Supplementary Information for

A Long Noncoding RNA, LncMyoD, Modulates Chromatin Accessibility to Regulate Muscle Stem Cell Myogenic Lineage Progression

Anqi Dong, Christopher B. Preusch, Wai-Kin So, Kangning Lin, Shaoyuan Luan, Ran Yi, Joyce W. Wong, Zhenguo Wu, Tom H. Cheung

a. Division of Life Science, Center for Stem Cell Research, Center for Systems Biology and Human Diseases, The State Key Lab in Molecular Neuroscience, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, 000000, Hong Kong, China.
b. Institute for Advanced Study, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, 000000, Hong Kong, China.
c. Molecular Neuroscience Center, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, 000000, Hong Kong, China.
d. Guangdong Provincial Key Laboratory of Brain Science, Disease and Drug Development, The Hong Kong University of Science and Technology, Shenzhen Research Institute, Shenzhen-Hong Kong Institute of Brain Science, 518057, Shenzhen, Guangdong, China

1 Correspondence: Tom H. Cheung
Email: tcheung@ust.hk

This PDF file includes:
Supplementary text
Figures S1 to S6
SI References

Other supplementary materials for this manuscript include the following:
Datasets S1 to S3
Supplementary Information Text
This is the extended description of the Materials and Methods section.

Materials and Methods

Cell Culture
C2C12 myoblasts, C3H 10T1/2 fibroblasts and HEK-293T cells were cultured in Dulbecco’s modified Eagle medium (DMEM, Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS, Invitrogen) and penicillin-streptomycin (P/S, Invitrogen). Freshly isolated satellite cells were cultured in Ham’s F-10 (F10, Sigma-Aldrich) supplemented with 10% horse serum (HS, Invitrogen) and P/S. Primary myoblasts were derived from freshly isolated satellite cells and cultured in DMEM supplemented with 20% FBS, basic fibroblast growth factor (bFGF, PeproTech), and P/S. Satellite cells were induced to differentiate in F10 supplemented with 2% HS and P/S. C2C12 myoblasts and adenovirus infected C3H 10T1/2 fibroblasts were induced to differentiate in DMEM supplemented with 2% HS and P/S. The medium was replenished every 24 h.

Satellite Cell Isolation
Satellite cells (SCs) were isolated as previously described (1). In brief, mouse hindlimb muscles were dissected and minced carefully. Muscles were digested with collagenase II (800 U/mL, Worthington Biochemical) in F10 with 10% HS and P/S at 37°C for 90 min. Digested muscles were then washed and incubated with dispase (1.1 U/mL, Thermo Fisher Scientific) and collagenase II (100 U/mL) at 37°C for 30 min for second digestion. Muscle suspension was syringed 10 times through a 20-G needle and subsequently filtered through a 40-μm strainer. SCs were purified from the cell suspension through negative selection for Sca1-APC, CD31-FITC, and CD45-FITC and positive selection for Vcam1-Biotin amplified by Streptavidin-PE-Cy7 using an Influx cell sorter.

ChIPmentation
ChIP-seq was performed according to the ChIPmentation protocol with minor modifications (2). In brief, 2-4 × 10^5 FACS-isolated SCs were fixed in 1% formaldehyde for 10 min at room temperature with rotation to cross-link chromatin proteins and DNA. Cross-linking was stopped by incubation with 0.125 M glycine for 5 min at room temperature. The cells were washed twice with ice-cold phosphate-buffered saline (PBS) containing 10 mg/mL bovine serum albumin (BSA) and lysed in sonication buffer (10 mM Tris-HCl [pH 8.0], 2 mM EDTA, 0.25% SDS, and protease inhibitor cocktail). The lysates were sonicated by a Covaris S220 Focused-ultrasonicator (Covaris) to shear genomic DNA into 200-600 bp fragments. The lysates were centrifuged to remove debris and were then diluted 1:1.5 in equilibration buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1.67% Triton X-100, 0.17% sodium deoxycholate, 233 mM NaCl, and protease inhibitor cocktail). Sonicated chromatin was pre-cleared with Dynabead Protein A/G (Invitrogen) at 4°C for 2 h. Antibody-protein A/G complexes were prepared by incubating Dynabead Protein A/G with antibodies against H3K4me3 (Abcam), H3K27me3 (Active Motif), H3K27ac (Active motif), or MyoD (Santa Cruz) at 4°C for 3 h. The complexes were incubated with the pre-cleared chromatin overnight at 4°C with rotation to immunoprecipitate the designated histone marks or MyoD. The immuno-complexes were washed for 5 min at 4°C with the following buffers: RIPA-low-salt wash buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 140 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, and 1% Triton X-100), RIPA-high-salt wash buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 500 mM NaCl, 0.1% SDS, 0.1% sodium deoxycholate, and 1% Triton X-100), RIPA-LiCl wash buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 250 mM LiCl, 0.5% Nonidet P-40, and 0.5% sodium deoxycholate), and TE buffer (10 mM Tris-HCl [pH 8.0] and 1 mM EDTA). To perform tagmentation directly on the bead-bound immunoprecipitated DNA, immuno-complexes were tagged in 10 μL tagmentation reaction containing 1 μL Tn5 (Vazyme) at 55°C for 10 min. Tn5 transposase was used to cleave double-stranded DNA and ligate adaptors at both ends. Tagmentation was stopped by adding 1 μL 1% SDS. Cross-linking was reversed and the tagged DNA was eluted in elution buffer (10 mM Tris-HCl [pH 8.0], 5 mM EDTA, 300 mM NaCl, 0.4% SDS, and proteinase K) at 55°C for 1 h, followed by 6-hour elution at 65°C. The eluted DNA was purified using Agencourt AMPure beads (Beckman Coulter). To prepare input DNA for each ChIP sample, the corresponding pre-cleared DNA also underwent cross-link reversal and 2.5 ng DNA was tagged at 55°C for 5 min in a 5-μL tagmentation reaction containing 1 μL 10× diluted Tn5. To prepare ChIP and input libraries, tagmented immunoprecipitated DNA and input DNA were amplified by PCR, each with a unique index incorporated. Libraries were selected by size using Agencourt AMPure beads. Fragments from 300-700 bp were pooled and sequenced on an Illumina Nextseq 500 (Illumina).
ATAC-seq
The ATAC-seq protocol for SCs was adapted from Buenrostro et al. (3). Approximately $1 \times 10^6$ freshly isolated or cultured SCs were collected by centrifugation at 500 × g for 5 min at 4°C and then washed in 50 μL cold PBS. Cells were collected by centrifugation at 500 × g for 5 min at 4°C and resuspended in 50 μL cold lysis buffer (10mM NaCl, 3mM MgCl₂, 0.1% NP-40, and 10mM Tris-Cl [pH 7.4]). Cell lysate was collected by centrifugation at 500 × g for 10 min at 4°C. Next, 22.5 μL UltraPure Distilled Water, 25 μL Tagment DNA Buffer, and 2.5 μL Tagment DNA Enzyme 1 (Illumina) were added to the cell lysate for tagmentation reaction. Samples were then incubated at 37°C for 30 min, followed by purification using an EconoSpin micro-volume DNA/RNA spin column (Epoch Life Science) in final elution volume of 10 μL. The subsequent PCR was set up with 10 μL transposed DNA, 25 μL NEBNext High-Fidelity 2× PCR Master Mix (NEB), and 1.25 μM ATAC-seq oligos. The PCR reaction was performed as described by Buenrostro et al. (3) for five cycles with odd or even probes and incubate at 37°C overnight. After the hybridization reaction is complete, Streptavidin magnetic C1 beads (Life Tech) was added into each ChIRP reaction and incubate at 37°C for 30 min. The beads were then washed with wash buffer (2x SSC, 0.5% SDS and protease inhibitor cocktail) for 5 times at 37°C. LncMyoD-associated chromatin were eluted in DNA elution buffer (50mM Tris-Cl [pH 7.4], 10mM EDTA, 1% SDS, protease inhibitor cocktail and RNase inhibitor) and sonicated by a Covaris S220 Focused-ultrasonicator (Covaris) to shear genomic DNA into 200-600 bp fragments. The lysates were centrifuged to remove debris and separated into two halves for odd and even probe pairs. Lysates were then added into hybridization buffer (750mM NaCl, 50mM Tris-Cl [pH 7], 1mM EDTA, 1% SDS, 15% Formamide, protease inhibitor cocktail and RNase inhibitor) with odd or even probes and incubate at 37°C overnight. After the hybridization reaction is complete, Streptavidin magnetic C1 beads (Life Tech) was added into each ChIRP reaction and incubate at 37°C for 30 min. The beads were then washed with wash buffer (2x SSC, 0.5% SDS and protease inhibitor cocktail) for 5 times at 37°C. LncMyoD-associated chromatin were eluted in DNA elution buffer (50mM NaHCO₃, 1% SDS) supplemented with RNase A and RNase H and crosslinking was reversed by adding Proteinase K into the reaction and incubate at 50°C for 45 min. Eluted DNA was extracted using an EconoSpin micro-volume DNA/RNA spin column (Epoch Life Science). Extracted DNA and input DNA were tagmented at 55°C for 5 min in a 5-μL tagmentation reaction containing 1 μL 10× diluted Tn5. To prepare ChIRP and input libraries, tagmented DNA were amplified by PCR, each with a unique index incorporated. Libraries were selected by size using Agencourt AMPure beads. Fragments from 300-700 bp were pooled and sequenced on a DNBSEQ-G400 sequencer (BGI).

5′ RACE and 3′ RACE
Both 5′ and 3′ RACE experiments were performed using the SMARTer RACE cDNA Amplification Kit (Clontech) according to the manufacturer’s instructions. RNA was extracted from C2C12 myoblast culture and used to determine the full length of LncMyoD transcript. The following primers were used:
5′ RACE: CCGGCGCCATGTGTGAGCTGCCC;
3′ RACE: CCCCATTTGTTACGGAATGTCAAGAGGG

RNA FISH
Probes for LncMyoD FISH were generated using the FISH Tag RNA Green Kit (Invitrogen) according to the manufacturer’s instructions. In brief, LncMyoD anti-sense RNA was in vitro transcribed from LncMyoD-containing plasmid using Sp6 promoter and amine-modified. Modified RNA was purified by ethanol precipitation and further labeled with fluorescent dye (Fluor 488). Labeled RNA was purified by ethanol precipitation for subsequent hybridization. RNA FISH was performed using Stellaris RNA FISH (Biosearch Technology) according to the manufacturer’s instructions. In brief, SCs cultured for different time were fixed with fixation buffer (3.7% [v/v] formaldehyde in 1× PBS) for 10 min at room temperature. Cells were then washed twice with 1× PBS and permeabilized in 70% [v/v] ethanol overnight at 4°C. For hybridization to RNA probes, cells were first rehydrated twice in wash buffer (10% formamide in 2× SSC buffer) at room temperature and then hybridized in hybridization buffer (100 mg/mL dextran sulfate and 10% formamide in 2× SSC buffer) with LncMyoD probe overnight at 37°C. Before imaging, cells were washed three times in wash buffer and stained with DAPI in 2× SSC buffer. Cells were subsequently washed in 2× SSC buffers and mounted for microscopic examination.
Cell fractionation
The cell fractionation was performed based on the protocol from Yu et al. (5). Cells were first lysed with PBS supplemented with 0.1% NP-40, protease inhibitor cocktail and RNase inhibitor. Supernatant was collected after brief centrifugation as the cytoplasmic fraction. The remaining pellet, following additional washing, was considered the nuclear fraction. RNAs were extracted using Direct-zol™ RNA MiniPrep kit (Zymo Research) for further analysis.

RNA Interference
All siRNAs were obtained from Dharmacon. FACS-sorted cells, C2C12 cells, and primary myoblasts were transfected with siRNA targeting LncMyoD using Lipofectamine 3000 reagent (Invitrogen) according to the manufacturer’s instructions. RNA was harvested 48 h after siRNA treatment using a NucleoSpin RNA XS kit (Clontech) for further analysis.

Immunofluorescence
For immunofluorescence microscopy, cells were grown on four-well chamber slides (SPL Life Science). Cells were rinsed in PBS and fixed in 2% paraformaldehyde for 5 min. Samples were then rinsed in PBS, permeabilized with 0.1% PBST, and blocked with 5% goat serum in 0.1% PBST for 30 min. Samples were then incubated with primary antibodies in 0.1% PBST overnight at 4°C. On the following day, samples were washed in 0.1% PBST and incubated with DAPI and secondary antibodies in 0.1% PBST at room temperature for 30 min. Samples were then washed in 0.1% PBST and mounted for microscopic examination.

Plasmid Construction
Myod1, Myogenin, Myf5, Mrf4, LncMyoD, and MyoD functional domains were cloned from the cDNA of C2C12 cells. The Gal4 DNA binding domain (Gal4-DBD) was cloned from the psG5-Gal4 plasmid. Gal4-DBD fused MRFs and MyoD truncations were cloned into the pcDNA3.0 vector for plasmid transfection. Myod1 was cloned into the pGEX-4T-1 vector for protein purification. The pGEX-4T-1 plasmid was a generous gift from Prof. Mingjie Zhang (HKUST). LncMyoD was cloned into pcDNA3.0 for in vitro transcription.

In brief, the pSG5-Gal4DBD plasmid was digested by BglII restriction enzyme as a backbone. Four MRFs were cloned into pSG5-Gal4DBD using the primers listed in Dataset S3 with an additional 20 nucleotides upstream for Gibson assembly. The pcDNA3.0 plasmid was digested with EcoRI and XbaI restriction enzymes. Then, Gal4-DBD-fused MRFs were cloned from the pSG5 vector into pcDNA3.0 vector using Gal4-U primers. For GST protein purification, the pGEX-4T-1 plasmid was digested with Xhol restriction enzyme. Myod1 was cloned into the backbone using the primers listed in Dataset S3 with an additional 20 nucleotides upstream for Gibson assembly.

Dual Luciferase Assay
The dual luciferase assay was performed according to the manufacturer’s instructions (Promega). In brief, C2C12 myoblasts and HEK-293T cells were collected 48 h after transfection with plasmids indicated in Figure 5 and SI Appendix, Figure S4 and lysed with 200 μL passive lysis buffer. Next, 20 μL lysate was transferred onto a luminometer plate, and luciferase activity was measured using a luminometer. After adding 50 μL LAR II, Firefly luciferase activity will be measured. After dispensing 50 μL Stop & Glo Reagent, Renilla luciferase activity can be measured and used as the internal control. By analyzing the relative luciferase activity, the induction of Firefly luciferase could be calculated.

Protein Purification
Plasmids containing GST and GST-MyoD were transformed into ROSSETA BL21-competent cells and grown to start culturing. Overnight culture was added to 1 L LB medium (1:20) and cultured for 2-3 h at 37°C until the optical density reached 0.5. IPTG was added to the culture at a final concentration of 1 mM and cells were cultured for additional 2 h at 37°C. Cells were centrifuged at 14,000 × g for 10 min at 4°C and were used immediately for protein purification or stored at -80°C.

Cell pellets were resuspended in 1× PBS with PMSF. After sonication, Triton-X was added to the solution to a final concentration of 1%. After incubation at 4°C for 30 min, the solution was centrifuged at 14,000 × g for 10 min at 4°C. The supernatant was then incubated with glutathione agarose beads (Sigma-Aldrich) for 30 min at 4°C. After three washes with 1% PBST, bound proteins were eluted with elution buffer (5-10 mM reduced glutathione [Sigma-Aldrich] in 50 mM Tris-HCl [pH 9.5]). Eluted proteins were then processed to dialysis for further experiments.
**GST Pulldown Assay**

The GST pulldown assay was performed as previously described by Wang et al. (6). Both proteins and glutathione agarose beads (Sigma-Aldrich) were blocked with excess yeast tRNA the night before the binding assay. LncMyoD and H2BG RNAs were in vitro transcribed and denatured at 95°C for 2 min, snap-cooled on ice, and refolded in folding buffer (100 mM KCl, 10 mM MgCl₂, and Tris-HCl [pH 7.0]) for at least 20 min before the binding assay. Proteins were incubated with LncMyoD or H2BG for 1 h at room temperature in PB100 buffer (20 mM HEPES [pH 7.6], 100 mM KCl, 0.05% NP40, 1 mM DTT, and 0.5 mM PMSF). The mixture was then incubated with glutathione agarose beads for 45 min at 4°C in PB100 buffer. After three washes in PB200 buffer (20 mM HEPES [pH 7.6], 200 mM KCl, 0.05% NP40, 1 mM DTT, and 0.5 mM PMSF), proteins were digested with proteinase K in proteinase K buffer (100 mM NaCl, 10 mM Tri-HCl [pH 7.0], 1 mM EDTA, and 0.5% SDS) for 1 h at room temperature. Bound RNA was extracted using TRI Reagent LS (Molecular Research Center) and analyzed by qRT-PCR.

**RNA immunoprecipitation (RIP)**

Cells were lysed in lysis buffer (50 mM Tris, pH 7.5, 100 mM KCl, 12 mM MgCl₂, 1% Nonidet P-40, 1mM DTT, 200 U/ml RNAse inhibitor, 1× protease inhibitor mixture). After centrifugation, supernatant was incubated with antibody targeting MyoD (BD Biosciences) or Gal4-DBD (Santa Cruz) at 4°C overnight. The mixture was further incubated with protein G magnetic beads (Invitrogen) at 4°C for 4 hours with rotation. Samples were then placed in a magnet and the pellets were washed three times in high salt buffer (50 mM Tris [pH 7.5], 300 mM KCl, 12 mM MgCl₂, 1% Nonidet P-40, 1 mM DTT). After washing, RNAs were extracted using Direct-zol™ RNA MiniPrep kit (Zymo Research) for further analysis.

**RNA-seq**

RNA was extracted from cultured SCs with siRNA treatment using a NucleoSpin RNA XS kit (Clontech) according to the manufacturer's instructions. RNA concentrations were determined using a Qubit RNA HS Assay Kit (Life Technologies) on a Qubit 2.0 Fluorometer (Life Technologies). RNA (1 ng) was used for first-strand cDNA preparation according to the SMART-seq v4 Ultra Low Input RNA Kit for Sequencing (Clontech) protocol and processed for library generation using the Ovation RNA-seq System V2 (NuGEN). Sequencing was carried out in 2 × 150 cycle paired-end mode on an Illumina NextSeq 500 (Illumina).

**Generation of LncMyoD-KO Cell Line**

The pSpCas9(BB)-2A-GFP (PX458) plasmid (Addgene) containing the human codon optimized SpCas9 gene with 2A-EGFP and the cloning backbone for sgRNA was used to generate the LncMyoD-KO cell line. sgRNAs were cloned into PX458 according to the Feng Zhang Lab CRISPR plasmid protocol (7). The following sgRNA sequences flanking exon 2 of LncMyoD were used:

- guide 2: AGTCCTAAGAGCACATGGGT
- guide 3: AAGTGGTCAAGCTTGGCGA.

All guides were designed following the guidelines from Ran et al. (7); and primer sequences are listed in Dataset S3.

Mouse C3H/10T1/2 fibroblasts and SV129 primary myoblasts were transfected with Cas9-2A-GFP-gRNA-encoding plasmids using Lipofectamine III (Invitrogen). Transfected cells were dissociated by trypsin 48 h after transfection and isolated using GFP reporter by FACS. The isolated GFP⁺ cells were subsequently serially diluted in DMEM to a final concentration of 0.5 cells per well in a 96-well plate. Colonies were allowed to expand for approximately 3 weeks, followed by genotyping to establish a new 10T1/2 clonal cell line. The WT clones are defined as clones that are successfully transfected with gRNA but do not have LncMyoD deletion. The KO clones are defined as clones that are successfully transfected with gRNA and have LncMyoD exon 2 deletion on both alleles.

**10T1/2 Fibroblast Transdifferentiation**

Adenovirus containing MyoD plasmid was allowed to infect 10T1/2 fibroblasts, which were then cultured for 2 days. Plasmids containing LncMyoD and empty vector were transfected 1 day after virus infection. Cells were then induced for differentiation until myotubes formed in control groups. Cells were then fixed and stained for immunofluorescence visualization.

**LncRNA Identification**

Nine public SC RNA-seq samples from Wüst et al. (8) and four SC RNA-seq samples generated by our laboratory were analyzed to identify differentially expressed LncRNAs. Reads of SC RNA-seq samples from Wüst et al. (8)
were downloaded and extracted for each sample from the NCBI Sequence Read Archive. For our laboratory's samples, low-quality 3′ base calls were trimmed using Trimmomatic v0.36 (9) with the following trimming parameters: SLIDINGWINDOW: 10:20, MINLEN: 30. Reads from all samples were then mapped to the mm10 reference genome using HISAT2 v2.1.0 (10) with the default parameters.

After read mapping, StringTie v1.3.3 (11) was used with default settings for genome-wide guided transcript assembly and expression level estimation. Using the Gencode version M19 gene annotations (12) and the mapped reads as input to StringTie, a consensus list of novel and annotated transcripts for both coding and non-coding genes in satellite cells were generated. The individual GTF files produced by StringTie for each input sample were merged using StringTie in merge mode. Then, the merged GTF file was reran once more for each sample using the new GTF file as input to StringTie with the -eB setting.

Next, possible novel IncRNAs were identified from this filtered set of expressed transcripts. To do this, we separated the transcripts in the merged GTF file into two sets, one annotated and one unannotated. The unannotated set of transcripts was used to search for possible novel IncRNAs. To avoid any possible low-quality unannotated transcripts that were assembled by StringTie, only those with more than one exon were considered; any novel transcripts <200 nucleotides in length were also excluded. For each remaining transcript, the corresponding mm10 reference genome sequence was extracted, and a coding potential score was calculated using the coding potential calculator CPC2 (13). Novel transcripts categorized by CPC2 as noncoding were kept and considered possible noncoding, novel IncRNAs.

Next, all annotated IncRNAs in the Gencode M19 gene annotations (12) were pulled and then merged with the identified novel IncRNAs into a master list of IncRNAs to be used for differential expression analysis. To identify IncRNAs that may be involved in myogenesis, the log2 FPKM fold change between quiescent satellite cell (FISC) and differentiated myotube sample groups was calculated for each IncRNA and those IncRNAs with a fold change of 1.5 or higher were kept. For these differentially expressed IncRNAs, the annotated protein-coding gene that was nearest to it was found. Of all IncRNA candidates discovered, only three were nearby a myogenic regulatory factor, NC 117, NC 3092, and NC 3093, and these were investigated further.

The exon sequences of the three identified novel IncRNAs were extracted from the mm10 reference genome and an NCBI BLAST nucleotide search was performed for highly similar sequences in the mouse genomic plus transcript sequence set. The novel IncRNA NC 117 reported a match to the predicted noncoding RNA XR_878548, NC 3093 reported a match to the lncRNA Munc, and NC3092 reported a match to the IncRNA LncMyoD.

ChIP-seq Analysis
ChIP-seq analysis was performed according to the ENCODE histone ChIP-seq data analysis guidelines (14). For all 1 × 75 bp ChIP-seq samples, reads had adapters, and low-quality bases were removed using Trimmomatic v0.36 (9) with the following trimming parameters: ILLUMINACLIP: Adapters.fa: 2:30:7:5, LEADING: 3, TRAILING: 3, SLIDINGWINDOW: 10:10, MINLEN: 20. The Adapters.fa file contained the following three entries for Nextera adapter trimming: CTGTCTCTTATACACATCTCCGAGCCCACGAGAC, TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG. After trimming, the reads were mapped to the mm10 genome using Bowtie2 v2.3.2 (15) with the default settings. The MarkDuplicates function in Picard v2.9.2 was used to mark reads identified as PCR and optical duplicates. All unmapped and duplicate reads as well as reads with a mapping quality <10 were removed from each sample using Samtools v1.3 (16) with the following parameters: -F 1540 -q 10.

For each treatment group, the mapped reads from the treatment sample replicates were pooled, and the mapped reads from the treatment control replicates were pooled. The run_spp.R script in Phantompeakqualtools v1.2 (17) was used to estimate the predominant fragment length of the pooled treatment reads for use as an input parameter to MACS2 during peak calling. Peaks were then called for each treatment group with MACS2 v2.1.1 (18) using that group's pooled treatment and control reads with the following parameters: -g mm -p 1e-2 –nomodel –shift 0 –extsize (estimated fragment length) –keep-dup all -B –SPMR. Fold-change signal tracks were then created for each treatment group from the output from the MACS2 peak calling by using the MACS2 bgdcmp command and the bedGraphToBigWig v332 tool (19).

ATAC-seq Alignment
Raw BCL sequencing data from the 2 × 75 bp paired-end ATAC-seq run were demultiplexed into individual sample replicates using Illumina’s bcl2fastq v2.19.0 software. Adapter trimming was enabled in bcl2fastq using the suggested Tn5 adapter, CTGTCTCTTATACACATCT. Adapter read through was still present in a fraction of the paired-end reads, so the FastQ files generated by bcl2fastq had an additional round of adapter trimming using Trimmomatic v0.36 (9) with parameters 'ILLUMINACLIP:NexteraPE-Custom.fa:1:13:8:1:True
SLIDINGWINDOW:10:10 MINLEN:15’. For the ILLUMINACLIP setting, a custom Nextera paired end adapter FASTA file was used containing the following four adapter entries in it:

PrefixA/1 – CTGTCCTCTTATACACATCTCCGAGCCCACGAGAC,
PrefixA/2 – CTGTCCTCTTATACACATCTGACGCTGCCGAGCT,
Trans1 – TCGTCGGCAGCGTCAGATGTATAGAGACAG,
Trans2 – GTCTCGTGGGCTCGGAGATGTATAGAGACAG.

After adapter trimming, the reads were mapped to the mm10 genome using Bowtie2 v2.3.2 (15) alignment software with the following settings: --minins 30 --maxins 2000 --dovetail -k 4 --very-sensitive-local. The MarkDuplicates function in Picard v2.9.2 was used to mark reads identified as PCR and optical duplicates. All unmapped and duplicate reads as well as reads with a mapping quality < 30 were removed from each sample using Samtools v1.3 (16) with the following parameters: -F 1804 -f 2 -q 30.

ATAC-seq Replicate Correlation and Peak Calling

Broad peak calling was performed on each individual replicate from the previous section using MACS2 v2.1.1 (18) with the following parameters: –nomodel –shift -100 –extsize 200 –broad. The output of all broad peak files from all replicates were pooled together into one BED file. This merged peak BED file was used to evaluate concordance between the individual replicates. The number of Tn5 insertion events within each region in the BED file was calculated for each replicate with the 5’ and 3’ mapping locations of paired end read pairs, each being considered an insertion event. Spearman’s rank correlation was performed and showed that the replicates were well correlated (SI Appendix, Figure S5A). After confirming the correlation of the replicates, the aligned reads from a sample’s two replicates were combined into one file. Broad peak calling was done again but this time on the combined reads using MACS2 with the same parameters listed above.

ChromVAR Analysis

Transcription factor binding motif accessibilities were assessed using ChromVAR v0.99.1 (20). Peak files called from the previous section were provided to get the counts for each peak in different samples. Mouse transcription factor motifs were obtained from the JASPAR 2016 database (21) and chromVARmotifs collection (20). The abundance of these motifs was calculated within each peak. Background peaks were computed using similar GC content and average accessibility with the input peaks. Finally, the deviation matrix was generated to give the accessibility for each set of peaks for different samples. The deviation z-scores from the matrix were used to generate the heatmap showing accessibility changes of different transcription.

NucleoATAC Analysis

To generate an enriched region file needed for NucleoATAC v0.3.4 (22) for analysis, the peak calls from both LncMyoD-KD and control samples were merged together into one file. The peak calls in this file were extended by 350 bp up- and downstream, and all peaks within 200 bp of each other were merged together. All Gencode-annotated TSSs were subsequently added to this BED file with an additional 1 kb up- and downstream of each TSS. In addition, the LncMyoD, Myogenin, Myod1, Pax7, Myf5, Acta1, Actc2, Srpk3, Unc45b, Csrp3, Tmem8c, and Myh1 loci were added to the BED file for analysis. These areas were chosen for deeper analysis based on differentially expressed genes determined in the gene expression analysis. In a similar fashion to the LncMyoD-KD and control samples, the peak calls from both ASC and FISC samples were merged into a single file as described above.

ChIRP-seq analysis

Reads from odd and even probe pools were mapped to the mm10 genome separately using Bowtie2 v2.3.2 (15) with the default settings. The MarkDuplicates function in Picard v2.9.2 was used to mark reads identified as PCR and optical duplicates. All unmapped and duplicate reads were removed. After filtering, only the reads that are shown in both the odd and even probe pools were kept for peak calling. Peaks were then called with MACS2 v2.1.1 (18) with the following parameters: -g mm -p 1e-2 –nomodel –shift 0 -B –SPMR. Fold-change signal tracks were then created for each treatment group from the output from the MACS2 peak calling by using the MACS2 bgdcmp command and the bedGraphToBigWig v332 tool (19).

Gene Expression Analysis in LncMyoD-KD RNA-seq

Samples from the 2 × 150 bp RNA-seq library were demultiplexed using bcl2fastq v2.19.0. Illumina sequencing often contains low-quality base calls towards the 3’ of reads, particularly in the reverse read of paired-end libraries. To increase the number of mapping reads, these low-quality base calls were trimmed using Trimmomatic v0.36 (9) with the following trimming parameters: SLIDINGWINDOW: 10:20, MINLEN: 30. The trimmed reads were then
mapped to mm10 using HISAT2 v2.1.0 (10) with the default parameters. After read mapping, StringTie v1.3.3 (11) was used with the default parameters to identify novel transcripts. The GTF files produced by StringTie for every sample were then merged using StringTie in merge mode, again with the default parameters. This merged GTF file was then used with StringTie with the -eB setting to quantify expression levels for all novel and annotated transcripts in each sample. Expression level estimates from StringTie were then used to determine differentially expressed genes between the LncMyoD-KD and control samples.
Figure S1. Chromatin Accessibility Changes during Satellite Cell Activation (Related to Figure 1)

A

| FlSC  | H3K4me3 | H3K27me3 | ATAC-seq |
|-------|---------|----------|----------|
|       |         |          |          |

| ASC   | H3K4me3 | H3K27me3 | ATAC-seq |
|-------|---------|----------|----------|
|       |         |          |          |

B

FISC

- Promoter (6.41%)
- 3' UTR (2.19%)
- 1st Exon (0.41%)
- 1st Intron (13.53%)
- Downstream (<10kb) (14.22%)
- Distal Intergenic (43.94%)

ASC

- Promoter (9.15%)
- 5' UTR (0.51%)
- 3' UTR (2.51%)
- 1st Exon (0.46%)
- Other Exon (5.41%)
- 1st Intron (13.25%)
- Other Intron (27.43%)
- Downstream (<10kb) (1.66%)
- Distal Intergenic (30.34%)

C

Open promoter regions

ASC: 1475
FISC: 1160

D

FISC enriched ATAC-seq peaks

E

ASC enriched ATAC-seq peaks
Figure S1. Chromatin Accessibility Changes during Satellite Cell Activation (Related to Figure 1)
(A) Genome tracks of H3K4me3 ChIPmentation, H3K27me3 ChIPmentation and ATAC-seq signals in FISCs and ASCs across Calcr and Heyl loci.
(B) Genomic annotations of open regions in FISCs and ASCs.
(C) Venn diagram showing the amount of overlapping open promoter regions between FISCs and ASCs.
(D, E) Reactome analysis of peaks identified in FISC (D) and ASC (E) ATAC-seq data.
(F, G) Reactome analysis of differentially expressed genes in FISCs (F) and ASCs (G).
Figure S2. Identification of LncMyoD (Related to Figure 2)

(A-B) Genome tracks of ATAC-seq signal in SCs across the XR_878548 (A) and Munc (B) loci.

(C) Genome tracks of ATAC-seq signal (upper panel) and ChIP-seq signals of H3K27me3, H3K4me3 and MyoD (lower panel) in satellite cells (SCs) across the LncMyoD locus.

(D) RACE results of full-length LncMyoD.

(E) Genome track of RNA-seq signal across LncMyoD locus in C2C12 myoblasts and myotubes.

(F) Genome track of RNA-seq signal across LncMyoD locus freshly isolated satellite cells (FISCs), proliferating satellite cells (pSC), and differentiating satellite cells (dSC).
Figure S3. LncMyoD Knockdown in SCs and Primary Myoblasts (Related to Figure 3)

(A) RT-qPCR result of LncMyoD knockdown (KD) in SCs, validating the effect of siRNA KD on LncMyoD expression (*p < 0.05). 

(B) RT-qPCR result of LncMyoD knockdown (KD) in primary myoblasts, validating the effect of siRNA KD on LncMyoD expression (**p < 0.01). 

(C-D) RT-qPCR results of Myod1 (C) and Myog (D) expression in LncMyoD knockdown SCs (**p < 0.01).
Figure S4. LncMyoD Interacts Exclusively with MyoD over Other Myogenic Regulatory Factors (Related to Figure 5)

A

|       | Interaction Propensity | Discriminative Power |
|-------|------------------------|----------------------|
| Pax7  | 35                     | 82%                  |
| MyoD  | 72                     | 98%                  |
| MyoG  | 4                      | 22%                  |
| Myf5  | 3                      | 20%                  |
| Mrf4  | 3                      | 20%                  |

B

C

D

E

F

G

13
Figure S4. *LncMyoD* Interacts Exclusively with MyoD over Other Myogenic Regulatory Factors (Related to Figure 5)

(A) Interaction propensity and discriminatory power calculated by the CatRAPID algorithm, predicting the interaction possibilities of *LncMyoD* with Pax7 and myogenic regulatory factors (MRFs). Higher interaction propensity indicates higher probability; higher discriminatory power indicates more confidence in how close the algorithm is to the interaction propensity.

(B) Heatmap showing the interaction hotspots of the *LncMyoD* RNA sequence and MyoD protein sequence.

(C) Transfection of the indicated constructs. No increase in luciferase activity was observed after transfection with Gal4-DBD fused MyoG, Myf5, or Mrf4.

(D) Increased MyoD expression led to enhanced luciferase activity.

(E) Increased *LncMyoD* expression led to enhanced luciferase activity.

(F) Purified GST and GST-MyoD were visualized by SDS-PAGE and Coomassie Blue staining.

(G) RT-qPCR results of retrieved *LncMyoD* level through Gal4-DBD RIP of Gal4-DBD fused MyoD truncations.
Figure S5. Genome Accessibility Changes after Loss of LncMyoD (Related to Figure 6)

A. Spearman rank correlation

B. H3K4me3 ChIP peaks in myoblast

C. H3K4me3 ChIP peaks in myotube

D. MyoD ChIP peaks in myoblast

E. MyoD ChIP peaks in myotube
Figure S5. Genome Accessibility Changes after Loss of LncMyoD (Related to Figure 6)
(A) Spearman correlation of ATAC-seq datasets in FISCs, ASCs, control (Ctrl), and LncMyoD-KD groups in biological replicates.
(B) Insertion size distribution distinguishing nucleosome-free regions (NFRs) and nucleosome-occupied regions in the Ctrl and LncMyoD-KD groups using the NucleoATAC program. Highlighted are single-nucleosome spanning reads (around 150bp long).
(C) Comparison of ATAC-seq signals between the Ctrl and LncMyoD-KD groups across regions enriched for H3K4me3 marks using C2C12 ChIP-seq data in myoblasts (left) and myotubes (right).
(D) Comparison of ATAC-seq signals between the Ctrl and LncMyoD-KD groups across regions bound by MyoD, MyoG, Usf1 and Cebp in C2C12 ChIP-seq data from myoblasts and myotubes.
(E) Changes in ATAC-seq signals between the Ctrl and LncMyoD KD groups across regions bound MyoD, MyoG, Usf1 and Cebp in C2C12 myoblasts and myotubes, showing the changes in the binding potential of these two TFs during myogenic differentiation after LncMyoD KD.
(F) Genome tracks of the Mybph and Mef2a loci showing ATAC-seq, H3K27ac ChIP-seq, and MyoD ChIP-seq across their promoter regions in the Ctrl and LncMyoD-KD groups.
(G) RT-qPCR results of ChIRP retrieved RNA.
(H) Genome tracks of the Cdh15 and LncMyoD loci showing LncMyoD ChIRP-seq and MyoD ChIP-seq across their promoter regions.
Figure S6. Generation of LncMyoD-KO Clones of 10T1/2 Fibroblasts (Related to Figure 7)

A

B

C

Myod1  Myog  Myl1

Relative expression level

D

E

Myod1  Myog  Myl1

Relative expression level

F

G

Day 0  Day 1  Day 2  Day 3
Figure S6. Generation of LncMyoD-KO Clones of 10T1/2 Fibroblasts (Related to Figure 7)
(A) Genotyping results of LncMyoD-knockout (KO) 10T1/2 fibroblast clones using the CRISPR/Cas9 system.
(B) Sequencing and mapping results of selected LncMyoD-KO HOMO clones.
(C-E) RNA-seq results showing the expression level changes of MyoD1, Myog and Myl1 in WT (C), LncMyoD-KO1 (D) and LncMyoD-KO2 (E) clones with different treatment.
(F) HeLa cells were induced to differentiate for 8 days after transfection of the indicated LncMyoD plasmid and MyoD adenovirus. Cells were then fixed to visualize myotube formation.
(G) Morphology of WT and LncMyoD-KO SV129 primary myoblasts induced to differentiate for 3 days in differentiation medium.

Supplementary Datasets
Dataset S1. Novel and annotated lncRNAs that are differentially expressed during satellite cell quiescence, activation, and differentiation
Dataset S2. Sequencing result of LncMyoD-KO clones
Dataset S3. Oligonucleotides used in this study
SI References

1. Liu L, Cheung TH, Charville GW & Rando TA (2015) Isolation of skeletal muscle stem cells by fluorescence-activated cell sorting. Nature Protocols
2. Schmidl C, Rendeiro AF, Sheffield NC & Bock C (2015) ChiPmentation: Fast, robust, low-input ChiP-seq for histones and transcription factors. Nature Methods
3. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods
4. Chu C, Quinn J & Chang HY (2012) Chromatin isolation by RNA purification (ChIRP). Journal of Visualized Experiments : JoVE
5. Yu X, et al (2017) Long non-coding RNA linc-RAM enhances myogenic differentiation by interacting with MyoD. Nature Communications
6. Wang KC, et al (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature
7. Ran FA, et al (2013) Genome engineering using the CRISPR-Cas9 system. Nature Protocols
8. Wüst S, et al (2018) Metabolic maturation during muscle stem cell differentiation is achieved by miR-1/133a-mediated inhibition of the Dlk1-Dio3 mega gene cluster. Cell Metabolism
9. Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: A flexible trimmer for illumina sequence data. Bioinformatics
10. Kim D, Langmead B & Salzberg SL (2015) HISAT: A fast spliced aligner with low memory requirements. Nature Methods
11. Pertea M, et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology
12. Frankish A, et al (2019) GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Research
13. Kang YJ, et al (2017) CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Research
14. Landt SG, et al (2012) ChiP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research
15. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nature Methods
16. Li H, et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics
17. Kharchenko PV, Tolstorukov MY & Park PJ (2008) Design and analysis of ChiP-seq experiments for DNA-binding proteins. Nature Biotechnology
18. Zhang Y, et al (2008) Model-based analysis of ChiP-seq (MACS). Genome Biology
19. Kent WJ, Zweig AS, Barber G, Hinrichs AS & Karolchik D (2010) BigWig and BigBed: Enabling browsing of large distributed datasets. Bioinformatics
20. Schep AN, Wu B, Buenrostro JD & Greenleaf WJ (2017) ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nature Methods
21. Mathelier A, et al (2016) JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Research
22. Schep AN, et al (2015) Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Research