CTCF mediates the cis-regulatory hubs in mouse hearts.

Mick Lee[1,2,*], Loïc Mangnier[3,4,5,6,*], Cory C. Padilla[7], Dominic Paul Lee[1,2], Wilson Tan[1,2], Wen Hao Zheng[1,2], Louis Hanqiang Gan[1,2], Ching Kit Chen[1,2,8,9], Yee Phong Lim[1,2,8,] Rina Miao Qin Wang[1,2,8], Peter Yiqing Li[1,2], Yonglin Zhu[12,13], Steve Bilodeau [5,6,10,11], Alexandre Bureau[3,4,5], Roger Sik-Yin Foo[1,2,], Chukwuemeka George Anene-Nzelu[1,2,12,13,]

[1]Cardiovascular Disease Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[2]Genome Institute of Singapore, A*STAR, Singapore

[3]Centre de Recherche CERVO, Quebec City, Canada.

[4]Département de Médecine Sociale et Préventive, Université Laval, Quebec City, Canada.

[5]Centre de Recherche en données Massives de l'Université Laval, Québec, Québec, Canada, G1V 0A6.

[6]Centre de Recherche du Centre Hospitalier Universitaire de Québec – Université Laval, axe oncologie, Québec, Québec, Canada, G1R 3S3.

[7]Cantata Bio, 100 Enterprise Way Suite A-101 Scotts Valley, CA 95066

[8]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

[9]Khoo Teck Puat-National University Children's Medical Institute, National University of Singapore, Singapore

[10]Centre de Recherche sur le Cancer de l'Université Laval, Québec, Québec, Canada, G1R 3S3.

[11]Département de Biologie Moléculaire, Biochimie Médicale et Pathologie, Faculté de Médecine, Université Laval, Québec, Québec, Canada, G1V 0A6.

[12]Montreal Heart Institute, Montreal, Quebec, Canada

[13]Department of Medicine, University of Montreal, Quebec, Canada

*These authors contributed equally

[#]Correspondence Chukwuemeka.anene-nzelu@umontreal.ca

ABSTRACT

The 3D chromatin architecture establishes a complex network of genes and regulatory elements necessary for transcriptomic regulation in development and disease. This network can be modelled by cis-regulatory hubs (CRH), which underscore the local functional interactions between enhancers and promoter regions and differ from other higher-order chromatin structures such as topologically associated domains (TAD). The Activity-by-contact (ABC) model of enhancer-promoter regulation has been recently used in the identification of these CRHs, but little is known about the role of CTCF on the ABC scores and the consequent impact on CRHs. Here we show that the loss of CTCF leads to a reorganization of the ABC-derived rankings of the putative enhancers in the mouse heart, a global reduction of the total number of CRHs and an increase in the size of the CRHs. Furthermore, CTCF loss leads to a higher percentage of CRHs that cross TAD boundaries. These results provide another layer of evidence to support the importance of CTCF in the formation of regulatory networks necessary for gene regulation.

Keywords: Cis-regulatory hubs, enhancers, CTCF, Activity-by-Contact, mouse cardiomyocytes (CM)

Summary: Deletion of CTCF in mouse cardiomyocytes led to reorganization of the activity-by-contact scores of the heart enhancers and changes in the cis-regulatory hubs

INTRODUCTION

Unravelling the role of the non-coding regulatory elements in gene regulation is important to gain mechanistic insight into key biological processes such as cell fate determination and disease pathogenesis [1] . Accordingly, tremendous effort has been deployed to annotate these regulatory elements in different tissues thanks to innovative sequencing techniques and computational tools [2-4]. These techniques include Chromatin immunoprecipitation with sequencing (ChIP-seq), Assay for Transposase Accessible Chromatin with sequencing (ATAC-seq), and Chromatin Capture conformation with high-throughput sequencing (Hi-C) [5,6]. In addition, computational methods such as the Activity by Contact (ABC) algorithm [3], and other tools [7,8] have been developed to integrate these datasets and identify enhancer-gene (E-G) pairs required for development and disease [9-11].

Besides the direct identification of E-G pairs, other studies have demonstrated that enhancers and promoters interact dynamically within a network called cis-regulatory hubs (CRH), which allows for multiple regulatory connections [12-15]. These CRHs can be defined as 3D regulatory networks that constitute a complex organization of multi-loci connections of enhancers and promoters, connected in 3D space within the nucleus. They highlight direct and indirect contact between genes and distal regulatory elements and can help predict target genes for disease-relevant non-coding SNPs [14]. They are often bound by a cluster of transcription factors (TF) and have been implicated in gene co-regulation, lineage specification and disease development [13,14,16]. These hubs differ from other known 3D structures, such as genomic compartments, which constitute euchromatin or heterochromatin regions. They also differ from topologically associated domains (TADs) defined as units within which most enhancer-promoter interactions occur [17]. Although the identification of these CRHs is often done solely through detailed analysis of 3D contacts from Hi-C data, [13,18-21] a recent study proposed building these CRHs using the ABC algorithm [14]. The ABC algorithm combines ATAC-seq, ChiP-seq and Hi-C to rank the E-G pairs based on their regulatory impact [3,22]. Thus in addition to identifying and ranking enhancers, the ABC algorithm can be used to generate a more functionally relevant CRH[14]. CRHs built using ABC

were strongly enriched for disease-relevant genes and also helped explain disease heritability [14,23].

Given that the 3D genome architecture through HI-C is a key component of the ABC algorithm, which is used to construct the CRHs, [14], it is of huge importance to understand the role of architectural proteins like CTCF in these CRHs. The CTCF protein often called the "genome weaver" plays a crucial function in the "folding" of the genome and bringing of enhancers into proximity to their distant target genes [17,24]. We and others have shown the importance of CTCF in regulating cardiac chromatin architecture and heart disease[25,26], However, CTCF's role in regulating CRHs has not been explored. In this study, using mouse cardiomyocytes as a model of terminally differentiated non-dividing cells, we employ the ABC model to identify top mCM gene-enhancers and then annotate the CRHs. We analyze the characteristics of the CRHs containing tissue-specific genes, showing that these genes are often found in multi-enhancer CRHs. We also show that in disease, there is a positive regulation of genes and enhancers within a CRH. Finally, we show that loss of CTCF leads to a merging of CRHs and the formation of new CRHs in the *Ctcf*-KO cells with resultant drastic change in the number and distribution of enhancers that cross the ABC threshold.

RESULTS

Mouse Cardiomyocyte Enhancer Landscape Identified Through Activity-By-Contact Algorithm

We performed ATAC-seq, H3K27ac ChiP-seq and HiChIP, to generate the ABC scores for the enhancer-gene pairs in the control adult mouse cardiomyocytes. Using an ABC cut-off of 0.01 [14], we identified ~7,000 out of 156,988 putative distal regulatory regions marked by H3K27ac peaks and/or ATAC peaks in the adult mouse CM that crossed the threshold. This represents about 5% of all putative regulatory regions predicted to have strong regulatory effects on their target genes and is similar to what was observed in another study [3]. With regards to the enhancer-gene (E-G) pairs, globally there were 34,496 E-G interactions with an average of 4.19 connections per gene and 5.29 connections per enhancer in these healthy adult mouse CMs. Figure 1A shows examples of the top enhancers for 2 CM disease-relevant genes *Mybpc3* and *Myh6*. The comprehensive list of ABC-linked enhancers and target genes can be found in supplementary Table 1. A literature search confirmed that at least 3 of the identified enhancers in our study have been validated. These include the *Nppa/Nppb* super-enhancer located upstream of both genes which has been shown to play a critical role in the stress gene response of *Nppa/Nppb* during pressure-overload-induced CM stress [25,27,28]. A separate enhancer for *Myh7* identified from the ABC scoring has also been validated previously in the mouse heart [29]. Deletion of this enhancer region led to the downregulation of *Myh7* in mouse hearts[29]. Taken together, these previous studies provide support for the validity of the ABC model.

For a better understanding of some of the features of highly ranked enhancers, we focussed our attention on the top 1% and bottom 1% of enhancers based on the ABC score. The enhancers in the top 1% had scores of 0.9 – 1, while those in the bottom 1% had scores of 0.01. First, we analyzed the average distance of these enhancers to the identified target genes. As expected, enhancers with high scores were more likely to be closer to the target genes, with mean and median distances of 83 kb and 54 kb, respectively, as compared to

the mean and median distances of the bottom 1% of enhancers which were 950 kb and 699 kb respectively (P value < 0.01). Next, we performed a gene ontology (GO) analysis of the genes linked to these enhancers. The GO biological processes found in the genes linked to top 1% of enhancers include terms such as striated muscle cell differentiation, regulation of blood circulation and actin cytoskeletal organization (Supplementary table 2). These are processes that are important for CM function and include genes such as *Gata4*, *Mybpc2*, *Cox7a1* and *Myl3*. On the other hand, the GO for genes in the bottom 1% had generic terms like RNA processing, protein modification and cellular metabolic processes (Supplementary Table 2).

To provide further evidence that these ABC-linked enhancers regulate their putative target genes, we studied the correlation between changes in gene expression versus changes in peak height of the ABC-linked enhancers during pressure overload-induced CM hypertrophy. Using RNA-seq and H3K27ac ChIP-seq datasets from mouse hearts subjected to transverse aortic constriction [25], we found that changes in gene expression correlate modestly with changes in H3K27ac peak signals (Pearson correlation= 0.34, $P < 2.2e-16$) (Figure 1B). The correlation was stronger when focusing only on genes and enhancers with greater or less than $\log_2$ 0.5 fold change after Transverse aortic constriction (Supplementary Figure 1A) (Pearson correlation = 0.51, $P < 2.2e-16$). This result shows that globally, upon external stimuli, gene expression is regulated in the same direction as their ABC-linked enhancers.

Loss Of CTCF Leads To Changes In The Enhancer Interactome And Changes In The ABC Scores Of Putative Enhancers

Next, to elucidate the role of CTCF in ABC scores and regulation of the CRHs, we also applied the ABC algorithm on the *Ctcf*-KO mouse CMs and observed that 36,226 regulatory elements crossed the 0.01 threshold. This resulted in 71,080 E-G interactions with a similar number of expressed genes (~8,000) (Supplementary Table 1) in control and *Ctcf*-KO samples. This led to an increase in the average number of connections per gene to 8.14 in *Ctcf*-KO samples. Of the 71,080 enhancer-gene pairs only 20% (14,200 E-G pairs) were shared between control and *Ctcf*-KO. This increase in the number of enhancers with ABC

score greater than 0.01 could be due to increased interactions between genes and ectopic enhancers upon CTCF loss. Interestingly we observed that while the highest-ranked ABC enhancer in the control had a score of 0.99, the highest-ranked enhancer in the *Ctcf*-KO sample had a score of 0.78 (scale of 0 – 1). Indeed the average score in the control was 0.054 while the average score in the *Ctcf*-KO was 0.028 (TTest P-value < 2.2e-16). This suggests that the loss of CTCF led to a gain of ectopic enhancers in the *Ctcf*-KO with a redistribution of the ABC scores of the enhancers.

Loss Of CTCF Alters The Cis-Regulatory Hubs

We integrated the ABC results to generate the CRHs focusing first on the control samples. We identified 1,522 hubs in these control cardiomyocytes with about 70% of the hubs containing fewer than 5 elements (Figure 2A). Figure 2B shows 2 examples of such hubs containing CM genes: *Myom1* a gene involved in the formation of myofibrils [30], and found in a simple hub. In contrast, *Tnni3* a CM-specific sarcomeric gene is found in a more complex hub. Consistent with the correlation between gene expression and ABC-linked enhancers in heart disease, we observed that genes and enhancers within the same hubs are positively correlated in the same direction of change during CM stress (Pearson correlation of 0.32 (P <2.2e-16)) (Figure 2C). This suggests a co-regulation of genes and enhancers found within the same hub upon external stimuli. We then ranked CM genes by FPKM and selected the top 50 genes, which were mostly CM-specific genes, to glean biological insights about the characteristics of CRHs that harbour these genes. Our data revealed that highly expressed CM genes were more likely to be in hubs with more regulatory elements than non-highly expressed genes (Figure 2D). Such multi-enhancer hubs have been proposed as a mechanism to buffer against external stresses and genetic perturbations and provide phenotypic robustness to disease-relevant genes [23,31].

Next, we analyzed the CRHs in the *Ctcf*-KO CMs and found a marked reduction in the total number of CRHs from 1522 in control to 660 CRHs in the KO, accompanied by an increase in the number of elements per hub. Figure 3A shows an example of the same hubs containing 2 CM genes shown in Figure 2A, showing an increase in the number of elements within each hub. Figure 3B shows the distribution of elements per hub, confirming a striking increase in

the percentage of hubs with 10 or more elements. Associated with this increase in the percentage of complex hubs, there was an increase in the average number of connections per gene in the *Ctcf*-KO cells, and a decrease in the number of connections per enhancer (Figure 3C, Supplementary Figure 1B). The decrease in the number of connections per enhancer is due to the overall increase in the number of enhancers that crossed the ABC threshold while maintaining the same number of expressed genes. Next, we asked if this reduction in the total number of CRHs in the *Ctcf*-KO was merely due to the merging of CRHs or the formation of new CRHs, and observed both cases. The merging of CRHs was observed primarily in the simple hubs (2-3 elements per hub) located next to each other as 403 (82%) of 492 simple hubs merged with 1 or more other hubs to form larger hubs. For the larger hubs, there was redistribution and formation of new hubs, as genes gained new interactions and lost other interactions. To further confirm the importance of CTCF in the organization of these hubs, we compared CTCF binding sites in the control hubs vs *Ctcf*-KO hubs. Our analysis showed an enrichment of CTCF in the control hubs (fisher exact test odds ratio of 1.69, two-sided $P < 2e-16$). Next, we analyzed the relationship between CRHs and TADs since TAD boundaries are enriched for CTCF [25], and earlier studies have shown that CRHs are generally constrained within TAD boundaries [14]. Indeed, using the TADs in the control as a reference, our data showed that about 80% of CRHs in the control were contained within the same TAD while only 20% spanned more than 1 TAD. In contrast, about 45% of CRHs in KO were contained within a TAD while 50% of CRHs in the KO spanned more than 1 TAD (Figure 3D) while 5% were not within TAD boundaries, showing that the loss of CTCF leads to a reorganization of the CRHs with the formation of cross-TAD enhancer-promoter interactions.

Finally, we sought to identify if CRH changes were directly correlated to gene expression after CTCF loss. We have previously shown that CTCF KO leads to CM dysfunction and global transcriptomic changes after 2 weeks [25], here we wanted to gain deeper insight into how that relates to CRH. Our analysis showed that despite the changes in the structure of CRHs, these changes were not directly reflected in gene expression. Indeed there was no correlation between gain or loss of CRH connections and gene expression, suggesting that other secondary effects of the CTCF KO may have come into play to regulate the gene expression profile at 2 weeks when the RNA-seq was performed. This has been observed in

other studies that showed that the direct effect of CTCF or Cohesin (another architectural protein) on RNA-seq is observed within a few hours of deletion [32,33]. As time progresses, other secondary factors come into play and obscure the direct effect of these proteins on gene expression.

DISCUSSION

Non-coding *cis*-regulatory elements (CRE) play crucial roles in development and disease and hold promise for next-generation therapeutic targets. While different studies by us and others have annotated these CREs and CRHs in different cellular models, the role of CTCF in the ABC-derived CRHs has yet to be studied. First, our study applied the ABC algorithm to the mouse heart, generating to the best of our knowledge, the first such ranked enhancer scores in mice hearts. Our findings confirm the positive correlation between genes and their ABC-linked enhancers in cardiac disease and affirm that genes and enhancers within the same hub are co-regulated during pressure overload-induced CM hypertrophy. Integrating ABC to identify CRHs may thus represent another method to analyze the effect of regulatory SNPs on target genes, as the SNP may affect a target gene when they are in the same hub, even when there is no direct link through pair-wise E-G analysis [14]. Analysis of the CRHs also suggests that tissue-specific genes are more likely to be contained in hubs with high connectivity and rich in distal elements. This increased number of E-P networks provides redundancy and phenotypic robustness while guaranteeing increased transcriptional activity thanks to the multi-enhancer networks.

Loss of CTCF affects the enhancer interactome, particularly impacting the H3K27ac interactions co-bound by CTCF. This leads to E-Ps that span larger distances and gain of new interactions. The loss of CTCF also leads to a re-organization of ABC-selected enhancers with a direct impact on the CRHs, resulting in fewer hubs but more elements per hub. This implies a crucial role for CTCF in determining which enhancers cross the ABC threshold and the organization of the ABC-derived CRH. While this study focussed on the effect of loss of CTCF on overall CRH structure and its relationship with TAD boundaries, future studies will examine individual CRHs and their relationship to CTCF binding to ascertain how loss of CTCF

can lead to local dysregulation of CRHs. Indeed, this finding has implications for mutations that affect local CTCF binding at regions that are not TAD boundaries, as it may lead to the merging of 2 or more CRHs and, thus, the establishment of an indirect connection between genes and enhancers within the newly formed hubs. In addition, given that the binding of CTCF to DNA is methylation-sensitive [34,35], this also supports findings that conditions that lead to DNA hypermethylation will reduce local or global CTCF binding [36]. This loss of CTCF binding will thus lead to the loss of insulation at TAD boundaries with a consequential gain of ectopic enhancers and reorganization of the ABC enhancers [36].

CONCLUSION

Our findings demonstrate that CTCF plays a critical role in determining the regulatory strength of enhancers as well as the membership and organization of the CRHs. Given that the CRHs play a role in disease and cell type-specific gene expression, processes that affect CTCF expression and genomic binding will lead to a disorganization of CRHs and a subsequent effect on transcriptional regulation upon external stimuli. Although we didn't find a correlation between the reorganization of CRHs and gene expression upon the loss of CTCF, this is mostly due to other secondary effects in the regulation of gene expression. We, however, believe that our findings reveal important principles of the role of CTCF in regulating CRHs and ABC scores. A major limitation of our study is that the resolution of 10kb for our HiChIP is not optimal as each 10kb window may contain various enhancers, and hence, our analysis may miss out on some enhancer-gene interactions. In addition, as the various components of the ABC analysis, namely ATAC, ChIP and HiC, may have technical variations, an independent ABC analysis would be valuable to confirm our findings. We, however, believe that our study forms the foundation for future analysis on the role of CTCF in the ABC enhancer scores and as such the relative contribution of enhancers to each gene. Our future studies will delve deeper into these questions.

METHODS

Animal Experiments

Animal experiments were performed under a license approved by the Institutional Animal Care and Use Committee (National University of Singapore). *Ctcf*$^{flox/flox}$ mice harboring LoxP sites flanking exons 3 and 12 of the *Ctcf* gene against a C57/bl6 strain background were used as previously published. AAV9-cTnt-eGFP and AAV9-cTnt-Cre-tdTomato vectors encoding codon improved Cre recombinase or eGFP under the control of cardiac troponin T promoter were used for control and *Ctcf* knock-out respectively [25]. Experiments were performed on adult 6- to 8-week-old mice. Cardiomyocytes were isolated from mouse hearts after 2 weeks of AAV9 injections for control and *Ctcf*-KO mice. Adeno-associated virus 9 (AAV9) virus for the delivery of Cre recombinase was generated as previously described [25].

H3k27ac HiChip Library Prep And Data Analysis.

Mouse adult cardiomyocytes were isolated following previously published protocol [25]. MNase-based HiChIP assay was performed on $2x10^6$ isolated CMs using the Dovetail Genomics HiChIP MNase kit protocol [37]. Frozen cells were resuspended in 1× PBS and crosslinked with 3 mM DSG and 1% formaldehyde. Washed cells were digested with 0.75 µl MNase in 100 µl of nuclease digest buffer with $MgCl_2$. Cells were lysed with 1×

RIPA, and clarified lysate was used for ChIP. H3K27ac antibody (Abcam 04729) was used to perform chromatin immunoprecipitation. The Protein A/G bead pulldown, proximity ligation, and libraries were prepared as described in the Dovetail protocol (Dovetail HiChIP MNase Kit). Libraries were sequenced on an Illumina Novaseq (Supplementary table 3 shows the sequencing QC).

HiChIP paired-end reads were aligned with BWA MEM [38] (version 0.7.17r1198-dirty) with the -5SP flag enabled with an index containing only the canonical chromosomes of the mm10 genome (available from the UCSC genome). The resulting alignments were then parsed with pairtools (versions 0.3.0) with the following options --min-mapq 40 –walks-policy 5unque –max-inter-align gap 30 and the thre –chroms-path file corresponding to the size of the chromosomes used for the alignment index. Parse pairs were deduplicated, sorted with pairtools. Valid pairs were identified and through pairtools select (pair type = 'UU', 'UR', 'RU', 'uu', 'uU' and 'Uu') and downsampled to the lowest number of valid pairs in a sample (126 million in LACZ) with pairtools sample. Contact matrices in .hic format were generated with juicetools pre function (version 1.22.01).

FitHiChIP [39] (version 9.1) was used to identify significant interactions ("loops") from the valid subsampled pairs at 10kb resolution with the following settings, loop type = all-to-all, coverage = bias correction, merge redundant loops = Yes, background model = FithHiChIP(L), FDR < 0.1, minimum interaction size =20kb, maximum interaction size = 2 mb. Conditionally unique (meaning only occurring in LACZ or CRE) and shared loops were identified with bedtools pairToPair (version 2.28.0) requiring both loop anchors to overlap the same coordinates be flagged as shared. Loop size was assessed by performing a two-sided Wilcoxson's rank-sum test on the distribution of loop ranges (distance between the two anchors). Loop anchors were then intersected with the LACZ CTCF binding sites to determine the proportion of loops resulting from CTCF presence. P-values for CTCF mediated loops were obtained through a two-sided Fischer's Exact test. Aggregate Peak Analysis (APA) [32] was computed with juicetools APA function. Loop anchors were used to as the sites to aggregate over at 10kb resolution. APA enrichment scores, loop center from the lower-left (LL) corner is shown as both an APA score (ratio of the mean of center to mean of LL) and a Z-score. APA normed output was used to plot the APA matrices.

H3K27ac HiChIP data were summarized and visualized with R, the package eulurr was used for the shared and unique loop Venn diagrams and ggplot2 for loop size and proportion of loops with CTCF binding site at the anchor. Additionally, ggplot2 was used to plot APA matrices with the geom_raster function, with the color scale set to the same min – max limits across all APA plots. At sites of interest, HiChIP loops files (in longrange format) were used to visualize loops along with ChIP-seq coverage and RNA-seq activity in the WashU Epigenome Browser (https://epigenomegateway.wustl.edu/). Supplementary Table 4 lists all the loops

RNA-seq

RNA was extracted from 3 biological replicates of control and CTCF KO ventricular CMs (1 million cells each). The paired-end libraries were constructed using Tru-seq kits (Illumina) and resulting libraries were sequenced on the Novaseq platform, generating 2×150-bp paired-end reads. The RNA reads were mapped to the mouse genome with Tophat version 2.0.11 with default parameters, and gene count was computed with htseq-count. Differential gene expression analysis was performed with EdgeR [40].

Chip-Seq Experiment

ChIP experiments on isolated CM (1 million cells each) were performed as described previously [25]. Briefly, CMs were cross-linked with 1% formaldehyde for 10 minutes at room temperature and quenched with glycine (0.125 mol/L final concentration) [25]. Cells were then washed in ice-cold PBS and pelleted and lysed in FA lysis buffer (10 mmol/L Tris-HCl, pH 8.0, 0.25% Triton X-100, 10 mmol/L EDTA, 0.1 mol/L NaCl). To facilitate cell lysis, the cell pellet was passed through a 27.5-gauge needle gently 20 times. Nuclei were pelleted by centrifugation resuspended in sonication buffer, and chromatin was fragmented via sonication to an average size of 200 to 300 bp (Bioruptor Plus, Diagenode). Chromatin was immunoprecipitated against H3K27ac (Abcam ab4729) or CTCF (EMD Millipore, catalog No. 07–729) overnight. Antibody-chromatin complexes were pulled down with Protein G Dynabeads (Invitrogen, catalog No. 10003D), washed with 0.1% SDS lysis buffer, and eluted with elution buffer (1% SDS, 10 mmol/L EDTA, 50 mmol/L Tris-HCl, pH 8). After cross-link reversal (4 hours of incubation at 65°C) and proteinase K treatment, immunoprecipitated

DNA was extracted. ChIP DNA was quantified by fluorometric quantification (Thermo Fisher Scientific, Qubit dsDNA HS assay kit, catalog No. 32851). Library preparation was performed with the New England Biolabs Ultra II Kit according to the manufacturer's specifications and sequenced on the Illumina NextSeq High platform.

ATAC-Seq Experiment And Analysis

The ATAC-seq was performed for both control and *Ctcf* KO CMs according to previously published Omni-ATAC protocol [41] Briefly, 50,000 adult cardiomyocytes were pelleted at 500 g for 5 minutes. The cells were resuspended in 100 µl ATAC-resuspension buffer containing 0.5% NP40, 0.5 % Tween-20 and 0.01% Digitonin. The cells were left on ice for 5 minutes after which 1 ml of cold ATAC-RSB containing 0.1% Tween 20 was added to wash out the lysis buffer. The nuclei were pelleted at 500 g for 5 minutes, the supernatant was carefully removed and the nuclei were resuspended in 50 µl transposition mixture containing 25 µl of 2x TD buffer, 16.5 µl of PBS, 5 µl nuclease free water, 0.5 µl of 1% digitonin, 0.5 µl of 10% Tween-20 and 2.5 µl of transposase (Illumina Tagment DNA enzyme 1, Catalogue number 20034198). The reaction was incubated at 37°C for 30 minutes in a thermoshaker with 1000 RPM. After the reaction, DNA was extracted using the NEB Monarch® PCR & DNA cleanup kit (Catalogue number T1030L), PCR was performed using the Illumina/Nextera primers after which Ampure XP beads were used for library cleanup. The resulting library was sequenced on Nextseq.

ChIP-Seq and ATAC-Seq Data Analysis

For both H3K27ac ChIP-Seq and ATAC-Seq, fastq files were aligned to the mm10 reference genome using the Burrows-Wheeler alignment tool in bwa software [38]. The corresponding bam files were obtained using samtools [42], while peak calling was done with macS2 software [43]. All peaks with a q-value less or equal to 0.1 were considered as statistically significant and used in the Activity-By-Contact Score [3] for CRH construction [14]. Supplementary table 5 lists H3K27ac peaks in Sham and TAC-treated mouse cardiomyocytes, as well as those in control vs *Ctcf*-KO cardiomyocytes.

ABC-Score And CRH Analysis

The ABC model defines active enhancers based on a quantitative score of DNAse or ATAC-seq, H3K27ac, and normalized Hi-C contact number[3]. This score is computed relative to a background activity over a 5-Mb window around a candidate element. Then, we set the threshold to 0.01; beyond which a candidate element is considered as a distal element. As an extension of the ABC-Score, CRHs [14] were defined as bipartite networks between promoters and distal elements (igraph R package) [44] TADs were called using the directionality index as previously described. Pearson correlation analysis was performed using R package

DATA AVAILABILITY

The raw data supporting the conclusion of this article has been deposited in a public repository.

ACKNOWLEDGEMENTS

We thank Cantata Bio for providing the HiChIP kits. We thank Dr. Charles Joly-Beauparlant and Dr. Tania Cuppens for helping with data analysis and result interpretation

AUTHORS' CONTRIBUTION

M.L, D.P.L, W.Z, L.G, C.C.K, Y.L, R.W, C.G.A.N performed the experiments, M.L, L.M, C.P, W.T, S.B, A.B, C.G.A.N provided data analysis. S.B, A.B, R.F, C.G.A.N Provided supervision for experiments and data analysis

CONFLICT OF INTEREST

Cory C. Padilla works for Cantata Bio, developers of the HiChIP kit used in the study. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

FUNDING

This work was supported by grants from the National Medical Research Council (to Dr Foo), and Start-up grants from the Montreal Heart Institute (Dr. Anene-Nzelu). Some data analyses were performed on computing resources from Compute Canada.

DECLARATIONS

Animal experiments were performed under a license approved by the Institutional Animal Care and Use Committee (National University of Singapore).

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable.

FIGURE LEGENDS

Figure 1:

A) USCS screenshot showing 2 examples of CM genes and their top ABC-linked enhancers
B) Pearson correlation coefficient showing correlation between differentially expressed genes and differential enhancer peaks for ABC-linked enhancers during cardiac disease

Figure 2: CRH in control Cardiomyocytes

A) Bar chart showing distribution of number of elements per CRH
B) Examples of 2 CRHs containing CM genes *Myom1* and *Tnni3* in the control cardiomyocytes
C) Pearson correlation coefficient between differential expression of genes and enhancer peaks contained within the same CRHs during CM stress
D) Box plot showing characteristics of CRHs containing CM-specific genes. These CRHs tend to have more enhancers and promoters

Figure 3: CRHs in the KO cells

A) Examples of 2 CRHs containing CM genes *Myom1* and *Tnni3* in the *Ctcf*-KO cardiomyocytes. Compared to the same CRHs in control, there is an increase in the number of elements in each hub
B) Bar chart showing distribution of elements per CRH in Control and *Ctcf*-KO. There is an increase in the percentage of CRHs with 10 or more elements from 15% in the Control to 60% in the CTCF KO cardiomyocytes

C) Bar chart showing number of connections per gene promoters and per enhancers. There is an increase in the average number of connections per gene in the *Ctcf*-KO CMs. $P < 0.01$

D) Bar chart showing overlap of CRHs with TADs, 80% of TADs are contained in one TAD in the control while only 40% of CRHs are contained within one TAD in the CTCF KO. In contrast 20% of CRHs span 2 or more TADs in the control, while 50% of CRHs span 2 or more TADs in CTCF KO.

Supplementary figure 1

A) Pearson analysis showing the correlation between differentially expressed genes and differential enhancer peaks for ABC-linked enhancers during cardiac disease when taking into account only genes and enhancers with a fold change of $\log_2$ 0.5.

B) A Box plot showing the number of connections per gene promoters and per enhancers in control and *Ctcf*-KO cells

REFERENCES

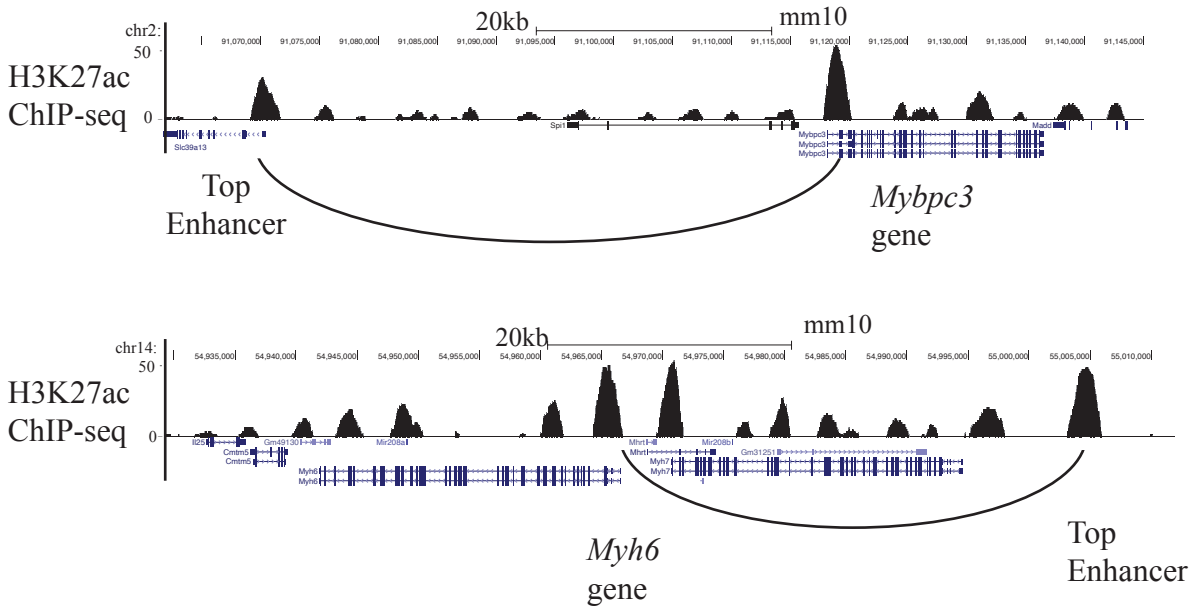
1. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195. doi: 10.1126/science.1222794
2. Boix CA, James BT, Park YP, Meuleman W, Kellis M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*. 2021;590:300-307. doi: 10.1038/s41586-020-03145-z
3. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51:1664-1669. doi: 10.1038/s41588-019-0538-0
4. Yao L, Liang J, Ozer A, Leung AK, Lis JT, Yu H. A comparison of experimental assays and analytical methods for genome-wide identification of active enhancers. *Nat Biotechnol*. 2022. doi: 10.1038/s41587-022-01211-7
5. Anene-Nzelu CG, Lee MCJ, Tan WLW, Dashi A, Foo RSY. Genomic enhancers in cardiac development and disease. *Nat Rev Cardiol*. 2021. doi: 10.1038/s41569-021-00597-2
6. Gasperini M, Tome JM, Shendure J. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat Rev Genet*. 2020;21:292-310. doi: 10.1038/s41576-019-0209-0
7. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet*. 2016;48:488-496. doi: 10.1038/ng.3539
8. Yao L, Shen H, Laird PW, Farnham PJ, Berman BP. Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol*. 2015;16:105. doi: 10.1186/s13059-015-0668-3
9. Alexanian M, Przytycki PF, Micheletti R, Padmanabhan A, Ye L, Travers JG, Gonzalez-Teran B, Silva AC, Duan Q, Ranade SS, et al. A transcriptional switch governs fibroblast activation in heart disease. *Nature*. 2021;595:438-443. doi: 10.1038/s41586-021-03674-1

10. Frangoul H, Altshuler D, Cappellini MD, Chen YS, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia. *N Engl J Med*. 2021;384:252-260. doi: 10.1056/NEJMoa2031054
11. Matharu N, Rattanasopha S, Tamura S, Maliskova L, Wang Y, Bernard A, Hardin A, Eckalbar WL, Vaisse C, Ahituv N. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science*. 2019;363. doi: 10.1126/science.aau0629
12. Di Giammartino DC, Kloetgen A, Polyzos A, Liu Y, Kim D, Murphy D, Abuhashem A, Cavaliere P, Aronson B, Shah V, et al. KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat Cell Biol*. 2019;21:1179-1190. doi: 10.1038/s41556-019-0390-6
13. Espinola SM, Gotz M, Bellec M, Messina O, Fiche JB, Houbron C, Dejean M, Reim I, Cardozo Gizzi AM, Lagha M, et al. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nat Genet*. 2021;53:477-486. doi: 10.1038/s41588-021-00816-z
14. Mangnier L, Joly-Beauparlant C, Droit A, Bilodeau S, Bureau A. Cis-regulatory hubs: a new 3D model of complex disease genetics with an application to schizophrenia. *Life Sci Alliance*. 2022;5. doi: 10.26508/lsa.202101156
15. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*. 2017;171:1573-1588 e1528. doi: 10.1016/j.cell.2017.11.008
16. Di Giammartino DC, Polyzos A, Apostolou E. Transcription factors: building hubs in the 3D space. *Cell Cycle*. 2020;19:2395-2410. doi: 10.1080/15384101.2020.1805238
17. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665-1680. doi: 10.1016/j.cell.2014.11.021
18. Li T, Jia L, Cao Y, Chen Q, Li C. OCEAN-C: mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol*. 2018;19:54. doi: 10.1186/s13059-018-1430-4
19. Madsen JGS, Madsen MS, Rauch A, Traynor S, Van Hauwaert EL, Haakonsson AK, Javierre BM, Hyldahl M, Fraser P, Mandrup S. Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. *Nat Genet*. 2020;52:1227-1238. doi: 10.1038/s41588-020-0709-z
20. Oudelaar AM, Harrold CL, Hanssen LLP, Telenius JM, Higgs DR, Hughes JR. A revised model for promoter competition based on multi-way chromatin interactions at the alpha-globin locus. *Nat Commun*. 2019;10:5412. doi: 10.1038/s41467-019-13404-x
21. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell*. 2018;71:858-871 e858. doi: 10.1016/j.molcel.2018.06.044
22. Anene-Nzelu CG, Tan WLW, Lee CJM, Wenhao Z, Perrin A, Dashi A, Tiang Z, Autio MI, Lim B, Wong E, et al. Assigning Distal Genomic Enhancers to Cardiac Disease-Causing Genes. *Circulation*. 2020;142:910-912. doi: 10.1161/CIRCULATIONAHA.120.046040
23. Tsai A, Alves MR, Crocker J. Multi-enhancer transcriptional hubs confer phenotypic robustness. *Elife*. 2019;8. doi: 10.7554/eLife.45325

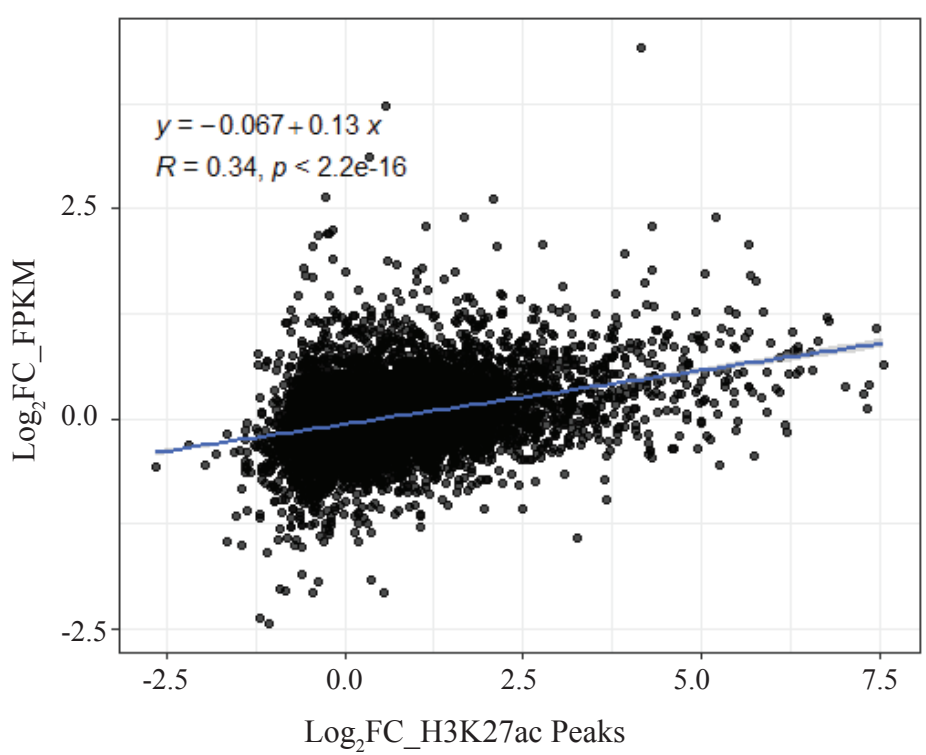
24. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*. 2011;12:283-293. doi: 10.1038/nrg2957
25. Lee DP, Tan WLW, Anene-Nzelu CG, Lee CJM, Li PY, Luu TDA, Chan CX, Tiang Z, Ng SL, Huang X, et al. Robust CTCF-Based Chromatin Architecture Underpins Epigenetic Changes in the Heart Failure Stress-Gene Response. *Circulation*. 2019;139:1937-1956. doi: 10.1161/CIRCULATIONAHA.118.036726
26. Rosa-Garrido M, Chapski DJ, Schmitt AD, Kimball TH, Karbassi E, Monte E, Balderas E, Pellegrini M, Shih TT, Soehalim E, et al. High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure. *Circulation*. 2017;136:1613-1625. doi: 10.1161/CIRCULATIONAHA.117.029430
27. Man JCK, van Duijvenboden K, Krijger PHL, Hooijkaas IB, van der Made I, de Gier-de Vries C, Wakker V, Creemers EE, de Laat W, Boukens BJ, et al. Genetic Dissection of a Super Enhancer Controlling the Nppa-Nppb Cluster in the Heart. *Circ Res*. 2021;128:115-129. doi: 10.1161/CIRCRESAHA.120.317045
28. Sergeeva IA, Hooijkaas IB, Ruijter JM, van der Made I, de Groot NE, van de Werken HJ, Creemers EE, Christoffels VM. Identification of a regulatory domain controlling the Nppa-Nppb gene cluster during heart development and stress. *Development*. 2016;143:2135-2146. doi: 10.1242/dev.132019
29. Dickel DE, Barozzi I, Zhu Y, Fukuda-Yuzawa Y, Osterwalder M, Mannion BJ, May D, Spurrell CH, Plajzer-Frick I, Pickle CS, et al. Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nat Commun*. 2016;7:12923. doi: 10.1038/ncomms12923
30. Hang C, Song Y, Li Y, Zhang S, Chang Y, Bai R, Saleem A, Jiang M, Lu W, Lan F, et al. Knockout of MYOM1 in human cardiomyocytes leads to myocardial atrophy via impairing calcium homeostasis. *J Cell Mol Med*. 2021;25:1661-1676. doi: 10.1111/jcmm.16268
31. Osterwalder M, Barozzi I, Tissieres V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*. 2018;554:239-243. doi: 10.1038/nature25461
32. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. 2017;169:930-944 e922. doi: 10.1016/j.cell.2017.05.004
33. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. Cohesin Loss Eliminates All Loop Domains. *Cell*. 2017;171:305-320 e324. doi: 10.1016/j.cell.2017.09.026
34. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*. 2000;405:486-489. doi: 10.1038/35013106
35. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012;22:1680-1688. doi: 10.1101/gr.136101.111
36. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016;529:110-114. doi: 10.1038/nature16490

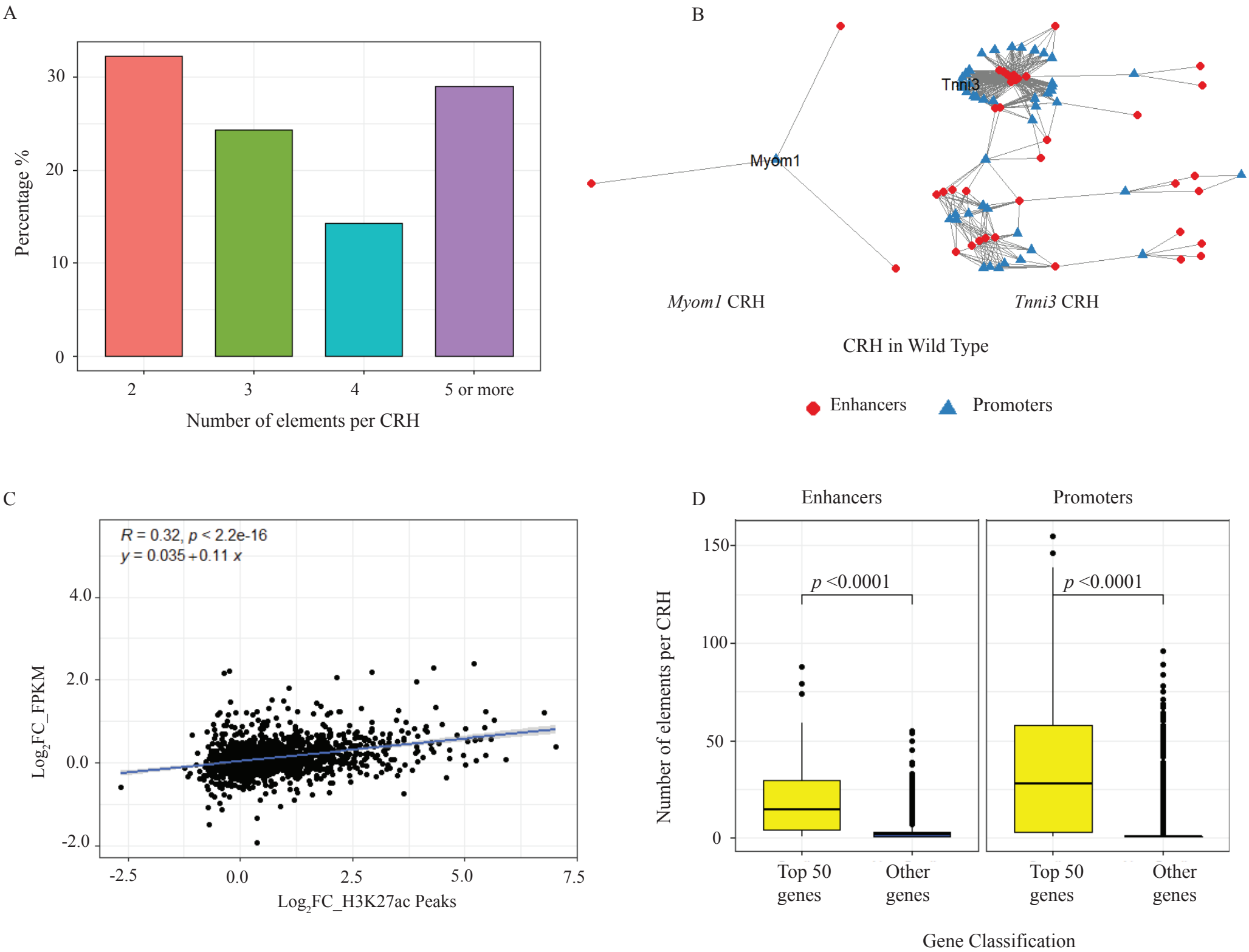
37. Sept CE, Tak YE, Cerda-Smith CG, Hutchinson HM, Goel V, Blanchette M, Bhakta MS, Hansen AS, Joung JK, Johnstone S, et al. High-resolution CTCF footprinting reveals impact of chromatin state on cohesin extrusion dynamics. *bioRxiv*. 2023. doi: 10.1101/2023.10.20.563340
38. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589-595. doi: 10.1093/bioinformatics/btp698
39. Bhattacharyya S, Chandra V, Vijayanand P, Ay F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat Commun*. 2019;10:4221. doi: 10.1038/s41467-019-11950-y
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139-140. doi: 10.1093/bioinformatics/btp616
41. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14:959-962. doi: 10.1038/nmeth.4396
42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079. doi: 10.1093/bioinformatics/btp352
43. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137. doi: 10.1186/gb-2008-9-9-r137
44. Csardi G NT. The igraph software package for complex network research. *Int J Complex Syst*. 2006.

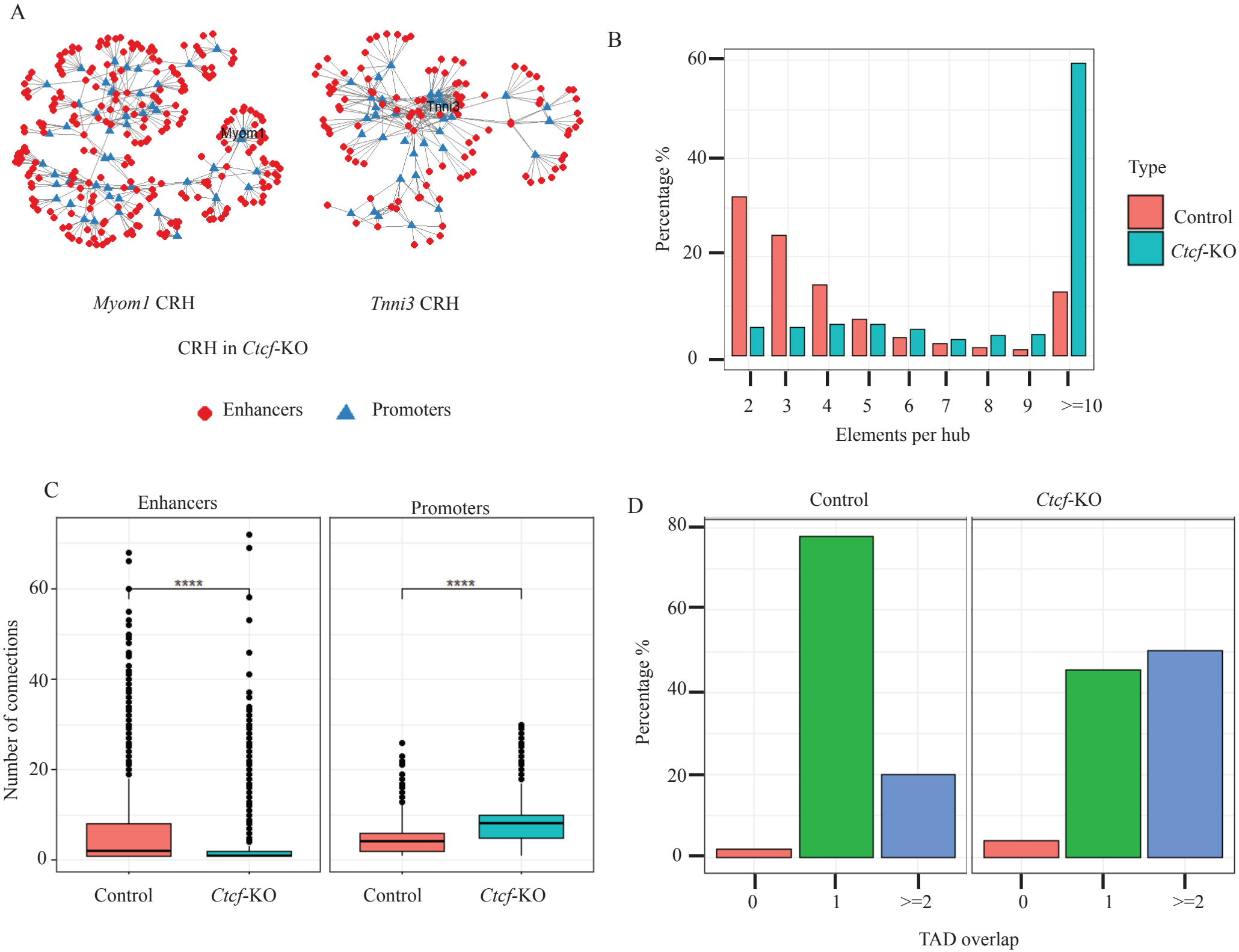
A


chr2:
20kb
mm10
H3K27ac
ChIP-seq
50
0
Top
Enhancer
Mybpc3
gene
chr14:
20kb
mm10
H3K27ac
ChIP-seq
50
0
Myh6
gene
Top
Enhancer

B


y = −0.067 + 0.13 x
R = 0.34, p < 2.2e-16
2.5
0.0
-2.5
Log2FC_FPKM
-2.5
0.0
2.5
5.0
7.5
Log2FC_H3K27ac Peaks


A
Percentage %
30
20
10
0
2
3
4
5 or more
Number of elements per CRH
B
Myom1
Tnni3
Myom1 CRH
Tnni3 CRH
CRH in Wild Type
Enhancers
Promoters
C
R = 0.32, p < 2.2e-16
y = 0.035 + 0.11 x
Log2FC_FPKM
4.0
2.0
0.0
-2.0
-2.5
0.0
2.5
5.0
7.5
Log2FC_H3K27ac Peaks
D
Enhancers
Promoters
Number of elements per CRH
150
100
50
0
p <0.0001
p <0.0001
Top 50 genes
Other genes
Top 50 genes
Other genes
Gene Classification


A
Myom1
Tnni3
Myom1 CRH
Tnni3 CRH
CRH in Ctcf-KO
Enhancers
Promoters
B
Percentage %
60
40
20
0
2
3
4
5
6
7
8
9
>=10
Elements per hub
Type
Control
Ctcf-KO
C
Enhancers
Promoters
Number of connections
60
40
20
0
****
****
Control
Ctcf-KO
Control
Ctcf-KO
D
Control
Ctcf-KO
Percentage %
80
60
40
20
0
0
1
>=2
0
1
>=2
TAD overlap