MyoD is a structure organizer of 3D genome architecture in muscle cells

Qian Chen[1,10], Fengling Chen[2,10], Ruiting Wang[1,10], Minglei Shi[3], Antony K. Chen[4], Zhao Ma[4], Guohong Li[5], Min Wang[5], Hu Li[6], Xu Zhang[7], Jinbiao Ma[8], Jiayun Zhong[8], Meihong Chen[1], Michael Q. Zhang[2,3,9]*, Yong Zhang[1]*, Yang Chen[2,3]* & Dahai Zhu[1,6]*

1. Chinese Academy of Medical Sciences and Department of Biochemistry and Molecular Biology, The State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, School of Basic Medicine, Peking Union Medical College, 5 Dong Dan San Tiao, Beijing, 100005, China.

2. MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Bioinformatics Division, BNRist, Department of Automation, Tsinghua University, Beijing 100084, China.

3. MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Medicine, Tsinghua University, Beijing 100084, China.

4. Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, China.

5. National Laboratory of Biomacromolecules, CAS Center for Excellent in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China; University of Chinese Academy of Science, Beijing, 100049, China.

6. Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou, 510320, China.

7. Beijing institute of collaborative innovation, Beijing, 100094, China.

8. State Key Laboratory of Genetic Engineering, Department of Biochemistry, School of Life Sciences, Fudan University, Shanghai 200438, PR China.

9. Department of Biological Sciences, Center for Systems Biology, The University of Texas, Dallas 800 West Campbell Road, RL11, Richardson, TX 75080-3021, USA.

10. These authors contributed equally.

Contact information of corresponding authors: Dahai Zhu dhzhu@pumc.edu.cn or zhu_dahai@grmh-gdl.cn

Abstract

The genome is not a linear molecule of DNA randomly folded in the nucleus, but exists as an organized, three-dimensional (3D) dynamic architecture. Intriguingly, it is now clear that each cell type has a unique and characteristic 3D genome organization that functions in determining cell identity during development. A currently challenging basic question is how cell-type specific 3D genome structures are established during development. Herein, we analyzed 3D genome structures in primary myoblasts and myocytes from MyoD knockout (MKO) and wild type (WT) mice and discovered that MyoD, a pioneer transcription factor (TF), can function as a "genome organizer" that specifies the proper 3D genome architecture unique to muscle cell development. Importantly, we genetically demonstrate that H3K27ac is insufficient for establishing MyoD-induced chromatin loops in muscle cells. The establishment of MyoD's "architectural role" should have profound impacts on advancing understanding of other pioneer transcription factors in orchestrating lineage specific 3D genome organization during development in a potentially very large number of cell types in diverse organisms.

Introduction

The recent Hi-C analysis from comprehensive interaction map over large regions or whole genomes have indicated that the genome is hierarchically organized into chromosome territories, A/B compartments, topologically associated domains (TAD), and chromatin loops[1-3]. To date, much evidence has shown that the 3D structure of the genome is highly different among different cell types[4-6], however, the molecular mechanisms underlying the establishment of cell-type specific 3D genome organization are largely unknown. It has been proposed that "pioneer transcription factors" act as anchor proteins to orchestrate cell-type specific 3D genome architecture[7,8]; however, there is no concrete genome-scale evidence to support this hypothesis. We selected MyoD for our exploration of pioneer transcription factors mediated specification of developmental-context specific 3D genome organization for the following reasons. First, MyoD is well-established as a pioneer transcription factor in myogenic cell

56 lineage specification during development[9-12] and trans-differentiation[13]. This protein is known
57 to regulate the expression of myogenic specific genes through its binding to E-box (CANNTG)
58 containing cis-regulatory elements[14,15] and there are more than fourteen million consensus E-
59 boxes in the genome[14]. Second, over forty thousand MyoD binding peaks identified by ChIP-
60 seq from independent studies[15-17] in muscle cells, only 15% of these MyoD binding peaks are
61 located in promoter regions (Extended Data Fig. 1a). Finally, besides binding and
62 transcriptional activation of genes during differentiation, MyoD also constitutively binds at tens
63 of thousands of additional sites throughout the genome in proliferating muscle stem cells[15,18]
64 (Extended Data Fig. 1a). Together, these data represent an attractively large empirical scope to
65 support a genome-wide functional analysis of MyoD's potential roles in muscle cell lineage
66 specific 3D genome organization. In this report, we provide computational and experimental
67 evidence to uncover the unappreciated role of MyoD as a genome organizer in establishing the
68 unique 3D genome architecture in muscle cells, beyond its well-known functions as a TF in
69 activating myogenic gene expression during development.

70

71 **Results**

72 **MyoD regulates A/B compartments switching and formation of contact domain**
73 **boundaries (CDBs) in muscle cells**

74 In light of the aforementioned MyoD data, we first performed RNA-seq analysis of
75 proliferating (GM, Pax7$^+$ and MyoD$^+$) and differentiating (DM, MyoG$^+$) muscle cells isolated
76 from hind limb skeletal muscle of WT and MKO mice (Extended Data Fig. 1b), which we
77 combined with MyoD ChIP-seq data to examine potential functional link(s) between MyoD
78 occupancy and transcriptional activation. Specifically, we extracted MyoD binding peaks
79 localized at promoter regions and identified their cognate genes within ± 3 kb (Fig. 1a):
80 intriguingly, we found that knockout of MyoD did not affect the expression of a majority of
81 these genes (76.7% for 2691/3510 in primary proliferation myoblasts, 56.6% for 1987/3509 in
82 differentiation myocytes) (Fig. 1a and Supplementary Table 1), suggesting that the presence of

83 MyoD at these genes (even in promoter regions) frequently does not directly impact their
84 transcriptional activation. This result further supports the hypothesis that, beyond the canonical
85 function of MyoD as an activator of myogenic gene expression during development, this protein
86 may exert broad functional impacts, potentially on higher-order genome structures.

87 To directly investigate potential genome architectural roles of the pioneer TF MyoD in
88 muscle cells, we further examined four muscle sample types (WT-GM, WT-DM, MKO-GM,
89 MKO-DM) with bridge-linker Hi-C (BL-Hi-C)[19] and ChIP-seq with antibodies against CTCF
90 and histone mark H3K27ac (Fig. 1b). For the BL-Hi-C data, 28 libraries were sequenced to a
91 total depth of over 12 billion reads; the high quality deep sequencing data were validated with
92 high cis interaction rate (Extended Data Fig. 1c) and reached 5 kb resolution (Extended Data
93 Fig. 1d), thereby ensuring the rigor of our subsequent computational analyses of each
94 hierarchical level of chromatin structure in muscle cells.

95 We then examined the impacts of MyoD depletion on A/B compartments and contact
96 domain boundaries (CDBs) in muscle cells by performing exploratory principal component
97 analysis (PCA) based on the compartment values and relative insulation scores[20] for the four
98 muscle sample types examined (Methods). Our results showed clear distinctions between the
99 WT and MKO genotypes for both proliferating and differentiating muscle cells (Fig. 1c).
100 Moreover, we detected a clear A/B compartment shift in proliferating MKO cells vs. WT cells
101 (1.69% B→A, 0.84% A→B) (Fig. 1d left and 1e left). For differentiating cells, there was a
102 clear compartment switch for MKO cells compared to differentiating WT cells (1.38% B→A
103 and A→B 1.28%) (Fig. 1d left, and 1e right).

104 We also analyzed CDB dynamics in muscle cells in response to MyoD knock out. A total
105 of 931 CDBs exhibited differential extents of insulation in MKO cells compared to WT cells
106 in proliferating and differentiating states (Fig. 1d right and 1f), implying that MyoD may
107 somehow facilitate CDB insulation. Pursuing this, we compared insulation scores for CDBs
108 bound by MyoD or by CCCTC-binding factor (CTCF). In general, the insulation scores of
109 MyoD-bound CDBs were similar to the scores of CTCF-bound CDBs (Fig. 1g). However, the

110 CDBs bound solely by MyoD exhibited significantly weaker insulation compared to the CTCF-
111 bound CDBs, indicating that MyoD may work in concert with CTCF to regulate the insulation
112 of CDBs in muscle cells.

113 Collectively, our analysis of the genome-wide distribution of MyoD alongside functional
114 studies of MKO proliferating and differentiating muscle cells support the idea that MyoD has
115 previously unappreciated functional roles in regulating A/B compartments and contact domain
116 boundaries in muscle cells.

117 **MyoD is an anchor protein for chromatin loop formation in muscle cells**

118 The most well-characterized anchor protein in vertebrates is CTCF, which mediates
119 chromatin looping in concert with cohesin probably through the loop extrusion mechanism[2,21-
120 25]. Considering our finding that MyoD apparently functions together with CTCF to regulate
121 insulation of CDBs, we wondered whether MyoD may control the formation of chromatin loops
122 in muscle cells. Using HiCCUPS algorithm[26] to assess our BL-Hi-C dataset, we ultimately
123 identified 24,492 chromatin loops in proliferating WT myoblasts and 31,243 chromatin loops
124 in differentiating WT myocytes (Fig. 2a and Supplementary Table 2, Methods). The accuracy
125 of the loop calls was supported by high-scoring aggregate peak analysis (APA) plots (Fig. 2a,
126 Methods) and by observable enrichment of CTCF-bound loops in muscle cells (Extended Data
127 Fig. 2a) wherein about 75% of them had convergent CTCF binding motifs (Extended Data Fig.
128 2b).

129 We then searched for enriched motifs within accessible regions in loop anchors and
130 detected significant enrichment for known myogenic E-box motifs[14,15](Fig. 2b). This
131 enrichment was further confirmed by our observation of enrichment for MyoD binding peaks
132 at loop anchors as analyzed using pseudo-peaks (Methods) with combined MyoD ChIP-seq
133 data from our present study and other published work[17] (Fig. 2c and Supplementary Table 3):
134 44.18% of chromatin loops in GM cells were MyoD-bound (10.87% with both anchors bound
135 by MyoD, 33.31% with one anchor bound by MyoD), while 43.93% of chromatin loops in DM
136 cells were MyoD-bound (10.75% with both anchors bound by MyoD, 33.18% with one anchor

137 bound by MyoD). Furthermore, MyoD binding peaks at loop anchors were concordant with
138 binding signals of CTCF and SMC3[27], a subunit of the cohesin complex (Extended Data Fig.
139 2c). We cataloged all of the identified chromatin loops into four types based on the binding of
140 MyoD and/or CTCF at loop anchors in muscle cells: MyoD-MyoD(noCTCF), MyoD-
141 MyoD(CTCF), MyoD-CTCF and CTCF-CTCF (Fig. 2d, e). This classification was further
142 supported by the observable MyoD co-localization with CTCF as well as non-overlaid MyoD
143 signals or CTCF signals when the two proteins were imaged by direct stochastic optical
144 reconstruction microscopy (dSTORM), a single-molecule localization microscopy technique
145 that can reveal the organization of specific proteins with a lateral resolution of ~20 nm[28] (Fig.
146 2f). In addition, we found that the MyoD-bound loops were significantly shorter than CTCF-
147 CTCF loops (288 kb on average), and the MyoD-MyoD(noCTCF) chromatin loops were even
148 shorter (113 kb on average) (Extended Data Fig. 2d). Together, these data substantiate MyoD
149 localization on anchors of chromatin loops in muscle cells.

150 We next assessed whether MyoD could indeed mediate chromatin loop formation *in vivo*
151 by investigating the effects of MyoD knockout on the MyoD-bound loops detected in WT cells.
152 APA plots indicated that MyoD knockout significantly decreased the loop strength values for
153 both MyoD-bound chromatin loops (MyoD-MyoD(noCTCF), MyoD-MyoD(CTCF) and
154 MyoD-CTCF loops) and CTCF-CTCF chromatin loops, with these decreases evident in both
155 GM and DM cells (Fig. 2g and Extended Data Fig. 2e). These results support that MyoD
156 functionally contributed to the formation of both MyoD-bound and CTCF-bound chromatin
157 loops *in vivo*. Consistent with this possibility, fluorescence *in situ* hybridization (FISH)
158 confirmed that the MyoD-bound chromatin loop connecting the *Myog* promoter and the *Mybph*
159 promoter was established at the *Myog* locus in differentiating WT cells but not in differentiating
160 MKO muscle cells (Fig. 2h, i). Moreover, *in vitro* DNA circularization assays with DNA
161 fragments bearing known MyoD binding motifs (Methods) showed that MyoD protein induced
162 the circularization of linear DNA with MyoD binding motifs *in vitro* (Fig. 2j and Extended Data
163 Fig. 2f, g), confirming MyoD-orchestrated chromatin looping.

164 Collectively, these genetic, biochemical, and cellular imaging results demonstrate that
165 MyoD functions as an anchor protein that orchestrates chromatin loop formation across the
166 genome in muscle cells. Moreover, our data support that MyoD impacts chromatin loop
167 formation by itself and through its physical interaction with CTCF.

168 **MyoD anchors formation of myogenic-specific chromatin loops**

169 Based on the above findings, we next asked whether MyoD can mediate the organization
170 of muscle cell-specific chromatin looping. To test this, Hi-C data from embryonic stem (ES)
171 and cortical neuron (CN) cells[5] were collected, and the same four types of chromatin loops
172 identified in muscle cells were examined in ES and CN cells. Only a few of MyoD-
173 MyoD(noCTCF) loops were identified in ES cells (only 8 loops, 1.8%) or CN cells (only 1 loop,
174 0.2%) (Fig. 3a), revealing that the MyoD-MyoD(noCTCF) loops detected in muscle cells
175 represent myogenic lineage specific chromatin loops. Strikingly, by assessing gene expression
176 trends associated with four types of loops across eight cell types (Methods), we demonstrated
177 that the genes from the MyoD-MyoD(noCTCF) loops exhibited the strongest extent of muscle
178 lineage specific expression (Fig. 3b). Accordantly, the genes associated with MyoD-
179 MyoD(noCTCF) chromatin loops also had the most enriched GO terms related to muscle cell
180 differentiation (Extended Data Fig. 3a).

181 To seek further molecular insights into MyoD-bound muscle cell-specific chromatin
182 looping, we examined ChIP-seq data for both MyoD and NeuroD2 — a pioneer TF for neuronal
183 cell lineage commitment and differentiation[18,29] in both muscle cells and neuron cells. Only 18%
184 of MyoD binding peaks in muscle cells were overlapped with NeuroD2 binding peaks in
185 cortical neuron cells, suggesting that MyoD and NeuroD2 might mediate the formation of
186 separate sets of chromatin loops that can subsequently regulate distinct sets of target genes to
187 properly specify lineage-specific cell identities. For example, we could identify that the *Myh*
188 gene cluster that are located within MyoD-bound, muscle-cell-specific chromatin loop were not
189 detected in neuronal cells, and there was no NeuroD2 binding at the same genomic region in
190 neuronal cells (Fig. 3c). *Vice versa*, chromatin loops bound by the NeuroD2 at the *Foxg1* loci

191 is neuron-specific and not observed in muscle cells (Extended Data Fig. 3b). Taken together,
192 these results illustrate that cell-lineage pioneer TFs exert genome-wide architectural roles for
193 modulating 3D genome organization, a functional impact considerably larger than the previous,
194 primarily gene-transcriptional-activation related understanding about the influence of pioneer
195 TFs on lineage specification.

196 **MyoD mediates primed architectural chromatin loops in proliferating muscle cells**

197 Recalling previously published studies showing that MyoD constitutively binds to tens of
198 thousands of sites genome-wide in proliferating cells[15,18] , together with our present finding that
199 such binding only infrequently results in transcriptional activation at this stage, and that MyoD
200 functions as an anchor protein for lineage-specific 3D genome organization, we sought to
201 further explore the nature of MyoD's genome architectural functions by re-assessing our BL-
202 Hi-C data. We were particularly interested in addressing the question of how the presence of
203 MyoD influences internal interactions occurring within chromatin loops.

204 Previous studies have indicated that CTCF and cohesin function together to establish
205 architectural loops and that the extent of enhancer-promoter (E-P) interactions is elevated
206 within such loops[30]. As our results identify a notable co-localization of MyoD binding peaks
207 with CTCF (Fig. 2f), we speculated that MyoD-bound loops may also feature increased intra-
208 loop interactions. To test this, we measured internal interactions for each loop by calculating
209 "domain score" (D-score) in WT and MKO proliferating cells. MyoD binding is enriched on
210 anchors of chromatin loops with reduced internal interactions (65%) in MKO-GM cells
211 compared with WT-GM cells (Fig. 4a, b), indicating that MyoD regulates the internal
212 interactions within chromatin loops by its direct binding on loop anchors like CTCF and cohesin.
213 Further analysis showed that interactions changes within loops were positively correlated with
214 expression changes of genes enclosed in the same chromatin loop (Fig. 4c).

215 We next focused on analyzing the internal interaction changes of MyoD-bound loops and
216 their cognate gene expression (Fig. 4d). We specifically examined those MyoD-bound loops
217 displaying reduced interactions in MKO proliferating cells because these loops might be

218 regulated directly by MyoD binding. We found that nearly 50% of the examined loops enclosed
219 down-regulated genes with known myogenic-related functions (Fig. 4d), indicating that
220 interactions within these MyoD-bound loops are necessary for the regulation of muscle identity
221 gene expression. For example, the *Rbm24* locus, a known gene that regulates myogenesis[31] and
222 its differential expression in WT-GM and MKO-GM cells offers an excellent illustration for a
223 gene enclosed within this type of MyoD-bound loop in muscle cells (Fig. 4e and Extended Data
224 Fig. 4a).

225 Intriguingly, we also found that most (89%) of the remaining MyoD-bound loops with
226 decreased interactions did not contain any differentially expressed genes upon MyoD knockout
227 (Fig. 4d). This finding underscored the apparent centrality of the architectural role (rather than
228 transcriptional activation) of MyoD for these MyoD-bound loops in proliferating muscle cells.
229 Although reduced interactions within these chromatin loops in MKO cells did not alter gene
230 expression in proliferating cells, we found that genes within these chromatin loops were not
231 turned on or off properly in the MKO-DM cells as compared with WT-DM cells (Fig 4f), and
232 this genetic dysregulation was clearly manifest in the defective differentiation phenotype of the
233 MKO cells (Extended Data Fig. 1b) [32-34]. As the loops were still maintained in differentiated
234 cells (Extended Data Fig. 4b), it appears that the interactions constrained within the loops in
235 proliferating cells are effectively "primed" for muscle cell differentiation by the presence of
236 MyoD, raising the possibility that some very early signal can somehow direct the occupancy of
237 MyoD to the sites across the genome that well subsequently permit both muscle-cell-
238 appropriate loop architecture and later rewiring in response to differentiation signals.
239 Supporting this idea, we found that *Lrmt1*, a known myogenesis related gene[35], which is only
240 expressed in differentiated muscle cells (Extended Data Fig. 4c), contains chromatin loops that
241 have already been defined by MyoD in proliferating cells (Fig. 4g). Thus mechanistically, our
242 analysis of the MyoD-bound loops and their internal interactions revealed molecular insights
243 about how MyoD exerts its architectural function, even in proliferating cells, to ensure the
244 correct trajectory towards eventual muscle cell identity.

245 **The regulatory loops specified by MyoD are functionally required for muscle cell**
246 **differentiation**

247 The genome is not only structurally organized within the nucleus but is also dynamically
248 orchestrated in response to various cellular signals[5,36-40]. However, the molecular mechanisms
249 underlying dynamic changes in 3D genome structure are largely unknown. Although the
250 biological function of MyoD in regulating muscle cell differentiation has been well documented,
251 and very recently, MyoD's transcription activity was found to be regulated by its ability to
252 mediate chromatin interactions during trans-differentiation[41], how MyoD impacts chromatin
253 conformation dynamics during muscle cell differentiation remains largely elusive. Using a
254 modified differential loop detection method[40] (Methods), we identified 6,242 differential
255 chromatin loops (~25%) between proliferating and differentiating muscle cells. Among these
256 loops, 5,754 were significantly enhanced whereas 488 were reduced in differentiating cells (Fig.
257 5a), and we detected significant enrichment for muscle cell differentiation related genes
258 associated with the enhanced chromatin loops (Extended Data Fig. 5a). These data also revealed
259 that nearly 25% of chromatin loops that undergo dynamic changes during muscle cell
260 differentiation contribute to 3D genome rewiring in a manner that functionally regulates
261 myogenic differentiation.

262 Moreover, similar analysis performed on MKO cells identified 585 differential chromatin
263 loops between WT-DM and MKO-DM cells. Consistent with their regulatory role, 81.5 %
264 (477/585) of these were significantly reduced in the differentiated MKO cells (Extended Data
265 Fig. 5b), and 71.57% (341/477) of the reduced loops identified in the differentiated MKO cells
266 had MyoD binding peaks support them as bona fide MyoD mediated chromatin loops in
267 differentiated WT cells (Extended Data Fig. 5c). Among the aforementioned enhanced loops in
268 differentiation, 285 loops might be directly induced by MyoD as they were bound by MyoD
269 and disappeared when MyoD was absent in differentiating muscle cells (Supplementary Table
270 4). The genes associated with those MyoD-induced loops significantly up-regulated during
271 differentiation and down-regulated when MyoD was absent, indicating the regulatory role of
272 MyoD-induced loops (Fig. 5b).

273 Next, we sought to understand the driving force for the enhancement of MyoD-induced
274 loops during muscle differentiation. One possibility might be due to increased MyoD binding
275 on loop anchors during muscle cell differentiation. By examining MyoD binding dynamics on
276 anchors of MyoD-induced loops, we found that majority of MyoD-induced loops did not show
277 novel MyoD binding during the differentiation (Fig. 5c). However, we found that 88% of
278 MyoD-induced loops were decorated with MyoD dependent H3K27ac modification (Fig. 5c).
279 Collectively, these findings raise possibility that H3K27ac may contribute to loop formation
280 during muscle cell differentiation and that MyoD somehow actively directs the deposition of
281 these modifications on the MyoD-induced loops.

282 Taking the *Mef2c* locus as an example of MyoD-induced loops to demonstrate their
283 regulatory role in differentiation, we found that enhancer-promoter loops linked to *Mef2c*
284 emerged during differentiation in a MyoD dependent manner (Fig. 5d). Concordantly,
285 H3K27ac modification was enhanced in WT-DM cells but diminished in MKO-DM cells at
286 this locus (Fig. 5d) while *Mef2c* gene expression was not induced in MKO-DM cells (Fig. 5f
287 upper). Similar results were obtained when we examined MyoD-induced loops within the
288 *Dyrk2* locus, from which the mRNA transcription product was only abundant in WT-DM cells
289 (Fig. 5e and Fig. 5f lower). Collectively, these results demonstrate MyoD orchestrates the
290 formation of myogenic lineage specific chromatin loops that are required for the transcriptional
291 regulation of myogenic genes during muscle cell differentiation.

292 **H3K27ac in the MyoD knock out cells is not sufficient to drive formation of the MyoD-**
293 **bound chromatin loops**

294 The dynamic organization of the 3D genome is known to facilitate sophisticated interplay
295 between aaccessible versus inaccessible chromatin states and TF occupancy[38]; however, the
296 interdependent impacts (and likely feedback) of chromatin status on 3D genome organization
297 have remained largely elusive. Aside from MyoD dependent H3K27ac modification on MyoD-
298 induced loops, we found that during muscle cell differentiation, the loop anchors which had
299 increased H3K27ac signals showed enrichment for consensus E-box sequences (i.e., the known

300 target motif of the MyoD protein) in general (Fig. 6a). When MyoD was knocked out, H3K27ac
301 modification was decreased on MyoD binding regions (Fig. 6b) which further support the
302 overall dependency of H3K27ac modification on MyoD during differentiation. As MyoD itself
303 stably binds around many myogenic genes in proliferating muscle cells, MyoD dependent
304 H3K27ac modification correlates even better with the loop formation or disruption. These
305 results and previously reported positive correlation between H3K27ac and loop[39,42] prompted
306 us to examine whether H3K27ac-mediated active chromatin state directly triggers the formation
307 of MyoD-bound chromatin loops during muscle cell differentiation.

308 We tested this possibility by treating MKO cells with the histone deacetylase I specific
309 inhibitor Tacedinaline (Tac) to enrich the H3K27ac signal in the cells throughout muscle cell
310 differentiation. It was found that Tac treatment could result in the accumulation of H3K27ac
311 modification in 18, 110 genomic regions including anchors of MyoD-induced loops (96/285)
312 (Fig. 6c). Furthermore, analysis of the BL-Hi-C data from the Tac-treated MKO and DMSO-
313 treated cells showed that, in the absence of MyoD, the accumulation of H3K27ac *per se* is not
314 sufficient to trigger the formation of new loops or to enhance the signal for MyoD-bound or
315 CTCF-CTCF chromatin loops (Fig. 6d, e). Thus, H3K27ac alone is insufficient for establishing
316 characteristic 3D genome organization in muscle cells, as simply accumulating H3K27ac
317 cannot rescue MyoD-induced loops or drive other loop formation. Together, our overall
318 findings reveal a critically determination role of the pioneer TF MyoD as a genome organizer
319 in establishing the unique 3D genome architecture in muscle cells.

320

321 **Discussion**

322 Each cell type acquired unique stage-specific organization of 3D genome during
323 development and cell-type specific organization of 3D genome can be regarded as an emergent
324 property that is mediated by interplay between transcription factors and chromatin-associated
325 proteins[43]. Based on this model, the components in regulating the emergent property of
326 chromatin should be coordinately orchestrated by a genome organizer to instruct specific

327 organization of 3D genome in each cell type at a specific developmental stage. In this report,
328 we provide compelling evidence to support this model wherein binding of the pioneer
329 transcription factor MyoD at sites across the genome in muscle stem cells exerts previously
330 unappreciated function as a genome organizer, which has broad functional impacts on genome
331 structure beyond its known functions in activating gene expression during development
332 (Extended Data Fig. 6). Mechanistically, we show that internal interaction changes in MyoD-
333 bound loops suggest a new insight into the potential mechanism of cell lineage pioneer TFs to
334 ultimately specify the positions and nature of the chromatin loops comprising the cell-type
335 specific internal interactions which ensures the subsequent differentiation process leading to
336 lineage specific cell identity. Further, we biochemically demonstrate that interaction between a
337 pioneer TF MyoD and structural protein CTCF drives different types of loops which ultimately
338 supports a lineage specific genetic programme. Finally, we show how the occupancy of pioneer
339 TF proteins coordinate the epigenetic mark H3K27ac to retain active expression of the pioneer-
340 TF-specified genome structures and developmental genetic programmes underlying proper
341 muscle stem development.

342 It will be interesting to determine which signals ultimately initiate MyoD's loop
343 specification layer of genetic regulation at the earliest possible stage of myogenic lineage-
344 specific cell fate determination. Excitingly, extending our insights beyond MyoD-mediated
345 architectural regulation of muscle cell fate, we have also provided additional evidence from
346 experiments with the well-studied neuron specific pioneer TF NeuroD2 in neuronal cells. This
347 result suggests a highly similar loop specification function of NeuroD2 which apparently
348 underlies the neuron lineage specific 3D genome architecture and attends (downstream) genetic
349 programming of future cell fate. We are eager to see if other lineage-specific pioneer TFs in
350 different developmental systems or organisms could confirm the general role, similar to that
351 played by MyoD in our discovery, as the genome organizers that orchestrate the lineage specific
352 chromatin and 3D genome structures required for cell fate determination during development.

353

Acknowledgments

This work was supported by grants from the National Key Research and Development Program of China (2016YFA0100703, 2017YFA0505503), the National Natural Science Foundation of China (Nos. 91540206, 31871343, 31671384, 31971080 and 91949106), the Natural Science Foundation in Beijing (No. 7192125), and the CAMS Initiative for Innovative Medicine (Nos. 2016-I2M-1-017, 2019-I2M-1-004). The funding agencies do not involve in any research activities in the present study.

We thank Professor Cheng Li and Professor Bing Zhu for useful discussion and critical reading of the manuscript.

Author contributions

Q.C. conducted cell culture, BL-Hi-C, ChIP-seq, RNA-seq libraries and wrote the manuscript. F.C. performed all computational and bioinformatics analysis, generated the figures and tables as well as wrote the manuscript. R.W. carried out 3D-FISH analysis and immunostaining of MyoD and CTCF. M.S. helped to perform BL-Hi-C libraries. M.C. helped data analysis. A.K.C. and Z.M. performed dSTORM analysis for co-localization of MyoD and CTCF. G.L. and M.W. conducted in vitro DNA circularization and transmission electron microscopy. H.L. helped data analysis. X.Z. helped to conduct 3D-FISH. J.M. and J.Z. purified recombinant MyoD protein. Y.C. and M.Q.Z. supervised computational and bioinformatics analysis. Y.Z. helped to design and supervised the experiments and wrote the manuscript. D.Z. conceived of and supervised the project and wrote the manuscript. The authors declare no competing financial interests.

References

1 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

2 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).

381 3 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by
382 analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
383 4 Zheng, H. & Xie, W. The role of 3D genome organization in development and
384 cell differentiation. *Nat Rev Mol Cell Biol* **20**, 535-550 (2019).
385 5 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural
386 Development. *Cell* **171**, 557-572 e524 (2017).
387 6 Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human
388 genome. *Nature* **583**, 737-743 (2020).
389 7 Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome
390 conformation in cell-fate decisions. *Nature* **569**, 345-354 (2019).
391 8 Kim, S. & Shendure, J. Mechanisms of Interplay between Transcription Factors
392 and the 3D Genome. *Mol Cell* **76**, 306-319 (2019).
393 9 Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected
394 cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000 (1987).
395 10 Tapscott, S. J. *et al.* MyoD1: a nuclear phosphoprotein requiring a Myc
396 homology region to convert fibroblasts to myoblasts. *Science* **242**, 405-411
397 (1988).
398 11 Weintraub, H. *et al.* Activation of muscle-specific genes in pigment, nerve, fat,
399 liver, and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad*
400 *Sci U S A* **86**, 5434-5438 (1989).
401 12 Rudnicki, M. A. *et al.* MyoD or Myf-5 is required for the formation of skeletal
402 muscle. *Cell* **75**, 1351-1359 (1993).
403 13 Lassar, A. B., Paterson, B. M. & Weintraub, H. Transfection of a DNA locus
404 that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**, 649-
405 656 (1986).
406 14 Lassar, A. B. *et al.* MyoD is a sequence-specific DNA binding protein requiring
407 a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell*
408 **58**, 823-831 (1989).
409 15 Cao, Y. *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential
410 for broad cellular reprogramming. *Dev Cell* **18**, 662-674 (2010).
411 16 Mousavi, K. *et al.* eRNAs promote transcription by establishing chromatin
412 accessibility at defined genomic loci. *Mol Cell* **51**, 606-617 (2013).
413 17 Umansky, K. B. *et al.* Runx1 Transcription Factor Is Required for Myoblasts
414 Proliferation during Muscle Regeneration. *PLoS Genet* **11**, e1005457 (2015).
415 18 Fong, A. P. *et al.* Genetic and epigenetic determinants of neurogenesis and
416 myogenesis. *Dev Cell* **22**, 721-735 (2012).
417 19 Liang, Z. *et al.* BL-Hi-C is an efficient and sensitive approach for capturing
418 structural and regulatory chromatin interactions. *Nat Commun* **8**, 1622 (2017).
419 20 Chen, F., Li, G., Zhang, M. Q. & Chen, Y. HiCDB: a sensitive and robust
420 method for detecting contact domain boundaries. *Nucleic Acids Res* **46**, 11239-
421 11250 (2018).

21 Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503-507 (2017).
22 Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e614 (2017).
23 Heidari, N. *et al.* Genome-wide map of regulatory interactions in the human genome. *Genome Res* **24**, 1905-1917 (2014).
24 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).
25 Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO journal* **32**, 3119-3129 (2013).
26 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
27 Tsai, P. F. *et al.* A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Mol Cell* **71**, 129-141.e128 (2018).
28 Heilemann, M. *et al.* Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angewandte Chemie (International ed. in English)* **47**, 6172-6176 (2008).
29 Bayam, E. *et al.* Genome-wide target analysis of NEUROD2 provides new insights into regulation of cortical projection neuron migration and differentiation. *BMC genomics* **16**, 681 (2015).
30 Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789-800 (2018).
31 Yang, J. *et al.* RBM24 is a major regulator of muscle-specific alternative splicing. *Dev Cell* **31**, 87-99 (2014).
32 Megeney, L. A., Kablar, B., Garrett, K., Anderson, J. E. & Rudnicki, M. A. MyoD is required for myogenic stem cell function in adult skeletal muscle. *Genes Dev* **10**, 1173-1183 (1996).
33 Sabourin, L. A., Girgis-Gabardo, A., Seale, P., Asakura, A. & Rudnicki, M. A. Reduced differentiation potential of primary MyoD-/- myogenic cells derived from adult skeletal muscle. *The Journal of cell biology* **144**, 631-643 (1999).
34 Montarras, D., Lindon, C., Pinset, C. & Domeyne, P. Cultured myf5 null and myoD null muscle precursor cells display distinct growth defects. *Biology of the cell* **92**, 565-572 (2000).
35 Li, H. K. *et al.* LRTM1 promotes the differentiation of myoblast cells by negatively regulating the FGFR1 signaling pathway. *Exp Cell Res*, 112237 (2020).
36 Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64 (2009).
37 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336 (2015).

38 Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* **50**, 238-249 (2018).
39 Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**, 1522-1528 (2017).
40 Phanstiel, D. H. *et al.* Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Mol Cell* **67**, 1037-1048 e1036 (2017).
41 Dall'Agnese, A. *et al.* Transcription Factor-Directed Re-wiring of Chromatin Architecture for Somatic Cell Nuclear Reprogramming toward trans-Differentiation. *Mol Cell* **76**, 453-472 e458 (2019).
42 Petrovic, J. *et al.* Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. *Mol Cell* **73**, 1174-1190 e1112 (2019).
43 Partridge, E. C. *et al.* Occupancy maps of 208 chromatin-associated proteins in one human cell type. *Nature* **583**, 720-728 (2020).

Methods

Cell line

C2C12 cell culture and differentiation. Mouse C2C12 cells (ATCC, CRL-1772) were cultured in growth medium consisting of Dulbecco's modified Eagle's medium (Gibco) supplemented with 4.5 g/L glucose, 10% fetal bovine serum (Ausbian), 1% penicillin and streptomycin at 37°C in a 5% CO_2 atmosphere. For the differentiation of C2C12 myoblasts, cells were transferred to Dulbecco's modified Eagle's medium containing 2% horse serum and 1% antibiotics, and then cultured for 24 hrs. C2C12 cells were grown to 80–90% confluence before induction of differentiation.

Mouse strains

MyoD KO mice (#002523) were obtained from the Jackson Laboratory. All the animal procedures were approved by the Animal Ethics Committee of Peking Union Medical College,

497 Beijing, China (ACUC-A01-2016-003). Mice were housed in the animal facility and had free
498 access to water and standard rodent chow.

499

500 **Primary myoblasts**

501 **Primary myoblast isolation, culture and differentiation.** Primary myoblasts were isolated
502 from hind limb skeletal muscle of MKO and WT littermates at 2~3 weeks old, minced, and
503 digested in a mixture of type II collagenase and Dispase. Cells were filtered from debris,
504 centrifuged, and cultured in growth medium (F-10 Ham's medium supplemented with 20%
505 fetal bovine serum, 10 ng/ml basic fibroblast growth factor and 1% antibiotics) on collagen-
506 coated cell culture plates at 37°C in 5% CO_2. For the differentiation of primary myoblasts, cells
507 were transferred to Dulbecco's modified Eagle's medium (Gibco) containing 2% horse serum
508 and 1% penicillin and streptomycin, and then cultured for 24 hrs. All cells were grown to 60–
509 70% confluence before induction of differentiation. For H3K27ac accumulation experiments,
510 MKO myoblasts were treated with 2.5 μM Tacedinaline (CI994) in differentiation medium for
511 12 hrs to enrich H3K27ac while DMSO treated MKO myoblasts in DM for 12 hrs served as
512 control. Then the cells were collected for ChIP-seq and BL-Hi-C. Tacedinaline (CI994) was
513 purchased from Selleck.

514

515 **Cell preparation for BL-Hi-C and library construction**

516 The BL-Hi-C library construction was performed as previously described[19] with some
517 modifications. In brief, the primary myoblasts were treated with 2% formaldehyde for 15 min
518 (R.T.) to crosslink protein–protein and protein–DNA in the cells, the reaction was quenched by
519 adding glycine solution (F.C. 0.2 M) for 5 min (R.T.). Scrape, transfer and centrifuge to collect
520 the cells, mix cells well and then split the crosslinked cell suspension into aliquots of 3-5 × 10^5
521 cells. Cell pellets were frozen in liquid nitrogen and stored at -80°C until use.

522 The cells were suspended using 1ml 1% SDS lysis buffer (50 mM HEPES-KOH, 150 mM
523 NaCl, 1 mM EDTA, 1% Triton X-100 and 1% SDS) at R.T. for 15 min. Wash the pellet with
524 1ml 0.1% SDS lysis buffer (50 mM HEPES-KOH, 150 mM NaCl, 1 mM EDTA, 1% Triton X-
525 100 and 0.1% SDS). Then centrifuge at 500 g for 2 min, discard the supernatant and add 50 μl
526 of 0.5% SDS for 30 min at 37°C. Add 145 μl H_2O and 12.5 μl 20% Triton X-100 to quench
527 SDS. Then, the genome was digested by 5 μl enzyme HaeIII (10 U/μl, NEB) with 81.5 μl H_2O,
528 10 μl 10 × NEB buffer 2, 2.5 μl 20% Triton X-100 and 1 μl BSA into fragments with blunt-
529 ends at 37°C for at least 2 hrs. Next, the blunt-ends of the DNA fragments were treated with
530 adenine and ligated with bridge linker containing biotin for 4 hrs (at R.T.). Next, add 2 μl 10
531 mM dATP and 5 μl Klenow Fragment (3' -> 5' exo-) (NEB) to digest the unligated DNA
532 fragments at 37°C for 40 min. Then pellet the cells and resuspend with 79 μl H_2O, 2 μl Bridge
533 Linker, 10 μl T4 DNA ligase buffer, 5 μl 20% Triton X-100, 1 μl BSA and 5 μl T4 DNA ligase
534 to perform the proximity ligation assay at R.T. for 4 hrs. Spin down and add 88 μl H_2O, 10 μl
535 lambda Exonuclease buffer, 1 μl lambda Exonuclease and 1 μl Exonuclease I to remove
536 unligated Bridge Linker. The cells were then digested by 2 μl 10% SDS, 5 μl 20mg/ml
537 Proteinase K (Ambion) and 4 μl 5M NaCl at 65°C for 6 hrs or overnight. The DNA was purified
538 following phenol–chloroform (Solarbio) extraction and ethanol precipitation. Further, the DNA
539 was fragmented to 300 bp average length using the S220 Focused-ultrasonicator (Covaris), and
540 the biotin-labeled DNA fragments were pulled by the streptavidin coated Dynabeads M280
541 (Thermo Fisher). Then discard the supernatant and add 78 μl H_2O, 10 μl T4 DNA ligase buffer,
542 1 μl 10mM dNTP, 5 μl T4 polynucleotide kinase (NEB), 1 μl T4 DNA Polymerase (NEB) and
543 5 μl Large (Klenow) fragment (NEB) to the DNA at R.T. for 30 min to repair end. Next, discard
544 the supernatant and add 83 μl H_2O, 10 μl NEB buffer 2, 2 μl 10 mM dATP and 5 μl Klenow
545 Fragment (3' -> 5' exo-) (NEB) for A-tailing at 37°C for 30 min. Then discard the supernatant
546 and add 7.6 μl H_2O, 10 μl 2 × Quick ligase buffer, 2 μl Quick ligase and 0.4 μl 20 μM Y-
547 adaptor for sequencing adaptor ligation at R.T. for 30min. The library from the beads was then
548 amplified by PCR of 12-14 cycles. After purification with the AMPure XP beads (Beckman,

549 Germany), the libraries were sequenced as double-end 150-bp reads. The sequences of Bridge
550 Linker and Y-adaptor were from the BL-Hi-C paper[19].

551

552 **ChIP and library preparation**

553 Primary myoblasts were fixed with 1% formaldehyde in F-10 at R.T. for 10 min and quenched
554 by adding glycine (F.C. 0.125 M) for 5 min (R.T.). Then the cells were scraped, pelleted, frozen
555 in liquid nitrogen and stored at 80°C until further use. Next, 1×10^6 cells per IP for H3K27ac,
556 2.5×10^6 cells for CTCF, 1×10^7 cells for MyoD were then thawed on ice, resuspended in cold
557 cell lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.5% NP-40) + 1 × EDTA-free Protease
558 Inhibitors. Cells were lysed for 15 min on ice with occasional inversion every 2 min and
559 centrifuged to pellet nucleus. Then the nuclear pellet was resuspended with cold nuclei lysis
560 buffer (50 mM Tris pH8.1, 10 mM EDTA, 1% SDS) + 1 × Protease Inhibitors. Nuclei were
561 lysed for 15 min on ice with occasional inversion every 2 min and then sonicated for 24 cycles
562 with 30s on and 30s off using Bioruptor (Diagenode). After sonication, 10 × volumes of IP
563 dilution buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100 + protease
564 inhibitors) was added, chromatin was precleared using 25 µl Protein A/G magnetic dynabeads
565 (ThermoFisher, Cat.N: 88802/03) / IP for 1 hr at 4°C with rotation. Then chromatin was
566 incubated with the antibody for 6 hrs or overnight at 4°C with rotation (The following
567 antibodies were used for ChIP: MyoD, sc32758; H3K27ac, ab4729; CTCF, CST2899S).
568 Meanwhile, 25 µl beads / IP were blocked with 1 ml PBS/BSA (1 × PBS + 5 mg/ml BSA-
569 fraction V) at R.T. for 3 hrs or overnight at 4°C. Beads were then added to the
570 chromatin/antibody and incubated 4 hrs at 4°C with rotation. Next, beads were washed twice
571 with cold low salt wash buffer (20 mM Tris pH 8.1, 2 mM EDTA, 150 mM NaCl, 1% Triton
572 X-100, 0.1% SDS), twice with high salt wash buffer (20 mM Tris pH 8.1, 2 mM EDTA, 500
573 mM NaCl, 1% Triton X-100, 0.1% SDS), twice with cold LiCl wash buffer (10 mM Tris pH
574 8.1, 1 mM EDTA, 250 mM LiCl, 1% NP- 40, 1% sodium deoxycholate) and once with cold 1
575 × TE buffer (10 mM Tris pH8.0, 1 mM EDTA). DNA: protein complexes were then eluted

576 twice for 10 min at 70°C in 150 µl elution buffer (100 mM $NaHCO_3$, 1 mM EDTA, 1% SDS)
577 each time. 5 M NaCl, RNase A and proteinase K were then added and samples + inputs were
578 reverse cross-linked at 65°C for 6 hrs or overnight and DNA was purified using
579 phenol/chloroform extraction and ethanol precipitation. The library construction was
580 performed using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina.

581

582 **Total RNA library preparation**

583 Total RNA was extracted using TRIZOL (Ambion, USA). The library construction was
584 performed by ANNOROAD and Novagene. Sequencing libraries were constructed using
585 NEBNext UltraTM RNA Library Prep Kit for Illumina (#E7530L, NEB, USA) following the
586 manufacturer's recommendations and index codes were added to attribute sequences to each
587 sample.

588

589 **Library QC and sequencing**

590 Before sequencing, the libraries were quantified by qPCR and the size distribution was assessed
591 using Agilent 2100 Bioanalyzer. Hi-C libraries were sequenced 2 × 150 bp pair-end on the
592 Illumina Nova seq S2 platform. ChIP-seq libraries were sequenced 1 × 50 bp single-end on the
593 Illumina Nextseq550/Nova seq S2 platforms. RNA-seq Libraries were sequenced 2 × 150 bp
594 paired-end run on the Illumina Nova seq S2/HiSeq X10 platforms. All the sequencing was
595 performed by ANNOROAD and Novagene.

596

597 **Recombinant MyoD purification**

598 The PCR-amplified DNA fragments encoding full-length MyoD (residues 1-318) was cloned
599 into a modified pET28a vector with an N-terminal His6-SUMO tag and Ulp1 protease site. The
600 constructed expression vector named pET28a-SUMO-MyoD-FL was transformed into the

601 Escherichia coli strain BL21 (DE3) (Agilent Technologies, Santa Clara, CA, USA). The cells
602 were grown in LB medium supplemented with 50 mg/ml kanamycin at 37°C until OD600
603 reached 0.6-0.8, and then cultured overnight at 18°C after adding 0.2 mM Isopropyl β-D-1-
604 thiogalactopyranoside (J&K) to induce protein expression.

605 The cells were harvested by centrifugation at 5000 rpm (Thermo Fisher Scientific) for 15
606 min. The pellet was resuspended in buffer containing 20 mM Tris-HCl pH 8.0, 500 mM NaCl,
607 25 mM imidazole pH 8.0 and lysed under high pressure via JN-02C cell crusher and further
608 clarified by centrifugation at 17,000 rpm, 60 min, 4℃ (Beckman Coulter). Supernatant was
609 loaded onto 5 ml HisTrap FF columns (GE Healthcare, Beijing, China) pre-equilibrated with
610 Buffer 1 (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM imidazole pH 8.0). The His6-SUMO-
611 MyoD-full-length (FL) was eluted from the column using elution buffer (Buffer2, 20 mM Tris–
612 HCl pH 8.0, 500 mM NaCl, 500 mM imidazole pH 8.0) with a stage-wise gradient on
613 ÄKTA™pure (GE Healthcare). His6-SUMO tags were cleaved by Ulp1 protease during
614 dialysis against buffer S (20 mM Tris-HCl, pH 8.0, 500 mM NaCl) and removed by a second
615 step HisTrap FF column (GE Healthcare). The MyoD-FL protein in flow through was diluted
616 with pre-cooling 20 mM Tris-HCl pH 8.0 to ensure a low NaCl concentration to remove excess
617 nucleic acids by HiTrap SP FF column (GE Healthcare, Beijing, China) with buffer A
618 containing 20 mM Tris-HCl pH 8.0, 200 mM NaCl and buffer B containing 20 mM Tris-HCl
619 pH 8.0, 1 M NaCl. Eluted protein was concentrated by centrifugal ultrafiltration (Millipore
620 Amicon Ultra, 10K), and loaded onto a pre-equilibrated HiLoad Superdex 75 16/60 column
621 (GE Healthcare) in buffer GF (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM DTT) for final
622 purification. All steps should perform on ice or at a low temperature.

623

624 ***In vitro* DNA looping assay and Transmission Electron Microscopy (TEM)**

625 First, plasmid pUC57-MyoD was constructed. pUC57-MyoD contains 10 MyoD binding sites
626 (E-box, CAGCTG) separated by ~2.8 kb of intervening DNA. The intervening DNA was
627 chosen as regions without MyoD occupancy or MyoD binding motifs based on ChIP-seq and

628 motif distribution in muscle cells. The MyoD binding motif (CAGCTG) was validated with
629 EMSA (our previous data). Next, the plasmid was digested with EcoRI (NEB R0101) and gel
630 purified.

631 The ligation assay was carried out as follows. The DNA and proteins were incubated in
632 binding buffer (20 mM Tris 7.9, 50 mM NaCl, 1 mM EDTA) at 25℃ for 20 min with a ratio
633 of 1:100, while the final concentration of DNA is 80 nM and the final concentration of proteins
634 is 8 μM. Then, DNA-protein complexes were purified by gravity-flow gel filtration (4 ml of 2%
635 agarose, ABT E-01508S-2B) using TE buffer (10 mM Tris 7.9, 1 mM EDTA). Peak fractions
636 of DNA-protein complexes were analyzed by TEM. Briefly, samples were fixed with 0.4%
637 glutaraldehyde in TE buffer on ice for 30 min. Then, DNA-protein complexes samples were
638 mixed with a buffer containing spermidine to a final concentration of 2 mM to enhance the
639 absorption of the chromatins to the grids. Samples were loaded to the glow-discharged carbon-
640 coated EM grids and incubated for 2 min and then blotted. The grids were washed stepwise in
641 20 ml baths of 0%, 25%, 50%, 75%, and 100% ethanol solution for 4 min (each at R.T.), air
642 dried and then shadowed with tungsten at an angle of 10° with rotation. Finally, the samples
643 were examined using a FEI Tecnai G2 Spirit 120-kV TEM. Micrographs are shown in reverse
644 contrast. The statistical significance was calculated using the two-tail student's t-test.

645

646 **DNA FISH**

647 DNA FISH probes were constructed according to the bioRxiv preprint (Tn5-FISH)[44].
648 Briefly, the probe library was generated with PCR amplification and recovery by DNA Cleanup
649 kit (D4014, Zymo research, U.S.A.). Then Tn5-FISH probes were amplified by a second PCR
650 with fluorescence-tagged primers. The *in situ* hybridization procedure of Tn5-FISH was similar
651 with traditional FISH as previously described[45]. Briefly, cells were seeded on coverslips and
652 fixed with 4% paraformaldehyde for 15 min, washed with PBS for 5 min, followed by
653 permeabilization with 0.1% Triton X-100/saponin solution for 10 min. Next, cells were washed
654 with PBS for 5 min and incubated in 20% glycerol for 20 min. The cells were then snap freeze-

655 thawed with liquid nitrogen for three times, air-dried and washed by PBS. Cell were further
656 treated with 0.1 M HCl, permeated with 0.5% Triton X-100/saponin and digested with 100
657 μg/mL RNase A for 30 min at 37℃, then balanced in 50% deionized formamide/2 × SSC
658 solution for 30 min. Each color of 10 ng Tn5-FISH probe was mixed with hybridization buffer,
659 then added on the slide and coverslip was placed on top, denatured at 75°C for 5 min, then 37°C
660 overnight. The coverslips were washed with wash buffer, stained with DAPI, and then mounted
661 on slides by AntiFade mounting medium. The slides were then imaged by Zeiss LSM780
662 confocal microscope. The images were processed by FIJI software (version 1.25 h, from NIH)
663 and quantified by Imaris 9.2.0 (Bitplane, Switzerland). The images were then processed with
664 deconvolution software Huygens Pro 17.10. The statistical analyses were achieved with build-
665 in two-tailed t-test models from Graphpad Prism7 software (Graphpad Software, San Diego,
666 U.S.A.).

667 The primer sequences for probe amplification were as follows:

668 *Mybph*-1F CCTACTGCTGACCCTTAAACACTCC,

669 *Mybph*-1R CCAGCCTCCATCCATCACCTCAAG,

670 *Mybph*-2F CTAGTCAAGGCCTCTAGGAGGAG,

671 *Mybph*-2R GATGCACGGTGAGCTCATAGCG,

672 *Mybph*-3F GAAGCCATACTCTTTAACATCAGCCAG,

673 *Mybph*-3R GGGTAGCACATTGTGAATCAGAAAGC,

674 *Mybph*-4F GACACACCATCTCTGGGACTTTCC,

675 *Mybph*-4R AGTTCGAGAAGGTTATCAGCATGGAG;

676 *Myog*-1F GCTACTGCCCTCAACTGTAGC,

677 *Myog*-1R CTGTGTCCTGTGCATTGGGTAC,

678 *Myog*-2F TCTGCCACCTCATTGGCTACAAC,

679 *Myog*-2R GCTCTTAATCATCTCTTCTGTGCT,

680 *Myog*-3F ATCTGGAGTGGTCCTGATGTGG,

681 *Myog*-3R GTGGCTATCCCTAGGTGAAGCC,

682 *Myog*-4F CACTCCTGAGTCTTCCTTACCCA,

683 *Myog*-4R TGAATCCCTGTCTGCAATGGTTTG.

684

685 **Immunofluorescent staining**

686 Cells were fixed in 4% paraformaldehyde in PBS, blocked in 3% BSA and stained according
687 to standard protocols using primary antibodies against MyoD (santa cruz, 1:500), CTCF
688 (millipore, 1:200), Pax7 (DSHB, 1:15), MyoG (DSHB, F5D, 1:500), MHC (DSHB, MF20,
689 1:500) and Alexa Fluor dye conjugated secondary antibodies (zhongshanjinqiao) and DAPI.
690 Secondary antibodies used for dSTORM: Goat anti-mouse Alexa Fluor 488 (Life Technologies,
691 A-11017) and goat anti-rabbit Alexa Fluor 647(Life Technologies A-21246).

692

693 **Direct stochastic optical reconstruction microscopy (dSTORM)**

694 Coverslips (Warner Instruments) with 25 mm 1.5 thickness were used. C2C12 cells were
695 cultured on cleaned coverslips coated with fibronectin (2 μg/mL in PBS, pH=7.4; Sigma).
696 Following immunostaining of MyoD and CTCF (See immunofluoroscent staining section
697 above), cells were blocked with 1 × PBS solution containing 50 mM Glycine for 10 min twice
698 and then washed once in 1 × PBS before being subjected to dual-color imaging in dSTORM
699 imaging buffer (1 × PBS, 100 mM β-mercaptoethylamine (Sigma), pH 7.4) containing
700 Tetraspek beads (Life Technologies), which were used as fiducial markers for x-y drift
701 correction and for overlaying two color images.

702 dSTORM images were acquired on an Olympus IX83 motorized inverted fluorescence
703 microscope equipped with the CellTIRF-4Line system, a 150 × 1.45 NA total internal reflection

704 objective, a back-illuminated EMCCD camera (Andor), the IX3-U-m4TIRSbx filter set
705 (Olympus) and 488 nm (150 mW) and 640 nm (140 mW) lasers. dSTORM images of Alexa
706 488 and Alexa 647 were acquired sequentially, with the acquisition of Alexa 647 preceding the
707 acquisition of Alexa 488. All images were acquired at 10 frames per second. Individual
708 molecules were localized and super-resolution images were reconstructed using the
709 ThunderSTORM analysis module available for ImageJ.

710

711 **RNA-seq data processing and differential gene expression analysis**

712 Paired-end RNA-seq reads were obtained from three biological duplicates of WT and MKO
713 primary myoblasts at proliferation (GM) and early differentiation (DM) stage, with an average
714 read depth as 4.5×10^7 read pairs per sample. HISAT2 v2.1.0[46] was used to align the reads to
715 mm9 genome, then HTSeq v0.6.0[47] was applied to calculate the counts per gene with default
716 parameters. Differential genes were identified with FDR <0.01 and foldchange>1.5 by DESeq2
717 v1.24.0[48].

718

719 **ChIP-seq data analysis**

720 H3K27ac and CTCF ChIP-seq analyses were performed with two replicates in WT and MKO
721 primary myoblasts at proliferation (GM) and early differentiation (DM) stage while MyoD
722 ChIP-seq analyses were in WT myoblasts and myocytes. H3K27ac ChIP-seq analyses were
723 also performed in MKO cells with or without Tacedinaline treatment. ChIP-seq reads were
724 mapped to the mm9 genome with Bowtie2 v2.3.5[49] using default parameters. Aligned reads
725 were filtered with minimum MAPQ of 20, and duplicates were removed using SAMtools v1.9[50].
726 Peaks were called using MACS2 v2.2.5[51] with default parameters for CTCF and MyoD.
727 H3K27ac peaks were called by MACS2 with broad peak mode. Signal tracks were generated
728 by using the -SPMR option in MACS2. Then, the UCSC Genome Browser utility[52]
729 bedGraphToBigWig was used to transform the bedgraph files to bigwig files. Differential Peaks

730 of ChIP-seq experiments were called with R package DiffBind v2.10.0[53] with default settings.
731 Heatmaps of ChIP-seq signal enrichment were generated by python package deeptools v3.0.1[54].
732 ChIP-seq peak annotation were done by ChIPseeker v1.20.0[55]. Public ChIP-seq data used in
733 this study were processed in the same manner. Specially, we merged our detected MyoD peaks
734 with those called from public MyoD ChIP-seq data (GSE56131). Then, the differential peaks
735 were determined by examining the MyoD binding strength on merged peaks comparing our
736 WT-DM and WT-GM Myod ChIP-seq data based on DiffBind v2.10.0. WT-GM pseudo-peaks
737 combined static MyoD peaks and MyoD peaks with reduced signal in WT-DM while WT-DM
738 pseudo-peaks combined static MyoD peaks and MyoD peaks with enhanced signal in WT-DM
739 cells.

740

741 **Motif enrichment analysis**

742 The homer[56] script findMotifsGenome was used with default parameters for enriching the motif
743 of ATAC-seq peaks overlapped with loop anchors in WT primary myoblasts. The homer script
744 findPeaks with option –nfr was used to find the nucleosome free regions in H3K27ac peaks.
745 Then, these nucleosome free regions in broad H3K27ac peaks were used to call motif by
746 findMotifsGenome with size as 200bp.

747

748 **Hi-C data processing, Hi-C loop calling and differential loop detection**

749 BL-Hi-C data were processed for WT and MKO primary myoblasts at proliferation (GM) and
750 early differentiation (DM) stage as well as MKO cells with or without Tacedinaline treatment.
751 Non-redundant uniquely mapped contacts were generated utilizing the in house HiCpipe
752 pipeline, which trims the bridge linker, aligns reads, filters artifact fragments, and removes
753 duplicates. Juicer[26] was used for following Hi-C map generation, compartment partition and
754 loop calling. For every 100 kb bin, A or B compartments were defined by the over 70%
755 replicates majority rule. Contact domain boundaries (CDBs) and their relative insulation scores

756 were calculated by HiCDB while insulation scores[57] were calculated on MyoD or CTCF bound
757 CDBs with insulation score method (parameters as -is 1000000 -ids 200000 -im mean -nt 0.1).
758 Differential CDBs were defined with a t-test FDR<0.01 and a difference between cases and
759 controls higher than 80% quantile of the overall difference. The domain score[38] (D-score) for
760 each loop was calculated by dividing the intra-loop interactions with all interactions connected
761 to the corresponding loop. Loops with differential internal interactions (D-score) were
762 determined with a t-test FDR<0.01 and fold change higher than 80% quantile of the overall fold
763 changes.

764 Loops were called under 3 kb, 5 kb and 10 kb resolution with HiCCUPS using parameters
765 (-p 6,4,2 -i 10,7,5) and scaled to 5 kb and 10 kb resolution. These parameters were also used to
766 determine the loop strength of requested loops with HiCCUPS. Loop strength was calculated
767 as relative Hi-C contacts at loop pixels compared with surrounding Donut-shaped background
768 pixels. The differential loop detection method was adapted from Douglas et al.[40] which
769 depended on DESeq2 to detect differential enrichment of the loop over background across
770 condition. For comparing loops in two conditions, loops were called on combined BL-Hi-C
771 matrixes for each condition and merged together at first. Then, the loop pixels and their
772 surrounding pixels were collected according to the previous work from each biological replicate
773 at both 5 kb and 10 kb resolution. As previous study described bias[39] based on contact distance
774 associated with the identification of differential interactions from Hi-C-like data, loops were
775 split into two distance regimes of greater or less than 150 kb for 5 kb and 10 kb resolution.
776 Differential loops called from each set were thresholded by FDR as 0.1 and were combined to
777 form the final set of dynamic loops. For dynamic loops detected under both 5 kb and 10 kb
778 resolution, 5 kb resolution loops were kept. Public *in situ* Hi-C data of *in vivo* ES and CN were
779 processed by HiC-Pro v2.11.1[58] in the mapping and filtering step and then treated with the same
780 processing framework as our BL-Hi-C data.

781

782 **Loop annotation, aggregation and functional analysis**

783 Loop anchors (scaled to 10 kb) were overlapped with MyoD and CTCF binding sites and
784 classified into four classes using bedtools toolset v2.27.1[59]. MyoD or CTCF occupancy on loops
785 were also determined by intersecting loop anchors and MyoD or CTCF peaks. Aggregate peak
786 analysis (APA) plots were generated to assess the quality of loop detection and explore the
787 characteristics of different loop classes by Juicer APA command. The aggregated maps
788 generated by Juicer were normalized by the actual supporting loop number and the cis
789 interaction pairs of each sample. The APA score was also calculated by Juicer APA command
790 as the ratio of Hi-C contacts at the central pixel to the mean Hi-C contacts of the lower left
791 pixels. Enrichment plots for differential CDBs and loops with differential internal interactions
792 were generated with coolpup.py[60]. To gain knowledge about the function of different loop types,
793 the loop anchors were analyzed by GREAT v3.0.0[61] with the association rule: single nearest
794 gene and 100kb max extension to identify the enriched biological process. Specially, for loop
795 internal interaction analysis, we examined the genes enclosed in the loops and the genes were
796 assigned to their most dynamic loops to get the explicit gene-loop pair. In other sections, genes
797 associated to loops were determined if their promoters (±3 kb around TSS) overlapped loop
798 anchors. To measure the muscle lineage specificity of genes associated with MyoD-bound and
799 CTCF-bound loops, we calculated the Z-score of gene expression in WT-DM cells compared
800 with other ENCODE collected cell types (embryonic stem cell, spleen, B cell, T cell,
801 megakaryocyte, neural progenitor and cortical neuron). ClusterProfiler v3.12.0[62] was used to
802 enrich and compare GO terms of different sets of loops bound by CTCF or MyoD.

803

804 **Visualization**

805 Tracks of Hi-C maps and ChIP-seq data were generated by pyGenomeTracks v3.1.2[63]. Hi-C
806 maps of each condition were normalized by cis interaction pairs.

807

808 **Statistics**

809 Analysis-specific statistics were applied as described in each subsection. For comparisons of
810 distributions, Wilcoxon rank sum test was employed in R. The statistical significance of the
811 other data was calculated using the two-tail student's t-test. The following annotation applies
812 for all figures: *P < 0.05, **P < 0.01, ***P < 0.001.

813

814 **Code availability**

815 Custom scripts described in the Online Methods will be made available upon request.

816

817 **Data availability.**

818 The raw sequence data reported in this paper have been deposited in the Genome Sequence
819 Archive (GSA)[64] under accession number CRA002490 that are publicly accessible at
820 https://bigd.big.ac.cn/gsa. The processed data of this paper have been deposited in the Gene
821 Expression Omnibus (GEO) database under the accession number: GSE157339. Accession
822 codes of the published data in GEO used in this study are as follows: MyoD ChIP-seq of wild-
823 type primary myoblasts, GSE56131; ATAC-seq of wild-type primary myoblasts, GSE63573;
824 NeuroD2 ChIP-seq of embryonic cortical neuron cells, GSE67539; Hi-C data of mouse neural
825 development as well as CTCF and H3K27ac ChIP-seq of embryonic cortical neuron cells,
826 GSE96107; SMC3 ChIP-seq of C2C12 myoblast and myotube, GSE113248.

827

828 **References**

829

830 44 Zhang, X. *et al.* Imaging chromatin interactions at sub-kilobase resolution Via
831 Tn5-FISH. *BioRxiv* (2019).
832 45 Bayani, J. & Squire, J. A. Fluorescence in situ Hybridization (FISH). *Current*
833 *protocols in cell biology* **Chapter 22**, Unit 22.24 (2004).
834 46 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based
835 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat*
836 *Biotechnol* **37**, 907-915 (2019).

47 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

48 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

49 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

50 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

51 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

52 Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Briefings in bioinformatics* **14**, 144-161 (2013).

53 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393 (2012).

54 Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).

55 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).

56 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).

57 Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-579 (2016).

58 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015).

59 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

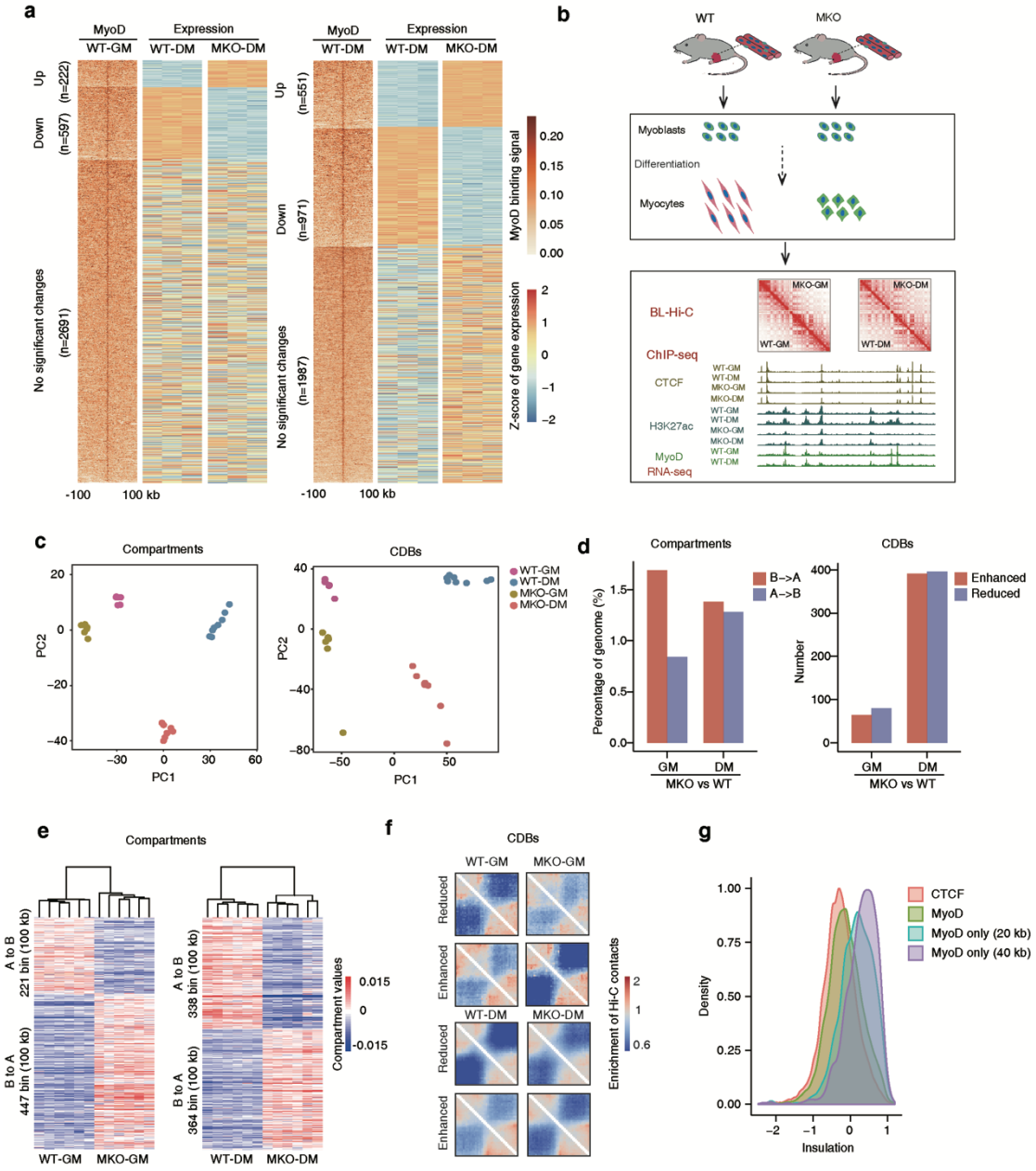
60 Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, 2980-2985 (2020).

61 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501 (2010).

62 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-287 (2012).

63 Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189 (2018).

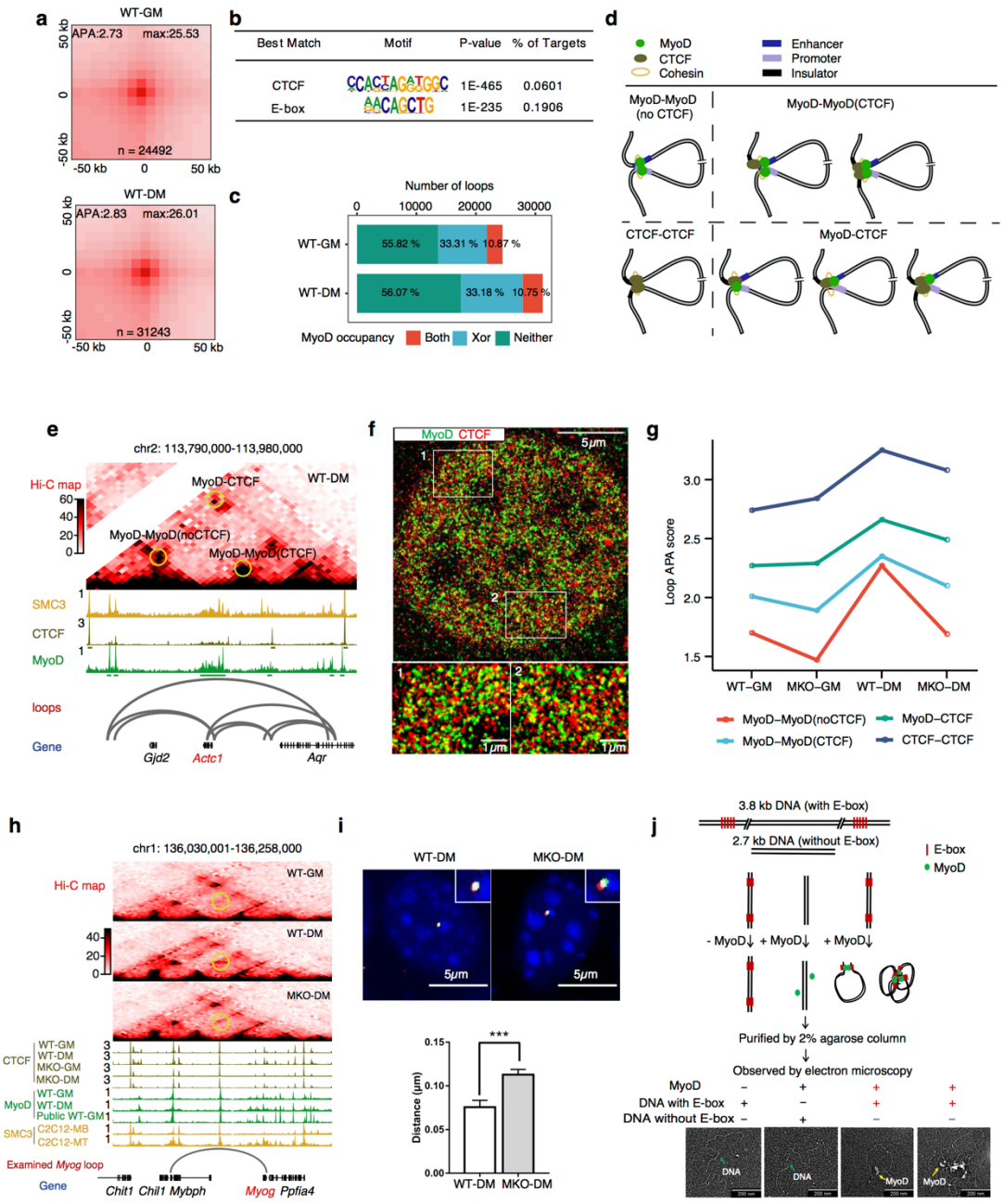
64 Wang, Y. *et al.* GSA: Genome Sequence Archive<sup/>. *Genomics, proteomics & bioinformatics* **15**, 14-18 (2017).



878

879 **Fig. 1| MyoD regulates A/B compartments switching and formation of contact domain**
880 **boundaries (CDBs) in muscle cells. a**, Association of MyoD binding peaks at promoter
881 regions and related gene expression trends. Heatmaps showing MyoD binding peaks at
882 promoters (centralized at the TSS within 3 kb distance) and related gene expression trends in
883 WT and MKO myoblasts was shown at left panel. Corresponding data for WT and MKO
884 myocytes was shown in the right panel. Pseudo-peaks of MyoD binding with combined ChIP-
885 seq data from public databases and our study were used for this analysis (Methods). **b**, Strategy
886 for genetic investigation of the chromatin architectural roles of MyoD. Briefly, muscle stem

887 cells isolated from WT or MKO mice were cultured in GM and induced to differentiation in
888 DM for 24 hrs. The proliferating myoblasts and differentiated myocytes were subjected to BL-
889 Hi-C analysis. To correlate 3D genome architecture to chromatin state and to gene expression
890 levels, multiple ChIP-seq and RNA-seq analyses were performed on myoblasts and myocytes
891 from WT and MKO mice. **c**, Principal Component Analysis (PCA) of the values of
892 compartments or the relative insulation scores of contact domain boundaries (CDBs) among
893 the four indicated sample types. Each dot represents an individual biological replicate. **d**,
894 Barplots showing the percentage of A/B compartment shifts between WT and MKO in
895 myoblasts and myocytes (left panel). Barplots showing the number of relative insulation
896 enhanced or reduced CDBs between WT and MKO in proliferating and differentiating muscle
897 stem cells (right panel). **e**, Heatmaps of the A/B compartment shifted regions between WT and
898 MKO in myoblasts and myocytes. **f**, Hi-C aggregation plots centered at differential CDBs
899 between WT and MKO in myoblasts and myocytes. **g**, Probability density of insulation score
900 at genetic regions bound by CTCF or MyoD. Four classes were assessed: CTCF-bound, MyoD-
901 bound, MyoD-bound and CTCF-unbound within 20 kb distance, MyoD-bound and CTCF-
902 unbound within 40 kb distance.



903

904 **Fig. 2| MyoD is an anchor protein for chromatin loop formation in muscle cells. a**,
905 Aggregate Peak Analysis (APA) plots showing the aggregated Hi-C contacts around chromatin
906 loops identified in WT-GM and WT-DM cells. n represents the total number of chromatin loops
907 called by HiCCUPS (Methods). APA score at upper left corner of each aggregation plot was
908 calculated as the ratio of Hi-C contacts at the central pixel to the mean Hi-C contacts of the
909 lower left pixels. The Hi-C contacts at the central pixel was shown at upper right corner of each
910 aggregation plot. **b**, Enriched motifs in accessible regions at anchors of the chromatin loops
911 identified in panel a. **c**, Percentage of chromatin loops with anchors bound by MyoD in WT-

912 GM and WT-DM cells. MyoD bound at both anchors (Both, red), one of two anchors (Xor,
913 blue), or neither of two anchors (Neither, green). Pseudo-peaks of MyoD binding with
914 combined ChIP-seq data from public databases and our study were used for this analysis
915 (Methods). **d**, Classification of the four types of chromatin loops identified in the study. MyoD-
916 MyoD(noCTCF) loops had both anchors bound only by MyoD. MyoD-MyoD(CTCF) loops
917 had both anchors bound by MyoD and either anchor concomitantly bound by CTCF. MyoD-
918 CTCF loops had only one anchor bound by MyoD and either anchor bound with CTCF. CTCF-
919 CTCF loops had both anchors bound only by CTCF. **e**, Representative examples of the MyoD-
920 bound chromatin loops. ChIP-seq of SMC3 were in C2C12 cells referenced from public
921 database (GSE113248). **f**, Representative dSTORM image showing co-localization of MyoD
922 (green) and CTCF (red) from immunofluorescent staining of C2C12 cells grown in DM for 24
923 hrs with antibodies against MyoD and CTCF. **g**, APA score dynamics for the four types of
924 chromatin loops among WT and MKO cells in GM and DM stages. **h**, A representative MyoD-
925 bound chromatin loop at the *Myog* locus that presents the Hi-C contact map depicting
926 normalized contact frequencies and ChIP-seq signal tracks. The examined chromatin loop with
927 reduced Hi-C contacts in MKO-DM sample (compared with WT-DM sample) was marked by
928 yellow circle. **i**, Chromatin loop at the *Myog* locus was validated by 3D-DNA fluorescent *in*
929 *situ* hybridization (3D-DNA FISH) in WT-DM and MKO-DM cells. The upper panel showed
930 *Myog* promoter (red) and enhancer (green) were more separated in the MKO cells (right) than
931 in the WT cells (left). The lower panel showed the average distances of *Myog* loop anchors.
932 ***$P < 0.001$, two-tail student's t-test. **j**, MyoD mediates chromatin looping *in vitro*. The upper
933 panel presents a schematic for the *in vitro* DNA circularization assay used to detect the ability
934 of MyoD (green) to mediate DNA looping with E-box (red) containing DNA (3.8 kb) and no
935 E-box control (2.7 kb). The lower panel presents representative views of circularized DNA
936 mediated by recombinant MyoD protein, imaged by Transmission Electron Microscopy (TEM).

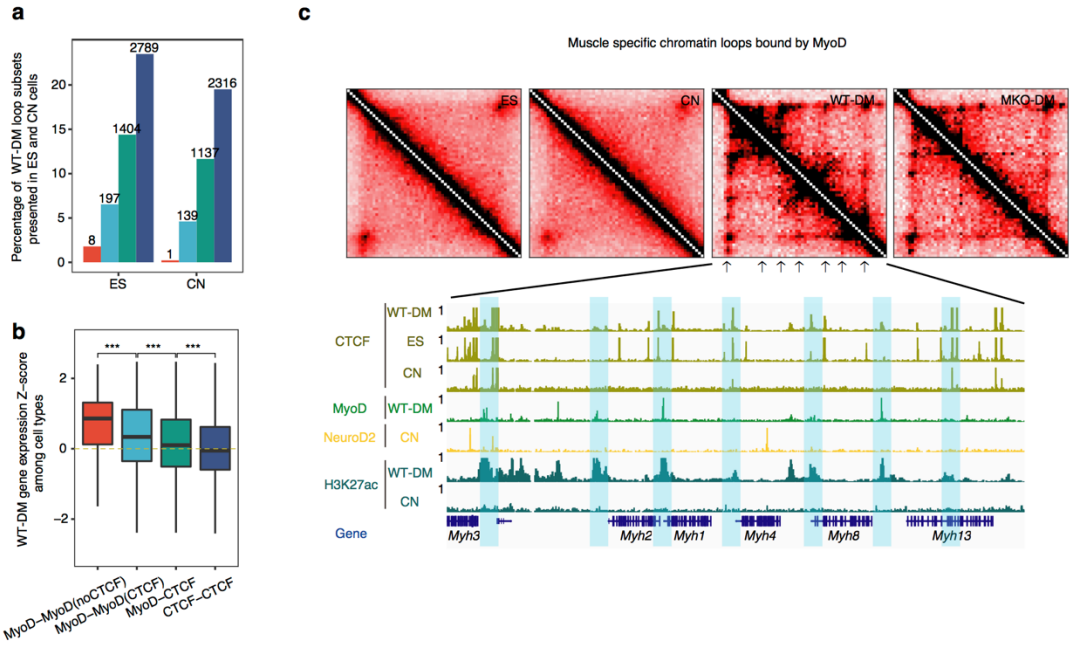


Fig. 3| MyoD anchors formation of myogenic-specific chromatin loops. a, Proportions of the muscle chromatin loop subsets presented in embryonic stem cell (ES) or cortical neuron (CN). The numbers of muscle chromatin loops found in ES or CN cells were indicated for each bar. **b**, Boxplots represent Z-score of gene expression in WT-DM cells compared with other ENCODE collected cell types (embryonic stem cell, spleen, B cell, T cell, megakaryocyte, neural progenitor and cortical neuron) for genes associated with each class of loops, showing MyoD-bound loops regulate lineage specific gene expression in muscle cells. ***P < 0.001, Wilcoxon rank sum test. **c**, Representative region showing muscle cell specific loops anchored on MyoD binding regions at *Myh* gene cluster.

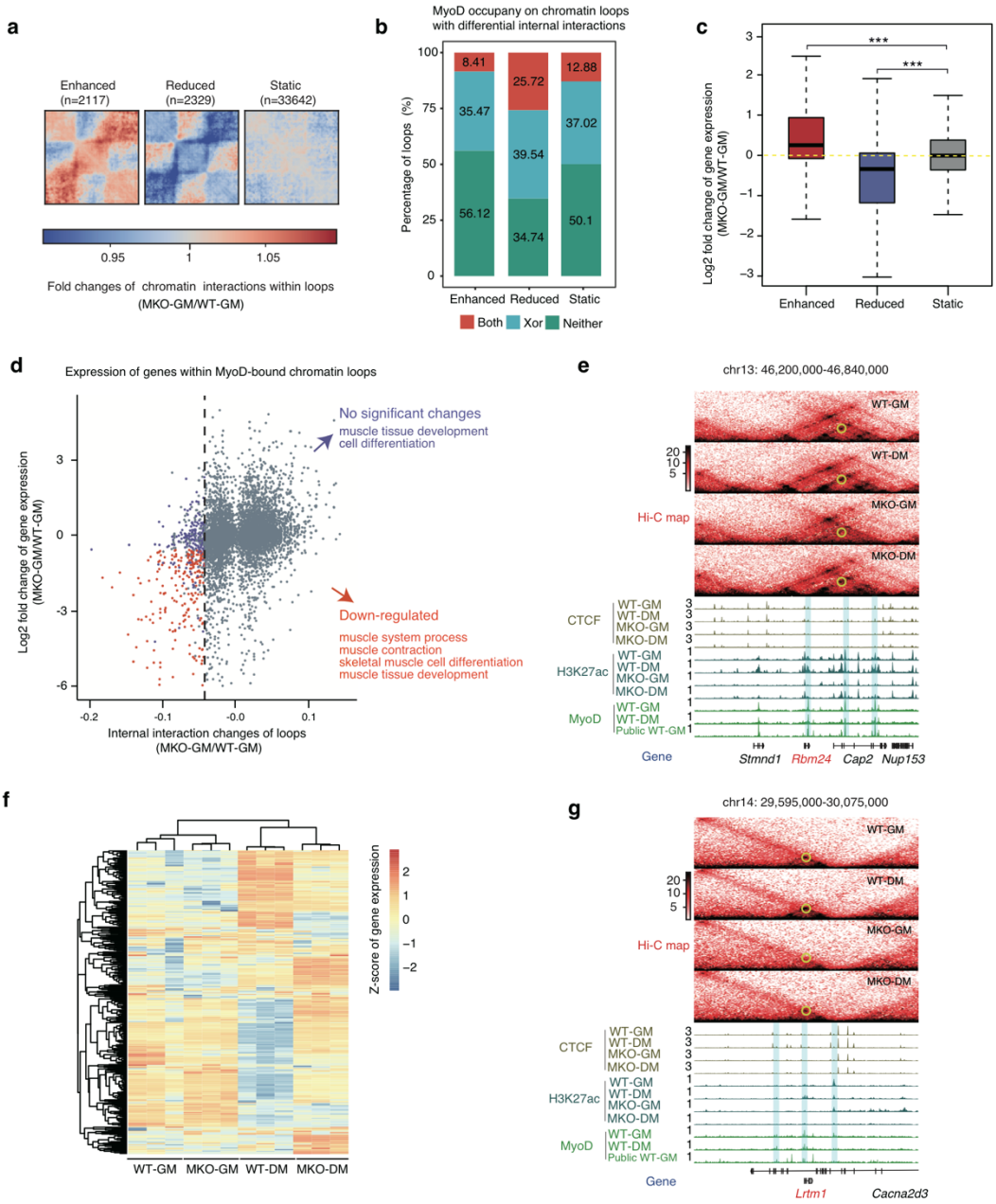
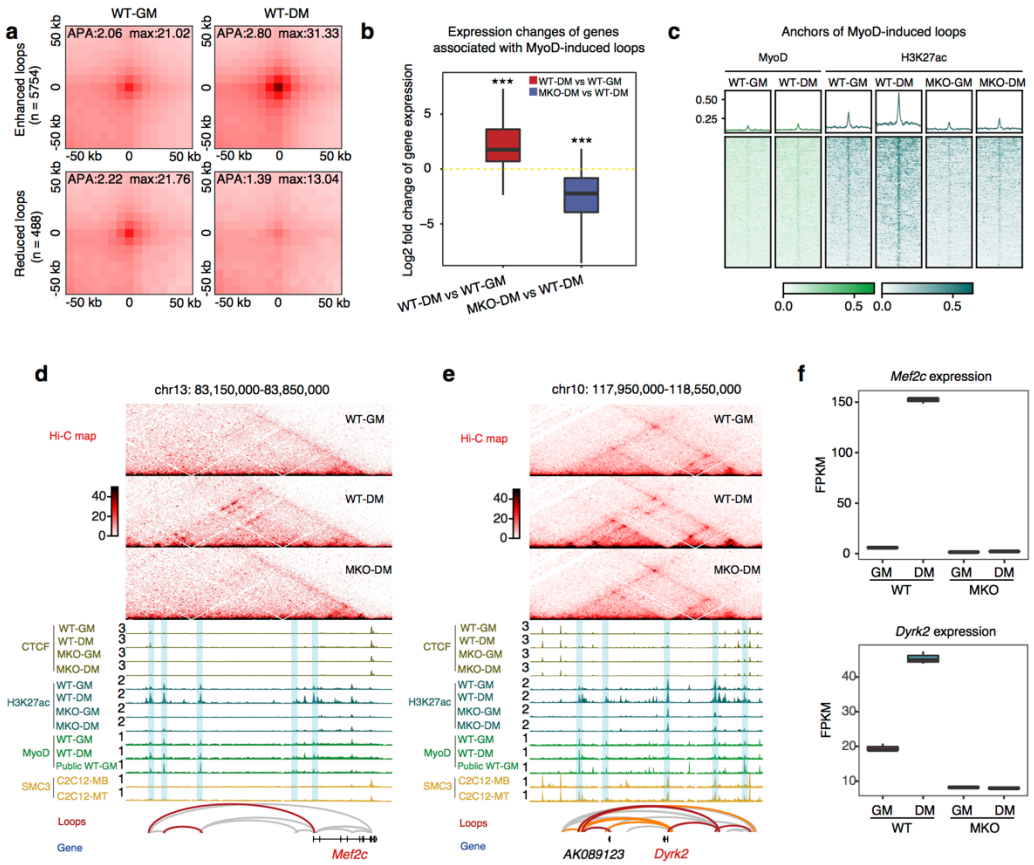


Fig. 4| MyoD mediates primed architectural chromatin loops in proliferating muscle cells. **a**, Aggregated Hi-C maps around chromatin loops with enhanced, reduced, or static internal interactions (measure by domain score) between WT and MKO myoblasts (MKO-GM vs. WT-GM). **b**, Percentage of chromatin loops with differential internal interactions (MKO-GM vs. WT-GM) that occupied by MyoD at anchors. MyoD bound at both anchors (Both, red), one of two anchors (Xor, blue), or neither of two anchors (Neither, green). Pseudo-peaks of MyoD binding with combined ChIP-seq data from public datasets and our study were used for this analysis (Methods). **c**, Boxplot showing the corresponding expression changes of genes within

956 chromatin loops having enhanced, reduced, or static internal interactions (domain score)
957 identified in panel b (MKO-GM vs. WT-GM). Considering that one loop may contain several
958 genes and domain scores of nested loops are associated, we assigned each gene to its most
959 dynamic nested loop. ***P < 0.001, Wilcoxon rank sum test. **d**, Scatter plot showing both the
960 domain score fold changes and gene expression fold changes of MyoD-bound chromatin loops
961 between WT-GM and MKO-GM cells. Internal interaction reduced MyoD-bound chromatin
962 loops without significantly differentially expressed genes in MKO-GM cells compared with
963 WT-GM cells were shown as blue dots, while internal interaction reduced MyoD-bound
964 chromatin loops with significantly down-regulated genes in MKO-GM cells compared with
965 WT-GM cells were shown as red dots. **e**, The *Rbm24* locus as a representative for an internal
966 interaction reduced MyoD-bound chromatin loop with significantly down-regulated genes in
967 MKO-GM cells compared with WT-GM cells. **f**, Heatmap showing expression (FPKM) Z-
968 scores among four indicated sample types of genes within internal interaction reduced MyoD-
969 bound chromatin that do not contain differentially expressed genes in MKO-GM cells
970 compared with WT-GM cells. **g**, The *Lmrt1* locus as a representative for an internal interaction
971 reduced MyoD-bound chromatin loop in MKO-GM cells compared with WT-GM cells that is
972 primed in proliferating muscle cells.



973

974 **Fig. 5| The regulatory loops specified by MyoD are functionally required for muscle cell**
975 **differentiation. a**, APA plots showing the aggregated Hi-C contacts around enhanced (n=5754)
976 and reduced (n=488) chromatin loops in WT-DM cells compared with WT-GM cells.
977 Differential loops were identified based on DESeq2 (Methods). **b**, Boxplot showing expression
978 changes of genes associated with MyoD-induced chromatin loops in WT-DM cells compared
979 with WT-GM cells (left) and in MKO-DM cells compared with WT-DM cells (right). ***P <
980 0.001, Wilcoxon rank sum test. **c**, MyoD and H3K27ac ChIP-seq signal at anchors of MyoD-
981 induced loops. Heatmaps showing signal enrichment centralized at the anchors within ± 100 kb
982 genomic region. **d** and **e**, The *Mef2c* and *Dyrk2* locus as representatives of MyoD-induced loops
983 during muscle cell differentiation. MyoD-induced loops detected by our algorithm were
984 presented in red in the loop track while differentiation-induced loops were in orange. Other
985 loops were presented in grey. **f**, *Mef2c* and *Dyrk2* gene expression among WT and MKO cells
986 in proliferating and differentiating stage.

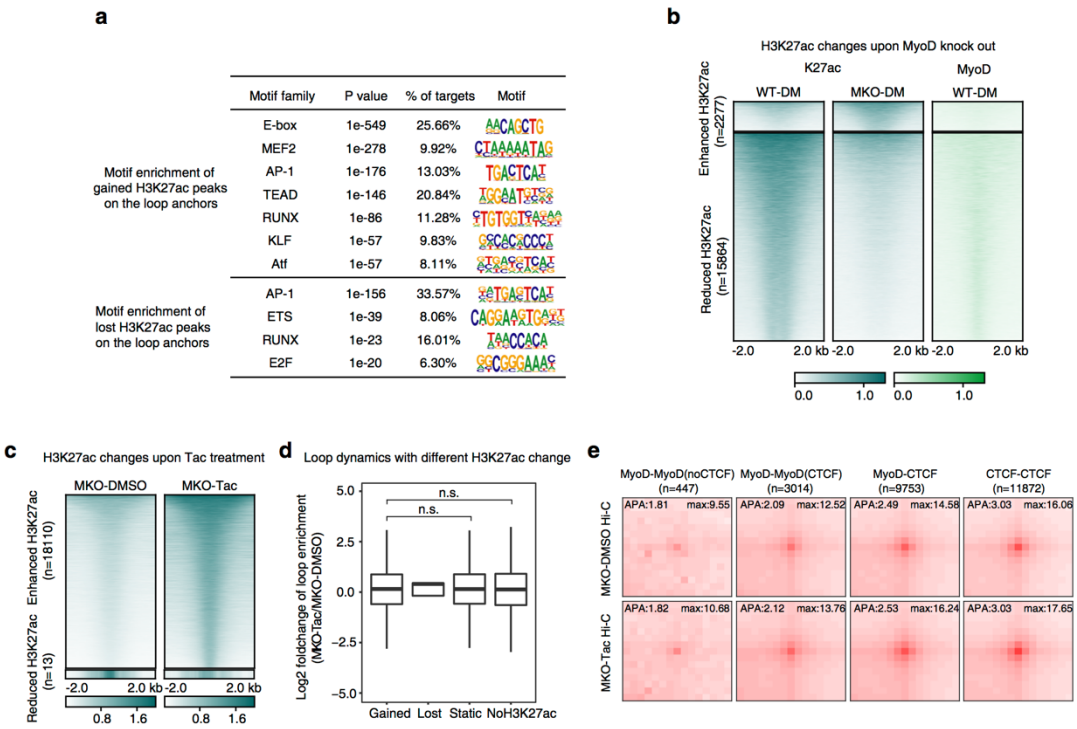
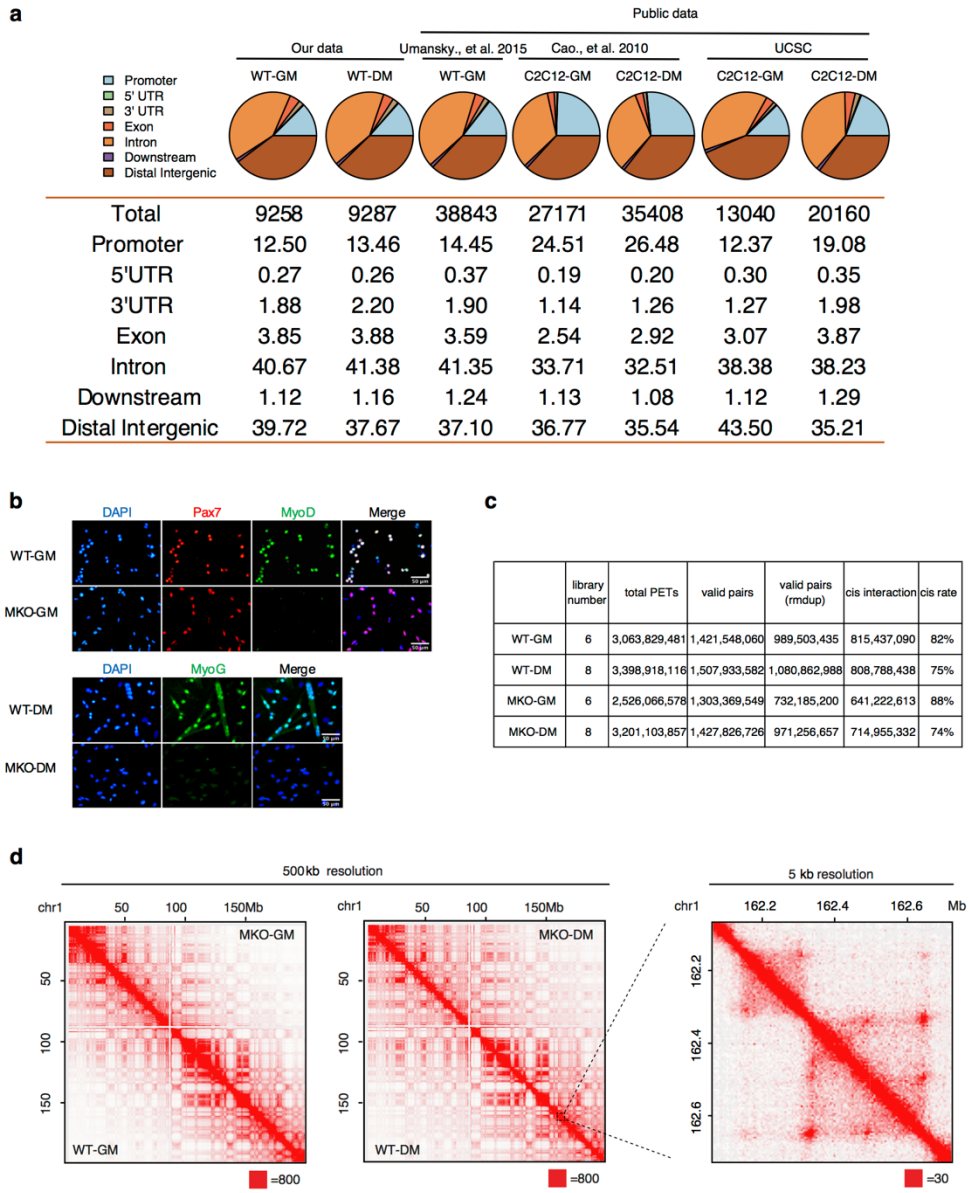


Fig. 6| H3K27ac in the MKO cells is not sufficient to drive formation of the MyoD-bound chromatin loops. a, Motif enrichment at gained or lost H3K27ac peaks on the loop anchors in WT-DM cells compared with WT-GM cells. **b**, Heatmaps showing enhanced and reduced H3K27ac peaks from ChIP-seq analysis in MKO-DM cells compared with WT-DM cells. MyoD binding peaks at those differential H3K27ac peaks were also shown. **c**, Heatmaps showing enhanced and reduced H3K27ac peaks from ChIP-seq analysis in MKO cells treated with a selective class I HDAC inhibitor Tacedinaline (Tac) in differentiation medium for 12 hrs. DMSO treated MKO cells served as control. **d**, Loop strength fold changes of the loops with gained, lost, static or no H3K27ac marks on their anchors between MKO cells treated with Tac or DMSO. Loop strength was defined as relative Hi-C contacts at loop pixels compared with surrounding Donut-shaped background pixels (Methods). ***P < 0.001, Wilcoxon rank sum test. **e**, APA plots showing aggregated Hi-C contacts around the four types of chromatin loops (classified in 2d) in MKO cells treated with Tac and DMSO control.

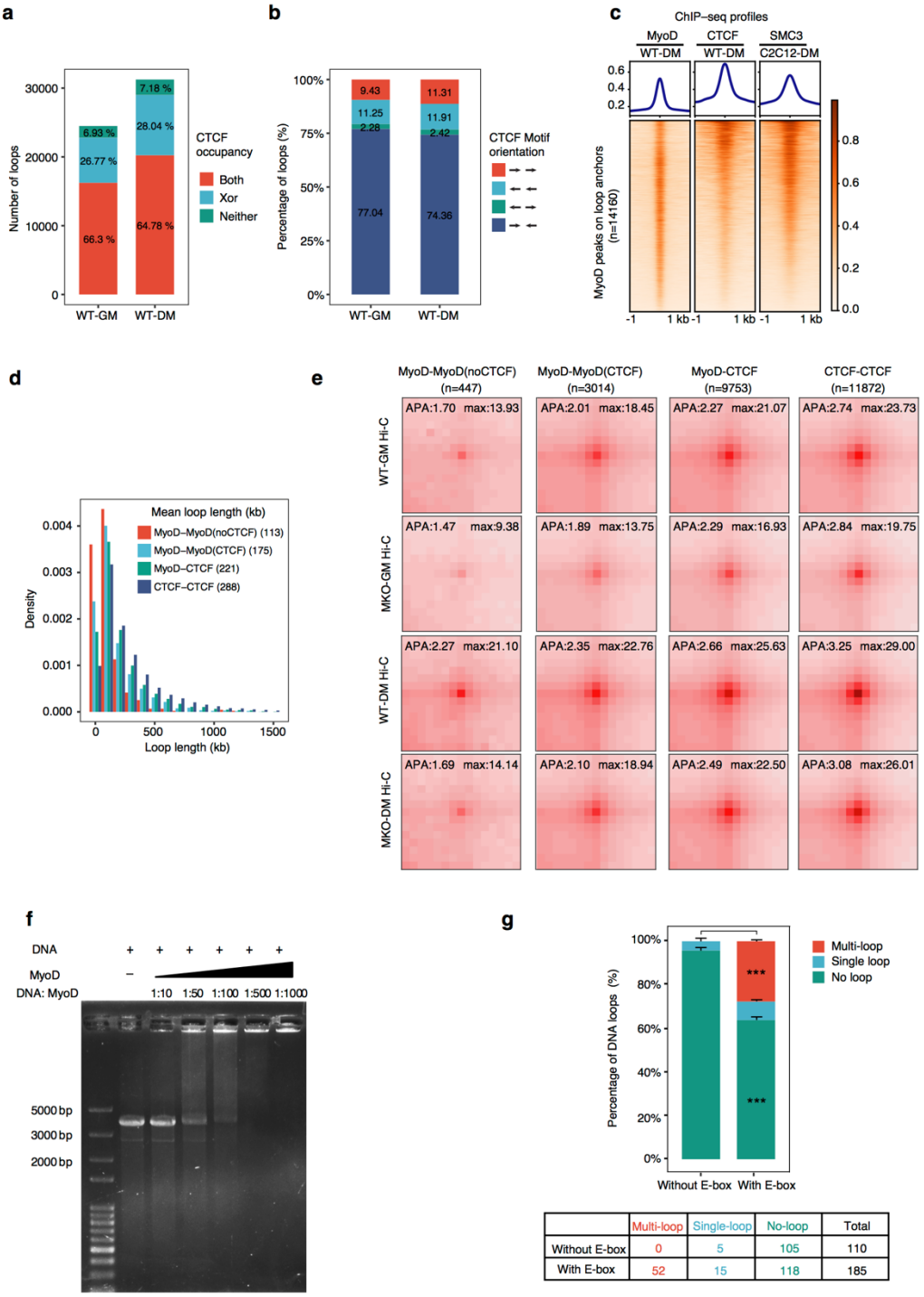


	WT-GM	WT-DM	WT-GM	C2C12-GM	C2C12-DM	C2C12-GM	C2C12-DM
Total	9258	9287	38843	27171	35408	13040	20160
Promoter	12.50	13.46	14.45	24.51	26.48	12.37	19.08
5'UTR	0.27	0.26	0.37	0.19	0.20	0.30	0.35
3'UTR	1.88	2.20	1.90	1.14	1.26	1.27	1.98
Exon	3.85	3.88	3.59	2.54	2.92	3.07	3.87
Intron	40.67	41.38	41.35	33.71	32.51	38.38	38.23
Downstream	1.12	1.16	1.24	1.13	1.08	1.12	1.29
Distal Intergenic	39.72	37.67	37.10	36.77	35.54	43.50	35.21

	library number	total PETs	valid pairs	valid pairs (rmdup)	cis interaction	cis rate
WT-GM	6	3,063,829,481	1,421,548,060	989,503,435	815,437,090	82%
WT-DM	8	3,398,918,116	1,507,933,582	1,080,862,988	808,788,438	75%
MKO-GM	6	2,526,066,578	1,303,369,549	732,185,200	641,222,613	88%
MKO-DM	8	3,201,103,857	1,427,826,726	971,256,657	714,955,332	74%

Extended Data Fig. 1| Unveiled genome architectural functions of MyoD in muscle cells. **a**, Genomic distribution of MyoD binding peaks in proliferating or differentiating muscle cells. WT represents primary muscle cells while C2C12 is a skeletal muscle cell line. MyoD ChIP-seq data were from our study and public datasets of indicated references. Promoter is defined as ± 3 kb from the transcription start site (TSS). **b**, Representative images for immunofluorescent staining of Pax7 (red), MyoD (green) and MyoG (green) in wild type (WT) and MyoD knockout (MKO) proliferating muscle stem cells (myoblasts, cultured in growth media, GM) was shown at upper panel. Corresponding data for differentiating muscle stem cells (myocytes, induced to differentiation in differentiation media, DM) was shown at lower

1011 panel. DAPI (blue) served to visualize nucleus. Scale bars represent 50 μm. **c**, Depth and quality
1012 of sequencing data for BL-Hi-C libraries on four types of samples (WT-GM, WT-DM, MKO-
1013 GM and MKO-DM). **d**, Hi-C map showing BL-Hi-C data in our study reached 5 kb resolution.
1014 Left heatmaps showing chromatin contact matrices from the whole chromosome 1 (chr1) at 500
1015 kb resolution in each of four cell samples. Right panel was a zoom-in heatmap for chr1: 162 ~
1016 162.8 Mb at 5 kb resolution. Maximum intensity was indicated in the lower right of each panel.

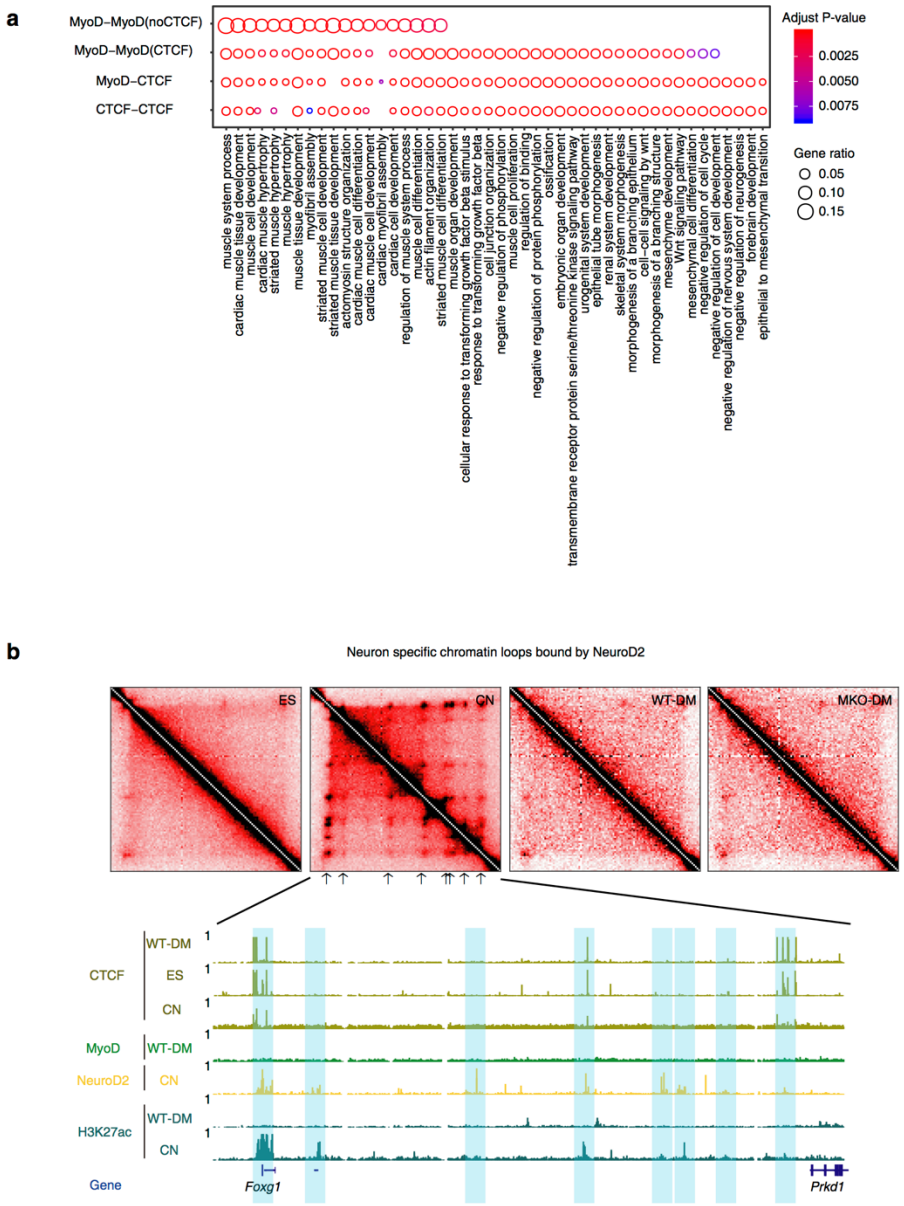


	Multi-loop	Single-loop	No-loop	Total
Without E-box	0	5	105	110
With E-box	52	15	118	185

1017

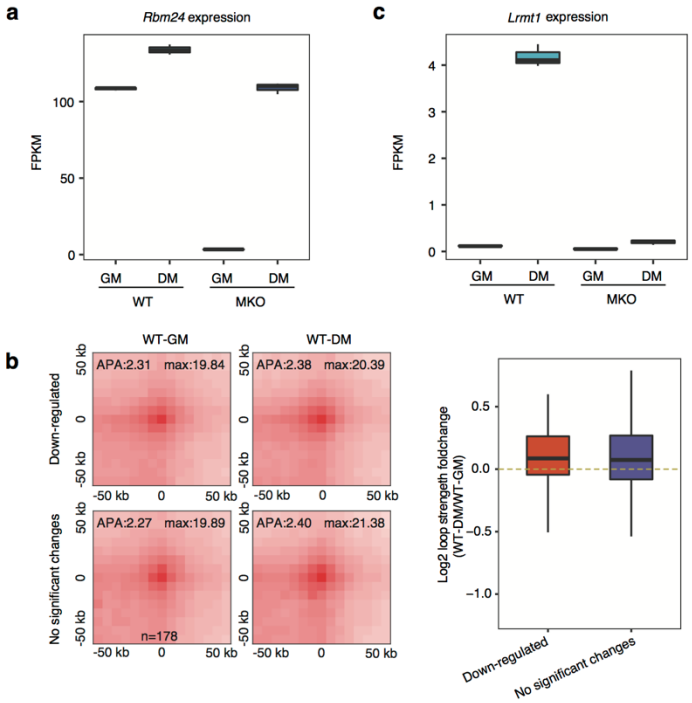
1018 **Extended Data Fig. 2| MyoD mediates chromatin loop formation *in vivo* and *in vitro*. a**,
1019 Percentages of chromatin loops with CTCF bound at both anchors (Both, red), one of two
1020 anchors (Xor, blue) or neither of two anchors (Neither, green) in WT-GM and WT-DM cells.
1021 **b**, Distribution CTCF motif orientation on loop anchors bound by CTCF in WT-GM and WT-
1022 DM cells. **c**, Heatmaps of ChIP-seq signal showing the enrichment of MyoD, CTCF and SMC3

1023 on MyoD-bound loop anchors, indicating the concordant binding of MyoD with CTCF and
1024 SMC3 on these loop anchors. MyoD and CTCF ChIP-seq were conducted in our study. ChIP-
1025 seq for SMC3 in C2C12 cells were referenced from public database (GSE113248). **d**, Loop
1026 length distribution of MyoD-MyoD(noCTCF), MyoD-MyoD(CTCF), MyoD-CTCF and
1027 CTCF-CTCF chromatin loops. The average loop length was shown at upper right corner. **e**,
1028 APA plot showed the aggregated Hi-C contacts around MyoD-bound three types of loops and
1029 CTCF-CTCF chromatin loops in MKO cells compared to WT cells. All four type of samples
1030 from GM and DM stages were shown. n represents number of each type of chromatin loops. **f**,
1031 Free and the recombinant MyoD protein-bound DNA fragment (3.7kb, with E-box) were shown
1032 and separated by polyacrylamide gel electrophoresis (PAGE). The binding of MyoD and DNA
1033 fragment also showed ratio-dependency. **g**, Statistics of the percentage and number of loops
1034 observed with DNA circularization in Fig. 2h. ***$P < 0.001$, two-tail student's t-test.

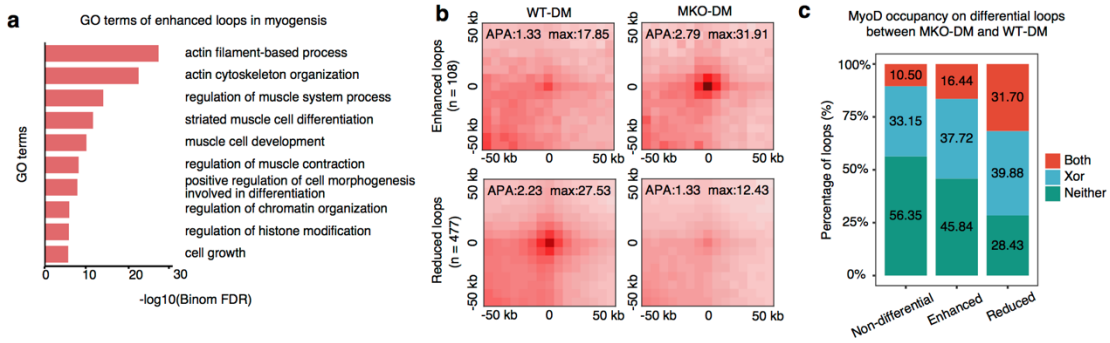


1035

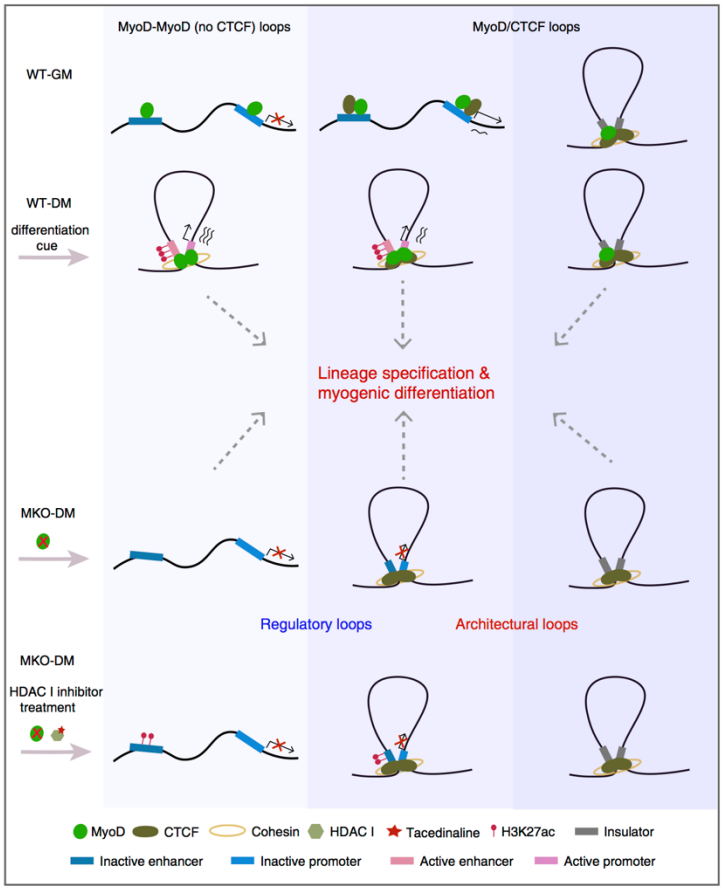
1036 **Extended Data Fig. 3| MyoD and NeuroD2 instructs lineage specific chromatin loops. a**,
1037 Enriched GO terms of genes with promoter overlapped with anchors of MyoD-
1038 MyoD(noCTCF), MyoD-MyoD(CTCF), MyoD-CTCF and CTCF-CTCF chromatin loops. **b**,
1039 Representative region of neurogenic lineage specific chromatin looping at *Foxg1* locus bound
1040 by NeuroD2.



Extended Data Fig. 4| MyoD mediates primed architectural loops in myoblasts. a, Boxplot showing the gene expression (FPKM) of *Rbm24* in four types of samples (WT-GM, WT-DM, MKO-GM and MKO-DM). **b,** Loop changes of MyoD-bound architectural loops during differentiation. APA plot showed the aggregated Hi-C contacts around internal interaction reduced MyoD-bound chromatin loops with down-regulated genes or not significantly changed genes in WT-GM and WT-DM cells. Boxplot showed the loop strength fold changes for loops with down-regulated genes or not significantly changed genes between WT-DM and WT-GM cells. **c,** Boxplot showing the gene expression (FPKM) of *Lrmt1* among four types of samples (WT-GM, WT-DM, MKO-GM and MKO-DM).



Extended Data Fig. 5| MyoD-induced regulatory loops are functionally required for muscle cell differentiation. a, Enriched GO terms of enhanced loops in WT-DM cells compared with WT-GM cells. **b,** APA plots showing aggregated Hi-C contacts around significantly reduced or enhanced chromatin loops in MKO-DM cells compared to WT-DM cells. n represents number of the differential chromatin loops between MKO-DM and WT-DM cells. Differential loops were identified with DESeq2 (Methods). **c,** Percentages of MyoD-bound loops in the non-differential, enhanced or reduced chromatin loops described in panel b. MyoD bound at both anchors (Both, red), one of two anchors (Xor, blue) or neither of two anchors (Neither, green) were shown. Pseudo-peaks of MyoD binding with combined ChIP-seq data from public and our study were used for this analysis.



Extended Data Fig. 6| Hypothetical model for MyoD functions as a structure organizer of 3D genome architecture in muscle cells. MyoD is expressed in proliferating cells and functions as genome organizer that specifies the proper 3D genome architecture unique to muscle cell via two distinct mechanisms: acting directly by itself and through interactions with CTCF protein. Intriguingly, the MyoD-bound loops including pre-existed in proliferating cells and differentiation-induced play dual roles of regulatory and architectural in differentiating cells, which underscore a novel idea that a given cell type specific pioneer TF may in fact hide architectural potential.