

## 特约综述



本课题组重点开展以下两个方面的研究: (1)基于染色质高级结构动态变化基础上的基因远程调控。通过遗传和表观遗传分析方法, 分析鉴定基因组中特定基因间区及基因非编码区的功能性调控元件, 并在此基础上探讨调控元件对靶基因的远距离调控机制及其生物学效应, 以期从染色质三维空间动态变化的角度认识远端调控元件在基因调控网络中的作用。(2)乳腺癌细胞耐药机制及生物标志物鉴定。运用分子生物学和表观遗传分析方法, 分析比较ER $\alpha$ (+)和ER $\alpha$ (-)乳腺癌耐药的分子机制及表观遗传学特征, 并在此基础上预测和评估耐药的标志物。

<http://gnjy.njmu.edu.cn/>

## 浅谈染色质高级结构与基因转录的远程调控

孙 鸾 孙玉洁\*

(江苏省人类功能基因组学重点实验室, 南京医科大学基础医学院细胞生物学系, 南京 210029)

**摘要** 精确的基因表达调控是细胞分化、个体发育和细胞维持正常生命活动的必要条件, 转录调控是真核细胞基因表达调控最关键的环节, 其神秘和精深吸引着无数科学家为之奋斗不已。染色质构象捕获及其衍生技术的建立和2003年启动的“DNA元件百科全书”计划, 将人们对基因转录调控的认识从二维层面推向三维空间。基因组中分布着众多调控元件, 它们与所调控的靶基因间可相距几万甚至几十万个核苷酸, 可以与靶基因位于相同或不同的染色体上。依据染色质环模型, 调控元件可通过染色质环高级结构, 与靶基因在空间上充分接近并相互作用, 发挥其调控功能。同一个调控元件可以调控不同的靶基因, 而相同的基因亦可能受不同调控元件的调节, 由此细胞在染色质高级结构层面形成了一个复杂的调节基因转录活性的三维网络。该文分别从基因远程调控现象的发现、研究方法、相关机制及面临的挑战等方面作一简要综述。

**关键词** 基因转录; 远距离调控; 染色质高级结构

## High-ordered Chromatin Structure and Long-range Regulation of Gene Transcription

Sun Luan, Sun Yujie\*

(Key Laboratory of Human Functional Genomics of Jiangsu Province, Department of Cell Biology, School of Basic Medical Sciences, Nanjing Medical University, Nanjing 210029, China)

**Abstract** The precise control of gene expression is essential for cell differentiation, individual development and normal life activity of a cell. Transcriptional regulation is a key step to control gene expression in eukaryocyte.

国家自然科学基金(批准号: 81301897)和南京医科大学科技发展基金重点项目(批准号: 2012NJMU012)资助的课题

\*通讯作者。Tel: 025-86862731, E-mail: yujiesun@njmu.edu.cn

This work was supported by the National Natural Science Foundation of China (Grant No.81301897) and the Key Project of Science and Technology Development Foundation of Nanjing Medical University (Grant No.2012NJMU012)

\*Corresponding author. Tel: +86-25-86862731, E-mail: yujiesun@njmu.edu.cn

网络出版时间: 2014-11-07 09:22

URL: <http://www.cnki.net/kcms/doi/10.11844/cjcb.2014.11.9002.html>

The development of chromatin conformation capture (3C) technique with its derived methods and the encyclopedia of DNA elements (ENCODE) project started from 2003 have revealed thousands of putative regulatory elements scattered throughout genome. These regulatory elements are far from their target genes and located either on the chromosome same as their target genes or on the different chromosomes. Transcription of one target gene can be regulated by multiple regulatory elements and one regulatory sequence can also control more than one target genes, which forms a complicated net work at the level of high-ordered chromatin structure. The commonly accepted chromatin loop model is that DNA sequences between enhancers and promoters form loops to allow the distal regulatory sequences to interact directly with the promoters. Here we briefly review the discovery of long-range regulation of gene expression, development of 3C and the derived techniques, related mechanism involved in long-range regulation and facing challenges within the area.

**Key words** gene transcription; long-range regulation; high-ordered chromatin structure

精确的基因表达调控是细胞分化、个体发育和细胞维持正常生命活动的必要条件。转录调控是真核细胞基因表达调控最关键的环节。基因组中编码蛋白质的基因总数仅占基因组的1.5%, 而调控元件的数目远远高于蛋白编码基因的数量。基因组中的调控元件可与所调控的靶基因位于相同或不同染色体上, 与靶基因相距几万甚至几十万个核苷酸, 在染色质高级结构层面形成了一个复杂的调节基因转录活性的三维网, 而人们对其认识还十分有限。探讨染色质三维空间结构的动态变化与基因远程调控的关系及其功能, 已成为后基因组时代一个活跃的领域。随着染色质空间捕获及相关技术的迅速发展, 基因表达远程调控的染色质环模型已经得到普遍认可。该模型认为, 调控元件可通过染色质环高级结构, 与靶基因在空间上充分接近并相互作用, 激活或抑制靶基因的活性, 发挥其远程调控功能。然而, 这个领域还存在许多未解之谜。本文将从远距离调控现象的发现、相关实验技术的发展、染色质高级结构形成的分子机制等方面简要综述染色质三维空间构象与基因表达调控之间的关系。

## 1 基因转录及远距离调控的发现

真核细胞转录起始是一个极其复杂的过程。在多种转录因子、组蛋白修饰酶、染色质重塑酶的作用下使局部染色质解螺旋呈现松散开放的状态, 便于招募TBP、TFIIB、TFIIE、TFIIF以及RNA聚合酶等几十种蛋白质形成巨大的转录起始复合物, 为起始基因转录做好充分准备。细胞核内转录的起始高效有序, 反式作用因子识别特殊的DNA序列(即顺式作用元件), 转录复合物中的各种蛋白质成分可直接

或间接结合在DNA元件上。因此, 以往对转录调控元件的搜寻通常局限在基因启动子附近。

一个偶然的的机会, 人们发现,  $\beta$ -globin基因簇远端一段非编码区的缺失或突变可引起地中海贫血症患者血红蛋白珠蛋白 $\beta$ 链表达受抑制。在人类和小鼠中, 血红蛋白基因簇高度保守且具有相似的表达模式。人 $\beta$ -globin基因簇位于11号染色体, 全长60 Kb, 包含了五个编码血红蛋白亚基的基因, 包括在胚胎期表达的 $\epsilon$ 、在胎儿期表达的 $G\gamma/A\gamma$ 和成年后表达的 $\beta/\delta$ 。在 $\epsilon$ 基因上游25 Kb处有一段长20 Kb、含5个组织特异性DNase I超敏位点(hypersensitive site, HS)的基座控制区(locus control region, LCR)<sup>[1]</sup>。转基因研究发现, 将编码 $\beta$ 链的基因及其附近区域导入小鼠体内或细胞中其表达很低, 如果将LCR片段插在基因上游则可上调其表达水平至100~1 000倍。反之, 在小鼠中敲除LCR则将基因转录水平抑制在很低的基础水平。随后, 人们在 $\alpha$ -globin、HoxD、 $T_H2$ 细胞因子、H19/Igf2等几个基因簇上也陆续观察到了类似的现象, 由此人们意识到, 具有调控功能的DNA元件并不一定要临近靶基因<sup>[2-4]</sup>。

问题接踵而至。如果调控元件位于靶基因的远端, 那么这些调控元件如何影响靶基因的转录活性? 染色质环(chromatin loop)模型认为间期细胞中线性的染色质纤维可以被弯曲形成环状结构, 在环的基部相距甚远的DNA元件同时锚着于蛋白质上, 使得远端调控元件上的反式作用因子能够接触到靶基因启动子从而发挥转录调控的功能。虽然这个假说合理地解释了远距离调控元件的作用方式, 但一直没有直接的实验证据可以证实, 直到2002年染色质空间捕获技术的出现。

## 2 染色质空间捕获技术的兴起与发展

2002年染色质空间捕获技术(Chromatin Conformation Capture, 3C)的出现使人们可以通过简单的酶切、连接等步骤将染色质三维空间上的相互作用转变成能被PCR检测的线性结构。运用3C技术分析不同发育阶段小鼠红细胞中 $\beta$ -globin基因簇的结构,首次为染色质环模型提供了直接证据<sup>[5]</sup>。3C技术的原理和主要步骤可概括为图1。简单地说,首先将细胞在活性状态下用甲醛处理固定交联,形成稳定的DNA-蛋白复合物。继而裂解细胞得到完整的细胞核,选用适当的限制性内切酶切割交联过的DNA,消化后的DNA片段在极低浓度下进行分子内连接,由此使得在空间上充分接近的两个DNA序列的断端被特异性地连接。将连接产物逆交联,使DNA和蛋白质分离,抽提出DNA,用于PCR检测。PCR引物分别设计在可能存在相互作用的两段DNA序列上,如果能得到相应的PCR产物,提示这两个远端DNA元件可能存在相互作用(图1)。在3C实验中,由于连接反应是随机发生的,因此容易产生假阳性结果。为了保证实验的可靠性,通常每个样品要配以严格的阴性对照,即不加入甲醛固定或者不加入连接酶连接都不应得到目的PCR产物。此外,反应体系中DNA的浓度需要被严格控制以保证连接反应发生在相互作用的DNA-蛋白复合物分子内部。最后,还需要将PCR产物进行测序以保证实验结果的真实

性。普通3C实验的结果能够表明两个DNA元件是否存在相互作用,但不能显示两者相互作用发生的频率,而相互作用频率的变化往往能提示其所蕴含的生物学功能。若要检测远端元件相互作用频率的动态变化,需要做定量3C(q3C),即通过构建3C实验中的严格的参照系统,用绝对定量的方式比较不同实验条件下PCR产物量的改变<sup>[6-7]</sup>。由于3C技术设计巧妙,结果直观,操作步骤相对简单,一经建立便得到广泛应用,染色质环学说也逐渐受到学术界的认可。

3C实验最后一步得到的DNA实际上是全基因组范围内的DNA酶切连接的产物。这其中不仅包含了预计可能发生相互作用的两个已知片段,还包含更多未知的但存在相互作用的DNA片段,被称为“3C文库”。为了探索这些大量的未知资源,2006年在3C的基础上又衍生出了4C(Circular Chromosome Conformation Capture)<sup>[8]</sup>和5C(Chromosome Conformation Capture Carbon Copy)<sup>[9]</sup>等多个用于大规模检测基因远距离相互作用的技术。正如之前所介绍的,3C实验首先需要预测可能发生相互作用的DNA序列,并将引物分别设计在两个已知DNA序列上。4C技术则是将引物设计在一个已知序列上,通过反向PCR联合芯片或测序筛选出基因组中所有与该已知序列相互作用的未知DNA片段。5C技术则是采用多对引物标记发生相互作用的DNA片段,最后通过

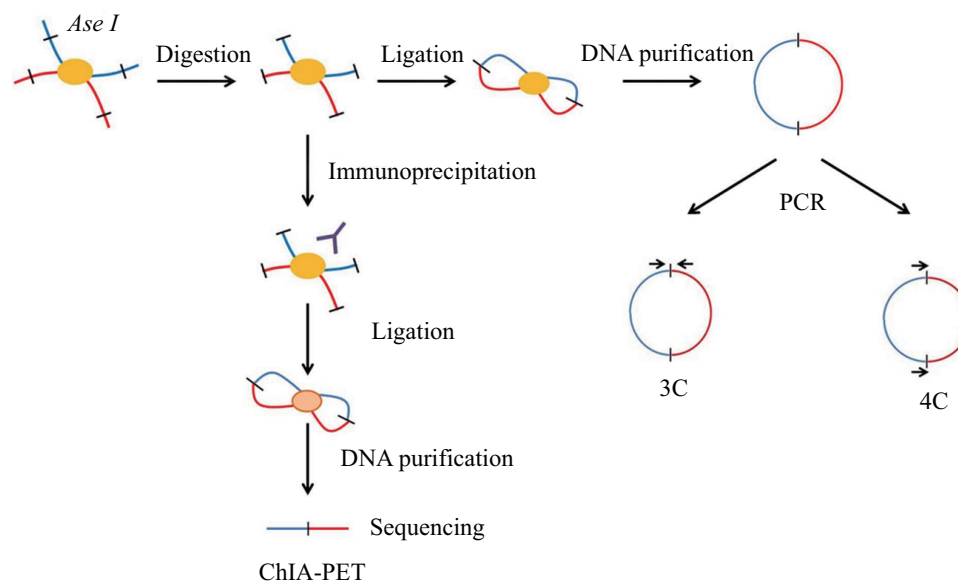


图1 3C及其衍生技术原理图

Fig.1 The schematic illustration of 3C and the derived technique



测序技术鉴定多个调控元件之间的相互作用。简而言之, 3C检测到的是一对一的相互作用, 4C能检测到一对多的相互作用, 5C则能检测到多对多的相互作用。由于5C技术在实际应用时特异性不够理想, 3C的发明人Job Dekker又将技术进一步优化产生了Hi-C。Hi-C在DNA片段进行分子内连接时将连接末端标记上生物素, 使得最后得到的文库更加具有特异性<sup>[10]</sup>。

以3C为基础的一系列技术虽然可以证实两个远端DNA元件在空间上相互靠近, 但是却无法检测锚着在染色质环基部的蛋白质, 而这对于认识远端调控元件与靶基因的相互作用的机制至关重要。为了解决这一问题, 相继发展出了ChIP-loop<sup>[11]</sup>及ChIA-PET(chromatin interaction analysis by paired-end tag sequencing)<sup>[12]</sup>技术。ChIP-loop实验在细胞固定交联后用超声或者酶切的方式使基因组断裂成小片段, 继而用相应的抗体进行免疫沉淀, 回收和特定抗体结合的DNA-蛋白复合物, 经连接、逆交联后采用PCR检测两个DNA片段是否具有相互作用。ChIA-PET在ChIP-loop的基础上联合深度测序技术在全基因组范围内检测由特定蛋白介导下发生相互作用的染色质片段。

### 3 染色质高级结构动态变化是基因表达远距离调控的基础

随着技术的日趋成熟, 众多远端调控元件相继被证实。与此同时, 人类基因组计划的研究结果也显示编码蛋白质的基因大约只占了全基因组DNA的1.5%, 另有45%是重复序列, 剩下超过50%的DNA序列由于功能不明确暂时被称为“垃圾序列”或“基因荒漠”<sup>[13-15]</sup>。在真核细胞基因组中, 基因编码区的DNA序列, 即外显子, 通常被数个内含子所间隔开, 而基因与基因之间存在大量的空白序列。因此, 继人类基因组测序计划完成之后, 2003年9月, 美国国立人类基因组研究院(US national human genome research institute, NHGRI)启动了“DNA元件百科全书”(encyclopedia of DNA elements, ENCODE)计划。来自英国、美国、西班牙、新加坡和日本的422名科学家参与其中, 通过定位基因组中转录区域、分析转录因子在全基因组的分布特征、染色质结构及组蛋白修饰特征, 详细解读了人类基因组中功能性元件。阶段性研究成果于2012年陆续发表在Nature

系列杂志上。首先, 根据对已知功能区域的分析, 总结出与功能相关的特征, 如具有活性的增强子通常结合了大量组蛋白修饰酶CBP/p300, 同时伴有组蛋白H3的特征性修饰, 包括H3K4me1、H3K4me2以及H3K27ac, 而绝缘子上则结合了大量的阻遏物CTCF。继而, 研究人员以蛋白质为“诱饵”, 运用ChIP-sequencing技术对全基因组序列进行高通量分析得到大量具有这些特征的片段, 提示这些片段可能具有潜在的调控功能<sup>[16-19]</sup>。与此同时, 4C、5C、ChIA-PET等技术也从染色质空间结构的角度证实, 全基因组范围内的远端DNA元件之间的相互作用是极其普遍的, 这当中不仅有调控元件和靶基因间的相互作用, 还包括许多调控元件与基因间序列的相互作用<sup>[20-21]</sup>。综合分析这些信息可以看出, 真核细胞的调控元件遍布于全基因组中, 细胞通过染色质高级结构的动态变化协调众多调控元件的功能。

下面以研究较为深入的 $\alpha$ - $\beta$ -globin基因簇和T<sub>H</sub>2细胞因子基因簇为例, 介绍基于染色质环高级结构基础上的基因远程调控。

#### 3.1 $\alpha$ / $\beta$ -globin基因簇与红细胞发育

人们对远距离激活基因转录的研究始于个体发育过程中血红蛋白基因的选择性表达。血红蛋白是高等生物体内负责运输氧气的一种蛋白质。人体内98%的血红蛋白由两分子 $\alpha$ 亚基和两分子 $\beta$ 亚基组成, 即 $\alpha_2\beta_2$ 。每个亚基又由一条肽链和一个血红素分子构成, 这条肽链即珠蛋白。编码 $\alpha$ -珠蛋白( $\alpha$ -globin)和 $\beta$ -珠蛋白( $\beta$ -globin)的基因分别位于人16号和11号染色体上, 多个编码珠蛋白的基因连续排列形成基因簇, 在人体发育的不同阶段顺序性表达。这一过程如果出现异常则会导致血红蛋白携氧能力低下, 引起恶性贫血。那么机体如何使珠蛋白基因簇上的众多基因在正确的时间表达或静息呢?

$\beta$ -globin基因簇是最早被鉴定出受到远端增强子基座控制区(LCR)调控的基因簇, 也是迄今为止机制研究最为深入的模型<sup>[22-23]</sup>。在不同发育阶段的红细胞中, LCR可与不同的靶基因形成染色质环, 激活需要被表达的基因, 而其他不与LCR靠近的基因则沉默。3C实验已证实, 胎儿期LCR与 $\gamma$ -globin基因靠近, 而在成体阶段LCR与 $\beta$ -globin基因发生相互作用。据报道, LCR每个HS位点上都存在多个转录因子的识别位点, 如GATA-1、EKLF、NF-E2、SP1及SATB1等<sup>[24-28]</sup>。结合在LCR上的转录因子通过染

色质环与靶基因启动子靠近并激活基因转录。敲除EKLf或干扰SATB1表达可显著减少LCR与基因启动子的相互作用<sup>[29]</sup>,而在GATA1<sup>-/-</sup>小鼠中过表达GATA1则可增加两者之间的相互作用频率<sup>[30-31]</sup>。此外,LCR还可以通过招募转录延伸因子、影响RNA聚合酶II磷酸化状态,促进靶基因转录的延伸<sup>[27]</sup>。可见,LCR远距离调控功能依赖于LCR上蛋白复合物对靶基因启动子的作用。此外,Ragoczy等<sup>[32-33]</sup>发现,红细胞分化过程中LCR的位置由细胞核外周逐渐向细胞核内部移动。伴随 $\beta$ -globin基因座定位改变的还有以RNA聚合酶II为核心的转录工厂。随着红细胞分化成熟,转录工厂数量逐渐减少,活跃的转录工厂与基因座一样逐渐向细胞核中间移动。因此,有人提出远端调控元件可以通过调节靶基因在细胞核内的定位或刺激靶基因向活跃的转录工厂靠近发挥转录调控功能。

与 $\beta$ -globin基因簇相似, $\alpha$ -globin基因簇由一系列编码血红蛋白 $\alpha$ 链的基因组成。同样,不同发育阶段的基因簇上基因的表达受到上游40~60 Kb处LCR的调控。 $\alpha$ -globin基因簇LCR上含有一系列HS位点,可以募集包括GATA-1、GATA-2以及EKLf等多种转录因子<sup>[34]</sup>。Vernimmen等<sup>[34]</sup>发现, $\alpha$ -globin LCR具有招募RNA聚合酶II及其他基础转录因子的功能,这些转录起始所必需的蛋白质因子通过染色质远距离相互作用由增强子转移到启动子,激活基因表达。因此,增强子与基因启动子相互作用促进转录机器的募集或稳定也被认为是远程调控的机制之一。

2010年,西班牙和美国的科研人员运用5C联合高精度建模平台(integrated modeling platform, IMP)技术分析了人16号染色体上一段长500 Kb、包含 $\alpha$ -globin基因簇及多个看家基因在内的片段ENm008,并模拟了其在细胞内部的3D结构模型<sup>[35]</sup>。研究结果显示,在淋巴样干细胞GM12878中,看家基因稳定表达而 $\alpha$ -globin基因不表达,此时染色质高度压缩形成一个类似球状的小体,称为“染色质小球(chromatin globule)”;而在表达 $\alpha$ -globin基因的慢性粒细胞性白血病细胞K562中,这一段染色质则比较松散,向外延伸,形成两个染色质小球。转录活跃的基因聚集在染色质小球的核心部位,不表达的基因则分布在小球的外周。研究人员推测,未知的远端调控元件与基因的相互作用可能是形成染色质小球的动力基础。在小球的中心位置可能是类似转录工

厂的蛋白质复合物,每个转录工厂转录不超过八个基因。因此,在转录活跃的K562细胞中就形成了两个染色质小球以满足基因表达的需求。该研究将人们从简单的远距离调控带入染色质三维空间结构的殿堂,使人们认识了基因的“形状”。

### 3.2 T<sub>H</sub>2细胞因子基因簇与T细胞活化

血红蛋白基因簇上的众基因在发育过程中受到位于同一条染色体上游增强子的调控从而顺序性表达的现象并不是个例,在T细胞的发育过程中也存在类似的情况。CD4<sup>+</sup> T细胞在不同细胞因子刺激下可以向T<sub>H</sub>1细胞或T<sub>H</sub>2细胞两个方向分化。T<sub>H</sub>1细胞主要分泌IL-2、TNF、IFN- $\gamma$ ,而T<sub>H</sub>2细胞主要分泌IL-4、IL-5、IL-6以及IL-13。其中,编码T<sub>H</sub>2细胞因子的基因*Il5*、*Il13*和*Il14*位于小鼠11号染色体上,连续排列,形成T<sub>H</sub>2细胞因子基因簇。尽管这三个基因前后相距达120 Kb,但是它们都受到LCR的调控。以小鼠为例,T<sub>H</sub>2细胞因子基因簇LCR长约25 Kb,包含了一系列HS位点。LCR位于*Il5*基因下游60 Kb处,距离下游的*Il13*基因15 Kb<sup>[36-37]</sup>。研究人员已证实,三个细胞因子基因与LCR在细胞中相互靠近,排列紧密。与 $\alpha/\beta$ -globin基因簇略有不同,研究者不仅检测到了三个细胞因子的启动子与LCR的相互作用,还检测到三个基因启动子之间的相互作用<sup>[38]</sup>。虽然在T<sub>H</sub>2细胞*Il4*、*Il5*、*Il13*表达活跃的状态下,三个启动子和LCR之间的相互作用显得更为持久,但是在非T细胞中,这些DNA元件之间依然可以被检测到有较弱的相互作用存在。这种处于准备状态的染色质空间结构,有利于基因的快速激活。同样,两种T细胞特异性表达的转录因子SATA6和GATA-3参与了这种相互作用的维持。2006年,一组研究人员发现,核基质附着蛋白SATB1也参与介导了T<sub>H</sub>2细胞因子基因簇染色质环的形成。当T<sub>H</sub>2细胞活化时,SATB1蛋白表达被迅速诱导,并与细胞因子基因座上的多个SATB1识别位点结合,将染色质折叠成数个环,形成一种转录激活状态下所特有的空间构象<sup>[39]</sup>。

有趣的是,在T<sub>H</sub>2细胞因子基因簇上不仅可以观察到同一条染色体上DNA的相互作用,还能检测到非同源染色体之间的相互作用。IFN- $\gamma$ 是一种由T<sub>H</sub>1细胞分泌的细胞因子,其基因*Ifng*位于小鼠10号染色体上。Spilianakis等<sup>[40]</sup>发现,在静息的T细胞中,10号染色体上的*Ifng*基因可以与11号染色体上T<sub>H</sub>2细胞因子基因簇相互靠近,转录静止。T细胞活化后,

这种染色体间的相互作用随即消失。Th2细胞中11号染色体上基因簇内部形成更为紧密的相互作用,使细胞因子转录被LCR激活。

#### 4 染色质环的亚核结构基础

亚核结构为远端染色质的相互作用提供结构基础。自二十世纪五十年代, Watson和Crick发现了DNA双螺旋结构, 开启了分子生物学的时代, 使遗传的研究深入到了分子水平。由直径2 nm的双链DNA分子到直径10 nm的核小体, 再到直径30 nm染色质纤维(螺线管), 到超螺线管, 直至细胞进入有丝分裂期形成染色体, 人们对遗传物质包装的认识飞速发展。

细胞分裂间期, 染色体去凝集, 呈细网状的染色质弥散在细胞核内。然而, 染色质在细胞核内的分布并不是随机的, 同一条染色体相对集中在一个区域, 称之为“染色体领土(chromosome territories)”。每条染色体的DNA被限制在相对固定的区域, 而染色质可在这个有限的小范围内快速运动。T. Cremer和C. Cremer按照一定的比例将染色体领域的模型粗略地画在一个取自活体HeLa细胞的细胞核的光学切片上, 就是目前被普遍接受的哺乳动物的核建构模型CT-IC模型(chromosome territory-interchromatin compartment model)<sup>[41]</sup>。该模型认为, 转录活跃的基因位于染色体领土的外围, 形成大的染色质环伸向间隔区或靠近核孔, 因为在间隔区或核孔附近可能包含了具有转录、复制、拼接、修复等功能的蛋白质因子, 并且便于基因的转录产物向细胞核外输送; 相反, 沉默的基因则位于染色体领土的内部。

结合染色质环学说和核建构CT-IC模型, Fraser等<sup>[42]</sup>认为, 间期细胞中30 nm染色质纤维周期性地附着在核骨架上, 在细胞核内部按照特定的模式分布, 其中转录活跃的染色质结构松散, 形成大的染色质环。因此, 位于环上的基因得以向染色体领土间的间隙伸展, 远距离染色质之间可以相互靠近, 甚至不同染色体之间的DNA元件也可以发生相互作用。

#### 5 蛋白质与染色质环的形成

在个体发育和不同生理或病理条件下, 细胞中远端调控元件与靶基因启动子的相互作用频率发生着动态变化, 从染色质高级结构层面调节和协调不

同基因的转录水平。是什么引起了远端染色质相互作用频率的动态变化呢? 促进形成染色质环的因素是什么? 有关染色质环的构建机制是基因远程调控研究领域关键问题之一。目前认为, 线性的染色质可通过结合在不同DNA区段上的蛋白之间的相互作用, 被弯曲折叠形成特殊的三维空间结构。近年来的研究显示, 核基质附着蛋白、转录因子和转录共激活因子、类固醇激素受体、CTCF以及Cohesin蛋白复合物均在染色质环高级结构的构建中发挥重要作用。另外也有观点认为, RNA聚合酶在促进远端调控元件与靶基因的相互作用方面发挥关键作用, 需要协同表达的基因聚集在一个以RNA聚合酶为核心的转录工厂上进行转录。以下我们简要阐述几个目前公认的参与构建染色质环的蛋白质。

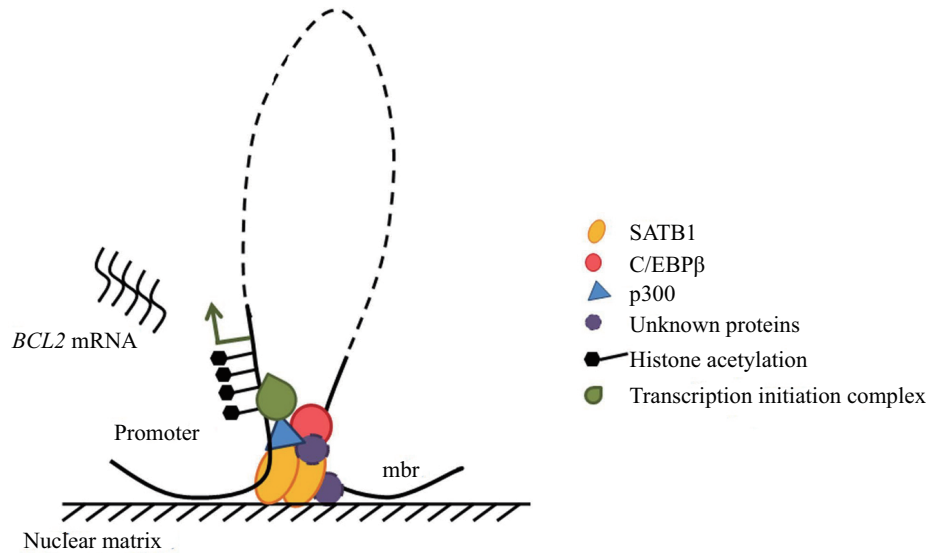
##### 5.1 核基质附着蛋白SATB1

核基质, 又称核骨架, 是细胞核内的一个由纤维蛋白构成的网架系统, 为DNA的复制和转录提供了平台。核基质附着蛋白可以识别基因组中核基质附着区序列, 将染色质锚定在核基质上, 在DNA复制和基因转录调控等重要细胞生命活动中发挥重要作用。

SATB1(special AT-rich binding protein 1), 是一种组织特异性表达的核基质附着蛋白, 通常在胸腺、脾脏、淋巴结组织内高表达。免疫荧光显示, 在胸腺T细胞内, SATB1蛋白主要分布在细胞核外周染色质区域呈“笼”状。同一条染色体上的基因可周期性锚着在SATB1蛋白上形成“花瓣”状。染色质环基部的SATB1蛋白可以招募组蛋白修饰酶、染色质重塑因子以及多种转录因子, 改变局部组蛋白和DNA的修饰水平, 调节核小体的相对位置, 使位于环基部的基因受到统一的调节<sup>[43]</sup>, 是一个重要的核基因组组织者。小鼠Th2细胞活化过程中, SATB1结合到11号染色体上200 Kb长的细胞因子基因簇, 与GATA3、STAT6一起将染色质折叠成环, 诱导细胞因子IL-4、IL-5、IL-13的产生<sup>[39]</sup>。在人T淋巴细胞白血病细胞中, SATB1参与PML小体的构成, 参与构建MHC I类分子基因簇高级结构, 抑制HLA-A、HLA-G、HCG4P6等基因的表达<sup>[44]</sup>。

本实验室前期研究发现, SATB1能够通与BCL2基因3'-UTR的mbr增强子元件及基因启动子上游的DNA序列相结合, 促使BCL2基因染色质形成环状结构, 由此使相距约200 Kb的启动子与mbr调控



图2 *BCL2*基因染色质环模型Fig.2 The model of chromatin loop of *BCL2* gene

元件在空间位置上充分接近,并在mbr上结合的另一转录因子C/EBP $\beta$ 的帮助下招募组蛋白乙酰化酶p300,促进*BCL2*基因启动子乙酰化,使启动子区的染色质形成更加开放的构象,增加转录因子CREB和RNA聚合酶II与启动子的结合,从而实现mbr增强子元件跨越200 Kb的转录激活功能<sup>[45]</sup>(图2)。SATB1介导的*BCL2*染色质的动态变化,是T细胞性白血病细胞调节*BCL2*基因应对外界信号刺激的重要环节。SATB1的作用是高度特异的,抑制caspase-6对SATB1蛋白的切割,可有效逆转早期凋亡中*BCL2*染色质环解聚,恢复*BCL2*的表达水平。

鉴于核基质附着蛋白在调节基因转录中的重要作用,进一步探讨SATB1以外的其他核基质附着蛋白,如LaminB、支架结合因子A(SAF-A)和SATB2等在染色质环高级结构介导的基因远程调控中的作用十分有意义。

## 5.2 转录因子

转录因子是一类可以识别特异性DNA序列并与其结合,进而招募一系列转录共激活因子帮助形成转录起始复合物促进基因转录的蛋白质。经典理论认为,真核细胞中转录因子识别启动子上游特征性序列是起始基因转录的最初步骤。近几年,多篇文献报道转录因子可能也是染色质环的缔造者之一,介导远端增强子的调控功能。

c-myc蛋白是myc家族成员之一,作为一种转录因子影响细胞增殖、周期以及凋亡等多项生命活动

过程。其在未分化细胞和肿瘤细胞中高表达,被认为是促进肿瘤发生发展的因素之一。2006年, Takahashi等<sup>[46]</sup>发现,将*Oct4*、*Sox2*、*Klf4*与*c-myc*四个编码转录因子的基因同时导入小鼠成纤维细胞后可以引起体细胞逆向分化成具有多项分化潜能的诱导多能干细胞(induced pluripotent stem cells, iPS)。这一发现使得c-myc一夜成名,备受关注。可见,在细胞分化过程中c-myc可以称得上是一个“掌控全局”的蛋白质。研究显示,全基因组中有多达15%的基因上有c-myc蛋白结合,此外还有众多基因间的序列上也结合了c-myc。在神经干细胞中敲除c-myc可导致明显的细胞核体积减小、形态异常以及广泛性的染色质凝集。此外, c-myc缺失还可以改变组蛋白H3/H4乙酰化和甲基化的修饰。可以推测, c-myc的功能不仅仅局限于调节转录,可能还参与构造染色质高级结构,维持常染色质的开放状态。

类固醇激素受体是一类重要转录因子,如雌激素受体 $\alpha$ (estrogen receptor  $\alpha$ , ER $\alpha$ )、糖皮质激素受体(glucocorticoid receptor, GCR)作为信号通路中最下游的环节,被上游的信号分子激活后与基因组中相应的结合元件相互识别,达到调控靶基因表达的作用。然而人们发现,一些受到激素调控的基因,其启动子区并未发现受体结合元件。例如, *BCL2*可以被雌激素激活,但是*BCL2*基因上的雌激素受体结合元件(estrogen response element, ERE)却位于启动子下游的外显子上。接受胞外雌二醇刺激后, ER被激活并

入核, 结合在ERE上, 促进远端的ERE元件与启动子相互作用, 激活*BCL2*基因转录<sup>[47-48]</sup>。随后, ER作为一个模型被用于研究染色质三维结构的动态变化。乳腺癌细胞MCF-7接受雌二醇刺激后, 利用ChIP-seq、ChIA-PET、RNA-seq等高通量研究技术发现, 被激活的ER广泛结合在基因组中的ERE上, 改变了染色质之间的相互作用模式, 并且这种远端染色质的相互作用改变了众多靶基因的转录<sup>[12]</sup>。可见, 类固醇受体在基因表达的远距离调控中发挥重要作用。

由于不同的转录因子有其相对特异的结合序列及共激活因子, 转录因子能否及如何影响远端调控元件与靶基因相互作用的特异性是个十分有趣和值得探究的方向。

### 5.3 CTCF和Cohesin

CTCF又称CCCTC结合因子, 属于具有锌指结构的转录因子。CTCF蛋白的表达极其普遍, 并且在基因组上其结合位点也非常丰富, 平均每隔15~25 Kb就有一个CTCF的结合位点。CTCF蛋白通常结合在基因间的绝缘子上, 将调控元件与靶基因以外的基因隔开, 防止增强子的过度作用。对小鼠 $\beta$ -globin、人MHC II类等几个基因簇研究显示, CTCF的绝缘功能是通过形成染色质环将活化的基因与沉默的基因隔离开。如小鼠红细胞中 $\beta$ -globin基因簇的上下游都分布了嗅觉感受器基因, CTCF通过形成染色质环将 $\beta$ -globin基因簇上的增强子LCR的功能局限在基因组小范围内, 以避免两侧的嗅觉感受器基因被LCR激活<sup>[49]</sup>。CTCF不仅介导同一条染色体上的基因簇成环, 还可以介导染色体之间的相互作用沉默印记基因的一条等位基因。譬如, CTCF介导人母源11号染色体上的基因*H19/Igf2*与父源11号染色体上的调控元件相互作用, 使母源染色体上*H19/Igf2*启动子高度甲基化, 沉默其表达<sup>[49-50]</sup>。可见, 蛋白质介导的染色质高级结构也是基因印迹的重要机制之一。

Cohesin是一个包含多个亚单位的蛋白质复合物, 其所包含的四个核心蛋白质形成一个“指环”状的结构, 将S期复制好的姐妹染色单体“绑”在一起, 直到有丝分裂结束。此外, Cohesin介导一些组织特异性的基因表达调控。在哺乳动物细胞中, Cohesin总是和CTCF共同定位在沉默子上发挥功能<sup>[51]</sup>。德朗热综合征(cornelia de lange syndrome, CLS)、罗伯茨综合征(roberts syndrome, RS)这两种先天发育异常综合征就是由于Cohesin复合物中蛋白组分的突

变导致的, 但是患者并未出现染色体的聚集或分离异常, 因此提示这两种综合征的病因可能与传统的Cohesin复合物的功能无关, 而是突变导致的Cohesin复合物不能稳定CTCF与基因组的结合, 使一些增强子的作用失去了控制, 超越边界, 导致众多基因转录异常。

## 6 eRNA与染色质高级结构及基因表达的远程调控

早在上世纪九十年代, 人们在研究 $\beta$ -globin基因簇上的增强子LCR时, 就首次描述了增强子RNA(enhancer RNA, eRNA)的存在<sup>[52]</sup>。近两年, 大规模的RNA-seq数据揭示, 在全基因组中相当一部分增强子上结合了RNA聚合酶II, 并转录一种长约500~2 000 nt、缺少poly A尾的非编码RNA, 称为enhancer RNA<sup>[53]</sup>。研究发现, 在某些刺激作用下, 增强子局部转录增加, 同时靶基因的转录水平也与之相呼应<sup>[54]</sup>。那么eRNA究竟是增强子发挥调控功能的介导者还是副产物? 为了探讨这一问题, 人们采用RNA干扰技术特异性地加速细胞中某一种eRNA的降解, 发现相应的靶基因转录也随之下降。荧光素酶报告系统得到的结果也与之相一致, 提示eRNA确实具有增强基因转录的功能。那么eRNA这种小分子又是如何调节基因转录的? 其中的机制又是什么? 经过实验证实, 远端增强子与靶基因相互作用形成染色质环需要大量的eRNA。核受体ER $\alpha$ 是一个众所周知的具有转录因子活性的蛋白质, 研究显示, ER $\alpha$ 可以招募复合物Cohesin介导远端ER $\alpha$ 的结合元件和靶基因启动子相互作用形成染色质环, 并且在某些ER $\alpha$ 的结合元件上有eRNA的转录<sup>[12]</sup>。RNA免疫沉淀实验发现, eRNA能招募形成染色质环所需的Cohesin蛋白质复合物, 敲减eRNA可以抑制Cohesin复合物的形成, 减弱增强子和启动子的相互作用, 从而抑制靶基因转录<sup>[55]</sup>。因此, eRNA被认为是维持远端染色质相互作用的“守护者”。对染色质高级结构的研究方兴未艾, 相信这其中还有更多的奥秘等待人们去揭示。

## 7 结语与展望

随着3C及其衍生技术的建立和应用, 基因远程调控的染色质环模型已得到广泛认可。从染色质高级结构层面解读基因转录的三维网络调控机制, 必



将为认识细胞分化、个体发育及人类疾病提供全新的视角。然而, 染色体远程调控模型的建立也提出了许多新问题, 如: 基因组中数量巨大的调控元件的靶基因是什么? 驱动染色质形成染色质环结构的因素是什么? 是什么因素决定了远端调控元件与靶基因相互作用的特异性? 此外, 调控元件往往具有组织特异性, 在不同类型的组织细胞中, 是什么因素决定了调控元件在不同组织中的活性状态? 在细胞分化和个体发育过程中, 是什么激活了调控元件从而决定了靶基因的开启或关闭? 在细胞有丝分裂过程中, 染色质环的结构如何稳定传递给子代细胞? 染色质高级结构与表观记忆效应之间的关系如何? 尽管在复杂的生命现象面前, 现有的研究都犹如盲人摸象, 但摸象的人多了, 分析思考的人多了, 便可最大限度接近大象的本来面目。生命的复杂与困惑, 也正是生命科学的迷人之处。

### 参考文献 (References)

- Forrester WC, Takegawa S, Papayannopoulou T, Stamatoyannopoulos G, Groudine M. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res* 1987; 15(24): 10159-77.
- Kleinjan DA, van Heyningen V. Long-range control of gene expression: Emerging mechanisms and disruption in disease. *Am J Hum Genet* 2005; 76(1): 8-32.
- Spitz F, Gonzalez F, Duboule D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 2003; 113(3): 405-17.
- West AG, Fraser P. Remote control of gene transcription. *Hum Mol Genet* 2005; 14 Spec No 1: R101-11.
- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002; 295(5558): 1306-11.
- Dekker J. The three 'C' s of chromosome conformation capture: Controls, controls, controls. *Nat Methods* 2006; 3(1): 17-21.
- Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2007; 2(7): 1722-33.
- Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 2006; 38(11): 1341-7.
- Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, *et al.* Chromosome conformation capture carbon copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res* 2006; 16(10): 1299-309.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; 326(5950): 289-93.
- Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* 2005; 37(1): 31-40.
- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 2009; 462(7269): 58-64.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; 409(6822): 860-921.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, *et al.* The sequence of the human genome. *Science* 2001; 291(5507): 1304-51.
- Baltimore D. Our genome unveiled. *Nature* 2001; 409(6822): 814-6.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007; 39(3): 311-8.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 2009; 457(7231): 854-8.
- Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* 2007; 3(8): e136.
- Zhang ZD, Paccanaro A, Fu Y, Weissman S, Weng Z, Chang J, *et al.* Statistical analysis of the genomic distribution and correlation of regulatory elements in the ENCODE regions. *Genome Res* 2007; 17(6): 787-97.
- Ferraiuolo MA, Sanyal A, Naumova N, Dekker J, Dostie J. From cells to chromatin: Capturing snapshots of genome organization with 5C technology. *Methods* 2012; 58(3): 255-67.
- Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, *et al.* Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 2013; 155(7): 1507-20.
- Engel JD, Tanimoto K. Looping, linking, and chromatin activity: New insights into beta-globin locus regulation. *Cell* 2000; 100(5): 499-502.
- Higgs DR, Garrick D, Anguita E, De Gobbi M, Hughes J, Muers M, *et al.* Understanding alpha-globin gene regulation: Aiming to improve the management of thalassemia. *Ann N Y Acad Sci* 2005; 1054: 92-102.
- Stamatoyannopoulos JA, Goodwin A, Joyce T, Lowrey CH. NF-E2 and GATA binding motifs are required for the formation of DNase I hypersensitive site 4 of the human beta-globin locus control region. *EMBO J* 1995; 14(1): 106-16.
- Tewari R, Gillemans N, Wijgerde M, Nuez B, von Lindern M, Grosveld F, *et al.* Erythroid Kruppel-like factor (EKLF) is active in primitive and definitive erythroid cells and is required for the function of 5'HS3 of the beta-globin locus control region. *EMBO J* 1998; 17(8): 2334-41.
- 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): 2924-9.
- Sawado T, Halow J, Bender MA and Groudine M. The beta

- globin locus control region (LCR) functions primarily by enhancing the transition from transcription initiation to elongation. *Genes Dev* 2003; 17(8): 1009-18.
- 28 Wang L, Di LJ, Lü X, Zheng W, Xue Z, Guo ZC, *et al.* Inter-MAR association contributes to transcriptionally active looping events in human beta-globin gene cluster. *PLoS One* 2009; 4(2): e4629.
- 29 Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipson S, *et al.* The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* 2004; 18(20): 2485-90.
- 30 Hong W, Nakazawa M, Chen YY, Kori R, Vakoc CR, Rakowski C, *et al.* FOG-1 recruits the NuRD repressor complex to mediate transcriptional repression by GATA-1. *EMBO J* 2005; 24(13): 2367-78.
- 31 Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, *et al.* Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 2005; 17(3): 453-62.
- 32 Ragozy T, Bender MA, Telling A, Byron R, Groudine M. The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes Dev* 2006; 20(11): 1447-57.
- 33 Ragozy T, Telling A, Sawado T, Groudine M, Kosak ST. A genetic analysis of chromosome territory looping: Diverse roles for distal regulatory elements. *Chromosome Res* 2003; 11(5): 513-25.
- 34 Vernimmen D, de Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* 2007; 26(8): 2041-51.
- 35 Bau D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, *et al.* The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol* 2011; 18(1): 107-14.
- 36 Lee GR, Fields PE, Griffin TJ, Flavell RA. Regulation of the Th2 cytokine locus by a locus control region. *Immunity* 2003; 19(1): 145-53.
- 37 Lee GR, Kim ST, Spilianakis CG, Fields PE, Flavell RA. T helper cell differentiation: Regulation by cis elements and epigenetics. *Immunity* 2006; 24(4): 369-79.
- 38 Spilianakis CG, Flavell RA. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 2004; 5(10): 1017-27.
- 39 Cai S, Lee CC, Kohwi-Shigematsu T. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* 2006; 38(11): 1278-88.
- 40 Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Inter-chromosomal associations between alternatively expressed loci. *Nature* 2005; 435(7042): 637-45.
- 41 Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2001; 2(4): 292-301.
- 42 Fraser P, Bickmore W. Nuclear organization of the genome and the potential for gene regulation. *Nature* 2007; 447(7143): 413-7.
- 43 Cai S, Han HJ, Kohwi-Shigematsu T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* 2003; 34(1): 42-51.
- 44 Kumar PP, Bischof O, Purbey PK, Notani D, Urlaub H, Dejean A, *et al.* Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol* 2007; 9(1): 45-56.
- 45 Gong F, Sun L, Wang Z, Shi J, Li W, Wang S, *et al.* The BCL2 gene is regulated by a special AT-rich sequence binding protein 1-mediated long range chromosomal interaction between the promoter and the distal element located within the 3'-UTR. *Nucleic Acids Res* 2011; 39(11): 4640-52.
- 46 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; 126(4): 663-76.
- 47 Perillo B, Ombra MN, Bertoni A, Cuzzo C, Sacchetti S, Sasso A, *et al.* DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science* 2008; 319(5860): 202-6.
- 48 Su CH, Tzeng TY, Cheng C, Hsu MT. An H2A histone isotype regulates estrogen receptor target genes by mediating enhancer-promoter-3'-UTR interactions in breast cancer cells. *Nucleic Acids Res* 2014; 42(5): 3073-88.
- 49 Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 2006; 20(17): 2349-54.
- 50 Kurukuti S, Tiwari VK, Tavooosidana G, Pugacheva E, Murrell A, Zhao Z, *et al.* CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci USA* 2006; 103(28): 10684-9.
- 51 Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 2008; 132(3): 422-33.
- 52 Tuan D, Kong S, Hu K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci USA* 1992; 89(23): 11219-23.
- 53 Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010; 465(7295): 182-7.
- 54 Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 2014; 39(4): 170-82.
- 55 Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 2013; 498(7455): 516-20.