Characterization of the global and locus-specific regulation of gene expression during early myogenic differentiation

Katherine Dixon

This thesis is submitted to the Faculty of Graduate and Postdoctoral Studies
as a partial fulfillment of the Master of Science program in
Cellular and Molecular Medicine

Department of Cellular and Molecular Medicine
Faculty of Medicine
University of Ottawa

Abstract

During cellular differentiation, gene expression is globally regulated through changes in the epigenome. How a single genome can give rise to a diversity of cell and tissue types remains a complex area of investigation, and here we sought to explore the molecular regulation of gene expression during the differentiation of skeletal muscle cells from committed myogenic progenitors. Using a systematic and integrated analysis of global transcriptional and epigenetic data, we characterized the regulation of gene expression in differentiating myoblasts and found that muscle-specific gene expression is regulated through differential activation of tissue-specific regulatory DNA elements by the myogenic transcription factor MyoD. In addition, the genome-wide localization of MyoD, and the mechanisms underlying its function in transcriptional regulation, varies between myogenic progenitors and differentiating myoblasts. Our study explores the recruitment and function of MyoD at regulatory elements of target genes and additionally describes a novel role for ligand-inducible signaling in the regulation of MyoD function and ultimately in myogenic differentiation.

Table of Contents

List of Figures

List of Abbreviations

AF-1	Activation function 1
AF-2	Activation function 2
Angptl4	Angiopoietin-like 4
APL	Acute promyelocytic leukemia
ATRA	All-trans retinoic acid
bHLH	Basic helix-loop-helix
BMP	Bone morphogenic protein
CBP	CREB-binding protein
ChIP	Chromatin immunoprecipitation
CREB	cAMP response element binding protein
CTCF	CCCTC-binding factor
CTCL	Cutaneous T cell lymphoma
DBD	DNA binding domain
DR	Direct repeat
ER	Estrogen receptor
ES	Embryonic stem
FGF	Fibroblast growth factor
FPKM	Fragments per kilobase per million
GERP	Genomic evolutionary rate profiling
HAT	Histone acetyltransferase
HCNCE	Highly conserved non-coding elements
HDAC	Histone deacetylase

HP1	Heterochromatin protein 1
HRE	Hormone response element
Id	Inhibitor of DNA binding
LBD	Ligand binding domain
LXR	Liver X receptor
Mef2	Myocyte enhancer factor 2
MRF	Myogenic regulatory factor
NGS	Next-generation sequencing
PCAF	p300/CBP-associated factor
PML	Promyelocytic leukemia
PPAR	Peroxisome proliferator-activated receptor
PRC2	Polycomb-repressive complex 2
RA	Retinoic acid
RAR	Retinoic acid receptor
RNA PolII	RNA polymerase II
RXR	Retinoid X receptor
TR	Thyroid receptor
TSS	Transcription start site

Acknowledgements

I would foremost like to thank my research supervisor, Dr. Qiao Li, for the opportunity she gave me to work with her on this project and for her continuous support and guidance throughout my program. Dr. Li's vision for the project, incredible scientific knowledge and passion for research helped me to develop my own excitement about my current area of research and interest in continuing academic research in the future. I would also like to acknowledge the funding agencies that made this research possible, the Natural Science and Engineering Research Council of Canada and the Canadian Institutes of Health Research.

I would especially like to thank the members of my thesis advisory committee, Dr. Ilya Ioshikhes and Dr. Barbara Vanderhyden, for their expertise and guidance throughout my program.

I am so grateful to have met many wonderful people from the Li lab who helped make my time in the lab enjoyable, including Munerah Hamed, Hamood Alsudais, Yuka Sai, Saadia Khilji, Anabel Bergeron, Howard Wong and Mark Xu. I also had the pleasure of working with members of the Wiper-Bergeron lab, who were endlessly optimistic and supportive about my research.

Finally, I would like to thank a few of the most important people in my life for their support during my program. My appreciation and love for my wonderful family, and Simon Cheesman for his amazing and endless positivity. My parents in particular have been the most supportive people throughout my life and especially towards my interest in science and scientific research. The opportunity to pursue education both in and outside of Ottawa would not have been possible without them, and they individually helped me through many times when I felt overwhelmed or discouraged. They have helped shape the person I am, and I am grateful for so many amazing things I have in my life because of them.

Introduction

In multicellular organisms, the ability to differentially regulate gene expression is fundamental to the development of different cell and tissue types. Global analyses of transcription and its regulation across mammalian cell lines has shown that despite vast differences in gene expression, these cells exhibit similar epigenetic features that distinguish genes being actively transcribed from genes that are not expressed. Functional DNA sequences in the non-coding portion of the genome have also been identified as playing fundamental roles in transcriptional regulation, and their differential usage across cell types accounts for specific transcriptional profiles. In the current study, we explore the molecular regulation of muscle-specific gene expression in differentiating myogenic progenitors, or myoblasts. The application of ligand-inducible signaling in the regulation of transcription may allow a better understanding of the mechanisms underlying cell type-specific gene expression, and how transcription can be modulated by therapeutics targeting genes that are deregulated in diseased muscle tissue.

1. Molecular regulation of tissue-specific gene expression

At the level of transcription, gene expression is regulated by a complex network of cell type-specific transcription factors and non-coding regulatory DNA sequences. The epigenome, which constitutes a venue for transcriptional regulation outside of the underlying DNA sequence, shows remarkable tissue-specificity and plays an important role in differential gene expression. Chemical modifications made to the higher order DNA structure, known as chromatin, alter the usage of regulatory DNA sequences used to guide transcription and ultimately allow the development of tissues with physiologically and functionally distinct characteristics. Which non-coding DNA sequences are involved in the regulation of transcription in specific cell types and how they regulate transcription through dynamic changes to the epigenome is a growing area of research that continues to be explored.

1.1. DNA organization and chromatin assembly

In eukaryotic cells, DNA is packaged through interactions with structural proteins called histones, which help to maintain the highly condensed and ordered structure of chromatin. The smallest structural unit in chromatin is the nucleosome, which consists of an octameric histone core wrapped in 146 bp of DNA. The protein core consists of four histone subtypes, named H2A, H2B, H3 and H4, each of which interacts with at least one other histone protein to form a dimer (in the case of H2A and H2B) or tetramer (in the case of H3 and H4) (Eickbush & Moudrianakis, 1978). An additional histone subtype, H1, serves a role in chromatin organization outside of the nucleosome and is found associated with regions of DNA acting as a link between nucleosomes (Shen & Gorovsky, 1996). The histone proteins are highly conserved, reflecting their fundamental role in the organization of eukaryotic DNA. Each histone has a globular structure containing a histone fold domain, which allows dimerization between histones within the core nucleosome particle, and a less structured N-terminal tail, which extends outside of the nucleosome core (Arents & Moudrianakis, 1995). The nucleosome structure is maintained by electrostatic interactions between the negatively charged DNA backbone and positively charged amino acid residues in the histone tail, and through a variety of post-translational modifications, the affinity of the tail for DNA can be altered in a dynamic fashion that affects that accessibility of functional DNA sequences to transcription factor binding and transcriptional activation.

Transcriptional regulation is guided in part by non-coding regulatory DNA sequences known as *cis*-regulatory elements. These elements, which include promoters, enhancers and insulators, actively regulate transcription at target genes when bound by transcription factors. Promoters span a small region directly upstream of the transcription start site (TSS) and play a direct role in recruiting members of the transcription initiation complex and RNA polymerase (RNA PolII) to the TSS. Active gene promoters are therefore often marked by chromatin modifications that increase

DNA accessibility, allowing efficient recruitment of RNA PolII. Different from promoters, enhancers may be located from a few to hundreds of kilobases away from their target TSS and regulate transcription by recruiting coregulator protein complexes that mediate chromatin state changes at target gene promoters. As enhancers lie far from their target TSS, the association between enhancer-promoter pairs may be guided by elements called insulators, which are bound by the CCCTC-binding factor (CTCF) to mediate chromatin looping (Phillips & Corces, 2009). Together, these regulatory DNA sequences form a complex network of genomic elements that play a part in regulating cell type-specific gene expression.

The function of promoters and enhancers in transcriptional regulation is dependent on their accessibility to transcription factor binding, which is mediated by covalent chemical modifications made to histone tails. Histone modifications such as acetylation, methylation, phosphorylation, ubiquitination and sumoylation, as well as the many enzymes that mediate their addition or removal from specific amino acid residues, have been identified (Kouzarides, 2007). The enzymes that mediate covalent modifications made to histone tails are therefore known as chromatin-modifying enzymes, and these ultimately function as coactivators and corepressors of transcription (Naar, Lemon & Tjian, 2001). However, chromatin-modifying enzymes often do not contain DNA binding domains and are recruited to the genome indirectly through interactions with DNA-bound transcription factors. A related group of enzymes known as chromatin-remodeling enzymes are similarly brought to the genome by transcription factors, and together these enzymes mediate local and global regulation of chromatin structure, DNA accessibility and transcriptional activity.

1.2. Chromatin modifications in transcriptional regulation

The idea that combinations of certain histone modifications could reflect specific chromatin-related activities was originally referred to as the histone code hypothesis (Jenuwein & Allis, 2001). It is now widely understood that the presence of specific modifications, or a combination of these, often correlates with specific functions in transcriptional regulation. Further, which modifications are present at a given locus, which transcription factors and coregulators are mediating the epigenetic changes and the location of the regulatory DNA element in respect to the target gene's transcription start site may all affect the transcriptional outcome. Notably, the epigenetic features associated with enhancers and promoters are found universally in cells from vastly different lineages, demonstrating that epigenetic modifications can help guide our understanding of gene expression during the development of many cell and tissue types (Heintzman et al., 2009).

Many types of covalent modifications have been associated with active gene transcription, but of these, histone acetylation and methylation are the best-characterized. Lysine acetylation in particular plays an important role in the activation of gene expression, as acetylation neutralizes the positive charge on lysine and weakens the interaction between DNA and histones, allowing chromatin decondensation and increasing the accessibility of DNA to transcription factor binding (Lee, Hayes, Pruss & Wolfe, 1993). Interestingly, histone methylation is a mark of active transcription only at certain amino acid residues. For example, methylation at lysine (K) 36 on histone 3 (H3) has been associated with the elongating form of RNA PolII and as a result is enriched in the coding regions and 3' end of genes being actively transcribed (Berger, 2007). Methylation of H3K4 has also been associated with transcriptional activation, and the state of methylation at this residue in particular has been used to distinguish between promoters and enhancers. Specifically, H3K4me3 is a mark of active and poised promoters, while regions of

H3K4me1 enrichment found outside of promoters have been used to identify tissue-specific enhancers (Hon, Hawkins & Ren, 2009).

With the exception of acetylation, most histone modifications are also associated with transcriptional repression. At the largest scale, chromatin is globally separated into two distinct states, a highly condensed form known as heterochromatin and a less condensed, transcriptionally active form known as euchromatin. Transcriptionally inactive heterochromatin is marked by hypoacetylation and high levels of methylation at H3K27 and H3K9. The high level of condensation in heterochromatin is largely maintained by the heterochromatin protein 1 (HP1), which promotes chromatin condensation at nearby genomic regions and is specifically recruited to methylated H3K9 through a protein domain called a chromodomain (Maison & Almouzni, 2004). The centromeres and telomeres of mitotic chromosomes are also enriched for H3K9me3 and HP1, suggesting that these factors play a distinct role in large-scale transcriptional repression (Lomberk, 2006). In euchromatin, the nucleosome organization is less condensed and chromatin is marked by greater levels of acetylation. The genes located in euchromatin may therefore be subject to local activation or repression, which occurs through recruitment of histone-modifying coregulators to the regulatory regions of target genes. H3K27me3 in particular mediates transcriptional repression in euchromatin, and notable examples of H3K27me3-mediated gene silencing occur at the Hox gene locus, in X chromosome inactivation and in genomic imprinting (Kouzarides, 2007). Unlike H3K27me3, the effect of H3K9 methylation on transcriptional regulation depends on cellular context, as H3K9me3 is enriched in heterochromatin but may also be found at the coding regions of actively transcribed genes (Vakoc, Mandat, Olenchock & Blobel, 2005). This observation demonstrates that the particular chromatin environment, including the presence of multiple histone modifications and position relative to a transcriptional target, allows for specific regulation of gene expression.

An interesting example of chromatin state complexity exists in embryonic stem cells, where regions may be marked by both active and repressive histone modifications. In particular, chromatin marked by H3K4me3 and H3K27me3 in embryonic stem (ES) cells was shown to be important for the maintenance of pluripotency. At the onset of differentiation, only one of the marks is maintained to allow for either transcriptional activation, associated with the presence of H3K4me3, or transcriptional repression, maintained by the presence of H3K27me3 (Bernstein et al., 2005). Examination of the genomic regions marked by bivalent histone modifications found that these were associated with key developmental transcription factors, suggesting that ES cells maintain their pluripotency by keeping the genes encoding lineage-determining transcription factors in a transcriptionally poised state, allowing for rapid activation or repression upon the initiation of differentiation. The regulators of chromatin modifications are therefore an important final consideration in exploring the regulation of transcription in specific cell lineages, including which enzymes are involved and how they function as transcriptional regulators for tissue-specific transcription factors.

1.3. Chromatin-modifying enzymes

Histone modifications are catalyzed by many families of modification- and often residue-specific enzymes. Globally associated with transcriptional activation, histone acetylation is catalyzed by coactivator proteins called histone acetyltransferases (HATs) and is removed through the action of histone deacetylases (HDACs). The first coactivator with HAT activity that was identified in higher eukaryotes was GCN5, which had previously been identified as a transcription coactivator (Brownell et al., 1996). In mammals, GCN5 and the highly similar p300/CBP-associated factor (PCAF) are both expressed, but only loss of GCN5 results in embryonic lethality (Yamauchi et al., 2000). GCN5/PCAF are distinct from other HATs, as they often form the catalytic subunit of larger coactivator complexes, namely the SAGA and ATAC complexes (Grant et al., 1997; Suganuma

et al., 2008). A related but distinct family of HATs that are ubiquitously expressed in mammals consists of the E1A binding protein p300 and the paralogous CREB-binding protein (CBP) (Goodman & Smolik, 2000). While both GCN5/PCAF and p300/CBP function as coactivators of gene transcription, the two families show differences in substrate specificity *in vivo*. Specifically, deletion of GCN5/PCAF has been associated with loss of acetylation at H3K9, while deletion of p300/CBP has been associated with loss of acetylation at H3K18 and H3K27 (Jin et al., 2011). The recruitment of GCN5/PCAF or p300/CBP to the regulatory regions of target genes is mediated by interactions with transcription factors bound to promoter or enhancer elements of target genes, and p300 specifically has been used as an enhancer-associated mark to help map the location of tissue-specific enhancers genome-wide (Visel et al., 2009). p300 has also been associated with the function of tissue-specific transcription factors during myogenic differentiation, and has been shown to mediate changes in histone acetylation at target gene promoters (Puri et al., 1997). Notably, the combined activities of PCAF and p300 function to activate transcription at muscle-specific genes, and therefore play important roles in myogenic differentiation.

The regulation of gene expression during development through local changes in histone acetylation also requires the action of HDACs. Similar to HATs, HDACs do not bind DNA directly, and are therefore recruited to target genetic loci by transcription factors or as part of larger corepressor protein complexes. The broad classification of HDACs results in three distinct families: class I HDACs, which are ubiquitously expressed; class II HDACs, which tend to be expressed in specific cell lineages; and class III HDACs, or NAD-dependent sirtuins (Haberland, Montgomery & Olson, 2009). A fourth class of HDACs contains only one member, HDAC11, although it is still unclear how its function differs from HDACs in other families. Notably, several HDACs have shown to be involved in muscle-specific gene expression. For example, deletion of HDAC1 and HDAC2 in the cardiac lineage results in upregulation of skeletal muscle-specific genes, while HDAC9 is a regulator of the skeletal muscle response to motor innervation (Mejat et

al., 2005; Montgomery et al., 2007). HATs and HDACs therefore maintain the balance of histone acetylation and deacetylation during development and during the differentiation of specific cell and tissue types.

As histone methylation is involved in both transcriptional activation and repression, histone methyltransferases often show greater substrate specificity than HATs. Methylation at H3K36 is mediated specifically by the histone methyltransferase NSD1, which associates with the phosphorylated C-terminal domain on RNA PolII during transcriptional elongation (Li, Moazed & Gygi, 2002). Although it was originally thought to be a stable modification, histone methylation can be reversed by histone demethylases. The demethylase LSD1 for example, can remove methylation from H3K4 and H3K9, however its ability to do so in vivo depends on its association with corepressor complexes such as Co-REST, in the case of H3K4me demethylation, or in a complex with the androgen receptor, in the case of H3K9me demethylation (Shi et al., 2005).

Methylation of H3K36 can additionally recruit a histone deacetylase to coding regions in order to prevent unwarranted histone acetylation and inappropriate recruitment of RNA polymerase to false transcription start sites, suggesting that histone modifications can confer multiple levels of regulation on gene expression (Carrozza et al., 2005). In a similar capacity, modifications with opposing functions can be prevented, such as in the case of H3K4me3 at active promoters that prevents binding of the general corepressor and histone deacetylase-containing complex NuRD (Nishioka et al., 2002). The genomic recruitment of coregulator complexes is mediated by protein subunits that contain domains that recognize histone tails with specific modifications, such as the chromodomain found in HP1 that recognizes methylation of H3K9. Similar to HP1, the polycomb-repressive complex 2 (PRC2) contains a chromodomain that recognizes methylated residues on H3 and mediates methylation of H3K27 (Min, Zhang & Xu, 2003). GCN5 can similarly be recruited to chromatin that has already been acetylated through a bromodomain, thus allowing

hyperacetylation at the locus and activation of transcription. A member of the ATP-dependent chromatin remodeling complex SWI/SNF, Swi2/Snf2 also contains a bromodomain domain that allows it to bind acetylated residues in histone tails and recruit SWI/SNF to allow the local rearrangement of nucleosomes in a way that promotes transcriptional activity (Hassan et al., 2002). Therefore, chromatin modifications function not only as final modifiers of chromatin structure, but also play an intermediate role in the recruitment of coregulators to target genetic loci. The activity of chromatin-modifying enzymes at a given locus can be limited by histone modifications previously deposited at the region, indicating that chromatin modifications act to establish and maintain the activity of regulatory DNA elements in a specific cell type.

1.4. Modern techniques in epigenetic research

Over the past two decades, the increasing efficiency and accessibility of nucleotide sequencing has allowed scientists to address a variety of biological questions concerning global gene expression and its regulation. The efficiency of next-generation sequencing (NGS) comes from recent advances in sequencing technology, where millions of DNA or RNA molecules are sequenced in parallel on microarray chips. Short sequencing reads of 30 to 50 bp are generated in multiple rounds of sequencing, and when compared to a reference genome, the genomic location of each read can be identified. In the case of RNA sequencing (RNA-seq), reads are generated from samples of total mRNA present in the cell. When these reads are aligned to the genome, the parent mRNA molecule can be identified after a process of transcript assembly and gene identification. NGS technology has also allowed the identification of regulatory DNA elements through a number of techniques aimed at studying the epigenetic features of DNA, such as genome-wide DNA accessibility using DNA-seq and transcription factor binding and chromatin modifications through chromatin immunoprecipitation followed by NGS (ChIP-seq) (Pepke, Wold & Mortazavi, 2009).

ChIP-seq in particular has been widely used to study the epigenetic regulation of gene expression. Initially, comparative genomics approaches could be used identify regulatory DNA elements based on high sequence constraint, suggesting a conserved function in transcriptional regulation, however this method could not predict which elements were active in a given cell type (Berman et al., 2004; Blow et al., 2010). In ChIP-seq, antibodies recognizing specific histone modifications and transcription factors are used to isolate regions of DNA associated with the factors of interest. DNA samples are then sequenced and mapped to a reference genome, allowing identification of regulatory DNA elements genome-wide. The mapping of chromatin modifications in many cell and tissue types using ChIP-seq has thus provided a solid framework to study the epigenome. The finding that certain regulatory elements are associated with specific histone modifications has additionally helped to identify networks of regulatory elements whose activity underlies cell type-specific gene expression. Integration of RNA-seq and ChIP-seq datasets therefore serves as a global approach to exploring the regulation of transcription in specific cell lineages and in the current study is applied to the characterization of transcriptional regulation during myoblast differentiation.

2. Transcriptional regulation during myogenesis

The specification, commitment and differentiation of myogenic progenitors from embryonic paraxial mesoderm is globally regulated by coordinated signaling pathways that induce sequential expression of several key developmental transcription factors. In particular, commitment to the muscle cell lineage and differentiation of myogenic progenitors is regulated by four master transcription factors known as the muscle regulatory factors (MRFs), each of which having the ability to convert non-muscle cell types into mature skeletal myocytes. MyoD was the first MRF discovered, but was soon followed by Myf5, myogenin and MRF4 (also known as Myf6 or

herculin), and the sequential expression of these transcription factors helps guide the progression of differentiation in myogenic progenitors (Braun, Buschhausen-Denker, Bober, Tannich & Arnold, 1989; Davis, Weintraub & Lassar, 1987; Edmonson & Olson, 1989; Rhodes & Konieczny, 1989).

2.1. Regulation of MRF expression

Myf5 and MyoD are the first MRFs to be expressed during development and have been shown to play overlapping roles in the commitment of mesodermal cells to the skeletal muscle cell lineage. While individual knockouts of Myf5 and MyoD did not have detrimental effects on the development of skeletal muscle tissue, double-null mice lacking both Myf5 and MyoD completely lacked skeletal muscle (Braun, Rudnicki, Arnold & Jaenisch 1992; Rudnicki, Braun, Hinuma & Jaenisch, 1992; Rudnicki et al., 1993). Although both factors are expressed during development, these experiments suggested that Myf5 and MyoD have functional redundancy and that expression of at least one is required for the development of muscle tissue. Subsequent to the original study documenting the phenotype of *Myf5:MyoD* null mice, it was found that the gene for MRF4, which lies approximately 8 kb from the *Myf5* locus, had also been impaired (Kassar-Duchossoy et al., 2004). The presence of a functional MRF4 however was not sufficient to restore normal tissue development in *Myf5:MyoD* null mice, suggesting that its function may largely overlap with the other MRFs. Notably, mice lacking myogenin do not generate skeletal myocytes and have an abundance of undifferentiated myoblasts (Hasty et al., 1993). These results suggested that myogenin and MRF4 lie downstream from Myf5 and MyoD, and that myogenin in particular is a fundamental regulator of the later stages of myogenic differentiation.

Upstream of the MRFs, the expression of the paired box transcription factor Pax3 plays an integral role in the specification of myogenic progenitors in embryonic mesoderm (Ridgeway & Skerjank, 2001). Following gastrulation, elongation of the vertebrate embryo along the anterior-posterior

axis occurs in parallel with the condensation of paraxial mesoderm into distinguishable blocks called somites, which are distributed symmetrically on either side of the neural tube. The process of somite formation, or somitogenesis, is regulated by opposing gradients of retinoic acid (RA) and fibroblast growth factor (FGF), which play critical roles during development in the establishment of body axis polarity (Niederreither & Dolle, 2008). The relative levels of RA and FGF thus help determine the pattern of gene expression in mesodermal cells during somitogenesis and affect the competence of these cells to respond to intercellular signals leading to the specification of somitic cells to the myogenic lineage. As Pax3 is initially ubiquitously expressed throughout the somite, the integration of these signals gradually restricts Pax3 expression to the dorsal region of the somite, known as the dermomyotome, thus marking the specification of mesodermal cells to the myogenic lineage (Goulding, Lumsden & Paquette, 1994).

Several signaling molecules secreted by the tissues surrounding the somite have been shown to play a role in myogenic specification, including Wnts from the surface ectoderm and neural tube, sonic hedgehog (Shh) from the notochord and bone morphogenic protein (BMP) from the lateral plate mesoderm. Several Wnts are involved in early myogenic specification, including Wnt1 and Wnt3 that are secreted from the dorsal neural tube, and Wnt4, Wnt6 and Wnt7a that are secreted by the surface ectoderm (Tajbakhsh et al., 1998). Wnts 1 and 3 bind to frizzled transmembrane receptors Fzd1 and Fzd6 on cells in the dorsomedial region of the somite, known as the epaxial dermoymotome, while Wnt7a binds to Fzd7 in cells in the dorsolateral region of the somite known as the hypaxial dermomyotome. Intriguingly, the epaxial and hypaxial domains of the dermomyotome are marked by preferential expression of Myf5 and MyoD, respectively, suggesting that the expression of Myf5 and MyoD is differentially regulated during development (Tajbakhsh, Rocancourt, Cossu & Buckingham, 1997).

The presence of two distinct populations of committed myogenic progenitors in the dermomyotome indicated that the molecular events leading to the expression of Myf5 and MyoD may differ. It was later found that Myf5 is regulated preferentially by Wnt1 and canonical, or beta-catenin-dependent, Wnt signaling in cells of the epaxial dermomyotome. In contrast, MyoD expression is regulated preferentially by Wnt7a in the hypaxial dermomyotome through non-canonical Wnt signaling. In support of these observations, the absence of the transcription coactivator beta-catenin, Myf5 expression in presomitic mesoderm is not induced by Wnt1, while the activation of MyoD expression by Wnt7a is unaffected (Brunelli, Relaix, Baesso, Buckingham & Cossu, 2007). The differences underlying the regulation of Myf5 and MyoD transcription through Wnt signaling was further characterized in mice lacking Pax3, as in the absence of both beta-catenin and Pax3, the expression of MyoD could not be activated by signals from the surface ectoderm. Mice lacking Myf5, Mrf4 and Pax3 also completely lack skeletal muscle, indicating that activation of MyoD through non-canonical Wnt signaling requires a functionally active Pax3 (Tajbakhsh, Rocancourt, Cossu & Buckingham, 1997).

In addition to Wnts, the activity of Shh is also required to activate the expression of the MRFs in the dermomyotome (Munsterberg, Kitajewski, Bumcrot, McMahon & Lassar, 1995). Paraxial mesoderm at different stages of somitogenesis shows varying competence to respond to Wnt signals from the dorsal neural tube and surface ectoderm. It was determined that competence to respond to different Wnt signals is dependent on Shh secreted from the notochord. Notably, Shh cooperates with Wnts from the dorsal neural tube to activate the expression of Myf5 in the epaxial dermomyotome, but it is does not cooperate with Wnt signaling from the surface ectoderm to activate MyoD expression in the hypaxial myotome (Cossu & Borello, 1999).

Unlike Wnt and Shh, BMP signaling is a negative regulator of MRF transcription and commitment of dermomyotomal cells to the myogenic lineage. When BMP4 is administered to paraxial

mesoderm cultured with Shh and Wnt1, the expression of the MRFs is blocked. Pax3 expression is not affected however, suggesting that BMP signaling prevents myogenesis downstream of the induction of Pax3 expression and myogenic specification (Reshef, Maroto & Lassar, 1998). Additionally, in the presence of the BMP antagonist Noggin, MyoD expression is induced in paraxial mesoderm by signals from the surface ectoderm, supporting a role for BMPs in the inhibition of early expression of myogenic markers in the developing somites. During development, inhibitors of BMP signaling, including noggin as well as chordin and follistatin, are secreted by somitic cells and promote the activation of Myf5 and MyoD expression. The specification of myogenic progenitors in the dermomyotome therefore requires coordination of Wnt, Shh and BMP signaling pathways to ultimately result in downregulation of the somitic marker Pax3 and expression of the MRFs.

Postnatal myogenesis, as observed during muscle regeneration, follows similar paths of regulation as during embryonic development. The ability of skeletal muscle to regenerate after injury is made possible by the presence of a population of committed muscle progenitors known as satellite cells. In adults, these cells are found adjacent to the myofiber plasma membrane, and they are mitotically quiescent under resting conditions. Upon injury, satellite cells rapidly enter the cell cycle to create a pool of proliferating myoblasts that can undergo differentiation and fuse with the existing myofiber. Notably, these cells undergo asymmetric division to give rise to both quiescent satellite cells and committed muscle progenitors leading to either self-renewal or differentiation, respectively (Kuang, Kuroda, Le Grand & Rudnicki, 2007). In addition, under certain conditions satellite cells can differentiate into non-muscle cell types such as adipocytes and osteocytes, indicating that these cells are less committed to the muscle lineage than proliferating myoblasts and maintain their stemness by preventing the progression of differentiation. Although several satellite cell markers have been described, the Pax3 paralogue Pax7 is a ubiquitous marker of satellite cells as it is specifically expressed in both quiescent

satellite cells and satellite cell-derived proliferating myoblasts (Seale et al., 2000). Satellite cells also express Pax3 and the MRFs, however these myogenic markers are only expressed in a subset of cells. Therefore, these cells differ substantially from embryonic myogenic progenitors in the regulation of gene expression during the early stages of myogenic commitment and differentiation.

2.2. Regulation of muscle-specific gene expression during myoblast differentiation

The sequential expression of the MRFs during myogenesis indicates that these transcription factors also have functional differences in the regulation of transcription. The MRFs are members of the basic helix-loop-helix (bHLH) family of transcription factors, which contain a basic domain that mediates DNA binding and a helix-loop-helix motif that allows dimerization with other bHLH proteins. A notable group of bHLH transcription factors involved in the regulation of myogenesis with the MRFs are the E proteins, and although they are ubiquitously expressed, the E proteins form heterodimers with MyoD and myogenin and mediate their DNA binding in the regulatory regions of muscle-specific genes (Lassar et al., 1991). However, as the MRFs recognize and bind highly similar DNA sequences, how each transcription factor is specifically recruited to regulatory regions of target genes and how each differentially regulates transcription at a given locus and stage of development is not well understood. A recent comparison of the activities of Myf5 and MyoD in myogenic specification and commitment found that the two factors bind a largely shared set of DNA binding sites, but differ in their potential to activate gene transcription (Conerly, Yao, Zhong, Groudine & Tapscott, 2016). As MyoD and myogenin show similar expression at the early stages of differentiation, these results suggest that MyoD and myogenin may also play distinct roles in the regulation of transcription.

Part of the diversity underlying the transcriptional activity of the MRFs results from their association with E proteins, as each protein contributes a unique DNA binding half site to the heterodimer. MRF-E protein complexes therefore have differing affinities for their shared DNA sequence motif, called an E box, which has the consensus sequence CANNTG. In addition to the E proteins, the myocyte enhancer factor-2 (Mef2) family of transcription factors has also been shown to interact with the MRFs and play an important role during myogenic differentiation. The Mef2 family consists of four members, Mef2a, Mef2b, Mef2c and Mef2d, each containing a conserved MADS DNA-binding domain. Although Mef2 proteins are expressed in many tissues, DNA binding activity is most prominent in skeletal muscle cells and neurons, suggesting that the Mef2 family plays a notable role during myogenesis. Accordingly, the DNA binding motif for Mef2 proteins has been identified in the regulatory regions of many muscle-specific genes, including essential binding sites in the promoters of MyoD and myogenin (Edmondson, Cheng, Cserjesi, Chakraborty & Olson, 1992; Leibham et al., 1994).

Although members of the Mef2 family are highly similar in their MADS domain, their amino acid sequences diverge at the C-terminus, suggesting variability among the Mef2 proteins in their transcriptional activation potential and association with the myogenic transcription factors. In support of differential roles for Mef2 subtypes in myogenesis, it was shown that Mef2c specifically is involved in a reciprocal regulatory circle with myogenin, as each was found to regulate the expression of the other (Ridgeway, Wilton & Skerjanc, 2000). Mef2c also functions to coregulate the activity of the MRFs at regulatory regions of myogenic genes, as Mef2c was shown to synergistically activate the expression of myogenic genes in fibroblasts transfected with MyoD or myogenin and Mef2c compared to fibroblasts transfected with MyoD or myogenin alone (Molkentin, Black, Martin & Olson, 1995). The cross-regulation between the MRFs and Mef2 proteins thus supports an important role for the latter in the regulation of MRF function.

In addition to the synergistic activity provided by the E proteins and Mef2 transcription factors, the activity of the MRFs can be negatively regulated directly or indirectly through the protein binding ability of other transcriptional regulators. For example, the bHLH protein Twist has been shown to negatively regulate myogenic differentiation by forming dimers with MyoD and preventing it from binding to DNA (Hamamori, Wu, Sartorelli & Kedes, 1997). The inhibitor of DNA binding (Id) proteins are also negative regulators of differentiation and function by binding to E proteins and the MRFs and preventing their dimerization. Id family members contain an HLH domain that mediates their interactions with bHLH proteins, but their lack of a basic N-terminus prevents them and the proteins they interact with from binding DNA (Langlands, Yin, Anand & Prochownik, 1997). Therefore, the activity of the MRFs is regulated in part by their association with other distinct families of transcription factors.

Transcriptional activation through the MRFs similarly depends on their association with transcriptional coactivators. In particular, p300 and PCAF are required for the activation of muscle-specific gene expression by MyoD, and the inhibition of either results in the inhibition of myogenic differentiation and MyoD-dependent gene transcription (Puri et al., 1997) (Figure 1). MyoD interacts with both p300 and PCAF, but while only the HAT activity of PCAF is required for transcriptional activation by MyoD at the early stages of differentiation, the HAT activity of p300 is required for terminal differentiation and cell fusion (Polesskaya et al., 2000). p300 and PCAF play an additional role in the regulation of myogenesis through direct acetylation of MyoD, and as a result, MyoD has a greater affinity for binding DNA at regulatory regions of target genes (Sartorelli et al., 1999). The role of p300 in myogenesis therefore extends beyond local changes in chromatin acetylation and is important for the direct transcriptional activity of MyoD.

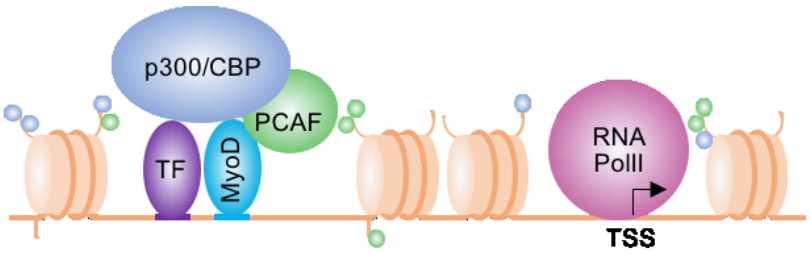

Figure 1. Transcriptional activation by MyoD. MyoD binds to DNA as dimers with other bHLH or Mef2 transcription factors. p300 and PCAF are recruited to enhancers by MyoD and mediate histone acetylation and transcriptional activation at target gene promoters.

3. Transcriptional regulation through nuclear receptor signaling

Intercellular signaling plays an integral role in the coordination of cellular differentiation and tissue formation. Many natural intercellular signaling molecules regulate transcription indirectly by binding ligand-specific membrane receptors and triggering the activation of intracellular signaling cascades. However, a group of ligand-inducible transcription factors known as nuclear receptors respond to lipid-soluble signaling molecules and allow the direct regulation of target genes involved in a variety of cellular processes. Accordingly, ligands targeting nuclear receptors have been used in the treatment of a number of diseases, such as retinoic acid in acute promyelocytic leukemia, tamixofen in breast cancer and thiazolidinediones in type II diabetes (Germain, Staels, Dacquet, Spedding & Laudet, 2006). The identification of nuclear receptor signaling pathways involved in skeletal muscle development may therefore allow targeted regulation of gene expression through the application of nuclear receptor ligands. We have previously found that a synthetic agonist specific for the retinoid X receptor (RXR) enhances myoblast differentiation (AlSudais et al., 2015). As a necessary dimerization partner for many nuclear receptors, RXR in particular is involved in the regulation of genes of many biological pathways, but it remains unclear how activation of RXR may mediate transcription under various conditions and how RXR signaling is specifically involved in myogenic differentiation.

3.1. Classification of nuclear receptors

Most nuclear receptors are classified as one of two main types, which are distinguishable by type of ligand and mode of activation. The general classification of nuclear receptors as type I or type II distinguishes between receptors bound by steroid and nonsteroid molecules, respectively. Steroid hormones such as estrogen, progesterone and testosterone are ligands for type I nuclear receptors and are secreted by endocrine glands, allowing them to travel towards target tissues through the bloodstream. In the absence of ligand, type I receptors are kept in the cytoplasm as monomers by chaperone proteins, but upon ligand binding, the receptors are released and translocate into the nucleus where they bind DNA as homodimers in the promoters of target genes.

Type II nuclear receptors differ substantially from type I receptors in their function as transcriptional regulators. Type II receptors are activated by nonsteroid molecules, such as RA for the retinoic acid receptor (RAR) and thyroid hormone for the thyroid hormone receptor (TR), which have important roles during embryonic development and in post-natal growth and metabolism, respectively. Other members of the type II nuclear receptors, including the farnesoid X receptor (FXR), liver X receptor (LXR), peroxisome proliferator-activated receptor (PPAR) and vitamin D receptor (VDR), also play important roles in development and cellular differentiation. Unlike type I receptors, type II nuclear receptors are bound constitutively to DNA with RXR, and in the absence of ligand binding, nuclear receptor heterodimers associate with transcriptional corepressors to inhibit transcription at the target gene (McKenna & O'Malley, 2002). The corepressor complexes associated with type II nuclear receptors contain histone deacetylases that prevent recruitment of RNA polymerase to target gene promoters and transcriptional activation, and include corepressors such as nuclear receptor corepressor (NCoR) and silencing mediator for retinoid and thyroid hormone receptors (SMRT). These complexes associate with

unliganded nuclear receptor dimers, particularly RAR and TR, and although they lack intrinsic enzymatic activity, they recruit HDACs that deacetylate histones and repress transcription at target gene promoters (Gronemeyer, Gustafsson & Laudet, 2004). When bound by a ligand, a conformational change is induced in the nuclear receptor dimer resulting in release of corepressors and association with coactivators, including p300, which is required for ligand-induced transcription of nuclear receptor target genes (Figure 2).

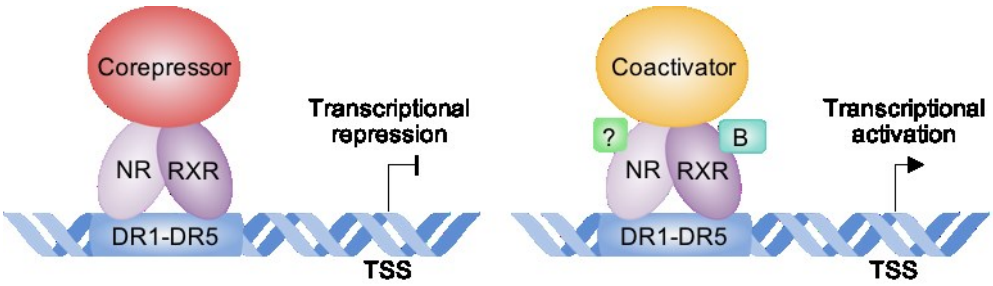

Figure 2. Transcriptional regulation through type II nuclear receptors. RXR and its partner receptor are bound constitutively to DNA at hormone response elements made of direct repeats (DR) separated by 1-5 bp. In the absence of ligand, the nuclear receptor dimer associates with transcriptional corepressors to inhibit transcription at target genes. In the presence of ligand, such as bexarotene (B) for RXR, a conformational shift of the dimer allows the release of corepressors and association with coactivators, leading to transcriptional activation at the target gene promoter.

3.2. Nuclear receptor structure

All nuclear receptors have a common structure composed of 5 or 6 domains. The most highly conserved domains are the ligand-binding domain (LBD) and DNA-binding domain (DBD), which show remarkable specificity towards a specific ligand and DNA sequence, respectively. It was shown that chimeric receptors containing the LBD from one receptor and the DBD from another are activated in response to the ligand for the LBD but regulate gene transcription at regulatory elements containing binding sites for the specific DBD (Green & Chambon, 1988). The LBD itself consists of four structurally distinct subunits, which mediate dimerization, ligand binding,

coregulator associations and receptor transactivation. The DBD binds DNA sequences called hormone response elements (HREs), which are specific to the nuclear receptor. The domain between the LBD and DBD is referred to as the D region and acts as a hinge between the two conserved domains. This domain may play an important role in the conformational shift observed in the protein upon ligand binding, leading to association or dissociation with coregulator complexes or other transcription factors. The N-terminal domain, also called activation function 1 (AF-1), is the site for post-translational modifications in several nuclear receptors including RAR. The transactivation domain, also called activation function 2 (AF-2), is located at the C-terminus of the receptor and undergoes a change in conformation following ligand binding. AF-1 and AF-2 are also referred to as the constitutive and ligand-dependent activation domains, respectively, allowing for tight control of nuclear receptor activity, differential recruitment of coregulators to nuclear receptor-bound regulatory elements and regulation of target gene transcription (Huang, Chandra & Rastinejad, 2010).

Almost all nonsteroid nuclear receptors, including RXR, bind to DNA at sites with the core sequence AGGTCA. HREs are composed of two repeats of this consensus sequence, and each is referred to as a half site and separated by a short DNA spacer. Half sites may be relatively positioned in a direct, indirect or inverted position, conferring additional specificity to the HRE for a given nuclear receptor dimer. Nuclear receptors in heterodimers with RXR bind to direct repeats (DRs) separated by one to five nucleotides, known as DR1 to DR5, and dimerize in a head-to-tail, or N-terminus to C-terminus, fashion (Rastinejad, Perlmann, Evans & Sigler, 1995). Given its characteristic property of acting as a DNA binding partner for many type II nuclear receptors, RXR in particular is involved in the regulation of genes involved in a wide variety of biological functions. Transcriptional regulation resulting from activation of RXR depends largely on its partner receptor. In certain heterodimers, RXR activation is permissive in activating transcription of target genes, however in other heterodimers, RXR activation alone is not sufficient to activate transcription and

requires activation of the partner receptor. The genetic targets of RXR can therefore vary in different cellular contexts, making their identification more difficult. In addition, the natural ligand for RXR, 9-*cis* retinoic acid, can also bind RAR. Although a ligand specific to RXR is not naturally synthesized in vivo, synthetic ligands called rexinoids were created to selectively target RXR. Bexarotene, also known as LGD 1069 or Targretin, is a RXR-specific agonist used clinically in the treatment of cutaneous T cell lymphoma (Duvic et al., 2001). Given the variety of genes targeted by nuclear receptor signaling and by RXR, it is unclear how bexarotene mediates transcription in differentiating myoblasts ultimately allowing for an increase in the rate of differentiation.

Hypothesis and Objectives

As the progression of myogenic differentiation is coupled with global changes in gene expression, and bexarotene was shown to promote the differentiation of skeletal myoblasts, we hypothesized that bexarotene may globally promote the regulation of muscle-specific gene expression.

Using a systematic analysis of global transcriptional and epigenetic data, our objectives were to (1) study the epigenetic landscape in committed muscle cell progenitors and classify observed patterns of chromatin modifications based on their potential role in the regulation of transcription, (2) quantify and characterize transcriptional regulation occurring during early myogenic differentiation and identify the genetic targets of bexarotene, and
(3) explore the molecular mechanisms and factors involved in bexarotene-enhanced myogenic differentiation.

Methods

Cell culture

Cells of the murine myoblasts cell line C2C12 were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum and penicillin/streptomycin at 37ºC and 5% CO_2 and were passaged at 60% confluency. Undifferentiated C2C12 cells were collected at the induction of differentiation, which was initiated when the cells reached 80-90% confluency. To initiate differentiation, growth media was replaced with low-mitogen media made from DMEM with 2% horse serum and penicillin/streptomycin, and the media of bexarotene-treated cells was supplemented with bexarotene dissolved in ethanol at 50 nM.

RNA-seq and data processing

Proliferating C2C12 myoblasts were collected prior to the induction of differentiation as described, and differentiating cells were collected after 12 and 24 hours. Total RNA was extracted using the RNeasy kit (Qiagen) according to manufacturer's instructions, and the quality and concentration of RNA samples was confirmed using the Agilent 2100 Bioanalyzer. Samples were sent to the McGill University and Genome Quebec Innovation Centre for cDNA library construction and sequencing.

RNA-seq data was processed following the Cufflinks pipeline. Sequencing reads were aligned to the mouse genome build mm9 using Tophat and guided by transcripts from the Ensembl 67 database. Transcript assembly was performed using Cufflinks and was followed by using Cuffdiff to estimate normalized gene expression and perform differential expression testing.

RNA-seq analysis

To identify genes showing differential expression between conditions for downstream analysis, genes with a fold change in expression greater than ±1.5-fold and Bonferroni-adjusted p-value less than 0.05 were selected. Gene ontology terms significantly associated with genes that were differentially expressed (Bonferroni-adjusted p-value < 0.05) were identified using the database for annotation, visualization and integrated discovery (DAVID) (Huang, Sherman & Lempicki, 2009). For visualization and hierarchical clustering, raw FPKM values were log2-transformed and scaled to the row mean and standard deviation using the following formula: (log2(FPKM)-mean of row)/(standard deviation of row). The Spearman correlation coefficient between gene expression across conditions was calculated using raw FPKM values.

ChIP-seq and data processing

Undifferentiated cells were collected as described, and differentiating cells were collected after 24 hours. Cells were crosslinked in growth media with 1% formaldehyde for 5 minutes and 1.25 mM glycine was used to stop the reaction. Cells were then washed in phosphate buffered saline and incubated on ice for 10 minutes in ChIP lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 1% SDS and 1X protease inhibitors). DNA shearing was performed by sonication using the Diagenode biorupter, and isolated chromatin was purified using the Cycle Pure Kit (Omega Bio-Tek) according to manufacturer's instructions. Purified chromatin was diluted, and pre-clearing with Dynabeads Protein A (Life Technologies) was performed prior to overnight incubation with antibodies for H4K8ac, H3K9ac, H3K18ac, H3K127ac, H3K27me3 and rabbit IgG. Chromatin precipitation was then performed using Dynabeads Protein A, which were then washed for 10 minutes in each of three washing buffers and TE buffer, and samples were incubated overnight at 65ºC for reverse-crosslinking. RNase A and proteinase K were added to digest RNA and

proteins, respectively. DNA was then purified using the Cycle Pure Kit (Omega) and sent to Génome Quebec for single end sequencing.

ChIP-seq reads were mapped to the mouse genome build mm9 using Bowtie, allowing for 3 mismatches and reporting the single best alignment per read. For visualization, aligned reads were extended by 125 bp at their 3' end and basewise signal intensity was computed. Local peaks in read enrichment were identified using MACS version 2.0 (Zhang et al., 2008) with default settings for p300. For RXR, modeling and estimation of ChIP fragment length was supressed and reads were extended manually by 150 bp at their 3' end.

Raw sequencing reads for the histone marks H3K4me1, H3K4me3 and H3K36me3, as well as for RNA polymerase II, were obtained from the NCBI Gene Expression Omnibus (GEO) under the accession numbers GSM721288 (H3K4me1), GSM918415 (H3K4me3, generated by ENCODE), GSM918417 (H3K36me3, generated by ENCODE) and GSM721286 (RNA PolII) (Asp et al., 2011; Yue et al., 2014). The corresponding input datasets for H3K4me1 and RNA PolII were obtained under the accession GSM721306, and input datasets for H3K4me3 and H3K36me3 under the accession GSM918421. All raw sequencing reads were aligned to the mouse genome build mm9 using Bowtie as described, and in the case of replicate samples, duplicate reads were removed for individual replicates before being merged into a single alignment. Genome-wide binding sites for CTCF, MyoD and myogenin were generated by ENCODE and obtained from GEO under the accessions GSM915188 (CTCF), GSM915186 and GSM915183 (MyoD) and GSM915159 (myogenin).

Chromatin state model

The chromatin state model was generated using ChromHMM (Ernst & Kellis, 2012), which uses a hidden Markov model-based approach to estimate the genome-wide co-occurrence and distribution of chromatin modifications. Briefly, the genome was partitioned into 200 bp bins and each bin was assigned a binary indicator for the presence or absence of each chromatin mark. Binarization of read alignments was computed using a Poisson background model normalized to the corresponding input data.

The enrichment of TSSs, transcription factor binding sites and highly conserved non-coding elements (HCNCEs) was calculated as a ratio between the fraction of nucleotides overlapping between the feature and state and the joint probability of observing the feature and state. Enrichment for a feature in a given state was therefore calculated as follows: (number of base pairs occupied by feature and state/number of base pairs in genome)/(number of base pairs covered by feature/number of base pairs in genome)(number of base pairs occupied by state/number of base pairs in genome). HCNCEs for the mm9 build were identified by the Sidow lab at Stanford University using genomic evolutionary rate profiling (GERP) and obtained from http://mendel.stanford.edu/SidowLab/downloads/gerp/ (Davydov et al., 2010).

Analysis of transcription factor binding sites

The enrichment of histone modifications at TSSs, p300-bound loci and transcription factor binding sites were calculated as the number of reads in 20 bp bins in a 2 kb region centered at the TSS or p300 or transcription factor peak centre, normalized to the total number of mapped reads (in millions) in the dataset. For enrichment profiles at TSSs, and MRF and RXR ChIP-seq peaks, the average read enrichment in each 20 bp bin was calculated individually for each histone mark in each condition. For heatmap visualization of p300 peaks, k-means clustering was performed

based on local enrichment of H3K18ac and H3K27ac in proliferating and differentiating myoblasts. p300 peaks in each cluster were assigned to their nearest regulatory element, and the overlap between a given p300 cluster and enhancer class was calculated as the number of peaks from the cluster that fell within 1 kb of an enhancer.

MyoD binding sites that were present in proliferating and differentiating myoblasts were identified if the centre of peaks from each condition fell within 100 bp. The identification of shared MyoD and myogenin binding sites in differentiating myoblasts was performed similarly. For detailed analysis of binding site co-occupancy, MyoD and myogenin peaks were first ranked by signal intensity. The highest ranking 20,000 peaks were grouped into bins of 2,000 to 20,000, increasing in size by intervals of 2,000, such that we isolated the top 2,000 peaks, and then the top 4,000, 6,000, etc. The fraction of shared binding sites between each group of peaks was calculated as the number of overlapping binding sites divided by the number of binding sites in the smallest bin. *De novo* motif analysis on p300 and MyoD ChIP-seq peaks was performed using Homer, allowing for motif identification within a 200 bp region at the peak centre.

Results

Gene expression is dynamically regulated during early myogenic differentiation

To quantify the global changes in gene expression occurring during the early stages of myogenic differentiation, we performed RNA-seq in proliferating and differentiating C2C12 myoblasts (Figure 3a). By 24 hours of differentiation, 7,102 genes showed a difference in expression greater than ±1.5-fold compared to their expression in proliferating cells, including 3,162 genes that were upregulated and 3,940 genes that were downregulated. Closer observation of these genes determined that many also showed significant changes in expression by 12 hours of differentiation. Notably, most of the changes in gene expression that were observed by 24 hours had occurred during the first 12 hours of differentiation, with only 9% of upregulated genes and 20% of downregulated genes showing a further change in expression between 12 and 24 hours, suggesting that gene regulation at this stage of differentiation occurs in two distinguishable phases (Figure 3b-c). While genes upregulated during the initial 12 hours of differentiation were found to be involved in many developmental processes, genes that were continuously upregulated or upregulated only between 12 and 24 hours were particularly involved in muscle cell differentiation and muscle tissue formation and function (Figure 3d). A complementary decrease in the expression of genes involved in cell cycle regulation and progression reflected the simultaneous exit from the cell cycle and activation of myogenic differentiation.

Global analysis of chromatin states identifies tissue-specific regulatory regions in muscle cell progenitors

To study the nature of *cis*-regulatory elements involved in the regulation of muscle-specific gene expression, we generated a chromatin state model based on genome-wide enrichment for several histone modifications associated with distinct activities in transcription regulation. Incorporating

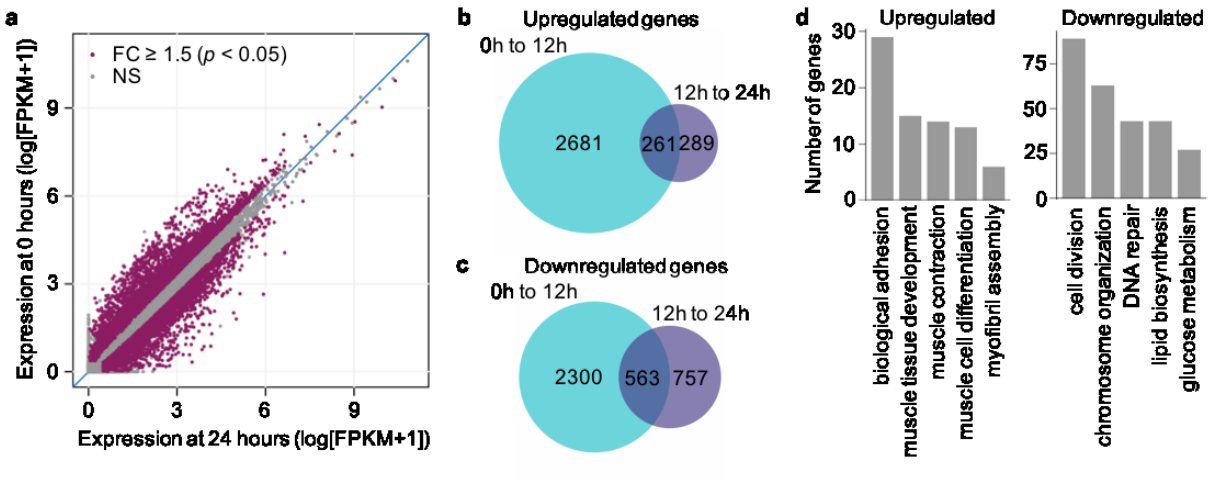

Figure 3. Global gene expression during early myogenic differentiation. **a.** Gene expression measured by RNA-seq in fragments per kilobase per million mapped reads (FPKM) for all annotated Ensembl genes. Genes showing a significant fold change (FC) in expression greater than 1.5 are shown in purple, and remaining genes are shown in gray. **b.** Overlap between genes that were upregulated during the first 12 hours of differentiation and genes that were upregulated between 12 and 24 hours of differentiation. Similarly shown is the overlap between genes that were downregulated during the first 12 hours of differentiation and genes that were downregulated between 12 and 24 hours of differentiation (**c.**). **d.** Gene ontology terms associated with genes that were upregulated and downregulated between 12 and 24 hours of differentiation.

ChIP-seq data from proliferating C2C12 myoblasts for the promoter-associated marks RNA PolII, H3K4me3 and H3K9ac, the enhancer-associated marks H4K8ac, H3K18ac, H3K27ac and H3K4me1, the transcription elongation mark H3K36me3 and the repressive mark H3K27me3, we used a hidden Markov model-based method to generate a 14-state model of chromatin states observed in proliferating skeletal muscle progenitors.

Each state in the model was characterized by a distinct combination of histone marks (Figure 4a). We noted four promoter states (states 1-4) that were marked by H3K4me3 and varying levels of histone acetylation and RNA PolII occupancy and that were generally also highly enriched in transcription start sites (TSSs). One state (state 5) was marked by moderate acetylation but not by methylation at H3K4, suggesting that this state may be functionally active in transcriptional regulation but was not representative of either promoter or enhancer states. This state showed relatively high enrichment for CTCF however, suggesting that it may characterize the chromatin state at insulator elements. Two states (states 6-7) showing enrichment for H3K36me3 identified regions of transcriptional elongation, while four states (states 8-11) were marked by the enhancer-associated mark H3K4me1 and by varying levels of histone acetylation. The four classes of enhancers included active enhancer states marked by enrichment of H3K27ac, as well as poised enhancers that were marked by H3K4me1 but that lacked H3K27ac. Interestingly, MyoD was found associated with both promoters and enhancers but was particularly enriched in active enhancer states, suggesting that MyoD is involved in maintaining the activity of tissue-specific enhancers at this stage of myogenesis. Finally, regions lacking any histone modifications were classified as inactive or as heterochromatin (states 12 and 13, respectively), while a polycomb-repressed state (state 14) was characterized by the presence of the PRC2-associated mark H3K27me3.

Similar to what has been reported previously, the majority of the genome (~80%) was not marked by any chromatin modifications (Ernst et al., 2011). Promoter and enhancer regions covered 1.3% and 6.8% of the genome, respectively, and had shorter median lengths than states that were inactive, repressed or identified as heterochromatin. While the latter results reflect the predictably smaller size of regulatory regions bound by transcription factors compared to inactive genomic loci, the former suggests that only a small portion of the genome is utilized to actively regulate transcription within the muscle cell lineage.

In mammals, non-coding regulatory DNA elements such as promoters and enhancers often show higher evolutionary conservation than non-functional genomic loci, due to their binding by sequence-specific transcription factors and importance in regulating gene expression during development. Here, we found that active promoters and enhancers were enriched in highly conserved non-coding elements (HCNCEs), which were previously identified for the mouse genome through genomic evolutionary rate profiling (GERP) using multiple mammalian sequence alignments (Davydov et al., 2010). Promoter states showed the highest enrichment in HCNCEs, perhaps reflecting the high evolutionary constraint on core promoter sequences that recruit the basal transcriptional machinery. Accordingly, promoters occupied by RNA PolII showed the highest enrichment in HCNCEs. We also found a moderate enrichment for HCNCEs at insulators, which may reflect the conservation of DNA binding sequences for CTCF. Enhancers showed a more modest but noticeable enrichment in conserved elements compared to inactive or heterochromatic chromatin regions, and especially at strong enhancers marked by high enrichment of MyoD. Together, our data supports the identification of tissue- and developmental stage-specific regulatory elements through distinct chromatin states observed in muscle cell progenitors.

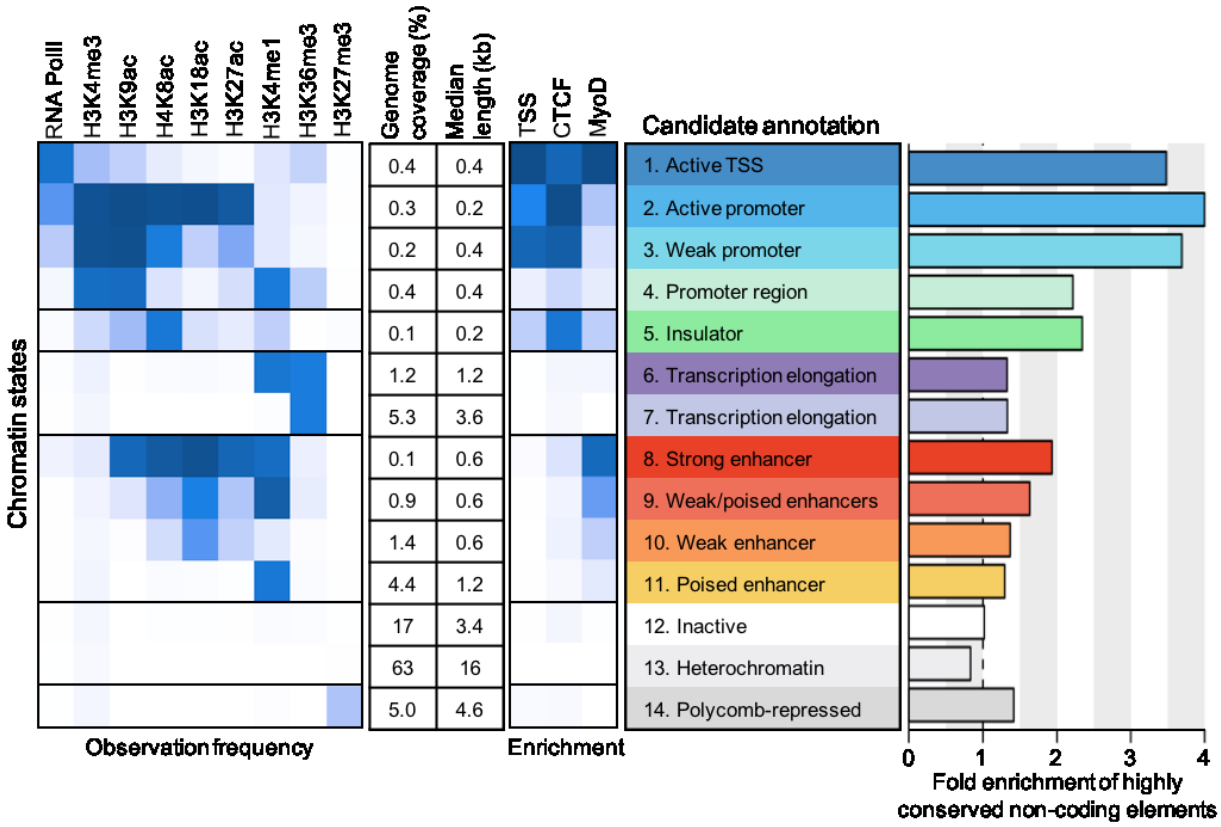

Figure 4. Characterization of the epigenome in proliferating myoblasts. A 14-state chromatin state model was generated based on global ChIP-seq read enrichment for several histone modifications in undifferentiated C2C12 myoblasts. Datasets for RNA PolII (GSE), H3K4me3 (GSE), H3K4me1 (GSE) and H3K36me3 (GSE) and corresponding input datasets were accessed from the NCBI Gene Expression Omnibus. The enrichment of transcription start sites (TSS), CTCF and MyoD was calculated as the ratio between the fraction of bases in the genome overlapping the feature and state and the joint probability that a base would overlap with the feature and state. Enrichment was similarly calculated for highly conserved non-coding elements (Davydov et al., 2010).

Poised enhancers become active following the initiation of differentiation

In order to assess the tissue-specific regulation of gene expression in differentiating myoblasts, we identified various classes of enhancers largely distinguishable by differential enrichment in histone acetylation. As histone acetylation is associated with gene expression, we hypothesized that poised enhancers marked by enrichment in H3K4me1 but not in H3K27ac may show an increase in acetylation following the initiation of differentiation. Similar to H3K4me1, the genome-wide localization of p300 is used to predict the location of tissue-specific enhancers. In addition, H3K4me1 and p300 may mark both active and poised enhancers (Zentner, Tesar & Scacheri, 2011). Therefore, to explore if these enhancers adopted an epigenetic signature characteristic of active enhancers at the early stages of differentiation, we performed ChIP-seq for p300 in C2C12 myoblasts and computed the normalized read enrichment for the p300-associated marks H3K18ac and H3K27ac in proliferating and differentiating myoblasts in a 2 kb region centered at the identified p300 peaks (Figure 5a). We further used k-means clustering to determine if the changes in acetylation observed at p300 peaks during myogenic differentiation were representative of the classes of enhancers identified in our chromatin state model. Consistent with our initial finding, four clusters of p300 peaks could be distinguished based on local enrichment of H3K18ac and H3K27ac. In proliferating myoblasts, the enrichment of H3K18ac and H3K27ac was highest in cluster 1, followed by clusters 3, 4 and 2. These clusters therefore showed similarity to the enrichment in acetylation observed in strong, weak/poised, weak and poised enhancers, respectively. Interestingly, the enrichment of H3K18ac and H3K27ac increased modestly in clusters 2 and 4 by 24 hours of differentiation, suggesting that the transcriptional activity of these p300-bound enhancers may increase at the early stages of differentiation.

To determine if differentiation-responsive p300-marked loci showed a particular association with a class of enhancer, we identified the fraction of peaks in each cluster that fell within 1 kb of an enhancer and found that each class of enhancer showed a preferential association with a specific p300 cluster (Figure 5b). Strong enhancers showed the greatest association with p300 peaks from cluster 1, which were highly enriched in histone acetylation in both proliferating and differentiating myoblasts. Although its association was more modest, weak/poised enhancers showed the greatest association with cluster 4, suggesting that these enhancers may respond to the initiation of differentiation with an increase in histone acetylation. In contrast, weak enhancers showed the greatest association with cluster 3, suggesting that these enhancers show a differentiation-dependent loss of activity and that enhancers marked by low enrichment in H3K4me1 may not play a particular role in transcriptional regulation during differentiation. Finally, poised enhancers marked only by high enrichment of H3K4me1, showed the greatest association with cluster 2 peaks. The modest increase in histone acetylation observed by 24 hours suggests that these enhancers may be activated only following the initiation of differentiation, and that they may play a greater role in transcription at the early stages of differentiation. Enhancers marked by H3K4me1 but lacking H3K27ac may therefore be amenable to changes in activity during differentiation, demonstrated specifically by an increase in H3K27ac. Given the differentiation-dependent increase in histone acetylation at a subset of p300 peaks, we confirmed the association of p300 with the myogenic transcription factors through *de novo* motif analysis. A putative E box motif was identified in 26% of p300 peaks, suggesting that p300 is recruited to the genome by MyoD and myogenin to promote muscle-specific gene expression during myoblast differentiation.

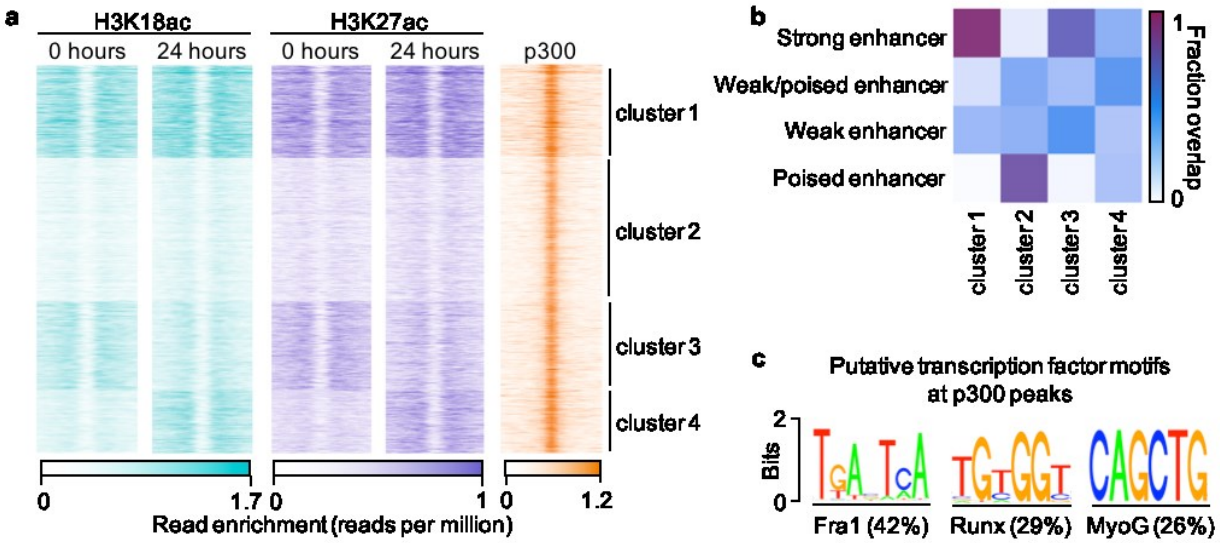

Figure 5. Differentiation promotes the activation of muscle-specific enhancers. **a.** Read enrichment for H3K18ac and H3K27ac prior to (0 hours) and 24 hours following the initiation of differentiation in C2C12 myoblasts. Normalized read enrichment was calculated as reads per million within 1 kb on either side of p300 ChIP-seq peaks. **b.** Association between co-regulated clusters identified in a. and enhancer states identified in Figure 3. The overlap between enhancer states and p300 clusters was calculated as the fraction of p300 peaks in the cluster that fell within 1 kb of an enhancer for a given enhancer-cluster pair. **c.** Putative transcription factor motifs identified at p300 peaks. Significantly enriched sequence motifs in 200 bp regions at p300 peak centres were identified *de novo* and annotated as putative transcription factor binding sites.

Bexarotene promotes a characteristic myogenic transcriptional program

We have previously shown that the RXR agonist bexarotene enhances the rates of differentiation and cell fusion in differentiating myoblasts. We therefore wished to characterize the global changes in gene expression occurring during the differentiation of bexarotene-treated myoblasts using RNA-seq, and further explore the molecular regulation of target gene transcription. Using a 1.5-fold threshold for differential expression, we identified 533 genes whose expression was affected at 12 and/or 24 hours of differentiation by bexarotene (Figure 6a). Hierarchical clustering on standardized values of expression revealed distinct groups of genes that appeared to be coregulated during differentiation and that showed a similar response to bexarotene. This included a large cluster of genes that appeared to be progressively downregulated during differentiation and in response to treatment with bexarotene, as well as a smaller cluster of genes that were progressively upregulated. To better assess the similarities in gene expression across conditions, we calculated the Spearman correlation coefficient between each condition using the untransformed values of expression for the identified bexarotene-responsive genes (Figure 6b). Gene expression in bexarotene-treated and control conditions from the same time point showed the highest correlation, and in particular between their expression at 12 hours of differentiation. Supporting our initial observations, the correlation between gene expression in proliferating myoblasts and myoblasts undergoing differentiation decreased over time and with bexarotene treatment, suggesting that bexarotene may globally promote the natural regulation of gene expression occurring during early myogenic differentiation.

By 24 hours of differentiation, bexarotene had upregulated the expression of many genes involved in muscle cell differentiation and muscle function, including contractile proteins and proteins involved in calcium signaling (Figure 6c-d). Bexarotene also further downregulated the expression of genes involved in the cell cycle, including genes involved in DNA packaging and cell division.

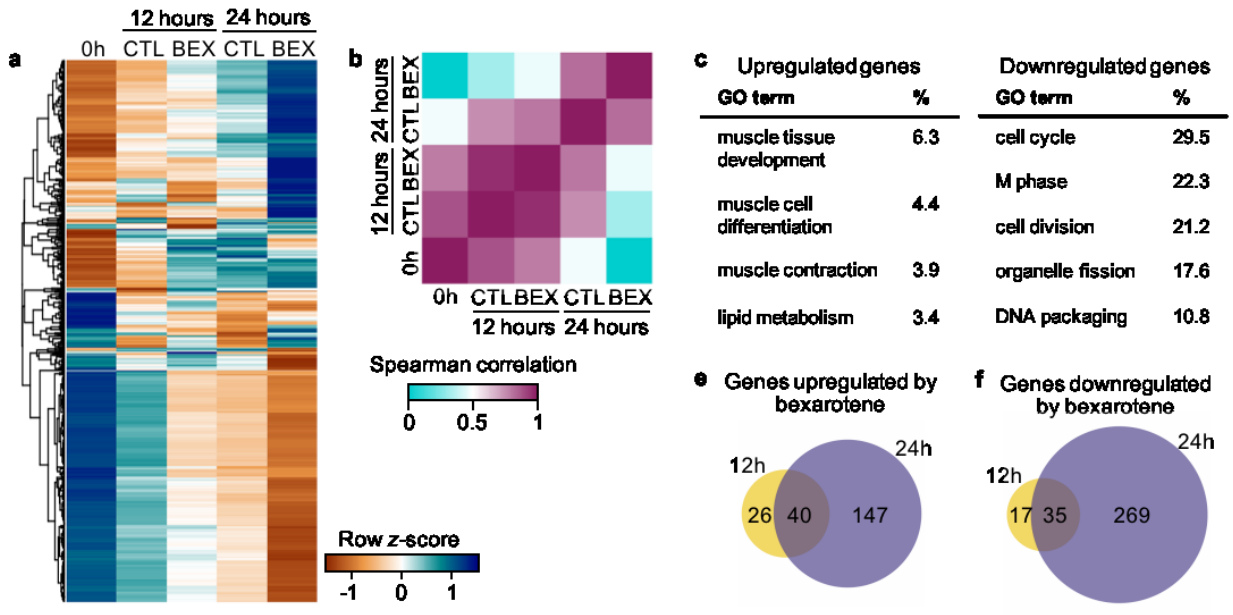

Figure 6. Bexarotene promotes muscle-specific gene expression. **a.** Hierarchical clustering of bexarotene-responsive genes on standardized values of gene expression in proliferating myoblasts (0h) and at 12 and 24 hours of differentiation in control (CTL) and bexarotene-treated (BEX) myoblasts. FPKM values were $\log_2$-transformed and scaled to the mean and standard deviation of the row to give row-based *z*-scores. **b.** Matrix of Spearman correlation coefficients computed between raw FPKM values for bexarotene-regulated genes in each condition described in (a). **c.** Gene ontology terms associated with genes that were upregulated by bexarotene, and similarly for genes that were downregulated by bexarotene (**d.**). **e.** Overlap between genes that were upregulated by bexarotene by 12 hours and between 12 and 24 hours of differentiation, and similarly for genes that were downregulated by bexarotene by 12 hours and between 12 and 24 hours of differentiation (**f.**).

Notably, the response of many of these genes was observed only after 24 hours of differentiation, suggesting that their expression may not be directly regulated by bexarotene (Figure 6e-f). Intriguingly, bexarotene promoted the expression of both MyoD and myogenin by 12 hours of differentiation, while bexarotene had similarly downregulated the expression of cyclin D1, a primary regulator of cell cycle progression (Fu, Wang, Sakamaki & Pestell, 2004). The coordinated regulation of muscle-specific and cell cycle-related gene expression suggested that the global transcriptional response to bexarotene during myogenic differentiation may be mediated by one or more key developmental transcription factors. Together, our results indicate that bexarotene may function upstream of MyoD and myogenin to promote the progression of myoblast differentiation.

MyoD-bound enhancers show differential activity in bexarotene-treated myoblasts

In order to assess the role that MyoD and myogenin may play in the differentiation of bexarotene-treated myoblasts, we were first interested in comparing the transcriptional activity of MyoD in proliferating and differentiating myoblasts. Comparing MyoD binding sites during proliferation and differentiation, we found that approximately half of the MyoD binding sites identified in proliferating myoblasts are maintained at 24 hours of differentiation. As the activity of MyoD is affected by its DNA binding partner, we searched for putative transcription factor motifs using *de novo* motif analysis at loci bound by MyoD in both proliferating and differentiating cells and at MyoD binding sites that were induced only following the initiation of differentiation (Figure 7a). Binding sites that were maintained during myoblast proliferation and differentiation showed an enrichment in E box sequence motifs, as well as motifs for the general transcription factor AP-1. While differentiation-specific binding sites showed a similar enrichment for E box motifs, they also showed enrichment for putative Mef2c binding sites. As Mef2c has been identified as a synergistic activator of muscle-

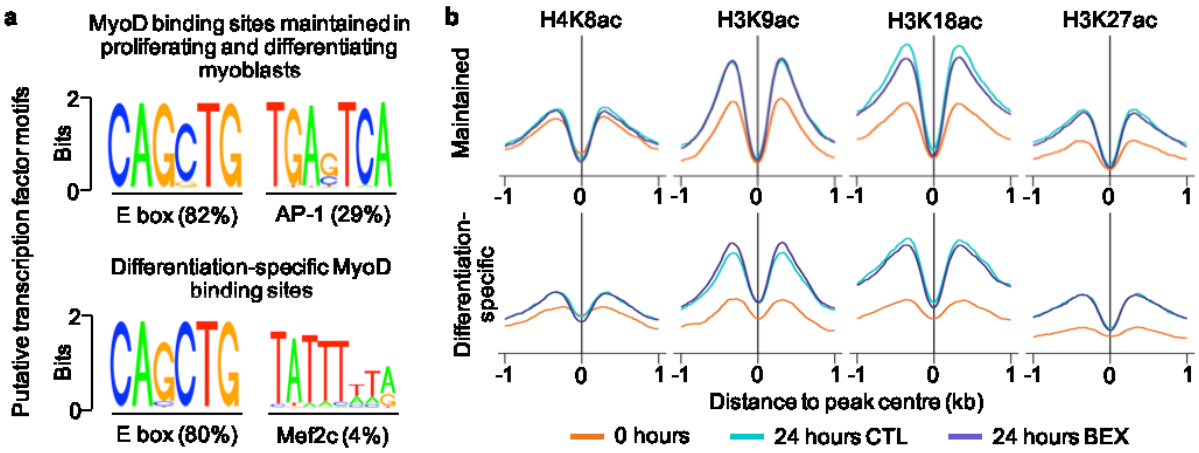

Figure 7. MyoD shows distinct transcriptional activity in differentiating myoblasts. **a.** Putative transcription factor motifs identified at MyoD binding sites that are maintained in both proliferating and differentiating myoblasts (top) and at MyoD binding sites specific to differentiating myoblasts (bottom). **b.** Histone acetylation MyoD binding sites that are maintained in both proliferating and differentiating myoblasts and at MyoD binding sites specific to differentiating myoblasts. Average read enrichment was calculated as the number of reads per million mapped reads in 200 bp bins across a 2 kb region centered at the MyoD peak. 0 hours: proliferating myoblasts; 24 hours CTL: 24 hours of differentiation in control myoblasts; 24 hours BEX: 24 hours of differentiation in bexarotene-treated myoblasts.

specific gene expression with MyoD, it is involved in differentiation-specific binding of MyoD to the regulatory elements of target genes.

To determine if the two classes of MyoD binding sites in differentiating myoblasts showed differences in their potential to activate transcription of muscle-specific genes, we compared enrichment of histone acetylation at MyoD binding sites that were maintained in proliferating and differentiating cells and those that were specific to the early stages of differentiation. As we previously found that a subset of enhancers showed a differentiation-dependent increase in histone acetylation, we computed the enrichment of histone acetylation at MyoD binding sites specifically associated with weak/poised enhancers. As observed for p300, MyoD binding sites found at weak/poised enhancers showed a differentiation-dependent increase in the enrichment of H3K18ac and H3K27ac, regardless of their classification (Figure 7b). The differentiation-dependent increase in histone acetylation was not limited to p300-associated marks however, demonstrating that MyoD associates with other coactivators with HAT activity during differentiation, including H3K9ac-associated PCAF. Interestingly, we observed a noticeable increase in H3K9ac in bexarotene-treated myoblasts only at differentiation-specific MyoD binding sites. This suggested that the potential of MyoD to activate transcription may increase in bexarotene-treated cells, and may therefore lead to increased activation of muscle-specific gene expression.

At the early stages of differentiation, the transcription of genes involved in myoblast differentiation is activated following the binding of MyoD and myogenin to E box sequences in the *cis*-regulatory regions of their targets. As an increase in the expression of myogenin was observed in bexarotene-treated myoblasts by 12 hours of differentiation, similar to the expression of MyoD, we hypothesized that myogenin may also play a role in the transcriptional response to bexarotene during myogenic differentiation. In order to assess the possibility of both MyoD and myogenin

mediating the activation of muscle-specific gene expression in bexarotene-treated myoblasts, we first compared the genome-wide localization of MyoD and myogenin in differentiating myoblasts. Although a similar number of binding sites were identified for MyoD and myogenin at 24 hours of differentiation (20,160 and 24,360, respectively), we found that the two factors largely differed in their recruitment to genome (Figure 8a). MyoD and myogenin ChIP-seq peaks were ranked by read enrichment and grouped into bins of increasing size ranging from 2,000 to 20,000 peaks, and the overlap in binding sites was calculated for each pair of bins. Regardless of transcription factor enrichment, approximately 45% of all MyoD and myogenin binding sites were co-localized across the genome, suggesting that MyoD and myogenin largely regulate different sets of genes at this stage of differentiation.

To further explore the relationship between MyoD and myogenin during differentiation and particularly in bexarotene-treated myoblasts, we grouped MyoD peaks based on their co-occupancy with myogenin and computed the normalized enrichment of histone acetylation in regions bound only by MyoD or bound by both MyoD and myogenin and specifically associated with poised enhancers (Figure 8b). Supporting our initial finding that MyoD DNA binding in differentiating myoblasts is associated with a local increase in histone acetylation, loci bound by MyoD alone or co-occupied by MyoD and myogenin showed a similar increase in enrichment in differentiating myoblasts compared to proliferating cells. However, the two classes of binding sites differed in their response to bexarotene, as MyoD binding sites co-occupied by myogenin showed a more pronounced increase in histone acetylation in bexarotene-treated cells compared to those bound only by MyoD. In addition to H3K9ac, shared MyoD and myogenin binding sites also showed a noticeable increase in each of H4K8ac, H3K18ac and H3K27ac, suggesting that myogenin may additively affect the bexarotene-dependent increase in histone acetylation at MyoD-bound loci.

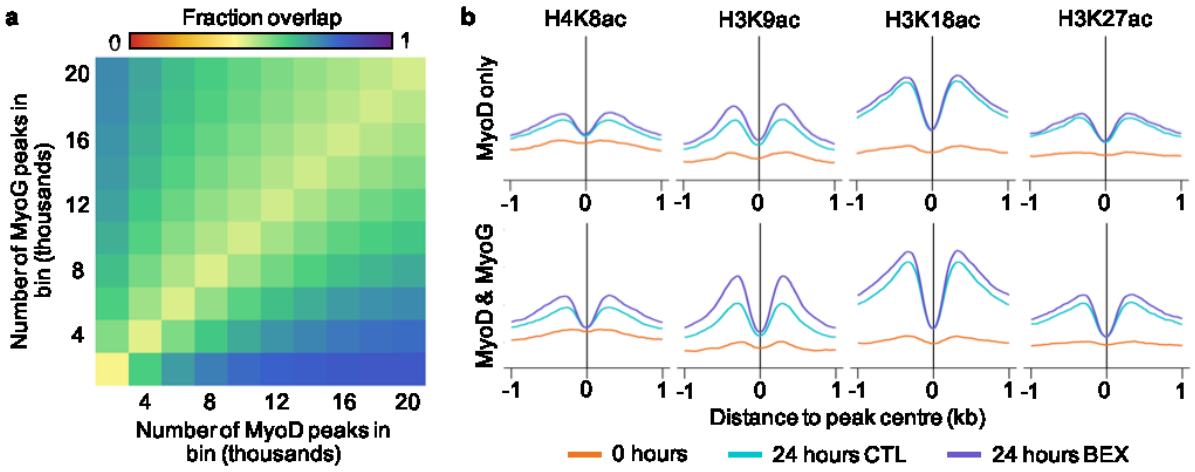

Figure 8. MyoD and myogenin differentially regulate transcription during early myogenic differentiation. **a.** Overlap between binding sites identified for MyoD and myogenin (MyoG) at 24 hours of myoblast differentiation. MyoD and MyoG peaks were ranked by read enrichment, and the highest ranking 20,000 peaks for each factor were grouped into bins in increments of 2,000. Shared binding sites were identified if a MyoD and MyoG peak fell within 100 bp, and the fraction overlap for each pair of bins was calculated as the number of peaks overlapping between the bins divided by the number of peaks in the smallest bin. **b.** Histone acetylation at poised enhancers bound by MyoD only or by MyoD and MyoG at 24 hours of differentiation. Read enrichment was computed as the number of reads per million mapped reads within 1 kb of the MyoD or MyoD/MyoG peak centre. 0 hours: proliferating myoblasts; 24 hours CTL: 24 hours of differentiation in control myoblasts; 24 hours BEX: 24 hours of differentiation in bexarotene-treated myoblasts.

Bexarotene-mediated transcription in differentiating myoblasts occurs primarily through RXR-independent regulation

As RXR was shown to mediate the effect of bexarotene during myogenic differentiation, we reasoned that bexarotene may elicit a direct transcriptional response through RXR found at the regulatory regions of target genes. To investigate the molecular regulation of bexarotene-responsive gene expression, we performed ChIP-seq for RXR in C2C12 myoblasts. A number of RXR binding sites were identified, including a p300-associated binding site in an intronic region of the gene encoding angiopoietin-like 4 (Angptl4), a characterized target of PPAR signaling and whose expression was upregulated by bexarotene (Figure 9a) (Yoon et al., 2000). Intriguingly, we found that bexarotene-treated myoblasts showed a particular increase in H3K9ac at the *Angptl4* locus compared to differentiating controls, suggesting that the increase in expression of Angptl4 may be mediated by an increase in H3K9 acetylation.

As p300 is a common nuclear receptor coactivator, we compared the localization of RXR and p300 in myoblasts and found that 66% of RXR binding sites showed p300 recruitment. To further explore the mechanism of RXR-mediated transcriptional regulation during myoblast differentiation, we computed the enrichment of histone acetylation at RXR binding sites in proliferating and differentiating myoblasts, as well as in differentiating myoblasts treated with bexarotene (Figure 9b). RXR-bound loci showed a differentiation-dependent increase in H3K18ac and modestly in H3K27ac, suggesting that the transcriptional activity of RXR during myogenic differentiation is associated with p300-dependent histone acetylation. However, histone acetylation did not increase in bexarotene-treated myoblasts compared to differentiating controls, indicating that bexarotene may not increase the average transcriptional activity of RXR in differentiating myoblasts and that's effect may depend on the specific gene locus.

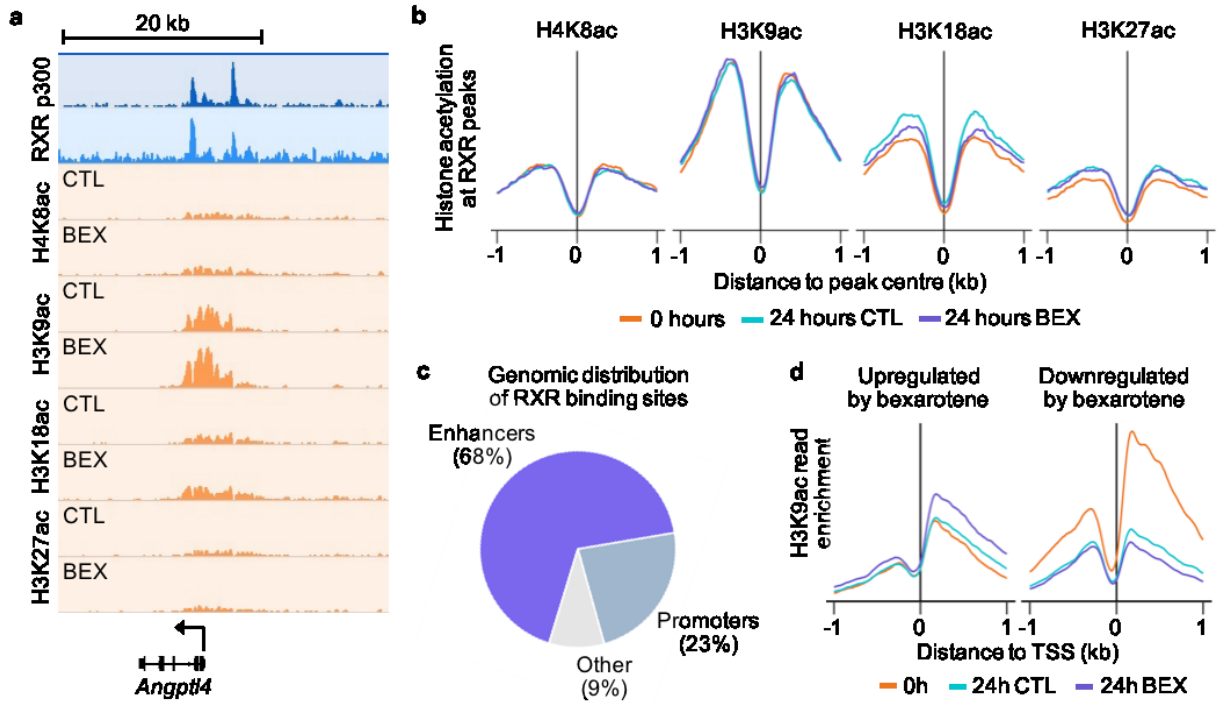

Figure 9. RXR may not directly regulate muscle-specific gene expression in bexarotene-treated myoblasts. **a.** RXR binding, p300 recruitment and histone acetylation at 24 hours of differentiation in control (CTL) and bexarotene-treated (BEX) myoblasts at the *Angptl4* locus. **b.** Average enrichment of histone acetylation in reads per million within 1 kb of RXR ChIP-seq peaks identified in myoblasts. 0 hours: proliferating myoblasts; 24 hours CTL: 24 hours of differentiation in control myoblasts; 24 hours BEX: 24 hours of differentiation in bexarotene-treated myoblasts. **c.** Distribution of RXR binding sites in regulatory DNA elements. **d.** Average enrichment of H3K9ac at the TSS of genes regulated by bexarotene in reads per million. 0 hours: proliferating myoblasts; 24 hours CTL: 24 hours of differentiation in control myoblasts; 24 hours BEX: 24 hours of differentiation in bexarotene-treated myoblasts.

Associating RXR binding sites with regulatory regions identified in our chromatin state analysis, we found that approximately 68% of RXR binding sites were found at enhancers, while only 23% were found at promoters (Figure 9c). This suggested that RXR may play a specific role in enhancer-mediated transcription and suggested further that RXR may bind to regulatory elements in a cell type- or tissue-specific manner. In order to explore whether the change in expression of bexarotene-responsive genes was reflected in local changes in chromatin state at promoters, we computed the enrichment of histone acetylation at the TSS of genes that were upregulated and downregulated by bexarotene (Figure 9d). It was previously found that the enrichment of histone acetylation at promoters decreases during myogenic differentiation, regardless of the regulation of gene expression (Asp et al., 2011). In contrast to these results, we found that the change in enrichment of the promoter-associated mark H3K9ac at bexarotene-responsive gene promoters was associated with the regulation of those genes. Specifically, we found that genes that were upregulated with bexarotene showed an average increase in H3K9ac in bexarotene-treated myoblasts compared to differentiating controls, while genes that were downregulated showed an average decrease in H3K9ac. These results suggest that a common mode of molecular regulation may mediate the change in gene expression observed in bexarotene-treated myoblasts. Given that RXR-bound loci did not show a bexarotene-dependent increase in acetylation as was observed at MyoD binding sites, our results further suggest that the regulation of bexarotene-responsive genes in differentiating myoblasts may not be mediated globally through an increase in the transcriptional activity of RXR but rather through the transcriptional activity of MyoD.

Discussion

In the current study, we explored the molecular regulation of gene expression during early myogenic differentiation. Characterization of the epigenetic landscape in committed myogenic progenitors reflected a tissue-specific distribution of regulatory DNA elements, and the transcriptional activity of muscle-specific enhancers was reflected by local changes in chromatin state observed at the early stages of differentiation. By 24 hours of differentiation, the RXR agonist bexarotene had promoted the upregulation of muscle-specific genes and downregulation of cell cycle regulators, suggesting that it had allowed the coordinated regulation of cellular differentiation and cell cycle exit. Given that the expression of MyoD and myogenin was upregulated in bexarotene-treated myoblasts compared to differentiating controls, these results further suggest that the global transcriptional response to bexarotene may be mediated by one or more key developmental transcription factors. An increase in histone acetylation at MyoD- and myogenin-bound enhancers, but not at regulatory regions bound by RXR, indicates that global transcriptional regulation in bexarotene-enhanced myogenic differentiation may be mediated by the transcriptional activity of the myogenic transcription factors.

In proliferating muscle cell progenitors, gene expression is regulated through tissue-specific enhancers in part by the activity of MyoD. Here we found that MyoD retains approximately half of its binding sites by 24 hours of differentiation, suggesting that the transcriptional targets of MyoD differ in proliferating and differentiating myoblasts. The identification of putative Mef2c DNA binding motifs at differentiation-specific MyoD binding sites indicated that although MyoD continues to regulate transcription throughout myoblast proliferation and differentiation, its potential to activate muscle-specific gene expression during differentiation depends on its association with differentiation-specific transcription factors. We additionally found that MyoD binding sites in differentiating myoblasts show a local increase in histone acetylation after 24

hours of differentiation, demonstrating that differential binding of MyoD at enhancers plays a role in the regulation of enhancer activity during myogenic differentiation.

Intriguingly, a specific increase in acetylation at H3K9 was observed at MyoD binding sites in bexarotene-treated cells compared to differentiating controls. This effect was particularly noticeable at MyoD binding sites that were induced only following the initiation of differentiation, suggesting that the effect of bexarotene on MyoD function is particular to its activity during differentiation. As bexarotene had increased the expression of MyoD by 12 hours of differentiation, the differentiation-dependent increase in H3K9ac may reflect a relative increase in the transcriptional activity of MyoD, resulting from greater MyoD binding at target loci at earlier stages of differentiation. We also noted that MyoD binding sites that were co-occupied by myogenin showed the greatest increase in histone acetylation, suggesting that myogenin additively affects transcriptional activity at a subset of regulatory regions. In this regard, determining how global MyoD and myogenin DNA binding is affected in bexarotene-treated cells may better our understanding of how these factors mediate local changes in transcriptional activity.

The specific increase in H3K9ac at both MyoD binding sites and at the TSS of genes regulated by bexarotene suggests that a particular coactivator may be globally involved in bexarotene-mediated transcriptional regulation. Acetylation at H3K9 has been associated with the coactivator PCAF, whose HAT activity is specifically required for MyoD-dependent activation of muscle-specific gene expression at the early stages of differentiation (Puri et al., 1997). Although p300 is also involved in the regulation of transcription through MyoD, the p300-associated marks H3K18ac and H3K27ac did not show a notable increase in acetylation in bexarotene-treated myoblasts compared to controls. These results suggest that p300 may not be a critical regulator of bexarotene-mediated transcription at this stage or that the HAT activity of p300 does not directly

mediate changes in chromatin state and transcriptional activity at the promoters of bexarotene-responsive genes. Therefore, the association between MyoD and PCAF during the differentiation of bexarotene-treated myoblasts should be explored.

The increase in histone acetylation at differentiation-specific MyoD binding sites suggested that the activity of MyoD-bound enhancers in differentiating myoblasts in particular is affected by bexarotene. This observation indicates that bexarotene may affect transcription, at least in part, upstream of MyoD. Although the effect of bexarotene on myogenic differentiation is dependent on RXR, we did not find that bexarotene allowed a specific increase in histone acetylation at RXR-bound loci, unlike the bexarotene-dependent increase in histone acetylation observed at MyoD binding sites. These results may reflect different molecular mechanisms underlying the mode of function of MyoD and RXR. Although both transcription factors recruit common coactivators such as p300 to the *cis*-regulatory regions of target genes, additional factors may mediate their transcriptional activity, such as their DNA binding partners and their association with other coactivators. The identification of RXR binding sites in the regulatory regions of muscle-specific genes is critical to identifying the primary genetic targets of RXR and to determining the foundation underlying differential transcription in bexarotene-treated myoblasts.

Previous studies exploring the genome-wide localization of nuclear receptors in various biological systems demonstrated that RXR binding sites show cell type and tissue specificity (Boergesen et al., 2012). Here we found that RXR was bound primarily to enhancers, thus showing a distribution of binding sites pertaining to a role in muscle-specific transcription. A number of other factors may also affect the role of RXR in target gene transcription however, including the stage of development and the partner nuclear receptor. For example, genome-wide profiling of RXR binding sites during adipocyte differentiation showed that the number of RXR binding sites may vary throughout differentiation (Nielsen et al., 2008). In addition, the association of RXR with

specific partners may also change during differentiation, suggesting that differential recruitment of RXR to regulatory regions of target genes and its potential to be activated by receptor-specific agonists depends on developmental stage. In particular, an increase in LXR expression during osteoblast differentiation, and resulting increase in binding sites, leads to a greater association of RXR with LXR (Menendez-Gutierrez et al., 2015). Therefore, an interesting consideration in identifying putative target genes of RXR is the cell type-specific expression of other nuclear receptors that may affect the location of RXR binding.

Clinically, bexarotene is used as a treatment for CTCL, a rare form of lymphoma characterized by abnormal T cell proliferation. An adverse effect of bexarotene is hypertriglyceridemia, which occurs in approximately 80% of patients but can be reduced when bexarotene is combined with other CTCL therapies or lipid-lowering agents such as statins. In support of previous studies, we identified several genes involved in fatty acid metabolism that were upregulated by bexarotene. The sterol regulatory element binding transcription factor 1 (Srebf1), for example, is a transcription factor that binds sterol regulatory element 1 sites in the regulatory regions of genes involved in sterol biosynthesis. Lipid homeostasis in the liver is regulated by many genes targeted by LXR and PPAR, which both form permissive heterodimers with RXR, and correspondingly, LXR response elements in the promoter of Srebf1 indicate that bexarotene may increase the expression of this gene through additional activation of LXR-RXR dimers (Wagner et al., 2003). Interestingly, Abca1 and Abcg1, two genes involved in cellular phospholipid transport, were not expressed in differentiating myoblasts in the absence of bexarotene but were upregulated in bexarotene-treated myoblasts, suggesting that bexarotene was sufficient to activate their expression.

We identified a RXR binding site in intron 3 of the gene encoding Angptl4, a characterized target of PPAR signaling involved in the regulation of angiogenesis and lipid metabolism. Interestingly,

injection of Angptl4 in mice promoted hypertriglyceridemia, suggesting that the adverse effect observed in CTCL patients treated with bexarotene may be partly mediated by increased expression of Angptl4 (Yoshida, Shimizugawa, Ono & Furukawa, 2002). In myotubes, the expression of Angptl4 is induced by long chain fatty acids specifically through PPARδ, resulting in elevated plasma levels of nonesterified fatty acids (Staiger et al., 2009). PPARδ is the predominant PPAR isoform in skeletal muscle, and overexpression of activated PPARδ in mice demonstrated that these gained less weight than wildtype mice (Wang et al., 2004). Bexarotene has also previously been shown to increase the expression of Angptl4 in myotubes through PPARδ-RXR heterodimers, suggesting that the change in gene expression in bexarotene-treated myoblasts may reflect both direct and indirect regulation through RXR activation (Robciuc et al., 2012).

Ligands targeting nuclear receptors have been used as therapeutics for a number of diseases. The ER antagonist tamoxifen, for example, has been widely used as an adjuvant therapy for estrogen receptor-positive (ER+) forms of breast cancer, and when used as a preventative therapeutic, tamoxifen reduces the occurrence of ER+ breast cancer in high risk patients by 49% (Fisher et al., 1998). Tamoxifen belongs to a class of drugs known as selective estrogen receptor modulators that block the effect of estrogen signaling by binding to ER and preventing its activation. More recent studies have found that agents called aromatase inhibitors are more effective treatments than tamoxifen in post-menopausal women with ER+ breast cancer, as these agents lower estrogen production by blocking the aromatase enzyme involved in the conversion of androgens into estrogen. However, the targeted inhibition of ER signaling in ER+ breast cancer patients was none the less an important achievement in the application of nuclear receptor signaling in the direct modulation of transcription.

Similar to the use of tamoxifen in the regulation of class I nuclear receptor signaling, thiazolidinediones and retinoids are agonists for PPAR and RAR, respectively, and have been used clinically in the treatment of disease. Thiazolidinediones have been used in the treatment of type II diabetes mellitus, as many PPAR target genes are involved in adipogenesis and glucose and lipid metabolism. Correspondingly, thiazolidinediones have shown to lower blood glucose levels and improve insulin sensitivity in diabetic patients (Diamant & Heine, 2003). Various thiazolidinediones exist and show different specificities PPAR subtypes, and as a result have different effects on lipids. Therefore, the nuclear receptor subtype is an important factor in determining the effect of agonists on nuclear receptor-regulated transcription. Retinoids were originally shown to induce terminal differentiation in leukemic cell cultures, and all trans-retinoic acid (ATRA) specifically has been used in combination with chemotherapy in the treatment of acute promyelocytic leukemia (APL), a subtype of acute myeloid leukemia characterized by an abnormal accumulation of immature granulocytes (Degos & Wang, 2001). The genetic basis of APL is a chromosomal translocation resulting in the fusion of the promyelocytic leukemia gene (PML) with the gene for RARα, resulting in a chimeric protein containing the N-terminus of the PML protein and the DNA and ligand binding domains of RARα. Unliganded PML-RARα binds corepressor complexes more tightly than unliganded RARα, such that physiological concentrations of ATRA are not sufficient to induce differentiation in leukemic cells. Clinical doses of ATRA however are sufficient to activate target gene transcription and induce differentiation, leading to 90% complete remission in APL patients (Puccetti & Ruthardt, 2004). Thus, an important consideration in developing therapeutics targeting nuclear receptor signaling is the specific context under which the nuclear receptor of interest exists. For RXR, this may include the associated coactivator or corepressor, the DNA binding partner and the specific genetic locus, demonstrating that achieving a specific cellular effect, such as an increase in the rate of cellular

differentiation, may be difficult without also invoking additional and unwarranted phenotypic effects.

In the context of health research, a better understanding of the mechanisms underlying the function of regulatory DNA elements in a specific cell or tissue type may ultimately allow treatments for diseased tissue to be designed to more accurately target the cell's own transcriptional network. In skeletal muscle, a variety of tissue-related diseases are characterized by muscle wasting, including cerebral palsy, inflammatory myopathies and muscular dystrophies. An interesting example of nuclear receptors in skeletal muscle disease relates to the muscle fiber type, as type II fibers are more prone to damage in patients with Duchenne muscular dystrophy (Webster, Silberstein, Hays & Blau, 1998). An increase in the number of type I myofibers found in slow skeletal muscle is observed with activation of PPARδ, suggesting that RXR agonists may also promote the formation of specific fiber types. A future application of the current study will therefore be to assess if bexarotene is able to similarly regulate transcription in differentiating myoblasts to promote their differentiation in a specific physiological context or disease model. An important pathway to explore will be the direct activation of gene expression through RXR, and which of its primary genetic targets are involved in the regulation of muscle-specific gene expression during myogenic differentiation. Although the promiscuity of RXR results in it being required for the regulation of genes involved in a variety of cellular functions, we demonstrate here that specific targeting of RXR signaling promotes the normal regulation of transcription occurring during myoblast differentiation, and therefore describe a novel role for RXR signaling in skeletal myogenesis.

References

AlSudais, H., Aabed, K., Nicola, W., Dixon, K., Chen, J., & Li, Q. (2016). Retinoid X receptor-selective signaling in the regulation of Akt/Protein kinase B isoform-specific expression. *The Journal of Biological Chemistry, 291*(6), 3090-3099.

Arents, G., & Moudrianakis, E. N. (1995). The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National Academy of Sciences of the United States of America, 92*(24), 11170-11174.

Asp, P., Blum, R., Vethantham, V., Parisi, F., Micsinai, M., Cheng, J., et al. (2011). Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proceedings of the National Academy of Sciences of the United States of America, 108*(22), E149-58.

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature, 447*(7143), 407-412.

Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B., & Celniker, S. E. (2004). Computational identification of developmental enhancers: Conservation and function of transcription factor binding-site clusters in drosophila melanogaster and drosophila pseudoobscura. *Genome Biology, 5*(9), R61. doi:10.1186/gb-2004-5-9-r61 [doi]

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*(2), 315-326.

Blow, M. J., McCulley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2010). ChIP-seq identification of weakly conserved heart enhancers. *Nature Genetics, 42*(9), 806-810.

Boergesen, M., Pedersen, T. A., Gross, B., van Heeringen, S. J., Hagenbeek, D., Bindesboll, C., et al. (2012). Genome-wide profiling of liver X receptor, retinoid X receptor, and peroxisome proliferator-activated receptor alpha in mouse liver reveals extensive sharing of binding sites. *Molecular and Cellular Biology, 32*(4), 852-867.

Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E., & Arnold, H. H. (1989). A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *The EMBO Journal, 8*(3), 701-709.

Braun, T., Rudnicki, M. A., Arnold, H. H., & Jaenisch, R. (1992). Targeted inactivation of the muscle regulatory gene myf-5 results in abnormal rib development and perinatal death. *Cell, 71*(3), 369-382.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell, 84*(6), 843-851.

Brunelli, S., Relaix, F., Baesso, S., Buckingham, M., & Cossu, G. (2007). Beta catenin-independent activation of MyoD in presomitic mesoderm requires PKC and depends on Pax3 transcriptional activity.*Developmental Biology, 304*(2), 604-614.

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K. et al. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell, 123*(4), 581-592.

Conerly, M. L., Yao, Z., Zhong, J. W., Groudine, M., & Tapscott, S. J. (2016). Distinct activities of Myf5 and MyoD indicate separate roles in skeletal muscle lineage specification and differentiation.*Developmental Cell, 36*(4), 375-385.

Cossu, G., & Borello, U. (1999). Wnt signaling and the activation of myogenesis in mammals. *The EMBO Journal, 18*(24), 6867-6872.

Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell, 51*(6), 987-1000.

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology, 6*(12),

Degos, L., & Wang, Z. Y. (2001). All trans retinoic acid in acute promyelocytic leukemia. *Oncogene, 20*(49), 7140-7145.

Diamant, M., & Heine, R. J. (2003). Thiazolidinediones in type 2 diabetes mellitus: Current clinical evidence. *Drugs, 63*(13), 1373-1405.

Duvic, M., Hymes, K., Heald, P., Breneman, D., Martin, A. G., Myskowski, P., et al. (2001). Bexarotene is effective and safe for treatment of refractory advanced-stage cutaneous T-cell lymphoma: Multinational phase II-III trial results. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 19*(9), 2456-2471.

Edmondson, D. G., & Olson, E. N. (1989). A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. *Genes & Development, 3*(5), 628-640.

Edmondson, D. G., Cheng, T. C., Cserjesi, P., Chakraborty, T., & Olson, E. N. (1992). Analysis of the myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2. *Molecular and Cellular Biology, 12*(9), 3665-3677.

Eickbush, T. H., & Moudrianakis, E. N. (1978). The histone core complex: An octamer assembled by two sets of protein-protein interactions. *Biochemistry, 17*(23), 4955-4964.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,473*(7345), 43-49.

Ernst, J., & Kellis, M. (2012). ChromHMM: Automating chromatin-state discovery and characterization. *Nature Methods, 9*(3), 215-216.

Fisher, B., Costantino, J. P., Wickerham, D. L., Redmond, C. K., Kavanah, M., Cronin, W. M., et al. (1998). Tamoxifen for prevention of breast cancer: Report of the national surgical adjuvant breast and bowel project P-1 study. *Journal of the National Cancer Institute, 90*(18), 1371-1388.

Fu, M., Wang, C., Li, Z., Sakamaki, T., & Pestell, R. G. (2004). Minireview: Cyclin D1: Normal and abnormal functions. *Endocrinology, 145*(12), 5439-5447.

Germain, P., Staels, B., Dacquet, C., Spedding, M., & Laudet, V. (2006). Overview of nomenclature of nuclear receptors. *Pharmacological Reviews, 58*(4), 685-704.

Goodman, R. H., & Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. *Genes & Development, 14*(13), 1553-1577.

Goulding, M., Lumsden, A., & Paquette, A. J. (1994). Regulation of pax-3 expression in the dermomyotome and its role in muscle development. *Development (Cambridge, England), 120*(4), 957-971.

Grant, P. A., Duggan, L., Cote, J., Roberts, S. M., Brownell, J. E., Candau, R., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: Characterization of an ada complex and the SAGA (Spt/Ada) complex. *Genes & Development, 11*(13), 1640-1650.

Green, S., & Chambon, P. (1988). Nuclear receptors enhance our understanding of transcription regulation. *Trends in Genetics : TIG, 4*(11), 309-314.

Gronemeyer, H., Gustafsson, J. A., & Laudet, V. (2004). Principles for modulation of the nuclear receptor superfamily. *Nature Reviews.Drug Discovery, 3*(11), 950-964.

Haberland, M., Montgomery, R. L., & Olson, E. N. (2009). The many roles of histone deacetylases in development and physiology: Implications for disease and therapy. *Nature Reviews.Genetics, 10*(1), 32-42.

Hamamori, Y., Wu, H. Y., Sartorelli, V., & Kedes, L. (1997). The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, twist.*Molecular and Cellular Biology, 17*(11), 6563-6573.

Hassan, A. H., Prochasson, P., Neely, K. E., Galasinski, S. C., Chandy, M., Carrozza, M. J., & Workman, J. L. (2002). Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell, 111*(3), 369-379.

Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N., & Klein, W. H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature,364*(6437), 501-506.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,459*(7243), 108-112.

Hon, G. C., Hawkins, R. D., & Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. *Human Molecular Genetics, 18*(R2), R195-201.

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research, 37*(1), 1-13.

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols, 4*(1), 44-57.

Huang, P., Chandra, V., & Rastinejad, F. (2010). Structural overview of the nuclear receptor superfamily: Insights into physiology and therapeutics. *Annual Review of Physiology, 72*, 247-272.

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science (New York, N.Y.), 293*(5532), 1074-1080.

Jin, Q., Yu, L. R., Wang, L., Zhang, Z., Kasper, L. H., Lee, J. E., et al. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *The EMBO Journal, 30*(2), 249-262.

Kassar-Duchossoy, L., Gayraud-Morel, B., Gomes, D., Rocancourt, D., Buckingham, M., Shinin, V., & Tajbakhsh, S. (2004). Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature,431*(7007), 466-471.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell, 128*(4), 693-705.

Kuang, S., Kuroda, K., Le Grand, F., & Rudnicki, M. A. (2007). Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell, 129*(5), 999-1010.

Langlands, K., Yin, X., Anand, G., & Prochownik, E. V. (1997). Differential interactions of id proteins with basic-helix-loop-helix transcription factors. *The Journal of Biological Chemistry, 272*(32), 19785-19793.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., et al. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell, 66*(2), 305-315.

Lee, D. Y., Hayes, J. J., Pruss, D., & Wolffe, A. P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell, 72*(1), 73-84.

Leibham, D., Wong, M. W., Cheng, T. C., Schroeder, S., Weil, P. A., Olson, E. N., & Perry, M. (1994). Binding of TFIID and MEF2 to the TATA element activates transcription of the xenopus MyoDa promoter.*Molecular and Cellular Biology, 14*(1), 686-699.

Li, J., Moazed, D., & Gygi, S. P. (2002). Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *The Journal of Biological Chemistry, 277*(51), 49383-49388.

Lomberk, G., Bensi, D., Fernandez-Zapico, M. E., & Urrutia, R. (2006). Evidence for the existence of an HP1-mediated subcode within the histone code. *Nature Cell Biology, 8*(4), 407-415.

Maison, C., & Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nature Reviews.Molecular Cell Biology, 5*(4), 296-304.

McKenna, N. J., & O'Malley, B. W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell, 108*(4), 465-474.

Mejat, A., Ramond, F., Bassel-Duby, R., Khochbin, S., Olson, E. N., & Schaeffer, L. (2005). Histone deacetylase 9 couples neuronal activity to muscle chromatin acetylation and gene expression. *Nature Neuroscience, 8*(3), 313-321.

Menendez-Gutierrez, M. P., Roszer, T., Fuentes, L., Nunez, V., Escolano, A., Redondo, J. M., et al. (2015). Retinoid X receptors orchestrate osteoclast differentiation and postnatal bone remodeling.*The Journal of Clinical Investigation, 125*(2), 809-823.

Min, J., Zhang, Y., & Xu, R. M. (2003). Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at lys 27. *Genes & Development, 17*(15), 1823-1828.

Molkentin, J. D., Black, B. L., Martin, J. F., & Olson, E. N. (1995). Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. *Cell, 83*(7), 1125-1136.

Montgomery, R. L., Davis, C. A., Potthoff, M. J., Haberland, M., Fielitz, J., Qi, X., et al. (2007). Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility.*Genes & Development, 21*(14), 1790-1802.

Munsterberg, A. E., Kitajewski, J., Bumcrot, D. A., McMahon, A. P., & Lassar, A. B. (1995). Combinatorial signaling by sonic hedgehog and wnt family members induces myogenic bHLH gene expression in the somite. *Genes & Development, 9*(23), 2911-2922.

Naar, A. M., Lemon, B. D., & Tjian, R. (2001). Transcriptional coactivator complexes. *Annual Review of Biochemistry, 70*, 475-501.

Niederreither, K., & Dolle, P. (2008). Retinoic acid in development: Towards an integrated view. *Nature Reviews.Genetics, 9*(7), 541-553.

Nielsen, R., Pedersen, T. A., Hagenbeek, D., Moulos, P., Siersbaek, R., Megens, E., et al. (2008). Genome-wide profiling of PPARgamma:RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes & Development, 22*(21), 2953-2967.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P., & Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes & Development, 16*(4), 479-489.

Pepke, S., Wold, B., & Mortazavi, A. (2009). Computation for ChIP-seq and RNA-seq studies. *Nature Methods, 6*(11 Suppl), S22-32.

Phillips, J. E., & Corces, V. G. (2009). CTCF: Master weaver of the genome. *Cell, 137*(7), 1194-1211.

Polesskaya, A., Naguibneva, I., Fritsch, L., Duquet, A., Ait-Si-Ali, S., Robin, P., et al. (2001). CBP/p300 and muscle differentiation: No HAT, no muscle. *The EMBO Journal, 20*(23), 6816-6825.

Puccetti, E., & Ruthardt, M. (2004). Acute promyelocytic leukemia: PML/RARalpha and the leukemic stem cell. *Leukemia, 18*(7), 1169-1175.

Puri, P. L., Sartorelli, V., Yang, X. J., Hamamori, Y., Ogryzko, V. V., Howard, B. H., et al. (1997). Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Molecular Cell, 1*(1), 35-45.

Rastinejad, F., Perlmann, T., Evans, R. M., & Sigler, P. B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature, 375*(6528), 203-211.

Reshef, R., Maroto, M., & Lassar, A. B. (1998). Regulation of dorsal somitic cell fates: BMPs and noggin control the timing and pattern of myogenic regulator expression. *Genes & Development, 12*(3), 290-303.

Rhodes, S. J., & Konieczny, S. F. (1989). Identification of MRF4: A new member of the muscle regulatory factor gene family. *Genes & Development, 3*(12B), 2050-2061.

Ridgeway, A. G., Wilton, S., & Skerjanc, I. S. (2000). Myocyte enhancer factor 2C and myogenin up-regulate each other's expression and induce the development of skeletal muscle in P19 cells. *The Journal of Biological Chemistry, 275*(1), 41-46.

Ridgeway, A. G., & Skerjanc, I. S. (2001). Pax3 is essential for skeletal myogenesis and the expression of Six1 and Eya2. *The Journal of Biological Chemistry, 276*(22), 19033-19039.

Robciuc, M. R., Skrobuk, P., Anisimov, A., Olkkonen, V. M., Alitalo, K., Eckel, R. H., et al. (2012). Angiopoietin-like 4 mediates PPAR delta effect on lipoprotein lipase-dependent fatty acid uptake but not on beta-oxidation in myotubes. *PloS One, 7*(10), e46212.

Rudnicki, M. A., Braun, T., Hinuma, S., & Jaenisch, R. (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene myf-5 and results in apparently normal muscle development. *Cell,71*(3), 383-390.

Rudnicki, M. A., Schnegelsberg, P. N., Stead, R. H., Braun, T., Arnold, H. H., & Jaenisch, R. (1993). MyoD or myf-5 is required for the formation of skeletal muscle. *Cell, 75*(7), 1351-1359.

Sartorelli, V., Puri, P. L., Hamamori, Y., Ogryzko, V., Chung, G., Nakatani, Y., et al. (1999). Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program. *Molecular Cell,4*(5), 725-734.

Seale, P., Sabourin, L. A., Girgis-Gabardo, A., Mansouri, A., Gruss, P., & Rudnicki, M. A. (2000). Pax7 is required for the specification of myogenic satellite cells. *Cell, 102*(6), 777-786.

Shen, X., & Gorovsky, M. A. (1996). Linker histone H1 regulates specific gene expression but not global transcription in vivo. *Cell, 86*(3), 475-483.

Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T., & Shi, Y. (2005). Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular Cell, 19*(6), 857-864.

Staiger, H., Haas, C., Machann, J., Werner, R., Weisser, M., Schick, F., et al. (2009). Muscle-derived angiopoietin-like protein 4 is induced by fatty acids via peroxisome proliferator-activated receptor (PPAR)-delta and is of metabolic relevance in humans. *Diabetes, 58*(3), 579-589.

Suganuma, T., Gutierrez, J. L., Li, B., Florens, L., Swanson, S. K., Washburn, M. P., et al. (2008). ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nature Structural & Molecular Biology, 15*(4), 364-372.

Tajbakhsh, S., Rocancourt, D., Cossu, G., & Buckingham, M. (1997). Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and myf-5 act upstream of MyoD. *Cell, 89*(1), 127-138.

Tajbakhsh, S., Borello, U., Vivarelli, E., Kelly, R., Papkoff, J., Duprez, D., et al. (1998). Differential activation of Myf5 and MyoD by different wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. *Development (Cambridge, England), 125*(21), 4155-4162.

Vakoc, C. R., Mandat, S. A., Olenchock, B. A., & Blobel, G. A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Molecular Cell, 19*(3), 381-391.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature, 457*(7231), 854-858.

Wagner, B. L., Valledor, A. F., Shao, G., Daige, C. L., Bischoff, E. D., Petrowski, M., et al. (2003). Promoter-specific roles for liver X receptor/corepressor complexes in the regulation of ABCA1 and SREBP1 gene expression. *Molecular and Cellular Biology, 23*(16), 5780-5789.

Wang, Y. X., Zhang, C. L., Yu, R. T., Cho, H. K., Nelson, M. C., Bayuga-Ocampo, C. R., et al. (2004). Regulation of muscle fiber type and running endurance by PPARdelta. *PLoS Biology, 2*(10), e294.

Webster, C., Silberstein, L., Hays, A. P., & Blau, H. M. (1988). Fast muscle fibers are preferentially affected in duchenne muscular dystrophy. *Cell, 52*(4), 503-513.

Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., et al. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America, 97*(21), 11303-11306.

Yoon, J. C., Chickering, T. W., Rosen, E. D., Dussault, B., Qin, Y., Soukas, A., et al. (2000). Peroxisome proliferator-activated receptor gamma target gene encoding a novel angiopoietin-related protein associated with adipose differentiation. *Molecular and Cellular Biology, 20*(14), 5343-5349.

Yoshida, K., Shimizugawa, T., Ono, M., & Furukawa, H. (2002). Angiopoietin-like protein 4 is a potent hyperlipidemia-inducing factor in mice and inhibitor of lipoprotein lipase. *Journal of Lipid Research,43*(11), 1770-1772.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, *515*(7527), 355-64.

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research, 21*(8), 1273-1283.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biology, 9*(9), R137-2008-9-9-r137.