

DOI: 10.3724/SP.J.1005.2010.00219

ChIP 技术及其在基因组水平上分析 DNA 与蛋白质相互作用

李敏俐¹, 王薇¹, 陆祖宏^{1,2}

1. 东南大学生物电子学国家重点实验室, 南京 210096;
2. 东南大学儿童发展与学习科学教育部重点实验室, 南京 210096

摘要: 染色质免疫沉淀(Chromatin immunoprecipitation, ChIP)技术是分析细胞内生理状态下 DNA 结合蛋白与基因组 DNA 相互作用的技术。ChIP 与高密度芯片(ChIP-chip)或高通量测序(ChIP-Seq)相结合能产生大量的研究数据, 在细胞的基因表达调控网络研究中发挥重要作用。文章主要介绍 ChIP、ChIP-chip 和 ChIP-Seq 的技术特点以及发展趋势, 重点讨论了 ChIP-Seq 数据分析方法及相关的实例。

关键词: 基因组; ChIP-Seq; ChIP-chip

Genomic analysis of DNA-protein interaction by chromatin immunoprecipitation

LI Min-Li¹, WANG Wei¹, LU Zu-Hong^{1,2}

1. The State Key Laboratory of Bioelectronics, Southeast University, Nanjing 210096, China;
2. Key Laboratory of Child Development and Learning Science, Ministry of Education, Southeast University, Nanjing 210096, China

Abstract: Chromatin immunoprecipitation (ChIP) is an effective technique to analyze the interactions of DNA binding proteins with the genome *in vivo*. ChIP coupled with high density microarray (ChIP-chip) or high-throughput sequencing (ChIP-Seq) has generated large amount of data and expected to allow the development of a network describing the cellular transcriptional regulation. Here, we reviewed the ChIP, ChIP-chip, and ChIP-Seq techniques as well as their perspectives. Focus is given to data analysis of ChIP-Seq and the applications of ChIP-chip and ChIP-Seq.

Keywords: genome; ChIP-Seq; ChIP-chip

真核生物基因组DNA以染色质形式存在, 研究蛋白质与DNA在染色质环境下的相互作用是阐明真核生物基因表达调控机制的基本途径。生物体内基因表达调控主要发生在转录过程中, 转录调控是顺式作用元件(*Cis*-acting elements)如启动子(Promoter)、增强子(Enhancer)与反式作用因子(*Trans*-acting factors)

相互作用的结果。基因组DNA的甲基化、组蛋白甲基化、乙酰化和磷酸化修饰, 核小体重新定位及染色体结构重建都影响调控。转录调控具有细胞类型、发育阶段和外界环境刺激的差异性, 哺乳动物转录调控序列分散在较大区域, 组蛋白修饰状态达 100 多种, 这些因素都增加了转录调控的复杂性。

收稿日期: 2009-06-23; 修回日期: 2009-08-27

基金项目: 国家自然科学基金项目(编号: 30871393)资助

作者简介: 李敏俐(1972-), 女, 硕士, 研究方向: 基因组学。Tel: 025-83793110; E-mail: lml@seu.edu.cn

王薇(1985-), 女, 硕士, 研究方向: 生物信息学。Tel: 025-83793110; E-mail: netasha@163.com

通讯作者: 陆祖宏(1960-), 男, 教授, 研究方向: 基因组学。Tel: 025-83793779; E-mail: zhlu@seu.edu.cn

ENCODE计划^[1]和表观基因组学研究计划^[2]是研究基因表达调控奥秘的国际合作项目。

了解基因转录调控的关键是识别蛋白质与DNA的相互作用,ChIP是全基因组范围内识别DNA与蛋白质体内相互作用的标准方法^[3],最初用于组蛋白修饰研究^[4],后来用于转录因子^[5]。ChIP与芯片(Chip, microarray)或测序(Sequencing)相结合的ChIP-chip和ChIP-Seq技术是在全基因组水平分析DNA与蛋白质相互作用的两大主要技术。本文将介绍ChIP、ChIP-chip和ChIP-Seq的技术特点及发展趋势,重点讨论ChIP-Seq的数据分析以及应用实例。

1 ChIP、ChIP-chip 和 ChIP-Seq 的技术特点

1.1 ChIP

ChIP技术由Orlando等^[6]于1997年创立。其基本原理为将处于适当生长时期的活细胞用甲醛交联后将细胞裂解,染色体分离并打碎为一定大小的片段;然后用特异性抗体免疫沉淀目标蛋白与DNA交联的复合物,对特定靶蛋白与DNA片段进行富集。采用低pH值条件反交联,DNA与蛋白质之间的Schiff键水解,释放DNA片段。通过对目标片段的纯化与检测,获得DNA与蛋白质相互作用的序列信息。

在上述ChIP过程中,甲醛能够进入细胞并使蛋白质与DNA或蛋白质与蛋白质之间通过希弗(Schiff)键交联,形成稳定结合的复合物。如果交联效果不太好,可以先用交联剂DMA(Dimethyl adipimide)^[7]或DSG(Disuccinimidyl glutarate)^[8]处理细胞,以加强后续甲醛交联的效果。破碎DNA可采用超声物理破碎或特定酶切消化,以获得所需长度的DNA片段。由于DNA片段长度将影响抗体免疫沉淀效率,因此破碎DNA是ChIP实验成功与否的重要因素。超声效果与细胞裂解是否充分、细胞浓度及裂解液成分等因素有关。超声处理后的液体应从浑浊状态变为透明状态^[9]。

选择专一性及亲和力较高的抗体是ChIP成功的关键。非特异性抗体将增加大量的非目标靶点DNA片段信息,从而掩盖了真实的蛋白质结合位点信息;而亲和力较差的抗体,则无法获得高信噪比靶点DNA片段^[10]。另一方面,在甲醛交联过程中可能会掩盖一些蛋白质的表位,这会影响到一部分蛋白质和DNA复合体的免疫沉淀反应。因此,用Western印

迹或免疫组化等常用的实验方法证明的能够对目标蛋白质进行免疫结合的抗体,并不能保证一定能够成功地进行ChIP实验。例如,Weitsman等^[11]检验了不同的雌激素受体 β 抗体在ChIP中的免疫沉淀能力,发现有的抗体虽然能够在标准免疫沉淀条件下与抗原结合,但不适合ChIP条件下使用,并且不同抗体的结合效率也有差异。对于某一目标蛋白采用多种抗体组合来进行ChIP实验是一种有效的方案。例如,Chen等^[12]针对CTCF在人类基因组中的结合位点研究,采用了多种不同单克隆抗体的混合物进行免疫沉淀。另外,有人利用表位附加标记技术表达融合蛋白进行免疫沉淀实验。常用的表位附加标记有红血球凝集素(Hemagglutinin)^[13]、生物素^[14]和c-Myc^[15]等。但哺乳动物细胞中难以引入表位附加标记的融合蛋白,并且其表达往往不同于其内源转录因子。

染色质免疫沉淀获得的DNA数量往往很多,其中包含大量的非特异结合造成的背景噪音,采用严格的洗涤条件可以降低背景。Ogryzko等^[16]将靶蛋白与生物素受体片段融合表达,通过抗生物素磁珠免疫沉淀,二者较强的结合力便于进行更为严谨的洗涤,达到降低背景的目的。

1.2 ChIP-chip

ChIP-chip是将生物芯片平台与ChIP实验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位点的方法。ChIP-chip获得的信息量主要取决于芯片表面固定的探针数量。探针的密度、分辨率与覆盖度是生物芯片的重要特征。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。最理想的芯片是能够覆盖全基因组,但是由于人的基因组巨大,尚不可能在一个芯片上固定覆盖人类全基因组所有的探针。芯片制作方法主要有两种:一种是将合成的寡核苷酸点到芯片上。点样方法的优点是制作方便,但存在可能由于点样针头洗涤不彻底而造成交叉污染及样点不均一等问题;另一种是原位合成法,即寡核苷酸探针在芯片上原位“生长”,如Affimetrix的“光蚀刻原位合成技术”,Nimblegen公司的“无掩膜芯片合成技术(Maskless array synthesis)”。原位合

成法具有较好的重复性和较高的密度。

芯片杂交所需样品量较大,要对免疫沉淀后的DNA进行PCR扩增以达到杂交要求。常用的扩增方法是LM-PCR(Ligation-mediated PCR),将反交联并纯化过的DNA用DNA聚合酶补平后再连接通用接头(Linker),然后用与通用接头互补的引物进行扩增,在扩增过程中引入荧光基团。由于免疫沉淀后的DNA片段长度不同,每种长度片段的百分比含量不同,PCR的扩增效率不同。对于中等片段DNA这种方法能够进行等比例扩增,但对于较长片段的DNA扩增效率将降低。一般通过控制PCR循环数,以减少扩增产生的偏向性(Bias)。另外, Liu等^[17]发展了一种基于T7的线性扩增方法,先用末端转移酶将一系列dT加到DNA末端,然后以polyA-T7 序列为引物进行DNA复制,双链DNA在T7 RNA聚合酶的作用下体外转录,然后反转录。这种方法虽然费时费力,但获得DNA的线性扩增度好,减少了不同序列之间扩增的偏向性。

芯片杂交检测可用双色竞争法。ChIP 获得的DNA 与对照(Control)DNA 分别用不同颜色标记并竞争性与芯片杂交,根据信号强弱判断哪些位点是免疫沉淀富集的位点。对照 DNA 通常来自超声破碎后不加特异抗体直接保留下的 DNA 或与特异抗体同物种的 IgG 免疫沉淀的 DNA 为对照。如果采用单色杂交检测,则需要分别检测染色质免疫沉淀 DNA 与对照 DNA 芯片杂交的结果。

1.3 ChIP-Seq

ChIP-Seq是将深度测序技术与ChIP实验相结合分析全基因组范围内DNA结合蛋白结合位点、组蛋白修饰、核小体定位或DNA甲基化的高通量方法,可以应用到任何基因组序列已知的物种,并能确切得到每一个片段的序列信息。传统的Sanger测序由于测序成本和速率限制,分析大量的免疫沉淀DNA序列从时间和花费上都是不现实的。在深度测序技术出现之前,为了减少测序量,人们提出基于SAGE(Serial analysis of gene expression)技术的测序方法,如GMAT(Genome-wide mapping technique)^[18], SACO(Serial analysis of chromatin occupancy)^[19], STAGE(Sequence tag analysis of genomic enrichment)^[20], SABE(Serial analysis of binding elements)^[21]。这些方法基本原理类似,都是将获得的免疫沉淀片段的末

端切割并相互连接形成一条相对较长的片段进行Sanger测序,用部分片段的信息代替整条片段,降低了测序量,但费用依然较高,并且无法对蛋白质的结合位点进行精确定位。

ChIP-PET(Paired-end diatag)首次应用深度测序技术—454 焦测序(Pyrosequencing)技术进行片段分析^[22],但免疫沉淀片段的处理方法依然是基于SAGE的原理,与上述4种方法的区别是将免疫沉淀片段的两个末端同时切割获得“双标签”并连接测序,以片段的两端信息定位整条片段优于“单标签”定位,提高了定位能力,解决了以往全长测序所带来的高花费、效率低的问题。基于SAGE获得的标签中,还有一小部分可能由于测序错误、克隆假象(Artifact)或基因组序列不完整等原因而无法定位到基因组上。能够定位到基因组上的标签的丰度及其在基因组上的分布用于估计蛋白质结合位点,一般认为标签重叠成“簇(Cluster)”的基因组区域为结合位点,但有些单一出现的也可能是结合位点,所以只考虑成簇位点可能会忽略某些真正的位点。理论上,较多标签的测序有利于最终识别所有的结合位点,但究竟多少标签才能真正获得专一性与特异性并无定论。

随着深度测序技术的迅速发展,染色质免疫沉淀与 DNA 测序直接结合已越来越多地应用在全基因组 DNA 与蛋白质作用分析。目前可用于 ChIP-Seq 分析的测序平台有 Genome Analyzer (Illumina, Solexa)、SOLiD (Applied Biosystem)、454-pyrosequencing (Roche)、HeliScope (Helicos)。每种测序方法在测序通量和成本上都在迅速地发展。可以预计,ChIP-Seq 技术将会很快发展成为该研究领域的主流技术。

染色质免疫沉淀技术是相对比较成熟的技术,ChIP-Seq的难点是测序后的生物信息学分析。DNA 打碎方法、染色质开放程度的不均一性、PCR扩增偏向性、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^[23],尽可能剔除假阳性并揭示出数据背后的机制是需要分子生物学与计算生物学工作者协同努力的问题。

测序后首先需要将序列比对到已知基因组上并确立真正的结合位点(峰, Peak)。对于转录因子,要寻找与“峰”对应的下游调控基因(靶基因),或者构建转录因子结合位点的保守结合序列(Motif);如果

转录因子的 motif 是已知的, 则可以计算“峰”序列中包含 motif 的序列百分比, 间接估计实验结果的可靠性。

1.3.1 确立真正的结合位点(峰)

ChIP-Seq 只能应用于基因组序列信息已知的物种, 并且由于个体基因组存在差异, 在将测序序列比对到基因组序列上要允许一定程度(小于等于 2 个)的错配^[24]。由于超声时随机打断染色质, 所以包含某个结合位点的超声片段可能有几条, 相互间起止点不同, 通过特异抗体免疫沉淀后, 这些片段被富集, 片段之间的重叠区域就是“峰”, 即 DNA 与蛋白质相互结合的位点。通常采用没有经过免疫沉淀的超声样品或物种特异 IgG 免疫沉淀样品作为负对照来剔除假阳性信号。Ji 等^[25]用 CisGenome 分析转录因子 OCT4 和 NANOG 有无负对照时“峰”的数量, 发现在不设立负对照时找到的“峰”较多, 但 motif 的百分比低, 可以推测其中包含有假阳性。

目前建立的适合找“峰”的统计分析软件有很多, 包括 PeakSeq^[26]、FindPeaks 3.1^[27]、F-Seq^[28]、SISSR^[29]、QuEST^[30]、MACS^[31]、Peak call^[24]、ChIPDiff^[32]、CisGenome^[25]、ChIP-seq processing pipeline^[33]。这些算法在结合位点的覆盖率、分辨率和假阳性控制方面各有特色。QuEST 和 SISSR 利用对照组数据剔除实验组中所找到的假阳性结合位点, 并把它系统地整合到其算法中, 而 FindPeaks 只将这种处理作为一种后续分析。另外, 这 3 个算法在统计某个区域内的标签序列数时, 所用的单位区间也不同: QuEST 采用的是核密度估计方法, 将标签序列所对应的位置为窗口中心, 赋予最高的密度值, 两侧则随核密度曲线平滑下降取密度值; 相比于 FindPeaks 直接将序列固定或动态向其下游延伸, 每个延伸位置赋予同样的分值要更合理; 核密度曲线的方法同时也比 SISSRs 用滑动窗口扫描的方法更加灵活省力, 不需要去关注那些没有任何标签序列数的区域, 从而减少了运算时间。

QuEST 未考虑单样本问题, 这与其相应的实验组的设计有紧密关系, 因为它全部采用对照组来控制实验结果; 而 FindPeaks 和 SISSRs 分别用蒙特卡罗模拟和泊松分布来估计背景噪声, 这两种方法的区别在于一种假设噪声在有效地可比的基因组范围

内服从均匀分布, 另一种则假设背景噪声服从泊松分布。在单样本情况下, 无论是假设背景噪声服从泊松分布还是用蒙特卡罗模拟背景噪声, 都是假设全基因组范围内噪声序列在各个位置出现的概率是相等的。然而经过 Zhang 等^[33]的拟合分析, 发现用幂函数分布模拟真实信号的同时用 gamma 分布模拟背景噪声信号的组合方式能够最佳地接近 STAT1 转录因子 ChIP-Seq 实验所得的真实信号模型。CisGenome 就是在考虑了局部的染色体结构差异很可能使得背景噪声不再均匀的情况下提出用负二项分布模型来取代泊松分布模型和蒙特卡罗模型, 这个模型允许随着基因组的坐标不同, 其背景序列数发生改变。可以说, 负二项分布模型更接近于真实的 ChIP-Seq 序列分布。与之类似的是, 考虑到扩增和测序、染色质开放性程度引入的局部偏差, MACS 用一种局部动态泊松分布来消除序列频数的局部差异, 提高了搜峰的特异性。PeakSeq 虽然也注意到背景噪声的不均匀性, 但是该算法用了一种间接的方法来解决这个问题。该算法作者认为 ChIP-Seq 实验中存在两类主要的偏向因素, 其一是基因组本身的序列重复性, 其二是染色质结构的开放性。针对第一种偏向性因素, PeakSeq 考察了基因组的某个特定长度的序列在每个位置的唯一性, 以此建立一张索引表用来修正染色体各片段之间的可比性差异。PeakSeq 用归一化来消除第二种偏向性因素的影响。用两种不同的方式来处理不同的偏差来源, 比起直接用某一种能捕获序列标签的全局分布(如负二项分布)来说, 可算是一种间接的方法。

ChIP-Seq 的生物信息处理除了对数据进行分析找到结合位点的富集区域并给出功能注释外, 还可以反馈出一些信息从而有助于 ChIP-Seq 实验的合理设计并优化。CisGenome 讨论了用单一正样本进行 ChIP-Seq 实验的优劣。Ji 等^[25]认为, 在整体的结合信号足够强的情况下, 单样本分析也可以提供合理的 FDR(False discovery rate)估计。反之, 在整体结合信号弱的情形下, 单样本的 FDR 估计会被低估。但是 Ji 也认为这种假阳性会被几个标准滤除, 如高重复片段区域, 缺乏 motif 富集区域将会增加这些峰的不可靠性从而将其剔除。Ji 给出的一个建议是在对某个转录因子的了解不够多或能够承担实验成本的情况下, 加上对照组的 ChIP-Seq 分析更加合适, 而当实验受

到成本限制时,对样本进行数据分析就应当谨慎并且应该用多重标准尽可能多地滤除假阳性。PeakSeq给出了ChIP-Seq实验的一些最优设计的思路,分析了序列饱和性,并通过数据分析得出结论认为必须至少进行两次ChIP-Seq重复来确保实验结果的可重复性。

1.3.2 “峰”与调控基因的对应

由于调控因子不仅以线性方式调控临近基因,而且可以通过染色质形成“loop”的方式调控相距较远的基因,甚至是不同染色体上的位点,所以确立与“峰”对应的下游调控基因比较困难。Johnson等^[24]将“峰”上下游 20 kb范围内的基因定义为具有调控关系的基因,而Wederell等^[34]发现大多数(68%)高度可信的“峰”位于基因转录起始位点上游 10 kb和 3' UTR下游 1 kb范围内。对染色质三维结构中不同位点的相互作用机制理解将有利于确立合理的“峰”与调控基因的对应关系。Dekker等^[35]创立的“chromosome conformation capture (3C)”及随后发展的chromosome conformation capture-on-chip (4C)^[36]、circular chromosome conformation capture (4C)^[37]、chromosome conformation capture carbon copy(5C)^[38]技术将为ChIP-Seq中“峰”与其调控基因对应问题提供依据。

2 ChIP-chip 与 ChIP-Seq 的应用

ChIP-chip 与 ChIP-Seq 是全基因组范围内分析转录因子或其它染色质蛋白结合位点及表观遗传机制的两大主要技术。相比而言,ChIP-Seq 具有定位分辨率高,成本不断降低及通量迅速增高,而且不受生物芯片上预先设定的基因组区域的限制等优势,越来越成为该领域的主流研究技术。

2.1 转录因子结合位点研究

ChIP-chip的应用受可获得的芯片种类的限制。最初用PCR扩增方法获得探针固定在芯片表面,这对于较小的基因组(如酵母)可以实现。Ren等^[39]用PCR方法扩增酵母 6 361 个基因间区域并将产物点样制作酵母基因组芯片,鉴定转录因子Gal4 和Ste12的结合位点,这是ChIP-chip首次在全基因组范围内的应用。随后,Iyer等^[40]和Horak等^[41]用相同的芯片构建方法相继鉴定了酵母中另外几个转录因子的结

合位点。Young研究组^[15]对 106 种转录因子进行c-myc附加标记并在 106 个酵母株系中表达,确定这些转录因子结合的启动子区域,确立了基因调控网络中复杂的相互作用,为构建酵母的基因调控网络提供了丰富的数据基础。

高等真核生物基因组较大,只能对局部区域采用PCR扩增的方法获得探针来构建芯片,如Horak等^[42]对哺乳动物 75 kb β 球蛋白区域构建芯片以研究GATA-1 的结合位点, Ren等^[43]设计包含 1 444 个人类基因上游序列的芯片来研究E2Fs的靶基因,随着人类基因组序列数据的进一步充实和注释的增加,他们又设计了覆盖当时已知的多数基因的 5' 端序列的芯片^[44]。Weimann等^[45]设计富含人类CpG岛的芯片。根据已有的序列设计探针具有偏向性,无法探知未知的没有注释区域的结合情况。为了克服偏向性, Martone等^[46]设计覆盖 22 号染色体剔除重复序列后的所有编码与非编码序列的芯片,鉴定Hela细胞在肿瘤坏死因子诱导下p65 的结合,第一次获得了一条染色体上转录因子无偏向的结合数据。

随着芯片技术的发展,现有各种商业化芯片可以选择,如不同物种的启动子芯片、CpG岛芯片、叠瓦芯片(Tiling array)等,也可以根据具体研究需要定制芯片。全基因组叠瓦芯片是按照一定的分辨率和探针长度设计的信息量较大的芯片。如Nimblegen公司提供的分辨率为 100 bp、探针长度为 50 bp的代表人类基因组的全套芯片共有 38 张。Kim等^[47]利用这套芯片鉴定转录起始复合体(PIC)4 种组分的基因组分布。

近几年ChIP-Seq迅速发展成为该领域的主流研究技术。Johnson等^[24]和Robertson等^[48]用Illumina测序平台测序分析转录因子NRSF和 STAT1 的免疫沉淀片段,这是最早的ChIP-Seq在转录因子研究中的应用。Chen等^[49]运用ChIP-Seq在小鼠胚胎干细胞中检测 13 个转录因子及 2 个转录调节子结合位点,发现有些位点是几个转录因子共同作用结合,如NANG、SOX2、OCT4、SMAD1 和STAT3 共同组成“胚胎干细胞增强子小体”。Marson等^[50]研究小鼠干细胞中NANG、SOX2、OCT4 和TCF3 的结合位点,识别了 14 230 个 4 种因子共同结合位点。ChIP-Seq起始研究材料可以是细胞系或组织。Wederell等^[34]研究成年小鼠肝脏组织中转录因子FOXA2 结合位点,识别 11

000 个位点。其中肝脏表达基因中 43.5%有相关的 FOXA 结合位点。

2.2 表观遗传学研究

“表观遗传”一词在不同文献中的定义有所差别^[51-53],但都包括组蛋白修饰、DNA 甲基化和核小体定位等方面,本文主要讨论 ChIP-chip 与 ChIP-Seq 在组蛋白修饰方面的应用。组蛋白修饰主要发生在 N 端氨基酸,包括甲基化、乙酰化、磷酸化、泛素化和 ADP 核糖基化等共价修饰。现有电荷中和(Charge neutralization)、组蛋白密码(Histone code)和信号通路(Signalling pathway)3 种主要的模型从不同角度解释组蛋白修饰在基因调控中的功能。

2002 年, Bernstein 等^[54]和 Robyl 等^[55]分别应用 ChIP-chip 研究酵母组蛋白修饰,发现其与特定的基因组区域和转录状态有关,这是 ChIP-chip 首次应用于全基因组蛋白修饰研究。随后以酵母为模型进行了大量的组蛋白修饰研究。Pokholok 等^[56]进行组蛋白甲基化和乙酰化研究发现:组蛋白乙酰化发生在基因转录起始部位, H3K36me3 遍布于编码区,这与转录激活子将乙酰转移酶招募到调控区域附近的理论模型^[18]及组蛋白乙酰化酶 Set2 与 RNA 聚合酶延伸相关的模型吻合。Liu 等^[57]发现 H3K9/14/18、H4K5/12、H2AK7 定位于活性基因的 5 端。酵母中组蛋白 H3K4 甲基化研究发现:该组蛋白的甲基化见于转录区域,沿 3 端到 5 端方向单甲基化、二甲基化到三甲基化数量渐增^[18],这与 H3K4 甲基转移酶 Set1 被招募到 RNA 聚合酶 活跃转录区域的 5 端的理论模型相符合^[58]。

高等真核生物组蛋白修饰研究结果与酵母基本一致,但有一些新发现。人类 T 细胞中 CpG 岛具有高水平组蛋白乙酰化和 H3K4me3^[59,60],这种组蛋白修饰与序列的相关性现象促使 Bernstein 等^[61]做了进化上的研究,发现人与小鼠直向同源位点组蛋白 H3 赖氨酸甲基化具有保守性,但相关 DNA 序列并不是十分保守。值得注意的是:这是组蛋白修饰在不同物种中位置上的保守,而不是相关序列保守。

Barski 等^[62]和 Mikkelsen 等^[63]首次应用 ChIP-Seq 分别在人类 T 细胞和胚胎干细胞中进行组蛋白修饰研究: Barski 等发现 H3K27/9/20/79 与 H2BK5 的单甲基化与基因活化相关,而 H3K27/9/79 的三甲基化与

基因抑制相关; Mikkelsen 等发现人类胚胎干细胞中几乎所有富含 CpG 岛的启动子区域都有 H3K4me3,这与 ChIP-chip 结果相互验证^[59,60]。H3K4me3 与基因活化相关,“成簇”存在, H3K27me3 与基因抑制相关,以“分散”状态存在,但二者有时也共存在某一区域,二者的绝对数量与相对数量决定基因的活性及胚胎干细胞的分化方向^[64]。

综合 ChIP-chip 与 ChIP-Seq 对组蛋白修饰的研究结果可以看出,特定修饰状态分布与增强子、启动子和绝缘子(Insulator)有相关性。如增强子与启动子区具有相似的组蛋白修饰,但活性启动子区与高水平的 H2BK5me1、H3K9me1、H3K27me1、H4K20me1、H3K36me3 有关,增强子区与高水平的 H3K4me1 相关。组蛋白修饰可以作为基因组注释的特征之一。

组蛋白修饰与 DNA 结合蛋白都影响基因表达,两方面数据比较可以得到更加可信的结论。例如, Cuddapah 等^[65]研究绝缘子结合蛋白 CTCF 在全基因的结合位点,发现这些位点也有 H3K27me3 修饰存在,而 H3K27me3 与异染色质的形成有关,从二者之间的一致性推测 CTCF 在染色质活性与抑制区域壁垒形成中发挥作用。Robertson 等^[48]研究 IFN 诱导前后 HelaS3 细胞中转录因子 STAT1 结合位点变化,发现诱导后位点比诱导前增加 28 996 个,这充分证明外界环境刺激下转录因子调控的变化。随后他们利用 ChIP-Seq 研究活性基因相关的组蛋白修饰 H3K4me1 与 H3K4me3,发现诱导后的结合位点中大多数在未诱导时已经具有 H3K4me1 与 H3K4me3 修饰^[66]。这说明激活的 STAT1 结合位点受预先存在的染色质状态影响。

3 展望

ChIP 是相对成熟的技术,但目前还存在一些技术难点。如需要大量的起始材料,对于难以大量培养的细胞如神经细胞和干细胞具有局限性,并且难以区分个别细胞与总体细胞的行为,解决办法是减少所用细胞数量。O'Neill 等^[67]创立了载体 ChIP (Carrier ChIP) 和 Dahl 等^[68]创立的 Q2(Quick and quantitative)ChIP,可以用少到 100 个细胞进行分析。ChIP-Seq 与 ChIP-chip 是目前全基因组分析染色质免疫沉淀产物的两大主要技术。相比而言, ChIP-Seq 具有以下 3 大主要优势:第一, ChIP-Seq 能实现真正的

全基因组分析。目前所能获得的芯片上固定的探针只能代表全基因组部分序列, 所获得的杂交信息具有偏向性; 第二, 对于结合位点分析, ChIP-Seq通过寻找“峰”, 结合分辨率可精确到 10~30 bp^[69], 而芯片上探针由于长度所限, 无法精确定位, 即使目前最高水平的商业芯片都无法提供可与ChIP-Seq媲美的分辨率^[70]; 第三是所需样本数量。ChIP-chip需要多达 4~5 μ g 的起始样本, 在杂交之前需要进行LM-PCR, 但可能导致背景增高, 竞争性扩增等导致假阳性^[71]。而ChIP-Seq仅需要纳克级起始材料, 如SOLiD起始材料可低至 20 ng。可以预计, ChIP-Seq技术将在细胞的基因表达调控网络研究中发挥更为重要的作用。

总之, 在全基因组范围内进行转录因子结合谱和表观遗传分析, 能够揭示活体细胞内基因表达调控的若干机制, 系统地整合这些 DNA 与蛋白质相互作用的数据, 将为构建越来越详细的基因表达调控网络模型提供依据。

参考文献(References):

- [1] Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhami P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SC, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hacker-muller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung WK, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammanna H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei CL, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaoz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Loytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglu S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CW, Ng P, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, Van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Xu M, Haidar JN, Yu Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, de Bakker PI, Kern AD, Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyraes E, Hallgrimsdottir IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VV, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, de Jong PJ. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 2007, 447(7146): 799–816. [\[DOI\]](#)
- [2] American Association for Cancer Research Human Epigenome Task Force; European Union, Network of Excellence, Scientific Advisory Board. Moving AHEAD with an international human epigenome project. *Nature*, 2008, 454(7205): 711–715. [\[DOI\]](#)
- [3] Massie CE, Mills IG. ChIPping away at gene regulation. *EMBO Rep*, 2008, 9(4): 337–343.

- [4] Chen H, Lin RJ, Xie W, Wilpitz D, Evans RM. Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell*, 1999, 98(5): 675–686.
- [5] Shang Y, Hu X, DiRenzo J, Lazar MA, Brown M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell*, 2000, 103(6): 843–852. [\[DOI\]](#)
- [6] Orlando V, Strutt H, Paro R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods*, 1997, 11(2): 205–214. [\[DOI\]](#)
- [7] Kurdistani SK, Robyr D, Tavazoie S, Grunstein M. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat Genet*, 2002, 31(3): 248–254. [\[DOI\]](#)
- [8] Tian B, Nowak DE, Jamaluddin M, Wang S, Brasier AR. Identification of direct genomic targets downstream of the nuclear factor-kappaB transcription factor mediating tumor necrosis factor signaling. *J Biol Chem*, 2005, 280(17): 17435–17448. [\[DOI\]](#)
- [9] Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods*, 2009, 48(3): 240–248. [\[DOI\]](#)
- [10] Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*, 2008, 40(7): 897–903. [\[DOI\]](#)
- [11] Weitsman GE, Skliris G, Ung K, Peng B, Younes M, Watson PH, Murphy LC. Assessment of multiple different estrogen receptor-beta antibodies for their ability to immunoprecipitate under chromatin immunoprecipitation conditions. *Breast Cancer Res Treat*, 2006, 100(1): 23–31.
- [12] Chen Q, Lin L, Smith S, Huang J, Berger SL, Zhou J. CTCF-dependent chromatin boundary element between the latency-associated transcript and ICP0 promoters in the herpes simplex virus type 1 genome. *J Virol*, 2007, 81(10): 5192–5201. [\[DOI\]](#)
- [13] Belyaev ND, Wood IC, Bruce AW, Street M, Trinh JB, Buckley NJ. Distinct RE-1 silencing transcription factor-containing complexes interact with different target genes. *J Biol Chem*, 2004, 279(1): 556–561. [\[DOI\]](#)
- [14] Kolodziej KE, Pourfarzad F, de Boer E, Krpic S, Grosveld F, Strouboulis J. Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Mol Biol*, 2009, 10: 6. [\[DOI\]](#)
- [15] Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 2002, 298(5594): 799–804. [\[DOI\]](#)
- [16] Viens A, Mechold U, Lehrmann H, Harel-Bellan A, Ogryzko V. Use of protein biotinylation *in vivo* for chromatin immunoprecipitation. *Anal Biochem*, 2004, 325(1): 68–76. [\[DOI\]](#)
- [17] Liu CL, Schreiber SL, Bernstein BE. Development and validation of a T7 based linear amplification for genomic DNA. *BMC Genomics*, 2003, 4(1): 19. [\[DOI\]](#)
- [18] Roh TY, Ngau WC, Cui K, Landsman D, Zhao K. High-resolution genome-wide mapping of histone modifications. *Nat Biotechnol*, 2004, 22(8): 1013–1016. [\[DOI\]](#)
- [19] Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH. Defining the CREB Regulon: A genome-wide analysis of transcription factor regulatory regions. *Cell*, 2004, 119(7): 1041–1054.
- [20] Kim J, Bhinge AA, Morgan XC, Iyer VR. Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat Methods*, 2005, 2(1): 47–53. [\[DOI\]](#)
- [21] Chen J, Sadowski I. Identification of the mismatch repair genes PMS2 and MLH1 as p53 target genes by using serial analysis of binding elements. *Proc Natl Acad Sci USA*, 2005, 102(13): 4813–4818. [\[DOI\]](#)
- [22] Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, Liu J, Zhao XD, Chew JL, Lee YL, Kuznetsov VA, Sung WK, Miller LD, Lim B, Liu ET, Yu Q, Ng HH, Ruan Y. A global map of p53 transcription-factor binding sites in the human genome. *Cell*, 2006, 124(1): 207–219. [\[DOI\]](#)
- [23] Nix DA, Courdy SJ, Boucher KM. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics*, 2008, 9:523. [\[DOI\]](#)
- [24] Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, 2007, 316(5830): 1497–1502. [\[DOI\]](#)
- [25] Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol*, 2008, 26(11): 1293–1300. [\[DOI\]](#)
- [26] Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol*, 2009, 27(1): 66–75. [\[DOI\]](#)
- [27] Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ. FindPeaks 3.1: A tool for identifying areas of

- enrichment from massively parallel short-read sequencing technology. *Bioinformatics*, 2008, 24(15): 1729–1730. [\[DOI\]](#)
- [28] Boyle AP, Guinney J, Crawford GE, Furey TS. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*, 2008, 24(21): 2537–2538. [\[DOI\]](#)
- [29] Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res*, 2008, 36(16): 5221–5231. [\[DOI\]](#)
- [30] Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, Batzoglou S, Myers RM, Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods*, 2008, 5(9): 829–834. [\[DOI\]](#)
- [31] Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 2008, 9(9): R137. [\[DOI\]](#)
- [32] Xu H, Wei CL, Lin F, Sung WK. An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. *Bioinformatics*, 2008, 24(20): 2344–2349. [\[DOI\]](#)
- [33] Zhang ZD, Rozowsky J, Snyder M, Chang J, Gerstein M. Modeling ChIP sequencing in silico with applications. *PLoS Comput Biol*, 2008, 4(8): e1000158. [\[DOI\]](#)
- [34] Wederell ED, Bilenky M, Cullum R, Thiessen N, Dagpinar M, Delaney A, Varhol R, Zhao Y, Zeng T, Bernier B, Ingham M, Hirst M, Robertson G, Marra MA, Jones S, Hoodless PA. Global analysis of *in vivo* Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res*, 2008, 36(14): 4549–4564. [\[DOI\]](#)
- [35] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*, 2002, 295(5558): 1306–1311. [\[DOI\]](#)
- [36] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 2006, 38(11): 1348–1354. [\[DOI\]](#)
- [37] Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 2006, 38(11): 1341–1347. [\[DOI\]](#)
- [38] Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. Chromosome conformation capture carbon copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 2006, 16(10): 1299–1309. [\[DOI\]](#)
- [39] Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. Genome-wide location and function of DNA binding proteins. *Science*, 2000, 290(5500): 2306–2309. [\[DOI\]](#)
- [40] Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, 2001, 409(6819): 533–538. [\[DOI\]](#)
- [41] Horak CE, Luscombe NM, Qian J, Bertone P, Piccirillo S, Gerstein M, Snyder M. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev*, 2002, 16(23): 3017–3033. [\[DOI\]](#)
- [42] Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM, Snyder M. GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIP-chip analysis. *Proc Natl Acad Sci USA*, 2002, 99(5): 2924–2929. [\[DOI\]](#)
- [43] Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev*, 2002, 16(2): 245–256. [\[DOI\]](#)
- [44] Li Z, van Calcar S, Qu C, Cavenue WK, Zhang MQ, Ren B. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc Natl Acad Sci USA*, 2003, 100(14): 8164–8169. [\[DOI\]](#)
- [45] Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev*, 2002, 16(2): 235–244. [\[DOI\]](#)
- [46] Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M. Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc Natl Acad Sci USA*, 2003, 100(21): 12247–12252. [\[DOI\]](#)
- [47] Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. A high-resolution map of active promoters in the human genome. *Nature*, 2005, 436(7052): 876–880. [\[DOI\]](#)
- [48] Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods*, 2007, 4(8): 651–657. [\[DOI\]](#)
- [49] Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX,

- Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 2008, 133(6): 1106–1117. [\[DOI\]](#)
- [50] Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, Sharp PA, Bartel DP, Jaenisch R, Young RA. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*, 2008, 134(3): 521–533. [\[DOI\]](#)
- [51] Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell*, 2007, 128(4): 635–638.
- [52] Ptashne M. On the use of the word 'epigenetic'. *Curr Biol*, 2007, 17(7): R233–236. [\[DOI\]](#)
- [53] Bird A. Perceptions of epigenetics. *Nature*, 2007, 447(7143): 396–398. [\[DOI\]](#)
- [54] Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci USA*, 2002, 99(13): 8695–8700.
- [55] Robyr D, Suka Y, Xenarios I, Kurdastani SK, Wang A, Suka N, Grunstein M. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell*, 2002, 109(4): 437–446. [\[DOI\]](#)
- [56] Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolzheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 2005, 122(4): 517–527. [\[DOI\]](#)
- [57] Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, Friedman N, Rando OJ. Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol*, 2005, 3(10): e328. [\[DOI\]](#)
- [58] Ng HH, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell*, 2003, 11(3): 709–719. [\[DOI\]](#)
- [59] Roh TY, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev*, 2005, 19(5): 542–552. [\[DOI\]](#)
- [60] Roh TY, Cuddapah S, Cui K, Zhao K. The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci USA*, 2006, 103(43): 15782–15787. [\[DOI\]](#)
- [61] Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, 2005, 120(2): 169–181. [\[DOI\]](#)
- [62] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, 129(4): 823–837. [\[DOI\]](#)
- [63] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 2007, 448(7153): 553–560. [\[DOI\]](#)
- [64] Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell*, 2009, 4(1): 80–93. [\[DOI\]](#)
- [65] Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*, 2009, 19(1): 24–32. [\[DOI\]](#)
- [66] Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, Thiessen N, Cezard T, Fejes AP, Wederell ED, Cullum R, Euskirchen G, Krzywinski M, Birol I, Snyder M, Hoodless PA, Hirst M, Marra MA, Jones SJ. Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res*, 2008, 18(12): 1906–1917. [\[DOI\]](#)
- [67] O'Neill LP, VerMilyea MD, Turner BM. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat Genet*, 2006, 38(7): 835–841. [\[DOI\]](#)
- [68] Dahl JA, Collas P. Q2ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. *Stem Cells*, 2007, 25(4): 1037–1046. [\[DOI\]](#)
- [69] Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*, 2008, 26(12): 1351–1359. [\[DOI\]](#)
- [70] Qi Y, Rolfe A, MacIsaac KD, Gerber GK, Pokholok D, Zeitlinger J, Danford T, Dowell RD, Fraenkel E, Jaakkola TS, Young RA, Gifford DK. High-resolution computational models of genome binding events. *Nat Biotechnol*, 2006, 24(8): 963–970. [\[DOI\]](#)
- [71] O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques*, 2006, 41(5): 577–580. [\[DOI\]](#)