

串联反向 CTCF 位点的系列删除揭示增强子调控 *HOXD* 基因簇表达的平衡

王玲, 李金环, 黄海燕, 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: 三维基因组染色质架构蛋白 CTCF (CCCTC-binding factor)能够介导增强子与基因启动子的远距离染色质相互作用,也可以结合调控区域的绝缘子发挥增强子绝缘功能,对发育中的基因表达调控具有重要作用。同源框基因家族(Homeobox gene family, *Hox*)编码一类控制动物发育的关键转录因子,在发育中主要沿胚胎首尾轴(head-to-tail axis)呈时空线性表达。在哺乳动物中,*Hox* 基因分为 *HoxA*、*HoxB*、*HoxC* 和 *HoxD* 四个基因簇,在中枢神经系统、骨骼和四肢发育中发挥重要功能。*HoxD* 基因簇主要调控四肢发育,受位于其两侧调控域内的增强子调节,沿肢体近远轴(proximal-distal axis)呈时空线性表达。在人类基因组中,*HOXD* 基因簇及其两侧的调控区域分布有串联排列的 CTCF 结合位点(简称 CTCF 位点),参与 9 个 *HOXD* 基因的表达调控。本研究以 *HOXD* 基因簇为模式基因,探究 CTCF 对发育基因(developmental genes)转录调控的影响。利用 CRISPR DNA 片段编辑技术在人 HEK293T 细胞中获得一系列的串联反向 CTCF 位点删除的单细胞克隆株。RNA-seq 实验揭示 CTCF 位点删除后 *HOXD* 基因表达下降。定量高分辨率染色体构象捕获实验显示,*HOXD* 与上游增强子簇的远距离染色质相互作用增强,与下游增强子簇的远距离染色质相互作用减弱。综上所述,串联反向的 CTCF 位点通过其绝缘子功能维持上下游增强子簇对 *HOXD* 基因簇表达调控的平衡,为探究动物发育过程中 *Hox* 基因表达的精准调控机制提供参考。

关键词: *HOXD* 基因簇; CTCF 位点; 调控平衡; 增强子; 绝缘子

收稿日期: 2021-04-11; 修回日期: 2021-05-12

基金项目: 国家自然科学基金项目(编号: 31800636, 31630039, 91940303)和上海市科学技术委员会项目(编号: 19JC1412500)资助[Supported by the National Natural Science Foundation of China (Nos. 31800636, 31630039, 91940303), and Science and Technology Commission of Shanghai Municipality (No. 19JC1412500)]

作者简介: 王玲, 硕士研究生, 专业方向: 生物学。E-mail: wanglingmail0613@163.com

通讯作者: 黄海燕, 博士, 副研究员, 研究方向: 药物分子遗传学。E-mail: hy_huang@sjtu.edu.cn

吴强, 博士, 教授, 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.21-132

网络出版时间: 2021/6/30 14:14:38

URI: <https://kns.cnki.net/kcms/detail/11.1913.R.20210630.1109.002.html>

Serial deletions of tandem reverse CTCF sites reveal balanced *HOXD* regulatory landscape of enhancers

Ling Wang, Jinhuan Li, Haiyan Huang, Qiang Wu

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

Abstract: The genome architectural protein CTCF (CCCTC-binding factor) not only mediates long-distance chromatin interactions between distal enhancers and target promoters, but also functions as an important insulator-binding factor to block improper enhancer activation of non-target promoters, and is thus of great significance to transcriptional regulation of developmental genes. The *Hox* (Homeobox) gene family plays an important role in the development of the brain, bones, and limbs. The spatiotemporal colinear expression of the *HOXD* cluster along the proximal-distal axis of limbs is regulated by two clusters of enhancers known as super-enhancers located in the flanking regulatory regions. We focused on the *HOXD* cluster to explore the architectural role of CTCF in transcriptional regulation of developmental genes. The *HOXD* cluster contains 9 paralogous genes intermixed with a series of CBS (CTCF-binding site) elements. Using the CRISPR DNA-fragment editing system, we generated a series of single-cell HEK293T clones with deletion of increasing numbers of reverse CBS elements. RNA-seq experiments revealed decreased levels of *HOXD* gene expression. In addition, chromosome conformation capture experiments revealed increased long-distance chromatin interactions between *HOXD* and the upstream enhancer cluster and corresponding decreased interactions between *HOXD* and the downstream enhancer cluster. Thus, tandem reverse CTCF sites function as insulators to maintain *HOXD* regulatory balance between the upstream and downstream enhancer clusters. This study has interesting implications on the precise gene expression control of the *Hox* family during animal development.

Keywords: *HOXD* gene cluster; CTCF site; balanced regulatory mechanism; enhancer; insulator

染色质架构蛋白 CTCF(CCCTC-binding factor)是一类结合 DNA 的锌指蛋白,在哺乳动物中可与基因组上调控区域内大量的绝缘子结合,对发育基因的精准调控起到关键作用^[1-6]。CTCF 最初在鸡中被发现是一种转录抑制因子或共抑制因子^[7],CTCF 结合位点(CTCF-binding site, CBS)序列在脊椎动物中高度保守,超过 30 多万的 CBS 广泛分布在基因组中^[8]。CTCF 与 cohesin 复合物通过环挤出模型(loop extrusion model)共同介导远距离染色质相互作用,参与基因的调控^[3,9,10]。CBS 序列的方向性直接影响染色质远距离互作,分别被正向和反向 CBS 元件锚定的增强子和启动子在 CTCF/cohesin 介导下形成远距离染色质环^[2,3,5,11-14]。成对的反向-正向 CBS (背靠背)通常分布在拓扑相关结构域(topological-associated domain, TAD)边界,发挥绝缘子的作用,阻止增强子跨区域异位激活基因表达^[3]。CTCF 除了

能与 cohesin 一起介导染色质环,还能与其他的非编码调控元件及架构蛋白(如 YY1)共同搭建三维基因组结构^[1,15,16]。CTCF 参与调控的染色质三维空间结构是生物体发挥正常生理功能的基础,包括神经系统中原钙粘蛋白(protocadherin, *Pcdh*)基因簇启动子的选择^[4,13,14,17]、尿苷二磷酸葡萄糖醛酸转移酶 1 (UDP-glucuronosyltransferase 1, *UGT1*)的表达^[12]、免疫球蛋白(Immunoglobulin, *Ig*)和 T 细胞受体(T cell receptor, *Tcr*)的 V(D)J 重组^[18-20]、四肢发育过程中的同源框基因家族(homeobox gene family, *Hox*)控制^[21-23]以及生物学的其他方面^[1,24]。

脊椎动物和果蝇 *Hox* 基因簇内的 CTCF 结合位点经过数亿年进化仍然存在,这暗示了在两侧对称动物进化历程中 CTCF 参与 *Hox* 基因表达调控的古老起源^[25]。*Hox* 基因簇编码的转录因子调控着细胞分化的多个方面,影响生物体四肢发育、器官形成

等^[26]。在哺乳动物中, *Hox* 基因分布在四个簇中, 分别命名为 *HoxA*、*HoxB*、*HoxC* 和 *HoxD*, 根据序列同源性及在簇中的位置, *Hox* 基因分为 13 组 (*Hox1~Hox13*) (图 1A), 躯干和四肢的不同区域表达 *Hox* 基因的不同组合^[27]。在发育的最初阶段, *Hox* 基因簇不表达, 在发育过程中主要沿胚胎首尾轴 (head-to-tail axis) 呈时空线性表达^[28]。

HoxD 基因簇包括 *HoxD1*、*HoxD3*、*HoxD4*、*HoxD8~HoxD13*, 整体位于两个拓扑结构域边界, 受上游增强子簇 (upstream enhancer cluster) 和下游增强子簇 (downstream enhancer cluster) 调控 (图 1, A 和 B)。在小鼠胚胎肢芽 (limb bud) 发育早期, 位于基因簇 3' 端粒侧 TAD (telomeric domain, T-DOM) 处于激活状态, 控制 *HoxD1~HoxD11* 在四肢的近端转录^[29]; 随后, 在胚胎肢芽 (limb bud) 远端细胞中, 位于 5' 中心粒侧 TAD (centromeric domain, C-DOM) 处于激活状态, 负责 *HoxD9~HoxD13* 在手指细胞中的表达^[30,31]。在小鼠胚胎肢芽中, 上游增强子簇中的 5 个增强子 (*Island1~5*) 包含许多保守的非编码调控元件, 同时富集有 H3K4me1、H3K27ac (活性增强子标记), 与 *HoxD10~HoxD13* 有很强的互作^[31]。删除部分增强子只会影响它们与 *HoxD10~HoxD13* 基因的互作, 但对 *HoxD10~HoxD13* 基因的表达影响不大; 当完全删除上游增强子簇, *HoxD10~HoxD13* 基因表达才会完全消失。这些研究表明这些 *Island* 之间可能存在冗余性和互补性^[31]。在小鼠生殖器结节 (genital tubercle, GT) 中, 上游增强子簇中的 GT1 和 GT2 与 *HoxD13* 启动子也有远距离 DNA 相互作用, 但 GT2 仅在生殖器中表现出较强的活性, 在肢端细胞中不能显著激活 *HoxD* 基因转录^[32]。在小鼠肢芽近端细胞中, 下游增强子簇中的 CS38-41 (conserved sequence element 38-41) 参与调控 *HoxD1~HoxD11* 的表达^[21]。这些上下游增强子簇通过转录因子或架构蛋白介导的染色质环与 *HoxD* 基因启动子实现远程互作, 从而完成特定的 *HoxD* 基因调控^[33~35]。然而, 目前对于 *HoxD* 基因簇调控机制的研究主要借助于在小鼠中对其调控区域或者基因进行大片段删除或反转^[21,22,31,32,36], 对于人细胞中 *HOXD* 基因簇调控机制的研究较少。

本研究以人 *HOXD* 基因簇为模式基因, 探索 CTCF 对发育基因 (developmental genes) 转录调控的

影响。利用 CRISPR DNA 片段编辑系统^[37~40]在人 HEK293T 细胞中获得一系列的串联反向 CBS 删除的单细胞克隆株。进行 RNA-seq 实验检测 CBS 删除后 *HOXD* 基因表达水平的变化, 最后通过定量高分辨率染色体构象捕获 (quantitative high resolution chromosome conformation capture copy, QHR-4C)^[13] 进一步观测 CBS 删除对 *HOXD* 基因簇三维基因组结构的影响, 发现成串反向排列的 CTCF 位点具有平衡上下游增强子簇对 *HOXD* 基因簇转录调控的作用, 为揭示哺乳动物 *HoxD* 基因转录的复杂调控提供参考。

1 材料与方法

1.1 实验材料

人胚胎肾细胞系 HEK293T 从中国科学院细胞库购买; 胰酶、青霉素/链霉素双抗购自美国 Gibco 公司; DMEM 培养基购自美国 HyClone 公司; 胎牛血清购自南美 ExCell 公司; Lipofectamine 3000 购自美国 Invitrogen 公司; pcDNA3.1-Cas9-WT 质粒来源于北京大学席建忠教授; *Bsa* I 限制性内切酶、T4 DNA 连接酶、*Dpn* II 限制性核酸内切酶、RNA 文库制备试剂盒、Q5 Hot Start HiFi PCR Master Mix 均从美国 NEB 公司采购; pGL3-U6-sgRNA-Puromycin-*Bsa* I 质粒由上海科技大学黄行许教授赠予; 质粒小量提取试剂盒、DNA 凝胶回收试剂盒均从美国 AXYGEn 公司采购; MinElute Gel Extraction kit 从德国 QIAGEN 公司采购; Green Taq Mix、dNTPs、高保真 Phanta DNA 聚合酶购自南京诺唯赞生物科技有限公司; pClone007 Versatile Simple Vector 购自北京擎科新业生物技术有限公司; sgRNA、测序引物由上海生物工程 (上海) 股份有限公司合成; Triton X-100、10% SDS、DNA 提取液、NP40 等从美国 SIGMA 公司采购; 三氯甲烷、硼酸和乙醇购自上海沪试实验室器材股份有限公司; 异丙醇和氯化钠购自上海凌峰化学试剂有限公司; 链霉素亲和素磁珠从美国 Invitrogen 公司采购; PCR 文库纯化试剂盒从瑞士 Roche 公司采购; 甲醛溶液、RNA 酶 A、糖原等均从美国 Thermo 公司采购; TRIzol Reagent 从美国 Ambion 公司采购; AMPure XP Beads 从美国 Beckman Coulter 公司采购; CTCF 抗体从英国 Abcam 公司采购。

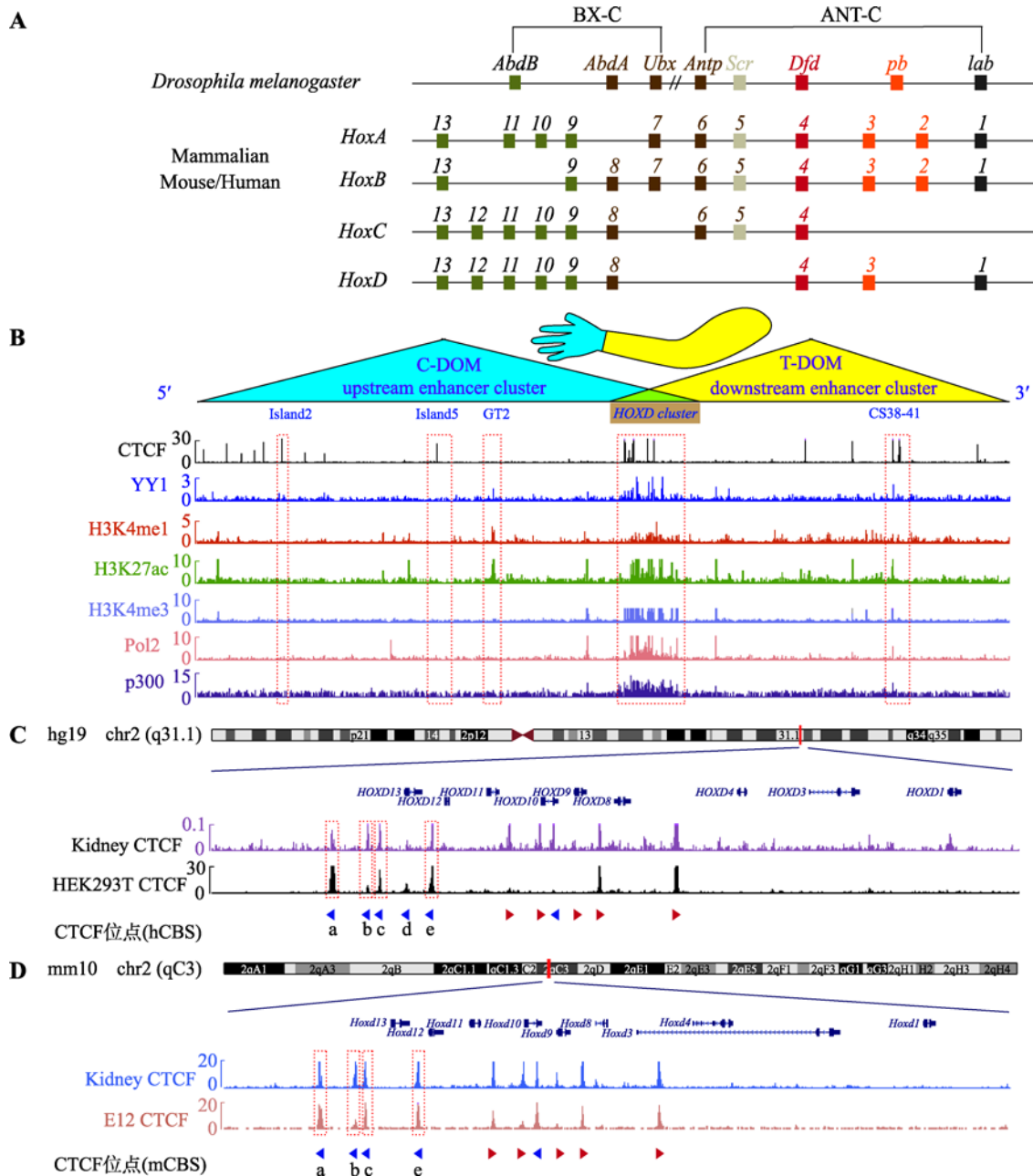


图 1 CTCF 位点在 *Hox* 基因簇及其调控区域内的分布

Fig. 1 Distribution of CTCF sites in the *Hox* cluster and its regulatory landscapes

A: 果蝇及哺乳动物 *Hox* 基因家族的基因组结构。果蝇 *Hox* 基因簇分为触足复合群(ANT-C)和双胸复合群(BX-C)，其中 ANT-C 包含 *Antp*、*Scr*、*Dfd*、*pb* 和 *lab* 基因，BX-C 包含 *AbdB*、*AbdA* 和 *Ubx* 基因。哺乳动物具有 4 个 *Hox* 基因簇：*HoxA*、*HoxB*、*HoxC* 和 *HoxD*，共包含 39 个基因，其中 *HoxD* 基因簇包括 *HoxD1*、*HoxD3*、*HoxD4*、*HoxD8*~*HoxD13*。果蝇及哺乳动物相同颜色的 *Hox* 基因为直系同源基因(orthologues)，它们起源于同一祖先基因。B: 人胚胎肾细胞系 HEK293T 中的 CTCF 蛋白、架构蛋白 YY1、增强子标记 H3K4me1 和 H3K27ac、启动子标记 H3K4me3、转录活性标记 PolII 和 p300 在人 *HOXD* 基因簇及其调控区域内的分布。红色虚线框指示增强子 Island2、Island5、GT2、CS38-41 和 *HOXD* 基因簇所在区域。哺乳动物的 *HoxD* 基因簇位于 3'端粒侧 TAD (T-DOM)和 5'中心粒侧 TAD (C-DOM)交界处，并受到上下游增强子簇(upstream enhancer cluster 和 downstream enhancer cluster)的调控。C: 人类 *HOXD* 基因簇区域 CTCF 位点的分布。人肾脏和胚胎肾细胞系 HEK293T 的 CTCF ChIP-seq 结合峰分布图：*HOXD* 基因簇 C-DOM 和 T-DOM 交界区域均具有串联排列的 CTCF 位点。红色虚线框依次指示 CBS a、b、c 和 e 所在区域。D: 小鼠 *HoxD* 基因簇区域 CTCF 位点的分布。小鼠肾脏和第 12.5 天胚胎肢芽 CTCF ChIP-seq 数据显示小鼠 *HoxD* 基因簇中心粒侧对应的位置具有串联反向排列的 4 个 CTCF 位点。图 C 和 D 中箭头代表 CTCF 位点，其中红色箭头代表正向 CTCF 位点，蓝色箭头代表反向 CTCF 位点。

1.2 细胞培养

HEK293T 细胞的培养基为 89% DMEM 完全培养液、10%胎牛血清和 1%青霉素/链霉素双抗的混合液, 在 37℃、5% CO₂ 的恒温培养箱中培养。

1.3 HEK293T 细胞 CTCF 的 ChIP-seq

收集 2×10⁷ 细胞后, 首先用 1%甲醛交联细胞, 再加入终浓度为 0.125 mol/L 的甘氨酸溶液终止反应, 然后用预冷的 ChIP 缓冲液裂解细胞两次(裂解条件均为 4℃, 缓慢旋转 10 min), 离心后再用预冷的 ChIP 缓冲液重悬细胞, 冰上孵育 10 min。用非接触式超声仪进行超声, 凝胶电泳鉴定超声后 DNA 片段长度为 100~10,000 bp。将细胞悬液离心后的上清转移至新的离心管中并加入 50 μL 的 protein A/G 磁珠孵育 2 h。用磁力架弃去磁珠, 将上清转移至新的离心管

中加入 4 μg 的 CTCF 抗体, 4℃缓慢旋转孵育过夜。第二天加入 50 μL 的 protein A/G 磁珠, 4℃缓慢旋转孵育 3 h。然后用 ChIP 缓冲液、高盐缓冲液、无盐缓冲液、LiCl 缓冲液清洗磁珠(每次 4℃缓慢旋转孵育 10 min), 最后用洗脱缓冲液洗脱 DNA。含有 DNA 的溶液经过 RNA 酶 A (37℃震荡孵育 2 h)、蛋白酶 K (55℃静置孵育 2 h)及酚氯仿试剂纯化后, DNA 沉淀用无核酸酶的水溶解, 最后用诺唯赞公司的 ChIP-seq 试剂盒 ND607-02 构建 DNA 文库。文库通过苏州金唯智生物科技有限公司的 Illumina HiSeq 平台进行测序, 根据 index 序列(P7-index 序列见表 1)对数据进行拆分, 通过 Bowtie2 比对到 GRCh37/hg19 基因组, 用 BamCoverage 进行归一化处理, 标准化到 RPKM (reads per kilobase per million mapped reads), 共做两个重复, 每个样品文库测序约 2000 万条序列。

表 1 引物序列

Table 1 Primers used in this study

类型	引物名称	序列(5'→3')
PCR	a1F	TTCCAGCACCTCGGCTTTGTC
	a1R	CCCACCTTCCACCTCTGTCCTG
	b1F	GTCCGCCCCGTGAGCTTCTGAA
	b1R1	CTCACAGCAGCCGAAACCG
	c1F	TGATGCAGCCTCTGTGACCG
	c1R	AGTTTTCCCGTGGCGTCTGA
	e1F	TTCCCTGTCCCAGCTTGATTTTC
	e1R	TCAACAGTGAAGGGCGGTGC
	e2F	CAAGCCACTCTCCCGCCACTA
	e2R	TCGCTCTCGTCCTCTCTTGGG
	e2R1	GGCTCCTGCACTGAGACCACA
sgRNA	sgRNA a1F	<u>ACCG</u> GAAGAGTGCGGGAGAACGG
	sgRNA a1R	<u>AAAC</u> CGTTCTCCCGCACTCTTCG
	sgRNA b1F	<u>ACCG</u> GGCGCATCAGGAATGTAAG
	sgRNA b1R	<u>AAAC</u> CTTACATTCTGATGCGCC
	sgRNA c1F	<u>ACCG</u> CAGGCGAAGTGCGGTTTCCA
	sgRNA c1R	<u>AAAC</u> TGGAACCGCACTTCGCCTG
	sgRNA e1F	<u>ACCG</u> AACTGTGCTCAAACGCTCTC
	sgRNA e1R	<u>AAAC</u> GAGAGCGTTTGAGCACAGTT
	sgRNA e2F	<u>ACCG</u> GAGGCGCAAAACAGCTGTTGT
	sgRNA e2R	<u>AAAC</u> ACAACAGCTGTTTGCGCCTC

续表

类型	引物名称	序列(5'→3')
Index	Island5-bioprimer	5'biotin-AAACACAAATGCATCAACCTG
	GT2-bioprimer	5'biotin-GAGCCAAACTGTACCCCTAGC
	<i>HOXD9</i> -bioprimer	5'biotin-ACCGACTAGTTCGCAGGCT
	Island5-P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCATCTCA TGAAGCTGGCATCT
	GT2-P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAACCTTTAG CTAAACCAAGGCCT
	<i>HOXD9</i> -P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGCAGCCTC CACCATTG
	P7-index-1	CAAGCAGAAGACGGCATACGAGAT <u>CGAGTAAT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-2	CAAGCAGAAGACGGCATACGAGAT <u>TCTCCGGA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-3	CAAGCAGAAGACGGCATACGAGAT <u>AATGAGCG</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-4	CAAGCAGAAGACGGCATACGAGAT <u>GGAATCTC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-5	CAAGCAGAAGACGGCATACGAGAT <u>TTCTGAAT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-6	CAAGCAGAAGACGGCATACGAGAT <u>ACGAATTC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-7	CAAGCAGAAGACGGCATACGAGAT <u>AGCTTCAG</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-8	CAAGCAGAAGACGGCATACGAGAT <u>GCGCATT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-9	CAAGCAGAAGACGGCATACGAGAT <u>CATAGCCG</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-10	CAAGCAGAAGACGGCATACGAGAT <u>TTCCGCGA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-11	CAAGCAGAAGACGGCATACGAGAT <u>GCGCGAGA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-12	CAAGCAGAAGACGGCATACGAGAT <u>CTATCGCT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-13	CAAGCAGAAGACGGCATACGAGAT <u>AGAGTACT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-14	CAAGCAGAAGACGGCATACGAGAT <u>GCTCCGTA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-15	CAAGCAGAAGACGGCATACGAGAT <u>CATGAGAG</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-16	CAAGCAGAAGACGGCATACGAGAT <u>TGAATCGC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-17	CAAGCAGAAGACGGCATACGAGAT <u>GTCTGACT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-18	CAAGCAGAAGACGGCATACGAGAT <u>CTGAATGC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-19	CAAGCAGAAGACGGCATACGAGAT <u>CGCTTCTG</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-20	CAAGCAGAAGACGGCATACGAGAT <u>TCGCATGA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-21	CAAGCAGAAGACGGCATACGAGAT <u>AATAGCAG</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-22	CAAGCAGAAGACGGCATACGAGAT <u>GTCCGCTA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-23	CAAGCAGAAGACGGCATACGAGAT <u>ACGCGATA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-24	CAAGCAGAAGACGGCATACGAGAT <u>TGATCGAT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-25	CAAGCAGAAGACGGCATACGAGAT <u>CCGCATGA</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-26	CAAGCAGAAGACGGCATACGAGAT <u>CCACAATC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-27	CAAGCAGAAGACGGCATACGAGAT <u>GATGTTCC</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

sgRNA 引物中带下划线部分为构建质粒所需的粘性末端。P7-index 中粗体带下划线部分为 index。

1.4 HEK293T 细胞中组蛋白修饰、转录活性标志及 CTCF 的 ChIP-seq 数据分析

人胚胎肾细胞系 HEK293T 的 H3K4me1 和

H3K27ac ChIP-seq 数据来源于 ENCODE 数据库, 数据编号分别为 ENCSR000FCG 和 ENCSR000FCH; YY1、H3K4me3、Pol II、p300 来源于 GEO, 数据编号分别为 GSM3636215、GSM945288、GSM935534、

GSM1239071; 人肾脏 CTCF ChIP-seq 数据的 GEO 编号为 GSM1006886; 小鼠肾脏及第 12.5 天胚胎肢芽 CTCF ChIP-seq 数据的 GEO 编号分别为 GSE91529、GSM4665698; 将数据分别比对到 GRCh37/hg19 及 GRCm38/mm10 基因组进行分析。

1.5 构建 pGL3-U6-sgRNA 质粒

按照表 1 合成单链 sgRNA (共 10 条, 分别合成双链 sgRNA a1、sgRNA b1、sgRNA c1、sgRNA e1 和 sgRNA e2), 按照如下体系(总体积为 20 μ L,)进行退火步骤: 2 μ L sgRNA-F, 2 μ L sgRNA-R, 2 μ L 10 \times NEB Buffer 2, 14 μ L ddH₂O; 退火条件: 95 $^{\circ}$ C 预变性 5 min, 缓慢降温(从第二个循环开始, 每个循环 20 s、降温 0.2 $^{\circ}$ C, 共 351 个循环)。环状 pGL3-U6-sgRNA-Puromycin-Bsa I 质粒经过 Bsa I 限制性核酸内切酶处理后, 得到线性的 pGL3-U6-sgRNA-Puromycin 载体。将退火后的双链 sgRNA 引物与线性化的 pGL3-U6-sgRNA-Puromycin 载体在室温进行连接, 经感受态细胞转化培养后, 挑取单菌落进行培养(每种 sgRNA 至少挑取 3 个单菌落), 用质粒小量提取试剂盒纯化质粒并测序。

1.6 细胞转染与单克隆化

待 12 孔板内的 HEK293T 细胞汇合度达到 70%~90% 左右时进行转染, 一个离心管中加入 Lipofectamine 3000 与 MEM 混合均匀室温静置 5 min, 另一个管中加入 Cas9 质粒、两种 sgRNA 质粒、p3000 试剂和 MEM 培养基混合均匀后室温静置 5 min (此步骤中 sgRNA 组合: sgRNA a1/sRNA e2 或 sgRNA b1/sRNA e2 或 sgRNA c1/sRNA e2 或 sgRNA e1/sRNA e2), 然后将 Lipofectamine 3000 加入其中, 混合均匀后室温孵育约 15 min, 滴入培养细胞中。

待转染时间达到 48 h 后, 换用含有 2 μ g/mL 嘌呤霉素的培养基对转染细胞进行药物筛选。药物筛选大约 4 天后, 更换为 DMEM 完全培养基继续培养 2 天, 然后用胰酶消化收集适量细胞至 PCR 管中, 加入 10 μ L 碱裂解液, 混匀置于 PCR 仪中, 98 $^{\circ}$ C 30 min 使细胞裂解, 再加入 10 μ L 中和液, 获得细胞基因组。在 sgRNA 上下游位置设计 PCR 引物, 鉴定转染后的细胞中是否含有目的片段删除的细胞, 若存在, 则对 12 孔板中剩余的转染细胞进行计

数, 用培养基稀释至浓度约为 1 个细胞每 100 μ L 培养液, 转移至 96 孔板中培养。待细胞生长一周左右, 显微镜下查看单一细胞团并做好标记。sgRNA 和 PCR 引物序列见表 1。

1.7 单细胞克隆株鉴定

将标记的单细胞克隆株(共 1056 个)用胰酶消化, 转移适量细胞到含有 10 μ L 碱裂解液的 PCR 小管中进行裂解反应获取细胞基因组。两个 sgRNA 对大片 DNA 进行删除时, 会造成两个切口, 需要通过特异性引物鉴定排除片段反转的情况^[37,41]。通过 PCR 反应(PCR 引物序列见表 1)及 DNA 琼脂糖凝胶电泳实验挑选出目的片段删除的纯合子单细胞克隆株, 将细胞目的片段删除型的扩增产物通过凝胶电泳分离, 切胶回收并送测。若测序结果为双峰, 则需要进行 TA 克隆测序鉴定单细胞克隆株的基因型。

按照 TA 克隆产品推荐用量将 PCR 产物与 pClone007 Versatile Simple Vector 混合均匀, 室温孵育约 10 min; 在离心管中将连接产物与 50 μ L 感受态细胞混合均匀, 冰上静置 30 min 后, 进行 42 $^{\circ}$ C 热激 45 s, 然后迅速转移至冰上, 2 min 后加入 500 μ L 的 LB 无抗培养基, 然后将离心管放在 37 $^{\circ}$ C 摇床上培养 1 h。离心后去除 450 μ L 上清, 重悬细胞后均匀涂布在含有氨苄抗性的 LB 固体培养基上, 37 $^{\circ}$ C 恒温箱中培养约 16 h, 次日挑取 8~10 个单菌落测序, 然后将桑格测序结果与目的片段序列比对, 确定单细胞克隆株基因型。最终获得 CBS e 删除、CBS c-e 删除、CBS b-e 删除和 CBS a-e 删除的单细胞克隆株各两株。

1.8 RNA-seq 实验

待 12 孔板中处于对数生长期的 HEK293T 细胞、单细胞克隆株汇合度达到 90% 时, 先用 1 \times PBS 清洗 2 遍, 然后加入 500 μ L/孔 TRIzol 溶液处理 15 min, 然后转移至无核酸酶的离心管中, 用三氯甲烷、异丙醇、乙醇提取纯化总 RNA。根据试剂说明书进行如下操作: 使用 Poly(A) mRNA Magnetic Isolation Module 试剂盒提取 mRNA, 利用磁珠对 mRNA 片段进行分离与片段化; 通过 PCR 反应先进行 cDNA 第一链合成, 再进行 cDNA 第二链合成反应; 双链 cDNA 连接 Adaptor 之后, 用 AMPure XP Beads

和新鲜配制的 80% 的乙醇对 DNA 进行纯化, 进行文库扩增反应; 扩增后产物用 AMPure XP Beads 和新鲜配制的 80% 的乙醇进行纯化, 最后用无核酸酶的水洗脱 DNA 文库。用 Qubit3 Fluorometer 测量终产物浓度后送至苏州金唯智生物科技有限公司的 Illumina HiSeq 平台进行测序, 根据特异性 index 序列(表 1)拆分数据, 用 STAR、cufflinks 等软件分析数据^[42]。本研究共有 30 个 RNA-seq 样品, 每个样品文库测序约 1000 万条序列。

1.9 定量高分辨率染色体构象捕获实验 QHR-4C

用 10 cm 培养皿扩增培养 HEK293T 细胞、单细胞克隆株, 待细胞汇合度达到 80%~90% 且生长良好时, 用胰酶消化离心收集细胞; 用 1% 甲醛交联细胞, 再加入终浓度为 0.125 mol/L 的甘氨酸溶液终止反应; 用预冷的 $1 \times$ PBS 重复清洗细胞, 然后配制 4C 裂解液($1 \times$ inhibitor、pH 7.5 50 mmol/L Tris-HCl、150 mmol/L NaCl、5 mmol/L EDTA、0.5% NP-40 和 1% Triton X-100)重复裂解细胞; 通过 Triton X-100、Dpn II 限制性内切酶处理使 DNA 片段化, 用 T4 DNA 连接酶捕捉相互邻近的 DNA 片段; 纯化连接后的 DNA, 然后进行超声处理(不同长度的 DNA 片段经琼脂糖凝胶电泳鉴定呈弥散状分布且主要片段长度小于 1000 bp)、PCR 线性扩增反应及接头连接反应, 最后进行 DNA 文库扩增。文库送至苏州金唯智生物科技有限公司的 Illumina HiSeq 平台进行测序, 根据 index(序列见表 1)对测序结果进行拆分, 通过 Bowtie2、Samtools、r3C-seq 等方法处理数据^[13]。本研究每个实验三个重复, 共有 78 个 4C 样品, 每个样品文库测序约 800 万条序列。

1.10 高通量测序数据信息

本研究的相关结果数据已收录在国家基因库生命大数据平台(CNGBdb)^[43]的国家基因库序列归档系统(CNSA)^[44], 项目编号: CNP0001773。

2 结果与分析

2.1 CTCF 位点在 *HOXD* 基因簇的分布高度保守

Hox 基因簇最初因引起黑腹果蝇翅膀数目改变

而被发现, 在两侧对称动物中高度保守^[45]。果蝇基因组中含有一个 *Hox* 基因簇, 由 8 个基因组成, 分为触足复合群(ANT-C)和双胸复合群(BX-C), 其中 ANT-C 包含 *Antp*、*Scr*、*Dfd*、*pb* 和 *lab* 基因, BX-C 包含 *AbdB*、*AbdA* 和 *Ubx* 基因(图 1A)。而哺乳动物基因组具有四个 *Hox* 基因簇: *HoxA*、*HoxB*、*HoxC* 和 *HoxD*, 共包括 39 个基因^[27](图 1A), 这可能是由进化过程中的两次 DNA 片段复制过程产生^[46,47]。本研究以人 HEK293T 细胞为模型, 通过分析 ChIP-seq 数据发现位于 C-DOM 和 T-DOM 边界的 *HOXD* 基因簇串联排列着多个 CTCF 位点, 提示 CTCF 可能对 *HOXD* 的转录调控发挥重要作用(图 1B)。

利用 ChIP-seq 数据进一步分析了与转录调控密切相关的转录因子和各组蛋白修饰在 *HOXD* 基因簇及其两侧调控区域内富集的情况, 包括 CTCF 蛋白、架构蛋白 YY1、增强子标记 H3K4me1、H3K27ac, 启动子标记 H3K4me3, 转录活性标记 Pol II、p300(图 1B): *HOXD* 基因簇所在的区域有大量的 CTCF 蛋白、YY1 蛋白富集以及 H3K27ac、H3K4me3 组蛋白修饰, 是一个转录活跃区域; Island2、Island5 处有 CTCF 蛋白结合峰, 但是没有相应的组蛋白修饰, 也没有转录活性标记(Pol II 和 p300), 提示 Island2、Island5 在 HEK293T 细胞中可能没有增强子活性。GT2 处没有 CTCF 蛋白结合, 有 YY1 的结合峰、H3K4me1 和 H3K27ac 等组蛋白修饰(图 1B), 说明在 HEK293T 中增强子 GT2 可能不完全依赖 CTCF 激活 *HOXD* 基因转录; CS38-41 处有 CTCF、YY1 结合以及 H3K27ac 组蛋白修饰, 同时富集 Pol II 和 p300, 因此在 HEK293T 细胞中具有较强的增强子活性。

为了分析 *HOXD* 基因簇处 CTCF 分布的保守性, 本研究比较了人肾脏和人胚胎肾细胞系 HEK293T 中 CTCF 的 ChIP-seq 结合峰分布, 发现 *HOXD* 基因簇 C-DOM 和 T-DOM 交界区域均分布着成簇排列的 CBS, 靠近 C-DOM 处具有 5 个串联排列的反向 CBS, 分别记为 CBS a、b、c、d 和 e(图 1C)。此外, 分析小鼠肾脏及第 12.5 天胚胎肢芽 CTCF 的 ChIP-seq 数据发现, 小鼠 *Hox* 基因簇中心粒侧对应的位置串联排列着类似的 CBS 簇(图 1D)。综上所述, C-DOM 和 T-DOM 边界处串联排列的反向 CBS 簇在哺乳动物中高度保守, CTCF 可能通过结合该 CBS 簇参与指导两侧调控区域内的增强子对 *HOXD* 基因簇的转

录表达调控。

2.2 边界处单个 CTCF 位点删除影响 *HOXD* 基因簇表达

为了研究反向串联排列的 CBS a、b、c、d 和 e 对人 *HOXD* 基因簇调控的作用, 首先用 CRISPR/

Cas9 编辑技术对 CBS e 进行删除。我们针对 CBS e 设计一对 sgRNA (sgRNA e1 和 sgRNA e2), Cas9 核酸酶在 sgRNA 介导下特异性识别 CBS e 两侧靶向序列后进行切割, 经过修复后能够获得 CBS e 删除的编辑细胞(图 2A)。利用 PCR 筛选单细胞克隆株(图 2B), 对获得的单细胞克隆株进行 TA 克隆并测序确

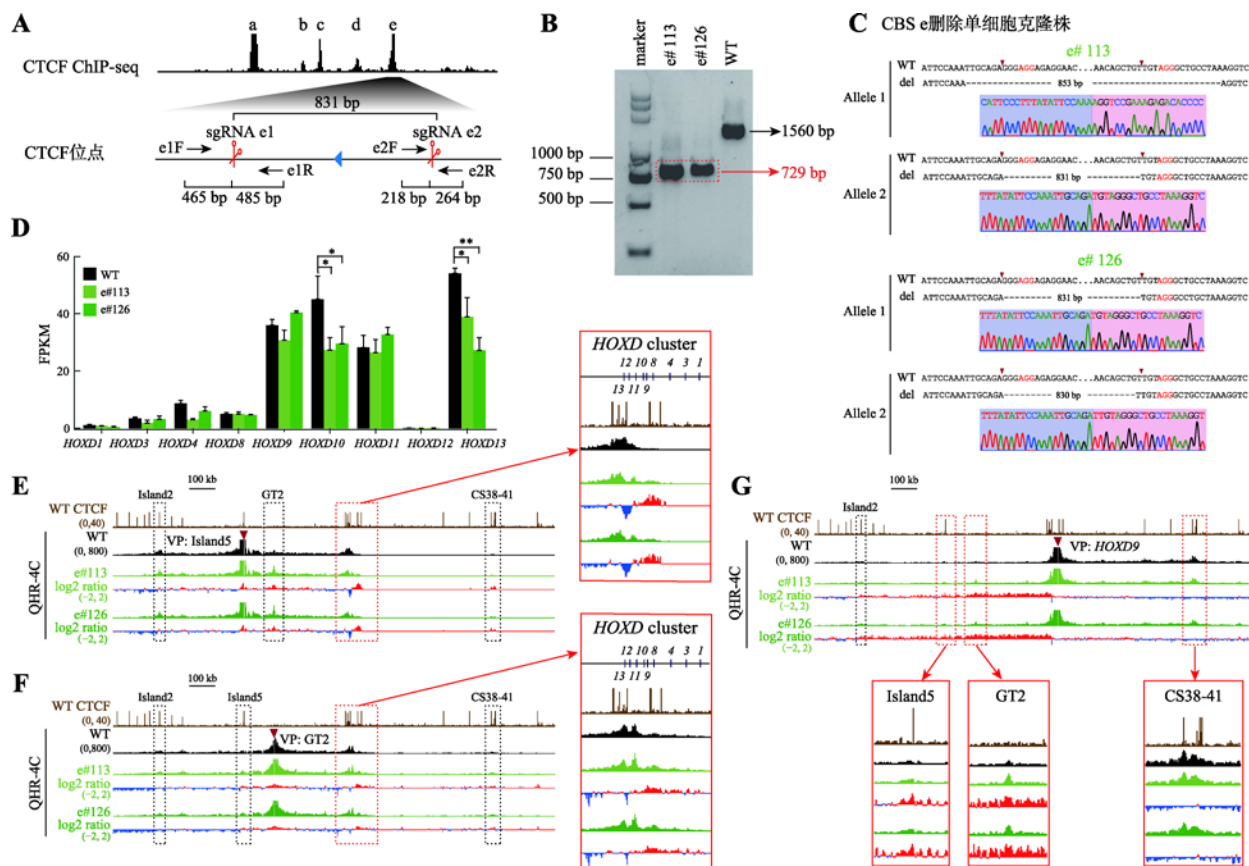


图 2 CBS e 删除改变增强子与启动子间的远程互作从而影响 *HOXD* 基因簇的基因表达

Fig. 2 Alterations of long-distance chromatin interactions between enhancers and promoters upon CBS e deletion affect gene expression of the *HOXD* cluster

A: 利用 CRISPR DNA 片段编辑技术获得 CBS e 删除的单细胞克隆株(e#113 和 e#126)的示意图。针对 CBS e 设计一对 sgRNA (sgRNA e1 和 sgRNA e2), Cas9 核酸酶在 sgRNA e1 和 sgRNA e2 的介导下特异性识别 CBS e 两侧靶向序列后进行切割, 形成的两个切口被修复后连接在一起获得 CBS e 删除的编辑细胞。B: 采用特异性引物 PCR 后进行凝胶电泳实验鉴定单细胞克隆株。引物对 e1F/e2R 在野生型(WT)细胞中扩增出 1560 bp 的片段, 在 e#113 和 e#126 细胞株中扩增出 CBS e 删除后的 729 bp 片段(红色虚线框指示目的条带)。C: TA 克隆并进行桑格测序确定单细胞克隆株的基因型。引物对 e1F/e2R 在 e#113 和 e#126 细胞株中扩增的产物经 TA 克隆后的测序结果图。D: RNA-seq 数据分析比较 WT 细胞、e#113 和 e#126 细胞株中 *HOXD* 基因簇表达水平。*: $P < 0.05$; **: $P < 0.01$; FPKM: fragments per kilobase of transcript per million mapped reads。E: 在染色质构象捕获(QHR-4C)实验中, 以增强子 Island5 为观测点(viewpoint, VP), 分析 e#113 和 e#126 细胞株中 Island5 与 *HOXD* 基因簇启动子及调控元件 Island2、GT2 和 CS38-41 的远程互作。将 e#113 和 e#126 细胞株的数据分别与 WT 进行 \log_2 处理, 红色实线框内为 Island5 与 *HOXD* 基因簇启动子之间的染色质相互作用, 黑色虚线框依次指示调控元件 Island2、GT2 和 CS38-41。F: 以增强子 GT2 为 VP, 分析 CBS e 删除后 GT2 与 *HOXD* 基因簇启动子的远程互作。红色实线框内为 GT2 与 *HOXD* 基因簇启动子间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、Island5 和 CS38-41。G: 以 *HOXD9* 启动子为 VP, 分析 e#113 和 e#126 细胞株中调控元件 Island2、Island5、GT2、CS38-41 与 *HOXD9* 启动子之间的染色质相互作用。黑色虚线框指示 Island2, 红色实线框内分别为 Island5、GT2 和 CS38-41 与 *HOXD9* 启动子之间的染色质相互作用放大图。

定其基因型(图 2C)。获得的 e#113 存在两种不同的染色体基因型, Allele 1 是 853 bp 片段删除的基因型, 其切口连接处有小片段 DNA 删除, Allele 2 是 Cas9 蛋白在 PAM 位点上游第 3 位切割后精确 DNA 修复产生的 831 bp 片段删除的基因型^[41,48,49]; e#126 也存在两种不同的染色体基因型, 其中一种是 Cas9 蛋白精准切割在 PAM 位点上游第 3 位后修复, 另一种可能是 Cas9 蛋白切割在 PAM 位点上游第 3 位和第 4 位^[48]; 基因型鉴定及测序结果均证明 e#113 和 e#126 细胞中 CBS e 已被删除(图 2C)。

为了研究单个 CBS e 删除是否会影响 *HOXD* 基因转录, 对获得的单细胞克隆株进行 RNA-seq 实验。在 HEK293T 细胞中, *HOXD9~HOXD11* 以及 *HOXD13* 表达量较高; CBS e 删除后 *HOXD10* 及 *HOXD13* 的表达水平降低最显著(图 2D)。为了探究 CBS e 删除对 *HOXD* 基因表达影响的机制, 对获得的 e#113 和 e#126 单细胞克隆株进行 QHR-4C 实验, 分别以 Island5、GT2 及 *HOXD9* 为观测点, 分析 CBS e 删除后细胞内 *HOXD* 基因簇染色质高级结构变化。在 WT 细胞中, 增强子 Island5、GT2 与 *HOXD* 基因簇之间存在显著的 DNA 相互作用(图 2, E 和 F), *HOXD9* 启动子主要与下游增强子簇中的 CS38-41 有远距离 DNA 相互作用(图 2G)。删除 CBS e 后, *HOXD8~HOXD11* 启动子与 Island5 之间相互作用有显著增加(图 2E)。除了 *HOXD12* 与 *HOXD13*, 其余 *HOXD* 基因启动子与 GT2 的相互作用略微增加(图 2F)。*HOXD9* 启动子与 Island5、GT2 的远程互作有明显增加, 与前述结果一致, 其与 CS38-41 的相互作用略微减少(图 2G)。综上所述, 单个 CBS e 删除引起 *HOXD* 基因簇启动子与增强子之间的远距离 DNA 相互作用改变, 导致 *HOXD10* 及 *HOXD13* 表达水平降低。

2.3 边界处多个 CTCF 位点删除对 *HOXD* 基因簇远距离 DNA 相互作用的影响具有叠加效应

为了探究多个 CBS 删除对 *HOXD* 基因转录水平的影响, 我们对 CBS c-e (图 3) 和 CBS b-e (图 4) 进行删除。分别设计一对 sgRNA (sgRNA c1/sgRNA e2 和 sgRNA b1/sgRNA e2) 对 HEK293T 细胞进行编辑(图

3A 和图 4A), 再用特异性引物 PCR 进行凝胶电泳鉴定(图 3B 和图 4B)。将目的条带切胶纯化、TA 克隆并进行桑格测序鉴定细胞的基因型, 最终获得了 CBS c-e 删除的 c-e#49 和 c-e#54 单细胞克隆株(图 3C) 以及 CBS b-e 删除的 b-e#59 和 b-e#80 单细胞克隆株(图 4C)。

通过 RNA-seq 实验发现, 在 CBS c-e 和 CBS b-e 删除的细胞(这些细胞的 *HOXD13* 已删除)中, *HOXD9*、*HOXD10* 及 *HOXD11* 的转录水平下降, 其中 *HOXD9* 下降最显著; *HOXD9* 的转录水平分别降低了约 40% 和 56%(图 3D 和图 4D)。QHR-4C 实验显示, 无论在 CBS c-e 还是 CBS b-e 删除的细胞中, *HOXD8~HOXD11* 与 Island5 之间的远程互作均增加(图 3E 和图 4E), 与 GT2 的互作也增加, 且幅度比 CBS e 单独删除型细胞大(图 3F 和图 4F)。在 CBS b-e 删除的细胞中, *HOXD9* 启动子与增强子 Island5 及 GT2 的远程互作相对 CBS c-e 删除型细胞进一步增强(图 3G 和图 4G), 并且在 Island5 介导的远程互作中新增了下流增强子簇(CS38-41)(图 3E 和图 4E); *HOXD9* 启动子与 CS38-41 之间的相互作用没有进一步减弱(图 3G 和图 4G)。综上所述, 边界处 CBS 删除的数量对上游增强子簇与 *HOXD* 基因簇远程互作的影响有叠加效应, 但不会递进式地影响整个 *HOXD* 基因(除了 *HOXD9*)的转录表达水平。

2.4 边界处 CTCF 位点的全部删除破坏了增强子调控 *HOXD* 基因簇表达的平衡

CTCF 常富集于 TAD 区域边界, 对维持相关拓扑结构域的稳定性具有重要作用。为了探究整个 CBS a-e 对 *HOXD* 基因簇表达调控的影响, 设计一对 sgRNA (sgRNA a1 和 sgRNA e2) 将 CBS a-e 删除(图 5A), 进行 PCR 鉴定(图 5B)、TA 克隆并进行桑格测序(图 5C)获得 a-e#93 和 a-e#108 单细胞克隆株。对 CBS a-e 删除的细胞进行 RNA-seq 实验发现, *HOXD9*、*HOXD10* 与 *HOXD11* 的基因表达显著降低, 其中 *HOXD9* 最显著, 降低了约 70% (图 5D)。QHR-4C 实验发现, Island5 与下游增强子簇(CS38-41)的互作进一步增加, 原本与 CBS a-e 整体的互作进一步转移至下游的一个反向 CBS 附近, 从而使其与 *HOXD* 基因簇内启动子的互作从 *HOXD8~HOXD13* 扩展至

HOXD4~HOXD11 (图 5E); GT2 与 *HOXD1~HOXD11* 启动子的互作增强, 同时与 T-DOM 的染色质远距离互作增强(图 5F)。以 Island5、GT2 为观测点发现一个有意思的现象: 有 CTCF 结合的 Island5 与 *HOXD* 基因簇的染色质相互作用区域具有明显的边界(图 2E、3E、4E 和 5E), 而没有 CTCF 结合的增强子 GT2 与 *HOXD* 基因簇区域的染色质相互作用呈现广泛弥散式分布(图 2F、3F、4F 和 5F), 没有明显的边界, 这提示具有 CTCF 结合的调控元件以及没有 CTCF 结合的调控元件与基因启动子之间的染色质相互作用

用模式是不同的。以 *HOXD9* 启动子为 VP 发现, 在 HEK293T 野生型细胞中, *HOXD9* 启动子主要与下游增强子簇有 DNA 相互作用; 而在 CBS a-e 删除的细胞中, *HOXD9* 启动子与上游增强子簇(Island2、Island5、GT2、GCR 和 Prox)的远程互作普遍显著增强, 与下游增强子簇(CS38-41)的远程互作普遍减弱, CBS a-e 删除改变了 *HOXD9* 的转录调控模式, 由最初主要受下游增强子簇调控转变为上下游两个增强子簇调控(图 5G), 原有的调控平衡被破坏, 这可能是导致 *HOXD9~HOXD11* 转录水平急剧降低的原因。

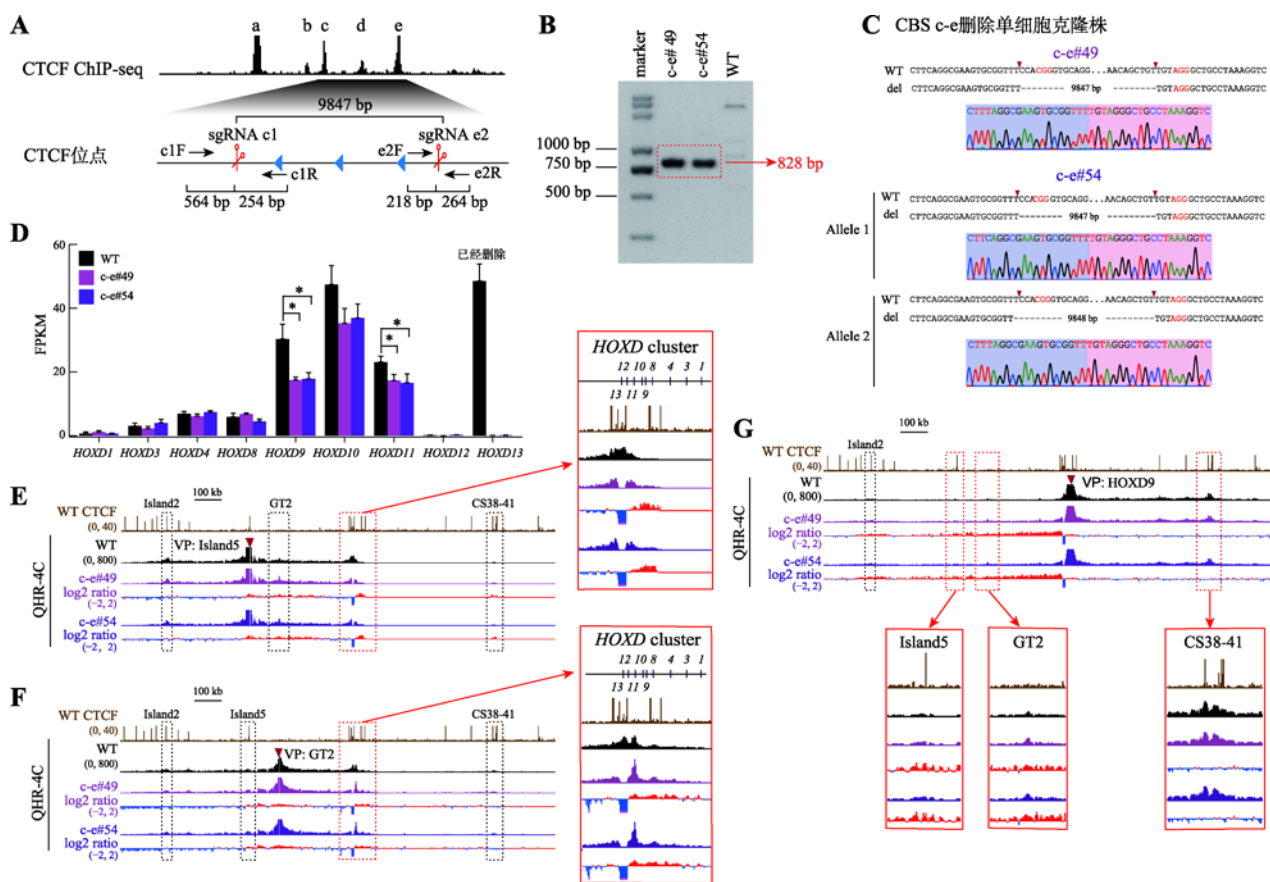


图3 CBS c-e 删除引起上游增强子与近端 *HOXD* 基因相互作用的增强

Fig. 3 Increases of long-distance chromatin interactions between enhancers and promoters upon CBS c-e deletion

A: 设计一对 sgRNA(sgRNA c1 和 sgRNA c2)对 HEK293T 野生型细胞进行编辑获得 CBS c-e 删除的单细胞克隆株(c-e#49 和 c-e#54)的示意图。B: 用引物 c1F 和 e2R PCR 鉴定 c-e#49 和 c-e#54 单细胞克隆株的凝胶电泳图。C: TA 克隆鉴定 c-e#49 和 c-e#54 单细胞克隆株基因型的测序结果图。D: RNA-seq 数据分析比较 WT 细胞、c-e#49 和 c-e#54 细胞中 *HOXD* 基因簇表达水平。E: QHR-4C 实验中, 以 Island5 为 VP, 分析 c-e#49 和 c-e#54 细胞株中 Island5 与 *HOXD* 基因簇启动子的远程互作。红色实线框内为 Island5 与 *HOXD* 基因簇启动子之间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、GT2 和 CS38-41。F: 以 GT2 为 VP, 分析 CBS c-e 删除后 GT2 与 *HOXD* 基因簇启动子的远程互作。红色实线框内为 GT2 与 *HOXD* 基因簇启动子间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、Island5 和 CS38-41。G: 以 *HOXD9* 启动子为 VP, c-e#49 和 c-e#54 细胞株中 Island2、Island5、GT2、CS38-41 与 *HOXD9* 启动子之间的染色质相互作用。黑色虚线框指示 Island2, 红色实线框内分别为 Island5、GT2 和 CS38-41 与 *HOXD9* 启动子之间的染色质相互作用放大图。

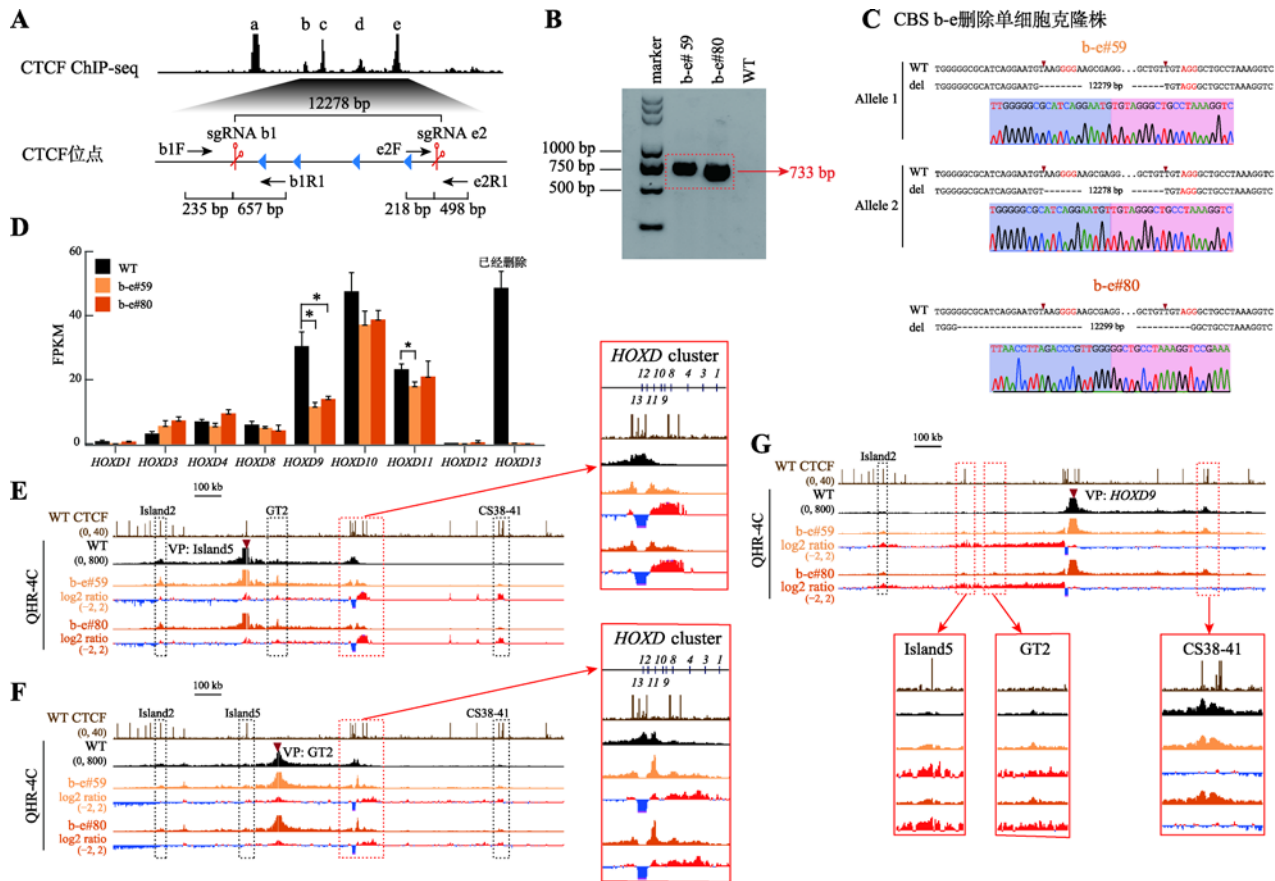


图 4 串联排列 CTCF 位点对增强子远程互作和 *HOXD* 基因表达具有叠加效应

Fig. 4 Additive effects on chromatin interactions and *HOXD* expression revealed by tandem CTCF sites deletion

A: 利用 CRISPR/Cas9 编辑系统获得 CBS b-e 删除的单细胞克隆株(b-e#59 和 b-e#80)的示意图。B: 用引物 b1F 和 e2R1 进行 PCR 鉴定 b-e#59 和 b-e#80 单细胞克隆株的凝胶电泳图。C: TA 克隆鉴定 b-e#59 和 b-e#80 单细胞克隆株基因型的测序结果图。D: RNA-seq 数据分析比较 WT 细胞、b-e#59 和 b-e#80 细胞株细胞中 *HOXD* 基因簇表达水平。E: QHR-4C 实验中, 以 Island5 为 VP, 分析 CBS b-e 删除对 Island5 与 *HOXD* 基因簇启动子远程互作的影响。红色实线框内为 Island5 与 *HOXD* 基因簇启动子之间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、GT2 和 CS38-41。F: 以 GT2 为 VP, 分析 b-e#59 和 b-e#80 细胞株中 GT2 与 *HOXD* 基因簇启动子的远程互作。红色实线框内为 GT2 与 *HOXD* 基因簇启动子间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、Island5 和 CS38-41。G: 以 *HOXD9* 启动子为 VP, 分析 b-e#59 和 b-e#80 细胞株中 Island5、GT2、CS38-41 与 *HOXD9* 启动子间的染色质相互作用。黑色虚线框指示 Island2, 红色实线框内分别为 Island5、GT2 和 CS38-41 与 *HOXD9* 启动子之间的染色质相互作用放大图。

以上结果说明, C-DOM 和 T-DOM 边界处的反向串联排列的 CBS 主要调控 *HOXD9~HOXD11* 及 *HOXD13* 的基因表达。在 HEK293T 细胞中, Island2、Island5 没有增强子活性, GT2 不完全依赖于 CTCF 调控 *HOXD9~HOXD11* 的表达, 上游增强子簇与 *HOXD* 基因簇启动子的远程互作增加, 削弱了下游增强子簇中的 CS38-41 对 *HOXD* 基因簇的激活强度。边界处单个 CBS 删除就会影响 *HOXD* 基因簇原有的三维基因组结构, 改变 *HOXD10* 和 *HOXD13* 的基因表达。多个 CBS 删除对 *HOXD* 基因簇三维基因组结

构以及 *HOXD9* 转录的影响具有叠加效应, 但是没有大幅度改变其他 *HOXD* 基因的表达。当边界处整个 CBS a-e 全部删除后, *HOXD* 基因簇与 CS38-41 的染色质相互作用进一步减弱, *HOXD9~HOXD11* 转录水平急剧降低, 并且上下游增强子簇之间出现了远距离相互作用。综上所述, *HOXD* 基因簇靠近中心粒侧区域的反向串联排列 CTCF 位点具有绝缘子功能, 通过阻断上游增强子簇与 *HOXD9~HOXD11* 的染色质远程互作, 维持 *HOXD* 基因簇表达调控的平衡, 使 *HOXD* 基因能够精准有序地表达。

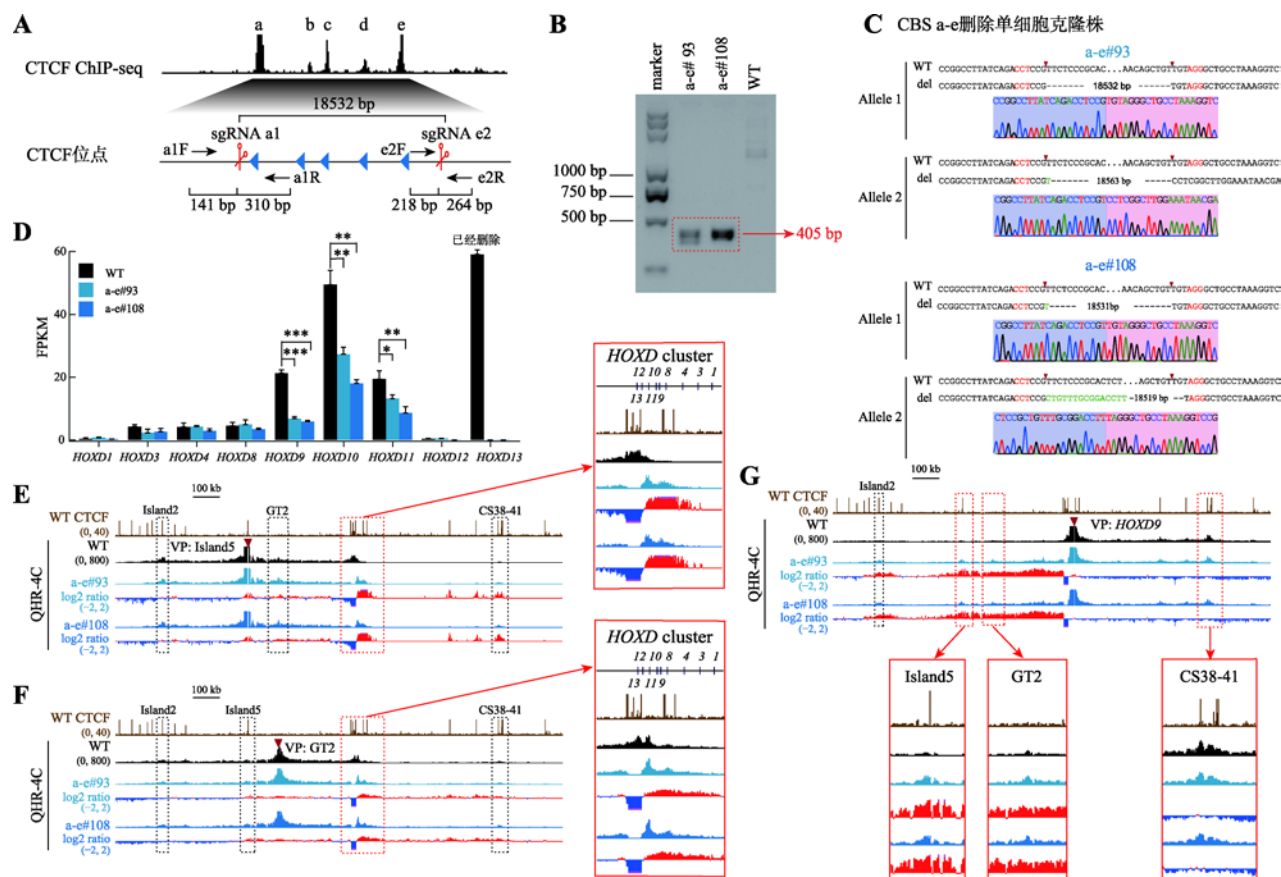


图 5 删除全部反向 CTCF 位点破坏增强子调控 *HOXD* 基因表达的平衡

Fig. 5 Regulatory balance of *HOXD* expression revealed by tandem reverse CTCF sites completely deletion

A: 获得 CBS a-e 删除的单细胞克隆株(a-e#93 和 a-e#108)的示意图。B: 用引物 a1F 和 e2R PCR 鉴定 a-e#93 和 a-e#108 细胞株的凝胶电泳图。C: TA 克隆鉴定 a-e#93 和 a-e#108 单细胞克隆株基因型的测序结果图。D: RNA-seq 数据分析比较 WT 细胞、a-e#93 和 a-e#108 细胞株细胞中 *HOXD* 基因簇转录水平。E: QHR-4C 实验中, 以 Island5 为 VP, 分析 CBS a-e 删除对 Island5 与 *HOXD* 基因簇启动子及两侧调控区域染色质相互作用的影响。红色实线框内为 Island5 与 *HOXD* 基因簇启动子之间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、GT2 和 CS38-41。F: 以 GT2 为 VP, 分析 a-e#93 和 a-e#108 细胞株中 GT2 与 *HOXD* 基因簇启动子的染色质相互作用。红色实线框内为 GT2 与 *HOXD* 基因簇启动子间的染色质相互作用放大图, 黑色虚线框依次指示调控元件 Island2、Island5 和 CS38-41。G: 以 *HOXD9* 启动子为 VP, 分析 CBS a-e 删除后 *HOXD9* 的表达调控模式。黑色虚线框指示 Island2, 红色实线框内为 Island5、GT2、CS38-41 与 *HOXD9* 基因簇启动子间的染色质相互作用放大图。

3 讨论

在哺乳动物中, 增强子的数量远远超过基因的数量, 复杂基因的表达往往受多个增强子组合调控, 防止在其适当范围之外的异位表达^[50]。增强子活性的经典模式是相对独立的、自主的并且可叠加的, 但是顺式调控区域内存在功能有冗余性(redundancy)的增强子^[51,52]。增强子激活需要结合多个转录因子^[53], 转录因子结合增强子后, 对基因表达起抑制作用或激活作用取决于结合的转录因子^[54]。例如,

转录因子 RFX5 与增强子的结合对原钙粘蛋白基因簇的表达起抑制作用^[55]。增强子活性的强弱取决于招募的共激活蛋白对染色质的修饰和增强子的重塑, 包括组蛋白乙酰基转移酶(例如 p300/CBP)、组蛋白甲基转移酶(例如 MLL3/4 和 CARM1)、染色质重塑因子(例如 Brg1 和 CHD7), 以及在启动子处促进转录的相关因子(例如中介体复合物)^[56]。本研究, CBS a-e 删除后, *HOXD* 基因簇与上游增强子簇中的 Island5、GT2 的远程互作显著增加, 而 *HOXD9*~*HOXD11* 的表达下降, 这表明 Island5、GT2 与启动子之间有互作, 但是它们没有增强子活性, 这可能

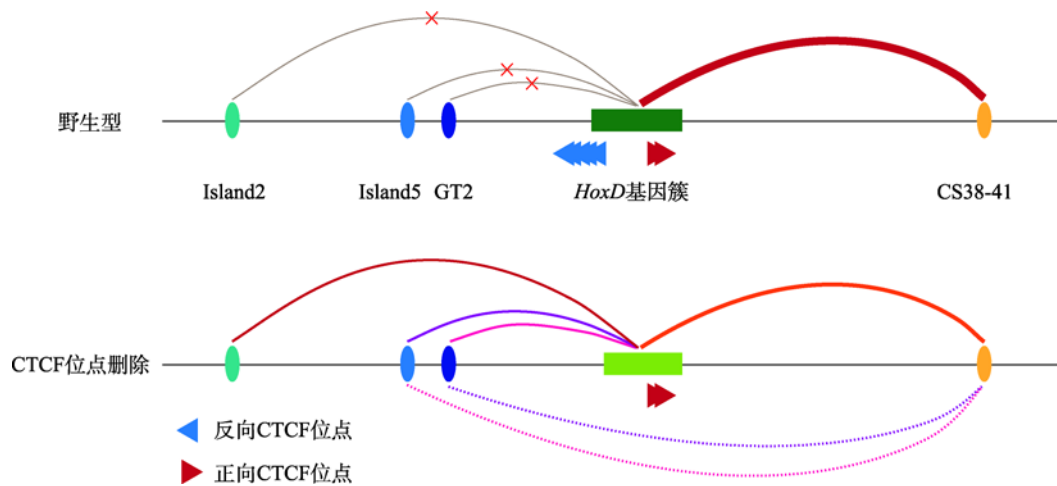


图 6 *HOXD* 基因簇的表达调控模式图

Fig. 6 A model for *HOXD* gene regulation

A: 野生型(WT)细胞中 *HOXD* 基因簇的表达调控示意图。在野生型细胞中, CTCF 位点具有绝缘子作用, 阻碍了上游增强子簇中的 Island2、Island5、GT2 与 *HOXD* 基因簇启动子的远程相互作用, *HOXD* 基因簇主要受下游增强子簇中的 CS38-41 调控。B: CTCF 位点删除的细胞中 *HOXD* 基因簇的表达调控示意图。在 CBS 删除的细胞中, CTCF 位点绝缘子作用消失, 上游增强子簇与 *HOXD* 基因簇之间产生远距离 DNA 相互作用, 导致 *HOXD* 基因簇双边调控的平衡被破坏、*HOXD* 基因表达降低。

是因为在这两个增强子上结合了抑制性的共激活蛋白或转录因子^[35,57]。小鼠胚胎发育过程中 *Hox* 基因的表达就需要结合不同的转录因子^[58]。

CTCF 与 cohesin 的结合是一个动态过程^[59], 成串的 CBS 可以阻碍渗透的 cohesin 滑动, 形成类似于“葫芦”结构的染色质环^[1,17], 由此介导的远距离增强子与启动子的相互作用可有效持久地调控基因的表达。我们先前利用小鼠和多种细胞模型对原钙粘蛋白和珠蛋白模式基因进行系统的研究^[13], 发现单个 CTCF 位点可确保适当的增强子绝缘和启动子激活, 而串联排列 CTCF 位点的绝缘作用可维持基因组三维空间的动态平衡及准确的启动子选择。在本研究中, 我们进一步利用 HEK293T 细胞模型对 *HOXD* 基因簇中的 CTCF 位点进行了系统的组合性删除实验, 发现串联排列 CTCF 位点通过其绝缘功能维持上下游增强子簇调控 *HOXD* 基因表达的平衡。另外我们观察到一个有趣的现象, 包含 CTCF 位点的增强子 Island5 与 *HOXD* 基因的染色质相互作用具有明显的边界, 而不包含 CTCF 位点的增强子 GT2 与 *HOXD* 基因的染色质相互作用区域则分布较广泛, 这和原钙粘蛋白基因簇中的增强子 HS5-1 (包含 CTCF 位点)、HS7 (不包含 CTCF 位点)与 *Pcdh* 基因染色质互作^[13]的特点一致。

在本研究中, 我们将 *HOXD* 基因簇 C-DOM 与 T-DOM 边界处 CTCF 位点进行逐步删除, 随着剩余的 CTCF 位点数量减少, *HOXD* 基因簇启动子与上游增强子簇 DNA 远程互作逐渐增强, 与下游增强子簇的 DNA 远程互作逐渐减弱。当边界串联反向的 CTCF 位点全部删除后, *HOXD9* 基因启动子由最初与下游增强子簇有相互作用转变为与上下游增强子簇均有相互作用, 并且上下游增强子簇之间出现了 DNA 远程互作(图 6)。

综上所述, *HOXD* 基因簇中串联反向 CTCF 位点扮演着绝缘子的角色, 阻碍上游增强子簇跨区域调控近端 *HOXD* 基因的表达, 对 *HOXD* 基因簇在肢体发育中时空线性表达的精细调控具有重要意义, 为深入研究 *Hox* 基因簇复杂调控网络提供了参考。

参考文献(References):

- [1] Wu Q, Liu PF, Wang LY. Many facades of CTCF unified by its coding for three-dimensional genome architecture. *J Genet Genomics*, 2020, 47(8): 407–424. [DOI]
- [2] 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 alpha promoter choice. *Proc Natl Acad Sci USA*, 2012, 109(51): 21081–

21086. [DOI]
- [3] Guo Y, Xu Q, Canzio D, Shou J, Li JH, Gorkin DU, Jung I, Wu HY, Zhai YN, 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]
- [4] Zhai YN, Xu Q, Guo Y, Wu Q. Characterization of a cluster of CTCF-binding sites in a protocadherin regulatory region. *Hereditas(Beijing)*, 2016, 38(4): 323–336. 翟亚男, 许泉, 郭亚, 吴强. 原钙粘蛋白基因簇调控区域中成簇的 CTCF 结合位点分析. *遗传*, 2016, 38(4): 323–336. [DOI]
- [5] Yin M, Wang J, Wang M, Li X, Zhang M, Wu Q, Wang Y. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res*, 2017, 27(11): 1365–1377. [DOI]
- [6] Guo Y, Wu Q. Inversion of CTCF binding sites by DNA fragment editing alters genome topology and enhancer/promoter functions. *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]
- [7] Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenko VV. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol*, 1996, 16(6): 2802–2813. [DOI]
- [8] Chen HB, Tian Y, Shu WJ, Bo XC, Wang SQ. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One*, 2012, 7(7): e41374. [DOI]
- [9] Nasmyth K. Disseminating the genome: Joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet*, 2001, 35: 673–745. [DOI]
- [10] Kim Y, Shi ZB, Zhang HS, Finkelstein IJ, Yu HT. Human cohesin compacts DNA by loop extrusion. *Science*, 2019, 366(6471): 1345–1349. [DOI]
- [11] Lu YJ, Shou J, Jia ZL, Wu YH, Li JH, Guo Y, Wu Q. Genetic evidence for asymmetric blocking of higher-order chromatin structure by CTCF/cohesin. *Protein Cell*, 2019, 10(12): 914–920. [DOI]
- [12] Zheng XF, Huang HY, Wu Q. Chromatin architectural protein CTCF regulates gene expression of the *UGT1* cluster. *Hereditas(Beijing)*, 2019, 41(6): 509–523. 郑晓飞, 黄海燕, 吴强. 染色质架构蛋白 CTCF 调控 *UGT1* 基因簇的表达. *遗传*, 2019, 41(6): 509–523. [DOI]
- [13] Jia ZL, Li JW, Ge X, Wu YH, Guo Y, Wu Q. Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol*, 2020, 21(1): 75. [DOI]
- [14] Wu YH, Jia ZL, Ge X, Wu Q. Three-dimensional genome architectural CCCTC-binding factor makes choice in duplicated enhancers at *Pcdha* locus. *Sci China Life Sci*, 2020, 63(6): 835–844. [DOI]
- [15] Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. *Cell*, 2016, 164(6): 1110–1121. [DOI]
- [16] Nichols MH, Corces VG. A tethered-inchworm model of SMC DNA translocation. *Nat Struct Mol Biol*, 2018, 25(10): 906–910. [DOI]
- [17] Wu Q, Jia ZL. Wiring the brain by clustered protocadherin neural codes. *Neurosci Bull*, 2021, 37(1): 117–131. [DOI]
- [18] Lin SG, Ba ZQ, Alt FW, Zhang Y. RAG chromatin scanning during V(D)J recombination and chromatin loop extrusion are related processes. *Adv Immunol*, 2018, 139: 93–135. [DOI]
- [19] Chen L, Carico Z, Shih HY, Krangel MS. A discrete chromatin loop in the mouse *Tcrα-Tcrδ* locus shapes the TCRδ and TCRα repertoires. *Nat Immunol*, 2015, 16(10): 1085–1093. [DOI]
- [20] Majumder K, Koues OI, Chan EAW, Kyle KE, Horowitz JE, Yang-Iott K, Bassing CH, Taniuchi I, Krangel MS, Oltz EM. Lineage-specific compaction of *Tcrβ* requires a chromatin barrier to protect the function of a long-range tethering element. *J Exp Med*, 2015, 212(1): 107–120. [DOI]
- [21] Rodríguez-Carballo E, Lopez-Delisle L, Zhan Y, Fabre PJ, Beccari L, El-Idrissi I, Huynh THN, Ozadam H, Dekker J, Duboule D. The *HoxD* cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev*, 2017, 31(22): 2264–2281. [DOI]
- [22] Rodríguez-Carballo E, Lopez-Delisle L, Yakushiji-Kaminatsui N, Ullate-Agote A, Duboule D. Impact of genome architecture on the functional activation and repression of *Hox* regulatory landscapes. *BMC Biol*, 2019, 17(1): 55. [DOI]
- [23] Rodríguez-Carballo E, Lopez-Delisle L, Willemin A, Beccari L, Gitto S, Mascres B, Duboule D. Chromatin topology and the timing of enhancer function at the *HoxD* locus. *Proc Natl Acad Sci USA*, 2020, 117(49): 31231–

31241. [DOI]
- [24] Jia ZL, Wu Q. Clustered protocadherins emerge as novel susceptibility loci for mental disorders. *Front Neurosci*, 2020, 14: 587819. [DOI]
- [25] Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci USA*, 2012, 109(43): 17507–17512. [DOI]
- [26] Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*, 1978, 276(5688): 565–570. [DOI]
- [27] Mallo M. Reassessing the role of *Hox* genes during vertebrate development and evolution. *Trends Genet*, 2018, 34(3): 209–217. [DOI]
- [28] Kmita M, Duboule D. Organizing axes in time and space; 25 years of colinear tinkering. *Science*, 2003, 301(5631): 331–333. [DOI]
- [29] Andrey G, Montavon T, Mascres B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science*, 2013, 340(6137): 1234167. [DOI]
- [30] Beccari L, Yakushiji-Kaminatsui N, Woltering JM, Necsulea A, Lonfat N, Rodríguez-Carballo E, Mascres B, Yamamoto S, Kuroiwa A, Duboule D. A role for *Hox13* proteins in the regulatory switch between TADs at the *HoxD* locus. *Genes Dev*, 2016, 30(10): 1172–1186. [DOI]
- [31] Montavon T, Soshnikova N, Mascres B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. A regulatory archipelago controls *Hox* genes transcription in digits. *Cell*, 2011, 147(5): 1132–1145. [DOI]
- [32] Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D. Convergent evolution of complex regulatory landscapes and pleiotropy at *Hox* loci. *Science*, 2014, 346(6212): 1004–1006. [DOI]
- [33] Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: Transcriptional enhancers in development and evolution. *Cell*, 2016, 167(5): 1170–1187. [DOI]
- [34] Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet*, 2019, 20(8): 437–455. [DOI]
- [35] Kim S, Shendure J. Mechanisms of interplay between transcription factors and the 3D genome. *Mol Cell*, 2019, 76(2): 306–319. [DOI]
- [36] Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. The dynamic architecture of *Hox* gene clusters. *Science*, 2011, 334(6053): 222–225. [DOI]
- [37] Li JH, Shou J, Guo Y, Tang YX, Wu YH, Jia ZL, Zhai YN, 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]
- [38] Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, Xiong JW, Xi JJ. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res*, 2013, 23(4): 465–472. [DOI]
- [39] Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 2014, 346(6213): 1258096. [DOI]
- [40] Liu PF, Wu Q. Probing 3D genome by CRISPR/Cas9. *Hereditas(Beijing)*, 2020, 42(1): 18–31.
- 刘沛峰, 吴强. CRISPR/Cas9 基因编辑在三维基因组研究中的应用. *遗传*, 2020, 42(1): 18–31. [DOI]
- [41] 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]
- [42] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with tophat and cufflinks. *Nat Protoc*, 2012, 7(3): 562–578. [DOI]
- [43] Guo XQ, Chen FZ, Gao F, Li L, Liu K, You LJ, Hua C, Yang F, Liu WL, Peng CH, Wang LN, Yang XX, Zhou FY, Tong JW, Cai J, Li ZY, Wan B, Zhang L, Yang T, Zhang MW, Yang LL, Yang YW, Zeng WJ, Wang B, Wei XF, Xu X. CNSA: A data repository for archiving omics data. *Database (Oxford)*, 2020; 2020: baaa055. [DOI]
- [44] Chen FZ, You LJ, Yang F, Wang LN, Guo XQ, Gao F, Hua C, Tan C, Fang L, Shan RQ, Zeng WJ, Wang B, Wang R, Xu X, Wei XF. CNGBdb: China National Genebank Database. *Hereditas(Beijing)*, 2020, 42(08): 799–809.
- 陈凤珍, 游丽金, 杨帆, 王丽娜, 郭学芹, 高飞, 华聪, 谈聪, 方林, 单日强, 曾文君, 王博, 王韧, 徐讯, 魏晓锋. CNGBdb: 国家基因库生命大数据平台. *遗传*, 2020, 42(8): 799–809. [DOI]
- [45] Pearson JC, Lemons D, McGinnis W. Modulating *Hox* gene functions during animal body patterning. *Nat Rev Genet*, 2005, 6(12): 893–904. [DOI]
- [46] Lonfat N, Duboule D. Structure, function and evolution of topologically associating domains (TADs) at *Hox* loci. *FEBS Lett*, 2015, 589(20): 2869–2876. [DOI]
- [47] Holland PW, Garcia-Fernández J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Dev Suppl*, 1994, 125–133. [DOI]

- [48] Shou J, Li J, Liu Y, Wu Q. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol Cell*, 2018, 71(4): 498–509 e4. [DOI]
- [49] Fu YF, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 2013, 31(9): 822–826. [DOI]
- [50] Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc Natl Acad Sci USA*, 2011, 108(33): 13570–13575. [DOI]
- [51] Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, 2010, 466(7305): 490–493. [DOI]
- [52] Barolo S. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays*, 2012, 34(2): 135–141. [DOI]
- [53] Buecker C, Wysocka J. Enhancers as information integration hubs in development: Lessons from genomics. *Trends Genet*, 2012, 28(6): 276–284. [DOI]
- [54] Jolma A, Yin YM, Nitta KR, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, 2015, 527(7578): 384–388. [DOI]
- [55] Wang N, Jia ZL, Wu Q. RFX5 regulates gene expression of the *Pcdha* cluster. *Hereditas(Beijing)*, 2020, 42(8): 760–774.
王娜, 甲芝莲, 吴强. RFX5 调控原钙粘蛋白 α 基因簇的表达. *遗传*, 2020, 42(8): 760–774. [DOI]
- [56] Malik S, Roeder RG. The metazoan mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet*, 2010, 11(11): 761–772. [DOI]
- [57] Bolt CC, Duboule D. The regulatory landscapes of developmental genes. *Development*, 2020, 147(3): dev171736. [DOI]
- [58] Neijts R, Deschamps J. At the base of colinear *Hox* gene expression: Cis-features and trans-factors orchestrating the initial phase of *Hox* cluster activation. *Dev Biol*, 2017, 428(2): 293–299. [DOI]
- [59] Xu DF, Ma RS, Zhang JH, Liu ZJ, Wu B, Peng JH, Zhai YN, Gong QG, Shi YY, Wu JH, Wu Q, Zhang ZY, Ruan K. Dynamic nature of CTCF tandem 11 zinc fingers in multivalent recognition of DNA as revealed by NMR spectroscopy. *J Phys Chem Lett*, 2018, 9(14): 4020–4028. [DOI]

(责任编辑: 李大力)