(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115087744 A
(43)申请公布日 2022.09.20

(21)申请号 202080092232.8
(22)申请日 2020.11.06
(30)优先权数据
62/931,411 2019.11.06 US
(85)PCT国际申请进入国家阶段日
2022.07.06
(86)PCT国际申请的申请数据
PCT/CA2020/051507 2020.11.06
(87)PCT国际申请的公布数据
WO2021/087615 EN 2021.05.14
(71)申请人 大学健康网络
地址 加拿大安大略省
申请人 范安德尔研究所
(72)发明人 萨曼莎·L·威尔逊 沈淑怡
丹尼尔·迪尼兹德卡尔瓦霍
迈克尔·M·霍夫曼
蒂莫西·J·特里谢
(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 贺淑东 武晶晶
(51)Int.Cl.
C12Q 1/6806(2018.01)
C12Q 1/6869(2018.01)
C12Q 1/6804(2018.01)
C12N 15/10(2006.01)

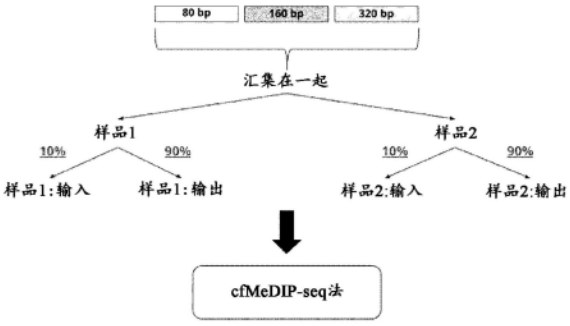
权利要求书3页 说明书33页
序列表17页 附图32页

(54)发明名称

用于无细胞MeDIP测序的合成加标对照及其使用方法

(57)摘要

本文描述了一种捕获并分析样品中无细胞甲基化DNA的方法。该方法包括对样品进行文库制备以允许随后对无细胞甲基化DNA进行测序。将预定量的对照合成DNA片段加入样品中。对照合成DNA片段各自具有不与目标基因组序列对齐的已知核酸序列，并且至少一些对照合成DNA片段被甲基化。使样品变性，并利用对甲基化多核苷酸具有选择性的结合剂捕获无细胞甲基化DNA和对照合成DNA片段。对捕获的DNA进行扩增和测序。



CN 115087744 A

1.一种捕获并分析样品中无细胞甲基化DNA的方法,所述方法包括以下步骤:

a.对所述样品进行文库制备以允许随后对所述无细胞甲基化DNA进行测序;

b.添加预定量的一组对照合成DNA片段,其中所述对照合成DNA片段各自具有基本上不与目标基因组序列对齐的已知核酸序列,并且其中所述一组对照合成DNA片段中的至少一些所述对照合成DNA片段被甲基化;

c.使所述样品变性;

d.使用对甲基化多核苷酸具有选择性的结合剂捕获无细胞甲基化DNA和所述对照合成DNA片段;以及

e.对所捕获的无细胞甲基化DNA和所述对照合成DNA片段进行扩增和测序。

2.根据权利要求1所述的方法,还包括:

f.基于经测序的所述对照合成DNA片段计算所述样品中所述无细胞甲基化DNA的量、浓度或重量摩尔浓度。

3.根据权利要求1或2所述的方法,其中所述一组对照合成DNA片段包含3至7个预定片段长度。

4.根据权利要求3所述的方法,其中所述一组对照合成DNA片段包含3个预定片段长度。

5.根据权利要求1至3中任一项所述的方法,其中所述对照合成DNA片段的长度为50至500个碱基对(bp),优选长度为80bp至320bp。

6.根据权利要求4所述的方法,其中所述3个预定片段长度分别是80bp、160bp和320bp。

7.根据权利要求1至6中任一项所述的方法,其中所述对照合成DNA片段具有0%至100%、优选25%至75%的鸟嘌呤和胞嘧啶(G+C)合并含量。

8.根据权利要求4所述的方法,其中所述3个预定片段长度分别具有35%、50%和65%的G+C含量。

9.根据权利要求1所述的方法,其中所述对照合成DNA片段中的每一个包含一数量的CpG二核苷酸,所述数量的范围在按碱基对计的所述片段长度的0到1/2之间。

10.根据权利要求8所述的方法,其中所述对照合成DNA片段中的每一个包含1至25个CpG二核苷酸,优选1至16个CpG二核苷酸。

11.根据权利要求1所述的方法,其中所述对照合成DNA片段具有SEQ ID NO:1-59中一个或多个所示的核酸序列。

12.根据权利要求1所述的方法,还包括:

i.对所述捕获的无细胞甲基化DNA和所述对照合成DNA片段进行测序;

ii.将经测序的无细胞甲基化DNA与所述对照合成DNA片段的已知核酸序列进行比较;以及

iii.将来自(ii)的任何不匹配的DNA与所述目标基因组序列进行比较。

13.根据权利要求1至12中任一项所述的方法,其中所述一组对照合成DNA片段中一半的所述对照合成DNA片段是甲基化的,而另一半是未甲基化的。

14.根据权利要求1至13中任一项所述的方法,其中所述一组对照合成DNA片段包含第一甲基化序列和第二未甲基化序列。

15.根据权利要求1至12中任一项所述的方法,其中所述一组对照合成DNA片段中的所有所述对照合成DNA片段都是甲基化的。

16.根据权利要求2所述的方法，包括在扩增前使用独特分子标识符(UMI)衔接子估计所捕获的无细胞甲基化DNA的量。

17.根据权利要求1至16中任一项所述的方法，其中所述结合剂是包含甲基-CpG结合结构域的蛋白质。

18.根据权利要求17所述的方法，其中所述蛋白质为MBD2蛋白质。

19.根据权利要求1至18中任一项所述的方法，其中步骤(d)包括使用抗体对所述无细胞甲基化DNA进行免疫沉淀。

20.根据权利要求19所述的方法，包括向所述样品加入至少0.05μg、优选至少0.16μg的抗体用于免疫沉淀。

21.根据权利要求20所述的方法，其中所述抗体是5-甲基胞嘧啶抗体、5-羟甲基胞嘧啶抗体、5-甲酰基胞嘧啶抗体或5-羧基胞嘧啶抗体。

22.根据权利要求1至21中任一项所述的方法，其中所述样品具有小于100ng的无细胞DNA，并且所述方法还包括向所述样品加入第一量的填充DNA，其中至少一部分所述填充DNA是甲基化的。

23.根据权利要求22所述的方法，其中所述第一量的填充DNA包含约5％、10％、15％、20％、30％、40％、50％、60％、70％、80％、90％或100％的甲基化填充DNA，其余为未甲基化填充DNA，优选5％至50％、10％至40％或15％至30％的甲基化填充DNA。

24.根据权利要求22所述的方法，其中填充DNA的所述第一量为20ng至100ng，优选30ng至100ng，更优选50ng至100ng。

25.根据权利要求22所述的方法，其中来自所述样品的所述无细胞DNA和所述第一量的填充DNA一起包含至少50ng总DNA，优选至少100ng总DNA。

26.根据权利要求22所述的方法，其中所述填充DNA的长度为50bp至800bp，优选100bp至600bp，更优选200bp至600bp。

27.根据权利要求22所述的方法，其中所述填充DNA为内源DNA或外源DNA。

28.根据权利要求27所述的方法，其中所述填充DNA是非人DNA，优选λDNA。

29.根据权利要求27所述的方法，其中所述填充DNA不与人DNA对齐。

30.一种鉴定用于捕获并分析无细胞甲基化DNA的对照合成DNA片段的序列的方法，所述方法包括以下步骤：

a.基于多个目标片段长度、目标的鸟嘌呤和胞嘧啶(G+C)合并含量和每个片段的CpG二核苷酸的目标数量生成核酸序列；以及

b.消除所生成的与人基因组对齐的序列；

其中所述多个目标片段长度包括在50至500个碱基对(bp)之间的3至7个目标片段长度；

其中所述目标G+C含量为0％至100％；并且

其中所述每个片段的CpG二核苷酸的所述目标数量在按碱基对计的所述片段长度的0到1/2之间。

31.根据权利要求30所述的方法，其中所述目标片段长度优选为80bp至320bp。

32.根据权利要求31所述的方法，其中所述目标片段长度是最短片段长度的倍数。

33.根据权利要求32所述的方法，其中所述多个目标片段长度包括80bp、160bp或320bp

的三个片段。

34.根据权利要求30所述的方法，其中所述目标G+C含量为25％至75％。

35.根据权利要求34所述的方法，其中所述目标G+C含量分别为35％、50％或65％。

36.根据权利要求30所述的方法，其中所述CpG二核苷酸的目标数量是每个片段中1至25、优选1至16个CpG二核苷酸。

37.根据权利要求30所述的方法，其中所述CpG二核苷酸的目标数量是每20bp、每40bp或每80bp一个CpG二核苷酸。

用于无细胞MeDIP测序的合成加标对照及其使用方法

技术领域

[0001] 本发明涉及甲基化DNA免疫沉淀测序领域,更具体地,涉及无细胞甲基化DNA的绝对定量方法。

背景技术

[0002] 甲基化DNA免疫沉淀测序(MeDIP-seq)在测量DNA甲基化方面大受欢迎。尽管无细胞甲基化DNA免疫沉淀测序(cfMeDIP-seq)对于测量高甲基化区域的DNA甲基化是稳健的,但可能存在可能影响结果的生物学和技术变异。另外,MeDIP-seq实验传统上定量了相对于实验的读段计数。这可能导致缺乏再现性并使得难以比较不同研究之间的结果。

发明内容

[0003] 根据一个方面,提供了一种捕获并分析样品中的无细胞甲基化DNA的方法,该方法包括以下步骤:a)对样品进行文库制备以允许随后对无细胞甲基化DNA进行测序;b)添加预定量的一组对照合成DNA片段,其中该对照合成DNA片段各自具有基本上不与目标基因组序列对齐的已知核酸序列,并且其中组中的至少一些对照合成DNA片段被甲基化;c)使样品变性;d)使用对甲基化多核苷酸具有选择性的结合剂捕获无细胞甲基化DNA和对照合成DNA片段;和e)对捕获的无细胞甲基化DNA和对照合成DNA片段进行扩增和测序。

[0004] 根据另一方面,提供了一种鉴定用于捕获并分析无细胞甲基化DNA的对照合成DNA片段的序列的方法,该方法包括以下步骤:a)基于多个目标片段长度、目标的鸟嘌呤和胞嘧啶(G+C)合并含量和每个片段的CpG二核苷酸的目标数量生成核酸序列;以及b)消除所生成的与人基因组对齐的序列;其中多个目标片段长度包括在50至500个碱基对(bp)之间的3至7个目标片段长度;其中目标G+C含量为0%至100%;并且其中每个片段的CpG二核苷酸的目标数量在按碱基对计的片段长度的0到1/2之间。

附图说明

[0005] 通过参考以下描述和附图可以最好地理解本发明的实施方案。在附图中:

[0006] 图1示出了一组合成加标(spike-in)对照片段的中试试验的实验设计。

[0007] 图2示出了通过加标至HCT116细胞系确定加标对照的量的实验设计。

[0008] 图3示出了片段长度的数据转换:(A.)转化前的片段长度;(B.)z评分归一化后的片段长度。

[0009] 图4示出了片段内的CpG的数量的数据转换:(A.)立方根转换前的CpG分布;(B.)立方根转换后的CpG分布。

[0010] 图5示出了cfMeDIP-seq法的DNA甲基化特异性。

[0011] 图6示出了仅对合成DNA片段进行测序的结果。图表示出了读段计数随片段长度、G+C含量和片段内CpG数量的分布:(A.)独特的甲基化输入样品的片段长度分布;(B.)独特的甲基化输出样品的片段长度分布;(C.)独特的甲基化输入样品的G+C含量分布;(D.)独特的

甲基化输出样品的G+C含量分布；(E.)按G+C含量划分的独特的甲基化输入样品的片段分布中的CpG数量；(F.)按G+C含量划分的独特的甲基化输出样品的片段分布中的CpG数量。

[0012] 图7示出了仅对合成DNA片段进行测序的结果。图表示出了读数计数随片段长度、G+C含量和片段内CpG数量的分布：(A.)独特的未甲基化输入样品的片段长度分布；(B.)独特的未甲基化输出样品的片段长度分布；(C.)独特的未甲基化输入样品的G+C含量分布；(D.)独特的未甲基化输出样品的G+C含量分布；(E.)按G+C含量划分的独特的未甲基化输入样品的片段分布中的CpG数量；(F.)按G+C含量划分的独特的未甲基化输出样品的片段分布中的CpG数量。

[0013] 图8示出了用于加标对照(黑色柱)的读段数量与用于生物样品HCT116(白色柱)的读段数量的比较。

[0014] 图9示出了cfMeDIP-seq的DNA甲基化特异性，其中在MiSeq上加标至HCT116，100万个读段。

[0015] 图10示出了cfMeDIP-seq的DNA甲基化特异性，其中在NovaSeq上加标至HCT116，每个样品6000万个读段。

[0016] 图11示出了合成的加标对照(黑色柱)与生物样品HCT116(白色柱)的总读段的比较。

[0017] 图12示出了描述与已知重量摩尔浓度(molality)相比的高斯广义混合模型的性能的Bland-Altman图。X轴：计算的和已知的重量摩尔浓度值之间的平均值。Y轴：计算的和已知的浓度值之间的方差。粗虚线：95％置信区间；浅虚线：95％置信区间边界。

[0018] 图13示出了使用合成的加标对照DNA的实验设计以(A)评估技术偏差和(B)优化合成的DNA量。

[0019] 图14示出了合成DNA的片段长度、G+C含量以及输入、输出和0.01ng加标中CpG分数的评估偏差。

[0020] 图15示出了(A)皮摩尔和标准偏差与(B)皮摩尔和可映射性评分之间的相关性。

[0021] 图16示出了计算的皮摩尔和M值之间以及读段计数和M值之间的相关性。

[0022] 图17示出了已知变量和主成分之间的关联。左)由每个主成分解释的方差比例；右)每个主成分的已知技术和临床变量之间的关联。***p<0.001。

具体实施方式

[0023] 开发了一种无细胞甲基化DNA免疫沉淀测序(cfMeDIP-seq)方法，用于处理低输入DNA和循环无细胞DNA(cfDNA)。cfMeDIP-seq法使用低输入cfDNA测量DNA甲基化，使其对于液体活检应用是理想的。从cfMeDIP-seq获得的DNA甲基化图谱有助于提供在循环肿瘤DNA研究中重要的原始组织信息。[1-6]类似于基于免疫沉淀的经典富集方案和测序方案如RNA-seq，解释需要参考或对照用于比较。参考对照由已知序列的加标参考DNA片段组成。[7-11]

[0024] 加标对照克服了DNA或RNA产量在不同实验条件下和在所有基因组区域上相等的假设。[8]结果，加标对照也调节生物学和技术偏差。加入加标对照显著改变了RNA-seq、ChIP-seq和基因组测序结果的解释。[7-11]已经表明所有全基因组分析将受益于加入加标对照。[8]通过每个样品的读段总数来归一化数据经常掩盖了感兴趣的变量的差异。归一化数据以假定参考对照DNA在样品之间是相同的，能够更精确地检测差值并调节可能影响结果的生物变

量。[8,9]尽管DNA和RNA测序实验已经使用了加标对照，但测量全基因组DNA甲基化的富集方法却未使用。

[0025] 在此，本文的发明人引入了用于cfMeDIP-seq的新合成的加标DNA对照。

[0026] 在一些实施方案中，加标对照校正了片段长度、G+C含量和CpG分数，并且可用于评估非特异性结合，这是cfMeDIP-seq分析的一个组成部分。在一些实施方案中，设计具有独特分子指数(UMI)的加标对照以调整聚合酶链反应(PCR)偏差、片段长度、鸟嘌呤和胞嘧啶(G+C)合并含量和每个片段的CpG二核苷酸(CpG)数量。这些修饰产生甲基化DNA的定量量度，而不是相对读段计数，并有助于减轻批次效应。

[0027] 加标对照用于测序方法如cfMeDIP-seq(无细胞甲基化DNA免疫沉淀和高通量测序)。cfMeDIP-seq用于使用无细胞DNA进行全基因组DNA甲基化映射。

[0028] 如本文所用，“甲基化DNA”是指具有添加的甲基的DNA，及其衍生物。例如，甲基化胞嘧啶(5-甲基胞嘧啶)的氧化衍生物是通过5mC氧化途径衍生的，并且包括5-羟甲基胞嘧啶、5-甲酰基胞嘧啶和5-羧基胞嘧啶(参见Song等人，Trends Biochem Sci.2013.10;38(10):480-484，其全部内容通过引用合并于本文)。

[0029] 根据一个方面，提供了一种捕获并分析样品中无细胞甲基化DNA的方法。该方法包括：

[0030] a.对样品进行文库制备以允许随后对无细胞甲基化DNA进行测序；

[0031] b.添加预定量的一组对照合成DNA片段，其中该对照合成DNA片段各自具有基本上不与目标基因组序列对齐的已知核酸序列，并且其中组中的至少一些对照合成DNA片段被甲基化；

[0032] c.使样品变性；

[0033] d.使用对甲基化多核苷酸具有选择性的结合剂捕获无细胞甲基化DNA和对照合成DNA片段；以及

[0034] e.对捕获的无细胞甲基化DNA和对照合成DNA片段进行扩增和测序。

[0035] 在一些实施方案中，该方法还包括基于经测序的对照合成DNA片段计算样品中无细胞甲基化DNA的量、浓度或重量摩尔浓度的步骤。

[0036] 无细胞甲基化DNA是在血流中自由循环并在DNA的各种已知区域甲基化的DNA。样品，例如血浆样品，可用于分析无细胞甲基化DNA。

[0037] 如本文所用，“文库制备”包括末端修复、A加尾、衔接子连接或在无细胞DNA上进行的任何其他制备以允许随后对DNA进行测序。

[0038] 如本文所用，“目标基因组序列”是指样品中无细胞甲基化DNA将针对其进行测序的基因组。在一些实施方案中，目标基因组是人基因组。在其他实施方案中，目标基因组是非人基因组。如本文所用，“基本上不与目标基因组序列对齐的核酸序列”是指在与目标基因组序列对齐中同一性小于30％、小于20％、小于10％、小于5％、小于3％或小于1％的序列。基本上不与目标基因组序列对齐的核酸序列含有不超过2个、不超过3个、不超过4个、不超过5个，不超过6个、不超过7个、不超过8个、不超过9个或不超过10个与目标基因组序列相同的对齐核苷酸。

[0039] 各种测序技术是本领域技术人员已知的，例如聚合酶链反应(PCR)后进行Sanger测序。还可用的是下一代测序(NGS)技术，也称为高通量测序，其包括各种测序技术，包括：

Illumina(Solexa)测序、Roche454测序、Ion torrent:Proton/PGM测序、SOLiD测序。与先前使用的Sanger测序相比,NGS可以更快、更便宜地对DNA和RNA进行测序。在一些实施方案中,所述测序被优化用于短读段测序。

[0040] DNA样品可以例如使用足够的热量进行变性。

[0041] 在一些实施方案中,该对照合成DNA片段组包含具有不同预定长度的多个片段。在一些实施方案中,该对照合成DNA片段组包含3至7个预定片段长度、3至6个预定片段长度或3至5个预定片段长度。在一个实施方案中,该对照合成DNA片段组包含3个预定片段长度。

[0042] 在一些实施方案中,对照合成DNA片段的长度为50至500个碱基对(bp),优选80至320bp。在一些实施方案中,一组合成DNA片段具有长度增加的片段。在一个实施方案中,一组合成DNA片段具有100bp、150bp和300bp的三个预定长度。在其他实施方案中,一组合成DNA片段具有最短片段长度倍数的片段。在一个实施方案中,一组具有80bp、160bp和320bp的三个预定长度。

[0043] 如本文所用,鸟嘌呤和胞嘧啶合并含量(G+C含量)是指DNA片段中鸟嘌呤或胞嘧啶核苷酸的百分比。在一些实施方案中,对照合成DNA片段具有0%至100%、优选25%至75%的G+C含量。在一个实施方案中,当一组具有三个预定片段长度时,所述3个预定片段长度分别具有35%、50%和65%的G+C含量。

[0044] 如本文所用,CpG二核苷酸(或CpG位点)是一种DNA区域,其中胞嘧啶核苷酸之后是沿着其5'→3'方向的线性碱基序列中的鸟嘌呤核苷酸。在一些实施方案中,对照合成DNA片段中的每一个包含一数量的CpG二核苷酸,该数量的范围按碱基对计的片段长度的0到1/2之间。在一些实施方案中,对照合成DNA片段中的每一个包含1至25个CpG二核苷酸,优选1至16个CpG二核苷酸。在一些实施方案中,对照合成DNA片段在每个最短片段长度上具有1、2或4个CpG位点。在一些实施方案中,对照合成DNA片段每20bp、每40bp或每80bp含有一个CpG位点。

[0045] 在一些实施方案中,对照合成DNA片段具有SEQ ID NO:1-59中一个或多个所示的核酸序列。

[0046] 在一些实施方案中,该方法还包括:

[0047] i.对捕获的无细胞甲基化DNA和对照合成DNA片段进行测序;

[0048] ii.将经测序的无细胞甲基化DNA与对照合成DNA片段的已知核酸序列进行比较;以及

[0049] iii.将来自(ii)的任何不匹配的DNA与目标基因组序列进行比较。

[0050] 在一些实施方案中,该组中的一些对照合成DNA片段是甲基化的,而该组中的一些对照合成DNA片段不是甲基化的。在一个实施方案中,该组中一半的对照合成DNA片段是甲基化的,另一半是未甲基化的。在一个实施方案中,所有对照合成DNA片段都是甲基化的。

[0051] 在一个实施方案中,该对照合成DNA片段组包含甲基化的第一序列和未甲基化的第二序列。

[0052] 在一个实施方案中,该方法还包括在扩增前使用独特分子标识符(UMI)衔接子来估计所捕获的无细胞甲基化DNA的量。

[0053] 在一些实施方案中,结合剂是包含甲基-CpG-结合结构域的蛋白质。一种这样的示例性蛋白质是MBD2蛋白质。如本文所用,“甲基-CpG-结合结构域(MBD)”是指长度为约70个

残基并结合含有一个或多个对称甲基化CpG的DNA的蛋白质和酶的某些结构域。MeCP2、MBD1、MBD2、MBD4和BAZ2的MBD介导与DNA的结合，并且在MeCP2、MBD1和MBD2的情况下，优先介导与甲基化CpG的结合。人蛋白质MECP2、MBD1、MBD2、MBD3和MBD4包含一个核蛋白家族，这些核蛋白与甲基-CpG结合域(MBD)中的每一个的存在相关。除了MBD3之外，这些蛋白质中的每一种都能够与甲基化DNA特异性结合。

[0054] 在其他实施方案中，结合剂是抗体，并且捕获无细胞甲基化DNA包括使用该抗体对无细胞甲基化DNA进行免疫沉淀。如本文所用，“免疫沉淀”是指使用特异性结合特定抗原的抗体从溶液中沉淀出抗原(如多肽和核苷酸)的技术。该方法可用于从样品中分离和浓缩特定的蛋白质或DNA，并要求抗体在程序中的某个时间点与固体基质偶联。固体基质包括例如珠，如磁珠。其他类型的珠和固体基质是本领域已知的。

[0055] 一种示例性抗体是5-甲基胞嘧啶抗体。对于免疫沉淀程序，在一些实施方案中，向样品加入至少0.05μg抗体；而在更优选的实施方案中，向样品加入至少0.16μg抗体。为了证实免疫沉淀反应，在一些实施方案中，本文所述的方法还包括在步骤(b)后向样品加入第二种量的对照DNA的步骤。

[0056] 其他示例性抗体是5-羟甲基胞嘧啶抗体、5-甲酰基胞嘧啶抗体和5-羧基胞嘧啶抗体。

[0057] 在一些实施方案中，样品具有小于100ng的无细胞DNA，并且该方法还包括向样品加入第一量的填充DNA，其中至少一部分填充DNA是甲基化的。填充DNA由大小与衔接子连接的cfDNA文库相似的扩增子组成，并且由不同甲基化水平的未甲基化和体外甲基化DNA组成。添加这种填充DNA具有实际应用，能够将输入DNA量归一化为100ng。这确保无论可用cfDNA的数量如何，所有样品的下游方案都保持不变。

[0058] 如本文所用，“填充DNA”可以是非编码DNA或其可由扩增子组成。

[0059] 在一些实施方案中，第一量的填充DNA包含约5％、10％、20％、30％、40％、50％、60％、70％、80％、90％或100％的甲基化填充DNA，其余为未甲基化填充DNA。在优选的实施方案中，第一量的填充DNA包含约50％的甲基化填充DNA。在一些实施方案中，5％至50％、10％至40％或15％至30％为甲基化填充DNA

[0060] 在一些实施方案中，第一量的填充DNA为20ng至100ng。在优选的实施方案中，填充DNA为30ng至100ng。在更优选的实施方案中，填充DNA为50ng至100ng。当来自样品的无细胞DNA和第一量的填充DNA组合在一起时，包含至少50ng总DNA，优选至少100ng总DNA。

[0061] 在一些实施方案中，填充DNA长度为50bp至800bp。在优选的实施方案中，长度为100bp至600bp；在更优选的实施方案中，长度为200bp至600bp。

[0062] 填充DNA是双链的。填充DNA也可以是内源或外源DNA。例如，填充DNA是非人DNA，在优选的实施方案中，是λDNA。如本文所用，“λDNA”是指肠杆菌(Enterobacteria)噬菌体λDNA。在一些实施方案中，填充DNA不与人DNA对齐。

[0063] 在其他实施方案中，提供了鉴定用于对照合成DNA片段的序列的方法。然后将对照合成DNA片段用于捕获并分析无细胞甲基化DNA。该方法包括：

[0064] a.基于多个目标片段长度、目标的鸟嘌呤和胞嘧啶(G+C)合并含量和每个片段的CpG二核苷酸的目标数量生成核酸序列；以及

[0065] b.消除所生成的与人基因组对齐的序列；

[0066] 多个目标片段长度包括3至7个不同的目标片段长度，这些长度是单位长度(也是最短片段)的倍数。片段长度为50至500个碱基对(bp)，优选80至320bp。目标G+C含量为0％至100％，优选25％至75％。每个片段的CpG位点的目标数量在按碱基对计的片段长度的0到1/2之间，并且每个最短片段长度1、2或4个CpG位点。在一些实施方案中，每个片段中CpG二核苷酸的目标数量是1至25、优选1至16个CpG二核苷酸。在一些实施方案中，对照合成DNA片段每20bp、每40bp或每80bp含有一个CpG位点。

[0067] 在一个实施方案中，该方法产生含有三个目标片段长度的核酸序列，这三个目标片段长度分别为80bp、160bp或320bp，并且目标G+C含量分别为35％、50％或65％。

[0068] 以下实施例说明了本发明的各个方面，而不限制本文公开的本发明的广义方面。

实施例

[0069] 方法

[0070] 为了满足cfMeDIP-seq实验中对参考对照的需要，设计了加标对照，其中整合使用独特分子指数(UMI)来调整聚合酶链反应(PCR)偏差以及由DNA片段的片段长度、G+C含量和CpG密度引起的免疫沉淀偏差。这使得能够以皮摩尔对甲基化DNA进行绝对定量，同时保留允许进行灵敏、组织特异性检测并且不同实验之间有可比结果的表观基因组信息。结合甲基化状态(甲基化和未甲基化)、碱基对(bp)片段长度(80bp、160bp、320bp)、G+C含量(35％、50％、65％)和片段内CpG分数(1/80bp、1/40bp、1/20bp)，设计了54个DNA片段。检查加标对照DNA序列以确保它们与人类基因组没有交叉对齐，并最大限度地减少二级结构的形成以避免扩增问题。在单独的加标DNA片段、加入到剪切的HCT116基因组DNA中的加标DNA或加入到来自急性髓性白血病(AML)患者样品的储存血浆的cfDNA中的加标DNA上进行cfMeDIP-seq，以分别评估技术和生物学偏差、确定实验所需的加标DNA的最佳量和评估批次效应。

[0071] 设计合成DNA加标对照.设计具有独特分子指数(UMI)的加标对照以调整聚合酶链反应(PCR)偏差、片段长度、G+C含量和每个片段的CpG数量，从而允许绝对定量而不是相对读段计数。

[0072] 使用先前由cfMeDIP-seq方案[6]产生的配对末端测序数据来评估不同性质的cfDNA以帮助设计合成对照[12]。CpG的数量被评估为按碱基对计的片段长度的整数(即80bp片段中的1个CpG与160bp片段中的2个CpG相当)。利用包括片段长度或大小、G+C含量和CpG分数的cfDNA性质的分布，设置如表1所示的以下加标片段参数。将片段内的CpG的数量设置为片段长度的整数[1/80、1/40、1/20]。

[0073] 表1.合成的加标对照片段的参数。

片段长度	80bp	160bp	320bp
G+C含量	35％	50％	65％
CpG	1/80	1/40	1/20

[0074] 首先，使用27种不同的一阶马尔可夫模型来生成具有这些精确参数的序列[5]。使用GenRGenS v.2.0生成序列。[13]使用BLASTn以确保不与人参考基因组(GRCh38/hg38)对齐，选择具有可能的最高E值的序列。使用Integrated DNA Technologies(IDT)UNAFold™软件(IDT，Coralville，USA)检查80bp和160bp片段的二级DNA结构。[4]UNAFold不支持超过280bp的序列，因此，使用RNAstructure™软件[8]检查320bp片段的二级DNA结构。对于每个马尔可夫模型(N＝27)，生成了许多序列，从每个模型中挑选两种满足缺乏与人类基因组对齐和缺乏

潜在二级结构的标准的序列。为每个参数组合设计两种不同的片段序列:一种是甲基化的,一种是未甲基化的,以评估与5-甲基胞嘧啶结合的特异性。使用52个合成DNA加标对照[(9(80bp)+8(160bp)+9(320bp))x 2＝52]来评估cfMeDIP-seq法中由于片段长度、G+C含量和CpG数量的变化引起的偏差。

[0075] 合成片段制备.80bp和160bp片段订购为4nmol Ultramer™ DNA Oligo,320bp片段订购为gBlocks Gene Fragments(Integrated DNA technologies,Coralville,USA)。序列如表2所列出。使用高保真2X Master Mix(New England Biolabs,Ipswitch,MA,USA,Cat M0492L)在其确定的最佳退火温度下对每个片段进行PCR扩增(参见表2)。使用QIAQuick PCR Purification Kit™(Qiagen,Hilden,Germany,Cat 28104)纯化扩增的片段。通过Nanodrop™测定浓度。对于每个甲基化片段,取1μg合成DNA片段用CpG甲基转移酶(M.SssI)(ThermoFisher Scientific,Waltham,MA,USA,Cat EM0821)进行甲基化。

[0076] 甲基化反应在37℃孵育30分钟,然后在65℃孵育20分钟。使用MinElute PCR Purification Kit™(Qiagen,Hilden,Germany,Cat:28004)纯化甲基化产物。为了测试片段被正确甲基化,用依赖于片段的HpyCH4IV、HpaII或AfeI限制酶消化原始PCR扩增子和甲基化PCR扩增子(表2)。当PCR扩增子在2%琼脂糖凝胶上运行时具有单一条带时,认为甲基化得到了验证。一旦验证了甲基化片段,使用Qubit测量合成片段的摩尔量。

[0077] 表2.合成DNA加标对照的序列。

名称	SEQ ID No.	序列	PCR 产物的退火温度(℃)	甲基化?(是/否)如果是,则使用限制酶确认甲基化
80b_1C_35G-2	1	TGTCTAAATTAAAGTTGTGATCTTTGACTTAGCATCGACTCACCCTATAGCCTACCAGACAAGAATTATGAAGAACATAT	50	
80b_1C_50G-2	2	GTACACCATCATTATCCTCATAGCTTAGTCTCCCGCAGGCCCAGGGTACATAAGGCTTGGAGATTCACTGTTAGCTGCTC	60	
80b_2C_50G-2	3	GCCTCCCCAACTATAGGGTCAGGAAGGATTATGGCACCCCACACGATTTTCACCCGATCTGTACCAGTAATCATACATGG	60	
80b_1C_65G-2	4	GCTACCAGTGGCCCCCCCCTACCGAGTCCCCCATTAACCTCACCCCCCTGACTGCTAACCTGGGATGGTGAAGCCTGGGC	60	
80b_2C_65G-2	5	GGTTATGCCCCCGCCCTGCATCCTCCCTGTCTACACGGCCCAACCCTAGCAATGTGTGGCCCCCCCTGCTGT	60	

		CTCCCATC		
80b_4C_65G-2	6	GCTGGTGCACCGCTGCCCCCACCCACCTCGCTTGTCACAGCCTCGGTAGGTCCTGATTTGATGCTTGGGTGCTCGGCTGG	60	
160b_4C_35G-2	7	GTATAATCATAACAAAGGCCTAATGAAAGACGCTGATTTGAAACTAGTTCCCTCATCATCTGATAGATTTCCTCGTGTCTTTTTTCGTGAATGGCACAATATGGTGTGAAGACCTATTACAATCAAAAAGTATAAACTAGCGACTAAGATCTCAGAATTA	60	
80b_1C_35G-1	8	TAGGATATAGGTTGTCCCCTAGTAGGAGATAAACTTTGATTAACATCCAATTGATCGTTAGTGTCCTTCAAAATTATGCT	60	否
80b_2C_35G-1	9	TCTAATACTCATCTTAGCTCGCGTGCTTTGTGATTTTAGTGCTGAAATTCTTAAATGTTAACCACTGTGAAATCCATAAG	60	否
80b_2C_35G-2	10	CTCAAATATAACAAGAGTAGCAAACTTACAAAGATCGCTGACAAGTATGTTATCCATTTCTAAGCGCTACCAATAACACT	60	是 AfeI
80b_4C_35G-1	11	AAGGCATTACTTATCTAATCAATCGACAAAACGTTAAGTCAGTGTTAGGATAGTGTCATTTGTACTCGTAGACGAAATTG	60	否
80b_4C_35G-2	12	TTATTATTGACCGTACACTATTTAACTAACAGATATGACGTATTACTATGATATGTTAATGACGCTGAGCTGCTCGGAGA	60	是 HpyCH4IV

80b_1C_50G-1	13	GAGGACCATATAGCTCGCACAGGAACCAGCTGAAGAATTGATTGGTAGTGCTGACCAGACACCAACCTTCAAACCTCTGC	60	否
80b_2C_50G-1	14	ACAACACCCTCCACCCAATACTTGTGAGTTGGTCGCAGCACGAGCCTAGTCTCCTTGTAAGTCAGTCAAATGCCTGTAAC	60	否
80b_4C_50G-1	15	AGTCATCAGCATATTGTCAGTACCCAGTGGTCTCTAGGAAAATCGGCCGGTACGTAAATACTCCTAGTGGGCTGCGTGGT	60	否
80b_4C_50G-2	16	GCTTCTTATGATACCAAAGTTGCCCAAAAAGGCTAGCGTTCTAGTTAGGGTGCAGCCGCGAGAAGACCGGGTTCATGAAG	60	是 HpaII
80b_1C_65G-1	17	GCAGGCTGCAGGGTTTGGCCCCTGGTTCCGTTCCAGCAGGTGGCATAGGTGGGGAAAGCCAGGTGCCTACAGTGGGGTGG	60	否
80b_2C_65G-1	18	CACCTTGAGACCTCCAGAGGGGGATCCACAACTTGCGCCCTCTGTGAAGTAGGCTCTGGTGCGCAGGGGGGAAGGGGGGC	60	否
80b_4C_65G-1	19	TTGGGAGGCTCTGGACTGGGGCAAACGACACCGTGCATCAACTGTGTGGTGGTGGCCTCGTCCCCCCCCATCCTCTCCGC	60	否
160b_2C_35G-1	20	TTAGTCGAGATTTTAGCCTAATTGAGAGATAGTCCGATGATATGTCTTTGATCTAACATGTCATCATGAAATATGAAGCCAACACACTCATATGTTCATGTGACAAA	60	否

		AGATCCAGTTAAGCCAGT ATTGAGGTTTGTCCATCA CTTAAGTACTATTCATT		
160b_2C_35G-2	21	CTTTACTACTGAATGTAA GCTCTTGCAGAGGATCTA ACAGGGATAGAATTATG AACACGTCTGTCACACAT AACTTCAAATGCAATTTA TTAATAAGGGTCAGAATG TGTGGTATCTTTCCAGAC TTATATCATTCCCTTTACT ATAACCGATTACACAT	60	是 HpyCH4IV
160b_4C_35G-1	22	ATGTGTAAGAAATAAAA TACTGGCTCATCATCATA AACTTGTCTATAATGTCA CTATTATCACAAAGAATG CAGGTACGACACGGTATC GGCAGCAGTGGATAGCT CGTATCTATATGAATAGG GGAAGTGAATAATATGA CAATAGTATACTTTGCTT A	60	否
160b_2C_50G-1	23	TTGTAGTACAGTCTAACC ATCTTGACCCAGTAGCTC CCCATCTGATATGCTCAG TAGCTAGGGTGGCCTGAG GGAACCGGTCAAACCCA CTTATTCTGAACCCCAGA GGTATGTTATGCGCAGGA ACCTGCCTTCTATTGGTA GTGTCTTGGGTCAACAG	60	否
160b_2C_50G-2	24	GTAACATGGTTACCACTG GGACCGGACCTTTTCACC TCCACTTTCAGGGAATAG GATTCAGTCCTGTATAGC AGTGTGACACCCCAAGG CCAATTCCACCCTACATT CAATGCCTGAGTGTATGT TGGCCATTGGGTAACTAG CCGTGTCCCAACCTCAT	60	是 HpaII

160b_4C_50G-1	25	AGCCTTGGACGTGAGTCTCTGTTTCTGACCCAACTGAGATCTTTTTACTGTCATTCTACCCCCTAGAGACTCGCGTTTCTAGAGAGGGGATGTATGTGAGGGGTTGTGATTTAGCCCGTGATGCCCTAGGATCTTGAGACAATTGTCAGGGCCCTCCAGT	60	是 HpyCH4IV
160b_4C_50G-2	26	GGCTCTAGGGGGTGATAAAGTCTCGGATTATGCTGTATGGAGTCCCATCAACTTCCAATGACAATATTGTACTCTAGGATAGCTAGATGACGCCCCAGGCAAAGAACCCTTTTCGTATGAGGCCAGCCTTCCAAGGTCCACTAGGCTCAGCTCCTCGATG	60	否
160b_8C_50G-1	27	GGTAAGTATGCAGCTCAACGAGGGTACCGGTAGCGACCCGCTGTTTGTTACTAGTAAGGACTCAGTATTGCGCTCTACTTGGTTCCTCATGACAGCTATGCAGGGATGTGTTCAGCCCGTTCTACCGAACCCTTCTAACATGAGCGTGCCTTTTGATTAG	55	否
160b_8C_50G-2	28	GCGAGTAACTGCTTCAATGGGACTACAATGTGCCACGGGTGCCCTACAGTCCTCAGCCCCAATTGCCCAAAACGAACCTTTCAACATCATCCCGGATTTTCACTCGAAGATTGTGACTGGGGGTTTTATGCAACAACCGAGCTATTACATGGTTGCGCGT	55	是 HpaII
160b_2C_65G-1	29	TTCTCAGGCAGCCCACCCCGGCAGTCCAGATCTAGCCCCCTCCCCTTGGTACTTGGGCATGGTGAGCCTCCGAGACCCCCTCCCTCTCCC	50	否

		CCCTCACCAGACCCCCCC CTATAGGTCCTGCAAGGT GCCTTCCCAAACACCCCA GTTAGGCATGGCCACC		
160b_2C_65G-2	30	GACTCCTCCCTAGGCCCC CATGGAGCCACCCCCTCA GGCCACTCCAGGCTACTA GGCCCAGGTTCCAGGCA AATGCCCTCTCTGCCAGT GCCACTAGCAACACCTCC CCTATCAAGGTGGCCCCA GGTCCTCACGTAGCATGC AGGCCCCCCGCTCCATC	60	是 HpyCH4IV
160b_4C_65G-1	31	CCCCAGAGGCAGGTGCC CTACCAAGCTCCCCCCAT GACCCCTAAATCCCCCAC CCTGCCCCGGCGGTTGCA GTGGTACCAACCAGTCAG GCCCCTCGCCAGTACCCT TCCATATCTTCAGCCTCC TGGCCATTCGATCAGGAG CCCCACAGCCCTAGGCC	60	是 HpaII
160b_4C_65G-2	32	TAGGGCCCGAGCCAGCCT GTACCTTGCGCCCCTGCC CCCCCTCTACCTGGGGAC CCCACGGTCATCCTTGAC AGGGTGCCCCTCGGCCCA CTCCCATTCTCCTTTTGTC TCCAGTAAACCCCCAGAG CCCAAGGTCAGCCTGCTG CAGGGTTTGCCTCCA	60	否
160b_8C_65G-1	33	ACTGCTGCGCGGCGCACC TCCCACATGTCCTACCCA TCACCTCCTCAGTGTTCA CTGGCTGGGTCTGTCCTC CTACAGGGTGCCAAGCG GGGCTCCATTGCCACTAG AAGCCCATGGTCCAGCGT GGCTAGATCCGAGCGGG GGGCCTCCACCAGCCGTC	60	否

160b_8C_65G-2	34	TGGAGGGGCTGGGCCTGCTCCCCTAGTGCGGAATCCTGCCCTCCGGTGGCTTGCTCTTTGGGTCCACGGGTACTAGAGGGGAATTATGACCAGAGCCCTGCAGCCCCGAAGCGGGGTGCGCCACAGTCCCCACGACTCCGCCAACCTTCATACCCTGTCC	60	是 HpaII
320b_4C_35G-1	35	AAATGTATAAATTTGGTGAGGACTGTAATTCTAGTTGTACTCCTATGTCTACAAGACCATCTCCTTACTATAGTGGGATTAATAATATTGTAAATCCGGCTATGATCTTAGACAGGGAAAATGAGTTGTAACCGATTGTTAAGTATCATTTTTCCTTGAATTGACATCACCTAGCTTGTCTTAATGTTCATGAGAATTTCAGGCTAACCACAATGTCAACTATGCGACACCATGTATCATCATTTCCACTTCACAACAGAACCGGGTCATTTTGTGTATTCCCATAGATTAAATGATTAACCTTATGCCACTATAATATA	55	否
320b_4C_35G-2	36	TAATGTATAAATATGGTGAGGACTGTAATTCTAGTTGTACTCCTATGTCTACAAGACCATCTCCTTACTATAGTGGGATTAATAATATTGTATATCCGGCTATGATCTTAGACAGGGAAAATGAGTTGTAACCGATTGTTAAGTATCATAATTCCTTGAATTGACATCACCTAGCTTGTCTTAATGTTCATGAGAATTTCAGGCTAACCACAATGTCAACTATGCGACACCAT	60	是 HpaII

		GTATCATCATTTCCACTT CACAACAGAACCGGGTC ATAATGTGTATTCCCATA GATTAAATGATTAACCTT ATGCCACTATAATATA		
320b_8C_35G-1	37	AGCATAAAAAGCCTATA ACTCGATTTTTTAACATT AAGCGGTACCGTTTCTGC ATCCAATGACATAATATA TATGGGAGCTTACTATTG AGATGCACTCTTAATACG GAATTACGTTACAAGGTA GAGGGCTATGACTAGAA TTGAGCTTTATATTAGCA GAAGTGTCTTGTCCAGTA GGGTCTTGAAAAGTTATT ATGTATGGTGTTCATGAG AATCGAGGTACATTAAG GTGAATCATTTAAATCCT AGTATGGATGTTACACTT CAATGCTTTTGTACAACT AACTCGGGTGCGTAAATA TATTGAAACAATGTTTA	60	是 HpyCH4IV
320b_8C_35G-2	38	ACTGGATTGTAGCTATGC CTAGCATTCCTTCTCTTG AGCCTCAAAGTCTTATCT GATGTTCATTCCAACACT CTTGAACGGATTTTAAAA ACAAATATGTATAACCGA CGCAAGAATTTAATATAT GAATAAGTCTCCTGTTCT AGATTTAATCTCAATAGT GATTATCGAAAATTAGTA TAGATTTAGTGAGAATAG AGATGTGTTCCTTCCTTA ATAGTCTTTAATCTTGGA CATGGTGAGCAGTATTAT ATCCGACTGTTAAGTCGA GACTGTCTTTCGTAATGT GACGCCTTACTTTTTGAG ATAGAGAACCTAAG	60	否

320b_16C_35G-1	39	TACTATAATTAGGCAGGTGATTGAAGAACCTGTTCTTTAACTATCATATACCAGTACATACTTCAACTATTTTTGTTGATCAGGATCTATTAACGAATCACGTTTGACTTAATTTACAACTTGCTCGCGGTATTAATAAATAGATTTTATTTGCCGTGTTTTCAGTACGTATAATGCGATCAAGCAGCAATGCCGGAACCTGTTAATTGTCCTCGCCTTACTGATATGTTTAAACTTCAACTTTTCGTCTCGAGGTTTAGTAAAATAATGTATTAAAGAATACGAACGTGATTATCCCGACGGCACACTGTCACTTCGTCT	60	否
320b_16C_35G-2	40	GAACAACTTATCTGAGAACAAGACTACGTCAACTTTTGTACGTGGGATAAGTTTTCCTGAATTCTAATTATAAATATGGACGTGCTCAATGAATTAACAGACGCCACACGACGTTTATATCGGACTGATTACAAGTTTATTCTGTGTAAGTAACAACAACGCTACGATCTCTGTATGATTGAATACAGTCAAACGGTTCTATAATCACTACTCATTCTACTGTTCGAACTAATATCAATCAAGTCGTTAATAAATCTTCATGATCACTCGCTATTTCTAGAAGTGACTCGCTAAGGACGATAATTATTCTTCGTTCAAAA	55	是 HpyCH4IV
320b_4C_50G-1	41	GCTAACACCATGGCTGCTAGAATTAAAGTATTGACTGACTCATACGTGGAATACCCAAGCCAACCCTGTCCCCTCAACTAAAGGTGATTC	60	否

		TGGACCTTTCAAGGGTTG GCAGGTATTTACCCCTAT CGACAAGAAGAGTGACC TACAGGAGAAATCGATC AGAGGCAGTTCAAGATC AACTCCCTGGTCCTCCTG GCCTGGAGCTCATGATGA AGAGTCCAGTGCCTCCTG CTCACCAGCATCCCATGT GACGCAGGTCACTAGGC CTTGTGGCCTTAAAAAGC CCAGCACATACTGACTAG GGCAGTTCAGCTTATACA		
320b_4C_50G-2	42	GGCTCACCAGTGGCTCCC CATCCGATCAAGCTACCC AACTAATGTACTCAAGGT ACATAATTAAGGTAGAA CCACCGAAAGTCTACTAG TGATGTTCACTCTGTCCC TGGGATAAGAGTAGCAT AAGGCCACTAAGCTCCAC TACCTCAGCCAACGTAAT GTCTCTTTCCAGCGTCTG TTCAAAATGCTCTTGGTA GTGGTTCTGCTAGGTAAG GGCAGTTCCTTTGCCAGG GTCTAGACCTAGCCCAGT GTGGTTACCCCCATGAAC AACAGGGTGGAGGCAGG GTCAACCCAGCTACTGGC CATAATTTCTAGAGTCA	60	是 HpyCH4IV
320b_8C_50G-1	43	GCCACTGGGGACCACCTC ATTACCAGGCTTTGGCAT GCTGTAATGTCTCCGATC CTTGAGATGCCCTGCCCC CAATCGCAGATGTCAGTA GGCAGCTAGCACAACTG AAACTACTCCACCCCAAC CGTGATCCTGGTGCAAGG CTTCCCAAAGAAAATATT TAACTAAATACAGAGAT	60	否

		GCTACCTAAATCCCGCTG GGAGTTAATCAGATTCGA TCCGCGACCCCCTTTGGT GTAAGAGTGTAGCTTGCT TCTTATACCCTCTCCCCG CTCCACAATAGGAGCCTA CTTCACCACACCAATAAG GTGAACATCCTATCTC		
320b_8C_50G-2	44	AGGGAGATCTAGCCTGG CTAAGCAGGGGCAACTA GTGCACTTCTTTCCACTC CAGCGCACTTCACCATTA ATGGATGTACAGATGATT GTTAGTTTGACTCTCATC AGCAATTCACTCCCTACT TACGTTTGTGGCCCCTTA CGTCTAGATATGGGGTCC GAGTAGCCCGACTGCTTT CCTCATCAGTTTTGGCAG CCTGCAACCAAACCCCAG TAGATAAAGGCAGTGTG CTACACGTCCGGGGGTAA GAAGCCTGGGTTCCTTTA ACTAAGTACTCCAACCAC ATTACAGGGGCATCCCGC TCAGTAGTTGATGGTCA	60	是 HpyCH4IV
320b_16C_50G-1	45	AGCGGGCCTAGGTTACAC CCCCGCAACATTTCTAAA TGCATCTAGGGACCTTAC TGGCACAGTTCAGCCCGC CAGTTAATTATCTTATTG AGATCCTGCAGAGGATA ATCTCCTTTCGGAATTAC CACAACGGATTCGCGTAA ACTAGTCTGCGATTGCTT TGTAAGCCAAGGGCTAC ACTGTATCCAGGGGCTGG AGGGTTTAGAAGTTTCCG TTCCTCATTTCCGAAACT AGACCTGATAGGCTGTCG TGTCCACGGATTCTACCA	60	否

		AGCATTCGCTCCGTACCC ACTTGCTACCGACTGTTG CCGTTGCCTACAGGTA		
320b_16C_50G-2	46	ATACCTCAGTTTATATTG GACCTCTAGCTGCCGGTT AGAGAAACTATTAGAAA GCACGGTTCTATACGGAC ATTCTTGGGCTGTATGAT ACAAACAGAGGCCGGTA CGACTACTTCCGCTCCAT GTAATTGACGAAGAGTCT TCTCCCTAGAGATCGCTT ATCCTTATTGCTAATGCT TGCCATAGGCCCCGCTTA GCACGACTGCCGATACGT GAATGCTATTGAGTACCG GCCCAGTCGTCCCGCATC TCCCACACTGAACCGGGT TCTCAGCTATGTCAACTG TCCTTCTTGTCTTTGGGTG CAGGGTCACCTGCCC	60	是 HpyCH4IV
320b_4C_65G-1	47	TGTATGATTTGAAGAGAT TTGTATATACACACATTG TTTTGTAGCATAGAAAAG GAGTTTTTGTCAACCGGT AGCCCACCCTGATTCTCA ACCAAGCCTGTAGATCTG TAATTGGGGTCTTAAGTC CTTGTTAAATTCTGGACA GCACTATGATTTTTTACA TTCTAAATCATTATACCA AGGTATCTTGTCTTATCT TCAGAGTGTCCAGCCTGT CGATAGATCGGAATACA ATCGTATAATTAATTGTT AAGCATGTTTCTTGTACA TACAGGTCAGTTACATCA ACATACTTATAAACAGTG CTGTAATATTTGTGA	50	是 HpaII
320b_4C_65G	48	GCAATTGATGATAACTGT GAGTGATTTTGTTTCTTTT	60	否

-2		GGAGACTACCTAACGCTT ATGACTTTGAGTTTCTGG TATGATTCAAAAGTAGAA TACCTGTAGCACGAGCAC CAATATCATTGTTAACTG AGGAAGTTCATGTACTTA CCGAATTATAGGAAAATT CAGTAGCTTTTCTTTGCC TACTAACTTAGGTTGTGT TCATCGAAATCATAACAT CACATAACTATTGTCTTC CATAAGCACTCAGGACTT CAAGTAAAAAGGATGAA GCCTATTCCATTTCACAT CTGAATAACTTTAGCAAA GTGTAAGAAGCTAA		
320b_8C_65G-1	49	CCCATGCATCAAACTGGC TCCCTCGTCTTCCGATTG TCTAGCTCATAGGCTCTC GAAGCATCTAAGGGCTAT TCCGGGTCCTGCAAGTAT AAGCTTCATTATTCCTAA GGATGTGGGGAGTGACT CAGAGGTCCAGATCGCA CTGTGGGCCAGACTGACA CAGCTTCAAAGGAAGGG CCTCCAAGTCCACTGCAC TCAGAATTAAGAATTCCC TGACGCATTATCTTGAGA AAGGCACTGGTCCATGTC TCTTGTATCTCCGAGACC AACTTAAAGGAGGGATG GGAGCTAACAGGCACCT CCCGACTTATCTTACCAG T	60	是 HpaII
320b_8C_65G-2	50	GAACTTCCAAATACACCG TCCCATCTGTTCAGTCAG GGATTGGGGTGAGAGAT ATATCGCATCAGGAATTA CGAAACCTTATGGGCAA GCAGTGATTAGCTAAGCC	60	否

		TGGAAACCTGGCAATTAA CACCTCAAACTGGATCCA TGCCTTTCTAACTATTTCC CACCCCTTGGCAGTCTAG GGCTGGCGAGAGGCCCT GCTAATACTTGAGGCGTA GATGGGGGGCGGCTTCTC TGCAAACTGGTGCCCGCT GGGGCTCTAATAATTATT CTTCCTTGTTCCACACAA CCAGCCCCTCGACTTTAA GCAGCTATGCTATGCA		
320b_16C_65G-1	51	TTCGGCACTTGTCTGCCC TCGTCAGAAAATGTTGGG TAAAACCCTAGGTTGTAG TTTGGGTCTGGCGAGCGG GAAAGTGCATGCTCGGCC CATGTGGGCTCCAAACTG AAGGTTATTAGATTCCTA GATGGTGAGACCGCATA CAAAAAGGGCCCTGGAA AGAGGTCACTTCAACGCA TCTCCTGATATTGGTCTG GTATCCACAGTAGAGCTA TTGTCGCCTAACAGTGAT GCCGCGCCGTCCTGTATT GGTGCGCGAGACAGCTT ATACGTACCTGAATGGCG ATAATTATCCGAGGGGCA GACTCAAGCTTAAGAAA	60	是 HpyCH4IV
320b_16C_65G-2	52	TTGGGCCGCCTTGTCCGC AACCAGCAACCGATAGC AGTCGGACTCCGAGTCAG TAGTGAAGTGCTTTAGCG TTAAGTGTTTATTGTGAA TGAGCCCTCTCTCCCCCA AATCACAAGAGGTGGCG GAAAAACACGAAGCCGA AGTACACCGACAAGGAA CGGTGCTCTCAAGAGTTG CCAGCCATTGCTAGACAG	60	否

		AGTAATTTCCTCCTCCAG GCGGAATTCAACAGTCCT CAGTCCCAGAATTATCTT GGGAAAGGATGGACACG AATATTTGGAACAGTGGA CGCCGACCCGTTTAATTA CAGGGTTCCCTGAGATTG T		
80b_1 C_35G - 2_mod	53	TGTCTAAATTAAAGTTGT GATCTTTGACTTAGCAAC GTCTCACCCTATAGCCTA CCAGACAAGAATTATGA AGAACATAT	50	是 HpyCH4I V
80b_1 C_50G - 2_mod	54	GTACACCATCATTATCCT CATAGCTTAGGCTCCACG TGCCTACAGGGCCATAAG GCTTGGAGATTCACTGTT AGCTGCTC	62	是 HpyCH4I V
80b_2 C_50G - 2_mod	55	GCCTCCCCAACTATAGGG TCAGGAAGGATTATGGC ACCCCACACGTATTTCAC CCGATCTGTACCAGTAAT CATACATGG	62	是 HpyCH4I V
80b_1 C_65G - 2_mod	56	GCTACCAGTGGCCCCCCC CTACCGGATCCATCCCTA ACCTCACCCCCCTGACTG CTAACCTGGGATGGTGAA GCCTGGGC	62	是 HpaII
80b_2 C_65G - 2_mod	57	GGTTATGCCCCCGCCCTG CATCCTCCCTGTCACACG TGCCCAACCCTAGCAATG TGTGGCCCCCCCTGCTGT CTCCCATC	62	是 HpyCH4I V
80b_4 C_65G - 2_mod	58	GCTGGTGCACCGCTGCCC CCACCCTCCACGTCTGTC ACAGCCTCGGTAGGTCCT GATTTGATGCTTGGGTGC TCGGCTGG	60	是 HpyCH4I V
160b_4 C_35G -	59	GTATAATCATAACAAAG GCCTAATGAAAGACGCT GATTTGAACATAGTTCCC	60	是

2_mod		TCATCATCTGATATTGTC CTACGTGTCTTTTTTCGAT GAGTGCACAATATGGTGT GAAGACCTATTACAATCA AAAAGTATAAACTAGCG ACTAAGATCTCAGAATTA		HpyCH4I V

[0078] 中试试验.为了评估加标对照是否有效,仅对未加标至生物样品的加标对照进行cfMeDIP-seq。在每组片段长度(80bp、160bp、320bp)内,合成片段的总数以等摩尔量加在一起。然后将样品以等量汇集在一起以构成10ng输入DNA(3:33ng/片段长度)。cfMeDIP-seq如Shen等人所述[7]。使用UMI衔接子来说明PCR扩增偏差,这需要衔接子连接在4℃下孵育过夜,最终衔接子浓度调节至0.09μmol。对于每个样品,在DNA变性后保存10%(1ng)的产物作为输入。扩增两个重复样品(总计N=4)的输入和输出,随后进行纯化,并使用AMPure Xp珠选择双重大小150bp至200bp。在MiSeq Nano flowcell(Illumina,San Diego,USA)上对样品进行测序,每个流动池在配对末端150bp处读取100万个读段(参见图1)。

[0079] HCT116的加标对照.为了测试合成片段的最佳浓度以用作加标对照,确保最终产物中存在足够的DNA以提供信息但不会淹没所有测序读段,通过将合成片段加标至HCT116细胞系中,剪切HCT116基因组DNA以模拟将存在于cfMeDIP-seq实验中的无细胞DNA(cfDNA)来测试。剪切的HCT116 cfDNA模拟物保持恒定的10ng,而合成片段库的不同浓度在0.1ng、0.3ng和1.0ng的DNA之间变化。以与中试试验中相同的方式制备样品并在MiSeq Nano上测序,每个流动池100万个读段,配对末端150bp(Illumina,San Diego,USA)。然后将0.1ng、0.05ng和0.01ng我们的对照加标至10ng的HCT116中,在NovaSeq上进行高分辨率测序,每个样品6000万个读段,配对末端2×100bp(参见图2)。

[0080] 评估技术偏差.为了评估合成片段作为加标对照的性能,使用加标对照DNA池作为输入进行cfMeDIP-seq。输入池由9.99ng合成的加标DNA组成,每种片段大小的量为等摩尔量,在每个片段大小池中,每种甲基化状态的量为等摩尔量(表2)。按照Shen等人(2018)[2]稍作修改进行cfMeDIP-seq。使用xGen Stubby Adapter和独特的双指数(UDI)引物对(Integrated DNA technologies,Coralville,IA,USA,Cat编号10005921)来说明PCR扩增偏差。在4℃过夜进行衔接子连接,通过稀释将最终衔接子浓度调节至0.09μmol。对于每个样品,保存1ng DNA变性产物作为输入。对于每个样品,我们扩增输入和输出,随后进行纯化,并使用AMPure Xp珠(Beckman Coulter,Brea,CA,USA)选择双重大小150bp至200bp。在MiSeq Nano flowcell(Illumina,San Diego,CA,USA)上对样品进行测序(Princess Margaret Genomics Centre,Toronto,ON,CA),配对末端2×150bp,每个流动池100万个读段(图13)。

[0081] 优化合成DNA量.通过向剪切的HCT116基因组DNA中添加不同量的加标对照,确定每个实验所需的加标对照DNA的最佳量(ATCC,Manassas,VA,USA,RRID:CVCL_0291)。使用LE220超声波发生器(Covaris,Woburn,MA,USA)剪切HCT116基因组DNA(一种结肠直肠癌细胞系),并使用AMPure XP珠(Beckman Coulter,Brea,CA,USA)选择大小以模拟cfDNA输入。用质量为0.1ng、0.05ng和0.01ng的合成的加标对照DNA生成3个重复样品,将它们中的每一个加入10ng剪切的HCT116cfDNA模拟物中。按照Shen等人(2018)[1]先前所述进行cfMeDIP-

seq实验,在Illumina NovaSeq 6000(Illumina,San Diego,CA,USA)上对样品进行测序(Princess Margaret Genomics Centre,Toronto,ON,CA),配对末端2×100bp,每个样品6000万个读段(图13)。

[0082] 生物信息学预处理.使用fastp 0.11.5版[2](参见式1)修剪衔接子,并移除评分小于20的读段。使用BowTie2 0.11.5版[3](参见式2)将读段与设计片段的序列对齐。随后,将未与本发明的合成DNA对齐的序列与人参考基因组(GRCh38/hg38)对齐。[15]在每个样品中,超过98%的读段与加标对照序列或人基因组对齐。当读段对中的至少一个读段未对齐或质量低时,删除读段对。低质量定义为Phred评分<20。通过匹配每个样品的每个读段的UMI序列,含有相同UMI的读段折叠计算为一个读段。

--umi--umi_loc=每个读段--umi_len=5

--衔接子序列=AATGATACGGCGACCACCGAGATCTACACATATGCGCACACTCTTTCCCTACACGAC

--衔接子序列_r2=CAAGCAGAAGACGGCATACGAGATACGATCAGGTGACTGGAGTTCAGACGTGT (1)

bowtie2--局部的-x[序列参考]--minins 80--maxins 320 (2)

[0083] 由加标对照数据计算绝对浓度.使用来自中试试验的去重复读段计数以及G+C含量、CpG、片段长度和重量摩尔浓度(pmol)来创建广义线性模型以计算原始样品内给定片段的重量摩尔浓度(pmol)。由于大小选择导致读段计数和片段长度之间的非单调关系,我们使用式3转换片段长度:

[0084] $x=(160-\text{片段长度})^2$ (3)

[0085] 这种转换导致数据左偏。因此,使用z评分来归一化这些数据(图3)。 (式4)

[0086] (x-μ片段长度/σ片段长度) (4)

[0087] 每个片段的CpG数量分布也是左偏的。为了将这些数据返回到正态分布,我们使用了立方根$\sqrt[3]{CpG数量}$(图4)。

[0088] $$\text{重量摩尔浓度(fmol/ng)/片段}=(x-\mu_{\text{片段长度}}/\sigma_{\text{片段长度}})+(\text{G+C含量})+\sqrt[3]{CpG数量}+(\text{读段计数}) \quad (5)$$

[0089] 使用具有对数连结(log link)的高斯广义线性模型来创建一个值(重量摩尔浓度),以说明可能影响这些数据结果的片段长度、G+C含量和CpG的数量。(式5)使用的最佳模型将基于每个实验而不同。

[0090] 来自加标对照数据的绝对定量.创建广义线性模型,以根据UMI共有序列、G+C含量、CpG分数、片段长度从去重复的加标对照读段计数中预测摩尔量。为此,使用了R 3.4.1版中的stats包。为了减少其左偏,使用CpG分数的立方根转换。通过使用为每个实验学习的回归系数(β),高斯广义线性模型(等式6)用于计算原始样品中存在的每个DNA片段的摩尔量(η)。该模型包括读段计数($x_{读段}$)、片段数量($x_{片段}$)、片段长度($x_{长度}$)、片段的G+C含量(x_{GC})和片段的CpG分数($\sqrt[3]{x}_{CpG分数}$)。每个实验和模型的回归系数(β)可以在表3中找到。

[0091] $$\eta = \beta_0 + \beta_{读段} x_{读段} + \beta_{片段} x_{片段} + \beta_{长度} x_{长度} + \beta_{GC} x_{GC} + \beta_{CPG分数} \sqrt[3]{x}_{CpG分数} (6)$$

[0092] 表3.回归系数(β)。所有实验均使用高斯广义线性模型。

	截距系数	片段长度系数	G+C 含量系数	CpG 分数系数	读段计数系数
10 ng HCT116 中含有 0.01 ng	0.0039210000	-0.0000105700	0.0000030070	0.0001706000	-0.0000001584
批次 1	0.0038410000	-1.1420000000	0.0000022400	0.0002940000	-0.0000001390
批次 2	0.0039900000	-0.0000110500	0.0000012400	0.0001415000	-0.0000000049
批次 3	0.0038670000	-0.0000112000	-0.0000014390	0.0004484000	-0.0000001343

[0093] 如前面的分析,[1,5,6]将基因组分箱(bin)到非重叠的300bp窗口中。使用Bedtools intersect22计算与定义的300bp窗口重叠的给定片段的比例。计算调整的摩尔量(η′)以仅考虑每个片段重叠的箱部分。该模型包括来自等式7的片段和基因组窗口之间的重叠(θ)、窗口大小(x)和摩尔量(η)。

[0094] $\eta' = (\theta/x) \times \eta$ (7)

[0095] 识别待过滤的区域.为了评估摩尔量和可映射性之间的关系,使用umap k100可映射性评分。[23]注释可映射性评分以呈现300bp窗口并使用每300bp窗口的最小可映射性评分。计算两个重复样品之间的标准偏差,其中0.01ng的合成DNA加标至10ng的HCT116基因组DNA中。评估了摩尔量和可映射性评分之间的关系,并且摩尔量和标准偏差不包括简单重复区域,[24]ENCODE黑名单中列出的区域,[25]可映射性评分≤0.5的区域和标准偏差≥0.25的区域。HOMER 4.10.4版用于研究特异性转录因子结合基序是否与我们的离群值相关。将窗口大小设定为300bp,并将离群值与HOMER产生的随机化基因组背景进行比较。

[0096] 皮摩尔和M值之间的相关性.片段长度、CpG分数、G+C含量和读段计数用于模拟摩尔量。使用高斯广义线性模型估计摩尔量(r^2=0.93)。对160bp片段表现更好的模型进行优先级区分,因为对这些片段进行了大小选择,并且这些是感兴趣的片段。

[0097] 为了显示将甲基化DNA定量为以皮摩尔计的摩尔量是DNA甲基化的有效量度,在Illumina EPIC阵列(Illumina,San Diego,CA,USA)上一式三份运行HCT116基因组DNA。在EPIC阵列上运行的HCT116基因组DNA样品是用0.01ng加标对照加标的HCT116基因组DNA的技术重复。使用sesame对EPIC阵列数据进行归一化和预处理。[27]在EPIC阵列上将CpG注释到300bp基因组窗口。当>1CpG的探针被注释到窗口时,探针M值在整个窗口取平均值。删除了映射到UCSC简单重复、[24]ENCODE黑名单、[25]低映射性≤0.50的区域以及重复之间的标准偏差≥0.25的区域的窗口。评估了EPIC阵列M值和皮摩尔与EPIC阵列M值和含≥3个CpG探针、≥5个CpG探针、≥7个CpG探针和≥10个CpG探针的窗口处读段计数之间的相关性。

[0098] 检查各实验批次的一致性.为了模拟已知的批次效应并测试我们的加标对照是否可以比不使用我们的加标对照的当前分析更好地减轻批次效应,将从5名AML患者的血浆获得的10ng cfDNA样品与0.01ng加标对照一起给予三名独立的研究者。按照由大学健康网络

的研究伦理委员会批准的程序(UHNREB 01-0573),在知情同意的情况下,从玛嘉烈公主癌症中心/大学健康网络(Princess Margaret Cancer Centre/University Health Network)的白血病组织库收集AML患者样品。使用AML样品,因为它们具有相对高的cfDNA量,允许我们将30ng的cfDNA分成三个技术重复。每个研究人员按照Shen等人(2018)[2]作出一些微小改变来进行cfMeDIP-seq法。其目的是模拟在不同实验室的不同研究的公开数据中常见的批次效应。因此,研究人员1和3使用与先前分析相同的UMI。研究人员2使用2bp简并UMI。[28]对于衔接子的连接,研究人员1和2在4℃下孵育过夜,而研究人员3在20℃下孵育2小时。用于扩增最终文库的PCR循环数在批次之间也有所不同。研究人员1运行15个周期,研究人员2运行13个周期,而研究人员3运行11个周期。研究人员1和3使用了抗体1(Diagenode,Denville,NJ,USA,Cat编号C15200081-100,批号RD004,RRID:AB_2572207),同时研究者2使用抗体2(Diagenode,Denville,NJ,USA,Cat编号C15200081-100,批号RD001,RRID:AB_2572207)。AML样品在Illumina NovaSeq 6000(Illumina,San Diego,CA,USA)上运行,配对末端2×100bp,每个样品6000万个读段。使用高斯广义线性模型来计算摩尔量并独立地调整每个批次中的片段长度、G+C含量和CpG分数。通过对我们计算摩尔量的样品进行主成分分析(PCA)来评估加标对照是否减轻了批次效应。也可以只使用读段计数以及使用QSEA[29](目前标准的MeDIP-seq数据处理流水线)预处理的读段计数,而不使用我们的加标对照对相同样品进行PCA。为了研究已知变量是否与主成分中的每一个相关,在每个主成分和每个分类变量之间进行双向ANOVA。分类变量包括:批次、测序仪、衔接子、样品和性别(通过Y染色体信号推断)。使用R 3.4.1.版中的computer.es包,将所得F统计量转化为Cohen's d的效应量。[30]使用Holm-Bonferroni方法调整多次测试校正的P值。[31]

[0099] 结果1

[0100] 中试试验.平均来说,51%的输入片段是甲基化的,49%是未甲基化的。在进行cfMeDIP后,平均97%的片段被甲基化,3%未被甲基化,代表非特异性结合(图5)。这与Shen等人[6]一致,其显示了与qPCR验证类似的非特异性结合。甲基化序列的富集进一步支持了cfMeDIP-seq法的有效性。

[0101] 为了评估对特定片段长度、G+C含量或CpG数量的扩增偏差,根据甲基化状态对输入和输出样品的独特读段(去重复)绘制读段计数分布。甲基化片段显示出160bp片段的富集,这是由于150至200bp片段的大小选择步骤所预期的。观察到优选较高的G+C含量。在cfMeDIP法之后,保持了160bp片段的富集,并观察到对具有较高G+C含量的片段的富集。每个片段的CpG数量没有出现影响cfMeDIP方案后富集片段的模式(图6)。

[0102] 未甲基化的片段对160bp片段显示相同的富集和更高的G+C含量。这表明cfMeDIP法的非特异性结合偏向于G+C含量较高的片段。片段中存在的CpG的数量和读段的数量之间没有关联(图7)。

[0103] 测试HCT116的加标对照.随着细胞系DNA的加入,对5-甲基胞嘧啶达到99.9%特异性,与非甲基化片段的非特异性结合≤0.1%(图9)。与仅对合成片段进行的中试试验类似,观察到对较小片段的PCR偏差效应,可通过使用UMI条形码识别哪些片段是PCR重复来减轻这种效应。根据150至200bp片段的大小选择步骤,富集160bp片段。还存在对G+C含量较高的片段的富集。在我们用于加标对照的各浓度下观察到这些模式。

[0104] 如图8所示,将在每种重量摩尔浓度的加标对照上使用的读段数与HCT116使用的

读段数进行比较。即使在0.1ng合成DNA加标至10ng HCT116时，约6％的读段用于对照。因此，在每个样品具有6000万个读段的更高分辨率测序下，在实验中将400万个读段用于加标对照。这比校正批次效应所需的更多。为此，测试了更低量的加标对照。

[0105] 优化加标对照的输入浓度.cfMeDIP-seq实验富集的甲基化DNA＞9.99％，与非甲基化片段的非特异性结合≤0.01％(图10)。还观察到对160bp片段的富集和更高的G+C含量。没有观察到片段中存在的CpG数量和读段计数的模式。

[0106] 与用于我们的生物样品HCT116的读段总数相比，评估了用于合成加标对照的读段总数。这允许优化将在后续实验中使用的加标对照的量，以最大化感兴趣的生物样品的读段，同时仍然获得关于对照片段的足够信息以校正生物学和技术偏差。加标0.01ng合成对照，允许使用≤0.01％的读段到对照中，而将剩余读段留给生物样品。仍有>650,000个对照序列读段用于分析(图11)。因此，决定在后续实验中使用0.01ng加标对照片段。

[0107] 计算加标对照的浓度.使用归一化片段长度和片段内CpG的数量，以及G+C含量和读段计数以模拟浓度(fmol/ng)(参见方法)。使用具有对数连结的高斯广义线性模型能够很好地估计浓度(r^2＝0.999)(图12)。80bp片段的性能低于160bp和320bp片段。然而，由于大小选择是针对150-200bp进行的，这对实际条件下的模型性能影响较小。

[0108] 结果2

[0109] cfMeDIP-seq优先富集高G+C含量区域.当在作为输入样品的合成加标对照上直接进行cfMeDIP-seq时(图13)，我们观察到51％的输入片段甲基化和49％未甲基化。片段进行cfMeDIP后，观察到片段丰度的偏移，97％的测序读段对应于甲基化片段。对甲基化序列的富集进一步支持了cfMeDIP-seq法的有效性。

[0110] 在cfMeDIP-seq之后，合成的加标对照输出以及10ng的HCT116中的加标对照显示160bp片段的富集，我们预期这是由于我们对150bp-200bp片段的大小选择步骤。在cfMeDIP法之后，我们保持了对160bp片段的富集，并观察到对具有较高G+C含量和高CpG分数的片段的富集(图14)。

[0111] 来自合成加标输入对照输出和0.01ng加标的未甲基化片段的信号与片段长度、G+C含量或CpG分数无关(图14)。这表明cfMeDIP法的非特异性结合是随机的。

[0112] 低输入加标对照证明足以说明生物学和技术差异.cfMeDIP-seq实验使用0.01ng加标对照DNA到10ng的HCT116基因组DNA中，以富集≥99.99％甲基化DNA，非特异性结合非甲基化片段≤0.01％(图14)。还观察到对160bp片段的富集和更高的G+C含量。观察到具有以1/80bp存在的CpG的片段的信号较弱。

[0113] 与用于生物样品HCT116基因组DNA的读段总数相比，评估了用于加标对照的读段总数。这允许优化在后续实验中使用的加标对照的量，最大化对感兴趣的生物样品的读段，同时从加标对照获得足够的读段以校正生物学和技术偏差。将0.01ng合成加标对照DNA加标至本发明的cfMeDIP-seq实验中，允许使用≤1％的读段到对照中，而将剩余读段留给生物样品。仍有>650,000个对照序列读段用于分析。因此，决定在后续实验中使用0.01ng加标对照片段。

[0114] 过滤有问题的区域消除了生物学和技术伪影的潜在来源.在过滤包含简单重复的区域、ENCODE黑名单区域、可映射性评分≤0.5的区域和重复之间的标准偏差≥0.25的区域之后，我们观察到摩尔量和标准偏差之间没有关系，并且摩尔量和可映射性评分之间没有

关系。这表明去除这些区域有利于减少生物学和技术伪影。有11个离群值窗口，如图15所示。这些区域都≥2pmol，如表4所示。

[0115] 表4.预测摩尔量为≥2pmol的300bp基因组窗口。

染色体	起始[a]	终点[a]	量	元件[b]	家族[b]	名称[b]
13	95,176,201	95,176,500	78.15 pmol	SINE	Alu	AluJo
2	120,383,401	120,383,700	55.01 pmol	SINE	MIR	MIR_Amn
12	95,476,201	95,476,500	11.70 pmol	SINE	Alu	AluSx
22	20,900,401	20,900,700	6.30 pmol	SINE	Alu	AluJb, AluY
17	1,025,101	1,025,400	5.33 pmol	SINE	Alu	AluYe5, AluSx1
X	44,613,001	44,613,300	4.23 pmol	SINE	Alu	AluSp, AluJr
1	44,582,701	44,583,000	3.80 pmol	SINE	Alu	AluSx1
8	139,704,301	139,704,600	3.74 pmol	低复杂度	-	G-富集
2	224,230,501	224,230,800	2.49 pmol	LTR	ERV1	HERVH-int
17	3,521,101	3,521,400	2.45 pmol	LINE, DNA转座子	L2, hAT-Blackjack	L2b, MER63D
16	11,578,801	11,579,100	2.36 pmol	LINE, SINE	L1, Alu	L1MD3, AluSp

a由hg38定义的基因组位置，1-起始，完全关闭。

b与我们的300bp基因组窗口重叠的所有元件、家族和名称。来自RepeatMasker 3.0版的UCSC基因组浏览器RepeatMasker轨道中定义的元件、家族和名称。

[0116] 所有这11个“离群值”窗口都是重复元件，主要是SINE元件(N＝8)，主要来自Alu家族，其起源追溯到灵长类。HOMER分析未产生显著基序。[26]

[0117] 绝对量化与M值相关性并与读段计数进行比较.线性模型用于计算每个300bp基因组窗口的摩尔量。在我们的HCT116基因组DNA样品中，在整个基因组的摩尔量和M值之间观察到显著的相关性。当分析限于高CpG密集区域时，观察到更高的相关性，该高CpG密集区域定义为300bp窗口，在300bp窗口内具有≥5个代表EPIC阵列上DNA甲基化的CpG探针。这并不

令人惊讶,因为cfMeDIP-seq技术优先测量高CpG密集区域的DNA甲基化。为了与当前标准进行比较,读段计数与M值相关(图16)。结果表明,摩尔量的表现与读段计数相似,但具有允许绝对量化的优势。

[0118] 加标对照减轻批次效应.以读段计数测量的原始数据上的PCA表明,包含76%方差的主成分1与处理批次相关(图17)。在QSEA归一化之后,主成分1仍然与批次相关,尽管在多次测试校正之后不显著,但测序仪此时与主成分1显著相关。这表明QSEA归一化实际上可能将没有生物学意义的方差引入数据。使用摩尔量测量极大地改善了批次效应。在不应用任何基因组过滤的情况下产生的基于以皮摩尔计的摩尔量的数据的归一化导致批次处理变量向主成分2偏移,构成方差的≤5%(图5)。对包含简单重复的区域、ENOCDE黑名单区域和可映射性低的区域加入建议过滤,看到了批次处理效果向主成分5进一步偏移,构成方差的1%(图17)。对主成分5的进一步研究,观察驱动该方差的前10%窗口,这些顶部区域的72%是由重复掩蔽器定义的重复元件,[33]主要是Alu元件。

[0119] 讨论

[0120] 数据显示了使用合成的加标对照DNA改善cfMeDIP-seq实验结果的有效性。实验1(在加标对照上直接进行的cfMeDIP-seq)和实验2(在0.01ng加标到HCT116基因组DNA上进行的cfMeDIP-seq)中与5-甲基胞嘧啶的非特异性结合的差异可以通过实验样品中甲基化CpG的比例的差异来解释。加标对照含有51%的甲基化CpG,而人基因组含有大约70%的甲基化CpG。[34]为了减少技术和生物学伪影,在分析前过滤可能有问题的区域是非常重要的。结果表明,去除具有简单重复的区域、与ENCODE黑名单重叠的区域、可映射性评分低的区域和在重复之间具有高标准偏差的区域有助于减少离群值以及技术和生物学伪影。结果表明,cfMeDIP-seq数据存在生物学和技术偏差,而使用我们的加标对照有助于减少这些偏差。

[0121] 尽管对可能有问题的区域进行严格的过滤,但是仍然存在离群值区域,其主要由重复元件、主要是SINE元件组成。虽然这些区域是CpG密集的,但是我们的加标对照针对CpG分数进行调整。因此,高CpG密度不太可能是这些区域成为离群值的原因。有趣的是,大多数Alu离群值是较早的Alu元件。[32]没有特异性转录因子结合基序与这些元件相关。有可能的是,因为这些元件是高度甲基化的,[32]所以可能看到这些区域的比例过高。根据实验问题,除了先前的过滤建议之外,一些人可以选择从分析中去除重复元件,例如LINE和SINE。然而,鉴于这些重复元件占据相对少的窗口,因此它们不可能严重影响结果。

[0122] 结果表明,使用加标对照有助于缓解批次之间因多种技术因素而产生的差异,包括:技术人员、衔接子、测序仪和衔接子连接孵育。使用加标对照,观察到与批次显著相关的主成分占数据方差的≤5%。而在原始数据中,或使用QSEA归一化的数据中,与处理批次显著相关主成分构成数据偏差的≥85%。

[0123] 本研究为MeDIP-seq实验中使用加标对照来说明生物学和技术偏差的有益影响提供了有力证据。它不仅会改善给定研究中的结果,将这些对照作为金标准将提高从许多实验室生成的数据的再现性。

参考文献1

[1]Stephen F Altschul等人“Basic local alignment search tool”.In:Journal of molecular biology 215.3(1990),pp.403-410.

[2]Shifu Chen等人“fastp:an ultra-fast all-in-one FASTQ preprocessor”

.In:Bioinformatics 34.17(2018),pp.i884-i890.

[3]Ben Langmead and Steven L Salzberg.“Fast gapped-read aligment with Bowtie 2”.In:Nature methods 9.4(2012),p.357.

[4]Richard Owczarzy等人“IDT SciTools:a suite for analysis and design of nucleic acid oligomers”.In:Nucleic acids research 36.suppl_2(2008),W163-W169.

[5]Yann Ponty,Michel Termier,and Alain Denise.“GenRGenS:software for generating random genomic sequences and structures”.In:Bioinformatics 22.12 (2006),pp.1534-1535.

[6]Shu Yi Shen等人“Sensitive tumour detection and classification using plasma cell-free DNA methylomes”.In:Nature 563.7732(2018),p.579.

[7]Shu Yi Shen等人“Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA”.In:Nature protocols 14.10(2019),pp.2749-2780.

[8]Zhenjiang Zech Xu and David H Mathews.“Secondary structure prediction of single sequences using RNAstructure”.In:RNA Structure Determination.Springer,2016,pp.15-34.

参考文献2

1.Shen,S.Y.,Singhania,R.,Fehringer,G.,Chakravarthy,A.,Roehrl,M.H.,Chadwick,D.,Zuzarte,P.C.,Borgida,A.,Wang,T.T.,Li,T.,等人Sensitive tumour detection and classification using plasma cell-free DNA methylomes.Nature 563,579(2018).

2.Shen,S.Y.,Burgener,J.M.,Bratman,S.V.&De Carvalho,D.D.Preparation of cfMeDIP-seq libraries for methylome profiling of plasma cell-free DNA.Nature protocols 14,2749-2780(2019).

3.Cao,F.,Wei,A.,Hu,X.,He,Y.,Zhang,J.,Xia,L.,Tu,K.,Yuan,J.,Guo,Z.,Liu,H.,等人Integrated epigenetic biomarkers in circulating cell-free DNA as a robust classifier for pancreatic cancer.Clinical Epigenetics 12,1-14(2020).

4.Lasseter,K.,Nassar,A.H.,Hamieh,L.,Berchuck,J.E.,Nuzzo,P.V.,Korthauer,K.,Shinagare,A.B.,Ogorek,B.,McKay,R.,Thorner,A.R.,等人Plasma cell-free DNA variant analysis compared with methylated DNA analysis in renal cell carcinoma.Genetics in Medicine,1-8(2020).

5.Nassiri,F.,Chakravarthy,A.,Feng,S.,Shen,S.Y.,Nejad,R.,Zuccato,J.A.,Voisin,M.R.,Patil,V.,Horbinski,C.,Aldape,K.,等人Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes.Nature Medicine 26,1044-1047(2020).

6.Nuzzo,P.V.,Berchuck,J.E.,Korthauer,K.,Spisak,S.,Nassar,A.H.,Abou Alaiwi,S.,Chakravarthy,A.,Shen,S.Y.,Bakouny,Z.,Boccardo,F.,等人Detection of renal cell carcinoma using plasma and urine cell-free DNA methylomes.Nature

Medicine 26,1041-1043(2020).

7.Jiang,L.,Schlesinger,F.,Davis,C.A.,Zhang,Y.,Li,R.,Salit,M.,Gingeras,T.R.&Oliver,B.Synthetic spike-in standards for RNA-seq experiments.Genome research 21,1543-1551(2011).

8.Chen,K.,Hu,Z.,Xia,Z.,Zhao,D.,Li,W.&Tyler,J.K.The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses.Molecular and cellular biology 36,662-667(2016).

9.Orlando,D.A.,Chen,M.W.,Brown,V.E.,Solanki,S.,Choi,Y.J.,Olson,E.R.,Fritz,C.C.,Bradner,J.E.&Guenther,M.G.Quantitative ChIP-Seq normalization reveals global modulation of the epigenome.Cell reports 9,1163-1170(2014).

10.Deveson,I.W.,Chen,W.Y.,Wong,T.,Hardwick,S.A.,Andersen,S.B.,Nielsen,L.K.,Mattick,J.S.&Mercer,T.R.Representing genetic variation with synthetic DNA standards.Nature methods 13,784(2016).

11.Blackburn,J.,Wong,T.,Madala,B.S.,Barker,C.,Hardwick,S.A.,Reis,A.L.,Deveson,I.W.&Mercer,T.R.Use of synthetic DNA spike-in controls(sequins) for human genome sequencing.Nature Protocols 14,2119(2019).

12.Mouliere,F.,Chandrananda,D.,Piskorz,A.M.,Moore,E.K.,Morris,J.,Ahlborn,L.B.,Mair,R.,Goranova,T.,Marass,F.,Heider,K.,等人Enhanced detection of circulating tumor DNA by fragment size analysis.Science translational medicine 10(2018).

13.Ponty,Y.,Termier,M.&Denise,A.GenRGenS:software for generating random genomic sequences and structures.Bioinformatics 22,1534-1535(2006).

14.Altschul,S.F.,Gish,W.,Miller,W.,Myers,E.W.&Lipman,D.J.Basic local alignment search tool.Journal of molecular biology 215,403-410(1990).

15.Graves-Lindsay,T.,Albracht,D.,Fulton,R.S.,Kremitzki,M.,Magrini,V.,Markovic,C.,McGrath,S.,Steinberg,K.M.,Wilson,R.K.,等人Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly.Genome research 27,849-864(2017).

16.Owczarzy,R.,Tataurov,A.V.,Wu,Y.,Manthey,J.A.,McQuisten,K.A.,Almabrazi,H.G.,Pedersen,K.F.,Lin,Y.,Garretson,J.,McEntaggart,N.O.,等人IDT SciTools:a suite for analysis and design of nucleic acid oligomers.Nucleic acids research 36,W163-W169(2008).

17.Xu,Z.Z.&Mathews,D.H.Secondary structure prediction of single sequences using RNA structure.RNA structure determination,15-34(Springer,2016).

18.Chen,S.,Zhou,Y.,Chen,Y.&Gu,J.fastp:an ultra-fast all-in-one FASTQ preprocessor.Bioinformatics 34,i884-i890(2018).

19.Ewing,B.&Green,P.Base-calling of automated sequencer traces using phred.II.Error probabilities.Genome research 8,186-194(1998).

20.Langmead,B.&Salzberg,S.L.Fast gapped-read alignment with Bowtie 2.Nature methods 9,357(2012).

21.R Core Team.R:A Language and Environment for Statistical Computing R Foundation for Statistical Computing(Vienna,Austria,2013).

22.Quinlan,A.R.&Hall,I.M.BEDTools:a flexible suite of utilities for comparing genomic features.Bioinformatics 26,841-842(2010).

23.Karimzadeh,M.,Ernst,C.,Kundaje,A.&Hoffman,M.M.Umap and Bismap: quantifying genome and methylome mappability.Nucleic acids research 46,e120 (2018).

24.Karolchik,D.,Hinrichs,A.S.,Furey,T.S.,Roskin,K.M.,Sugnet,C.W., Haussler,D.&Kent,W.J.The UCSC Table Browser data retrieval tool.Nucleic acids research 32,D493-D496(2004).

25.Amemiya,H.M.,Kundaje,A.&Boyle,A.P.The ENCODE blacklist: identification of problematic regions of the genome.Scientific reports 9,1-5 (2019).

26.Heinz,S.,Benner,C.,Spann,N.,Bertolino,E.,Lin,Y.C.,Laslo,P.,Cheng, J.X.,Murre,C.,Singh,H.&Glass,C.K.Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.Molecular cell 38,576-589(2010).

27.Zhou,W.,Triche Jr,T.J.,Laird,P.W.&Shen,H.SeSAMe:reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions.Nucleic acids research 46,e123-e123(2018).

28.Wang,T.T.,Abelson,S.,Zou,J.,Li,T.,Zhao,Z.,Dick,J.E.,Shlush,L.I., Pugh,T.J.&Bratman,S.V.High efficiency error suppression for accurate detection of low-frequency variants.Nucleic acids research 47,e87-e87(2019).

29.Lienhard,M.,Grasse,S.,Rolff,J.,Frese,S.,Schirmer,U.,Becker,M., Börno,S.,Timmermann,B.,Chavez,L.,Sültmann,H.,等人QSEA—modelling of genome-wide DNA methylation from sequencing enrichment experiments.Nucleic acids research 45,e44-e44(2017).

30.Re,A.C.D.compute.es:Compute Effect Sizes(2013).

31.Holm,S.A simple sequentially rejective multiple test procedure.Scandinavian journal of statistics,65-70(1979).

32.Deininger,P.Alu elements:know the SINEs.Genome biology 12,236 (2011).

33.Tarailo-Graovac,M.&Chen,N.Using RepeatMasker to identify repetitive elements in genomic sequences.Current protocols in bioinformatics 25,4-10(2009).

34.Strichman-Almashanu,L.Z.,Lee,R.S.,Onyango,P.O.,Perlman,E.,Flam,F., Frieman,M.B.&Feinberg,A.P.A genome-wide screen for normally methylated human

CpG islands that can identify novel imprinted genes.Genome research 12,543-554(2002).

序列表

<110> 大学健康网络

<120> 用于无细胞MeDIP测序的合成加标对照及其使用方法

<130> 05014971-220PCT

<150> US 62/931411

<151> 2019-11-06

<160> 59

<170> PatentIn版本 3.5

<210> 1

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 1

tgtctaaatt aaagttgtga tctttgactt agcatcgact cacctatag cctaccagac 60

aagaattatg aagaacatat 80

<210> 2

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 2

gtacaccatc attatcctca tagcttagtc tcccgcaggc ccagggtaca taaggcttgg 60

agattcactg ttagctgctc 80

<210> 3

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 3

gcctccccaa ctatagggtc aggaaggatt atggcacccc acacgatttt cacccgatct 60

gtaccagtaa tcatacatgg 80

<210> 4

<211> 80

<212> DNA

<213> 人工序列
<220>
<223> 对照序列
<400> 4
gctaccagtg gccccccccct accgagtccc ccattaacct caccccccctg actgctaacc 60
tgggatggtg aagcctgggc 80
<210> 5
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 5
ggttatgccc ccgccctgca tcctccctgt ctacacggcc caaccctagc aatgtgtggc 60
ccccccctgct gtctcccatc 80
<210> 6
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 6
gctggtgcac cgctgccccc acccacctcg cttgtcacag cctcggtagg tcctgatttg 60
atgcttgggt gctcggctgg 80
<210> 7
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 7
gtataatcat aacaaaggcc taatgaaaga cgctgatttg aaactagttc cctcatcatc 60
tgatagattt cctcgtgtct tttttcgtga atggcacaat atggtgtgaa gacctattac 120
aatcaaaaag tataaactag cgactaagat ctcagaatta 160
<210> 8
<211> 80
<212> DNA
<213> 人工序列
<220>

<223> 对照序列
<400> 8
taggatatag gttgtcccct agtaggagat aaactttgat taacatccaa ttgatcgtta 60
gtgtccttca aaattatgct 80
<210> 9
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 9
tctaatactc atcttagctc gcgtgctttg tgattttagt gctgaaattc ttaaatgtta 60
accactgtga aatccataag 80
<210> 10
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 10
ctcaaatata acaagagtag caaacttaca aagatcgctg acaagtatgt tatccatttc 60
taagcgctac caataacact 80
<210> 11
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 11
aaggcattac ttatctaatc aatcgacaaa acgttaagtc agtgttagga tagtgtcatt 60
tgtactcgta gacgaaattg 80
<210> 12
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 12
ttattattga ccgtacacta tttaactaac agatatgacg tattactatg atatgttaat 60

gacgctgagc tgctcggaga 80

<210> 13

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 13

gaggaccata tagctcgcac aggaaccagc tgaagaattg attggtagtg ctgaccagac 60

accaaccttc aaacctctgc 80

<210> 14

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 14

acaacaccct ccacccaata cttgtgagtt ggtcgcagca cgagcctagt ctccttgtaa 60

gtcagtcaaa tgcctgtaac 80

<210> 15

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 15

agtcatcagc atattgtcag tacccagtgg tctctaggaa aatcggccgg tacgtaaata 60

ctcctagtgg gctgcgtggt 80

<210> 16

<211> 80

<212> DNA

<213> 人工序列

<220>

<223> 对照序列

<400> 16

gcttcttatg ataccaaagt tgcccaaaaa ggctagcgtt ctagttaggg tgcagccgcg 60

agaagaccgg gttcatgaag 80

<210> 17

<211> 80

<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 17
gcaggctgca gggtttggcc cctggttccg ttccagcagg tggcataggt ggggaaagcc 60
aggtgcctac agtggggtgg 80
<210> 18
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 18
caccttgaga cctccagagg gggatccaca acttgcgccc tctgtgaagt aggctctggt 60
gcgcaggggg gaaggggggc 80
<210> 19
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 19
ttgggaggct ctggactggg gcaaacgaca ccgtgcatca actgtgtggt ggtggcctcg 60
tccccccccа tcctctccgc 80
<210> 20
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 20
ttagtcgaga ttttagccta attgagagat agtccgatga tatgtctttg atctaacatg 60
tcatcatgaa atatgaagcc aacacactca tatgttcatg tgacaaaaga tccagttaag 120
ccagtattga ggtttgtcca tcacttaagt actattcatt 160
<210> 21
<211> 160
<212> DNA
<213> 人工序列

<220>
<223> 对照序列
<400> 21
ctttactact gaatgtaagc tcttgcagag gatctaacag ggatagaatt atgaacacgt 60
ctgtcacaca taacttcaaa tgcaatttat taataagggt cagaatgtgt ggtatctttc 120
cagacttata tcattccctt tactataacc gattacacat 160
<210> 22
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 22
atgtgtaaga aataaaatac tggctcatca tcataaactt gtctataatg tcactattat 60
cacaaagaat gcaggtacga cacggtatcg gcagcagtgg atagctcgta tctatatgaa 120
taggggaagt gaataatatg acaatagtat actttgctta 160
<210> 23
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 23
ttgtagtaca gtctaaccat cttgacccag tagctcccca tctgatatgc tcagtagcta 60
gggtggcctg agggaaccgg tcaaacccac ttattctgaa ccccagaggt atgttatgcg 120
caggaacctg ccttctattg gtagtgtctt gggtcaacag 160
<210> 24
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 24
gtaacatggt taccactggg accggacctt ttcacctcca ctttcaggga ataggattca 60
gtcctgtata gcagtgtgac accccaaggc caattccacc ctacattcaa tgcctgagtg 120
tatgttggcc attgggtaac tagccgtgtc ccaacctcat 160
<210> 25
<211> 160
<212> DNA

<213> 人工序列
<220>
<223> 对照序列
<400> 25
agccttggac gtgagtctct gtttctgacc caactgagat ctttttactg tcattctacc 60
ccctagagac tcgcgtttct agagagggga tgtatgtgag gggttgtgat ttagcccgtg 120
atgccctagg atcttgagac aattgtcagg gccctccagt 160
<210> 26
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 26
ggctctaggg ggtgataaag tctcggatta tgctgtatgg agtcccatca acttccaatg 60
acaatattgt actctaggat agctagatga cgccccaggc aaagaaccct tttcgtatga 120
ggccagcctt ccaaggtcca ctaggctcag ctcctcgatg 160
<210> 27
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 27
ggtaagtatg cagctcaacg agggtaccgg tagcgacccg ctgtttgtta ctagtaagga 60
ctcagtattg cgctctactt ggttcctcat gacagctatg cagggatgtg ttcagcccgt 120
tctaccgaac ccttctaaca tgagcgtgcc ttttgattag 160
<210> 28
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 28
gcgagtaact gcttcaatgg gactacaatg tgccacgggt gccctacagt cctcagcccc 60
aattgcccaa aacgaacctt tcaacatcat cccggatttt cactcgaaga ttgtgactgg 120
gggttttatg caacaaccga gctattacat ggttgcgcgt 160
<210> 29
<211> 160

<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 29
ttctcaggca gcccaccccg gcagtccaga tctagccccc tccccttggt acttgggcat 60
ggtgagcctc cgagaccccc tccctctccc ccctcaccag acccccccct ataggtcctg 120
caaggtgcct tcccaaacac cccagttagg catggccacc 160
<210> 30
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 30
gactcctccc taggccccca tggagccacc ccctcaggcc actccaggct actaggccca 60
ggttccaggc aaatgccctc tctgccagtg ccactagcaa cacctcccct atcaaggtgg 120
ccccaggtcc tcacgtagca tgcaggcccc ccgctccatc 160
<210> 31
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 31
ccccagaggc aggtgcccta ccaagctccc cccatgaccc ctaaatcccc caccctgccc 60
cggcggttgc agtggtacca accagtcagg cccctcgcca gtacccttcc atatcttcag 120
cctcctggcc attcgatcag gagccccaca gccctaggcc 160
<210> 32
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 32
tagggcccga gccagcctgt accttgcgcc cctgcccccc ctctacctgg ggaccccacg 60
gtcatccttg acagggtgcc cctcggccca ctcccattct ccttttgtct ccagtaaacc 120
cccagagccc aaggtcagcc tgctgcaggg tttgcctcca 160
<210> 33

<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 33
actgctgcgc ggcgcacctc ccacatgtcc tacccatcac ctcctcagtg ttcactggct 60
gggtctgtcc tcctacaggg tgccaagcgg ggctccattg ccactagaag cccatggtcc 120
agcgtggcta gatccgagcg gggggcctcc accagccgtc 160
<210> 34
<211> 160
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 34
tggaggggct gggcctgctc ccctagtgcg gaatcctgcc ctccggtggc ttgctctttg 60
ggtccacggg tactagaggg gaattatgac cagagccctg cagccccgaa gcggggtgcg 120
ccacagtccc cacgactccg ccaaccttca taccctgtcc 160
<210> 35
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 35
aaatgtataa atttggtgag gactgtaatt ctagttgtac tcctatgtct acaagaccat 60
ctccttacta tagtgggatt aataatattg taaatccggc tatgatctta gacagggaaa 120
atgagttgta accgattgtt aagtatcatt tttccttgaa ttgacatcac ctagcttgtc 180
ttaatgttca tgagaatttc aggctaacca caatgtcaac tatgcgacac catgtatcat 240
catttccact tcacaacaga accgggtcat tttgtgtatt cccatagatt aaatgattaa 300
ccttatgcca ctataatata 320
<210> 36
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 36

taatgtataa atatggtgag gactgtaatt ctagttgtac tcctatgtct acaagaccat 60
ctccttacta tagtgggatt aataatattg tatatccggc tatgatctta gacagggaaa 120
atgagttgta accgattgtt aagtatcata attccttgaa ttgacatcac ctagcttgtc 180
ttaatgttca tgagaatttc aggctaacca caatgtcaac tatgcgacac catgtatcat 240
catttccact tcacaacaga accgggtcat aatgtgtatt cccatagatt aaatgattaa 300
ccttatgcca ctataatata 320

<210> 37
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 37

agcataaaaa gcctataact cgatttttta acattaagcg gtaccgtttc tgcatccaat 60
gacataatat atatgggagc ttactattga gatgcactct taatacggaa ttacgttaca 120
aggtagaggg ctatgactag aattgagctt tatattagca gaagtgtctt gtccagtagg 180
gtcttgaaaa gttattatgt atggtgttca tgagaatcga ggtacattaa ggtgaatcat 240
ttaaatccta gtatggatgt tacacttcaa tgcttttgta caactaactc gggtgcgtaa 300
atatattgaa acaatgttta 320

<210> 38
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 38

actggattgt agctatgcct agcattcctt ctcttgagcc tcaaagtctt atctgatgtt 60
cattccaaca ctcttgaacg gattttaaaa acaaatatgt ataaccgacg caagaattta 120
atatatgaat aagtctcctg ttctagattt aatctcaata gtgattatcg aaaattagta 180
tagatttagt gagaatagag atgtgttcct tccttaatag tctttaatct tggacatggt 240
gagcagtatt atatccgact gttaagtcga gactgtcttt cgtaatgtga cgccttactt 300
tttgagatag agaacctaag 320

<210> 39
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 39

tactataatt aggcaggtga ttgaagaacc tgttctttaa ctatcatata ccagtacata 60
cttcaactat ttttgttgat caggatctat taacgaatca cgtttgactt aatttacaac 120
ttgctcgcgg tattaataaa tagattttat ttgccgtgtt ttcagtacgt ataatgcgat 180
caagcagcaa tgccggaact gttaattgtc ctcgccttac tgatatgttt aaacttcaac 240
ttttcgtctc gaggtttagt aaaataatgt attaaagaat acgaacgtga ttatcccgac 300
ggcacactgt cacttcgtct 320

<210> 40
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 40

gaacaactta tctgagaaca agactacgtc aacttttgta cgtgggataa gttttcctga 60
attctaatta taaatatgga cgtgctcaat gaattaacag acgccacacg acgtttatat 120
cggactgatt acaagtttat tctgtgtaag taacaacaac gctacgatct ctgtatgatt 180
gaatacagtc aaacggttct ataatcacta ctcattctac tgttcgaact aatatcaatc 240
aagtcgttaa taaatcttca tgatcactcg ctatttctag aagtgactcg ctaaggacga 300
taattattct tcgttcaaaa 320

<210> 41
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 41

gctaacacca tggctgctag aattaaagta ttgactgact catacgtgga atacccaagc 60
caaccctgtc ccctcaacta aaggtgattc tggacctttc aagggttggc aggtatttac 120
ccctatcgac aagaagagtg acctacagga gaaatcgatc agaggcagtt caagatcaac 180
tccctggtcc tcctggcctg gagctcatga tgaagagtcc agtgcctcct gctcaccagc 240
atcccatgtg acgcaggtca ctaggccttg tggccttaaa aagcccagca catactgact 300
agggcagttc agcttataca 320

<210> 42
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 42

ggctcaccag tggctcccca tccgatcaag ctacccaact aatgtactca aggtacataa 60
ttaaggtaga accaccgaaa gtctactagt gatgttcact ctgtccctgg gataagagta 120
gcataaggcc actaagctcc actacctcag ccaacgtaat gtctctttcc agcgtctgtt 180
caaaatgctc ttggtagtgg ttctgctagg taagggcagt tcctttgcca gggtctagac 240
ctagcccagt gtggttaccc ccatgaacaa cagggtggag gcagggtcaa cccagctact 300
ggccataatt tctagagtca 320

<210> 43
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 43

gccactgggg accacctcat taccaggctt tggcatgctg taatgtctcc gatccttgag 60
atgccctgcc cccaatcgca gatgtcagta ggcagctagc acaactgaaa ctactccacc 120
ccaaccgtga tcctggtgca aggcttccca aagaaaatat ttaactaaat acagagatgc 180
tacctaaatc ccgctgggag ttaatcagat tcgatccgcg acccccctttg gtgtaagagt 240
gtagcttgct tcttataccc tctccccgct ccacaatagg agcctacttc accacaccaa 300
taaggtgaac atcctatctc 320

<210> 44
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 44

agggagatct agcctggcta agcaggggca actagtgcac ttctttccac tccagcgcac 60
ttcaccatta atggatgtac agatgattgt tagtttgact ctcatcagca attcactccc 120
tacttacgtt tgtggcccct tacgtctaga tatggggtcc gagtagcccg actgctttcc 180
tcatcagttt tggcagcctg caaccaaacc ccagtagata aaggcagtgt gctacacgtc 240
cgggggtaag aagcctgggt tcctttaact aagtactcca accacattac aggggcatcc 300
cgctcagtag ttgatggtca 320

<210> 45
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 45

agcgggccta ggttacaccc ccgcaacatt tctaaatgca tctagggacc ttactggcac 60
agttcagccc gccagttaat tatcttattg agatcctgca gaggataatc tcctttcgga 120
attaccacaa cggattcgcg taaactagtc tgcgattgct ttgtaagcca agggctacac 180
tgtatccagg ggctggaggg tttagaagtt tccgttcctc atttccgaaa ctagacctga 240
taggctgtcg tgtccacgga ttctaccaag cattcgctcc gtacccactt gctaccgact 300
gttgccgttg cctacaggta 320

<210> 46
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 46

atacctcagt ttatattgga cctctagctg ccggttagag aaactattag aaagcacggt 60
tctatacgga cattcttggg ctgtatgata caaacagagg ccggtacgac tacttccgct 120
ccatgtaatt gacgaagagt cttctcccta gagatcgctt atccttattg ctaatgcttg 180
ccataggccc cgcttagcac gactgccgat acgtgaatgc tattgagtac cggcccagtc 240
gtcccgcatc tcccacactg aaccgggttc tcagctatgt caactgtcct tcttgtcttt 300
gggtgcaggg tcacctgccc 320

<210> 47
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 47

tgtatgattt gaagagattt gtatatacac acattgtttt gtagcataga aaaggagttt 60
ttgtcaaccg gtagcccacc ctgattctca accaagcctg tagatctgta attggggtct 120
taagtccttg ttaaattctg gacagcacta tgatttttta cattctaaat cattatacca 180
aggtatcttg tcttatcttc agagtgtcca gcctgtcgat agatcggaat acaatcgtat 240
aattaattgt taagcatgtt tcttgtacat acaggtcagt tacatcaaca tacttataaa 300
cagtgctgta atatttgtga 320

<210> 48
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 48

gcaattgatg ataactgtga gtgattttgt ttcttttgga gactacctaa cgcttatgac 60
tttgagtttc tggtatgatt caaaagtaga atacctgtag cacgagcacc aatatcattg 120
ttaactgagg aagttcatgt acttaccgaa ttataggaaa attcagtagc ttttctttgc 180
ctactaactt aggttgtgtt catcgaaatc ataacatcac ataactattg tcttccataa 240
gcactcagga cttcaagtaa aaaggatgaa gcctattcca tttcacatct gaataacttt 300
agcaaagtgt aagaagctaa 320

<210> 49
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 49

cccatgcatc aaactggctc cctcgtcttc cgattgtcta gctcataggc tctcgaagca 60
tctaagggct attccgggtc ctgcaagtat aagcttcatt attcctaagg atgtggggag 120
tgactcagag gtccagatcg cactgtgggc cagactgaca cagcttcaaa ggaagggcct 180
ccaagtccac tgcactcaga attaagaatt ccctgacgca ttatcttgag aaaggcactg 240
gtccatgtct cttgtatctc cgagaccaac ttaaaggagg gatgggagct aacaggcacc 300
tcccgactta tcttaccagt 320

<210> 50
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 50

gaacttccaa atacaccgtc ccatctgttc agtcagggat tggggtgaga gatatatcgc 60
atcaggaatt acgaaacctt atgggcaagc agtgattagc taagcctgga aacctggcaa 120
ttaacacctc aaactggatc catgcctttc taactatttc ccacccccttg gcagtctagg 180
gctggcgaga ggccctgcta atacttgagg cgtagatggg gggcggcttc tctgcaaact 240
ggtgcccgct ggggctctaa taattattct tccttgttcc acacaaccag cccctcgact 300
ttaagcagct atgctatgca 320

<210> 51
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 51

ttcggcactt gtctgccctc gtcagaaaat gttgggtaaa accctaggtt gtagtttggg 60
tctggcgagc gggaaagtgc atgctcggcc catgtgggct ccaaactgaa ggttattaga 120
ttcctagatg gtgagaccgc atacaaaaag ggccctggaa agaggtcact tcaacgcatc 180
tcctgatatt ggtctggtat ccacagtaga gctattgtcg cctaacagtg atgccgcgcc 240
gtcctgtatt ggtgcgcgag acagcttata cgtacctgaa tggcgataat tatccgaggg 300
gcagactcaa gcttaagaaa 320

<210> 52
<211> 320
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 52
ttgggccgcc ttgtccgcaa ccagcaaccg atagcagtcg gactccgagt cagtagtgaa 60
gtgctttagc gttaagtgtt tattgtgaat gagccctctc tcccccaaat cacaagaggt 120
ggcggaaaaa cacgaagccg aagtacaccg acaaggaacg gtgctctcaa gagttgccag 180
ccattgctag acagagtaat ttcctcctcc aggcggaatt caacagtcct cagtcccaga 240
attatcttgg gaaaggatgg acacgaatat ttggaacagt ggacgccgac ccgtttaatt 300
acagggttcc ctgagattgt 320

<210> 53
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 53
tgtctaaatt aaagttgtga tctttgactt agcaacgtct caccctatag cctaccagac 60
aagaattatg aagaacatat 80

<210> 54
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 54
gtacaccatc attatcctca tagcttaggc tccacgtgcc tacagggcca taaggcttgg 60
agattcactg ttagctgctc 80

<210> 55
<211> 80

<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 55
gcctccccaa ctatagggtc aggaaggatt atggcacccc acacgtattt cacccgatct 60
gtaccagtaa tcatacatgg 80
<210> 56
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 56
gctaccagtg gccccccccc accggatcca tccctaacct caccccccctg actgctaacc 60
tgggatggtg aagcctgggc 80
<210> 57
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 57
ggttatgccc ccgccctgca tcctccctgt cacacgtgcc caaccctagc aatgtgtggc 60
ccccccctgct gtctcccatc 80
<210> 58
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 对照序列
<400> 58
gctggtgcac cgctgccccc accctccacg tctgtcacag cctcggtagg tcctgatttg 60
atgcttgggt gctcggctgg 80
<210> 59
<211> 160
<212> DNA
<213> 人工序列
<220>

<223> 对照序列

<400> 59

gtataatcat aacaaaggcc taatgaaaga cgctgatttg aacatagttc cctcatcatc 60

tgatattgtc ctacgtgtct tttttcgatg agtgcacaat atggtgtgaa gacctattac 120

aatcaaaaag tataaactag cgactaagat ctcagaatta 160

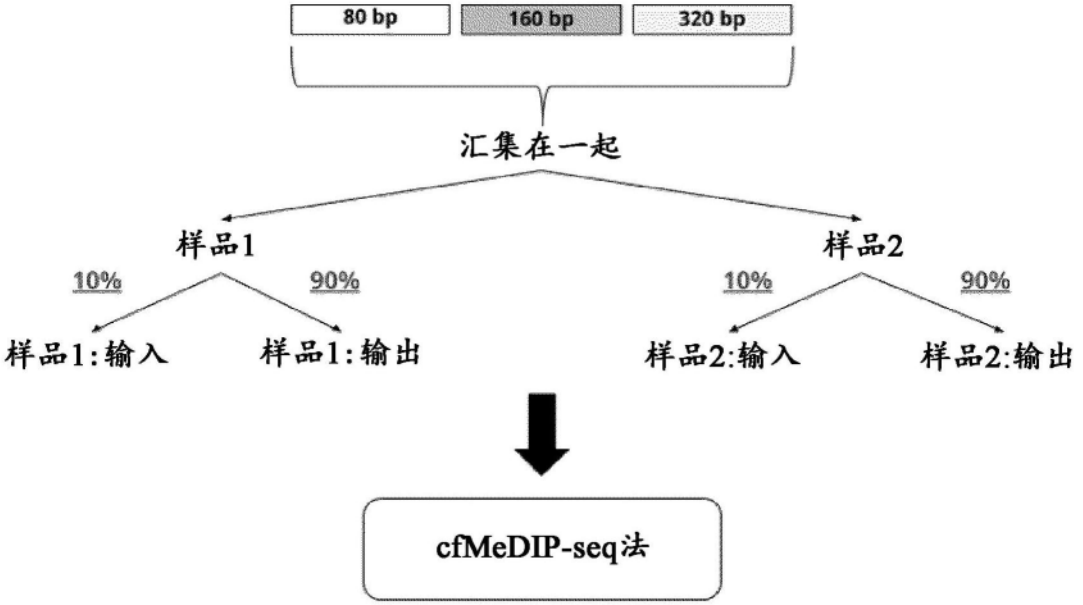

80 bp
160 bp
320 bp
汇集在一起
样品1
样品2
10%
90%
10%
90%
样品1:输入
样品1:输出
样品2:输入
样品2:输出
cfMeDIP-seq法

图1

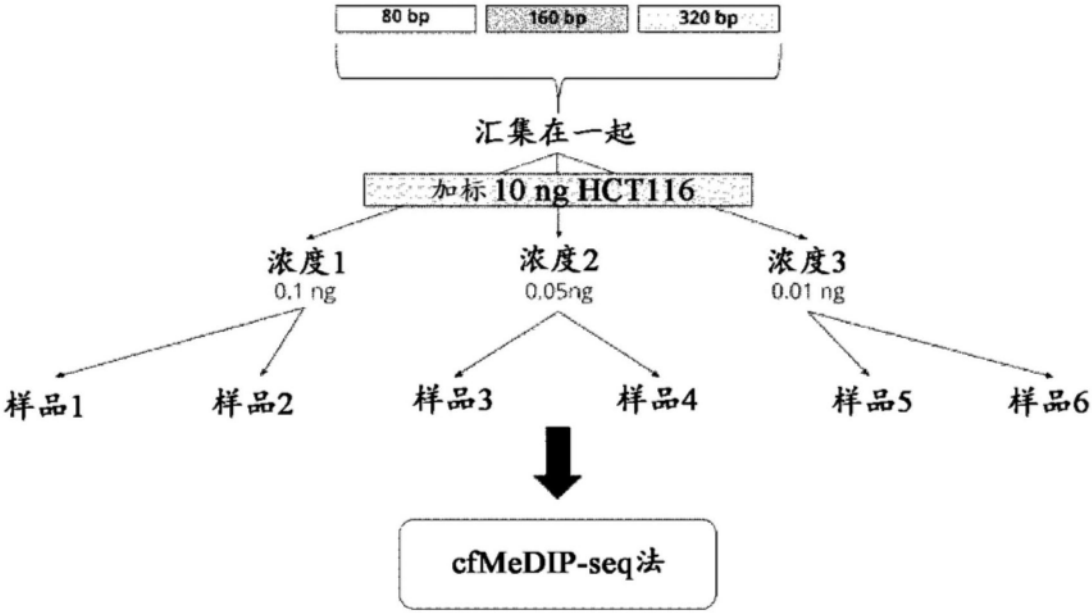

80 bp
160 bp
320 bp
汇集在一起
加标 10 ng HCT116
浓度1
0.1 ng
浓度2
0.05ng
浓度3
0.01 ng
样品1
样品2
样品3
样品4
样品5
样品6
cfMeDIP-seq法

图2

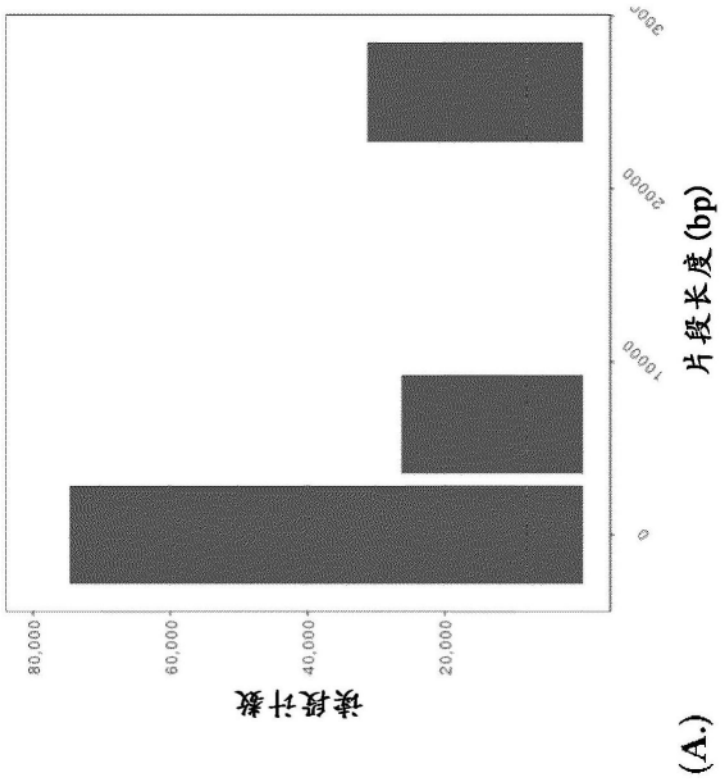

读段计数
80,000
60,000
40,000
20,000
片段长度(bp)
0
10000
20000
300
(A.)

图3A

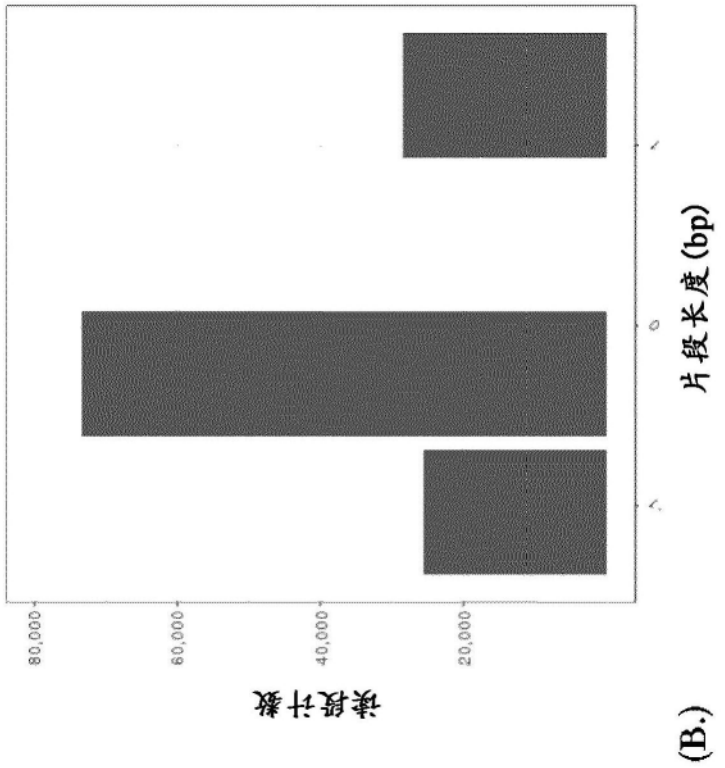

读段计数
80,000
60,000
40,000
20,000
片段长度(bp)
0
(B.)

图3B

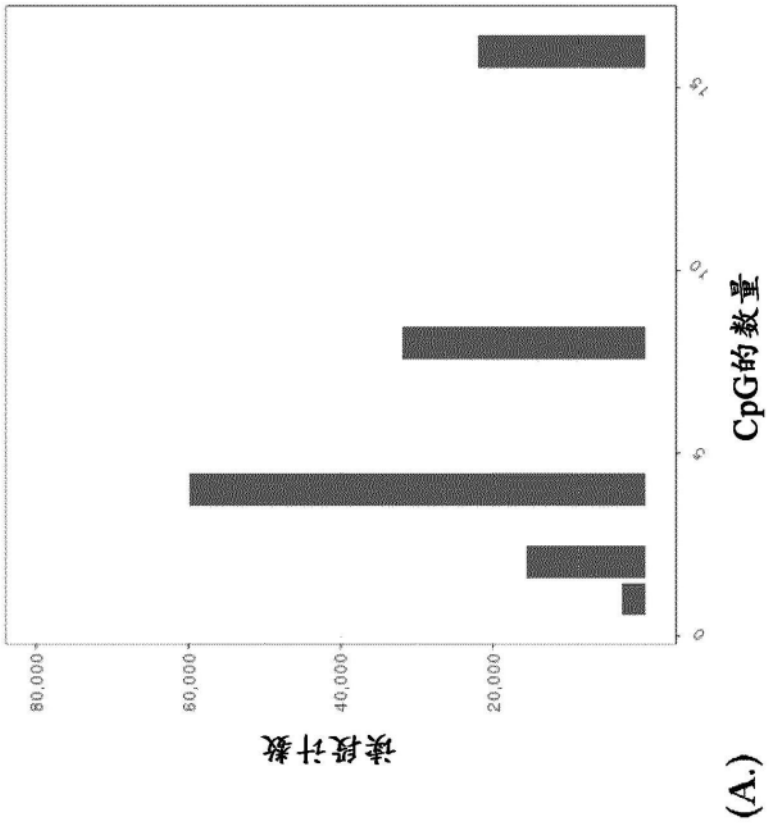


图4A

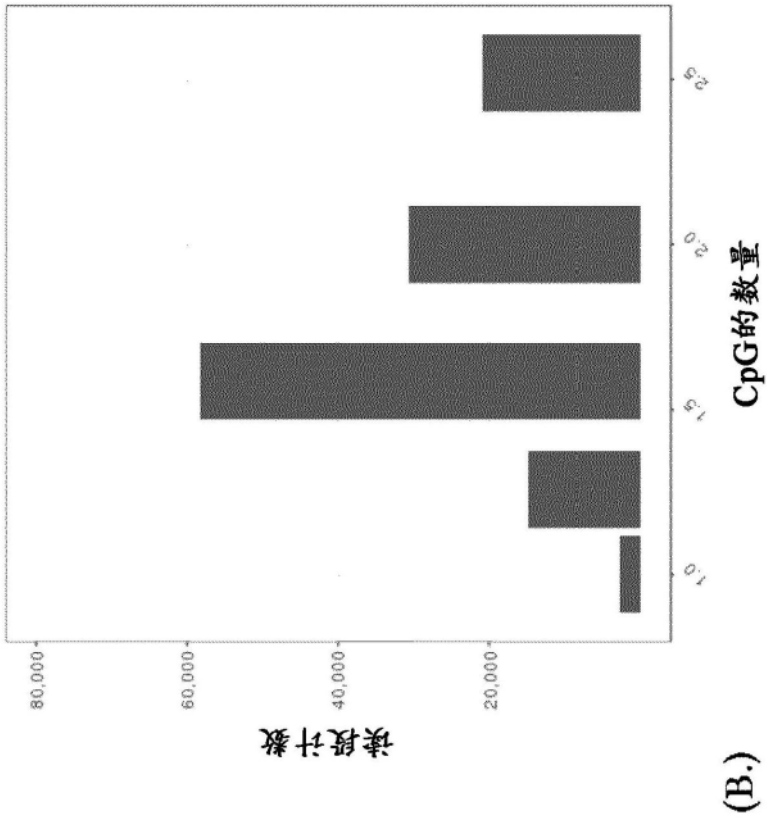


图4B

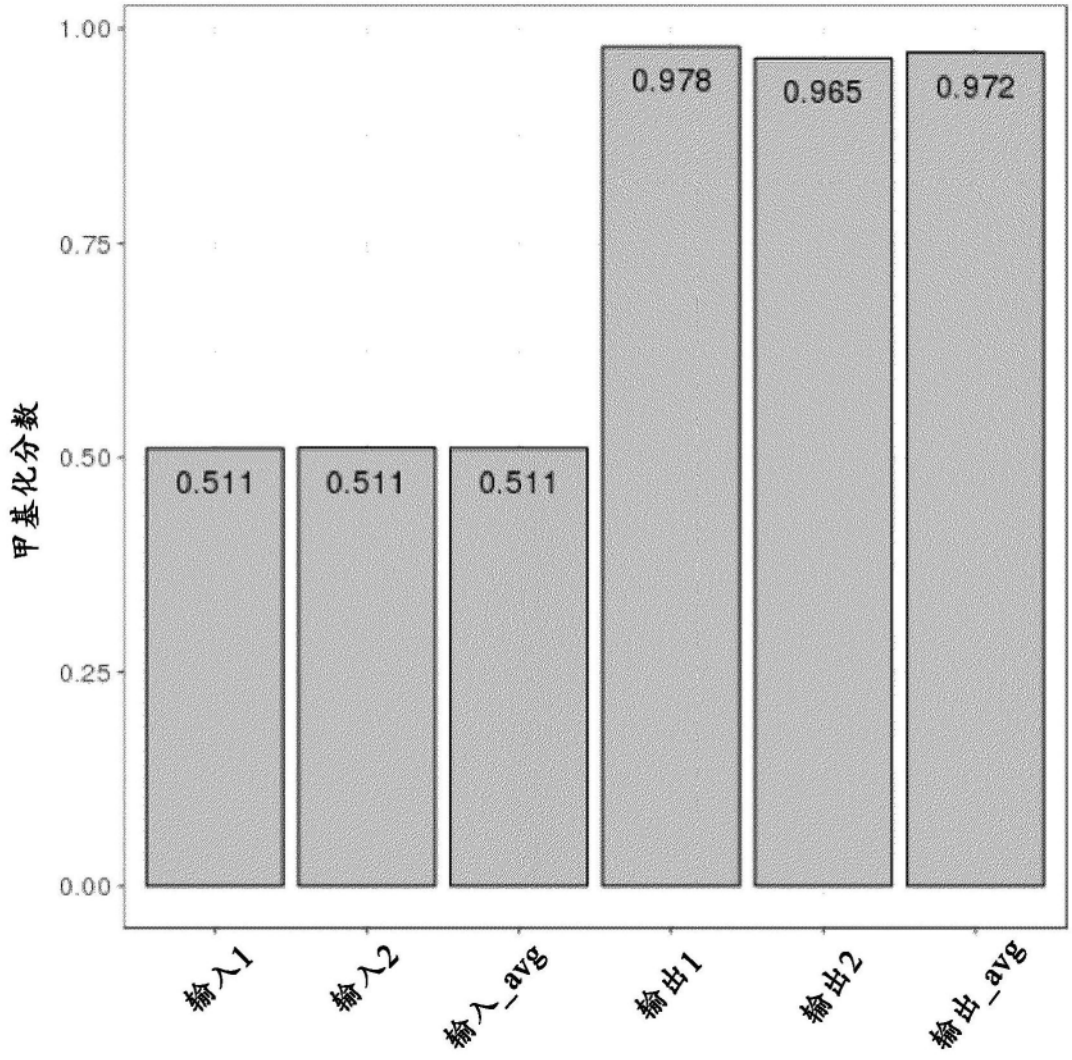

1.00
0.75
0.50
0.25
0.00
甲基化分数
0.511
0.511
0.511
0.978
0.965
0.972
输入1
输入2
输入_avg
输出1
输出2
输出_avg

图5

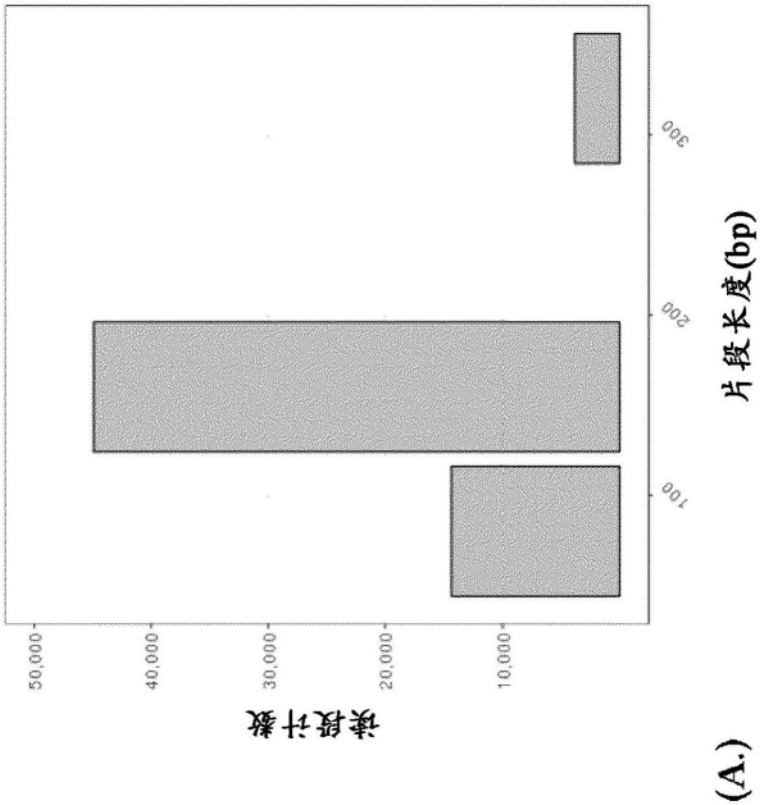


图6A

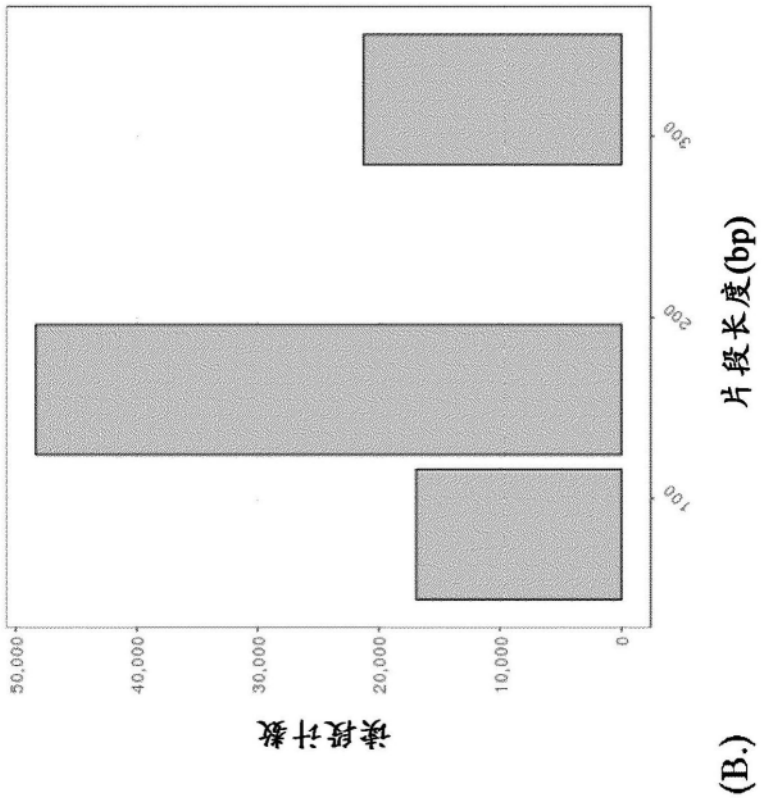


图6B

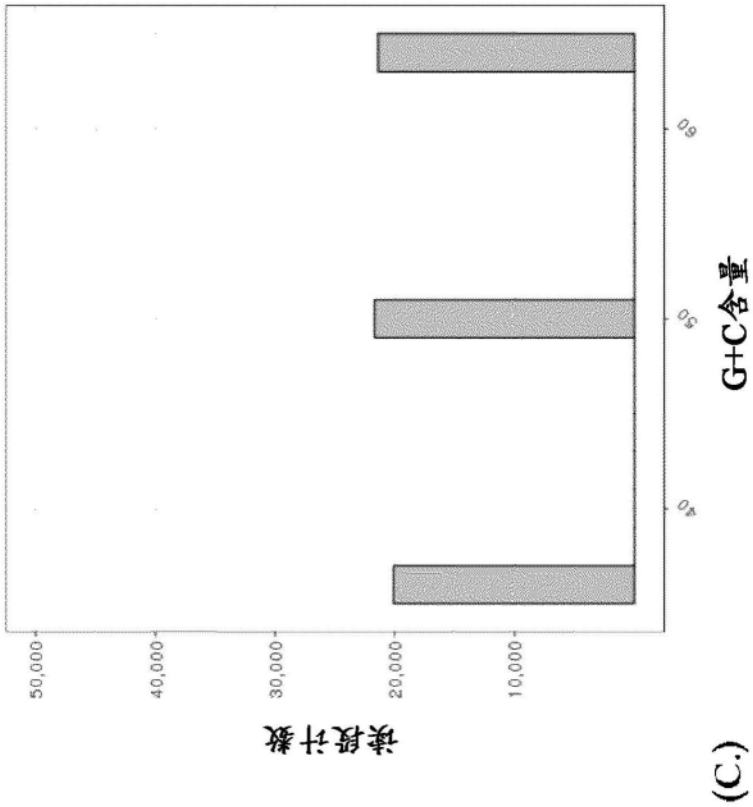


图6C

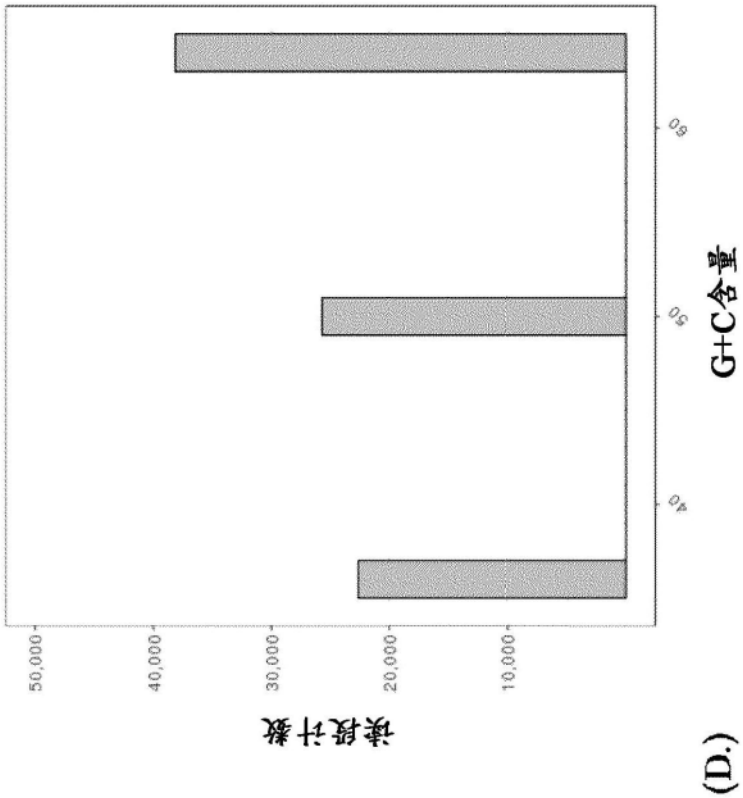


图6D

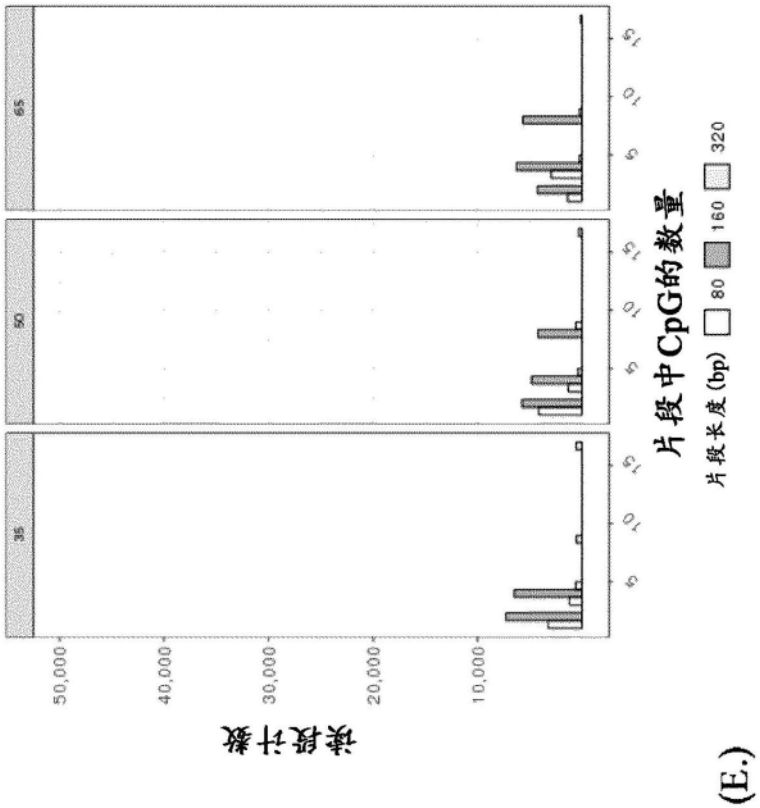

读段计数
50,000
40,000
30,000
20,000
10,000
35
50
65
5
10
15
片段中CpG的数量
片段长度(bp) 80 160 320
(E.)

图6E

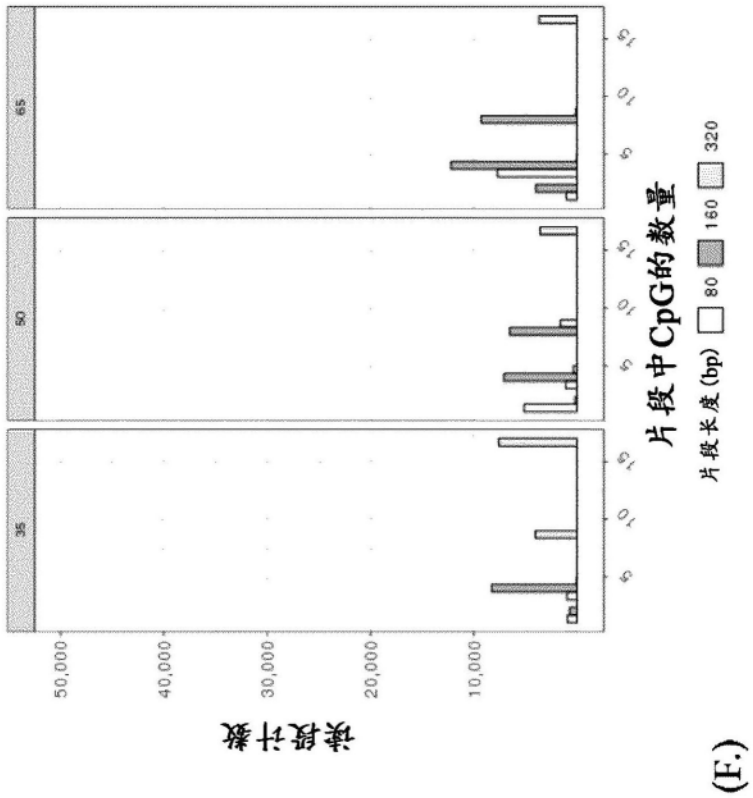

读段计数
50,000
40,000
30,000
20,000
10,000
35
50
65
5
10
15
片段中CpG的数量
片段长度(bp)
80
160
320
(F.)

图6F

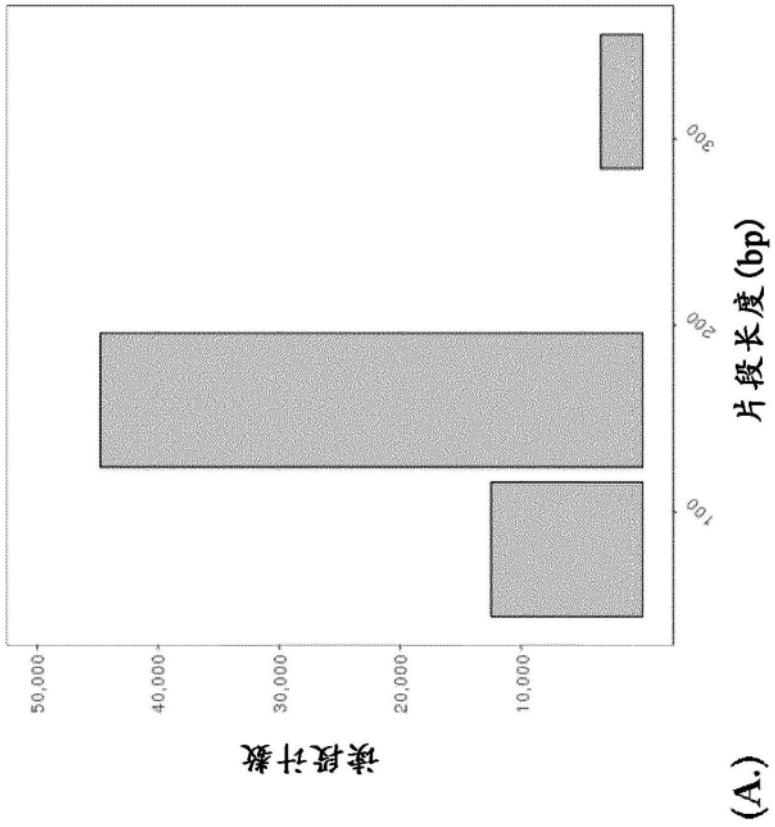


图7A

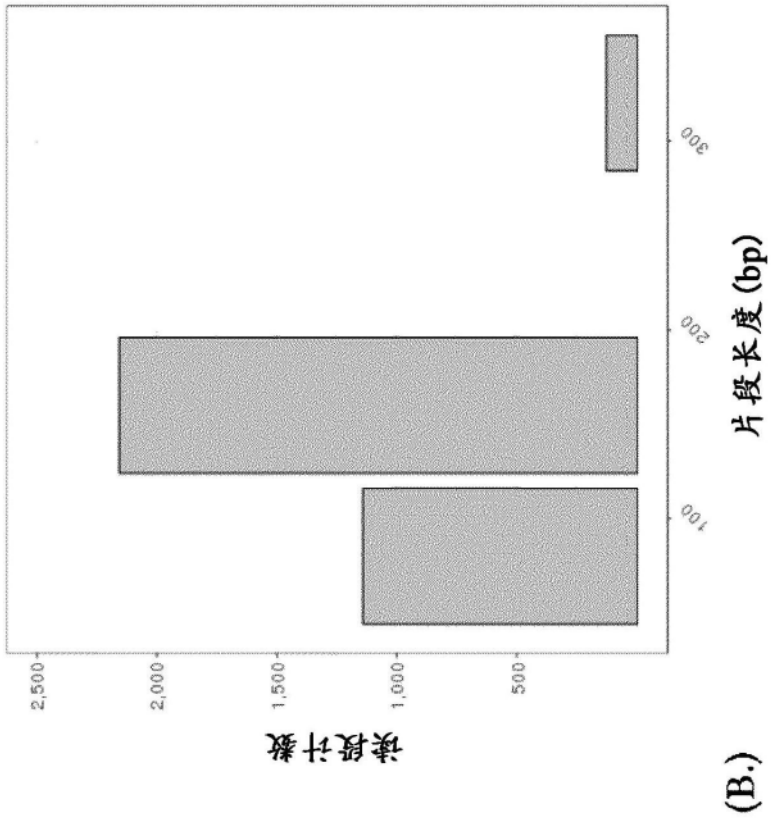


图7B

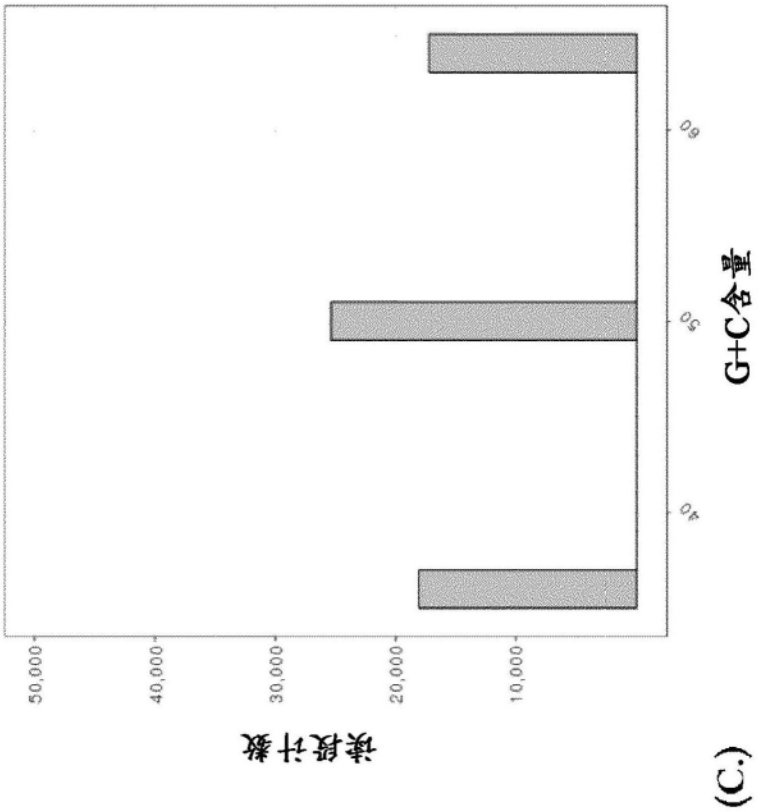


图7C

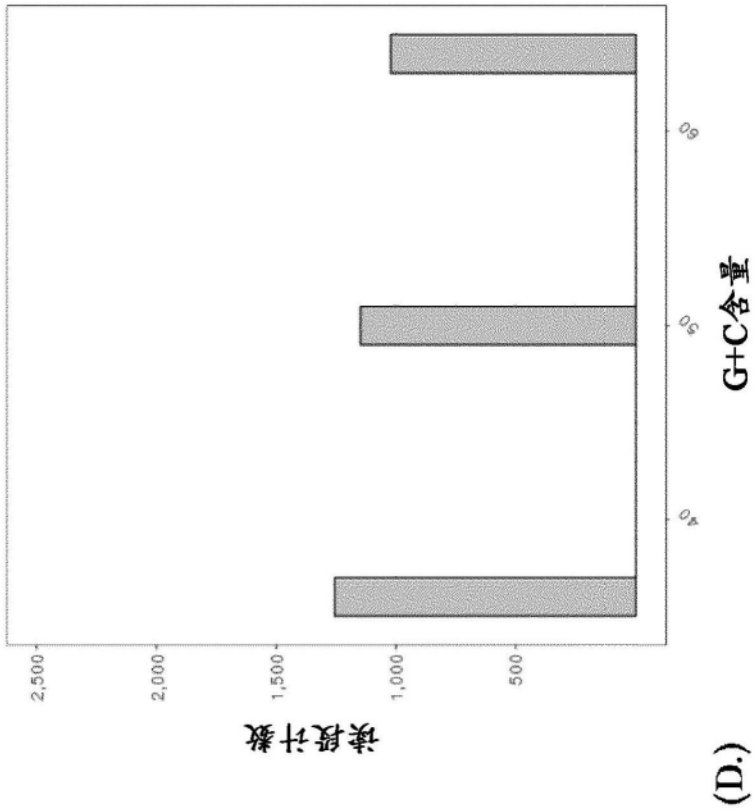


图7D

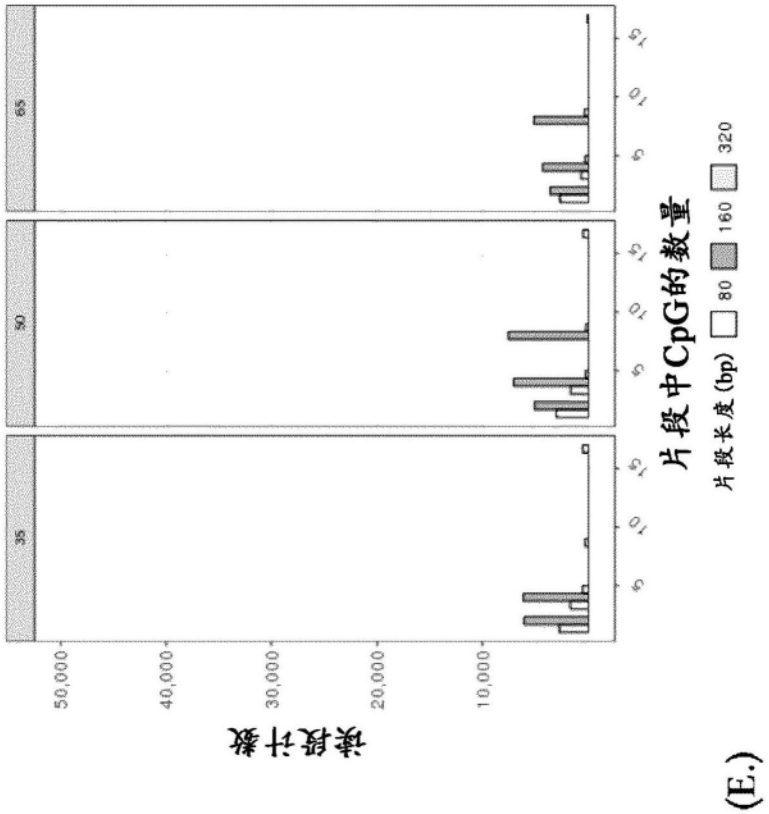

35
50
65
50,000
40,000
30,000
20,000
10,000
读段计数
片段中CpG的数量
片段长度(bp) 80 160 320
(E.)

图7E

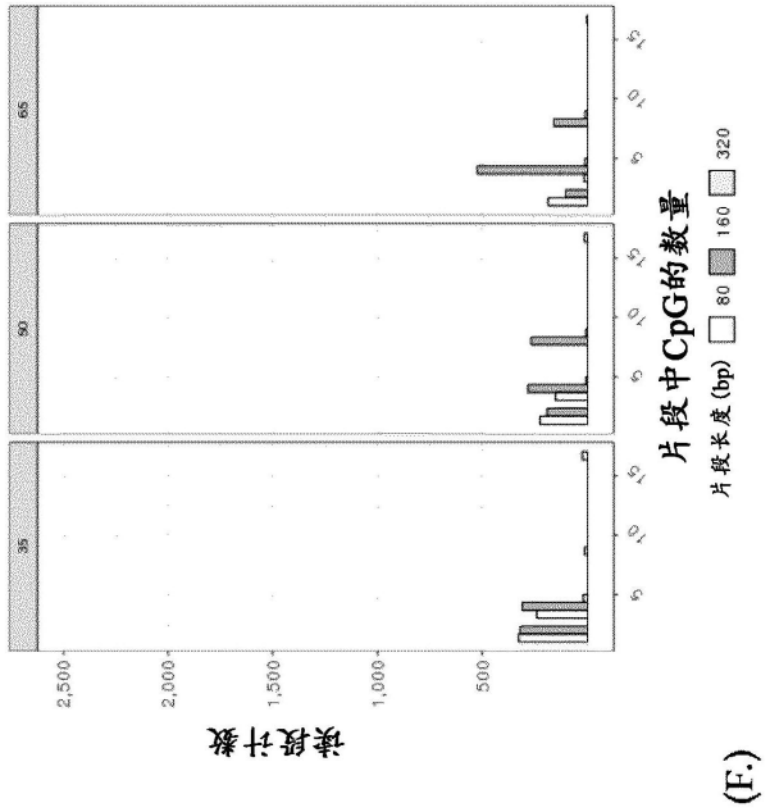

读段计数
2,500
2,000
1,500
1,000
500
35
90
65
5
10
15
片段中CpG的数量
片段长度(bp)
80
160
320
(F.)

图7F

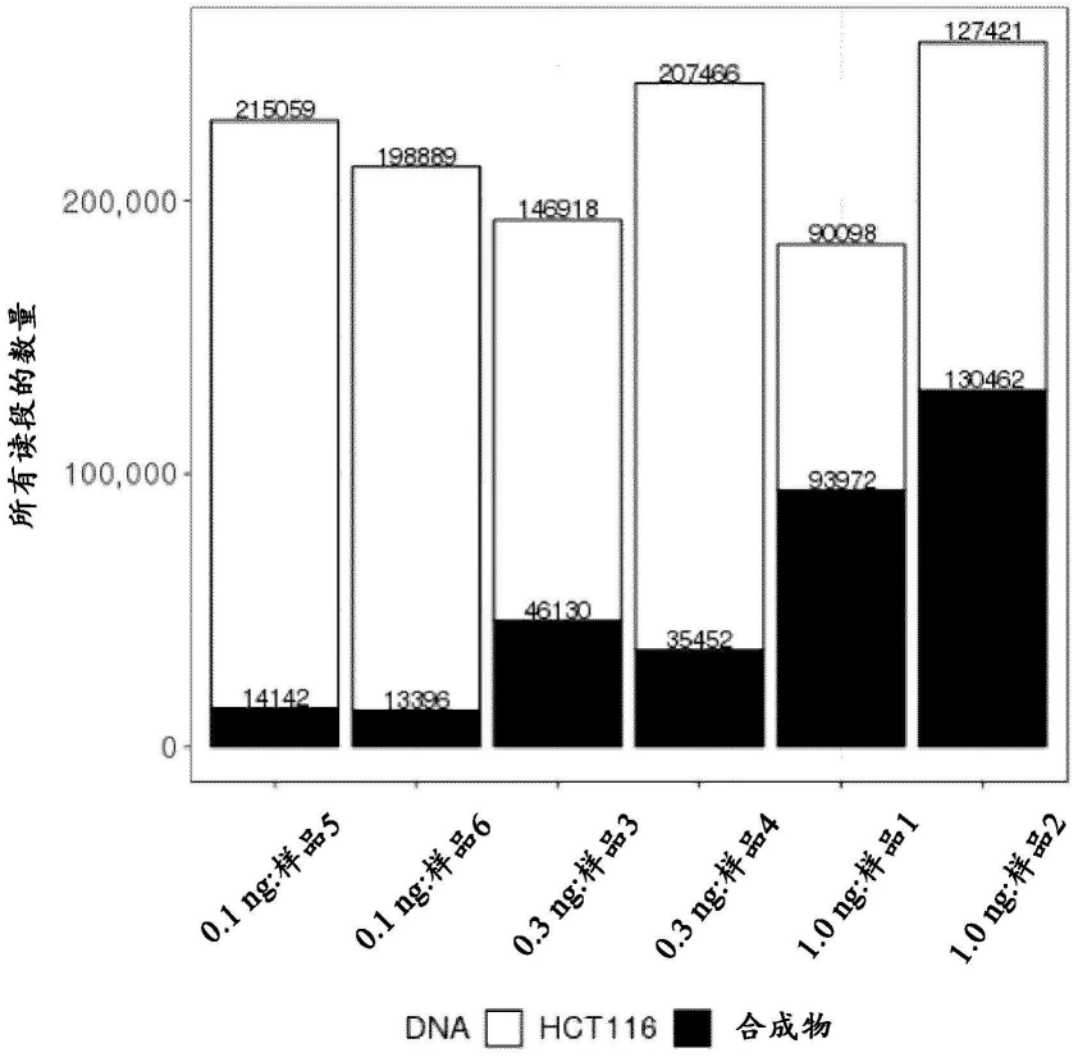

所有读段的数量
200,000
100,000
0
215059
198889
146918
207466
90098
127421
14142
13396
46130
35452
93972
130462
0.1 ng:样品5
0.1 ng:样品6
0.3 ng:样品3
0.3 ng:样品4
1.0 ng:样品1
1.0 ng:样品2
DNA
HCT116
合成物

图8

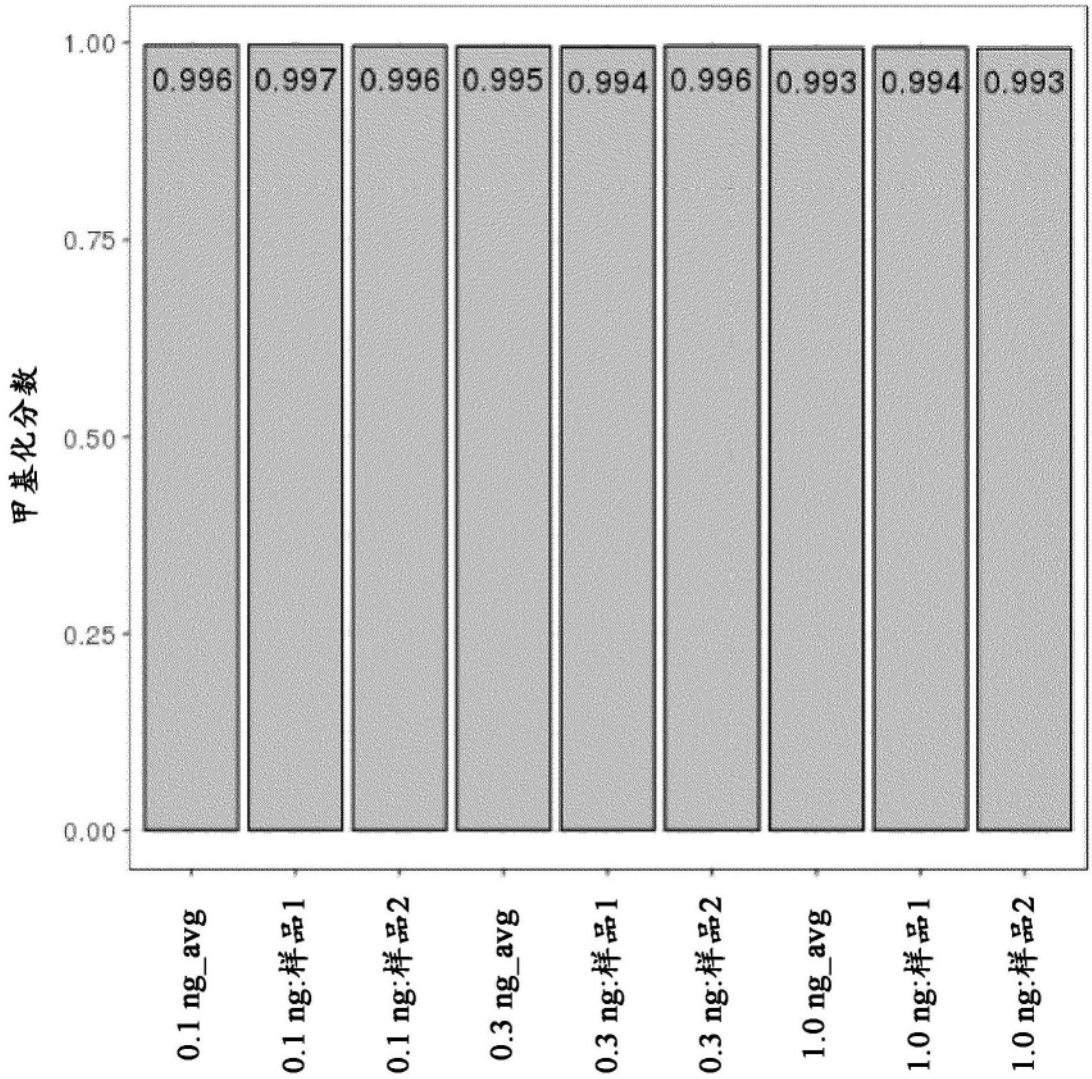

甲基化分数
1.00
0.75
0.50
0.25
0.00
0.996
0.997
0.996
0.995
0.994
0.996
0.993
0.994
0.993
0.1 ng_avg
0.1 ng:样品1
0.1 ng:样品2
0.3 ng_avg
0.3 ng:样品1
0.3 ng:样品2
1.0 ng_avg
1.0 ng:样品1
1.0 ng:样品2

图9

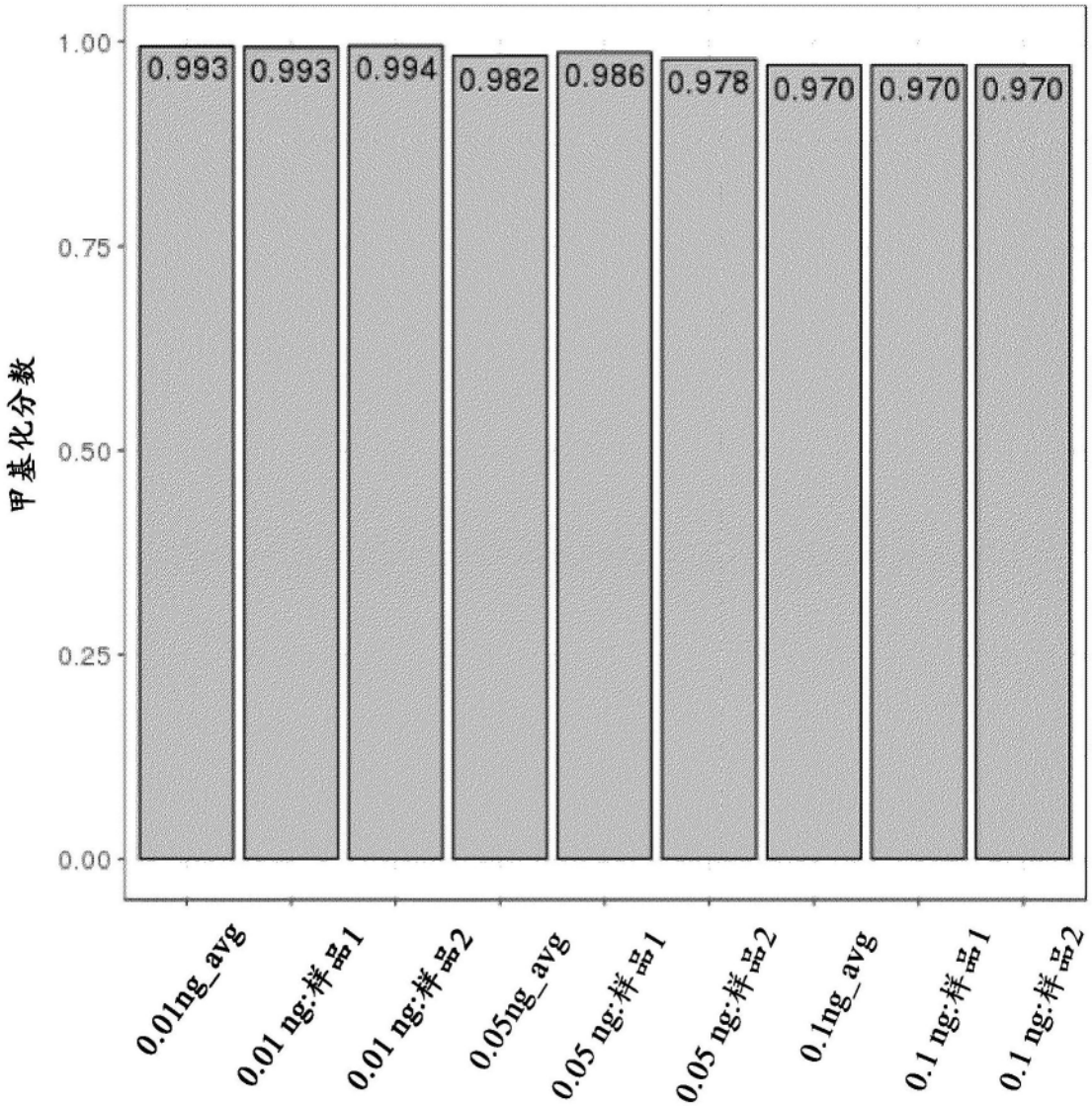

甲基化分数
1.00
0.75
0.50
0.25
0.00
0.993
0.993
0.994
0.982
0.986
0.978
0.970
0.970
0.970
0.01ng_avg
0.01 ng:样品1
0.01 ng:样品2
0.05ng_avg
0.05 ng:样品1
0.05 ng:样品2
0.1ng_avg
0.1 ng:样品1
0.1 ng:样品2

图10

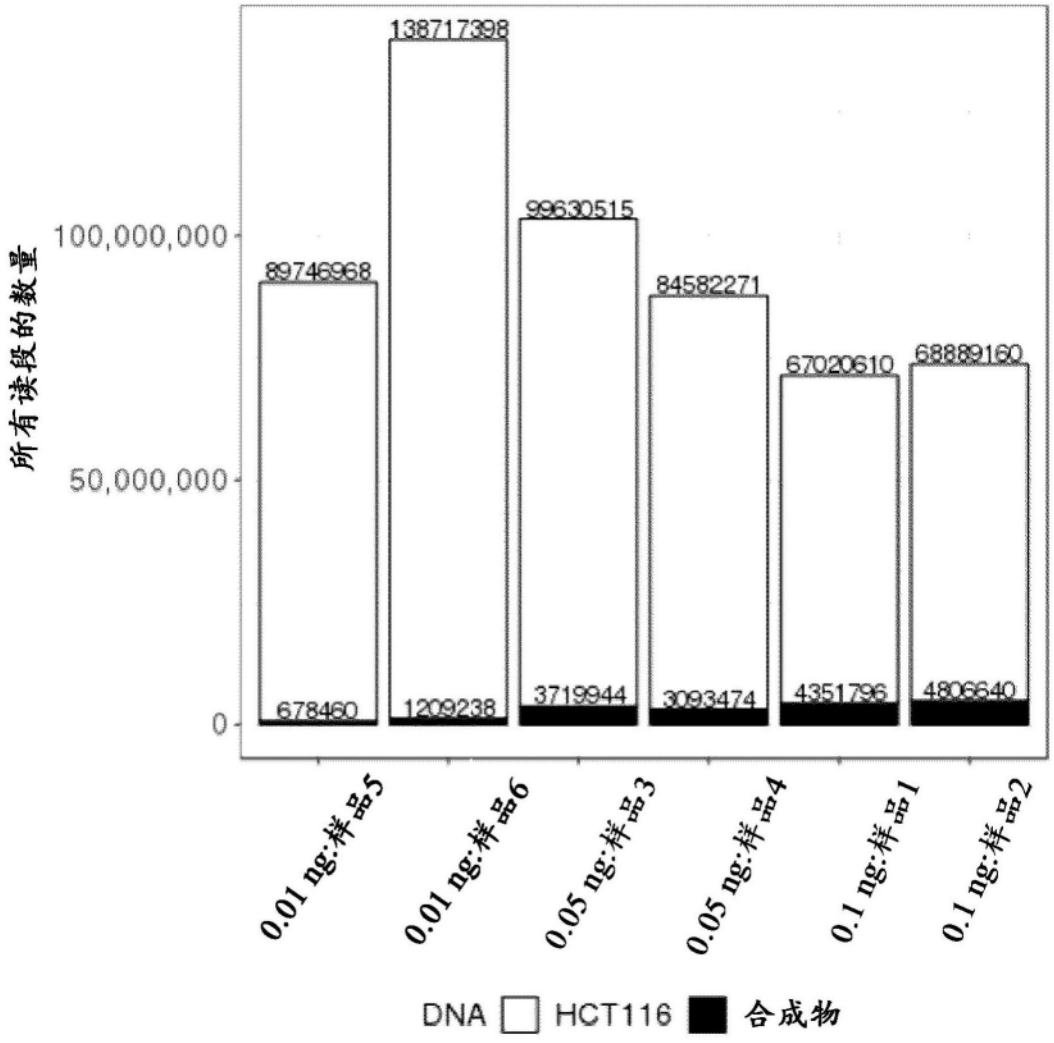

所有读段的数量
100,000,000
50,000,000
0
89746968
138717398
99630515
84582271
67020610
68889160
678460
1209238
3719944
3093474
4351796
4806640
0.01 ng:样品5
0.01 ng:样品6
0.05 ng:样品3
0.05 ng:样品4
0.1 ng:样品1
0.1 ng:样品2
DNA
HCT116
合成物

图11

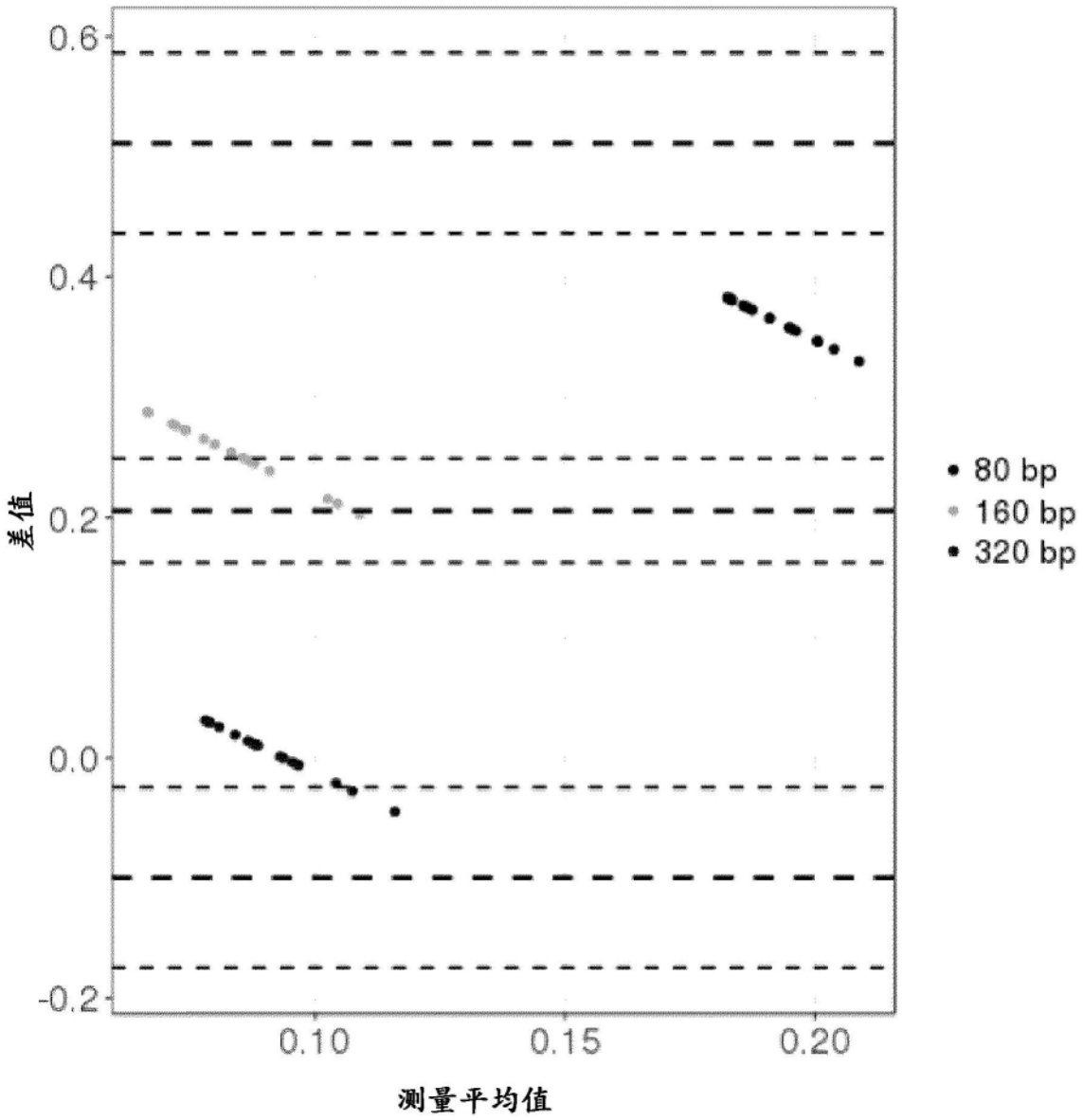


图12

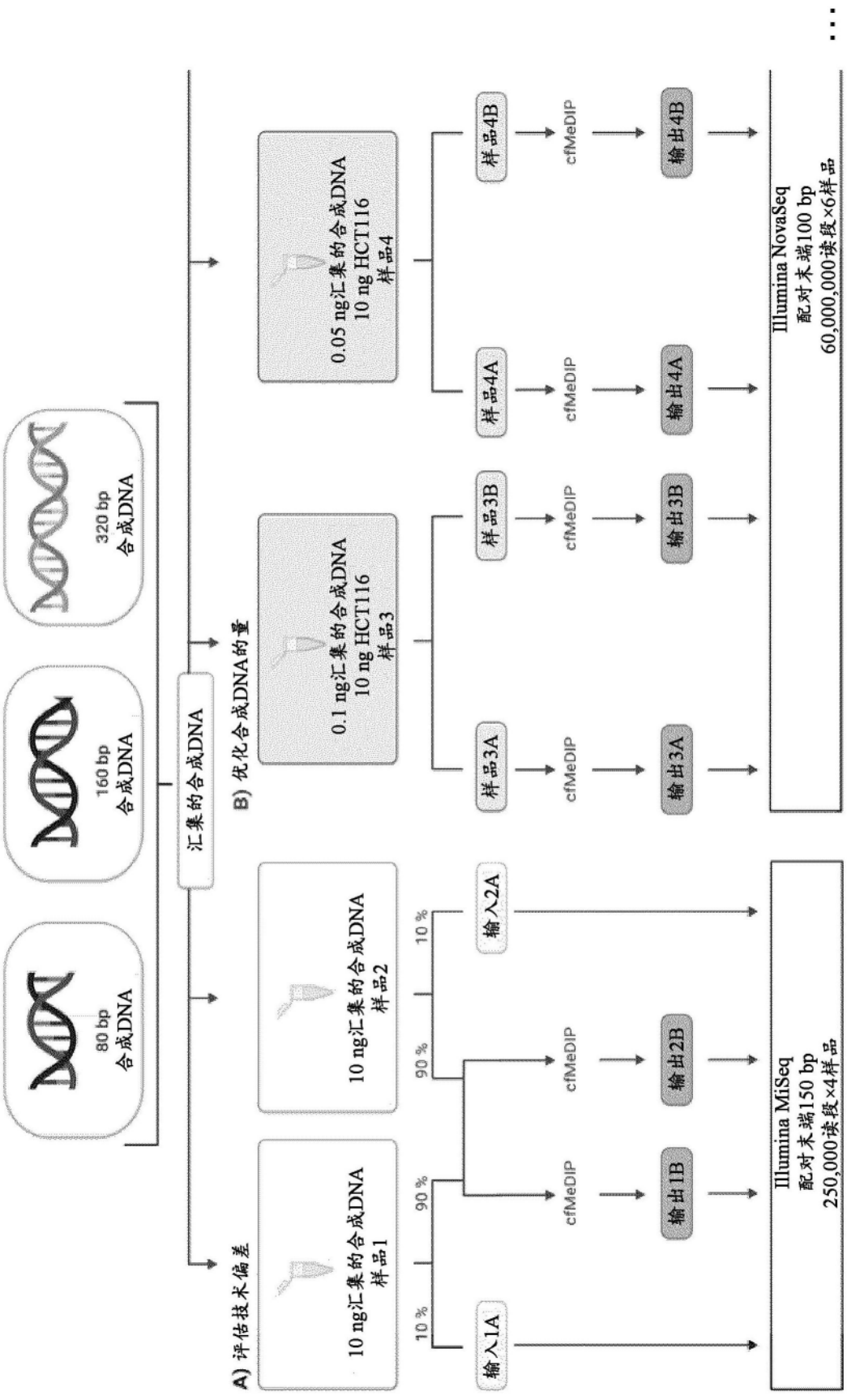

80 bp
合成DNA
160 bp
合成DNA
320 bp
合成DNA
汇集的合成DNA
A) 评估技术偏差
B) 优化合成DNA的量
10 ng汇集的合成DNA
样品1
10 ng汇集的合成DNA
样品2
0.1 ng汇集的合成DNA
10 ng HCT116
样品3
0.05 ng汇集的合成DNA
10 ng HCT116
样品4
10 %
90 %
输入1A
输入2A
cfMeDIP
样品3A
样品3B
样品4A
样品4B
输出1B
输出2B
输出3A
输出3B
输出4A
输出4B
Illumina MiSeq
配对末端150 bp
250,000读段×4样品
Illumina NovaSeq
配对末端100 bp
60,000,000读段×6样品

图13

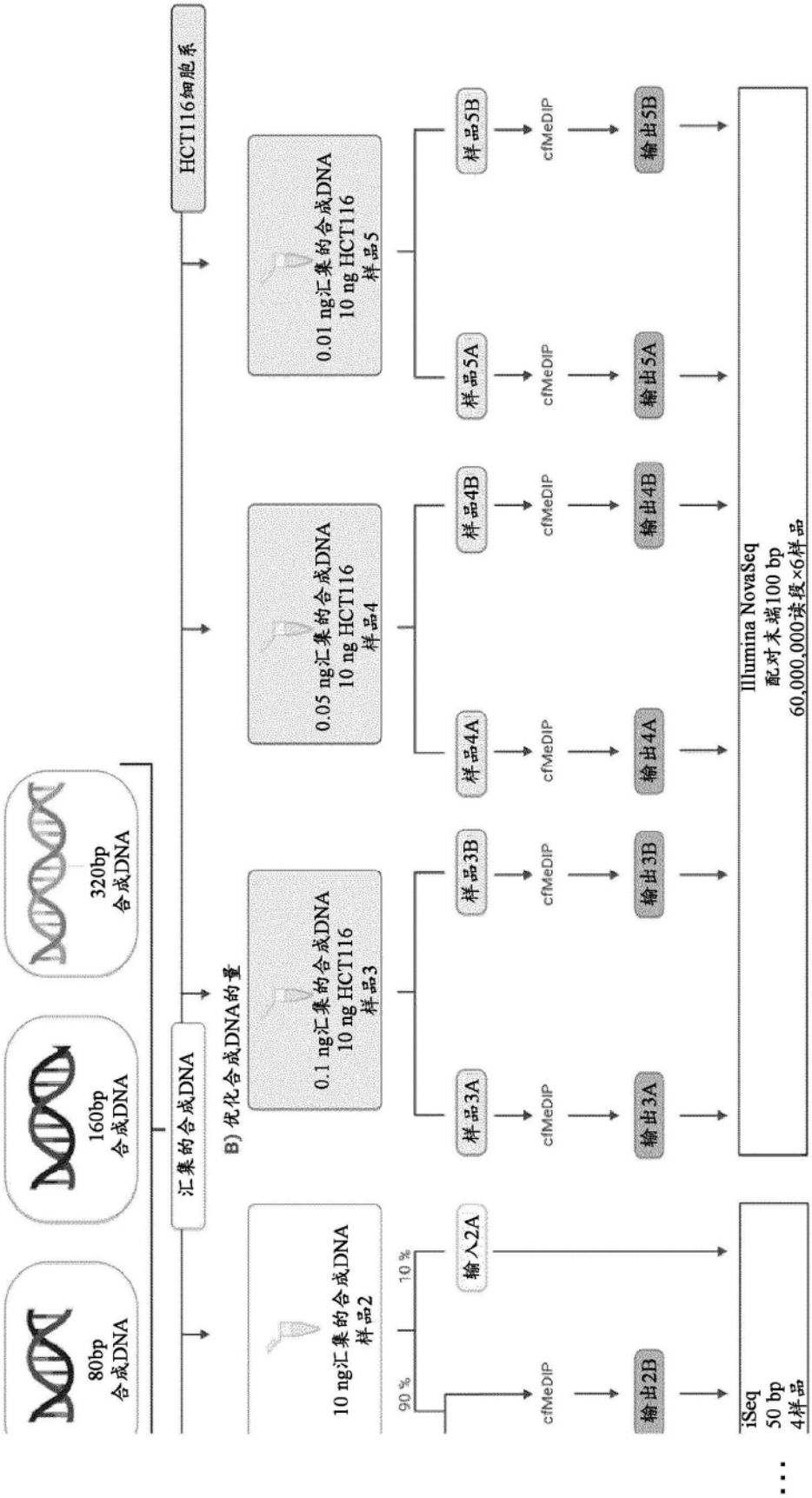

80bp
合成DNA
160bp
合成DNA
320bp
合成DNA
汇集的合成DNA
HCT116细胞系
B) 优化合成DNA的量
10 ng汇集的合成DNA
样品2
0.1 ng汇集的合成DNA
10 ng HCT116
样品3
0.05 ng汇集的合成DNA
10 ng HCT116
样品4
0.01 ng汇集的合成DNA
10 ng HCT116
样品5
90 %
10 %
输入2A
样品3A
样品3B
样品4A
样品4B
样品5A
样品5B
cfMeDIP
输出2B
输出3A
输出3B
输出4A
输出4B
输出5A
输出5B
iSeq
50 bp
4样品
Illumina NovaSeq
配对末端100 bp
60,000,000读段×6样品
...

图13续

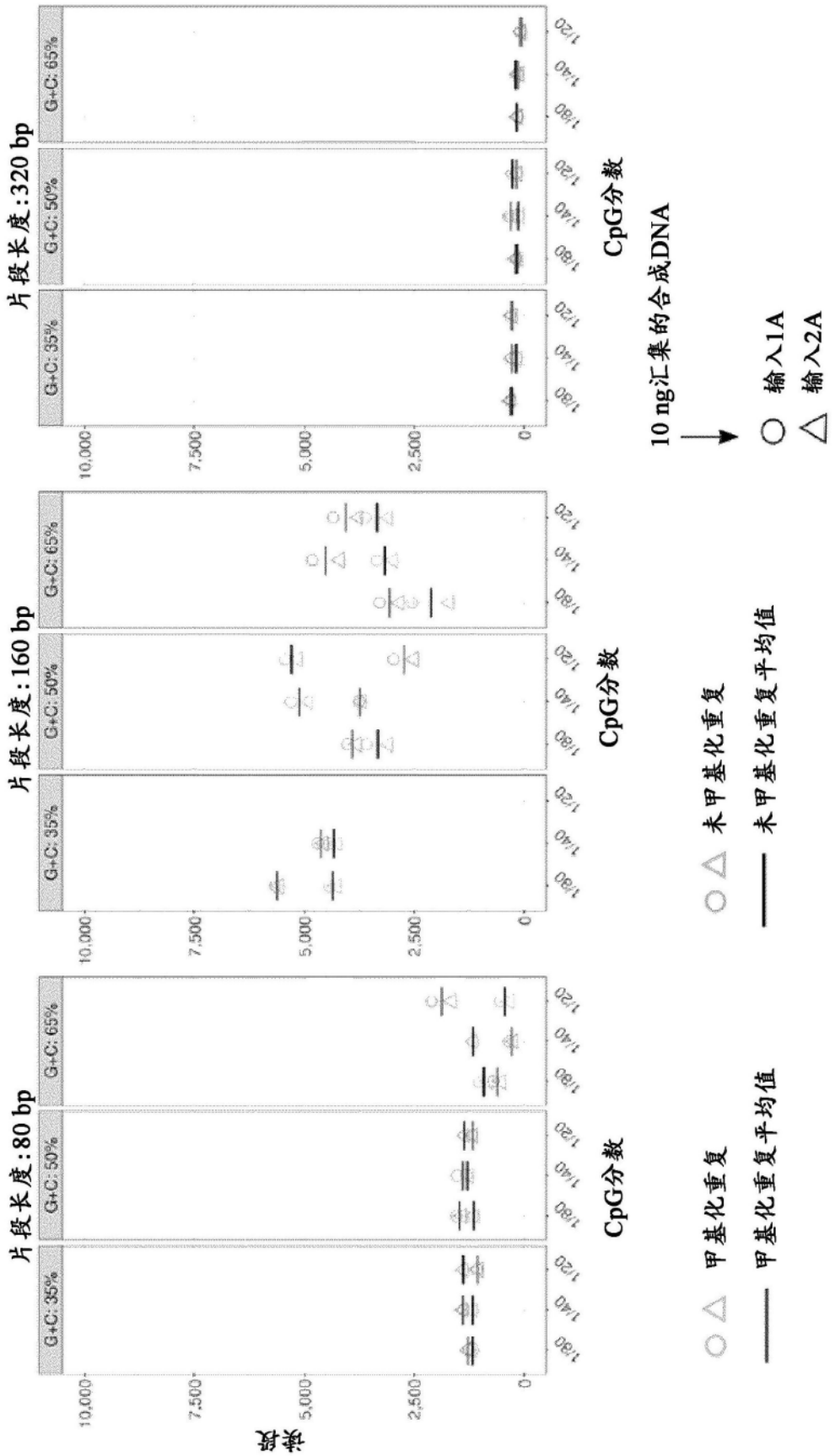

片段长度:80 bp
片段长度:160 bp
片段长度:320 bp
G+C: 35%
G+C: 50%
G+C: 65%
读段
CpG分数
10,000
7,500
5,000
2,500
0
1/20
1/40
1/80
10 ng汇集的合成DNA
输入1A
输入2A
甲基化重复
甲基化重复平均值
未甲基化重复
未甲基化重复平均值

图14A

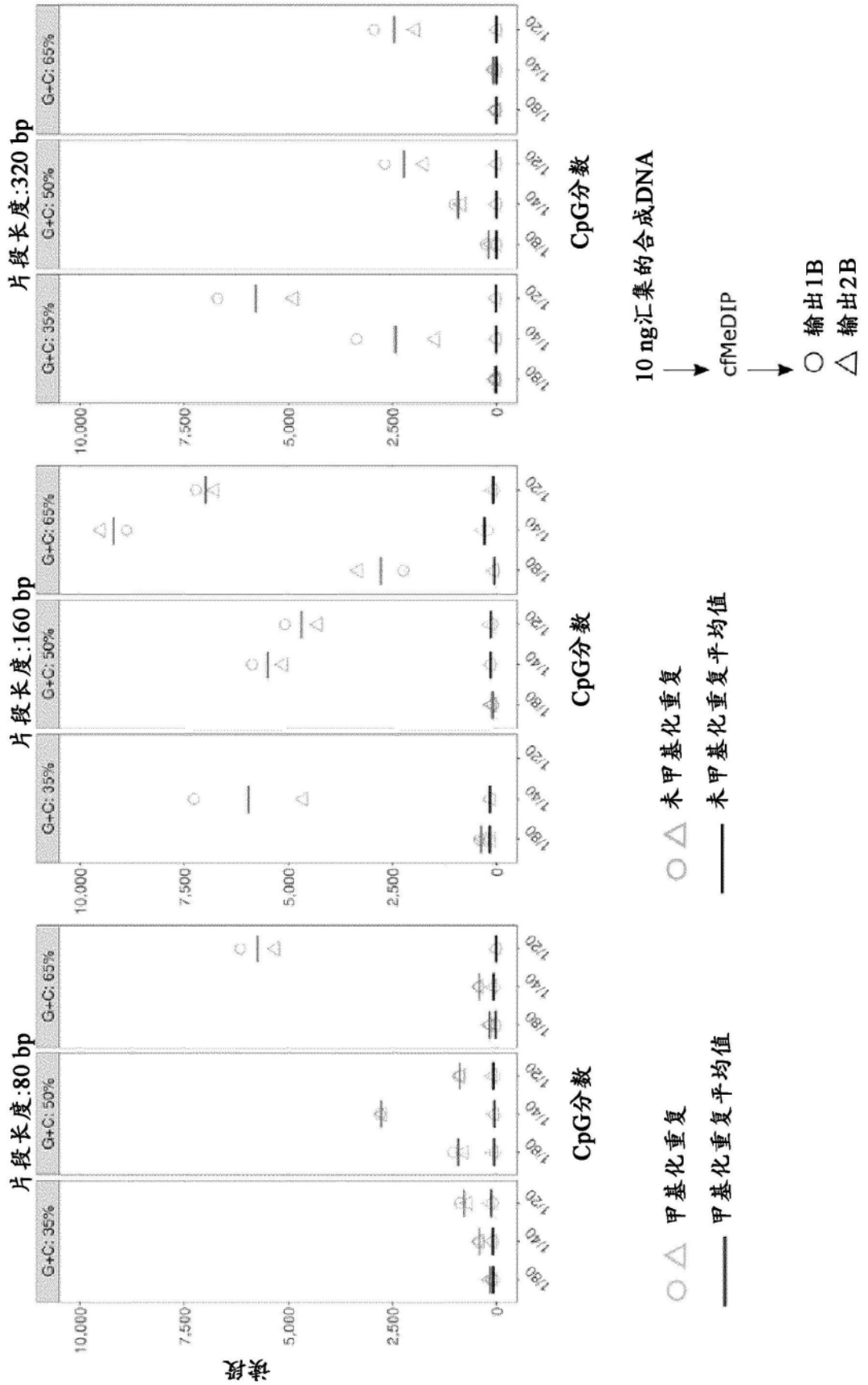

片段长度:80 bp
片段长度:160 bp
片段长度:320 bp
G+C: 35%
G+C: 50%
G+C: 65%
读段
CpG分数
10,000
7,500
5,000
2,500
0
1/80
1/40
1/20
10 ng汇集的合成DNA
cfMeDIP
输出1B
输出2B
甲基化重复
甲基化重复平均值
未甲基化重复
未甲基化重复平均值

图14续

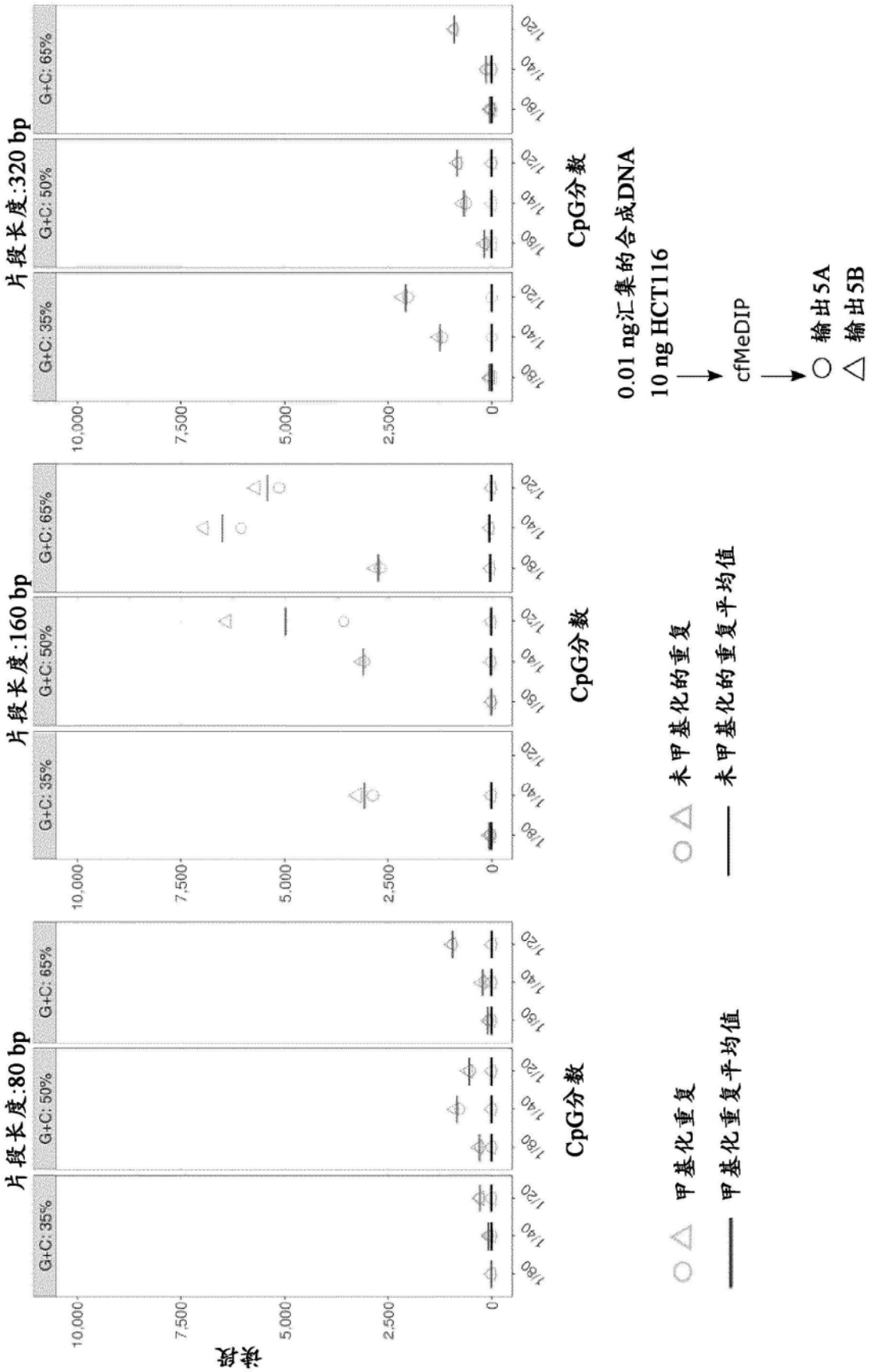

片段长度:80 bp
片段长度:160 bp
片段长度:320 bp
G+C: 35%
G+C: 50%
G+C: 65%
读段
CpG分数
10,000
7,500
5,000
2,500
0
1/80
1/40
1/20
甲基化重复
甲基化重复平均值
未甲基化的重复
未甲基化的重复平均值
0.01 ng汇集的合成DNA
10 ng HCT116
cfMeDIP
输出5A
输出5B

图14续

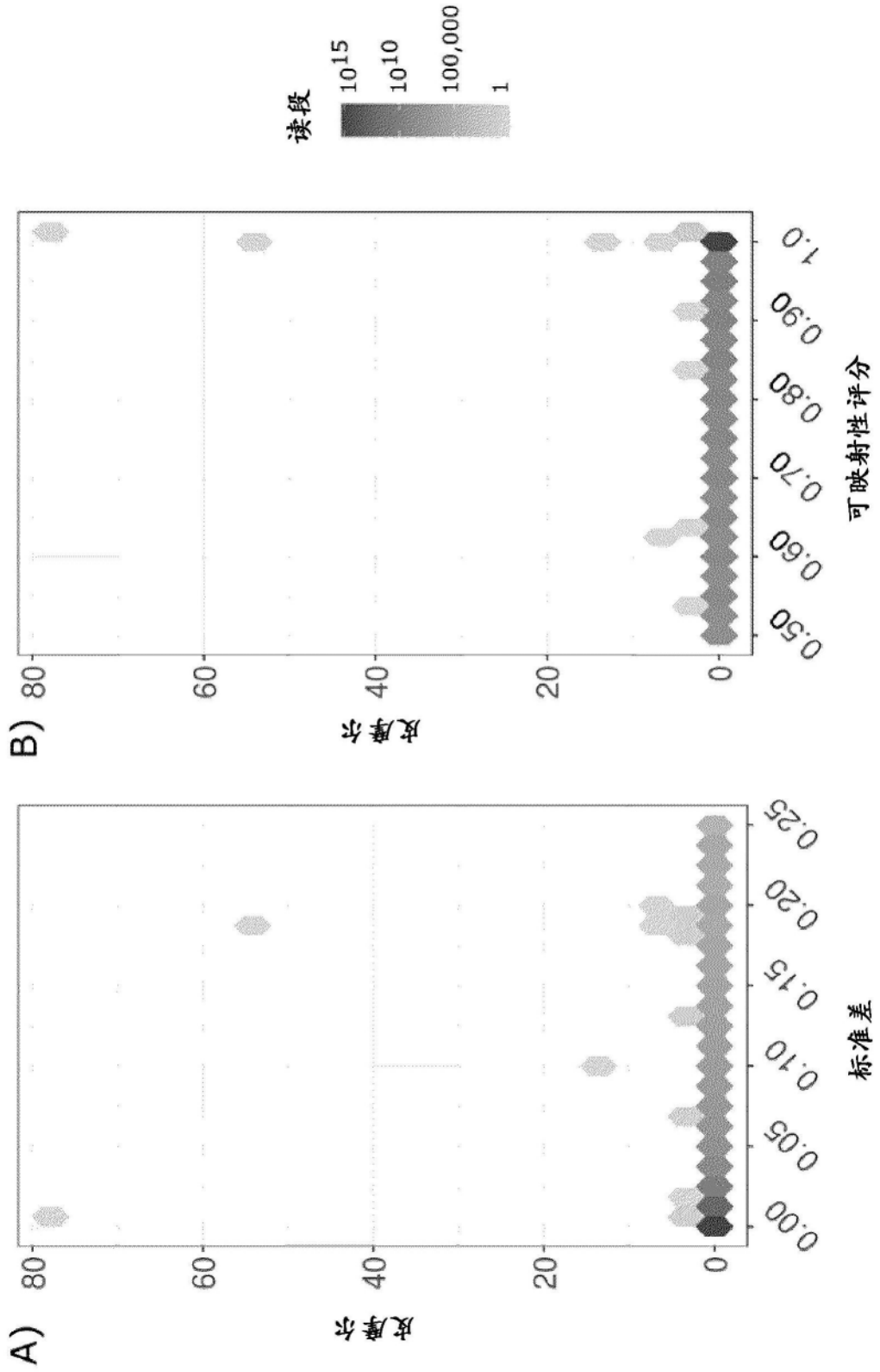

读段
10^{15}
10^{10}
100,000
1
A)
B)
皮摩尔
0
20
40
60
80
0.00
0.05
0.10
0.15
0.20
0.25
标准差
0.50
0.60
0.70
0.80
0.90
1.0
可映射性评分

图15

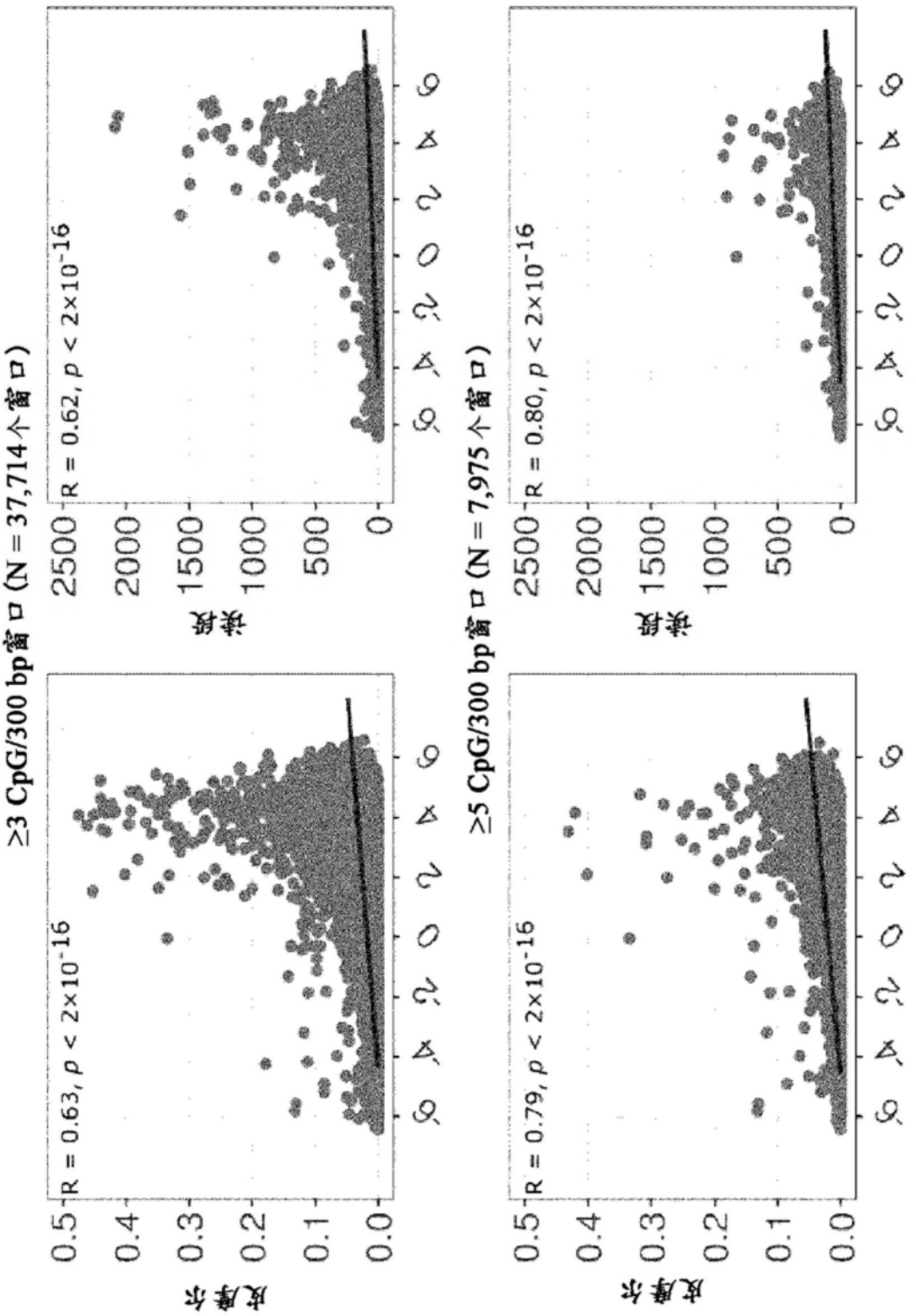

≥3 CpG/300 bp窗口 (N = 37,714个窗口)
≥5 CpG/300 bp窗口 (N = 7,975个窗口)
R = 0.62, p < 2×10-16
R = 0.80, p < 2×10-16
R = 0.63, p < 2×10-16
R = 0.79, p < 2×10-16
读段
皮摩尔

图16

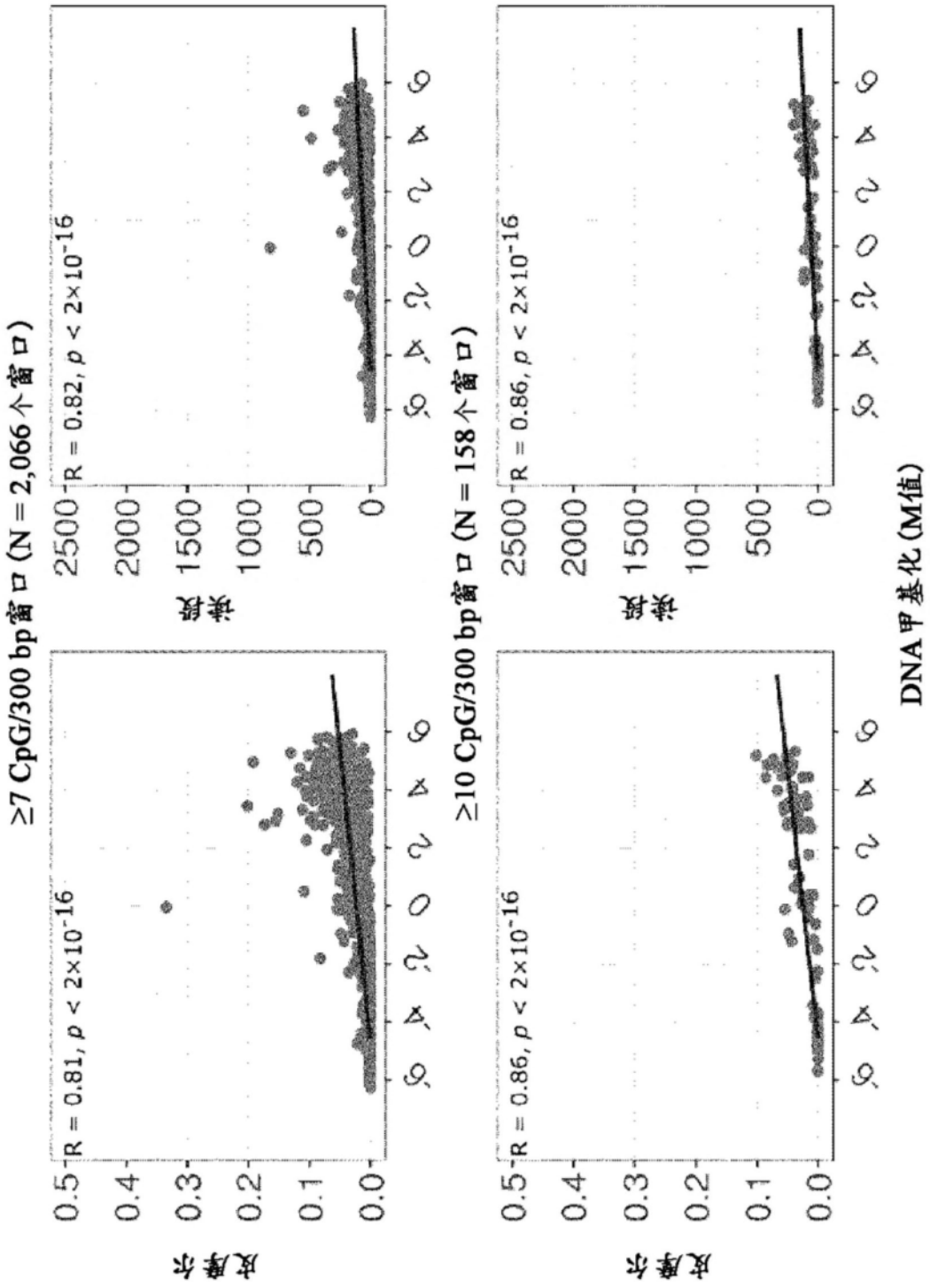

≥7 CpG/300 bp窗口 (N = 2,066个窗口)
≥10 CpG/300 bp窗口 (N = 158个窗口)
R = 0.82, p < 2×10-16
R = 0.81, p < 2×10-16
R = 0.86, p < 2×10-16
R = 0.86, p < 2×10-16
读段
皮摩尔
DNA甲基化 (M值)

图16续

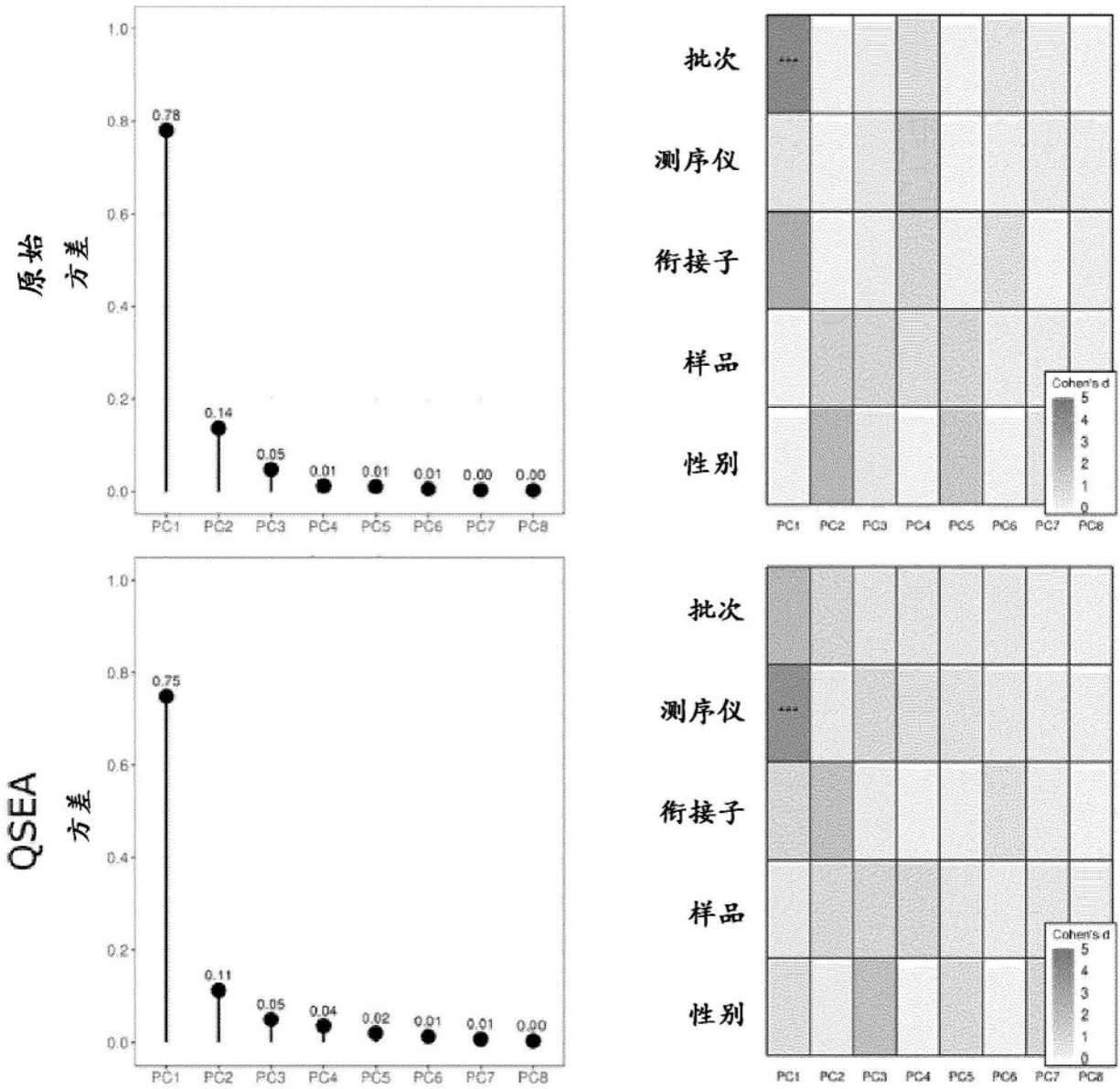

原始
方差
QSEA
方差
0.78
0.14
0.05
0.01
0.01
0.01
0.00
0.00
0.75
0.11
0.05
0.04
0.02
0.01
0.01
0.00
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
批次
测序仪
衔接子
样品
性别
Cohen's d
5
4
3
2
1
0

图17

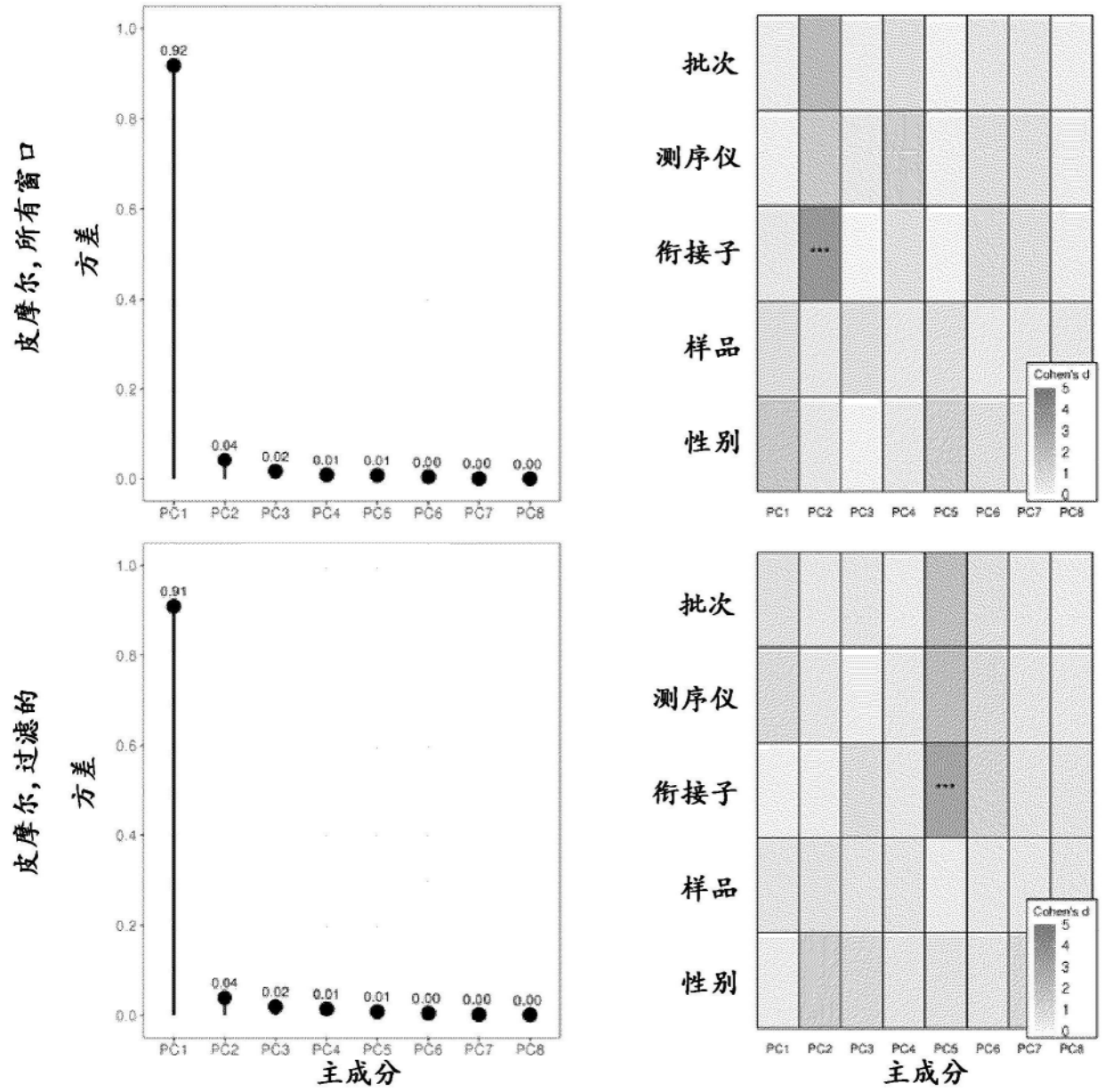

皮摩尔,所有窗口
方差
皮摩尔,过滤的
方差
主成分
批次
测序仪
衍接子
样品
性别
Cohen's d
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
0.92
0.04
0.02
0.01
0.01
0.00
0.00
0.00
0.91
0.04
0.02
0.01
0.01
0.00
0.00
0.00
***

图17续