Tissue-specific epigenetics in gene neighborhoods: myogenic transcription factor genes

Sruti Chandra1, Jolyon Terragni2, Guoqiang Zhang2, Sriharsa Pradhan2, Stephen Haushka3, Douglas Johnston4, Carl Baribault5, Michelle Lacey5 and Melanie Ehrlich1,*

1Program in Human Genetics, Tulane Cancer Center, and Center for Bioinformatics and Genomics, Tulane Health Sciences Center, New Orleans, LA 70122, USA, 2New England Biolabs, Ipswich, MA 01938, USA, 3Department of Biochemistry, University of Washington, Seattle, WA 98109, USA, and 4Department of Microbiology, Immunology and Parasitology, LSU Health Sciences Center, New Orleans, LA 70112, USA, 5Tulane Cancer Center and Department of Mathematics, Tulane Health Sciences Center and Tulane University, New Orleans, LA 70122, USA

*To whom correspondence should be addressed at: Human Genetics Program, SL31, Tulane Health Sciences Center, New Orleans, LA 70112, USA. Tel: +1 5049882446; Fax: +1 5049881763; Email: ehrlich@tulane.edu

Abstract

Myogenic regulatory factor (MRF) genes, MYOD1, MYOG, MYF6 and MYF5, are critical for the skeletal muscle lineage. Here, we used various epigenome profiles from human myoblasts (Mb), myotubes (Mt), muscle and diverse non-muscle samples to elucidate the involvement of multigene neighborhoods in the regulation of MRF genes. We found more far-distal enhancer chromatin associated with MRF genes in Mb and Mt than previously reported from studies in mice. For the MYF5/MYF6 gene pair, regions of Mb-associated enhancer chromatin were located throughout the adjacent 236-kb PTPRQ gene even though Mb expressed negligible amounts of PTPRQ mRNA. Some enhancer chromatin regions inside PTPRQ in Mb were also seen in PTPRQ mRNA-expressing non-myogenic cells. This suggests dual-purpose PTPRQ enhancers that upregulate expression of PTPRQ in non-myogenic cells and MYF5/MYF6 in myogenic cells. In contrast, the myogenic enhancer chromatin regions distal to MYOD1 were intergenic and up to 19 kb long. Two of them contain small, known MYOD1 enhancers, and one displayed an unusually high level of 5-hydroxymethylcytosine in a quantitative DNA hydroxymethylation assay. Unexpectedly, three regions of MYOD1-distal enhancer chromatin in Mb and Mt overlapped enhancer chromatin in umbilical vein endothelial cells, which might upregulate a distant gene (PIK3C2A). Lastly, genes surrounding MYOG were preferentially transcribed in Mt, like MYOG itself, and exhibited nearby myogenic enhancer chromatin. These neighboring chromatin regions may be enhancers acting in concert to regulate myogenic expression of multiple adjacent genes. Our findings reveal the very different and complex organization of gene neighborhoods containing closely related transcription factor genes.

Introduction

MYOD1, MYF5, MYF6 and MYOG genes, which are known as myogenic regulatory factor (MRF) genes, encode transcription factors (TFs) that are part of an intricate, extraordinarily tissue-specific transcriptional regulatory network driving skeletal muscle formation, maintaining normal muscle activity, and repairing damaged muscle (1,2). In 1987, transfected murine MyoD1 cDNA was demonstrated to convert non-myogenic cells to myogenic cells (3). The major developmental roles for the four MRF genes, their very high specificity for the skeletal muscle lineage and the ability to study them in myogenic cell cultures undergoing differentiation and in rodent embryos make them excellent models for...
analysis of development- and expression-linked epigenetic changes. MRF proteins heterodimerize with ubiquitous E-proteins and have distinct functions during embryogenesis and/or postnatally, but they are partially redundant (1,4). Transcriptional regulation of MRFs has a major role in programming myogenesis (1) although post-transcriptional regulation is also critical (5,6). There have been numerous reports about the control of expression of MRF genes prenatally and postnatally, including some epigenetic studies (7–10), but most dealt with mouse genes and none have included comparisons of the DNA and chromatin epigenetics of all four MRF genes and their broad gene neighborhoods, as done in this study for the human genes. The importance of long-range chromatin interactions in regulating gene expression has become especially clear from genomewide chromosome conformation capture analyses and related studies (11,12).

In mice, a highly conserved 258-bp sequence (core enhancer), which is ~20 kb upstream of the transcription start site (TSS) of Myod1, can confer muscle lineage-specific expression on a reporter transgene in embryos, although not later in development, and strongly upregulates reporter gene expression upon transient transfection of murine myoblasts (Mb) (13,14). One other enhancer associated with Myod1, the distal regulatory region (DDR, 714 bp), has been genetically identified and is only ~5 kb upstream of the MyoD1 TSS (14). Like the Myod1 core enhancer, the MyoD1 DDR is highly conserved between mice and humans. It can be regulated independently of the core enhancer (14).

From studies of reporter genes in transgenic mice, a complicated interdigitated set of enhancers that extend up to 140 kb from the Myf5 TSS has been described for Myf5/Myf5 and Myf6/Myf6, which are about 7 kb apart in the mouse (and human) genome (15–17). Most of these enhancers overlap the large Ptprq/Ptpnrq gene (203 and 236 kb in the mouse and human genomes, respectively), which is the nearest upstream neighbor to the MRF gene-pair (18). Distal regulatory sequences for the fourth MRF gene, Myog/Myog, have been much less examined. However, evidence of epigenetic marks indicative of enhancer activity within 7 kb of the TSS has been presented (19) and enhancer activity in a reporter gene assay was shown for sequences 4 and 1.5 kb upstream of its TSS (20). Autoregulation and cross-regulation are seen among the MRF genes, e.g. the Myod Tf (Myod1.E-protein heterodimer) is implicated in the regulation of Myog and autoregulation of Myod1, including at the core enhancer during development and in adult muscle fibers (1,21,22).

In the current study, we compared high-resolution profiles for histone modifications, DNasel hypersensitivity, binding of the chromatin-looping Ccctc binding factor (Ctcf), reduced-representation bisulfite sequencing-determined DNA methylation and RNA-seq data for Myod1, Myog, Myf5 and Myf6 and their surrounding gene neighborhoods in order to elucidate relationships between distal chromatin and MRF gene expression in the skeletal muscle lineage. Human Mb, multinucleated myotubes (Mt) derived from the Mb, skeletal muscle tissue and many non-muscle cell and tissue types were studied. We used profiles of cell-type specific H3 lysine-4 monomethylated (H3K4me1), H3 lysine-27 acetylation (H3K27ac) and DNasel hypersensitivity, which, in concert, are highly predictive of enhancers (23), to identify new candidate enhancer regions for regulating these genes. In addition, we quantified 5-hydroxymethylcytosine (5hmC) at selected sites within the vicinity of these genes using an assay that distinguishes 5-methylcytosine (5mC) and 5hmC (Epimark, 24) and address a recent study that reported prevalent differences in DNA methylation in human Mb versus Mt (25). Our analyses have uncovered evidence for much more extensive enhancer regions distal to the Myod1 and Myog in both Mb and Mt and for Myf5 and Myf6 in Mb than previously reported, for sharing parts of these regions among myogenic and specific non-myogenic cell types, and for unusually high levels of genomic 5hmC in Mb and Mt at enhancer chromatin far upstream of Myod1.

Results

Promoter-type chromatin covering the bodies of MRF genes

Our RNA-seq profiling of Mb and Mt (26) indicated that Myod1, Myf5, Myf6 and Myog were expressed specifically and at high levels in Mb and/or Mt, as expected. RNA-seq FPKM (fragments per kilobase of exon per million fragments mapped) for Mb were 80, 307, 44 and 76, respectively, and for Mt: 36, 9, 24 and 421, respectively. We compared the chromatin epigenetics of these highly expressed MRF genes within their gene bodies and in immediately adjacent regions using genome-wide profiles of predicted enhancer states based upon histone modifications (ENCODᵉ/Cromatin state segmentation/Broad Institute) (27,28). These genes displayed active canonical promoter-type chromatin (strong H3K4me3 and H3K27ac) along their gene bodies. Myod1, Myf6 and Myog exhibited weaker signals for H3K4me3 immediately upstream of their TSS than within the body of these genes in Mb and Mt, and this was also true for the H3K27ac and H3K4me2 signals at Myod1 and Myf6 (Fig. 1; Supplementary Material, Figs. S1–S3). Extensive H3K4me3 signal in the body of genes was previously noted in a minority of actively transcribed genes and overrepresented among cell-identity determining genes, like MRF genes, by Benayoun et al. (30). Moreover, H3K4me3 enrichment can peak immediately downstream of the TSS (31), and the MRF genes consist of only 1.8–2.9 kb. Furthermore, upstream H3K4me3 may not be indispensable for transcription of all eukaryotic genes (32). In the examined non-myoicgenic cell types, repressive H3K27me3 chromatin was observed within Myod1 and its upstream region (Figs 2 and 3). Within Myog, Myf5 and Myf6, H3K27me3-enriched chromatin was seen in only some of the non-myoicgenic cell cultures (Figs 4 and 5).

The observed changes in steady-state RNA levels in Mt versus Mb for Myod1, Myf5, Myf6 and Myog (Mt/Mb RNA-seq FPKM ratios of 0.45, 0.03, 0.55 and 5.5, respectively) are consistent with the intragenic signal for H3K27ac or H3K4me3 at Myod1, Myf5 and Myf6 being lower in Mt versus Mb while that for Myog was much higher in Mt (Supplementary Material, Figs S1E, S2G and S3F, pink highlighting). A distinct feature of transcription in the immediate vicinity of Myod1 in RNA-seq profiles was an unusually long region of sense-strand transcription past the 3′ end of the gene (Fig. 3A; Supplementary Material, Fig. S4A and D). This extended transcription was seen almost up to the strong constitutive Ctcf binding site ~8 kb downstream of the end of the RefSeq gene (Supplementary Material, Fig. S1E), although the extension of the Myod1 transcript had a much lower signal than the RNA over the gene itself.

A hypomethylated myogenic DMR and lincRNA gene 5 kb from the Myod1 core enhancer

A human sequence with 92% identity to the genetically determined murine Myod1 258-bp core enhancer (13) is located 20 kb upstream of Myod1 (Fig. 2C) and similarly situated in the mouse genome. The gene neighborhoods of Myod1 and the
The DNA methylation analysis came from reduced representation bisulfite sequencing (RRBS, 37) profiling of Mb and Mt versus many types of non-myoogenic cell cultures and skeletal muscle tissue versus diverse non-muscle tissues in which only methylation at CpG sites, the predominant site of DNA methylation, as in the study of Miyata et al. (25). This is unlike the recent finding of Miyata et al. (25) on commercially derived cell cultures, which have not been examined to determine the percentages of contaminating fibroblasts and conversion of Mt to Mb. Immunocytochemical characterization of batches of Mb and Mt is important because contaminating fibroblasts can overtake the cultures with continued incubation and conditions need to be optimized for maximal Mt formation from Mb. Determining relative amounts of myoblasts, fibroblasts and myotubes cannot be done by phase microscopy. Increasing levels of fibroblast contamination and inefficient Mt formation can be problems when inducing Mt formation from low-density Mb cultures and employing unusually prolonged incubation times for differentiation, as in the study of Miyata et al. (25). This is a major concern in studies relying on average signal levels, such as DNA methylation profiling, and the problem is exacerbated when comparing Mt and Mb cultures.

The MbMt- and muscle-hypomethylated DMR that we newly identified (Fig. 3B) is located within a very large, myogenesis-specific enhancer chromatin (EnhC) region (19 kb wide) centered 28 kb upstream of MYOD1 (~28 EnhC). This DMR overlapped an Mb-specific DnaseI hypersensitive site (DHS; Fig. 3D, blue arrow, and Supplementary Material, Fig. S4G) and a sequence orthologous to a MyoD binding site detected in ChIP-seq profiles of murine C2C12 Mt (Supplementary Material, Figs S4C and S5C). It was ~0.15 kb upstream of the 5′ end of a gene encoding a myogenesis-associated 0.6-kb long non-coding (linc) RNA that corresponded to a rhabdomyosarcoma-derived ENSEMBL transcript (ENST00000524885; Fig. 2A). In the upstream intergenic region of MYOD1, the only substantial poly(A)+ RNA-seq signal of >0.2 kb was this transcript (Fig. 3A, Supplementary Material, Fig. S4D). Analysis of total RNA by RNA-seq revealed only one other Mb-specific RNA transcript, a ~0.2 kb RNA, that overlapped the region of ~45 EnhC (data not shown, ENCODE/Cold Spring Harbor Laboratories; 27).

Previously, hypomethylation of the MyoD1 core enhancer DNA (38,39) and also in the Myog promoter region (9,40) was described in the skeletal muscle lineage. However, RRBS analysis gave no data in these regions probably because of their low CpG density, which biases against RRBS detection (37). Good coverage of the CpG island (CGI) overlapping the MYOD1 promoter region and gene body was obtained in the RRBS profile, and indicated that this CGI was predominantly unmethylated in all samples (Fig. 3B). In addition, there was MbMt hypermethylation of the first exon of KCNC1, a potassium-channel gene preferentially
expressed in the brain (41) and the nearest downstream neighbor of MYOD1.

**The 19-kb enhancer chromatin upstream of MYOD1**

The very large, partly conserved (Supplementary Material, Figs S1D and S4E) region of −28 EnhC mentioned earlier was observed in Mb and Mt (Supplementary Material, Fig. S1E, H3K4me1 and H3K27ac) within the 74 kb of intergenic DNA between MYOD1 and the inner-ear specific OTOG (Fig. 3A). This EnhC was not seen in a lymphoblastoid cell line (LCL), normal human mammary epithelial cells (HMEC), embryonic stem cells (ESC) or normal human epidermal keratinocytes (NHEK; Figs 2C and 3C). Ten kilobases of this 19-kb EnhC region is orthologous to enhancer-like sequences mentioned in a report by Blum et al. (10) about genomic profiles for H3K4me1, H3K27ac, PolII binding and p300 binding from murine C2C12 Mb and Mt (Supplementary Material, Fig. S1B, ChIP-seq). Throughout much of the −28 EnhC region in the human genome, we observed DHS specifically in adult-derived or fetal myogenic cultures and DNA sequences orthologous to high-affinity C2C12 murine Mb or Mt MyoD binding sites (Fig. 3D; Supplementary Material, Fig. S1D). The reduction in DHS peak heights in adult-derived Mt versus Mb in much of this EnhC region might reflect the lower MYOD1 RNA levels in Mt. Only one non-myogenic cell culture, human umbilical vein endothelial cells (HUVEC; discussed subsequently), displayed a small region of EnhC and a strong DHS overlapping the −28 EnhC of Mb. This limited overlap was only at the MYOD1-distal end of the −28 EnhC (Fig. 3C; Supplementary Material, Fig. S4). In contrast, Taberlay et al. (42) found that nucleosome depletion at the core enhancer, which is in the MYOD1-proximal end of −28 EnhC, was similar in a MYOD1-expressing rhabdomyosarcoma cell culture and a MYOD1-non-expressing fetal bladder fibroblast cell line. The lack of specificity for open chromatin at the core enhancer in the study of Taberlay et al. compared with the myogenic specificity for open chromatin in the same region in this study could be due to different experimental techniques or types of examined myogenic and fibroblast cell cultures.

Most of the −18 kb immediately upstream of the MYOD1 TSS displayed histone marks typical of weakly transcribed chromatin unlike the super-enhancer designation by Whyte et al. (43) based upon dense MyoD binding to orthologous DNA sequences in mouse C2C12 Mt (Supplementary Material, Fig. S1B). There was one prominent region of EnhC (1.8 kb wide) proximal to MYOD1−5 EnhC, which contains the human ortholog to the mouse 714-bp DRR enhancer (44). This −5 EnhC region overlapped one of only three CTCF binding in the vicinity of MYOD1 that displayed preferential binding in myogenic cells (Supplementary Material, Fig. S1E, boxes). These sites are not predicted to function as insulators because they overlap active enhancer- or promoter-type chromatin, as do many CTCF sites (28). They might be important in organizing the topological domain in which MYOD1 resides.

In the MYOD1 intergenic region close to OTOG in both Mb and Mt, there was the −67 EnhC (6 kb wide), which overlapped a cluster of Mb- and Mt-specific DHS peaks (Fig. 3). Enhancer cluster was identified in the orthologous region in C2C12 Mt by Blum et al. (10) (Supplementary Material, Fig. S1B, ChIP-seq) as were C2C12 Mb- and Mt-deduced (8) MYOD-binding sites (Supplementary Material, Fig. S1D). In contrast, the −45 EnhC (2.2 kb wide), which was not observed by Blum et al., overlapped only a small DHS in Mb and Mt that was not specific for myogenic cells and did not overlap any predicted MYOD binding sites (Fig. 3D; Supplementary Material, Fig. S1D). In skeletal muscle tissue, DNaseI- and histone modification profiles of skeletal muscle tissue did not reveal −67 and −45 EnhC regions although evidence for −28 and −5 EnhC regions was seen (Fig. 3D; Supplementary Material, Fig. S6D). The −5 EnhC, which contains the DRR enhancer and a strong Mb- and Mt-specific DHS peak, was also seen as a strong DHS in skeletal muscle tissue while the core enhancer region’s DHS was diminished in skeletal muscle (Fig. 3D,
highlighted peaks). These findings are consistent with lower but appreciable expression levels of MYOD1 in postnatal muscle tissue (41), where the DRR is important for MyoD1/MYOD1 expression to help prevent atrophy and facilitate muscle repair (21,45). A muscle tissue-specific region displaying active promoter-like chromatin was seen from 0.2–0.8 kb upstream of
MYOD1 (Supplementary Material, Fig. S6D, arrow) that was more prominent than in Mb or Mt (Supplementary Material, Fig. S1E). This suggests a previously undescribed cis-acting transcription control element for MYOD1.

Despite the strong association of the epigenetic marks in the MYOD1 upstream regions with the skeletal muscle lineage, HUVEC displayed partial overlap of DHS and EnhC with those of Mb, LCL, normal human lung fibroblasts (NHLF) and HMEC samples was PIK3C2A (RNA-seq, ENCODE/CalTech and ENCODE/Cold Spring Harbor Laboratories) confirming Mb-specific antisense RNA signals (blue and green arrows). Lung fibroblasts in (C) are IMR-90, a fetal lung fibroblast cell line, and in (B) are adult lung fibroblasts. (D) DNase-seq and (E) one RRBS track, as in Fig. 3. For orientation, pink bars above tracks in (C) and (D) indicate the position of the MYF5/MYF6 gene-pair and intervening sequence.
Cufflinks analysis, data not shown). PIK3C2A is located 0.5 Mb from MYOD1 and encodes a phosphoinositide 3-kinase catalytic subunit. Another example of EnhC in non-muscle cells overlapping myogenesis-related EnhC in the neighborhood of MYOD1 involves the MYOD1 core enhancer. In the K562 leukemia cell line, both EnhC and a DHS were centered over the known core enhancer region (Supplementary Material, Fig. S7A), although not over the DRR enhancer (Supplementary Material, Fig. S7B). In addition, FOS- and JUN-binding peaks (subunits of the AP-1 TF) were centered over this small enhancer region in K562 cells (Supplementary Material, Fig. S7A) as well as being present in the −67, −45 and −28 EnhC regions in HUVEC (ENCODE/TF ChIP-seq, data not shown) (27). Comparable TF ChIP-seq data are not available for Mb or Mt for comparison but, nonetheless, the ubiquitous...
presence of AP-1 can help explain these exceptions to the strict myogenic specificity of EnhC located between MYOD1, a skeletal muscle lineage-specific gene and OTOG, an ear-specific gene.

**MYF5 and MYF6: myogenic enhancer chromatin in adjacent genes**

MYF5 and MYF6 are close to each other in both the human and mouse genomes and share much human/mouse sequence similarity in their upstream regions (Supplemental Material, Fig. S8A). This sequence similarity includes regions where distal and far-distal enhancers (18) have been described in the mouse within the large neighboring Ptprq gene, which encodes a tyrosine phosphatase that is expressed at its highest levels in fetal kidney, lung and cochlea (41). Human Mb in the MYF5/MYF6 neighborhood exhibited scattered regions of EnhC, mostly in introns of PTPRQ, and extending up to the 5′ end of this 236-kb gene (Fig. 4B). Previously, mouse intragenic Ptprq enhancers active on Myf5 and Myf6 had been described up to ~140 kb upstream of Myf5 (equivalent to ~150 kb upstream of the human MYF5). The PTPRQ intragenic regions of EnhC included human orthologs to previously described myogenic enhancers, such as the 70-bp enhancer element associated with expression of Myf5 in adult quiescent satellite cells (16) and a 145-bp enhancer element linked to expression of Myf5 during embryological hypaxial muscle formation in the mouse (Fig. 4, SE and HE, respectively). Upstream of the MYF5/MYF6 gene-pair in Mb were also several regions consisting of interspersed enhancer-type and active-promoter chromatin (Fig. 4B, red and yellow bars). Many of the regions of PTPRQ-intragenic EnhC were seen not only in Mb but also in NHLF (Fig. 4B, black-outlined yellow or orange regions in the NHLF track), osteoblasts and skin fibroblasts (Supplemental Material, Fig. S2G). In contrast, most of these PTPRQ-intragenic EnhC regions were absent in HUVEC, HMEC, NHK, an LCL and ESC (Fig. 4B; Supplemental Material, Fig. S2G).

The neighborhood of MYF5/MYF6 exhibited only two C2C12-deduced MYOD binding sites, both of which overlapped EnhC seen preferentially in Mb (Supplemental Material, Fig. S2F and I, arrows). The substantial DHS peaks seen in this neighborhood that were specific for myogenic cells were in the immediate vicinity of MYF5/MYF6. Other DHS seen in Mb were either very small or not specific for myogenic cells (Fig. 4D). A MbMt-hypomethylated DMR was observed 0.2 kb downstream of the MYF6 TSS (Fig. 4A). The CTCF sites mapped in Mb in the MYF5/MYF6 neighborhood were not specific for myogenic cells with the exception of a very weak MYF6-proximal site seen in Mb and Mt (Supplemental Material, Fig. S2G, arrow), which like some of the other CTCF sites, overlapped enhancer or promoter chromatin.

RNA-seq data indicated that human Mb samples did not display detectable mRNA from the whole PTPRQ gene (Supplemental Material, Fig. S2D). NHLF, osteoblasts and skin fibroblasts, which exhibited multiple intragenic regions of EnhC within the gene, expressed considerable levels of this mRNA, unlike the five non-muscle cell types that displayed little EnhC overlapping PTPRQ (Fig. 4B, C; Supplemental Material, Fig. S2). Mb samples did generate ncRNA from this gene region with a unique pattern of transcripts relative to12 disparate types of non-myogenic cell cultures, according to two independent sets of ENCODE RNA-seq profiles (Fig. 4C, layered RNA-seq profiles and strand-specific RNA-seq). Only Mb exhibited an asymmetrical distribution of total RNA-seq signal along the introns in the large PTPRQ gene such that the signal was mostly plus-strand RNA from the MYF5/MYF6-proximal half of PTPRQ and minus-strand RNA from the MYF5/MYF6-distal half (whole-cell total RNA-seq, Supplementary Material, Fig. S2C, boxes). Two antisense RNA peaks and one sense RNA peak overlapped regions of active-promoter chromatin or EnhC observed preferentially in Mb in the middle or 3′ end of PTPRQ (Fig. 4B and C, long arrows). Genomewide analysis of 5′ mRNA caps (ENCODE/CAGE/Riken Omics) further confirmed the Mb-specific production of PTPRQ-intragenic ncRNA and indicated multiple discrete ends of RNA molecules within PTPRQ preferentially in Mb, including for the afore-mentioned three RNA peaks (Supplementary Material, Fig. S2B, box).

**MYOG and surrounding genes: all preferentially expressed in Mt**

Human Mb lacked repressive chromatin in the vicinity of MYOG (Fig. 5) analogous to that observed immediately upstream of Myog in the immortal murine C2C12 Mb cell line, but not in C2C12 Mt preparations (20). Accordingly, unlike C2C12 Mb, human Mb expressed this gene abundantly (Fig. 1C) although at a lower level than in Mt. Upstream of MYOG are ADORA1, MYBPP and CHI3L1 and downstream is PPFIA4. All these MYOG neighbors were preferentially expressed in Mb compared with non-myogenic cell cultures (Fig. 5A) and at even higher (3–4 times) levels in Mt than in Mb (RNA-seq data not shown). The expression data for MYBPP are consistent with its encoding a myofibril structural protein (46). However, levels of CHI3L1 RNA, which codes for a TNFα-antagonistic glycoprotein, were previously reported to decrease upon differentiation of primary human Mb (47). CHI3L1 and the adenosine receptor ADORA1 have been implicated in skeletal muscle repair or protection from injury (47,48). PPFIA4, encodes a little-studied protein tyrosine phosphatase-interacting protein, that has not been reported to have a specific role in the skeletal muscle lineage, although it is preferentially expressed in skeletal muscle, cardiac muscle and brain (49). In addition to RNA-seq analysis, promoter and enhancer chromatin profiling (Supplementary Material, Fig. S3B and F) also indicated its preferential expression in myogenic cells.

There was Mb- and Mt-associated EnhC centered 5, 14 and 17 kb upstream of MYOG that was closer to MYOG than to other genes (Fig. 5C; Supplemental Material, Fig. S3F). The MYOG –5 EnhC overlapped a cluster of myogenic DHS peaks (Fig. 5D) and a myogenic-specific CTCF binding site (Supplemental Material, Fig. S3F). The –14 and –17 EnhC regions showed only barely visible DHS peaks (Fig. 5D). All three of these chromatin regions displayed stronger enhancer-like histone modifications in Mt than in Mb, which is consistent with upregulation of MYOG in Mt at the transcriptional level (Supplemental Material, Fig. S3F, H3K4me3, H3K27ac and H3K79me2). The MYOG –5 and –14 EnhC regions as well as many other myogenic EnhC regions in this gene neighborhood had sequences orthologous to C2C12 Mb or Mt MYOD-binding sites (Supplemental Material, Fig. S3C). One of the myogenic DHS in the middle of the MYOG-downstream PFFIA4 gene exhibited MbMt- and muscle-specific DNA hypomethylation (Fig. 5B and D). An ESC-specific DHS overlapped MbMt hypermethylation (Fig. 5B and D). In skeletal muscle tissue, histone modification (H3K36me3 and H3K4me3) indicated that in the MYOG was the most transcriptionally active gene in its neighborhood (Fig. 5D and E) while analogous profiling of Mb and Mt showed that MYOG neighbors were also transcriptionally active at the myogenic progenitor stage (Supplemental Material, Fig. S3F). Similar profiles indicate that in their respective gene neighborhoods in skeletal muscle, only MYOD1 (Supplemental Material, Fig. S6D) and MYF6 (Supplemental Material, Fig. S8C), but not MYF5, were transcribed in the examined muscle tissue. A caveat for these conclusions about skeletal muscle is...
that there can be skeletal muscle subtype differences in gene expression (50), and the skeletal muscle subtype is unknown for the ENCODE data other than for the DNase-seq profiling (psosas muscle).

DNA hydroxymethylation in MYOD1-distal regions in both myogenic and non-myogenic samples

Because RRBS detects both ShmC and SmC as ‘methylated’, we distinguished between these bases and quantified them at selected CCGG sites with a 5mC and ShmC assay (Epimark, New England BioLabs). This assay involves T4 phage (β-glucosyltransferase, restriction endonucleases and six quantitative real-time PCR determinations per sample at a given CCGG site (24). Two to seven independent biological replicates of skeletal muscle, myogenic progenitor cell strains and diverse samples of non-myogenic cells and tissues were compared at five different sites. As expected, the Epimark assays confirmed the RRBS-determined hypomethylation at the site 25 kb upstream of MYOD1 and 236 bp downstream of the MYF6 TSS. For example, seven Mb or Mt samples from different cell strains had 0–36% 5mC while six skin fibroblast cell strains had 93–99% 5mC and four HUVEC cell strains had 95–99% 5mC at the MYOD1 –25 kb site. Surprisingly, with respect to ShmC levels, a comparison of MYOD1 –9, –25, –45.5 and –48 kb sites and the MYF6 +238 bp site, revealed that only the MYOD1 –25 and –45.5 kb sites had >1% ShmC in Mb and Mt (Fig. 6A). Both of these sites were also the only ones in this study that overlapped EnhC in Mb. In eight assayed Mb and Mt samples at the MYOD1 –25 kb site, on average, 32% of C was ShmC (range, 12–42%; Fig. 6A, asterisk). In contrast, the three analyzed muscle samples had an average of only 0.6% ShmC (range, 0–1.7%). RRBS revealed that although this site was hypomethylated in skeletal muscle lineage samples versus other samples, it still retained substantial C modification. The enzymatic assay indicated this partial modification of C was actually a mixture of ShmC and SmC for Mb and Mt, but almost exclusively 5mC for skeletal muscle (Fig. 6A and B). Cerebellum had more ShmC than skeletal muscle at the MYOD1 –25 and –9 kb sites with ratios of ShmC to SmC of about one compared with adult brain ShmC sites that have on average about twice as much 5mC as ShmC (51). The high level of ShmC at these two sites in cerebellum should not be due to regulation of MYOD1 expression in that tissue because transcription of this gene is almost completely specific for the skeletal muscle lineage (34). However, it might reflect poised EnhC in cerebellum that regulates expression of the neighboring KCNC1 gene, which is transcribed selectively in some tissues (41) including cerebellum (52). Alternatively, it might simply be due to brain tissue having very high levels of genomic ShmC relative to other tissues (53,54) resulting in an elevated ShmC content at many transcription-control sites irrelevant to the neural lineage.

Discussion

Gene deserts can provide a buffer around key developmental genes protecting them from heterologous enhancers and promoters and can harbor far distal enhancers targeted to such genes. This is especially advantageous for small genes. However, none of the functionally and structurally related MRF genes (MYOD1, MYOG, MYF5, and MYF6) is bordered by a gene desert (>0.5 Mb free of canonical genes) (55) even though MRF genes are all small, highly expressed and differentiation-determinants. These genes, which have little room in their two undersized introns for intragenic enhancers, use various strategies for ensuring highly cell-type specific, high-level transcription. MYOD1, while not adjacent to a true gene desert, was next to a 74-kb upstream region devoid of mRNA-encoding genes but populated with four regions of myogenic EnhC, two of which have within them known <1-kb MYOD1-specific enhancers. In many regions in the MYOD1 and MYOG neighborhoods, there was myogenic EnhC overlapping Mb- and Mt-specific DHS peaks and murine-inferred binding sites for the

![Figure 6](https://academic.oup.com/hmg/article-abstract/24/16/4660/746396/746396)

Figure 6: Quantification by enzymatic assay of ShmC and SmC at selected CCGG sites in the MYOD1 neighborhood and in MYF6. The average percentages of ShmC (A) and SmC (B) at the indicated CCGG sites in biological replicate samples were determined by an enzymatic and real-time PCR assay. The positions of the sites are given relative to the TSS of the respective gene and shown in Fig. 2 and Supplementary Material, Figure S6A. Note the different scale for ShmC and SmC. EnhC, the studied site overlapped EnhC in Mb; non-enhancer, the site did not overlap EnhC. Asterisk, highly significantly elevated ShmC levels in the set of Mb and Mt relative to 49 other examined sites in various genomic locations, as described in the text.
MYOD TF. In its 155-kb neighborhood, MYOG, unlike MYOD1 and the MYF5/MYF6 gene-pair, has upstream and downstream gene neighbors that are preferentially expressed in myogenic progenitor cells. Concomitantly, much of the myogenesis-linked EnhC and many overlapping myogenic DHS peaks were near or within these MYOG-neighboring genes. In contrast to the MYOG and MYOD1 neighborhoods, the MYF5/MYF6 neighborhood generated many myogenesis-specific ncRNAs of >2 kb in both the sense and antisense orientations. These originated throughout the 236-kb gene body of FPTPRQ, a MYF6-adjacent non-myogenic gene, which also had multiple regions of Mb-associated EnhC from its 5′ to its 3′ end. Many of these EnhC regions in the 3′ half of FPTPRQ contain known development-specific enhancers of MYF5 and/or MYF6 (56). However, there was little myogenesis-specific DNaseI hypersensitivity and only a couple of mouse-deduced MYOD binding sites in the MYF5/MYF6 neighborhood of Mb and Mt. Therefore, the three MRF gene neighborhoods use very different strategies for far-distal cis-acting transcription control.

Although most of the intergenic EnhC found up to 71 kb upstream of MYOD1 was specific for Mb and Mt, there were short regions within the MYOD1 −67, −45 and −28 kb EnhC that were also observed in HUVEC. These regions might upregulate the expression of the 0.5 Mb-distant PIK3C2A specifically in HUVEC through the type of very long-distance interactions described for other genes (11, 12). PIK3C2A has an essential role in endothelial cells (57) and is the closest gene to MYOD1 with preferential expression in HUVEC. This might, therefore, represent an example of enhancer dual-purposing such that the enhancer upregulates a gene in one cell type (MYOD1 in Mb and Mt) and a very different type of gene in a specific heterologous cell type (PIK3C2A in HUVEC).

Related to this observation was the finding that within the critically important, 258-bp MYOD1 core enhancer (34) in the 19-kb wide MYOD1 −25 EnhC, there was a short region of overlap between EnhC in K562 cancer cells (but not in seven other types of non-myogenic cell cultures) and EnhC in Mb. In this case, only a small region of EnhC overlapping the known MYOD1 core enhancer was shared by the two cell types, and there were no apparent K562-upregulated genes even in the distant neighborhood. Nonetheless, for this example, there is TF binding data that illustrates how cell type-specific enhancers might be shared by dissimilar cell types for different purposes. The small EnhC in K562 cells and the accompanying specific DHS were centered over a binding site for AP-1 TF subunits, as seen in ChIP-seq profiles (Supplementary Material, Fig. S7A). Although there are few TF ChIP-seq Mb databases available for comparison to K562, individual enhancer analyses provide evidence that AP-1 is involved in upregulation of MYOD1/MyoD1 by the core enhancer in myogenic cells (13, 58). AP-1, like MYOD, can be found at enhancers and act combinatorially with other transcription factors, including MYOD (10), to generate enhancers (8, 59, 60). The lack of binding of MYOD protein and certain other development-associated proteins (22) at the core enhancer in K562 cells could explain why this region of EnhC in K562 might not be acting as an enhancer, as it does in Mb and Mt.

Although ncRNAs play major roles in enhancer function and myogenesis (35, 61), only two myogenesis-associated ncRNAs (MYOD1 −25 lincRNA, ENST00000524885, >0.2 kb and one unnamed RNA <0.2 kb at the MYOD1 −45 EnhC) were identified by poly(A)+ or total RNA-seq (ENCODE, 27) at EnhC in the MYOD1 neighborhood. Other enhancer ncRNAs may not have been observed because of their generally very low steady-state concentrations, including for a previously detected MYOD1 core enhancer ncRNA transcribed 5 kb downstream of the MYOD1 −25 lincRNA (35). In the presumed promoter region of the gene encoding the −25 lincRNA were two CpG sites that were hypomethylated in Mb, Mt and skeletal muscle tissue versus non-muscle samples. A quantitative enzymatic assay applied to one of these sites (CCGG) revealed that in Mb and Mt, the CpG was partially modified. At this site, there were equal amounts of 5hmC and 5mC in the persisting, modified C residues (average of 27% 5hmC and 27% 5mC in the set of Mb and Mt). We compared the average 5hmC content of this site with that of 49 other CpG sites in various gene regions analyzed because of overlapping or nearby myogenesis-associated DNAs. The MYOD1 −25 kb site was unusual in its moderately high 5hmC content in myogenic progenitor cells [Fig. 6, (24, 26, 62) and unpublished data]. In pairwise comparisons, only this site and one site upstream of another gene (2NF556) had significantly higher average 5hmC levels in Mb and Mt than did most other sites (adjusted P < 0.0001). Although global genomic levels of 5hmC in Mb and Mt are about half of that in skeletal muscle (24), skeletal muscle had negligible amounts of this modified base at the MYOD1 −25 kb site. 5hmC is associated with tissue-specific, active or poised enhancers and also is overrepresented in the bodies of very highly transcribed genes (63, 64). The unusual enrichment in 5hmC in myogenic progenitor cells immediately upstream of the MYOD1 −25 lincRNA gene may indicate a role for this lincRNA in upregulating MYOD1 especially in Mb, which have the highest level of expression of MYOD1 and display a DHS at the −25 kb site.

The myogenesis-specific epigenetic features distal to MYOD1 strongly suggest that regulatory elements for MYOD1 in Mb and Mt extend far beyond the previously analyzed 256-bp core enhancer and 714-bp DRR enhancer 20 and 5 kb upstream of MYOD1, respectively (14, 34). Not all of these myogenic EnhC regions may function as canonical enhancers. For example, although −45 EnhC encoded a <0.2-kb RNA and displayed an elevated 5hmC content, it overlapped only a very small nonspecific DHS in Mb, unlike typical enhancers (65). In contrast, MYOD1 −5, −28 and −67 kb EnhC, including the sites of the known core and DRR enhancers, exhibited substantial DHS peaks with strong myogenic specificity. In adult skeletal muscle tissue, the importance of the DRR enhancer 5 kb upstream of MYOD1/MyoD1 (14) was consistent with histone modification and DHS profiles (Fig. 3; Supplementary Material, Fig. S6D). In addition, a strong promoter-like signature (or possibly atypical enhancer-like signature) (66) was seen immediately upstream of the gene, specifically in muscle tissue but not in myogenic progenitor cells that had not been reported earlier.

The epigenetic features that we describe in this study suggest many more distal cis-acting regulatory sites for MRF genes than previously reported. The 24 kb of DNA upstream of the human MYOD1 gene, which contains the DRR and core enhancers, sufficed to drive expression of a reporter gene in transgenic mice in a muscle lineage-specific manner during prenatal development and in adult mice (21, 34). However, there were some differences in the timing of transgene expression in independent transgenic mice embryos versus endogenous MyoD1 expression within certain skeletal muscle lineages, and it was proposed that additional, unelucidated enhancer elements fine-tune expression of MyoD1 (21, 34). When extrapolating from studies of these transgenes to the endogenous genes, their unknown chromosomal environment and their multiple copies in many of the examined mouse embryos need to be considered (34). In the context of MYOD1’s natural chromatin neighborhood in Mb and Mt, the large regions of myogenic EnhC upstream of the gene and the low-level transcription throughout the 9-kb MYOD1-downstream region in concert with the core and DRR
enhancers might be necessary to optimize MYOD1/MyoD1 transcrip-
tion levels, especially during postnatal muscle repair. The adult-derived Mb and Mt samples in this study mimic intermediates in such repair. In addition, these large EnhC regions might help suppress inappropriate expression of neighboring genes through effects on higher-order chromatin structure. The need for myogenesis-specific repression of non-myogenic neighboring genes is suggested by the MbMt hypermethylation in exon 1 of KCNC1, which is silent in all examined cell types and was associated with generally repressive H3K27me3 in non-myogenic cell types but not in Mb.

For MYOG, there has been little discussion of possible distal enhancers in previous reports other than the finding of more H3K4me1 throughout the 17 kb upstream of Myog in C12C12 Mt versus Mb (19,20) and the demonstration by Asp et al. (20) of transfection-assessed enhancer activity for <1-kb regions situated 1.5 and 4 kb upstream of the murine Myog TSS. The immediate upstream promoter region of murine Myog (TSS −0.2 kb) suffices to confer skeletal muscle-specific expression on a reporter gene in transgenic mice or Mt in culture. This differs from the promoter region of MyoD1, which by itself has little activity and no muscle-lineage specificity (13,67). We observed EnhC associated with Mb and especially Mt at many regions within the MYOG neighborhood, including at −5, −14 and −17 EnhC. The myogenesis-specific DHS and a myogenic CTCF binding site at the MYOG−5 EnhC are consistent with the findings of Asp et al. (20) and the enrichment of tissue-specific CTCF binding sites at tissue-specific enhancers (68). In contrast, MYOG−14 and −17 EnhC regions displayed only tiny DHS peaks and thus may not have enhancer activity in Mb and Mt. Alternatively, they might be weak enhancers working in concert with other stronger enhancers, including those closer to other genes in the MYOG neighborhood that are preferentially expressed in myogenic progenitors, especially in Mt. We propose that Mt-associated enhancers throughout this gene neighborhood encompassing MYOG, PPFAIA4, ADORA1, MYBP and CH3L1 act in concert to establish a topological domain that favors their expression in a highly differentiation-specific manner, such as seen in subdomains of HOX gene clusters (11).

For the murine Myf5/Myf5 gene pair, it has been demonstrated that the large adjacent Ptprq gene harbors interdigitated enhancers, some of which act in concert, and function in specific skeletal muscle sublineages and/or developmental stages to regulate Myf5, Myf6 or both (18). Our finding of many EnhC regions in the similar human PTPRQ gene in Mb are consistent with the proven enhancer function for the murine Myf5 and Myf6 genes at many subregions within the adjacent Ptprq gene, as determined by transgenic mouse studies, and help explain these surprising results. As for Ptprq in mouse Mb (69), we observed negligible expression of PTPRQ mRNA in human Mb. In the current study, we report remarkable bidirectional expression of ncRNA from PTPRQ specifically in Mb. Bidirectional transcription is a frequent characteristic of enhancers (70). Moreover, in non-myogenic cells, we observed an association of fibroblast and osteoblast PTPRQ mRNA expression with intragenic PTPRQ EnhC regions. Those EnhC regions in PTPRQ in fibroblasts and osteoblasts often overlapped EnhC regions in Mb despite the lack of PTPRQ mRNA in Mb. Of the eight examined non-myogenic cell types, only NHLF, skin fibroblasts, and, especially, osteoblasts displayed this myogenic/non-myogenic EnhC overlap in the gene body of PTPRQ. These were the only examined cell types with evidence for appreciable PTPRQ mRNA synthesis. Such observations suggest that the intragenic PTPRQ EnhC in NHLF, skin fibroblasts and osteoblasts upregulates transcription of PTPRQ mRNA in these non-myogenic cells, while the partly overlapping but larger set of Mb-associated EnhC regions within PTPRQ introns together with EnhC closer to MYF5 and MYF5 upregulates the adjacent MYF5 or MYF6 genes in Mb. This appears to be another example of dual-purpose enhancers and is reminiscent of enhancers in exons of one gene upregulating transcription of nearby genes (71).

In conclusion, this study of the epigenetics and transcriptome in the neighborhoods of the four related human MRF genes demonstrates the very different organization of gene neighborhoods containing structurally and functionally related differentiation-determining genes. These differences include far-distal intergenic enhancers (the MYOD1 neighborhood), neighboring genes having similar tissue specificity and possibly sharing enhancers (the MYOG neighborhood) or enhancers within neighboring genes that have different tissue specificity and nonetheless possibly share enhancers (the MYF5 and MYF6 gene neighborhood). A comparison of various DNA and chromatin epigenetic profiles in specific regions for a wide variety of cell types has revealed evidence for highly nuanced epigenetics, such as enhancers in PTPRQ that may upregulate expression of PTPRQ in certain non-myogenic cell types and of MYF5/MYF6 in myogenic cells. The type of complicated and multilayered epigenetics of the MRF gene neighborhoods that we have described may be necessary for differential control of gene expression to establish proper spatiotemporal patterns during development, efficient repair after muscle damage and appropriate responses to differing physiological demands (7,34,45,50).

Materials and Methods
DNA samples, RRBS data and DHS profiles were obtained as previously described (24,36). Quality control of Mb (70% confluent) and Mt samples used for RRBS, DHS and 5hmC/5mC assays was done by desmin immunostaining as previously described (24,72). More than 90% of the cells in Mb preparations used for analysis were Mb, and >70% of nuclei in Mt preparations were in multinucleated cells. Mt were obtained by initiation of differentiation by serum deprivation in medium with 2% horse serum for 1 day followed by 3-4 days of incubation in medium containing 15% horse serum (72). Two fetal Mb cell strains (H246 and H275) used just for DNase-seq were derived by one of the authors (SH) in 1980–1981 from the pooled leg muscles of normal E-79 and E-62 day human fetuses (73,74). Donors were anonymous and permission to use the therapeutically aborted fetal tissues for research purposes was provided under a signed patient consent form approved by the University of Washington’s Human Subjects Committee. About 75% (E-79) and 65% (E-62) of clones derived from these cell strains formed differentiated skeletal muscle and the remaining clones exhibited various types of fibroblastic morphologies (73,74).

DHS, RRBS, chromatin state, histone modification and RNA-seq profiles are available at http://ucsc.genome.edu (ENCODE: DNA methylation by RRBS, Richard Myers, HudsonAlpha Institute for Biotechnology; Open Chromatin, DNasel HS, Greg Crawford, Duke University; Chromatin State Segmentation by HMM, Histone Modifications by ChIP-seq, and CTCF ChIP-seq, Bradley Bernstein, Broad Institute; Transcription Levels by RNA-seq, non-strand-specific, >200 nt poly(A)+ RNA, Barbara Wold, California Institute of Technology; Long RNA-seq for poly(A)+ Whole-Cell RNA by strand-specific analysis, >200 nt poly(A)+ RNA, Tom Gingeras, Cold Spring Harbor Laboratories). For quantification of RNA-seq data, we used ENCODE non-strand-specific profiles or, for Mb/Mt comparisons, independent poly(A)+ RNA-seq data generated in collaboration with Greg Crawford (26). Quantification of RNA-seq data was done with the Cufflinks CuffDiff
tool (75). Tissue histone modification profiles were from http://epigenomebrowser.org, part of the ROADMAP Epigenomics Project.

Myogenic hypomethylation refers to statistically significant differences between myogenic and non-myogenic samples. Statistical analysis of myogenic differentially methylated sites (a change in methylation of at least 50% at a significance level of \(\leq 0.01\) from RRBS data on 18 types of cell culture or 15 types of tissues) was done as previously described (24). Myogenic DMRs from the same RRBS data sets were detected using our UPQ algorithm as recently described (33). Both 5mC and 5hmC were quantified at specific CCGG sites by previously described methods (Epimark 5-hmC and 5-mC Analysis Kit, New England BioLabs) (24) using real-time PCR with 2-to-5 (usually 3) biological replicates. The samples used for the 5mC/5mC analysis were independent of those used for RRBS. For determination of the significance of unusually high levels of 5mC at specific sites in the set of Mb and Mt, a mixed effects ANOVA model was independent of any sites and pairwise comparisons of site-specific estimates were conducted using the Tukey method.

**Supplementary Material**

Supplementary Material is available at HMG online.

**Conflict of Interest statement.** None declared.

**Funding**

We thank Melody Badoo and the Tulane Cancer Center (COBRE grant NIGMS P20GM103518) for help with the Cufflinks analysis of the ENCODE RNA-seq data. This research was supported in part by a grant from the National Institutes of Health to M.E. (NS04885).

**References**


