

DOI:10.3969/j.issn.1672-5565.201701008

# Hi-C 的干细胞染色质多重相互作用特征分析

张峰,刘亚军,谢建明,孙 啸,刘宏德\*

(生物电子学国家重点实验室(东南大学),南京 210096)

**摘要:**染色质高级结构是基因转录调节的重要因素,染色质多重相互作用是高级结构中的一种,是多个( $\geq 3$ )染色质片段在空间上相互接触而形成的紧凑结构。为了解染色质多重相互作用这类高级结构的特征及其在干细胞中分化中起到的作用,通过对 Hi-C 数据进行相关分析并计算基因的 FPKM 表达量,研究了染色质多重相互作用。分析发现:多重相互作用约占所有作用的 30%,包含近 70% 的基因;此类作用区域的高表达基因多于低表达基因;且与组蛋白乙酰化相关性高。在分化过程中,多重作用位点数目和比例减少;位于多重作用区域的基因的表达略有降低;组蛋白乙酰化(H3K27ac 和 H3K23ac)在多重作用区域的减弱,而组蛋白甲基化(H3K4me3 和 H3K27me3)倾向于增强。结果表明,染色质多重相互作用是一种广泛存在的染色质高级结构,在干细胞分化中有重要作用,此类结构多具有 H3K27ac 修饰,调节基因的表达。总之,染色质多重相互作用是一种重要的基因转录调节因素,在细胞分化中具有调控作用。

**关键词:**染色质多重相互作用;Hi-C;基因表达;组蛋白修饰

**中图分类号:**R737.9 **文献标志码:**A **文章编号:**1672-5565(2017)03-137-05

## Characteristic analysis of chromosome multiple interactions in stem cell based on Hi-C

ZHANG Feng, LIU Yajun, XIE Jianming, SUN Xiao, LIU Hongde\*

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

**Abstract:** Higher-order chromatin structure is an important factor for transcriptional regulation. Chromatin multiple interaction (CMI) is a kind of compact high-level structure in which more than three chromatin fragments spatially interconnect together. In order to understand the characteristic of the CMI and its change in stem cell differentiation, the CMI was investigated by both correlation analysis of Hi-C data and gene expression (FPKM). The results show that CMI sites occur in 30% of Hi-C reads, covering nearly 70% of human genes, in which the highly expressed genes are more than those with low expression. Meanwhile, the CMI is highly correlated with histone acetylation. During stem cell differentiation, the number and the proportion of CMI sites decrease, and the expression slightly decreases in the CMI regions. The histone acetylation (H3K27ac and H3K23ac) is weakened in the CMI regions, whereas histone methylation (H3K4me3 and H3K27me3) tends to increase. These results suggest that the CMI is widely distributed, mostly modified with histone H3K27ac and plays a role in differentiation and gene regulation. In conclusion, chromatin multiple interaction is kind of important transcriptional regulator and contributes to cell differentiation regulation.

**Keywords:** Chromosome multiple interaction (CMI); Hi-C; Gene expression; Histone modification

真核生物染色质会通过核小体与 DNA 链的多重缠绕折叠形成紧密的三维高级结构<sup>[1-2]</sup>,这种结构在基因表达中远程调节的物理实现上起到了关键

作用<sup>[3-5]</sup>。胚胎干细胞(ES cells)是早期胚胎的原始细胞,具有分化为三胚层各种细胞<sup>[6]</sup>的潜在能力。在干细胞分化过程中,细胞核内染色质的三维

收稿日期:2017-01-22;修回日期:2017-05-02.

基金项目:国家自然科学基金(31371339);国家重点基础研究发展计划(973计划)(2012CB316501).

作者简介:张峰,男,硕士研究生,研究方向:生物信息学;E-mail: zf564js@163.com.

\* 通信作者:刘宏德,男,副教授,研究方向:生物信息学;E-mail: 101100344@seu.edu.cn.

结构会发生相应的构象变化<sup>[7]</sup>。在哺乳动物<sup>[8]</sup>中通过表达某一基因能够激活其他成体细胞特征性基因,致使细胞命运发生改变,而染色质三维结构变化可对这一过程进行调控。研究干细胞与其衍生细胞的染色质结构差异可以更深入地了解这种调控关系。

利用 Hi-C 技术<sup>[9]</sup>,研究者可以高通量地获得染色质高级结构中相互作用的 DNA 片段。通过检测这些 DNA 片段在染色质上的分布,可以绘制染色质相互作用图谱<sup>[10]</sup>,此图谱展示了染色质上各位点间的相互作用情况。然而,除了存在单个位点仅与单个位点作用的情况外,还存在单个位点与多个位点作用以及多个位点与多个位点相互作用的情况(图 1),本文将这种位点称为染色质多重相互作用位点(chromosome multiple interaction sites, CMI sites)。这些位点可能涉及某些复杂的生物学过程,甚至是其中的关键节点。

如果将 CMI 位点所涵盖的范围称之为 CMI 区域,那么染色质上其他区域则可称为非 CMI 区域。比较不同细胞系中这两种区域内的表观遗传修饰富集情况、基因数目及其表达情况可以粗略的反映出 CMI 位点与细胞分化调控的关联性。目前,尚未见此分析,本文通过分析 Hi-C 数据,系统分析了染色质多重相互作用的特征以及在细胞分化中的变化。

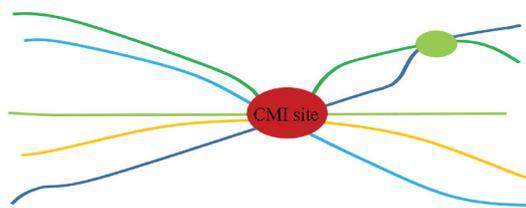


图 1 CMI 位点示意

Fig.1 An illustration of CMI site (red oval)

注:细胞中染色质(线条)通过 DNA 位点的接触来相互作用,包括发生在两个 DNA 片段间的作用(绿色椭圆),和发生在多个 DNA 片段间的作用(红色椭圆)。

## 1 数据和方法

### 1.1 数据

本文利用文献[11]提供的 Hi-C Reads 回帖后的数据,包括人类 ES 细胞系以及中内胚层细胞(ME cells)、间充质干细胞(MS cells)、神经祖细胞(NP cells)、滋养层细胞(TR cells)等 4 种 ES 衍生细胞系,获取存在多次相互作用的位点。同时,结合 mRNA 和人类参考基因组数据,以及组蛋白甲基化

与乙酰化、CTCF、DNA 甲基化(DNA Meth)、DNaseI 高敏感位点(DHS)等表观遗传信息,本文进行了进一步的分析。所有数据的 GEO 登录号为 GSE52457。

### 1.2 方法

首先统计各细胞系中相同位置 Hi-C Reads 的出现频率,频率超过 2 就意味着此位点在 Hi-C 数据中重复出现,即视为 CMI 位点。然后进一步统计 CMI 位点数目与其作用频率,可得到 CMI 位点在所有 Hi-C 位点中的比例以及同一作用频率位点的数目。最后,通过汇总 CMI 位点在染色质中的位置信息,计算 CMI 区域所涵盖的范围(以 bp 为单位)。通过将 CMI 位点匹配到人类基因组上,计算与 CMI 区域有接触的基因数目。

结合 RNA-Seq 数据,进一步计算各基因的 FPKM 值,从而反映基因的表达水平。本文取整个细胞系中 FPKM 值的中位数为标准,来判定某个基因是否高表达,FPKM 值高于中位数的基因被认为是高表达的基因,FPKM 值低于中位数的基因则被认为是低表达的基因。

以多个分辨率(10~100 kbp,步长为 10 kbp)计算 CMI 区域和所有相互作用区域上表观遗传修饰 reads 数目,即两区域中表观遗传修饰发生频率的总和,两者之间的比率作为 CMI 区域发生表观遗传修饰的相对概率。然后,计算各个 CMI 位点的 Hi-C 作用频率以及染色质各位点上组蛋白修饰的频率,对染色质上非 CMI 位点以及未发生组蛋白修饰位点的作用频率记为 0,进而对两组数据进行相关(皮尔森相关系数),并在不同分辨率下求平均相关性系数,以此反映 CMI 频率与组蛋白修饰频率的相关性强弱。

对比 ES 细胞系与其 4 种衍生细胞系的 CMI 位点数目、CMI 区域相关基因数目及表达量、CMI 区域相关表观遗传修饰数目、CMI 与组蛋白乙酰化与甲基化的相关性系数,从而比较细胞分化过程中人类 ES 细胞与其分化后细胞的差异。

## 2 结果与讨论

### 2.1 染色质多重相互作用(CMI)的数目及分布

CMI 位点的统计结果表明其占有所有相互作用位点数目的近 30%(图 2(a)),单点作用频率最高达 500 次,所有 CMI 位点遍布整个染色质。可以看出 CMI 位点在所有染色质相互作用中占据了一定的重要性。其中 ES 细胞系的所有相互作用位点总数以及 CMI 位点数均明显多于其他 4 种衍生细胞系

(见表 1)。同时,不同作用频率的 CMI 位点的出现次数也会随着作用频率的增加而衰减,且 ES 细胞具有

更多的复杂相互作用(图 2(b))。这说明 CMI 会随着细胞分化而变化,但总体分布的趋势相对保守。

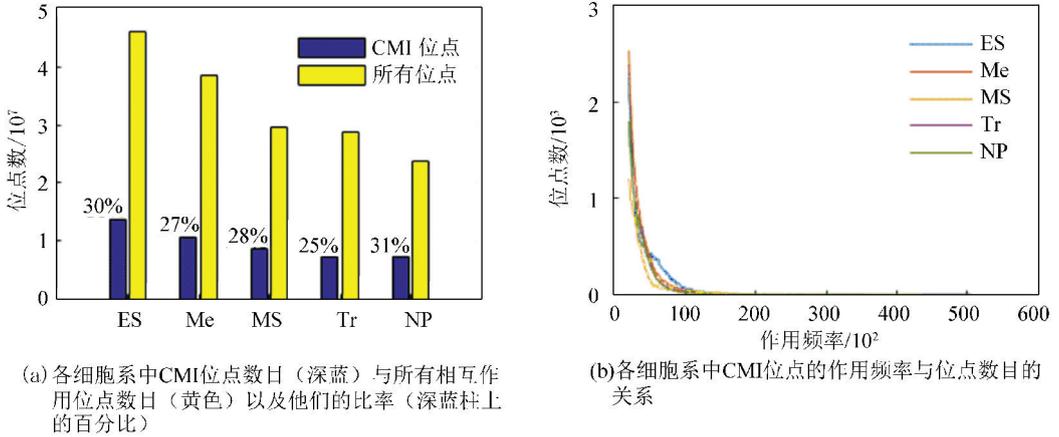


图 2 CMI 位点数目与频率

Fig.2 The frequency and count of CMI sites

表 1 CMI 位点数目统计

Table 1 Count of CMI sites

细胞系	胚胎干细胞	中内胚层细胞	间充质干细胞	神经祖细胞	滋养层细胞
CMI 位点数目	13 737 681	10 519 216	8 444 522	7 278 627	7 199 033
所有位点数目	46 320 680	38 547 004	29 768 456	23 789 733	28 868 083

### 2.2 染色质多重相互作用区域的基因表达分析

通过统计所有 CMI 位点及 Hi-C 位点占据的染色质区域,本文发现 CMI 区域约占所有相互作用区域的 15%,但是与人类基因的匹配结果显示其与近 70%的基因有接触。这表明 CMI 可能参与了许多复杂的生物学过程。

值得注意的是,通过计算各个细胞系中不同区域的基因平均 FPKM 表达量,发现 CMI 区域中基因表达量明显高于非 CMI 区域(图 3(a))。进一步计算

高表达与低表达基因数目在两区域中的比例,发现 CMI 区域的高表达基因明显多于低表达基因,而非 CMI 区域的高表达基因则明显少于低表达基因,并且,CMI 区域高表达基因显著的多于非 CMI 区域(图 3(b))。可以推测,与 CMI 区域相关联的基因倾向于高表达,与非 CMI 区域相关联的基因倾向于低表达,但是这种倾向性随着细胞分化有减弱的趋势。上述结果说明,在干细胞中,有更多的 CMI 形成,且随分化而逐渐减弱,暗示 CMI 是干细胞的标识性结构。

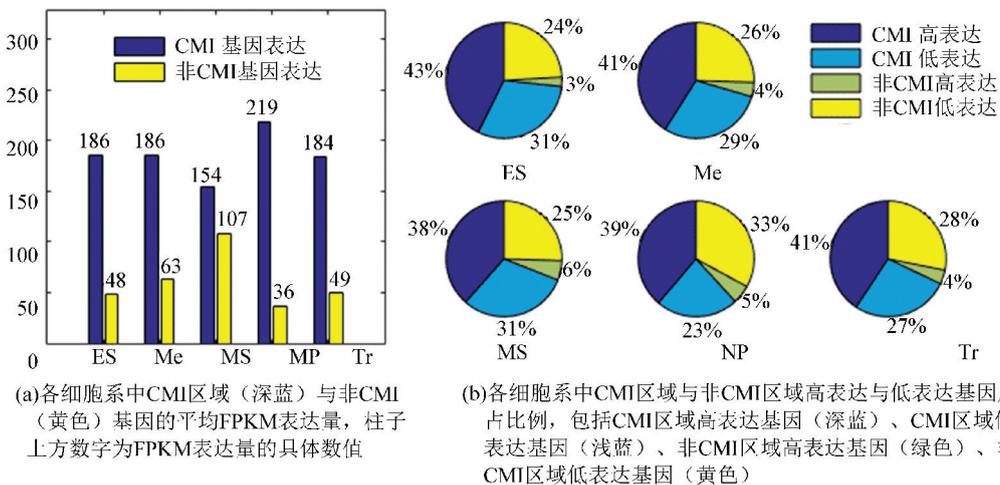


图 3 CMI 区域基因表达情况

Fig.3 Gene expression in CMI regions

### 2.3 染色质多重相互作用区域的表观遗传修饰

通过比较 CMI 区域与所有染色质相互作用区域表观遗传修饰发生的频率,发现仅占染色质相互作用区域 15% 的 CMI 区域发生表观遗传修饰的相对概率能达到 30%~60%,表明表观遗传修饰在 CMI 区域存在富集现象。但是 ES 细胞系中 CMI 区域发生表观遗传修饰的相对概率明显大于其他 4 种衍生细胞系(图 4(a)),这表明干细胞分化可能会导致这种富集效应减弱。为了进一步研究 CMI 区域与表观遗传修饰之间的关系,将 CMI 位点的相互作用频率与表观遗传修饰的发生频率进行相关性分析,发现 CMI 的作用频率与某些表观遗传修饰发生频率具有较强相关性(相关性系数介于 0.6~0.8 之间),也有一些具有弱相关性(相关性系数介于 0.4~0.6 之间),不同细胞系之间的频率相关性(以下简称相关性)具有显著差异(图 4(b))。

ES 细胞系中 CMI 整体与组蛋白乙酰化的相关性较强,而与组蛋白甲基化的相关性较弱,比如与 H3K23ac、H3K4ac、H2BK120ac 的相关性均大于 0.6,而与组蛋白甲基化相关性最大仅 0.5,与其他组蛋白甲基化更是几乎不相关(相关性系数低于 0.4);Me 细胞系与 H3K4 甲基化的相关性相比于 ES 细胞系显著增加,与其他乙酰化相关性均有下降,与组蛋白

乙酰化相关性明显增强;MS 细胞系与组蛋白修饰的相关性整体明显下降;NP 细胞系与组蛋白甲基化的相关性显著上升,同时与组蛋白乙酰化相关性显著增强;TR 细胞系与组蛋白乙酰化的相关性急剧减弱,整体变化比较复杂。如果 H3K27ac 是增强子的重要标志,那么该结果表明此类 Enhancer 在空间上可能和多个染色质的区域靠近并调控相关基因<sup>[12-13]</sup>。

总体来看,随着分化过程中 CMI 区域在所有相互作用区域中表观遗传修饰的占比减少,虽然 4 种 ES 衍生细胞系与组蛋白乙酰化的相关性相对于 ES 细胞系趋于弱化,但是与组蛋白甲基化的相关性趋于强化,其中强化最明显的是与 H3K4me3 的相关性。可以推测,干细胞分化过程中与组蛋白乙酰化相关的 CMI 位点趋于减少,而与组蛋白甲基化相关的 CMI 位点趋于增多,从而共同维护了染色质相互作用中 CMI 位点与非 CMI 位点以及区域范围大小的动态平衡。H3K4me3 和 H3K27me3 的双重修饰是分化基因的重要标志<sup>[14-15]</sup>,本文结果表明分化中这两种甲基化修饰在 CMI 区域中增强(图 4(b)),这说明分化过程中 H3K4me3 和 H3K27me3 的双重修饰可能凭借多重作用结构,一次调控多个区域的基因转录。

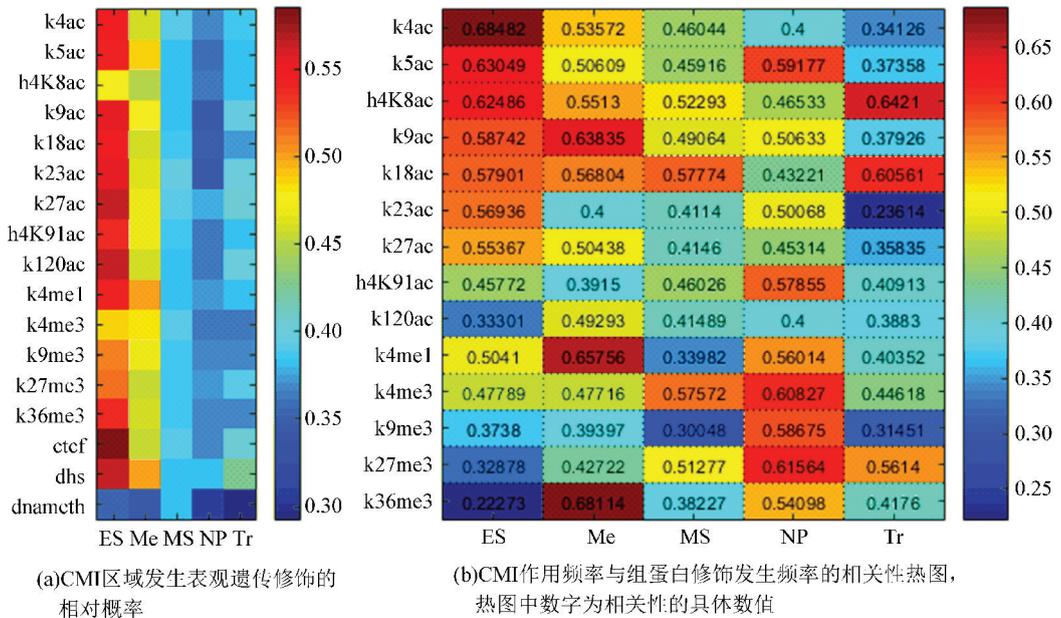


图 4 各细胞系中 CMI 区域的表观遗传修饰情况

Fig.4 Epigenetic modifications in CMI regions of cell lines

## 3 结论

本文基于 Hi-C 数据,系统分析了一类特殊的染

色质高级结构——染色质多重相互作用的特征。结果表明:染色质多重相互作用在所有染色质高级结构中占有较高的比例(约 30%),而且此类结构关联的基因占有所有基因的 70%。染色质多重相互作用

富集组蛋白修饰,具有高乙酰化的特征,位于此类区域的高表达基因比例较大。在干细胞分化期间,染色质多重相互作用数量和比例有变小的趋势,H3K27ac的修饰有减弱的趋势,而甲基化却有增加的趋势(H3K4me3和H3K27me)。推测这些修饰可能凭借多重作用结构,一次调控多个区域的基因转录。总之,染色质多重相互作用在干细胞分化过程中动态变化,具有重要的基因转录调节作用。

## 参考文献 (References)

- [1] HU Ming, DENG Ke, QIN Zhaohui, et al. Understanding spatial organizations of chromosomes via statistical analysis of Hi-C data[J]. *Quantitative Biology*, 2013, 1(2):156-174. DOI: 10.1007/s40484-013-0016-0.
- [2] MIELE A, DEKKER J. Long-range chromosomal interactions and gene regulation[J]. *Molecular BioSystems*, 2008, 4(11):1046-1057. DOI: 10.1039/b803580f.
- [3] GORKIN D U, LEUNG D, REN Bing. The 3D genome in transcriptional regulation and pluripotency[J]. *Cell Stem Cell*, 2014, 14(6):762-775. DOI: 10.1016/j.stem.2014.05.017.
- [4] LI Guoliang, RUAN Xiaolan, AUERBACH R K, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation[J]. *Cell*, 2012, 148(1/2):84-98. DOI: 10.1016/j.cell.2011.12.014.
- [5] HANDOKO L, XU Han, LI Guoliang, et al. CTCF-mediated functional chromatin interactome in pluripotent cells[J]. *Nature Genetics*, 2011, 43(7):630-638. DOI: 10.1038/ng.857.
- [6] THOMSON J A, ITSKOVITZ-ELDOR J, SHAPIRO S S, et al. Embryonic stem cell lines derived from human blastocysts[J]. *Science*, 1998, 282(5391):1145-1147.
- [7] JOFFE B, LEONHARDT H, SOLOVEI I. Differentiation and large scale spatial organization of the genome[J]. *Current Opinion in Genetics & Development*, 2010, 20(5):562-569. DOI: 10.1016/j.gde.2010.05.009.
- [8] DAVIS R L, WEINTRAUB H, LASSAR A B. Expression of a single transfected cDNA converts fibroblasts to myoblasts[J]. *Cell*, 1987, 51(6):987-1000. DOI: 10.1016/0092-8674(87)90585-X.
- [9] LIEBERMAN-AIDEN E, van BERKUM N L, WILLIAMS L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the Human genome[J]. *Science*, 2009, 326(5950):289-293. DOI: 10.1126/science.1181369.
- [10] DEKKER J, MARTI-RENO M A, MIRNY L A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data[J]. *Nature Reviews Genetics*, 2013, 14(6):390-403. DOI: 10.1038/nrg3454.
- [11] DIXON J R, JUNG I, SELVARAJ S, et al. Chromatin architecture reorganization during stem cell differentiation[J]. *Nature*, 2015, 518(7539):331-336. DOI: 10.1038/nature14222.
- [12] ZHU Yun, SUN Lin, CHEN Zhao, et al. Predicting enhancer transcription and activity from chromatin modifications[J]. *Nucleic Acids Research*, 2013, 41(22):10032-10043. DOI: 10.1093/nar/gkt826.
- [13] CREYGHTON M P, CHENG A W, WELSTEAD G G, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107(50):21931-21936. DOI: 10.1073/pnas.1016071107.
- [14] BOGLIOTTI Y S, ROSS P J. Mechanisms of histone H3 lysine 27 trimethylation remodeling during early mammalian development[J]. *Epigenetics*, 2012, 7(9):976-981. DOI: 10.4161/epi.21615.
- [15] LIU Xiaoyu, WANG Chenfei, LIU Wenqiang, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos[J]. *Nature*, 2016, 537(7621):558-562. DOI: 10.1038/nature19362.