Investigating the Failure of Myogenesis in the Pediatric Tumor Rhabdomyosarcoma

Kyle L. MacQuarrie

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2011

Stephen Justice Tapscott, Chair
Edith Wang
Peter S. Nelson

Program Authorized to Offer Degree:
Molecular and Cellular Biology Graduate Program

University of Washington

Abstract

Investigating the Failure of Myogenesis in the Pediatric Tumor Rhabdomyosarcoma

Kyle L. MacQuarrie

Chair of the Supervisory Committee:
Professor Stephen Justice Tapscott
Department of Neurology

Rhabdomyosarcoma (RMS) is a pediatric tumor of skeletal muscle that fails to undergo terminal differentiation, even though it expresses the myogenic regulatory factor MyoD, which should be sufficient for that process. We have previously provided evidence for multiple inhibitory transcription factors in the tumors acting to oppose the activity of MyoD. Even so, we have found that it is possible to restore MyoD activity by forcing it to interact with one of its protein dimer partners, which leads to differentiation of the cells, and downregulation of the inhibitors. This work now demonstrates that differentiation of RMS can be achieved by expressing other transcription factors that play positive roles in myogenesis, RUNX1 and RP58, and that all these mechanisms of differentiation result in the increase of a single microRNA, miR-206, that is itself sufficient to differentiate RMS. One of the inhibitory factors we previously found as opposing MyoD, MSC, acts at the promoter of miR-206 to interfere with the MyoD activity necessary for the microRNA's expression. Other analyses of expression and gene regulation suggest the existence of an epistatic relationship between MyoD, RUNX1, RP58, and miR-206, with MyoD positively regulating all the other targets, RUNX1 assisting with the activation of RP58 and miR-206, and miR-206 the target of all the other factors. Genome-wide analysis of DNA binding by MyoD and MSC demonstrates that both factors bind throughout the genome of RMS, with both distinct and overlapping binding. Comparison of MyoD binding in RMS to that of MyoD in primary

human cells reveals differences in the binding sites for possible cooperative factors, including RUNX1, but an overall similarity in the MyoD binding between RMS and human myotubes. Taken as a whole, the data suggests that RMS represent an arrested state of development balanced between myoblast and myotube, and that manipulation of components of the myogenic gene program can 'tip the balance' and restore their ability to differentiate.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The author wishes to thank members of the Tapscott lab, the members of his doctoral committee, as well as members of other groups at the Fred Hutchinson and elsewhere who have offered invaluable assistance and support. This work would not have been possible without the support and guidance of Dr. Tapscott.

DEDICATION

To R.A.M.

Chapter 1: Introduction

This chapter should be considered in the context of the following publications, and Figure 1.1 has been reproduced from the third publication:

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., **MacQuarrie, K.L.**, Davison, J., Morgan, M.T., Ruzzo, W.L., Gentleman, R.C., and Tapscott, S.J. (2010) Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev. Cell *18(4)*, 662-674.

Yang, Z., **MacQuarrie, K.L.**, Analau, E., Tyler, A.E., Dilworth, F.J., Cao, Y., Diede, S.J., and Tapscott, S.J. (2009) MyoD and E-protein heterodimers switch rhabdomyosarcoma cells from an arrested myoblast phase to a differentiated state. Genes Dev. *23(6)*, 694-707.

MacQuarrie, K.L. and Tapscott, S.J. (2011) Stuck in a Balancing Act: Histone Methyltransferase Activity of KMT1A Traps Alveolar Rhabdomyosarcomas in an Undifferentiated State. Cell Cycle, *10(19)*.

Skeletal Muscle Development

Development in the embryo

The process of skeletal muscle development in vertebrates begins in the embryo, and must shepherd cells through processes of lineage commitment, terminal differentiation, and tissue maturation, to make a functional muscle. While some variability has been described in earlier steps of the process, as described below, ultimately skeletal muscle development converges at the level of giving rise to a population of cells that express the myogenic regulatory factors (MRFs). Though few in number, the MRFs are remarkable in their role as the transcription factors that serve to regulate the gene networks that control skeletal muscle fate, structure and function.

The majority of the cells that will become skeletal muscle originate in the somites, structures that lie laterally to the neural tube and notochord and give rise to numerous cell types. Somites produce not only the myogenic cells, but those of dermal, skeletal and cartilaginous lineages as well (Mok and Sweetman, 2011). The majority of the myogenic cells in developed tissues come from the somites, though there are some exceptions. Somites serves as the point of origin for all skeletal muscles of the limbs and the trunk, but certain muscles of the head and neck come from myogenic cells that originate at non-somitic locations. Regardless of the point of origin, all cells that give rise to skeletal muscle are mesodermal in nature. Somites bud off from the pre-somitic mesoderm as mesenchymal cores surrounded by epithelial coverings, and develop sequentially in an anterior-posterior fashion. Somites that are located more anteriorly are therefore more developed than the somites farther in the posterior direction at the same timepoint.

Somites divide into sub-structures over time, with the first dorsal-ventral division resulting in the ventrally located sclerotome, originator of the axial skeleton, and the dorsally located dermomyotome. The dermomyotome, as its name suggests, includes a mixture of both dermal and myogenic precursor cells, and is the first point when myogenic precursors can be detected. Further development results in the delamination of cells from the lips of the

dermomyotome, the migration of those cells ventrally, and establishment of the myotome. The myotome is further divided into the epaxial and hypaxial myotomes, which receive migrating cells from different areas of the dermomyotome and eventually go on to give rise to different muscles (Ordahl and Le Douarin, 1992). The epaxial myotome gives rise to more medial muscles, specifically a subset of the muscles of the back, while the hypaxial myotome goes on to form muscles of the limbs and body wall. It has also been determined that a second wave of migration of cells into the myotome occurs, resulting in the creation of the satellite cell population, which serve as the pool of muscle stem-like cells in the adult (Gros et al., 2005; Lepper and Fan, 2010; Relaix et al., 2005).

Though complex in the details, and showing variability in effect between epaxial and hypaxial cells, numerous signaling pathways impact on the process of specifying cells to become skeletal muscle. Bone morphogenetic protein (BMP) signaling has been identified as a negative regulator of myogenesis, and BMP signaling in the dorsal somite is inhibited by noggin (Hirsinger et al., 1997; Marcelle et al., 1997; Reshef et al., 1998). The Notch pathway has also been shown to be an inhibitor of myogenesis, both *in vivo* and *in vitro* (Kopan et al., 1994). Components of the Wnt pathway have been shown to have a positive role in the acquisition of myogenesis, with *Wnts 1*, *3a*, and *4* being identified as specific positive regulators (Wagner et al., 2000), and the diffusible Wnt receptors *Frzb* and *Sfrp2* possibly titrating and modulating their effects (Ladher et al., 2000). Similarly, sonic hedgehog (Shh) signaling has been implicated in having a positive myogenic role, but myogenesis in Shh-null mice is specifically compromised in the epaxial myotome, suggesting the possibility of a more region-specific effect (Borycki et al., 1999).

Transcriptional regulation of skeletal muscle development and the myogenic regulatory factors

While the myogenic cells of non-somitic origin can be specified by factors such as *Pitx2* (Dong et al., 2006; Shih et al., 2007), the myogenic precursor cells that arise from the somite are characterized by the expression of the paired box transcription factors *Pax3* and *Pax7* (Gros et al., 2005; Relaix et al., 2005). Mice that lack PAX3 do not develop limb muscles (Goulding et al., 1994), and the myogenic defects are even more severe in

Pax3/Pax7 double knock-out mice (Relaix et al., 2005). The *Pax* genes have been shown to play a role in the survival of myogenic cells before they undergo differentiation (Collins et al., 2009), and are upstream, controlling factors in the expression of the myogenic regulatory factors (Bajard et al., 2006; Maroto et al., 1997; Sato et al., 2010). *Pax*-expressing cells from the somites also serve as an important source of myogenic precursors in developed muscles (Gros et al., 2005; Schienda et al., 2006) further demonstrating the crucial role that *Pax* genes play in the development of skeletal muscle.

The MRFs consist of four related basic helix-loop-helix (bHLH) Class II transcription factors that are capable of homo- or heterodimerization through their HLH domain and DNA binding through their basic region. *MyoD* was the first MRF identified, discovered through a cDNA subtractive screen (Davis et al., 1987), and the other three MRFs (*Myf5*, *Myog*, and *MRF4*) were identified soon thereafter (Braun et al., 1990; Braun et al., 1989; Miner and Wold, 1990; Rhodes and Konieczny, 1989; Wright et al., 1989). Extensive work both *in vitro* and *in vivo* has identified overlapping but distinct activities and roles for the four factors. *Myf5* and *MyoD* are known to control the process of commitment to the myogenic lineage, *Myog* is a key regulator of terminal differentiation, and *MRF4* exhibits a complicated role that is involved both in the earlier commitment functions and in differentiation.

All skeletal muscle cells are characterized by the expression of the MRFs, regardless of their location of origin and dependency, or lack thereof, on the *Pax* genes. The timing of expression of the MRFs in mice relates to their described roles and activities – *Myf5* is expressed and present before *MyoD*, with both expressed prior to detectable *Myog* expression. *MRF4* expression shows a biphasic pattern, with expression in the mouse embryo first detectable shortly after the onset of *Myf5* expression (embryonic day 9.0), followed by a decrease 2.5 days later and then an increase again at day 16.0 (Bober et al., 1991; Hinterberger et al., 1991).

Individual deletions of the *Myf5* and *MyoD* genes result in relatively normal appearance of muscle in adult mice, suggesting considerable redundancy between the two factors. However, there are observable defects in *Myf5* and *MyoD* knockout myogenic cells

of epaxial and hypaxial origin, respectively, demonstrating the ability of MRFs to substitute for each other to a certain extent, while still possessing unique roles (Kablar et al., 1998). The phenotype of *Myf5/MyoD* double knock-out mice is one of a complete lack of myogenic cells, demonstrating the necessity of these MRFs for the process of determination of the myogenic lineage (Rudnicki et al., 1993). Detailed study of the role of *Myf5* has been complicated by the fact that it possesses regulatory elements dispersed across more than 100 kb of DNA upstream of its transcription start site that is interspersed with the regulatory elements of both *MRF4* and another gene, but its role in commitment and determination is clear (Carvajal et al., 2001; Olson et al., 1996). Both *MyoD* and *Myf5* have been described as being nodal points in the process of myogenesis, integrating multiple signals to result in a decision about the mutually exclusive processes of proliferation versus differentiation (Weintraub et al., 1991).

Mice that lack *Myog* exhibit cells that are committed to the myogenic lineage, but an absence of differentiated cells. This phenotype demonstrates the necessity of *Myog* for the process of terminal differentiation, a role that agrees with its expression pattern – later in time compared to *MyoD* and *Myf5* (Hasty et al., 1993; Nabeshima et al., 1993). As with *MyoD* and *Myf5*, *Myog* also possesses functions that cannot be replicated by the other MRFs; *MyoD* is incapable of substituting for *Myog* in the process of generating differentiated myogenic cells from murine embryonic stem cells (Myer et al., 2001). Conversely, *Myog* cannot completely substitute for the role of the earlier MRFs; expression of *Myog* under the control of the *Myf5* regulatory elements in *Myf5/MyoD* double knock-out mice was not able to completely rescue the double knock-out phenotype (Wang and Jaenisch, 1997).

The details of the role of *MRF4* in differentiation and commitment are significantly less clear. Its expression pattern in mouse would suggest roles in both early and later muscle processes, but the complicated nature of its regulatory elements has made detailed understanding of its role difficult. From experiments that have shown an ability of MRF4 to compensate for *Myog* in murine ES cells (Sumariwalla and Klein, 2001) as well as the presence of skeletal muscle in *MyoD/Myf5* double-null mice that have preserved MRF4 function (Kassar-Duchossoy et al., 2004), it can be concluded that *MRF4* does have roles in

both the processes of determination and differentiation, though its role in the context of the other MRFs is still unclear.

MyoD and the control of myogenesis

Background on MyoD function in myogenesis

MyoD was the first myogenic regulatory factor to be identified and for reasons ranging from its simpler promoter structure compared to *Myf5* and *MRF4*, to its role as a direct regulator of *Myog* expression, to its ability to drive the entire process of myogenic terminal differentiation, it continues to serve as the exemplar of the MRFs. *MyoD* has been termed a 'master regulator' for its ability to turn cells of a non-myogenic origin into differentiated myotubes. In the nearly two decades since its identification, much has been determined of the molecular mechanisms of both how MyoD affects both cells on a global level and how it affects individual gene targets.

Heterodimers of MyoD and one of the E-proteins (E2A, HEB, E2-2) are understood to be the functional form of MyoD in a cell. While the relative abundance of such heterodimers and their preference for specific sequence contexts are still unclear, bHLH proteins are known to bind the sequences termed 'E-boxes' (CANNTG). MyoD:E heterodimers function as transactivators, meaning that after binding to DNA, they lead to the increased expression of their gene targets (Lassar et al., 1991). Experiments have provided evidence that the presence and occupancy of multiple E-boxes at a given regulatory area results in greater stability of binding and enhanced target activation (Gilmour et al., 1991; Piette et al., 1990; Weintraub et al., 1990; Wentworth et al., 1991).

The introduction of MyoD into a cell results in differential expression of hundreds of genes in distinct temporal clusters, some increasing in expression, and others decreasing, suggesting the action of indirect mechanisms (Bergstrom et al., 2002). The process of myogenesis requires temporal control of target expression, as cells are first committed but continue to proliferate, and then later shift to differentiation and cell-cycle withdrawal. While organogenesis models in simpler systems and organisms, such as the *C. elegans* pharynx, have shown evidence of temporal regulation through the relative affinity of factors

for specific DNA sequences (Gaudet and Mango, 2002), *MyoD* has been shown to control timing through a complex feed-forward regulation of its targets. After induction of its earlier gene targets, such as p38 MAPK or the *Mef2* protein family members, those targets cooperate with *MyoD* at the regulatory elements of later genes to activate them, ensuring that, even though MyoD can bind widely throughout the genome, its targets are activated in a regulated fashion (Penn et al., 2004).

A more recently described role for *MyoD* in affecting myogenesis has been that of microRNA (miRNA) expression. miRNAs are small (~21 nt) non-coding RNAs that bind to target mRNAs and mediate downregulation of their targets (Ge and Chen, 2011). miRNAs that affect myogenesis, both positively (the mir-1/-206 family), and negatively (the miR-133 family), have been identified and found to be induced by the action of MyoD (Kim et al., 2006; Rao et al., 2006; Rosenberg et al., 2006). While a selection of direct targets have been identified for both types of miRNAs and offers some explanation of how they can impact the process of myogenesis – for instance, miR-206 directly targets the p180 subunit of DNA polymerase alpha, a component of cellular machinery that would clearly be unnecessary in terminally differentiated cells (Kim et al., 2006) – there are likely numerous, as of yet unknown, additional targets.

Proteins can affect MyoD function positively and negatively

Numerous proteins have been implicated as co-factors in some capacity for *MyoD* during the process of myogenesis. Apart from the aforementioned p38 MAPK and Mef proteins, roles have also been described for the Hox protein co-factors *Pbx* and *Meis* (Berkes et al., 2004), the *Six* proteins (Spitz et al., 1998), and the ubiquitous *Sp1* factor (Biesiada et al., 1999). The exact manner by which such factors cooperate with MyoD vary, but both Pbx/Meis and the Six proteins Six1 and Six4 have been shown to impact on the ability of MyoD to activate myogenin. The role of the members of the *Mef2* family has been described more broadly; they appear to act by synergistically cooperating with MyoD at regulatory elements (Molkentin et al., 1995). Adding support to this model is the finding that E-boxes and Mef2 binding sites are closely positioned at many muscle-specific genes (Wasserman and Fickett, 1998).

In addition to being positively regulated by cooperating transcription factors, MyoD activity can be impaired by diverse factors. One of the most classic family of such factors, functional details of which have been known for a few decades, are the Id proteins. There are multiple *Id* genes, and they all share the common characteristic of possessing the HLH domain necessary for factor dimerization and lacking the basic region that confers DNA binding. The model for their function is that they dimerize with factors such as MyoD and/or E-proteins, fail to bind to DNA, and therefore functionally disrupt functional protein heterodimers by titrating away their component parts (Benezra et al., 1990). It has been demonstrated that forcing heterodimerization between MyoD and an E-protein, E47, diminishes the inhibitory action of Ids, offering further support for the idea that the Ids function at the level of dimerization interference (Neuhold and Wold, 1993).

Other protein factors have also been shown to play various inhibitory roles during myogenesis, many of them, though not all, belonging to the bHLH family themselves. The bHLH protein Musculin (MSC), also known as MyoR, is expressed in proliferating myoblasts and decreases during differentiation both *in vitro* and *in vivo*. MSC heterodimerizes with E-proteins and binds to E-boxes, and inhibits the process of myogenesis when co-expressed with MyoD in fibroblasts (Lu et al., 1999). The bHLH *Mist1* operates in a similar fashion, forming Mist1:MyoD heterodimers that lack activating potential, and occupying certain E-boxes as Mist1:Mist1 homodimers, presumably blockading them from functional MyoD occupancy (Lemercier et al., 1998). Dec1 is another bHLH protein that acts by occupying E-boxes as a homodimer and mediating transcriptional repression at bound sites (St-Pierre et al., 2002). The bHLH protein Twist titrates away E-proteins and interferes with myogenesis both at the level of MyoD and the Mef2 proteins (Spicer et al., 1996), but also has the unusual property of interfering with MyoD and myogenic activity in a protein-protein interaction that is mediated through the basic region of MyoD, rather than the HLH region (Hamamori et al., 1997). Though not belonging to the bHLH family, *Mdfi* is strongly expressed in the sclerotome during development and sequesters MRFs in the cytoplasm of cells when co-expressed, preventing their activity (Chen et al., 1996). *Mdfi* also affects

Tcf/Lef protein binding, suggesting an additional role in regulation of Wnt signaling and the ability to affect myogenesis through that action (Snider et al., 2001).

MyoD and chromatin remodeling

MyoD is capable of mediating significant chromatin remodeling at locations at which it binds. MyoD has extensively described interactions with two different histone acetyltransferases (HATs): p300 and p300/CBP-associated factor (PCAF). p300 and MyoD directly interact, and PCAF is then subsequently recruited to the complex in a p300 dependent manner, with each of the HATs responsible for a unique role in the process of acetylation (Puri et al., 1997a; Puri et al., 1997b; Sartorelli et al., 1997; Sartorelli et al., 1999). p300 acts to hyperacetylate residues in the tails of core histone members H3 and H4, while PCAF acts to acetylate MyoD itself on two residues located near the basic region. In the context of chromatin, *in vitro* assays have shown that both hyperacetylation events are necessary for strong transactivation (Dilworth et al., 2004).

Evidence from multiple MyoD-regulated gene targets - *Myog*, muscle-specific creatine kinase (*ckm*) and *MyoD* itself - have shown that prior to the action of MyoD, the DNA at those genomic locations is inaccessible as judged by nuclease accessibility assays, indicating a closed or restrictive chromatin structure. In response to MyoD, even in the presence of the protein synthesis inhibitor cycloheximide, accessibility increases, indicating that chromatin remodeling has taken place (Gerber et al., 1997). Recently, we have shown that the histone hyperacetylation that occurs in response to MyoD action is global in nature, and occurs throughout the genome at locales bound by MyoD, even those that are located distant to any known transcriptional target (Cao et al., 2010).

MyoD has also been shown to recruit the chromatin-remodeling SWI/SNF complex, a complex comprised of enzymes that perform ATP-dependent remodeling of chromatin and play roles in gene expression, development, cell fate decision, and cancer (reviewed in(Hargreaves and Crabtree, 2011). In the case of myogenic cells, the MyoD recruitment of SWI/SNF occurs through mechanisms that are dependent on p38 signaling (Simone et al., 2004). SWI/SNF complexes recruited at the *Myog* locus are done so after histone acetylation

has taken place, but actually prior to stable MyoD binding (de la Serna et al., 2005). Pbx has been shown to bind constitutively in that area, suggesting that Pbx interaction with MyoD permits first histone acetylation, than SWI/SNF activity, and finally stable MyoD occupancy (Berkes et al., 2004), suggesting a specific model for how MyoD, chromatin remodeling complexes, and myogenic co-factors can interact to regulate target expression. Interestingly, SWI/SNF activity is still required in developed cells for expression of myogenic genes, emphasizing the importance of chromatin remodeling in myogenesis (Ohkawa et al., 2007).

Other evidence has also identified negative regulators of chromatin accessibility as associating with MyoD and affecting its function. The histone methyltransferase KMT1A is responsible for methylation of histone 3, lysine 9 (H3K9), a histone tail modification associated with gene silencing. In myoblasts, KMT1A has been shown to associate with MyoD at the myogenin locus, and then decrease both in occupancy and protein level as myogenic differentiation occurs (Mal, 2006). Similarly, histone deacetylase 1 (HDAC1) has been implicated in preventing MyoD action prior to terminal differentiation. Biochemical evidence showed an association between MyoD and HDAC1 in myoblasts, suggesting that HDAC1 could be utilizing MyoD to target areas of the genome to deacetylate and thus render silent, preventing premature gene activation. During the process of differentiation to myotubes, hypophosphorylation of the retinoblastoma protein (pRb) leads to an increased association between pRb and HDAC1 and a concomitant decrease in the MyoD:HDAC1 association (Puri et al., 2001). Together, these data suggest a model in which MyoD associates in undifferentiated myoblasts with chromatin remodelers that function as silencers or repressors and lead to less accessible structures. As myogenic cells differentiate, MyoD then transitions away from this association. The experiments demonstrating a relationship between MyoD and negative chromatin regulators have been done on single targets and through biochemical approaches, and so it is still unclear at this time how generalizable these findings are across the genome.

Recent work from our lab, briefly mentioned above in the context of global histone acetylation in response to MyoD binding, has utilized the technique of chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) to identify MyoD

binding at tens of thousands of sites throughout the mammalian genome, even at very high levels of statistical stringency (Cao et al., 2010). Motif analysis of DNA located adjacent to locations bound by MyoD identified potential binding sites for well-described (eg. Meis, AP1, SP1) and less characterized (Runx1) co-factors, as well as motifs for DNA-binding factors that may themselves be interfering with MyoD binding (eg. RP58). The unexpected finding that MyoD binds at a multiplicity of sites remote to any gene target raises many questions however. One of the more obvious is how and why MyoD 'chooses' binding sites, given 1) that there are many more potential binding sites (E-boxes) in the genome than actual sites of MyoD binding, and 2) there are many more MyoD bound sites than direct gene targets. While answers to these issues are still elusive, possible explanations for these observations are addressed conceptually in Chapter 4.

Rhabdomyosarcoma

Background on rhabdomyosarcoma

Rhabdomyosarcoma (RMS) is a soft tissue sarcoma of skeletal muscle that arises mainly in pediatric populations, and is characterized by expression of myogenic regulatory factors, especially *MyoD*, and varying amounts of other skeletal muscle genes (Merlino and Helman, 1999; Merlino and Khanna, 2007; Sebire and Malone, 2003; Xia et al., 2002). Despite the expression of one or more MRFs, all RMS fail to terminally differentiate, and therefore continue their inappropriate growth. RMS ranks as the most common of all soft tissue sarcomas in children, accounting for one-half of all such cases, and approximately 5% of all cancers in children. Approximately 350 new cases are diagnosed in the US each year. Current overall survival rates are reported as approximately 70%, but prognosis is strongly dependent on tumor location (for review, see (Paulino and Okcu, 2008).

Rhabdomyosarcomas are grouped into three major subclasses – alveolar (ARMS), embryonal (ERMS), and pleomorphic. Further distinctions have been made among the subtypes themselves (eg, botryoid is a further subclass of embryonal), but classification on the level of the three major subtypes reveals substantial differences between them. Embryonal RMS is the most common subtype found in pediatric populations, making up two-thirds or more of all cases. ERMS also tends to appear in younger populations, with

tumors often located in the retroperitoneal, genitourinary, and orbital areas. Alveolar tumors comprise the bulk of the remainder of diagnosed RMS, often appearing in an older patient population than ERMS. ARMS exhibits differences in the areas it tends to affect, and is more likely to be found in the limbs in comparison to ERMS. Pleomorphic tumors hold the unusual distinction of appearing mainly in adult populations and, unsurprisingly, make up a very small proportion of all diagnosed RMS. Those few pediatric patients that are diagnosed as having disease with pleomorphic characteristics often have tumors of mixed embryonal-pleomorphic qualities (Newton et al., 1988).

As mentioned above, all rhabdomyosarcomas share the characteristic of a failure to undergo myogenic differentiation appropriately, preventing their terminal withdrawal from the cell cycle, and guaranteeing their continued proliferation. The majority of diagnosed ARMS share one of two chromosomal translocations – an aberrant joining of chromosome 13 with either chromosome 1 or 2 (t(1;13)(p36;q14) and t(2;13)(q35;q14), respectively). This translocation results in a fusion between the *FKHR* (*FOX01A*) gene, a member of the family of fork head transcription factors, and either *PAX3* or *PAX7* (Barr et al., 1993; Buckingham, 2007; Davis et al., 1994). The newly formed *PAX-FKHR* fusion gene possesses the DNA binding characteristics of PAX, but transcriptional activity up to 100 times as strong as that of non-fusion PAX (Fredericks et al., 1995). Studies comparing PAX3-FKHR and PAX3 demonstrated that the fusion protein is more capable of repressing myogenic differentiation in cultured cells than the wild-type protein (Epstein et al., 1995), and therefore is hypothesized to contribute heavily to the pathogenesis of ARMS.

In contrast to the situation with ARMS, ERMS have never been identified as possessing a characteristic chromosomal rearrangement that could help explain their evasion of terminal myogenic differentiation. An area on the small arm of chromosome 11 (11p15) has been identified as a region that often undergoes allelic loss in the tumors (Koufos et al., 1985), but chromosomal transfer experiments suggest that the chromosomal region contains a tumor suppressor, not a regulator of myogenesis (Koi et al., 1993). Indeed, none of the identified gene products from the region play a skeletal muscle specific role, but instead

include such products as the cyclin dependent kinase inhibitor *p57/Kip2* (reviewed in (Xia et al., 2002).

Transcription factors, microRNAs, and chromatin remodelers can affect RMS proliferation and differentiation

It has been demonstrated that various cellular pathways function in the maintenance of the undifferentiated state of both alveolar and embryonal RMS. Pathways that have been implicated in the block of differentiation include that of p38 MAP kinase (Puri et al., 2000), the myogenic regulatory factor *MRF4* (Sirri et al., 2003), a NF-κB/YY-1/miR-29b circuit (Wang et al., 2008), RAS (Langenau et al., 2007), c-MET and the INK4a/ARF locus (Sharp et al., 2002), myostatin (Rossi et al., 2011) and insulin-like growth factor 2 (Hahn et al., 2000). No common link has been demonstrated between these pathways however, making it unlikely to be able to purposefully develop any single curative therapy, and obscuring any common molecular origins of the tumor.

More recently, the bHLH factor *HES1* (hairy enhancer of split 1) was found to play a role in the differentiation block in RMS. Work at the Fred Hutchinson identified *HES1* as playing a key role in preventing cells from permanently withdrawing from the cell cycle – an event that occurs both in cellular senescence and in differentiation. Specifically, HES1 activity was necessary to keep cells competent to re-enter the cell cycle; abrogation of HES1 activity led to an inappropriate and permanent exit from the cycle. Since RMS themselves are a cell type that fails to permanently exit the cell cycle when it is appropriate for them to do so, a dominant negative HES1 was introduced into a cell culture model and found to lead to restoration of appropriate myogenic differentiation in the cells tested (Sang et al., 2008). Apart from the impact on RMS models, HES1 is of special interest since it has been described as functioning as an inhibitor of MyoD function; it dramatically decreases MyoD activity on myogenic reporters when co-expressed with MyoD and E-proteins. HES1 RNA is found at high levels in undifferentiated skeletal muscle and then is drastically decreased in level in mature skeletal muscle (Sasai et al., 1992).

Work in cell culture models of ARMS has identified KMT1A, the histone methyltransferase found to associate with MyoD in myoblasts, as being aberrantly regulated in RMS cells. When cells were shifted to culture conditions that would induce differentiation in normal myogenic cells – low-serum media – they actually upregulated KMT1A, and its methyltransferase activity was found to increase, not only generally, but at its known MRF target *Myog*. shRNA-mediated knockdown of KMT1A led to a differentiated cellular phenotype, increase in markers of myogenesis and, most strikingly, reduced tumor formation in xenograft mice models (Lee et al., 2011). This study offers an interesting contrast to the results with HES1, demonstrating that differentiation in RMS could be achieved not only through the actions of transcription factors, but the chromatin factors associated with them.

Other recent work has shown the impact of additional members of the myogenic network on the biology of RMS. Multiple groups have shown that the microRNA miR-206 is underrepresented in RMS, and that increasing its levels in tumors, both *in vitro* and *in vivo*, lead to differentiation of the cells (Missiaglia et al., 2010; Rao et al., 2010; Taulli et al., 2009). In the case of xenograft mouse models of RMS, the increase in miR-206 expression, induced by injections of a miR-206 expressing lentivirus, even led to reduced tumor growth while the injections continued. Taken as a whole, the data from individual players in the myogenic network, whether they are transcription factors, chromatin remodelers, or small RNAs, suggest that not only is the process of myogenesis impaired in RMS, but that manipulation of myogenesis is both possible and desirable from the perspective of treatment.

A model of rhabdomyosarcoma as trapped at a 'tipping point' in the myogenic process

Initial work from our lab demonstrated that MyoD target activation was compromised in RMS, while MyoD itself exhibited no defects. Specifically, MyoD showed low activity in measures of transcriptional activity on myogenic reporters in RMS cell culture systems, but chimeric proteins made up of the DNA binding region of a Gal protein and MyoD activated targets many times more strongly. Further, formation of heterokaryons between RMS cell lines and normal fibroblasts led to a rescue of *MyoD* activity and a restoration of the ability of RMS to differentiate into skeletal muscle (Tapscott et al., 1993). Taken as a whole, this work suggests two non-exclusive possibilities - that RMS 1) lack a necessary activating

factor that is not skeletal muscle or myoblast specific and is provided by the fibroblasts, and/or 2) lack one or more inhibitors that themselves function to downregulate or antagonize myogenic inhibitors in RMS. It is also noteworthy that heterokaryon fusion between RMS cells, even those of different subtypes, failed to rescue differentiation, suggesting some sort of common thread between cell types.

Tying in with this identification of the role of a myogenic bHLH inhibitor contributing to the state of RMS, we have recently expanded the initial work on compromised MyoD activity in RMS. We have identified multiple other bHLH myogenic inhibitors as being present in RD cells, a cell culture model of the embryonal subtype (McAllister et al., 1969), and contributing to the block in MyoD-mediated differentiation. We identified both MSC, a factor previously mentioned in the section on MyoD and myogenesis, as well as a previously uncharacterized splice form of E2A, one of the E-proteins, that we termed E2A-2/5. Both factors act in a repressive manner on MyoD activity in RD cells and, in the case of MSC, compete with MyoD for a limiting quantity of E-proteins to serve as a heterodimerization partner (Yang et al., 2009). When considered with the results from the HES1 and KMT1A studies mentioned above, this suggests a model for the molecular mechanisms that keep RMS trapped in a proliferative state.

Specifically, the data suggests that RMS are balanced between the states of proliferation and differentiation (**Figure 1.1**). They possess both MyoD and E-proteins, which should act to drive differentiation, but also possess a variety of inhibitory factors that include, but are not limited to, MSC, E2A-2/5, HES1, and KMT1A. This balance, rather than representing a tumor-specific state, is reminiscent of a state found during normal development. In normal cells, this balance between inhibition and activation serves to suspend MyoD activity until the point at which the cells have filled a sufficient anatomic space and can coordinately differentiate. In RMS, this balance is inappropriately maintained and permits the cells to continue to proliferate. When the balance is disrupted, such as by downregulating or interfering with the negative factors, or increasing the positive factors, differentiation is favored and proceeds.

Offering further support for our model, we have demonstrated that forced protein heterodimers of MyoD and the E-protein E12 differentiate RD cells when introduced retrovirally (Yang et al., 2009). As would be expected, the cells shifted to the differentiated myotube morphology, upregulated myogenic markers, and withdrew from the cell cycle. Further, in agreement with the predictions of our model, they also coordinately downregulated the levels of numerous myogenic inhibitors. E2A-2/5 decreased at the RNA level, MSC was found to decrease at the protein level, and both *Mdfi* and *Dec1* were found by expression array analysis and subsequent confirmatory RT-PCR to decrease. This suggests that the effect of the MyoD~E12 forced dimer is not simply to overwhelm the inhibitory factors that exist in RMS, but to restore the normal process of myogenesis in which inhibitory factors are downregulated as the process occurs. This results in a negative feedback loop that ensures that the activity of MyoD will proceed without interference and lock in the differentiated state.

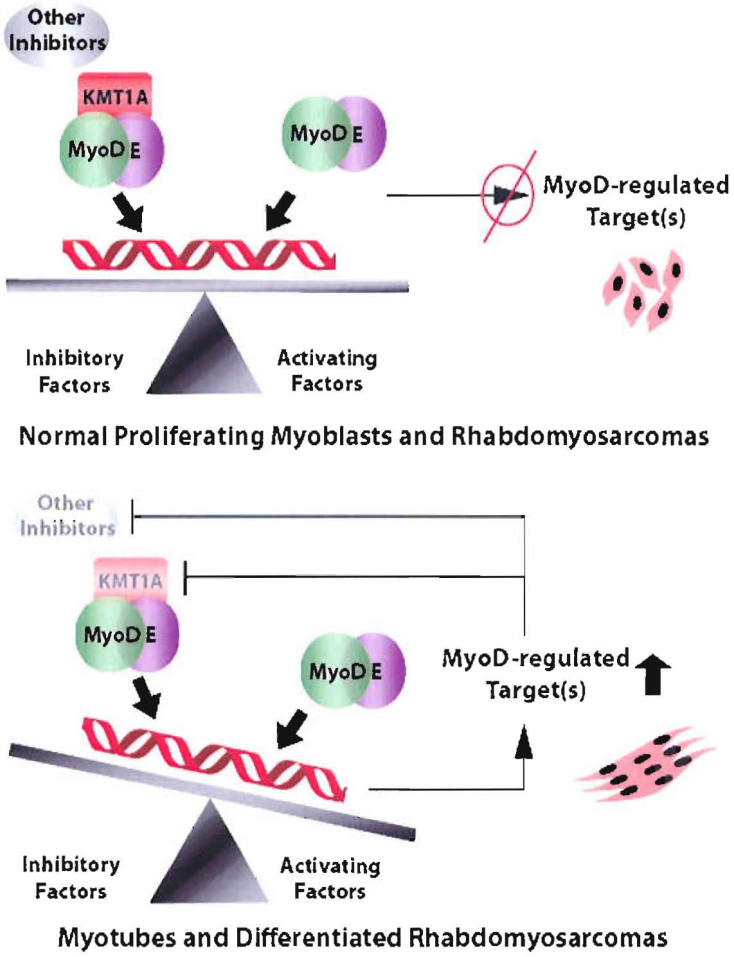


Figure 1.1. 'Tipping point' model of switch from proliferating myoblasts to differentiated myotubes in normal development and rhabdomyosarcoma. (Top) In normal, proliferating myoblasts, as well as the skeletal muscle tumor rhabdomyosarcoma, a competition exists between the activating factors MyoD and its E-protein dimerization partner and inhibitory complexes in which MyoD:E is associated with KMT1A. Numerous other inhibitory factors exist as well, resulting in the prevention of competent myogenic target activation and the cell being trapped in an undifferentiated state. (Bottom) Upon a shift in the balance of the factors such that activation predominates, myogenic targets act negatively upon the inhibitory factors and complexes and thereby perpetuate their own expression, ensuring that the process of differentiation will go to completion.

Chapter 2: Diverse means to differentiate rhabdomyosarcoma cells function through a single myogenic microRNA

Summary

The pediatric tumor of skeletal muscle, rhabdomyosarcoma (RMS), expresses the myogenic bHLH protein MyoD – a factor that should be sufficient to cause the cells to differentiate – but continues to proliferate. Restoration of myogenic activity in RMS by expression of a forced protein dimer comprised of MyoD and a dimer partner E-protein results in differentiation of the cells, and the downregulation of numerous transcriptional inhibitors of myogenesis present in the RMS cells. We now show that the action of the forced dimer increases the expression of two other transcription factors, RUNX1 and RP58, that are themselves sufficient to differentiate RMS cells when expressed in them, and that all of the pro-differentiation factors lead to an increase in the levels of the microRNA miR-206. RUNX1 appears to act to activate RP58 and miR-206 in conjunction with MyoD, suggesting an epistatic relationship between the factors. Further, the inhibitory bHLH protein MSC appears to compete with MyoD at the miR-206 promoter to prevent its strong activation by occluding a DNA binding site needed by MyoD. Together, the data suggests that multiple factors that can differentiate RMS cells function through miR-206, and its activity is regulated by a competition between bHLH factors and their DNA occupancy.

Introduction

Rhabdomyosarcoma (RMS) is a soft tissue sarcoma of skeletal muscle that arises mainly in pediatric populations, and is characterized by expression of myogenic regulatory factors (MRFs), especially MyoD, and varying amounts of other skeletal muscle genes (Merlino and Khanna, 2007; Sebire and Malone, 2003; Xia et al., 2002). MyoD is a basic helix-loop-helix (bHLH) transcription factor that serves as a transcriptional activator when bound to E-box sequences (CANNTG) in a heterodimer with one of the ubiquitous E-proteins (HEB, E2-2, or E2A) (Lassar et al., 1991). MyoD acts in a promoter-specific manner to regulate its targets, and is responsible for both activating targets directly, as well as leading to downregulation of a subset of genes (Bergstrom et al., 2002).

MyoD is capable of converting multiple cell types into terminally differentiated skeletal muscle when expressed (Davis et al., 1987), and is capable of positively regulating

the expression of the MRFs *Myog* (Cao et al., 2006) and *MRF4* (Black et al., 1995). The process of myogenesis in skeletal muscle has been described as being 'all-or-none' in nature, meaning that cells integrate multiple signals to determine whether to continue to proliferate, or switch to a terminally differentiated state, but there are no characterized 'transitional' cell states. MyoD, and the highly related factor Myf5, seem to serve as nodal points in the process, integrating multiple signals to make decisions about the mutually exclusive processes of growth and differentiation (Weintraub et al., 1991). Given the failure of RMS to differentiate while expressing *MyoD*, this suggests one or more mechanisms are operational in RMS that interfere with MyoD activity and function, trapping the tumor cells in a proliferative state.

Initial work from our lab demonstrated that MyoD target activation was compromised in RMS, while MyoD itself exhibited no defects. Specifically, MyoD showed low activity in measures of transcriptional activity on myogenic reporters in RMS cell culture systems, but chimeric proteins made up of the DNA binding region of a Gal protein and MyoD activated targets many times more strongly. Further, formation of heterokaryons between RMS cell lines and normal fibroblasts, but not between RMS cell lines themselves, led to a rescue of MyoD activity and a restoration of the ability of RMS to differentiate into skeletal muscle (Tapscott et al., 1993). Taken as a whole, this work suggests two, non-exclusive possibilities: that RMS 1) lack a necessary activating factor that is not skeletal muscle or myoblast specific and is provided by the fibroblasts, and/or 2) lack one or more inhibitory factors that act to downregulate other inhibitors and thus contribute to activation.

We have recently expanded the initial work on compromised MyoD activity in RMS and identified multiple other bHLH myogenic inhibitors as being present in RD cells, a cell culture model of the embryonal subtype (McAllister et al., 1969), and contributing to the block in MyoD-mediated differentiation. We identified both MSC (musculin, aka MyoR), a transcription factor initially described as inhibiting myogenesis (Lu et al., 1999), as well as a previously uncharacterized splice form of E2A, one of the E-proteins, that we termed E2A-2/5. Both factors act in a repressive manner on MyoD activity in RD cells and, in the case of

MSC, compete with MyoD for a limiting quantity of E-proteins to serve as a heterodimerization partner (Yang et al., 2009).

Remarkably, even given multiple bHLH inhibitors of myogenesis present in RMS cells, we have found that introduction of a forced protein heterodimer of MyoD and a full-length E2A (termed MyoD~E) leads to differentiation of the RD cells in which it is expressed. Notably, the activity of the forced dimer led to the downregulation of numerous myogenic inhibitors at the RNA and protein level. Our findings have led us to propose a model in which RMS are trapped in a proliferative state due to a 'balancing act' between the competing actions of a variety of inhibitory factors and the activating ability of MyoD and a full-length E-protein. When the balance is shifted in the direction of activation, through either a sufficient level of interference with the inhibitory factors or a sufficient enhancement to the activating potential of MyoD, a negative feedback loop mediated by MyoD is established, tips the balance, and locks in differentiation.

If RMS are poised in a balanced state between the activity of inhibitory and activating factors, then other manipulations of factors that affect the balance should also result in differentiation. Indeed, in agreement with this model, recent publications have demonstrated that downregulation of a histone methyltransferase associated with inhibition (Lee et al., 2011), interference with a transcription factor that inhibits myogenesis (Sang et al., 2008), and introduction of a pro-myogenic microRNA (Taulli et al., 2009) are all capable of pushing differentiation in rhabdomyosarcomas. Taken together, it suggests not only that RMS are representative of a point in normal myogenic differentiation, but that examination of other factors involved in normal differentiation should serve as a means to identify additional differentiating agents for RMS and expand the possibilities for therapeutic targets.

To search for additional factors of importance in the proliferation-differentiation balance, we sought to determine 1) if potential myogenic co-factors that have recently come to our attention due to our determination of the genome-wide occupancy of MyoD in myogenic cells (Cao et al., 2010) can affect this balance, and 2) what factor(s) induced by the forced MyoD~E dimer could establish a negative feedback loop and lock in the

differentiation process. We report here that the MyoD-targeted transcription factors RUNX1 and RP58 are both individually sufficient to drive myogenic differentiation in RMS culture models, and seem to support the process of myogenesis in distinct, as well as overlapping, fashions. Further, we report that both they, as well as the forced MyoD dimer, upregulate the pro-myogenic microRNA miR-206 when expressed in RMS, suggesting a common integration point for myogenesis. Finally, we offer evidence that the inhibitory bHLH protein MSC, an indirect miR-206 target, interferes with miR-206 expression by occluding an E-box that MyoD needs to occupy to drive high levels of miR-206 expression. Taken as a whole, our data suggests the existence of a multi-factorial epistatic relationship between MyoD, RUNX1 and RP58 that serves to regulate the expression of a microRNA key to making the decision of myogenic and rhabdomyosarcoma cells to switch to a state of terminal differentiation.

Results

RUNX1 and RP58, cooperative transcription factors for myogenesis, are sufficient to differentiate rhabdomyosarcoma cells

Our recent analysis of MyoD chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) has identified potential transcription factor binding sites for factors that could affect myogenesis, including both those with well-known (eg. Meis) and less-known (eg. Runx1, RP58) roles in myogenic cells (Cao et al., 2010). We hypothesized that one or more of these factors might be aberrantly regulated or otherwise lacking in RMS, and focused on factors with less- or unknown roles in myogenesis.

RUNX1 is a runt-related transcription factor best known for its role in hematopoiesis and acute myeloid leukemia (AML) (Cohen, 2009), and while it is expressed in developing skeletal muscle cells (Zhu et al., 1994) and plays a role in denervated muscle (Wang et al., 2005), its role in developing muscle is uncharacterized. The RUNX1 binding motif is associated with MyoD-bound sites in differentiated myotubes and we have previously shown that *RUNX1* expression is induced both by the myogenic feed-forward network (Penn et al., 2004), and the action of the forced MyoD~E dimer in RMS (Yang et al., 2009). qPCR confirmed that *RUNX1* levels increased with the forced dimer, and that it is expressed at

higher levels in both normal fibroblasts and myotubes created by MyoD action in fibroblasts (**Fig 2.1A**).

The binding motif for RP58, also known as ZNF238, is associated with MyoD bound sites that decrease in occupancy during differentiation, and, like *RUNX1*, increases in level with the action of the MyoD~E dimer and is expressed in myotubes (**Fig 2.1B**). Recent work has identified RP58 as being a crucial factor in myogenesis that directly downregulates the inhibitory Id factors (Yokoyama et al., 2009), factors that can interfere with the formation of functional MyoD-containing heterodimers.

Lentiviral expression vectors for each factor were cloned and transduced into RD cells. Myotubes that stained strongly for myosin heavy chain (MHC) expression, a marker of myogenesis, were observed to form specifically in the RUNX1 and RP58 infected conditions as compared to either non-infected cells or cells infected with a titer-matched GFP virus (**Fig 2.1C**). Muscle-specific creatine kinase (*CKM*), a myogenic marker, increased in both cases (**Fig 2.1D**), and EdU labeling over a 24-hour period showed a significant decrease in the number of labeled RD cells when expressing either factor, though the effect was more dramatic with those infected with RUNX1 (**Fig 2.1E**). This differentiation does not appear to be cell-type specific, as RhJT cells, an alveolar subtype cell culture model, expressing RUNX1 show an increase of MHC at the protein level and *CKM* at the RNA level (data not shown). As in normal myogenesis, expression of RP58 causes both *ID2* and *ID3* to decrease (**Fig 2.1F**). As a further control, increased expression of RP58 and RUNX1 in response to introduction of the virus was confirmed (**Fig 2.2**).

A forced MyoD~E protein dimer, RUNX1, and RP58 all increase the pro-myogenic microRNA miR-206 when expressed in RMS cells

Since the forced MyoD~E dimer induces both *RUNX1* and *RP58*, and both factors are sufficient to differentiate RMS cells, we sought to determine if all the factors act through a common mechanism. Given that we have previously identified the downregulation of multiple myogenic inhibitors in response to expression of the MyoD~E dimer, we hypothesized that a microRNA would be the most parsimonious mechanism by which it

could act. We performed microRNA microarrays with RNA from RD cells infected with either MyoD~E or control retrovirus to determine what microRNAs were altered by MyoD~E expression.

A relatively small number of microRNAs changed expression, but miR-206, a microRNA that has been shown to induce myogenic differentiation (Kim et al., 2006; Taulli et al., 2009), was the most consistently increased (**Table 2.1**). Increase of mature miR-206 in response to MyoD~E was confirmed by miRNA Northern blotting (**Fig 2.3A, upper panel**), as was miR-133b, a miRNA from the same primary transcript and another positive hit on the array (**Fig 2.3A, second panel**). Other microRNAs are maintained either at a constant level (**Fig 2.3A, middle panels**), or are decreased (**Fig 2.3A, bottom panel**) in response to the forced dimer. The constant level of miR-29b in response to the forced dimer is particularly notable, since miR-29b has previously been described as participating in a NFκB-mediated transcriptional program in RMS and driving differentiation (Wang et al., 2008). RT-PCR using primers in the presumptive human primary transcript containing miR-206 showed a substantial increase, further suggesting the increase is at the level of transcription (**Fig 2.3B**). microRNA Northern blots of RD cells differentiated through RUNX1 and RP58 expression found that, as with the forced dimer, miR-206 levels were increased in both cases (**Fig 2.3C**), with an increase in primary transcript (data not shown). miR-206 levels in C2C12 cells, a myogenic cell culture model, showed that miR-206 expression changes in proliferative versus differentiated RMS resembled the changes as C2C12 cells shift from beginning myogenesis (90% confluency) to myotubes (DM) (**Fig 2.3D**).

While miR-206 is known to be directly regulated by MyoD binding (Rosenberg et al., 2006), its relationship to *RUNX1* and *RP58* is unknown. To test the hypothesis that RUNX1 directly cooperated with MyoD in its regulation, the response of a miR-206 promoter luciferase reporter to RUNX1 was examined (**Fig 2.3E, black bars**). RUNX1 alone leads to a minor activation of the reporter, while RUNX1 combined with MyoD leads to a synergistic activation of the reporter compared to either individual transcription factor. Experiments using a reporter in which a putative RUNX1 binding site has been mutated demonstrates not only that RUNX1 fails to activate the reporter, either alone or when combined with MyoD,

but that the ability of MyoD, either by itself or when combined with E12, to activate that reporter is reduced (**Fig 2.3E, grey bars**). ChIP experiments also identified RUNX1 as binding in the miR-206 promoter (**Fig 2.3F**). Taken together, this suggests that RUNX1 binding at the miR-206 promoter not only enhances the expression of miR-206, but is necessary for a fully competent activation by MyoD. RP58 did not lead to activation of the reporter (data not shown), suggesting that RP58 could be acting on miR-206 indirectly, possibly through its effect on the ID proteins and MyoD dimerization.

In agreement with previous reports demonstrating that miR-206 alone is sufficient to differentiate RMS cells, transfection of pre-miR-206 constructs into RD cells resulted in dramatic myotube formation (**Fig 2.4A**), an increase at the RNA level of the myogenic marker *CKM* (**Fig 2.4B**), and a withdrawal of such cells from the cell cycle (**Fig 2.4C**), with similar results in the alveolar RMS model, RhJT cells (**Fig 2.4D and data not shown**). As would be expected from prior reports of its effect on myogenic cells (Chen et al., 2006), introduction of miR-133b did not lead to RMS differentiation as judged by either morphology or gene expression (**Fig 2.4E, F**).

MyoD activity positively regulates RUNX1 and RP58 and RUNX1 positively regulates RP58

To further understand the relationship between *MyoD*, *RUNX1*, *RP58*, and miR-206, we sought to identify controlling factors of *RUNX1* and *RP58* expression. Previously, it has been shown that MyoD activates reporters driven by the *RP58* promoter (Yokoyama et al., 2009). In agreement with this, induction with beta-estradiol of MyoD activity in human fibroblasts stably expressing an estradiol-inducible MyoD demonstrated an increase in *RP58* expression at 6 and 12 hours post-induction (**Fig 2.5A**). Interestingly, *RP58* expression also increased in response to expression of RUNX1 in RD cells, but the converse was not true; RP58 did not upregulate *RUNX1* expression and actually led to a slight decrease (**Fig 2.5B**). ChIP data also identifies RUNX1 as bound at the first intron of *RP58*, suggesting it functions directly to activate RP58 (**Fig 2.5C**).

Bisulfite sequencing reveals no evidence for methylation of a *RUNX1* promoter (**Fig 2.6**), so the possibility that the specific E-protein that serves as the MyoD dimer partner

controls *RUNX1* expression was tested. RD cells transduced with the MyoD~E dimer were compared directly to those transduced with a dimer comprised of MyoD and the E2/5 splice form of the E protein.

An obvious morphological difference is apparent between MD~E and MD~E2/5 expressing RD cells. Those expressing the forced dimer that included the full-length E12 protein exhibit formation of myotubes in the great majority of cells on the plate, while those expressing the splice form of the dimer formed myotubes at a substantially reduced level (**Fig 2.7A**). Western blots were used to confirm relatively equivalent expression of the forced dimers in infected cells (**Fig 2.7B**). MD~E expressing cells were found to express substantially more *CKM* relative to MD~E2/5 expressing cells (**Fig 2.7C**), and higher *RUNX1* levels (**Fig 2.7D**), demonstrating that the specific bHLH protein partnered with MyoD is critical for competent expression of its downstream targets. Taken as a whole, the data suggests a unidirectional relationship proceeding from *MyoD* to *RUNX1* to *RP58* and then miR-206, with RUNX1 cooperating with MyoD at *RP58* and miR-206, and RP58 functioning indirectly at miR-206.

miR-206 target genes are a subset of the genes affected by RUNX1 and RP58

To determine the genes regulated by RUNX1, RP58, and miR-206 in the context of the RMS differentiation, and further test our model that they are in an epistatic relationship, gene expression arrays were performed on RNA from RD cells differentiated by each one of the aforementioned factors and compared to RD cells infected with a GFP-expressing virus. GO analysis of the effects of each individual factor ranked by most significant p-values identified multiple muscle related categories for upregulated genes, with 5 of the 10 most significant categories shared between all factors (**Table 2.2**). In agreement with our hypothesis of an epistatic relationship, the number of genes that were identified as being significantly regulated (fold change > 2, FDR <0.05) by each factor became sequentially reduced from RUNX1 (735) to RP58 (617) to miR-206 (355). Also as predicted, the target overlap between individual factors was substantial, as was the overlap between all three (**Fig 2.8A**), and there was substantial correlation between target genes (**Fig 2.8B**). As would be expected if the effects of RUNX1 and RP58 on RMS are mediated through miR-206, gene

targets identified as being 'unique' to the miR-206 condition were also found in the RUNX1 and RP58 arrays, but at lower fold-changes (**Fig 2.8C, top**). A similar, though weaker, effect was seen with the RP58 'unique' targets appearing in the RUNX1 array at lower fold-changes (**Fig 2.8C, bottom**).

RUNX1, RP58 and miR-206 were notable for having significant effects on a few key transcription factors and cellular signaling cascades involved in myogenesis (**Table 2.3**), as confirmed by RT-PCR (**Figure 2.8D**). All three led to a significant upregulation of the MRF, *MYOG*, a target of MyoD (Cao et al., 2006). RUNX1 upregulated *MEF2C* and *MEF2D*, additional cooperative factors for MyoD activity (Penn et al., 2004). RP58 downregulated transcription factors of two groups of interest: 1) positive regulators of cell cycle (*MYCN, RCOR2, E2F2*) and 2) members of the *HES/HEY* family (*HEY1, HES6, HEYL, HES1*). It has previously been demonstrated that interference with HES1 contributes to RMS proliferation (Sang et al., 2008), and the *HES/HEY* family is known to be Notch responsive (Fischer and Gessler, 2007), a signaling pathway with myogenic inhibitory effects (Buas et al., 2009; Kopan et al., 1994; Lindsell et al., 1995). Among miR-206's most strongly downregulated targets were two members of the Notch signaling pathway, *DLL3* and *NOTCH3*.

The bHLH protein MSC occupies an E-box in the miR-206 promoter that MyoD requires for strong activation of miR-206 expression

The above data suggests that miR-206 expression is a crucial decision point for myogenic differentiation, with sufficient expression capable of forcing differentiation. We have previously identified murine miR-206 as being regulated by MyoD binding (Rosenberg et al., 2006), but it was unclear whether MyoD itself is misregulated at miR-206 in RMS, or if another factor is responsible for the insufficient expression. We have previously shown that MyoD can bind targets in RMS (Tapscott et al., 1993), but that the bHLH protein MSC interferes with MyoD activity (Yang et al., 2009). Since MSC has been shown to be downregulated by miR-206 activity (Kim et al., 2006), we hypothesized MyoD and MSC might operate in opposing fashion at the level of miR-206 regulation. The miR-206 luciferase reporter was tested with MSC, and co-transfection of MSC along with MyoD and E12 almost completely ablated the ability of MyoD and E12 to activate the reporter (**Fig 2.9,**

black bars). This repression was also found when MSC was co-transfected with the forced MyoD~E12 dimer, suggesting the effect of MSC is due to binding on the DNA, not interference with the formation of MyoD:E dimers (**Fig 2.10**).

MyoD was found by chromatin immunoprecipitation (ChIP) in the promoter region of the primary transcript for miR-206, in agreement with our previous findings in murine cells (**Fig 2.11A),** and ChIP for acetylated H4 histones suggested the locus was open (**Fig 2.11B**). In agreement with the ability of MSC to suppress the miR-206 reporter, ChIP demonstrated MSC also bound at the miR-206 promoter (**Fig 2.11C**), and at a comparable enrichment to that of MyoD.

Sequence analysis of the area assessed by ChIP located three potential MyoD- and/or MSC-binding E-boxes. We have recently performed ChIP-Seq using MyoD and MSC in RD cells, as well as MyoD in human fibroblasts differentiated to myotubes through the action of MyoD (see Chapter 3). The promoter region of miR-206 was examined and a strong MyoD peak observed at the same location interrogated by site-specific ChIP (**Fig 2.12, top panel**). MSC was also found in that area, though its peak of occupancy was at a different E-box compared to MyoD (**Fig 2.12, middle panel**). Electrophoretic mobility shift assays demonstrate that both types of heterodimers can bind both of the E-boxes. Competition assays demonstrate that both heterodimers prefer the E-box at which MyoD binds, but that the relative level of that preference is greater for MyoD than MSC (**Fig 2.13**).

To test the hypothesis that the MSC binding was interfering with MyoD activation of miR-206, the miR-206 reporter was mutated to scramble the MSC occupied E-box. Rather than becoming insensitive to MSC-mediated repression, the reporter instead became insensitive to activation by MyoD and E12 (**Fig 2.9, grey bars**). Taken together with the ChIP-Seq data identifying distinct sites of occupancy, this suggests that MSC is repressing the reporter by physically occluding an E-box that MyoD needs to occupy to fully activate miR-206. The ChIP-Seq data on MyoD in the human myotubes offers further support for this model. Compared to the MyoD peak in RD cells, there is a broadening of the MyoD peak in myotubes that appears to widen to include E-boxes located more proximally to the

start of the miR-206 transcript (**Fig 2.12, bottom panel, arrow**), suggesting that in myotubes, MyoD occupies additional positions. In addition, site-specific ChIPs identify a reduced level of MSC occupancy at the miR-206 promoter in RD cells that undergo RUNX1-mediated differentiation (**Fig 2.14A**), and MD~E differentiation (**Fig 2.14B**).

Discussion

We have previously proposed a model of RMS as being a balancing act between the activities of repressive and activating bHLH protein dimers, a balance that, when tipped by the MyoD~E dimer in the direction of activation, induces one or more factors 'X' to downregulate the myogenic repressors (Yang et al., 2009). Our present data offers further support for this model, demonstrating that the transcription factors *RUNX1* and *RP58* can also drive RMS differentiation, and that miR-206 can function as 'X'. RUNX1 enhances MyoD activity, at a minimum at miR-206, and judging from the expression array data, possibly at some of the key MyoD downstream targets (eg. *MYOG*, *MEF* genes). In contrast, RP58 activity downregulates not only multiple members of the inhibitory *HES* and *HEY* protein family, but multiple factors that drive cellular proliferation as well. But despite appearing to act directly at different targets, RUNX1 and RP58 both ultimately serve to increase miR-206 transcription and lead to a terminally differentiated state, supporting our hypothesis that, regardless of the precise mechanism used to tip the balance in myogenic cells to favor differentiation, the end result is the same.

MyoD acts in a complex feed-forward network to regulate its target genes, requiring some of its direct target genes to cooperate with it at later targets, allowing for fine temporal control. The data on RUNX1 activity suggests that it fits into this network as a cooperating factor for MyoD to assist in regulating a subset of targets, such as *RP58* and miR-206. While the regulation of *RUNX1* in myogenic cells will need further exploration, our data demonstrate that the E-protein partner of MyoD is an important factor in its regulation. RP58, on the other hand, does not appear to cooperate directly with MyoD at targets. Our data on *RP58* regulation, motif analysis, and gene targets in this and previous work (Cao et al., 2010) suggests that the induction of this inhibitory factor serves two purposes: 1) to downregulate genes that inhibit myogenesis and promote proliferation, and 2) to interfere

with MyoD binding itself, possibly assisting to shift the genes regulated by MyoD during development through that mechanism. Genome-wide exploration of the direct targets of RUNX1 and RP58 will be of great interest to further delineate their respective roles in myogenesis.

Musculin was originally identified as a bHLH protein that inhibits the process of myogenesis (Lu et al., 1999), and our data suggest that, at least at miR-206, it functions by physically occluding an E-box that MyoD needs to occupy for full activation. A requirement for multiple MyoD-bound E-boxes to drive full target activation has been described before (Gilmour et al., 1991; Lassar et al., 1989; Piette et al., 1990; Weintraub et al., 1990; Wentworth et al., 1991), but this is the first evidence that MSC can operate in this fashion and the first evidence for this sort of relationship in miR-206 regulation. Future work will be necessary to determine if this is a widespread mechanism at other myogenic targets, and if the close, but distinct pattern of MyoD and MSC binding is a common occurrence. Given the fact that previous work has identified MSC as being downregulated by miR-206 activity (Kim et al., 2006), and MSC occupancy of the miR-206 promoter decreases after RMS differentiate, this suggests MSC and miR-206 act in a classic negative regulatory loop, though possibly in an indirect fashion.

Our work in rhabdomyosarcomas has demonstrated striking levels of similarity between the tumors and the biology in normal myogenic cells. This suggests, therefore, that mechanisms of differentiation utilized in normal cells and model systems might be translatable to approaches to differentiate RMS. Certainly, the finding that multiple factors capable of differentiating RMS all share regulation of miR-206 as a common point suggests that screens to identify druggable targets that affect miR-206 regulation could negate the need for attempts at differentiation therapy to optimize microRNA mimetics or delivery systems.

Hematological malignancies have long been categorized and described on the basis of cell differentiation state. Our work suggests that the same may be possible for solid tumors, with rhabdomyosarcomas appearing as an arrested point of the myoblast to myotube

transition, trapped on the brink of the decision point to differentiate. While myogenic cells have been notable for more than two decades for the sharp demarcation between the processes of proliferation and differentiation, it is possible that this is a lesson more broadly applicable to other cell and tumor types. bHLH factors control cell fate and differentiation in multiple cell types, and experiments in tumors from such systems may identify not only the existence of other 'tipping points', but a more widespread utility to manipulating the controlling factors in such systems to stop the growth of tumor cells.

Materials and Methods

Trizol – Acid Phenol RNA Isolation

Cells from one 10 cm plate for each biological condition were rinsed once with PBS, then scraped up into 1 mL of Trizol (Invitrogen). After vortexing to the point of no visible clumps, tubes were incubated at room temperature for five minutes. If necessary, Trizol solutions were frozen at -80° C before further processing. To continue RNA isolation, 200 ul of chloroform were added to each tube. After 15 seconds of vortexing, tubes were incubated at room temperature for 2 minutes, then spun at 10,600 RPM for 15 minutes at 4° C in a tabletop microcentrifuge. The aqueous layer (approximately 600 ul) was removed to a new tube, then an equal volume of isopropanol was added, mixed, and then incubated at -20° C for 30 minutes. After incubation, tubes were spun at 4° for 20 minutes at 14,000 RPM in a tabletop centrifuge. After visualization to ensure a visible RNA pellet had been precipitated, the solution was removed. One mL of 75% ethanol was added to the tube, vortexed and then spun at 9500 RPM at 4° C for 5 minutes in a tabletop centrifuge. The RNA was then resuspended gently, with pipetting, in 400 uL of DEPC-treated water at room temperature. An equal volume of acid phenol (Ambion, pH 4.5) was then added to the resuspended RNA, vortexed for 3-5 seconds to mix thoroughly, and the tube then spun at 12000 RPM at 4° C for 15 minutes. An aqueous layer of approximately 380 uL was recovered and an equal volume of chloroform added. Vortexing and spinning was repeated exactly as in the acid-phenol addition step, and an aqueous layer of approximately 360 uL recovered. 3 M sodium acetate solution was then added to reach a final concentration of 0.3 M (40 uL to 360 uL recovered RNA solution), and, after mixing, 1 mL of 100% cold ethanol added. After thorough mixing

by inversion, the tubes were then placed in a cold bath made of dry ice and 95% ethanol for 5 minutes. Tubes were then spun at 14,000 RPM for 20 minutes at 4° C in a tabletop centrifuge. The precipitated pellet was visualized, the supernatant removed and the pellet washed in 1 mL of 75% cold ethanol with brief vortexing. The tube was spun for 5 minutes at 4° at 9500 RPM and all the ethanol carefully removed from the pellet. After air drying for approximately 10 minutes, to a point where there was no visible moisture, but before the center portion of the pellet turned transparent, the pellet was resuspended in 50-75 uL of nuclease-free water. RNA concentration was checked by UV spectroscopy or Nanodrop (Thermo Scientific) and stored at -80° C.

microRNA Northern Blots

Northern blot ladder preparation: Radioactively labeled ladders were prepared fresh for each experiment. 100 ng of 10 bp DNA ladder (Ambion) was combined with 1 uL of 10x PNK Buffer (NEB), 6 uL of water, 1 uL of T4 PNK enzyme (NEB) and 1 uL of a 1:500 dilution of γP^{32}-ATP (PerkinElmer). The mixture was incubated at 37° C for 30 minutes, and then incubated on a heat block at approximately 95° C for 5 minutes to denature the enzyme.

Acrylamide gel preparation: A 15% polyacrylamide gel was prepared fresh for each experiment the same day. A small (15 mL) Erlenmeyer flask was used to combine 5.5 mL of 40% 1:19 acrylamide, 1.1 mL of 10x TBE (Tris-Borate EDTA solution), 330 uL of DEPC-treated water, and 5.28 g of urea. To make a homogenous solution, the flask was microwaved at 100% power in a standard microwave for 5 second intervals three times, with 5 - 10 seconds of swirling of the flask after each interval. After the solution was completely homogenous, it was allowed to cool for approximately 5 minutes at room temperature and then filtered through a 0.45 um syringe filter into a 15 mL conical tube with a screw cap (Falcon). To that filtered solution, 66 uL of 10% APS (ammonium persulfate) was added first, and then 6.6 uL of TEMED. After inversion of the tube to mix, the solution was then poured into a gel casting apparatus that had been set up using 0.75 mm combs and spacers, and a comb with 10 wells. After allowing the gel to set for approximately 30 minutes, the comb was carefully removed and wells cleaned out with a small piece of filter paper and gentle pipetting of the running buffer (see below).

Running the samples: 25 ug of RNA, prepared using the Trizol-acid phenol approach, was used for each lane. Samples were concentrated in a speed vacuum as necessary to reach a volume of 5 - 8 uL. To each sample, an equal volume of 2x loading dye (Ambion Gel Loading Buffer II) was added. Dye was also added as with the RNA samples to the ladder prepared above. All samples and the ladder were then incubated in a 65° C water bath for 15 minutes, and then moved immediately to ice for 10 minutes. During this time, the solidified gel prepared above was placed in a gel-running apparatus, as used standardly for Western blots, in 1X TBE buffer and pre-run at 100 V for at least 10 minutes. After samples and ladder were on ice for 10 minutes, they were immediately loaded (the entirety of their volume) onto the gel, taking care to not load anything in the two wells on either edge. The gel was run first at 100 V for 1.5 hours then, after the lower dye front had reached approximately halfway the distance down the gel, run at 150 V for another 1.5 hours, until the lower dye front just reached the bottom.

Transferring and hybridizing the samples: The samples were transferred to a nytran SPC membrane in 1X TBE buffer at 250 mA for 45 minutes in standard Biorad wet transfer apparatus. The transfer components were assembled in the 1X TBE buffer in the following order, listed in the order from closest to the clear side of the transfer cassette to closest to the black side of the cassette: sponge, 2 Whatman papers, membrane (presoaked in water before being placed into the TBE), gel, 2 Whatman papers, sponge. The assembled components were gently pressured in a rolling fashion with a pipette tip to make sure no bubbles were present. The cassette was then latched closed, and placed into the electrical apparatus so that the clear side of the cassette faced the positive pole (the red side on Biorad transfer materials). After 45 minutes of transferring, the cassette was disassembled, and the membrane was UV-crosslinked for 2 minutes on an automatic setting (Stratagene UV Stratalinker 1800). The blot was washed 3 times with normal water and then placed into a hybridization tube. The blot was prehybridized for 2 hours at 35° C in a rotating hybridization oven with 12 mL of Ultra-hybe buffer (Ambion) that had been heated briefly to 65° C to go into solution then cooled to room temperature before addition to the blot. During this prehybridization, the probe(s) was/were prepared. The following components were combined: 14 uL of water, 1 uL of 10 uM oligo, 2 uL of T4 PNK enzyme, 2 uL of 10x PNK buffer, 1 uL of undiluted γP^{32}-ATP and incubated for 30 minutes at 37° C. Probes were then

heated on the heating block for 5 minutes, as with ladder preparation. Each probe was then brought to a final volume of 50 uL with double distilled water, and run through a G-25 column (GE Healthcare) as per manufacturer's directions to purify and the elution collected. After prehybridization was completed, purified probe was added to the blot and hybridized with rotation overnight at 35° C. The next morning, the blot was washed twice in 2X SSC/0.5% SDS for 25 minutes each time at room temperature with rocking. Blots were removed from wash solution, allowed to drip dry, and then wrapped in plastic wrap before being taped in a film cassette, blue X-ray film added, and allowed to expose for varying lengths of time at -80° C.

Stripping blots for serial probing: To re-probe blots, they were first stripped for 2 hours at 85° C in an excess of 1% SDS with vigorous rocking. Blots were then rinsed once with normal water, and prehybridized as above before the addition of new, freshly prepared probe.

Probe Sequences: microRNA probes consisted of the reverse complement of the sequence of mature microRNA of interest. Probe sequences were: miR-206: CCACACACTTCCTTACATTCCA; miR-133b: TAGCTGGTTGAAGGGGACCAAA; miR-29b: AACACTGATTTCAAATGGTGCTA; miR-16: CGCCAATATTTACGTGCTGCTA; miR-199a*: TAACCAATGTGCAGACTACTGT

microRNA Transient Transfections

pre-microRNA constructs for miR-206, miR-1 and miR-133b were purchased from Ambion. Reverse transfections were done at final concentration of 25 uM of the pre-miRNA using siPORT NeoFX (Ambion) as per manufacturer's directions. An appropriate volume of pre-miR was diluted into 100 uL of Opti-mem media (Invitrogen) and then mixed with 5 uL of siPORT that had been diluted in the same manner. After a 10 minute incubation at room temperature, 1 x 10^5 RD cells were mixed in a 9:1 ratio of cells to transfection material and placed at 37° C. After 72 hours of growth, cells were shifted to low-serum differentiation media for 24 hours before harvested or fixed for further analysis.

Chromatin Immunoprecipitation

All ChIPs were performed on RD cells transduced with Babe-based retroviruses. After 16 – 24 hours of infection in the presence of 8 ug/mL polybrene, and 24 hours of

recovery, cells were selected for 40 – 48 hours in 1.5 ug/mL puromycin to eliminate non-infected cells. Cells were then washed repeatedly and shifted to low-serum differentiation media for 24 - 28 hours before harvesting. Cells were washed twice in PBS + 2% serum, and then fixed for 11 minutes at room temperature in a 1% PBS formaldehyde solution (5 mM HEPES pH 8.0, 10 mM NaCl, 0.1 mM EDTA, 0.05 mM EGTA). The formaldehyde was quenched with the addition of glycine to a final concentration of 0.25 M. After removal of the solution, cells were washed twice with ice-cold PBS, and scraped into PBS with the addition of PMSF and complete protease inhibitors (Roche). Cells were spun down at 1100 RPM at 4° C, and then resuspended in 2 mL of lysis buffer (1% SDS, 5 mM EDTA, 1% Deoxycholate, 50 mM Tris pH 8.0) for 10 minutes on ice. Sonication was performed using a probe sonicator (Fisher, Sonic Dismembrator Model 500) at 45% amplitude for 3 minutes total, with cycles of 30 seconds on and 59 seconds off. This results in chromatin sheared to sizes of roughly 150 – 600 bp long. Chromatin was then spun at 14000 RPM for 10 minutes at 4° C and 100 uL of the chromatin set aside at -20° to use as input. The remaining chromatin was diluted 1:10 in dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris pH 8.0) and precleared for 2 hours rocking in the cold with 200 uL of Protein A/G agarose beads (Upstate). Beads were spun down for 1 minute at 700g and the chromatin then divided to immunoprecipitate overnight at 4° with nutation with 20 uL of antibodies. For transcription factor ChIPs, 5 – 6 mLs of chromatin were used, and 1 mL was used for histone modification ChIPs. Control precipitations were performed using equal volumes of chromatin. The following day, 20 uL of Protein A/G beads were added, and the mixture rocked in the cold for an additional 2 hours. Beads were harvested by centrifugation, and then rocked sequentially in Paro Buffer 1 (0.1% SDS, 1% Triton X-100, 2mM EDTA, 20 mM Tris pH 8.1 150 mM NaCl), 2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris pH 8.1, 500 mM NaCl), and 3 (0.25 M LiCl, 1% NP-40, 1% Deoxycholate, 1 mM EDTA, 10 mM Tris pH 8.1) before a final wash in TE (10 mM Tris pH 8.0, 1mM EDTA). All washes were 10 minutes in length at room temperature, and the beads were harvested between each wash with centrifugation in a table-top centrifuge at 2500 RPM for 1 minute. After the TE wash, beads were resuspended in 150 uL elution buffer (1% SDS, 0.1 M $NaHCO_3$) and allowed to sit at room temperature for 15 – 20 minutes with occasional gentle tapping before being placed in a 65° water bath overnight. Inputs were also placed at 65° after having SDS

and $NaHCO_3$ added to the same final concentration. The following day, the elutions were separated from the beads and purified using the PCR purification kit (Qiagen) according to manufacturer's directions. Elutions were done in 50 uL Buffer EB.
Antibodies used were as follows: Runx1 (Abcam, ab23980), MyoD (Tapscott et al., 1988), MSC (Santa Cruz, sc-9556X), Acetylated Histone H4 (Upstate 06-866). Primers used for site-specific amplification were: miR-206: CAACAAGCACCCAAAACAGA, TTCCACATTCACGCAGAGAG; HBB control locus: AACGGCAGACTTCTCCTCAGG, AGTCAGGGCAGAGCCATCTA; miR-206 locus for Runx binding: TGGCATATGTTTCCCCATTT, GTTGAGCCACTCAGGGTCTG; RP58: CCACAGTCAGCTGGATCAGA, GAGGGCAGCTCACAAGGTAG; RUNX control locus (MYH8 enhancer): TGTGGCTATCTCTGTGTGCAG, TTAGATTTTGGGGGATGGTG

Expression microarrays

RNA was isolated using the RNeasy mini kit (Qiagen) from RD cells infected with either RUNX1-, RP58-, miR-206- or GFP-expressing lentiviruses and allowed to differentiate for 72 hours. Each condition was performed with 3 independent biological replicates. RNA was hybridized to Illumina Human HT-12 v4 BeadChips. Analysis was performed in R/Bioconductor using the lumi and limma packages with annotations found in the lumiHumanAll.db package. p-values were adjusted to account for multiple testing using Benjamini and Hochberg's method, and cut-offs for significant changes were a FDR <0.05 and a fold-change >2. GO category enrichment tests were performed using the conditional algorithm of the GOstats package and a gene "universe" of any gene with a GO annotation that was called as "present" in at least one of the 12 array datasets.

EdU labeling, Western blots, and cell stains

After 24 hours in low-serum differentiation media, cells were shifted to differentiation media supplemented with EdU at a final concentration of 50 uM (Invitrogen) and incubated for a further 24 hours. Cells were then fixed and stained according to the manufacturer's protocols using the Click-iT kit, and total nuclei and EdU positive nuclei counted by hand.

Western blots were performed on whole cell lysates collected in Laemelli buffer containing 10% beta-mercaptoethanol. All blots were blocked in 3% milk (w/v) in 0.5% Tween-20-containing PBS before incubation with primary antibody (MHC: MF-20, MyoD: 5.8A, Runx1: Abcam, ab23980), a HRP-conjugated secondary antibody, and chemiluminescent detection (Amersham).

Cells were fixed with 2% paraformaldehyde for 6 minutes at room temperature before permeabilization with Triton X-100. Myosin heavy chain was detected with the MF-20 antibody, and nuclei detected with DAPI.

microRNA microarrays

RNA was isolated using acid-phenol purification from RD cells transduced with either MD~E or empty vector retroviruses and differentiated for 24 hours after puromycin selection. miRNAs were labeled using Exiqon's miRCURY labeling kit, and then competitively hybridized to in-house spotted miRNA arrays (FHCRC core facility). Cut-offs for significant changes were a FDR <0.05 and a fold-change >2.

qPCR and RT-PCR

All qPCR was performed using SybrGreen from Bio-Rad on an Applied Biosystems 7900HT. Relative expression levels were calculated using cDNA dilution standard curves or delta-delta Ct calculations. All values are reported as the mean $\pm$ SEM of at least 3 independent biological experiments. Primers used for amplification were as follows: CKM: CCAAGTTCGAGGAGATCCTC, AGCTGCACCTGTTCTACTTCG; TIMM17b: GGAGCCTTCACTATGGGTGT, CACAGCATTGGCACTACCTC; ID2: CCCAGAACAAGAAGGTGAGC, ATAGTGGGATGCGAGTCCAG; ID3: CTGGACGACATGAACCACTG, GTAGTCGATGACGCGCTGTA; MYOG: GGCCACAGATGCCACTACTT, GCTTTACCTCCCTGGAAAGG; MEF2D: CTCTTTGCCGTGACAACACC, CTCATGAACGGTCTGGGAAC; MYCN: CACAAGGCCCTCAGTACCTC, CACAGTGACCACGTCGATTT; E2F2: CTACACACCGCTGTACCCG, CCAGATCCAGCTTCCTTTTG; RCOR2: TCAGCTCATCTCCCTCAAGC, TAGTGGATCAATACCGCCCT; HEYL:

ATCGACGTGGGCCAAGAG, ATCCCTCTGCGTTTCTTCCT; HEY1: TGGATCACCTGAAAATGCTG, CGAAATCCCAAACTCCGATA; pri-miR-206 (regular RT-PCR): GTTTCGGCAAGTGCCTCCT, CTCTTGCTTCCTTGGTGAGG; (qPCR): TGCTGTGAGTGAGGTTCAGG, CAGGGTTGTGGTGTGAAGTG; NOTCH: TGTGCAAATGGAGGTCGTT, CCTGAGTGACAGGGGTCCT; DLL3: CATCGAAACCTGGAGAGAGG, CCTGCGCGCTGAATGTC.

Plasmid construction

The coding sequences of RUNX1 and RP58 were cloned into pRRLSIN.cPPT.PGK/GFP.WPRE in the BamHI/SalI sites. The miR-206 lentivirus was purchased from Open Biosystems. Lentiviral supernatant was produced by the FHCRC core viral facility. MD~E2/5 was cloned into the pCLBabe backbone and packaged using BBS-mediated calcium precipitation into Phoenix cells. For the miR-206 promoter luciferase reporter, a ~2.5 kb piece of DNA upstream of human miR-206 was amplified using the primers GAATGCTAGCCTGTCCTTGATTTTACCC and CAATAGATCTTTGTGCAGCTACAGTCTA and cloned into the NheI/BglII sites in pGL3 basic.

Cell culture, transient transfections and luciferase assays

RD cells were maintained in DMEM with 10% bovine calf serum and 1% Pen-Strep (Gibco). Low-serum differentiation media consisted of DMEM with 1% horse serum, 1% Pen-Strep and 10 ug/mL insulin and transferrin. Transient transfections for luciferase assays were performed using Superfect and a total of 3 ug of DNA – a total of 1.5 ug of plasmids being tested for their effect on the reporter, 1 ug of luciferase reporter, and 0.5 ug of renilla internal control (renilla-CS2) (Qiagen). Luciferase assays used the Dual-Luciferase Assay kit (Promega) according to manufacturer's directions. All results were corrected to co-transfected Renilla-pCS2 and are reported as the mean $\pm$ SEM of at least 3 independent experiments, each experiment having 3 biological replicates of all conditions.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays were performed as described previously (Davis et al., 1990). Briefly, proteins were translated in vitro using a rabbit reticulocyte lysate system (Promega) and synthesis of a single protein product of the correct size confirmed using parallel ^{35}S-labeled translations. Equal volumes of translation product were used in each lane, and balanced with empty CS2 translated reactions or reticulocyte lysate alone. Mixtures were incubated at 37°C for 20 min in a DNA binding cocktail solution (either 20 mM HEPES (pH 7.6), 3 mM $MgCl_2$, 1 mM DTT, 1 mM EDTA or 20 mM HEPES (pH 7.6), 1.5 mM $MgCl_2$, 1mM DTT, 1 mM EDTA, 50 mM KCl) before γ-^{32}P ATP-labeled oligonucleotide probes were added at room temperature for 15 min. Probe sequences were as follows (forward probes only listed): MSC-bound E-box: TGGATGGGCAGCTGCTGCCCAT; MyoD-bound E-box: TGGCTCAACAGCTGCCAATGTC. Complexes were resolved on 6% polyacrylamide gels and exposed to radiographic film.

Bisulfite conversion and sequencing

Bisulfite reactions and sequencing were performed as has been described previously (Diede et al., 2010). Primer sequences were: set 1: GGTAGGAGTTGTTTGTAGGGTTTTA, CCCACATCCCAAACTAAAAAAA; set 2: GGAGATTTGGAAAAAGAAAGTAGGT, AAAATCTTTCCTAACTAAAAAACTCTTC; set 3: GAGTTAAGTTTTAGGAATAGGGGTTT, CCCTCCCCCAAAACTAAAATACTA; set 4: GGTGTATGTAAGGTTGGGATTAATTT, CCACTTTCTAACTCTATCCCTAAAAAAA.

Table 2.1 **miRNA changes in response to MyoD~E12 expression in RD cells.**

Experiment Number	miRNA ID§	FC (log_2)*
Biological Replicate #1	**hsa-miR-206, mmu-miR-206, rno-miR-206, dre-miR-206, gga-miR-206, mdo-miR-206, mne-miR-206, ppy-miR-206, xtr-miR-206**	1.345; 1.3225
	hsa-miR-663	1.103
	mml-miR-133a, ppy-miR-133a	1.015; 1.0385; 1.0005
	hsa-miR-133a-133b	1.032; 1.0155
	cel-miR-243	1.0215
	hsa-miR-199a*, mmu-miR-199a*, bta-miR-199a*, dre-miR-199*, gga-miR-199*, xtr-miR-199a*	-1.015
	hsa-miR-143	-1.025
Biological Replicate #2	**hsa-miR-206, mmu-miR-206, rno-miR-206, dre-miR-206, gga-miR-206, mdo-miR-206, mne-miR-206, ppy-miR-206, xtr-miR-206**	1.081; 1.021
	hsa-miR-125a, mmu-miR-125a, rno-miR-125a, bta-miR-125a	-1.004
	hsa-miR-335, mmu-miR-335, rno-miR-335	-1.021
	mghv-miR-M1-6	-1.025
	hsa-miR-196b, mmu-miR-196b, rno-miR-196b, mdo-miR-196b	-1.0305
	ptc-miR478a-ptc-miR478b-ptc-miR478c	-1.0325
	dme-miR-6,dps-mir-6	-1.0865
	hsa-miR-199a*, mmu-miR-199a*, bta-miR-199a*, dre-miR-199*, gga-miR-199*, xtr-miR-199a*	-1.106
	hsa-miR-653	-1.11
	hsa-miR-99b, mmu-miR-99b, rno-miR-99b	-1.1225; -1.137
	ath-miR160a-, ath-miR160b-, ath-miR160c, gma-miR160, mtr-miR160, osa-miR160a-, osa-miR160b-, osa-miR160c-, osa-miR160d, ptc-miR160a-, ptc-miR160b-, ptc-miR160c-, ptc-miR160d, sbi-miR160d-, sbi-miR160a-, sbi-miR160c-, sbi-miR160b-, sbi-miR160e, zma-miR160a-, zma-miR160c-, zma-miR160d-, zma-miR160b-, zma-miR160e	-1.182
	cbr-miR-249	-1.2025
	osa-miR169f-, osa-miR169g, ptc-miR169r, sbi-miR169c-, sbi-miR169d, zma-miR169f-, zma-miR169g-, zma-miR169h	-1.436
Biological Replicate #3	hsa-miR-302a, mmu-miR-302	1.3945
	osa-miR164c	1.2045
	hsa-miR-206, mmu-miR-206, rno-miR-206, dre-miR-206, gga-miR-206, mdo-miR-206, mne-miR-206, ppy-miR-206, xtr-miR-206	1.0395

Table 2.1, continued

	hsa-miR-204, mmu-miR-204, rno-miR-204, dre-miR-204, fru-miR-204, gga-miR-204-211, ggo-miR-204, mdo-miR-204, mne-miR-204, ppa-miR-204, ppy-miR-204, ptr-miR-204, sla-miR-204, ssc-miR-204, tni-miR-204a, xtr-miR-204	-1.0145
	hsa-miR-548b	-1.016
	mmu-miR-684	-1.025
	hsa-miR-376b	-1.045
	mmu-miR-505	-1.0675
	hsa-miR-214, mmu-miR-214, rno-miR-214, age-miR-214, bta-miR-214, dre-miR-214, fru-miR-214, ggo-miR-214, mdo-miR-214, mml-miR-214, mne-miR-214, ppa-miR-214, ppy-miR-214, ptr-miR-214, sla-miR-214, ssc-miR-214, tni-miR-214, xtr-miR-214	-1.088; -1.016
	ath-miR159a, gma-miR159, ptc-miR159a-, ptc-miR159b-, ptc-miR159c	-1.137
	mmu-miR-467a	-1.184
	cel-miR-84	-1.2475
	hsa-miR-589	-1.258
	dre-miR-27c, fru-miR-27c, tni-miR-27c	-1.4145
	hsa-miR-549	-1.75
Biological Replicate #4	ath-miR394a-, ath-miR394b, osa-miR394, ptc-miR394a-, ptc-miR394b, sbi-miR394a-, sbi-miR394b, zma-miR394a-, zma-miR394b	1.801
	aga-miR-9c, dme-miR-9c, dps-miR-9c	1.252
	hsa-miR-206, mmu-miR-206, rno-miR-206, dre-miR-206, gga-miR-206, mdo-miR-206, mne-miR-206, ppy-miR-206, xtr-miR-206	1.247; 1.2055
	rlcv-miR-rL1-11	1.14
	cel-miR-243	1.0905; 1.0415
	mmu-miR-679	-1.0035
	dme-miR-9b, dps-miR-9b	-1.0105
	hsa-miR-20b, mmu-miR-20b, rno-miR-20b, gga-miR-20bxla-miR-20, xtr-miR-20b	-1.119
	hcmv-miR-US25-2-5p	-1.1675
	osa-miR441a-, osa-miR441b-, osa-miR441c	-1.4745
	mmu-miR-291a-5p-291b-5p, rno-miR-291-5p	-1.526

§All miRNA names generated from the miRNA expression array are listed, regardless of whether the sequences are identical amongst species.

*All changes are listed as the $\log_2$ value in order from most upregulated to most downregulated, within the restrictions of FDR <0.05 and $|\log_2$ fold-change$|$ >1 for each independent biological replicate. If more than one change was identified for a single miRNA, they are listed individually.

Bold miRNA names indicate those miRNAs found in more than one biological replicate.

Table 2.2. **GO Categories of genes upregulated by RUNX1, RP58 and miR-206.**

	GO Term[a]	P-value[b]	GO Category Size[c]	Gene Count[d]
RUNX1-regulated genes				
	Muscle filament sliding	1.49E-19	30	18
	Actin-mediated cell contraction	1.49E-19	30	18
	Actin filament-based movement	5.53E-18	40	19
	Structural constituent of muscle	1.93E-15	29	15
	Z disc	5.21E-14	33	15
	Heart process	5.77E-12	51	16
	Sarcomere	7.94E-12	32	13
	Muscle organ development	1.05E-09	175	25
	Blood circulation	1.33E-09	138	22
	Actin cytoskeleton	8.89E-09	168	23
RP58-regulated genes				
	Structural constituent of muscle	8.23E-11	29	11
	Myofibril	9.40E-11	84	17
	Muscle filament sliding	1.45E-10	30	11
	Actin-mediated cell contraction	1.45E-10	30	11
	Actin filament-based movement	4.84E-09	40	11
	Actin cytoskeleton	1.07E-08	223	24
	Muscle organ development	4.13E-08	175	20
	Myosin filament	8.04E-08	10	6
	Cell differentiation	9.43E-08	1241	64
	Developmental process	1.22E-07	2251	97
miR-206-regulated genes				
	Muscle filament sliding	1.06E-16	30	13
	Actin-mediated cell contraction	1.06E-16	30	13
	Actin filament-based movement	9.40E-15	40	13
	Sarcomere	2.17E-13	32	11
	Actin cytoskeleton	1.08E-12	223	22
	Muscle cell development	6.10E-12	63	13
	Heart process	7.75E-12	51	12
	Striated muscle cell differentiation	1.06E-10	78	13

Table 2.2, continued

	Structural constituent of muscle	1.27E-10	29	9
	Cardiac muscle contraction	2.27E-10	20	8

Bold categories indicate those that are common between RUNX1, RP58, and miR-206.
[a]The name of the indicated GO category
[b]P-value associated with the GO category
[c]The total number of genes contained within the GO category
[d]The number of genes contained with the GO category found to be significantly regulated in the analysis.

Table 2.3. **Select potential myogenic regulators affected by RUNX1, RP58, and miR-206.**

	Gene Symbol	RUNX1 FC (log_2)[a]	RP58 FC (log_2)	miR-206 FC (log_2)
Strong RUNX1 regulation				
	MYOG	**2.01**	1.28	1.17
	MEF2C	**1.81**	1.10	0.84
	MEF2D	**1.50**	0.37	0.69
Strong RP58 regulation				
	MYCN	-0.71	**-3.34**	-1.29
	RCOR2	-1.11	**-2.03**	-0.86
	HEYL	-0.69	**-1.95**	-0.24
	HES6	-0.23	**-1.63**	-0.39
	E2F2	-0.43	**-1.67**	-0.01
	HEY1	-0.53	**-1.05**	0.23
	*HES1**	-1.09	**-0.72**	-0.41
Strong miR-206 regulation				
	NOTCH3	-0.39	-1.5	**-1.75**
	DLL3	-2.07	-1.88	**-1.93**

*Though it didn't reach the 2-fold change cut-off, *HES1* was included for its known role in RMS.

[a]All fold-changes are reported as the log2 value.

Bold numbers indicate the fold-change in the analysis that originally identified the genes as being of potential interest.

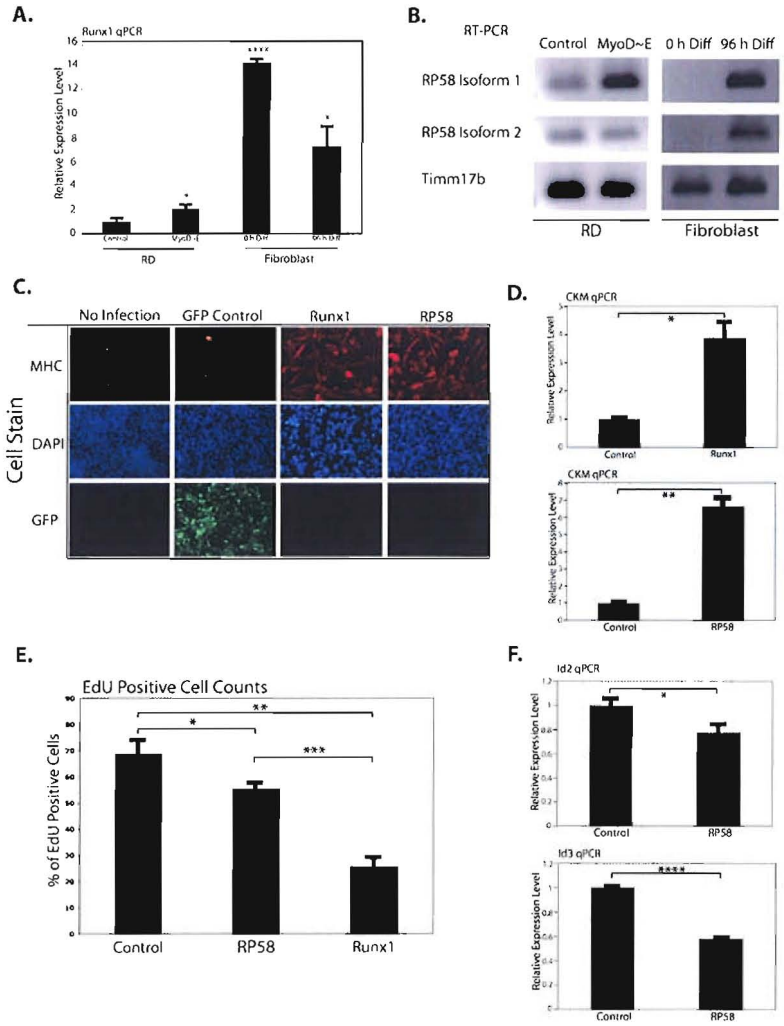


Figure 2.1. Expression of *RUNX1* or *RP58* leads to terminal differentiation of RMS cells. A) qPCR for *RUNX1* was performed in RD cells infected with a control virus, or the forced MyoD~E dimer (MyoD~E) as well as control (0 h) human fibroblasts and fibroblasts differentiated into myotubes (96 hr). B) RT-PCR for the two isoforms of *RP58* in RD cells and fibroblasts as in 1A. C) Myosin heavy chain (MHC) immunostains in RD cells either not infected, infected with a control GFP-expressing lentivirus (GFP control) or RUNX1 or RP58 expressing lentivirus. All cells were infected at equivalent MOIs, and cells differentiated for 72 hours before staining. GFP was detected directly. D) qPCR for muscle-specific creatine kinase (*CKM*) in RD cells infected with either RP58 or RUNX1 viruses. E) After 24 hours of differentiation, RD cells were pulsed for a further 24 hours with EdU-containing differentiation media, before fixation and quantification of the percentage of EdU positive cells. D) qPCR for *ID2* and *ID3* in control and *RP58* expressing RD cells. All qPCR data are normalized to *TIMM17b* expression, and the level in control cells is set to 1. All bar graphs represent the mean $\pm$ SEM of at least 3 independent experiments. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$; ****: $p< 1x10^{-4}$

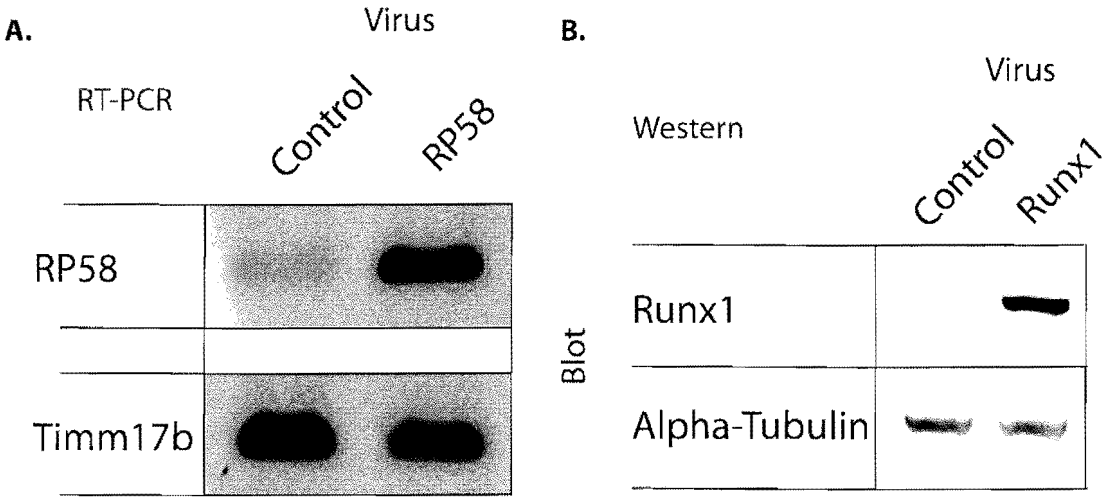


Figure 2.2. RD cells infected with RP58 and RUNX1 viruses increase expression of the appropriate factor. A) RT-PCR for *RP58* in RD cells infected with either a control virus or the *RP58*-containing virus. *TIMM17b* is used as a loading control. B) Western blot for RUNX1 in control and RUNX1 virus infected RD cells. The blot was then stripped and reprobed for alpha-tubulin as a loading control. Bands were confirmed to be of the correct size.

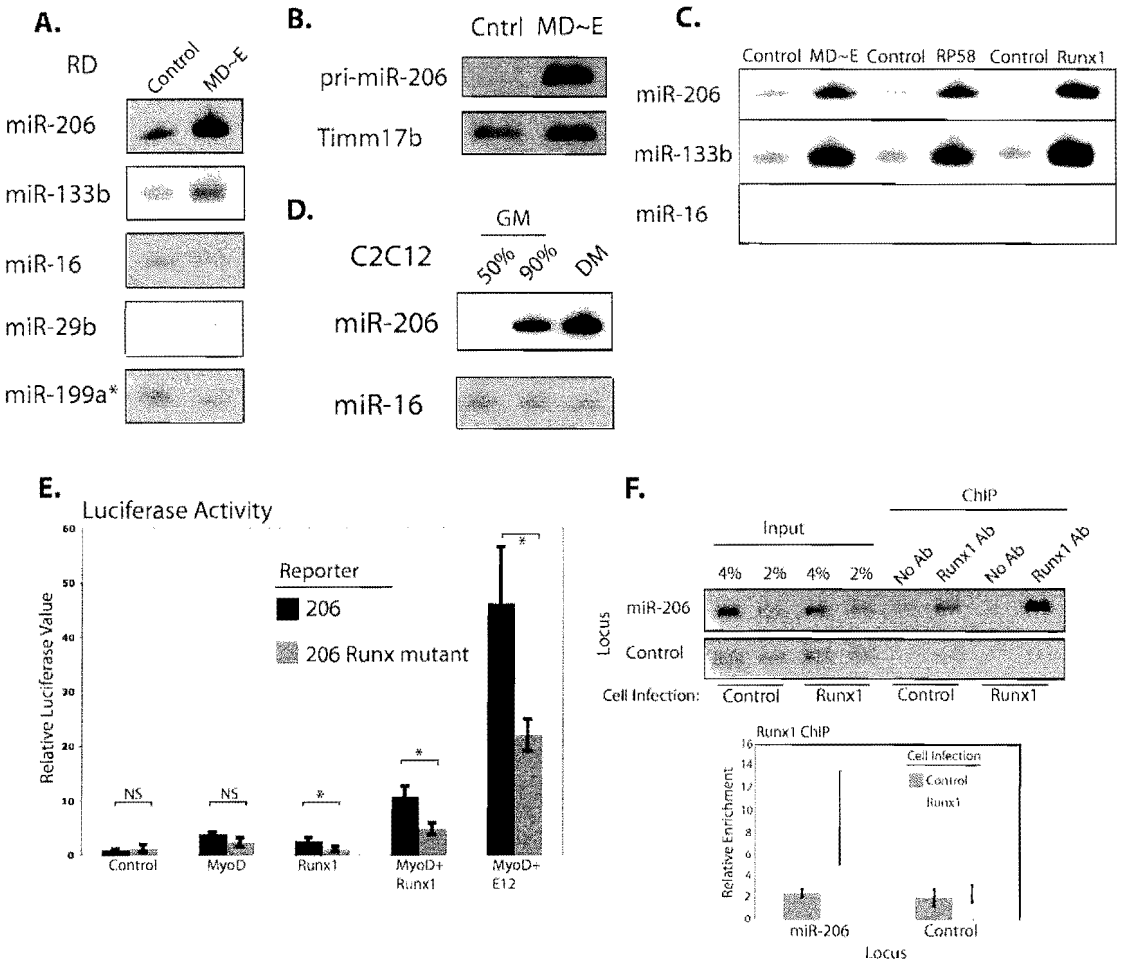


Figure 2.3. MyoD~E, RUNX1, and RP58 increase miR-206. A) microRNA Northern blots to detect the mature form of the indicated microRNAs in either control or MD~E expressing RD cells. B) RT-PCR using primers located in the pre- and pri-miR-206 sequence to detect the primary miR-206 transcript. *TIMM17b* is an internal control. C) microRNA Northerns as in 2A, in RD cells infected with either empty (control) retrovirus, or retrovirus expressing a transcription factor as indicated. D) microRNA Northerns for the indicated miRNAs in C2C12 cells at various stages of differentiation ranging from undifferentiated myoblasts (50% GM), through beginning differentiation (90% GM) to myotubes (DM). E) Luciferase activity in RD cells using a miR-206 promoter driven reporter and transiently transfected factors as indicated. '206 Runx mutant' indicates that the reporter has had a putative RUNX1 binding site mutated to prevent RUNX1 binding. Luciferase experiments were performed in triplicate, and are reported as the mean $\pm$ SEM. F) RUNX1 ChIP assays at the miR-206 promoter and a control locus before (control) and after (RUNX1) infection of the cells with empty or RUNX1-expressing retrovirus. PCRs were performed for the same number of cycles. The graph represents the mean $\pm$ SEM of 2 independent experiments. * : $p<0.05$

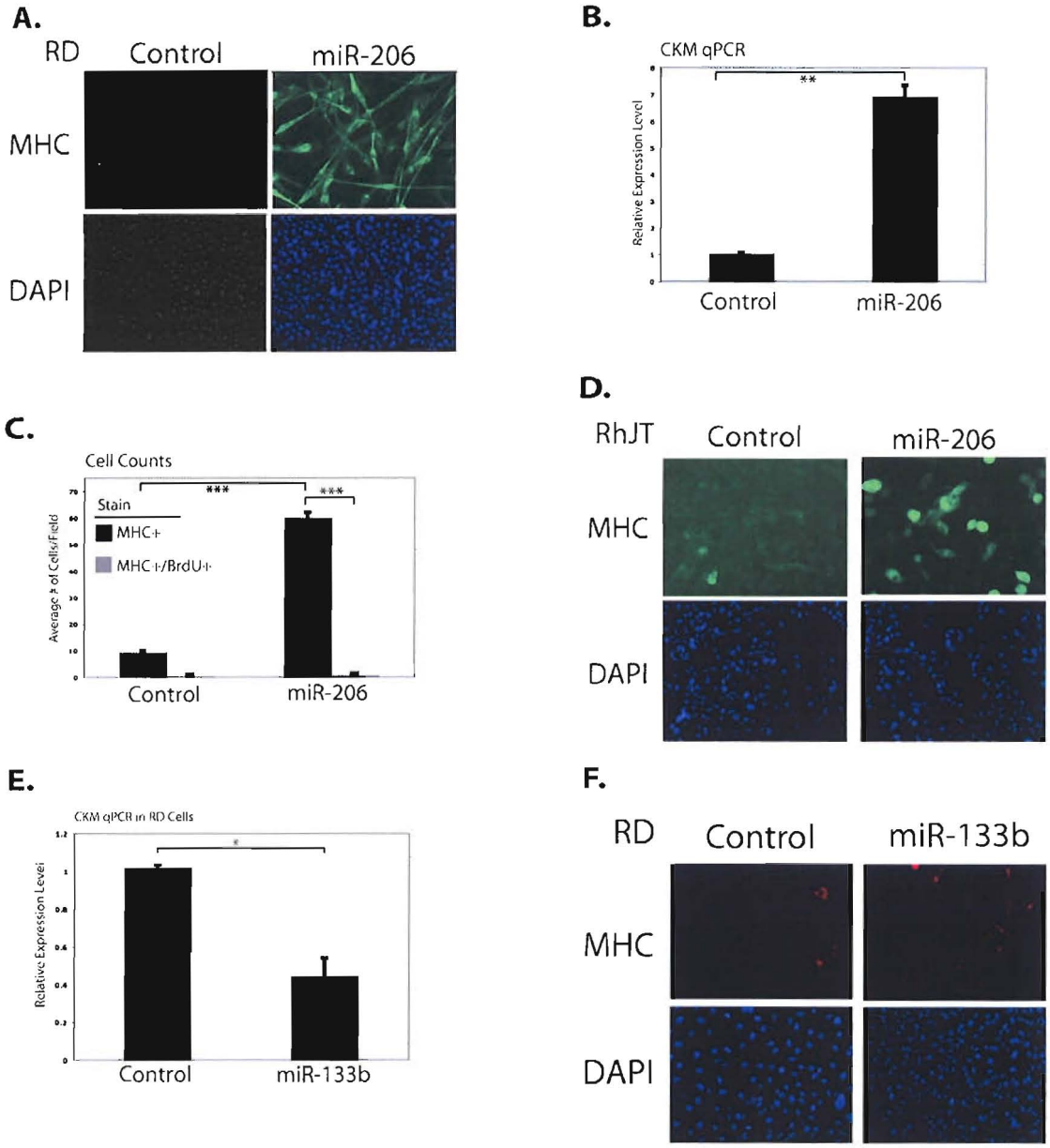


Figure 2.4. miR-206, but not miR-133b, differentiates RMS cells. A) Immunostains for MHC in RD cells transfected with either a pre-miR-206 RNA construct, or a negative control construct. Nuclei were stained with DAPI. B) qPCR for *CKM* in RD cells treated as in A. C) RD cells treated as in A were pulsed with BrdU for 24 hours and then stained and counted by hand to determine the extent of co-localization of MHC-expressing myotubes, and nuclei with BrdU. D) Immunostains for MHC in RhJT cells that were treated as in A. E) qPCR for *CKM* in RD cells transfected with pre-miR-133b or control. F) Immunostains, as in part A, in RD cells transfected with either pre-miR-133b or control. All bar graphs are the mean $\pm$ SEM of at least 3 independent experiments, and qPCR results were normalized to *TIMM17b*. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

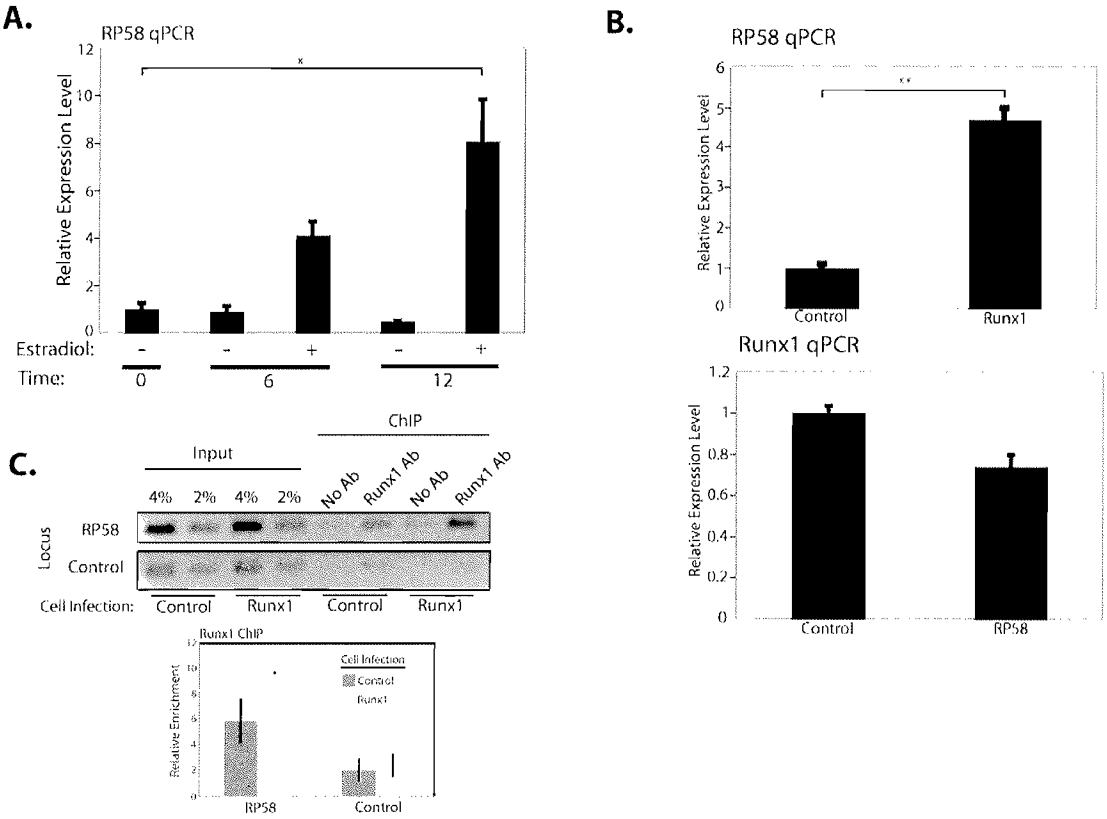


Figure 2.5. RUNX1 and MyoD both positively regulate *RP58* expression. A) 10T1/2 fibroblast cells stably expressing an estradiol-inducible version of MyoD were induced to undergo myogenesis by addition of beta-estradiol to the culture medium. RNA was taken at the indicated timepoints and conditions, and qPCR performed to quantitate the relative levels of *RP58* over time. B) qPCR for *RUNX1* and *RP58* in RD cells transduced with virus expressing the converse factor. C) RUNX1 ChIP assays at the intron of *RP58* and a control locus before (control) and after (Runx1) infection of the cells with empty or RUNX1-expressing retrovirus. PCRs were performed for the same number of cycles. The graph represents the mean ± SEM of 2 independent experiments.

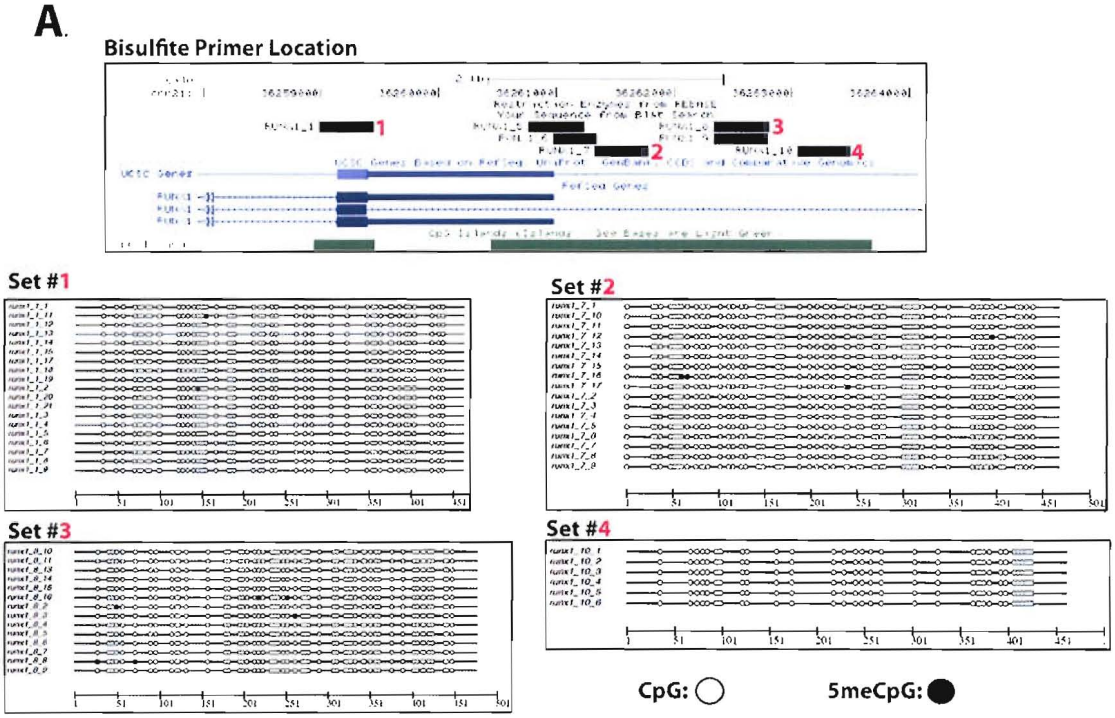


Figure 2.6. *RUNX1* is not hypermethylated in RD cells. A) The upper panel indicates the location of primers designed to interrogate the methylation status of this *RUNX1* promoter region in a UCSC browser shot. Numbered panels below correspond to the numbered bands above. Darkened circles indicate a methylated CpG, while empty circles indicate an unmethylated CpG. Each horizontal row indicates an individual sequenced clone.

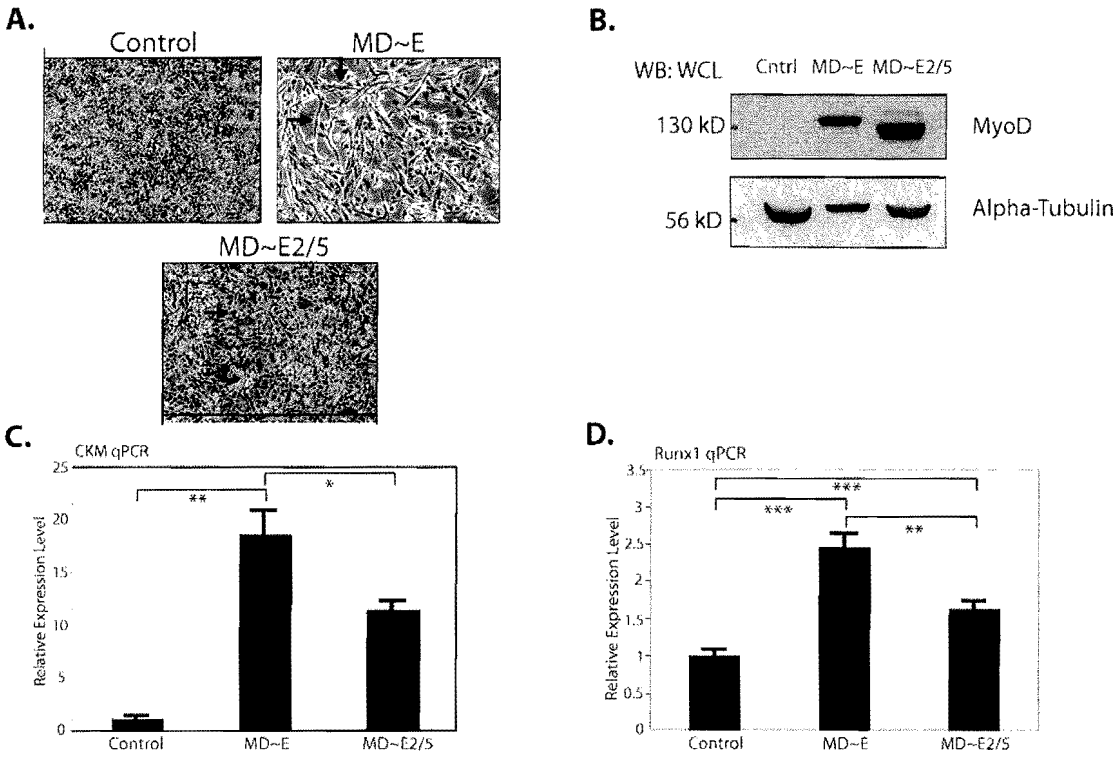
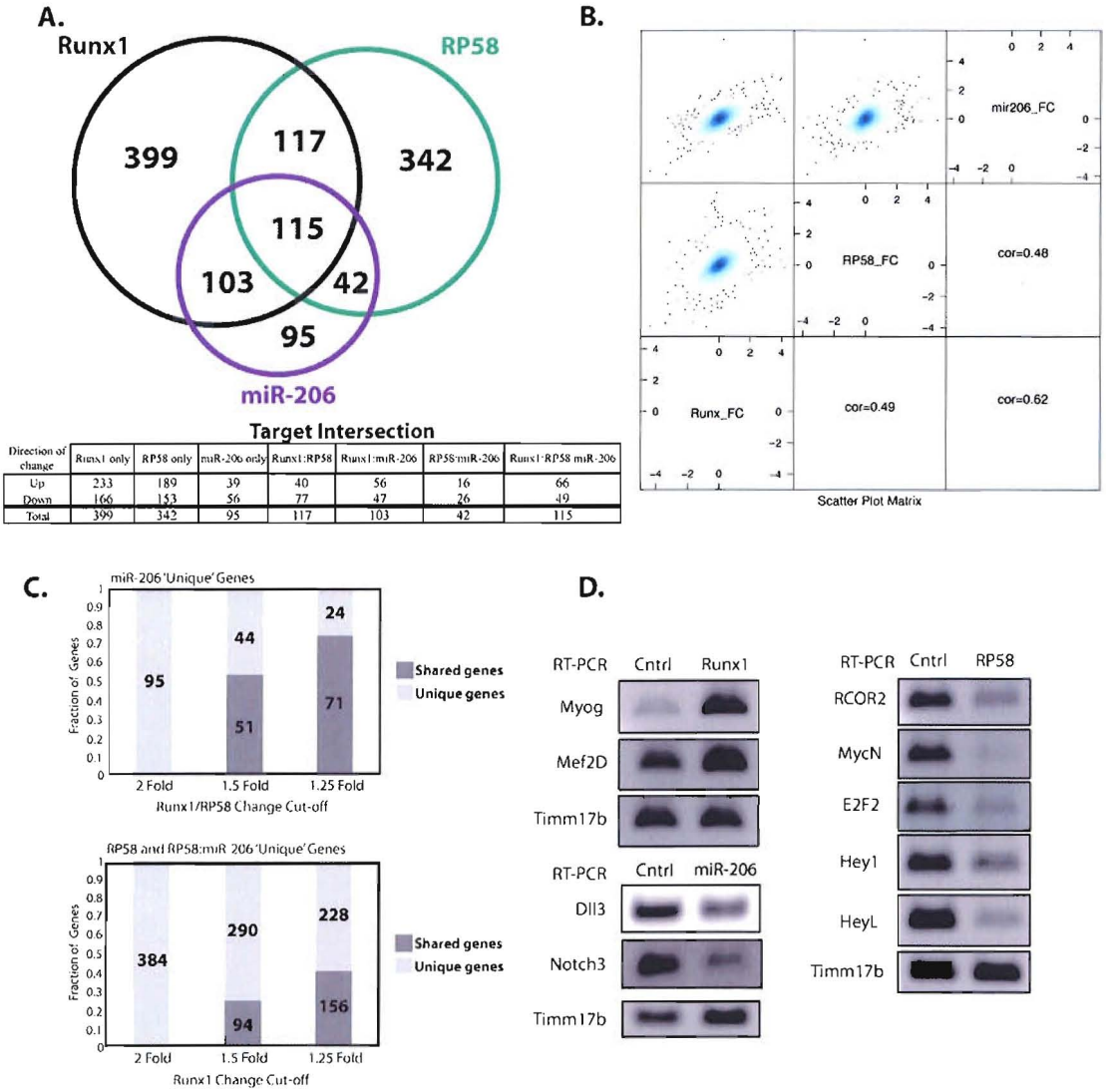


Figure 2.7. A forced MyoD~E2/5 dimer does not fully activate myogenic targets. A) Light microscopy images of RD cells infected with either control virus or virus expressing either the MyoD~E or MyoD~E2/5 forced protein dimer and allowed to differentiate for 24 hours. Arrows indicate representative cells that have appeared to form myotubes. B) Western blot for MyoD and alpha-tubulin, as a loading control, from cells treated as in 5A. The size of the bands detected in MD~E and MD~E2/5 lanes correspond roughly to the calculated size of the MyoD~E dimer. C) qPCR for *CKM* in RD cells treated as in 5A. D) qPCR for *RUNX1* in RD. All qPCRs are represented as the mean ± SEM of at least 3 independent experiments. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.



Target Intersection

Direction of change	Runx1 only	RP58 only	miR-206 only	Runx1:RP58	Runx1:miR-206	RP58:miR-206	Runx1:RP58 miR-206
Up	233	189	39	40	56	16	66
Down	166	153	56	77	47	26	49
Total	399	342	95	117	103	42	115

Figure 2.8. RUNX1, RP58, and miR-206 function through common mechanisms. A) 3-way Venn diagram representing the overlap between significantly regulated (fold-change >2, FDR <0.05) gene targets in RD cells differentiated either through RUNX1, RP58, or miR-206 expression relative to GFP-infected controls. B) Scatter plots showing pairwise comparisons of gene expression from the expression data used in 5A. 'Cor' indicates the correlation for each comparison. C) Clustered bar graph demonstrating that the majority of genes listed as being 'uniquely' regulated by miR-206 in 5A, are also regulated by RUNX1 and/or RP58, but at lower levels of expression change. FDR was kept constant (<0.05) in this analysis, and to be included as a 'shared' target, the change must occur in the same direction (either up- or down-regulated) in RUNX1 and/or RP58 as in miR-206. D) RT-PCR for various gene targets from Table 2. *TIMM17b* serves as the internal control.

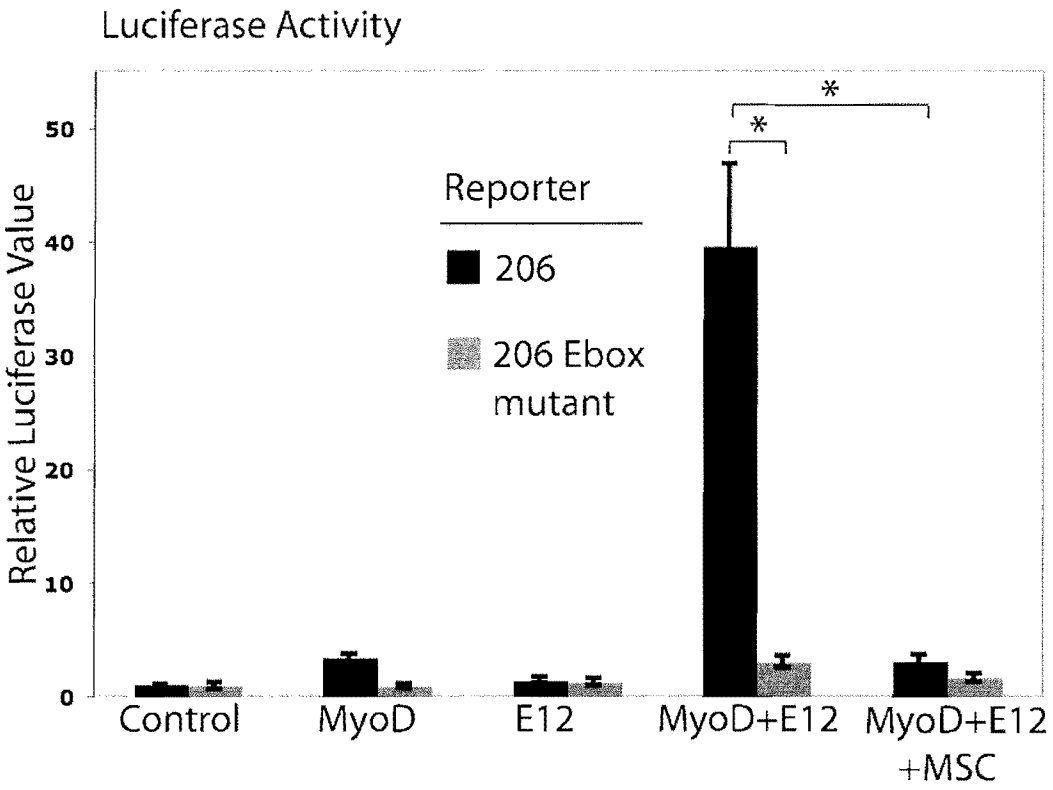


Figure 2.9. MSC represses MyoD activation of miR-206 and occupies an E-box MyoD requires. Luciferase assays in RD cells with constructs as indicated below the figure using either the miR-206 promoter luciferase reporter (206) or one in which the E-box that the peak of MSC occupancy is located over has been mutated (206 Ebox mutant).

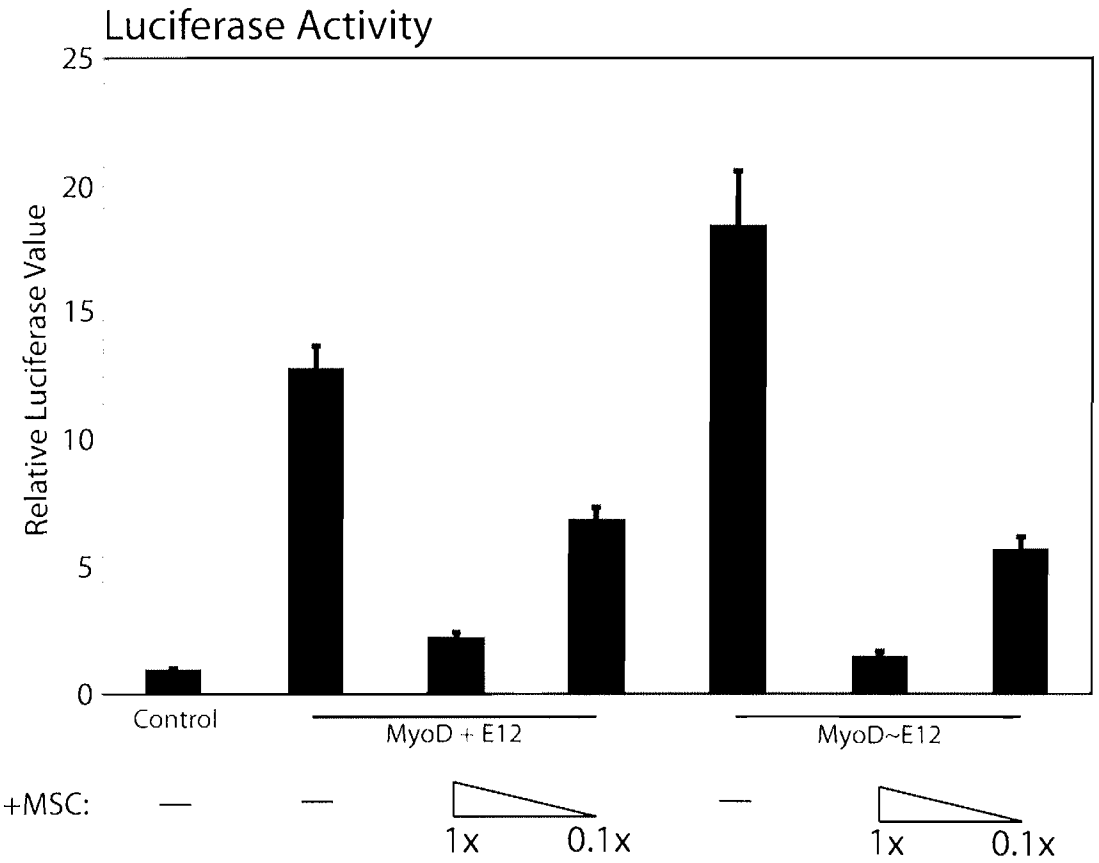


Figure 2.10. MSC inhibits the activation of the miR-206 reporter by the forced MyoD~E dimer. Luciferase assay using the miR-206 reporter with either MyoD and E12 introduced individually or as the forced dimer, in the presence of two different amounts of co-introduced MSC. 1x indicates that the MSC transfected was equal to amount of MyoD (or MyoD~E), and 0.1x indicates that the MSC transfected was 1/10th that amount. Values are represented as the mean $\pm$ SEM of 3 independent experiments.

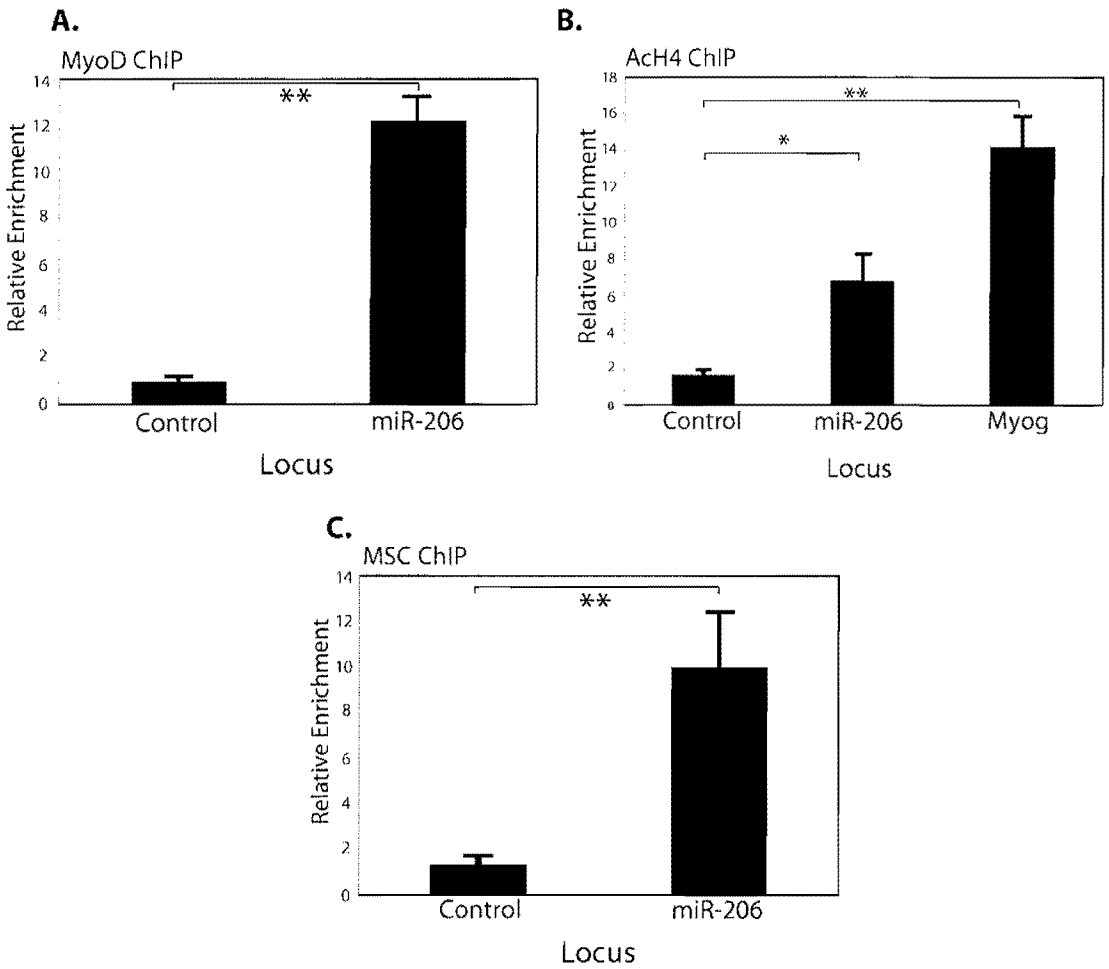


Figure 2.11. Site specific ChIPs in RD cells. A) ChIP for MyoD in RD cells in differentiation media shows MyoD enrichment upstream of miR-206, but none at hemoglobin beta (control). B) Site-specific ChIPs in RD cells for acetylated histone H4, a marker of histone acetyltransferase activity and an open chromatin structure, at hemoglobin beta (control), miR-206, and the myogenin promoter (Myog). C) ChIP for MSC using the same primers as used for Part A. All ChIPs are represented as the mean $\pm$ SEM of at least 3 independent experiments. * : $p<0.05$; ** : $p<0.01$.

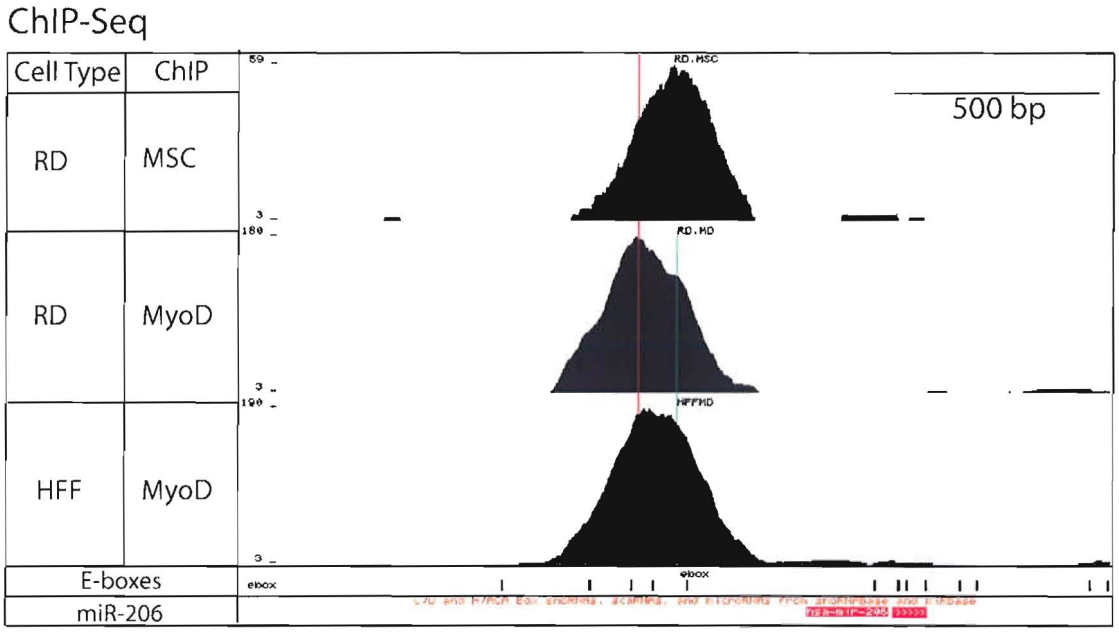


Figure 2.12. MyoD and MSC occupy distinct E-boxes in the miR-206 promoter. Screenshot from the human UCSC Genome Browser of the region that corresponds to the miR-206 promoter region. Mapped reads from ChIP-Seq for MyoD in RD and HFF cells are indicated, with the number on the left-hand y-axis indicating the number of reads mapped at the peak of occupancy. The location of E-boxes are indicated at the bottom of the panel by the black rectangles. Vertical lines are drawn through the apparent highest points of occupancy for MyoD and MSC in RD cells.

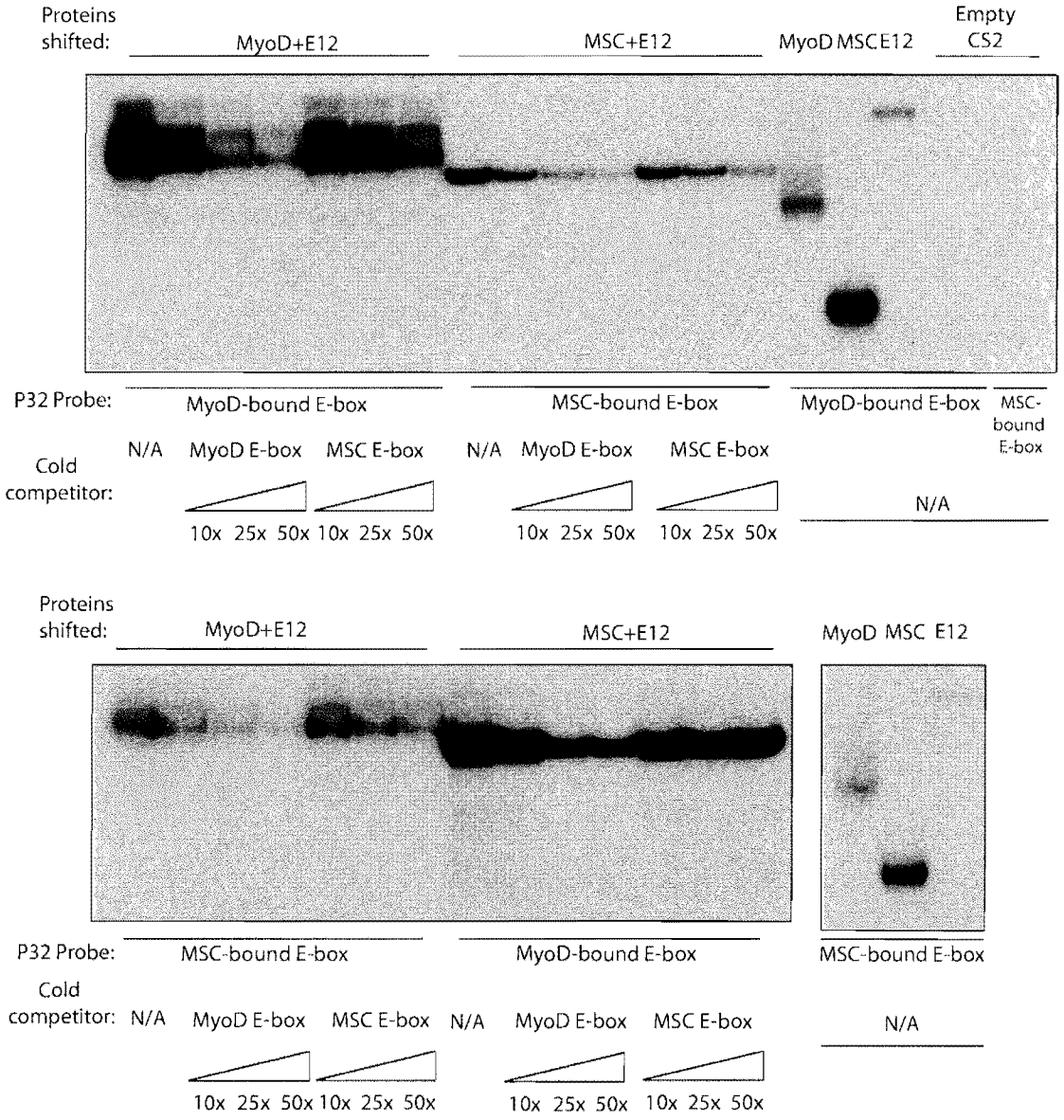


Figure 2.13. *In vitro* assessment of MyoD and MSC binding in the miR-206 promoter. Electrophoretic mobility shift assays were performed using proteins as indicated and probes that represented the DNA sequence under either the E-box occupied most strongly by MyoD in RD cells (MyoD-bound E-box), or the E-box under the peak of MSC occupancy (MSC-bound E-box). Bound complexes were competed with cold competitor probes prepared at the indicated excess.

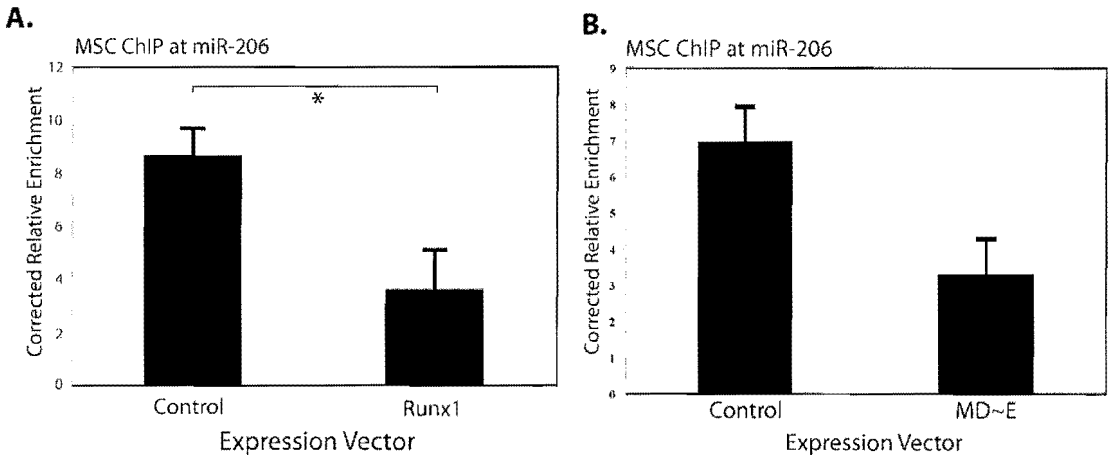


Fig 2.14. Differentiation of RD cells results in reduced MSC occupancy at the miR-206 promoter. A) Site-specific MSC ChIP in RD cells infected either with an empty retrovirus (Control), or RD cells differentiated through the action of RUNX1 (Runx1). Values represent the mean ± SEM from at least 3 independent replicates. B) ChIP for MSC at the miR-206 promoter in RD cells either treated with empty virus (control), or differentiated through the expression of the forced MyoD~E protein dimer (MD~E). Values are the average ± Std Dev of 2 independent experiments. Corrected relative enrichment equals relative enrichment at miR-206/relative enrichment at the control locus. *: $p<0.05$.

Chapter 3: Genome-wide binding of myogenic bHLH factors in human myogenic cells and rhabdomyosarcomas

Summary

Rhabdomyosarcomas (RMS) are a pediatric tumor of skeletal muscle that express the myogenic bHLH protein MyoD but fail to undergo terminal differentiation. Previous experiments have determined that the DNA binding of MyoD in RMS occurs, but that there is a defect in its ability to activate myogenic targets. bHLH inhibitors of myogenesis have also been identified in RMS, but the relation of their binding to that of MyoD is unknown. Here, we use chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-Seq) to demonstrate that both MyoD and the bHLH inhibitor MSC bind widely throughout the genome of RMS cells, at both overlapping and unique sites, an effect driven partially by binding site sequence. Further, comparison of MyoD bound sites between RMS and normal human myotubes demonstrates differences in bound sites at a subset of locations with potential functional implications. One of these differences indicates that DNA binding of the transcription factor RUNX1 associated with MyoD binding is underrepresented in RMS - a finding in agreement with our demonstration that expression of RUNX1 differentiates RMS cells.

Introduction

High-throughput sequencing coupled to chromatin immunoprecipitation (ChIP-Seq) of transcription factors permits genome-wide assessment of DNA binding by the factors and the possibility to draw new conclusions about their functionality. While some factors have been found to bind in a restricted pattern that suggests a fairly direct correlation with gene activation, others have been found to bind in a surprisingly promiscuous fashion, exhibiting binding throughout a substantial proportion of the mappable genome (MacQuarrie et al., 2011). Such findings suggest that, for certain factors, it could be important to re-consider data on DNA binding from older experiments that did not have high-throughput approaches available.

We have recently performed ChIP-Seq for the myogenic regulatory factor (MRF) MyoD in murine cells of the skeletal muscle lineage, and described widespread binding of MyoD both in intra- and intergenic regions of the genome (Cao et al., 2010). MyoD is a

member of the basic helix-loop-helix (bHLH) family of transcription factors, a large group of factors that all possess a basic region that permits DNA binding and amphipathic helices that permit dimerization with other bHLH family members (Lassar et al., 1989; Murre et al., 1989). In myogenic cells, MyoD heterodimerizes with members of the E-protein family of bHLH proteins, binds DNA in a sequence specific fashion, and leads to target transactivation (Lassar et al., 1991). We found MyoD bound extensively both in undifferentiated, proliferating myoblasts, as well as in terminally differentiated myotubes. Genes that had increased expression with differentiation were associated with MyoD ChIP-Seq peaks that increased during differentiation, and genes that decreased expression were associated with decreasing MyoD peaks. Analysis of the neighboring areas to MyoD bound sites revealed potential binding sites for a variety of other factors that are known or believed to play roles during myogenesis (eg. Ap1, Meis, Runx, Sp1).

Rhabdomyosarcoma (RMS) is a pediatric tumor of skeletal muscle that resembles cells of the myogenic lineage both at the level of molecular markers, as well as morphology (Merlino and Helman, 1999; Sebire and Malone, 2003). Puzzlingly, the tumors routinely express MyoD, even though expression of that factor is normally sufficient to cause terminal differentiation of the skeletal muscle cells it is expressed in. Previous work identified no defect in the ability of MyoD in RMS to bind to DNA, but rather in its ability to activate myogenic target genes (Tapscott et al., 1993), but the binding of MyoD in these tumors has never been investigated in a genome-wide fashion.

We have also recently identified multiple transcriptional inhibitors of myogenesis as being present in RMS, including the bHLH factor MSC (or MyoR) (Yang et al., 2009). MSC has been shown to also heterodimerize with E-proteins and bind E-boxes *in vitro* (Lu et al., 1999), and in RMS, it inhibits myogenic activity. Numerous bHLH myogenic inhibitors have been described (reviewed in (Berkes and Tapscott, 2005)), but the extent of their binding throughout the genome and their relation to MyoD binding has not been explored directly. While it is possible that individual inhibitors could bind at a small subset of sites bound by MyoD and have a relatively direct relationship between binding and inhibitory

function, it is also possible that, as with MyoD, inhibitors could bind at an excess of sites compared to those they act at directly.

To address these two questions, the relation of MyoD binding in RMS to normal myogenic cells, as well as a direct comparison of a myogenic bHLH inhibitor to MyoD, we have performed ChIP-Seq for 1) MyoD in an embryonal cell culture model of RMS, RD cells; 2) MyoD in primary human myoblasts and myotubes; 3) MyoD in human fibroblasts converted to differentiated myotubes through viral expression of MyoD; and 4) MSC in RD cells. We have found that MyoD binds widely throughout the genome of RD cells and, surprisingly, MSC binds at a comparable number of sites. MyoD and MSC exhibit both unique and overlapping binding sites, a pattern partially driven by sequence preference, suggesting a complicated functional interplay between the two factors. Comparison of MyoD binding between RD cells and normal human myotubes identifies many shared binding sites, but specific differences with potential functional implications. These differences include a different proportion of MyoD-favored E-boxes bound by MyoD in each cell type, as well as differences in the motifs for potential cooperative myogenic factors adjacent to MyoD-occupied sites. One of the binding motifs identified is that of RUNX1, a factor capable of differentiating RMS cells when expressed exogenously in them, demonstrating the potential therapeutic implications of these differences between cell types.

Results

MSC binding in RD cells

ChIP-Seq of endogenous MSC and endogenous MyoD in RD cells was performed to explore the role of MSC in the tumor cells, as well as to compare the binding profile of one of the many myogenic bHLH inhibitors to MyoD itself. Somewhat surprisingly, MSC binds at a comparable number of sites throughout the genome as does MyoD (**Table 3.1**). Biologically independent site-specific ChIPs were performed and confirmed a selection of MyoD and MSC-specific bound locations identified by the ChIP-Seq as being specifically enriched by the appropriate antibody (**Fig 3.1).**

Motif analysis of all MSC-bound sites in the genome identifies a preferred binding site very similar to MyoD, with MSC preferring a GC E-box with an additional 3' flanking 'G', giving a binding site of CAGCTGG (**Fig 3.2A**). Overlap analysis of p-value ranked sites was performed to compare MyoD to MSC in RD cells, and identified a surprisingly high overlap, suggesting that MSC binds at many of the same sites throughout the genome as MyoD (**Fig 3.2B**). Examination of screenshots from the ChIP-Seq data demonstrates that there are sites bound only by MyoD or by MSC, sites that bind both factors in an apparently identical pattern, as well as closely overlapping binding patterns (**Fig 3.2C**). Bound sites specific for MSC are enriched for GC E-boxes with specific flanking nucleotides, giving a preference for CCAGCTGG (**Fig 3.2D**), a binding site that, given its completely palindromic nature, may possibly indicate a homodimer binding site. Electrophoretic mobility shift assays confirmed the ChIP-Seq motifs, with MSC binding to the CCAGCTGG site strongly, either as a homo- or heterodimer, and MyoD binding poorly (**Fig 3.3, lanes 1 - 4**). Inversion of the flanking nucleotides to give GCAGCTGC resulted in a restoration of MyoD binding, while still permitting MSC binding, though possibly at a slightly reduced level (**Fig 3.3, compare lanes 4 and 8 and 3 and 7**).

The overlap in binding between MyoD and MSC is substantial, but the distribution of bound locations shows some differences. All the sites bound by MyoD and MSC in RD cells were grouped into categories based on their relationship to annotated genes. While the distribution of MyoD and MSC peaks in introns, intergenic areas, the 3' end of genes and areas farther up- or downstream of gene bodies appear basically identical for MyoD and MSC, MSC peaks are more abundant in promoters and the exons of gene bodies (**Fig 3.4**).

We have previously reported the use of tandem affinity purification coupled to LC-MS/MS (liquid chromatography – tandem mass spectrometry) to identify the protein complexes that MyoD and E-proteins form in RD cells (Yang et al., 2009), and the same technique was performed in RD cells using tagged MSC. Tagged MSC was tested for functionality and binding and was found to both repress myogenic reporters and bind E-boxes *in vitro* with a comparable pattern to untagged MSC (**Fig 3.5**). LC-MS/MS identified a relatively small number of proteins as associating specifically with MSC (**Table 3.2**). MSC

associates with all of the E-proteins with a high coverage rate of their sequences and, in agreement with our findings with tagged MyoD, does not appear to associate directly with MyoD, offering further evidence that our ChIP-Seq data represents distinct MyoD- and MSC-containing complexes bound to DNA.

MyoD binding in human myoblasts and myotubes and human rhabdomyosarcoma cells

ChIP-Seq was performed for endogenous MyoD in 1) the embryonal cell culture model of RMS, RD cells; 2) primary human myoblasts and myotubes; and 3) human foreskin fibroblasts (HFFs) converted to myotubes via the expression of MyoD from a lentivirus, using polyclonal antibodies specific to MyoD that have been previously described (Tapscott et al., 1988). In agreement with our findings in normal myogenic cells of murine origin, MyoD peaks, indicating bound MyoD, were at thousands of locations throughout the human genome in both myoblasts and myotubes, either those created from fibroblasts, or from primary cells. Also, as would be predicted from the aforementioned finding that the DNA binding of MyoD in rhabdomyosarcoma cells is not impaired, a comparable number of MyoD-bound sites were found in RD cells as in the primary human myoblasts and myotubes at a variety of p-value cutoffs (**Table 3.1**). HFFs had considerably more peaks at all cutoffs compared to any other cell type, presumably due to the overexpression of the lentivirally expressed MyoD binding to low-affinity MyoD sites in significant quantities (Y. Cao, Z. Yao, unpublished observations). In all cases, MyoD binds throughout the genome, with comparable proportions bound in all genomic areas examined (eg. promoter, introns, intergenic) (**Fig 3.6**).

The p-value ranking of MyoD-bound sites was used to rank sites and then examine the extent of overlap in specific MyoD peaks between different cell types and conditions, therefore identifying the extent to which MyoD binds at identical locations in the two compared conditions. Occupied sites in HFF+MyoD cells showed considerable overlap with occupied sites in myotubes from primary cells (**Fig 3.7A**), with myotubes of either type showing a somewhat reduced overlap with occupied sites in primary myoblasts (**Fig 3.7B**). Comparison of the overlap between RD cells and myoblasts and myotubes showed that in the primary human cells, more MyoD-bound sites were shared between myotubes and RDs than

myoblasts and RDs, and the overlap between HFF+MyoD cells and RDs was comparable to that seen with the primary myotubes (**Fig 3.7C**). We have previously proposed that RMS represent an arrested transitional state between myoblasts and myotubes, with expression of certain cellular factors reminiscent of normal cells going through that transition, and the similarity of the MyoD binding pattern to the myotube binding supports that model.

MyoD binding in RMS cells and primary human cells differs at a subset of sites with potential functional implications

Given the overall high degree of similarity in bound MyoD locations between RMS and non-tumor cells, and the finding that RUNX1 and RP58 can differentiate RMS, we hypothesized that comparison of MyoD peaks between myotubes and RD cells would reveal subtler, but potentially functionally important, differences in bound sites and potential co-factors. To address this issue, we first examined the sequence specificity of the E-boxes bound by MyoD in RDs and in myotubes. We have previously shown that the majority of MyoD in blasts and tubes in murine cells are bound at E-boxes with either a central dinucleotide of GC or GG (Cao et al., 2010), translating to E-boxes with either the sequence CA**GC**TG or CA**GG**TG. In the comparison in human cells, RDs exhibit a shift in those specific E-boxes relative to the primary human myotubes, with a relatively higher proportion of GG E-boxes and a relatively lower proportion of GC E-boxes (**Fig 3.8A**). Since MyoD binds in a heterodimer with E-proteins to E-boxes, but the relative affinity of all the possible MyoD:E heterodimers for specific E-boxes is unknown, this suggests that there may be a difference in one or more MyoD:E heterodimers in RMS. RT-PCR for each of the E-proteins, as well as some specific isoforms of them, demonstrates that myotubes express dramatically different levels of some E-proteins compared to RD cells (**Fig 3.8B**).

To address the question of whether RMS exhibit a difference in factors that cooperate to regulate myogenesis, we performed a motif analyses to look for positively or negatively enriched sequence motifs adjacent to MyoD peaks. The analysis compared the MyoD peaks found specifically in primary human myotubes to those in RD cells, to determine if any motifs would be found that could explain the difference between the cells. Interestingly, the analysis identified the binding site for the runt-related transcription factor RUNX1 as

enriched adjacent to primary myotube-specific peaks, a finding in agreement with the data presented in Chapter 2 (**Fig 3.8C, row 1**). The analysis also found potential binding sites for a nuclear factor (NFIC) that been proposed to cooperate with bHLH proteins in myogenic cells (Hebert et al., 2007) (**Fig 3.8C, row 3**), and the binding site for a JUN protein (**Fig 3.8C, row 2**). In agreement with the results of the E-box analysis mentioned above, the motif analysis identified a depletion of GG core E-boxes in RD cells relative to myotubes (**Fig 3.8C, row** 7).

Discussion

MSC was initially described as a myogenic inhibitor with the ability to bind E-boxes and repress both myogenic reporters and the process of MyoD-mediated myogenic conversion (Lu et al., 1999). MSC knock-out mice were found to have no discernable phenotype however, until crossed with knock-outs for the MSC homolog, capsulin, leaving it an open question on how many gene targets are strongly affected by MSC activity (Lu et al., 2002). In the nervous system, another transcriptional repressor, termed NRSF, was found by ChIP-Seq to be bound at a more restricted number of sites (~2000), and seems to have a relatively direct relationship between binding and activity (Johnson et al., 2007). In stark contrast to that, MSC binds throughout the genome, and shares a surprisingly large number of bound sites with MyoD itself, suggesting that it is not the relatively simple situation of MSC binding at some subset of MyoD targets to regulate them. The results with the favored binding motifs suggests that sequence accounts for some of the difference in binding, especially MSC-specific sites, but the exact relationship between the ability to bind *in vitro* and the locations that are actually bound *in vivo* will likely require further investigation of the influences exerted by both chromatin accessibility and cooperative factors to be explained completely. Similarly, the greater presence of MSC peaks in promoters and exons of genes may have functional implications, but factors such as GC-content will need further investigation to determine if they account for the observed differences. While we initially identified MSC in the RD cells by LC-MS/MS using tagged E2A, the mass spectrometry results with tagged MSC indicate it associates with all of the E-proteins. Relative binding affinities and sequence preferences will need to be determined for the various MyoD-, MSC-

and E-protein dimer combinations to better understand the relationship between them and the functional implications.

The genome-wide binding of MyoD in RD cells agrees completely with our prior findings that MyoD activity is compromised in RMS, but its DNA binding itself is generally unaffected. Our findings with the ChIP-Seq now expand these conclusions in two respects: 1) they demonstrate that, as we have previously proposed, RMS cells appear to be perched on the verge of terminal differentiation, and 2) while MyoD binds widely throughout the genome, there are detectable differences in both the sequence of the E-boxes it binds and the potential adjacent binding sites for other factors. Such differences that would be likely to affect MyoD functionality are of special interest, as they may suggest ways to rescue the differentiation defect in RMS.

The fraction of bound MyoD across various genomic locations is almost indistinguishable between the RD cells, primary cells, and converted fibroblasts. The p-value ranked analysis of bound locations shows that, while the general pattern of binding is consistent, the similarity in bound sites is higher between the myotubes and RD cells than between myoblasts and RDs, as would be predicted by our model, and our findings with miR-206 in Chapter 2.

The differences in bound E-boxes and motif analysis suggest that there are more subtle defects or differences in MyoD binding between RMS cells and human myogenic cells that successfully differentiate. Modeling of bHLH complexes binding to E-boxes (A. Fong, P. Bradley, unpublished observations) has suggested that E-proteins bind to the 3' end of the E-box, which is the half of the E-box that demonstrates the difference in proportional occupation between RDs and myotubes (CAG**CTG** versus CAG**GTG**). Further experiments would need to be performed to determine if the differences in E-protein expression correlate with differences in their preferred E-box binding sequence when heterodimerized with MyoD. If so, it may be that expression of a single E-protein could shift the pattern of MyoD binding on a wide scale, and possibly lead to terminal differentiation.

The functional impact of the motif analysis is clearer than that of the E-box composition. The finding that the RUNX1 motif is underrepresented near RD specific MyoD peaks compared to primary myotubes is in perfect agreement with the finding that RUNX1 differentiates RMS cells when expressed in them. While the results in Chapter 2 found a role for RUNX1 in cooperating with MyoD at myogenic targets to increase their expression, it is possible that it also serves an additional role by altering MyoD binding, which could explain its presence near myotube-specific MyoD peaks. ChIP-Seq for RUNX1 in RD cells and myotubes could solve that question, as could MyoD ChIP-Seq in RD cells expressing RUNX1. The other motifs identified by the analysis, such as for NFIC, might also be of functional significance in RMS, possibly serving as additional mechanisms by which the cells can be driven to differentiation, and representing defective pathways in the tumors.

Materials and Methods

Chromatin immunoprecipitation

ChIP was carried out as described in Chapter 2. ChIP-Seq was performed as has been described previously (Cao et al., 2010), with antibodies as listed in Chapter 2. The primer sequences for the MyoD and MSC-specific sites checked for independent ChIP-Seq confirmation are as follows: A – gcttgatgatgcttgcagaa, cggagaggatcatgtaactgc; B – ctggtccctttcaggagaca, gccgtccatctaaaggtcaa; C – aatgacaagcactcgcacaa, atcgagaagttgcgtgcttt; D – atctggaatgccttctgtgg, attgcctaggaagggacaca; E – gcgacgagctccacatctac, aggatgcccatgactttgag; F – ctcaccatccgaccaagagt, ggggtcacgtgtgtatgaga. Real-time was performed using Sybrgreen, as in Chapter 2, and relative enrichment calculated as % of Input in samples with antibody/% of Input in samples with no antibody.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays were performed as described in Chapter 2. Probe sequences were as follows (forward probes only listed): Probe 1: CGGCCGA**CCAGCTGG**AGATCCT; Probe 2: CGGCCGA**GCAGCTGC**AGATCCT; Probe B1: GATCCCCCCAA**CACCTG**CTGCCTGA. Complexes were resolved on 6% polyacrylamide gels and exposed to radiographic film.

Motif analysis

Motif analysis was performed as in (Cao et al., 2010).

RT-PCR

RT-PCR was performed as in Chapter 2, on cDNA prepped from total RNA isolated using the RNeasy mini kit (Qiagen) from RD cells transduced with an empty retroviral construct and placed in differentiation media for 24 hours after selection or from human fibroblast cells expressing an estradiol-inducible MyoD (MyoD-ER), and placed in differentiation media with 10^{-7} M beta-estradiol for a period of 96 hours. Primers were as follows: E2-2: CCAACTTCTTTGGCAAGTGG, TCTCCATAGTTCCTGGACGG; HEBisoA+B: GACCAACTACACTGGGAAGCA, GGAAGGACTTGGTTGACCACT; HEBisoC: TGCTTATCCTGTCCCTGGAA, ATCTGAATTTGGGGATGGTG; E12: GTGACATCAACGAGGCCTTT, AGTTTGGTCTGGGGCTTCTC; E47: GAGGACGAGGAGAACACGTC, GACAGCACCTCGTCCGTACT; TIMM17b: GGAGCCTTCACTATGGGTGT, CACAGCATTGGCACTACCTC.

Protein purification and mass spectrometry

Five nearly confluent 24.5 cm x 24.5 cm tissue culture dishes of RD cells stably expressing nTAP-tagged MSC were scraped into PBS and spun down at 1100 RPM for 5 minutes. The cell pellet was resuspended in 5x volume of Buffer A (10 mM Hepes, 1.5 mM $MgCl_2$, 10 mM KCl, 500 uM DTT) with the addition of complete protease inhibitors (Roche) and placed on ice for 10 minutes. Cells were spun at 600 g for 10 minutes at 4° then resuspended in 2x volume Buffer A, then passed through 22G11/2 needles once, then 25G11/2 needles four times while being kept cold. Cells were spun at 15000g for 20 min and the upper layer of supernatant and cellular debris removed, taking care to preserve the nuclei at the bottom. The nuclei were resuspended in 1x volume of Buffer B (20 mM Hepes, 0.5 mM EDTA, 100 mM KCl, 10% glycerol, 2 mM DTT, 3 mM $CaCl_2$, 1.5 mM $MgCl_2$, 0.25 mM $NaOVO_3$, 10 mM NaF, 50 mM beta-glycerophosphate) with complete protease inhibitors. 2 ul of the suspension was quantitated by UV spec for DNA content, and the nuclei subjected to MNase digestion for 10 min at 37° (0.25 U of MNase for every 40 ug of

DNA). Cells were then placed on ice, and 20 ul of 0.5 M EDTA added for each 500 ul of lysate, and cells rocked at 4° for 1.5 to 2 hrs. Material was then spun at 14000 RPM for 15 min at 4° and then supernatant saved. Rabbit IgG beads equal to 1/10 of the volume of the supernatant were added and rocked in the cold for 1.5 to 2 hrs. Beads were spun down gently, then placed in a chromatograph column (Bio-rad) and washed with Buffer B 1x as the above composition and then 2x with Buffer B at 150 mM KCl. Beads were then washed 3x with TEV buffer (10 mM Hepes, pH 7.6, 150 mM KCl, 10% glycerol, 0.1% NP40, 0.5 mM EDTA, 1 mM DTT). TEV protease (Invitrogen) was resuspended in TEV buffer at 1U/uL then 100 U of TEV added to each column for every 50 ul of packed beads and the columns capped. After 4 hrs sitting at 4°, the cap was removed, the eluate collected and then 3x washes of 100 ul each with calmodulin binding buffer (10 mM Hepes-KOH pH 8.0, 150 mM NaCl, 1 mM MgOAc, 1 mM imidazole, 0.1% NP-40, 2 mM $CaCl_2$, 10 mM beta-mercaptoethanol) done over the beads and combined with the eluate. 1M CaCl2 was added at $1/250^{th}$ of the volume and mixed by inversion. 100 ul of calmodulin sepharose beads (Stratagene) were added and rocked in the cold for 1.5 hrs. Beads were spun down, placed on the chromatograph column, rinsed 2x with 1 mL of calmodulin binding buffer, then 2x with 1 mL of calmodulin rinsing buffer (50 mM Amm. Bicarb. pH 8.0, 75 mM NaCl, 1 mM MgOAc, 1 mM imidazole, 2 mM $CaCl_2$). Complexes were then eluted with a total of 6x 100 ul rinses of calmodulin elution buffer (50 mM Amm bicarb pH 8.0, 25 mM EGTA). Elution was then taken to the Fred Hutchinson Protein Core Facility for trypsinization and subjected to LC-MS/MS.

Table 3.1 **Number of identified ChIP-Seq peaks in human cells at specific p-values.**

	p-value		
ChIP	**1.00E-05**	**1.00E-07**	**1.00E-10**
RD.MyoD	52762	36267	25231
RD.MSC	57002	39868	26392
1° Myoblasts	44364	32313	23158
1° Myotubes	58449	43334	31630
HFF+MyoD	131203	111313	73284

Table 3.2 **Proteins identified by LC-MS/MS as associated with MSC in RD cells.**

Protein Symbol	# of Unique Peptides*	Percent of A.A.'s**
TDP43 (A4GUK4)	1	2
ARHGEF10	1	1
CERKL	1	2
EBF3	**2**	**6**
gi|14783413|hypothetical_prot	1	2
HIST1H2AE	2	22
HIST2H2BE	2	23
HMGA2	1	21
HNRPC	1	5
HNRPH1	1	4
HNRPM	2	4
HRNR	2	3
HSPA8	3	7
IPI00221261	1	2
KRT14	1	3
KRT2	5	10
LDB2	**1**	**3**
MATR3	1	1
OTTHUMP00000028832	2	2
PBX2	**1**	**7**
PRKDC	1	1
MSC (Q53XZ2)	**1**	**5**
RBMX	1	4
SNRPE	1	17
TCF12 (HEB)	**11**	**29**
TCF3 (E2A)	**21**	**40**
TCF4 (E2-2)	**13**	**34**

* # of Unique peptides lists the number of distinct peptides belonging to the indicated protein determined by the MS analysis to be present in the sample.
** Percent of A.A.'s indicates the percentage of the total number of amino acids in each identified protein that are actually detected in the MS data.
Bold entries indicate known transcription factors.

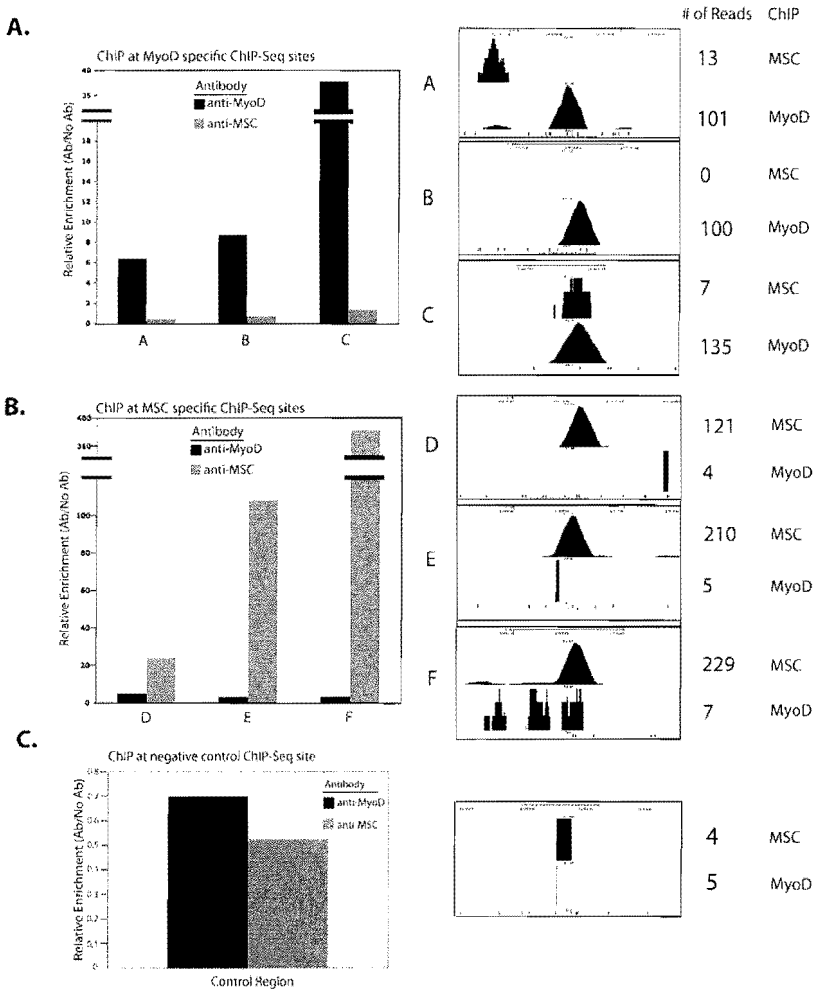


Figure 3.1. Validation of MyoD and MSC ChIP-Seq results. A) ChIPs on biologically independent RD samples were performed using anti-MyoD antibody and anti-MSC antibody as for the ChIP-Seq. Individual loci to be tested were chosen based on the ChIP-Seq identifying them as having a strong, MyoD-specific enrichment. B) ChIPs were performed for MSC and individual sites tested as for 1A, only with MSC-specific sites being chosen for this set of loci. C) ChIPs for MyoD and MSC were tested using primers for a gene that is not expressed in RD cells and should be essentially lacking any MyoD or MSC signal, according to the ChIP-Seq. All values are represented as the 'relative enrichment', which is calculated as % of Input with antibody/% of Input without antibody using qPCR.

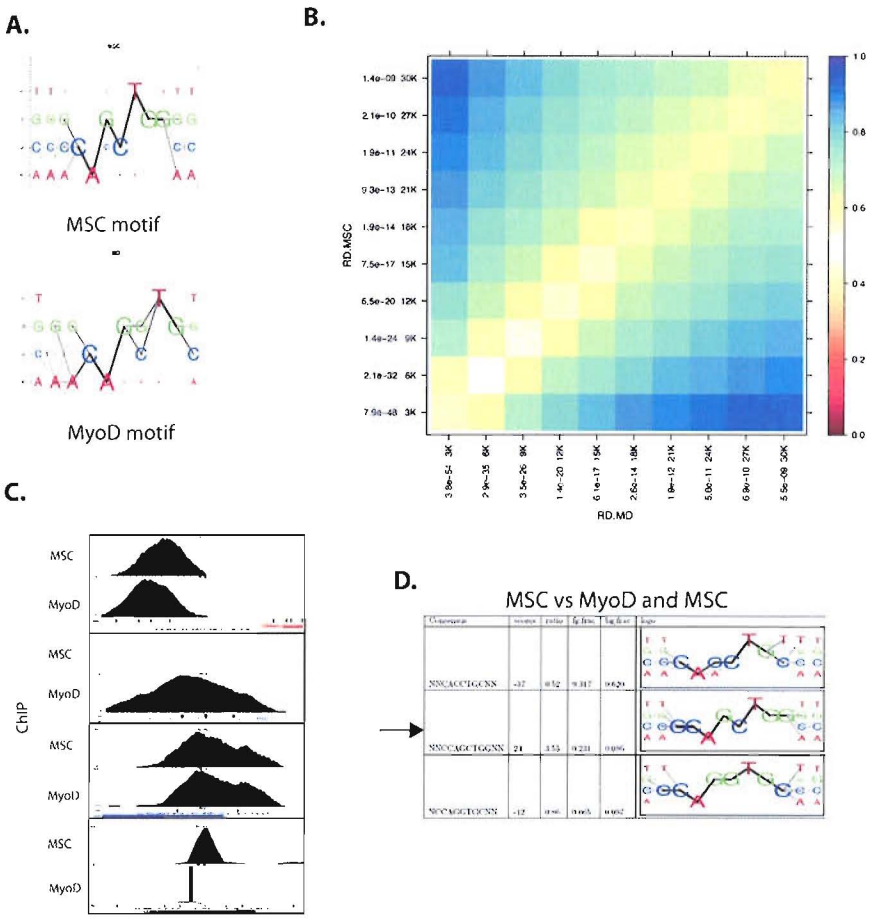


Figure 3.2. DNA binding characteristics of MSC in RD cells. A) The binding motifs of MSC and MyoD in RD cells show a preference in both instances for a central GC dinucleotide, but MSC favors an additional G at the +1 position compared to MyoD. B) MyoD and MSC ChIP-Seq peaks were ranked based on p-value and the proportion of identical peaks at various p-values determined and graphed as indicated. Colors approaching closer to blue indicate a higher proportion of identical peaks. C) Select UCSC Genome Browser screenshots from the MyoD and MSC ChIP-Seq data show genomic locations with either specific binding of one or the other factor, apparently identical binding by both factors, or slightly offset binding of the factors. The x-axis indicates DNA position, the y-axis indicates the number of sequencing reads found at those locations, and the position of E-boxes are represented by small black marks at the bottom of each panel. D) A DNA binding motif analysis similar to what was done in 2A, but this time comparing peaks found specifically in the MSC sample to peaks shared between MyoD and MSC.

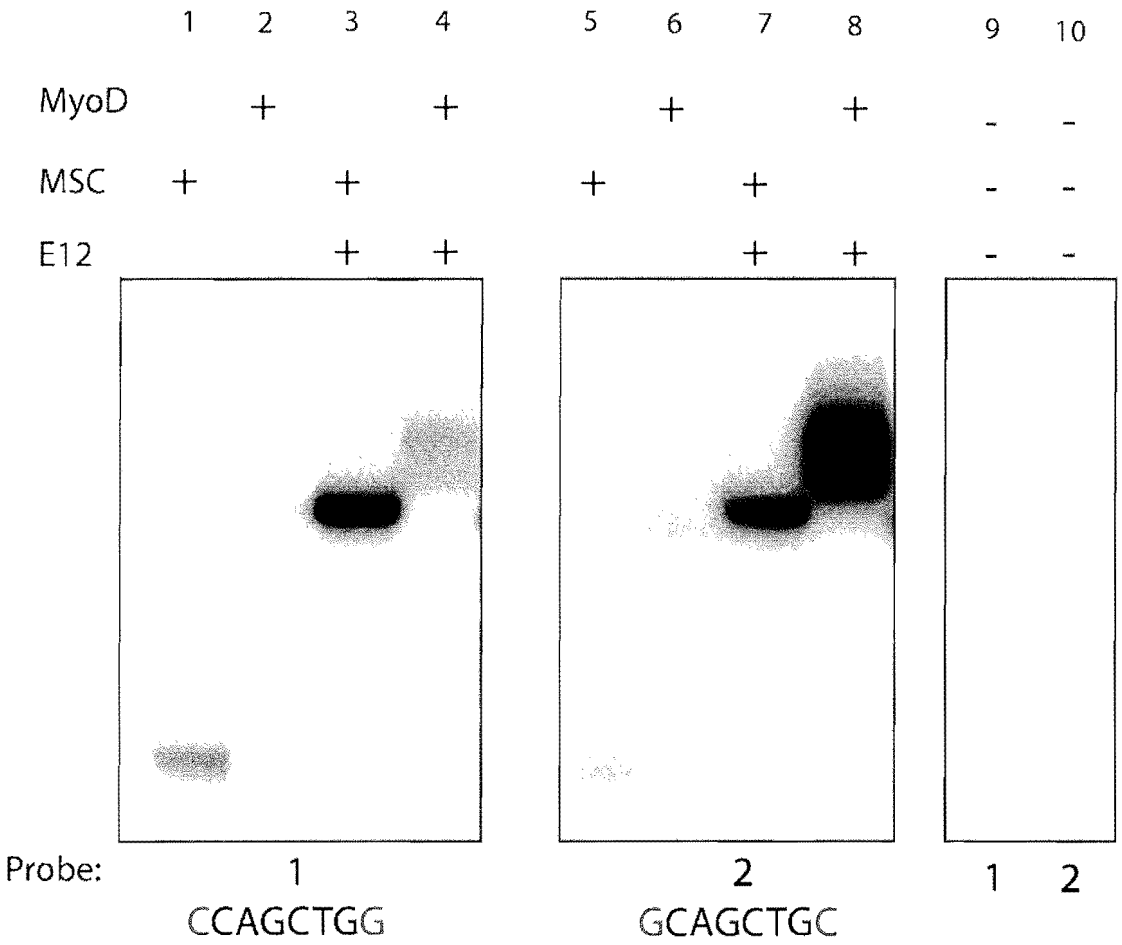


Figure 3.3. Binding preferences identified by ChIP-Seq are reflected in *in vitro* DNA binding assays. Electrophoretic mobility shift assays were performed to determine if the MSC specific binding site identified in Figure 2D correlated with any observable differences in DNA binding. *In vitro* translated proteins were mixed alone and together as indicated with either of two radioactive probes, differing only in which side the E-box-flanking C and G were present on.

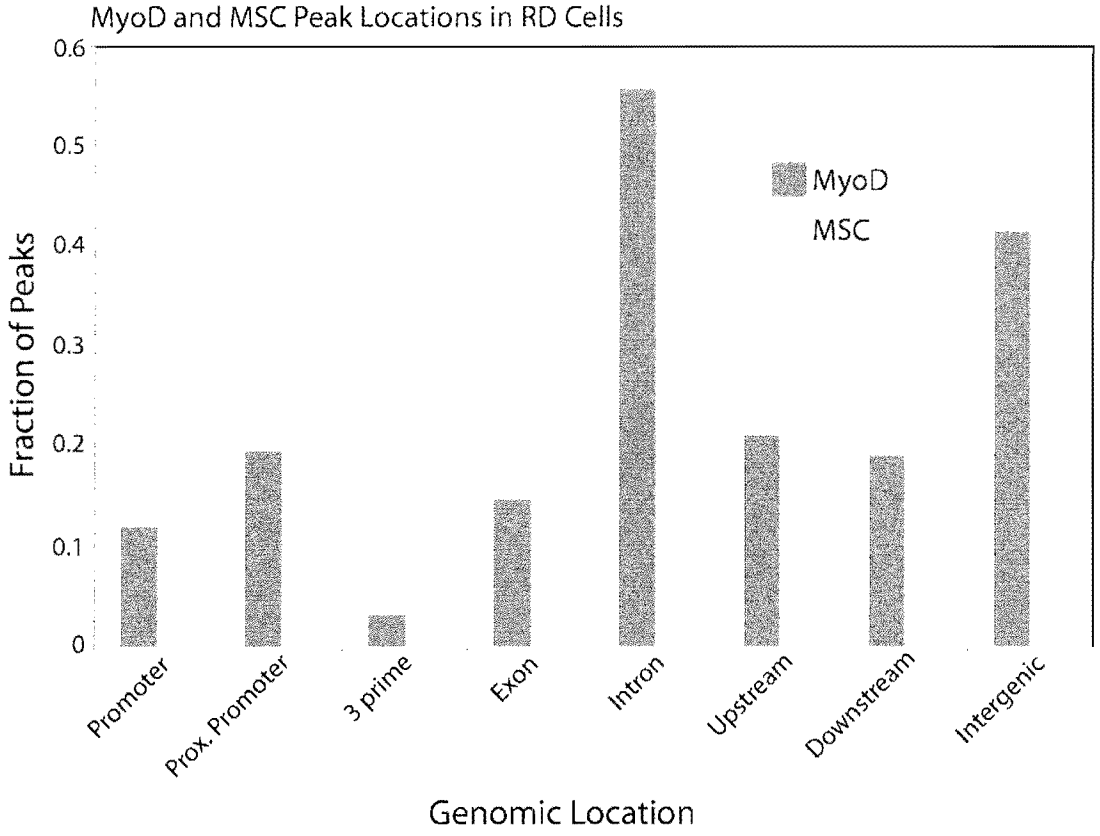


Figure 3.4. Genomic distribution of MyoD and MSC binding. All MyoD and MSC peaks in RD cells were examined to determine which of the indicated, non-mutually exclusive categories they fell under. The fraction of peaks that qualify as each category are graphed on the y-axis. Promoter: +/-500 bp from the transcription start site (TSS); Prox promoter: +/-2 kb from the TSS; 3 prime: +/-500 nt from the end of the transcript; upstream: -2 kb to -10 kb upstream of the TSS; downstream: +2 kb to +10 kb from the end of the transcript; intergenic: >10 kb from any annotated gene.

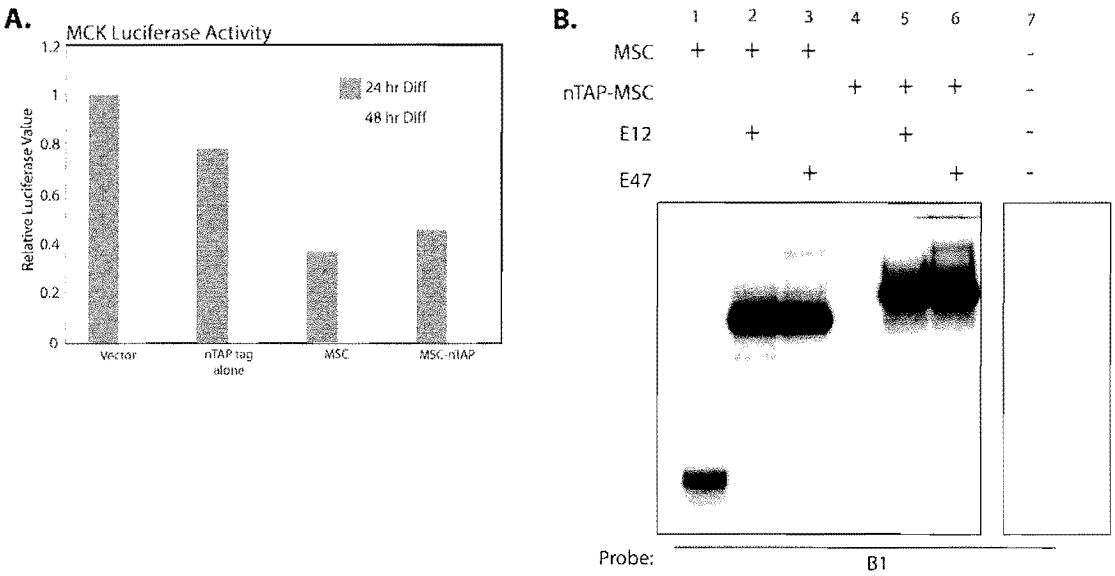


Figure 3.5. nTAP-tagged MSC functions like untagged MSC in functional assays and DNA binding of heterodimers. A) Luciferase assays were performed in C2C12 cells at two different time points after transfection with a muscle-specific creatine kinase (*MCK*) luciferase reporter, a beta-galactosidase internal control and constructs as indicated. Results are from a single experiment with each condition performed in duplicate. B) Electrophoretic mobility shift assay using proteins as indicated and an E-box containing radioactive probe. Note that nTAP-tagged MSC is larger than MSC and is expected to appear at a higher location on the exposure.

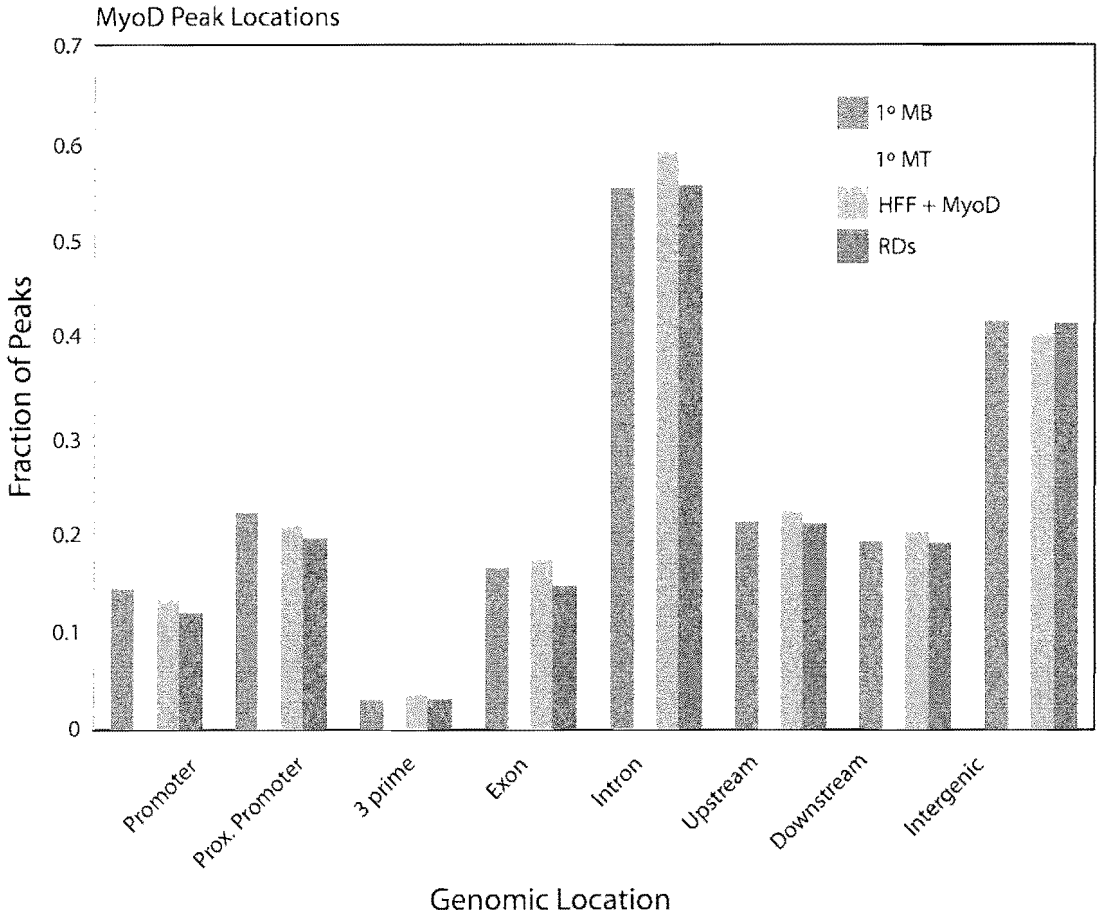


Figure 3.6. Genomic distribution of MyoD bound sites in RD cells compared to human myoblasts and myotubes. As in figure 3.4, the fraction of MyoD peaks in each indicated genomic region is graphed for primary myoblasts (1° MB), primary myotubes (1° MT), fibroblasts converted to myotubes by MyoD (HFF +MyoD), and RD cells (RDs).

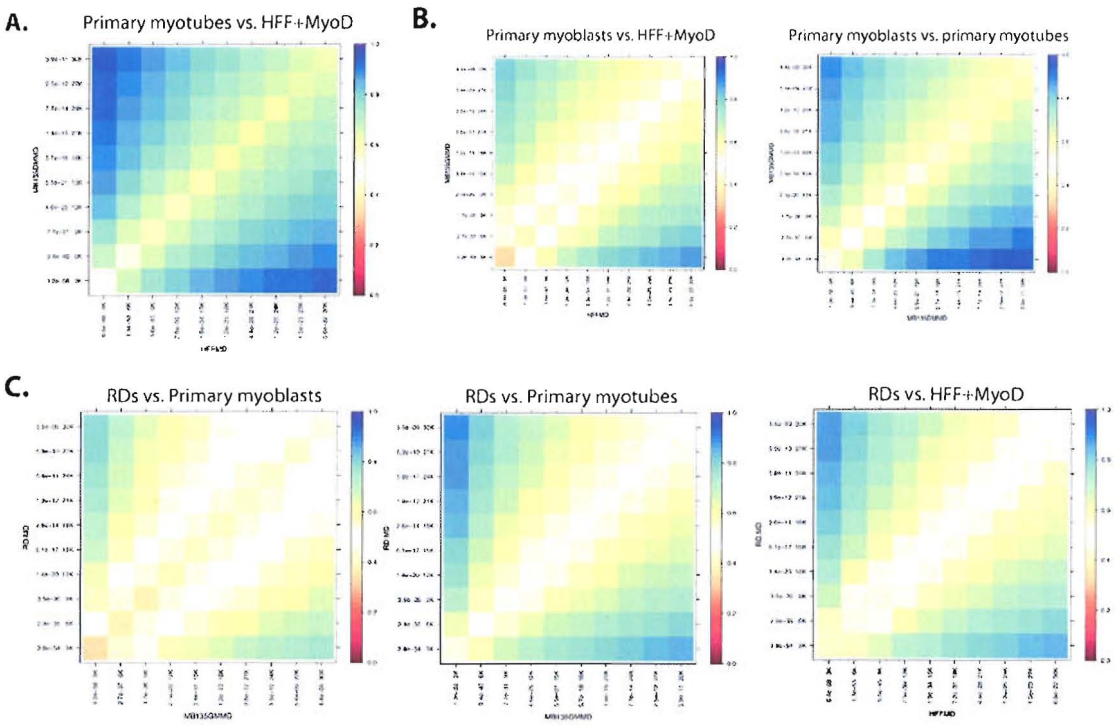
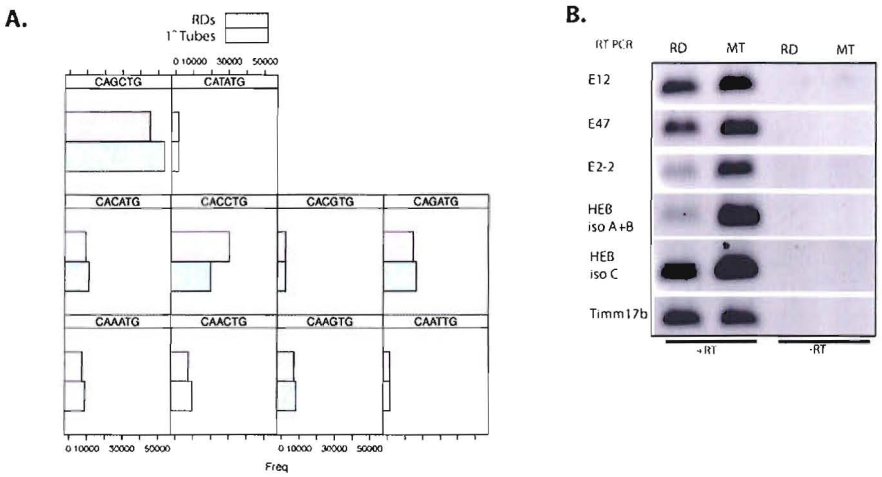


Figure 3.7. The sites bound by MyoD in RD cells overlap to a larger extent with sites bound by MyoD in myotubes than in myoblasts. A) MyoD peaks were ranked based on p-value, and then the proportion of identical peaks between human myotubes from primary samples and from MyoD-expressing fibroblasts. The percentage of identical peaks was calculated at various cut-off points by rank, as indicated by the y-axis values. As the proportion of identical peaks increases toward one, the color shades further to blue. B) The extent of identical MyoD peaks is visibly lower between either HFF+MyoD cells or primary human myotubes and the primary human myoblasts, regardless of peak p-value cut-off. C) The same analysis was performed as in A and B, but comparing MyoD peaks in RD cells to MyoD peaks in either of the two types of myotubes or in primary myoblasts.



C.

	Consensus	scores	ratio	fg.frac	bg.frac	logo	DB match	DB Evalue	DB logo
1	NNRCCACANN	21	1.71	0.45	0.29		RUNX1	2.5e-07	
2	NNTGANTCANN	20	2.49	0.25	0.12		Jundm2	6.6e-06	
3	NNTGCCAGNN	19	1.61	0.46	0.32		NFIC	5.2e-05	
4	NNHAAATANN	19	1.95	0.32	0.19				
5	NNTGVCCANN	16	1.50	0.47	0.35		Tcfap2a	0.00017	
6	NNAVCCAANN	15	1.60	0.33	0.22				
7	NNCAGGTGNN	-46	0.30	0.24	0.61		Tcfe2a	1.9e-06	
8	NNDCCCCCNN	-20	0.56	0.24	0.36		SP1	3.2e-05	
9	NNCCTCCHNN	-19	0.67	0.48	0.60				
10	NNCAGGARNN	-15	0.76	0.48	0.58				

Figure 3.8. Human myotubes have a subset of MyoD-bound sites that differ from the sites bound in RD cells with potential functional differences. A) MyoD peaks found specifically in primary human myotubes were compared to those found specifically in RD cells, and the composition of E-boxes under those peaks determined and quantified. CAGCTG and CACCTG are the favored MyoD bound E-boxes. B) RT-PCR for various E-proteins and specific isoforms of the E-proteins in RD cells cultured in low-serum differentiation media and human fibroblasts converted to myotubes through the action of an estradiol-inducible version of MyoD. HEB iso A+B and iso C refer to different isoforms of the E-protein, HEB. TIMM17b is an internal control. C) Primary human myotube specific MyoD peaks were compared to RD specific MyoD peaks and an analysis performed on the DNA immediately (± 100 basepairs) surrounding the peaks to determine over- and under-represented DNA motifs. Motifs were compared to a transcription factor database to determine what DNA binding factor likely accounts for the discovered motif (DB match, DB e-value, DB logo).

Chapter 4: Conceptual models of genome-wide transcription factor binding

This chapter has been published as: **MacQuarrie, K.L.***, Fong, A.P.*, Morse, R.H., and Tapscott, S.J. (2011) Genome-wide transcription factor binding: beyond direct target regulation. Trends Genet. *27(4),* 141-148.

*co-authors

Summary

The binding of transcription factors to specific DNA target sequences is the fundamental basis of gene regulatory networks. Chromatin immunoprecipitation combined with DNA tiling arrays or high-throughput sequencing—ChIP-chip and ChIP-Seq—has produced many recent studies that detail the binding sites of various transcription factors. Surprisingly, data from a variety of model organisms and tissues have demonstrated that transcription factors vary greatly in their number of genomic binding sites, and that binding events can significantly exceed the number of known or possible direct gene targets. Thus, our current understanding of transcription factor function must expand to encompass what role, if any, binding might play outside of direct transcriptional target regulation.

The finding, in both normal models of myogenic differentiation and in rhabdomyosarcoma cells, that MyoD binds at many thousands of places throughout the genome beyond its direct targets suggests that the models of transcription factor binding leading directly to nearby gene regulation are not adequate. We propose alternate possibilities that could account for the genome-wide binding observed for certain transcription factors, both myogenic and non-myogenic. The alternate, non-exclusive, roles that we discuss include action at a distance, large-scale chromatin remodeling, the site accessibility model, and the selective advantage model. The novel model we propose, the selective advantage model, hypothesizes that widespread genome binding of certain transcription factors offers an evolutionary advantage. If factors bind throughout the genome, then a single mutation in that factor can affect numerous downstream targets in a single step, rather than requiring iterative alterations or additions of binding sites. While the models proposed here are speculative, they give a framework in which to consider further large-scale data on transcription factor binding.

Regulatory networks and the core model of gene regulation

The complex interactions between multiple transcription factors and gene targets across various tissues, cellular contexts, and time points are termed 'transcriptional regulatory networks'. It has been stated that a truly thorough understanding of such interactions should theoretically explain how an organism is 'computed' from its DNA

(Weintraub, 1993). The core model of gene regulation posits that transcription factors recruit a polymerase complex to the transcriptional start site (Ptashne and Gann, 2002). Transcription factors initiate this by binding at nearby or distant DNA sequences and directly interacting with components of the polymerase complex or with complexes that indirectly mediate the polymerase interaction. In eukaryotes, the latter may include chromatin remodelers or modifiers that facilitate access or increase protein-protein affinities via histone modifications (Cosma, 2002; Fry and Peterson, 2001). The simplest view of the core model would suggest that factor binding directly correlates with transcriptional regulation. However, numerous examples of the separate regulation of factor binding and transcriptional activation suggest otherwise (Davis et al., 1990; Guarente et al., 1982; Turcotte and Guarente, 1992). For example, recent studies indicate that the sequence of the DNA binding site can induce conformational changes in the bound transcription factor that permits transcriptional regulation by subsets of a transcription factor family that can bind to similar sites (Leung et al., 2004; Meijsing et al., 2009).

Defining the relationship between transcription factor binding and target regulation across the entire genome of various species has become an attainable goal with the recent explosion in advanced computing and information processing tools. These advances have resulted in some remarkable progress in reconstructing and predicting regulatory networks (Lee et al., 2002). The advent of ChIP-chip (chromatin immunoprecipitation coupled to microarray hybridization) and ChIP-Seq (chromatin immunoprecipitation coupled to high-throughput sequencing) have now allowed for determination of the precise, genome-wide distribution of transcription factor binding sites. The results of numerous studies employing these techniques have been at times predictable and at other times surprising. While some studies have shown the expected correlation between factor binding and gene regulation, others have observed binding events that vastly exceed the number of expected gene targets (**Table 4.1**). Given these findings, it is timely to reconsider the relationship between transcription factors and gene regulation and the role, if any, that widespread transcription factor binding may play outside of direct gene target regulation.

Transcriptional Regulatory Networks

Transcription factors interact in a sequence-specific fashion with DNA to either increase or decrease transcription of gene targets. Transcription factors often bind and regulate multiple targets simultaneously, and targets, in turn, are frequently regulated by multiple factors. Regulatory networks can be constructed to describe these interactions, and represent the interactions that occur at multiple factor-target levels. Networks can be comprised of various motifs, which represent the regulatory approaches taken by one or more factors at specific targets. Multiple types of motifs have been described, but two common ones include the feed-forward loop and multi-input motif (**Fig 4.1**). Using these and other commonly found motifs (eg. auto-regulatory loops in which a gene product downregulates its own production), transcription factors are able to establish complex and dynamic mechanisms of gene regulation.

Transcription Factor Binding and Direct Gene Regulation

Several genome-wide transcription factor binding studies in various model organisms have supported a relatively direct connection between factor binding and gene regulation. One of the first genome-wide assessments of transcription factor binding in yeast reported transcription factor binding in promoter regions, in spite of the presence of binding motifs in both coding and intergenic regions (Lieb et al., 2001). Another report evaluating over 100 tagged factors in yeast identified more than 4,000 promoter-transcription factor interactions and described numerous regulatory circuits (Lee et al., 2002). The subset of circuits that comprised feed-forward networks (**Fig 4.1A**) alone was extensive, involving 39 factors, 49 distinct networks, and greater than 10% of all bound areas. This study emphasized both the importance of regulatory networks in controlling gene expression, as well as the ability of ChIP studies to uncover such networks.

A later study looking at an individual transcription factor in yeast, with roles in both filamentous growth and mating behavior, also found that DNA binding tightly correlated with function. Under cellular conditions that activated either growth or mating functions individually, the factor was found to occupy approximately 60 unique binding sites that were located in the promoters of genes with appropriate corresponding functions (Zeitlinger et al.,

2003). This binding was noted to be dependent on another transcription factor for the process of filamentation, an example of the importance of cooperative factor binding (**Fig 4.1B**) in mediating transcription factor activity.

The forkhead box A homolog *pha-4* regulates organogenesis of the pharynx in *C. elegans,* and provides an example of factor binding correlating closely with direct gene target effects in a multicellular organism. Initial studies demonstrated that expression of its targets correlated with PHA-4 binding sites in promoter regions, and that the timing of target expression correlated with binding affinity between transcription factor and its target sequence (Gaudet and Mango, 2002). Follow-up studies refined this model, providing evidence for other factors that cooperated with PHA-4 binding to modulate timing of target expression (Gaudet et al., 2004). Taken together, the data suggested that pharyngeal organ development is regulated by a combination of PHA-4 binding affinity and cooperating factors to temporally regulate gene expression. It also suggested that it should be possible to predict the time of expression of a putative *pha-4* target gene solely from analysis of its DNA sequence.

Recent ChIP-seq data for *pha-4* has been in agreement with this assessment. The great majority (>90%) of the bound sites identified in either embryos or larvae can be designated as 'gene-associated' using a distance cut-off of 2 kb or less between a bound site and nearest gene (Zhong et al., 2010). Overlapping the binding with gene expression data (high-throughput sequencing of RNA), most (87%) of the associated genes were expressed when PHA-4 binding was present, and a decrease in factor binding was associated with a reduction in expression for most (60%) presumptive targets, suggesting that binding of the factor activated the expression of those genes.

Studies in *Drosophila melanogaster* have identified the importance of cis-regulatory modules (CRMs), short DNA sequences (~300-500 nucleotides in length) that integrate multiple input signals to control gene expression. For example, the binding of Mef2, an important factor in mesodermal development, changes temporally during the course of muscle development (Sandmann et al., 2006). At the time points evaluated, different factor

motifs were noted at Mef2 binding regions, suggesting a cooperative factor mechanism used to temporally regulate the expression of various Mef2 targets. Further complexity in regulation is also suggested by a study comparing the binding profiles of *Mef2* and *lameduck* (*Lmd*) (Cunha et al., 2010). Mutants of *Mef2* and *Lmd* demonstrate a similar defect in myoblast fusion, suggesting similar or overlapping biological roles; however, while their DNA binding profiles overlap significantly, the effect of binding is widely variable. Depending on the enhancer target, co-binding can lead to additive, synergistic, or repressive effects, as demonstrated in reporter assays using eight different characterized enhancers. For example, co-expression of *Lmd* and *Mef2* activates the *blow* enhancer while expression of *Lmd* counteracts the positive effect of *Mef2* on the *CG9416* enhancer. While these results reveal the potential complexity of regulatory networks, a relatively direct relationship can still be inferred between DNA binding and target gene effects.

The close relationship between DNA binding and gene target effect has also been observed in mammalian systems. In one of the first studies to use ChIP-Seq, the binding of the zinc-finger protein neuron-restrictive silencer factor (NRSF) was mapped to only ~2000 sites in the human genome (Johnson et al., 2007). It was found that a few hundred potential target genes showed relatively 'low' gene expression compared to average cellular transcript expression when a NRSF peak was located nearby (≤1 kb), suggesting that NRSF was exerting its transcriptionally repressive effects at those genes when bound nearby. Studies of other factors, such as Pregnane X receptor (PXR) (Cui et al., 2010) and calcium-response factor (CaRF) (Pfenning et al., 2010), have also demonstrated a direct correlation of factor binding with gene regulation in mammalian cells.

Transcription Factor Binding in Excess of Known Direct Targets

In contrast to the model of direct gene regulation, several studies have demonstrated transcription factor binding at a large number of sites, many of which cannot be clearly connected with target gene regulation. In *Drosophila*, several ChIP-chip studies using whole genome tiling arrays have been performed for developmental transcription factors (Li et al., 2008; Zeitlinger et al., 2007). These studies have identified a large number of binding regions, on the order of several thousands, for individual factors in the developing embryo,

indicating a greater amount of DNA binding by developmental factors than had been anticipated. For example, over 2,000 binding regions were observed for Twist in the *Drosophila* genome in two separate studies utilizing distinct microarray designs (Sandmann et al., 2007; Zeitlinger et al., 2007), vastly exceeding the number of known Twist targets and including many intronic and intergenic sites. Also unexpectedly, Twist binding overlaps significantly with both Dorsal and Snail binding sites, and many of these sites possess highly conserved motifs. Their conservation suggests they are likely to be functional sites, but the significance of them is still unclear.

While widespread binding of early developmental transcription factors is perhaps not entirely surprising (Liang and Biggin, 1998), the unexpected finding has been the identification of numerous binding sites of unclear function, including for other factors as well. Studies of the binding and gene regulation of Myc and other proteins of the dMax family in *Drosophila* and human cells have shown extensive binding across the genome, but that binding did not necessarily correlate with transcriptional regulation of the nearby target genes (Fernandez et al., 2003; Orian et al., 2003).

In an early ChIP-seq study examining the interferon-γ (IFN-γ) responsive transcription factor *STAT1* in human cells, a strikingly large number of bound sites was observed (Robertson et al., 2007). In unstimulated cells, over 10,000 binding sites were identified, and this increased more than four-fold after stimulation with IFN-γ. In both conditions, approximately 50% of the total sites were intragenic and 25% intergenic. While there was a strong overlap with sites of known STAT1 activity, the majority of binding sites were not located adjacent to STAT1 regulated genes, suggesting that many, or most, bound sites were not directly regulating a nearby gene target. The authors suggested that many of the STAT1 sites might correspond to weaker, less favored binding sites, or possibly functional sites with STAT1 bound in only a subset of the total cell population.

As another example of widespread binding, the hematopoietic factor *GATA1* was reported to have over 15,000 DNA binding sites in a mouse erythroblast line (Cheng et al., 2009). GATA1-factor binding is apparently necessary for the binding of another

hematopoietic factor, the basic helix-loop-helix (bHLH) factor Tal1, to an adjacent E-box motif, the consensus binding site for bHLH factors. There is a strong association of Tal1 binding with erythroid gene regulation (Frankel et al., 2010; Kassouf et al., 2010; Palii et al., 2011), with over 2000 genes, most of which (90%) were categorized as related to erythroid development, having Tal1 binding within putative regulatory elements in one study, and over half of Tal1-regulated genes containing Tal1 bound within a proximal or distal regulatory element in another study (Palii et al., 2011). In this case, the widespread binding of GATA1 might be identifying the sites that can be bound by Tal1, and possibly other factors at different times or in different cells, to execute cell-type specific programs of gene expression.

The myogenic bHLH factor *MyoD* is another transcription factor that offers potential insight into genome-wide binding. MyoD directly regulates genes expressed during skeletal muscle differentiation (Bergstrom et al., 2002) and orchestrates a temporal pattern of gene expression through a feed-forward circuit (Penn et al., 2004). ChIP-seq on MyoD in skeletal muscle cells identified approximately 30,000-60,000 MyoD binding sites (Cao et al., 2010). As anticipated, genes regulated by MyoD during myogenesis had associated MyoD binding sites. However, almost 75% of all genes were associated with a MyoD binding site and about 25% of the MyoD sites were in intergenic regions. Therefore, the majority of MyoD binding events were not directly associated with gene regulation. Although regional transcription was not detected at these intergenic sites, MyoD binding was demonstrated to induce local chromatin modifications, specifically acetylation of histone H4 that is generally associated with active and/or accessible regions of the genome.

Together with the studies discussed above, these findings demonstrate that some transcription factors have binding events that are vastly in excess of the genes that they directly regulate. The remainder of this review will discuss the possible significance of these large number of transcription factor binding events that are not directly related to gene transcription. One proposed explanation for large-scale genome-wide transcription factor binding is the presence of 'non-functional' binding sites that serve no biological purpose (Li et al., 2008). Alternatively, it has been proposed that transcription factors may bind to many low affinity sites in the genome and contribute to gene expression at levels that are low but

sufficient to allow evolutionary conservation, an idea proposed from a large scale ChIP-chip study in yeast (Tanay, 2006). Presuming that these sites are functional, other possibilities include roles in affecting the functional concentration of factors, induction of chromatin looping, changing chromatin and nuclear structure, or the evolution of new transcriptional regulatory networks.

Site Accessibility Model

It has been suggested that binding sites occurring outside of areas directly involved in gene regulation may be 'non-specific,' or random. However, these intergenic sites contain the factor-specific binding motifs and have been validated both experimentally and statistically, the latter by passing very strict statistical cutoffs (Cao et al., 2010; Robertson et al., 2007). Thus, it seems more appropriate to conclude that the observed genome-wide binding of some transcription factors is a biologically specific event; however, the biological role at many of the sites remains largely undetermined.

Based on the binding of the *lac* repressor to bacterial DNA, it was suggested that genome-wide binding at non-regulatory sites might function to maintain an optimum amount of available transcription factor in the nucleus (Lin and Riggs, 1975). In this model, some of the transcription factor binding sites that are located in intergenic regions or repetitive elements might serve that function, helping to fine-tune gene expression by limiting the concentration of unbound factors and preventing binding to sites that need to be regulated by co-factor occupancy and cooperative binding. In this model, the genome-wide binding serves as a reservoir for factors, sequestering them in a manner analogous to other biological buffering systems.

Some studies provide support for this model. For example, in the *Drosophila* studies that show binding at thousands of sites in the genome in addition to binding at regulated genes (Li et al., 2008; MacArthur et al., 2009), higher-affinity binding occurred at regulated genes, and lower-affinity binding occurred in regions not regulated by the factors. This is consistent with the model that accessible DNA serves as a low-affinity reservoir for

transcription factors and that these sites are not directly regulating regional gene transcription.

Other studies provide additional support for the notion that transcription factors will bind to any available sites genome-wide. ChIP-seq of 15 transcription factors and regulators involved in mouse embryonic stem (ES) cell biology demonstrated binding for multiple factors at the same 3,583 sites in both promoter and intergenic regions (Chen et al., 2008). Similarly, in *Drosophila* several of the patterning factors exhibit notable overlap in their binding sites, although there is variability in the degree of overlap. And while analyses of binding site sequences demonstrate, in general, factor specificity for preferred DNA-binding motifs previously identified *in vitro*, many regions also exist which lack consensus binding motifs (Li et al., 2008). Therefore, some genome-wide binding might reflect factor interaction with accessible DNA regions that have not been specifically selected for a role in regional gene transcription.

Although likely correct in many instances, this model does not explain why there is an order of magnitude, or more, difference in genome-wide binding for factors with equivalently complex binding motifs. As noted above, MyoD has ~30,000-60,000 binding sites whereas Tal1 is reported to have ~3,000-6,000 sites in erythroid cells (Cao et al., 2010; Frankel et al., 2010; Kassouf et al., 2010; Palii et al., 2011). Both are bHLH factors that dimerize with an E-protein and recognize the core CANNTG E-box motif. The substantial difference in their genome-wide binding, however, suggests that sequence complexity is not the only determinant of binding. One possibility is that some factors are more constrained by site accessibility than others. MyoD can initiate chromatin remodeling at inaccessible sites and can bind independently of other factors, whereas the related bHLH factor Myogenin is more constrained to bind to accessible sites (Bergstrom and Tapscott, 2001; Cao et al., 2006; Cao et al., 2010; Penn et al., 2004) and the Tal1 bHLH factor might require GATA1 or other factors to bind (Palii et al., 2011). This suggests that the difference in the number of MyoD and Tal1 binding sites might, at least in part, reflect their relative ability to make new sites accessible for binding and to bind independently of other factors.

Chromosome Looping and Changes in Nuclear Architecture

Another, non-exclusive, model is that intergenic binding sites regulate gene transcription at a distance. Chromatin looping provides a mechanism for transcriptional control by bringing regulatory elements into proximity with target genes. Chromosome conformation capture studies indicate that the interaction of the distant locus control region (LCR) with the beta globin gene is required for high-level transcription. Interestingly, this interaction is dependent on GATA1 acting as an anchor (Vakoc et al., 2005). Given that GATA1 binds to over 15,000 sites, it is plausible that some proportion of these may effect transcription by inducing chromatin loops. In agreement with this idea, the LCR is necessary for globin genes to associate with transcriptionally-engaged PolII sites (Ragoczy et al., 2006), while other experiments demonstrated the association of hundreds of specific genomic loci with the murine globin genes in 'transcription factories' (Schoenfelder et al., 2010). In another specific example of chromatin looping leading to gene regulation, a Wnt-responsive enhancer downstream of the *Myc* gene has been shown to loop to cooperate with a 5' enhancer in a beta-catenin/TCF dependent fashion to regulate *Myc* expression (Yochum et al., 2010). These studies suggest that genome-wide binding might establish productive long-range interactions, either by looping to bring distant enhancers together with promoters, or in more complex interactions such as the co-regulation found in transcription factories.

Genome-wide Binding Affecting Global Chromatin and Nuclear Structure

As noted above, many of the MyoD binding events are not directly associated with regional gene transcription, but rather with regional histone modifications associated with active or accessible chromatin (Cao et al., 2010). Genome-wide changes in chromatin also occur in response to Myc binding (Knoepfler et al., 2006). Therefore, a major biological role of these factors, and perhaps other genome-wide binding factors, might not be to directly regulate transcription, but rather to re-organize the chromatin to make regions generally more accessible for factors expressed later in development. Such a role is supported by several studies of genome-wide influence on chromatin structure of general regulatory factors in yeast (Badis et al., 2008; Ganapathi et al., 2010; Hartley and Madhani, 2009).

Although it might seem unusual to suggest that some transcription factors have a role in regional chromatin organization at some sites and function as typical transcription factors at others, these represent two related functions of many transcription factors and it is reasonable to imagine that they can be deployed independently. For example, at genes transcriptionally regulated by MyoD, MyoD recruits histone acetyltransferases and chromatin remodeling complexes prior to mediating transcriptional iniitaition, which often occurs following the binding of an additional transcription factor (Aziz et al.; Penn et al., 2004; Tapscott, 2005). Therefore, the initial steps of transcription factor-mediated chromatin modifications can be distinguished from subsequent steps of transcriptional activation.

The suggestion that some transcription factors might have a role in regional chromatin organization that is independent of regional transcription is reminiscent of *CTCF*, which was originally identified as a transcription factor and is now recognized to have a broad role in chromatin organization. CTCF has also been found to have tens of thousands of binding sites in human and mouse cells (Chen et al., 2008; Kim et al., 2007). The greatest portion of CTCF sites were located in intergenic regions and many were at the border of distinct chromatin regions, consistent with a role in demarcating different chromatin domains (Barski et al., 2007; Kim et al., 2007). Furthermore, CTCF binding sites were flanked by arrays of well-positioned nucleosomes enriched in specific histone types (H2A.Z) and specific histone modifications, suggesting additional roles in broad changes in chromatin composition and structure (Fu et al., 2008).

Related to the model that some transcription factors might influence chromatin on a global scale is the idea that some of these factors might contribute to other aspects regional nuclear organization. Apart from its role in affecting chromatin structure, CTCF may also mediate long-range chromatin interactions (Hadjur et al., 2009; Mishiro et al., 2009). Also, as previously noted, both MyoD and Myc mediate broad epigenetic reprogramming within the nucleus, and it is reasonable to speculate that this activity might alter nuclear architecture and be important for their biological function. The ability to study changes in nuclear organization has recently become more accessible through the development of techniques

such as Hi-C (Lieberman-Aiden et al., 2009), and it will be interesting to determine whether the major role of some transcription factors is to re-organize the architecture of the nucleus.

Selective advantage model to explain widespread binding

The relationship between the feed-forward network motif and the evolution of new transcriptional regulatory networks is another theoretical model for understanding a potential biological role for genome-wide binding. Feed-forward regulation is the dominant motif for regulating complex biological pathways, with the ability to temporally regulate the expression of its targets while retaining the ability to rapidly cease target expression (Cordero and Hogeweg, 2006; Lee et al., 2002; Shen-Orr et al., 2002). Feed forward circuits have been found to occur repeatedly in *S. cerevisiae*, and have arisen via convergent evolution, suggesting their widespread utility (Conant and Wagner, 2003).

Genome-wide transcription factor binding and feed-forward mechanisms might have led to the evolution of distinct regulatory networks from a common network, a theory that can be understood using *MyoD* as an example. MyoD directly binds and regulates genes expressed throughout the program of skeletal myogenesis. At many targets, binding alone is not sufficient for transcriptional activation, but instead requires cooperation with factors that MyoD also regulates, thereby achieving temporal patterning through the feed-forward circuit. The evolution of a feed-forward circuit can be easily understood as the refinement of an initial single-input motif (**Fig 4.2**). For example, a primitive *MyoD*-like factor might have initially activated all the genes necessary for a primitive muscle cell phenotype, providing some selective advantage for this initial event. Subsequently, feed-forward regulation could be superimposed on the single-input motif to gradually improve and regulate the final output.

One prediction of this model is that factors with the potential to regulate complex transcriptional programs would bind throughout the genome because mutations in factors that sample a large portion of the genome would have the highest probability of generating a new network by changing the expression of large numbers of genes. Again using *MyoD* as an example, MyoD binds within a regulatory distance of more than one-half of all genes (Cao et al., 2010). Altering the activation potential of *MyoD* through a translocation or mutation

could drastically alter genome-wide transcription and potentially generate a novel complex phenotype from a single genetic event. In this model, genome-wide binding of a subset of transcription factors might reflect an evolutionary advantage rather than a cell-type specific function.

Comparing the findings from genome-wide transcription factor binding studies supports two general types of transcription factor binding. In some studies, the transcription factors tend to bind in the neighborhood of genes that they regulate, whereas in others the factors bind throughout the genome and relatively equivalently at both regulated and apparently non-regulated genes. A major caveat in suggesting that these might represent different biological strategies is the problem inherent to comparing results from different studies. Differences in sample preparation, data acquisition, and data processing can result in dramatically different conclusions that do not directly reflect the biology of the factors studied. Having acknowledged this important caveat, some factors appear to have binding profiles that reflect their regulatory network. For these factors it should be possible to infer their function based on knowledge of their binding sites, and, ultimately, it might be possible to compute their regulatory networks directly from knowledge of the organism's DNA sequence. The binding profiles of other factors appear much too dispersed across the genome to accurately correlate binding with regional transcription. For these factors, it might be impossible to infer their regulatory networks from DNA sequence, or even from knowledge of where they are physically bound. It remains to be determined whether these genome-wide binding events have one or more biological functions that are distinct from regulating regional transcription. Although speculative, this raises the intriguing possibility that the majority of binding events of some transcription factors might not be the direct regulation of transcription, but rather a currently unrecognized role in genome-wide biology.

Table 4.1. **Numbers of Transcription Factor Bound Sites from Select ChIP-chip and ChIP-Seq Experiments**

Transcription Factor	Species	Technique	Reported # of Bound Sites	Ref
Ste12	*S. cerevisiae*	ChIP-chip	65/57[a]	Zeitlinger, 2003
Pha-4	*C. elegans*	ChIP-Seq	4350/4808[b]	Zhong, 2010
Twist	*D. melanogaster*	ChIP-chip	2096	Sandmann, 2006
Twist	*D. melanogaster*	ChIP-chip	3000	Zeitlinger, 2007
NRSF	human	ChIP-Seq	1946	Johnson, 2007
Tal1	mouse	ChIP-Seq	2994	Kassouf, 2010
Tal1	human	ChIP-Seq	6315	Palii, 2010
PXR	mouse	ChIP-Seq	3812/6446[c]	Cui, 2010
CaRF	mouse	ChIP-Seq	176	Pfenning, 2010
STAT1	human	ChIP-Seq	11004/41582[d]	Robertson, 2007
GATA1	mouse	ChIP-Seq	15360	Cheng, 2009
CTCF	human	ChIP-chip	13804	Ganapathi, 2010
CTCF	human	ChIP-Seq	20262	Kim, 2007
CTCF	mouse	ChIP-Seq	39609	Chen, 2008
MyoD	mouse	ChIP-Seq	25956/59267[e]	Cao, 2010

a: Binding sites are those specifically identified in either mating or filamentous growth conditions, respectively.

b: Binding sites are in embryos, and L1 larvae, respectively.

c: Binding sites are listed for basal conditions, and conditions in which a synthetic activator of PXR was used, respectively.

d: Binding sites are listed for conditions of non-stimulated and interferon-γ-stimulated cells, respectively.

e: Binding sites are listed at two different statistical cutoffs (false discovery rates of 10^{-7} and 0.018, respectively)

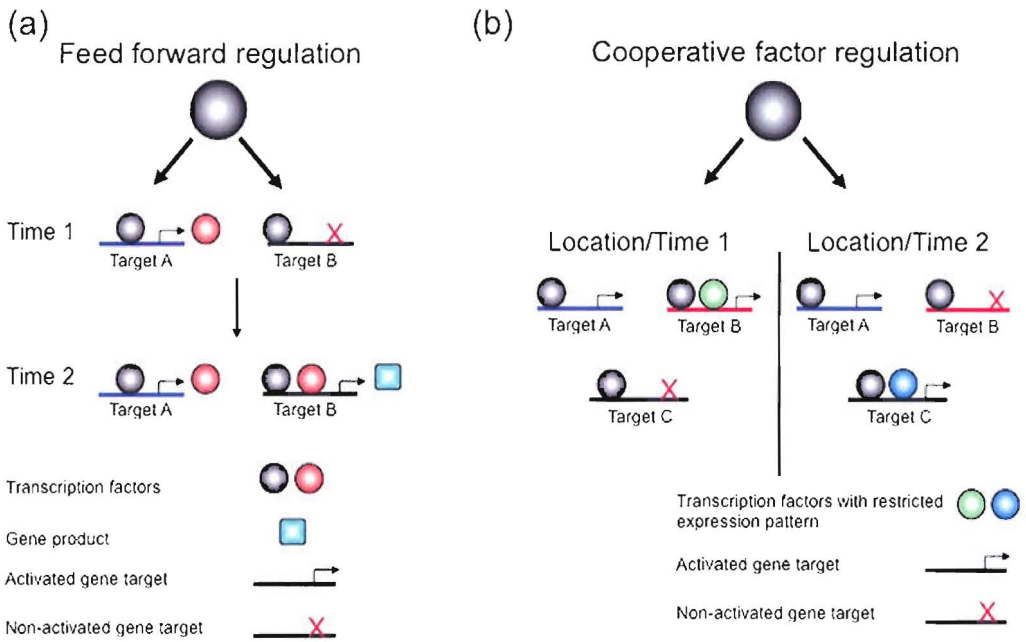


Figure 4.1. Examples of regulatory motifs used to control transcription. A variety of mechanisms, or regulatory motifs, are used to control the expression of specific gene targets over unique spacial (eg. specific tissue types) and/or temporal contexts. (A) Feed forward regulation permits temporal control of the targets of a single transcription factor. A transcription factor, represented by the grey circle, binds to multiple DNA targets (blue and black targets), but only activates one of them (top half of Figure 4.1a). The gene target that it activates (red circle) can then also bind to one of the same gene targets as the original factor (black), and together they activate transcription (bottom half of Figure 4.1a). (B) The use of cooperative factors permits transcription factors to be expressed widely, but discriminately activate gene targets. A single transcription factor, again represented by the grey circle, binds to multiple gene targets, activating one (the blue line) consistently, regardless of the cellular context (either tissue type or time). Other targets that it binds to in both cases (black and red targets), are activated only if they are also bound by another factor (compare activation of black and red targets between left and right side of Figure 4.1b), expressed specifically in that cellular condition.

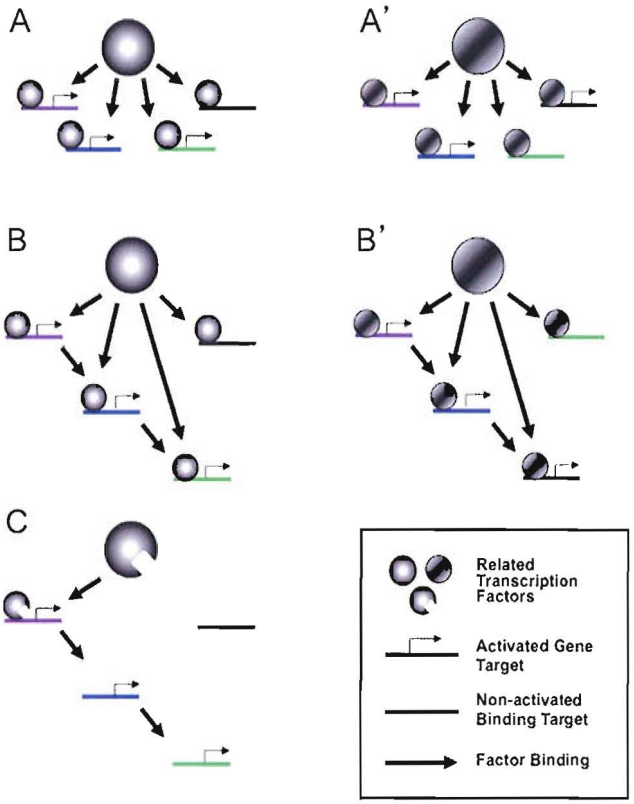


Figure 4.2. Genome-wide binding and the evolution of transcriptional networks. The ability of certain transcription factors to bind widely throughout the genome could permit the evolution of new transcriptional regulatory networks in a relatively limited number of events. This could mean that genome-wide binding might actually serve an evolutionary advantage in cells, permitting them to more easily acquire new networks and phenotypes, as a result of the different genes involved in those networks. (A) Schematic representation of a transcription factor that binds to many sites throughout the genome and regulates transcription at a subset of these sites in a single input motif, in which it alone regulates the expression of the targets at which it binds. (A') Duplication and sequence divergence of this factor can give rise to a family member with similar DNA binding characteristics but transcriptional regulation of an overlapping yet distinct set of genes. The more promiscuous the binding of factor A and A', the greater the subset of genes they have the potential to influence and the greater potential for target diversity between A and A'. Therefore, changing from A to A' could lead to the generation of a new complex program by a single factor modification. (B and B') If the cellular phenotype conferred by the set of genes regulated in A and A' have some selective advantage, then the single input motif can be refined by the gradual super-imposition of a feed-forward motif to achieve temporal regulation and more robust kinetics. (C) It is also possible for feed-forward motifs to degenerate into simple cascades of regulated genes over time if subsequent mutations in the original factor limit the set of genes that can be directly bound, further separating the two networks that originally came from a common progenitor.

Chapter 5: Discussion

miR-206 in rhabdomyosarcomas

The finding that miR-206 increases in response to the effects that MyoD~E, RUNX1, RP58 all have on RMS suggests that it is a key point of integration for the process of differentiation. It is somewhat surprising to consider that a single microRNA could be of such crucial importance, but there are still many unanswered questions about the function of miR-206 and the finer details of its effects. miR-206 is highly similar to the microRNA miR-1, which also leads to differentiation in myogenic cells, but they have different expression patterns in organisms and apparent differences in regulation as well (Callis et al., 2008). The miR-206/1 family is highly conserved, with diverse organisms down to zebrafish having an identical mature miR-206 sequence to that of human, offering further credence to the idea that it is a crucial part of the myogenic program.

It had previously been established that MyoD controls miR-206 expression, and I have now expanded that work to demonstrate an inhibitory role for both the bHLH protein MSC as well as an activating role for RUNX1. The expression of RP58 also leads to an increase in miR-206 levels, but there is no evidence for the effect being direct, suggesting that it likely is affecting miR-206 indirectly by its ability to alter bHLH dimer balance. Other groups have provided evidence that factors such as YY1 and AP1 also affect miR-206 expression (Song and Wang, 2009), further expanding the list of factors that appear to regulate its expression. If miR-206 expression is as crucial to myogenesis as its effects on RMS implies, having it be regulated by so many factors may be an evolutionary advantage, allowing it to respond in a carefully modulated fashion to a variety of effects on the cell. It is also worth noting that the evidence that YY1 negatively regulates miR-206 may also explain the aforementioned ability of the NF-κB/YY-1/miR-29b circuit to differentiate RMS. In my experiments, I could find no evidence that the forced protein dimer increased miR-29b levels, and it is possible that the effects on YY1 by the other groups actually led to a derepression of miR-206 expression, and an increase in miR-206 expression was the causative factor for the observed differentiation.

It is still unclear at this point what the precise mechanism is by which increased miR-206 expression leads to differentiation. It has been shown to directly and indirectly

downregulate a variety of targets (eg. a specific DNA polymerase subunit, MSC) in myogenic cells that make logical sense given its effects, but it is unknown if there is a single crucial target or if the effects of miR-206 are due to a cumulative effect on a variety of factors. Certainly, the finding that *NOTCH3*, and the Notch ligand *DLL3*, are downregulated in response to the lentiviral expression of miR-206 is of interest and warrants further study. NOTCH3 has been identified as a direct miR-206 target (Song et al., 2009), but the effect of specific interference with NOTCH3 in RMS or other myogenic cells is unknown. It is possible that interference with the inhibitory Notch pathway in the proper manner may result in RMS differentiation, but it also is possible that downregulation of the Notch pathway is permissive, but not sufficient, for differentiation in the cells. If so, it would be interesting to determine if Notch downregulation could potentiate manipulations like MSC siRNA and result in successful differentiation of the cells.

Another outstanding puzzle about the effect of miR-206 in RMS, as well as its role in myogenesis in general, is the relationship between it and miR-133b. In both my experiments, as well as previously published data, miR-206 and miR-133b appear to have opposing effects – one pushing differentiation, and the other interfering with that process, respectively. It appears paradoxical then, that they are processed from the same primary transcript and, in my experiments, both increase sharply in response to RMS differentiation. Given the results of my transient transfections of each microRNA into RD cells, it is clear that the pro-differentiation effects are specific to miR-206, as would be expected from all prior data. While it is not necessarily of direct bearing on RMS, this system might offer a tractable manner to begin to address the question of how miR-206 and -133b interact and the relative strength of their effects. It is possible that, while increased levels of miR-206 lead to differentiation, that its effect is actually being diminished or modulated by the increased levels of miR-133b. If so, it raises the possibility that interference with miR-133b activity, expression, or processing might be sufficient to cause RMS differentiation; my microRNA Northern blots clearly demonstrate that miR-206 is being expressed at easily detectable levels in 'normal' RD cells and, unopposed, it may be sufficient to differentiate the cells.

RUNX1 and RP58 in rhabdomyosarcomas and myogenesis

My findings with RUNX1 and RP58 place both factors downstream of MyoD in human cells, and as positive regulators of myogenesis, but there are still many unanswered questions about the specifics of their roles and effects. The findings with RUNX1 in Chapter 2 suggest that it functions directly at important downstream targets of MyoD, including miR-206 and RP58, to increase their expression. The arrays results suggest it may also function in a similar manner at the transcription factors *MYOG* and *MEF2C* and *MEF2D* as well, making it have a remarkably extensive effect on the later stages of myogenesis that are affected by such factors.

The ChIP-Seq data in Chapter 3 also points to RUNX1 playing an important role in differentiated myotubes, but in that analysis its binding site is found to be associated with MyoD binding sites found only in differentiated human myotubes. While the motif analysis does not explain causation, it could mean that RUNX1 is actually responsible for helping to recruit MyoD to bind at some subset of locations throughout the genome. Given the effect that expression of RUNX1 has on RMS, it would suggest that this subset of sites is of critical functional importance to myogenesis. It is possible, however, that there is another reason that MyoD cannot bind at those locations, and that the locations it is not binding at are ones that later require RUNX1 binding for activation. ChIP experiments for MyoD in RUNX1-differentiated cells would be able to address that question, determining if RUNX1 expression shifts MyoD to be bound at those sites. The effect that RUNX1 has on sites where MyoD is already bound (such as miR-206 and RP58) will also need to be investigated, to further understand how it assists MyoD in activating some of its targets.

RP58 has been described as directly downregulating *Id2* and *Id3*, a role that ties in perfectly with our model of a tipping point in RMS. By downregulating those two inhibitory factors, RP58 could enhance the association of productive MyoD:E-protein dimers, and increase myogenic activity in RMS. Indeed, the qPCR data after expression of RP58 suggests that *ID* downregulation is occurring in the RD cells. However, our ChIP-Seq MyoD data in murine cells also suggests another role for RP58 binding. The RP58 motif is associated with MyoD peaks that decrease during the process of differentiation – a decrease in occupancy that is associated with genes that decrease in expression. Since RP58 has a

binding motif that is a variant of a type of E-box (ACATCTG), this would suggest that RP58 binding of E-boxes might even be directly interfering with or displacing bound MyoD. Given that RP58 is clearly a downstream target of MyoD according to both my findings and those of other groups (Yokoyama et al., 2009), this would suggest that MyoD induces the expression of factors that then interfere with MyoD binding itself at some subset of targets to cause target downregulation. Certainly, this effect and the possibility suggested by my arrays that RP58 downregulates both positive cell cycle regulators and the *HES/HEY* family warrants further study of its direct effects, but my attempts to ChIP it have all failed. Two antibodies to RP58 as well as an antibody against a tagged version of the protein have all failed to enrich RP58 at either of the *ID* genes, and it appears that other reagents or approaches will be necessary to further investigate the role and action of RP58 in myogenesis.

Genome-wide binding of bHLH factors in rhabdomyosarcomas

The data reported in Chapter 3 that MyoD binds widely throughout the genome of RD cells, and at a comparable number of sites to MyoD in primary human cells, is in agreement with our previous findings that the DNA binding capabilities of MyoD are not compromised in RMS. The finding that is significantly more surprising is the widespread nature of MSC binding. While it is unknown how many genes MSC regulates, the results from the knockout mice make it reasonable to speculate that it would be a relatively small set of genes, and certainly smaller than the set of MyoD-controlled genes. Granted, as the ChIP-Seq with MyoD has demonstrated, and as discussed in Chapter 4, DNA binding does not have to correlate with function at an immediately adjacent target. Even given that, it is startling to think that a factor that affects some subset of MyoD activity, and that is not the only bHLH myogenic inhibitor that has been described, is bound at a comparable number of locations as MyoD throughout the genome.

Future work will need to more carefully define the targets of MSC activity and compare that to the ChIP-Seq data, to determine if there are defining or distinguishing characteristics to targets that would explain how MSC and MyoD coordinate their activities. The finding that MSC-specific peaks are found at E-boxes with a sequence that would be

consistent with a homodimer binding (CCAGCTGG) could be of particular functional significance. *In vitro* experiments and the mass spectrometry data make it clear that MSC associates with E-proteins in complexes, but its relative affinity for heterodimerization versus homodimerization, if MSC homodimerization is even something that occurs *in vivo*, is completely unknown. Experiments including ChIPs could begin to shed light not only on that question, but on the question of whether widespread binding at and near MyoD-bound sites is a common occurrence for any type of transcription factor that negatively regulates myogenesis, or something more unique to MSC or bHLH inhibitors specifically.

One of the most unexpected findings from the MyoD ChIP-Seq in RD cells is the finding that the proportion of 'GC' versus 'GG' E-boxes is different between RDs and primary human myotubes. The two sequences are the most highly bound E-boxes by MyoD, but ChIP-Seq with other cell types in our lab has demonstrated that the 'GG' E-box is more MyoD specific, while the 'GC' E-box is bound by other bHLH factors as well. The functional significance of this E-box shift is therefore unclear but, as mentioned in Chapter 3, the single nucleotide difference in the E-boxes appears to be on the side occupied by the E-protein. Given the difference in expression levels of HEB and E2-2 that I see between myotubes and RD cells, the E-box sequence preference of each E-protein when heterodimerized with MyoD needs to be determined – experiments that can be accomplished through EMSA. If one of the E-proteins that is noticeably underrepresented in RD cells is found to have a strong GC E-box preference, this may explain the observed difference in occupied E-boxes. In that case, the effect on RD cells of expression of that specific E-protein will need to be investigated to determine if it affects the state of myogenesis in the cells or leads to differentiation.

The arrested state of rhabdomyosarcomas

MyoD is a potent driver of the myogenic fate, and possesses a remarkable ability to drive a variety of cells to become terminally differentiated myotubes. The process of myogenesis has been described as being 'all-or-nothing' in nature, referring to the fact that when a cell undergoes the process of myogenesis, it proceeds fully to the state of myotube, without clearly defined transitional states or stages along the way. But the fact that such a

state has not been clearly defined or described does not necessarily imply that it does not exist.

The data described herein, both from myself and others from our group, suggests that rhabdomyosarcoma cells may represent or be trapped in just such a transitional state. While retaining a morphology more closely reminiscent of myoblasts than myotubes, and continuing to proliferate rapidly, RMS cells possess many characteristics of myogenic cells that are beginning to differentiate. From expressing later myogenic genes and factors such as *MYOG* and miR-206, to having a genome-wide binding pattern of MyoD that more closely resembles myotubes than myoblasts, to the multiple means that we have found are capable of driving differentiation in the cells, RMS appear to be on the brink of completing myogenesis even before experimental manipulations. The fact that the means I have found that differentiate the cells – RUNX1, RP58, and miR-206, in addition to the forced MyoD~E-protein dimer – all appear to play some role in the normal process of myogenesis implies that the regulatory relationships that exist in normal myogenesis are relatively preserved in RMS.

Studies of RMS biology then, offer the interesting benefit of not only understanding the biology of a type of tumor, but providing information about normal myogenesis at a stage that has generally been inaccessible in studies of normal myogenic cells. Experiments with most myogenic cell culture models or animal models tend to compare cells when they are as undifferentiated as is possible to myotubes at various time points. Isolating a population of cells that are, as a whole, at the same point in between those two ends is technically difficult, and can suffer from problems with inconsistent cell density or passaging in culture models. RMS offer the possibility of studying that otherwise problematic point in the process of differentiation with some consistency.

Our current model for the state of RMS (**Figure 1.1**) suggests that, as a result of the cells being on the verge of differentiation – at a 'tipping point' - and since multiple factors impact on their state, multiple points of manipulation exist that can result in differentiation. This effect should be possible not only by enhancing the activity of MyoD and myogenesis in general, but by interfering with the activity of the inhibitory factors as well. The factors that

I have found all appear to function by assisting, or functioning as a downstream effector, of MyoD, and it would be of great interest to explore what other inhibitors could be depleted or inhibited to drive differentiation. We have previously reported that siRNA-mediated knockdown of MSC was not sufficient to drive differentiation, but that expression of a DNA-binding dominant negative MSC resulted in the formation of myotubes in RD cells (Lee et al., 2011; Yang et al., 2009). Experiments by other groups that have affected myogenic inhibitors and caused RMS differentiation have used shRNA and dominant negative approaches (Lee et al., 2011; Sang et al., 2008), and the technical details of these experiments may be key to their success or failure. Certainly, given the widespread binding of MSC throughout the genome described in Chapter 3, transient transfection of siRNA constructs may simply be inadequate for depleting the DNA-bound MSC at sufficient levels to allow the myogenic balance to tip and differentiation to proceed. Approaches that have a longer-term effect and/or greater efficacy may be crucial for success when attempting to interfere with other myogenic inhibitors of interest in future experiments.

As mentioned above, inhibitory factors that affect the 'tipping point' in RMS include KMT1A and HES1. My findings with RP58 suggest that the ID proteins also play a role on the side of inhibition, and we have already described E2A-2/5 and MSC as doing so. It is also possible that the *PAX* genes, lying genetically upstream of the MRFs, may function as inhibitors in this model and these tumors when existing as the *PAX-FKHR* fusion. My work expands the factors that act in a positive fashion on the process of myogenesis in this model to include RUNX1, RP58, and miR-206. Future work will need to further explore the relationship between the positive factors and the inhibitory ones, and especially the molecular mechanisms that are responsible for the downregulation and inhibition of the inhibitory factors that lock in the process of terminal differentiation. Certainly, the fact that multiple means exist that all cause RMS differentiation and withdrawal from the cell cycle is encouraging for the possibility to find a druggable target that could be used to leverage a novel differentiation-based therapy for these tumors.

Finally, the model of RMS as a solid tumor that represents an arrested state of development and possesses a 'tipping point' that can be manipulated may be more broadly

applicable than to just these specific tumors. bHLH factors control differentiation in other cell types, such as neurons, and it is possible that other pediatric tumors may be trapped in an analogous state to RMS – even those tumors that come from cell types where bHLH factors are not in a controlling role. Future experiments examining tumors other than rhabdomyosarcoma will be needed to determine if the lessons from these cells can be of use in other systems.

References

Aziz, A., Liu, Q.C., and Dilworth, F.J. (2010). Regulating a master regulator: Establishing tissue-specific gene expression in skeletal muscle. Epigenetics *5*, 691-695.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., *et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Bajard, L., Relaix, F., Lagha, M., Rocancourt, D., Daubas, P., and Buckingham, M.E. (2006). A novel genetic hierarchy functions during hypaxial myogenesis: Pax3 directly activates Myf5 in muscle progenitor cells in the limb. Genes & Development *20*, 2450-2464.

Barr, F.G., Galili, N., Holick, J., Biegel, J.A., Rovera, G., and Emanuel, B.S. (1993). Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma. Nat Genet *3*, 113-117.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Benezra, R., Davis, R.L., Lockshon, D., Turner, D.L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49-59.

Bergstrom, D.A., Penn, B.H., Strand, A., Perry, R.L., Rudnicki, M.A., and Tapscott, S.J. (2002). Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. Mol Cell *9*, 587-600.

Bergstrom, D.A., and Tapscott, S.J. (2001). Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family. Molecular and Cellular Biology *21*, 2404-2412.

Berkes, C.A., Bergstrom, D.A., Penn, B.H., Seaver, K.J., Knoepfler, P.S., and Tapscott, S.J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Molecular Cell *14*, 465-477.

Berkes, C.A., and Tapscott, S.J. (2005). MyoD and the transcriptional control of myogenesis. Semin Cell Dev Biol *16*, 585-595.

Biesiada, E., Hamamori, Y., Kedes, L., and Sartorelli, V. (1999). Myogenic basic helix-loop-helix proteins and Sp1 interact as components of a multiprotein transcriptional complex required for activity of the human cardiac alpha-actin promoter. Molecular and Cellular Biology *19*, 2577-2584.

Black, B.L., Martin, J.F., and Olson, E.N. (1995). The mouse MRF4 promoter is trans-activated directly and indirectly by muscle-specific transcription factors. The Journal of biological chemistry *270*, 2889-2892.

Bober, E., Lyons, G.E., Braun, T., Cossu, G., Buckingham, M., and Arnold, H.H. (1991). The muscle regulatory gene, Myf-6, has a biphasic pattern of expression during early mouse development. The Journal of Cell Biology *113*, 1255-1265.

Borycki, A.G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C., and Emerson, C.P., Jr. (1999). Sonic hedgehog controls epaxial muscle determination through Myf5 activation. Development *126*, 4053-4063.

Braun, T., Bober, E., Winter, B., Rosenthal, N., and Arnold, H.H. (1990). Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. The EMBO journal *9*, 821-831.

Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E., and 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*, 701-709.

Buas, M.F., Kabak, S., and Kadesch, T. (2009). Inhibition of myogenesis by Notch: evidence for multiple pathways. J Cell Physiol *218*, 84-93.

Buckingham, M. (2007). Skeletal muscle progenitor cells and the role of Pax genes. C R Biol *330*, 530-533.

Callis, T.E., Deng, Z., Chen, J.-F., and Wang, D.-Z. (2008). Muscling through the microRNA world. Exp Biol Med (Maywood) *233*, 131-138.

Cao, Y., Kumar, R.M., Penn, B.H., Berkes, C.A., Kooperberg, C., Boyer, L.A., Young, R.A., and Tapscott, S.J. (2006). Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters. EMBO J *25*, 502-511.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., MacQuarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L., *et al.* (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev Cell *18*, 662-674.

Carvajal, J.J., Cox, D., Summerbell, D., and Rigby, P.W. (2001). A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. Development *128*, 1857-1868.

Chen, C.M., Kraut, N., Groudine, M., and Weintraub, H. (1996). I-mf, a novel myogenic repressor, interacts with members of the MyoD family. Cell *86*, 731-741.

Chen, J.-F., Mandel, E.M., Thomson, J.M., Wu, Q., Callis, T.E., Hammond, S.M., Conlon, F.L., and Wang, D.-Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet *38*, 228-233.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cheng, Y., Wu, W., Kumar, S.A., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.B., Zhang, Y., Drautz, D., *et al.* (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res *19*, 2172-2184.

Cohen, M.M. (2009). Perspectives on RUNX genes: an update. Am J Med Genet A *149A*, 2629-2646.

Collins, C.A., Gnocchi, V.F., White, R.B., Boldrin, L., Perez-Ruiz, A., Relaix, F., Morgan, J.E., and Zammit, P.S. (2009). Integrated functions of Pax3 and Pax7 in the regulation of proliferation, cell size and myogenic differentiation. PLoS ONE *4*, e4475.

Conant, G.C., and Wagner, A. (2003). Convergent evolution of gene circuits. Nat Genet *34*, 264-266.

Cordero, O.X., and Hogeweg, P. (2006). Feed-forward loop circuits as a side effect of genome evolution. Mol Biol Evol *23*, 1931-1936.

Cosma, M.P. (2002). Ordered recruitment: gene-specific mechanism of transcription activation. Molecular Cell *10*, 227-236.

Cui, J.Y., Gunewardena, S.S., Rockwell, C.E., and Klaassen, C.D. (2010). ChIPing the cistrome of PXR in mouse liver. Nucleic Acids Research.

Cunha, P.M., Sandmann, T., Gustafson, E.H., Ciglar, L., Eichenlaub, M.P., and Furlong, E.E. (2010). Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity. PLoS Genet *6*, e1001014.

Davis, R.J., D'Cruz, C.M., Lovell, M.A., Biegel, J.A., and Barr, F.G. (1994). Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma. Cancer Res *54*, 2869-2872.

Davis, R.L., Cheng, P.F., Lassar, A.B., and Weintraub, H. (1990). The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell *60*, 733-746.

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

de la Serna, I.L., Ohkawa, Y., Berkes, C.A., Bergstrom, D.A., Dacwag, C.S., Tapscott, S.J., and Imbalzano, A.N. (2005). MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. Mol Cell Biol *25*, 3997-4009.

Diede, S.J., Guenthoer, J., Geng, L.N., Mahoney, S.E., Marotta, M., Olson, J.M., Tanaka, H., and Tapscott, S.J. (2010). DNA methylation of developmental genes in pediatric medulloblastomas identified by denaturation analysis of methylation differences. Proceedings of the National Academy of Sciences of the United States of America *107*, 234-239.

Dilworth, F.J., Seaver, K.J., Fishburn, A.L., Htet, S.L., and Tapscott, S.J. (2004). In vitro transcription system delineates the distinct roles of the coactivators pCAF and p300 during MyoD/E47-dependent transactivation. Proc Natl Acad Sci USA *101*, 11593-11598.

Dong, F., Sun, X., Liu, W., Ai, D., Klysik, E., Lu, M.F., Hadley, J., Antoni, L., Chen, L., Baldini, A., *et al.* (2006). Pitx2 promotes development of splanchnic mesoderm-derived branchiomeric muscle. Development *133*, 4891-4899.

Epstein, J.A., Lam, P., Jepeal, L., Maas, R.L., and Shapiro, D.N. (1995). Pax3 inhibits myogenic differentiation of cultured myoblast cells. J Biol Chem *270*, 11719-11722.

Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. Genes Dev *17*, 1115-1129.

Fischer, A., and Gessler, M. (2007). Delta-Notch--and then? Protein interactions and proposed modes of repression by Hes and Hey bHLH factors. Nucleic Acids Res *35*, 4583-4596.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490-493.

Fredericks, W.J., Galili, N., Mukhopadhyay, S., Rovera, G., Bennicelli, J., Barr, F.G., and Rauscher, F.J., 3rd (1995). The PAX3-FKHR fusion protein created by the t(2;13) translocation in alveolar rhabdomyosarcomas is a more potent transcriptional activator than PAX3. Mol Cell Biol *15*, 1522-1535.

Fry, C.J., and Peterson, C.L. (2001). Chromatin remodeling enzymes: who's on first? Curr Biol *11*, R185-197.

Fu, Y., Sinha, M., Peterson, C.L., and Weng, Z. (2008). The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet *4*, e1000138.

Ganapathi, M., Palumbo, M.J., Ansari, S.A., He, Q., Tsui, K., Nislow, C., and Morse, R.H. (2010). Extensive role of the general regulatory factors, Abf1 and Rap1, in determining genome-wide chromatin structure in budding yeast. Nucleic Acids Res.

Gaudet, J., and Mango, S.E. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821-825.

Gaudet, J., Muttumu, S., Horner, M., and Mango, S.E. (2004). Whole-genome analysis of temporal gene expression during foregut development. PLoS Biol *2*, e352.

Ge, Y., and Chen, J. (2011). MicroRNAs in skeletal myogenesis. Cell Cycle *10*, 441-448.

Gerber, A.N., Klesert, T.R., Bergstrom, D.A., and Tapscott, S.J. (1997). Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. Genes & Development *11*, 436-450.

Gilmour, B.P., Fanger, G.R., Newton, C., Evans, S.M., and Gardner, P.D. (1991). Multiple binding sites for myogenic regulatory factors are required for expression of the acetylcholine receptor gamma-subunit gene. J Biol Chem *266*, 19871-19874.

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

Gros, J., Manceau, M., Thomé, V., and Marcelle, C. (2005). A common somitic origin for embryonic muscle progenitors and satellite cells. Nature *435*, 954-958.

Guarente, L., Nye, J.S., Hochschild, A., and Ptashne, M. (1982). Mutant lambda phage repressor with a specific defect in its positive control function. Proc Natl Acad Sci U S A *79*, 2236-2239.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Hahn, H., Wojnowski, L., Specht, K., Kappler, R., Calzada-Wack, J., Potter, D., Zimmer, A., Muller, U., Samson, E., and Quintanilla-Martinez, L. (2000). Patched target Igf2 is indispensable for the formation of medulloblastoma and rhabdomyosarcoma. J Biol Chem *275*, 28341-28344.

Hamamori, Y., Wu, H.Y., Sartorelli, V., and 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*, 6563-6573.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell Res *21*, 396-420.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445-458.

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

Hebert, S.L., Simmons, C., Thompson, A.L., Zorc, C.S., Blalock, E.M., and Kraner, S.D. (2007). Basic helix-loop-helix factors recruit nuclear factor I to enhance expression of the NaV 1.4 Na+ channel gene. Biochimica et biophysica acta *1769*, 649-658.

Hinterberger, T.J., Sassoon, D.A., Rhodes, S.J., and Konieczny, S.F. (1991). Expression of the muscle regulatory factor MRF4 during somite and skeletal myofiber development. Developmental Biology *147*, 144-156.

Hirsinger, E., Duprez, D., Jouve, C., Malapert, P., Cooke, J., and Pourquie, O. (1997). Noggin acts downstream of Wnt and Sonic Hedgehog to antagonize BMP4 in avian somite patterning. Development *124*, 4605-4614.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497-1502.

Kablar, B., Asakura, A., Krastel, K., Ying, C., May, L.L., Goldhamer, D.J., and Rudnicki, M.A. (1998). MyoD and Myf-5 define the specification of musculature of distinct embryonic origin. Biochem Cell Biol *76*, 1079-1091.

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

Kassouf, M.T., Hughes, J.R., Taylor, S., McGowan, S.J., Soneji, S., Green, A.L., Vyas, P., and Porcher, C. (2010). Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. Genome Res *20*, 1064-1083.

Kim, H.K., Lee, Y.S., Sivaprasad, U., Malhotra, A., and Dutta, A. (2006). Muscle-specific microRNA miR-206 promotes muscle differentiation. J Cell Biol *174*, 677-687.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.

Knoepfler, P.S., Zhang, X.-y., Cheng, P.F., Gafken, P.R., McMahon, S.B., and Eisenman, R.N. (2006). Myc influences global chromatin structure. EMBO J *25*, 2723-2734.

Koi, M., Johnson, L.A., Kalikin, L.M., Little, P.F., Nakamura, Y., and Feinberg, A.P. (1993). Tumor cell growth arrest caused by subchromosomal transferable DNA fragments from chromosome 11. Science *260*, 361-364.

Kopan, R., Nye, J.S., and Weintraub, H. (1994). The intracellular domain of mouse Notch: a constitutively activated repressor of myogenesis directed at the basic helix-loop-helix region of MyoD. Development *120*, 2385-2396.

Koufos, A., Hansen, M.F., Copeland, N.G., Jenkins, N.A., Lampkin, B.C., and Cavenee, W.K. (1985). Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism. Nature *316*, 330-334.

Ladher, R.K., Church, V.L., Allen, S., Robson, L., Abdelfattah, A., Brown, N.A., Hattersley, G., Rosen, V., Luyten, F.P., Dale, L., *et al.* (2000). Cloning and expression of the Wnt antagonists Sfrp-2 and Frzb during chick development. Developmental Biology *218*, 183-198.

Langenau, D.M., Keefe, M.D., Storer, N.Y., Guyon, J.R., Kutok, J.L., Le, X., Goessling, W., Neuberg, D.S., Kunkel, L.M., and Zon, L.I. (2007). Effects of RAS on the genesis of embryonal rhabdomyosarcoma. Genes Dev *21*, 1382-1395.

Lassar, A.B., Buskin, J.N., Lockshon, D., Davis, R.L., Apone, S., Hauschka, S.D., and Weintraub, H. (1989). MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. Cell *58*, 823-831.

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

Lee, M.-H., Jothi, M., Gudkov, A.V., and Mal, A.K. (2011). Histone Methyltransferase KMT1A Restrains Entry of Alveolar Rhabdomyosarcoma Cells into a Myogenic Differentiated State. Cancer Research *71*, 3921-3931.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., *et al.* (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.

Lemercier, C., To, R.Q., Carrasco, R.A., and Konieczny, S.F. (1998). The basic helix-loop-helix transcription factor Mist1 functions as a transcriptional repressor of myoD. EMBO J *17*, 1412-1422.

Lepper, C., and Fan, C.-M. (2010). Inducible lineage tracing of Pax7-descendant cells reveals embryonic origin of adult satellite cells. Genesis *48*, 424-436.

Leung, T.H., Hoffmann, A., and Baltimore, D. (2004). One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. Cell *118*, 453-464.

Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., *et al.* (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol *6*, e27.

Liang, Z., and Biggin, M.D. (1998). Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. Development *125*, 4471-4482.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet *28*, 327-334.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, S., and Riggs, A.D. (1975). The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eucaryotes. Cell *4*, 107-111.

Lindsell, C.E., Shawber, C.J., Boulter, J., and Weinmaster, G. (1995). Jagged: a mammalian ligand that activates Notch1. Cell *80*, 909-917.

Lu, J., Webb, R., Richardson, J.A., and Olson, E.N. (1999). MyoR: a muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of MyoD. Proc Natl Acad Sci U S A *96*, 552-557.

Lu, J.-R., Bassel-Duby, R., Hawkins, A., Chang, P., Valdez, R., Wu, H., Gan, L., Shelton, J.M., Richardson, J.A., and Olson, E.N. (2002). Control of facial muscle development by MyoR and capsulin. Science *298*, 2378-2381.

MacArthur, S., Li, X.Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keranen, S.V., *et al.* (2009). Developmental roles of 21 Drosophila

transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol *10*, R80.

MacQuarrie, K.L., Fong, A.P., Morse, R.H., and Tapscott, S.J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. Trends Genet, 1-8.

Mal, A.K. (2006). Histone methyltransferase Suv39h1 represses MyoD-stimulated myogenic differentiation. EMBO J *25*, 3323-3334.

Marcelle, C., Stark, M.R., and Bronner-Fraser, M. (1997). Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. Development *124*, 3955-3963.

Maroto, M., Reshef, R., Munsterberg, A.E., Koester, S., Goulding, M., and Lassar, A.B. (1997). Ectopic Pax-3 activates MyoD and Myf-5 expression in embryonic mesoderm and neural tissue. Cell *89*, 139-148.

McAllister, R.M., Melnyk, J., Finkelstein, J.Z., Adams, E.C., Jr., and Gardner, M.B. (1969). Cultivation in vitro of cells derived from a human rhabdomyosarcoma. Cancer *24*, 520-526.

Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407-410.

Merlino, G., and Helman, L. (1999). Rhabdomyosarcoma--working out the pathways. Oncogene *18*, 5340-5348.

Merlino, G., and Khanna, C. (2007). Fishing for the origins of cancer. Genes Dev *21*, 1275-1279.

Miner, J.H., and Wold, B. (1990). Herculin, a fourth member of the MyoD family of myogenic regulatory genes. Proceedings of the National Academy of Sciences of the United States of America *87*, 1089-1093.

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., and Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. EMBO J *28*, 1234-1245.

Missiaglia, E., Shepherd, C.J., Patel, S., Thway, K., Pierron, G., Pritchard-Jones, K., Renard, M., Sciot, R., Rao, P., Oberlin, O., *et al.* (2010). MicroRNA-206 expression levels correlate with clinical behaviour of rhabdomyosarcomas. Br J Cancer, 1-9.

Mok, G.F., and Sweetman, D. (2011). Many routes to the same destination: lessons from skeletal muscle development. Reproduction *141*, 301-312.

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

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell *56*, 777-783.

Myer, A., Olson, E.N., and Klein, W.H. (2001). MyoD cannot compensate for the absence of myogenin during skeletal muscle differentiation in murine embryonic stem cells. Developmental Biology *229*, 340-350.

Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., and Nonaka, I. (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. Nature *364*, 532-535.

Neuhold, L.A., and Wold, B. (1993). HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. Cell *74*, 1033-1042.

Newton, W.A., Jr., Soule, E.H., Hamoudi, A.B., Reiman, H.M., Shimada, H., Beltangady, M., and Maurer, H. (1988). Histopathology of childhood sarcomas, Intergroup Rhabdomyosarcoma Studies I and II: clinicopathologic correlation. J Clin Oncol *6*, 67-75.

Ohkawa, Y., Yoshimura, S., Higashi, C., Marfella, C.G., Dacwag, C.S., Tachibana, T., and Imbalzano, A.N. (2007). Myogenin and the SWI/SNF ATPase Brg1 maintain myogenic gene expression at different stages of skeletal myogenesis. The Journal of biological chemistry *282*, 6564-6570.

Olson, E.N., Arnold, H.H., Rigby, P.W., and Wold, B.J. (1996). Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. Cell *85*, 1-4.

Ordahl, C.P., and Le Douarin, N.M. (1992). Two myogenic lineages within the developing somite. Development *114*, 339-353.

Orian, A., van Steensel, B., Delrow, J., Bussemaker, H.J., Li, L., Sawado, T., Williams, E., Loo, L.W.M., Cowley, S.M., Yost, C., *et al.* (2003). Genomic binding by the Drosophila Myc, Max, Mad/Mnt transcription factor network. Genes & Development *17*, 1101-1114.

Palii, C.G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., Atkins, H., Allan, D., Dilworth, F.J., Gentleman, R., *et al.* (2011). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. The EMBO journal *30*, 494-509.

Paulino, A.C., and Okcu, M.F. (2008). Rhabdomyosarcoma. Curr Probl Cancer *32*, 7-34.

Penn, B.H., Bergstrom, D.A., Dilworth, F.J., Bengal, E., and Tapscott, S.J. (2004). A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. Genes Dev *18*, 2348-2353.

Pfenning, A.R., Kim, T.-K., Spotts, J.M., Hemberg, M., Su, D., and West, A.E. (2010). Genome-wide identification of calcium-response factor (CaRF) binding sites predicts a role in regulation of neuronal signaling pathways. PLoS ONE *5*, e10870.

Piette, J., Bessereau, J.L., Huchet, M., and Changeux, J.P. (1990). Two adjacent MyoD1-binding sites regulate expression of the acetylcholine receptor alpha-subunit gene. Nature *345*, 353-355.

Ptashne, M., and Gann, A. (2002). Genes & signals (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Puri, P.L., Avantaggiati, M.L., Balsano, C., Sang, N., Graessmann, A., Giordano, A., and Levrero, M. (1997a). p300 is required for MyoD-dependent cell cycle arrest and muscle-specific gene transcription. EMBO J *16*, 369-383.

Puri, P.L., Iezzi, S., Stiegler, P., Chen, T.T., Schiltz, R.L., Muscat, G.E., Giordano, A., Kedes, L., Wang, J.Y., and Sartorelli, V. (2001). Class I histone deacetylases sequentially interact with MyoD and pRb during skeletal myogenesis. Molecular Cell *8*, 885-897.

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

Puri, P.L., Wu, Z., Zhang, P., Wood, L.D., Bhakta, K.S., Han, J., Feramisco, J.R., Karin, M., and Wang, J.Y. (2000). Induction of terminal differentiation by constitutive activation of p38 MAP kinase in human rhabdomyosarcoma cells. Genes Dev *14*, 574-584.

Ragoczy, T., Bender, M.A., Telling, A., Byron, R., and Groudine, M. (2006). The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. Genes & Development *20*, 1447-1457.

Rao, P.K., Kumar, R.M., Farkhondeh, M., Baskerville, S., and Lodish, H.F. (2006). Myogenic factors that regulate expression of muscle-specific microRNAs. Proc Natl Acad Sci USA *103*, 8721-8726.

Rao, P.K., Missiaglia, E., Shields, L., Hyde, G., Yuan, B., Shepherd, C.J., Shipley, J., and Lodish, H.F. (2010). Distinct roles for miR-1 and miR-133a in the proliferation and differentiation of rhambomyosarcoma cells. FASEB J, 1-11.

Relaix, F., Rocancourt, D., Mansouri, A., and Buckingham, M. (2005). A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. Nature *435*, 948-953.

Reshef, R., Maroto, M., and 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*, 290-303.

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

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., *et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods *4*, 651-657.

Rosenberg, M.I., Georges, S.A., Asawachaicharn, A., Analau, E., and Tapscott, S.J. (2006). MyoD inhibits Fstl1 and Utrn expression by inducing transcription of miR-206. J Cell Biol *175*, 77-85.

Rossi, S., Stoppani, E., Puri, P.L., and Fanzani, A. (2011). Differentiation of human rhabdomyosarcoma RD cells is regulated by reciprocal, functional interactions between myostatin, p38 and extracellular regulated kinase signalling pathways. EUROPEAN JOURNAL OF CANCER, 1-11.

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

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E.E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev *21*, 436-449.

Sandmann, T., Jensen, L.J., Jakobsen, J.S., Karzynski, M.M., Eichenlaub, M.P., Bork, P., and Furlong, E.E. (2006). A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. Dev Cell *10*, 797-807.

Sang, L., Coller, H.A., and Roberts, J.M. (2008). Control of the reversibility of cellular quiescence by the transcriptional repressor HES1. Science *321*, 1095-1100.

Sartorelli, V., Huang, J., Hamamori, Y., and Kedes, L. (1997). Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. Molecular and Cellular Biology *17*, 1010-1026.

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

Sasai, Y., Kageyama, R., Tagawa, Y., Shigemoto, R., and Nakanishi, S. (1992). Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. Genes Dev *6*, 2620-2634.

Sato, T., Rocancourt, D., Marques, L., Thorsteinsdottir, S., and Buckingham, M. (2010). A Pax3/Dmrt2/Myf5 regulatory cascade functions at the onset of myogenesis. PLoS genetics *6*, e1000897.

Schienda, J., Engleka, K.A., Jun, S., Hansen, M.S., Epstein, J.A., Tabin, C.J., Kunkel, L.M., and Kardon, G. (2006). Somitic origin of limb muscle satellite and side population cells. Proceedings of the National Academy of Sciences of the United States of America *103*, 945-950.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., *et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet *42*, 53-61.

Sebire, N.J., and Malone, M. (2003). Myogenin and MyoD1 expression in paediatric rhabdomyosarcomas. J Clin Pathol *56*, 412-416.

Sharp, R., Recio, J.A., Jhappan, C., Otsuka, T., Liu, S., Yu, Y., Liu, W., Anver, M., Navid, F., Helman, L.J., *et al.* (2002). Synergism between INK4a/ARF inactivation and aberrant HGF/SF signaling in rhabdomyosarcomagenesis. Nat Med *8*, 1276-1280.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet *31*, 64-68.

Shih, H.P., Gross, M.K., and Kioussi, C. (2007). Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch. Proceedings of the National Academy of Sciences of the United States of America *104*, 5907-5912.

Simone, C., Forcales, S.V., Hill, D.A., Imbalzano, A.N., Latella, L., and Puri, P.L. (2004). p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. Nature genetics *36*, 738-743.

Sirri, V., Leibovitch, M.P., and Leibovitch, S.A. (2003). Muscle regulatory factor MRF4 activates differentiation in rhabdomyosarcoma RD cells through a positive-acting C-terminal protein domain. Oncogene *22*, 5658-5666.

Snider, L., Thirlwell, H., Miller, J.R., Moon, R.T., Groudine, M., and Tapscott, S.J. (2001). Inhibition of Tcf3 binding by I-mfa domain proteins. Molecular and Cellular Biology *21*, 1866-1873.

Song, G., and Wang, L. (2009). Nuclear receptor SHP activates miR-206 expression via a cascade dual inhibitory mechanism. PLoS ONE *4*, e6880.

Song, G., Zhang, Y., and Wang, L. (2009). MicroRNA-206 targets notch3, activates apoptosis, and inhibits tumor cell migration and focus formation. The Journal of biological chemistry *284*, 31921-31927.

Spicer, D.B., Rhee, J., Cheung, W.L., and Lassar, A.B. (1996). Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. Science *272*, 1476-1480.

Spitz, F., Demignon, J., Porteu, A., Kahn, A., Concordet, J.P., Daegelen, D., and Maire, P. (1998). Expression of myogenin during embryogenesis is controlled by Six/sine oculis homeoproteins through a conserved MEF3 binding site. Proceedings of the National Academy of Sciences of the United States of America *95*, 14220-14225.

St-Pierre, B., Flock, G., Zacksenhaus, E., and Egan, S.E. (2002). Stra13 homodimers repress transcription through class B E-box elements. The Journal of biological chemistry *277*, 46544-46551.

Sumariwalla, V.M., and Klein, W.H. (2001). Similar myogenic functions for myogenin and MRF4 but not MyoD in differentiated murine embryonic stem cells. Genesis *30*, 239-249.

Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res *16*, 962-972.

Tapscott, S., Thayer, M.J., and Weintraub, H. (1993). Deficiency in rhabdomyosarcomas of a factor required for MyoD activity and myogenesis. Science *259*, 1450-1453.

Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Development *132*, 2685-2695.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science *242*, 405-411.

Taulli, R., Bersani, F., Foglizzo, V., Linari, A., Vigna, E., Ladanyi, M., Tuschl, T., and Ponzetto, C. (2009). The muscle-specific microRNA miR-206 blocks human rhabdomyosarcoma growth in xenotransplanted mice by promoting myogenic differentiation. J Clin Invest *119*, 2366-2378.

Turcotte, B., and Guarente, L. (1992). HAP1 positive control mutants specific for one of two binding sites. Genes Dev *6*, 2001-2009.

Vakoc, C.R., Letting, D.L., Gheldof, N., Sawado, T., Bender, M.A., Groudine, M., Weiss, M.J., Dekker, J., and Blobel, G.A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol Cell *17*, 453-462.

Wagner, J., Schmidt, C., Nikowits, W., Jr., and Christ, B. (2000). Compartmentalization of the somite and myogenesis in chick embryos are influenced by wnt expression. Developmental Biology *228*, 86-94.

Wang, H., Garzon, R., Sun, H., Ladner, K., Singh, R., Dahlman, J., Cheng, A., Hall, B., Qualman, S., Chandler, D., *et al.* (2008). NF-kappaB-YY1-miR-29 regulatory circuitry in skeletal myogenesis and rhabdomyosarcoma. Cancer Cell *14*, 369-381.

Wang, X., Blagden, C., Fan, J., Nowak, S.J., Taniuchi, I., Littman, D.R., and Burden, S.J. (2005). Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. Genes Dev *19*, 1715-1722.

Wang, Y., and Jaenisch, R. (1997). Myogenin can substitute for Myf5 in promoting myogenesis but less efficiently. Development *124*, 2507-2513.

Wasserman, W.W., and Fickett, J.W. (1998). Identification of regulatory regions which confer muscle-specific gene expression. Journal of molecular biology *278*, 167-181.

Weintraub, H. (1993). Summary: genetic tinkering--local problems, local solutions. Cold Spring Harb Symp Quant Biol *58*, 819-836.

Weintraub, H., Davis, R., Lockshon, D., and Lassar, A. (1990). MyoD binds cooperatively to two sites in a target enhancer sequence: occupancy of two sites is required for activation. Proc Natl Acad Sci U S A *87*, 5623-5627.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T.K., Turner, D., Rupp, R., and Hollenberg, S. (1991). The myoD gene family: nodal point during specification of the muscle cell lineage. Science *251*, 761-766.

Wentworth, B.M., Donoghue, M., Engert, J.C., Berglund, E.B., and Rosenthal, N. (1991). Paired MyoD-binding sites regulate myosin light chain gene expression. Proc Natl Acad Sci U S A *88*, 1242-1246.

Wright, W.E., Sassoon, D.A., and Lin, V.K. (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell *56*, 607-617.

Xia, S.J., Pressey, J.G., and Barr, F.G. (2002). Molecular pathogenesis of rhabdomyosarcoma. Cancer Biol Ther *1*, 97-104.

Yang, Z., MacQuarrie, K.L., Analau, E., Tyler, A.E., Dilworth, F.J., Cao, Y., Diede, S.J., and Tapscott, S.J. (2009). MyoD and E-protein heterodimers switch rhabdomyosarcoma cells from an arrested myoblast phase to a differentiated state. Genes Dev *23*, 694-707.

Yochum, G.S., Sherrick, C.M., Macpartlin, M., and Goodman, R.H. (2010). A beta-catenin/TCF-coordinated chromatin loop at MYC integrates 5' and 3' Wnt responsive enhancers. Proc Natl Acad Sci USA *107*, 145-150.

Yokoyama, S., Ito, Y., Ueno-Kudoh, H., Shimizu, H., Uchibe, K., Albini, S., Mitsuoka, K., Miyaki, S., Kiso, M., Nagai, A., *et al.* (2009). A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. Dev Cell *17*, 836-848.

Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R., and Young, R.A. (2003). Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell *113*, 395-404.

Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., and Levine, M. (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev *21*, 385-390.

Zhong, M., Niu, W., Lu, Z.J., Sarov, M., Murray, J.I., Janette, J., Raha, D., Sheaffer, K.L., Lam, H.Y.K., Preston, E., *et al.* (2010). Genome-wide identification of binding sites defines distinct functions for Caenorhabditis elegans PHA-4/FOXA in development and environmental response. PLoS Genet *6*, e1000848.

Zhu, X., Yeadon, J.E., and Burden, S.J. (1994). AML1 is expressed in skeletal muscle and is regulated by innervation. Mol Cell Biol *14*, 8051-8057.

VITA

Kyle L. MacQuarrie was born in Boston, Massachusetts and called that state his home until he moved west to Seattle for the Medical Scientist Training Program in 2005. He earned a Bachelor of Science in Biochemistry/Molecular Biology and Psychology from the University of Massachusetts at Amherst in 2003. He earned a Ph.D. in Molecular and Cellular Biology at the University of Washington and the Fred Hutchinson Cancer Research Center in 2011.