MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis

Marc Santolini, Iori Sakakibara, Morgane Gauthier, Francesc Ribas-Aulinas, Hirotaka Takahashi, Tatsuya Sawasaki, Vincent Mouly, Jean-Paul Concordet, Pierre-Antoine Defossez, Vincent Hakim, et al.

Published online 14 June 2016

Nucleic Acids Research, 2016, Vol. 44, No. 18 ***8621–8640***
doi: 10.1093/nar/gkw512

MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide *cis*-regulatory module analysis

Marc Santolini[1,2,3,†], Iori Sakakibara[1,2,4,†], Morgane Gauthier[1,2], Francesc Ribas-Aulinas[1,2], Hirotaka Takahashi[5], Tatsuya Sawasaki[5], Vincent Mouly[6], Jean-Paul Concordet[1,2], Pierre-Antoine Defossez[7], Vincent Hakim[3,*] and Pascal Maire[1,2,*]

[1]Institut Cochin, Université Paris-Descartes, Centre National de la Recherche Scientifique (CNRS), UMR 8104, Paris, France, [2]Institut National de la Santé et de la Recherche Médicale (INSERM) U1016, Paris, France, [3]Ecole Normale Supérieure, CNRS, Laboratoire de Physique Statistique, PSL Research University, Université Pierre-et-Marie Curie, Paris, France, [4]Division of Integrative Pathophysiology, Proteo-Science Center, Graduate School of Medicine, Ehime University, Ehime, Japan, [5]Proteo-Science Center, Ehime University, Ehime 791-8577, Japan, [6]Sorbonne Universités, UPMC Univ Paris 06, INSERM UMRS974, CNRS FRE3617, Center for Research in Myology, 75013 Paris, France and [7]University Paris Diderot, Sorbonne Paris Cité, UMR 7216 CNRS, 75013 Paris, France

Received December 15, 2015; Revised May 24, 2016; Accepted May 26, 2016

ABSTRACT

Myogenic regulatory factors of the MyoD family have the ability to reprogram differentiated cells toward a myogenic fate. In this study, we demonstrate that Six1 or Six4 are required for the reprogramming by MyoD of mouse embryonic fibroblasts (MEFs). Using microarray experiments, we found 761 genes under the control of both Six and MyoD. Using MyoD ChIPseq data and a genome-wide search for Six1/4 MEF3 binding sites, we found significant co-localization of binding sites for MyoD and Six proteins on over a thousand mouse genomic DNA regions. The combination of both datasets yielded 82 genes which are synergistically activated by Six and MyoD, with 96 associated MyoD+MEF3 putative *cis*-regulatory modules (CRMs). Fourteen out of 19 of the CRMs that we tested demonstrated in Luciferase assays a synergistic action also observed for their cognate gene. We searched putative binding sites on these CRMs using available databases and *de novo* search of conserved motifs and demonstrated that the Six/MyoD synergistic activation takes place in a feedforward way. It involves the recruitment of these two families of transcription factors to their targets, together with partner transcription factors, encoded by genes that are themselves activated by Six and MyoD, including Mef2, Pbx-Meis and EBF.

INTRODUCTION

In vivo, myogenesis is triggered by myogenic regulatory factors such as MyoD, which initiates a transcriptional cascade leading to the acquisition of muscle cell fate (1,2). This cellular myogenic induction can be reproduced *in vitro*: expressing MyoD in different cell types, including fibroblasts, turns them into muscle cells (1,3). These seminal observations of transcription factor (TF)-induced reprogramming paved the way for the subsequent demonstration that forced expression of the pioneer TFs Oct4, Sox2 and Klf4 together with c-Myc can convert differentiated cells into induced pluripotent stem cells (4,5). More recently, reprogramming of fibroblasts to neurons was achieved by coexpression of the pioneer TF Ascl1 together with Brn2 and Mytl1, supporting the model of TF cooperativity based on factor co-occupancy on *cis*-regulatory modules (CRMs) to achieve cell reprogramming (6,7).

The nature of the transcription factors present in cells at the onset of *Myod1* expression, directly activated by MyoD, or colocalizing with MyoD on CRMs is far from being elucidated. While most cell types can be reprogrammed to the myogenic pathway by forced MyoD expression (8,9), there are exceptions such as P19 (10) and human ES cells (11) which are refractory to MyoD reprogramming. The

*To whom correspondence should be addressed. Tel: +33 1 44 41 24 13; Fax: 33 1 44 41 24 21; Email: pascal.maire@inserm.fr
Correspondence may also be addressed to Vincent Hakim. Email: hakim@lps.ens.fr
† These authors contributed equally to the work as first authors.
Present addresses:
Marc Santolini, Center for Complex Network Research, Northeastern University, Boston, MA 02115, USA.
Jean-Paul Concordet, INSERM U1154, CNRS UMR7196, Museum National d'Histoire Naturelle, 75005 Paris, France.

failure of MyoD reprogramming in these cells may be the consequence of its inability to reach its specific targets: MyoD binding profiles differ in mouse embryonic fibroblasts (MEFs) and P19 cells, showing only 30% binding site overlap across ChIPseq experiments (10). This suggests that TFs specific to MEFs and P19 cells may influence MyoD binding. Efficient MyoD reprogramming may thus require co-occupancy with one or more TFs on specific MyoD CRMs. Indeed, MyoD needs 'beacon factors', such as the Pbx and Meis homeoproteins, to be recruited to otherwise inaccessible E- boxes (12). In addition, TFs encoded by genes controlled by MyoD may participate in this reprogramming, as exemplified by Mef2 proteins, which have been shown to synergize with MyoD to activate a battery of target genes (13,14). The precise role of these various transcription factors in MyoD cell reprogramming remains to be tested.

During embryogenesis, myogenic cell fate choice is driven by the activation of the myogenic determination genes *Myf5*, *Myod1* or *MRF4* in competent cells of the embryo (2). Upstream regulators contributing to the activation of these determination genes have been characterized in vertebrates, among which Gli, LEF/TCF, Pax and Six (2). Downstream, Myf5, MyoD and MRF4 share a common target, *Myogenin*, with its early expression also under the control of Six and Mef2 (15–17). The homeoproteins of the Six family are expressed in various cell types in the embryo including dermomyotomal cells of the somites and mesodermal cells of the branchial arches where they are required to activate *Myod1* and *Myogenin* (17–19) by binding to DNA elements called MEF3 (20). Mutation of the MEF3 sites in the CRMs of *Myod1* or of *Myogenin* impairs the activity of a *LacZ* reporter in transient transgenic assays, demonstrating the requirement of Six binding for *Myod1* and *Myogenin* CRM activation *in vivo*. Six homeoproteins thus appear as good candidates to cooperate with MyoD at the onset of its expression to trigger a myogenic fate in uncommitted cells.

Four *Six* genes (*Six1*, *Six2*, *Six4* and *Six5*) are expressed in embryonic myogenic lineages, before and concomitantly with *Myod1* expression. Testing the potential synergy between Six and MRFs *in vivo* has been hampered by the overlapping expression pattern of the four Six proteins, which share a common MEF3 binding site. Moreover, *Myod1* is itself a direct target of Six proteins (19,21), and it would have to be expressed independently of them to test a potential synergy with Six proteins. In cellular models of myogenesis (myogenic cells, MyoD expressing C3H/10T1/2 fibroblast cells), Six1, Six4 and Six5 have been found to participate in muscle gene activation together with MyoD (22–26). Although no direct biochemical interaction between MyoD and Six has been characterized, both Six1 and Six4 have been shown to synergize with MyoD to activate *Myogenin* and other MyoD targets during myoblast differentiation, and Six4 is present on 36% of MyoD ChIP peaks identified in C2C12 myotubes (24,26). *Myogenin* expression and myogenic differentiation efficiency are reduced in C3H/10T1/2 cells overexpressing MyoD after knockdown for *Six4* (26), and in myogenic C2C12 cells after knockdown for *Six1* plus *Six4* (24). However myogenesis still occurs in the absence of *Six1* and *Six4*, suggesting compensatory rescue mechanisms during myogenic fate acquisition.

What could these compensatory mechanisms be? During embryogenesis the presence of Six2 and Six5 may be responsible for the expression of *Myod1* observed in *Six1Six4dKO* (19,27). In cellular models, *Six1* and *Six4* have been expressed before their knockdown, and the genome might have kept some epigenetic memory of previous *Six1* or *Six4* expression. Moreover, it has been shown that Six4 can recruit the H3K27me3 UTX demethylase in myogenic cells and modulate their epigenetic landscape (26,28). Consequently, the chromatin landscape of these knockdown cells may be more permissive for MyoD accessibility than cells that have never expressed these two homeoproteins. Furthermore, it cannot be excluded that after *Six1* or *Six4* knockdown in C2C12 or C3H/10T1/2 cells, some of their downstream targets may still be expressed, or that Six5 maintains a basal level of Six activity sufficient to synergize with MyoD.

Therefore, a fundamental question is: are Six homeoproteins required for MyoD to activate the myogenic program during embryogenesis, and are they required for MyoD reprogramming of non-myogenic cells? A way to address this question is to analyze the consequences of *Six1/2/4/5* loss once *Myod1* has been activated *in vivo*, or to use a cellular model devoid of Six activity and analyze the consequences of forced MRF expression on myogenic commitment. *Six1/4* mutant mouse embryonic fibroblasts (MEFs) from the ventral body of E13.5 embryos with very low *Six2* and *Six5* expression provide an ideal model system to investigate cell reprogramming by ectopic MyoD expression in the absence of Six activity, and thus identify roadblocks in the cell reprogramming properties of MyoD.

In this manuscript, we explore the hypothesis that Six and MyoD cooperate at the CRM level to activate a battery of downstream target genes required for myogenic fate acquisition. We show that in *Six1/4* mutant MEFs MyoD reprogramming ability is impaired, and that MyoD targets remain repressed in these mutant cells expressing MyoD. Nevertheless, rescue experiments with either Six1 or Six4 and MyoD allow *Myogenin* expression in these *Six1/4* mutant MEFs, showing that they did not lose the abilility to adopt a myogenic fate. Using bioinformatics approaches, we identify three transcription factor families, namely, EBF, Mef2, and Pbx/Meis, the genes of which are under the control of Six and MyoD, and that colocalize with Six and MyoD on CRMs to activate a battery of genes expressed during MEF reprogramming.

MATERIALS AND METHODS

Obtaining a MEF3 PWM

To build a PWM for MEF3, we used a list of 46 binding sites that were previously tested by electrophoretic mobility shift assay on the basis of their proximity to the *Myogenin* MEF3 consensus GAAACCTGA. The EMSA experiments segregated these sites into 32 positive and 14 negative sites (see Supplementary Table S1). For the learning procedure, we divided the positive sites into a training set of 14 sequences for which alignments were available in other eutharian species, and a test set of 18. We then sought to iden-

tify the PWM that would best distinguish the positive from negative sites. Because all these sites were tested on the basis of their proximity to the MEF3 consensus, we note that this procedure is more stringent than a simple distinction of positive sites from random sites. We generated two types of PWMs: a 'reference' PWM built from the 14 training set sequences, and a 'refined' PWM built using Imogene (29), a bayesian *de novo* motif finder that uses phylogeny, on the alignments of the training sequences with their orthologues in 11 other mammals. The information content of the PWM generated was varied between 7 and 12 bits by unit of 1, and both evolutionary models used in Imogene were employed. For each PWM generated, a ROC curve was built, indicating for varying detection thresholds (or sites 'scores') the proportion of positive sequences detected in the test set (TPR for true positive rate) against the proportion of negative sequences detected in the negative set (FPR for false positive rate). We show in Supplementary Figure S1 the optimal ROC curves obtained for both the reference and evolutionary cases. Both curves are relatively similar and allow a stringent detection of 45% of positive sites without any detection of a negative site, for a detection threshold of 7 bits. The refined PWM shows slightly better performance than the reference PWM at high scores (lows FPRs), and was finally selected for the rest of the work.

Microarrays

After validation of RNA quality with the Bioanalyzer 2100 (using Agilent RNA6000 nano chip kit), 50 ng of total RNA were reverse transcribed following the Ovation PicoSL WTA System (Nugen). Briefly, the resulting double-strand cDNA was used for amplification based on SPIA technology. After purification according to the Nugen protocol, 5 μg of single strand DNA were used for generation of Sens Target DNA using the Ovation Exon Module kit (Nugen). 2.5 μg of Sens Target DNA were fragmented and labelled with biotin using Encore Biotin Module kit (Nugen). After control of fragmentation using a Bioanalyzer 2100, the cDNA was then hybridized to GeneChip® Mouse Gene 1.0 ST (Affymetrix) at 45°C for 17 h. After overnight hybridization, the ChIPs were washed using the fluidic station FS450 following specific protocols (Affymetrix) and scanned using the GCS3000 7G. The scanned images were then analyzed with Expression Console software (Affymetrix) to obtain raw data (cel files) and metrics for Quality Controls. The analysis of some of these metrics and the study of the distribution of raw data show no out-lier experiments. RMA normalization was performed using R and normalized data were subjected to statistical tests.

Definition of Six only, MyoD only and Six+MyoD categories of genes

We distinguished different classes of genes (Supplementary Figure S4 and Table S4) as follows: A gene was considered as upregulated (or downregulated) by Six only, under the proliferating (or differentiating) condition, if it was upregulated (or downregulated) more than 1.5 times in wild-type (wt) MEFs compared with mutant MEFs in that condition, and was upregulated (or downregulated) less than 1.5 times in wt MEFs expressing MyoD compared to wt MEFs. This criterion was used to define in the same way genes upregulated or downregulated by MyoD only under proliferating or differentiating conditions. Finally, genes upregulated or downregulated by Six+MyoD under proliferating or differentiating conditions were defined by their upregulation (or downregulation) in wt MEFs expressing MyoD >1.5 compare to their maximum (or minimum) expression value in wt MEFs, mutant MEFs and mutant MEFs expressing MyoD under the same conditions.

MyoD ChIPseq reads

MyoD ChIPseq reads were downloaded from the Sequence Read Archive (SRA010854). We focused on three conditions: C2C12 cells at 50% confluence, 95% confluence and 48h after differentiation. We used MAQ v0.6.6 to align reads to the mouse reference genome (NCBI v37, mm9) as detailed in (30). We then used MACS v1.4.2 to define peaks, with parameters gsize = 1.87e9, mfold = 10, tsize = 35, *P* value = 1e−2, bw = 200. Peaks were extended to 1 kb regions by taking −500/+500 bp around the peak center. Finally, peaks from the three conditions were merged to provide the final set of peaks used here.

Prediction of conserved MEF3 binding sites

The previously defined optimal MEF3 PWM was used to scan the genome. A stringent scanning threshold of 10 bits was used, ensuring a low false positive rate. To study conservation of the binding sites in other related species, we used 12 eutherian EPO alignments from the Ensembl website (http://www.ensembl.org/). The 12 species were divided into five groups of related species: (i) mouse and rat, (ii) human, chimpanzee, gorilla, marmoset, orangutan and macaque, (iii) cow and pig, (iv) dog and (v) horse. A binding site was defined as conserved when it was found in at least three of these groups in a ±20 bp region around the original binding site on the mouse sequence, as described in (29).

Hybrid *de novo* motif search

The motif search was conducted using both *de novo* motifs from Imogene, an algorithm initially devised for the detection of TF binding sites in *Drosophila* (31) and extended to vertebrates genomes (29), and motifs from Jaspar (32) and Transfac (33) databases. First, Imogene was used to identify motifs from the sequences of interest using thresholds of 7, 8 and 9 nats and the Felsenstein evolutionary model. Then, motifs from Jaspar and Transfac databases were used as prior motifs in Imogene and were refined for the sequences of interest using similar thresholds. For each motif, an optimal scanning threshold was obtained by requesting a stringent false discovery rate (FDR) of 1% of background sequences containing at least one binding site. More precisely, background sequences of 1kb (Imogene) were scanned and the FDR was computed as the number of sequences with at least one binding site. These optimal scanning thresholds were used to compute a TPR for each motif as the proportion of training enhancers with at least one binding site. Mo-

tifs were finally ranked by TPR. The 15 motifs with TPR >25% were finally selected for further analysis.

Cell culture

Wt and *SixdKO* MEFs were isolated from the ventral skin of E13.5 embryos and cultured until naturally immortalized. Fibroblasts were cultured in DMEM (Invitrogen), supplemented with 10% FBS (Invitrogen) and 0,1% penicillin–streptomycin (Invitrogen). All cultures were grown in humidified incubators at 37°C under 5% CO_2.

Stable MyoD transductions were performed using lentiviral vectors for 6 h in the presence of Polybrene (4 mg/ml; Sigma-Aldrich) with a multiplicity of infection (MOI) of 10 (i.e. using 500 000 transducing units [TU] of vectors to transduce 50 000 cells). MyoD induced cell lines were established using a doxycycline inducible *Myod1* lentivirus (34) and selected clonally. One wt and one *SixdKO* clone were used in this study, both of which produced MyoD after doxycycline induction in 100% of the cells. For myoconversion, doxycycline (2 mg/ml; Sigma-Aldrich) was added in the differentiation medium, composed of DMEM with 2% Horse Serum (Invitrogen), for 6 h. After 6 h cells were lysed and nuclear proteins were prepared with the NE-PER kit (78833, Thermo Fisher scientific). For establishment of polyclonal *SixdKOs1* and *SixdKOs4, SixdKO* fibroblasts were infected with human Six1 or Six4 lentiviral particles and selected with blasticidin (10μg/ml) for 4 passages.

Immunohistochemistry

Cultured cells were fixed in 4% paraformaldehyde, then incubated for 1 h in blocking buffer (Horse serum 2.5%; Vector, triton 0.1%; PBS) then with the following antibodies: MyoD (1/100^e; SCBT), Myogenin (1/100^e; SCBT), MF20 (1/100^e; DSHB) at RT for 2 h. Appropriate species specific fluorescent secondary antibodies (1/1000e; Jackson ImmunoResearch) were used along with DAPI (4′,6′-diamidino-2-phénylindole) at RT for 1 h. Images were captured using a Zeiss fluorescent microscope.

Electroporation

Five hundred ng of expression plasmid for *Myod1* or for *GFP* were transfected into 2.0×10^5 MEFs using the Neon transfection system (Invitrogen), according to the manufacture's instructions. Two days after electroporation, MEFs were sampled as proliferating fibroblasts. Myogenic differentiation was induced by replacement of the growth medium with the differentiation medium (DMEM high glucose Glutamax (31966, Invitrogen), 2% HS, 1% Penicillin-streptomycin) 2 days following electroporation.

Luciferase assays

MEFs were co-transfected with a tk-Renilla plasmid that allows the normalization of transfection efficiency. MEFs were lysed in Passive Lysis Buffer (Dual-Luciferase Reporter Assay System, Promega) and shaken for 15 min. Lysates were used for measurement of Luciferase and Renilla activities (Dual-Luciferase Reporter Assay System, Promega). The ratio between Luciferase and Renilla activities was further used to compare the induction of Luciferase activity between samples. Triplicates ($n = 3$) were obtained for the experiments presented in Figure 4, Supplementary Figures S5 and S6. Two to four experiments ($N = 2$–4) were performed in triplicate for the experiments presented in Figure 6 and Supplementary Figure S7.

RT-qPCR

Total RNA was extracted from cell cultures with TRIzol reagent (Sigma-Aldrich) and treated with DNAseI (Ambion Life Technologies 'DNA-free'). 500 ng of RNA was reverse transcribed into cDNA, using High Capacity cDNA Reverse Transcription Kits (Applied Byosystems). RT products were subjected to qPCR amplification with Sybr Green Master Mix (Roche) on a LightCycler 480 (Roche). Primers used are listed in Supplementary Table S2.

Conversion from mouse symbols to human entrez IDs

In order to compute pathway enrichment, we first computed a table converting mouse gene symbols to human entrez IDs. We used the UCSC genome browser mm9.kgXref, mm9.hgBlastTab and hg19.kgXref conversion tables available on the mySQL host genome-mysql.cse.ucsc.edu. The kgXref tables were used for conversion between symbols and entrez IDs while the Blast table was used to obtain the human orthologs of mouse genes.

Pathway enrichment

Pathways were taken from both MSigDB v3.1 (35), focusing on canonical pathways from KEGG, BIOCARTA, REACTOME and the STKE signaling collection, and Wikipathways (36), with a total of 1511 pathways.

Statistical analyses

The data were processed using Prism6.0e and R software. Values were reported as mean + standard error. Comparisons between two groups were made with Student's *t*-test or ANOVA. Pathway enrichment was computed using hypergeometric *P*-values. For Figure 3B, departure of the histogram from a uniform distribution was assessed using a chi-square test (chisq.test function in R). For Figure 3D, *Z* scores of the mean FC of the first peak were computed by comparing to a random distribution of mean FCs expected when averaging over the same number of randomly selected genes. We then computed $Z = (\mathrm{FC} - \mu)/\sigma$ after 10 000 realizations, where μ is the average and σ the standard deviation of the expected mean FC of randomly selected CRMs. For Figure 3E, null distributions were obtained by randomizing the fold-changes across genes and Z scores were computed after 10 000 randomizations.

Gel mobility shift assays

Nuclear proteins from adult muscles were purified as reported previously (37). Nuclear proteins from *Myod1*-inducible wt and *SixdKO* MEFs were prepared 6 h after Doxycycline addition (or without Doxycycline) in

low serum medium to allow cell differentiation and isolated with the NE-PER kit (78833, Thermo Fisher scientific). Recombinant mouse Six1 and Six4 proteins were obtained from *Six1* and *Six4* full-length mouse cDNA cloned into the pCR3 vector (Clontech) with a T7 transcription/translation kit (Promega). 4 μg of nuclear proteins or recombinant Six1 and Six4 proteins were incubated with α-P32 radiolabelled oligonucleotides in the presence of 1 μg of polydI-dC, and the presence of hundred fold molar excess of oligonucleotide for competition experiments as described previously (37). The sequences of the oligonucleotides used are presented in Supplementary Table S3.

Plasmid construction

For the construction of pGL3-*Dact1-enh*, pGL3-*Tnnc1-enh,* pGL3-*Tnnc2-enh,* pGL3-*Mtap4-enh,* pGL3-*Etv1-enh,* pGL3-*Murc-enh,* pGL3-*Lrrc17-enh1,* pGL3-*Lrrc17-enh2,* pGL3-*Lrrc17-enh3,* pGL3-*Srl-enh,* pGL3-*Mef2c-enh1,* pGL3-*Mef2c-enh2,* pGL3-*Mef2c-enh3,* pGL3-*Ttn-enh,* mouse genomic DNA was first used as a template to clone each DNA fragment with the following primers:

Dact1-enh forward 5′-TCTATCGATAGGTACCCC CCAGACACAGTCAACCTT-3′ and *Dact1-enh* reverse, 5′-GCTTGATCTCGGTACGGGAGTCAGGGTCAT TGAGA -3′,

Tnnc1-enh forward 5′-TCTATCGATAGGTACCGG TCTCAGCTTTCAGCATCC-3′ and *Tnnc1-enh* reverse, 5′-GCTTGATCTCGGTACTGAAGGGACTCACCATC CTC-3′,

Tnnc2-enh forward 5′-TCTATCGATAGGTACCAC TTCTCTTGGGTGGGGAAG-3′ and *Tnnc2-enh* reverse, 5′-GCTTGATCTCGGTACTGCCATTCACTGGTGAT CTT-3′,

Mtap4-enh forward 5′-TCTATCGATAGGTACCCTTT ACTTGAGATTGCCTCCTG -3′ and *Mtap4-enh* reverse, 5′-GCTTGATCTCGGTACAAACCCCTGAAAGATGG TAGG-3′,

Etv1-enh forward 5′-TCTATCGATAGGTACCCG ACAAGTAGGCGGATGTTT-3′ and *Etv1-enh* reverse, 5′-GCTTGATCTCGGTACAGCCACTGGTCGTAA GCAAC-3′,

Murc-enh forward 5′-TCTATCGATAGGTACCTGAG CTATGTGAGACCCTGTC-3′ and *Murc-enh* reverse, 5′-GCTTGATCTCGGTACCTGCCCGGTGAATTATG ACT-3′,

Lrrc17-enh1 forward 5′-TCTATCGATAGGTACCAG CATCCTCTGGTTTGGATG-3′ and *Lrrc17-enh1* reverse, 5′-GCTTGATCTCGGTACATGCAACCAATTACCCC AAA-3′, *Lrrc17-enh2* forward 5′-TCTATCGATAGGTA CCCTCGGTTCTCCACCAAGTTT-3′ and *Lrrc17-enh2* reverse 5′-GCTTGATCTCGGTACGAGGTCACCAA GAGGCTGTG-3′, *Lrrc17-enh3* forward 5′- TCTATC GATAGGTACCCCCTGTCCCTTCATCTTCAA-3′ and *Lrrc17-enh3* reverse 5′- GCTTGATCTCGGTACACCCT CCAGCCAAGAAAAGT-3′,

Srl-enh forward 5′-TCTATCGATAGGTACCGACTCC CAAGACTGGCTTCC-3′ and

Srl-enh reverse 5′-GCTTGATCTCGGTACGGGGA AATTCAACCACTAGC-3′,

Mef2c-enh1 forward 5′-TCTATCGATAGGTACCGC TTTCTAATTTGGGAGCATGA-3′ and *Mef2c-enh1* reverse 5′-GCTTGATCTCGGTACTCAAATGTTGT TGGCGTTGT-3′, *Mef2c-enh2* forward 5′-TCTATCGA TAGGTACCCACCTATTGGAAGCAGATATTTTG-3′ and *Mef2c-enh2* reverse 5′-GCTTGATCTCGGTACACC GTTTGCTCATGTTTGAA-3′, *Mef2c-enh3* forward 5′-TCTATCGATAGGTACCAGTCGTGCCAAAGAAA ATGG-3′ and *Mef2c-enh3* reverse 5′- GCTTGATCTCGG TACTCAGGAGTCCTTAGCTGCATT-3′,

Ttn-enh forward 5′-TCTATCGATAGGTACCAAAG GTTTCATCTTGCTGACC-3′ and

Ttn-enh reverse 5′-GCTTGATCTCGGTACTTTCCTG TCTGTAAGCCAGTGT-3′, respectively. Amplified DNA fragments were subsequently inserted into KpnI digested pGL3 TATA plasmid using the GeneArt kit (Life Technologies).

For the construction of the mutation series of pGL3-*Lrrc17-3* MEF3, E-box, mot14, mot5 sites, mot7 sites, Meis1 or mot10 sites of the enhancer were mutated as shown in Supplementary Figure S2. For the construction of the mutation series of pGL3-*Tnnc2*, MEF3 sites, E-box sites, Mef2 sites, mot4 or mot5 sites of the enhancer were mutated as shown in Supplementary Figure S2. All plasmid sequences were confirmed by sequencing.

RESULTS

The expression of myogenic genes depends on both Six1/4 and MyoD

To study the genome-wide synergy between Six and MyoD during myogenesis, we developed a model system to monitor the expression of MyoD and Six1/Six4 proteins. Mouse embryonic fibroblasts (MEFs) can be reprogrammed by MyoD, leading to the activation of hundreds of genes specifically expressed in skeletal muscle cells. Since *Six1/4* genes are essential for all the steps of muscle development, we questioned whether Six was needed for MyoD reprogramming properties. *Six1;Six4dKO* (*SixdKO*) and wt MEFs were isolated from E13.5 embryos, put in culture and transfected with an expression plasmid coding for MyoD. Two days after transfection, proliferating MEFs were placed in differentiation medium (low serum) for another 2 days and assessed for *Myogenin* expression. Wt MEFs express *Six1* and *Six4* but not *Myod1* or *Six2*, while the expression of *Six5* is low, as estimated by qPCR experiments (Figure 1A, and data not shown). *Myogenin* mRNA was detected in wt MEFs transfected by MyoD, in *SixdKO* MEFs transfected by MyoD + Six1 or by MyoD + Six1-VP16, a chimeric transdominant positive Six1 protein, but not in *SixdKO* MEFs transfected by MyoD only (Figure 1B). A similar observation was made after lentiviral infection of wt and *SixdKO* MEFs with a Doxycycline inducible *Myod1* expression vector: expression of MyoD was unable to activate *Myogenin* expression in mutant MEFs as shown by qPCR experiments (Supplementary Figure S3A), or by immunofluorescence (Figure 1C). We detected *Myogenin* mRNA (Supplementary Figure S3A) and nuclear protein accumulation (Figure 1D) in *SixdKO* MEFs rescued by Six1 or Six4 in the presence of MyoD, showing that *SixdKO*

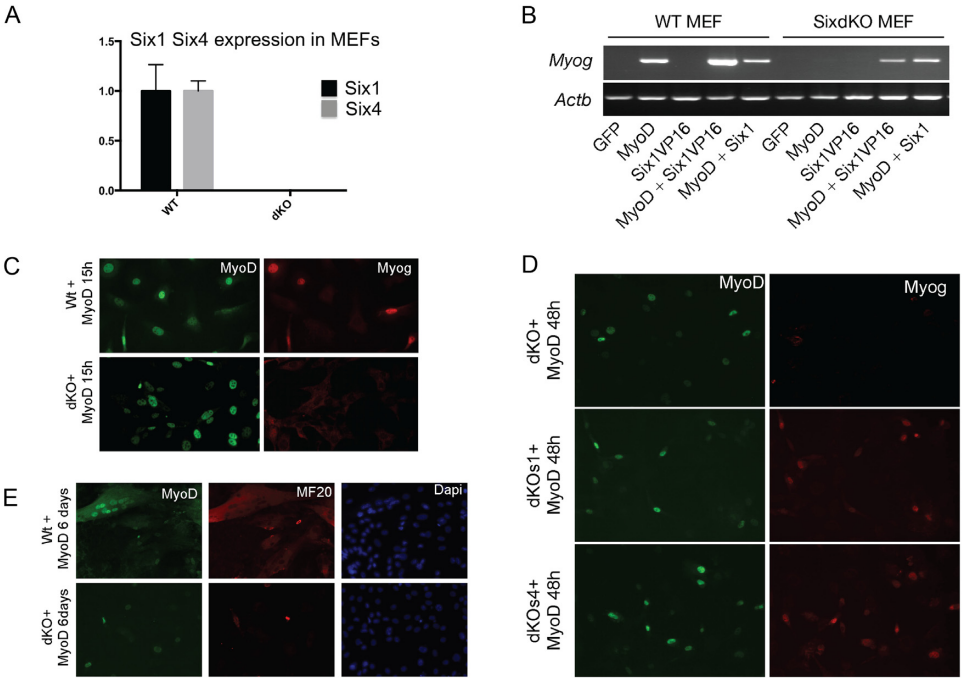

Figure 1. *SixdKO* MEFs are unable to engage the myogenic program in the presence of MyoD. (**A**) *Six1* and *Six4* mRNA levels in wt and *SixdKO* MEFs as estimated by qPCR experiments. (**B**) *Myogenin* (*Myog*) expression in wt or *SixdKO* MEFs after forced expression of MyoD, and in *SixdKO* MEFs after forced expression of MyoD + Six1 or MyoD + Six1-VP16, as estimated by PCR experiments with cDNA from transfected cells with the expression vectors mentioned. *β-Actin* (*Actb*) was used as an internal control for the PCR experiments. (**C**–**E**) Immunocytochemistry of stable wt and *SixdKO* MEFs transduced with *Myod1* cDNA, and expressing MyoD after Doxycycline addition. (**C**) MyoD and Myogenin immunodetection 15 h after Doxycycline addition in wt and mutant cells. (**D**) MyoD and Myogenin immunodetection 48 h after Doxycycline addition in stable mutant MEFs rescued by *Six1* (dKOs1) or by *Six4* (dKOs4) and further transduced by *Myod1*. (**E**) MyoD and sarcomeric Myosin heavy chain (revealed by MF20 antibodies) immunodetection 6 days after Doxycycline addition. Note that in wt MEFs, endogenous *Myod1* expression has been turned on after Doxycycline induction.

MEFs did not lose their ability for myogenic reprogramming. Myosin heavy chain expression, detected by MF20 antibodies, was obvious in multinucleated myotubes formed after expression of MyoD in wt MEFs, but was not detected in *Six1/4* mutant MEFs expressing MyoD, in which no myotube formation was observed (Figure 1E). We then transfected an expression vector coding for a dominant positive MyoD-E12 chimeric protein (38) in *SixdKO* MEFs, but detected no *Myogenin* expression, suggesting that it was not the absence of MyoD partners in mutant MEFs that prevented MyoD trans-activation of *Myogenin* (Supplementary Figure S3B).

As the expression of many muscle genes, including *Myogenin*, is controlled by the regulatory factor Mef2, we hypothesised that the reprogramming of *SixdKO* MEFs by MyoD could be rescued by forced MyoD and Mef2 expression. As shown in Supplementary Figure S3C, expression of both MyoD and Mef2 failed to activate *Myogenin* expression in *SixdKO* MEFs, demonstrating that Six1/4 is essential for MyoD and Mef2 mediated muscle gene expression in MEFs.

Altogether, these data indicate that *Myogenin*, a known early target of MyoD and Mef2 (15,16) and Six (17) cannot be activated by MyoD and Mef2 in the absence of Six proteins. Whether absence of *Myogenin* activation in *SixdKO* MEFs + MyoD is an isolated example or reflects the more general inability of MyoD to reprogram *SixdKO* MEFs remained to be determined.

To identify MyoD targets under the control of both Six and MyoD other than *Myogenin*, we electroporated wt and *Six1/4* mutant MEFs with an expression vector coding for MyoD to analyze gene expression levels using Affymetrix arrays (Figure 2A). Two days after electroporation, proliferating MEFs were either harvested and their RNA extracted, or placed in differentiation medium (low serum) for another 2 days before RNA extraction. The importance of the synergistic dependence between Six and MyoD to activate the myogenic program was illustrated by the expression of endogenous *Myogenin* expression revealed by qPCR assays. It is strictly correlated with the presence of both *Six1/4* and *Myod1* and significantly higher than expected based on the observed increase in *Six1* and *Myod1* (Figure 2B). Contrary to what was observed in *SixdKO* MEFs, expression of MyoD in wt MEFs led to the upregulation (>1.5-fold) of the expression of 761 genes in growth and differentiating conditions, and the downregulation (>1.5-fold) of 391 genes (Figure 2C, Supplementary Table S4). Pathway enrichment analysis indicated that the 'striated muscle contraction' pathway was predominant for upregulated genes, while the interferon signaling pathway was highly represented for down-regulated genes (Supplementary Figure S4).

Pathway enrichment analysis revealed that in both proliferating and differentiating conditions MyoD was unable to activate muscle type specific genes in *SixdKO* MEFs (Supplementary Figure S4 and Table S4). In *SixdKO* MEFs, a

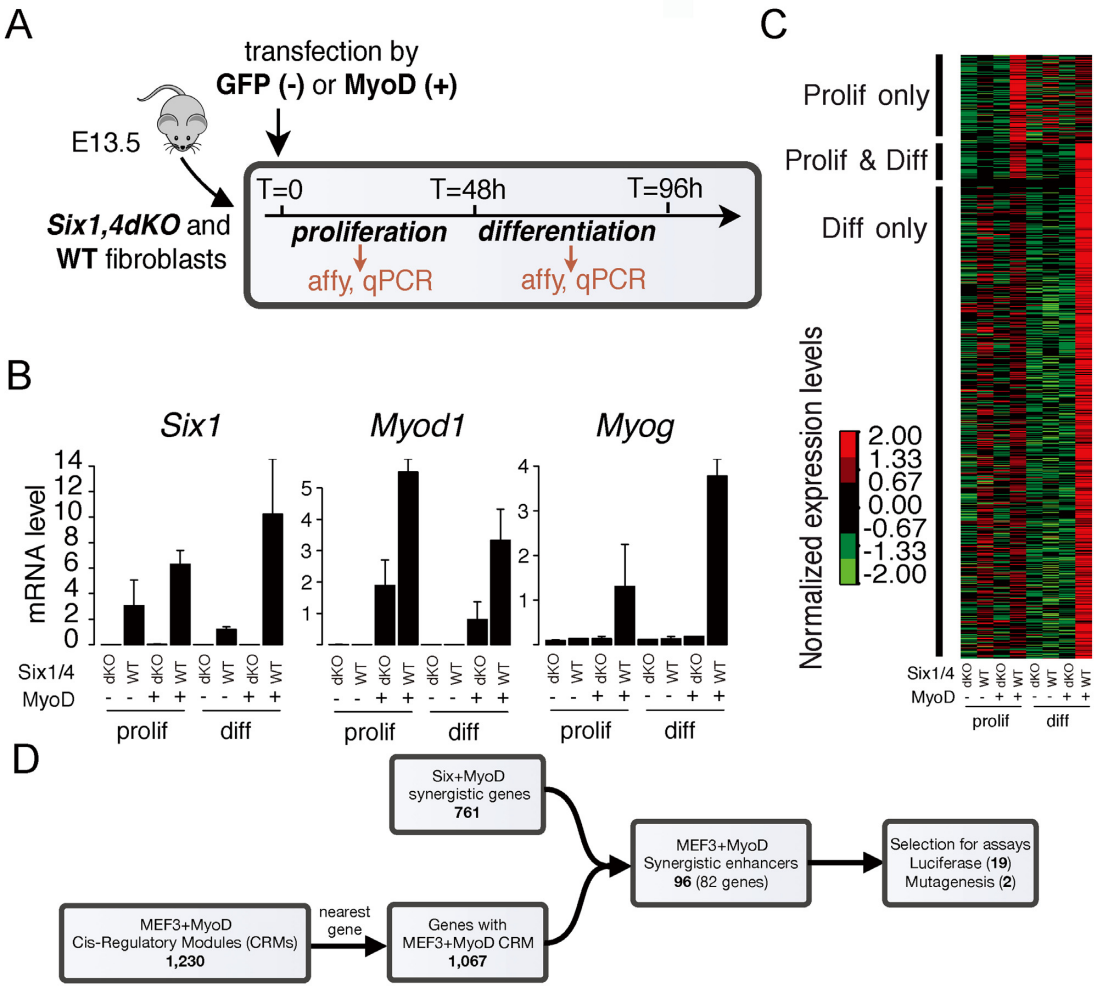

Figure 2. (**A**) Overall strategy for isolation and culture of wt and of *SixdKO* (dKO) MEFs from wt and mutant E13.5 embryos, and their electroporation with an expression vector coding for MyoD (+) or GFP (−). (**B**) Barplots showing qPCR evaluation of the expression level of *Six1*, *Myod1* and *Myogenin (Myog)* mRNA under the eight conditions used. *Myod1* expression is the sum of exogenous *Myod1* (expression vector of the mouse *Myod1* cDNA) and endogenous *Myod1*. ($N = 3$). Prolif: proliferating conditions, diff: differentiating conditions. Wt MEFs are indicated as Six1/4 (Wt) and mutant MEFs as Six1/4 (dKO) transfected by *Myod1* (+) or by *GFP* alone (−). (**C**) Heatmap of microarray data showing 761 synergistic genes as revealed by Affymetrix experiments, and showing genes up- (red) and down- (green) regulated under the eight conditions used in this study. (**D**) Workflow of the bioinformatic analyses in this study.

few genes are activated by MyoD >1.5-fold under proliferating conditions, among which *Rfc2, Psmd10, Ssna1* and *Polr2i*, and a few genes are activated by MyoD >1.5-fold under differentiating conditions, among which *Mthfd2*, *Ahcyl2*, *Nup43* and *Aaas*.

In *SixdKO* MEFs, very few genes are downregulated on proliferation (down regulation <2-fold for *Dcp1a* for example) or differentiation (downregulation below two fold for *Tubg1* for example). We conclude from these data that MyoD is completely impaired in its ability to act as a master myogenic regulatory factor in MEFs deprived of Six1/4 activity.

Altogether, these data indicate that MyoD can reprogram wt MEFs and activate known myogenic genes such as *Myog*, *Tnnc1*, *Tnnc2*, *Myh3*, *Actc1* or *Mef2c,* and that this property of MyoD depends on the presence of Six.

Six binding sites are enriched around MyoD ChIPseq peaks, and synergistic genes are regulated by Six+MyoD CRMs

To investigate genome-wide the synergy of Six and MyoD to activate the myogenic program we looked for DNA regions that could be bound by both TFs. In order to do this, we retrieved MyoD ChIPseq data (30). This dataset consisted of MyoD bound regions of ∼300 bp in C2C12 cells under three different conditions: myoblasts at 50% of confluence, myoblasts at 95% of confluence, and differentiated myotubes. The analysis of all three conditions with fusion of overlapping regions (see Methods) yielded a total of 34,732 bound regions. We then looked for the presence of binding sites for Six1 and Six4 around the MyoD peaks. ChIPseq data for Six1 in C2C12 cells was only available for 20% of the mouse genome (24). Genome-wide ChIPseq data for Six4 has become available more recently (26). However, we noted that the overlap between the two data sets was only about 31% suggesting that many binding sites are unique to Six1

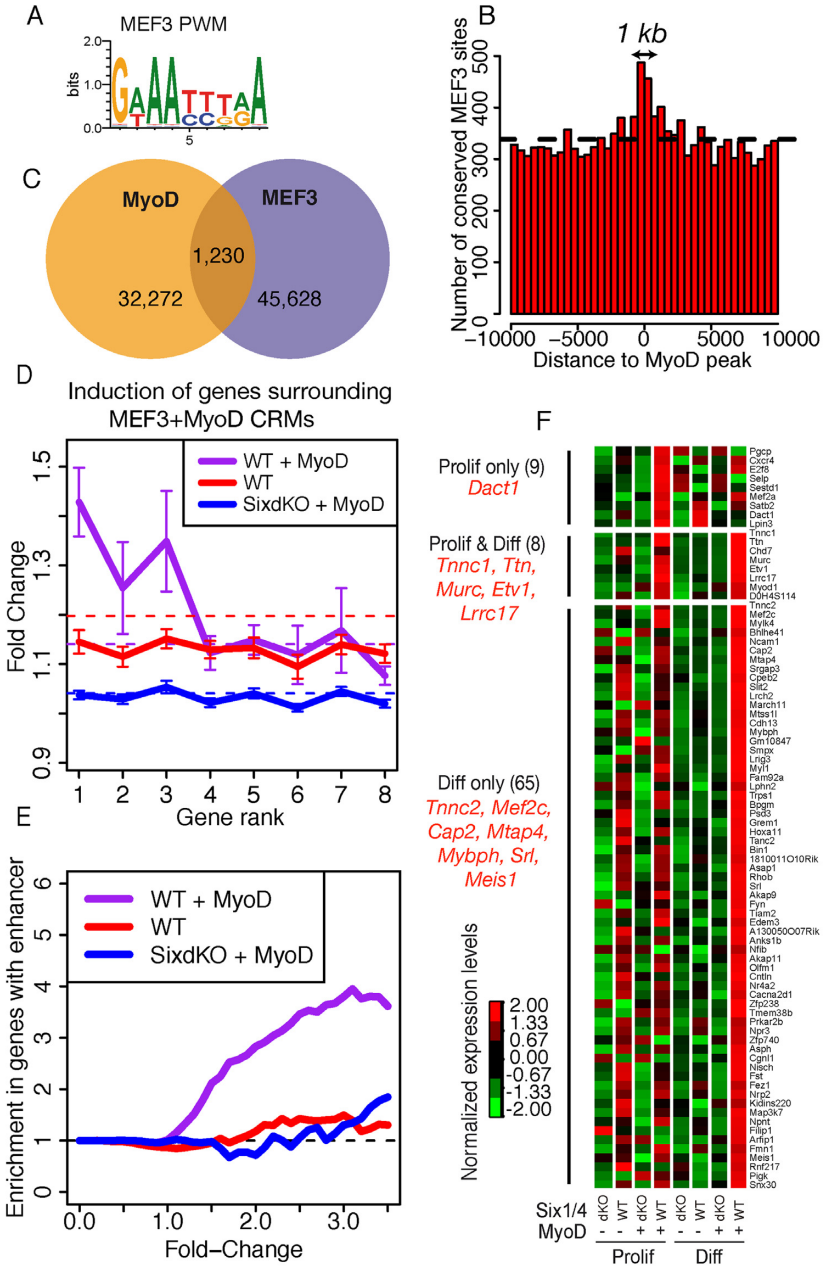

Figure 3. Identification of MyoD + Six CRMs. (**A**) Optimal MEF3 PWM. (**B**) Enrichment of conserved MEF3 sites around MyoD peak centers (position 0) showing a significant signal in a ±500 bp region around the peaks ($P = 5.4e{-}22$, chi-square test). The average count number is shown as a dashed line. (**C**) Venn diagram comparing MyoD ChIPseq data with MEF3 conserved binding sites in the ±500 bp region around the center of the MyoD peak. (**D**) Six+MyoD CRMs are associated on average with a positive synergistic effect on the transcription of the nearest gene. Gene rank corresponds to the absolute rank position of the gene from the identified CRMs. We note that genes in rank 1 (i.e. the gene with the closest TSS from the CRM) show the highest activation ($Z = 5.1$, see Materials and Methods). Dashed lines represent genome-wide averages. (**E**) The proportion of genes with a Six + MyoD CRM increases significantly with their expression Fold Change (as measured with Affymetrix microarrays). (**F**) Affymetrix data for the genes activated synergistically by Six and MyoD. More than 90% of MEFs were transfected by *Myod1* in our electroporation/transfection conditions. For each gene, expression data are normalized and centered separately under proliferating or differentiating conditions. Genes in red were further selected for Luciferase validation of their predicted CRMs.

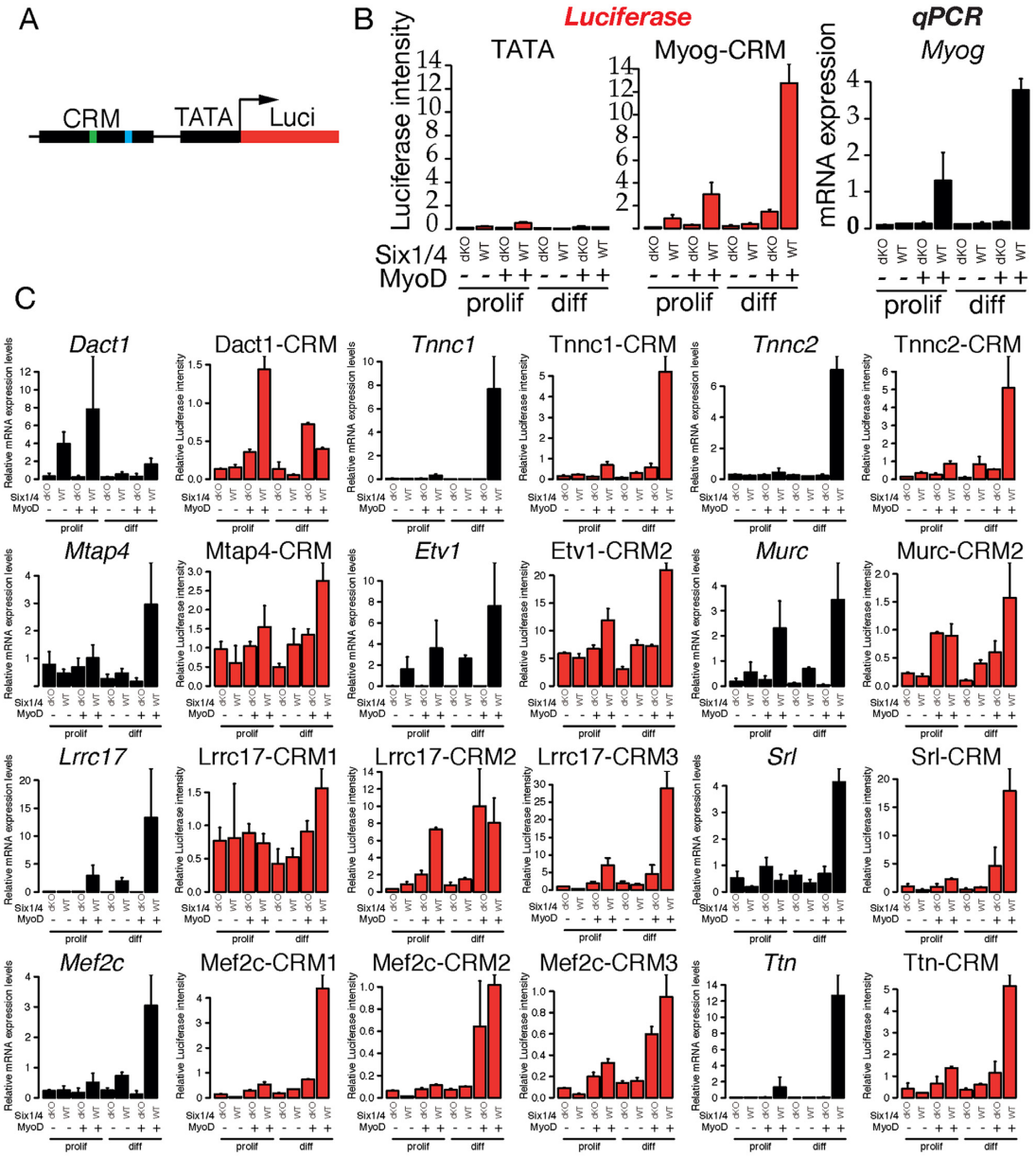

Figure 4. Validation of the Six + MyoD CRM synergy. (**A**) Fourteen potential CRM sequences of 1 kb were cloned 30 bases upstream of a minimal promoter (−35 to +45 of the *Aldolase A* muscle promoter consisting of a TATA box and associated 5′ UTR). (**B** and **C**) Wt and *SixdKO* MEFs were transfected with the CRM-Luciferase construct of the genes mentioned, and Luciferase activity (red bars) was measured under the 8 conditions described in Figure 1. qRT-PCR experiments were performed to measure mRNA levels of the corresponding genes, as an indicator of endogenous gene activity. (**B**) The activity of the empty *TATA-Luciferase* and of the −183 bp *Myogenin* promoter were used as negative and positive controls respectively. (**C**) Comparison of endogenous mRNA accumulation of the associated nearest gene and Luciferase activity under the 8 conditions for the 14 synergistic MyoD+MEF3 CRMs. Note that several potential CRMs were associated with *Etv1*, *Murc*, *Lrrc17* and *Mef2c*. Error bars correspond to the standard deviation across 3 technical triplicates.

and Six4 rather than shared by these two proteins. Thus, we decided to follow a bioinformatics PWM approach to obtain a genome-wide map of Six binding sites, independently of the particular identity of the Six protein and independently as well of cell differentiation states and culture conditions that impinge on the binding of transcription factors to their targets. We built a MEF3 PWM (29,31,39–40) using known binding sites taken from EMSA data and from the literature (see Material and Methods, Supplementary Table S1 and Figure 3A and Supplementary Figure S1). This PWM was then used to predict putative binding sites in the whole genome, with the requirement of conservation among 12 eutherian species (see Material and Methods). This conservation requirement has proven to be a good predictor of functionality for non-coding sequences (41). The extent of the association between MyoD and Six was then studied by computing the enrichment in conserved MEF3 sites around MyoD ChIPseq peaks. MEF3 sites were significantly en-

riched in the ±500 bp regions around the centers of MyoD peaks, as shown by the enrichment signal (Figure 3B). This led us to define putative Six+MyoD CRMs as DNA regions of 1kb centered on MyoD ChIPseq peaks (Figure 3B) and containing at least one conserved MEF3 site. This procedure yielded 1,230 Six+MyoD putative CRMs (Figure 3C, Supplementary Table S5).

We then searched for a link between the previously defined Six + MyoD CRMs and a change in nearby gene transcription. With this aim, each of the 1230 CRMs was associated to the nearest TSS, defining a set of 1067 putatively regulated genes. In agreement with the idea that these CRMs have a regulatory effect on the nearest gene, the CRMs were on average associated with a synergistic expression behavior of the nearest gene (Figure 3D). In addition, the proportion of genes with a Six+MyoD CRM increased with the fold-change of corresponding Affymetrix mRNA values as compared with the Six alone (wt) or MyoD alone (*SixdKO* + MyoD) conditions (Figure 3E). This was evaluated by computing an enrichment as the ratio of the proportion of genes with an associated CRM above a given threshold divided by the genome-wide proportion of genes with an associated CRM. Statistical significance was assessed by computing the Zscore of the observed enrichment at FC = 1.5 (see Materials and Methods). We found $Z = -0.79$ (WT), $Z = -0.11$ (SixdKO + MyoD) and $Z = 11$ (WT + MyoD), showing that there was a significant enrichment only in the Six + MyoD case. Overall, we found 82 'synergistic genes' with an associated Six+MyoD CRM, for which the expression under proliferating or differentiated conditions was higher by a factor of at least 1.5 in the WT fibroblasts transfected with MyoD compared with those deprived of Six1/4 but expressing MyoD or with those expressing Six1 and Six4 only (Figures 2D and 3F). Surprisingly, some known genes submitted to a synergistic activation by Six and MyoD, like *Myogenin* or *Tnnt3,* are absent from this list. This can be explained by the fact that no conserved MEF3 site was found around the MyoD peak using our criteria. In the case of *Myogenin*, for example, the conservation criterion is not met because the sequence cannot be aligned in enough species. In addition, the threshold used for detection of MEF3 sites might simply be too stringent in some cases.

These 82 synergistic genes have further been divided into three categories: genes showing a synergy under proliferating conditions only (nine genes among which *E2f8*, *Cxcr4*, *Satb2*, *Mef2a, Sestd1* and *Dact1*), in both proliferating and differentiating conditions (eight genes among which *Etv1/Ets1*, *Lrrc17*, *Murc* and *Myod1*), and in differentiating conditions only (65 genes among which *Tnnc2*, *Mef2c*, *Grem1*, *Nfib*, *Fst*, *Meis1*…). It is notable that two *Mef2* genes are under the control of Six + MyoD expression, and that Mef2 transcriptional activity is much higher in wt MEFs expressing Six and MyoD than in mutant MEFs expressing MyoD (Supplementary Figure S5). Interestingly it should also be noted that *Sestd1* and *Dact1* belong to the same planar cell polarity (PCP) genetic pathway (42) and may be important to amplify myogenic progenitors (43) and that *Grem1* and *Fst* are two secreted proteins with BMP antagonist activity, which may be required for non-myogenic cells to adopt a muscle fate (44).

Validation of the predicted Six+MyoD CRMs

We then tested whether the Six + MyoD CRMs that we had identified adjacent to the 82 genes were responsible for the transcriptional activation of these associated genes (Figure 2D). In order to do this we selected 19 CRMs in the vicinity of 12 genes among the three above-defined categories (proliferation only, proliferation & differentiation, differentiation only). We first confirmed by qPCR the synergistic behavior of the associated genes (Figure 4 and Supplementary Figure S6) observed in the Affymetrix experiments in these eight conditions. We confirmed, for example, the increased expression of *Dact1* under proliferating conditions in the presence of both MyoD and Six1/4 (Figure 4C), the increased expression of *Tnnc1* and *Tnnc2* under differentiating conditions only when both Six1/4 and MyoD are expressed, and the increased expression under both proliferating and differentiating conditions of *Murc* and *Lrrc17* when MyoD and Six1/4 are present. Then, for each of these 19 CRMs, we built a plasmid containing one associated CRM cloned upstream of a minimal promoter and the Luciferase gene as reporter. Recombined plasmids were transfected in wt or *SixdKO* MEFs and studied under the same eight conditions previously used to follow endogenous gene expression. Luciferase results are presented in Figure 4 and Supplementary Figure S6, where they are compared with endogenous gene expression. The identified *Dact1* CRM was synergistically activated by the presence of both Six1/4 and MyoD under proliferating conditions only, the *Mef2c*-CRM1, *Ttn*, *Srl*, *Tnnc1* and *Tnnc2* CRMs were activated synergistically by MyoD and Six1/4 only under differentiating conditions, while *Lrrc17*-CRM3 was activated synergistically by MyoD and Six1/4 under both proliferating and differentiating conditions. For some genes, the Six + MyoD CRMs that had been identified did not recapitulate the synergistic effect observed with the endogenous gene (*Cap2* and *Mybph* CRMs, Supplementary Figure S6) suggesting either that specific interactions between the CRM and its own minimal promoter are required, or that other CRMs are required. Altogether, 14/19 (74%) CRMs showed a synergistic activity in our cellular model (Figure 4 for positive examples and S6 for negative ones). Conversely, 10/12 genes (83%) have at least one associated synergistic CRM. Comparison of our data and the ChIP experiments with C2 myogenic cells using Six1 and Six4 antibodies published earlier (24,26), shows that among the 14 positive CRMs, 3 (21%) are ChIPed by Six4 (Mtap4-CRM, Lrrc17-CRM3, Srl-CRM), 6 (42%) by Six1 (Mef2c-CRM2, Tnnc1-CRM, Etv1-CRM2, Murc-CRM2, Lrrc17-CRM3, Srl-CRM), and that among the five negative CRMs, 3 (60%) are ChIPed by Six1 (Cap2-CRM, Etv1-CRM1, Mef2c-CRM), and 1 (20%) by Six4 (Cap2-CRM). From this, we can draw several conclusions. First, our bioinformatics approach is able to predict functional CRMs with a high positive success rate (74%). Second, only 50% of our 14 positive CRMs are predicted using either Six1 or Six4 ChIP data, showing that our method picks up new, previously unidentified, regulatory elements. Finally, the experimental Six1 and Six4 binding data do not discriminate between negative and positive CRMs since they predict respectively

60% and 50% of them. Taken together, this shows that our method is more predictive than simple physical binding.

Motif analysis of the synergistic CRMs

We have established that Six1/4 and MyoD act synergistically to reprogram MEFs toward a myogenic fate, and that CRM sequences bound by both transcription factors participate in this reprogramming by activation of neighbouring transcription units. Moreover, this activation is impaired when one of the two transcription factors is lacking. Since many other ubiquitous or muscle-specific transcription factors could participate with Six1/4 and MyoD in this reprogramming, we searched for functional binding for TFs in MEFs within the CRMs. We investigated which of these TFs were present in MEFs or in MEFs + MyoD and controlled the positive synergistic CRMs, and whether their expression was impaired in *SixdKO* MEFs. In order to do this we first bioinformatically explored the TF motif content of the 14 synergistic CRMs. We used both Imogene, a *de novo* motif generator which relies on phylogeny and inter-species motif conservation (29), and the available set of PWMs from Transfac and Jaspar databases (see Materials and Methods). The advantage of using a *de novo* approach is that it finds putative motifs that are not present in current databases. *De novo* motifs were ranked together with refined database motifs (see Materials and Methods), (31,45). The 15 top-ranked motifs are shown in Figure 5. While AP1, Meis1 and MEF2 are enriched in MyoD bound regions (30), the other motifs correspond to TFs expressed in muscle cells (EBF, NFAT) or to motifs with no known associated protein for which a regulatory role in myogenic cells remains to be demonstrated. To test their potential role in the synergy observed between MyoD and Six1/4, we selected two synergistic CRMs rich in these new motifs, the *Lrrc17*-CRM3 and *Tnnc2* CRM. We then mutated the predicted binding sites in these DNA sequences and measured mutant CRM activity in eight similar conditions (Figure 6, Supplementary Figure S7, and S2 showing the DNA sequence of these two CRMs).

Lrrc17-CRM3 is able to drive the expression of the Luciferase transgene in transfected MEFs. We observed an up-regulation of this CRM activity during proliferation in wt cells expressing MyoD compared to mutant cells expressing MyoD. This synergy was based on the presence of MEF3 and E-boxes, since their mutation impaired this activation, while mutation of the other motifs did not significantly impair it (Figure 6A, Supplementary Figure S7A). Lrrc17-CRM3 was further activated during the differentiation process in wt MEFs expressing MyoD, but not in mutant MEFs. Full activity of this CRM in wt MEFs expressing MyoD was dependent on the presence of mot7, mot14, MEF3 and E-boxes, since the mutation of any of these elements significantly decreased its activity. On the contrary, mutation of mot10, mot3, mot15 or mot5 did not lead to a statistically-significant decrease of the CRM activity, suggesting that the binding of transcription factors on these elements, if existing, was not *per se* important in these transient transfection assays. We concluded from this analysis that Six/MyoD synergy is based on the presence of both proteins binding to DNA and on co-occupancy of other transcription factors, the activity of which is mainly detected under differentiating conditions in wt MEFs expressing MyoD.

Contrary to the *Lrrc17*-CRM3, the *Tnnc2* CRM activity is mainly observed under differentiating conditions (Figure 6B), similarly to the expression of the endogenous gene (Figure 4C). Contrary to the combined mutations of the Lrrc17-CRM3, individual mutations of this CRM did not lead to a severe activation decrease except for the MEF2 mutation under differentiating conditions in MEFs expressing Six1/4 and MyoD. Indeed, mutation of MEF3, E-boxes or of the *de novo* motifs mot4 and mot5 did not greatly impair its activity (Figure 6B, Supplementary Figure S7B). It had been previously shown that Mef2 and MyoD proteins can interact, and that binding of Mef2 or MyoD to DNA is sufficient to recruit both proteins in an active transcription complex (13). Accordingly, mutation of both MEF2 sites and E-boxes led to a further reduction of CRM activity (Figure 6B). We conclude that in contrast to Lrrc17-CRM3, the activity of the *Tnnc2* CRM relies mainly on Mef2 activity, and that the synergy observed between MyoD and Six on this CRM requires a MEF2 element.

Newly identified DNA motifs within the CRMs are recognized by nuclear proteins that accumulate preferentially in MEFs expressing MyoD and Six1/4

To test whether the binding of nuclear proteins on the DNA motifs that we have identified could explain their activation effects (Figure 6), we prepared nuclear extracts from wt and *SixdKO* MEFs in the presence or not of MyoD under differentiating conditions. To obtain sufficient amounts of nuclear proteins (which was not possible after MEF electroporation) we produced stable wt and *SixdKO* MEFs expressing a Doxycycline inducible *Myod1* expression vector. Both in wt and *SixdKO* lines, MyoD efficiently accumulated in the nucleus 6 h after Doxycycline addition whereas *Myogenin* was activated in wt cells only (Supplementary Figure S8). We also observed that nuclear MyoD accumulation was higher in *SixdKO* MEFs than in wt MEFs (Supplementary Figure S8C, Figure 7B). In a first step, we validated the binding of Six1/4 proteins on the three MEF3 sites identified on these two CRMs (Figure 7A and Supplementary Figure S9A) and the binding of MyoD on the two E-boxes present in Lrrc17-CRM3 and on two of the four E-Boxes identified in Tnnc2 CRM (Figure 7B). Efficient MyoD binding was observed with nuclear extracts of *SixdKO* MEFs producing MyoD (Figure 7B), demonstrating that absence of reprogramming of *SixdKO* MEFs by MyoD is not the result of an impairment of MyoD to bind DNA as shown by GMSA experiments. It should be noted that the two MEF3 sites present in the *Lrrc17*-CRM3 have differential binding affinity for Six proteins, the *Lrrc17*-CRM3-1 presenting a higher affinity for the nuclear proteins accumulating in MEFs expressing Six1/4 and MyoD or for *in vitro* produced Six1 and Six4 proteins (Figure 7A, Supplementary Figure S9A).

The MEF2 sequence present in the *Tnnc2* CRM was recognized by the same nuclear proteins that bind to the *Desmin* MEF2 site (Figure 7C). We observed a preferential accumulation of Mef2 proteins in the nuclei of wt MEF + MyoD (Figure 7C). This higher accumulation correlated

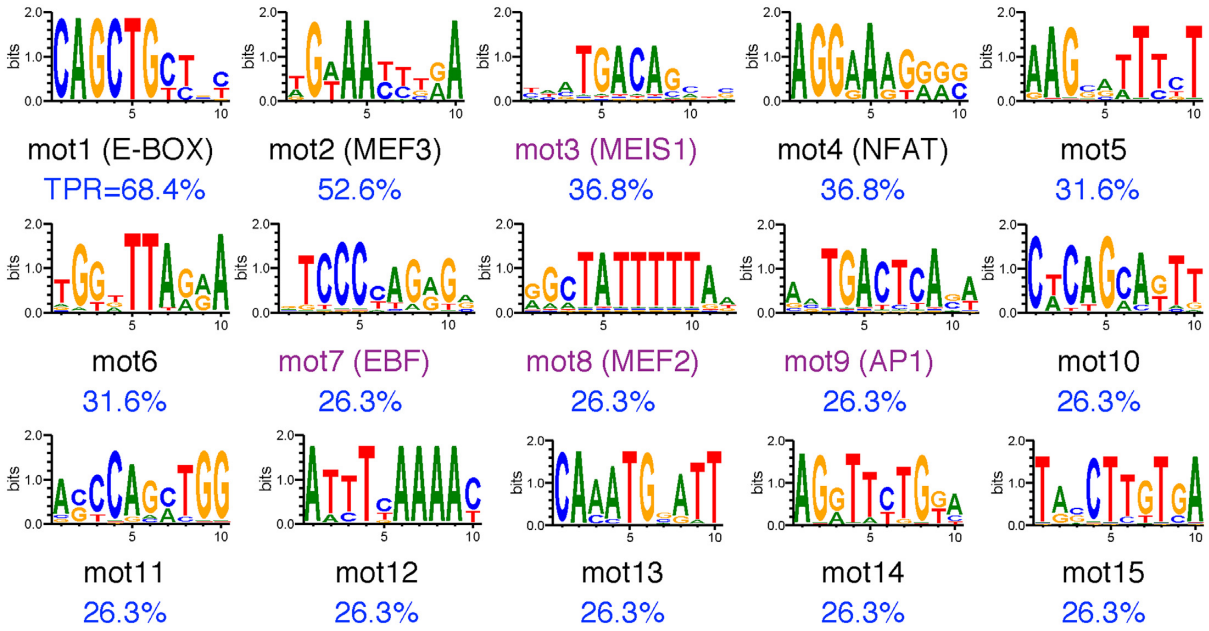

Figure 5. TF motif analysis. TF motifs found at a False Positive Rate (FPR) of 1%, with an enrichment of at least 25-fold (True Positive Rate or TPR > 25%), see Materials and Methods. The binding sites identified are conserved in several mammal species (see Materials and Methods). Motifs identified by refining known motifs from databases (corresponding TF indicated) are shown in purple.

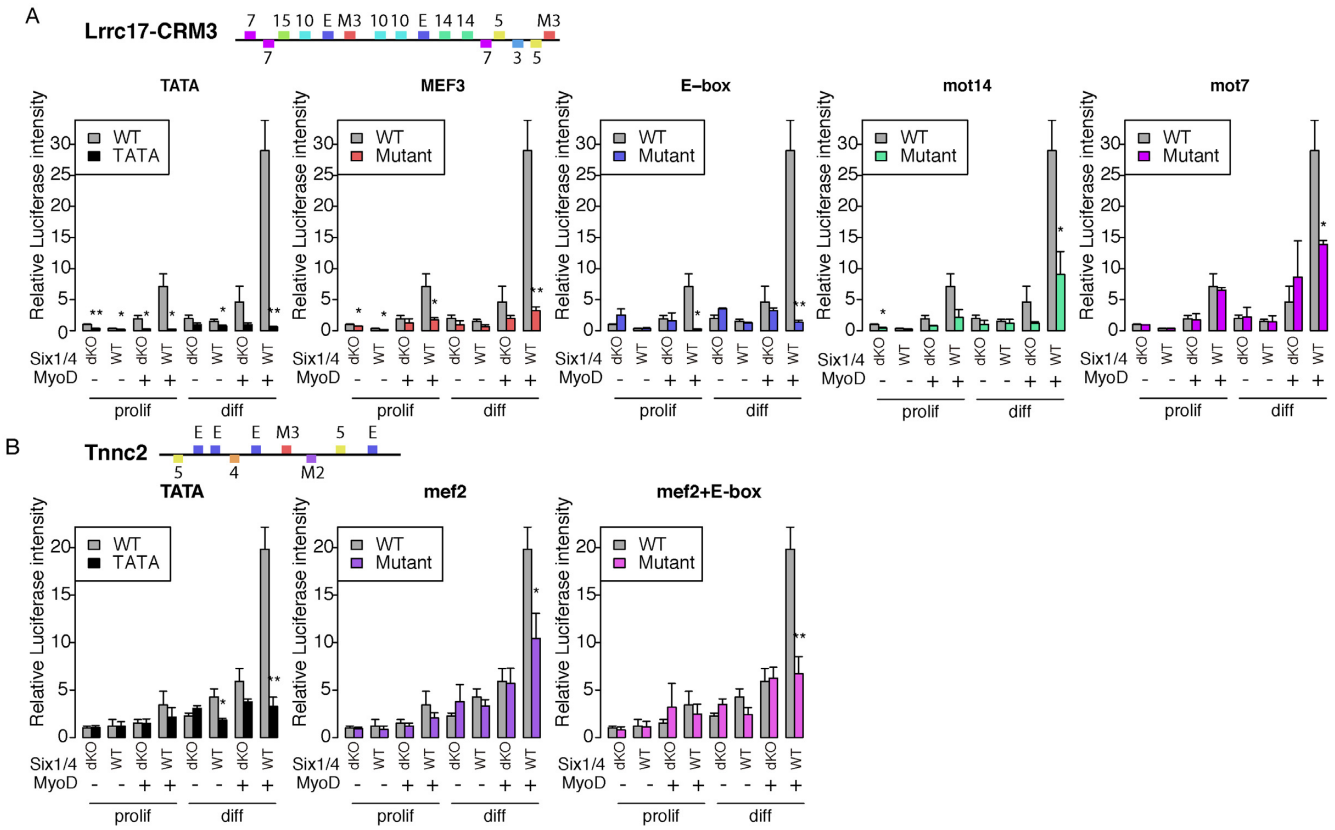

Figure 6. Luciferase activity of mutant CRMs. (**A** and **B**) *Lrrc17*-CRM3 and *Tnnc2* CRM were mutated in identified TF motifs, and mutant CRM Luciferase activity was measured under the 8 conditions after wt and *SixdKO* MEF transfection. The different TF motifs in the two CRMs are also schematically shown (upper boxes, sense, lower boxes, antisense motifs). Wt, mutant and TATA corresponds to the presence of the corresponding CRM (wt), mutant CRM (mutant) or empty vector with only the TATA box (TATA) in the Luciferase reporter. *$P < 0.05$, **$P < 0.01$; Student's *t*-tests.

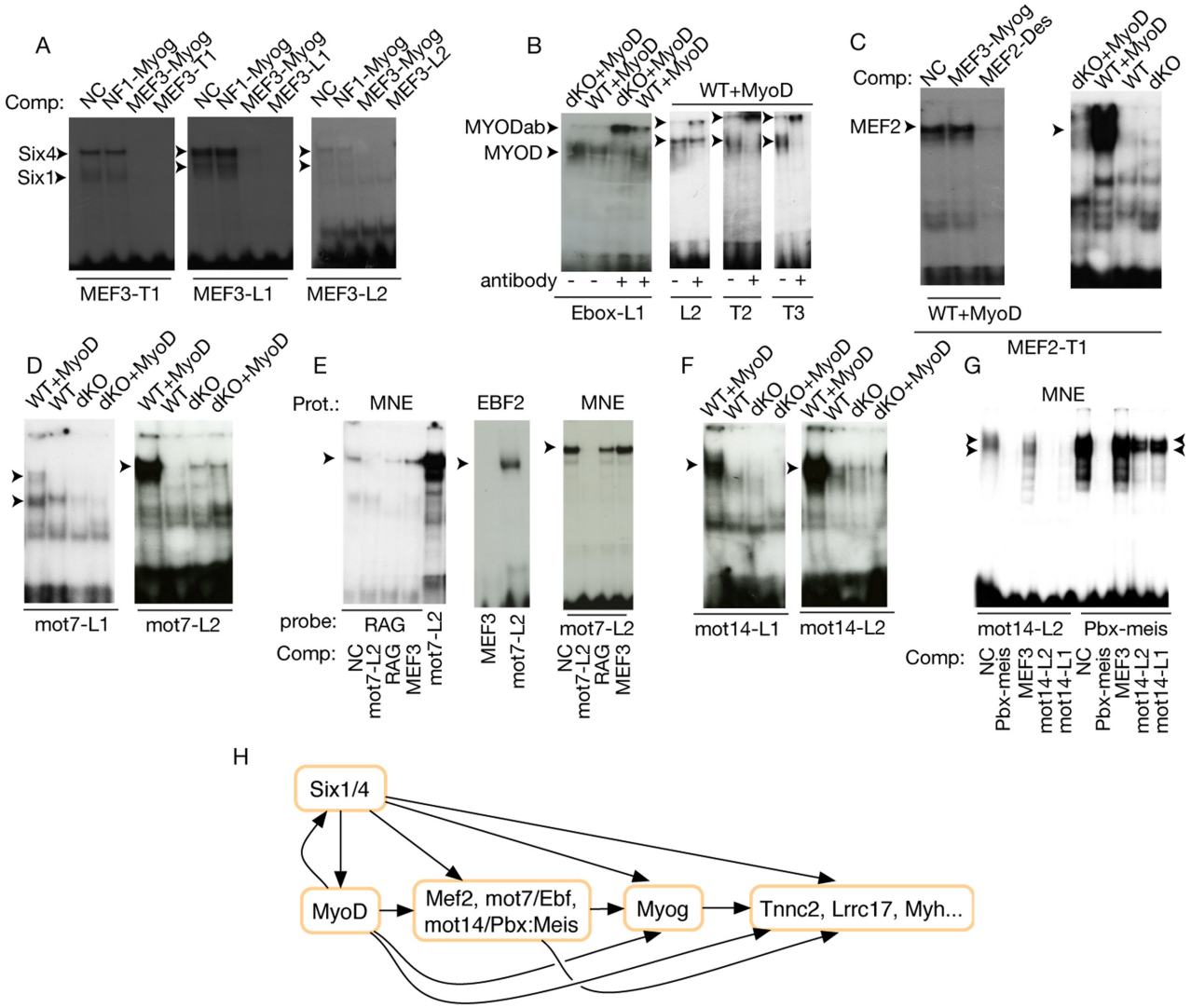

Figure 7. Identification of the nuclear proteins binding to the DNA motifs. (**A**–**G**) Gel mobility shift assays were performed with nuclear proteins extracted from wt, wt+MyoD, *dKO*, *dKO* + MyoD cells or with adult muscle nuclear extracts as indicated. (**A**) Wt nuclear extracts were used to detect MEF3 binding activity on the *Tnnc2* MEF3 (T1) and on the two MEF3 sites present in the *Lrrc17* CRM3 (L1 and L2). No competition (NC) or hundred fold molar excess of DNA competitor were used, with unrelated sequences (NF1-Myog), with the *Myogenin* MEF3 site (MEF3-Myog) or with the same sequence, showing that the three sequences bind Six1 and Six4 proteins. (**B**) E-boxes L1, L2, T2 and T3 efficiently bind MyoD produced in wt or *SixdKO* MEFs as demonstrated by the ability of MyoD antibodies (MYODab) to displace the DNA–protein complex formed in both wt+MyoD and *SixdKO*+MyoD MEFs. (**C**) The MEF2 motif present in the *Tnnc2* CRM is recognized by Mef2 proteins as demonstrated by the specific displacement of the DNA–protein complexes observed with a hundred fold molar excess of the *Desmin* MEF2 site but not with the *Myogenin* MEF3 site. Mef2 protein accumulation in *SixdKO* +MyoD, wt+MyoD, wt or *SixdKO* MEFs. (**D**) Mot7 DNA motifs (L1 and L2) are recognized by protein complexes that accumulate preferentially in the nuclei of wt + MyoD cells. (**E**) The EBF site of the *Rag1* promoter recognizes the same muscle nuclear proteins as mot7-L2, and a hundred fold molar excess of mot7-L2 disrupts EBF-RAG interactions. *In vitro* synthesized EBF2 binds mot7-L2 but not the *Myogenin* MEF3 site. The muscle nuclear protein complexes formed with mot7-L2 are competed specifically with a hundred fold molar excess of mot7-L2 and of the EBF site of the *Rag1* promoter but not with the *Myogenin* MEF3 site. (**F**) Mot14-L1 and mot14-L2 DNA motifs present in the Lrrc17-CRM3 are bound by proteins that accumulate preferentially in wt+MyoD MEFs. (**G**) Adult muscle nuclear proteins recognizing mot14-L2 and the *Myogenin* Pbx-Meis site have the same migration properties, and hundred fold molar excess of Pbx-Meis site competes the mot14-L2-protein complex formation. A hundred fold molar excess of mot14-L1 or of mot14-L2 but not of *Myogenin* MEF3 competes for Pbx-Meis protein complex formation. Note that nuclear extracts from wt +MyoD and *SixdKO* MEFs +MyoD were prepared 6 h after Doxycycline addition. MNE, adult muscle nuclear extracts. (**H**) Schematic summary of the interactions between the different genes studied.

with higher Mef2 transcriptional activity (Supplementary Figure S5), and with a major effect of the MEF2 mutation on the activity of the *Tnnc2* CRM, when associated with the E-box mutation only in wt MyoD expressing MEFs. This demonstrates that MyoD can induce nuclear Mef2 protein accumulation only in MEFs producing Six1/4 proteins, in agreement with the induction of *Mef2a* and *Mef2c* gene expression observed in wt MEFs by MyoD (Figure 3F).

The three mot7 sequences present in the *Lrrc17*-CRM3 are recognized by nuclear proteins that are enriched in wt MEFs + MyoD (Figure 7D, data not shown). Since these sequences are predicted to be EBF binding sites, we tested the ability of recombinant EBF2 proteins to bind to mot7-L2 by AlphaScreen assays which confirmed efficient specific binding (Supplementary Figure S10). We next validated these data by direct GMSA with recombinant EBF2 proteins on mot7-L2 (Figure 7E). In adult mouse myonuclei, proteins binding to mot7 motifs were also detected (Figure 7E); mot7-L2 has more affinity than mot7-L1 and mot7-L3 for these specific proteins (data not shown). Addition of several oligonucleotide competitors failed to disrupt the mot7-L2-protein complexes except for the addition of a known binding site for EBF proteins from the RAG1 promoter (46) (Figure 7E). GMSA with this known EBF binding site led to the same migrating DNA-protein complexes as those formed with mot7-L2, and was efficiently competed by an excess of mot7-L2 (Figure 7E). We conclude that mot7-L2 efficiently binds EBF proteins, and that the binding activity of these proteins is higher in cells expressing MyoD and Six than in cells devoid of Six or of MyoD.

We next showed that the two mot14 sites present on the Lrrc17-CRM3 are recognized by nuclear proteins accumulating preferentially in MyoD- producing wt MEFs and that mot14-L2 has more affinity for these proteins than mot14-L1 (Figure 7F). A very low level of mot14 binding proteins was observed in MyoD-producing *SixdKO* MEFs or in wt and mutant MEFs without MyoD. This showed that the presence of Six and MyoD in MEFs allows the accumulation of specific transcription factors involved in the control of shared downstream CRM sequences, reinforcing the myogenic potential of MyoD. Proteins binding to mot14 also accumulate in nuclear extracts from adult skeletal muscles (Figure 7G). Competition experiments with several oligonucleotides known to bind specific transcription factors failed to identify the nuclear proteins binding to these sites (Supplementary Figure S9B). Since we observed that *Meis1* mRNA expression was increased in wt MEFs expressing MyoD, we tested the ability of Meis1 proteins to bind mot14-L2 by AlphaScreen assays, but failed to detect Meis1 binding (Supplementary Figure S10). We observed that 9 out of 13 bases were conserved between the *Myogenin* Pbx-Meis motif (**TGATGTGCAGCAA**) and mot14-L2 (**TGAGGTTCTGTAA**). Based on the inability of Meis1 to bind mot14-L2, and the known role of the Pbx-Meis heterodimer in specific binding to the *Myogenin* promoter, we tested the ability of the *Myogenin* Pbx/Meis motif to compete binding of nuclear proteins on mot14 in gel mobility shift assays. We showed that the *Myogenin* Pbx/Meis motif was recognized by adult muscle nuclear extracts, that the DNA Pbx-Meis protein complexes were similar to the ones detected on mot14, and that *Myogenin* Pbx/Meis motif efficiently competed for the formation of mot14 protein complexes (Figure 7G). We further showed that mot14 also competed for *Myogenin* Pbx-Meis protein complexes formation (Figure 7G). We concluded that mot14 and the *Myogenin* Pbx/Meis motif were recognized by the same nuclear proteins, most probably composed of Pbx-Meis heterodimers. However, we cannot exclude that the Y-box protein Msy3, another protein binding to the *Myogenin* Pbx/Meis motif (47), does not participate in the DNA protein complexes observed on both mot14 and *Myogenin* Pbx/Meis sequences.

Overall, *Lrrc17*-CRM3, which is activated in both proliferating and differentiating myogenic cells, contains several DNA motifs, at least four of which (MEF3, E-box, mot7/EBF, mot14/Pbx-Meis) participate in its activity and are required for the full synergy between Six and MyoD.

Although no significant effect of the mutation of the mot4 motif was obtained in our Luciferase assay, the resemblance of this motif to the NFAT consensus sequence led us to test whether nuclear proteins were able to bind it in our assay. As shown on Supplementary Figure S9C, mot4 is recognized by nuclear proteins present in wt MEFs expressing MyoD and in adult skeletal muscles (not shown). Nevertheless, addition of a hundred-fold molar excess of the *Myogenin* NFAT binding site (48) does not compete for the binding, suggesting that transcription factors other than NFAT are responsible for this specific binding.

Altogether our experiments show that SixdKO MEFs expressing MyoD are unable to adopt a muscle fate and to activate MyoD targets. In other words, MyoD needs to cooperate with Six1/4 proteins to activate its downstream targets. This synergy leads to the initiation of MyoD downstream target gene activation, among which genes encoding Mef2, EBF, Pbx-Meis and other as yet unidentified transcription factors. These newly synthesized myogenic transcription factors act with Six and MyoD to fully activate hundreds of MyoD and Six/MyoD targets required to progress into the myogenic lineage. The Six/MyoD synergistic activation thus takes place in a feedforward way (49), recruiting these two families of transcription factors to their target sites along with the recruitment to adjacent sites of transcription factors such as Mef2, Pbx-Meis and EBF, encoded by genes that are themselves activated by Six and MyoD. (Figure 7H).

DISCUSSION

General strategy

In this work, we have combined gene expression, ChIPseq data and bioinformatic analyses to dissect the genome-wide transcriptional synergy between the TFs Six1 and Six4 and MyoD, during cell reprogramming.

Using an expression vector to control MyoD expression, we found that MyoD cannot transdifferentiate fibroblasts into muscle cells without the presence of Six1,4. At the gene expression level MyoD targets are not activated by MyoD in the absence of Six homeoproteins. This extends the previous results of Chakroun *et al.* who previously reported that regulation of MyoD targets depends on Six4 in C3H/10T1/2 cells expressing MyoD (26). We showed that the synergy between MyoD and Six is a widespread phenomenon, leading

under our experimental conditions to the activation of 761 genes.

By combining ChIPseq data and bioinformatics predictions, we found that MyoD and Six are co-localized on 1230 regulatory elements which tend to be in the vicinity of synergic genes. Among the 1230 CRMs, we extracted a stringent set of 96 'synergistic' CRMs meeting the conditions that (i) it binds MyoD, (ii) it contains a strong Six DNA binding site, conserved across several mammalian species, that is located in the vicinity of the site binding MyoD, (iii) the closest gene shows synergistic influence of these two TFs. We tested 19 such CRMs by Luciferase reporter assays and found that the majority (14/19 or 74%) led to a synergistic activation pattern similar to the nearest associated gene, therefore confirming the involvement of these CRMs in the direct regulation of their associated myogenic genes.

Finally, we investigated the role of other TF partners at the CRM level. We used a hybrid TF motif search method on the 14 experimentally validated CRMs and found 15 overrepresented motifs including the Six MEF3 and MyoD E-box binding sites, along with consensus binding sites for Mef2, Ebf, and new motifs binding yet unidentified TFs.

We note that the *de novo* motif finding algorithm, Imogene (29), had previously been experimentally validated in *Drosophila* (31,50), but had only been shown bioinformatically to have predictive power for mammals (29). Our findings that the predicted mot14 binding sites are recognized by nuclear proteins in Six + MyoD expressing MEFs and that their mutation decreased reporter CRM activity, provide this missing experimental test of Imogene.

Overall, our bioinformatic analysis extends previous attempts to identify combinations of transcription factors involved in myogenesis (51,52) by two approaches. First, we conducted a genome-wide study where CRMs can be localised distally to the nearest gene, instead of focusing on promoter regions as previously reported. Second, previous work used PWMs from publicly available databases while we generated *de novo* motifs with Imogene (29) and refined existing database motifs to increase their specificity. Overall our results validate the usefulness of such an unbiased method to dissect regulatory elements.

Mechanistic basis of the MyoD/Six synergy: three possible explanations

We bio-informatically identified 45 628 evolutionary conserved MEF3 binding sites in the mouse genome. ChIPseq performed with Six4 antibodies revealed 58 528 peaks in myogenic cells (26). We find that 1129 of our MEF3 binding sites are in a Six4 peak, compared to an expected 774 $\pm$ 28 for a 2.7 Gb genome (P = 1e-33, binomial test). Even though there is a significant enrichment, the two datasets are still largely distinct. Six proteins are expressed in kidney, thymus, neural crest cells, fibroblasts and neurons where they bind MEF3 sites that are not ChIPed in C2 myogenic cells, but that should be present in our identification of evolutionary conserved MEF3 binding sites. However, when this is restricted to the 1230 MEF3 sites associated with MyoD peaks (Figure 3C), the overlap between the two data sets becomes much more significant, with 311 MEF3 sites bound by Six4 as compared to a random expected number of 20 $\pm$ 4 (P = 1.2e−262). This is also the case for Six1: out of 2624 ChIPed regions, 157 fall in a MyoD+MEF3 peak, the random expected number being at most 8 $\pm$ 3 if we consider that all MEF3+MyoD CRMs are in the 17% of the genome covered by the ChIP-on-chip experiment. Together with the inability of MyoD to reprogram fibroblasts in the absence of Six proteins, the co-occurrence of MyoD and Six binding sites suggests that Six cooperate with MyoD on DNA to determine MyoD activity.

At least three non-exclusive mechanistic hypotheses may explain the role of Six in determining the properties of MyoD.

First, DNA binding of Six homeoproteins in MEFs may create specific permissive chromatin patterns required for MyoD binding.

Second, the simultaneous binding of transcription factors of these two families may be required to activate common downstream genes. According to this hypothesis, even when bound to its specific targets, MyoD would be unable to recruit the general transcription machinery and its target genes would remain unexpressed in the absence of Six proteins.

Third, Six proteins may allow recruitment of cofactors necessary for the reprogramming activity of MyoD.

We now discuss these three hypotheses in turn starting with the last two that were directly addressed in the present study.

Is the simultaneous binding of Six and MyoD required for MEF reprogramming?

We have shown that MyoD alone enters the nucleus of *SixdKO* MEFs and activates the expression of Luciferase reporters under the control of several CRMs (*Murc-CRM2, Dact1-CRM, LRRC17-CRM2, Cap2-CRM*, see Supplementary Figure S6) in these mutant cells. Moreover, nuclear extracts isolated from *SixdKO* MEFs expressing MyoD revealed MyoD binding activity in GMSA experiments. Therefore, MyoD proteins produced in *SixdKO* MEFs are able to bind episomal DNA, and suggest that the ability of MyoD to recruit the general transcription machinery is preserved in these mutant cells. However, we have shown that this is not sufficient to activate the downstream myogenic program in *SixdKO* cells (Figure 2, Supplementary Figure S4).

We have presented evidence that MyoD-binding regions with a conserved adjacent MEF3 site are associated with a synergistic transcriptional activation of the nearest gene. Moreover, experimental testing of 19 predicted CRMs showed a high success rate of 74%, including the stringent requirement that we observe a synergy between Six and MyoD at the CRM level (negative cases like *Cap2-CRM* show reporter activity, although not synergistically). Thus, simultaneous binding of Six and MyoD is required for MEF reprogramming and our approach is able to pick up functionally relevant regulatory elements rather than only physical binding.

We wondered whether this simultaneous binding of Six would be promoted by a certain type of E-box bound by MyoD. Indeed, two classes of MyoD binding sites have recently been characterized by genome-wide ChIP-seq exper-

iments. The first class comprises so called 'private' sites that MyoD binds specifically and that are not shared with other bHLH proteins, and the second class corresponds to binding sites 'shared' between MyoD and other bHLH factors such as Neurogenin. The nature of the genes activated by one or the other type of E-box is not known, and the kinetics of gene activation under their control remains to be determined (53). On the one hand, we observed that most of the E-boxes present in the characterized Six/MyoD synergic CRMs are of the shared CAGCTG type, in particular the MyoD bound E-boxes of *Lrrc17-CRM3* and *Tnnc2 CRM* (Supplementary Figure S11). On the other hand, most E-boxes present in the ChIPseq data (30) we used, are also of the shared type in a proportion not statistically significantly different ($P = 0.11$) from that obtained for the synergistic CRMs. We therefore cannot conclude at this stage that Six/MyoD synergy is promoted by one particular type of E-box, but the absence of muscle gene activation observed in *SixdKO* MEFs expressing MyoD suggests that Six is required for the activation of muscle genes associated with both types of E-box.

Do other transcription factors or cofactors enhance the synergy between six and MyoD?

Enrichment for MEF3, mot7/EBF and mot14/Pbx-Meis motifs around MyoD peaks had not been identified in previous studies (30,52,54). More recently, Chakroun *et al.* identified, by genome-wide ChIPseq experiments, the presence of Six4 in 36% of the peaks bound by MyoD in C2C12 myotubes, and showed that many CRMs with MEF3 and E-boxes present in muscle genes are cooperatively activated by MyoD and Six4 in the presence of a MEF3 element (26). Our analysis of DNA motif conservation between the 14 synergic CRMs identified DNA motifs that are present in many of them, among which the MEF2 motif, the EBF motif mot7, and the novel mot14 motif. Mot7 sites were assigned to EBF TF by Transfac and Jaspar databases. The mot14 motif does not resemble any characterized DNA motif present in Jaspar, Transfac, HTSELEX or Uniprobe data banks, but we provide evidence here that it is recognized by the same proteins that bind to the *Myogenin* Pbx-Meis motif. Nuclear proteins present in adult skeletal muscles and in wt MEFs expressing MyoD bind these two motifs. Furthermore we show that mot7/EBF and mot14/Pbx-Meis binding proteins, as well as Mef2, are faintly detected in wt MEFs, in *Six1/4* mutant MEFs or in *Six1/4* mutant MEFs expressing MyoD. We demonstrate that the amount of the corresponding nuclear proteins is strongly increased in wt MEF in the first 6 h following MyoD induction, and that it may participate in MEF reprogramming. In keeping with this, Mef2 transcriptional activity is higher in these cells as compared with MEFs expressing neither MyoD nor Six or in cells expressing only MyoD. Furthermore, mutation of the mot7/EBF, mot14/Pbx-Meis, or MEF2 motifs present in *Tnnc2* or *Lrrc17* CRMs strongly reduced their activity, showing that binding of the corresponding proteins to these CRMs participates in the Six/MyoD synergy observed during MEF reprogramming. This demonstrates that the co-occupancy of CRMs by Six, MyoD and by transcription factors - themselves activated by Six and MyoD in a feedforward manner—is required for their activity. We attempted to rescue the inability of mutant MEFs to be reprogrammed by MyoD by co-expressing Mef2 and MyoD expression vectors, but failed to detect any significant increase of *Myogenin* expression. This means that Mef2 proteins cannot synergize with MyoD to reprogram MEF in the absence of Six1 or Six4 proteins. This last result is consistent with the finding in transgenic mice that the MEF3 mutant *Myogenin* promoter is inactive in embryonic myogenic cells. The MRF, Pbx, NFI, NFAT and Mef2 proteins which are known to interact with the *Myogenin* promoter (17,48,55), are expressed in these cells but they are unable to activate the *Myogenin* promoter *in vivo* without the help of Six proteins.

We show that LRRC17-CRM3 activity is decreased if its mot7/EBF binding site is mutated. EBF zinc finger proteins are expressed in myogenic cells in *Ciona* (56) in *Drosophila* (57) and also in embryonic and adult vertebrates (58,59). In *Drosophila*, Collier/EBF proteins expressed in muscle progenitor cells are required for the specification of muscle identity, and control the expression of *Eya*, a Six1/4 cofactor, revealing a new parallel between the myogenic regulatory networks operating in *Drosophila* and vertebrates (27,57). In *Ciona*, EBF controls pharyngeal muscle founder cells and triggers MRF associated differentiation (56). EBF proteins have also been shown to synergize with MyoD in mouse myogenic cells to activate *Atp2a1* gene expression (59). It is interesting to note that the *Atp2a1* gene is also under the direct control of Six proteins and that in embryos and adult myofibers of compound *Six* mutant mice, *Atp2a1* is severely downregulated (21,40,60). Whether EBF proteins are still expressed in axial and branchial arches myogenic progenitors of compound *Six* mutant embryos remain to be determined. We show here an induction of *Ebf2* mRNA expression in wt MEFs expressing MyoD and show that EBF binding to regulatory regions of genes expressed in muscle cells is required for the activation of a subset of muscle genes. In the present study we identified such EBF binding sites in *Etv1-CRM2*, *Lrrc17-CRM1*, *Lrrc17-CRM3*, *Mtap4*, *Srl* and *Tnnc1* CRMs and propose that all these genes are under the control of a common *Six/MyoD/EBF* pathway.

We show that LRRC17-CRM3 activity is decreased if its mot14/Pbx-Meis binding site is mutated. In vertebrates, the TALE homeoprotein family includes four *Pbx*, three *Meis* and two *Prep* genes. Pbx proteins can associate with Hox, Meis and Prep proteins to bind DNA. The Pbx/Meis heterodimer interacts with the *Myogenin* promoter and allows further MyoD binding to this promoter (61). Direct involvement of Pbx/Meis proteins during embryonic myogenesis in mammals has not yet been addressed. In Zebrafish embryos Pbx2 and Pbx4 control a subset of MyoD targets (62). It has been established that Six1/4 are bound on *Pbx1*, *Pbx2*, *Meis1* and *Meis2* regulatory sequences in myogenic cells (24) and may thus participate in their activation. We show here that *Meis1* is activated in wt MEFs expressing MyoD, suggesting that the increase of mot14 binding proteins observed during wt MEF reprogramming by MyoD may depend on *Meis1* induction. This could allow Pbx-Meis heterodimer formation and specific target recognition. Alternatively, we cannot exclude the presence of MSY3 protein in the complexes bound to the *Myogenin* Pbx/Meis or mot14

motifs, since MSY3 can bind the *Myogenin* Pbx-Meis motif (47). We identified in the present study Pbx-Meis/MSY3 binding sites on *Lrrc17-CRM1*, *Lrrc17-CRM2*, *Lrrc17-CRM3*, *Mybph*, *Srl, Myogenin* and *Mef2c-CRM3* and propose that all these genes are under the control of a common *Six/MyoD/Pbx-Meis/Msy3* pathway.

Chakroun et al (26) observed that the synergy between MyoD and Six4 in HEK 293T cells transfected with expression vectors for *Six4* and *Myod1* was strictly dependent on the presence of an intact MEF3 site in their CRM-Luciferase reporter assays. In contrast, we observed that the strong synergy between Six and MyoD during MEF reprogramming relies not only on their binding site occupancy but also on other surrounding sequences (Mef2, mot7, mot14). This discrepancy could be due to a different genetic cascade operating in HEK293T cells overexpressing Six4 and MyoD from that in MEFs overexpressing MyoD (Figure 7H). The precise role of EBF, Mef2 and Pbx-Meis/MSY3 proteins during MEF reprogramming by MyoD remains to be established through analysis of the consequences of their respective knockdown on endogenous muscle gene expression.

Do Six proteins allow for the formation of a permissive chromatin state for MyoD binding?

MyoD can bind nucleosomal DNA (63), suggesting that even in compacted chromatin it may reach its targets. Alternatively 'beacon transcription factors', like Pbx/Meis (61), may be bound to nucleosomal DNA prior to MyoD, allowing its binding and subsequent chromatin remodeling, as suggested at the *Myogenin* promoter. The ability of a *Myogenin* CRM mutated in the Pbx/Meis motif to drive correctly *LacZ* transgene expression in transient transgenic embryos (47) suggests that Pbx/Meis proteins themselves are not absolutely necessary to open chromatin at the *Myogenin* locus. We show here the complete absence of any muscle gene activation in *Six* mutant MEFs expressing MyoD. This suggests that Six proteins may act as one type of obligatory 'beacon' transcription factor required for E-box dependent genes to be activated (10). This hypothesis is supported by the absence of activity in somitic and limb muscle cells of transgenic embryos with a MEF3 mutant *Myogenin* promoter (17).

Whether Six protein binding to DNA is required to open chromatin at specific loci was not directly addressed in the present study and remains to be tested. ChIPseq with MyoD antibodies in wt and *Six1/4* mutant MEFs would be hepful to unravel genome occupancy, and provide information on how Six proteins modulate binding of MyoD to its thousands of genomic target regions (30,64).

The hypothesis that Six binding can direct MyoD to occupy specific loci is nonetheless interesting. It would explain the requirement of Six upstream of bHLH MRF determination genes in most myogenic territories during embryogenesis, and later on the requirement for Six to act synergistically with bHLH MRFs in the same pathway. Transgenic analyses of several MEF3 readouts have established the robustness of Six transcriptional activity in the myogenic lineage from early stages of embryogenesis (27) to the adult muscle fiber itself (65,66). Furthermore, transgenic analyses of several wt and MEF3 mutant CRMs over the years has shown that the presence of Six1/4 binding sites is crucial for their proper activity. Examples include the hypaxial *Myf5* enhancer (67), *Myod1* enhancers (19), and the *Myogenin* promoter (17) in the embryo, and *Aldolase A* muscle promoter expression in the adult myofiber (65,68). Mutation of the MEF3 sites in the CRMs mentioned impaired *LacZ* or *CAT* reporter transgene expression. This shows that the complete muscle specific transcription factor machinery is unable to activate the above mentioned MEF3 mutant transgenes, most probably because it is unable to bind the mutant promoters. Such an important role of Six proteins in chromatin opening may also explain the apparently contradictory results showing that in some cases Six and MyoD are absolutely required to activate their endogenous targets, while in other cases, they are dispensable (i.e. Tnnc2 CRM) to activate isolated CRM in transient transfection assays. This also may explain why we were unable to activate *Myogenin* expression in *SixdKO* MEFs after forced Mef2c and MyoD expression. This leads us to hypothesize that Six homeoproteins need to be bound to DNA to allow chromatin opening, further transcription factor binding and chromatin landscape reorganization. In agreement with this hypothesis, Six proteins associate with several epigenetic protein chromatin remodelers among which Brg1 (69), CBP (70) and UTX (26,28) that may participate in Six dependent chromatin dynamics.

Six proteins can also activate striated muscle genes in the absence of MRFs as observed in cardiac cells deprived of *Ezh2* (71), suggesting that Six proteins may function with other bHLH proteins expressed in cardiomyocytes to activate direct MyoD targets like *Tnnt3*, *Mylpf*, *Ryr1* or *Actininα3*, although this remains to be established. Six proteins are also required to activate neuronal bHLH determination genes in *Drosophila* (72) and in mammals through the recruitment of SWI/SNF associated proteins (69,73). Interestingly, forced expression of Six1 in cooperation with Eya1, Brg1 and Baf170 is able to reprogram 3T3 fibroblasts into the neuronal lineage, and co-expression of Six1 and Eya1 is sufficient to activate the proneuronal bHLH factor *Neurog1* and later on *Neurod1* during otic neurogenesis (69). One can thus suspect a more general role of Six homeoproteins in establishing competence for specific cell fates in pluripotent embryonic cells, and in maintaining competence by modulating the genomic accessibility of master determination transcription factors specific to these fates. In the muscle lineage, Six homeoproteins are not only expressed in the commitment phase, but remain expressed and are required throughout muscle development and in the adult myofiber (40,66) and its associated myogenic stem cells (23,25–26).

In conclusion, we show that Six proteins play a central role in establishing the ability of MyoD to activate myogenesis. Six homeoproteins participate with MyoD in the activation of a genetic cascade leading to increased expression of EBF, Mef2 and Pbx-Meis/MSY3, that allows MyoD to initiate a sequence of molecular events leading to cell reprogramming by switching on a battery of downstream common target genes associated with acquisition of skeletal muscle fate (12).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the sequencing and genomic platform of Institute Cochin for microarray experiments and F. Dumont and S. Jacques for advice. We thank Dr E.Olson for the gift of the poly-MEF2-Luciferase and of the Mef2C expression plasmids, Dr P.Seale for the gift of EBF2 expression plasmid and Dr S. Tapscott for the MyoD-E12 plasmid. We thank Dr A. Sotiropoulos, Dr S. Gautron and specially Dr G. Butler-Brown and Dr M.Buckingham for critical reading of the manuscript.

FUNDING

University of Paris VII and by the 'Association Française contre les Myopathies' (AFM to M.S.); Agence National de la Recherche (ANR to I.S.); The Uehara Memorial Foundation and JSPS Postdoctoral Fellowships for Research Abroad; labex 'Who am I?' (to M.G.); AFM; Institut National de la Santé et de la Recherche Médicale (INSERM); Centre National de la Recherche Scientifique (CNRS) for the PEPS ≪Bio-Maths-Info≫; ANR [RPV09108KKA]; Idex 'Sorbonne Paris Cité'; AFM [17167 to P.A.D.]; labex 'Who am I?', ANR-11-LABX-0071 under program ANR-11-IDEX-0005-01. Funding for open access charge: INSERM.
Conflict of interest statement. None declared.

REFERENCES

1. Davis,R.L., Weintraub,H. and Lassar,A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, **51**, 987–1000.
2. Buckingham,M. and Rigby,P.W. (2014) Gene regulatory networks and transcriptional mechanisms that control myogenesis. *Dev. Cell*, **28**, 225–238.
3. Schafer,B.W., Blakely,B.T., Darlington,G.J. and Blau,H.M. (1990) Effect of cell history on response to helix-loop-helix family of myogenic regulators. *Nature*, **344**, 454–458.
4. Yamanaka,S. (2009) A fresh look at iPS cells. *Cell*, **137**, 13–17.
5. Soufi,A., Garcia,M.F., Jaroszewicz,A., Osman,N., Pellegrini,M. and Zaret,K.S. (2015) Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, **161**, 555–568.
6. Wapinski,O.L., Vierbuchen,T., Qu,K., Lee,Q.Y., Chanda,S., Fuentes,D.R., Giresi,P.G., Ng,Y.H., Marro,S., Neff,N.F. *et al.* (2013) Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell*, **155**, 621–635.
7. Chen,J., Zhang,Z., Li,L., Chen,B.C., Revyakin,A., Hajj,B., Legant,W., Dahan,M., Lionnet,T., Betzig,E. *et al.* (2014) Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, **156**, 1274–1285.
8. Weintraub,H., Tapscott,S.J., Davis,R.L., Thayer,M.J., Adam,M.A., Lassar,A.B. and Miller,A.D. (1989) Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 5434–5438.
9. Delfini,M.C. and Duprez,D. (2004) Ectopic Myf5 or MyoD prevents the neuronal differentiation program in addition to inducing skeletal muscle differentiation, in the chick neural tube. *Development*, **131**, 713–723.
10. Fong,A.P., Yao,Z., Zhong,J.W., Cao,Y., Ruzzo,W.L., Gentleman,R.C. and Tapscott,S.J. (2012) Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev. Cell*, **22**, 721–735.
11. Albini,S., Coutinho,P., Malecova,B., Giordani,L., Savchenko,A., Forcales,S.V. and Puri,P.L. (2013) Epigenetic reprogramming of human embryonic stem cells into skeletal muscle cells and generation of contractile myospheres. *Cell Rep.*, **3**, 661–670.
12. Fong,A.P. and Tapscott,S.J. (2013) Skeletal muscle programming and re-programming. *Curr. Opin. Genet. Dev.*, **23**, 568–573.
13. Molkentin,J.D., Black,B.L., Martin,J.F. and Olson,E.N. (1995) Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell*, **83**, 1125–1136.
14. Penn,B.H., Bergstrom,D.A., Dilworth,F.J., Bengal,E. and Tapscott,S.J. (2004) A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev.*, **18**, 2348–2353.
15. Cheng,T.C., Wallace,M.C., Merlie,J.P. and Olson,E.N. (1993) Separable regulatory elements governing myogenin transcription in mouse embryogenesis. *Science*, **261**, 215–218.
16. Yee,S.P. and Rigby,P.W. (1993) The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.*, **7**, 1277–1289.
17. Spitz,F., Demignon,J., Porteu,A., Kahn,A., Concordet,J.P., Daegelen,D. and Maire,P. (1998) Expression of myogenin during embryogenesis is controlled by Six/sine oculis homeoproteins through a conserved MEF3 binding site. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 14220–14225.
18. Grifone,R., Demignon,J., Houbron,C., Souil,E., Niro,C., Seller,M., Hamard,G. and Maire,P. (2005) Six1 and Six4 homeoproteins are required for Pax3 and MRF expression during myogenesis in the mouse embryo. *Development*, **132**, 2235–2249.
19. Relaix,F., Demignon,J., Laclef,C., Pujol,J., Santolini,M., Niro,C., Lagha,M., Rocancourt,D., Buckingham,M. and Maire,P. (2013) Six homeoproteins directly activate Myod expression in the gene regulatory networks that control early myogenesis. *PLoS Genet.*, **9**, e1003425.
20. Parmacek,M.S., Ip,H.S., Jung,F., Shen,T., Martin,J.F., Vora,A.J., Olson,E.N. and Leiden,J.M. (1994) A novel myogenic regulatory circuit controls slow/cardiac troponin C gene transcription in skeletal muscle. *Mol. Cell. Biol.*, **14**, 1870–1885.
21. Niro,C., Demignon,J., Vincent,S., Liu,Y., Giordani,J., Sgarioto,N., Favier,M., Guillet-Deniau,I., Blais,A. and Maire,P. (2010) Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. *Dev. Biol.*, **338**, 168–182.
22. Sato,S., Nakamura,M., Cho,D.H., Tapscott,S.J., Ozaki,H. and Kawakami,K. (2002) Identification of transcriptional targets for Six5: implication for the pathogenesis of myotonic dystrophy type 1. *Hum. Mol. Genet.*, **11**, 1045–1058.
23. Yajima,H., Motohashi,N., Ono,Y., Sato,S., Ikeda,K., Masuda,S., Yada,E., Kanesaki,H., Miyagoe-Suzuki,Y., Takeda,S. *et al.* (2010) Six family genes control the proliferation and differentiation of muscle satellite cells. *Exp. Cell Res.*, **316**, 2932–2944.
24. Liu,Y., Chu,A., Chakroun,I., Islam,U. and Blais,A. (2010) Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. *Nucleic Acids Res.*, **38**, 6857–6871.
25. Le Grand,F., Grifone,R., Mourikis,P., Houbron,C., Gigaud,C., Pujol,J., Maillet,M., Pagès,G., Rudnicki,M., Tajbakhsh,S. *et al.* (2012) Six1 regulates stem cell repair potential and self-renewal during skeletal muscle regeneration. *J. Cell Biol.*, **198**, 815–832.
26. Chakroun,I., Yang,D., Girgis,J., Gunasekharan,A., Phenix,H., Kaern,M. and Blais,A. (2015) Genome-wide association between Six4, MyoD and the histone demethylase Utx during myogenesis. *Faseb J.*, **29**, 4738–4755.
27. Grifone,R., Demignon,J., Giordani,J., Niro,C., Souil,E., Bertin,F., Laclef,C., Xu,P.X. and Maire,P. (2007) Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. *Dev. Biol.*, **302**, 602–616.
28. Seenundun,S., Rampalli,S., Liu,Q.C., Aziz,A., Palii,C., Hong,S., Blais,A., Brand,M., Ge,K. and Dilworth,F.J. (2010) UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. *EMBO J.*, **29**, 1401–1411.
29. Rouault,H., Santolini,M., Schweisguth,F. and Hakim,V. (2014) Imogene: identification of motifs and cis-regulatory modules underlying gene co-regulation. *Nucleic Acids Res.*, **42**, 6128–6145.
30. Cao,Y., Yao,Z., Sarkar,D., Lawrence,M., Sanchez,G., Parker,M., MacQuarrie,K., Davison,J., Morgan,M., Ruzzo,W. *et al.* (2010) Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell.* , **18**, 662–674.

31. Rouault,H., Mazouni,K., Couturier,L., Hakim,V. and Schweisguth,F. (2010) Genome-wide identification of cis-regulatory motifs and modules underlying gene coregulation using statistics and phylogeny. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 14615–14620.
32. Portales-Casamar,E., Thongjuea,S., Kwon,A.T., Arenillas,D., Zhao,X., Valen,E., Yusuf,D., Lenhard,B., Wasserman,W.W. and Sandelin,A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **38**, D105–D110.
33. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
34. Chaouch,S., Mouly,V., Goyenvalle,A., Vulin,A., Mamchaoui,K., Negroni,E., Di Santo,J., Butler-Browne,G., Torrente,Y., Garcia,L. *et al.* (2009) Immortalized skin fibroblasts expressing conditional MyoD as a renewable and reliable source of converted human muscle cells to assess therapeutic strategies for muscular dystrophies: validation of an exon-skipping approach to restore dystrophin in Duchenne muscular dystrophy cells. *Hum. Gene Ther.*, **20**, 784–790.
35. Liberzon,A., Subramanian,A., Pinchback,R., Thorvaldsdottir,H., Tamayo,P. and Mesirov,J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.
36. Kelder,T., van Iersel,M.P., Hanspers,K., Kutmon,M., Conklin,B.R., Evelo,C.T. and Pico,A.R. (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.*, **40**, D1301–D1307.
37. Salminen,M., Spitz,F., Fiszman,M.Y., Demignon,J., Kahn,A., Daegelen,D. and Maire,P. (1995) Myotube-specific activity of the human aldolase A M-promoter requires an overlapping binding site for NF1 and MEF2 factors in addition to a binding site (M1) for unknown proteins. *J. Mol. Biol.*, **253**, 17–31.
38. Yang,Z., MacQuarrie,K.L., Analau,E., Tyler,A.E., Dilworth,F.J., Cao,Y., Diede,S.J. and Tapscott,S.J. (2009) MyoD and E-protein heterodimers switch rhabdomyosarcoma cells from an arrested myoblast phase to a differentiated state. *Genes Dev.*, **23**, 694–707.
39. Stormo,G.D. (2010) Motif discovery using expectation maximization and Gibbs' sampling. *Methods Mol. Biol.*, **674**, 85–95.
40. Sakakibara,I., Santolini,M., Ferry,A., Hakim,V. and Maire,P. (2014) Six homeoproteins and a Iinc-RNA at the fast MYH locus lock fast myofiber terminal phenotype. *PLoS Genet.*, **10**, e1004386.
41. Pennacchio,L.A., Ahituv,N., Moses,A.M., Prabhakar,S., Nobrega,M.A., Shoukry,M., Minovitsky,S., Dubchak,I., Holt,A., Lewis,K.D. *et al.* (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, **444**, 499–502.
42. Yang,X. and Cheyette,B.N. (2013) SEC14 and spectrin domains 1 (Sestd1) and Dapper antagonist of catenin 1 (Dact1) scaffold proteins cooperatively regulate the Van Gogh-like 2 (Vangl2) four-pass transmembrane protein and planar cell polarity (PCP) pathway during embryonic development in mice. *J. Biol. Chem.*, **288**, 20111–20120.
43. Le Grand,F., Jones,A.E., Seale,V., Scime,A. and Rudnicki,M.A. (2009) Wnt7a activates the planar cell polarity pathway to drive the symmetric expansion of satellite stem cells. *Cell Stem Cell*, **4**, 535–547.
44. Chal,J., Oginuma,M., Al Tanoury,Z., Gobert,B., Sumara,O., Bousson,F., Zidouni,Y., Mursch,C., Moncuquet,P. *et al.* (2015) Differentiation of pluripotent stem cells to muscle fiber to model Duchenne muscular dystrophy. *Nat. Biotechnol.*, **33**, 962–969.
45. Santolini,M., Mora,T. and Hakim,V. (2014) A general pairwise interaction model provides an accurate description of in vivo transcription factor binding sites. *PLoS One*, **9**, e99015.
46. Zandi,S., Mansson,R., Tsapogas,P., Zetterblad,J., Bryder,D. and Sigvardsson,M. (2008) EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors. *J. Immunol.*, **181**, 3364–3372.
47. Berghella,L., De Angelis,L., De Buysscher,T., Mortazavi,A., Biressi,S., Forcales,S.V., Sirabella,D., Cossu,G. and Wold,B.J. (2008) A highly conserved molecular switch binds MSY-3 to regulate myogenin repression in postnatal muscle. *Genes Dev.*, **22**, 2125–2138.
48. Armand,A.S., Bourajjaj,M., Martinez-Martinez,S., el Azzouzi,H., da Costa Martins,P.A., Hatzis,P., Seidler,T., Redondo,J.M. and De Windt,L.J. (2008) Cooperative synergy between NFAT and MyoD regulates myogenin expression and myogenesis. *J. Biol. Chem.*, **283**, 29004–29010.
49. Milo,R., Itzkovitz,S., Kashtan,N., Levitt,R., Shen-Orr,S., Ayzenshtat,I., Sheffer,M. and Alon,U. (2004) Superfamilies of evolved and designed networks. *Science*, **303**, 1538–1542.
50. Menoret,D., Santolini,M., Fernandes,I., Spokony,R., Zanet,J., Gonzalez,I., Latapie,Y., Ferrer,P., Rouault,H., White,K.P. *et al.* (2013) Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome Biol.*, **14**, R86.
51. Wasserman,W.W. and Fickett,J.W. (1998) Identification of regulatory regions which confer muscle-specific gene expression. *J. Mol. Biol.*, **278**, 167–181.
52. Blais,A., Tsikitis,M., Acosta-Alvear,D., Sharan,R., Kluger,Y. and Dynlacht,B.D. (2005) An initial blueprint for myogenic differentiation. *Genes Dev.*, **19**, 553–569.
53. Bergstrom,D.A., Penn,B.H., Strand,A., Perry,R.L., Rudnicki,M.A. and Tapscott,S.J. (2002) Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. *Mol. Cell*, **9**, 587–600.
54. Blum,R., Vethantham,V., Bowman,C., Rudnicki,M. and Dynlacht,B.D. (2012) Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. *Genes Dev.*, **26**, 2763–2779.
55. Faralli,H. and Dilworth,F.J. (2012) Turning on myogenin in muscle: a paradigm for understanding mechanisms of tissue-specific gene expression. *Comp. Funct. Genomics*, 836374.
56. Razy-Krajka,F., Lam,K., Wang,W., Stolfi,A., Joly,M., Bonneau,R. and Christiaen,L. (2014) Collier/OLF/EBF-dependent transcriptional dynamics control pharyngeal muscle specification from primed cardiopharyngeal progenitors. *Dev. Cell*, **29**, 263–276.
57. de Taffin,M., Carrier,Y., Dubois,L., Bataille,L., Painset,A., Le Gras,S., Jost,B., Crozatier,M. and Vincent,A. (2015) Genome-Wide Mapping of Collier In Vivo Binding Sites Highlights Its Hierarchical Position in Different Transcription Regulatory Networks. *PLoS One*, **10**, e0133387.
58. Green,Y.S. and Vetter,M.L. (2011) EBF proteins participate in transcriptional regulation of Xenopus muscle development. *Dev. Biol.*, **358**, 240–250.
59. Jin,S., Kim,J., Willert,T., Klein-Rodewald,T., Garcia-Dominguez,M., Mosqueira,M., Fink,R., Esposito,I., Hofbauer,L.C., Charnay,P. *et al.* (2014) Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. *Nat. Commun.*, **5**, 3793.
60. Richard,A.F., Demignon,J., Sakakibara,I., Pujol,J., Favier,M., Strochlic,L., Le Grand,F., Sgarioto,N., Guernec,A., Schmitt,A. *et al.* (2011) Genesis of muscle fiber-type diversity during mouse embryogenesis relies on Six1 and Six4 gene expression. *Dev. Biol,*. **359**, 303–320.
61. Berkes,C.A., Bergstrom,D.A., Penn,B.H., Seaver,K.J., Knoepfler,P.S. and Tapscott,S.J. (2004) Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol. Cell*, **14**, 465–477.
62. Maves,L., Waskiewicz,A.J., Paul,B., Cao,Y., Tyler,A., Moens,C.B. and Tapscott,S.J. (2007) Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development*, **134**, 3371–3382.
63. Gerber,A.N., Klesert,T.R., Bergstrom,D.A. and Tapscott,S.J. (1997) Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev.*, **11**, 436–450.
64. Soleimani,V.D., Yin,H., Jahani-Asl,A., Ming,H., Kockx,C.E., van Ijcken,W.F., Grosveld,F. and Rudnicki,M.A. (2012) Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. *Mol. Cell*, **47**, 457–468.
65. Spitz,F., Salminen,M., Demignon,J., Kahn,A., Daegelen,D. and Maire,P. (1997) A combination of MEF3 and NFI proteins activates transcription in a subset of fast-twitch muscles. *Mol. Cell. Biol.*, **17**, 656–666.
66. Grifone,R., Laclef,C., Spitz,F., Lopez,S., Demignon,J., Guidotti,J.E., Kawakami,K., Xu,P.X., Kelly,R., Petrof,B.J. *et al.* (2004) Six1 and Eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype. *Mol. Cell. Biol.*, **24**, 6253–6267.
67. Giordani,J., Bajard,L., Demignon,J., Daubas,P., Buckingham,M. and Maire,P. (2007) Six proteins regulate the activation of Myf5

expression in embryonic mouse limbs. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 11310–11315.
68. Salminen,M., Lopez,S., Maire,P., Kahn,A. and Daegelen,D. (1996) Fast-muscle-specific DNA-protein interactions occurring in vivo at the human aldolase A M promoter are necessary for correct promoter activity in transgenic mice. *Mol. Cell. Biol.*, **16**, 76–85.
69. Ahmed,M., Xu,J. and Xu,P.X. (2012) EYA1 and SIX1 drive the neuronal developmental program in cooperation with the SWI/SNF chromatin-remodeling complex and SOX2 in the mammalian inner ear. *Development*, **139**, 1965–1977.
70. Ikeda,K., Watanabe,Y., Ohto,H. and Kawakami,K. (2002) Molecular interaction and synergistic activation of a promoter by Six, Eya, and Dach proteins mediated through CREB binding protein. *Mol. Cell. Biol.*, **22**, 6759–6766.
71. Delgado-Olguin,P., Huang,Y., Li,X., Christodoulou,D., Seidman,C.E., Seidman,J.G., Tarakhovsky,A. and Bruneau,B.G. (2012) Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for postnatal cardiac homeostasis. *Nat. Genet.*, **44**, 343–347.
72. Zhang,T., Ranade,S., Cai,C.Q., Clouser,C. and Pignoni,F. (2006) Direct control of neurogenesis by selector factors in the fly eye: regulation of atonal by Ey and So. *Development*, **133**, 4881–4889.
73. Zou,D., Silvius,D., Fritzsch,B. and Xu,P.X. (2004) Eya1 and Six1 are essential for early steps of sensory neurogenesis in mammalian cranial placodes. *Development*, **131**, 5561–5572.