

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.Sciencedirect.com)

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Distant cis-regulatory elements in human skeletal muscle differentiation

Rachel Patton McCord ^{a,e,1}, Vicky W. Zhou ^{a,d}, Tiffany Yuh ^{a,f}, Martha L. Bulyk ^{a,b,c,e,*}

^a Division of Genetics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

^b Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

^c Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, USA

^d Biological and Biomedical Sciences Graduate Program, Harvard Medical School, Boston, MA 02115, USA

^e Harvard University Graduate Biophysics Program, Cambridge, MA 02138, USA

^f Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:

Received 12 August 2011

Accepted 17 August 2011

Available online 23 August 2011

Keywords:

Cis-regulatory modules

DNA looping interactions

Transcriptional regulation

Transcription factors

DNA binding sites

ABSTRACT

Identifying gene regulatory elements and their target genes in human cells remains a significant challenge. Despite increasing evidence of physical interactions between distant regulatory elements and gene promoters in mammalian cells, many studies consider only promoter-proximal regulatory regions. We identify putative cis-regulatory modules (CRMs) in human skeletal muscle differentiation by combining myogenic TF binding data before and after differentiation with histone modification data in myoblasts. CRMs that are distant (>20 kb) from muscle gene promoters are common and are more likely than proximal promoter regions to show differentiation-specific changes in myogenic TF binding. We find that two of these distant CRMs, known to activate transcription in differentiating myoblasts, interact physically with gene promoters (*PDLIM3* and *ACTA1*) during differentiation. Our results highlight the importance of considering distal CRMs in investigations of mammalian gene regulation and support the hypothesis that distant CRM-promoter looping contacts are a general mechanism of gene regulation.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The identification of genomic sequence regions that regulate genes in a condition-specific manner is essential to understanding how the same genome sequence can give rise to the diversity of cell types and functions observed in an organism. In an organism with a small genome, such as yeast, the majority of gene regulation can be explained by transcription factor (TF) binding and chromatin modifications within approximately 600 bp to 1 kb of DNA sequence upstream of the regulated gene [1,2]. In metazoans, numerous prior studies in a range of organisms from sea urchin to mammals have identified cis-regulatory modules (CRMs), consisting of clusters of TF binding sites, located next to (or within the introns of) the genes whose expression they regulate [3]. However, as compared to the yeast genome, metazoan genomes, in particular mammalian genomes, have a much higher proportion of noncoding sequence, and recent research has highlighted the importance of more distant CRMs in gene regulation within these genomes [4–6]. Much detailed work on distantly located transcriptional enhancers in *Drosophila* has

shown the importance of such distant regulatory elements and the ways that they can be directed to their target genes by insulator boundaries and promoter targeting sequences [7]. In mammalian systems, the rules governing enhancer–promoter links are less evident, but certain examples of distant regulatory elements have been studied in detail. For example, the locus control region (LCR) of the murine β -globin locus forms a GATA-1-dependent looping interaction with actively transcribed globin gene promoters located approximately 40–60 kb away in erythroid cells [8,9]. Such results suggest that a complete understanding of gene regulation will require searching for CRMs distant from target genes and further studies of how these CRMs are directed to and regulate their target genes.

The differentiation of human skeletal myoblasts into mature muscle fibers requires the coordinated regulation of many genes and is essential for the development and maintenance of proper muscle function. This differentiation process can be easily induced and monitored in cell culture, making it a tractable model system for investigating the regulatory mechanisms underlying dynamic, tissue-specific gene expression in human. Prior studies have measured changes in gene expression and TF binding during differentiation in primary human skeletal muscle cells or in the similar C2C12 mouse skeletal muscle cell line [10–14]. These studies have shown that a master regulatory TF, MyoD (encoded by *MYOD1*), initiates the cascade of gene regulation that leads to fusion of undifferentiated myoblast cells into multinucleated, elongated myotubes with developed contractile elements [10,15]. Other TFs in the myogenic regulatory factor (MRF) basic helix-loop-helix (bHLH) family – myogenin

Abbreviations: CRM, cis-regulatory module; 3C, Chromosome Conformation Capture; TF, transcription factor.

* Corresponding author at: Harvard Medical School New Research Building, Room 466D, 77 Avenue Louis Pasteur, Boston, MA 02115, USA. Fax: +1 617 525 4705.

E-mail address: mlbulyk@receptor.med.harvard.edu (M.L. Bulyk).

¹ Current affiliation: Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, MA 01605, USA.

(MyoG, encoded by *MYOG*), Myf5 (*MYF5*), and MRF4 (*MYF6*) – some of which are activated directly by MyoD, work together with MyoD and other TFs such as Serum Response Factor (SRF) and the Myocyte Enhancer Factor 2 (Mef2) family to regulate the expression of genes involved in the continuation and completion of this differentiation process [10,15,16]. Using the sets of differentially expressed genes, conservation of sequence across species, and the DNA binding site motifs of myogenic TFs, CRMs responsible for coordinating changes in expression during myogenic differentiation have been predicted computationally [12,17,18]. However, unlike the β -globin locus in which some of the classic distant regulatory interactions have been characterized previously, little is known about the role of distant regulatory modules in human skeletal muscle differentiation.

Previous efforts to characterize the transcriptional regulatory network in skeletal muscle differentiation have measured binding of myogenic TFs using chromatin immunoprecipitation followed by microarray hybridization or sequencing (ChIP-chip or ChIP-Seq, respectively) in murine C2C12 cell lines [10,11,14] and have tested the regulatory function of some predicted TF-bound elements with reporter assays [18]. While the results of these studies have revealed previously unknown regulatory connections between TFs and differentiation processes, they tended to focus on TF binding near promoters (approximately 1–4 kb upstream of transcriptional start sites (TSSs)) [10,11,18]. In contrast, methods to predict CRMs involved in myogenic differentiation have identified candidate regulatory modules by searching sequences up to 10 kb [17] or 50 kb [12] away from TSSs. A handful of the more distantly located predicted CRMs (approximately 20–30 kb upstream or downstream of TSSs) were found to activate expression in reporter assays specifically during myogenic differentiation and to be bound by myogenic TFs [12]. These results suggest that gene regulatory myogenic TF binding occurs at locations distant from the proximal promoter regions that have been the focus of most prior studies.

In this study, we identified genomic regions with a potential regulatory role in myogenic differentiation by combining experimental measurements of myogenic TF binding before and after muscle differentiation with previously published locations of enhancer-associated chromatin modifications in myoblasts. We found such putative CRMs located both proximal to and distant from muscle genes, and observed that distant CRMs showed changes in TF binding during differentiation more often than did CRMs located in proximal promoter regions. Investigating the potential regulatory roles of such distant candidate CRMs, we found that a few are located near known microRNAs, which we propose might be involved in myogenic differentiation. To determine whether distant CRMs with no microRNA or protein-coding gene promoters nearby might form long-range looping interactions with muscle gene promoters, we examined two distant CRMs previously found to drive gene expression specifically during myogenic differentiation as case examples. We found that these CRMs form differentiation-specific physical interactions with their closest target genes. Our results provide further support that a complete understanding of transcriptional regulation in mammals will require consideration of CRMs located distant from their target genes in the genome sequence.

2. Results

2.1. A combination of myogenic transcription factor binding and enhancer-associated histone marks best distinguishes positive and negative control CRMs

To search for candidate cis-regulatory regions utilized in human skeletal muscle differentiation, we characterized the binding of the known myogenic TFs MyoG, MyoD, and SRF to regions surrounding muscle genes before and after differentiation. ChIP-chip analysis of TF binding in 100 kb regions surrounding each of 104 muscle genes

resulted in 1781 significant (Bonferroni-corrected; $\alpha = 0.05$), biologically replicated 1-kb regions of DNA bound by either MyoG or SRF immediately before (0 h) differentiation and 3702 1-kb regions bound by either MyoG or SRF 48 h after differentiation.

Prior studies have shown that considering clusters of TF binding sites and certain chromatin modifications can aid in the identification of functional regulatory regions, narrowing the set of candidate CRMs from the large set of regions bound by single factors to a smaller, higher-confidence set [19–21]. Therefore, we investigated which combinations of TF binding and chromatin features might be predictive of CRMs involved in myogenic differentiation. We considered the overlap of MyoG and SRF binding from our ChIP-chip data and the changes we measured in this binding from 0 h to 48 h of differentiation, as well as supplementary single-replicate ChIP-chip data that we collected for MyoD binding at 0 h and 48 h and for the enhancer associated protein p300 (*EP300*) [22] binding at 48 h after differentiation. In addition to our own experimental data, we considered previously published data on histone modifications in human myoblasts (corresponding approximately to the 0 h of differentiation timepoint), focusing on marks that have been shown to be associated with enhancers both near and far from gene TSSs (H3K4me1 and H3K27Ac) [23]. As an initial guide for future predictions of putative CRMs, combinations of the above features were evaluated for their ability to distinguish between 9 positive control CRMs and 9 negative control genomic regions. These control regions were identified previously as either active or inactive, respectively, in regulating gene expression during myogenic differentiation, as evidenced by their being occupied by MyoD, MyoG, and SRF by ChIP-qPCR and by their ability to activate reporter gene expression during differentiation [12] (Table 1).

We observed MyoG binding at 48 h after differentiation for all 9 positive controls and only 2 of the negative controls, making this a strong indicator of CRMs active during myogenic differentiation. The binding of SRF, MyoD, and p300 at 48 h post-differentiation co-localized with MyoG binding at subsets of the known CRMs, but including the ChIP data for these factors did not help to further distinguish the positive control CRMs from the negative control regions (Table 1). The binding of MyoD or SRF alone at 48 h post-differentiation was less specific than MyoG binding for distinguishing these sets of regions: we observed SRF binding at 7 positive and 7 negative controls and MyoD binding at 7 positive and 3 negative controls.

Combining TF binding information with data on enhancer-associated histone marks H3K4me1 and H3K27Ac in human myoblasts [23] further helped to distinguish the positive and negative controls. We found that the CRMs bound by MyoG at 48 h post-differentiation and that had either H3K4me1 or H3K27Ac modifications in myoblasts (i.e., immediately before differentiation) include all of the positive controls and none of the negative controls. MyoG binding at regions with enhancer-associated chromatin modifications not only identified positive control CRMs located in proximal promoter regions, but also distinguished previously validated distant (greater than 20 kb away from TSS) CRMs from negative control genomic regions (Figs. 1A and B). Since this combination of enhancer marks and MyoG binding best discriminates the positive control CRMs from the negative controls, in subsequent analyses we considered all regions with these features as a set of putative muscle differentiation CRMs.

The promoter-associated histone mark H3K4me3 [24] also provided a good distinction between the known CRMs and the negative controls, but the levels of this mark were only marginal for validated muscle differentiation enhancers more distant from promoters (Fig. 1A), in agreement with previous observations that H3K4me3 is generally not observed at enhancers located more than 2 kb away from TSS [23]. Therefore, we did not include this feature in subsequent analyses of CRMs since our analyses broadly encompassed both proximal and distant regulatory regions.

Table 1

'X' indicates binding of a TF to a region.

	Region name	Region coordinates (Hg18)	MyoG 0 h	MyoG 48 h	SRF 0 h	SRF 48 h	MyoD 0 h	MyoD 48 h	p300 48 h	H3K4me1 myoblasts	H3K27Ac myoblasts
Positive controls	ACTA1 promoter	chr1: 227635616–227637614	X	X	X	X		X	X	X	X
	ACTA1 CRM	chr1: 227612394–227614392	X	X	X	X	X	X		X	X
	PDLIM3 promoter	chr4: 186693000–186695000		X	X	X	X	X	X	X	X
	PDLIM3/ SORBS2 CRM	chr4: 186724927–186726925		X	X	X	X	X	X	X	X
	CAV3	chr3: 8749393–8751391		X	X	X	X	X	X	X	X
	TNNT2	chr1: 199612744–199614742		X	X	X	X	X		X	X
	COX6A2	chr16: 31346483–31348481		X		X		X	X	X	X
	HSPB3	chr5: 53787965–53789963	X	X			X		X	X	X
CSRP3	chr11: 19179244–19181242		X			X		X	X	X	
Negative controls	GAP43	chr3: 116777470–116779468							X		
	CPM	chr12: 67659795–67661793					X				
	MGLL	chr3: 129031127–129033125				X					
	BDKRB2	chr14: 95762300–95764298				X				X	X
	IGFBP4	chr17: 35868970–35870968				X	X		X	X	
	S1PR2	chr19: 10199535–10201533				X		X	X		
	KRT77	chr12: 51369942–51371940				X	X	X	X	X	
	CLC	chr19: 44947497–44949495		X	X	X	X				
HBZ	chr16: 148530–150528		X		X		X	X			

2.2. Comparison to computational prediction of CRMs

MyoG binding during differentiation and enhancer-associated histone marks in myoblasts not only distinguished positive control CRMs from negative control genomic regions, but also overlapped significantly with computational predictions of CRMs identified by searching for clusters of conserved MRF DNA binding site motif matches with the algorithm PhylCRM [12]. This MRF motif (the motif recognized by MyoG and MyoD) was more enriched among a set of genes upregulated during myogenic differentiation than any other combination of MRF, Mef2, SRF, or Tead motifs in a previous study [12]. Of the 1975 regions enriched for MyoG binding and enhancer histone marks, 1128 contain a CRM predicted by PhylCRM (Supplementary Fig. S1). This significant overlap ($p < 1 \times 10^{-14}$; Hypergeometric test) between computationally predicted CRMs and experimentally identified putative CRMs further demonstrates that PhylCRM can be used successfully to predict likely CRMs and that the observed MyoG binding events are likely mediated by direct interactions with MyoG binding sites within the predicted CRMs. Neither correlation is perfect, however: 43% of regions with experimentally observed MyoG binding and chromatin modifications were not PhylCRM-predicted CRMs, and 67% of PhylCRM-predicted CRMs do not show these *in vivo* indicators of regulatory function. TF binding at regions not predicted by PhylCRM could be indirect association through other TFs, rather than by the myogenic factor binding motif. Likewise, PhylCRM predicted regions that are not bound could be CRMs in other conditions. Given the limitations of both experimental TF binding measurements and computational predictions, it is important to combine such different sources of evidence to identify putative CRMs.

As a larger set of putative elements than the few positive and negative controls considered earlier, PhylCRM predictions allow us to further evaluate the patterns of TF binding at putative regulatory regions. Interestingly, the binding of some TFs, such as SRF, that were not as specific as MyoG for positive control CRMs, do show significant overlap with PhylCRM-predicted CRMs. As measured by the sum of sensitivity and specificity, SRF binding at 48 h is nearly as effective as MyoG binding at 48 h in discriminating PhylCRM-predicted CRMs from genomic regions that did not contain PhylCRM hits (Supplementary Table S1). This suggests that SRF

binding is an informative feature of many muscle differentiation CRMs, even though it is not highly specific in our small set of validated positive and negative controls. The degree of overlap between computational predictions and different combinations of TF binding and/or chromatin marks across different timepoints is shown in Supplementary Table S1.

2.3. Changes in TF binding during differentiation

Many of the genes whose predicted CRMs were included on our custom-designed ChIP-chip microarray were chosen because they were upregulated over a timecourse of human skeletal muscle differentiation [12]. To evaluate whether changes in TF binding might be a hallmark of cis-regulatory regions of muscle genes and might effect changes in their gene expression levels, we compared TF binding at 0 and +48 h relative to the induction of myoblast differentiation by serum removal. Changes in TF occupancy occurred at many genomic regions during differentiation; for example, 1033 of the 3007 regions bound by MyoG at 48 h were not bound by any of the examined myogenic TFs (MyoG, MyoD, or SRF) before differentiation (0 h). However, in most cases, including information about these changes in TF binding did not help to distinguish the positive control CRMs from the negative control regions since the patterns of TF binding changes during differentiation vary across the known regulatory regions. For example, the *ACTA1* promoter is bound by MyoG and SRF at both 0 and 48 h while the *COX6A2* promoter is unbound before differentiation and then bound by MyoG, MyoD, and SRF at 48 h. The validated distant enhancer region between the *PDLIM3* and *SORBS2* genes is bound by MyoD and SRF at 0 h and then by an expanded set of factors (MyoG, MyoD, and SRF) with stronger peak intensities at 48 h (Fig. 1A). Only in the case of SRF binding did the ChIP timecourse data help to distinguish the positive control CRMs from negative control regions: of the 7 known CRMs bound by SRF at 48 h, 6 were also bound at 0 h, while in contrast only 1 of the 7 negative control regions bound at 48 h was also bound at 0 h.

In order to determine the value of any of the measured ChIP datasets in predicting the locations of TF binding events at 48 h during differentiation, we constructed a dependency network using the WinMine toolkit [25]. A dependency network, like a Bayesian network, finds predictive

relationships between variables using conditional probability distributions [26]. The derived network indicates which variables (in this case, TF binding events at a given genomic region) are well predicted by other variables. The dependency network for our data suggests that

MyoD binding at 0 h is more important for predicting the binding patterns of MyoG, SRF, MyoD, and p300 at 48 h after induction of differentiation than are either MyoG or SRF binding at 0 h or H3K4me1 marks in myoblasts (Supplementary Fig. S2). These findings are consistent with

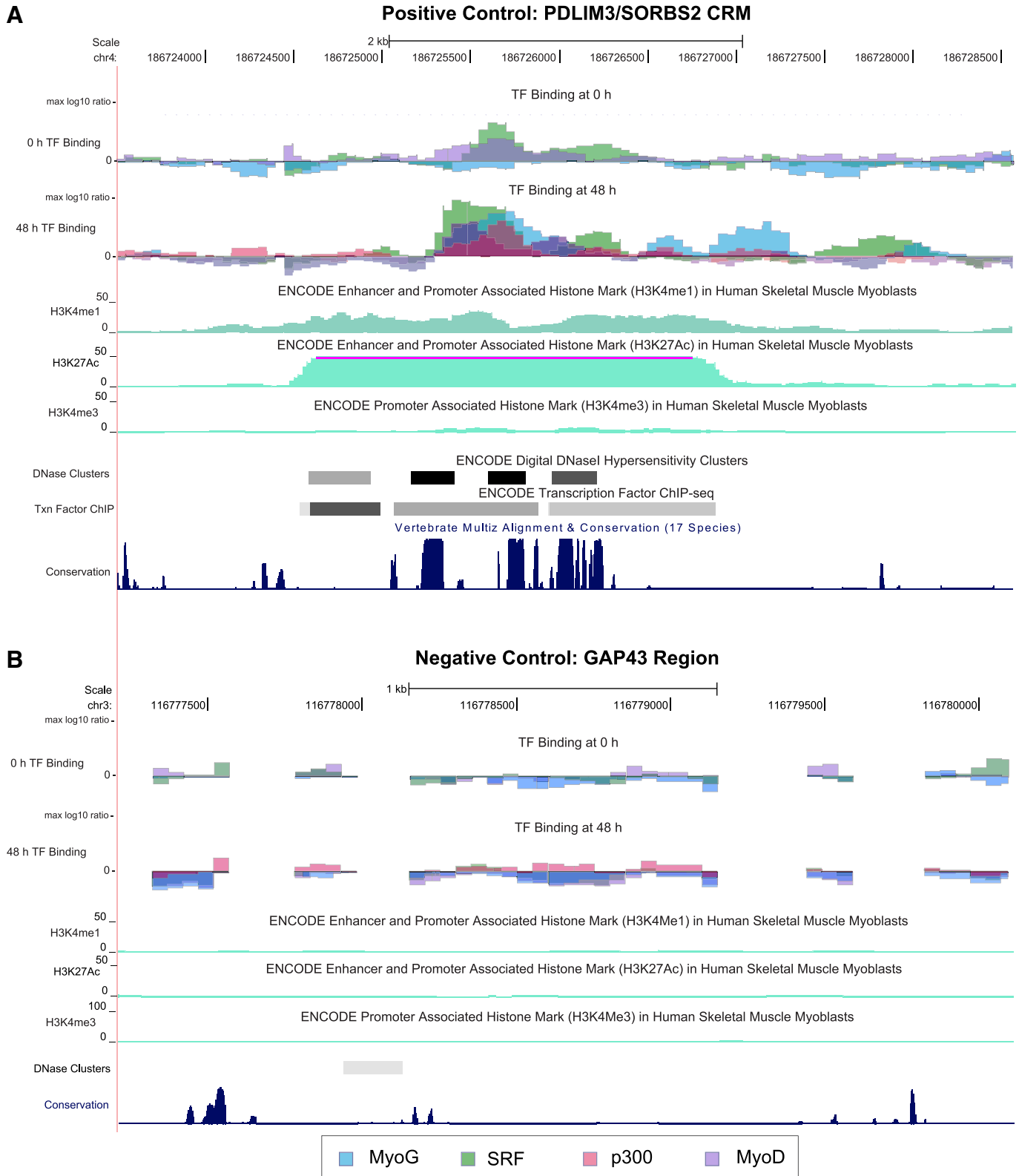


Fig. 1. ChIP-chip TF binding distinguishes positive and negative control regions. All data shown are log₁₀ ratios of IP to Input DNA signal and are normalized to the maximum ratio for each TF dataset. A) Myogenic TFs MyoG, MyoD, and SRF are bound, and enhancer-associated histone marks are observed, at the location of a validated CRM driving expression during muscle differentiation. B) No significant peaks of TF binding are seen for a negative control genomic region (near the *GAP43* gene).

prior literature on the pioneering role of MyoD in establishing cis-regulatory sites in myogenic differentiation [10].

2.4. Locations of TF-bound, putative regulatory regions relative to genes

Most regions previously known to have a role in myogenic differentiation (including most of the positive controls considered in the previous analyses) are located within proximal promoter regions of protein-coding genes (within 1–2 kb of TSS), but some of the putative regulatory

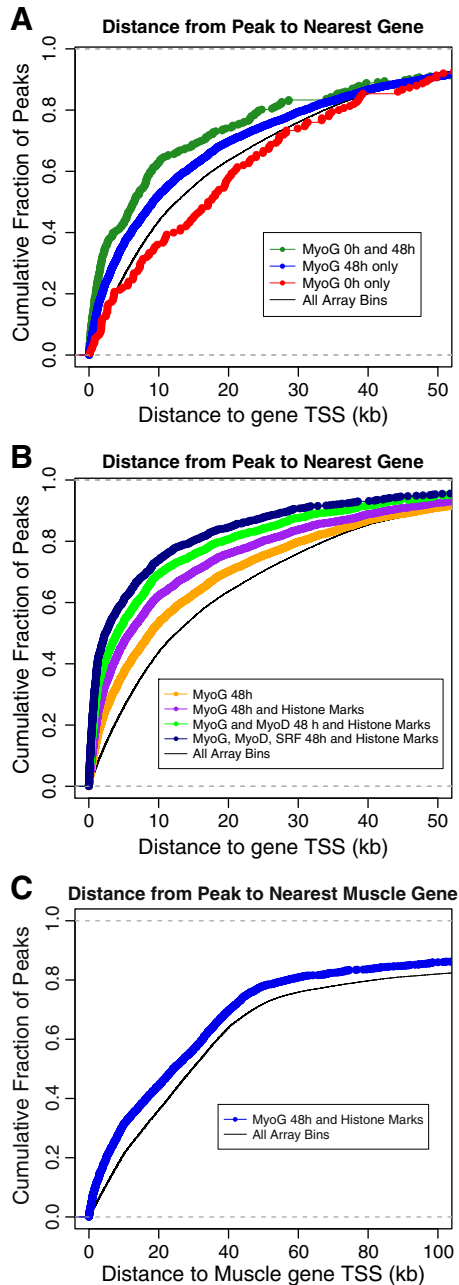


Fig. 2. Cumulative distributions of distances between putative CRMs and gene transcription start sites (TSS). A) Genomic regions bound by MyoG at both 0 and 48 h relative to differentiation are located closer to gene promoters than those whose MyoG binding status changes between 0 h and 48 h. B) Regions bound by multiple myogenic TFs and with enhancer-associated “Histone Marks” (H3K4me1 and H3K27Ac) are located closer to promoter regions than regions with fewer bound factors C) Many putative CRMs bound by MyoG and with enhancer-associated histone marks are located distant from the promoters of “muscle genes” that are differentially expressed during muscle differentiation or which have an annotated muscle function. “All Array Bins” in each case shows the cumulative distribution of distances to TSS for all 1-kb genomic regions covered on the ChIP-chip array.

elements we identified are more distant from any protein-coding genes (>20 kb away from TSS; Fig. 1A). To further investigate the role of distant genomic regions in the regulation of myogenic differentiation, we examined the distribution of distances between genes and putative regulatory elements enriched for myogenic TF binding and enhancer-associated histone modifications. There is an enrichment of these putative regulatory regions close to gene TSS (within 1 kb) as compared with the locations of all possible 1-kb regions on our custom ChIP-chip array ($p < 1 \times 10^{-14}$, Hypergeometric test; Fig. 2). Despite this skew toward proximal promoter regions, about 80% of regions bound by MyoG at 48 h with H3K4me1 or H3K27Ac modifications fall outside of traditional promoters (1 kb from TSS) and, strikingly, 25% of such putative CRMs are more than 20 kb away from any TSS. Interestingly, genomic regions occupied by different TFs and with different binding dynamics during differentiation differ in their distances from gene TSS. Genomic regions that either lose or gain MyoG binding from 0 h to 48 h tend to be located further from TSSs than those that have constant MyoG binding at 0 and 48 h ($p = 1.39 \times 10^{-4}$ for constant binding versus gain of binding at 48 h; Kolmogorov–Smirnov (KS) test; Fig. 2A). Increased distance from gene promoters is also observed for regions that change in both MyoG and SRF binding between 0 and 48 h of differentiation as compared with those that are bound by both MyoG and SRF at both timepoints. In addition, regions bound by all 3 of these TFs (MyoG, MyoD, SRF) tend to lie closer to genes than regions with only one or two TFs bound ($p < 2.2 \times 10^{-16}$ for MyoG at 48 h vs. MyoG, SRF, and MyoD at 48 h; KS test; Fig. 2B).

For 69% of the regions with enriched MyoG binding at 48 h and H3K4me1 or H3K27Ac marks in myoblasts, the closest gene is not known to be involved in myogenic differentiation and is not differentially expressed during this process. In fact, 90% of such putative CRMs are located outside of 1 kb muscle gene promoter regions and 56% are more than 20 kb away from a muscle gene TSS (Fig. 2C). Furthermore, for 30% of these MyoG-bound, H3K4me1 or H3K27Ac modified regions, associating with the closest muscle gene along the linear chromosome requires skipping over intervening non-muscle genes.

We specifically included 100 genomic regions on our custom ChIP-chip array design that contained 287 high-scoring PhylCRM predictions of myogenic CRMs with clusters of evolutionarily conserved myogenic TF binding site matches, but were located far (>50 kb) from any annotated muscle gene TSS. These distant predicted CRMs exhibited MyoG binding and enhancer-associated histone marks at a somewhat lower frequency (26%) than PhylCRM-predicted CRMs <50 kb away from muscle gene TSS (35%). However, permutation testing showed that this level of overlap was greater than expected ($p = 0.02$) for randomly selected genomic regions represented on the array and much greater than expected ($p < 1 \times 10^{-4}$) for genomic regions with the same median distance to the nearest TSS of any protein-coding gene (21.8 kb). The significant overlap between MyoG binding events, H3K4me1 or H3K27Ac histone marks, and the PhylCRM computational CRM predictions indicates that these putative distant CRMs may indeed be biologically functional CRMs.

2.5. Potential regulation of microRNAs by CRMs distant from muscle protein-coding genes

To explore the possible functions of putative myogenic CRMs that are distant from muscle protein-coding genes, we searched for microRNA genes annotated in miRBase [27] that are located adjacent to these CRMs. We identified 14 different putative CRMs that are located within 25 kb of 10 different microRNA genes and that are closer to these microRNAs than to any known muscle protein-coding genes (Table 2). In 5 cases, the putative CRMs are closer to one of these microRNAs than to any protein-coding gene (Table 2). Some of these microRNAs that may be regulated by the myogenic TFs bound at these putative CRMs already have an annotated function in muscle differentiation (Table 2). For example, miR-133 and miR-1 (encoded

Table 2
All potential CRMs (bound by MyoG at 48 h and with enhancer-associated histone marks) that are within 25 kb of a miRNA gene and closer to a miRNA gene than to any muscle gene TSS. Some CRMs are closest to a miRNA gene while others are closer to some other protein-coding gene. Grey highlighting indicates miRNAs with a previously documented role in muscle differentiation.

	ChIP peak location (MyoG48h filtered by histone marks)	miRNA gene symbol	Position relative to miRNA 5' end	Closest gene	Position relative to gene TSS	Closest muscle gene	Position rel. to muscle gene TSS
miRNA closer than any gene	chr19: 10201001–10203000	MIR4322	89 bp upstream	NM_004230 (S1PR2)	948 bp downstream	NM_015719 (COL5A3)	219.9 kb upstream
	chr18: 17664001–17666000	MIR1-2	1.95 kb upstream	NM_020774 (MIB1)	89.5 kb downstream	NM_005257 (GATA6)	338.4 kb upstream
		MIR133A1	5.26 kb upstream				
	chr20: 33035001–33036000	MIR499	5.84 kb upstream	NM_020884 (MYH7B)	28.6 kb downstream	AB040945 (MYH7B)	9.1 kb downstream
	chr20: 33047001–33048000	MIR499	7.66 kb downstream	NM_020884 (MYH7B)	42.1 kb downstream	AB040945 (MYH7B)	22.6 kb downstream
chr6: 36720001–36722000	MIR3925	22.7 kb upstream	NM_001220777 (CDKN1A)	31.2 kb upstream	U03106 (CDKN1A)	33.5 kb upstream	
miRNA closer than any muscle gene	chr7: 32733001–32736000	MIR550A2	4.62 kb upstream	NR_003502 (ZNRFP1)	414 bp downstream	NM_198098 (AQP1)	1.81 Mb downstream
		MIR550B2	4.71 kb downstream				
	chr11: 1831001–1832000	MIR4298	5.84 kb downstream	NM_002339 (LSP1)	726 bp downstream	NM_003282 (TNNI2)	14.7 kb downstream
	chr11: 1842001–1846000	MIR4298	6.65 kb upstream	NM_001013253 (LSP1)	2.5 kb upstream	NM_003282 (TNNI2)	27.2 kb downstream
	chr11: 1847001–1848000	MIR4298	10.2 kb upstream	NM_001013253 (LSP1)	1.02 kb downstream	NM_003282 (TNNI2)	30.7 kb downstream
	chr11: 1848001–1849000	MIR4298	11.2 kb upstream	NM_001013254 (LSP1)	174 bp upstream	NM_003282 (TNNI2)	31.7 kb downstream
	chr11: 1849001–1851000	MIR4298	12.7 kb upstream	NM_001013255 (LSP1)	1.3 kb downstream	NM_003282 (TNNI2)	33.2 kb downstream
	chr5: 88014001–88018000	MIR9-2	17.5 kb upstream	NR_015436 (LINC00461)	376 bp downstream	NM_001193349 (MEF2C)	139.5 kb downstream
	chr11: 1955001–1959000	MIR675	17.6 kb downstream	NR_024471 (MRPL23-AS1)	10.7 kb downstream	NM_001042780 (TNNT3)	59.6 kb downstream
	chr11: 1852001–1862000	MIR4298	19.7 kb upstream	NM_001013255 (LSP1)	8.3 kb downstream	NM_003282 (TNNI2)	37.8 kb downstream

by MIR133A1 and MIR1-2, respectively) influence skeletal muscle proliferation and differentiation, with miR-133 specifically downregulating the myogenic TF SRF, and miR-1 directly downregulating HDAC-4 (*HDAC4*) and thus indirectly modulating Mef2 gene expression [28]. Additionally, miR-499 (encoded by MIR499) is differentially expressed during murine skeletal muscle differentiation [29] and is known to be involved in cardiac myocyte differentiation [30] and other processes in skeletal muscle [31]. The 7 remaining microRNAs (miR-9-2, miR-550a-2, miR-550b-2, miR-675, miR-3925, miR-4298, and miR-4322) we identified based on their proximity to putative myogenic CRMs do not yet have known regulatory functions in muscle differentiation and would be interesting to investigate further for their potential myogenic regulatory roles in future studies.

2.6. Differentiation-specific looping interactions between distant CRMs and muscle gene promoters

We used Chromosome Conformation Capture (3C) [32,33] experiments to investigate whether the putative myogenic CRMs that we found located distant from muscle genes form physical interactions with their adjacent, putative target genes during myogenic differentiation. We selected two CRMs that were bound by MyoG at 48 h after induction of differentiation, exhibited H3K4me1 and H3K27Ac enhancer-

associated histone marks, were found to drive expression of a reporter gene during differentiation [12], and were located at least 20 kb away from the TSS of the closest known protein-coding gene. The “ACTA1 CRM” (Fig. 3A) is located 23 kb downstream of the TSS of *ACTA1*, a gene which encodes the skeletal muscle alpha actin protein, an essential part of the contractile element in muscle. The “PDLIM3/SORBS2 CRM” (Fig. 3B) is located 32 kb upstream of *PDLIM3*, which encodes a cytoskeletal organizing protein localized to the Z line of the sarcomere and which may also regulate the myogenic differentiation transcriptional network by affecting the nuclear localization of SRF [34]. This CRM is also located 240 kb downstream of *SORBS2*, another gene differentially expressed during myogenic differentiation.

To assay the physical interactions between these two CRMs and their adjacent genes, we performed 3C experiments on human muscle cells 48 h before, immediately (“0 h”) before, and 48 h after induction of differentiation by serum removal. We identified “differentiation-specific interactions” by searching for any peaks of interaction in the 0 h or 48 h post-differentiation cells, outside of the neighboring +/–15 kb that may interact highly by random collisions with the fixed bait, that were higher than the 48 h pre-differentiation interaction level (see Materials and methods). The results of these experiments show evidence for a differentiation-specific interaction between the “ACTA1 CRM” and the proximal promoter region spanning from 1.6 kb

48 h before differentiation, but then become evident at 0 h, when the cells are confluent and elongating and differentiation is induced by serum removal. These CRM–promoter interaction peaks become more pronounced at 48 h after induction of differentiation (Fig. 3). These results suggest that a differentiation-induced chromosome conformation is established at these gene loci as the cells are approaching confluence and may be initiating the differentiation process even before stimulation of differentiation by serum removal. These physical interaction data parallel the previously published observation that the expression of the *ACTA1* and *PDLIM3* genes begins to increase from –48 h to 0 h [12] and our observations described above that predicted CRMs are often already bound by myogenic TFs at 0 h. Interestingly, the *PDLIM3/SORBS2* CRM was observed to interact only with the *PDLIM3* promoter region 23 kb downstream of the CRM but not with either of the two alternative promoters of *SORBS2* located 240 kb and 385 kb upstream of the *PDLIM3/SORBS2* CRM (Supplementary Fig. S5). Thus, while the *PDLIM3/SORBS2* CRM acts on its target promoter over a distance, the CRM selectively interacts with the closer promoter rather than the farther promoter. These observations suggest a model in which distant enhancers participate in activating gene expression by looping to contact the promoters of the genes closest to them. Alternatively, the promoter selectivity could be due to promoter sequence elements that differ between genes to specify the appropriate interaction or cell-type specific insulator architecture [35]. Interestingly, while the *PDLIM3/SORBS2* CRM did not interact with the *SORBS2* promoters, it did show evidence of differentiation-specific interaction with an intronic region of one of the *SORBS2* isoforms where a peak of H3K4me1 enrichment is observed (Supplementary Fig. S4). This intronic region may be an additional regulatory region, and the observed contacts between these two noncoding regions raise the interesting possibility of combinatorial regulation of genes by multiple CRMs analogous to combinatorial TF input into transcriptional enhancers.

3. Discussion

In this study we found a variety of evidence that CRMs involved in myogenic differentiation can be located distant from the proximal promoter regions of muscle genes. When 50 kb of sequence upstream and downstream of muscle genes are considered, many genomic regions with the same TF binding patterns and histone modifications as those observed in known myogenic CRMs are located more than 20 kb away from muscle genes (56% of the putative set), and 90% of them fall outside of 1-kb proximal promoters.

We found that distant CRMs were more likely than proximal CRMs to show changes in MyoG and SRF binding after induction of human muscle differentiation. This finding is in agreement with a prior study that characterized the genome-wide binding of MyoD during murine myogenic differentiation with ChIP-Seq [14]. In that study, Cao et al. noted that 90% of promoter proximal MyoD peaks were bound by MyoD both before and after differentiation and that most of the “differentially bound” regions tended to be located further from gene TSS. “Differentially bound” peaks of MyoD binding were found in that study to be more likely to regulate differentiation-specific gene expression changes [14]. This suggests that the distant differentially bound MyoD peaks in that study and the distant putative CRMs we observed here to be differentially bound by MyoG and SRF are likely to be differentiation-specific regulatory regions. These results are also in agreement with the previous general observation that the chromatin state at promoters is often shared across cell types while distant putative regulatory regions often have highly cell-type specific histone modifications [36]. Beyond trends in binding dynamics with distance from gene TSS, we also observed that regions bound by three myogenic TFs tend to be located closer to promoters than regions bound by one or two TFs. This may relate to the phenomenon of TF binding “hotspots” (genomic regions where nearly every assayed TF is observed to be bound) observed in fly and worm [37,38]. In this scenario, the regions bound within proximal promoters may be “hotspots” for TF binding,

while the distant regions are not hotspots but rather are bound selectively across timepoints and cell types.

The need to search for CRMs outside of proximal promoter regions increases the difficulty of a computational search by increasing the amount of DNA sequence to be examined. However, we previously found that the PhylCRM algorithm, which searches for clusters of conserved TF binding sites, was able to predict CRMs distant from muscle genes; these predicted CRMs were subsequently validated experimentally by ChIP-qPCR and by reporter assays [12]. Here, we found that two of these predicted and validated CRMs, the *ACTA1* CRM and the *PDLIM3/SORBS2* CRM, both located more than 20 kb from muscle genes, interact with adjacent gene promoters as assayed by 3C. Additionally, PhylCRM predicted a CRM 340 kb away from the nearest muscle gene TSS near a microRNA with a known role in muscle differentiation; we found this region to be bound by the myogenic TFs MyoG, MyoD, and SRF 48 h after the initiation of differentiation. Thus, computational predictions combined with experimental measurements of cell-type specific histone modifications and TF binding should provide a strong starting point for identifying active transcriptional enhancers both within and far from promoters in future studies of other cell types and processes. The DNA binding motif for the myogenic factor Mef2, in combination with MyoG and SRF motifs, was found previously to help in identifying CRMs using PhylCRM [12]. The Mef2 antibodies available at the time of the ChIP-chip experiments presented here did not produce high-quality DNA binding profiles, but future experiments characterizing the binding of Mef2 during differentiation might provide further evidence for CRM locations.

After distant CRMs are identified, a remaining challenge is to determine which genes they regulate and the mechanisms by which they affect gene expression. The differentiation-specific interactions observed in this study by 3C for the CRMs located 20–35 kb away from *ACTA1* and *PDLIM3*, combined with the myogenic TF binding observed with ChIP-chip and the upregulation of expression of these genes during differentiation, suggest a model in which these distal enhancers assist in activating gene expression by looping to contact the promoters of the closest genes. These are the first such distal interactions between a CRM and promoter that have been identified in mammalian muscle differentiation. Combined with the mounting evidence of CRM–promoter interactions in other cell types and systems, these data suggest that long-distance physical interaction is a general mechanism of enhancer action rather than a mechanism specific to a particular system. That these enhancer–promoter interactions begin to form at 0 h of differentiation suggests that a differentiation-specific set of looping interactions may be established before the resulting changes in gene expression and cell phenotype begin. This is consistent with findings in other systems: in mouse erythroid progenitors, the β -globin LCR region exhibits interactions that are established before changes in gene expression occur [39], and interactions between enhancers and promoters implicated in hormone-responsive gene expression in mouse adenocarcinoma cells are already present at lower levels before hormone treatment [40].

When identifying target genes for CRMs outside proximal promoter regions, some previous studies have used the simplifying assumption that CRMs are likely to interact with gene promoters lying within a domain bounded by CTCF sites, on the basis that CTCF often acts as an insulator across which boundary interactions are less likely to occur [14,41]. This assumption matches experimental results in the case of the *PDLIM3/SORBS2* CRM, which interacts with the *PDLIM3* promoter in the same CTCF domain (as measured by a previous study in human skeletal muscle myoblasts [23]), but does not interact with the *SORBS2* promoters that are separated from the CRM by several CTCF sites. However, the interaction we observed between the *ACTA1* CRM and the *ACTA1* promoter crosses a CTCF site, as does the interaction between the *PDLIM3/SORBS2* CRM and the predicted CRM located 27 kb upstream of a *SORBS2* promoter. Thus, our results suggest that the links between CRMs and

their target genes cannot always be predicted accurately by CTCF sites.

How enhancer–promoter interactions are established and why they are advantageous to myogenic differentiation are yet to be determined. It is possible that the myogenic TFs observed to bind to these interacting regions are also responsible for mediating these interactions. However, a full understanding of the interaction mechanism will require future work to search for sequence motifs recognized by other factors potentially involved in establishing chromosomal domains (CTCF, for example [39]), to experimentally purify factors associated with the interacting genomic regions [42], and to disrupt implicated DNA binding sites to test whether they are required for the physical interactions.

Although this study primarily focused on the regulatory links between distant CRMs and target genes, we observed a differentiation-specific interaction between one distant CRM (the *PDLIM3/SORBS2* CRM) and another region exhibiting the H3K4me1 enhancer-associated histone modification 265 kb away in the intron of *SORBS2*. This interaction raises the interesting possibility of combinatorial regulation of genes by multiple CRMs, analogous to combinatorial TF input into transcriptional enhancers. Interestingly, while the *PDLIM3/SORBS2* CRM interaction with the *PDLIM3* promoter seems to be already present at the initiation of differentiation (0 h), the interaction with this other putative CRM does not appear until 48 h into differentiation. This may indicate a distinct role of a CRM–CRM interaction in the timecourse of myogenic differentiation. Multiple distant (>50 kb away) enhancer regions have been previously shown to interact to select the appropriate promoter for expression of the major histocompatibility complex regulator *CIITA* [43]. Future work will be needed to determine the prevalence of such interactions involving multiple CRMs and what role they might play in a process such as myogenic differentiation.

This study identifies putative CRMs for myogenic differentiation distant from muscle gene promoters and demonstrates that some of these CRMs physically interact with their target genes. These results highlight the importance of looking beyond the proximal promoter to understand the transcriptional regulation of genes involved in differentiation. These distant CRM–promoter interactions may relate to the global changes in genome organization observed as cells undergo differentiation [44]. In the future, combining computational enhancer predictions, experimental tests of distant CRMs, and global measurements of chromosome conformation [45] will help to further elucidate the mechanisms of gene regulation during this differentiation process.

4. Materials and methods

4.1. Proliferation and differentiation of human myoblasts in cell culture

Adult primary human skeletal myoblasts (Lonza) were grown in SkGM-2 medium (Lonza). Myogenic differentiation was stimulated by switching the culture medium to DMEM-F12 with 2% horse serum (Sigma) when the cells reached about 70% confluence. All time points referred to in this study are with respect to the time of switching to differentiation medium. The majority of the results presented here were obtained using adult male skeletal myoblasts; female skeletal myoblasts were used in one 3C experiment to confirm the reproducibility, and lack of gender-specificity, of the differentiation-specific physical interaction of the *PDLIM3/SORBS2* CRM.

4.2. Design of the custom ChIP-chip array

We designed a custom ChIP-chip array using the Agilent 1x244K oligonucleotide array format (Agilent Technologies Inc., AMADID # 015037) to achieve 25-bp resolution tiling of 100 kb surrounding each of 104 selected human genes. The genes were chosen because they were either upregulated over the timecourse of myogenic

differentiation [12], or annotated by Gene Ontology [46] or other literature as being a part of the muscle sarcomere. For each of these genes, the array covers –40 kb to +10 kb relative to transcription start and –10 kb to +40 kb relative to transcription stop. Additionally, the array sequence covers: 5 kb around each of 9 negative control regions previously found not to be bound by myogenic TFs [12]; 5 kb around each of 35 potentially regulatory noncoding SNPs (i.e., polymorphisms predicted to disrupt myogenic TF binding sites located within a computationally predicted CRM); and 5 kb around each of 100 predicted CRMs which are distant (i.e., greater than 50 kb away from the TSS) from annotated muscle-specific or differentially expressed genes [12,46].

4.3. Chromatin immunoprecipitation with microarray hybridization (ChIP-chip)

ChIP experiments were performed using a modification of the ChIP protocol described previously [12]. Briefly, cells at various timepoints relative to differentiation were crosslinked with 1% formaldehyde for 20 min and then quenched by glycine addition. Cells were harvested by scraping and then lysed with detergents and mechanical stress. After cell lysis, crosslinked chromatin was sheared by sonication to an average length of ~600 bp, and then immunoprecipitated with relevant TF-specific antibodies (MyoD: sc-760; Myogenin: sc-576; SRF: sc-335; p300: sc-585; all from Santa Cruz Biotechnology, Inc.). A separate, mock ‘no antibody’ (i.e., protein A beads without antibody) sample was processed as a negative control. The supernatant from the mock IP was saved as ‘Input’ DNA. The purified DNA from each ChIP was amplified and labeled with Cy5 (IPed DNA) or Cy3 (Input DNA) according to a published protocol [47]. Arrays were hybridized, washed, and scanned at 5-micron resolution according to recommendations from Agilent Technologies, Inc.

4.4. Computational analyses of ChIP-chip data

Agilent Feature Extraction software (version 9.1.3.1; Agilent Technologies Inc., Santa Clara, CA) was used to calculate \log_{10} ratios of IP DNA/Input DNA (Cy5/Cy3) for each microarray spot in each experiment and to remove spots flagged as outliers because of debris on the slide or excessively high background. These log ratios were averaged across technical replicate experiments. ChIPOTle software [48] was then used to identify peaks of statistically significant TF binding enrichment for the averaged technical replicate datasets (while keeping each biological replicate separate) using a sliding window size (200 bp) that accounts for the density of tiling and the average length of the immunoprecipitated DNA fragments. A threshold on the calculated p values for each peak was set so that the resulting FWER (α) was 0.05 by the Bonferroni correction. The final set of significantly bound regions was then determined for each TF and each timepoint by finding genomic regions with binding peak overlap between biological replicates (peaks with overlapping genomic coordinates or centers separated by less than 800 bp were counted as overlapping) and without peaks in the negative control ‘no antibody’ experiment.

Distances between significant ChIP-chip binding peaks and gene TSSs were calculated using RefSeq gene annotations, excluding microRNA genes. For the purpose of the analyses described in this study, we define “muscle genes” as genes that meet at least one of the following criteria: a) previously found to be differentially upregulated (FDR 5%) over the timecourse of human myogenic differentiation (clusters 0–5 in the Warner et al. publication) [12]; b) annotated with one or more of the following Gene Ontology (GO) terms: myofibril, sarcomere, contractile fiber, muscle cell differentiation, muscle contraction, or structural constituent of muscle (GO database release 2011-01-06) [46]. The locations of microRNA genes were obtained from miRBase Release 16 [27]. All genomic coordinates are from Hg18.

The WinMine toolkit [25] was used to create a dependency network describing the value of ChIP binding data for certain TFs at 0 or 48 h in

predicting other TF binding at 48 h. The model was learned on a 70% training set of the data using a kappa value of 0.01 and resulted in 63% probability of correctly predicting the values in the remaining (30%) test set of data.

4.5. Identifying predicted CRMs within genomic sequence using PhylCRM

The PhylCRM algorithm was run with default parameters as described previously [12], searching for clusters of conserved MRF binding sites using the MYF motif (representing the specificities of MyoG and MyoD) from a prior study [49]. We chose to use MRF alone for the PhylCRM predictions because this motif was more enriched among a set of genes upregulated during myogenic differentiation than any other combination of MRF, Mef2, SRF, or Tead motifs in a previous study [12]. The phylogenetic tree used to evaluate binding site conservation in PhylCRM included human, chimp, macaque, mouse, rat, dog, cow, and opossum. From evaluation of the PhylCRM window scores for some previously experimentally tested CRMs, we set the threshold PhylCRM window score for calling a cluster of conserved MRF sites a predicted CRM at 1.5.

4.6. Chromosome Conformation Capture (3C)

3C experiments were performed following previously described protocols [6]. Briefly, approximately 5×10^7 human muscle cells were cross-linked in a final concentration of 1% formaldehyde and harvested at each of two timepoints: 48 h before and 48 h after differentiation by serum deprivation. Chromatin extracted from these cells was digested with either an AflIII or BglIII restriction enzyme (New England Biolabs) and then incubated with T4 DNA ligase (New England Biolabs), such that regions of chromosome contact were ligated together. After reversing the formaldehyde crosslinks, physical interactions between genomic regions were detected with PCR primers specific for each ligated interaction product. Primer pairs with melting temperatures of 56–60 °C, within 2 °C of each other, and of 35–50% GC content, were designed to uniquely amplify an approximately 100–300 bp region straddling the location of potential ligation between two restriction fragments. PCR-amplified products were detected by quantitative real-time PCR (qPCR) on an iCycler (BioRad) with SybrGreen Supermix for iQ (BioRad) and standard protocols. The quantification by qPCR was validated by visualization on an agarose gel and quantification using QuantityOne gel image quantification software (BioRad). The qPCR quantification results were utilized for the final analysis.

To control for differences in primer efficiency, DNA fragments generated from bacterial artificial chromosomes (BACs) spanning the genomic regions of interest were digested and randomly ligated (ACTA1: RP11-1111E20, PDUM3/SORBS2: CTD-2559A19, CTD-2194A4, and RP11-78H20). Quantified PCR products from each 3C reaction were then normalized by the quantified results from the same PCR amplifications performed on this BAC control library. For those regions for which an ANOVA test of interaction frequencies within a given genomic region (excluding fragments within 15 kb of the fixed primer, which are likely to show high levels of interactions from random collisions [50]) indicated the presence of a significant peak ($p < 0.05$) within the data, we tested the observed peaks for statistically significant differences in interaction level between timepoints using a Student's *t*-test.

4.7. ChIP-chip data accession number

The raw and processed ChIP-chip microarray data are available in GEO series accession number GSE29865 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29865>).

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2011.08.003.

Acknowledgments

The authors thank Job Dekker and Ye Zhan for providing training in the 3C technique, Jason Warner for assistance with array design and training in ChIP-chip experiments, Cherelle Walls for assistance with cell culture and ChIP-chip experiments, and Anthony Philippakis for assistance with PhylCRM and array design.

This work was supported by NIH/NHGRI grant # R21 HG005149 (M.L.B.). R.P.M. was supported in part by a National Science Foundation Graduate Research Fellowship.

References

- [1] C. Zhu, K.J. Byers, R.P. McCord, Z. Shi, M.F. Berger, D.E. Newburger, K. Saulrieta, Z. Smith, M.V. Shah, M. Radhakrishnan, A.A. Philippakis, Y. Hu, F. De Masi, M. Pacek, A. Rolf, T. Murthy, J. Labaer, M.L. Bulyk, High-resolution DNA-binding specificity analysis of yeast transcription factors, *Genome Res.* 19 (2009) 556–566.
- [2] G. Chua, M.D. Robinson, Q. Morris, T.R. Hughes, Transcriptional networks: reverse-engineering gene regulation on a global scale, *Curr. Opin. Microbiol.* 7 (2004) 638–646.
- [3] E.H. Davidson, *Genomic Regulatory Systems: Development and Evolution*, Academic Press, San Diego, 2001.
- [4] T. Sexton, F. Bantignies, G. Cavalli, Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation, *Semin. Cell Dev. Biol.* 20 (2009) 849–855.
- [5] R.I. Kumaran, R. Thakar, D.L. Spector, Chromatin dynamics and gene positioning, *Cell* 132 (2008) 929–934.
- [6] J. Dostie, T.A. Richmond, R.A. Arnaout, R.R. Selzer, W.L. Lee, T.A. Honan, E.D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R.D. Green, J. Dekker, Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements, *Genome Res.* 16 (2006) 1299–1309.
- [7] Q. Lin, L. Lin, J. Zhou, Chromatin insulator and the promoter targeting sequence modulate the timing of long-range enhancer–promoter interactions in the *Drosophila* embryo, *Dev. Biol.* 339 (2010) 329–337.
- [8] B. Tolhuis, R.J. Palstra, E. Splinter, F. Grosveld, W. de Laat, Looping and interaction between hypersensitive sites in the active beta-globin locus, *Mol. Cell* 10 (2002) 1453–1465.
- [9] C.R. Vakoc, D.L. Letting, N. Gheldof, T. Sawado, M.A. Bender, M. Groudine, M.J. Weiss, J. Dekker, G.A. Blobel, Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1, *Mol. Cell* 17 (2005) 453–462.
- [10] Y. Cao, R.M. Kumar, B.H. Penn, C.A. Berkes, C. Kooperberg, L.A. Boyer, R.A. Young, S.J. Tapscott, Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters, *EMBO J.* 25 (2006) 502–511.
- [11] A. Blais, M. Tsikitis, D. Acosta-Alvear, R. Sharan, Y. Kluger, B.D. Dynlacht, An initial blueprint for myogenic differentiation, *Genes Dev.* 19 (2005) 553–569.
- [12] J.B. Warner, A.A. Philippakis, S.A. Jaeger, F.S. He, J. Lin, M.L. Bulyk, Systematic identification of mammalian regulatory motifs' target genes and functions, *Nat. Methods* 5 (2008) 347–353.
- [13] D.A. Bergstrom, B.H. Penn, A. Strand, R.L. Perry, M.A. Rudnicki, S.J. Tapscott, Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression, *Mol. Cell* 9 (2002) 587–600.
- [14] Y. Cao, Z. Yao, D. Sarkar, M. Lawrence, G.J. Sanchez, M.H. Parker, K.L. MacQuarrie, J. Davison, M.T. Morgan, W.L. Ruzzo, R.C. Gentleman, S.J. Tapscott, Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming, *Dev. Cell* 18 (2010) 662–674.
- [15] C.A. Berkes, S.J. Tapscott, MyoD and the transcriptional control of myogenesis, *Semin. Cell Dev. Biol.* 16 (2005) 585–595.
- [16] P.J. Gianakopoulos, V. Mehta, A. Voronova, Y. Cao, Z. Yao, J. Coutu, X. Wang, M.S. Waddington, S.J. Tapscott, I.S. Skerjanc, MyoD directly up-regulates premyogenic mesoderm factors during induction of skeletal myogenesis in stem cells, *J. Biol. Chem.* 286 (2011) 2517–2525.
- [17] W. Thompson, M.J. Palumbo, W.W. Wasserman, J.S. Liu, C.E. Lawrence, Decoding human regulatory circuits, *Genome Res.* 14 (2004) 1967–1974.
- [18] Q. Sun, G. Chen, J.W. Streb, X. Long, Y. Yang, C.J. Stoeckert Jr., J.M. Miano, Defining the mammalian CArGome, *Genome Res.* 16 (2006) 197–207.
- [19] M.C. Frith, J.L. Spouge, U. Hansen, Z. Weng, Statistical significance of clusters of motifs represented by position specific scoring matrices in nucleotide sequences, *Nucleic Acids Res.* 30 (2002) 3214–3224.
- [20] N.D. Heintzman, R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, B. Ren, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, *Nat. Genet.* 39 (2007) 311–318.
- [21] B.P. Berman, Y. Nibu, B.D. Pfeiffer, P. Tomancak, S.E. Celniker, M. Levine, G.M. Rubin, M.B. Eisen, Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome, *Proc. Natl. Acad. Sci. U.S.A.* 99 (2002) 757–762.
- [22] A. Visel, M.J. Blow, Z. Li, T. Zhang, J.A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E.M. Rubin, L.A. Pennacchio, ChIP-seq accurately predicts tissue-specific activity of enhancers, *Nature* 457 (2009) 854–858.

- [23] J. Ernst, P. Kheradpour, T.S. Mikkelsen, N. Shores, L.D. Ward, C.B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, B.E. Bernstein, Mapping and analysis of chromatin state dynamics in nine human cell types, *Nature* 473 (2011) 43–49.
- [24] M.G. Guenther, S.S. Levine, L.A. Boyer, R. Jaenisch, R.A. Young, A chromatin landmark and transcription initiation at most promoters in human cells, *Cell* 130 (2007) 77–88.
- [25] D.M. Chickering, The WinMine Toolkit, in: Microsoft, , 2002.
- [26] D. Heckerman, D.M. Chickering, C. Meek, R. Rounthwaite, C. Kadie, Dependency networks for inference, collaborative filtering, and data visualization, *J. Mach. Learn. Res.* 1 (2000) 49–75.
- [27] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data, *Nucleic Acids Res.* 39 (2011) D152–157.
- [28] J.F. Chen, E.M. Mandel, J.M. Thomson, Q. Wu, T.E. Callis, S.M. Hammond, F.L. Conlon, D.Z. Wang, The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation, *Nat. Genet.* 38 (2006) 228–233.
- [29] Y. Chen, J. Gelfond, L.M. McManus, P.K. Shireman, Temporal microRNA expression during in vitro myogenic progenitor cell proliferation and differentiation: regulation of proliferation by miR-682, *Physiol. Genomics* 43 (2010) 621–630.
- [30] T. Hosoda, H. Zheng, M. Cabral-da-Silva, F. Sanada, N. Ide-Iwata, B. Ogorek, J. Ferreira-Martins, C. Arranto, D. D'Amario, F. Del Monte, K. Urbanek, D.A. D'Alessandro, R.E. Michler, P. Anversa, M. Rota, J. Kajstura, A. Leri, Human cardiac stem cell differentiation is regulated by a micrine mechanism, *Circulation* 123 (2011) 1287–1296.
- [31] E. van Rooij, D. Quiat, B.A. Johnson, L.B. Sutherland, X. Qi, J.A. Richardson, R.J. Kelm Jr., E.N. Olson, A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance, *Dev. Cell* 17 (2009) 662–673.
- [32] J. Dekker, K. Rippe, M. Dekker, N. Kleckner, Capturing chromosome conformation, *Science* 295 (2002) 1306–1311.
- [33] H. Hagege, P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat, T. Forne, Quantitative analysis of chromosome conformation capture assays (3C-qPCR), *Nat. Protoc.* 2 (2007) 1722–1733.
- [34] P. Pomies, M. Pashmforoush, C. Vegezzi, K.R. Chien, C. Auffray, M.C. Beckerle, The cytoskeleton-associated PDZ-LIM protein, ALP, acts on serum response factor activity to regulate muscle differentiation, *Mol. Biol. Cell* 18 (2007) 1723–1733.
- [35] A. Dean, In the loop: long range chromatin interactions and gene regulation, *Brief. Funct. Genomics* 10 (2011) 3–10.
- [36] N.D. Heintzman, G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee, R.K. Stuart, C.W. Ching, K.A. Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D. Green, V.V. Lobanenkov, R. Stewart, J.A. Thomson, G.E. Crawford, M. Kellis, B. Ren, Histone modifications at human enhancers reflect global cell-type-specific gene expression, *Nature* 459 (2009) 108–112.
- [37] M.B. Gerstein, Z.J. Lu, E.L. Van Nostrand, C. Cheng, B.I. Arshinoff, T. Liu, K.Y. Yip, R. Robilotto, A. Rechtsteiner, K. Ikegami, P. Alves, A. Chateigner, M. Perry, M. Morris, R.K. Auerbach, X. Feng, J. Leng, A. Vielle, W. Niu, K. Rhrissorrakrai, A. Agarwal, R.P. Alexander, G. Barber, C.M. Brdlik, J. Brennan, J.J. Brouillet, A. Carr, M.S. Cheung, H. Clawson, S. Contrino, L.O. Dannenberg, A.F. Dernburg, A. Desai, L. Dick, A.C. Dose, J. Du, T. Egelhofer, S. Ercan, G. Euskirchen, B. Ewing, E.A. Feingold, R. Gassmann, P.J. Good, P. Green, F. Gullier, M. Gutwein, M.S. Guyer, L. Habegger, T. Han, J.G. Henikoff, S.R. Henz, A. Hinrichs, H. Holster, T. Hyman, A.L. Iniguez, J. Janette, M. Jensen, M. Kato, W.J. Kent, E. Kephart, V. Khivansara, E. Khurana, J.K. Kim, P. Kolasinska-Zwierz, E.C. Lai, I. Latorre, A. Leahey, S. Lewis, P. Lloyd, L. Lochovsky, R.F. Lowdon, Y. Lubling, R. Lyne, M. MacCoss, S.D. Mackowiak, M. Mangone, S. McKay, D. Meenas, G. Merrihew, D.M. Miller III, A. Muroyama, J.I. Murray, S.L. Ooi, H. Pham, T. Phippen, E.A. Preston, N. Rajewsky, G. Ratsch, H. Rosenbaum, J. Rozowsky, K. Rutherford, P. Ruzanov, M. Sarov, R. Sasidharan, A. Sboner, P. Scheid, E. Segal, H. Shin, C. Shou, F.J. Slack, C. Slightam, R. Smith, W.C. Spencer, E.O. Stinson, S. Taing, T. Takasaki, D. Vafeados, K. Voronina, G. Wang, N.L. Washington, C.M. Whittle, B. Wu, K.K. Yan, G. Zeller, Z. Zha, M. Zhong, X. Zhou, J. Ahringer, S. Stromer, K.C. Gunsalus, G. Micklem, X.S. Liu, V. Reinke, S.K. Kim, L.W. Hillier, S. Henikoff, F. Piano, M. Snyder, L. Stein, J.D. Lieb, R.H. Waterston, Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project, *Science* 330 (2010) 1775–1787.
- [38] C. Moorman, L.V. Sun, J. Wang, E. de Wit, W. Talhout, L.D. Ward, F. Greil, X.J. Lu, K.P. White, H.J. Bussemaker, B. van Steensel, Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*, *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 12027–12032.
- [39] J.E. Phillips, V.G. Corces, CTCF: master weaver of the genome, *Cell* 137 (2009) 1194–1211.
- [40] O. Hakim, M.H. Sung, T.C. Voss, E. Splinter, S. John, P.J. Sabo, R.E. Thurman, J.A. Stamatoyannopoulos, W. de Laat, G.L. Hager, Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements, *Genome Res.* 21 (2011) 697–706.
- [41] N.D. Heintzman, B. Ren, Finding distal regulatory elements in the human genome, *Curr. Opin. Genet. Dev.* 19 (2009) 541–549.
- [42] J. DeJardin, R.E. Kingston, Purification of proteins associated with specific genomic Loci, *Cell* 136 (2009) 175–186.
- [43] Z. Ni, M. Abou El Hassan, Z. Xu, T. Yu, R. Bremner, The chromatin-remodeling enzyme BRG1 coordinates CITA induction through many interdependent distal enhancers, *Nat. Immunol.* 9 (2008) 785–793.
- [44] I. Rajapakse, M.D. Perlman, D. Scalzo, C. Kooperberg, M. Groudine, S.T. Kosak, The emergence of lineage-specific chromosomal topologies from coordinate gene regulation, *Proc. Natl. Acad. Sci. U.S.A.* 106 (2009) 6679–6684.
- [45] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragozcy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science* 326 (2009) 289–293.
- [46] M.A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G.M. Rubin, J.A. Blake, C. Bult, M. Dolan, H. Drabkin, J.T. Eppig, D.P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J.M. Cherry, K.R. Christie, M.C. Costanzo, S.S. Dwight, S. Engel, D.G. Fisk, J.E. Hirschman, E.L. Hong, R.S. Nash, A. Sethuraman, C.L. Theesfeld, D. Botstein, K. Dolinski, B. Feuerbach, T. Berardini, S. Mundodi, S.Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E.M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, R. White, The Gene Ontology (GO) database and informatics resource, *Nucleic Acids Res.* 32 (2004) D258–261.
- [47] T.I. Lee, S.E. Johnstone, R.A. Young, Chromatin immunoprecipitation and microarray-based analysis of protein location, *Nat. Protoc.* 1 (2006) 729–748.
- [48] M.J. Buck, A.B. Nobel, J.D. Lieb, ChIPOTIE: a user-friendly tool for the analysis of ChIP-chip data, *Genome Biol.* 6 (2005) R97.
- [49] W.W. Wasserman, M. Palumbo, W. Thompson, J.W. Fickett, C.E. Lawrence, Human–mouse genome comparisons to locate regulatory sites, *Nat. Genet.* 26 (2000) 225–228.
- [50] J. Dekker, The three 'C' s of chromosome conformation capture: controls, controls, controls, *Nat. Methods* 3 (2006) 17–21.