遗传 Hereditas (Beijing) 2019 年 6 月, 41(6): 509－523
www.chinagene.cn

研究报告

染色质架构蛋白 CTCF 调控 *UGT1* 基因簇的表达

郑晓飞，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：尿苷二磷酸葡萄糖醛酸转移酶(UDP-glucuronosyltransferase, UGT)是一类重要的II相药物代谢酶，通过葡萄糖醛酸接合反应代谢大量内外源小分子化合物，对机体维持内部动态平衡具有重要意义。*UGT* 基因突变或表达异常会造成高胆红素血症等多种疾病、影响药物疗效或减弱代谢药物能力，因此探索 *UGT* 表达调控机制将会为人类疾病的预防和个体化医疗以及精准医学提供科学依据。脊椎动物 *UGT* 分为 *UGT1* 和 *UGT2* 两个亚家族，*UGT1* 基因簇结构与原钙粘蛋白(protocadherin, *Pcdh*)、免疫球蛋白或 B 细胞受体(immunoglobulin or B-cell receptor)、T 细胞受体(T-cell receptor)基因簇类似，但与 *UGT2* 结构不同，分为可变区和恒定区，可变区包含成串排列的外显子，任意一个外显子都可以被可变剪接到下游同一套恒定区外显子上，形成 9 种 *UGT1* 信使 RNA 并翻译成不同 UGT1 葡醛酸转移酶亚型。本实验室前期工作发现，染色质架构蛋白 CTCF 结合 DNA 的方向性在染色质三维结构构建中发挥重要作用。基于此，为了进一步解析 *UGT1* 复杂基因簇的三维转录调控机制，本研究分析和比较了人和小鼠 *UGT1* 基因簇的 CTCF 结合位点(CTCF binding site, CBS)的方向性分布，发现人和小鼠的 *UGT1* 基因簇中 CBS 分布差异很大。以人肺癌细胞系 A549 为模型，通过 RNAi 敲低细胞中 CTCF 和 SMC3 (cohesin 亚基)，证明了 CTCF 和 cohesin 蛋白参与调控人 *UGT1* 基因簇的转录表达。进一步采用 CRISPR 介导的 DNA 片段编辑技术对 *hCBS1* 进行了原位反转(*in situ* inversion)和删除，并通过 RNA-seq 分析技术发现 *hCBS1* 删除能够显著降低 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 的表达水平，然而 *hCBS1* 反转仅仅显著降低 *UGT1A7* 的表达水平。上述研究表明 *hCBS1* 参与 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 的转录调节，是人 *UGT1* 基因簇的潜在转录调控元件。本研究为未来进一步探索 *UGT1* 基因簇的三维基因转录调控机制提供了实验基础。

关键词：*UGT1*；CTCF；CRISPR/Cas9；DNA 片段编辑；转录调控

收稿日期：2019−03−15；**修回日期**：2019−04−08

基金项目：国家自然科学基金项目(编号：81872944, 31470820, 81302861)资助[Supported by the National Natural Science Foundation of China (Nos. 81872944, 31470820, 81302861)]

作者简介：郑晓飞，硕士研究生，专业方向：遗传学。E-mail: 1159171993@qq.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.yczz.19-072

网络出版时间：2019/5/28 15:50:37

URI: http://kns.cnki.net/kcms/detail/11.1913.R.20190528.1550.002.html

Chromatin architectural protein CTCF regulates gene expression of the *UGT1* cluster

Xiaofei Zheng, 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: UDP-glucuronosyltransferases (UGTs) are an important family of phase II drug-metabolizing enzymes that catalyze the glucuronidation of numerous endogenous or exogenous small compounds. The aberrant expression of *UGT* isoforms causes many diseases, such as hyperbilirubinemia and affect drug efficacy or toxicity. Understanding mechanisms of *UGT* gene regulation will provide scientific foundations for disease prevention and personalized or precision medicine. Vertebrate *UGT* family genes can be divided into *UGT1* and *UGT2* subfamilies. Similar to the protocadherin, immunoglobulin, and T-cell receptor gene clusters and different from the *UGT2* gene cluster, the *UGT1* gene cluster is organized into variable and constant regions. The *UGT1* variable region contains a tandem array of variable exons, each of which can be alternatively spliced to a single set of 4 downstream constant exons, generating at least nine *UGT1* mRNAs that could be translated into different UGT1 glucuronyltransferase isoforms. Our previous work reveals that the relative orientations and locations of CTCF binding sites play a key role in the three-dimensional organization of the mammalian genomes in cell nuclei. Thus in order to study the transcriptional mechanisms of *UGT1* gene cluster, the distributions and orientations of CTCF binding sites (CBSs) are analyzed and compared between human and mouse *UGT1* gene clusters. We find that the CBSs in the *UGT1* gene cluster are not conserved between human and mouse species. We show that CTCF and cohesin regulate the transcription of the *UGT1* gene cluster by knocking down the CTCF or the cohesin subunit SMC3 in the human A549 cell line. By using CRISPR DNA-fragment editing, we deleted and inverted *hCBS1*. By RNA-seq experiments, we find that *hCBS1* deletion results in a significant decrease of levels of the *UGT1A6*, *UGT1A7*, and *UGT1A9* gene expression and that *hCBS1* inversion results in a significant decrease of levels of the *UGT1A7* gene expression. Our data suggest that the CTCF binding site *hCBS1* plays an important regulatory role in the regulation of *UGT1* gene expression, providing an experimental basis for further mechanistic studies of the 3D genome regulation of the *UGT1* gene cluster.

Keywords: *UGT1*; CTCF; CRISPR-Cas9; DNA fragment editing; transcriptional regulation

脊椎动物基因组包含一系列能够编码多种蛋白质的特殊基因簇，它们由可变区和恒定区组成[1,2]，这些基因簇包括免疫球蛋白(*Ig*)、T 细胞受体(*Tcr*)、原钙粘蛋白(*Pcdh*)和尿苷二磷酸葡糖醛酸转移酶 1 (*UGT1*)等[1,3]。其中, *UGT1* 基因簇编码的 UGT (UDP-glucuronosyltransferase)蛋白质家族是一类II相药物代谢酶，这类酶能够将葡萄糖醛酸供体基团转移到各种内外源疏水小分子化合物受体底物上，将它们转化为水溶性物质而排出体外，或者影响小分子药物的药代动力学和药物最终的生物学效应[4]。内源性底物包括许多在机体内起重要作用的小分子化合物，如胆红素、胆汁酸、脂酸、类固醇、甲状腺激素和脂溶性维生素等；外源性底物则包括小分子药物、环境污染物和致癌化合物等。它们先经过I相代谢酶 P450 氧化成极性化合物, 再被 UGT 等II相代谢酶进一步极化后排出体外。

哺乳动物 UGT 分为两大类：UGT1 和 UGT2[3,5]，分别被 *UGT1 和 UGT2* 两个基因簇编码。其中 UGT2

又可分为两个亚家族：UGT2A 和 UGT2B。*UGT2A* 分为可变区和恒定区，可变区内有 2 个外显子，恒定区内有 5 个外显子，而 *UGT2B* 有 7 个基因，每个基因都由一串外显子单独构成[6,7]。*UGT1* 基因簇结构与 *UGT2* 基因簇明显不同，但却类似于原钙粘蛋白基因簇[8]，由可变区和恒定区组成，可变区包含 9 个成串排列的外显子，恒定区包含 4 个外显子，每一个可变区外显子都有其自身的启动子，启动子激活后转录出的信使 RNA 前体能够通过可变剪接形成 9 种不同的转录本[1,5]。每个可变区外显子编码信号肽和 N 端糖苷受体结合域，4 个恒定区外显子编码高度保守的 UDPGA 供体结合域和 C 端内质网锚定跨膜段。因此，*UGT1* 基因簇编码的葡醛酸转移酶家族可以催化大量亲脂内外源化合物与 UDP-葡萄糖醛酸的接合反应，将它们转化为亲水的葡萄糖醛酸化合物[9]。尽管 *UGT1* 基因簇结构类似于原钙粘蛋白基因簇，但其架构蛋白 CTCF (CCCTC-结合因子)结合位点的分布特征与原钙粘蛋白基因簇截然不同，解析 *UGT1* 复杂基因簇的三维转录调控机制将会进一步拓宽人们对拓扑结构域(topological domain)内这一类复杂基因簇的转录调控的认知。

基因的表达调控和三维基因组中染色质高级结构密切相关，特别是远端增强子和目标启动子之间的特异性远程相互作用对于启动子激活至关重要。CTCF 是一种高度保守和广泛表达的转录因子，是组织发育必需的一种染色质架构蛋白[10]，在染色质折叠中起关键作用[11]，依靠 11 个锌指蛋白域(ZFs)方向性结合到人类基因组约 4 万个特定序列元件(CTCF binding sites, CBSs)上[12,13]，其中 60%以上的结合都没有组织特异性(tissue-invariant)[14~16]。早期认为 CTCF 起到转录抑制作用，后来发现它也具有转录激活功能，近年来发现其在三维基因组架构中起到关键作用。总之，CTCF 蛋白具有多种复杂甚至相反的功能，包括转录激活和抑制、基因印迹、RNA 聚合酶暂停、选择性剪接、DNA 复制和修复、染色体缩合或凝结(chromosome condensation)和易位(translocation)、X 染色体失活、肿瘤发生、免疫系统 V(D)J 重组和神经系统启动子选择等[12,13]。CTCF 蛋白可能通过自二聚化[17]，或与 cohesin 蛋白质复合体相互作用参与形成染色质环[18~22]，一般成环的一对 CTCF 结合位点的方向是相向的[13,23~25]。染色质架构蛋白 CTCF 通过方向性结合决定着染色质的环化方向[25]。在与 *UGT1* 基因簇结构高度相似的原钙粘蛋白基因簇中，CTCF 通过方向性识别各增强子和启动子，形成长距离染色质环化结构，从而实现了对原钙粘蛋白基因时空表达的精密调控[23]。

利用 II 型成簇规律间隔短回文重复系统 CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated nucelase 9)发展起来的基因组编辑技术为探索染色质环化机制和研究染色质高级结构提供了良好的技术平台[26,27]。CRISPR 基因组编辑依赖于 sgRNA (single-guide RNA)将 Cas9 酶定位到靶基因并进行切割，然后通过非同源末端连接(nonhomologous end-joining, NHEJ)、同源重组(homologous recombination, HR)、微同源介导的末端连接(microhomology-mediated end joining, MMEJ)或单链退火(single-strand annealing, SSA)方式修复切割末端[13,28,29]。采用双位点的 DNA 片段编辑(DNA fragment editing)方法可以删除、反转或重复所研究的目标片段[13,30]。本研究通过敲低细胞中 CTCF 或 SMC3(cohesin 亚基)的转录水平并利用 CRISPR 介导的 DNA 片段编辑技术对人类 *UGT1* 基因簇中保守的 CBS 元件进行编辑，旨在探索 *UGT1* 基因簇中 CTCF 结合位点对 *UGT1* 基因簇转录调控的影响，对研究 *UGT1* 基因表达调控网络的结构基础具有重要意义。

1　材料与方法

1.1　实验材料

人肺癌细胞系 A549 由上海交通大学电信学院仪器系付华林老师提供；人胚肾细胞系 293T 购买于中国科学院细胞库；胎牛血清、青霉素/链霉素双抗和 DMEM 购自美国 Gibico 公司；psPax2 和 pMD2.G 质粒购自美国 Addgene 公司；pLKO.1 质粒购自美国 Sigma 公司；CTCF、SMC3 抗体购自英国 Abcam 公司；Cas9 质粒由北京大学席建忠教授馈赠；pGL3-U6-sgRNA-puro 质粒由南京大学黄行许教授馈赠；逆转录试剂购自南京诺唯赞生物科技有限公司；荧光定量 PCR 试剂购自瑞士 Roche 公司；Annealing bu-

ffer、T4 DNA ligase 以及 RNA-seq 试剂盒购自美国 NEB 公司；Lipo3000 购自美国 Invitrogen 公司；rTaq 酶购自日本 TaKaRa 公司；质粒中抽试剂盒购自德国 Qiagen 公司；Trizol 购自美国 Ambion 公司；所有引物均由上海生物工程股份有限公司合成。

1.2 CTCF 方向性及结合基序分析

ChIP-seq 数据来自 NCBI，抽取富集 CTCF 的 DNA 序列。所分析的人源细胞系数据包括人肺癌细胞系 A549 (GSM822289)、人肝癌细胞系 HepG2 (GSM733645)、人慢性髓系白血病细胞系 K562 (GSM749733)、人胚肾细 CBS 胞系 HEK293 (GSM-749668)、正常人表皮角质形成细胞 NHEK (GSM-749707)、人乳腺癌细胞系 MCF7 (GSM1022658)和人子宫内膜腺癌细胞系 HEC-1-B (来自于本课题组)。所分析的小鼠细胞系包括 B 细胞淋巴瘤 CH12 (GSM923568)、GATA 1-红系祖细胞 G1E (GSM923570)和第 0 天的胚胎干细胞 ES-B4 E0 (GSM918748)；小鼠组织包括 8 周龄心脏(GSM918756)、肾脏(GSM-918731)、肝脏(GSM918715)和肺(GSM918722)。

利用在线软件 CTCF BSDB2.0 预测每个 CTCF 结合的 motif，并判别其方向[31] (http://insulatordb.uthsc.edu/storm_new.php)。利用软件 Clustal X 和在线软件 BoxShade 对 CTCF 结合基序进行序列分析比对(https://embnet.vital-it.ch/software/BOX_form.html)。

1.3 cohesin 与增强子标志分析

ChIP-seq 数据来自 NCBI，分析 cohesin 亚基 SMC3 及增强子标志 p300 和 H3K27ac 在不同人源细胞系的分布。人肺癌细胞系 A549 的 SMC3、p300 和 H3K27ac GEO 序列号分别是 GSM3106366、GSM-1010827 和 GSM2421872；人肝癌细胞系 HepG2 的 SMC3、p300 和 H3K27ac GEO 序列号分别是 GSM-935542、GSM935545 和 GSM733743；人宫颈癌细胞系 HeLa-S3 的 CTCF、SMC3、p300 和 H3K27ac GEO 序列号分别是 GSM822285、GSM935384、GSM-935500 和 GSM733684。

1.4 细胞培养

A549 细胞和 293T 细胞的培养基配方为 DMEM 中加入 10%的胎牛血清和 1%的青霉素/链霉素双抗。培养条件为 37℃、5% CO_2。

1.5 shRNA 慢病毒包装和感染细胞

shRNA (short hairpin RNA)引物序列采用文献[23]中的序列(表 1)。寡核苷酸的两条链经过变性、退火后连接到 pLKO.1 载体，通过转化感受态细菌和质粒抽提得到 shRNA 的重组质粒。

293T 细胞用无抗培养基传代到 10 cm 培养皿，密度为 8×10^5 cells/mL。第 2 天用 Lipo3000 转染质粒 DNA (5.4 μg psPax2、0.6 μg pMD2.G、6 μg 载体质粒)，24 h 后换为正常培养基，分两次回收病毒悬液，过滤得到病毒初始液，加入 5×PEG 母液，混匀、4℃放置过夜后 4℃、4000×*g* 离心 30 min，得到 shRNA 病毒沉淀，加入适量 DMEM 溶解沉淀并分装存于−80℃。

在六孔板准备 A549 细胞，密度为 1×10^5 cells/mL。第 2 天用慢病毒感染细胞，培养 24 h 后换培养基，48 h 后收集细胞。

1.6 Western blot 鉴定

收集细胞样品中加入 100 μL RIPA 裂解液，冰上放置 30 min，4℃、13000×*g* 离心 15 min，收集上清并分装，液氮速冻后放−80℃冰箱待用。配制 10%分离胶：4 mL MilliQ 水、2.4 mL 1.5 mol/L Tris-HCl (pH 8.8)、3.4 mL 30% Arc-Bis、100 μL 10% SDS、100 μL 10%过硫酸铵、10 μL 四甲基乙二胺。4%积层胶包括：3 mL MilliQ 水、500 μL 1 mol/L Tris-HCl (pH 6.8)、522 μL 30% Arc-Bis、40 μL 10% SDS、40 μL 10%过硫酸铵、6 μL 四甲基乙二胺。取 10 μL 蛋白样品于 PCR 管中，补 MilliQ 水至 16 μL，加入 4 μL 5× SDS Loading buffer 混匀后，95℃ 5 min。点样并电泳，首先 80 V 电泳 30 min，然后 120 V 电泳 2 h，剪切所需蛋白胶，按照滤纸-硝酸纤维素滤膜-凝胶-滤纸的顺序叠加，在电泳槽中 0.16 A 电泳 2 h。然后将印记后的膜转入 5%的脱脂牛奶中，室温避光摇晃 1 h 后，按照目的条带大小剪切膜，分别加入一抗稀释液，4℃摇晃过夜。加入 PBST，室温摇晃 5 min，重复 3 次。加入 1 mL 二抗稀释液，室温摇晃 1 h，再用 PBST 洗膜 3 次。最后用 Odyssey 双色红外激光成像系统扫膜成像。

表 1　本研究使用的引物序列

Table 1　Primers sequences used in this study

类型	引物	序列(5′→3′)
qPCR	hCTCF F	TCTGGACGACAATGAGGATGAG
	hCTCF R	GCACCTGTATTCTGGTCTTCAAC
	1A6F1	CAACTGTAAGAAGAGGAAAGAC
	1A7F1	CCCCTATTTTTTCAAAAATGTCTT
	1A9F1	GAACATTTATTATGCCACCG
	1AR	ATTGATCCCAAAGAGAAAACCAC
	hSMC3 F	AATGATGTGATGAACCTCCTTG
	hSMC3 R	TGAGAATCTGGTGCTGTTGC
PCR	C1F	GGCTGCTCCTCCTATTTCCTG
	C1F2	GACCAGGACAAGGAGGCGT
	C1F5	TCTCCCCAGTGCTCCCATGT
	C1F8	AGGTTTGGGTGGAAAGAGGT
	C1R	GCAAGTCTGCTTAGAGACTGAAGTG
	C1R2	ACCTGGGAGGGTGCTCTCA
	C1R6	CCTTAAAACTGTATGCATCCTAAGC
	C1R9	GTGAGGTCACCCCAGAAAAA
sgRNA	sgC1fF	ACCGAATCATTCAGATCAGGCTAT
	sgC1fR	AAACATAGCCTGATCTGAATGATT
	sgC1uF	ACCGGGTCTAGGAGTCCTAGACGT
	sgC1uR	AAACACGTCTAGGACTCCTAGACC
	sgC1dF	ACCGTTAAACGTGTGCCTTTCTAA
	sgC1dR	AAACTTAGAAAGGCACACGTTTAA
shRNA	pLKO.1-hCTCF-1F	CCGGGAGGAGCCTGCCGTAGAAATTCTCGAGAATTTCTACGGCAGGCTCCTCTTTTTG
	pLKO.1-hCTCF-1R	AATTCAAAAAGAGGAGCCTGCCGTAGAAATTCTCGAGAATTTCTACGGCAGGCTCCTC
	pLKO.1-hSMC3-1F	CCGGGCAGTGCAACACAGAATTAAACTCGAGTTTAATTCTGTGTTGCACTGCTTTTTG
	pLKO.1-hSMC3-1R	AATTCAAAAAGGCAGTGCAACACAGAATTAAACTCGAGTTTAATTCTGTGTTGCACTGC
P7-index	P7-index-wyh14	CAAGCAGAAGACGGCATACGAGATGCTCCGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-wyh15	CAAGCAGAAGACGGCATACGAGATCATGAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-wyh18	CAAGCAGAAGACGGCATACGAGATCTGAATGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-wyh20	CAAGCAGAAGACGGCATACGAGATTCGCATGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-wyh21	CAAGCAGAAGACGGCATACGAGATAATAGCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
	P7-index-wyh22	CAAGCAGAAGACGGCATACGAGATGTCGCGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

sgRNA 引物序列中下划线部分为构建质粒所需的粘性末端。P7-index 引物序列中下划线部分为 index。

1.7　各细胞系 RNA 的提取

对 12 孔板中长满的细胞进行胰酶消化后，3000×*g* 离心 10 min，收集沉淀加入 500 μL Trizol 溶液，室温放置 5 min，再加入 100 μL 氯仿，剧烈摇 15 s, 室温放置 3 min 后, 4℃、12000×*g* 离心 15 min，取上层透明相于干净 1.5 mL 的 EP 管中。在该管中加入 250 μL 异丙醇，充分混匀，室温放置 10 min 后，4℃、12000×*g* 离心 10 min，去上清收集沉淀，用 75%的乙醇洗沉淀，4℃、7500×*g* 离心 5 min，收集沉淀，倒置 10 min，最后加入无核酸酶的水 30 μL 溶解沉淀得到 RNA 溶液。用 NanoDrop2000 测 RNA

浓度，并放−80℃冰箱待用。

1.8 荧光定量 PCR 检测表达量

RNA 逆转录为 cDNA 的方法为：在无核酸酶的 PCR 管加入 4 μL gDNA wiper 混合液、0.5 μg RNA 模板, 补无核酸酶水至总体积 16 μL, 轻轻吹打混匀, 42℃ 2 min。继续加入 4 μL 5×HiScript II qRT SuperMix II, 轻轻吹打混匀, 50℃ 15 min；85℃ 5 s。产物即可用于 qPCR 反应或存于−20℃冰箱待用。

qPCR 反应体系：5 μL 2×SYBR Green Master 混合液、0.3 μL 正向引物(10 μmol/L)、0.3 μL 反向引物(10 μmol/L)、2 μL 模板、2.4 μL MilliQ 水。所用到正、反向引物序列见表 1。qPCR 扩增条件：95℃ 10 min；95℃ 15 s, 60℃ 33 s, 40 个循环；95℃ 15 s, 60℃ 1 min, 95℃ 15 s。分析其 ΔΔCt 值及 RQ (relative quantity)值，$RQ=2^{-\Delta\Delta Ct}$。显著性差异分析使用 *t*-test。

1.9 CBS 元件的基因片段编辑

根据 CRISPR/Cas9 基因编辑原理设计 sgRNA 并合成(表 1)。每一对 sgRNA 的两条链经过变性、退火后连接到 pGL3-U6-sgRNA-puro 载体上。通过转化感受态细菌和质粒抽提得到 sgRNA 的重组质粒。

hCBS1 反转和删除的单克隆细胞株的获得采用了两组 sgRNA。在 24 孔板准备细胞，密度为 8×10^5 cells/mL。第 2 天用 Lipo3000 转染 sgRNAs (均为 169.9 ng)和 Cas9 (940.392 ng)。转染 18 h 后换培养基，24 h 后加嘌呤霉素(1 mg/mL)，每孔 1 μL，连续 4 天，将存活下来的细胞继续培养。待细胞足够多时，取一部分做模板。设计 PCR 引物，进行细胞鉴定。鉴定到目标基因型后，将该混合细胞逐渐放大培养。按照 30 cells/mL 的浓度接种到 96 孔板进行单克隆化。两周后对单克隆细胞进行鉴定。在此过程，要保证这些单克隆细胞生长状态良好。

1.10 CRISPR 单克隆细胞的鉴定

采用细胞裂解法和 PCR 方法鉴定单克隆细胞。细胞 DNA 模板的制备方法：取部分细胞于 PCR 管，3000×*g* 离心 10 min，留下细胞沉淀；加入 20 μL 裂解液，沸水煮 5 min；再加入 20 μL 中和液，混匀即可。可存于−20℃。裂解液配方为 25 mmol/L NaOH、0.2 mmol/L EDTA；中和液配方为 40 mmol/L Tris-HCl。

PCR 反应体系：6.3 μL MilliQ 水、1 μL 10× Buffer (含 Mg^{2+})、1 μL dNTP 混合液(各 2.5 mmol/L)、0.3 μL 正向引物(10 μmol/L)、0.3 μL 反向引物(10 μmol/L)、0.1 μL rTaq 酶(5 U/μL)和 1 μL DNA 模板。所用到正、反向引物序列见表 1。PCR 扩增条件：94℃ 3 min；94℃ 15 s, 60℃ 30 s, 72℃ 1min, 38 个循环；72℃ 7 min (退火温度根据引物 Tm 适当调整，延伸时间根据产物长度调整)。PCR 扩增产物用 2%琼脂糖凝胶进行电泳分析鉴定。

1.11 RNA-seq 建库和高通量测序

待检测细胞系 RNA 的抽提方法和上面方法相同。poly(A) mRNA 的分离、cDNA 合成和文库构建根据 NEB RNA-sequencing 试剂盒说明书操作, 所用 P7-index 序列见表 1。文库采用 Invitrogen 公司的 Qubit 3 Fluorometer 仪器进行定量, 并送至苏州金唯智生物科技有限公司进行质检和高通量测序，测序仪器为 Illumina HiSeq 平台，测序长度双端 150 bp。RNA-seq 数据用 TopHat 和 Cufflinks 进行分析[32]，显著性差异分析使用 *t*-test。

2 结果与分析

2.1 CTCF 和 cohesin 调控 *UGT1* 基因簇表达

CTCF 在细胞中的多样性功能与其在三维基因组中长距离染色质环化作用相关。本研究分析了多类人源细胞系的 CTCF ChIP-seq (chromatin immunoprecipitation and massive parallel sequencing)数据，发现人 *UGT1* 基因簇中存在 3 个 CTCF 可结合位点：*hCBS1~3*，其中 *hCBS1* 位点在所分析的人肺癌细胞系(A549)、人肝癌细胞系(HepG2)、人慢性髓系白血病细胞系(K562)、人胚肾细胞系(HEK293)、正常人表皮角质形成细胞系(NHEK)、人乳腺癌细胞系(MCF7)和人子宫内膜腺癌细胞系(HEC-1-B)中均高度富集 CTCF 蛋白(图 1A)。

CBS 基序通常被分为 4 个模块(Module#1~4)[25,33]，

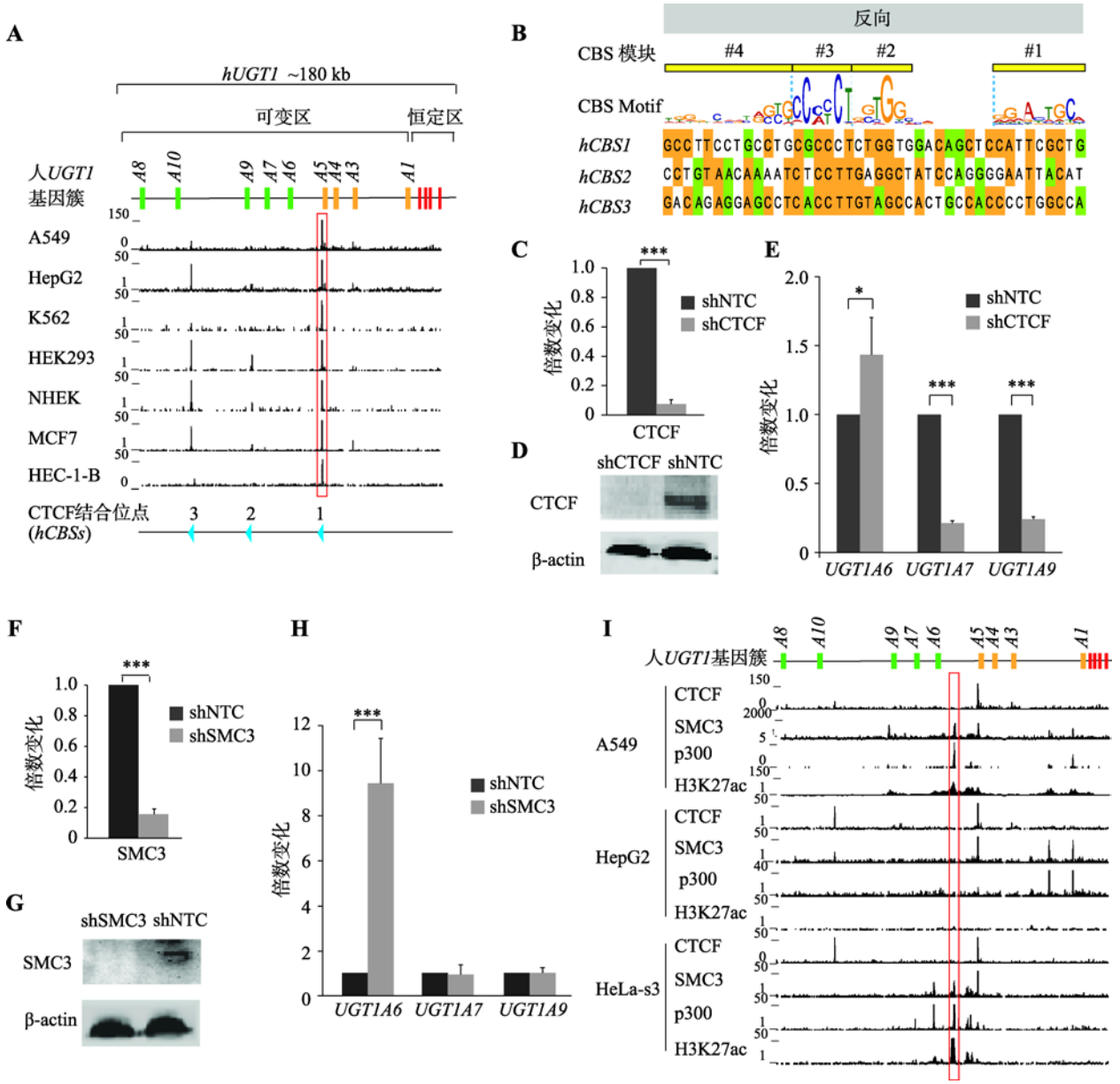

图 1　CTCF 和 cohesin 参与人 *UGT1* 基因簇的转录调控

Fig. 1　CTCF and cohesin regulate the transcription of human *UGT1* gene cluster

A：人 *UGT1* 基因簇内 CBS 分布。人 *UGT1* 位于 2 号染色体，类似于原钙粘蛋白基因簇，由可变区和恒定区组成，可变区内串联排列着 9 个高度相似的可变区外显子，下游恒定区包含 4 个恒定区外显子。垂直框代表 *UGT1* 基因簇外显子并用不同颜色标记，绿色表示苯酚组可变外显子，橘黄色表示胆红素组可变外显子，红色表示恒定外显子；人肺癌细胞系 A549、人肝癌细胞系 HepG2、人慢性髓系白血病细胞 K562、人胚肾细胞 HEK293、正常人表皮角质形成细胞 NHEK、人乳腺癌细胞 MCF7 和人子宫内膜腺癌细胞 HEC-1-B 的 CTCF ChIP-seq 数据显示人 *UGT1* 基因簇内 3 个 CTCF 可结合位点：*hCBS1~3*，每个 CBS 方向由下方蓝色三角形指示；B：人 *UGT1* 基因簇 3 个 CBS 序列及方向性预测；Module#4~1 的 CBS 定义为反向 CBS；C：荧光定量 PCR 技术检测到 A549 细胞中敲低 CTCF 后 CTCF 的转录水平显著降低，shNTC 是对照组；D：蛋白免疫印迹实验证明 CTCF 蛋白被敲低；E：荧光定量 PCR 技术检测敲低 CTCF 对 A549 细胞中高度表达的 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 三个基因的转录的影响；F：荧光定量 PCR 技术检测 A549 细胞中敲低 SMC3 后 SMC3 的转录水平显著降低；G：蛋白免疫印迹实验证明 SMC3 蛋白被敲低；H：荧光定量 PCR 技术检测敲低 SMC3 对 A549 细胞中高表达基因 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 的转录影响；I：人肺癌细胞系 A549、肝癌细胞系 HepG2 和宫颈癌细胞系 HeLa-S3 的 cohesin 蛋白亚基 SMC3 以及增强子标志 p300 和 H3K27ac 在人 *UGT1* 基因簇的分布。*：$P<0.05$ 表示有统计学差异，**：$P<0.01$ 表示有显著的统计学差异，***：$P<0.001$ 表示有非常显著的统计学差异。

核心序列为 Module#2~4, 其中 Module#2~3 由 CTCF 的 ZF4~7 特异识别, Module#4 由 ZF3 识别。核心序列上有一段保守的 11 bp 回文序列“CCACCAGGTGG”, 位于 Module#2~3 内, 中心的核苷酸碱基“A”被 CTCF 蛋白 ZF6 上的 Gln418 残基特异性识别, 对鉴别 CBS 的方向至关重要。约 15%的 CBS 在核心序列上游存在 Module#1, CTCF 的 ZF9~11 缠绕在 Module#1 的 DNA 双链上并将它们的 α-螺旋插入 Module#1 的大沟里, 对 CTCF 方向性结合也至关重要[34]。根据上述 CBS 的结合特征, 结合 BSDB2.0 分析软件, 预测人 *UGT1* 基因簇中 *CBS1~3* 的方向为反向(图 1B), 它们在核心序列 Module#2~3 处保守度较高(图 1B)。

A549 细胞系中 *UGT1* 基因簇各基因的表达差异较大, *UGT1A6*、*UGT1A7* 和 *UGT1A9* 高度表达, *UGT1A1* 和 *UGT1A3* 低度表达, 而 *UGT1A2*、*UGT1A4*、*UGT1A5*、*UGT1A8* 和 *UGT1A10* 则处于沉默状态。为研究 CTCF 是否参与调控 *UGT1* 基因簇的转录表达, 以人肺癌细胞系 A549 为模型, 通过向细胞感染表达 shRNA 的慢病毒来敲低 CTCF。CTCF 靶向的 shRNA 病毒感染 A549 细胞后, CTCF 的转录水平降低了 93% (图 1C), CTCF 的蛋白表达水平显著降低(图 1D)。进一步采用荧光定量 PCR 技术检测了 *UGT1* 的表达, 发现 *UGT1A6* 转录水平升高了 43%, *UGT1A7* 和 *UGT1A9* 分别降低了 79%和 76% (图 1E)。表明 CTCF 显著影响 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 的表达, 参与调控 *UGT1* 基因簇的转录表达。

CTCF 在染色质高级架构中的绝缘作用依赖于 cohesin 蛋白复合体, 它们往往共定位于染色质上的 CBS 处, 本研究进一步利用慢病毒介导的 shRNA 敲低了 SMC3 的表达, SMC3 的转录水平降低 86.7% (图 1F)、蛋白量显著降低(图 1G)。SMC3 敲低后, *UGT1A6* 转录水平升高 8.25 倍, 而 *UGT1A7* 和 *UGT1A9* 转录水平几乎没有变化(图 1H), 这和 CTCF 敲低时的情况不一样, 可能与 cohesin 在组织特异性转录中的 CTCF 非依赖性功能相关[20,35,36]。于是, 进一步分析了人肺癌细胞系 A549、肝癌细胞系 HepG2 和宫颈癌细胞系 HeLa-S3 的 cohesin 亚基 SMC3 蛋白以及增强子标志 p300 和 H3K27ac 在 *UGT1* 基因簇的结合分布, 结果发现 SMC3 除了在 CTCF 位点富集, 还在 p300 和 H3K27ac 位点处富集(图 1I)。

2.2 CRISPR 删除 *hCBS1* 片段对人细胞系 A549 中 *UGT1* 基因簇转录的影响

为进一步研究人 *UGT1* 基因簇内保守的 CTCF 结合元件 *hCBS1* 对 *UGT1* 基因簇的转录调控的影响, 以人肺癌细胞系 A549 为模型, 利用 CRISPR/Cas9 片段编辑技术原位删除包含 *hCBS1* 的 DNA 片段。

本研究采用使用最广泛的酿脓链球菌(*Streptococcus pyogenes*)来源的 SpCas9 酶。该酶被 sgRNA 带入靶标 DNA 处, 其两个核酸酶结构域 HNH 和 RuvC 在 5′ PAM (protospacer adjacent motif)序列(NGG)上游 3 bp 处分别对互补链和非互补链进行切割, 产生具有平头末端的 DSBs (double strand breaks)[37~40]。最新的研究发现, SpCas9 切割位点不局限于 PAM 上游 3 bp 处, 非互补链上的切割可能发生在更上游位置, 进而产生突出末端[27]。在 A549 细胞中转入 SpCas9 和 sgRNAs (sgC1F 和 sgC1D, 图 2A)表达质粒后, 在转染细胞群中检测到 *hCBS1* 删除的基因型, 进一步对转染细胞群进行单克隆化, 获得 89 个单克隆细胞株, 对这些单克隆细胞进行 PCR 鉴定, 发现 C1D2 和 C1D3 是 *hCBS1* 删除纯合型单克隆细胞株。*hCBS1* 片段删除示意图如图 2B 所示, SpCas9 在 sgCIF 和 sgC1D 引导下, 特异性识别 *hCBS1* 片段两侧靶标位置并切割产生两个切口, 两个切口连接到一起形成 *hCBS1* 片段删除的编辑细胞。

C1D2 和 C1D3 的基因型鉴定结果如图 2C 所示, 切口处原野生型片段(上游 C1F8/C1R9 引物对 PCR 产物 151 bp、下游 C1F2/C1R 引物对 PCR 产物 301 bp)全部消失, 反转检测条带(C1F8/C1F2 引物对 PCR 产物 219 bp、C1R/C1R9 引物对 PCR 产物 233 bp)呈阴性, *hCBS1* 片段两侧引物 C1F8/C1R 扩增产物从野生型细胞的 1229 bp 缩短为 290 bp, 表明编辑片段被删除。对 *hCBS1* 片段两侧引物 C1F8/C1R 扩增的 290 bp 产物进一步测序分析, 发现 C1D2 和 C1D3 单克隆细胞株中删除片段上下游的切割位点均恰好在 PAM 上游 3 bp 处(图 2D), 这可能因为即使 SpCas9 在非互补链造成的切割位点超过 PAM 上游 3 bp, 细胞内修复体系也会使其 3′端补齐, 然后再连接到一起。

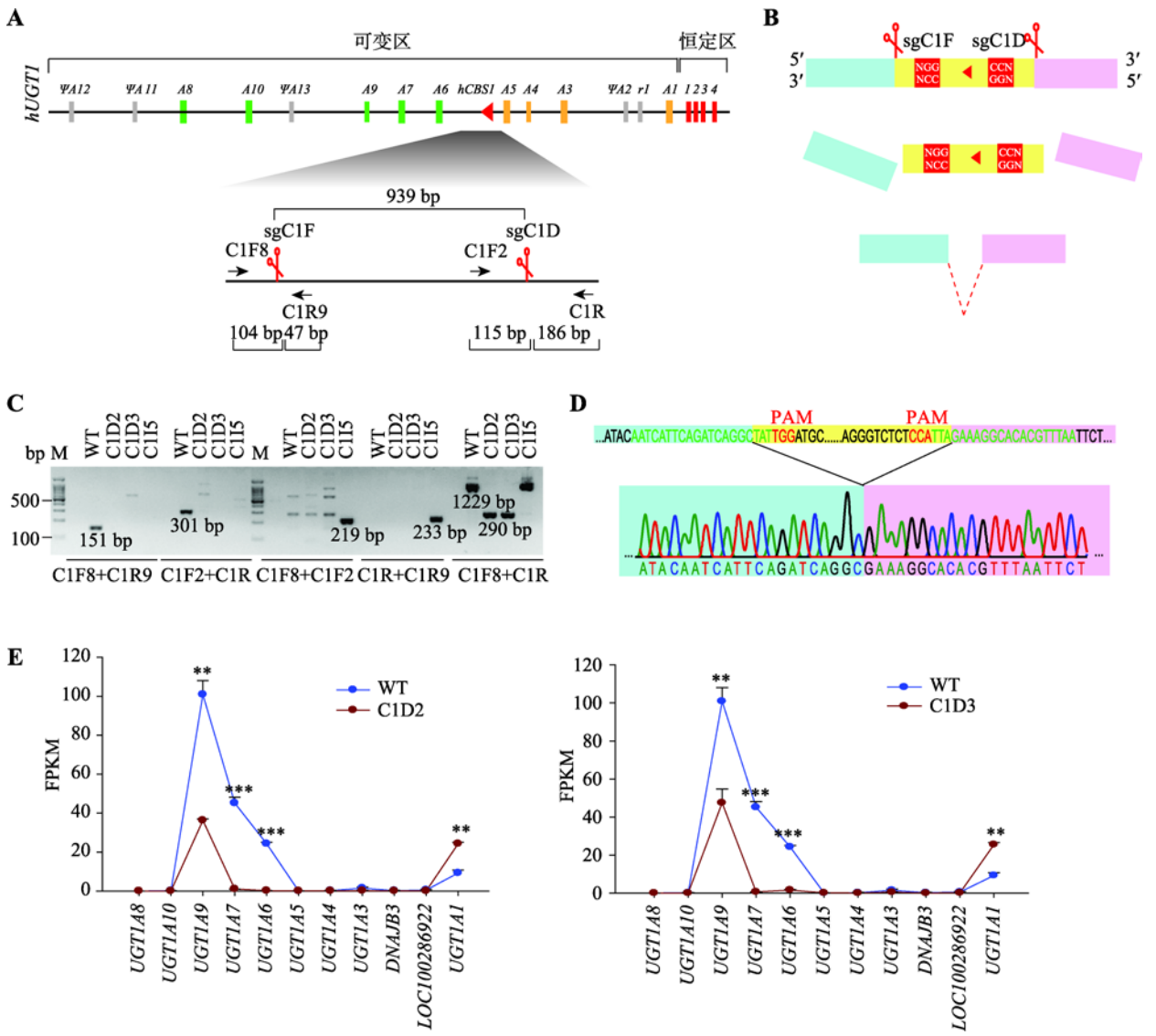

图 2　CRISPR 删除 *hCBS1* 片段对人细胞系 A549 中 *UGT1* 基因簇转录的影响

Fig. 2　Effect of CRISPR deletion of *hCBS1* fragment on the transcription of human *UGT1* gene cluster in A549 cell line

A：CRISPR 删除 *hCBS1* 片段所使用的成对 sgRNA (sgCIF 和 sgC1D)以及各鉴定引物的位置示意图；B：*hCBS1* 片段删除示意图。Cas9 在 sgCIF 和 sgC1D 引导下，特异性识别 *hCBS1* 片段两侧靶标位置并切割产生两个切口，两个切口连接到一起形成 *hCBS1* 片段删除的编辑细胞；C：*hCBS1* 片段删除的单克隆 A549 细胞株 C1D2 和 C1D3 的基因型鉴定结果。野生型 A549 细胞(WT)和 *hCBS1* 片段原位反转型 A549 细胞(C1I5)被作为 PCR 鉴定对照；用 C1F8/C1R9 或 C1F2/C1R 引物对可在 WT 中扩增出切点处 151 bp 或 301 bp 片段，用 C1F8/C1F2 或 C1R/C1R9 可在 C1I5 中扩增出反转后剪接处的 219 bp 或 233 bp 片段，C1D2 和 CID3 中未检测到以上 4 个片段；用 *hCBS1* 片段两侧引物 C1F8/C1R 可在 WT 和 C1I5 中扩增出 1229 bp 条带，C1D2 和 CID3 中扩增条带缩短为 290 bp；D：C1F8/C1R 在 C1D2 和 C1D3 细胞中的扩增产物的测序结果示意图；E：RNA-seq 数据分析比较 WT 和 C1D2 或 C1D3 细胞中 *UGT1* 基因簇的转录水平。

采用 RNA-seq 分析方法检测比较 C1D2 和 C1D3 单克隆与野生型 A549 细胞的转录水平(图 2E)，发现在 C1D2 单克隆细胞中，*UGT1A9* 表达降低 64%，*UGT1A6* 表达降低 99%，*UGT1A7* 表达降低 97.6%，但是 *UGT1A1* 表达升高 1.62 倍，*UGT1* 基因簇其他基因没有变化。在 C1D3 单克隆细胞中，*UGT1A9* 表达降低 52%，*UGT1A6* 表达降低 93%，*UGT1A7* 表达降低 99%，但是 *UGT1A1* 表达升高 1.75 倍，*UGT1* 基因簇其他基因没有变化，这与 C1D2 克隆结果一致。总之，*hCBS1* 删除显著影响 *UGT1* 基因簇

的基因转录。

2.3 CRISPR 反转 *hCBS1* 片段对人细胞系 A549 中 *UGT1* 基因簇转录的影响

CTCF 在 DNA 上的结合具有方向性，参与形成染色质环化，在三维基因组复杂的空间结构的形成和维持中起关键作用，上述研究证明删除 *hCBS1* 显著影响 *UGT1* 基因簇的转录，表明 *hCBS1* 是 *UGT*1 基因簇的一个重要的转录调控元件。*hCBS1* 是反向的，那么如果把 *hCBS1* 反转，变为正向，*UGT1* 基因簇局部的基因表达会有怎样的变化呢？

本研究在 A549 细胞中转入 SpCas9 和 sgRNAs (sgC1U 和 sgC1D，图 3A)表达质粒后，在转染细胞群中检测到 *hCBS1* 反转的基因型，进一步对转染细胞群进行单克隆化，获得 90 个单克隆细胞株，对这些单克隆细胞进行基因型鉴定，发现 C1I4 单克隆细胞株为 *hCBS1* 反转纯合型。*hCBS1* 片段反转示意图如图 3B 所示，Cas9 在 sgCIU 和 sgC1D 引导下，特异性识别 *hCBS1* 片段两侧靶标位置并切割产生两个切口，*hCBS1* 反转后两个切口分别连接到染色体断点处形成 *hCBS1* 片段反转的编辑细胞。

C1I4 的基因型鉴定结果如图 3C 所示，切口处原野生型片段(上游 C1F5/C1R6 引物对 PCR 产物 244 bp、下游 C1F2/C1R 引物对 PCR 产物 301 bp)检测为阴性，表明野生基因型不存在；用 *hCBS1* 反转检测引物对 C1F/C1F2 或 C1R/C1R2 可在 C1I4 中扩增出 257 bp 或 289 bp 大小的反转条带，表明 *hCBS1* 发生反转；用 *hCBS1* 片段两侧引物 C1F/C1R 可扩增出和野生型条带大小相当约 977 bp 的条带。对 *hCBS1* 片段两侧引物 C1F/C1R 扩增的产物进一步测序分析，发现 C1I4 单克隆细胞株中编辑片段上下游的切割位点均恰好在 PAM 上游 3 bp 处(图 3D)。

采用 RNA-seq 分析方法检测比较 C1I4 单克隆与野生型 A549 细胞中 *UGT1* 基因簇的转录水平(图 3E)，发现当 *hCBS1* 反转后，C1I4 单克隆中 *UGT1A6* 和 *UGT1A9* 表达没有明显变化，但是 *UGT1A7* 表达降低 74%，低表达的 *UGT1A1* 和 *UGT1A3* 变化幅度并不显著，*UGT1* 其余基因(*UGT1A4*、*UGT1A5*、*UGT1A8*、*UGT1A10*)依然处于沉默状态，几乎不转录。总之，*hCBS1* 反转主要影响 *UGT1A7* 基因的转录。

2.4 人和小鼠 *UGT1* 基因簇中 CBS 分布比较

人和小鼠 *UGT1* 基因簇分别长 180 kb 和 200 kb，基因结构相似，由高度同源且串联排列的 9 个可变外显子组成的可变区和 4 个恒定外显子组成的恒定区组成(图 4A)，可变区任意外显子可被选择性剪接到下游全套恒定外显子上，共可形成 9 种转录本，表达 9 种 UGT1 同工酶。根据催化底物特异性，UGT1 同工酶被分为胆红素组和苯酚组，其编码基因的可变区分别对应图 4A 中的橘黄色和绿色外显子。胆红素组中，人 *UGT1A1* 和小鼠 *Ugt1a1* 基因为直系同源关系，蛋白氨基酸序列相似度为 66%。人 *UGT1A3-5* 与小鼠 *Ugt1a2*、*1a5* 为直系同源关系，它们的平均氨基酸序列同源性约 61%。苯酚组也有两个分支，人 *UGT1A6* 与小鼠 *Ugt1a6a-b* 为直系同源关系，它们之间相似度为 70%，小鼠的 *Ugt1a6a* 和 *1a6b* 基因是在两个物种分开后复制衍生形成的，两者编码的氨基酸相似度为 95%。人 *UGT1A7-10* 与小鼠 *Ugt1a7-10* 基因直系同源，平均氨基酸序列相似度约 67%[1,41]。除了 9 个功能性可变外显子外，人和小鼠 *UGT1* 基因簇可变区内还有多片分散排列的 *UGT1* 假基因和残骸(relic)序列。

为解析 *UGT1* 基因簇内分布的 CBS 在人和模式动物小鼠间是否高度保守，本研究分析和比较了小鼠不同细胞和组织中 *Ugt1* 基因簇的 CTCF ChIP-seq 数据(图 4B)，包括 B 细胞淋巴瘤细胞系(CH12)、GATA 1-红系祖细胞系(G1E)、第 0 天的胚胎干细胞(ES-B4 E0)和 8 周龄的心脏、肾脏、肝脏、肺，发现小鼠 *Ugt1* 基因簇含有至少 11 个 CTCF 可结合位点：*mCBS1~11*，其中 8 个为正向、3 个为反向，在核心序列 Module#2~3 处保守性较高(图 4，C 和 D)。比较人和小鼠 *UGT1* 基因簇中的 CBS 分布和方向性，发现小鼠 *Ugt1* 基因簇含有 CBS 数目比人 *UGT1* 基因簇更多。

3 讨论

脊椎动物 *UGT1* 表达具有组织特异性，同一个体不同器官的 *UGT1* 表达水平有很大的差别[1,42]，这和组织特异性转录因子及其配体激活密切相关[7]。

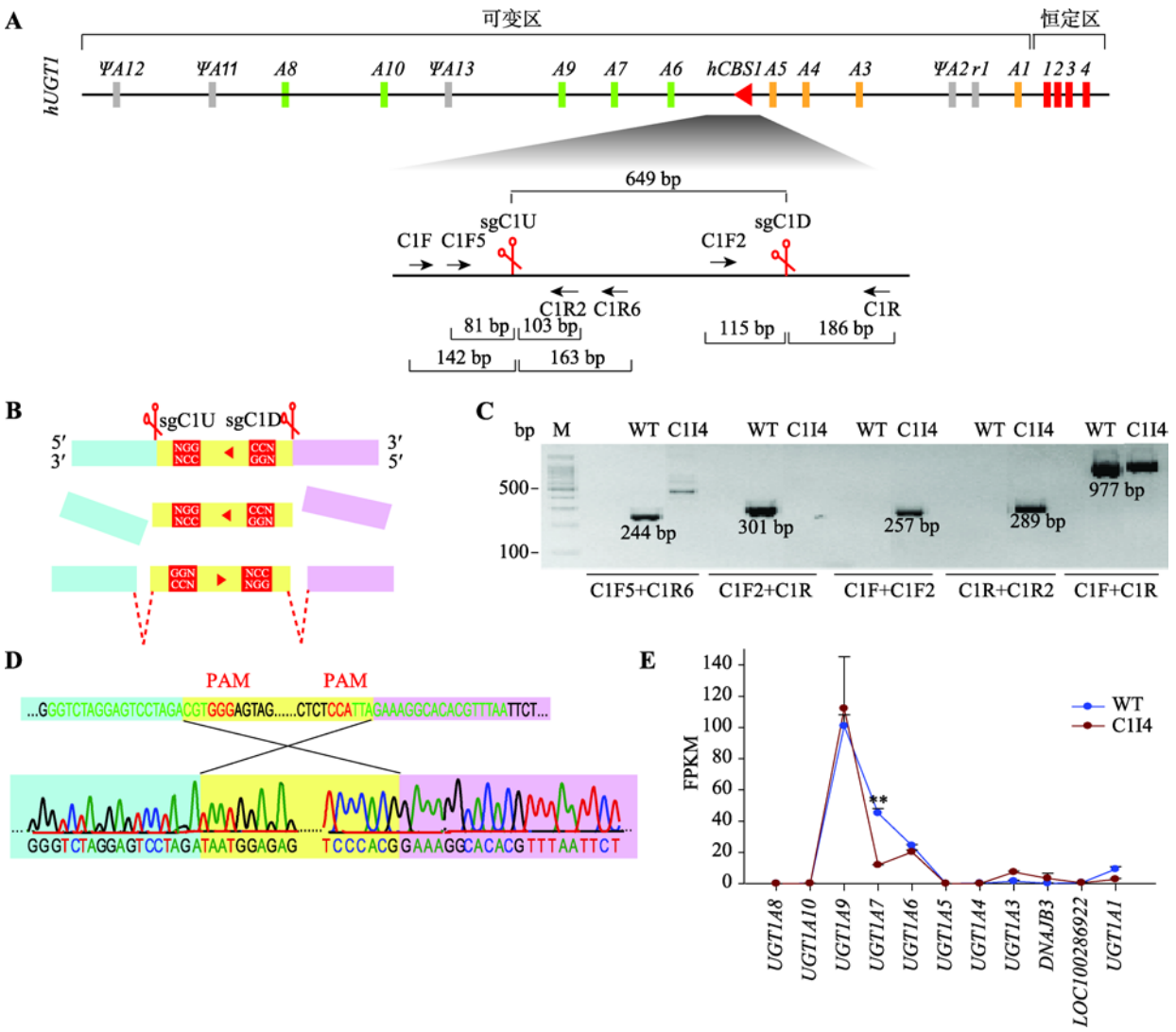

图 3　CRISPR 反转 *hCBS1* 片段对人细胞系 A549 中 *UGT1* 基因簇转录的影响

Fig. 3　Effect of CRISPR inversion of *hCBS1* fragment on the transcription of human *UGT1* gene cluster in A549 cell line

A：CRISPR 反转 *hCBS1* 片段所使用的成对 sgRNA (sgCIU 和 sgC1D)以及各鉴定引物的位置示意图；B：*hCBS1* 片段反转示意图。Cas9 在 sgCIU 和 sgC1D 引导下，特异性识别 *hCBS1* 片段两侧靶标位置并切割产生两个切口，*hCBS1* 反转后两个切口分别连接到染色体断点处形成 *hCBS1* 片段反转的编辑细胞；C：*hCBS1* 片段反转的单克隆 A549 细胞株 C1I4 的基因型鉴定结果。野生型 A549 细胞(WT)被作为 PCR 鉴定对照；用 C1F5/C1R6 或 C1F2/C1R 引物对可在 WT 中扩增出切点处 244 bp 或 301 bp 的野生型条带，C1I4 中检测结果均为阴性；用 CBS1 反转检测引物对 C1F/C1F2 或 C1R/C1R2 可在 C1I4 中扩增出 257 bp 或 289 bp 片段；用 *hCBS1* 片段两侧引物 C1F8/C1R 可在 WT、C1I4 中扩增出约 977 bp 条带；D：C1F/C1R 在 C1I4 细胞中的扩增产物的测序结果示意图；E：RNA-seq 数据分析比较 WT 和 C1I4 细胞中 *UGT1* 基因簇的转录水平。

UGT 葡醛酸转移酶的多样性和其在各种组织中的表达差异与特定疾病以及治疗药物的疗效或毒性有关，新生儿黄疸、克里格勒-纳贾尔(Crigler-Najjar)综合征(Ⅰ型和Ⅱ型)和吉尔伯氏(Gilbert)综合征等遗传性高胆红素疾病[43,44]均是由 *UGT1* 多态性或基因突变造成的胆红素代谢功能降低或缺失引起，其中大部分变异位点位于非编码区。非编码区在 *UGT1* 基因簇的转录调控中发挥重要作用[45]。为了解析这些非编码区如何在复杂的染色质三维空间结构上调控 *UGT1* 基因簇的表达以及在染色体拓扑结构域内各 *UGT1* 启动子与远端 DNA 调控元件之间的特异性成环机制，本研究首次聚焦 *UGT1* 基因簇中染色质

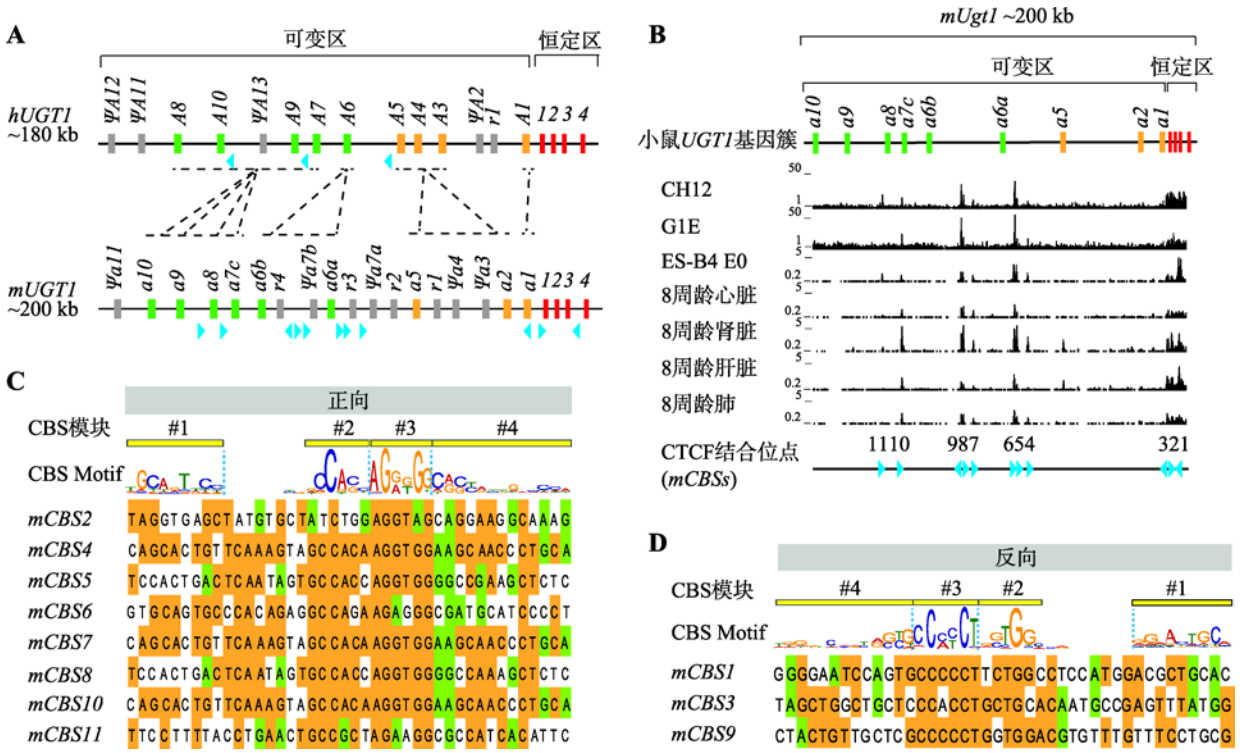

图 4 小鼠 *Ugt1* 基因簇内 CTCF 结合位点分布和方向性预测

Fig. 4 Distributions and orientations of the CTCF binding sites(CBSs)in mouse *Ugt1* gene cluster

A：人和小鼠 *UGT1* 基因簇内各同工酶基因的线性进化关系示意图。人 *UGT1* 位于 2 号染色体，小鼠 *Ugt1* 位于 1 号染色体，它们都类似于原钙粘蛋白基因簇，由可变区和恒定区组成，可变区内串联排列着 9 个高度相似的可变区外显子，下游恒定区包含 4 个恒定区外显子。人 *UGT1* 基因簇可变区还有 4 个假基因和 1 个残骸(relic)序列，而小鼠 *Ugt1* 可变区含有 5 个假基因和 4 个残骸序列。垂直框代表 *UGT1* 基因簇外显子并用不同颜色标记，绿色表示苯酚组可变外显子，橘黄色表示胆红素组可变外显子，红色表示恒定外显子，灰色表示假基因或残骸序列，分别用 Ψ 和 r 表示，直系同源基因之间由黑色虚线相连，蓝色三角形表示 CBS 方向；B：小鼠 *Ugt1* 基因簇内 CBS 分布。小鼠 B 细胞淋巴瘤 CH12、GATA 1-红系祖细胞 G1E 和第 0 天的胚胎干细胞 ES-B4 E0, 以及小鼠(C57BL/6)8 周龄心脏、肾脏、肝脏和肺的 CTCF ChIP-seq 数据显示小鼠 *Ugt1* 基因簇内至少有 11 个 CTCF 可结合位点：*mCBS1~11*，每个 CBS 方向由下方蓝色三角形指示；C：小鼠 *Ugt1* 基因簇内正向 CBS 序列及方向性预测。Module#1~4 的 CBS 定义为正向 CBS；D：小鼠 *Ugt1* 基因簇内反向 CBS 序列及方向性预测。Module#4~1 的 CBS 定义为反向 CBS。

架构蛋白 CTCF 及其结合元件 CBS, 探索其对 *UGT1* 基因簇表达的影响。

本研究分析比较了人和模式动物小鼠的不同细胞或组织中 CTCF 在 *UGT1* 基因簇的富集情况，发现人和小鼠的 *UGT1* 基因簇中 CBS 分布差异较大。通过对人和小鼠 *UGT1* 基因簇内的 CBS 方向以及保守性的初步分析, 发现这些 CBS 基序在 Module#2-3 处保守度较大。然后以人肺癌细胞系 A549 为模型，通过 CTCF 和 cohesin 蛋白靶向 shRNA 病毒感染细胞以敲低细胞中的 CTCF 和 cohesin 蛋白, 发现 CTCF 敲低后, A549 细胞中高度表达的 *UGT1A7* 和 *UGT1A9* 的转录水平显著降低，而 *UGT1A6* 升高。cohesin 蛋白敲低后，*UGT1A6* 表达显著升高，而 *UGT17* 和 *UGT1A9* 没有变化。该现象表明 CTCF 和 cohesin 蛋白参与调控 *UGT1* 基因簇的转录表达，可能通过影响 *UGT1* 局部基因簇的染色质构象而改变各启动子和增强子之间的相互作用。最后，本研究发现人和小鼠的所有 *UGT1* 基因启动子附近没有 CBS 位点，这与基因结构类似的原钙粘蛋白基因簇[8,46,47]存在明显不同：原钙粘蛋白基因簇中几乎每个启动子附近都有 CBS 位点，且 CBS 位点的位置和方向在人和小鼠原钙粘蛋白基因簇中高度保守[48]。

进一步通过 CRISPR 介导的 DNA 片段编辑技术研究了人 *UGT1* 基因簇内高度保守的 CTCF 结合位点 *hCBS1* 对 *UGT1* 转录的影响。同样以人肺癌细胞系 A549 为研究模型，利用 CRISPR/Cas9 系统将

hCBS1 片段删除。从 RNA-seq 数据可以看出, *hCBS1* 删除明显降低了 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 的表达。当把 *hCBS1* 方向反转为正向时，*UGT1A7* 转录水平明显降低。这可能是因为删除或反转 *hCBS1* 会改变染色质空间构象，影响增强子和启动子之间远距离相互作用，例如，*hCBS1* 删除使得调控 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 启动子的增强子空间上远离这些启动子，使它们的转录活性降低。A549 细胞中主要分布有 2 个 CBS：*hCBS1* 和 *hCBS3*，两者方向均向左(图 1A)，*hCBS1* 反转后仅 *UGT1A7* 表达下调, 可能是因为 *hCBS1* 反转后与 *hCBS3* 方向相背，造成 *hCBS1* 与下游 CBS 成环、*hCBS3* 与上游 CBS 成环，使得 *hCBS1* 和 *hCBS3* 之间的 3 个基因 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 不再处于同一个调控拓扑域内，它们具体受哪些增强子调控有待进一步研究。尽管如此，该现象表明 *hCBS1* 参与调节 *UGT1A6*、*UGT1A7* 和 *UGT1A9* 的转录，是 *UGT1* 基因簇的潜在转录调控元件。

综上所述, 本研究证明了 CTCF 和 cohesin 蛋白参与调控人 *UGT1* 基因簇的转录，删除或反转该基因簇内高度保守的 CTCF 结合位点 *hCBS1* 显著改变该基因簇活性基因的转录，这可能源于 CTCF 带来的染色质构象改变。本研究工作将为后续对 *UGT1* 的三维基因转录调控机制研究奠定基础，为临床合理用药和人类疾病的预防提供科学依据。

参考文献(References):

[1] Zhang T, Haws P, Wu Q. Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. *Genome Res*, 2004, 14(1): 79–89. [DOI]

[2] Huang H, Wu Q. Cloning and comparative analyses of the zebrafish Ugt repertoire reveal its evolutionary diversity. *PLoS One*, 2010, 5(2): e9144. [DOI]

[3] Li C, Wu Q. Adaptive evolution of multiple-variable exons and structural diversity of drug-metabolizing enzymes. *BMC Evol Biol*, 2007, 7: 69. [DOI]

[4] Nagar S, Blanchard RL. Pharmacogenetics of uridine diphosphoglucuronosyltransferase (UGT) 1A family members and its role in patient response to irinotecan. *Drug Metab Rev*, 2006, 38(3): 393–409. [DOI]

[5] Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genom*, 2005, 15(10): 677–685. [DOI]

[6] Yang N, Sun R, Liao X, Aa J, Wang G. UDP-glucuronosyltransferases (UGTs) and their related metabolic cross-talk with internal homeostasis: a systematic review of UGT isoforms for precision medicine. *Pharmacol Res*, 2017, 121: 169–183. [DOI]

[7] Rowland A, Miners JO, Mackenzie PI. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. *Int J Biochem Cell B*, 2013, 45(6): 1121–1132. [DOI]

[8] 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]

[9] Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol*, 2000, 40: 581–616. [DOI]

[10] Heath H, de Almeida CR, Sleutels F, Dingjan G, van de Nobelen S, Jonkers I, Ling KW, Gribnau J, Renkawitz R, Grosveld F, Hendriks RW, Galjart N. CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J*, 2008, 27(21): 2839–2850. [DOI]

[11] de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF binding polarity determines chromatin looping. *Mol Cell*, 2015, 60(4): 676–684. [DOI]

[12] Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*, 2014, 15(4): 234–246. [DOI]

[13] Huang H, Wu Q. CRISPR double cutting through the labyrinthine architecture of 3D genomes. *J Genet Genomics*, 2016, 43(5): 273–288. [DOI]

[14] Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 2007, 128(6): 1231–1245. [DOI]

[15] Nakahashi H, Kieffer Kwon KR, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, Qian J, Dubois W, Welsh S, Phair RD, Pugh BF, Lobanenkov V, Hager GL, Casellas R. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep*, 2013, 3(5): 1678–1689. [DOI]

[16] Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD,

Xu H, Ramsay RG, Odom DT, Flicek P. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res*, 2012, 22(11): 2163–2175. [DOI]

[17] Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell*, 2004, 13(2): 291–298. [DOI]

[18] Watrin E, Kaiser FJ, Wendt KS. Gene regulation and chromatin organization: relevance of cohesin mutations to human disease. *Curr Opin Genet Dev*, 2016, 37: 59–66. [DOI]

[19] Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 2010, 467(7314): 430–435. [DOI]

[20] Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, 2008, 451(7180): 796–801. [DOI]

[21] Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*, 2006, 20(17): 2349–2354. [DOI]

[22] Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan CS, Kunarso G, Shahab A, Bourque G, Cacheux-Rataboul V, Sung WK, Ruan Y, Wei CL. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet*, 2011, 43(7): 630–638. [DOI]

[23] Guo Y, Monahan K, Wu H, 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]

[24] Rao SS, 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]

[25] Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, 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]

[26] 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]

[27] 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]

[28] Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 2014, 346(6213): 1258096. [DOI]

[29] Fellmann C, Gowen BG, Lin PC, Doudna JA, Corn JE. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat Rev Drug Discov*, 2017, 16(2): 89–100. [DOI]

[30] Li J, Shou J, Guo Y, Tang Y, Wu Y, Jia Z, Zhai Y, Chen Z, 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]

[31] Ziebarth JD, Bhattacharya A, Cui Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res*, 2012, 41(Database issue): D188–194. [DOI]

[32] 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]

[33] Rhee HS, Pugh BF. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, 2011, 147(6): 1408–1419. [DOI]

[34] 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]

[35] Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res*, 2010, 20(5): 578–588. [DOI]

[36] Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci*

USA, 2008, 105(24): 8309–8314. [DOI]

[37] Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 2013, 154(6): 1380–1389. [DOI]

[38] Bao A, Burritt DJ, Chen H, Zhou X, Cao D, Tran LP. The CRISPR/Cas9 system and its applications in crop genome editing. *Crit Rev Biotechnol*, 2019, 39(3): 321–336. [DOI]

[39] Jiang F, Doudna JA. CRISPR-Cas9 structures and mechanisms. *Annu Rev Biophys*, 2017, 46: 505–529. [DOI]

[40] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337(6096): 816–821. [DOI]

[41] Owens IS, Basu NK, Banerjee R. UDP-Glucuronosyltransferases: gene structures of UGT1 and UGT2 families. *Methods Enzymol*, 2005, 400: 1–22. [DOI]

[42] Wang Y, Huang H, Wu Q. Characterization of the zebrafish Ugt repertoire reveals a new class of drug-metabolizing UDP glucuronosyltransferases. *Mol Pharmacol*, 2014, 86(1): 62–75. [DOI]

[43] Bartlett MG, Gourley GR. Assessment of UGT polymerphisms and neonatal jaundice. *Semin Perinatol*, 2011, 35(3): 127–133. [DOI]

[44] Che F, Luo ZY. Research progress of UGT1A1 gene in the pathogenesis of Gilbert and Crigler-Najjar syndromes. *J Chin Prac Diagn Ther*, 2015, 29(3): 219–222.
车芳, 骆子义. UGT1A1 基因在 Gilbert 综合征及 Crigler-Najjar 综合征发病机制中研究进展. 中华实用诊断与治疗杂志, 2015, 29(3): 219–222. [DOI]

[45] 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(04): 313–325. [DOI]

[46] 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]

[47] Wu Q. Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. *Genetics*, 2005, 169(4): 2179–2188. [DOI]

[48] Zhai Y, 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(04): 323–336. [DOI]

(责任编委: 方向东)