

原钙粘蛋白基因簇调控区域中成簇的 CTCF 结合位点分析

翟亚男, 许泉, 郭亚, 吴强

上海交通大学系统生物医学研究院比较生物医学研究中心, 系统生物医学教育部重点实验室, 上海 200240

摘要: 哺乳动物中原钙粘蛋白(Protocadherin, *Pcdh*)基因簇包含 50 多个串联排列的基因, 这些基因形成 3 个紧密相连的基因簇(*Pcdh α* 、*Pcdh β* 和 *Pcdh γ*), 所编码的原钙粘蛋白质群在神经元多样性(Neural diversity)和单细胞特异性(Single cell identity)以及神经突触信号转导中发挥重要作用。前期的工作已证实转录因子 CTCF(CCCTC-binding factor)与 CTCF 结合位点(CTCF-binding site, CBS)的方向性结合能够决定增强子和启动子环化的方向以及其远距离交互作用的特异性, 并进一步在 *Pcdh* 基因座(Locus)形成两个(*Pcdh α* 和 *Pcdh γ*)染色质拓扑结构域(CTCF/cohesin-mediated chromatin domain, CCD), 而且染色质拓扑结构域对于控制基因表达调控至关重要。本文通过生物信息学方法对比人类和小鼠序列, 发现 *Pcdh β* 染色质拓扑结构域调控区域中的 DNase I 超敏位点(DNase I hypersensitive sites, HSs)较为保守。染色质免疫沉淀及大规模测序实验(Chromatin immunoprecipitation and massive parallel sequencing, ChIP-Seq)揭示 CBS 位点在 *Pcdh β* 调控区域中成簇分布并且具有相同的方向。凝胶电泳迁移实验(Electrophoresis mobility shift assay, EMSA)确定 *Pcdh β* 调控区域内具体的 42 bp CBS 位点并且发现一个 CTCF 峰包含两个 CBS 位点。在全基因组范围内, 运用计算生物学方法分析 CTCF 和增强子、启动子等调控元件的关系, 发现 CBS 位点在调控元件附近有较多分布, 推测 CTCF 通过介导增强子和启动子的特异性交互作用, 在细胞核三维基因组内形成活性转录枢纽调控基因精准表达。

关键词: 原钙粘蛋白基因簇; 转录因子 CTCF; CTCF 结合位点; 染色质拓扑结构域 CCD; 基因表达调控

Characterization of a cluster of CTCF-binding sites in a protocadherin regulatory region

Yanan Zhai, Quan Xu, Ya Guo, Qiang Wu

Key laboratory of Systems Biomedicine (Ministry of Education), Center for Comparative Biomedicine, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract: The mammalian clustered protocadherin (*Pcdh*) locus contains more than 50 highly-similar genes arrayed in tandem. These *Pcdh* genes are organized into three closely-linked clusters (*Pcdh α* , *Pcdh β* , and *Pcdh γ*). The encoded PCDH proteins play critical roles in neuronal diversity, single cell identity, and synaptic connectivity. Recent

收稿日期: 2016-01-21; 修回日期: 2016-03-01

基金项目: 国家自然科学基金项目(编号: 91519302)资助[Supported by the National Natural Science Foundation of China (No. 91519302)]

作者简介: 翟亚男, 硕士研究生, 专业方向: 遗传学。E-mail: zhaichn@gmail.com

通讯作者: 吴强, 博士, 教授, 博士生导师, 研究方向: 基因表达调控及神经发育。E-mail: qwu123@gmail.com

DOI: 10.16288/j.ycz.16-037

网络出版时间: 2016/3/2 15:05:03

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20160302.1505.002.html>

studies revealed that directional CTCF (CCCTC-binding factor) binding to CBS (CTCF-binding site) determines the specific interaction between enhancers and promoters, and the three *Pcdh* clusters form two CCDs (CTCF/cohesin-mediated chromatin domain) which is important for gene regulation. Here, we characterized a regulatory region, which contains several HSs (DNase I hypersensitive sites), downstream of the *Pcdhβγ* clusters. Specifically, CTCF ChIP-Seq (Chromatin immunoprecipitation-sequencing) experiments revealed several peaks in the *Pcdhβγ* regulatory regions. In addition, we performed comprehensive EMSA (Electrophoresis mobility shift assay) experiments and identified exact 42-bp CBS sites for each CTCF peak. Interestingly, we found that single CTCF peaks can contain two CBS sites. Finally, genome-wide computational analyses revealed many CBS sites close to chromatin marks of regulatory elements, such as enhancers and promoters. We propose that CTCF may mediate specific interactions between enhancers and promoters to form active transcription hub for gene expression.

Keywords: protocadherin; transcription factor CTCF; CTCF-binding site; CTCF/cohesin-mediated chromatin domain; regulation of gene expression

人类大脑负责学习、记忆和认知等重要功能, 包含由上千万不同神经元形成的几十万亿特异性突触连接的复杂网络, 原钙粘蛋白群在这复杂的神经系统网络中发挥重要的作用^[1]。近年研究发现, 原钙粘蛋白群在神经发育和树突生成中至关重要^[2-6]。哺乳动物原钙粘蛋白基因(Protocadherin, *Pcdh*)由 50 多个基因构成的 3 个紧密相连的基因簇(*Pcdhα*、*Pcdhβ* 和 *Pcdhγ*)组成^[7-11]。例如, 人类 *Pcdh* 基因簇包含 53 个基因, *Pcdhα* 基因簇和 *Pcdhγ* 基因簇由多个相似的随机表达的可变区外显子、组成型表达的 C 类型的可变区外显子以及其下游的恒定区外显子构成, 而 *Pcdhβ* 基因簇仅由多个相似的可变区外显子构成^[8,10] (图 1A)。每个可变区外显子有其自己的启动子, 这些启动子在不同类型神经元中选择性激活, 将可变区外显子转录的 mRNA 顺式剪切到恒定区外显子上^[12,13]。每类神经元随机选择表达几种原钙粘蛋白群成员^[14,15], 这些蛋白可以相互作用形成许多复合物^[16,17], 构成了神经元细胞表面蛋白的多样性^[17]。*Pcdhα* 基因簇下游有 2 个 DNase I 超敏位点(DNase I hypersensitive sites, HSs): HS7 和 HS5-1, 这 2 个 HS 超敏位点负责调控 *Pcdhα* 基因簇的表达^[18]; *Pcdhβγ* 基因簇有一个共同的调控区域位于 *Pcdhγ* 下游, 包含 HS7L、HS5-1aL^[19]、HS5-1bL、HS18 和 HS19-20^[15,20] 等位点 这些 HS 超敏位点有很多转录因子结合序列, CTCF (CCCTC-binding factor) 是其中比较重要的因子。

CTCF 是一种在进化过程中高度保守的锌指蛋白转录因子, 在染色质架构中起到关键作用^[21]。CTCF 蛋白是 1990 年在研究鸡 Myc 5' 序列结合蛋白

时发现的, 这个蛋白结合的核心序列是“CCCTC”, 后来便命名为 CCCTC-binding factor^[22]。β-珠蛋白基因的增强子和启动子之间有一段包含 CTCF 蛋白结合位点的序列可以阻断增强子对启动子的激活作用, 因此认为 CTCF 结合位点是绝缘子^[23]。CTCF 的结合序列长度约为 42 bp, 可以分为 4 个模块(Module): Module1、Module2、Module3 和 Module4^[24,25]。CTCF 结合位点(CTCF-binding site, CBS)在全基因组范围内分布广泛^[26], 并且 CTCF 可以通过招募 Cohesin 来介导 DNA 调控元件之间的相互作用^[27,28]。最新研究发现, CBS 位点具有方向性, 这种方向性使得基因组染色质形成一个个 CTCF/Cohesin 介导的拓扑结构域 CCD(CTCF/cohesin-mediated Chromatin Domain), 例如 *Pcdh* 基因簇包含两个 CCD 拓扑结构域: *Pcdhα* 基因簇形成一个 CCD, *Pcdhβγ* 基因簇形成另外一个 CCD^[24] (图 1A)。

本文利用 CTCF 和 Rad21 特异性抗体通过染色质免疫沉淀结合大规模平行测序(Chromatin immunoprecipitation and massive parallel sequencing, ChIP-Seq)实验发现 *Pcdhβγ* 调控区域中包含一系列 CBS 位点, 进一步通过凝胶电泳迁移实验(Electrophoretic mobility shift assay, EMSA)发现了 CTCF 蛋白具体结合的 42 bp 序列, 证明这些 CBS 位点的方向都是反向的。运用计算生物学方法分析了全基因组范围内启动子标志和增强子标志与基因组范围内 CBS 位点的关系。本文确定了 *Pcdhβγ* 下游调控区域的所有 CBS 位点, 并发现了 CTCF 蛋白在基因组范围内与启动子-增强子相互作用的关系。因

此推测在 CTCF 蛋白作用下增强子和启动子形成活性转录枢纽, 调控基因的精准表达。

1 材料和方法

1.1 实验材料

人类神经母细胞瘤细胞系 SK-N-SH 购自上海细胞库; 体外转录表达试剂盒购自美国 Promega 公司; 凝胶电泳迁移实验试剂盒购自美国 Thermo 公司; Quant-iT PicoGreen dsDNA Reagent 购自美国 Invitrogen 公司; ChIP 试剂盒(17-371)、CTCF 抗体(07-729)、RNAPII 抗体(05-623B)、H3K4me3 抗体(17-614)和 H3K27me3 抗体(17-622)购自美国 Millipore 公司; Rad21 抗体(ab922)购自英国 Abcam 公司; CTCF 单抗(612149)购自美国 BD 公司; 羊抗兔二抗购自上海麦约尔生物技术有限公司; Protein G Agarose 购自美国 Sigma 公司。

1.2 方法

1.2.1 序列比对分析

人类和小鼠 *Pcdhβγ* 调控区域的序列比对来自 <http://genome.ucsc.edu/>。利用 Clustal W 软件进行人类 *Pcdh* 基因簇中的 CBS 位点序列分析, 网址为 <http://www.genome.jp/tools/clustalw/>。人类和小鼠 *Pcdhγ* 调控区域中的 CBS 位点比对通过 DNAMAN(Version 7.0)软件。

1.2.2 染色质免疫沉淀及高通量测序

细胞固定与裂解: 培养人类神经母细胞瘤细胞系 SK-N-SH 至约 80%~90% 融合, 添加终浓度 1% 甲醛固定 10 min; 加入终浓度 125 mmol/L 的甘氨酸, 室温孵育 5 min, 终止交联; 用预冷的 PBS(含终浓度为 1 mmol/L 的 PMSF)洗涤 2 次; 加入 2 mL 预冷的 PBS(含 1× Protease Inhibitor Cocktail), 使用细胞刮收集细胞, 1000×g 在 4℃ 离心 5 min, 弃上清; 加入 SDS 裂解缓冲液(含 1× Protease Inhibitor Cocktail), 冻存于 -70℃ 备用。超声: 超声 10 s, 暂停 35 s, 振幅为 20, 共 20 个循环; 14000×g 在 4℃ 离心 10 min, 保留上清。

染色质免疫共沉淀: 将超声后的蛋白-DNA 复合物用稀释缓冲液(含 1× Protease Inhibitor Cocktail)稀释 10 倍; 加入 60 μL Protein G Agarose, 4℃ 旋转

孵育 1 h; 4000×g 在 4℃ 离心 1 min, 取上清, 取其中 10 μL 作为 Input, 4℃ 保存; 加入 2 μL CTCF、Rad21、RNAPII、H3K4me3 或 H3K27me3 抗体, 4℃ 旋转孵育过夜; 加入 60 μL Protein G Agarose, 4℃ 旋转孵育 1 h; 4000×g 在 4℃ 离心 1 min, 弃上清; 分别用 Low Salt Immune Complex Wash Buffer、High Salt Immune Complex Wash Buffer 和 LiCl Immune Complex Wash Buffer 洗涤 1 次, 每次 3~5 min; TE Buffer 洗涤 2 次, 每次 3~5 min; Input 对照组中加入 200 μL Elution Buffer 备用; 上述洗涤后的 Beads 中加入 100 μL Elution Buffer, 混匀后室温孵育 15 min; 4000×g 在室温离心 1 min, 收集上清, 重复一次, 共收集 200 μL; 在 Input 对照组和实验组中加入终浓度 200 mmol/L 的 CBS, 65℃ 孵育 4~5 h 进行交联; 分别加入 1 μL RNase A, 在 37℃ 孵育 30 min; 加入 0.5 mol/L EDTA 4 μL、1 mol/L Tris-HCl 8 μL 和 1 μL 蛋白酶 K, 在 45℃ 孵育 1~2 h; 加入 1 mL Bind Reagent A, 混匀后转移至 Spin Filter, 12000×g 在室温离心 30 s, 弃滤液; 加入 500 μL Wash Reagent B, 12000×g 在室温离心 30 s, 弃滤液, 12000×g 在室温离心 30 s; 将 Spin Filter 移至新的收集管中, 洗脱 DNA, -20℃ 存放备用。

二代测序与数据分析: 通过 Quant-iT PicoGreen dsDNA Reagent 检测 ChIP 沉淀后的 DNA 浓度, 采用 ChIP-qPCR 检测富集 DNA 的特异性; 取 10 ng DNA 按照 Illumina 公司 HiSeq 2000 平台文库制备方法制备文库, 进行 50 bp 单末端测序; 与 hg19 参考基因组序列比对后, 通过 MACS (Model-based analysis of ChIP-Seq) 软件^[29]分析富集信号。ChIP-Seq 数据已经提交至 NCBI 数据库, GEO (Gene Expression Omnibus) accession number 为 GEO:GSE76815, 可进入 NCBI 网站查看, 网址为: <http://www.ncbi.nlm.nih.gov/geo/>。

1.2.3 凝胶电泳迁移实验中的质粒构建

通过 PCR 方法扩增小鼠和人类基因组中 *Pcdhβγ* 调控区域中 CBS 位点序列, 引物序列见表 1。以克隆好的质粒为模板, 使用携带突变位点的引物进行 PCR, 构建 CBS 位点突变质粒, 所用引物见表 2。扩增小鼠突变位点时, 先用 FM 引物和 T 载体上的 SP6 引物扩增出小片段 DNA, 再以小片段 DNA 作

表 1 验证 CTCF 在 *Pcdhβ9* 调控区域的结合分布所用引物Table 1 Primers for identifying the distribution of CTCF in *Pcdhβ9* regulatory region

引物名称	序列(5'→3')	引物名称	序列(5'→3')
ma-F	CTTCTCAGAAGGCCACTGC	ha-FB	*CTTCCGTGAAGGCCACTG
ma-R	AAGGAGGCACTGGGGGGCTTTAACC	hb-F	GCCAGATCCAAGGCTGGG
ma-FB	*CTTCTCAGAAGGCCACTGC	hb-R	TGGAATGTAGGTCAGCTGCAC
mb-F	TACTAGCTGAAGACTAGGGTT	hb-FB	*GCCAGATCCAAGGCTGGG
mb-R	TGGAATGTAGGTCAGCTGCACT	hψa-F	GCTGAGCATCACCAGTGCAG
mb-FB	*TACTAGCTGAAGACTAGGGTT	hψa-R	ATGGACCTTGCCAGTTTC
mc-F	ACCCAGGCA CTTAGACTAG	hψa-FB	*GCTGAGCATCACCAGTGCAG
mc-R	CCAAATGACATTTGGGATAAGG	hc-F	ACCCGGGTACTGGAAACTGG
mc-FB	*ACCCAGGCA CTTAGACTAG	hc-R	TCCCGCTGACATGTGGGATG
md-F	GGGAGTGCCAGGCAGTGAC	hc-FB	*ACCCGGGTACTGGAAACTGG
md-R	ACAGAAAGGAAAAAACCCGGAGT	hd-F	GGGAGTGCCAGACAGTAAAGC
md-FB	*GGGAGTGCCAGGCAGTGAC	hd-R	TAGTGAAGAACATAAACAGACTCC
me-F	AATCTGGAACCTTGAGCCAAAATGG	hd-FB	*GGGAGTGCCAGACAGTAAAGC
me-R	TACTTTTCTGATCTGCTTGCACT	he-F	AATCTGGAACCTGAGCCAA
me-FB	*AATCTGGAACCTTGAGCCAAAATGG	he-R	GTTTTTCTATTCTGCTTGC
mf-F	ATTTTACTGAACAGCTCGC	he-FB	*AATCTGGAACCTGAGCCAA
mf-R	GCCGCTGTTGTGAGATGCAG	hf-F	AACTGCTGATCAGTCTCCAAAT
mf-FB	*ATTTTACTGAACAGCTCGC	hf-R	CTGTCCTTTCCAAATACAAGCT
mg-F	ACAATGACACATCAAAAATAAGAG	hf-FB	*AACTGCTGATCAGTCTCCAAAT
mg-R	GAATCTCTAAGCAACAGTGCTCCTC	hg-F	ACAATGACACATCTAAAATAAAAAGG
mg-FB	*ACAATGACACATCAAAAATAAGAG	hg-R	TTTTATTCTAGCAGTGTGCTTTAC
mh-F	TGTATAAACACTAATAAAGATACAAGTGG	hg-FB	*ACAATGACACATCTAAAATAAAAAGG
mh-R	GAGGGTATTGACATTTCCCTGTG	hh-F	AAAAATTCATAATAAAGATGCAAC
mh-FB	*TGTATAAACACTAATAAAGATACAAGTGG	hh-R	GTCTTTTCTTTTCCCTGTG
mψb-F	ATACCCTCA ACTTTTGATGGG	hh-FB	*AAAAATTCATAATAAAGATGCAAC
mψb-R	CTGGCTTTGTGTTAGCTTTAGGC	hψb-F	AAGAAAAAGACCCTCAGATTC
mψb-FB	*ATACCCTCA ACTTTTGATGGG	hψb-R	CTGCCCTGACATAATTTCTTG
ha-F	CTTCCGTGAAGGCCACTG	hψb-FB	*AAGAAAAAGACCCTCAGATTC
ha-R	AAGGAGGCACTGGGGAAC		

注：*代表生物素(Biotin)标记。

为引物和 F 引物扩增出完整的带有突变位点的 CBS 位点序列。

PCR 扩增条件：94℃ 3 min；94℃ 30 s，60℃ 30 s，72℃ 30 s，25 个循环；72℃ 3 min。将 PCR 产物在 2% 琼脂糖凝胶上进行电泳，对照 DNA marker 将片段大小符合预期的切下。利用 DNA 凝胶回收离心法将目的片段进行回收并连接到 pGEM-T-Easy-Vector 上。所有克隆质粒均经 Sanger 测序确定其正确性。

1.2.4 凝胶电泳迁移实验

使用生物素(Biotin)标记引物和高保真 DNA 聚

合酶 KOD, 通过 PCR 从构建好的质粒上扩增探针。将探针在 2% 琼脂糖凝胶上进行电泳，切胶回收并测浓度。用 MilliQ 水将探针稀释至 20 fmol/μL。6% 聚丙烯酰胺凝胶(15 mL)包括：1.5 mL 5×TBE 缓冲液、9.63 mL MilliQ 水、3 mL 30% Arc-Bis、750 μL 50% 甘油、112.5 μL 10% 过硫酸铵和 9.38 μL 四甲基乙二胺 TEMED。将制好的凝胶 80V 预电泳 20 min。在预电泳过程中进行 EMSA 结合反应(10 μL)，依次向 PCR 管中加入以下试剂：3 μL MilliQ 水、1 μL 10×Binding Buffer、0.5 μL 50% 甘油、0.5 μL 100 mmol/L MgCl₂、0.5 μL 1 μg/μL poly(dI.dC)、1 μL 1% NP-40、0.5 μL

表 2 突变 *Pcdhβy* 调控区域中 CBS 位点所用引物

Table 2 Primers for mutating the sequences of CBSs in *Pcdhβy* regulatory region

引物名称	序列(5'→3')
ma-FM	CCAGGCCGAACAATTGTGGCCAG
mb-FM	CTGCCAGCGAAAGATGCTGTCTG
mc-FM	GTCCACTACGAAAAATGCCGGCTG
md-RM	GTTAGACACCATTTTTTGCCCTTTCTC
me-FM	CAACTGCGCAACAACCGTGGC
mf-FM	GGTCGGCATTTTTCGGTGTCGTAC
mg-FM	GCTCCATGTGAACGATGGTGGCG
mh-FM	CTGGTGAGTGAACAATAGTGGGCAG
SP6	ATTTAGGTGACACTATAG
ha-FM	CTTCCGTGAAGGCCACTGCCAGGTCCCCAGTGCGAAAAATAGTGCCAT
hb-FM	GCCAGATCCAAGGCTGGGGCTCACTGCTCCAGCGAACAATGCTGGCTGAGCATC
hc-FM	ACCCGGGTACTGGAACTGGGCAAGGTCCATTACGAACAATGTTGACCGTGTGC
hd-RM	GGGAGTGCCAGACAGTAAAGCGAAGGGAAAGGGCAAAAAATGGCGGCCAAGATAG
he-FM	AATCTGGAACCTGAGCCAAAATAGCTCAACTGCGCTACAACCGTGGCCGAAAAG
hf-FM	CTGACAGGTATCTTTCAGTATACTAAT
hg-FM	ACAATGACACATCTAAAATAAAAGGGATCCATGTGAACGATGGTGGCTGAAAGC
hh-FM	AAAAATTCACTAATAAAGATGCAACTGGGTGAGTGAACAATAGTGGACAGAAGAG

2 mmol/L ZnSO₄, 吹打混匀。加入 1 μL 体外转录表达蛋白, 室温孵育 20 min。再加入 2 μL 20 fmol/μL 的探针, 室温孵育 20 min。每个样品中加入 2.5 μL 5× Loading Buffer, 吹打混匀后, 吸取 10 μL 上样。在冰浴中, 80 V 电泳 100 min。在冰浴中将蛋白质和探针复合物转至尼龙膜上, 380 mA 电转 40 min。转膜完成后, 紫外交联 10 min。将尼龙膜放入干净的小盒子中, 加入 10 mL Blocking Buffer, 摇床上孵育 15 min。加入 33.3 μL 亲和素, 混匀后, 摇床上孵育 15 min。加入 1×Wash Buffer, 摇床上孵育 5 min, 重复 4 次。加入 10 mL Substrate Equilibration Buffer, 摇床上孵育 5 min。将尼龙膜取出, 与配好的底物反应液室温避光反应 5 min。反应完成后, 用保鲜膜将尼龙膜裹好, 使用 ChemiDoc XRS+成像系统检测结果。

1.2.5 体外转录翻译实验

体外转录翻译反应体系(25 μL), 向无核酸酶污染的 PCR 管中依次加入以下试剂: 20 μL TNT Quick Master Mix、0.5 μL 甲硫氨酸、500 ng DNA(含表达序列的质粒), 用去除核酸酶的水补至 25 μL, 轻轻吹打混匀。30℃孵育 90 min。反应结束后, 立即分

装为 2.2 μL/管。分装完成后, 立即液氮速冻, 并放置 -70℃ 保存。

1.2.6 蛋白免疫印记杂交实验(Western blot)

15%分离胶(5 mL)包括: 1.15 mL MilliQ 水、2.5 mL 30%Arc-Bis、1.25 mL 1.5 mol/L Tris-HCl(pH = 8.8)、50 μL 10% SDS、50 μL 10%过硫酸铵、2 μL 四甲基乙二胺 TEMED。4%压缩胶(3 mL)包括: 1.79 mL MilliQ 水、400 μL 30%Arc-Bis、750 μL 0.5 mol/L Tris-HCl(pH = 6.8)、30 μL 10% SDS、30 μL 10%过硫酸铵、3 μL 四甲基乙二胺 TEMED。各样品中加入 10 μL 1×SDS-PAGE 上样缓冲液(用 RIPA 裂解液将 SDS-PAGE 上样缓冲液稀释至 1×, 稀释后加入 1% 体积裂解液的 PMSF), 98℃孵育 5 min 使蛋白变性。80 V 电泳 20 min, 110 V 电泳 120 min。冰浴中将蛋白质转至 PVDF 膜上, 380 mA 电转 25 min。转膜完毕后, 立即将膜放置 MilliQ 水中, 漂洗 1~2 min。加 10 mL 封闭液, 室温温和摇晃 1 h。加入 CTCF 抗体作为一抗, 室温孵育 3 h 或 4℃孵育 10 h。加 1×TBST, 室温温和摇晃 15 min, 重复 3 次。加入羊抗兔二抗, 室温孵育 2 h。加入 1×TBST, 室温温和

摇晃 15 min, 重复 3 次。用 Odyssey 双色红外激光成像系统扫描膜成像。

1.2.7 CTCF 结合基序分析

利用人类和小鼠在 *Pcdh β* 调控区域的 CBS 位点序列信息, 通过 Weblogo 软件^[30]得出 CTCF 结合基序, 网址为 <http://weblogo.threeplosone.com/>。

1.2.8 生物信息学分析

从 ENCODE 数据库^[31]中下载人类慢性骨髓性白血病细胞系 K562 的 CTCF、增强子标志^[32] H3K4Me1 和 H3K27ac 以及启动子标志^[32] H3K4Me3 和 RNAP II 的 ChIP-Seq 数据, 运用 Python 语言分析蛋白 CTCF 和这些调控元件标志的关系。

2 结果与分析

2.1 人类 *Pcdh* 基因簇中 CBS 位点分布

染色质架构蛋白 CTCF 在人类 *Pcdh* 基因簇中有着广泛的分布。在人类 *Pcdh α* 基因簇中, *Pcdh α 1* 和 *Pcdh α 1* 基因的启动子区域包含 1 个 CBS 位点, *Pcdh α 2~ α 13* 基因的启动子区域包含 2 个 CBS 位点, *Pcdh α 2* 基因的启动子区域不包含 CBS 位点, 并且所有的 *Pcdh α* 可变区中的 CBS 位点都是正向的^[15,24] (图 1, B 和 C)。除了 *Pcdh β 1*、*Pcdh γ 4* 和 *Pcdh γ 5* 外的每一个 *Pcdh β* 基因簇成员的启动子区域也都包含有 CBS 位点, 这些位点也都是正向的^[15,24] (图 1, D 和 E)。

每一个 CBS 位点包含 4 个序列模块 (Sequence module), 本文把具有模块 #1~#4 (Module #1~#4) 的 CBS 位点定义为正向^[15,24]。在人类神经母细胞瘤 (Neuroblastoma) SK-N-SH 细胞系中, CTCF 介导增强子 HS5-1 和 *Pcdh α 4*、 α 8、 α 12 和 α 1 启动子的相互作用, 并且突变启动子处的 CBS 位点导致启动子活性的降低^[15]。在人类和小鼠增强子 HS5-1 位点发现有 HS5-1a 和 HS5-1b 两个 CBS 位点, 这 2 个 CBS 位点都具有序列模块 #4~#1 (Module #4~#1), 因此其方向是反向的 (图 1F)。在人类子宫内膜癌细胞系 HEC-1B 中, 将包含 2 个 CBS 位点的增强子 HS5-1 进行反转, 会减少增强子 HS5-1 与上游 *Pcdh α* 启动子的环化, 并增加增强子 HS5-1 与下游 *Pcdh β* 启动子的环化, 因此改变了 *Pcdh* 基因簇的 2 个 CCD 拓

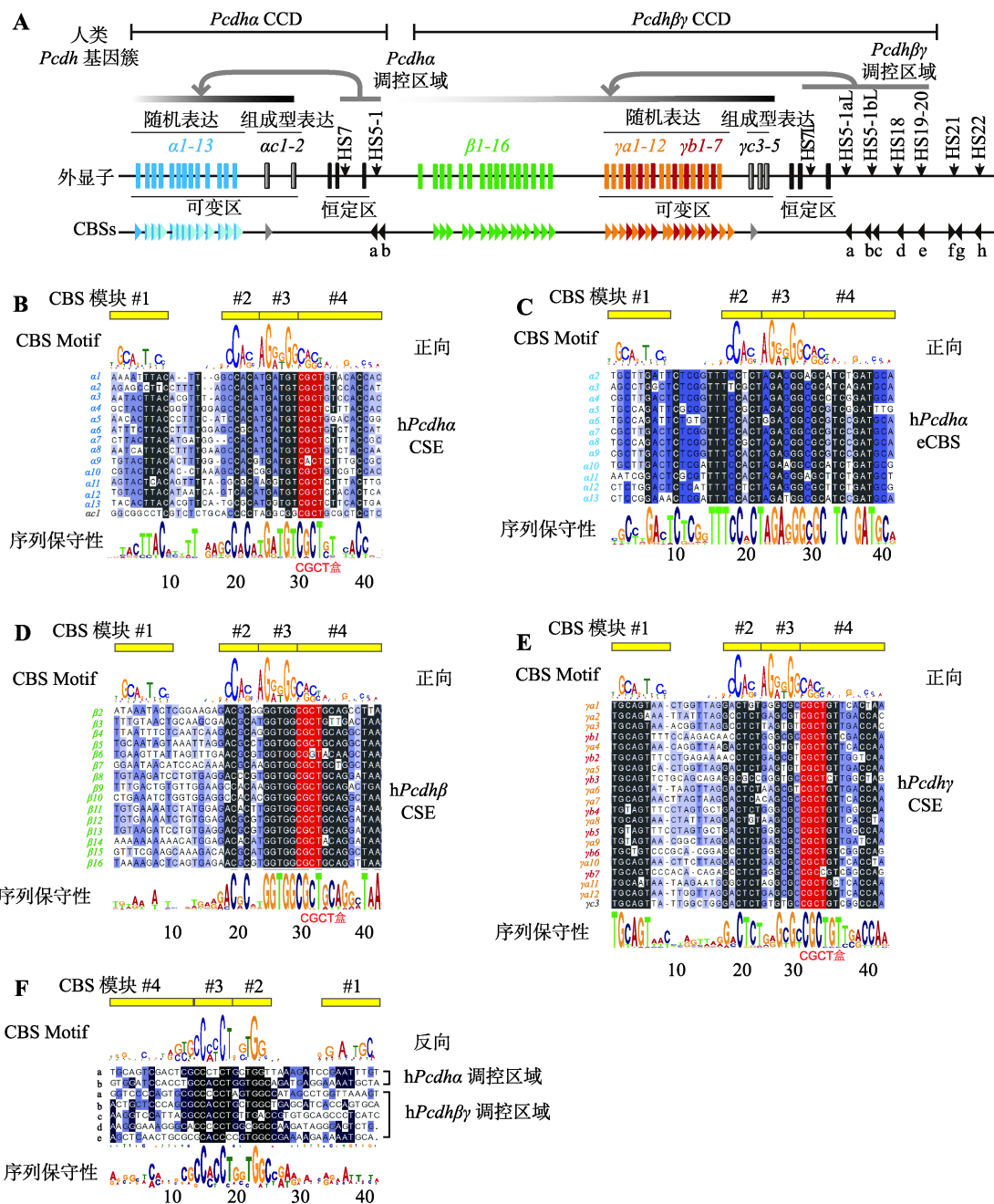
扑结构域^[24]。

2.2 *Pcdh β* 下游调控区域中的染色质标志 (Chromatin marks) 和 CBS 位点的验证

不同物种间的基因组调控元件具有高度保守性^[33]。最近的研究发现 *Pcdh β* 下游调控序列包含 HS5-1aL、HS5-1bL、HS18、HS19-20、HS21 和 HS22 等 DNase I 超敏位点^[15,20,24]。对比人类和小鼠 *Pcdh β* 下游的这些超敏位点序列 (图 1A 和图 2A), 本文发现该区域上所有的 HS 超敏位点都较保守 (图 2A)。在人类神经母细胞瘤 SK-N-SH 细胞系中, 利用 CTCF 或 Rad21 特异性抗体做的 ChIP-Seq (GSE76815) 结果显示每个 HS 超敏位点附近均有 CTCF 蛋白富集 (图 2A)。

CTCF 和 Rad21 常结合在基因组的启动子和增强子区域^[24,34,35]。为了排除这些 HS 超敏位点不是启动子, 本文利用 H3K4Me3 和 RNAP II (H3K4Me3 和 RNAP II 是启动子标志^[32]) 特异性抗体做了 ChIP-Seq 实验 (GSE76815), 数据显示这些 HS 超敏位点无启动子标志 (图 2A)。进一步分析增强子标志^[32] H3K4Me1、H3K27ac 和 p300 的 ChIP-Seq 数据, 发现以上增强子标志在每个 HS 超敏位点均有富集 (图 2A), 说明这些 HS 超敏位点可能是增强子簇, 形成一个超级增强子区域^[36]。本文利用 H3K27Me3 (异染色质标志^[37]) 特异性抗体的 ChIP-Seq 实验 (GSE76815) 结果证实这一区域不含异染色质 (图 2A)。

本研究发现在 HS5-1aL、HS5-1bL、HS18、HS19-20、HS21 和 HS22 等超敏位点都存在潜在的 CBS 位点。在人类和小鼠的 HS5-1aL、HS18 和 HS19-20 分别找到 1 个潜在的 CBS 位点, 在 HS21 和 HS22 则分别找到 2 个潜在的 CBS 位点, 在人类 HS5-1bL 找到 3 个潜在的 CBS 位点, 在小鼠 HS5-1bL 找到 2 个潜在的 CBS 位点 (图 2B)。特别重要的是, 像 *Pcdh α* 基因簇的增强子 HS5-1 区域的两个 CBS 位点 (HS5-1a 和 HS5-1b) 反向的方向性一样, 本文发现 *Pcdh β* 基因簇调控区域 HS5-1aL、HS5-1bL、HS18 和 HS19-20 超敏位点的 CBS 位点都是反向的^[24] (图 1F)。将以上这些 CBS 位点的序列在人类和小鼠中进行保守性分析, 可以看出在人类和小鼠中这些位点的 CBS 序列是极为相似的 (图 2C)。基于上面的结论, 本文发现在人类和小鼠 *Pcdh β* 下游调控区域中有大量潜在的 CBS 位点, 并且这些位点的方向有一

图 1 人类 *Pcdh* 基因簇中 CBS 位点方向性分析**Fig. 1** Characteristics and orientations of the CBS sites in the three human *Pcdh* gene clusters

A: 人类 *Pcdh* 基因簇示意图。人类 *Pcdh* 基因座包含 53 个高度相似的基因形成 3 个基因簇: *Pcdha* 基因簇 ($\alpha 1 \sim \alpha 13$ 和 $\alpha c 1 \sim \alpha c 2$), *Pcdhb* 基因簇 ($\beta 1 \sim \beta 16$) 和 *Pcdhy* 基因簇 ($\gamma 1 \sim \gamma 12$, $\gamma b 1 \sim \gamma b 7$ 和 $\gamma c 3 \sim \gamma c 5$)。 *Pcdha* 和 γ 基因簇由可变区 (Variable region) 和恒定区 (Constant region) 构成。 *Pcdha* 基因簇可变区包含 13 个随机表达的 A 型 ($\alpha 1 \sim \alpha 13$) 和两个组成型表达的 C 型 ($\alpha c 1$ 和 $\alpha c 2$) 可变区外显子和 3 个恒定区外显子。 *Pcdhy* 基因簇可变区包含 12 个随机表达的 A 型 ($\gamma 1 \sim \gamma 12$)、7 个随机表达的 B 型 ($\gamma b 1 \sim \gamma b 7$) 和 3 个组成型表达的 C 型 ($\gamma c 3 \sim \gamma c 5$) 可变区外显子以及 3 个恒定区外显子。 *Pcdhb* 基因簇只有 16 个可变区外显子而没有恒定区外显子。垂直框代表 *Pcdh* 外显子并用不同颜色标记, 其对应的 CTCF 结合位点 (CBS) 用不同深浅三角标记在下方, 三角标记的方向代表 CBS 位点的方向, 垂直的黑色箭头代表 DNase I 超敏位点; B~F: 人类 *Pcdh* 基因簇中 CBS 位点序列及方向性; 具有 Module1~#4 的 CBS 位点定义为正向, 而具有 Module4~#1 的 CBS 位点则为反向。 *Pcdh* 基因簇编码区上游的 CSE (Conserved sequence elements)^[10] 及 *Pcdha* 基因簇外显子 CBS (Exonic CTCF-binding sites, eCBSs) 位点均为正向, *Pcdh* 基因簇调控区域中的 CBS 位点为反向; *Pcdha* CSE (B), *Pcdha* eCBS (C), *Pcdhb* CSE (D), *Pcdhy* CSE (E), *Pcdh* 调控区域 (F), 红色区域为 CGCT 盒^[10]。

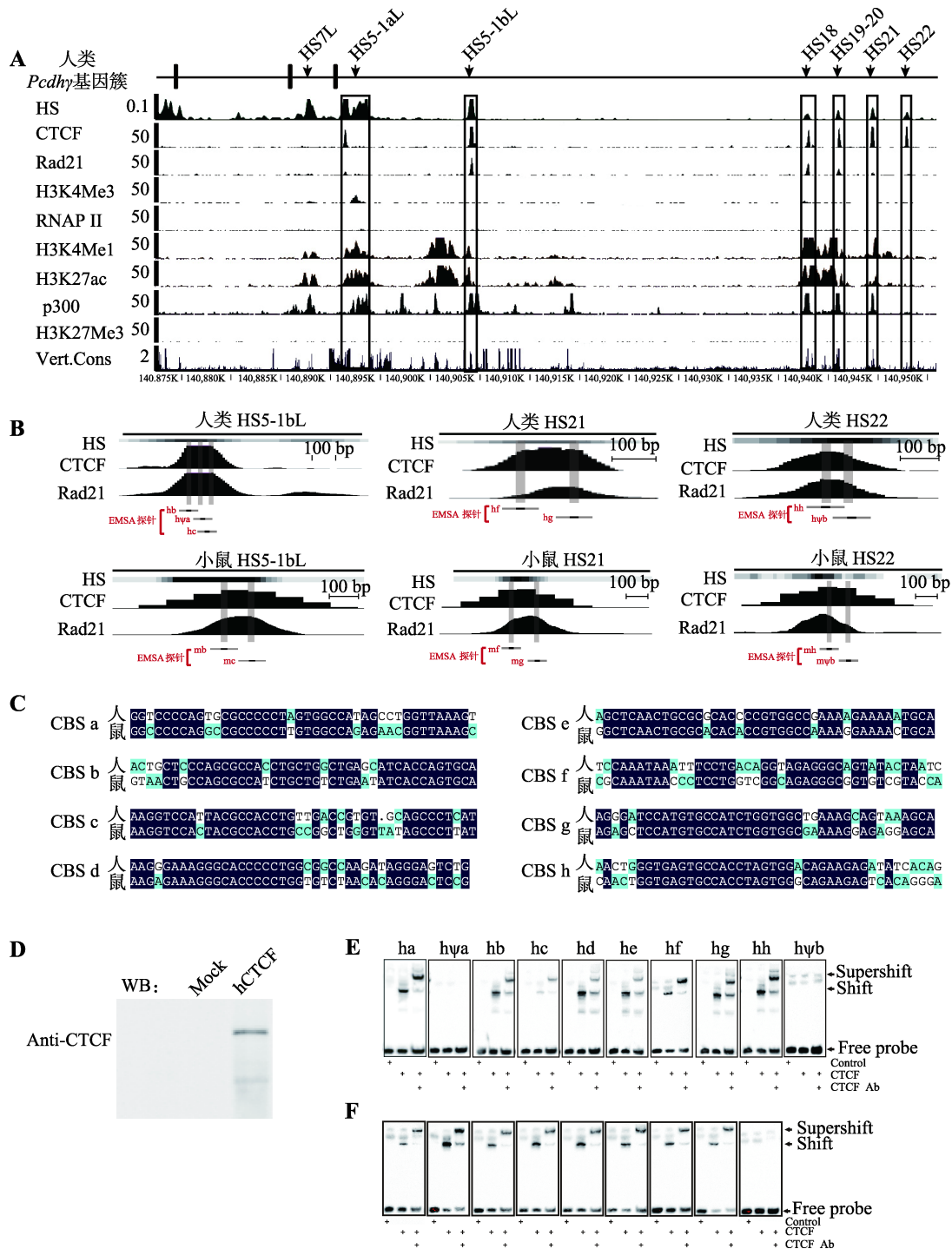


图 2 CTCTF 在 *Pcdhβγ* 调控区域的分布

Fig. 2 The distribution of CTCTF in the regulatory region of the *Pcdhβγ* gene clusters

A: *Pcdhβγ* 下游 HS 超敏位点保守性分析。CTCF、Rad21、H3K4Me3、RNAP II 和 H3K27Me3 的数据来自前期的 ChIP-Seq 实验(GSE76815)。人类胚胎干细胞 H1-hESC (HS)、人类脐静脉管内皮细胞系 HUVEC(H3K4Me1 和 H3K27ac)以及人类神经母细胞瘤细胞系 SK-N-SH(p300)的 ChIP-Seq 数据来自 ENCODE 项目^[31]; B: HS 超敏位点中 CBS 位点的确定。灰色垂直框代表 CBS 位点在 CTCF ChIP-Seq 峰中的位置,底部水平框代表 EMSA 实验中的探针位置; C: 人类和小鼠 CBS 位点(a~h)序列保守性分析; D: 体外表达的人源 CTCTF 蛋白的免疫印迹; E: 人类 *Pcdhβγ* 调控区域内潜在在 CBS 位点的 EMSA 分析; F: 小鼠 *Pcdhβγ* 调控区域内潜在在 CBS 位点的 EMSA 分析。Free probe 代表没有结合蛋白的 DNA 探针迁移的条带,Shift 代表和 CTCTF 蛋白结合后的 DNA 探针迁移的条带,Supershift 代表和 CTCTF 蛋白及其抗体结合后的 DNA 探针迁移的条带。

定的规律,这可能在 *Pcdh β* 基因表达调控中发挥重要作用。

通过前面的分析本文找到了 *Pcdh β* 的调控区域中每个 HS 区域内的潜在 CBS 位点。通过一系列凝胶电泳迁移实验(EMSA)来验证这些 CBS 位点与 CTCF 的结合,从而确定每个 ChIP-Seq 峰中 CTCF 蛋白结合的具体位置,以进一步研究 CTCF 蛋白在基因组中的功能。EMSA 是一种体外研究蛋白质和 DNA 探针相互作用的实验,DNA 探针带有亲和素标记,在化学物质作用下会发荧光,在相同的电场强度下结合了蛋白的 DNA 探针比自由 DNA 探针泳动速度慢,因此会出现一条迁移的条带(Shift),加入相应抗体就会出现一条迁移更慢的条带(Supershift),可以进一步验证蛋白结合 DNA 的特异性。

本文利用兔网织红细胞系统表达重组的人源 CTCF 蛋白(图 2D)。EMSA 实验证明 hHS5-1aL、

hHS18 和 hHS19-20 区域中所预测的 1 个潜在 CBS 位点都能与 CTCF 结合,在 hHS5-1bL 的 3 个潜在 CBS 位点中,hb 和 hc 能与 CTCF 结合,h ψ a 不能与 CTCF 结合;hHS21 的两个 CBS 位点(hf 和 hg)都能与 CTCF 结合;在 hHS22 的两个潜在位点中,hh 能与 CTCF 结合,但 h ψ b 不能与 CTCF 结合(图 2E)。在小鼠中,除了 mHS22 位点中第 2 个 m ψ b 之外所有预测的潜在 CBS 位点都能与 CTCF 结合(图 2F)。

2.3 CBS 位点序列对蛋白 CTCF 结合的影响

将前面确定并且验证得到的人类和小鼠 *Pcdh β* 调控区域的 CBS 位点序列进行对比,并使用 Weblogo 软件^[30]确定人类(图 3A)和小鼠(图 3B)的 CTCF 结合基序。可以看出,人类和小鼠中,42 bp 的 CTCF 结合序列都较为保守,特别是模块 3^[25]的序列更加保守(图 3, A 和 B),这些保守碱基可能对

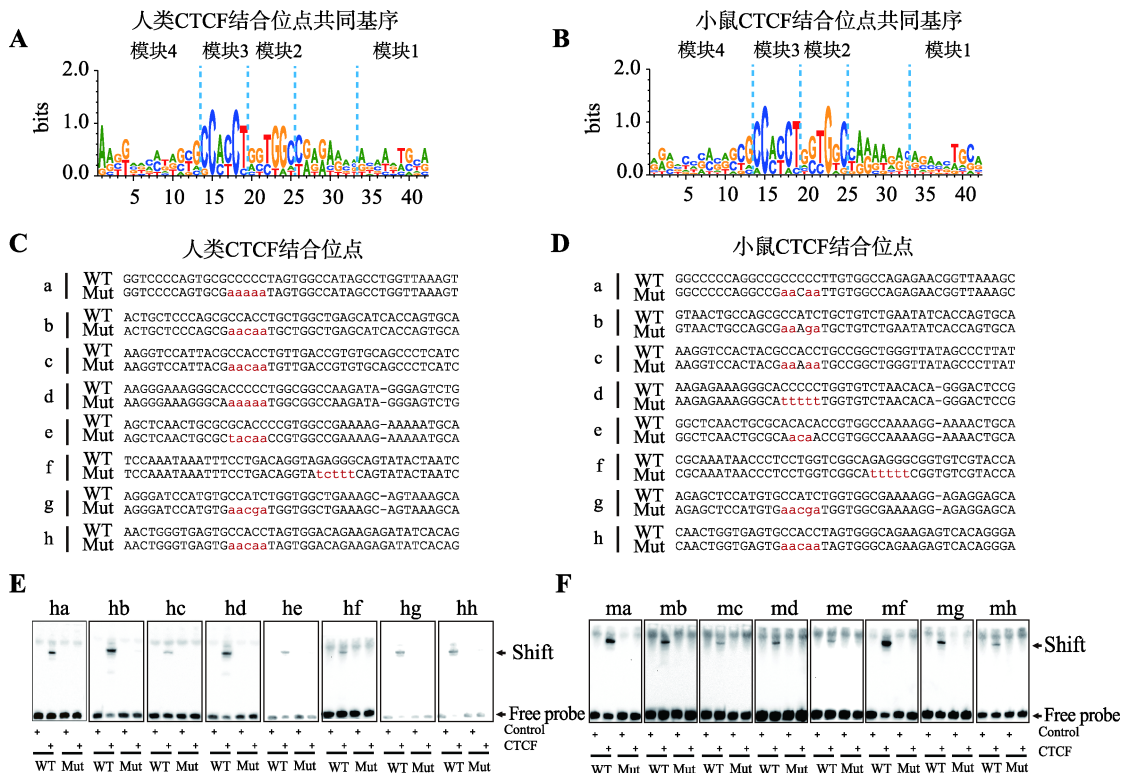


图 3 CBS 位点序列对 CTCF 结合的影响

Fig. 3 The effect of CBS sequence mutations on CTCF binding

A: 人类 *Pcdh β* 下游调控区域中 CBS 位点的共同基序(Consensus motif); B: 小鼠 *Pcdh β* 下游调控区域中 CBS 位点的共同基序; C: 人类 *Pcdh β* 下游调控区域中的 CBS 位点序列。D: 小鼠 *Pcdh β* 下游调控区域中的 CBS 位点序列。WT 表示野生型序列, Mut 表示突变序列, 红色字体表示突变位点; E: 人类 *Pcdh β* 调控区域中 CBS 位点的 EMSA 分析; F: 小鼠 *Pcdh β* 调控区域中 CBS 位点的 EMSA 分析。人类和小鼠的 *Pcdh β* 调控区域一共有 8 个 CBS 位点(CBS a~h), CBS f 的方向是正向的, 与其他 7 个反向的位点相反。WT 表示野生型序列, Mut 表示突变序列。

CTCF 的结合较为重要。本文对人类和小鼠 *Pcdhβγ* 下游 HS 超敏位点的 CBS 位点序列进行一系列的突变来研究 CBS 位点序列中的起决定性作用的结合序列。所做突变的 CBS 位点的野生型序列 WT 及突变序列 Mut 详见图 3C 和图 3D, 其中红色字体为突变后的碱基序列。对以上所有的 WT 和 Mut 序列进行 EMSA 实验, 发现将 CBS 位点的 5 bp 序列突变后, DNA 序列不再具有结合 CTCF 蛋白的能力(图 3, E 和 F), 由此说明 Module3 对于 CTCF 蛋白的结合是至关重要的, 并进一步证明了本文所找到的 CTCF 具体结合位点的准确性。

2.4 全基因组分析 CTCF 在调控区域的分布

通过前面实验本文发现在 CTCF 蛋白的 ChIP-Seq

峰附近有一个或多个 CBS 位点, 这些 CBS 位点可能位于基因的启动子和增强子等调控区域。最近发现很多增强子标志(Enhancer Mark)p300 富集区域有 CBS 位点, 并且他们可能与启动子和增强子之间长距离染色质环化有关^[24]。本文在全基因组范围内利用 ENCODE 计划^[31]中人类慢性骨髓性白血病细胞系 K562 的 ChIP-Seq 数据分析了 CTCF 和增强子标志^[32] H3K4Me1, H3K27ac 以及启动子标志^[32] RNAP II, H3K4Me3 的关系。本文发现有 44% 的 H3K4Me1 峰(图 4A)和 45% 的 H3K27ac 峰(图 4B)附近 2 kb 内至少有一个 CBS 位点; 79% 的 RNAP II 峰(图 4C)和 72% 的 H3K4Me3 峰(图 4D)附近 2 kb 内至少有一个 CBS 位点。本文推测 CTCF 蛋白结合在增强子和启动子等调控元件附近, 可能参与增强子和启动子之间的

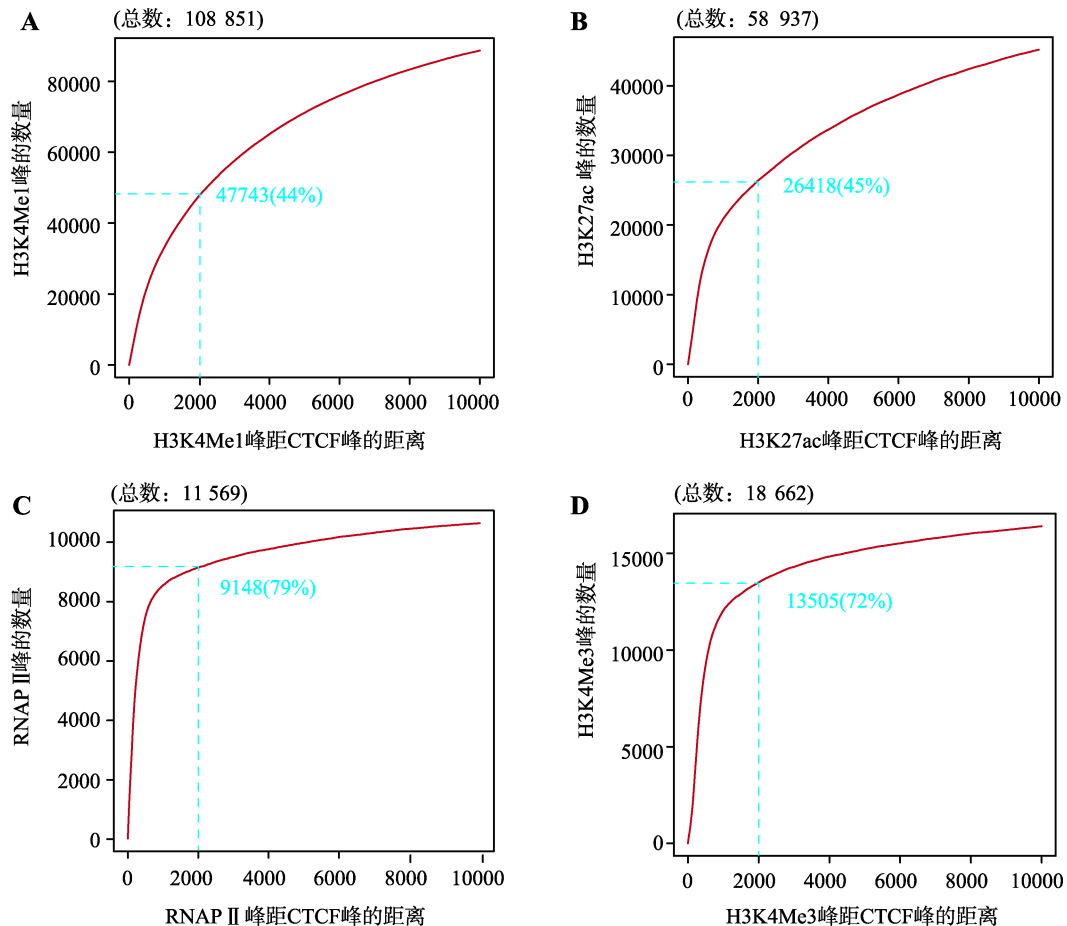


图 4 人类全基因组中 CTCF 结合位点和调控区域标志的关系

Fig. 4 The relationship between CBS and chromatin marks in the regulatory regions of the human genome

CTCF 结合位点和以下 4 种基因调控区域标志的关系。A: 增强子标志 H3K4Me1; B: 增强子标志 H3K27ac; C: 启动子标志 RNAP II; D: 启动子标志 H3K4Me3。横坐标指标志峰(H3K4Me1、H3K27ac、RNAP II 和 H3K4Me3)到 CTCF 峰的距离; 纵坐标指在横坐标所示的标志峰到 CTCF 峰的距离以内的标志峰的个数。

染色质长距离环化^[38]。例如, ENCODE 计划^[31]中 K562 的 ChIA-PET(Chromatin interaction analysis by paired-end tag sequencing)数据^[39,40]显示在 CTCF 蛋白的介导下, *Pcdhβγ*的调控区域和 *Pcdhβγ*基因启动子区域发生相互作用(图 5A)。

3 讨论

神经元是组成神经系统的基本单位, 原钙粘蛋白对于神经元迁移、树突和轴突发育、神经元之间的通讯及自我回避至关重要^[6,34]。因此, *Pcdh* 基因簇表达调控的研究是神经系统及大脑功能研究的一个重要组成部分。但是, 目前 *Pcdh* 基因簇的表达调控机制并不完全清楚, 该基因簇的增强子调控元件发挥作用的机制也未完全确定。本课题组最近的研究表明 CBS 位点的方向性在基因表达调控和染色体三维结构形成中起到了重要作用^[24,35]。本文详细研究了转录因子 CTCF 在 *Pcdhβγ*基因簇下游调控区域内的结合和保守性, 这对于探索增强子调控元件的

作用机制极其重要。

本文通过人类神经母细胞瘤细胞系 SK-N-SH 的 CTCF ChIP-Seq 实验, 发现 CBS 位点在 *Pcdh* 基因簇的启动子和增强子区域广泛分布。在本研究中, 利用一系列的 EMSA 实验找到了 *Pcdhγ*下游的 HS 超敏位点(HS5-1aL、HS5-1bL、HS18、HS19-20、HS21 和 HS22)中 CTCF 的具体结合位点。有趣的是, 在一些单个 CTCF ChIP-Seq 峰中找到了 2 个 CBS 位点, 如 HS5-1bL 和 HS21。CBS 位点序列突变的 EMSA 实验显示 CBS 位点的模块 3 序列对 CTCF 的结合极其重要, 5 bp 碱基突变后的序列不能结合 CTCF, 该结论也进一步表明本文所找到的具体 CBS 位点的准确性。*Pcdh* 基因簇中, 几乎每个启动子都有 CBS 位点^[15], 增强子区域包含成簇的 CBS 位点, 并且这些 CBS 位点序列具有一定的方向性(图 1), 表明 CTCF 可能参与 *Pcdh* 基因簇的表达调控。但 3C(Chromosome conformation capture)实验表明 HS21 和 HS22 超敏位点不能和 *Pcdhβγ*启动子区域形成染色质环, 因此这

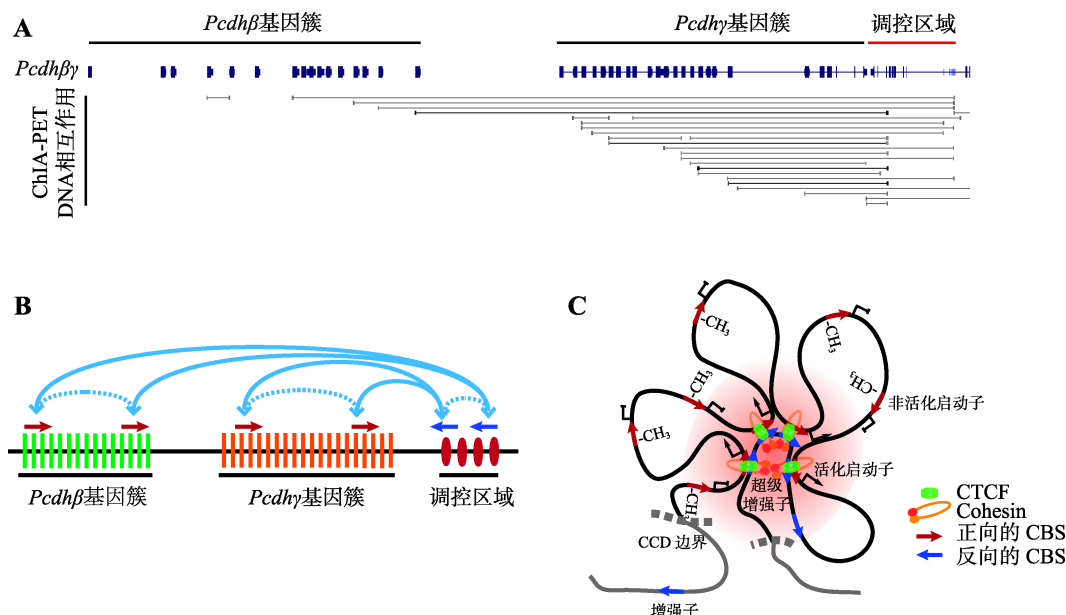


图 5 CTCF 介导的调控区域与启动子之间的染色质相互作用模型

Fig. 5 The CCD model of CTCF-mediated chromatin interactions between CBS clusters in the regulatory region and promoters

A: *Pcdhβγ*基因簇中人类慢性骨髓性白血病细胞系 K562 的 CTCF ChIA-PET 数据(来自 ENCODE 计划); B: *Pcdh* 基因簇调控区域与启动子相互作用示意图。垂直框表示 *Pcdh* 基因, 红色椭圆表示 HS 超敏位点, 红蓝箭头表示 CBS 位点和其方向, 蓝色弧状箭头表示 DNA 间相互作用, CBS 位点正向反向成对之间有强的相互作用(实线), 同向的 CBS 位点之间有较弱的相互作用(虚线); C: CCD (CTCF-cohesin mediated chromatin domain)模型。红色区域内表示超级增强子, 灰色虚线表示 CCD 界限, 黑色缠绕的线表示 CCD, 其他部分如右下角图例所示; 靠近超级增强子区域的启动子被激活进行转录, 外围的启动子未被激活则相应基因不进行转录。

2 个 HS 超敏位点不是 *Pcdh β* 的增强子^[24]。这可能与 HS21 超敏位点的 CBS 位点序列的方向有关, HS21 超敏位点中的 CBS 位点方向是正向的, 与其余位点的 CBS 位点方向相反, 这可能导致 HS21 区域不能作用于 *Pcdh β* 的启动子。此外, CBS 位点序列异常, 如单核苷酸多态性(SNP), 导致一些重要基因的表达异常, 并进一步引发某些疾病^[41]。以上现象进一步说明了 CTCF 对于基因表达的重要性。

在 *Pcdh β* 基因簇中, 除了 *Pcdh β 1*, *Pcdh γ 4* 和 *Pcdh γ 5* 之外的所有原钙粘蛋白启动子上都有 CBS 位点, 并且这些 CBS 位点的方向都是正向的(图 1A, 图 5B)。同时在 *Pcdh β* 下游的调控区域有 4 个连续的 HS 区域, 这些 HS 区域共包含有 5 个方向都是反向的 CBS 位点。ENCODE 计划^[31]中 K562 的 CTCF ChIA-PET 数据^[39,40]表明 *Pcdh β* 基因簇的调控区域和启动子区域有相互作用(图 5A)。综合前面的研究, 本文提出 *Pcdh β* 拓扑结构域中通过超级增强子^[42]染色质环化形成活性转录中心的基因调控模型(图 5C), *Pcdh β* 启动子的正向 CBS 位点和调控区域超级增强子的反向 CBS 位点通过长距离染色质环化形成 *Pcdh β* 拓扑结构域(CCD)的调控中心。*Pcdh β* 基因簇中需要表达的基因的启动子通过染色体环化使得启动子上的 CBS 位点与超级增强子接触, 形成启动子和超级增强子上的正向反向 CBS 对形成远程交互作用, 这些启动子被激活并转录表达。*Pcdh β* 基因簇中不需要表达的基因的启动子与超级增强子没有相互作用, 于是启动子是非激活的。

在基因组范围内利用已有的 ENCODE 数据进行了一些验证, 分析发现基因组上增强子标志^[32] H3K4Me1 和 H3K27ac 的附近大都有 CTCF 结合, 启动子标志^[32] RNAP II 和 H3K4Me3 附近也有 CTCF 的结合, 本文认为 CTCF 多结合在调控元件如启动子和增强子附近。因此, 在 CTCF 及其他因子的介导下增强子可能和启动子在空间距离上相互靠近形成远距离染色质环(Long-distance chromatin loops), 使得增强子上结合的蛋白因子可以促进基因转录^[38]。本研究一方面揭示了转录因子 CTCF 方向性结合在 *Pcdh β* 调控区域的重要生物作用, 另一方面为探索增强子和绝缘子的作用机制提供了新的方向, 例如可以利用 CRISPR 介导的基因编辑技术^[43,44]研究它们在基因表达调控中的作用机制。

参考文献(References):

- [1] Li W, Wu Q. Protocadherin and the diversity of neurons. *Sci Technol Vision*, 2013, (27): 14.
李伟, 吴强. 原钙粘蛋白分子与神经元的多样性. *科技视界*, 2013, (27): 14. [DOI]
- [2] Chen WV, Alvarez FJ, Lefebvre JL, Friedman B, Nwakeze C, Geiman E, Smith C, Thu CA, Tapia JC, Tasic B, Sanes JR, Maniatis T. Functional significance of isoform diversification in the protocadherin gamma gene cluster. *Neuron*, 2012, 75(3): 402–409. [DOI]
- [3] Garrett AM, Schreiner D, Lobas MA, Weiner JA. γ -protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron*, 2012, 74(2): 269–276. [DOI]
- [4] Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T. CTCF is required for neural development and stochastic expression of clustered *Pcdh* genes in neurons. *Cell Rep*, 2012, 2(2): 345–357. [DOI]
- [5] Lefebvre JL, Kostadinov D, Chen WV, Maniatis T, Sanes JR. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature*, 2012, 488(7412): 517–521. [DOI]
- [6] Suo L, Lu HN, Ying GX, Capecchi MR, Wu Q. Protocadherin clusters and cell adhesion kinase regulate dendrite complexity through Rho GTPase. *J Mol Cell Biol*, 2012, 4(6): 362–376. [DOI]
- [7] Wu Q. Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. *Genetics*, 2005, 169(4): 2179–2188. [DOI]
- [8] Wu Q, Maniatis T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*, 1999, 97(6): 779–790. [DOI]
- [9] Wu Q, Maniatis T. Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes. *Proc Natl Acad Sci USA*, 2000, 97(7): 3124–3129. [DOI]
- [10] Wu Q, Zhang T, Cheng JF, Kim Y, Grimwood J, Schmutz J, Dickson M, Noonan JP, Zhang MQ, Myers RM, Maniatis T. Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res*, 2001, 11(3): 389–404. [DOI]
- [11] Zou C, Huang W, Ying G, Wu Q. Sequence analysis and expression mapping of the rat clustered protocadherin gene repertoires. *Neuroscience*, 2007, 144(2): 579–603. [DOI]
- [12] Tasic B, Nabholz CE, Baldwin KK, Kim Y, Rueckert EH, Ribich SA, Cramer P, Wu Q, Axel R, Maniatis T.

- Promoter choice determines splice site selection in protocadherin α and γ pre-mRNA splicing. *Mol Cell*, 2002, 10(1): 21–33. [DOI]
- [13] Wang XZ, Weiner JA, Levi S, Craig AM, Bradley A, Sanes JR. Gamma-protocadherins are required for survival of spinal interneurons. *Neuron*, 2002, 36(5): 843–854. [DOI]
- [14] Esumi S, Kakazu N, Taguchi Y, Hirayama T, Sasaki A, Hirabayashi T, Koide T, Kitsukawa T, Hamada S, Yagi T. Monoallelic yet combinatorial expression of variable exons of the protocadherin- α gene cluster in single neurons. *Nat Genet*, 2005, 37(2): 171–176. [DOI]
- [15] Guo Y, Monahan K, Wu HY, Gertz J, Varley KE, Li W, Myers RM, Maniatis T, Wu Q. CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc Natl Acad Sci USA*, 2012, 109(51): 21081–21086. [DOI]
- [16] Schreiner D, Weiner JA. Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc Natl Acad Sci USA*, 2010, 107(33): 14893–14898. [DOI]
- [17] Thu CA, Chen WV, Rubinstein R, Chevee M, Wolcott HN, Felsovalyi KO, Tapia JC, Shapiro L, Honig B, Maniatis T. Single-cell identity generated by combinatorial homophilic interactions between α , β , and γ protocadherins. *Cell*, 2014, 158(5): 1045–1059. [DOI]
- [18] Ribich S, Tasic B, Maniatis T. Identification of long-range regulatory elements in the protocadherin- α gene cluster. *Proc Natl Acad Sci USA*, 2006, 103(52): 19719–19724. [DOI]
- [19] Wu HY, Guo Y, Li W, Wu Q. Cloning and functional analysis of the regulatory elements in the human protocadherin gene cluster. *Life Sci Res*, 2014, 18(2): 95–99.
吴海洋, 郭亚, 李伟, 吴强. 人类原钙粘蛋白基因簇调控元件的克隆及对其启动子活性的影响. *生命科学研究*, 2014, 18(2): 95–99. [DOI]
- [20] Yokota S, Hirayama T, Hirano K, Kaneko R, Toyoda S, Kawamura Y, Hirabayashi M, Hirabayashi T, Yagi T. Identification of the cluster control region for the protocadherin- β genes located beyond the protocadherin- γ cluster. *J Biol Chem*, 2011, 286(36): 31885–31895. [DOI]
- [21] Nichols MH, Corces VG. A CTCF code for 3D genome architecture. *Cell*, 2015, 162(4): 703–705. [DOI]
- [22] Lobanenko VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene*, 1990, 5(12): 1743–1753. [DOI]
- [23] Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 1999, 98(3): 387–396. [DOI]
- [24] Guo Y, Xu Q, Canzio D, Shou J, Li JH, Gorkin DU, Jung I, Wu HY, Zhai Y, Tang YX, Lu YC, Wu YH, Jia ZL, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*, 2015, 162(4): 900–910. [DOI]
- [25] Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, 2011, 147(6): 1408–1419. [DOI]
- [26] Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenko VV, Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 2007, 128(6): 1231–1245. [DOI]
- [27] Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, 2014, 15(4): 234–246. [DOI]
- [28] Vietri Rudan M, Hadjur S. Genetic Tailors: CTCF and cohesin shape the genome during evolution. *Trends Genet*, 2015, 31(11): 651–660. [DOI]
- [29] Feng JX, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*, 2012, 7(9): 1728–1740. [DOI]
- [30] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*, 2004, 14(6): 1188–1190. [DOI]
- [31] The ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. *Science*, 2004, 306(5696): 636–640. [DOI]
- [32] Heintzman ND, Stuart RK, Hon G, Fu YT, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu CX, Ching KA, Wang W, Weng ZP, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 2007, 39(3): 311–318. [DOI]
- [33] Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA. Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, 2000, 288(5463): 136–140. [DOI]
- [34] Jia ZL, Guo Y, Tang YX, Xu Q, Li BJ, Wu Q. Regulation of the protocadherin *Celsr3* gene and its role in globus pallidus development and connectivity. *Mol Cell Biol*, 2014, 34(20): 3895–3910. [DOI]
- [35] Guo Y, Wu Q. The architectural rule of CTCF-mediated topological genome looping and enhancer insulation.

- Hereditas (Beijing)*, 2015, 37(10): 1073–1074.
- 郭亚, 吴强. 采用 DNA 片段编辑技术反转 CTCF 结合位点改变基因组拓扑结构和增强子与启动子功能. *遗传*, 2015, 37(10): 1073–1074. [DOI]
- [36] Downen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, Young RA. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 2014, 159(2): 374–387. [DOI]
- [37] Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappaport J, Lerdrup M, Helin K. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol*, 2008, 10(11): 1291–1300. [DOI]
- [38] Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*, 2012, 489(7414): 109–113. [DOI]
- [39] Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han YY, Ooi HS, Ariyaratne PN, Vega VB, Luo YQ, Tan PY, Choy PY, Wansa KDSA, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li HX, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan XA, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan YJ. An oestrogen-receptor- α -bound human chromatin interactome. *Nature*, 2009, 462(7269): 58–64. [DOI]
- [40] Li GL, Ruan XA, Auerbach RK, Sandhu KS, Zheng MZ, Wang P, Poh HM, Goh Y, Lim J, Zhang JY, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong SZ, Zhang ZZ, Landt S, Raha D, Euskirchen G, Wei CL, Ge WH, Wang HE, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan YJ. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, 2012, 148(1): 84–98. [DOI]
- [41] Tang ZH, Luo OJ, Li XW, Zheng MZ, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Włodarczyk J, Ruszczycki B, Michalski P, Piecuch E, Wang P, Wang DJ, Tian SZ, Penrad-Mobayed M, Sachs LM, Ruan XA, Wei CL, Liu ET, Wilczynski GM, Plewczynski D, Li GL, Ruan YJ. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, 2015, 163(7): 1611–1627. [DOI]
- [42] Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*, 2015, 16(3): 144–154. [DOI]
- [43] Li JH, Shou J, Guo Y, Tang YX, Wu YH, Jia ZL, Zhai Y, Chen ZF, Xu Q, Wu Q. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J Mol Cell Biol*, 2015, 7(4): 284–298. [DOI]
- [44] Li JH, Shou J, Wu Q. DNA fragment editing of genomes by CRISPR/Cas9. *Hereditas(Beijing)*, 2015, 37(10): 992–1002.
- 李金环, 寿佳, 吴强. CRISPR/Cas9 系统在基因组 DNA 片段编辑中的应用. *遗传*, 2015, 37(10): 992–1002. [DOI]

(责任编辑: 方向东)