**MyoD is a structure organizer for emergent property of 3D genome architecture in muscle cells**

Qian Chen[1]*, Fengling Chen[2]*, Ruiting Wang[1]*, 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#] and 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

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

## Results

### MyoD regulates A/B compartments switching and formation of contact domain boundaries (CDB) in muscle cells

In light of the aforementioned MyoD data, we first performed an RNA-seq analysis of proliferating (GM, Pax7$^{+}$ and MyoD$^{+}$) and differentiating (DM, MyoG$^{+}$) muscle cells isolated from hind limb skeletal muscle of wild type (WT) and MyoD knockout (MyoD$^{-/-}$, MKO) mice, which we combined with publicly available MyoD ChIP-Seq data to examine potential functional link(s) between MyoD occupancy and transcriptional activation. Specifically, we extracted MyoD binding peaks localized at promoter regions and identified their cognate genes within ± 3 kb (Fig. 1A and Extended Data Fig. 1B): knockout of MyoD did not affect the expression of a majority of these genes (76.7% for 2691/3510 in primary proliferation myoblasts, 56.6% for 1987/3509 in differentiation myocytes) (Fig. 1A), suggesting that the

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

To directly investigate potential genome architectural roles of the pioneer TF MyoD in muscle cells, we further examined four muscle sample types (WT-GM, WT-DM, MKO-GM, MKO-DM) with bridge-linker Hi-C (BL-Hi-C)[19] and ChIP-seq with antibodies against CTCF and several histone marks (H3K27ac, H3K4me1, H3K4me3, and H3K27me3) (Fig. 1B). For the BL-Hi-C data, 28 libraries were sequenced to a total depth of over 12 billion reads; the high quality deep sequencing data were validated with high cis interaction rate (Extended Data Fig. 1C) and reached 5 kb resolution (Extended Data Fig. 1D), thereby ensuring the rigor of our subsequent computational analyses of each hierarchical level of chromatin structure in muscle cells.

We then examined the impact(s) of MyoD depletion on A/B compartments and contact domain boundaries (CDBs) in muscle cells by performing exploratory principal component analysis (PCA) to determine the compartment values and relative insulation scores for the four muscle sample types examined (see Methods). Our results showed clear distinctions between the WT and MKO genotypes for both proliferating and differentiating muscle cells (Fig. 1C). Moreover, we detected a clear A/B compartment shift in proliferating MKO cells vs. WT cells (1.69% B→A, 0.84% A→B) (Fig. 1D left and 1E left). For differentiating cells, there was a clear compartment switch for MKO cells compared to differentiating WT cells (1.38% B→A and A→B 1.28%) (Fig. 1D left, and 1E right).

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

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

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

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

The most well-characterized anchor protein in vertebrates is CTCF, which mediates chromatin looping in concert with cohesion probably through the loop extrusion mechanism[2,20-24]. Considering our finding that MyoD apparently functions together with CTCF to regulate insulation of CDBs, we wondered whether MyoD may control the formation of chromatin loops in muscle cells. We therefore used HiCCUPs[25] to assess our BL-Hi-C dataset and ultimately identified 24,492 chromatin loops in proliferating WT myoblasts and 31,243 chromatin loops in differentiating WT myocytes (Fig. 2A, see Methods). The accuracy of the loop calls was supported by high-scoring aggregate peak analysis (APA) plots (Fig. 2A, see Methods) and by detected enrichment of CTCF-bound loops in muscle cells (Extended Data Fig. 2A) wherein about 75% of them had convergent CTCF binding motifs (Extended Data Fig., 2B).

We then searched for enriched motifs within accessible regions in loop anchors and detected significant enrichment for known myogenic E-box motifs[14,15](Fig. 2B). This enrichment was further confirmed by our observation of enrichment for MyoD binding peaks at loop anchors as analyzed using pseudopeaks (see Methods) with combined MyoD ChIP-seq data from the present study and other published work[17] (Fig. 2C and Supplementary Table 1): 44.18% of chromatin loops in GM cells were MyoD-bound (10.87% with both anchors bound by MyoD, 33.31% with either anchor bound by MyoD), while 43.93% of chromatin loops in DM cells were MyoD-bound (10.75% with both anchors bound by MyoD, 33.18% with either anchor bound by MyoD). Furthermore, MyoD binding peaks at loop

anchors were concordant with binding signals of CTCF and SMC3[26], a subunit of the cohesin complex (Extended Data Fig. 2C). We cataloged all of the identified chromatin loops into four types based on the binding of MyoD and/or CTCF at loop anchors in muscle cells: MyoD-MyoD(noCTCF), MyoD-MyoD(CTCF), MyoD-CTCF and CTCF-CTCF (Fig. 2D,E). This classification was further supported by the observable MyoD co-localization with CTCF as well as non-overlaid MyoD signals or CTCF signals when the two proteins were imaged by direct stochastic optical reconstruction microscopy (dSTORM), a single-molecule localization microscopy technique that can reveal the organization of specific proteins with a lateral resolution of ~20 (Fig. 2F). In addition, we found that the MyoD-bound loops were significantly shorter than CTCF-CTCF loops (288 kb on average), and the MyoD-MyoD(noCTCF) chromatin loops even shorter (113 kb on average) (Extended Data Fig. 2D). Together, these data substantiate MyoD localization on anchors of chromatin loops in muscle cells.

We next assessed whether MyoD could indeed mediate chromatin loop formation *in vivo* by investigating the effects of MyoD knockout on the MyoD-bound loops we earlier detected in WT cells. APA plots indicated that MyoD knockout significantly decreased the loop strength values for both MyoD-bound chromatin loops (MyoD-MyoD(noCTCF), MyoD-MyoD(CTCF) and MyoD-CTCF loops) and CTCF-CTCF chromatin loops, with these decreases evident in both GM and DM cells (Fig. 2G and Extended Data Fig. 2E). These results support that MyoD functionally contributes to the formation of both MyoD-bound and CTCF-bound chromatin loops *in vivo*. Taking MyoG locus as an example, the FISH analysis further confirmed that the MyoD-bound chromatin loop connecting MyoG promoter and Mybph promoter were not established in differentiating MKO muscle cells (Fig. 2H,I). Finally, we validated MyoD-orchestrated chromatin looping by performing *in vitro* DNA circularization assays with DNA fragments bearing known MyoD binding motifs (Methods). These assays confirmed that the MyoD protein can induce the circularization of linear DNA with MyoD binding motifs *in vitro* (Fig. 2J and Extended Data Fig. 2F,G).

Collectively, these genetic, biochemical, and cellular imaging results demonstrate that MyoD functions as an anchor protein that orchestrates chromatin loop formation across the

genome in muscle cells. Moreover, our data support that MyoD impacts chromatin loop formation by itself and through its physical interaction with CTCF.

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

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

Seeking further molecular insights into MyoD-bound muscle cell-specific chromatin looping, we examined ChIP-seq data for both MyoD and NeuroD—a pioneer transcription factor for neuronal cell lineage commitment and differentiation[18,27]. Comparison of their respective binding peaks in muscle cells and neuron cells indicated that only 18% of MyoD binding peaks in muscle cells were overlapped with NeuroD binding peaks in cortical neuron cells. These data indicate that MyoD and NeuroD mediate the formation of separate sets of chromatin loops that can subsequently regulate distinct sets of target genes to properly specify lineage-specific cell identities; for example, the *Myh* gene cluster located within a MyoD-bound, muscle-cell-specific chromatin loop was not detected in neuronal cells, and there was no NeuroD binding at the same genomic region in neuronal cells (Fig 3C). *Vice versa*, chromatin loops bound by the NeuroD at the *Foxg1* loci in CN cells was neuron-specific and not observed in muscle cells (Extended Data Fig. 3B). Taken together,

these results illustrate that cell-lineage pioneer TFs exert genome-wide architectural roles for modulating 3D genome organization, a functional impact considerably larger than the previous, primarily gene-transcriptional-activation related understanding about the influence of pioneer TFs on lineage specification.

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

Recalling previously published studies showing that MyoD constitutively binds to tens of thousands of sites genome-wide in proliferating cells, our finding that such binding only infrequently results in transcriptional activation at this stage, and our demonstration that MyoD functions as an anchor protein for lineage-specific 3D genome organization, we re-assessed our MKO knockout data to further explore the nature of MyoD's genome architectural functions (*i.e.*, how the presence of MyoD influences internal interactions occurring within chromatin loops).

Previous studies have indicated that CTCF and cohesin function together to establish architectural loops and that the extent of enhancer-promoter (E-P) interactions is elevated within such loops[28]. As a proportion of the MyoD binding peaks co-localize with CTCF and with cohesin, we speculated that MyoD-bound loops may also feature increased intra-loop interactions. Therefore, we measured internal interactions for each loop by calculating a "domain score" (D-score) in MKO proliferating cells. Comparative analysis with the WT and MKO data revealed that MyoD binding is enriched on chromatin loops that display reduced internal interactions (65%) upon MyoD knock out (Fig. 4A,B), indicating that MyoD regulates the internal interactions within chromatin loops by its direct binding on loop anchors like CTCF and cohesin. Further analysis showed that interactions changes within loops were positively correlated with expression changes of genes enclosed in the same chromatin loop (Fig. 4C).

We next focused on analyzing the internal interaction changes of MyoD-bound loops and their cognate gene expression (Fig. 4D). We specifically examined those MyoD-bound loops displaying reduced interactions in MKO proliferating cells to determine if these are regulated directly by MyoD binding. Nearly 50% of these loops enclosed down-regulated

genes with known myogenic-related functions (Fig. 4D), indicating that interactions within these MyoD-bound loops are apparently necessary for the regulation of muscle identity gene expression. For example, the *Rbm24* locus, a known gene that regulates myogenesis[29] and its differential expression in WT and MKO cells offers an excellent illustration for a gene enclosed within this type of MyoD-bound loop in muscle cells (Fig 4E and Extended Data Fig. 4A).

Intriguingly, we also found that most (89%) of the remaining MyoD-bound loops with decreased interactions did not contain any differentially expressed genes upon MyoD knockout (Fig. 4D). This finding underscored the apparent centrality of the architectural role (rather than transcriptional activation) of MyoD for these MyoD-bound loops in muscle proliferating cells. Although reduced interactions within these chromatin loops in MKO cells did not alter gene expression in proliferating cells, we found that genes within these chromatin loops were not turned on or off properly in the MKO-DM cells as compared with WT-DM cells (Fig 4F), and this genetic dysregulation was clearly manifest in the defective differentiation phenotype of the MKO cells (Extended Data Fig. 1B) [30-32]. As these chromatin loops remained in differentiated cells (Extended Data Fig. 4B), we interpret these data to indicate that the interactions constrained within MyoD-bound loops in proliferating cells are effectively "primed" for muscle cell differentiation by the presence of MyoD. That is, some very early signal must somehow direct the occupancy of MyoD to the sites across the genome that well subsequently permit both muscle-cell-appropriate loop architecture and later rewiring in response to differentiation signals. As one example, Lrmt1 is a known myogenesis related gene[33] which is only expressed in differentiated muscle cells (Extended Data Fig. 4C), but we found that its corresponding chromatin loop was already defined by MyoD in proliferating cells (Fig 4G). Thus mechanistically, our analysis of the MyoD-bound loops and their internal interactions revealed molecular insights about how MyoD exerts its architectural function, even in proliferating cells, to ensure the correct trajectory towards eventual muscle cell identity.

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

The genome is not only structurally organized within the nucleus but is also dynamically orchestrated in response to various cellular signals[5,34-38]. However, the molecular mechanism(s) underlying dynamic changes in 3D genome structure are largely unknown. Although the biological function of MyoD in regulating muscle cell differentiation has been well documented, and very recently, MyoD transcription activity was regulated by its ability to mediate chromatin interaction[39], how MyoD impacts chromatin conformation dynamics during muscle cell differentiation remains largely elusive. Using a modified differential loop detection method[38] (see Methods), we identified 6,242 differential chromatin loops (~25%) between proliferating and differentiating muscle cells (Supplementary Table 2). Among these loops, 5754 were significantly enhanced whereas 488 were reduced in differentiating cells (Fig. 5A), and we detected significant enrichment for muscle cell differentiation related genes in the enhanced chromatin loops (Extended Data Fig. 5A). These data support that nearly 25% of chromatin loops that undergo dynamic changes during muscle cell differentiation contribute to 3D genome rewiring in a manner that functionally regulates myogenic differentiation.

A similar analysis of MKO cells identified 585 differential chromatin loops between WT-DM and MKO-DM cells (Supplementary Table 2). Consistent with their regulatory impacts, 81.5 % (477/585) of these were significantly reduced in the differentiated MKO cells (Extended Data Fig. 5B), and 71.57% (341/477) of these reduced loops identified in the differentiated MKO cells had MyoD ChIP-seq data support as bona fide MyoD-bound chromatin loops in differentiated WT cells (Extended Data Fig. 5C). Among the aforementioned enhanced loops in differentiation, 285 loops might be directly induced by MyoD as they were bound by MyoD and disappeared when MyoD was knock out in differentiating muscle cells. The genes associated with those MyoD induced loops significantly upregulated during differentiation and downregulated when MyoD was knock out,indicating the regulatory role of MyoD induced loops (Fig. 5B).

Next, we sought to understand the driving force for the enhancement of MyoD induced loops during muscle differentiation. One possibility might be due to increased MyoD binding on loop anchors during muscle cell differentiation. By examining MyoD binding dynamics on anchors of MyoD induced loops, we found that majority of MyoD induced loops did not show

increased MyoD binding during the differentiation (Fig. 5C). However, we found that 88% of MyoD induced loops were decorated with MyoD dependent H3K27ac modification (Fig. 5C), suggesting that histone modifications may contribute to loop formation during muscle cell differentiation and that MyoD somehow actively directs the deposition of these modifications.

Taking the *Mef2c* locus as an example of MyoD induced loops to demonstrate their regulatory role in differentiation; we found that differentiated MKO cells had no enhancer or promoter loops at this locus (Fig. 5D). Concordantly, H3K27ac modification was diminished at this locus (Fig. 5D) and *Mef2c* gene expression was not induced in MKO cells cultured in DM (Fig. 5F upper). Another example is the *Gadd45a* locus, from which the mRNA transcription product was only expressed in WT-DM cells (Fig. 5E and Fig. 5F lower). Collectively, these results demonstrate MyoD orchestrates the formation of myogenic lineage specific chromatin loops that are required for the transcriptional regulation of myogenic genes during muscle cell differentiation.

**H3K27ac in the MKO cells is not sufficient to drive formation of the MyoD-bound chromatin loops**

The dynamic organization of the 3D genome is known to facilitate sophisticated interplay between aaccessible versus inaccessible chromatin states and TF occupancy[36]; however, elucidating the interdependent impacts (and likely feedback) of chromatin status on 3D genome organization remains a complex and challenging open question. Aside from MyoD dependent H3K27ac modification on MyoD induced loops, we found that during muscle cell differentiation, the loop anchors which had increased H3K27ac signals showed enrichment for consensus E-box sequences (i.e., the known target motif of the MyoD protein) in general (Fig. 6A). When MyoD is knocked out, H3K27ac modification was decreased on MyoD binding regions (Fig. 6B) which further support the overall dependency of H3K27ac modification on MyoD during differentiation. As MyoD itself stably binds around many myogenic genes in proliferating muscle cells, MyoD dependent H3K27ac modification correlates even better with the loop formation or disruption. Thus, we asked whether

H3K27ac-mediated active chromatin state triggers the formation of MyoD-bound chromatin loops during muscle cell differentiation.

We tested this possibility by treating MKO cells with the histone deacetylase I inhibitor Tacedinaline (Tac) to enrich the H3K27ac signal in the cells throughout muscle cell differentiation. With Tac treatment, 18110 genomic regions including anchors of MyoD induced loops (96/285) accumulated the H3K27ac modification (Fig. 6C). The effect of the accumulated H3K27ac on chromatin loops was analyzed by using BL-Hi-C data from the Tac-treated MKO and DMSO-treated cells. We showed that in the absence of MyoD, the accumulation of H3K27ac *per se* is not sufficient to trigger the formation of new loops or to enhance the signal for MyoD-bound or CTCF-CTCF chromatin loops (Fig. 6D, E). These results establish that simply accumulating H3K27ac in MKO muscle cells cannot rescue MyoD induced loops or drive other loop formation. Our genetic results clearly demonstrate that H3K27ac are insufficient for establishing characteristic 3D genome organization in muscle cells. Together, our overall findings reveal a critically determination role of the pioneer TF MyoD as a genome organizer in establishing the unique 3D genome architecture in muscle cells.

**Discussion**

Each cell type acquired unique organization of 3D genome during development and cell-type specific organization of 3D genome can be regarded as an emergent property that is mediated by interplay between transcription factors and chromatin-associated proteins[40]. Based on this model, the components in regulating the emergent property of chromatin should be coordinately orchestrated by a genome organizer to instruct specific organization of 3D genome in each cell type during development. In this report, we provide compelling evidence to support this model wherein binding of the pioneer transcription factor MyoD at sites across the genome in muscle stem cells exerts previously unappreciated function as a genome organizer, which has broad functional impacts on genome structure beyond its known functions in activating gene expression during development (Extended Data Fig. 6C).

Mechanistically, we show that internal interaction changes in MyoD-bound loops suggests a new insight into the potential mechanism of cell lineage pioneer TFs to ultimately specify the positions and nature of the chromatin loops comprising the cell-type specific internal interactions which ensures the subsequent differentiation process leading to lineage specific cell identity. Further, we demonstrate that biochemical idea that interaction between a pioneer TF MyoD and structural protein CTCF explains the chromatin accessibility states which ultimately support the lineages specific genetic developmental programmes. Finally, we show how the occupancy of pioneer TF proteins coordinate the epigenetic impacts of histone modifications including H3K27ac to retain active expression of the pioneer-TF-specified genome structures and developmental genetic programmes underlying proper muscle stem development.

It will be interesting to determine which signals ultimately initiate MyoD's loop specification layer of genetic regulation at the earliest possible stage of myogenic lineage-specific cell fate determination. Excitingly, extending our insights beyond MyoD-mediated architectural regulation of muscle cell fate, we provide additional evidence from experiments with the well-studied neuron specific pioneer TF NeuroD in neuronal cells. This suggests highly a similar loop specification function for NeuroD that apparently underlies the neuron lineages specific 3D genome architecture and attendant (downstream) genetic programming of cell fate. Future investigations which discovery (or fail to identify) the non-transcriptional regulatory, genome-wide architectural impacts of other lineage-specific pioneer TFs from a large number of cell types in diverse organisms will reveal whether our discoveries about MyoD represent a general process for pioneer TFs as genome organizers that orchestrate the lineage specific chromatin and 3D genome structures required for cell fate determination during development.

**Funding**

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.

**Author contributions**

Q.C. conducted cell culture, BL-Hi-C library, ChIP-seq, RNA-seq libraries and wrote the manuscript. F.C. performed all computational and bioinformatics analysis and 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.

**Acknowledgments**

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

**Reference**

1 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (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, doi:10.1016/j.cell.2014.11.021 (2014).

3 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

4 Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol* **20**, 535-550, doi:10.1038/s41580-019-0132-4 (2019).
5 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 e524, doi:10.1016/j.cell.2017.09.043 (2017).
6 Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737-743, doi:10.1038/s41586-020-2151-x (2020).
7 Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345-354, doi:10.1038/s41586-019-1182-7 (2019).
8 Kim, S. & Shendure, J. Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol Cell* **76**, 306-319, doi:10.1016/j.molcel.2019.08.010 (2019).
9 Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000, doi:10.1016/0092-8674(87)90585-x (1987).
10 Tapscott, S. J. *et al.* MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* **242**, 405-411, doi:10.1126/science.3175662 (1988).
11 Weintraub, H. *et al.* Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad Sci U S A* **86**, 5434-5438, doi:10.1073/pnas.86.14.5434 (1989).
12 Rudnicki, M. A. *et al.* MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351-1359, doi:10.1016/0092-8674(93)90621-v (1993).
13 Lassar, A. B., Paterson, B. M. & Weintraub, H. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* **47**, 649-656, doi:10.1016/0092-8674(86)90507-6 (1986).
14 Lassar, A. B. *et al.* MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell* **58**, 823-831, doi:10.1016/0092-8674(89)90935-5 (1989).
15 Cao, Y. *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* **18**, 662-674, doi:10.1016/j.devcel.2010.02.014 (2010).
16 Mousavi, K. *et al.* eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol Cell* **51**, 606-617, doi:10.1016/j.molcel.2013.07.022 (2013).
17 Umansky, K. B. *et al.* Runx1 Transcription Factor Is Required for Myoblasts Proliferation during Muscle Regeneration. *PLoS Genet* **11**, e1005457, doi:10.1371/journal.pgen.1005457 (2015).
18 Fong, A. P. *et al.* Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev Cell* **22**, 721-735, doi:10.1016/j.devcel.2012.01.015 (2012).

19 Liang, Z. *et al.* BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat Commun* **8**, 1622, doi:10.1038/s41467-017-01754-3 (2017).
20 Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503-507, doi:10.1038/nature22063 (2017).
21 Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e614, doi:10.1016/j.cell.2017.04.013 (2017).
22 Heidari, N. *et al.* Genome-wide map of regulatory interactions in the human genome. *Genome Res* **24**, 1905-1917, doi:10.1101/gr.176586.114 (2014).
23 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922, doi:10.1016/j.cell.2017.05.004 (2017).
24 Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO journal* **32**, 3119-3129, doi:10.1038/emboj.2013.237 (2013).
25 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).
26 Tsai, P. F. *et al.* A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Mol Cell* **71**, 129-141.e128, doi:10.1016/j.molcel.2018.06.008 (2018).
27 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, doi:10.1186/s12864-015-1882-9 (2015).
28 Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789-800, doi:10.1038/s41576-018-0060-8 (2018).
29 Yang, J. *et al.* RBM24 is a major regulator of muscle-specific alternative splicing. *Dev Cell* **31**, 87-99, doi:10.1016/j.devcel.2014.08.025 (2014).
30 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, doi:10.1101/gad.10.10.1173 (1996).
31 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, doi:10.1083/jcb.144.4.631 (1999).
32 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, doi:10.1016/s0248-4900(00)01110-2 (2000).

33 Li, H. K. *et al.* LRTM1 promotes the differentiation of myoblast cells by negatively regulating the FGFR1 signaling pathway. *Exp Cell Res*, 112237, doi:10.1016/j.yexcr.2020.112237 (2020).
34 Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64, doi:10.1038/nature08497 (2009).
35 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).
36 Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* **50**, 238-249, doi:10.1038/s41588-017-0030-7 (2018).
37 Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**, 1522-1528, doi:10.1038/ng.3935 (2017).
38 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, doi:10.1016/j.molcel.2017.08.006 (2017).
39 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, doi:10.1016/j.molcel.2019.07.036 (2019).
40 Partridge, E. C. *et al.* Occupancy maps of 208 chromatin-associated proteins in one human cell type. *Nature* **583**, 720-728, doi:10.1038/s41586-020-2023-4 (2020).

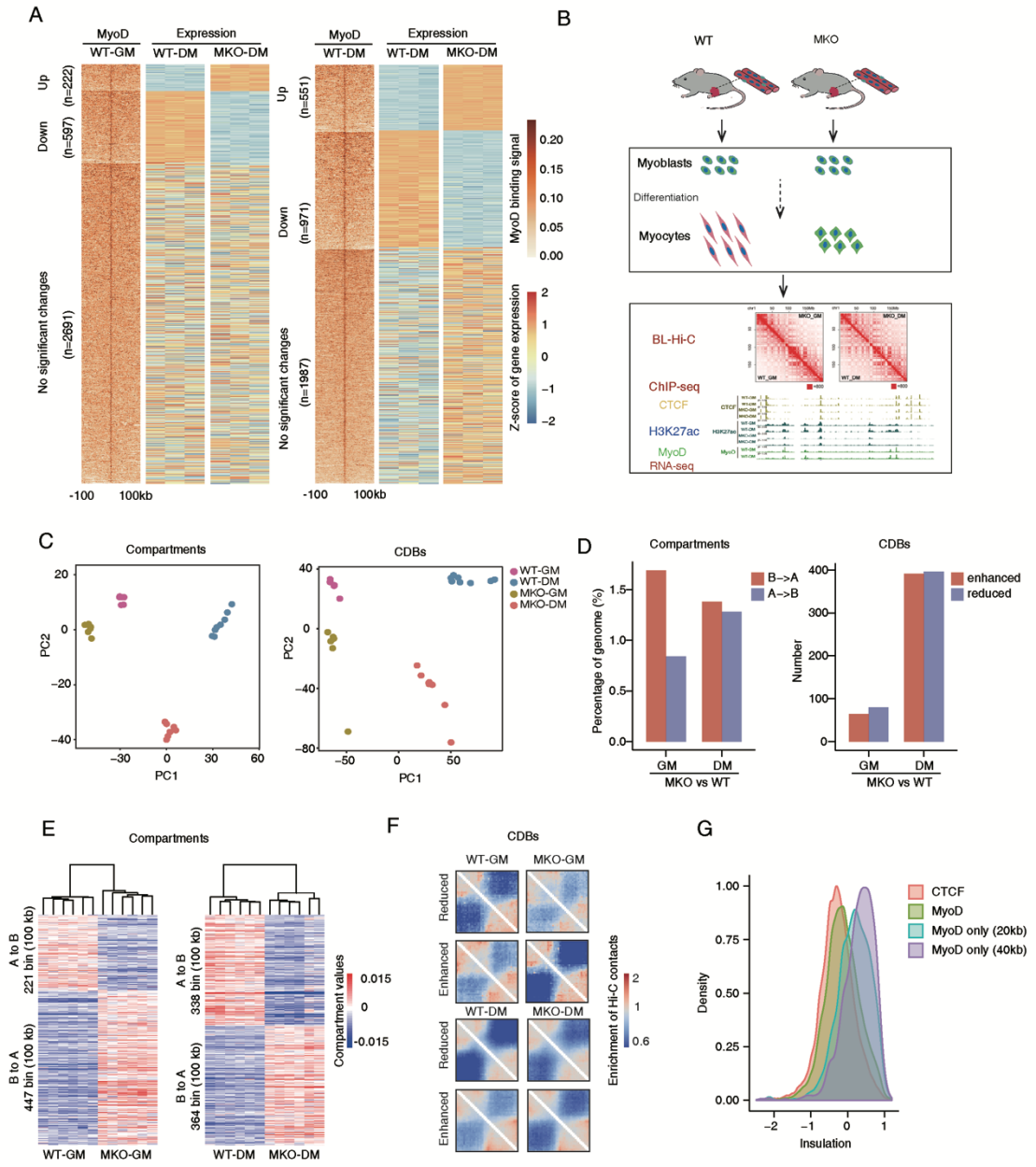


**Figure 1. MyoD regulates A/B compartments switching and formation of contact domain boundaries (CDBs) in muscle cells.**

(A) Association of MyoD binding peaks at promoter regions and related gene expression trends. Heatmaps showing MyoD binding peaks at promoters (centralized at the TSS within 3 kb distance) and related gene expression trends in wild type (WT) and MyoD knockout (MKO) proliferating muscle stem cells (myoblasts, cultured in growth media, GM) was shown at left panel. Corresponding data for differentiating muscle stem cells (myocytes, induced to differentiation in differentiation media, DM) was shown in the right panel. Pseudo-peaks of MyoD binding with combined ChIP-seq data from public databases and our study were used

for this analysis (see methods).

(B) Strategy for genetic investigation of the chromatin architectural roles of MyoD. Briefly, muscle stem cells isolated from WT or MKO mice were cultured in GM and induced to differentiation in DM for 24 hrs. The proliferating myoblasts and differentiated myocytes were subjected to BL-Hi-C analysis. To correlate 3D genome architecture to chromatin state and to gene expression levels, multiple ChIP-seq and RNA-seq analyses were performed on myoblasts and myocytes from WT and MKO mice.

(C) Principal Component Analysis (PCA) of the values of compartments or the insulation scores of contact domain boundaries (CDBs) among the four indicated sample types. Each dot represents an individual biological replicate.

(D) Barplots showing the percentage of A/B compartment shifts between WT and MKO in myoblasts and myocytes (left panel). Barplots showing the number of enhanced or reduced CDBs between WT and MKO in proliferating and differentiating muscle stem cells (right panel).

(E) Heatmaps of the A/B compartment shifted regions between WT and MKO in myoblasts and myocytes.

(F) Hi-C aggregation plots centered at differential CDBs between WT and MKO in myoblasts and myocytes.

(G) Probability density of insulation score at genetic regions bound by CTCF or MyoD. Four classes were assessed: CTCF-bound, MyoD-bound, MyoD-bound but CTCF-unbound within 20 kb distance; MyoD-bound and CTCF-unbound within 40 kb distance.

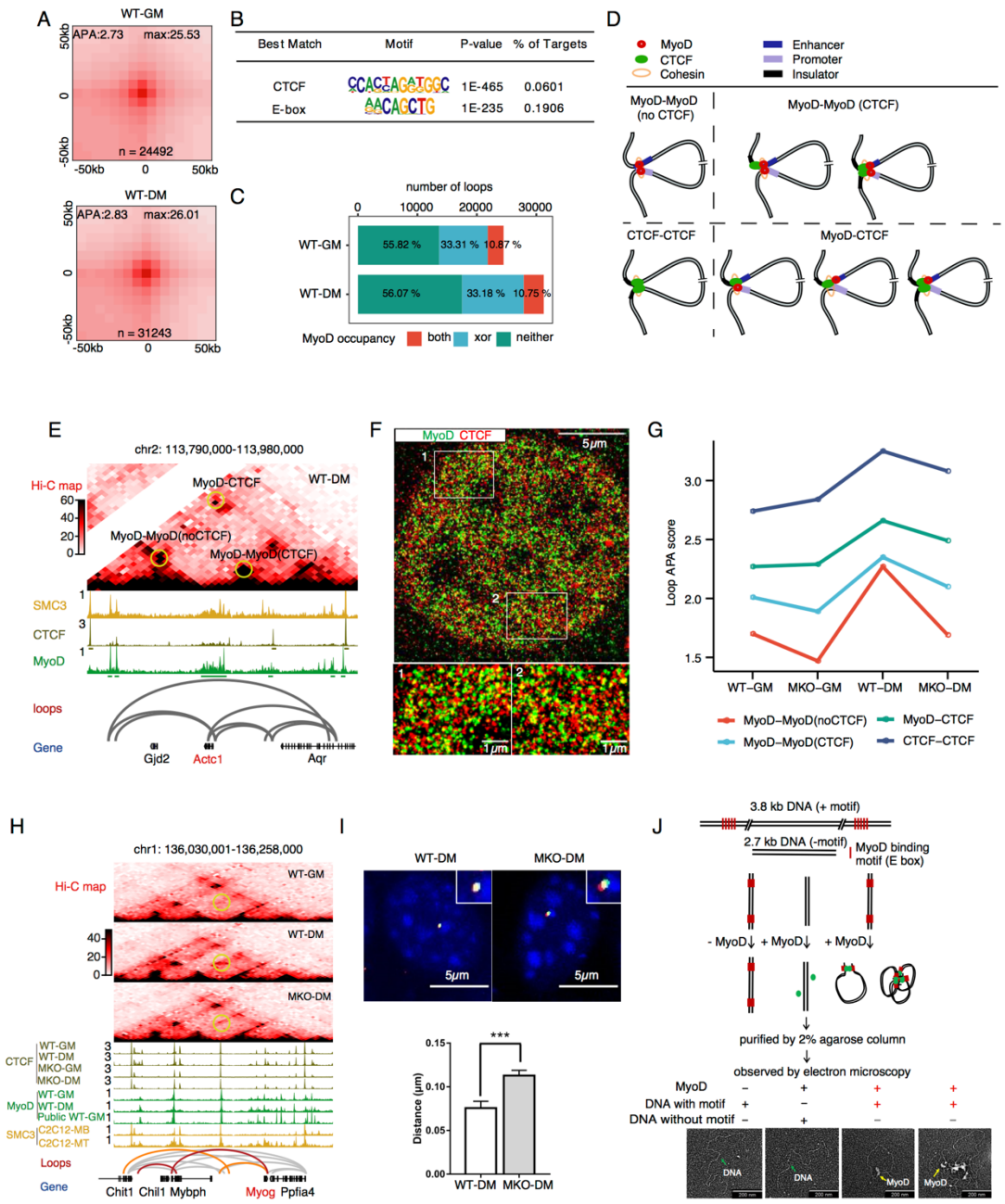


**Figure 2. MyoD is an anchor protein for chromatin loop formation in muscle cells.**

(A) Aggregate Peak Analysis (APA) plots showing the aggregated Hi-C contacts around chromatin loops identified in WT-GM and WT-DM cells. n represents the total number of chromatin loops called by HiCCUPS (see methods). APA score at top left corner of each aggregation plot was calculated as the ratio of Hi-C contacts at the central pixel to the mean Hi-C contacts of the lower left pixels. The Hi-C contacts at the central pixel was shown at top right corner of each aggregation plot.

(B) Enriched motifs in accessible regions at anchors of the chromatin loops identified in panel A.

(C) Percentage of chromatin loops with anchors bound by MyoD in WT-GM and WT-DM cells. 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 databases and our study were used for this analysis (see methods).

(D) Classification of the four types of chromatin loops identified in the study. MyoD-MyoD(noCTCF) loops had both anchors bound only by MyoD. MyoD-MyoD(CTCF) loops had both anchors bound by MyoD and either anchor concomitantly bound by CTCF. MyoD-CTCF loops had only one anchor bound by MyoD and either anchor bound with CTCF. CTCF-CTCF loops had both anchors bound by CTCF only.

(E) Representative examples of the MyoD-bound chromatin loops.

(F) Representative dSTORM image showing co-localization of MyoD (green) and CTCF (red); immunofluorescent staining of C2C12 cells grown in DM for 24 hrs with antibodies against MyoD and CTCF.

(G) APA score dynamics for the four types of chromatin loops among WT and MKO cells in GM and DM stages.

(H) A representative MyoD-bound chromatin loop at the *MyoG* locus that presents the Hi-C contact map depicting normalized contact frequencies and ChIP-seq signal tracks. The examined chromatin loop with reduced Hi-C contacts in MKO-DM sample (compared with WT-DM sample) was marked by yellow circle.

(I) The chromatin loop at the *MyoG* locus was validated by 3D-DNA fluorescent *in situ* hybridization (3D-DNA FISH) in WT-DM and MKO-DM cells. The upper panel showed *MyoG* promoter (red) and enhancer (green) were more separated in the MKO cells (right) than in the WT cells (left). The lower panel showed the average distances of *MyoG* loop anchors, measured using 3D-DNA FISH.

(J) MyoD mediates chromatin looping *in vitro*. The upper panel presents a schematic for the *in vitro* DNA circularization assay used to detect the ability of MyoD (green) to mediate DNA looping with E-box (red) containing DNA (3.8 kb) and no E-box motif control (2.7 kb). The lower panel presents representative views of circularized DNA mediated by recombinant

MyoD protein, imaged by Transmission Electron Microscopy (TEM).

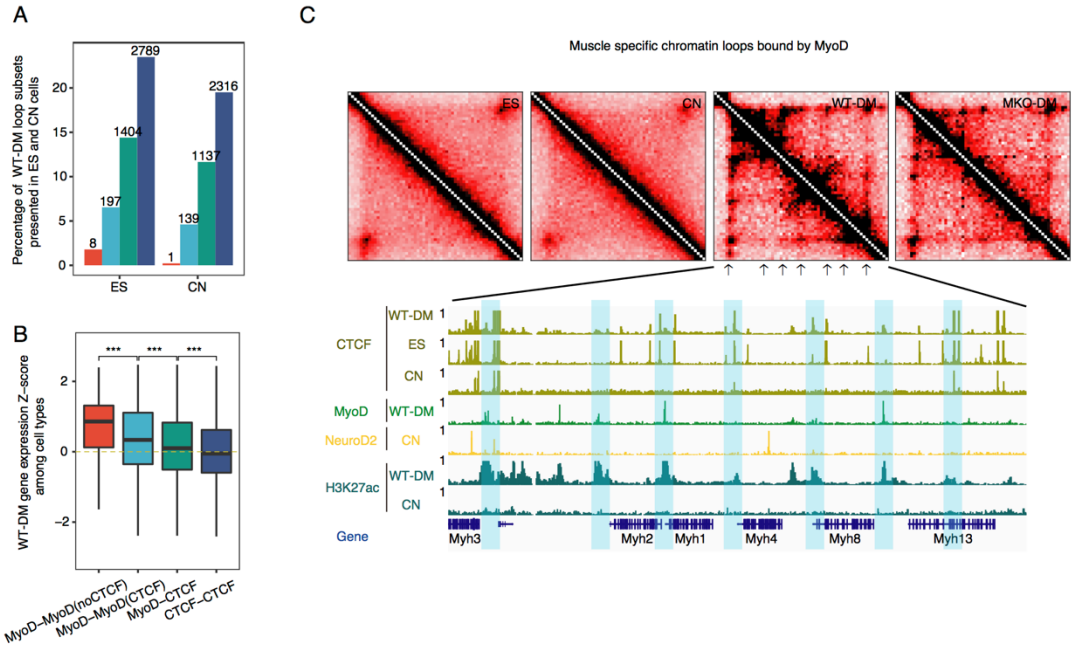


**Figure 3. MyoD anchors formation of myogenic-specific chromatin loops.**

(A) Relative 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, megakaryote, 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.

(C) Representative region showing muscle cell specific loops anchored on MyoD binding regions at *Myh* gene cluster.

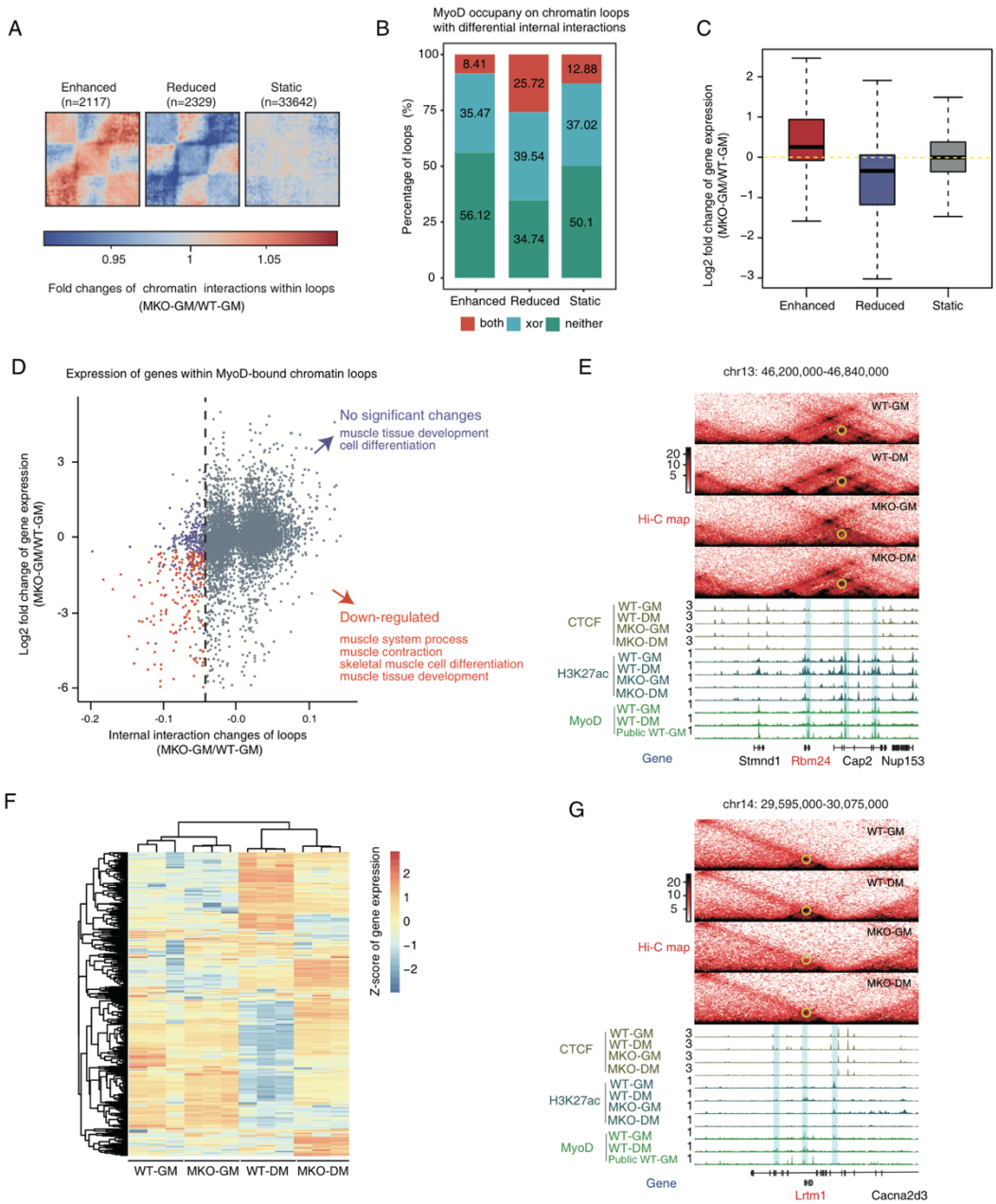


**Figure 4. MyoD mediates primed architectural chromatin loops in proliferating muscle cells.**

(A) Aggregate Hi-C maps around chromatin loops with enhanced, reduced, or static internal interactions (domain score) between MKO and WT 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 (see methods).

(C) Boxplot showing the corresponding expression changes of genes within chromatin loops having enhanced, reduced, or static internal interactions (domain score) identified in 4B (MKO-GM vs. WT-GM). Considering that one loop may contain several genes and D-scores of nested loops are associated, we assigned each gene to its most dynamic nested loop.

(D) Scatter plot showing both the domain score fold changes and gene expression fold changes of MyoD-bound chromatin loops between MKO-GM and WT-GM cells. Reduced MyoD-bound chromatin loops without significantly differentially expressed genes in MKO-GM cells compared with WT-GM cells were shown as blue dots, while reduced MyoD-bound chromatin loops with significantly down-regulated genes in MKO-GM cells compared with WT-GM cells were shown as red dots.

(E) The *Rbm24* locus as a representative for a reduced MyoD-bound chromatin loop without significantly differentially expressed genes in MKO-GM cells compared with WT-GM cells.

(F) Heatmap showing z-scores based on expression (FPKM) of genes within MyoD-bound chromatin loops.

(G) The *Lmrt1* locus as a representative for a reduced MyoD-bound chromatin loop with significantly down-regulated genes in MKO-GM cells compared with WT-GM cells.

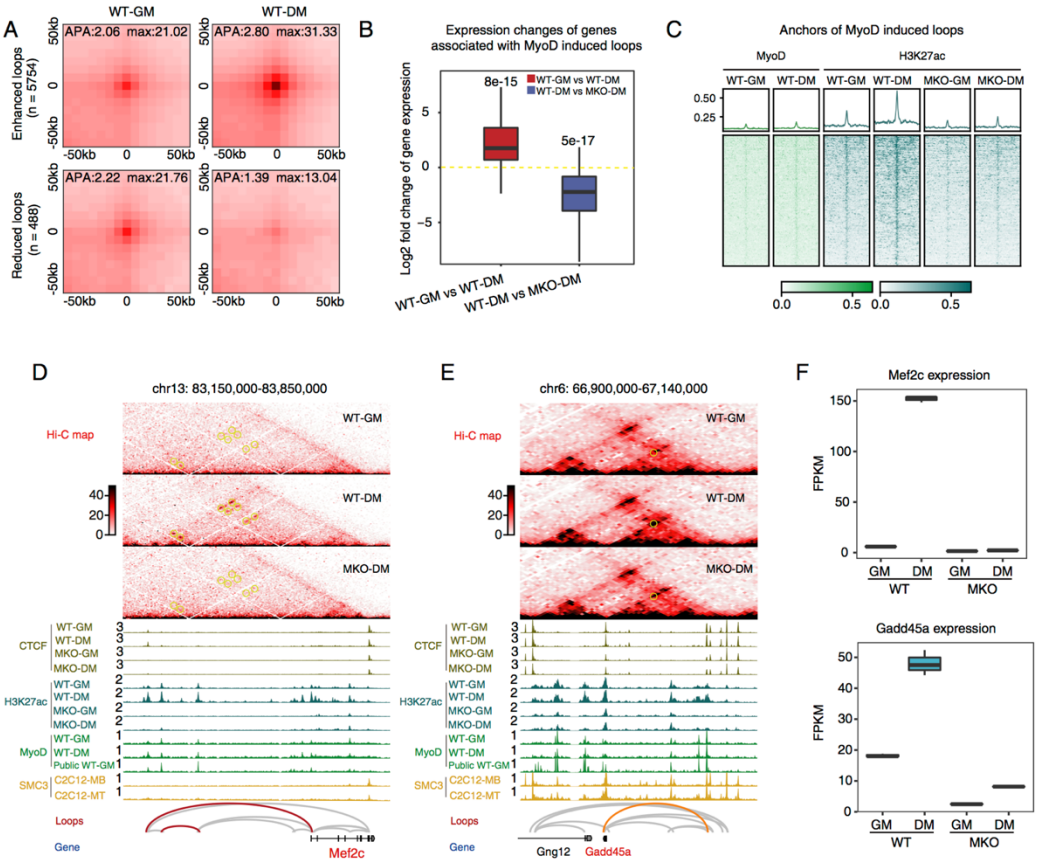


**Figure 5. The regulatory loops specified by MyoD are functionally required for muscle cell differentiation.**

(A) APA plots showing the aggregated Hi-C contacts around enhanced (n=5754) and reduced (n=488) chromatin loops between WT-DM and WT-GM cells. Differential loops were identified based on DESeq2 (see methods).

(B) Boxplot showing expression changes of genes associated with MyoD induced chromatin loops in WT-DM cells compared with WT-DM cells (left) and in MKO-DM cells compared with WT-DM cells (right).

(C) MyoD and H3K27ac ChIP-seq signal at anchors of MyoD induced loops. Heatmaps showing signal enrichment centralized at the TSS within 100 kb distance.

(D) The *Mef2c* locus as a representative of MyoD induced loops during muscle cell differentiation.

(E) The *Gadd45a* locus as a representative of a differentiation-induced loop bound by MyoD.

(F) *Mef2c* and *Gadd45a* gene expression among WT and MKO cells in proliferating and differentiating stage.

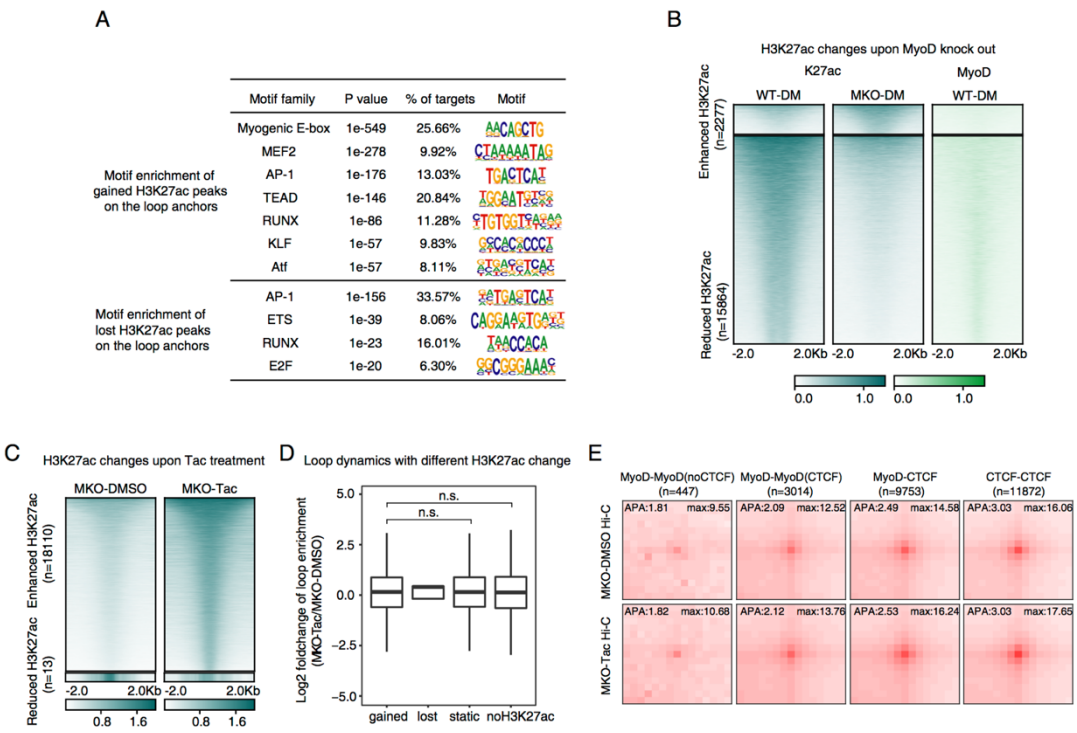


**Figure 6. H3K27ac in the MKO cells is not sufficient to drive formation of the MyoD-bound chromatin loops.**

(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 (see methods).

(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.