CTCF-Mediated Functional Chromatin Interactome in Pluripotent Cells

Lusy Handoko[1,*], **Han Xu**[1,*], **Guoliang Li**[1,*], **Chew Yee Ngan**[1], **Elaine Chew**[1], **Marie Schnapp**[1], **Charlie Wah Heng Lee**[1], **Chaopeng Ye**[1], **Joanne Lim Hui Ping**[1], **Fabianus Mulawadi**[1], **Eleanor Wong**[1,2], **Jianpeng Sheng**[3], **Yubo Zhang**[1], **Thompson Poh**[1], **Chee Seng Chan**[1], **Galih Kunarso**[4], **Atif Shahab**[1], **Guillaume Bourque**[1], **Valere Cacheux-Rataboul**[1], **Wing-Kin Sung**[1,2], **Yijun Ruan**[1,#], and **Chia-Lin Wei**[1,2,#,¥]

[1] Genome Institute of Singapore, Singapore 138672

[2] National University of Singapore, Singapore 117543

[3] Nanyang Technological University, Singapore 637551

[4] Duke-NUS Graduate Medical School Singapore, Singapore 169857

Abstract

Mammalian genomes are viewed as functional organizations that orchestrate spatial and temporal gene regulation. CTCF, the most characterized insulator-binding protein, has been implicated as a key genome organizer. Yet, little is known about CTCF-associated higher order chromatin structures at a global scale. Here, we applied Chromatin Interaction Analysis by Paired-End-Tag sequencing to elucidate the CTCF-chromatin interactome in pluripotent cells. From this analysis, 1,480 *cis* and 336 *trans* interacting loci were identified with high reproducibility and precision. Associating these chromatin interaction loci with their underlying epigenetic states, promoter activities, enhancer binding and nuclear lamina occupancy, we uncovered five distinct chromatin domains that suggest potential new models of CTCF function in chromatin organization and transcriptional control. Specifically, CTCF interactions demarcate chromatin-nuclear membrane attachments and influence proper gene expression through extensive crosstalk between promoters and regulatory elements. This highly complex nuclear organization offers insights towards the unifying principles governing genome plasticity and function.

[#] Corresponding authors: Chia-Lin Wei, Tel: 1-(925) 927-2593, cwei@lbl.gov; Yijun Ruan, (65) 6808-8073, ruanyj@gis.a-star.edu.sg.
[*]These authors contributed equally
[¥]Current address: Joint Genome Institute, Walnut Creek, California, U.S.A.

Keywords

insulator; enhancer; chromatin organization; epigenetic regulation; nuclear lamina

Introduction

Eukaryotic genomes are organized into functional architectures [1]. These higher order genome structures and their associated sub-nuclear compartments are increasingly recognized as the key components contributing to many aspects of nuclear activities, including DNA transcription[2,3]. Among the different types of structures, long-range chromatin interaction is the most distinguishable feature [4]. Its spatial distribution and temporal regulation are critical for cell identity and cell fate decision [5]. Therefore, the elucidation of the three-dimensional chromatin interactome is essential for understanding genome organization in lineage specification and transcription regulation[6-9].

Among all the epigenetic modulations, insulator-mediated chromatin remodeling, enhancer occupancy and sub-nuclear chromatin localization are known to affect genome configuration [10]. CTCF, one of the most extensively studied insulator binding proteins in vertebrate[11,12], is known to demarcate boundaries between euchromatin and heterochromatin [13]. Several models, including enhancer blocking [14] and domain barrier [13], have been proposed for CTCF-mediated insulation. In the enhancer blocking model, CTCF can block the communication between adjacent regulatory elements such as enhancers and gene promoters. In the domain barrier model, CTCF binding is proposed to function as a buffer from the effects caused by the adjacent chromatin domains. To address its function as an insulator, CTCF binding profiles have been determined from the genomes of several different cell types and across species [15-17]. Based on their largely invariant patterns and the associated genes, it is still not clear how CTCF functions to influence the three dimensional structures and global genome activities.

While the genome-wide role of CTCF is not yet understood, its function in mediating chromatin interactions has been extensively studied in several selected gene regions, particularly the imprinting loci[18,19]. In the best characterized *H19/Igf2* locus, CTCF binding on *H19* imprinting control regions (ICR) can result in the expression of paternal *Igf2* and maternal *H19* through physical chromatin interactions with the differentially methylation regions (DMR) upstream of the *Igf2* promoter [20,21]. CTCF is also implicated in sub-nuclear localization [22] and mediating *Xist* interactions during X chromosome inactivation (XCI) [23]. Beyond these well characterized interactions, additional CTCF-dependent inter- and intra-chromosomal interactions were identified from these imprinting regions using Circular Chromosome Conformation Capture (4C) [24,25]. With these available information, CTCF has been proposed as one of the leading candidates as a global genome organizer to coordinate high order chromatin structures and regulate gene expression [12]. Despite this apparent role, CTCF-associated chromatin interactions have only been studied at limited numbers of loci. Little is known about the CTCF-associated chromatin organization and its impacts on the epigenetic states and transcriptional activities at a global scheme.

Here, we present the first global and high resolution CTCF-associated chromatin interactome map in murine embryonic stem (ES) cells. Representing a primary state of genome architecture and epigenetic conformation, the pluripotent nature of ES cell lies in its unique genome plasticity and distinctive transcription program [26]. Our results suggest that CTCF configures the genome into distinct chromatin domains and sub-nuclear compartments that exhibit unique epigenetic states and diverse transcriptional activities. Interrogated by the whole genome p300 enhancer binding, RNAPII activity and Nuclear Lamin (NL) occupancy, this interactome map offers unprecedented information content and resolution on the 3D genome structure of ES cells. Contrary to the enhancer blocking model, CTCF associated interactions potentially promote communications between functional regulatory elements to regulate gene expression. Our data also indicates that CTCF loops can feature as domain barriers by demarcating NL-chromatin interactions and delineating the chromosomes sub-nuclear localizations. This chromatin organization map in a genome-wide context can extend our knowledge of insulator-directed transcriptional regulation and serve as a framework to reveal mechanisms critical for genome plasticity in pluripotency and development.

Results

CTCF-associated chromatin interactome characterized by ChIA-PET analysis

To gain insights into CTCF-associated global chromatin organization, we performed ChIA-PET analysis [6]. Here, juxtaposed interacting chromatins brought together in close spatial proximity by CTCF *in vivo* are enriched by chromatin immunoprecipitation (ChIP) and connected through DNA linkers followed by paired end tag (PET) extraction and sequencing (Supplementary Fig.1a). A ChIA-PET analysis generates two types of genome-wide datasets, the binding sites defined by ChIP enrichment and the interactions between two binding loci revealed by proximity ligation events (Supplementary Note). Tags from each side of PETs mapped onto the same chromosome within 10 Kb are considered as self-ligating PETs and used to define CTCF binding sites. PETs with two ends mapped to different chromosomes (inter-chromosomal interacting PETs) or to the same chromosomes spanned longer than 10Kb (intra-chromosomal interacting PETs) are candidates to define long-range chromatin interaction loci [27]. These PETs are further clustered based on the overlapping anchor regions and a random distribution model is applied to determine the significance of the interactions [27]. CTCF binding sites and interaction data with false discovery rate (FDR)   0.05 are subjected for further analysis.

Two CTCF chromatin preparations were used to construct independent biological (BR1 *vs.* BR2) and technical (TR1 *vs.* TR2) replicates (Supplementary Table 1). High reproducibility between TR1 *vs.* TR2 was observed (Supplementary Note); indicating that these libraries are of good quality and the current sequencing depth has reached near saturation to explore existing library complexity. The overlap between BR1 *vs.* BR2 is only 38%; this suggests that the diversity is large and the interactions captured here represent only a subset of the complex CTCF-associated chromatin interactome (Supplementary Note). To achieve higher coverage, all PET sequences from different replicates were pooled. From a total of 44.7 million unique PET sequences, 18.8 million PETs (42%) were mapped to unique locations

on the mm8 genome, which defined 39,371 CTCF binding sites (Supplementary Table 2) and 5,384 inter-ligation PET clusters (Supplementary Table 3a,b).

To verify the fidelity of these 39,371 CTCF binding sites identified by the ChIA-PET approach, we compared them with CTCF binding sites determined by an independent ChIP-seq analysis. From near saturated sampling of $\sim$ 30 million individual ChIP fragments by ChIP-seq, we identified 68,292 CTCF sites (FDR < 0.05). More than 98% (38,700/39,371) of the CTCF binding peaks defined by ChIA-PET overlap with those identified by ChIP-seq. The peak intensities between these common sites show high degree of correlation (Pearson correlation of 0.79) and ChIP-qPCR on selective sites also confirmed they are *bona fide* binding sites (Supplementary Fig. 1b, c). Therefore, ChIA-PET mapping and processing appear to be accurate and robust.

We next focused our attention on the interaction PET clusters derived from the 10.1 million uniquely mapped inter-ligating PETs. Based on occurrence frequency (multiple overlapping clusters; FDR 0.05), presence of CTCF binding sites and absence of homologous sequences between anchor regions; we defined 1,816 high confidence interactions, including 1,480 intra-chromosomal and 336 inter-chromosomal interactions. In total, 3,306 CTCF binding sites are involved in these chromatin interactions. When compared with the binding sites that do not anchor the interactions, these 3,306 sites do not exhibit any obvious different genomic features like the consensus CTCF binding motif or histone modification signals. However, they do show higher binding intensities (p<10E-308 in KS-test; Supplementary Fig. 1d, e).

These 1,816 CTCF-mediated interactions were then used for subsequent analysis on their associated genes, chromatin environments and transcriptional activities (Fig. 1a). A detailed view of interactions detected on chr. 13-15 and *cis*-interactions on chr. 10 are shown as Figure 1b and 1c, respectively. From this chromatin interactome map, we observed many complex CTCF interaction patterns, for examples, loops sequestering multiple genes (such as the cytochrome P450 family 2 *Cyp2* and clusters of *Keratin* genes) or connecting multiple promoters (such as *Pcdh* α, β and γ loci). We also specifically examined the well characterized *Igf2-H19* region. Our data confirmed the previously known *Igf2* DMR-*H19* ICR interaction locus (chr7:142,348,445-142,493,110) [20] and identified putative new inter- and intra- chromosomal interactions associated with CTCF bindings in this locus (Table 1).

To validate the intra- and inter-chromosomal interactions, we applied 4C-based molecular method[25,28] and Fluorescent In Situ Hybridization (FISH) cytogenetic assay, respectively. *Cyp2b10* region (chr7: 25,625,755-25,626,491) and *Pcdhga12* promoter region (chr18: 37,890,974-37,892,946) were selected as anchors for 4C validations (Supplementary Table 4a). In the *Cyp2* locus, we identified two intra-chromosomal interactions by ChIA-PET and both were confirmed by 4C (Fig. 2a). Apart from these two confirmed sites; 4C also discovered additional two intra-chromosomal interactions. In the *Pcdhg* locus, we identified five intra-chromosomal interactions and confirmed two of them by 4C (Supplementary Fig. 2a). Using FISH, we systematically validated 14 and confirmed 9 inter-chromosomal interactions (Supplementary Fig. 2b-d). Seven of the nine confirmed interactions have CTCF binding sites at both anchors while two contain CTCF binding sites at only one of the

anchors. An example of specific co-localization between two alleles on chr.13 (chr13:13,658,687-13,660,696) and chr.15 (chr15:74,912,106-74,914,727) is shown in Figure 2b. Their co-localization frequency was 2 folder higher than that from the random negative control loci (14.61% *vs.* 7.49%, p-value < 1.28E-05). To determine whether CTCF is directly involved in tethering these chromatin loops, CTCF expression in the ES cells was knocked down by siRNA and the co-localization frequencies of interacting loci were measured by FISH and 3C experiments (Fig. 2c-d, Supplementary Fig. 2e, f) [6]. In the CTCF knock down cells, the interaction frequencies of all four inter- and one intra-chromosomal interactions examined were reduced to the background control level (Supplementary Table 4b-c). To further verify that such redcution of interaction frequency was not influenced by the viability of CTCF knock down cells, we performed the fluorescence activated cell sorting (FACS) assays and confirmed that these cells were viable and no significant level of apoptosis or cell cycle arrest was detected when compared to the control cells and untreated cells (data not shown). Taken together, these results suggest that the majority of the interactions detected by ChIA-PET are genuine and reliable and CTCF binding is required in the formation of these loops.

Collectively, we identified 1,480 intra- and 336 inter-chromosomal interactions. The majority of the interactions and particularly, the strong interactions, are the intra-chromosomal interactions; significant numbers of inter-chromosomal interactions were also validated. Such extensive crosstalk between chromosomes prompted us to explore chromosomal spatial organization. Using the normalized inter-chromosomal interaction frequencies as the measurements of relative vincinity, spatial chromosome proximities are detected between chr15-16-18, chr6-9, chr2-7 and chr1-19 (Supplementary Fig. 3a). Hierarchical clustering shows chromosomes are distributed among 3 major sub-clusters (Supplementary Fig. 3b). The specificity of these high affinity clusters was confirmed through their high reproducibility among two biological replicates (Supplementary Fig. 3c, r = 0.736, p-value <6E-34).

CTCF interactions correlate with distinct chromatin domains featured with specific histone modification patterns

To further delineate the nature of CTCF-defined chromatin loops, we analyzed histone modification signatures within the intra-chromosomal interaction loops. Seven histone H3 modifications (H3K4me1, me2, me3, K36me3, K27me3, K9me3 and K20me3; data from http://genome.ucsc.edu/broadChromatinChIPSeq) that represent different types of chromatin activities from murine ES cells were examined within the loops and compared with the signal intensities from immediately outside of the loops (to the left and to the right, respectively) (Fig. 3a). When the cumulated normalized intensities from all CTCF loops were compared with the intensity profiles determined from control simulated loops constructed from randomly paired CTCF binding sites, CTCF interactions exhibit unique enrichment patterns of H3K4me1 and H3K27me3 within the loops and at the boundaries, respectively as well as depletion pattern of H3K36me3 (Fig. 3b-d). These observations suggest CTCF interactions harbor specific chromatin domains. To further elucidate the chromatin states associated with each individual CTCF interaction, we applied a non-supervised clustering algorithm and clustered loops exhibiting similar combinatory histone

modification patterns relative to their neighboring regions (Fig. 4a). Using this strategy, we uncovered 5 distinct patterns; Category I, accounting for 12% of the CTCF-defined chromatin loops, featured chromatin signatures of active characteristics, *i.e.* active marks H3K4me1&me2 and H3K36me3 are enriched inside the loops while the repressive K9, K20, K27 methylation marks are depleted; Category II (11%) demarcates repressive domains with extensive K9, K20 and K27 methylations but active chromatin marks are under-represented. Category III loops (19%) are hubs for enhancer and promoter activities with an enrichment of H3K4me1 and me2 signals within the loop regions and H3K4me3 peaking at the boundaries of the loops, while the active transcription mark H3K36me3 and repressive K27me3 mark are only observed outside of the loops on opposite sides; Category IV, the most abundant type (31%), displays opposite chromatin states flanking the boundaries of chromatin loops, but does not exhibit any specific pattern within the loops. Specifically, active K4 methylation marks are observed on one side and repressive K9, K20 and K27 methylation marks are observed on the other side. These loops are likely to act as domain barriers to physically separate active and repressive chromatin domains. The remaining 27% of the loops (Category V) do not display any specific chromatin patterns based on the seven histone signatures surveyed here. Supplementary Table 5 lists the category assigned for each interaction loop.

To correlate such unique histone patterns with genomic feature like loop span, CTCF loops were sorted in ascending order based on their spans, and the average cumulated intensities of each histone mark were determined and smoothed using a sliding window of 100 interactions, respectively. As shown in Fig. 4b, a clear transition of histone modification patterns is observed at loop size ∼200 Kb. Such correlation between enrichment pattern and span size is uniquely found in CTCF-associated loops and not found in the random simulated control loops (p-value < 5E-05) or loops associated with other transcription factors such like SALL4 and RNAP II (Supplementary Fig. 4a). 994 CTCF loops with span < 200 Kb are mostly over-represented with the active chromatin features (H3K4 methylations) while the loops with span > 200 Kb clearly demonstrate repressive domains characteristics indicated by extensive K9, K20 and K27 methylations. In particular, the distributions of H3K4, K36, K9 and K20 methylations across variable span sizes are clearly distinct in the CTCF-associated loops (p-value = 1E-03 to 1E-05; Supplementary Fig. 4b). Among them, H3K4me1 signals within the active CTCF loops are moderately reduced in CTCF knock-down cells (p-value < 0.005; Supplementary Fig. 5). Taken together, these data revealed the specificity of CTCF loop in defining the chromatin domains.

CTCF-associated chromatin domains are closely correlated with transcription status

The impacts of distinct CTCF-associated chromatin domains on transcriptional control were measured through the promoter activities from genes within each category of loops by RNAP II binding signals. We performed a genome-wide RNAP II binding profiling by ChIP-seq and identified 35,853 RNAP II occupied regions in ES cells (Supplementary Table 6a). As expected, RNAP II signals are significantly enriched within the loop regions in Category I (active chromatin hubs) and depleted in the repressive chromatin hub loops. RNAP II signals are packed around the boundaries of the predicted enhancer loops (Category III). In Category IV, RNAP II signals are mainly found from the active chromatin

domain side of the loop regions (Supplementary Fig. 6). Fig. 4c shows the proposed models for each category. Examples of individual loops, their corresponding histone modification profiles and associated genes are shown in Supplementary Fig. 7.

In conclusion, our results demonstrate that the CTCF-associated chromatin interactome defines distinct chomatin domains with unique epigenetic states. Moreover, these CTCF-associated long range interactions could be involved in regulating RNAP II activities within the corresponding domains, and thus affect gene expression.

Enhancer p300 targeting and nuclear Lamin B anchoring are imperative features of distinct chromatin domains

p300, a key tissue specific enhancer binding protein, can direct active transcription across long distance from its corresponding promoters [29]. Nuclear lamina (NL), a fibrous protein network underlying the inner nuclear membrane, is known to interact with CTCF to tether chromatin regions into sub-nuclear periphery for gene silencing[22]. Given their apparent roles in modulating transcription and chromatin structure [29-31] as well as known connections with CTCF, we characterized their genomic distributions in relation to the CTCF-defined chromatin interactome. Genome-wide p300 and Lamin B occupancies in ES cells were defined by ChIP-seq. A total of 5,033 p300 binding sites and 1,344 Lamin B-associated domains (LADs) were discovered (Supplementary Table 6b, c) and confirmed by ChIP-qPCR (Supplementary Table 7 and Fig. 8a, b).

p300 peaks are found to associate with the genome in the gene-rich regions but under-represented in proximal promoters (Supplementary Fig. 8c). Further analysis shows high co-occurrence of p300 with open chromatin, enhancer signals H3K4me1 & me2 and loci targeted by cell specific multiple-transcription binding-loci (MTL) [32] recruiting core pluripotency factors like Oct4-Sox2-Nanog (Supplementary Fig. 8d, e). When p300 binding is surveyed in the context of different CTCF chromatin domains, p300 binding is found to specifically target in the loops harnessing active and enhancer marks. The binding densities of p300 are 4.5 fold and 2.5 fold higher than average in the active (Category I) and enhancer (Category III) domains, respectively (p=1.95E-36; Supplementary Fig. 8f). Furthermore, many of the p300-containing CTCF mediated chromatin loops may be involved in connecting p300 with promoters over long distances. For example, promoter of *Pofut2*, a highly expressed gene in ES cells, is connected with p300 located 124 Kb away (chr10:76,573,384-76,575,577), potentially by a CTCF enhancer loop (Fig. 5a). *Zbtb2*, an active gene in ES cells (chr10:5,958,433-5,979,467), its promoter could potentially contacts p300 bound 23 Kb downstream of the TSS through a CTCF associated interaction.

This finding prompted us to study global p300-promoter associations through CTCF interactions. Using genome-wide p300 binding sites and their distances to the nearest promoters defined by CTCF chromatin interactions, we found that 28% of the genes whose promoters are distal but brought into close contact (<10Kb) with p300 through CTCF mediated looping are up-regulated in ES cells compared with neuronal stem cells. This level of up-regulation is similar to the 32% up-regulation from genes whose promoters located in proximity (<10Kb) from p300 binding sites; and is significantly higher when compared with 14% of up-regulated genes whose promoters are located distal (>10 Kb) from p300 binding

(Fig. 5b). To further determine whether CTCF associated loops directly regulate the expressions of these genes, CTCF expression was down-regulated by siRNA and gene expressions and RNAP II binding were examined by quantitative RT-PCR and ChIP-seq. Among the five genes examined, four show decreased expressions (Fig. 5c) and the RNAP II binding profiles around the detected interaction loci also exhibit reduced intensities (Supplementary Fig. 9). Thus, CTCF-associated DNA looping can facilitate communications between regulatory elements, like enhancers and promoters and influence transcription, as previously suggested [33]. Such interaction events can reinforce the effects of distal regulatory elements specifically on their corresponding promoters through organizing proper chromatin configuration.

The 1,344 LADs are wide and discrete. Their spans vary from 10 Kb to 2 Mb and in total, occupy 283 Mb (∼11%) of the mouse ES cell genome. Their patterns are highly similar with the LADs defined by the DamID method [34] (Fig. 6a, Supplementary Fig. 10a) and 250 out of 283 Mb (88%) of the LADs identified here overlap with the regions determined by DamID; suggesting that the ChIP-based sequencing method should be feasible for genome-wide Lamin study. When CTCF looping signals were investigated in relation to LADs, they were found to be depleted within the LADs but located close to the borders (Fig. 6a; Supplementary Fig. 10b). To verify their relative position in demarcating the LADs, CTCF loop occupancy was profiled along LADs and up to 1 Mb outside of LAD borders (Fig. 6b). The loop density is found to peak around the borders or in the sub-telomeric regions (Supplementary Fig. 10c).

The presence of LADs and p300 binding in relation to different loops are mostly non-overlapping (p-value 1E-20). Most of the LADs appear to associate with low transcriptional activity and gene content, but enriched with repeats (Supplementary Fig. 10d-e). The expression levels of these LAD-associated genes are on average ∼ 10-20 folds lower than the expression levels of genes outside of the LADs (Fig. 6c). CTCF loops tend to be located proximal to the boundaries of LADs. Therefore, LAD-containing CTCF loops may act as insulators by establishing boundary structures. Such structures can restrict the spreading of the silenced nature of the LADs into the neighboring regions or *vice versa*. Because of the effect of LADs on gene repression, the genes within ES-LADs could be, at least partially, variable and preferentially up-regulated in other cell types. As shown in Fig. 6d, 64% of LAD-associated genes are up-regulated   2 fold in NS cells, suggesting that some portions of the chromatins may alter their associations with NL in different cell types; consistent with previous finding [34].

Discussion

In this study, we present for the first time a genome-wide CTCF-mediated chromatin interactome map and its configured chromatin domains. Our data implies that CTCF can function as a genome organizer through several modes: creating local chromatin hubs and harnessing clusters of genes with coordinated expression; facilitating communication between regulatory elements and their corresponding promoters, and demarcating boundaries between chromatin and sub-nuclear compartments like NL as domain anchors/barriers. As a result, the spatial distribution of these interactions and their interplay with key

epigenetic components delineate broad genome properties and specify unique transcriptional programs (Supplementary Fig. 11).

Systematic interrogations of CTCF-associated loops with histone modifications, RNAPII occupancy, enhancer p300 binding and NL association define five categories of chromatin domains with unique epigenetic states. These diverse types of chromatin interactions offer attractive models for CTCF to coordinate gene regulations. Specifically, we uncover many of the CTCF interactions directing multiple promoter-promoter and promoter-enhancer communications. Enhancers are known to activate transcription in a distance and direction independent manner [35]. Therefore, mechanisms have to be in place to ensure their specificity and avoid promiscuity. Our finding suggests a potential new function of CTCF in promotor-enhancesome interaction which is different from the previously proposed enhancer blocking model. Such an "enhancer-bridging" activity could potentialy engage CTCF in generating cell specific chromatin interactions and guide gene expression programs.

In some selective loci, CTCF-directed chromatin contacts are found to occur prior to the gene activation and not to be disrupted by inhibition of RNAP II activity [36]. Therefore, it is likely that the impacts on transcription could be the outcome of the CTCF mediated higher ordered structure. As such, the regulation of CTCF function can be established through the formation of CTCF loop structures. Comparing with regulation at the level of primary binding, this mode of regulation offers better versatility. Extending from this concept, factors such as the genome context and co-factor recruitment within the CTCF anchoring regions could potentially manipulate the formation or stability of these long-range interactions. Recent efforts in searching the CTCF interacting partners have identified proteins functioning in nuclear architecture and transcription, including RNAPII [37], Lamins, nucleolar protein NPM [22], Yy1[38] and Cohesin [39]. Furthermore, special nucleosome distributions are found to array near the chromatin linker of CTCF binding [40]. Specifically, Cohesin was found to be required for CTCF mediated intra-chromosomal contacts around the interferon-γ locus [41] and genome-wide studies in different cell lines have shown that 80% of Cohesin binding sites co-localizes with CTCF binding sites [42,43]. Furthermore, mediator and cohesin were shown recently to exhibit similar enhancer-promoter bridging activity [44] as shown here for CTCF. It is conceivable that Cohesin, mediator together with CTCF, may play key roles in the formation of these loops and function as an integral part of the transcription apparatus to influence cell specific chromatin structures and transcriptions. The potential associations between CTCF and its partner proteins, their consequences on dynamic chromatin structures and gene expression will be exciting areas to explore.

As expected to be a key genome organizer, CTCF binding was found to be involved in both intra- and inter-chromosomal interactions [24,45]. It is anticipated that the interactions identified in this study only represent a snapshot of a very large and complex CTCF-associated interactome and we are far from ascertaining the comprehensive interaction events. With the continous improvement on the robustness of the ChIA-PET approach and advances in the throughput, accuracy and coverage of the sequencing technologies, it is likely that many more of the dynamic and transient interactions can be captured to reveal the truly comprehensive and high resolution nature of the CTCF-associated 3D genome structures.

Collectively, our results reveal the CTCF-mediated genome organization, its imperative feature and correlation with gene expression. The unique chromatin domains found here offer testable hypotheses to examine the molecular mechanisms by which CTCF directs its various insulating activities. Furthermore, regulation of the looping events by controlling the CTCF interactions with chromatin and other auxillary partners will be important for the establishment of gene expression cascades throughout development. With this knowledge, one can start to dissect the unifying principles governing the dynamic natures of these high order chromatin architectures and means to manipulate them in the future.

Methods

Cell culture, siRNA and expression analysis

Mouse ES cells E14 were maintained as previously described [32]. CTCF or control siRNA (Dharmacon) was transfected using DharmaFect2 as in Lim *et al.*, [48] and according to DharmaFect protocol. The cells were harvested 48h after the 3rd transfection to generate CTCF knockdown and control cells. For gene expression analysis, RNA was isolated using TRIZOL Reagent (Invitrogen) and RNeasy mini kit (Qiagen). cDNA was then generated using SuperScript™ III reverse transcriptase (Invitrogen) and used for qPCR (Roche).

Chromatin based analyses: ChIA-PET, 3C, 4C and ChIP-Seq

CTCF ChIA-PET—CTCF ChIA-PET libraries were constructed accordingly to Fullwood *et al.*, [6] (Supplementary Fig. 1a). PET sequences were processed with ChIA-PET Tool [27], subject to an additional rescue procedure in the tag mapping step. For multiple mapped tags, if only one of the mapping locations is within 1Kb from a CTCF binding site, this location is then uniquely assigned to this tag and the PET will be used for further analysis.

ChIP-Seq—ChIP-Seq were performed as described previously[31,32,46] using known antibodies against p300 (Santa Cruz sc-585), CTCF (Upstate, 07-729), Lamin B (Santa Cruz, sc-6216) and RNAP II (8WG16 clone, Covance, MMS-126R), respectively. Sequence data was analyzed using the peak finding algorithm CCAT [49]. The p300 sites and CTCF sites were identified with FDR cut-off 0.2 and 0.05, respectively. The Lamin-B regions were identified with FDR cut-off 0.2.

Modified circular chromosome conformation capture (4C)—Modified circular chromosome conformation capture (4C) was performed as described previously[25]. 400-600 bp PCR products were purified from the gel and sequenced by GSFLX. For each 4C library, we generated 20-30K raw sequences. Sequences were aligned with both primer sequences and the aligned parts were trimmed from the reads. The post-trimmed unique reads longer than 25bp were mapped to the mouse genome mm8 with BLAT and the ones with the best unique mapping loci were used to define interaction. The interactions were clustered and cluster size 2 were selected.

3C-qPCR—3C-qPCR (chromosome conformation capture-qPCR) was carried out as previously described [50, 6]. A region within an independent *Ercc3* locus on chr18 was

selected as internal control. 3C preparation was repeated 2-4 replicates. Around 30 ng DNA were used for qPCR (2-4 times, each with duplicates).

FISH (fluorescent in situ hybridization) assay

FISH in ES cells was done as previously described[6]. Co-localization is defined if the distance between two signals (center to center) is less than the average of the diameters of both signals (1.12 μm). The percentage of co-localization between two potentially interacting anchors (experimental probe) was compared with that between one anchor and the negative control region on chr.16. 2-tailed p-value was calculated using Fisher's exact test. An interaction was confirmed if p-value is < 0.05 and the co-localization ratio is > 1.5 fold.

Primers, linkers and interactions used

RNAP II and SALL4 interactions used for comparison of the histone modification profiles are listed in Supplementary Tables 8-9. All the linkers, primers and siRNA sequences used for validation experiments RNAP II are listed in Supplementary Table 10.

Clustering of CTCF-associated intra-chromosomal interactions through histone signature profiles

To discover the histone modification patterns of CTCF loops, we performed k-mean clustering for the 1,295 CTCF loops in 4 steps:

a. Measurement of histone modification intensity—To measure the intensities of different histone modifications within or proximal to these CTCF loops, we adopted the publicly available 7 histone modification datasets (H3K4me1, me2, me3, H3K36me3, H3K27me3, H3K9me3 and H3K20me3) derived from ES cells[46,51]. Given a CTCF loop with the length l, we extended l base pairs to its left and right. The extended region with the total length $3l$, was partitioned into 30 bins (Fig. 3a). For each bin, the reads from each of the histone modification data were counted, denoted c_{ijk} ($i=1,\ldots,1295;j=1,\ldots,7;k=1,\ldots,30$). The read counts were log-transformed such that: $x_{ijk}=\log(c_{ijk}+1)$ where a pseudo-count of 1 was added to the read count to handle the situation of $c_{ijk}=0$. As such, each of the 7 histone modification intensities was determined in these 30 bins and a matrix was generated.

b. Normalization of histone modification intensity—The intensity computed from read counts depends on the loop length, the sequencing depth, and the intrinsic ChIP-Seq bias due to copy number variation, read mappability and chromatin structures[52]. Therefore, we further normalized the log-transformed intensity in two steps:

i. We normalized against the average value of the 30 bins, for each loop and each histone modification type: $y_{ijk}=x_{ijk} - \sum_{u=1}^{30} x_{iju}/30$

ii. We normalized against the average value among all histone modification types, for each loop and each bin: $z_{ijk}=y_{ijk} - \sum_{u=1}^{7} y_{iuk}/7$

c. Generating empirical negative controls of CTCF loops—We generated empirical negative controls from pairs of CTCF binding sites with no interactions. For a fair comparison, the control pairs were chosen to follow the same multivariate distribution as the identified loops in the 3D space by i) genomic distance between two binding sites; ii) binding intensities of both binding sites represented by read counts in ChIP-Seq library. We took 3 steps to generate the negative control set.

    i. Random pairs of CTCF ChIP-Seq binding sites with genomic distance < 1Mb were selected. Among them, pairs with at least one ChIA-PET linking the two binding sites were filtered out. This resulted in a large control set with 2.12M pairs;

    ii. The large control set was randomly divided into 10 subsets, each with 212k pairs of binding sites. The 10 subsets were denoted $S_1, S_2, \ldots, S_{10}$;

    iii. The 1,295 identified loops were projected to the 3D space. For each loop, we selected the nearest control pair in one of the subsets in the 3D space based on Mahalanobis distance. This step was repeated for all $S_1 = (i=1,2,\ldots,10)$. As the result, 10 subsets of negative controls, each with 1,295 pairs of binding sites, were selected. The 10 subsets of negative controls were processed as described (*a*). Standard deviations were computed based on the 10 subsets for each bin.

d. Clustering of CTCF loops based on histone modification patterns—For each CTCF loop, 7×30 features were extracted based on the histone modification intensity measure z_{ijk}. K-mean clustering was employed using the software Cluster 3.0 (http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm). As result, seven clusters were generated. Among them, the 3rd and the 4th clusters as well as the 5th and the 6th clusters were symmetric around the center of the loop. We further grouped the symmetric clusters into five categories.

Define RNAP II binding signal distributions along the CTCF loop regions

RNAP II ChIP-seq signals were determined for each CTCF loop and their proximal regions as in *a* above. The log-transformed intensities of RNAP II were further normalized using the methods in *b*. For each category of CTCF loops defined based on histone modification signatures, we computed the overall RNAP II profile using the method in *b*.

Define promoter-p300 enhancer interactomic distance

From the 13,049 Refseq genes, we annotated their promoters with the nearest p300 binding sites based on **a).** linear genomic distance and **b).** interactomic distance through the CTCF mediated interactions. Suppose we have a list of CTCF interactions

$\left\{ (s_1^{head}, s_1^{tail}), (s_2^{head}, s_2^{tail}), \ldots, (s_n^{head}, s_n^{tail}) \right\}$, where n is the number of CTCF interactions, s_i^{head} and s_i^{tail} are the genomic locations of the two binding sites in the *i*th interaction. Give two genomic locations x and y, the interactomic distance between x and y is defined as:

$$d'(x,y) = \min \left\{ \min_i \left\{ d(x, s_i^{head}) + d(y, s_i^{tail}) \right\}, \min_i \left\{ d(y, s_i^{head}) + d(x, s_i^{tail}) \right\} \right\}; \text{ where } d(x,y)$$

is the linear genomic distance. By this definition, two remote genomic locations that are brought together through a CTCF loop would correspond to a short interactomic distance

while their linear genomic distance is relatively longer. Using 10 Kb as a cutoff, genes were categorized as linear genomic distances < 10 Kb, linear genomic distance > 10 Kb but interactomic distance < 10 Kb and both linear and interactomic genomic distance >10Kb.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Francastel C, Schubeler D, Martin DI, Groudine M. Nuclear compartmentalization and gene activity. Nat Rev Mol Cell Biol. 2000; 1:137–43. [PubMed: 11253366]

2. Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007; 128:787–800. [PubMed: 17320514]

3. Bolzer A, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol. 2005; 3:e157. [PubMed: 15839726]

4. Misteli T, Soutoglou E. The emerging role of nuclear architecture in DNA repair and genome maintenance. Nat Rev Mol Cell Biol. 2009; 10:243–54. [PubMed: 19277046]

5. Meshorer E, Misteli T. Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol. 2006; 7:540–6. [PubMed: 16723974]

6. Fullwood MJ, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009; 462:58–64. [PubMed: 19890323]

7. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–93. [PubMed: 19815776]

8. Duan Z, et al. A three-dimensional model of the yeast genome. Nature. 2010; 465:363–7. [PubMed: 20436457]

9. Schoenfelder S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet. 2010; 42:53–61. [PubMed: 20010836]

10. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature. 2007; 447:425–32. [PubMed: 17522676]

11. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 1999; 98:387–96. [PubMed: 10458613]

12. Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009; 137:1194–211. [PubMed: 19563753]

13. Felsenfeld G, et al. Chromatin boundaries and chromatin domains. Cold Spring Harb Symp Quant Biol. 2004; 69:245–50. [PubMed: 16117655]

14. Valenzuela L, Kamakaka RT. Chromatin insulators. Annu Rev Genet. 2006; 40:107–38. [PubMed: 16953792]

15. Barski A, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007; 129:823–37. [PubMed: 17512414]

16. Cuddapah S, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 2009; 19:24–32. [PubMed: 19056695]

17. Kim TH, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell. 2007; 128:1231–45. [PubMed: 17382889]

18. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000; 405:482–5. [PubMed: 10839546]

19. Majumder P, Gomez JA, Chadwick BP, Boss JM. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. J Exp Med. 2008; 205:785–98. [PubMed: 18347100]

20. Murrell A, Heeson S, Reik W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet. 2004; 36:889–93. [PubMed: 15273689]

21. Yoon YS, et al. Analysis of the H19ICR insulator. Mol Cell Biol. 2007; 27:3499–510. [PubMed: 17339341]

22. Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell. 2004; 13:291–8. [PubMed: 14759373]

23. Augui S, et al. Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. Science. 2007; 318:1632–6. [PubMed: 18063799]

24. Ling JQ, et al. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science. 2006; 312:269–72. [PubMed: 16614224]

25. Zhao Z, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006; 38:1341–7. [PubMed: 17033624]

26. O'Shea KS. Self-renewal vs. differentiation of mouse embryonic stem cells. Biol Reprod. 2004; 71:1755–65. [PubMed: 15329329]

27. Li G, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. 2010; 11:R22. [PubMed: 20181287]

28. Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006; 38:1348–54. [PubMed: 17033623]

29. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009; 457:854–8. [PubMed: 19212405]

30. Guelen L, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008; 453:948–51. [PubMed: 18463634]

31. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39:311–8. [PubMed: 17277777]

32. Chen X, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008; 133:1106–17. [PubMed: 18555785]

33. Splinter E, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 2006; 20:2349–54. [PubMed: 16951251]

34. Peric-Hupkes D, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell. 2010; 38:603–13. [PubMed: 20513434]

35. Fiering S, Whitelaw E, Martin DI. To be or not to be active: the stochastic nature of enhancer action. Bioessays. 2000; 22:381–7. [PubMed: 10723035]

36. Palstra RJ, et al. Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. PLoS One. 2008; 3:e1661. [PubMed: 18286208]

37. Chernukhin I, et al. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. Mol Cell Biol. 2007; 27:1631–48. [PubMed: 17210645]

38. Zlatanova J, Caiafa P. CTCF and its protein partners: divide and rule? J Cell Sci. 2009; 122:1275–84. [PubMed: 19386894]

39. Parelho V, et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell. 2008; 132:422–33. [PubMed: 18237772]

40. Fu Y, Sinha M, Peterson CL, Weng Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet. 2008; 4:e1000138. [PubMed: 18654629]

41. Hadjur S, et al. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature. 2009; 460:410–3. [PubMed: 19458616]

42. Schmidt D, et al. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. 2010; 20:578–88. [PubMed: 20219941]

43. Wendt KS, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 2008; 451:796–801. [PubMed: 18235444]

44. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–5. [PubMed: 20720539]

45. Sandhu KS, et al. Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. Genes Dev. 2009; 23:2598–603. [PubMed: 19933149]

46. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007; 448:553–60. [PubMed: 17603471]

47. Ivanova NB, et al. A stem cell molecular signature. Science. 2002; 298:601–4. [PubMed: 12228721]

48. Lim LS, et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol Biol Cell. 2007; 18:1348–58. [PubMed: 17267691]

49. Xu H, et al. A signal-noise model for significance analysis of ChIP-seq with negative control. Bioinformatics. 2010; 26:1199–204. [PubMed: 20371496]

50. Hagège H, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nature Protocol. 2007; 2:1722–33.

51. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008; 454:766–70. [PubMed: 18600261]

52. Vega VB, Cheung E, Palanisamy N, Sung WK. Inherent signals in sequencing-based Chromatin-ImmunoPrecipitation control libraries. PLoS One. 2009; 4:e5241. [PubMed: 19367334]
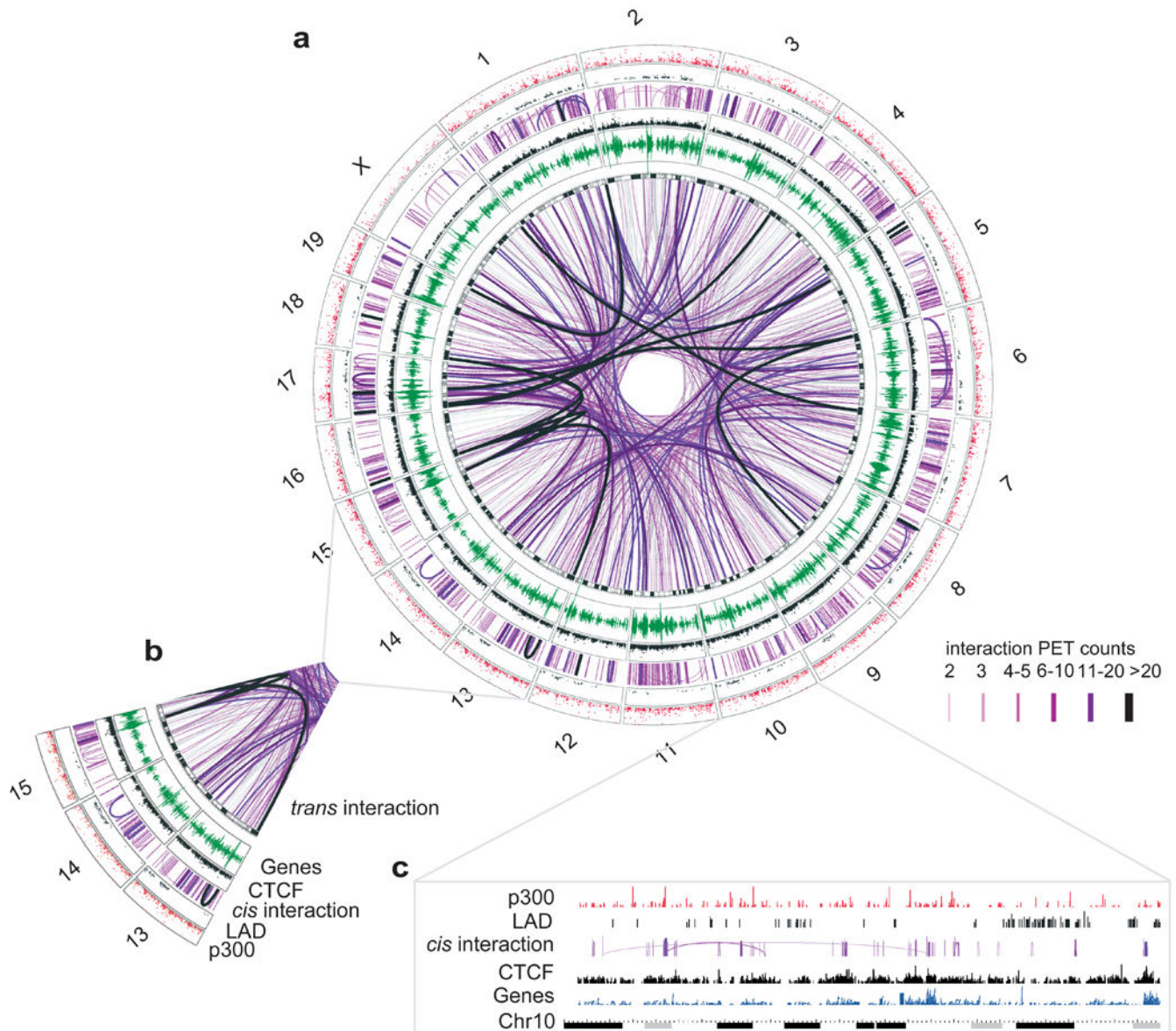
Figure 1.
Genome-wide CTCF mediated chromatin interactome.

(a) Circos map of whole genome CTCF chromatin interactome, associated genes, p300 and Lamin B occupancies from chr.1 to chr.X. The circos was generated using the Circos software package (http://mkweb.bcgsc.ca/circos/). Inter-chromosomal interactions are drawn in the inner-most ring. This is followed by the gene density track (dark green). The CTCF track (black) shows the peak signals of CTCF, followed by the intra-chromosomal interactions. The Lamin B track (dark grey) represents the peak signals of Lamin B. The p300 track (red) shows the fold change between the sample and the control.
(b) An expanded view of the interactions found from chr. 13, 14 and 15. Profiles shown in different tracks from inner to outer rings are listed accordingly. The intensity of the color is proportional to the PET counts in the cluster as shown.
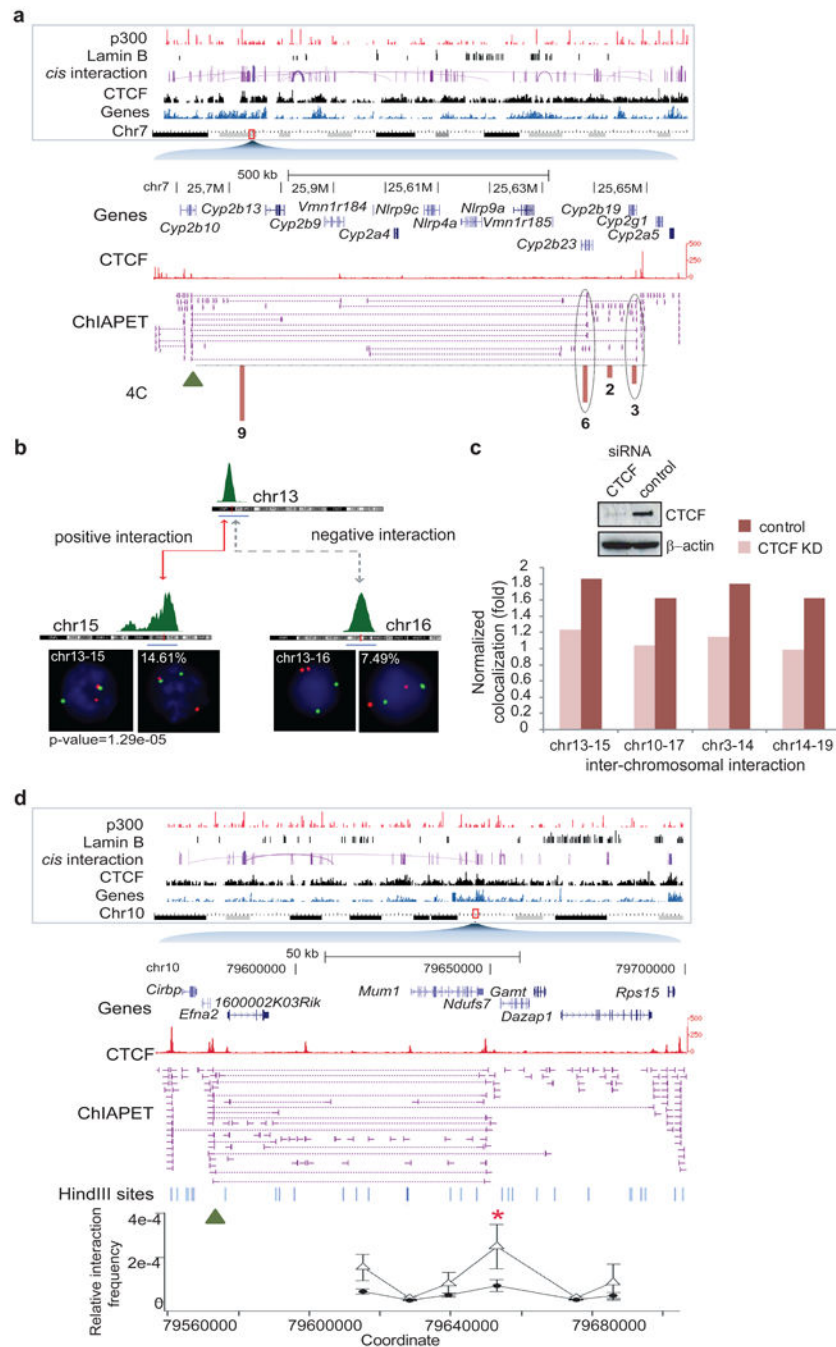
(c) Chromosome-wide view of all the intra-chromosomal interactions detected on chr.10. The content shown each track is labeled on the side.
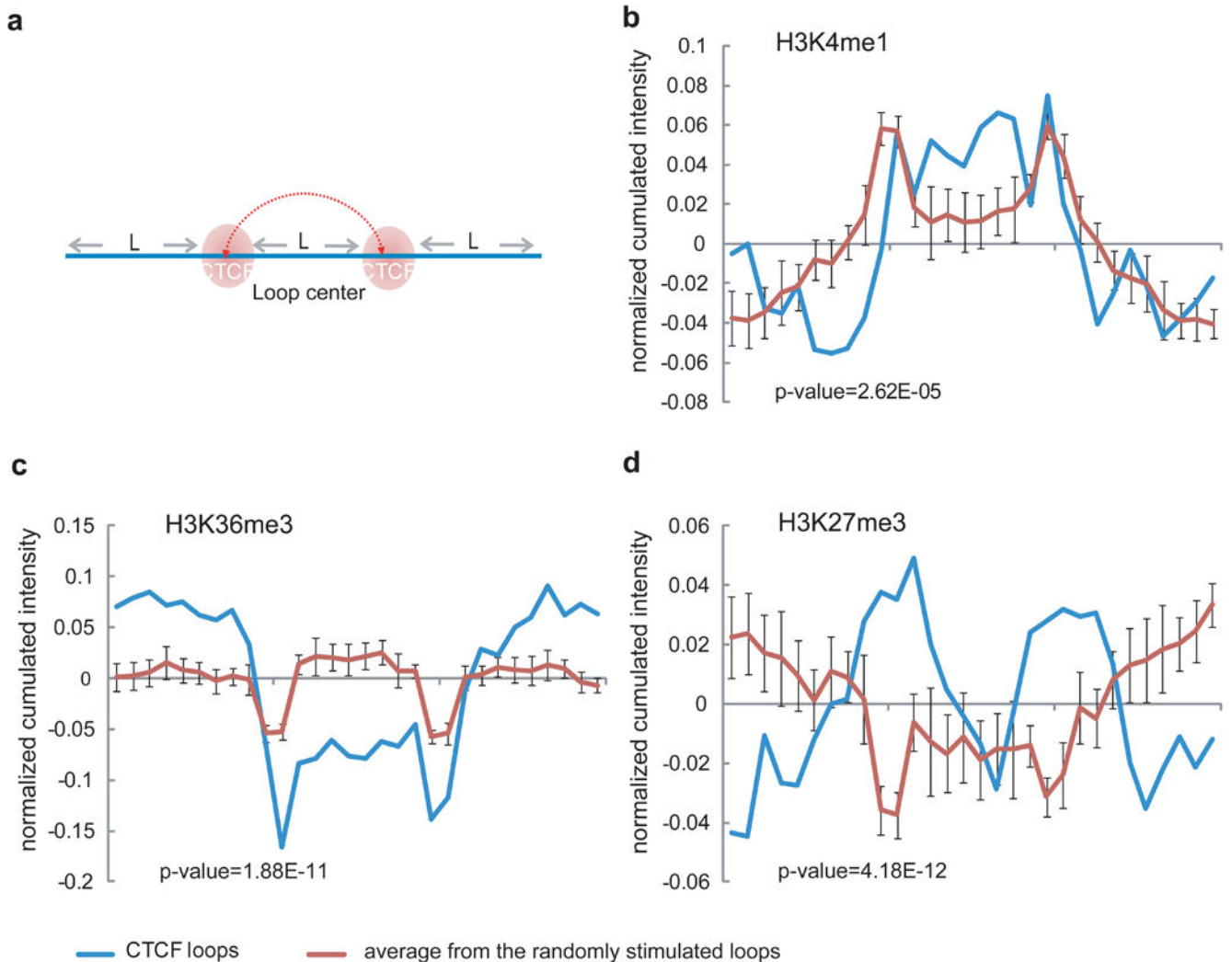
Figure 2.
Validation of CTCF-mediated chromatin interactions

(a) A region on chr7:25,586,953- 26,569,774 harboring a *cis*-interaction cluster with *Cyp2* gene family is shown as purple lines connecting CTCF binding sites (red peaks). Numbers on the lower panel show the frequencies of the interactions detected by independent 4C sequence reads. Triangle mark indicates the anchored primer location. The confirmed interactions are circled.

(b) FISH analysis confirms the *trans*-interaction between chr.13:13,658,687 and chr. 15:74,912,106 (red line). The co-localization is shown as staining of red and green fusion spots. In the control experiment (dotted line), the co-localization percentage of the negative control region (chr.16:52,100,818) is significantly lower (7.5% *vs.* 14.6%).

(c) FISH on CTCF knock-down cells. mES cells were transfected with CTCF siRNA (CTCF kd cells) or control siRNA (control cells). Western blot shows that CTCF protein in the CTCF kd cells was less than 10% of that in the control cells. The co-localization ratio were tested for 4 interaction loci (X axis). The Y axis is the fold change of the co-localization frequency between the interaction and the negative control loci.

(d) Chromosome-wide view of the *cis*-interactions detected on chr.10 (top). Middle: detailed view of a 70 Kb loop harboring *Efna2* and *Mim1* genes between 79,564,519-79,700,518. Bottom: 3C validation between the anchor (green triangle) and the distal site (red star) in the control cells (white triangle) and the CTCF kd cells (black square). Y axis is the relative interaction frequency and X axis displays the genomic coordinates. HindIII sites are marked in blue.

Figure 3.
Cumulative histone modification patterns of CTCF loops

(a) The loop model. The 1,295 loops with span less than 1Mb and their neighboring regions were first aligned; the same distance *L* of the loop span is extended to the upstream and downstream outside of the loop boundaries. For each histone modification, normalized histone modification signals are determined from the total distance 3*L*. and then normalized by their sequencing depth. A set of 1,295 negative control loops randomly selected from 2M + pairs of CTCF binding sites with similar distribution of spans and CTCF binding affinities were constructed as control. The cumulative normalized signals of H3K4me1 **(b)**, H3K36me3 **(c)** and H3K27me3 **(d)** were plotted for the comparison between CTCF loops and control loops. The intensity plots exhibit the significant different patterns between the CTCF loops and simulated control loops. The random process was repeated for 10 times to calculate the p-value as shown.
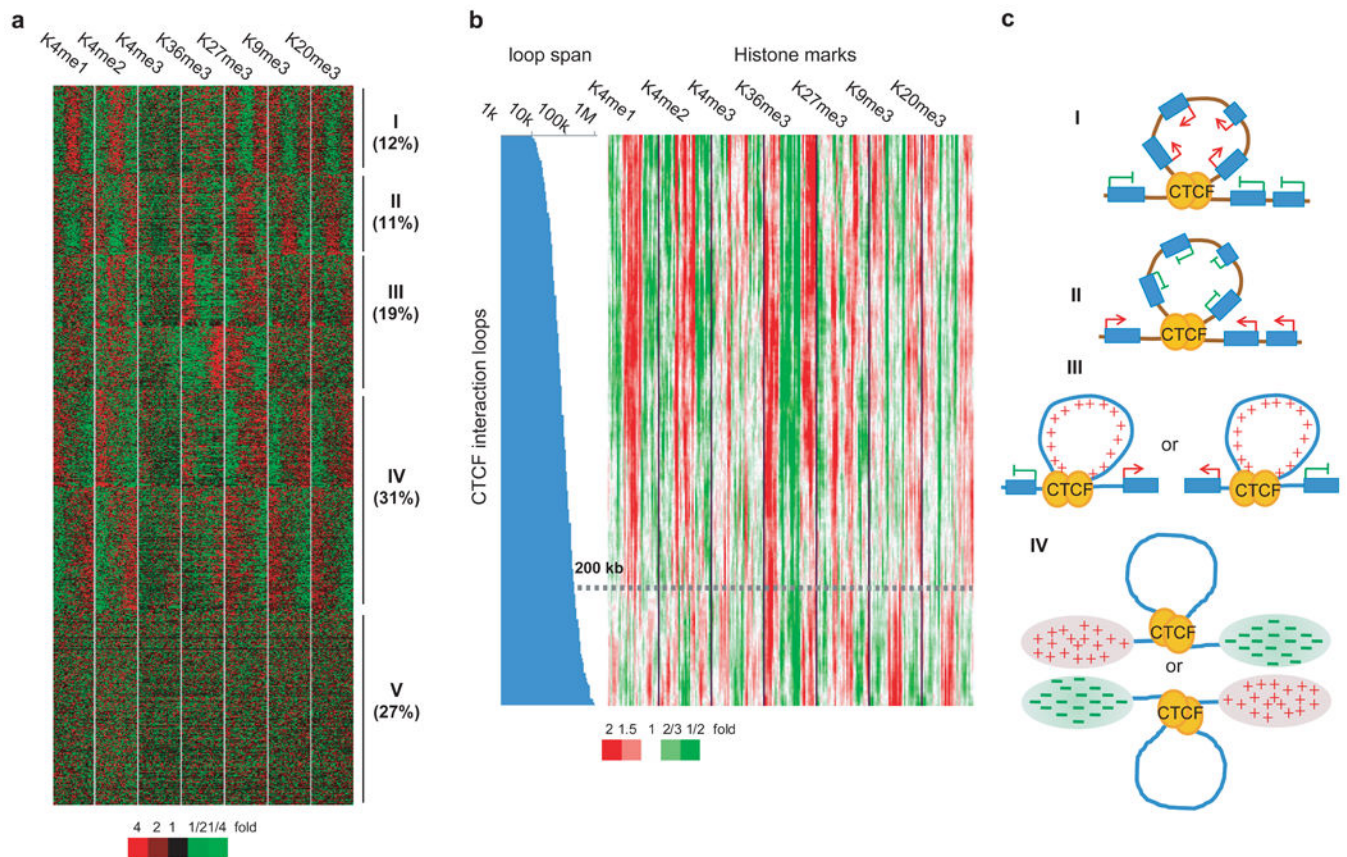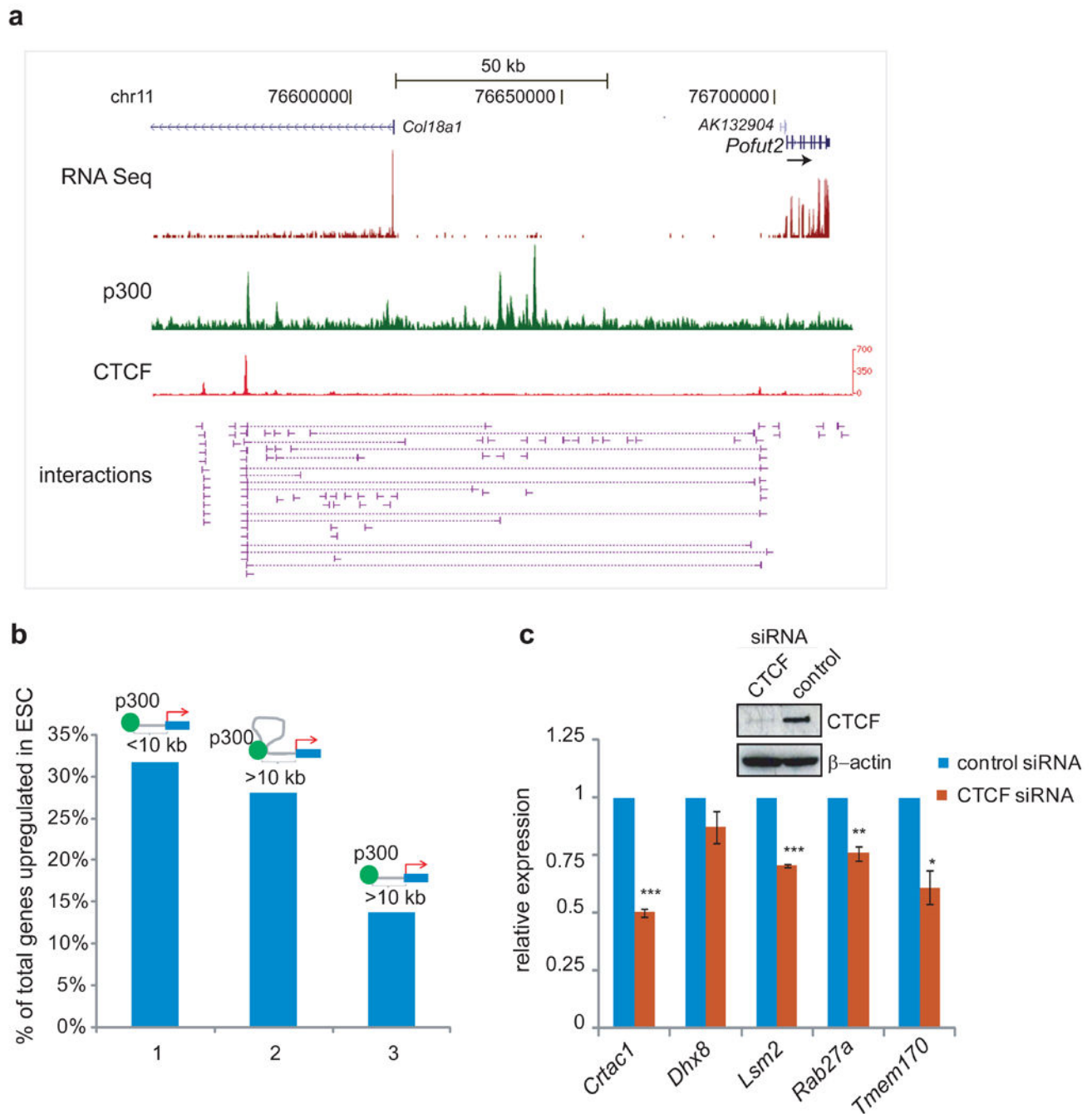
Figure 4.
Distinct types of chromatin domains defined by CTCF tethered interactions.
(a) Five chromatin domains demarcated by CTCF loops through clustering of seven histone modification signatures. For each loop region, the same distance *L* is extended upstream and downstream of the loop boundaries. Normalized signals of 7 histone modifications (listed on the top) from the total distance 3*L* (left, within and right of the loops) are determined for each loop and shown as one row. Loops exhibiting the similar combinatory signatures are clustered and symmetric clusters are grouped. The percentage of loops found in each category is listed on the left.
(b) Correlation of unique histone modification signatures with loop span. CTCF loops are sorted in ascending order of span and the hisone patterns associated with different spans are shown. Each column coresponds to an aligned bin and each row corresponds to CTCF-associated loops. A window containing 100 CTCF loops is moved vertically to average the signal. A clear transition of the histone patterns that differentiates active signals from inactive signals is observed at ~200k.
(c) The proposed models of Category I - IV based on the histone and RNAP II intensity profiles.

Figure 5.
Promoter-p300 communications facilitated by CTCF-associated chromatin interactions.
(a) A *cis*-interaction cluster (chr10:76,552,662-76,718,442; shown in connecting purple
lines) between CTCF binding sites (red peaks) connects the promoter of *Pofut2* (blue gene
track) with its p300 enhancer binding site (green) located 124 Kb 5′. RNA-Seq signal (dark
red) detected for *Pofut2* expression is shown as the track on top panel.
(b) Percentage of genes up-regulated in ES cells compared to NS cells from **1).** Genes
whose promoters and nearest p300 binding distances are less than 10 Kb, **2).** Genes whose

promoters and nearest p300 are brought in close proximity by CTCF interactions to less than 10 Kb and **3).** Genes whose promoters with distal p300 binding are more than 10 Kb apart. The ES cells and NS cells gene expression data were derived from the microarray experiments done by Mikkelsen *et al.*,[46]

(c) Expression of the genes whose promoters are connected to p300 enhancer by CTCF-associated loops. Gene expression was assessed by Real time PCR from the CTCF knock-down and control cells. Western blot shows the CTCF protein level in CTCF KD cells compared to the control cells. The expression levels were normalized against the house-keeping gene, *Gapdh* and calculated from two independent qPCR reactions of 2 independent siRNA experiments (total 2×2 qPCR). The expression levels of the genes in the CTCF knock-down cells were then normalized against those of in the control cells (=1). p-value calculated by *t*-test is indicated. ***: p-value 9.54E-05, **: p-value 5.01E-03, *: p-value 1.26E-02

Figure 6.

<u>L</u>amin-<u>A</u>ssociated <u>D</u>omains (LADs) in ES cells

(a) Multiple LADs found in a 47.4 Mb interval (chr3:97,107,505-144,477,019) represented by the profile of fold change between ChIP-Seq signal and input background (light blue track). As a comparison, LADs determined by DamID from murine ES cells are shown in black bars and CTCF loops are shown above the LAD tracks. Gene density and histone modification profiles are shown in the lower panels.

(b) CTCF loop occupancy profile within LADs and up to 1 Mb outside of LAD boundaries.

(c) Genes associated with LADs are transcriptional repressed. The distributions of gene expression levels with different locations relative to LADs are plotted. X-axis is the log value of gene expression based on Ivanova's microarray dataset [47] and the Y-axis is the percentage of genes.

(d) Percentage of genes up-regulated in ES (blue bars) or NS cells (red bars) relative to their locations to LADs are shown. Genes located within or in close proximity with LADs are preferentially up-regulated. 64% of the LAD-associated genes are up-regulated in NS cells.

Table 1

CTCF-associated interactions surrounding *Igf2-H19* locus

Interaction sites	Type of interaction	CTCF sites	Neighbor Gene	Anchor	p-value
chr1:60079609-60082500	trans	Yes	*Carf*	chr7:142387788-142388922	9.84E-05
chr1:89039633-89041093	trans	Yes	*Chrng*	chr7:142389297-142390623	8.03E-04
chr1:94749061-94750944	trans	Yes	*Rnpepl1*	chr7:142387788-142388922	7.50E-05
chr10:126566539-	trans	Yes	*B4galnt1*	chr7:142387788-142388922	5.04E-04
chr10:79197777-79199974	trans	Yes	*Palm*	chr7:142389297-142390623	6.14E-04
chr12:55898693-55900968	trans	Yes	*Baz1a*	chr7:142390695-142392310	9.00E-04
chr14:58509763-58511166	trans	Yes	*AK005323*	chr7:142389297-142390623	9.82E-04
chr14:70210649-70211883	trans	Yes		chr7:142390695-142392310	7.02E-04
chr15:63600834-63602014	trans	Yes		chr7:142387788-142388922	2.37E-04
chr16:17683117-17686871	trans	Yes	*Klhl22*	chr7:142389297-142390623	9.93E-04
chr16:91592289-91593902	trans	Yes	*Cryzl1, AK160609*	chr7:142389297-142390623	3.57E-04
chr18:56788593-56791015	trans	Yes		chr7:142389297-142390623	4.13E-04
chr2:80320088-80321852	trans	Yes	*Nckap1*	chr7:142387788-142388922	6.67E-05
chr2:91093876-91095948	trans	Yes	*1110051M20Rik*	chr7:142390695-142392310	4.58E-04
chr3:130173171-130176018	trans	Yes	*Col25a1*	chr7:142390695-142392310	9.00E-04
chr4:136734421-136736158	trans	Yes		chr7:142389297-142390623	3.79E-04
chr4:140630821-140634433	trans	Yes	*Fam131c*	chr7:142387788-142388922	3.85E-04
chr4:97886540-97887658	trans	Yes	*Inadl*	chr7:142387788-142388922	3.50E-05
chr5:123176256-123177952	trans	Yes	*Fbxl10*	chr7:142389297-142390623	7.03E-04
chr6:122217499-122218680	trans	Yes		chr7:142390695-142392310	5.34E-04
chr7:111250584-111252190	cis	Yes		chr7:142389297-142390623	2.35E-05
chr7:117803429-117804915	cis	No		chr7:142389297-142390623	4.31E-05
chr7:131874490-131875099	cis	No	*AK212323*	chr7:142389297-142390623	4.31E-05
chr7:142140090-142141395	cis	No	*6330512M04Rik*	chr7:142389297-142390623	2.16E-05
chr7:142236751-142239685	cis	Yes		chr7:142389297-142390623	6.46E-08
chr7:142289981-142291608	cis	Yes	*Lsp1*	chr7:142389297-142390623	1.72E-04
chr7:142346303-142348621	cis	Yes	*Mrpl23*	chr7:142387788-142388922	3.53E-05
chr7:142346303-142348621	cis	Yes	*Mrpl23*	chr7:142389297-142390623	3.88E-04

Interaction sites	Type of interaction	CTCF sites	Neighbor Gene	Anchor	p-value
chr7:142493990-142493990	cis	No		chr7:142390695-142392310	5.88E-06
chr7:142718721-142720636	cis	Yes	*Th*	chr7:142389297-142390623	1.94E-04
chr7:4669920-4670465	cis	No	*U2af2*	chr7:142389297-142390623	2.16E-05
chr7:48636678-48636804	cis	No		chr7:142390695-142392310	5.88E-06
chr7:80258483-80260771	cis	Yes	*Fes*	chr7:142389297-142390623	4.31E-04
chr7:96218771-96219753	cis	No	*Odz4*	chr7:142389297-142390623	2.16E-05
chr9:55998012-55999800	trans	Yes	*C230081A13Rik*	chr7:142387788-142388922	2.93E-04
chrX:109019796-109022418	trans	Yes		chr7:142390695-142392310	7.48E-04
chrX:131320743-131322427	trans	Yes		chr7:142389297-142390623	9.59E-04