

表观遗传学新标记—5-羟甲基胞嘧啶检测方法的研究进展

方科, 张凯翔, 王建, 付志猛, 赵湘辉

第四军医大学基础医学部神经生物学教研室暨脑科学协同创新中心, 西安 710032

摘要: 被称为“第六种碱基”的 5-羟甲基胞嘧啶(5-hydroxymethylcytosine, 5hmC), 广泛分布于多种哺乳动物的组织和细胞中, 与胚胎发育、神经系统功能以及肿瘤研究高度相关。与 5-甲基胞嘧啶(5-methylcytosine, 5mC)相比, 5hmC 在组织中含量更低, 难以精确的检测。随着研究的深入, 5hmC 参与的重要生物学作用逐渐被人们发现, 同时也促使着 5hmC 的检测和定量方法不断发展。为了区分 5hmC 与其他胞嘧啶衍生物, 很多利用化学或者酶学修饰实现靶向检测或非靶向富集 5hmC 的方法应运而生。因此, 选择并发展灵敏、准确、可靠的 5hmC 检测技术对于表观遗传研究至关重要。本文重点综述了近年来发展起来的 5hmC 检测和测序技术, 通过比较分析各种方法的优缺点, 为研究人员选择特定合适的方法开展相关研究提供重要的参考。

关键词: 表观遗传学; 5-羟甲基胞嘧啶; 检测方法; 高通量测序

Advances on the profiling of 5-hydroxymethylcytosine

Ke Fang, Kaixiang Zhang, Jian Wang, Zhimeng Fu, Xianghui Zhao

Department of Neurobiology and Collaborative Innovation Center for Brain Science, School of Basic Medicine, Fourth Military Medical University, Xi'an 710032, China

Abstract: 5-hydroxymethylcytosine (5hmC) is a naturally existing component in mammalian genomic DNA and is regarded as the sixth DNA base. Accumulating studies have revealed the essential role of 5hmC in embryonic development, brain function and cancer research. Compared to another well-known cytosine methylation derivative, 5-methylcytosine (5mC), the detection of 5hmC is difficult for its lower level existing in most tissues. To distinguish 5hmC from other cytosine derivatives, the methods using chemical or enzymatic DNA treatment, have been applied in targeted 5hmC detection or non-targeted 5hmC enrichment. Therefore, profiling DNA hydroxymethylcytosine by sensitive, accurate and reliable method is crucial for epigenetic study. This review discusses the principles behind recently developed techniques for 5hmC quantification and mapping. By comparing the advantages and shortcomings of these methods, the general guidelines were provided on how to select appropriate methods for specific experimental contexts.

收稿日期: 2015-11-26; 修回日期: 2015-12-25

基金项目: 国家自然科学基金(编号: 31371077, 31571050)和陕西省科学技术研究发展计划项目(编号: 2014KJXX-57)资助[Supported by the National Natural Science Foundation of China (Nos. 31371077, 31571050) and Shaanxi Science and Technology Research and Development Program (No. 2014KJXX-57)]

作者简介: 方科, 本科, 临床医学专业。E-mail: fxzbzhyh@foxmail.com

通讯作者: 赵湘辉, 博士, 副教授, 研究方向: 神经发育的表观遗传调控。E-mail: xianghuizhao@fmmu.edu.cn

DOI: 10.16288/j.ycz.15-480

网络出版时间: 2016-3-2 15:05:09

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20160302.1505.006.html>

Keywords: epigenetic; hydroxymethylcytosine; profiling; high-throughput sequencing

5-羟甲基胞嘧啶(5-hydroxymethylcytosine, 5hmC)最早于1952年在噬菌体DNA中被发现。它被糖基转移酶介导发生糖基化修饰,从而使噬菌体DNA在进入宿主后能抵抗宿主限制酶的降解^[1]。但当时5hmC的功能并没有引起更多的重视。直到2009年,Heintz和Rao研究组分别在同期Science杂志上报道了在人和小鼠的脑、胚胎干细胞DNA中含有丰富的5hmC^[2,3],这才让更多的研究人员重新着手认识这个修饰碱基。但与经典的5-甲基胞嘧啶(5-methylcytosine, 5mC)相比,5hmC的作用与功能并不完全清楚。作为重要的表观遗传标记,5hmC可能与DNA去甲基化和转录调控相关;同时人们发现5hmC在基因组中的分布不仅表现出不同于5mC的普遍性和动态特征,并且参与了重要的细胞功能调控,涉及干细胞的多能性、细胞发育分化和癌症的发生与发展等方面^[4-6]。此外,5hmC的分布和含量已经开始被作为许多癌症疾病早期诊断和预后的重要标志之一^[7,8]。因此,5hmC的研究在生物医学领域具有极其重要的意义。

随后,Rao和Zhang研究组发现,5hmC可以被TET家族氧合酶(Ten-eleven-translocation protein,包括TET1、TET2和TET3)转化5mC产生^[3,9]。TET蛋白是Fe²⁺和 α -酮戊二酸依赖的双加氧酶,通过C端的CD结构域(Cys-rich and DSBH regions)发挥功能。除TET2分子外,其他两个家族分子在N端还包含一个可与染色质相互作用的CXXC锌指结合域^[4]。更有研究表明TET蛋白可以将5hmC氧化为5-甲酰胞嘧啶(5-formyldeoxycytidine, 5fC)和5-羧基胞嘧啶(5-carboxylcytosine, 5caC)^[10]。两种新的碱基修饰方式的发现,也进一步拓宽了表观遗传学的研究领域和范畴。

若要了解5hmC在生理过程中的作用,掌握其在基因组DNA中的定量和定位信息是非常必要的。但这项工作最初却面临着巨大的技术挑战:一方面是因为当时标准的生物化学方法不容易区分5mC和5hmC^[11,12];另一方面是由于哺乳动物基因组中的胞嘧啶衍生物在胞嘧啶中所占的比例很低,因此精确检测其含量十分困难。近几年来,相关分析与测序技术快速发展,大大促进了人们对5hmC的探索研究和其生物学功能的认识。

在最初发现5hmC时,大量的研究利用传统的色谱、质谱以及特异性抗体识别的方法对不同来源组织和细胞的基因组DNA中5hmC和5mC的水平进行了比较分析;后来出现了各种针对5hmC的化学修饰、酶学修饰方法,大大提高了检测5hmC的灵敏度和特异性。随着高通量测序技术的逐渐普及应用,结合新的5hmC富集、修饰方法的出现,研究人员实现了在单碱基水平检测5hmC在基因组定位的目标;此外,第三代测序技术不需要经过DNA的扩增与修饰,也已经被应用于检测各种胞嘧啶衍生物的基因组定位研究。这些方法各有所长(表1),在实际应用中可以相互补充。本文主要针对5hmC检测技术取得的研究进展进行综述,以期为开展相关研究的工作人员提供重要的参考。

1 基因组中5hmC整体水平的初步检测

1.1 基于特异性抗体检测DNA整体的羟甲基化程度

对于研究5hmC在特定组织中的分布,可以利用商品化5hmC特异性的抗体,结合免疫荧光化学技术(Immunofluorescence, IF)显示组织中5hmC的分布^[13];或通过点杂交,半定量显示组织或细胞DNA中5hmC的总体水平^[13]。Leonice等应用上述两种方法对系统性肥大细胞增生病人的骨髓细胞中5hmC的水平进行了研究^[13]。目前已有Active Motif、Diagenode、Epigentek、Abcam、GeneTex等公司都提供了可用于免疫组化、点杂交和亲和沉淀的5hmC抗体。Epigentek和Enzo公司还分别利用抗体研发了类似于ELISA检测的方法,可以实现微量样品中5hmC水平的快速、高通量检测。

另一种方法是利用重亚硫酸盐将5hmC转化成5-磺基胞嘧啶(Cytosine-5-methylsulphonate, CMS),再利用抗CMS的抗体间接检测5hmC。这样不仅可以提高检测的精度,而且不受5hmC分布丰度的制约^[13,14]。

1.2 基于薄层色谱(Thin-layer chromatography, TLC)、高压液相色谱法(High pressure liquid chromatography, HPLC)和质谱(Mass spectrometry, MS)的定性和初步定量方法

在早期的研究中,研究者通过色谱或质谱等灵

表 1 5hmC 的定量和定位分析方法
Table 1 The methods on the quantification and mapping of 5hmC

方法			适用目的	优点	缺点	参考文献
基 因 组 中 5hmC 水平 的初步检测	抗体检测法	5hmC 抗体应用于 免疫荧光化学或点 杂交	以定量、 半定量为 目的	易于开展	很难准确定量,不能达到 单碱基分辨率	[13]
		重亚硫酸盐转化 后应用 CMS 抗体 检测				[13, 14]
	色谱、质谱	薄层色谱		可以检测含有 CpG 序列的 基因组 DNA 中 5hmC 总量	[2, 3,9]	
		高压液相色谱		灵敏度高、选择性好,可提供 DNA 各个组分的组成信息	不能达到单碱基分辨率 的水平	[15] [16~19]
		液相色谱-质谱连用				
	葡萄糖基化标记-液闪分析		不产生序列偏倚,并且所需 样本量小	不能确保完全的酶学反应 仅限于酶切位点	[20~21] [22]	
特定 DNA 序 列中的 5hmC 检测	葡萄糖基转移酶修饰-限制性内切酶消化					
非 靶 向 性 富 集 的 方 法 检 测 基 因 组 中 的 5hmC	抗体富集检测	5hmC 抗体富集 DNA (hMeDIP-Seq)	以了 解 5 hmC 在 基因序 列 中的分 布 为目的	可以检测 5hmC 丰度较高的 CpG 位置的羟甲基化 DNA	有一定的序列偏好,且不 能达到单碱基分辨率	[28, 29]
		重亚硫酸盐转化后富 集(CMS 抗体免疫共 沉淀 DNA 测序)		对 5hmC 丰度的依赖性更低	有抗体非特异结合的风 险,且不能达到单碱基分 辨率	[31]
	基于葡萄糖基修饰 的富集	hMe-Seal 化学标记法		方法特异性高	不能达到单碱基分辨率 的水平	[32, 33]
		JBP-1 沉淀法		简单直接	JBP-1 捕获 5hmC 的效率 较差,灵敏度不高	[35]
		GLIB 沉淀法	以了解 5 hmC 在 基因序 列 中的精 确 定位为目的	结合单分子测序技术,不需要 PCR 扩增,避免了 PCR 偏向	测序价格昂贵	[31,34]
	限制性内切酶依 赖的相关检测	RRHP		测序所需的样品量极少,并 且可以达到单碱基分辨率	只有 Msp I 识别区域的 5hmC 可以被识别、检验	[37]
		内切酶 AbaS I 结合高 通量测序(Aba-Seq)		可以敏感的检测低丰度位 点的 5hmC,并且可以在单 碱基水平确定 5hmC 的分布	内切酶的位置偏好	[38]
		Pvu-Seal-Seq		无位置偏好,灵敏度高,适 合在样本量少或者胞嘧啶 修饰丰度低的情况下使用	目前不清楚	[40]
全 基 因 组 水 平 测 序 检 测 5hmC 分布	基于亚硫酸氢钠 法测序	TET 协同的亚硫酸氢 钠法测序(TAB-Seq)		达到单碱基分辨率	亚硫酸氢盐处理会产生 严重的 DNA 降解	[41, 42]
		氧化重亚硫酸盐测 序(OxBS-Seq)			氧化反应的控制	[43, 44]
	第三代测序技术	单分子实时测序技术 (SMRT)		不需要扩增 DNA,降低了因 为扩增而对修饰位点的影响	碱基识别错误率较高,测 序价格昂贵	[45, 46]
		纳米孔测序		高的读长、快速测序	错误率还不清楚	[51, 52]

敏度有限的检测方法确定了 5hmC 在不同组织的含量存在差异性。这些方法可以同时检测 DNA 样本中 5mC 和 5hmC 的水平,但是均不能提供 DNA 序列中特定位点上单个碱基的信息。

将 DNA 样品酶解成单核苷酸后,其中的各种类

型碱基由于在 TLC 薄层固定相(如硅胶)上的迁移距离不同,经过层析缸内展开后可实现样品内不同组分的分离,结合对定量用的标准品同时分离,再比较测定各斑点的位置及其密度就可以实现对样品中 5hmC 的定性和定量分析^[2,3,9]。但该方法灵敏度有

限, 因此定量准确度不高。在此基础上, HPLC 将流动相改为高压输送, 色谱柱以特殊的方法用小粒径的填料填充, 从而使柱效大大高于经典液相色谱; 同时柱后连有高灵敏度的 UV 检测器, 可对流出的核酸类物质, 包括 5hmC, 实现连续检测^[15]。

研究人员还利用分离度、灵敏度和重复性更高的液相色谱(Liquid chromatography, LC)-质谱联用的方法检测了 5hmC 的含量。该方法利用样品中各个组分质荷比(m/z)不同, 在通过电场或是电磁场时将它们区分开来, 以检测 5hmC 的总量^[16-19]。例如, Globisch 等^[18]观察到 5hmC 在小鼠大脑 DNA 中所占比率最高, 达到总量的 0.3%~0.7%, 在脊髓中为 0.47%; 而在肾脏、鼻上皮、膀胱、心脏、骨骼肌和肺中, 为 0.15%~0.17%; 含量最低的组织为肝、脾、内分泌腺等器官, 仅为 0.03%~0.06%。目前 Zymo Research 公司提供两种 DNA Degradase™酶混合物, 可以将 DNA 降解为单核苷酸, 分别为后续 HPLC 和 LC/MS 测量准备样品。尽管该方法灵敏度高、选择性好, 可提供各个组分组成和结构信息, 但不能达到单碱基分辨率的水平。此外, 这 3 种方法因不适用于大样本实验检测, 在实际应用中也受到一定限制。

1.3 基于葡糖基化标记-液闪分析定量 5hmC

除了上述色谱-质谱分析的方法, 利用化学修饰也可以检测 DNA 样本中的 5hmC 水平。葡糖基化是使用 T4 噬菌体 β -葡糖基转移酶(β -glycosyltransferase, β -GT)将^{[3]H}标记的葡萄糖由尿核苷二磷酸葡萄糖(UDP-^{[3]H}-glucose)供体转移至 5hmC 位点上, 生成 β -葡糖基-5-羟甲基胞嘧啶(5ghmC), 进而通过放射性计数分析进行整体水平的 5hmC 的测定^[20]。Szwagierczak 等利用这种方法也检测了小鼠不同器官组织 DNA 中的 5hmC 比率, 且均略高于上述 HPLC-MS 方法观察到的结果^[20]。比如大脑中为 1%~1.3%, 肾脏中接近 0.8%。说明这种方法灵敏度更高, 所需样本量更小, 并且它不依赖于特定的设备, 不产生序列偏倚。最近, 汪海林研究组提出了一种基于糖基化的高通量的 5hmC 检测方法^[21]。研究人员发现, 硼酸可与 5ghmC 中的邻二醇发生共价作用, 并有效抑制聚合酶扩增反应。利用该特点, 在 PCR 反应中可特异性地识别特定基因/片段中的 5hmC。该方法不受酶切位点限制, 可适用于各种细胞、各种序列

中 5hmC 的分析鉴定, 具有广阔的应用前景。

2 靶向 DNA 序列中的 5hmC 检测

如果需要大致判断特定感兴趣基因组 DNA 序列中是否有 5hmC 修饰, 可以利用限制性内切酶消化, 并联合上述的葡糖基转移酶修饰, 然后利用针对靶向片段 DNA 设计引物进行 PCR 扩增检测实现^[22]。

同裂酶 *Msp*I 和 *Hpa*I 可以识别同样的序列(CCGG), 但它们对甲基化状态的敏感性不同。后者只切割完全未修饰的位点, 胞嘧啶上的任何修饰(5mC、5hmC 或 5ghmC)都阻碍切割; 而前者则识别并切割 5mC 和 5hmC, 5ghmC 除外。利用这一特点, 基因组 DNA 先经糖基化修饰将 5hmC 转化为 5ghmC, 使其避免被 *Msp*I 消化; 接着 *Msp*I 消化后若 PCR 扩增有特异性产物, 就说明该序列中含有 5hmC。同样, *Hpa*I 消化后, 若序列中存在 5mC 和 5ghmC, 将阻碍切割, 因此 PCR 扩增可以检测出 5mC 和 5hmC。未经酶消化直接进行 PCR 检测, 获得的是 DNA 总量的信息。如果使用实时荧光定量 PCR, 就可以大致估计 5hmC 修饰的多少。这一方法使得 5hmC 在低丰度区域的检测得以实现。目前, NEB 公司已经投放了基于上述方法专用于 5hmC 分析的 EpiMark 商业化试剂盒, Zymo Research 和 Enzo 公司也提供原理相似的 Quest 5hmC Detection Kit 和 BIOAR-RAY™ 5hmC Methylation Kit。虽然这一技术极大地方便了对于特定 DNA 序列进行 5hmC 定量研究的相关实验, 但值得注意的是该方法仅限于对包含酶切位点的 DNA 序列进行检测, 在应用上有一定的局限性, 而且也不能达到单碱基分辨率水平。

3 非靶向性富集的方法检测基因组中的 5hmC

此策略主要是通过各种 DNA 富集的方法, 如抗体、化学修饰、限制性内切酶等, 先将含有 5hmC 的 DNA 片段富集后再进行下游分析, 通常是高通量测序实现对基因组中 5hmC 分布的测定。这种策略比后面将要提到的全基因组测序的方法更加经济, 因此也被大多数实验室选用。

3.1 抗体富集检测法

利用高度特异性的商品化抗体, 结合 DNA 免疫共沉淀测序(DNA immunoprecipitation and sequencing,

DIP-Seq)的方法,可以完成对 5hmC、5fC 和 5caC 测量^[9,23~27]。羟甲基化 DNA 免疫共沉淀测序(5hmC DIP-Seq, 即 hMeDIP-Seq)是利用 5hmC 特异抗体,形成抗体/DNA 复合物,进而使得含有 5hmC 的 DNA 片段被富集,再通过高通量测序或芯片技术进行基因组定位^[28,29]。小鼠和人的胚胎干细胞中 5hmC 的基因组分布情况就是用这种方法获得的^[24,28]。前面提到的生产商品化 5hmC 抗体的公司几乎都推出了 hMeDIP 试剂盒。利用这种方法,可以检测 5hmC 丰度较高的 CpG 位置的羟甲基化 DNA,但仍然无法得到单碱基水平的信息,或者说无法得知 5hmC 的序列信息^[30]。

经亚硫酸氢盐处理可以将 5hmC 转化成 CMS,这种形式的碱基可以使用 CMS 特异性抗体,通过免疫沉淀的方法加以富集,进而检测 5hmC 的分布^[31]。因为 CMS 带有负电荷,因此 CMS 抗体富集的方法相较于 5hmC 抗体富集的方法更加灵敏,在测序时对 5hmC 含量的要求较低,同时对 5hmC 丰度的依赖性更低。

3.2 基于葡糖基修饰的富集方法

3.2.1 hMe-Seal 化学标记法

该方法在葡糖基化法基础上发展起来,将酶反应底物上的葡萄糖进行化学修饰转变成 UDP-6-叠氮葡萄糖(UDP-6-N₃-glucose),从而取代了最初的放射性标记。在 β -GT 作用下将 6-N₃-glucose 转移到羟甲基位置,生成 N₃-5ghmC。随后,通过点击化学方法在叠氮基团标记的 5hmC 上添加一分子生物素。然后利用磁珠捕获包含 5hmC 的 DNA 片段,结合高通量 DNA 测序技术或单分子测序技术,可分析 5hmC 在基因组 DNA 中的分布情况^[32,33]。这种方法减少了测序所需的样本量,且 5hmC 上的生物素标签使其在测序中显示出更高的动力学信号,提高了核苷酸识别的准确性。另外,采用链酶亲和素-HRP 反应还可以测量样品整体 5hmC 的含量^[32]。何川研究组利用这种方法对发育阶段小鼠小脑基因组 DNA 中 5hmC 的分布进行了研究,还鉴定了一些与年龄相关的神经退行性疾病基因位点中 5hmC 修饰的改变情况^[32]。

3.2.2 GLIB 沉淀法

GLIB 是葡糖基化(Glucosylation)、高碘酸氧化(Periodateoxidation)和生物素化(Biotinylation)三种

修饰反应结合方法的缩写^[31,34]。该方法首先利用 β -GT 在每个 5hmC 上添加一个葡萄糖分子,然后用高碘酸钠氧化葡萄糖将邻近的羟基转化成醛基,之后用醛基反应进一步修饰,在每个 5hmC 上添加两个探针标记的生物素分子,并可利用链酶亲和素来 pull-down。Pastor 等^[31,34]将 GLIB 处理的 DNA 进行单分子 DNA 测序(Helicos),完成了小鼠胚胎干细胞中 5hmC 的全基因组定位检测。这种方法不需要 PCR 扩增,从而避免了 PCR 偏向,但是测序价格昂贵。

3.2.3 JBP-1 沉淀法

Robertson 等^[35]发现锥虫(*Trypanosoma*)的 J 结合蛋白 1(J-binding protein 1, JBP1)与糖基化 hmC 紧密结合,可以使用磁珠将含有 hmC 的 DNA 加以分离。将 JBP1 与磁珠偶联,就可以直接识别并结合葡糖基化的产物 5ghmC。含有 5hmC 的 DNA 片段被富集后,可应用 qPCR、LC/MS、芯片、测序等方法进一步分析。尽管 JBP1 方法简单直接,但其捕获 5hmC 的效率,或者说灵敏度比上述的 hMe-Seal、GLIB 等方法较差。Zymo 公司通过制作重组 JBP1 蛋白,提高了其捕获 5hmC 的效率,并发展了 JBP1-Seq 方法^[36]。

3.3 限制性内切酶依赖的相关检测

3.3.1 RRHP

近来,Zymo Research 公司建立了 RRHP(Reduced representation 5hmC profiling, RRHP)方法^[37],检测了几种肿瘤组织 DNA 中 5hmC 的定位。该方法将酶切反应与高通量测序相结合,可以实现在单核苷酸水平对 5hmC 的检测。在该方法中,先利用限制性内切酶 *Msp*I 消化切割基因组 DNA 中 GC 富集的区域,再在消化末端添加测序接头。经 β -GT 处理后,接头部位的 5hmC 被保护起来并不再受 *Msp*I 再次消化的作用。这样只有两末端都有接头的 DNA 片段才能被扩增、测序。5hmC 的位置,也就是接头连接的区域,可以通过测序得知。在这种方法中,虽然只有 *Msp*I 识别区域的 5hmC 可以被识别、检验,但该方法可以区别 5mC 和 5hmC,并将测序所需的样品量降低至 100ng 左右,比前面提到的其他富集 DNA 方法测序所需的 DNA 量少了很多。

3.3.2 Aba-Seq

这是一种在葡糖基修饰基础上,DNA 修饰依赖

的限制性核酸内切酶 *AbaS* 结合高通量测序的方法(DNA-modification-dependent restriction endonuclease *AbaS* I coupled with Sequencing, *Aba*-Seq)^[38], 可以高分辨率检测基因组中 5hmC 的分布。在经过 β -GT 处理后, *AbaS* 可以识别并且在 5ghmC 的 3' 末端的固定距离切割 dsDNA(大约为 11~13 或者 9~11 个核苷酸)。处理后的 DNA 片段可以被建成测序文库进行深度测序。5hmC 位置的测定是由断裂的末端决定的。*Aba*-Seq 可以敏感的检测低丰度位点的 5hmC, 并且可以在单碱基水平确定 5hmC 的分布。但是 *AbaS* 酶的切割活性与切割位点两边的 DNA 链上存在胞嘧啶高度相关, 如果只有一端有胞嘧啶将大大降低该酶的活性^[39]。

3.3.3 Pvu-Seal-Seq

与 *AbaS* 同属一个内切酶家族的 *PvuRts1* 酶, 不具有上述的位置偏好, 在理论上能够识别全基因组中的 5hmC。但是该酶的特异性不如 *AbaS*, 还可以识别 5mC 和 C, 因此 NEB 公司的研究人员将 5hmC 化学标记富集法与限制性内切酶结合起来, 先用 *PvuRts1* 酶切割 DNA, 再用 Seal 的方法在叠氮基团标记的 5hmC 上添加一分子生物素, 利用磁珠富集包含 5hmC 的 DNA 片段, 结合高通量测序可以检测到低丰度的 5hmC 位点^[40]。这项研究展示了一种无需重亚硫酸盐转化的灵敏方法, 适合在样本量少或者胞嘧啶修饰丰度低的情况下使用。

4 全基因组水平测序检测 5hmC 分布

这种检测 5hmC 的策略一方面主要是基于对全基因组的亚硫酸氢钠法测序(Bisulphite sequencing, BS-Seq)完成的。不同的化学修饰手段在测序前将 5hmC 与 5mC 区别开, 通过比对同一样品在正常 BS-Seq 和化学修饰后 BS-Seq 的序列, 确定 5hmC 的基因定位。相关原理见图 1。另一方面, 第三代测序技术也可以实现全基因组水平的 5hmC 检测。

4.1 TET 协同的亚硫酸氢钠法测序(TET-assisted bisulphate sequencing, TAB-Seq)

何川课题组发现, 和胞嘧啶一样, 5-羧基胞嘧啶(5-Carboxylcytosine, 5caC)在 BS-Seq 中会显示为

胸腺嘧啶 T^[41]。因此, 通过利用 β -GT 将糖环转移到 5hmC 上, 转变为 5ghmC; 然后利用过量的 TET 氧化酶将 5mC、5fC 氧化为 5caC。在经过糖基转移酶和 TET 氧化酶处理之后, 在最终的亚硫酸氢钠法测序结果中, 5mC 处被显示为胸腺嘧啶 T, 而 5hmC 被显示为胞嘧啶 C。这样, 通过与未经处理样品的 BS-Seq 结果做对比, 即可识别出 5mC 和 5hmC 的分布位点。这种方法被称为 TAB-Seq^[41, 42], 该方法有效解决了传统亚硫酸氢盐测序无法区分 5mC 和 5hmC 的问题, 为检测样品中 5hmC 的分布提供了一种非常可靠的手段。人类和小鼠胚胎干细胞基因组中 5hmC 的分布和水平就是通过该方法实现测量的^[42]。

4.2 氧化重亚硫酸盐测序(Oxidative bisulphate sequencing, OxBS-Seq)

这种方法利用 5hmC 的氧化反应, 实现在单个核苷酸水平检测 5hmC^[43]。5hmC 在高钌酸钾 KRuO_4 的氧化作用下转变成 5fC, 在基于重亚硫酸氢盐的测序方法中, 该位点作为胸腺嘧啶 T 被识别, 而未被处理的 5mC 则被识别为胞嘧啶 C。最终 5hmC 的确定由同一样品的 BS-Seq 测序中的 5mC 和 5hmC 信号与 OxBS-Seq 测序中的 5mC 信号的差值决定。当利用差值法计算两随机样品的修饰位点的百分数时, 必须进一步计算每一测量的可信区间; Booth 等^[44]发展了一种优化的 OxBS-Seq 方法, 该方法可以减少氧化物诱导的 DNA 降解。

4.3 利用第三代测序技术检测 5hmC

第三代测序技术的发展, 使得直接检测 DNA 修饰位点而不经化学和酶的修饰成为可能。Pacific Bioscience 公司发展了单分子实时(Single-molecule real-time, SMRT)测序技术, 使得通过记录核苷酸掺入时间或动力学特征而可以直接测量 5mC 及其氧化衍生物^[45, 46]。该系统利用单个 DNA 聚合酶进行 DNA 合成, 在合成过程中, 系统会监控不同的核苷酸掺入, 而通过掺入核苷酸的记录即可反映出核苷酸的序列。而且该系统不需要扩增 DNA, 降低了因为扩增而对修饰位点的影响。SMRT 测序的缺点在于不同衍生物之间的动力学特征差异太小, 以致影响检测的准确性, 碱基识别错误率较高(18%~20%)^[47]。

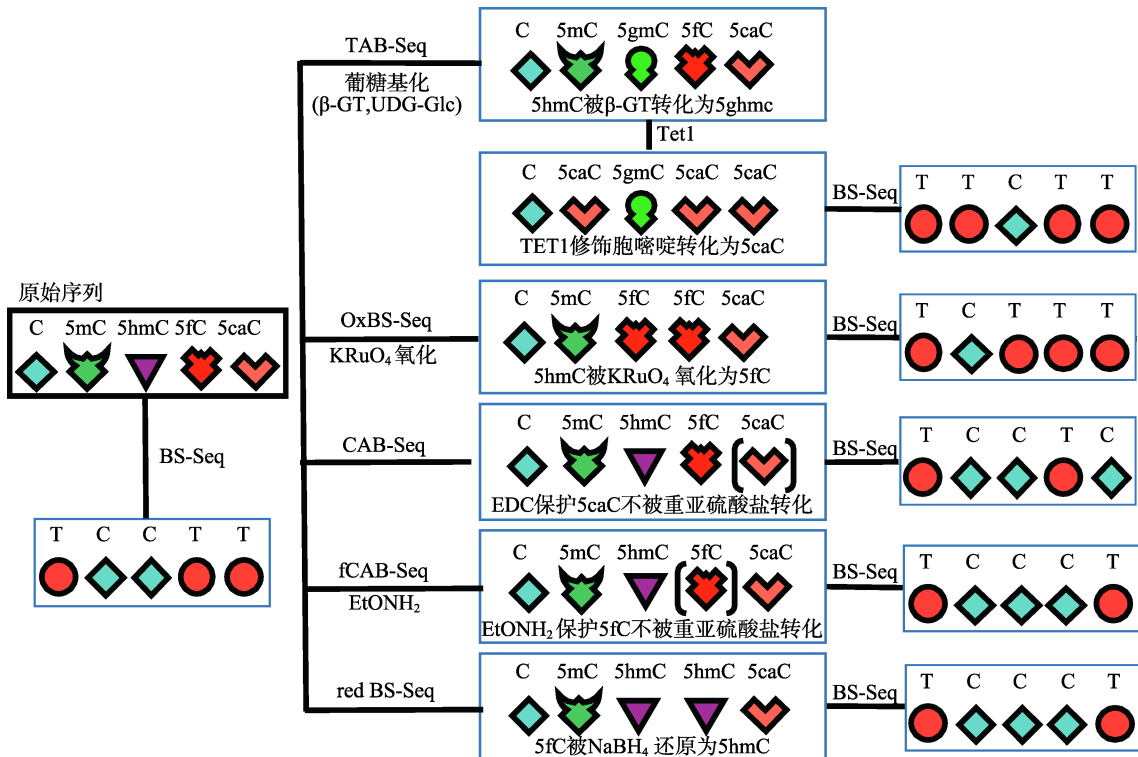


图 1 基于重亚硫酸盐测序检测 5mC 衍生物的方法示意图

Fig. 1 Schematic diagram of mapping 5-methylcytosine (5mC) oxidation derivatives at single-base resolution based on bisulphite sequencing

单核苷酸水平的 5-羟甲基胞嘧啶(5hmC)的检测方法有两种, TAB-Seq 和 oxBS-Seq。在 TAB-Seq 中, DNA 中的 5hmC 被 β -糖基转移酶(β -GT) 转化为 5ghmC (5-glucosylmethylcytosine) 后, 再经过 TET1 蛋白处理后将所有的修饰胞嘧啶转化为 5-羧基胞嘧啶(5caC)。经过重亚硫酸盐测序 BS-Seq 后, 5ghmC(原来的 5hmC) 被识别为胞嘧啶。在 OxBS-Seq 中, 5hmC 被 KRuO_4 氧化成为 5fC; 在重亚硫酸盐处理后, 被转化为胸腺嘧啶 T。5hmC 的含量测定由 BS-Seq 中胞嘧啶的含量减去 OxBS-Seq 中胞嘧啶的含量所得。在检测 5caC 的方法中(CAB-Seq), 利用 EDC 以保护 5caC 在随后的 BS-Seq 反应中不被脱氨基。5caC 的含量由 CAB-Seq 中胞嘧啶的含量减去 BS-Seq 中胞嘧啶确定。检测 5fC 的有两种方法, fCAB-Seq 和 red BS-Seq。fCAB-Seq, 利用 EtONH_2 保护 5fC 不被重亚硫酸盐转化, 而 5fC 的含量由 fCAB-Seq 中胞嘧啶的含量减去传统的 BS-Seq 中胞嘧啶的量。在 red BS-Seq 中, 5fC 被 NaBH_4 还原为 5hmC, 在亚硫酸盐处理后显示为胞嘧啶, 5fC 的含量由 red BS-Seq 中的胞嘧啶减去 BS-Seq 中的胞嘧啶表示。

然而, 在某些研究中, 化学标记和限制酶修饰可以增强信号的检测。比如在测序前, 使用前面提到的方法将 5hmC 转化为 N_3 -5ghmC 或 HS-N_3 -5ghmC 可以提高与 5mC 的动力学差异, 从而提高定量检测的准确性^[48]。还有研究发现, 5caC 的动力学信号相比 5mC 更加强烈; 将 5mC 氧化成为 5caC 可以增强细菌基因组中 5mC 的检测信号^[49]。另外, 该方法测序价格昂贵, 目前很难被大多数实验室接受。

纳米孔是另一种单核苷酸 DNA 测序技术^[50], 采用电泳技术, 借助电泳驱动单个分子逐一通过纳米孔来实现测序。由于纳米孔的直径(1~2 nm)非常细小, 允许单个核苷酸聚合物通过, 保证了检测具有良好的持续性。固态和蛋白纳米孔均已被用于检测 5hmC^[51,52]。

与 SMRT 相似的是, 纳米孔道技术具有高的读长、快速测序以及无需化学修饰的优点。然而纳米孔测序的错误率还不清楚。因此第三代测序技术在广泛使用之前, 还有很多的改进工作需要完成。

5 5fC 和 5caC 的检测方法

除了 5hmC, 哺乳动物基因组中的 5fC、5caC 在单核苷酸水平的检测方法近来也已经实现。这两个核苷酸经过重亚硫酸氢盐处理脱氨基后在测序中被检测为胸腺嘧啶 T。为了将 5caC 与其他修饰位点相区别, 发展了化学修饰协同的亚硫酸氢钠法测序 (Chemical modification assisted bisulphite sequencing, CAB-Seq)^[53]。反应利用化合物 1-乙基-3-(3-二甲基氨

基丙基)碳酰二亚胺盐酸盐(EDC)可以选择性的标记 5caC 中的羧基这一特点, 催化 5caC 的羧基和伯胺基之间形成酰胺键, 以保护 5caC 在随后的 BS-Seq 反应中不被脱氨基。被保护的 5caC 位点在 BS-Seq 中不转化为 T。最终, 通过比较 BS-Seq 测序的信号与 CAB-Seq 的信号水平的差值就可以判定 5caC 的位置分布。最近两项研究报道了在单核苷酸水平检测 5fC 的方法, 即 fCAB-Seq (5fC quantification using chemical modification-assisted bisulphate sequencing)方法和还原法 BS-Seq (Reduced bisulphite sequencing, red BS-Seq)方法^[54,55]。在 fCAB-Seq 中, 5fC 经过 EtONH₂ 修饰后, 再利用 BS-Seq 进行测序。被修饰的 5fC 被保护而不与重亚硫酸氢盐反应, 而 5caC 和胞嘧啶则被胸腺嘧啶 T 所取代。5fC 的确定是由同一样品的 BS-Seq 的标准结果与 fCAB-Seq 的差值决定的。在 red BS-Seq 中, NaBH₄ 可将 5fC 还原诱导成为 5hmC, 在接下来的 BS-Seq 中被识别为 C 而不再是 T。

与传统的 BS-Seq 测定全基因组水平 5mC 的方法相比, 测定这些超低含量胞嘧啶衍生物的工作量远远高出若干倍, 而且对样本量的要求在微克以上。对于 CpG 区的检测, 可以采用我们上面提过的利用限制性内切酶 *Msp* I 切割后, 添加测序接头, 建立测序文库后再做 BS-Seq, 这样可以减少测序量^[43,54]。尽管如此, 目前只有少数实验室可以开展上述方法。由于测序数据有限, 也还不知道酶切修饰和化学修饰对测序准确性的影响。因此这些实验必须有高质量的对照做质控。

尽管依赖亚硫酸氢盐处理的测序技术能够实现单碱基分辨率的 5fC 检测, 但亚硫酸氢盐处理会产生严重的 DNA 降解、且测序成本昂贵。前面提及的 Pvu-Seal-Seq, 在酶切反应前, 在 β -GT 作用下先将 5hmC 转化成 5ghmC, 以防后续反应将 5hmC 富集; 再利用 NaBH₄ 可将 5fC 还原诱导成为 5hmC 的特点, 采用 Pvu-Seal-Seq 可以间接检测 5fC^[40]。另外在 2015 年 9 月, 北京大学生命科学学院伊成器研究组在 *Nature Methods* 杂志上报道了另一种通过化学标记和富集手段, 并且免亚硫酸氢盐处理(Bisulfite-free)的 5fC 单碱基分辨率、全基因组水平测序方法(Cyclization-enabled C-to-T transition of 5fC, fC-CET)。这一技术经过选择性化学标记后, 可以在后续的 PCR 中将 5fC 转化为 T, 而且不对基因组 DNA 造成明显

降解, 较适用于小量、珍贵核酸样本的分析与测序^[56]。

6 结 语

随着表观遗传学研究的不断深入, 人们对于 5hmC 的生物学功能认识更加完善^[5,57], 而这些进展在很大程度上得益于不同层次的针对 5hmC 的检测和测序技术的发展。在实际的科研工作中, 应该结合研究的需要和方法的可行性, 首先选择定性或半定量的检测技术, 如利用特异性抗体完成免疫荧光, 或者采用薄层色谱大致了解特定组织中 5hmC 分布和含量情况。还可以采用抗体、限制性内切酶及其他化学修饰的方法富集含有 5hmC 的 DNA 片段, 再进行各种下游分析, 比如 PCR 检测、芯片、测序等等。当需要详尽了解特定细胞或组织中 5hmC 的基因组分布时, 就可以选择使用几种可以实现单核苷酸分辨率的测序方法了, 如 TAB-Seq、OxBS-Seq、Aba-Seq 和 Pvu-Seal-Seq 和 RRHP。而在 5hmC 的研究中, 各种新的化学修饰方法不断出现, 包括第三代测序技术, 如 SMRT, 也应用于检测 5hmC 的分布和水平。这些方法在很大程度上提高了 5hmC 检测的效率。与此同时, 另外两种胞嘧啶碱基修饰—5caC 和 5fC, 在如火如荼的表观遗传学研究中逐渐引人注目。相信通过研究人员的不断应用发展, 各种技术方法逐渐成熟, 将会吸引更多的科学家进入这一重要的研究领域。

参考文献(References):

- [1] Wyatt GR, Cohen SS. A new pyrimidine base from bacteriophage nucleic acids. *Nature*, 1952, 170(4338): 1072–1073. [DOI]
- [2] Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 2009, 324(5929): 929–930. [DOI]
- [3] Tahiliani M, Koh KP, Shen YH, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 2009, 324(5929): 930–935. [DOI]
- [4] Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev*, 2011, 25(23): 2436–2452. [DOI]
- [5] Tan L, Shi YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*,

- 2012, 139(11): 1895–1902. [DOI]
- [6] Zhang YX, Gao KR, Yu SY. Progress of research on 5-hydroxymethylcytosine. *Hereditas (Beijing)*, 2012, 34(5): 509–518.
- 张燕霞, 高可润, 禹顺英. 5-羟甲基胞嘧啶的研究进展. *遗传*, 2012, 34(5): 509–518. [DOI]
- [7] Jin SG, Jiang Y, Qiu RX, Rauch TA, Wang YS, Schackert G, Krex D, Lu Q, Pfeifer GP. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with *IDH1* mutations. *Cancer Res*, 2011, 71(24): 7360–7365. [DOI]
- [8] Lian CG, Xu YF, Ceol C, Wu FZ, Larson A, Dresser K, Xu WQ, Tan L, Hu YG, Zhan Q, Lee CW, Hu D, Lian BQ, Kleffell S, Yang YJ, Neiswender J, Khorasani AJ, Fang R, Lezcano C, Duncan LM, Scolyer RA, Thompson JF, Kakavand H, Houvras Y, Zon LI, Mihm MC Jr, Kaiser UB, Schatton T, Woda BA, Murphy GF, Shi YG. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell*, 2012, 150(6): 1135–1146. [DOI]
- [9] Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 2010, 466(7310): 1129–1133. [DOI]
- [10] Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 2011, 333(6047): 1300–1303. [DOI]
- [11] Nestor C, Ruzov A, Meehan RR, Dunican DS. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *Biotechniques*, 2010, 48(4): 317–319. [DOI]
- [12] Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res*, 2010, 38(11): e125. [DOI]
- [13] Leoni C, Montagner S, Deho' L, D'Antuono R, De Matteis G, Marzano AV, Merante S, Orlandi EM, Zanotti R, Monticelli S. Reduced DNA methylation and hydroxymethylation in patients with systemic mastocytosis. *Eur J Haematol*, 2015, 95(6): 566–575. [DOI]
- [14] Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, Liu XS, Aravind L, Agarwal S, Maciejewski JP, Rao A. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, 2010, 468(7325): 839–843. [DOI]
- [15] He YF, Li BZ, Li Z, Liu P, Wang Y, Tang QY, Ding JP, Jia YY, Chen ZC, Li L, Sun Y, Li XX, Dai Q, Song CX, Zhang KL, He C, Xu GL. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 2011, 333(6047): 1303–1307. [DOI]
- [16] Ruoppolo M, Nitti G, Valsasina B, Malorni A, Marino G, Pucci P. Disulfide isoform intermediates in the reoxidation of recombinant human basic fibroblast growth factor. *Biochemistry*, 1993, 32(19): 4991–4996. [DOI]
- [17] Thomas B, Akoulitchev AV. Mass spectrometry of RNA. *Trends Biochem Sci*, 2006, 31(3): 173–181. [DOI]
- [18] Globisch D, Münzel M, Müller M, Michalakakis S, Wagner M, Koch S, Brückl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One*, 2010, 5(12): e15367. [DOI]
- [19] Le T, Kim KP, Fan GP, Faull KF. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Anal Biochem*, 2011, 412(2): 203–209. [DOI]
- [20] Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res*, 2010, 38(19): e181. [DOI]
- [21] Zhao C, Wang HL, Zhao BL, Li CP, Yin RC, Song MY, Liu BD, Liu Z, Jiang GB. Boronic acid-mediated polymerase chain reaction for gene- and fragment-specific detection of 5-hydroxymethylcytosine. *Nucleic Acids Res*, 2014, 42(9): e81. [DOI]
- [22] Davis T, Vaisvila R. High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. *J Vis Exp*, 2011(48): 2661. [DOI]
- [23] Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res*, 2011, 21(12): 1670–1676. [DOI]
- [24] Ficiz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 2011, 473(7347): 398–402. [DOI]
- [25] Adey A, Shendure J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome Res*, 2012, 22(6): 1139–1143. [DOI]
- [26] Shen L, Zhang Y. 5-hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr Opin Cell Biol*, 2013, 25(3): 289–296. [DOI]
- [27] Szulwach KE, Li XK, Li YJ, Song CX, Wu H, Dai Q, Irier H, Upadhyay AK, Gearing M, Levey AI, Vasanthakumar A, Godley LA, Chang Q, Cheng XD, He C, Jin P. 5-hmC-

- mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci*, 2011, 14(12): 1607–1616. [DOI]
- [28] Stroud H, Feng SH, Morey Kinney S, Pradhan S, Jacobsen SE. 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol*, 2011, 12(6): R54. [DOI]
- [29] Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Michael Dixon J, Harrison DJ, Meehan RR. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res*, 2012, 22(3): 467–477. [DOI]
- [30] Thomson JP, Hunter JM, Nestor CE, Dunican DS, Terranova R, Moggs JG, Meehan RR. Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Research*, 2013, 41(22): e206. [DOI]
- [31] Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 2011, 473(7347): 394–397. [DOI]
- [32] Song CX, Szulwach KE, Fu Y, Dai Q, Yi CQ, Li XK, Li YJ, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang BC, Godley LA, Hicks LM, Lahn BT, Jin P, He C. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol*, 2011, 29(1): 68–72. [DOI]
- [33] Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, Lo A, Codelli JA, Bertozzi CR. Copper-free click chemistry for dynamic *in vivo* imaging. *Proc Natl Acad Sci USA*, 2007, 104(43): 16793–16797. [DOI]
- [34] Pastor WA, Huang Y, Henderson HR, Agarwal S, Rao A. The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat Protoc*, 2012, 7(10): 1909–1917. [DOI]
- [35] Robertson AB, Dahl JA, Ougland R, Klungland A. Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nat Protoc*, 2012, 7(2): 340–350. [DOI]
- [36] Cui LB, Chung TH, Tan D, Sun XG, Jia XY. JBP1-seq: a fast and efficient method for genome-wide profiling of 5hmC. *Genomics*, 2014, 104(5): 368–375. [DOI]
- [37] Petterson A, Chung TH, Tan D, Sun XG, Jia XY. RRHP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. *Genome Biol*, 2014, 15(9): 456. [DOI]
- [38] Sun ZY, Terragni J, Borgaro JG, Liu YW, Yu L, Guan SX, Wang H, Sun DP, Cheng XD, Zhu ZY, Pradhan S, Zheng Y. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep*, 2013, 3(2): 567–576. [DOI]
- [39] Wang H, Guan SX, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, Roberts RJ, Zhu ZY, Zheng Y. Comparative characterization of the PvuRts11 family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res*, 2011, 39(21): 9294–9305. [DOI]
- [40] Sun ZY, Dai N, Borgaro JG, Quimby A, Sun DP, Corrêa IR Jr, Zheng Y, Zhu ZY, Guan SX. A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. *Mol Cell*, 2015, 57(4): 750–761. [DOI]
- [41] Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, He C. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat Protoc*, 2012, 7(12): 2159–2170. [DOI]
- [42] Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li XK, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, 2012, 149(6): 1368–1380. [DOI]
- [43] Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, Reik W, Balasubramanian S. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, 2012, 336(6083): 934–937. [DOI]
- [44] Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, Balasubramanian S. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc*, 2013, 8(10): 1841–1851. [DOI]
- [45] Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*, 2010, 7(6): 461–465. [DOI]
- [46] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, De Winter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong XX, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma CC, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. Real-time DNA sequencing from single polymerase molecules. *Science*, 2009, 323(5910): 133–138. [DOI]
- [47] Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. Characterizing and

- measuring bias in sequence data. *Genome Biol*, 2013, 14(5): R51. [DOI]
- [48] Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, Turner SW, He C, Korlach J. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat Methods*, 2012, 9(1): 75–77. [DOI]
- [49] Clark TA, Lu XY, Luong K, Dai Q, Boitano M, Turner SW, He C, Korlach J. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol*, 2013, 11: 4. [DOI]
- [50] Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang XH, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrogel CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggins M, Schloss JA. The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 2008, 26(10): 1146–1153. [DOI]
- [51] Wanunu M, Cohen-Karni D, Johnson RR, Fields L, Benner J, Peterman N, Zheng Y, Klein ML, Drndic M. Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J Am Chem Soc*, 2011, 133(3): 486–492. [DOI]
- [52] Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, Bartlett JJ, Pavlenok M, Gundlach JH. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci USA*, 2013, 110(47): 18904–18909. [DOI]
- [53] Lu X, Song CX, Szulwach K, Wang Z, Weidenbacher P, Jin P, He C. Chemical modification-assisted bisulfite sequencing (CAB-seq) for 5-Carboxylcytosine detection in DNA. *J Am Chem Soc*, 2013, 135(25): 9315–9317. [DOI]
- [54] Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem*, 2014, 6(5): 435–440. [DOI]
- [55] Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li YJ, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, 2013, 153(3): 678–691. [DOI]
- [56] Xia B, Han DL, Lu XY, Sun ZZ, Zhou AK, Yin QZ, Zeng H, Liu MH, Jiang X, Xie W, He C, Yi CQ. Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat Methods*, 2015, 12(11): 1047–1050. [DOI]
- [57] Wen L, Tang FC. Genomic distribution and possible functions of DNA hydroxymethylation in the brain. *Genomics*, 2014, 104(5): 341–346. [DOI]

(责任编辑: 张飞雄)