

DOI: 10.3724/SP.J.1005.2012.00509

5-羟甲基胞嘧啶的研究进展

张燕霞, 高可润, 禹顺英

上海交通大学医学院附属精神卫生中心, 上海 200030

摘要: CpG 二核苷酸中胞嘧啶的甲基化形式 5-甲基胞嘧啶(5-methylcytosine, 5mC)在哺乳动物中是一种常见的表观遗传修饰, 在基因表达调控、发育调节、基因组印迹等方面发挥重要作用。近 3 年来研究发现, 除了 5mC 外, 胞嘧啶碱基的另一种修饰—5-羟甲基胞嘧啶(5-hydroxymethylcytosine, 5hmC)在哺乳动物的多种组织中有着丰富的表达, 它可能与 5mC 有着不同的生物学功能。文章就近年来 5hmC 的研究进展进行了综述。

关键词: 5-羟甲基胞嘧啶; TET 蛋白; 表观遗传

Progress of research on 5-hydroxymethylcytosine

ZHANG Yan-Xia, GAO Ke-Run, YU Shun-Ying

Shanghai Mental Health Center, School of Medicine, Shanghai Jiao Tong University, Shanghai 200030, China

Abstract: 5-methylcytosine (5mC) in cytosine-guanine dinucleotide (CpG) is a usual epigenetic modification in mammals. It plays crucial roles in gene regulation, development, genomic imprinting and so on. In the last three years, it was discovered that in addition to 5mC, another modified cytosine base—5-hydroxymethylcytosine (5hmC) was abundant in many mammalian tissues, which may have different biological function from 5mC. This paper reviews the recent progresses in the studies of 5hmC.

Keywords: 5-hydroxymethylcytosine; TET protein; epigenetics

5-羟甲基胞嘧啶(5-hydroxymethylcytosine, 5hmC)是 5-甲基胞嘧啶(5-methylcytosine, 5mC)的羟甲基化形式, 最早于 1952 年在噬菌体 DNA 中被发现, 它可能被糖基转移酶介导糖基化修饰, 从而使噬菌体基因组在进入宿主后能抵抗宿主限制酶的降解^[1]。1972 年 Penn 等^[2]在哺乳动物大鼠(*Rattus norvegicus*)、小鼠(*Mus musculus*)以及卵生动物牛蛙(*Rana chatesbiana*)脑组织提取的 DNA 中也发现了羟

甲基化修饰的胞嘧啶, 约占 DNA 总胞嘧啶的 15% 左右, 但此后的研究未能重复这一结果^[3], 直到 2009 年 Tahiliani^[4]和 Kriaucionis^[5]的两个研究团队在同一期 *Science* 中分别报道了 5hmC 在人、小鼠大脑及胚胎干细胞中有着丰富表达, 羟甲基化的概念才又一次进入人们的视野并得到了重视。目前 5hmC 是继 5mC 被称为“第 5 个碱基”之后的“第 6 个碱基”。虽然 5hmC 的功能目前尚未完全明了, 但它可能也是

收稿日期: 2011-08-29; 修回日期: 2011-12-04

基金项目: 国家自然科学基金项目(编号: 81071094)资助

作者简介: 张燕霞, 硕士研究生, 专业方向: 精神病遗传学。Tel: 13764690291; E-mail: 998179_zyx@163.com

通讯作者: 禹顺英, 博士, 副教授, 硕士生导师, 研究方向: 精神病遗传学。E-mail: yushuny@yahoo.com

网络出版时间: 2012-3-16 02:50:48

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20120316.1450.004.html>

一个重要的表观遗传学标记,可能与去甲基化过程或转录调控有关^[6-10]。

1 5hmC 的生成

DNA氧化损伤可使碱基产生多种修饰形式,如8-羟基鸟嘌呤、5-羟基胞嘧啶等,故而早期认为细胞内5hmC的产生可能与氧化应激有关,但Kriaucionis等^[5]发现5hmC的产生并不伴随其他DNA氧化损伤产物的增加,并且其在浦肯野细胞和颗粒细胞中的总量与成年小鼠的年龄之间没有相关性,因此氧化应激并不是导致细胞内5hmC形成的直接原因,可能只是一种驱动因素。

2009年Tahiliani等^[4]通过生物信息学技术在哺乳动物中发现了与锥体虫(*Trypanosoma*)J结合蛋白1(J binding protein 1, JBP1)和J结合蛋白2(J binding protein 2, JBP2)同源的TET蛋白家族(Ten-Eleven-Translocation protein, 包括TET1、TET2和TET3),这一重要发现为进一步理解5hmC的生成带来了希望。锥体虫含有的 β -D-糖化-羟甲基尿嘧啶(β -D-glucosyl-hydroxymethyluracil, 简称碱基J)是胸腺嘧啶(Thymine, 简称为T)的甲基依序羟基化和糖基化后形成的。碱基J的生物合成需要两种蛋白酶JBP1、JBP2,这两种酶是二价铁离子(Fe^{2+})和 α -酮戊二酸(2-oxoglutarate, 2OG)依赖性氧化酶家族成员。与5mC相似,碱基J也与基因沉默有关。Tahiliani等^[4]通过体外细胞培养发现TET1能将5mC修饰成5hmC,这种修饰既能作用于完全甲基化的DNA,又能作用于半甲基化的DNA。Ito等^[11]和Ko等^[12]发现另外两种TET蛋白(TET2和TET3)也具有相似的催化反应(图1)。

*TET1*位于染色体10q21, mRNA总长度约9.6kb, 包含12个外显子, 编码2136个氨基酸(Amino acids, aa)。*TET2*位于染色体4q24, mRNA总长度9.7kb, 包含11个外显子, 由于切割位点的不同可有3个不同的亚型:第1个亚型无外显子3b, 可编码2002个aa;第2个亚型无外显子3b和3c, 可编码1194个aa;第3个亚型缺乏外显子3c之后的所有外显子, 外显子3b处是个终止密码子, 最终编码1164个aa。这3个亚型在不同的组织细胞内表达存在差异。*TET3*位于染色体2p13, mRNA长10.9kb, 包含9个外显子, 编码1660个aa, 通过互补DNA筛查发现它可能包括了

3个亚型:第1个亚型编码了全部aa(1660), 第2个亚型缺失了1440~1555位的aa, 第3个亚型缺失了728~1660位的aa, 目前尚未有关于这3个亚型的表达数据^[13]。

所有的TET蛋白家族都含有3个 Fe^{2+} 结合位点(TET1: aa1672, 1674, 2028; TET2: aa1382, 1384, 1884; TET3: aa942, 944, 1538)和一个2OG结合位点(TET1: aa2043, TET2: aa1896, TET3: aa1553), 前者与 Fe^{2+} 结合后可表现出蛋白的催化活性。蛋白二级结构预测显示TET蛋白都含有一个富含半胱氨酸的结构域(Cysteine-rich domain, CD)和几条 α 链, 这些 α 链构成了一个具有2OG- Fe^{2+} 依赖性氧化酶特征的双链螺旋结构(Double-stranded α -helix, DSBH), CD和DSBH这两个结构合在一起形成了该蛋白的催化结构域。除了上述结构域外, TET1还包括了3个核定位信号(Nuclear localization signals, NLS)(AA20-50, 620-653, 1158-1162), 提示了TET1在核内的潜在功能。而细胞核定位信号的预测和分析显示TET2和TET3都不存在NLS区^[14]。另外, TET1还含有一个保守的CXXC结构域^[4, 15], 涉及DNA甲基化和组蛋白甲基化/去甲基化因子, 如DNA甲基转移酶1(DNA methyltransferase 1, Dnmt1)、甲基CpG结合蛋白1(Methyl-CpG-binding domain protein 1, MBD1)、赖氨酸特异性去甲基酶1(Lysine-specific demethylase 2A, KDM2A)及CXXC指蛋白1(CXXC finger protein 1, CFP1)等也含有该结构域, 目前认为CXXC结构域可使蛋白能够与非甲基化的CpG DNA结合从而修饰组蛋白或DNA甲基化(图2)。

目前研究发现这3个TET蛋白能在不同组织及细胞内将5mC转变成5hmC^[16-18]。Koh等^[18]发现*Tet1*和*Tet2*在小鼠胚胎干细胞内丰富表达, *Tet3*则低表达。进一步功能分析发现*Tet1*和*Tet2*表达受到了多能性调控网络及胚胎干细胞分化潜能的调控, 5hmC随着*Tet1*和*Tet2*的变化发生相应的改变, 提示5hmC在胚胎干细胞维持和分化过程中具有重要作用。在人类疾病研究中发现*TET1*及*TET2*可能与血液恶性疾病密切相关, 如*TET1*在急性髓性白血病(acute myeloid leukemia, AML)中与骨髓/淋巴或混合白血病基因(Myeloid/lymphoid or mixed-lineage leukemia gene, MLL)融合^[19], *TET2*在髓系癌前或恶性血液病中(如骨髓发育不良、骨髓增生综合征和急性髓性白

血病等)频发突变^[12, 20]。另外, 在前列腺癌、乳腺癌和结肠癌等亦发现 5hmC水平显著下降^[21]。这些发现促进了学术界对 5hmC研究的日益关注和重新认识^[22]。

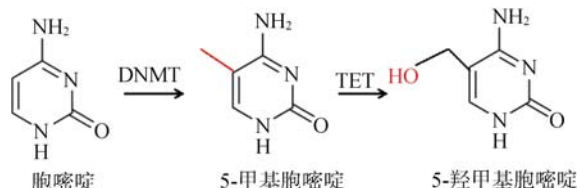


图1 5mC 和 5hmC 的生成图示

2 5hmC 在哺乳动物中的组织分布特点

DNA 碱基修饰在哺乳动物的不同组织间存在差异性分布, 这种差异还会随着年龄、疾病等因素发生变化。5hmC 作为 5mC 的特殊修饰形式是否与 5mC 存在一样的组织分布特点呢?

早期, 研究者们通过色谱或质谱等灵敏度有限的检测方法发现 5hmC 的分布存在明显的组织差异性。在动物中, Kriaucionis 等^[5]基于薄层色谱 (Thin-layer chromatography, TLC) 及质谱分析 (Mass spectrometry, MS) 发现 5hmC 在小鼠脑组织中有着丰富表达, 尤其是皮质与脑干区, 其中 5hmC 在浦肯野

细胞中约占总核苷酸的 0.6%, 而在颗粒细胞中约占 0.2%, 在肌肉、心肺中低于 0.2%, 在肝、脾、肾、睾丸中甚至低于 0.1%, 在小鼠神经母细胞瘤细胞系 (Neuro2A) 和人宫颈癌细胞系 (HeLa 细胞) 中未发现 5hmC 的存在。随后 Munzel 等^[23]利用液相色谱串联质谱法 (Liquid chromatography-mass spectrometry, LC-MS) 发现小鼠脑组织中 5hmC 约占所有脱氧鸟嘌呤的 0.4%~0.7%, 但其分布特点却与 Kriaucionis 等^[5]的研究结果有所不同, 5hmC 在皮层、海马高表达, 脑干与嗅球区中度表达, 而小脑及视网膜低表达, 浦肯野细胞中无表达。

随着 5hmC 检测技术的不断发展及其检测敏感性的提高, 5hmC 在小鼠不同组织内的表达量高于早期的报道数据。Szwagierczak 等^[17]通过放射性的 hmC 糖基化检测分析了小鼠胚胎干细胞及不同成年组织基因组内 5hmC 的含量, 结果显示 5hmC 含量较以往报告的数据更高, 如胚胎干细胞中高达 0.3%, 脑组织 1%~1.3%, 肝肾肺 0.3% 以上, 脾、胸腺、睾丸 0.1% 左右。然而与 5mC 在不同组织间的表达差异不大相比, 5hmC 在不同组织之间的表达差异很大。Globisch 等^[16]基于同位素的 HPLC-MS 的数据结果, 根据 5hmC 在小鼠不同组织间的表达水平将组织大致分为三类: (1) 高表达组织 (0.3%~0.7%), 包

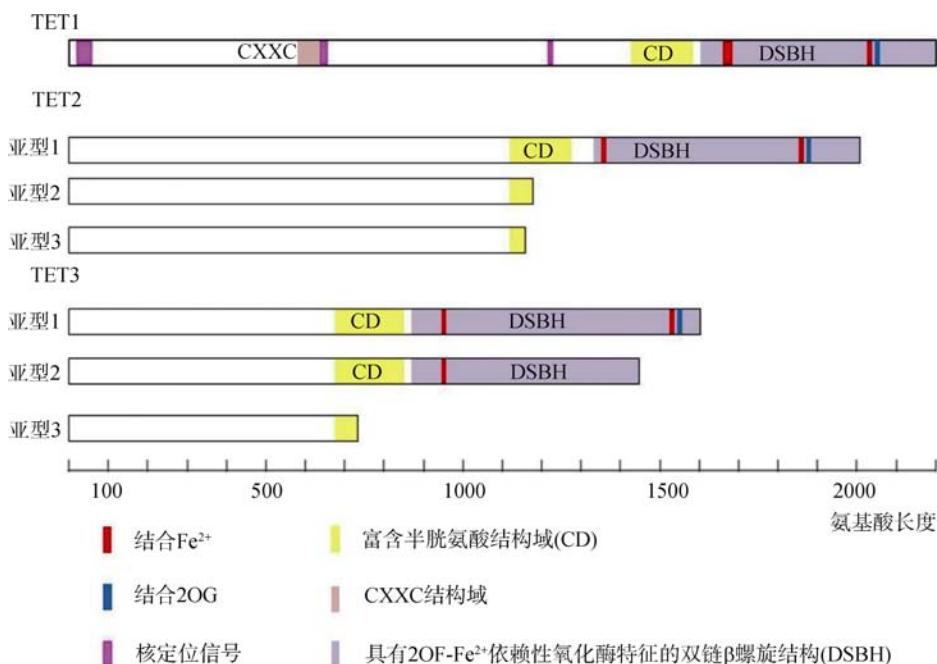


图2 TET 蛋白家族结构示意图

括中枢神经系统和脊髓,虽然脊髓不是大脑的直接组成部分,但 5hmC 的表达量高达 0.47%;(2)中度表达组织(0.15%~0.17%),包括肾、鼻粘膜上皮、膀胱、心、骨骼肌和肺;(3)低表达组织(0.03%~0.06%),包括肝、脾、内分泌腺(睾丸和垂体)。可以看出,不同组织之间 5hmC 的表达量可以相差 10 倍以上(图 3)。

在人类,5hmC 在不同组织间的表达也存在明显的差异性。Li 等^[21]通过免疫沉淀法发现 5hmC 在人脑组织高达 0.67%,肝、肾、结直肠组织相对高表达,约 0.38%~0.57%,而在肺相对低表达,约 0.14%,心、乳腺、胎盘组织内极低表达,约 0.05%~0.06%,在 4 个癌细胞系 HeLa(人宫颈癌)、ANCA3(子宫内膜癌)、HCT116 和 SW620(结肠癌)内几乎未发现 5hmC 的存在,可见 5hmC 在人类不同组织间的表达差异很大(图 4)。Li 等^[21]同期检测了上述组织内 5mC 的含量,发现各组织间 5mC 的含量相差不大(约 1~2.5 倍),这与以往的研究结果相似^[24],也说明 5hmC 和 5mC 在组织间的分布不同。

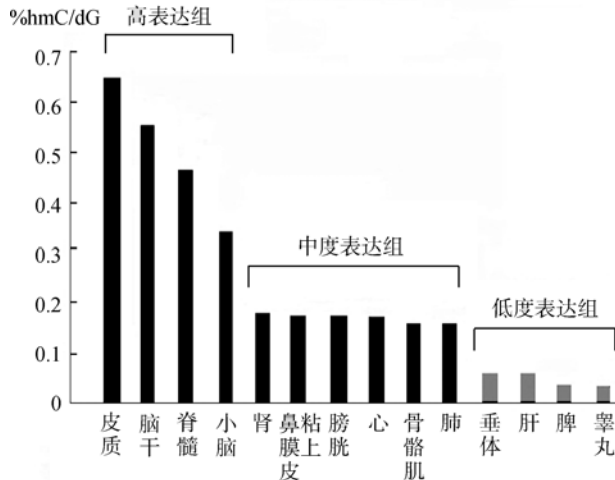


图 3 小鼠组织中 5hmC 表达情况^[16]

值得注意的是,5hmC 的表达水平在胚胎发育的不同时期也不同。有研究发现随着胚胎干细胞的分化,5hmC 表达水平逐渐下调,在诱导性多潜能干细胞产生期间又再次上调,说明 5hmC 的表达可能与多潜能细胞状态有关^[25]。出生后 5hmC 的表达又会随着脑的发育成熟而上调,例如出生 90 d 后的成年小鼠海马内 5hmC 表达量是出生后 1 d 小鼠的 2 倍^[23],随后的研究亦发现小脑 5hmC 的表达量也表现出类似变化,由出生后 7 d 的 0.1% 增加到成年后的 0.4%,

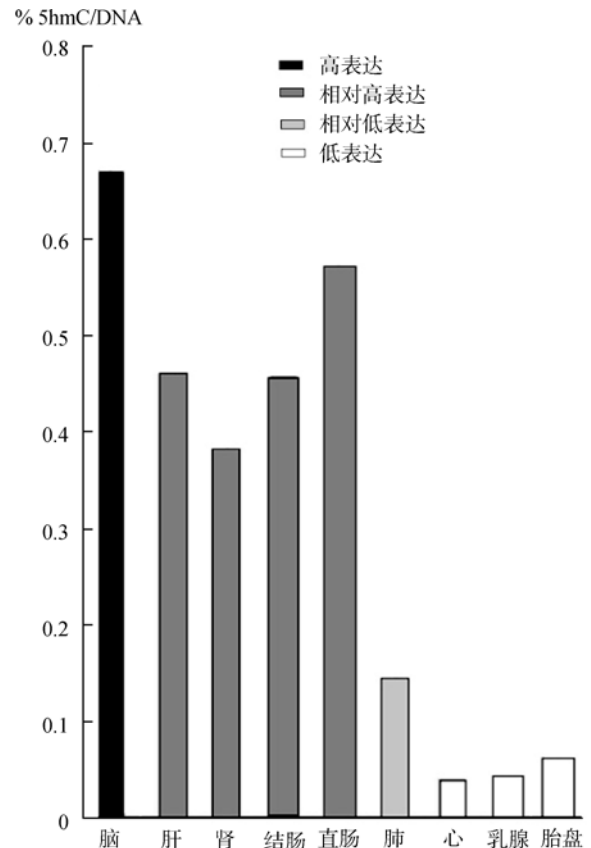


图 4 人不同组织中 5hmC 的表达情况^[21]

说明 5hmC 在脑组织成熟及神经发育过程中可能起着重要作用^[26]。

总之,目前 5hmC 表达谱的研究尚处于起始阶段,主要集中在人、小鼠胚胎干细胞及器官组织的研究,虽然不同研究团队之间的研究结果可能存在一定的差异,但是可以基本肯定的是,无论是小鼠还是人类 5hmC 在胚胎干细胞及脑组织中都有着丰富表达。

3 5hmC 的基因组分布特点

了解 5hmC 在基因组中的分布特点对于了解其生物学功能是非常重要的。Song 等^[26]报道了小鼠脑内 5hmC 的全基因组图,发现 5hmC 特异性的富含于基因内部(Genebody)以及转录起始位点(Transcriptional start site, TSS)、转录终点(Transcription termination sites, TTS)的近端上游及下游区域。徐国良研究小组^[27]对小鼠胚胎干细胞的研究也发现 5hmC 主要集中于基因富集区(Gene-rich areas),尤其在基

因内部高度富集, 可高达 51.5%, 启动子区达 7.6%, 基因间隔达 40.9%, 而基因沙漠区(Gene deserts)罕见, 并且 5hmC与CpG含量之间存在相关性。在启动子区域, 5hmC的含量与该区域内CpG含量呈负相关, 即当CpG含量低中度时 5hmC富集。另外, 5hmC在具有独特的组蛋白修饰二价区域 (H3K4me3+/H3K27me3+)的启动子区也有明显富集。5hmC在TSS及 5'非编码区(5'untranslated regions, 5'UTR)也有富集, 其富集度由TSS位点逐渐向TTS位点增加, 在TTS位点周围又陡然下降^[27,28]。在基因内部, 5hmC多富集于CpG中度密集区, 在许多蛋白-DNA相互作用位点, 如绝缘子及多潜能型转录因子结合位点(Nanog、Sox2 和Oct4)、与增强子相关的组蛋白修饰位点(H3K4me1 和H3K27ac)等, 5hmC也较 5mC相对富集, 但细胞分化后这些区域内富集的 5hmC逐渐消失而 5mC渐现^[28,29]。

Jin等^[30]分析了人脑额叶组织内 5hmC的分布情况, 发现 5hmC主要富集于启动子(55%~59%)及基因内部(35%~38%), 而基因间隔区只占了不到 6%。在人胚胎干细胞中, Szulwach等^[31]和Stroud等^[32]也发现 5hmC除主要集中在基因内部尤其是外显子之外, 还集中在其他区域, 如基因间隔、基因内的CpG岛区、5'UTR区、启动子区, 以及与增强子相关的组蛋白修饰位点, 如H3K4me1 和H3K27ac, 同时 5hmC还

富集于其他的蛋白-DNA相互作用位点, 如OCT4 和NANOG结合位点, 这一结果与Pastor等^[28]报道的小鼠胚胎干细胞中 5hmC的基因组分布相似。另外Szulwach等^[31]发现 5hmC富集区通常含中度水平的CG含量, 在高密度的CpG区 5hmC水平反而降低, 而Stroud等^[32]发现 5hmC区富含CG含量, 但该区域存在GC含量偏斜, 即 5'端G大于C, 3'端则相反, 这种现象在与基因和增强子重叠的 5hmC区域以及未与上述元件重叠的区域都能观察到, 提示这种序列组成可能是所有 5hmC区域的共同特征, 因此 5hmC可能会出现在CG偏斜区域。

综上所述, 目前的研究都支持 5hmC 与 5mC 都高度富集于基因内部, 尤其是外显子区, 然而只有 5hmC 在基因 TSS、5'UTR 区富集。5mC 虽然也能定位于 TSS 的上下游, 但其主要富集于 TSS 3'端, 并且在其他基因转录调控元件区如增强子、启动子等 5hmC 表达水平较 5mC 相对高, 可见 5hmC 在基因调控元件区具有独特的分布特点, 这些都不是简单的用 5mC 的分布特点可以解释的, 另一方面也提示 5hmC 在调控基因转录中可能起着重要作用(表 1)。

4 5hmC 生物学功能

5mC 与基因表达抑制有关, 其羟基化后形成的 5hmC 具有什么功能呢? 目前虽然还不甚清楚, 但

表 1 5hmC 在基因组中的分布

种属	组织	增强子	启动子	基因内部	基因间隔	文献
鼠	脑	/	/	5hmC 特异性的富含于基因内部以及 TSS、TTS 的近端上游及下游区域	/	[26]
		在组蛋白 H3K4me1、H3K27ac 修饰位点及蛋白-DNA 相互作用位点(Nanog、Sox2 和 Oct4)富集	6%(远端启动子 3.9%+近端启动子 2.1%), 5hmC在组蛋白 H3K4me3 修饰位点富集	58%(5'UTR 0.6%, 外显子 6.4%, 内含子 48.4%, 3'UTR 2.6%), 5hmC 水平与中度的 CpG 含量有关	30.4%	[29]
		/	7.6%, 5hmC 水平与低中度的 CpG 含量及组蛋白 H3K27me3 修饰有关	51.5%(外显子 23.9%+内含子 27.6%), 5hmC 密度由 TSS 位点逐渐增加, 在 TTS 位点周围又陡然下降	40.9%	[27]
人	脑	/	55%~59%	35%~38%	6%	[30]
		在组蛋白修饰位点(H3K4me1 和 H3K27ac)及其他蛋白-DNA 相互作用位点(OCT4 和 NANOG)富集	/	5hmC 趋向定位于上游近端区域及基因内部	/	[32]
		在组蛋白修饰位点富集, (如, H3K4me1、H3K27ac 等)	与组蛋白修饰有关(如, H3K4me1、H3K4me2 等)	基因内的 CpG 岛、5'UTR、外显子等 5hmC 含量丰富, 内含子区相对较少	稀少	[31]

注: “/”代表本研究中未涉及该方面的内容。

它可能与去甲基化过程和基因活性增加有关。

关于 5hmC 在去甲基化过程中的作用机制目前主要有两种假设(图 5): 第一种假设是 5hmC 参与了被动去甲基化过程, 在细胞分裂过程中 5hmC 的存在干扰了 DNMTs 介导的 CpG 二核苷酸中胞嘧啶的甲基化过程, 从而导致被动去甲基化^[33]; 第二种假设是 5hmC 参与了主动去甲基化过程, 基于哺乳动物 5hmC 特异性 DNA 转葡萄糖基酶的发现^[34], 故推测 5hmC 可能是与 DNA 修复机制相关的主动去甲基化通路中的一个重要的中间产物^[4], 通过一种至今尚未发现的酶机制可将 5hmC 转变成胞嘧啶。

值得一提的是, 近期 Ito 等^[35]又在 *Science* 上报到 TET 蛋白也可在体外将 5hmC 转变成 5-甲酰胞嘧啶(5-formylcytosine, 5fC)和 5-羧基胞嘧啶(5-carboxyleytosine, 5caC), 随后在小鼠基因组中也发现了上述两个碱基修饰的存在, 并且 5hmC、5fC 和 5caC 的基因组含量随 TET 蛋白的过度表达和耗尽而增减。徐国良研究小组^[35]也在 *Science* 上报到 TET 在体内和体外实验中都能将 5mC 和 5hmC 氧化成 5caC, 之后 5caC 会被胸腺嘧啶 DNA 糖基化酶(Thymine- DNA glycosylase, TDG)识别并消化, 进一步 TDG 敲除实验也说明 5caC 的积累。这些研究都表明 5mC 在 TET

蛋白氧化后经过 TDG 介导的 5caC 碱基切割过程是 DNA 主动去甲基化的一条重要途径。因此 5hmC 很有可能只是 DNA 去甲基化过程中的一个中间产物, 但是即使 TDG 的存在也不能排除 TET 蛋白与其他去甲基化途径如碱基切割修复途径等共同作用的可能性, Cortellino 等^[36]和 Guo 等^[37]的研究恰好支持了这种可能性。

5hmC 在哺乳动物基因组中的稳定表达及其在基因调控元件区独特的分布特点显示了它与 5mC 不同的基因调控作用。在调控基因转录的机制方面, 有研究认为由于羟甲基化的 5mC 不能识别甲基化 CpG 结合蛋白, 如 MBD1、MBD2、MBD4 和 MECP2 等, 故能招募组蛋白去乙酰化酶(Histone deacetylases, HDAC), 从而使基因处于转录活性状态^[31, 33]。目前该方面主要来自胚胎干细胞的研究, 但各研究团队间的结果不尽相同。

Williams 等^[38]认为 5hmC 富集与基因转录激活没有相关性, 并且 TET1 蛋白调节基因表达变化并不依赖其羟化酶活性。

Pastor 等^[28]和 Ficiz 等^[39]虽然认为羟甲基化水平与基因表达水平有关, 但两人的研究结果却显示出 5hmC 对基因转录的不同调控方向: Pastor 等^[28]发

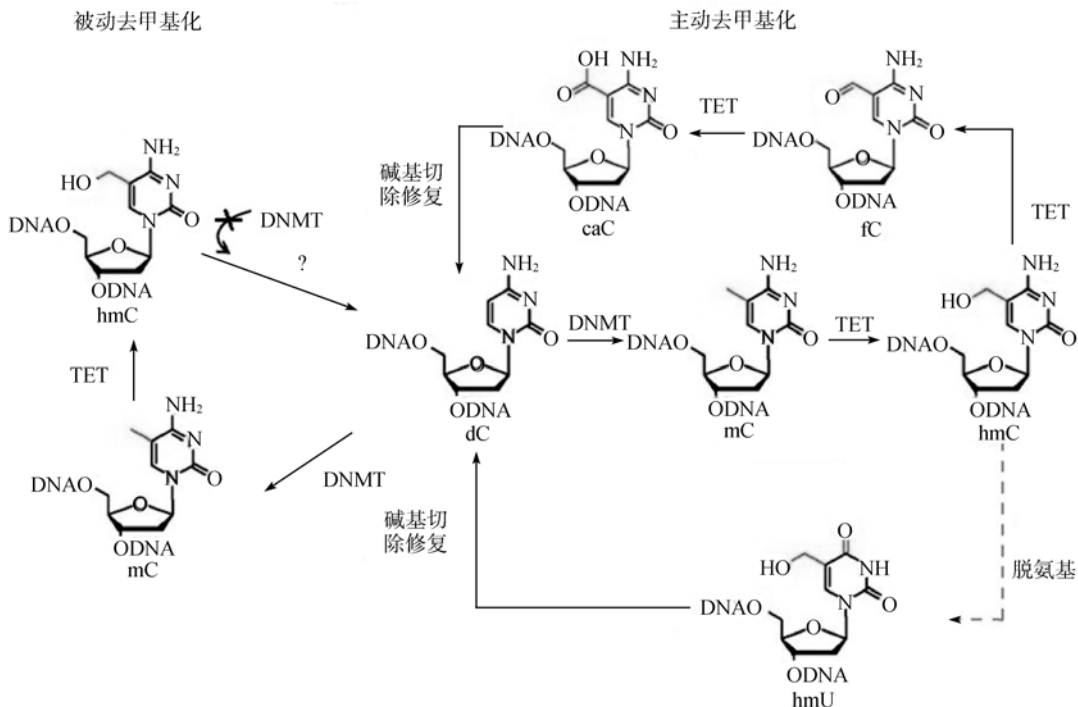


图 5 5hmC 在 DNA 去甲基化过程中的作用

现胚胎干细胞中启动子区(含二价结构域)富集 5hmC 的基因通常比没有 5hmC 富集的基因表达水平低, 随着胚胎干细胞分化这些基因的表达水平又逐渐上升, 故认为在胚胎干细胞中这些基因启动子区域的 5hmC 对基因转录可能具有负调控的作用, 虽然该特点与 5mC 相似, 但是启动子区含二价结构域的基因通常并不富集 5mC, 并且富集 5mC 的基因表达水平并不会随胚胎干细胞分化而上调; 相反, Ficiz 等^[39]通过对全基因组分析发现基因启动子区或外显子区富含 5hmC 与基因表达增加呈正相关, 与 Song 等^[26]在 小鼠小脑中的发现一致, 却与 5mC 主要分布于基因内部及基因远端区域沉默基因表达不同, 提示了 5hmC 可能在激活或维持基因表达时起重要作用, 而随着胚胎干细胞的分化原下调基因的启动子区的 5hmC 水平则会显著下调, 并伴随着 5mC 水平的上调^[39]。

Wu 等^[26]及 Xu 等^[28]却认为 5hmC 对基因的转录调控作用取决于其在基因组中的位置, 研究发现 5hmC 在低表达基因的启动子区以及高或中度表达基因的基因内部相对富集, 在高表达基因的 TSS 位点极度低水平, 故认为启动子区域的 5hmC 可能起着转录沉默的作用, 而基因内部的 5hmC 与转录激活有关。这与人类额叶组织 5hmC 的表达研究类似, Jin 等^[31]发现基因内部 5hmC 的表达水平与基因表达水平呈正相关, 但是含低度 CpG 含量的启动子区 5hmC 水平与该基因表达水平也呈正相关。

这些研究结果不一致的原因可能是分析方法以及基因研究范围的差异导致。此外, 5hmC 与基因转录活性之间的相关性是否存在直接的因果关系也还有待于进一步的验证。

5 5hmC 的检测方法及面临的技术挑战

要了解 5hmC 在生物学通路中的作用, 掌握 5hmC 在基因组 DNA 中的定量及定位的可靠数据是非常有必要的, 但这项工作如今仍面临着巨大技术挑战, 因为标准的生物化学方法不容易区分 5mC 和 5hmC。目前研究中所用的 hmC 检测方法主要包括以下几种:

(1) 基于薄层色谱(TLC)和质谱(MS)的定性及定量方法: TLC 是利用样本各组分在薄层固定相(如硅

胶)上的不同移动距离, 在展开罐内展开后即可实现样品内不同组分的分离, 继而通过测定各斑点的位置及其密度就可以完成对样品的定性、定量分析的色谱法。在 TLC 的基础上还发展了分离度、灵敏度和重现性更高的二维 TLC、高性能(效)液相色谱法(High-pressure liquid chromatography, HPLC)等。MS 分析与 TLC 原理不同, 由于不同分子 m/z (质量电荷比)的不同, 即可在通过电场(或电磁场)时将它们区分开来, 该方法除灵敏度高、选择性好之外更可提供丰富的组成和结构信息。2009 年 Tahiliani 等^[4]和 Kriaucionis 等^[5]相继通过 TLC 或二维 TLC、HPLC 及 MS 法对 5hmC 进行了定性及定量分析。薄层色谱与质谱的联用实现了优势互补, 更利于复杂样品的定性。Munzel 等^[23]通过 HPLC-MS 联用精确测定了小鼠组织内 5hmC 的分布情况, Le 等^[40]则设计了一种更快速、稳健、准确的 MS 分析方法即多重反应监测的液相色谱电喷射离子串联质谱法(Liquid chromatography electrospray ionization tandem mass spectrometry with multiple reaction monitoring, LC-ESI-MS/MS-MRM), 它可同时检测基因组 DNA 中 5mC 和 5hmC 的水平, 比目前 5hmC 定量分析方法(如 TLC 和下面介绍的糖基化的 5hmC 放射性标记法)更敏感, 但不能提供 DNA 序列中 hmC 的某个特定位置上的单个碱基的信息。

(2) 单分子实时(SMRT)测序^[41]: SMRT 是一种边合成边测序的方法, 在这项技术中, DNA 聚合酶可催化含有荧光标记的核苷酸合成完整的 DNA 链, 荧光脉冲的到达和持续时间可以提供关于聚合酶动力学的信息从而可区分出 DNA 模板中的修饰性核苷酸, 包括 N6 甲基腺嘌呤、5mC 和 5hmC。这项方法可以经得起长阅读长度检验, 并且可能在高度重复基因组区域绘制甲基化类型图谱。

(3) 5hmC 酶标记法^[42]: 基于 T4 噬菌体 β -葡萄糖转移酶(T4 β -glycosyltransferase, β -GT)在葡萄糖的供体尿核苷二磷酸葡萄糖(Uridine Diphosphoglucose, UDP-Glu)存在的情况下能将 5hmC 糖基化形成 β -葡萄糖基-5-羟甲基胞嘧啶(β -glucosyl-5-hydroxymethylcytosine, 5-ghmC), Szwagierczak 等^[17]设计了一种定量方法: 即将含有放射性标记的葡萄糖残基 UDP-[3H]-Glu 转移至 hmC 形成 [3H]-ghmC, 然后进行放射性计数的定量方法。虽然该方法敏感性较高,

不受序列偏倚的影响而且可以同时进行大样本的实验,不需要特别专业及昂贵的设备,但是该方法不能确保完整的酶学反应,也不能同时检测 5mC 水平。Song等^[26]用化学标记取代了放射性标记,通过 β -GT将含化学修饰的葡萄糖残基UDP-6-N3-Glu转移至 5hmC形成N3-5ghmC,再通过点击化学(Click chemistry)方法在每个hmC残基上添加一个生物素单位,从而使它们从含有hmC的DNA链中分离出来,结合测序分析了解 5hmC在基因组中的分布规律,也可以通过生物素特异性的抗体进行 5hmC定量分析。

(4)限制性内切酶法^[43]:限制性内切酶法也是基于 β -GT的 5hmC的糖基化反应进行的。甲基化依赖的限制性内切酶*MspI*和*HpaII*可识别同样的序列(CCGG),但它们对甲基化状态的敏感性是不同的。*HpaII*只切割完全未修饰的位点,胞嘧啶上的任何修饰(5-mC、5-hmC或 5-ghmC)都阻碍切割。*MspI*则识别并切割 5-mC和 5-hmC, 5-ghmC除外。如果CpG位点含有 5hmC,那么糖基化和消化之后能检测到条带,但是未糖基化的对照反应中没有。如果使用实时定量PCR,那么就可以估计这个位点大概有多少羟甲基胞嘧啶。此外,这种方法还能与现有的分析技术兼容,能够扩展到高通量。

(5)沉淀法:包括有 5-亚甲基磺酸胞嘧啶(CMS)沉淀法、GLIB 沉淀法、羟甲基化 DNA 免疫沉淀(5hmC antibody-based DNA immunoprecipitation, hMeDIP)法及 JBP1 沉淀法。

CMS 沉淀法,即通过亚硫酸氢钠处理基因组DNA后,非甲基化的胞嘧啶通过脱氨基和脱磺酸基作用可以转变成尿嘧啶,5mC不变,5hmC由于不能脱氨基而形成CMS,然后利用CMS特异的抗血清将包含CMS的DNA免疫沉淀, DNA斑点杂交分析(Dot blot assay)时抗CMS抗体较抗 5hmC抗体更敏感,且对抗体浓度依赖性较低,通过对包含 5hmC的DNA片段进行高通量测序可进行 5hmC的全基因组定位^[28]。

GLIB沉淀法, GLIB是糖基化(Glucosylation)、高碘酸氧化(Periodateoxidation)和生物素化(Biotinylation)3种反应结合的方法的缩写,它是利用一些酶学和化学反应的组合来分离出包含单个 5hmC的DNA片段。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。GLIB处理的DNA可结合使用Helicos单分子DNA测序进行全基因组定位,这种方法不需要扩增步骤,从而避免了PCR偏向^[28]。

hMeDIP法,利用 5hmC特异抗体使得 5hmC的直接免疫沉淀。再结合高通量测序或芯片技术进行基因组定位^[27, 29, 39],然而Pastor等^[28]发现选择性标记 5hmC 后再对标记物进行免疫沉淀比直接采用 5hmC抗体免疫沉淀羟甲基化DNA 具有更高的灵敏度。

JBP1 沉淀法,糖基化的 5hmC可以被JBP1 蛋白有效并选择性的沉淀到磁性珍珠上,再结合定量PCR、微阵列和测序分析可识别或定量 5hmC^[41]。

6 结语与展望

5hmC 在哺乳动物组织中的“重新发现”只有短短的 3 年,目前的检测技术也还有待完善,对其功能的了解还处于起步状态,但它对我们认识表观遗传调控在人体发育和疾病发生中的作用增添了一个新的重要维度。但是,5hmC 的分布是否像 5mC 一样局限于 CpG 二核苷酸? TET 蛋白的靶基因和功能是什么? 5hmC 的功能及其在疾病中的作用机制是怎样的? 等等问题都亟待解决,鉴于该领域的技术平台和实验模型已取得的巨大进步,这些问题的答案可能会很快揭晓。

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] Penn NW, Suwalski R, O'Riley C, Bojanowski K, Yura R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J*, 1972, 126(4): 781-790. DOI
- [3] Kothari RM, Shankar V. 5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. *J Mol Evol*, 1976, 7(4): 325-329. DOI
- [4] 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](#)
- [5] 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](#)
- [6] Liutkevičiūtė Z, Lukinavičius G, Masevičius V, Daujotytė D, Klimašauskas S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nat Chem Biol*, 2009, 5(6): 400–402. [DOI](#)
- [7] Miller CA, Sweatt JD. Covalent modification of DNA regulates memory formation. *Neuron*, 2007, 53(6): 857–869. [DOI](#)
- [8] Popp C, Dean W, Feng SH, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*, 2010, 463(7284): 1101–1105. [DOI](#)
- [9] Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science*, 2001, 293(5532): 1089–1093. [DOI](#)
- [10] Zhu JK. Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet*, 2009, 43: 143–166. [DOI](#)
- [11] 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](#)
- [12] 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](#)
- [13] Mohr F, Döhner K, Buske C, Rawat VPS. TET genes: new players in DNA demethylation and important determinants for stemness. *Exp Hematol*, 2011, 39(3): 272–281. [DOI](#)
- [14] Cokol M, Nair R, Rost B. Finding nuclear localization signals. *EMBO Rep*, 2000, 1(5): 411–415. [DOI](#)
- [15] Zhang HK, Zhang X, Clark E, Mulcahey M, Huang S, Shi YG. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res*, 2010, 20(12): 1390–1393. [DOI](#)
- [16] 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](#)
- [17] 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](#)
- [18] Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, 2011, 8(2): 200–213. [DOI](#)
- [19] Lorschach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, 2003, 17(3): 637–641. [DOI](#)
- [20] Holmfeldt L, Mullighan CG. The role of TET2 in hematologic neoplasms. *Cancer Cell*, 2011, 20(1): 1–2. [DOI](#)
- [21] Li WW, Liu M. Distribution of 5-hydroxymethylcytosine in different human tissues. *J Nucleic Acids*, 2011, 2011: 870726. [DOI](#)
- [22] Ndlovu MN, Denis H, Fuks F. Exposing the DNA methylome iceberg. *Trends Biochem Sci*, 2011, 36(7): 381–387. [DOI](#)
- [23] Münzel M, Globisch D, Brückl T, Wagner M, Welzmler V, Michalakakis S, Müller M, Biel M, Carell T. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl*, 2010, 49(31): 5375–5377. [DOI](#)
- [24] Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res*, 1982, 10(8): 2709–2721. [DOI](#)
- [25] Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, Chebotareva T, Pells S, Hannoun Z, Sullivan G, Chandran S, Hay DC, Bradley M, Wilmot I, De Sousa P. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell Res*, 2011, 21(9): 1332–1342. [DOI](#)
- [26] 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](#)
- [27] Xu YF, Wu FZ, Tan L, Kong LC, Xiong LJ, Deng J, Barbera AJ, Zheng LJ, Zhang HK, Huang S, Min JR,

- Nicholson T, Chen TP, Xu GL, Shi Y, Zhang K, Shi YG. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell*, 2011, 42(4): 451–464. [DOI](#)
- [28] 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](#)
- [29] Wu H, D'Alessio AC, Ito S, Wang ZB, Cui KR, Zhao KJ, Sun YE, Zhang Y. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev*, 2011, 25(7): 679–684. [DOI](#)
- [30] Jin SG, Wu XW, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res*, 2011, 39(12): 5015–5024. [DOI](#)
- [31] Szulwach KE, Li XK, Li YJ, Song CX, Han JW, Kim SS, Namburi S, Hermetz K, Kim JJ, Rudd MK, Yoon YS, Ren B, He C, Jin P. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet*, 2011, 7(6): e1002154. [DOI](#)
- [32] Stroud H, Feng SH, Kinney SM, 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](#)
- [33] Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res*, 2007, 67(3): 946–950. [DOI](#)
- [34] Cannon SV, Cummings A, Teebor GW. 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. *Biochem Biophys Res Commun*, 1988, 151(3): 1173–1179. [DOI](#)
- [35] 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](#)
- [36] Cortellino S, Xu JF, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A. Thymine DNA glycosylase is essential for active DNA demethylation by linked deaminationbase excision repair. *Cell*, 2011, 146(1): 67–79. [DOI](#)
- [37] Guo JU, Su YJ, Zhong C, Ming GL, Song HJ. Hydroxylation of 5methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, 2011, 145(3): 423–434. [DOI](#)
- [38] Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PAC, Rappsilber J, Helin K. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, 2011, 473(7347): 343–348. [DOI](#)
- [39] 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](#)
- [40] 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](#)
- [41] 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](#)
- [42] Höbartner C. Enzymatic labeling of 5-hydroxymethylcytosine in DNA. *Angew Chem Int Ed Engl*, 2011, 50(19): 4268–4270. [DOI](#)
- [43] Davis T, Vaisvila R. High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. *J Vis Exp*, 2011, (48): e2661. [DOI](#)