Molecular Cell, Volume *47*

Supplemental Information

Snail Regulates MyoD Binding-Site Occupancy to Direct Enhancer Switching and Differentiation-Specific Transcription in Myogenesis

Vahab D. Soleimani, Hang Yin, Arezu Jahani-Asl, Hong Ming, Christel E.M. Kockx, Wilfred F.J. van Ijcken, Frank Grosveld, and Michael A. Rudnicki

INVENTORY OF SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURES

SUPPLEMENTAL TABLES

Table S4. GO Term Analysis of MyoD Targets in Primary Myotubes, Related to Figure 1

Table S5. GO Term Analysis of Myf5 Targets in Primary Myoblasts, Related to Figure 1

Table S6. GO Term Analysis of MyoD and Myf5 Common Targets in Primary Myoblasts, Related to Figure 2

Table S7. GO Term Analysis of MyoD Targets in C3H10T1/2 Fibroblasts, Related to Figure 1

Figure S1. Validation of Chromatin Tandem Affinity Purification and Functional Validation of TAP-tagged MyoD, Related to Figure 1.

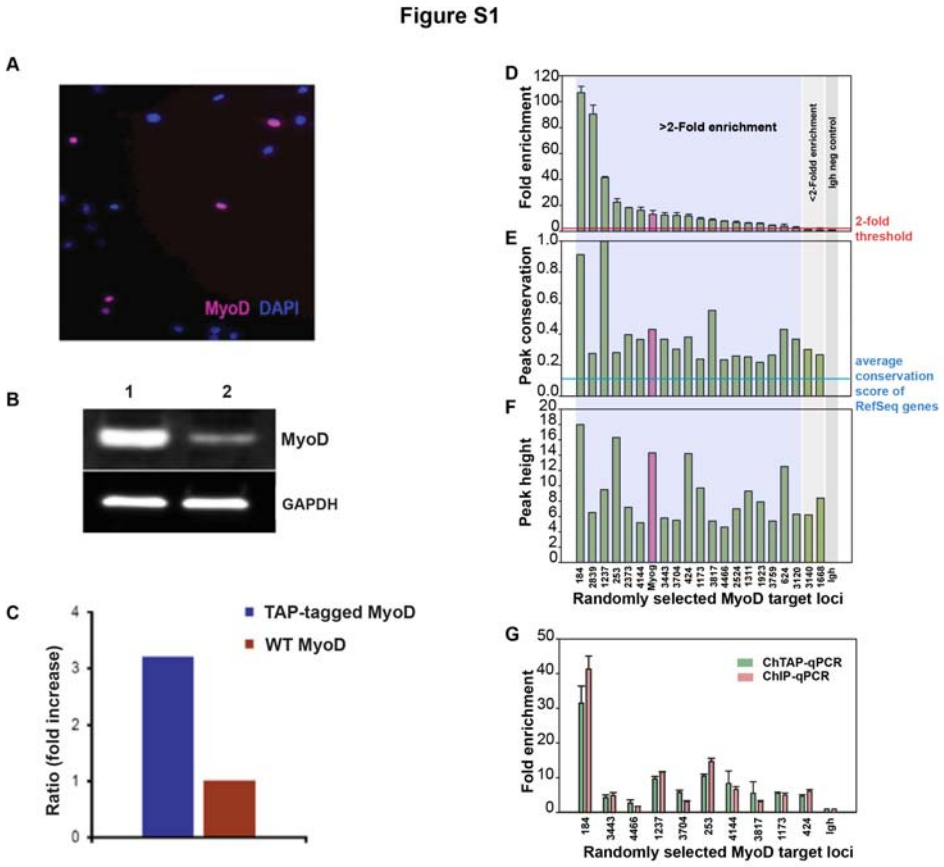


Figure S2. Genome Wide binding Site and Conservation Analysis of MyoD and Myf5 Binding Sites, Related to Figure 2

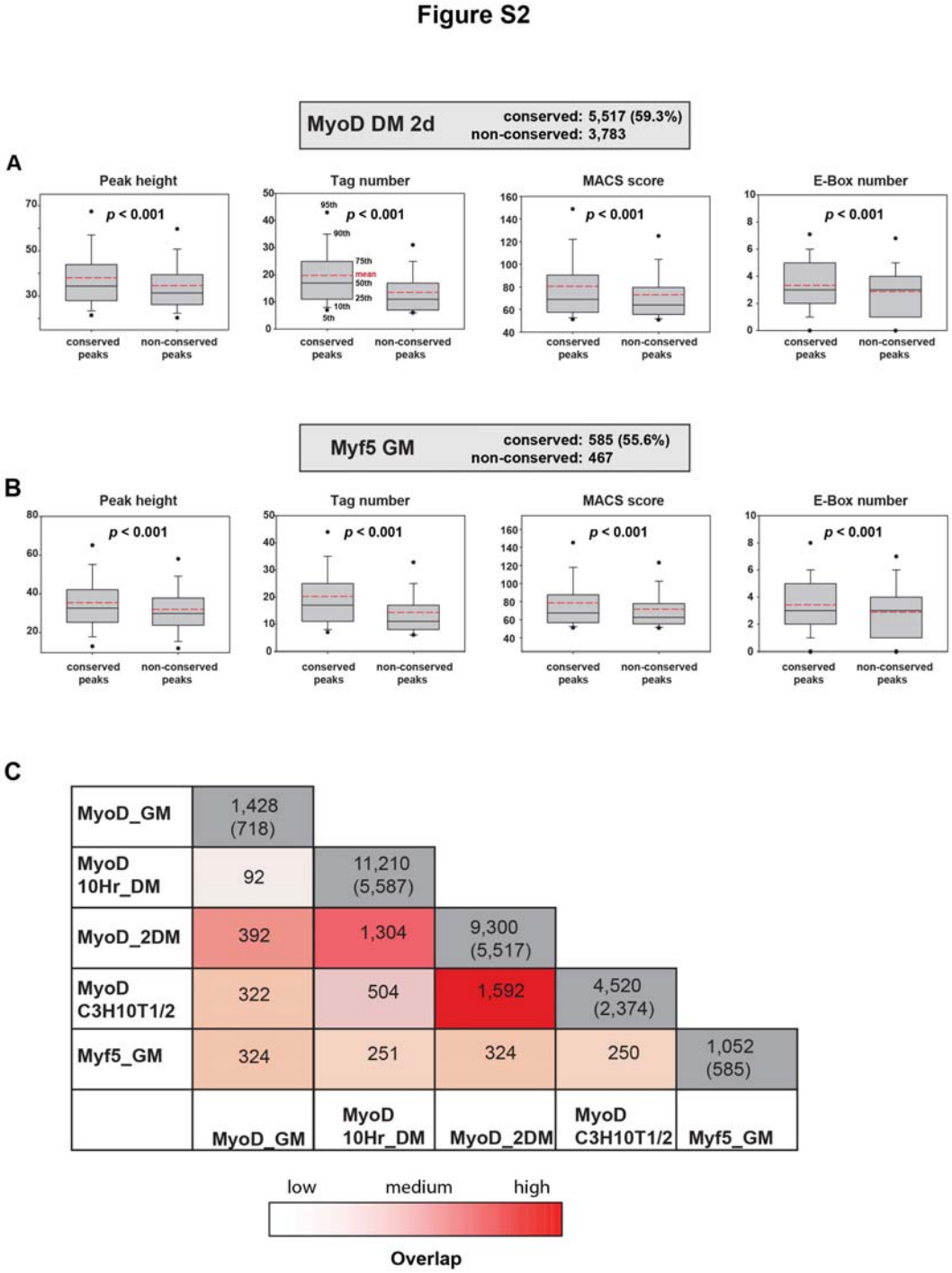


Figure S3. Quantification of MyoD Molecules at The Single Cell Level in Primary Myoblasts and C2C12 Myogenic Cell Line, Related to Figure 2

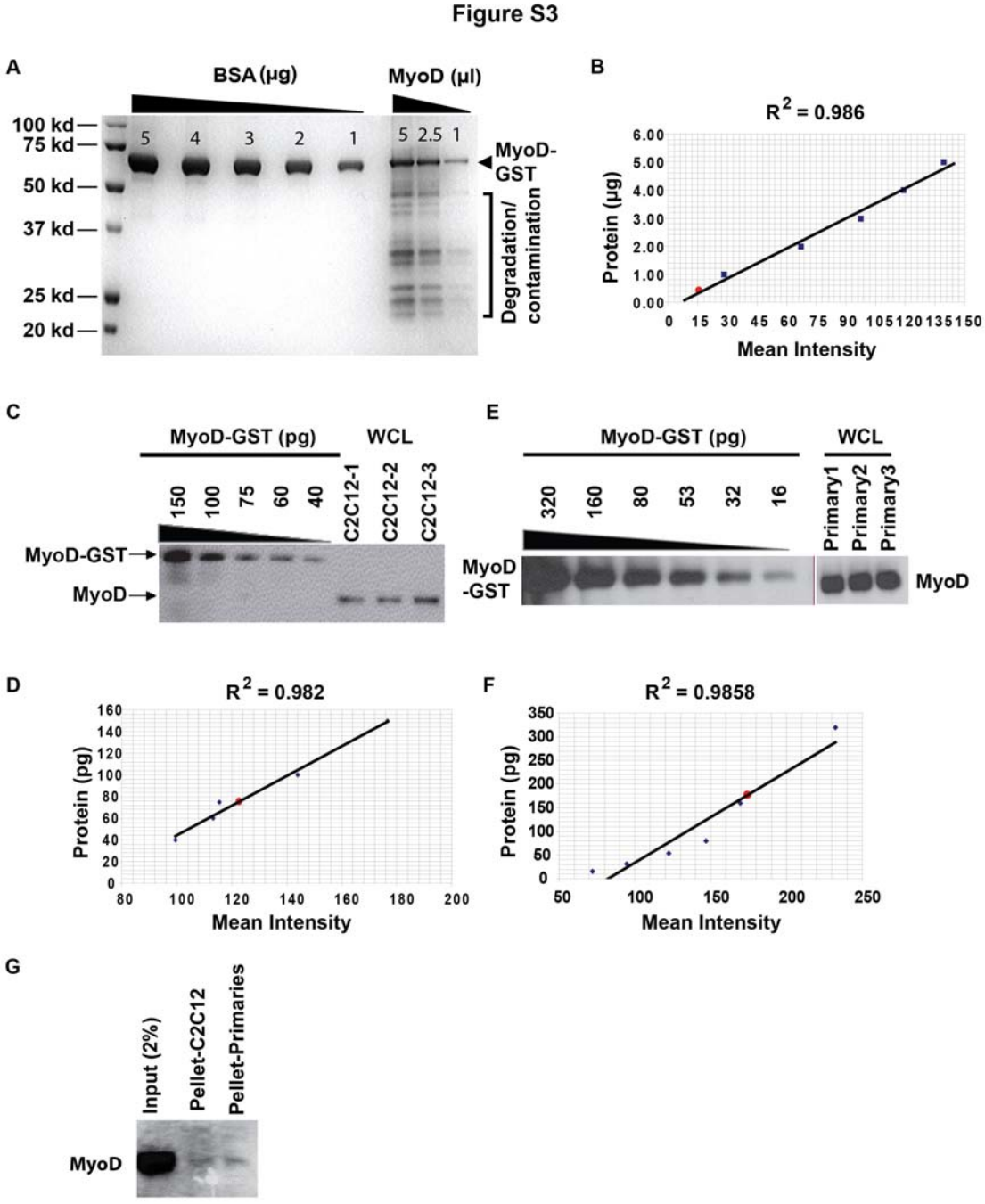


Figure S4. Snai1/2 Blocks MyoD-Mediated Differentiation in Muscle Cells, Related to Figure 4

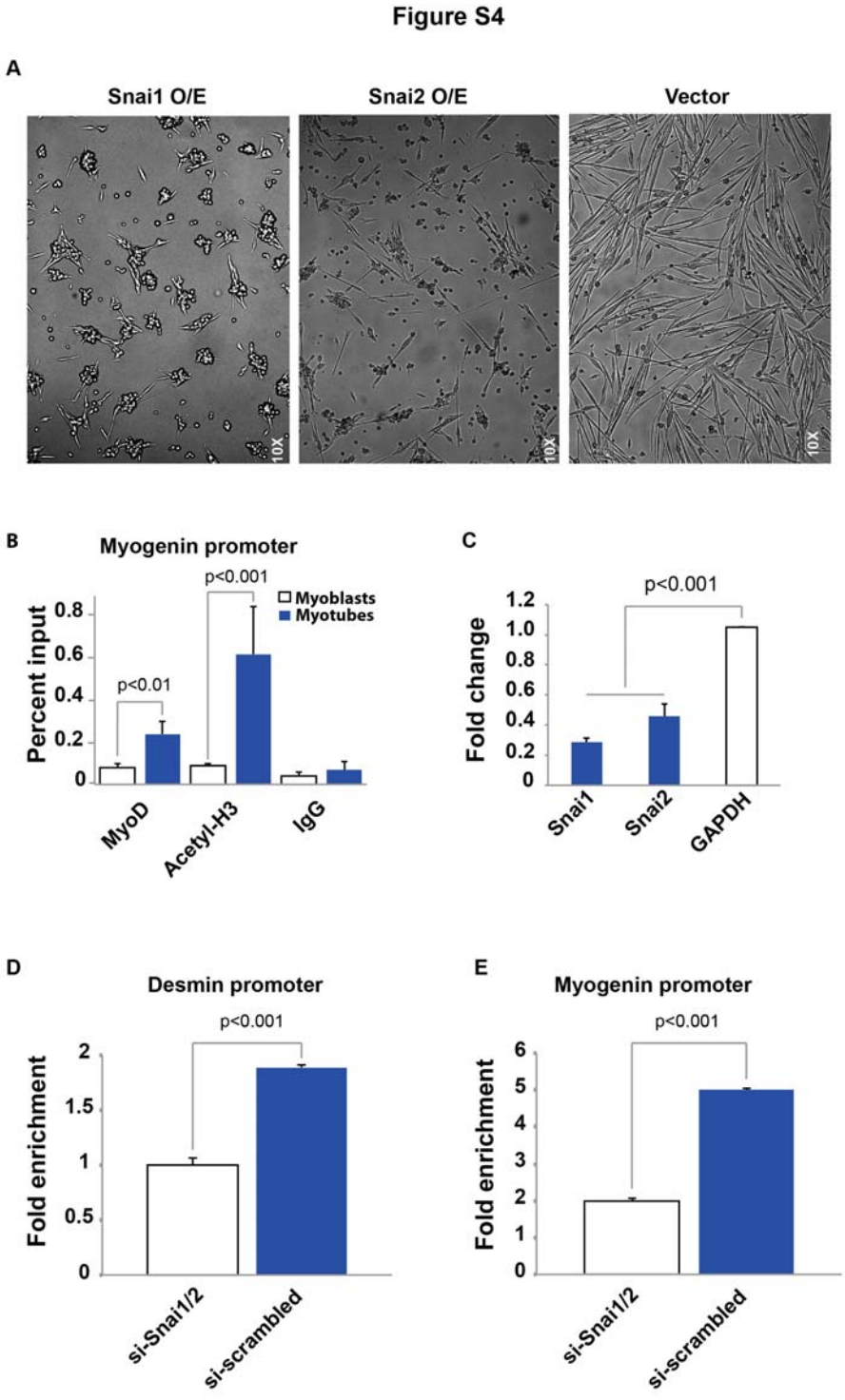
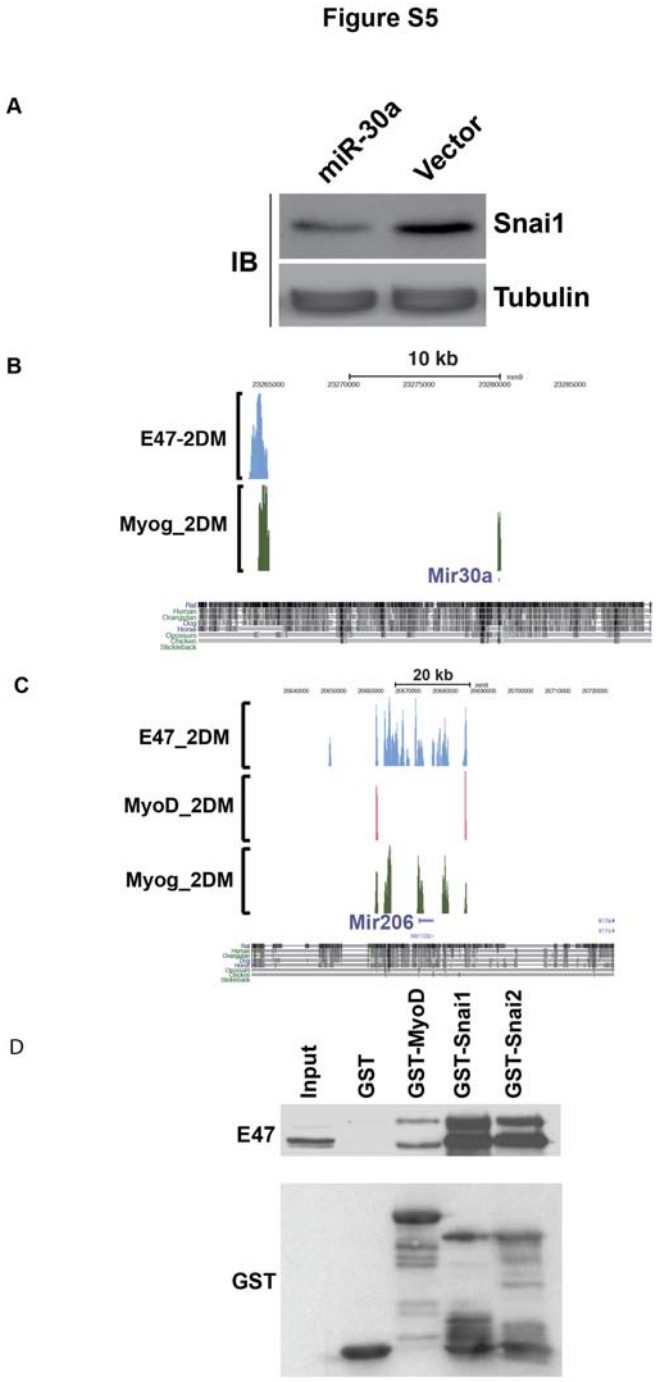


Figure S5. The Relationship Between Snails, MRFs/E-Proteins and miR30a/miR-206 in Regulation of Muscle Cell Differentiation, Related to Figure 7



SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Validation of Chromatin Tandem Affinity Purification and Functional Validation of TAP-tagged MyoD, Related to Figure 1.

(A) MyoD-TAP fusion protein retains function and is localized to cell nucleus. MyoD null myoblasts were infected with retrovirus containing MyoD-CTAP tagged construct. Functional localization of MyoD-TAP protein in MyoD-/- myoblasts by immunostaining with an anti MyoD antibody showed the nuclear localization of TAP-tagged MyoD. **(B)** comparison between the expression level of endogenous MyoD versus retroviral infection of MyoD-TAP by Western blot analysis using an anti MyoD antibody. Lanes: 1) myoblasts infected with MyoD-TAP, lane 2) uninfected (wild type) myoblasts **(C)** Quantification of the level of endogenous MyoD protein versus TAP-tagged MyoD (exogenous) from C.

(D) ChIP-qPCR validation of a random set of MyoD targets from MyoD ChTAP-Seq dataset of differentiated myotubes using locus-specific primers. Myog promoter, a known MyoD target was used as a positive control (red). Error bars represent standard deviations (n=3). **(E)** The conservation scores of the same set of MyoD target sites calculated from phastcon scores of placental mammals. The horizontal blue line indicates average conservation scores of RefSeq genes. **(F)** ChTAP-Seq peak heights of the same set of MyoD targets. **(G)** Comparison between ChTAP-qPCR and traditional ChIP-qPCR shows that ChTAP can robustly identify bona fide binding sites, as seen by strong correlation between ChTAP- and ChIP-qPCR. Error bars represent standard deviations (n=3).

Figure S2. Genome Wide binding Site and Conservation Analysis of MyoD and Myf5 Binding Sites, Related to Figure 2

(A-B) Box plots showing that that evolutionary conserved MyoD and Myf5 binding sites have higher quality scores than non-conserved counterparts. Peak quality scores were measured by peak height, tag number per peak, MACS score and the number of E-boxes per peak. Dashed red lines within the box shows mean, solid black line inside the box shows the median. **(C)** Overlap of MyoD and Myf5 targets in different datasets and during growth and differentiation of muscle cells as well as fibroblasts (C3H10T1/2) forced to undergo myogenic differentiation by the ectopic expression of MyoD. Diagonal entries indicate total number of peaks per dataset. Numbers in the bracket show the number of conserved binding sites. Shading intensity reflects the extent of overlap based on the number of shared target sites. GM: growth; DM: differentiation.

Figure S3. Quantification of MyoD Protein at The Single Cell Level in Both C2C12 Myogenic Cell Line and The Mouse Primary Skeletal Muscle Cells, Related to Figure 2

(A) MyoD-GST recombinant protein purified from *E. coli* was quantified based on known concentrations of BSA. The gel was stained with Comassie Brilliant Blue (CBB). Measurements and quantifications were done using ImageJ program. **(B)** BSA standard curve was used to determine the concentration of full length MyoD-GST recombinant protein. **(C-D)** Standard curve of MyoD-GST protein was used to determine the amount of MyoD protein from the C2C12 cell lysates (see the Extended Experimental Procedures for details). **(E-F)** Standard curve of MyoD-GST protein was used to determine the

amount of MyoD protein from primary myoblasts cell lysates as above. **(G)** Western blot showing the extent of MyoD protein loss in cell pellet (cell derbies). Mouse monoclonal Anti MyoD antibody (5.8A) was used in Western blots. Extrapolation of data from Figure S3 showed that mouse primary myoblasts contain approximately 98,000 ± 5,100 MyoD molecules per single cell, while C2C12 myogenic cell line contains approximately 38,000 ± 4,600 MyoD molecules per cell.

Figure S4. Snai1/2 Blocks MyoD-Mediated Differentiation in Myoblasts, Related to Figure 4

(A) Ectopic expression of Snai1/2 blocks muscle cell entry into differentiation. Primary myoblasts were infected with Snai1 or Snai2 retrovirus containing full-length mouse Snai1 or Snai2 Open Reading Frame (ORF) fused with CTAP tag (6His-TEV-3FLAG). A control (Vector) retrovirus of the same backbone harboring only the CTAP tag was used to assess the effect of the over expression (O/E) of Snai1 and Snai2 on differentiation of myoblasts. Stable cells were obtained by puromycin selection. Bright field images taken after 48 hours in differentiation media (DMEM supplemented with 5% horse serum) show that continuous expression of Snai1 or Snai2 significantly reduces differentiation. **(B)** Enrichment of acetylated histone H3 and MyoD on Myogenin promoter during differentiation. **(C)** Snai1/2 transcript in myoblasts can be efficiently depleted by siRNA interference (**D**-**E**) Promoter recruitment of HDAC1 is dependent on Snai1/2. Removal of Snai1/2 results in the reduction of HDAC1 on Desmin and Myogenin promoters in myoblasts. Error bars represent SD (n=3).

Figure S5. The Relationship Between Snails, MRFs/E-Proteins and miR30a/miR-206 in Regulation of Muscle Cell Differentiation, Related to Figure 7

(A) Transient transfection of Cos7 cells with miR-30 expression vector results in significant reduction in Snai1 protein level. **(B)** miR-30a is target of Myogenin/E47 in primary myotubes. **(C)** Multiple binding sites of MRFs/E47 on miR-206 during differentiation of muscle cells. **(D)** Snai1 and Snai2 directly bind E47. GST pull down assays showing the interaction of Snai1 and Snai2 with E47, a heterodimerization partner of bHLH proteins. Recombinant proteins were produced in BL20 *E. coli*. Pull down assays were done with nuclear extracts obtained from primary mouse skeletal myoblasts under growth condition.

SUPPEMENTAL TABLES

Table S1. Sample Description of ChIP-Seq Datasets, Related to Figure 1.

ChIP-seq Sample	Number of tags (non-redundant)	Illumina Sequencing Platform	Description
Myf5_GM	8,792,116	GAIIX	Myf5 in myoblasts
MyoD_GM	13,353,158	GAIIX	MyoD in myoblasts
EV_GM	10,456,283	GAIIX	Empty vector control myoblasts
MyoD_10h	4,263,527	GAIIX	MyoD in myocytes (10 hours DM)
EV_10H	12,022,005	GAIIX	Empty vector control 10 hours DM
MyoD_10T	12,862,941	GAIIX	MyoD in C3H10T1/2 fibroblasts
EV_10T	6,983,662	GAIIX	Empty vector control C3H10T1/2
MyoD_2DM	7,569,520	GAIIX	MyoD in myotubes (48 hours DM)
EV_2DM	42,907,935	HI-SEQ 2000	Empty vector control myotubes
Snai1_GM	5,622,639	GAIIX	Snai1 in myoblasts
EV_GMs	9,559,105	GAIIX	Empty vector control for Snai1
HDAC1_GM	22,371,510	HI-SEQ 2000	HDAC1 in myoblasts
HDAC2_GM	53,583,467	HI-SEQ 2000	HDAC2 in myoblasts
E47_2DM	59,535,426	HI-SEQ 2000	E47 in myotubes

Binding data is available at GEO GSE24852 series, part of the super series GSE24904

Table S2. Validation of a Randomly Selected Set of MyoD Targets in Myotubes by ChTAP- and ChIP qPCR, Related to Figure 2

Peak ID	Chrm	Start	End	Conservation score	Peak hight-intensity	Location
624	chr7	53596779	53597336	4.303	12.5	Intergenic
4144	chr1	187565338	187565682	3.653	5.2	Intergenic
4466	chr19	26424226	26424400	2.341	4.6	Intergenic
3817	chr11	58882503	58882687	5.517	5.4	Obscn:Intron|Exon
3759	chr5	134829126	134829281	2.648	5.4	Intergenic
2839	chr17	21638416	21638563	2.745	6.5	Zfp53:Intron
1173	chr9	40913455	40913752	2.383	9.7	Ubash3b:Intron
3140	chr4	58565958	58566141	3.007	6.2	Lpar1:Intron|Exon
424	chr8	125170303	125170691	3.805	14.2	Cbfa2t3:Intron
1311	chr1	17135626	17135919	2.526	9.3	Gdap1:Exon|Intron
1237	chr11	88047475	88047633	9.975	9.5	Intergenic
2373	chr6	71673706	71673892	3.96	7.2	Reep1:Intron
2524	chr3	53435144	53435416	2.584	7	Frem2:Intron
253	chr6	100139376	100139729	2.808	16.3	Intergenic
1668	chr8	127857666	127857859	2.66	8.4	Intergenic
3120	chr9	108589067	108589279	3.677	6.3	Intergenic
1923	chr11	69804331	69804592	2.172	7.9	Intergenic
184	chr3	20693250	20693478	9.123	18	Intergenic
3704	chr11	77599936	77600158	3.027	5.5	Myo18a:Intron
3443	chr17	14669773	14670034	3.667	5.8	Intergenic

Table S3. GO Term Analysis of MyoD Targets in Growing Primary Myoblasts, Related to Figure 1

GO term	P value
GO:0016311~dephosphorylation	0.001071293
GO:0006333~chromatin assembly or disassembly	0.001149766
GO:0018101~peptidyl-citrulline biosynthetic process from peptidyl-arginine	0.001548151
GO:0019240~citrulline biosynthetic process	0.001548151
GO:0006836~neurotransmitter transport	0.001948599
GO:0019320~hexose catabolic process	0.00228203
GO:0006007~glucose catabolic process	0.00228203
GO:0030036~actin cytoskeleton organization	0.002724788
GO:0046365~monosaccharide catabolic process	0.002967452
GO:0000052~citrulline metabolic process	0.002969085
GO:0044275~cellular carbohydrate catabolic process	0.003805797
GO:0030029~actin filament-based process	0.005437025
GO:0043269~regulation of ion transport	0.007463228
GO:0019794~nonprotein amino acid metabolic process	0.007647174
GO:0006334~nucleosome assembly	0.008035456
GO:0051924~regulation of calcium ion transport	0.008616121
GO:0031497~chromatin assembly	0.009636792

Table S4. GO Term Analysis of MyoD Targets in Primary Myotubes (48 Hours in Differentiation Media), Related to Figure 1. (Table is Provided in Excel Format).

Table S5. GO Term Analysis of Myf5 Targets in Growing Primary Myoblasts (GM), Related to Figure 1

GO term	P value
GO:0044238~primary metabolic process	7.10E-04
GO:0008152~metabolic process	0.001180576
GO:0046686~response to cadmium ion	0.001413229
GO:0044237~cellular metabolic process	0.001528753
GO:0007517~muscle organ development	0.001604305
GO:0043283~biopolymer metabolic process	0.001606876
GO:0043170~macromolecule metabolic process	0.00181799
GO:0007519~skeletal muscle tissue development	0.002717083
GO:0060538~skeletal muscle organ development	0.003262796
GO:0014706~striated muscle tissue development	0.003769728
GO:0014823~response to activity	0.004260087
GO:0009968~negative regulation of signal transduction	0.00472626
GO:0010648~negative regulation of cell communication	0.005113096
GO:0019538~protein metabolic process	0.005344334
GO:0060537~muscle tissue development	0.006579724
GO:0034960~cellular biopolymer metabolic process	0.006929614
GO:0030036~actin cytoskeleton organization	0.007826095
GO:0044260~cellular macromolecule metabolic process	0.008529953

Table S6. GO Term Analysis of Common Target Genes Between Myf5 and MyoD in Myoblasts (GM), Related to Figure 1

GO Term	p value
GO:0044249~cellular biosynthetic process	0.01255917
GO:0009058~biosynthetic process	0.013906703
GO:0030036~actin cytoskeleton organization	0.018619587
GO:0007517~muscle organ development	0.019233821
GO:0046580~negative regulation of Ras protein signal transduction	0.020023947
GO:0051058~negative regulation of small GTPase mediated signal transduction	0.020023947
GO:0015804~neutral amino acid transport	0.020023947
GO:0030029~actin filament-based process	0.025214585
GO:0010467~gene expression	0.0276267
GO:0016053~organic acid biosynthetic process	0.027844941
GO:0046394~carboxylic acid biosynthetic process	0.027844941
GO:0009611~response to wounding	0.029317322
GO:0042060~wound healing	0.033716347
GO:0006399~tRNA metabolic process	0.035958917
GO:0055002~striated muscle cell development	0.039298494
GO:0007205~activation of protein kinase C	0.039298494
GO:0030239~myofibril assembly	0.039298494
GO:0014706~striated muscle tissue development	0.039833144
GO:0034660~ncRNA metabolic process	0.042649743
GO:0008610~lipid biosynthetic process	0.04380238
GO:0006633~fatty acid biosynthetic process	0.044211024
GO:0006868~glutamine transport	0.046736488
GO:0006670~sphingosine metabolic process	0.046736488
GO:0060537~muscle tissue development	0.049783588
GO:0006082~organic acid metabolic process	0.049988128
GO:0043436~oxoacid metabolic process	0.050021707
GO:0019752~carboxylic acid metabolic process	0.050021707
GO:0002253~activation of immune response	0.056320027
GO:0042180~cellular ketone metabolic process	0.057818237
GO:0034645~cellular macromolecule biosynthetic process	0.060264942
GO:0009059~macromolecule biosynthetic process	0.063060255
GO:0031032~actomyosin structure organization	0.063217585
GO:0002252~immune effector process	0.063403694
GO:0055001~muscle cell development	0.072045367
GO:0007507~heart development	0.072153735
GO:0055007~cardiac muscle cell differentiation	0.076600111
GO:0009987~cellular process	0.077559319
GO:0016044~membrane organization	0.078014266
GO:0007059~chromosome segregation	0.083771523
GO:0002377~immunoglobulin production	0.085971215
GO:0034961~cellular biopolymer biosynthetic process	0.088341204

GO Term	P value
GO:0043284~biopolymer biosynthetic process	0.0902436
GO:0048305~immunoglobulin secretion	0.09129777
GO:0002440~production of molecular mediator of immune response	0.095665554
GO:0008104~protein localization	0.098678803
GO:0051605~protein maturation by peptide bond cleavage	0.099278465

Table S7. GO Term Analysis of MyoD Targets in C3H10T1/2 Fibroblasts Undergoing Myogenic Differentiation By Forced Expression of MyoD, Related to Figure 1. (Table Provided in Excel Format)

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Plasmids and Cell Culture

TAP tagged fusion constructs were produced by GateWay cloning (Invitrogen San Diego, CA). The complete Open Reading Frame (ORF) of MyoD, Myf5, Snai1, Snai2, HDAC1, HDAC2 and E47 were fused with a C-terminus TAP tag composed of 6His-TEV-3xFLAG in pBRIT-CTAP retroviral expression vector backbone (McKinnell et al., 2008). Retroviral particles were produced in Phoenix-eco packaging cell line, a kind gift from Nolan's lab (http://www.stanford.edu/group/nolan/retroviral_systems/phx.html). miR-206 was cloned by inserting a pri-miR-206 sequence between EcoRI and NotI sites on a pCDH-CMV-MCS-EF1-Puro plasmid (System Bioscience) using 5' GGTCCCAGAGATTCTTCACAC sense and 5' AGCATAGTTGACCTGAAACCC antisense primers. pCMV-miR-30 plasmid (Cat #20670) was obtained from the Addgene (http://www.addgene.org).

Muscle satellite cells were isolated from the hind limb muscles of 4- to 6-week old wild type mice using Fluorescence Activated Cell Sorting (FACS) as described previously (Kuang et al., 2007). Cells were propagated in growth medium (Ham's F10 supplemented with 20% FBS and 2.5 ng/ml bFGF) for three passages. Pure populations of myoblasts were infected with a retroviral vector (MOI ~10) harboring the gene of interest fused with a CTAP tag. Puromycin selection (2.5 µg/ml) was applied to eliminate non-infected cells for up to one week. Low-level expression and functionality of retrovirally-introduced factor were assessed by standard biochemical and immunofluorescence techniques, as illustrated for MyoD (Fig S1). Thereafter, stably infected myoblasts were propagated in growth medium containing 1.5 µg/ml of

puromycin until harvest. For differentiation of myoblasts Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 5% horse serum was used.

Antibodies

The following antibodies were used throughout this work: anti MyoD antibody Catalogue number sc-304x (Santa Cruz Biotechnology) for ChIP, anti MyoD antibody Catalogue number 554130 (BD Biosciences) for immunostaining and western blot, anti GAPDH antibody Catalogue number 4300 (Ambion) and anti alpha tubulin antibody catalogue number T9026 (Sigma). Anti Snai1 antibody, Catalogue number 3895 (Cell Signaling Technologies) and anti Snai2 antibody catalogue number 9585 (Cell Signaling Technologies), anti-3xFLAG M2 affinity gel (Sigma Aldrich) was used for the first immuno pull down of TAP-tagged protein and His-Select nickel beads (Sigma-Aldrich) were used for the second immuno pull down of the TAP tagged protein. Anti E47, sc-763; anti HDAC1, sc-7872 (Santa Cruz Biotechnologies); anti HDAC2 sc-7899 (Santa Cruz Technologies). Anti myogenin, F5D and anti MyHC, MF20 are from (Developmental Studies Hybrodoma Bank). Anti Acetyl-histone H3, Catalogue number 06-599 (Millipore) was used for ChIP-qPCR.

Primers and Probes

ChIP Primers:

Peak ID	Sequence
624F	TGCCTCAGATCAGTGTGGAC
624R	GTGCAGAGAGGCCAATTCAG
4144F	CGAGGCAGTTGCTCTTTACC
4144R	CTGCAAGCGATTTCAGTGAG
4466F	CAATGCTCTAGCATGGCTCA
4466R	TTGTCCCTGGATTTGAGGAG
3817F	ATCTGGCACTCACCATGGAC

3817R	GGAAGACACTGAGGCCGTCT
3759F	ACCTCTGCTTCACAGGTGCT
3759R	AGCTCCAGCAAAATCAATGC
2839F	AACAGCTTCTTGTGGCATCC
2839R	CTGTCCAGTCCCCTCACAAG
1173F	TTAAAAGGACCTGCGGTGAC
1173R	TGGGGAGATGGTGGCTATAA
3140F	CGATCCAGGTAAGCAACAGG
3140R	ACCAGCCGGTGGAACTCA
424F	GACGCTTGGATGGGAGTTAG
424R	ATTGTTCCGTGCCTTGCTTA
1311F	GACGCACTCCTTCAGCTCTC
1311R	AGTGCACCCCAGTTCTGAAT
1237F	AATTGGCAAGTTGTCACAAAAG
1237R	AAGCGATGTCTTTGTGTCAGATAA
2373F	CACAGCAGGCTCAGCAACT
2373R	CCTTATGCTCCTGCCAGTGT
2524F	GAAGTGGCTCTCAGGGACAG
2524R	CTGCTTGAGCAGAGGAGGTC
253F	AGCTGTGGCTTCCATGTTCT
253R	CATGCCACAGGAAACAGATG
1668F	CTGTCTGCATGGCTGGAAT
1668R	CGCTGACTAAGCCACTCCTC
3120F	GCTTGTGAAAGGAACCAAGC
3120R	GGCTTCCATTTGACTCTGACA
1923F	GACGTGGACAGGCTAAGAGG
1923R	AGAAAAGGGCGGAGAATCAC
184F	TTCAGAACTCAGAGCCAGACAG
184R	TACCTCCATCAAGCAGTTGG
3704F	AGGCCGTGCTTTCAGTAGAA
3704R	GTGATGGGTCTGTGCATGAG
3443F	TACTCCCCAGCAGCTGTTTC
3443R	CTGGGATGAGAACCCATTTG

Expression Primers

Primer Name	Sequence
Desmin_F	ATGAAGAGGAGATCCGTGAG
Desmin_R	GGTATTCCATCATCTCCTGCT
Tnnc1_F	GAGCAGAAGAATGAGTTCAAGG
Tnnc1_R	TCGTCTACTTCGTCAATCATCTC
Smyd1_F	TGGCTACATGAAACTCTACC
Smyd1_R	CCTCTAAGTCTTTGGTGATAGG
Mypn_F	CCAGAATGCAGGAAAGAGAC
Mypn_R	CCACTCACCTTACAGTCCAG
Snai1_F	AGCCCAACTATAGCGAGCTG

Snai1_R	GGGGTACCAGGAGAGAGTCC
Snai2_F	AGCGAACTGGACACACACAC
Snai2_R	AGGAGAGTGGAGTGGAGCTG
miR-206_F	TGGAATGTAAGGAAGTGTGTGG
miR-206_R	GCATACGAGCTCTTCCGATCT
U6_F	TCGCTTCGGCAGCAC
U6_R	TTTGCGTGTCATCCTTGC
Myog_F	CAGTGAATGCAACTCCCACAG
Myog_R	ATGGACGTAAGGGAGTGCAGA
CKM_F	AAGTACTACCCTCTGAAGAGC
CKM_R	AAGATCTCCTCAATCTTCTGC
Mef2a_F	CGGATATCGTTGAGTGTATT
Mef2a_R	CATTGAGAAGTTCTGAGGTG
GAPDH_F	TGTCCGTCGTGGATCTGA
GAPDH_R	GGTCCTCAGTGTAGCCCAAG

siRNAs (Sigma)
Snai1: *SASI_Mm01_00097354
SASI_Mm01_00097355

Snai2:
*SASI_Mm01_00143833
SASI_Mm01_00143834
SASI_Mm01_00143835

*siRNAs that showed the most efficient knockdown. They were used in all subsequent experiments.

Luciferase Constructs

ID	Sequence
GC:	GCGTATAGATCTGGTCCTGACAGCTGGGCTGGGCTAGGGTCCTGACAGCT GGGCTGGGCTAGGGTCCTGACAGCTGGGCTGGGCTAAGCTTAGTGAGTC
AA:	GCGTATAGATCTGGTCCTGACAAATGGGCTGGGCTAGGGTCCTGACAAAT GGGCTGGGCTAGGGTCCTGACAAATGGGCTGGGCTAAGCTTAGTGAGTC
AT	GCGTATAGATCTGGTCCTGACAATTGGGCTGGGCTAGGGTCCTGACAATT GGGCTGGGCTAGGGTCCTGACAATTGGGCTGGGCTAAGCTTAGTGAGTC
AC	GCGTATAGATCTGGTCCTGACAACTGGGCTGGGCTAGGGTCCTGACAACT GGGCTGGGCTAGGGTCCTGACAACTGGGCTGGGCTAAGCTTAGTGAGTC
GG	GCGTATAGATCTGGTCCTGACAGGTGGGCTGGGCTAGGGTCCTGACAGGT GGGCTGGGCTAGGGTCCTGACAGGTGGGCTGGGCTAAGCTTAGTGAGTC

EMSA Probes

Probe ID	Sequence
GC	AGGGTCCTGACAGCTGGGCTGGGGC
TG	AGGGTCCTGACATGTGGGCTGGGGC
AT	AGGGTCCTGACAATTGGGCTGGGGC
AA	AGGGTCCTGACAAATGGGCTGGGGC

Chromatin Tandem Affinity Purification (ChTAP)

At least 120 million cells were used for each ChIP experimental. Myoblasts or differentiated myotubes stably expressing a C-terminus TAP tagged fusion proteins were cross-linked with 1% formaldehyde in 1x PBS for 10 minutes. ChTAP was performed as described previously (Soleimani et al., 2012).

ChIP-Seq

The ChIP DNA library was prepared from DNA obtained by ChTAP (Soleimani et al., 2012) according to the Illumina protocol (www.illumina.com) and as described previously (Soler et al., 2010). Sequencing was done on a Genome Analyzer IIX or Hi-Seq 2000 Sequencing System (Illumina). Short read sequences were mapped to mm9 mouse genome assembly (NCBIM37) using Illumina pipeline ELAND. The binding data is available at GEO, GSE24852, also see Table S1.

Solexa Read Alignment and Peak Calling

Solexa reads (36 bp) were aligned to mouse genome (NCBI37) by ELAND (Illumina) with up to two mismatches. Sequences tags with same 5' ends were stacked and counted as one tag to avoid PCR-induced sequencing depth bias. Aligned sequencing tags from ChTAP were fed to MACS ver1.3 (Model-based Analysis for ChIP-Seq (Zhang et al., 2008) together with tags from control experiments using empty vectors. Peak calling was done based on a p-value cut-off of 10^{-5}. Background subtraction was done by calculating

the raw sores of 50 base pair windows across genome for tags from the control and the ChIP datasets. We then subtracted the raw scores of control from the raw scores of ChIP to generate the subtracted scores. Tag-enriched genomic regions in the control experiments were determined using Global Identifier of Target Region (GLITR) (Tuteja et al., 2009) and hereafter called pseudo peaks. Pseudo peaks were used as negative control peaks in downstream comparative analyses such as Figure 1.

Analysis of Conservation

PhastCon conservation scores of 20-way placental mammals were downloaded from the UCSC Genomic Bioinformatics (http://hgdownload.cse.ucsc.edu/goldenPath/mm9/phastCons30way/placental/).

Averaged conservation scores were calculated for each ChIP-Seq peak and each pseudo peak, excluding nucleotide bases within coding sequences of RefSeq genes (UCSC refGene table) and repetitive regions (UCSC rmsk table). Meanwhile, the maximum length of local conserved regions (defined as more than 10 continuous bases with conservation scores > = 0.4) was calculated. ChIP-Seq peaks that had average conservation scores of less than 0.1 and contained no local conserved region were grouped into non-conserved binding sites.

Conserved E-box and E-box Classification

E-Boxes with average conservation scores more than or equal to 0.4 were defined as conserved E-Boxes. Nucleotides within coding sequences of RefSeq genes and repetitive regions were not included in the conservation score calculation as described above. Both conserved and non-conserved E-Boxes were classified by on the center dinucleotide sequences. Given the palindromic nature of E-boxes, each dinucleotide combination and

its reverse complement were grouped together and classified into 10 non-redundant E-box groups.

Gene Expression Data During The Time Course of Primary Muscle Cells Differentiation

We performed a three-point time series of muscle cells differentiation using microarrays analysis to study the gene expression profile during muscle cell differentiation. Briefly, muscle stem cells-derived myoblasts were maintained in growth media (Ham's F10 supplemented with 20% FBS, 1% penicillin/streptomycin, 2.5 ng/ml bFGF) until harvest. For differentiation (2 days and 5 days in DM), Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% horse serum was used. Total RNA was isolated in triplicates from FACS sorted mouse primary myoblasts and differentiated myotubes using RNeasy mini kit (Qiagen). The purity of RNA was analyzed by Bioanalyzer (Agilent Technologies, Santa Clara, CA). Samples that showed RNI > 9.0 were used for subsequent labeling and hybridization. The output expression data was analyzed using the Significance Analysis of Microarray (SAM). The microarray data is available at GEO, GSE24811.

Global Analysis of Snai1 and Snai2 Targets Genes

We used siRNAs for *in vivo* depletion of *Snai1* and *Snai2* transcript in mouse primary myoblasts followed by the whole transcriptome analysis using RNA-seq. The RNA-seq data is available at GEO, GSE38236. Briefly, mRNA was purified from 2 µg of total RNA and fragmented according to the manufacturer's (Illumina) recommendations. The first and the second strand cDNA were synthesized followed by end repairing and 3' end

adenylation and adaptor ligation. Ligation products were purified followed by PCR amplification and the resulting library was analyzed by Bioanalyzer (Agilent Technologies). Cluster generation was carried out following Illumina TruSeq SR Cluster Kit v5 and Illumina Cluster Station was used for cluster generation. Single-read sequencing was done using TruSeq SBS Kit v5 – GA (36-cycle) on Illumina GAIIx Genome Analyzer. Reads from RNA-seq samples were mapped to the mm9 (NCBI build 37) mouse genome assembly and to splice sites predicted from UCSC splicing models (refFlat.txt.gz) as described (Fujita et al., 2011) using the eland_rna analysis option of the Illumina GERALD pipeline v1.7 (Illumina) using default parameters. The Illumina CASAVA pipeline (v1.7) was used to aggregate the mapped reads and quantify the transcripts present in the original sample, using the readBases method. The EdgeR Bioconductor package (Robinson et al., 2010) was used to compare expression between pairs of samples using the negative binomial exact test. Fold change and p-values were corrected by multiple testing using the Bonferroni method.

Identification of MyoD Target Genes

Gene expression information of *MyoD*$^{-/-}$ primary myoblasts undergoing differentiation at 2 days time point were previously reported (Porter et al., 2007) and collected from GEO (GSE10424 and GSE10430). MyoD regulated genes were determined by Significance Analysis of Microarrays via R packages: GCRMA and siggenes.

Pair-Wise Overlap Between Two Datasets by Common Peaks

ChTAP-Seq peaks from two datasets were matched by pair-wise comparison regarding their genomic locations. A peak was defined common if it overlapped with another peak

from the second dataset in such a way that the overlapped region was longer than half the length of the narrower peak.

GO Enrichment Analysis

Genes within -100 kb ~ +100 kb range of peak centers were extracted GO enrichment analysis of genes from each dataset and common genes between two datasets were conducted via the functional annotation module of DAVID 6.7 (http://david.abcc.ncifcrf.gov) (Dennis et al., 2003; Huang da et al., 2009).

GST Pull-Down Assay

N-terminus GST fusion proteins of mouse full length MyoD, Snai1 and Snai2 were produced by sub cloning their respective ORF into pGEX4T-1 plasmid using *EcoRI* and *NotI* restriction enzymes sites. The recombinant protein was produced in BL20 *E. coli* and purified by Glutathione sepharose beads (GH Health Care).

Gel Shift Assay

Electrophoretic mobility shift assay (EMSA) were carried out as described previously (Soleimani et al., 2012) using a ^{32}P radio labeled synthetic DNA probe containing the canonical E-box sequence with various combination of the degenerate center dinucleotide as outlined above.

Luciferase Reporter Assay

Cos7 and HEK293T cells were cotransfected with pGL4.23-E-box containing a multimerized (3x) E-box motif. A CMV driven MyoD, E47 or Snai1 in pBRIT backbone was used for competition assay. Luciferase assays were done using Dual-Luciferase Assay System (Promega) following manufacture's recommendations. For the competition

assay between MyoD/E47, Snai1 and Snai2, a 1:1 molar ratio of plasmids was used. Transfection were carried out using Linear Polyethylenimine (Reed et al., 2006).

Quantification of MyoD Protein at The Single Cell Level

Cells were counted in triplicates by two independent methods; 1) Fluorescent Activated Cell Sorting (FACS) (Fig S3) and by an automated Trypan Blue method using Countess automated cell counter (Invitrogen). MyoD-GST fusion protein was purified from *E. coli* as described previously. The concentration of the full length MyoD-GST fusion protein was calculated from a standard curve via BSA (Figure S3). C2C12 myogenic cells and primary mouse skeletal myoblasts were lysed in nuclear lysis buffer containing 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.5% Triton X-100 plus protease inhibitors. The lysate was spun at 14K for 15 minutes and supernatant was transferred to a new eppendorf tube. The pellet containing cell derbies was boiled in SDS loading buffer for 5 minutes and run on SDS-PAGE and blotted to estimate MyoD protein loss in pellet. A standard curve for MyoD-GST fusion protein was used to quantify the total amount of MyoD protein per cell lysate. The number of MyoD molecules per cell was determined from the MyoD-GST standard curve based on a molecular size of 45 kd, the band size of endogenouse MyoD on an SDS-PAGE gel blotted with anti MyoD 5.8A antibody.

siRNA Knockdowns

Transient, *in vivo* depletion of *Snai1* and *Snai2* was carried out by lipofectamine-mediated transfection of independent siRNAs and a non-specific scrambled siRNA as a control in primary myoblasts. The knockdown efficiency was measured by Real-Time RT-qPCR using gene-specific primers.

SUPPLEMENTAL REFERENCES

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., *et al.* (2011). The UCSC Genome Browser database: update 2011. Nucleic acids research *39*, D876-882.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Kuang, S., Kuroda, K., Le Grand, F., and Rudnicki, M.A. (2007). Asymmetric self-renewal and commitment of satellite stem cells in muscle. Cell *129*, 999-1010.

McKinnell, I.W., Ishibashi, J., Le Grand, F., Punch, V.G., Addicks, G.C., Greenblatt, J.F., Dilworth, F.J., and Rudnicki, M.A. (2008). Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex. Nature cell biology *10*, 77-84.

Porter, C.J., Palidwor, G.A., Sandie, R., Krzyzanowski, P.M., Muro, E.M., Perez-Iratxeta, C., and Andrade-Navarro, M.A. (2007). StemBase: a resource for the analysis of stem cell gene expression data. Methods Mol Biol *407*, 137-148.

Reed, S.E., Staley, E.M., Mayginnes, J.P., Pintel, D.J., and Tullis, G.E. (2006). Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors. Journal of virological methods *138*, 85-98.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England) *26*, 139-140.

Soleimani, V.D., Punch, V.G., Kawabe, Y., Jones, A.E., Palidwor, G.A., Porter, C.J., Cross, J.W., Carvajal, J.J., Kockx, C., van Ijcken, W., *et al.* (2012). Transcriptional dominance of Pax7 in adult myogenesis is due to high affinity recognition of homeodomain motifs. Developmental Cell *In press*.

Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.J., Stevens, M., Kockx, C., van Ijcken, W., *et al.* (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes & development *24*, 277-289.

Tuteja, G., White, P., Schug, J., and Kaestner, K.H. (2009). Extracting transcription factor targets from ChIP-Seq data. Nucleic acids research *37*, e113.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.