

HOXD 基因簇内一系列 CTCF 位点反转揭示绝缘子功能

何象龙, 李金环, 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: CTCF (CCCTC-binding factor) 是一种重要的染色质架构蛋白, 其与绝缘子的方向性结合在哺乳动物基因组三维空间结构形成和维持中起着至关重要的作用。正向-反向相对方向的 CTCF 结合位点(简称 CTCF 位点)可以在染色质黏连蛋白(cohesin)的协助下, 形成染色质环, 介导远距离 DNA 元件之间的相互作用; 而在染色质拓扑结构域边界区域的 CTCF 位点呈现反向-正向相背方向分布, 发挥绝缘子的功能。为进一步研究 CTCF 介导染色质环的形成与其绝缘功能之间的关系, 本研究采用 DNA 片段编辑方法通过设计成对 sgRNA (dual sgRNA)构建了一系列 HOXD 基因簇区域 CTCF 位点反转的单细胞克隆。定量高分辨率染色质构象捕获实验显示边界区域 CTCF 位点反转会改变原有的染色质环方向, 通过环挤出模型(loop extrusion)形成新的染色质环, 引起染色质拓扑结构域边界漂移至新形成的一对反向-正向 CTCF 位点处。此外, 串联排列的 CTCF 位点可以通过阻碍反方向渗透的黏连蛋白继续滑动发挥绝缘子的功能。RNA-seq 实验发现 CTCF 位点反转引起的局部基因组空间结构变化会进一步影响基因的表达。上述研究表明相邻两个染色质拓扑结构域边界区域的反向-正向 CTCF 位点可以通过与各自所在拓扑结构域内相向的 CTCF 位点形成染色质环, 阻碍黏连蛋白滑动, 该发现为进一步研究 CTCF 的绝缘功能和其对基因组拓扑结构的影响提供了参考。

关键词: CTCF; 染色质环; HOXD; 绝缘子; CTCF 位点反转

Combinatorial CRISPR inversions of CTCF sites in HOXD cluster reveal complex insulator function

Xianglong He, Jinhuan Li, 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: CTCF (CCCTC-binding factor) is a zinc-finger protein which plays a vital role in the three-dimensional (3D) genome architecture. A pair of forward-reverse convergent CTCF binding sites (CBS elements) mediates long-distance DNA interactions to form chromatin loops with the assistance of the cohesin complex, while CBS elements at the chromatin

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 Nature Science Foundation of China (Nos. 31800636, 31630039, 91940303) and Science and Technology Commission of Shanghai Municipality program (No. 19JC1412500)]

作者简介: 何象龙, 在读硕士研究生, 专业方向: 生物学。E-mail: esroom@163.com

通讯作者: 吴强, 博士, 教授, 研究方向: 三维基因组架构与功能。E-mail: qwu123@gmail.com

DOI: 10.16288/j.ycz.21-131

网络出版时间: 2021/7/7 08:47:52

URI: <https://kns.cnki.net/kcms/detail/11.1913.r.20210706.1111.001.html>

domain boundaries show reverse-forward divergent patterns and function as insulators to discriminate against DNA interaction between chromatin domains. However, there are still many unresolved problems regarding CTCF-mediated insulation function. In order to study the connections between chromatin loops and the insulation function of CBS elements, we combinatorically inverted CBS elements at the *HOXD* locus by using CRISPR/Cas9 DNA-fragment editing methods in the HEK293T cell line and obtained five different kinds of single-cell CRISPR clones. By performing quantitative high-resolution chromosome conformation capture copy (QHR-4C) experiments, we found that boundary CBS inversions abolish original chromatin loops and establish new loops from the opposite direction, thus shifting the insulator boundary to the new divergent CTCF sites. Furthermore, tandem CBS elements block cohesin permeated from the opposite orientation to function as insulators. RNA-seq experiments showed that alterations of local three-dimensional genome architecture would further influence gene expression of the *HOXD* cluster. In conclusion, a pair of divergent CBS elements function as insulators by forming chromatin loops within chromatin domains to block cohesin sliding.

Keywords: CTCF; chromatin loop; *HOXD*; insulator; CBS inversion

大约 2 m 长的人类基因组通过多次盘曲折叠, 最终被包裹在直径约 5 μm 的细胞核中, 形成了复杂的染色质三维空间结构^[1~3]。特定的染色质空间结构对于 DNA 复制^[4]、基因转录调控^[5,6]和 DNA 修复^[7,8]等生物学过程有着重要作用。早期通过光学显微镜研究发现, 有丝分裂间期染色体有规律的分布在细胞核的不同区域, 被称为染色质领地(chromosome territories)^[9]; 随着染色质构象捕获(chromosome conformation capture, 3C)^[10]、环状染色质构象捕获(circular chromosome conformation capture, 4C)^[11]、Hi-C^[12]和配对末端标签测序分析染色质相互作用(chromatin interaction analysis by paired-end tag sequencing, ChIA-PET)^[13]等技术的发展, A/B 隔室(compartment)、拓扑相关结构域(topologically associated domain, TAD)和染色质环(chromatin loop)等概念被相继提出^[2,12,14~18], 促进人类从多个层次对染色质高级结构的认识。基于 Hi-C 数据, 基因组上平均大小约为 1 Mb、内部 DNA 相互作用紧密的区域, 被定义为 TAD^[13,14]。在人的不同细胞系中, TAD 表现出较好的保守性^[15]。通常认为在 TAD 内部, 增强子(enhancer)、绝缘子(insulator)与沉默子(silencer)等顺式作用元件可以通过 CTCF(CCCTC-binding factor)和 YY1 等染色质架构蛋白介导的染色质环与基因的启动子(promoter)发生相互作用^[5,19~22], 来调控基因的表达。

CTCF 蛋白是一种参与染色质折叠的重要架构蛋白。CTCF 蛋白最早发现于对鸡(*Gallus gallus*)

c-myc 基因的研究, 由于其与 DNA 结合的序列含有高度保守的 CCCTC 序列, 因而被称为 CCCTC 结合因子^[23]。CTCF 蛋白通过 11 个串联的锌指结构与人类基因组上成千上万的绝缘子调控元件结合^[19,24~27]。在哺乳动物中, CTCF 结合位点(CTCF-binding site, CBS)是最主要的绝缘子调控元件, 包含 4 个模块(module), 即 module 1~4, 其具有方向性, module 1~4 的方向是正向, module 4~1 的方向则为反向^[5]。CTCF 在黏连蛋白的协助下, 通过环挤出模型介导远距离的 DNA 相互作用^[16,18]。在 TAD 边界处分布着大量反向-正向的 CBS, 通过阻断 cohesin 的滑动, 起到绝缘子的功能^[2,5,28~30]。删除 CBS 导致 CTCF 无法结合, 会破坏原有 DNA 之间的染色质环, 引起局部基因组构象的改变^[28,31~33]。反之, 插入异位 CBS 会形成新的染色质环, 同时作为绝缘子阻碍原有染色质环的形成^[34,35]。在 *Pchda* 基因簇的增强子和启动子之间插入正向或者反向的 CBS, 都可以起到绝缘子的作用, 显示绝缘子没有方向性^[34]。在原钙粘蛋白(protoadherin, *Pcdh*)、免疫球蛋白(immunoglobulin, *Ig*)、*HoxD* 等基因簇存在着成串同向的 CBS, 这些 CBS 均有机会形成染色质环^[5,36,37], 显示正向-反向 CBS 对 cohesin 的阻碍作用是有渗透性的, 这可能是由于 CTCF 与 DNA 的动态结合^[34,38]。然而, 当前对于成串 CBS 绝缘子功能的研究主要侧重于启动子与增强子的相互作用, 在染色质拓扑结构域边界处的成串 CBS 的绝缘子功能还缺少系统性的研究。

HoxD 基因簇是控制四肢发育时空顺序的重要

基因簇, 被广泛应用于对基因组三维空间结构的研究^[19,39]。*HoxD* 基因簇区域存在一系列的 CBS, 与两侧调控区域内的 CBS 通过环挤压模型分别形成两个 TAD^[19,39~42]。*HoxD* 基因有序的表达受到两侧 TAD 内增强子簇的调控, 其表达分为两个阶段; 在四肢发育早期, 3'端粒侧 TAD (telomeric domain, T-DOM) 内由 CS38-41 (conserved sequence elements 38-41) 和 CS65 (conserved sequence element 65) 组成的下游增强子簇或超级增强子(downstream enhancer cluster or super-enhancer)共同调节 *Hoxd1-11* 的表达, 并在四肢近端维持这种表达模式; 在肢端发育后期, 5'中心粒侧 TAD (centromeric domain, C-DOM) 内由 Island 1-5、GCR 和 Prox 组成的上游增强子簇或超级增强子(downstream enhancer cluster or super-enhancer)协同调节四肢远端 *Hoxd9-13* 的表达^[19,42,43]。*HoxD* 基因簇区域大范围的片段反转破坏原来的基因组三维结构, 影响 *HoxD* 基因簇的正常表达^[39,41,44], 但这些大范围的片段反转主要针对增强子调控元件, 针对绝缘子 CTCF 位点的 DNA 片段反转研究较少。

成簇规律间隔短回文重复系统(clustered regularly interspaced short palindromic repeats/CRISPR-associated nuclease 9, CRISPR/Cas9)源于古菌和细菌中的一种适应性免疫系统, 近年来被广泛应用于 DNA 片段反转、DNA 片段删除和染色体异位等编辑实验^[22,45~52]。HEK293T 细胞系是一种常用的模式细胞系, 常被用于对染色质三维空间结构的研究^[5,53,54]。本实验室之前通过对 HEK293T 细胞系以及 HEC-1B 等细胞系中 CBS 位点反转的研究发现, CBS 方向性是哺乳动物细胞中普遍存在的规律^[5]。本研究利用 CRISPR DNA 片段编辑技术在 HEK293T 细胞系中对 *HOXD* 基因簇所在区域中串联排列的 CBS 元件进行一系列组合性的反转, 利用定量高分辨率染色质构象捕获(quantitative high-resolution chromosome conformation capture copy, QHR-4C)技术^[34]研究 CBS 反转后染色质空间构象的变化, 发现 CBS 反转会改变该 CBS 原有的远距离 DNA 相互作用, 引起染色质拓扑结构域边界漂移至新形成的一对反向-正向的 CBS 处; 同时, 相邻两个染色质拓扑结构域边界区域的反向-正向 CBS 可以通过与各自所在拓扑结构域内的相向 CBS 形成染色质环, 发挥绝缘子功能, 这种绝缘子的作用有助于稳定基因组三维空间构象。

1 材料与方法

1.1 实验材料

人胚肾细胞系 HEK293T 细胞从中国科学院细胞库购买; pcDNA3.1-Cas9-WT 质粒由北京大学席建忠教授惠赠; pGL3-U6-sgRNA 质粒由上海科技大学黄行许教授惠赠; DMEM 培养基从美国 Hyclone 公司购买; 胎牛血清购自上海 ExCell 公司; *Bsa* I 内切酶、*Dpn* II 内切酶、T4 DNA 连接酶、蛋白酶 K 和 ChIP-Seq 建库试剂盒均购自美国 NEB 公司; DNA 凝胶回收试剂盒和质粒 DNA 小量试剂盒购自美国 AXYGENT 公司; 嘌呤霉素、NP-40、Triton X-100、DMSO、Tris base、SDS、醋酸钠购自美国 SIGMA 公司; Lipofectamine 3000、RNaseA、糖原、16% 甲醛溶液购自美国 Thermo Fisher Scientific 公司; 胰酶和青链霉素购自美国 Gibco 公司; Streptavidin Magnetic Beads 购自美国 Invitrogen 公司; RNA-seq 试剂盒和 Phanta 酶购自南京诺维赞生物科技有限公司; TRIzol 购自美国 Ambion 公司; AMPure XP beads 购自美国 Beckman Coulter 公司; CTCF 抗体从英国 Abcam 公司购买; 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; 引物均由上海生物工程股份有限公司和北京擎科新业生物技术有限公司提供。

1.2 *HOXD* 基因簇 TAD 分析

从 SRA 数据库下载 HEK293 细胞的 Hi-C 数据^[54]: SRR710075、SRR710076、SRR710077, 以 hg19 为参考基因组, 利用 HiC-Pro^[55]软件进行分析, 获得标准化矩阵; 截取 chr2:175,760,000~178,200,000 范围的矩阵, 利用 TADtool 绘制热图^[56]。

1.3 CTCF 结合序列分析

分析 CTCF 蛋白 ChIP-seq 数据, 确定 *HOXD* 基因簇区域 CBS 位点分布情况; 利用 MEME^[57]软件分析获得 CTCF 的结合基序(motif), 以此基序作为参考, 使用 FIMO^[57]软件查找 *HOXD* 基因簇区域 CTCF 结合序列, 并确定 CBS 方向。

1.4 细胞培养

HEK293T 细胞用 89% DMEM、10%胎牛血清和 1%青链霉素双抗配制的完全培养基培养,在 5% CO₂ 浓度、37℃恒温培养箱中培养。

1.5 构建 sgRNA 质粒

使用 *Bsa* I 酶将 pGL3-U6-sgRNA 质粒线性化,并用琼脂糖凝胶进行分离纯化。将购买的 sgRNA 正反脱氧核苷酸单链溶解并稀释到 20 μmol/L (表 1),

表 1 引物序列
Table 1 Primers used in this study

类型	引物名称	序列(5'→3')
sgRNA	a-sgRNA1F	<u>accg</u> CGAAGAGTGCGGGAGAACGG
	a-sgRNA1R	<u>aaac</u> CCGTTCTCCCGCACTCTTCG
	b-sgRNA1F	<u>accg</u> GGCGCATCAGGAATGTAAG
	b-sgRNA1R	<u>aaac</u> CTTACATTCCTGATGCGCC
	b-sgRNA2F	<u>accg</u> TTCCAGAGATTATGAGCCAG
	b-sgRNA2R	<u>aaac</u> CTGGCTCATAATCTCTGGAA
	c-sgRNA2F	<u>accg</u> AGTAAACTGGGTGTTGCC
	c-sgRNA2R	<u>aaac</u> GGCAACACCCAGTTTTACT
	e-sgRNA1F	<u>accg</u> AACTGTGCTCAAACGCTCTC
	e-sgRNA1R	<u>aaac</u> GAGAGCGTTTGAGCACAGTT
	e-sgRNA2F	<u>accg</u> GAGGCGAAACAGCTGTTGT
	e-sgRNA2R	<u>aaac</u> ACAACAGCTGTTTGCGCCTC
PCR	a-1F	TTCCAGCACCTCGGCTTTGTC
	a-1R	CCCACTTTCCACCTCTGTCCTG
	b-1F	GTCCGCCCCGTGAGCTTCTGAA
	b-1R	GCGTTGGCAGGGCTGGACTC
	b-1F1	TTCCCTGACCCTCCAAGCAC
	b-1R1	CTCACAGCAGCCGAAACCG
	b-2F	GACACCTTGGTTCAGGCTCG
	b-2R	CAGAGGCTGCATCAAACCAC
	c-1F	TGATGCAGCCTCTGTGACCG
	c-1R	AGTTTTCCCGTGGCGTCTGA
	c-2F	GGACAACCCCTCACCTTGAA
	c-2R	TATGGTGCCCAAAGTCCCAC
	e-1F	TTCCCTGTCCCAGCTTGATTTT
	e-1R	TCAACAGTGAAGGGCGGTGC
	e-2F	CAAGCCACTCTCCCGCCACTA
	e-2R	TCGCTCTCGTCTCTCTTGGG
bio-primer	upstream-a-bioprimer	5'biotin-GGACGAGGGACATAGAGAGTTC
	downstream-e-bioprimer	5'biotin- AGGAGCCCAGGCATAGAGAC
	CS38-bioprimer	5'biotin-CCGAATTAAATCCCCGTGAC
	Island5-bioprimer	5'biotin-AAACACAAATGCATCAACCTG
P5 primer	upstream-a-P5	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTCTGTCTAGAGATTCAAGTATTTCCCA

续表

类型	引物名称	序列(5'→3')
P5 primer	downstream-e-P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTACGCGATACTCAAGGGAATACAAGCC
	CS38-P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTTACAGTGTCATTTTCCCTG
	Island5-P5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCATCTCATGAAGCTGGCATCT
P7 index	P7-index-1	CAAGCAGAAGACGGCATAACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-2	CAAGCAGAAGACGGCATAACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-3	CAAGCAGAAGACGGCATAACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-4	CAAGCAGAAGACGGCATAACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-5	CAAGCAGAAGACGGCATAACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-6	CAAGCAGAAGACGGCATAACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-7	CAAGCAGAAGACGGCATAACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-8	CAAGCAGAAGACGGCATAACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-9	CAAGCAGAAGACGGCATAACGAGATCATAGCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-10	CAAGCAGAAGACGGCATAACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-11	CAAGCAGAAGACGGCATAACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-12	CAAGCAGAAGACGGCATAACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-13	CAAGCAGAAGACGGCATAACGAGATAGAGTACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-14	CAAGCAGAAGACGGCATAACGAGATGCTCCGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-15	CAAGCAGAAGACGGCATAACGAGATCATGAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-16	CAAGCAGAAGACGGCATAACGAGATTGAATCGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-17	CAAGCAGAAGACGGCATAACGAGATGTCTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-18	CAAGCAGAAGACGGCATAACGAGATCTGAATGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-19	CAAGCAGAAGACGGCATAACGAGATCGCTTCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-20	CAAGCAGAAGACGGCATAACGAGATTTCGCATGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-21	CAAGCAGAAGACGGCATAACGAGATAATAGCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-22	CAAGCAGAAGACGGCATAACGAGATGTCGCGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-23	CAAGCAGAAGACGGCATAACGAGATACGCGATAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

续表

类型	引物名称	序列(5'→3')
P7 index	P7-index-24	CAAGCAGAAGACGGCATAACGAGATTGATCGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-25	CAAGCAGAAGACGGCATAACGAGATCCGCATGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-26	CAAGCAGAAGACGGCATAACGAGATCCACAATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-27	CAAGCAGAAGACGGCATAACGAGATGATGTTCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-28	CAAGCAGAAGACGGCATAACGAGATGAATACGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-29	CAAGCAGAAGACGGCATAACGAGATTAGATACCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-30	CAAGCAGAAGACGGCATAACGAGATTATGGTTCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
adapter	adapter-upper	GACGTGTGCTCTTCCGATCTGNNNNNN-3' NH2 C6
	adapter-lower	5'P-CAGATCGGAAGAGCACACGTC-3' NH2 C6

sgRNA 中下划线标示小写字母部分为构建质粒所需的粘性末端, P7 index 中下划线标示部分为 index。

配制引物退火溶液, 体系如下: 正反单链 sgRNA 各 2 μ L, 2 μ L NEB buffer2, 14 μ L ddH₂O, 并置于水浴锅中 95℃ 5 min, 然后自然冷却至室温。带有悬挂端的双链 sgRNA 与获得的线性化载体连接, 然后用感受态细胞转化摇菌。利用质粒小量试剂盒提取质粒, 并用桑格测序(Sanger sequencing)确认质粒是否正确连接。

1.6 筛选 CTCF 位点反转的单细胞克隆

待到 12 孔板中 HEK293T 细胞长到约 60%时, 利用 Lipofectamine 3000 试剂将 Cas9 质粒、sgRNA 质粒一起转染到细胞中。2 天后, 用含有 2 ng/ μ L 嘌呤霉素的 DMEM 培养基培养 4 天。换用普通 DMEM 完全培养基培养 2 天后, 进行单克隆铺板。

约 10 天后, 用胰酶消化单克隆细胞, 取部分细胞使用碱裂法裂解获得基因组。使用特异性引物进行 PCR 鉴定, 鉴定各单细胞克隆类型, 鉴定引物如表 1 所示。PCR 产物经 TA 克隆后, 进行桑格测序, 确定单细胞克隆基因型。

1.7 定量高分辨率染色质构象捕获实验 QHR-4C

QHR-4C 实验方法参考本实验室之前的研究^[34], 步骤如下: 培养鉴定好的单细胞克隆, 收取处于对数生长期的细胞 500 万个左右, 用 4C 裂解液(配方为: 50 mmol/L Tris-HCl, 150 mmol/L NaCl, 5 mmol/L

EDTA, 0.5% NP-40, 1% Triton X-100)裂解细胞两次。获得的细胞核沉淀经 *Dpn* II 内切酶过夜酶切后, 用 T4 DNA 连接酶在 16℃过夜连接, 之后利用超声将 DNA 片段打断为 200~1000 bp 长度范围。超声后的 DNA 片段利用酚氯仿进行纯化, 后使用带有 biotin 的引物进行线性扩增。扩增产物加上接头后进行 PCR 扩增和产物纯化, 获得 QHR-4C 文库。文库用 Illumina HiSeq 测序获得原始数据, 根据 index 和 barcode 拆分数据。用 fastuniq 去重, 通过 bowtie2 比对到 hg19 参考基因组。比对文件经过 samtools 排序, 用 r3Cseq 进行计数和标准化, 获得每个酶切片段的 RPKM 值。每个单细胞克隆株至少做两个重复实验, 共计 99 个 QHR-4C 文库, 每个文库测序约 800 万条序列。所用 index 见表 1。

1.8 RNA-seq 实验

待细胞长到 12 孔板约 80%~90%时, 使用 TRIzol 处理细胞, 并提取细胞总 RNA。随后用诺维赞 RNA-seq 建库试剂盒进行 RNA-seq 文库构建。步骤如下: 用 beads 捕获 mRNA, 热打断至 150~200 bp, 使用随机引物逆转录获得 cDNA, 加 A 和加接头, 最后进行 PCR 扩增。文库用 Illumina HiSeq 进行测序, 数据经 index 拆分后, 利用 trim_galore 去除低质量 reads, 使用 STAR 比对到 hg19 参考基因组, 利用 cufflinks 对转录本进行计数和标准化, 得到每

个基因的 FPKM 值。每个单细胞克隆株做三个重复实验, 共计 33 个 RNA-seq 文库, 每个文库测序约 1000 万条序列。

1.9 ChIP-seq 实验

待细胞汇合度达到 10 cm 培养皿约 80%~90% 时收取细胞, 用甲醛交联固定, 用超声将基因组 DNA 打断为 100~1000 bp, 孵育 protein G 凝胶珠。4℃过夜孵育一抗后, 再次加入 protein G 凝胶珠孵育, 经洗脱 DNA 和用酚氯仿纯化 DNA 后, 利用 DNA 建库试剂盒进行建库, 包括末端修复、末端加 A、加接头和 PCR 扩增等步骤, 获得 ChIP-seq 文库。文库用 Illumina HiSeq 进行测序, 测序数据经过 index 拆分获得双端测序序列, 使用 trim_galore 进行数据过滤, bowtie2 比对到 hg19 参考基因组。之后用 samtools 对比对结果进行排序, 使用 macs2 成峰。实验做两个重复, 每个重复测序约 2000 万条序列。

1.10 高通量测序数据信息

该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)^[58]的国家基因库序列归档系统(CNSA)^[59], 项目编号: CNP0001772。

2 结果与分析

2.1 HOXD 基因簇区域 CTCF 位点序列分析

为了研究 TAD 边界处成串 CBS 的绝缘子作用, 本研究选定 *HOXD* 基因簇为研究对象。Hi-C 数据显示 *HOXD* 基因簇位于两个 TAD 边界处, CTCF ChIP-seq 数据证实 *HOXD* 基因簇以及两侧的 C-DOM 和 T-DOM 分布着大量 CTCF 结合位点(图 1A)。为了判断 CBS 的方向性, 对每一个 CTCF 结合峰进行序列分析(图 1B)。这些 CBS 基序共包括 4 个 module, 其中 module 2~4 包含最为保守的 20 bp 核心基序, 少部分在核心基序上游含有 module 1, 正向 CBS 方向为 module 1 到 module 4 (图 1B)。在 module 2~3 中最为保守的 C20、G25 和 G28 是判断 CBS 存在的重要参考, 而 module 1 和位于 24 位的碱基是否为 A 则可以用于判断 CBS 方向性; 这些 CBS 序列特性也与之之前发表文章相符^[2,5]。在 HEK293T 细胞中,

HOXD 基因簇分布着 5 个反向 CBS 和 2 个正向 CBS, 分别命名为 CBS a-e 和 f-g (图 1A)。在 C-DOM 中大部分是正向 CBS, 只有 3 个反向 CBS (图 1B); T-DOM 中 CBS 呈反向串联排布(图 1B)。

2.2 构建 *HOXD* 基因簇一系列 CTCF 位点反转的单细胞克隆株

利用 CRISPR/Cas9 DNA 片段编辑技术, 针对预反转的 DNA 片段设计一对 sgRNA, 将 Cas9 和 sgRNA 质粒转染到 HEK293T 细胞中, Cas9 核酸酶在成对 sgRNA 的介导下对目的 DNA 片段进行切割。断裂的双链 DNA 在细胞 DNA 修复系统的作用下, 可以产生 DNA 片段原位反转, 反转后的 DNA 片段具有两个不同的接口(上游接口和下游接口)(图 1C)。利用此策略, 我们对 *HOXD* 基因簇内 CBS a-e 进行组合反转(图 1D)。共获得 5 种 CBS 反转单细胞克隆株, 分别为 e 反转、b-e 反转、a-e 反转、a-c 反转和 a-b 反转(图 1D), 由于不同 sgRNA 具有不同的编辑效率, 单克隆筛选数相差较大(图 2A), 其中 a-b 反转共筛选 1307 个单细胞克隆只获得了一个预期反转克隆。最后, e 反转获得 5 个纯合子克隆株, b-e 反转、a-e 反转、a-c 各获得 2 个纯合子克隆株, a-b 反转获得 1 个纯合子克隆株(图 2A)。全部单细胞克隆株通过 PCR 进行鉴定, DNA 片段反转的接口还进行 TA 克隆和桑格测序确定其具体的基因型, 大部分接口为两种基因型, 除了在 PAM 位点上游 3 bp 处精确切割后的连接产物外, 单碱基插入占据相当的比例(图 2, B~J), 推测 Cas9 核酸酶在 PAM 位点上游 4 bp 处切割导致^[48,60]。在 b-e#127 单细胞克隆株的一个等位基因下游接口存在 518 bp 的碱基删除, 但删除区域不包括 CBS 核心基序(图 2G); 在 a-c#81 单细胞克隆株的一个等位基因上游接口存在 258 bp 的碱基插入, 经比对发现插入序列为质粒序列(图 2I), 因此这两个细胞株仍是纯合的 CBS 反转单细胞克隆株。

2.3 CTCF 位点反转引起染色质拓扑结构域边界漂移

为了研究 CBS 的绝缘子作用, 对获得的 CBS 反转单细胞克隆株进行 QHR-4C 实验。在野生型细

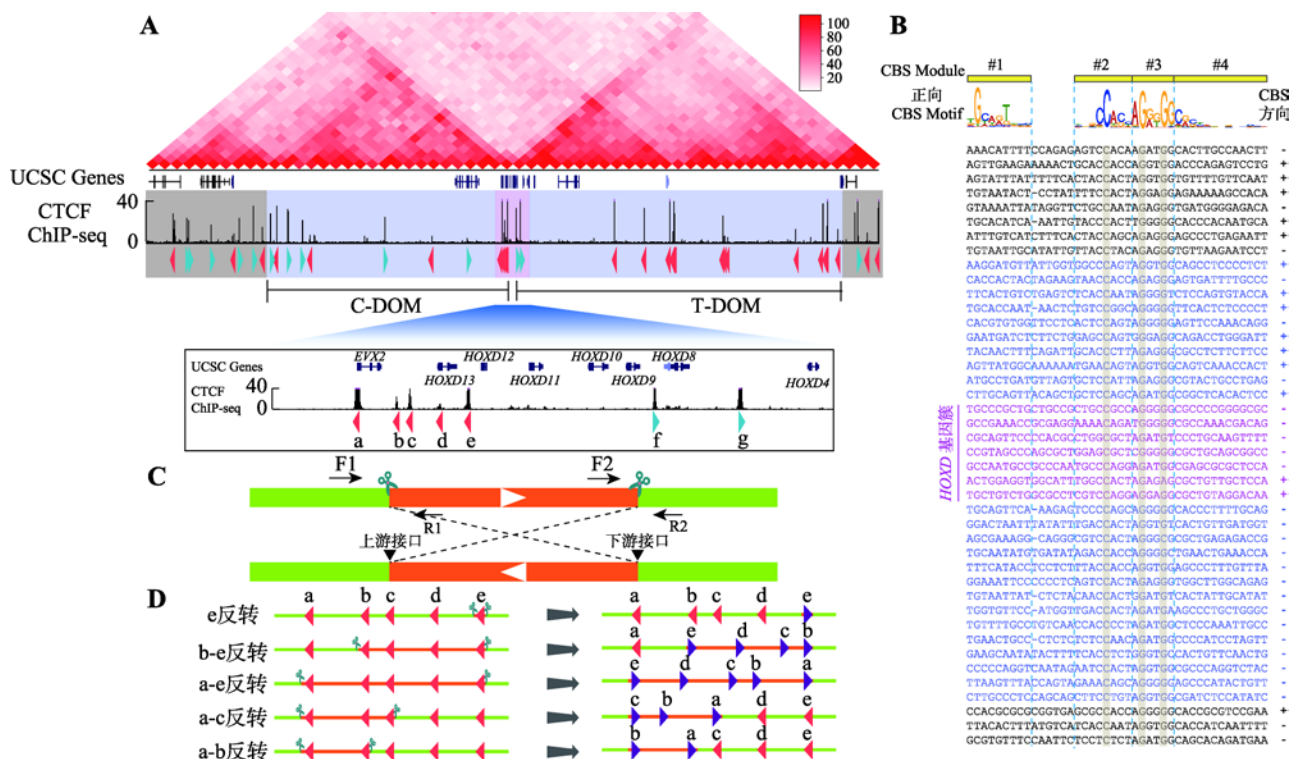


图 1 人类 *HOXD* 基因簇区域 CTCF 位点序列分析

Fig. 1 Sequence analyses of CTCF sites of the human *HOXD* cluster

A: *HOXD* 基因簇区域基因组三维空间结构示意图。*HOXD* 基因簇位于 C-DOM 和 T-DOM 两个 TAD 边界处, 位于 *HOXD* 基因簇附近的 CBS a-e 与 C-DOM 中的正向 CBS 形成染色质环, CBS f-g 与 T-DOM 中的成串反向 CBS 形成染色质环。**B:** *HOXD* 基因簇区域 CBS 序列比对图。正向 CBS 的方向为 module 1~4, 其中 module 1 与 module 2 之间相隔 5 bp 或者 6 bp。DNA 序列中的 “-” 表示该 CBS 的 module 1 与 module 2 间隔为 5 bp, 故用 “-” 补齐。右侧 CBS 方向列, “+” 表示 CBS 为正向, “-” 表示该 CBS 为反向且显示 DNA 序列是反向互补序列。**C:** 利用 CRISPR/Cas9 系统进行 DNA 片段反转示意图。针对预反转片段设计一对 sgRNA 进行片段编辑, 会产生上游接口和下游接口; 在切割点附近设计引物, 用引物 F1+F2、R1+R2 进行 PCR 鉴定, 若出现目的条带且 F1+R1、F2+R2 和 F1+R2 无目的条带, 则为纯合反转单细胞克隆。**D:** *HOXD* 基因簇内 CBS 组合性反转示意图, 共构建五种反转单细胞克隆株。

胞中, 位于 T-DOM 中的增强子 CS38 主要与正向 CBS f-g 有较强的 DNA 相互作用(图 3A), 而位于 C-DOM 中的增强子 Island5 主要与反向的 CBS a-e 具有 DNA 相互作用(图 3B), 与 Hi-C 结果一致(图 1A)。CBS e 反转后, CBS e 与 CS38 的远距离 DNA 相互作用增强(图 3A), 显示反转后 CBS e 可以作为绝缘子阻碍 CBS f-g 渗透的 cohesin 继续滑动; 与之相反, CBS e 与 Island5 的远距离 DNA 相互作用减弱(图 3B)。CBS d 和反转后的 CBS e 呈反向-正向分布, 形成了一个新的染色质拓扑结构域边界。

CBS b-e 反转后, CBS a 与 CBS e 形成一对反向-正向的 CBS (图 1D), CBS b-e 与 Island5 的远距离 DNA 相互作用下降(图 3B), 与 CS38 形成新的远距离 DNA 相互作用(图 3A), 阻碍从 CBS f-g 方向渗透

的 cohesin 继续滑动。CBS a 和反转后的 CBS e 形成了新的染色质拓扑结构域边界, 与 CBS e 反转结果相似, 即边界 CBS 反转引起染色质拓扑结构域边界的漂移。CBS a-e 全部反转后, CBS a-e 与 Island5 的 DNA 相互作用明显减弱(图 3B), 与 CS38 形成新的 DNA 相互作用(图 3A), 由于缺乏成对反向-正向 CBS, 破坏了 *HOXD* 区域的染色质拓扑边界。

为了进一步确定这些染色质空间构象的改变, 选用 CBS a 和 CBS e 附近的 DNA 片段作为观测点 (upstream-a 和 downstream-e, upstream-a 观测点位于 CBS a 的上游, downstream-e 观测点位于 CBS e 的下游), 为了保证观测点不随 CBS 反转而改变, 观测点的位置不在反转片段内。在 CBS e 反转的细胞中, CS38 与 CBS a 的远距离 DNA 相互作用没有变

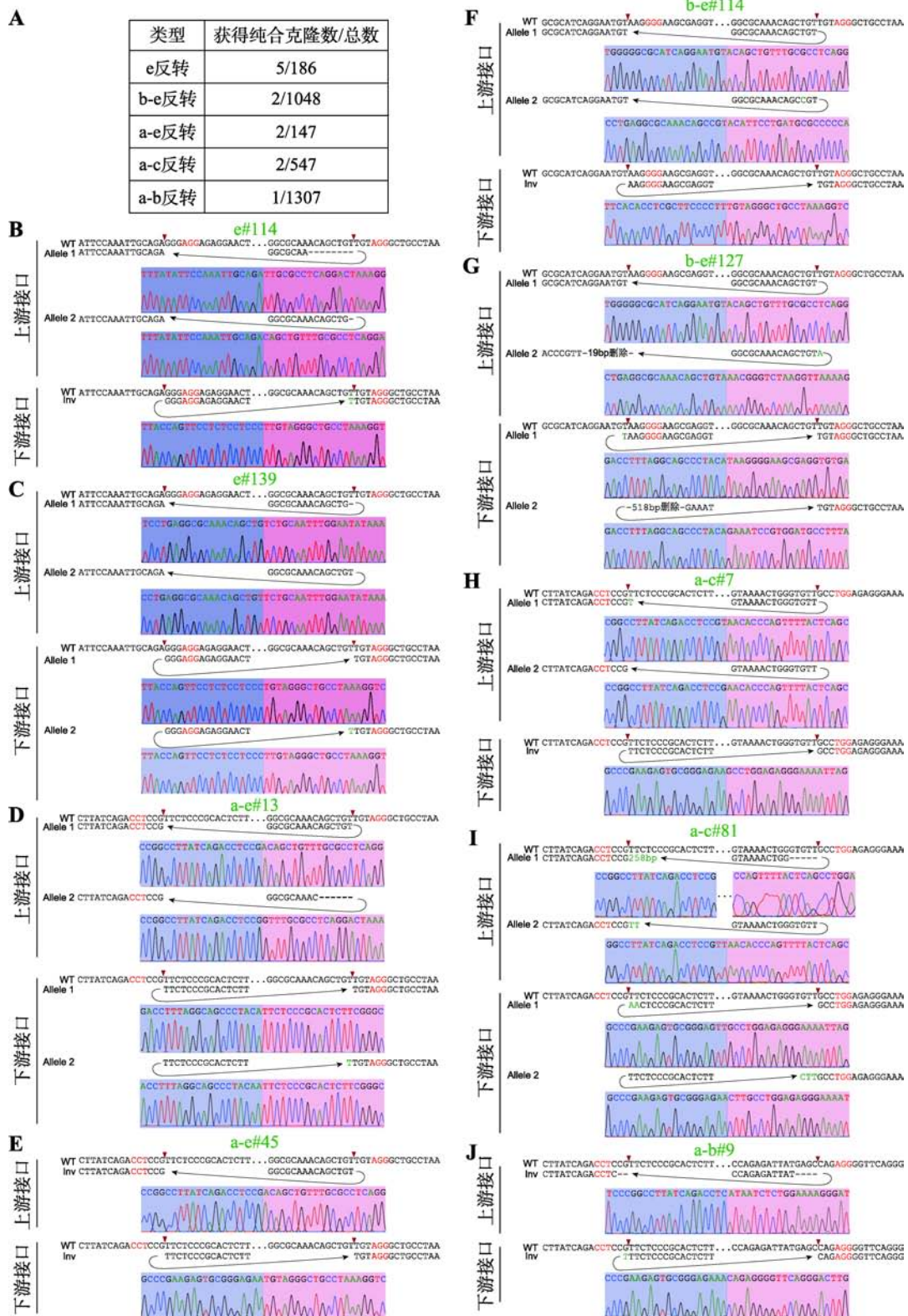


图 2 反转单细胞克隆株基因型鉴定

Fig. 2 Genotyping of single-cell clones of combinatorial CBS inversions

A: 获得纯合 CBS 反转单细胞克隆株个数以及筛选克隆总数目。B~J: 每一个纯合反转单细胞克隆株的基因型鉴定(TA 克隆桑格测序图)。序列中红色表示 PAM 所在位置, 红色箭头指示 PAM 上游-3 位。序列中“-”表示存在碱基删除, 绿色表示存在碱基增加。

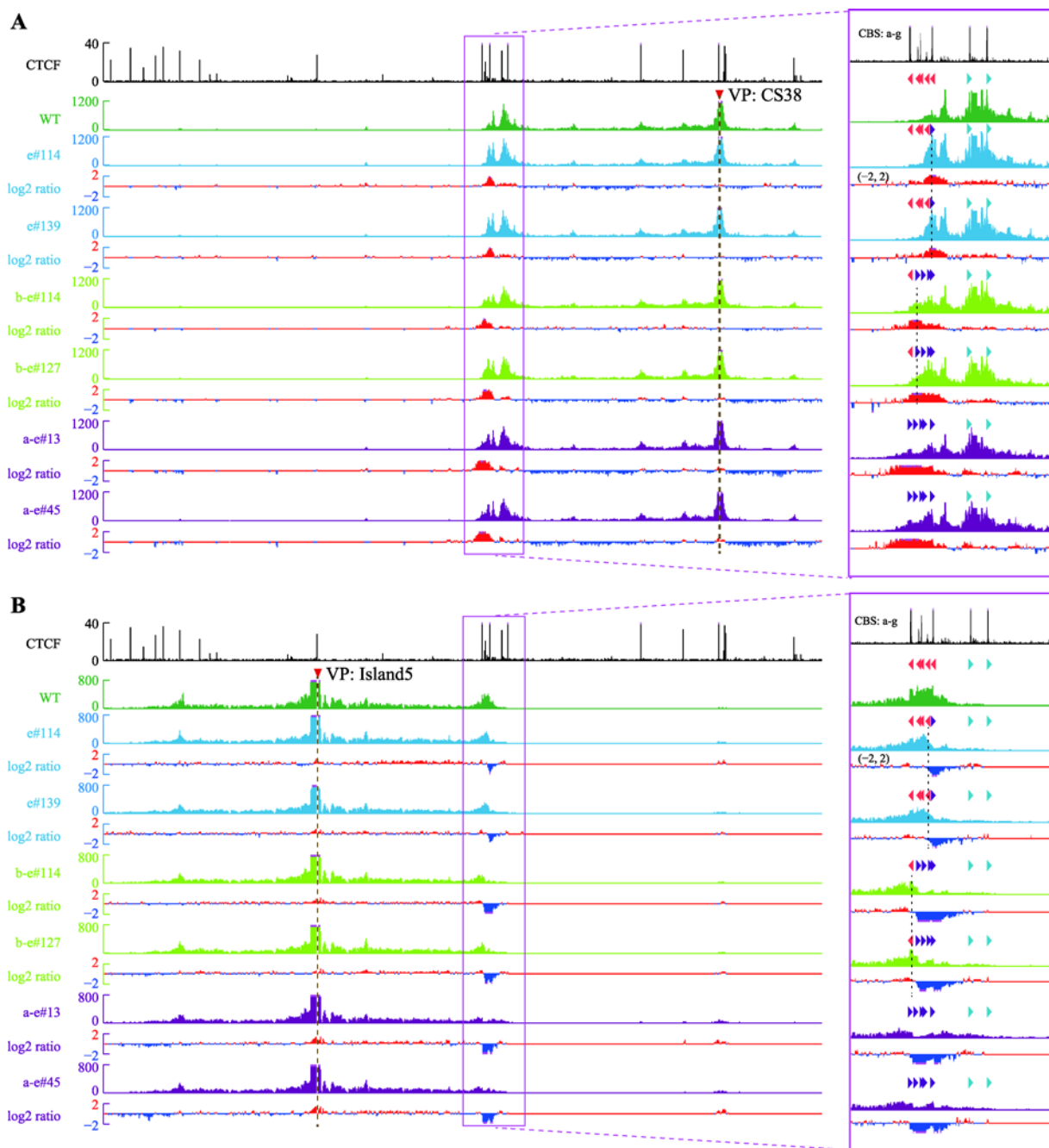


图 3 从两侧增强子观察 CBS 反转后局部三维基因组构象变化

Fig. 3 Alterations of the local 3D chromatin structure observed from viewpoints of flanking enhancers in CBS inversion clones

A: 以 CS38 为观测点(viewpoint, VP)显示, CBS 反转后与 CS38 的染色质相互作用显著增强。CBS 反转单细胞克隆株数据分别与野生型(WT)进行 log2 转化, 获得两者差异数据。B: 以 Island5 为观测点显示, CBS 反转后与 Island5 的原有染色质相互作用减弱。

化(图 4A); 而 CBS b-e 反转后, CS38 与 CBS a 的远距离 DNA 相互作用有轻微增强(图 4A); 反转 CBS a-e 后 CBS a 主要与 CS38 形成较强的 DNA 相互作用(图 4A)。从 downstream-e 为观测点来看, 所有反

转克隆均与 CS38 有较强的 DNA 相互作用(图 4B), 与上述结果相同。这些结果说明, 反转 CBS 后, 该 CBS 原有的远距离 DNA 相互作用被破坏, 即 CBS 方向可以决定染色质环的形成, 反转的 CBS 可以阻

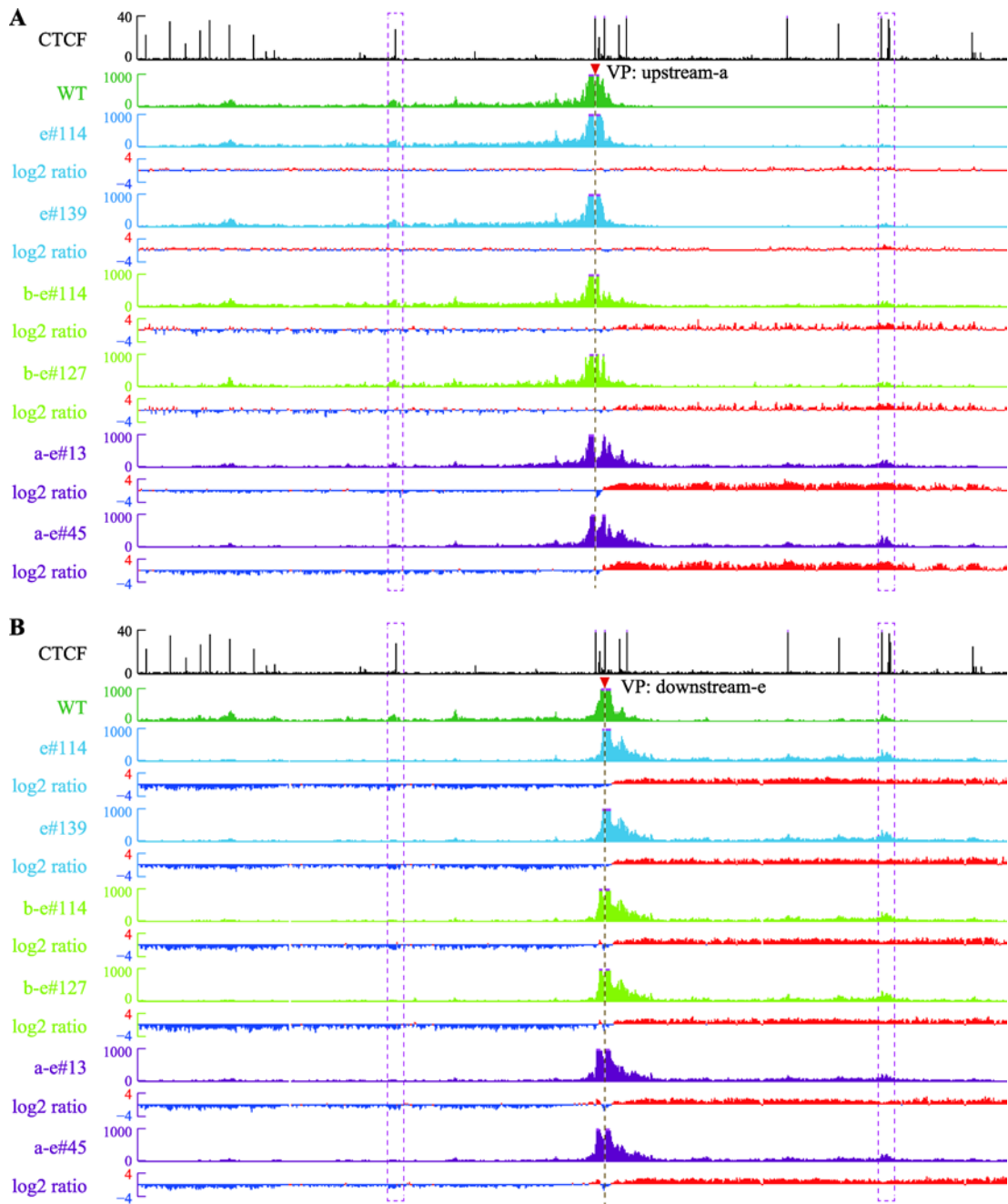


图 4 从反转位点附近观察 CBS 反转后局部三维基因组构象变化

Fig. 4 Alterations of the local 3D chromatin structure observed from viewpoints near the inverted CBS

A: 以 upstream-a 为观测点, 在 CBS e 反转细胞中观测点与两侧远距离 DNA 相互作用无明显变化, 在 CBS b-e 反转细胞中观测点与 CS38 的远距离 DNA 相互作用增强, 在 CBS a-e 反转细胞中观测点与 CS38 相互作用明显增强, 伴随着与 Island5 的远距离 DNA 相互作用减弱。B: 以 downstream-e 为观测点, 在三种包含 CBS e 反转的细胞株中, 观测点与 CS38 的染色质相互作用均增强, 伴随着与 Island5 染色质相互作用减弱。

碍反方向渗透的 cohesin 继续滑动, 发挥绝缘子的功能; 同时, 边界 CBS 元件反转会引起染色质拓扑结构域边界的漂移。

2.4 非边界 CTCF 位点也发挥绝缘子功能

本实验室之前的研究发现, 位于 *Pcdha* 基因簇

启动子区域的正向 CBS 可以与位于远端的增强子 HS5-1 形成染色质环, 阻碍更远处的启动子 CBS 与增强子 HS5-1 的远距离 DNA 相互作用, 具有绝缘子功能^[34]。为了研究 *HOXD* 基因簇区域 CBS 是否

同样具有绝缘子功能, 本研究进一步反转了 CBS a-b 和 CBS a-c (图 1D)。以 Island5 为观测点显示, 反转 CBS a-b 或 CBS a-c 后, Island5 与 CBS e 的远距离 DNA 相互作用增强(图 5A), 而以 CS38 为观测点

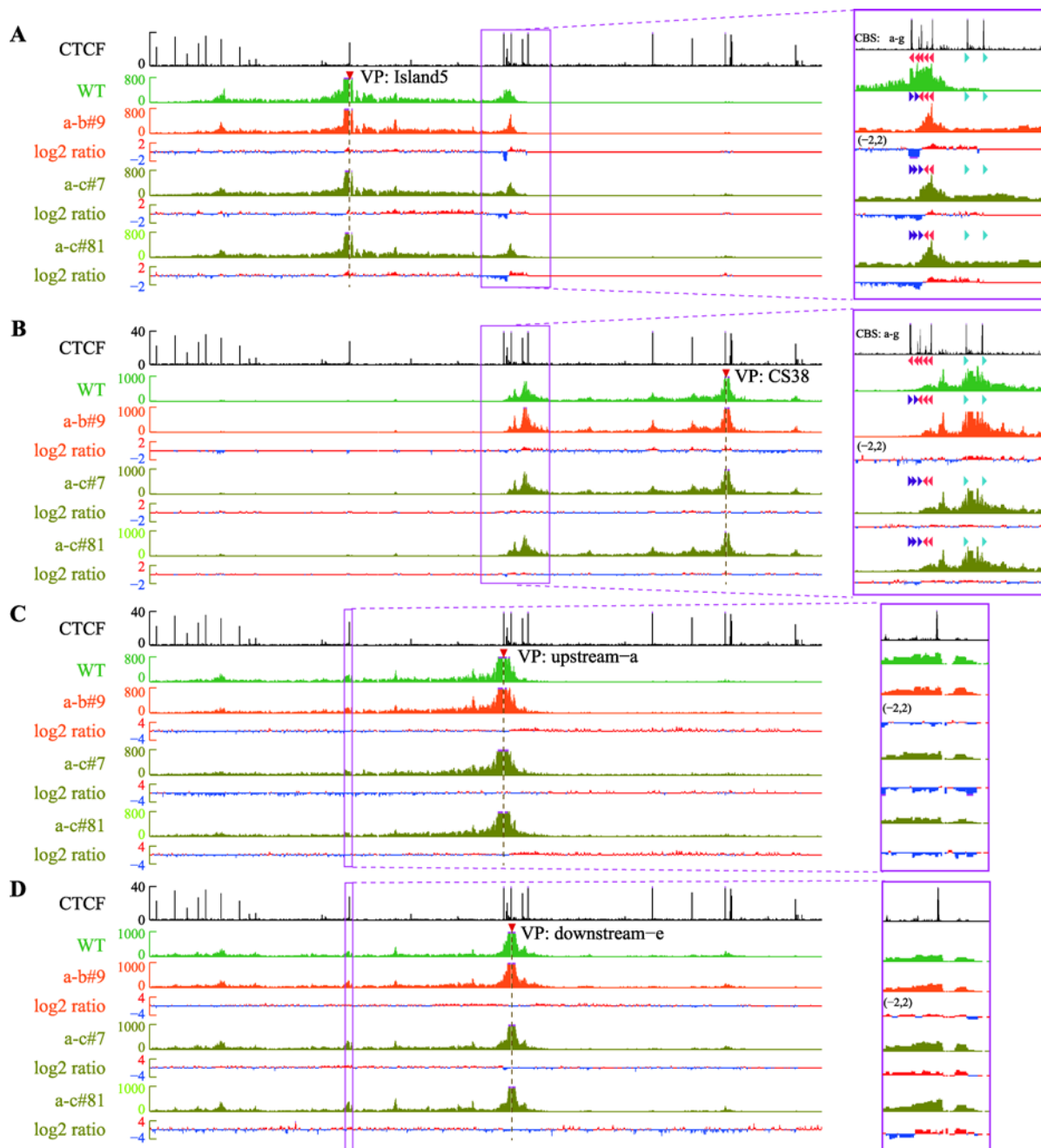


图 5 非边界的成串 CTCF 位点也能发挥绝缘子功能

Fig. 5 Non-boundary CBS elements function as insulators

A: 在 CBS a-b 或 CBS a-c 反转细胞株中, Island5 与被反转 CBS 原有的远距离 DNA 相互作用减弱, 与 CBS e 的远距离 DNA 相互作用增强。B: 以 CS38 为观测点显示, 观测点与被反转 CBS 的染色质相互作用无明显变化。C: 以 upstream-a 为观测点, 在 CBS a-b 或 CBS a-c 反转细胞中, 可见 CBS a 与 Island5 的染色质相互作用减弱。D: 以 downstream-e 为观测点, 在 CBS a-b 或 CBS a-c 反转细胞中, 可见 CBS e 与 Island5 的染色质相互作用增强。

显示 CS38 与 CBS a-e 的远距离 DNA 相互作用无明显变化(图 5B)。相应地, upstream-a 与上游的 Island5 的远距离 DNA 相互作用减弱(图 5C), downstream-e 与 Island5 的远距离 DNA 相互作用增强(图 5D), 与 Island5 为观测点的结果相同。但是, 染色质拓扑结构域边界并没有发生变化。因此, CBS a-b 或 a-c 可以作为绝缘子, 通过与 Island5 形成染色质环, 分别阻碍 CBS c-e 或 d-e 与 Island5 的远距离 DNA 相互作用。反转细胞株的 RNA-seq 实验显示 *HOXD* 基因簇区域局部染色质构象变化后, 基因表达水平会发生改变(图 6, A~D)。其中, 在所有的 CBS 反转细胞株中, *HOXD13* 的表达水平均显著下降; 除 CBS b-e 反转细胞株外, *HOXD3* 与 *HOXD10* 的表达水平均下降; 有趣的是, 在 CBS e 反转和 CBS a-e 反转的细胞中, *HOXD9* 的表达水平显著上升(图 6, A~D)。

3 讨论

CTCF 通过阻碍 cohesin 蛋白在染色质上滑动形成染色质环, 介导远距离的 DNA 相互作用, 对于基因组三维空间结构的形成与维持有重要作用^[29,61,62]。多个位点的 CBS 反转和插入实验显示染色质环形成依赖于成对正向-反向 CBS^[5,28,29], 而 CBS 的绝缘子功能是不依赖方向的^[34]。研究 CBS 的方向性有助于理解 CBS 在染色质高级拓扑结构中的功能, 但是在成串 CBS 中, 理解 CBS 的功能则更加复杂。原钙粘蛋白基因簇的 CBS 研究显示, 成串 CBS 可能通过环挤出模型, 呈现类似于“葫芦”结构的平衡分布模式^[34]。这显示成对正向-反向 CBS 对于 cohesin 的阻碍不是完全的, 存在 cohesin 渗透出正向-反向 CBS 对(CBS pair)限制的染色质范围的情况, 可能是由于 CTCF 与 DNA 结合是动态的^[38], 因而, 在成

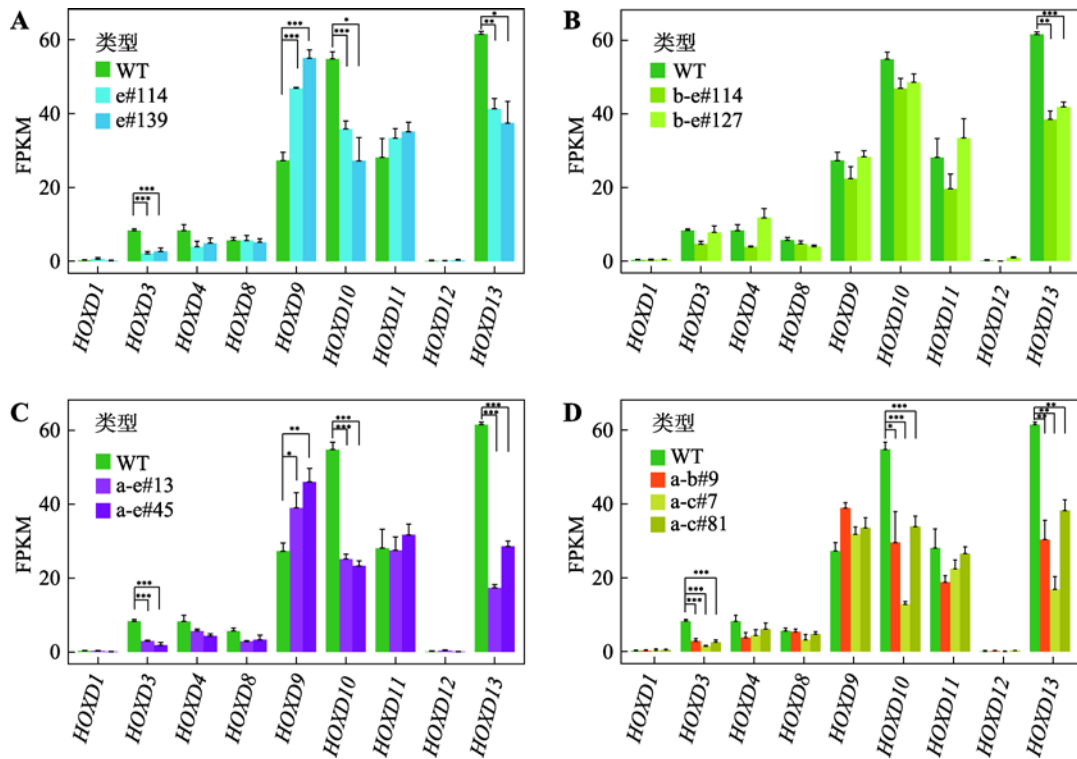


图 6 CTCF 位点反转改变 *HOXD* 基因表达

Fig. 6 CBS inversions alter expression levels of *HOXD* genes

A: CBS e 反转后, *HOXD3*、*HOXD10* 和 *HOXD13* 的表达水平明显降低, *HOXD9* 的表达水平升高; *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$; FPKM, fragments per kilobase of transcript per million mapped reads. B: CBS b-e 反转后, *HOXD13* 表达水平显著降低。C: CBS a-e 反转后, *HOXD3*、*HOXD10* 和 *HOXD13* 的表达水平降低, 而 *HOXD9* 的表达水平上升。D: CBS a-b 或 CBS a-c 反转后, *HOXD3*、*HOXD10* 和 *HOXD13* 的表达水平降低。

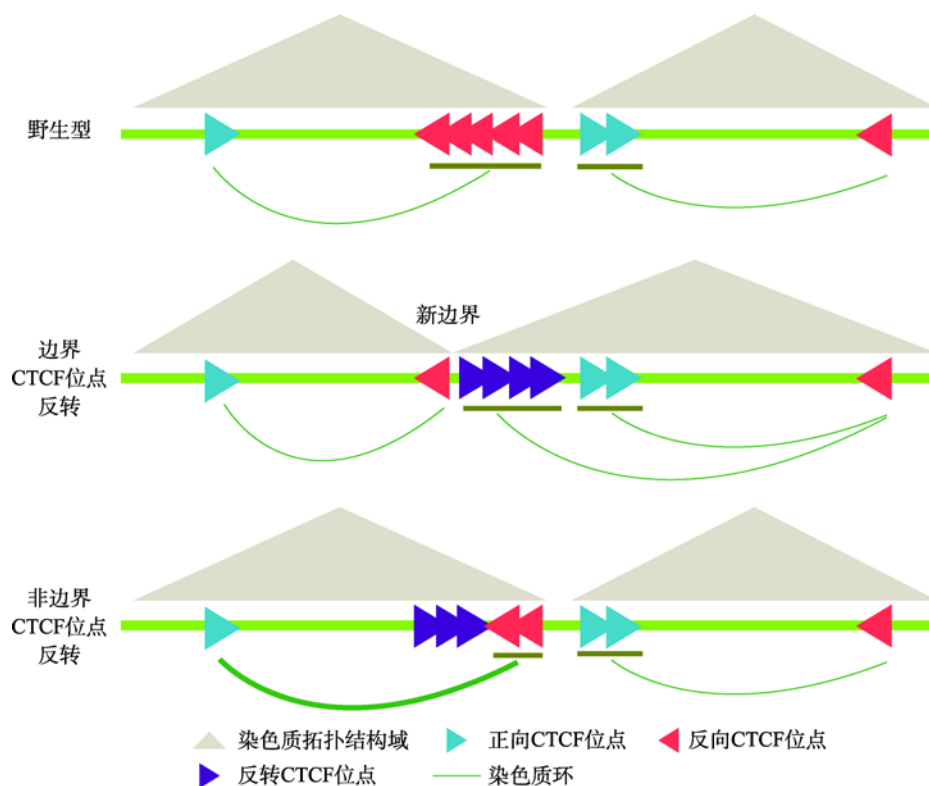


图 7 染色质拓扑结构域边界成串 CTCF 位点发挥绝缘子功能的模式图

Fig. 7 A model for insulator function of tandem boundary CBS sites

A: 在野生型细胞中, 位于染色质拓扑结构域边界成串反向-正向排列的 CBS 分别与两侧拓扑结构域内的 CBS 形成染色质环。边界 CBS 反转后, 可以阻碍反方向渗透的 cohesin 滑动, 形成染色质环; 染色质结构域边界也漂移至形成的一对反向-正向 CBS 处, 而非边界 CBS 反转不引起染色质结构域边界漂移。非边界 CBS 反转后边界 CBS 与拓扑结构域内的相向 CBS 的远距离 DNA 相互作用增强, 显示非边界 CBS 也能发挥绝缘子功能。

串 CBS 中, CBS 呈“葫芦”状分布^[34]。但是, 位于染色质拓扑结构域边界的成串 CBS 如何发挥绝缘子功能, 目前仍缺少系统研究。

本研究通过对 *HOXD* 基因簇区域的 CBS 进行 CRISPR 组合反转, 获得了五种纯合反转单细胞克隆株。QHR-4C 实验显示, 在 CBS e、CBS b-e、CBS a-e 反转细胞中, 由于包含了 TAD 边界的 CBS 元件, 反转的 CBS 引起了远距离 DNA 互作的巨大变化, 证实了边界单个 CBS 反转可以改变染色质环形成方向^[29]。在这三种反转细胞中, 多个观测点的 QHR-4C 数据显示, 反转后 CBS 可以阻碍渗透的 cohesin 继续滑动, 发挥绝缘子功能, 染色质拓扑结构域边界也漂移至新形成的一对反向-正向 CBS 处(图 7)。其次, CBS a-b 和 CBS a-c 反转结果显示, CBS a-b 或 a-c 可以作为绝缘子, 分别阻碍 CBS c-e 或 d-e 与增强子 Island5 的远距离 DNA 相互作用, 证实在成串

CBS a-e 中, 非边界的 CBS 也可以作为绝缘子阻碍边界 CBS 与增强子的染色质环的形成; 反转非边界 CBS 不改变染色质拓扑结构域边界(图 7)。CBS 反转引起的局部染色质构象变化会进一步影响 *HOXD* 基因簇的表达, 但是对于不同的基因, 引起的表达水平变化不尽相同, 显示出 *HOXD* 基因簇存在复杂的表达调控机制。综上所述, 本研究验证了环挤出模型, 同时进一步完善了对 CTCF 绝缘功能的理解。

参考文献(References):

- [1] Kumar Y, Sengupta D, Bickmore W. Recent advances in the spatial organization of the mammalian genome. *J Biosci*, 2020, 45: 18. [\[DOI\]](#)
- [2] 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. [\[DOI\]](#)

- [3] Ning CY, He MN, Tang QZ, Zhu Q, Li MZ, Li DY. Advances in mammalian three-dimensional genome by using Hi-C technology approach. *Hereditas(Beijing)*, 2019, 41(3): 215–233.
宁椿游, 何梦楠, 唐茜子, 朱庆, 李明洲, 李地艳. 基于 Hi-C 技术哺乳动物三维基因组研究进展. *遗传*, 2019, 41(3): 215–233. [DOI]
- [4] Marchal C, Sima J, Gilbert DM. Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol*, 2019, 20(12): 721–737. [DOI]
- [5] 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]
- [6] de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, 2013, 502(7472): 499–506. [DOI]
- [7] Sanders JT, Freeman TF, Xu Y, Gollosi R, Stallard MA, Hill AM, San Martin R, Balajee AS, McCord RP. Radiation-induced DNA damage and repair effects on 3D genome organization. *Nat Commun*, 2020, 11(1): 6178. [DOI]
- [8] Arnould C, Rocher V, Finoux AL, Clouaire T, Li K, Zhou F, Caron P, Mangeot PE, Ricci EP, Mourad R, Haber JE, Noordermeer D, Legube G. Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature*, 2021, 590(7847): 660–665. [DOI]
- [9] Cremer T, Cremer M. Chromosome territories. *Cold Spring Harb Perspect Biol*, 2010, 2(3): a003889. [DOI]
- [10] Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*, 2002, 295(5558): 1306–1311. [DOI]
- [11] Zhao ZH, Tavoosidana G, Sjölander M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 2006, 38(11): 1341–1347. [DOI]
- [12] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, 326(5950): 289–293. [DOI]
- [13] Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EGY, Huang PYH, 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 RKM, 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]
- [14] Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Bluthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 2012, 485(7398): 381–385. [DOI]
- [15] Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 2012, 485(7398): 376–380. [DOI]
- [16] Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 2014, 159(7): 1665–1680. [DOI]
- [17] 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]
- [18] Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci USA*, 2015, 112(47): E6456–E6465. [DOI]
- [19] Andrey G, Montavon T, Mascrez 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]
- [20] Dixon JR, Gorkin DU, Ren B. Chromatin domains: The unit of chromosome organization. *Mol Cell*, 2016, 62(5): 668–680. [DOI]
- [21] Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett

- NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, Guo YE, Hnisz D, Jaenisch R, Bradner JE, Gray NS, Young RA. YY1 is a structural regulator of enhancer-promoter loops. *Cell*, 2017, 171(7): 1573–1588.e28. [DOI]
- [22] 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]
- [23] Lobanenkov 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]
- [24] Yin ML, Wang JY, Wang M, Li XM, Zhang M, Wu Q, Wang YL. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res*, 2017, 27(11): 1365–1377. [DOI]
- [25] Nanni L, Ceri S, Logie C. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol*, 2020, 21(1): 197. [DOI]
- [26] Ghirlando R, Felsenfeld G. CTCF: Making the right connections. *Genes Dev*, 2016, 30(8): 881–891. [DOI]
- [27] Phillips JE, Corces VG. CTCF: Master weaver of the genome. *Cell*, 2009, 137(7): 1194–1211. [DOI]
- [28] de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, de Laat W. CTCF binding polarity determines chromatin looping. *Mol Cell*, 2015, 60(4): 676–684. [DOI]
- [29] 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]
- [30] Wu Q, Jia ZL. Wiring the brain by clustered protocadherin neural codes. *Neurosci Bull*, 2021, 37(1): 117–131. [DOI]
- [31] Narendra V, Rocha PP, An DS, Raviram R, Skok JA, Mazzoni EO, Reinberg D. CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation. *Science*, 2015, 347(6225): 1017–1021. [DOI]
- [32] Ren G, Jin WF, Cui KR, Rodriguez J, Hu GQ, Zhang ZY, Larson DR, Zhao KJ. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol Cell*, 2017, 67(6): 1049–1058.e6. [DOI]
- [33] Gómez-Marín C, Tena JJ, Acemel RD, López-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, Bovolenta P, Nobrega MA, Carvajal J, Gómez-Skarmeta JL. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci USA*, 2015, 112(24): 7542–7547. [DOI]
- [34] 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]
- [35] Zhang D, Huang P, Sharma M, Keller CA, Giardine B, Zhang HY, Gilgenast TG, Phillips-Cremens JE, Hardison RC, Blobel GA. Alteration of genome folding via contact domain boundary insertion. *Nat Genet*, 2020, 52(10): 1076–1087. [DOI]
- [36] 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]
- [37] Zhang XF, Zhang Y, Ba ZQ, Kyritsis N, Casellas R, Alt FW. Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature*, 2019, 575(7782): 385–389. [DOI]
- [38] 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]
- [39] Amândio AR, Lopez-Delisle L, Bolt CC, Mascrez B, Duboule D. A complex regulatory landscape involved in the development of mammalian external genitals. *eLife*, 2020, 9: e52962. [DOI]
- [40] Bianco S, Annunziatella C, Andrey G, Chiariello AM, Esposito A, Fiorillo L, Prisco A, Conte M, Campanile R, Nicodemi M. Modeling single-molecule conformations of the *HoxD* region in mouse embryonic stem and cortical neuronal cells. *Cell Rep*, 2019, 28(6): 1574–1583.e4. [DOI]
- [41] Fabre PJ, Leleu M, Mormann BH, Lopez-Delisle L, Noordermeer D, Beccari L, Duboule D. Large scale genomic reorganization of topological domains at the *HoxD* locus. *Genome Biol*, 2017, 18(1): 149. [DOI]
- [42] 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]

- [43] Montavon T, Soshnikova N, Mascrez 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]
- [44] Rodríguez-Carballo E, Lopez-Delisle L, Willemin A, Beccari L, Gitto S, Mascrez 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]
- [45] 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]
- [46] 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]
- [47] 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]
- [48] Shou J, Li JH, Liu YB, 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] Wang LY, Huang HY, Wu Q. The diversity of DNA fragment editing by CRISPR/Cas9 in highly homologous or repetitive sequences. *Hereditas(Beijing)*, 2017, 39(4): 313–325.
汪乐洋, 黄海燕, 吴强. 利用 CRISPR/Cas9 对基因组中高度同源 DNA 片段编辑多样性的遗传学研究. 遗传, 2017, 39(4): 313–325. [DOI]
- [50] Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, Hsu PD, Wu XB, Jiang WY, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas9 systems. *Science*, 2013, 339(6121): 819–823. [DOI]
- [51] Wu Q, Shou J. Toward precise CRISPR DNA fragment editing and predictable 3D genome engineering. *J Mol Cell Biol*, 2021, 12(11): 828–856. [DOI]
- [52] 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]
- [53] 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]
- [54] Zuin J, Dixon JR, van der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch TA, van IJcken WFJ, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci USA*, 2014, 111(3): 996–1001. [DOI]
- [55] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol*, 2015, 16: 259. [DOI]
- [56] Kruse K, Hug CB, Hernández-Rodríguez B, Vaquerizas JM. TADtool: Visual parameter identification for TAD-calling algorithms. *Bioinformatics*, 2016, 32(20): 3190–3192. [DOI]
- [57] Bailey TL, Johnson J, Grant CE, Noble WS. The meme suite. *Nucleic Acids Res*, 2015, 43(W1): W39–W49. [DOI]
- [58] 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(8): 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]
- [59] 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]
- [60] Shi X, Shou J, Mehryar MM, Li JW, Wang LY, Zhang M, Huang HY, Sun XF, Wu Q. Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di- and tri-nucleotide CRISPR insertions without template donor. *Cell Discov*, 2019, 5: 53. [DOI]
- [61] Gligoris TG, Scheinost JC, Burmann F, Petela N, Chan KL, Uluocak P, Beckouët F, Gruber S, Nasmyth K, Löwe J. Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science*, 2014, 346(6212): 963–967. [DOI]
- [62] Nishiyama T. Cohesion and cohesin-dependent chromatin organization. *Curr Opin Cell Biol*, 2019, 58: 8–14. [DOI]