

RESEARCH

Open Access

# DNA methylation and differentiation: *HOX* genes in muscle cells

Koji Tsumagari<sup>1</sup>, Carl Baribault<sup>2</sup>, Jolyon Terragni<sup>3</sup>, Sruti Chandra<sup>1</sup>, Chloe Renshaw<sup>1</sup>, Zhiyi Sun<sup>3</sup>, Lingyun Song<sup>4</sup>, Gregory E Crawford<sup>4</sup>, Sriharsa Pradhan<sup>3</sup>, Michelle Lacey<sup>2</sup> and Melanie Ehrlich<sup>1,5\*</sup>

## Abstract

**Background:** Tight regulation of homeobox genes is essential for vertebrate development. In a study of genome-wide differential methylation, we recently found that homeobox genes, including those in the *HOX* gene clusters, were highly overrepresented among the genes with hypermethylation in the skeletal muscle lineage. Methylation was analyzed by reduced representation bisulfite sequencing (RRBS) of postnatal myoblasts, myotubes and adult skeletal muscle tissue and 30 types of non-muscle-cell cultures or tissues.

**Results:** In this study, we found that myogenic hypermethylation was present in specific subregions of all four *HOX* gene clusters and was associated with various chromatin epigenetic features. Although the 3' half of the *HOXD* cluster was silenced and enriched in polycomb repression-associated H3 lysine 27 trimethylation in most examined cell types, including myoblasts and myotubes, myogenic samples were unusual in also displaying much DNA methylation in this region. In contrast, both *HOXA* and *HOXC* clusters displayed myogenic hypermethylation bordering a central region containing many genes preferentially expressed in myogenic progenitor cells and consisting largely of chromatin with modifications typical of promoters and enhancers in these cells. A particularly interesting example of myogenic hypermethylation was *HOTAIR*, a *HOXC* noncoding RNA gene, which can silence *HOXD* genes in trans via recruitment of polycomb proteins. In myogenic progenitor cells, the preferential expression of *HOTAIR* was associated with hypermethylation immediately downstream of the gene. Other *HOX* gene regions also displayed myogenic DNA hypermethylation despite being moderately expressed in myogenic cells. Analysis of representative myogenic hypermethylated sites for 5-hydroxymethylcytosine revealed little or none of this base, except for an intragenic site in *HOXB5* which was specifically enriched in this base in skeletal muscle tissue, whereas myoblasts had predominantly 5-methylcytosine at the same CpG site.

**Conclusions:** Our results suggest that myogenic hypermethylation of *HOX* genes helps fine-tune *HOX* sense and antisense gene expression through effects on 5' promoters, intragenic and intergenic enhancers and internal promoters. Myogenic hypermethylation might also affect the relative abundance of different RNA isoforms, facilitate transcription termination, help stop the spread of activation-associated chromatin domains and stabilize repressive chromatin structures.

**Keywords:** Alternative splicing, DNA methylation, Enhancers, H3K4 trimethylation, *HOTAIR*, *HOX* genes, 5-Hydroxymethylcytosine, Muscle, Myoblasts, Polycomb repression

\* Correspondence: ehrlich@tulane.edu

<sup>1</sup>Hayward Human Genetics Program and Tulane Cancer Center, Tulane Health Sciences Center, New Orleans LA, USA

<sup>5</sup>Center for Bioinformatics and Genomics, Tulane Health Sciences Center, New Orleans, LA, USA

Full list of author information is available at the end of the article

## Background

*HOX* genes are a subset of homeobox genes found in four highly conserved gene clusters on different chromosomes. They encode transcription factors essential for determining the vertebrate body axes during embryonic development and for guiding other aspects of prenatal and postnatal differentiation and postnatal homeostasis [1,2]. Probably as a derangement of these normal roles, *HOX* genes are often hypermethylated in cancer [3]. During embryogenesis, the genes within a given *HOX* cluster are activated sequentially in a collinear manner corresponding to the body plan. Because of their pivotal differentiation-linked roles, *HOX* genes must be regulated in a precise spatiotemporal manner, which makes their cell type-specific epigenetics of particular interest. The collinear activation of *HOX* genes during embryogenesis is mediated by the remodeling of chromatin from a repressive to a transcription-permissive state through changes in histone modifications, especially repressive histone H3 trimethylation at lysine 27 (H3K27me3) and activation-associated H3K4 tri-, di- and mono-methylation (H3K4me3, 2 and 1) [4].

We have been studying the epigenetic markers associated with the skeletal muscle lineage, with emphasis on DNA methylation but also incorporating analysis of chromatin epigenetics. DNA methylation is known to vary markedly among different tissues and cell types [5-9]. Human myoblasts (Mb) are an attractive model for analysis of differentiation because they can be efficiently differentiated into very large, multinucleated, postmitotic myotubes (Mt) *in vitro* and can be compared with skeletal muscle tissue, which is largely derived from such myogenic progenitors. The differentiation of Mb to Mt is relevant not only to the formation of skeletal muscle during embryogenesis but also to postnatal repair of muscle [10].

By reduced representation bisulfite sequencing (RRBS) [6], we recently profiled CpG methylation throughout the genome in the muscle lineage using Mb, Mt and skeletal muscle for comparison to 17 nonmyogenic cell cultures and 14 normal nonmuscle tissues [11]. RRBS, which has single-base resolution, detects approximately 5% of genomic CpGs in a wide variety of sequences, namely, gene bodies and intergenic regions; CpG islands, which account for approximately 50% of RRBS-detected CpGs [6], and nonisland sequences; and single-copy and repeated sequences. Using stringent criteria, we identified differentially methylated CpG sites by comparing the set of myoblasts plus myotubes (MbMt) with many diverse nonmuscle cell cultures derived from normal tissues [11]. We similarly mapped CpGs differentially methylated in skeletal muscle vs. nonmuscle tissue. The RRBS-detected CpG sites in Mb and Mt were much more similar to each other than to other cell lineages. When sites with myogenic differential

methylation were mapped to the nearest gene and then these genes were examined for related functional terms, homeobox genes were found to be one of the most strongly overrepresented classes among the MbMt-hypermethylated genes.

Homeobox genes include the *HOX* genes, which are oriented in the same direction in a given *HOX* gene cluster so that their intracluster location can be referred to as 5' or 3' according to the direction of transcription [12]. This uniform directionality reflects the generation of the archetypal cluster by gene duplication. The ancestral *HOX* gene cluster was in turn replicated to produce four gene clusters. These contain paralogous genes related by sequence similarity and intracluster position and were assigned to the same number group. Paralogous *HOX* genes have many similarities in function but can also display distinct functionality [12,13].

The *HOXA/Hoxa* cluster is implicated in regulating mouse limb bud development (especially *Hoxa9-Hoxa13*) [14]. *Hoxa9* and *Hoxa10* are expressed in the murine C2C12 Mb cell line and in limb muscles during embryogenesis and postnatally, but *Hoxa10* was repressed during muscle regeneration following injury [15]. Targeted disruption of *Hoxa13* increased the level of expression of the myogenic transcription factor MyoD in embryonic murine forelimb [16]. *Hoxa1* coordinates the expression of other *Hoxa* genes in murine embryonic stem cells upon induction by retinoic acid, leading to demethylation of H3K27me3 [17]. *HOXA/Hoxa* genes are expressed in some postnatal lineages, including hematopoietic cells [18], adult lung [19] and endometrium [20]. Unlike *HOXA/Hoxa* genes, *HOXB/Hoxb* genes are not detectably expressed in murine limb muscle during embryogenesis [15]. However, *Hoxb5* is implicated in determining limb positions along the anteroposterior axis [21]. Among other functions, *HOXB/Hoxb* genes are likely to play a role in lung development [19] and hematopoiesis [22].

Murine *Hoxc* genes are also expressed in the skeletal muscle lineage, including *Hoxc12* in embryonic myoblasts [23] and *Hoxc9-Hoxc13* in the embryonic muscle hindlimb, but not in the forelimb [15]. *Hoxc6*, *Hoxc9*, *Hoxc10* and *Hoxc11* are expressed in murine C2C12 Mb and Mt [15] and during the formation of other organ systems, such as the nervous system [24]. Among the postnatal tissues with specific expression of *HOXC/Hoxc* genes are muscle [15], lymphocytes [25], mammary glands [26], skin and keratinocytes [27]. *HOXD/Hoxd* genes, like *HOXA/Hoxa* genes, appear to play especially important roles in limb and digit formation [14,28] as well as in the development of other organs, such as the formation of the terminal regions of the digestive and urogenital tracts [12]. However, *Hoxd11* is expressed in embryonic muscle, but not in postnatal muscle or C2C12 Mb or Mt [15].

Differential expression of *HOX/Hox* genes in a spatially and temporally specific manner is associated with chromatin modification [29-31], expression of ncRNAs (including miRNAs) in cis or trans [32-34], long-distance enhancers outside the *HOX* clusters as well as local enhancers [35] and three-dimensional chromatin architecture [4,36]. Studies of specific *HOX/Hox* genes have revealed tissue-specific DNA methylation, which is likely to help lock in complicated expression patterns for *HOX* genes and possibly help to establish these expression patterns [37-40]. In a whole-genome analysis of DNA methylation, the four *HOX* gene clusters were found to be hypomethylated in human embryonic stem cells (ESCs) relative to fibroblast-like derivatives of ESC, neonatal foreskin fibroblasts and blood monocytes [41]. To the best of our knowledge, the present study is the first to use single-base resolution profiling of DNA methylation to investigate all the *HOX* clusters in a wide variety of normal cell cultures and tissues. We also correlated DNA epigenetic differences with differential chromatin epigenetics and gene expression. We found that the variety of *HOX* genes' functions is reflected in their developmentally associated DNA methylation patterns, which had diverse relationships with gene expression.

In addition, we examined whether DNA hypermethylation in myogenic progenitor cells involves 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC) because they cannot be distinguished by RRBS or most other types of DNA methylation analysis [42]. In mammalian DNA, 5hmC is the sixth genetically programmed base. It is usually very much less abundant than 5mC and serves as an intermediate in DNA demethylation as well as a stable DNA base [43,44]. Increases in 5hmC and decreases in 5mC have been reported in *HOXA1* and *HOXA2* upon induction of differentiation of the NT2 embryonal carcinoma cell line by retinoic acid, which derepresses *HOX* genes in a collinear manner [45]. Discriminating between 5mC and 5hmC is important because they seem to typically play very different roles in the control of gene expression, usually repression at cis-acting transcription control elements for 5mC and activation at enhancers for 5hmC [42,46]. Therefore, we quantified 5mC and 5hmC at five representative CpG sites in the four *HOX* clusters of muscle and nonmuscle samples by enzymatic assay.

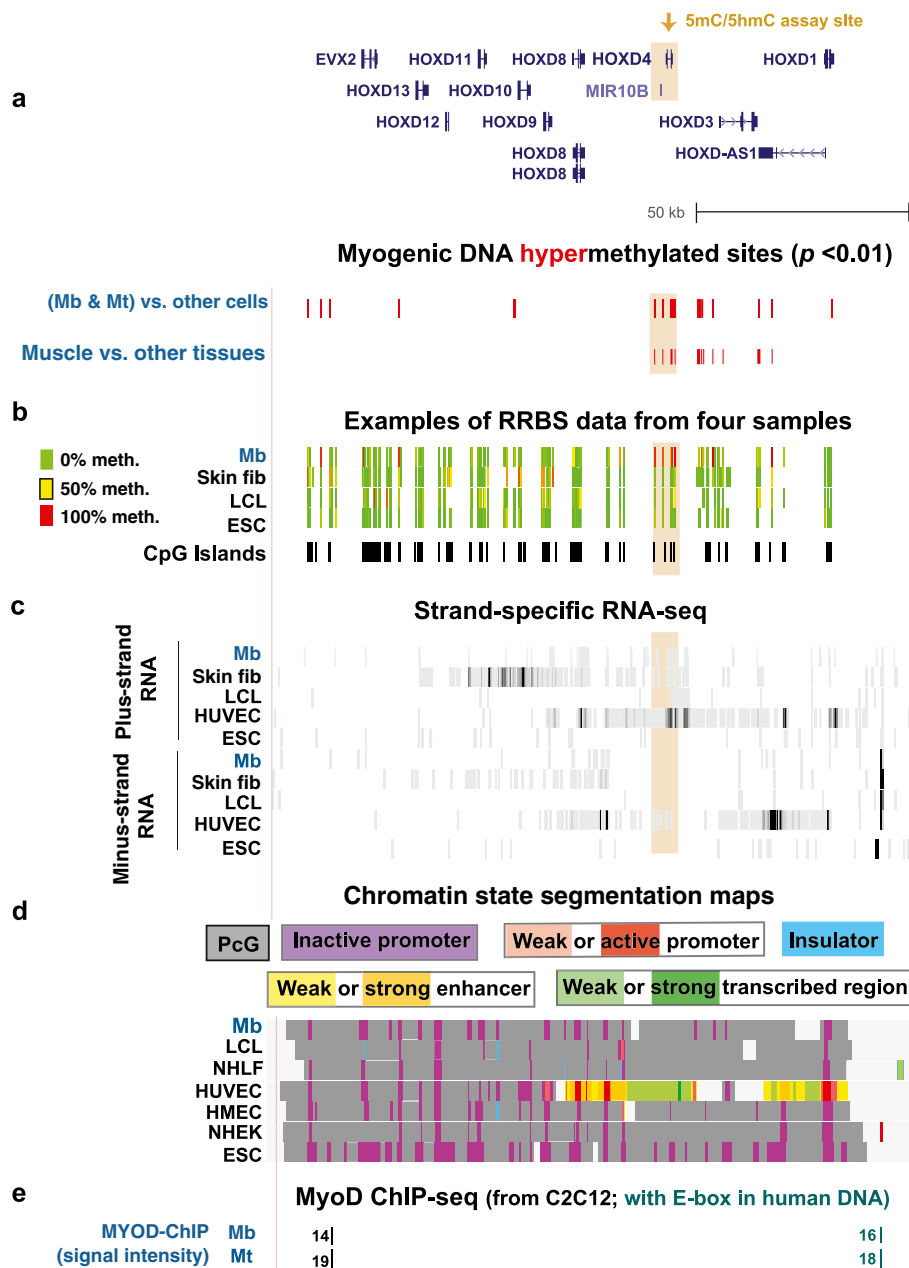
## Results and discussion

### Myogenic DNA hypermethylation at *HOXD* genes vs. H3K27me3 in many cell types

To identify myogenic differential methylation in *HOX* gene clusters, we analyzed RRBS data from the ENCODE project ([11]; <http://genome.ucsc.edu/>; DNA methylation by RRBS; HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA). The methylome profiles that we used were generated from our Mb and Mt samples plus 16 other types of cell culture and skeletal muscle plus 14 types

of normal tissue. The Mb samples were derived from biopsies, and aliquots were differentiated to Mt. Importantly, all had been characterized immunocytochemically as previously described [11]. The nonmuscle cultures were untransformed cells, with the exception of lymphoblastoid cell lines (LCLs). We determined significant myogenic hypermethylation or hypomethylation using stringent criteria, namely, at least a 50% difference in methylation in Mb and Mt (as a set, MbMt) vs. the nonmyogenic cell cultures or in skeletal muscle tissue vs. nonmuscle tissue at a significance level of  $P < 0.01$  using fitted binomial regression models at each monitored CpG site [11]. This analysis involved our recently developed algorithm that adjusts single-site  $P$  values for coverage score and sample size. We then plotted the sites with myogenic differential methylation to the nearest gene and subgene region as illustrated for *HOX* genes in Additional file 1. All our references to myogenic differential methylation met the above requirements for statistical significance.

In the *HOXD* gene cluster, many sites were hypermethylated in the MbMt set vs. nonmuscle cell cultures or in skeletal muscle tissue vs. nonmuscle tissues, as shown in Figure 1a. Figure 1b displays the coverage of RRBS in this region by exhibiting DNA methylation data tracks from the UCSC Genome Browser for representative samples. One of the subregions with the most myogenic hypermethylation in both progenitor cells and tissues was in the vicinity of *HOXD4* and had 38 MbMt-hypermethylated sites and 33 skeletal muscle-hypermethylated sites (Figure 1a, tan highlighting, and Additional file 2). The two clusters of MbMt-hypermethylated sites in the *HOXD4* upstream region surround a retinoic acid-sensitive mesodermal enhancer [47] and are near the adjacent *MIR10B* gene (Figure 1), whose methylation was implicated in gene silencing in cis in gastric cancer [32]. Both DNA methylation and H3K27me3 were seen at the *MIR10B* promoter region in human mammary epithelial cells (HMEC) in a previous study [48] as well as in the present study (Figure 1 and Additional file 2). Our analysis of RNA-seq data (ENCODE/California Institute of Technology; <http://genome.ucsc.edu/>; [49]) by Cufflinks [50], a program which evaluates RNA-seq profiles to determine steady-state amounts of different RNA isoforms, indicated that human umbilical vein endothelial cells (HUVEC) expressed this gene abundantly, whereas less than 200 times as much *HOXD4* RNA was detected in Mb, epidermal keratinocytes (NHEK), lung fibroblasts (NHLF), ESC and an LCL (Additional file 1). Only HUVEC did not have the repressive polycomb group chromatin marks at *HOXD4* and throughout most of the *HOXD* cluster (Figure 1d). However, the predominant, 5.1-kb HUVEC transcript began upstream of *HOXD4* near the *MIR10B* gene and extended past the 3' end of *HOXD4*. A second,



**Figure 1 Myogenesis-associated hypermethylation in the 3' half of the *HOXD* gene cluster, which displayed polycomb silencing in most cell types.** (a) Red bars, the 55 CpG sites significantly hypermethylated in Mb plus Mt vs. 16 types of non-muscle-cell cultures and 61 CpG sites significantly hypermethylated in skeletal muscle tissue vs. 14 types of nonmuscle tissues in the chr2:176,921,692 -177,074,604 region. At this scale, many differentially methylated sites cannot be discriminated. (b) Examples of RRBS data (a). Using an 11-color semicontinuous scale (see color guide), these tracks indicate the average DNA methylation levels at each monitored CpG site from the quantitative sequencing data (ENCODE/HudsonAlpha Institute for Biotechnology). Data are shown for only a few of the cell culture samples evaluated for this study. Skin fib, neonatal foreskin fibroblasts. (c) Strand-specific RNA-seq profiling at the *HOXD* gene cluster for Mb, neonatal foreskin fibroblasts, HUVEC and ESC. Each track displays the signal from RNA-seq (ENCODE/Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA) from these cell cultures. The vertical viewing range for the strand-specific RNA-seq was 1 -100 in the UCSC Genome Browser for this and subsequent figures unless otherwise noted. Tan highlighting, the *HOXD4* region shown in Additional file 2. (d) The predicted type of chromatin structure in subregions of the *HOXD* gene cluster is displayed in chromatin state segmentation maps (ENCODE/Broad Institute, Cambridge, MA, USA) based mostly on histone modifications [54]. The predicted local chromatin states are shown with the indicated colors. PcG, polycomb group protein -associated H3K27me3. (e) MyoD binding from C2C12 ChIP-seq [59] and identification of orthologous human sequences. The relative binding strength is indicated, and sites shown in blue overlapped CAGCTG, which is present in approximately 75% of MyoD ChIP-seq peaks and is part of the degenerate consensus sequence for MyoD binding [59].

noncoding transcript was seen in HUVEC (ENST00000465649), whose transcription begins within the single *HOXD4* intron. The myogenic intragenic hypermethylated sites in *HOXD4* surround or overlap this alternative transcription start site (TSS; pink triangle, Additional file 2). Myogenic hypermethylation of the intron might help suppress the use of a secondary, intronic promoter.

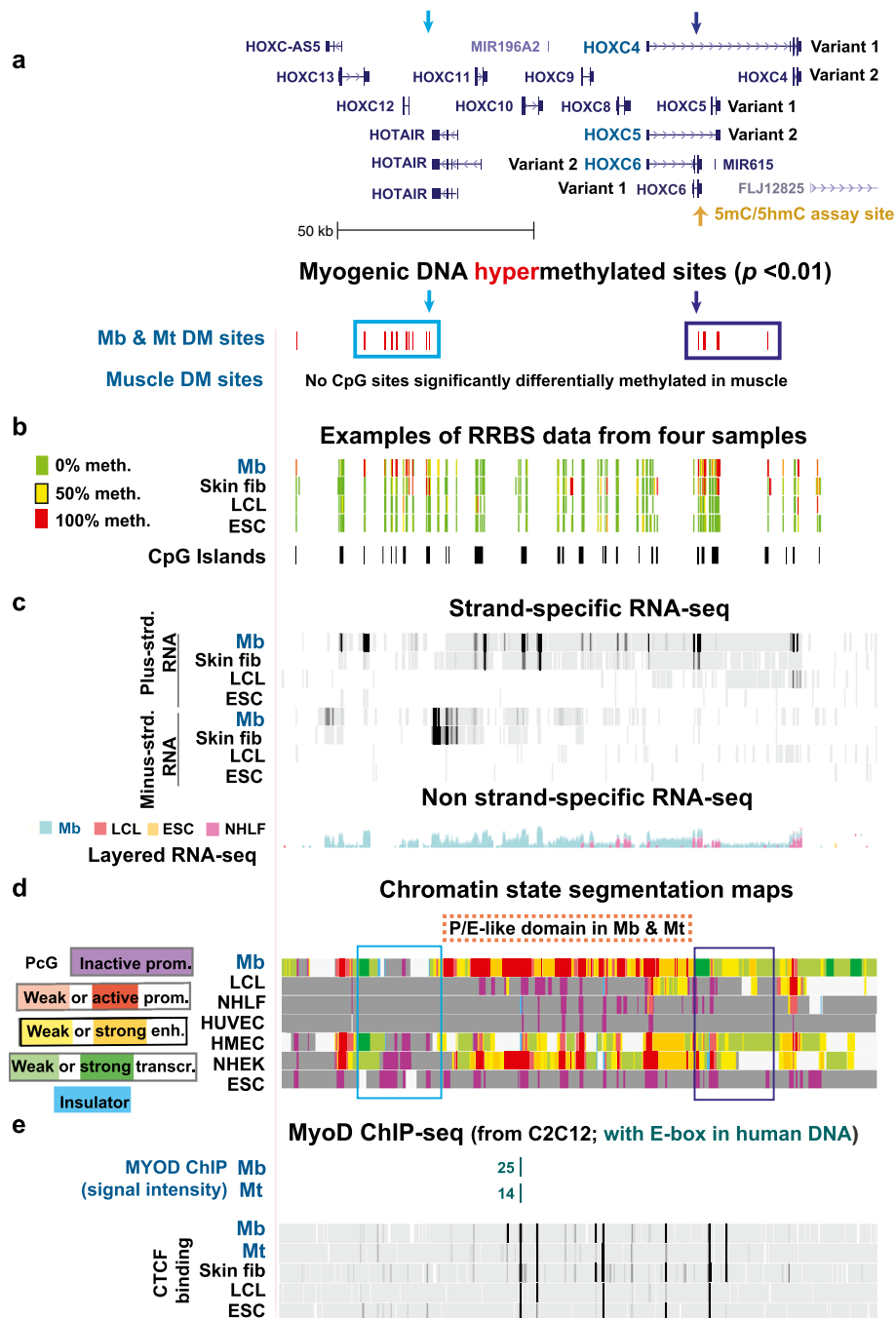
Not only was *HOXD4* silent in most of the examined cells types, including Mb, but this was also the case for the rest of the *HOXD* cluster, especially the 3' half of the cluster (Figure 1c). Similarly, there was silencing-associated H3K27me3 throughout the gene cluster in Mb, Mt and most examined nonmyogenic cell types (Figure 1d, PcG, and Additional file 3) as determined by whole-genome chromatin immunoprecipitation/next-generation DNA sequencing (H3K27me3 ChIP-seq; ENCODE/Broad Institute, <http://genome.ucsc.edu/>). There was an unusually high concentration of CpG islands in the *HOXD* cluster and the other three *HOX* clusters (Figures 1, 2, 3, 4, 5 and 6), but this cannot explain the myogenic hypermethylation in *HOX* gene clusters. For example, there was a much higher density of MbMt-hypermethylated sites in the 3' half of the *HOXD* gene cluster relative to the 5' half, but not a higher density of CpG islands (Figure 1a).

An important question is raised by our finding of much myogenesis-associated DNA hypermethylation in the 3' half of the *HOXD* cluster while H3K27me3 was seen throughout this region in most examined cell types, including Mb and Mt. Why did the many examined populations of myogenic cells display DNA hypermethylation in this multigene region compared with other cell types, even though the myogenic and nonmyogenic cells shared polycomb silencing which might have sufficed for repression of genes in this region [51]? Although regions of DNA methylation and H3K27me3 sometimes overlap, the relationships between these two epigenetic markers are varied and region-specific [52]. Our findings could be explained most easily by the hypothesis that, for the 3' half of *HOXD*, polycomb group silencing at the chromatin level does not suffice for repression of 3' *HOXD* genes in Mb and Mt, and, specifically in these cells, H3K27me3 needs to be supplemented with DNA methylation. Without the DNA hypermethylation, myogenic progenitor cells might be more susceptible to leaky expression of 3' *HOXD* genes than are most other cell types. Alternatively, *HOXD*-encoded proteins or ncRNAs generated from the 3' half of the cluster might be deleterious specifically to myogenic progenitor cells. Consistent with a combined role of DNA methylation and H3K27me3 in some types of *HOX/Hox* gene regulation, recently it was shown that experimentally induced DNA hypomethylation in mouse embryonic fibroblasts led to decreased H3K27me3 at *Hox* genes, including genes in

the 3' half of the *Hoxd* gene cluster [53]. Some *Hox* genes were shown to be derepressed upon DNA demethylation. Our study suggests that the roles of DNA methylation of *HOX* genes during development are more nuanced than can be seen in a study of one cell type because NHLF (IMR90), ESC and LCL samples exhibited much H3K27me3 in the 3' half of the *HOXD* cluster despite very little DNA methylation there (Figures 1b and 1d, Additional file 3). In contrast, Mb and Mt displayed both H3K27me3 and much DNA methylation in this region.

#### **Myogenic hypermethylation in the *HOXC* cluster bordering an H3K4me3-rich multigenic region**

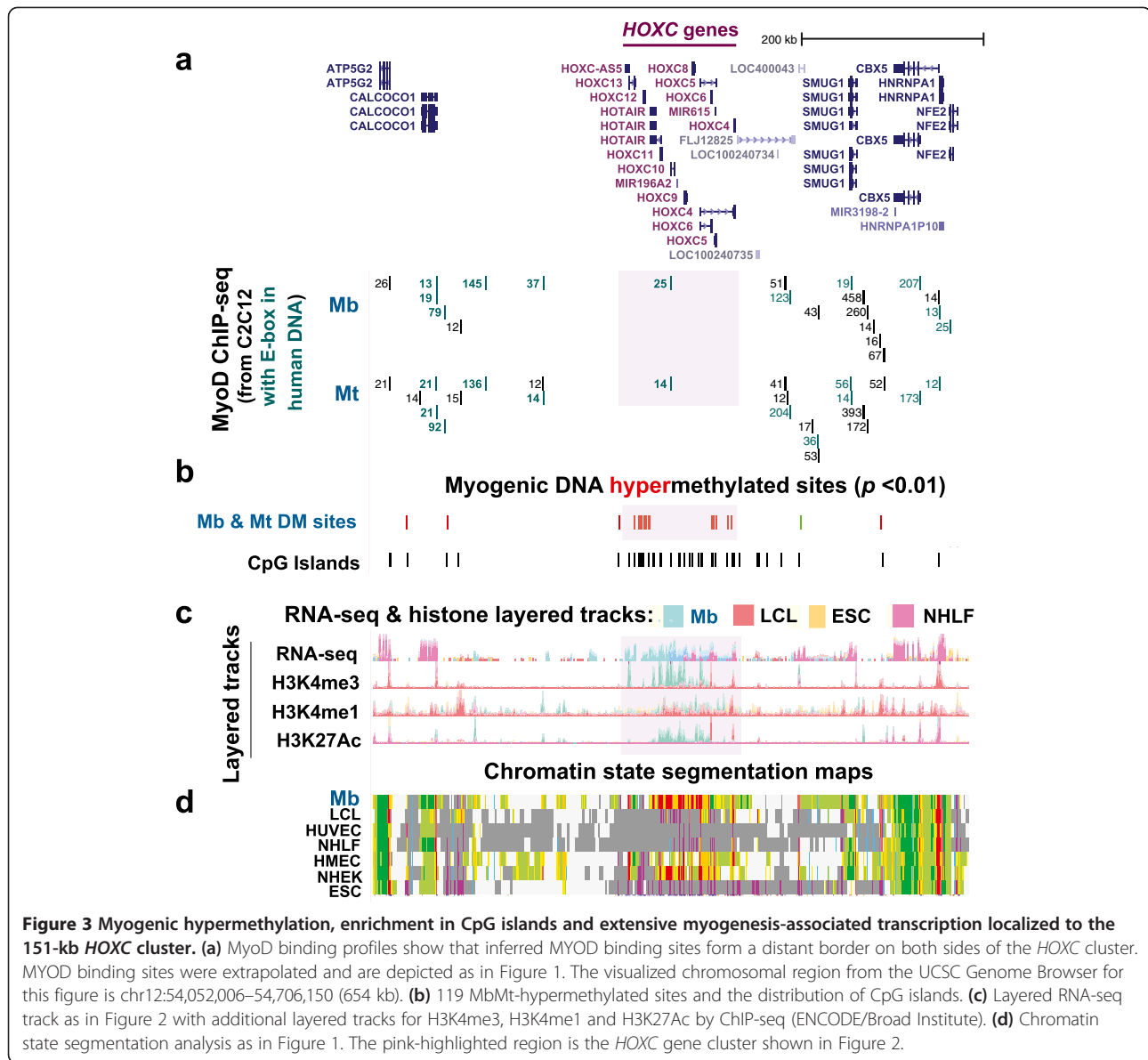
MbMt-hypermethylation was also seen in the *HOXC* cluster (Figure 2a), but, unlike *HOXD* genes, many *HOXC* genes were moderately or strongly expressed in Mb and Mt but not in NHLF, LCL, ESC and HUVEC samples (ENCODE/RNA-seq, California Institute of Technology and Cold Spring Harbor Laboratory; Figure 2c and Additional file 1). Foreskin fibroblasts were the other examined cell type that displayed considerable expression of *HOXC* genes, although less than for Mb (Figure 2c, skin fib). Figure 2d shows a distillation of ChIP-seq chromatin epigenetic data (H3K4me1, 2 or 3; H3K27Ac; H3K9Ac; H3K27me3; H3K36me3; H4K20me1; and CTCF binding) by Ernst *et al.* using a multivariate hidden Markov model (ENCODE/Broad Institute; [54]) to predict chromatin states (color-coded chromatin state segmentation maps). In much of the central, multigenic region of the *HOXC* cluster in Mb and Mt, the chromatin state segmentation map shows chromatin with the features of a strong promoter, especially a strong signal for H3K4me3. The H3K4me3 was present in broad intragenic and intergenic subregions in Mb (Figure 2d, Mb, red subregions), as was found for transcribed *HOX* gene clusters in murine embryonic fibroblasts [30]. This active promoter-like (or active enhancer-like [55]) chromatin rich in H3K4me3 in the middle of the *HOXC* gene cluster was interspersed with a type of chromatin typically associated with active enhancers (H3K27Ac plus H3K4me1; Figure 2d, Mb, orange subregions). We refer to such a multigenic region consisting largely of chromatin with the typical characteristics of active promoters and enhancers as a P/E-like domain. The P/E-like domain probably reflects, in part, the high density of ncRNA genes, including undocumented ones, and alternative transcription start sites within *HOX* gene clusters [56-58]. This P/E domain is also predicted to contain a MYOD binding site (Figure 2e) because it contains a sequence which is orthologous to a genomic sequence in C2C12 mouse Mb and Mt samples that bound MyoD in MyoD-ChIP-seq profiles [59]. Moreover, this site in the human genome has a centrally located CAGCTG E-box, which is found in many MYOD/MyoD binding sites [59].



**Figure 2** Myogenic hypermethylated sites at both ends of the *HOXC* gene cluster, which was preferentially transcribed in myogenic cells. (a) 119 CpG sites with significant hypermethylation in Mb plus Mt vs. 16 types of non-muscle-cell cultures at chr12:54,318,064–54,468,880. (b) Representative RRBS tracks with the location of CpG islands beneath them. (c) Strand-specific RNA-seq profiling (as in Figure 1) for the *HOXC* gene cluster and standard RNA-seq (not strand-specific; ENCODE/California Institute of Technology). The layered RNA-seq shows the superimposed profiles from Mb, LCL, ESC and NHLF cells in the indicated color code. (d) Chromatin state segmentation analysis as in Figure 1. (e) MyoD binding site profiles as in Figure 1 and CTCF binding from ChIP-seq profiling of the indicated cell types (ENCODE/Broad Institute). Arrows and empty boxes denote features mentioned in the text.

Within a 650-kb region centered over the approximately 130 kb *HOXC* cluster, Mb displayed a P/E-like domain only at the *HOXC* cluster, and this cluster was the most prominent gene region with myogenesis-associated

gene expression (Figures 3c and 3d). Many strong C2C12-inferred MYOD sites were located outside the *HOXC* cluster (Figure 3a). We hypothesize that these may be part of long-distance, tissue-specific *HOX* enhancers, like those

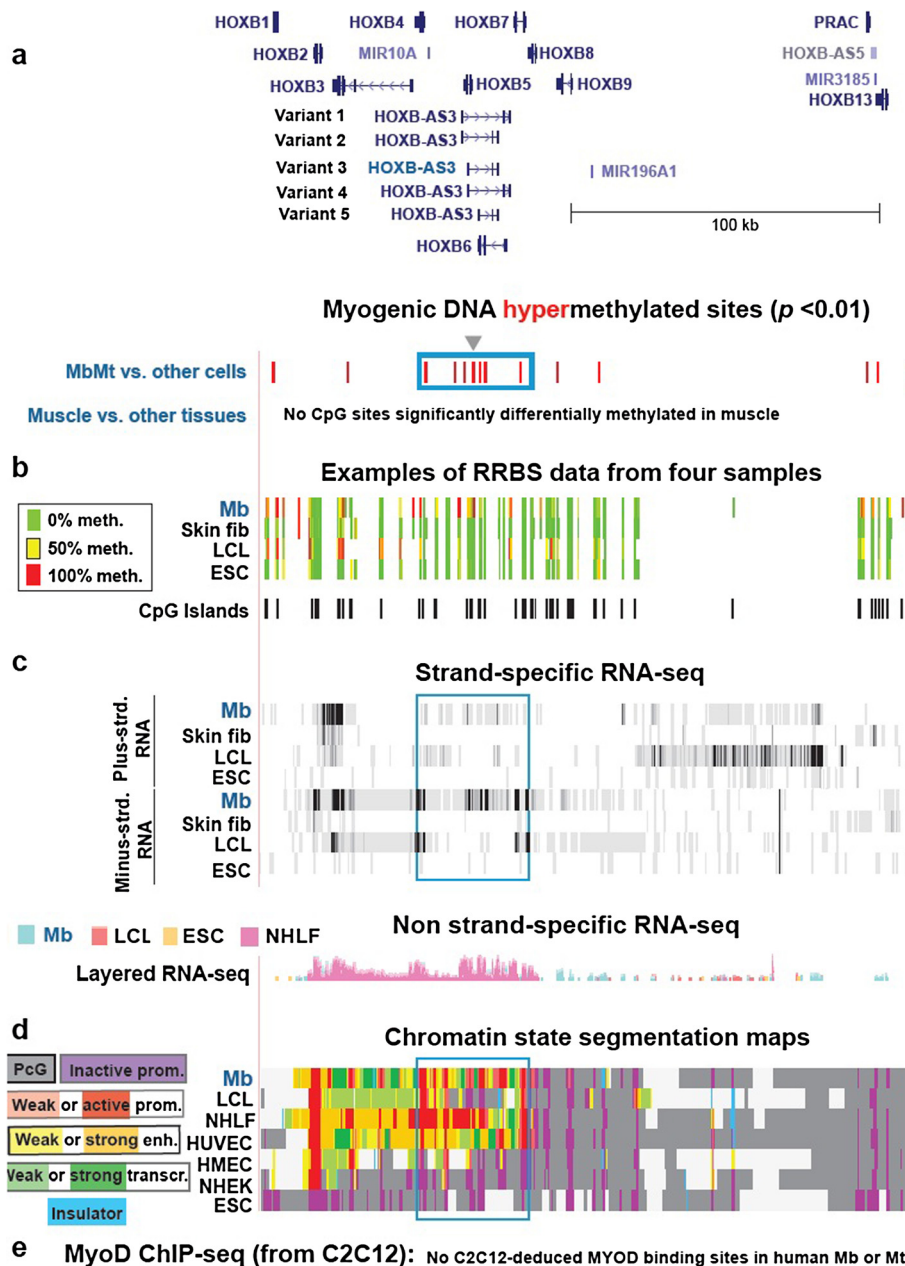


previously described [35], or may help organize long-range chromatin structure around the *HOXC* cluster. Moreover, as for all the *HOX* clusters, the *HOXC* region with its high concentration of CpG islands and RRBS-detected, MbMt-hypermethylated sites was surrounded by DNA that had a low density of both (Figure 3b).

Many of the MbMt-hypermethylated sites within the *HOXC* gene cluster (Figure 2a, blue boxes) surround the myogenesis-associated P/E-like domain. Their location relative to chromatin epigenetic marks suggests that they are part of a boundary element preventing the spread of the central P/E-like domain and the associated high levels of expression into the adjacent chromatin (Figures 2a and 2d and Additional file 4). This hypothesis would be consistent with observed negative relationships between

DNA methylation and H3K4 methylation [60]. CTCF sites often function as boundary elements or insulators [61]. There were no strong CTCF sites at the 5' end of the P/E-like domain in Mb, and there was only a constitutive CTCF site near the other end (Figure 2e, bottom, and Additional file 4).

The cluster of myogenic DM sites at the 3' border of the P/E-like domain (Figure 2a, dark blue arrow) overlapped a CpG island in *HOXC4* variant 1 (intron 1) and *HOXC6* variant 2 (last exon; Figure 3a and ENCODE/RNA-seq, California Institute of Technology). These genes were both preferentially expressed in Mb and share the same TSS. *HOXC5* variant 2 shares this TSS, too, but it had no detectable transcripts according to RNA-seq data (ENCODE/California Institute of Technology and Cold



**Figure 4** Myogenic hypermethylation in the central region of the *HOXB* gene cluster, which is preferentially transcribed in myogenic cells. **(a)** 88 MbMt-hypermethylated sites in the chr17:46,602,904–46,814,469 region. **(b)** Examples of RRBS data. **(c)** Strand-specific RNA-seq as in Figure 1, except that the vertical viewing ranges were 1–10 for the plus strand and 1–100 for the minus strand. **(d)** Chromatin state segmentation analysis. **(e)** The MyoD binding site track shows no C2C12-extrapolated MYOD sites in this region. Arrows, empty boxes and the triangle denote features mentioned in the text.

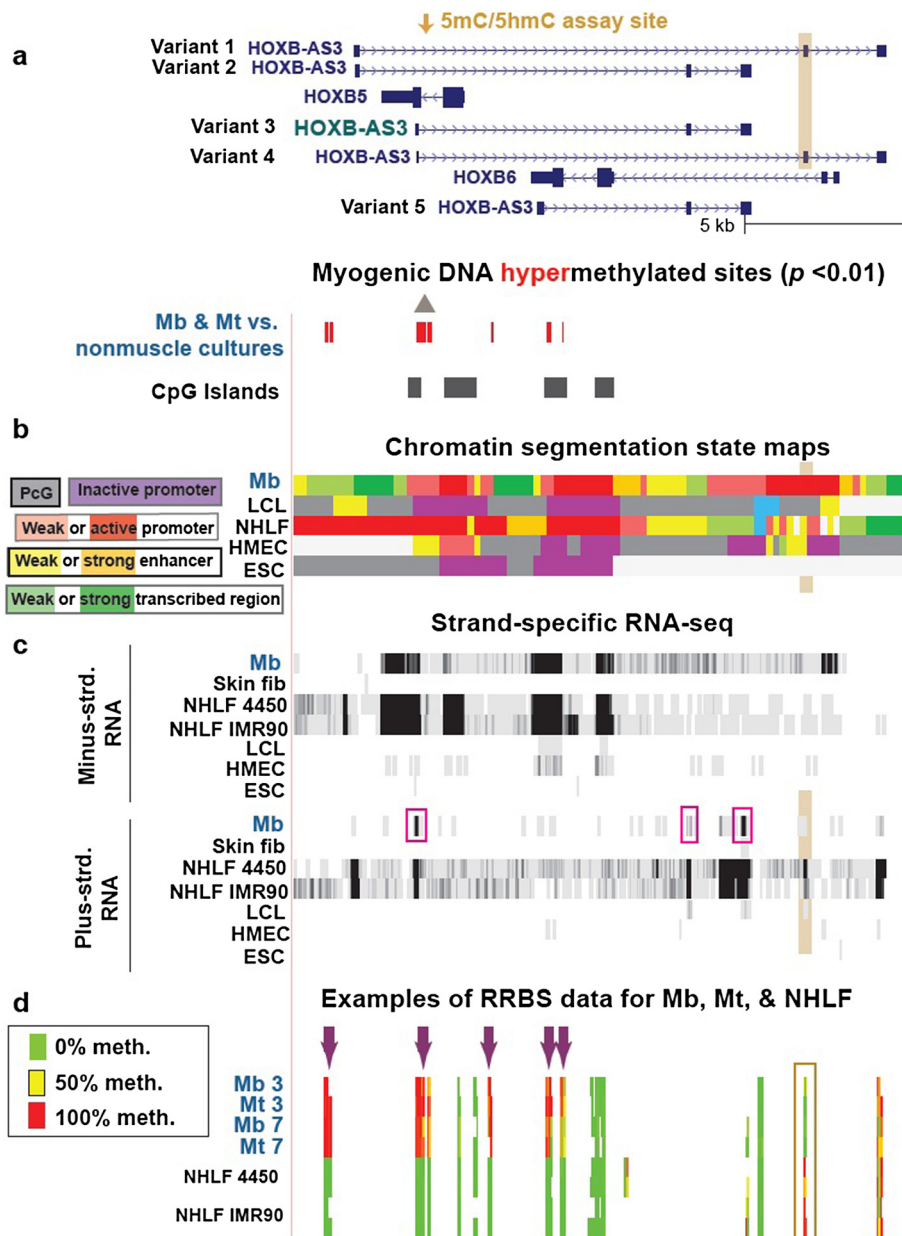
Spring Harbor Laboratory). Differential splicing will help determine relative expression of these overlapping *HOXC4*, *HOXC5* and *HOXC6* genes. Because DNA methylation can affect the relative steady-state levels of RNA by modulating the rate of progression of the RNA polymerase II (RNA Pol II) complex in diverse ways [62,63], we hypothesize that the 41 intragenic MbMt-hypermethylated DNA sites in these three overlapping genes help regulate differential

splicing in this region through effects on RNA Pol II elongation. Such effects of intragenic DNA methylation are likely to be gene-specific and/or cell type-specific [64,65].

#### Myogenic hypermethylation downstream of *HOTAIR*

Near the 5' end of the *HOXC* P/E-like domain in Mb, there were eight MbMt-hypermethylated sites in a CpG island approximately 1 kb downstream of *HOTAIR*, a

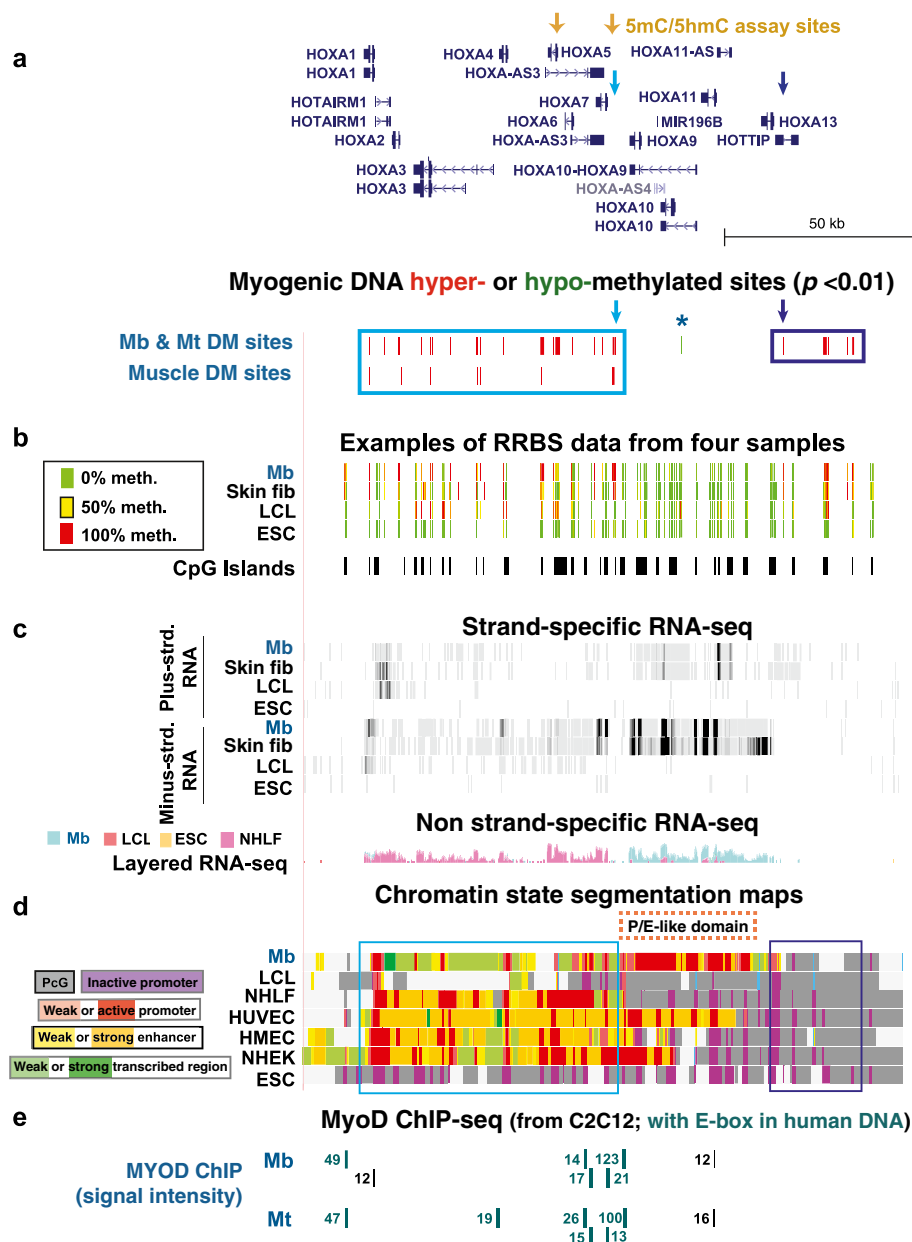




**Figure 5** Cell type-specific differences in DNA methylation and transcription in the region containing *HOXB5*, *HOXB6* and *HOXB-AS3* variant genes. **(a)** 42 MbMt-hypermethylated sites in a subregion of *HOXB* (chr17:46,665,998–46,684,371). **(b)** Chromatin segmentation state maps. **(c)** Strand-specific RNA-seq as in Figure 4. The pink boxes indicate the RNA-seq evidence for *HOXB-AS3* variant 3 as the predominant variant expressed in Mb. **(d)** RRBS data for two control Mb cell strains and Mt preparations derived from them, as well as two fetal lung fibroblast cell strains analyzed as technical duplicates. Arrows and highlighted subregions are described in the text.

long noncoding RNA (lncRNA) gene (Figure 2a, light blue arrow, and Additional file 4). Mb displayed a moderate level of expression of *HOTAIR*, whereas there was little or no expression in most of the other studied cell types, with the exception of foreskin fibroblasts (Figure 2c and Additional file 1), which were highly methylated in the *HOTAIR* downstream region like Mb and Mt (Figure 2b). Hypermethylation of *HOTAIR*'s 3' downstream CpG island

was also seen by Lu *et al.* [66] in breast cancer and correlated with expression. The MbMt hypermethylation downstream of *HOTAIR* included the 3' half of *HOXC12* (Figure 2 and Additional file 4). Lu *et al.* proposed that one of the functions of *HOTAIR* downstream methylation was to facilitate transcription termination at *HOXC12*. They reported no good matches in this region to DNA poly(A) signals [66] as detected by a program for predicting



**Figure 6** Peripheral myogenic hypermethylation and a central myogenic hypomethylated site in the *HOXA* gene cluster. **(a)** 187 MbMt-hypermethylated and 20 muscle-hypermethylated sites as well as one MbMt-hypomethylated site in the chr7:27,116,782–27,273,459 region. **(b)** Examples of RRBS data. **(c)** RNA-seq profiles as in Figure 1. **(d)** Chromatin state segmentation analysis. **(e)** MyoD binding sites from C2C12.

optimal AATAAA poly(A) termination signals [67]. However, we found that the program indicated two individually low-rated poly(A) signals, 9 bp apart, 2.5 kb downstream from the 3' end of the RefSeq *HOXC12* sequence. We noted that the sense transcript from *HOXC12* extends approximately to these two poly(A) signals downstream of the canonical RefSeq sequence (Additional file 4, RNA-seq, orange triangle).

We propose that DNA methylation in this *HOXC* sub-region not only acts as part of a boundary element but

also facilitates transcription termination through RNA Pol II pausing at *HOXC12*. This in turn may favor expression of the oppositely oriented *HOTAIR*. In a study of mouse *Hoxc6* and *Hoxc8* in embryonic fibroblasts, Tao *et al.* provided several lines of evidence for DNA methylation causing repression at the transcription elongation step due to long-lived pausing of RNA Pol II [63]. They showed that the effect of demethylation on *Hox* transcription was tissue-specific and specific to individual *Hox* genes [65]. Consistent with the results of Tao *et al.*,

the P/E-like domain in *HOXC* in Mb includes the *HOXC6* promoter region and *HOXC8*, both of which were mostly or completely unmethylated and were upregulated in Mb and Mt vs. other cell types (Figure 2, Additional file 1, and data not shown).

*HOTAIR* RNA in trans represses genes across the whole *HOXD* gene cluster by recruiting chromatin-modifying polycomb group proteins, which results in extensive H3K27 trimethylation of the *HOXD* cluster [68]. We hypothesize that the preferential expression of *HOTAIR* that we found in myogenic progenitor cells is partly responsible for their *HOXD* DNA hypermethylation. This would be consistent with the recent report that knockdown of *HOTAIR* caused a decrease in DNA methylation of the promoter region of the *PTEN* gene in laryngeal squamous carcinoma cells [69].

#### Hypermethylation and transcription in a *HOXB* subregion of myogenic progenitor cells

The *HOXB* cluster, unlike the *HOXC* and *HOXD* clusters, displayed most of its MbMt hypermethylation in a subregion with considerable gene expression in myogenic progenitor cells, namely, the subregion containing *HOXB4*, *HOXB5*, *HOXB6* and *HOXB7*, and *HOXB-AS3* (Figures 4a–4c and Additional file 1). There was even higher expression of these genes in HUVECs or NHLFs. *HOXB-AS3* transcripts in Mb were mostly variant 3 (Figure 5c, pink boxes). There were 20 MbMt-hypermethylated sites from approximately 40 to 400 bp after the TSS of *HOXB-AS3* variant 3, which overlapped the single intron and last exon of *HOXB5* as well as a CpG island (Figures 4a and 5a, gray triangle). The higher levels of expression of *HOXB5* and several variant *HOXB-AS3* transcripts in NHLF vs. Mb were paralleled by a lack of methylation in this subregion in lung fibroblasts at sites that were hypermethylated in Mb and Mt (Figure 5d, purple arrows). The opposite DNA methylation pattern was seen at exon 2 in *HOXB-AS3* variants 1 and 4, where NHLF displayed DNA methylation, whereas Mb, Mt and most other cell types had little or no methylation (Figure 5, tan highlighting and data not shown). These findings could be explained most easily by the following hypothesis. DNA methylation in Mb close to the *HOXB-AS3* variant 3 TSS and further upstream may be downmodulating its transcription and suppressing transcription of the other *HOXB-AS3* variants in Mb, whereas DNA methylation at exon 2 in NHLF might control splicing of *HOXB-AS3* transcripts specifically in those cells. Moreover, the results from NHLF suggest that antisense *HOXB-AS3* transcription favors the sense *HOXB5* expression, as indicated by other studies of antisense vs. sense genes in *HOX* gene clusters [58,70]. *HOXB5* expression might be fine-tuned by the effects of differential methylation on the level of transcription of overlapping *HOXB-AS3* gene isoforms.

#### Myogenic hypermethylated sites may serve as a boundary at the end of highly transcribed *HOXA* chromatin in myogenic cells

Like the *HOXC* cluster, the *HOXA* cluster displayed much hypermethylation on both sides of a P/E-like domain in Mb and Mt (Figures 6a and 6d, boxed regions). Genes in the *HOXA* P/E-like domain, including *HOXA9*, *HOXA10*, *HOXA11* and *HOXA11-AS*, were preferentially expressed in Mb and HUVEC vs. NHLF, an LCL, and ESC (ENCODE/RNA-seq, California Institute of Technology) and were expressed at higher levels in Mb than were genes bordering this domain, namely, *HOXA7* and *HOXA13* (Figure 6c and Additional files 1 and 5). Consistent with the findings for Mb and HUVEC, 5' *Hoxa* genes are involved in the skeletal muscle and endothelial cell lineages during mouse embryo development and in the adult mouse [15,71]. Between *HOXA7* and *HOXA9* was a cluster of 15 MbMt-hypermethylated sites and four muscle-hypermethylated sites (Figure 6a, light blue arrow, and Additional file 5). Surrounding this subregion was the highest concentration of orthologous sites for MyoD binding in the *HOX* gene clusters (ChIP-seq profiles of C2C12 mouse Mb and Mt [59]), and all of these contained centrally located, MYOD/MyoD-like CAGCTG E-box sequences (Figure 6e and Additional file 5). We propose that the clusters of MbMt-hypermethylated sites here and at the other border of the P/E-like domain help establish the boundaries of this myogenesis-associated domain either alone or in conjunction with nearby constitutive CTCF binding sites (Additional file 5 and data not shown).

#### Myogenic hypomethylation in *HOXA* and extensive undermethylation in ESC and several nonembryonic cell types

The only MbMt-hypomethylated site found in the *HOX* gene clusters was located in the middle of the MbMt-associated P/E-like domain of the *HOXA* cluster (Figure 6a and Additional file 5, asterisk). This site is 1.7 kb upstream of the protein-encoding isoform of *HOXA10* and inside the single intron of the ncRNA-encoding isoform of this gene. *Hoxa10* is implicated in limb muscle development and expressed in murine hindlimb progenitor muscle cells from neonatal muscle [72]. Strand-specific RNA-seq indicates that both the lncRNA and mRNA isoforms of *HOXA10* were expressed in Mb and HUVEC (Figure 6, and data not shown). The MbMt-hypomethylated site may be part of an extended myogenesis-associated enhancer for the *HOXA10* gene in the P/E-like domain.

One of the sample types with the least DNA methylation throughout the *HOX* clusters was ESC (Figures 1, 2, 4 and 6). Moreover, *HOX* clusters in ESC had less DNA methylation than fibroblasts and monocytes [41]. This exceptional lack of *HOX* DNA methylation was also seen for astrocytes, choroid plexus epithelial cells, iris pigment

epithelial cells and retinal pigment epithelial cells (data not shown). The similar *HOX* DNA epigenetics of these four cell types is probably due to their common derivation from the neuroectoderm.

#### Similarities and differences in methylation of paralogous *HOX* genes and comparison of Mb and ESC epigenetic marks

*HOX* clusters provide the opportunity to compare the epigenetics of paralogous sets of genes. Paralog group 4 *HOX* genes all had RRBS data. Of these genes, *HOXA4*, *HOXB4* and *HOXD4* had MbMt-hypermethylated sites in the coding sequences of the last exon (Additional file 1), which encodes the homeodomain. *HOXC4* was also methylated in this subregion in Mb and Mt, as were a number of other types of cell cultures, so that this subregion was not scored as hypermethylated (data not shown). Four other *HOX* genes also had clusters of hypermethylated sites in the coding sequences of the last exon (Additional file 1).

*HOX* gene myogenic hypermethylation was also found in gene subregions without much sequence similarity. This includes the 3'-untranslated region (3'-UTR) of *HOXB6* and *HOXC5*, exon 1 of *HOXA6*, the 2-kb upstream region of *HOXC12* and an internal exon (exon 3 of four exons) of *HOXA3* (Additional file 1). All of these genes had mostly unmethylated CpGs detected by RRBS in Mb and Mt in their vicinity, so that the detected MbMt hypermethylation was not just due to large, continuous blocks of DNA methylation. *HOXA6* and *HOXC6*, both of which had two exons, illustrate variety in the DNA methylation of paralogs. They exhibited, respectively, hypermethylation (and gene silencing) and little or no methylation in their first exon (and moderate gene expression) in myogenic progenitor cells (Additional file 1).

We found that subregions of H3K4me2 in ESCs often were located at MbMt-hypermethylated sites (Additional files 3, 4, 5 and 6, purple triangles). H3K4me2 marks (transcription promoting) in ESCs often overlap H3K27me3 signals (transcription repressing) and hence are called *bivalent chromatin subregions* that are paused for activity [17]. We hypothesize that the frequent overlap of ESC H3K4me2 with MbMt hypermethylation is due to the resolution of a bivalent chromatin mark to a univalent H3K27me3 mark with the addition of *de novo* DNA methylation early in the differentiation of the skeletal muscle lineage.

#### Unusually high 5hmC levels at a hypermethylated site in *HOXB5* in skeletal muscle

Because Mb and Mt have particularly high levels of the RNA encoding TET1 and TET2, enzymes that generate 5hmC from 5mC residues [11], and RRBS cannot

distinguish between 5hmC and 5mC [42], it was important to determine relative amounts of these modified C residues at representative *HOX* cluster sites. We quantified 5mC and 5hmC at a MbMt-hypermethylated *MspI* site (5'-CCGG-3') in the single introns of *HOXB5* and *HOXD4*, exon 1 of *HOXA5*, exon 2 of *HOXC6*, and 1.7 kb upstream of the TSS of *HOXA7* (Figures 1, 2, 4 and 6) by an enzymatic assay that involves glucosylation of 5hmC by T4 phage  $\beta$ -glucosyltransferase ( $\beta$ -GT; Epimark; New England Biolabs, Ipswich, MA, USA), digestion with *MspI* or *HpaII* and real-time PCR [11]. Using sets of samples independent from those for RRBS, the hypermethylation of these five sites in Mb and of the *HOXD4* and *HOXA7* sites in skeletal muscle was verified by this assay (Table 1). Moreover, we found that all or almost all of the hypermethylation at these sites in Mb was due to 5mC rather than to 5hmC.

Surprisingly, only the skeletal muscle samples at the assayed CpG site in *HOXB5* intron 1 displayed considerable levels of 5hmC (27% or 41% of all C as 5hmC), and, remarkably, these samples exhibited more 5hmC than 5mC (5hmC and no 5mC or mostly 5hmC; Table 1). In a previous study of genomic DNA 5hmC mapping in mouse embryonic stem cells (E14) ([73] and unpublished data), only about 2% of the mapped 5hmC sites were found to contain higher levels of 5hmC compared to 5mC using *MspI* and *HpaII* differential digestion after  $\beta$ -glucosylation, as in this study. At the *HOXB5* site analyzed in the present study, all the detected modified C was 5hmC in the heart samples, one of the two assayed cerebellum samples and the foreskin fibroblast sample. However, the overall levels of modified C in these samples were much lower than in skeletal muscle: only 1% to 6% vs. 41% to 43%, respectively (Table 1). In a study of the *HOXA* gene cluster in NT2 embryonal carcinoma cells before and after retinoic acid-induced differentiation, Bocker *et al.* [45] found that gene activation was accompanied by conversion of much 5mC to 5hmC. Their analysis involved immunoprecipitation using antibodies to 5hmC or 5mC, which does not allow comparisons of relative amounts of 5hmC to 5mC. Our results also indicate that some *HOX* genes can have more genomic 5hmC in differentiation products than in precursor cells, although, in this case, the comparison is adult tissue to progenitor cells. This finding is also consistent with our previous demonstration that skeletal muscle had twice the average genomic 5hmC content of Mb or Mt in an assay of overall levels of genomic 5hmC [11].

#### Conclusions

Our profiling of differential DNA methylation in *HOX* gene clusters suggests that myogenesis-associated hypermethylation plays diverse roles in controlling cell type-specific expression of *HOX* genes and does not simply

**Table 1 Quantification of 5mC and 5hmC at five tested CCGG sites in the four HOX clusters**

Location of tested site	(5mC + 5hmC)/All C (ratio) <sup>a</sup>										
	5hmC/(5mC + 5hmC) (%)										
	Mb1	Mb2	Skel muscle1	Skel muscle2	Heart1	Heart2	Brain1	Brain2	Leuk1	Leuk2	Fib
<i>HOXB5</i> in1	<b>0.93<sup>b</sup></b>	<b>0.80</b>	<b>0.43</b>	<b>0.41</b>	0.06	0.01	0.04	0	0.05	0.00	0.03
	0%	15%	<b>63%<sup>c</sup></b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>		0%	0%	<b>100%</b>
<i>HOXA5</i> ex1	<b>0.99</b>	<b>0.96</b>	0.19	0.04	0.05	0	0.02	0.01	<b>0.97</b>	<b>0.86</b>	0.02
	0%	1%	17%	0	0		<b>100%</b>	<b>62%</b>	0%	0%	0%
<i>HOXA7</i> upstr	<b>0.99</b>	<b>0.65</b>	<b>0.55</b>	<b>0.48</b>	0.01	0	0.02	0.01	0.07	0.02	0.02
	0%	1%	15%	0%	0%		<b>78%</b>	<b>62%</b>	<b>71%</b>	0%	4%
<i>HOXC6</i> ex2	<b>0.98</b>	<b>0.46</b>	0.26	0.01	0.01	0	0.04	0	0.18	0.03	0.02
	0%	2%	31%	0%	0%		<b>100%</b>		0%	0%	1%
<i>HOXD4</i> in1	<b>0.98</b>	<b>0.92</b>	<b>0.78</b>	<b>0.54</b>	0.08	0.03	0.05	0	0.08	0.01	1.00
	0%	0%	10%	0%	<b>48%</b>	0%	<b>100%</b>		0%	0%	3%

<sup>a</sup>5hmC and 5mC were quantified by an enzymatic assay (Epimark; New England Biolabs) coupled with real-time PCR as described in Methods. Two biological replicates each were tested for myoblasts, skeletal muscle biopsy samples, heart and brain autopsy samples, and leukocytes. One sample of foreskin fibroblasts was assayed. The tested positions of CpG sites (in CCGG sequences) were as follows (hg19): *HOXB5* intron 1, chr17:46670028; *HOXA5* exon 1, chr7:27183175; *HOXA7* upstream, chr7: 27197926; *HOXC6* upstream, chr12:54423458; and *HOXD4* intron 1, chr2:177017078. These positions are also indicated in Figures 1, 2, 4 and 6.

<sup>b</sup>Samples with ratios of (5mC + 5hmC)/total C greater than 0.4 are shown in boldface.

<sup>c</sup>Samples with more than 40% of modified C as 5hmC are shown in boldface.

mirror chromatin epigenetics. Specific roles for developmentally associated, differential methylation of *HOX* gene regions would be consistent with the unusually high density of the sense and antisense genes, alternative promoters and alternative transcription termination sites in *HOX* gene clusters [56-58] and the need for tight control of expression of these key developmental regulatory genes. For example, we found extensive DNA hypermethylation in the 3' half of *HOXD* selectively in myogenic cells and skeletal muscle tissue, whereas H3K27me3 was present throughout the *HOXD* gene cluster in many cell types, including Mb and Mt. This finding is consistent with the hypothesis that the skeletal muscle lineage needs especially tight or stable silencing of transcription of the *HOXD* gene cluster conferred by DNA methylation plus polycomb silencing. Moreover, our results indicate that myogenic DNA hypermethylation was often localized to bivalent ESC subregions, which may have been resolved to stably repressed subregions by *de novo* DNA methylation during differentiation. This is similar to a model for DNA hypermethylation of polycomb protein-controlled genes in cancer [74,75]. At the *HOXA* and *HOXC* gene clusters, the pattern of tissue-specific epigenetic marks suggests another function of myogenic DNA hypermethylation. In these gene clusters, subregions rich in myogenic DNA hypermethylation appear to be part of boundaries around a central multigenic region consisting of mostly enhancer- or promoter-type histone modifications. This DNA hypermethylation might help prevent the spreading of activating histone modifications from the central region of *HOXA* and *HOXC* gene clusters to their periphery.

Our study also suggests that myogenic hypermethylation of DNA might partly downregulate in cis the level of transcription of some *HOX* antisense ncRNA genes that positively control expression of overlapping protein-encoding *HOX* genes, such as *HOXB-AS3* and *HOXB5*. Myogenic hypermethylation from intragenic or intergenic locations could exert its effects on enhancers by decreasing transcription of the enhancer itself as well as by repression of canonical promoters of protein-encoding or lncRNA genes [76,77]. Moreover, our results are consistent with the hypotheses that hypermethylation within gene bodies affects which RNA isoforms are generated by modulating differential splicing and the use of alternate promoters [62,76,78,79].

Yet other relationships between differential methylation and transcription were indicated by the association of upregulation of *HOTAIR* in the *HOXC* cluster and hypermethylation of *HOTAIR*'s immediate downstream sequences in myogenic progenitor cells and foreskin fibroblasts. In addition, the one example of myogenic hypomethylation in the *HOX* gene clusters was in the single intron of *HOXA10* in a region with the chromatin features of a myogenesis-associated enhancer. This tissue-specific DNA hypomethylation might activate or help maintain the activity of a tissue-specific enhancer, consistent with the positive association of DNA hypomethylation and inducible enhancers [80]. The dynamic nature of developmentally linked changes in DNA was evidenced by our finding that, at five tested representative CpG sites displaying myogenic hypermethylation, the levels of 5hmC were low or negligible in the skeletal muscle lineage, with the prominent exception of an intronic region in *HOXB5*

in skeletal muscle tissue but not in muscle progenitor cells. In summary, our study of myogenesis-associated differences in DNA methylation indicates the importance of considering a wide variety of possible roles for differential DNA methylation when studying disease-linked epigenetic changes.

## Methods

### Samples

All the Mb cell strains used for methylation analysis were propagated from muscle biopsy samples that were previously described [11]. Mt samples were obtained from these myoblast cell strains by serum limitation for 5 days [11]. By immunostaining [81], we demonstrated that all batches of myoblasts contained more than 90% desmin-positive cells and that myotube preparations had more than 75% of their nuclei in multinucleated, desmin-positive and myosin heavy chain-positive cells. Four of the nine Mb and Mt samples were Mb-Mt pairs from two normal controls, and five were from two facioscapulohumeral muscular dystrophy patients or an inclusion body myositis patient; however, all Mb and Mt samples predominantly shared the same myogenesis-associated epigenetic marks [11]. The other cell cultures for DNA methylation profiling and assessment of differential methylation in myogenic vs. nonmyogenic cell cultures and the tissue samples and the two primary (not passaged) cell cultures (hepatocytes and pancreatic islets) used for DNA methylation profiling to identify skeletal muscle-associated differential methylation were previously described normal samples [11]. All cell cultures were untransformed, except for the LCLs, which had been transformed *in vitro* by Epstein-Barr virus. Three control Mb or Mt samples were used for the combined data shown in DNase-seq profiles (Additional files 3, 4, 5 and 6). Two of these Mb samples and two of the Mt samples were from the same batches of cells used for RRBS. Although different sources of Mb and Mt were used for ChIP-seq, peaks of myogenesis-associated H3K4me3 or H3K4me2 usually overlapped DNase-seq peaks for these cell types (Additional files 3, 4, 5 and 6), as expected.

### DNA methylation profiling and statistical analyses

For the methylation analysis, high-molecular-weight DNA was extracted, digested with *MspI* and used for bisulfite-based RRBS, including next-generation sequencing on an Illumina platform (Illumina Inc, San Diego, CA, USA) as previously described using the same samples as we used for our last study of myogenic differential methylation [11]. DNA methylation data for cell cultures and tissues are from the ENCODE project and available from the UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=292099017&c=chr1&g=wgEncodeHaibMethylRrbs>). BED files containing DNA methylation data for 5 Mb cell strains, 5 Mt samples

(including one technical duplicate) and 47 nonmyogenic samples representing 23 unique cell lines from a variety of tissue types were aggregated into a single matrix. Rows were included for each site detected in any sample and used for assessing statistically significant differential methylation in the set of Mb plus Mt vs. non-muscle-cell cultures or skeletal muscle vs. non-muscle-cell cultures as previously described [11]. To increase the specificity of our analyses, we restricted our attention to those sites for which a change in methylation percentage of at least 50% was observed at a significance level of 0.01 or below. The closestBed program, (<http://bedtools.readthedocs.org/en/latest/>), a member of the bedtools suite [82], was used to map each DM site to the nearest gene using both protein-coding (NM\*) and noncoding (NR\*) genes and one isoform per gene as described previously [11].

### Chromatin epigenetic and transcription profiling

Data sets and sample information for histone modifications and CTCF binding, non-strand-specific RNA-seq and strand-specific RNA-seq were obtained from the ENCODE project (<http://genomes.ucsc.edu/>) via the laboratories of Bradley Bernstein (Broad Institute), Barbara Wold (California Institute of Technology) and Tom Gingeras (Cold Spring Harbor Laboratory), respectively. RNA-seq data were available for Mb, but not for Mt. RNA isoform analysis and quantification were done with the CuffDiff tool [83] using the above-mentioned ENCODE non-strand-specific RNA-seq database. For cell cultures that were represented in both RNA-seq profiles (Mb, HUVEC, GM12878, NHEK, NHLF and H1 ESC), the relative expression of different cell types was similar. Mb and Mt samples for ENCODE histone modification and CTCF profiling and Mb for RNA-seq were commercially obtained, and no immunostaining was described for them. For DNaseI hypersensitivity profiling, intact nuclei were treated with DNaseI and the DNaseI-hypersensitive fraction was analyzed by next-generation sequencing as previously described ([11,84]; ENCODE (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=292099619&c=chr20&g=wgEncodeOpenChromDnase>)). In addition, we mined data from our previous expression profiling of Mb and Mt vs. 19 types of non-muscle-cell cultures on microarrays ([81]; GeneChip Exon 1.0 ST Array; Affymetrix, Santa Clara, CA, USA). There was overlap of three of the five control Mb or Mt samples for RRBS with the samples used for expression microarray profiling.

### Quantification of 5hmC and 5mC by enzymatic assay

For analysis of the levels of 5hmC and 5mC at a given *MspI* site, we used an assay involving  $\beta$ -GT (Epimark; New England Biolabs). After incubation or a mock enzyme incubation, aliquots were digested with *MspI*, which can cleave CCGG sites whether or not they are

methylated or hydroxymethylated at the internal C residue, but not if they contain glucosylated 5hmC. In parallel, digestions were done with *HpaII*, which can cleave CCGG sites only if they are unmodified at the internal C, and other aliquots were used as uncut controls according to the manufacturer's instructions. Next, real-time PCR was performed and methylation status was calculated by subtraction of Ct values. The respective forward and reverse primers for PCR were as follows (5' to 3'): *HOXC6*, ATCTTTAGGGGTCGGCTACG and CGCGTTAGGTAGCGATTGA; *HOXB5*, AGATGCCACATTCAAGCTC and CAAGGGTGAGGCACTAGGAG; *HOXA7* upstream, GGTGTGGAGTGAGGACAAC and CGATGCGACTGGGATTATTT; *HOXA5*, TTGCTCGCTCACGGA ACTAT and TATAGACGCACAAACGACCG; and *HOXD4*, GGGATTTCAAAATGCTTGA and ACCTCCTCAAACACACCCAC.

## Additional files

**Additional file 1: Table S1.** Correlations between myogenic differential methylation and transcriptional up- or downregulation in myogenic vs. nonmyogenic cells.

**Additional file 2: Figure S1.** Myogenesis-associated DNA hypermethylation in the silenced *HOXD4* region.

**Additional file 3: Figure S2.** Myogenic DNA hypermethylation and chromatin epigenetic marks in the *HOXD1*-to-*MIR10B* subregion of the *HOXD* gene cluster.

**Additional file 4: Figure S3.** Myogenic DNA hypermethylation and chromatin epigenetic marks in the *HOXA-AS5*-to-*HOXC11* subregion.

**Additional file 5: Figure S4.** Myogenic DNA hyper- and hypo-methylation and chromatin epigenetic marks in the *HOXA-AS3*-to-*HOXA11* subregion.

**Additional file 6: Figure S5.** Myogenic DNA hypermethylation and chromatin epigenetic marks in the *HOXB4*-to-*HOXB9* subregion.

## Abbreviations

3'-UTR: 3'-untranslated region; 5hmC: 5-hydroxymethylcytosine; 5mC: 5-methylcytosine; ChIP-seq: Chromatin immunoprecipitation/next-generation DNA sequencing; DM: Differentially methylated; ESC: H1 embryonic stem cell; Fib2: Foreskin fibroblasts; GM12878: Lymphoblastoid cell line produced from blood by Epstein-Barr virus transformation;  $\beta$ -GT:  $\beta$ -glucosyltransferase; H3K27Ac: Histone H3 lysine 27 acetylation; H3K27me3: histone H3 lysine 27 trimethylation; H3K4me1: Histone H3 lysine 4 monomethylation; H3K4me3: Histone H3 lysine 4 trimethylation; HMEC: Human mammary epithelial cell; HUVEC: Human umbilical cord endothelial cells; LCL: Lymphoblastoid cell line; lncRNA: long noncoding RNA; Mb: myoblast; MbMt: Set of myoblasts plus myotubes; Mt: Myotube; ncRNA: Noncoding RNA; NHEK: Normal human epidermal keratinocyte; NHLF: Normal human lung fibroblast; PcG: Polycomb group; P/E-like domain: Promoter- and enhancer-type chromatin in a multigenic region; RNA Pol II: RNA polymerase II; RRBS: Reduced representation bisulfite sequencing; SA epith: Small airway epithelial cell; TSS: transcription start site.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

KT prepared and immunocytochemically evaluated the Mb and Mt samples and helped with the Cufflinks analysis. CB and ML did the statistical analyses. JT, ZS and SP were responsible for the 5hmC and 5mC quantification. SC and CR helped with the bioinformatics analyses. LS and GEC did the DNase hypersensitivity profiling. LS did the liftover of C2C12 MyoD coordinates to

the human genome. ME did bioinformatics analyses and wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgments

For making their data (<http://genome.ucsc.edu/>) available prepublication, we thank the ENCODE Chromatin Group at the Broad Institute and Massachusetts Institute of Technology (Bradley Bernstein's and Manolis Kellis's groups) and the ENCODE Transcription Group at the California Institute of Technology (Barbara Wold's group) and the Cold Spring Harbor Laboratory (Tom Gingeras's group). We are grateful to the ENCODE DNA Methylation Group at the HudsonAlpha Institute for Biotechnology (Rick Myers's Group) for the initial collaboration on RRBS [11] and to Melody Badoo for help with the Cufflinks program update. This research was supported in part by a grant from the National Institutes of Health to ME (NS04885).

## Author details

<sup>1</sup>Hayward Human Genetics Program and Tulane Cancer Center, Tulane Health Sciences Center, New Orleans LA, USA. <sup>2</sup>Tulane Cancer Center and Department of Mathematics, Tulane Health Sciences Center and Tulane University, New Orleans, LA, USA. <sup>3</sup>New England Biolabs, Ipswich MA, USA. <sup>4</sup>Institute for Genome Sciences & Policy, Duke University, Durham, NC, USA. <sup>5</sup>Center for Bioinformatics and Genomics, Tulane Health Sciences Center, New Orleans, LA, USA.

Received: 5 April 2013 Accepted: 21 June 2013

Published: 2 August 2013

## References

- Mallo M, Wellik DM, Deschamps J: *Hox* genes and regional patterning of the vertebrate body plan. *Dev Biol* 2010, **344**:7–15.
- Foronda D, de Navas LF, Garaulet DL, Sánchez-Herrero E: Function and specificity of *Hox* genes. *Int J Dev Biol* 2009, **53**:1404–1419.
- Barber BA, Rastegar M: Epigenetic control of *Hox* genes during neurogenesis, development, and disease. *Ann Anat* 2010, **192**:261–274.
- Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D: The dynamic architecture of *Hox* gene clusters. *Science* 2011, **334**:222–225.
- Ladd-Acosta C, Pevsner J, Sabuncuyan S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP: DNA methylation signatures within the human brain. *Am J Hum Genet* 2007, **81**:1304–1315.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES: Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 2008, **454**:766–770.
- De Bustos C, Ramos E, Young JM, Tran RK, Menzel U, Langford CF, Eichler EE, Hsu L, Henikoff S, Dumanski JP, Trask BJ: Tissue-specific variation in DNA methylation levels along human chromosome 1. *Epigenetics Chromatin* 2009, **2**:7.
- Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP: Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 2010, **467**:338–342.
- Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, Absher DM, Wold BJ, Myers RM: Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* 2013, **23**:555–567.
- Kang JS, Krauss RS: Muscle stem cells in developmental and regenerative myogenesis. *Curr Opin Clin Nutr Metab Care* 2010, **13**:243–248.
- Tsumagari K, Baribault C, Terragni J, Varley KE, Gertz J, Pradhan S, Baddoo M, Crain CM, Song L, Crawford GE, Myers RM, Lacey M, Ehrlich M: Early de novo DNA methylation and prolonged demethylation in the muscle lineage. *Epigenetics* 2013, **8**:317–332.
- Maconochie M, Nonchev S, Morrison A, Krumlauf R: Paralogous *Hox* genes: function and regulation. *Annu Rev Genet* 1996, **30**:529–556.
- Akbas GE, Taylor HS: *HOXC* and *HOXD* gene expression in human endometrium: lack of redundancy with *HOXA* paralogs. *Biol Reprod* 2004, **70**:39–45.
- Zakany J, Duboule D: The role of *Hox* genes during vertebrate limb development. *Curr Opin Genet Dev* 2007, **17**:359–366.
- Houghton L, Rosenthal N: Regulation of a muscle-specific transgene by persistent expression of *Hox* genes in postnatal murine limb muscle. *Dev Dyn* 1999, **216**:385–397.

16. Yamamoto M, Kuroiwa A: *Hoxa-11 and Hoxa-13 are involved in repression of MyoD during limb muscle development.* *Dev Growth Differ* 2003, **45**:485–498.
17. Kashyap V, Gudas LJ, Brenet F, Funk P, Viale A, Scandura JM: **Epigenomic reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid.** *J Biol Chem* 2011, **286**:3250–3260.
18. Novotny E, Compton S, Liu PP, Collins FS, Chandrasekharappa SC: **In vitro hematopoietic differentiation of mouse embryonic stem cells requires the tumor suppressor menin and is mediated by Hoxa9.** *Mech Dev* 2009, **126**:517–522.
19. Golpon HA, Geraci MW, Moore MD, Miller HL, Miller GJ, Tuder RM, Voelkel NF: **HOX genes in human lung: altered expression in primary pulmonary hypertension and emphysema.** *Am J Pathol* 2001, **158**:955–966.
20. Kelly M, Daftary G, Taylor HS: **An autoregulatory element maintains HOXA10 expression in endometrial epithelial cells.** *Am J Obstet Gynecol* 2006, **194**:1100–1109.
21. Rancourt DE, Tsuzuki T, Capocchi MR: **Genetic interaction between hoxb-5 and hoxb-6 is revealed by nonallelic noncomplementation.** *Genes Dev* 1995, **9**:108–122.
22. Björnsson JM, Larsson N, Brun AC, Magnusson M, Andersson E, Lundström P, Larsson J, Repetowska E, Ehinger M, Humphries RK, Karlsson S: **Reduced proliferative capacity of hematopoietic stem cells deficient in Hoxb3 and Hoxb4.** *Mol Cell Biol* 2003, **23**:3872–3883.
23. Biressi S, Tagliafico E, Lamorte G, Monteverde S, Tenedini E, Roncaglia E, Ferrari S, Ferrari S, Cusella-De Angelis MG, Tajbakhsh S, Cossu G: **Intrinsic phenotypic diversity of embryonic and fetal myoblasts is revealed by genome-wide gene expression analysis on purified cells.** *Dev Biol* 2007, **304**:633–651.
24. Lacombe J, Hanley O, Jung H, Philippidou P, Surmeli G, Grinstein J, Dasen JS: **Genetic and functional modularity of Hox activities in the specification of limb-innervating motor neurons.** *PLoS Genet* 2013, **9**:e1003184.
25. Lawrence HJ, Stage KM, Mathews CH, Detmer K, Scibienski R, MacKenzie M, Migliaccio E, Boncinelli E, Largman C: **Expression of HOX C homeobox genes in lymphoid cells.** *Cell Growth Differ* 1993, **4**:665–669.
26. Garcia-Gasca A, Spyropoulos DD: **Differential mammary morphogenesis along the anteroposterior axis in Hoxc6 gene targeted mice.** *Dev Dyn* 2000, **219**:261–276.
27. Rieger E, Bijl JJ, van Oostveen JW, Soyer HP, Oudejans CB, Jiwa NM, Walboomers JM, Meijer CJ: **Expression of the homeobox gene HOXC4 in keratinocytes of normal skin and epithelial skin tumors is correlated with differentiation.** *J Invest Dermatol* 1994, **103**:341–346.
28. Delpretti S, Zakany J, Duboule D: **A function for all posterior Hoxd genes during digit development?** *Dev Dyn* 2012, **241**:792–802.
29. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES: **Genomic maps and comparative analysis of histone modifications in human and mouse.** *Cell* 2005, **120**:169–181.
30. Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, Ge K, Krumlauf R, Shilatifard A: **Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II.** *Mol Cell Biol* 2009, **29**:6074–6085.
31. Zhang Y, Liu Z, Medrzycki M, Cao K, Fan Y: **Reduction of Hox gene expression by histone H1 depletion.** *PLoS One* 2012, **7**:e38829.
32. Kim K, Lee HC, Park JL, Kim M, Kim SY, Noh SM, Song KS, Kim JC, Kim YS: **Epigenetic regulation of microRNA-10b and targeting of oncogenic MAPRE1 in gastric cancer.** *Epigenetics* 2011, **6**:740–751.
33. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY: **A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.** *Nature* 2011, **472**:120–124.
34. Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY: **Anatomic demarcation by positional variation in fibroblast gene expression programs.** *PLoS Genet* 2006, **2**:e119.
35. Tschopp P, Duboule D: **A genetic approach to the transcriptional regulation of Hox gene clusters.** *Annu Rev Genet* 2011, **45**:145–166.
36. Chambeyron S, Da Silva NR, Lawson KA, Bickmore WA: **Nuclear re-organisation of the Hoxb complex during mouse embryonic development.** *Development* 2005, **132**:2215–2223.
37. Avraham A, Sandbank J, Yarom N, Shalom A, Karni T, Pappo I, Sella A, Fich A, Walfisch S, Gheber L, Evron E: **A similar cell-specific pattern of HOXA methylation in normal and in cancer tissues.** *Epigenetics* 2010, **5**:41–46.
38. Hershko AY, Kafri T, Fainsod A, Razin A: **Methylation of HoxA5 and HoxB5 and its relevance to expression during mouse development.** *Gene* 2003, **302**:65–72.
39. Yamagishi T, Ozawa M, Ohtsuka C, Ohyama-Goto R, Kondo T: **Evx2-Hoxd13 intergenic region restricts enhancer association to Hoxd13 promoter.** *PLoS One* 2007, **2**:e175.
40. Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A: **A novel CpG island set identifies tissue-specific methylation at developmental gene loci.** *PLoS Biol* 2008, **6**:e22.
41. Laurent L, Wong E, Li G, Huynh T, Tsigirgos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL: **Dynamic changes in the human methylome during differentiation.** *Genome Res* 2010, **20**:320–331.
42. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE: **5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells.** *Genome Biol* 2011, **12**:R54.
43. Guo JU, Su Y, Zhong C, Ming GL, Song H: **Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain.** *Cell* 2011, **145**:423–434.
44. Globisch D, Munzel M, Müller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T: **Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates.** *PLoS One* 2011, **5**:e15367.
45. Bocker MT, Tuorto F, Raddatz G, Musch T, Yang FC, Xu M, Lyko F, Breiling A: **Hydroxylation of 5-methylcytosine by TET2 maintains the active state of the mammalian HOXA cluster.** *Nat Commun* 2012, **3**:818.
46. Szulwach KE, Li X, Li Y, Song CX, Han JW, Kim S, Namburi S, Hermetz K, Kim JJ, Rudd MK, Yoon YS, Ren B, He C, Jin P: **Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells.** *PLoS Genet* 2011, **7**:e1002154.
47. Moroni MC, Vigano MA, Mavilio F: **Regulation of the human HOXD4 gene by retinoids.** *Mech Dev* 1993, **44**:139–154.
48. Vrba L, Garbe JC, Stampfer MR, Futscher BW: **Epigenetic regulation of normal human mammary cell type-specific miRNAs.** *Genome Res* 2011, **21**:2026–2037.
49. ENCODE Project Consortium, Myers RM, Stamatoyannopoulos J, Snyder M, Dunham I, Hardison RC, Bernstein BE, Gingeras TR, Kent WJ, Birney E, Wold B, Crawford GE, Bernstein BE, Epstein CB, Shores N, Ernst J, Mikkelsen TS, Kheradpour P, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Lin MF, Kellis M, Gingeras TR, Davis CA, et al: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**:e1001046.
50. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: **Improving RNA-Seq expression estimates by correcting for fragment bias.** *Genome Biol* 2011, **12**:R22.
51. Soshnikova N, Duboule D: **Epigenetic regulation of vertebrate Hox genes: a dynamic equilibrium.** *Epigenetics* 2009, **4**:537–540.
52. Hagarman JA, Motley MP, Kristjansdottir K, Soloway PD: **Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells.** *PLoS One* 2012, **8**:e53880.
53. Reddington JP, Perricone SM, Nestor CE, Reichmann J, Youngson NA, Suzuki M, Reinhardt D, Dunican DS, Prendergast JG, Mjoseng H, Ramsahoye BH, Whitelaw E, Grealley JM, Adams IR, Bickmore WA, Meehan RR: **Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes.** *Genome Biol* 2013, **14**:R25.
54. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473**:43–49.
55. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau JC, Ferrier P, Spicuglia S: **H3K4 tri-methylation provides an epigenetic signature of active enhancers.** *EMBO J* 2011, **30**:4198–210.
56. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL: **Targeted RNA sequencing reveals the deep complexity of the human transcriptome.** *Nat Biotechnol* 2012, **30**:99–104.
57. Coulombe Y, Lemieux M, Moreau J, Aubin J, Joksimovic M, Berube-Simard FA, Tabaries S, Boucherat O, Guillou F, Larochelle C, Tuggle CK, Jeannotte L: **Multiple promoters and alternative splicing: Hoxa5 transcriptional complexity in the mouse embryo.** *PLoS One* 2010, **5**:e10600.
58. Sessa L, Breiling A, Lavorgna G, Silvestri L, Casari G, Orlando V: **Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster.** *RNA* 2007, **13**:223–239.



59. Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, Gentleman RC, Tapscott SJ: **Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming.** *Dev Cell* 2010, **18**:662–674.
60. Cheng X, Blumenthal RM: **Coordinated chromatin control: structural and functional linkage of DNA and histone methylation.** *Biochemistry* 2010, **49**:2999–3008.
61. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137**:1194–1211.
62. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.** *Nature* 2011, **479**:74–79.
63. Tao Y, Xi S, Briones V, Muegge K: **Lsh mediated RNA polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation.** *PLoS One* 2011, **5**:e19163.
64. Nguyen CT, Gonzales FA, Jones PA: **Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation.** *Nucleic Acids Res* 2001, **29**:4598–4606.
65. Xi S, Zhu H, Xu H, Schmidtman A, Geiman TM, Muegge K: **Lsh controls Hox gene silencing during development.** *Proc Natl Acad Sci USA* 2007, **104**:14366–14371.
66. Lu L, Zhu G, Zhang C, Deng Q, Katsaros D, Mayne ST, Risch HA, Mu L, Canuto EM, Gregori G, Benedetto C, Yu H: **Association of large noncoding RNA HOTAIR expression and its downstream intergenic CpG island methylation with survival in breast cancer.** *Breast Cancer Res Treat* 2012, **136**:875–883.
67. Liu H, Han H, Li J, Wong L: **DNAFMiner: a web-based software toolbox to recognize two types of functional sites in DNA sequences.** *Bioinformatics* 2005, **21**:671–673.
68. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.** *Cell* 2007, **129**:1311–1323.
69. Li D, Feng J, Wu T, Wang Y, Sun Y, Ren J, Liu M: **Long intergenic noncoding RNA HOTAIR is overexpressed and regulates PTEN methylation in laryngeal squamous cell carcinoma.** *Am J Pathol* 2013, **182**:64–70.
70. Sasaki YT, Sano M, Kin T, Asai K, Hirose T: **Coordinated expression of ncRNAs and HOX mRNAs in the human HOXA locus.** *Biochem Biophys Res Commun* 2007, **357**:724–730.
71. Trivedi CM, Patel RC, Patel CV: **Homeobox gene HOXA9 inhibits nuclear factor- $\kappa$ B dependent activation of endothelium.** *Atherosclerosis* 2007, **195**:e50–e60.
72. Porter JD, Israel S, Gong B, Merriam AP, Feuerman J, Khanna S, Kaminski HJ: **Distinctive morphological and gene/protein expression signatures during myogenesis in novel cell lines from extraocular and hindlimb muscle.** *Physiol Genomics* 2005, **24**:264–275.
73. Sun Z, Terragni J, Borgaro JG, Liu Y, Yu L, Guan S, Wang H, Sun D, Cheng X, Zhu Z, Pradhan S, Zheng Y: **High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells.** *Cell Rep* 2013, **3**:567–576.
74. Baylin SB: **Stem cells, cancer, and epigenetics.** In *The Stem Cell Research Community, Stem Book*. Edited by Jaenisch R, Laird P. 2009. doi:10.3824/stembook.1.50.1. Available at: [http://www.stembook.org/sites/default/files/pubnode/59012d58e397f06325596210a30ad7790cd314c0/Baylin\\_Final\\_proofs/Stem\\_cells\\_cancer\\_and\\_epigenetics/Stem\\_cells\\_cancer\\_and\\_epigenetics.pdf](http://www.stembook.org/sites/default/files/pubnode/59012d58e397f06325596210a30ad7790cd314c0/Baylin_Final_proofs/Stem_cells_cancer_and_epigenetics/Stem_cells_cancer_and_epigenetics.pdf).
75. Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbrugger T, Wang Q, Aryee MJ, Joyce P, Ahuja N, Weisenberger D, Collisson E, Zhu J, Yegnasubramanian S, Matsui W, Baylin SB: **A DNA hypermethylation module for the stem/progenitor cell signature of cancer.** *Genome Res* 2012, **22**:837–849.
76. Maunakea AK, Nagarajan RP, Bilienky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Schorks K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF: **Conserved role of intragenic DNA methylation in regulating alternative promoters.** *Nature* 2010, **466**:253–257.
77. Kolovos P, Knoch TA, Grosveld FG, Cook PR, Papantonis A: **Enhancers and silencers: an integrated and simple model for their function.** *Epigenetics Chromatin* 2012, **5**:1.
78. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP: **A human B cell methylome at 100-base pair resolution.** *Proc Natl Acad Sci USA* 2009, **106**:671–678.
79. Sati S, Tanwar VS, Kumar KA, Patowary A, Jain V, Ghosh S, Ahmad S, Singh M, Reddy SU, Chandak GR, Raghunath M, Sivasubbu S, Chakraborty K, Scaria V, Sengupta S: **High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region.** *PLoS One* 2012, **7**:e31621.
80. Taube JH, Allton K, Duncan SA, Shen L, Barton MC: **Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells.** *J Biol Chem* 2010, **285**:16135–16144.
81. Tsumagari K, Chang SC, Lacey M, Baribault C, Chittur SV, Sowden J, Tawil R, Crawford GE, Ehrlich M: **Gene expression during normal and FSHD myogenesis.** *BMC Med Genomics* 2011, **4**:67.
82. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26**:841–842.
83. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7**:562–578.
84. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, Sheffield NC, Gräf S, Huss M, Keefe D, Liu Z, London D, McDaniel RM, Shibata Y, Showers KA, Simon JM, Vales T, Wang T, Winter D, Zhang Z, Clarke ND, Birney E, Iyer VR, Crawford GE, Lieb JD, Furey TS: **Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity.** *Genome Res* 2011, **21**:1757–1767.

doi:10.1186/1756-8935-6-25

**Cite this article as:** Tsumagari et al.: DNA methylation and differentiation: HOX genes in muscle cells. *Epigenetics & Chromatin* 2013 **6**:25.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

