

## RNA 中 6-甲基腺嘌呤的研究进展

李语丽<sup>1,3</sup>, 于军<sup>1</sup>, 宋述慧<sup>2</sup>

1.中国科学院北京基因组研究所重点实验室, 北京 101001;

2.中国科学院北京基因组研究所级中心, 北京 101001;

3.中国科学院大学, 北京 100049

**摘要:** RNA 酶促共价修饰研究, 尤其是 m<sup>6</sup>A (6-甲基腺嘌呤), 是 RNA 生物学研究的一个新兴领域。m<sup>6</sup>A 是真核生物 mRNA 内部序列中最常见的一种转录后修饰形式, 由包含 3 个独立组分的复合物 RNA: m<sup>6</sup>A 甲基转移酶催化生成。最新研究发现肥胖相关蛋白 FTO 可以脱掉 m<sup>6</sup>A 上的甲基, 表明该甲基化过程是可逆的。抑制或敲除 m<sup>6</sup>A 甲基转移酶会引起重要的表型变化, 但是由于过去的检测方法受限, m<sup>6</sup>A 确切的作用机制目前为止还不甚清楚。二代测序技术结合免疫沉淀方法为大规模检测 m<sup>6</sup>A 修饰并研究其作用机制提供了可能。文章主要综述了 m<sup>6</sup>A 的发现史、生成机制、组织和基因组分布、检测方法、生物学功能等及其最新研究进展, 并通过比较 3 种 IP-seq 技术和数据分析的异同及优缺点, 对 m<sup>6</sup>A 这种 RNA 表观修饰研究中尚未解决的问题进行了讨论。

**关键词:** RNA 修饰; 6-甲基腺嘌呤; IP-seq

## Recent progresses in RNA N<sup>6</sup>-methyladenosine research

LI Yu-Li<sup>1,3</sup>, YU Jun<sup>1</sup>, SONG Shu-Hui<sup>2</sup>

1. CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China;

2. Core Genomic Facility, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** RNA modifications, especially methylation of the N<sup>6</sup> position of adenosine (A)—m<sup>6</sup>A, represent an emerging research territory in RNA biology. m<sup>6</sup>A is a post-transcriptional modification of RNAs, which is catalyzed by the RNA:m<sup>6</sup>A methyltransferase complex containing three individual components and is the most common form found in the internal sequences of mRNAs in eukaryotes. Latest study showed that the fat mass and obesity-associated protein could remove the methyl group, indicating that the modification is reversible. Importantly, inhibiting or silencing the methyltransferase will cause significant changes of phenotypes. However, due to limited detection methods, the mechanism of m<sup>6</sup>A has not been figured out yet. Next-generation sequencing combining with IP (immunoprecipitation) technologies makes it possible to detect m<sup>6</sup>A modifications in a large scale. Here, we reviewed recent progresses of m<sup>6</sup>A studies including the discovery of m<sup>6</sup>A, mechanism of biosynthesis, tissue and genome distribution, detection methodology and possible biological functions. We also compared three IP-seq technologies that are currently widely used, and summarized the challenges in m<sup>6</sup>A studies.

**Keywords:** RNA modifications; m<sup>6</sup>A; IP-seq

收稿日期: 2013-04-28; 修回日期: 2013-08-25

**基金项目:** 国家自然科学基金项目 (编号: 30900831) 和北京市科技新星计划 (编号: Z121105002512060) 资助

**作者简介:** 李语丽, 博士研究生, 研究方向: 生物信息学。Tel: 010-84097620; E-mail: liyl@big.ac.cn

**通讯作者:** 于军, 研究员, 研究方向: 生物信息学。E-mail: junyu@big.ac.cn

宋述慧, 副研究员, 研究方向: 生物信息学。Tel: 010-84097620; E-mail: songshh@big.ac.cn

RNA 同时具有调控和信息分子的双重功能，在众多细胞机制中发挥着核心的作用，例如作为遗传信息的信使，催化生化反应，作为细胞结构骨架等<sup>[1]</sup>。RNA 转录后的修饰为 RNA 功能的多样化奠定了化学基础，2011 年更新的 RNA 修饰数据库 RNAMDB 共收录了 109 种 RNA 的修饰形式，其中甲基化修饰占 80%<sup>[2]</sup>。6-甲基腺嘌呤 ( $N^6$ -methyladenosine,  $m^6A$ )，是发生在碱基 A 第六位 N 原子上的甲基化，作为真核生物中最常见的一种 RNA 转录后修饰，近年来得到了广泛的关注和研究。本文综述了  $m^6A$  的发现历史、生成机制、分布特点、生物学功能、检测方法、IP-seq 技术的发展以及对未来的展望。

## 1 $m^6A$ 的发现

$m^6A$  最初在细菌 DNA 中发现。1955 年 Dunn 等<sup>[3]</sup>在观察胞嘧啶(T)的结构类似物对需 T 菌株 *Bacterium coli* 15 T 的影响时，发现 DNA 复制过程中出现了一种数量可观的新碱基。将 DNA 水解后利用二维纸层析进行碱基分离时，除了得到 4 种常见的碱基 (A、T、G、C) 外，还发现了第 5 个紫外吸收光点 X。经多重实验鉴定，对应于该吸收光点的新碱基为 6-methylaminopurine (即  $N^6$ -methyladenosine,  $m^6A$ )。同时，在其他一些菌株和噬菌体中也检测到了  $m^6A$  的微量存在<sup>[4]</sup>。

1958 年，Dunn 等<sup>[5]</sup>在研究 *B.coli* RNA 中的非自然碱基时也同样分离到了  $m^6A$ ，并同时在多个物种中进行了检测和定量，包括细菌、酵母、病毒、小麦的胚芽、兔子的肝脏及小鼠肝脏的微粒体等。有趣的是，小麦胚芽中  $m^6A$  的含量尤其突出<sup>[5]</sup>。1974 年，Desrosiers 等<sup>[7]</sup>借助真核生物中 Poly(A)结构的发现<sup>[6]</sup>研究了 Novikoff 肝癌细胞中 mRNA 的甲基化情况。研究发现 mRNA 中核糖的甲基化和碱基的甲基化各占一半，核糖的甲基化比较均衡地分布在 4 个自然核苷中，而碱基的甲基化中主要的修饰形式则为  $m^6A$  (比例约为 80%)。rRNA 和 tRNA 中碱基修饰形式的种类比 mRNA 要复杂的多。另外，在多种高等真核生物<sup>[8~13]</sup>以及病毒<sup>[14~16]</sup>的 mRNA 中都检测到  $m^6A$  的存在。 $m^6A$  如此广泛的存在让我们很难忽视它的生物学意义和重要性，从而进行更深入的研究。

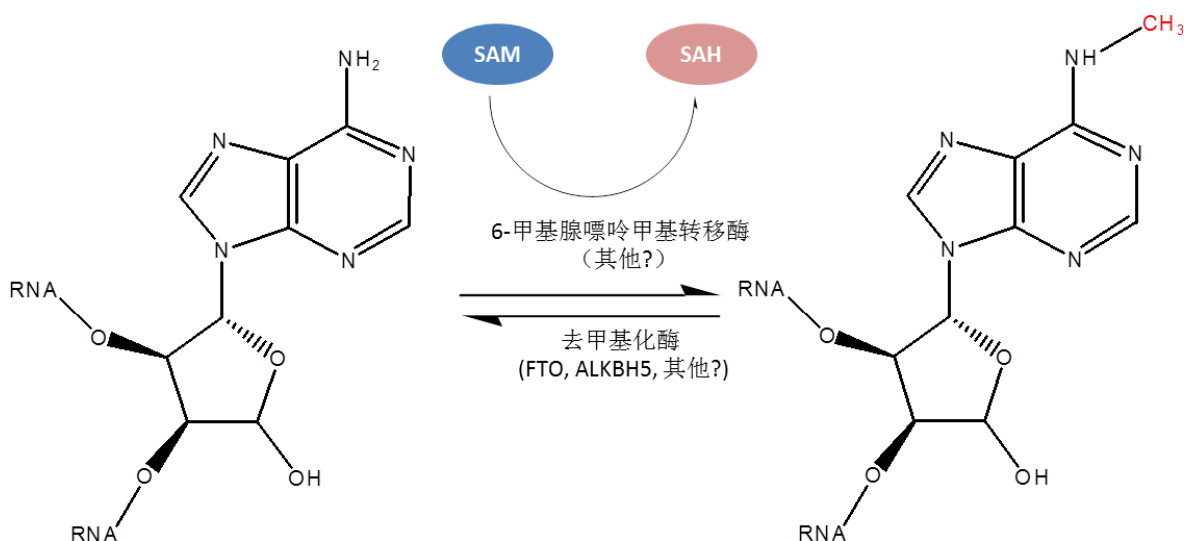


图 1 mRNA 腺嘌呤甲基化/去甲基化的酶促反应

$m^6A$  甲基转移酶将甲基供体 S-腺苷甲硫氨酸 (SAM) 上的甲基转移到 mRNA 腺嘌呤的第六位 N 原子上，SAM 脱甲基转变为 S-腺苷高半胱氨酸 (SAH)；而甲基化之后的腺嘌呤也能够去甲基化酶，如 FTO、ALKBH5 等的催化下脱掉甲基。

## 2 m<sup>6</sup>A 的生成

虽然在 1958 年就发现了 RNA 中 m<sup>6</sup>A 的存在,但是其生成的酶学研究相对滞后。直到 1988 年 Narayan 和 Rottman 发展了一套 m<sup>6</sup>A 体外甲基化系统,为 m<sup>6</sup>A 生成的酶学研究奠定了重要的基础<sup>[17]</sup>。1992 年首次部分分离纯化到了催化这一过程的甲基转移酶<sup>[18]</sup>,Martin 在 HeLa 细胞中以小鼠二氢叶酸还原酶的 mRNA 为甲基化的底物在 HeLa 细胞的细胞核提取物中分离到了两种形式的酶,体外实验证实它们都能催化这一反应。关于酶的亚细胞结构定位,其主要分布在细胞核内,在核糖体也有少量存在,而在其他亚细胞结构中未检测到酶活。碱基 A 在甲基转移酶的催化下生成 m<sup>6</sup>A,其反应式见图 1。1994 年 Bokar 等<sup>[19]</sup>分离到了酶的 3 个独立组分 MT-A1、MT-A2 和 MT-B,其中 MT-A1 是最小的组分,只有 30 kDa,其功能未知;MT-A2 的分子量为 200 kDa,在它的 70 kDa 的亚基上含有甲基供体 S-腺苷甲硫氨酸的结合位点;而 MT-B 是酶复合物中最大的组分,其分子量为 875 kDa,实验证明其可能含有 mRNA 的结合位点。70 kDa 的亚基 MT-A70 被分离纯化后通过微测序得到氨基酸的序列,进一步得到该亚基的 cDNA 克隆<sup>[20]</sup>。MT-A70 的预测蛋白有 580 个氨基酸,其中含有原核生物 DNA 的 m<sup>6</sup>A 甲基转移酶所共有的甲基化 motif 1 和 motif 2 的相似序列,说明其甲基化的功能在进化中具有很强的保守性<sup>[20]</sup>。虽然之后的研究对 MT-A70 有了较深入的了解,但是对该酶所在的复合物的其他组分还知之甚少。

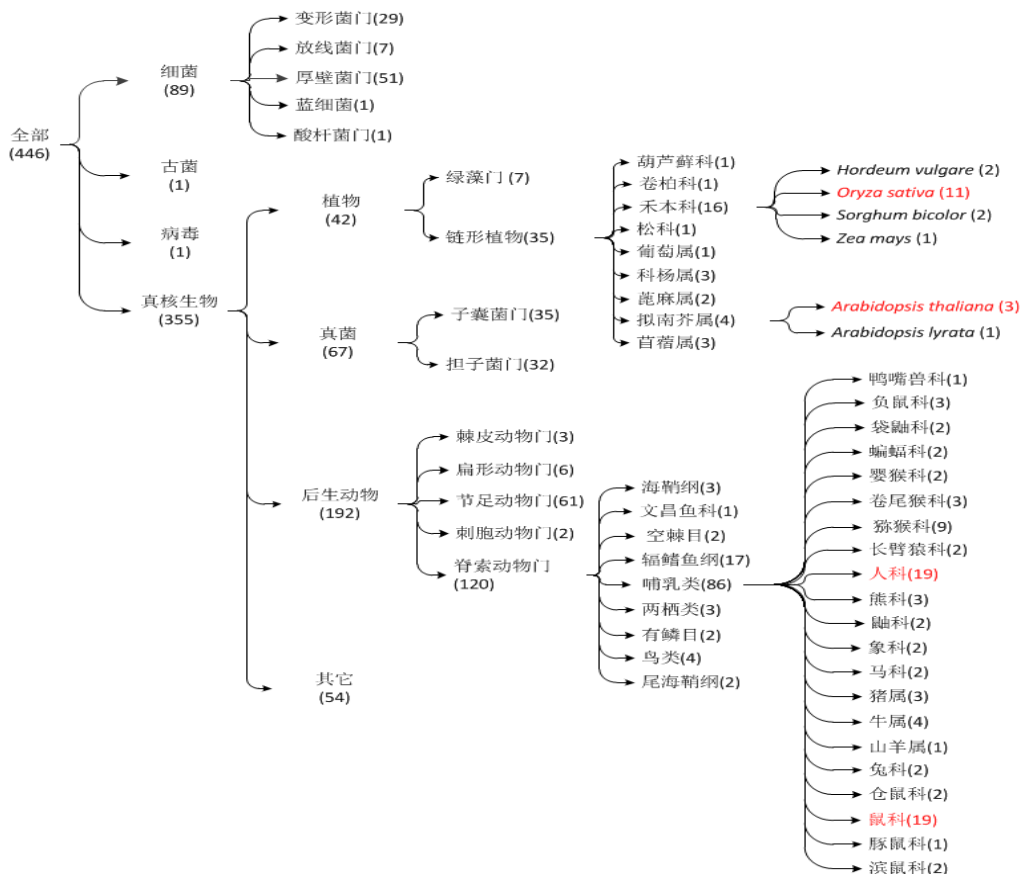


图 2 人 MT-A70 同源基因在整个生物界中的分布

MT-A70 的同源基因广泛存在于细菌、古菌、病毒及真核生物界,而约有 80% 的序列分布在真核生物中,其中在人、小鼠、拟南芥和水稻等重要的高等生物中都有存在(红色字体标注)。

目前已知编码人 MT-A70 蛋白的 *METTL3* 基因位于 14 号染色体上。此基因共有 17 条转录本,其中

有蛋白产物的有 9 条。已验证具有 mRNA:m<sup>6</sup>A 甲基转移酶活性的转录本，长度为 2 015 bp，其基因含有 11 个外显子，共编码 580 个氨基酸，对应于 MT-A70 功能域的氨基酸位置为 389~550。搜索该蛋白的同源基因（将此蛋白序列提交到 HMMER 数据库（<http://hmmer.janelia.org/>），设置 significance e-value 为 1e-10），共得到 446 条同源序列，其在各个物种中的分布见图 2。由图可知，其在细菌、古菌、病毒及真核生物界广泛存在，其中约有 80% 的序列分布在真核生物中。早在 2002 年，Bujnicki 等<sup>[21]</sup>已经对 MT-A70 进行了系统进化分析，并将该蛋白家族分为 4 个亚家族，按字母排序将其分别命名为 A、B、C、D，它们有相似的 motif 和一个较清晰的共同祖先。其中，进化分支 A、B 和 C 分别包含人 MT-A70 蛋白以及从人的大脑和睾丸组织中分离的 cDNA 克隆预测的两个蛋白的直系同源基因，因此这 3 个分支的同源基因只存在于真核生物中。分支 D 是距离最远的，它含有一小类细菌 DNA 中 m<sup>6</sup>A 的甲基转移酶，这一类酶参与限制/修饰系统。同时，本文也对这 446 条同源序列的功能域的结构进行归类，可分为 12 类（图 3），其中超过 90% 的序列只含有一个 MT-A70 domain，剩余不到 10% 的序列含有 2 个或多个 MT-A70 结构域或混杂其他的结构域，表明 MT-A70 蛋白在进化过程中是极其保守的。根据以上的研究结果，可以推测 RNA 中 m<sup>6</sup>A 的碱基修饰在真核生物的生命活动中扮演着重要角色。

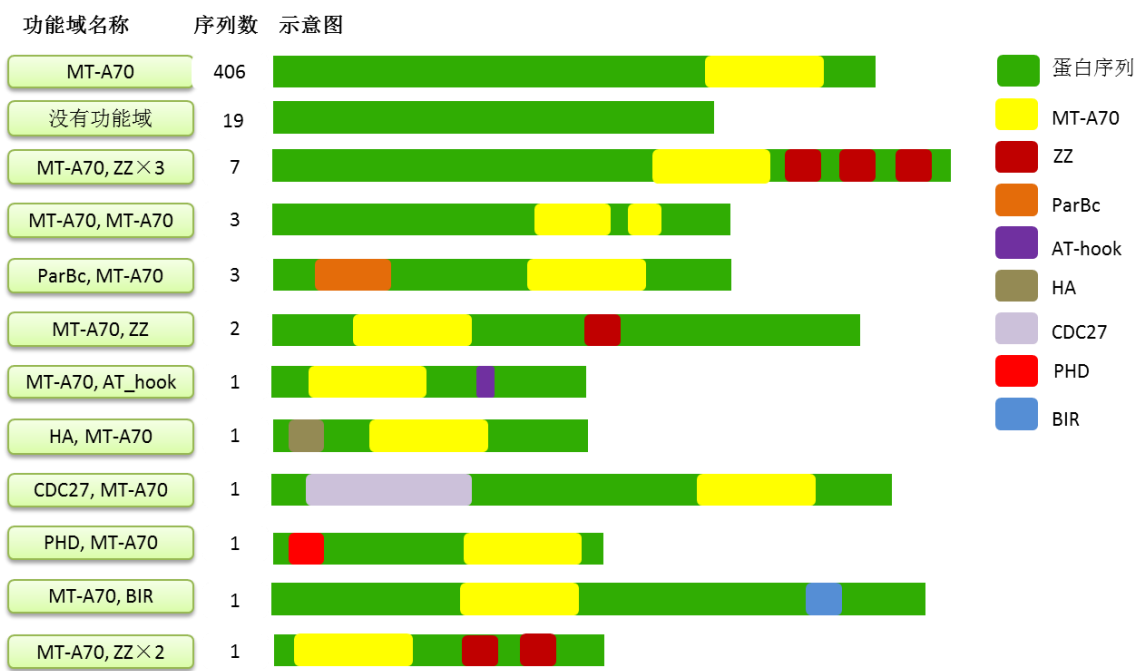


图 3 MT-A70 蛋白的功能域结构分类

MT-A70 的同源蛋白的 446 条序列中超过 90% 的序列只含有一个 MT-A70 结构域，剩余不到 10% 的序列含有 2 个或多个 MT-A70 结构或混杂其他的功能域，表明 MT-A70 蛋白在进化过程中是极其保守的。

虽然对 RNA:m<sup>6</sup>A 甲基转移酶其中一个亚基 MT-A70 的序列结构及系统进化有了较清晰的了解，但是对整个酶的催化机制还不甚清楚。关于该酶所识别的 RNA 的特异序列已有众多研究，但是目前为止只在两个基因——*Rous sarcomavirus* 和牛的催乳素 *bPRL* 的 mRNA 中找到了精确的甲基化位点<sup>[22, 23]</sup>。*Rous sarcomavirus* 共有 7 个甲基化位点，绝大多数的位点出现在 GGACU 序列中，少数在 UGACU 和 AGACU 序列中；而 *bPRL* mRNA 的甲基化修饰只发生在 AGACU 序列中，以上这几种序列可归结为 RRACH

(R=purine, H=A, C 或 U)<sup>[24]</sup>。Rottman 等<sup>[25]</sup>将共有序列的核心序列 GAC 在 bRPL 中进行复制发现该位点完全不能甲基化, 而且发现 m<sup>6</sup>A 甲基化的频率远小于共有序列的频率, 因而推断要确定一个真正的 m<sup>6</sup>A 位点除了共有序列可能还需要茎环、圆环等特定二级结构的存在。近期的研究<sup>[26]</sup>将共有序列进一步延长并找到了更多的变体, 可见酶结合的 RNA 序列呈现一定的多态性, 或者说特异性不强, 因此猜测除了可能会存在二级结构外, 还有可能是存在其他的基因或酶来共同催化这一过程<sup>[27]</sup>。

与 DNA 和组蛋白的甲基化/去甲基化的可逆过程一样, m<sup>6</sup>A 的甲基化也是一个动态的过程, 即细胞中同时存在甲基化酶和去甲基化酶来共同调控这一可逆过程<sup>[28]</sup>。目前为止已报道的编码去甲基化酶的基因因为 AlkB 家族的两个成员: *FTO*(fat mass and obesity-associated)和 *ALKBH5*(Homo sapiens alkB, alkylation repair homolog 5 (*E. coli*)), 实验证明这两个基因被敲除后 m<sup>6</sup>A 的水平都会显著提高<sup>[29, 30]</sup>。已知人类肥胖相关基因 *FTO* 的过表达会导致肥胖、糖尿病以及增加老年痴呆症的风险, 但是其作为去甲基化酶发挥作用的机制研究较少<sup>[31-36]</sup>; *ALKBH5* 是一种 mRNA 结合蛋白, 敲除后会导致小鼠生精小管细胞中 mRNA 的 m<sup>6</sup>A 修饰水平增高, 进而导致睾丸萎缩和精子数量及活力的骤减, 从而表明 *ALKBH5* 能够通过 mRNA 中 m<sup>6</sup>A 的去甲基化调控精子的形成<sup>[29]</sup>。RNA m<sup>6</sup>A 的甲基化/去甲基化机制目前仍在研究中, 但是参与此过程的 (去) 甲基化酶在进化过程中往往具有高度的保守性, 普遍分布在各个物种中, 猜测其表达调节机制也可能类似。

### 3 m<sup>6</sup>A 的分布特点

#### 3.1 m<sup>6</sup>A 的组织分布特点

由于 m<sup>6</sup>A 绝大部分分布的多细胞真核生物中, 它的分布是否有组织特异性或偏好性呢? 为此人们在植物和动物中分别进行了相关的研究和分析。在对植物 m<sup>6</sup>A 的研究中, Zhong 等<sup>[37]</sup>利用放射性标记和二维 TLC 方法, 首次对拟南芥中不同组织 (根、叶片及花) 的 m<sup>6</sup>A 进行了定量分析。分析显示, 在 3 个组织中, 花中 m<sup>6</sup>A 的含量最高, 其 m<sup>6</sup>A:A 的比例值为 1.4%, 而在叶片和根中分别为 0.9% 和 0.6%。同时, 实验检测 MTA (拟南芥 MT-A70 同源蛋白) 在花芽中的表达量也是各个组织中表达最高的。在哺乳动物中, m<sup>6</sup>A 也广泛分布在多个组织中。Meyer 等<sup>[26]</sup>的研究显示, 在肝脏、肾脏及大脑中 m<sup>6</sup>A 的表达要高于其他组织, 如心脏、肺等。另外他们还发现在不同的癌细胞系中, m<sup>6</sup>A 的含量也有很大不同。可见, 不论在植物还是动物中, m<sup>6</sup>A 的分布首先是具有普遍性, 分布在多个组织中; 同时也具有偏好性, 即在某些组织中 (植物的分生组织, 动物的肝脏、肾脏及大脑等) 分布要明显高于其他组织。

#### 3.2 m<sup>6</sup>A 的基因组分布特点

借助高通量测序技术, 已经得到粗略的 m<sup>6</sup>A 修饰谱。Meyer 等<sup>[26]</sup>通过研究小鼠大脑中 m<sup>6</sup>A 表观修饰, 发现其主要分布在基因内部 (94.8%), 其中在 CDS、UTRs、intron 中的比例分别为 50.9%、41.9% 和 2.0%, 而 UTRs 区域的 m<sup>6</sup>A 更倾向于富集在 3' UTRs, 见图 4a。如将所有 m<sup>6</sup>A 峰值在 mRNA 中的百分位位置的分布进行统计, 发现在 CDS 的末端区域其百分比迅速上升, 约为 CDS 起始区域的 5~6 倍, 而随着 3' UTRs 的延长, m<sup>6</sup>A 的含量迅速下降; 在 CDS 的上下游各 1 kb 的统计, 发现 CDS 区域的 m<sup>6</sup>A 甲基化主要富集在终止密码子的附近。由此可知, m<sup>6</sup>A 在基因组上的分布尤以 mRNA 的终止密码子和 3'UTR 区最为富集。本文利用文献<sup>[26]</sup>中的一组数据 (GSM854223 和 GSM854224), 按照文章所述的方法进行了测

试，得到的峰值在 mRNA 上的分布见图 4b，与上述结论基本一致。而 Dominissini 等<sup>[38]</sup>的研究则表明，m<sup>6</sup>A 修饰位点除了富集在终止密码子周围外，其峰值还倾向于出现在基因内部的长外显子中：落在基因内部外显子的 2 838 个峰值中，87%的峰值落在长度大于 400 bp 的外显子中。

m<sup>6</sup>A 在基因组上出现特定区域富集的特点与 DNA 甲基化修饰集中分布的特点一致，已知 DNA 甲基化对基因表达具有重要的调控作用，而 m<sup>6</sup>A 作为一种转录后的修饰形式，有可能会在基因的转录后调控中发挥重要作用<sup>[10, 39]</sup>。

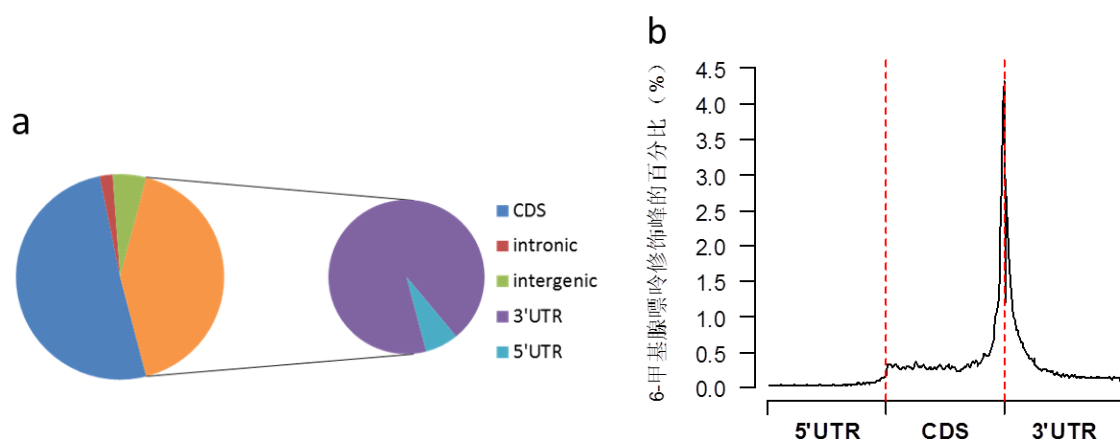


图 4 小鼠大脑中 m<sup>6</sup>A 修饰在基因组上的分布特点<sup>[26]</sup>

a: m<sup>6</sup>A 在基因内部各个区域包括 CDS、intron、UTRs，分布比例分别为 50.9%、41.9%和 2.0%，而 UTRs 区域的 m<sup>6</sup>A 更倾向于富集在 3' UTRs；b: 将找到的 m<sup>6</sup>A 的修饰峰定位到基因的各个位置（CDS、intron、UTRs）上，并转换成百分位的位置，作图结果显示修饰峰主要富集在 CDS 末端和 3' UTR 的起始位置，与图 4a 的结果一致。

#### 4 m<sup>6</sup>A 的生物学功能

表观遗传学是在研究与经典孟德尔遗传法则不相符的生命现象过程中逐步发展起来的一门学科，又称为“拟遗传学”、“表遗传学”、“外遗传学”以及“后遗传学”，主要研究在细胞核 DNA 没有改变的情况下，基因功能的可逆的、可遗传的改变。DNA 的甲基化是表观遗传学的一种机制，它通过特定的甲基化酶将甲基加在 DNA 分子上，如胞嘧啶环的 5' 碳上，甲基化之后的 DNA 序列在复制、转录及翻译的过程中都不会发生改变，但是却会改变最终的遗传表现。甲基化修饰同样也发生在 RNA 水平上，那么 RNA 的甲基化，尤其是人们关注的 m<sup>6</sup>A 修饰会有哪些生物学功能呢？

MT-A70 还未被纯化之前，研究 m<sup>6</sup>A 功能主要有两种思路：一是把已鉴定到的主要的甲基化位点突变后观察其对细胞的影响；二是加入甲基化的抑制剂，主要为甲基供体 S-腺苷甲硫氨酸的类似物，然后观察其引起的效应。Csepany 等<sup>[24]</sup>和 Kane 等<sup>[40]</sup>在病毒 RSV 中的研究采用了第一种思路，相比于野生型，已知的 m<sup>6</sup>A 甲基转移酶的底物 DHFR(编码二氢叶酸还原酶)在 m<sup>6</sup>A 位点被突变后会降低翻译的效率<sup>[41]</sup>。已知的甲基化抑制剂主要有三类，分别是环亮氨酸、同型半胱氨酸及腺苷类似物。用环亮氨酸处理 CHO 细胞会导致 mRNA 甲基化大幅度减少<sup>[42]</sup>；而处理被 B77 鸟肉瘤病毒感染的细胞，会使次基因组 env 的 RNA 与基因组 RNA 的比值发生变化<sup>[43]</sup>。腺苷类似物的加入会导致未剪切的 pre-mRNA 在细胞核内的累积。同型半胱氨酸作为 S-腺苷甲硫氨酸依赖的甲基转移酶反应的强效抑制剂，能够显著降低 m<sup>6</sup>A 的水平，



延迟 mRNA 在胞质中的出现时间，从而推测 m<sup>6</sup>A 在 mRNA 的加工及转运过程中发挥作用<sup>[44]</sup>。由于已经验证的甲基化位点非常少，因此第一种思路的开展比较受限。而第二种思路的特点是从化学的角度改变甲基化的化学反应，并观察可能引起的细胞生物学过程的改变，其缺陷在于 S-腺苷甲硫氨酸是细胞内众多甲基转移酶的甲基供体，因此改变化学反应中 S-腺苷高半胱氨酸与 S-腺苷甲硫氨酸的平衡，除了 m<sup>6</sup>A 甲基转移酶，还会影响到其他的甲基化酶的反应，因此很难区分不同酶反应导致的效应<sup>[45]</sup>。

甲基转移酶腺苷甲硫氨酸结合亚基的克隆，为研究不同物种中 MT-A70 的作用机制以及 m<sup>6</sup>A 修饰本身的调节作用奠定了重要的基础。1997 年 Bokar 首次在 Hela 细胞中克隆到甲基转移酶腺苷甲硫氨酸结合亚基的全长 cDNA，即 MT-A70 亚基，并证实了这个亚基是甲基转移酶发挥作用的一个非常重要的组分<sup>[20]</sup>。酵母中的 *IME4* (Inducer of meiosis 4) 基因与 *MT-A70* 基因有较高的同源性，诱导酵母的出芽生殖，而出芽的激活可能来源于该蛋白的甲基转移酶活性，间接提示了 m<sup>6</sup>A 的碱基修饰具有调控酵母出芽生殖的生理功能<sup>[45]</sup>。*IME4* 的同源基因 *Dmime4* 在果蝇的卵巢和睾丸中表达，表明它在配子生成和发育过程中功能保守，同时还发现 *Dmime4* 基因靶向调节后生动物发育的多个方面的 Notch 信号通路<sup>[46]</sup>。*METTL3* 和 *FTO* 在细胞核颗粒中与剪切蛋白具有共定位关系，而多亚型基因的甲基化与差异剪切的外显子及内含子的相关性也表明 m<sup>6</sup>A 在 RNA 剪切调控中的作用<sup>[20, 30]</sup>，甲基化与“剪切密码”的整合可能会提高对可变剪切事件的预测能力<sup>[38]</sup>。同样在植物界的多个物种中也找到了 *MT-A70* 的同源基因，如拟南芥中的 *MTA*，将 *MTA* 突变或沉默后，会引起植株的多种发育缺陷，如在胚胎发育中不能由球形时期过渡到心形时期、叶子褶皱、花序变短、顶端优势降低以及毛状体分支增加<sup>[37]</sup>，总之，m<sup>6</sup>A 在多细胞真核生物中具有重要作用。

将甲基转移酶突变或沉默来观测表型的变化在一定程度上能说明酶的功能或 m<sup>6</sup>A 修饰的作用，但是就目前的研究现状来看，并不能完全确定以下两点。一方面，因为甲基转移酶是一个非常复杂的复合物，并不能确定它是否只具备添加甲基的功能，所以单纯的把它沉默掉并不能说明表型的改变是由 m<sup>6</sup>A 修饰的变化引起的。另一方面，该方式不能获悉这个酶的靶基因到底是什么，也就是说研究者并不清楚敲除掉这个酶之后会引起哪些基因甲基化的改变，而这些改变又会影响到哪些生物学过程。无疑，只有通过最直接的方式才能解决这些问题，即发生 m<sup>6</sup>A 修饰的基因到底有哪些，具体的修饰位点是什么，在不同状态下这些基因的甲基化水平会怎么变化，进而推测这些变化所带来的生物学意义。但是这种最直接的方式给我们的实验技术提出了巨大的挑战。目前借助二代测序技术，已经能够抓取到发生 m<sup>6</sup>A 修饰的基因，如人的肝癌细胞系 HepG2 中 7 000 多个基因上存在着超过 12 000 个甲基化位点<sup>[38]</sup>，但是对每个基因精确的甲基化位点以及转录组中的甲基化水平（有多少基因的多少位点）和程度（多少个可甲基化的位点被甲基化）还不甚明确，因此需要进行更深入的思考，也需要更先进的检测技术和更严谨的数据分析方法来做支撑。

## 5 m<sup>6</sup>A 的检测方法及 IP-seq 技术概述

由于被甲基化的碱基 A 并没有改变它的碱基配对能力，目前还不能通过直接测序来检测，也没发现它有独特的化学反应。如，DNA 中的 5mC 经过 bisulfite 处理后会发​​生碱基变化从而可以通过测序检测。所以之前的检测方法一直是基于放射性标记技术，即将甲基供体 S-腺苷甲硫氨酸进行放射性标记，并与待处理的细胞进行孵育，随后通过 TLC 或 HPLC 来检测<sup>[30, 47, 48]</sup>。但是这种方法既昂贵又繁琐，效率又低，

而且无法实现大规模的基因检测及确切的甲基化位点的定位。借助二代测序技术通量高、快速高效的特点，研究人员发展了一种能够快速、高效地检测全基因组  $m^6A$  修饰谱的方法——MeRIP-seq<sup>[26]</sup>，与之类似方法还有  $m^6A$ -seq<sup>[38]</sup>，它们的实验原理基本一致，主要差别在 MeRIP-seq 研究的是总 RNA 水平上的  $m^6A$ ，而  $m^6A$ -seq 研究的是信使 RNA (mRNA) 水平上的碱基修饰。下面主要以 MeRIP-seq 为例简单介绍这种方法的实验原理及数据分析方法，并从 IP-seq 的视角来比较说明 MeRIP-seq 与我们已经熟知的 ChIP-seq<sup>[49, 50]</sup>、PAR-CLIP<sup>[51, 52]</sup>等 IP-seq 技术的异同，进而就 IP-seq 技术检测  $m^6A$  修饰尚存在的问题及挑战进行讨论。




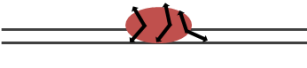
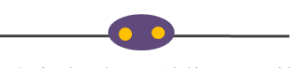



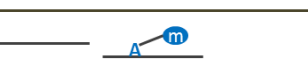
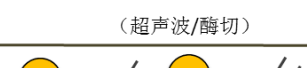
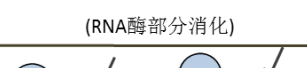
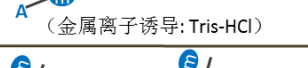



技术	ChIP-seq	PAR-CLIP	MeRIP-seq
原始序列	 DNA 结合蛋白 DNA	 RNA 结合蛋白 RNA	 $m^6A$ mRNA
步骤1 交联	 (促交联剂:甲醛等化学试剂)	 (促交联方式: 4SU 培养, UV照射)	 (无交联反应)
步骤2 片段化	 (超声波/酶切)	 (RNA酶部分消化)	 (金属离子诱导: Tris-HCl)
步骤3 免疫沉淀	 (抗DNA结合蛋白的抗体)	 (抗RNA结合蛋白的抗体)	 (抗m <sup>6</sup> A的抗体)
步骤4 反交联	 (洗脱磁珠; 蛋白酶 消化DNA 结合蛋白)	 (RNA酶再次消化; SDS-PAGE 分离;洗脱磁珠; 蛋白酶消化RNA结合蛋白)	 (洗脱磁珠)
步骤5 测序	测序读段		

图5 3种 IP-seq 技术的实验处理及操作方法的异同

### 5.1 MeRIP-seq 技术的实验原理

免疫沉淀法 (Immunoprecipitation, IP) 与二代测序技术 (Next generation sequencing, NGS) 结合起来研究核酸/脱氧核酸和蛋白质相互作用的技术称之为 IP-seq 技术，其中已知的 ChIP-seq、PAR-CLIP 等发展较为完善，应用较为成熟。MeRIP-seq 作为一种新的 IP-seq 技术，目前已被成功地用于检测全基因组的  $m^6A$  修饰。其主要的实验原理为：利用免疫共沉淀的方法即用抗  $m^6A$  碱基的抗体与被随机打断的 RNA 片段进行孵育，抓取有  $m^6A$  修饰的片段并进行测序；同时平行测序一个对照样本以此消除抗体抓取过程中的背景<sup>[26]</sup>。与 ChIP-seq、PAR-CLIP 技术相比 (图 5)，它们的共同点都是将免疫沉淀法与二代测序技术结合起来检测全基因组 (或全转录组) 范围内的结合区域或修饰区域。不同点主要体现在以下几个方面：1) ChIP-seq 与 PAR-CLIP 是检测蛋白与核酸相互作用的区域，通过交联与反交联的技术实现，而 MeRIP-seq 是直接检测 RNA 上修饰碱基的区域；2) ChIP-seq 与 PAR-CLIP 的特异性抗体所要抓取的抗原是 DNA/RNA 结合蛋白，而 MeRIP-seq 中的抗体抓取的是 RNA 上被甲基化了的碱基，如  $m^6A$ ；3) 由于 MeRIP-seq 不能像 CLIP 那样可以在实验中用 RNA 酶来消化非特异结合的 RNA 片段<sup>[53]</sup>，所以需要



同时测定一个对照样本，来去除由于抗体的非特异性结合所引入的噪音信息。因此，MeRIP-seq 较前两种 IP-seq 技术在实验步骤上更为简单快捷。

## 5.2 MeRIP-seq 技术的数据分析方法

如何在海量的测序数据中找到  $m^6A$  的修饰区域呢？首先 RNA 的随机片段化保证了修饰位点的高覆盖率，然后在高分辨率下以小范围的固定窗口在基因组上扫描，比较 IP 样本和对照样本中落在每个窗口里的读段数，并通过统计学的方法计算  $p$  值和 FDR，然后将具有显著差异的窗口连结起来并筛选合适长度的连续窗口即为  $m^6A$  修饰的 peak 区，基于此思路开发的软件 MeRIPPeR<sup>[26]</sup>和 MeRIP-PF<sup>[54]</sup>简单易用，能够较准确的找到  $m^6A$  peaks 所在的区域，方便研究者在全转录组范围内进行 RNA 表观遗传学的探索和研究。MeRIPPeR 和 MeRIP-PF 软件的差别主要在比对软件 BWA 的版本以及单一比对的读段的定义上，因此所定义的修饰峰也稍有不同<sup>[54]</sup>。

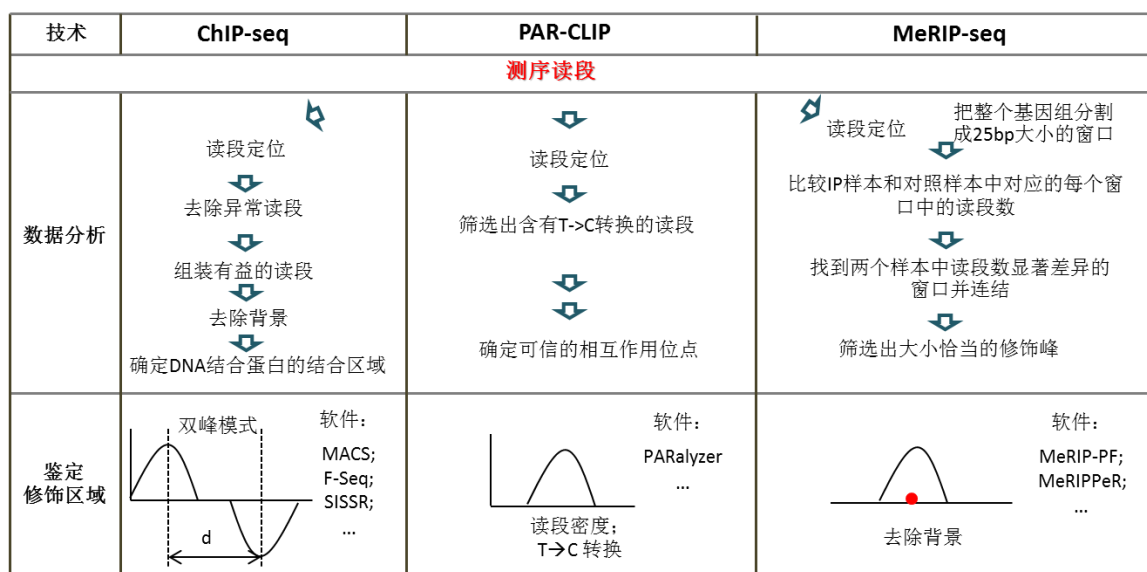


图 6 3 种 IP-seq 技术的数据分析流程及分析方法的比较<sup>[49,52,54]</sup>

与 ChIP-seq、PAR-CLIP 相比（图 6），它们处理数据的共同点在于都需要去除背景噪音，比如由抗体的非特异性结合反应或抗体的亲和力较差以及测序错误等因素引入的背景或误差。不同点主要有 3 个方面：1）ChIP-seq 和 MeRIP-seq 数据都会平行测定一个对照样本，通过与对照样本的比较，设定相应的参数，筛选出具有显著性差异的 peaks，而 PAR-CLIP 由于应用了 4-硫尿核苷（4SU），一方面加强了 RNA 和 RNA 结合蛋白的交联效应，另一方面使得交联的序列发生了 T 到 C 的转变，因此可以借助这一突变并结合测序读段的密度来很好地鉴定结合区域；2）前两种技术主要检测的是 DNA/RNA 与蛋白的结合区域，是几十甚至几百个碱基的区段，而 MeRIP-seq 需要检测的是单个的修饰位点，但是目前检测方法的分辨率还停留在修饰区域上，只能通过寻找共同的 motif 来确定修饰位点，因此要想通过最直接的途径确定修饰位点，就必须对实验技术进一步改进；3）前两种技术只涉及到是否有结合的问题，而不涉及量的问题，但是，由于  $m^6A$  修饰发生在 RNA 水平上，就需要考虑甲基化水平的问题，即基因转录的 mRNA 是否完全被甲基化或者说被甲基化的转录本的比例是多少。

## 5.3 检测 $m^6A$ 修饰谱的 IP-seq 技术所面临的问题及挑战

虽然借助于 MeRIP-seq 或 m<sup>6</sup>A-seq 技术已经能够给出全转录组的 m<sup>6</sup>A 的修饰谱,但是这种技术仍存在一些问题,主要有以下几个方面:1)采用不同的数据分析方法,所得结果会有很大的差异;2)序列比对软件的不同会导致找到的峰的数目不同<sup>[55]</sup>。Saletore 等<sup>[55]</sup>分别利用 BWA、TopHat 2 以及 GSNAP3 种软件做序列比对,基于 MeRIP-seq 的方法,控制相同的 FDR (0.05),找到的峰的数目分别为 19 617、135 706 和 45 738;3)测序的深度也会严重影响到峰的数目,体现在峰的数目会随着测序深度的增大而增加。究其原因可能是,对于一条转录本其只有一定比例的拷贝会被甲基化,而在测序深度较低时,修饰比例低的转录本无法被检测到。因此如何高效准确的定义一个真正的修饰峰仍是目前的一个挑战。

## 6 展 望

MeRIP-seq 技术能够在全转录组范围内定位有 m<sup>6</sup>A 修饰的基因,使人们对 m<sup>6</sup>A 碱基修饰的认知有了较大的飞跃,但是目前仍然面临着几个重要的问题需要解决。第一,甲基化的催化机制。关于甲基化的催化机制,由于纯化到的甲基转移酶是个非常复杂的复合物,目前除了识别亚基 MT-A70,整个酶复合物还不甚清晰。第二,如何突破技术瓶颈,达到可以直接检测 m<sup>6</sup>A 修饰位点的水平。第三代测序技术已经能够实现单分子实时测序,为直接检测 m<sup>6</sup>A 位点提供了一个好的技术平台,但是目前的第三代测序仪测序成本高、测序错误率较高,而且直接以 RNA 为模版进行测序还存在众多问题<sup>[55, 56]</sup>,因此要真正实现单碱基分辨率的检测还有待测序技术的进一步提高;第三,根据目前的研究,还很难确定出高度一致的甲基化基序,因此有理由猜想是否还存在其他的甲基转移酶来参与 m<sup>6</sup>A 修饰<sup>[27]</sup>;最后,探索有 m<sup>6</sup>A 修饰的基因的功能,这些被修饰基因所引起的改变也是生物体表型发生变化的直接原因,因此有必要发展实验手段对特定基因的修饰位点进行功能研究<sup>[27]</sup>。

## 参考文献(References):

1. Sharp PA. The Centrality of RNA. *Cell*, 2009, 136(4): 577-580.
2. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res*, 2010, 39(Database): D195-D201.
3. Dunn DB, Smith JD. Occurrence of a new base in the deoxyribonucleic acid of a strain of *Bacterium Coli*. *Nature*, 1955, 175(4451): 336-337.
4. Dunn DB, Smith JD. The occurrence of 6-methylaminopurine in deoxyribonucleic acids. *Biochem J*, 1958, 68(4): 627-636.
5. Littlefield JW, Dunn DB. Natural occurrence of thymine and three methylated adenine bases in several ribonucleic acids. *Nature*, 1958, 181(4604): 254-255.
6. Jelinek W, Adesnik M, Salditt M, Sheiness D, Wall R, Molloy G, Philipson L, Darnell JE. Further evidence on the nuclear origin and transfer to the cytoplasm of polyadenylic acid sequences in mammalian cell RNA. *J Mol Biol*, 1973, 75(3): 515-532.
7. Desrosie R, Frideric K, Rottman F. Identification of methylated nucleosides in messenger-RNA from Novikoff hepatoma-cells. *Proc Natl Acad Sci USA*, 1974, 71(10): 3971-3975.
8. Adams JM, Cory S. Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature*, 1975, 255(5503): 28-33.
9. Desrosiers RC, Friderici KH, Rottman FM. Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5' terminus. *Biochemistry-US*, 1975, 14(20): 4367-4374.
10. Dubin DT, Taylor RH. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic*

- Acids Res*, 1975, 2(10): 1653-1668.
11. Furuichi Y, Morgan M, Muthukrishnan S, Shatkin AJ. Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: m-7G(5')ppp(5')G-MpCp-. *Proc Natl Acad Sci USA*, 1975, 72(1): 362-366.
  12. Perry RP, Kelley DE, Friderici K, Rottman F. The methylated constituents of L cell messenger RNA: evidence for an unusual cluster at the 5' terminus. *Cell*, 1975, 4(4): 387-394.
  13. Haugland RA, Cline MG. Post-transcriptional modifications of oat coleoptile ribonucleic acids. 5'-Terminal capping and methylation of internal nucleosides in poly(A)-rich RNA. *Eur J Biochem*, 1980, 104(1): 271-277.
  14. Beemon K, Keith J. Localization of N6-methyladenosine in the Rous sarcoma virus genome. *J Mol Biol*, 1977, 113(1): 165-179.
  15. Moss B, Gershowitz A, Stringer JR, Holland LE, Wagner EK. 5'-Terminal and internal methylated nucleosides in herpes simplex virus type 1 mRNA. *J Virol*, 1977, 23(2): 234-239.
  16. Chen-Kiang S, Nevins JR, Darnell JE Jr. N-6-methyl-adenosine in adenovirus type 2 nuclear RNA is conserved in the formation of messenger RNA. *J Mol Biol*, 1979, 135(3): 733-752.
  17. Narayan P, Rottman FM. An *in vitro* system for accurate methylation of internal adenosine residues in messenger RNA. *Science*, 1988, 242(4882): 1159-1162.
  18. Tuck MT. Partial purification of a 6-methyladenine mRNA methyltransferase which modifies internal adenine residues. *Biochem J*, 1992, 288( Pt 1): 233-240.
  19. Bokar JA, Rath-Shambaugh ME, Ludwiczak R, Narayan P, Rottman F. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J Biol Chem*, 1994, 269(26): 17697-17704.
  20. Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA*, 1997, 3(11): 1233-1247.
  21. Bujnicki JM, Feder M, Radlinska M, Blumenthal RM. Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. *J Mol Evol*, 2002, 55(4): 431-444.
  22. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM. Mapping of N6-methyladenosine residues in bovine prolactin mRNA. *Proc Natl Acad Sci USA*, 1984, 81(18): 5667-5671.
  23. Kane SE, Beemon K. Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Mol Cell Biol*, 1985, 5(9): 2298-2306.
  24. Csepány T, Lin A, Baldick CJ Jr, Beemon K. Sequence specificity of mRNA N6-adenosine methyltransferase. *J Biol Chem*, 1990, 265(33): 20117-20122.
  25. Rottman FM, Bokar JA, Narayan P, Shambaugh ME, Ludwiczak R. N6-adenosine methylation in mRNA: substrate specificity and enzyme complexity. *Biochimie*, 1994, 76(12): 1109-1114.
  26. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 2012, 149(7): 1635-1646.
  27. Pan T. N6-methyl-adenosine modification in messenger and long non-coding RNA. *Trends Biochem Sci*, 2013, 38(4): 204-209.
  28. Niu Y, Zhao X, Wu YS, Li MM, Wang XJ, Yang YG. N(6)-methyl-adenosine (m(6)A) in RNA: an old modification with a novel epigenetic function. *Genomics Proteomics Bioinformatics*, 2013, 11(1): 8-17.
  29. Zheng GQ, Dahl JA, Niu YM, Fedoresak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, Lu Z, Bosmans RP, Dai Q, Hao YJ, Yang X, Zhao WM, Tong WM, Wang XJ, Bogdan F, Furu K, Fu Y, Jia G, Zhao X, Liu J, Krokan HE, Klungland A, Yang YG, He C. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell*, 2013, 49(1): 18-29.
  30. Jia GF, Fu Y, Zhao X, Dai Q, Zheng GQ, Yang Y, Yi CQ, Lindahl T, Pan T, Yang YG, He C. N6-methyladenosine in

- nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol*, 2011, 7(12): 885-887.
31. Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, Carlsson LMS, Kiess W, Vatin V, Lecoecur C, Delplanque J, Vaillant E, Pattou F, Ruiz J, Weill J, Levy-Marchal C, Horber F, Potoczna N, Hercberg S, Le Stunff C, Bougnères P, Kovacs P, Marre M, Balkau B, Cauchi S, Chèvre JC, Froguel P. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet*, 2007, 39(6): 724-726.
  32. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JRB, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 2007, 316(5826): 889-894.
  33. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, 2007, 316(5829): 1341-1345.
  34. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, Ruther U. Inactivation of the Fto gene protects from obesity. *Nature*, 2009, 458(7240): 894-898.
  35. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I, Jonsdottir T, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Jonsson F, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Lauritzen T, Aben KK, Verbeek AL, Roeleveld N, Kampman E, Yanek LR, Becker LC, Tryggvadottir L, Rafnar T, Becker DM, Gulcher J, Kiemenev LA, Pedersen O, Kong A, Thorsteinsdottir U, Stefansson K. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet*, 2009, 41(1): 18-24.
  36. Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, Wells S, Bruning JC, Nolan PM, Ashcroft FM, Cox RD. Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet*, 2010, 42(12): 1086-1092.
  37. Zhong S, Li H, Bodi Z, Button J, Vespa L, Herzog M, Fray RG. MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 2008, 20(5): 1278-1288.
  38. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G. Topology of the human and mouse m(6)A RNA methylomes revealed by m(6)A-seq. *Nature*, 2012, 485(7397): 201-206.
  39. Narayan P, Rottman FM. Methylation of mRNA. *Adv Enzymol Relat Areas Mol Biol*, 1992, 65: 255-285.
  40. Kane SE, Beemon K. Inhibition of methylation at 2 internal N-6-methyladenosine sites caused by Gac to Gau mutations. *J Biol Chem*, 1987, 262(7): 3422-3427.
  41. Tuck MT, Wiehl PE, Pan T. Inhibition of 6-methyladenine formation decreases the translation efficiency of dihydrofolate reductase transcripts. *Int J Biochem Cell Biol*, 1999, 31(8): 837-851.
  42. Bachelier JP, Amalric F, Caboche M. Biosynthesis and utilization of extensively undermethylated poly(A)<sup>+</sup> RNA in CHO cells during a cycloleucine treatment. *Nucleic Acids Res*, 1978, 5(8): 2927-2943.
  43. Stoltzfus CM, Dane RW. Accumulation of spliced avian retrovirus mRNA is inhibited in S-adenosylmethionine-depleted chicken embryo fibroblasts. *J Virol*, 1982, 42(3): 918-931.
  44. Camper SA, Albers RJ, Coward JK, Rottman FM. Effect of undermethylation on mRNA cytoplasmic appearance and half-life. *Mol Cell Biol*, 1984, 4(3): 538-543.
  45. Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA. Induction of sporulation in *Saccharomyces cerevisiae* leads to the

- formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res*, 2002, 30(20): 4509-4518.
46. Hongay CF, Orr-Weaver TL. *Drosophila* Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc Natl Acad Sci USA*, 2011, 108(36): 14855-14860.
  47. Munns TW, Liszewski MK, Sims HF. Characterization of antibodies specific for N6-methyladenosine and for 7-methylguanosine. *Biochemistry-US*, 1977, 16(10): 2163-2168.
  48. Bringmann P, Luhrmann R. Antibodies specific for N6-methyladenosine react with intact snRNPs U2 and U4/U6. *FEBS Lett*, 1987, 213(2): 309-315.
  49. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*, 2008, 26(12): 1351-1359.
  50. 高山, 张宁, 李勃, 徐硕, 叶彦波, 阮吉寿. 下一代测序中 ChIP-seq 数据的处理与分析. *遗传*, 2012, 34 (6): 773-783.
  51. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 2010, 141(1): 129-141.
  52. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol*, 2011, 12(8): r79.
  53. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods*, 2005, 37(4): 376-386.
  54. Li Y, Song S, Li C, Yu J. MeRIP-PF: An easy-to-use pipeline for high-resolution peak-finding in MeRIP-Seq data. *Genomics Proteomics Bioinformatics*, 2013, 11(1): 72-75.
  55. Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol*, 2012, 13(10): 175.
  56. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res*, 2012, 40(4): e29.