位点，非连接端为一条链突出结构，连接端为粘末端；

g 为不含酶切位点，非连接端为树杈结构，连接端为粘末端；

h 为含酶切位点，非连接端为树杈结构，连接端为粘末端。

步骤 B）中的辅助接头，步骤 F）中的引物和步骤 G）中切割酶如下设计：

当辅助接头为 a 和 c 时，引物设计成互补于经过 E）翻转的辅助接头的序列，且在 5’末端额外连接有限制性内切酶的识别位点，所述的限

制性内切酶的识别位点位于引物上，并保证该酶的切割位点位于辅助接头与待测 DNA 连接点的上下游 5bp；所述的限制性内切酶选自：EcoP15I 和 MmeI。当进行步骤 G 时，采用针对所设计的限制性内切酶的酶切位点的酶进行单酶切；

当辅助接头为 b 和 d 时，在辅助接头中靠近辅助接头与待测 DNA 连接点的上游 5bp 内位置设计酶切位点，该酶切位点需符合以下三个特点：1）该酶切位点至少包含一个甲基化胞嘧啶；2）酶切位点内不含有非甲基化的胞嘧啶；3）该酶切位点不含有 CpG 二核苷酸位点。所述的限制性内切酶选自：AluI, BclI, BfaI, BglII, BsrGI, BspHI, CviAII, FatI, HindIII, HpyCH4V, NlaIII, NsiI, PciI, ScaI, SpeI, XbaI 等等。在引物设计成与经过 E）翻转的辅助接头互补的序列，当进行步骤 G 时，采用针对所设计的甲基化的酶切位点进行切割的酶进行单酶切；或者，在引物 5’ 端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合；当进行步骤 G 时，采用针对所设计的限制性内切酶的酶切位点的酶进行双酶切。

当辅助接头为 e 和 g 时，辅助接头和引物的设计原理与 a、c 方案相同，不同处在于与目的 DNA 片段连接方式不同：a、c 连接方式为“T-A”连接，e 和 g 为粘性末端连接。因此，对辅助接头和引物的设计原理不再赘述。

当辅助接头为 f 和 h 时，辅助接头和引物的设计原理与 b、d 方案相同，不同处在于与目的 DNA 片段连接方式不同：b、d 连接方式为“T-A”连接，f 和 h 为粘性末端连接。因此，对辅助接头和引物的设计原理不再赘述。

3.项目 1 或 2 的测序方法，其中步骤 A）包括：

a 将基因组 DNA 片段化成为双链 DNA 片段；

b 将 a 中的双链 DNA 片段末端修复为平末端;

c 将 b 中的平末端的 3’ 末端加上 A 碱基。

4. 项目 1 或 2 的测序方法，其中步骤 C）的甲基化免疫共沉淀选自 MeDIP 和 MBD。

5. 项目 1 或 2 的测序方法，其中步骤 d)的重复序列去除使用 COT1 DNA。

6. 项目 2 的测序方法，其中步骤 F）中酶切位点甲基化是指 b、d、f、h 方案，具体如以上方案所述。

7. 项目 1 的测序方法，其中的辅助接头设计成防止自体连接的结构，在 5’ 末端标记有生物素。

8. 一种执行项目 1 或 2 的方法的自动装置。

9. 项目 8 的装置，其包括

A)基因组 DNA 的片段化单元和末端修复单元;

B）将 A）得到的双链 DNA 与辅助接头连接的连接单元;

C)将 B）的产物进行甲基化免疫共沉淀的单元;

D)去除 C）的产物中的中度，高度重复序列的去除单元;

E)将 D）得到的产物进行亚硫酸盐翻转处理的单元;

F)将 E）得到的单链 DNA 通过引物进行 PCR 扩增获得双链 DNA 的单元;

G）酶切去除辅助接头的单元;

H）对 G）中获得的双链 DNA 进行末端修复的单元

J）对 H）中修复的 DNA 进行测序接头连接的单元;

K）对 J）中的产物进行甲基化 DNA 测序的测序单元。

附图说明:

图 1: 功能区甲基化 DNA 精细图测序样本制备原理图。主要实验步

骤依次为基因组 DNA 的片段化、DNA 片段末端修复、DNA 片段 3’ 端加 A 碱基、连接辅助接头、甲基化免疫共沉淀、去除甲基化 DNA 片段中的重复序列、亚硫酸氢盐翻转 DNA、使用根据辅助接头设计的引物进行 PCR 将经翻转的 DNA 变成双链、去除辅助接头、末端修去除辅助接头的 PCR 产物并在 3’ 端加 A 碱基、连接测序接头、测序。

图 2：辅助接头的设计示意图。辅助接头的设计按是否含有酶切位点分为两类，含有酶切位点的辅助接头和不含有酶切位点的接头（b, d, f, h）；辅助接头的非连接端可以设计成三种形式：树杈（forked）结构（图 c 和 d）、一条链突出结构（a 和 b）和平端结构。结合辅助接头连接端的 TA 连接或其他粘性连接特性，共分八种情况。分别如下所示：

a 为不含酶切位点，非连接端为一条链突出结构，连接端为平末端形式；

b 为含酶切位点，非连接端为一条链突出结构，连接端为平末端形式；

c 为不含酶切位点，非连接端为树杈结构，连接端为平末端形式；

d 为含酶切位点，非连接端为树杈结构，连接端为平末端形式；

e 为不含酶切位点，非连接端为一条链突出结构，连接端为粘末端形式；

f 为含酶切位点，非连接端为一条链突出结构，连接端为粘末端形式；

g 为不含酶切位点，非连接端为树杈结构，连接端为粘末端形式；

h 为含酶切位点，非连接端为树杈结构，连接端为粘末端形式。

图 3：平末端的辅助接头的连接和移除方式（1），本方式中的辅助接头使用其中图 2 中的两种：a 和 c。因为设计成采用酶切方式移除辅助接头 1，而且移除的时间点在经过亚硫酸氢盐翻转之后，所要移除的

辅助接头 1 以单链的形式存在于待测 DNA 的末端，此时为了有效移除所述辅助接头 1 同时使单链 DNA 成为双链 DNA 以方便后续测序接头连接，发明人设计出与经过亚硫酸氢盐翻转后辅助接头 1 序列相匹配的引物（A），同时在靠近引物的 5'位点处设计酶切识别位点（B），酶切识别位点不与辅助接头序列相匹配，处于引物的突出延伸部分。选用的酶切位点是与限制切割酶相适应的，所选用的限制性内切酶的特点为能在酶切识别位点下游 20-30bp 处对 DNA 进行酶切，通过控制辅助接头的长度将酶切位点设计在 DNA 片段末端连接辅助接头处 5bp 范围内。

图 4：平末端的辅助接头 1 的连接和移除方式（2）。辅助接头 1 中包含甲基化酶切位点方式的辅助接头 1 的连接和去除。本方式中的辅助接头 1 可以使用图 2 中的两种：b 和 d。辅助接头 1 中含有甲基化的酶切位点，通过甲基化酶切位点可以保证在酶切去除辅助接头时 DNA 片段的完整性（但可允许 DNA 片段末端 5bp 内的损伤）和经亚硫酸氢盐翻转之后酶切识别序列不改变。

图 5：平末端的辅助接头连接和移除方式（3）。辅助接头中包含甲基化酶切位点和引物中包含酶切位点方式的辅助接头的连接和去除。辅助接头中同样在靠近连接处设计甲基化的酶切位点。引物中在 5' 端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合。通过两次酶切保证酶切的彻底性。

图 6：粘性末端的辅助接头的连接和去除方式（1）。本方式中的辅助接头可以使用其中两种：e 和 g。设计与 bisulfite 翻转后辅助接头序列相匹配的引物，同时在靠近引物的 5' 位点处设计酶切识别位点。

图 7：粘性末端的辅助接头的连接和去除方式（2）。辅助接头中包含甲基化酶切位点方式的辅助接头的连接和去除。本方式中的辅助接头可以使用其中两种：f 和 h。辅助接头中含有甲基化的酶切位点，保证 DNA 片段不被切割和经亚硫酸氢盐翻转之后酶切识别序列不改变。

图 8：粘性末端的辅助接头的连接和去除方式（3）。辅助接头中包含甲基化酶切位点和引物中包含酶切位点方式的辅助接头的连接和去除。辅助接头中同样在靠近连接处设计甲基化的酶切位点。引物中在 5’端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合。通过两次酶切保证酶切的彻底性。

图 9：去除甲基化的重复序列示意图。COT-1 DNA 标记生物素后，利用亲和素磁珠与生物素结合的原理，将标记有生物素的 COT-1 DNA 标记在包被有亲和素的磁珠上并去除未标记生物素的 COT-1 DNA 片段。将与生物素-DNA 结合的亲和素磁珠与经甲基化免疫共沉淀后的甲基化 DNA 片段进行杂交，甲基化的重复序列将会被杂交在标记生物素的 COT-1 DNA 上，分离磁珠并舍弃磁珠，将溶液中的 DNA 收回，回收的 DNA 即功能区甲基化的 DNA 片段。

图 10:全基因磺酸化处理测序与本技术测序结果共同比对的单个 CpG 位点的测序深度分析。比对说明单个 CpG 位点的测序深度趋势一致。其中左图为 MRERB 测序结果；右图为全基因磺酸化测序结果。

图 11-图 18 为所有类型的辅助接头和引物序列说明：

图 11：平末端的辅助接头连接和去除方式（1）的接头序列和引物序列，本辅助接头连接和去除方式中的接头序列不包含特殊的碱基序列，但要保证该接头经磺酸化后能成功扩增。本接头的连接端的 3‘端设计了“T”碱基的突出、5’端修饰了磷酸基团以便和 DNA 连接时能以“T-A”连接式相连接。在 PCR 产物中接头去除方面，我们在 PCR 引物的 5‘端引入酶识别序列，该酶的特点是能在酶识别位点下游的大于 20bp 的地方切割 DNA 片段，可以将辅助接头完全切去。另外可以引物的 5’端引入生物素基团（也可以不修饰生物素基团，不修饰的情况下去除小片段时可以用 Ampure Beads 去除。）。因此，去除接头前的 PCR 产物片段包括待测 DNA 片段、对称的两个酶切位置、对称的匹配接头序列及

其对称的两个酶识别位点序列。

图 12：平末端的辅助接头连接和去除方式（2）的接头序列和引物序列。本辅助接头连接和去除方式中的接头序列包含要保证该接头经磺酸化后能成功扩增的 20-30bp 的序列和一个甲基化的酶识别序列。该酶的特征是酶识别位点和切割位点相同。本接头的连接端的 3‘端设计了“T”碱基的突出、5’端修饰了磷酸基团以便和 DNA 连接时能以“T-A”连接式相连接。在 PCR 产物中接头去除方面，由于该酶在接头中做了甲基化修饰，因为经过磺酸化之后序列不变，用该酶进行酶切可把接头去除。另外可以引物的 5’端引入生物素基团（也可以不修饰生物素基团，不修饰的情况下去除小片段时可以用 Ampure Beads 去除。）。因此，去除接头前的 PCR 产物片段包括待测 DNA 片段、对称的两个酶识别切割序列、对称的匹配接头序列。

图 13：平末端的辅助接头连接和去除方式（3）的接头序列和引物序列。本辅助接头连接和去除方式中的接头序列包含要保证该接头经磺酸化后能成功扩增的 20-30bp 的序列和一个甲基化的酶识别序列。该酶的特征是酶识别位点和切割位点相同。本接头的连接端的 3‘端设计了“T”碱基的突出、5’端修饰了磷酸基团以便和 DNA 连接时能以“T-A”连接式相连接。在 PCR 产物中接头去除方面，由于该酶在接头中做了甲基化修饰，因为经过磺酸化之后序列不变，用该酶进行酶切可把接头去除；另外，我们同时在 PCR 产物中接头去除方面，我们在 PCR 引物的 5‘端引入酶识别序列，该酶的特点是能在酶识别位点下游的大于 20bp 的地方切割 DNA 片段，可以将辅助接头完全切去。另外可以引物的 5’端引入生物素基团（也可以不修饰生物素基团，不修饰的情况下去除小片段时可以用 Ampure Beads 去除。）。因此，去除接头前的 PCR 产物片段包括待测 DNA 片段、对称的两个酶切位共同切割的位置、对称的匹配接头序列及酶 2 对称的两个酶识别位点序列。

图 14：粘性末端的辅助接头连接和去除方式（1）的接头序列和引物序列。本辅助接头连接和去除方式类同平末端的辅助接头连接和去除的方式（1）。不同的是本接头的连接端的 3‘端非“T-A”连接式相连接。而且设计和 DNA 末端的粘性末端序列匹配的序列。

图 15：粘性末端的辅助接头连接和去除方式（2）的接头序列和引物序列。本辅助接头连接和去除方式类同平末端的辅助接头连接和去除的方式（2）。不同的是本接头的连接端的 3‘端非“T-A”连接式相连接。而且设计和 DNA 末端的粘性末端序列匹配的序列。

图 16：粘性末端的辅助接头连接和去除方式（3）的接头序列和引物序列。本辅助接头连接和去除方式类同平末端的辅助接头连接和去除的方式（3）。不同的是本接头的连接端的 3‘端非“T-A”连接式相连接。而且设计和 DNA 末端的粘性末端序列匹配的序列。

图 17：其他接头设计方案说明。

系列一（b-1, d-1, f-1, h-1）辅助接头设计方案类似 b、d 接头设计，是在辅助接头中靠近辅助接头与待测 DNA 连接点的上游 20-30bp 内位置设计酶切位点，该酶切位点需符合以下四个特点：1）该酶切位点至少包含一个甲基化胞嘧啶；2）酶切位点内不含有非甲基化的胞嘧啶；3）该酶切位点不含有 CpG 二核苷酸位点；4）该酶的切割位点位于辅助接头与待测 **DNA** 连接点的上、下游 **5bp** 内。所述的限制性内切酶选自：**EcoP15I** 和 **MmeI** 等等。在引物设计成与经过 **E**）翻转的辅助接头互补的序列，当进行步骤 **G** 时，采用针对所设计的甲基化的酶切位点进行切割的酶进行单酶切。

系列二（b-2, d-2, f-2, h-2）辅助接头设计方案是系列一方案和 b、d、f、h 的结合，通过调整辅助接头的长度使两个酶切位点重合，该方案可以使用双酶切去除辅助接头。在引物设计成与经过 E）翻转的辅助接头互补的序列，当进行步骤 G 时，采用针对所设计的甲基化

的酶切位点进行切割的酶进行双酶切。

发明详述

定义

为了更清楚地阐述本发明，在本发明中的各个定义解释如下。

DNA片段化及修复

DNA 片段化是指：将基因组用机械或者酶切等方法将基因组 DNA 剪切成小的 DNA 片段（如几百碱基或几千碱基）。

DNA 片段的修复是指：剪切后的 DNA 片段，尤其是机械方法打断的 DNA 片段，DNA 片段双链末端可能会出现以下情况：双链同侧末端断裂位置不一致、3’末端或者 5’末端有损伤，DNA 片段修复的意义在于将这些末端补齐或者 3’、5’有损伤的片段修复成 5’连接磷酸基团和 3’连接羟基的平末端结构。

根据高通量测序仪对测序文库的要求（WO2008096146A1），需制备合适长度的 DNA 片段文库。首先将 DNA 片段化成适合测序仪测序长度大小的 DNA 片段。片段化效果一般以所要求制备的文库片段大小在所片段化 DNA Smear 的主带位置较为理想。例如，若要求制备插入 100 bp 的 Paired End 文库，则断裂后 DNA Smear 主带在 100 bp 处即可，若断裂效果不理想则需要进行重新断裂。样品片段化步骤也可使用其他断裂系统，具体参数可根据仪器的要求进行调整。

用于断裂 DNA 的方法包括酶切和超声断裂的方法。

酶切的方法可选用单一内切酶或多种不含 CpG 位点的内切酶联合酶切。酶切方法由于酶切识别位点的固定性，可能致使部分 DNA 片段长度不适合而舍弃在文库之外；而超声断裂 DNA 是随机断裂 DNA，断裂位点不固定，理论上整个基因组可包含在测序文

库之中。本发明实验方案中片段化基因组既可用酶切的方法，也可选用超声断裂的方法，但实例用超声断裂方法。

如果是由超声断裂的 DNA，需对 DNA 片段末端修复，修复成平末端。该修复的目的是将双链 DNA 片段损伤的 5’ 端磷酸化，损伤的 3 ‘端羟基化，修复成平末端的 DNA 片段 3’ 末端加 “A” 碱基使连接辅助接头时形成 “TA” 的粘性末端连接；如果是多种酶联合酶切方法片段化的基因组 DNA，由于酶切末端序列不一样，首先将 DNA 片段修饰成平末端后在 3’ 端加 “A” 碱基使连接辅助接头时形成 “TA” 的粘性末端连接；如果是单一酶切，若酶切末端是平末端在末端的 3’ 端加“A”碱基使连接辅助接头时形成“TA”的粘性末端连接；如果是粘性末端，设计辅助接头的时候可设计成与 DNA 片段粘性末端匹配的接连末端，进行粘性末端的连接。

辅助接头（adapter）的连接和切除

辅助接头是指：加入辅助接头的目的是在DNA在进行MeDIP和bisulfite磺酸化后能经过PCR反应后使单链DNA变成双链，从而能进行常规文库制备

辅助接头的类型：根据DNA片段化的方法不同以及碱基修饰的不同的分类。目前，甲基化的bisulfite测序利用的甲基化的接头（WO2009024019A1）。由于该接头的所有胞嘧啶位点都是甲基化的，DNA片段在连接甲基化的测序接头后经bisulfite翻转之后，接头的序列不发生改变，PCR之后仍和测序引物匹配。

但是甲基化接头影响了 MeDIP 实验中甲基化抗体绑定甲基化片段，因此该接头不适用于 MeDIP 后 bisulfite 甲基化测序。而如果在 MeDIP 前连接没有甲基化的接头，bisulfite 翻转会致使接头序列发生改变而和测序引物不匹配。如果在 bisulfite 前不

加接头，双链 DNA 经 MeDIP 和 bisulfite 之后变成单链 DNA，并且序列发生改变，常规建库（以 solexa 测序文库为例）所用的 Paired End adapter 与 Paired End primer 无法在本实验中使用。因此本发明引入辅助接头（图 2、图 17）解决 MeDIP 与 bisulfite 翻转导致序列改变之间的矛盾，能在不影响 MeDIP 的情况下，经 bisulfite 后成功扩增 DNA 片段库并能进行后续的常规建库。

辅助接头的的接连和剪切

辅助接头的作用是解决甲基化免疫共沉淀与 bisulfite 致使接头序列改变导致测序引物无法匹配的矛盾，因此在 bisulfite PCR 之后要把辅助接头剪切去掉，以减少测序成本。

辅助接头的接连和剪切结合起来考量可分为以下几种情况:

1. 多酶切 DNA 片段或超声断裂片段的连接辅助接头及酶切方式一---引物中包含酶切位点（图 3）。

多酶切 DNA 片段或超声断裂片段经末端修复和 3’ 端加 “A” 碱基后需连接辅助接头，本方式中的辅助接头可以使用其中两种: a 和 c。该辅助接头除了在其中一条 5’端修饰磷酸基团便于和 DNA 片段中 3’ 的羟基连接外不必做任何其他修饰。

连接辅助接头之后的 DNA 片段经 MeDIP 等一系列处理后，设计与 bisulfite 翻转后辅助接头序列相匹配的引物（图 3A），同时在靠近引物的 5’ 位点处设计酶切识别位点（图 3B），酶切识别位点不与辅助接头序列相匹配，处于引物的延伸部分。该酶的特点为能在酶切识别位点下游 20-30bp 处对 DNA 进行酶切，通过控制辅助接头的长度将酶切位点设计在 DNA 片段末端连接辅助接头处 5bp 范围内。DNA 片段经 PCR 之后在 DNA 的两端多了酶切识别序列，通过酶对双链序列的识别便可在酶切处酶切 DNA 片段。由于在引物的 5’ 标记生物素，酶切掉的辅助接头部分和未被酶

切的 DNA 片段可与亲和素的结合，能迅速、彻底的被去除。可用于引物中酶包括：EcoP15I 和 MmeI。

在本发明的一个实施方式中，采用引物中包含酶切位点方式的辅助接头的连接和去除（图 2）。本方式中的辅助接头可以使用其中两种：a 和 c。设计与 bisulfite 翻转后辅助接头序列相匹配的引物（A），同时在靠近引物的 5’ 位点处设计酶切识别位点（B），酶切识别位点不与辅助接头序列相匹配，处于引物的延伸部分。该酶的特点为能在酶切识别位点下游 20-30bp 处对 DNA 进行酶切，通过控制辅助接头的长度将酶切位点设计在 DNA 片段末端连接辅助接头处 5bp 范围内。

2. 多酶切 DNA 片段或超声断裂片段的连接辅助接头及酶切方式二---辅助接头中包含甲基化酶切位点（图 3）。

本方式中多酶切 DNA 片段或超声在断裂片段经末端修复和 3’ 端加“A”碱基后连接辅助接头可用（图 2）b 和 d 两种。b 和 d 辅助外链子在靠近连接端设计酶切位点，设计该位点的目的是为了 PCR 之后能在此处将辅助接头酶切掉去除。因此，该酶切位点设计在辅助接头的靠近连接处。由于经过亚硫酸氢盐转化之后未甲基化的胞嘧啶会转化为鸟嘧啶，并在 PCR 后变成胸腺嘧啶。发明人在设计酶切位点时如果酶识别序列中含有胞嘧啶，需将此处的胞嘧啶做甲基化修饰以保证经亚硫酸氢盐翻转后酶识别序列不发生改变。由于亚硫酸氢盐转化后的 PCR 产物序列中腺嘌呤（A）和胸腺嘧啶（T）增多，发明人在设计酶切位点时为了不切割到 DNA 片段，选择的酶切识别位点中含有一个非 CpG 位点的 C，并且该胞嘧啶不在酶切识别序列的末位。可用在此处的酶包括：AluI, BclI, BfaI, BglII, BsrGI, BspHI, CviAII, FatI, HindIII, HpyCH4V, NlaIII, NsiI, PciI, ScaI, SpeI, XbaI 等等。另外，

该辅助接头也在其中一条 5’ 端修饰磷酸基团便于和 DNA 片段中 3’ 的羟基连接。

连接辅助接头之后的 DNA 片段经 MeDIP 等一系列处理后，设计与 bisulfite 翻转后辅助接头序列相匹配的引物并在 5’ 端修饰生物素（图 4）。由于在引物 5’ 端修饰了生物素，酶切掉的辅助接头部分和未被酶切的 DNA 片段可与亲和素的结合，能迅速、彻底的被去除。

在本发明的一个实施方式中，采用辅助接头中包含甲基化酶切位点方式的辅助接头的连接和去除（图 4）。本方式中的辅助接头可以使用其中两种：图 4 的 b 和 d。辅助接头中含有甲基化的酶切位点，保证 DNA 片段不被切割和经亚硫酸氢盐翻转之后酶切识别序列不改变。

3. 多酶切 DNA 片段或超声断裂片段的连接辅助接头及酶切方式三---辅助接头中包含甲基化酶切位点和引物中包含酶切位点（图 5）。

单一种酶进行酶切时往往不够彻底，有时候有必要利用两种或多种酶进行酶切处理。本方式中多酶切 DNA 片段或超声在断裂片段经末端修复和 3’ 端加 “A” 碱基后连接辅助接头如同方式二用（图 2）b 和 d 两种，并同样在靠近连接处设计甲基化的酶切位点。

设计与 bisulfite 翻转后辅助接头序列相匹配的引物时如同方式一，在靠近引物的 5’ 位点处设计酶切识别位点（图 5），该酶在酶切识别位点下游 20-30bp 处对 DNA 进行酶切，通过控制辅助接头的长度将两个酶切位点设计在 DNA 片段的同一个位置。通过两次酶切，保证切割的彻底性。在引物的 5’ 标记生物素，酶切掉的辅助接头部分和未被酶切的 DNA 片段可与亲和素的结合，

能迅速、彻底的被去除。

在本发明的一个实施方式中，采用辅助接头中包含甲基化酶切位点和引物中包含酶切位点方式的辅助接头的连接和去除（图5）。辅助接头中同样在靠近连接处设计甲基化的酶切位点。引物中在5’端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合。通过两次酶切保证酶切的彻底性。

4. 单酶切DNA片段的连接辅助接头及酶切方式----引物中包含酶切位点（图6）。

单酶切DNA片段为粘性末端的DNA片段不必做末端修复和3’端加“A”碱基，可直接设计和酶切粘性末端匹配的辅助接头，本方式中的辅助接头可以使用其中两种：e和g。该辅助接头除了在其中一条5’端修饰磷酸基团便于和DNA片段中3’的羟基连接外不必做任何其他修饰。

连接辅助接头之后的DNA片段经MeDIP等一系列处理后，设计与bisulfite翻转后辅助接头序列相匹配的引物（图6），同时在靠近引物的5’位点处设计酶切识别位点（图6），酶切识别位点不与辅助接头序列相匹配，处于引物的延伸部分。该酶的特点为能在酶切识别位点下游20-30bp处对DNA进行酶切，通过控制辅助接头的长度将酶切位点设计在DNA片段末端连接辅助接头处5bp范围内。DNA片段经PCR之后在DNA的两端多了酶切识别序列，通过酶对双链序列的识别便可在酶切处酶切DNA片段。由于在引物的5’标记生物素，酶切掉的辅助接头部分和未被酶切的DNA片段可与亲和素的结合，能迅速、彻底的被去除。可用于引物中酶包括：EcoP15I和MmeI。

在本发明的一个实施方式中，采用粘性末端的辅助接头的连接和去除（引物中包含酶切位点方式）（图6）。本方式中的辅

助接头可以使用其中两种：e 和 g。设计与 bisulfite 翻转后辅助接头序列相匹配的引物，同时在靠近引物的 5’ 位点处设计酶切识别位点。

5. 单酶切 DNA 片段的连接辅助接头及酶切方式一---辅助接头中包含甲基化酶切位点（图 7）。

单酶切 DNA 片段为粘性末端的 DNA 片段不必做末端修复和 3’ 端加“A”碱基，可直接设计和酶切粘性末端匹配的辅助接头，本方式中的辅助接头可以使用其中两种：f 和 h。该辅助接头除了在其中一条 5’ 端修饰磷酸基团便于和 DNA 片段中 3’ 的羟基连接外不必做任何其他修饰。

辅助外链子在靠近连接端设计包含一个甲基化酶切位点，酶切位点时为了不切割到 DNA 片段，选择的酶切识别位点中含有一个非 CpG 位点的 C，并且该胞嘧啶不在酶切识别序列的末位。为了经亚硫酸氢盐翻转后酶切识别序列不发生改变，发明人将酶切识别序列里的胞嘧啶甲基化，其他碱基不做修饰。可用在此处的酶包括：AluI, BclI, BfaI, BglII, BsrGI, BspHI, CviAII, FatI, HindIII, HpyCH4V, NlaIII, NsiI, PciI, ScaI, SpeI, XbaI 等等。另外，该辅助接头也在其中一条 5’ 端修饰磷酸基团便于和 DNA 片段中 3’ 的羟基连接。

连接辅助接头之后的 DNA 片段经 MeDIP 等一系列处理后，设计与 bisulfite 翻转后辅助接头序列相匹配的引物并在 5’ 端修饰生物素（图 4）。由于在引物 5’ 端修饰了生物素，酶切掉的辅助接头部分和未被酶切的 DNA 片段可与亲和素的结合，能迅速、彻底的被去除。

在本发明的一个实施方式中，采用辅助接头中包含甲基化酶切位点方式的辅助接头的连接和去除（图 7）。本方式中的辅助接

头可以使用其中两种:f 和 h。辅助接头中含有甲基化的酶切位点，保证 DNA 片段不被切割和经亚硫酸氢盐翻转之后酶切识别序列不改变。

6. 单酶切 DNA 片段的连接辅助接头及酶切方式三---辅助接头中包含甲基化酶切位点和引物中包含酶切位点（图 8）。

在本发明的一个实施方式中，单一种酶进行酶切时往往不够彻底，有时候有必要利用两种或多种酶进行酶切处理。本方式中多酶切 DNA 片段或超声在断裂片段经末端修复和 3’ 端加“A”碱基后连接辅助接头如同方式二用（图 8）f 和 h 两种，并同样在靠近连接处设计甲基化的酶切位点。

设计与 bisulfite 翻转后辅助接头序列相匹配的引物时如同方式一，在靠近引物的 5’ 位点处设计酶切识别位点（图 8），该酶在酶切识别位点下游 20-30bp 处对 DNA 进行酶切，通过控制辅助接头的长度将两个酶切位点设计在 DNA 片段的同一个位置。通过两次酶切，保证切割的彻底性。在引物的 5’ 标记生物素（biotin），酶切掉的辅助接头部分和未被酶切的 DNA 片段可与亲和素的结合，能迅速、彻底的被去除。

在本发明的一个实施方式中采用辅助接头中包含甲基化酶切位点和引物中包含酶切位点方式的辅助接头的连接和去除（图 8）。辅助接头中同样在靠近连接处设计甲基化的酶切位点。引物中在 5’ 端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合。通过两次酶切保证酶切的彻底性。

甲基化免疫共沉淀

甲基化免疫共沉淀（MeDIP）是通过 5-甲基胞嘧啶抗体可用来进行免疫沉淀高特异性的富集甲基化 DNA 片段，结合新一代测序技术可以高通量的筛选异常甲基化的基因，此法避免了应用限

制性酶在酶切位点上的局限性。本发明所使用甲基化免疫共沉淀技术根据 Michael Weber 的免疫共沉淀技术略有改动。为了测序 DNA 片段长度的需要，将 DNA 片段大小由原来的 200-1000bp 调整为 100-300bp；另外， 5-甲基胞嘧啶抗体孵育时间经过优化后发现孵育效果最优时间为 12 小时而非文献报道的 2 小时。

亚硫酸盐翻转

亚硫酸盐翻转是采用亚硫酸氢盐对单链 DNA 分子进行化学修饰，导致未甲基化胞嘧啶(C)可被亚硫酸氢盐脱去氨基而转变成尿嘧啶(U)，而 5mC 不能被修饰，仍保持为 5mC，在 PCR 反应过程中，尿嘧啶与腺嘌呤(A)配对，尿嘧啶则被胸腺嘧啶(T)取代。这一化学反应过程首先由 Frommer 等(M. Frommer 等人，A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89 (1992) 1827-31)报导。具体过程为：第一步，亚硫酸氢钠使胞嘧啶磺酸化；第二步，对苯二酚脱氨基；第三步，在碱性环境下，磺基消失，成为尿嘧啶。目前，该实验可由天漠（zymo）bisulfite-Golden 试剂盒进行实施。

辅助接头的酶切和去除

根据辅助接头的不同设计方案中选择不同的酶，利用合适量的酶和通过延长酶切时间使接头尽可能彻底被酶切掉。考虑到酶切的不彻底性以及短片段辅助接头用普通试剂盒难去除干净，发明人在引物的 5’端修饰生物素，通过用亲和素包被的磁珠结合生物素的方法去掉短片段的辅助接头和未被酶切的 DNA 片段。

UDNA末端修复和测序接头连接

经内切酶酶切的DNA片段末端首先进行修复成可以和测序 adapter（Solexa, 454, Solid）匹配的末端。以 Solexa 为例，经酶酶切的 DNA 片段末端，若为粘性末端，首先修复成平末端，然后在 DNA 平末端的 3’

端加“A”碱基，和Solexa公司提供的PE adapter形成“T-A”连接；若酶切末端为平末端，直接在DNA平末端的3’端加“A”碱基后和Solexa公司提供的PE adapter形成“T-A”连接。

PCR

在DNA经过磺酸化（bisulfite）之后，Bisulfite-PCR的目的是将单链的DNA片段变成双链便于测序接头的连接。同时扩增DNA的量。Bisulfite PCR扩增用引物和磺酸化之后的辅助接头的序列进行匹配。

PE PCR在辅助接头酶切去除之后，加连接PE接头后扩增，扩增的目的是将DNA片段两端的PE分叉接头（forked PE adapter）变成可与测序引物匹配的序列。

测序

去除辅助接头后的DNA片段可直接用于后续Genome Sequencer FLX系统高通量测序的常规实验步骤如使用罗氏公司的GS FLX Standard DNA文库制备试剂盒；或Illumina公司Genome Analyzer（GA）系统高通量测序的常规实验步骤如使用Paired End DNA文库制备试剂盒构建常规Paired End DNA测序文库；也可以用于AB公司SOLiD系统的SOLiD Library Oligos试剂盒常规实验步骤进行高通量测序。

实施例

下面通过实施例对本发明的具体实施方式进行阐述，所述的实施方式仅仅用来解释和说明本发明，其并不限制本发明的保护范围。任何本领域技术人员根据公知的知识和现有技术的教导能够想到的等价的变体都包含在本发明的保护范围中。

实施例1

实施例中接头、qPCR检测引物与磺酸化处理后DNA的PCR扩增引物为人工合成序列由Invitrogen公司合成；所使用的COT1 DNA购买于Invitrogen公司，qPCR检测使用AB公司的SYBR与配套试剂；EcoP15I

酶购买于NEB公司。实施例的操作原理如图5所示。

1.获取基因组DNA

取血样1（取自志愿者）10毫升，用QIAamp DNA Blood Mini Kit（Qiagen公司）提取样品DNA。将提取得到的DNA编号为YH-1，取其中10微克DNA样品作为起始材料，参照图1的流程构建文库，本实施例中构建的是Illumina GA System Paired End文库。

2. 片段化基因组DNA

采用Covaris system（AB公司）将前述步骤1中的DNA样品进行断裂，样品断裂结束后，取约总量的1/30（体积比）的断裂后的样品在2%琼脂糖凝胶上进行1×TAE电泳检测，从所述电泳凝胶回收待测DNA片段。为将样品DNA片段控制在100~500 bp范围内，Covaris system断裂具体操作为:

1)双击打开Covaris主程序“SonoLAB S-Series V2.54”后点击“Start [Enter]”，待仪器排气30分钟，且水温降为10℃左右后方可使用。打开主程序之前须先确定打碎仪和冷却仪的电源开关是否打开，否则主程序会提示错误并重试。

2）将样品加入待用的100ul的covaris micro Tube中，样品加入量5μg，最后将终体积用TE补至100 μl，用移液器将样品溶液吹打混匀。 点击“Configure”，设置参数:

处理1（Treatment 1）	任务/循环(Duty/cycle) (%)	10
	强度(Intensity)	5
	循环(Cycle)/冲击度(burst)	200
	时间(Time)秒(s)	960
循环(Cycle)		16

选择模式为“Frequency Sweeping”。所有设置完成后点击“Save”或“Save As...”保存好程序，再点击“Return to Main Panel”回到主界面。将断裂管放入Covaris装置中，选择已保存程序，开始打碎。按所设置的程序打碎完成后关闭打碎仪和制冷

仪的电源，然后关闭打碎主程序和电脑。

3)将前述 2)中断裂后的样品从玻璃管中吸出，放入 1.5ml EP 管中。取出 3 μl 的样品用于 2%琼脂糖凝胶进行 1×TAE 电泳检测，并用 QIAquick PCR 纯化试剂盒进行回收，产物溶于 32 μl 的洗脱缓冲液(Elution buffer, EB)中。

3.末端修复

先从-20℃保存的试剂盒（Illumina Paired End 测序文库构建试剂盒)中取出 10×多聚核苷酸激酶缓冲液和 10mM dNTPs mix，将其置于冰上融解并充分混匀 10×多聚核苷酸激酶缓冲液。在 1.5 mL 的离心管中配制 100μL 末端修复反应体系：取前述步骤 2 最后获得的 30 μL 的片段化回收产物、45uL 超纯水、10 μL 10×多聚核苷酸激酶缓冲液（B904）、4 μL dNTP solution set（稀释混合为 10mM each）、5 μL T4 DNA 聚合酶、1 μL Klenow Fragment 与 5 μL T4 多聚核苷酸激酶。20℃温浴 30 分钟后，用 QIAquick PCR 纯化试剂盒回收纯化经过末端修复的 DNA，将产物 DNA 溶于 34μL 的 EB 中。

4.添加 A 碱基

先从-20℃保存的试剂盒（Illumina Paired End 测序文库构建试剂盒）中取出 10x blue buffer 和 1 mM dATP，将其置于冰上融解并使其充分混匀。在 1.5 mL 的离心管中配制 50μL 的 DNA 末端加“A”反应的体系：32μL 前述步骤 3 获得的末端修复纯化回收产物、5 μL 10x blue buffer、10 μL dATP(稀释为 1mM，GE 公司)与 3 μL Klenow (3’-5’ exo-)。37℃温浴 30 分钟后用 QIAquick PCR 纯化试剂盒回收纯化加 A 体系中的 DNA，将产物溶于 32μL 的 EB 中。取 1μL 于 NanoDrop 1000 上测 OD 值并记录样品浓度、OD260/280 比值以及 OD260/230 比值等参数。

5.辅助接头 1 的连接

从-20℃保存的试剂盒（Illumina Paired End 测序文库构建试

剂盒）中取出 2×Rapid 连接缓冲液和Alu Linker，将其置于冰上融解并充分混匀 2×Rapid 连接缓冲液。在 1.5 mL的离心管中配制 100μL的连接反应体系：30μL步骤 3 得到的加A回收产物、50 μL 2x Rapid连接缓冲液、6μL（辅助接头）(50uM)（5'-AGCTGGGCACCGCTCATGCCACTCCGGCT，5'-pGCCGGAGTGGCATGAGCGGTGCCCAG）、10μLT4 DNA连接酶与 4μL超纯水。20℃温浴 15 分钟后用ZYMO DNA Clean & ConcentratorPTMP-5（ZYMO公司）回收纯化加有辅助接头 1 的DNA，产物溶于 40 μL的TE中。

6.甲基化 DNA 的免疫共沉淀（MeDIP）

(1)取来自步骤 5 的样品，参考步骤 4 NanoDrop 1000 测定结果，此时 DNA 为 4μg 左右，加 TE 补至体积为 450μL，充分混匀。在 98℃的水浴锅中处理 10 分钟，然后置于冰上 10 分钟。

(2)在样品中加入 51μL 的 10×IP 缓冲液和 10μL Anti- 5 - Methylcytosine Mouse mAb(Abcam公司)，充分混匀；将样品竖直放置在 rotator 上，4℃孵育 12 小时。

(3)取 30μL Dynabeads M-280 山羊抗小鼠 IgG(Invitrogen公司)，用 800μL 0.1% PBS-BSA 将磁珠洗涤 2 次，加入 30μL 1×IP 缓冲液混匀。

(4)将 Dynabeads 加到 DNA-Antibody 混合物中，4℃震荡混合 3 小时。

(5)用 800μl 1×IP 缓冲液将磁珠洗涤 3 次，第一次洗时在 vertex 上涡旋 2-4 秒，涡旋 3 次；后两次洗涤用移液器反复吹打混合。用磁力架收集后将上清移除。

(6)加入 200μL 蛋白水解酶 K 消化缓冲液和 3μL 蛋白水解酶 K（50μg/μl），50℃烘箱中旋转 3 小时。

(7)用ZYMO DNA Clean & ConcentratorPTMP-5 回收步骤(6)这一反应体系中的DNA，用 7 倍DNA溶液体积（步骤 6）的结合缓冲液（binding buffer），溶于 20μL经 60℃水浴过的超纯水中。1μL于NanoDrop 1000

上测得OD值并记录样品浓度，并Q-PCR检测MeDIP效果。

7.COT1 DNA去除重复序列

(1)重复序列的生物素标记：在 1.5mL 管中加入 4μL COT 1 DNA（100ng/μL）(Invitrogen 公司)、8μL 随机引物(8N, 1ug/μL)(Invitrogen)与25μL 超纯水后97℃处理10分钟并置于冰上10分钟。然后继续加入 5μL 10×Klenow 缓冲液、5μL 生物素/dNTP mix（Biotin-16-dUTP: 0.35mM; dTTP: 0.65mM; dCTP: 1mM; dGTP: 1mM; dATP: 1mM.）与 3μL Klenow Enzyme(exo-, Fermentas)使最终反应体积为50μL。37℃处理过夜，第二天继续反应并补加1μL Klenow Enzyme，继续反应 3-5 小时。反应结束用 QIAquick PCR 纯化试剂盒回收纯化反应体系中的DNA，产物溶于100μL 的EB中并用 Nano-drop1000 检测浓度。若总质量小于2μg，可认为标记失败，需重新标记。

(2)吸取 100μL of Streptavidin-Dynabeads® M-280(Invitrogen)到1.5mL不粘管,用800μL TE 缓冲液或者1×B&W 缓冲液(5mM Tris-HCL pH 7.5, 0.5mM EDTA, 1M NaCl)洗磁珠两次。

(3)将磁珠重新悬浮于100μL 2×B&W缓冲液(10mM Tris-HCL pH 7.5, 1mM EDTA, 2M NaCl)。

(4)将用生物素标记好的COT-1 DNA（预先变性，97℃处理10分钟，后置于冰上10分钟）。 加入 悬浮磁珠 中（步骤（3））；

(5)常温低速旋转1小时;

(6)移除上清，1×B&W 缓冲液轻洗两次。

(7)抽干MeDIP后的DNA样品，加入65℃预热的杂交液100μL，移至含有磁珠的不粘管，65℃翻转过夜。

(8)将磁珠上清移至新的 1.5mL Eppendorf 管，用 ZYMO-5 纯化，结合缓冲液量为7倍，22μL 洗脱。1μL 于NanoDrop 1000上测得OD值并记录样品浓度。

8.亚硫酸氢盐处理（Bisulfite treatment）

本实验使用 ZYMO EZ DNA Methylation-Gold Kit PTMP（ZYMO 公司）进行亚硫酸氢盐处理

(1)在 PCR 管中添加 130 μL 的 CT 翻转试剂到每 20μL DNA 样品中。如果 DNA 样品的体积小于 20μL， 则用水来弥补差量。通过轻弹试管或移液器操作来混合样品。

(2)将样品管放到循环变温器并按以下步骤操作:98℃放置 10 分钟，64℃放置 2.5 小时后立刻进行下述操作或者在 4℃下存储(最多 20 小时)。

(3)吸取 600μL 的 M 结合缓冲液到 Zymo-Spin ICPTMP 柱中，并将柱放如试剂盒所提供的收集管中。

(4)装填样品 (从步骤 2)到 Zymo-Spin ICPTMP 柱含有 M 结合缓冲液。 盖上盖子将柱颠倒数次来混合样品。

(5)全速 (>10,000×g)离心 30 秒，去除流出液。

(6)吸取 200μL 的 M 洗涤缓冲液到柱中，全速离心 30 秒。

(7)吸取 200 μL 的 M-Desulphonation 缓冲液到柱中并且在室温(20℃-30℃)下放置 15-20 分钟。在培养后，全速离心 30 秒。

(8)吸取 200μL 的 M 洗涤缓冲液到柱中，全速离心 30 秒。再添加 200 μL 的 M 洗涤缓冲液并且离心 30 秒。

(9)吸取 20μL 的 M 洗脱缓冲液到柱基质中,将柱放置在 1.5 mL 的管中，全速离心来洗脱 DNA。

9. 亚硫酸氢盐磺酸化处理后 DNA 的 PCR 扩增

(1)从-20℃保存的试剂盒中取出，将其置于冰上化冻并充分混匀 JumpStartPTMP Taq DNA 聚合酶。在 0.2 mL 的 PCR 管中配制 50μL 的 PCR 反应体系: 5μL 亚硫酸氢盐处理纯化后产物、5μL 10×PCR 缓冲液(Sigma)、4μL dNTP、0.5μL JumpStartPTMP Taq DNA 聚合酶(Sigma)、

2μL辅助接头引物1.0 (10 pmol/μL)(Invitrogen)、 2μL辅助接头引物(10 pmol/μL) (Biotin-5' -GGTCAGCAGCTAAACACCACTCATACCACTCCA,Biotin- 5' - GGTCAGCAGTTGGGTATTGTTTATGTTATTTTGGT)(Invitrogen)与35.5μL超纯水。

(2)在热循环仪中运行下列扩增程序: 94℃ 10秒、10个循环的 94℃ 30秒/ 52℃30秒/72℃30秒、72℃5分钟于4℃存储。

(3)反应结束后用 QIAquick PCR 纯化试剂盒回收纯化反应体系中的DNA, 产物溶于50μL的EB中。取1μL于NanoDrop 1000上测得OD值并记录样品浓度。

10.酶切去除辅助接头1

(1)在 1.5 mL 的离心管中配制 100μL 反应体系: PCR 扩增产物 (~350ng, 不超过400ng), 10μL 10×缓冲液 3(NEB), 1μL 100×BSA, 2μL sinefungine(10mM), 20μL 10×ATP, 3μL EcoP15I(10U/μL), 补超纯水至100μL;

(2)37℃酶切过夜,第二天补1μL酶 继续酶切2小时。ZYMO-5 纯化, 结合缓冲液用量为5倍, 50μL洗脱;

(3)吸取50μL of Streptavidin-DynabeadsP®P M-280 (Invitrogen)到1.5mL 不粘管, 用800μL TE缓冲液或1×B&W 缓冲液洗磁珠两次;

(4)将磁珠重悬于50μL 2×B&W缓冲液,加50μL酶切产物(step 2);

(5)常温旋转 30分钟 (转速小不超过600rpm);

(6)将磁珠上清移至新的1.5mL Eppendorf 管中,用 Qiagen(Mini) 纯化, 取1μL纯化产物于NanoDrop 1000上测得OD值并记录样品浓度。

11. 测序

将步骤10中去除辅助接头1后的DNA片段直接用于后续 Illumina 公司 Genome Analyzer (GA)系统, 进行高通量测序, 其采用的实验步骤参见使用 Paired End DNA 文库制备试剂盒构建常规 Paired End DNA

测序文库。

实施例 1 的结果评价和分析

1. MeDIP 富集效果的检测

对本发明使用的 MeDIP 方法的富集效果与商业 MeDIP 试剂盒(Diagenode)的富集效果进行比较(表 1),其中甲基化富集效果(4994)为 62.71%,大于 Diagenode 的甲基化富集效果,非甲基化的富集率(8804)为 0.33%,小于 Diagenode 的非甲基化富集率。

因此,本 MeDIP 方法的回收率达到商业 MeDIP 试剂盒(Diagenode)的回收率要求范围,并相对商业 MeDIP 试剂盒富集量大,因此认为该 MeDIP 方法可与商用试剂盒相媲美,其试验结果平行可信。

表 1. YH-1 DNA 样品的 MeDIP 富集效率

引物		Diagenode	YH-adapter-U
4994	回收率 (%)	43.65	62.71
8804		0.76	0.33

2. 对 EcoP15I 酶切效果的检测

对 EcoP15I 酶切产物分别进行挑克隆后利用 Sanger 法测序,检测结果表明(表 2),EcoP15I 的酶切效率达 86%。该酶的高酶切效率结果表明该 Linker 可被高效切除,对后续接头连接等实验没有干扰。

表 2. EcoP15*I* 酶切效率检测

克隆	EcoP15*I*
酶切	24
未酶切	4
总量	28
酶切率 (%)	86

3. Sanger 法测序库检

对用本次发明的测序库样本制备技术所制备的测序库进行克隆处理并使用 Sanger 法进行小范围的测序检测。库检结果(表 3)中有 76%

的 reads 能比对回基因组区，且其甲基化化率高达 87%；亚硫酸盐（磺酸化）对非甲基化胞嘧啶的翻转率为 99%；片段扩增重复率为零。

表 3：YH-1 基因组甲基化免疫共沉淀后重复序列去除结合磺酸化处理文库 Sanger 法测序结果

文库	测序数量	比对数量	比对比率	甲基化率	转换率	扩增重复率	GC 碱基含量
YH-1	71	54	0.76	0.87	0.99	0	27.81%

实施例 2

实施例 2 的原理如图 6 所示，其中除了下述步骤与实施例一不同外，所有步骤与实施例 1 相同，用下述步骤 5、9 和 10 分别替代实施例一中的步骤 5、9 和 10：

5. 辅助接头的连接

从-20℃保存的试剂盒中取出 2x Rapid连接缓冲液和Alu Linker，将其置于冰上融解并充分混匀 2x Rapid连接缓冲液。在 1.5 mL的离心管中配制 100μL的连接反应体系：30μL加A回收产物、50 μL 2x Rapid 连接缓冲液、6μL 辅助接头(50uM) (5'- CTGGGCACCGCTCATGCCACTCCGGC TAAG^{5m}CT, 5'-pG^{5m}C TTAGCCGGAGTGGCATGAGCGGTGCCCAG)、10μLT4 DNA连接酶与 4μL超纯水。20℃温浴 15 分钟后用ZYMO DNA Clean & ConcentratorPTMP-5 回收纯化连接有辅助接头中的DNA，产物溶于 40 μL 的TE中。

9. 亚硫酸氢盐处理后 DNA 的 PCR 扩增（AluI linker 介导的 PCR 反应）

(1)从-20℃保存的试剂盒中取出，将其置于冰上化冻并充分混匀 JumpStartPTMP Taq DNA聚合酶。在 0.2 mL的PCR管中配制 50μL 的PCR 反应体系：5μL 亚硫酸氢盐处理纯化后产物、5μL 10×PCR buffer(Sigma)、4μL dNTP、0.5μL JumpStartPTMP Taq DNA 聚合酶（Sigma）、2μL Alu I 引物 1.0 (10 pmol/μL)(Invitrogen)、 2μ

L 辅助接头引物(10 pmol/μL) (Biotin-5' -CTAAACACCACTCATACCACT CCA, Biotin- 5' -TTGGGTATTGTTTATGTTATTTTGGT) (Invitrogen)与 35.5 μL超纯水。

(2)在热循环仪中运行下列扩增程序:94℃ 10秒、10个循环的 94℃ 30秒/ 52℃30秒/72℃30秒、72℃5分钟与 4℃存储。

(3)反应结束后用 QIAquick PCR 纯化试剂盒回收纯化反应体系中的 DNA, 产物溶于 50μL 的 EB 中。取 1μL 于 NanoDrop 1000 上测得 OD 值并记录样品浓度。

10.酶切去除辅助接头

(1)在 1.5 mL 的离心管中配制 100μL 反应体系: PCR 扩增产物 (<400ng)、10μL 10×缓冲液 2(NEB)、3μL Alu I, 补超纯水至 100 μL;

(2) 37℃酶切过夜, 第二天补 1μL 酶 继续酶切 2 小时。ZYMO-5 纯化, 结合缓冲液用量为 5 倍, 50μL 洗脱;

(3)吸取 100μL 的Streptavidin - Dynabeads P ® P M - 280 (Invitrogen) 到 1.5mL不粘管, 用 800μL TE缓冲液或者 1×B&W缓冲液洗磁珠两次;

(4)将磁珠重悬于 50μL 2×B&W 缓冲液。加入 50μL(2)的酶切产物;

(5)常温旋转 30 分钟 (转速小不超过 600rpm);

(6)将磁珠上清移至新的 1.5mL Eppendorf 管, Qiagen (Mini) 纯化, 取 1μL 于 NanoDrop 1000 上测得 OD 值并记录样品浓度。

实施例 2 的结果评价和分析

1. 辅助接头效率检测

由于辅助接头中含有一个甲基化胞嘧啶, 因此首先检测该甲基化位点是否对 MeDIP 有影响。对比辅助接头, 设计和其对应的非甲基化接头。使用甲基化检测引物编号为 4994 与 GAPDH 两对、非甲基化检测引物编

号为 8804 与 TSH2B 两对分别在 MeDIP 后对甲基化与非甲基化的辅助接头含量进去检测。检测结果（表 4）所示，甲基化的辅助接头与非甲基化对应的富集率一致。因此对甲基化的辅助接头对 MeDIP 的本底影响可忽略。

表 4. 甲基化与非甲基化辅助接头对 MeDIP 的影响验证

引物	辅助接头-U 回收率 (%)	辅助接头-M 回收率 (%)
4994	3.560	3.560
8804	0.023	0.028
TSH2B	0.014	0.015
GAPDH	13.810	11.390

2. MeDIP 富集效果的检测

分别利用检测甲基化富集效果的引物（4994）和非甲基化富集效果的引物（8804）对 MeDIP 富集效果进行检测（表 5），其中甲基化富集效果为 45.6%，和 Diagenode 的甲基化富集效果（表 2）相差无几，非甲基化的富集率为 2.9%，尽量略高于 Diagenode 的非甲基化富集率，但远远低于甲基化的富集效果。因此，本次 MeDIP 富集 DNA 效率成功。

表 5. YH-1 DNA 样品的 MeDIP 富集效率

引物		YH-adapter-M
4994	回收率 (%)	45.6
8804		2.9

3. 对 AluI 酶切效果的检测

对 AluI 酶酶切产物分别进行挑克隆后利用 Sanger 法测序，检测结果表明（表 6），两次 AluI 酶的酶切效率分别为 94%和 100%。该酶切效率结果表明该 Linker 可被高效切除，对后续接头连接等实验没有干扰。

表 6. AluI 的酶切效率检测

克隆	AluI(1st)	AluI(2nd)
酶切	30	79
未酶切	2	0
总量	32	79
酶切率（%）	94	100

4. Sanger 法测序库检

对用本次发明的测序库样本制备技术所制备的测序库进行克隆处理并使用 Sanger 法进行小范围的测序检测。库检结果（表 7）中有 85% 的 reads 能比对回基因组区，且其甲基化化率高达 93%; 说明连接 Alu I 接头的 DNA 片段经过 MeDIP 处理能富集到基因组中大部分甲基化区域的序列。另外，高达 97%的转化率说明亚硫酸盐（磺酸化）处理的效果也是非常明显的。因此，通过 Alu I 接头连接的 DNA 片段经过 MeDIP 处理、重复序列去除、亚硫酸盐处理与 Alu I 酶切后仍可得到有效的序列信息。文库检测合格并用于构建 Illumina 公司 GA 系统高通量测序的 Paired End DNA 文库与进行相应的测序。

表 7: YH-1 基因组甲基化免疫共沉淀后重复序列去除结合磺酸化处理文库 Sanger 法测序结果

文库	测序数量	比对数量	比对比率	甲基化率	转换率	扩增重复率	GC 碱基含量
YH-1	75	64	0.85	0.93	0.97	0	30.64%

发明人仅仅在实施例 1 和 2 是仅用了平末端的辅助接头连接和去除方式（1）和（2）进行了说明。具有一定的代表性。由于平末端的辅助接头连接和去除方式（3）是平末端的辅助接头连接和去除方式（1）和（2）的综合，因此，平末端的辅助接头连接和去除方式（1）和（2）实施的成功可以说明（3）也可以成功。平末端相对粘性末端的原理在于连接接头的方式的不同，而平末端相对粘性末端连接效率低，因此，在平末端实施成功的情况下粘性末端的辅助接头连接和去除方式(4-6)

定能实施成功。

5. 与亚硫酸氢盐测序法的有效数据比较

在 Illumina 公司 GA 系统进行高通量测序，所述的测序是 45cycle Paired End 测序。具体测序流程与试剂均采用 Illumina 公司标准流程与试剂盒。测序结果产生 2.53Gb 数据，能比对（mapping）到参考基因组（homo sapienss）达一半，和其他亚硫酸氢盐测序比对率近似[H. Xiang 等人， Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nat Biotechnol 28 516-20]，具体结果可以参见表 8。其中 DNA 平均甲基化水平达 84.2%，全基因组甲基化测序的约 70%甲基化水平相比说明 MeDIP 结合甲基化胞嘧啶的特异性较高。

表 8: YH-1 基因组甲基化免疫共沉淀后重复序列去除结合磺酸化处理文库 GA 系统测序结果

文库	YH-1
总数据	57,510,098（2.53Gb）
过滤后总数据	57,049,532
过滤率	0.80%
比对数	30,287,175
比对率	53.09%
扩增重复率	5.83%
有效数据	28,619,720（1.27Gb）
数据有效率	50.17%
转化率	97.18%
甲基化率	84.20%

6. 与 Sanger 测序法比较文库的重复序列含量

对 45cycle Paired End 测序数据的比对至基因组的序列进行高级生物信息分析，结果如表 9 所示，重复序列为 30.87%，和 Sanger 测序法检测的 MeDIP 的重复序列含量比（75%）降低了一半多。以上结果说明该 MeDIP 结合 bisulfite 方法有很高的特异性，能有效的分离 DNA 甲

基化片段和非甲基化片段。GA 测序序列比对到重复序列区比率(30.8%)，重复序列去除将近一半，降低了测序文库大小和冗余数据。

表 9 YH-1 基因组甲基化免疫共沉淀后重复序列去除结合磺酸化处理文库 GA 系统测序其他参数结果

基因特征	数据量	百分比 (%)
重复序列	2927604	30.83
转座子	2173324	22.89
CpG 岛	144614	1.52
基因元件	1678027	17.66
ncRNA	24037	
总数	9493842	

7.全基因磺酸化处理测序与本技术测序结果共同比对的单个 CpG 位点的测序深度分析

对 1 号染色体采用 MRERB 方法和全基因组磺酸化处理测序结果进行比较，针对在基因组功能区共同比对到的 CpG 位点做测序深度比较发现，二者测序深度的趋势基本一致，具体参见图 10。但是测序深度差 4 倍，因此，利用本发明的 MRERB 方法只需 10G 的数据将达到 70G 全基因组亚硫酸氢盐测序的水平，整体数据量降低了 7 倍。

因此，通过对 MRERB 文库制备技术的 Solexa 测序数据初步分析表明该方法能够从全基因组中特异性的分离出甲基化 DNA 片段，并能去除大部分重复序列；经亚硫酸氢盐翻转之后，能对 GA 系统读取的序列进行比对、定位，并可以进行单个胞嘧啶的甲基化分析，可用于不同样本的甲基化检测。

综上所述，本技术既能检测到基因组功能区DNA甲基化，又能精准比对单个胞嘧啶甲基化状态。在技术成本上，在功能区甲基化测序数据饱和的情况下可比全基因组亚硫酸氢盐测序的数据量少 80%-90%，节约每个样本检测成本 70-80%。

权 利 要 求

1. 一种甲基化 DNA 的高通量测序方法，包括建库步骤和测序步骤，其中建库步骤是指获得待检测的甲基化 DNA 文库的步骤，所述建库步骤包括:

A)基因组 DNA 的片段化及双链 DNA 片段末端修复;

B)将 A)得到的双链 DNA 与辅助接头连接,所述的辅助接头是被设计成能与修复后的双链 DNA 末端相连接的双链 DNA 序列，经由该辅助接头使经后续 E)步骤处理的单链 DNA 经 PCR 后变成双链 DNA,辅助接头与修复的双链 DNA 连接的一端为连接末端，另一端为非连接末端;

C)将 B)的产物进行甲基化免疫共沉淀;

D)去除 C)的产物中的中度，高度重复序列;

E)将 D)得到的产物进行亚硫酸盐翻转处理;

F)将 E)得到的单链 DNA 通过引物进行 PCR 扩增获得双链 DNA;

G)酶切去除辅助接头;

其中测序步骤是指对前述建库步骤获得的文库进行序列测定，包括步骤:

H)对 G)中获得的双链 DNA 进行末端修复，连接测序接头;

J)对 H)中的产物进行甲基化 DNA 测序。

2. 权利要求 1 的测序方法:

其中，步骤 B）中辅助接头选自下述 a-h 中的至少一种:

a 为不含酶切位点，非连接端为一条链突出结构，连接端为平末端;

b 为含酶切位点，非连接端为一条链突出结构，连接端为平末端;

c 为不含酶切位点，非连接端为树杈结构，连接端为平末端;

d 为含酶切位点，非连接端为树杈结构，连接端为平末端;

e 为不含酶切位点，非连接端为一条链突出结构，连接端为粘末端;

f 为含酶切位点，非连接端为一条链突出结构，连接端为粘末端；

g 为不含酶切位点，非连接端为树杈结构，连接端为粘末端；

h 为含酶切位点，非连接端为树杈结构，连接端为粘末端。

步骤 B)中的辅助接头，步骤 F)中的引物和步骤 G)中切割酶的设计和组合应用选自下述(I)-(VI)中的至少一种：

(I)当辅助接头为 a 和 c 时，引物设计成互补于经过 E)翻转的辅助接头的序列，且在 5'末端额外连接有限制性内切酶的识别位点，所述的限制性内切酶的识别位点位于引物上，并保证该酶的切割位点位于辅助接头与待测 DNA 连接点的上下游 5bp 的范围内；当进行步骤 G)时，采用针对所设计的限制性内切酶的酶切位点的酶进行单酶切；

(II)当辅助接头为 b 和 d 时，在辅助接头中靠近辅助接头与待测 DNA 连接点的上游 5bp 内位置设计酶切位点，该酶切位点需符合以下三个特点：1)该酶切位点至少包含一个甲基化胞嘧啶；2)酶切位点内不含有非甲基化的胞嘧啶；3)该酶切位点不含有 CpG 二核苷酸位点；引物设计成与经过 E)翻转的辅助接头互补的序列，当进行步骤 G)时，采用针对所设计的甲基化的酶切位点进行切割的酶进行单酶切；任选地，在所述引物 5' 端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合；当进行步骤 G)时，采用针对所设计的限制性内切酶的酶切位点的酶进行双酶切；

(III)当辅助接头为 e 和 g 时，辅助接头和引物的设计与 a、c 方案相同，不同处在于与待测 DNA 片段连接方式不同：a、c 连接方式为“T-A”连接，e 和 g 为粘性末端连接；

(IV)当辅助接头为 f 和 h 时，辅助接头和引物的设计与 b、d 方案相同，不同处在于与待测 DNA 片段连接方式不同：b、d 连接方式为“T-A”连接，f 和 h 为粘性末端连接；

(V)辅助接头设计同 b、d 接头设计，进一步在辅助接头中靠近辅助

接头与待测 DNA 连接点的上游 20-30bp 内位置设计酶切位点，该酶切位点同时满足以下四点：1）该酶切位点至少包含一个甲基化胞嘧啶；2）酶切位点内不含有非甲基化的胞嘧啶；3）该酶切位点不含有 CpG 二核苷酸位点；4）该酶的切割位点位于辅助接头与待测 DNA 连接点的上、下游 5bp 内；在引物设计成与经过 E）翻转的辅助接头互补的序列，当进行步骤 G)时，采用针对所设计的甲基化的酶切位点进行切割的酶进行单酶切；

(VI)是(V)和 b、d、f、h 的组合，通过调整辅助接头的长度使两个酶切位点重合，该方案可以使用双酶切去除辅助接头；在引物设计成与经过 E）翻转的辅助接头互补的序列，当进行步骤 G)时，采用针对所设计的甲基化的酶切位点进行切割的酶进行双酶切。

3. 权利要求 1 或 2 的测序方法，其中步骤 A)包括：

a 将基因组 DNA 片段化成为双链 DNA 片段；

b 将 a 中的双链 DNA 片段末端修复为平末端；

c 将 b 中的平末端的 3’末端加上 A 碱基。

4. 权利要求 1 或 2 的测序方法，其中步骤 C)的甲基化免疫共沉淀选自 MeDIP 和 MBD。

5. 权利要求 1 或 2 的测序方法，其中步骤 D)的重复序列去除使用 COT1 DNA。

6. 权利要求 2 的测序方法，其中步骤 F)中酶切位点甲基化是指辅助接头中设定的酶切位点里的的胞嘧啶位点是甲基化的。

7. 权利要求 1 的测序方法，其中的辅助接头设计成防止自体连接的结构，在 5’末端标记有生物素。

8. 一种执行权利要求 1 或 2 的方法的自动装置。

9. 权利要求 8 的装置，其包括

A)基因组 DNA 的片段化单元和末端修复单元；

B)将 A)得到的双链 DNA 与辅助接头连接的连接单元;

C)将 B)的产物进行甲基化免疫共沉淀的单元;

D)去除 C)的产物中的中度，高度重复序列的去除单元;

E)将 D)得到的产物进行亚硫酸盐翻转处理的单元;

F)将 E)得到的单链 DNA 用引物进行 PCR 扩增获得双链 DNA 的单元;

G)酶切去除辅助接头的单元;

H)对 G)中获得的双链 DNA 进行末端修复的单元

J)对 H)中修复的 DNA 进行测序接头连接的单元;

K)对 J)中的产物进行甲基化 DNA 测序的测序单元。

经修改的权利要求
国际局收到日：2011年7月22日(22.07.2011)

1. 一种建立甲基化DNA文库的方法，包括以下步骤：

A)样品双链DNA的片段化；

B)将A)得到的片段化DNA两端与可剪切的辅助接头连接；

C)将B)的产物进行甲基化免疫共沉淀，得到单链DNA片段；

D)去除C)得到的单链DNA片段中的含中度、高度重复序列的片段；

E)将D)得到的产物中未甲基化修饰的胞嘧啶碱基转化为尿嘧啶碱基；

F)将E)得到的单链DNA通过引物进行PCR扩增获得双链DNA；

G)去除可剪切的辅助接头。

2.一种甲基化DNA的测序方法，包括以下步骤：

A)样品双链DNA的片段化；

B)将A)得到的双链DNA的两端分别与可剪切的辅助接头连接；

C)将B)的产物进行甲基化免疫共沉淀，得到单链DNA；

D)去除C)的产物中的中度，高度重复序列；

E)将D)得到的产物中未甲基化修饰的胞嘧啶碱基转化为尿嘧啶碱基；

F)将E)得到的单链DNA通过引物进行PCR扩增获得双链DNA；

G)去除可剪切的辅助接头；

H)对G)中获得的双链DNA进行测序；

任选地，还包括根据H）步骤的测序结果与样品DNA的序列或参考序列进行比对，确认甲基化胞嘧啶碱基的个数和位置信息。

3. 权利要求 1 或 2 的方法，其中步骤 E)是将产物进行重亚硫酸盐翻转处理。

4、权利要求 1 或 2 的方法，其中辅助接头的可剪切通过在辅助接头和/或引物中设计酶切识别位点来实现。

5. 权利要求 4 的方法，

其中，步骤 B）中辅助接头与双链 DNA 连接的一端为连接末端，另一端为非连接末端，选自下述 a-h 中的至少一种:

a 为不含酶切位点，非连接端为一条链突出结构，连接端为平末端;

b 为含酶切位点，非连接端为一条链突出结构，连接端为平末端;

c 为不含酶切位点，非连接端为树杈结构，连接端为平末端;

d 为含酶切位点，非连接端为树杈结构，连接端为平末端;

e 为不含酶切位点，非连接端为一条链突出结构，连接端为粘末端;

f 为含酶切位点，非连接端为一条链突出结构，连接端为粘末端;

g 为不含酶切位点，非连接端为树杈结构，连接端为粘末端;

h 为含酶切位点，非连接端为树杈结构，连接端为粘末端;

步骤 B)中的辅助接头，步骤 F)中的引物和步骤 G)中切割酶的设计和组合应用选自下述(I)-(VI)中的至少一种:

(I)当辅助接头为 a 和 c 时，引物设计成互补于经过 E)翻转的辅助接头的序列，且在 5'末端额外连接有限制性内切酶的识别位点，所述的限制性内切酶的识别位点位于引物上，并保证该酶的切割位点位于辅助接头与待测 DNA 连接点的上下游 5bp 的范围内；当进行步骤 G)时，采用针对所设计的限制性内切酶的酶切位点的酶进行单酶切;

(II)当辅助接头为 b 和 d 时，在辅助接头中靠近辅助接头与待测 DNA 连接点的上游 5bp 内位置设计酶切位点，该酶切位点需符合以下三个特点：1)该酶切位点至少包含一个甲基化胞嘧啶；2)酶切位点内不含有非甲基化的胞嘧啶；3)该酶切位点不含有 CpG 二核苷酸位点；引物设计成

与经过 E)翻转的辅助接头互补的序列，当进行步骤 G)时，采用针对所设计的甲基化的酶切位点进行切割的酶进行单酶切；任选地，在所述引物 5’ 端设计另外一个酶切位点，通过调整辅助接头的长度使两个酶切位点重合；当进行步骤 G)时，采用针对所设计的限制性内切酶的酶切位点的酶进行双酶切；

(III)当辅助接头为 e 和 g 时，辅助接头和引物的设计与 a、c 方案相同，不同处在于与待测 DNA 片段连接方式不同：a、c 连接方式为“T-A”连接，e 和 g 为粘性末端连接；

(IV)当辅助接头为 f 和 h 时，辅助接头和引物的设计与 b、d 方案相同，不同处在于与待测 DNA 片段连接方式不同：b、d 连接方式为“T-A”连接，f 和 h 为粘性末端连接；

(V)辅助接头设计同 b、d 接头设计，进一步在辅助接头中靠近辅助接头与待测 DNA 连接点的上游 20-30bp 内位置设计酶切位点，该酶切位点同时满足以下四点：1）该酶切位点至少包含一个甲基化胞嘧啶；2）酶切位点内不含有非甲基化的胞嘧啶；3）该酶切位点不含有 CpG 二核苷酸位点；4）该酶的切割位点位于辅助接头与待测 DNA 连接点的上、下游 5bp 内；在引物设计成与经过 E）翻转的辅助接头互补的序列，当进行步骤 G)时，采用针对所设计的甲基化的酶切位点进行切割的酶进行单酶切；

(VI)是(V)和 b、d、f、h 的组合，通过调整辅助接头的长度使两个酶切位点重合，该方案可以使用双酶切去除辅助接头；在引物设计成与经过 E）翻转的辅助接头互补的序列，当进行步骤 G)时，采用针对所设计的甲基化的酶切位点进行切割的酶进行双酶切。

6. 权利要求 1 或 2 的方法，其中步骤 C)的甲基化免疫共沉淀选自 MeDIP 和 MBD。

7. 权利要求 1 或 2 的方法，其中步骤 D)的重复序列去除使用 COT1 DNA。

8. 权利要求 5 的方法，其中步骤 F)中酶切位点甲基化是指辅助接头中设定的酶切位点里的的胞嘧啶位点是甲基化的。

9. 权利要求 1 或 2 的方法，其中的辅助接头设计成防止自体连接的结构，在 5’末端标记有生物素。

10. 一种甲基化 DNA 文库，由权利要求 1 的方法构建得到。

11. 一种试剂盒，其中至少包含用于 DNA 末端修复的试剂，辅助接头试剂、甲基化免疫共沉淀试剂，用于甲基化胞嘧啶转化为尿嘧啶的试剂，用于 PCR 扩增的试剂和辅助接头剪切试剂；

辅助接头试剂包括部分或全部互补的正反向寡聚核苷酸，经退火可形成双链的辅助接头； 所述的辅助接头是被设计成能与的双链 DNA 末端相连接的双链 DNA 序列，辅助接头与双链 DNA 连接的一端为连接末端，另一端为非连接末端，

所述 PCR 扩增试剂包括引物；

辅助接头的序列设计配合引物的序列设计使得辅助接头是可剪切的。

12、权利要求 11 的试剂盒，其中：

所述辅助接头与双链 DNA 连接的一端为连接末端，另一端为非连接末端；

所述剪切试剂为限制性内切酶；

所述辅助接头选自下述 a-h 中的至少一种的辅助接头:

a 为不含酶切位点,非连接端为一条链突出结构,连接端为平末端;

b 为含酶切位点,非连接端为一条链突出结构,连接端为平末端;

c 为不含酶切位点,非连接端为树杈结构,连接端为平末端;

d 为含酶切位点,非连接端为树杈结构,连接端为平末端;

e 为不含酶切位点,非连接端为一条链突出结构,连接端为粘末端;

f 为含酶切位点,非连接端为一条链突出结构,连接端为粘末端;

g 为不含酶切位点,非连接端为树杈结构,连接端为粘末端;

h 为含酶切位点,非连接端为树杈结构,连接端为粘末端;

所述 PCR 扩增试剂引物选自下述(I)-(VI)中的至少一种:

(I)当辅助接头为 a 和 c 时,PCR 扩增试剂中的引物设计成互补于经过重亚硫酸盐翻转的辅助接头的序列,且在 5'末端额外连接有限制性内切酶的识别位点,所述的限制性内切酶的识别位点位于引物上,并保证该酶的切割位点位于辅助接头与待测 DNA 连接点的上下游 5bp 的范围内;

(II)当辅助接头为 b 和 d 时,在辅助接头中靠近辅助接头与待测 DNA 连接点的上游 5bp 内位置设计酶切位点,该酶切位点需符合以下三个特点:1)该酶切位点至少包含一个甲基化胞嘧啶;2)酶切位点内不含有非甲基化的胞嘧啶;3)该酶切位点不含有 CpG 二核苷酸位点;引物设计成与经过重亚硫酸盐翻转的辅助接头互补的序列;

(III)当辅助接头为 e 和 g 时,辅助接头和引物的设计与 a、c 方案相同,不同处在于与待测 DNA 片段连接方式不同:a、c 连接方式为"T-A"连接,e 和 g 为粘性末端连接;

(IV)当辅助接头为 f 和 h 时,辅助接头和引物的设计与 b、d 方案相同,不同处在于与待测 DNA 片段连接方式不同:b、d 连接方式为"T-A"连接,f 和 h 为粘性末端连接;

(V)辅助接头设计同 b、d 接头设计,进一步在辅助接头中靠近辅助

接头与待测 DNA 连接点的上游 20-30bp 内位置设计酶切位点，该酶切位点同时满足以下四点：1）该酶切位点至少包含一个甲基化胞嘧啶；2）酶切位点内不含有非甲基化的胞嘧啶；3）该酶切位点不含有 CpG 二核苷酸位点；4）该酶的切割位点位于辅助接头与待测 DNA 连接点的上、下游 5bp 内；在引物设计成与经过重亚硫酸盐翻转的辅助接头互补的序列；

(VI)是(V)和 b、d、f、h 的组合，通过调整辅助接头的长度使两个酶切位点重合，该方案可以使用双酶切去除辅助接头;在引物设计成与经过重亚硫酸盐翻转的辅助接头互补的序列。

13. 权利要求 11 的试剂盒，其中所述甲基化免疫共沉淀试剂至少包含用于识别甲基化胞嘧啶的单克隆抗体。

14. 权利要求 11 的试剂盒，其中所述用于甲基化胞嘧啶转化为尿嘧啶的试剂为重亚硫酸盐。

15. 权利要求 11 的试剂盒用于建立甲基化 DNA 文库的或者用于甲基化 DNA 测序的用途。

16. 一种执行权利要求 2 的方法的自动装置，其包括:

A)样品双链 DNA 的片段化单元;

B)将 A)得到的双链 DNA 的两端与辅助接头连接的连接单元;

C)将 B)的产物进行甲基化免疫共沉淀的单元;

D)去除 C)的产物中的中度，高度重复序列的去除单元;

E)将 D)得到的产物中的未甲基化胞嘧啶转化为尿嘧啶的单元;

F)将 E)得到的单链 DNA 用引物进行 PCR 扩增获得双链 DNA 的单元;

G)去除可剪切辅助接头的单元;

H)对 G)中获得的双链 DNA 测序的测序单元。

按条约第 19 条（1）所作的声明

申请人现在将原权利要求 1 分成新的权利要求 1 和 2，同时合并了原权利要求 8 和 9，此外新增加了有关试剂盒的权利要求 11－15，所做修改没有超出原始权利要求书和说明书记载的范围。

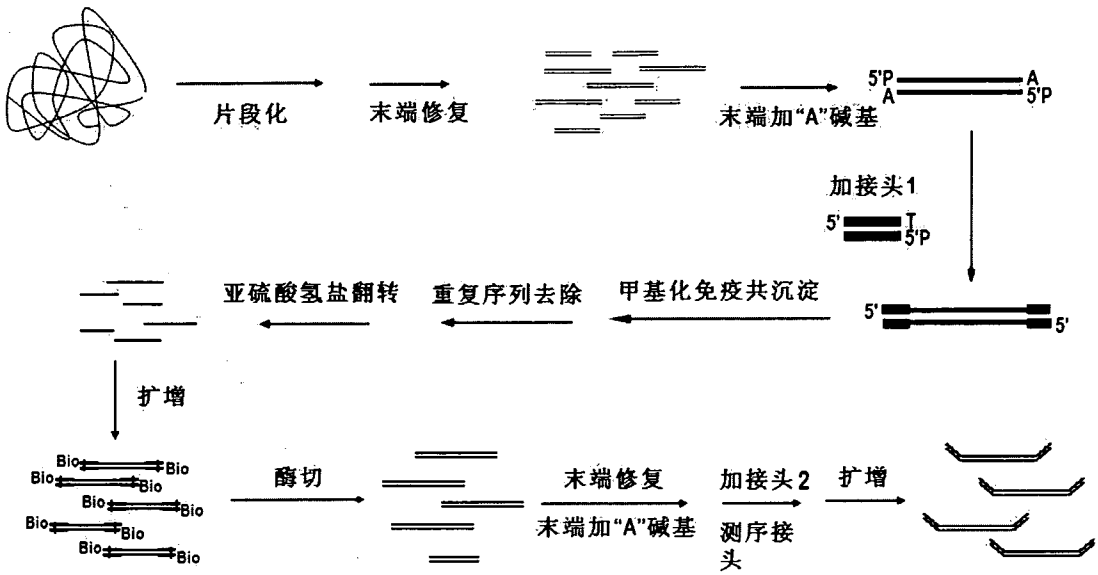

片段化
末端修复
末端加"A"碱基
5'P
A
A
5'P
加接头1
5'
T
5'P
甲基化免疫共沉淀
5'
5'
重复序列去除
亚硫酸氢盐翻转
扩增
Bio
酶切
末端修复
末端加"A"碱基
加接头2
测序接头
扩增

Fig.1

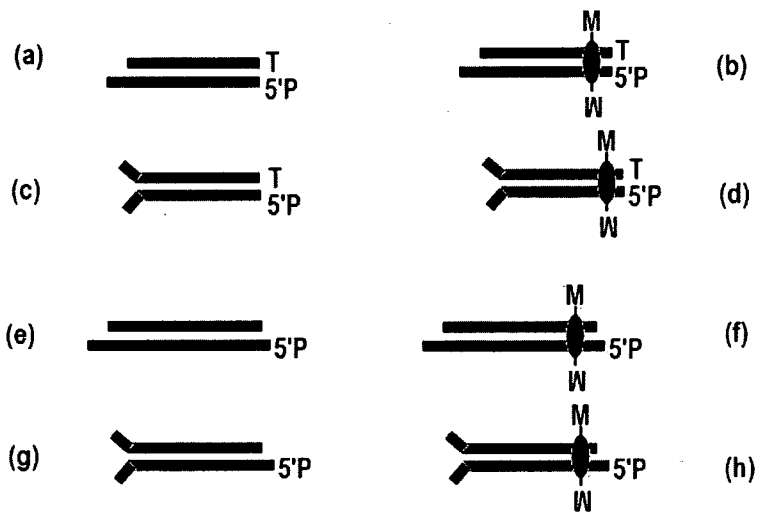

(a)
T
5'P
(b)
M
T
5'P
W
(c)
T
5'P
(d)
M
T
5'P
W
(e)
5'P
(f)
M
5'P
W
(g)
5'P
(h)
M
5'P
W

Fig.2

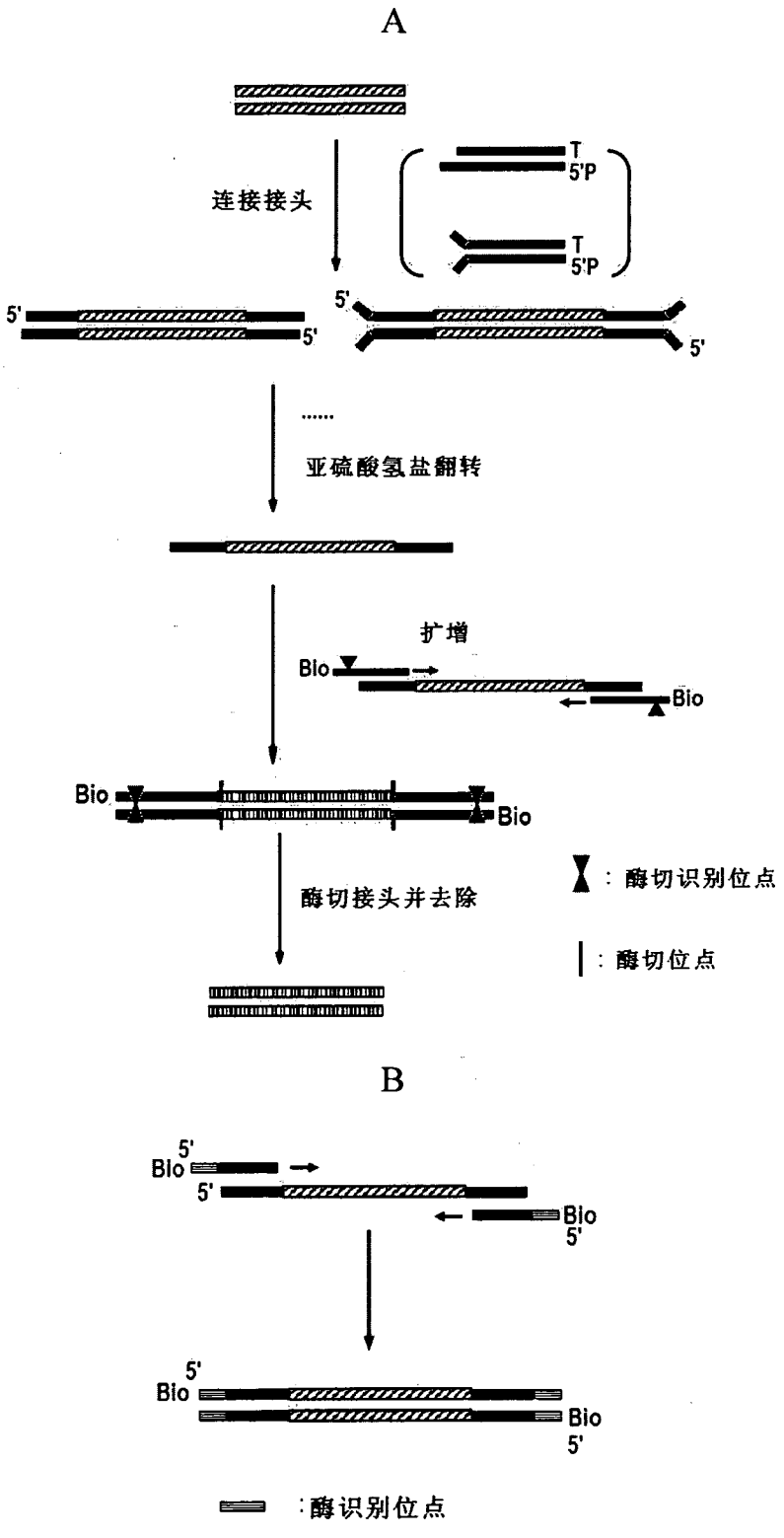

A
连接接头
T
5'P
T
5'P
5'
5'
5'
5'
.....
亚硫酸氢盐翻转
扩增
Bio
Bio
Bio
Bio
酶切接头并去除
：酶切识别位点
：酶切位点
B
5'
Bio
5'
Bio
5'
5'
Bio
Bio
5'
：酶识别位点

Fig.3

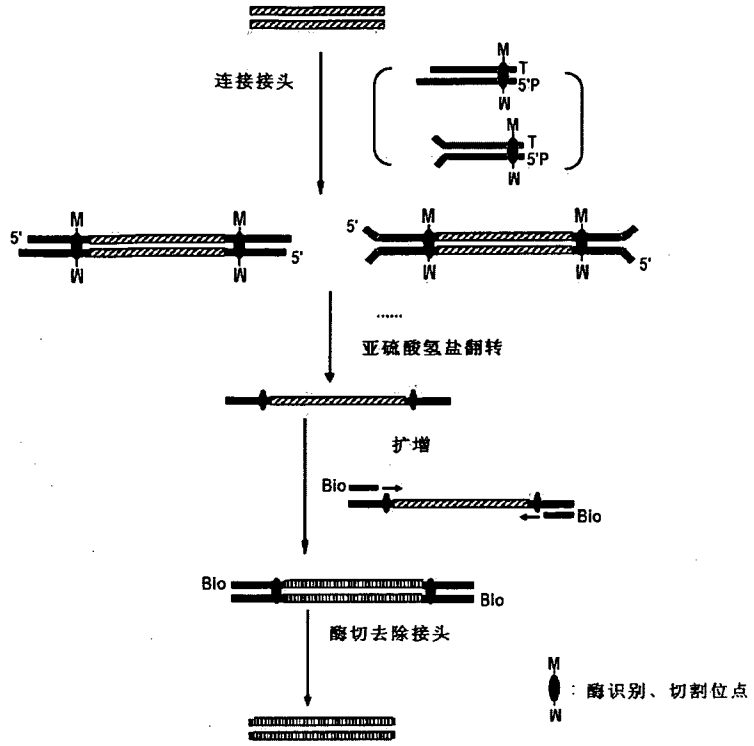

连接接头
M
T
5'P
W
5'
亚硫酸氢盐翻转
扩增
Bio
酶切去除接头
：酶识别、切割位点

Fig.4

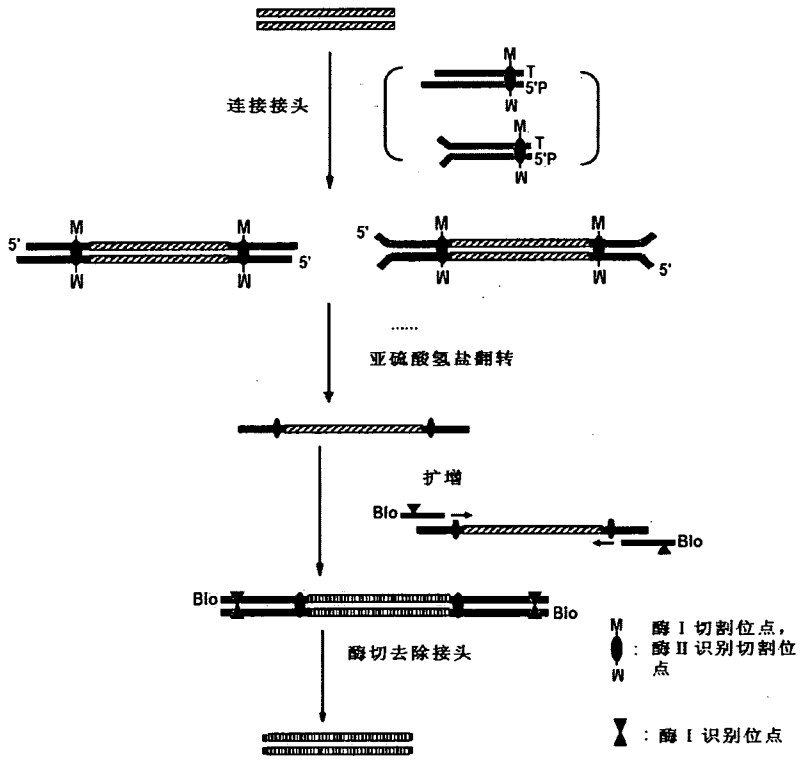

连接接头
M
T
5'P
W
5'
亚硫酸氢盐翻转
扩增
Bio
酶切去除接头
：酶 I 切割位点，酶 II 识别切割位点
：酶 I 识别位点

Fig.5

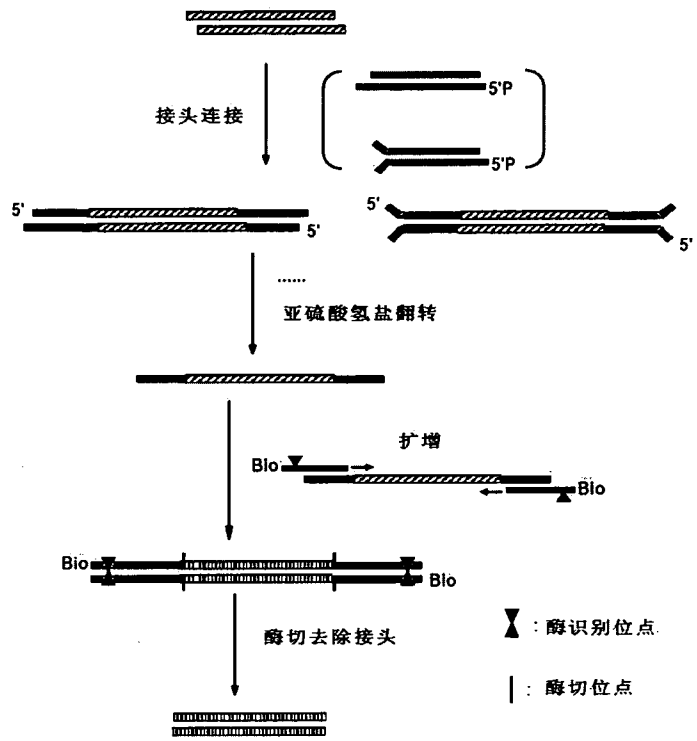

接头连接
5'P
5'P
5'
5'
5'
5'
.....
亚硫酸氢盐翻转
扩增
Bio
Bio
Bio
Bio
酶切去除接头
：酶识别位点
：酶切位点

Fig.6

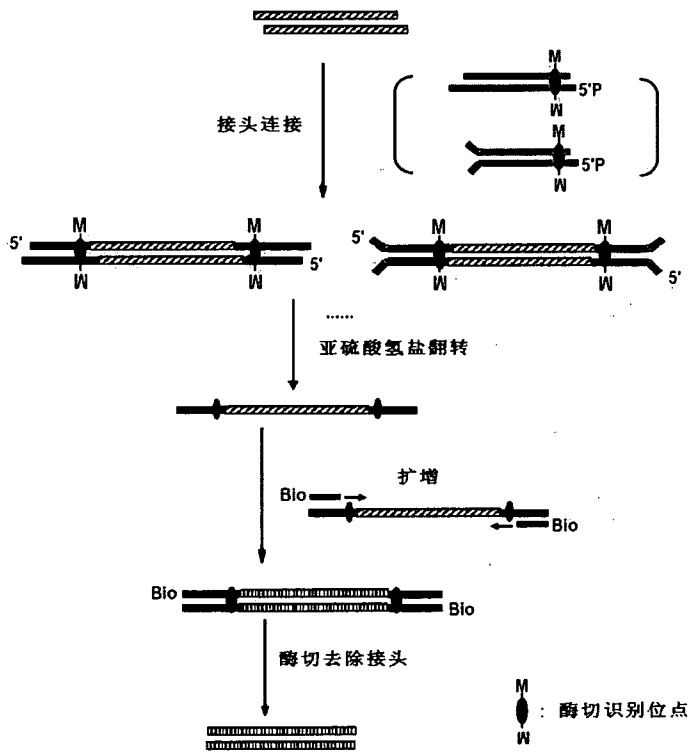

接头连接
M
5'P
W
M
5'P
W
5'
M
M
W
W
5'
5'
M
M
W
W
5'
.....
亚硫酸氢盐翻转
扩增
Bio
Bio
Bio
Bio
酶切去除接头
M
W
：酶切识别位点

Fig.7

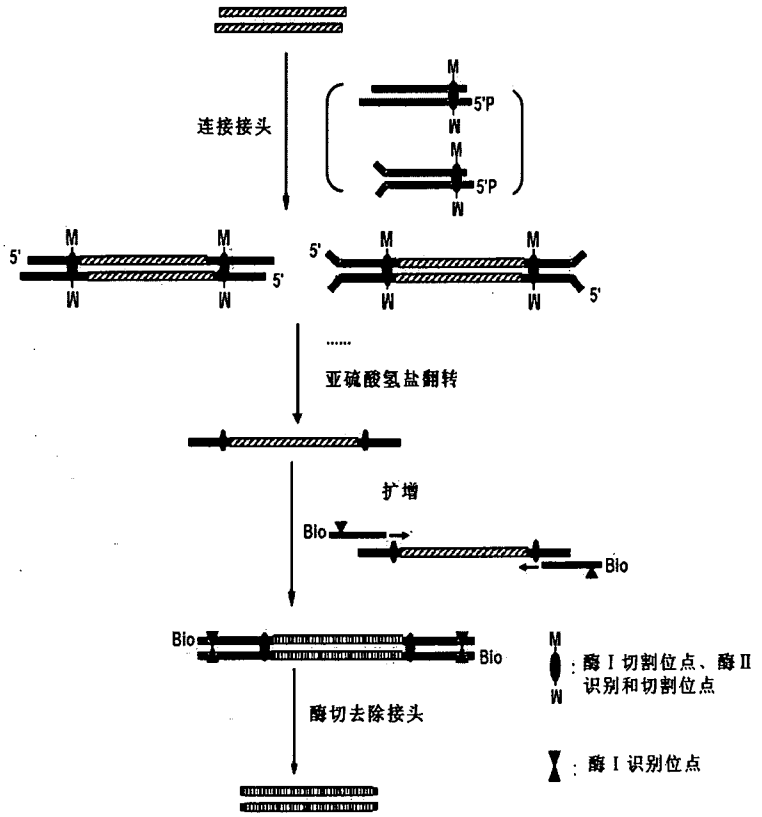

连接接头
M
5'P
W
M
5'P
W
5'
M
M
W
W
5'
5'
M
M
W
W
5'
.....
亚硫酸氢盐翻转
扩增
Bio
Bio
Bio
Bio
M
W
：酶Ⅰ切割位点、酶Ⅱ识别和切割位点
酶切去除接头
：酶Ⅰ识别位点

Fig.8

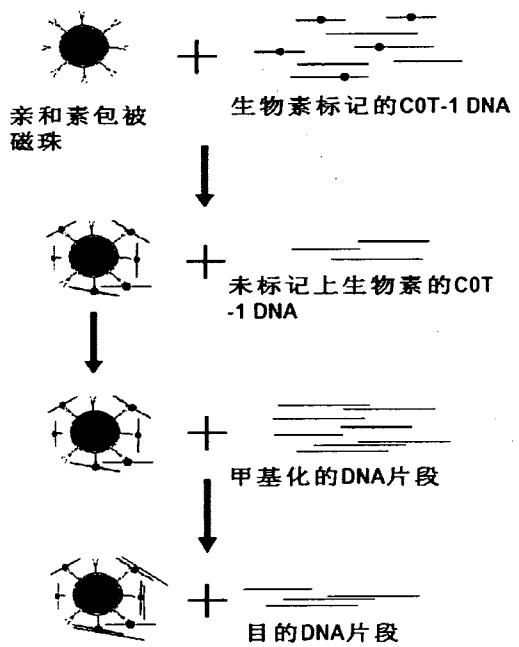

亲和素包被磁珠
生物素标记的C0T-1 DNA
未标记上生物素的C0T-1 DNA
甲基化的DNA片段
目的DNA片段

Fig.9

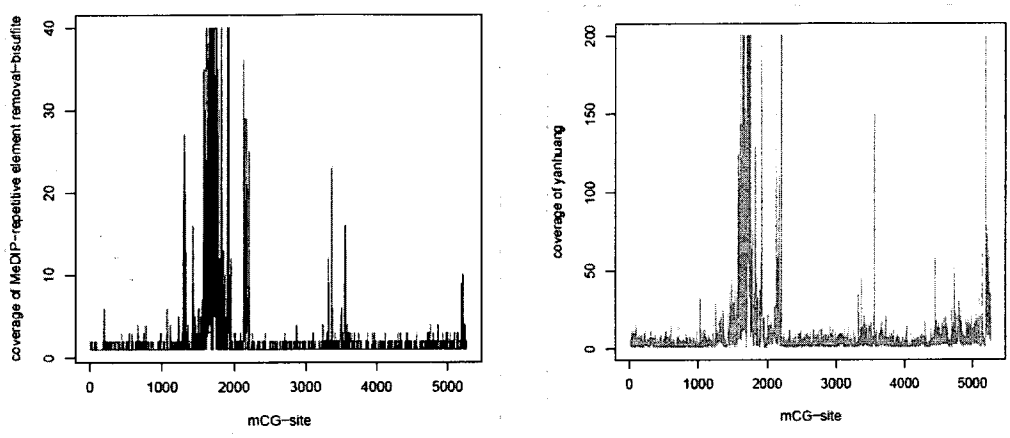

coverage of MeDIP--repetitive element removal--bisulfite
0
10
20
30
40
0
1000
2000
3000
4000
5000
mCG-site
coverage of yanhuang
0
50
100
150
200
0
1000
2000
3000
4000
5000
mCG-site

Fig.10

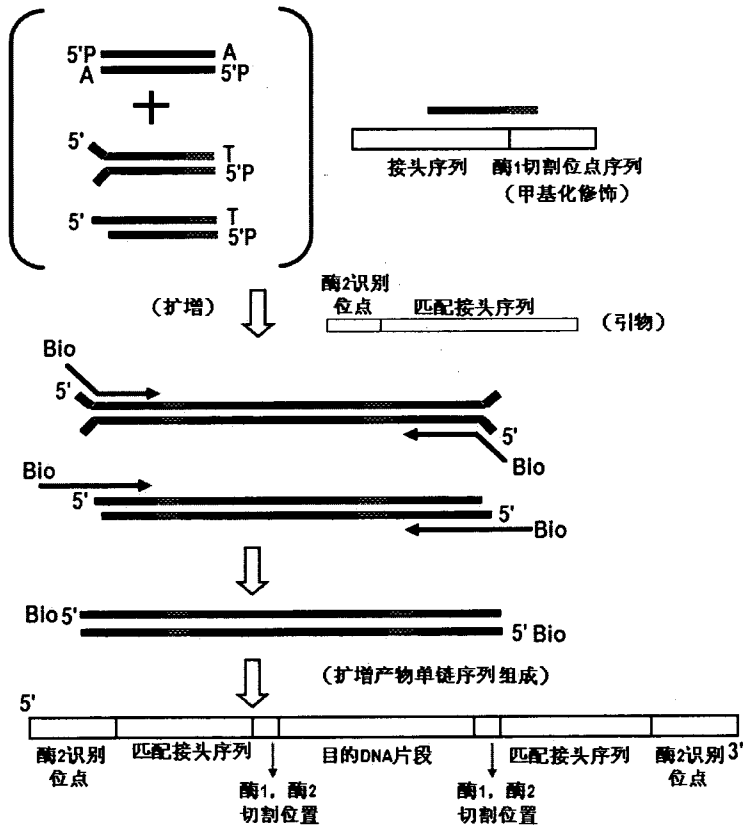

5'P
A
A
5'P
5'
T
5'P
5'
T
5'P
接头序列
酶1切割位点序列
（甲基化修饰）
（扩增）
酶2识别
位点
匹配接头序列
（引物）
Bio
5'
5'
Bio
Bio
5'
5'
Bio
Bio 5'
5' Bio
（扩增产物单链序列组成）
5'
酶2识别
位点
匹配接头序列
目的DNA片段
匹配接头序列
酶2识别
位点
3'
酶1，酶2
切割位置
酶1，酶2
切割位置

Fig.11

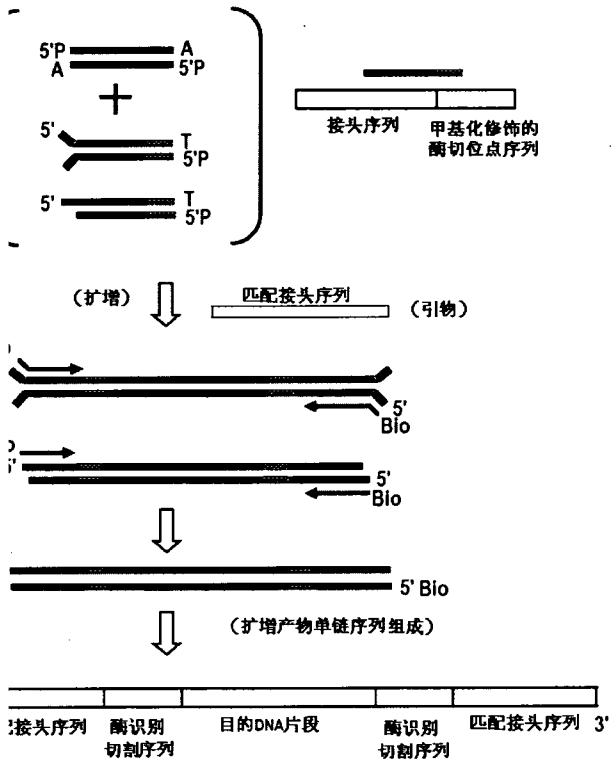

5'P
A
A
5'P
5'
T
5'P
5'
T
5'P
接头序列
甲基化修饰的
酶切位点序列
(扩增)
匹配接头序列
(引物)
5'
Bio
5'
Bio
5' Bio
(扩增产物单链序列组成)
接头序列
酶识别
切割序列
目的DNA片段
酶识别
切割序列
匹配接头序列
3'

Fig.12

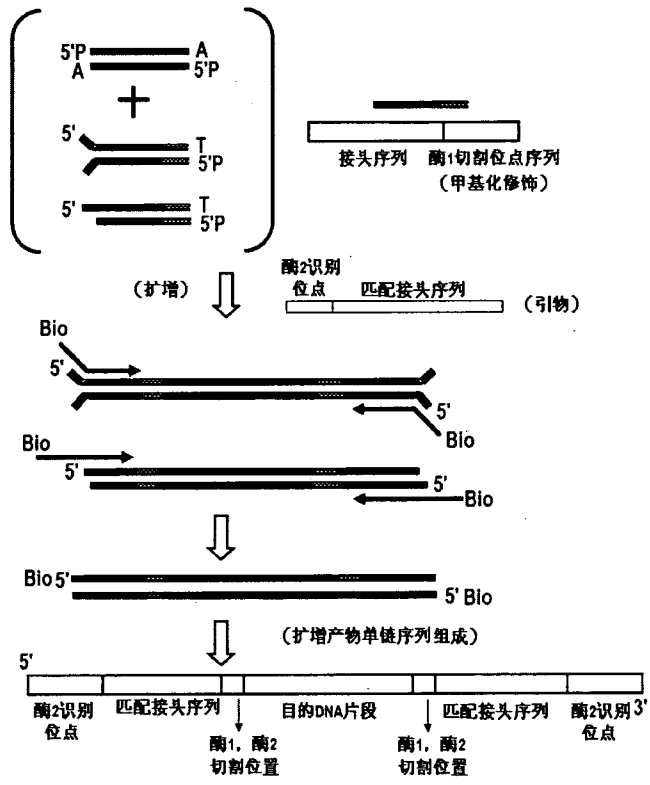

5'P
A
A
5'P
5'
T
5'P
5'
T
5'P
接头序列
酶1切割位点序列
(甲基化修饰)
(扩增)
酶2识别
位点
匹配接头序列
(引物)
Bio
5'
5'
Bio
Bio
5'
5'
Bio
Bio 5'
5' Bio
(扩增产物单链序列组成)
5'
酶2识别
位点
匹配接头序列
酶1，酶2
切割位置
目的DNA片段
匹配接头序列
酶1，酶2
切割位置
酶2识别
位点
3'

Fig.13

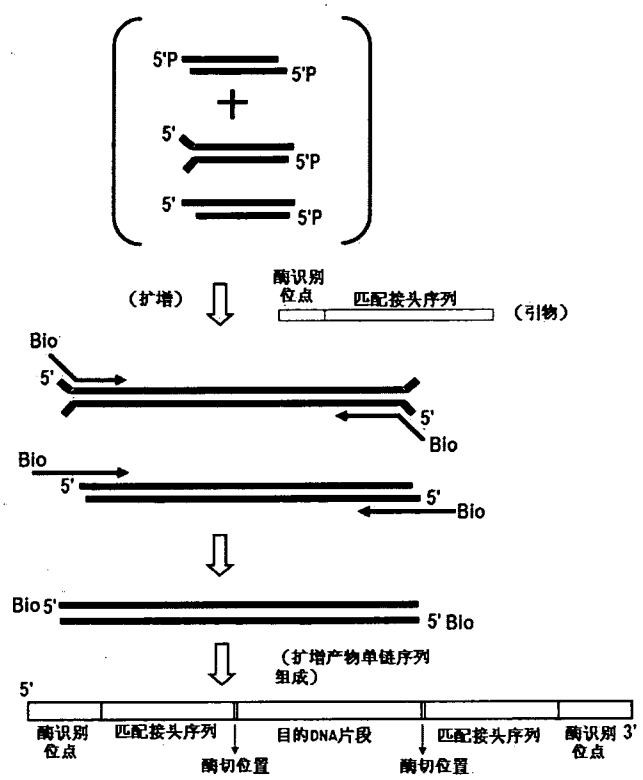

5'P
5'P
5'
5'P
5'
5'P
(扩增)
酶识别位点
匹配接头序列
(引物)
Bio
5'
5'
Bio
Bio
5'
5'
Bio
Bio 5'
5' Bio
(扩增产物单链序列组成)
5'
酶识别位点
匹配接头序列
目的DNA片段
匹配接头序列
酶识别位点 3'
酶切位置
酶切位置

Fig.14

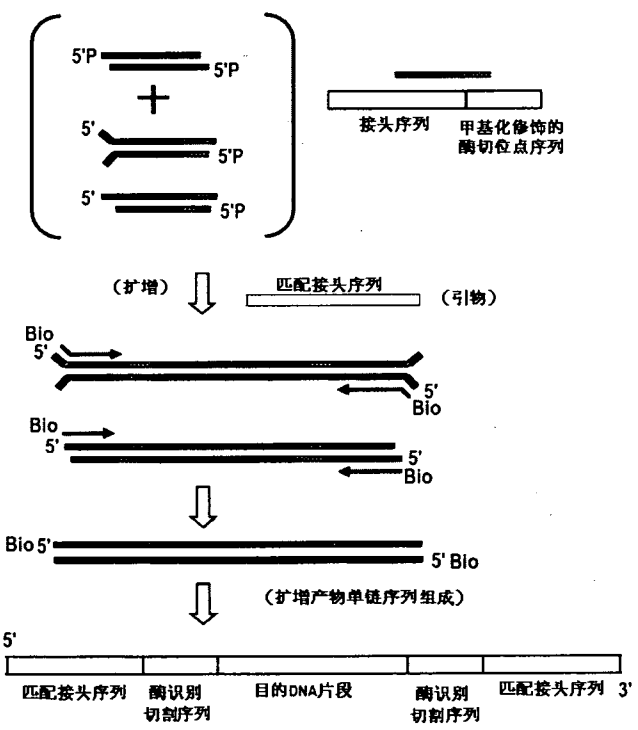

5'P
5'P
5'
5'P
5'
5'P
接头序列
甲基化修饰的酶切位点序列
(扩增)
匹配接头序列
(引物)
Bio
5'
5'
Bio
Bio
5'
5'
Bio
Bio 5'
5' Bio
(扩增产物单链序列组成)
5'
匹配接头序列
酶识别切割序列
目的DNA片段
酶识别切割序列
匹配接头序列 3'

Fig.15

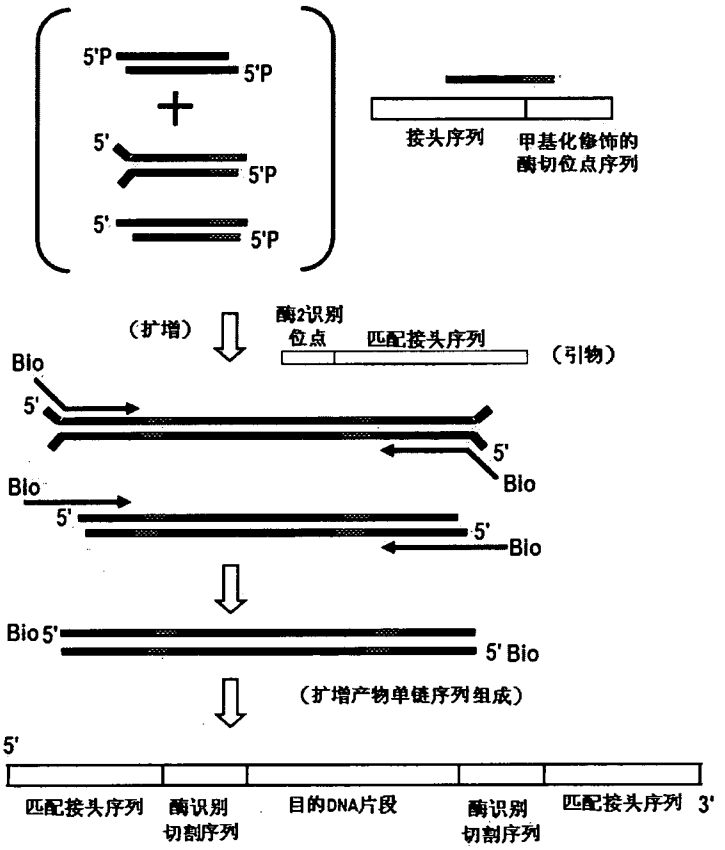

5'P
5'P
5'
5'P
5'
5'P
接头序列
甲基化修饰的
酶切位点序列
(扩增)
酶2识别
位点
匹配接头序列
(引物)
Bio
5'
5'
Bio
Bio
5'
5'
Bio
Bio 5'
5' Bio
(扩增产物单链序列组成)
5'
匹配接头序列
酶识别
切割序列
目的DNA片段
酶识别
切割序列
匹配接头序列
3'

Fig.16

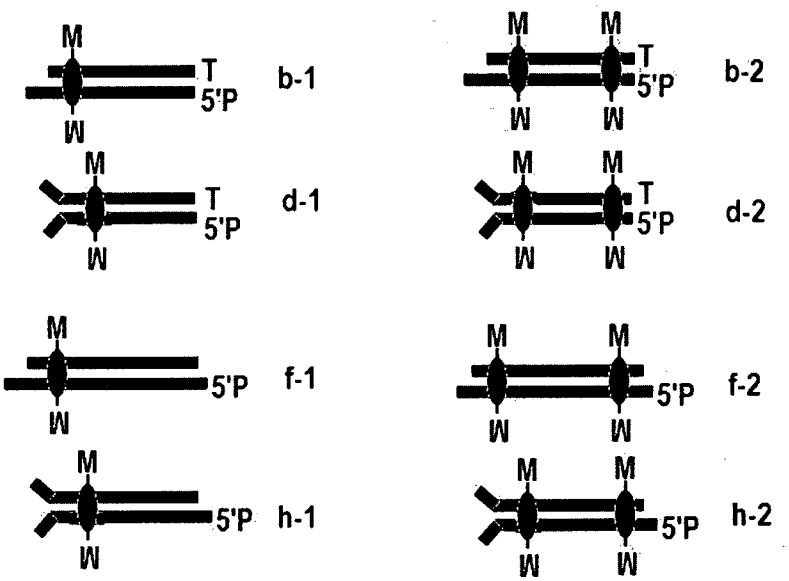

M
T
5'P
W
b-1
b-2
d-1
d-2
f-1
f-2
h-1
h-2

Fig.17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2010/001219

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (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)

IPC: C12Q

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)

CNPAT, CNKI, WPI, EPODOC, ISI-WEB OF KNOWLEDGE,EMBASE: CpG， methyl DNA, methylation, methylated, sequencing, adapter, bisulfite, DNA immunoprecipitation, MeDIP, MBD, COT-1, repetitive

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO2007010004 A1 (EPIGENOMICS AG) 25 January 2007 (25.01.2007), see the whole document	1-7, 9
A	Tim Hui-Ming Huang, et al., Methylation profiling of CpG islands in human breast cancer cells, Human Molecular Genetics, 1999, Vol. 8, No. 3, pages 459-470	1-7, 9

☒ 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
3 May 2011 (03.05.2011)	**26 May 2011 (26.05.2011)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer QIU,Fuen Telephone No. (86-10)62411034

Form PCT/ISA /210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2010/001219

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Yoshinao Ruike, et al., Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. BMC Genomics, Feb.2010, Vol. 11, No. 137, pages 1471-2164	1-7, 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2010/001219

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 8
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
The subject-matter of claim 8 is unclear. Claim 8 claims an automatic device for executing the method of claim 1 or 2, however, the device is not defined in terms of the structural features of the product. Thus, claim 8 does not meet the requirements of Article 6 PCT in that the matter for which protection is sought is not clearly defined.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2010/001219

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2007010004 A1	25.01.2007	EP 1904650 A1	02.04.2008
		DE 102005034628 A1	01.02.2007
		US 2009208941 A1	20.08.2009
		DE 102005034628 B4	23.08.2007

国际检索报告

国际申请号
PCT/CN2010/001219

A. 主题的分类

C12Q 1/68 (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)

IPC: C12Q

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）)

CNPAT, CNKI, WPI, EPODOC, ISI-WEB OF KNOWLEDGE, EMBASE: 甲基化，CpG，测序，接头，亚硫酸氢盐，甲基化免疫共沉淀，重复序列，methyl DNA, methylation, methylated, sequencing, adapter, bisulfite, DNA immunoprecipitation, MeDIP, MBD, COT-1, repetitive

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	WO2007010004 A1 (EPIGENOMICS AG) 25.1 月 2007 (25.01.2007), 参见全文	1-7, 9
A	Tim Hui-Ming Huang等, Methylation profiling of CpG islands in human breast cancer cells, Human Molecular Genetics, 1999, 卷8, 第3期, 第459-470页	1-7, 9
A	Yoshinao Ruike等, Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. BMC Genomics, 2月2010, 卷11, 第137期, 第1471-2164页	1-7, 9

☐ 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
3. 5 月 2011 (03.05.2011)	**26.5 月 2011 (26.05.2011)**
ISA/CN 的名称和邮寄地址： 中华人民共和国国家知识产权局 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 **邱福恩** 电话号码：(86-10) **62411034**

国际检索报告

国际申请号
PCT/CN2010/001219

第II栏　某些权利要求被认为是不能检索的意见(续第 1 页第 2 项)

根据条约第 17 条(2)(a)，对某些权利要求未做国际检索报告的理由如下：

1. ☐ 权利要求：

因为它们涉及不要求本单位进行检索的主题，即：

2. ☒ 权利要求：8

因为它们涉及国际申请中不符合规定的要求的部分，以致不能进行任何有意义的国际检索，

具体地说：权利要求 8 的技术方案不清楚。权利要求 8 要求保护一种执行权利要求 1 或 2 的方法的自动装置，但是没有根据产品的结构特征来限定该装置。因此，该权利要求所寻求保护的范围不清楚，不符合 PCT 条约第 6 条的规定。

3. ☐ 权利要求：

因为它们是从属权利要求，并且没有按照细则 6.4(a)第 2 句和第 3 句的要求撰写。

第III栏　缺乏发明单一性的意见(续第 1 页第 3 项)

本国际检索单位在该国际申请中发现多项发明，即：

1. ☐ 由于申请人按时缴纳了被要求缴纳的全部附加检索费，本国际检索报告涉及全部可作检索的权利要求。

2. ☐ 由于无需付出有理由要求附加费的劳动即能对全部可检索的权利要求进行检索，本单位未通知缴纳任何附加费。

3. ☐ 由于申请人仅按时缴纳了部分被要求缴纳的附加检索费，本国际检索报告仅涉及已缴费的那些权利要求。具体地说，是权利要求：

4. ☐ 申请人未按时缴纳被要求缴纳的附加检索费。因此，本国际检索报告仅涉及权利要求书中首先提及的发明；包含该发明的权利要求是：

关于异议的说明：
☐ 申请人缴纳了附加检索费，同时提交了异议书，适用时，缴纳了异议费。
☐ 申请人缴纳了附加检索费，同时提交了异议书，但未在通知书规定的时间期限内缴纳异议费。
☐ 缴纳附加检索费时未提交异议书。

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2010/001219

检索报告中引用的专利文件	公布日期	同族专利	公布日期
WO 2007010004 A1	25.01.2007	EP 1904650 A1	02.04.2008
		DE 102005034628 A1	01.02.2007
		US 2009208941 A1	20.08.2009
		DE 102005034628 B4	23.08.2007