(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
2016 年 6 月 30 日 (30.06.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/101258 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2014/095120

(22) 国际申请日: 2014 年 12 月 26 日 (26.12.2014)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申 请 人: **北 京 大 学** **(PEKING UNIVERSITY)** [CN/CN]; 中国北京市海淀区颐和园路 5 号, Beijing 100871 (CN)。 **首都医科大学附属北京世纪坛医院 (BEIJING SHIJITAN HOSPITAL, CAPITAL MEDICAL UNIVERSITY)** [CN/CN]; 中国北京市海淀区铁医路 10 号, Beijing 100038 (CN)。

(72) **发明人**: **文路 (WEN, Lu)**; 中国北京市海淀区颐和园路 5 号北京大学, Beijing 100871 (CN)。 **彭吉润 (PENG, Jirun)**; 中国北京市海淀区铁医路 10 号首都医科大学附属北京世纪坛医院, Beijing 100038 (CN)。 **黄岩谊 (HUANG, Yanyi)**; 中国北京市海淀区颐和园路 5 号北京大学, Beijing 100871 (CN)。 **汤富酬 (TANG, Fuchou)**; 中国北京市海淀区颐和园路 5 号北京大学, Beijing 100871 (CN)。 **刘晓萌 (LIU, Xiaomeng)**; 中国北京市海淀区颐和园路 5 号北京大学, Beijing 100871 (CN)。 **李静宜 (LI, Jingyi)**; 中国北京市海淀区颐和园路 5 号北京大学, Beijing 100871 (CN)。 **郭化虎 (GUO, Huahu)**; 中国北京市海淀区铁医路 10 号首都医科大学附属北京世纪坛医院, Beijing 100038 (CN)。

(74) **代理人**: **北京君尚知识产权代理事务所（普通合伙） (BEIJING JOYSHINE INTELLECTUAL PROPERTY OFFICE)**; 中国北京市海淀区北四环西路 68 号左岸工社 1316-1317 室, Beijing 100080 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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,

[见续页]

(54) **Title**: METHOD FOR DETECTING DIFFERENTIALLY METHYLATED CPG ISLANDS ASSOCIATED WITH ABNORMAL STATE OF HUMAN BODY

(54) **发明名称**: 一种检测与人体异常状态相关的差异甲基化 CpG 岛的方法

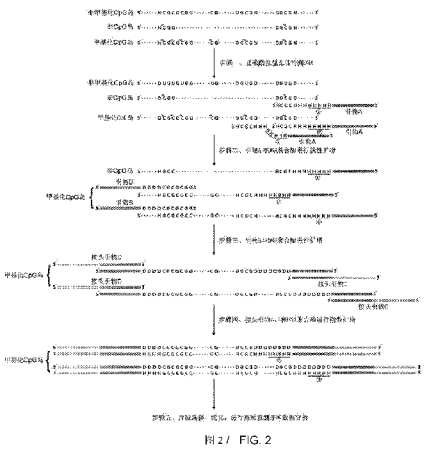


图 2 / FIG. 2

AA NON METHYLATED CPG ISLAND
BB NON CPG ISLAND
CC METHYLATED CPG ISLAND
DD STEP 1: TREATMENT OF DNA FOR DETECTING WITH HYDROSULFITE
EE NON METHYLATED CPG ISLAND
FF NON CPG ISLAND
GG PRIMER A
HH METHYLATED CPG ISLAND
II PRIMER A
JJ PRIMER A
KK STEP 2: LINEAR AMPLIFICATION OF PRIMER A AND DNA POLYMERASE
LL NON CPG ISLAND
MM PRIMER B
NN METHYLATED CPG ISLAND
OO PRIMER B
PP STEP 3: AMPLIFICATION OF PRIMER B AND DNA POLYMERASE
QQ ADAPTER PRIMER D
RR METHYLATED CPG ISLAND
SS ADAPTER PRIMER D
TT ADAPTER PRIMER C
UU ADAPTER PRIMER C
VV STEP 4: EXPONENTIAL AMPLIFICATION OF ADAPTER PRIMER C,D AND DNA POLYMERASE
WW METHYLATED CPG ISLAND
XX STEP 5: FRAGMENTS SELECTION, FRAGMENTS PURIFICATION, HIGH THROUGHPUT SEQUENCING AND DATA ANALYSIS OF FRAGMENTS

(57) **Abstract**: Disclosed is a method for detecting differentially methylated CpG islands associated with an abnormal state of a human body, characterized by detecting very minute amounts of methylated CpG short tandem nucleic acid sequences in highly fragmented DNA samples with genome scale, identifying differentially methylated CpG islands associated with abnormal state of human body and determining the corresponding abnormal state of human body. Sequencing libraries are constructed by using CpG short tandem sequences as primers to perform three steps of PCR reactions on DNAs which are conversed by modifiers, and detections of very minute amounts of methylated CpG short tandem nucleic acid sequences are implemented with high throughput sequencing technology. A group of genome sequences and methylation patterns of differentially methylated CpG islands which are associated with hepatocellular carcinoma are also disclosed; they may be used for distinguishing between hepatocellular carcinoma and non-cancerous state.

(57) **摘要**: 本发明公开了一种检测与人体异常状态相关的差异甲基化 CpG 岛的方法，其特征在于，以基因组规模检测高度片段化 DNA 样品中极微量甲基化 CpG 短串联核酸序列，从而鉴定与人体异常状态相关的差异甲基化 CpG 岛并确定相应人体异常状态。以 CpG 短串联序列为引物对修饰剂转换的 DNA 进行三步 PCR 反应构建测序文库并通过高通量测序技术实现对极微量甲基化 CpG 短串联核酸序列的检测。本发明还公开了一组与肝细胞癌相关的差异甲基化 CpG 岛的基因组序列和甲基化模式可用于区分肝细胞癌与无癌状态。

WO 2016/101258 A1

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, 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))。

一种检测与人体异常状态相关的差异甲基化 CpG 岛的方法

技术领域

本发明涉及生物医学领域。更具体的，本发明涉及一种通过 DNA 甲基化信息确定人体具有异常状态如肿瘤的系统和方法。

背景技术

DNA 甲基化是指胞嘧啶（C）被添加甲基而成为甲基胞嘧啶（mC）的表观遗传修饰过程，主要发生在 CpG 位点（CpG 表示核苷酸对，其中鸟嘌呤（G）在 DNA 链上紧随 C 后）。CpG 在人类基因组的大部分区域散在分布并处于高甲基化状态，但在一些被称为 CpG 岛的区段（全人类基因组有近 3 万个），CpG 聚集成簇而在人体正常生理状态下大多数处于去甲基化状态。CpG 岛富集于基因启动子，大约有 60%的人类基因启动子含有 CpG 岛。去甲基化状态是保证 CpG 岛启动子发挥转录活性的先决条件，而反之，其甲基化则通过募集甲基结合蛋白导致启动子长期沉默。在正常生理状态下，这种机制参与了包括 X 染色体失活和基因印记在内的一些重要生理机制(Jones and Baylin., Nat Rev Genet,2002)。

CpG 岛异常高甲基化是人类肿瘤的一个普遍现象，不仅存在于几乎所有常见人类肿瘤中，包括肝细胞癌、结直肠癌、肺癌、胃癌、乳腺癌和前列腺癌等，而且发生在肿瘤早期乃至癌前病变阶段，并累及大量基因(综述请见 Baylin and Jones, Nat Rev Cancer, 2011; Feinberg and Tycko, Nat Rev Cancer, 2004)。近年来的基因组学研究表明，全基因组有数千个 CpG 岛可在不同肿瘤中出现异常高甲基化，据估算，平均每个肿瘤在早期阶段可有数百个 CpG 岛发生异常高甲基化(Ammerpohl et al., Int J Cancer 2012; Costello et al., Nat Genet 2000; Hinoue et al., Genome Res 2012; Keshet et al., Nat Genet 2006)。一些 CpG 岛的异常高甲基化还与肿瘤的预后以及耐药有着密切关系。此外，CpG 岛甲基化异常还可发生于其它的人体异常状态，例如系统性红斑狼疮(Robertson. Nat Rev Genet, 2005)。人的体液（包括血液、尿液、唾液和排泄物等）中存在游离核酸（cell-free DNA），检测游离核酸能够尽量减少对病人进行组织取样所造成的伤害，从而无创性地获取人体的状态信息(综述请见 Schwarzenbach et al., Nat Rev Cancer, 2011)。研究表明，肿瘤细胞 DNA 在肿瘤发生早期即可释放入血，因此，游离肿瘤 DNA（cell-free tumor DNA）检测作为肿瘤性疾病的一种无创性筛选、监测与追踪方法具有广阔的应用前景。与肿瘤 DNA 的序列性改变相比，CpG 岛异常高甲基化发生更为普遍，因而在肿瘤筛查方面更具优势(Heyn and Esteller, Nat Rev Genet, 2012; Schwarzenbach et al., 2011)。

然而，由于游离肿瘤 DNA 含量极低（皮克级别）、只占全部游离核酸的极小部分（0.1%至 1%）且已高度片段化（70 至 200 碱基对），其检测具有极大的技术挑战性(Schwarzenbach et al., Nat Rev Cancer, 2011)。目前所报道的游离肿瘤 DNA 的 CpG 岛甲基化检测方法——包括甲基化特异性 PCR(Herman, et al., PNAS, 1999)、限制性酶切-PCR、焦磷酸测序法(Korshunova, et al., Genome Research)、Methylight (Eads et al., Nucleic Acids Res, 2000)和 MethylBeaming (Li et al., Nat Biotechnol，2009)——具有较高的灵敏度，适用于检测高度片段化 DNA 样品，但每个反应只能检测单个或少数位点。而另一方面，DNA 甲基化的组学检测方法(综述请见 Laird, Nat Rev Genet, 2010)，包括 Infinium 甲基化芯片(Bibikova et al., Genome Research, 2006)和简化代表性亚硫酸氢盐测序（RRBS, Meissner et al., Nature, 2008)），能以基因组规模同时检测大量位点，已

被广泛地用于来自肿瘤细胞或组织的基因组 DNA，却因为技术方法上的限制难以对游离肿瘤 DNA 进行检测分析。例如，Infinium 甲基化芯片需要微克级别的 DNA 量(Bibikova et al., Genome Research, 2006)，RRBS 可以检测皮克级别的 DNA 量，但所采用的限制性酶切后筛取小片段的 CpG 岛富集策略不适于分析已经高度片段化的游离肿瘤 DNA (Gu et al., Nature Protocol, 2011)。有报道直接对全部血浆游离 DNA 进行亚硫酸氢盐测序分析用于肿瘤检测，这种方法可用于分析在肿瘤细胞基因组中更为广泛分布的低甲基化改变，但由于每个 CpG 岛覆盖度很浅（只大约覆盖 1 次），因此难以对肿瘤细胞基因组的 CpG 岛异常高甲基化进行分析(Chan, et al., PNAS, 2013)。

综上所述，目前尚缺乏一种能够以基因组规模高灵敏地检测游离 DNA 中极微量的能反映人体异常状态的甲基化 CpG 岛的分析方法。

发明内容

在第一方面，本发明提供了一种检测与人体异常状态相关的差异甲基化基因组位点的方法，包括下列步骤：

1） 获得来自人的 DNA 样品。

所述 DNA 样品来自人的细胞、组织、血液、体液、尿液、唾液、排泄物或其组合；优选来自人的血浆或尿液游离 DNA。外周血中的游离 DNA 主要来自造血细胞系；血清与血浆游离 DNA 比较，含有更多来自造血细胞系的 DNA，因此采用来自血浆的游离 DNA 样品能够减少检测背景。外周血中的游离 DNA 通过尿液排出体外，因此尿液中也含有大量游离 DNA，尿液检验较外周血检验是一种更无创伤性的检验方式。

2）以基因组规模检测所述 DNA 样品中 CpG 短串联核酸序列的甲基化程度。

所述基因组规模是指同时检测基因组中 100 个以上 CpG 岛的高通量分析，目前已有的游离 DNA 的 CpG 岛甲基化检测方法均只能检测单个或少数 CpG 岛，因此本发明方法与已有方法相比有显著的改进。本发明方法改进的关键是采用了 CpG 短串联核酸序列作为靶向检测位点，所述 CpG 短串联核酸序列是指在一段较短的核酸序列（7 至 9 个碱基对）中含有较多（3 个或以上）CpG 核苷酸对。这些 CpG 短串联核酸序列具有 3 个特征：a）在人类基因组中有大量拷贝，人类基因组中共有数万个 CpG 短串联序列，因此以它们为靶位点可以进行基因组规模的检测；b）这些 CpG 短串联序列高度富集于 CpG 岛，因此它们含有重要的能够用于判断人体异常状态的 DNA 甲基化信息；c）这些 CpG 短串联序列具有较高的解链温度，因此能够被有效地扩增。以 CGCGCGG 为例，如图 1 所示，人类基因组中共有 23,726 个 CGCGCGG 序列，其中 20,525（86.5%）位于 CpG 岛；全基因组中 27,435 个 CpG 岛中有 9,373（34.2%）个含有 CGCGCGG 序列；其解链温度为 43 度，已经接近常规 PCR 反应的退火温度。因此单以 CGCGCGG 为靶向序列，也足以对 CpG 岛进行基因组规模的检测。如图 1 所示，按短串联在人类基因组中的数目与其在 CpG 岛富集程度排序，CGCGCGG 是最优选的 CpG 短串联；除此以外，较优好的 CpG 短串联包括 CGGCGGCGG（全基因组 11,276 个，其中 87.6%位于 CpG 岛），CGCGGCGG（全基因组 12,322 个，其中 83.6%位于 CpG 岛），CGGCGCGG（全基因组 11,354 个，其中 80.8%位于 CpG 岛），CGCGGCGC（全基因组 9,947 个，其中 90.3%

位于 CpG 岛），CGGCGCGC（全基因组 7,885 个，其中 86.6%位于 CpG 岛），CGCGCGC（全基因组 33,818 个，其中 60.7%位于 CpG 岛），CGCGCGA（全基因组 5,923 个，其中 76.8%位于 CpG 岛），CGCGCGT（全基因组 10,553 个，其中 57%位于 CpG 岛）。同时针对多个 CpG 短串联核酸序列，可以进一步增加检测的规模，并提高检测的灵敏度。在短串联序列中，除了 CpG 以外含有尽量多的 G，这样不仅能够进一步增加解链温度和提高扩增效率，而且能够增加在 CpG 岛中的富集程度。本领域技术人员可以理解，核酸序列杂交与扩增的特异性主要取决于其 3'端序列，而 5'端 CpG 核苷酸对的作用逐渐减少。对于 CpG 短串联序列来说，3'端前 3 个核苷酸的作用最为重要，其次是 4 至 6 个核苷酸，再次为 7 至 9 个核苷酸，3'端起前 5 至 6 个核苷酸优选八种形式：CGCGG、CGCGA、CGCGT、CGCGC、CGGCGG、CGGCGA、CGGCGT 和 CGGCGC，其中优选 CGCGG 与 CGGCGG 两种形式，更优选 CGCGG 形式。不典型的 CpG 短串联序列，3'端序列与典型的 CpG 短串联序列相同但总体含有较少（2 个）CpG 核苷酸对，如 GGCGCGG 序列，也可以达到以基因组规模检测 CpG 岛甲基化的目的，但是检测效率要低于典型的 CpG 短串联序列。增加 CpG 短串联序列的长度（大于 9 个碱基对）和相应的 CpG 核苷酸对的个数，能够进一步增加解链温度，但是靶位点的数量会显著减少；如上所述，在 5'端增加的 CpG 核苷酸对对于检测的特异性影响实际较少。

3）将所述 CpG 短串联核酸序列在所述 DNA 样品中的甲基化程度与其在来自正常人群的 DNA 样品中的甲基化程度进行比较，鉴定与人体异常状态相关的差异甲基化 CpG 岛；以及，

4）所述差异甲基化 CpG 岛被用于诊断相应的人体异常状态。

本发明方法一方面可以应用于临床研究，大规模分析鉴定与人体异常状态相关的差异甲基化 CpG 岛；另一方面可以应用于临床分子检验，通过所鉴定的差异甲基化 CpG 岛预测个体是否具有异常状态。

已有的 DNA 甲基化组学分析方法如 Infinium 甲基化芯片和 RRBS 可以大规模分析鉴定组织或细胞 DNA 样品中与人体异常状态相关的差异甲基化 CpG 岛，但是所得结果无法直接转化为游离 DNA 的临床分子检验应用。本发明方法的优势在于能够直接大规模鉴定游离 DNA 样品中与人体异常状态相关的差异甲基化 CpG 岛，因此所得结果能与游离 DNA 的临床分子检验直接衔接。

肿瘤性疾病是本发明方法实施的重要方面之一。根据本发明的一个实施例，通过比较分析肝细胞癌患者与无肿瘤个体的血浆游离 DNA，鉴定获得 132 个能够显著区分肝细胞癌患者与无肿瘤个体的差异甲基化 CpG 岛，其核酸序列如序列 1~132 所示；血浆游离 DNA 中检测到这些 CpG 岛的异常甲基化提示个体患有肝细胞癌。该 132 个差异甲基化 CpG 岛分为两类：I 型血浆差异甲基化 CpG 岛（68 个，序列号：1~68）同时在组织和血浆游离 DNA 中显示差异甲基化；II 型血浆差异甲基化 CpG 岛（64 个，序列号：69~132）仅在血浆游离 DNA 中显示差异甲基化。

部分 I 型血浆差异甲基化 CpG 岛的与肝细胞癌组织的相关性已有报道(Shen, et al., 2012, Hepatology; Ammerpohl, et al., 2012, Int J Cancer; Song, et al, 2013, PLoS One)，但是已有的甲基化组分析方法未能得知它们中哪些在肝细胞癌患者和正常人的游离 DNA 样本之间也显现差异甲基化。本发明方法公开的 I 型血浆差异甲基化 CpG 岛直接将游离 DNA 甲基化检测与个体是否患有肝细胞癌相关联。

II 型血浆差异甲基化 CpG 岛与肝细胞癌的关系之前未被揭示，因其甲基化水平在肝细胞癌组织与非癌肝组织之间无显著差异，而仅在肝细胞癌患者与非癌个体的血浆游离 DNA 之间显现明显差异。这是由于该类 CpG 岛在正常肝组织中已在一定程度上被高甲基化；而其在外周血白细胞中，因而也在正常人血浆游

离 DNA 中，显示低甲基化水平；在肝细胞癌发生时，它们会随着肿瘤细胞坏死与凋亡释放入血，因而在游离 DNA 样本中被检测为异常高甲基化。除了肝细胞癌以外，在其它肝组织被破坏的情况下也可能在外周血游离 DNA 中检测到这些 CpG 岛的异常高甲基化。因而，本发明方法公开的 II 型血浆差异甲基化 CpG 岛在游离 DNA 中的异常高甲基化提示个体患有肝细胞癌或其他肝组织被破坏的异常状态。

本发明还提供了一组在无癌个体血浆中甲基化背景很低的 CpG 岛（序列号：133~187），这组 CpG 岛虽然不属于血浆差异甲基化 CpG 岛，但其仍以一定频率在肝细胞癌组织中发生异常高甲基化（属于肿瘤组织差异甲基化 CpG 岛），而其具有低甲基化背景（无癌个体平均 uMePM 值<0.2）和高甲基化检测效率（FMG 结果中平均 MePM 值大于 75），使其能广泛地用于检测早期肿瘤性疾病。

在第二方面，本发明提供了一种以基因组规模检测所述 CpG 短串联核酸序列甲基化程度的方法：甲基化 CpG 短串联扩增与测序法（Methylated CpG Tandems Amplification and Sequencing, MCTA-Seq），如图 2 所示，包括下列步骤：

步骤一、用修饰剂处理 DNA 样品，以便将 DNA 样品中的 C 转换为尿嘧啶（U），而 mC 不变，获得经过转换的 DNA 片段。

处理 DNA 样品的修饰剂在形成单链 DNA 的条件下修饰 C，但不修饰 mC。可以采用亚硫酸氢盐、乙酸盐或柠檬酸盐，优选地，修饰剂是亚硫酸氢盐。可以采用商品化试剂盒将 DNA 样品进行亚硫酸氢盐处理，任选地，可以采用 MethylCode Bisulfite Conversion Kit (Invitrogen)、EZ DNA methylation-Gold Kit (ZYMO) 或 EpiTect Bisulfite Kit (Qiagen)。

步骤二、将所述经过转换的 DNA 片段用引物 A 和 DNA 聚合酶进行至少 1 次线性扩增，以便获得一端能够锚定接头引物 C 的半扩增子；

其中引物 A 由 3’端和 5’端两部分组成，3’端部分用于结合和扩增所述经过转换的 DNA 片段，其特征为，一组能结合所述经过转换的 DNA 片段的随机序列；优选地，除了 CpG 以外，引物 A 的 3’端部分只包含 C、腺嘌呤（A）和胸腺嘧啶（T），经过修饰剂处理的 DNA 样品中大部分 C 已经转化为 U，因此采用这种形式的引物可以提高结合效率；优选地，引物 A 的 3’端起前 7 个核苷酸中包含 1 个 CpG 核苷酸对，一方面，由于非甲基化 CpG 岛中的 CpG 被修饰剂转换为 UpG，这种设计可以达到初步富集甲基化 CpG 岛的作用，另一方面，CpG 可以增加引物的解链温度，提高扩增效率；优选地，引物 A 的 3’端第二个核苷酸为 C，一方面，这可以增加引物的解链温度，提高扩增效率，另一方面，由于引物 3’端部分除了 CpG 以外，只包含 C、A 和 T，3’端起第二个核苷酸为 C 的引物可以抑制引物二聚体的形成。优选的，引物 A 的 3’端和 5’端部分之间包含一段独特分子识别标签序列（unique molecular identifiers），该序列的作用是给每个扩增子在步骤四的 PCR 指数扩增之前加上一条标签，从而得以识别并消除 PCR 过度扩增的产物，优选的，该独特分子识别标签序列为 HHHHH（H = A 或 T 或 C），可以识别 243 种不同的分子。

其中所述引物 A 的 5’端部分用于锚定接头引物 C，其特征为，反向互补序列能够被接头引物 C 结合并进行 PCR 扩增。锚定是指其引物 A 能够通过其 5’端部分将接头引物 C 通过 PCR 反应连接到待测目的片段上。

DNA 聚合酶可以是任何合适的聚合酶，如 Taq 聚合酶、ExTaq、LATaq DNA 聚合酶、AmpliTaq、Amplitaq

Gold、Titanium Taq 聚合酶、KlenTaq DNA 聚合酶、Platinum Taq 聚合酶、Accuprime Taq 聚合酶、Pyrobest DNA 聚合酶、Pfu 聚合酶、Pfu 聚合酶 turbo、Phusion 聚合酶、Pwo 聚合酶、Vent 聚合酶、Vent Exo-聚合酶、SequenaseTM 聚合酶、9。Nm DNA 聚合酶、Therminator DNA 聚合酶、Expand DNA 聚合酶、rTth DNA 聚合酶、DyNazymeTM EXT 聚合酶、DNA 聚合酶 1、T7 DNA 聚合酶、T4 DNA 聚合酶、Bst DNA 聚合酶、phi-29 DNA 聚合酶和 Klenow 片段。

优选地，DNA 聚合酶具有链置换能力。单条 DNA 模板上可能会结合多条引物 A，尤其是 CpG 密度较高的 CpG 岛区域，具有链置换能力的 DNA 聚合酶能够使 5’端的引物得以延伸，使得即使只进行 1 次扩增也会产生类似线性扩增的效果，因此具有链置换能力的 DNA 聚合酶能够获得更高的扩增效率。可以是具有链置换能力的任何合适的聚合酶，包括但不限于 DNA 聚合酶 I (Klenow)大片段（New England Biolabs (NEB)目录号 M0210S）、Klenow 片段(exo-)（NEB 目录号 M0212S）、Bst DNA 聚合酶大片段（NEB 目录号 M0275S）、Vent(exo-) （NEB 目录号 M0257S）、Deep Vent (exo-) （NEB 目录号 M0259S）、M-MulV 逆转录酶（NEB 目录号 M0253S）、9。Nm DNA 聚合酶（NEB 目录号 M0260S）和 Phi-29 DNA 聚合酶（NEB 目录号 M0269S）。优选地，DNA 聚合酶是核酸外切酶缺损的。在一个优选形式中，DNA 聚合酶为 Klenow 片段(exo-)。

线性扩增是指扩增产物量随着扩增次数增加呈线性而非指数增加。如前所述，采用具有链置换能力的 DNA 聚合酶即使只进行 1 次扩增，也会产生类似线性扩增的效果；可以进行 2~30 次线性扩增，但是许多具有链置换能力的 DNA 聚合酶如 Klenow 片段在 DNA 变性过程会失活，需要重新添加。

步骤三、将所述半扩增子用引物 B 和 DNA 聚合酶对甲基化 CpG 短串联核酸序列进行选择性扩增并获得两端能够分别锚定接头引物 C 和接头引物 D 的全扩增子。

其中引物 B 由 3’端和 5’端两部分组成，3’端部分用于选择性扩增甲基化 CpG 短串联核酸序列，其即为所述 CpG 短串联核酸序列。本领域技术人员可以理解，位于引物的 3’端的 CpG 或者 CpG 串联序列容易形成引物二聚体，为了避免引物二聚体的形成，CpG 短串联核酸序列引物的 3’末端第一个核酸设计为 G、A 或 T。优选的，3’端和 5’端部分之间包含一段长度为 2 至 10 个核苷酸的随机序列，优选的，该随机序列只包含 G、A 和 T。加入随机序列的目的是增加引物的解链温度，经过修饰剂处理的 DNA 样品中大部分 C 已经转化为 U，因此第一轮扩增得到的互补链中大部分 G 已转化为 A，因此随机序列只需包含 G、A 和 T。解链温度的增加以 CGCGCGG 为例，其解链温度为 43 度，在 5’端增加 4 个 D（D = G/A/T），解链温度增加至 53 度（AAAACGCGCGG 或 TTTTCGCGCGG）至 66 度（GGGGCGCGCGG），已经能够在常规退火温度下进行有效扩增。引物 B 的 5’端部分用于锚定接头引物 D，其特征为，其反向互补序列能够被接头引物 D 结合并进行聚合酶链式反应扩增。

DNA 聚合酶可以是前述任何合适的聚合酶。优选地，DNA 聚合酶是热启动的。本领域技术人员可以理解，热启动 DNA 聚合酶可以增加扩增反应的特异性，尤其有助于阻止 CpG 短串联引物在扩增之前形成二聚体。

步骤四、将所述全扩增子用接头引物 C、接头引物 D 和 DNA 聚合酶进行聚合酶链式反应指数扩增，以便获得甲基化 CpG 短串联核酸序列的扩增产物；

接头引物此处是指引物的作用与常规高通量测序文库构建方法（如 Illumnia 的 TruSeq DNA Sample Prep

Kit 和 Applied Biosystems (ABI)的 The SOLiD™ Fragment Library Construction Kit）中接头（adaptor）的作用一样，即将待测 DNA 片段结合到测序芯片上并使其能被 PCR 扩增富集以及测序。与常规高通量测序文库构建过程中利用连接反应加接头序列的方法不同，本发明的方法利用引物 A 和引物 B 的 5'部分的锚定序列将接头序列通过 PCR 反应连接到待测 DNA 片段的两端。本领域技术人员可以理解，接头引物 C 与接头引物 D 之一中可以包含标签序列，以便在测序芯片上同时检测多个文库。接头引物 C 与引物 D 对应特定高通量测序平台的接头序列，这些高通量测序平台包括，但不局限于 Illumina 的 Genome Analyzer IIx、HiSeq 和 MiSeq，ABI 的 SoLiD、5500 W Series Genetic Analyzer 和 Ion Torrent PGM，Roche454 的 GS Junior 和 GS FLX+。

步骤五、将所述扩增产物进行分离纯化，所述扩增产物构成高通量测序文库；以及，将所述高通量测序文库上机测序及数据分析。

对扩增产物进行分离纯化的方法可以是任何合适的方法，包括但不局限于磁珠纯化、纯化柱纯化和琼脂糖胶电泳纯化。优选地，纯化方法能够对目的片段的长度进行选择，可采用 3~4%琼脂糖胶电泳的方法选择 160-250bp 之间目的片段并进行胶回收纯化。

用于分析测序文库的高通量测序平台包括，但不局限于 Illumina 的 Genome Analyzer IIx、HiSeq 和 MiSeq 测序平台，ABI 的 SoLiD、5500 W Series Genetic Analyzer 和 Ion Torrent PGM 测序平台，Roche454 的 GS Junior 和 GS FLX+测序平台。

数据分析的方法不受限制，可以应用任何合适的数据分析和序列比对软件，包括但不局限于 Bismark、BSMAP、Bowtie、SOAP 和 R 软件包等。

在第三方面，本发明提供了 MCTA-Seq 的试剂盒，该试剂盒包含：引物 A、引物 B、接头引物 C 和接头引物 D、DNA 聚合酶以及使用该试剂盒的说明书。

在第四方面，本发明提供了一种检测肝细胞癌的方法，包括如下步骤：1）获得来自人的 DNA 样品，所述 DNA 样品来自人的细胞、组织、血液、体液、尿液、唾液、排泄物或其组合，优选来自人的血浆或尿液游离 DNA；2）检测核酸序列 1~68 或序列 69~132 中至少 1 种或其组合是否存在异常甲基化；这些核酸序列的甲基化提示存在肝细胞癌。核酸序列 1~68 为前述 I 型肝细胞癌血浆差异甲基化 CpG 岛，核酸序列 69~132 为前述 II 型肝细胞癌血浆差异甲基化 CpG 岛。所用于检测甲基化的方法包括但不局限于 MCTA-Seq、甲基化特异性 PCR、限制性酶切-PCR、焦磷酸测序法、Methylight 和 MethylBeaming 等。

附图说明

图 1 各种 CpG 短串联核酸序列在人类基因组中的数目及在 CpG 岛区域的比例。

图 2 甲基化 CpG 短串联扩增与测序法（MCTA-Seq）示意图。引物 A 的下划直线部分指示独特分子识别标签序列部分，①、②和③指示不同的标签序列，注意不同的扩增位点也可用于分子识别。

图 3 MCTA-Seq 琼脂糖胶图。泳道 1 为未超声人类 WBC 基因组 DNA，泳道 2 为超声至平均长度为 200 碱基对的人类 WBC 基因组 DNA。M：DNA 碱基长度标准（maker，为 TAKARA 的 20bp ladder）；bp：碱基对；NTC：无模板对照。

图 4 MCTA-Seq 高通量测序文库的 Fragment Analyzer 检测结果。

图 5 MCTA-Seq 的测序结果与 CpG 短串联的相关性。X 轴显示含有不同个数 CGCGCGG 短串联的 CpG 岛，Y 轴显示人类基因组完全甲基化 DNA 标准品中的甲基化检测值（MePM：methylated alleles per million mapped reads，表示每一百万个回帖到基因组的读数中回帖至该 CpG 岛中读数）。插入的小图显示 MCTA-Seq 的 Read2 读数起始 3~13 个碱基对应的基因组序列，其中 5~11 对应 CGCGCGG 短串联序列；bits 表示序列的富集程度。

图 6 MCTA-Seq 的检测通量。人类基因组完全甲基化 DNA 标准品的 MCTA-Seq 的技术重复 1（X 轴）与技术重复 2（Y 轴）的甲基化值（MePM）。显示在两个技术重复的平均 MePM 大于 8 的 CpG 岛数目以及两个技术重复之间的相关系数。

图 7 不同基因组 DNA 样本中的甲基化 CpG 岛检测数量。FMG：人类基因组完全甲基化 DNA 标准品；HePG2：人类肝细胞癌细胞系；Hela：人类子宫颈癌细胞系；WBC：人类外周血白细胞。显示均值，误差棒为均值±标准差。

图 8 CDKN2A 基因区域的 MCTA-Seq 结果。灰框：CDKN2A (CYCLIN-DEPENDENT KINASE INHIBITOR 2A)基因区域中 4 个 CpG 岛，i1 和 i2 为两个基因间 CpG 岛，p 为启动子 CpG 岛；三角形：CGCGCGG 短串联序列，其中位于灰框上方的三角形指示正向链的 CGCGCGG，位于灰框下方的三角形指示反向链的 CGCGCGG；基因间 CpG 岛 i2 中有 2 个 CGCGCGG 短串联序列。标尺为 MePM，显示 HePG2、Hela 和 WBC 三种细胞中所有 MCTA-Seq 测序读数。

图 9 MCTA-Seq 的重复性与检测差异甲基化。WBC 与 HePG2 技术重复（replicate）的 MCTA-Seq 结果显示可重复性很好；比较 WBC 与 HePG2 的结果显示二者有大量差异甲基化位点，其中主要是 HePG2 中高甲基化的位点。

图 10 MCTA-Seq 的灵敏度。图下方显示不同 FMG 量稀释至 WBC 基因组，每个稀释实验有两个技术重复。右侧显示检测到一些基因的启动子 CpG 岛或基因间 CpG 岛（CDKN2A(i2)），热图中，白色表示高甲基化值，黑色表示低甲基化值，标尺上显示 $\log_2$(MePM)值。

图 11 VIM 基因区域的 MCTA-Seq 结果。灰框：VIM (VIMENTIN)的启动子 CpG 岛（序列号：2）；三角形：CGCGCGG 短串联序列，其中位于灰框上方的三角形指示正向链的 CGCGCGG，位于灰框下方的三角形指示反向链的 CGCGCGG。M（%）：甲基化频率；GE：单倍体基因组当量；rep1：技术重复 1； rep2: 技术重复 2 ；Reads：测序读数；UMIs：独特分子识别标签读数；箭头：最短扩增子（30-bp）。

图 12 VIM 启动子 CpG 岛的甲基化值与 FMG 比例的线性关系。误差棒为均值±标准差。

图 13 SEPT9 基因区域的 MCTA-Seq 结果。灰框：SEPT9 (SEPTIN9) 转录变异体 2 的启动子 CpG 岛；三角形：CGCGCGG 短串联序列，其中位于灰框上方的三角形指示正向链的 CGCGCGG，位于灰框下方的三角形指示反向链的 CGCGCGG；M（%）：甲基化频率；GE：单倍体基因组当量；rep1：技术重复 1； rep2: 技术重复 2 ；Reads：测序读数；UMIs：独特分子识别标签读数。

图 14 SEPT9 启动子 CpG 岛的甲基化值与 FMG 比例的线性关系。误差棒为均值±标准差。

图 15 独特分子识别标签序列减少扩增偏差。FMG7.5pg 稀释组的测序读数（a，MePM）与独特分子

识别标签读数（b）比较。uMePM：unique methylated alleles per million mapped reads，每一百万个回帖到基因组的读数中回帖至该 CpG 岛中独特分子识别标签读数。

图 16 组成分分析 MCTA-Seq 的肝细胞癌组织检测结果。

图 17 组织差异甲基化 CpG 岛。火山图中 X 轴显示肝细胞癌组织（27 例）与癌旁组织（27 例）的平均 MePM 值之比的 Log2 值，Y 轴显示 P 值（双尾 t 检验）的 FDR 校正值 q 值，共有 1,605 个 CpG 岛在肝细胞癌组织与癌旁组织之间显示差异甲基化（$q < 0.05$, 甲基化比值>2）。

图 18 甲基化 CpG 岛数目。箱线图显示在（左）肝细胞癌患者（P(HCC)，17 例）、肝硬化患者（P(Ci) ，11 例）和健康志愿者（P(N)，9 例）的血浆游离 DNA 以及（右）肝细胞癌组织（T(HCC)，27 例）、癌旁组织（T(Aj)，27 例）和正常肝组织（T(NL)，3 例）中的在不同 MePM 阈值下得甲基化 CpG 岛数目。双尾 MWW 检验。

图 19 血浆差异甲基化 CpG 岛。火山图中 X 轴显示肝细胞癌患者（13 例，肿瘤最大直径小于 5cm）血浆与无癌个体血浆（20 例，包含肝硬化患者与健康志愿者）的平均 MePM 值之比，Y 轴显示 P 值（双尾 t 检验）的 FDR 校正值 q 值，共有 635 个 CpG 岛在肝细胞癌患者血浆与无癌个体血浆之间显示差异甲基化（$q < 0.05$, 甲基化比值>2）。

图 20 维恩图比较组织差异甲基化 CpG 岛与血浆差异甲基化 CpG 岛。376 个 CpG 岛同时在组织和血浆中显示差异甲基化，称为 I 型血浆差异甲基化 CpG 岛，259 个 CpG 岛仅在血浆中显示差异甲基化，称为 II 型血浆差异甲基化 CpG 岛。

图 21 II 型血浆差异甲基化 CpG 岛分类。聚类分析显示 102 个 CpG 岛不仅在肝细胞癌中显示高甲基化，而且在癌旁与正常肝组织中也显示高甲基化（但在 WBC 中为低甲基化），称为 IIa 型血浆差异甲基化 CpG 岛；其余为 IIb 型。热图中白色为高甲基化，黑色为低甲基化，由于色彩原因仅显示部分 CpG 岛。

图 22 IIa 型血浆差异甲基化 CpG 岛的组织差异性甲基化。RRBS 结果显示 IIa 型血浆差异甲基化 CpG 岛 (94 个能被 RRBS 检测到)为在正常肝组织（T(NL)）特异性高甲基化，而在 WBC 中为低甲基化。双尾 MWW 检验。

图 23 CpG 岛 SHANK2(i2)结果。SHANK2 基因内的第二个 CpG 岛（SHANK2(i2)，序列号：79）是 IIa 型血浆差异甲基化 CpG 岛的一个例子。（a）RRBS 结果，T(NL)：正常肝组织组织，双尾 t 检验，误差棒为均值±标准差；（b）MCTA-Seq 组织检测结果，T(HCC)：肝细胞癌组织、T(Aj)：癌旁组织；（c）MCTA-Seq 血浆检测结果，P(HCC)：肝细胞患者血浆，P(Ci)：肝硬化患者血浆，P(N)：健康志愿者血浆。

图 24 高性能 I 型血浆差异甲基化 CpG 岛举例。68 个高 AUC 值（>0.9）I 型血浆差异甲基化 CpG 岛中 4 个——包括 VIM 启动子 CpG 岛（序列号：2）、RIMS 启动子 CpG 岛（序列号：64）、TBX2 启动子 CpG 岛（序列号：25）和 KCNQ4 基因内 CpG 岛（序列号：38）——的 ROC 曲线和血浆甲基化值的箱线图。

图 25 高性能 II 型血浆差异甲基化 CpG 岛举例。64 个高 AUC 值（>0.9）II 型血浆差异甲基化 CpG 岛中 4 个——包括 APAK2 启动子 CpG 岛（序列号：130）、SH3PXD2A 基因内 CpG 岛（序列号：70）、IGF2 基因内 CpG 岛（序列号：75）和 APBB2 启动子 CpG 岛（序列号：116）的 ROC 曲线和血浆甲基化值的

箱线图。

图 26 SEPT9 启动子 CpG 岛的 ROC 曲线，以及血浆和组织甲基化值的箱线图。

图 27 高性能 CpG 岛同步增高检验。超过 90%特异性阈值为阳性，132 个高性能 CpG 岛同时超过阈值的个数，其的 ROC 曲线和血浆甲基化值箱线图。

图 28 肝细胞癌的多位点检测模式。共设立 3 个指标，包括 1)超过 90%特异性阈值的 CpG 岛数目（上）；2）超过 90%阈值最大标准差（最大 M 值，中）；3）超过 90%阈值的标准差总和（总 M 值，底）；对 68 个 I 型血浆差异甲基化 CpG 岛（左）和 64 个 II 型血浆差异甲基化 CpG 岛（右）均检测这 3 个指标，共 6 个指标。对于每个指标，设立一个阈值（水平虚线），该阈值能够尽量肝细胞癌患者与无癌个体。区分超过阈值为阳性。两个检测标准，高特异性标准为 6 个指标中有 5 个或 6 个指标阳性；高灵敏度标准为 6 个指标中有 2 个或 3 个指标阳性，其中至少 1 个指标为 I 型血浆差异甲基化 CpG 岛。P(HCC)：肝细胞癌患者血浆，病例 1~17 为训练组，按肿瘤大小排序：1(P50), 2(P29), 3(P41), 4(P3), 5(P37), 6(P57), 7(P62), 8(P77), 9(P10), 10(P39), 11(P60), 12(P36), 13(P44), 14(P59), 15(P74), 16(P58), 17(P51)；病例 18~27 为测试组，按肿瘤大小排序：18(P92), 19(P119), 20(P109), 21(P101), 22(P114), 23(P85), 24(P115), 25(P80), 26(P81), 27(P113)。P(Ci)：肝硬化患者血浆，病例 28~38 为训练组，39~44 为测试组。P(N)：健康志愿者血浆，病例 45~53 为训练组，54~72 为测试组。

图 29~图 39 显示全部 132 个 I 型和 II 型高性能血浆差异甲基化 CpG 岛在全部 72 例研究对象血浆游离 DNA 中的 MCTA-Seq 检测结果，研究对象的排序同图 28。

图 40 总 M 值与肝细胞癌大小呈线性关系。I 型（左）和 II 型（右）血浆差异甲基化 CpG 岛的总 M 值均与肝细胞癌最大直径呈线性相关（显示公式与相关系数）。

图 41 比较甲胎蛋白血清学检测结果。热图显示采用 MCTA-Seq 对肝细胞癌患者（27 例）和肝硬化患者（17 例）血浆游离 DNA 进行检测的结果。上方热图显示 68 个高性能 I 型血浆差异甲基化 CpG 岛和 64 个高性能 II 型血浆差异甲基化 CpG 岛的 M 值；下方显示两种类型 CpG 岛的各 3 个检测指标。

图 42 不同 CpG 短串联检测 CDKN2A 基因区域。p：CDKN2A 的启动子 CpG 岛；i1 和 i2：CDKN2A 的两个基因内 CpG 岛；三角：CGCGCGG 短串联；箭头：CGGCGGCGG 短串联。

具体实施方式

以下结合具体实施方式和附图，对本发明作进一步说明。本领域技术人员可以对具体实施方案中所示的本发明作出多种变化和修改而不脱离广义描述时的本发明的精神和范围。

实施例 1: 根据本发明的甲基化 CpG 短串联扩增与测序法（MCTA-Seq）以如下方式发生（附图 2）。

1. 步骤一、亚硫酸氢盐处理 DNA 样品。

采用试剂盒 MethylCode™ Bisulfite Conversion Kit（Invitrogen），并按照制造商提供的说明书进行操作，具体步骤如下：

1.1 制备 CT 转换试剂（CT Conversion Reagent）溶液：从试剂盒中取出 CT 转换试剂，加入 900 μl 水、50 μl 重悬缓冲液和 300μl 稀释缓冲液，室温短时震荡混匀 10 分钟待其溶解，室温避光保存；

1.2 将 500 pg 至 500 ng DNA 样品共 20 μl 加入 PCR 管中；

1.3 将 130 μl CT 转换试剂溶液加入 DNA 样品，轻弹或用枪头吹打混匀；

1.4 将 PCR 管在一台热循环仪上进行如下程序：98 度----10 分钟，64 度----2.5 小时，4 度保存（不超过 20 小时）备用；

1.5 将 DNA 纯化柱放入收集管中，加入 600 μl 结合缓冲液；

1.6 将 1.4 步骤中的 DNA 样品加入结合结合缓冲液中，避盖上下颠倒混匀数次；

1.7 最大转速（>10,000g）离心 30 秒，弃过柱液；

1.8 加入 100 μl 洗涤缓冲液（已加乙醇），最大转速离心 30 秒，弃过柱液；

1.9 加入 200 μl Desulphonation 缓冲液，将纯化柱在室温站立 15~20 分钟；

1.10 最大转速离心 30 秒，弃过柱液；

1.11 加入 100 μl 洗涤缓冲液（已加乙醇），最大转速离心 30 秒，弃过柱液；

1.12 重复 1.11 一次，将纯化柱放置到一个新的 1.5ml 离心管中；

1.13 加入 10 μl 溶解缓冲液，最大转速离心 30 秒以洗脱 DNA。

2. 步骤二、引物 A 和 DNA 聚合酶进行扩增。

2.1 将步骤 1 中得到的 DNA 在一个 PCR 管中在冰上配置如下扩增反应体系：

内容	体积
DNA 样品与水	10.5 μl
NEBuffer 2	1.5 μl
dNTP (2.5 mM)	1.5 μl
引物 A (5 μM) *	1 μl
Klenow (exo-) **	0.5 μl
总体积	15 μl

*: 引物 A：为以下四种引物的等摩尔混合物，总浓度为 5 μM：

5’-TTTCCCTACACGACGCTCTTCCGATCTHHHHHHHHHCGCH-3’；

5’-TTTCCCTACACGACGCTCTTCCGATCTHHHHHHHHCGHCH-3’

5’-TTTCCCTACACGACGCTCTTCCGATCTHHHHHHHCGHHCH-3’,

5’-TTTCCCTACACGACGCTCTTCCGATCTHHHHHHCGHHHCH-3’;

其中 H=A/T/C，下划波浪线部分为引物的 3’端随机序列部分，下划直线部分为引物的 5’端锚定序列部分，下划双直线部分为引物的独特分子识别标签序列部分。

**: 在 2.3 步骤中加入

2.2 将 PCR 管放入 PCR 热循环仪中进行如下程序：95 度----2 分钟，4 度----暂停；

2.3 离心，4 度 13000 rpm 共 1 分钟；

2.4 在冰上加入 0.5 μl Klenow 片段(exo-) (NEB 目录号 M0212S)，震荡混匀；

2.5 离心，4 度 13000 rpm 共 1 分钟；

2.6 在 PCR 热循环仪中进行如下程序进行 1 轮扩增反应：4 度----50 秒，10 度----1 分钟，20 度----4 分钟，30 度----4 分钟，37 度----4 分钟；

2.7 在 PCR 热循环仪中进行如下程序以灭活 Klenow 片段：75 度----20 分钟，4 度----暂停；

2.8 离心，4 度 13000 rpm 共 1 分钟；

3. 步骤三、引物 B 和 DNA 聚合酶进行扩增。

3.1 在一个 PCR 管中在冰上配置如下扩增反应体系：

内容	体积
Ex Taq Buffer	0.5 μl
引物 B (10 μM) *	0.5 μl
热启动 HS Ex Taq (Takara 目录号 DRR006B)	0.3 μl
水	3.7 μl
总体积	5 μl

*：引物 B：(5’- GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTDDDDCGCGCGG-3’, D = A/T/G，其中下划波浪线部分为引物的 3’端 CpG 短串联序列部分，下划直线部分为引物的 5’端锚定序列部分，下划双直线部分为随机序列部分。

3.2 将上述混合物加入 2.8 的扩增反应产物中，震荡混匀；

3.3 离心，4 度 13000 rpm 共 1 分钟；

3.4 在 PCR 热循环仪中进行如下程序：95 度---3 分钟，50 度----2 分钟，72 度----1 分钟，4 度----暂停；

3.5 离心，4 度 13000 rpm 共 1 分钟；

4. 步骤四、接头引物 C、接头引物 D 和 DNA 聚合酶进行指数扩增。

4.1 在一个 PCR 管中在冰上配置如下扩增反应体系：

内容	体积
Ex Taq Buffer	3 μl
接头引物 C (100 uM) *	0.6 μl
接头引物 D (100 uM) **	0.6 μl
dNTP (2.5 mM)	3 μl
热启动 HS Ex Taq (Takara 目录号 DRR006B)	0.3 μl
水	22.5 μl
总体积	30 μl

*：接头引物 C：

5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3’其中下

划直线部分与引物 A 的 5'端部分相同；

**：接头引物 D：

5'-CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCT-3' 其中下划直线部分与引物 B 的 5'端部分相同，下划双直线处为标签序列（Illumina index 9）；

4.2 将混合物加入 3.5 的扩增反应产物中，震荡混匀；

4.3 离心，4 度 13000 rpm 共 1 分钟；

4.4 在 PCR 热循环仪中进行如下程序：95 度----3 分钟；

4.5 在 PCR 热循环仪中进行如下程序：95 度----30 秒，65 度----30 秒，72 度----1 分钟，共 17 个循环，4 度暂停。

5. 步骤五、片段选择、纯化、进行高通量测序和数据分析。

5.1 制备 4%琼脂糖凝胶，加入 1×SYBR Safe（Invitrogen）；

5.2 将 4.5 的扩增产物在 4%琼脂糖凝胶中进行电泳分离；

5.3 对凝胶中的 DNA 进行成像分析（图 3，结果显示超声的 DNA 样品与未超声样品相比有相同的扩增效率，说明 MCTA-Seq 能够扩增高度片段化的 DNA 样品）；

5.4 切胶回收 180-250bp 的目的片段；

5.5 将目的片段进行胶回收纯化（Qiagen，QIAquick Gel Extraction Kit），获得高通量测序文库；

5.6 用 Fragment Analyzer 分析系统（Agilent）检测高通量测序文库插入片段的大小（图 4，结果显示测序文库的片段长度介于 180-250bp 之间），并利用 QPCR 对文库的浓度进行绝对定量分析；

5.7 在 Illumina HiSeq2000/2500 测序仪上，按照读长为 100bp 的双末端测序，对文库进行高通量测序分析，获得大约 1 千万条原始测序数据；

5.8 数据分析：首先去除任何接头序列和低质量序列，然后应用 Bismark 软件将数据与参考人类基因组序列（Hg19）进行比对，并以此为基础进行后续的生物信息学分析。

实施例 2: MCTA-Seq 的检测通量、可重复性、灵敏度和独特分子识别标签。

采用 MCTA-Seq 对人类基因组完全甲基化 DNA 标准品（fully methylated human genomic DNA , FMG，Chemicon/Millipore 目录号 S7821, CpGenome Universal Methylated DNA）、人类外周血白细胞（human white blood cell, WBC）基因组 DNA 和两种肿瘤细胞株（HePG2 和 Hela 细胞株）基因组 DNA 进行检测。这些与随后实施例中的样本，平均每个样本检测 750 万条配对末端序列；通过掺入 Lamda DNA 序列估计亚硫酸氢盐转换平均效率为 98.7%（97.53~99.31%，表一）。

结果显示回帖序列主要起始至 CGCGCGG 短串联序列（图 5）。FMG 结果显示，9,393 个含有一个或以上 CGCGCGG 短串联序列的 CpG 岛中，8,748 个（93.3%）能够被 MCTA-Seq 以较高效率检测（每一百万回帖至基因组序列中回帖至该 CpG 岛的序列数大于 8，methylated alleles per million mapped reads, MePM>8，图 6）；检测效率与该 CpG 岛中的 CGCGCGG 数量呈正比（图 5）。在 WBC、HePG2 与 Hela 细

胞中，分别检测到2849、3726和3773个甲基化CpG岛（图7）。技术重复之间的相关性非常高，平均相关系数为0.974（图6与图9）。WBC与肿瘤细胞之间的差异甲基化CpG岛被清晰显示（图8与图9）。图8显示CDKN2A基因区域的MCTA-Seq结果：1）几乎全部测序读数均来自这3个CGCGCGG短串联序列，说明MCTA-Seq扩增具有高度特异性；2）这两个CpG岛只在HePG2与Hela细胞基因组中有强阳性扩增，而在WBC中几乎无扩增，说明该CpG岛为差异甲基化位点，被MCTA-Seq所显示。

为了确定MCTA-Seq的灵敏度，将FMG以一定比例（1:10, 1:20, 1:50, 1:100, 1:200和1:400）掺入WBC之中。结果显示MCTA-Seq可以检测到少至甲基化频率为0.25%的甲基化CpG岛（图10）。甲基化频率的检测下限可随着测序深度的增加而继续下移，其最终决定与对甲基化CpG岛绝对分子数量的检测下限。为了确定该值，进行了另一个稀释实验：将FMG的绝对数量进行系列的稀释（150, 60, 30, 15 and 7.5 皮克），而保持FMG:WBC的相对比例为1:100。结果显示MCTA-Seq可以检测到少至7.5皮克基因组DNA（约为2.5个单倍体人类基因组当量）。3,579个在WBC中无信号的CpG岛中，3,005（84%）在7.5皮克组两个技术重复之一被检测到，1,687 （47%）在两个技术重复中均被检测到（图10）。检测数目在15皮克组均明显上升，提示7.5皮克组的取样随机性。

以VIM（VIMENTIN）的启动子CpG岛（序列号：2）和SEPT9（SEPTIN9）的启动子CpG岛为例展示检测结果。这两个CpG岛中分别有2个和4个CGCGCGG短串联序列，扩增子几乎完全起始至短串联序列，最短扩增子只有30个碱基对（图11和图13），且甲基化程度检测值与稀释度在最低的稀释范围呈线性相关（图12和图14）。另外，结果表明，独特分子识别标签可以提高7.5皮克组（PCR过度扩增程度最高）的相关性（图15，图11和图13）。

实施例3：MCTA-Seq检测肝细胞癌组织。

获取27例肝细胞癌组织、27例配对的癌旁组织以及3例正常肝脏组织（表二），正常肝脏组织获取自手术治疗的肝血管瘤患者，所有个体的样本收集和后续分析均经过本人签字同意，研究获得首都医科大学附属北京世纪坛医院伦理委员会批准同意。对组织进行MCTA-Seq检测分析，组成分分析表明大部分肝细胞癌组织（27例中23例）能够与癌旁组织以及正常肝组织显著区分（图16）。共发现1,605个在肝细胞癌组织中异常高甲基化的CpG岛（组织差异甲基化CpG岛，t检验, FDR < 0.05, 平均甲基化程度差别2倍以上, 图17）。

实施例4：MCTA-Seq检测肝细胞癌患者血浆游离DNA。

获取肝细胞癌患者（训练组17例，测试组10例，临床信息见表三）、肝硬化患者（训练组11例，测试组6例）健康志愿者（训练组9例，测试组19例）的外周血，所有个体的样本收集和后续分析均经过本人签字同意，研究获得首都医科大学附属北京世纪坛医院伦理委员会批准同意。分离血浆：每个个体以EDTA抗凝管获取5ml外周静脉血，在室温放置6小时内分离获得血浆：1）、离心，室温，1350g共12分钟，将血浆转移至一个新的15ml离心管，注意不要吸到细胞；2）、离心，室温，1350g共12分钟，将血浆转移至新的1.5ml或2ml离心管，残余底部少量血浆不吸；3）、离心，室温，>10000g共5分钟以彻底

去除细胞成分，将血浆转移至新的 1.5ml 或 2ml 离心管，冻存于-80 度。游离 DNA 采用 QIAamp DNA Blood Midi 试剂盒抽提获得。

进行 MCTA-Seq 分析，在肝细胞癌患者血浆中检测到 3,465 个甲基化 CpG 岛（中位数，MePM > 1），显著高于无癌个体（中位数 2,837，P < 0.01, MWW 检验，图 18）。与组织样本的检测情况明显不同的是，肝细胞癌患者与无癌个体之间的差异随着 MePM 阈值的增加的逐渐减少；肝细胞癌组织与癌旁组织之间的差异随着 MePM 阈值的增加而逐渐增加（图 18）。这说明血浆游离 DNA 中的肿瘤相关的甲基化改变主要发生于低甲基化频率范围，显著低于组织中的甲基化改变。共发现 635 个在肝细胞患者血浆游离 DNA 中异常高甲基化的 CpG 岛（血浆差异甲基化 CpG 岛，T 检验, FDR < 0.05, 平均甲基化程度差别 2 倍以上, 图 19）。其中，376 个也在组织样本中显示肿瘤特异性高甲基化，称为 I 型血浆差异甲基化 CpG 岛；259 个未在组织样本中显示差异甲基化，称为 II 型血浆差异甲基化 CpG 岛（图 20）。聚类分析显示 II 型 CpG 岛又可分为两组，其中 102 个（称为 IIa 型）不仅在肝细胞癌组织中呈现高甲基化，在癌旁组织和正常肝组织中也呈现高甲基化；但在 WBC 中呈现低甲基化，表现为组织特异性甲基化的特征（图 21，表四）。RRBS 结果证实这些 CpG 岛在正常肝组织的甲基化水平显著高于 WBC（中位数 32.8%比 8%, *P* = 5E-18, MWW 检验, 图 22，表四）。IIa 型血浆差异甲基化 CpG 岛中，54 个（53%）为基因内 CpG 岛，与组织特异性甲基化 CpG 岛中较高比例为基因内 CpG 岛一致。IIa 型血浆差异甲基化 CpG 岛的一个例子是 SHANK2 的基因内 CpG 岛（序列号：79，图 23）

通过受试者工作曲线（Receiver Operating Characteristic，ROC）方法分析血浆差异甲基化 CpG 岛并计算曲线下值（Area Under the Curve，AUC）。635 个血浆差异甲基化 CpG 岛中，132 个具有高性能区分肝细胞癌与无癌个体的特征（AUC > 0.9）。其中 68 个为 I 型血浆差异甲基化 CpG 岛，64 个为 II 型血浆差异甲基化 CpG 岛，大多数（56 个）为 IIa 型。两种血浆差异甲基化 CpG 岛中的一些例子如图 24 和图 25 所示。它们在肝细胞患者及对照的组织与血浆中的 MCTA-Seq 检测结果见表四。

另外，以 SEPT9 为例显示在其他肿瘤中性能较好的差异甲基化 CpG 岛（如 SEPT9 在结直肠癌中）在肝细胞癌中由于发生率较低而 AUC 较差（SEPT9 为 0.771），提示不同肿瘤具有不同的甲基化模式，因此通过基因组规模检测差异甲基化 CpG 岛有可能通过甲基化模式区分不同部位与类型的肿瘤。

实施例 5：MCTA-Seq 通过血浆游离 DNA 检测肝细胞癌患者。

对于每个高 AUC 值血浆差异甲基化 CpG 岛，设定一个相当于 90%特异性的阈值（即对于 90%的无肿瘤个体，该 CpG 岛在血浆游离 DNA 中的甲基化水平位于该阈值以下），分析 132 个高性能 CpG 岛中超过 90%特异性阈值的 CpG 岛数目。结果显示在大多数肝细胞癌患者的血浆游离 DNA 中这些 CpG 岛同时释放入血，而在大多数无癌个体中，它们不表现出这种同步模式。超过 90%阈值的 CpG 岛数目这一检测指标将检测特异性从每个 CpG 岛的 90%增加至 100%，AUC 值为 0.982（95%CI：0.952-1，图 28）。

在该检测指标的基础上，进一步增加定量信息，最终建立肝细胞癌多位点检测模式（图 29）。对每个 CpG 岛，将超过 90%特异性阈值的标准差个数设为 M 值，然后计算最大 M 值和总 M 值。最大 M 值提供最显著异常高甲基化 CpG 岛的信息，通过极值确定指标是否为阳性，因此被期望能够将检测的灵敏度最大化。总 M 值用于衡量异常高甲基化总体水平。因此，一共有三个指标：超过阈值的位点数目，最大 M 值

和总 M 值。对于每个指标，根据无癌患者中的最大值定义阳性阈值（超过阈值的 CpG 岛位点数目≥30，最大 M 值≥5，总 M 值≥10），并分别分析 I 型和 II 型血浆差异甲基化 CpG 岛；因此，共设立 6 个指标。进一步，建立两类标准：高特异性标准为 5 个或 6 个指标为阳性，高灵敏度标准为 3 个或 4 个指标为阳性，其中至少 1 个来自 I 型 CpG 岛。

采用该检测模式分析训练组。高特异性标准的灵敏度为 88% (15/17) ,特异性 100%(1/20)，高灵敏度标准灵敏度为 94% (16/17) ，特异性 95%(1/20)。进一步以双盲方式分析测试组共 35 例血浆样本，包括 10 例肝细胞癌、6 例肝硬化和 19 例健康志愿者（表二、三）。两种标准均鉴定出 9 例肝细胞癌患者（灵敏度为 90%），1 例肝硬化患者在高特异性标准显示阳性（特异性 96%），另增 1 例健康志愿者在高灵敏标准显示阳性 (特异性 92%)。总体灵敏度在高灵敏标准为 93%，在高特异性标准为 89%；总体特异性在在高灵敏标准为 96%，在高特异性标准为 98%。图 29~图 39 显示全部 132 个 I 型和 II 型高性能血浆差异甲基化 CpG 岛在全部 72 例研究对象血浆游离 DNA 中的 MCTA-Seq 检测结果。

进一步分析 MCTA-Seq 结果与肿瘤大小的关系：I 型和 II 型位点的总 M 值均与肿瘤大小呈线性相关(I 型: $r = 0.656$; II 型: $r = 0.508$, 图 40). 进一步比较 MCTA-Seq 结果与血清 AFP 检测结果。27 例肝细胞癌患者中，有 10 例为 AFP 阴性结果（<20 ng/mL，图 41，表三）。这 10 例患者中有 9 例能被 MCTA-Seq 通过高特异性标准检测到，而通过高灵敏度标准，10 例均能被检测到（图 42）。

实施例 6：靶向 CGCGCGG 以外 CpG 短串联序列的 MCTA-Seq。

以 CDKN2A 基因区域为例显示靶向不同 CpG 短串联核酸序列可以增加检测的范围（图 32）。CDKN2A 的启动 CpG 岛中不含 CGCGCGG 短串联序列，因此在上述采用 CGCGCGG 为引物的 MCTA-Seq 中不能被检测到。但是其含有 1 个 CGGCGGCGG 位点，采用 CGGCGGCGG 或 CGCGGCGG 短串联序列为引物，可以以很高的效率检测到该 CpG 岛。

表一、高通量测序信息。

名称	分组	样本数	原始读数(双端配对，百万，均值±标准差)	唯一回帖读数(百万，均值±标准差)	亚硫酸氢盐转换效率(%，均值±标准差)
	FMG, WBCs, HePG2, Hela 和稀释实验	30	7.9 ± 2.8	3.2 ± 1.2	98.86 ± 0.19
T(HCC)	组织	27	6.3 ± 2.3	2.1 ± 1.0	98.80 ± 0.18
T(Aj)	组织	27	6.7 ± 3.0	2.2 ± 0.8	98.63 ± 0.47
T(NL)	组织	3	5.3 ± 0.8	2.8 ± 0.3	98.90 ± 0.16
P(HCC)-tr	血浆，训练组	17	8.4 ± 2.4	3.0 ± 0.7	98.70 ± 0.33
P(HCC)-te	血浆，测试组	10	9.6 ± 3.1	3.8 ± 1.1	98.53 ± 0.42
P(Ci)-tr	血浆，训练组	11	7.1 ± 0.8	2.6 ± 0.3	98.34 ± 0.24
P(Ci)-te	血浆，测试组	5	9.4 ± 1.4	3.5 ± 0.7	98.79 ± 0.30

P(N)-tr	血浆，训练组	9	7.4 ± 3.1	2.9 ± 1.0	98.65 ± 0.54
P(N)-te	血浆，测试组	20	8.9 ± 2.1	3.1 ± 0.8	98.68 ± 0.40

T(HCC)：肝细胞癌组织；T(Aj)：癌旁组织；T(NL)：正常肝组织；P(HCC)：肝细胞癌患者血浆；P(Ci)：肝硬化患者血浆；P(N)：健康志愿者血浆； tr：训练组；te：测试组。

表二、肝细胞癌患者与对照个体的临床信息。

	分组	年龄 (岁，均值 ±标准差)	性别		肿瘤大小 (cm)		
			男（例数/百分比）	女（例数/百分比）	> 5	3~5	≤ 3
T(HCC)	组织	59 ± 9	22(81%)	5(19%)	10(37%)	6(22%)	11(41%)
T(Aj)	组织	59 ± 9	22(81%)	5(19%)	10(37%)	6(22%)	11(41%)
T(NL)	组织	55 ± 8	1(33%)	2(67%)			
P(HCC)-tr	血浆，训练组	61 ± 10	15(88%)	2(12%)	4(24%)	7(41%)	6(35%)
P(HCC)-te	血浆，测试组	63 ± 11	8(80%)	2(20%)	4(40%)	3(30%)	3(30%)
P(Ci)-tr	血浆，训练组	56 ± 11	8(73%)	3(27%)			
P(Ci)-te	血浆，测试组	50 ± 8	4(75%)	1(25%)			
P(N)-tr	血浆，训练组	62 ± 15	5(56%)	4(44%)			
P(N)-te	血浆，测试组	55 ± 17	16(80%)	4(20%)			

表三、 肝细胞癌患者临床与病理数据。

编号	分组	年龄	性别	乙肝病毒感染	结节类型	肿瘤大小	是否有血管侵犯	TNM 分期	AFP (ng/mL)
p50	训练组	68	男	Yes	Satellites	15	yes	T3aN0M0	>1210
p92	测试组	77	男	No	Satellites	15	yes	T3bN0M0	>1210
p29	训练组	56	男	No	Unique	11.2	yes	T4N0M0	>1210
p119	测试组	50	男	Yes	Unique	11	yes	T3bN0M0	4.41
p41	训练组	57	男	No	Satellites	8	No	T3aN0M0	59
p109	测试组	74	男	No	Unique	8	No	T1N0M0	2.61
p101	测试组	76	女	Yes	Unique	7	yes	T2N0M0	>1210
p3	训练组	42	男	Yes	Satellites	5.6	No	T3aN0M0	78.5
p37	训练组	64	男	Yes	Unique	5	No	T1N0M0	675.9
p114	测试组	39	男	Yes	Satellites	5	yes	T4N0M0	>1210
p85	测试组	63	男	Yes	Unique	4.5	yes	T2N0M0	16.2
p57	训练组	76	女	No	Unique	4	yes	T2N0M0	2.29
P62	训练组	51	男	Yes	Unique	4	yes	T4N0M0	20.23

P77	训练组	61	男	Yes	Unique	4	yes	T4N0M0	>1210
p10	训练组	56	男	Yes	Satellites	3.5	No	T2N0M0	768.5
p39	训练组	72	男	Yes	Satellites	3.5	No	T2N0M0	3.65
p60	训练组	70	男	Yes	Unique	3.5	yes	T3bN0M0	446.5
p115	测试组	60	男	Yes	Unique	3.3	No	T1N0M0	27.4
p36	训练组	44	男	Yes	Unique	3	No	T1N0M0	5.74
p44	训练组	63	男	Yes	Satellites	3	yes	T4N0M1	>1210
p59	训练组	73	男	No	Unique	3	No	T1N0M0	>1210
P74	训练组	68	女	No	Unique	3	yes	T2N0M0	12.6
p80	测试组	70	男	Yes	Unique	3	No	T1N0M0	36.8
p81	测试组	61	女	No	Unique	2.8	No	T1N0M0	8.23
p113	测试组	63	男	No	Unique	2.1	yes	T4N0M0	126.1
p58	训练组	60	男	No	Unique	2	No	T1N0M0	3.64
p51	训练组	51	男	Yes	Unique	1.5	No	T1N0M0	5.12

表四、I 型与 II 型血浆差异甲基化 CpG 岛的 MCTA-Seq 检测数据。

序列号	基因简称	类型	AUC	T(HCC) MePM	T(Aj) MePM	T(NL) MePM	T(WBC) MePM	P(HCC) uMePM	P(Con) uMePM	NL RRBS (%)	WBC RRBS (%)
1	INTER1	I	0.94	95(106)	32(17)	34(13)	1(1)	10(11)	1(1)	41(16)	31(7)
2	VIM	I	0.96	52(47)	4(6)	6(3)	0(0)	3(6)	0(0)	3(1)	1(0)
3	ST8SIA6	I	0.96	88(111)	5(6)	2(0)	0(0)	5(8)	0(0)	4(1)	1(0)
4	LOC 100130992	I	0.92	113(93)	33(15)	22(7)	1(0)	17(14)	2(2)	4(2)	1(0)
5	RET(i)	I	0.93	56(72)	8(5)	11(3)	0(0)	6(8)	1(1)	4(1)	1(0)
6	RASGEF1A(i)	I	0.96	75(83)	11(9)	9(5)	1(0)	6(7)	0(0)	5(3)	1(0)
7	INTER2	I	0.91	21(33)	3(2)	1(1)	0(0)	2(3)	0(0)	1(2)	3(2)
8	INTER3	I	0.92	43(36)	16(7)	16(8)	1(1)	12(8)	2(2)	21(13)	10(3)
9	CACNA1C	I	0.92	104(102)	18(9)	30(17)	0(0)	14(14)	1(1)	10(6)	1(1)
10	FAR2	I	0.93	46(48)	4(2)	3(1)	0(0)	7(10)	1(1)	7(1)	2(1)
11	EFCAB4B	I	0.95	124(112)	13(16)	14(3)	0(0)	12(14)	0(1)	30(1)	1(1)
12	KCNH3(i)	I	0.96	100(65)	43(17)	24(6)	1(0)	14(12)	2(1)	29(4)	9(8)
13	GRASP	I	0.92	36(35)	8(12)	1(0)	0(0)	4(5)	0(0)	8(2)	1(0)
14	POU4F1-AS1	I	0.9	135(115)	11(7)	4(1)	1(0)	10(21)	1(1)	8(1)	3(1)
15	PARP6	I	0.91	75(94)	5(6)	7(5)	0(0)	7(11)	0(0)	9(3)	0(0)

16	ISL2(i)	I	0.93	66(55)	20(11)	23(7)	1(0)	14(15)	2(2)	13(2)	3(0)
17	TM6SF1	I	0.92	23(25)	2(1)	0(0)	0(0)	4(7)	0(0)	8(6)	2(0)
18	PRKCB	I	0.94	72(81)	5(5)	2(1)	0(0)	4(8)	0(0)	3(1)	0(0)
19	MMP25	I	0.91	32(53)	3(3)	2(1)	0(0)	3(3)	1(0)	4(2)	2(1)
20	INTER4	I	0.97	46(35)	22(13)	6(5)	1(0)	5(5)	1(0)	48(13)	3(0)
21	SSH2(i)	I	0.93	162(187)	23(33)	25(11)	0(0)	13(15)	1(1)	15(9)	0(0)
22	INTER5	I	0.9	61(70)	14(7)	13(2)	2(0)	10(8)	3(2)	19(12)	6(1)
23	SLC25A39	I	0.92	109(143)	14(11)	28(3)	0(1)	9(17)	0(1)	11(1)	1(0)
24	MAP3K14(i)	I	0.93	67(94)	18(10)	23(1)	9(6)	21(14)	6(3)	1(1)	1(0)
25	TBX2	I	0.94	577(487)	181(50)	117(22)	7(1)	107(86)	14(6)	19(5)	6(1)
26	INTER6	I	0.93	61(57)	30(17)	29(6)	1(0)	14(11)	4(2)	43(28)	ND
27	ACTG1	I	0.96	141(144)	64(45)	110(33)	1(1)	22(24)	2(3)	6(1)	1(0)
28	SLC16A3	I	0.91	42(55)	9(6)	5(3)	1(0)	3(4)	1(1)	12(0)	5(1)
29	PIK3R5	I	0.98	41(36)	19(9)	12(5)	0(0)	7(6)	1(1)	10(2)	1(0)
30	NEDD4L(i)	I	0.96	82(68)	23(22)	26(4)	0(0)	9(7)	1(1)	24(13)	3(3)
31	NFATC1	I	0.91	33(43)	10(3)	4(1)	3(2)	4(6)	1(1)	23(1)	20(10)
32	MATK	I	0.94	53(53)	7(5)	6(6)	0(0)	5(6)	1(1)	5(1)	2(0)
33	PTCHD2	I	0.9	36(52)	5(4)	3(1)	1(0)	3(3)	1(0)	3(1)	2(0)
34	KIF21B	I	0.95	60(76)	6(5)	8(2)	0(0)	6(9)	0(0)	10(5)	1(0)
35	PPFIA4(i)	I	0.95	53(40)	22(10)	15(5)	0(0)	12(9)	1(1)	35(2)	3(1)
36	INTER7	I	0.91	61(103)	5(6)	2(1)	1(0)	6(8)	1(1)	5(1)	2(2)
37	INTER8	I	0.98	94(77)	47(23)	87(32)	0(0)	8(6)	1(1)	18(9)	3(1)
38	KCNQ4(i)	I	0.95	60(46)	27(13)	17(5)	0(0)	14(15)	0(0)	35(10)	14(13)
39	ERG	I	0.91	57(40)	19(7)	16(2)	1(1)	9(7)	2(2)	7(2)	5(1)
40	INTER9	I	0.91	72(48)	28(15)	16(3)	0(1)	8(6)	3(2)	9(1)	7(2)
41	PAX8(i)	I	0.9	147(99)	32(15)	10(4)	1(0)	9(8)	1(1)	27(4)	10(2)
42	PTPN18(i)	I	0.97	119(104)	26(31)	9(5)	0(0)	9(11)	1(1)	12(3)	2(0)
43	SP9	I	0.94	145(141)	32(11)	35(11)	3(0)	13(10)	3(2)	9(2)	3(0)
44	NGEF	I	0.93	62(49)	26(22)	7(1)	1(0)	7(8)	1(1)	33(7)	10(1)
45	SYN2	I	0.94	75(56)	31(13)	32(9)	2(0)	16(11)	3(2)	7(0)	6(2)
46	GNB4	I	0.91	61(44)	17(10)	15(3)	0(0)	9(10)	0(1)	11(15)	1(1)
47	BDH1	I	0.92	95(69)	43(25)	92(33)	0(0)	10(10)	1(2)	26(6)	1(0)
48	LOC285548	I	0.93	29(42)	5(4)	2(1)	0(0)	3(3)	1(1)	6(2)	3(1)
49	INTER10	I	0.96	52(36)	7(6)	7(3)	0(0)	4(3)	0(0)	19(13)	7(8)
50	CPLX1(i)	I	0.92	113(69)	41(34)	62(18)	2(0)	13(10)	3(2)	57(2)	33(8)
51	ADRA1B(i)	I	0.93	63(63)	21(9)	21(9)	0(0)	10(9)	1(1)	43(26)	5(0)
52	GDNF(i)	I	0.92	47(38)	16(12)	16(7)	1(1)	7(6)	1(1)	26(1)	5(1)
53	ULBP3	I	0.92	51(55)	17(16)	18(10)	0(0)	9(10)	1(1)	1(1)	1(0)

54	INTER11	I	0.95	65(65)	19(9)	14(6)	3(0)	15(11)	3(2)	12(1)	6(1)
55	INTER12	I	0.94	215(127)	67(33)	24(17)	1(1)	24(25)	2(2)	18(1)	3(1)
56	TRIM15(i)	I	0.93	234(176)	116(47)	95(13)	4(1)	40(32)	6(4)	43(4)	12(1)
57	INTER13	I	0.91	209(163)	77(43)	51(31)	2(1)	16(15)	3(3)	17(4)	9(2)
58	SND1	I	0.95	249(175)	76(50)	55(14)	0(1)	20(21)	2(2)	55(2)	2(0)
59	INTER14	I	0.92	38(58)	1(1)	1(1)	0(0)	5(7)	0(0)	6(0)	2(1)
60	INTER15	I	0.92	75(55)	23(10)	25(5)	4(2)	11(5)	5(2)	12(3)	7(4)
61	AKR1B1	I	0.94	136(125)	12(10)	17(2)	0(1)	9(12)	0(0)	5(1)	0(0)
62	INTER16	I	0.91	28(36)	7(4)	11(1)	0(0)	3(3)	0(0)	9(0)	1(0)
63	HOXA10 -HOXA9(i)	I	0.9	61(66)	7(4)	3(0)	0(0)	6(6)	1(1)	9(0)	2(1)
64	RIMS2	I	0.91	146(159)	7(4)	3(1)	0(0)	6(9)	0(0)	7(3)	3(0)
65	TDH(i)	I	0.91	39(53)	4(3)	2(1)	0(0)	5(13)	1(1)	8(1)	2(1)
66	TRPS1(i)	I	0.92	314(230)	84(40)	133(23)	4(1)	31(26)	5(3)	21(29)	2(3)
67	FAM110B(i)	I	0.92	171(134)	32(21)	29(8)	1(0)	17(16)	2(2)	96(0)	92(1)
68	INTER17	I	0.91	33(46)	6(3)	6(3)	0(1)	6(7)	1(1)	5(2)	3(1)
69	TLX1NB(i)	IIa	0.92	76(39)	48(28)	57(4)	0(0)	12(9)	2(1)	34(2)	4(2)
70	SH3PXD2A(i)	IIa	0.95	40(20)	33(18)	27(7)	0(0)	6(5)	1(1)	68(7)	1(0)
71	GSTO2	IIa	0.98	118(93)	78(45)	92(35)	0(0)	17(16)	1(1)	28(2)	0(0)
72	FAM196A	IIa	0.92	136(106)	81(26)	69(13)	2(1)	21(15)	6(3)	10(2)	5(2)
73	ZMIZ1(i)	IIa	0.91	151(84)	129(59)	97(41)	0(0)	22(18)	4(3)	33(6)	3(0)
74	INTER18	IIa	0.91	106(87)	111(54)	49(11)	2(2)	19(13)	4(2)	51(2)	10(0)
75	IGF2(i)	IIa	0.98	40(29)	45(21)	72(16)	0(0)	10(7)	1(1)	49(14)	4(3)
76	SLC1A2	IIb	0.92	41(39)	23(8)	16(7)	1(0)	11(11)	1(1)	11(2)	3(1)
77	SCT	IIa	0.98	47(35)	43(18)	31(5)	1(1)	21(19)	2(2)	36(14)	11(5)
78	MACROD1(i)	IIa	0.91	156(127)	79(35)	48(19)	7(0)	24(15)	8(4)	34(3)	11(1)
79	SHANK2(i)	IIa	0.94	147(71)	140(63)	158(37)	1(0)	26(19)	3(4)	82(7)	11(3)
80	FIGNL2(i)	IIa	0.91	99(83)	72(30)	60(12)	4(1)	31(25)	10(8)	32(6)	4(0)
81	LAG3(i)	IIa	0.92	65(36)	59(22)	53(9)	2(1)	11(7)	3(2)	55(33)	8(6)
82	INTER19	IIa	0.93	284(245)	176(80)	183(65)	2(0)	52(45)	8(5)	37(52)	ND
83	INTER20	IIa	0.94	315(197)	245(73)	179(37)	2(1)	93(65)	8(5)	0(0)	0(0)
84	RAI1(i)	IIa	0.92	127(91)	99(55)	157(39)	4(0)	26(20)	9(4)	54(1)	28(1)
85	ULK2	IIb	0.92	39(80)	7(8)	3(3)	0(0)	3(4)	0(1)	9(4)	8(0)
86	SARM1	IIa	0.96	139(100)	86(40)	68(17)	1(0)	18(17)	3(3)	14(3)	1(0)
87	RAP1GAP2(i)	IIb	0.9	11(19)	8(4)	2(0)	0(0)	3(3)	1(0)	6(1)	2(0)
88	WNK4(i)	IIa	0.96	70(47)	65(30)	57(9)	7(4)	15(9)	5(3)	ND	3(4)

89	VMO1	IIa	0.91	80(54)	110(39)	118(31)	7(2)	26(18)	6(4)	87(19)	ND
90	ERN1	IIa	0.93	153(188)	71(46)	115(21)	2(0)	14(14)	2(2)	2(0)	1(0)
91	TMC8(i)	IIb	0.97	17(22)	22(8)	16(6)	0(0)	11(9)	0(0)	21(3)	1(0)
92	INTER21	IIa	0.91	120(91)	101(33)	106(13)	13(2)	18(10)	7(4)	35(11)	8(1)
93	NFIX(i)	IIa	0.97	68(38)	72(31)	99(21)	3(1)	24(16)	4(2)	55(10)	25(11)
94	NFIX(i)	IIa	0.94	49(54)	62(37)	49(30)	0(0)	12(7)	2(1)	48(7)	18(1)
95	LINGO3(i)	IIa	0.92	50(60)	32(14)	49(5)	9(0)	21(16)	5(4)	9(2)	7(3)
96	NFKBID(i)	IIa	0.96	75(54)	43(28)	73(21)	0(0)	12(11)	1(1)	ND	ND
97	RCN3(i)	IIa	0.91	127(90)	73(50)	55(12)	0(0)	13(14)	1(2)	38(11)	1(1)
98	ESPNP(i)	IIa	0.94	283(172)	159(105)	151(50)	2(0)	34(32)	6(4)	30(1)	15(3)
99	NR5A2(i)	IIa	0.91	165(94)	140(74)	292(88)	7(1)	32(21)	6(4)	38(4)	11(0)
100	PRDM16	IIa	0.91	150(80)	152(35)	93(19)	6(1)	36(21)	10(7)	18(1)	13(2)
101	SH3D21(i)	IIa	0.92	132(88)	71(21)	54(17)	17(2)	42(23)	13(6)	42(6)	44(8)
102	AGRN(i)	IIa	0.93	149(132)	92(41)	30(8)	1(1)	26(17)	5(3)	42(5)	18(3)
103	RUNX1(i)	IIb	0.92	11(10)	11(4)	14(6)	0(0)	3(2)	0(0)	47(13)	0(0)
104	TMPRSS6(i)	IIa	0.92	185(101)	173(62)	171(32)	8(3)	22(16)	6(4)	50(4)	11(3)
105	WNT7B(i)	IIa	0.93	36(47)	51(30)	9(7)	4(1)	16(13)	3(2)	13(2)	13(0)
106	ACOXL	IIb	0.93	20(12)	11(5)	2(2)	0(0)	3(3)	0(0)	7(0)	1(0)
107	INTER22	IIa	0.94	105(58)	99(56)	97(18)	2(1)	21(13)	3(2)	63(10)	13(4)
108	SP5	IIa	0.93	113(53)	82(36)	98(12)	4(1)	21(12)	5(3)	14(4)	5(2)
109	IHH(i)	IIa	0.91	234(110)	184(64)	236(65)	7(3)	39(26)	9(6)	31(3)	5(1)
110	INTER23	IIb	0.94	44(57)	21(12)	17(15)	1(0)	4(3)	1(1)	9(8)	1(0)
111	ABCG5(i)	IIa	0.96	113(63)	98(39)	119(22)	1(0)	14(9)	2(2)	29(12)	3(3)
112	NR1I2(i)	IIa	0.94	192(96)	145(102)	125(33)	0(0)	21(15)	2(2)	64(51)	0(0)
113	GRK7(i)	IIa	0.94	95(60)	71(28)	59(16)	0(0)	19(15)	2(2)	ND	ND
114	KIF15	IIb	0.92	17(28)	44(31)	94(28)	0(0)	8(9)	1(1)	13(1)	1(0)
115	ZMYND10	IIa	0.94	82(48)	42(23)	31(10)	0(0)	9(8)	1(1)	26(19)	2(0)
116	APBB2	IIa	0.93	163(99)	113(72)	123(12)	1(1)	37(28)	6(4)	43(13)	0(0)
117	SHROOM1(i)	IIa	0.9	126(53)	109(36)	196(31	10(	28(16)	9(4)	68(3)	17(0)

						)	1)				
118	NRG2(i)	IIb	0.96	17(23)	7(5)	2(1)	0(0)	5(4)	0(0)	4(4)	1(0)
119	PCDHGA1(i)	IIa	0.91	53(35)	34(11)	31(10)	0(1)	11(7)	3(2)	42(9)	13(3)
120	F12(i)	IIa	0.96	165(111)	114(56)	122(21)	0(0)	21(19)	2(2)	41(22)	1(1)
121	MIR548AO(i)	IIa	0.93	543(256)	324(100)	343(61)	15(1)	101(65)	27(12)	44(2)	12(5)
122	SOBP(i)	IIa	0.9	127(64)	105(33)	76(5)	4(2)	28(16)	8(3)	51(8)	17(0)
123	ESR1(i)	IIa	0.91	69(42)	55(32)	97(25)	1(1)	11(8)	1(2)	47(1)	2(0)
124	TFR2(i)	IIa	0.96	82(48)	64(29)	40(13)	1(0)	14(17)	2(2)	34(9)	1(1)
125	PRKAG2(i)	IIa	0.94	160(145)	102(51)	75(3)	1(0)	24(18)	3(2)	35(39)	2(3)
126	SHH(i)	IIa	0.91	261(122)	175(66)	137(33)	4(2)	27(15)	8(4)	39(2)	9(2)
127	GATA4(i)	IIa	0.96	124(109)	165(60)	234(43)	2(1)	40(31)	6(3)	33(8)	11(4)
128	OPLAH(i)	IIa	0.97	919(691)	479(202)	404 (145)	9(2)	105(93)	13(6)	27(3)	6(2)
129	GDF6(i)	IIb	0.94	29(45)	21(12)	29(16)	0(0)	11(10)	1(1)	25(1)	7(1)
130	AKAP2	IIa	0.93	125(87)	65(48)	82(4)	0(0)	22(21)	2(3)	56(15)	1(1)
131	SARDH(i)	IIa	0.92	163(119)	144(76)	227(44)	7(1)	13(5)	5(3)	34(8)	6(2)
132	ARID3C(i)	IIa	0.92	63(58)	44(24)	56(12)	0(0)	14(12)	2(2)	46(26)	5(6)

注释：MePM 与 uMePM 值：均值（标准差）。T(HCC)：肝细胞癌组织；T(Aj)：癌旁组织；T(NL)：正常肝组织； P(HCC)：肝细胞癌患者血浆；P(Con)：无癌个体（包括肝硬化患者与健康志愿者）血浆。

以上所述仅为本发明的一种具体实施方式，但本发明的保护范围并不局限于此，任何熟悉该技术的人在本发明揭露的技术范围以内，可以轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求书的保护范围为准。

权 利 要 求 书

1. 一种基因组规模检测与人体异常状态相关的差异甲基化 CpG 岛的方法，包括如下步骤：1）获得来自人的 DNA 样品；2）以基因组规模检测所述 DNA 样品中 CpG 短串联核酸序列的甲基化程度；3）将所述 CpG 短串联核酸序列在所述 DNA 样品中的甲基化程度与其在来自正常人群的 DNA 样品中的甲基化程度进行比较，鉴定与人体异常状态相关的差异甲基化 CpG 岛；

其中所述 DNA 样品来自人的细胞、组织、血液、体液、尿液、唾液、排泄物或其组合；优选来自人的血浆或尿液游离 DNA；

其中所述基因组规模是指同时检测基因组中 100 个以上 CpG 岛。

2. 根据权利要求 1 所述的方法，其特征在于，所述差异甲基化 CpG 岛被用于确定相应人体异常状态。

3. 根据权利要求 1~2 任一项所述的方法，其特征在于，所述 CpG 短串联核酸序列是指长度等于 7 个核苷酸并包含 3 个 CpG 核苷酸对的核酸序列。

4. 根据权利要求 1~2 任一项所述的方法，其特征在于，所述 CpG 短串联核酸序列是指长度等于 8 个核苷酸并包含 3 个或以上 CpG 核苷酸对的核酸序列。

5. 根据权利要求 1~2 任一项所述的方法，其特征在于，所述 CpG 短串联核酸序列是指长度等于 9 个核苷酸并包含 3 个或以上 CpG 核苷酸对的核酸序列。

6. 根据权利要求 1~2 任一项所述的方法，其特征在于，所述 CpG 短串联核酸序列是指核酸序列 CGCGCGG、CGCGCGA、CGCGCGT、CGCGCGC、CGGCGCGG、CGGCGCGA、CGGCGCGT、CGGCGCGC、CGCGGCGG、CGCGGCGA、CGCGGCGT、CGCGGCGC、CGGCGGCGG、CGGCGGCGA、CGGCGGCGT、CGGCGGCGC 中的 1 种或其组合；优选至少 1 种为 CGCGCGG。

7. 根据权利要求 1~6 任一项所述的方法，其特征在于，所述人体异常状态为肿瘤。

8. 根据权利要求 1~6 任一项所述的方法，其特征在于，所述 DNA 样品来自人的血浆或尿液游离 DNA；其中所述差异甲基化 CpG 岛如序列 1~68 所示的核酸序列；这些核酸序列的异常高甲基化提示肝细胞癌。

9. 根据权利要求 1~6 任一项所述的方法，其特征在于，所述 DNA 样品来自人的血浆或尿液游离 DNA；其中所述差异甲基化 CpG 岛如序列 69~132 所示的核酸序列；这些核酸序列的异常高甲基化提示肝细胞癌或肝组织被破坏的异常状态。

10. 根据权利要求 1~9 任一项所述的方法，其特征在于，采用甲基化 CpG 短串联扩增与测序法进行检测，包括下列步骤：

步骤一、用修饰剂处理上述 DNA 样品，以便将 DNA 样品中的胞嘧啶转换为尿嘧啶，而 5’甲基胞嘧啶不变，获得经过转换的 DNA 片段；

步骤二、将所述经过转换的 DNA 片段用引物 A 和 DNA 聚合酶进行至少 1 次线性扩增，以便获得一端能够锚定接头引物 C 的半扩增子；

其中引物 A 由 3’端和 5’端两部分组成：3’端部分用于结合和扩增所述经过转换的 DNA 片段，其特征为，一组能结合所述经过转换的 DNA 片段的随机序列；5’端部分用于锚定接头引物 C，其特征为，其反向互补序列能够被接头引物 C 结合并进行聚合酶链式反应扩增；优选的，所述引物 A 的 3’端部

分特征为，除了 CpG 以外，只包含 C、A 和 T；

步骤三、将所述半扩增子用引物 B 和 DNA 聚合酶对甲基化 CpG 短串联核酸序列进行选择性扩增并获得两端能够分别锚定接头引物 C 和接头引物 D 的全扩增子；

其中引物 B 由 3’端和 5’端两部分组成：3’端部分用于选择性扩增甲基化 CpG 短串联核酸序列，其即为所述 CpG 短串联核酸序列；5’端部分用于锚定接头引物 D，其特征为，其反向互补序列能够被接头引物 D 结合并进行聚合酶链式反应扩增；

步骤四、将所述全扩增子用接头引物 C、接头引物 D 和 DNA 聚合酶进行聚合酶链式反应指数扩增，以便获得甲基化 CpG 短串联核酸序列的扩增产物；

步骤五、将所述扩增产物进行分离纯化，所述扩增产物构成高通量测序文库；以及，将所述高通量测序文库上机测序及数据分析。

11. 根据权利要求 10 所述的方法，其特征在于，所述修饰剂为亚硫酸氢盐。

12. 根据权利要求 10~11 任一项所述的方法，其特征在于，所述引物 A 的 3’端起前 7 个核苷酸中包含 1 个 CpG 核苷酸对。

13. 根据权利要求 10~12 任一项所述的方法，其特征在于，所述引物 A 的 3’端起第 2 个核苷酸为 C。

14. 根据权利要求 10~13 任一项所述的方法，其特征在于，所述引物 A 的 3’端和 5’端部分之间包含一段独特分子识别标签序列。

15. 根据权利要求 10~14 任一项所述的方法，其特征在于，所述步骤二中的 DNA 聚合酶具有链置换功能。

16. 根据权利要求 10~15 任一项所述的方法，其特征在于，所述步骤二中的 DNA 聚合酶是外切核酸酶活性缺损的。

17. 根据权利要求 10~16 任一项所述的方法，其特征在于，所述步骤二进行 2~30 次线性扩增。

18. 根据权利要求 10~17 任一项所述的方法，其特征在于，所述引物 B 的 3’端和 5’端部分之间包含一段长度为 2 至 10 个核苷酸的随机序列；优选的，该随机序列只包含 G、A 和 T。

19. 根据权利要求 10~18 任一项所述的方法，其特征在于，所述步骤三中的 DNA 聚合酶是热启动的。

20. 一种采用甲基化 CpG 短串联扩增与测序法进行检测的试剂盒，该试剂盒包含：权利要求 10~19 任一项所述的方法中所述引物 A、引物 B、接头引物 C 和接头引物 D、DNA 聚合酶以及使用该试剂盒的说明书。

21. 一种检测肝细胞癌的方法，包括如下步骤：1）获得来自人的 DNA 样品；2）检测核酸序列 1~68 或序列 69~132 中至少 1 种或其组合是否存在甲基化；这些核酸序列的甲基化提示肝细胞癌；所述 DNA 样品来自人的细胞、组织、血液、体液、尿液、唾液、排泄物或其组合，优选来自人的血浆或尿液游离 DNA。

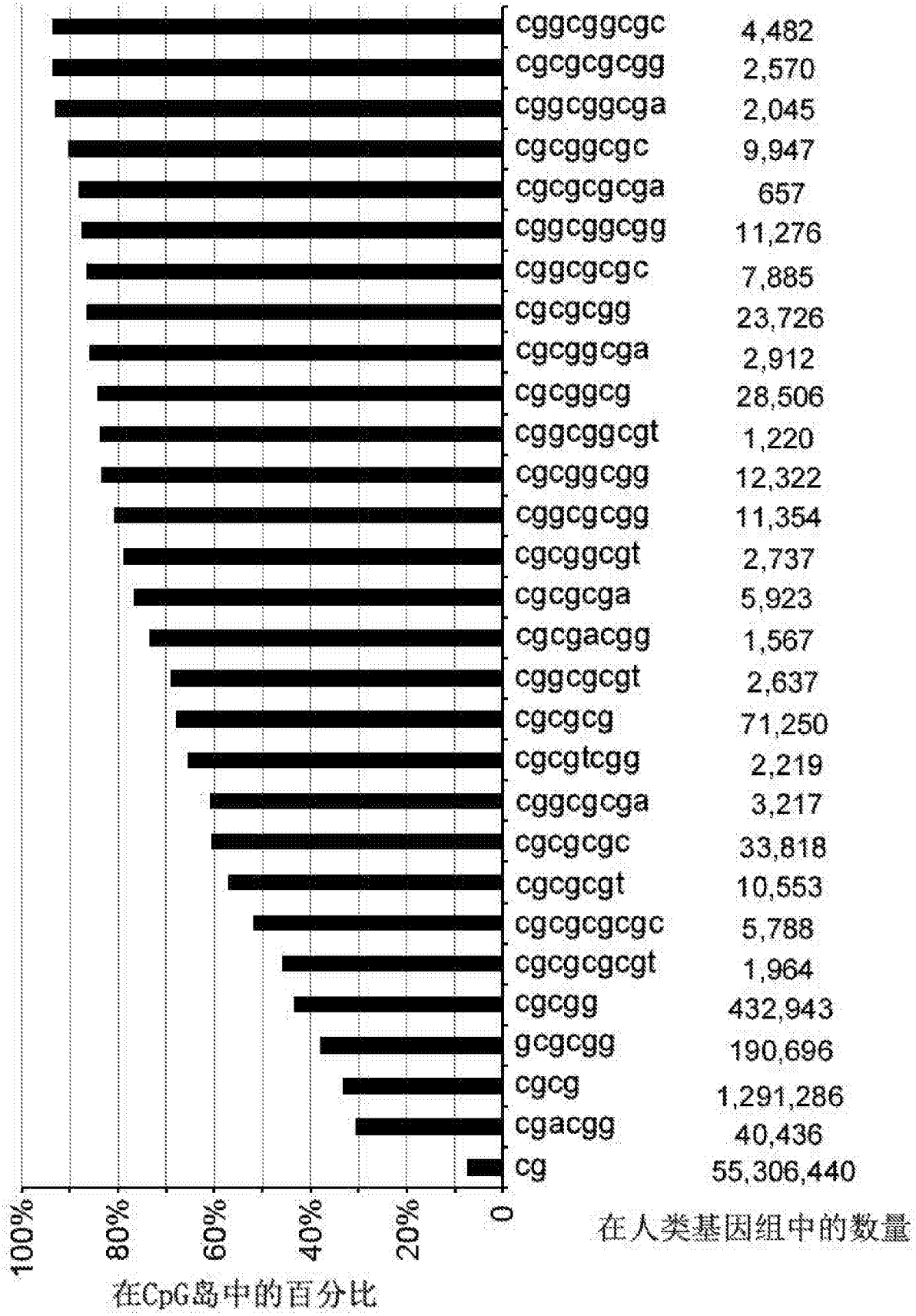

cggcggcgc 4,482
cgcgcgcgg 2,570
cggcggcga 2,045
cgcggcgc 9,947
cgcgcgcga 657
cggcggcgg 11,276
cggcgcgc 7,885
cgcgcgg 23,726
cgcggcga 2,912
cgcggcg 28,506
cggcggcgt 1,220
cgcggcgg 12,322
cggcgcgg 11,354
cgcggcgt 2,737
cgcgcga 5,923
cgcgacgg 1,567
cggcgcgt 2,637
cgcgcg 71,250
cgcgtcgg 2,219
cggcgcga 3,217
cgcgcgc 33,818
cgcgcgt 10,553
cgcgcgcgc 5,788
cgcgcgcgt 1,964
cgcgg 432,943
gcgcgg 190,696
cgcg 1,291,286
cgacgg 40,436
cg 55,306,440
100%
80%
60%
40%
20%
0
在CpG岛中的百分比
在人类基因组中的数量

图 1

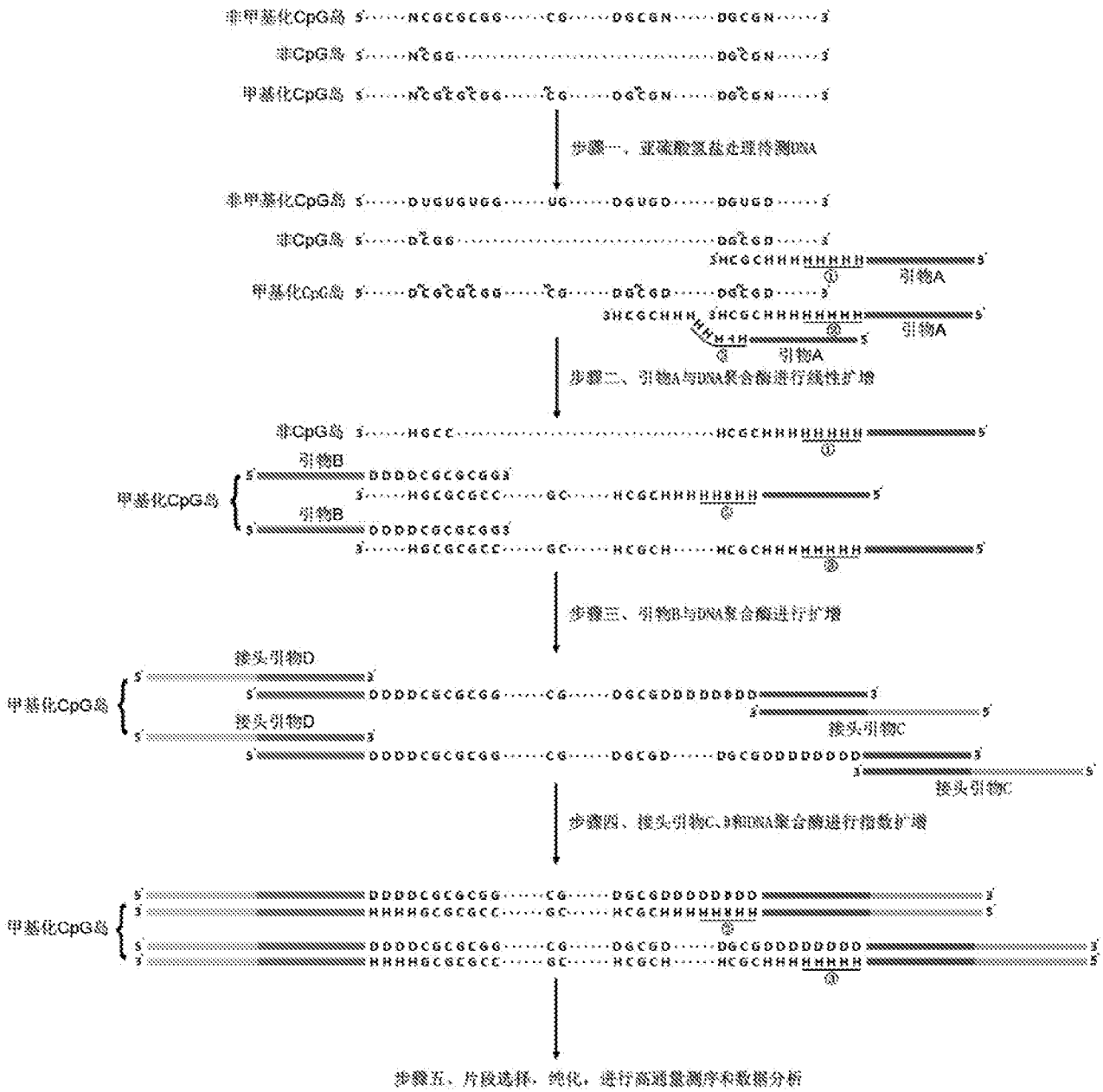

非甲基化CpG岛
非CpG岛
甲基化CpG岛
步骤一、亚硫酸氢盐处理待测DNA
引物A
步骤二、引物A与DNA聚合酶进行线性扩增
引物B
步骤三、引物B与DNA聚合酶进行扩增
接头引物D
接头引物C
步骤四、接头引物C、D和DNA聚合酶进行指数扩增
步骤五、片段选择、纯化，进行高通量测序和数据分析

图 2

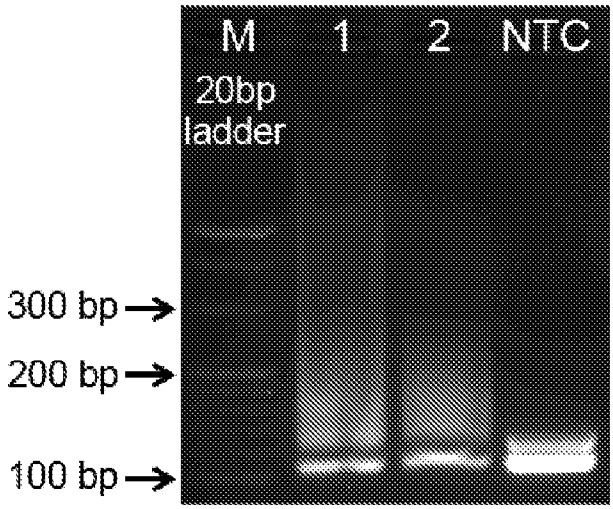

M
1
2
NTC
20bp
ladder
300 bp
200 bp
100 bp

图 3

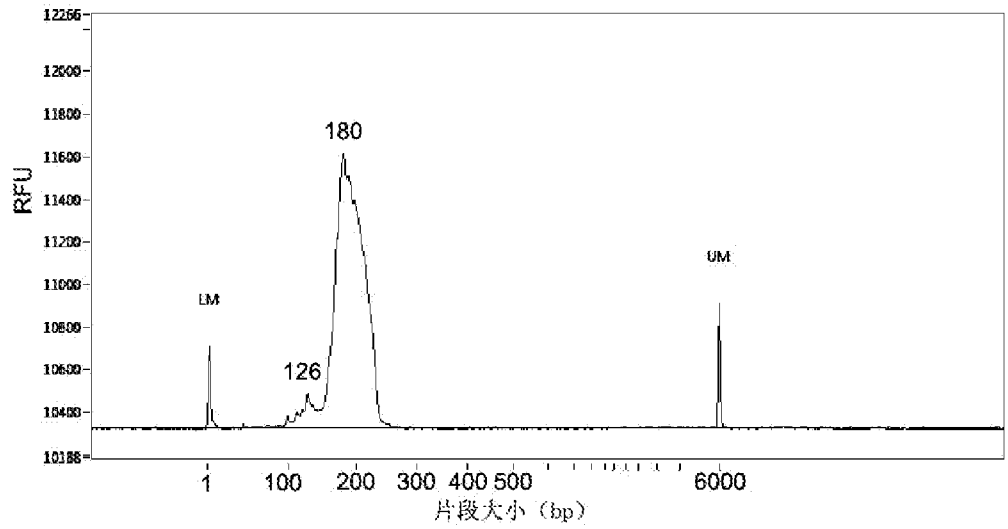

RFU
180
126
LM
UM
1
100
200
300
400
500
6000
片段大小（bp）

图 4

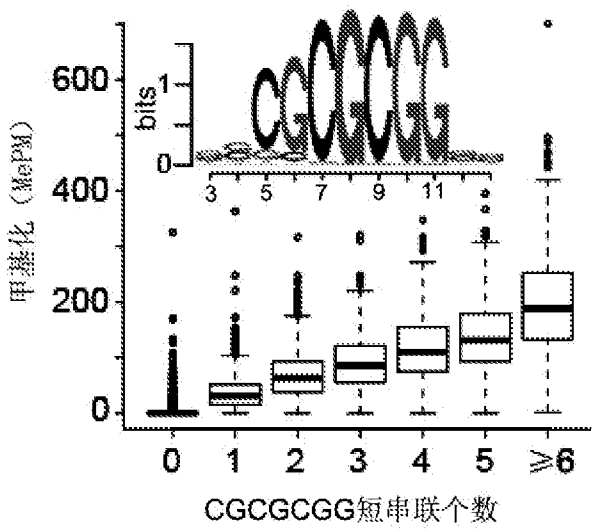

甲基化（MePM）
600
400
200
0
bits
1
0
3
5
7
9
11
0
1
2
3
4
5
≥6
CGCGCGG短串联个数

图 5

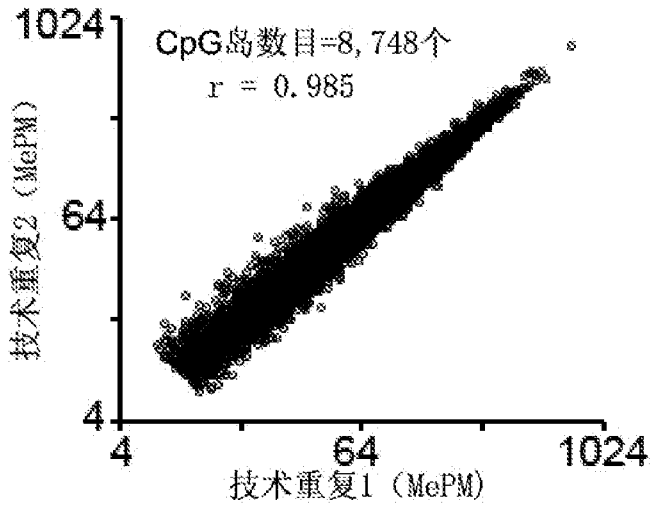

CpG岛数目=8,748个
r = 0.985
技术重复2（MePM）
技术重复1（MePM）
1024
64
4
4
64
1024

图 6

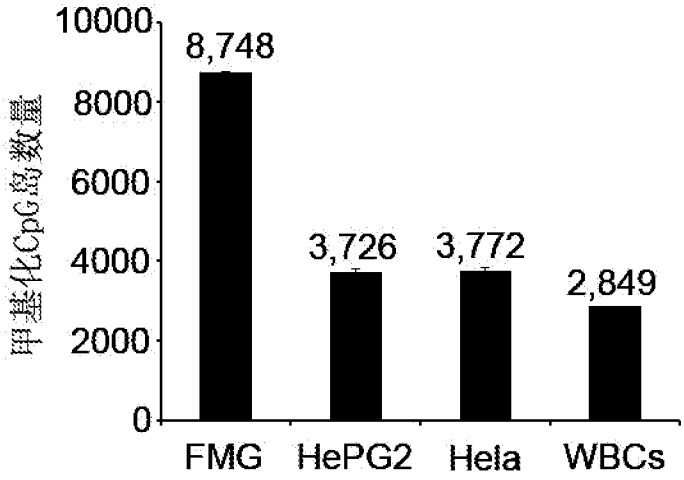

甲基化CpG岛数量
10000
8000
6000
4000
2000
0
8,748
3,726
3,772
2,849
FMG
HePG2
Hela
WBCs

图 7

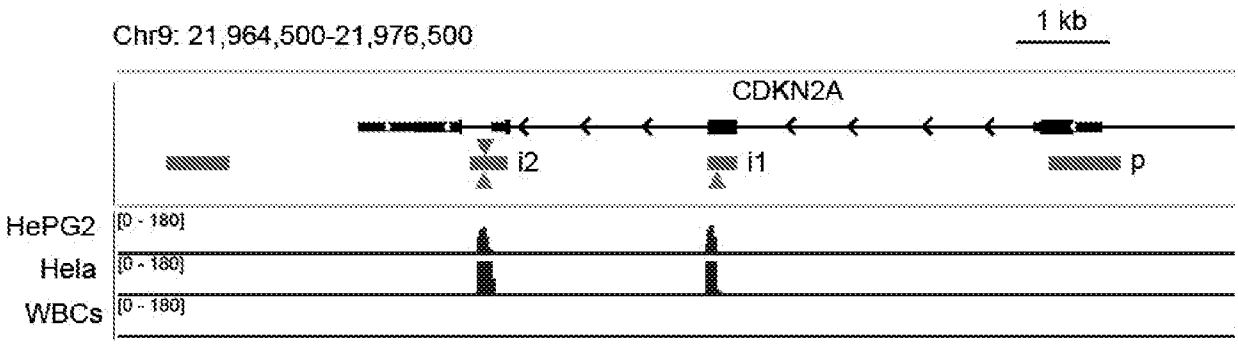

Chr9: 21,964,500-21,976,500
1 kb
CDKN2A
i2
i1
p
HePG2
Hela
WBCs
[0 - 180]
[0 - 180]
[0 - 180]

图 8

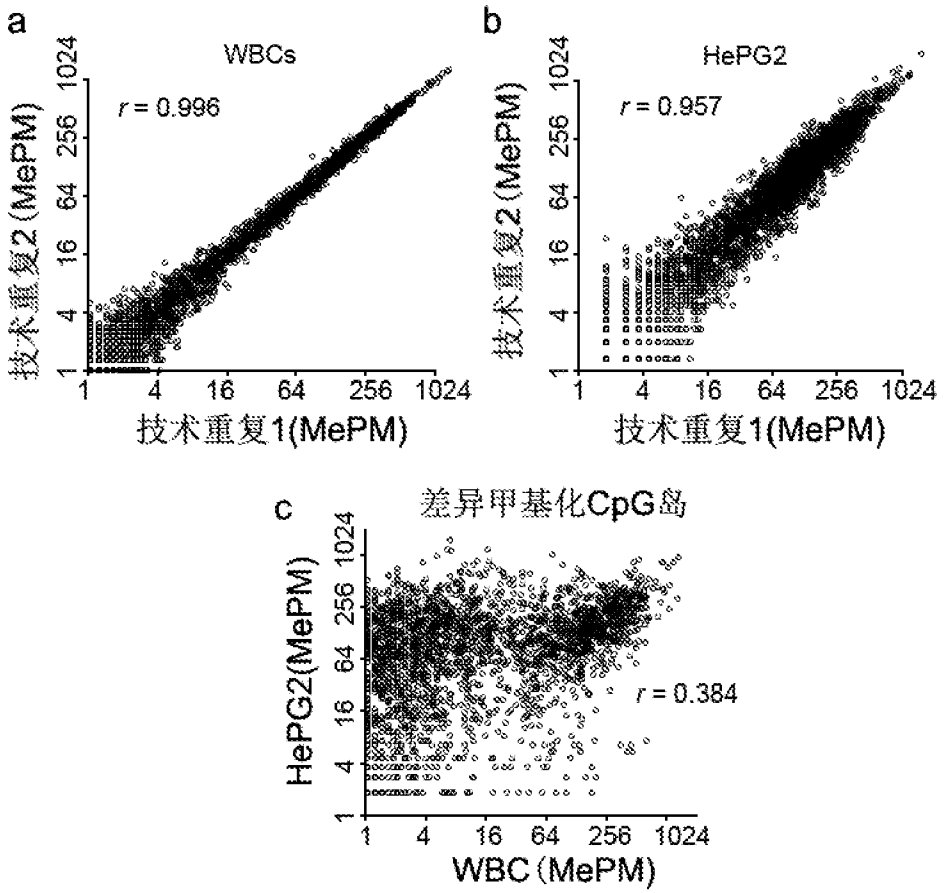

a
WBCs
r = 0.996
技术重复2 (MePM)
技术重复1(MePM)
b
HePG2
r = 0.957
技术重复2 (MePM)
技术重复1(MePM)
c
差异甲基化CpG岛
r = 0.384
HePG2(MePM)
WBC (MePM)
1
4
16
64
256
1024

图 9

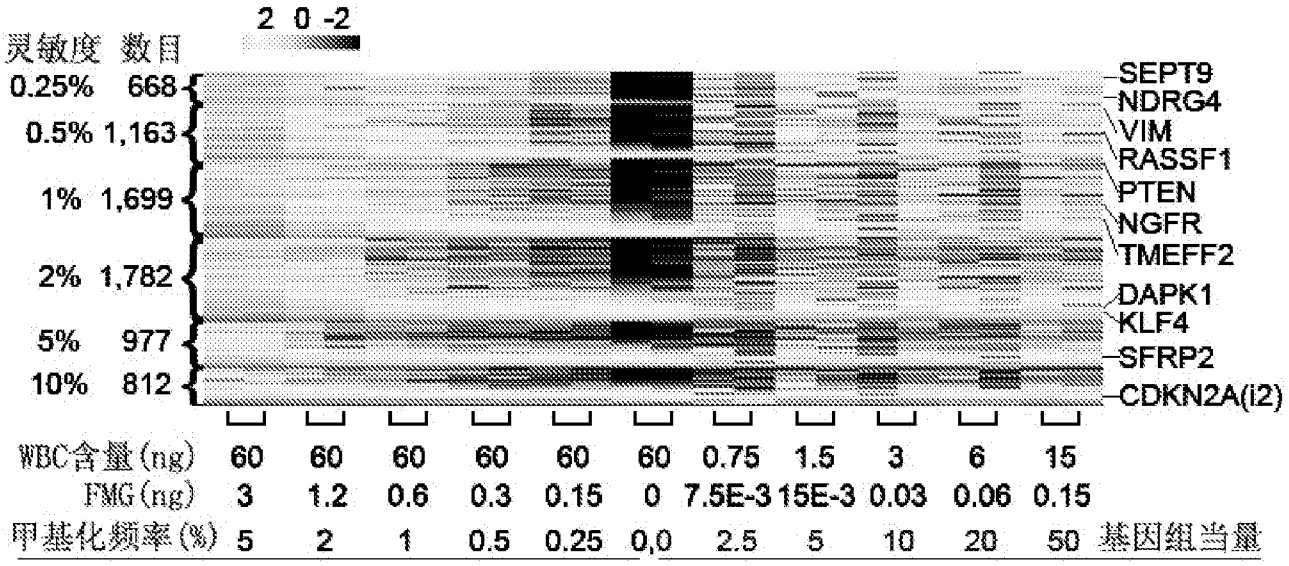

2 0 -2
灵敏度 数目
0.25% 668
0.5% 1,163
1% 1,699
2% 1,782
5% 977
10% 812
SEPT9
NDRG4
VIM
RASSF1
PTEN
NGFR
TMEFF2
DAPK1
KLF4
SFRP2
CDKN2A(i2)
WBC含量(ng) 60 60 60 60 60 60 0.75 1.5 3 6 15
FMG(ng) 3 1.2 0.6 0.3 0.15 0 7.5E-3 15E-3 0.03 0.06 0.15
甲基化频率(%) 5 2 1 0.5 0.25 0,0 2.5 5 10 20 50 基因组当量

图 10

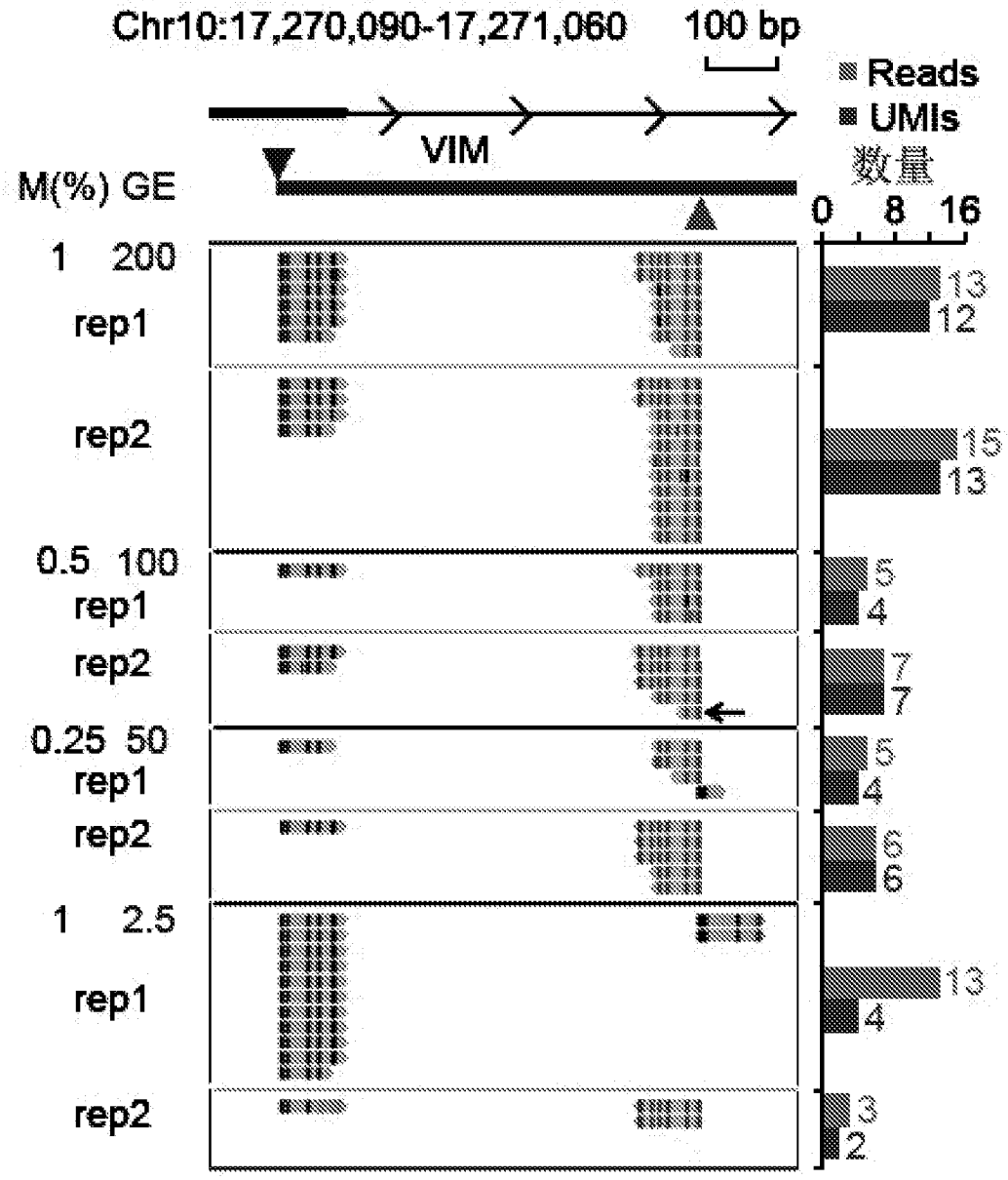

Chr10:17,270,090-17,271,060
100 bp
Reads
UMIs
数量
VIM
M(%) GE
0 8 16
1 200
rep1
rep2
0.5 100
rep1
rep2
0.25 50
rep1
rep2
1 2.5
rep1
rep2
13
12
15
13
5
4
7
7
5
4
6
6
13
4
3
2

图 11

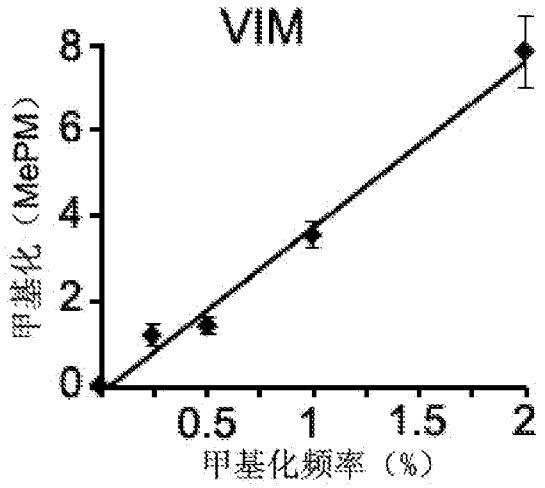

VIM
甲基化（MePM）
甲基化频率（%）
0
2
4
6
8
0.5
1
1.5
2

图 12

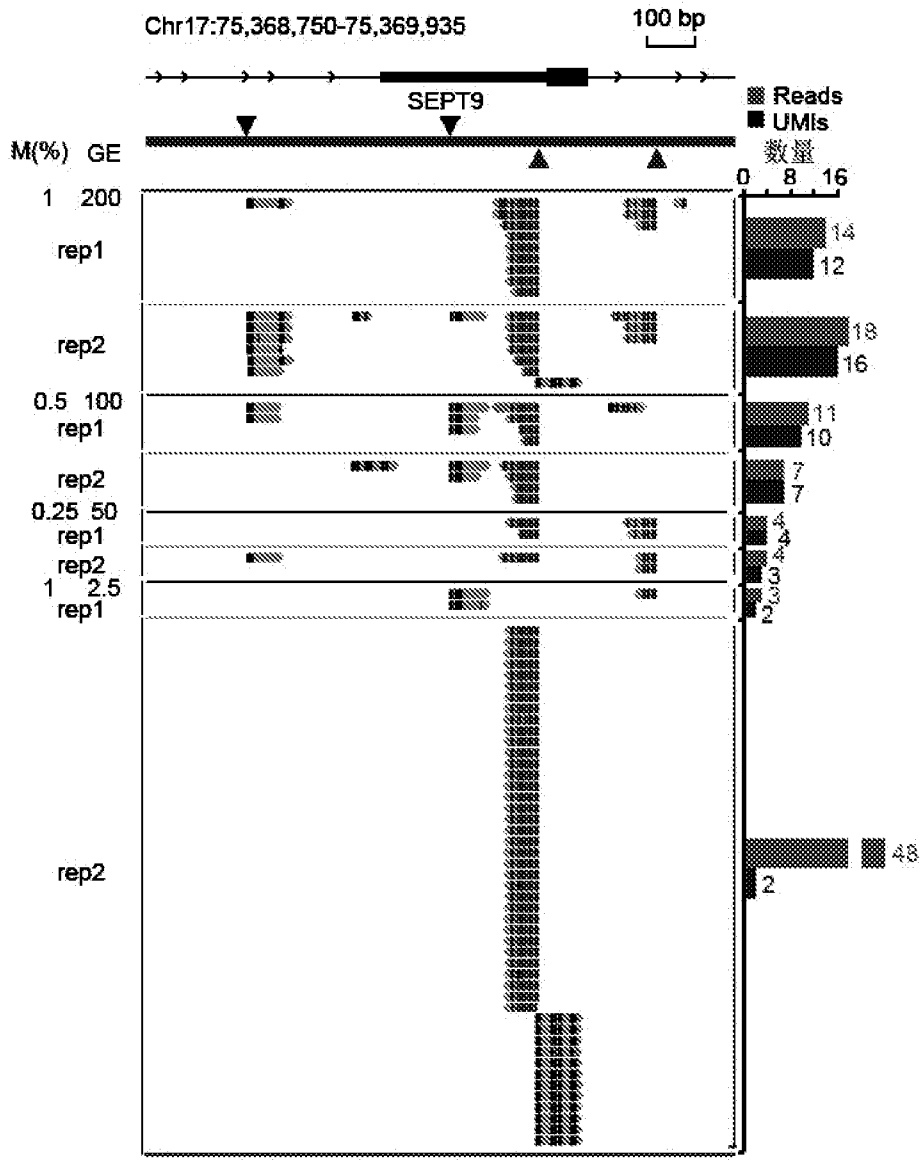

Chr17:75,368,750-75,369,935
100 bp
SEPT9
Reads
UMIs
M(%) GE
数量
0 8 16
1 200
rep1
14
12
rep2
18
16
0.5 100
rep1
11
10
rep2
7
7
0.25 50
rep1
rep2
1 2.5
rep1
rep2
48
2

图 13

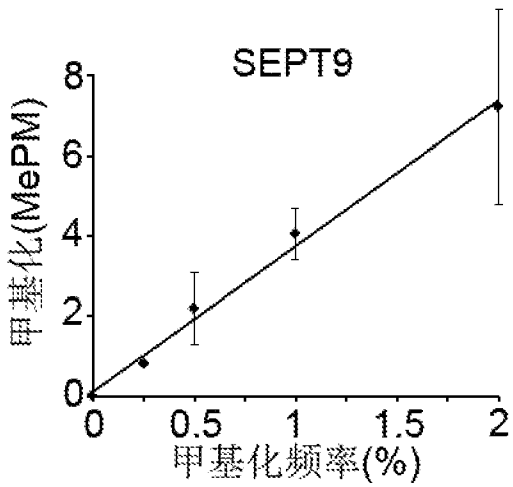

SEPT9
甲基化(MePM)
甲基化频率(%)
0
2
4
6
8
0
0.5
1
1.5
2

图 14

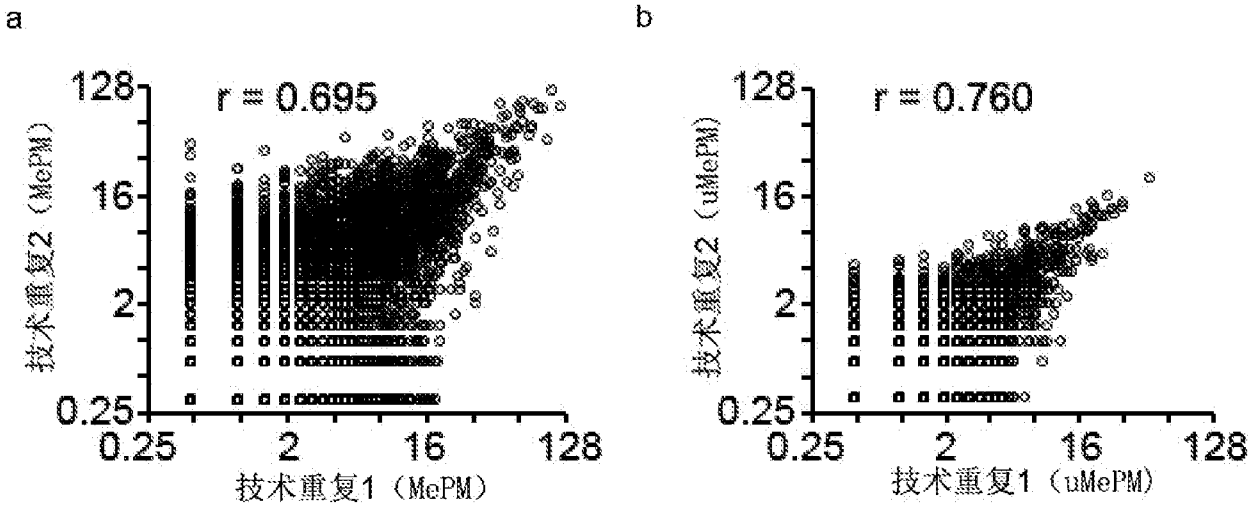

a
b
r = 0.695
r = 0.760
技术重复2（MePM）
技术重复1（MePM）
技术重复2（uMePM）
技术重复1（uMePM）
128
16
2
0.25
0.25
2
16
128

图 15

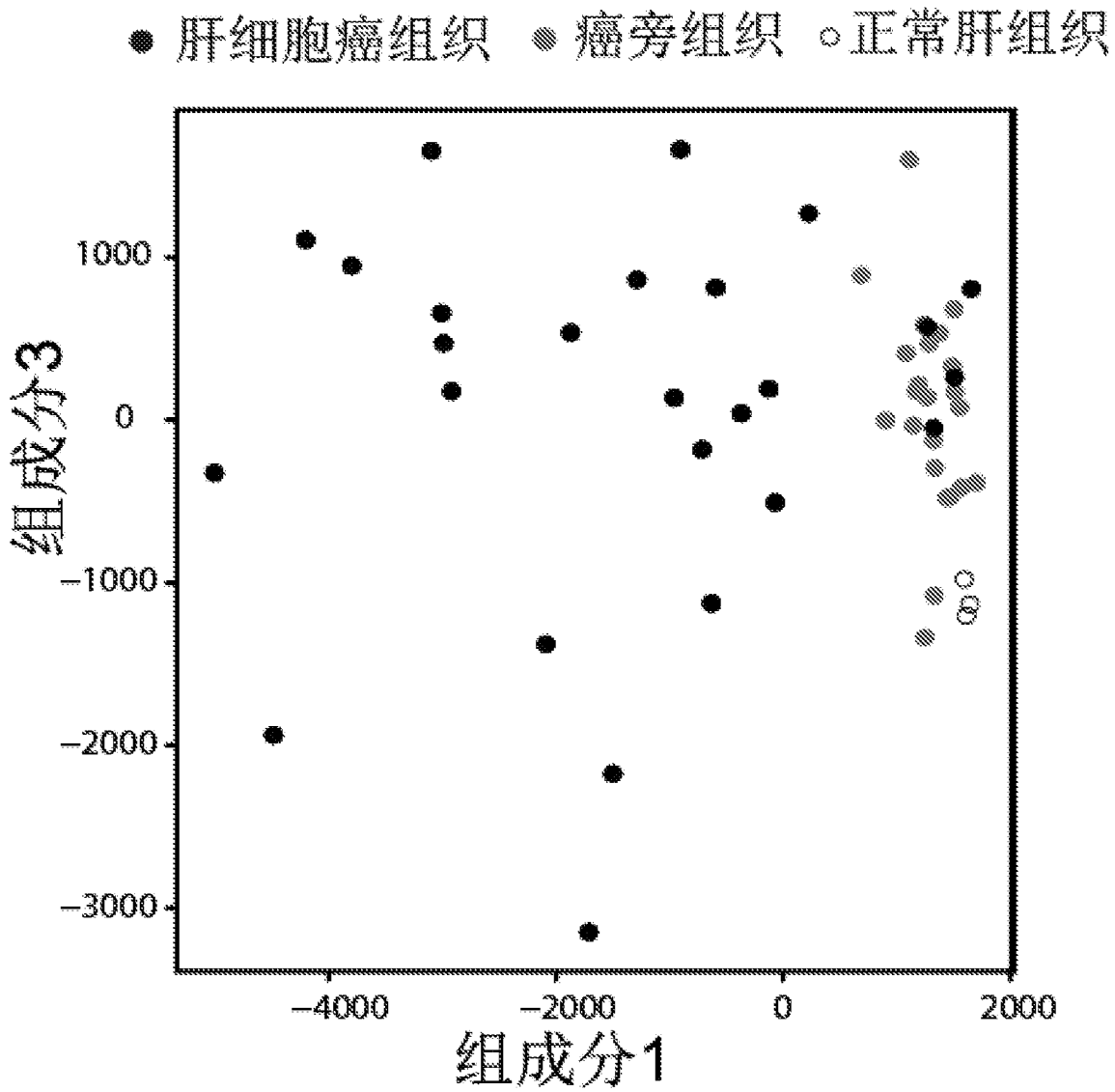

肝细胞癌组织
癌旁组织
正常肝组织
组成分3
组成分1
1000
0
−1000
−2000
−3000
−4000
−2000
0
2000

图 16

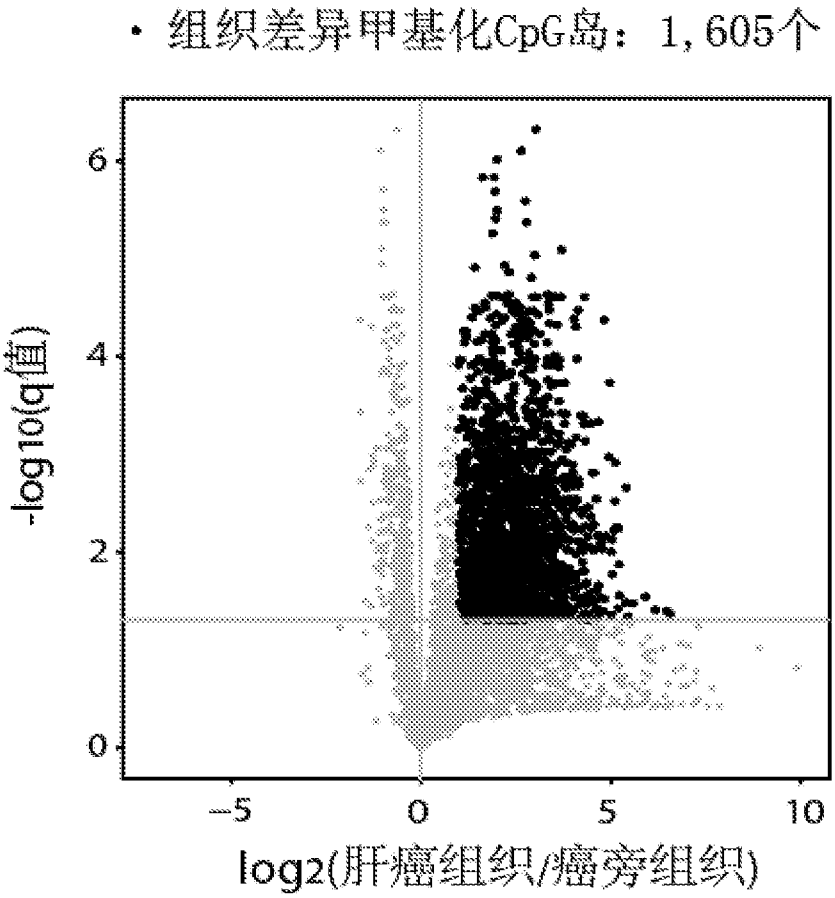

• 组织差异甲基化CpG岛：1,605个
-log10(q值)
log2(肝癌组织/癌旁组织)

图 17

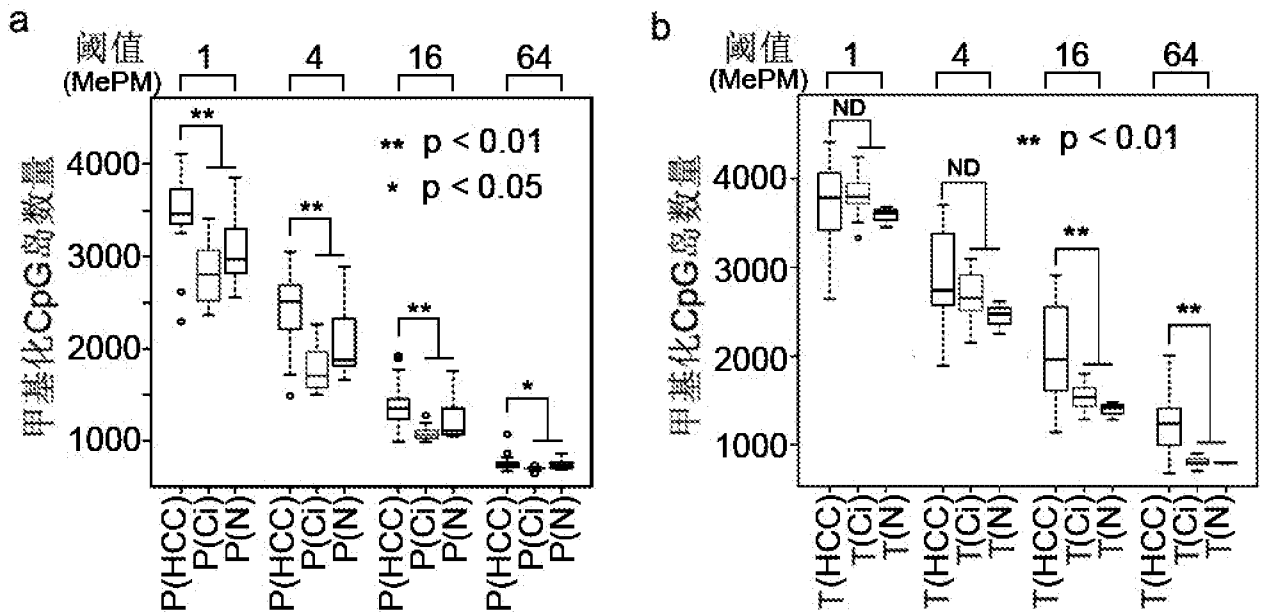

a
阈值 (MePM)
1
4
16
64
** p < 0.01
* p < 0.05
甲基化CpG岛数量
P(HCC)
P(Ci)
P(N)
b
ND
T(HCC)
T(Ci)
T(N)

图 18

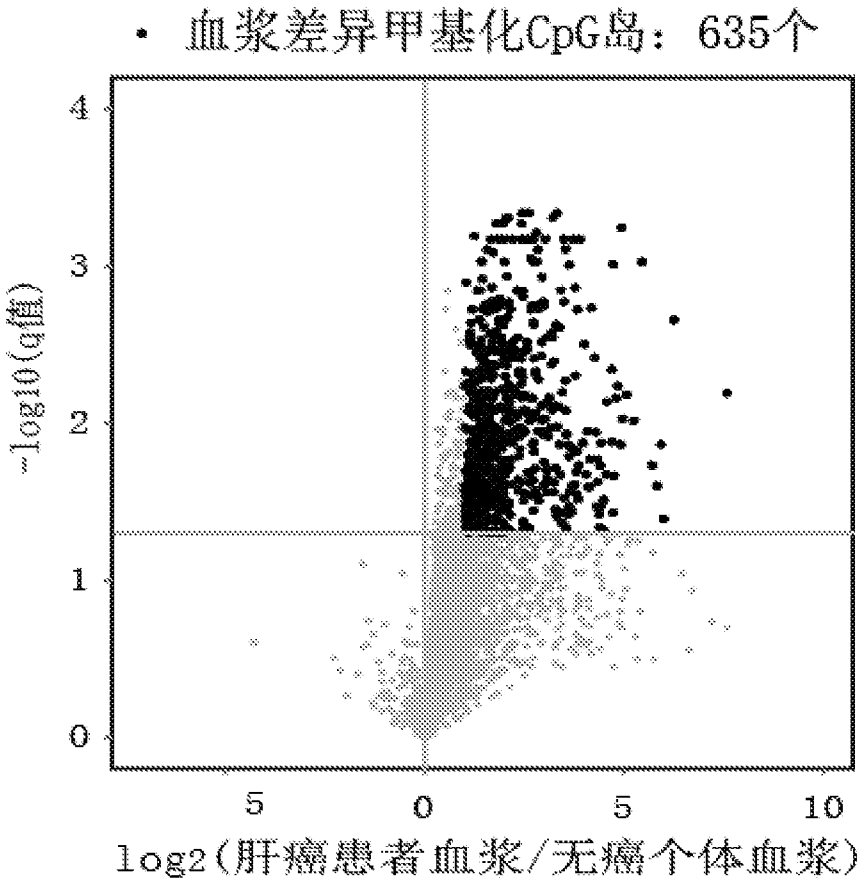

• 血浆差异甲基化CpG岛：635个
-log10(q值)
4
3
2
1
0
5
0
5
10
log2(肝癌患者血浆/无癌个体血浆)

图 19

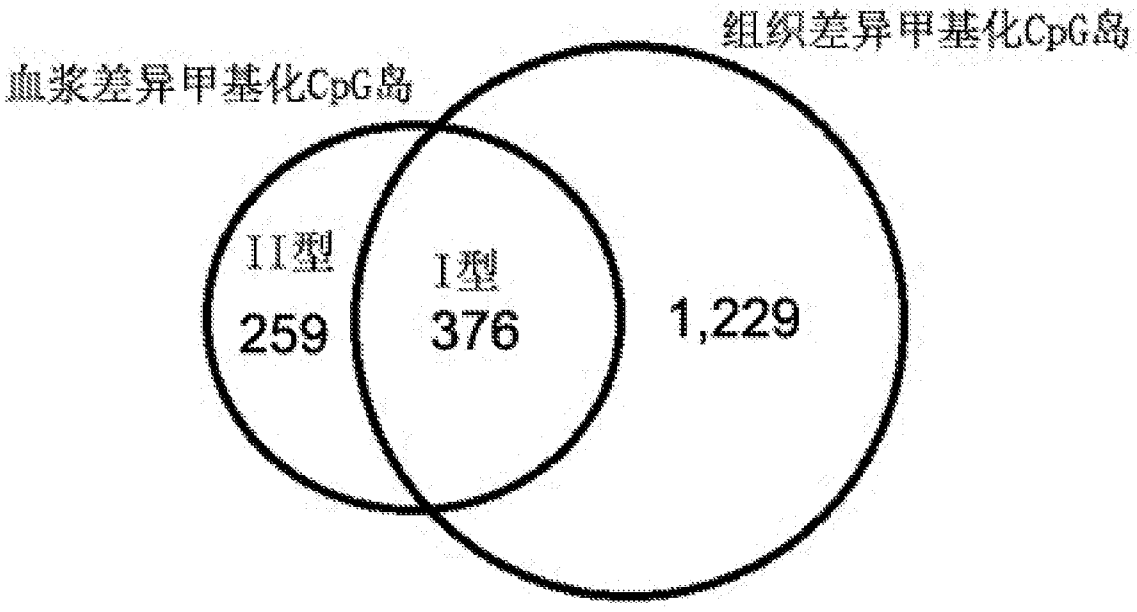

组织差异甲基化CpG岛
血浆差异甲基化CpG岛
II型
259
I型
376
1,229

图 20

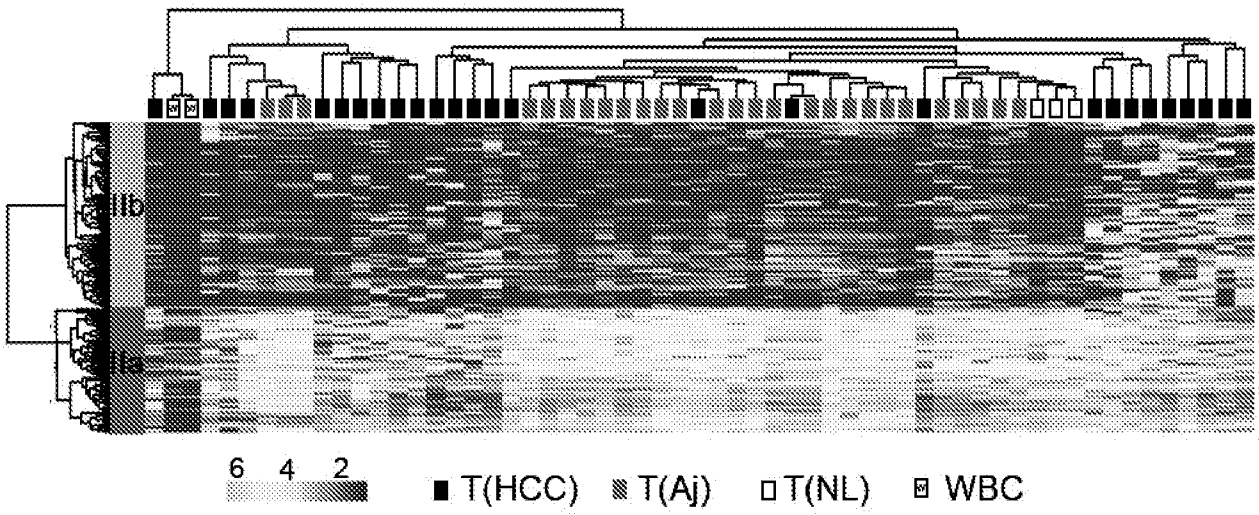

IIb
IIa
6 4 2
T(HCC)
T(Aj)
T(NL)
WBC

图 21

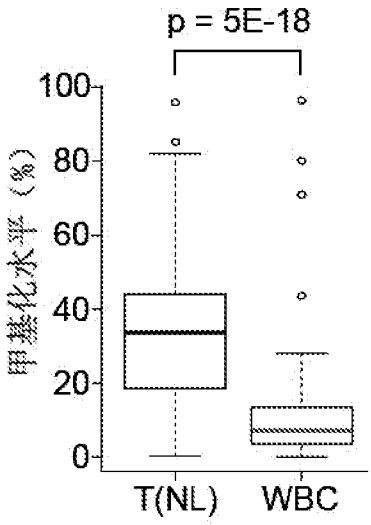

p = 5E-18
甲基化水平（%）
100
80
60
40
20
0
T(NL)
WBC

图 22

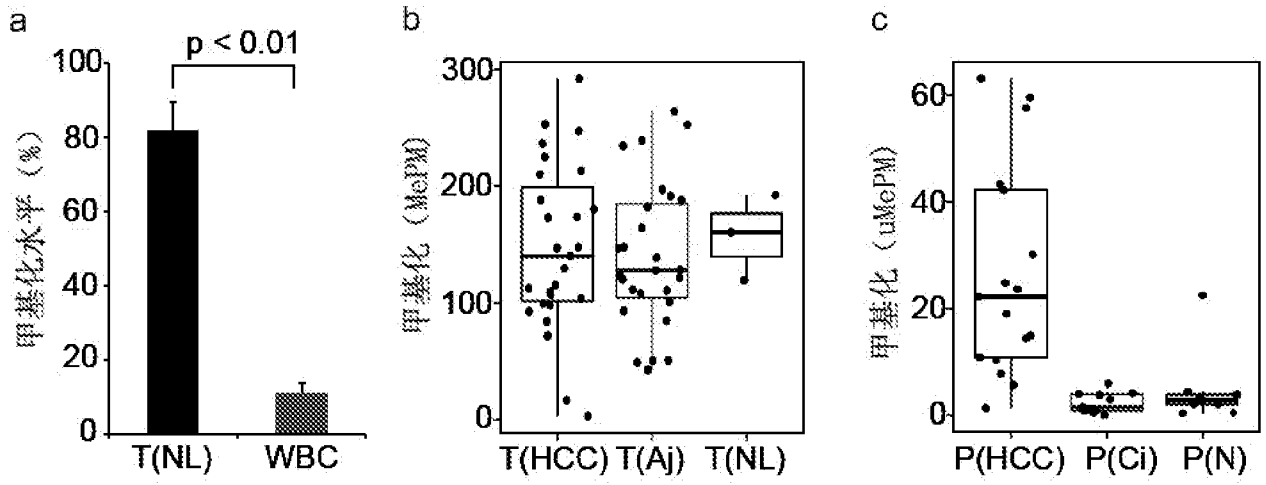

a
p < 0.01
甲基化水平（%）
100
80
60
40
20
0
T(NL)
WBC
b
甲基化（MePM）
300
200
100
0
T(HCC)
T(Aj)
T(NL)
c
甲基化（uMePM）
60
40
20
0
P(HCC)
P(Ci)
P(N)

图 23

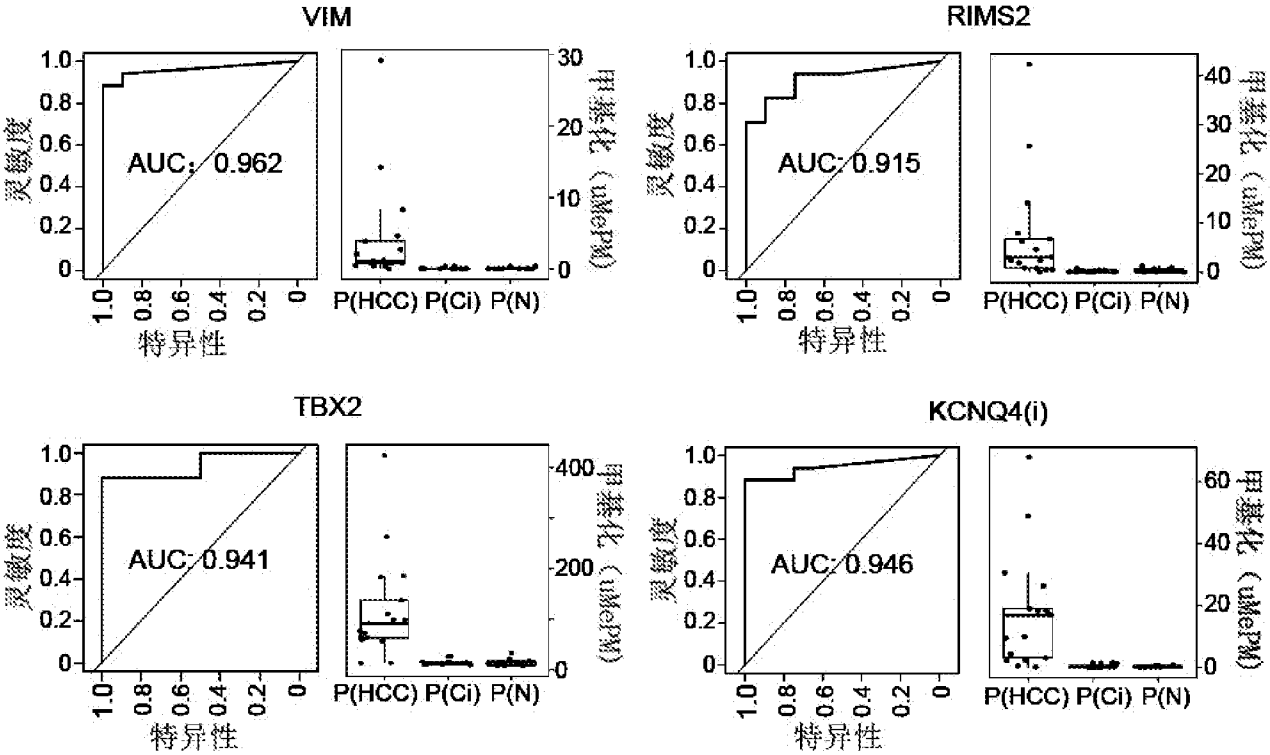

VIM
AUC：0.962
RIMS2
AUC: 0.915
TBX2
AUC: 0.941
KCNQ4(i)
AUC: 0.946
灵敏度
特异性
甲基化（uMePM）
P(HCC) P(Ci) P(N)

图 24

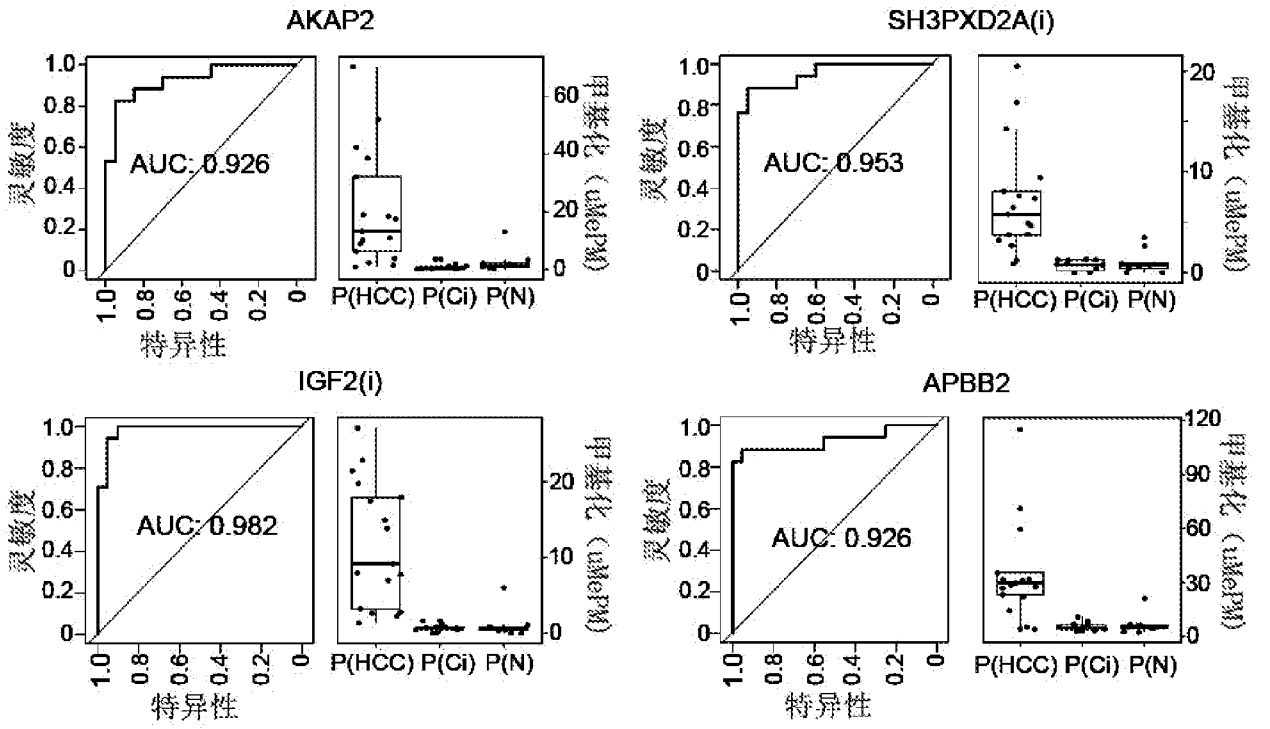

AKAP2
AUC: 0.926
SH3PXD2A(i)
AUC: 0.953
IGF2(i)
AUC: 0.982
APBB2
AUC: 0.926
灵敏度
特异性
甲基化（uMePM）
P(HCC) P(Ci) P(N)

图 25

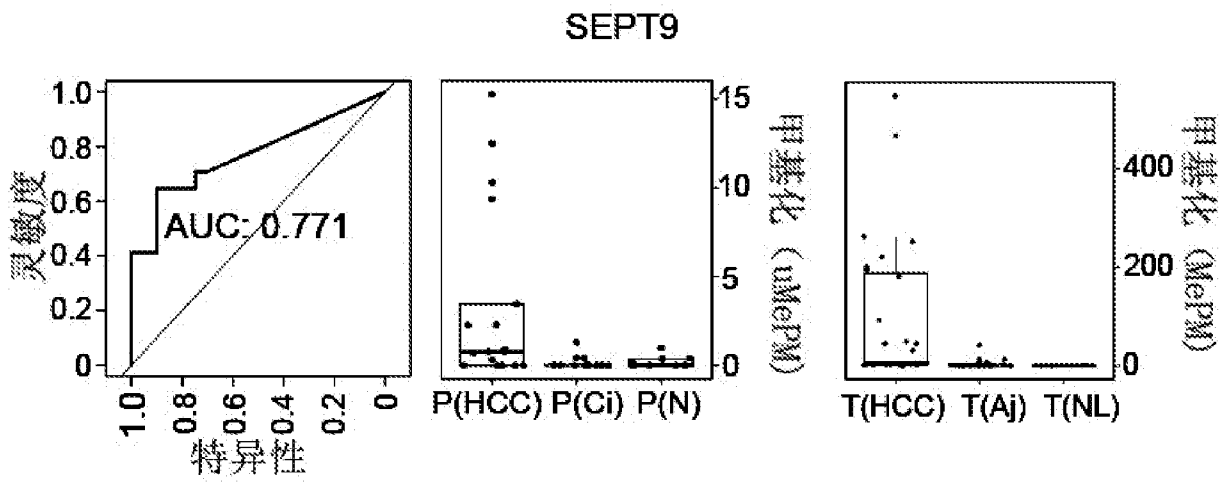

SEPT9
灵敏度
1.0
0.8
0.6
0.4
0.2
0
AUC: 0.771
1.0 0.8 0.6 0.4 0.2 0
特异性
15
10
5
0
甲基化 (uMePM)
P(HCC) P(Ci) P(N)
400
200
0
甲基化 (MePM)
T(HCC) T(Aj) T(NL)

图 26

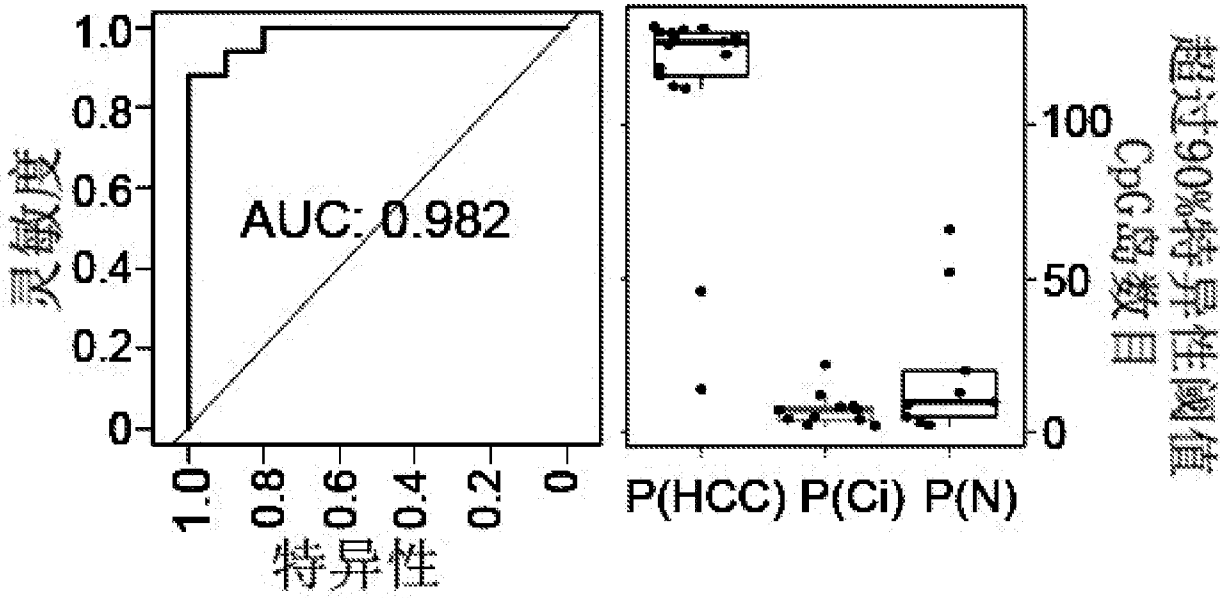
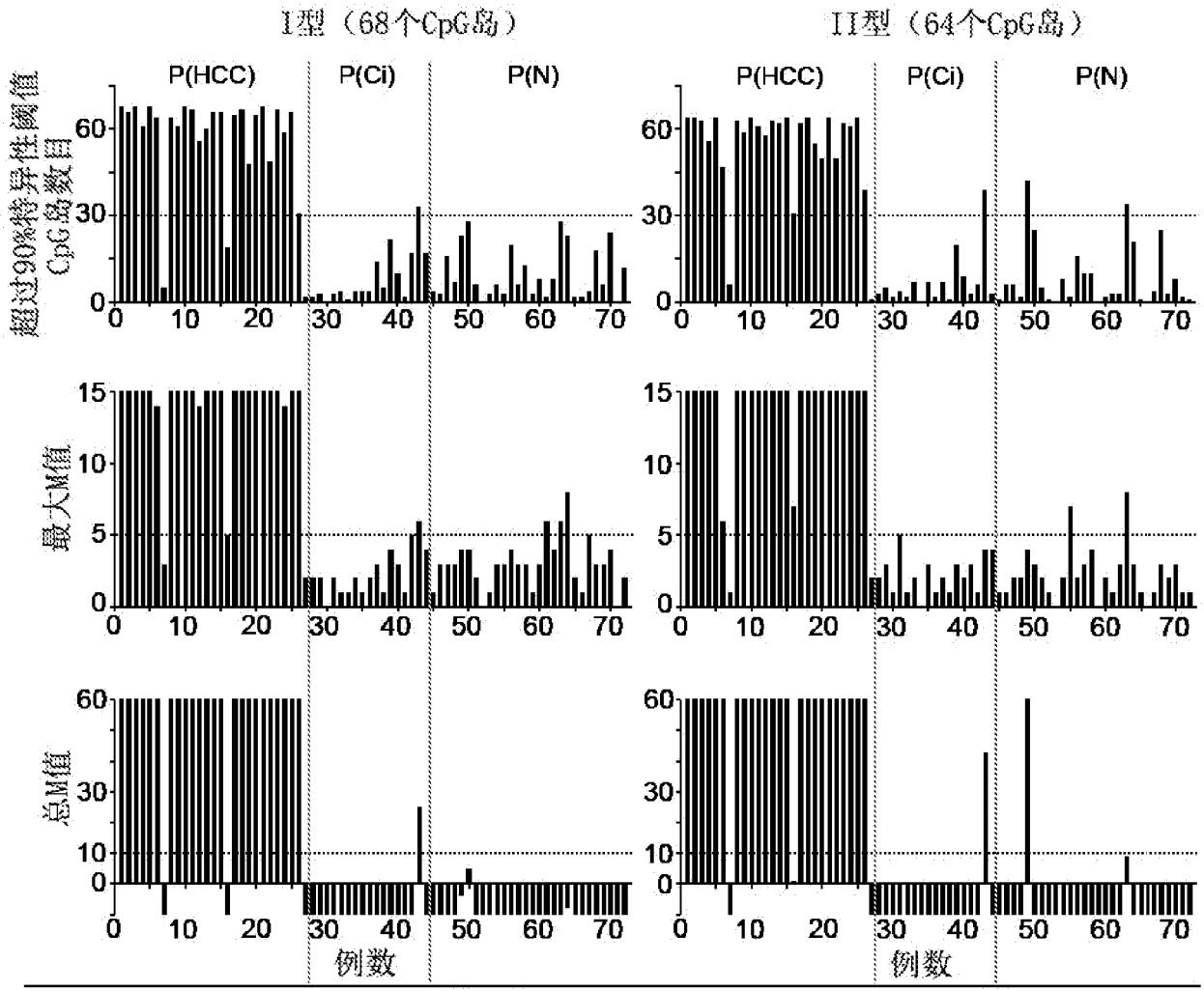

灵敏度
1.0
0.8
0.6
0.4
0.2
0
AUC: 0.982
1.0 0.8 0.6 0.4 0.2 0
特异性
100
50
0
CpG甲基化数目
最优90%特异性阈值
P(HCC) P(Ci) P(N)

图 27



	灵敏度			特异性		
	训练组	测试组	总体	训练组	测试组	总体
高特异性标准（≥5个阳性指标）	88%(15/17)	90%（9/10）	89%(24/27)	100%(17/17)	96%(24/25)	98%(44/45)
高灵敏度标准（≥3个阳性指标，其中至少1个为I型CpG岛指标）	94%(16/17)	90%(9/10)	93%(25/27)	100%(20/20)	92%(23/25)	96%(43/45)

图 28

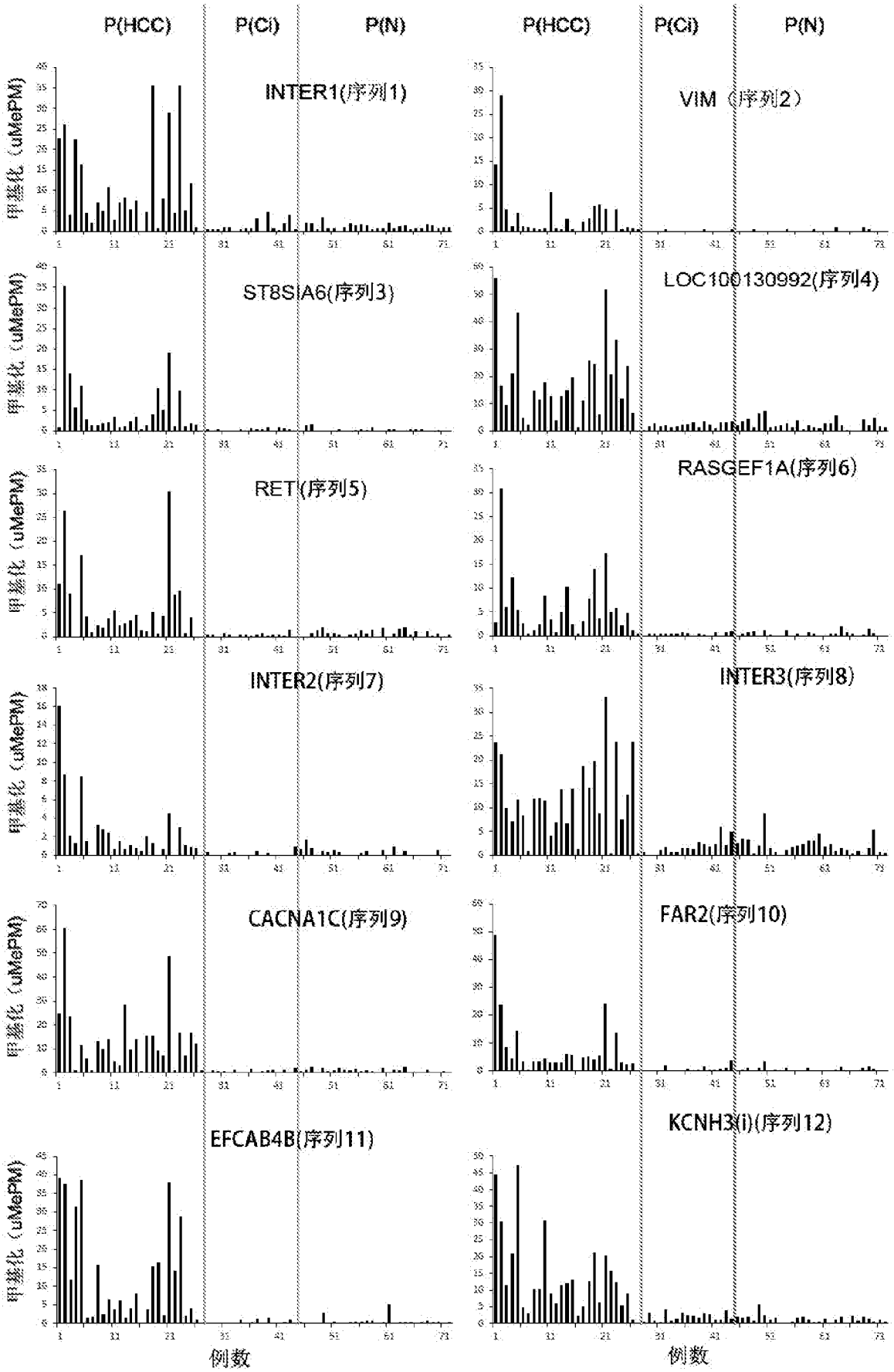

P(HCC)
P(Ci)
P(N)
INTER1(序列1)
VIM（序列2）
ST8SIA6(序列3)
LOC100130992(序列4)
RET(序列5)
RASGEF1A(序列6)
INTER2(序列7)
INTER3(序列8)
CACNA1C(序列9)
FAR2(序列10)
EFCAB4B(序列11)
KCNH3(i)(序列12)
甲基化（uMePM）
例数

图 29

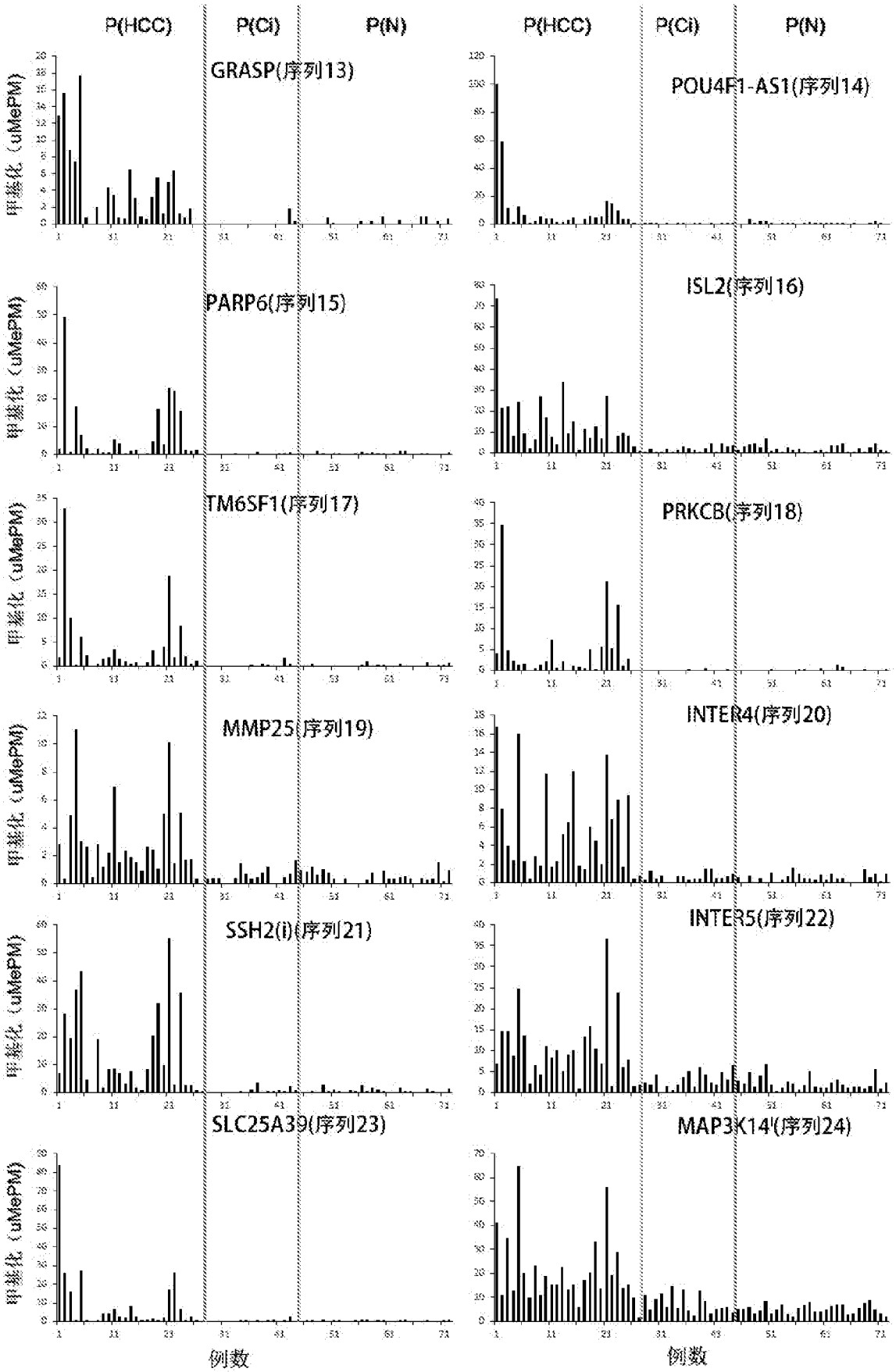

P(HCC)
P(Ci)
P(N)
GRASP(序列13)
POU4F1-AS1(序列14)
PARP6(序列15)
ISL2(序列16)
TM6SF1(序列17)
PRKCB(序列18)
MMP25(序列19)
INTER4(序列20)
SSH2(i)(序列21)
INTER5(序列22)
SLC25A39(序列23)
MAP3K14'(序列24)
甲基化（uMePM）
例数

图 30

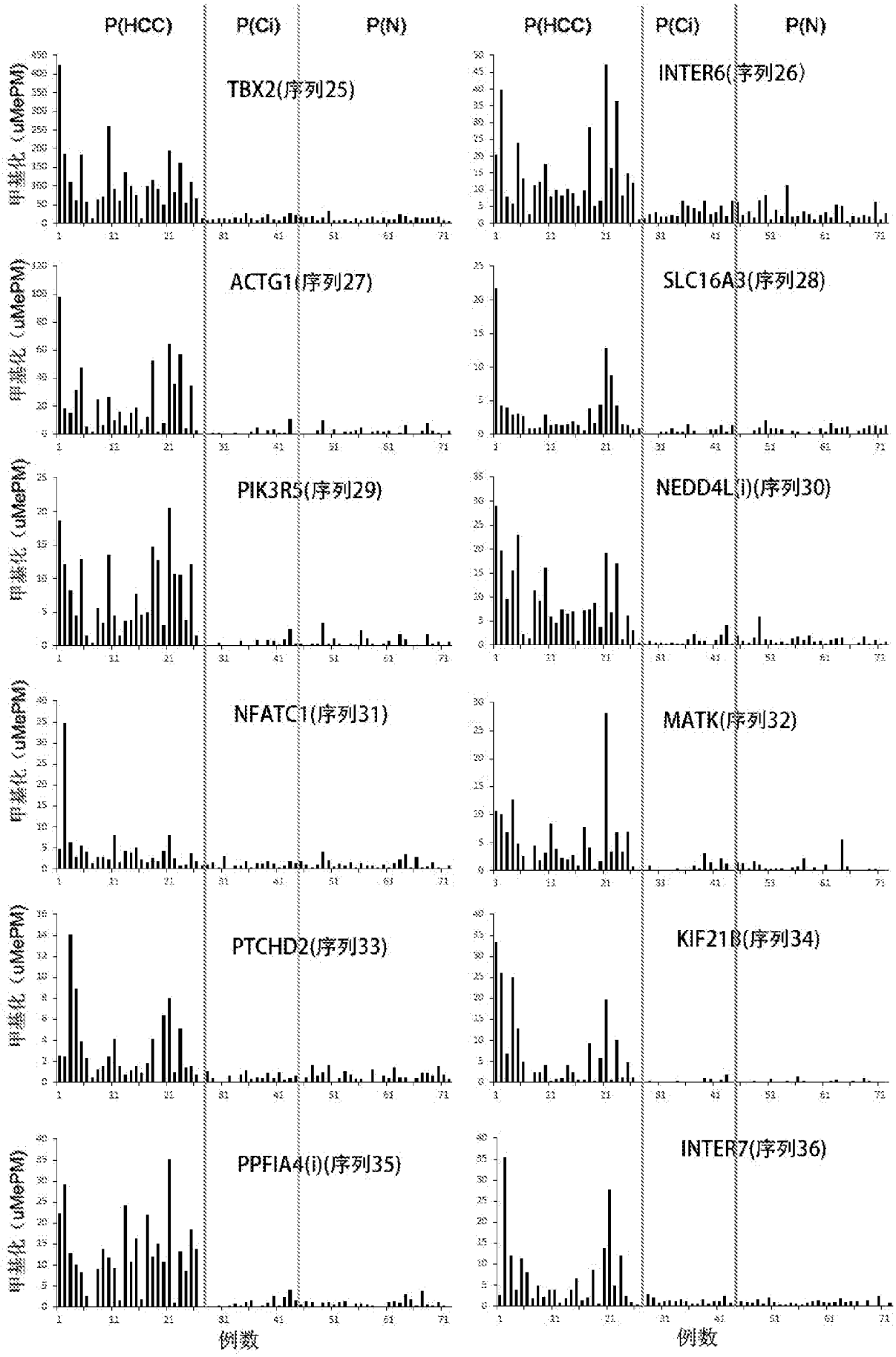

P(HCC)
P(Ci)
P(N)
TBX2(序列25)
INTER6(序列26)
ACTG1(序列27)
SLC16A3(序列28)
PIK3R5(序列29)
NEDD4L(i)(序列30)
NFATC1(序列31)
MATK(序列32)
PTCHD2(序列33)
KIF21B(序列34)
PPFIA4(i)(序列35)
INTER7(序列36)
甲基化（uMePM）
例数

图 31

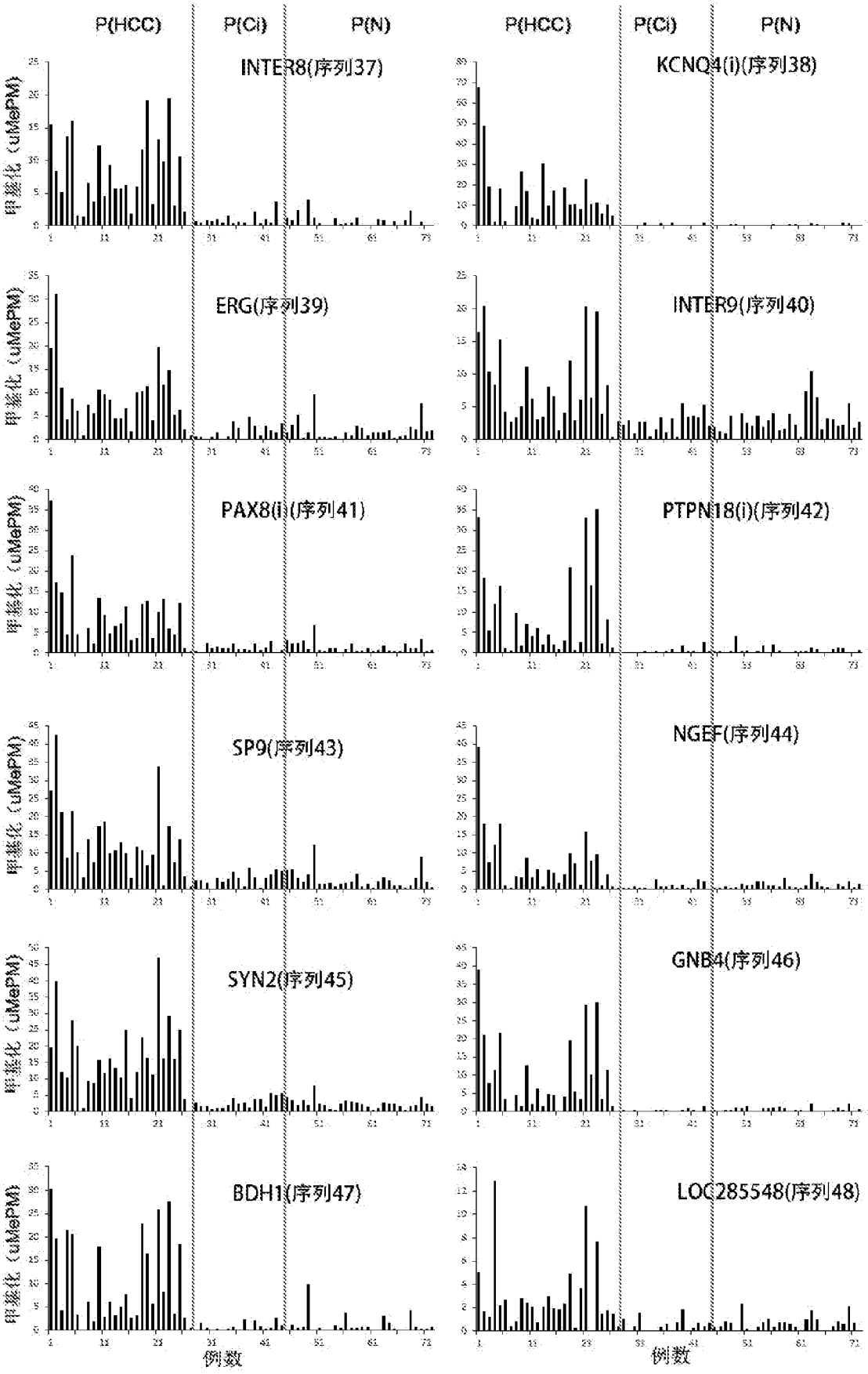

P(HCC)
P(Ci)
P(N)
INTER8(序列37)
KCNQ4(i)(序列38)
ERG(序列39)
INTER9(序列40)
PAX8(i)(序列41)
PTPN18(i)(序列42)
SP9(序列43)
NGEF(序列44)
SYN2(序列45)
GNB4(序列46)
BDH1(序列47)
LOC285548(序列48)
甲基化（uMePM）
例数

图 32

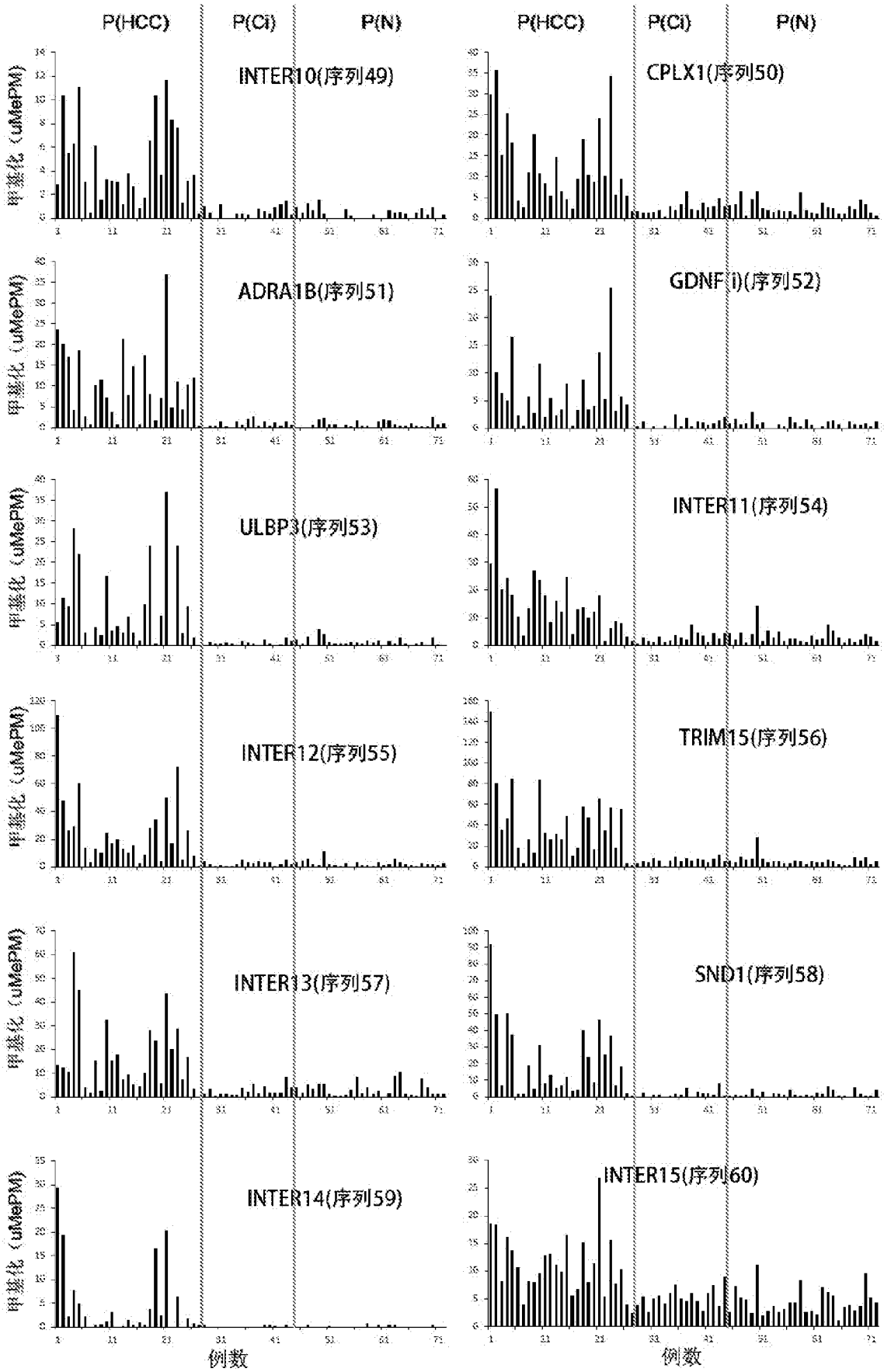

P(HCC)
P(Ci)
P(N)
甲基化（uMePM）
INTER10(序列49)
CPLX1(序列50)
ADRA1B(序列51)
GDNF i)(序列52)
ULBP3(序列53)
INTER11(序列54)
INTER12(序列55)
TRIM15(序列56)
INTER13(序列57)
SND1(序列58)
INTER14(序列59)
INTER15(序列60)
例数
例数

图 33

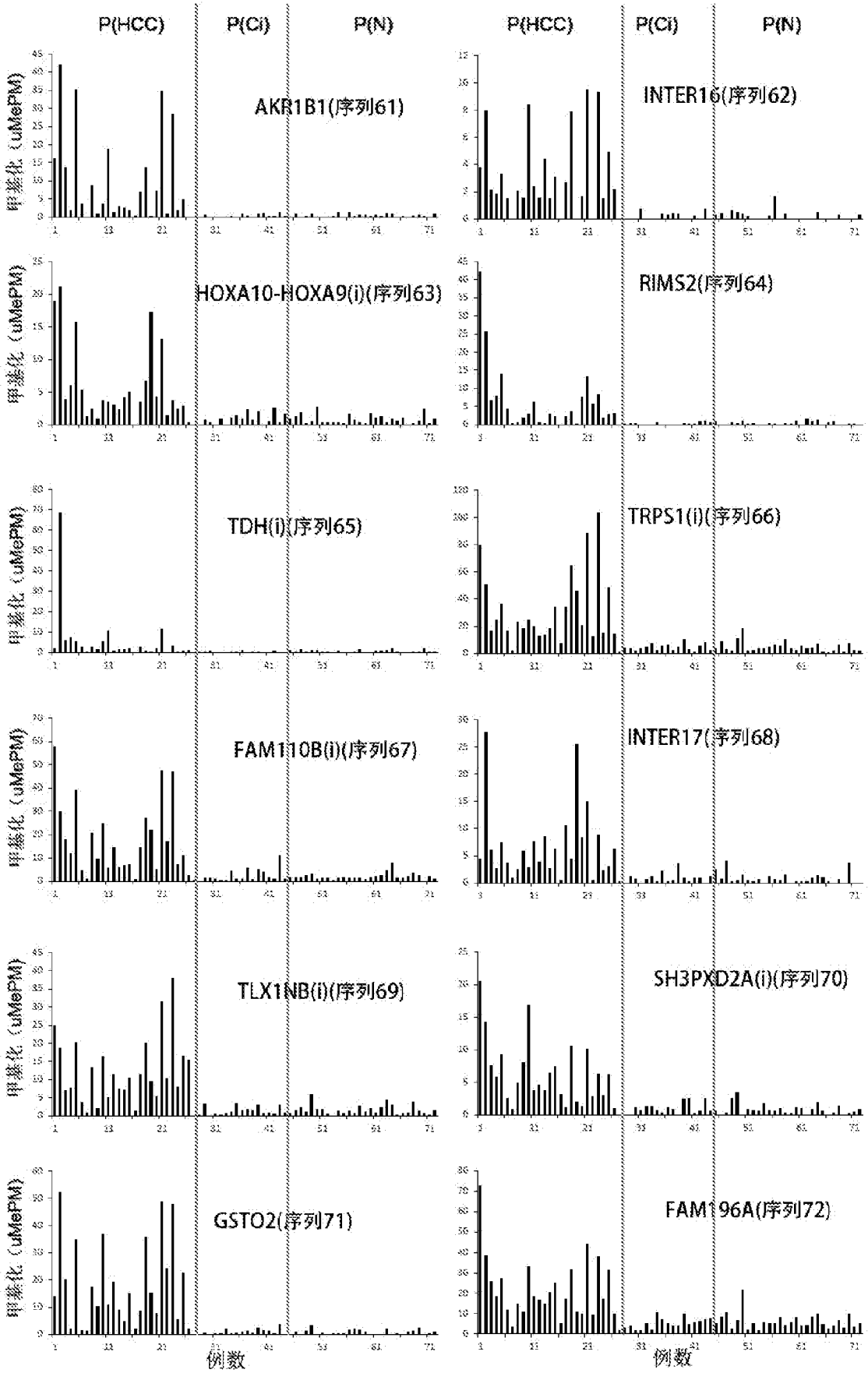

P(HCC)
P(Ci)
P(N)
甲基化（uMePM）
AKR1B1(序列61)
INTER16(序列62)
HOXA10-HOXA9(i)(序列63)
RIMS2(序列64)
TDH(i)(序列65)
TRPS1(i)(序列66)
FAM110B(i)(序列67)
INTER17(序列68)
TLX1NB(i)(序列69)
SH3PXD2A(i)(序列70)
GSTO2(序列71)
FAM196A(序列72)
例数

图 34

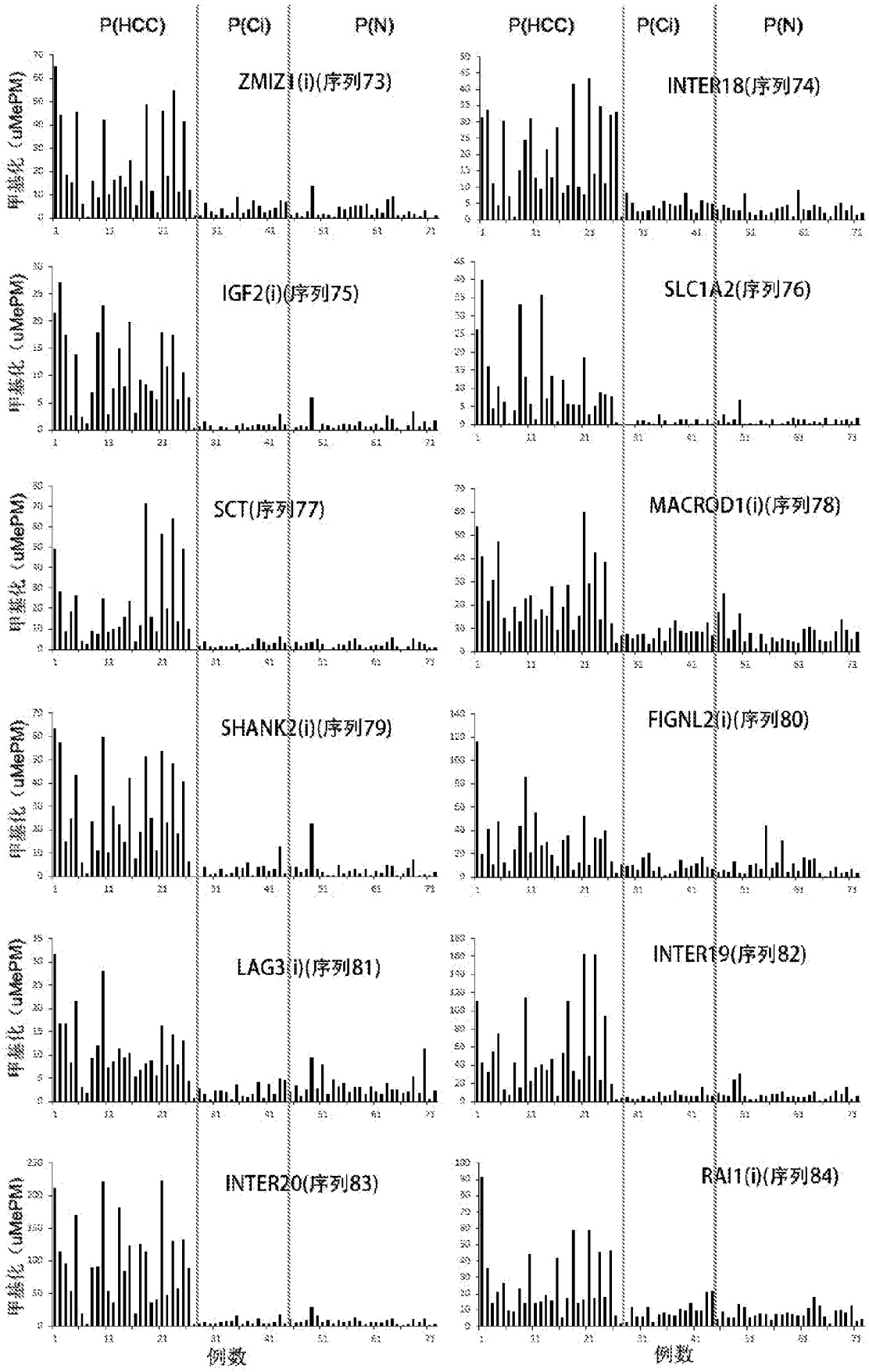

P(HCC)
P(Ci)
P(N)
甲基化（uMePM）
ZMIZ1(i)(序列73)
INTER18(序列74)
IGF2(i)(序列75)
SLC1A2(序列76)
SCT(序列77)
MACROD1(i)(序列78)
SHANK2(i)(序列79)
FIGNL2(i)(序列80)
LAG3(i)(序列81)
INTER19(序列82)
INTER20(序列83)
RAI1(i)(序列84)
例数

图 35

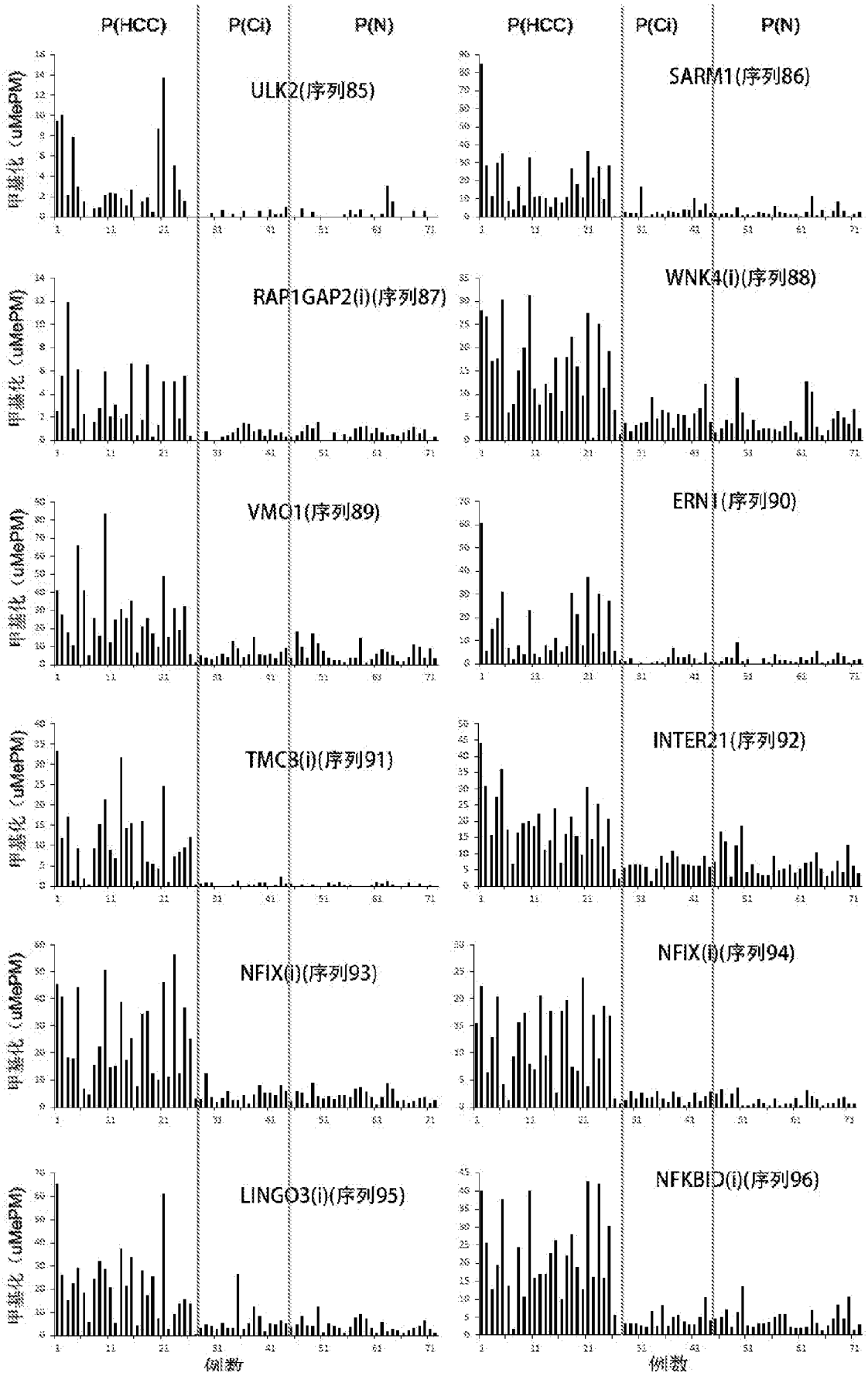

P(HCC)
P(Ci)
P(N)
甲基化（uMePM）
ULK2(序列85)
SARM1(序列86)
RAP1GAP2(i)(序列87)
WNK4(i)(序列88)
VMO1(序列89)
ERN1(序列90)
TMC8(i)(序列91)
INTER21(序列92)
NFIX(i)(序列93)
NFIX(i)(序列94)
LINGO3(i)(序列95)
NFKBID(i)(序列96)
例数

图 36

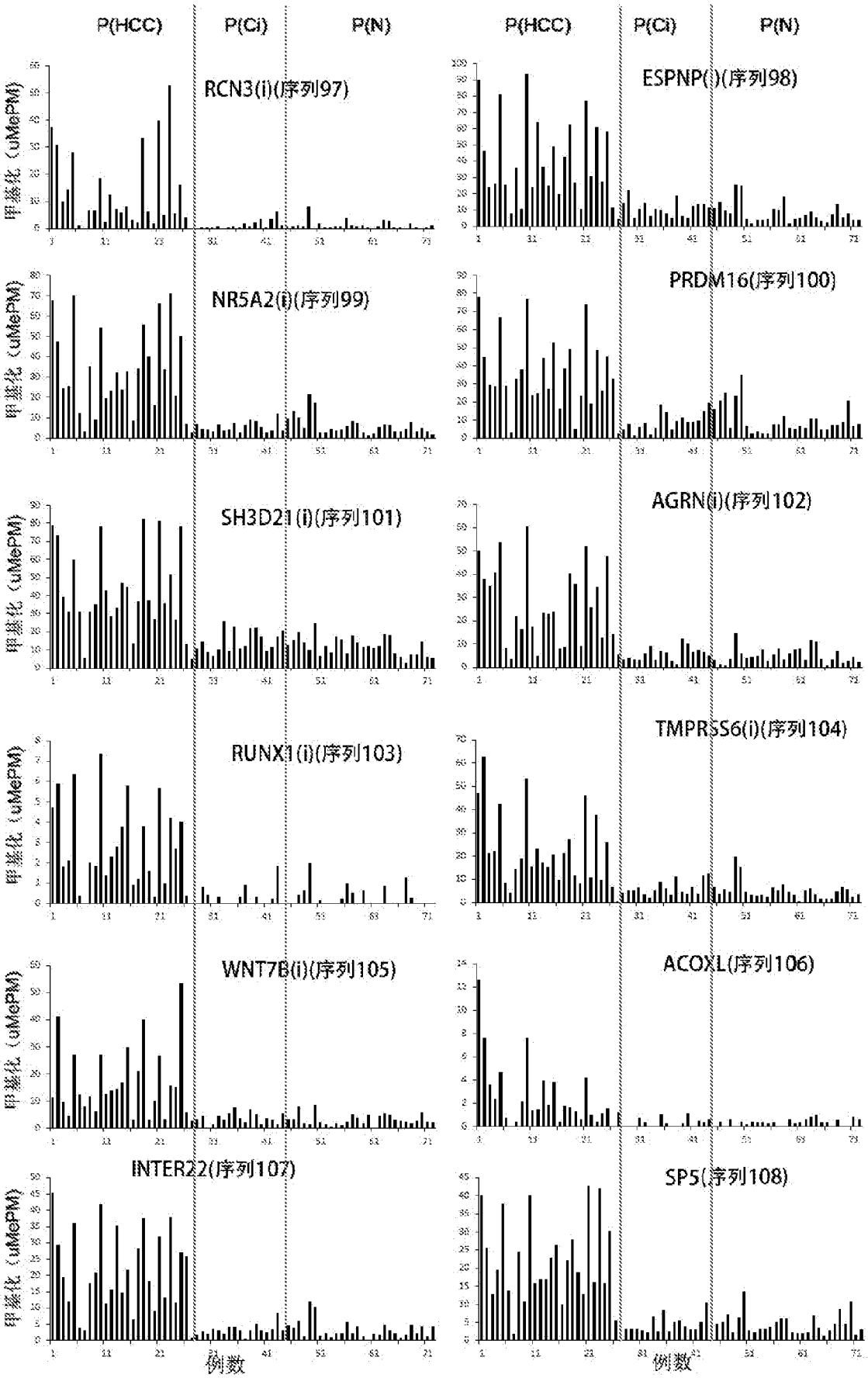

P(HCC)
P(Ci)
P(N)
RCN3(i)(序列97)
ESPNP(i)(序列98)
NR5A2(i)(序列99)
PRDM16(序列100)
SH3D21(i)(序列101)
AGRN(i)(序列102)
RUNX1(i)(序列103)
TMPRSS6(i)(序列104)
WNT7B(i)(序列105)
ACOXL(序列106)
INTER22(序列107)
SP5(序列108)
甲基化（uMePM）
例数

图 37

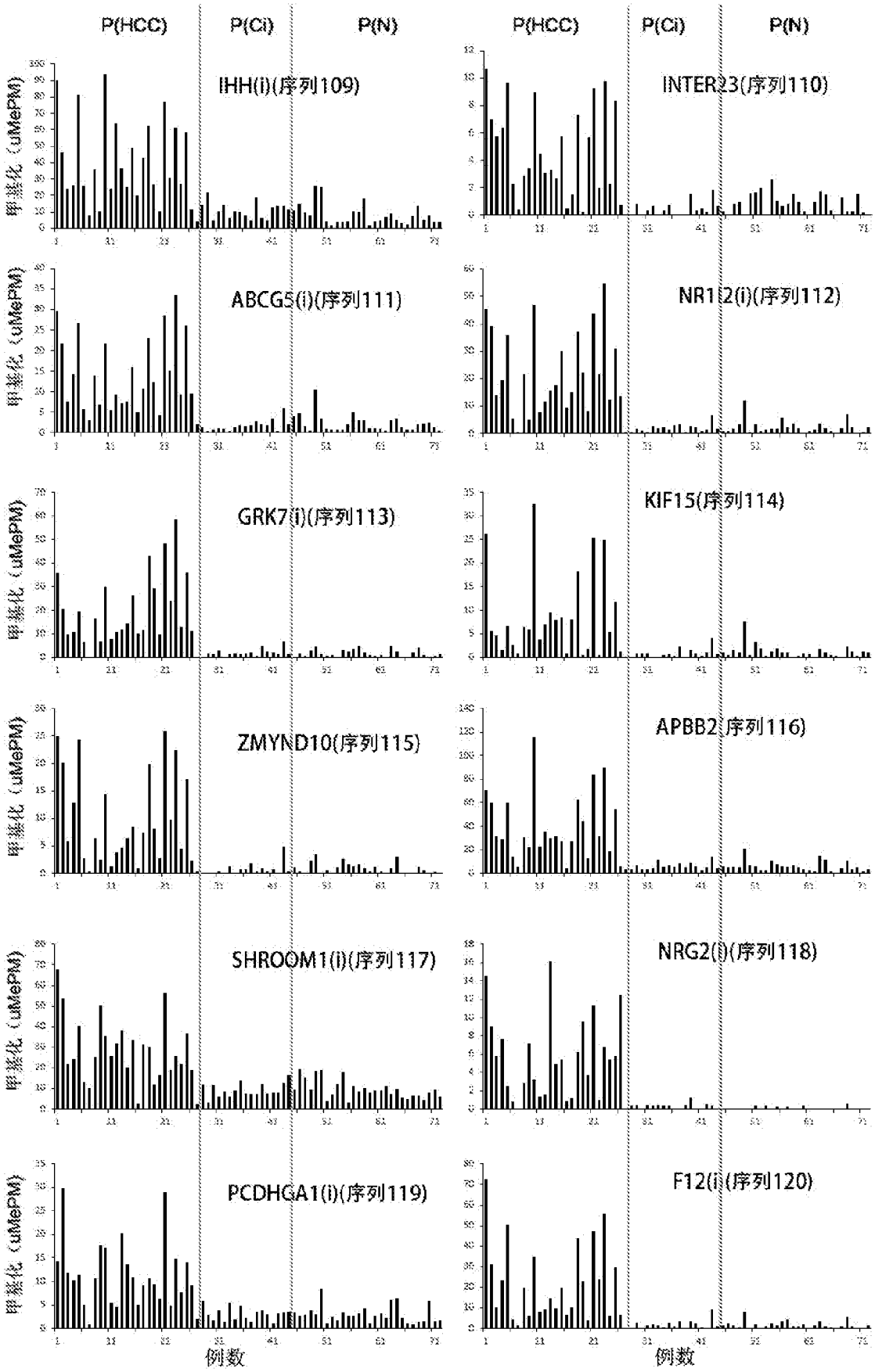

P(HCC)
P(Ci)
P(N)
IHH(i)(序列109)
INTER23(序列110)
ABCG5(i)(序列111)
NR1I2(i)(序列112)
GRK7(i)(序列113)
KIF15(序列114)
ZMYND10(序列115)
APBB2(序列116)
SHROOM1(i)(序列117)
NRG2(i)(序列118)
PCDHGA1(i)(序列119)
F12(i)(序列120)
甲基化（uMePM）
例数

图 38

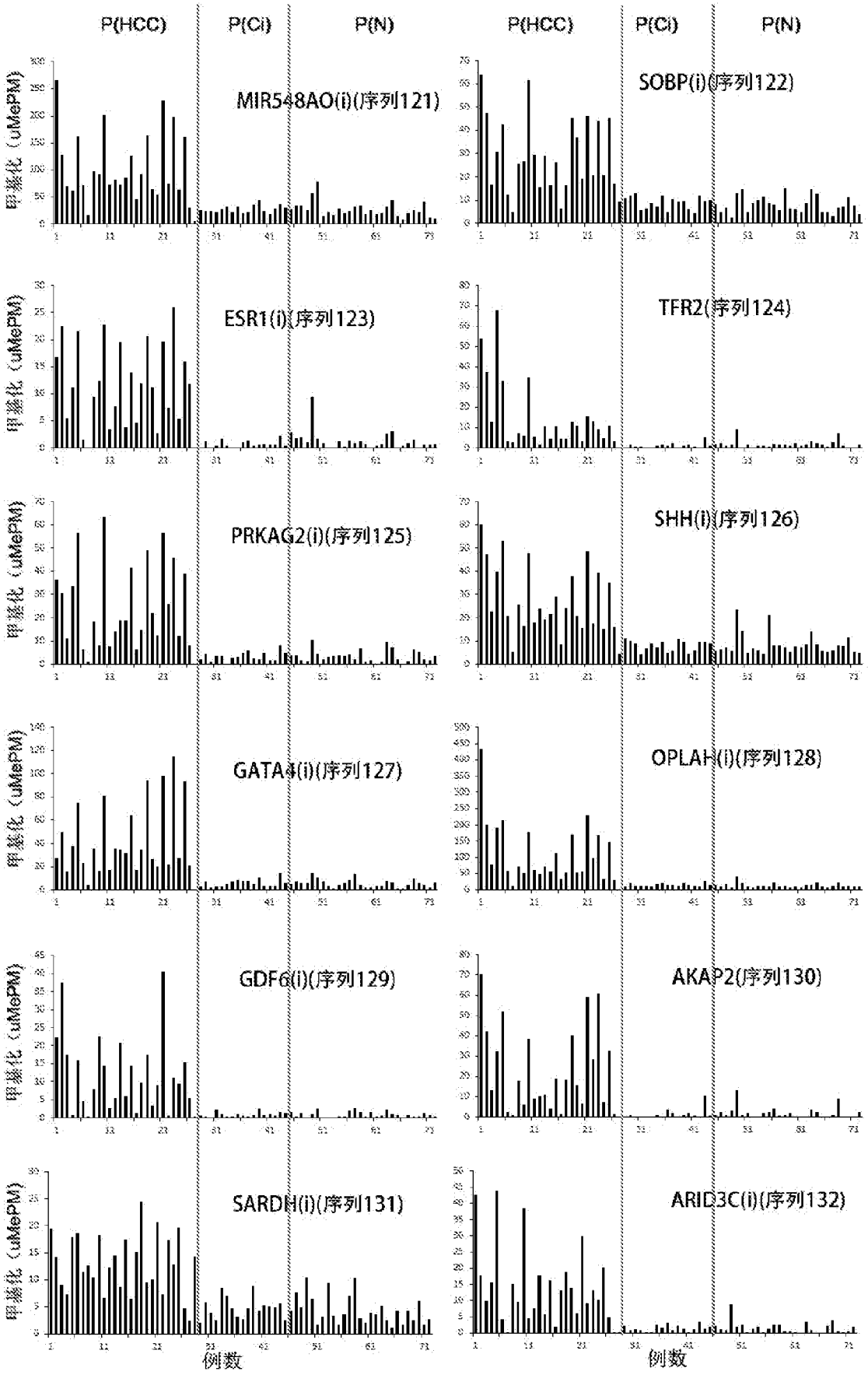

P(HCC)
P(Ci)
P(N)
甲基化（uMePM）
MIR548AO(i)(序列121)
SOBP(i)(序列122)
ESR1(i)(序列123)
TFR2(序列124)
PRKAG2(i)(序列125)
SHH(i)(序列126)
GATA4(i)(序列127)
OPLAH(i)(序列128)
GDF6(i)(序列129)
AKAP2(序列130)
SARDH(i)(序列131)
ARID3C(i)(序列132)
例数

图 39

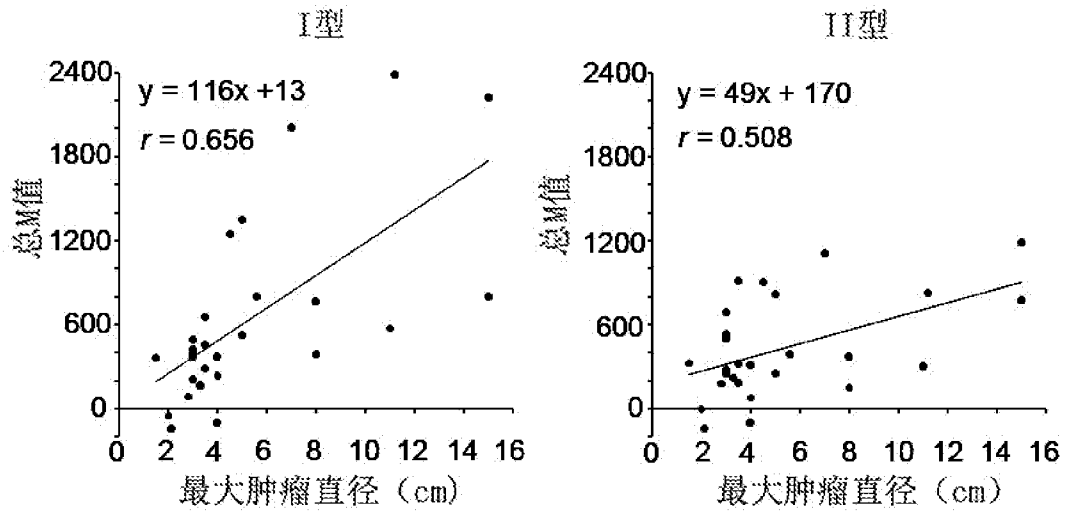

I型
y = 116x +13
r = 0.656
总M值
最大肿瘤直径（cm）
II型
y = 49x + 170
r = 0.508
总M值
最大肿瘤直径（cm）

图 40

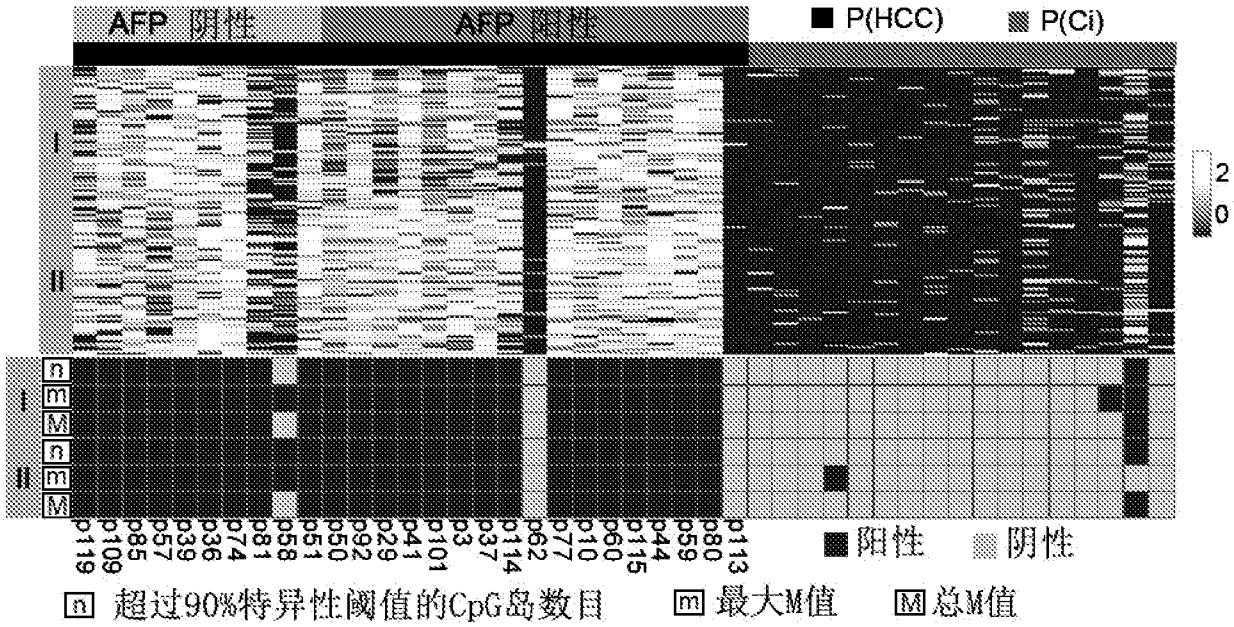

AFP 阴性
AFP 阳性
P(HCC)
P(Ci)
阳性
阴性
超过90%特异性阈值的CpG岛数目
最大M值
总M值

图 41

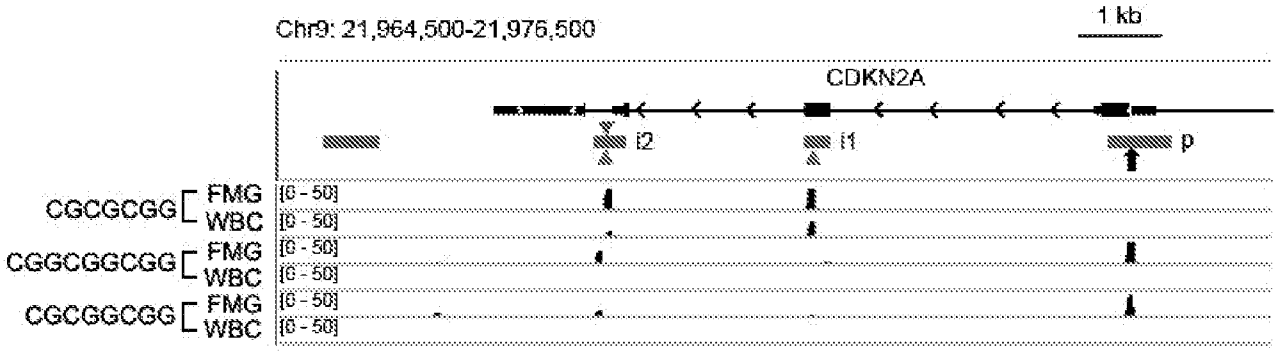

Chr9: 21,964,500-21,976,500
1 kb
CDKN2A
i2
i1
p
CGCGCGGG
CGGCGGCGGG
CGCGGCGGG
FMG
WBC

图 42

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/095120

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)

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)

CNABS, CPRSABS, SIPOABS, DWPI, CNTXT, WOTXT, EPTXT, USTXT, CNKI, GOOGLE SCHOLAR, Elsevier Science, PubMed: CpG, methylation, genome, compare, evaluate, cytosine, uracil, semi-amplification, whole amplification, SEMI-amplicon, CpG short tandem amplification, Short tandem, library

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101160411 A (EPIGENOMICS AG), 09 April 2008 (09.04.2008), see claims 1-20, description, page 1, paragraph 1 and page 24, paragraph 4 to page 26, paragraph 4	1-9, 21
A	CN 101160411 A (EPIGENOMICS AG), 09 April 2008 (09.04.2008), see claims 1-20, description, page 1, paragraph 1 and page 24, paragraph 4 to page 16paragraph 4	10-20
A	CN 101230399 A (INSTITUTE OF CHEMISTRY CHINESE ACADEMY OF SCIENCES), 30 July 2008 (30.07.2008), see abstract, and claims 1-10	1-21

☐ 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
22 September 2015 (22.09.2015)	**08 October 2015 (08.10.2015)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **ZHAO, Yanhao** Telephone No.: (86-10) **62411043**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/095120

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 filed or furnished:

   a. (means)

   ☐ on paper

   ☒ in electronic form

   b. (time)

   ☒ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purposes of search

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 in 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/CN2014/095120

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101160411 A	09 April 2008	CN 101160411 B	14 May 2014
		ZA 200708701 A	26 November 2008
CN 101230399 A	30 July 2008	CN 101230399 B	02 June 2010

国际检索报告

国际申请号 PCT/CN2014/095120

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)

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）)

CNABS,CPRSABS,SIPOABS,DWPI,CNTXT,WOTXT,EPTXT,USTXT,CNKI,GOOGLE SCHOLAR, Elsevier Science, PubMed: CpG,甲基化，基因组，比，评，胞嘧啶，尿嘧啶，半扩增，全扩增，SEMI-amplicon，CpG短串联扩增，文库，Short tandem，library

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
X	CN 101160411 A (EPI基因组股份公司) 2008年 4月 9日 (2008-04-09) 参见权利要求1-20，说明书第1页第1段，第24页第4段-第26页第4段	1-9, 21
A	CN 101160411 A (EPI基因组股份公司) 2008年 4月 9日 (2008-04-09) 参见权利要求1-20，说明书第1页第1段，第24页第4段-第26页第4段	10-20
A	CN 101230399 A (中国科学院化学研究所) 2008年 7月 30日 (2008-07-30) 参见摘要及权利要求1-10	1-21

☐ 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
2015年 9月 22日	2015年 10月 8日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国 传真号 (86-10)62019451	赵彦豪 电话号码 (86-10)62411043

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告

国际申请号
PCT/CN2014/095120

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对要求保护的发明必要的核苷酸和/或氨基酸序列，国际检索是在下列基础上进行的：

a. (提交提供)

☐ 纸件形式

☑ 电子形式

b. (提交时间)

☑ 含在申请提交时的国际申请中

☐ 以电子形式与国际申请一起提交

☐ 为检索之用随后提交本单位

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申请中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/CN2014/095120

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	101160411	A	2008年 4月 9日	CN	101160411	B	2014年 5月 14日
				ZA	200708701	A	2008年 11月 26日
CN	101230399	A	2008年 7月 30日	CN	101230399	B	2010年 6月 2日