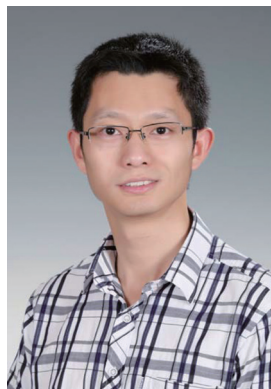


DOI: 10.13376/j.cbils/2018053

文章编号: 1004-0374(2018)04-0440-07



伊成器, 北京大学生命科学学院研究员、博士生导师。2010年于美国芝加哥大学化学系获得博士学位。2012年以“青年千人计划”(第二批)入选者的身份引进北京大学。2014年获得北京大学“绿叶生物医药杰出青年学者奖”, 2015年获得国家自然科学基金委“优秀青年科学基金项目”, 2016年获得中国化学会青年化学家奖、北京市青年岗位能手、第十届药明康德生命化学研究奖, 2018年获得第16届霍英东青年教师基金等荣誉称号。伊成器研究员通过化学生物学新技术新方法的发展, 致力于核酸化学修饰调控重要生命过程的机理研究, 在表观遗传学领域获得一系列突破性进展。课题组成立五年来, 已在高水平学术期刊上以通讯作者身份发表论文20余篇。

## RNA修饰检测技术

马士清<sup>1,2</sup>, 彭金英<sup>1\*</sup>, 伊成器<sup>1,2,3\*</sup>

(1 北京大学生命科学学院蛋白质与植物基因研究国家重点实验室, 北京 100871; 2 北京大学-清华大学生命联合中心 北京大学前沿交叉学科研究院, 北京 100871; 3 北京大学化学与分子工程学院合成与功能分子生物中心, 北京100871)

**摘要:** 目前在RNA上已发现超过一百种化学修饰, 这些修饰相关的测序技术开发和功能研究发展迅速。关于RNA修饰的相关研究领域被称为“RNA表观遗传学”。为了对RNA修饰的含量和位置进行检测, 研究人员利用液相色谱、质谱和高通量测序等技术开发了多种定量或定点的RNA修饰检测方法。该文对这些方法进行了归纳整理, 并对每种方法的基本原理和应用进行了介绍。

**关键词:** RNA修饰; 定量检测; 定点检测; 高通量测序

**中图分类号:** Q52; Q71 **文献标志码:** A

## The detection methods of RNA modifications

MA Shi-Qing<sup>1,2</sup>, PENG Jin-Ying<sup>1\*</sup>, YI Cheng-Qi<sup>1,2,3\*</sup>

(1 State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China; 2 Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China; 3 Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China)

**Abstract:** Over 100 different types of chemical modifications have been characterized in cellular RNA to date. Studies on detection methods and functional analysis of these RNA modifications have led to the emergence of a new field, called “RNA epigenetics”. Several quantitative and site-specific methods have been developed to detect the RNA modifications. In this paper, we summarized these detection methods and introduced the basic principles and their application.

**Key words:** RNA modification; quantitative detection; distribution detection; high-throughput sequencing

收稿日期: 2017-12-27

基金项目: 北京市自然科学基金项目(5162012); 国家自然科学基金委优秀青年科学基金项目(21522201); 国家自然科学基金委重大研究计划项目(91740112)

\*通信作者: E-mail: chengqi.yi@pku.edu.cn (伊成器); jypengpku@pku.edu.cn (彭金英); Tel: 010-62752895

迄今为止, 科学家们在 RNA 上已发现了上百种化学修饰, 这些修饰对于 RNA 的结构和功能起到了重要的调控作用。类比于主要研究 DNA 上的 5-甲基胞嘧啶和蛋白质上的磷酸化、乙酰化等修饰的传统表观遗传学, 人们将 RNA 修饰的相关研究称为“RNA 表观遗传学”(RNA epigenetics)<sup>[1]</sup>。

RNA 修饰种类繁多, 各种修饰的含量和功能也不尽相同, 其中 RNA 核苷上的甲基化是 RNA 修饰的主要形式, 约占了 RNA 修饰总量的三分之二。目前在 mRNA 上已经发现的甲基化修饰主要有 5-甲基胞嘧啶(5mC)、1-甲基腺嘌呤(m<sup>1</sup>A)、6-甲基腺嘌呤(m<sup>6</sup>A)、假尿嘧啶( $\psi$ )和 2' 氧甲基(2-OMe)等, 对这些修饰的合成机制和互作蛋白等方面的研究是目前 RNA 领域的热点问题。

在 RNA 修饰的研究过程中, 对 RNA 修饰含量的测定和合适的检测方法的选取是首先要解决的问题, 本文将分为定量检测技术和高通量检测技术两部分介绍 RNA 修饰的检测技术。

## 1 RNA修饰的定量检测技术

RNA 的检测主要分为定量检测和定位检测。在 RNA 化学修饰的定量测定中, 一般先利用 RNA 酶处理将 RNA 消化成单个核糖核苷酸或者寡聚核糖核苷, 然后再进行检测。目前常用的检测技术, 主要包括二维纤维素薄层析(two-dimensional cellulose thin-layer chromatography, 2D-TLC)、高效液相色谱(high performance liquid chromatography, HPLC)和液相色谱与质谱联用(coupling of liquid chromatography to mass spectrometry, LC-MS)等 3 种方法。

### 1.1 二维纤维素薄层析技术分离法(2D-TLC)

2D-TLC 对 RNA 修饰进行检测的原理是基于碱基极性的不同, 结合 <sup>32</sup>P 标记的高灵敏性, 实现对于不同核苷酸修饰的分离以及定量<sup>[2]</sup>。实验中, 首先使用 RNase A、T1 或 T2 将 RNA 链消化成寡聚核糖核苷酸, 然后用 T4 多聚核糖核苷酸激酶将 <sup>32</sup>P 标记到 5' 端, 最后使用核酸酶 P1 消化得到 5'-<sup>32</sup>P-NMP, 之后不同的单磷酸核糖核苷酸可以使用 2D-TLC 进行分离。分离时, 首先利用异丁酸-氨水-水的混合物作为展开液, 依据碱基电荷的不同进行第一次分离; 之后利用异丙醇-盐酸-水混合物或者磷酸钠-硫酸铵-异丙醇混合物根据碱基极性的不同进行第二次分离<sup>[2]</sup>。不同的核苷酸由于各自电荷和极性的差异, 经过二维分离后在层析板上位于不同的位置, 从而可以通过检测相应位置是否有 <sup>32</sup>P 信号来确定

样品含有的修饰。此外, 还可以通过将该修饰的信号强度与常见碱基(A、G、C、U)进行比较来确定该修饰的相对含量。

虽然 2D-TLC 具有便于操作和灵敏度高的优点, 但是, 实验中需要使用 RNase 进行酶切, 而 RNase 酶在消化时有碱基偏好性, 并且 <sup>32</sup>P 对于不同的修饰标记效率也有所差异, 因此, 2D-TLC 的检测结果有碱基偏好性的缺点。

### 1.2 高效液相色谱检测技术(HPLC)

高效液相色谱是一种可以快速检测和定量 RNA 修饰的研究手段, 也是基于不同核苷的极性差异的特点来实现分离。

实验中, 首先用核酸酶 P1 和碱性磷酸酶将 RNA 链消化为单个的核苷, 然后使用 C18 反向色谱柱, 在不同的水/甲醇或乙腈梯度下对不同的核苷进行分离, 同时利用 UV 分光光度计对核苷的吸光度进行检测<sup>[3]</sup>。不同的核苷具有不同的保留时间, 从而可以通过 HPLC 观察相应的保留时间点是否有峰来确定目标 RNA 中是否含有感兴趣的 RNA 修饰; 同时, 利用 UV 检测到的吸光度可以对所研究的修饰进行定量分析, 但是由于不同核苷的吸光系数有所差异, 因此, UV 检测到的吸光度并不能直接代表核苷含量。

与 2D-TLC 相比, 这种方法可以大大节省单次反应的时间, 但是由于 UV 分光光度计的灵敏度有限, HPLC 只适用于高丰度 RNA 修饰的研究。此外, HPLC 利用保留时间对核苷进行分离及鉴定, 因此, 对于一些保留时间相近的核苷, 该方法并不适用。

### 1.3 液相色谱与质谱联用检测技术(LC-MS)

液相色谱与质谱联用是一种基于 HPLC 的研究手段, 但与 HPLC 利用 UV 分光光度计对核苷进行检测不同, 本方法是在 HPLC 分离后, 使用质谱对核苷进行检测, 从而大大提高了检测的灵敏度。

在使用质谱对核苷进行检测的过程中, 首先对核苷进行电离产生离子碎片, 不同的核苷具有不同的核质比及不同的离子碎片, 从而可以利用保留时间与离子对的不同对核苷进行分离与鉴定, 并且通过对离子强度的检测来对核苷进行定量<sup>[4]</sup>。利用液相色谱与质谱联用技术, 不仅可以对高丰度的 tRNA 上的修饰在不同刺激条件下的动态变化进行定量检测, 还可以对含量较低的 mRNA 上的修饰进行检测<sup>[5]</sup>。

## 2 RNA修饰的高通量测序技术

对 RNA 修饰的定点检测技术有很多种, 以前

人们一般使用引物延伸技术、SCARLET 技术 (site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography) 等<sup>[6-7]</sup>。这些技术的通量较低，同一时间只能检测一个或几个点的修饰情况，实验操作复杂，并且需要使用到同位素。

目前较为常用的是高通量测序技术 (high-throughput sequencing)，又称下一代测序技术 (next-generation sequencing technology, NGS)，这种技术利用边合成边测序的方法，可以一次并行对几十万到几百万条 DNA 分子进行序列测定<sup>[8]</sup>。目前主要的测序平台包括美国 Illumina 公司的 HiSeq 2500、HiSeq X10 测序平台等和美国 Life Technology 公司的 Ion Torrent™ 平台。

通过结合特异性抗体富集、化学小分子标记等手段，使用高通量测序技术，针对不同的 RNA 修饰人们开发了多种高通量检测方法。

## 2.1 基于抗体富集的RNA修饰检测技术

将免疫沉淀技术 (immunoprecipitation, IP) 和二代测序技术结合起来研究核酸 / 脱氧核酸和蛋白质相互作用的技术称为 IP-seq 技术，其中像 ChIP-seq、RIP-seq、CLIP、PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) 等技术的发展已十分完善和成熟，并被广泛使用<sup>[9-11]</sup>。对于有 IP 级别抗体的 RNA 修饰，可以类似地开发相应的 IP-seq 方法。目前比较成熟的技术有针对 m<sup>6</sup>A 的 MeRIP-seq (methylated RNA immunoprecipitation sequencing)、m<sup>6</sup>A-seq、PA-m<sup>6</sup>A-seq (photo-crosslinking-assisted m<sup>6</sup>A sequencing) 和 miCLIP 等技术，以及针对 m<sup>1</sup>A 的 m<sup>1</sup>A-seq 和 m<sup>1</sup>A-ID-seq 技术。

### 2.1.1 MeRIP-seq和m<sup>6</sup>A-seq

MeRIP-seq 和 m<sup>6</sup>A-seq 技术是开发较早和应用广泛的 m<sup>6</sup>A 测序技术，下面将以 MeRIP-seq 为例进行介绍<sup>[12-14]</sup>。

在 MeRIP-seq 技术中，首先对总 RNA 通过 oligo-dT 筛选得到有 poly(A) 尾的 mRNA，然后利用二价阳离子在高温下将 mRNA 随机打断为 100~200 nt 的小片段，同时，将适量的 m<sup>6</sup>A 抗体与交联了羊抗兔免疫球蛋白 (immunoglobulin G, IgG) 的磁珠进行孵育。在用 IP 缓冲液洗掉多余的抗体之后，将片段化的 RNA 与磁珠混匀孵育过夜，洗脱数次后，在洗脱缓冲液中用蛋白酶 K 将抗体消化掉，酚氯仿抽提纯化富集到 RNA 片段，然后连接两端的测序接头，通过 PCR 反应扩增得到最终的

测序文库，最后对文库进行高通量测序并进行生物信息学分析。

由于在进行 IP 之后，只有含有 m<sup>6</sup>A 的片段能被拉下来，因此在后续的生物信息分析中，通过和使用 IgG 做 IP 的对照样本进行对比，可以看到 m<sup>6</sup>A 附近区域的富集峰。峰的宽度和片段的长度有关，通常在 100~200 nt。由于 m<sup>6</sup>A 在碱基配对中也是和胸腺嘧啶配对，因此，只能通过之前已知的 m<sup>6</sup>A 的特异周围序列 RRACH (R = purine, H = A, C 或 U) 来进行推测，通常在一个峰中会存在 1~3 个可能的 m<sup>6</sup>A 位点。

### 2.1.2 PA-m<sup>6</sup>A-seq技术

虽然 MeRIP-seq 和 m<sup>6</sup>A-seq 技术成功地测绘了 m<sup>6</sup>A 的全转录组修饰谱图 (ref)，但这两种方法只能达到 100~200 nt 的分辨率。为提高 m<sup>6</sup>A 的测序精度，何川教授课题组通过将 PAR-CLIP 技术和传统的 m<sup>6</sup>A-seq 技术结合，开发出了高分辨率的 m<sup>6</sup>A 测序技术——PA-m<sup>6</sup>A-seq<sup>[15]</sup>。

在 PAR-CLIP 技术中，通过采用具有光活性的核糖核苷类似物，如 4- 硫尿苷 (4-thiouridine, 4sU) 或者 6- 硫鸟嘌呤 (6-thioguanosine, 6sG) 处理细胞，新生成的 mRNA 上会被掺入这些类似物<sup>[11]</sup>。经 365 nm 的紫外光短时间照射后，这些具有光活性的类似物会与 RNA 结合蛋白上位置接近的氨基酸残基发生反应，该反应产物会在反转录反应中导致碱基错配，据此就可以比较精确地获知 RNA 与蛋白的作用位点。

受 PAR-CLIP 技术的启发，在 PA-m<sup>6</sup>A-seq 技术中，研究人员首先用 4sU 对 HeLa 细胞进行适当时间的处理，然后用 oligo-dT 的磁珠纯化得到含有 4sU 的 mRNA。不像 MeRIP-seq 或 m<sup>6</sup>A-seq 处理的是片段化的 mRNA，该项技术直接针对全长的 mRNA，用 m<sup>6</sup>A 抗体进行免疫沉淀。在抗体和 RNA 孵育之后，用 365 nm 的紫外光照射 IP 体系来诱导交联，然后用 RNase T1 消化 RNA 到约 30 nt 左右；之后使用蛋白酶 K 处理消化掉剩余的 m<sup>6</sup>A 抗体，只剩下和 RNA 共价交联位点处的肽段；接着酚氯仿抽提纯化处理后的 RNA 片段，两端与测序接头连接，经 PCR 扩增后得到最终测序文库。

根据 4sU 交联后会产生突变的特性，在数据分析时，可以根据富集峰中是否存在 T 到 C 的突变来去除假阳性；此外，交联后 RNase T1 的高效切割使得 PA-m<sup>6</sup>A-seq 可以达到 23 nt 左右的分辨精度，相较于 MeRIP-seq 和 m<sup>6</sup>A-seq 技术，PAR-CLIP 在

测序精度上有了突破性的进步。

### 2.1.3 miCLIP技术和m<sup>6</sup>A-CLIP技术

虽然 PA-m<sup>6</sup>A-seq 技术大大提高了测序精度, 但由于其仍需要通过 m<sup>6</sup>A 的特殊周边碱基序列来推测确定 m<sup>6</sup>A 的位点, 所以, 依旧不能在实验水平上做到单碱基分辨率。为了达到精确测定 m<sup>6</sup>A 位点的目的, 研究人员利用目前已有的高特异性的 m<sup>6</sup>A 抗体, 开发出了单碱基分辨率的 m<sup>6</sup>A 测序方法 miCLIP (m<sup>6</sup>A individual-nucleotide-resolution crosslinking and immunoprecipitation) 和 m<sup>6</sup>A-CLIP<sup>[16-17]</sup>。这两种方法可以准确地测定 m<sup>6</sup>A 具体位置, 为后续研究提供了强有力的技术支持。

基于在 iCLIP 技术中, 紫外交联后共价结合在碱基上的氨基酸残基可以导致反转录终止的特点, 研究人员猜测 m<sup>6</sup>A 抗体在紫外交联后也可以产生类似的效应, 从而可以作为特异识别标记用来确定 m<sup>6</sup>A 的位点。基于此猜测, 在 miCLIP 技术开发过程中, 首先体外转录合成了含有单个 m<sup>6</sup>A 修饰位点的模式 RNA 序列, 然后用几种不同的商业化 m<sup>6</sup>A 抗体进行 IP。使用 254 nm 的紫外光照射进行交联, 然后用蛋白酶 K 消化 RNA- 抗体复合物, 只留下共价交联于 RNA 的氨基酸残基。之后纯化 RNA 片段, 进行反转录和测序, 统计每种不同的抗体造成的碱基错配或终止反转录的情况。通过实验, 研究人员发现, SySy 公司的抗体有大概率会在 m<sup>6</sup>A 的 3' 端后一位引入特异性的反转录终止, 而 Abcam 公司的抗体则会在该位置引入 C 到 T 的碱基错配, 因此, 在 miCLAP 技术中一般采用 SySy 公司的抗体。类似地, 在 m<sup>6</sup>A-CLIP 技术中, 研究人员也选择了 SySy 公司的抗体进行测序。

在筛选得到合适的抗体之后, 研究人员利用这两种抗体, 通过和在模式 RNA 序列上类似的操作手段, 对真实 mRNA 样品进行处理和测序文库构建。生物信息学分析显示, 这种方法可以准确得到 m<sup>6</sup>A 的位置, 同时具有很高的信噪比和敏感性。更为重要的是, 对于其他有高特异性抗体的修饰, 也可以参考 miCLIP 建立类似的单碱基分辨率测序方法。

### 2.1.4 m<sup>1</sup>A 测序技术

m<sup>1</sup>A 很早就被发现存在于 tRNA 的 9 位和 58 位以及 rRNA 上, 但直到 2016 年人们才证实其存在于 mRNA 上; 同时, 利用商业化的 IP 级别的 m<sup>1</sup>A 特异性抗体, 研究人员开发了 m<sup>1</sup>A-seq 和 m<sup>1</sup>A-ID-seq 两种技术<sup>[18-19]</sup>。

m<sup>1</sup>A-seq 和 m<sup>1</sup>A-ID-seq 的基本思路与 MeRIP-

seq 一样, 都是利用抗体对含有 m<sup>1</sup>A 的片段进行富集, 然后建库测序。与 m<sup>6</sup>A 不影响碱基配对或者反转录过程有所不同, m<sup>1</sup>A 在配对过程中会主要产生 A 到 T 的碱基错配, 同时有一定的可能性使逆转录酶停在其 3' 端后一位碱基处, 这导致了在直接 IP 测序到的峰中会出现“落峰”现象, 即一个峰的中间部分相对于两边反而较低。此外, 在 m<sup>1</sup>A-seq 中, 研究人员利用碱性条件下 m<sup>1</sup>A 会发生 Dimroth 重排反应转变为不影响碱基配对的 m<sup>6</sup>A 的性质, 进一步寻找到了若干个具体的 m<sup>1</sup>A 修饰位点。而在 m<sup>1</sup>A-ID-seq 中, 研究人员利用了一个去甲基化酶使 m<sup>1</sup>A 去甲基化转变为正常的 A, 从而提高检测的可信度和分辨率。两种选择的基本思想相同, 都是通过将 m<sup>1</sup>A 进行转变来降低或消除其对碱基配对和反转录的影响, 从而在数据分析的过程中可以与 IP 组进行对比, 减少假阳性信号和提高测序的可靠性。但上述两种测序方法都没有做到单碱基分辨率, 目前的检测精度在 50~200 nt。

## 2.2 利用化学手段的RNA修饰测序方法

有很多种 RNA 修饰, 目前并没有比较好的特异性抗体, 这种时候就需要利用化学手段来进行区分和测序。一般是筛选可以和修饰发生特异性反应的化学小分子, 利用小分子及其衍生物进行富集或者反应产物的特殊性质来进行测序。

### 2.2.1 假尿嘧啶的测序方法

假尿嘧啶 (pseudouridine,  $\psi$ ) 是目前已知的在总 RNA 中含量最多的化学修饰<sup>[20]</sup>。假尿嘧啶的化学性质和尿嘧啶接近, 都与腺嘌呤配对, 而且目前没有特异性较好的假尿嘧啶抗体, 因此, 只能利用化学手段进行区分和测序。化合物 CMCT (N-cyclohexyl-N'-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate) 可以和尿嘧啶、假尿嘧啶以及鸟嘌呤发生反应, 但在经过碱性条件 (pH = 10.4) 处理之后, 只有假尿嘧啶的 3 号位氮原子和 CMCT 的反应产物会得到保留。由于 N3 位在 Watson-Crick 碱基配对面, 因此, CMCT 的加成会阻止碱基配对, 导致在反转录过程中逆转录酶会停在假尿嘧啶的 3' 端, 产生截短的 cDNA 片段。这一反应最早在 1993 年由 Bakin 和 Ofengand<sup>[21]</sup> 用于引物延伸实验, 实现了 rRNA 上的假尿嘧啶位点的检测。

利用 CMCT 的特异性反应, 结合高通量测序手段, 近年来研究人员开发了多种假尿嘧啶的单碱基分辨率的全转录组高通量测序方法。下面将分为两类进行介绍, 即直接利用 CMCT 进行测序和利

用 CMCT 衍生物进行测序的方法。

### 2.2.2 直接利用 CMCT 进行测序的方法——Psi-seq、Pseudo-seq 和 $\psi$ -seq

Psi-seq、Pseudo-seq 和  $\psi$ -seq 这 3 种方法的基本原理和实验过程较为相似，因此，这里选择  $\psi$ -seq 作为代表来进行介绍<sup>[22-24]</sup>。

在  $\psi$ -seq 中，首先提取总 RNA 并进行 oligo-dT 筛选得到 mRNA，然后对 mRNA 进行随机片段化到约 100~150 nt。将片段化的 mRNA 与 CMCT 在 37 °C 反应 30 min，然后碱性条件 (pH = 10.4) 处理 6 h 来去除 CMCT 和 G 与 U 的反应产物。回收反应后的 mRNA，连接 3' 接头序列，用与接头序列互补的反转引物进行反转录，此时逆转录酶会停在反应后的假尿嘧啶的 3' 端一个碱基的位置，产生截短的 cDNA 片段；再连接 5' 接头序列，进行 PCR 扩增后即可得到测序文库。

由于逆转录酶会被 CMC- $\psi$  的反应产物终止，因此，在生物信息学分析中，可以统计每个尿嘧啶停在 3' 端的序列个数和通读的序列个数的比值。如果某一位置处存在假尿嘧啶修饰，则该位置处的比值会显著高于背景，通过设定合适的阈值即可确定修饰位点。

利用这一方法，研究人员成功地对酵母和人 HeLa 细胞系等进行了全转录组测序，鉴定出了数百个 mRNA 的假尿嘧啶修饰位点；同时，通过对热休克处理条件下的样品进行测序，发现在这种条件下 mRNA 和 ncRNA 会诱导产生新的修饰位点，而且这些产生了新位点的 mRNA 会变得更加稳定，说明假尿嘧啶可能在生物体应对环境刺激的反应中有重要作用。

### 2.2.3 利用 CMCT 衍生物进行测序的方法——CeU-seq

虽然上述 3 种方法成功地实现了假尿嘧啶的单碱基分辨率测序，但由于测序过程中没有富集过程，因此，对于修饰比例较低或者表达量较低的转录本上的位点，不能够很好地进行区分，而且测序背景也比较高，可能会引入一些假阳性位点。

在最新开发的 CeU-seq (CMC-enriched pseudouridien sequencing) 方法中，研究人员通过合成 CMCT 衍生物和利用点击化学反应很好地解决了上述问题<sup>[18]</sup>。点击化学 (click chemistry) 的概念由 Sharpless 在 2001 年提出，通过以碳-杂原键 (C-X-C) 合成为基础的组合化学方法简单高效地获得分子多样性，代表反应为铜催化的叠氮-炔基 Huisgen 环加成反应。

在这一方法中，研究人员首先对 CMCT 进行了化学改造，在其上引入了一个叠氮基团 ( $N_3$ -CMCT)，之后将 mRNA 与  $N_3$ -CMCT 进行反应和进一步的碱处理。将处理后的 mRNA 与 DBCO-biotin 混合，DBCO 基团可以不需要铜离子即可和叠氮基团发生点击化学反应，从而生成  $\psi$ -CMC-biotin 的反应产物。再利用偶联了亲和素的磁珠对带有生物素的 mRNA 片段进行 IP，即可实现对含有  $\psi$  的片段的富集。之后的建库方法和前面的方法类似，也是先连接 3' 接头序列进行反转录，然后对 cDNA 进行成环和 PCR 扩增得到最终文库。

通过对含  $\psi$  片段的富集，可以提高测序后停在  $\psi$  位点后一位的序列与通读序列的相对比值，从而提高信噪比和减少假阳性。此外，富集过程还可以提高修饰比例较低的位点和表达量较低的 mRNA 上的位点相对于背景的信号强度，从而在生物信息学的分析过程中将这些位点筛选出来；同时，富集操作还可以有效地降低得到可信位点所需的测序深度，从而大大节省了测序费用。

通过将 CeU-seq 应用于人 HEK293T 细胞系和小鼠组织，研究人员分别得到了 2 000 个和 1 800 个  $\psi$  位点，与测定的  $\psi$  在 mRNA 中的含量 ( $\psi/U = 0.2\% \sim 0.4\%$ ) 更为符合。当采用与  $\psi$ -seq 同样的测序深度和生信阈值对几种方法的数据进行分析时，CeU-seq 仍可得到约 1 000 个  $\psi$  位点，证明了该方法可以提高对  $\psi$  检测的灵敏度，为研究  $\psi$  的生物学功能提供了更为强大的工具。以上 4 种假尿嘧啶测序技术原理对比见图 1。

## 3 三代测序技术

二代测序技术虽然目前得到了极大的发展和应用，但其方法中的 PCR 扩增过程会在一定程度上增加测序的错误率，而且读长较短，有一定的系统偏向性。为了解决上述缺点，近些年来开发了三代测序技术，其具有单分子测序、不需要 PCR 扩增、读长长、测序错误随机及无系统偏差等特点。

三代测序以 PacBio 公司的 SMRT 技术 (single-molecule amplification and sequencing technology) 和 Oxford Nanopore Technologies 公司的纳米孔单分子测序技术为代表<sup>[25-26]</sup>。SMRT 技术的基本原理是将 DNA 聚合酶固定于约 100 nm 的小孔中，使其和模板结合，同时加入分别用不同荧光标记的 4 种核苷酸，在碱基配对阶段，不同碱基的加入会发出不同的光，根据收集到的光的波长与峰值可判断掺入

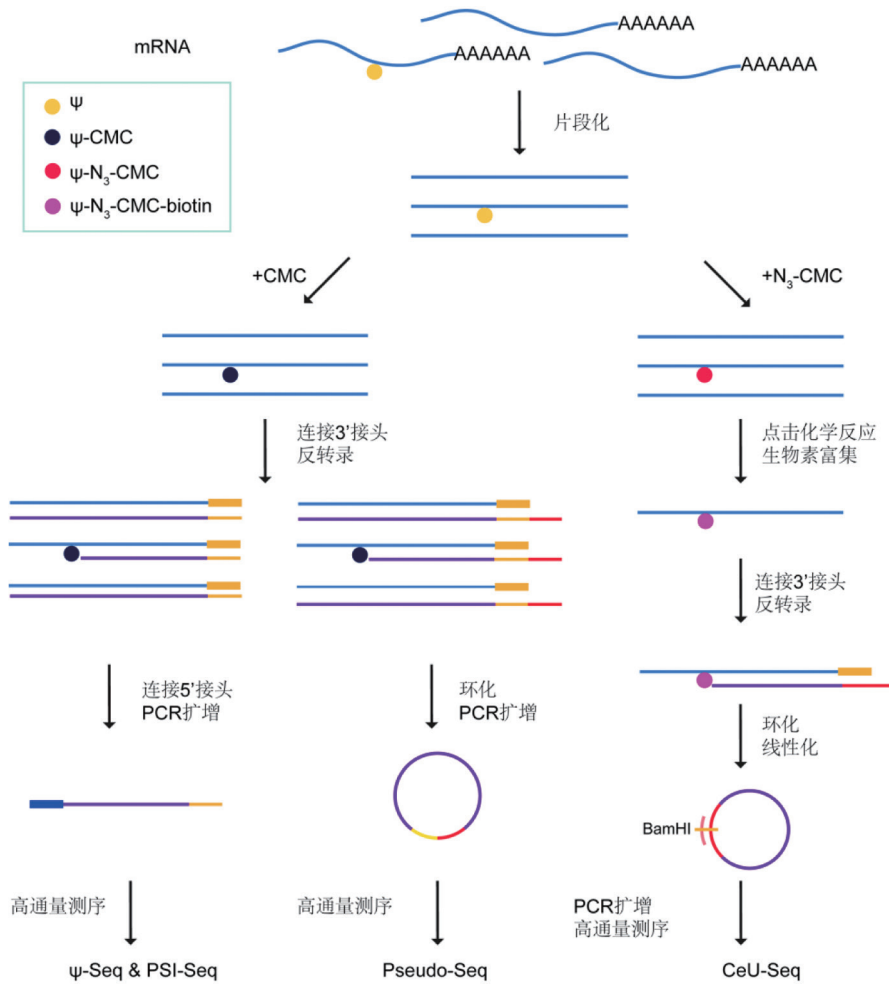


图1 4种假尿嘧啶测序技术原理对比示意图

的碱基类型。如果某一碱基处存在修饰(如5mC),其通过聚合酶时的速度会减慢,此时检测相邻两个碱基之间信号相隔的时间,可发现相邻两峰之间的距离增大,从而可以用来检测甲基化等信息。SMRT目前已被成功应用于对DNA上的5mC、5hmC和m<sup>6</sup>A等几种修饰的检测,但目前还没有被用于检测RNA上化学修饰的报道。Oxford Nanopore Technologies公司所开发的纳米单分子测序技术与以往的测序技术皆不同,它是基于电信号而不是光信号。当单链DNA或者RNA通过特殊设计的纳米孔时,由于碱基的电荷影响,会短暂地改变流过纳米孔的电流强度(每种碱基所影响的电流变化幅度不同),通过检测电流的变化即可鉴定出所通过的碱基,理论上对于不同的DNA或RNA修饰都可进行区分。此方法目前仍处于技术开发完善阶段,暂时还没有应用该技术进行RNA修饰测定的相关报道。

综上,本文系统介绍了目前对RNA修饰进行定量和定位检测的常用的一些方法,这些方法的开发和应用对推动整个RNA修饰研究领域的发展起到了至关重要的作用。借助于这些方法,人们可以对不同的生物样本和材料中的RNA修饰含量与位置进行测定,通过对比可以更好地研究这些修饰的功能。此外,受到目前已有方法的启发,更多的修饰正在被发现,研究人员们也在努力开发新的更加高效和精确的测序方法,这些新技术的研发必将引起整个RNA表观遗传学领域的新一波研究热潮。

[参 考 文 献]

[1] He C. Grand challenge commentary: RNA epigenetics? Nat Chem Biol, 2010, 6: 863-5  
 [2] Grosjean H, Keith G, Droogmans L. Detection and quantification of modified nucleotides in RNA using thin-layer chromatography. Methods Mol Biol, 2004, 265: 357-91

- [3] Nees G, Kaufmann A, Bauer S. Detection of RNA modifications by HPLC analysis and competitive ELISA. *Methods Mol Biol*, 2014, 1169: 3-14
- [4] Brandmayr C, Wagner M, Bruckl T, et al. Isotope-based analysis of modified tRNA nucleosides correlates modification density with translational efficiency. *Angew Chem Int Ed Engl*, 2012, 51: 11162-5
- [5] Chan CT, Dyavaiah M, DeMott MS, et al. A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. *PLoS Genet*, 2010, 6: e1001247
- [6] Stern S, Moazed D, Noller HF. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol*, 1988, 164: 481-9
- [7] Liu N, Parisien M, Dai Q, et al. Probing *N*<sup>6</sup>-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA*, 2013, 19: 1848-56
- [8] Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*, 2010, 11: 31-46
- [9] Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, 2009, 10: 669-80
- [10] Zhao J, Ohsumi TK, Kung JT, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell*, 2010, 40: 939-53
- [11] Hafner M, Landthaler M, Burger L, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 2010, 141: 129-41
- [12] Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature*, 2012, 485: 201-6
- [13] Meyer KD, Saletore Y, Zumbo P, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 2012, 149: 1635-46
- [14] Schwartz S, Agarwala SD, Mumbach MR, et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell*, 2013, 155: 1409-21
- [15] Chen K, Lu Z, Wang X, et al. High-resolution *N*<sup>6</sup>-methyladenosine m<sup>6</sup>A map using photo-crosslinking-assisted m<sup>6</sup>A sequencing. *Angew Chem Int Ed Engl*, 2015, 54: 1587-90
- [16] Linder B, Grozhik AV, Olarerin-George AO, et al. Single-nucleotide-resolution mapping of m<sup>6</sup>A and m<sup>6</sup>A<sub>m</sub> throughout the transcriptome. *Nat Methods*, 2015, 12: 767-72
- [17] Ke S, Alemu EA, Mertens C, et al. A majority of m<sup>6</sup>A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev*, 2015, 29: 2037-53
- [18] Li X, Xiong X, Wang K, et al. Transcriptome-wide mapping reveals reversible and dynamic *N*<sup>1</sup>-methyladenosine methylome. *Nat Chem Biol*, 2016, 12: 311-6
- [19] Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, et al. The dynamic *N*<sup>1</sup>-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, 2016, 530: 441-6
- [20] Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*, 2000, 49: 341-51
- [21] Bakin A, Ofengand J. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry*, 1993, 32: 9754-62
- [22] Schwartz S, Bernstein DA, Mumbach MR, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, 2014, 159: 148-62
- [23] Carlile TM, Rojas-Duran MF, Zinshteyn B, et al. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, 2014, 515: 143-6
- [24] Lovejoy AF, Riordan DP, Brown PO. Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One*, 2014, 9: e110799
- [25] Chin CS, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*, 2013, 10: 563-9
- [26] Derrington IM, Butler TZ, Collins MD, et al. Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci USA*, 2010, 107: 16060-5