The cooperation of cis-elements during M-cadherin promoter activation

Yung-Jui Lin#, Chien-Han Kao#, Sheng-Pin Hsiao, and Shen-Liang Chen*

From:
Department of Life Sciences, National Central University, Jhongli 32001, Taiwan.

Running title: MyoD targets multiple E-boxes in M-cadherin promoter

#These two authors contribute equally to this study.

*To whom all correspondence should be addressed:
Dr. Shen Liang Chen, Department of Life Sciences, National Central University, 300 Jhongda Rd, Jhongli 32001, Taiwan (R. O. C.). Tel: 886-3-4227151 ext 65069; Fax: 886-3-4228482; E-mail: slchen@cc.ncu.edu.tw.

Keywords: muscle, M-cadherin, MyoD, Meis1, promoter, Pbx1
Abstract:
M-cadherin is a skeletal muscle-specific transmembrane protein mediating the cell-cell adhesion of myoblasts during myogenesis. It is expressed in the proliferating satellite cells and highly induced by myogenic regulatory factors (MRFs) during terminal myogenic differentiation. Several conserved cis-elements, including 5 E-boxes, 2 GC boxes, and 1 conserved downstream element (CDE) were identified in the M-cadherin proximal promoter. We found that E-box-3 and -4 close to the transcription initiation site (TIS) mediated most of its transactivation by MyoD, the strongest myogenic MRF. Including of any one of the other E-boxes restored the full activation by MyoD, suggesting an essential collaboration between E-boxes. Stronger activation of M-cadherin promoter than that of muscle creatine kinase (MCK) by MyoD was observed regardless of culture conditions and the presence of E47. Furthermore, MyoD/E47 heterodimer and MyoD–E47 fusion protein achieved similar levels of activation in differentiation medium (DM), suggesting high affinity of MyoD/E47 to E-boxes 3/4 under DM. We also found that GC boxes and CDE positively affected MyoD mediated activation. The CDE element was predicted to be the target of the chromatin-modifying factor Meis1/Pbx1 heterodimer. Knockdown of Pbx1 significantly reduced the expression level of M-cadherin, but increased that of N-cadherin. Using ChIP assay, we further found significant reduction of MyoD recruitment to M-cadherin promoter when CDE was deleted. Taken together, these observations suggest that the chromatin-modifying function of Pbx1/Meis1 is critical to M-cadherin promoter activation before MyoD is recruited to E-boxes to trigger transcription.
Introduction

Skeletal muscle is formed by bundles of parallel myocytes that are elongated, fiber-shaped, and multinucleated giant cells (also called myofibers in vivo or myotubes in vitro) running from tendon to tendon. Myocytes are functional mature cells and they acquire multinucleated phenotype through plasma membrane fusion of dozens to hundreds of mononucleated myoblasts in the process of myogenic terminal differentiation (1). Trunk myoblasts are derived from somitic dermomyotome precursors that are confined to the SKM lineage by the expression of either MyoD or Myf5. Upon the stimuli of terminal myogenic differentiation signals, either factor can direct the expression of Myogenin and Mef2c in myoblasts to facilitate the acquiring of myofiber phenotypes, including cell cycle exit, contractile machinery, and multinucleation (2). Many adhering factors, including Cadherin and Ig families, have been found to participate in the fusion of myoblasts into myofibers (3).

Cadherins are trans-membrane adhering proteins that connect neighboring cells in a Ca^{2+}-dependent manner (4, 5). Among them, M- and N-cadherin are the major members expressed in the developing SKM cells and they allow myoblasts to recognize, and adhere to one another before fusing into multinucleated myofibers. The homotypic binding of extracellular domain from M- and N-cadherins ensures that fusion only happens between myoblasts (6). Inter-cellular Cadherin-binding transduces signals into the nucleus to regulate the expression and function of MRFs and to promote cell cycle exit (7). The importance of Cadherins in myogenesis can be demonstrated by the finding that the fusion of myoblasts can be blocked by antibodies targeting the extracellular domain of either M-cadherin or N-cadherin (8-11), and similar effect can also been achieved by incubating with peptides blocking the binding
between M-cadherins on neighboring cells (12).

Although SKM cells express M-, N-, and R-cadherin, M-cadherin is the only muscle-specific one and its expression is highly induced during myogenesis (13, 14). We have demonstrated the strong activation of M-cadherin promoter by MyoD and other MRFs and this pathway might be responsible for its muscle-specific expression as all myogenic regulatory factors are muscle-specific (15). M-cadherin is moderately expressed in proliferating satellite cells and myoblasts, and its highest expression is observed during early myogenic differentiation when myoblasts adhere to fuse together, then, the expression declines in maturing myocytes, and similar phenomenon is observed in developing somites in utero (16, 17).

The proximal promoter of M-cadherin contains 5 E-boxes and some other cis-elements, including GC-boxes and a conserved downstream element (CDE). E-boxes 3 and 4 have been found to mediate most of the activation effects by MyoD and the conserved downstream element (CDE) also plays important role during this process (15). The sequence context of CDE is predicted to be the target site of the heterodimeric factor Pbx/Meis, and the involvement of this factor in M-cadherin transcriptional activation has been demonstrated before (18). Pbx/Meis belongs to the TALE (three amino acid loop extension) subfamily of the homeodomain DNA-binding transcription factors and they are important cofactors regulating/modulating DNA-binding and transactivation of homeodomain transcription factors (19, 20). A subset of myogenic genes, including Myogenin, are targeted by Pbx/Meis for initial relaxation of chromatin structure before recruitment of the MyoD for their transcription (18). Direct interaction on target sites between MyoD and Pbx/Meis has been examined (21, 22). It is of interest to elucidate how the Meis/Pbx recruited to CDE might cross-talk with
MyoD or other MRFs bound to E-box to determine the chromatin for transcriptional activation.

In adult skeletal muscle, *M-cadherin* expression is restricted to the apical side of the myogenic stem cells (called satellite cells) that are located within the basal lamina of each myocyte (23). The expression of *M-cadherin* is activated after damage of skeletal muscle and the highest level is found in the fusing myoblasts. However, its expression is not induced by denervation that leads to SKM atrophy, implying its participation in the stem cells mediated regeneration (8, 17, 23). As *M-cadherin* is expressed in the proliferating myoblasts and quiescent satellite cells where most myogenic genes are not expressed, which implies the existence of a unique mechanism by which *M-cadherin* is activated but without inducing other myogenic genes participating in terminal differentiation. Elucidating this *M-cadherin*-specific activation mechanism in myogenic precursors will certainly gain insight into the mechanism by which MyoD activates gene transcription and the acquisition of myogenic stem cell characteristics. Therefore, a detailed analysis of the cooperation between cis-elements in the proximal promoter of *M-cadherin* was performed here.
Materials and Methods:

**Cell culture and transient transfection assay**

Proliferating C3H10T1/2 fibroblasts and C2C12 myoblasts were kept at low confluence in growth medium (GM, DMEM supplemented with 10% and 20%, respectively, fetal calf serum (FCS)). To induce myogenic differentiation, C2C12 at confluence myoblast stage (CMB) were changed to differentiation medium (DM, DMEM with 2% horse serum) for 4-6 days until myotubes formed. Transient transfection assay was done in 12-well plates and has been described before (24). Briefly, aliquots of DNA containing promoter-driven reporters (0.67 μg/well) and expression vectors (0.17 μg/well) were mixed with non-liposome transfection reagent (T-Pro, T-Pro biotechnology) in 1X Hepes buffer and incubated at room temperature for 10-15 min. Then, the mixture was transferred onto cells (70-80% confluence) and the transfection was allowed to proceed overnight before the medium was replaced by fresh medium and incubated for another 24 hr before harvested for determining their luciferase activity using the Bio-Tek Clarity 2 luminometer. All reactions were done in triplicate and repeated at least 3 times.

**Plasmids**

The cloning of *M-cadherin* promoter of full length and proximal regions into pGL3-basic vector has been reported before (15). Mutation of each E-box was performed by overlapping PCR using primers carrying mutated E-box sequence (AAAAAAA) and the mutant PCR product was inserted into the Smal site of pGL3-basic vector. Mutant promoters carrying multiple E-box mutations were created by overlapping PCR using previously E-box (s) mutated promoter as template to make 2nd, 3rd, and 4th mutations sequentially. For stable cloning of wildtype and mutant promoters, they were
subcloned into the HindIII (blunted) site of the pStable-luc vector. The GC-boxes and CDE were deleted with overlapping PCR and cloned into p-Stable-luc directly. The CDS of E47 was amplified from cDNA derived from Sol8 myoblasts and was inserted into the EcoRV site of the pCDNA3.1 vector. The CDS of MyoD was inserted into the NdeI site of pET32a-LRH to create pET32a-MyoD-LRH, in which MyoD was linked to a flexible linker sequence (RSESGGGGSPG) that allows free moving of connected proteins (25). The MyoD-linker CDS was amplified and the insert was subcloned into the EcoRV site of the pCDNA3.1 vector. Then, E47 CDS was inserted into the 3’-EcoRV site to make MyoD–E47 chimeric gene. All PCR was performed with pfu DNA polymerase and the PCR products derived clones were sequence to ensure the sequence integrity. The primer sets used in cloning are listed in Supplementary table II.

**Chromatin immunoprecipitation (ChIP) assay**

The detailed procedure of CHIP assay has been described in our previous works (26-28). Briefly, C2C12 cells were washed, fixed in formaldehyde (1%), and sonicated to shear chromatin. Specific antibody was added to the cleared lysate and the binding was allowed to proceed at 4°C overnight before Staph A (Sigma, #P7155) was added to capture the immune complex and then washed extensively. Then, immune complex was eluted and the released DNA was extracted with phenol/chloroform twice and further purified using a PCR purification Kit (Geneaid; Taipei, Taiwan). The primer sets used in ChIP assay are listed in Supplementary table III. Antibodies (0.5 μg) against MyoD (Clone MoAb 5.8A, BD Pharmingen) and Flag-tag (clone M1 # F3040, Sigma) were purchased from BD Pharmingen and Sigma respectively. Meis1 and Pbx1 antibodies were purchased from abcam.
Quantitative RT-PCR (qRT-PCR)

The detailed protocol of qRT-PCR has been described in our previous works (29, 30). Briefly, cells and embryos were solubilized in Solution D and then, total RNA was extracted with series of phenol/chloroform mixture. Their cDNA was synthesized by the Superscript III kit (Invitrogen) according to the manufacturer’s protocol. The qPCR product was detected by SYBR Green reaction mix (RealQ Plus 2X Master Mix Green, AMPLIQON). All reactions were performed in a Rotor-Gene Q real-time PCR cycler (QIAGEN) with an amplification program of 45 cycles. The qRT-PCR primer sequences used in this study were described in Supplementary table I.

Western blot

The detailed protocol of Western blot has been described in our previous papers (29, 31). Briefly, C2C12 cells were lysed in RIPA buffer containing protease inhibitors. Total protein (50 μg) was resolved on SDS-PAGE gel, and then transferred onto PVDF membrane. After blocking with blocking buffer (PBS containing 0.5% Tween 20 and 5% milk), the membrane was hybridized by primary antibody at 4°C overnight. HRP-conjugated secondary antibodies were added into the membrane and incubated at room temperature for 1 hr. After washed with PBS, the HRP signal was detected by an enhanced chemiluminescence kit and viewed with X-ray film. Here is the list of primary antibodies: MyoD (Clone MoAb 5.8A, BD Pharmingen); Mes1 (ab19876, abcam); Pbx1 (ab192606, abcam; #4342, Cell Signaling); N-cadherin (ab18203-100, abcam); M-cadherin (78090, abcam); Gapdh (GTX100118, GeneTex).

Blocking the CDE element function in vivo

HEK 293FT cells were transfected with the plasmids pHR-SFFV-KRAB-dCas9-p2a-mCherry (6.25 μg), pCMV-Δ8.91 (5.625 μg), and pMD.G (0.625 μg) to generate
infectious lentivirus expressing KRAB-dCas9-P2A-mCherry fusion gene. Virus was harvested from medium 24 h after transfection and transferred to C2C12 myoblasts for infection. Infected C2C12 myoblasts were later serially diluted in 96-well titer plate to identify red clones expressing KRAB-dCas9-P2A-mCherry. Cells with bright mCherry fluorescence were further expanded and then transfected with PX459ΔCas9 vector (a derivative of the PX459 vector with the Cas9 gene deleted) carrying sgRNA (5'-'CCCCAATGCTCTCAGCGTGG-3') targeting the sequence immediately downstream of CDE site and selected with puromycin (2.5 μg/ml) for 2-3 weeks. Monoclones with bright mCherry fluorescence and puromycin resistance were picked up and expanded for analyzing the expression of M-cadheirn and other genes with qRT-PCR. Both pHR-SFFV-KRAB-dCas9-p2a-mCherry and PX459 vectors were acquired from Addgene originally, and PX459ΔCas9 was a generous gift from Dr. Wei-Yi Chen (Department of Biochemistry and Molecular Biology, National Yang Ming University).

**Gene knockdown by lentivirus expressed shRNA**

For establishing Pbx1 knockdown clone, pCMV-Δ8.91 (5.625 μg), pMD.G (0.625 μg), and the shRNA expressing pLKO1 vectors (6.25 μg) were transfected into 293FT cells (on a 10 cm dish) to generate infectious lentivirus, which was collected 48 hr later before added to C2C12 culture medium. Infected C2C12 myoblasts were selected with puromycin (2 μg/ml) for at least 2 weeks to generate monoclononal colonies and these clones were expanded and examined for their Pbx1 levels with RT-PCR. Clones derived from pLKO1-274077 plasmid (clone ID TRCN0000274077) showed reduced Pbx1 level were used for further experiments. All vectors used for the knockdown experiments were purchased from the RNAi core facility (http://rnai.genmed.sinica.edu.tw) of the Academia Sinica.
Results:

Factors regulating M-cadherin are induced during myogenesis

The proximal promoter of M-cadherin contains several predicted cis-elements, including 5 E-boxes, 2 GC-boxes, and a conserved downstream element (CDE; Fig. 1A and Supplementary fig. 1A). Among them, cis-elements around transcription initiation site (TIS) are highly conserved between mammals (Fig. 1A & B). As M-cadherin expression is SKM-specific and highly induced during myogenesis, it is of interest to examine the expression of factors targeting its cis-elements during this process. This process can be recapitulated precisely in vitro by myoblast lines, such as C2C12 and Sol8, that are widely used for studying gene expression and other changes. The E-boxes are targeted by MRFs and the GC-boxes are binding sites for the ubiquitously expressed SP1. The CDE was predicted as the target for Mes1/Pbx1 heterodimer by the Genomatix software. Here we found the mRNA levels of both M-cadherin and MyoD are highly induced by cell-cell contact in confluent myoblasts (CMB) and further increase was observed as cells fused to form multinucleated myotubes (Fig. 1C). Myf5 expression was induced in CMB but dropped to basal level in MT. Cell-cell contact induced the expression of both Pbx1 and Meis1, and Pbx1 was further induced in MT. SP1 is a ubiquitously expressed transcription factor and its expression was not much affected during myogenesis.

The expression of these factors was further examined in myogenesis of embryos of 11.5-12.5 dpc. As somitogenesis proceeds in the direction of rostral to caudal, tissues in the rostral part contain mature somites and those in the caudal part contain early somites and presomitic mesoderm; therefore, somites at various developing stages can be seen in the trunk of a single embryo. Here the embryos were separated into rostral, interlimb, and tail parts to represent somites from mature to early stages (Fig. 1D). We
found *M-cadherin* and *MyoD* were highly expressed in the rostral and interlimb somites but not in the head region, where nervous tissue was the major component at this stage. *N-cadherin*, *Pbx1*, and *Meis1* were similarly induced during somitogenesis, but the former two also showed high expression in the head region probably due to their high expression in the CNS. The observations in C2C12 differentiation and embryo somitogenesis demonstrated the correlation between the expression of these factors and myogenesis, and it further suggest that Pbx1 and Meis1 might participate in the transcriptional regulation of *M-cadherin* as MyoD and Myf5 do.

**Multiple E-boxes are required for full activation of *M-cadherin* expression**

There are five conserved E-boxes (E1~E5) found in the *M-cadherin* proximal promoter (-252~+200; Fig. 1A). To identify the role played by each E-box, they were mutated individually and we found that mutation of either E1 or E5 did not negatively affect its activation by MyoD (Fig. 2A). Mutation of E2 or E3 progressively reduced MyoD mediated activation, and mutation of E4 seriously retarded MyoD mediated activation (down to 16.1%). These observations suggest that E4 plays the most critical role in the activation of *M-cadherin* promoter by MyoD and E2 and E3 might facilitate this process, implying the possible cooperation of E-boxes during *M-cadherin* promoter activation. Since the relative activation of these E-box mutants, as compared to that of wildtype, by MyoD in cells kept in growth medium (GM) or differentiation medium (DM) is similar, it suggest that MyoD might target the same E-boxes to activate *M-cadherin* expression regardless of the culture condition or differentiation stage.

The cooperation of companion E-boxes with E4 was dissected with mutation of multiple E-boxes, and we found that double E-box mutants Mt1/2, Mt1/5, or Mt2/5 did not affect its activation but mutations of E-boxes 1/3, 2/3, 3/4, or 3/5 significantly
reduced its activation, suggesting the important facilitating role of E3 in the M-cadhrin promoter activation by MyoD (Fig. 2B). As double E-box mutations without involving E3 and E4 (mutants Mt1/2, Mt1/5, or Mt2/5) had conserved its activation, it suggests that E3 and E4 boxes form the core cis-elements recognized by MyoD and addition of any one of the rest E-boxes (up to 3 E-boxes in total) is required to retain full activation by MyoD.

We further found that simultaneous mutation of E1 or E2, but not E5, with E3 enhanced the negative effect of E3 mutation (comparing Mt1/3, Mt2/3, and Mt3/5 with Mt3, p<0.01; Fig. 2A & B), which demonstrates that, in the absence of E3, both E1 and E2 are important cis-elements for supporting the function of E4 (Fig. 2B). Therefore, it was not surprising to find that simultaneous mutation of E-boxes 1, 2, and 3 significantly enhanced the repressive effect of E3 mutation (Fig. 2C), but which was not significantly different from Mt1/3 or Mt2/3, suggesting the redundant role played by E1 and E2. We were surprised to see full activation of the mutants Mt1/2, Mt1/5, and Mt2/5 as single mutants Mt2 reduced the activation (Fig. 2B). The full activation of these 3 double mutants suggests the rescue of Mt2 by either Mt1 or Mt5, and that E-boxes 3 and 4 plus any one of the other E-boxes are sufficient for full activation by MyoD.

Triple E-box mutants Mt1/3/5 or Mt2/3/5 did not further enhance the repressive effect of Mt1/3 and Mt2/3, confirming again for the requirement of E1 or E2 when E3 was absent. The activation mediated solely by E-boxes 3 and 4 was examined by the triple mutant E1/2/5, but which only achieved about 50% activation, demonstrating again the requirement of other E-boxes for full activation (Fig. 2C). The activation mediated by single E4 or E3 was examined in quadruplet E-box mutant Mt1/2/3/5 or
Mt1/2/4/5, which was about 23% and 6.2% respectively. The activation of some of these E-box mutants was also examined in C2C12 myoblasts, and we found that Mt1/2/5, Mt1/2/3/5, and Mt1/2/4/5 all shown slightly higher activation in C2C12 (Fig. 2D) than that in 10T1/2 fibroblasts, indicating a better working environment for MyoD in myoblasts. However, the requirement of multiple E-boxes for full activation and the importance of E3 and E4 boxes were conserved in both cell types.

**E-proteins are required for full activation of *M-cadherin* promoter.**

MyoD forms heterodimers with ubiquitously expressed class I bHLH factors (or called E-proteins), including products of *E2A* (*E12*, and *E47*) and *HEB* (*HEBα* and *HEBβ*) genes (32), for high affinity binding to E-boxes in myogenic gene promoters/enhancers to trigger their expression. This heterodimerization is critical to the function of MyoD transactivation as competitive binding of an activation domain-lacking bHLH protein, Id (inhibitor of differentiation), inhibits its transactivational activity. The requirement of multiple E-boxes for activating *M-cadherin* promoter raises the possibility that multiple E-boxes might facilitate dimerization with E-proteins and bypass the inhibitory effect of Id. The expression of Id remains is high in proliferating myogenic cells kept in growth medium but is gradually down-regulated during differentiation (33, 34). We found that MyoD could highly activate the activity of both *M-cadherin* and *Myogenin*, but not *MCK*, promoters in cells kept in growth medium (Fig. 3A), demonstrating their lower sensitivity to the presence of Id. Consequently, co-transfection of E47 or expression of MyoD-E47 fusion protein did not further potentiate this activation. As expected, the activation of Mt1/2/5 was significantly lower than that of wildtype in GM (Fig. 3B).

In cells kept in differentiation medium we found that supplementation of E47
significantly enhanced MyoD mediated activation of \textit{M-cadherin}, but not \textit{MCK} and \textit{Myogenin}, promoter (Fig. 3C); however, the activation of Mt1/2/5 was still significantly lower than that of wildtype. Fusion of MyoD and E47 into a single chimeric protein significantly enhanced the activation of \textit{MCK} promoter, but only achieved an activation level lower than that of \textit{M-cadherin} promoter triggered by MyoD and E47 co-transfection. These observations suggest that \textit{M-cadherin} promoter has different requirement/sensitivity for E47 or other E-proteins before and after the initiation of terminal differentiation: before differentiation, \textit{M-cadherin} promoter has low sensitivity to E47, but after the initiation of differentiation, it becomes highly sensitive to the presence of E47, and MyoD/E47 heterodimer efficiently formed and activates \textit{M-cadherin} promoter with efficiency similar to that of MyoD-E47 chimeric protein. Since supplementation of either E47 or MyoD-E47 failed to rescue the activation of Mt1/2/5 to the level of wildtype promoter (Fig. 3B), it confirms again that binding of MyoD/E47 heterodimer to only E3/4 boxes is not sufficient for full activation of \textit{M-cadherin} promoter and multiple E-boxes might cooperate to recruit multiple MyoD/E47 heterodimers to achieve higher activation.

\textbf{E3- and E4-boxes play differential role during promoter activation}

Our previous study has shown that the E3 box in the \textit{M-cadherin} promoter is a perfect target site of the Bhlhe40 protein (15), and studies from other groups also indicate that Bhlhe40 interferes with the dimerization of MyoD and E47, resulting in the down-regulation of \textit{Myogenin}, \textit{Mef2c}, and \textit{myosin heavy chain (MHC)} and prevention of myogenic differentiation (35). Therefore, it is of interest to test whether Bhlhe40 can still repress MyoD transactivation on mutant \textit{M-cadherin} promoters. We found that Bhlhe40 mediated repression was highly dependent on the E3 as its mutation abolished this repression (Fig. 4A). On the contrary, E4 is essential to MyoD transactivation as...
it is abolished by E4 mutation. The differential dependence of MyoD and Bhlhe40 on E3 and E4 prompted us to ask whether their reciprocal mutation will enhance the activity of either transcription factor. When the sequence of E4 was changed to that of E3 in the reporter called double E3 (DB3), MyoD transactivation was significantly reduced and could be totally repressed by Bhlhe40 (Fig. 4B). Surprisingly, mutation of E3 to E4 (DB4) did not enhance but reduced MyoD transactivation, and at the same time the repression by Bhlhe40 was still reserved. Furthermore, the binding of both MyoD and Bhlhe40 to the promoter was increased in the DB3 but the binding of Bhlhe40 to DB4 probe was totally lost (Fig. 4C). These observations suggest that E4 is essential for MyoD transactivation but the presence of E3 is helpful, which is confirmed by the reduced activation of DB4 by MyoD. The high binding affinity of DB3 by MyoD and Bhlhe40 implies that E3 might be the initial binding site for transcription factors, such as Bhlhe40 and MyoD, that regulate the assembly/functions of RNA polymerase II machinery around the TIS in E4.

To further identify the role of E3, the DNA-binding domain of Bhlhe40 was co-transfected with MyoD and we found significant repression of MyoD transactivation, demonstrating the competitive binding effect of DNA-binding domain of Bhlhe40 (Fig. 4D). Surprisingly, the C-terminal region without DNA-binding activity had the same effect. As the C-terminal region of Bhlhe40 contains strong HDAC-recruiting activity, it suggests that E3 bound factors can regulate MyoD activity by both competitive binding and enzymatic activity.

**The CDE element and GC-boxes are also important for M-cadherin activation**

In addition to the 5 E-boxes in the promoter, 2 GC-boxes and a conserved downstream element (CDE) was predicted by the Genomatix software. These *cis*-elements were
deleted individually and we found these deletion mutants lost most of their activation by MyoD in C3H10T1/2 cells (Fig. 5A & B). However, the repressed activation was eased when they were transfected into C2C12 myoblasts under growth medium condition (Fig. 5C & D), suggesting some factors in proliferating myoblasts might facilitate MyoD transactivation so the requirement of GC-boxes and CDE element is reduced. Nevertheless, these observations demonstrate the important roles of GC-box and CDE element.

**Pbx1 is an important regulator of M-cadherin expression**

The CDE element was predicted as the target site of the Pbx/Meis1 heterodimer and previous studies have pointed out that Pbx/Meis1 play important role in myogenesis, especially in the activation of Myogenin expression (18). Therefore, it prompted us to examine whether Pbx/Meis1 can activate M-cadherin promoter directly. Surprisingly, no significant activation of M-cadherin promoter was observed, regardless of culture condition and cell types (Fig. 6A & Supple. fig. 2A). As Pbx/Meis1 heterodimer regulates gene expression by modulating the epigenetic modifications around target gene promoters, these observations suggest that M-cadherin promoter not integrated into chromosome might not be affected by Pbx/Meis1 heterodimer. This issues was further investigated by knocking down Pbx1 expression in stable clones of C2C12 myoblasts (C2-shPbx1) with lentivirus expressed shRNA. We found the expression of M-cadherin mRNA was concomitantly reduced significantly in C2-shPbx1 cells (Fig. 6B), confirming the requirement of Pbx/Meis1 for endogenous M-cadherin expression. Surprisingly, the expression of N-cadherin was strongly activated in Pbx-knockdown cells (C2-shPbx1), suggesting a compensation for reduced M-cadherin.

Concomitant reduction of Pbx1 and M-cadherin expression was also observed at the
protein level (Fig. 6C). Interestingly, the shRNA used here targets 3’-UTR sequence conserved in Pbx1a, -c, -d, but not -b, mRNAs, so all but b variants should be knocked down. Among all variants, Pbx1a encodes the largest protein (about 47 Kd) but both mRNA variants b and c encode the same protein Pbx1b (about 42 Kd) due to the same coding sequence is used. As the antibody used here targets the N-terminal sequence conserved in all variants, the detected protein pattern suggests that Pbx1b is the major Pbx1 variant (encoded by Pbx1b and -c) in C2C12 myoblasts and the residual lower band should be the product of un-targeted Pbx1b mRNA.

**CDE element is important for MyoD recruitment**

Although Pbx1/Mes1 heterodimer does not seem to participate in the activation of transiently transfected M-cadherin promoter, its role in the epigenetic regulation of this promoter needs to be examined. The wildtype and ΔCDE promoter were cloned into the pStable-luc vector (Supple. Fig. 1B) that allowed integration of the reporter by G418 selection (27). Stable clones of C2C12 myoblasts carrying wildtype and ΔCDE reporters were generated and the recruitment of MyoD to reporters was examined by ChIP assay. Using a primer set specific to ectopic M-cadherin promoter, we found the recruitment of MyoD to chromatin packed ΔCDE reporter was sharply reduced as compared to that of wildtype reporter (Fig. 7C). This observation confirmed the importance of CDE element in the recruitment of MyoD to M-cadherin promoter, and it also suggests that the chromatin modifying function of Pbx1/Mes1 is critical to M-cadherin promoter activation before MyoD is recruited to E-boxes to trigger transcription.

To explore the involvement of CDE in M-cadherin expression in vivo, a CRISPER/dCass9 approach was adopted. C2C12 myoblasts were over-expressed
with the fusion gene \textit{KRBAB-dCas9-P2A-mCherry}, in which P2A protease self-digestion will release the transcriptional repressor KRAB-dCas9 and the fluorescent mCherry. Over-expression of a single guide RNA (sgRNA) targeting sequence (+18~+36) immediately downstream of \textit{CDE} can recruit KRAB-dCas9 to CDE to block its function (Fig. 8A). We found that \textit{M-cadherin} was significantly repressed by KRAB-dCas9 in confluent myoblasts (Fig. 8B & C), demonstrating the importance of \textit{CDE} element in \textit{M-cadherin} transcriptional activation.
Discussion:

*M-cadherin is expressed in myogenic stem cells*

*M-cadherin* is expressed in developing myogenic precursors *in utero* and adult satellite cells before the terminal differentiation triggering genes *Myogenin* and *Mef2c* are expressed (14, 17). This early expression suggests that its regulation by MyoD and Myf5 should be different from those genes, such as *MCK*, and *MHC*, that are targeted by both Myogenin and Mef2c to drive their high expression only after the initiation of terminal differentiation. *M-cadherin* expression is also significantly increased during early terminal differentiation but gradually reduced in the late half of terminal differentiation. Shortly after birth, the expression of *M-cadherin* in SKM is down-regulated, implying its major function in myoblasts fusion into multinucleated myofibers. Surprisingly, mice with homozygous *M-cadherin* knockout are viable, fertile, and show no gross developmental defects (36), suggesting that other *Cadherins*, especially *N-cadherin*, can be activated to compensate for the function of M-cadherin. The activated *N-cadherin* expression is also seen in our *M-cadherin* knockdown study (Fig. 6B).

The early expression in quiescent satellite cells also suggests that the initial expression phase of *M-cadherin* might be triggered by other mechanisms before its induction by MyoD or Myf5 as the activity of either factor is quite low in these cells. The *CDE* element next to the TIS was supposed to mediate this function as its targeting factors Meis1/Pbx1 have been shown to open up the chromatin structure of some myogenic genes, such as *Myogenin*, before their activation by MRFs. However, our results showed that weak *M-cadherin* expression (6x10^-5 fold of m36b4 level) was induced in the tail region where both MyoD and Myf5 was also weakly expressed (1.8 x10^-4 and 1.1x10^-3 fold of m36b4 level, respectively), suggesting the involvement of MRFs in its
induction in the tail region. Actually, the embryos used in this study contained a few developing somites in their tail region, so the expression of both *MyoD* and *Myf5* was already detectable there. In the future, embryos of earlier stages or removal of developing somites in the tail region might be needed to detect *M-cadherin* expression in the presomitic mesoderm.

Knockdown of *Pbx1* in C2C12 did sharply reduce *M-cadherin*, but not *N-cadherin*, expression level (Fig. 6B); confirming its participation in regulating *M-cadherin* specifically. The compensative activation of *N-cadherin* in C2C12-shPbx cells suggests a precise coordination between these two genes to ensure that total amount of Cadherins in myoblasts is sufficient to adhere cells together. It has been shown that *M-cadherin* adhesion activates Rac1-GTPase via the Rho-GEF Trio. On the other hand, N-cadherin signaling activates RhoA-GTPase activity, which is required for Catenin accumulation under plasma membrane, but decreases Rac1-GTPase activity (11, 37, 38). Therefore, Rac1- and RhoA-GTPase might be the regulatory core mediating the compensative expression of these two genes. It will be interesting to examine if the promoters of both genes are affected directly by the signaling of Rac1- and RhoA-GTPase.

Intriguingly, *M-cadherin* promoter activity was not affected in the C2-shPbx cells in transient transfection assays, probably due to lacks of proper chromatin structure. However, stable clones of *M-cadherin-ΔCDE*-luc could not recruit MyoD efficiently to the ectopic *M-cadherin* promoter (Fig. 7), confirming that *CDE* recruited Mes1/Pbx1 plays key role in opening up chromatin structure for subsequent binding of MyoD/Myf5 in satellite cells and myoblasts. The *CDE* function *in vivo* was also demonstrated by the *sgRNA* recruited Krab-dCAS9, in which *M-cadherin* was highly repressed (Fig. 8).
It will be interesting to know whether Krab-dCAS9 recruited to other \textit{cis}-elements, such as E3 and E4, can have the same effects.

**Collaboration and substitution between E-boxes**

The myogenic bHLH domain targets MRFs to the class II (tissue-specific) E-boxes (consensus sequence: CANNTG) in the control regions of most, but not all, muscle-specific genes (39, 40). Multiple E-boxes are found in the promoters of many myogenic genes and their collaboration and substitution during transcriptional activation is an issue that has not been resolved.

The E-boxes 3 and 4 in \textit{M-cadherin} promoter have been identified as the key \textit{cis}-elements mediating its activation by MyoD in our studies. Mutation of E3 and E4 reduced the activation to 40% and 20%, respectively (Fig.2), indicating differential roles played by either E-box and can not be substituted by each other. E3 box is targeted by the repressor Bhlhe40 and its repressing effect was indeed relieved when E3 was mutated. As E4 was targeted by MyoD, it prompted us to change the sequence of E3 to that of E4 in the hope that it will increase the activation by MyoD. In contrary to our expectation, the activation of DB4 was lower than that of wildtype promoter and it was still repressible by Bhlhe40 (Fig. 4B). The lower activation might be explained by the weaker binding affinity by MyoD and it suggests a novel sequence requirement for MyoD binding to target site, in which a Bhlhe40 site and a MyoD site form an optimal activation/binding site for MyoD. Although DB3 increased the binding of both Bhlhe40 and MyoD, its activation by MyoD was very poor, indicating the irreplaceable role of E4 (Fig. 4C) and an optimal binding affinity is requirement for higher activation.
Participation of E-proteins and other bHLH factors

As the expression of Id is high in proliferating myoblasts but is low in differentiating myoblasts (33, 34), the formation of functional heterodimers between MyoD and E proteins in cells kept in GM should be difficult than those kept in DM. Therefore, it is expectable that the activation of M-cadherin promoter by MyoD was much higher in DM than in GM and supplement of E47 only showed significant effect in DM (Fig. 3). Fusion of MyoD and E47 bypassed the inhibiting effect of Id, so higher activation could be seen in GM. However, in DM, MyoD–E47 fusion protein achieved similar activation as MyoD/E47 co-transfection, indicating that heterodimerization is as efficient as both factors are fused.

It was a surprise to find that, even with MyoD–E47, the activation of M-cadherin promoter in GM was still much lower than that in DM (Fig. 3), suggesting that additional mechanisms are implicated in the full activation of M-cadherin promoter during terminal differentiation. Several transcriptional coactivators, such as p300 and p/CAF, are required for MyoD to efficiently activate downstream target genes (41, 42). Besides, other coactivators, including GRIP-1, CARM1, and PGC-1α, are necessary for Mef2C to function efficiently on target gene promoters (43-45). Since MyoD and Mef2c usually cooperate on target gene promoters to activate transcription, the amount of these coactivators might play important role in determining the efficiency of transcriptional activation. The expression of these MyoD or Mef2C associated coactivators has been documented before and most of them showed much higher expression during/after differentiation. These increased coactivators might be part of the reason for the much higher activation in DM in addition to the reduced Id competition/inhibition.
The role of Pbx1/Meis1 in myogenic gene regulation

The milestone study by Berkes et al. found that Pbx1/Meis1 heterodimer is constitutively bound to the promoter of *Myogenin* before MyoD is recruited to activate its expression, thus establishing the role of PBX1/Meis1 complex as the pioneer factor in myogenic differentiation (18). Later studies demonstrated that Pbx1/Meis1 heterodimer can penetrate repressed chromatin of myogenic genes to recruit chromatin remodeling complex and MyoD or other MRFs to activate myogenic gene expression (46). Differential recruitment of coactivator and corepressor complexes were observed between *Myogenin* and *muscle creatine kinase (MCK)* promoters, where MyoD and HDAC2 are bound to the inactive *MCK* promoter when *Myogenin* is already actively expressed under the stimulus of MyoD and coactivators (47). Taken together, these studies point out the critical role played by Pbx1/Meis1 in myogenic differentiation and this finding is further corroborated in this study by the essential requirement of the CDE element during MyoD mediated *M-cadherin* gene activation.

As myogenic lineage is determined by the expression of either *Myf5* or *MyoD* in different embryonic tissues, the binding of Pbx1/Meis1 to their important cis-elements, such as distal enhancer in *MyoD* gene, should establish the involvement of Pbx1/Meis1 in myogenic determination and that should be the next priority in this field.

There are 4 Pbx isoforms (*Pbx1*–4) in mammals and the expression of *Pbx1–3* can be found in most embryonic tissues but with that of *Pbx4* found majorly in testes. The *Pbx1*−/− homozygous mutant embryos die in utero with dramatic abnormality in multiple organs, while no detectable phenotype can be found in *Pbx2*−/− homozygous mutant mice. *Pbx3*−/− homozygous mutant mice die perinatally due to central respiratory failure (reviewed in (48)). These *Pbx* genes knockout studies suggest that distinct roles might be played by different *Pbx* isoforms during embryogenesis and their functional...
redundancy is limited. Actually, in C2C12 myoblasts, Pbx1 showed the strongest expression and followed by that of Pbx2 and -3, and much weaker level of Pbx4 was found (Supple. fig. 2B), indicating that Pbx1 is the major isoform in C2C12 myoblasts. However, the involvement of other isoforms in myogenesis remains to be determined. In the near future, isoform-specific shRNA or KRAB-dCas9 mediated gene knockdown approach should be employed to specifically reduce the expression of each isoform so their function in myogenic gene activation can be revealed.

Acknowledgements:
The authors like to thank Dr. Hsueh Yi-Ping (Institute of Molecular Biology, Academia Sinica) and Drs. Won-Jin Wang and Wei-Yi Chen (Institute of Biochemistry and Molecular Biology, National Yang-Min University) for their assistance during the revision of this manuscript. Yung-Jui Lin was supported by the undergraduate student research fund from the Ministry of Science and Technology, Taiwan, ROC. This study was supported by funding from the Ministry of Science and Technology, Taiwan, ROC to Shen Liang Chen [MOST 107-2320-B-008-001 and MOST 108-2311-B-008-002]. The authors have no conflict of interest to declare.
References:

1. Tajbakhsh S. Skeletal muscle stem cells in developmental versus regenerative myogenesis. Journal of internal medicine. 2009;266(4):372-89.
2. Pownall ME, Gustafsson MK, Emerson CP, Jr. Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. Annu Rev Cell Dev Biol. 2002;18:747-83.
3. Krauss RS, Cole F, Gaio U, Takaesu G, Zhang W, Kang JS. Close encounters: regulation of vertebrate skeletal myogenesis by cell-cell contact. J Cell Sci. 2005;118(Pt 11):2355-62.
4. Goodwin M, Yap AS. Classical cadherin adhesion molecules: coordinating cell adhesion, signaling and the cytoskeleton. J Mol Histol. 2004;35(8-9):839-44.
5. Tepass U, Truong K, Godt D, Ikura M, Peifer M. Cadherins in embryonic and neural morphogenesis. Nat Rev Mol Cell Biol. 2000;1(2):91-100.
6. Yaffe D, Feldman M. The Formation of Hybrid Multinucleated Muscle Fibers from Myoblasts of Different Genetic Origin. Dev Biol. 1965;11:300-17.
7. Cole F, Zhang W, Geyra A, Kang JS, Krauss RS. Positive regulation of myogenic bHLH factors and skeletal muscle development by the cell surface receptor CDO. Dev Cell. 2004;7(6):843-54.
8. Wrobel E, Brzoska E, Moraczewski J. M-cadherin and beta-catenin participate in differentiation of rat satellite cells. Eur J Cell Biol. 2007;86(2):99-109.
9. George-Weinstein M, Gerhart J, Blitz J, Simak E, Knudsen KA. N-cadherin promotes the commitment and differentiation of skeletal muscle precursor cells. Dev Biol. 1997;185(1):14-24.
10. Knudsen KA, Myers L, McElwee SA. A role for the Ca2(+)-dependent adhesion molecule, N-cadherin, in myoblast interaction during myogenesis. Exp Cell Res. 1990;188(2):175-84.
11. Charrasse S, Meriane M, Comunale F, Blangy A, Gauthier-Rouviere C. N-cadherin-dependent cell-cell contact regulates Rho GTPases and beta-catenin localization in mouse C2C12 myoblasts. J Cell Biol. 2002;158(5):953-65.
12. Zeschnigk M, Kozian D, Kuch C, Schmoll M, Starzinski-Powitz A. Involvement of M-cadherin in terminal differentiation of skeletal muscle cells. J Cell Sci. 1995;108 (Pt 9):2973-81.
13. Pouliot Y, Gravel M, Holland PC. Developmental regulation of M-cadherin in the terminal differentiation of skeletal myoblasts. Dev Dyn. 1994;200(4):305-12.
14. Donalies M, Cramer M, Ringwald M, Starzinski-Powitz A. Expression of M-cadherin, a member of the cadherin multigene family, correlates with differentiation of skeletal muscle cells. Proc Natl Acad Sci U S A. 1991;88(18):8024-8.
15. Hsiao SP, Chen SL. Myogenic regulatory factors regulate M-cadherin expression
by targeting its proximal promoter elements. Biochem J. 2010;428(2):223-33.

16. Rose O, Rohwedel J, Reinhardt S, Bachmann M, Cramer M, Rotter M, et al. Expression of M-cadherin protein in myogenic cells during prenatal mouse development and differentiation of embryonic stem cells in culture. Dev Dyn. 1994;201(3):245-59.

17. Moore R, Walsh FS. The cell adhesion molecule M-cadherin is specifically expressed in developing and regenerating, but not denervated skeletal muscle. Development. 1993;117(4):1409-20.

18. Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell. 2004;14(4):465-77.

19. Longobardi E, Penkov D, Mateos D, De Florian G, Torres M, Blasi F. Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. Dev Dyn. 2014;243(1):59-75.

20. Moens CB, Selleri L. Hox cofactors in vertebrate development. Dev Biol. 2006;291(2):193-206.

21. Heidt AB, Rojas A, Harris IS, Black BL. Determinants of myogenic specificity within MyoD are required for noncanonical E box binding. Mol Cell Biol. 2007;27(16):5910-20.

22. Knoepfler PS, Bergstrom DA, Uetsuki T, Dac-Korytko I, Sun YH, Wright WE, et al. A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic Acids Res. 1999;27(18):3752-61.

23. Irintchev A, Zeschning M, Starzinski-Powitz A, Wernig A. Expression pattern of M-cadherin in normal, denervated, and regenerating mouse muscles. Dev Dyn. 1994;199(4):326-37.

24. Teng HF, Kuo YL, Loo MR, Li CL, Chu TW, Suo H, et al. Valproic acid enhances Oct4 promoter activity in myogenic cells. J Cell Biochem. 2010;110(4):995-1004.

25. Lu TC, Zhao GH, Chen YY, Chien CY, Huang CH, Lin KH, et al. Transduction of Recombinant M3-p53-R12 Protein Enhances Human Leukemia Cell Apoptosis. J Cancer. 2016;7(10):1360-73.

26. Chung SY, Huang WC, Su CW, Lee KW, Chi HC, Lin CT, et al. FoxO6 and PGC-1alpha form a regulatory loop in myogenic cells. Bioscience reports. 2013;33(3).

27. Lang KC, Lin IH, Teng HF, Huang YC, Li CL, Tang KT, et al. Simultaneous overexpression of Oct4 and Nanog abrogates terminal myogenesis. Am J Physiol Cell Physiol. 2009;297(1):C43-54.

28. Teng HF, Li PN, Hou DR, Liu SW, Lin CT, Loo MR, et al. Valproic acid enhances Oct4 promoter activity through PI3K/Akt/mTOR pathway activated nuclear receptors. Mol
29. Chung SY, Kao CH, Villarroya F, Chang HY, Chang HC, Hsiao SP, et al. Bhlhe40 Represses PGC-1alpha Activity on Metabolic Gene Promoters in Myogenic Cells. Mol Cell Biol. 2015;35(14):2518-29.

30. Pan YC, Wang XW, Teng HF, Wu YJ, Chang HC, Chen SL. Wnt3a signal pathways activate MyoD expression by targeting cis-elements inside and outside its distal enhancer. Biosci Rep. 2015;35(2).

31. Chang HC, Kao CH, Chung SY, Chen WC, Aninda LP, Chen YH, et al. Bhlhe40 differentially regulates the function and number of peroxisomes and mitochondria in myogenic cells. Redox Biol. 2019;20:321-33.

32. Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, et al. Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell. 1991;66(2):305-15.

33. Kumar D, Shadrach JL, Wagers AJ, Lassar AB. Id3 is a direct transcriptional target of Pax7 in quiescent satellite cells. Mol Biol Cell. 2009;20(14):3170-7.

34. Jen Y, Weintraub H, Benezra R. Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. Genes Dev. 1992;6(8):1466-79.

35. Azmi S, Ozog A, Taneja R. Sharp-1/DEC2 inhibits skeletal muscle differentiation through repression of myogenic transcription factors. J Biol Chem. 2004;279(50):52643-52.

36. Hollnagel A, Grund C, Franke WW, Arnold HH. The cell adhesion molecule M-cadherin is not essential for muscle development and regeneration. Mol Cell Biol. 2002;22(13):4760-70.

37. Charrasse S, Comunale F, Fortier M, Portales-Casamar E, Debat A, Gauthier-Rouviere C. M-cadherin activates Rac1 GTPase through the Rho-GEF trio during myoblast fusion. Mol Biol Cell. 2007;18(5):1734-43.

38. Charrasse S, Comunale F, Grumbach Y, Poulat F, Blangy A, Gauthier-Rouviere C. RhoA GTPase regulates M-cadherin activity and myoblast fusion. Mol Biol Cell. 2006;17(2):749-59.

39. Buskin JN, Hauschka SD. Identification of a myocyte nuclear factor that binds to the muscle-specific enhancer of the mouse muscle creatine kinase gene. Mol Cell Biol. 1989;9(6):2627-40.

40. Lassar AB, Buskin JN, Lockshon D, Davis RL, Apone S, Hauschka SD, 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. 1989;58(5):823-31.

41. Sartorelli V, Puri PL, Hamamori Y, Ogryzko V, Chung G, Nakatani Y, et al. Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program. Mol
Cell. 1999;4(5):725-34.

42. Yuan W, Condorelli G, Caruso M, Felsani A, Giordano A. Human p300 protein is a coactivator for the transcription factor MyoD. J Biol Chem. 1996;271(15):9009-13.

43. Chang JH, Lin KH, Shih CH, Chang YJ, Chi HC, Chen SL. Myogenic basic helix-loop-helix proteins regulate the expression of peroxisomal proliferator activated receptor-gamma coactivator-1alpha. Endocrinology. 2006;147(6):3093-106.

44. Chen SL, Wang SC, Hosking B, Muscat GE. Subcellular localization of the steroid receptor coactivators (SRCs) and MEF2 in muscle and rhabdomyosarcoma cells. Mol Endocrinol. 2001;15(5):783-96.

45. Chen SL, Dowhan DH, Hosking BM, Muscat GE. The steroid receptor coactivator, GRIP-1, is necessary for MEF-2C-dependent gene expression and skeletal muscle differentiation. Genes Dev. 2000;14(10):1209-28.

46. de la Serna IL, Ohkawa Y, Berkes CA, Bergstrom DA, Dacwag CS, Tapscott SJ, et al. MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. Mol Cell Biol. 2005;25(10):3997-4009.

47. Cho OH, Mallappa C, Hernandez-Hernandez JM, Rivera-Perez JA, Imbalzano AN. Contrasting roles for MyoD in organizing myogenic promoter structures during embryonic skeletal muscle development. Dev Dyn. 2015;244(1):43-55.

48. Selleri L, Zappavigna V, Ferretti E. 'Building a perfect body': control of vertebrate organogenesis by PBX-dependent regulatory networks. Genes Dev. 2019;33(5-6):258-75.
Figure legends:

**Figure 1: M-cadherin regulators are activated during myogenesis**

(A) Schematics showing the locations of cis-elements in the mouse M-cadherin promoter (-252~+200). E1~E5: E-boxes; GC: GC box; CDE: conserved downstream element. (B) Alignment of several mammalian M-cadherin promoter sequences around transcriptional initiation site (TIS, +1). The expression levels of M-cadherin regulators during differentiation of C2C12 myoblasts (C) and embryogenesis (D) were determined by qRT-PCR. In C2C12 myoblasts, the expression level of each gene in proliferating myoblasts (PMB) was set as 1 fold expression. The morphology of cells at PMB, confluent myoblast (CMB), and myotubes (MT) stages are shown at the left panel in (C). The embryos (11.5~12.5 dpc, left panel in (D)) were dissected into head, rostral, interlimb and tail parts and the expression level of each gene in tail was set as 1 fold. **: p<0.01 vs. PMB or tail.

**Figure 2: Multiple E-boxes are required for full activation of M-cadherin expression**

Wildtype and E-box mutant M-cadherin promoters were transfected into C3H10T1/2 (A~C) and C2C12 (D) cells in the absence or presence of MyoD-expression vector overnight and their luciferase activity was determined after kept in growth or differentiation medium (GM or DM as indicated) for 24 hr. The activity of each promoter in the presence of GFP was set as the basal level and its activation by MyoD was compared with that of the wildtype promoter (Wt, set as 100% activation). * and **: p<0.05 and p<0.01 respectively vs. wildtype.
Figure 3: E-proteins are required for full activation of M-cadherin promoter.

Myogenic gene promoters were co-transfected with MyoD, MyoD and E47, or MyoD–E47 fusion gene expression vector into C3H10T1/2 cells and kept in GM (A, B) or DM (C) for 24 hr before harvested for determining luciferase activity. In (B), the wildtype promoter activation was set 100% activation. * and **: p<0.05 and p<0.01 respectively vs. GFP in (A, C) or vs. wt in (B).

Figure 4: E3- and E4-boxes play differential roles during promoter activation

(A, B) M-cadherin promoters were co-transfected with MyoD, Bhlhe40, or both, into C3H10T1/2 cells and kept in DM for 24 hr before harvested for determining luciferase activity. The sequences of double E3 (DB3) and double E4 (DB4) mutants are shown at the top panel of (B). (C) Bacteria expressed GST-MyoD and MBP-Bhlhe40 proteins were used to bind 32P-labeled M-cadherin promoter probes in the EMSA assay. (D) M-cadherin promoters were co-transfected with expression vectors for MyoD, Bhlhe40 mutants, or both, into C3H10T1/2 cells and kept in DM for 24 hr before harvested for determining luciferase activity. * and **: p<0.05 and p<0.01 respectively vs. GFP. #: p<0.05 vs. wildtype activation by MyoD.

Figure 5: Both GC-boxes and CDE elements are critical for activation of M-cadherin expression

Wildtype and mutant M-cadherin promoters were transfected into C3H10T1/2 (A–B) and C2C12 (C–D) cells in the absence or presence of MyoD-expression vector overnight and their luciferase activity was determined after kept in growth or differentiation medium (GM or DM) for 24 hr. The activity of each promoter in the presence of GFP was set as 1 fold activation. GC: GC-box deleted; DGC: 2 GC-boxes deleted; CDE: CDE element deleted. * and **: p<0.05 and p<0.01 vs. GFP. # and
Figure 6: Pbx1/Meis1 heterodimer is important to M-cadherin expression

(A) Wildtype M-cadherin promoter was co-transfected with expression vector of GFP, MyoD, Pbx, or Meis1 into C3H10T1/2 and C2C12 cells overnight and their luciferase activity was determined after kept in growth medium (GM) for 24 hr. The activity of each promoter in the presence of GFP was set as 1 fold activation. * and **: p<0.05 and p<0.01 vs. GFP. The mRNA of Pbx1 was knocked down by lentivirus expressing shRNA (5’-CGAAGCAATCAGCAAACACAA-3’) and the expression of Pbx1, M-cadherin, N-cadherin, and Meis1 in C2-Control and –shPbx1 cells was determined by qRT-PCR (B). The protein levels of Pbx1 and M-cadherin were also determined by Western blot (C). The antibody for Pbx1 is from Cell Signaling (#4342, Cell Signaling). **: p<0.01 vs. C2-Control.

Figure 7: The CDE element is important for MyoD recruitment to M-cadherin

(A) Schematics describing the locations of vector-specific primer RV3 and primers 013005/013004 for M-cadherin promoter (-252~+200) used in ChIP assay. RV3/013004 amplifies ecotopic M-cadherin promoter, and 003005/003004 amplifies ecotopic and endogenous (all) M-cadherin promoters. Both wildtype and ΔCDE promoters in pStable-luc vector were stably integrated into C2C12 chromosomes and their recruitment of MyoD was examined by ChIP assay. ChIP assay products were analyzed with both traditional (B) and quantitative (C) PCRs. **: P<0.01 vs. wildtype.

Figure 8: M-cadherin expression is repressed by cis-element recruited repressor

(A) A schematics showing the recruited KRAB-dCas9 repressor to the sequence (+18~+36) immediately downstream of the CDE element via binding to the single guide
RNA (sgRNA). KRAB-dCas9-P2A-mCherry was expressed as a single fusion protein by lentivirus infection but later digested into 3 proteins (KRAB-dCas9, P2A, and mCherry) by the self-digestion of the P2A protease. Cells expressing mCherry are transfected with the vector expressing sgRNA and selected with puromycin to form stable clones. The images of stable clone of cells with empty vector or sgRNA expression are shown in (B), and their mRNA levels of dCas9, mCherry, and M-cadherin are shown in (C). **: p<0.01 vs. vector control.

Supplementary figure 1: The sequence and reporter map of mouse M-cadherin proximal promoter

The sequence of mouse M-cadherin proximal promoter (-300~+200) is shown in (A) and the plasmid map of pStable-M-cadherin-luc reporter, in which M-cadherin promoter (-252~+200) was inserted into HindIII site of pStable-luc vector, is shown in (B). Neo’: the neomycin-resistant gene expression cassette.

Supplementary figure 2: Effects of Pbx1/Meis1 on M-cadherin expression

(A) Wildtype M-cadherin promoter was co-transfected with expression vector of GFP, MyoD, Pbx, or Meis1 into C2C12 cells overnight and their luciferase activity was determined after kept in growth medium (GM) for 24 hr. The activity of each promoter in the presence of GFP was set as 1 fold activation. * and **: p<0.05 and p<0.01 vs. GFP. (B) The expression of Pbx1~4 isoforms in confluent C2C12 myoblasts was determined by qRT-PCR and the results are shown as levels relative to that of m36b4. The Y-axis is shown in log scale due to the large difference between Pbx1~4 expression levels. **: p<0.01 vs. Pbx1.
Fig. 1

(C) PMB
CMB
MT

(D) head
tail
inter-limb
rostral

** PMB
CMB
MT

Relative mRNA level

M-cad
Pbx1
Meis1
Sp1
MyoD
Myf5

(****)

Relative mRNA level

M-cad
N-cad
Pbx1
Meis1
MyoD
Myf5

(****)

Downloaded from http://portlandpress.com/biochemj/article-pdf/doi/10.1042/BCJ20200535/903371/bcj-2020-0535.pdf by guest on 05 February 2021

Biochemical Journal. This is an Accepted Manuscript. You are encouraged to use the Version of Record that, when published, will replace this version. The most up-to-date version is available at https://doi.org/10.1042/BCJ20200535
Fig. 2

(A) Activation by MyoD (%) for different samples:
- WT
- 3.7k
- Mt1
- Mt2
- Mt3
- Mt4
- Mt5

(B) Activation by MyoD (%) for 10T-DM:
- WT
- Mt1/2
- Mt1/3
- Mt1/5
- Mt2/3
- Mt2/5
- Mt3/4
- Mt3/5

Bars with asterisks indicate significant differences between groups.

* p < 0.05
** p < 0.01
*** p < 0.001
**** p < 0.0001
Fig. 3

(A) Fold activation by MyoD

- Vector
- MyoD
- E47
- MyoD+E47
- MyoD~E47

10T- GM

M-cadherin
MCK
Myogenin

(B) Activation by MyoD (%)

- Wt
- Mt1/2/5

10T- GM
Fold activation by MyoD

- ** ** **
- ** **
- **
- **
- **

Fig. 3

10T-DM
Fig. 4
Fig. 4

(B)
| Probe          | Wt (-52~+24) | DB4 (-52~+24) | DB3 (-52~+24) |
|----------------|--------------|--------------|--------------|
| GST            | -            | +            | -            |
| GST-MyoD       | -            | -            | +            |
| MBP            | -            | +            | -            |
| MBP-Bhlhe40    | -            | +            | -            |

MyoD binds 2 boxes
Bhlhe40 binds E3 only

Free probe
Fig. 4

Fold activation by MyoD

Bhlhe40

1-135

136-412

- -

- -

- -

- -

- -

++

++

++

**

**

##

---

Downloaded from http://portlandpress.com/biochemj/article-pdf/doi/10.1042/BCJ20200535/903371/bcj-2020-0535.pdf by guest on 05 February 2021
Fig. 5

(A) Fold activation of WT, ΔGC1, ΔGC1/2, and ΔCDE in 10T/DM cells. (B) Fold activation in 10T/GM cells. (C) Fold activation in C2C12/DM cells. (D) Fold activation in C2C12/GM cells.

WT ∆GC1 ∆GC1/2 ∆CDE

GFP MyoD

Fold activation

0 5 10 15 20 25 30 35 40 45 50

(A) 10T/DM

(B) 10T/GM

(C) C2C12/DM

(D) C2C12/GM

WT ∆GC1 ∆GC1/2 ∆CDE

GFP MyoD

Fold activation

0 5 10 15 20 25 30 35 40 45 50

* * * * * * * * * *
**Fig. 6**

(A) Fold of activation

- GFP
- MyoD
- Pbx1
- Meis1

(B) Relative mRNA level

- C2-Control
- C2-ShPbx1

(C) Western blot analysis

- shRNA
- Control 1
- Control 2

- M-cadherin
- Pbx1a
- Pbx1b/c
- Gapdh
Fig. 7

(A) M-cad-P and Luciferase

(B) Reporter

WT ΔCDE
Ecotopic All

(C) Relative binding

WT ΔCDE
Ecotopic All

**
Fig. 8

(A) Schematic diagram of the CRISPR-Cas9 system targeting the M-cadherin gene. The sgRNA (sgRNA) is designed to cleave the M-cadherin gene at the indicated genomic locations. M-cadherin +1, E3, E4, CDE, and +18~+36 are the reference sequences.

(B) Fluorescence microscopy images of C2-KRAB-dCas9-P2A-mCherry transfected with vector or sgRNA. The images show the expression of mCherry in the cells. Bright field images are also provided for comparison.

(C) Bar graph showing the relative mRNA levels of dCas9, mCherry, and M-cadherin. The graph compares the expression levels between the vector and sgRNA groups. Statistical significance is indicated by **.

**Relative mRNA level**
Supplementary Fig. 1

(A) M-cadherin proximal promoter

-300 GGAGCCCT  ACAACACATG  TGCTGACCTC  AAAACTTTAC  AGAGCCTCCC
-250  TCAGGCACTG  ATCA[CATATG]  CTGTCACTCA  GATCGCCCCC  CCCCCCAT
-200  CACTTTGCTC  TGGCTCTGCC  TTGATCTTTT  CACACTACAC  TCTTGTTGGCT
-150  GTCCAGGCCC  ACCTA[CATGT]  GTGTCACTC  ATCTCTGATC  CCGCCCATGA
-100  CTCTGCCCCC  CATGCAGTCA  CTCAGCCCTA  GCCCTGCCCA  GA[CCCCGCCC]
-50   TGG[CCCGCC]  CCACCACA[CA]  CGTGCACTCC  CTTGGCCTGC  TCCCCAAGCA
+1   [CTTG]GCTGT  CACTCA[GCC]  AATGCTCTCA  GCGTGGGGGT  AGAAGAAGCA
+51  CCACCCGGGG  TTCCACCGGC  CGCCCCAGTCG  GC[ACTTG]C  CTCGGCTCTT
+101 CCGCCACTGC  CTCAATGGGT  TCTGCTCTGC  TCCTCGCCCT  CGGGCTGCTT
+151 GCCCAGGTAA  GCTTGTAAGC  CCTGTGCGAT  CCCTCCTCT  AGCTTTGGGA

E1  E2  E3  E4  E5  GC1  GC2  CDE
Supplementary Fig. 1

(B)
