

核小体及染色质修饰的基因组分布模式和染色质状态*

刘宏德^{1)**} 罗坤²⁾ 马昕³⁾ 翟金城¹⁾ 谢建明¹⁾ 孙啸¹⁾ 万亚坤⁴⁾

(¹⁾ 东南大学生物电子学国家重点实验室, 南京 210096; (²⁾ 新疆医科大学第一附属医院神经外科, 循证医学研究所, 乌鲁木齐 830054;

(³⁾ 南京审计学院金审学院, 南京 211815; (⁴⁾ 东南大学生命科学研究院, 南京 210096)

摘要 染色质是真核 DNA 的存在方式, 可以通过影响 DNA 的可及性调节基因转录, 其基本单元为核小体, 系由约 147 bp 的 DNA 缠绕在组蛋白八联体上形成的结构, 核小体之间以连接 DNA 相连. 核小体组蛋白上能发生甲基化和乙酰化等化学修饰. 核小体位置、DNA 的甲基化和组蛋白的修饰等对染色质状态(常染色质或异染色质)及基因组之间的长程相互作用有重要影响. 近年, 基于高通量测序技术, 核小体位置和染色质修饰在多种细胞中的基因组分布已被测定. 结果显示, 这些标记的分布模式具有位点特异、动态变化、相互偶联和高度复杂的特征. 本文详细回顾并评述了核小体位置和染色质修饰的分布模式、对应生物学功能、修饰之间的关联、实验测定技术、染色质状态的计算分析等内容. 该工作对于深入认识和理解染色质的表观遗传调节机制有重要意义.

关键词 染色质, 核小体, 组蛋白修饰, 分布模式, 高通量 DNA 测序

学科分类号 Q7

DOI: 10.3724/SP.J.1206.2012.00434

真核生物 DNA 以染色质形式存在, 其基本单元为核小体. 核小体系由约 147 bp 的 DNA 缠绕在组蛋白核上形成的结构(图 1), 核小体之间以连接 DNA 相连^[1-2]. 组蛋白会发生甲基化或者乙酰化等化学修饰. 构成 DNA 的胞嘧啶(C)也可在 5'端发生甲基取代修饰. 在间期核中, 染色质可有常染色质(euchromatin)和异染色质(heterochromatin)两种状态, 前者伸展且呈透明状态, 后者卷曲凝缩. 有丝分裂期, 染色质高度螺旋、折叠形成染色体. 染色质是一种动态结构, 核小体的位置、DNA 甲基化和组蛋白的修饰、染色质结合蛋白等能深刻影响染色质的状态, 这些因素的改变能导致染色质状态的改变. 染色质状态决定了 DNA 可接近性及 DNA 在核内的空间位置^[1], 调节蛋白质与 DNA 的结合, 进而调控转录、DNA 复制、DNA 损伤修复等基本生物过程, 关于此内容的研究是表观基因组学分析的核心内容之一^[3].

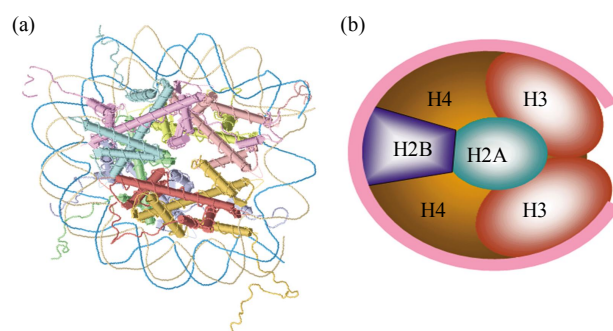


Fig. 1 Three-dimensional structure of nucleosome(a) and configuration of histones in nucleosome(b)

图 1 核小体的结构

(a)核小体 3D 结构(PDB ID: 1EQZ), N 端的组蛋白尾部从两圈 DNA 中伸展至核小体外部. (b)核小体中组蛋白的分布模型, (H3-H4)₂ 四聚体提供了核心, 在俯视图中可见一对 H2A-H2B, 另一对位于下方.

* 新疆维吾尔自治区自然科学基金(2012211A076), 国家重点基础研究发展计划(973)(2012CB316501), 国家自然科学基金(31240080, 31371339)和江苏省自然科学基金(12KJD520006)资助项目.

** 通讯联系人.

Tel: 025-83795174, E-mail: liuhongde@seu.edu.cn

收稿日期: 2012-12-25, 接受日期: 2013-03-19

近年, 借助于二代测序技术(NGS), 关于染色质的分析进展迅速, 多种生物的特定细胞核小体位置、组蛋白修饰(主要为组蛋白甲基化和乙酰化修饰)位置已经被测定^[2-5]. 这些染色质修饰在基因组特殊位点周围的分布具有特定的模式, 带有复杂的生物学意义. 本文将从核小体位置、DNA 甲基化和组蛋白修饰的基因组分布模式及生物学意义、染色质修饰的实验测定技术、染色质状态的相关计算方法等方面阐明该领域的研究进展.

1 核小体、DNA 甲基化和组蛋白修饰 (乙酰化和甲基化) 的基因组分布模式

在常染色质和异染色质中, 核小体位置^[1, 4-6]、DNA 甲基化^[4]、组蛋白修饰(甲基化、乙酰化、磷

酸化、泛素化和组蛋白变体替换)^[2, 7-10]等在基因组上的分布模式不同, 下面分别回顾该方面近年的进展.

1.1 核小体的分布模式

核小体是染色质的基本构建单元, 核小体在基因组特殊位点有特异的分布模式(图 2). 通常, 在活性转录起始位点(TSS)周围, 有一个约 200 bp 宽的核小体缺失区域(NFR)^[5, 11], 此区域正是一些转录因子的结合位置, TSS 周围 NFR 对于转录起始必不可少^[6]. 在紧邻 NFR 的两侧, 有两个定位的核小体^[2, 5, 11], 这种组织模式或可能给转录因子的结合提供空间位置的匹配. NFR 下游大约 1 000 bp 的范围内, 核小体的分布具有等间隔性^[6, 12-13].

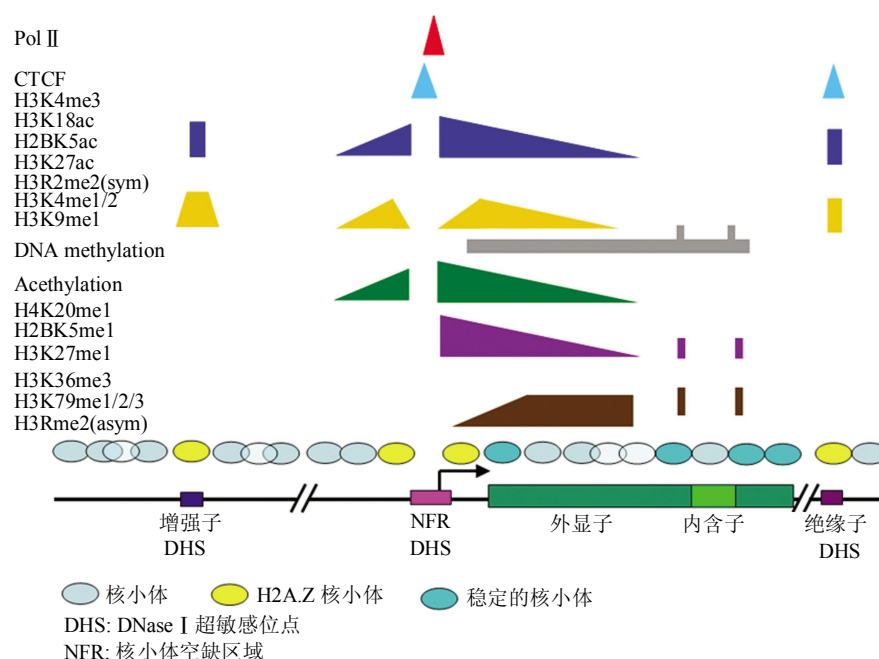


Fig. 2 The patterns of chromatin modification distributions in expressed genes of human genome

图 2 染色质修饰在人类活性基因中的分布模式

NFR 两侧的核小体极易发生 H2A.Z 组蛋白替换修饰(尤其在转录时), H2A.Z 的替换会导致核小体结构的不稳定. 在 CD4⁺ T 细胞中, NFR 上游的第一个核小体(-1)容易解体, 而下游第一个核小体(+1)在基因活化和沉默时动态定位^[9]. 在果蝇的基因组中, 包含两个 H2A.Z 组蛋白的同型核小体富集在活化启动子 TSS 下游, 且在转录延长时解体; 而异型的核小体(只包含一个 H2A.Z 的核小体)在此区域分布很少^[14], 说明这种同型的核小体更加不稳

定. 新近的研究表明, H2A.Z 的变体 H2A.Z.2.2 能使得核小体更加不稳定^[15].

核小体壁垒模型(nucleosome barrier model)认为, TSS 下游的第一个核小体(+1)稳定定位, 是一个障碍, 这使得下游的核小体等间隔分布. 这种现象在酵母基因组中尤其明显. TSS 周围 NFR 的跨度似乎与基因表达水平有某种关联, 在酵母中发现, NFR 的宽度会有变化, 高表达的基因具有更宽的 NFR^[16], 但明确量化的关系仍需要进一步分析.

核小体定位是指 DNA 双螺旋相对于组蛋白八联体的相对位置, 多种因素影响核小体定位^[17], 包括: DNA 序列本身对核小体的偏好性^[18]; DNA 结合蛋白(如转录因子)的作用^[19]; 染色质重塑复合物的作用, 以及邻近核小体对定位的作用^[20]. 然而, 在核小体动态定位过程中, 这些因素的相互作用机制仍然需要进一步查明. 核小体定位与转录因子的结合密切相关, 但核小体的水平与基因表达水平之间的相关性并不显著^[21], 说明核小体在 TSS 空缺是转录的必要条件, 但基因表达水平却是多方因素调节的结果.

1.2 组蛋白修饰的基因组分布模式

基因组不同区域在活化和失活两种状态下, 其染色质修饰的组合模式不同^[2, 22-23]. 各种组蛋白修饰对应的染色质活性见表 1. 我们根据目前的文献, 绘制了部分染色质修饰的分布模式(图 2).

Table 1 The relationship between histone modifications and chromatin states

表 1 组蛋白修饰(包括两个组蛋白变体)与染色质状态的对应关系

组蛋白修饰位置	修饰类型			
	单甲基化(me1)	二(双)甲基化(me2)	三甲基化(me3)	乙酰化(ac)
H3K4	A/R	A	A	A
H3K9	R	R	R	A
H4K20	A			A
H3K27	\	R	R	A
H3K36	\	A/R	A	A
H3K56	A	\	\	A
H3K79	A	A	R	A
H3R2	\	R	\	A
H2A.Z	A	A	A	A
H2BK5	A	\	\	A
H4K20	R	R	R	A
H4R3	\	R	\	A
H3.3	A	A	A	A
H2A UB1	A	A	A	A
H2B UB1	R	R	R	R

A: 活化或染色质的疏松; R: 抑制或染色质的固缩; \: 功能待分析; me: 甲基化; ac: 乙酰化.

通常, 活化启动区具有 H3K4me3 和组蛋白乙酰化(ac)修饰(图 2), 其分布模式为: 转录起始

位置两侧的 2 个核小体具有最高的修饰水平; 远离转录起始位点, 修饰水平逐渐减弱. 这种分布模式在人和小鼠的同类型细胞中高度保守, 可能在哺乳动物进化中保守^[24]. 沉默的启动子常带有 H3K27me3、H3K9me3 和 DNA 甲基化修饰.

增强子处富集组蛋白乙酰化和 H3K4me3, 这成为识别增强子的重要标识^[25-26], 而且这些修饰具有细胞类型特异性^[25, 27]. 增强子区域和转录起始位区域, 一般为 DNase 敏感位点(DHS), 最近研究证实: H3K4 的甲基化水平和 H3K9 和 H3K27 的乙酰化水平在 DHS 周围急剧升高. H3K9me1 和 H4K20me1 的水平在 DHS 轻度升高, 但在泛素修饰的 DHS 周围却减弱^[28].

在外显子两端定位有稳定的核小体^[29-30], 同时带有 H3K36me3、H3K79me1、H2BK5me1 和 H3K27me1/2/3 修饰. 此处的组蛋白修饰与剪切和可变剪切密切相关, 如: 纤维母细胞生长因子 2 (FGFR2)在人类间叶干细胞和前列腺上皮细胞中可变剪切, 研究发现 H3K36me3 可能是剪切接头系统的阅读标记^[29].

基因的转录区带有 H3K36me3 修饰^[31], 而启动子区带有 H3K4me3 修饰, 这种特征(被称为“K3K4 规则”)对于非编码基因的转录起始位点的识别特别有用.

常染色质常与对称的 H3R2me2(“对称”是指精氨酸两侧链各发生一个甲基取代)与 H3K4me3 修饰关联^[32]. 组成型异染色质带有 H3K9 和 H4K20 的甲基化修饰, H3K9me3 是异染色质结合蛋白 HP1 的结合标记, 而 H4K20me1 被发现在诱导染色质固缩中有作用^[2, 22-23, 33]. 新近发现, 不对称的 H3R2me2(两次甲基取代发生在侧链同一个 N 原子)也与异染色质和基因的沉默关联^[28]. 着丝粒区域富集 H4R3me2、H3R2me2 和 H4K20me3^[34].

H3K4me3 和 H3K27me3 分别是活化和抑制性质的修饰, 这两种修饰同时发生在发育有关的转录因子基因和细胞状态调节基因的启动子区^[2, 10, 35-36], 即“双价修饰”, 其中, H3K27me3 被 PRC2 识别并结合. 这些发育相关的基因在多能干细胞中沉默, 但是能够迅速被激活. 这两种修饰在形成异染色质的过程中存在一个非编码 RNA 介导的偶联过程: HOTAIR 是一种长非编码 RNA, 在确立异染色质过程中充当支架^[9], HOTAIR 的 5' 端与 PRC2 结合, 3' 端与 LSD1/CoREST/REST 复合物结合,

这样 HOTAIR 就使得 H3K27 的甲基化和 H3K4 的去甲基化偶合. 在诱导多能干细胞(iPSC)和多能胚胎干细胞(ESCs)中, H3K4me3 和 H3K27me3 的分布模式几乎没有差异^[37].

可转录区域伴有 H3K36me2 和 H3K79me3 标记. H3K4me3/2 能够招募 TFIID 亚基 3(TAF3)、V(D)J 重组酶 RAG2 和其他的染色质结构重塑因子, 即这两种修饰与基因转录、B 细胞和 T 细胞发育中的 V(D)J 重组等有关, 似乎扮演染色质结构索引的角色^[25].

基因表达水平与组蛋白的修饰水平密切关联, 在拟南芥中, H3K9ac 和 H3K4me3 关联, 在组成性基因的 TSS 周围, 这两种修饰的信号高并且窄; 而在特异性表达基因的 TSS 周围, 这些修饰宽且向编码区扩展^[38]. 某些组蛋白的修饰水平可以预测基因表达水平, 对于低 GC 含量的启动子, H3K4me3 和 H3K79me1 的水平能很好地预测基因表达, 而对于高 GC 含量的启动子, H3K27ac 和 H4K20me1 与基因表达水平关系密切^[39].

所有表观修饰之间具有复杂的相互作用^[2-37,10,21,40-45],

活化相关的组蛋白修饰与 DNA 甲基化互斥, 如 H3K4me3 与 DNA 的甲基化在基因的启动子区域呈现分布上的互斥(H3K4 的甲基化修饰能阻碍 DNA 转甲基酶 DNMT3L 的结合), 在拟南芥中, H2A. Z 的替换也与 DNA 甲基化互斥. 在人类细胞增强子周围, H3K4me1 富集, 但是 H3K4me3 比较缺乏^[40]. TrxG 和 PcG 是在基因组上具有相同结合位点但作用效果却完全相反的两类染色质结合蛋白. PcG 蛋白包括组蛋白修饰剂(如甲基化 H3K4 的 MLL1)、染色质重塑蛋白(如 CHD1-8、BRM 和 BRG1 等)^[10]. 核小体与邻近核小体 DNA 的静电作用会被组蛋白乙酰化削弱^[31].

组蛋白修饰之间也存在复杂的偶联, 已发现的偶联见表 2. 这些关系主要发生在两类性质不同的修饰之间, 如具有活化作用的 H3K4 甲基化和具有抑制作用的 H3K27 甲基化. 这种偶联作用实现了染色质状态从活化状态向抑制状态(或者抑制状态向活化状态)的转化. 也有的偶联实现了某种状态的加强, 遗憾的是, 虽然对这些相互作用有所了解, 但尚不完全清楚.

Table 2 The coupling between histone modifications and its effect for the chromatin state changes			
Table 2 组蛋白修饰之间的偶联关系及对应的染色质状态的变化			
组蛋白修饰之间的偶联关系	染色质状态	涉及的酶或者机制等	物种
H3S10 Pho→	R→A	Snf1 kinase	
H3k14 ac		Gcn5	
(H3k18ac & H3k23ac)→	A+	CARM1	
H3R17me			
H3R36me→	R+	Rpd3S	
H3 和 H4 去乙酰化			
H2B Ub→	R	Set1	<i>S. cerevisiae</i>
(H3K4me & H3K79me)		Dot1	
H3K27me2/3 去甲基化→	R→A	MLL3/4 Set1-H3K4	Mammalian
H3K4me			
H3K36 去甲基化 & H2AK119 Ub)→	R	PRC2、PRC1	<i>Drosophila</i>
H3K27me			
H3K4 去甲基化→	A→R	RNAi	<i>S. pombe</i>
H3K9me			
H3K27me→	R→A		Mammalian
H3 裂解			Stem cell

A: 活化或染色质的疏松; R: 抑制或染色质的固缩; +: 强化当前状态.

1.3 DNA 甲基化的分布模式

DNA 甲基化是指 DNA 胞嘧啶 C5 位置上的 H

原子被甲基所取代的修饰过程, 通常发生在 CpG 二核苷上. 启动子区的 DNA 甲基化是基因活性的

抑制因素. 在哺乳动物中, 约有 60%~90% CpG 二核苷被甲基化^[47-48]. 非甲基化的 CpG 二核苷多成簇出现在基因的 5' 调控区域(CpG 岛), 抑癌基因的 CpG 岛的甲基化常在癌细胞中被发现^[48]. DNA 甲基化常与异染色质伴随出现^[49]. 已经鉴定出了人类基因组的 DNA 甲基化在部分细胞中的分布^[41, 50]. 在活性转录因子结合位点(TFBS)周围, 甲基化水平低^[29]. 核小体定位和 DNA 甲基化分布之间存在关联: 在人类基因组编码区的边界、核小体和 DNA 甲基化水平都高^[42]. 在拟南芥中, 核小体 DNA 的甲基化分布具有 10 bp 的周期性 (DNA 序列 10 bp 周期被认为是核小体的定位密码^[13, 18]), 而且, 甲基化更倾向于发生在核小体 DNA 上, 连接 DNA 具有较低的甲基化水平^[43-44]. 一般地, DNA 甲基化与 H3K27me3 化共存, 但在 CpG 岛上, 这两种标记互相排斥^[51].

Lister 等^[52]的研究表明: DNA 甲基化在细胞分化过程中动态变化, 相比于胚胎干细胞, 胎儿纤维母细胞的甲基化有一个较广范围的减少, 且与低的转录活性关联. 而且, 在胚胎干细胞中, 近 1/4 的甲基化发生在非 CG 核苷(non-CG)上, 这类修饰在蛋白质结合位点缺乏, 而在基因体(gene body)区富集. 在分化的干细胞, 非 CG 核苷的甲基化中消失, 而在诱导多能干细胞中重现, 这说明胚胎干细胞可能有另外一种甲基化机制.

1.4 染色质标记与染色体构象和基因组长程相互作用

在真核细胞中, 基因组存在长程的相互接触(长程相互作用), 形成染色(质)体的 3D 构象. 基因的转录等过程受到染色体构象和长程作用的影响. 组蛋白修饰等染色质标记与染色体构象、基因组的长程相互作用之间密切相关.

在 2009 年, Lieberman 等^[53]发现, 在人类基因组中, 富集基因的染色体区域在空间上邻近; 基因组在空间上被分割为开放的染色质和致密的染色质两大类, 在百万碱基尺度上, 染色质构象是一个没有结点的聚合物“分形球”, 这种组织方式一方面实现了对大量的 DNA 的密集压缩包装, 同时保证可以折叠或者打开任何基因组的局部区域.

Guillaume 等^[22]利用 53 种染色质结合蛋白信息, 将果蝇的染色质分为 5 种类型, 各类状态染色

质富集的基因功能不同, 基因表达水平各异. 同样是在果蝇中, Sexton 等^[54]利用组蛋白修饰和染色质结合蛋白信息, 将基因组分为 4 种类型的结构域, 这 4 种结构域分别为: 富集活化表观遗传标记的结构域、富集 HP1 标记的结构域、富集 Polycomb 标记的结构域和没有明显标记的结构域. 果蝇的这些结构域线性地分布在整个基因组上, 基因失活的区域在空间上聚集在一起, 而具有基因活性的区域在失活区域之外, 这些区域中包含有跨染色体或者染色体内的远距离接触, 这些结构域的边界存在绝缘子结合位点.

基因组的长程相互作用是一个复杂网络, 在 GM12878、K562 和 HeLa-S3 细胞中^[55], 有超过 1000 个来自于启动子和远距离位点的相互作用, 这些远距离位点包含增强子、启动子和绝缘子结合位点. 基因表达、启动子-增强子相互作用、增强子 RNA 三者之间存在显著相关性. 长程相互作用在基因组分布上有明显的不对称性, 这种作用主要位于转录起始约 120 kb 的上游. 在前列腺癌细胞中^[56], 促癌转录因子的过表达与染色质的全局性和功能性的连续改变相关.

2 核小体位置、DNA 甲基化、组蛋白修饰以及基因组长程相互作用的实验测定

目前, 实验方法测定核小体位置、DNA 甲基化和组蛋白修饰区域、基因组长程作用等的主要手段是基于第二代测序技术(图 3)^[57].

染色质免疫共沉淀(ChIP)是最直接的识别蛋白质在 DNA 上的结合位点或者某种组蛋白修饰位点的方法^[25, 33, 57]. ChIP 结合测序(ChIP-Seq)方法测定组蛋白修饰主要通过以下步骤: 交联组蛋白和 DNA、用 MNase 消化连接 DNA、免疫沉淀、纯化 DNA、(放大)测序等. DNA 元件百科全书计划(ENCODE)推荐 ChIP-Seq 的流程中, 对于组蛋白修饰的测定, 抗体至少需要免疫印迹验证. 如果测定目标属于较宽的区域, 对于人类细胞, 唯一比对到参考基因组的 Reads 数目不少于 4 千万个, 至少 80% 的 Reads 比对到基因组不同位置上. 至少要有两次重复实验, 对人类细胞, 每次实验唯一比对的 Reads 需要超过 2 千万个.

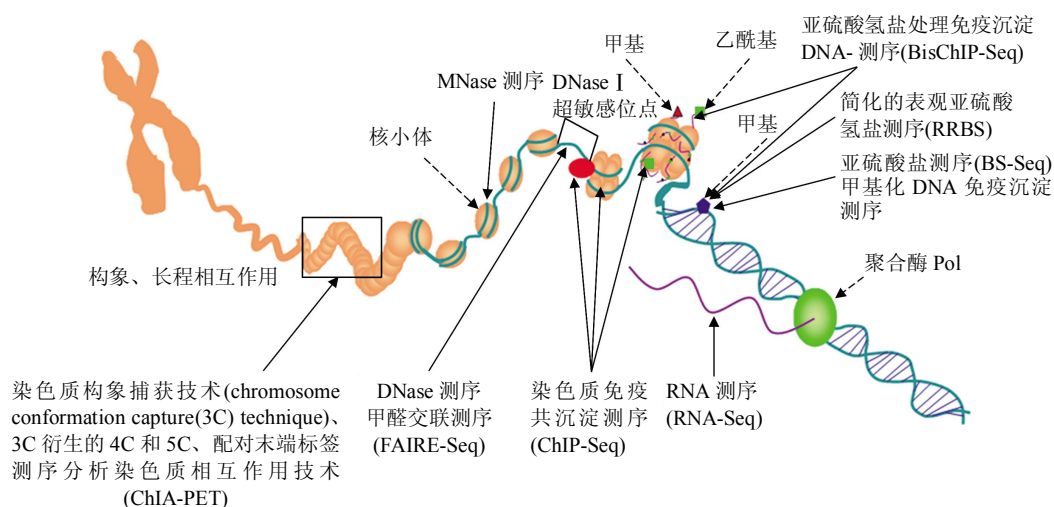


Fig. 3 The techniques of determinating nucleosome position, DNA methylation, histone modifications and chromosome conformation

图 3 核小体位置、DNA 甲基化、组蛋白修饰、染色体构象等的实验测定技术

核小体的测定可以通过微球菌核酸酶(micrococcal nuclease, MNase)酶切测序(MNase-Seq)、ChIP-Seq 和酶切结合芯片的方法测定. MNase-Seq 通过 MNase 的消化确定基因组上哪些区域存在核小体^[33, 57]. DNase-Seq 通过 DNase I 核酸酶的消化识别缺乏核小体的染色质开放区域. 在核小体位置测定中, 新近有一技术可以直接从核小体的二分位点(dyad site)裂解核小体 DNA, 从而直接测定核小体中心部分的 DNA 序列^[58]. 具体的技术过程为: 首先在组蛋白 H4 的第 47 个氨基酸残基(S)位置引入一个半胱氨酸(C)(H4S47C), 该半胱氨酸能在一个铜螯合物(N-(1, 10-phenanthroline-5-yl)iodoacetamide)和过氧化氢存在的条件下形成一个短期的羟基自由基, 该自由基能裂解临近的 DNA 骨架, 因此核小体 DNA 在二分对称位点被切断. 随后, 这些核小体 DNA 片段被测序. 这项技术已经成功地测定了酵母的核小体, 揭示了更细节更清晰的核小体分布.

DNA 甲基化测定主要有全基因组亚硫酸盐测序(BS-Seq)、甲基化 DNA 免疫沉淀测序(MeDIP-Seq), 以及基于芯片的方法. BS-Seq 测定 DNA 甲基化的实验原理是^[59]: 通过亚硫酸氢盐(bisulfite)处理, 将基因组中未发生甲基化的 C 碱基转换成碱基 U, 随后 PCR 扩增转化为 T, 使其与发生甲基化的碱基 C 区分开来, 再结合高通量测序技术, 测定具体的甲基化的基因组位置. 该技术适用于绘制单碱基分辨率的全基因组 DNA 甲基

化图谱, 费用相对较高. MeDIP-seq^[60]适合于测定基因组上的高甲基化的区域, 该技术通过 5' 甲基胞嘧啶抗体特异性地富集基因组上甲基化的 DNA 片段, 然后测序获取甲基化片段的位置信息, 但不能达到单碱基分辨率. 简化的表观亚硫酸氢盐测序技术(RRBS)主要以通过 *Msp* I 酶切富集 CpG 位点, 用较小的数据量得到包含最多 CpG 位点甲基化信息的单碱基精度的甲基化图谱^[61].

新近有一项新的称作“亚硫酸氢盐处理免疫沉淀 DNA- 测序(bisulfite sequencing of chromatin immunoprecipitated DNA, BisChIP-seq)”的技术可以直接测定基因组中 H3K27me3 和 DNA 甲基化之间的分布关系^[62], 这完全不同于以往推断两种表观修饰之间的共依存关系, 仅依靠计算两者分布的相关系数.

通过染色体构象捕获技术(chromosome conformation capture, 3C), 可分析染色质在核内的空间构象, 该技术主要经过染色质甲醛交联染色体邻近区域、酶切交联 DNA 片段对、片段对分子内交联、反向连接、扩增测序等步骤(图 4)^[57, 63-65]. Hi-C 是新近发展的一种染色体构象分析技术, 可实现全基因组染色体构象分析, 基本流程为甲醛交联 DNA、利用能使 5' 端悬长的酶降解 DNA(如 *Hind* III)、以生物素标记核苷填补末端、连接末端、剪切连接后的 DNA、富集分离带有生物素标记的 DNA 片段、制备文库、测序. Hi-C 已经成功应用于染色体构象的分析^[66]. ChIA-PET 技术^[57-67]可以分

析某种转录因子介导的染色体之间的长程相互作用, 其基本过程为: 染色体甲醛交联、超声片段化、以特定转录因子的抗体进行染色质免疫共沉淀、以 DNA 连接子连接交联在一起的 DNA 片段、

提取 DNA 片段、测序. 翟侃等^[68]综述了染色质构象捕获、环状染色质构象捕获、3C 碳拷贝、ChIA-PET 和 Hi-C 等技术的基本原理及发展历程, 对影响实验结果准确性的主要因素进行了分析.

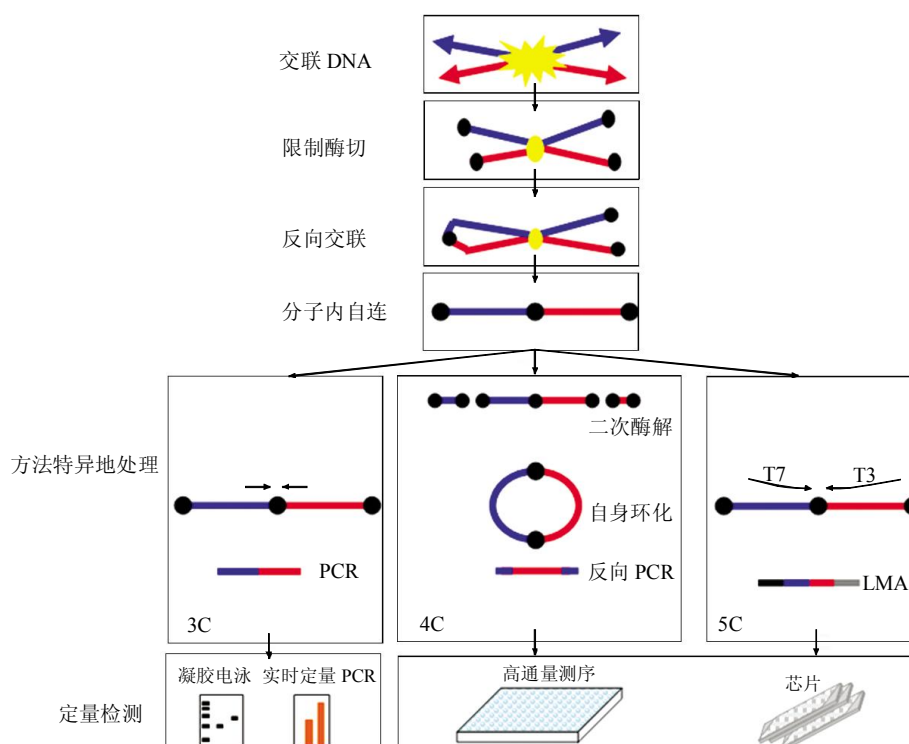


Fig. 4 The technique flowchart of chromosome conformation capture

图 4 染色体构象捕获技术(chromosome conformation capture)的流程示意图

另外, 对于染色质结合蛋白(如转录因子、PcG 蛋白等), 还可以使用 DNA 腺嘌呤甲基酶识别 (DamID) 技术^[48](一种基于测量甲基转移酶融合蛋白表达量的方法).

基于抗体测序数据的一般分析流程为(图 5): 首先过滤低质量的 Reads 和接头 Reads. 然后利用比对工具比对到参考基因组, 可用的比对工具包括 Bowtie^[69]、SOAP^[70]、RazerS^[71]、BFAST^[72]等. 过滤 PCR 中可能人工引入的 Reads, 过滤比对位置不唯

一的 Reads. 利用“peak calling”工具计算并识别信号富集区域. 峰识别的过程相对较复杂, 目前常用的工具有 MACS^[73]、FindPeaks^[74]、BayesPeak^[75]、GLITR^[76]、HPeak^[77]、F-Seq^[78]等. 针对染色体构象及基因组长程相互作用测序数据的处理流程见文献[79-81]. BS-Seq 的 Reads 比对可利用 Merman aligner(<http://rafalab.jhsph.edu/bsmooth>)^[82], 后续的差异甲基化区域识别分析可利用基于 R 语言的工具包 bsseq^[82].

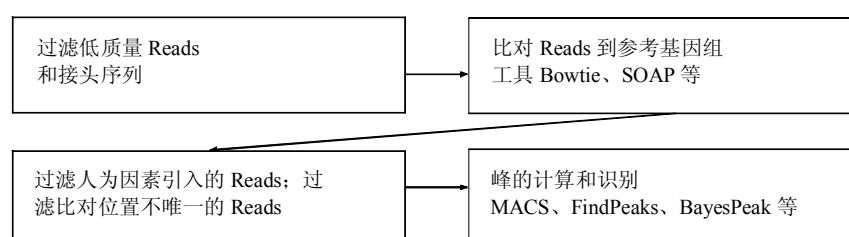


Fig. 5 The flowchart of analyzing the reads dataset from antibody-based sequencing

图 5 基于抗体产生的测序 Reads 数据的一般分析流程

3 染色质状态的计算分析

染色质状态可粗略地分为常染色质和异染色质, 染色质的修饰与染色质的状态密切相关, 不同类型的染色质表观遗传修饰对于染色质的状态具有不同的作用, 一些修饰(H3K27me3、DNA 甲基化等)会使得染色质固缩, 位于固缩染色质区域的基因沉默, 而另一些修饰却能使染色质疏松、活化, 如组蛋白的乙酰化和 H3K4me3 等. 随着对染色质研究的深入, 可以将染色质的状态进行进一步细分. 染色质表观修饰的不同组合对应不同状态的染色质状态, 或者说染色质状态可以由位于其上的表观遗传标记来刻画表征, 因此, 在计算上, 可以通过分析染色质表观修饰的类型和组合方式来细化染色质的状态, 这种细化可进一步明确处于各类染色质状态的特点, 有利于分析处于这些染色质中基因的转录活性, 也有助于解析表观遗传修饰(尤其是染色质修饰)的基因调节作用.

目前, 基于染色质修饰信息计算分析染色质状态的报道非常有限, Guillaume 等^[20]通过分析基于 DamID 技术测定的 53 种染色质结合蛋白的信息, 将果蝇的染色质分为 5 种类型, 不同类型的染色质状态和功能各异. 其做法是, 首先将基因组划分为一定长度的片段, 根据 53 种蛋白(包括 PcG 蛋白、HP1、SU(VAR)3-9 和 LHR 等)的结合信息, 将每个片段表示为一个向量, 并进行主成分分析. 结果发现, 这些片段在前 3 个主成分平面上自动聚为 5 类. 然后, 利用状态数目为 5 的隐马尔科夫(HMM)模型识别基因组的各区域的染色质状态类别. 通过分析这 5 种类别染色质所对应的功能, 发现有两种类型对应于异染色质: 经典的异染色质被 SU(VAR)3-9、HP1 和 HP1 的结合蛋白 LHR 和 HP6 等标记, 主要出现在着丝粒区域; 另一类异染色质结合有 PcG 蛋白, Hox 基因族分布在此类染色质中; 第三类染色质约占基因组长度一半的区域, 分布有较少的基因, 结合有 H1、D1、IAL、SUUR 和 LAM, 基因处于抑制状态, 少数有活性的基因与发育有关. 第四和第五类染色质属于常染色质, 大多数基因分布于此二类染色质, H3K4me2 和 H3K79me3 的水平高, 而 H3K27me3 和 H3K9me2 的水平低. 活性基因带有 H3K36me2.

Ernst 等^[83]利用 41 种染色质表观遗传修饰(组蛋白修饰、Pol II 和 CTCF 等)信息, 将人类 CD4⁺ T 细胞的染色质划分为状态不同的 51 种类型. 该研

究与 Guillaume 等的方法有相似之处, 首先将每条染色体划分为等间隔片段, 用一个 0 和 1 构成的向量表示每个片段中各类染色质表观遗传修饰标记的富集程度, 具体方法为: 以泊松分布计算表观遗传修饰标记数目的背景值, 实际的染色质表观遗传修饰标记的数目与背景相比, 大于为 1, 小于为 0. 然后以 k-means 法初测可能的染色质的类型数目, 并用 HMM 计算每种类型(HMM 的状态)中各类标记的组合模式, 以及各种类型的染色质在基因组上的分布. 人类染色质组总计可分 51 种具有不同染色质表观遗传修饰标记组合的类型, 每种类型的基因功能、基因表达量、遗传变异的富集程度等都有所不同. 此研究所解析的染色质的类型以及各类的基因组分布和注释信息均已经嵌在 UCSC 基因组浏览器中(track 类型为 ChromHMM). 新近, 该研究组开发了基于上述方法的染色质状态自动注释工具(ChromHMM)^[84].

新近, Barbieri 等^[85]提出了一种“线珠切换”模型来模拟染色体构象的起始及动态变化, 结果表明, 这种模型可以瞬间重现染色体构象的“分形球”^[85]特征.

4 总结和展望

染色质在基因转录、DNA 复制、DNA 损伤修复等基本生物学过程中具有至关重要的调节作用. 染色质状态决定于 DNA 甲基化、组蛋白修饰、核小体定位、非编码 RNA、染色质结合蛋白、DNA 序列等因素的共同作用. 核小体定位和组蛋白修饰等在基因组的启动子区、增强子区、基因的编码区(尤其是两端)等处具有特异的分布模式, 而且, 这些分布模式具有高度的动态性, 即细胞类型的特异性、环境的特异性. 这些因素之间存在复杂的相互作用, 共同决定染色质状态. 在活性基因和沉默基因的区域, 其染色质状态差异很大. 不正常的染色质状态与一些疾病(如癌症)的发生发展密切相关. 目前测定染色质表观遗传修饰的技术主要基于二代测序技术.

在此领域, 尚有诸多问题需要进一步探索, 包括:

a. 以染色质表观遗传修饰信息解析染色质状态的动态特征. 染色质表观遗传修饰的分布模式在各类细胞中存在差异, 这些差异实际最终体现为染色质状态的差异, 即染色质的动态特征. 关于染色质动态特征的分析目前尚存在一些挑战. 首先, 需

要足够的全基因组各类细胞的染色质状态信息,目前尚没有技术能直接单碱基分辨地测定染色质状态,因此,这项任务在目前技术条件下仍然是一个巨大的挑战。其次,在计算分析层面,高通量数据存储、高复杂度计算仍然是单个实验组很难完成的任务。最后,由于细胞状态的时间连续性,如何选择合适时间点去分析依然需要考虑。

b. 分析染色质表观遗传修饰信息、染色质状态以及基因表达之间的关系。细胞的特征和行为取决于基因表达模式,因此解析染色质状态与基因表达水平的关系至关重要,虽然目前有研究尝试用组蛋白的修饰水平预测基因表达水平,但是全面地揭开这种关联,并明确机理,需要进一步的探索和努力。

c. 分析细胞癌变过程染色质表观遗传修饰的改变过程。染色质修饰的不正常改变与癌细胞的发生发展密切关联,然而,生物学上细节的机理和全局性规律仍然需要查明,由于癌细胞涉及遗传和表观遗传的共同失调,以及两者之间的共同作用,再加上癌细胞类型多等因素,使得该类问题变得异常复杂。

d. 分析环境信号刺激下染色质状态改变的分子途径。染色质状态改变的分子途径在细胞分化、药物作用机理等分析中非常重要,如果能全面掌握染色质状态改变的分子途径,对于识别新的药物分子靶标、定向控制分化等意义重大。

综上,未来关于染色质的分析可能的方向为:

a. 快速灵敏的染色质状态测定技术。该类技术要能实现染色质状态的高通量、高分辨(单碱基分辨)、高速、低成本地测定。b. 染色质状态的动态解析算法和软件。对于高通量的染色质状态数据,必须有算法和工具能够解析染色质状态的动态变化特征和规律,分析不同组织类型细胞之间染色质状态的差异和变化,以及同类型细胞在不同环境下的染色质状态变化,进而分析染色质的细胞功能的调节机制。c. 癌细胞染色质的表观遗传分析。主要包括对癌细胞染色质状态的特征分析,发现新的基于染色质状态的癌细胞检测标识,发现新的针对异常染色质的治疗靶点等。d. 染色质的表观遗传调控机制分析。主要为系统分析染色质相关表观遗传调节机制。

总之,染色质状态的分析是解析表观遗传机制,阐明真核细胞 DNA 生物学过程,建立重大疾病诊断和治疗新方法等的重要研究内容,未来该领

域的研究充满挑战且将富有成果。

参 考 文 献

- [1] Lewin B. Gene VIII. 8th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2004
- [2] Zhou W V, Goren A, Bernstein B E. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, 2011, **12**(1): 7-18
- [3] Bird A. Perceptions of epigenetics. *Nature*, 2005, **447**(7143): 396-398
- [4] Manel E. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*, 2005, **8**(4): 286-298
- [5] Schones D E, Cui K, Cuddapah S, *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 2008, **132**(5): 887-898
- [6] Jiang C, Pugh B F. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*, 2009, **10**(3): 161-172
- [7] Kouzarides T. Chromatin modifications and their function. *Cell*, 2007, **128**(4): 693-705
- [8] Barski A, Cuddapah S, Cui K R, *et al.* High-resolution profiling of histone methylations in the human genome. *Cell*, 2007, **129**(4): 823-837
- [9] Tsai M C, Manor O, Wan Y, *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 2010, **329**(5992): 689-693
- [10] Schuettengruber B, Martinez A M, Iovino N, *et al.* Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol*, 2011, **12**(12): 799-814
- [11] Assaf W, Hughes A, Yassour M, *et al.* High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res*, 2009, **20**(1): 90-100
- [12] Liu H D, Duan X Y, Yu S X, *et al.* Analysis of nucleosome positioning determined by DNA helix curvature in the human genome. *BMC Genomics*, 2011, **12**: 72
- [13] Liu H D, Lin S H, Cai Z W, *et al.* Role of 10-11 bp periodicities of eukaryotic DNA sequence in nucleosome positioning. *Biosystems*, 2011, **105**(3): 295-299
- [14] Weber C M, Henikoff J G, Henikoff S. H2A.Z nucleosomes enriched over active genes are homotypic. *Nat Struct Mol Biol*, 2010, **17**(12): 1500-1507
- [15] Clemens B, Katrin S, Sebastian P, *et al.* H2A.Z.2.2 is an alternatively spliced histone H2A.Z variant that causes severe nucleosome Destabilization. *Nucl Acids Res*, 2012, **40**(13): 5951-5964
- [16] Zaugg B J, Luscombe M N. A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome Res*, 2012, **22**(1): 84-94
- [17] Vishwanath R I. Nucleosome positioning: bringing order to the eukaryotic genome. *Trends Cell Biol*, 2012, **22**(5): 250-256
- [18] Liu H D, Wu J S, Xie J M, *et al.* Characteristics of nucleosome core DNA and their applications in predicting nucleosome positions. *Biophysical J*, 2008, **94**(12): 4597-4604

- [19] You J S, Kelly T K, Carvalho D D D, *et al.* OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proc Natl Acad Sci USA*, 2011, **108**(35): 14497–14502
- [20] Jansen A, Zande E, Meert W, *et al.* Distal chromatin structure influences local nucleosome positions and gene expression. *Nucl Acids Res*, 2012, **40**(9): 3870–3885
- [21] Huebert D J, Kuan P F, Keles S, *et al.* Dynamic changes in nucleosome occupancy are not predictive of gene expression dynamics but are linked to transcription and chromatin regulators. *Mol Cell Biol*, 2012, **32**(9): 1645–1653
- [22] Guillaume J F, Joke G V B, Ulrich B, *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, 2010, **143**(2): 212–224
- [23] Rakyan K V, Down A T, Balding J D, *et al.* Epigenome-wide association studies for common human diseases. *Nat Rev Genet*, 2011, **12**(8): 529–541
- [24] Woo Y H, Li W H. Evolutionary conservation of histone modifications in mammals. *Mol Biol Evol*, 2012, **29**(7): 1757–1767
- [25] Heintzman N D, Hon G C, Hawkins R D, *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 2009, **459**(7243): 108–112
- [26] Boyle A P, Davis S, Shulha H P, *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 2008, **132**(2): 311–322
- [27] Cotney J, Leng J, Sunghee O, *et al.* Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res*, 2012, **22**(6): 1069–1080
- [28] Shu W, Chen H, Bo X, *et al.* Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains. *Nucl Acids Res*, 2011, **39**(17): 7428–7443
- [29] Andersson R, Enroth S, Rada-Iglesias A, *et al.* Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res*, 2009, **19**(10): 1732–1741
- [30] Luco R F, Pan Q, Tominaga K, *et al.* Regulation of alternative splicing by histone modifications. *Science*, 2010, **327**(5968): 996–1000
- [31] Guttman M. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 2009, **458**(7235): 223–227
- [32] Yuan C, Matthews A G W, Jin Y, *et al.* Histone H3R2 symmetric dimethylation and histone H3K4 trimethylation are tightly correlated in eukaryotic genomes. *Cell*, 2012, **1**(2): 83–90
- [33] Bell O, Tiwari K V, Thoma H N, *et al.* Determinants and dynamics of genome accessibility. *Nat Rev Genet*, 2011, **12**(8): 554–564
- [34] Rosenfeld J A, Wang Z B, Schones D E, *et al.* Determination of enriched histone modifications in non-genic portions of the human genome. *BMC Genomic*, 2009, **10**: 143
- [35] Bernstein B E. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 2006, **125**(2): 315–326
- [36] Vastenhouw N L, Schier A F. Bivalent histone modifications in early embryogenesis. *Curr Opin Cell Biol*, 2012, **24**(3): 374–386
- [37] Matthew G G, Frampton G M, Soldner F, *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell*, 2010, **7**(2): 249–257
- [38] Misook H, Danny W-K N, Li W, *et al.* Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res*, 2011, **21**(4): 590–598
- [39] Karlić R, Chung H R, Lasserre J, *et al.* Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci USA*, 2010, **107**(7): 2926–2931
- [40] Gkikopoulos T, Schofield P, Singh V, *et al.* A Role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science*, 2011, **333**(6050): 1758–1760
- [41] Manel E. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet*, 2005, **8**(4): 286–298
- [42] Zilberman D, Coleman-Derr D, Ballinger T, *et al.* Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*, 2008, **456**(7218): 125–129
- [43] Choi J K, Bae J B, Lyu J, *et al.* Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol*, 2009, **10**: R89
- [44] Chodavarapu R K, Feng S, Bernatavichute Y V, *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature*, 2010, **466**(7304): 388–392
- [45] Suganuma T, Workman J L. Crosstalk among histone modifications. *Cell*, 2008, **135**(4): 604–607
- [46] Heintzman N D, Stuart R K, Hon G, *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet*, 2007, **39**(3): 311–318
- [47] Xu C R, Cole A P, Meyers J D, *et al.* Chromatin "Prepattern" and histone modifiers in a fate choice for liver and pancreas. *Science*, 2011, **332**(6032): 963–966
- [48] Hansen K D, Timp W, Bravo H C, *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*, 2011, **43**(8): 768–775
- [49] Kouzarides T. Chromatin modifications and their function. *Cell*, 2007, **128**(4): 693–705
- [50] Gaetano G, Saverio M. Epigenomic profiling of cancer cells. *Internat J Biochem Cell Biol*, 2009, **41**(1): 127–135
- [51] Brinkman A B, Gu H, Bartels S J J, *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res*, 2012, **22**(6): 1128–1138
- [52] Lister R, Pelizzola M, Downen R H, *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 2009, **462**(7271): 315–322
- [53] Lieberman A E, van Berkum N L, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, **326**(5950): 289–293
- [54] Sexton T, Yaffé E, Kenigsberg E, *et al.* Three-dimensional folding

- and functional organization principles of the *Drosophila* genome. *Cell*, 2012, **148**(3): 458–472
- [55] Sanyal A, Lajoie B R, Jain G, *et al.* The long-range interaction landscape of gene promoters. *Nature*, 2012, **489**(7414): 109–113
- [56] Rickman D S, Soong T D, Moss B, *et al.* Oncogene-mediated alterations in chromatin conformation. *Proc Natl Acad Sci USA*, 2012, **109**(23): 9083–9088
- [57] Furey T S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet*, 2012, doi: 10.1038/nrg3306
- [58] Brogaard K, Xi L, Wang J P, *et al.* A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 2012, **486**(7404): 496–501
- [59] Li N, Ye M, Li Y, *et al.* Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods*, 2010, **52**(3): 203–212
- [60] Grimm C, Adjaye J. Analysis of the methylome of human embryonic stem cells employing methylated DNA immunoprecipitation coupled to next-generation sequencing. *Methods Mol Biol*, 2012, **873**: 281–295
- [61] Meissner A, Gnirke A, George W B, *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucl Acids Res*, 2005, **33**(18): 5868–5877
- [62] Brinkman A B, Gu H, Bartels S J J, *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res*, 2012, **22**(6): 1128–1138
- [63] Lieberman-Aiden E, van Berkum N L, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, **326**(5950): 289–293
- [64] Josée D, Todd A R, Ramy A A, *et al.* Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 2006, **16**(10): 1299–1309
- [65] Splinter E, Wit E D, Werken H J G, *et al.* Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Method*, 2012, **58**(3): 221–230
- [66] Lieberman-Aiden E, van Berkum N L, Williams L, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, **326**(5950): 289–293
- [67] Fullwood M J, Liu M H, Pan Y F, *et al.* An oestrogen-receptor-a-bound human chromatin interactome. *Nature*, 2009, **462**(7269): 58–64
- [68] 翟 侃, 武治印, 于典科. 染色质构象捕获及其衍生技术. *生物化学与生物物理进展*, 2010, **37**(9): 939–944
- Zhai K, Wu Z Y, Yu D K. *Prog Biochem Biophys*, 2010, **37**(9): 939–944
- [69] Langmead B, Trapnell C, Pop M, *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 2009, **10**(3): R25
- [70] Liu C M, Wong T, Wu E, *et al.* SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics*, 2012, **28**(6): 878–879
- [71] Weese D, Holtgrewe M, Reinert K. RazerS 3: Faster, fully sensitive read mapping. *Bioinformatics*, 2012, **28**(20): 2592–2599
- [72] Homer N, Merriman B, Nelson S F. BFAST: An alignment tool for large scale genome resequencing. *PLoS One*, 2009, **4**(11): e7767
- [73] Zhang Y, Liu T, Meyer C, *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol*, 2008, **9**(9): R137
- [74] Fejes A P, Robertson G, Bilenky M, *et al.* FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics*, 2008, **24**(15): 1729–1730
- [75] Spyrou C, Stark R, Lynch A G, *et al.* BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*, 2009, **10**: 299
- [76] Tuteja G, White P, Schug J, *et al.* Extracting transcription factor targets from ChIP-seq data. *Nucl Acids Res*, 2009, **37**(17): e113
- [77] Gentleman R C, Carey V J, Bates D M, *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 2004, **5**(10): R80
- [78] Boyle A P, Guinney J, Crawford G E, *et al.* F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics*, 2008, **24**(21): 2537–2538
- [79] Splinter E, de Wit E, van de Werken H J, *et al.* Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: From fixation to computation. *Methods*, 2012, **58**(3): 221–230
- [80] Belton J M, McCord R P, Gibcus J H, *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods*, 2012, **58**(3): 268–276
- [81] Nativio R, Ito Y, Murrell A. Quantitative chromosome conformation capture. *Methods Mol Biol*, 2012, **925**: 173–185
- [82] Kasper D H, Langmead B, Irizarry R A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, 2012, **13**(10): R83
- [83] Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*, 2010, **28**(8): 817–825
- [84] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods*, 2012, **9**(3): 215–216
- [85] Barbieri M, Chotalia M, Fraser J, *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci USA*, 2012, **109**(40): 16173–16178

Patterns of Nucleosome and Chromatin Modifications and Their Effects on Chromatin States*

LIU Hong-De^{1)**}, LUO Kun²⁾, MA Xin³⁾, ZHAI Jin-Cheng¹⁾, XIE Jian-Ming¹⁾, SUN Xiao¹⁾, WAN Ya-Kun⁴⁾

¹⁾ State Key Laboratory of Bioelectronics, Southeast University, Nanjing 210096, China;

²⁾ Department of Neurosurgery, Xinjiang Evidence-based Medicine Research Institute,
The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China;

³⁾ Jinshen College, Nanjing Audit University, Nanjing 211815, China;

⁴⁾ Institute of Life Science, Southeast University, Nanjing 210096, China)

Abstract Chromatin packages eukaryotic DNA, it can affect accessibility of DNA, thus chromatin has important roles in gene transcription regulation, DNA replication, DNA repair, and so on. The basic unit of chromatin is nucleosome consisting of ~ 147 DNA which sharply bends and tightly wraps on surface of octamer of histones. Between two nucleosomes is linker DNA. Histones can be modified by methylation and acetylation. Chromatin state (euchromatin and heterochromatin) is determined by the combinatorial effects of nucleosome positioning, histone modifications, and chromatin-binding proteins. Recently, by virtue of high throughput sequencing technique, the distributions of the chromatin marks were genome-widely determined in various cells. Results indicated that the distribution mode is specific at key sites of genome and is dynamically changed with both cell types and environment stimulus, showing a complex plot. In this paper, we detailedly reviewed the distribution modes of chromatin marks, biological meanings of the modes, interactions of the chromatin marks (nucleosome, DNA and histone modifications, histone variants, chromatin-binding proteins), techniques that were used to determine the marks, and the computational analysis for chromatin states. This work is important in understanding epigenetic regulation mechanism.

Key words chromatin, nucleosome, histone modifications, distribution mode, high throughput DNA sequencing

DOI: 10.3724/SP.J.1206.2012.00434

* This work was supported by grants from The Natural Science Foundation of Xinjiang Uygur Autonomous Region (2012211A076), The National Basic Research Program of China (973) (2012CB316501), The National Natural Science Foundation of China (31240080, 31371339) and Jiangsu Natural Science Foundation(12KJD520006).

**Corresponding author.

Tel: 86-25-83795174, E-mail: liuhongde@seu.edu.cn

Received: December 25, 2012 Accepted: March 19, 2013