

表观遗传学研究方法进展

郑小国^{1,2} 陈亮¹ 楼巧君¹ 罗利军^{1,2}

(¹上海市农业生物基因中心, 上海 201106; ²华中农业大学, 武汉 430070)

摘要: 表观遗传调控是基因表达调控的重要组成部分, 已成为当前研究的热点。目前其研究主要集中在 DNA 甲基化和组蛋白修饰。针对这两种表观修饰, 其研究方法也取得了较大进展, 一方面方法的灵敏度和特异性都在不断提高; 另一方面表观修饰的检测正在逐步从定性检测向定量分析方向发展, 从个别位点向高通量检测发展。此外, 新一代测序技术的应用将大大推动表观遗传研究的发展, 包括单分子实时测序法、单分子纳米孔测序法等。综述目前常用的 DNA 甲基化、组蛋白修饰研究方法以及最新的单分子测序技术, 并对它们在表观遗传修饰检测中的应用作了简要对比分析。

关键词: 表观遗传学 DNA 甲基化 组蛋白修饰 限制性内切酶酶切法 重亚硫酸盐法 染色质免疫共沉淀 单分子实时测序 单分子纳米孔测序

Research Method Progress of Epigenetics

Zheng Xiaoguo^{1,2} Chen Liang¹ Lou Qiaojun¹ Luo Lijun^{1,2}

(¹Shanghai Agrobiological Gene Center, Shanghai 201106; ²Huazhong Agricultural University, Wuhan 430070)

Abstract: Epigenetic regulation has become a hotspot in the study on gene expression regulation. Epigenetic regulation mainly focuses on DNA methylation and histone modification. A great deal of techniques has been developed to detect these epigenetic modifications and great progresses have also been achieved in terms of detection sensitivity and specificity, qualitative and quantitative assay. The next-generation sequencing technologies will greatly promote the development of epigenetic research providing a chance for rapid and high throughput epigenetic detection, which include the single-molecular real-time sequencing, single-molecule nanopore DNA sequencing. This review introduced the most commonly used and newly developed techniques, and made a brief comment about their applications in the detection of DNA methylation and histone modification.

Key words: Epigenetic DNA methylation Histone modification Restriction endonuclease enzyme digestion Bisulfate Chromatin immunoprecipitation(ChIP) Single-molecular real-time sequencing Single-molecule nanopore DNA sequencing

表观遗传(epigenetic)是指 DNA 序列不发生变化, 但基因表达却发生了可遗传的改变。这种改变是细胞内除了遗传信息以外的其他可遗传物质发生的改变, 且这种改变在发育和细胞增殖过程中能稳定遗传^[1,2]。

表观遗传学的现象很多, DNA 甲基化(DNA methylation)、组蛋白修饰(histone modification)、基因组印记(genomic imprinting)、RNA 编辑(RNA editing)、基因沉默、核仁显性、休眠转座子激活和性别相关性基因剂量补偿效应等都是典型的表观遗传现象^[3,4]。表观遗传研究目前主要集中在 DNA 甲基

化和组蛋白共价修饰两方面。

DNA 甲基化(DNA methylation)是常见的表观遗传现象, 它是指在 DNA 甲基转移酶(DNA methyltransferases, DNMTs)的作用下将甲基添加在 DNA 分子中的碱基上。常见的 DNA 甲基化发生在 DNA 链上的胞嘧啶(C)第 5 位碳原子和甲基间的共价结合形成 5 甲基胞嘧啶(5mC)^[5,6]。除此以外, 在腺嘌呤(A)N6 位置可以引入一个甲基形成 N6 甲基腺嘌呤(N6-mA)。在真核生物中, DNA 甲基化主要发生在 CpG 二联核苷酸上, 也常在植物基因组 CNG 和 CNN 序列上发现^[7,8]。在某些区域, CpG 序列密

收稿日期: 2011-04-07

作者简介: 郑小国, 男, 博士研究生, 研究方向: 表观遗传学; E-mail: zxcg08@sagc.org

通讯作者: 罗利军, 男, 研究员, 博士, 研究方向: 水稻节水抗旱; E-mail: lijun@sagc.org.cn

度很高,可达均值的 5 倍以上,成为 C 和 G 的富集区,就像在 DNA 序列汪洋大海中浮现的一座座孤岛,称为 CpG 岛(CpG islands)^[9]。作为表观遗传学最重要的现象之一,DNA 甲基化对基因的表达发挥着重要的调控功能,涉及异染色质形成、转座子扩散防御、基因印迹、源基因表达调控、转基因沉默和细胞分化等,在生命活动中起着重要的作用^[3,4,10]。

组蛋白共价修饰(histone modification)包括^[11]:乙酰化(acetylation)、甲基化(methylation)、磷酸化(phosphorylation)、核糖化(ribosylation)、泛素化(ubiquitylation)和类泛素化(sumoylation)。这些修饰共同构成了可以通过特殊解码蛋白解读的组蛋白密码,在解码过程中影响基因表达^[12,13]。其中乙酰化和甲基化主要发生在组蛋白 lys 残基上,乙酰化与转录激活相关;甲基化既可能与转录激活相关,也可能与转录抑制相关,激活或抑制取决于甲基化 lys 位点以及添加的甲基基团数目^[8,14,15],每个 lys 残基可以加上 1-3 个甲基基团,形成单甲基化、二甲基化、三甲基化 3 种形式,每种形式有其独特的功能^[16]。磷酸化是另外一种重要的调控方式^[17],在基因转录、DNA 修复、细胞凋亡以及染色质凝聚等过程中起重要作用。核糖化^[18]是指组蛋白 H1、H2A、H2B 及 H3 和多聚 ADP-核糖的共价结合,ADP-核糖基化被认为是在真核细胞内启动复制过程的扳机;泛素化^[19]是指将被降解的蛋白质连接上泛素标记,一些可以诱导基因启动子区域的 H2B 组蛋白会被修饰以启动基因表达;类泛素蛋白^[20](small ubiquitin-like modifiers, SUMO)对蛋白质的翻译后修饰被认为与蛋白质的细胞内定位、稳定性和转录活性有关,也可能参与异染色质结构的调节。

近年来,人们越来越认识到表观遗传在基因表达调控方面的重要性,并开发出一系列检测表观遗传修饰的方法,尤其是 DNA 甲基化和组蛋白修饰检测方法取得了较大进展,一方面方法的灵敏度和特异性都在不断提高;另一方面表观修饰的检测正在逐步从定性检测向定量分析方向发展,从个别位点向高通量检测发展。

1 DNA 甲基化研究技术

目前全基因组或接近全基因组甲基化检测技术较多,主要分为 3 类:第 1 类技术以限制性内切酶酶

切为基础,用一个或多个酶限制性切割未甲基化 DNA(如 *Hpa* II 和 *Not* I)或甲基化 DNA(如 *Mcr*BC)。这些方法结合芯片、毛细管测序^[21]等技术已经检测了多种生物的全基因组甲基化,但仅限于内切酶能够识别的 CpG 位点。第 2 类技术依赖于全基因组 DNA 重亚硫酸盐转换,经重亚硫酸盐处理后基因组 DNA 未甲基化胞嘧啶(C)转换为尿嘧啶(U)(经扩增后最终为 T),甲基化 C 保持不变^[22]。此方法可以进行单 CpG 位点解析并且可以结合芯片或高通量测序。但此方法的局限性在于经过重亚硫酸盐转换后,序列特异性降低,在芯片上难以设计足够的特异性探针进行全基因组分析。此外,如果用于较大基因组日常分析,此方法十分昂贵。第 3 类技术以免疫学为基础,用 5-甲基胞嘧啶特异性抗体或者用含有甲基结合结构域的蛋白通过免疫沉淀富集基因组甲基化或未甲基化片段^[23],称为甲基化 DNA 免疫共沉淀(methylated DNA immunoprecipitation, MeDIP or mDIP)^[10,24-28],MeDIP 结合芯片(MeDIP-chip)技术可以对任何物种进行高通量全基因组 DNA 甲基化作图^[29]。

以上 3 类技术可以与测序或芯片技术组合形成多种方法:如酶切结合测序、酶切结合芯片技术、亚硫酸氢钠结合测序、亚硫酸氢钠结合芯片技术、免疫沉淀结合测序和免疫沉淀结合芯片技术等。

1.1 以限制性内切酶为基础的 DNA 甲基化检测技术

1.1.1 甲基化敏感扩增多态性(methylation sensitive amplification polymorphism, MSAP) 本方法是在扩增片段长度多态性(amplified fragment length polymorphism, AFLP)技术的基础上建立起来的^[30,31]。其基本程序是:提取基因组 DNA,根据 *Hpa* II、*Msp* I 两种酶对甲基化敏感程度不同,分别用 *Eco*R I/*Hpa* II、*Eco*R I/*Msp* I 两组酶组合对基因组 DNA 进行双酶切,连上接头后,用按照接头序列设计的引物加上一个选择性碱基进行 PCR 扩增,称为预扩增。预扩增产物稀释后,再加入另外两个选择性碱基引物进行第二次 PCR 扩增,称为选择性扩增。扩增产物变性后在 6% 的 PAGE 胶上电泳,最后采用银染或同位素放射自显影方法处理 PAGE 胶,统计和分析 DNA 条带即可得出基因组 CpG 位点甲基化状态。还可以对差

异扩增片段回收克隆测序 通过比对可鉴定甲基化差异位点。MSAP 技术相对其他 DNA 甲基化测定技术有如下优点: 不需知道被测 DNA 的序列信息; 操作相对简便, 在 AFLP 基础上可直接操作; 可在全基因组范围检测 CCGG 位点的胞嘧啶甲基化变化。MSAP 技术的局限性在于只能检测 CCGG 位点的胞嘧啶甲基化。此方法常用于研究动植物全基因组甲基化状态。

1.1.2 McrBC 酶切法 McrBC 是一个 DNA 甲基化特异性限制性内切酶, 识别 (A/G)C 基序中甲基化的 C, 切割一条或两条链上含有甲基化胞嘧啶的 DNA。McrBC 不作用于非甲基化的 DNA。McrBC 的 DNA 识别位点包含两个 (G/A)mC 形式的半位点。半位点的间距最长可达 3 kb, 但最适宜切割的间距是 55–103 bp。McrBC 切割时需要 GTP 存在, 当存在 GTP 的非水解类似物时, 该酶可以特异性地结合到甲基化的 DNA 上, 而不产生切割反应。McrBC 在一对 PvuCG 序列元件上进行切割, 因此可以检测相当大一部分的甲基化 CpGs。

通常该方法结合芯片技术检测甲基化 DNA, 其基本程序是: 首先用 McrBC 酶对全基因组 DNA 进行酶切, 然后将酶切片段与芯片杂交, 对数据统计分析检测甲基化 DNA。该方法的局限性在于: McrBC 不能识别内部胞嘧啶甲基化了的 *Hpa* II/*Msp* I 位点 (CCGG), 而且该酶在每一对半位点之间切割一次, 切割位点靠近这一对半位点的一侧, 但通常距最近的甲基化碱基大约 30 bp。当 DNA 片段含有多个 McrBC 识别半位点时 (如含有甲基化胞嘧啶的基因组 DNA), 因该酶的切割位点具有不确定性, 易导致切割位点重叠^[32]。

1.2 重亚硫酸盐法

重亚硫酸盐法的原理是单链 DNA 在 HSO₃ 存在的条件下能有效地将未甲基化 C 转变成 U, 两轮 PCR 扩增后, 该位点变为 T, 而 m5C 则保持不变, 仍然为 C。通过扩增克隆测序比对, 可以鉴别待测位点是否发生了甲基化^[22]。该技术很容易确定个体基因组 DNA 分子中各个单链的甲基化状态及甲基化位置。

在此基础上, 可以结合 DNA 直接测序、限制性内切酶分析、甲基化特异性 PCR 扩、甲基化敏感的

单核苷酸引物延伸法、高效液相色谱法以及锁式探针技术等方法测出相应序列的甲基化情况^[33]。

1.2.1 重亚硫酸盐结合 DNA 直接测序法 该方法由 Formmer 等^[22]提出, 其基本过程是: 提取高质量基因组 DNA 平均分为两组, 一组经重亚硫酸盐处理转换, 另一组作为对照。用引物进行扩增, 对分别获得的两组扩增片段进行测序对比, 如果处理后一组的 C 变为 T, 则这些转变的位点为没有发生甲基化的位点, 若处理后仍然为 C, 则这些位点发生了甲基化。该方法适用于特定基因甲基化检测, 如鉴定某个基因在某种状态下的甲基化状态。

1.2.2 联合重亚硫酸盐限制性内切酶分析法 联合重亚硫酸盐限制性内切酶分析法 (combined bisulfite restriction analysis, COBRA) 由 Xiong 和 Peter 报道^[33, 34], 其基本程序是: 先对样本 DNA 进行重亚硫酸盐处理, PCR 扩增目的片段, 随后用限制性内切酶消化, 此酶识别序列中需包含 CG 序列, 如 BstU I (CGCG)。若其识别序列中的 C 发生完全甲基化 (5mCG5mCG), 则 PCR 扩增后保留为 CGCG, BstU I 能够识别并进行切割; 若待测序列中 C 未发生甲基化, 则 PCR 后转变为 TGTG, BstU I 识别位点丢失, 不能进行切割。这样酶切产物再经电泳分离、探针杂交、扫描定量后即可得出原样本中甲基化的比例^[35, 36]。用 COBRA 和 Agilent 2100 Bioanalyzer 对酶切产物进行直接分析, 使 COBRA 的定量更快速、准确且无放射性污染^[37]。此方法相对简单, 不需预先知道 CpG 位点及样本序列; 可对 DNA 甲基化水平进行定量研究; 需要样本量少, 可用于石蜡包埋样本的分析。本方法的局限性在于只能获得特殊酶切位点甲基化情况, 因此检测阴性不能排除样品 DNA 中甲基化存在的可能。此方法常用于样本量少, DNA 甲基化定量检测试验中。

1.2.3 亚硫酸盐结合甲基化特异性 PCR 法 甲基化特异性 PCR (methylation-specific PCR/MS-PCR) 是 Herman 等^[38]首先提出的一种检测基因组 DNA 甲基化水平的常用方法^[33, 38]。该法是将 DNA 经重亚硫酸盐处理, 非甲基化的胞嘧啶转变为尿嘧啶, 而甲基化的胞嘧啶保持不变。在 PCR 反应时, 设计两套不同的引物对: 一对引物序列针对经亚硫酸氢钠处理后的甲基化 DNA 链设计, 若用该对引物能扩增出

片段,说明该检测位点发生了甲基化;另一引物针对经亚硫酸氢钠处理后的非甲基化 DNA 链设计,若用该对引物能扩增出片段,说明该检测位点没有甲基化。两套引物都具有很高的特异性^[39]:引物末端均设计至检测位点结束;两套引物分别只能与重亚硫酸盐处理后的序列互补配对,即一对结合处理后的甲基化 DNA 链,另一对结合处理后的非甲基化 DNA 链,与未经处理的 DNA 序列无互补配对; MSP 引物覆盖序列中必须含有一个或一个以上的 CpG 岛,以保证引物的特异性,经克隆测序就检测出引物所覆盖序列的甲基化位点,引物覆盖序列中 CpG 岛所占比例越高,甲基化 DNA 检出率越高。该法不足之处在于^[40]:引物的选择和设计非常关键,否则易导致假阳性;如果亚硫酸氢钠对 DNA 处理不完全,也易导致假阳性。可通过限制性内切酶酶解法检验 PCR 产物作进一步判断。

1.2.4 甲基化敏感的单核苷酸引物延伸法 甲基化敏感的单核苷酸引物延伸法 (methylation-sensitive single-nucleotide primer extension, Ms-SNuPE) 可用于快速判断 DNA 片段中某些具体位点的甲基化状况。由 Gonzalgo and Jones 提出^[41]。其基本原理是:样本 DNA 经重亚硫酸盐处理后 PCR 扩增目的片段,产物电泳分离后平均分为两份作为 Ms-SNuPE 模板。分别针对所检测的 CpG 位点设计上游引物,使 3'端引物紧邻 C。扩增时加³²P 标记的 dCTP 或 dTTP 进行单核苷酸延伸反应。如果待测位点被甲基化,则同位素标记的 dCTP 会在反应延伸时连于引物末端;若是未被甲基化,则标记的 dTTP 参与反应^[42-44]。再电泳分离产物,成像。根据某一 CpG 位点的 C/T 信号强度比,可定量其甲基化程度。也可不经电泳,PCR 产物直接转移到尼龙膜上,成像后即可得到所测多个 CpG 位点的平均甲基化程度。Ms-SNuPE 可快速定量多个 CpG 位点的甲基化程度,但序列较长时此方法不适用,且此方法具有放射性。

1.2.5 高效液相色谱法 (high performance liquid chromatography, HPLC) Oefner 等^[45]首先提出变性高效液相层析 (denaturing HPLC, DHPLC) 用于单核苷酸和 DNA 分析。邓大君等^[46]将其改进并与 PCR 联用建立了一种检测甲基化程度的分析方法,其原理是将经重亚硫酸盐处理的 DNA 进行差异性

扩增,由于原甲基化的胞嘧啶经重亚硫酸盐处理被保留,因此在随后的 PCR 扩增时,其变性温度也相应上升,使 PCR 产物在色谱柱中保留的时间明显延长,从而判定出 PCR 产物的 DNA 序列甲基化状况。

1.2.6 亚硫酸盐结合锁式探针法 锁式探针 (padlock probe) 及其应用是 Nilsson 等首次报道的^[47-48],其原理是只有当锁式探针和相应的检测靶 DNA 同时存在于连接体系中且完全互补配对时,线型锁式探针才能在连接酶的作用下被有效地连接成环型,而在没有相应靶 DNA 存在时锁式探针不被连接酶连接,仅以线性形式存在。在核酸外切酶的作用下,线性形式存在的锁式探针被消化水解,而连接成环状的锁式探针就可以在接下来的扩增反应中得到扩增^[49]。对扩增产物或扩增信号进行分析,就可以判断连接体系中是否存在检测靶 DNA。锁型探针以前是用来获取外显子并进行测序的^[50]。

亚硫酸盐结合锁式探针法的基本程序是^[51]:首先合成长约 150 nt 的寡核苷酸链,酶切裂解转换成锁式探针;然后锁式探针文库与经重亚硫酸盐处理的 DNA 退火,延伸,再与 5'端连接环化,并酶切除线状 DNA 分子;最后用一对常用引物对所有的环状锁型探针进行 PCR 扩增,鸟枪法测序。

此方法有两个主要困难^[51]:亚硫酸盐处理 DNA 使所有未甲基化的胞嘧啶转换成了尿嘧啶,使序列复杂性显著减少,因此从亚硫酸盐处理的 DNA 中获得特异性的序列比天然 DNA 困难;此方法获取灵敏度低、偏差高及等位基因随机缺失。因此,用目前存在的方法获得精确高效的 DNA 甲基化分析是不可能的。尤其是无法解决等位基因的缺失问题。此外合成探针数量较多,工作量较大,价格昂贵。

1.2.7 结合新一代测序技术的高通量单碱基解析 新一代测序技术的出现,使全基因组测序单碱基分析成为可能。可以通过高通量测序对比检测基因组 DNA 甲基化程度及位点^[52]。此方法的优势在于能够对全基因组甲基化进行精确分析。最近,重亚硫酸盐结合新一代测序技术完成了拟南芥约 120 Mb 基因组 DNA 甲基化图谱分析^[53]。

1.3 甲基化 DNA 免疫共沉淀 (methyl-DNA immunoprecipitation, MeDIP) MeDIP 方法是用 5-甲基胞嘧啶特异性抗体或者

用含有甲基结合结构域的蛋白通过免疫沉淀富集基因组甲基化或未甲基化片段^[23, 54], 其基本程序是: (1) 提取全基因组 DNA, 经超声波打断成长度为 400–500 bp 的片段; (2) 加热变性, 将变性的单链 DNA 样品平均分为两份; (3) 其中一份加入甲基化 DNA 特异性抗体, 另一份作为 Total input DNA 样品; (4) 使用亲和层析分离上一步样品中的甲基化 DNA 片段的抗体复合物, 样品中其余的非甲基化 DNA 片段被洗脱, 纯化得到甲基化 DNA 片段 (MeDIP DNA)。得到纯化的甲基化 DNA 片段后可以结合荧光定量 PCR 技术定量检测 DNA 甲基化, 也可结合芯片技术检测基因甲基化。

1.3.1 MeDIP 结合实时荧光定量 PCR 技术 (MeDIP-qPCR) 实时荧光定量 PCR 技术是在 PCR 反应体系中加入荧光试剂, 利用荧光信号实时检测 PCR 进程, 最后通过标准曲线对未知模板进行定量分析的方法。实时荧光定量 PCR 能够专一、灵敏、快速、高重复地精确定量起始模板的浓度。

实时定量 PCR 技术与 MeDIP 技术结合, 先通过 MeDIP 技术富集甲基化的 DNA 片段, 再利用实时荧光 PCR 技术进行检测, 经过数据分析处理得到检测的生物样品中特定定位点的甲基化水平的精确数值。

此方法的特点是: 高特异性; 不需要标记探针; 精确度高、重复性好; 灵敏度高。

1.3.2 MeDIP 结合芯片技术 (MeDIP-chip) 甲基化免疫共沉淀结合芯片技术基本程序是^[29]: 用 MeDIP 的方法富集甲基化的 DNA 片段, 与全基因组对照分别标记后 (MeDIP – Cy3, Input – Cy5) 混合与设计的芯片杂交, 用高解析度芯片扫描仪检测杂交信号; 对杂交结果进行数据提取、标准化、峰值分析和报告。

但到目前为止不能通过 MeDIP 方法准确估计甲基化水平, 而且低 CpG 密度区域甲基化分析存在问题。

1.4 单分子实时测序法直接检测 DNA 甲基化 (single-molecular, real-time sequencing, SMRT)

SMRT 是最新开发的方法, 由 Flusberg 等^[55]提出, 本方法不需要亚硫酸盐处理, 而利用 DNA 聚合酶进行边合成边收集荧光信号的方法进行测序。其原理是: DNA 聚合酶催化荧光标记的核苷酸结合到

核苷酸链上, 当核苷酸掺入时, 通过荧光脉冲到达和持续的时间检测可以获得聚合酶动力学信息, 从而可以直接测定 DNA 模板上的核苷酸修饰, 包括: N6-甲基腺嘌呤, 5-甲基胞嘧啶, 5-羟甲基化胞嘧啶。

Pacific Biosciences 公司发明了一种直径只有几十纳米的纳米孔 (zero-mode waveguides, ZMWs), 单分子的 DNA 聚合酶被固定在这个孔内。聚合酶链式反应开始时由于荧光标记的 A、T、C 和 G 较快速地从外面进入到孔内又出去, 它们形成了较稳定的背景荧光信号。当这些荧光标记的脱氧核苷酸被掺入 DNA 链的时候, 它的荧光就同时能在 DNA 链上探测到。当它与 DNA 链形成化学键的时候, 它的荧光基团就被 DNA 聚合酶切除, 荧光消失。这种荧光标记的脱氧核苷酸不会影响 DNA 聚合酶的活性, 并且在荧光被切除之后, 合成的 DNA 链和天然的 DNA 链完全一样。

该技术还需进行优化, 第一个是把双链 DNA 环化反复测序, DNA 聚合酶能够以环化的 DNA 作为模板滚环复制, 反复测一段 DNA 序列。这种反复测序, 纠正了偶尔出现的复制错误, 从而使测序精度非常高, 准确率达 99.99%。第二个是激发光中断测序法: DNA 聚合酶虽然很稳定, 但是在强大的激发光作用下酶也是有一定寿命的。如果把激发光中断一段时间, 在这段时间内 DNA 聚合酶继续复制 DNA, 当激发光重新开启以后, 就可以测到长 DNA 链后面的序列。

1.5 依赖于单分子纳米孔技术的测序 (single-molecule nanopore DNA sequencing)

该方法由英国 Oxford Nanopore Technologies 的科学家在 2010 年初提出^[56], 单分子纳米孔测序仪能直接分辨出未修饰的胞嘧啶和甲基化胞嘧啶。其原理是当核酸外切酶消化单链 DNA 后, 单个碱基落入孔中, 它们瞬间与环式糊精相互作用, 并阻碍了穿过孔中的电流。每个碱基 A、T、G、C 以及甲基胞嘧啶都有自己特有的电流振幅, 每个碱基也有特有的平均停留时间, 它的解离速率常数是电压依赖的, +180 mV 的电位能确保碱基从孔的另一侧离开, 因此很容易转化成 DNA 序列。

以往对甲基胞嘧啶进行测序, 都要先进行重亚硫酸盐转化, 纳米孔技术就能直接读出这第 5 种碱

基。目前, 纳米孔测序仅限于短的寡核苷酸。此外, Oxford Nanopore 的测序仪仍面临两个重要的技术问题: 一是如何将核酸外切酶更好地附着在孔上, 让它每次以正确的序列顺序只进入一个碱基, 然后从另一侧离开, 这是一个大挑战; 另一个是并行化, 这个问题可能简单一些, 可以开发出一个芯片, 上面有数十万个孔, 来确保整个测序过程更快速。不过, 一旦成功, 第 5 种碱基 5mC 的直接测序将会产生重大影响。

2 组蛋白修饰研究技术

组蛋白修饰较多, 包括甲基化、乙酰化、磷酸化、泛素化、SUMO 化和 ADP 核糖基化等。但目前组蛋白修饰研究方法较少, 目前最常用的为染色质免疫共沉淀技术。

染色质免疫沉淀技术 (chromatin Immunoprecipitation, ChIP) 由 O'Neill 和 Turner^[57] 提出, 是研究体内蛋白质与 DNA 相互作用的一种技术。它的基本原理是在活细胞状态下固定蛋白质—DNA 复合物, 然后通过超声或酶处理将染色质切断为一定长度范围内的染色质小片段, 然后通过免疫学方法沉淀此复合体, 特异性地富集与目的蛋白结合的 DNA 片段, 通过对目的片段的纯化与检测, 从而获得蛋白质与 DNA 相互作用的信息。染色质免疫沉淀技术一般包括细胞固定, 染色质断裂, 染色质免疫沉淀, 交联反应的逆转, DNA 的纯化, 以及 DNA 的鉴定。

ChIP 不仅可以检测体内反式因子与 DNA 的动态作用, 还可以用来研究组蛋白的各种共价修饰与基因表达的关系。而且, ChIP 与其他方法的结合, 扩大了其应用范围: ChIP 与基因芯片相结合建立的 CHIP-chip 方法已广泛用于特定反式因子靶基因的高通量筛选^[58, 59]; ChIP 与体内足迹法相结合, 用于寻找反式因子的体内结合位点^[60, 61]; RNA-ChIP 用于研究 RNA 在基因表达调控中的作用^[62, 63]。染色质免疫共沉淀结合芯片技术 (CHIP-chip) 和染色质免疫共沉淀结合新一代短序列测序技术 (ChIP-seq) 是目前检测组蛋白修饰最常用的方法。

2.1 染色质免疫共沉淀结合芯片技术 (ChIP-chip)

ChIP-chip 是将 ChIP 与生物芯片相结合, 在全基因组或基因组较大区域上高通量分析 DNA 结合位点或组蛋白修饰的方法^[64]。该技术获得的信息

量主要取决于芯片的探针的密度、分辨率与覆盖度。探针密度指生物芯片表面所固定的 DNA 探针的数量。分辨率指设计生物芯片时两个相邻探针的 DNA 序列在基因组上相隔的距离, 分辨率越高, 相邻探针之间的距离越短。覆盖度指固定在生物芯片上 DNA 序列占基因组全序列的比例。

该技术的基本程序是: 先通过染色质免疫共沉淀技术 (ChIP) 富集组蛋白被修饰的 DNA 片段, 然后加上通用接头进行 PCR 扩增, 在扩增过程中引入荧光基团。由于富集的片段长短不同, 所以扩增效率不同, 通过控制循环数来减少偏好性。最后将扩增的片段与设计的芯片杂交。

杂交可通过两种方法, 一种是单杂交法: 对照组 (未经免疫沉淀富集的基因组 DNA) 与试验组分别与芯片杂交, 然后对比; 另一种是双色竞争法: 用另一种颜色的荧光标记对照组, 对照组和试验组同时与设计的芯片竞争性杂交, 通过两种信号强弱对比得出该位点的修饰程度。

2.2 染色质免疫共沉淀结合短序列测序技术 (ChIP-seq)

ChIP-seq 是将 ChIP 与测序技术相结合, 在全基因组范围内检测 DNA 组蛋白修饰的高通量方法, 可以应用到任何基因组序列的物种, 并能确切得到每一个片段的序列信息^[64]。

该技术的基本程序是: 通过 ChIP 富集目的片段, 纯化后加上通用接头进行 PCR 扩增, 最后加 sol-exa 接头进行测序。目前该技术比较成熟, 通量也在不断提高, 成本随着新一代测序技术的出现和发展逐步降低, ChIP 和测序技术的结合越来越广泛的应用到 DNA 与互作蛋白分析。

该技术的主要困难在于测序完成后对海量数据的分析。并且各个环节的差别, 如 DNA 质量、获取的片段长短不同导致的扩增效率差异、基因组的重复程度以及测序和序列比对过程中的错误都会引入系统误差造成假阳性^[65]。

2.3 组蛋白泛素化和 SUMO 化修饰研究

组蛋白泛素化和 SUMO 化研究中, 至今没有针对这种组蛋白修饰的特异性抗体, 虽然有报道 H2B 泛素化多克隆抗体, 但没有商用^[66]。Beger 实验室尝试用分支肽开发 H2B-SUMO 抗体也未获得成

功^[20],而且两种修饰在原生质细胞萃取物中不稳定且易被降解。所以必须开发独特的方法用于组蛋白泛素化和 SUMO 化分析,目前已有利用组蛋白、泛素上含有抗原决定簇标签的独特酵母菌株研究组蛋白泛素化和 SUMO 化的报道^[12]。

3 讨论

表观遗传调控是生物生长发育重要的调控机制,其中 DNA 甲基化和组蛋白修饰是表观遗传学研究的重要内容。目前 DNA 甲基化检测方法较多,主要分为 3 类^[21-25]: 依赖于限制性内切酶的检测技术; 依赖亚硫酸盐转换的检测技术; 结合免疫方法的检测技术。在这 3 类技术中依赖于限制性内切酶的方法比较简便易行,且不需知道被测 DNA 的序列信息,在不同生物上具有通用性,可用于 DNA 序列背景知识未知的生物全基因组范围检测 CCGG 位点的胞嘧啶甲基化变化,但局限性较大,只能检测部分 DNA 甲基化,如 CpG 位点或者 CpG 岛的甲基化。依赖亚硫酸盐转换的检测技术被视为检测 DNA 甲基化的金标准,亚硫酸盐处理后未甲基化 C 转换为 U,甲基化胞嘧啶保持不变,经扩增后未甲基化 C 变为 T,甲基化 C 保持不变,通过测序对比即可即可定性定量检测 DNA 甲基化^[22]。以亚硫酸盐转换为基础开发的方法较多,各有优缺点。优点是此方法检测比较准确,很容易确定个体基因组 DNA 分子中各个单链的甲基化状态及甲基化位置,结合不同技术可灵活检测全基因组或特殊序列位点甲基化状态。缺点是亚硫酸盐转换后序列特异性降低,不易设计特异性引物,获取特异性序列比未处理的 DNA 困难; 在结合某些方法时容易出现假阳性,如 COBRA 方法^[33-34]; 且全基因组检测价格比较昂贵。结合免疫方法的技术能准确检测全基因组任何位点 DNA 甲基化,不受序列限制,也不会像亚硫酸盐转换方法降低序列的特异性,可以进行高通量检测,不需要标记探针,精确度高,灵敏度高。此方法结合芯片或高通量测序是目前最常用的 DNA 甲基化检测方法。缺点是步骤较多,易出错。目前市场有较多试剂盒可用,大大减少了步骤出错率。此方法结合芯片技术目前尚不能准确估计甲基化水平,低 CpG 密度区域甲基化分析也存在问题^[29]。

这 3 类检测方法可以相互结合或者与芯片、高

通量测序等其他技术结合灵活检测样品 DNA 甲基化。如限制性内切酶结合高通量测序技术、限制性内切酶结合芯片技术、亚硫酸盐转换结合高通量测序技术、亚硫酸盐转换结合芯片技术、亚硫酸盐转换结合限制性内切酶技术、免疫沉淀结合高通量测序技术、免疫沉淀结合芯片技术和免疫沉淀结合荧光定量技术等。

除了这 3 类较成熟的技术以外,最近开发的两种新型测序技术也可用于 DNA 甲基化检测。一个是单分子、实时测序技术 (SMRT)^[55],根据合成 DNA 链时聚合酶停留的时间检测 DNA 是否甲基化,目前仪器已经问世; 另一个是单分子纳米孔测序技术^[56]。根据各种碱基通过单分子纳米孔时都有自己独特的电位变化进行 DNA 测,当 5mC 通过纳米孔引起的电位跟其他 4 种碱基不同,可以精确检测 DNA 链上甲基化位点。这两类技术的发展大大缩短了测序时间,减少了测序成本。

组蛋白修饰研究方法相对较少,而且研究重点主要集中在组蛋白甲基化和乙酰化上。目前主要用染色质免疫共沉淀结合芯片的方法 (ChIP-chip) 和结合短序列测序技术 (ChIP-seq)^[64]。ChIP-chip 方法获得的信息量主要取决于芯片的探针密度、分辨率与覆盖度。ChIP-seq 技术比较成熟,可以应用到任何基因组序列的物种,并能确切得到每一个片段的序列信息,但是数据量较大,且易出现假阳性。

组蛋白泛素化和 SUMO 化研究中,至今没有它们的特异性抗体,且这两种修饰对原生质萃取物中的蛋白质水解酶敏感,所以必须开发独特的方法用于组蛋白泛素化和 SUMO 化分析^[12]。

参考文献

- [1] Egger G, Liang C, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, 2004, 429 (6990): 457-63.
- [2] Bird A. Perceptions of epigenetics. *Nature*, 2007, 447 (7143): 396-8.
- [3] Bender J. DNA methylation and epigenetics. *Annu Rev Plant Biol*, 2004, 55: 41-68.
- [4] Paszkowski J, Whitham SA. Gene silencing and DNA methylation processes. *Curr Opin Plant Biol*, 2001, 4(2): 123-9.
- [5] Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*, 2002, 16(1): 6-21.

- [6] Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem*, 2005, 74: 481-514.
- [7] Cao X, Jacobsen SE. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. *Proc Natl Acad Sci USA*, 2002, 99(Suppl 4): 16491-8.
- [8] Hsieh TF, Fischer RL. Biology of chromatin dynamics. *Annu Rev Plant Biol*, 2005, 56: 327-51.
- [9] Chen CM, Chen HL, Hsiau TH, et al. Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes. *Am J Pathol*, 2003, 163(1): 37-45.
- [10] Zhang X, Yazaki J, Sundaresan A, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 2006, 126(6): 1189-201.
- [11] Li X, Wang X, He K, et al. High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell*, 2008, 20(2): 259-76.
- [12] Trujillo KM, Tyler RK, Ye C, et al. A genetic and molecular toolbox for analyzing histone ubiquitylation and sumoylation in yeast. *Methods*, 2011, 54(3): 296-303.
- [13] Mellor J. It takes a PHD to read the histone code. *Cell*, 2006, 126(1): 22-4.
- [14] Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol*, 2005, 6(11): 838-49.
- [15] Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, 129(4): 823-37.
- [16] Sims RJ 3rd, Reinberg D. Histone H3 Lys 4 methylation: caught in a bind?. *Genes Dev*, 2006, 20(20): 2779-86.
- [17] 赵树靛, 房静远. 组蛋白磷酸化的机制及其作用研究进展. *细胞生物学杂志*, 2009(2): 178-182.
- [18] Boulukas T. DNA strand breaks alter histone ADP-ribosylation. *Proc Natl Acad Sci USA*, 1989, 86(10): 3499-503.
- [19] Tanny JC, Erdjument-Bromage H, Tempst P, et al. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev*, 2007, 21(7): 835-47.
- [20] Nathan D, Ingvarsdottir K, Sterner DE, et al. Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes Dev*, 2006, 20(8): 966-76.
- [21] Rollins RA, Haghghi F, Edwards JR, et al. Large-scale structure of genomic methylation patterns. *Genome Res*, 2006, 16(2): 157-63.
- [22] Frommer M, McDonald LE, Millar DS, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA*, 1992, 89(5): 1827-31.
- [23] Nair SS, Coolen MW, Stirzaker C, et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain(MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics*, 2011, 6(1): 34-44.
- [24] Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*, 2005, 37(8): 853-62.
- [25] Keshet I, Schlesinger Y, Farkash S, et al. Evidence for an instructive mechanism of *de novo* methylation in cancer cells. *Nat Genet*, 2006, 38(2): 149-53.
- [26] Gebhard C, Schwarzfischer L, Pham TH, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res*, 2006, 66(12): 6118-28.
- [27] Rauch T, Li H, Wu X, et al. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res*, 2006, 66(16): 7939-47.
- [28] Illingworth R, Kerr A, Desousa D, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol*, 2008, 6(1): e22.
- [29] Down TA, Rakyan VK, Turner DJ, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol*, 2008, 26(7): 779-85.
- [30] Cervera MT, Ruiz-Garcia L, Martinez-Zapater JM. Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP markers. *Mol Genet Genomics*, 2002, 268(4): 543-52.
- [31] McClelland M, Nelson M, Raschke E. Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res*, 1994, 22(17): 3640-59.
- [32] Sutherland E, Coe L, Raleigh EA. MspBC: a multisubunit GTP-dependent restriction endonuclease. *J Mol Biol*, 1992, 225(2): 327-48.
- [33] 顾婷婷, 张志明, 郑鹏生. DNA 甲基化研究方法的回顾与评价. *中国妇幼健康研究*, 2006(6): 555-560 564.
- [34] Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res*, 1997, 25(12): 2532-4.
- [35] Boyko A, Kovalchuk I. Analysis of locus-specific changes in methylation patterns using a COBRA(combined bisulfite restriction analysis) assay. *Methods Mol Biol*, 2010, 631: 23-31.
- [36] Eads CA, Laird PW. Combined bisulfite restriction analysis(COBRA). *Methods Mol Biol*, 2002, 200: 71-85.
- [37] Brena RM, Auer H, Kornacker K, et al. Accurate quantification of

- DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. *Nucleic Acids Res*, 2006, 34(3): e17.
- [38] Herman JG, Graff JR, Myohanen S, et al. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA*, 1996, 93(18): 9821-6.
- [39] Licchesi JD, Herman JG, Methylation-specific PCR. *Methods Mol Biol*, 2009, 507: 305-23.
- [40] Derks S, Lentjes MH, Hellebrekers DM, et al. Methylation-specific PCR unraveled. *Cell Oncol*, 2004, 26(5-6): 291-9.
- [41] Gonzalgo ML, Jones PA. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension(Ms-SNuPE). *Nucleic Acids Res*, 1997, 25(12): 2529-31.
- [42] 李克勇, 肖武生, 常秀丽, 等. 基于 PCR 技术的 DNA 甲基化检测方法研究进展. *国外医学(卫生学分册)*, 2008(5): 314-318.
- [43] Gonzalgo ML, Liang G. Methylation-sensitive single-nucleotide primer extension(Ms-SNuPE) for quantitative measurement of DNA methylation. *Nat Protoc*, 2007, 2(8): 1931-6.
- [44] Kuppaswamy MN, Hoffmann JW, Kasper CK, et al. Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B(factor IX) and cystic fibrosis genes. *Proc Natl Acad Sci USA*, 1991, 88(4): 1143-7.
- [45] Liu WO, Oefner PJ, Qian C, et al. Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders. *Genet Test*, 1997, 1(4): 237-42.
- [46] 邓大君, 邓国仁, 吕有勇, 等. 变性高效液相色谱法检测 CpG 岛胞嘧啶甲基化. *中华医学杂志*, 2001, 81(3): 158-161.
- [47] Nilsson M, Malmgren H, Samiotaki M, et al. Landegren, Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science*, 1994, 265(5181): 2085-8.
- [48] 黄冠军, 殷幼平, 张仑, 等. 柑桔溃疡病菌滚环扩增检测体系的建立. *微生物学报*, 2008(3): 375-379.
- [49] Szemes M, Bonants P, de Weerd M, et al. Diagnostic application of padlock probes—multiplex detection of plant pathogens using universal microarrays. *Nucleic Acids Res*, 2005, 33(8): e70.
- [50] Porreca GJ, Zhang K, Li JB, et al. Multiplex amplification of large sets of human exons. *Nat Methods*, 2007, 4(11): 931-6.
- [51] Deng J, Shoemaker R, Xie B, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol*, 2009, 27(4): 353-60.
- [52] Lister R, O'Malley RC, Tonti-Filippini J, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 2008, 133(3): 523-36.
- [53] Cokus SJ, Feng S, Zhang X, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 2008, 452(7184): 215-9.
- [54] Jacinto FV, Ballestar E, Esteller M. Methyl-DNA immunoprecipitation(MeDIP): hunting down the DNA methylome. *Biotechniques*, 2008, 44(1): 35, 37, 39 passim.
- [55] Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*, 2010, 7(6): 461-5.
- [56] Clarke J, Wu HC, Jayasinghe L, et al. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*, 2009, 4(4): 265-70.
- [57] O'Neill LP, Turner BM. Immunoprecipitation of chromatin. *Methods Enzymol*, 1996, 274: 189-97.
- [58] Miura H, Tomaru Y. [ChIP on chip for transcriptional regulatory network analysis]. *Tanpakushitsu Kakusan Koso*, 2004, 49(17 Suppl): 2710-6.
- [59] Blais A, Dynlacht BD. Devising transcriptional regulatory networks operating during the cell cycle and differentiation using ChIP-on-chip. *Chromosome Res*, 2005, 13(3): 275-88.
- [60] Testa A, Donati G, Yan P, et al. Chromatin immunoprecipitation (ChIP) on chip experiments uncover a widespread distribution of NF- κ B binding CCAAT sites outside of core promoters. *J Biol Chem*, 2005, 280(14): 13606-15.
- [61] Sandmann T, Jakobsen JS, Furlong EE. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nat Protoc*, 2006, 1(6): 2839-55.
- [62] Nagino K, Nomura O, Takii Y, et al. Ultrasensitive DNA chip: gene expression profile analysis without RNA amplification. *J Biochem*, 2006, 139(4): 697-703.
- [63] Li T, Xie Z, Wang Y, et al. A new combination of RNA-mediated DNA ligation and on-chip elongation for detecting viral RNA. *Diagn Microbiol Infect Dis*, 2008, 62(1): 26-33.
- [64] 李敏俐, 王薇, 陆祖宏. ChIP 技术及其在基因组水平上分析 DNA 与蛋白质相互作用. *遗传*, 2010, 32(3): 219-228.
- [65] Nix DA, Courdy SJ, Boucher KM. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics*, 2008, 9: 523.
- [66] Nakanishi S, Lee JS, Gardner KE, et al. Shilatifard, Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. *J Cell Biol*, 2009, 186(3): 371-7.

(责任编辑 狄艳红)