CTCF mediates the Activity-by-contact derived cis-regulatory hubs

Mick Lee[1,2*], Loïc Mangnier[3,4,5,6*], Cory C. Padilla[7], Dominic Paul Lee[1,2], Wilson Tan[1,2], Peter Yiqing Li [1,2], Wen Hao Zheng[1,2], Louis Gan[1,2], Ching Kit Chen[1,2,8,9], Yee Phong Lim[1,2,8,] Rina Miao Qin Wang[1,2,8], 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 mdccgoa@nus.edu.sg, roger.foo@nus.edu.sg

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 modeled 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 topological 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 enhancer-promoter interactions resulting in a re-distribution of ABC scores of the putative enhancers in C57/bl6 mouse hearts. The loss of CTCF also leads to a global reduction of the total number of CRHs, and an increase in the size of the CRHs due to increase in the number of elements within each hub. In addition, CTCF loss led to more 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.

71

72 **INTRODUCTION**

73

74 The regulation of gene expression by the non-coding genome is an area of rapidly evolving
75 research globally as it is implicated in different biological phenomena including organ
76 development, cell-fate determination, tissue response to external stimuli and overall cellular
77 homeostasis [1]. Furthermore, given the localization of the majority of disease-relevant
78 genomic variations in non-coding regulatory regions [1], their involvement in disease
79 pathogenesis is well documented. Accordingly, tremendous effort is deployed to annotate
80 these regulatory elements in different species, tissues and cells [2-4]. To this end, different
81 experimental protocols and computational algorithms have been applied to identify genome-
82 wide regulatory elements responsible for controlling gene expression [5]. These include
83 techniques such as ChIP-seq, ATAC-seq, DNase-seq, Hi-C, and functional genomics using
84 CRISPR activation or inhibition studies etc., each with their strengths and limitations[5,6]. In
85 addition, computational methods such the Activity by Contact (ABC) algorithm[3], TargetFinder,
86 ELMER and others have been developed to integrate these datasets and identify gene-
87 enhancer pairs. Similarly, large consortiums have been formed to annotate these regulatory
88 elements in different tissues [7,8]. These studies provide a readily accessible resource and
89 reference catalogue of the non-coding regulatory elements to the scientific community and
90 help scientists unravel the role of regulatory regions in disease development and evaluate
91 potential therapeutic strategies [9-11].

92

93 Besides the direct identification of enhancer-gene (E-G) pairs, other studies have
94 demonstrated that enhancers and promoters often interact dynamically within a network
95 often called cis-regulatory hubs (CRH), allowing for multiple regulatory connections [12,13].
96 These CRHs can be defined as 3D regulatory networks that constitute a complex organization
97 of multi-loci connections of enhancers and promoters, connected in 3D space within the
98 nucleus. They highlight direct and indirect contact between genes and distal regulatory
99 elements and can help predict target genes for disease-relevant non-coding SNPs[13]. They are
100 often bound by a cluster of transcription factors and have been implicated in gene co-
101 regulation, lineage specification and disease development [12-14]. These hubs differ from other
102 known 3D structures such as genomic compartments which constitute euchromatin or

103 heterochromatin regions. They also differ from TADs defined as units within which most
104 enhancer-promoter interactions occur and loops which are pair-wise connections [15].
105 Identification of these CRHs is often done through detailed analysis of 3D contacts based on
106 Hi-C data or using chromatin accessibility (ATAC or DNAse) to study co-accessible regions and
107 thus construct these hubs [12,16-19]. A recent study proposed building these CRHs as a bipartite
108 network using the Activity-by contact (ABC) algorithm [13]. The ABC algorithm combines ATAC-
109 seq, ChiP-seq and Hi-C to rank the gene enhancers based on their regulatory impact and was
110 shown to outperform other models such as HiChIP or TargetFinder in identifying regulatory
111 enhancers [3,20]. Thus in addition to identifying and ranking enhancers, the ABC algorithm is
112 used to generate a more predictive and physiologically relevant CRH[13], and these hubs were
113 strongly enriched for disease-relevant genes and also helped explain disease heritability [13,21].
114
115 Given that the 3D genome architecture is a key component of the ABC algorithm, which is
116 used to construct the CRHs, [13], it is of huge importance to understand the role of architectural
117 proteins like CTCF. The CTCF protein often called the "genome weaver" plays a crucial
118 function in the "folding" of the genome and bringing of enhancers into proximity to their
119 distant target genes [15,22]. However, their role in regulating CRHs have not been explored. In
120 this study, using mouse cardiomyocytes as a model of terminally differentiated non-dividing
121 cells, we employ the ABC model to identify top cardiac gene-enhancers and then to study
122 CRHs. We analyse the characteristics of CRHs containing tissue-specific genes, showing that
123 these genes are enriched in multi-enhancer CRHs. We also show that in disease, there is a
124 positive regulation of genes and enhancers within a CRH. Finally we show that loss of CTCF
125 leads to a merging of CRHs and formation of new CRHs in the *Ctcf*-KO cells with resultant
126 drastic change in the number and distribution of enhancers that cross the ABC threshold.
127
128 **RESULTS**
129
130 *Cardiac Enhancer Landscape Identified Through Activity-By-Contact Algorithm*
131
132 We performed ATAC-seq, H3K27ac ChiP-seq and HiChIP , to generate the ABC scores for the
133 enhancer-gene pairs in the control adult mouse cardiomyocytes. Using an ABC cut-off of 0.01
134 [13], we identified 7000 out of 156,988 putative distal regulatory regions marked by H3K27ac

135 peaks and/or ATAC peak in the adult mouse CM that crossed the threshold. This represents
136 about 5% of all putative regulatory regions predicted to have regulatory effect on their target
137 genes and is similar to what was observed in another study [3]. With regards to the enhancer-
138 gene (E-G) pairs, globally there were 34,496 E-G interactions with an average of 4.19
139 connections per gene, and 5.29 connections per enhancer in these control adult mouse CMs.
140 Figure 1A shows examples of top-linked enhancers and their target genes for 2 cardiac
141 disease-relevant genes *Mybpc3* and *Myh6*. The comprehensive list of ABC linked enhancers
142 and target genes can be found in Supplementary Table 1. A literature search confirmed that
143 at least 3 of the identified enhancers in our study have been validated. These include the
144 *Nppa/Nppb* super enhancer located upstream of both genes and which has been shown to
145 play a critical role in stress gene response of *Nppa/Nppb* during pressure-overload induced
146 cardiac stress [23-25]. A separate enhancer for *Myh7* identified from the ABC scoring has also
147 been validated previously in the mouse heart [26]. Deletion of this enhancer region led to
148 downregulation of *Myh7* in mouse hearts[26]. Taken together, these previous studies provide
149 support to the validity of our prediction model.
150 To provide further evidence that these ABC-linked enhancers regulate their putative target
151 genes, we studied correlation between changes in gene expression versus changes in peak
152 height of the ABC-linked enhancers during pressure overload-induced cardiac hypertrophy.
153 Using RNA-seq and H3K27ac ChIP-seq datasets from mouse hearts subjected to transverse
154 aortic constriction [23], we found that changes in gene expression correlates modestly with
155 changes in H3K27ac peak signals (Pearson correlation= 0.34, $P < 2.2e-16$) (Fig 1B). The
156 correlation was stronger when focussing only on genes and enhancers with greater or less
157 than $\log_2$ 0.5 fold change after Transverse aortic constriction (Suppl Fig 1A) (Pearson
158 correlation = 0.51, $P < 2.2e-16$). This result shows that globally, upon external stimuli, gene
159 expression is regulated in the same direction as their ABC linked enhancers.
160
161 *Loss Of CTCF Leads To Changes In The Enhancer Interactome And Changes In The ABC Scores*
162 *Of Putative Enhancers*
163
164 Next, to elucidate the role of CTCF in ABC scores and regulation of the CRHs, we also applied
165 the ABC algorithm to the *Ctcf*-Knock out (*Ctcf*-KO) mouse cardiomyocytes. Given that previous
166 publications have shown that the overall number of regulatory regions identified through

167 ChIP-seq or ATAC-seq after CTCF deletion remain mostly unchanged [27,28], with changes only
168 observed in the peak heights, we focussed our attention on the HiChIP differences. We have
169 previously published a genome-wide Hi-C in mouse cardiomyocytes [23]. A deeper analysis of
170 the H3K27ac interactions through the first Micro-C/MNase mediated H3K27ac HiChIP in
171 mouse CMs unravelled the enhancer interactome in the mouse heart. HiChIP presents
172 advantages over Hi-C by providing better resolution with equal or lower sequencing depth [29]
173 while the use of MNase in Micro-C provides better resolution for enhancer-promoter
174 interactions compared to restriction-enzyme digest methods [30]. The H3K27ac HiChIP
175 experiment was performed on two independent biological replicates each, and the data were
176 merged to give a total of 160 million unique pairs for the *Ctcf*-KO and 126 million unique pairs
177 for the control samples sufficient for HiChIP analysis[29]. For comparative analysis, the *Ctcf*-KO
178 data was down sampled to 126M to match the control data. Using a window size of 10kb and
179 an FDR of <0.1, we performed FitHiChIP [31] to identify the significant H3K27ac-anchored loops
180 in our dataset. The results revealed 12,843 H3K27ac loops unique to control cells, 7,070
181 H3K27ac loops unique to *Ctcf*-KO cells and 483 shared H3K27ac loops between both
182 conditions (Figure 2A, Supplementary Table 2). The KO-unique H3K27ac loops spanned longer
183 distances than control-unique and shared H3K27ac loops (Figure 2B), consistent with previous
184 studies using H3K4me3 HiChIP in zebrafish [27]. We further analysed these H3K27ac interaction
185 anchors for the presence of CTCF using the Control CTCF ChIP-seq as reference. We
186 discovered that 14% of these H3K27ac loop anchors occurred in regions with CTCF binding at
187 both anchors in the control-unique group, while only 4.2% of the H3K27ac loops in the *Ctcf*-
188 KO occurred in anchors with native CTCF binding. Concordantly, 56% of the H3K27ac loops
189 had no native CTCF binding at the anchors in control, with the number increasing to 65% in
190 the KO samples. The percentage of H3K27ac loops with native CTCF binding at one anchor
191 was consistent at about 30% for both control and *Ctcf*-KO (Fig 2C). This shows that the loss of
192 CTCF led to a 70% decrease in the H3K27ac loops that co-localize with CTCF at both anchors
193 in the *Ctc*f-KO cells, with an increase in ectopic interactions between H3K27ac anchors
194 without CTCF. Next we performed the aggregate peak analysis (APA) for the control and *Ctcf*-
195 KO samples, the APA reveals enrichment for loop domains often bound by CTCF and Cohesin
196 at the anchors [15]. Consistent with our previously published Hi-C data, and other studies that
197 have performed aggregate analysis in different *Ctcf*-KO models [32], there was a marked
198 decrease in the APA score from 3.33 in the control to 0.86 in the *Ctcf*-KO for all H3K27ac loops,

199 suggesting indeed the loss or weakening of loop domains in the *Ctcf*-KO samples [23].The APA
200 score was also higher for the control unique and shared loops in the control samples, while
201 the APA score for *Ctcf*-KO unique group was higher in the *Ctcf*-KO samples as anticipated,
202 possible representing some re-arrangement of Cohesin in the *Ctcf*-KO group (Figure 2D)
203
204 We applied the ABC algorithm on the *Ctcf*-KO sample and observed that there were 36,226
205 regulatory elements that crossed the 0.01 threshold, resulting in 71,080 E-G interactions with
206 similar number of expressed genes (~8,000) (Supplementary Table 1) in control and *Ctcf*-KO
207 samples. This led to an increase in the average number of connections per gene to 8.14 in
208 *Ctcf*-KO samples. Of the 71,080 enhancer-gene pairs only 20% (14,200 E-G pairs) were shared
209 between control and Ctcf-KO. This increase in ABC scores could be due to increased
210 interaction frequency and/or increased H3K27ac ChIP-seq upon CTCF loss. Interestingly we
211 observed that the highest ranked ABC enhancer in the control had a score of 0.99, while the
212 highest ranked enhancer in the *Ctcf*-KO sample had a score of 0.78 (scale of 0 – 1). Indeed the
213 average score in the control was 0.054 while the average score in the *Ctcf*-KO was 0.028 (TTest
214 P-value < 2.2e-16) . This suggests that the loss of CTCF led to gain of ectopic enhancers in the
215 *Ctcf*-KO with a redistribution of the ABC scores of the enhancers.
216 Next we integrated the ABC results to generate the CRHs focussing first on the control
217 samples. We identified 1,522 hubs in these control cardiomyocytes with about 70% of the
218 hubs containing fewer than 5 elements (Fig 3A). Figure 3B shows 2 examples of such hubs
219 containing cardiac genes: *Myom1* a gene involved in formation of myofibrils [33], and found in
220 a simple hub while *Tnni3* a cardiac-specific sarcomeric gene is found in a more complex hub.
221 Consistent with the correlation between gene expression and ABC-linked enhancers in heart
222 disease, we also verified that genes and enhancers within the same hubs are positively
223 correlated in same direction of change during cardiac stress, showing Pearson correlation of
224 0.32 (P <2.2e-16) (Figure 3C). This suggests a co-regulation of genes and enhancers found
225 within the same hub upon external stimuli. We then ranked cardiac genes by FPKM and
226 selected the top 50 genes, mostly cardiac-specific genes, to glean biological insights about the
227 characteristics of CRHs that harbour these highly expressed genes. Our data revealed that
228 highly expressed cardiac genes were more likely to be in hubs with more regulatory elements
229 than non-highly expressed genes (Figure 3D). These cardiac-specific genes were also more
230 likely to be found in hubs with higher number of promoters. Such multi-enhancer hubs has

231 been proposed as a mechanism to buffer against environmental stresses and genetic
232 perturbations and thus provide phenotypic robustness to key genes [21,34]. On the other hand,
233 hubs with high number of promoters are characteristic of lineage specific genes [13,21].
234
235 *Loss Of CTCF Alters The Cis-Regulatory Hubs*
236
237 Finally, we analysed the CRHs in the *Ctcf*-KO CMs and found a marked reduction in total
238 number of CRHs from 1522 in control to 660 CRHs in the KO, accompanied with an increase
239 in the number of elements per hub. Fig 4A shows example of the same hubs containing 2
240 cardiac genes shown in Fig 3A. Describe observation., Figure 4B shows the distribution of
241 elements per hub, confirming a striking increase in the percentage of hubs with 10 or more
242 elements. Associated with this increase in percentage of complex hubs, there was an increase
243 in the average number of connections per gene in the *Ctcf*-KO cells, and a decrease in number
244 of connections per enhancer (Fig 4C, Supplementary Fig 1B). We asked if this reduction in
245 total number of CRHs was merely due to merging of CRHs or formation of new CRHs, and
246 observed both cases. The merging of CRHs was observed primarily in the simple hubs (2-3
247 elements per hub) located within close proximity as 403 (82%) of 492 such simple hubs
248 merged with 1 or more other hubs to form larger hubs. The proportion of the number of CRHs
249 merging to form a new one varied, ranging from 2 merged CRHs to 15 merged CRHs
250 Supplementary Fig 1C. For the larger hubs, there was mostly formation of new hubs, as genes
251 gained new interaction and lost other interactions. To further confirm the importance of CTCF
252 in the organization of these hubs, we compared CTCF binding sites in the control hubs vs *Ctcf*-
253 KO hubs based on control CTCF ChIP-seq, and observed an enrichment of CTCF in the control
254 hubs (fisher exact test odds ratio of 1.69, two sided $P < 2e-16$). Next, we analyzed the
255 relationship between CRHs and TADs since TAD boundaries are enriched for CTCF [23], and
256 earlier studies have shown that CRHs are generally constrained within TAD boundaries [13].
257 Indeed, using the TADs in the control as reference, our data showed that about 80% of CRHs
258 in the control were contained within the same TAD while only 20% spanned more than 1 TAD.
259 In contrast, about 45% of CRHs in KO were contained within a TAD while 50% of CRHs in the
260 KO spanned more than 1 TAD (Fig 4D) while 5% were not within TAD boundaries, showing
261 that the loss of CTCF leads to a reorganization of the CRHs with the formation of cross-TAD
262 enhancer-promoter interactions.

263

264 **DISCUSSION**

265

266 Non-coding regulatory elements play crucial roles in development and disease and hold
267 promise for next-generation therapeutic targets. While different studies by us and others
268 have annotated these elements and CRHs in different cellular models, the role of CTCF in the
269 ABC derived CRHs was yet to be studied. First, our study applied the ABC algorithm to the
270 mouse heart, generating to the best of our knowledge, the first such ranked enhancer scores
271 in mice hearts. Our findings confirm the positive correlation between genes and their ABC-
272 linked enhancers in cardiac disease, and affirms that genes and enhancers within the same
273 hub are co-regulated during pressure overload-induced cardiac hypertrophy. Integrating ABC
274 to identify CRHs may thus represents another method to analyze for effect of regulatory SNPs
275 on target genes, as the SNP may have effect on a target gene when they are in the same hub,
276 even when there is no direct link through pair-wise enhancer-promoter analysis [13]. Analysis
277 of the CRHs also suggests that tissue-specific genes are more likely to be contained in hubs
278 with high connectivity and rich in distal elements. This increased number of E-P networks
279 provides redundancy and phenotypic robustness while guaranteeing increased
280 transcriptional activity thanks to the multi-enhancer networks.

281

282 Loss of CTCF affects the enhancer interactome, particularly impacting the H3K27ac
283 interactions co-bound by CTCF. This leads to E-Ps that span larger distances and gain of new
284 interactions. The loss of CTCF also leads to a re-organization of ABC-selected enhancers with
285 a direct impact on the CRHs, resulting in fewer hubs but with more elements per hub. This
286 implies a crucial role for CTCF in not only determining which enhancers cross the ABC
287 threshold but also in the organization of the ABC-derived CRH. While this study focussed on
288 the effect of loss of CTCF on overall CRH structure and its relationship with TAD boundaries,
289 future studies will examine individual CRHs and their relationship to CTCF binding to ascertain
290 how loss of CTCF can lead to local dysregulation of CRHs. Indeed, this finding has implications
291 for mutations that affect local CTCF binding at regions that are not TAD boundaries, as it may
292 lead to merging of 2 or more CRHs and thus the establishment of indirect connection between
293 genes and enhancers within the newly formed hubs.

294

295 Our findings therefore demonstrate that CTCF plays a critical role in determining the
296 membership of the CRHs as well as their organization, and thus plays a role in disease and cell
297 type-specific gene expression through organization of these cis-regulatory hubs.
298
299 **ABBREVIATIONS**
300 ChIP-seq – Chromatin immunoprecipitation with sequencing
301 ATAC-seq – Assay for Transposase Accessible Chromatin with sequencing
302 Hi-C – High throughput chromatin conformation capture with sequencing
303 ABC – Activity by contact algorithm
304
305 **DECLARATIONS**
306 **ETHICS APPROVAL:** Animal protocol approved by the Institutional Animal Care and Use
307 Committee (National University of Singapore)
308 **CONSENT FOR PUBLICATION** – The authors consent to publication of this article
309 **AVAILABILITY OF DATA AND MATERIALS**
310 The raw data supporting the conclusion of this article will be made available by the authors,
311 without undue reservation.
312 **COMPETING INTEREST**
313 Cory C. Padilla works for Cantata Bio, developers of the HiChIP kit used in the study. The other
314 authors declare that the research was conducted in the absence of any commercial or
315 financial relationships that could be construed as a potential conflict of interest
316 **FUNDING**
317 This work was supported by grants from the National Medical Research Council (to Dr Foo),
318 Start-up grants from the Montreal Heart Institute (Dr. Anene-Nzelu). The funding was used
319 to purchase the reagents. Some data analyses were performed on computing resources
320 from Compute Canada.
321 **AUTHOR CONTRIBUTION**
322 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,
323 S.B, A.B, C.G.A.N provided data analysis. S.B, A.B, R.F, C.G.A.N Provided supervision for
324 experiments and data analysis. All authors read and approved the manuscript
325 **ACKNOWLEDGEMENTS**
326 We thank Cantata Bio for providing the HiChIP kits

327 **METHODS**

328

329 *Animal Experiments*

330 Animal experiments were performed under a license approved by the Institutional Animal
331 Care and Use Committee (National University of Singapore). *Ctcf*$^{flox/flox}$ mice harboring LoxP
332 sites flanking exons 3 and 12 of the *Ctcf* gene against a C57/bl6 strain background were used
333 as previously published [23]. Details of the source of the original strain has been documented
334 in our earlier publication [23]. AAV9-cTnt-eGFP and AAV9-cTnt-Cre-tdTomato vectors encoding
335 codon improved Cre recombinase or eGFP under the control of cardiac troponin T promoter
336 were used for control and *Ctcf* knock-out respectively [23]. Experiments were performed on
337 adult 6- to 8-week-old mice. 3 mice were used for each group, housed in 2 different cages.
338 The numbers were based on our previously published study [23]. Cardiomyocytes were
339 isolated from mouse hearts after 2 weeks of AAV9 injections for control and *Ctcf*-KO mice.
340 CTCF knock-out was assessed by western blot, and upon confirmation, cells were used for
341 subsequent experiments. Adeno-associated virus 9 (AAV9) virus for the delivery of Cre
342 recombinase was generated as previously described [23]. For cardiomyocyte isolation,
343 isoflurane anaesthesia is used before performing a thoracotomy to harvest the hearts as
344 previously described [23].

345

346 *H3k27ac HiChip Library Prep And Data Analysis.*

347 Mouse adult cardiomyocytes were isolated following previously published protocol [23].
348 MNase-based HiChIP assay was performed on $2x10^6$ isolated CMs. Frozen cells were
349 resuspended in 1× PBS and crosslinked with 3 mM DSG and 1% formaldehyde. Washed cells
350 were digested with 0.75 µl MNase in 100 µl of nuclease digest buffer with $MgCl_2$. Cells were
351 lysed with 1× RIPA, and clarified lysate was used for ChIP. H3K27ac antibody (Abcam 04729)
352 was used to perform chromatin immunoprecipitation. The Protein A/G bead pulldown,
353 proximity ligation, and libraries were prepared as described in the Dovetail protocol (Dovetail
354 HiChIP MNase Kit). Libraries were sequenced on an Illumina Novaseq.

355

356 HiChIP paired-end reads were aligned with BWA MEM (version 0.7.17r1198-dirty) with the -
357 5SP flag enabled with an index containing only the canonical chromosomes of the mm10
358 genome (available from the UCSC genome). The resulting alignments were then parsed with

359 pairtools (versions 0.3.0) with the following options --min-mapq 40 –walks-policy 5unque –
360 max-inter-align gap 30 and the thre –chroms-path file corresponding to the size of the
361 chromosomes used for the alignment index. Parse pairs were deduplicated, sorted with
362 pairtools. Valid pairs were identified and through pairtools select (pair type = 'UU', 'UR', 'RU',
363 'uu', 'uU' and 'Uu') and downsampled to the lowest number of valid pairs in a sample (126
364 million in LACZ) with pairtools sample. Contact matrices in .hic format were generated with
365 juicetools pre function (version 1.22.01).
366
367 FitHiChIP (version 9.1) was used to identify significant interactions ("loops") from the valid
368 subsampled pairs at 10kb resolution with the following settings, loop type = all-to-all,
369 coverage = bias correction, merge redundant loops = Yes, background model = FithHiChIP(L),
370 FDR < 0.1, minimum interaction size =20kb, maximum interaction size = 2 mb. Conditionally
371 unique (meaning only occurring in LACZ or CRE) and shared loops were identified with
372 bedtools pairToPair (version 2.28.0) requiring both loop anchors to overlap the same
373 coordinates be flagged as shared. Loop size was assessed by performing a two-sided
374 Wilcoxson's rank-sum test on the distribution of loop ranges (distance between the two
375 anchors). Loop anchors were then intersected the LACZ CTCF binding sites to determine the
376 proportion of loops resulting from CTCF presence. P-values for CTCF mediated loops were
377 obtained through a two-sided Fischer's Exact test. Aggregate Peak Analysis (APA) was
378 computed with juicetools APA function. Loop anchors were used to as the sites to aggregate
379 over at 10kb resolution. APA enrichment scores, loop center from the lower-left (LL) corner is
380 shown as both an APA score (ratio of the mean of center to mean of LL) and a Z-score. APA
381 normed output was used to plot the APA matrices.
382
383 H3K27ac HiChIP data were summarized and visualized with R, the package eulurr was used
384 for the shared and unique loop Venn diagrams and ggplot2 for loop size and proportion of
385 loops with CTCF binding site at the anchor. Additionally, ggplot2 was used to plot APA
386 matrices with the geom_raster function, with the color scale set to the same min – max limits
387 across all APA plots. At sites of interest, HiChIP loops files (in longrange format) were used to
388 visualize loops along with ChIP-seq coverage and RNA-seq activity in the WashU Epigenome
389 Browser (https://epigenomegateway.wustl.edu/).
390

391 *Chip-Seq Experiment And Data Analysis*

392

393 ChIP experiments on isolated CM were performed as described previously [23]. Briefly, CMs
394 were cross-linked with 1% formaldehyde for 10 minutes at room temperature and quenched
395 with glycine (0.125 mol/L final concentration). Cells were then washed in ice-cold PBS and
396 pelleted and lysed in FA lysis buffer (10 mmol/L Tris-HCl, pH 8.0, 0.25% Triton X-100, 10
397 mmol/L EDTA, 0.1 mol/L NaCl). To facilitate cell lysis, the cell pellet was passed through a 27.5-
398 gauge needle gently 20 times. Nuclei were pelleted by centrifugation resuspended in
399 sonication buffer, and chromatin was fragmented via sonication to an average size of 200 to
400 300 bp (Bioruptor Plus, Diagenode). Chromatin was immunoprecipitated against H3K27ac
401 (Abcam ab4729) or CTCF (EMD Millipore, catalog No. 07–729) overnight. Antibody-chromatin
402 complexes were pulled down with Protein G Dynabeads (Invitrogen, catalog No. 10003D),
403 washed with 0.1% SDS lysis buffer, and eluted with elution buffer (1% SDS, 10 mmol/L EDTA,
404 50 mmol/L Tris-HCl, pH 8). After cross-link reversal (4 hours of incubation at 65°C) and
405 proteinase K treatment, immunoprecipitated DNA was extracted. ChIP DNA was quantified
406 by fluorometric quantification (Thermo Fisher Scientific, Qubit dsDNA HS assay kit, catalog
407 No. 32851). Library preparation was performed with the New England Biolabs Ultra II Kit
408 according to the manufacturer's specifications and sequenced on the Illumina NextSeq High
409 platform.

410

411 Paired-end reads were aligned to the mm10 genome using Bowtie2. The uniquely aligned
412 reads were PCR-duplicate removed. Peak-calling was performed using Dfilter, and
413 parameters were optimized for different antibodies. For H3K27ac, the parameters are: ks=60,
414 lpval=6. For CTCF, ks=10, lpval=6, -nonzero, -refine. Reads were counted in 100-bp windows
415 for each library and scaled to normalize for sequencing depth. Read counts in each window
416 were then adjusted by normalizing their GCcontent against the average GC-content of all
417 libraries. Peak height was obtained by summing the bin-wise normalized counts. Peak height
418 was quantile-normalized to reduce technical variation. Supplementary Table 3 lists H3K27ac
419 peaks in Sham and TAC treated mouse cardiomyocytes.

420 *ATAC-Seq Experiment And Analysis*

421

422 The ATAC-seq was performed according to previously published Omni-ATAC protocol [35]
423 Briefly, 50,000 adult cardiomyocytes were pelleted at 500 g for 5 minutes. The cells were
424 resuspended in 100 µl ATAC-resuspension buffer containing 0.5% NP40, 0.5 % Tween-20 and
425 0.01% Digitonin. The cells were left on ice for 5 minutes after which 1 ml of cold ATAC-RSB
426 containing 0.1% Tween 20 was added to wash out the lysis buffer. The nuclei were pelleted
427 at 500 g for 5 minutes, the supernatant was carefully removed and the nuclei were
428 resuspended in 50 µl transposition mixture containing 25 µl of 2x TD buffer, 16.5 µl of PBS, 5
429 µl nuclease free water, 0.5 µl of 1% digitonin, 0.5 µl of 10% Tween-20 and 2.5 µl of transposase
430 (Illumina Tagment DNA enzyme 1, Catalogue number 20034198). The reaction was incubated
431 at 37°C for 30 minutes in a thermoshaker with 1000 RPM. After the reaction, DNA was
432 extracted using the NEB Monarch® PCR & DNA cleanup kit (Catalogue number T1030L), PCR
433 was performed using the Illumina/Nextera primers after which Ampure XP beads were used
434 for library clean up. The resulting library was sequenced on Nextseq. Note that while ChIP-
435 seq and HiChIP were performed on control and CTCF-KO cells, the ATAC-seq experiments
436 were performed on a wild type C57/bl6 strain mice. ATACseq peaks were called on each
437 sample using MACS2 (callpeak function with these parameters: -nomodel and -B). BedGraph
438 files generated by MACS2 where normalized to read counts per million reads sequenced
439
440 *ABC-Score And CRH Analysis*
441
442 The ABC model defines active enhancers based on a quantitative score of DNAse, H3K27ac,
443 and normalized Hi-C contact number[3]. This score is computed relative to a background activity
444 over a 5-Mb window around a candidate element. Then, we set the threshold to 0.01; beyond
445 which a candidate element is considered as a distal element. As an extension of the ABC-
446 Score, CRHs were defined as bipartite networks between promoters and distal elements
447 (igraph R package) [36] TADs were called using the directionality index as previously described.
448 Pearson correlation analysis performed using R package
449
450 Figure 1:
451 A) USCS screenshot showing 2 examples of cardiac genes and their top ABC-linked
452 enhancers

453 B) Pearson analysis showing correlation between differential expressed genes and
454 differential enhancer peaks for ABC-linked enhancers during cardiac disease
455
456 Figure 2: H3K27ac HiChIP in Control and CTCF KO
457 A) Venn diagram showing number of FitHiChIP interactions in Control-unique and *Ctcf*-
458 KO unique and shared loops
459 B) Box plot showing the loop sizes in the various groups. The loops in CTCF-unique
460 interactions span longer distance than the interactions in Control unique CM
461 C) Bar chart showing distribution of CTCF at the FitHiChIP anchors. 14% of the loops have
462 CTCF at both anchors in the control while only 4.2% of loops in the CTCF KO have CTCF
463 on both anchors using the Control CTCF biding as reference
464 D) Aggregate Peak Analysis for the loops in Control and CTCF KO cells, showing a marked
465 reduction of the APA and Z-score in the CTCF KO samples for all loops
466
467 Figure 3: CRH in control Cardiomyocytes
468 A) Bar chart showing distribution of number of elements per CRH
469 B) Examples of 2 CRHs containing cardiac genes *Myom1* and *Tnni3* in the control
470 cardiomyocytes
471 C) Pearson correlation between differential expression of genes and enhancer peaks
472 contained within the same CRHs during cardiac stress
473 D) Box plot showing characteristics of CRHs containing cardiac specific genes. These CRHs
474 tend to have more enhancers and promoters
475
476 Figure 4: CRHs in the KO cells
477 A) Examples of 2 CRHs containing cardiac genes *Myom1* and *Tnni3* in the *Ctcf*-KO
478 cardiomyocytes. Compared to the same CRHs in control, there is an increase in the
479 number of elements in each hub
480 B) Bar chart showing distribution of elements per CRH in Control and *Ctcf*-KO. There is
481 an increase in the percentage of CRHs with 10 or more elements from 15% in the
482 Control to 60% in the CTCF KO cardiomyocytes

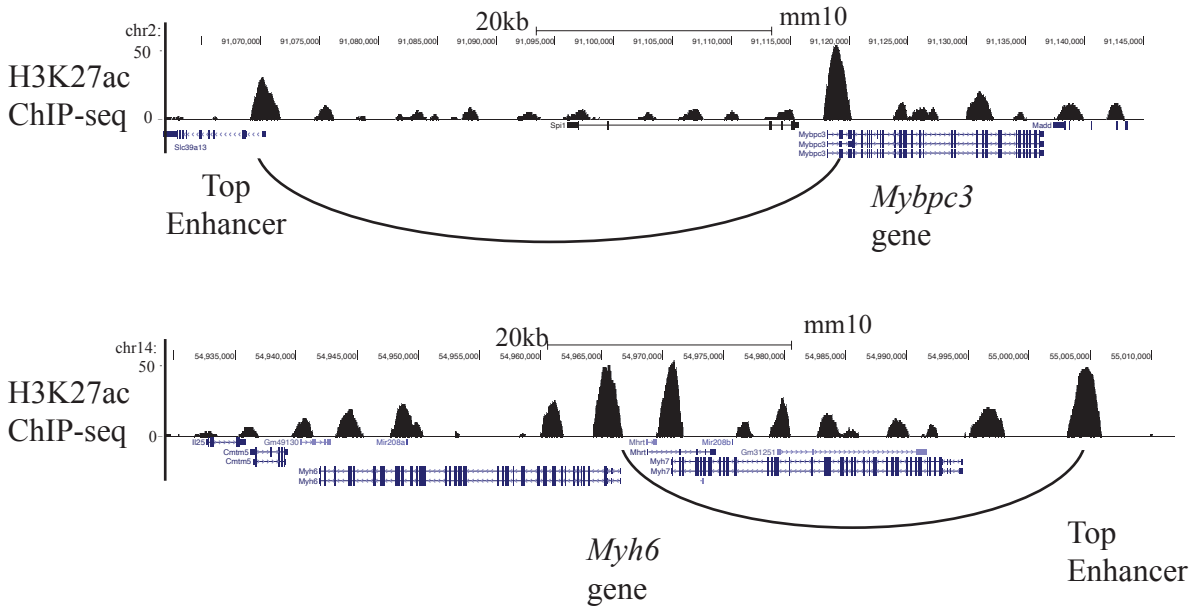
483 C) Bar chart showing number of connections per gene promoters and per enhancers.
484 There is an increase in the average number of connections per gene in the *Ctcf*-KO
485 CMs
486 D) Bar chart showing overlap of CRHs with TADs, 80% of TADs are contained in one TAD
487 in the control while only 40% of CRHs are contained within one TAD in the CTCF KO.
488 In contrast 20% of CRHs span 2 or more TADs in the control, while 50% of CRHs span
489 2 or more TADs in CTCF KO.
490
491 Supplementary figure 1
492 A) Pearson analysis showing correlation between differential expressed genes and
493 differential enhancer peaks for ABC-linked enhancers during cardiac disease when
494 taking into account only genes and enhancers with a fold change of log2 0.5
495 B) Bar chart showing number of connections per gene promoters and per enhancers in
496 control and *Ctcf*-KO cells
497 C) Bar chart showing how many CRHs were merged to form a bigger CRH in *Ctcf*-KO cells,
498 and the percentage of each category
499
500
501
502

503 1 Maurano, M. T. *et al.* Systematic localization of common disease-associated variation
504 in regulatory DNA. *Science* **337**, 1190-1195, doi:10.1126/science.1222794 (2012).
505 2 Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic
506 circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300-307,
507 doi:10.1038/s41586-020-03145-z (2021).
508 3 Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from
509 thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669, doi:10.1038/s41588-
510 019-0538-0 (2019).
511 4 Yao, L. *et al.* A comparison of experimental assays and analytical methods for genome-
512 wide identification of active enhancers. *Nat Biotechnol*, doi:10.1038/s41587-022-
513 01211-7 (2022).
514 5 Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of
515 validated and target-linked human enhancers. *Nat Rev Genet* **21**, 292-310,
516 doi:10.1038/s41576-019-0209-0 (2020).
517 6 Anene-Nzelu, C. G., Lee, M. C. J., Tan, W. L. W., Dashi, A. & Foo, R. S. Y. Genomic
518 enhancers in cardiac development and disease. *Nat Rev Cardiol*, doi:10.1038/s41569-
519 021-00597-2 (2021).

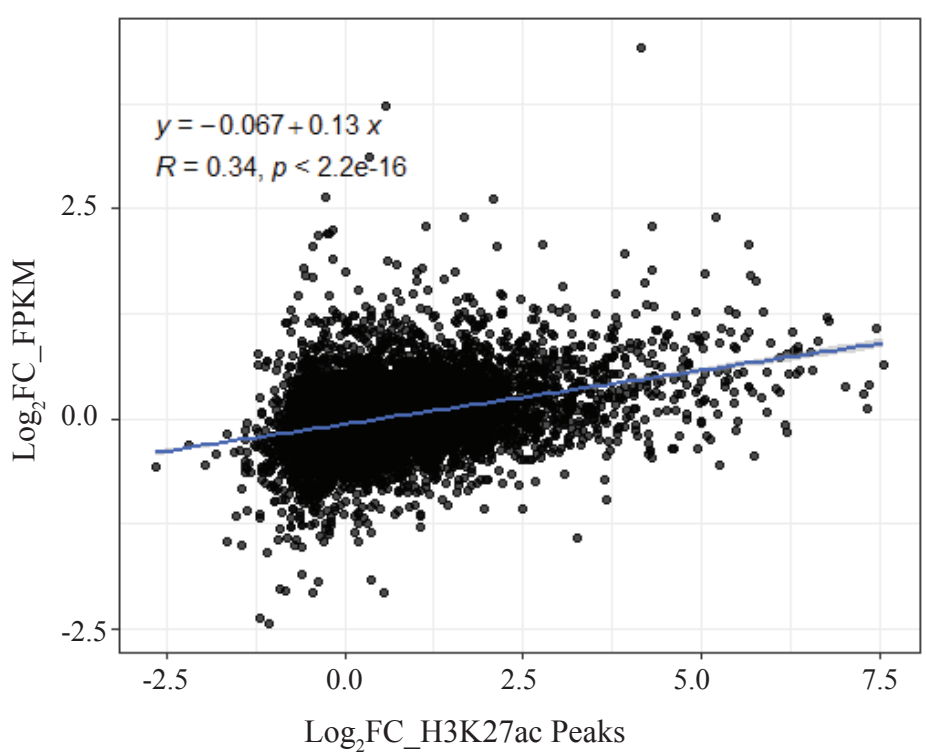
520 7 Consortium, E. P. *et al.* Expanded encyclopaedias of DNA elements in the human and
521 mouse genomes. *Nature* **583**, 699-710, doi:10.1038/s41586-020-2493-4 (2020).
522 8 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human
523 epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
524 9 Alexanian, M. *et al.* A transcriptional switch governs fibroblast activation in heart
525 disease. *Nature* **595**, 438-443, doi:10.1038/s41586-021-03674-1 (2021).
526 10 Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-
527 Thalassemia. *N Engl J Med* **384**, 252-260, doi:10.1056/NEJMoa2031054 (2021).
528 11 Matharu, N. *et al.* CRISPR-mediated activation of a promoter or enhancer rescues
529 obesity caused by haploinsufficiency. *Science* **363**, doi:10.1126/science.aau0629
530 (2019).
531 12 Espinola, S. M. *et al.* Cis-regulatory chromatin loops arise before TADs and gene
532 activation, and are independent of cell fate during early Drosophila development. *Nat*
533 *Genet* **53**, 477-486, doi:10.1038/s41588-021-00816-z (2021).
534 13 Mangnier, L., Joly-Beauparlant, C., Droit, A., Bilodeau, S. & Bureau, A. Cis-regulatory
535 hubs: a new 3D model of complex disease genetics with an application to
536 schizophrenia. *Life Sci Alliance* **5**, doi:10.26508/lsa.202101156 (2022).
537 14 Di Giammartino, D. C., Polyzos, A. & Apostolou, E. Transcription factors: building hubs
538 in the 3D space. *Cell Cycle* **19**, 2395-2410, doi:10.1080/15384101.2020.1805238
539 (2020).
540 15 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals
541 principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021
542 (2014).
543 16 Li, T., Jia, L., Cao, Y., Chen, Q. & Li, C. OCEAN-C: mapping hubs of open chromatin
544 interactions across the genome reveals gene regulatory networks. *Genome Biol* **19**,
545 54, doi:10.1186/s13059-018-1430-4 (2018).
546 17 Madsen, J. G. S. *et al.* Highly interconnected enhancer communities control lineage-
547 determining genes in human mesenchymal stem cells. *Nat Genet* **52**, 1227-1238,
548 doi:10.1038/s41588-020-0709-z (2020).
549 18 Oudelaar, A. M. *et al.* A revised model for promoter competition based on multi-way
550 chromatin interactions at the alpha-globin locus. *Nat Commun* **10**, 5412,
551 doi:10.1038/s41467-019-13404-x (2019).
552 19 Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell
553 Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e858,
554 doi:10.1016/j.molcel.2018.06.044 (2018).
555 20 Anene-Nzelu, C. G. *et al.* Assigning Distal Genomic Enhancers to Cardiac Disease-
556 Causing Genes. *Circulation* **142**, 910-912, doi:10.1161/CIRCULATIONAHA.120.046040
557 (2020).
558 21 Tsai, A., Alves, M. R. & Crocker, J. Multi-enhancer transcriptional hubs confer
559 phenotypic robustness. *Elife* **8**, doi:10.7554/eLife.45325 (2019).
560 22 Ong, C. T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-
561 specific gene expression. *Nat Rev Genet* **12**, 283-293, doi:10.1038/nrg2957 (2011).
562 23 Lee, D. P. *et al.* Robust CTCF-Based Chromatin Architecture Underpins Epigenetic
563 Changes in the Heart Failure Stress-Gene Response. *Circulation* **139**, 1937-1956,
564 doi:10.1161/CIRCULATIONAHA.118.036726 (2019).

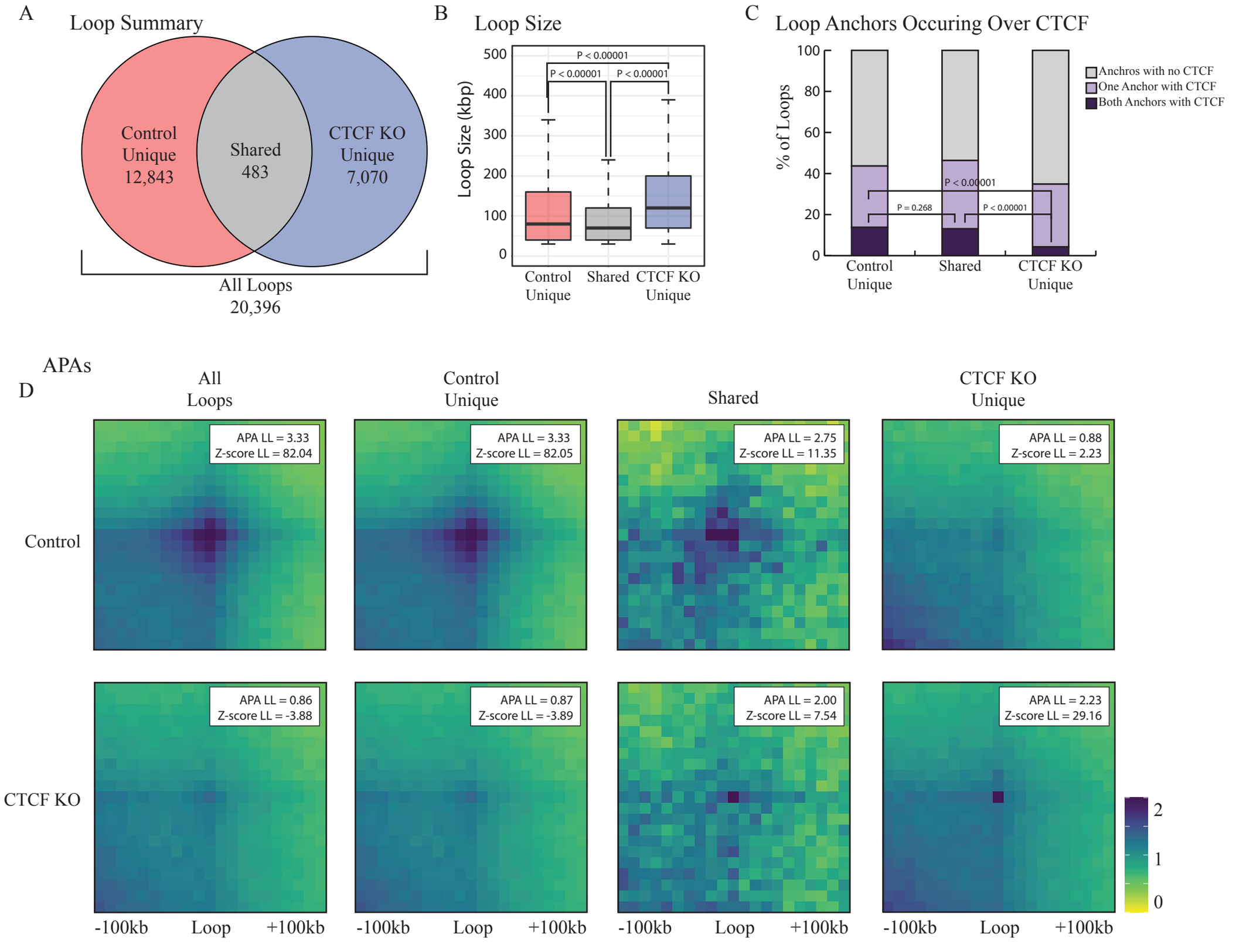
565 24 Man, J. C. K. *et al.* Genetic Dissection of a Super Enhancer Controlling the Nppa-Nppb
566 Cluster in the Heart. *Circ Res* **128**, 115-129, doi:10.1161/CIRCRESAHA.120.317045
567 (2021).
568 25 Sergeeva, I. A. *et al.* Identification of a regulatory domain controlling the Nppa-Nppb
569 gene cluster during heart development and stress. *Development* **143**, 2135-2146,
570 doi:10.1242/dev.132019 (2016).
571 26 Dickel, D. E. *et al.* Genome-wide compendium and functional assessment of in vivo
572 heart enhancers. *Nat Commun* **7**, 12923, doi:10.1038/ncomms12923 (2016).
573 27 Franke, M. *et al.* CTCF knockout in zebrafish induces alterations in regulatory
574 landscapes and developmental gene expression. *Nat Commun* **12**, 5415,
575 doi:10.1038/s41467-021-25604-5 (2021).
576 28 Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent
577 gene activation. *Nat Struct Mol Biol* **28**, 152-161, doi:10.1038/s41594-020-00539-5
578 (2021).
579 29 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed
580 genome architecture. *Nat Methods* **13**, 919-922, doi:10.1038/nmeth.3999 (2016).
581 30 Hsieh, T. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian
582 Chromatin Folding. *Mol Cell* **78**, 539-553 e538, doi:10.1016/j.molcel.2020.03.002
583 (2020).
584 31 Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant
585 chromatin contacts from HiChIP data by FitHiChIP. *Nat Commun* **10**, 4221,
586 doi:10.1038/s41467-019-11950-y (2019).
587 32 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of
588 Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922,
589 doi:10.1016/j.cell.2017.05.004 (2017).
590 33 Hang, C. *et al.* Knockout of MYOM1 in human cardiomyocytes leads to myocardial
591 atrophy via impairing calcium homeostasis. *J Cell Mol Med* **25**, 1661-1676,
592 doi:10.1111/jcmm.16268 (2021).
593 34 Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in
594 mammalian development. *Nature* **554**, 239-243, doi:10.1038/nature25461 (2018).
595 35 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables
596 interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396
597 (2017).
598 36 Csardi G, N. T. The igraph software package for complex network research. *Int. J.*
599 *Complex Syst.* (2006).
600

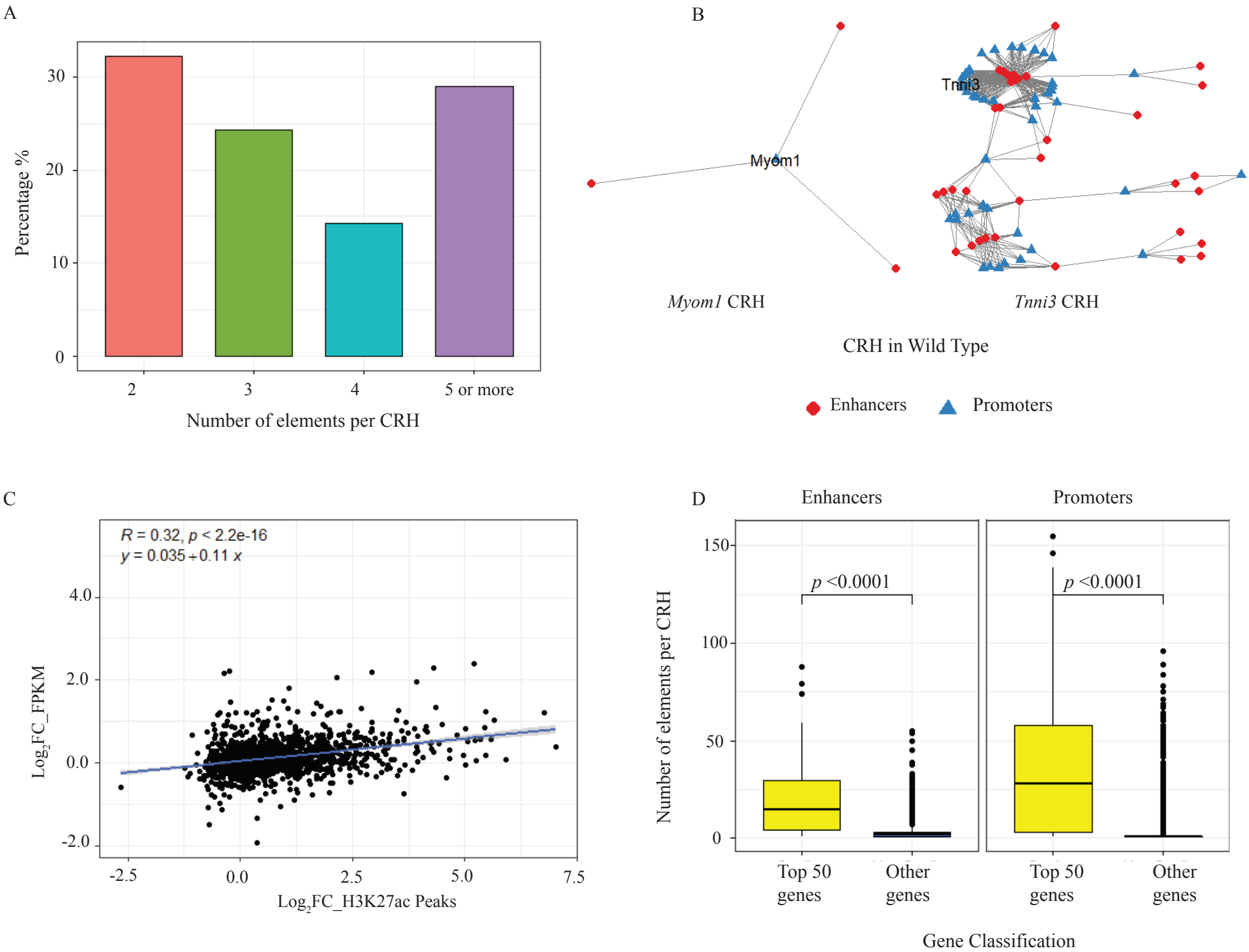
A

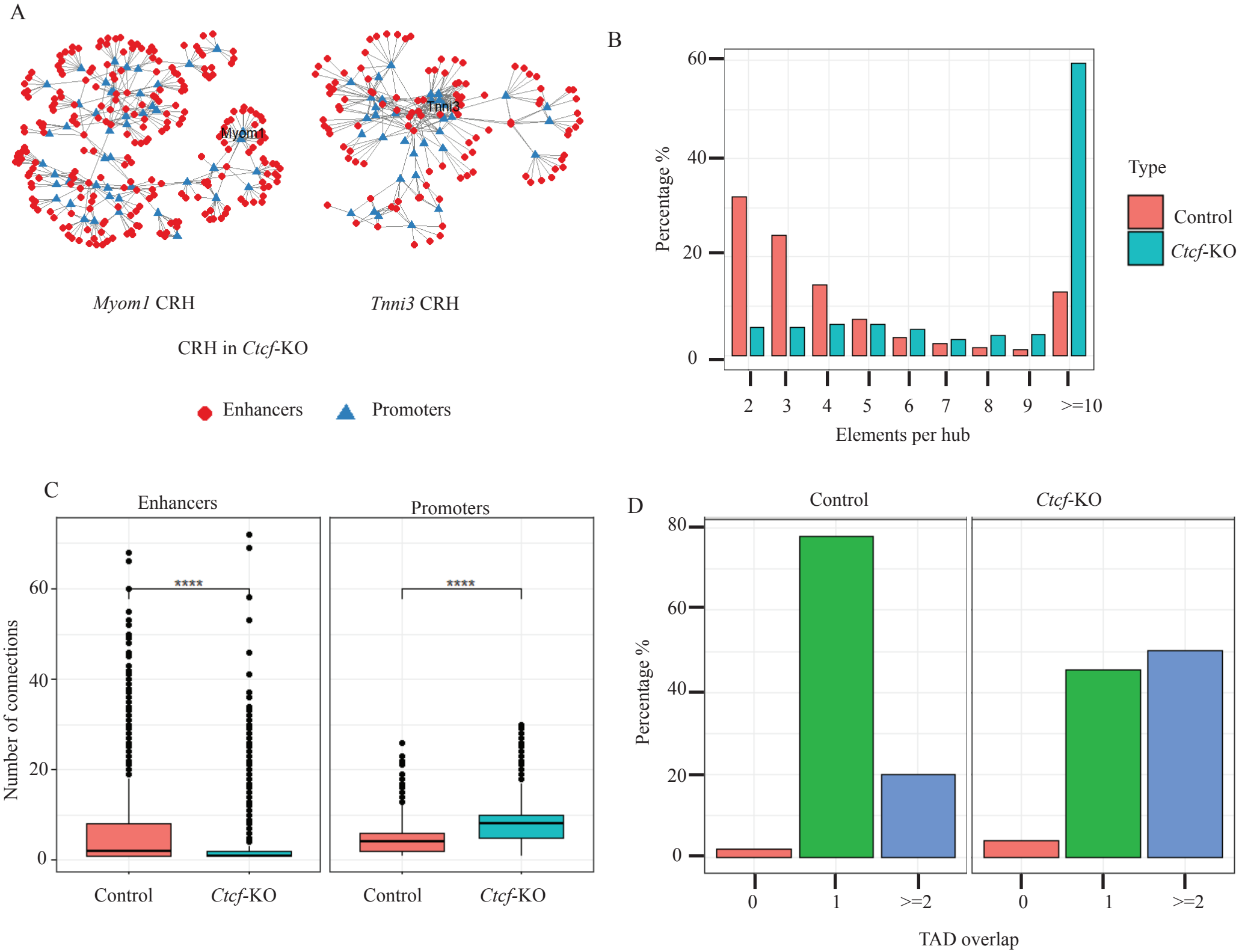

chr2:
20kb
mm10
H3K27ac
ChIP-seq
50
0
Top
Enhancer
Mybpc3
gene
chr14:
20kb
mm10
H3K27ac
ChIP-seq
50
0
Myh6
gene
Top
Enhancer

B

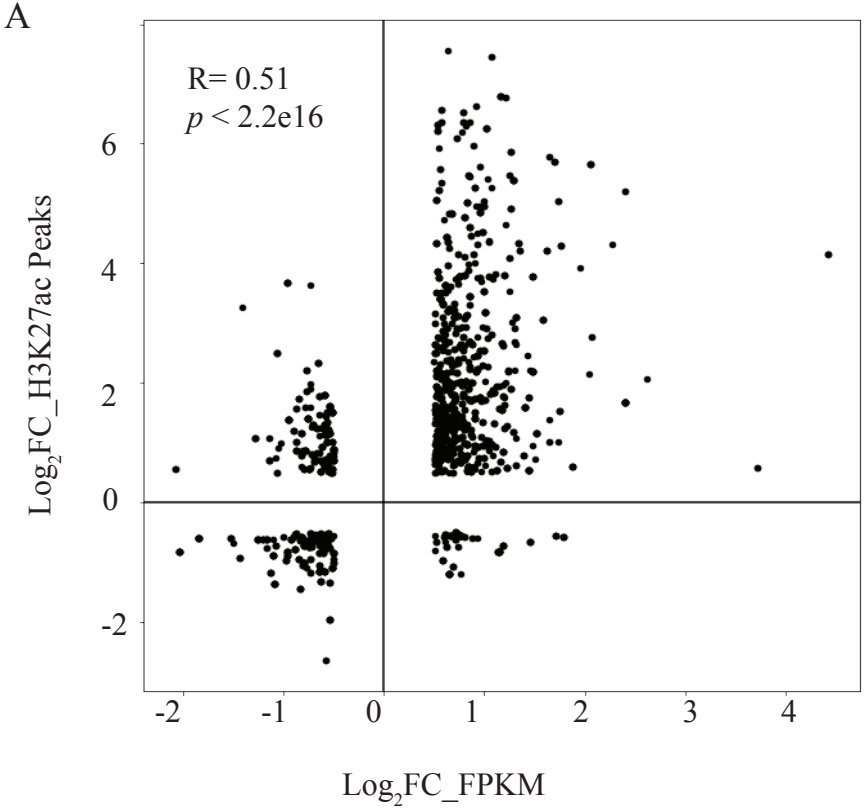

y = −0.067 + 0.13 x
R = 0.34, p < 2.2e-16
2.5
0.0
-2.5
Log2FC_FPKM
-2.5
0.0
2.5
5.0
7.5
Log2FC_H3K27ac Peaks


A
Loop Summary
Control Unique 12,843
Shared 483
CTCF KO Unique 7,070
All Loops 20,396
B
Loop Size
Loop Size (kbp)
P < 0.00001
P < 0.00001
P < 0.00001
Control Unique
Shared
CTCF KO Unique
C
Loop Anchors Occuring Over CTCF
% of Loops
Anchros with no CTCF
One Anchor with CTCF
Both Anchors with CTCF
P = 0.268
P < 0.00001
P < 0.00001
Control Unique
Shared
CTCF KO Unique
D
APAs
All Loops
Control Unique
Shared
CTCF KO Unique
Control
CTCF KO
APA LL = 3.33
Z-score LL = 82.04
APA LL = 3.33
Z-score LL = 82.05
APA LL = 2.75
Z-score LL = 11.35
APA LL = 0.88
Z-score LL = 2.23
APA LL = 0.86
Z-score LL = -3.88
APA LL = 0.87
Z-score LL = -3.89
APA LL = 2.00
Z-score LL = 7.54
APA LL = 2.23
Z-score LL = 29.16
-100kb
Loop
+100kb

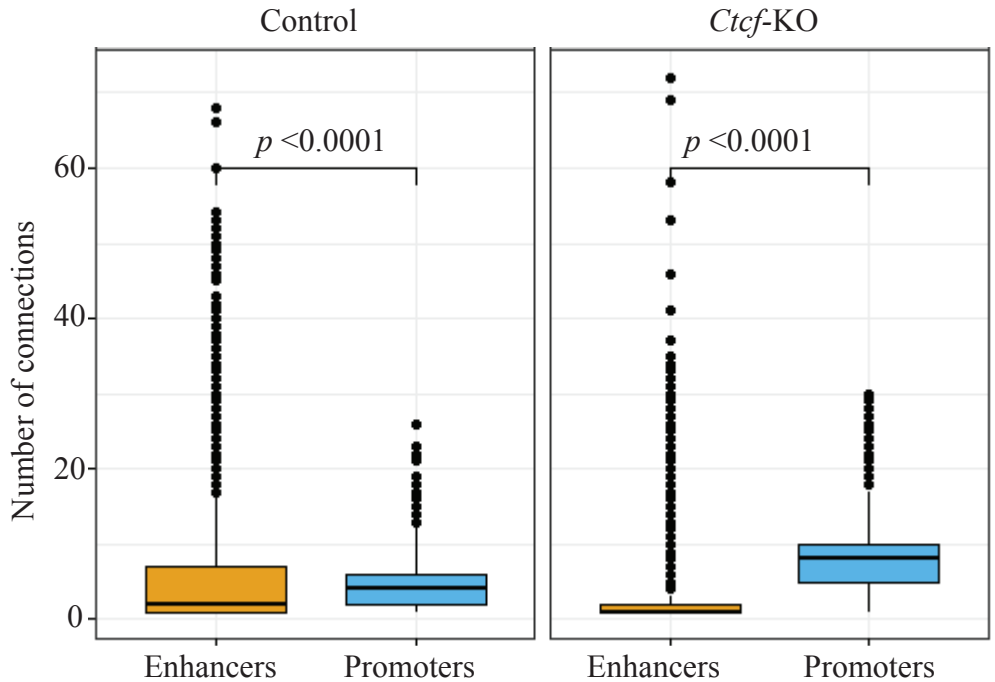

A
Percentage %
30
20
10
0
2
3
4
5 or more
Number of elements per CRH
B
Myom1
Tnni3
Myom1 CRH
Tnni3 CRH
CRH in Wild Type
Enhancers
Promoters
C
R = 0.32, p < 2.2e-16
y = 0.035 + 0.11 x
Log2FC_FPKM
4.0
2.0
0.0
-2.0
-2.5
0.0
2.5
5.0
7.5
Log2FC_H3K27ac Peaks
D
Enhancers
Promoters
Number of elements per CRH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

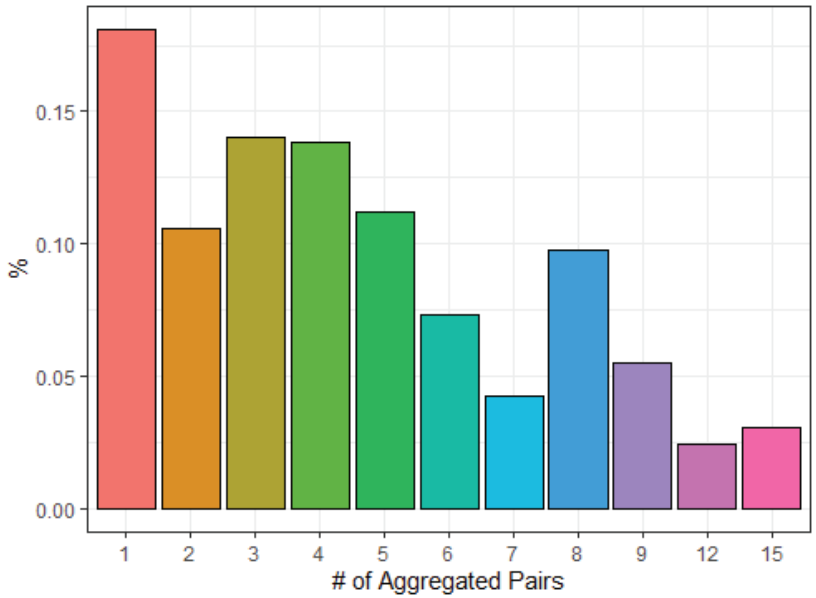
A



B



C



Supplementary Figure