Human Molecular Genetics, 2015, Vol. 24, No. 2 **410–423**
doi:10.1093/hmg/ddu457
Advance Access published on September 4, 2014

DNA methylation analysis of human myoblasts during *in vitro* myogenic differentiation: *de novo* methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation

Kohei Miyata[1,3,†], Tomoko Miyata[2,4,†], Kazuhiko Nakabayashi[2,†,*], Kohji Okamura[1], Masashi Naito[5], Tomoko Kawai[2], Shuji Takada[1], Kiyoko Kato[4], Shingo Miyamoto[3], Kenichiro Hata[2] and Hiroshi Asahara[1,5,6,*]

[1]Department of Systems BioMedicine and [2]Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan, [3]Department of Obstetrics and Gynecology, Faculty of Medicine, Fukuoka University, Fukuoka 814-0180, Japan, [4]Department of Obstetrics and Gynecology, Graduate School of Medical Sciences, Kyushu University, Fukuoka 812-8582, Japan, [5]Department of Systems BioMedicine, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan and [6]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA

Although DNA methylation is considered to play an important role during myogenic differentiation, chronological alterations in DNA methylation and gene expression patterns in this process have been poorly understood. Using the Infinium HumanMethylation450 BeadChip array, we obtained a chronological profile of the genome-wide DNA methylation status in a human myoblast differentiation model, where myoblasts were cultured in low-serum medium to stimulate myogenic differentiation. As the differentiation of the myoblasts proceeded, their global DNA methylation level increased and their methylation patterns became more distinct from those of mesenchymal stem cells. Gene ontology analysis revealed that genes whose promoter region was hypermethylated upon myoblast differentiation were highly significantly enriched with muscle-related terms such as 'muscle contraction' and 'muscle system process'. Sequence motif analysis identified 8-bp motifs somewhat similar to the binding motifs of ID4 and ZNF238 to be most significantly enriched in hypermethylated promoter regions. ID4 and ZNF238 have been shown to be critical transcriptional regulators of muscle-related genes during myogenic differentiation. An integrated analysis of DNA methylation and gene expression profiles revealed that *de novo* DNA methylation of non-CpG island (CGI) promoters was more often associated with transcriptional down-regulation than that of CGI promoters. These results strongly suggest the existence of an epigenetic mechanism in which DNA methylation modulates the functions of key transcriptional factors to coordinately regulate muscle-related genes during myogenic differentiation.

INTRODUCTION

DNA methylation plays a critical role in gene silencing in vertebrates and is essential for mammalian development (1) during which genome-wide alterations in DNA methylation patterns occur. In the human genome, ~15% of the CpG dinucleotides are found in CpG islands (CGIs), which are typically defined as genomic regions of >200 bp in size, a GC content greater

*To whom correspondence should be addressed at: 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan. Tel: +81 34160181; Fax: +81 34172864; Email: nakabaya-k@ncchd.go.jp (K.N.); 1-5-45 Yushima, Bunkyou-ku, Tokyo 113-8519, Japan. Tel: +81 358035015; Fax: +81 358035810; Email: asahara.syst@tmd.ac.jp (H.A.)
†K.M., T.M. and K.N. contributed equally to the work.

than 50%, and an observed-to-expected ratio of CpG ≥ 0.6. Previous studies have shown that tissue-specific transcription is controlled, in part, by tissue-specific differentially methylated regions, which contain both CGIs and non-CGIs (2–4). *Dnmt3a* and *Dnmt3b*, which function in *de novo* cytosine methylation, are responsible for the methylation pattern at the early stage of embryogenesis (5,6).

Skeletal muscle regeneration is mediated by satellite cells, which reside between the muscle fiber and basal lamina and behave as adult muscle stem cells (7,8). Satellite cells undergo symmetric cell divisions for self-renewal and asymmetrical cell divisions for myogenic differentiation. Myogenesis is regulated by a serial cascade of transcription factors. Transcriptional regulators involved in this process include Pax7 in activated satellite cells, MyoD and Myf5 in proliferating myoblasts and myogenin (*Myog*) in fusion competent myocytes (9). Previous reports suggest that the gene expression of these key transcriptional regulators is affected, in part, by DNA methylation. The relation between DNA methylation and the onset of myogenesis was established by the observation that artificial demethylation of the genome with 5-azacytidine (5AC) induced myogenic differentiation of murine C3H10T1/2 fibroblasts (10). This observation led to the cloning of *MyoD*, one of the master genes for myogenesis, which have been shown to induce myogenesis in a number of non-myogenic cell lines (11). DNA methylation of the *Myog* promoter region has also been shown to play a crucial role in the negative regulation of *Myog* expression. During early mouse development, the *Myog* promoter is reported to be initially methylated but becomes demethylated as the development proceeds (12). Furthermore, culture of murine mesenchymal progenitor cells with 5AC has been shown to up-regulate *Myog* expression at the myoblast stage and to promote myogenesis (13).

In vitro myoblast differentiation experiments have been performed as a model of skeletal muscle regeneration and development. A recent genome-wide methylation study using reduced representation bisulfite sequencing has indicated that *de novo* DNA methylation occurs in the myoblast genome during differentiation into myotubes, especially in promoters of genes that encode homeodomain and T-box proteins (14). The same group also reported that human skeletal tissues were globally hypomethylated compared with *in vitro* differentiated myotubes (14). Furthermore, DNA hypomethylation in skeletal muscle has been shown to be important for the maturation of myotubes (13). Although these studies have provided many new insights, genome-wide DNA methylation patterns during myogenesis have not been fully understood.

DNA methylation of the promoter can affect the gene expression because methylated CpG dinucleotides act as binding sites for methyl-CpG-binding domain proteins, which inhibit the binding of transcription factors to the promoters (15). It is important to consider the CpG density of promoters when analyzing the correlation of gene expression and DNA methylation (16). In this study, we analyzed genome-wide alterations in DNA methylation during myogenesis using the Infinium HumanMethylation450 BeadChip array. We observed a tendency of increased DNA methylation at the genome-wide level during myogenic differentiation and a highly significant enrichment of muscle-related genes among the genes whose promoter regions become hypermethylated upon differentiation. DNA motif analysis for the hypermethylated promoter regions identified 8-bp motifs similar to the binding motifs of key myogenic transcriptional regulators, suggesting the existence of an epigenetic mechanism, in which *de novo* DNA methylation functions to coordinately regulate muscle-related genes during myogenic differentiation. Our data provide a basis toward elucidating epigenetic principles that regulate myogenic differentiation.

RESULTS

Genome-wide DNA methylation changes during *in vitro* skeletal myoblast differentiation

Human myoblasts (MBs) were stimulated with low-concentration serum to induce myogenic differentiation. Cells were collected at four time points, i.e. pre-stimulation (MB_day1), mid-stimulation (MB_day3 and MB_day8) and post-differentiation (MB_day15) (Supplementary Material, Fig. S1) and subjected to real-time polymerase chain reaction (PCR) analysis to validate the expression of myogenesis markers (Supplementary Material, Fig. S2). The expression levels of *MYOD1* (Supplementary Material, Fig. S2A) and *MYF5* (Supplementary Material, Fig. S2B) were the highest in MB_day1 and gradually decreased in a time-dependent manner. *MYOG* (Supplementary Material, Fig. S2C) and *CKM* (Supplementary Material, Fig. S2D) were up-regulated in MB_day3 and MB_day8. These results were similar to those of a previously published study (17).

Using Infinium HumanMethylation450 BeadChip arrays, we obtained DNA methylation profiles of all skeletal myoblast samples (MB_day1, MB_day3, MB_day8 and MB_day15) as well as those of human mesenchymal stem cells (hMSCs) and two human skeletal muscle tissues (SKm1 and SKm2), which were used for comparison. In the hierarchical clustering analysis for the entire methylation dataset using complete linkage with Euclidean distance, the four myoblast samples clustered together, whereas hMSCs, SKm1 and SKm2 branched off from the cluster of myoblasts (Fig. 1A). In the cluster of myoblasts, the four samples were placed in a chronological order, indicating gradual accumulation of DNA methylation changes as myogenic differentiation proceeded. Boxplot representation of the β-values of the myoblast samples showed that the median β-values tended to increase as the culture time increased (Fig. 1B), indicating that the global DNA methylation level increased during myogenic differentiation.

To investigate whether DNA methylation levels in myoblasts differ depending on the CpG content, we categorized 450K probes into four groups, CGI, shore (0–2 kb distance from CGI), shelf (2–5 kb distance from CGI) and non-CGI ('open sea'), according to Illumina's annotation. The DNA methylation level of the CGI category was the lowest in all four samples (Fig. 1C–F), which is consistent with the fact that the majority of CGIs are unmethylated (18). The tendency of increased DNA methylation during myogenesis was most pronounced in the shore category. Next, we investigated the differential DNA methylation patterns during myogenesis in each of the seven gene feature groups according to Illumina's annotation: intergenic, TSS1500 [within 1500 bp of a transcription start site (TSS)], TSS200 (within 200 bp of a TSS), 5′ untranslated region (UTR), first exon, gene body and 3′ UTR. Scatter plot representation of the β-values of MB_day3, MB_day8 and

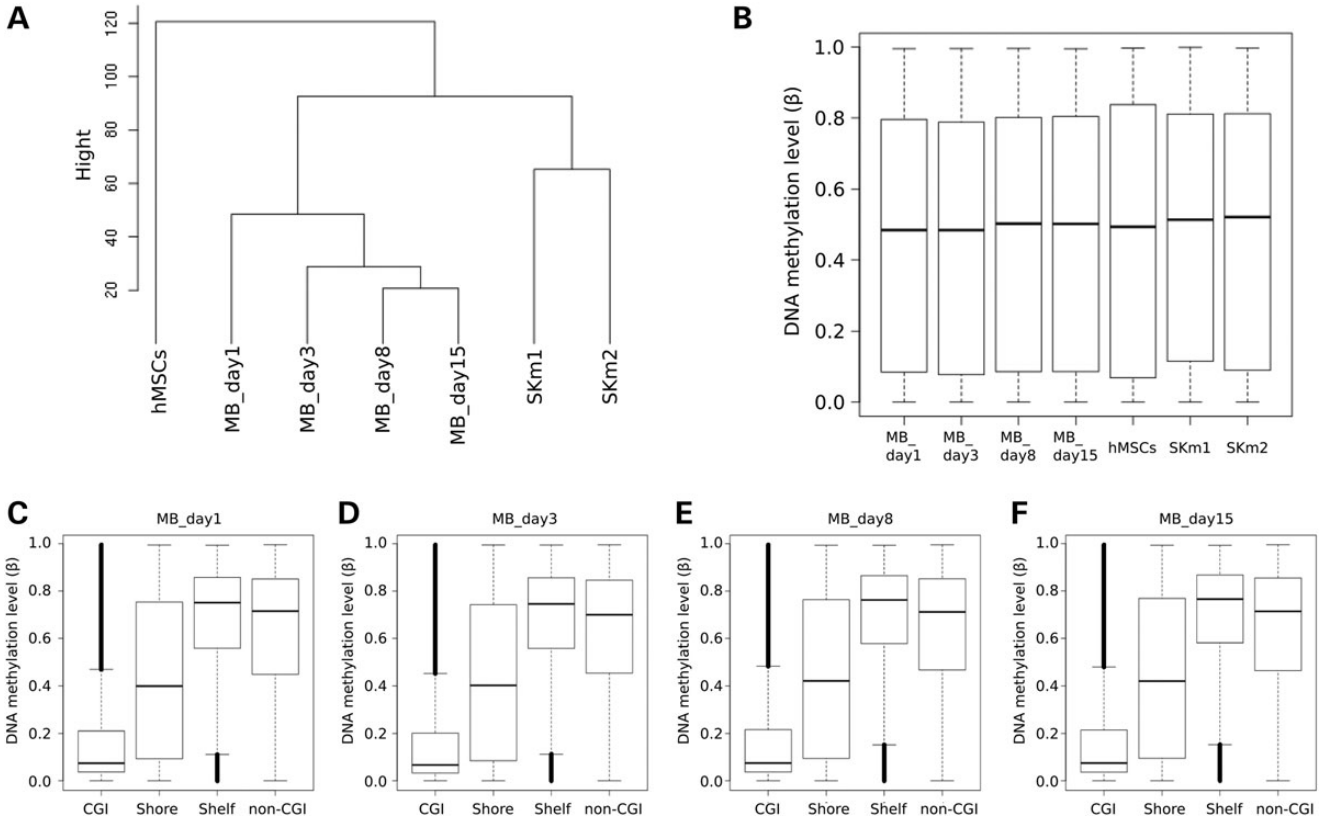


Figure 1. DNA methylation profiles of myoblasts (MBs), hMSCs and human skeletal muscle samples (SKm1 and SKm2). Hierarchical clustering (**A**) and boxplots (**B**) for all β-values. Boxplots of β-values for four subgroups categorized by CpG density and distance from CpG island (CGI): CGI, shore, shelf and non-CGI. MB_day1 (**C**), MB_day3 (**D**), MB_day8 (**E**) and MB_day15 (**F**). Black horizontal lines in boxplots indicate the median β-values.

MB_day15 compared with MB_day1 and calculation of the ratios of hyper- and hypo-methylated probes [delta β ($\Delta\beta$) $\geq$ 0.2 and $\Delta\beta \leq -0.2$, respectively] revealed that hypermethylation was observed more frequently than hypomethylation in all seven groups (Fig. 2). This result is consistent with the increase in global DNA methylation during myogenesis shown in Figure 1B. The hypermethylation tendency was in particular evident in the following three groups: intergenic, gene body and 3′ UTR (Fig. 2). We subsequently examined the DNA methylation patterns during myogenesis in subcategories by the CpG content: CGI, shelf and shore, and non-CGI. Scatter plot representation of the β-values in these three categories revealed that the DNA methylation level changed less frequently in the CGI category than that in the other two categories (Fig. 3). We also assessed whether certain types of repetitive sequences contribute to the genome-wide hypermethylation tendency during myoblast differentiation. However, as far as LINE, SINE, LTR elements were examined, no striking hypermethylation was observed (Supplemental Material, Fig. S3). Taken together, these results demonstrate that, although *de novo* DNA methylation at the genome-wide level occurs more frequently than demethylation, alterations in DNA methylation patterns tend to occur less frequently in CGI and/or regions proximal to TSS than those in other regions during myogenic differentiation. This is likely because the majority of promoter regions of housekeeping genes are CGIs, and such CGIs are, in general, resistant to *de novo* DNA methylation.

Functional annotation of differentially methylated regions during myogenic differentiation

To assess whether genomic regions with altered DNA methylation are functionally linked to myogenesis, gene ontology (GO) analysis was performed for hyper- and hypo-methylated probes ($\Delta\beta \geq 0.2$ and $\Delta\beta \leq -0.2$, respectively) located in six gene feature groups, respectively. Gene ontology analysis was performed using the Genomic Region Enrichment of Annotations Tool (GREAT, http://bejerano.stanford.edu/great/public/html/), which is a publicly available tool to analyze the functional significance of *cis*-regulatory regions (19). Compared with MB_day1, 4103 and 1191 probes were extracted to be hyper- and hypo-methylated in MB_day8, and 10 066 and 3097 probes to be hyper- and hypo-methylated in MB_day15 compared with MB_day1 among 485 577 probes (Fig. 4A). We compiled the BED-format lists of hyper- and hypo-methylated regions by binning two or more probes that are located in the same gene feature groups of the same gene as one region (the complete lists are provided in Supplementary Material, Tables S1 and S2) and subjected them to the GREAT annotation. In the GO analysis for MB_day15, hypermethylated regions in TSS1500, TSS200 and first exon were significantly associated with the GO biological process term 'muscle system process' (binominal raw *P*-values 1.56×10^{-10}, 8.94×10^{-8} and 4.37×10^{-7}, respectively) and also highly associated with several other muscle-related terms (Fig. 4B–D and Supplementary Material,

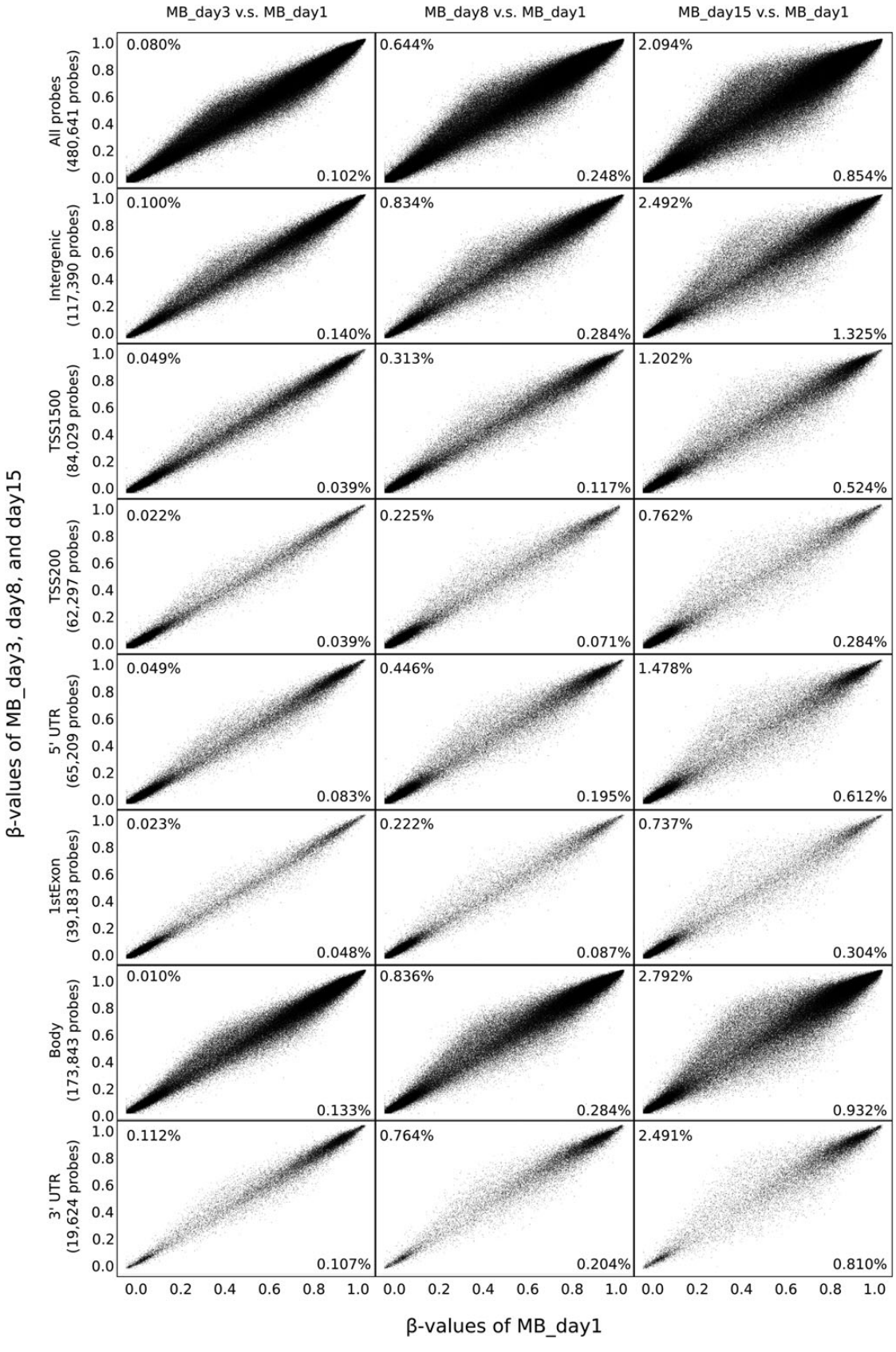


Figure 2. DNA methylation differences in MB_day3, MB_day8 and MB_day15 (y-axis) compared with MB_day1 (x-axis). Scatter plots of β-values for all probes and each of the seven gene feature groups: intergenic, TSS1500, TSS200, 5′ UTR, first exon, gene body and 3′ UTR. Percentages of the probes whose $\Delta\beta$ is ≥ 0.2 (upper left) and ≤ -0.2 (lower right) are displayed in each panel.

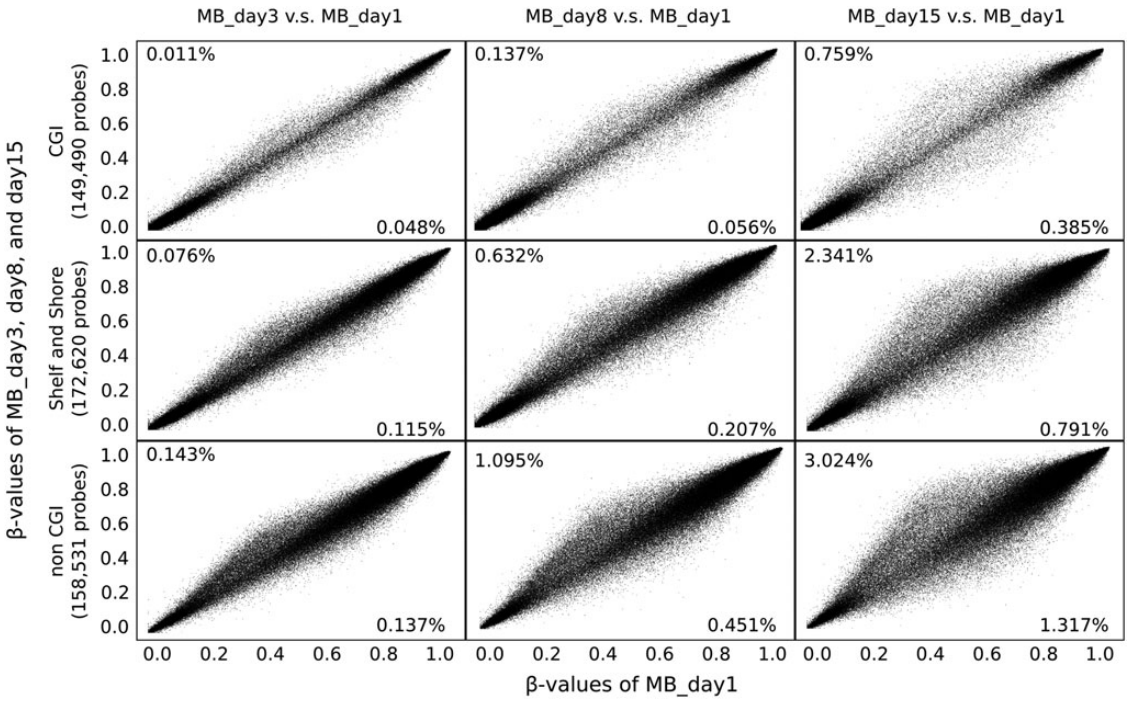


Figure 3. Scatter plots of β-values for CGI, shelf and shore, and non-CGI probes [MB_day3, MB_day8 or MB_day15 (y-axis) versus MBs_day1 (x-axis)]. Percentages of the probes whose $\Delta\beta$ is ≥ 0.2 (upper left) and ≤ -0.2 (lower right) are displayed in each panel.

Table S3). In contrast, none of the hypomethylated regions (in each of the six gene feature groups) was found to be enriched with muscle-related GO terms, except the term 'actin filament-based process' in 5′ UTR (binominal raw P-values 3.03×10^{-6}) (Supplementary Material, Table S3). In the GO analysis for MB_day8, there was no association between hypermethylated regions and muscle-related GO biological process terms. However, hypermethylated regions in TSS1500 and gene body were found to be enriched with muscle-related GO terms categorized into 'Cellular Compornent', such as 'myofibril' and 'contractile fiber' (Supplementary Material, Tables S2 and S4). Overall, hypermethylated regions proximal to TSS (especially TSS200 and TSS1500) are most strikingly enriched with muscle-related GO terms. These results suggest that the promoter regions of particular muscle-related genes are regulated epigenetically through *de novo* DNA methylation during myoblast differentiation.

Functional annotation of differentially expressed genes during myogenesis

To assess whether the genes whose expression was altered during myogenesis are functionally linked to myogenesis, GO analysis was performed for up- and down-regulated genes (fold change ≥ 2.0 and ≤ 0.5, respectively) using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov). In MB_day3, MB_day8 and MB_day15, compared with MB_day1, 1626, 1622 and 1537 genes were found to be up-regulated and 1173, 1456, and 1515 genes to be down-regulated (Supplementary Material, Table S5). The numbers of differentially expressed genes were similar among these three time points during myoblast differentiation. This pattern of differential gene expression is in contrast with that of DNA methylation changes, in which the extent of differential methylation at Day 3 was still small and became more and more evident as culture time went by (Figs. 2 and 3).

In GO analysis, down-regulated genes at all three time points were found to be strikingly enriched with cell cycle-related GO terms, such as 'cell cycle' and 'cell cycle phase' with Benjamini P-values of $< 10^{-55}$ (Supplementary Material, Fig. S4A–C and Tables S5 and S6), consistent with the cell cycle arrest that immediately occurs after low-serum treatment to myoblasts for the induction of myogenic differentiation. Up-regulated genes were found to be enriched with muscle-related GO terms such as 'muscle organ development' and 'muscle tissue development' as well as other terms such as 'regulation of cell proliferation' and 'extracellular matrix organization' (Fig. 5B, C and Supplementary Material, Table S6). However, the statistical significance of the enrichment of these terms was much weaker than that of cell cycle-related GO terms observed in down-regulated genes.

Taken together with the GO annotation results for the differentially methylated regions, these results suggest that the majority of differential gene expression during myogenic differentiation occurs independently of DNA methylation and is unrestricted to muscle-related genes and that the global alteration of DNA methylation is a subsequent event to alteration of gene expression.

Motif-based sequence analyses identify 8-bp motifs highly enriched in hypermethylated promoter regions

To identify candidate transcriptional factors that regulate promoters that are targets of *de novo* DNA methylation during myogenic differentiation, we searched for enriched DNA motifs in

A

485,577 probes

QC: probes with a detection P-value of < 0.05 or a blank β value were excluded

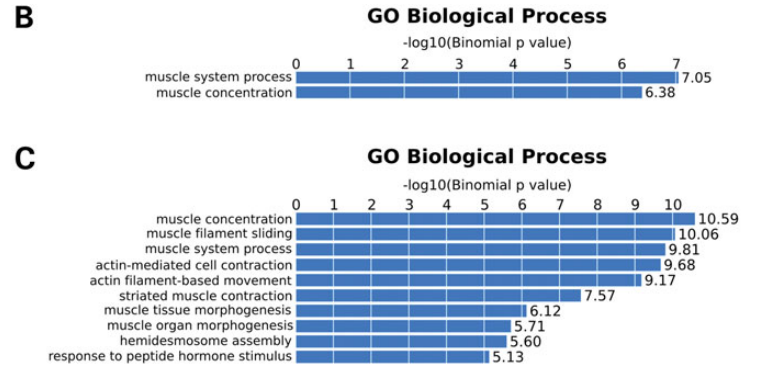
480,641 probes

Hypermethylated probes (Δβ ≥ 0.2)

	v.s. day 3	v.s. day 8	v.s. day 15
Intergenic	116	979	2,925
TSS1500	41	263	1,011
TSS200	16	140	475
5' UTR	32	291	964
1st exon	9	87	289
Gene body	173	1,453	4,854
3' UTR	22	150	493
Total	384	3,097	10,066

Hypomethylated probes (Δβ ≤ -0.2)

	v.s. day 3	v.s. day 8	v.s. day 15
Intergenic	164	457	1,555
TSS1500	33	98	440
TSS200	24	45	177
5' UTR	54	127	399
1st exon	19	34	119
Gene body	232	494	1,620
3' UTR	21	40	159
Total	491	1,191	4,103



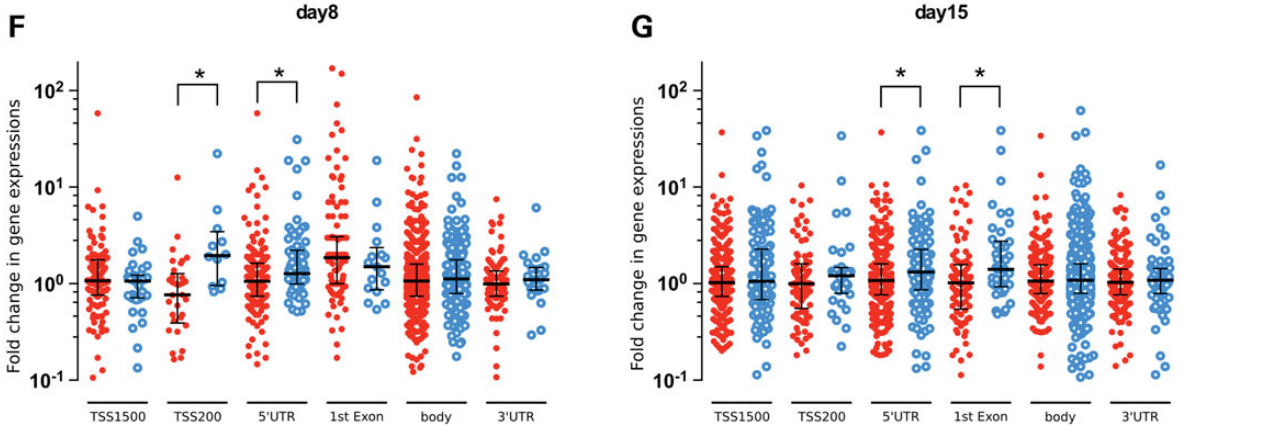


Figure 4. Analysis of DNA methylation status and gene expression in each of the seven gene feature groups. Flowchart for the processing of DNA methylation array data (**A**). The numbers of differentially methylated ($\Delta\beta \geq 0.2$ or $\Delta\beta \leq -0.2$) probes in MB_day3, MB_day8 and MB_day15 compared with MB_day1 in total and in each of the seven gene feature groups are shown. Gene ontology analysis for hypermethylated genomic intervals in MB_day15 using GREAT (**B**–**E**). Enriched terms in the GO biological process are shown for TSS1500 (B), TSS200 (C), 5′ UTR (D) and first exon (E). Blue bar charts represent binomial *P*-values (in $-\log_{10}$ scale). Dot plots of fold changes in gene expression in MB_day8 (**F**) and MB_day15 (**G**) compared with MB_day1 for hypermethylated ($\Delta\beta \geq 0.2$, red closed circles) and hypomethylated ($\Delta\beta \leq -0.2$, blue open circles) probes in six gene feature groups. Black horizontal lines in the boxplots indicate the median β-values. * *P*-value < 0.05.

sequences flanking the sides of hypermethylated probes using Multiple Em for Motif Elicitation (MEME, http://meme.nbcr.net/meme/) and a motif comparison tool, Tomtom (http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi). We compiled 50-bp DNA sequences flanking both sides of hypermethylated probes in the categories TSS1500, TSS200, 5′ UTR and first exon using a custom script. When two 101-bp sequences overlapped each other, they were merged to one continuous, longer sequence. The resultant genomic intervals from the categories TSS1500, TSS200, 5′ UTR and first exon (Supplementary Material, Table S7) were subjected jointly or separately to MEME using the parameters described in the section Materials and Methods. Only two motifs (Motif 1 C[AT]GCTG[CGT] [CTG] and Motif 2 [CG][AT]G[GC]C[TA]G[GC]) were found to be significantly over-represented (*E*-values 1.4E−94 and 5.5E−41, respectively) in close proximity of hypermethylated probes, when 1327 genomic intervals in all four categories were jointly subjected to the MEME analysis (Fig. 6A and Supplementary Material, Table S8). These two 8-bp motifs were found 444 and 596 times in 711 out of 1327 hypermethylated genomic intervals around TSS (53.5%) (Supplementary Material, Table S8). Separate analyses of genomic intervals in each

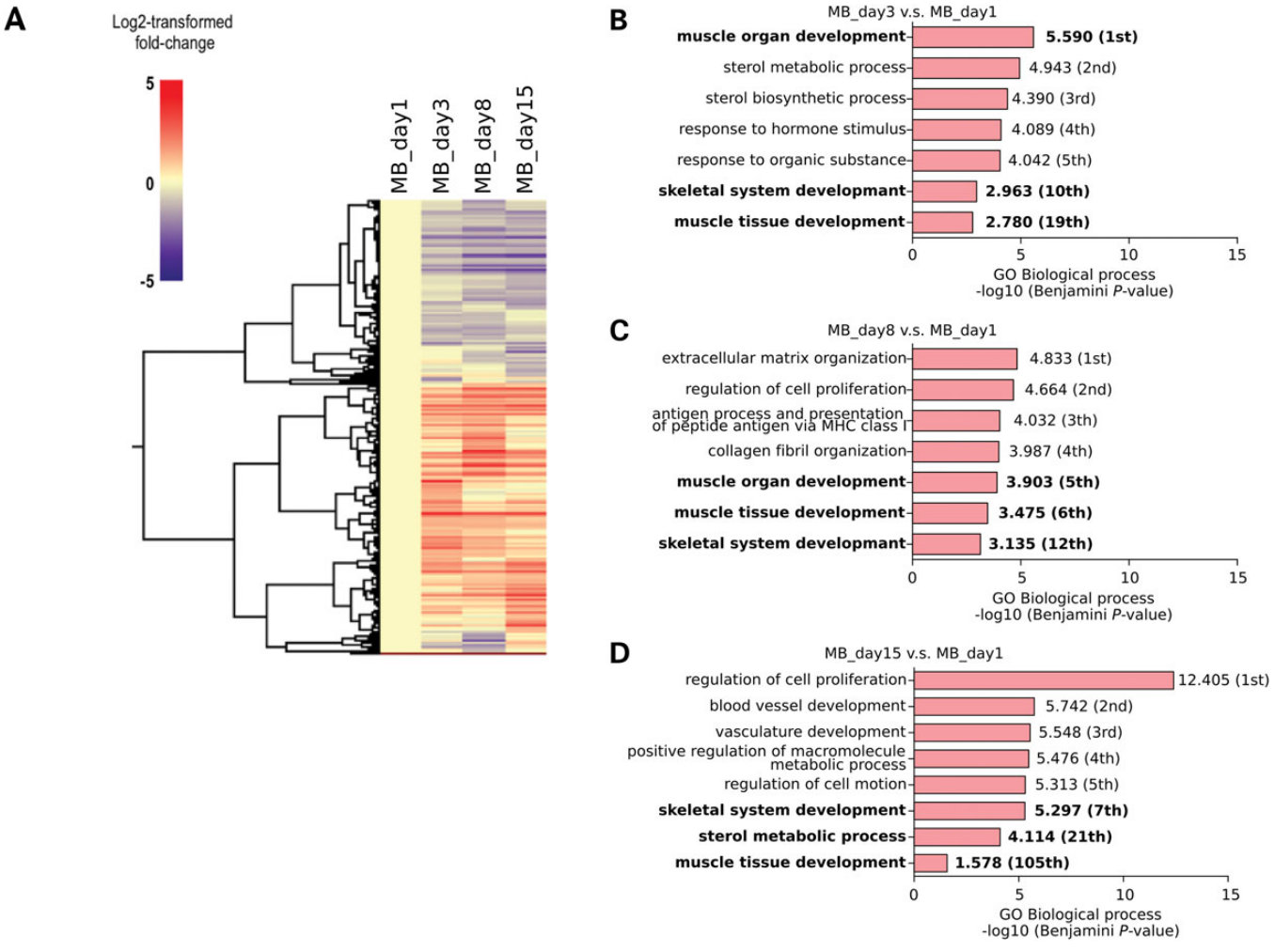


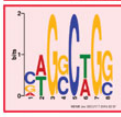
Figure 5. Heatmap representation of the fold changes of differentially expressed genes during myoblast differentiation (**A**). The fold change levels of the genes that are differentially expressed (fold change ≥ 2.0 or ≤ 0.5) in MB_day3, MB_day8 and MB_day15 compared with MB_day1 are shown (using log2-transformed fold-change values). Only the signal intensities with the expression call flag of 'Present' were subjected to extract differentially expressed genes. Gene ontology analysis for up-regulated (fold change ≥ 2.0) genes using DAVID (**B**–**D**). Enriched terms in the GO biological process are shown for MB_day3 (B), MB_day8 (C) and MB_day15 (D) compared with MB_day1 (muscle-related terms are shown in bold). Red bar charts represent Benjamini *P*-values (in $-\log_{10}$ scale). The rank of the Benjamini *P*-value in the list of enriched GO terms (Supplementary Material, Table S6A) is shown in parentheses.

of four gene feature groups also discovered motifs similar to Motif 1 in the joint analysis (Motif 1 in Fig. 6B–E). Interestingly, further analysis using Tomtom revealed that the motifs identified in joint and separate analyses were somewhat similar to the 'ID4_DBD' motif (Fig. 6A–E). 'ZNF238_full' and 'ZNF238_DBD' motifs were also found in the separate analyses for TSS1500 and TSS200 (Fig. 6B and C). The ID families and ZNF238 (also known as ZBTB18 or RP58) have been shown to be key transcriptional factors in myogenic differentiation (20–23). These results strongly suggest the existence of epigenetic mechanism in which DNA methylation modulates the functions of key transcriptional factors to coordinately regulate muscle-related genes during myogenic differentiation.

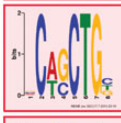
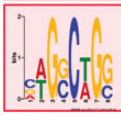
Integrated analysis of DNA methylation and gene expression

To investigate the relationship between DNA methylation and gene expression during myogenic differentiation, we conducted an integrated analysis of these data. Hyper- and hypo-methylated probes summarized in Figure 4A except for the intergenic probes were subjected to the analysis. The corresponding normalized signal intensities of the genes with differentially methylated probes were extracted from the expression array data using gene symbols for matching. The fold changes in expression of MB_day8 and MB_day15 compared with MB_day1 were plotted for hyper- and hypo-methylated genes separately (shown by red and blue dots, respectively) for each of the six categories (Fig. 4F and G). The Mann–Whitney *U*-tests showed that the fold changes of expression of hypomethylated genes were significantly greater than those of hypermethylated genes in the categories TSS200 ($P = 0.0013$) and 5′ UTR ($P = 0.0020$) in MB_day8, and 5′ UTR ($P = 0.0040$) and first exon ($P = 0.0034$) in MB_day15: DNA hyper- and hypo-methylation in the TSS200 regions at MB_day8 were modestly but most significantly associated with the down- and up-regulation of gene expression.

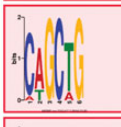
A

	E-value	Length	Sites	Logo	Similar motif registered in TOMTOM
motif1	1.4E-94	8	444		Ascl2_DBD, NHLH1_DBD, NHLH1_full, TGIF1_DBD, TGIF2LX_full, TGIF2_DBD, MEIS2_DBD, MYF6_full, ID4_DBD, PKNOX1_DBD, Tcf21_DBD, Meis3_DBD_2, Pknox2_DBD, SNAI2_DBD, PKNOX2_DBD, MEIS3_DBD_2, Meis2_DBD_2, TFAP4_full, TFAP4_DBD, SCRT_DBD, MESP1_DBD, TCF4_DBD, TCF3_DBD, FIGLA_DBD, Atoh1_DBD, SCRT2_DBD, HEY1_DBD
motif 2	5.5E-41	8	596		ID4_DBD, TFAP2C_full_3

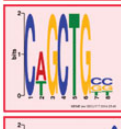
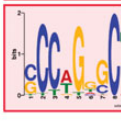
B

	E-value	Length	Sites	Logo	Similar motif registered in TOMTOM
motif1	1.1E-37	8	262		Ascl2_DBD, NHLH1_DBD, NHLH1_full, TCF3_DBD, TCF4_DBD, TCF4_full, MESP1_DBD, ID4_DBD, SNAI2_DBD, TGIF1_DBD, TGIF2LX_full, MYF6_full, TGIF2_DBD, FIGLA_DBD, Tcf21_DBD, MEIS2_DBD, PKNOX1_DBD, SCRT2_DBD, Meis3_DBD_2, Pknox2_DBD, PKNOX_DBD, SCRT1_DBD, Meis2_DBD, ZIC3_full, HEY_DBD, MEIS_DBD, Zic3_DBD, MNT_DBD, ZNF238_full, ZIC1_full, HEY2_full, MAX_DBD_2, ZNF238_DBD, HES5_DBD, ZIC2_DBDHEY_DBD
motif 2	2.9E-08	8	256		ID4_DBD, TFAP2C_full_3, Tcfap2a_DBD_2

C

	E-value	Length	Sites	Logo	Similar motif registered in TOMTOM
motif1	9.9E-16	6	135		Ascl2_DBD, NHLH1_DBD, NHLH1_full, TFAP4_DBD, TFAP4_full, Tcf21_DBD, TGIF1_DBD, MYF6_full, TGIF2_DBD, TGIF2LX_full, MEIS2_DBD, PKNOX1_DBD, Meis_DBD_2, Pknox2_DBD, Meis2_DBD_2, MEIS3_DBD_2, FIGLA_DBD, PKNOX2_DBD, TCF3_DBD, ID4_DBD, TCF4_DBD, SNAI2_DBD, MESP1_DBD, TCF4_full, ZNF238_full, ZNF238_DBD, SCRT2_DBD
motif 2	8.3E-05	8	134		ID4_DBD, MEIS2_DBD_2, PKNOX2_DBD, Pknox2_DBD

D

	E-value	Length	Sites	Logo	Similar motif registered in TOMTOM
motif1	3.1E-31	8	162		NHLH1_DBD, Ascl2_DBD, NHLH1_full, TGIF1_DBD, TGIF2_DBD, TGIF2LX_full, MYF6_full, MEIS2_DBD, PKNOX1_DBD, ID4_DBD, Tcf21_DBD, Meis3_DBD, Pknox2_DBD, SNAI_DBD, MEIS_DBD_2, Meis_DBD_2, PKNOX2_DBD, TFAP4_full, MESP1_DBD, TFAP4_DBD, SCRT_DBD, SCRT_DBD, TCF_DBD, TCF_DBD, ZIC3_full, FIGLA_DBD
motif 2	4.5E-03	11	72		Hic1_DBD, TFAP2C_full, Tcfap2a_DBD_3, TFAP2C_DBD, TFAP2B_DBD

E

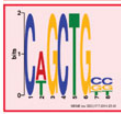
	E-value	Length	Sites	Logo	Similar motif registered in TOMTOM
motif1	4.4E-13	8	53		Ascl2_DBD, SNAI2_DBD, ID4_DBD, TCF3_DBD, TCF4_DBD, MNT_DBD, NHLH1_full, Tcf21_DBD, MESP1_DBD, NHLH1_DBD, MAX_DBD_2, HES5_DBD, HES7_DBD, TCF4_full, SCRT2_DBD, CTCF_full, FIGLA_DBD

Figure 6. DNA sequence motifs identified to be enriched in the flanking region (−50 to +50 bp) of hypermethylated CpG sites in all four categories (from TSS1500 to first exon) (**A**), TSS1500 (**B**), TSS200 (**C**), 5′ UTR (**D**) and first exon (**E**) by MEME analysis. *E*-values represent the statistical significance of the motifs. The rightmost column contains the list of proteins detected by Tomtom analysis to have a DNA-binding motif similar to each of the motifs enriched nearby hypermethylated CpG sites.

Because it has been suggested that the extent to which gene expression correlates with DNA methylation of the promoter region, in part, depends on its CpG density (24), we evaluated whether this is also the case in myogenic differentiation. We categorized genes into CGI genes and non-CGI genes according to RefSeq database (see Materials and Methods) and annotated the CGI status of each of the 124 925 probes in the categories TSS1500 and TSS200 that we regarded as those with which DNA methylation levels of promoters can be measured. The numbers of the genes hosting hyper- and hypo-methylated probes in the TSS1500 and TSS200 categories were summarized in the tables in Figure 7A (full lists are provided as Supplementary Material, Table S9). The numbers of up- and down-regulated genes among those were counted and further categorized into

CGI- and non-CGI promoter genes (Fig. 7A). Using gene symbols for matching, the expression signal intensities of those genes with hypermethylated and hypomethylated promoters were retrieved for Mann–Whitney *U*-test to determine differences in the gene expression levels. Among the tests for all, CGI promoter and non-CGI promoter genes, statistically significant differences were detected in the tests for all genes ($P = 0.0013$) and non-CGI genes ($P = 0.0006$) in MB_day8 (Fig. 7C), and for non-CGI genes in MB_day15 ($P = 0.0183$) (Fig. 7D). These results demonstrate that, during myogenic differentiation, DNA methylation of non-CGI promoters suppresses gene expression more frequently than DNA methylation of CGI promoters.

To explore functional features of the DNA methylation-mediated control of the expression of genes with non-CGI promoters during myogenic differentiation, we performed DAVID GO analysis for two categories of genes with non-CGI promoters: (1) 20 up-regulated genes (fold change ≥ 2.0) whose promoters (TSS1500 and TSS200) were hypomethylated ($\Delta\beta \leq -0.2$) in MB_day15 compared with MB_day1 and (2) 27 down-regulated genes (fold change ≤ 0.5) whose promoters were hypermethylated ($\Delta\beta \geq 0.2$) (Supplementary Material, Table S10). Whereas no term was found to be enriched among the former genes when corrected *P*-value (by Benjamini–Hochberg procedure) of 0.05 is applied as a significance threshold, the term 'muscle organ development' was found to be significantly enriched ($Pc = 0.0030$) among the latter genes (Supplementary Material, Table S10). Out of 27 genes, 6 genes, *MYF6*, *MUSK*, *MSTN*, *TRIM72*, *CHRNA1* and *SGCA*, were assigned with this term. To validate these alterations of DNA methylation and gene expression during myogenic differentiation, bisulfite sequencing and quantitative RT–PCR were undertaken for the above six genes (Fig. 8A–F). All of six gene promoters were confirmed to become hypermethylated in MB_day15 compared with MB_day1. All genes except for MUSK were confirmed to be transcriptionally repressed in MB_day15 than in MB_day1. Interestingly, five genes other than *MYF6* were found to be up-regulated immediately after myogenic stimuli (MB_day3) and subsequently down-regulated (MB_day8 and MB_day15) (Fig. 8A–F). The observed transient up-regulation of muscle-related genes at MB_day3, which is consistent with the detection of 'muscle organ development' as the top term in GO analysis for up-regulated genes at MB_day3 (Fig. 5B), indicates that hypermethylation of muscle-related gene promoters at least occasionally occurs after their transient activation of gene expression. It should be also noted that the promoter regions of all six genes contain Motif 1 and/or Motif 2 discovered in the MEME analysis (Supplementary Material, Table S8 and Fig. S5).

Collectively, our results of integrated analyses of DNA methylation and gene expression profiles suggest that DNA methylation, especially *de novo* DNA methylation, of certain non-CGI promoters contributes to the transcriptional down-regulation of muscle-related genes during myogenic differentiation.

DISCUSSION

In vitro skeletal muscle formation is among the most investigated models for studying molecular mechanisms of tissue differentiation and development. In this study, by obtaining a chronological profile of genome-wide DNA methylation in a human myoblast differentiation model, we revealed that the global DNA methylation level of myoblasts gradually increases as myogenesis proceeds. We then performed a GO analysis as a means to assess functional features of promoter regions that are targets of *de novo* DNA methylation and revealed that those of muscle-related genes were most significantly enriched. Our findings suggest critical roles of *de novo* DNA methylation in myogenic differentiation and, at the same time, raise new questions regarding how particular muscle-related genes are targeted by the *de novo* methylation machinery. Which of the DNA methyltransferases DNMT3A and DNMT3B is responsible for the *de novo* methylation of muscle-related genes? By what mechanism is the *de novo* methylation machinery recruited to the promoter regions of particular muscle-related genes? Future studies that would provide answer to these questions will significantly increase our understanding of the epigenetic principles for the commitment and differentiation of myoblasts.

As a bioinformatics approach to identify candidate transcription factors that bind to promoters that are targeted by *de novo* methylation during myoblast differentiation, we conducted a DNA motif analysis and identified two 8-bp motifs (C[AT]GCTG[CGT][CTG], [CG][AT]G[GC]C[TA]G[GC]), which were frequently observed in the proximity of hypermethylated probes in the TSS1500 to first exon category. Furthermore, these and other similar motifs were found to be somewhat similar to the binding motifs of ID4 and ZNF238. Id proteins, including ID4, are known to be members of helix-loop-helix transcription factors that inhibit the differentiation of muscle cells by antagonizing the transcriptional activity of myogenic regulatory factors such as MyoD and MYF5 (20–22). In this regard, we recently reported that ZNF238, also known as ZBTB18 and RP58, is a critical transcriptional repressor inhibiting the expression of Id genes (23). Because functional inhibition of Id proteins is a critical step during myogenesis, DNA methylation of binding sites of Id proteins may contribute to the promotion of myogenic differentiation in addition to the RP58-Id regulatory network. It is also noteworthy that RP58 is known to bind to DNMT3A (25). Therefore, RP58 represents a strong candidate transcriptional repressor, which may recruit DNMT3A to the promoters of certain muscle-related genes for *de novo* methylation.

To further clarify roles of *de novo* DNA methylation in the transcriptional regulation of muscle-related genes, we conducted an integrated analysis of gene expression and DNA methylation data. Gene ontology term analysis showed that down-regulated genes with hypermethylated non-CGI promoters were enriched with the term 'muscle organ development'. Among six genes that were assigned to this term, *MYF6*, also known as myogenic regulatory factor 4 (*MRF4*), has been known to regulate fusion of myoblasts into multinucleated muscle cells. It is also known that *MYF6* expression decreases in terminally differentiated muscle cells, in part, via down-regulation of *Myog* (26). Oikawa and colleagues reported that CIBZ, a methyl-CpG-binding protein, inhibits *Myog* expression by binding to its methylated promoter region (27). This report is supported by our study in which we showed that *Myog* expression is down-regulated by *de novo* methylation of its promoter region in terminally differentiated muscle cells. Our data also suggest that *MYF6* and other myogenic genes that need to be down-regulated during terminal differentiation may be regulated by *de novo* methylation of their promoters in addition to the inhibition of the master-transcription factor '*Myog*'.

A

Methylation array
Illmina HumanMethylation450 BeadChip

485,577 probes

QC: probes with a detection *P*-value of < 0.05 or a blank β-value were excluded

480,641 probes

Annotation with RefSeq database
Probes located between TSS (0 bp) and -1500 bp
Categolized into CGI or non-CGI promoter gene probes

124,925 probes

Integration of expression data
Affymetrix GeneChip
Human Genome U133 Plus 2.0 Array

MB_day3 v.s. MB_day1

		All	CGI	non-CGI
Δβ ≥ 0.2 16 genes	Fold change ≥ 2.0	2 genes	1 gene	1 gene
	Fold change ≤ 0.5	3 genes	1 gene	2 genes
Δβ ≤ -0.2 13 genes	Fold change ≥ 2.0	3 genes	2 genes	1 gene
	Fold change ≤ 0.5	5 genes	3 genes	2 genes

MB_day8 v.s. MB_day1

		All	CGI	non-CGI
Δβ ≥ 0.2 90 genes	Fold change ≥ 2.0	13 genes	5 genes	8 genes
	Fold change ≤ 0.5	25 genes	10 genes	15 genes
Δβ ≤ -0.2 37 genes	Fold change ≥ 2.0	17 genes	5 genes	12 genes
	Fold change ≤ 0.5	5 genes	4 genes	1 genes

MB_day15 v.s. MB_day1

		All	CGI	non-CGI
Δβ ≥ 0.2 264 genes	Fold change ≥ 2.0	36 genes	20 genes	16 genes
	Fold change ≤ 0.5	51 genes	24 genes	27 genes
Δβ ≤ -0.2 128 genes	Fold change ≥ 2.0	35 genes	15 genes	20 genes
	Fold change ≤ 0.5	27 genes	13 genes	14 genes

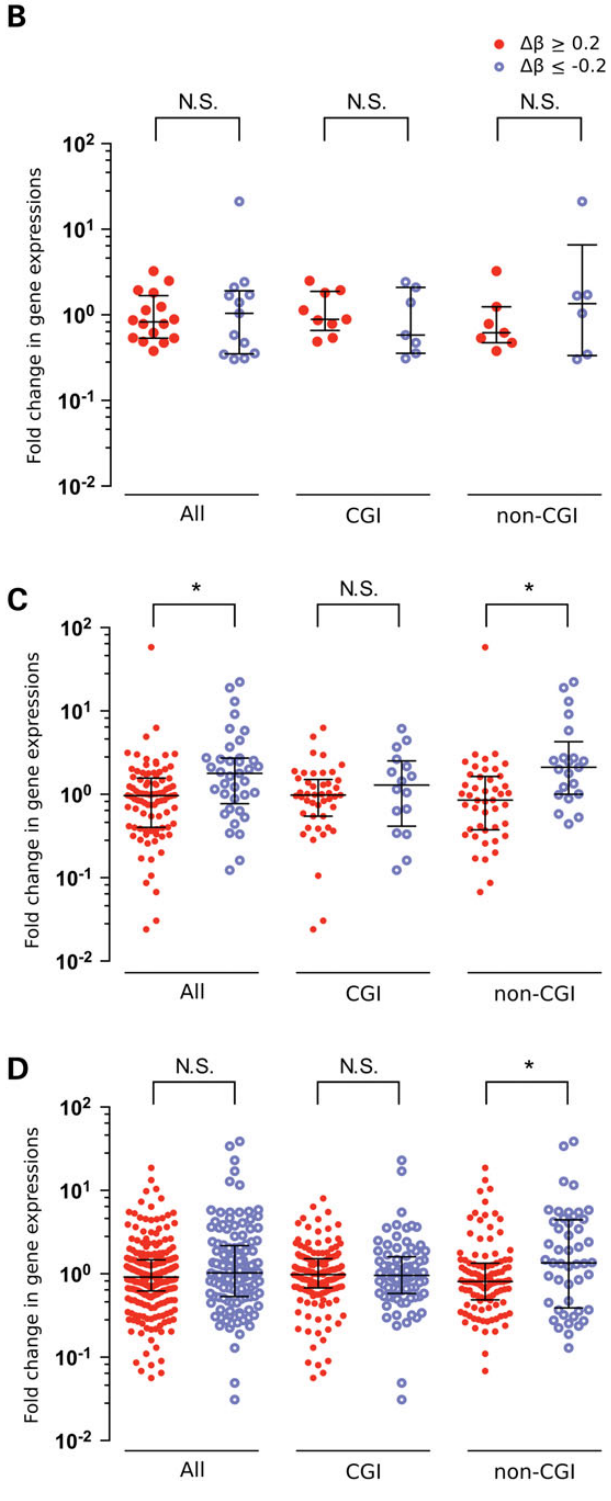
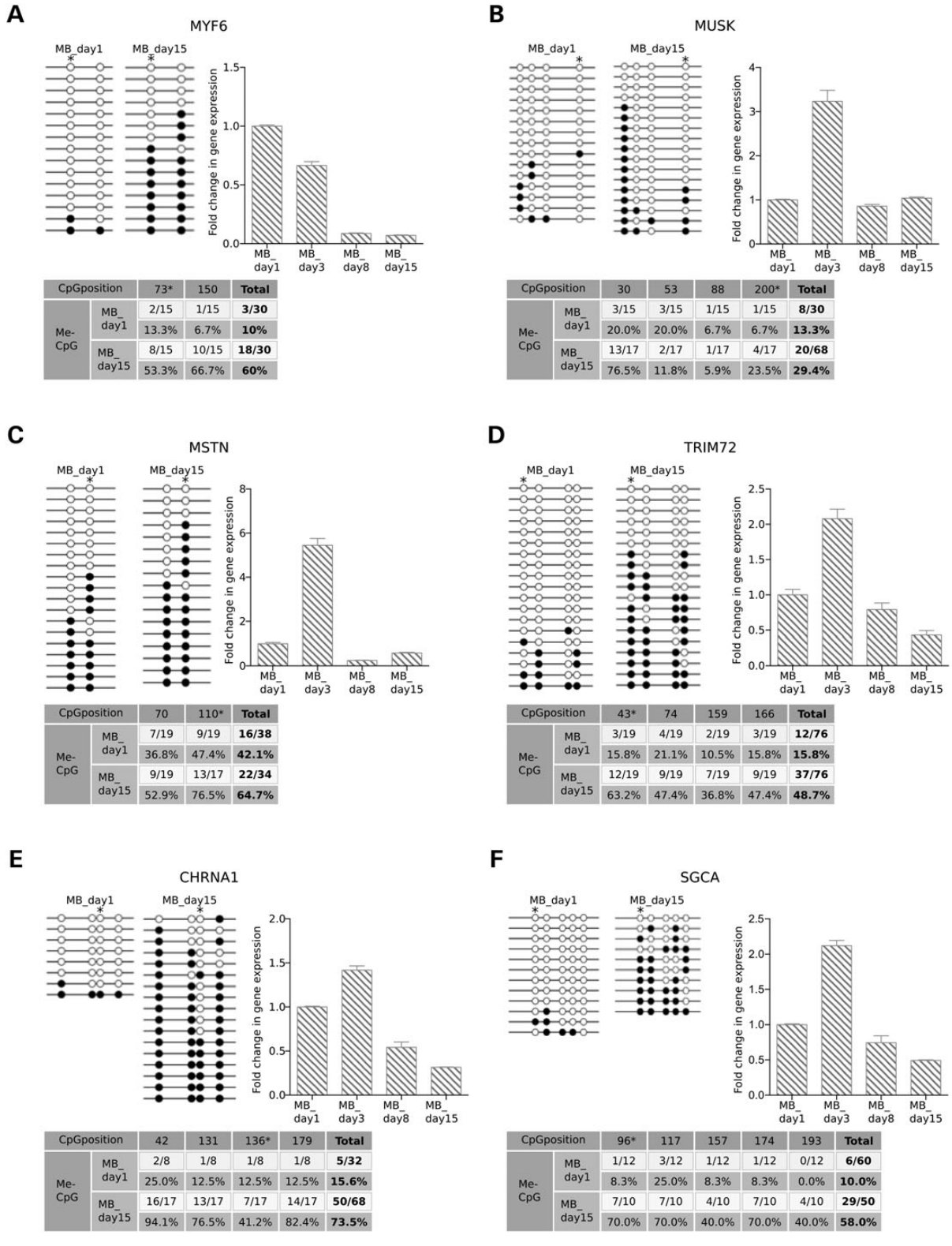


Figure 7. The integrated analysis of DNA methylation data of promoter regions (TSS1500 and TSS200) and gene expression data. Flowchart and the numbers of the genes hosting hyper- and hypo-methylated probes in the TSS1500 and TSS200 categories (**A**). Beeswarm plots of fold changes in gene expression in MB_day3 (**B**), MB_day8 (**C**) and MB_day 15 (**D**) compared with MB_day1 of hypermethylated ($\Delta\beta \geq 0.2$) and hypomethylated ($\Delta\beta \leq -0.2$) probes located in TSS1500 and TSS200 categories. Plots for all probes (CGI and non-CGI probes), CGI probes and non-CGI probes are shown. Black horizontal lines in the plots indicate the median and the first/third quantiles of fold-change values. *P-value < 0.05; N.S., not significant.

Previous studies suggested the importance of considering the CpG density of promoters when analyzing the correlation of gene expression and DNA methylation (16). CpG islands are known to associate with ~50% of a TSS (18). Approximately 70% of annotated promoters are linked to a CGI (16). The majority of CGI promoters are unmethylated and located in



A	CpGposition	73*	150	Total
Me-CpG	MB_ day1	2/15	1/15	**3/30**
		13.3%	6.7%	**10%**
	MB_ day15	8/15	10/15	**18/30**
		53.3%	66.7%	**60%**

B	CpGposition	30	53	88	200*	Total
Me-CpG	MB_ day1	3/15	3/15	1/15	1/15	**8/30**
		20.0%	20.0%	6.7%	6.7%	**13.3%**
	MB_ day15	13/17	2/17	1/17	4/17	**20/68**
		76.5%	11.8%	5.9%	23.5%	**29.4%**

C	CpGposition	70	110*	Total
Me-CpG	MB_ day1	7/19	9/19	**16/38**
		36.8%	47.4%	**42.1%**
	MB_ day15	9/19	13/17	**22/34**
		52.9%	76.5%	**64.7%**

D	CpGposition	43*	74	159	166	Total
Me-CpG	MB_ day1	3/19	4/19	2/19	3/19	**12/76**
		15.8%	21.1%	10.5%	15.8%	**15.8%**
	MB_ day15	12/19	9/19	7/19	9/19	**37/76**
		63.2%	47.4%	36.8%	47.4%	**48.7%**

E	CpGposition	42	131	136*	179	Total
Me-CpG	MB_ day1	2/8	1/8	1/8	1/8	**5/32**
		25.0%	12.5%	12.5%	12.5%	**15.6%**
	MB_ day15	16/17	13/17	7/17	14/17	**50/68**
		94.1%	76.5%	41.2%	82.4%	**73.5%**

F	CpGposition	96*	117	157	174	193	Total
Me-CpG	MB_ day1	1/12	3/12	1/12	1/12	0/12	**6/60**
		8.3%	25.0%	8.3%	8.3%	0.0%	**10.0%**
	MB_ day15	7/10	7/10	4/10	7/10	4/10	**29/50**
		70.0%	70.0%	40.0%	70.0%	40.0%	**58.0%**

Figure 8. Bisulfite sequencing and quantitative RT–PCR analyses to validate alterations of DNA methylation and gene expression levels for six muscle-related genes, *MYF6* (**A**), *MUSK* (**B**), *MSTN* (**C**), *TRIM72* (**D**), *CHRNA1* (**E**) and *SGCA* (**F**), during myogenic differentiation. Open and closed circles represent unmethylated and methylated CpG sites, respectively. Asterisks indicate the position of the CpG site that was found to be hypermethylated in MB_day15 compared with MB_day1 in the analysis of Infinium HumanMethylation450 BeadChip data [the corresponding name (IlmnID) of the probe is listed in Supplementary Material, Table S12]. Each row of circles corresponds to an individual clone sequenced. For each gene, the methylation ratios (%) of the CpG sites examined are shown in the table underneath. Gene expression levels in MB_day3, MB_day8 and MB_day15 relative to MB_day1 were displayed as bar charts. Each bar represents the mean fold change and the standard error of technical replicates ($n = 3$).

ubiquitously expressed housekeeping genes. In contrast, the promoters of tissue-specific genes tend to be non-CGIs (28,29). In our integrated analysis of DNA methylation and gene expression data for MB_day1 and MB_day15, *de novo* DNA methylation of non-CGI promoters was more frequently accompanied with the suppression of gene expression than *de novo* DNA methylation of CGI promoters. Furthermore, GO analysis for the 27 genes with non-CGI promoters that were down-regulated (expression fold change ≤ 0.5) and hypermethylated ($\Delta\beta \geq 0.2$) upon myogenic differentiation revealed significant enrichment ($P = 8.567E-06$) for muscle-related genes. These findings support the hypothesis that muscle-related genes, especially those with non-CGI promoter, are regulated by *de novo* DNA methylation during myogenic differentiation.

In this study, we described DNA methylation changes occurring during myoblast differentiation revealed by detailed and comprehensive analyses of the Infinium HumanMethylation 450 BeadChip array data. This method uses bisulfite-converted DNA and does not distinguish 5-methylcytosine and 5-hydroxymethylcytosine (5 hmC), the latter of which is generated by oxidation of the former by ten eleven translocation (TET) enzymes. It has been suggested that the TET-medicated 5 hmC not only triggers DNA demethylation but also participates in epigenetic regulation *per se* in essential biological processes, such as cell pluripotency, embryonic development and cellular differentiation (30,31). Examination of genome-wide distribution and alteration of 5 hmC during myoblast differentiation is a promising future study to broaden our understanding of epigenetic mechanisms regulating myoblast differentiation.

In summary, analysis of the DNA methylation profile in a human myoblast differentiation model revealed the frequent occurrence of *de novo* DNA methylation of muscle-related genes with non-CGI promoter. Furthermore, it suggested the existence of key transcriptional factor(s) involved in the coordinated regulation of such muscle-related genes during myogenic differentiation. Our findings provide a basis to better understand the epigenetic principles that control myogenesis.

MATERIALS AND METHODS

Cell culture and samples

Human primary skeletal muscle myoblasts (Lonza, CC-2580) were cultured in SkGM-2 BulletKit medium (Lonza, CC-3160). Differentiation into myotubes was stimulated by switching the culture medium to DMEM-F12 (Thermo scientific) supplemented with 2% horse serum (Sigma–Aldrich) when cells reached ~50% confluence (Day 1). Medium was replaced every other day and changed to SkGM-2 BulletKit medium at Day 8. SkGM-2 BulletKit medium was replaced every other day until Day 15. Cells were collected on Days 1, 3, 8 and 15 for subsequent analyses. Human normal skeletal muscle tissues from two individuals were obtained from Tissue Solutions (SK052009, SKm1; SKM102210A, SKm2). hMSCs were grown in DMEM (Thermo Scientific) supplemented with 10% fetal bovine serum (Thermo Scientific). All cell lines and human tissues were derived from male individuals.

RNA isolation, RT–PCR and expression array analysis

Total RNA was isolated from myoblasts (Days 1, 3, 8 and 15, $n = 1$ for each time point) using ISOGEN (Nippon Gene) and reverse-transcribed using SuperScript Reverse Transcriptase II (Invitrogen). Quantitative real-time PCR was performed with SYBR Green PCR Master Mix (Applied Biosystems). Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was used as an internal control for mRNA expression. Gene expression changes were quantified using the delta–delta CT method. Primer sequences for real-time PCR are listed in Supplementary Material, Table S11. For gene expression array analysis, 200 ng of RNA isolated from myoblasts (Day 1, 3, 8, or 15) was labeled with biotin using GeneChip 3′ IVT Express Kit (Affymetrix). Samples were hybridized on Affymetrix Human Genome U133 Plus 2.0 arrays. Expression array data were summarized with MAS5 algorithm using the Affymetrix GeneChip Command Console Software. Probes recognized as 'Absent' in myoblasts on Day 1, 3, 8 or 15 were excluded. Signal intensities of multiple probes for one gene were averaged, and fold changes in gene expression were calculated. Statistical analysis was performed using Mann–Whitney *U*-test.

DNA isolation and Illumina HumanMethylateion450 BeadChip

Myoblasts, hMSCs and human SKm_1 and SKm_2 tissues were harvested and eluted in 0.1% sodium dodecyl sulfate/Tris–EDTA buffer (pH 8.0). Cell lysates were incubated with Proteinase K (Sigma–Aldrich) and RNase A (Invitrogen) at 50°C for 2 h. Phenol/chloroform extraction and ethanol precipitation were performed to purify genomic DNA (gDNA), and precipitates were dissolved with distilled water. The concentration of gDNA was measured with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies). Bisulfite reactions were performed using the Epitect Plus DNA Bisulfite Kit (QIAGEN) with an input of 1.5 μg of gDNA. After bisulfite conversion, 300 ng of each sample was whole-genome-amplified, enzymatically fragmented and hybridized to the Illumina HumanMethylateion450 BeadChip array, which contains probes to determine the DNA methylation levels of >480 000 CpG sites in a quantitative manner. After hybridization, the BeadChip array was processed for the single-base extension reaction, stained and imaged on an Illumina iScan.

Cloned-based bisulfite sequencing

The sequences of the primers used for bisulfite-PCR reactions are listed in Supplementary Material, Table S12. Bisulfite-PCR products were cloned using the StrataClone PCR Cloning Kit (Agilent Technologies) and transformed into StrataClone Competent Cells. Single colonies were picked up and used as starting material to amplify plasmid DNA within them using the TempliPhi DNA Amplification Kit (GE Healthcare). Sequencing reactions for individual amplified clones were conducted using the BigDye Terminator version 3.1 Cycle Sequencing kit (Applied Biosystems) with the M13 Rev primer. Sequence data were obtained using ABI3130xl Genetic Analyzer and analyzed using the QUMA website (http://quma.cdb.riken.jp/).

DNA methylation data analysis

The methylation level of each of the >480 000 CpG sites was calculated using the GenomeStudio Methylation Module Software ver. 1.0 as methylation β-value ranging from 0 if completely unmethylated to 1 if completely methylated [β-value = intensity of the methylated allele/(intensity of the unmethylated allele + intensity of the methylated allele + 100)]. We excluded probes with a detection P-value of >0.05 or blank β-value from further analyses. Differences in the β-values (delta-beta, $\Delta\beta$) between target and control samples were interpreted as follows: $\Delta\beta \geq 0.2$ and $\Delta\beta \leq -0.2$ were regarded as hyper- and hypo-methylated, respectively. This is based on the fact that a $\Delta\beta$ detection sensitivity of 0.2 (95% confidence level) was previously estimated for >90% of 27 000 infinium I assay probes (32). CpG sites were categorized into seven gene feature groups or into three groups according to the probe annotation provided by Illumina. When a single CpG site was assigned to multiple gene symbols or gene features, the corresponding β-value was used for multiple times. R packages were used for data processing, statistical analysis and graphic visualization.

CGI annotation for RefSeq transcriptional start sites

Chromosomal and nucleotide positions of RefSeq transcriptional start sites were retrieved from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz. For each of the RefSeq transcripts, 100-bp DNA sequences flanking both sides of the TSS were compiled, and the resultant 201-bp sequence was assessed whether it fulfills CGI criteria of a GC content of ≥50% and an observed-to-expected ratio of CpG of ≥0.6.

GO analysis and motif discovery

Genomic regions identified as hyper- or hypo-methylated regions were tested for enrichment of GO terms using the Genomic Regions Enrichment of Annotations Tool (GREAT, http://bejerano.stanford.edu/great/public/html/) version 2.0.2 (19). Parameters were set to 'Single nearest gene, within 1000.0kb'. Binomial P-values were displayed as bar charts. Gene ontology term analyses using gene symbols were performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID; http://david.abcc.ncifcrf.gov) (33). As a cut-off for functional categories, we chose a P-value of 0.05. The motif analysis was conducted using the MEME algorithm (http://meme.nbcr.net/meme/) on the super-computing resource provided by Human Genome Center, Institute of Medical Science, University of Tokyo. Flanking sequences (from −50 to +50 bp) of hyper- or hypo-methylated CpG sites were obtained from the reference human genome sequence (hg19) using custom scripts and used as input sequences in multi-FASTA format. Parameters used were 'zero or one per sequence' for distributed, 'minimum 6 and maximum 15' for the width of the motif, and '10' for the maximum number of motifs to find (34). Motifs identified by MEME were subsequently subjected to a motif comparison tool, Tomtom (http://meme.nbcr.net/meme/cgi-bin/tomtom.cgi), for comparison with known motifs in databases. Significant thresholds were set to 'E-value < 10' and 'Human and Mouse (Jolma2013)' was selected as a database in the Tomtom analysis (35,36).

Data deposition

The data used in publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE55571 for DNA methylation array and GSE55034 for gene expression array.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Hiromi Kamura for her assistance to obtain HumanMethylation450 BeadChip data.

Conflict of Interest statement. None declared.

FUNDING

This work was supported, in part, by The Grant of National Center for Child Health and Development, (#24-3 to K.N. and #25-1 to H.A.); JSPS KAKENHI (Grant Numbers 23249071, 24115707 and 24659669 to H.A.); Japan Science and Technology Agency (Core Research for Evolutional Science and Technology to H.A.); Takeda Science Foundation, Bristol-Myers RA Research Fund (to H.A.); Mochida Memorial Foundation for Medical and Pharmaceutical Research (to H.A.); National Institute of Biomedical Innovation (to H.A.) and National Institutes of Health (AR050631 to H.A.).

REFERENCES

1. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
2. Ghosh, S., Yates, A.J., Frühwald, M.C., Miecznikowski, J.C., Plass, C. and Smiraglia, D. (2010) Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. *Epigenetics*, **5**, 527–538.
3. Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R. and Bird, A.P. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.*, **6**, e1001134.
4. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
5. Li, E., Bestorm, T.H. and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
6. Okano, M., Bellm, D.W., Haber, D.A. and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
7. Chen, J.C. and Goldhamer, D.J. (2003) Skeletal muscle stem cells. *Reprod. Biol. Endocrinol.*, **1**, 101.
8. Péault, B., Rudnicki, M., Torrente, Y., Cossu, G., Tremblay, J.P., Partridge, T., Gussoni, E., Kunkel, L.M. and Huard, J. (2007) Stem and progenitor cells in skeletal muscle development, maintenance, and therapy. *Mol. Ther.*, **15**, 867–877.
9. Dilworth, F.J. and Blais, A. (2011) Epigenetic regulation of satellite cell activation during muscle regeneration. *Stem Cell Res. Ther.*, **2**, 18.
10. Taylor, S.M. and Jones, P.A. (1979) Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell*, **17**, 771–779.

11. Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, **51**, 987–1000.
12. Palacios, D., Summerbell, D., Rigby, P.W. and Boyes, J. (2010) Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse Myogenin gene. *Mol. Cell Biol.*, **30**, 3805–3815.
13. Hupkes, M., Jonsson, M.K., Scheenen, W.J., van Rotterdam, W., Sotoca, A.M., van Someren, E.P., van der Heyden, M.A., van Veen, T.A., van Ravestein-van Os, R.I., Bauerschmidt, S. *et al.* (2011) Epigenetics: DNA demethylation promotes skeletal myotube maturation. *FASEB J.*, **25**, 3861–3872.
14. Tsumagari, K., Baribault, C., Terragni, J., Varley, K.E., Gertz, J., Pradhan, S., Badoo, M., Crain, C.M., Song, L., Crawford, G.E. *et al.* (2013) Early de novo DNA methylation and prolonged demethylation in the muscle lineage. *Epigenetics*, **8**, 317–332.
15. Bird, A. (1992) The essentials of DNA methylation. *Cell*, **70**, 5–8.
16. Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, **39**, 457–466.
17. Mofarrahi, M. and Hussain, S.N. (2011) Expression and functional roles of angiopoietin-2 in skeletal muscles. *PLoS One*, **6**, e22882.
18. Deaton, A.M. and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.*, **25**, 1010–1022.
19. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
20. Jen, Y., Weintraub, H. and Benezra, R. (1992) Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Genes Dev.*, **6**, 1466–1479.
21. Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L. and Weintraub, H. (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell*, **61**, 49–59.
22. Langlands, K., Yin, X., Anand, G. and Prochownik, E.V. (1997) Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J. Biol. Chem.*, **272**, 19785–19793.
23. Yokoyama, S., Ito, Y., Ueno-Kudoh, H., Shimizu, H., Uchibe, K., Albini, S., Mitsuoka, K., Miyaki, S., Kiso, M., Nagai, A. *et al.* (2009) A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. *Dev. Cell*, **17**, 836–848.
24. Hsieh, C.L. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol. Cell Biol.*, **14**, 5487–5494.
25. Fuks, F., Burgers, W.A., Godin, N., Kasai, M. and Kouzarides, T. (2001) Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.*, **20**, 2536–2544.
26. Mastroyiannopoulos, N.P., Nicolaou, P., Anayasa, M., Uney, J.B. and Phylactou, L.A. (2012) Down-regulation of myogenin can reverse terminal muscle cell differentiation. *PLoS One*, **7**, e29896.
27. Oikawa, Y., Omori, R., Nishii, T., Ishida, Y., Kawaichi, M. and Matsuda, E. (2011) The methyl-CpG-binding protein CIBZ suppresses myogenic differentiation by directly inhibiting myogenin expression. *Cell Res.*, **21**, 1578–1590.
28. Vinson, C. and Chatterjee, R. (2012) CG methylation. *Epigenomics*, **4**, 655–663.
29. Slieker, R.C., Bos, S.D., Goeman, J.J., Bovée, J.V., Talens, R.P., van der Breggen, R., Suchiman, H.E., Lameijer, E.W., Putter, H., van den Akker, E.B. *et al.* (2013) Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin*, **6**, 26.
30. Delatte, B. and Fuks, F. (2013) TET proteins: on the frenetic hunt for new cytosine modifications. *Brief Funct. Genomics.*, **12**, 191–204.
31. Delatte, B., Deplus, R. and Fuks, F. (2014) Playing TETris with DNA modifications. *EMBO J.*, **33**, 1198–1211.
32. Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. and Gunderson, K.L. (2009) Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*, **1**, 177–200.
33. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
34. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
35. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G. *et al.* (2013) DNA-binding specificities of human transcription factors. *Cell*, **152**, 327–339.
36. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L. and Noble, W.S. (2007) Quantifying similarity between motifs. *Genome Biol.*, **8**, R24.