The Role of Six1 in Transcriptional Regulation during Myogenesis

Yubing Liu

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements

for the Doctorate in Philosophy degree in Biochemistry

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

University of Ottawa

ABSTRACT

Skeletal myogenesis is under the control of a combinatorial network of transcription factors. It has been shown that the homeobox protein Six1 is required for embryonic myogenesis. Using functional genomics approaches, I determined that Six1 is required for myoblasts differentiation through direct binding to a cluster of genes that are related to muscle function and muscle structure during my Master's studies. However, it was still not fully understood how Six1 selects its genomic targets and whether Six1 regulates the expression of Myod directly. I devoted my PhD work to study three central aspects of Six1 function: through what DNA motif it binds to DNA, how it regulates the expression of the myogenic regulatory factor MyoD, and how it might regulate chromatin structure at the enhancer regions of muscle genes. A more degenerate MEF3-like DNA sequence consensus has been identified from Six1 ChIP-on-chip experiments. This MEF3 motif was further optimized using bioinformatic methods and was proved to discover Six1 binding sites with improved specificity and sensitivity. Myod, a member of myogenic regulatory factors (MRFs), is a master regulator in the myogenic lineage. Multiple MEF3 sites were identified on the regulatory regions of Myod, including two MEF3 sites within its core enhancer region (CER). Six1 was able to bind to the CER directly through these two MEF3 sites and regulated the Myod expression in cultured primary myoblasts. Previous work has suggested that the CER is also bound by Myod in myoblasts. I demonstrated that the binding of Myod to the CER depended on the presence of Six1. Six1 was also involved in maintaining a relatively 'open' chromatin structure at the CER, suggesting that Six1 may play a direct or indirect role in chromatin remodeling. During my Master's studies, I demonstrated a synergistic regulation by the Six and MRF families. This synergistic function gains potential importance by the fact that

~25% of Six1 genomic targets are also bound by Myod. I decided to study whether the co-occupancy of Six1 and Myod was essential to maintain the proper global chromatin structure at these loci. Six1 and Myod co-bound genomic regions correlated with more accessible chromatin, which was detected by the formaldehyde-assisted isolation of regulatory elements (FAIRE) assay followed by DNA deep sequencing (FAIRE-seq). When combined with small interfering RNA-mediated gene knockdown of Six1 or Myod, FAIRE-seq data suggested that Six1, but not Myod, was involved in regulating the chromatin accessibility at these co-bound DNA loci. To shed light on the mechanism by which Six1 functions, proteomics approaches were used and revealed that proteins involved in "regulation of transcription" and "chromatin organization" were enriched among Six1-bound proteins. Cdk9 and its partner cyclin T have been shown to stimulate gene expression by releasing RNA polymerase II from transcriptional pause, but they can also function at gene enhancers. I determined that Six1 and Cdk9 participated in the same protein complex, and that the Cdk9 activity appeared to mediate the effect of Six1 on the chromatin accessibility at the CER to regulate the Myod expression. Taken together, these results demonstrate that Six1 regulates the expression of Myod through its direct binding on the CER which facilitates transcriptional elongation.

ACKNOWLEDGEMENTS

First of all, I would like to thank my PhD thesis supervisor, **Dr. Alexandre Blais**, for recruiting me to his lab as a Master's candidate and then a PhD candidate. I would like to thank Dr. Blais for his unconditional guidance and support during these years. Dr. Blais not only introduced the Six1 transcription factor to me, but also led me to the fascinating "epigenetic" world. I would like to thank Dr. Blais for pushing me when I was hesitating, encouraging me when I was uncertain and helping me out when I was in trouble shooting. Without the support and understanding from Dr. Blais, I would not have been able to write this thesis.

Next, I would like to thank Dr. Imane Chakroun, who is both an altruistic colleague and a supportive friend. Dr. Chakroun is always available for fruitful discussion and lending a hand for experiments. I would like to thank the past and present Blais lab members for their cooperative attitude and collaborative contribution over the years. Blais lab members together cultivate a helping environment that inspires me to go forward. I would like to thank Skerjanc, Couture, Brand, Dilworth, Ioshikhes, Kothary and Nemer labs for exchanging reagents, ideas and technical support.

I would like to thank my thesis advisory committee members Dr. Ilona Skerjanc, Dr. Marjorie Brand, and Dr. Jean-Francois Couture for their priceless suggestion and advisement over the years. I would also like to thank my PhD comprehensive exam members Dr. Ilona Skerjanc, Dr. David Picketts, and Dr. Mary-Ellen Harper.

I would like to thank funding agencies for supporting my PhD research, especially the Postgraduate Scholarships-Doctoral from NSERC and the Ontario Graduate Scholarship from the government of Ontario.

Last but not least, I would like to thank my family for their unconditional support during the years. I would like to thank my parents and my parents-in-law who helped me to take care of my twins.

Contents

LIST OF ABBREVIATIONS

A

ARMS	alveolar rhabdomyosarcoma

B

Baf170	SWI/SNF complex 170 kDa subunit
bHLH	basis helix-loop-helix domain
Bmal1	Arntl (aryl hydrocarbon receptor nuclear translocator-like)
BOR syndrome	Branchio-oto-renal syndrome
bp	base pair
Brd4	bromodomain containing 4

C

CAGE	cap analysis of gene expression
CBP	CREB-binding protein
CDK9	Cyclin-dependent Kinase 9
CER	core enhancer region
ChIP-on-chip	ChIP combined with DNA microarray
ChIP-seq	ChIP followed by DNA deep sequencing
Chrna1	cholinergic receptor, nicotinic, alpha polypeptide 1
Chrng	cholinergic receptor, nicotinic, gamma polypeptide
c-jun	Jun proto-oncogene
CKM	creatine kinase M-type
Clock	circadian locomotor output cycles kaput
c-MET	MET proto-oncogene
CRM	*cis*-regulatory modules
CTD	C-terminal repeat domain

D

Dach1	dachshund 1
Dach2	dachshund 2
DRR	distal regulatory region
DRB	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	DRB Sensitivity Inducing Factor
DSix4	CG3871 gene product from transcript CG3871-RA
E	
ECM	Extracellular matrix
EMSA	Electrophoretic Mobility Shift Assay
EMT	Epithelial-mesenchymal transition
eRNA	enhancer RNA
ESC	embryonic stem cells
ey	eyeless
eya	eyes absent
Eya1	EYA transcriptional coactivator and phosphatase 1
Eya2	EYA transcriptional coactivator and phosphatase 2
Eya3	EYA transcriptional coactivator and phosphatase 3
Eya4	EYA transcriptional coactivator and phosphatase 4
F	
FAIRE	Formaldehyde-Assisted Isolation of Regulatory Elements
FKHR	Forkhead Homolog in Rhabdomyosarcoma
Foxo3	forkhead box O3
G	
G9a	Ehmt2 (euchromatic histone lysine N-methyltransferase 2)
GnRH1	gonadotropin-releasing hormone
Grg4	Tle4 (transducin-like enhancer of split 4)
Grg5	Aes (amino-terminal enhancer of split)

Gro-seq	global run-on sequencing
H	
HD	homeodomain
Hdac3	histone deacetylase 3
H3K4me3	trimethylation of lysine 4 of Histone H3
H3K4me1	monomethylation of lysine 4 of Histone H3
H3K4me2	dimethylation of lysine 4 of Histone H3
H3K27ac	acetylation of lysine 27 of Histone H3
K	
Kb	kilo basepair
kD	kilo Dalton
KD	knockdown
KO	knockout
L	
Lbx1	ladybird homeobox homolog 1 (Drosophila)
lncRNA	long non-coding RNA
LPS	lipopolysaccharides
M	
MASTR	MEF2-interacting transcription factor
Mef2	myocyte enhancer factor-2
Meis	Meis homeobox 1
MHC	myosin heavy chain
MRFs	Myogenic Regulatory Factors
Mrf4	Myf6 (myogenic factor 6)
Msx1	msh homeobox 1
Myf5	myogenic factor 5

Myod	myogenic differentiation 1
Myog	Myogenin
N	
NDR	nucleosome-depleted region
NELF	negative elongation factor
NeuroD	neurogenic differentiation 1
O	
optix	CG18455 gene product from transcript CG18455-RC
P	
p27	Cdkn1b (cyclin-dependent kinase inhibitor 1B)
p300	E1A binding protein p300
Pax3	paired box 3
Pax7	paired box 7
Pbx	Pbx/knotted 1 homeobox
pCAF	p300/CBP-associated factor
Pitx2	paired-like homeodomain transcription factor 2
PRR	proximal regulatory region
p-TEFb	positive transcription elongation factor b
R	
RMS	Rhabdomyosarcoma
RNA pol II	RNA polymerase II
RNA-seq	RNA deep sequencing
S	
S1P	the soluble form of sphingolipid signalling
SD	Six domain
SF/HGF	Scatter factor/hepatocyte growth factor
SHD	Six-type homeodomain

SRF	serum response factor
so	Sine oculis
SWI/SNF	SWItch/Sucrose Non-Fermentable
T	
TAD	Transcriptional activation domain
TLR4	toll-like receptor 4
Tnnc1	troponin C, cardiac/slow skeletal
Y	
YY1	YY1 transcription factor

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

CHAPTER 2. DISCOVERY, OPTIMIZATION AND VALIDATION OF AN OPTIMAL DNA BINDING SEQUENCE FOR THE SIX1 HOMEODOMAIN TRANSCRIPTION FACTOR

CHAPTER 3. SIX1 REGULATES MYOD EXPRESSION IN ADULT MUSCLE PROGENITOR CELLS

CHAPTER 4. SIX1 INTERACTS WITH CDK9 TO REGULATE MYOD EXPRESSION IN MYOBLASTS

CHAPTER 5. DISCUSSION

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1. Embryonic Development of Skeletal Muscle

Skeletal muscle is responsible for voluntary movements of the body, which are mediated by the cooperation between motor neurons and the skeletal muscle. Skeletal muscle also plays a role in metabolism (Dahlmans et al., 2016; Shadrach and Wagers, 2011). The huge expenditure of energy in skeletal muscle puts this tissue as a major player in the obesity-related metabolic disorder. In obese subjects, intramuscular deposition of fats in skeletal muscle is associated with reduced mitochondrial metabolism, which is characterized by smaller mitochondria and less mitochondria content in this tissue (Dahlmans et al., 2016).

In pathological conditions or upon injury, skeletal muscle can regenerate extensively and differentiate to repair damaged muscle. There are a number of muscle-related diseases with defective muscle differentiation and/or muscle regeneration. Rhabdomyosarcoma (RMS) is a soft tissue cancer, which displays myogenic phenotype but is defective in terminal muscle differentiation (Keller and Guttridge, 2013). Muscle dystrophies are a heterogeneous collection of inherited muscle disorders. They often display progressive muscle weakening in different locations and dystrophic histological changes, such as continuous muscle regeneration and fibrosis (Mercuri and Muntoni, 2013; Shieh, 2013). Muscle wasting is also associated with natural ageing, namely sarcopenia. Sarcopenia is the major cause of body weakening, social limitations and health service burden in elders (Shadrach and Wagers, 2011).

Skeletal muscle development is a highly orchestrated process and has many evolutionarily conserved players (Bismuth and Relaix, 2010; Buckingham et al., 2003). The development of the

trunk, limb and head has distinct embryonic origins and divergent gene regulatory networks (reviewed in (Bismuth and Relaix, 2010)). Somites, the paired mesodermal segments along the neural tube, give rise to trunk and limb muscles (Bismuth and Relaix, 2010). Head muscles derive from cranial unsegmented mesoderm (Sambasivan et al., 2011a). Muscle precursor cells from the ventralateral myotome first undergo the epithelial-mesenchymal transition *(*EMT*)*, and then migrate to the forming limb buds. A few transcription factors and signalling pathways have been identified to play important roles in the migration of muscle precursor cells, such as the transmembrane tyrosine kinase receptor c-MET and its ligand scatter factor/hepatocyte growth factor (SF/HGF), and the Lbx1 transcription factor (Bladt et al., 1995; Christ and Brand-Saberi, 2002; Gross et al., 2000). Disrupted expression of the c-MET receptor or its ligand, or loss of function of Lbx1, all lead to the absence or reduction of muscles in the limbs.

At the later developmental stage, some muscle precursor cells continually contribute to embryonic and fetal muscle growth in developing limb buds, while some will take the position underneath the basal lamina as the satellite cell population (Chang and Rudnicki, 2014; Relaix et al., 2005). Satellite cells contribute to postnatal muscle growth and adult muscle regeneration (Seale et al., 2000).

1.2. Adult Muscle Regeneration

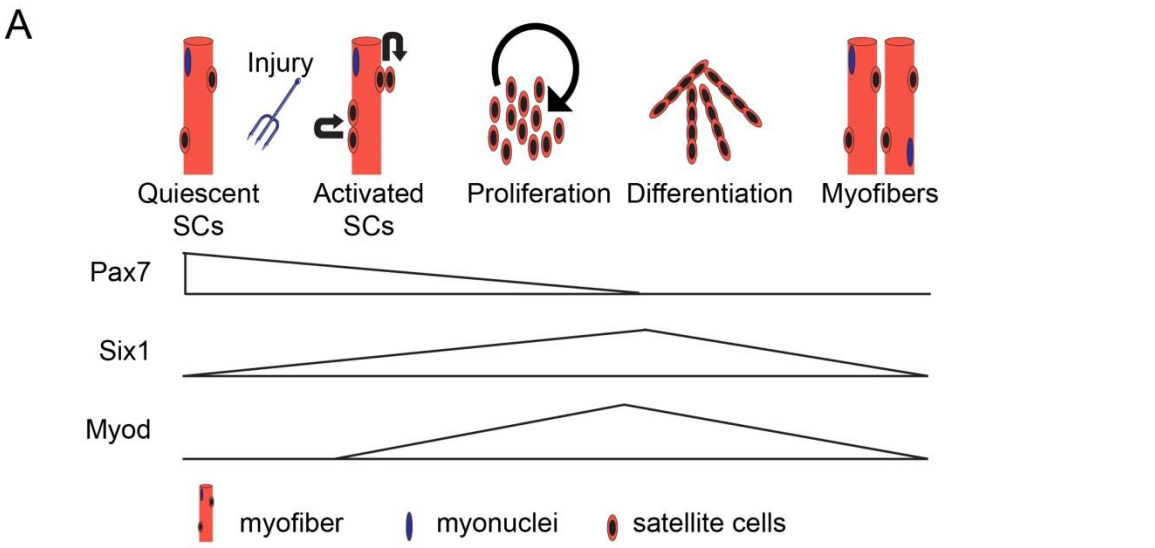
Satellite cells, considered as muscle stem cells, locate between the sarcolemma of their residing myofibers and the basal lamina, hence their name (Mauro, 1961). Several combinations of cell surface markers and transgenic mouse lines have been used to isolate satellite cells from tissues (Blanco-Bose et al., 2001; Bosnakovski et al., 2008; Fukada et al., 2004; Montarras et al., 2005; Pasut et al., 2012; Sacco et al., 2008; Sambasivan et al., 2009). The transcription factor Pax7 is

expressed specifically in satellite cells and their descendants in the myogenic lineage (Seale et al., 2000). Satellite cells are vital for homeostasis and regeneration of skeletal muscle. Specific and induced ablation of Pax7 positive cells in adult mice leads to poor regeneration, demonstrating that satellite cells are the chief mediators of muscle regeneration (Lepper et al., 2011; Sambasivan et al., 2011b).

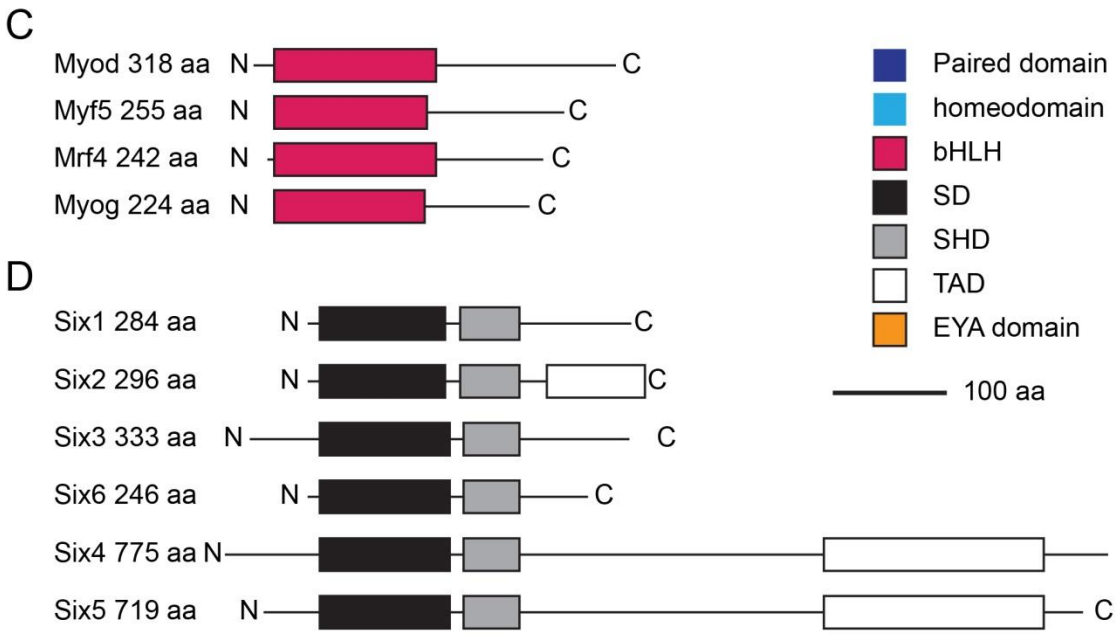
Satellite cells are a heterogeneous mix of quiescent stem cell and activated proliferating myoblasts (Chang and Rudnicki, 2014). In resting muscle, they are dormant in their niche and are predisposed to activation. Numerous studies suggest that quiescence is actively maintained despite their low metabolism, and the disturbance in genes that are required for quiescence leads to premature proliferation and eventually early exhaustion of the satellite cell pool (Bjornson et al., 2012; Chakkalakal et al., 2012; Cheung et al., 2012; Crist et al., 2012; Jiang et al., 2014; Liu et al., 2013a; Montarras et al., 2013; Mourikis et al., 2012; Shea et al., 2010). When triggered, satellite cells can go through planar division (symmetric) and basal-apical division (asymmetric) (Chang and Rudnicki, 2014) (Figure 1.1A). Symmetric division gives rise to two identical daughter cells, which are either both committed to proliferate or both replenishing the stem cell reservoir. Asymmetric division generates one committed daughter cell and one going back to quiescence (Dumont et al., 2015). The decision to enter the cell cycle, how to divide and the fate of daughter cells are regulated, at least in part, by signals originating from the satellite cell niche.

Although satellite cells have been recognized as the primary source of muscle regeneration, cells from other origins also show myogenic potential (Chang and Rudnicki, 2014; Judson et al., 2013). It is believed that mesenchymal progenitor cells (LaBarge and Blau, 2002), pericytes (Dellavalle et al., 2011), progenitor interstitial cells (PICs) (Mitchell et al., 2010), mesoangioblasts (Minasi et al., 2002) and cells from skeletal muscle side population (Meeson et

al., 2004) can be manipulated to generate functional myofibers under certain conditions. However, their contribution to normal muscle regeneration remains under investigation.


A
Injury
Quiescent SCs
Activated SCs
Proliferation
Differentiation
Myofibers
Pax7
Six1
Myod
myofiber
myonuclei
satellite cells


B
Paired box
Homeobox
Pax3 479 aa N C
Pax7 503 aa N C


C
Myod 318 aa N C
Myf5 255 aa N C
Mrf4 242 aa N C
Myog 224 aa N C
Paired domain
homeodomain
bHLH
SD
SHD
TAD
EYA domain
D
Six1 284 aa N C
Six2 296 aa N C
Six3 333 aa N C
Six6 246 aa N C
Six4 775 aa N C
Six5 719 aa N C
100 aa

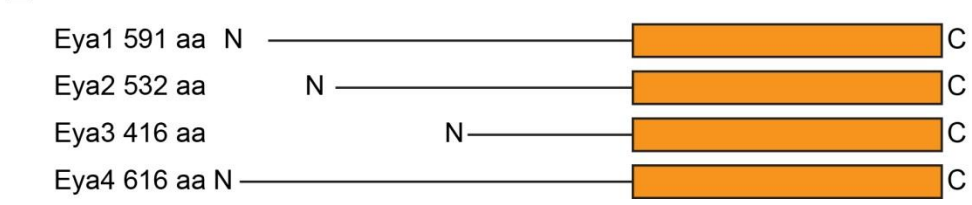

E
Eya1 591 aa N C
Eya2 532 aa N C
Eya3 416 aa N C
Eya4 616 aa N C

Figure 1. 1 Transcription factors and co-factors in muscle development

A) Model of satellite cell activation and muscle regeneration. Quiescent satellite cells (SCs), expressing Pax7 and Six1, are activated upon stimuli, undergoing symmetrical or asymmetrical cell division. Activated SCs, with enhanced transcription of Six1 and Myod, start to proliferate and eventually differentiate to repair damaged myofibers. The expression patterns of Pax7, Six1 and Myod are shown at the bottom. B) Pax3 and Pax7 share a Paired domain (126 aa) and a homeodomain (54 aa). C) The MRF family transcription factors share a basic-helix-loop-helix (bHLH) domain (132-142 aa). D) The Six family transcription factors share a conserved Six domain (SD, 110-114 aa) and a non-canonical Six-type homeodomain (SHD, 49-55 aa). Six4 and Six5 share a transactivation domain (TAD, approximately193 aa), which is not found in other family members. The carboxyl-terminus of Six2 is also shown to have transactivation function. E) The Eya family cofactors share a conserved Eya domain (ED, 271 aa). The length of each protein is indicated.

1.3. The Satellite Cell Niche

The satellite cell niche is the micro-environment that they reside in and it is essential to maintain the satellite cell properties: proliferation and self-renewal while maintaining the myogenic potential. The satellite cell niche, not only is a supportive anatomical structure that separates the stem cells from their neighbouring tissues, but also transmits mechanical and biochemical signals that are needed to maintain their proper stem cell characteristics. The niche includes autocrine and paracrine signals carried by satellite cells themselves and their residing myofibers, motor neurons, extracellular matrix and endocrine signals from the circulatory system (Dumont et al., 2015; Rezza et al., 2014). For example, the Notch signalling pathway has been shown as an important player during muscle regeneration. A delicate balance between Notch1 signalling and its inhibitor Numb controls satellite cell function and myoblasts differentiation. Enhanced activation of Notch1 signalling maintains satellite cells in a more primitive proliferating state, while attenuated Notch1 signalling by induced Numb expression leads muscle progenitor cells to a more committed state (Conboy and Rando, 2002; Conboy et al., 2003).

Satellite cells are surrounded by the extracellular matrix (ECM) in three-dimensions and the ECM can transmit various factors that are important to maintain the stem cell property of satellite cells (Yin et al., 2013). In cases of injury or exercise, growth factors are released and presented to satellite cells through the ECM (Yin et al., 2013). In addition, cell surface receptors and adhesion molecules are distributed differentially on satellite cells, which respond to distinct signals from their surrounding niche (Kuang et al., 2008). Integrin $\alpha 7\beta 1$, expressed by satellite cells and located on the basal lamina side of satellite cells, interacts with collagen IV and laminin that are components of the basal lamina (Thomas et al., 2015; Yin et al., 2013). Moreover, as components of the basement membrane, collagen VI is required for proper muscle regeneration

and robust satellite cell self-renewal, probably due to its contribution to the niche stiffness (Urciuolo et al., 2013; Zou et al., 2008). Furthermore, although fibronectin is normally present in the reticular lamina, distant from satellite cells, it is also transiently presented to and expressed in satellite cells during muscle regeneration, where it couples with the Frizzled-7 and Syndecan-4 co-receptor, regulating symmetric division and expansion of satellite cells (Bentzinger et al., 2013).

Of note, the crosstalk between satellite cells and myofibers is essential for muscle homoeostasis during postnatal life. M-cadherin is expressed on the myofiber side of satellite cells and on their residing myofibers (Irintchev et al., 1994). The interaction of M-cadherins between satellite cells and their myofibers could activate myogenic cell proliferation *in vitro* (Marti et al., 2013). Similarly, connective tissue is also required for proper muscle regeneration (Mathew et al., 2011). When Tcf4-expressing fibroblasts in connective tissue are genetically ablated, muscle regeneration is intensively delayed, leading to premature satellite cell differentiation and smaller myofibers after injury (Murphy et al., 2011). It is suggested that TCF-dependent signals from fibroblasts promote the fetal to adult muscle transition and TCF-independent secreted factors elevate myoblast fusion. In addition, both vasculature and systemic signals also contribute to the regulation of satellite cell function. As an example, S1P (Sphingosine-1-phosphate), which could be released from various cell types to the blood, can promote satellite cell proliferation and muscle regeneration through an autocrine/paracrine mode (reviewed in (Donati et al., 2013)).

1.4. Slow- and Fast- twitch muscle fibers

Adult skeletal muscles are composed of myofibers that exhibit different metabolic enzymes and specific types of contractile proteins. In general, myofibers are classified into two broad

categories: fast-twitch glycolytic fibers and slow-twitch oxidative fibers. Fast-twitch myofibers are made for intense but transient movement, while they are not resistant to fatigue due to their glycolytic nature. Instead, slow-twitch myofibers are tolerant to fatigue and they are designed for prolonged function. The function of a certain muscle defines its composition of fiber types. For example, gastrocnemius muscles, the bulky and superficial layer of "calf" muscles, are enriched for fast-type fibers to accommodate their function in walking and running. By contrast, soleus muscles are petite in size and they are buried underneath of gastrocnemius muscles (Brooke and Kaiser, 1970). Soleus muscles are enriched for slow-type fibers, which are intended for sustaining standing posture (Brooke and Kaiser, 1970). The composition of fiber types has been shown to be influenced by many factors, including signaling from motor neurons (Olson and Williams, 2000). A few transcription factors have been shown to mediate these signals and regulate the expression of certain contractile proteins to have an impact on the fiber type composition (Ekmark et al., 2007; Grifone et al., 2004; Niro et al., 2010; Salminen et al., 1996).

1.5. Transcription Factors in Myogenesis

A combinatorial regulatory network of transcription factors and co-factors coordinately controls myogenesis through the embryonic stage, to the postnatal period and adulthood. Many of those transcription factors are well conserved from invertebrates to vertebrates. A large amount of the current knowledge of skeletal myogenesis was gained from mouse, fruit fly, fish and avian models, through gain-of-function, loss-of-function, and transplantation assays.

1.5.1. Pax Transcription Factors: Pax3 and Pax7

The Pax transcription factor family consists of nine proteins, Pax1 to Pax9 (reviewed in (Lang et al., 2007)). Based on the protein structures, nine members are divided into four sub-groups:

Pax1/9; Pax2/5/8; Pax3/7 and Pax4/6. They all share a PAIRED domain which enables their binding to DNA and some of them also contain a homeodomain also implicated in DNA binding. The expression of genes in the same sub-group is highly overlapping. Pax3 and Pax7 are detected in central nervous system, neural crest, somites and muscles (Figure 1.1B). Pax3 and Pax7 together mark the myogenic population in the dermomyotome that gives rise to fetal myogenic precursor cells and postnatal satellite cells (Gros et al., 2005; Kassar-Duchossoy et al., 2005; Relaix et al., 2005).

Mutations in *PAX3* are related to the Waardenburg syndrome in human, characterized by mild to severe deafness, pigmentation anomalies, and some defects in tissues derived from neural crest (Epstein et al., 1991). In the *Splotch* mouse model where functional Pax3 is lacking, the size and organization of dermomyotome and later on the myotome are severely affected and limb muscles are ablated (Tajbakhsh et al., 1997; Tremblay et al., 1998). Moreover, the absence of ventral musculature and limb muscles suggests that Pax3 has a role in the long-range migration of muscle progenitor cells to their destination, possibly through the regulation of c-Met and Lbx1 (Epstein et al., 1996; Lagha et al., 2010; Tremblay et al., 1998).

Furthermore, the *PAX3*/*FKHR* (*FOXO1*) gene fusion is detected in 70% of alveolar rhabdomyosarcoma (ARMS) patients and correlates with a less favorable prognostic outcome (Barr, 2001). The *PAX3*/*FKHR* fusion gene is over-expressed in ARMS cell lines compared to the wild-type *PAX3* (Davis and Barr, 1997), and the fusion protein acts as a potent transcriptional activator when the C-terminal *FKHR* transactivation domain is fused to the N-terminal DNA binding domain of *PAX3* (Bennicelli et al., 1996). The mouse Pax3 itself can also induce the transcription of target genes. Ectopic expression of Pax3 can induce myogenesis in non-myogenic tissues of chick embryos, manifested by the expression of myogenic regulatory factors

and terminal differentiation markers (Maroto et al., 1997). Over-expression of Pax3 in P19 embryonic carcinoma cells can induce the expression of several transcription factors important for myogenesis, including Myod and Six1 (Ridgeway and Skerjanc, 2001).

Pax7, a nuclear satellite cell marker, is essential for satellite cell specification and survival (Seale et al., 2000). Pax7 is expressed in quiescent and activated satellite cells and its expression is then diminished at the onset of differentiation (Olguin and Olwin, 2004). Pax7-null mice exhibit strikingly reduced muscle mass after birth and die within two weeks, which is a result of complete loss of satellite cells that are responsible for the embryonic and post-natal muscle growth (Seale et al., 2000). Pax7 over-expression inhibits muscle differentiation through blocking the expression of Myog and the cell cycle exit (Olguin and Olwin, 2004; Olguin et al., 2007; Zammit et al., 2006). It is suggested that the ratio of PAX7 and MYOD proteins has an impact on the cell fate decision between proliferation and differentiation (Olguin et al., 2007). Pax7 may affect Myod function in two ways: transcriptional regulation and interfering with the stability of the MYOD protein (Olguin et al., 2007). Taken together, this evidence strongly supports the overlapping but also distinct functions of Pax3 and Pax7: the involvement of Pax3 in the specification of the muscle lineage and the migration of muscle precursor cells in the embryonic stage, and the requirement of Pax7 in satellite cell function in the adulthood.

1.5.2. Myod and Myogenic regulatory factors (MRFs)

Myod was the first transcription factor to be discovered that can convert non-muscle cells to the skeletal muscle lineage (Davis et al., 1987). Later on, three related genes, *Myf5*, *Mrf4* and *Myog*, were identified as Myod family members and showed variable myogenic conversion potential (Braun et al., 1989, 1990; Edmondson and Olson, 1989; Rhodes and Konieczny, 1989; Wright et

al., 1989). Genetic deletion of MRFs proteins revealed their redundant but also distinct roles in muscle development (reviewed in (Fong and Tapscott, 2013; Megeney and Rudnicki, 1995)). Myf5, Myod and MRF4 are functionally redundant and are required for muscle lineage determination, since mice lacking either gene alone do not yield apparent skeletal muscle defect while disruption of these three MRFs together results in the complete absence of skeletal muscle (Braun and Arnold, 1995; Kassar-Duchossoy et al., 2004; Patapoutian et al., 1995; Rudnicki et al., 1992, 1993; Zhang et al., 1995). Gene ablation studies in mice revealed that Myog is absolutely required for muscle differentiation at the embryonic stage but it is dispensable during postnatal life (Hasty et al., 1993; Meadows et al., 2008, 2011; Nabeshima et al., 1993).

MRFs are a class of basic helix-loop-helix (bHLH) transcription factors (Murre et al., 1994), which function through heterodimerization with E proteins (Lassar et al., 1991). They share three conserved domains, a transcriptional activation domain (TAD) in the amino terminal region, a bHLH domain in the middle, and an α-helix domain in the carboxyl terminal region (Singh and Dilworth, 2013; Tapscott, 2005) (Figure 1.1C). The TAD confers transcription enhancement when it is fused to GAL4 DNA-binding domain (Weintraub et al., 1991). The bHLH domain mediates their interaction with E-proteins and the subsequent binding to the E-box DNA consensus sequences that are present among muscle-specific promoters and enhancers (Blackwell and Weintraub, 1990; Weintraub et al., 1990). The muscle-specific di-amino acids, an alanine at position 86 and a threonine at position 87 of Myod, in the bHLH domain of MRFs are necessary to activate myogenic transcription (Brennan et al., 1991; Davis et al., 1990; Ma et al., 1994; Weintraub et al., 1991). These di-amino acids were referred as the “myogenic code” (Heidt et al., 2007). The analysis of genomic binding data of Myod and NeuroD suggests that the E-box DNA sequences are presented differentially between myogenic and neuronal genes, which may

also contribute to the specificity of target gene selection (Fong et al., 2012, 2015). NeuroD is a master regulator of neurogenesis, which also contains a bHLH domain and binds the E-box motif (Chae et al., 2004).

Myod also interacts with other transcription factors and co-factors to fine-tune its function. Genes with similar function, such as those genes encoding muscle structure proteins, tend to be regulated by a group of transcription factors through their *cis*-regulatory modules (CRM) (Jeziorska et al., 2009). The CRM is a cluster of transcription factor binding sites that integrates the input of transcription factors and their associated co-factors into a transcriptional output of the target gene. Myod binds to genome-wide targets in myoblasts and a subset of genes are also co-bound by Six1 or Six4, based on genomic binding profile analysis (Cao et al., 2010; Chakroun et al., 2015; Liu et al., 2010). In addition, over a thousand CRMs were identified to be co-occupied by Myod and Six (Santolini et al., 2016). Furthermore, Myod and Six family members cooperatively activate the expression of a number of genes, such as Myog (Chakroun et al., 2015; Liu et al., 2010; Santolini et al., 2016). Interestingly, CRMs that have been validated to be synergistically regulated by Myod and Six proteins also contain consensus binding sites for other transcription factors, such as the myocyte enhancer factor-2 (Mef2). Mef2 proteins are able to bind to the Myog promoter in differentiating myocytes and their binding is required for the expression of Myog *in vivo* and in *vitro* (Cheng et al., 1993; Rampalli et al., 2007). It has been reported that the MEF2 and MRFs families cooperatively activate the expression of certain myogenic genes (Blais et al., 2005; Molkentin et al., 1995; Naidu et al., 1995).

The transcriptional gene regulation requires the contact of transcription factors to their DNA targets, which encounters the organization of nucleosomes. DNA on nucleosomes forms tight association with histones and is difficult to access by DNA-binding transcription factors. In the

case of myogenic conversion, exogenously expressed Myod encounters the silent chromatin architecture at many myogenic genes in these non-muscle cells (Tapscott, 2005). Similarly, muscle differentiation requires opening up chromatin at regulatory elements of many myogenic genes that are induced dramatically in differentiated myotubes (Tapscott, 2005). Various strategies have been shown to be involved in facilitating access to DNA by transcription factors (Figure 1.2). Firstly, histone variants can replace canonical histones in a replication-independent manner and contribute to permissive chromatin environment at regulatory elements, such as H3.3 and H2A.Z (Jin et al., 2009) (Figure 1.2A). Secondly, histones are substrates of a range of reversible post-translational modifications (Figure 1.2A). Histone acetylation, which is catalyzed by a series of histone acetyltransferases (HATs), interferes with the affinity between DNA and histones and then destabilizes nucleosomes (Roth et al., 2001). The impact of histone methylation on the chromatin structure depends on the position of histone residues. The lysine 4 at the histone H3 (H3K4) can be mono- (H3K4me1), di- (H3K4me2) or tri-methylated (H3K4me3), by certain methyltransferases (Khare et al., 2012). H3K4me3 is capable of recruiting chromatin-remodeling proteins that enhance the chromatin accessibility at promoters (Bell et al., 2011). By contrast, tri-methylation of the lysine 27 at the histone H3 is a repressive histone mark and contributes to nucleosome condensation (Bell et al., 2011). Thirdly, ATP-dependent chromatin remodeling complexes generate nucleosome-depleted regions by evicting or sliding nucleosomes (Hargreaves and Crabtree, 2011). Last but not the least, the binding of pioneer factors or a few cooperative transcription factors could compete with histones to bind DNA and thus gives rise to regions with low occupancy of nucleosomes at regulatory elements (Figure 1.2B).

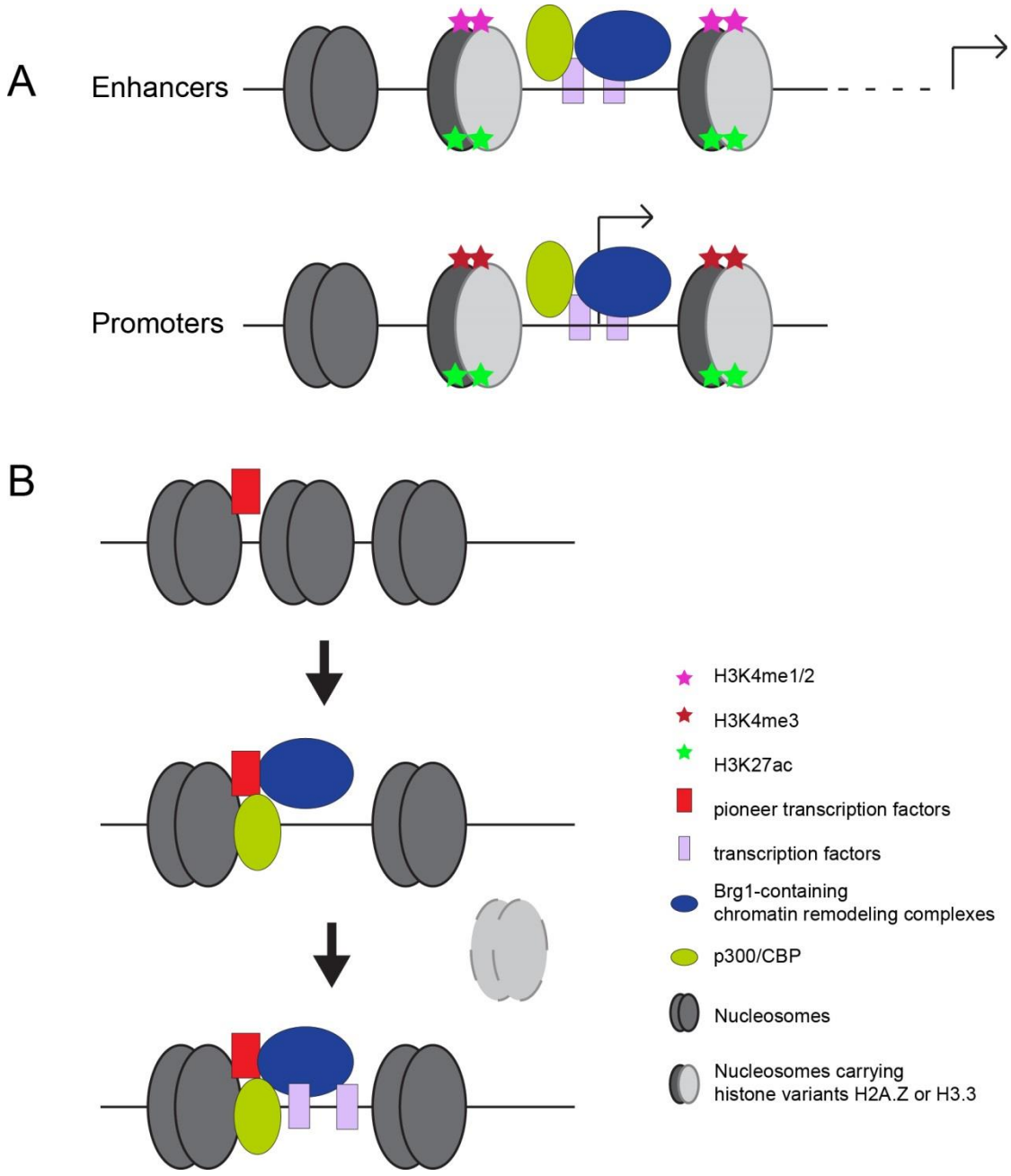

A
Enhancers
Promoters
B
H3K4me1/2
H3K4me3
H3K27ac
pioneer transcription factors
transcription factors
Brg1-containing
chromatin remodeling complexes
p300/CBP
Nucleosomes
Nucleosomes carrying
histone variants H2A.Z or H3.3

Figure 1. 2 Modifiers of chromatin architecture

A) Nucleosomes that contain histone variants, such as H3.3 and/or H2A.Z, display instable chromatin structure, resulting in permissive chromatin for transcription factors to bind. The recruitment of chromatin modifiers, including p300/CBP protein acetyltransferases and Brg1-containing chromatin remodeling complexes, modulates chromatin architecture. H3K4me1/2 and H3K27ac are associated with functional enhancers, while H3K4me3 and H3K27ac are enriched at active promoters. B) Pioneer transcription factors are a class of master regulators in lineage specification. Pioneer factors recognize DNA sequences on nucleosomes and mark these sites for change of the chromatin structure, and thus further facilitate the binding of other transcription factors that can only bind to more exposed DNA. In some cases, cooperative binding of a few transcription factors can promote chromatin remodeling through association with various factors.

For example, Myog is a MRF family member whose expression is initiated at the early steps of differentiation (Andrés and Walsh, 1996; Wright et al., 1989). The *Myog* promoter is not permissive in myoblasts but becomes more accessible and bound by Myod in differentiating myoblasts (Cao et al., 2010; Serna et al., 2005). A few mechanisms have been associated with the recruitment of Myod to the *Myog* promoter (reviewed in (Faralli et al., 2012)). The Pbx/Meis heterodimer binds to regulatory elements of a subset of muscle-specific genes in proliferating myoblasts including the *Myog* promoter, and upon differentiation it tethers the heterodimer of Myod and E-proteins to these loci (Berkes et al., 2004). In the meantime, Pbx1 interacts with Brg1, which is the catalytic subunit of the large, multi-protein SWI/SNF chromatin-remodeling complex, at the onset of myoblast differentiation, while the association of Myod and Brg1 is only detected in differentiating cells (Forcales et al., 2012; Serna et al., 2005). The Myod-mediated Myog expression depends on the functional Brg1 since dominant negative Brg1 abolishes its expression (Serna et al., 2005). In addition, Myod directly interacts with Baf60c, which is one of three alternative subunits of the SWI/SNF complex, and their association is persistent from myoblasts to myotubes (Forcales et al., 2012). Upon differentiation, the phosphorylation of BAF60C by p38 kinase promotes the incorporation of BAF60C and MYOD into the rest of the SWI/SNF complex, through which the chromatin remodeling happens and the induction of Myog transcription is achieved (Forcales et al., 2012). In line with this, embryonic stem cells (ESCs) are resistant to the Myod-mediated myogenic conversion and this has been shown to be due to a lack of Baf60c expression (Albini et al., 2013). Concomitant over-expression of Baf60c with Myod enables Myod to induce the commitment of human embryonic stem cells into the skeletal muscle lineage (Albini et al., 2013). Taken together, evidence suggests that a few parallel mechanisms are involved in Myod-mediated gene activation.

Due to the strong myogenic conversion potential of Myod, the regulation of its transcription has been extensively studied. Three regulatory elements upstream of the Myod promoter have been identified through mutagenesis and transgenic mice studies: the distal regulatory region (DRR) that is 5 kb upstream of the Myod gene, the core enhancer region (CER) that is 24 kb upstream, and the proximal regulatory region (PRR) which comprises the core promoter (Tapscott, 2005). During embryonic development, the expression of Myod is restricted to the skeletal muscle lineage, including myotomes and developing limb muscles (Buckingham et al., 2003). Myod transcription has been shown under the control of transcription factors and signaling pathways (Buckingham and Rigby, 2014). Combined mutation of Myf5 and Pax3 in mice leads to loss of Myod expression (Tajbakhsh et al., 1997). Similarly, delayed and reduced Myod expression is also observed in Six1 mutants, Six1 and Six4 double mutants, and Eya1/Eya2 double mutants in mice (Grifone et al., 2005, 2007; Laclef et al., 2003a). MASTR (MEF2-interacting transcription factor) also induces Myod expression *in vivo*, probably through direct regulation on the DRR (Creemers et al., 2006; Mokalled et al., 2012). Pitx2 (Paired-like homeodomain transcription factor 2) binds to the CER and drives the early expression of Myod in developing embryos (L'honoré et al., 2010).

During muscle regeneration, Myod is not detected in quiescent satellite cells but its expression is quickly induced upon satellite cell activation (Rocheteau et al., 2012). Both primary myoblasts and C2C12 cells, a murine myoblast cell line, have been broadly used to study the transcriptional regulation of Myod. Pax3, Pax7 and Foxo3 form a protein complex and then locate to the PRR to facilitate the recruitment of RNA polymerase II in primary myoblasts and C2C12 myoblasts (Hu et al., 2008). SRF (serum response factor) and YY1 complexes bind to the DRR and activate a DRR-driven reporter in proliferating C2 myoblasts, and their binding is competed out by MEF2

transcription factors to ensure proper differentiation (L'honore et al., 2003, 2007). The heterodimer of BMAL1 and CLOCK, circadian rhythm regulators, binds to the CER in C2C12 myoblasts on a non-canonical E-box, activating the circadian oscillation of Myod expression in C2C12 myotubes (Zhang et al., 2012). Ectopically expressed Msx1 binds to the CER in C2C12 myoblasts and represses its expression (Lee et al., 2004b). The Six1 homeodomain transcription factor can bind directly to the DRR, regulating Myod expression in primary myotubes (Le Grand et al., 2012). Myod protein can also bind to its own regulatory elements: the PRR in myoblasts (Zingg et al., 1994) and the DRR in myotubes (Le Grand et al., 2012), suggesting a Myod auto-regulation loop. Taken together, evidence suggests that Myod is tightly controlled by an elaborated regulatory network.

1.5.3. Six transcription factor family

Homeodomain (HD) proteins bind to DNA mainly acting as transcription factors that may have positive or negative impact on the expression of target genes (Bürglin and Affolter, 2016). Since the discovery of homeobox genes in *Drosophila*, our knowledge of the conserved homeobox gene family has expanded substantially in both animals and plants (Bürglin and Affolter, 2016). *sine oculis* (*so*), a homeobox gene, was discovered first in *Drosophila* where it exerts an important function on compound eye development within a gene regulatory network (Cheyette et al., 1994; Fischbach and Technau, 1984; Kumar, 2009a). Two additional genes, *optix* and *DSix4* were identified subsequently in fruit flies, comprising three founding members of the Six family that are believed to have undergone gene duplication during evolution (Kawakami et al., 2000; Seimiya and Gehring, 2000; Seo et al., 1999). In human and mouse, the Six transcription factor family consists of six members, from Six1 to Six6. Based on the protein sequences, these family members divide into three subgroups and each subgroup is thought to have originated from one

homologue gene in *Drosophila*, Six1/Six2 (*so*), Six3/Six6 (*optix*), and Six4/Six5 (*DSix4*) (Kawakami et al., 2000; Seo et al., 1999).

All six members share a Six-type homeodomain (SHD) and a Six-specific domain (SD) (Kawakami et al., 2000) (Figure 1.1D). The SHD domain bears key residues that are different from a typical HD, such as an arginine at the position 5, a glutamine at the position 12 and a lysine at the position 50, allowing their binding to specific DNA sequences (Kumar, 2009a; Salminen et al., 1996; Serikaku and O'Tousa, 1994). The SD mediates their interaction with other proteins and also contributes to DNA binding specificity (Kawakami et al., 2000; Ohto et al., 1999). Six1, Six2, Six4 and Six5 can recognize the MEF3 consensus DNA sequence, while Six3 and Six6 have been reported to bind the typical "ATTA" consensus DNA sequence recognized by most HD factors (Kawakami et al., 1996a; Kumar, 2009a; O'Brien et al., 2016; Xie et al., 2015). Their N-terminal sequences are variable in length with low conservation. Intrinsic TADs are identified in the Six4/Six5 subgroup and Six2 (Brodbeck and Englert, 2004; Kawakami et al., 1996b). These TADs are able to activate the transcription of reporters when tethered to the reporter promoters.

All Six members are expressed in many tissues during development, and they often show overlapping expression patterns (Kawakami et al., 2000; Kumar, 2009a). In mouse embryos, the expression of Six1, Six2, Six4 and Six5 is detected in skeletal muscle, kidneys, eyes and ears, with almost identical patterns (reviewed in (Kumar, 2009a)). Six3 and Six6 are similarly expressed in the developing eyes and forebrain (Jean et al., 1999; Oliver et al., 1995). Gene ablation experiments in the mouse model and gene mutations identified in various human genetic diseases suggested different requirement of these genes for embryonic development and adult regeneration.

Six1 knockout (KO) mice die at birth due to their inability to breath with the absence of diaphragm and severe rib malformations, in addition to severe hypaxial muscle hypoplasia (Laclef et al., 2003a). Further examination on other organs revealed that the mutant neonates lack kidney, thymus and show disorganized craniofacial structure, especially in the inner ears (Laclef et al., 2003b; Li et al., 2003; Ozaki et al., 2004; Xu et al., 2003). It is believed that Six1 regulates the proliferation and apoptosis of progenitor cells in those developing organs. The expression of Myod and Myog is largely delayed in the developing limb buds in the absence of Six1 (Laclef et al., 2003a). Interestingly, deletion of Six4 does not yield overt phenotypes (Ozaki et al., 2001). In human beings, Six5 is found to associate with myotonic dystrophy that is characterized by progressive muscle wasting and weakness, as well as cataract and cardiac conduction defects, with a late, adulthood onset (Boucher et al., 1995). However, Six5 KO mice show no evident abnormalities of skeletal muscle function (Klesert et al., 2000; Sarkar et al., 2000). Of note, Six1 and Six4 double-mutant mice display more severe phenotype than Six1 mutation alone: the expression of all MRFs is delayed and reduced further; the myotomes are more disorganized; the expression of migration factors is decreased, including Pax3, Lbx1 and c-Met, resulting in a lack of muscle cells that migrate from the myotome (Grifone et al., 2005).

Mutations in the homeodomain of Six3 are directly related to holoprosencephaly in human patients, likely due to the disruption of DNA binding (Wallis et al., 1999). Holoprosencephaly is a common birth defect of the forebrain, which often affects the facial features along with other birth defects. Inactivation of Six3 in the mouse results in postnatal death with the forebrain defect that is similar to that observed in human patients (Lagutin et al., 2003). Morpholino-mediated knockdown of Six3 in fish embryos also leads to deficiency in the forebrain and eye development (Carl et al., 2002). Based on gene ablation experiments in mice, Six6 is shown to

control the proliferation of progenitor cells during development of eyes and the pituitary gland, acting as a repressor of an anti-proliferation gene, Cdkn1b (p27Kip1) (Larder et al., 2011). Six6-null mice display infertility due to the reduced number of neurons that express gonadotropin-releasing hormone (GnRH) and the subsequent disruption of the fertility hormone axis (Larder et al., 2011; Xie et al., 2015). Similarly, combined gene deletion of Six1and Six4 in mice leads to defects in gonadal development including reduced proliferation of gonadal precursor cells and the impaired male differentiation pathway (Fujimoto et al., 2013).

Similar to the failure in kidney development in Six1 KO mice, Six2-null mice also develop severe renal hypoplasia (Self et al., 2006). Six2 is necessary for the maintenance of the mesenchymal progenitor population and to prevent premature epithelial differentiation (Self et al., 2006). Taken together, this information suggests that Six family transcription factors retained functional redundancy during evolution, so that compensation mechanism may take place among different family members (Klesert et al., 2000; Ozaki et al., 2001; Sarkar et al., 2000). The DNA binding specificity may explain their involvement in the development of different organs: the Six3/6 subgroup stands alone in neurogenesis; Six1/2 participates in nephrogenesis; and Six1 and Six4 take part in myogenesis.

1.6. Co-regulators of the Six family transcription factors

The *eyes absent* (*eya*) gene was identified in *Drosophila* as a requisite factor for the survival of eye progenitor cells during development (Bonini et al., 1998). In mammals, four homologues of *eya* were identified later, from Eya1 to Eya4 (Borsani et al., 1999; Xu et al., 1997; Zimmerman et al., 1997) (Figure 1.1E). In mammals, Eya proteins represent intrinsic protein phosphatase activities that is either specific to phospho-threonine or phosph-tyrosine (reviewed in (Rebay,

2016)). The catalytic activity of Eya proteins on phospho-tyrosine has been shown to be involved in transcriptional regulation of their target genes (Ahmed et al., 2012a; Eisner et al., 2015; Li et al., 2003; Wu et al., 2013a). Three proteins have been identified as functional substrates of their tyrosine phosphatase, including the histone H2A variant H2AX, the Protein Kinase C zeta (aPKCζ) and the estrogen receptor beta (ERβ) (reviewed in (Rebay, 2016)). In addition, Eya1 has been shown to catalyze the dephosphorylation of a threonine residue on the transcription factor Myc, which may have an impact on its protein stability (Xu et al., 2014).

The Drosophila *eya* protein is found localized constitutively in the nucleoplasm whereas the mammalian Eya proteins accumulate in the cytoplasm during embryonic development (Bonini et al., 1998; Fougerousse et al., 2002). Exogenously expressed Eya proteins that locate mainly in the cytoplasm can be translocated to the nucleus by the concomitant over-expression of Six proteins (Grifone et al., 2004; Ohto et al., 1999). In the nucleus, Eya proteins are not able to bind to DNA directly, but instead they are recruited to Six-bound DNA loci via a direct interaction between their conserved EYA domain and the SD domain of Six1 (Ikeda et al., 2002; Ohto et al., 1999). The crystal structure of the SD and HD domains of the human Six1 and the Eya domain of Eya2 has been resolved. Six1 directly interacts with Eya2 with a single α-helix in its SD (Patrick et al., 2013).

In *Drosophila*, forced expression of *so* and *eya* synergistically induce the formation of ectopic compound eye (Pignoni et al., 1997). In the mouse model, Eya1/Eya2 double-mutants exhibit similar phenotype to the Six1-null mouse: lack of diaphragm and the absence of limb muscles with impaired expression of MRFs in myotomes and limb buds, suggesting that their functions are genetically related (Grifone et al., 2007). In mammals, Six and Eya proteins have been shown to cooperatively regulate development and homeostasis in the context of different tissues

(Ahmed et al., 2012b; Grifone et al., 2004; Heanue et al., 1999; Zou et al., 2004). Six1 and Eya2 can synergistically induce the expression of Myod and Myog in somite explant culture (Heanue et al., 1999). The forced expression of Six1 and Eya1 favors muscle fiber type switch from slow-twitch to fast-twitch (Grifone et al., 2004). Moreover, Six1 and Eya1 have been shown to cooperatively induce the expression of target genes to drive neurogenesis in developing mammalian cranial placodes (Schlosser et al., 2008; Zou et al., 2004). Interestingly, Six1 and Eya1 are capable of converting fibroblasts into neuronal cells expressing neurogenic markers, a phenomenon that is enhanced by co-expression of Brg1 (Ahmed et al., 2012b).

Similar to the function of *so* and *eya* in *Drosophila*, a few genes exhibit similar properties: mutations on those genes lead to defective compound eye development while forced expression in other tissues can induce ectopic compound eyes, such as *ey* and *dac* (homologues of Pax and Dach in vertebrates, respectively) (Kawakami et al., 2000; Kumar, 2009b; Relaix and Buckingham, 1999). They constitute a regulatory network that synergistically controls eye development in fruit flies. Interestingly, the gene network is fairly conserved in vertebrates and their homologous genes also play pivotal roles in development of many other organs, in addition to eye development (Lang et al., 2007; Popov et al., 2010; Xu, 2013). For instance, in the peripheral sensory nervous system, members of the Six and Eya families play a central role in the specification of sensory progenitor cells and certain members of the Pax family act downstream to further confine the placode progenitors (Grocott et al., 2012). In addition, members of the Pax, Six and Eya families also constitute a gene network that orchestrates nephrogenesis (Brodbeck and Englert, 2004; Dressler, 2006).

Six proteins not only cooperate with Eya proteins and function as transcription activators, but also interact with transcriptional co-repressor complexes to repress gene transcription. Six3 is

shown to directly interact with Grg4 and Grg5, the murine counterparts of the *Drosophila* repressor *Groucho* (Smith and Jaynes, 1996; Zhu et al., 2002). Co-expression of Six3 and Grg4 (or Grg5) represses the transcription of a reporter gene driven by a Six3-responsive promoter and the active repression depends on their functional interaction (Zhu et al., 2002). Six6 has been shown to repress the expression of a reporter gene driven by the promoter of Cdkn1b ($p27^{kip1}$), which is a cyclin-dependent kinase inhibitor (Li et al., 2002). The Six6-mediated repression is further potentiated by the co-expression of Dach1 or Dach2. It is suggested that Dach1 interacts with Six6 and further recruits other co-repressors, such as Hdac3, a histone deacetylase that leads to more compact chromatin (Li et al., 2002). Similarly, the co-repressor Dach1 has been shown to interact with Six1 and repress the Six1-mediated activation of the reporter driven by the *Myog* promoter (Li et al., 2003). It is reasonable to speculate that the outcome of Six1 binding to target genes depends on its associated co-factors.

1.7. Target Genes of Six1

Chromatin immunoprecipitation (ChIP) combined with DNA microarray (ChIP-on-Chip) or followed deep DNA sequencing (ChIP-seq) has been widely performed to locate the global binding sites of transcription factors and the distribution of specific histone marks. Compared to ChIP-on-Chip, ChIP-seq generates more precise peaks with better signal-to-noise ratio. Six1 is a homeodomain transcription factor that is functional in many tissues (Ho et al., 2011). Taking the advantage of DNA microarray or DNA deep sequencing, the global targets of Six1 in different tissues have been explored by different research groups. During my Master's studies, I contributed to this by mapping the genome-wide binding sites of Six1 using ChIP-on-chip in C2C12 cells (Liu et al., 2010). Six1 genomic targets have also been mapped using ChIP-seq in embryonic kidneys (O'Brien et al., 2016) and in Wilms' tumors (Wegert et al., 2015).

Genes regulated by Six1 have been identified with loss of function assays, and in certain circumstances with gain of function assays (Ando et al., 2005; Liu et al., 2010; Niro et al., 2010; Yan et al., 2015). Global transcription has been studied using gene expression microarray (GEM), RNA deep sequencing (RNA-seq), cap analysis of gene expression sequencing (CAGE-seq) and global run-on sequencing (GRO-seq) (Li et al., 2016). GEM and RNA-seq examine the total stable RNA, including or excluding ribosomal RNA. RNA-seq allows better detection of low abundance transcripts and identification of biogical isoforms compared to GEM (Zhao et al., 2014). CAGE-seq captures the capped RNA in a transcriptome and is broadly explored to study transcription initiation. GRO-seq is incorporated with a specific enrichment of nascent RNA, providing better identification of dynamic transcriptional activity (Li et al., 2016). GEMs combined with RNA interference that specifically targets Six1 have discovered a set of genes whose expression is under the control of Six1 in fibroblasts (Adrados et al., 2015) and in myoblasts (Liu et al., 2010).

A number of genes are potentially direct targets of Six1. During myogenesis, Six1 is shown to regulate the expression of Pax3 via direct binding to its hypaxial enhancer (Grifone et al., 2007). Similarly, Six1 modulates Myod expression through direct binding to the DRR in differentiating myoblasts (Le Grand et al., 2012; Relaix et al., 2013). Six1 can also cooperatively bind to the promoter of Myog with Myod and regulate its expression through a conserved MEF3 DNA motif (Le Grand et al., 2012; Spitz et al., 1998). In addition, Six1, together with Pax3, controls the expression of Myf5, through direct binding on its 145-base pair (bp) element that is 57.5 kb upstream of the Myf5 gene (Bajard et al., 2006; Giordani et al., 2007). Furthermore, Six1 has been shown to regulate the expression of a large set of genes that are related to muscle structure, and Six1 binding on regulatory elements of many of these genes has been validated by ChIP in

C2C12 cells, such as Tnnc1 and components of the muscle acetylcholine receptor, Chrng and Chrna1 (Liu et al., 2010). Lastly, Six1 and Six4 have been shown to be necessary for the expression of fast-twitch muscle genes, based on gene profiling assays in developing trunk muscles (Niro et al., 2010).

1.8. The role of Six1 in health and diseases

As mentioned above, the functional Six/Eya complex is essential for mammalian development. Disruption of the binary complex, such as genetic mutations in *SIX1* or *EYA1*, is associated with the Branchio-oto-renal (BOR) syndrome (Kochhar et al., 2007, 2008; Patrick et al., 2009). The BOR syndrome is an autosomal dominant disorder with incomplete penetrance, showing variable defects in branchial arches, otic structure, and renal formation. *EYA1* is the most common gene that is linked to the BOR syndrome (Kochhar et al., 2007). An update of known *EYA1* mutations has been reported (http://deafnessvariationdatabase.org/letter/e). *SIX1* mutations that correlate with the BOR syndrome exhibit defects either in DNA binding through the HD domain or in translocating EYA proteins into the nucleus through the SD domain (Patrick et al., 2009; Ruf et al., 2004). Gene ablation studies in the mouse model provide the functional connection between the BOR syndrome and the Six1/Eya1 complex. Similar phenotypes in affected BOR patients and in mouse mutants reflect the functional conservation between these two families.

Satellite cells are the primary contributor of muscle regeneration in both physical and pathological conditions of muscle wasting, such as ageing and muscular dystrophy (Grand and Rudnicki, 2007). Six1 is transcribed in quiescent satellite cells and its expression is further induced upon satellite cell activation (Le Grand et al., 2012; Yajima et al., 2010). When Six1 is conditionally ablated in satellite cells, both muscle differentiation and muscle regeneration are

significantly impaired, in spite of increased satellite cell self-renewal (Le Grand et al., 2012). The implication of Six1 in satellite cell homeostasis suggests a potential role of Six1 in muscle wasting.

In addition to its physiological role in many tissues, over-expression of SIX1 correlates with many cancer types, such as ovarian cancer, prostate cancer, and lung cancer, partly through direct activation of cyclin A1 (Christensen et al., 2008; Coletta et al., 2004; Reichenberger et al., 2005). In addition, Six1 is implicated in the metastasis of rhabdomyosarcoma, and this is in part through transcriptional regulation of Ezrin (Yu et al., 2004, 2006). Besides the BOR syndrome, the recurrent mutation Q177R in the HD of SIX1 and SIX2 is often present in Wilms' tumors (Walz et al., 2015; Wegert et al., 2015). The Q177R mutant exhibits lower affinity at some DNA binding sites, while having higher affinity at new sites that are not bound by the wild-type SIX1 (Wegert et al., 2015). c-Myc, a transcription factor whose persistent expression is found in many cancers, is also a target of Six1 (Li et al., 2003; Yu et al., 2006). Furthermore, Six1 and Eya2 interaction is required for Six1-mediated tumor metastasis (Patrick et al., 2013). The physiological and oncogenic roles of Six1 in different tissues are recently reviewed and a few approaches have been suggested to therapeutically target the Six1/Eya complex to potentially inhibit tumor progression (Blevins et al., 2015).

1.9. The regulation of Six1

There are several layers of Six1 regulation: transcriptional, post-transcriptional, translational, and post-translational modulations. DNA methylation at the promoter region is negatively correlated with gene transcription. It is suggested that Six1 is under the control of DNA methylation at its promoter region (Wu et al., 2013c). Using P19 embryonic carcinoma cells as a model, studies

suggest that Six1 expression is transcriptionally regulated by Pax3, Wnt/beta-catenin and retinoid signalling pathways, suggesting that these pathways are at play for the induction of Six1 expression during development (Petropoulos and Skerjanc, 2002; Ridgeway and Skerjanc, 2001; Wong et al., 2013). At the post-transcriptional level, microRNA-185 (miR-185) is shown to bind to the three prime untranslated region (3'-UTR) of Six1 and negatively regulates a reporter gene expression driven by the Six1 3'-UTR in 293 cells (Imam et al., 2010). In many cancers, miR-185 is expressed at lower levels and this is associated with higher levels of Six1 (Imam et al., 2010). Six1 protein levels are regulated according to the cell cycle phase and its levels stay low in G_1 phase, gradually accumulate at the G_1/S boundary and peak in G_2/M transition (Christensen et al., 2007). Aside from the regulation of Six1 mRNA or protein levels, the function of the protein has been shown to be controlled by various means. Six1 is a nuclear phosphoprotein and becomes hyperphosphorylated at mitosis (Ford et al., 2000). In addition to association with various co-factors, the ability of Six1 to bind to DNA is affected by CK2, a serine/threonine protein kinase, whereby phosphorylation reduces its binding affinity (Ford et al., 2000).

1.10. Chromatin and Six1-mediated transcription

1.10.1. Six1-interacting chromatin modifiers

DNA is assembled and compacted with histone proteins to form nucleosomes, which are the basic unit of chromatin. Chromatin is the scaffold where DNA is compacted into high order structure, acting as a barrier to transcription. Chromatin structure must therefore be accessible to allow transcription factor binding and gene expression. There are various ways this can happen. The core histones are subjected to a variety of post-translation modifications which in turn have an impact on chromatin structure. In many cases, there are correlations between the presence of

these histone modification “marks” and gene regulation, and in fact for some of them we know they can influence gene transcription. In addition, active regulatory elements, such as promoters and enhancers, are characterized as nucleosome-depleted regions (NDRs) (Ercan et al., 2004; Lee et al., 2004a). NDRs are enriched for transcription factor binding motifs which facilitate the recruitment of tissues-specific transcription factors and subsequently the general transcription machinery (Bell et al., 2011; Bernstein et al., 2004; Hartley and Madhani, 2009; Radman-Livaja and Rando, 2010). The chromatin can be regulated with different mechanisms: 1) histone variants that give rise to less stable nucleosomes (Chen et al., 2013; Jin et al., 2009); 2) histone modifications that generate a less compact association between the core histones and DNA (Couture and Trievel, 2006; García-González et al., 2016; Hon et al., 2009); 3) ATP-dependent chromatin remodeling complexes that actively slide or evict nucleosomes (Hargreaves and Crabtree, 2011; Narlikar et al., 2002; Vignali et al., 2000); and 4) occupancy by RNA pol II that maintains an open chromatin structure by preventing the deposition of nucleosomes (Adelman and Lis, 2012; De Santa et al., 2010).

As a transcription factor, Six1 has been shown to interact with components of chromatin remodeling complexes in the process of facilitating transcription. Six1 and Eya1 interact directly with Baf170 and Brg1 to cooperatively induce a neurogenic conversion in 3T3 fibroblasts (Ahmed et al., 2012b). Baf170 is a functional core subunit of the SWI/SNF complex (Phelan et al., 1999). An enzymatically functional Brg1 is required for the EYA1- and SIX1-induced neurogenic conversion as a catalytically inactive Brg1 prevents it (Ahmed et al., 2012b). Six1 is also shown to recruit CREB-binding protein (CBP) through its interaction with Eya3 (Ikeda et al., 2002; Li et al., 2003). CBP is a protein acetyltransferase, which shares a great structural similarity with E1A binding protein p300. CBP and p300 have been shown to acetylate a variety

of proteins, including histone H3, histone H4 and non-histone proteins (Wang et al., 2008). The intact acetyltransferase activity of p300 is required for the myogenic commitment, through its transcriptional regulation of Myod and Myf5 (Roth et al., 2003). Interestingly, Myod has also been shown to be acetylated by p300/CBP and pCAF (p300/CBP-associated factor), and its acetylation contributes to enhanced transcriptional function of Myod (Dilworth et al., 2004; Polesskaya et al., 2000). Despite significant achievements in understanding mechanisms of the chromatin remodeling, it is still under intensive investigation how transcription factors, such as Six1, gain access to compact chromatin and maintain an open chromatin structure on their regulatory elements including in the context of myogenesis.

1.10.2. Properties of Transcriptional Enhancers

Six1 regulates target gene expression through direct, or perhaps also indirect, binding to their promoters and enhancers. In general, active promoter regions share the following characteristics: nucleosome-depletion, proximity to a transcription start site, deposition of active histone marks (H3K4me3 and acetylation of the lysine 27 at the histone H3 (H3K27ac)), incorporation of unstable histone variants (H3.3 and H2A.Z), occupancy of general and sequence-specific transcription factors, and binding of RNA pol II (Kim and Shiekhattar, 2015) (Figure 2.1A). Enhancers are non-coding DNA modules that are distal to the promoter region as *cis*-elements. In general, enhancers locate long linear distances, ranging from a few kilo bases (kb) to a few million bases (Mb), from their target promoters and regardless of their orientation, or even from different chromosomes (Bulger and Groudine, 2010; Heintzman et al., 2007; Kim and Shiekhattar, 2015). Enhancers can be bound by transcription factors that positively regulate their target gene transcription (Stadhouders et al., 2012a). Given the fact that enhancers are distant from their target promoters, a model of enhancer-promoter looping has been proposed to bring

them in proximity as a general mechanism. Enhancer-promoter looping is believed to be mediated, at least partially, by CCCTC-binding factors (CTCFs), cohesins and/or the Mediator complex (reviewed in (Plank and Dean, 2014)). Similar to promoters, enhancers are also evolutionarily conserved. Functional enhancers share many features with active promoters, such as nucleosome-depletion, accumulation of H3K27ac, incorporation of histone variants, occupancy of transcription factors, and binding of RNA pol II (Chen et al., 2013; Heintzman et al., 2007; Kim and Shiekhattar, 2015; Lee et al., 2004a). Enhancers also display unique characteristics, such as enrichment of H3K4me1/2 and occupancy of p300 and c-Jun (Blum et al., 2012; Heintzman et al., 2007; Visel et al., 2009) (Figure 2.1A). RNA-seq, GRO-seq and CAGE-seq have revealed that active enhancers are also transcribed into what are known as enhancer-associated RNAs (eRNA) (reviewed in (Kim et al., 2015a)). eRNAs can be bi-directionally or uni-directionally transcribed from enhancer elements. In general, eRNAs are unstable, short (< 2 kb), expressed at a low level, mostly unspliced and they lack poly-adenylate tails (Kim et al., 2015a).

Similar to the genes they control, eRNAs are transcribed in a tissue and time-specific manner, especially for genes that are important for development or differentiation (Lam et al., 2014). eRNAs have been studied mostly in the context of stimulus-responsive genes, such as in breast cancer cells receiving estrogen treatment or in macrophages following exposure to lipopolysaccharides (LPS). Such strategy takes the advantage of synchronous activation and potent induction of transcription from the baseline. It is technically harder to study eRNAs in the context of development. Upon stimulus, the production of eRNA often precedes or at least synchronizes with the target gene regulation (Arner et al., 2015; Hsieh et al., 2014; IIott et al., 2014; Li et al., 2013a). Recently a systematic study of CAGE on a number of cell types has

detected dynamic transcription change along the stimuli time course, finding that eRNA transcription is the most common early response during cell state transition (Arner et al., 2015). Accumulating evidence suggests that eRNAs are involved in transcriptional regulation of their target genes. Genomic data discovered a clear correlation between the production of eRNAs and the transcription of their target genes (Hah et al., 2013; Kim et al., 2010, 2015b). Loss-of-function of several eRNAs by RNA interference has shown detrimental effect on the recruitment of RNA pol II at promoters and the expression of their target genes (Kim et al., 2015b; Lam et al., 2013; Li et al., 2013a; Melo et al., 2013). In addition, enhancers that express eRNA are more enriched for the enhancer-promoter looping than those without eRNA transcription (Hah et al., 2013). eRNAs may help to establish the enhancer-promoter looping, probably through recruitment of the Mediator coactivator complex (Hsieh et al., 2014; Lai et al., 2013; Li et al., 2013a; Melo et al., 2013).

In addition, some studies also suggest that the action of transcription or the presence of the transcription machinery is essential for the expression of target genes and establishing the chromatin architecture of functional enhancers (reviewed in (Kim et al., 2015b)). For example, inhibition of transcription elongation by different chemicals leads to reduced deposition of H3K4me1/2 at *de novo* enhancers that are stimulus-responsive to the TLR4 signaling (Kaikkonen et al., 2013). The transcription elongation complex has been shown to carry chromatin modifiers through direct interaction (reviewed in (Selth et al., 2010)). For instance, pCAF, a protein acetyltransferase, has been shown to associate with elongation-competent RNA pol II and facilitate productive transcription (Cho et al., 1998; Obrdlik et al., 2008). It has been proposed that functional transcription at enhancers could have a positive impact on their chromatin architecture and degree of DNA accessibility. The connection could be through the

eRNA transcripts which may interact with unknown co-factors or through chromatin modifiers that are associated with the transcription elongation complex. The two mechanisms may converge at some enhancers.

A few approaches have been used to study chromatin accessibility globally, including DNAse-seq, MNase-seq, FAIRE-seq and ATAC-seq. DNAse-seq and MNase-seq involve in enzymatic digestion of exposed DNA. MNase-seq produces signals of nucleosome-protected DNA. DNAse-seq, ATAC-seq and FAIRE-seq are enriched for nucleosome-free DNA (Meyer and Liu, 2014). FAIRE-seq is enzyme-free but also less sensitive, exhibiting higher background noise (Davie et al., 2015; Simon et al., 2012). ATAC-seq measures transposase-accessible chromatin. Compared to other approaches, ATAC-seq requires low amount of biological samples (Buenrostro et al., 2015).

1.10.3. Transcription elongation and p-TEFb

Promoter-proximal pausing of RNA pol II is a general mechanism of gene expression regulation which is applied to around 30% of all genes across different species (Adelman and Lis, 2012). Accumulating studies have shown that poised genes are already transcriptionally initiated at the promoter-proximal region for about 20-60 bp downstream of transcription start site (Gaertner and Zeitlinger, 2014; Jonkers and Lis, 2015). RNA pol II is stalled by negative elongation factors and transcribed RNA is degraded quickly by the system due to its instability. The transcription pausing is actively maintained by the cooperative interactions of DRB Sensitivity Inducing Factor (DSIF) and negative elongation factor (NELF) (Zhou et al., 2012). Although no functional transcripts are produced, transcription pausing primes the gene for prompt activation (Jonkers and Lis, 2015). The transcription pause and elongation are mediated by the phosphorylation

patterns of RNA pol II, DSIF and NELF. The largest subunit of RNA pol II contains an unusual extension at its carboxyl-terminus, named as the C-terminal repeat domain (CTD). The CTD serves as a structural scaffold mediating its interaction with a series of nuclear factors. The CTD comprises 25-52 tandem heptapeptide (YSPTSPS) and its phosphorylation pattern is associated with the function of RNA pol II (Phatnani and Greenleaf, 2006). Paused RNA pol II in the initiating complex is phosphorylated at Serine 5 residues and elongation-engaged RNA pol II is further phosphorylated at Serine 2 residues (Phatnani and Greenleaf, 2006). Cdk7 and Cyclin H play an essential role in phosphorylating serine 5 residues (Morales and Giordano, 2016; Shapiro, 2006), while phosphorylation of Serine 2 residues (Ser2P) is catalyzed by the positive transcription elongation factor b (p-TEFb) (Marshall et al., 1996). p-TEFb is composed of a regulatory cyclin subunit (Cyclin T1 or Cyclin T2) and the catalytic cyclin-dependent kinase 9 (Cdk9) subunit (Fu et al., 1999; Peng et al., 1998). NELF is dislodged from the complex after being phosphorylated by Cdk9 to release the transcription pausing (Fujinaga et al., 2004). DSIF is also phosphorylated by p-TEFb and then travels with RNA pol II during transcription elongation (Yamada et al., 2006). In summary, p-TEFb prompts productive transcriptional elongation.

Acting as a sort of “gas pedal” of transcriptional elongation, the function of p-TEFb is under tight control in many ways. Cdk9 and its partner cyclins can be regulated by transcriptional, translational and post-translational modification, as reviewed in (Peterlin and Price, 2006; Zhou et al., 2012). In mouse and human tissues, two isoforms of Cdk9 (42 kD and 55 kD) have been identified, which are generated from different transcription start sites (Shore et al., 2003). The Cdk9-55 has been shown to play an important role in muscle regeneration (Giacinti et al., 2008). Many inhibitors have been discovered to inhibit or block the catalytic activity of Cdk9, such as

DRB and Flavopiridol (Blachly and Byrd, 2013). Flavopiridol is the first CDK inhibitor that was applied to clinical trials and demonstrated some anti-cancer effect in hematologic malignancy (Bose et al., 2013). Flavopiridol originates from an Indian plant, *Dysoxylum binectariferum*. Flavopiridol is a pan-CDK inhibitor and its targets include Cdk9 and Cdk7, while it shows much greater inhibitory activity on Cdk9 than on Cdk7 (Chao and Price, 2001; Fischer and Gianella-Borradori, 2005; Kim et al., 2000; Krystof and Uldrijan, 2010). When applied to cells, Flavopiridol can potently inhibit RNA pol II-mediated gene transcription, including the production of mRNAs and eRNAs (Chao and Price, 2001; Kaikkonen et al., 2013). This indicates that Cdk9 and possibly transcriptional pause regulate the production of eRNAs, in addition to mRNAs.

A few proteins have been shown to recruit Cdk9 to target genes via protein-protein interaction. Brd4, a member of the BET (bromodomain and extra terminal domain) family, interacts with a variety of proteins, including acetylated histones and DNA-binding transcription factors (Asangani et al., 2014; Strahl and Allis, 2000; Wu et al., 2013b). Brd4 and p-TEFb form a protein complex *in vivo*, and Brd4 is required for the recruitment of Cdk9 to activate the subsequent transcription of target genes (Jang et al., 2005; Liu et al., 2013b; Yang et al., 2005). In muscle cells, p-TEFb and Mef2 proteins synergistically stimulate Mef2-dependent reporter transcription, probably through their direct interaction (Nojima et al., 2008). Similarly, Myod has been shown to interact with Cdk9 directly and they co-occupy a few regulatory elements of myogenic genes, including promoters and enhancers (Giacinti et al., 2006; Simone et al., 2002). The function of p-TEFb is clearly involved in myogenesis (Giacinti et al., 2006, 2008; Marchesi et al., 2013; O'Brien et al., 2012). However, the mechanisms for gene-specific recruitment of Cdk9 have not been intensively studied during myogenesis.

Rationale and Hypotheses

Firstly, the reported TRANSFAC MEF3 motif is inflexible given the fact that it was derived from regulatory elements of a few myogenic genes (Parmacek et al., 1994; Spitz et al., 1997). Another Six1-binding DNA motif was deduced based on *in vitro* protein-binding microarrays and it is very different from the TRANSFAC MEF3 motif. Taking the advantage of genome-wide binding sites of Six1, **I hypothesized that Six1 binds to greater variety of DNA sequences than what is reflected by the TRANSFAC MEF3 motif.**

Secondly, Six1 is functionally upstream of Myod during embryonic muscle development. However, it remained unclear whether Myod was still under the control of Six1 during muscle regeneration and what modes of regulation were in molecular details. **I hypothesized that Six1 directly regulates the expression of Myod in muscle progenitor cells.**

Lastly, the CER is a functional enhancer of Myod during muscle regeneration. Six1 regulates the expression of Myod through direct binding to two MEF3 sequences at the CER. Nevertheless, it remained unresolved through which mechanisms Six1 exerted its transcriptional modulation on the Myod gene. **I hypothesized that Six1 promotes Myod transcription through recruiting the p-TEFb complex to the CER.**

To address these hypotheses, I have performed molecular biology experiments combined with bioinformatics analyses to determine the DNA binding property of Six1 and the underlying mechanisms how Six1 regulates Myod expression.

The following three manuscripts focus on **1)** Six1 recognizes a variety of DNA sequences; **2)** Six1 regulates the expression of Myod in muscle progenitor cells; **3)** Six1 recruits Cdk9 to the Myod CER to maintain Myod expression.

CHAPTER **2.** DISCOVERY, OPTIMIZATION AND VALIDATION OF AN OPTIMAL DNA BINDING SEQUENCE FOR THE SIX1 HOMEODOMAIN TRANSCRIPTION FACTOR

The following manuscript is published in *Nucleic Acids Research*.

Liu, Y., Nandi, S., Martel, A., Antoun, A., Ioshikhes, I., & Blais, A. (2012). Discovery, optimization and validation of an optimal DNA-binding sequence for the Six1 homeodomain transcription factor. *Nucleic Acids Research*, *40*(17), 8227–8239.

I generated genomic Six1 binding profiles using ChIP-on-Chip at three time points during C2C12 myoblasts differentiation for Figure 2.1, Figure 2.2 and Table 2.1. Another student (Andre) helped with Six1 purification. Using purified Six1 protein, I performed gel shift assays (EMSA) on variant sequences that contain MEF-like sequences, and made Figure 2.3B and 2.3C. Based on the flexibility of the MEF3 DNA motif, Dr. Alexander Blais helped to design different sequences in Table 2.2. I performed gel shift assays and data analysis to make Table 2.2. I compared the detection sensitivity between Six1-opti (computationally optimized) and Six1_MB+MT (discovered from real data), and made Figure 2.5B and 2.5C. Lastly, I contribute to the editing of the manuscript.

Discovery, optimization and validation of an optimal DNA binding sequence for the Six1 homeodomain transcription factor

Yubing Liu[1,2], Soumyadeep Nandi[1,2], André Martel[1,2], Alen Antoun[1,2], Ilya Ioshikhes[1,2] and Alexandre Blais[1,2]*

Affiliations

1, Ottawa Institute of Systems Biology

2, Biochemistry, Microbiology and Immunology Department, Faculty of Medicine, University of Ottawa

Keywords: Six1, homeodomain, electrophoretic mobility shift assays, position weight matrix, PWM optimization

Short title: DNA binding properties of Six1.

Liu, Y., Nandi, S., Martel, A., Antoun, A., Ioshikhes, I., & Blais, A. (2012). Discovery, optimization and validation of an optimal DNA-binding sequence for the Six1 homeodomain transcription factor. *Nucleic Acids Research*, *40*(17), 8227–8239.

ABSTRACT

The Six1 transcription factor is a homeodomain protein involved in controlling gene expression during embryonic development. Six1 establishes gene expression profiles that enable skeletal myogenesis and nephrogenesis, among others. While several homeodomain factors have been extensively characterized with regards to their DNA binding properties, relatively little is known of the properties of Six1. We have used the genomic binding profile of Six1 during the myogenic differentiation of myoblasts to obtain a better understanding of its preferences for recognizing certain DNA sequences. DNA sequence analyses on our genomic binding dataset, combined with biochemical characterization using binding assays, reveal that Six1 has a much broader DNA binding sequence spectrum than had been previously determined. Moreover, using a position weight matrix optimization algorithm, we generated a highly sensitive and specific matrix that can be used to predict novel Six1 binding sites with highest accuracy. Furthermore, our results support the idea of a mode of DNA recognition by this factor where Six1 itself is sufficient for sequence discrimination, and where Six1 domains outside of its homeodomain contribute to binding site selection. Together, our results provide new light on the properties of this important transcription factor, and will enable more accurate modeling of Six1 function in bioinformatic studies.

INTRODUCTION

A defining characteristic of transcription factors (TFs) is their ability to recognize and bind to specific DNA sequences in the genome, within the regulatory region of the target genes they control. Multiple structural classes of TFs exist, among which are homeodomain factors. The primary function of the homeodomain is DNA binding. This group of regulatory factors were first discovered due to their involvement in homeotic conversions in *Drosophila*, but were later found to exist in essentially all eukaryote species, from yeast to humans. Bioinformatics analyses indicate that 235 homeodomain TFs are encoded in the human genome (not counting splice variants and pseudogenes)(1). This diversity is manifested by various expression profiles, protein domain composition and interaction partners. However, a large number of homeodomains seemingly recognize the same DNA sequence TAAT, or close variants, leading to the question of how redundancy in binding site selection is avoided by these proteins. The prototypical 60 amino acids long homeodomain assumes a three-dimensional structure composed of an unstructured N-terminal arm followed by three alpha-helices. The N-terminal arm and the third helix contribute most of the DNA sequence binding specificity, and their amino acid composition are thought to contribute to this property (2, 3). Two recent large-scale surveys considered the question of binding specificity of homeodomain TFs from mouse (4) or Drosophila (5). It was found that indeed, when homeodomains are considered globally, their amino acid sequence correlates with their DNA sequence binding preferences within and adjacent to the TAAT core. However, domains residing outside of the homeodomain can also influence DNA binding, either through direct DNA contacts or interaction with dimerization partners (6, 7, 8, 9). Therefore, it remains to be established whether homeodomain TF binding site predictions based solely on *in vitro* DNA binding preferences are sufficiently accurate to allow to predict which target genes they regulate.

The Six family of homeodomain TFs is conserved from flies to humans, and in mammalians counts six members, from Six1 to Six6. Like most homeodomain TFs, they are involved in controlling the development of various tissue types; for instance, Six1 and Six4 are involved in the development of the eyes, ears, kidneys and skeletal muscle (reviewed in (10, 11)). We became interested in this group of TFs when we found that the DNA motif they recognize, the MEF3 sequence element, is enriched within the promoter region of Myogenic Regulatory Factors (MRFs) target genes, in muscle precursor cells (myoblasts) (12). The MEF3 element is a phylogenetically conserved motif (DNA consensus TCAGGTTTC) that was originally identified within the regulatory region of only a few muscle-specific genes (13, 14). The Six1 and Six4 members of the Six family, two factors essential to myogenesis, were subsequently found to bind specifically to the MEF3 sequence within the myogenin (Myog) gene promoter and activate this gene's expression during embryogenesis (15, 16, 17, 18, 19). MRFs control myogenesis by activating the expression of a large cohort of genes necessary for differentiation and function of muscle cells, and in some instances they accomplish this task by cooperating with transcription factors of the Mef2 and Pbx families (20, 21). The connection between the MRF target genes and MEF3 sites led us to postulate that Six family members can also cooperate with the MRFs to regulate myogenesis. This was confirmed by the genome-wide identification of Six1 target genes in mouse myoblasts and myotubes, and by functional assays which showed that Six factors can activate transcription with the MRFs in a synergistic fashion (22).

The wealth of DNA binding information contained within our genomic Six1 binding profile (22) gave us the opportunity to examine the DNA sequence binding preferences of this TF in myoblasts and myotubes and to analyze it in the context of the previously identified MEF3 sequence motif. Here, using *de novo* motif finding and position weight matrix (PWM)

optimization, we report that Six1 has a broader than anticipated DNA binding profile, which extends well beyond the canonical MEF3 consensus sequence. *In vitro* binding data corroborate these *in vivo* DNA binding sequence preferences, suggesting that sequence-specific DNA binding by Six1 does not require interaction with dimerization partners. However, we find a discrepancy between our results and those of an *in vitro* DNA binding screen for the Six1 homeodomain, performed by others (4), suggesting that Six1 homeodomain sequence is not the sole determinant of its binding to DNA.

MATERIALS AND METHODS

De novo motif finding

The Amadeus program was used for *de novo* DNA sequence motif discovery (23). The "bound" sets corresponded to the sequences bound by Six1 in mouse myoblasts (MB, total of 1022 sites) or myotubes (MT, total of 1853 sites) with an FDR less than 10% (22). These two sets of bound sequences were broken down into 3 (for MB) or 5 (for MT) randomly assigned subgroups, respectively (*i.e.* MB-A, MB-B, MB-C and so on). The purpose of this sub-grouping was to obtain one subgroup for motif discovery, and to reserve the other sequences for subsequent validation of the discovered motifs. The Amadeus program was run on each of these 8 subgroups, performing "large" searches of 12 base pairs motifs and examining both DNA strands. For these 8 searches, the background sequence set corresponded to a randomly selected subset of the genomic regions surveyed in our ChIP-on-chip analyses (20% of surveyed loci, approximately 108 Mb of sequence). Both sets of sequences were repeat-masked using RepeatMasker (24). The top ranking sequence motif (lowest corrected p-value after 25 cycles of boot-strapping where "bound" and "background" sequences are randomly interchanged) was retained. Searches for motifs less than 12 base pairs in length were also run and yielded very similar motifs (not shown).

Finally, an averaged PWM (Six1_MB+MT) was calculated by aligning the 8 PWMs and averaging the frequencies at each position. PWMs were represented graphically using the Weblogo program (25).

Motif abundance within sets of sequences was estimated using the CisGenome program (26). The number of PWM "hits" within "bound" and "background" sequences was determined with a likelihood ratio setting of 500. The bound set of sequences for this purpose was composed of those excluded from the initial motif discovery (*e.g.* for the matrix identified using subgroup MB-A, the sequences in MB-B and MB-C were combined and used to test its performance). Here the background sequences were a distinct set of 20% of all surveyed loci. The relative enrichment score (the ratio of frequency of PWM hits among bound regions over that in background regions) was calculated. We also calculated the cumulative hypergeometric probability, the chance of finding at least a certain number of hits to a PWM among the bound regions (sample) given that a certain number of hits exist among the background (population) sequences (sample and population sizes expressed as total base pair length). Finally, searches were also repeated on the top 5% most phylogenetically conserved portion of the sequences.

PWM optimization

The methods described by Staden (27) and Bucher (28) were utilized to calculate the PWM. We adopted the base frequencies reported by Amadeus in the *de novo* motif searches (above). These base frequencies were converted into odds scores by dividing the frequencies by expected frequency which is calculated from the Database of Transcription Start Sites (DBTSS) for mouse using the formula described in (27):

$$e_{bi} = \frac{\sum_{i=1}^{L} n_{bi}}{L}$$

where *b* is one of 4 nucleotides (A,C,G or T) at position *i*, n_{bi} is the number of times base *b* occurs at the i^{th} position of the motif and *L* is the length of the sequence.

The sequences used from DBTSS were of the length 1201 base pairs (-1000 to +201 from the TSS) and were aligned with respect to the TSS.

The weight for each position of the matrix is derived using the formula described in (29):

$$w_{bi} = \ln\left(\frac{n_{bi}}{e_{bi}} + s_i\right) + c_i$$

where *b* is one of the 4 nucleotides, n_{bi} is the number of times base *b* occurs at the i^{th} position of the motif, c_i is a constant providing column maximum value to be zero, s_i is a smoothing parameter preventing the logarithm of zero (or too small a value).We adopted the criteria as:

$s_i = 0$ if the first term under logarithm in Formula is larger than $0.01 \times \frac{n}{4 \times e_{bi}}$ and

$s_i = 0.01 \times \frac{n}{4 \times e_{bi}}$ otherwise, where $n = \sum_{b=1}^{4} n_b$

To calculate the similarity score for a specific sequence within the PWM we used the formula as:

$$S = \sum_{i=1}^{L} w_{bi}$$

where *L* is the length of PWM, w_{bi} is the log-odds weight of nucleotide *b* at position *i* in the PWM. To optimize the derived PWM we have used correlation coefficient (CC) also known as

Simple Matching Coefficient (SMC) (30) as the objective function. This function takes into account sensitivity and specificity of the predicted TFBS. The process of optimization started with evaluating the performance of the PWM by calculating CC at each cutoff. The CC is calculated as:

$$CC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP) \times (TP + FP) \times (TN + FN)}}$$

CC is calculated for each cutoff starting from a very stringent threshold and relaxing the threshold until we get the maximal CC. To calculate the CC we have divided the sequences into two different datasets depending on the binding preference of Six1 in myoblasts and myotubes. The sequences where Six1 is found experimentally to bind are regarded as positive and the sequences where Six1 did not bind are regarded as negative dataset. These two datasets were utilized to find out the four parameters to calculate CC: true positives, true negative, false positives and false negatives (TP, TN, FP and FN respectively). From the above mentioned datasets we designated TP as the number of sites from the experimental positive dataset positively identified by the PWM with a given cutoff, and we regarded any sites computationally identified from the experimental negative dataset where Six1 do not bind as FP. TN is calculated as the difference between the total number of sequences in the negative dataset and FP, while FN is calculated as the difference between the total number of sequences in the positive dataset and TP. The FN was calculated assuming each sequence in the positive datasets should have at least one binding site. The above step is repeated for each cutoff with the increment of 0.1, and maximal CC with respective cutoff was identified. We selected the corresponding cutoff and we further refined the performance of the PWM as follows. The PWM with the previously optimized cutoff shows sensitivity of 54% and specificity of 76%. This matrix becomes our

initial PWM for the next step of optimization. Again we start with the stringent cutoff and refined the motif list used to build the PWM at each 0.1 increment in the cutoff. At each cutoff, the matrix was used to find motifs from the positive dataset. The list of motifs thus obtained from positive dataset was utilized to build the new PWM. With this PWM, we searched motifs in the negative dataset and compared the search results with those obtained from the positive dataset. We subtracted the motifs from the positive search list and rebuilt the PWM with the remaining list. The new PWM was evaluated with the function CC. We repeated the latter step for a large range of cutoffs, from stringent to relaxed, and selected the cutoff where the CC attained the maximum. The resulting matrix provides better discrimination between the positive (bound) and negative (surveyed) datasets. The Respective algorithm of PWM optimization is hence named "Bound/Surveyed sequence Discrimination" (BSD) algorithm. The PWM was optimized using binding data for Six1 in C2C12 myoblasts and in fully differentiated myotubes. For further validation on an independent dataset, we used genomic regions bound by Six1 only at 24 hours of differentiation, but not bound in myoblasts nor in myotubes. This corresponds to a total of 187 DNA sequences with an average length of 1061 bp which have no overlap with the myoblasts and myotubes datasets.

Recombinant Six1 purification

The full-length coding sequence of mouse Six1 was amplified from C2C12 cells and cloned in frame in the pHIS2 plasmid, which codes for an N-terminal hexa-histidine tag followed by a linker region. The protein was produced in the E. coli STAR strain. Cells were grown to an OD^{600} of 0.6 and induced to produce the protein with 0.1 mM Isopropyl β-D-1-thiogalactopyranoside for 1h30 at 37°C. Cells were pelleted, resuspended in binding buffer (50 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM DTT, 10 mM imidazole and 2 mM

phenylmethylsulfonyl fluoride) and sonicated using a microtip sonifier. The lysate was spun at 17,000 x *g* for 15 minutes at 4°C, and the supernatant was applied to Nickel-Sepharose beads (GE Healthcare). Beads were washed in binding buffer containing 40 mM imidazole, and elution was carried out using binding buffer containing 500 mM imidazole. The eluate was immediately bound to heparin-Sepharose beads (GE Healthcare). The beads were washed with wash buffer (50 mM Tris pH 8.0, 650 mM NaCl, 0.5 mM DTT and 2 mM phenylmethylsulfonyl fluoride) and eluted in a similar buffer containing 750 mM NaCl. The eluate was dialyzed for 18 hours against a similar buffer reduced to 150 mM NaCl and containing no imidazole, then concentrated using Amicon Ultra (30 kD cut-off, Millipore), aliquoted, frozen in liquid nitrogen and stored at -80°C. Coomassie blue staining of the purified Six1 protein indicated an estimated purity of 90% (Figure 3A). For work with the homeodomain of Six1, amino acids 110 to 201 of mouse Six1 were cloned using the same strategy as for the full-length protein. This Six1-HD protein therefore contains the homeodomain and 15 amino acids of flanking sequence on each side, which conforms to what Berger et al. have used. Purification of the histidine-tagged Six1-HD protein from E. coli was performed in the same was as for the full-length protein, except that dialysis and concentration were performed using devices with smaller pore sizes (3 kD cut-off).

Electrophoretic mobility shift assays and calculation of dissociation constant (K_d^{app})

EMSA experiments were performed using His-Six1 and fluorescently-labeled double-stranded DNA probes, which were prepared by end labeling of double-stranded DNA containing a G nucleotide overhang at each end of the molecule, with the Klenow enzyme Exo- (NEB) and Cy5-labeled dCTP (GE Healthcare). In all cases, the probe sequence context was that of the mouse myogenin MEF3 site, with the sequence gTTAGAGGGGGGCTCAGGTTTCTGTGGCGTTGGC as the top strand. The initial small script "g" represents the additional nucleotide overhang used for

labeling, while the underlined nucleotides constitute the MEF3 site. In order to test the influence of various MEF3 nucleotide substitutions on Six1 binding, and to disregard the putative influence of surrounding nucleotides, we changed the sequence of the MEF3 site while retaining the same surrounding sequence context. EMSA binding reactions contained varying amounts of His-Six1 protein in a fixed volume (4 µl in 50 mM Tris, 150 mM NaCl and 1 mM DTT) and 20 fmoles of probe in a final volume of 10 µl. The binding buffer was composed of Hepes 25 mM pH 7.6, KCl 8 mM, dIdC 1 µg, $MgCl_2$ 5 mM, Glycerol 10% v/v. The reactions were set up on ice, then incubated at 37ºC for 5 minutes, and loaded on a 5% w/v acrylamide gel (29:1 ratio acrylamide to bis-acrylamide) containing 2% glycerol, with TGE 0.5x (12.5 mM Tris, 95 mM Glycine, 0.5 mM EDTA) as the running buffer. After separation, the gels were rinsed in water and the fluorescent signal was quantitated using a Typhoon Phosphorimager (GE Healthcare), adjusting the photomultiplier tube voltage so that none of the signal is saturated. The ratio of the volume of shifted probes over that of the total (shifted and free probes) was calculated using ImageQuant software (GE Healthcare) for each concentration of Six1. To determine the K_d^{app} of protein-DNA binding, we determined the concentration of Six1 protein (in nanomolar) required to reach half maximal binding, using the function of one site saturation in SigmaPlot and following recommendations outlined in (31).

RESULTS

Identification of a novel MEF3-like motif

We have previously obtained ChIP-on-chip data for Six1 binding in the C2C12 cell line of mouse myoblasts (22). The experiment was performed in proliferating myoblasts and in differentiated myotubes and led to the identification of 1022 and 1853 high-confidence bound genomic loci in these two cell types, respectively. The average length of Six1-bound loci is 1230 bp; we used Amadeus, a *de novo* motif finding program, to precisely identify the DNA sequence motif most likely recognized by Six1 in these genomic regions. Our assumption is that the most abundant motif in these bound regions should be the one directly bound by Six1. For the purpose of motif discovery and subsequent testing, the search was run multiple times on the Six1 ChIP-on-chip target sequences and partitioned in eight subsets of approximately 330-370 sequences (see materials and methods for details). As shown in Fig.1, we obtained eight similar PWMs with little if any difference between the PWMs of Six1-bound genomic sequences detected in myoblasts and myotubes. Additionally, to summarize these results we also combined these eight PWMs to obtain an average matrix, called Six1_MB+MT (Fig. 1 and Table S1).

Fig. 1 clearly shows that our novel matrices share a strong resemblance with the previously identified MEF3 sequence motif represented in the TRANSFAC database. However, the new PWMs are clearly more degenerate, since at multiple positions more than one nucleotide is allowed; this is most obvious near the center of the motif, at positions 6-8, where the preference for the AGG nucleotides is weaker than in the TRANSFAC motif. In contrast, positions 4 (T) and 12 (C) display the lowest variety. These differences have important implications for the prediction of target gene binding by Six1: using the inflexible TRANSFAC MEF3 element would possibly overlook a large number of true targets.

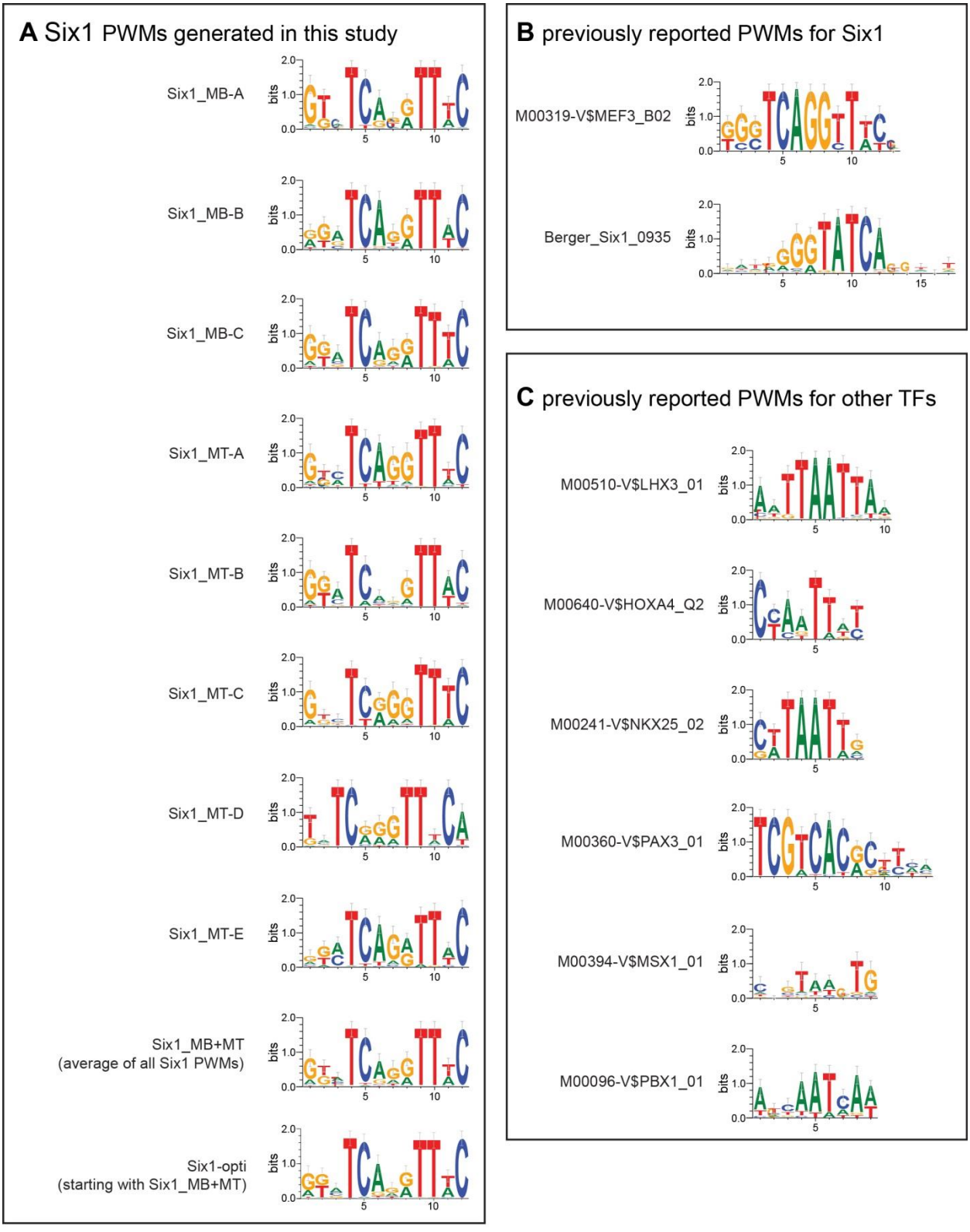

A Six1 PWMs generated in this study
Six1_MB-A
Six1_MB-B
Six1_MB-C
Six1_MT-A
Six1_MT-B
Six1_MT-C
Six1_MT-D
Six1_MT-E
Six1_MB+MT
(average of all Six1 PWMs)
Six1-opti
(starting with Six1_MB+MT)
B previously reported PWMs for Six1
M00319-V$MEF3_B02
Berger_Six1_0935
C previously reported PWMs for other TFs
M00510-V$LHX3_01
M00640-V$HOXA4_Q2
M00241-V$NKX25_02
M00360-V$PAX3_01
M00394-V$MSX1_01
M00096-V$PBX1_01
bits

Figure 2. 1 Position weight matrices for Six1 and other homeodomain TFs.

A) PWMs generated in this study. **B)** Previously reported PWMs for Six1. **C)** The PWMs of other well-characterized homeodomain TFs The PWMs were represented graphically using sequence logos.

In order to determine if these differences are significant and if the increased degeneracy in our novel PWMs has an impact of their predictive value, we evaluated the sensitivity and specificity of the PWMs. We compared the number of matches to each PWM that can be found in bound sequences and in control sequences using the Cisgenome motif mapping program: a larger number of matches indicate higher sensitivity, while the specificity is given by the enrichment ratio (frequency of matches in bound sequences divided by that in control sequences). We first verified that hits to the Six1_MB+MT PWM are substantially enriched among all strata of the Six1-bound loci found by ChIP-on-chip, not solely among the top (highest confidence) loci (Fig. 2). Table 1 gives the results of the comparison of our *de novo* identified motifs and reveals that all nine motifs we generated from our ChIP-on-chip data are present in large numbers among the bound genomic regions (*i.e.* the PWMs are sensitive) and are characterized by substantial enrichment levels (*i.e.* they are also specific). Importantly, the results presented in Table 1 clearly show that the novel matrices outperform the TRANSFAC motif in both specificity (higher enrichment levels) and sensitivity (larger number of sites). These analyses were performed giving equal consideration to all genomic sequences. When only the phylogenetically conserved regions of the bound loci were studied, the enrichment level of the PWMs increased, as can be expected for the binding sites of a developmentally important TF (32, 33). Here again, our *de novo* PWMs outperformed the TRANSFAC matrix with their higher enrichment levels (Table 1, enrichment – conserved sites only).

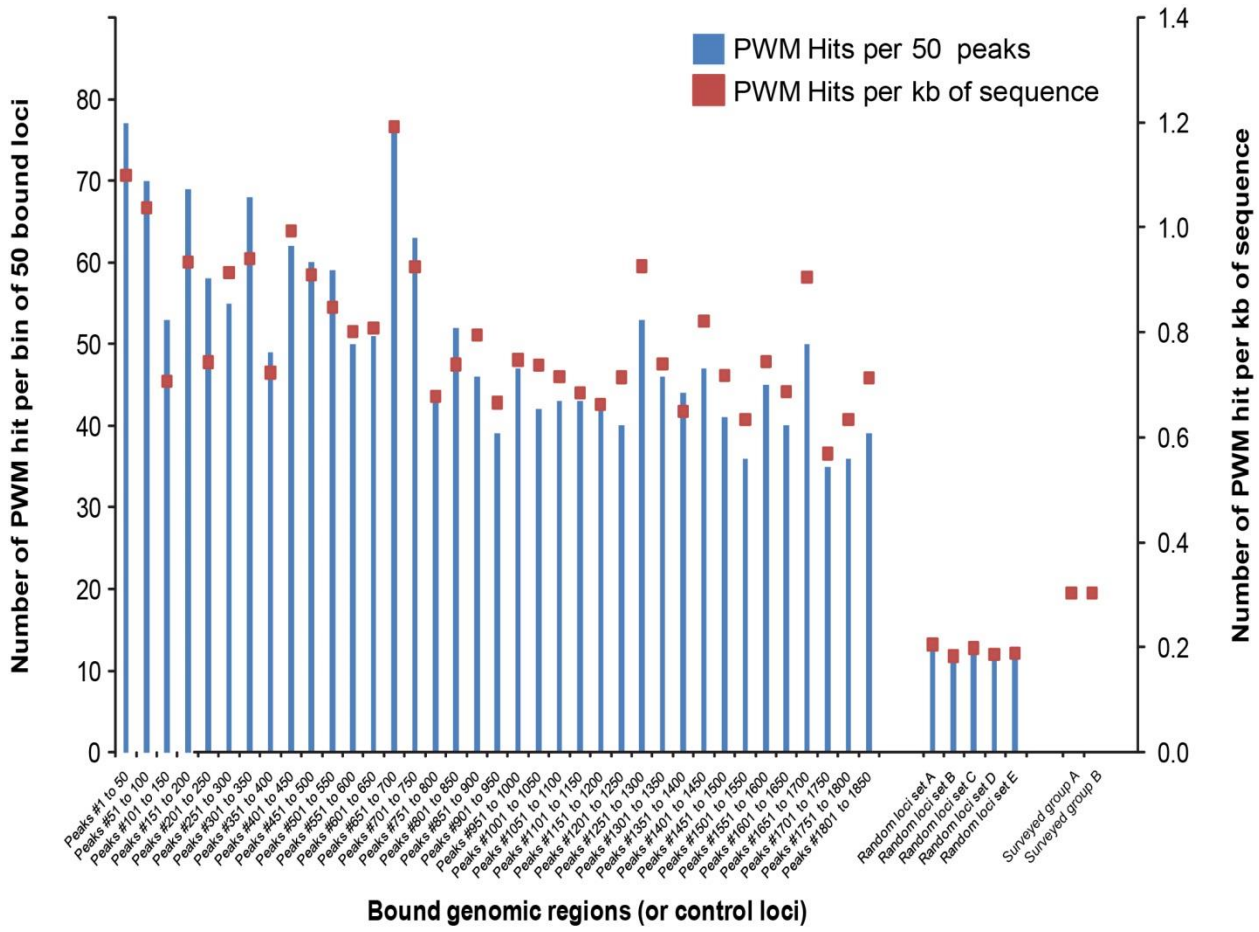

PWM Hits per 50 peaks
PWM Hits per kb of sequence
Number of PWM hit per bin of 50 bound loci
Number of PWM hit per kb of sequence
Bound genomic regions (or control loci)
Peaks #1 to 50
Peaks #51 to 100
Peaks #101 to 150
Peaks #151 to 200
Peaks #201 to 250
Peaks #251 to 300
Peaks #301 to 350
Peaks #351 to 400
Peaks #401 to 450
Peaks #451 to 500
Peaks #501 to 550
Peaks #551 to 600
Peaks #601 to 650
Peaks #651 to 700
Peaks #701 to 750
Peaks #751 to 800
Peaks #801 to 850
Peaks #851 to 900
Peaks #901 to 950
Peaks #951 to 1000
Peaks #1001 to 1050
Peaks #1051 to 1100
Peaks #1101 to 1150
Peaks #1151 to 1200
Peaks #1201 to 1250
Peaks #1251 to 1300
Peaks #1301 to 1350
Peaks #1351 to 1400
Peaks #1401 to 1450
Peaks #1451 to 1500
Peaks #1501 to 1550
Peaks #1551 to 1600
Peaks #1601 to 1650
Peaks #1651 to 1700
Peaks #1701 to 1750
Peaks #1751 to 1800
Peaks #1801 to 1850
Random loci set A
Random loci set B
Random loci set C
Random loci set D
Random loci set E
Surveyed group A
Surveyed group B

Figure 2. 2 Enrichment of the Six1_MB+MT matrix hits within the loci bound by Six1 in myotubes, as a function of their score in the ChIP-on-chip experiment.

The 1853 genomic loci bound by Six1 in myotubes were ranked in decreasing order of ChIP enrichment and subdivided in bins of 50 regions. Matches to the Six1_MB+MT matrix were identified, and the numbers of PWM hit per bin (left-hand y axis) or per base pair (right-hand y axis) were calculated. As controls, 5 sets of randomly selected genomic regions were also scanned for PWM hits. Additionally, two sets of sequences totaling 108 Mb and originally surveyed in the ChIP-on-chip experiments were also scanned here.

Name	Target list[a]	No. of sites (all)[b]	Enrichment (all sites)[c]	*P*-value (all sites)[d]	No. of sites (conserved only)[e]	Enrichment (conserved sites only)	*P*-value (conserved sites)
Six1_MB-A_m01.mat	MB_BC	485	4.73	<1E-16	139	6.49	<1E-16
Six1_MB-B_m01.mat	MB_AC	449	4.45	<1E-16	141	7.21	<1E-16
Six1_MB-C_m01.mat	MB_AB	471	5.23	<1E-16	144	7.31	<1E-16
Six1_MT-A_m01.mat	MT_BCDE	977	2.58	<1E-16	246	4.60	<1E-16
Six1_MT-B_m01.mat	MT_ACDE	1043	2.96	<1E-16	260	4.52	<1E-16
Six1_MT-C_m01.mat	MT_ABDE	888	2.73	2.2E-16	240	4.70	<1E-16
Six1_MT-D_m01.mat	MT_ABCE	891	3.08	<1E-16	248	4.92	1.3E-15
Six1_MT-E_m01.mat	MT_ABCD	895	2.36	<1E-16	242	3.93	<1E-16
Six1_MB+MT.mat	MB_ABC	1144	3.50	<1E-16	321	4.99	<1E-16
	MT_ABCDE	1873	2.93	3.3E-16	489	4.67	7.8E-16
Berger_Six1_0935.mat	MB_ABC	308	1.38	2.8E-08	51	1.52	2.6E-03
	MT_ABCDE	544	1.25	2.5E-07	84	1.54	9.5E-05
M00319-V$MEF3_B02.mat	MB_ABC	26	2.29	1.2E-04	9	5.02	8.6E-05
	MT_ABCDE	51	2.29	8.5E-08	11	3.77	1.7E-04
M00510-V$LHX3_01-Lhx3a.mat	MB_ABC	350	0.67	1.0E+00	116	0.67	1.0E+00
	MT_ABCDE	782	0.77	1.0E+00	203	0.72	1.0E+00
M00640-V$HOXA4_Q2-HOXA4.mat	MB_ABC	697	0.89	1.0E+00	185	0.99	5.9E-01
	MT_ABCDE	1429	0.93	1.0E+00	317	1.04	2.6E-01
M00241-V$NKX25_02-Nkx2-5.mat	MB_ABC	472	0.77	1.0E+00	116	0.69	1.0E+00
	MT_ABCDE	866	0.72	1.0E+00	188	0.68	1.0E+00
M00360-V$PAX3_01-Pax-3.mat	MB_ABC	27	1.13	2.9E-01	2	0.39	9.7E-01
	MT_ABCDE	63	1.35	1.3E-02	14	1.67	4.5E-02
M00394-V$MSX1_01-Msx-1.mat	MB_ABC	241	1.00	5.2E-01	52	0.80	9.5E-01
	MT_ABCDE	528	1.12	6.4E-03	105	1.00	5.2E-01
M00096-V$PBX1_01-Pbx1a.mat	MB_ABC	541	0.80	1.0E+00	147	0.94	7.7E-01
	MT_ABCDE	997	0.76	1.0E+00	211	0.83	1.0E+00

Table 2. 1 enrichment of binding sites predicted by PWMs discovered for Six1, for existing Six1 PWMs, and for other homeodomain transcription factors.

1: List of target genomic regions scanned with a given PWM. MB indicates Six1-bound targets in myoblasts, and MT those bound in myotubes. Subgroups of targets are given as letters (*e.g.* MB_AB refers to the combination of myoblast targets subgroups A and B).

2: Number of sites corresponding to "hits" to the PWM, irrespective of their phylogenetic conservation.

3: The enrichment is given as the ratio of hits found in the indicated target set over those found in a fraction of the ChIP-surveyed sequence space, pro-rated by the length of each group of sequences in base pairs.

4: The p value represents the cumulative hypergeometric probability subtracted from 1.

5: Same as for 2, but limited to genomic regions among the top 5% most phylogenetically conserved among 45 vertebrate species.

Secondly, we also compared our novel matrices to that identified by Berger et al. (4) by probing protein binding microarrays with the bacterially expressed Six1 homeodomain in isolation, excluding the N-terminal Six domain as well as the C-terminal region (11). Again, our *de novo* matrices outperform this motif, both when all sequences or only the conserved subset were considered (Table 1). We note that the similarity between our and Berger's matrices is limited to positions 4-6 of our motif (10-12 of their motif, consensus TCA).

Finally, we also verified whether the binding sites of other homeodomain transcription factors, including some that are involved in controlling myogenesis (Nkx2.5, Msx1, Pbx1, Pax3), are enriched among the genomic sites bound by Six1. None of these were enriched to a significant level within the Six1-bound genomic regions (Table 1). The canonical "ATTA" (reverse-complement of TAAT) DNA sequence motif recognized by homeodomain transcription factors (e.g. Nkx2.5, in Fig. 1) is observed in the Six1_MB+MT matrix at positions 8 to 11 (consensus (A/G)TT(T/A)). However, among the matches to the PWM that we have identified, only 222 out of 1873 conform to the canonical TAAT sequence at these positions; this sub-motif ranks fourth in frequency, behind GTTT, GTTA and ATTT (536, 374 and 268 hits respectively, Table S2). Together, the results of this analysis suggest that Six1-bound DNA elements is not limited to the canonical "ATTA" DNA sequence motif shared by several homeodomain TFs and provide a PWM that characterizes Six1 DNA binding preferences with improved accuracy over all other existing matrices.

Broad sequence specificity of DNA binding by Six1

The Six1 PWM that we established is substantially different from other homeodomain TFs as well as from the PWMs previously reported for Six1 (TRANSFAC, and Berger et al.). We were

especially intrigued by the rather degenerate nature of the central portion of the matrix. Consequently, we used electrophoretic mobility shift assays (EMSA) to probe Six1's ability to bind a range of DNA sequences that is wider than previously expected, avoiding the contribution of other confounding factors.

First, we set out to determine the apparent equilibrium dissociation constant (K_d^{app}) of purified Six1 for the MEF3 site present in the Myog proximal promoter (15, 22, 34, 35), a site that is identical to the consensus DNA motif established by the TRANSFAC PWM, and which is to date the best characterized Six1 binding site. EMSA reactions were performed in the presence of fixed amounts of Cy5-labeled probe and increasing amounts of the Six1 protein (Fig. 3A). We found that Six1 binds to the mouse Myog MEF3 site with a K_d^{app} of 35nM (Fig. 3B). Next, we aimed to verify whether the sequence preferences given by our PWM reflect the affinity of the protein to DNA. Accordingly, we designed a library of Cy5 labeled DNA duplexes corresponding to MEF3 site derivatives with a focus on the sequences diverging between Six1_MB+MT and the MEF3 PWM from TRANSFAC (Table 2). We found that, as suggested by the relative degeneracy of our novel MEF3-like matrix, many sequences differing from the TRANSFAC motif can be bound with high avidity by Six1 (Table 2, Myog_mut 07-11). These results further support the ideas that the TRANSFAC PWM is too stringent, and that *de novo* motif more accurately captures the DNA sequence preference of the Six1 transcription factor. Interestingly, we found that the C nucleotide "suffix" of the motif (TCAGGTTTC) is essential for high affinity binding of Six1 to DNA; mutation to any other nucleotide leads to a sharp decrease in binding (Table 2, Myog_mut 14 to 16, and Fig. 3C). This is an important observation considering that a shorter MEF3 element, amputated of this cytosine suffix, has often been described (36, 37, 38). Other variants are also indicative of Six1 binding preference. For example,

even though the Six1_MB+MT has a high level of degeneracy at positions 2-3 and 6, changing the prefix GGC to GAT (Mut03, at positions 2-3) or position 6 from an A to a T (Mut06) abolishes binding.

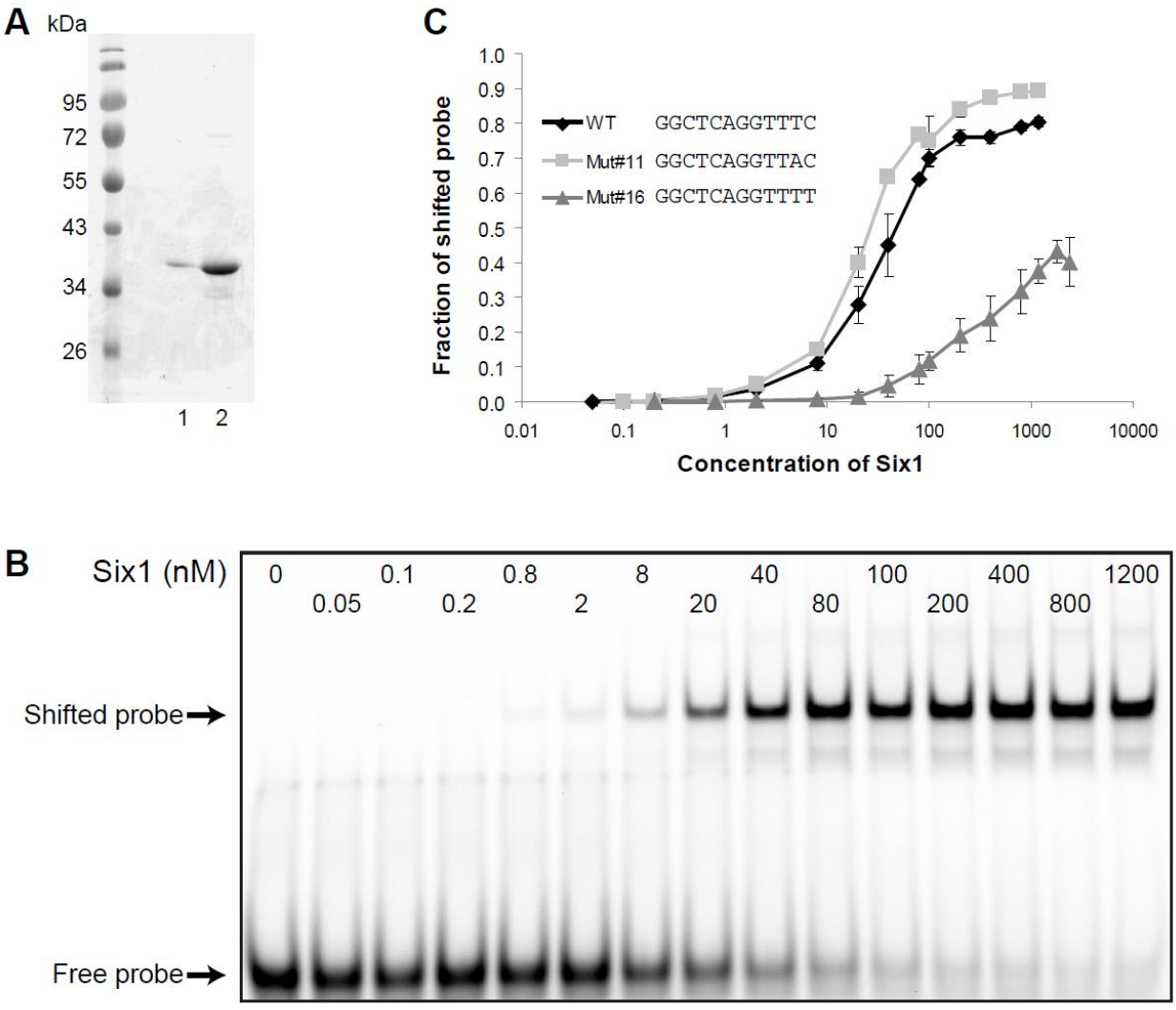

A
kDa
95
72
55
43
34
26
1 2
C
Fraction of shifted probe
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
WT GGCTCAGGTTTC
Mut#11 GGCTCAGGTTAC
Mut#16 GGCTCAGGTTTT
0.01 0.1 1 10 100 1000 10000
Concentration of Six1
B
Six1 (nM)
0 0.05 0.1 0.2 0.8 2 8 20 40 80 100 200 400 800 1200
Shifted probe
Free probe

Figure 2. 3 EMSA gels using recombinant Six1 on selected DNA sequences.

A) SDS-PAGE gel stained with coomassie blue, showing 200 ng (lane 1) or 2.0 μg (lane 2) of Six1 protein. **B)** Increasing amounts of Six1 were incubated with fixed amounts of fluorescently labeled double-stranded DNA probes, and electrophoresed on non-denaturing polyacrylamide gels. The gel shown is from an experiment performed with the wild-type Myog MEF3 site. **C)** The fluorescent signals corresponding to the free and shifted probes were measured in each lane, and the proportion of shifted probe over shifted free probe were plotted as a function of the concentration of Six1 present in each lane. The concentration of Six1 to reach half maximal binding represents the K_d^{app} value for that probe sequence. The complete set of results is reported in Table 2.

Name[a]	Sequence[b]	Rationale[c]	K_d^{app} (nM)[d]
WT			
Myog_WT	gTTAGAGGGG**GGCTCAGGTTTC**TGTGGCGTTGGC	Wild-type sequence	34.7 ± 7.9
Prefix changes			
Myog_mut01	gTTAGAGGGGGGATCAGGTTTCTGTGGCGTTGGC	Most frequent prefix	28.7 ± 4.3
Myog_mut02	gTTAGAGGGGGTATCAGGTTTCTGTGGCGTTGGC	Prefix conforms to Berger *et al.*	16.8 ± 2.4
Myog_mut03	gTTAGAGGGGGATTCAGGTTTCTGTGGCGTTGGC	Least frequent prefix	>350
Core changes			
Myog_mut04	gTTAGAGGGGGGCTCGGGTTTCTGTGGCGTTGGC	G is second most frequent after A	40.9 ± 4.1
Myog_mut05	gTTAGAGGGGGGCTCCGGTTTCTGTGGCGTTGGC	C is least frequent nucleotide	63.8 ± 7.7
Myog_mut06	gTTAGAGGGGGGCTCTGGTTTCTGTGGCGTTGGC	T is third most frequent after A	>350
Myog_mut07	gTTAGAGGGGGGCTCAGATTTCTGTGGCGTTGGC	Very frequent dinucleotide	52.3 ± 5.6
Myog_mut08	gTTAGAGGGGGGCTCAAGTTTCTGTGGCGTTGGC	Very frequent dinucleotide	34.2 ± 1.8
Myog_mut09	gTTAGAGGGGGGCTCATGTTTCTGTGGCGTTGGC	Very frequent dinucleotide	29.4 ± 2
Myog_mut10	gTTAGAGGGGGGCTCAAATTTCTGTGGCGTTGGC	Very frequent dinucleotide	32.4 ± 2.1
Myog_mut11	gTTAGAGGGGGGCTCAGGTTACTGTGGCGTTGGC	A is second most frequent after T	24.2 ± 1.3
Myog_mut12	gTTAGAGGGGGGCTCAGGTTCCTGTGGCGTTGGC	Rare nucleotide	81.6 ± 10
Myog_mut13	gTTAGAGGGGGGCTCAGGTTGCTGTGGCGTTGGC	Rare nucleotide	72.2 ± 3.1
Suffix changes			
Myog_mut14	gTTAGAGGGGGGCTCAGGTTTATGTGGCGTTGGC	Very rare nucleotide	>350
Myog_mut15	gTTAGAGGGGGGCTCAGGTTTGTGTGGCGTTGGC	Very rare nucleotide	>350
Myog_mut16	gTTAGAGGGGGGCTCAGGTTTTTGTGGCGTTGGC	Very rare nucleotide	>350
Multiple changes			
Myog_mut32	gTTAGAGGGGGTATCAGGGTTCTGTGGCGTTGGC	mut02 with change near 3' end	>350
Myog_mut33	gTTAGAGGGGGTATCAGGTGTCTGTGGCGTTGGC	mut02 with change near 3' end	130 ± 50
Myog_mut34	gTTAGAGGGGGTATCAGGTTTGTGTGGCGTTGGC	mut02 with change at 3' end	>350
Tests of Six1-opti PWM prediction of binding sites (changes to the core, suffix and/or prefix)			
Myog_mut17	gTTAGAGGGGATCTCATATTACTGTGGCGTTGGC	Unique to Six1-opti	25 ± 3.3
Myog_mut18	gTTAGAGGGGAGATCACATTTCTGTGGCGTTGGC	Unique to Six1-opti	39.4 ± 2.1
Myog_mut19	gTTAGAGGGGAGATCACATTACTGTGGCGTTGGC	Unique to Six1-opti	48.2 ± 1.1
Myog_mut20	gTTAGAGGGGTTCTCAAATTACTGTGGCGTTGGC	Unique to Six1-opti	46.7 ± 1.1
Myog_mut21	gTTAGAGGGGGTATAAAATTTCTGTGGCGTTGGC	Unique to Six1-opti	74.3 ± 9
Myog_mut22	gTTAGAGGGGAGCTCTGGTTACTGTGGCGTTGGC	Unique to Six1-opti	85.4 ± 10
Myog_mut23	gTTAGAGGGGAGATCAGGTTTATGTGGCGTTGGC	Unique to Six1-opti	69.3 ± 8.2
Myog_mut24	gTTAGAGGGGGGGTCAGGTGACTGTGGCGTTGGC	Unique to Six1-opti	>350
Myog_mut25	gTTAGAGGGGATATCAGATATCTGTGGCGTTGGC	Unique to Six1-opti	29.1 ± 5.3
Myog_mut26	gTTAGAGGGGGTATCAAATAACTGTGGCGTTGGC	Unique to Six1-opti	10.8 ± 3
Myog_mut27	gTTAGAGGGGGCCTCGGGTTTCTGTGGCGTTGGC	Unique to Six1_MB+MT	>350
Myog_mut28	gTTAGAGGGGGGATCGGGTTCCTGTGGCGTTGGC	Unique to Six1_MB+MT	42.4 ± 7.6
Myog_mut29	gTTAGAGGGGGTTTCAGGTTCCTGTGGCGTTGGC	Unique to Six1_MB+MT	75.4 ± 8.3
Myog_mut30	gTTAGAGGGGGTCTCGGCTTTCTGTGGCGTTGGC	Unique to Six1_MB+MT	>350
Myog_mut31	gTTAGAGGGGGATTCAGGTTTCTGTGGCGTTGGC	Unique to Six1_MB+MT	>350

Table 2. 2 Summary of EMSA experiments

a, Myog_WT is the Myog probe with wild type MEF3 consensus in the center. Myog_mut01 to Myog_mut30 are probes with various mutations in the MEF3 consensus. Myog_mut31 is the same probe as Myog_mut03 cited as a different rationale.

b, Mutated nucleotides in the MEF3 consensus are highlighted in black. The lower cap "g" nucleotide was added for fluorescent labeling purposes. The natural sequence would be a "C" at that position.

c, Rationales to choose the corresponding sequences are listed. Mut01 the most frequent MEF3 sequence found in Six1_MB and Six1_MT binding data. Mut02 contains TA at position 2 and 3, which is found in the Berger et al. study. Myog03 has the least frequency of dinucleotides (AT) at position 2 and 3. The MEF3 in Mut05 is found in the Myod core enhancer region. Mut07 to10 are selected with different dinucleotide combination at position 7 and 8. Mut04, 06, and 11 to 16 are chosen based on the frequency of the nucleotide at a certain position. Mut17 to 26 are MEF3 sequences found only using Six1-opti MEF3 motif. Mut27 to 31 are MEF3 sequences found only using Six1_MB+MT MEF3 motif. Of note, Mut03 and Mut31 contain the same MEF3 sequence.

d, Dissociation constant ($\mathbf{K_d^{app}}$) and standard error of mean are calculated for each probe based on at least 3 independent experiments. >350nM, not accurately determined due to very weak binding.

Regions outside the homeodomain contribute to DNA binding sequence specificity

In their large-scale study of mouse homeodomains, Berger et al. reported a DNA sequence motif preferred by the Six1 homeodomain (Six1-HD, Fig. 1B) that is rather different from the one we report here for the full-length protein (Fig. 1A). This has important implications for the possible mode of DNA binding by Six1, and suggests that protein regions outside of its homeodomain may participate in binding site selection. We therefore addressed this question using EMSA, by comparing the affinities of Six1-HD and Six1 for certain DNA sequences.

First, we tested binding of the two proteins on the Myog wild-type and mut02 probes, since the latter conforms to the Berger et al. preferred sequence. We observed that while binding of Six1-HD on the mut02 probe occurs with a K_d^{app} of 690 nM, only very weak binding occurred between the homeodomain and the wild-type Myog MEF3 site (K_d^{app} 7700 nM, Fig. 4A). This is consistent with the fact that the Berger et al. PWM gives a substantial importance to the GTA prefix, which is present in mut02 but absent in the wild-type probe. In contrast, the full-length Six1 protein binds both sequences with comparable affinities (Table 2, 34.7 and 28.7 nM for WT and mut02, respectively), in accordance with the fact that our Six1_MB+MT PWM attributes considerably less importance to the prefix sequence (Fig. 1).

Secondly, we designed new Myog mutant probes (mut32, mut33 and mut34), which conform to the Berger et al. PWM at the prefix (GTA instead of GGC), but they deviate from the wild-type sequence near the 3' end (Table 2). Using these new probes, we again compared Six1-HD and Six1 by EMSA. We found that Six1-HD is mostly unaffected by these mutations, including replacement of the suffix C by a G (mut34, Fig. 4B, top row). This is consistent with the reported PWM, which does not dictate any sequence preference at these positions. On the other hand, two

of these mutations greatly affect full-length Six1 binding (Fig. 4B, bottom row). Based on these substantial differences in DNA sequence preferences between Six1-HD and Six1, we conclude that binding site selection by Six1 involves not only its homeodomain but also regions outside of it.

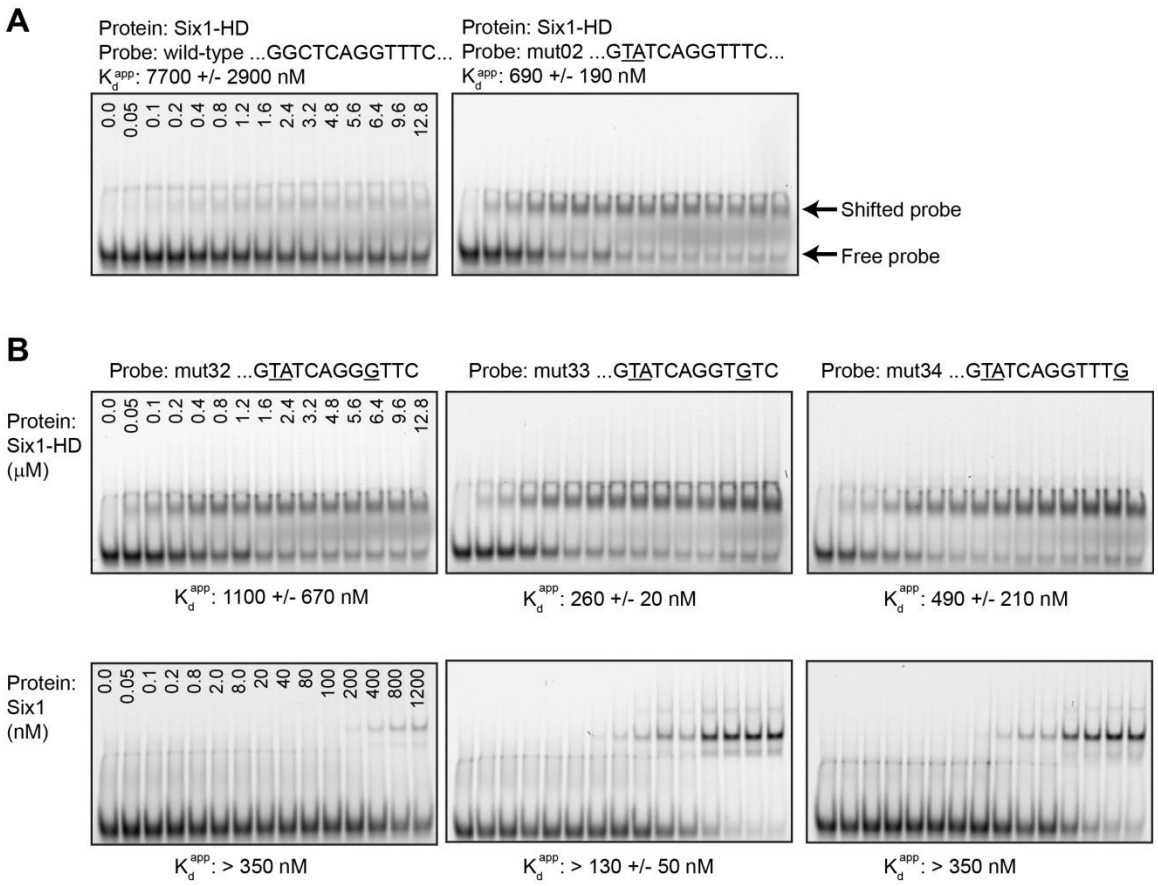

A
Protein: Six1-HD
Probe: wild-type ...GGCTCAGGTTTC...
K_d^{app}: 7700 +/- 2900 nM
Protein: Six1-HD
Probe: mut02 ...GTATCAGGTTTC...
K_d^{app}: 690 +/- 190 nM
0.0 0.05 0.1 0.2 0.4 0.8 1.2 1.6 2.4 3.2 4.8 5.6 6.4 9.6 12.8
Shifted probe
Free probe
B
Probe: mut32 ...GTATCAGGGTTC
Probe: mut33 ...GTATCAGGTGTC
Probe: mut34 ...GTATCAGGTTTG
Protein: Six1-HD (μM)
0.0 0.05 0.1 0.2 0.4 0.8 1.2 1.6 2.4 3.2 4.8 5.6 6.4 9.6 12.8
K_d^{app}: 1100 +/- 670 nM
K_d^{app}: 260 +/- 20 nM
K_d^{app}: 490 +/- 210 nM
Protein: Six1 (nM)
0.0 0.05 0.1 0.2 0.8 2.0 8.0 20 40 80 100 200 400 800 1200
K_d^{app}: > 350 nM
K_d^{app}: > 130 +/- 50 nM
K_d^{app}: > 350 nM

Figure 2. 4 Substantial differences in DNA sequence selectivity between Six1-HD and Six1.

A) EMSA gels were performed with increasing amounts of Six1 homeodomain (Six1-HD) using the Myogenin WT probe (left), or the mut02 probe (right), which conforms to the consensus reported by Berger et al. using protein binding microarrays. **B)** EMSA experiments performed with the Six1-HD (top row) or Six1 (bottom row) proteins, on three derivatives of the mut02 probe (mutated positions are underlined, compared to the WT probe). Note that because the Six1-HD has a relatively low affinity for DNA in these assays, the amounts of protein used are higher than those used for the full-length Six1 protein. The K_d^{app} values (all in nanomolars) for each protein on each probe are given underneath the respective gel images.

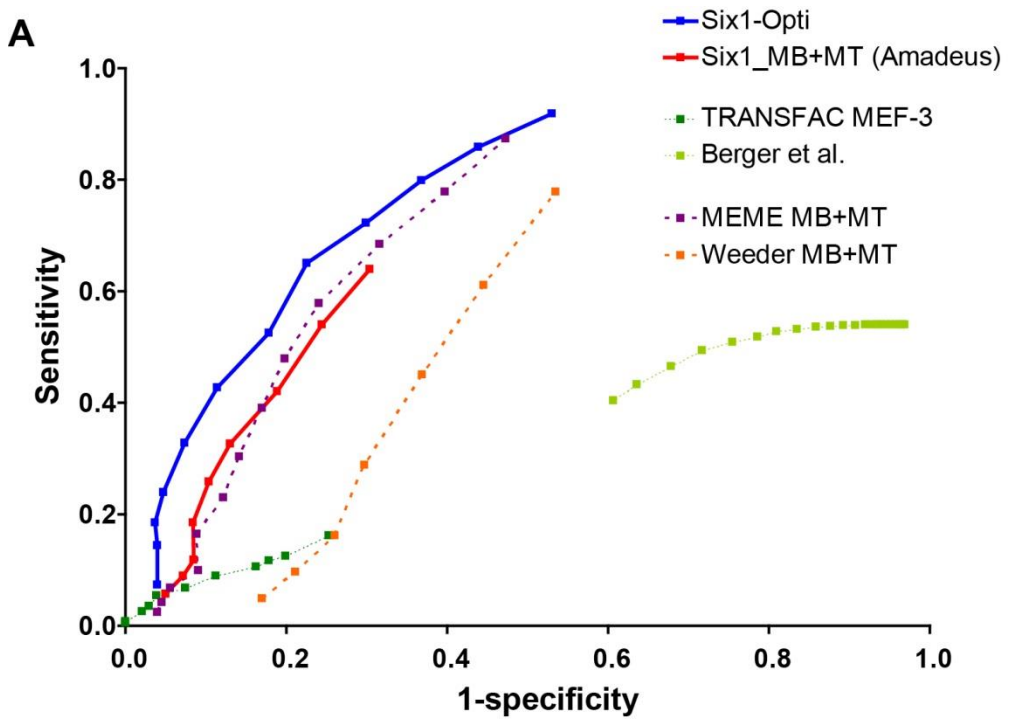
Computational optimization of the Six1 PWM

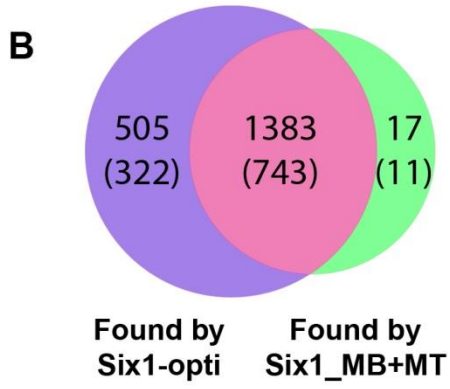
As outlined above the novel Six1_MB+MT PWM discovered *de novo* outperforms other existing matrices. However, we postulated that it could still be improved using bioinformatic approaches, considering a potential "dilution effect" due to the length of peaks from the ChIP-on-Chip technique (a stretch of 12 bp motif in peaks of 1230 bp on average). We employed for this purpose a novel BSD approach (Bound/Surveyed Discrimination) developed in this study (see Methods for details). Using our ChIP-on-chip data, we sought to determine if changes to the nucleotide weight values within the matrix could be introduced and further increases its specificity and sensitivity to discriminate between the "bound" and "surveyed" sets of sequences. In doing this, we made the assumption that matches to the PWM in "bound" sequences represent biologically true binding sites, while matches to the "surveyed" regions represent mostly "not-bound" sites. We note however that the "bound" set of sequences is a subset of the "surveyed" sequences and so surveyed sequences do contain truly bound sites. This led to the generation of the Six1-opti matrix (Fig. 1A, bottom). A receiver-operator characteristics curve analysis of the two PWMs reveals that indeed, the optimized PWM performs better than the original Six1_MB+MT matrix (or the TRANSFAC and Berger PWMs, Fig. 5A), since at any given level of specificity, the optimized PWM has enhanced sensitivity. The benefits of using our optimization procedure are unlikely to originate from having started with a poor initial PWM generated by Amadeus, since PWMs obtained by MEME-ChIP (39) and Weeder (40), two popular motif finding programs, did not perform any better than the one obtained with Amadeus (Six1_MB+MT).

The improved Six1-opti PWM allowed us to identify new potential binding sites that may have been missed using the original Six1_MB+MT matrix. At similar sensitivity and specificity (~60%

and ~73%, respectively), the Six1-opti matrix identified 322 novel putative binding sequences (occurring a total of 505 times among our Six1-bound genomic loci) that were missed with the starting matrix (Fig. 5B). On the other hand, the starting matrix identified only 11 sequences (17 occurrences) not found by the optimized PWM. It is also noteworthy to consider these results in terms of putative target gene identification, since this is a common use of PWM scanning programs. Using the Six1-opti PWM would allow to identify 1051 target genes (i.e. sequences with at least one hit to the PWM), while the original matrix would only recognize 747 of them. This represents a 40.7 % increase in sensitivity.

In order to determine if Six1 binds the new sequences identified by the optimized matrix, we used EMSA to assess the binding affinity of Six1 to them. We selected 10 novel sites unique to Six1-opti and 5 sites uniquely identified with Six1_MB+MT for validation (Table 2). Interestingly, 9 out of 10 sites unique to Six1-opti have comparable affinity to the Myog probe, whereas only 2 out of 5 sites unique to Six1_MB+MT are bound by Six1 with a measurable K_d^{app}, suggesting that our optimization approach improved the discriminatory power of our Six1 PWM. Finally, as an ultimate test of the relevance of the binding site predictions made by the optimized PWM, we repeated the search on an independent set of Six1 target loci which are bound by Six1 24 hours after the onset of myoblast differentiation, but not in myoblasts or myotubes (see Materials and Methods). As expected, with 49 common binding sequences out of 187 Six1-bound regions, 23 binding sites were identified solely with the Six1-opti PWM, while only 2 binding sequences were uniquely defined with Six1_MB+MT (Fig. 5C). This confirms the superiority of the optimized PWM over the original matrix.


A
Six1-Opti
Six1_MB+MT (Amadeus)
TRANSFAC MEF-3
Berger et al.
MEME MB+MT
Weeder MB+MT
Sensitivity
1-specificity
0.0
0.2
0.4
0.6
0.8
1.0


B
505
(322)
1383
(743)
17
(11)
Found by
Six1-opti
Found by
Six1_MB+MT

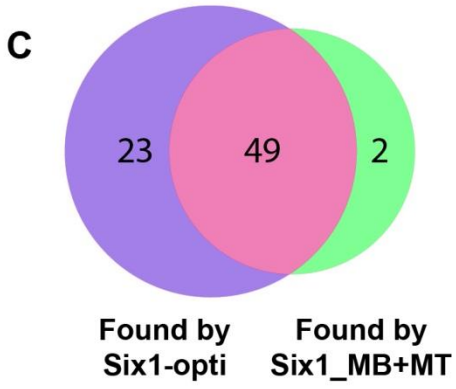

C
23
49
2
Found by
Six1-opti
Found by
Six1_MB+MT

Figure 2. 5 Performance comparison of the Six1_MB+MT and Six1-opti PWMs.

A) ROC curves. The y-axis is the sensitivity and the x-axis is the 1-specificity value. Performances of the original TRANSFAC MEF3 and Berger et al. PWMs for Six1 are also given for comparison. **B)** Venn diagram indicating the number of hits to each PWM, and their overlap, at their respective optimal thresholds, among loci targeted by Six1 in myotubes. Numbers in parentheses are the number of unique sequences (one sequence can occur more than once). **C)** Results for the comparison between predictions made with the Six1-opti and Six1_MB+MT PWMs, on sequences bound by Six1 only at 24 hours post-differentiation.

DISCUSSION

Using bioinformatic analysis of the genomic binding profile of Six1 in muscle cells, we have found that this transcription factor has the ability to bind to a much broader range of DNA sequences than anticipated. While the previously reported MEF3 sequence motif is clearly enriched among genomic regions bound by Six1, other sequences that diverge substantially from this DNA element are also found preferentially at Six1-bound loci. We used an *in vitro* binding assay with recombinant Six1 to confirm that the protein indeed binds to these divergent sequences with high affinity, thereby ruling out an artifactual effect of the *de novo* sequence discovery algorithm we used. The novel PWM should prove to be useful to studies that employ TF target site prediction to elucidate the structure of regulatory networks (41). For Six1, a TF involved in the genesis of multiple tissue types, a more precise DNA binding motif may contribute to discovering novel direct targets, elucidating composite regulatory networks and rationalizing its implication in diseases such as breast cancer or Branchio-oto-renal syndrome (7, 42, 43).

The results of our analyses provide useful information that may guide structure-function studies of Six1-DNA interactions. The Six family homeodomains, including that of Six1, belong to the K50 class of homeodomains, as they differ from the majority of other homeodomains classes by the change of a key DNA-binding residue, asparagine at position 50 of the HD, to a lysine. The importance of that residue in encoding DNA binding specificity has been highlighted by a number of biochemical and structural studies (44, 45, 46, 47, 48). Other residues implicated in DNA binding specificity or stability, for example those within the N-terminal arm, differ between Six and other homeodomains or within Six family members (discussed in (49, 50, 51)). Yet, it remains to be determined precisely how DNA binding specificity is established by Six

family TFs, and by Six1 in particular. The various family members have been shown to bind to different sequences: Six3 and Six6 can bind to the canonical TAAT sequence that is bound by most homeodomain TFs, while Six1/2/4/5 have all been shown to be able to bind to sequences resembling the MEF3 element (reviewed in (11)). Some Six proteins can also bind DNA as heterodimers with Eya family proteins, and these interactions are thought to enhance their affinity for DNA (7, 51). Most TFs, including several homeodomain factors, bind to DNA *in vivo* as homo- or heterodimers, and stabilization of such oligomeric states has been put forth as one possible mechanism to explain the influence of Eya proteins in regulating Six proteins (51). In our experiments, one predominant oligomeric state of Six1 was detected in EMSA, however, the oligomeric state of Six1 in solution remains unclear (*i.e.* whether Six1 binds DNA as a monomer or as a multimer in our assays), although at very high protein concentrations slower migrating complexes became visible on EMSA gels (data not shown). We cannot tell at this point if these species represent aggregates or functionally and physiologically relevant oligomers. Further analysis of the precise mode of DNA recognition by Six1, the influence of interaction partners, and possible involvement of oligomeric states, should help us understand the function of this protein, and rationalize its implication in diseased states.

ChIP-on-chip analysis captures a snapshot of protein-DNA interactions as they occur in live cells, and although chromatin is immunoprecipitated with an antibody against Six1, putative DNA-binding partners of Six1 were possibly involved in the interactions we have discovered. It would be interesting to determine the genomic binding profiles of Eya proteins in myoblasts to see if indeed these proteins tend to bind DNA along with Six1, and if co-binding with Eya proteins alters DNA binding preferences in any way. However these experiments may prove excessively difficult to perform *in vivo*: while it is possible to ChIP a specific Eya-Six chromatin-bound

complex (using sequential ChIP assays), it would be much more challenging to devise a way of pulling-down only Six1 chromatin complexes that do not contain Eya proteins. Interestingly, however, the results of our *in vitro* experiments corroborate those of our ChIP-on-chip experiments: the variety of sequence motifs enriched among Six1-bound loci is reflected in EMSA using DNA probes and Six1 alone. This leads us to postulate that protein domain(s) within Six1 itself are the main determinant of DNA sequence selection by Six1 *in vivo* (at least in muscle cells, where our analysis was done). As shown by others and noted above, Eya binding could influence predominantly the binding affinity rather than sequence selectivity (7, 51).

The PWM we have generated is fairly close to the initial MEF3 PWM reported more than 15 years ago. The stringency (and therefore low sensitivity) of the TRANSFAC MEF3 PWM comes from the fact that it was derived from the DNA sequences of only five sites within muscle gene promoters. One can easily imagine that with larger sampling, the PWM would perform better at predicting Six1 binding sites.

The significant discrepancy between our data and those reported by Berger *et al.* was more puzzling. The authors used protein binding microarrays to determine the sequence-specificity of the mouse Six1 homeodomain and reported a Six1 PWM that has only limited resemblance to our PWM or to the TRANSFAC MEF3 element (4). We reason that these differences originate from the fact that only the homeodomain region of Six1 was studied, whereas our ChIP-on-chip and binding studies were performed with full-length Six1. Indeed, we confirmed with EMSA experiments that Six1-HD exhibits a DNA sequence preference that is in line with what Berger et al. reported, but that is substantially different from that exhibited by the full-length protein. We therefore conclude that regions outside of the Six1 homeodomain participate in DNA

binding, either through direct DNA contacts or indirectly, perhaps by enabling structural stabilization.

SUPPLEMENTARY DATA

Table 2. S 1 Different matrixes of Six1 binding DNA motif

>Six1_MB+MT_nucleotide_frequency_matrix

0.1665	0.0382	0.7501	0.0491
0.0429	0.0253	0.4525	0.4833
0.4569	0.3149	0.1564	0.0757
0.0048	0.0041	0.0012	0.9939
0.0218	0.9326	0.0012	0.0483
0.7068	0.0584	0.1908	0.048
0.1728	0.0549	0.6499	0.1264
0.291	0.0208	0.6789	0.0133
0.0089	0.0012	0.0039	0.99
0.0012	0.0012	0.0113	0.9902
0.3918	0.0465	0.0159	0.5499
0.0088	0.9797	0.0012	0.0142

>Six1_MB+MT_PWM

-1.43463	-2.95711	0	-2.67064
-2.40685	-2.2016	-0.12211	0
0	-0.42347	-1.14378	-1.81264
-3.3336	-3.4056	-3.51473	0
-2.83604	0	-3.41439	-2.92335
0	-2.54618	-1.38296	-2.70464
-1.25064	-2.4505	0	-1.57844
-0.77251	-2.56931	0	-2.69012
-3.21844	-3.49246	-3.43077	0
-3.43843	-3.49308	-3.23466	0
-0.32371	-2.51009	-2.54173	0
-3.17029	0	-3.46444	-3.05564

>Six1-opti_nucleotide_frequency_matrix

0.25815	0.01359	0.6712	0.05707
0.02446	0.01087	0.48098	0.4837
0.44293	0.30978	0.2038	0.04348
0	0.00272	0.00272	0.99457
0.01087	0.97283	0.00272	0.01359
0.80163	0.01902	0.12772	0.05163

0.26087	0.06793	0.5	0.1712
0.39946	0.01902	0.56793	0.01359
0.00815	0.00272	0	0.98913
0.00815	0	0.00815	0.9837
0.42935	0.0462	0.00543	0.51902
0	0.9837	0	0.0163

>Six1-opti_PWM

-0.88493	-2.71404	0	-2.40916
-2.16914	-2.50696	-0.0619	0
0	-0.40881	-0.84799	-2.33611
-3.4861	-3.45115	-3.47172	0
-3.11756	0	-3.41294	-3.06825
0	-2.84713	-1.91027	-2.75763
-0.57654	-1.97525	0	-1.01287
-0.27726	-2.42881	0	-2.50779
-3.24003	-3.4479	-3.55562	0
-3.23448	-3.52952	-3.31022	0
-0.17441	-2.45887	-2.7453	0
-3.43464	0	-3.51153	-3.01589

Table 2. S 2 1873 Six1 binding peaks in myotubes is uploaded on line.

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067762

.

FUNDING

This work was supported by Discovery grants from the Canadian Natural Science and Engineering Research Council (NSERC) to A.B., S.N. and I.I., by Leaders Opportunity Fund grant of Canadian Foundation of Innovation and Ontario Research Foundation to I.I. and S.N., and by a NSERC graduate scholarship to Y.L.

ACKNOWLEDGEMENTS

We thank Jean-François Couture and Sabina Sarvan for advice and help with protein purification and K_d^{app} determination.

REFERENCES

1. Holland,P.W., Booth,H.A. and Bruford,E.A. (2007) Classification and nomenclature of all human homeobox genes. *BMC Biol.,* **5,** 47.

2. Gehring,W.J., Qian,Y.Q., Billeter,M., Furukubo-Tokunaga,K., Schier,A.F., Resendez-Perez,D., Affolter,M., Otting,G. and Wuthrich,K. (1994) Homeodomain-DNA recognition. *Cell,* **78,** 211-223.

3. Svingen,T. and Tonissen,K.F. (2006) Hox transcription factors and their elusive mammalian gene targets. *Heredity,* **97,** 88-96.

4. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Pena-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T., et al. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell,* **133,** 1266-76.

5. Noyes,M.B., Christensen,R.G., Wakabayashi,A., Stormo,G.D., Brodsky,M.H. and Wolfe,S.A. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell,* **133,** 1277-1289.

6. Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science,* **270,** 262-269.

7. Patrick,A.N., Schiemann,B.J., Yang,K., Zhao,R. and Ford,H.L. (2009) Biochemical and functional characterization of six SIX1 branchio-oto-renal syndrome mutations. *J. Biol. Chem.,* **284,** 20781-20790.

8. Chang,C.P., Brocchieri,L., Shen,W.F., Largman,C. and Cleary,M.L. (1996) Pbx modulation of hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the hox locus. *Mol. Cell. Biol.,* **16,** 1734-1745.

9. Shen,W.F., Montgomery,J.C., Rozenfeld,S., Moskow,J.J., Lawrence,H.J., Buchberg,A.M. and Largman,C. (1997) AbdB-like hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol. Cell. Biol.,* **17,** 6448-6458.

10. Kawakami,K., Sato,S., Ozaki,H. and Ikeda,K. (2000) Six family genes--structure and function as transcription factors and their roles in development. *Bioessays,* **22,** 616-26.

11. Kumar,J.P. (2009) The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. *Cell Mol Life Sci,* **66,** 565-83.

12. Blais,A., Tsikitis,M., Acosta-Alvear,D., Sharan,R., Kluger,Y. and Dynlacht,B.D. (2005) An initial blueprint for myogenic differentiation. *Genes Dev,* **19,** 553-69.

13. Parmacek,M.S., Ip,H.S., Jung,F., Shen,T., Martin,J.F., Vora,A.J., Olson,E.N. and Leiden,J.M. (1994) A novel myogenic regulatory circuit controls slow/cardiac troponin C gene transcription in skeletal muscle. *Mol Cell Biol,* **14,** 1870-85.

14. Spitz,F., Salminen,M., Demignon,J., Kahn,A., Daegelen,D. and Maire,P. (1997) A combination of MEF3 and NFI proteins activates transcription in a subset of fast-twitch muscles. *Mol Cell Biol,* **17,** 656-66.

15. 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. *Proc Natl Acad Sci U S A,* **95,** 14220-5.

16. Laclef,C., Hamard,G., Demignon,J., Souil,E., Houbron,C. and Maire,P. (2003) Altered myogenesis in Six1-deficient mice. *Development,* **130,** 2239-52.

17. Grifone,R., Demignon,J., Houbron,C., Souil,E., Niro,C., Seller,M.J., Hamard,G. and Maire,P. (2005) Six1 and Six4 homeoproteins are required for Pax3 and mrf expression during myogenesis in the mouse embryo. *Development,* **132,** 2235-49.

18. Li,X., Oghi,K.A., Zhang,J., Krones,A., Bush,K.T., Glass,C.K., Nigam,S.K., Aggarwal,A.K., Maas,R., Rose,D.W., et al. (2003) Eya protein phosphatase activity regulates Six1-dach-eya transcriptional effects in mammalian organogenesis. *Nature,* **426,** 247-54.

19. Bessarab,D.A., Chong,S.W., Srinivas,B.P. and Korzh,V. (2008) Six1a is required for the onset of fast muscle differentiation in zebrafish. *Dev Biol,* **323,** 216-28.

20. Molkentin,J.D. and Olson,E.N. (1996) Combinatorial control of muscle development by basic helix-loop-helix and MADS-box transcription factors. *Proc Natl Acad Sci U S A,* **93,** 9366-73.

21. 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. *Mol Cell,* **14,** 465-77.

22. Liu,Y., Chu,A., Chakroun,I., Islam,U. and Blais,A. (2010) Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. *Nucleic Acids Res.,* **38,** 6857-6871.

23. Linhart,C., Halperin,Y. and Shamir,R. (2008) Transcription factor and microRNA motif discovery: The amadeus platform and a compendium of metazoan target sets. *Genome Res,* **18,** 1180-9.

24. Tarailo-Graovac,M. and Chen,N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics,* **Chapter 4,** Unit 4.10.

25. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: A sequence logo generator. *Genome Res.,* **14,** 1188-1190.

26. Ji,H., Jiang,H., Ma,W., Johnson,D.S., Myers,R.M. and Wong,W.H. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol,* **26,** 1293-300.

27. Staden,R. (1989) Methods for calculating the probabilities of finding patterns in sequences. *Comput. Appl. Biosci.*, **5,** 89-96.

28. Bucher,P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.,* **212,** 563-578.

29. Gershenzon,N.I., Stormo,G.D. and Ioshikhes,I.P. (2005) Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites. *Nucleic Acids Res.,* **33,** 2290-2301.

30. Burset,M. and Guigo,R. (1996) Evaluation of gene structure prediction programs. *Genomics,* **34,** 353-367.

31. Carey,M. and Smale,S.T. (2000) Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

32. Prabhakar,S., Poulin,F., Shoukry,M., Afzal,V., Rubin,E.M., Couronne,O. and Pennacchio,L.A. (2006) Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res,* **16,** 855-63.

33. Woolfe,A., Goodson,M., Goode,D.K., Snell,P., McEwen,G.K., Vavouri,T., Smith,S.F., North,P., Callaway,H., Kelly,K., et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.*, **3,** e7.

34. Yee,S.P. and Rigby,P.W. (1993) The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev,* **7,** 1277-89.

35. Seenundun,S., Rampalli,S., Liu,Q.C., Aziz,A., Palii,C., Hong,S., Blais,A., Brand,M., Ge,K. and Dilworth,F.J. (2010) UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. *Embo J,* **29,** 1401-11.

36. Bai,L. and Merchant,J.L. (2000) Transcription factor ZBP-89 cooperates with histone acetyltransferase p300 during butyrate activation of p21waf1 transcription in human cells. *J Biol Chem,* **275,** 30725-33.

37. Dong,Y., Walsh,M.D., McGuckin,M.A., Gabrielli,B.G., Cummings,M.C., Wright,R.G., Hurst,T., Khoo,S.K. and Parsons,P.G. (1997) Increased expression of cyclin-dependent kinase inhibitor 2 (CDKN2A) gene product P16INK4A in ovarian cancer is associated with progression and unfavourable prognosis. *Int J Cancer,* **74,** 57-63.

38. Donjerkovic,D., Zhang,L. and Scott,D.W. (1999) Regulation of p27Kip1 accumulation in murine B-lymphoma cells: Role of c-myc and calcium. *Cell Growth Differ,* **10,** 695-704.

39. Machanick,P. and Bailey,T.L. (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics,* **27,** 1696-1697.

40. Pavesi,G., Mereghetti,P., Mauri,G. and Pesole,G. (2004) Weeder web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.,* **32,** W199-203.

41. Vavouri,T. and Elgar,G. (2005) Prediction of cis-regulatory elements using binding site matrices--the successes, the failures and the reasons for both. *Curr. Opin. Genet. Dev.*, **15,** 395-402.

42. Coletta,R.D., Christensen,K., Reichenberger,K.J., Lamb,J., Micomonaco,D., Huang,L., Wolf,D.M., Muller-Tidow,C., Golub,T.R., Kawakami,K., et al. (2004) The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. *Proc Natl Acad Sci U S A,* **101,** 6478-83.

43. Ruf,R.G., Xu,P.X., Silvius,D., Otto,E.A., Beekmann,F., Muerb,U.T., Kumar,S., Neuhaus,T.J., Kemper,M.J., M.,R.R.,Jr, et al. (2004) SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. *Proc Natl Acad Sci U S A,* **101,** 8090-5.

44. Wilson,D.S., Sheng,G., Jun,S. and Desplan,C. (1996) Conservation and diversification in homeodomain-DNA interactions: A comparative genetic analysis. *Proc. Natl. Acad. Sci. U. S. A.,* **93,** 6886-6891.

45. Tucker-Kellogg,L., Rould,M.A., Chambers,K.A., Ades,S.E., Sauer,R.T. and Pabo,C.O. (1997) Engrailed (Gln50-->Lys) homeodomain-DNA complex at 1.9 A resolution: Structural basis for enhanced affinity and altered specificity. *Structure,* **5,** 1047-1054.

46. Chaney,B.A., Clark-Baldwin,K., Dave,V., Ma,J. and Rance,M. (2005) Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause rieger syndrome. *Biochemistry,* **44,** 7497-7511.

47. Ades,S.E. and Sauer,R.T. (1994) Differential DNA-binding specificity of the engrailed homeodomain: The role of residue 50. *Biochemistry,* **33,** 9187-9194.

48. Grant,R.A., Rould,M.A., Klemm,J.D. and Pabo,C.O. (2000) Exploring the role of glutamine 50 in the homeodomain-DNA interface: Crystal structure of engrailed (Gln50 --> ala) complex at 2.0 A. *Biochemistry,* **39,** 8187-8192.

49. Seo,H.C., Curtiss,J., Mlodzik,M. and Fjose,A. (1999) Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. *Mech. Dev.,* **83,** 127-139.

50. Suh,C.S., Ellingsen,S., Austbo,L., Zhao,X.F., Seo,H.C. and Fjose,A. (2010) Autoregulatory binding sites in the zebrafish six3a promoter region define a new recognition sequence for Six3 proteins. *FEBS J.*, **277,** 1761-1775.

51. Hu,S., Mamedova,A. and Hegde,R.S. (2008) DNA-binding and regulation mechanisms of the SIX family of retinal determination proteins. *Biochemistry,* **47,** 3586-3594.

CHAPTER 3. SIX1 REGULATES MYOD EXPRESSION IN ADULT MUSCLE PROGENITOR CELLS

The following manuscript is published in *PLOS ONE*.

Liu, Y., Chakroun, I., Yang, D., Horner, E., Liang, J., Aziz, A., … Blais, A. (2013). Six1 regulates MyoD expression in adult muscle progenitor cells. *PloS One*, *8*(6), e67762.

I helped to optimize the immunofluorescent staining at the early stage of experiments for Figure 3.1. I also contributed to generating the frozen TA muscles injured by CTX. I helped to generate the Six1 ChIP-seq data for Figure 3.3A. I performed ChIP assays for Figure 3.3B and gel shift assays (EMSA) for Figure 3.3D. With the help with Yves (a technician in Dr. Kathory's lab), I performed the beta-gal staining for transgenic mice carrying the Myod CER for Figure 3.4B. I also performed luciferase assays for Figure 3.4C in differentiating C2C12 myoblasts and for Figure 3.4E in regenerating TA muscles. I established stable C2C12 cell lines carrying the WT CER and MEF3-mutated (Mutant) CER. I also performed ChIP assays on these two stable cell lines for Figure 3.5D and 3.5E. Lastly, I contribute to the editing of the manuscript.

Six1 regulates MyoD expression in adult muscle progenitor cells.

Yubing Liu[1,2,5], Imane Chakroun[1,2,5], Dabo Yang[1,2], Ellias Horner[1,2], Jieyi Liang[1,2], Arif Aziz[3], Alphonse Chu[1,2], Yves De Repentigny[3], F. Jeffrey Dilworth[3,4], Rashmi Kothary[3,4] and Alexandre Blais[1,2,*].

1, Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

2, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada

3, Ottawa Hospital Research Institute, Ottawa, ON, Canada

4, Departments of Medicine, and Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada

5, The first two authors contributed equally to this work.

Liu, Y., Chakroun, I., Yang, D., Horner, E., Liang, J., Aziz, A., … Blais, A. (2013). Six1 regulates MyoD expression in adult muscle progenitor cells. *PloS One*, *8*(6), e67762.

ABSTRACT

Quiescent satellite cells are myogenic progenitors that enable regeneration of skeletal muscle. One of the early events of satellite cell activation following myotrauma is the induction of the myogenic regulatory factor MyoD, which eventually induces terminal differentiation and muscle function gene expression. The purpose of this study was to elucidate the mechanism by which MyoD is induced during activation of satellite cells in mouse muscle undergoing regeneration. We show that Six1, a transcription factor essential for embryonic myogenesis, also regulates MyoD expression in muscle progenitor cells. Six1 knock-down by RNA interference leads to decreased expression of MyoD in myoblasts. Chromatin immunoprecipitation assays reveal that Six1 binds the Core Enhancer Region of *MyoD*. Further, transcriptional reporter assays demonstrate that Core Enhancer Region reporter gene activity in myoblasts and in regenerating muscle depends on the expression of Six1 and on Six1 binding sites. Finally, we provide evidence indicating that Six1 is required for the proper chromatin structure at the Core Enhancer Region, as well as for MyoD binding at its own enhancer. Together, our results reveal that MyoD expression in satellite cells depends on Six1, supporting the idea that Six1 plays an important role in adult myogenesis, in addition to its role in embryonic muscle formation.

INTRODUCTION

Skeletal muscle is a plastic organ that can regenerate itself after injury. This property relies on the presence of resident adult stem cells termed satellite cells (reviewed in [1]). In resting adult muscle, satellite cells are quiescent and are found in low numbers positioned between the basal lamina and the myofiber sarcolemma. Muscle injury leads to the activation of satellite cells: they are released from their anatomic position, initiate several rounds of cell division, and eventually undergo myogenic differentiation to create new muscle mass [2]. Recent work has demonstrated the importance of satellite cells in muscle regeneration [3-6].

At the gene expression level, quiescent satellite cells are characterized by the expression of the Paired-box transcription factor Pax7 [7], and by their lack of expression of muscle structural genes. Although the messenger RNA for the myogenic regulatory factor (MRF) Myf5 transcription factor is expressed by most quiescent satellite cells, the Myf5 protein itself is absent from these cells [8,9]. Likewise, the other proteins of this family, MyoD, myogenin and MRF4 are unexpressed in quiescent cells. Instead, MyoD expression is strongly induced early after injury, as satellite cells become activated [10,11]. After proliferation, satellite cells initiate their terminal differentiation: MyoD activates the expression of the MRF myogenin, as well as a host of muscle function genes, and the cells undergo fusion and exit the cell cycle.

Very little is known about the mode of up-regulation of MyoD expression in satellite cells; what we know of this gene's regulation concerns embryonic development. Three important *cis*-regulatory elements active during embryogenesis have been identified: the core enhancer region (CER), the distal regulatory region (DRR), and the proximal regulatory region (PRR, reviewed in [12]). Transgenic analyses have revealed that the CER, which is located 23 kb upstream of the transcription start site of the MyoD gene, drives expression in muscle precursor cells of the limb

buds and the myotome [13,14]. The essential role of the CER in controlling the proper temporal expression of MyoD has been demonstrated by deletion analyses: MyoD transcription is delayed by one to two days in the limb buds and branchial arches of mouse embryos lacking the enhancer, although myotomal expression, which depends on Myf5 and Pax3 expression, is not affected [15]. The DRR on the other hand, appears to drive expression of a LacZ reporter gene later, in differentiated cells of the muscle lineage, and possibly also in adult muscles [16,17]. Deletion analyses have confirmed the important role of the DRR in adult muscle expression of MyoD, but have also revealed that the DRR is in fact necessary, albeit not sufficient, to drive full MyoD expression levels in the limb buds and branchial arches at E10.5 [18]. Finally, the PRR, which in mice has been defined as a 275-bp sequence immediately preceding the transcription start site, does not have muscle-specific activity in mammalian cells, but cooperates specifically with the DRR to confer muscle specificity [19].

Various factors that act upon these three regulatory elements have been identified. AP-1, the dimer formed by Jun and Fos family members, activates transcription of a reporter gene containing the MyoD PRR [20]. The PRR is also bound by a CCAAT-binding activity in a muscle-specific manner, and its transcriptional activity depends on a consensus Sp1 binding site. The MRFs Myf5 and MyoD themselves are suspected to act upon the CER and DRR in various developing structures, as transgenic reporter activities are diminished in $MyoD^{-/-};Myf5^{-/-}$ embryos [21]. Although direct binding of these MRFs to the CER or DRR has not been assessed in the embryos, both elements contain E-box motifs [14,19], and MyoD has been shown by ChIP-seq to bind to both elements in C2C12 myoblasts [22], and to bind the DRR in differentiating myoblasts [23]. The homeobox factor Msx1 represses MyoD expression and delays the execution of the myogenic program in the distal portion of the developing limb bud,

by binding to the CER and inducing the formation of a repressive chromatin environment [24-26]. In an analogous manner, the bHLH-PAS factor Sim2 can also repress MyoD expression by binding to the CER in the developing limb bud [27]. In contrast, the homeodomain transcription factor Pitx2 is involved in the initiation of MyoD expression in the limb bud, and directly binds to the CER [28]. In the adult, binding of Bmal1 and Clock circadian bHLH-PAS transcription factors to the CER is responsible for the circadian rhythmicity of MyoD expression in skeletal muscle [29,30]. In adult muscle regeneration, Mef2 factors and their co-regulator MASTR have been shown to bind to the DRR to activate MyoD expression in satellite cells [31]. Likewise, Serum Response Factor (SRF) can bind to the DRR, and DRR-dependent transgene expression in regenerating adult muscle depends in part on the presence of the SRF binding site [32,33].

The Six family of homeodomain transcription factor has also been shown to play a role in regulating MyoD expression. The Six family consists of 6 members, from Six1 to Six6. They are involved in controlling the differentiation of various tissue types [34]. Six1 and Six4 are important for embryonic myogenesis: $Six1^{-/-}$ mouse embryos have important myogenesis defects that are exacerbated in $Six1^{-/-}$;$Six4^{-/-}$ animals [35-37]. Importantly, MyoD expression is severely impaired in the absence of Six1, in the limb bud as well as in the epaxial extension of the dermomyotome and the ventral myotomal extension. In animals lacking both Six1 and Six4, a Six family factor that may partially compensate for the absence of Six1, MyoD expression is further affected and persists only in the hypaxial-ventral portion of the interlimb somites myotome. $Six1^{-/-}$ and $Six1^{-/-}$;$Six4^{-/-}$ fetuses die at birth from respiratory failure and it is only recently that the role they play in adult satellite cells during regeneration has been revealed. Conditional ablation of Six1 in Pax7-expressing satellite cells leads to a regeneration failure following injury, due to myoblast differentiation impairment. MyoD expression is diminished in

these cells, and this is thought to be mediated at least in part by a binding site for Six1 in MyoD's DRR [23].

We have previously reported a genome-wide profiling of Six1 binding sites in C2C12 mouse myoblasts and myotubes. We found that Six1 and Six4 are necessary for the myogenic differentiation of these cells, and that they accomplish their function in part by activating shared target gene expression in cooperation with the MRFs [38]. We have now investigated the function of Six1 in primary adult mouse myoblasts, and focused on the role it plays in regulating the expression of MyoD.

MATERIALS AND METHODS

Ethics statement

Surgical procedures were performed using aseptic techniques and in complete agreement with the University of Ottawa Animal Care and Use Committee in compliance with the Guidelines of the Canadian Council on Animal Care and the Animals for Research Act. The University of Ottawa Animal Care and Use Committee approved this study.

Muscle injury and immunostaining

Seven-week-old C57BL/6 female mice (Charles River, Canada) were first anesthetized with isoflurane. Both legs were shaved, and 30µl of cardiotoxin (CTX, Latoxan, France, 10mM in phosphate-buffered saline (PBS)) were injected into the left TA, and 30µl of PBS were injected into the right TA muscle as control. A total of 5 mice were used for each time point to account for biological variability. Mice were sacrificed by cervical dislocation 2, 3, 4 or 7 days after cardiotoxin injection. Their TA muscles were dissected and fixed in 2 % (w/v) paraformaldehyde in PBS overnight at 4ºC. Muscles were then washed 2 times for 5 minutes in cold PBS (4ºC), and quenched with 0.25 M glycine in PBS 2 times for 10 minutes on ice. Muscles were then incubated in 5% sucrose (in PBS) for a minimum of 6 hours, and then incubated in 20% sucrose (in PBS) overnight at 4ºC. Muscles were frozen with Tissue-Tek O.C.T. (Optimal Cutting Temperature) Compound in 2-methylbutane chilled in liquid nitrogen, and stored in -80ºC freezer. Muscles were cryosectioned at -28ºC into 10 µm sections on Fisherbrand Superfrost Plus slides and stored at -80ºC until antigen retrieval was performed.

Slides were dried at 37ºC for 10 minutes, then incubated in citrate buffer (10mM citric acid, 0.05% Tween 20 (v/v), pH 6.0) for 20 minutes at 98ºC, and slides were cooled down at room temperature in citrate buffer for 20 minutes. Sections were subsequently rinsed with PBS-T (PBS, 0.1% Tween 20 (v/v)) and then permeabilized with 0.5% Triton X-100 (v/v) in PBS for 10 minutes at room temperature. Sections were rinsed again with PBS-T and blocked with blocking buffer (3% bovine serum albumin (w/v) in PBS-T) for 1 hour at room temperature (or overnight at 4ºC). Sections were further blocked for 1 hour at room temperature with 3.6% (v/v) M.O.M. Mouse Ig Blocking Reagent solution in PBS (Vector Laboratories), then washed twice with PBS. Sections were incubated with primary antibodies against Six1 (1:100 v/v rabbit anti-mouse, made in-house), MyoD (1:100 v/v Santa Cruz, sc-32758X, 5.8A mouse monoclonal), Pax7 (1:100 v/v Developmental Studies Hybridoma Bank) in blocking buffer for 1 hour at room temperature, and rinsed with PBS-T. Adjacent sections without primary antibodies were also prepared, to ensure the specificity of the resulting fluorescent signals. Sections were incubated with 1:200 biotin-conjugated donkey anti-mouse IgG F(ab') fragment (Jackson Immunoresearch) in blocking buffer for 1 hour at room temperature, and washed in PBS-T. Sections were incubated with secondary antibodies conjugated to Alexa594 (1:1000, Invitrogen, donkey anti-rabbit IgG) and with Steptavidin-Alexa488 (1:1000, Invitrogen) in blocking buffer for 1 hour at room temperature in the dark, and rinsed with PBS-T. Sections were rinsed with PBS, then de-ionized water, and with 70% ethanol before incubation with 0.3% Sudan Black (w/v, Sigma) in 70% ethanol for 10 minutes. Slides were washed with PBS and washed with de-ionized water. Slides were air-dried and mounted with ProLong Gold antifade reagent with DAPI (Invitrogen).

Images were acquired with a Carl Zeiss Axiovision Observer D1 microscope operated with the Axiovision Rel 4.8 software, using a 20x objective. Six images, each representing fields of

670µm by 896µm (0.6 mm^2) were acquired per sample, at different locations on the muscle section. For the purpose of preparing figures, Adobe Photoshop was used to adjust levels (black and white points, but not gamma), taking care to avoid clipping pixels and applying changes to the entire image. Cell counting was performed with the ImageJ 1.45 software and the Cell Counter plug-in (W. Rasband, National Institutes of Health). Nuclei positive for either of the proteins of interest were counted in each image, and the average was calculated. The average and standard error of those counts over the 5 replicates were finally calculated.

Cell culture

The C2C12 cells (American Type Culture Collection) were grown in Growth Medium (GM) containing 88% Dulbecco's Modified Eagle's Medium (DMEM); 10% Fetal Bovine Serum (FBS); 1% L-Glutamine and 1% Penicillin/Streptomycin (P/S)) until confluent. Once confluent, the cells were induced to differentiate by replacement with Differentiation Medium (DM) containing 96% DMEM; 2% Horse Serum; 1% L-Glutamine and 1% P/S. The cells were grown in a humidified water-jacketed incubator at 37ºC with 5% CO2. The C2iFRT cell line that constitutively expresses the Tet-repressor and contains a single genomic FRT recombination site insertion has been previously described [39]. A cDNA encoding N-terminally Flag-tagged mouse MyoD was cloned into the MCS of the pCDNA5/FRT/TO plasmid. The pCDNA5/FRT/TO-FL-MyoD plasmid was co-transfected with the Flp recombinase into C2iFRT cells and selected for hygromycin expression as outlined in the Flp-In system protocol (Invitrogen). Batch cultures of C2iFRT-FL-MyoD cells were then screened for their ability to differentiate to form myotubes and for MyoD expression after induction with doxycycline at a concentration of 0.5 µM. Primary myoblasts were prepared as described by Rando et al. [40] with the modifications outlined below. The primary myoblasts were isolated from 60 day old C57BL/6 female mice.

Gastrocnemius, *tibialis anterior* (TA) and *quadriceps* muscles were pooled and digested with 0.2% Collagenase I (Sigma) and 625 μg/mL Dispase II (Roche) for 1.5 - 2.0 hours at 37°C. Cells were then diluted with DMEM and passed through a 70 μm nylon mesh filter (BD Falcon) to remove undigested connective tissues. Cells were rinsed twice with DMEM and resuspended in plating medium containing 90% DMEM; 10% donor equine serum and 5 ng/mL of bFGF (Peprotech). Cells were then pre-plated twice for 1h in a 10 cm tissue culture treated dish (Corning), transferred to Matrigel (BD Biosciences) coated cell culture dishes and allowed to adhere to the plate for 48h. Cells were subsequently maintained in DMEM (American Type Culture Collection) with 20% Fetal Bovine Serum (HyClone), 10% Donor Serum (HyClone) and 1% P/S (HyClone) supplemented with 10 ng/mL of bFGF and 2 ng/mL of bHGF (Peprotech).

RNAi

Primary and C2C12 myoblasts were transfected with siRNA duplexes (non-silencing or Six1-specific) using Lipofectamine RNAiMax (Invitrogen), essentially as described previously [38].

Six1 antibody preparation

New Zealand white rabbits were immunized with full-length histidine-tagged recombinant mouse Six1 protein, as described previously [38]. The serum was purified by running rabbit serum through a column of immobilized GST-Six1 protein (amino acids 198-248, sharing no homology to other family members), washed several times with tris-buffered saline and eluted with a solution of 0.1 M glycine pH 2.5 to elute specific anti-Six1 antibodies. Specificity was confirmed by western blot on all six full-length proteins of the murine Six family, produced in rabbit reticulocyte lysates: only Six1 is detected by the purified antibodies (data not shown).

Western blots

Total protein lysates were prepared by rinsing the cells in PBS and lysing in 20 mM Tris pH 6.8, 6 M urea and 0.1% SDS. Lysates were sonicated briefly and spun down to remove debris. Primary antibodies used are anti-Six1 (rabbit, home-made), anti beta-actin (mouse monoclonal, Sigma), anti-MyoD (mouse monoclonal 5.8A, Santa-Cruz) or anti-beta-tubulin (mouse monoclonal, E7, Developmental Studies Hybridoma Bank).

ChIP assays and ChIP-sequencing

Chromatin immunoprecipitation assays were performed as described before [38], for both C2C12 and primary myoblasts. The antibodies used were rabbit anti-Six1 [38] and normal rabbit IgG (Jackson Immunoresearch). For ChIP assays on transfected cells, a 9:1 mixture of either wild-type or mutant reporter plasmid (described below) and puromycin resistance plasmid was transfected in C2C12 myoblasts using the polyethylenimine method [41]. Cells were selected with puromycin for 7 days, resistant clones were pooled to generate polyclones and expanded, and chromatin was prepared as described before [38]. PCR primer sequences used in ChIP are: CER-F: TGCTTCTTTCGGCCAAGTAT; CER-R: CCAACTGGCTGTGTTGTGAG; HoxD10-F: GAGAAATCGGACTCACCTTCC; HoxD10-R: CACATACCCAGGCAGAACG. For PCR to distinguish the endogenous CER and the CER transgene, primers were a-GTTGGGGGAAGGGGACAG; b- GACTCCAGGAAGGAAGAAGAGG; c-ACCCGTGACTCACAACACAG; d- TCTCCAGTGTCTACTCGAG. Quantitative PCR was performed on input chromatin from the wild-type and mutant polyclones and on a titration curve made with the pure reporter plasmid to ensure that the transgene copy numbers are comparable for both polyclones; the wild-type and mutant polyclones contain respectively 2.1 and 2.2 copies

of transgene per cell (data not shown). Transgene ChIP data were analyzed as follows. First, qPCR titration curves made of input chromatin from the wild-type or the mutant polyclones were run in parallel to the ChIP samples, so "percent-of-input" values could be ascribed to each ChIP sample. Second, the percent of input values obtained with the non-specific antibody control (normal IgG) were subtracted from the percent-of-input values obtained with the other antibodies. Third, those IgG-subtracted percent-of-input values were reported as fractions of the values obtained with the wild-type CER polyclone.

ChIP-seq experiments were performed on 50 million primary myoblasts in growth phase. Chromatin was fragmented to an average size of 200 bp and immunoprecipitated using rabbit anti-Six1 or a control rabbit IgG (Jackson). An input chromatin sample (prior to immunoprecipitation enrichment) was also prepared. After purification, sequencing libraries were prepared by the McGill University and Génome Québec Innovation Centre, and sequenced at 1 sample per lane on HiSeq2000. Sequencing reads were aligned to the mm9 mouse genome assembly using Bowtie in –n mode [42], allowing 0 mismatches in the first 36 nucleotides of each read, and removing reads that align at more than one location in the genome. Picard was used to filter out replicated reads (http://picard.sourceforge.net/). SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) was used to extend reads to a length of 200 bp, to normalize read counts to the total number of retained reads in each sample, and to calculate normalized read densities for each sample in contiguous, non-overlapping bins of 25 base pairs. The read density in the input sample was finally subtracted from the immunoprecipitation samples. These read densities are given as wiggle format files as Supplementary table S1. A full description of the ChIP-seq results will be published elsewhere (Y.L. and A.B., in preparation).

EMSA

Full-length CER DNA probes were prepared by PCR amplification from wild-type or mutant plasmids, restriction digestion and fill-in with Klenow enzyme (Promega) in the presence of Cy5-labelled dCTP. Recombinant Six1 was produced in rabbit reticulocyte lysates (Promega). Gels were scanned on a Typhoon imager (GE Healthcare).

Reporter constructs and reporter assays

PCR on C2C12 genomic DNA was used to amplify the murine CER and PRR regions separately, adding restriction sites for cloning: CER-F: GACGACGCTAGCTGAGCCCCACAGCATTTGGG, CER-R: GACGACCTCGAGCCCCAGCCCTAGGCCTGAGC; PRR-F: GACGACCTCGAGTAGACACTGGAGAGGCTTGGG; PRR-R: GACGACAGATCTAGGCGCCCTGGGCTATTTATCC. The PRR was cloned upstream of the luciferase gene in pGL3-Basic (Promega) and the CER was subsequently cloned upstream of the PRR. Mutations were introduced in the CER region by sequential overlapping mutagenic PCR reactions. The wild-type CER+PRR were additionally cloned upstream of the LacZ reporter in the p1230 plasmid [43], from which the beta-globin minimal promoter had been removed by restriction and re-ligation. The CMV-Renilla luciferase plasmid (Promega) was used as an internal control. The Myogenin promoter wild-type and MEF3 site mutant constructs were described previously [38].

Mice were injected with cardiotoxin, and three days later were injected with 25 uL of a saline solution containing 12.5 ug of reporter DNA (10 ug luciferase, and 2.5 ug CMV-Renilla). The DNA was then electroporated. Four days later, the animals were sacrificed by cervical

dislocation, their TA muscles were removed and quickly frozen in liquid nitrogen. The tissue was crushed to a fine powder and resuspended in passive lysis buffer (Promega). C2C12 cells were transfected using the polyethylenimine method, and harvested 48 hours after transfection, by rinsing in PBS and lysing in passive lysis buffer. Lysates were assayed for firefly and renilla luciferase activities using the Dual Luciferase assay kit (Promega). To normalize across samples, firefly luciferase activity values were divided by those of renilla luciferase. Where indicated, this ratio was further normalized by dividing by the normalized luciferase readings obtained in control conditions.

Transgenic reporter mice

We tested in mouse embryos the fidelity of our mouse CER-PRR construct as a fusion with the LacZ reporter gene, by pronuclear injection of DNA using standard methods [44]. Founder embryos were harvested at 11.5 days of gestation, and were fixed and stained with X-Gal to reveal beta-galactosidase activity. A total of 15 founder embryos, from 2 separate litters, gave comparable results.

RESULTS

Six1 is expressed in satellite cells of regenerating muscle

To study the function of Six1 during adult muscle regeneration, and to address the question whether it is involved in the regulation of MyoD expression, we started by determining the profile of Six1 and MyoD protein expression in the TA muscle of adult mice at various time points following injury. Intra-muscular injection of cardiotoxin was used as the injury model and immuno-fluorescence was performed on frozen muscle cross-sections to detect protein expression. First, sections were co-stained with antibodies against Six1 and the satellite cell marker Pax7 [7] (Fig. 1A and 1B). Six1 was undetectable in quiescent satellite cells from resting, uninjured muscles. Instead, we found the protein only in the myonuclei of myofibers. However, Six1 protein was clearly detected in Pax7-positive cells starting approximately three days following injury. This coincides with the surge in satellite cell numbers that occurs as these cells initiate proliferation following injury: at day 3 post-CTX, essentially all Pax7-positive cells are also Six1-positive. At later time-points, the situation is different: Six1 is highly expressed in the centrally-located nuclei of nascent or newly regenerated myofibers which are Pax7-negative, and Pax7-positive cells remain in high numbers but return to a state where they are Six1-negative.

MyoD has been reported to be expressed in satellite cells following injury [10,11]. To determine if Six1 and MyoD are co-expressed, which is a logical requirement for a role of Six1 as regulator of MyoD expression, we co-stained regenerating muscle sections with antibodies against those two factors (Fig. 1C and 1D). We found that MyoD is undetectable in myonuclei or mononucleated cells from resting muscle, but that it is expressed along with Six1 at three and four days following CTX injection: virtually all MyoD-positive mononucleated cells are also

Six1-positive at these two time-points. Seven days following injury, the numbers of MyoD-positive cells decrease but the cells remain mostly Six1-positive. Considering the staining pattern and the results in Fig. 1A, showing that at this time point the majority of Pax7-positive cells are Six1-negative, we reason that the double MyoD-positive/Six1-positive cells at day 7 are not satellite cells but rather represent the differentiating progeny of satellite cells (*i.e.* myocytes) or small, nascent myofibers. From these experiments, we conclude that Six1 is expressed in satellite cells as they become activated following injury, and that MyoD-positive satellite cells of regenerating muscle are also Six1-positive. This strong correlation between Six1 and MyoD expression is consistent with a role of Six1 as regulator of MyoD expression in activated satellite cells.

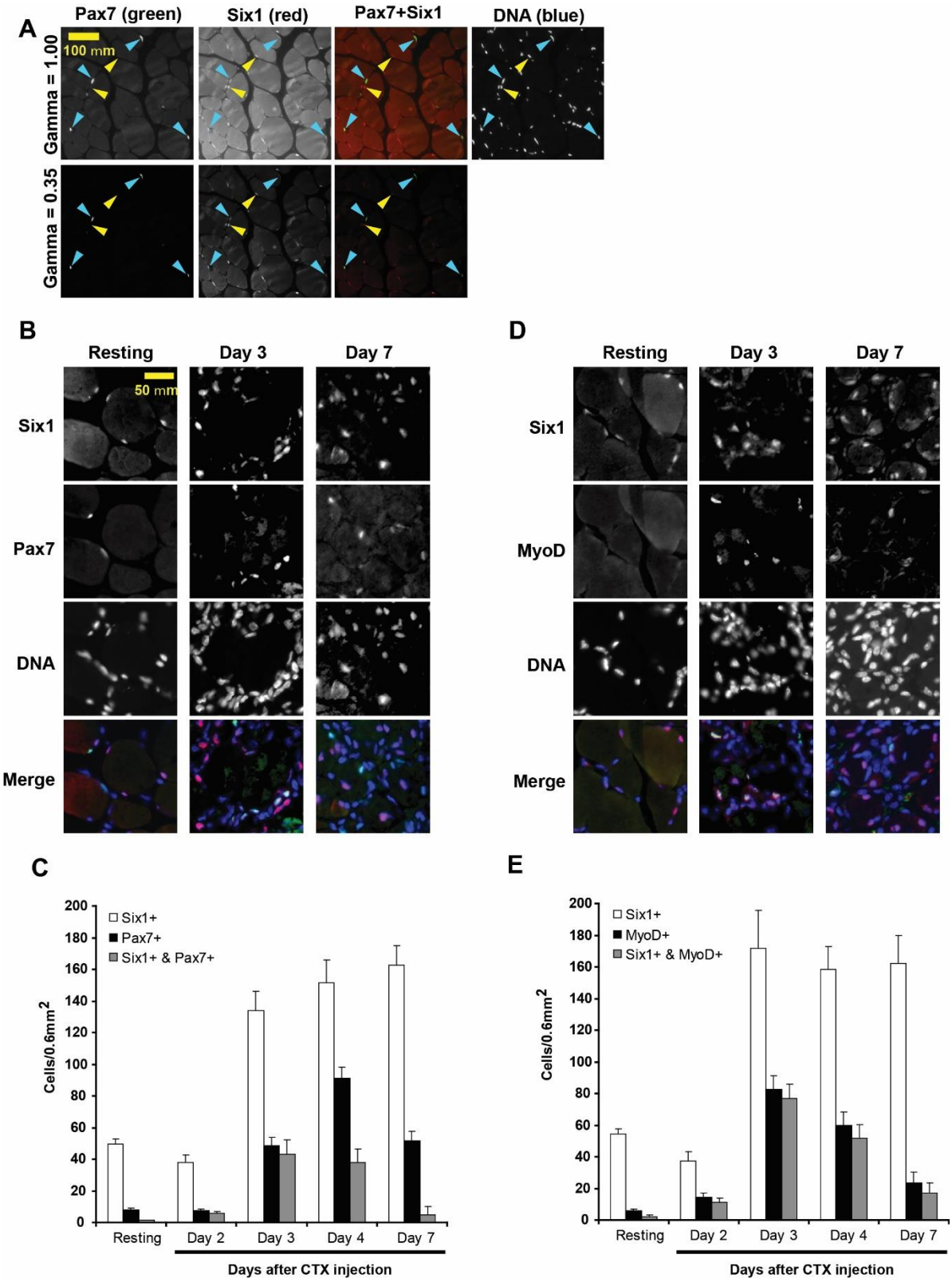

A
Pax7 (green)
Six1 (red)
Pax7+Six1
DNA (blue)
100 mm
Gamma = 1.00
Gamma = 0.35
B
Resting
Day 3
Day 7
50 mm
Six1
Pax7
DNA
Merge
D
Resting
Day 3
Day 7
Six1
MyoD
DNA
Merge
C
Six1+
Pax7+
Six1+ & Pax7+
Cells/0.6mm2
Resting
Day 2
Day 3
Day 4
Day 7
Days after CTX injection
E
Six1+
MyoD+
Six1+ & MyoD+
Cells/0.6mm2
Resting
Day 2
Day 3
Day 4
Day 7
Days after CTX injection

Figure 3. 1 Six1 is expressed along with Pax7 and MyoD in activated satellite cells of regenerating muscles.

A) Immunostaining of sections from paraffin-embedded resting TA muscles, using antibodies against Pax7 (green) and Six1 (red). Pax7-positive satellite cells are marked by blue arrowheads, while Six1-positive cells are marked by yellow arrows. Gamma settings were adjusted to 1.00 or 0.35 to increase the signal-to-noise ratio. DAPI was used as a counterstain to label nuclei. **B)** Immunostaining of frozen sections of resting TA, or muscles after 3 or 7 days following cardiotoxin injection, with antibodies against Six1 (red signal) and Pax7 (green signal). DAPI was used as a counterstain to label nuclei. **C)** Quantification of the anti-Six1 and anti-Pax7 staining signal in resting muscles and at 2, 3, 4 or 7 days post-injury, as shown in panel A. Bars indicate the average number of positively stained nuclei counted in 0.6 mm^2 fields of view, using 5 mice per condition. Error bars, S.E.M.. **D)** Immunostaining performed in samples identical as in A, but using antibodies against Six1 and MyoD. Magnification as shown in A. **E)** Quantification of the anti-Six1 and anti-MyoD staining as shown in C.

Six1 is required for MyoD expression in myoblasts

To more directly address the question of whether Six1 is required for MyoD expression, we used primary myoblasts freshly isolated from adult mice as a model. Western blots reveal that proliferating primary myoblasts express appreciable levels of Six1 and MyoD, and that the levels decrease over time as the cells are induced to differentiate (Fig. 2A). We next tested whether MyoD protein levels in primary myoblasts depend on Six1 expression, by knocking down Six1 in growth phase myoblasts with siRNA against Six1 (siSix1), or non-silencing duplexes (siNS), and extracting total proteins from the cells 48 hours later, still in growth conditions. Western blot results in primary cells agree with those obtained with C2C12 cells (Fig. 2B): MyoD requires Six1 for its expression. An appreciable decrease in MyoD protein levels was also observed when Six1 expression was knocked-down using two different lentiviruses expressing distinct short hairpins against the Six1 mRNA (data not shown). These results indicate that MyoD expression in primary myoblasts depends on Six1 function, and suggest that Six1 performs a similar regulatory role *in vivo* in regenerating muscle.

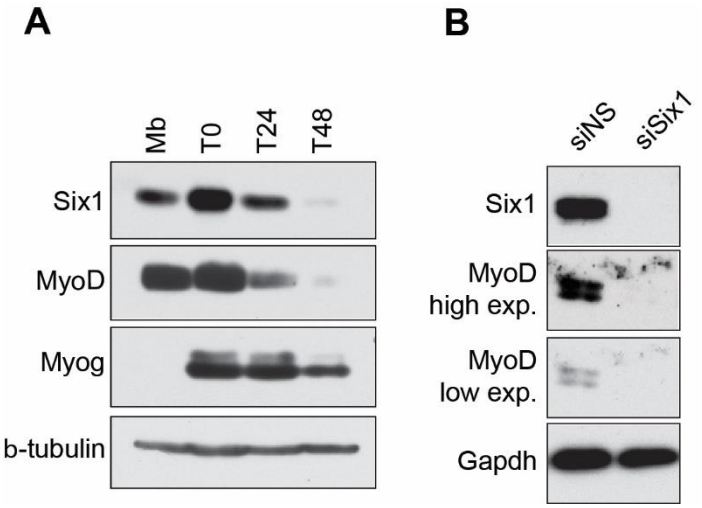
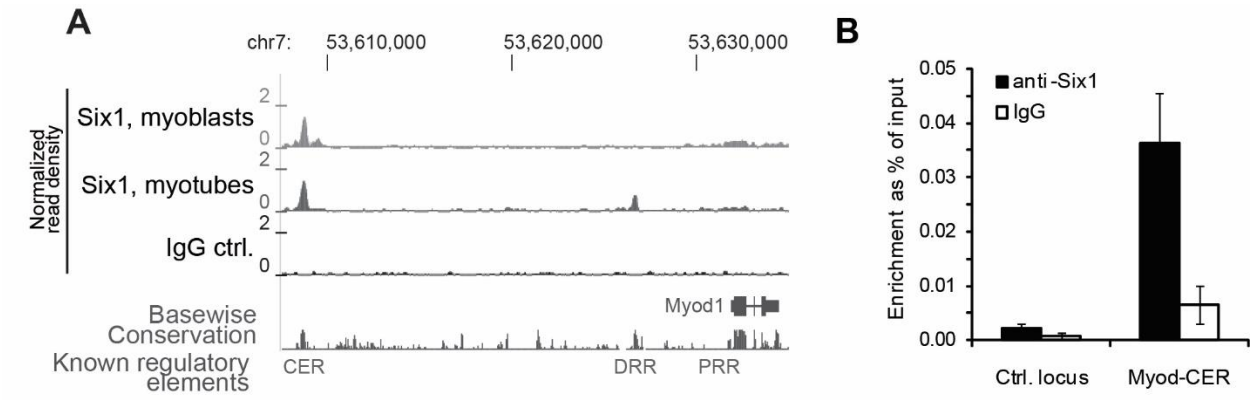

A
Mb
T0
T24
T48
Six1
MyoD
Myog
b-tubulin
B
siNS
siSix1
Six1
MyoD
high exp.
MyoD
low exp.
Gapdh

Figure 3. 2 Six1 is expressed in primary myoblasts and is necessary for MyoD expression.

A) Western blot on total protein lysates of primary myoblasts in growth phase (Mb), at confluence (T0), differentiated for 24 (T24) or 48 hours (T48). The antibodies used were anti-Six1, anti-MyoD and anti-myogenin. Anti-β-tubulin was used as a loading control. **B)** Western blot showing the expression of Six1, MyoD and GAPDH on total protein lysates of primary myoblasts in growth phase, 48 hours after their transfection with siRNA duplexes targeting Six1 (siSix1), or with a non-silencing siRNA (siNS). A low and a high film exposure are shown for the anti-MyoD western blot. Comparable results were obtained in three independent experiments.

Six1 binds the MyoD CER in primary myoblasts

We have previously reported ChIP-on-chip analysis of Six1 binding in C2C12 myoblasts. We repeated similar experiments, this time using ChIP-sequencing (ChIP-seq) on chromatin prepared from primary myoblasts in their growth phase. A full analysis of these results will be described elsewhere (Y. Liu et al., in preparation). We analyzed the binding profile of Six1, reported as normalized read density, across the MyoD locus, and found that Six1 binds to the CER in myoblasts and in myotubes (Fig. 3A). We confirmed the binding of Six1 to this enhancer, using ChIP on independent primary myoblast chromatin preparations, thereby ruling out biases potentially introduced by the high-throughput sequencing approach (Fig. 3B). Examining the CER sequence, we found two MEF3-like elements that comply with the sequence elements we have shown Six1 is able to bind [45] (Fig. 3C). To determine if Six1 is able to directly bind to these elements, we performed electrophoretic mobility shift assays (EMSAs) using recombinant mouse Six1 produced in rabbit reticulocyte lysates and a fluorescently-labeled probe representing the full-length CER sequence. Competition using an excess of unlabelled oligonucleotides representing the MEF3 site of the Myogenin promoter [46], or a mutant version, was used to assess the specificity of protein binding. Six1 is indeed able to bind the CER directly and specifically, and this depends on the presence of at least one of the two MEF3 sites, since mutation of both sites was necessary to completely abolish protein binding in this assay (Fig. 3D).



C

	MEF3 site #1 (rev. comp.)	MEF3 site #2
Mouse	AGCAGGATAAAATTACTGTC	GGTCTTCTCCGGTTTC-TC------TA
Rat	AGCAGGATAAAATTACTGTC	GGTCTTCTCCGGTTTC-TC------TA
Human	AGCAGGATAAAgTTACccTC	GGcCTTCTCtGGTTTC-TC------cA
Dog	AGCAGGATAAAgTTACccTC	GGcCTTCTCgGGTTTCtTC------cA
Horse	AGCAGGATAAAgTTACccTC	GGcCTTCTCgGGTTTC-TC------cA
Opossum	AGCgGGATAAAgTTACcagC	GcaCaTCTCaGGTTTC-CCAGAACATA
Mutant	AGCAGGAcccAAccACTGTC	GGTCTTCaaaGGTTaa-TC------TA

D

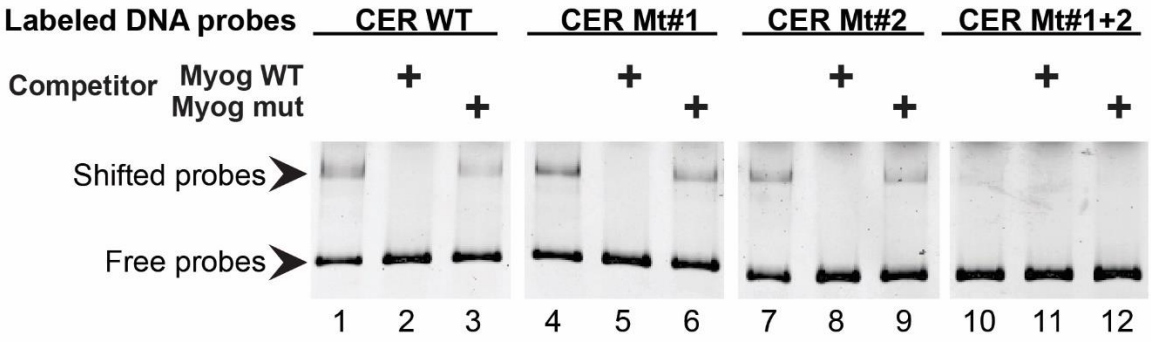


Figure 3. 3 Six1 directly binds to the core enhancer region of MyoD at two conserved MEF3 sites.

A) Profile of genomic binding of Six1 in primary myoblasts in growth phase, showing binding at the CER. The read density is expressed as reads per million mappable reads in bins of 25 base pairs above the read density in the input sample. The signal obtained with a non-specific antibody (non-immune rabbit IgG) is shown for comparison. **B)** Conventional gene-specific ChIP assays followed by real-time PCR were used to confirm the binding of Six1 to the CER, in proliferating myoblasts. n=3 biological replicates (independent chromatin preparations). By one-tailed paired t test, the signal for anti-Six1 at the CER is significantly above that obtained on the negative control locus, and above that obtained at the CER with normal rabbit IgG, with $p < 0.05$. Error bars, S.E.M.. **C)** Sequences of the two MEF3 sites identified within the mouse core enhancer. The murine MyoD gene is on the + strand; the reverse-complement of site #1 is shown. Conservation across mammalian species is shown, along with the mutations created in the EMSA probes and reporter constructs. Positions in small script indicate divergent sequences using mouse as reference. Dots indicate sequence not shown; dashes indicate missing sequences in certain species. **D)** Direct binding of Six1 to the CER, shown by EMSA experiments using recombinant Six1 protein incubated with a wild-type CER probe, or with versions mutated at either or both MEF3 sites identified. Specificity of binding was assessed by competition with a 50-fold molar excess of unlabelled myogenin MEF3 site oligonucleotides, either wild-type sequence or mutated.

MyoD CER activity requires Six1 binding

Considering that Six1 binds to the CER directly, we next aimed to determine if the enhancer's activity depends on Six1. We cloned the murine CER and PRR elements one after the other in front of the LacZ and luciferase reporter genes (Fig. 4A). Since the murine CER has never been studied in this context, we first verified that our construct drives reporter gene expression faithfully in transgenic mouse embryos, based on the published literature on the human enhancer [14,15,47] and endogenous mouse gene [14]. Using LacZ as the reporter, we detected β-galactosidase activity in the somites and limb buds of mice at embryonic day 11.5 with a pattern similar to that described previously by others (Fig. 4A,B). Among other features, β-galactosidase staining anterior to the forelimb bud is most visible at the dorsal part of the myotome, while it is most obvious in the ventral myotome posterior to the forelimb bud. We next created a luciferase reporter construct, moving the mouse CER and PRR to the pGL3-Basic plasmid backbone. Transfection of this construct in C2C12 cells subsequently transfected with control or Six1-targeting siRNA duplexes revealed that the CER is active in C2C12 cells, and that this activity depends on Six1 expression (Fig. 4C). In contrast, a related construct containing only the PRR had a lower activity that was not dependent on Six1. We used the myogenin promoter, a well-known Six1 target gene, as a positive control in these assays; the Myog reporter behaved the expected way by showing a reduced activity upon Six1 knock-down. In order to determine which of the two MEF3 elements contributes to the enhancer activity, we also constructed mutant versions of the CER+PRR reporter where the two MEF3 sites were mutated singly or in combination. The reporters were transfected in primary myoblasts, and the cells were harvested in growth phase or after two days in differentiation medium to induce myotube formation. The results show that MEF3 site #2 is by far the most active since its mutation caused the greatest

reduction in luciferase activity (Fig. 4D). In contrast, site #1 mutation had a slight but not statistically significant effect on reporter activity, both in the contexts of the wild-type or mutated site #2. Finally, to assess the *in vivo* significance of the Six1 binding sites on CER activity, we transfected resting or regenerating TA muscles with the wild-type or double MEF3 sites mutant CER+PRR luciferase constructs, and determined reporter activity four days later. We found that the CER reporter activity is significantly higher in regenerating muscle, and that this depends on the presence of the Six1 binding sites (Fig. 4E). Although the luciferase activity may originate from various cell types in addition to satellite cells, our results indicate that the CER of MyoD is a functional binding site for Six1 and that enhancer activity depends on Six1 function.

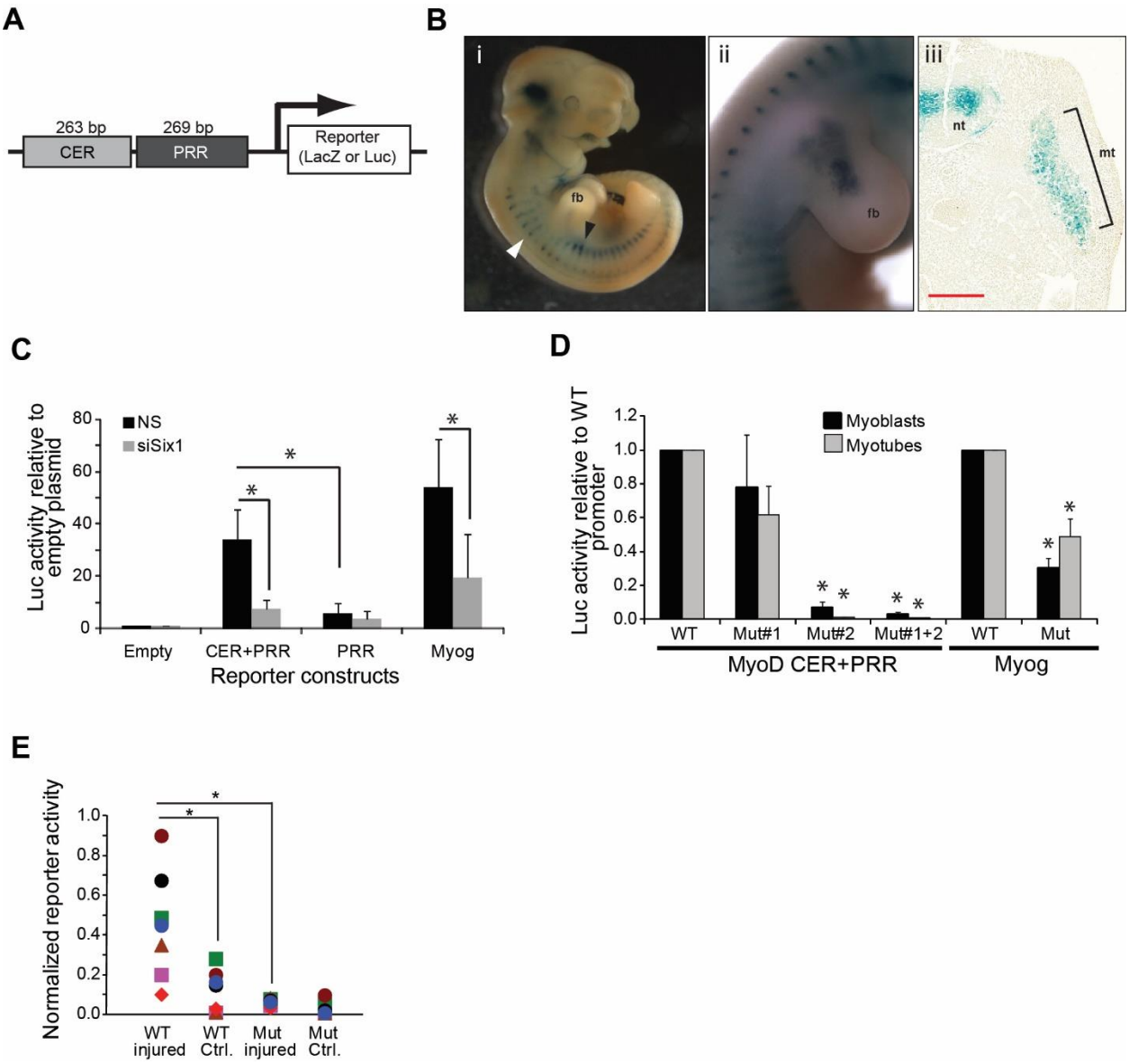

A
263 bp
269 bp
CER
PRR
Reporter
(LacZ or Luc)
B
i
ii
iii
fb
fb
nt
mt
C
Luc activity relative to empty plasmid
NS
siSix1
Empty
CER+PRR
PRR
Myog
Reporter constructs
D
Luc activity relative to WT promoter
Myoblasts
Myotubes
WT
Mut#1
Mut#2
Mut#1+2
WT
Mut
MyoD CER+PRR
Myog
E
Normalized reporter activity
WT injured
WT Ctrl.
Mut injured
Mut Ctrl.

Figure 3. 4 CER reporter constructs depend on Six1 expression and on Six1 binding sites for maximal activity.

A) Schematic representation of the reporter constructs used for these experiments. The backbone for luciferase is pGL3-Basic, while that for LacZ is p1230. **B)** The murine CER+PRR LacZ construct drives reporter gene expression at the expected locations, in E11.5 transgenic founder mouse embryos. ii) The white arrowhead points to the dorsal part of the somites, with enhanced reporter activity. The black arrowhead points to the ventral part of the myotomes. fb, forelimb bud. ii) Forelimb bud signal on a different embryo. Signal is often seen in the hindlimb bud as well, on other embryos (not shown). iii) A cross-section at the inter-limb level reveals that the ventral signal seen in i) comes from the myotome (mt). Non-specific transgene expression is also detected in the neural tube (nt). **C)** CER enhancer activity depends on Six1 expression, as shown by promoter reporter assays performed in C2C12 myoblasts transfected first with the indicated luciferase reporter plasmids, and 24 hours later with the indicated siRNAs: control non-silencing or targeting Six1. The normalized luciferase activity readings are reported as fold over the numbers obtained with the empty pGL3-Basic plasmid and non-silencing siRNA. Bars represent the average of 3 biological replicates; error bars, S.E.M.. Asterisks indicate significance ($p < 0.05$) by two-tailed paired t test. **D)** CER enhancer activity depends on Six1 binding sites, as shown by reporter assays performed in primary myoblasts and myotubes. Myoblasts were transfected with the indicated reporter constructs, and either harvested as myoblasts or induced to differentiate for 48 hours prior to harvest. For comparison, the effect of MEF3 site mutation is also shown for the myogenin promoter. In each case, reporter activity is reported as fraction of the activity of the wild-type reporter. n= 3, asterisks indicate significance by two-tailed paired t test ($p < 0.05$). Error bars, S.E.M.. **E)** CER enhancer activity increases in regenerating muscle

and this depends on the Six1 binding sites. Wild-type or MEF3-mutated reporter constructs were injected and electroporated in uninjured TA muscles, or in TA muscles 3 days post-injury by cardiotoxin injection. The transfected tissues were harvested 4 days later for luciferase assays. Values reported are normalized luciferase readings for each individual mouse leg harvested (n = 8, each depicted by a different symbol). Significance of reporter activity differences was assessed by Wilcoxon rank-sum test, with $p < 0.05$ as threshold.

Six1 is necessary for proper chromatin structure and for MyoD binding at the CER

Since Six1 has been shown to interact with proteins that can alter chromatin structure [37,48], we reasoned that chromatin remodeling might underlie the regulatory role of Six1 at the CER. We analyzed ChIP-seq data for the H3K4me1 (mono-methylated histone H3 lysine 4) in C2C12 myoblasts [49], since this is a mark associated with transcriptional enhancers [50]. Figure 5A shows the Six1, MyoD, H3K4me1 and mono-nucleosomes location profiles in myoblasts. Interestingly, the peak of Six1 binding is located very close to the peak of MyoD binding [22]. Furthermore, Six1 localizes to an area possessing the typical enhancer element architecture where a nucleosome poor domain is flanked on both sides by nucleosomes bearing the H3K4me1 mark. To determine if Six1 is responsible for establishing this enhancer structure, we first devised a strategy that involves ChIP assays on genome-integrated CER+PRR reporter genes. We stably transfected C2C12 myoblasts with our CER+PRR luciferase constructs, in their wild-type or double MEF3-site mutant versions, by co-transfecting either plasmid with limited amounts of a puromycin-resistance gene expression plasmid. Drug-resistant cells were pooled together to constitute either wild-type or double MEF3-site mutant CER "polyclones". Using quantitative PCR, we ensured that the wild-type and mutant polyclones contained equal numbers of transgenes (data not shown). Comparing ChIP assay results between the wild-type and mutant polyclones allowed us to ascertain the role of Six1 in establishing the CER enhancer architecture without having to use Six1 loss-of-function. This is important because we have shown that Six1 controls MyoD expression (Fig. 2) and because MyoD controls its own transcription [51]. In fact, MyoD can activate a transiently-transfected CER-luciferase reporter transgene in heterologous cells (Fig. 5B). We designed PCR primer pairs to be used after ChIP that would allow us to distinguish protein binding at the endogenous CER and at the CER transgene

separately (Fig. 5C). We first performed ChIP assays using antibodies against Six1, MyoD, H3K4me1 or an antibody recognizing all forms of histone H3, and performed quantitative PCR for the endogenous CER locus (primers a and b). This allowed us to confirm that the presence of these proteins or histone marks at the endogenous locus is not overtly different in the wild-type and the mutant CER polyclones (Fig. 5D). On the same ChIP samples, we then performed quantitative PCR using primers c and d, to detect specifically enrichment at the CER transgene. All proteins or marks are detected at the wild-type CER transgene as at the endogenous enhancer (Fig. 5E, inset). However, binding of Six1 to the double MEF3-site mutant transgene is greatly impaired, as expected (Fig. 5E). Moreover, MyoD recruitment to the MEF3-site mutant is also severely diminished. Strikingly, the presence of H3K4me1, and of nucleosomes in general (evidenced with the histone H3 antibody), is markedly increased when Six1 binding is prohibited.

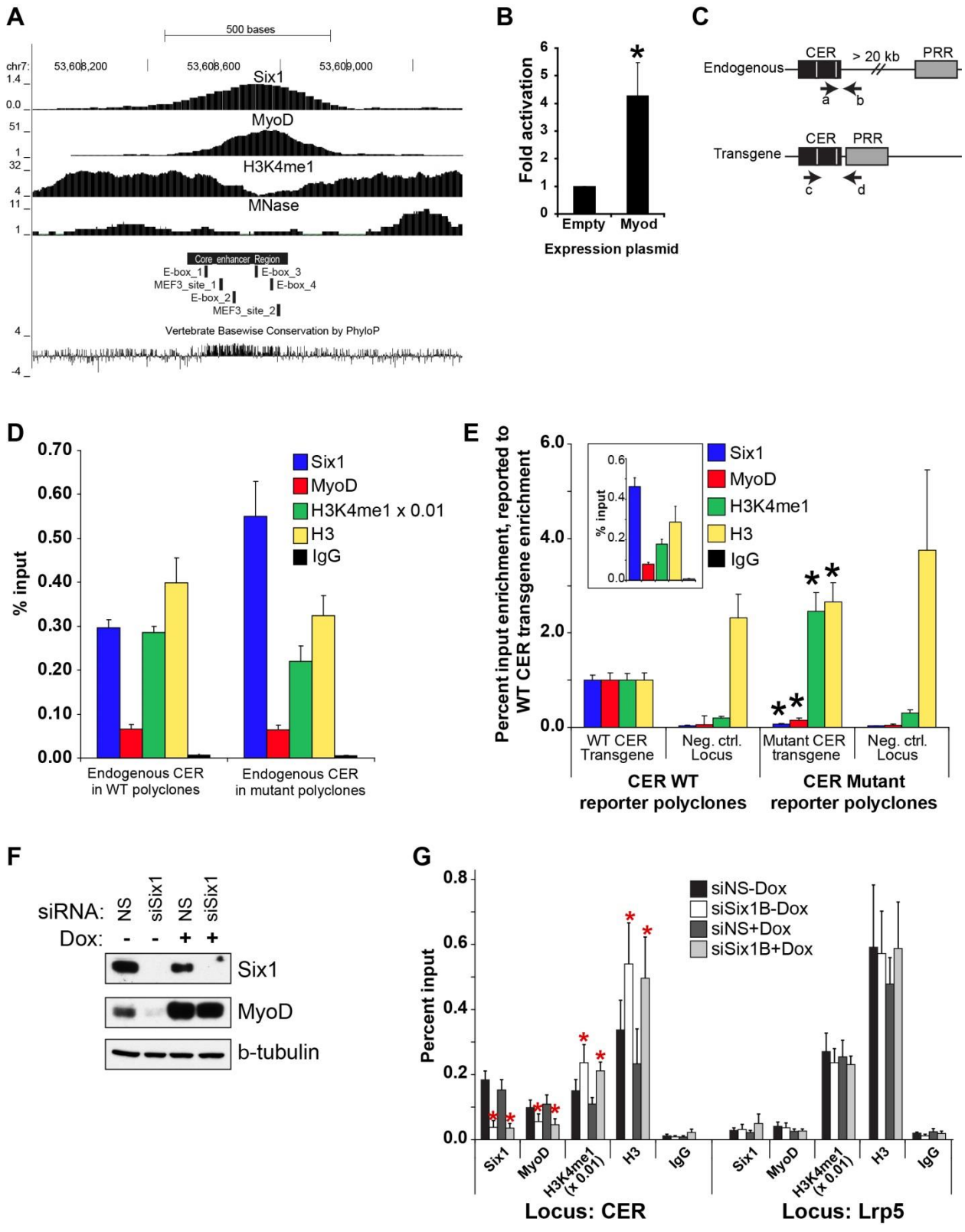

A
500 bases
chr7:
53,608,200
53,608,600
53,609,000
Six1
MyoD
H3K4me1
MNase
Core enhancer Region
E-box_1
E-box_3
MEF3_site_1
E-box_4
E-box_2
MEF3_site_2
Vertebrate Basewise Conservation by PhyloP
B
Fold activation
Empty
Myod
Expression plasmid
C
Endogenous
CER
> 20 kb
PRR
a
b
Transgene
CER
PRR
c
d
D
% input
Six1
MyoD
H3K4me1 x 0.01
H3
IgG
Endogenous CER in WT polyclones
Endogenous CER in mutant polyclones
E
Percent input enrichment, reported to WT CER transgene enrichment
% input
Six1
MyoD
H3K4me1
H3
IgG
WT CER Transgene
Neg. ctrl. Locus
Mutant CER transgene
Neg. ctrl. Locus
CER WT reporter polyclones
CER Mutant reporter polyclones
F
siRNA:
NS
siSix1
NS
siSix1
Dox:
-
-
+
+
Six1
MyoD
b-tubulin
G
Percent input
siNS-Dox
siSix1B-Dox
siNS+Dox
siSix1B+Dox
Six1
MyoD
H3K4me1 (x 0.01)
H3
IgG
Locus: CER
Locus: Lrp5

Figure 3. 5 Chromatin structure and MyoD binding at the CER depend on Six1 function.

A) Genomic binding profiles of Six1 in primary myoblasts (this study), and of MyoD [22], H3K4me1 and mononucleosomes [49] in C2C12 myoblasts. The position of the MEF3 sites and of three E-boxes (CANNTG) is also shown. **B)** Over-expression of MyoD in HEK293T cells leads to activation of the CER+PRR luciferase reporter gene. The results are reported over the luciferase activity obtained with an empty expression plasmid, and reflect the average of three independent replicates. Asterisk, $p < 0.05$ by one-tailed paired t test. **C)** Schematic representation of the PCR strategy used to distinguish the endogenous CER and the exogenous CER+PRR transgene. Primers a and b together can only amplify the endogenous CER sequences, while primers c and d will only give an amplification product on the CER+PRR reporter gene, since the endogenous CER and PRR are separated by more than 20 kb of sequence. The white bars in the CER represent the two Six1 binding sites. **D) and E)** ChIP assays performed on chromatin from stable polyclones of the CER+PRR reporters (wild-type or double MEF3 sites mutant). **D)** The real-time PCR quantities of the endogenous CER (primers a + b) were expressed as percentage of input chromatin. The level of binding of Six1, MyoD, H3K4me1 and H3 on the endogenous CER is not significantly different in wild type and mutant polyclones. Because the H3K4me1 signal is very high, we divided the values plotted for this mark by 100. **E)** The real-time PCR quantities of the transgene (primers a + c) were expressed as percentage of input chromatin and normalized over the quantities of the endogenous CER locus (primers a + b, used as internal control). Enrichment at a control locus (not targeted by Six1 or MyoD) is also given, in each set of polyclones. The results are reported as a fraction of the enrichment obtained on the wild-type CER+PRR construct. The inset shows the enrichment of Six1, MyoD, H3K4me1 and H3 as percentage of input on the wild type transgene. n=3 replicates

for each reporter gene construct; bars: S.E.M.. Asterisks, $p < 0.05$ by unpaired two-tailed t test. **F)** Western blot showing the levels of Six1 and MyoD in C2iFRT-FL-MyoD cells transfected with the control (siNS) or Six1 knock-down (siSix1) siRNA duplexes, and treated or not with doxycycline to induce MyoD expression. Beta-tubulin is shown as loading control. **G)** ChIP assays performed on chromatin isolated from C2iFRT-FL-MyoD cells treated as in F. The enrichment is shown as percent of input chromatin. Because the H3K4me1 signal is very high, we divided the values plotted for this mark by 100. The Lrp5 locus serves as a control locus not targeted by Six1 or MyoD. Asterisks indicate $p < 0.05$ by one-tailed paired T test.

Secondly, in order to confirm that loss of Six1 causes a remodeling of chromatin at the endogenous CER, we performed the knock-down of Six1 in a C2C12 cell line where the expression of MyoD can be induced by treatment with doxycycline (C2iFRT-FL-MyoD cells). Using this system, induction of exogenous MyoD expression allows us to maintain MyoD protein levels in the absence of Six1. As such, we can determine the impact of loss of Six1 without the complication of concomitant loss of MyoD protein. Western blot analysis confirmed that Six1 knock-down leads to lower MyoD expression levels, and revealed that treatment with doxycycline indeed rescues the MyoD expression defect (Fig. 5F). ChIP assays were then performed on similar samples using antibodies against Six1, MyoD, H3K4me1 and H3 (and normal IgG). As expected, Six1 binding decreases to background levels after its knock-down. Furthermore, we detected lower recruitment of MyoD at the CER after Six1 knock-down, which is consistent with the lower MyoD protein levels in this condition. Interestingly, as was the situation with the transgenic assays (Fig. 5E), the results showed that the global abundance of nucleosomes, and of nucleosomes bearing the H3K4me1 mark, increases when Six1 is knocked down (Fig. 5G). Importantly, the impact of Six1 knock-down was not due to the lower expression of MyoD since rescue by over-expression of Flag-MyoD failed to return the H3 or H3K4me1 signals to normal. Finally, our results confirmed that MyoD binding to the CER requires the presence of Six1, since induction of exogenous MyoD expression by the addition of doxycycline does not rescue its ability to bind to this enhancer. None of these effects were detected at the Lrp5 locus, where a robust H3K4me1 signal can be detected but where neither Six1 nor MyoD bind. These results combined indicate that Six1 binding is required for establishing or maintaining the appropriate structure of chromatin at the CER and for allowing MyoD to bind its enhancer.

DISCUSSION

We have shown that Six1 is expressed in satellite cells of adult muscle in regeneration, and that its expression and function are consistent with its role in regulating MyoD expression: the Six1 protein is detected in activated satellite cells, and its expression coincides with the presence of MyoD. Further, the expression of MyoD is attenuated in myoblasts where Six1 expression is knocked-down, suggesting that Six1 accomplishes a similar function in activated satellite cells. We have also demonstrated that Six1 exerts its function *in vitro* and *in vivo* through two MEF3 sites within the CER enhancer of MyoD, and that it acts at least in part by contributing to the specific chromatin architecture of the enhancer. We reason that the action of Six1 towards chromatin remodeling contributes to the action of additional transcription factors, such as MyoD itself, which we have shown can directly activate transcription from its CER.

Our findings suggest that Six1 exerts its effect on MyoD expression via the core enhancer region: the CER activity depends on Six1 and on its binding sites, in cultured cells and in regenerating muscle. We cannot rule out the involvement of other regulatory regions in the regulation of MyoD expression. However, our results show that the increased CER *in vivo* activity in regenerating muscle parallels the increased expression levels of MyoD after injury, so the CER appears a relevant enhancer to control MyoD upregulation in regenerating muscle. We and others have found that Six1 can also bind to the DRR enhancer [23], but we found that this binding is limited to primary myoblasts that are undergoing differentiation: binding of Six1 to the DRR in proliferating cells was undetectable. The situation is similar for MyoD, since it binds its DRR only in myotubes, not in myoblasts [22]. These observations regarding transcription factor binding at the DRR are in line with findings made with transgenic reporter mouse embryos, which have revealed that DRR-LacZ reporter genes are mostly active in differentiated

muscle cells [16,17]. Based on this, we conclude that of those two enhancers, the CER is the most relevant to the induction of MyoD expression in activated satellite cells.

A mechanistic clue as to how Six1 regulates MyoD expression came from examining the structure of chromatin at the CER. This enhancer, like many others, is characterized by a relative paucity of nucleosomes and by its flanking by nucleosomes bearing the H3K4me1 mark. In the absence of MEF3 sites, Six1 binding fails to occur and the structure of chromatin at the enhancer is altered: nucleosomes are more abundant, and they bear the H3K4me1 mark. This situation is reminiscent to that recently reported for Pax7 and Tpit target genes in the pituitary gland: the pioneering action of Pax7 is associated with a conversion of its target enhancers from a unimodal H3K4me1 distribution centered at its binding sites, to one that is bimodal, flanking the Pax7 binding sites on both sides and which allows Tpit binding [52]. We observed a similar situation in muscle precursor cells: our results reveal that in the absence of Six1, the MyoD protein is unable to bind its own enhancer. Based on our observations, we therefore propose that Six1 might analogously act as a pioneer factor that enables MyoD recruitment to the CER, by establishing a chromatin environment that enables MyoD to access DNA. The concept of pioneer factor-facilitated MyoD binding at target genes has a precedent: the homeodomain factor Pbx1 has been shown to constitutively bind the Myogenin promoter, and to facilitate the recruitment of MyoD at that locus upon differentiation [53,54]. Considering that Six1 is known to directly interact with components of the nucleosome-displacing SWI/SNF complex during inner ear neurogenesis, it is possible that Six1 might contribute to recruit an analogous complex at the CER in activated satellite cells. This would serve to open up the chromatin structure and allow MyoD to bind its own enhancer. According to this model, the effect of Six1 on MyoD induction can occur only in the presence of a certain amount of pre-existing MyoD protein. Assuming that

MyoD protein levels are absolutely null in quiescent satellite cells, the initial appearance of MyoD protein would be Six1-independent, and instead could rely on other mechanisms such as microRNA regulation as has been shown for the related gene Myf5 in satellite cells [9]. Once MyoD protein levels reach a certain threshold, a Six1- and MyoD-dependent boost of MyoD gene transcription would occur. Our observation that MyoD can indeed activate transcription from its CER enhancer, together with the well-established fact that MyoD regulates its own expression, is consistent with such a model. Another, non-mutually exclusive possibility is that Six1 function at the CER permits the binding of other transcription factors, in addition to MyoD.

Our model that Six1 functions at the CER by facilitating the recruitment of MyoD through the remodeling chromatin at this locus adds to the possible mechanisms by which chromatin regulation controls muscle cell differentiation [55]. We have previously reported that close to 40 percent of the loci bound by Six1 overlap to a highly significant degree with MyoD binding sites in C2C12 myoblasts [38], which suggests that other MyoD targets may also be regulated in a similar fashion by Six1. It will therefore be interesting to determine whether this mechanism for the combinatorial regulation of transcription of muscle genes is a general feature of MyoD-Six1 joint targets.

Yajima et al. have previously reported that Six1 is expressed in satellite cells: they found that Six1 was present in quiescent satellite cells of resting muscle, and in those of muscles in regeneration[56]. In addition, Le Grand et al. have detected the presence of the Six1 protein in seemingly quiescent satellite cells on explanted myofibers [23]. We failed to detect Six1 protein expression in quiescent satellite cells on frozen muscle sections, and we postulate that this could be due to technical differences (e.g. sensitivity of the method and/or specificity of the immunological reagents used). Based on previous reports, we also surmise that the isolation of

single myofibers may constitute a stress sufficiently strong to cause the activation of satellite cells [9,57]. Nevertheless, our main conclusions that Six1 protein expression increases following activation, and that the number of Six1-expressing cells increases as well following injury, are consistent with the previous reports. We found that the majority of MyoD-positive cells maintain Six1 expression during satellite cell activation (2, 3, 4 days after injury) before MyoD expression goes down with myofiber maturation (7 days after injury), suggesting that Six1 contributes to muscle regeneration by augmenting and maintaining MyoD expression in the satellite cells in their proliferative and early differentiation phases. This is consistent with the results of Le Grand, who showed that explanted myofibers of satellite cell-specific Six1 knock-outs express less MyoD-positive satellite cells [23]. Our molecular analyses suggest that Six1 favors this commitment of muscle stem cells by facilitating MyoD self-regulation, at least in part by enabling MyoD binding and participating in the remodeling of chromatin at the core enhancer region.

ACKNOWLEDGEMENTS

We thank Ilona Skerjanc, Hervé Faralli, and Nadine Wiper-Bergeron for helpful suggestions and comments. We also thank John Lunde, Aymeric Ravel-Chapuis, Adel Amirouche, Bernard Jean Jasmin for advice and technical help with muscle injury and electroporation, and François Marchildon for advice on immuno-staining.

REFERENCES

1. Peault B, Rudnicki M, Torrente Y, Cossu G, Tremblay JP, et al. (2007) Stem and progenitor cells in skeletal muscle development, maintenance, and therapy. Mol Ther 15: 867-77.

2. Zammit PS. (2008) All muscle satellite cells are equal, but are some more equal than others? J Cell Sci 121: 2975-82.

3. Lepper C, Partridge TA, Fan CM. (2011) An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. Development 138: 3639-3646.

4. McCarthy JJ, Mula J, Miyazaki M, Erfani R, Garrison K, et al. (2011) Effective fiber hypertrophy in satellite cell-depleted skeletal muscle. Development 138: 3657-3666.

5. Murphy MM, Lawson JA, Mathew SJ, Hutcheson DA, Kardon G. (2011) Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. Development 138: 3625-3637.

6. Sambasivan R, Yao R, Kissenpfennig A, Van Wittenberghe L, Paldi A, et al. (2011) Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. Development 138: 3647-3656.

7. Seale P, Sabourin LA, Girgis-Gabardo A, Mansouri A, Gruss P, et al. (2000) Pax7 is required for the specification of myogenic satellite cells. Cell 102: 777-86.

8. Gayraud-Morel B, Chretien F, Jory A, Sambasivan R, Negroni E, et al. (2012) Myf5 haploinsufficiency reveals distinct cell fate potentials for adult skeletal muscle stem cells. J Cell Sci 125: 1738-1749.

9. Crist CG, Montarras D, Buckingham M. (2012) Muscle satellite cells are primed for myogenesis but maintain quiescence with sequestration of Myf5 mRNA targeted by microRNA-31 in mRNP granules. Cell Stem Cell 11: 118-126.

10. Zammit PS, Heslop L, Hudon V, Rosenblatt JD, Tajbakhsh S, et al. (2002) Kinetics of myoblast proliferation show that resident satellite cells are competent to fully regenerate skeletal muscle fibers. Exp Cell Res 281: 39-49.

11. Grounds MD, Garrett KL, Lai MC, Wright WE, Beilharz MW. (1992) Identification of skeletal muscle precursor cells in vivo by use of MyoD1 and myogenin probes. Cell Tissue Res 267: 99-104.

12. Chen JC, Goldhamer DJ. (1999) Transcriptional mechanisms regulating MyoD expression in the mouse. Cell Tissue Res 296: 213-219.

13. Goldhamer DJ, Faerman A, Shani M, Emerson CP,Jr. (1992) Regulatory elements that control the lineage-specific expression of myoD. Science 256: 538-542.

14. Goldhamer DJ, Brunk BP, Faerman A, King A, Shani M, et al. (1995) Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. Development 121: 637-649.

15. Chen JC, Goldhamer DJ. (2004) The core enhancer is essential for proper timing of MyoD activation in limb buds and branchial arches. Dev Biol 265: 502-12.

16. Asakura A, Lyons GE, Tapscott SJ. (1995) The regulation of MyoD gene expression: Conserved elements mediate expression in embryonic axial muscle. Dev Biol 171: 386-398.

17. Chen JC, Love CM, Goldhamer DJ. (2001) Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. Dev Dyn 221: 274-88.

18. Chen JC, Ramachandran R, Goldhamer DJ. (2002) Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. Dev Biol 245: 213-223.

19. Tapscott SJ, Lassar AB, Weintraub H. (1992) A novel myoblast enhancer element mediates MyoD transcription. Mol Cell Biol 12: 4994-5003.

20. Andreucci JJ, Grant D, Cox DM, Tomc LK, Prywes R, et al. (2002) Composition and function of AP-1 transcription complexes during muscle cell differentiation. J Biol Chem 277: 16426-16432.

21. Kablar B, Krastel K, Ying C, Tapscott SJ, Goldhamer DJ, et al. (1999) Myogenic determination occurs independently in somites and limb buds. Dev Biol 206: 219-231.

22. Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, et al. (2010) Genome-wide MyoD binding in skeletal muscle cells: A potential for broad cellular reprogramming. Dev Cell 18: 662-74.

23. Le Grand F, Grifone R, Mourikis P, Houbron C, Gigaud C, et al. (2012) Six1 regulates stem cell repair potential and self-renewal during skeletal muscle regeneration. J Cell Biol 198: 815-832.

24. Woloshin P, Song K, Degnin C, Killary AM, Goldhamer DJ, et al. (1995) MSX1 inhibits myoD expression in fibroblast x 10T1/2 cell hybrids. Cell 82: 611-620.

25. Lee H, Habas R, Abate-Shen C. (2004) MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. Science 304: 1675-8.

26. Song K, Wang Y, Sassoon D. (1992) Expression of hox-7.1 in myoblasts inhibits terminal differentiation and induces cell transformation. Nature 360: 477-481.

27. Havis E, Coumailleau P, Bonnet A, Bismuth K, Bonnin MA, et al. (2012) Sim2 prevents entry into the myogenic program by repressing MyoD transcription during limb embryonic myogenesis. Development 139: 1910-1920.

28. L'honore A, Ouimette JF, Lavertu-Jolin M, Drouin J. (2010) Pitx2 defines alternate pathways acting through MyoD during limb and somitic myogenesis. Development 137: 3847-3856.

29. Zhang X, Patel SP, McCarthy JJ, Rabchevsky AG, Goldhamer DJ, et al. (2012) A non-canonical E-box within the MyoD core enhancer is necessary for circadian expression in skeletal muscle. Nucleic Acids Res 40: 3419-3430.

30. Andrews JL, Zhang X, McCarthy JJ, McDearmon EL, Hornberger TA, et al. (2010) CLOCK and BMAL1 regulate MyoD and are necessary for maintenance of skeletal muscle phenotype and function. Proc Natl Acad Sci U S A 107: 19090-19095.

31. Mokalled MH, Johnson AN, Creemers EE, Olson EN. (2012) MASTR directs MyoD-dependent satellite cell differentiation during skeletal muscle regeneration. Genes Dev 26: 190-202.

32. L'honore A, Lamb NJ, Vandromme M, Turowski P, Carnac G, et al. (2003) MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. Mol Biol Cell 14: 2151-2162.

33. L'honore A, Rana V, Arsic N, Franckhauser C, Lamb NJ, et al. (2007) Identification of a new hybrid serum response factor and myocyte enhancer factor 2-binding element in MyoD enhancer required for MyoD expression during myogenesis. Mol Biol Cell 18: 1992-2001.

34. Kumar JP. (2009) The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. Cell Mol Life Sci 66: 565-83.

35. Laclef C, Hamard G, Demignon J, Souil E, Houbron C, et al. (2003) Altered myogenesis in Six1-deficient mice. Development 130: 2239-52.

36. Grifone R, Demignon J, Houbron C, Souil E, Niro C, et al. (2005) Six1 and Six4 homeoproteins are required for Pax3 and mrf expression during myogenesis in the mouse embryo. Development 132: 2235-49.

37. Li X, Oghi KA, Zhang J, Krones A, Bush KT, et al. (2003) Eya protein phosphatase activity regulates Six1-dach-eya transcriptional effects in mammalian organogenesis. Nature 426: 247-54.

38. Liu Y, Chu A, Chakroun I, Islam U, Blais A. (2010) Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. Nucleic Acids Res 38: 6857-6871.

39. Sebastian S, Faralli H, Yao Z, Rakopoulos P, Palii C, et al. (2013) Tissue-specific splicing of a ubiquitously expressed transcription factor is essential for muscle differentiation. Genes Dev (In press) .

40. Rando TA, Blau HM. (1994) Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. J Cell Biol 125: 1275-1287.

41. Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, et al. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine. Proc Natl Acad Sci U S A 92: 7297-7301.

42. Langmead B, Trapnell C, Pop M, Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

43. Yee SP, Rigby PW. (1993) The regulation of myogenin gene expression during the embryonic development of the mouse. Genes Dev 7: 1277-89.

44. Nagy A. (2003) Manipulating the mouse embryo: A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 764 p.

45. Liu Y, Nandi S, Martel A, Antoun A, Ioshikhes I, et al. (2012) Discovery, optimization and validation of an optimal DNA-binding sequence for the Six1 homeodomain transcription factor. Nucleic Acids Res 40: 8227-8239.

46. Spitz F, Demignon J, Porteu A, Kahn A, Concordet JP, et al. (1998) Expression of myogenin during embryogenesis is controlled by six/sine oculis homeoproteins through a conserved MEF3 binding site. Proc Natl Acad Sci U S A 95: 14220-5.

47. Kucharczuk KL, Love CM, Dougherty NM, Goldhamer DJ. (1999) Fine-scale transgenic mapping of the MyoD core enhancer: MyoD is regulated by distinct but overlapping mechanisms in myotomal and non-myotomal muscle lineages. Development 126: 1957-65.

48. Ahmed M, Xu J, Xu PX. (2012) EYA1 and SIX1 drive the neuronal developmental program in cooperation with the SWI/SNF chromatin-remodeling complex and SOX2 in the mammalian inner ear. Development 139: 1965-1977.

49. Asp P, Blum R, Vethantham V, Parisi F, Micsinai M, et al. (2011) Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. Proc Natl Acad Sci U S A 108: E149-58.

50. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311-318.

51. Thayer MJ, Tapscott SJ, Davis RL, Wright WE, Lassar AB, et al. (1989) Positive autoregulation of the myogenic determination gene MyoD1. Cell 58: 241-8.

52. Budry L, Balsalobre A, Gauthier Y, Khetchoumian K, L'honore A, et al. (2012) The selector gene Pax7 dictates alternate pituitary cell fates through its pioneer action on chromatin remodeling. Genes Dev 26: 2299-2310.

53. Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, et al. (2004) Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell 14: 465-77.

54. de la Serna IL, Ohkawa Y, Berkes CA, Bergstrom DA, Dacwag CS, et al. (2005) MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. Mol Cell Biol 25: 3997-4009.

55. Dilworth FJ, Blais A. (2011) Epigenetic regulation of satellite cell activation during muscle regeneration. Stem Cell Res Ther 2: 18.

56. Yajima H, Motohashi N, Ono Y, Sato S, Ikeda K, et al. (2010) Six family genes control the proliferation and differentiation of muscle satellite cells. Exp Cell Res 316: 2932-2944.

57. Jones NC, Tyner KJ, Nibarger L, Stanley HM, Cornelison DD, et al. (2005) The p38alpha/beta MAPK functions as a molecular switch to activate the quiescent satellite cell. J Cell Biol 169: 105-16.

SUPPLEMENTARY INFORMATION

Table 3.S 1 Binding profile of Six1 in primary myoblasts at the MyoD locus

The data are in WIG format and can be uploaded as custom tracks on the UCSC genome browser, using the mouse mm9 genome release. The file contains the read density obtained with anti-Six1 and the read density obtained using normal rabbit IgG as negative control.

Uploaded on http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0067762#s5

CHAPTER 4. SIX1 INTERACTS WITH CDK9 TO REGULATE MYOD EXPRESSION IN MYOBLASTS

Liu, Y., Chakroun, I., Bergamin, E., Yang, D., and Blais, A.

I analyzed the genomic binding properties of Six1 in primary myoblasts, which was generated in our lab (Figure 4.1A, B and C). I performed bioinformatics analyses of Six1 ChIP-seq (from our lab) and Myod ChIP-seq (from the Tapscott lab) data for Figure 4.1D and 4.1E. I performed FAIRE-seq experiments and the following bioinformatics analyses of FAIRE-seq data with other published genomic data (Figures 4.2 and 4.3). I also performed bioinformatics analyses of FAIRE-seq data and the validation FAIRE-qPCR (Figure 4.4). In addition, I performed co-immunoprecipitation (co-IP) experiments for the mass spectrometry identification and the validation co-IP of Cdk9 and Six1 (Figure 4.5). I also performed FAIRE experiments and western blots after Flavopiridol treatment (Figure 4.6A and 4.6B). I carried out the ChIP assays for Figure 4.6C and 4.6D. Lastly, I contributed to the writing and the editing of the manuscript.

Six1 interacts with Cdk9 to regulate Myod expression in myoblasts

Yubing Liu[1,2], Imane Chakroun[1,2], Elisa Bergamin[1,2], Dabo Yang[1,2] and Alexandre Blais[1,2].

1, Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada

2, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada

ABSTRACT

Six1 is essential for embryonic muscle development and for the homeostasis of muscle stem cells, satellite cells, during muscle regeneration in adulthood. Myod is a master regulator in the myogenic lineage whose expression is absent in quiescent satellite cells but quickly induced upon satellite cell activation. Previously we and others have reported that the expression of Myod is under the control of Six1 during muscle regeneration. Six1 directly binds to the core enhancer region (CER) of Myod and the binding of Six1 is required to recruit Myod and to maintain proper chromatin structure at the CER. Here we report that Six1 has an impact on the global chromatin architecture. The co-binding of Six1 and Myod at distal sites correlates with functional enhancers, as exemplified by the deposition of H3K4me1 and H3K27ac and the enriched binding of RNA polymerase II, p300 and c-Jun on these distal loci. Furthermore, we demonstrate that Six1 and Cdk9 form a protein complex and they are functionally related in regulating Myod expression. Together, we manifest that Six1 recruits Cdk9 to the CER, promoting the expression of Myod in myoblasts.

INTRODUCTION

Satellite cells are muscle stem cells that reside on the myofiber between sarcolemma and basal lamina (reviewed in (Chang and Rudnicki, 2014)). Satellite cells are indispensable for postnatal muscle growth and adult skeletal muscle regeneration (Lepper et al., 2009; Murphy et al., 2011; Sambasivan et al., 2011b). Upon activation by injury or exercise, satellite cells can undergo cell division either to differentiate and repair damaged myofibers or to self-renew to replenish the satellite cell reservoir (Kuang et al., 2007).

The process of muscle regeneration is under the regulation of a combinatorial network of transcription factors (Blais, 2015). Quiescent and proliferating satellite cells express the Paired-box transcription factor Pax7 (Seale et al., 2000). Myod and Myf5, two determinative members of the Myogenic regulatory factors (MRFs), are expressed in committed myoblasts (Gunther et al., 2013; Wood et al., 2013). The expression of Myogenin is induced in differentiating myocytes (Faralli et al., 2012; Liu et al., 2013c). Although it is controversial whether the Myod gene is transcribed in quiescent satellite cells (Cornelison and Wold, 1997; Fukada et al., 2007; Liu et al., 2013a; Pallafacchina et al., 2010), it is well accepted that the Myod protein is not detectable in quiescent satellite cells but is induced quickly upon satellite cell activation (Liu et al., 2013a; Zammit et al., 2004). A large proportion of daughter cells, expressing Myod, proliferate and then undergo differentiation to generate new muscle mass, while a minor proportion, expressing Pax7 but not Myod, return to quiescent state as self-renewal (Olguin and Olwin, 2004; Yoshida et al., 1998; Zammit et al., 2004).

During muscle regeneration, the expression of Myod is tightly controlled to ensure proper balance between satellite cell renewal and muscle differentiation. Three regulatory elements have

been identified upstream of the Myod gene, the proximal regulatory region (PRR), the distal regulatory region (DRR) and the core enhancer region (CER) (Asakura et al., 1995; Goldhamer et al., 1995). The PRR, which is adjacent to the transcription start site (TSS), bears only weak activation on a reporter gene (Goldhamer et al., 1992; Tapscott et al., 1992). The CER is a 258 base pairs (bp) element that is 24 kilobases (kb) upstream of the gene and the DRR is a 700 bp element that is 5 kb upstream (Asakura et al., 1995; Goldhamer et al., 1995). Both the CER and the DRR are sufficient to direct the expression of the PRR-driven lacZ reporter during embryonic myogenesis, which recapitulates the temporal expression of Myod (Asakura et al., 1995; Goldhamer et al., 1995). Studies using deletion mutagenesis in the chromosomal context showed that: 1) the DRR is dispensable for Myod expression in myotomes but is required to maintain normal Myod expression in adult muscles (Asakura et al., 1995; Chen et al., 2002); and 2) the CER is required for the proper temporal and spatial activation of Myod in both the myotome and limb buds (Chen et al., 2001).

Several transcription factors have been shown to directly regulate Myod expression. Pax3, Pax7 and Foxo3 can bind to the PRR and coordinately recruit RNA polymerase II (RNA pol II) to activate Myod expression in C2C12 myoblasts and satellite cells (Hu et al., 2008). Fra-2, a subunit of the AP-1 transcription factor, activates a reporter driven by the Myod promoter and loss of function of Fra-2 leads to reduced expression of Myod in C2C12 myoblasts (Alli et al., 2013; Andreucci et al., 2002). SRF and YY1 complexes compete with Mef2 transcription factors to bind the DRR and activate the expression of the DRR-driven reporter in C2 myoblasts (L'honore et al., 2003, 2007). The Myocardin family of transcriptional co-activators, Mastr and Mrtf-a, bind to the DRR through associations with Mef2c in C2C12 myoblasts and regulate Myod expression *in vivo* (Creemers et al., 2006; Mokalled et al., 2012). The CER is bound by

Bmal1 (brain and muscle ARNT-like 1) and Clock (circadian locomotor output cycles kaput) *in vivo* and a non-canonical E-box in the CER is necessary for the full activation and circadian oscillation of the CER reporter gene transcription in C2C12 myotubes (Andrews et al., 2010; Zhang et al., 2012). When it is artificially expressed, Msx1 binds to the CER in C2C12 myoblasts and represses Myod expression through favouring the recruitment of the linker histone H1b and the euchromatic histone-lysine N-methyltransferase G9a (Lee et al., 2004b; Wang and Abate-Shen, 2012). Pitx2 binds to the CER in embryonic limbs and mediates the early activation of Myod in developing embryos (L'honoré et al., 2010). Myod itself positively regulates its own expression (Thayer et al., 1989), possibly through direct binding to its regulatory elements, such as the PRR in G8 myoblasts (Zingg, Pedraza-Alva, & Jost, 1994), the DRR in primary myotubes (Le Grand et al., 2012) and the CER in C2C12 myoblasts and myotubes (Liu et al., 2013).

Six1, a homologue of *sine oculis* in *Drosophila*, is a member of the Six family transcription factors, which consist of six members, from Six1 to Six6 (Kawakami et al., 2000). Six1 is essential for embryonic myogenesis, along with its involvement in the development of other tissues (Grifone et al., 2005; Laclef et al., 2003a, 2003b; Xu et al., 2003). In addition, Six1 expression is induced during adult muscle regeneration, and Six1 gene ablation in adult satellite cells results in impaired muscle regeneration and disrupted satellite cell homeostasis (Le Grand et al., 2012). Six1 is shown to regulate Myod expression in developing limb buds and during muscle regeneration (Laclef et al., 2003a; Le Grand et al., 2012).

Enhancers typically contain a cluster of DNA elements that are bound by transcription factors, and consequently facilitate the expression of target genes in a cell-type and time-specific manner (García-González et al., 2016). In the context of myogenesis, several features have been shown to accumulate at active enhancers, including H3K4me1 (mono-methylation of histone 3 lysine 4)

and H3K27ac (acetylation of histone 3 lysine 27), binding of p300, c-Jun and RNA pol II (Blum et al., 2012; Heintzman et al., 2009; Lam et al., 2014; Visel et al., 2009).

Here we report that co-binding of Six1 and Myod correlate with more accessible chromatin and enrichment of active enhancer marks at distal regions. Six1 is required to maintain chromatin accessible at a subset of Myod-bound distal enhancers in myoblasts. Using functional proteomics studies, Cdk9 is identified as a Six1-interacting protein. Cdk9, a cyclin-dependent kinase, and its regulatory cyclin subunit (Cyclin T) form the positive transcription elongation factor b (p-TEFb). The kinase activity of Cdk9 is required to release transcription pause and to promote transcription elongation at proximal regions (Fu et al., 1999; Peng et al., 1998; Pirngruber et al., 2009), and has also been shown to act at enhancers, to facilitate the production of enhancer RNA (eRNA) (Hah et al., 2013). eRNA transcripts and/or the action of active transcription have been shown to be required for the induction of target genes, in *cis* or in *trans* (Arner et al., 2015; Kaikkonen et al., 2013; Kim et al., 2015a; Melo et al., 2013; Mousavi et al., 2013). The interaction between Cdk9 and Six1 has been validated in 293 cells. Pharmaceutical inhibition of the Cdk9 kinase activity results in decreased expression of Myod, probably through affecting chromatin structure on the CER. CDK9 is able to bind to the Myod CER and its binding depends on the presence of Six1 binding. Together, we described that Six1 binding is correlated with high frequency of accessible chromatin at Myod-distal targets, which is mediated through recruitment of Cdk9.

MATERIALS AND METHODS

Cell culture, stable cell lines and RNA interference

C2C12 myoblasts were acquired from the American Type Culture Collection and cultured in DMEM supplemented with 10% fetal bovine serum, glutamine and antibiotics. Cells were incubated at 37 °C with 5% CO_2. C2C12 stable cell lines bearing the wild type and mutant CER sequences were established and cultured as reported previously (Liu et al., 2013c). For RNA interference, C2C12 cells were seeded at 50 000 cells in 6 well plates (Corning) and were transfected with Stealth siRNA (Invitrogen) using RNAiMAX (Invitrogen), following the recommended protocol (Invitrogen). siRNA oligo sequences are listed in supplementary Table 1.

Antibodies

Six1 specific antibody was produced as reported previously (Liu et al., 2010). Normal rabbit IgG was used as a negative control for immunoprecipitation experiments (Jackson ImmunoResearch). Cdk9 antibody was a gift from Dr. Ali Shilatifard (Northwestern University). Both Six1 and Cdk9 antibodies were raised in rabbits. Other antibodies used in western blot are anti-Myod (5.8A, Santa Cruz, sc-32758), anti-beta-tubulin (clone E7, DSHB) and anti-Flag (F1804, Sigma). Secondary antibodies against mouse or rabbit IgG, coupled to horseradish peroxidase (Promega).

Chromatin immunoprecipitation (ChIP) and ChIP-seq assay

ChIP was performed on C2C12 myoblasts as described previously (Liu et al., 2010). ChIP-seq was performed on primary myoblasts as described previously (Chakroun et al., 2015; Liu et al., 2010). Briefly, C2C12 myoblasts and primary myoblasts were fixed with 1% Formaldehyde for 30 mins at room temperature; quenched by 125 mM glycine for 5 mins at room temperature. Chromatin was sonicated with sonifier (Branson) to achieve an average of 150-200 base pairs fragments. Six1 antibody was used to pull down Six1-associated chromatin (Liu et al., 2010).

DNA was purified with QiaQuick PCR purification columns (Qiagen) and was analyzed with BioAnalyzer (Agilent). DNA with a size of 100 to 200 bp was selected at the gel selection step and sequenced with Génome Québec on the Sanger / Illumina 1.9 platform. The sequence length was 50 bp. Unique reads were mapped to the mouse genome mm9 using Bowtie 2 (Langmead and Salzberg, 2012). Myod and Six1-bound peaks were called using MACS v2.0 (Feng et al., 2011). Peaks bound by Six1 were annotated using Pavis (Huang et al., 2013). Gene annotation was performed by GREAT (McLean et al., 2010). Read density was plotted using Easeq (Lerdrup et al., 2016).

Co-immunoprecipitation and mass spectrometry

Six1 specific antibody was used to precipitate Six1-interacting proteins in the nuclear extract of C2C12 myoblasts (Liu et al., 2010). Nuclear extraction was performed as previously (Palii et al., 2011). Six1 and Six1-interacting proteins were separated with SDS-polyacrylamide gradient gel (Bio-Rad, 4% - 20%) and were stained with blue silver staining method (Candiano et al., 2004). Enriched gel bands were cut and analyzed by LC-MS/MS (liquid chromatography – tandem mass spectrometry). Target proteins were identified with MASCOT software.

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) assay, quantitative PCR and FAIRE-seq

FAIRE assay was performed as described in (Simon et al., 2012). Briefly, C2C12 myoblasts were cross-linked with 1% Formaldehyde for 10 mins and then quenched by 125 mM Glycine for 5 mins at room temperature. Chromatin was sonicated with sonifier (Branson) to achieve an average of 200-300 base pairs fragments. DNA was purified with QiaQuick PCR purification columns (Qiagen) and was analyzed by quantitative PCR (Agilent) or deep DNA sequencing.

DNA with a size of 150 to 250 bp was sequenced with Génome Québec on the Sanger / Illumina 1.9 platform. The sequence length was 50 bp. Unique reads were mapped to the mouse genome mm9 using Bowtie 2 (Langmead and Salzberg, 2012). Read density was plotted using Easeq (Lerdrup et al., 2016).

Plasmids and Protein pull-down assay

Coding sequences of Cdk9, Eya3 and GFP were cloned to pCMV vectors which bear the coding sequence of the VAP tag. The VAP tag contains three tandem Flag tags, two TEV cleavage sites, the Twin-Strep tag, and six tandem His tags, as described in (Ni et al., 2011). VAP-tagged CDK9, EYA3 or GFP were co-expressed with Myc-SIX1 in 293T cells. StrepTactin Sepharose resin was used to pull down Strep-tagged proteins and their interacting proteins (Ni et al., 2011).

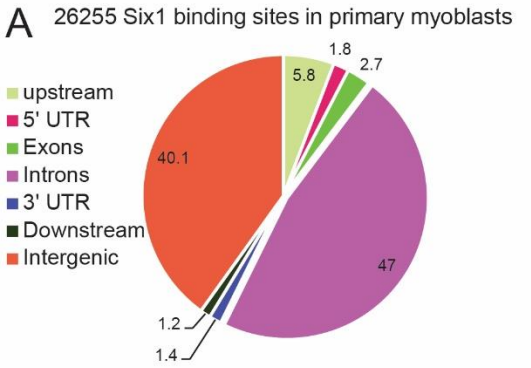
RESULTS

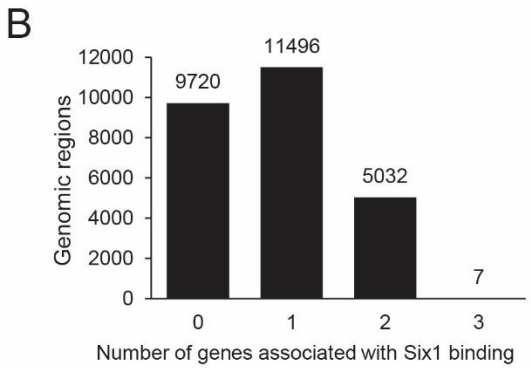
Six1 and Myod shared targets are associated with more accessible chromatin

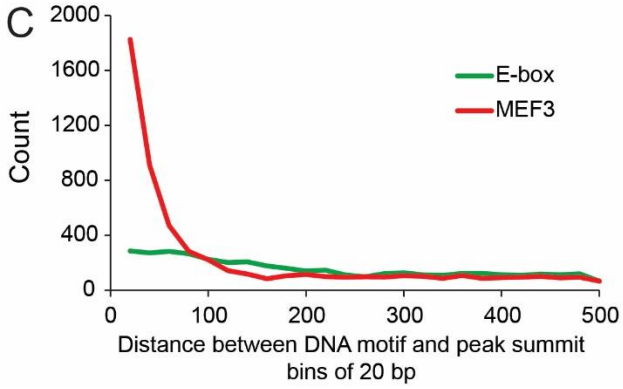
In order to characterize the role of Six1 in adult muscle precursors, we started by identifying its genomic binding sites. We had previously performed ChIP-on-chip in C2C12 myoblasts (Liu et al., 2010), and here we used a similar strategy employing ChIP-seq in primary myoblasts. This led to the identification of 26255 Six1-binding sites, which is much more than 1022 Six1-binding peaks discovered in ChIP-on-Chip experiment (Fig. S1A). Comparison of two methods found that the vast majority of ChIP-on-Chip peaks are also identified by ChIP-seq and ChIP-on-Chip locates the highest confidence ChIP-seq peaks (Fig. S1B). Of 26255 Six1 peaks, 47% and 40% fell into introns and intergenic regions, respectively (Fig. 1A). 63% of binding sites were associated with one or more genes (Fig. 1B). A binding site is assigned to a gene following these rules: 1) within 5 kb upstream of a TSS; 2) within 1 kb downstream of a TSS and 3) if no gene is in this range, within 50 kb from a TSS. Six1 recognizes the MEF3 DNA motif (Liu et al., 2012). 5682 MEF3 sites were identified in 4809 Six1-binding peaks in primary myoblasts and it was found frequently in the proximity of the peak summit (Fig. 1C). The MEF3 motif was enriched at Six1-bound loci with the highest peak summits (Fig. S2).

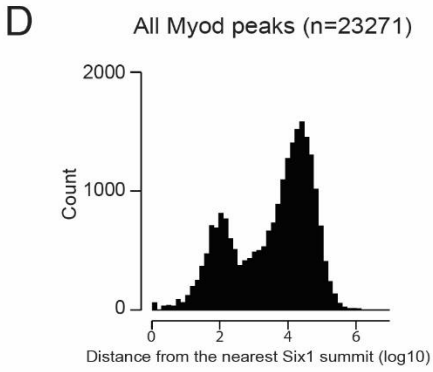
Because synergy between Six factors and the MRFs has previously been reported, we evaluated the extent of co-occupancy by MyoD and Six1 (Chakroun et al., 2015; Liu et al., 2010; Santolini et al., 2016). Myod is able to occupy a large set of genomic regions in C2C12 myoblasts (Cao et al., 2010). The summits of Myod and Six1 peaks were collected and extended to 1 kb as the core binding sites. The distance between Myod and Six1 peak summits was examined. We found that 7356 Myod peaks coincided with Six1 binding sites within 1000 bp, termed as shared Myod+Six1 sites, which accounted for 31.6% of 23271 Myod binding sites and 28% of all Six1

binding sites (Fig. 1D). The other Myod or Six1 peaks were named unique Myod or unique Six1 sites. Out of 16857 distal Myod sites (a peak summit more than 2.5 kb away from any known TSS), 5803 (34%) were co-occupied by Six1 as shared Myod+Six1 distal sites (Fig. 1E). 1553 (24%) of 6414 proximal Myod sites (a peak summit within 2.5 kb of a gene TSS) as shared Myod+Six1 proximal sites (Fig. 1E).


A
26255 Six1 binding sites in primary myoblasts
upstream
5' UTR
Exons
Introns
3' UTR
Downstream
Intergenic
5.8
1.8
2.7
40.1
47
1.2
1.4


B
Genomic regions
12000
10000
8000
6000
4000
2000
0
9720
11496
5032
7
0
1
2
3
Number of genes associated with Six1 binding


C
Count
2000
1600
1200
800
400
0
E-box
MEF3
0
100
200
300
400
500
Distance between DNA motif and peak summit
bins of 20 bp


D
All Myod peaks (n=23271)
Count
2000
1000
0
0
2
4
6
Distance from the nearest Six1 summit (log10)

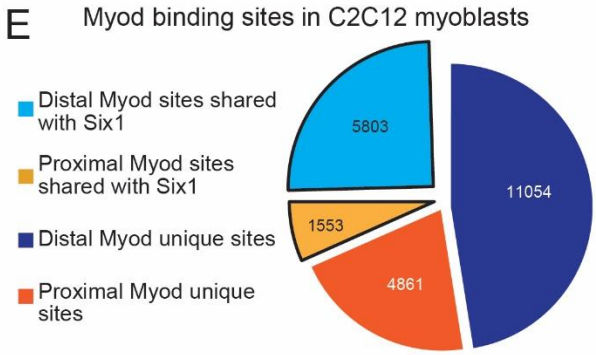

E
Myod binding sites in C2C12 myoblasts
Distal Myod sites shared with Six1
Proximal Myod sites shared with Six1
Distal Myod unique sites
Proximal Myod unique sites
5803
11054
1553
4861

Figure 4. 1 Six1 and Myod co-occupy a large set of genomic loci

A) Characterization of Six1 binding peaks. Upstream: less than 2.5 kb from a TSS. 5′ UTR: 5′-untranslated region. 3′ UTR: 3′-untranslated region. Downstream: less than 1 kb from a TES (transcription end site). Intergenic: genomic loci that are more than 2.5 kb from a TSS and more than 1 kb from a TES. B) Genome annotation of Six1 binding peaks. C) Distribution of DNA motifs around the summits of Six1 binding peaks. E-box is recognized by Myod and MEF3 is recognized by Six1. D) Distribution of the distance between Myod peak summits and the nearest Six1 peak summits. E) Myod binding peaks were grouped based on their genomic locations and their distance from Six1. Distal Myod sites: the distance between Myod peak summits and the nearest TSSs was more than 2.5 kb. Proximal Myod sites: the distance between Myod peak summits and the nearest TSSs was less than 2.5 kb. Myod sites shared with Six1: Myod peaks with Six1 binding in 1 kb (summit-summit distance). Myod unique sites: Myod peaks without Six1 binding in 1 kb (summit-summit distance).

A central, nucleosome-depleted region is one of the characteristics of active promoters and enhancers (Lee et al., 2004a). Six1 was shown to be implicated in the chromatin structure at the Myod CER (Liu et al., 2013c). Without the presence of Six1, more histone H3 was accumulated at the exogenous CER (Liu et al., 2013c). In order to explore the potential role of Six1 in the chromatin structure, FAIRE-seq was performed on C2C12 myoblasts. The average FAIRE signal was comparable between Six1 and Myod target loci (Fig. 2A). The co-binding of Myod and Six1 was associated with substantially more accessible chromatin (Fig. 2B), suggesting that both proteins may participate in establishing an open chromatin structure. Myod binding is superimposed with the locations of muscle enhancers (Blum and Dynlacht, 2013; Blum et al., 2012). We found that chromatin at shared Myod+Six1 distal sites was more accessible than that at unique Myod distal sites (Fig. 2C), while this phenomenon was not as prominent at shared Myod+Six1 proximal sites (Fig. 2D). Micrococcal nuclease (MNase) cleaves exposed DNA until it encounters an obstacle, such as a nucleosome. As an alternative way to study the chromatin structure, MNase-seq showed results that were concordant with those of the FAIRE-seq analysis (Fig. 2E, 2F) (Asp et al., 2011). Taken together, these data suggest that Six1 and Myod may cooperatively regulate the expression of target genes through establishing accessible chromatin at distal sites, the potential enhancers.

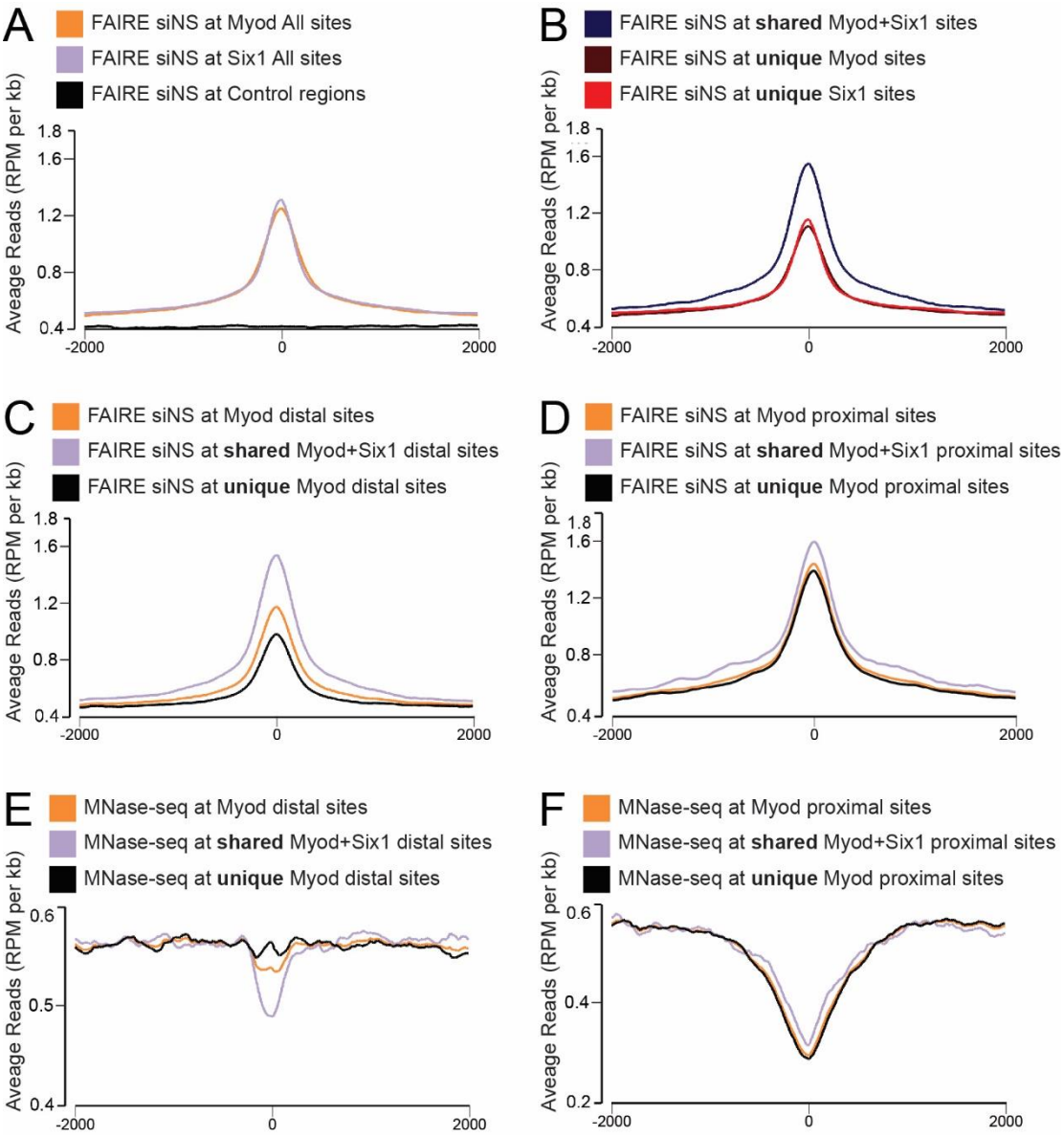

A
FAIRE siNS at Myod All sites
FAIRE siNS at Six1 All sites
FAIRE siNS at Control regions
Aveage Reads (RPM per kb)
1.8
1.6
1.2
0.8
0.4
-2000
0
2000
B
FAIRE siNS at shared Myod+Six1 sites
FAIRE siNS at unique Myod sites
FAIRE siNS at unique Six1 sites
Aveage Reads (RPM per kb)
1.8
1.6
1.2
0.8
0.4
-2000
0
2000
C
FAIRE siNS at Myod distal sites
FAIRE siNS at shared Myod+Six1 distal sites
FAIRE siNS at unique Myod distal sites
Aveage Reads (RPM per kb)
1.8
1.6
1.2
0.8
0.4
-2000
0
2000
D
FAIRE siNS at Myod proximal sites
FAIRE siNS at shared Myod+Six1 proximal sites
FAIRE siNS at unique Myod proximal sites
Aveage Reads (RPM per kb)
1.8
1.6
1.2
0.8
0.4
-2000
0
2000
E
MNase-seq at Myod distal sites
MNase-seq at shared Myod+Six1 distal sites
MNase-seq at unique Myod distal sites
Aveage Reads (RPM per kb)
0.6
0.5
0.4
-2000
0
2000
F
MNase-seq at Myod proximal sites
MNase-seq at shared Myod+Six1 proximal sites
MNase-seq at unique Myod proximal sites
Aveage Reads (RPM per kb)
0.6
0.4
0.2
-2000
0
2000

Figure 4. 2 Co-binding of Six1 on Myod distal targets correlates with more accessible chromatin

Average FAIRE-seq or MNase-seq signals were represented for a 4000 bp window centered at the binding peak summits of each group of ChIP-seq targets. Y-axis is the read density normalized to reads per million (RPM) per 1000 bp. A) FAIRE-seq signals at all Myod peaks or all Six1 peaks. Control regions were 5 kb upstream of the 10 000 most expressed genes in C2C12 myoblasts that were not bound by Six1 or Myod, showing -7 kb to -3 kb of TSSs of 3110 genes. B) FAIRE-seq signals at shared Myod+Six1 sites, or sites that were unique to each transcription factor. C) FAIRE-seq signals at Myod distal sites with (shared) or without (unique) Six1 binding. D) FAIRE-seq signals at Myod proximal sites with (shared) or without (unique) Six1 binding. E) MNase-seq signals at Myod distal sites with (shared) or without (unique) Six1 binding. F) MNase-seq signals at Myod proximal sites with (shared) or without (unique) Six1 binding.

Myod distal targets shared by Six1 correlate with marks of active enhancers

Active enhancers display accumulation of H3K4me1 and H3K27ac (Heintzman et al., 2007; Visel et al., 2009). Other features of enhancers are the binding of these loci by p300 co-activator, the AP-1 member c-Jun, and RNA polymerase II (Asp et al., 2011; Blum et al., 2012). Compared to unique Myod distal regions, shared Myod+Six1 distal regions depicted a more pronounced bimodal distribution of H3K4me1 (Fig. 3A) and H3K27ac (Fig. 3B), along with more pronounced recruitment of RNA polymerase II (Fig. 3C), p300 (Fig. 3D) and c-Jun (Fig. 3E). The co-binding of Six1 and Myod on the distal sites also correlated with more production of newly transcribed RNA as detected by GRO-seq (global run on transcription followed by deep sequencing) (Fig. 3F) (Ji et al., 2011). Taken together, these observations indicate that the presence of Six1 is associated to several marks of active enhancers, suggesting that it may participate in establishing or maintaining this type of chromatin structure.

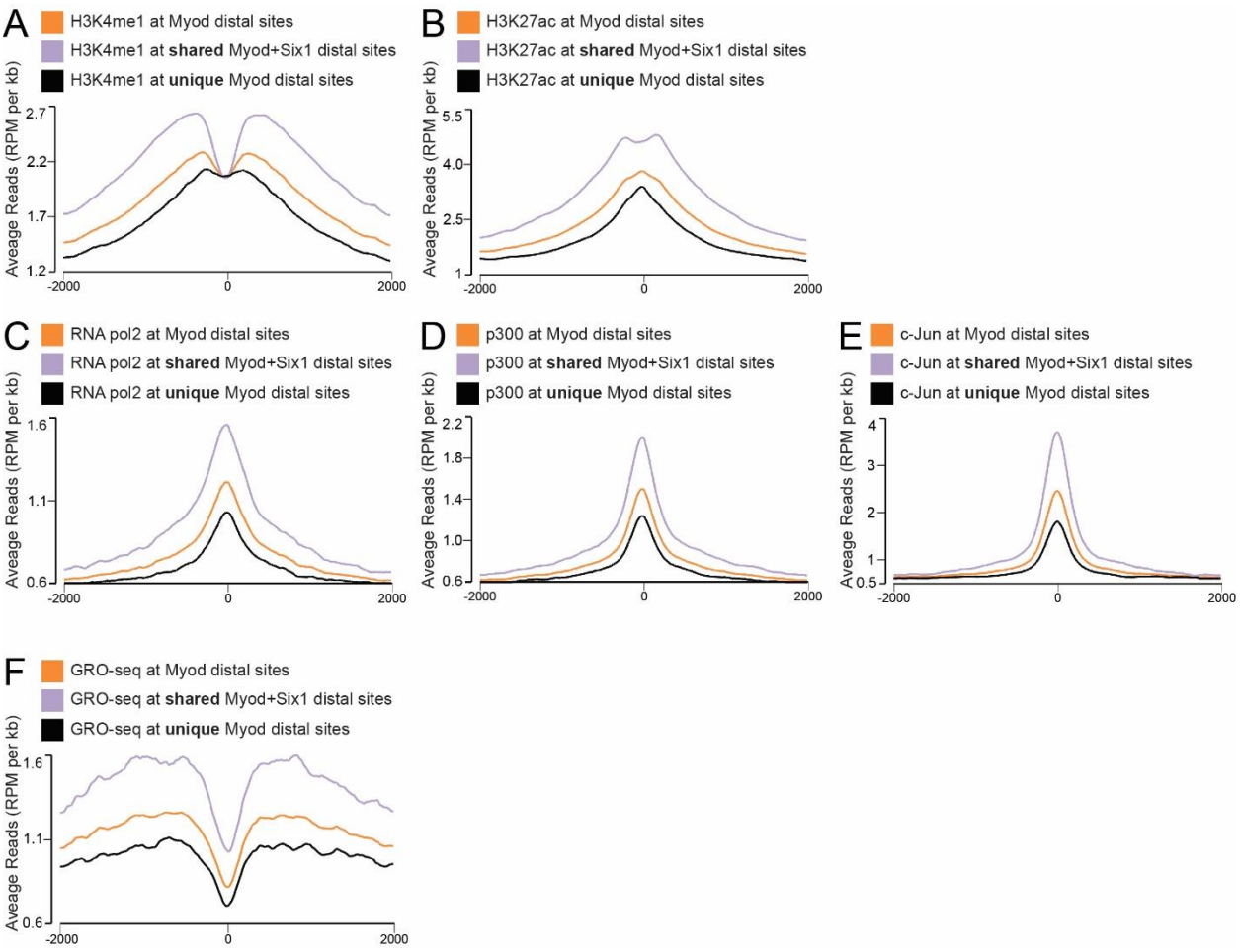

A
H3K4me1 at Myod distal sites
H3K4me1 at shared Myod+Six1 distal sites
H3K4me1 at unique Myod distal sites
Average Reads (RPM per kb)
2.7
2.2
1.7
1.2
-2000
0
2000
B
H3K27ac at Myod distal sites
H3K27ac at shared Myod+Six1 distal sites
H3K27ac at unique Myod distal sites
Average Reads (RPM per kb)
5.5
4.0
2.5
1
-2000
0
2000
C
RNA pol2 at Myod distal sites
RNA pol2 at shared Myod+Six1 distal sites
RNA pol2 at unique Myod distal sites
Average Reads (RPM per kb)
1.6
1.1
0.6
-2000
0
2000
D
p300 at Myod distal sites
p300 at shared Myod+Six1 distal sites
p300 at unique Myod distal sites
Average Reads (RPM per kb)
2.2
1.8
1.4
1.0
0.6
-2000
0
2000
E
c-Jun at Myod distal sites
c-Jun at shared Myod+Six1 distal sites
c-Jun at unique Myod distal sites
Average Reads (RPM per kb)
4
3
2
1
0.5
-2000
0
2000
F
GRO-seq at Myod distal sites
GRO-seq at shared Myod+Six1 distal sites
GRO-seq at unique Myod distal sites
Average Reads (RPM per kb)
1.6
1.1
0.6
-2000
0
2000

Figure 4. 3 Co-binding of Six1 at Myod distal sites is associated with marks of active enhancers

ChIP-seq or GRO-seq signals were represented for a 4000 bp window centered at the binding peak summits of each group of ChIP-seq targets. Y-axis is the read density normalized to reads per million (RPM) per 1000bp. A) H3K4me1 ChIP-seq, B) H3K27ac ChIP-seq, C) RNA pol II ChIP-seq, D) p300 ChIP-seq, E) c-Jun ChIP-seq and F) GRO-seq were compared at Myod distal sites with (shared) or without (unique) Six1 binding.

Six1 is necessary to maintain an open chromatin configuration at enhancers

Myod has been shown to recruit several enhancer-associated transcription factors to modulate the assembly of active muscle enhancers (Blum et al., 2012). We previously reported that the recruitment of Myod on the Myod CER depended on the presence of Six1. In order to distinguish the role of Myod and Six1 in the chromatin structure, we combined FAIRE-seq with small interfering RNA (siRNA) to evaluate the global change of the chromatin accessibility upon Myod or Six1 knockdown (KD). The chromatin accessibility at shared Myod+Six1 proximal sites was only slightly affected by Six1 KD or Myod KD (Fig. 4A). However, the chromatin structure at shared Myod+Six1 distal sites was substantially less "open" after Six1 KD (Fig. 4B). To better evaluate the impact of Six1 on the chromatin structure, 5803 shared Myod+Six1 distal sites were clustered into 12 groups based on FAIRE signals of siNS, siSix1 and siMyod samples (Fig. 4C). Loss of function of Six1 led to reduced chromatin accessibility at several subgroups (Cluster 1, 10, and 11), totaling 1438 enhancers (Fig. 4C), suggesting that Six1 is indispensable for the proper chromatin structure at those subgroups. By comparison, only cluster 0 (total of 498 enhancers) showed a decrease in FAIRE signals following Myod knock-down.

To validate the genomic observation, FAIRE was performed on C2C12 myoblasts, and analyzed by quantitative PCR. Chromatin on the Myod CER was less accessible after Six1 KD (Fig. 4D, Fig. S4A). This effect was not mediated by subsequent loss of Myod after Six1 KD, in that Myod KD did not induce the same effect on this locus (Fig. 4D). FAIRE assay was also performed on C2C12 stable cell lines carrying either the wild type CER (WT) or the mutant CER (Mut) that was defective for Six1 binding (Liu et al., 2013c). The mutant plasmid CER was less open in the absence of Six1 binding (Fig. 4E, Fig. S4B). Taken together, these results suggest that Six1 expression and DNA binding are required to maintain the open chromatin structure at a subset of

shared Myod+Six1 distal regions.

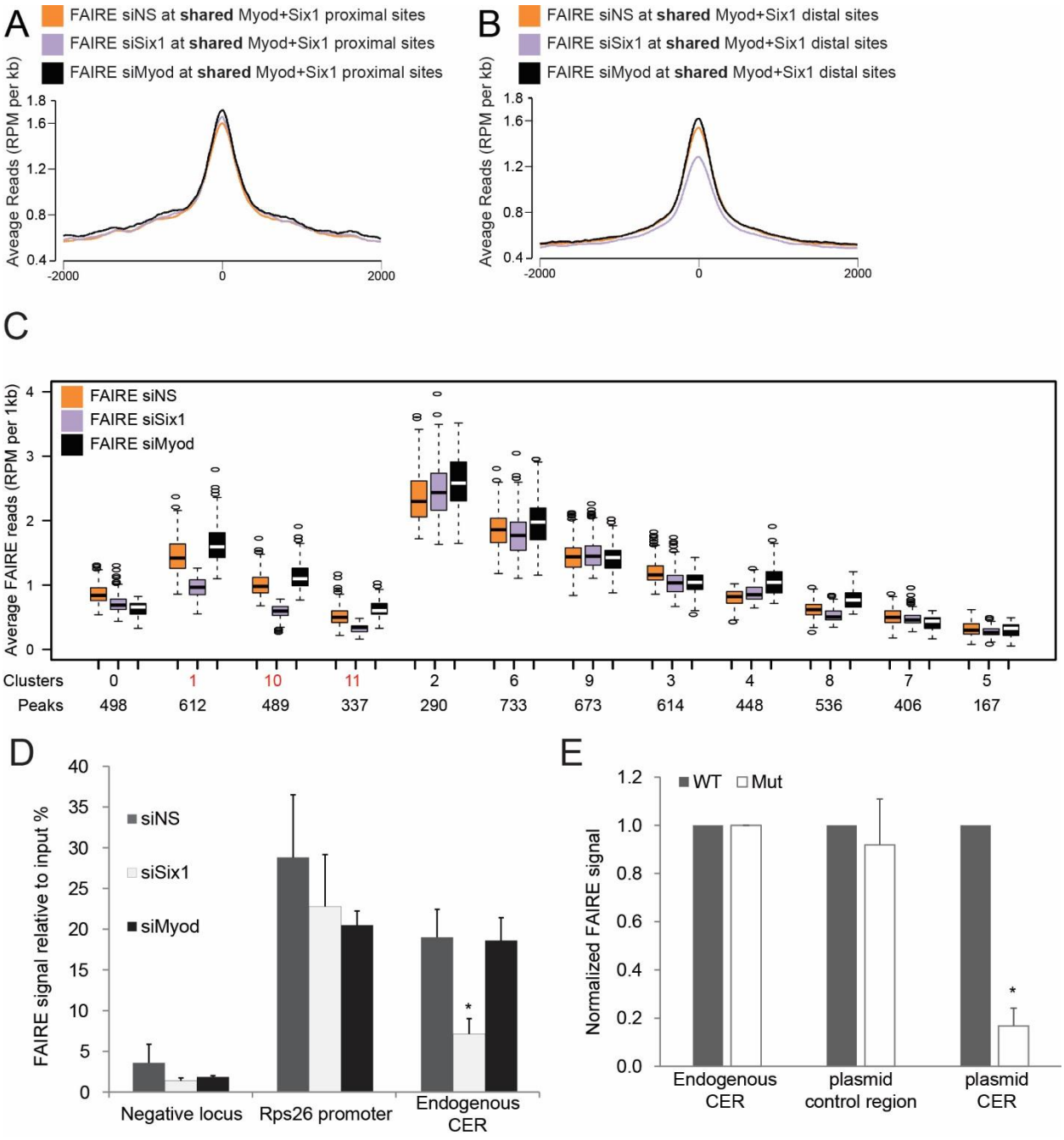

A
FAIRE siNS at shared Myod+Six1 proximal sites
FAIRE siSix1 at shared Myod+Six1 proximal sites
FAIRE siMyod at shared Myod+Six1 proximal sites
Average Reads (RPM per kb)
B
FAIRE siNS at shared Myod+Six1 distal sites
FAIRE siSix1 at shared Myod+Six1 distal sites
FAIRE siMyod at shared Myod+Six1 distal sites
Average Reads (RPM per kb)
C
FAIRE siNS
FAIRE siSix1
FAIRE siMyod
Average FAIRE reads (RPM per 1kb)
Clusters 0 1 10 11 2 6 9 3 4 8 7 5
Peaks 498 612 489 337 290 733 673 614 448 536 406 167
D
siNS
siSix1
siMyod
FAIRE signal relative to input %
Negative locus
Rps26 promoter
Endogenous CER
E
WT
Mut
Normalized FAIRE signal
Endogenous CER
plasmid control region
plasmid CER

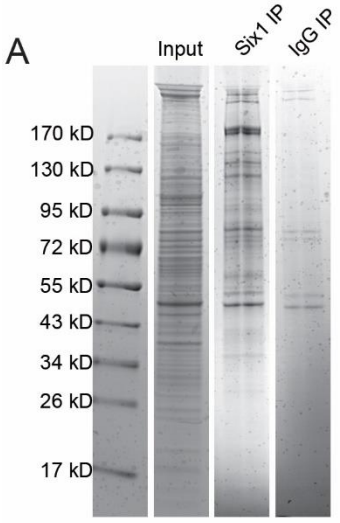
Figure 4. 4 Six1 is necessary to maintain the open chromatin structure at Myod distal regions

FAIRE-seq was performed after small interfering RNA (siRNA) transfection targeting non-specific, Six1 or Myod sequences. Average FAIRE-seq signals were compared among siNS, siSix1 and siMyod at A) proximal Myod and Six1 common targets, or B) distal Myod and Six1 common targets. Y-axis is the read density normalized to reads per million (RPM) per 1000 bp. C) 5803 distal Myod+Six1 sites were grouped into 12 clusters using Easeq based on FAIRE-seq signals after siNS, siSix1 and siMyod. FAIRE-seq signals were collected in a 400 bp region centered at Myod peak summits. The number of peaks in each cluster was shown at the bottom. D) FAIRE assay was performed in C2C12 myoblasts that were treated with siRNA, followed by quantitative PCR. Y-axis is the FAIRE signal relative to the input. Negative locus is a gene desert on the chromosome 15. Endogenous CER is the Myod core enhancer region. Rps26 is the promoter region of Rps26. $n = 4$. * $p < 0.05$. E) FAIRE assays were performed on stable C2C12 cell lines carrying either the WT CER or the mutant CER. Y-axis is the FAIRE signal normalized to the endogenous CER. The plasmid control region amplifies the plasmid sequence after the poly-A sequences. The plasmid CER specifically amplifies either the WT or the mutant CER from the plasmid sequences. $n = 3$. * $p < 0.05$.

Six1 forms a protein complex with Cdk9

Six1 is a homeodomain transcription factor and seems insufficient to remodel compact chromatin, in that a large proportion of DNA loci targeted by Six1 exhibited low DNA accessibility. We aimed at discovering proteins that could potentially mediate the influence of Six1 on the chromatin structure in myoblasts. Co-immunoprecipitation (co-IP) assay was performed to identify Six1-interacting proteins. Compared to the IgG control, a few proteins were co-immunoprecipitated with Six1 using C2C12 nuclear extracts (Fig. 5A). We identified 217 proteins with high confidence, and a few gene categories were enriched, such as DNA-templated transcription and chromatin (Fig. 5B). CDK9 was a Six1-interacting protein with high confidence. The activity of Cdk9, a cyclin-dependent kinase, is required for pause-release of transcription elongation (Marshall et al., 1996; Peterlin and Price, 2006). Furthermore, Cdk9 activity has been implicated on enhancers, such as its binding on enhancers of pluripotency-related genes and muscle-specific genes (Di Micco et al., 2014; Giacinti et al., 2006). To validate the interaction between CDK9 and SIX1, we performed co-IP in 293 cells. Recombinant over-expressed CDK9 and SIX1 were found in the same protein complex (Fig. 5C). We also tried to detect interaction of endogenous CDK9 and SIX1 for a few trials without success. Our lack of success in these assays may be caused by the possibly transient interaction between the two proteins when expressed at their natural levels, or by the lack of sensitivity of the assay.

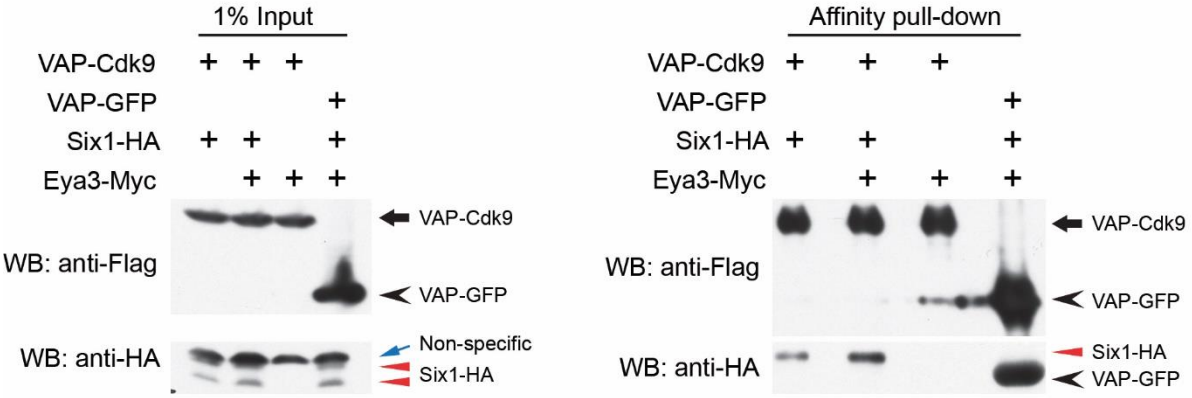
A



B

Term	Count	Gene list
GO:0006351~transcription, DNA-templated	19	**Six1, Cdk9**, Pelp1, Kdm1a, Ctnnd1, Hdac1, Znf687, Trim28, L3mbtl3, Xrn2, Znf592, Tp53bp1, Dido1, Mbd3, Ppp1r10, Polr2a, Chd4, Nab2, Nab1
GO:0003723~RNA binding	19	Prpf31, Edc3, Fip1l1, Ddx6, Dcp2, Xrn1, Cdkn2aip, Marf1, Rbmxl1, Ppp1r10, Matr3, Cpsf1, Cpsf2, Frg1, Pabpc1, Prpf8, Sart3, Hnrnpf, Skiv2l2
GO:0003676~nucleic acid binding	14	Ddx6, Xrn2, Zcchc8, Xrn1, Marf1, Rbmxl1, Cpsf1, Matr3, Chd4, Prpf8, Pabpc1, Sart3, Hnrnpf, Skiv2l2
GO:0006397~mRNA processing	18	Prpf31, Fip1l1, Prpf4, Srrm2, Xrn2, Zcchc8, Rbmxl1, Cpsf1, Cpsf2, Sart1, Frg1, Prpf3, Prpf8, Pabpc1, Sart3, Hnrnpf, Skiv2l2, Wdr33
GO:0000785~chromatin	6	Mbd3, Ppp1r10, Hdac1, Tox4, Trim28, Wdr82

C



Note: VAP-CDK9 52 kDa; VAP-GFP 37 kDa; Six1-HA 40 kDa

Figure 4. 5 Six1 forms protein complexes with Cdk9

A) Six1-interacting proteins were co-immunoprecipitated in C2C12 myoblasts. Proteins were separated by SDS-PAGE and stained by coomassie blue. B) Putative Six1-interacting proteins were classified using gene ontology analysis. Selected GO terms were listed with proteins identified in this study. C) VAP-tagged Cdk9 was over-expressed with HA-tagged Six1 and/or Myc-tagged Eya3 in 293T cells. VAP-tagged GFP was used as a negative control. The VAP tag is a combination of a few protein tags, allowing affinity-pull-down using Strep-Tactin® Sepharose® beads through the Twin-Strep tag and protein detection using anti-FLAG antibody through three tandem Flag tags. 1% Input was run as the positive control. Anti-FLAG antibody was used to detect recombinant Cdk9 or GFP. Anti-HA antibody was used to detect recombinant Six1. The strong autofluorescence of VAP-GFP gave a band that was not mediated by primary or secondary antibody in the pull-down.

Cdk9 activity is needed for Myod expression and its recruitment on the CER depends on Six1

Cdk9 and Myod have been shown to co-occupy regulatory elements of muscle-specific genes (Giacinti et al., 2006). We determined to evaluate whether Cdk9 is a functional regulator of Myod. Flavopiridol is a chemical inhibitor of Cdk9 kinase activity and has been shown to inhibit RNA polymerase II-mediated transcription very potently (Chao & Price, 2001). Notably, it shows ten to twenty times greater inhibitory activity on CDK9 compared to other CDKs (Chao and Price, 2001; Kim et al., 2000; Krystof and Uldrijan, 2010). First we tested whether inhibiting the function of Cdk9 by Flavopiridol had an influence on the chromatin structure at the CER. Indeed, Flavopiridol treatment led to reduced chromatin accessibility at the CER, which was similar to the observation after Six1 KD (Fig. 6A).

The CER is required for the proper spatiotemporal expression of Myod during embryonic development (Chen and Goldhamer, 2004; Goldhamer et al., 1995). Moreover, the CER is also transcribed to an enhancer RNA (^{CE}RNA), and the ^{CE}RNA transcript has been shown to be essential for the expression of Myod (Mousavi et al., 2013). However, it was not clear whether Myod is regulated by Cdk9. We decided to examine the Myod expression after the inhibition of Cdk9 activity. Flavopiridol was able to inhibit the expression of Myod at 50 nM after 4 hours of incubation (Fig. 6B), suggesting that MyoD is regulated by transcription pausing. The expression of Six1 was also affected by Flavopiridol but to a lesser extent. We cannot exclude that reduced Myod expression is also mediated by partial loss of Six1.

Furthermore, we tested whether CDK9 was recruited to the regulatory regions of Myod. ChIP assay was performed on C2C12 myoblasts and CDK9 was enriched at the CER but not at the

DRR or the PRR, suggesting that the CER confers the regulation by CDK9 (Fig. 6B). Next, we examined whether the recruitment of CDK9 depends on the presence of Six1. We performed ChIP assays on stable C2C12 cell lines carrying either the WT CER (WT) or the mutant CER (Mut) (Liu et al., 2013c). While the recruitment of CDK9 was not statistically different at the endogenous CER between two CER cell lines, CDK9 was not able to occupy the mutant CER when Six1 binding sites were missing (Fig. 6B). Taken together, our results suggest that Six1 mediates the function of Cdk9 at the Myod locus, which is required to maintain accessible chromatin at the CER and is essential for Myod expression.

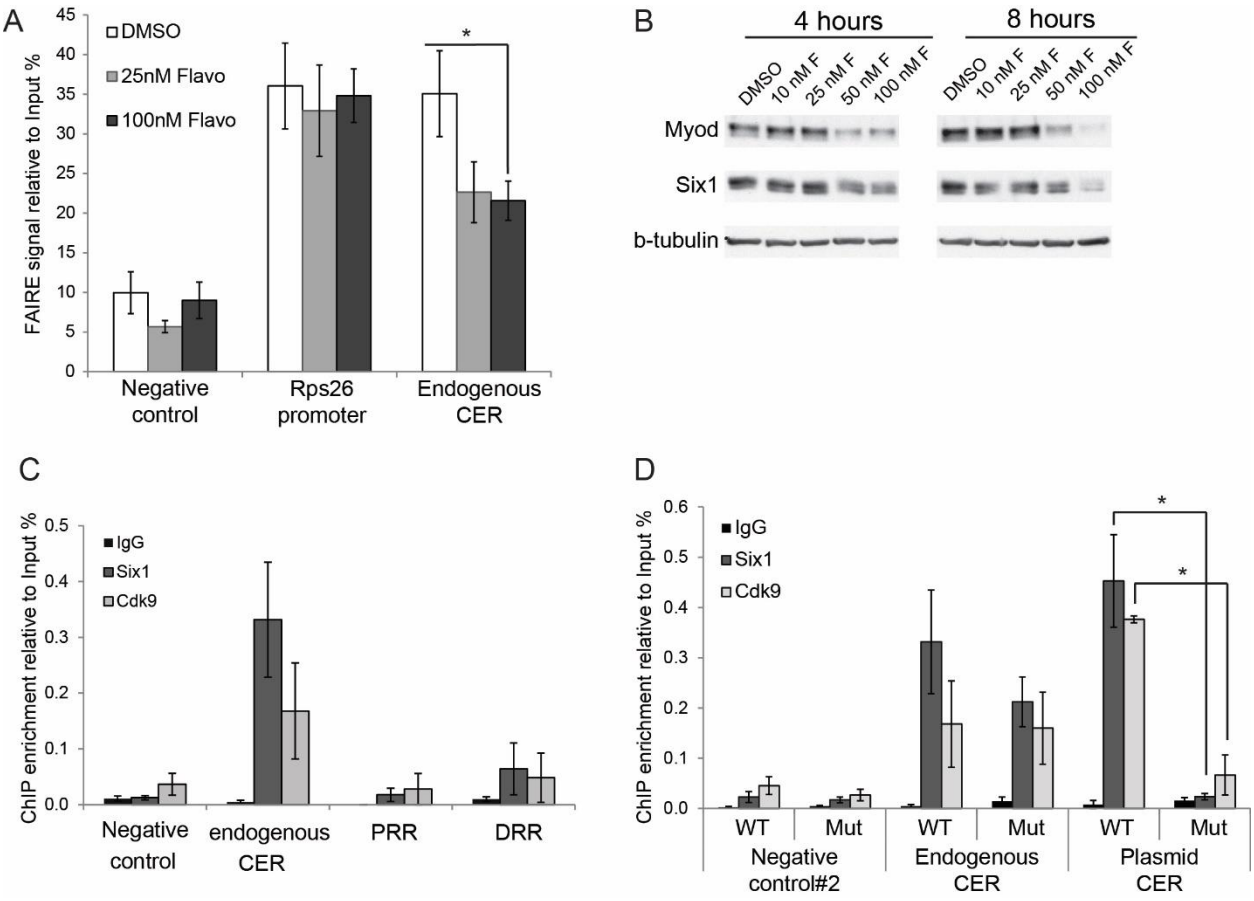

A
DMSO
25nM Flavo
100nM Flavo
FAIRE signal relative to Input %
Negative control
Rps26 promoter
Endogenous CER
*
B
4 hours
8 hours
DMSO
10 nM F
25 nM F
50 nM F
100 nM F
Myod
Six1
b-tubulin
C
IgG
Six1
Cdk9
ChIP enrichment relative to Input %
Negative control
endogenous CER
PRR
DRR
D
IgG
Six1
Cdk9
ChIP enrichment relative to Input %
WT
Mut
Negative control#2
Endogenous CER
Plasmid CER
*
*

Figure 4. 6 Cdk9 regulates the expression of Myod and its recruitment on the CER depends on Six1

A) FAIRE assay was performed on C2C12 myoblasts that were treated with Flavopiridol. Y-axis is the FAIRE signal relative to the input. C2C12 myoblasts were incubated with Flavopiridol at 25 nM or 100 nM for 24 hours. DMSO was used as the vehicle control at 0.05%. * $p < 0.05$; $n >= 3$. B) Western blots were performed using Myod, Six1 and beta-tubulin antibodies. C2C12 myoblasts were incubated with Flavopiridol at 10 nM, 25 nM, 50 nM or 100 nM (10 nM F, 25 nM F, 50 nM F or 100 nM F) for 4 or 8 hours. DMSO was used as the vehicle control at 0.05%. C) ChIP assay was performed on C2C12 myoblasts. ChIP signals were shown as relative to the input. $n = 3$. D) ChIP assay was performed on C2C12 stable cell lines carrying either the WT CER or the mutant CER, using Cdk9 and Six1 antibodies. ChIP signals were shown as relative to the input. $n = 3$. * $p < 0.05$. Negative control is a gene desert region on chromosome 15, as a closed DNA locus. Endogenous CER amplifies the Myod core enhancer region. PRR amplifies the proximal regulatory region of Myod. DRR amplifies the distal regulator region of Myod. The plasmid CER specifically amplifies either the WT or the mutant CER from the plasmid sequences. Rps26 promoter is the promoter region of Rps26. Negative control #2 is the promoter region of HoxD10.

DISCUSSION

Accessible chromatin is prerequisite for gene activation in mammalian cells (Lee et al., 2004a). Six1 and Myod co-occupied a significant number of targets and their co-occurrence correlated with more open chromatin at these loci (Fig. 2B). At distal regions, the co-binding of Six1 and Myod was also associated with marks of active enhancers (Fig. 3). Interestingly, nearly 20% of Myod peaks had Six1 binding within 150 bp, which roughly matches the length of nucleosomal DNA, suggesting that they often function through the same enhancer (Fig. 1D). Moreover, loss of function of Six1 led to reduced chromatin accessibility at nearly 30% of shared Myod+Six1 distal regions (Fig. 4B and 4C). Myod, as a master regulator in the myogenic lineage, was shown to be able to recruit co-factors, such as Baf60c, a subunit of the SWI/SNF chromatin remodelling complex, and p300, a protein acetyltransferase, to activate target gene expression (Forcales et al., 2012; Magenta et al., 2003; Sartorelli et al., 1997). Previously, our lab showed that Myod expression in myoblasts is under the control of Six1 through the CER enhancer (Liu et al., 2013). Therefore, it is important to differentiate the impact of Six1 on the chromatin structure from that of Myod. FAIRE assays combined with siRNA gene knockdown confirmed that the chromatin structure at a subset of shared Myod+Six1 distal sites was affected by Six1 but not by Myod (Fig. 4B). We propose that Six1 is required to maintain accessible chromatin at these loci.

It has been reported that the Six and MRF families synergistically regulate the expression of their target genes (Chakroun et al., 2015; Liu et al., 2010; Santolini et al., 2016). Six1 and Six4 are required to activate a cluster of myogenic genes mediated by Myod (Chakroun et al., 2015; Santolini et al., 2016). Here we demonstrated that Six1 was necessary to maintain accessible chromatin at many potential enhancers that were co-bound by Six1 and Myod. Considering that open chromatin is a mark of gene activation, it is likely that Six1 is involved in Myod-mediated

gene transcription through affecting chromatin accessibility at potential enhancers.

The regulation of enhancers is tightly controlled in a spatiotemporal manner. Genomics studies have found that active enhancers that are associated with gene activation possess characteristic histone marks and protein binding. Here, Myod distal targets co-bound by Six1 displayed prominent features of active enhancers, including the recruitment of RNA pol II. RNA pol II mediates the transcription of eRNA at enhancers, which is often induced before or at least at the same time with their target gene expression (Hah et al., 2013; Hsieh et al., 2014; Kim et al., 2015b). The production of eRNAs has been reported to be required for the transcription of their target gene in certain tissues, such as eRNA at the CER (Mousavi et al., 2013). KD of eRNA at the CER leads to decreased Myod expression, which is similar to KD of Six1. In macrophage cells, the action of active transcription on TLR4 signaling-induced *de novo* enhancers is necessary to maintain the H3K4me1 histone mark (Kaikkonen et al., 2013). In our study, when the transcription elongation was blocked by Flavopiridol, the transcription action was halted and the eRNA products were likely reduced as well, which led to reduced Myod expression and chromatin accessibility at the CER. Myod expression is sensitive to Flavopiridol treatment, compared to incubation of the drug either for a few days or at a concentration of 1 μM in C2C12 cells (Mousavi et al., 2012; O'Brien et al., 2012). Therefore, the function of Cdk9 is required for the Myod expression in myoblasts.

The activity of Cdk9 has been shown to be involved in the Myod-mediated fibroblast conversion (Simone et al., 2002). Cdk9 is able to interact with Myod and they co-occupy myogenic regulatory regions, such as the Myog promoter and the MCK (creatine kinase, muscle) enhancer (Giacinti et al., 2006; Simone et al., 2002). However, it was not discussed whether the recruitment of Cdk9 depended on Myod at these DNA loci. Here, we showed for the first time

that the binding of Cdk9 at an enhancer depends on the presence of a transcription factor, Six1. Six1 and Cdk9 were detected in the same protein complex. But whether the interaction is direct was not tested here. Considering the associations between Cdk9 and several other transcription factors such as Myod, Myc and Stat3, it cannot be ruled out that Six1 recruits these transcription factors which further tether Cdk9 on the CER and other enhancers (Giacinti et al., 2006; Giraud et al., 2004; Kanazawa et al., 2003).

The existing open regulatory regions were analyzed in this study. However, it would be interesting to test whether Six1 is required to initiate the chromatin remodelling at compact loci. For example, the Myog promoter is inaccessible in myoblasts and becomes accessible upon differentiation (Faralli et al., 2012; Giacinti et al., 2006). As a common target of Six1, Cdk9 and Myod, the expression of Myog is induced upon differentiation (Giacinti et al., 2006; Liu et al., 2010). To our knowledge, whether Six1 participates in remodeling the Myog promoter for the assembly of the transcription pre-initiation complex has not been tested.

As a potential therapeutic approach, satellite cell transplantation encounters a few obstacles, including premature differentiation, partially mediated by Myod induction (Asakura et al., 2007; Montarras et al., 2005). Satellite cell activation is concomitant with a wave of chromatin remodeling and gene expression. The following muscle differentiation requires the temporal expression of Myod and activation of its downstream genes. Our preliminary data showed that chromatin at the Myod CER became more accessible upon satellite cell activation (Fig. S5). Is Six1 essential for this change at the CER? Is it through recruiting chromatin remodeling complexes or stimulating active transcription at the CER? Cdk9 was shown to play a role in muscle regeneration upon injury (Giacinti et al., 2008). Cdk9 and Myod have been shown to co-regulate the expression of target genes, and here we demonstrated that Cdk9 could also function

upstream of Myod (Giacinti et al., 2006; Simone et al., 2002). Temporal control of Myod expression in satellite cells through specific Cdk9 inhibitors could be beneficial to prevent premature differentiation and to implement the following therapeutic applications.

SUPPLEMENTARY DATA

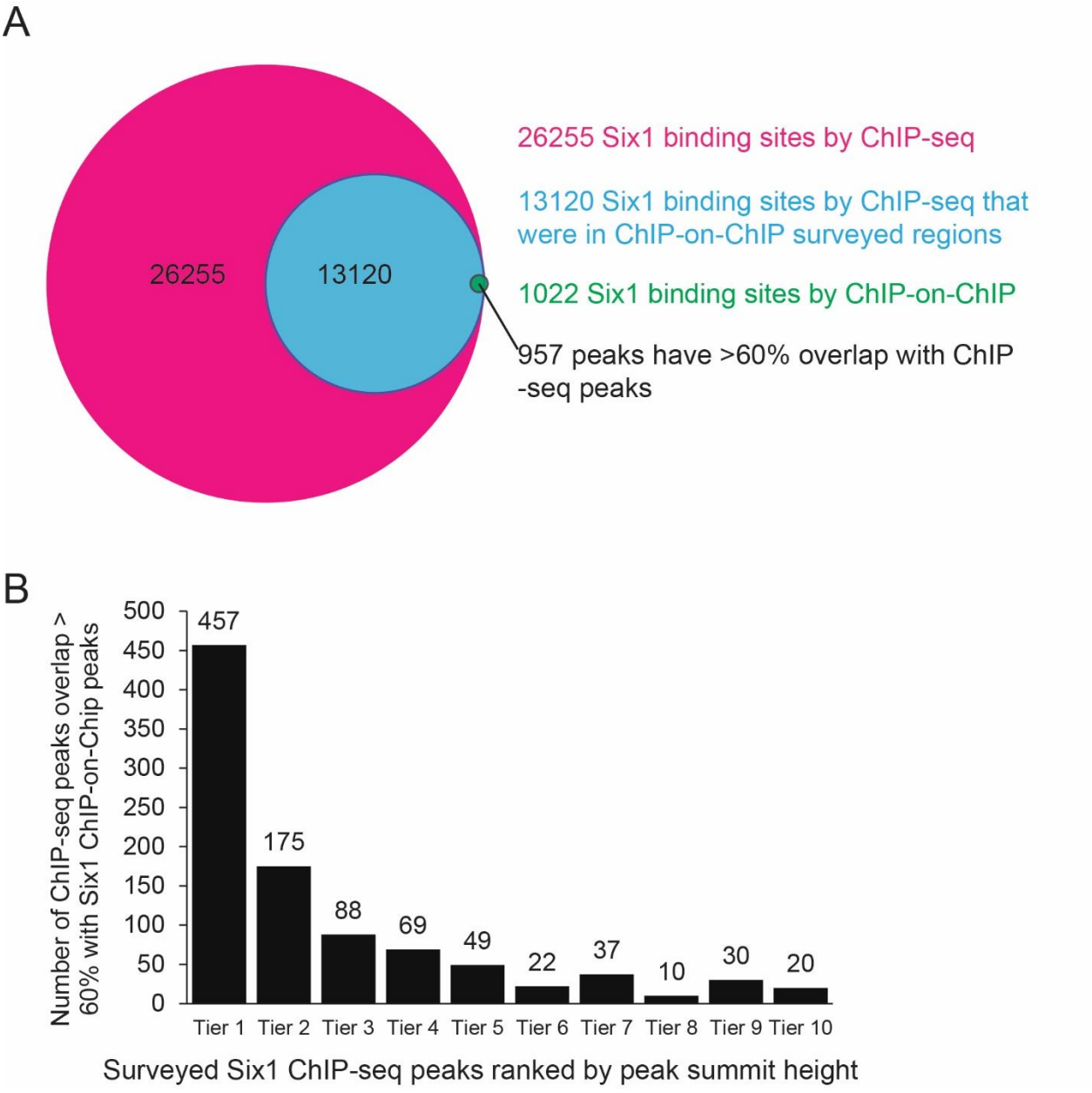


Figure 4.S 1 Comparison of Six1 ChIP-seq and ChIP-on-Chip data

A) Six1 ChIP-on-Chip was performed on C2C12 myoblasts. Six1 ChIP-seq was performed on primary myoblasts. A) Peaks identified in ChIP-seq and ChIP-on-Chip were shown in pink and green respectively. Considering that ChIP-on-Chip arrays represent a proportion of the genome (surveyed regions), Six1 ChIP-seq peaks that fell into survey regions were shown in blue. The majority of ChIP-on-Chip peaks were also found in ChIP-seq with good accuracy. B) ChIP-seq peaks in surveyed regions were sorted by the height of peak summit and were grouped into 10 tiers. The number of ChIP-on-Chip peaks that had more than 60% overlap with ChIP-seq peaks in each tier was shown here.

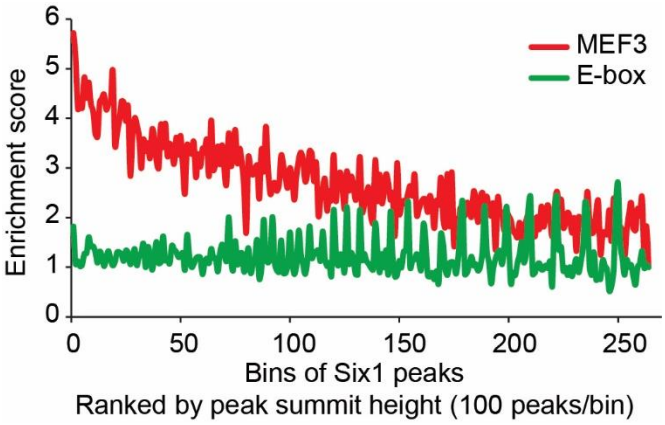

MEF3
E-box
Enrichment score
0
1
2
3
4
5
6
0
50
100
150
200
250
Bins of Six1 peaks
Ranked by peak summit height (100 peaks/bin)

Figure 4.S 2 MEF3 DNA motif is enriched among Six1 binding sites

The enrichment of MEF3 DNA motif was analyzed using Cisgenome (Ji et al., 2006). The control regions were selected using Cisgenome. Six1 peaks were sorted according to the height of peak summit. Each bin contains 100 Six1 peaks.

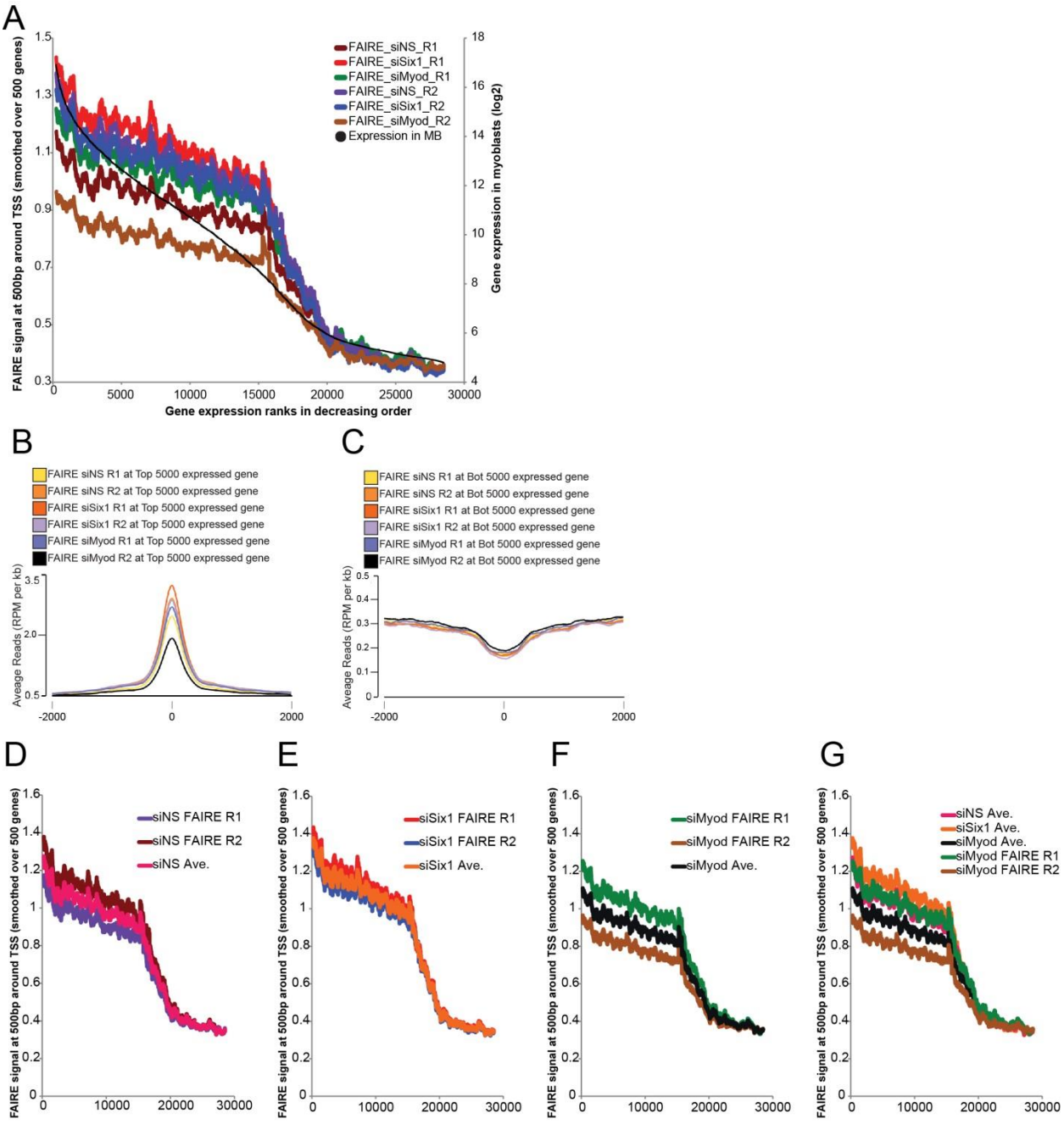
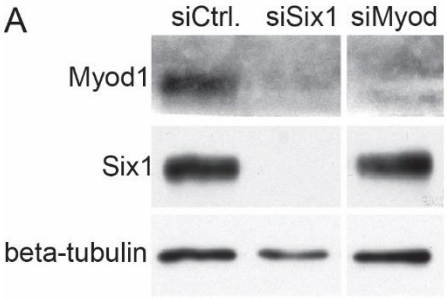

A
FAIRE_siNS_R1
FAIRE_siSix1_R1
FAIRE_siMyod_R1
FAIRE_siNS_R2
FAIRE_siSix1_R2
FAIRE_siMyod_R2
Expression in MB
FAIRE signal at 500bp around TSS (smoothed over 500 genes)
Gene expression in myoblasts (log2)
Gene expression ranks in decreasing order
B
FAIRE siNS R1 at Top 5000 expressed gene
FAIRE siNS R2 at Top 5000 expressed gene
FAIRE siSix1 R1 at Top 5000 expressed gene
FAIRE siSix1 R2 at Top 5000 expressed gene
FAIRE siMyod R1 at Top 5000 expressed gene
FAIRE siMyod R2 at Top 5000 expressed gene
Aveage Reads (RPM per kb)
C
FAIRE siNS R1 at Bot 5000 expressed gene
FAIRE siNS R2 at Bot 5000 expressed gene
FAIRE siSix1 R1 at Bot 5000 expressed gene
FAIRE siSix1 R2 at Bot 5000 expressed gene
FAIRE siMyod R1 at Bot 5000 expressed gene
FAIRE siMyod R2 at Bot 5000 expressed gene
Aveage Reads (RPM per kb)
D
siNS FAIRE R1
siNS FAIRE R2
siNS Ave.
FAIRE signal at 500bp around TSS (smoothed over 500 genes)
E
siSix1 FAIRE R1
siSix1 FAIRE R2
siSix1 Ave.
FAIRE signal at 500bp around TSS (smoothed over 500 genes)
F
siMyod FAIRE R1
siMyod FAIRE R2
siMyod Ave.
FAIRE signal at 500bp around TSS (smoothed over 500 genes)
G
siNS Ave.
siSix1 Ave.
siMyod Ave.
siMyod FAIRE R1
siMyod FAIRE R2
FAIRE signal at 500bp around TSS (smoothed over 500 genes)

Figure 4.S 3 FAIRE-seq of R1 siNS, R2 siNS, R1 siSix1, R2 siSix1, and R1 siMyod are included in the analysis

A) Transcription start sites (TSS) of 28756 genes with detectable expression in C2C12 myoblasts were collected and ranked by their expression level (black line). FAIRE-seq signals were quantitated in a 400-bp window centered at the TSS for each data set. The Y-axis represents the FAIRE signal. The secondary Y-axis represents the gene expression level which is $\log_2$ transformed. B-C) Average FAIRE signals were represented at the promoter regions of top 5000 (B) or bottom 5000 (C) of expressed genes in C2C12 myoblasts. Y-axis is the read density normalized to reads per million (RPM) per 1000 bp. FAIRE-seq signals were shown for two independent replicates and the average of D) siNS, E) siSix1 and F) siMyod. G) The average FAIRE signals were compared among siNS, siSix1 and siMyod. Two replicates of FAIRE-seq after siMyod were also shown here.

.


A
siCtrl.
siSix1
siMyod
Myod1
Six1
beta-tubulin

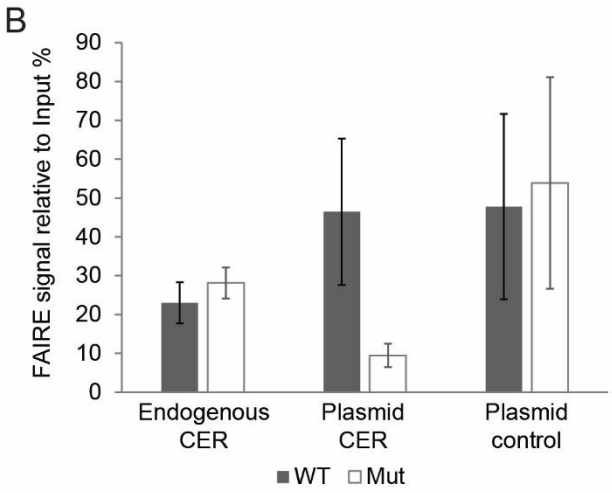

B
FAIRE signal relative to Input %
90
80
70
60
50
40
30
20
10
0
Endogenous CER
Plasmid CER
Plasmid control
WT
Mut

Figure 4.S 4 Validation of knockdown for FAIRE-experiments and the raw FAIRE signal before normalization on stable cell lines

A) Representative western blot showed efficient knock-down of Six1 and Myod. The whole cell extract of C2C12 myoblasts with different siRNA treatment was probed with anti-Six1, anti-Myod or anti-beta tubulin antibody. B) FAIRE assay was performed on WT and Mut CER stable C2C12 cell lines. Y-axis is the FAIRE signal relative to the Input. n = 3. Endogenous CER amplifies the Myod core enhancer region. The plasmid CER specifically amplifies either the WT or the mutant CER from the plasmid sequences. The plasmid control region amplifies the plasmid sequence after the poly-A sequences.

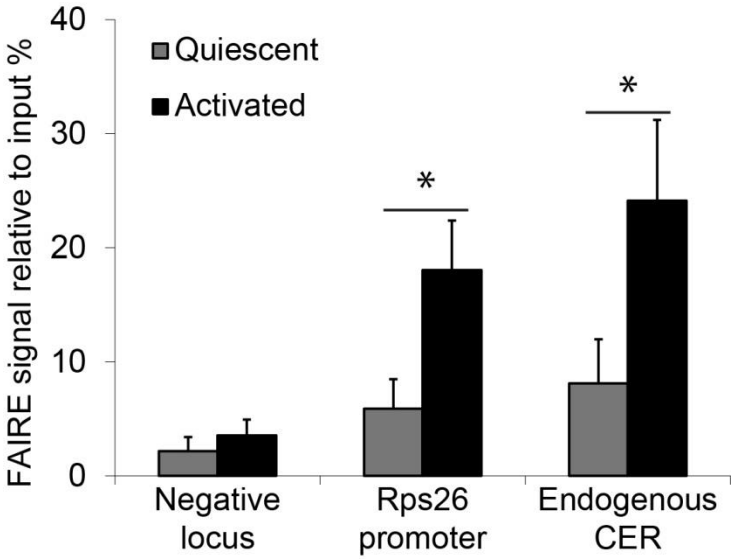

40
30
20
10
0
FAIRE signal relative to input %
Quiescent
Activated
*
*
Negative locus
Rps26 promoter
Endogenous CER

Figure 4.S 5 The Myod CER becomes more accessible upon satellite cell activation

FAIRE assay followed by quantitative PCR was performed on freshly isolated satellite cells (quiescent) or cultured satellite cells (activated). Y-axis is the FAIRE signal relative to the Input. Satellite cells were collected by Flow Cytometry from *Pax7-ZsGreen* reporter mice. Hind limb muscles were collected and digested as reported previously (Pasut et al., 2012). Negative locus is a gene desert on chromosome 15. Endogenous CER is the Myod core enhancer region. Rps26 is the promoter region of Rps26. $n = 3$. * $p < 0.05$.

ACKNOWLEDGEMENTS

This research was supported by CIHR grant MOP-119458 to A.B. and NSERC Postgraduate Scholarships-Doctoral to Y.L.. We thank Drs. Marjorie Brand and Lawrence Puente who provided insight and expertise to the proteomic work. We thank Dr. Mads Lerdrup for assistance with Easeq. We thank Dr. Eric Milot for the Cdk9 expression plasmids. We thank Drs. Shilatifard and Smith for informative discussion.

REFERENCES

1. Alli, N.S., Yang, E.C., Miyake, T., Aziz, A., Collins-Hooper, H., Patel, K., and McDermott, J.C. (2013). Signal-dependent fra-2 regulation in skeletal muscle reserve and satellite cells. Cell Death Dis. *4*, e692.

2. Andreucci, J.J., Grant, D., Cox, D.M., Tomc, L.K., Prywes, R., Goldhamer, D.J., Rodrigues, N., Bedard, P.A., and McDermott, J.C. (2002). Composition and function of AP-1 transcription complexes during muscle cell differentiation. J. Biol. Chem. *277*, 16426–16432.

3. Andrews, J.L., Zhang, X., McCarthy, J.J., McDearmon, E.L., Hornberger, T.A., Russell, B., Campbell, K.S., Arbogast, S., Reid, M.B., Walker, J.R., et al. (2010). CLOCK and BMAL1 regulate MyoD and are necessary for maintenance of skeletal muscle phenotype and function. Proc. Natl. Acad. Sci. U. S. A. *107*, 19090–19095.

4. Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science *347*, 1010–1014.

5. Asakura, A., Lyons, G.E., and Tapscott, S.J. (1995). The regulation of MyoD gene expression: conserved elements mediate expression in embryonic axial muscle. Dev. Biol. *171*, 386–398.

6. Asakura, A., Hirai, H., Kablar, B., Morita, S., Ishibashi, J., Piras, B.A., Christ, A.J., Verma, M., Vineretsky, K.A., and Rudnicki, M.A. (2007). Increased survival of muscle stem cells lacking the MyoD gene after transplantation into regenerating skeletal muscle. Proc Natl Acad Sci U A *104*, 16552–16557.

7. Asp, P., Blum, R., Vethantham, V., Parisi, F., Micsinai, M., Cheng, J., Bowman, C., Kluger, Y., and Dynlacht, B.D. (2011). Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. Proc. Natl. Acad. Sci. U. S. A. *108*, E149-58.

8. Blais, A. (2015). Myogenesis in the Genomics Era. J. Mol. Biol. *427*, 2023–2038.

9. Blum, R., and Dynlacht, B.D. (2013). The role of MyoD1 and histone modifications in the activation of muscle enhancers. Epigenetics Off. J. DNA Methylation Soc. *8*, 778–784.

10. Blum, R., Vethantham, V., Bowman, C., Rudnicki, M., and Dynlacht, B.D. (2012). Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. Genes Dev. *26*, 2763–2779.

11. Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G.M., Carnemolla, B., Orecchia, P., Zardi, L., and Righetti, P.G. (2004). Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis *25*, 1327–1333.

12. 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.

13. Chakroun, I., Yang, D., Girgis, J., Gunasekharan, A., Phenix, H., Kærn, M., and Blais, A. (2015). Genome-wide association between Six4, MyoD, and the histone demethylase Utx during myogenesis. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *29*, 4738–4755.

14. Chang, N.C., and Rudnicki, M.A. (2014). Satellite cells: the architects of skeletal muscle. Curr. Top. Dev. Biol. *107*, 161–181.

15. Chao, S.H., and Price, D.H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J. Biol. Chem. *276*, 31793–31799.

16. Chen, J.C., and Goldhamer, D.J. (2004). The core enhancer is essential for proper timing of MyoD activation in limb buds and branchial arches. Dev. Biol. *265*, 502–512.

17. Chen, J.C., Love, C.M., and Goldhamer, D.J. (2001). Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *221*, 274–288.

18. Chen, J.C., Ramachandran, R., and Goldhamer, D.J. (2002). Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. Dev. Biol. *245*, 213–223.

19. Cornelison, D.D., and Wold, B.J. (1997). Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells. Dev. Biol. *191*, 270–283.

20. Creemers, E.E., Sutherland, L.B., Oh, J., Barbosa, A.C., and Olson, E.N. (2006). Coactivation of MEF2 by the SAP Domain Proteins Myocardin and MASTR. Mol. Cell *23*, 83–96.

21. Di Micco, R., Fontanals-Cirera, B., Low, V., Ntziachristos, P., Yuen, S.K., Lovell, C.D., Dolgalev, I., Yonekubo, Y., Zhang, G., Rusinova, E., et al. (2014). Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. Cell Rep. *9*, 234–247.

22. Faralli, H., Dilworth, F.J., Faralli, H., and Dilworth, F.J. (2012). Turning on Myogenin in Muscle: A Paradigm for Understanding Mechanisms of Tissue-Specific Gene Expression, Turning on Myogenin in Muscle: A Paradigm for Understanding Mechanisms of Tissue-Specific Gene Expression. Int. J. Genomics Int. J. Genomics *2012*, *2012*, e836374.

23. Feng, J., Liu, T., and Zhang, Y. (2011). Using MACS to identify peaks from ChIP-Seq data. Curr. Protoc. Bioinforma. Ed. Board Andreas DBaxevanis Al *Chapter 2*, Unit 2.14.

24. Forcales, S.V., Albini, S., Giordani, L., Malecova, B., Cignolo, L., Chernov, A., Coutinho, P., Saccone, V., Consalvi, S., Williams, R., et al. (2012). Signal-dependent

incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex. EMBO J. *31*, 301–316.

25. Fu, T.J., Peng, J., Lee, G., Price, D.H., and Flores, O. (1999). Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. J. Biol. Chem. *274*, 34527–34530.

26. Fukada, S., Uezumi, A., Ikemoto, M., Masuda, S., Segawa, M., Tanimura, N., Yamamoto, H., Miyagoe-Suzuki, Y., and Takeda, S. 'ichi (2007). Molecular signature of quiescent satellite cells in adult skeletal muscle. Stem Cells Dayt. Ohio *25*, 2448–2459.

27. García-González, E., Escamilla-Del-Arenal, M., Arzate-Mejía, R., and Recillas-Targa, F. (2016). Chromatin remodeling effects on enhancer activity. Cell. Mol. Life Sci. CMLS.

28. Giacinti, C., Bagella, L., Puri, P.L., Giordano, A., and Simone, C. (2006). MyoD recruits the cdk9/cyclin T2 complex on myogenic-genes regulatory regions. J. Cell. Physiol. *206*, 807–813.

29. Giacinti, C., Musarò, A., De Falco, G., Jourdan, I., Molinaro, M., Bagella, L., Simone, C., and Giordano, A. (2008). Cdk9-55: a new player in muscle regeneration. J. Cell. Physiol. *216*, 576–582.

30. Giraud, S., Hurlstone, A., Avril, S., and Coqueret, O. (2004). Implication of BRG1 and cdk9 in the STAT3-mediated activation of the p21waf1 gene. Oncogene *23*, 7391–7398.

31. Goldhamer, D.J., Faerman, A., Shani, M., and Jr, C.P.E. (1992). Regulatory elements that control the lineage-specific expression of myoD. Science *256*, 538–542.

32. Goldhamer, D.J., Brunk, B.P., Faerman, A., King, A., Shani, M., and Jr, C.P.E. (1995). Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. Dev. Camb. Engl. *121*, 637–649.

33. Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M.J., Hamard, G., and Maire, P. (2005). Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Dev. Camb. Engl. *132*, 2235–2249.

34. Gunther, S., Kim, J., Kostin, S., Lepper, C., Fan, C.M., and Braun, T. (2013). Myf5-positive satellite cells contribute to Pax7-dependent long-term maintenance of adult muscle stem cells. Cell Stem Cell *13*, 590–601.

35. Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. *23*, 1210–1223.

36. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Calcar, S.V., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

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

38. Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.-S.M., Chen, S., et al. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc. Natl. Acad. Sci. U. S. A. *111*, 7319–7324.

39. Hu, P., Geles, K.G., Paik, J.H., DePinho, R.A., and Tjian, R. (2008). Codependent activators direct myoblast-specific MyoD transcription. Dev. Cell *15*, 534–546.

40. Huang, W., Loganantharaj, R., Schroeder, B., Fargo, D., and Li, L. (2013). PAVIS: a tool for Peak Annotation and Visualization. Bioinforma. Oxf. Engl. *29*, 3097–3099.

41. Ji, H., Vokes, S.A., and Wong, W.H. (2006). A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. Nucleic Acids Res. *34*, e146.

42. Ji, Z., Luo, W., Li, W., Hoque, M., Pan, Z., Zhao, Y., and Tian, B. (2011). Transcriptional activity regulates alternative cleavage and polyadenylation. Mol. Syst. Biol. *7*, 534.

43. Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell *51*, 310–325.

44. Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., and Peterlin, B.M. (2003). c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene *22*, 5707–5711.

45. Kawakami, K., Sato, S., Ozaki, H., and Ikeda, K. (2000). Six family genes–structure and function as transcription factors and their roles in development. BioEssays News Rev. Mol. Cell. Dev. Biol. *22*, 616–626.

46. Kim, K.S., Sack, J.S., Tokarski, J.S., Qian, L., Chao, S.T., Leith, L., Kelly, Y.F., Misra, R.N., Hunt, J.T., Kimball, S.D., et al. (2000). Thio- and oxoflavopiridols, cyclin-dependent kinase 1-selective inhibitors: synthesis and biological effects. J. Med. Chem. *43*, 4126–4134.

47. Kim, T.-K., Hemberg, M., and Gray, J.M. (2015a). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb. Perspect. Biol. *7*, a018622.

48. Kim, Y.W., Lee, S., Yun, J., and Kim, A. (2015b). Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the β-globin locus. Biosci. Rep. *35*.

49. Krystof, V., and Uldrijan, S. (2010). Cyclin-dependent kinase inhibitors as anticancer drugs. Curr. Drug Targets *11*, 291–302.

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

51. Laclef, C., Hamard, G., Demignon, J., Souil, E., Houbron, C., and Maire, P. (2003a). Altered myogenesis in Six1-deficient mice. Dev. Camb. Engl. *130*, 2239–2252.

52. Laclef, C., Souil, E., Demignon, J., and Maire, P. (2003b). Thymus, kidney and craniofacial abnormalities in Six 1 deficient mice. Mech. Dev. *120*, 669–679.

53. Lam, M.T.Y., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. *39*, 170–182.

54. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

55. Le Grand, F., Grifone, R., Mourikis, P., Houbron, C., Gigaud, C., Pujol, J., Maillet, M., Pagès, G., Rudnicki, M., Tajbakhsh, S., et al. (2012). Six1 regulates stem cell repair potential and self-renewal during skeletal muscle regeneration. J. Cell Biol. *198*, 815–832.

56. Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004a). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat. Genet. *36*, 900–905.

57. Lee, H., Habas, R., and Abate-Shen, C. (2004b). MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. Science *304*, 1675–1678.

58. Lepper, C., Conway, S.J., and Fan, C.M. (2009). Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements. Nature *460*, 627–631.

59. Lerdrup, M., Johansen, J.V., Agrawal-Singh, S., and Hansen, K. (2016). An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat. Struct. Mol. Biol. *23*, 349–357.

60. L'honore, A., Lamb, N.J., Vandromme, M., Turowski, P., Carnac, G., and Fernandez, A. (2003). MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. Mol. Biol. Cell *14*, 2151–2162.

61. L'honore, A., Rana, V., Arsic, N., Franckhauser, C., Lamb, N.J., and Fernandez, A. (2007). Identification of a new hybrid serum response factor and myocyte enhancer factor 2-binding element in MyoD enhancer required for MyoD expression during myogenesis. Mol. Biol. Cell *18*, 1992–2001.

62. L'honoré, A., Ouimette, J.-F., Lavertu-Jolin, M., and Drouin, J. (2010). Pitx2 defines alternate pathways acting through MyoD during limb and somitic myogenesis. Dev. Camb. Engl. *137*, 3847–3856.

63. Liu, L., Cheung, T.H., Charville, G.W., Hurgo, B.M., Leavitt, T., Shih, J., Brunet, A., and Rando, T.A. (2013a). Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. Cell Rep. *4*, 189–204.

64. Liu, Y., Chu, A., Chakroun, I., Islam, U., and Blais, A. (2010). Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. Nucleic Acids Res. *38*, 6857–6871.

65. Liu, Y., Nandi, S., Martel, A., Antoun, A., Ioshikhes, I., and Blais, A. (2012). Discovery, optimization and validation of an optimal DNA-binding sequence for the Six1 homeodomain transcription factor. Nucleic Acids Res. *40*, 8227–8239.

66. Liu, Y., Chakroun, I., Yang, D., Horner, E., Liang, J., Aziz, A., Chu, A., Repentigny, Y.D., Dilworth, F.J., Kothary, R., et al. (2013b). Six1 regulates MyoD expression in adult muscle progenitor cells. PloS One *8*, e67762.

67. Magenta, A., Cenciarelli, C., De Santa, F., Fuschi, P., Martelli, F., Caruso, M., and Felsani, A. (2003). MyoD stimulates RB promoter activity via the CREB/p300 nuclear transduction pathway. Mol. Cell. Biol. *23*, 2893–2906.

68. Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J. Biol. Chem. *271*, 27176–27183.

69. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

70. Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A.F., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., et al. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol. Cell *49*, 524–535.

71. Mokalled, M.H., Johnson, A.N., Creemers, E.E., and Olson, E.N. (2012). MASTR directs MyoD-dependent satellite cell differentiation during skeletal muscle regeneration. Genes Dev. *26*, 190–202.

72. Montarras, D., Morgan, J., Collins, C., Relaix, F., Zaffran, S., Cumano, A., Partridge, T., and Buckingham, M. (2005). Direct isolation of satellite cells for skeletal muscle regeneration. Science *309*, 2064–2067.

73. Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol. Cell *45*, 255–262.

74. Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol. Cell *51*, 606–617.

75. Murphy, M.M., Lawson, J.A., Mathew, S.J., Hutcheson, D.A., and Kardon, G. (2011). Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. Dev. Camb. Engl. *138*, 3625–3637.

76. Ni, Z., Olsen, J.B., Emili, A., and Greenblatt, J.F. (2011). Identification of mammalian protein complexes by lentiviral-based affinity purification and mass spectrometry. Methods Mol. Biol. Clifton NJ *781*, 31–45.

77. O'Brien, S.K., Knight, K.L., and Rana, T.M. (2012). Phosphorylation of histone H1 by P-TEFb is a necessary step in skeletal muscle differentiation. J. Cell. Physiol. *227*, 383–389.

78. Olguin, H.C., and Olwin, B.B. (2004). Pax-7 up-regulation inhibits myogenesis and cell cycle progression in satellite cells: a potential mechanism for self-renewal. Dev. Biol. *275*, 375–388.

79. 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. EMBO J. *30*, 494–509.

80. Pallafacchina, G., François, S., Regnault, B., Czarny, B., Dive, V., Cumano, A., Montarras, D., and Buckingham, M. (2010). An adult tissue-specific stem cell in its niche: a gene profiling analysis of in vivo quiescent and activated muscle satellite cells. Stem Cell Res. *4*, 77–91.

81. Pasut, A., Oleynik, P., and Rudnicki, M. (2012). Isolation of Muscle Stem Cells by Fluorescence Activated Cell Sorting Cytometry. In Myogenesis, J.X. DiMario, ed. (Humana Press), pp. 53–64.

82. Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. Genes Dev. *12*, 755–762.

83. Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

84. Pirngruber, J., Shchebet, A., and Johnsen, S.A. (2009). Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. Cell Cycle Georget. Tex *8*, 3636–3642.

85. Sambasivan, R., Yao, R., Kissenpfennig, A., Wittenberghe, L.V., Paldi, A., Gayraud-Morel, B., Guenou, H., Malissen, B., Tajbakhsh, S., and Galy, A. (2011). Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. Dev. Camb. Engl. *138*, 3647–3656.

86. Santolini, M., Sakakibara, I., Gauthier, M., Ribas-Aulinas, F., Takahashi, H., Sawasaki, T., Mouly, V., Concordet, J.-P., Defossez, P.-A., Hakim, V., et al. (2016). MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis. Nucleic Acids Res.

87. 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. Mol. Cell. Biol. *17*, 1010–1026.

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

89. Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat. Protoc. *7*, 256–267.

90. Simone, C., Stiegler, P., Bagella, L., Pucci, B., Bellan, C., De Falco, G., De Luca, A., Guanti, G., Puri, P.L., and Giordano, A. (2002). Activation of MyoD-dependent transcription by cdk9/cyclin T2. Oncogene *21*, 4137–4148.

91. Tapscott, S.J., Lassar, A.B., and Weintraub, H. (1992). A novel myoblast enhancer element mediates MyoD transcription. Mol. Cell. Biol. *12*, 4994–5003.

92. Thayer, M.J., Tapscott, S.J., Davis, R.L., Wright, W.E., Lassar, A.B., and Weintraub, H. (1989). Positive autoregulation of the myogenic determination gene MyoD1. Cell *58*, 241–248.

93. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

94. Wang, J., and Abate-Shen, C. (2012). The MSX1 homeoprotein recruits G9a methyltransferase to repressed target genes in myoblast cells. PloS One *7*, e37647.

95. Wood, W.M., Etemad, S., Yamamoto, M., and Goldhamer, D.J. (2013). MyoD-expressing progenitors are essential for skeletal myogenesis and satellite cell development. Dev. Biol. *384*, 114–127.

96. Xu, P.X., Zheng, W., Huang, L., Maire, P., Laclef, C., and Silvius, D. (2003). Six1 is required for the early organogenesis of mammalian kidney. Dev. Camb. Engl. *130*, 3085–3094.

97. Yoshida, N., Yoshida, S., Koishi, K., Masuda, K., and Nabeshima, Y. (1998). Cell heterogeneity upon myogenic differentiation: down-regulation of MyoD and Myf-5 generates "reserve cells." J. Cell Sci. *111*, 769–779.

98. Zammit, P.S., Golding, J.P., Nagata, Y., Hudon, V., Partridge, T.A., and Beauchamp, J.R. (2004). Muscle satellite cells adopt divergent fates a mechanism for self-renewal? J. Cell Biol. *166*, 347–357.

99. Zhang, X., Patel, S.P., McCarthy, J.J., Rabchevsky, A.G., Goldhamer, D.J., and Esser, K.A. (2012). A non-canonical E-box within the MyoD core enhancer is necessary for circadian expression in skeletal muscle. Nucleic Acids Res. *40*, 3419–3430.

CHAPTER 5. DISCUSSION

5.1. Six1 and Myod are Key Players in Myogenesis

Six1 is indispensible for embryonic muscle development in the mouse model, and its deficiency is accompanied by defects in the formation of other tissues. Six1 KO mice die at birth and display broad muscle hypoplasia (Laclef et al., 2003a). During muscle regeneration, the expression of Six1 is quickly induced. Our work showed that after 3 days of injury, most of Pax7-positive satellite cells co-express Six1 (Liu et al., 2013c). Recent work has further revealed the necessity of Six1 during muscle regeneration: specific ablation of Six1 in satellite cells results in disrupted muscle regeneration and disturbed homeostasis of satellite cells (Le Grand et al., 2012). Six1 exerts its function in myogenesis partially through its transcriptional regulation of Myod. The expression of Myod in developing limb buds is delayed in Six1 KO mice (Laclef et al., 2003a), and it is also reduced in regenerating muscle when Six1 is conditionally inactivated in satellite cells (Le Grand et al., 2012). We and others showed that Six1 is able to bind the DRR in differentiating myoblasts and the CER in both proliferating and differentiating myoblasts (Le Grand et al., 2012). Therefore, Six1 is functionally upstream of Myod in both embryonic muscle development and adult muscle regeneration.

Myod is well-known for its capacity of converting non-muscle cells into muscle lineage, as a master regulator of myogenesis. Six1 and Myod have been found to function in parallel: they co-regulate the expression of downstream targets. Firstly, the Six1 DNA binding motif, the MEF3 element, is enriched in regulatory elements of Myod and Myog target genes that are induced upon differentiation (Blais et al., 2005). Secondly, nearly 25% of Six1 targets in myoblasts are also bound by Myod. The DNA motif recognized by Myod, the E-box, is also enriched among

Six1-bound targets (Figure 5.1). Thirdly, Myod requires Six1 and Six4 to convert primary fibroblasts into muscle cells (Chakroun et al., 2015; Santolini et al., 2016). Last but not least, Six4, another member of the Six family, and Myod synergistically regulate the expression of many target genes in C2C12 myoblasts, 293T cells or primary fibroblasts (Chakroun et al., 2015; Liu et al., 2010). Six4 and Six1 recognize MEF3 DNA motif and they co-regulate many muscle-specific genes (Spitz et al., 1998). Six1 and Myod also showed synergistic effect on the reporter expression driven by the Myog promoter in HepG2 cells (Figure 5.2).

A

B

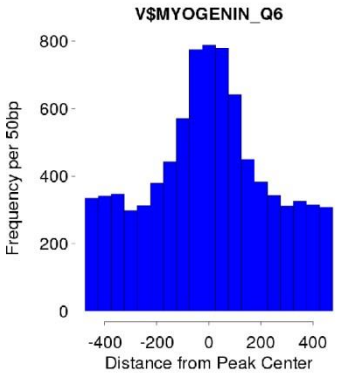
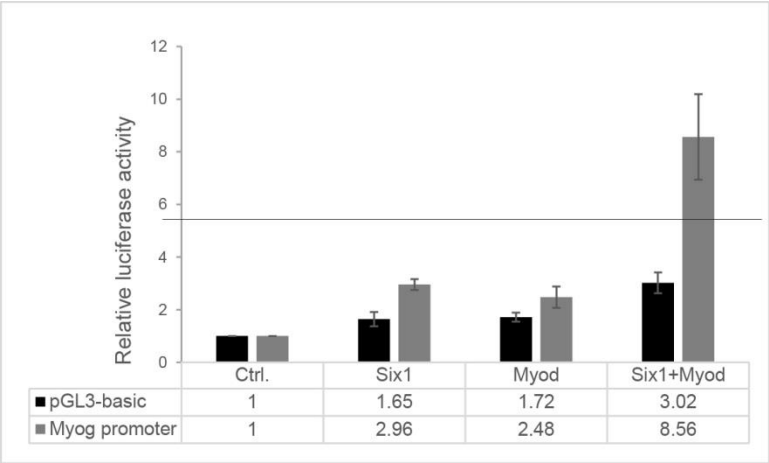

V$MYOGENIN_Q6
800
600
400
200
0
Frequency per 50bp
-400
-200
0
200
400
Distance from Peak Center

Figure 5. 1 The E-box DNA motif is enriched among Six1 binding peaks

A) The DNA motif logo of TRANSFAC "V$MYOGENIN_Q6". B) The E-box DNA motif is enriched at the center of 26255 Six1 binding peaks. X-axis is the distance from the peak summit (0). Y-axis is the frequency of E-box motif per 50 bp using Cendist (http://biogpu.ddns.comp.nus.edu.sg/~chipseq/webseqtools2/).



	Ctrl.	Six1	Myod	Six1+Myod
pGL3-basic	1	1.65	1.72	3.02
Myog promoter	1	2.96	2.48	8.56

Figure 5. 2 Six1 and Myod synergistically activate the Myog-promoter reporter in HepG2 cells

pGL3-basic and Myog promoter reporter plasmids were co-transfected with p3XFlag-Six1-Myc, or p3XFlag-Myod-Myc, or both plasmids together. The renilla luciferase reporter pRL-CMV was used as the internal transfection control. The firefly luciferase activity in each condition was normalized to the baseline level without over-expression plasmid. The grey line indicates the predicted additive effect of each protein alone.

Six1 is a homeodomain transcription factor that binds to DNA directly. Typical homeodomain proteins are characterized by their capacity to recognize the "TAAT" DNA motif (Kornberg, 1993). The α3-helix of the homeodomain contacts the major groove of DNA and three basic residues in the N-terminal arm are required to bind to the minor groove of DNA (Ades and Sauer, 1995). However, Six1 lacks these three basic residues in its N-terminal arm, suggesting other regions of Six1 may participate in DNA binding (Patrick et al., 2013). The crystal structure of Six1 and its co-activator Eya2 has been resolved. It is suggested that the N-terminal arm of Six1 interacts with Eya2 (Patrick et al., 2013). The α6-helix of Six-domain contains five basic residues and has the potential to contact the major groove of DNA, which could cooperate with the homeodomain to recognize the MEF3 DNA motif that differs from the "TAAT" motif (Patrick et al., 2013). Using quantitative gel shift assay, we demonstrated that the full length of Six1 shows superior binding affinity to the homeodomain alone, further indicating that regions outside the homeodomain is indispensable for DNA binding.

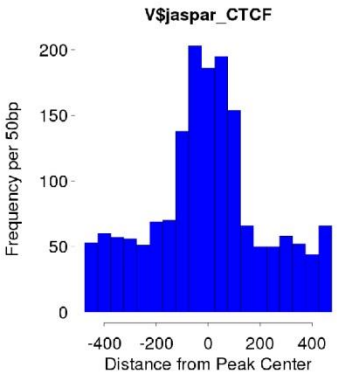
The reported MEF3 motif from the TRANSFAC database was derived from five regulatory regions of muscle-related genes, typified with a stringent selection of DNA sequences, TCAGGTTTC. We have reported the binding profile of Six1 by ChIP-on-Chip, with 1022 sites identified in C2C12 myoblasts and 1853 sites in myotubes. Rich information of Six1 binding sites provides a broad representation of Six1-prefered DNA sequences. Six1_MB+MT, TCANNTTA/TC, calculated from genomic binding data, is more flexible in the middle positions and is more sensitive in localizing Six1 binding sites. For example, the CER of Myod contains two putative MEF3 sites when searched with Six1_MB+MT, which would have been missed with the TRANSFAC MEF3. Six1 binding on these two MEF3 sites was validated with *in vitro* EMSA. It is clear that the CER has two, and only two, Six1 binding sites. Six1_MB+MT was

further computationally optimized, named Six1_opti, which shows even further improved specificity and sensitivity. The diversity of nucleotides in the MEF3 motif was evaluated extensively *in vitro* by fluorescent EMSA using recombinant Six1 protein. By comparing the binding affinity of Six1 to variant MEF3 sequences, EMSA greatly reflects the property of the MEF3 motif generated *in silico*.

In addition to the MEF3 motif, a few other DNA motifs are also enriched in Six1-bound sites. The co-occurrence of E-box provides the physical feasibility of synergistic function between transcription factors of the Six and MRF families. Similarly, the enrichment of other DNA motifs may suggest a broad cooperation between Six1 and these transcription factors. For example, CTCF, a zinc finger protein that is a key regulator in genome organization, has been shown to modulate muscle development and differentiation, including enhancer-promoter looping. CTCF knockdown mediated by morpholino oligonucleotides (MOs) in zebrafish resulted in somite disorganization and reduced expression of Myod and Myog (Delgado-Olguín et al., 2011). It was reported that CTCF interacts with Myod in C2C12 myoblasts (Delgado-Olguín et al., 2011). CTCF and Myod co-bind a few regulatory regions, such as a p57 enhancer that is 150 kb downstream and the core promoter of Sgca, and co-regulate the expression of target genes (Battistelli et al., 2014; Delgado-Olguín et al., 2011). Interestingly, the Myod CER also contains a putative CTCF binding site, which is between two MEF3 sites (Figure 5.3). It is appealing to propose that CTCF may facilitate looping of the chromatin fiber to bring the enhancer and proximal promoter in close physical proximity, which may in turn enhance the expression of Myod.

A

B


V$jaspar_CTCF
Frequency per 50bp
200
150
100
50
0
-400
-200
0
200
400
Distance from Peak Center

C

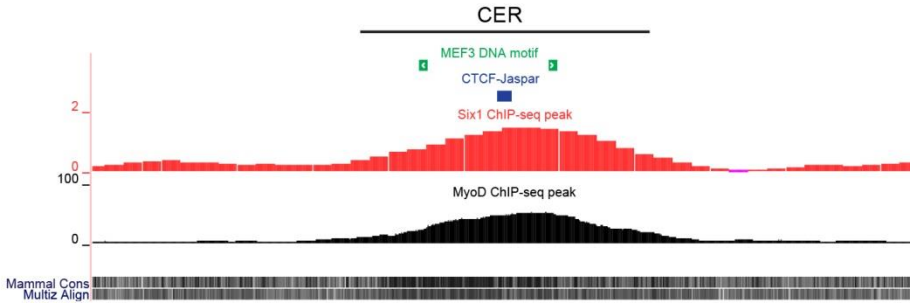

CER
MEF3 DNA motif
CTCF-Jaspar
Six1 ChIP-seq peak
2
0
100
MyoD ChIP-seq peak
0
Mammal Cons
Multiz Align

Figure 5. 3 The Myod CER contains one putative CTCF binding site and two MEF3 sites

A) The DNA motif logo of "V$jaspar_CTCF". B) The CTCF DNA motif is enriched at the center of 26255 Six1 binding peaks. X-axis is the distance from the peak summit (0). Y-axis is the frequency of the CTCF DNA motif per 50 bp using Cendist (http://biogpu.ddns.comp.nus.edu.sg/~chipseq/webseqtools2/). C) A snapshot from the UCSC genome browser. Six1 and Myod ChIP-seq data were uploaded (Six1 in red and Myod in black). The locations of MEF3 DNA motifs are labeled in green. The location of the CTCF DNA motif is labeled in blue. The Y-axis is the normalized ChIP-seq read density. Mammalian conservation is shown at the bottom.

We have mapped 26255 Six1 binding sites in primary myoblasts by ChIP-seq. The number of peaks is a drastic increase from 1022 peaks in C2C12 myoblasts that were identified by ChIP-on-Chip. The difference is not solely a result of cell types, as Six1 ChIP-seq in C2C12 myoblasts yielded a comparable number of peaks and these peaks are highly overlapping (Figure 5.4A). We reason that it is due to different sensitivity of two techniques. Firstly, the majority of ChIP-on-Chip peaks were also identified by ChIP-seq with high confidence. Secondly, Six1_opti, which is computed based on ChIP-on-Chip data, is significantly enriched among Six1 ChIP-seq targets. In addition, the increased resolution afforded by the ChIP-seq method, compared to ChIP-on-Chip, allows more precise location of Six1 binding, resulting in a larger number of high confidence peaks (Figure 5.4B).

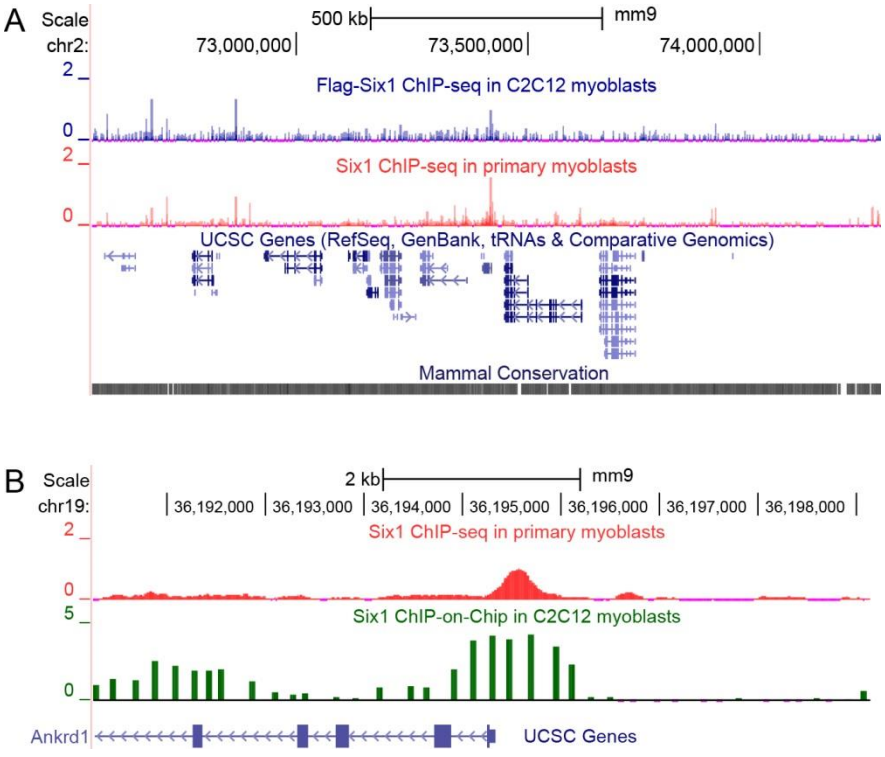

A
Scale
500 kb
mm9
chr2:
73,000,000
73,500,000
74,000,000
2
Flag-Six1 ChIP-seq in C2C12 myoblasts
0
2
Six1 ChIP-seq in primary myoblasts
0
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
Mammal Conservation
B
Scale
2 kb
mm9
chr19:
36,192,000
36,193,000
36,194,000
36,195,000
36,196,000
36,197,000
36,198,000
2
Six1 ChIP-seq in primary myoblasts
0
5
Six1 ChIP-on-Chip in C2C12 myoblasts
0
Ankrd1
UCSC Genes

Figure 5. 4 Comparisons of Six1 binding sites between C2C12 myoblasts and primary myoblasts and between ChIP-seq and ChIP-on-Chip

A) A snapshot from the UCSC genome browser. Six1 ChIP-seq in primary myoblasts and Flag-Six1 ChIP-seq in C2C12 myoblasts were uploaded (Six1 in red and Flag-Six1 in blue). An example region on chromosome 2 is shown here. A list of known genes is also shown. The Y-axis is the normalized ChIP-seq read density. Mammalian conservation is shown at the bottom. B) A snapshot from the UCSC genome browser. Six1 ChIP-seq in primary myoblasts and Six1 ChIP-on-Chip data in C2C12 myoblasts were uploaded (ChIP-seq in red and ChIP-on-Chip in green). The promoter region of Ankrd1 is shown as a representative example.

As a transcription factor, Six1 lacks an evident transactivation domain. It has been reported that Six1 could interact with co-factors to either repress or activate target gene expression in other cell lineages (Bricaud and Collazo, 2011; Brugmann et al., 2004; Li et al., 2003). In C2C12 myoblasts, Six1 was capable of activating the luciferase reporter driven by the Myog promoter (Spitz et al., 1998). Eya proteins are expressed in C2C12 cells and they have been shown to interact with Six1 to regulate the expression of specific targets (Grifone et al., 2004; Li et al., 2003). Eya3 was shown to further recruit co-activator CREB-binding protein CBP, which is a protein acetyltransferase (Ikeda et al., 2002; Li et al., 2003). Dach1 and Dach2 are homologues of Drosophila *dachsund* (*dach*) that plays an essential role in compound eye development in flies (Mardon et al., 1994). Dach1 has been recognized as a co-repressor and was shown to repress Six1-mediated reporter transcription and this repression was reversed in the presence of Eya3 (Li et al., 2003; Wu et al., 2003, 2007). We have reported the gene expression profiling after Six1 knockdown in C2C12 myoblasts and genes directly bound by Six1 could be up-regulated or down-regulated by Six1 knockdown, indicating that Six1 could function as both an activator and a repressor (Liu et al., 2010). However, it is still difficult to differentiate whether the change of gene transcription is a primary effect or secondary effect through a Six1 downstream target. The dual role of Six1 as both a repressor and an activator requires further investigation regarding the composition of Six1-containing protein complexes at a specific developmental stage.

In order to elucidate the function of Six1, we conducted a screen for Six1-binding proteins in C2C12 myoblasts and proteins involved in a few protein complexes were identified, including activator and repressor complexes. There are three SAGA-like complexes in mammalian cells, and they function as chromatin-acetylating transcription co-activators (Martinez et al., 2001). We found that TATA-box binding protein associated factor 9 (TAF9) potentially interacts with Six1.

The SWI/SNF complex is an ATP-dependent nucleosome remodeling complex that contains multiple subunits (Vignali, Hassan, Neely, & Workman, 2000). The SWI/SNF complex is capable of reconstructing the position of nucleosomes in an ATP-dependent manner (Phelan, Sif, Narlikar, & Kingston, 1999). Although the alternative enzymatic subunits, Brg1 and Brm, were not detected in this study, Baf170 and Baf57 were identified as potential Six1-interactors. Interestingly, Brg1 has been shown to directly interact with Six1 and promotes neuronal conversion driven by Six1 and Eya1 (Ahmed et al., 2012b). Certain subunits of the Mediator complex were found to interact with Six1, such as Med1 and Med16. The Mediator complex is a giant platform containing 31 subunits, which is believed to participate in connecting transcription factors with RNA pol II to positively regulate gene expression (Bourbon et al., 2004; Taatjes, 2010). On the other hand, components of co-repressor complexes, such as the Mi-2/NuRD complex, were pulled down with Six1. The NuRD complex is capable of remodeling chromatin in an ATP-dependent mode and deacetylating histones (Xue et al., 1998). The expression of the Chd4 subunit (chromodomain helicase DNA binding protein 4) is highly induced during muscle regeneration and specific inhibition of Chd4 leads to premature muscle differentiation, as exampled by elevated expression of Myog and MHC (myosin heavy chain) (Mammen et al., 2009). It would be interesting to validate the interaction and to test whether Six1 recruits these protein complexes into different DNA loci to modulate the expression of specific genes. Moreover, Six1 may interact with them at specific developmental stages. The broad interaction spectrum of Six1 suggests that it may participate in many steps of transcription regulation, such as chromatin remodeling, transcription initiation, transcription elongation and mRNA processing.

5.2. Who is the Pioneer Factor, Six1 or Myod?

Gene transcription happens in the context of chromatin. DNA and histone proteins constitute nucleosomes which are further condensed by linker histones. Nucleosomes act as a barrier to transcription and nucleosomal DNA tends to be a poor substrate for most transcription factors (Sekiya et al., 2009; Zaret and Mango, 2016). How transcription factors gain access to nucleosomes has been a subject undergoing intense study. To decipher the relationship between chromatin change and the binding of Six1 and Myod, three models are proposed here (Figure 5.5). In the first model, Six1 binds to compact chromatin, serving a similar role as pioneer transcription factors, which in turn leads to chromatin remodelling and facilitates the binding of Myod and other transcription factors. Eventually, the gene activation is achieved through functional transcription elongation. In the second model, the roles are inverted and Myod is the pioneer transcription factor. In the third model, the co-binding of Six1 and Myod coordinately induce chromatin structure change. The three models are not mutually exclusive and may be applied at different genomic loci.

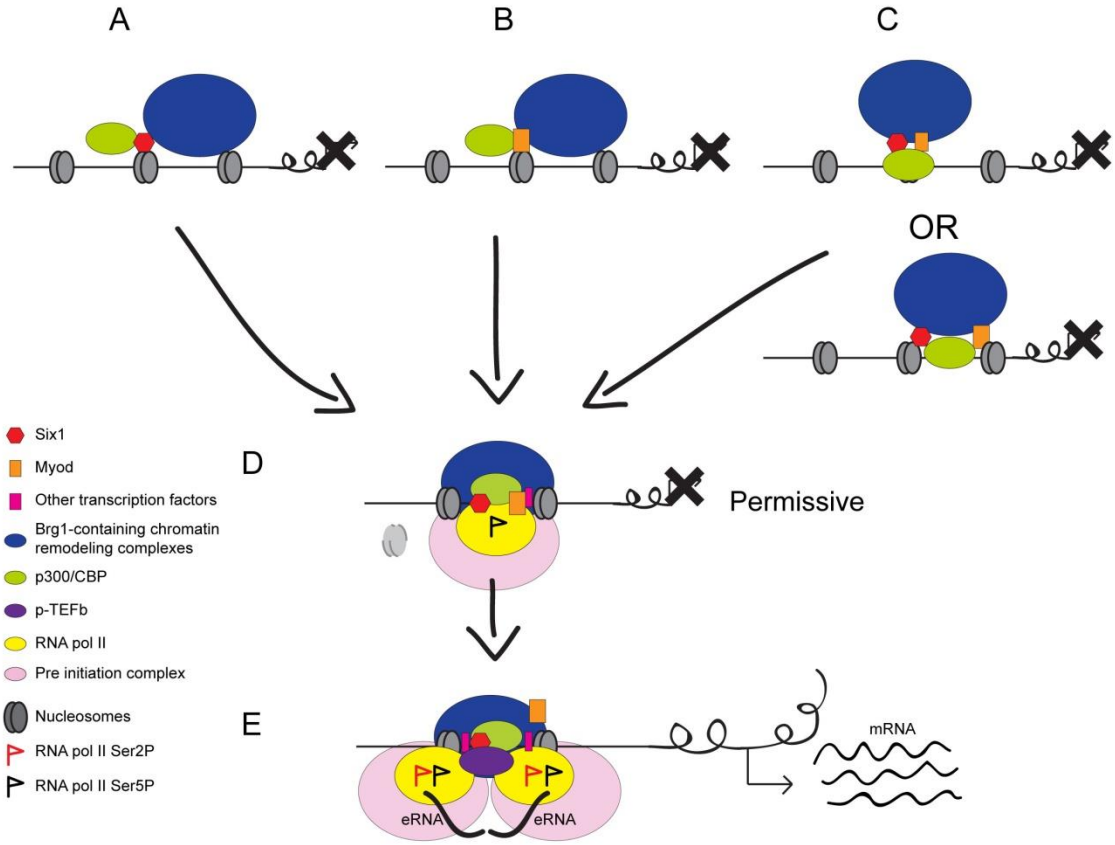

A
B
C
OR
Six1
Myod
Other transcription factors
Brg1-containing chromatin remodeling complexes
p300/CBP
p-TEFb
RNA pol II
Pre initiation complex
Nucleosomes
RNA pol II Ser2P
RNA pol II Ser5P
D
Permissive
E
mRNA
eRNA
eRNA

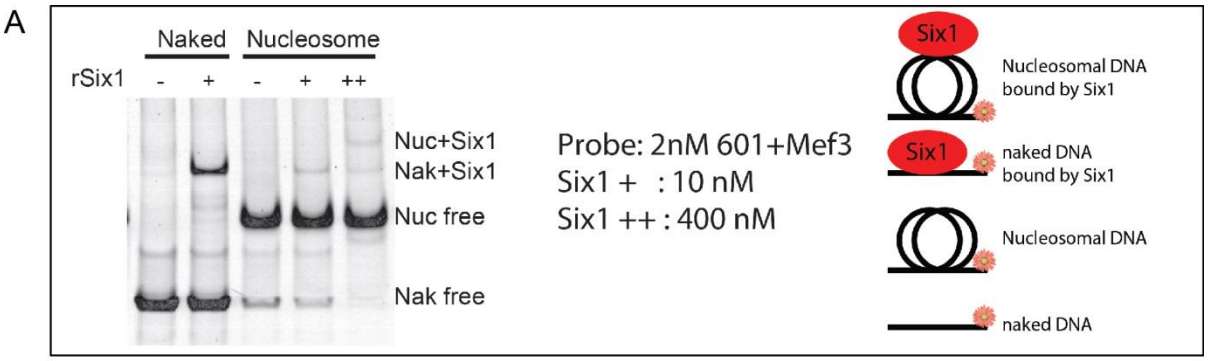
Figure 5. 5 Proposed models of Six1 transcriptional regulation

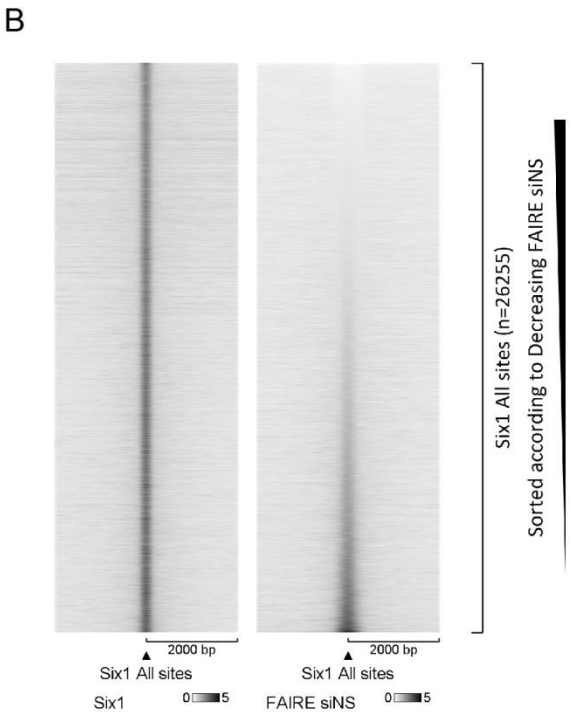
A) Six1 binds to DNA sequences on nucleosomes and recruits chromatin modifiers. B) Myod binds to nucleosomal DNA and recruits chromatin modifiers. C) Six1 and Myod cooperatively bind to DNA sequences on nucleosomes and together recruit chromatin modifiers. D) The consequent chromatin opening facilitates the recruitment of other transcription factors and RNA pol II. E) Cdk9 is recruited to the sites to promote productive gene transcription.

Pioneer factors are a subclass of master regulators that are fundamental to cell fate decisions. Pioneer factors are characterized by their intrinsic ability to recognize full or partial DNA sequences on nucleosomes, where they mark these silent loci for cooperative binding of a collective of regulatory factors to induce changes in gene expression (Iwafuchi-Doi and Zaret, 2016; Zaret and Mango, 2016). For instance, FoxA1 can gain access to nucleosomal DNA *in vitro* and to silent, liver-specific genes *in vivo* during embryonic development (Gualdi et al., 1996; Iwafuchi-Doi and Zaret, 2016; Sekiya et al., 2009; Zaret et al., 2010). FoxA1 could displace linker histone H1 through their structure similarity and maintain the enhancer DNA accessible for other liver-specific transcription factors to bind (Zaret et al., 2010).

Is Six1 a pioneer factor in muscle precursors? Firstly, Six1 has the potenital to bind to nucleosomal DNA *in vitro* and *in vivo*. Using EMSA, Six1 is able to shift a MEF3 containing DNA that was artificially wrapped on the reconstituted histone core (Figure 5.6A). Out of 26255 Six1 peaks in primary myoblasts, a large proportion of loci resides in "closed" chromatin as determined by FAIRE-seq, suggesting that, at least in principle, Six1 should be able to access to compact chromatin *in vivo* (Figure 5.6B). Secondly, Six1 does not seem capable of binding DNA during mitosis, termed as mitotic bookmarking (Figure 5.6C). Mitotic bookmarking has been reported as a potential mechanism of transmitting the transcriptional program by several transcription factors, including some pioneer factors such as Sox2, Oct4, Gata1 and Foxa1 (Caravaca et al., 2013; Kadauke et al., 2012; Teves et al., 2016). However, it is not a property of some predefined pioneer factors, such as Foxo1 (Hatta and Cirillo, 2007; Teves et al., 2016; Yalley et al., 2016). Moreover, the lack of mitotic bookmarking of Six1 could be a fixation artefact, as suggested by a comparison between live cell imaging and imaging after fixation (Teves et al., 2016). Thirdly, Six1 seems to have an impact on global chromatin structure in

myoblasts. Knockdown of Six1 in myoblasts resulted in less accessible chromatin at about 30% of Six1 and Myod co-bound sites. Fourthly, Six1 is required for Myod binding at the CER and may potentiate other regulatory regions which wait to be confirmed. Therefore, combined evidence suggests that Six1 may have the potential to be a pioneer factor.


A
Naked
Nucleosome
rSix1
- + - + ++
Nuc+Six1
Nak+Six1
Nuc free
Nak free
Probe: 2nM 601+Mef3
Six1 + : 10 nM
Six1 ++ : 400 nM
Six1
Nucleosomal DNA bound by Six1
Six1
naked DNA bound by Six1
Nucleosomal DNA
naked DNA


B
Six1 All sites (n=26255)
Sorted according to Decreasing FAIRE siNS
2000 bp
Six1 All sites
Six1
0 5
2000 bp
Six1 All sites
FAIRE siNS
0 5

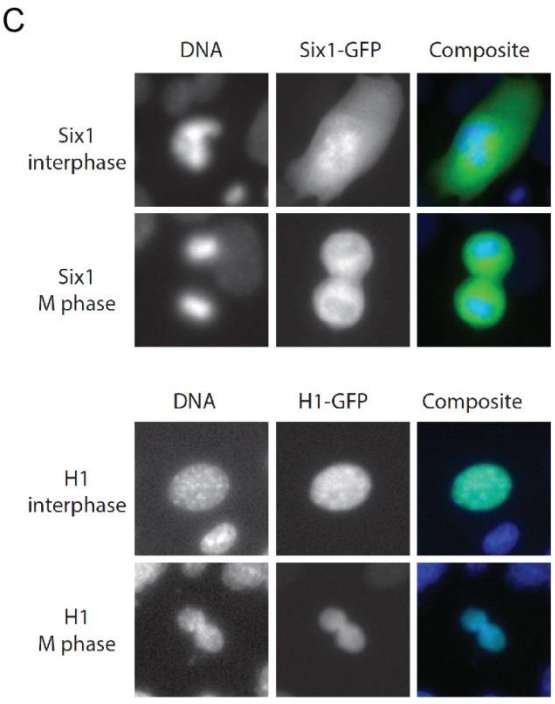

C
DNA
Six1-GFP
Composite
Six1 interphase
Six1 M phase
DNA
H1-GFP
Composite
H1 interphase
H1 M phase

Figure 5. 6 Six1 has the potential to bind DNA sequences on nucleosomes

A) Electrophoretic mobility shift assay was performed to examine the binding of Six1 to nucleosomal DNA. Artificial nucleosomes were assembled with recombinant core histones and fluorescent 601-DNA inserted with MEF3 sequence (601+MEF3). Core histones and recombinant Six1 were purified from E.coli. 601-MEF3 probe was generated from PCR and labeled with Cy5. Increasing amount of Six1 was used to test its binding to either naked 601-MEF3 DNA or nucleosomal 601-MEF3 DNA. Six1 is able to bind naked DNA and DNA assembled on nucleosomes. B) The read density of Six1 ChIP-seq and FAIRE-seq in myoblasts is shown for all 26255 Six1 binding peaks in a 4000 bp window centered at the peak summits. Peaks were sorted according to average FAIRE signals in a 400 bp window centered at peak summits. The colorimetric scale is shown at the bottom. C) Six1 or histone H1 coding sequences were cloned into a GFP fusion protein expression plasmid. Six1-GFP and H1-GFP coding plasmids were transfected into C2C12 myoblasts treated with nocodazole. Examples of cells in interphase and M phase are shown.

As mentioned above, Six1 may interact with the SWI/SNF protein complex. Six1 has been shown to directly interact with Brg1 and the two proteins cooperatively drive neurogenesis in fibroblasts (Ahmed et al., 2012b). Brg1 participates in establishing accessible chromatin structure at enhancers. In leukemia cells, Brg1 was shown to bind an enhancer region of Myc and was required for occupancy of lineage-specific TFs (Shi et al., 2013). During mesodermal differentiation of embryonic stem cells, it has been reported that Brg1 was involved in the establishment of differentiation-induced enhancers, marked by an active enhancer mark H3K27ac, suggesting that Brg1 has an impact on enhancer function (Alexander et al., 2015). In addition, the enzymatic function of Brg1 was necessary for a subset of Myod-induced genes during myogenic conversion in fibroblasts (Serna et al., 2005). Moreover, we found that Brg1 was able to bind to the CER in C2C12 myoblasts, indicating a possible involvement of Brg1 in regulating the expression of Myod (Figure 5.7). It would be interesting to test whether Six1 recruits Brg1 to the CER and potentially to other active enhancers, especially of which the chromatin accessibility was lowered by Six1 knockdown.

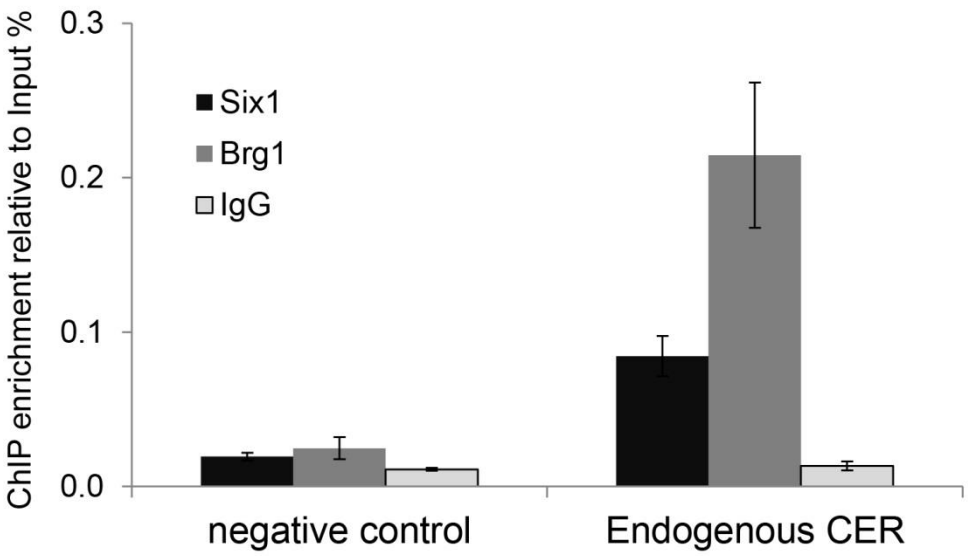

ChIP enrichment relative to Input %
0.3
0.2
0.1
0.0
Six1
Brg1
IgG
negative control
Endogenous CER

Figure 5. 7 Brg1 binds to the CER of Myod in C2C12 myoblasts

ChIP assays were performed in C2C12 myoblasts using Six1 and Brg1 antibodies. Normal rabbit IgG was used as a negative control. The negative control region is a gene desert on chromosome 15. Endogenous CER is the Myod core enhancer region. n = 3.

Myod is a potent master regulator that can reprogram a variety of cells to myogenic lineage, inducing the expression of silent muscle-specific genes in non-muscle cells (Weintraub et al., 1989). It has been a debated question how Myod gains access to compact chromatin and potentially remodels chromatin permanently to sustain the myogenic program (Gerber et al., 1997). Myod was shown to interact with many protein modifiers and chromatin remodelers, such as p300, Baf60c, Brg1 and Cdk9. p300, E1A binding protein p300, directly associates with and acetylates Myod, which potentiates the transactivation ability of Myod (Polesskaya et al., 2000; Sartorelli et al., 1997). p300 can relax the chromatin structure by acetylating core histones and enhance transcription as a co-activator that recruits the basal transcriptional machinery (Jin et al., 2011; Tropberger et al., 2013). In addition, Myod forms a protein complex with Baf60c, a subunit of SWI/SNF complexes, on regulatory regions of silent Myod target genes (Forcales et al., 2012; Serna et al., 2005). The phosphorylation of Baf60c by p38α kinase promotes the recruitment of the Brg1-containing SWI/SNF complex to Myod target genes, which in turn leads to chromatin remodelling and gene activation (Forcales et al., 2012). Furthermore, Myod has been shown to interact with Cdk9 directly and they co-occupy regulatory elements of muscle-specific genes to stimulate their expression, indicating a possibility of Myod-dependent recruitment of Cdk9 to specific loci (Giacinti et al., 2006; Simone et al., 2002).

However, knockdown of Myod in C2C12 myoblasts only resulted in mild change on chromatin structure, suggesting that Myod could be dispensable for maintaining accessible chromatin on most of target genes. Of note, the C2C12 model used here has intrinsic myogenic capacity and many of myogenic regulatory elements are already programmed to “opened” states. It remains unclear whether Myod is required for decompacting silent DNA loci during satellite cell activation. In addition, Myod-mediated myogenic conversion requires the presence of Six1 and

Six4 transcription factors, at least in fibroblasts, suggesting that Six transcription factors are prerequisite for myogenic reprogramming (Chakroun et al., 2015; Santolini et al., 2016). The requirement of Six1 and Six4 in myogenic conversion challenges Myod as a pioneer factor. Moreover, Six1 and Six4, as potentially obligatory factors, are widely expressed in many cell types that have been shown to be compatible with Myod-mediated myogenic conversion, such as fibroblasts, smooth muscle and adipocytes (Choi et al., 1990; Weintraub et al., 1989). Reprogramming is less efficient between cell types derived from different germ layers, generally considered as a consequence of inherent chromatin structure, transmissible histone marks and available co-factors (Vierbuchen and Wernig, 2011). For example, HepG2 is a human liver carcinoma cell line that is refractory to Myod-mediated myogenic reprogramming (Pomerantz et al., 2009; Weintraub et al., 1989). HepG2 cells have very limited expression of Six1 and Eya3, without detectable expression of other Six and Eya members (Figure 5.8). It would be interesting to evaluate whether ectopic expression of Six1 and Eya3 could cooperate with Myod to induce the expression of muscle-specific genes in conversion-resistant cells.

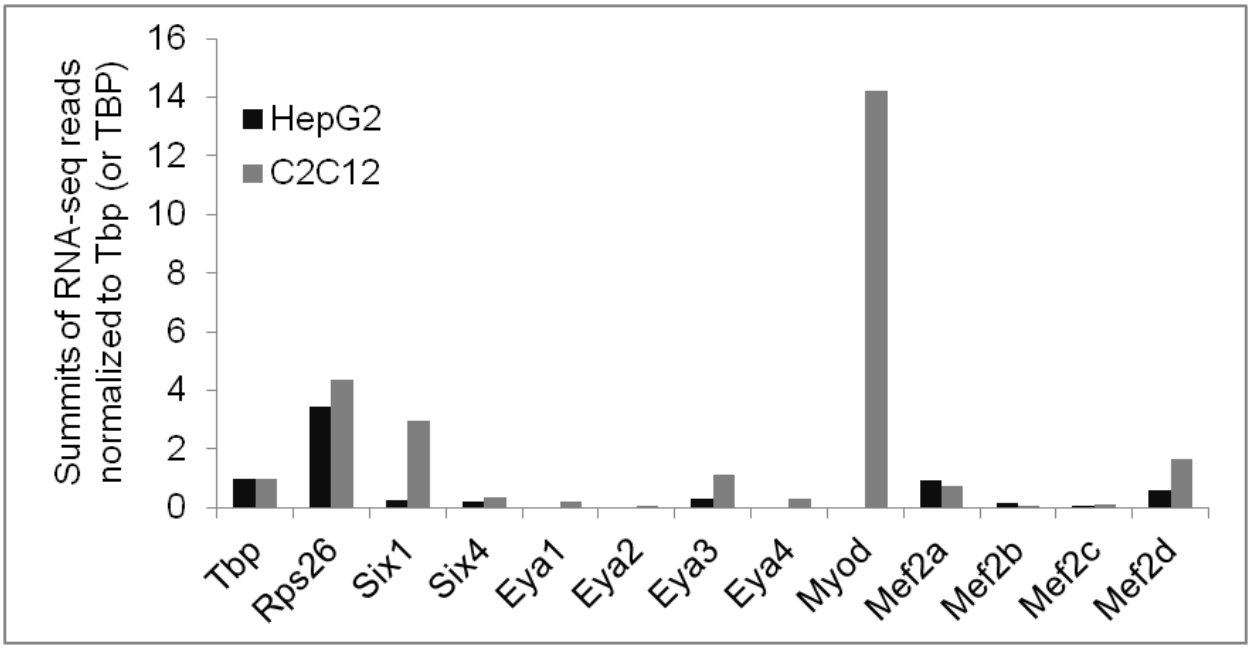

Summits of RNA-seq reads normalized to Tbp (or TBP)
16
14
12
10
8
6
4
2
0
HepG2
C2C12
Tbp
Rps26
Six1
Six4
Eya1
Eya2
Eya3
Eya4
Myod
Mef2a
Mef2b
Mef2c
Mef2d

Figure 5. 8 The summits of RNA-seq reads at different genes in C2C12 myoblasts and HepG2 cells

RNA-seq data in C2C12 myoblasts and HepG2 cells from ENCODE/Caltech were visualized using UCSC genome browser. The summit of RNA-seq reads at each gene was normalized to that at Tbp of the respective cell line.

Similar to Six1, Myod binds to a large number of genomic loci, 23271 peaks in myoblasts (Cao et al., 2010). A proportion of these peaks display low chromatin accessibility or even compact chromatin, especially at Myod target loci without Six1 binding nearby. It cannot be excluded that the binding of transcription factors on closed chromatin as detected by ChIP is a result of fixation artefact or mediated by unknown co-factors. Artificially tethered Myod/E47 heterodimer has been shown to induce the expression of a reporter harbouring the CKM (creatine kinase M-type) minimal promoter that is assembled on nucleosome arrays (Dilworth et al., 2004). However, there is no direct evidence showing that Myod can access to nucleosomal DNA by itself. In addition, it has not been tested whether Myod has the property of mitotic bookmarking. Taken together, evidence suggests that Myod is essential to execute the myogenic conversion program by inducing the expression of muscle-specific genes, but it may not be the pioneer factor that recognizes silent genomic loci at the first places.

Enhancers often contain a cluster of transcription factor binding sites and combinatorial binding of selective transcription factors contributes to precise regulation of tissue-specific genes (Chen et al., 2008; Natoli, 2010; Spitz and Furlong, 2012). It is appealing to propose that the collaborative binding of Six1 and Myod integrates intracellular and extracellular signals to precisely modulate the transcription of their target genes. A defined set of transcription factors is capable of reprogramming fibroblasts into pluripotent stem cells, while either alone is incapable of doing so, suggesting that a compound effect of these transcription factors together initiates the reprogramming (Maherali et al., 2007; Okita et al., 2007; Takahashi and Yamanaka, 2006; Wernig et al., 2007). Further studies demonstrated that the reprogramming is accompanied by massive chromatin structure changes and histone mark rewriting, indicating that these transcription factors cooperatively recruit chromatin remodelling complexes and other co-factors

to induce specific gene activation and repression (Gaspar-Maia et al., 2011; Guenther et al., 2010; Tao et al., 2014; Watanabe et al., 2013). Similarly, myogenic conversion demands the presence of both Six and MRF proteins. While the function of Myod is restricted in the muscle lineage, Six1 is implicated in the development of other tissues, such as kidneys and inner ears, suggesting Six1 may cooperate with other lineage-specific transcription factors.

In primary myoblasts, Six1 and Myod co-occupy 5804 distal DNA loci that display features of active enhancers, such as enrichment of H3K27ac, H3K4me1, and RNA pol II. Six1 and Myod also co-bind 1553 proximal regions that exhibit bimodal distribution of H3K3me3. Overall, around 30% of Myod targets were also bound by Six1 within 1000 bp. Without a protein-protein interaction between Six1 and Myod detected (negative data not shown), Six1 and Myod likely synergistically regulate target genes through changing the structure of chromatin or regulating various steps of the transcriptional program. Interestingly, over 60% of Six1 and Myod common targets have the binding of two transcription factors within 150 bp. The binding proximity points to the physical feasibility of cooperation between the two transcription factors. In fact, we and others have reported the synergistic effect on gene regulation between Six and MRF families (Chakroun et al., 2015; Liu et al., 2010; Santolini et al., 2016). Taken together, evidence indicates that Six1 and Myod may collaboratively bind to specific targets to induce the maximum transcriptional output upon stimuli.

Six1 and Myod may concomitantly bind to certain genomic loci. However, Six1 is genetically upstream of Myod during embryonic muscle development and Six1 is broadly expressed in many cell types that have propensity to be converted to muscle cells, including fibroblasts (Kawakami et al., 2000; Laclef et al., 2003a, 2003b; Relaix et al., 2013). Current evidence favours a model that Six1, as a pioneer factor, marks a large set of genomic loci for Myod to bind. NeuroD, a

bHLH transcription factors, is a master regulator in the neurogenic lineage. Six1 and NeuroD have been shown to cooperate in driving neurogenesis (Ahmed et al., 2012b). Considering the essential role of Six1 in the development of both muscles and sensory neurons, it is reasonable to propose that lineage-specific transcription factors are recruited by Six1, through direct or indirect interactions, to drive the expression of lineage-specific target genes (Schlosser et al., 2008; Zou et al., 2004).

It is possible that Myod can have an influence on Six1 expression as a forward-loop. In $Myod^{-/-}$ /$Myf5^{-/-}$ embryonic fibroblasts, forced expression of Myod induces Six1 expression by 3-4 fold and the induction depends on the intact α-helix domain in the carboxy terminal region of Myod (Berkes and Tapscott, 2005; Ishibashi et al., 2005). Similarly, forced expression of Myod in P19 embryonic carcinoma cells also induces Six1 expression, and dominant negative Myod abolishes the enhanced expression of Six1 (Gianakopoulos et al., 2011). However, Myod does not bind directly to the promoter region of Six1 in C2C12 and primary myoblasts (Cao et al., 2010). The regulation of Six1 expression by Myod could be through a plausible enhancer, located downstream of the Six1 gene, or through post-transcriptional mechanisms. Six1 could also be modulated by a factor that is a target of Myod. There are two Myod ChIP-seq peaks downstream of the Six1 gene, but further investigation is required to determine whether these binding events are directly associated with transcription of the Six1 gene.

5.3. Enhancers and Cdk9

Recent work has revealed that many active enhancers are transcribed by RNA pol II, generating unstable and bidirectional short eRNAs. In fact, eRNA production has been recognized as an independent mark of functional enhancers. Transcription pausing at proximal regions has been

appreciated as a general mechanism of transcription regulation, by which nearly 30% of genes in flies and mammalian cells are modulated (Adelman and Lis, 2012; Giacinti et al., 2006; Sigova et al., 2015). Pausing of RNA pol II has been shown to preclude the nucleosome formation and maintain accessible chromatin structure at target promoters, in order to facilitate gene activation upon stimulus (Gilchrist et al., 2010). The act of transcription, through RNA pol II-associated protein complexes, helps to maintain chromatin in a nucleosome-depleted state at proximal regions, and perhaps distal regions as well (Barton and Crowe, 2001; Lee et al., 2015; Stadhouders et al., 2012a).

The Price group set the groundwork that transcriptional elongation is mediated by p-TEFb which is composed of the kinase subunit Cdk9 and the cyclin T regulatory subunit (Marshall et al., 1996; Peterlin and Price, 2006). Studies have focused on the transcription pausing at promoter-proximal regions, while it is not clear whether functional enhancers are controlled by the same manner (Zhou et al., 2012). Some labs have found little evidence of Cdk9 activity at enhancers, while other labs saw clear signs of it. A transcription elongation mark, RNA pol II-Ser2P, has been shown to accumulate at promoter-proximal regions but not at enhancers in T-cells (Descostes et al., 2014; Koch et al., 2011). By contrast, the transcription elongation at stimulus-activated enhancers is regulated by p-TEFb, as inhibition of the p-TEFb function by Flavopiridol results in reduced eRNA transcripts in macrophage cells (Kaikkonen et al., 2013). The adenosine analogue 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB), a Cdk9 kinase inhibitor, inhibits RNA pol II-mediated transcription elongation similarly to Flavopiridol (Bensaude, 2011). DRB treatment in murine embryonic stem cells leads to reduced transcription at both promoters and enhancers, suggesting Cdk9 function is required for both mRNA and eRNA production (Sigova et al., 2015). Cdk9 has been shown to participate in releasing transcription-pausing at

both promoters and enhancers through its kinase function (Giacinti et al., 2006; Peterlin and Price, 2006; Pirngruber et al., 2009; Sigova et al., 2015). In addition, our analysis shows that RNA pol II-Ser2P is enriched among distal regions that were co-bound by Six1 and Myod in myoblasts (Figure 5.9). Taken together, evidence suggests that transcription pausing could be a mechanism that controls the function of a subset of enhancers in myogenic cells.

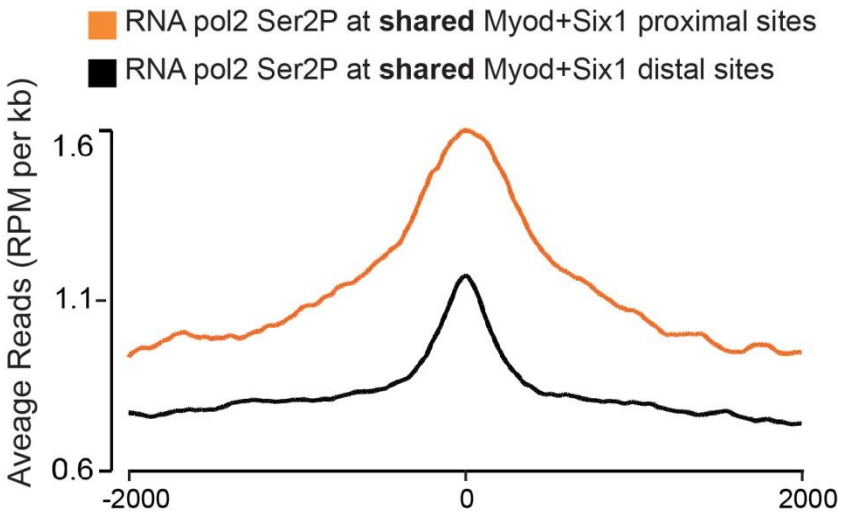

RNA pol2 Ser2P at **shared** Myod+Six1 proximal sites
RNA pol2 Ser2P at **shared** Myod+Six1 distal sites
Aveage Reads (RPM per kb)
1.6
1.1
0.6
-2000
0
2000

Figure 5. 9 The RNA pol II Ser2P ChIP-seq signal is enriched at both promoters and enhancers of Myod and Six1 targets

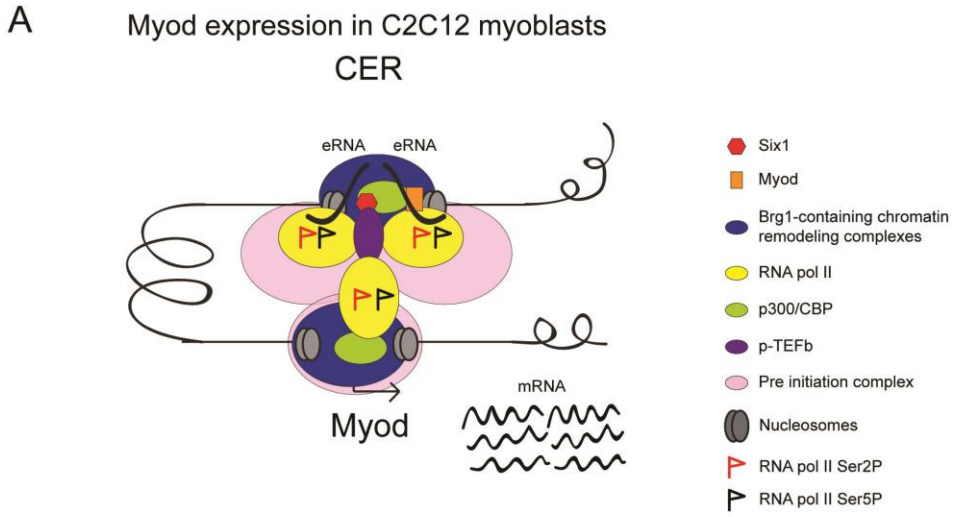
The ChIP-seq data of RNA pol II Ser2P (phosphorylated serine 2 residues of the heptapeptide repeats of RNA pol II) have been reported in C2C12 cells at the early stage of differentiation (GSM628015) (Mousavi et al., 2012). ChIP-seq signals of the RNA pol II Ser2P were represented in a 4000 bp window centered at the peak summits of proximal or distal shared Myod+Six1 sites, using Easeq. Y-axis is the read density normalized to reads per million (RPM) per 1000 bp.

Transcription factors, such as c-Myc, have been proposed to promote the recruitment of p-TEFb to specific genomic loci, mostly at proximal regions (Barboric et al., 2001; Core and Lis, 2008; Kanazawa et al., 2003; Martinez-Fernandez et al., 2009; Peterlin and Price, 2006; Rahl et al., 2010). The function of Cdk9 in stimulating RNA pol II elongation was diminished in the absence of certain transcription factors. In haematological cells ,some distal enhancers have been shown to stimulate transcription elongation by providing Cdk9 to the proximal regions (Jing et al., 2008; Song et al., 2010; Stadhouders et al., 2012b). Brd4 regulates transcription elongation of both coding RNA and enhancer RNA and it has been shown to recruit Cdk9 to specific regulatory elements, such as super-enhancers (Di Micco et al., 2014; Kanno et al., 2014; Yang et al., 2005). However, it has not been directly evaluated whether the recruitment of Cdk9 depends on the presence of these regulatory proteins.

Interestingly, Cdk9 does not only interact with Myod, but also forms a protein complex with Six1. Cdk9 and Myod were shown to co-occupy enhancers and promoters of muscle-specific genes (Giacinti et al., 2006; Simone et al., 2002). Our work found the co-binding of Cdk9, Myod and Six1 at the CER of Myod, but not at the promoter region. The CER is essential for the temporal and spatial expression of Myod during embryonic development (Chen and Goldhamer, 2004; Chen et al., 2005; Goldhamer et al., 1992, 1995). In addition, the reporter induction driven by the CER depends on the presence of Six1 in C2C12 myoblasts. Furthermore, we demonstrated that the recruitment of both Myod and Cdk9 to the CER relied on the binding of Six1. Given the fact that Cdk9 inhibition by Flavopiridol yielded a lower FAIRE signal at the CER, while Myod knockdown had no effect, it is likely that Cdk9 acts prior to MyoD and thus that Six1 is the recruiter of Cdk9 at that locus. However, it requires further investigation whether Cdk9 is recruited directly by Six1 or subsequently by Myod to other genomic loci.

It is well accepted that both Six1 and the CER are essential for the expression of Myod. Of note, knockdown of either Six1 or the eRNA transcript at the CER (^{CE}RNA) diminished the expression of Myod (Mousavi et al., 2013) and (Figure 3.2B). Inhibition of the Cdk9 kinase activity by Flavopiridol led to reduced chromatin accessibility at the CER and decreased Myod expression, which resembles the loss of Six1, suggesting the functional link between Cdk9 and Six1. We propose a model that the CER is presented to the proximal promoter through chromatin looping. The binding of Six1 recruits Cdk9 to the CER, where CDK9 can catalyze the phosphorylation of Serine 2 residues of RNA pol II at both the CER and the nearby Myod promoter, resulting in sustained Myod expression (Figure 5.10). The Myod promoter region is less accessible than the CER and is not enriched for transcription factor binding sites. Therefore, the Myod promoter may provide the assembly of RNA polymerase initiation complex and the CER conveys the elongation signal through residing Cdk9. This model can be tested by examining the accumulation of RNA pol II S2P at the three prime end of the Myod gene after Six1 knock-down, or Flavopiridol treatment.

f


A
Myod expression in C2C12 myoblasts
CER
eRNA
eRNA
mRNA
Myod
Six1
Myod
Brg1-containing chromatin remodeling complexes
RNA pol II
p300/CBP
p-TEFb
Pre initiation complex
Nucleosomes
RNA pol II Ser2P
RNA pol II Ser5P

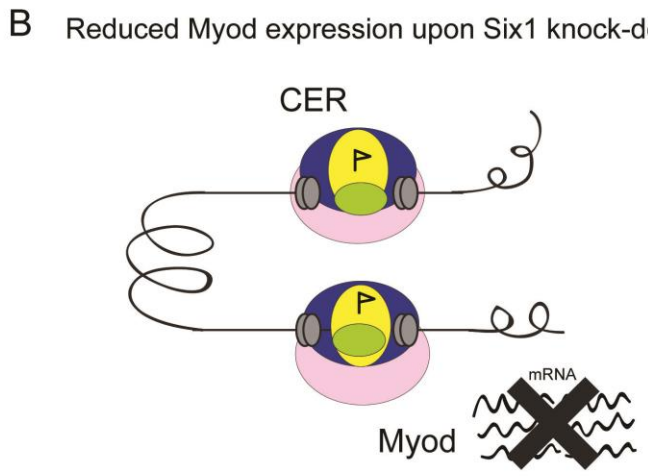

B
Reduced Myod expression upon Six1 knock-down
CER
mRNA
Myod

Figure 5. 10 Proposed model of Six1-mediated Myod expression

A) In C2C12 myoblasts, Six1 binds to the CER and facilitates the binding of Cdk9 and Myod to the CER. Cdk9 phosphorylates Serine 2 residues of the CTD on the largest subunit of RNA pol II, which promotes the transcription of both the CER and the Myod gene. B) When Six1 is knocked-down in C2C12 myoblasts, the recruitment of Cdk9 and Myod to the CER is impaired, resulting in reduced expression of Myod and potentially decreased eRNA production.

The function of Cdk9 is also involved in maintaining the proper chromatin structure at the CER (Figure 4.5C). We reasoned three possible mechanisms. Firstly, the act of active transcription at enhancers may be responsible for its highly accessible chromatin. Elongation-competent RNA pol II has been shown to carry chromatin modifiers through interactions with its CTD, and thus it may have a positive effect on chromatin architecture at both promoters and enhancers (Selth et al., 2010). Secondly, eRNA production may recruit or trap unknown chromatin-determination factors to maintain the CER DNA accessible. As mentioned above, ^{CE}RNA is required for the expression of Myod in myoblasts (Mousavi et al., 2013). Knockdown of eRNA at estrogen-induced enhancers also resulted in reduced transcription of target genes in breast cancer cells (Li et al., 2013a). Furthermore, eRNA transcripts can activate the expression of reporters when they are artificially tethered to the promoters (Lam et al., 2013; Li et al., 2013a; Melo et al., 2013). It was suggested that eRNA may promote the promoter-enhancer looping through binding to the Cohesin complex. Lastly, the proximal RNA polymerase II-associated chromatin modifiers and chromatin remodelers may have an influence on the enhancer chromatin structure.

5.4. Clinical Relevance and Future Applications

Branchio-oto-renal (BOR) syndrome is an autosomal dominant inherited disease, affecting approximately 1 in 40 000 population and 1 in 50 deaf children. Reviewed in (Kochhar et al., 2007). Around 40% BOR patients carry Eya1 mutations and mutations in Six1 are less frequent. To date, six mutations in Six1 have been reported (Chang et al., 2004; Kochhar et al., 2008; Mosrati et al., 2011). Six1 is also over-expressed in many cancers and it is considered as an oncogene (Blevins et al., 2015; Christensen et al., 2008). Six1 could participate in the tumorigenesis through regulating cell cycle-related genes (Christensen et al., 2007; Coletta et al., 2004; Li et al., 2013b), disturbing genome stability (Coletta et al., 2008), and promoting

metastasis (McCoy et al., 2009; Ng et al., 2009; Yu et al., 2006). Recently Six1 and its co-factor Eya1 are considered as putative therapeutic targets in cancer (Blevins et al., 2015). Blocking the interface of protein-protein interaction by mimicking peptides and inhibition of involving enzymatic protein complexes may provide therapy to treat Six1-related diseases. Therefore, identifying other Six1- and/or Eya1- interacting proteins may help to dissect functional-related targets in the diseases. Here we reported Six1 binding profile and Six1-interacting protein spectrum in mammalian myogenic cells, which will enrich our understanding of Six1 function to screen interesting targets. More genomic binding profiles of transcription factor and co-factors, transcriptome and interactome data would provide more information in terms of dynamic Six1 functions.

5.5. Concluding remarks

During my PhD, I have shown that 1) both the SHD and the SD of Six1 contribute to its binding to a degenerate MEF3 DNA motif that contains a flexible center, 2) Six1 regulates the expression of Myod through its direct binding on the CER in muscle progenitor cells and 3) Six1 regulates the expression of Myod through its positive impact on the chromatin structure at the CER which is mediated at least in part by the recruitment of Cdk9. Altogether, my PhD thesis on understanding the DNA binding mode of Six1 and the underlying mechanisms of its transcriptional regulation has improved our knowledge about the regulatory role of Six1 during muscle regeneration as well as indicated future direction for discovering potential therapeutic candidates that target Six1-containing protein complexes.

REFERENCES (FOR CHAPTER 1 AND 5)

1. Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. 13, 720–731.

2. Ades, S.E., and Sauer, R.T. (1995). Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. Biochemistry (Mosc.) 34, 14601–14608.

3. Adrados, I., Larrasa-Alonso, J., Galarreta, A., López-Antona, I., Menéndez, C., Abad, M., Gil, J., Moreno-Bueno, G., and Palmero, I. (2015). The homeoprotein SIX1 controls cellular senescence through the regulation of p16INK4A and differentiation-related genes. Oncogene.

4. Ahmed, M., Wong, E.Y.M., Sun, J., Xu, J., Wang, F., and Xu, P.-X. (2012a). Eya1-Six1 interaction is sufficient to induce hair cell fate in the cochlea by activating Atoh1 expression in cooperation with Sox2. Dev. Cell 22, 377–390.

5. Ahmed, M., Xu, J., and Xu, P.X. (2012b). EYA1 and SIX1 drive the neuronal developmental program in cooperation with the SWI/SNF chromatin-remodeling complex and SOX2 in the mammalian inner ear. Dev. Camb. Engl. 139, 1965–1977.

6. Albini, S., Coutinho, P., Malecova, B., Giordani, L., Savchenko, A., Forcales, S.V., and Puri, P.L. (2013). Epigenetic reprogramming of human embryonic stem cells into skeletal muscle cells and generation of contractile myospheres. Cell Rep. 3, 661–670.

7. Alexander, J.M., Hota, S.K., He, D., Thomas, S., Ho, L., Pennacchio, L.A., and Bruneau, B.G. (2015). Brg1 modulates enhancer activation in mesoderm lineage commitment. Dev. Camb. Engl. 142, 1418–1430.

8. Ando, Z.-I., Sato, S., Ikeda, K., and Kawakami, K. (2005). Slc12a2 is a direct target of two closely related homeobox proteins, Six1 and Six4. FEBS J. 272, 3026–3041.

9. Andrés, V., and Walsh, K. (1996). Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. J. Cell Biol. 132, 657–666.

10. Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rönnerblad, M., Hrydziuszko, O., Vitezic, M., et al. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science 347, 1010–1014.

11. Asangani, I.A., Dommeti, V.L., Wang, X., Malik, R., Cieslik, M., Yang, R., Escara-Wilke, J., Wilder-Romans, K., Dhanireddy, S., Engelke, C., et al. (2014). Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. Nature 510, 278–282.

12. 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 Dev. 20, 2450–2464.

13. Barboric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. Mol. Cell 8, 327–337.

14. Barr, F.G. (2001). Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. Oncogene 20, 5736–5746.

15.Barton, M.C., and Crowe, A.J. (2001). Chromatin alteration, transcription and replication: What's the opening line to the story? Oncogene 20, 3094–3099.

16. Battistelli, C., Busanello, A., and Maione, R. (2014). Functional interplay between MyoD and CTCF in regulating long-range chromatin interactions during differentiation. J. Cell Sci. 127, 3757–3767.

17. Bell, O., Tiwari, V.K., Thoma, N.H., and Schubeler, D. (2011). Determinants and dynamics of genome accessibility. Nat. Rev. 12, 554–564.

18. Bennicelli, J.L., Edwards, R.H., and Barr, F.G. (1996). Mechanism for transcriptional gain of function resulting from chromosomal translocation in alveolar rhabdomyosarcoma. Proc. Natl. Acad. Sci. U. S. A. 93, 5455–5459.

19. Bensaude, O. (2011). Inhibiting eukaryotic transcription. Transcription 2, 103–108.

20. Bentzinger, C.F., Wang, Y.X., von Maltzahn, J., Soleimani, V.D., Yin, H., and Rudnicki, M.A. (2013). Fibronectin regulates Wnt7a signaling and satellite cell expansion. Cell Stem Cell 12, 75–87.

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

22. 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. Mol. Cell 14, 465–477.

23. Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol. 5, R62.

24. Bismuth, K., and Relaix, F. (2010). Genetic regulation of skeletal muscle development. Exp. Cell Res. 316, 3081–3086.

25. Bjornson, C.R., Cheung, T.H., Liu, L., Tripathi, P.V., Steeper, K.M., and Rando, T.A. (2012). Notch signaling is necessary to maintain quiescence in adult muscle stem cells. Stem Cells Dayt. Ohio 30, 232–242.

26. Blachly, J.S., and Byrd, J.C. (2013). Emerging drug profile: cyclin-dependent kinase inhibitors. Leuk. Lymphoma 54, 2133–2143.

27. Blackwell, T.K., and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250, 1104–1110.

28. Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., and Birchmeier, C. (1995). Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. Nature 376, 768–771.

29. Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B.D. (2005). An initial blueprint for myogenic differentiation. Genes Dev. 19, 553–569.

30. Blanco-Bose, W.E., Yao, C.C., Kramer, R.H., and Blau, H.M. (2001). Purification of mouse primary myoblasts based on alpha 7 integrin expression. Exp. Cell Res. 265, 212–220.

31. Blevins, M.A., Towers, C.G., Patrick, A.N., Zhao, R., and Ford, H.L. (2015). The SIX1-EYA transcriptional complex as a therapeutic target in cancer. Expert Opin. Ther. Targets 19, 213–225.

32. Blum, R., Vethantham, V., Bowman, C., Rudnicki, M., and Dynlacht, B.D. (2012). Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. Genes Dev. 26, 2763–2779.

33. Bonini, N.M., Leiserson, W.M., and Benzer, S. (1998). Multiple roles of the eyes absent gene in Drosophila. Dev. Biol. 196, 42–57.

34. Borsani, G., DeGrandi, A., Ballabio, A., Bulfone, A., Bernard, L., Banfi, S., Gattuso, C., Mariani, M., Dixon, M., Donnai, D., et al. (1999). EYA4, a novel vertebrate gene related to Drosophila eyes absent. Hum. Mol. Genet. 8, 11–23.

35. Bose, P., Simmons, G.L., and Grant, S. (2013). Cyclin-dependent kinase inhibitor therapy for hematologic malignancies. Expert Opin. Investig. Drugs 22, 723–738.

36. Bosnakovski, D., Xu, Z., Li, W., Thet, S., Cleaver, O., Perlingeiro, R.C., and Kyba, M. (2008). Prospective isolation of skeletal muscle stem cells with a Pax7 reporter. Stem Cells Dayt. Ohio 26, 3194–3204.

37. Boucher, C.A., King, S.K., Carey, N., Krahe, R., Winchester, C.L., Rahman, S., Creavin, T., Meghji, P., Bailey, M.E., and Chartier, F.L. (1995). A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable (CTG)n repeat. Hum. Mol. Genet. 4, 1919–1925.

38. Braun, T., and Arnold, H.H. (1995). Inactivation of Myf-6 and Myf-5 genes in mice leads to alterations in skeletal muscle development. EMBO J. 14, 1176–1186.

39. 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. EMBO J. 8, 701–709.

40. 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. EMBO J. 9, 821–831.

41. Brennan, T.J., Chakraborty, T., and Olson, E.N. (1991). Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. Proc. Natl. Acad. Sci. U. S. A. 88, 5675–5679.

42. Bricaud, O., and Collazo, A. (2011). Balancing cell numbers during organogenesis: Six1a differentially affects neurons and sensory hair cells in the inner ear. Dev. Biol. 357, 191–201.

43. Brodbeck, S., and Englert, C. (2004). Genetic determination of nephrogenesis: the Pax/Eya/Six gene network. Pediatr. Nephrol. Berl. Ger. 19, 249–255.

44.Brooke, M.H., and Kaiser, K.K. (1970). Muscle Fiber Types: How Many and What Kind? Arch. Neurol. 23, 369–379.

45. Brugmann, S.A., Pandur, P.D., Kenyon, K.L., Pignoni, F., and Moody, S.A. (2004). Six1 promotes a placodal fate within the lateral neurogenic ectoderm by functioning as both a transcriptional activator and repressor. Dev. Camb. Engl. 131, 5871–5881.

46. Buckingham, M., and Rigby, P.W.J. (2014). Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev. Cell 28, 225–238.

47. Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D., and Relaix, F. (2003). The formation of skeletal muscle: from somite to limb. J. Anat. 202, 59–68.

48. Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. In Current Protocols in Molecular Biology, (John Wiley & Sons, Inc.), p.

49. Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev. Biol. 339, 250–257.

50. Bürglin, T.R., and Affolter, M. (2016). Homeodomain proteins: an update. Chromosoma 125, 497–521.

51. 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.

52. Caravaca, J.M., Donahue, G., Becker, J.S., He, X., Vinson, C., and Zaret, K.S. (2013). Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes Dev. 27, 251–260.

53. Carl, M., Loosli, F., and Wittbrodt, J. (2002). Six3 inactivation reveals its essential role for the formation and patterning of the vertebrate eye. Development 129, 4057–4063.

54. Chae, J.H., Stein, G.H., and Lee, J.E. (2004). NeuroD: the predicted and the surprising. Mol. Cells 18, 271–288.

55. Chakkalakal, J.V., Jones, K.M., Basson, M.A., and Brack, A.S. (2012). The aged niche disrupts muscle stem cell quiescence. Nature 490, 355–360.

56. Chakroun, I., Yang, D., Girgis, J., Gunasekharan, A., Phenix, H., Kærn, M., and Blais, A. (2015). Genome-wide association between Six4, MyoD, and the histone demethylase Utx during myogenesis. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 29, 4738–4755.

57. Chang, N.C., and Rudnicki, M.A. (2014). Satellite cells: the architects of skeletal muscle. Curr. Top. Dev. Biol. 107, 161–181.

58. Chang, E.H., Menezes, M., Meyer, N.C., Cucci, R.A., Vervoort, V.S., Schwartz, C.E., and Smith, R.J. (2004). Branchio-oto-renal syndrome: the mutation spectrum in EYA1 and its phenotypic consequences. Hum. Mutat. 23, 582–589.

59. Chao, S.H., and Price, D.H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J. Biol. Chem. 276, 31793–31799.

60. Chen, J.C., and Goldhamer, D.J. (2004). The core enhancer is essential for proper timing of MyoD activation in limb buds and branchial arches. Dev. Biol. 265, 502–512.

61. Chen, J.C., Mortimer, J., Marley, J., and Goldhamer, D.J. (2005). MyoD-cre transgenic mice: a model for conditional mutagenesis and lineage tracing of skeletal muscle. Genes. N. Y. N 2000 41, 116–121.

62. Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X., et al. (2013). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes Dev. 27, 2109–2124.

63. 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.

64. Cheng, T.C., Wallace, M.C., Merlie, J.P., and Olson, E.N. (1993). Separable regulatory elements governing myogenin transcription in mouse embryogenesis. Science 261, 215–218.

65. Cheung, T.H., Quach, N.L., Charville, G.W., Liu, L., Park, L., Edalati, A., Yoo, B., Hoang, P., and Rando, T.A. (2012). Maintenance of muscle stem-cell quiescence by microRNA-489. Nature 482, 524–528.

66. Cheyette, B.N., Green, P.J., Martin, K., Garren, H., Hartenstein, V., and Zipursky, S.L. (1994). The Drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system. Neuron 12, 977–996.

67. Cho, H., Orphanides, G., Sun, X., Yang, X.-J., Ogryzko, V., Lees, E., Nakatani, Y., and Reinberg, D. (1998). A Human RNA Polymerase II Complex Containing Factors That Modify Chromatin Structure. Mol. Cell. Biol. 18, 5355–5363.

68. Choi, J., Costa, M.L., Mermelstein, C.S., Chagas, C., Holtzer, S., and Holtzer, H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. Proc. Natl. Acad. Sci. U. S. A. 87, 7988–7992.

69. Christ, B., and Brand-Saberi, B. (2002). Limb muscle development. Int. J. Dev. Biol. 46, 905–914.

70. Christensen, K.L., Brennan, J.D., Aldridge, C.S., and Ford, H.L. (2007). Cell cycle regulation of the human Six1 homeoprotein is mediated by APC(Cdh1). Oncogene 26, 3406–3414.

71. Christensen, K.L., Patrick, A.N., McCoy, E.L., and Ford, H.L. (2008). The six family of homeobox genes in development and cancer. Adv. Cancer Res. 101, 93–126.

72. Coletta, R.D., Christensen, K., Reichenberger, K.J., Lamb, J., Micomonaco, D., Huang, L., Wolf, D.M., Muller-Tidow, C., Golub, T.R., Kawakami, K., et al. (2004). The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. Proc. Natl. Acad. Sci. U. S. A. 101, 6478–6483.

73. Coletta, R.D., Christensen, K.L., Micalizzi, D.S., Jedlicka, P., Varella-Garcia, M., and Ford, H.L. (2008). Six1 overexpression in mammary cells induces genomic instability and is sufficient for malignant transformation. Cancer Res. 68, 2204–2213.

74. Conboy, I.M., and Rando, T.A. (2002). The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. Dev. Cell 3, 397–409.

75. Conboy, I.M., Conboy, M.J., Smythe, G.M., and Rando, T.A. (2003). Notch-mediated restoration of regenerative potential to aged muscle. Science 302, 1575–1577.

76. Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319, 1791–1792.

77. Couture, J.F., and Trievel, R.C. (2006). Histone-modifying enzymes: encrypting an enigmatic epigenetic code. Curr. Opin. Struct. Biol. 16, 753–760.

78. Creemers, E.E., Sutherland, L.B., Oh, J., Barbosa, A.C., and Olson, E.N. (2006). Coactivation of MEF2 by the SAP Domain Proteins Myocardin and MASTR. Mol. Cell 23, 83–96.

79. Crist, C.G., Montarras, D., and Buckingham, M. (2012). Muscle satellite cells are primed for myogenesis but maintain quiescence with sequestration of Myf5 mRNA targeted by microRNA-31 in mRNP granules. Cell Stem Cell 11, 118–126.

80. Dahlmans, D., Houzelle, A., Schrauwen, P., and Hoeks, J. (2016). Mitochondrial dynamics, quality control and miRNA regulation in skeletal muscle: implications for obesity and related metabolic disease. Clin. Sci. Lond. Engl. 1979 130, 843–852.

81. Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., and Aerts, S. (2015). Discovery of Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by ATAC-seq and FAIRE-seq Open Chromatin Profiling. PLOS Genet. 11, e1004994.

82. Davis, R.J., and Barr, F.G. (1997). Fusion genes resulting from alternative chromosomal translocations are overexpressed by gene-specific mechanisms in alveolar rhabdomyosarcoma. Proc. Natl. Acad. Sci. U. S. A. 94, 8047–8051.

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

84. 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.

85. De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. 8, e1000384.

86. Delgado-Olguín, P., Brand-Arzamendi, K., Scott, I.C., Jungblut, B., Stainier, D.Y., Bruneau, B.G., and Recillas-Targa, F. (2011). CTCF Promotes Muscle Differentiation by Modulating the Activity of Myogenic Regulatory Factors. J. Biol. Chem. 286, 12483–12494.

87. Dellavalle, A., Maroli, G., Covarello, D., Azzoni, E., Innocenzi, A., Perani, L., Antonini, S., Sambasivan, R., Brunelli, S., Tajbakhsh, S., et al. (2011). Pericytes resident in postnatal skeletal muscle differentiate into muscle fibres and generate satellite cells. Nat. Commun. 2, 499.

88. Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife 3, e02105.

89. Di Micco, R., Fontanals-Cirera, B., Low, V., Ntziachristos, P., Yuen, S.K., Lovell, C.D., Dolgalev, I., Yonekubo, Y., Zhang, G., Rusinova, E., et al. (2014). Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. Cell Rep. 9, 234–247.

90. 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. U. S. A. 101, 11593–11598.

91. Donati, C., Cencetti, F., and Bruni, P. (2013). Sphingosine 1-phosphate axis: a new leader actor in skeletal muscle biology. Front. Physiol. 4, 338.

92. Dressler, G.R. (2006). The cellular basis of kidney development. Annu. Rev. Cell Dev. Biol. 22, 509–529.

93. Dumont, N.A., Wang, Y.X., and Rudnicki, M.A. (2015). Intrinsic and extrinsic mechanisms regulating satellite cell function. Dev. Camb. Engl. 142, 1572–1581.

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

95. Eisner, A., Pazyra-Murphy, M.F., Durresi, E., Zhou, P., Zhao, X., Chadwick, E.C., Xu, P.-X., Hillman, R.T., Scott, M.P., Greenberg, M.E., et al. (2015). The Eya1 phosphatase promotes Shh signaling during hindbrain development and oncogenesis. Dev. Cell 33, 22–35.

96. Ekmark, M., Rana, Z.A., Stewart, G., Hardie, D.G., and Gundersen, K. (2007). De-phosphorylation of MyoD is linking nerve-evoked activity to fast myosin heavy chain expression in rodent adult skeletal muscle. J. Physiol. 584, 637–650.

97. Epstein, D.J., Vekemans, M., and Gros, P. (1991). Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. Cell 67, 767–774.

98. Epstein, J.A., Shapiro, D.N., Cheng, J., Lam, P.Y., and Maas, R.L. (1996). Pax3 modulates expression of the c-Met receptor during limb muscle development. Proc. Natl. Acad. Sci. U. S. A. 93, 4213–4218.

99. Ercan, S., Carrozza, M.J., and Workman, J.L. (2004). Global nucleosome distribution and the regulation of transcription in yeast. Genome Biol. 5, 243.

100. Faralli, H., Dilworth, F.J., Faralli, H., and Dilworth, F.J. (2012). Turning on Myogenin in Muscle: A Paradigm for Understanding Mechanisms of Tissue-Specific Gene Expression, Turning on Myogenin in Muscle: A Paradigm for Understanding Mechanisms of Tissue-Specific Gene Expression. Int. J. Genomics Int. J. Genomics 2012, 2012, e836374.

101. Fischbach, K.F., and Technau, G. (1984). Cell degeneration in the developing optic lobes of the sine oculis and small-optic-lobes mutants of Drosophila melanogaster. Dev. Biol. 104, 219–239.

102. Fischer, P.M., and Gianella-Borradori, A. (2005). Recent progress in the discovery and development of cyclin-dependent kinase inhibitors. Expert Opin. Investig. Drugs 14, 457–477.

103. Fong, A.P., and Tapscott, S.J. (2013). Skeletal muscle programming and re-programming. Curr. Opin. Genet. Dev. 23, 568–573.

104. Fong, A.P., Yao, Z., Zhong, J.W., Cao, Y., Ruzzo, W.L., Gentleman, R.C., and Tapscott, S.J. (2012). Genetic and epigenetic determinants of neurogenesis and myogenesis. Dev. Cell 22, 721–735.

105. Fong, A.P., Yao, Z., Zhong, J.W., Johnson, N.M., Farr III, G.H., Maves, L., and Tapscott, S.J. (2015). Conversion of MyoD to a Neurogenic Factor: Binding Site Specificity Determines Lineage. Cell Rep. 10, 1937–1946.

106. Forcales, S.V., Albini, S., Giordani, L., Malecova, B., Cignolo, L., Chernov, A., Coutinho, P., Saccone, V., Consalvi, S., Williams, R., et al. (2012). Signal-dependent incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex. EMBO J. 31, 301–316.

107. Ford, H.L., Landesman-Bollag, E., Dacwag, C.S., Stukenberg, P.T., Pardee, A.B., and Seldin, D.C. (2000). Cell cycle-regulated phosphorylation of the human SIX1 homeodomain protein. J. Biol. Chem. 275, 22245–22254.

108. Fougerousse, F., Durand, M., Lopez, S., Suel, L., Demignon, J., Thornton, C., Ozaki, H., Kawakami, K., Barbet, P., Beckmann, J.S., et al. (2002). Six and Eya expression during human somitogenesis and MyoD gene family activation. J. Muscle Res. Cell Motil. 23, 255–264.

109. Fu, T.J., Peng, J., Lee, G., Price, D.H., and Flores, O. (1999). Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription. J. Biol. Chem. 274, 34527–34530.

110. Fujimoto, Y., Tanaka, S.S., Yamaguchi, Y.L., Kobayashi, H., Kuroki, S., Tachibana, M., Shinomura, M., Kanai, Y., Morohashi, K.-I., Kawakami, K., et al. (2013). Homeoproteins Six1 and Six4 regulate male sex determination and mouse gonadal development. Dev. Cell 26, 416–430.

111. Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol. Cell. Biol. 24, 787–795.

112. Fukada, S., Higuchi, S., Segawa, M., Koda, K., Yamamoto, Y., Tsujikawa, K., Kohama, Y., Uezumi, A., Imamura, M., Miyagoe-Suzuki, Y., et al. (2004). Purification and cell-surface marker characterization of quiescent satellite cells from murine skeletal muscle by a novel monoclonal antibody. Exp. Cell Res. 296, 245–255.

113. Gaertner, B., and Zeitlinger, J. (2014). RNA polymerase II pausing during development. Dev. Camb. Engl. 141, 1179–1183.

114. García-González, E., Escamilla-Del-Arenal, M., Arzate-Mejía, R., and Recillas-Targa, F. (2016). Chromatin remodeling effects on enhancer activity. Cell. Mol. Life Sci. CMLS.

115. Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. 12, 36–47.

116. 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 Dev. 11, 436–450.

117. Giacinti, C., Bagella, L., Puri, P.L., Giordano, A., and Simone, C. (2006). MyoD recruits the cdk9/cyclin T2 complex on myogenic-genes regulatory regions. J. Cell. Physiol. 206, 807–813.

118. Giacinti, C., Musarò, A., De Falco, G., Jourdan, I., Molinaro, M., Bagella, L., Simone, C., and Giordano, A. (2008). Cdk9-55: a new player in muscle regeneration. J. Cell. Physiol. 216, 576–582.

119. Gianakopoulos, P.J., Mehta, V., Voronova, A., Cao, Y., Yao, Z., Coutu, J., Wang, X., Waddington, M.S., Tapscott, S.J., and Skerjanc, I.S. (2011). MyoD directly up-regulates premyogenic mesoderm factors during induction of skeletal myogenesis in stem cells. J. Biol. Chem. 286, 2517–2525.

120. Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell 143, 540–551.

121. Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M., and Maire, P. (2007). Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. Proc. Natl. Acad. Sci. U. S. A. 104, 11310–11315.

122. Goldhamer, D.J., Faerman, A., Shani, M., and Jr, C.P.E. (1992). Regulatory elements that control the lineage-specific expression of myoD. Science 256, 538–542.

123. Goldhamer, D.J., Brunk, B.P., Faerman, A., King, A., Shani, M., and Jr, C.P.E. (1995). Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. Dev. Camb. Engl. 121, 637–649.

124. Grand, F.L., and Rudnicki, M.A. (2007). Skeletal muscle satellite cells and adult myogenesis. Curr. Opin. Cell Biol. 19, 628–633.

125. Grifone, R., Laclef, C., Spitz, F., Lopez, S., Demignon, J., Guidotti, J.E., Kawakami, K., Xu, P.X., Kelly, R., Petrof, B.J., et al. (2004). Six1 and Eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype. Mol. Cell. Biol. 24, 6253–6267.

126. Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M.J., Hamard, G., and Maire, P. (2005). Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. Dev. Camb. Engl. 132, 2235–2249.

127. Grifone, R., Demignon, J., Giordani, J., Niro, C., Souil, E., Bertin, F., Laclef, C., Xu, P.X., and Maire, P. (2007). Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. Dev. Biol. 302, 602–616.

128. Grocott, T., Tambalo, M., and Streit, A. (2012). The peripheral sensory nervous system in the vertebrate head: a gene regulatory perspective. Dev. Biol. 370, 3–23.

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

130. Gross, M.K., Moran-Rivard, L., Velasquez, T., Nakatsu, M.N., Jagla, K., and Goulding, M. (2000). Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. Dev. Camb. Engl. 127, 413–424.

131. Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., and Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev. 10, 1670–1682.

132. Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. (2010). Chromatin Structure and Gene Expression Programs of Human Embryonic and Induced Pluripotent Stem Cells. Cell Stem Cell 7, 249–257.

133. Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. 23, 1210–1223.

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

135. Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that Specify Promoter Nucleosome Location and Identity. Cell 137, 445–458.

136. 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.

137. Hatta, M., and Cirillo, L.A. (2007). Chromatin Opening and Stable Perturbation of Core Histone:DNA Contacts by FoxO1. J. Biol. Chem. 282, 35583–35593.

138. Heanue, T.A., Reshef, R., Davis, R.J., Mardon, G., Oliver, G., Tomarev, S., Lassar, A.B., and Tabin, C.J. (1999). Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation. Genes Dev. 13, 3231–3243.

139. Heidt, A.B., Rojas, A., Harris, I.S., and Black, B.L. (2007). Determinants of myogenic specificity within MyoD are required for noncanonical E box binding. Mol. Cell. Biol. 27, 5910–5920.

140. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Calcar, S.V., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. 39, 311–318.

141. Ho, J.W., Bishop, E., Karchenko, P.V., Nègre, N., White, K.P., and Park, P.J. (2011). ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis. BMC Genomics 12, 134.

142. Hon, G.C., Hawkins, R.D., and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. Hum. Mol. Genet. 18, R195-201.

143. Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.-S.M., Chen, S., et al. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc. Natl. Acad. Sci. U. S. A. 111, 7319–7324.

144. Hu, P., Geles, K.G., Paik, J.H., DePinho, R.A., and Tjian, R. (2008). Codependent activators direct myoblast-specific MyoD transcription. Dev. Cell 15, 534–546.

145. IIott, N.E., Heward, J.A., Roux, B., Tsitsiou, E., Fenwick, P.S., Lenzi, L., Goodhead, I., Hertz-Fowler, C., Heger, A., Hall, N., et al. (2014). Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. Nat. Commun. 5, 3979.

146. Ikeda, K., Watanabe, Y., Ohto, H., and Kawakami, K. (2002). Molecular interaction and synergistic activation of a promoter by Six, Eya, and Dach proteins mediated through CREB binding protein. Mol. Cell. Biol. 22, 6759–6766.

147. Imam, J.S., Buddavarapu, K., Lee-Chang, J.S., Ganapathy, S., Camosy, C., Chen, Y., and Rao, M.K. (2010). MicroRNA-185 suppresses tumor growth and progression by targeting the Six1 oncogene in human cancers. Oncogene.

148. Irintchev, A., Zeschnigk, M., Starzinski-Powitz, A., and Wernig, A. (1994). Expression pattern of M-cadherin in normal, denervated, and regenerating mouse muscles. Dev. Dyn. Off. Publ. Am. Assoc. Anat. 199, 326–337.

149. Ishibashi, J., Perry, R.L., Asakura, A., and Rudnicki, M.A. (2005). ram. J. Cell Biol. 171, 471–482.

150. Iwafuchi-Doi, M., and Zaret, K.S. (2016). Cell fate control by pioneer transcription factors. Dev. Camb. Engl. 143, 1833–1837.

151. Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol. Cell 19, 523–534.

152. Jean, D., Bernier, G., and Gruss, P. (1999). Six6 (Optx2) is a novel murine Six3-related homeobox gene that demarcates the presumptive pituitary/hypothalamic axis and the ventral optic stalk. Mech. Dev. 84, 31–40.

153. Jeziorska, D.M., Jordan, K.W., and Vance, K.W. (2009). A systems biology approach to understanding cis-regulatory module function. Semin. Cell Dev. Biol. 20, 856–862.

154. Jiang, C., Wen, Y., Kuroda, K., Hannon, K., Rudnicki, M.A., and Kuang, S. (2014). Notch signaling deficiency underlies age-dependent depletion of satellite cells in muscular dystrophy. Dis. Model. Mech. 7, 997–1004.

155. Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant–containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. Nat. Genet. 41, 941–945.

156. Jin, Q., Yu, L.-R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.-E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. 30, 249–262.

157. Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Mol. Cell 29, 232–242.

158. Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16, 167–177.

159. Judson, R.N., Zhang, R.-H., and Rossi, F.M.A. (2013). Tissue-resident mesenchymal stem/progenitor cells in skeletal muscle: collaborators or saboteurs? FEBS J. 280, 4100–4108.

160. Kadauke, S., Udugama, M.I., Pawlicki, J.M., Achtman, J.C., Jain, D.P., Cheng, Y., Hardison, R.C., and Blobel, G.A. (2012). Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. Cell 150, 725–737.

161. Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell 51, 310–325.

162. Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., and Peterlin, B.M. (2003). c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. Oncogene 22, 5707–5711.

163. Kanno, T., Kanno, Y., LeRoy, G., Campos, E., Sun, H.-W., Brooks, S.R., Vahedi, G., Heightman, T.D., Garcia, B.A., Reinberg, D., et al. (2014). BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. Nat. Struct. Mol. Biol. 21, 1047–1057.

164. 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.

165. Kassar-Duchossoy, L., Giacone, E., Gayraud-Morel, B., Jory, A., Gomès, D., and Tajbakhsh, S. (2005). Pax3/Pax7 mark a novel population of primitive myogenic cells during development. Genes Dev. 19, 1426–1431.

166. Kawakami, K., Ohto, H., Takizawa, T., and Saito, T. (1996a). Identification and expression of six family genes in mouse retina. FEBS Lett. 393, 259–263.

167. Kawakami, K., Ohto, H., Ikeda, K., and Roeder, R.G. (1996b). Structure, function and expression of a murine homeobox protein AREC3, a homologue of Drosophila sine oculis gene product, and implication in development. Nucleic Acids Res. 24, 303–310.

168. Kawakami, K., Sato, S., Ozaki, H., and Ikeda, K. (2000). Six family genes–structure and function as transcription factors and their roles in development. BioEssays News Rev. Mol. Cell. Dev. Biol. 22, 616–626.

169. Keller, C., and Guttridge, D.C. (2013). Mechanisms of impaired differentiation in rhabdomyosarcoma. FEBS J. 280, 4323–4334.

170. Khare, S.P., Habib, F., Sharma, R., Gadewal, N., Gupta, S., and Galande, S. (2012). HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 40, D337–D342.

171. Kim, T.-K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell 162, 948–959.

172. Kim, K.S., Sack, J.S., Tokarski, J.S., Qian, L., Chao, S.T., Leith, L., Kelly, Y.F., Misra, R.N., Hunt, J.T., Kimball, S.D., et al. (2000). Thio- and oxoflavopiridols, cyclin-dependent kinase 1-selective inhibitors: synthesis and biological effects. J. Med. Chem. 43, 4126–4134.

173. Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182–187.

174. Kim, T.-K., Hemberg, M., and Gray, J.M. (2015a). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb. Perspect. Biol. 7, a018622.

175. Kim, Y.W., Lee, S., Yun, J., and Kim, A. (2015b). Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the β-globin locus. Biosci. Rep. 35.

176. Klesert, T.R., Cho, D.H., Clark, J.I., Maylie, J., Adelman, J., Snider, L., Yuen, E.C., Soriano, P., and Tapscott, S.J. (2000). Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. Nat. Genet. 25, 105–109.

177. Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat. Struct. Mol. Biol. 18, 956–963.

178. Kochhar, A., Fischer, S.M., Kimberling, W.J., and Smith, R.J. (2007). Branchio-oto-renal syndrome. Am. J. Med. Genet. A 143A, 1671–1678.

179. Kochhar, A., Orten, D.J., Sorensen, J.L., Fischer, S.M., Cremers, C.W., Kimberling, W.J., and Smith, R.J. (2008). SIX1 mutation screening in 247 branchio-oto-renal syndrome families: a recurrent missense mutation associated with BOR. Hum. Mutat. 29, 565.

180. Kornberg, T.B. (1993). Understanding the homeodomain. J. Biol. Chem. 268, 26813–26816.

181. Krystof, V., and Uldrijan, S. (2010). Cyclin-dependent kinase inhibitors as anticancer drugs. Curr. Drug Targets 11, 291–302.

182. Kuang, S., Gillespie, M.A., and Rudnicki, M.A. (2008). Niche regulation of muscle satellite cell self-renewal and differentiation. Cell Stem Cell 2, 22–31.

183. Kumar, J.P. (2009a). The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. Cell. Mol. Life Sci. CMLS 66, 565–583.

184. Kumar, J.P. (2009b). The molecular circuitry governing retinal determination. Biochim. Biophys. Acta 1789, 306–314.

185. LaBarge, M.A., and Blau, H.M. (2002). Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. Cell 111, 589–601.

186. Laclef, C., Hamard, G., Demignon, J., Souil, E., Houbron, C., and Maire, P. (2003a). Altered myogenesis in Six1-deficient mice. Dev. Camb. Engl. 130, 2239–2252.

187. Laclef, C., Souil, E., Demignon, J., and Maire, P. (2003b). Thymus, kidney and craniofacial abnormalities in Six 1 deficient mice. Mech. Dev. 120, 669–679.

188. Lagha, M., Sato, T., Regnault, B., Cumano, A., Zuniga, A., Licht, J., Relaix, F., and Buckingham, M. (2010). Transcriptome analyses based on genetic screens for Pax3 myogenic targets in the mouse embryo. BMC Genomics 11, 696.

189. Lagutin, O.V., Zhu, C.C., Kobayashi, D., Topczewski, J., Shimamura, K., Puelles, L., Russell, H.R.C., McKinnon, P.J., Solnica-Krezel, L., and Oliver, G. (2003). Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. Genes Dev. 17, 368–379.

190. Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494, 497–501.

191. Lam, M.T.Y., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature 498, 511–515.

192. Lam, M.T.Y., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. 39, 170–182.

193. Lang, D., Powell, S.K., Plummer, R.S., Young, K.P., and Ruggeri, B.A. (2007). PAX genes: roles in development, pathophysiology, and cancer. Biochem. Pharmacol. 73, 1–14.

194. Larder, R., Clark, D.D., Miller, N.L.G., and Mellon, P.L. (2011). Hypothalamic dysregulation and infertility in mice lacking the homeodomain protein Six6. J. Neurosci. Off. J. Soc. Neurosci. 31, 426–438.

195. 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.

196. Le Grand, F., Grifone, R., Mourikis, P., Houbron, C., Gigaud, C., Pujol, J., Maillet, M., Pagès, G., Rudnicki, M., Tajbakhsh, S., et al. (2012). Six1 regulates stem cell repair potential and self-renewal during skeletal muscle regeneration. J. Cell Biol. 198, 815–832.

197. Lee, C.-K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004a). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat. Genet. 36, 900–905.

198. Lee, H., Habas, R., and Abate-Shen, C. (2004b). MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. Science 304, 1675–1678.

199. Lee, K., Hsiung, C.C.-S., Huang, P., Raj, A., and Blobel, G.A. (2015). Dynamic enhancer-gene body contacts during transcription elongation. Genes Dev. 29, 1992–1997.

200. Lepper, C., Partridge, T.A., and Fan, C.M. (2011). An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. Dev. Camb. Engl. 138, 3639–3646.

201. L'honore, A., Lamb, N.J., Vandromme, M., Turowski, P., Carnac, G., and Fernandez, A. (2003). MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. Mol. Biol. Cell 14, 2151–2162.

202. L'honore, A., Rana, V., Arsic, N., Franckhauser, C., Lamb, N.J., and Fernandez, A. (2007). Identification of a new hybrid serum response factor and myocyte enhancer factor 2-binding element in MyoD enhancer required for MyoD expression during myogenesis. Mol. Biol. Cell 18, 1992–2001.

203. L'honoré, A., Ouimette, J.-F., Lavertu-Jolin, M., and Drouin, J. (2010). Pitx2 defines alternate pathways acting through MyoD during limb and somitic myogenesis. Dev. Camb. Engl. 137, 3847–3856.

204. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013a). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498, 516–520.

205. Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat. Rev. Genet. 17, 207–223.

206. Li, X., Perissi, V., Liu, F., Rose, D.W., and Rosenfeld, M.G. (2002). Tissue-specific regulation of retinal and pituitary precursor cell proliferation. Science 297, 1180–1183.

207. Li, X., Oghi, K.A., Zhang, J., Krones, A., Bush, K.T., Glass, C.K., Nigam, S.K., Aggarwal, A.K., Maas, R., Rose, D.W., et al. (2003). Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. Nature 426, 247–254.

208. Li, Z., Tian, T., Lv, F., Chang, Y., Wang, X., Zhang, L., Li, X., Li, L., Ma, W., Wu, J., et al. (2013b). Six1 promotes proliferation of pancreatic cancer cells via upregulation of cyclin D1 expression. PloS One 8, e59203.

209. Liu, L., Cheung, T.H., Charville, G.W., Hurgo, B.M., Leavitt, T., Shih, J., Brunet, A., and Rando, T.A. (2013a). Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. Cell Rep. 4, 189–204.

210. Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2013b). Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell 155, 1581–1595.

211. Liu, Y., Chu, A., Chakroun, I., Islam, U., and Blais, A. (2010). Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation. Nucleic Acids Res. 38, 6857–6871.

212. Liu, Y., Chakroun, I., Yang, D., Horner, E., Liang, J., Aziz, A., Chu, A., Repentigny, Y.D., Dilworth, F.J., Kothary, R., et al. (2013c). Six1 regulates MyoD expression in adult muscle progenitor cells. PloS One 8, e67762.

213. Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell 77, 451–459.

214. Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1, 55–70.

215. Mammen, A.L., Casciola-Rosen, L.A., Hall, J.C., Christopher-Stine, L., Corse, A.M., and Rosen, A. (2009). Expression of the dermatomyositis autoantigen Mi-2 in regenerating muscle. Arthritis Rheum. 60, 3784–3793.

216. Marchesi, I., Nieddu, V., Caracciolo, V., Maioli, M., Gaspa, L., Giordano, A., and Bagella, L. (2013). Activation and function of murine Cyclin T2A and Cyclin T2B during skeletal muscle differentiation. J. Cell. Biochem. 114, 728–734.

217. Mardon, G., Solomon, N.M., and Rubin, G.M. (1994). dachshund encodes a nuclear protein required for normal eye and leg development in Drosophila. Dev. Camb. Engl. 120, 3473–3486.

218. 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.

219. Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J. Biol. Chem. 271, 27176–27183.

220. Marti, M., Montserrat, N., Pardo, C., Mulero, L., Miquel-Serra, L., Rodrigues, A.M.C., Vaquero, J.A., Kuebler, B., Morera, C., Barrero, M.J., et al. (2013). M-cadherin-mediated intercellular interactions activate satellite cell division. J Cell Sci 126, 5116–5131.

221. Martinez-Fernandez, A., Nelson, T.J., Yamada, S., Reyes, S., Alekseev, A.E., Perez-Terzic, C., Ikeda, Y., and Terzic, A. (2009). iPS programmed without c-MYC yield proficient cardiogenesis for functional heart chimerism. Circ. Res. 105, 648–656.

222. Mathew, S.J., Hansen, J.M., Merrell, A.J., Murphy, M.M., Lawson, J.A., Hutcheson, D.A., Hansen, M.S., Angus-Hill, M., and Kardon, G. (2011). Connective tissue fibroblasts and Tcf4 regulate myogenesis. Dev. Camb. Engl. 138, 371–384.

223. Mauro, A. (1961). Satellite cell of skeletal muscle fibers. J. Biophys. Biochem. Cytol. 9, 493–495.

224. McCoy, E.L., Iwanaga, R., Jedlicka, P., Abbey, N.S., Chodosh, L.A., Heichman, K.A., Welm, A.L., and Ford, H.L. (2009). Six1 expands the mouse mammary epithelial stem/progenitor cell pool and induces mammary tumors that undergo epithelial-mesenchymal transition. J. Clin. Invest. 119, 2663–2677.

225. Meadows, E., Cho, J.H., Flynn, J.M., and Klein, W.H. (2008). Myogenin regulates a distinct genetic program in adult muscle stem cells. Dev. Biol. 322, 406–414.

226. Meadows, E., Flynn, J.M., and Klein, W.H. (2011). Myogenin regulates exercise capacity but is dispensable for skeletal muscle regeneration in adult mdx mice. PloS One 6, e16184.

227. Meeson, A.P., Hawke, T.J., Graham, S., Jiang, N., Elterman, J., Hutcheson, K., Dimaio, J.M., Gallardo, T.D., and Garry, D.J. (2004). Cellular and molecular regulation of skeletal muscle side population cells. Stem Cells Dayt. Ohio 22, 1305–1320.

228. Megeney, L.A., and Rudnicki, M.A. (1995). Determination versus differentiation and the MyoD family of transcription factors. Biochem. Cell Biol. Biochim. Biol. Cell. 73, 723–732.

229. Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A.F., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., et al. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol. Cell 49, 524–535.

230. Mercuri, E., and Muntoni, F. (2013). Muscular dystrophies. Lancet Lond. Engl. 381, 845–860.

231. Meyer, C.A., and Liu, X.S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat. Rev. Genet. 15, 709–721.

232. Minasi, M.G., Riminucci, M., De Angelis, L., Borello, U., Berarducci, B., Innocenzi, A., Caprioli, A., Sirabella, D., Baiocchi, M., De Maria, R., et al. (2002). The meso-angioblast: a multipotent, self-renewing cell that originates from the dorsal aorta and differentiates into most mesodermal tissues. Dev. Camb. Engl. 129, 2773–2783.

233. Mitchell, K.J., Pannerec, A., Cadot, B., Parlakian, A., Besson, V., Gomes, E.R., Marazzi, G., and Sassoon, D.A. (2010). Identification and characterization of a non-satellite cell muscle resident progenitor during postnatal development. Nat. Cell Biol. 12, 257–266.

234. Mokalled, M.H., Johnson, A.N., Creemers, E.E., and Olson, E.N. (2012). MASTR directs MyoD-dependent satellite cell differentiation during skeletal muscle regeneration. Genes Dev. 26, 190–202.

235. 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.

236. Montarras, D., Morgan, J., Collins, C., Relaix, F., Zaffran, S., Cumano, A., Partridge, T., and Buckingham, M. (2005). Direct isolation of satellite cells for skeletal muscle regeneration. Science 309, 2064–2067.

237. Montarras, D., L'honore, A., and Buckingham, M. (2013). Lying low but ready for action: the quiescent muscle satellite cell. FEBS J. 280, 4036–4050.

238. Morales, F., and Giordano, A. (2016). Overview of CDK9 as a Target in Cancer Research. Cell Cycle Georget. Tex 0.

239. Mosrati, M.A., Hammami, B., Rebeh, I.B., Ayadi, L., Dhouib, L., Ben Mahfoudh, K., Hakim, B., Charfeddine, I., Mnif, J., Ghorbel, A., et al. (2011). A novel dominant mutation in SIX1, affecting a highly conserved residue, result in only auditory defects in humans. Eur. J. Med. Genet. 54, e484-488.

240. Mourikis, P., Sambasivan, R., Castel, D., Rocheteau, P., Bizzarro, V., and Tajbakhsh, S. (2012). A critical requirement for notch signaling in maintenance of the quiescent skeletal muscle stem cell state. Stem Cells Dayt. Ohio 30, 243–252.

241. Mousavi, K., Zare, H., Wang, A.H., and Sartorelli, V. (2012). Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol. Cell 45, 255–262.

242. Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol. Cell 51, 606–617.

243. Murphy, M.M., Lawson, J.A., Mathew, S.J., Hutcheson, D.A., and Kardon, G. (2011). Satellite cells, connective tissue fibroblasts and their interactions are crucial for muscle regeneration. Dev. Camb. Engl. 138, 3625–3637.

244. Murre, C., Bain, G., van Dijk, M.A., Engel, I., Furnari, B.A., Massari, M.E., Matthews, J.R., Quong, M.W., Rivera, R.R., and Stuiver, M.H. (1994). Structure and function of helix-loop-helix proteins. Biochim. Biophys. Acta 1218, 129–135.

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

246. Naidu, P.S., Ludolph, D.C., To, R.Q., Hinterberger, T.J., and Konieczny, S.F. (1995). Myogenin and MEF2 function synergistically to activate the MRF4 promoter during myogenesis. Mol. Cell. Biol. 15, 2707–2718.

247. Narlikar, G.J., Fan, H.-Y., and Kingston, R.E. (2002). Cooperation between Complexes that Regulate Chromatin Structure and Transcription. Cell 108, 475–487.

248. Natoli, G. (2010). Maintaining cell identity through global control of genomic organization. Immunity 33, 12–24.

249. Ng, K.T., Lee, T.K., Cheng, Q., Wo, J.Y., Sun, C.K., Guo, D.Y., Lim, Z.X., Lo, C.M., Poon, R.T., Fan, S.T., et al. (2009). Suppression of tumorigenesis and metastasis of hepatocellular carcinoma by shRNA interference targeting on homeoprotein Six1. Int. J. CancerJournal Int. Cancer.

250. Niro, C., Demignon, J., Vincent, S., Liu, Y., Giordani, J., Sgarioto, N., Favier, M., Guillet-Deniau, I., Blais, A., and Maire, P. (2010). Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. Dev. Biol. 338, 168–182.

251. Nojima, M., Huang, Y., Tyagi, M., Kao, H.-Y., and Fujinaga, K. (2008). The positive transcription elongation factor b is an essential cofactor for the activation of transcription by myocyte enhancer factor 2. J. Mol. Biol. 382, 275–287.

252. Obrdlik, A., Kukalev, A., Louvet, E., Farrants, A.-K.Ö., Caputo, L., and Percipalle, P. (2008). The Histone Acetyltransferase PCAF Associates with Actin and hnRNP U for RNA Polymerase II Transcription. Mol. Cell. Biol. 28, 6342–6357.

253. O'Brien, L.L., Guo, Q., Lee, Y., Tran, T., Benazet, J.-D., Whitney, P.H., Valouev, A., and McMahon, A.P. (2016). Differential regulation of mouse and human nephron progenitors by the Six family of transcriptional regulators. Dev. Camb. Engl. 143, 595–608.

254. O'Brien, S.K., Knight, K.L., and Rana, T.M. (2012). Phosphorylation of histone H1 by P-TEFb is a necessary step in skeletal muscle differentiation. J. Cell. Physiol. 227, 383–389.

255. Ohto, H., Kamada, S., Tago, K., Tominaga, S.I., Ozaki, H., Sato, S., and Kawakami, K. (1999). Cooperation of six and eya in activation of their target genes through nuclear translocation of Eya. Mol. Cell. Biol. 19, 6815–6824.

256. Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature 448, 313–317.

257. Olguin, H.C., and Olwin, B.B. (2004). Pax-7 up-regulation inhibits myogenesis and cell cycle progression in satellite cells: a potential mechanism for self-renewal. Dev. Biol. 275, 375–388.

258. Olguin, H.C., Yang, Z., Tapscott, S.J., and Olwin, B.B. (2007). Reciprocal inhibition between Pax7 and muscle regulatory factors modulates myogenic cell fate determination. J. Cell Biol. 177, 769–779.

259. Oliver, G., Mailhos, A., Wehr, R., Copeland, N.G., Jenkins, N.A., and Gruss, P. (1995). Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. Dev. Camb. Engl. 121, 4045–4055.

260. Olson, E.N., and Williams, R.S. (2000). Remodeling muscles with calcineurin. BioEssays 22, 510–519.

261. Ozaki, H., Watanabe, Y., Takahashi, K., Kitamura, K., Tanaka, A., Urase, K., Momoi, T., Sudo, K., Sakagami, J., Asano, M., et al. (2001). Six4, a putative myogenin gene regulator, is not essential for mouse embryonal development. Mol. Cell. Biol. 21, 3343–3350.

262. Ozaki, H., Nakamura, K., Funahashi, J., Ikeda, K., Yamada, G., Tokano, H., Okamura, H., Kitamura, K., Muto, S., Kotaki, H., et al. (2004). Six1 controls patterning of the mouse otic vesicle. Dev. Camb. Engl. 131, 551–562.

263. Pasut, A., Oleynik, P., and Rudnicki, M. (2012). Isolation of Muscle Stem Cells by Fluorescence Activated Cell Sorting Cytometry. In Myogenesis, J.X. DiMario, ed. (Humana Press), pp. 53–64.

264. Patapoutian, A., Yoon, J.K., Miner, J.H., Wang, S., Stark, K., and Wold, B. (1995). Disruption of the mouse MRF4 gene identifies multiple waves of myogenesis in the myotome. Dev. Camb. Engl. 121, 3347–3358.

265. Patrick, A.N., Schiemann, B.J., Yang, K., Zhao, R., and Ford, H.L. (2009). Biochemical and Functional Characterization of Six SIX1 Branchio-oto-renal Syndrome Mutations. J. Biol. Chem. 284, 20781–20790.

266. Patrick, A.N., Cabrera, J.H., Smith, A.L., Chen, X.S., Ford, H.L., and Zhao, R. (2013). Structure-function analyses of the human SIX1-EYA2 complex reveal insights into metastasis and BOR syndrome. Nat. Struct. Mol. Biol. 20, 447–453.

267. Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. Genes Dev. 12, 755–762.

268. Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell 23, 297–305.

269. Petropoulos, H., and Skerjanc, I.S. (2002). Beta-catenin is essential and sufficient for skeletal myogenesis in P19 cells. J. Biol. Chem. 277, 15393–15399.

270. Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 20, 2922–2936.

271. Phelan, M.L., Sif, S., Narlikar, G.J., and Kingston, R.E. (1999). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol. Cell 3, 247–253.

272. Pignoni, F., Hu, B., Zavitz, K.H., Xiao, J., Garrity, P.A., and Zipursky, S.L. (1997). The Eye-Specification Proteins So and Eya Form a Complex and Regulate Multiple Steps in Drosophila Eye Development. Cell 91, 881–891.

273. Pirngruber, J., Shchebet, A., and Johnsen, S.A. (2009). Insights into the function of the human P-TEFb component CDK9 in the regulation of chromatin modifications and co-transcriptional mRNA processing. Cell Cycle Georget. Tex 8, 3636–3642.

274. Plank, J.L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. Mol. Cell 55, 5–14.

275. Polesskaya, A., Duquet, A., Naguibneva, I., Weise, C., Vervisch, A., Bengal, E., Hucho, F., Robin, P., and Harel-Bellan, A. (2000). CREB-binding protein/p300 activates MyoD by acetylation. J. Biol. Chem. 275, 34359–34364.

276. Pomerantz, J.H., Mukherjee, S., Palermo, A.T., and Blau, H.M. (2009). Reprogramming to a muscle fate by fusion recapitulates differentiation. J. Cell Sci. 122, 1045–1053.

277. Popov, V.M., Wu, K., Zhou, J., Powell, M.J., Mardon, G., Wang, C., and Pestell, R.G. (2010). The Dachshund gene in development and hormone-responsive tumorigenesis. Trends Endocrinol. Metab. TEM 21, 41–49.

278. Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: how is it established, and why does it matter? Dev. Biol. 339, 258–266.

279. Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell 141, 432–445.

280. Rampalli, S., Li, L., Mak, E., Ge, K., Brand, M., Tapscott, S.J., and Dilworth, F.J. (2007). p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. Nat. Struct. Mol. Biol. 14, 1150–1156.

281. Rebay, I. (2016). Multiple Functions of the Eya Phosphotyrosine Phosphatase. Mol. Cell. Biol. 36, 668–677.

282. Reichenberger, K.J., Coletta, R.D., Schulte, A.P., Varella-Garcia, M., and Ford, H.L. (2005). Gene amplification is a mechanism of Six1 overexpression in breast cancer. Cancer Res. 65, 2668–2675.

283. Relaix, F., and Buckingham, M. (1999). From insect eye to vertebrate muscle: redeployment of a regulatory network. Genes Dev. 13, 3171–3178.

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

285. Relaix, F., Demignon, J., Laclef, C., Pujol, J., Santolini, M., Niro, C., Lagha, M., Rocancourt, D., Buckingham, M., and Maire, P. (2013). Six Homeoproteins Directly Activate Myod Expression in the Gene Regulatory Networks That Control Early Myogenesis. PLoS Genet 9, e1003425.

286. Rezza, A., Sennett, R., and Rendl, M. (2014). Adult stem cell niches: cellular and molecular components. Curr. Top. Dev. Biol. 107, 333–372.

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

288. Ridgeway, A.G., and Skerjanc, I.S. (2001). Pax3 is essential for skeletal myogenesis and the expression of Six1 and Eya2. J. Biol. Chem. 276, 19033–19039.

289. Rocheteau, P., Gayraud-Morel, B., Siegl-Cachedenier, I., Blasco, M.A., and Tajbakhsh, S. (2012). A subpopulation of adult skeletal muscle stem cells retains all template DNA strands after cell division. Cell 148, 112–125.

290. Roth, J.F., Shikama, N., Henzen, C., Desbaillets, I., Lutz, W., Marino, S., Wittwer, J., Schorle, H., Gassmann, M., and Eckner, R. (2003). Differential role of p300 and CBP acetyltransferase during myogenesis: p300 acts upstream of MyoD and Myf5. EMBO J. 22, 5186–5196.

291. Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone Acetyltransferases. Annu. Rev. Biochem. 70, 81–120.

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

293. 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.

294. Ruf, R.G., Xu, P.X., Silvius, D., Otto, E.A., Beekmann, F., Muerb, U.T., Kumar, S., Neuhaus, T.J., Kemper, M.J., Jr, R.M.R., et al. (2004). SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. Proc. Natl. Acad. Sci. U. S. A. 101, 8090–8095.

295. Sacco, A., Doyonnas, R., Kraft, P., Vitorovic, S., and Blau, H.M. (2008). Self-renewal and expansion of single transplanted muscle stem cells. Nature 456, 502–506.

296. Salminen, M., Lopez, S., Maire, P., Kahn, A., and Daegelen, D. (1996). Fast-muscle-specific DNA-protein interactions occurring in vivo at the human aldolase A M promoter are necessary for correct promoter activity in transgenic mice. Mol. Cell. Biol. 16, 76–85.

297. Sambasivan, R., Gayraud-Morel, B., Dumas, G., Cimper, C., Paisant, S., Kelly, R.G., Kelly, R., and Tajbakhsh, S. (2009). Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. Dev. Cell 16, 810–821.

298. Sambasivan, R., Kuratani, S., and Tajbakhsh, S. (2011a). An eye on the head: the development and evolution of craniofacial muscles. Dev. Camb. Engl. 138, 2401–2415.

299. Sambasivan, R., Yao, R., Kissenpfennig, A., Wittenberghe, L.V., Paldi, A., Gayraud-Morel, B., Guenou, H., Malissen, B., Tajbakhsh, S., and Galy, A. (2011b). Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. Dev. Camb. Engl. 138, 3647–3656.

300. Santolini, M., Sakakibara, I., Gauthier, M., Ribas-Aulinas, F., Takahashi, H., Sawasaki, T., Mouly, V., Concordet, J.-P., Defossez, P.-A., Hakim, V., et al. (2016). MyoD reprogramming requires Six1 and Six4 homeoproteins: genome-wide cis-regulatory module analysis. Nucleic Acids Res.

301. Sarkar, P.S., Appukuttan, B., Han, J., Ito, Y., Ai, C., Tsai, W., Chai, Y., Stout, J.T., and Reddy, S. (2000). Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. Nat. Genet. 25, 110–114.

302. 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. Mol. Cell. Biol. 17, 1010–1026.

303. Schlosser, G., Awtry, T., Brugmann, S.A., Jensen, E.D., Neilson, K., Ruan, G., Stammler, A., Voelker, D., Yan, B., Zhang, C., et al. (2008). Eya1 and Six1 promote neurogenesis in the cranial placodes in a SoxB1-dependent fashion. Dev. Biol. 320, 199–214.

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

305. Seimiya, M., and Gehring, W.J. (2000). The Drosophila homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism. Dev. Camb. Engl. 127, 1879–1886.

306. Sekiya, T., Muthurajan, U.M., Luger, K., Tulin, A.V., and Zaret, K.S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 23, 804–809.

307. Self, M., Lagutin, O.V., Bowling, B., Hendrix, J., Cai, Y., Dressler, G.R., and Oliver, G. (2006). Six2 is required for suppression of nephrogenesis and progenitor renewal in the developing kidney. EMBO J. 25, 5214–5228.

308. Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annu. Rev. Biochem. 79, 271–293.

309. Seo, H.C., Curtiss, J., Mlodzik, M., and Fjose, A. (1999). Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. Mech. Dev. 83, 127–139.

310. Serikaku, M.A., and O'Tousa, J.E. (1994). sine oculis is a homeobox gene required for Drosophila visual system development. Genetics 138, 1137–1150.

311. Serna, I.L. de la, 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.

312. Shadrach, J.L., and Wagers, A.J. (2011). Stem cells for skeletal muscle repair. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366, 2297–2306.

313. Shapiro, G.I. (2006). Cyclin-dependent kinase pathways as targets for cancer treatment. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 24, 1770–1783.

314. Shea, K.L., Xiang, W., LaPorta, V.S., Licht, J.D., Keller, C., Basson, M.A., and Brack, A.S. (2010). Sprouty1 regulates reversible quiescence of a self-renewing adult muscle stem cell pool during regeneration. Cell Stem Cell 6, 117–129.

315. Shi, J., Whyte, W.A., Zepeda-Mendoza, C.J., Milazzo, J.P., Shen, C., Roe, J.-S., Minder, J.L., Mercan, F., Wang, E., Eckersley-Maslin, M.A., et al. (2013). Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes Dev. 27, 2648–2662.

316. Shieh, P.B. (2013). Muscular dystrophies and other genetic myopathies. Neurol. Clin. 31, 1009–1029.

317. Shore, S.M., Byers, S.A., Maury, W., and Price, D.H. (2003). Identification of a novel isoform of Cdk9. Gene 307, 175–182.

318. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science 350, 978–981.

319. Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat. Protoc. 7, 256–267.

320. Simone, C., Stiegler, P., Bagella, L., Pucci, B., Bellan, C., De Falco, G., De Luca, A., Guanti, G., Puri, P.L., and Giordano, A. (2002). Activation of MyoD-dependent transcription by cdk9/cyclin T2. Oncogene 21, 4137–4148.

321. Singh, K., and Dilworth, F.J. (2013). Differential modulation of cell cycle progression distinguishes members of the myogenic regulatory factor family of transcription factors. FEBS J. 280, 3991–4003.

322. Smith, S.T., and Jaynes, J.B. (1996). A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo. Dev. Camb. Engl. 122, 3141–3150.

323. Song, S.-H., Kim, A., Ragoczy, T., Bender, M.A., Groudine, M., and Dean, A. (2010). Multiple functions of Ldb1 required for beta-globin activation during erythroid differentiation. Blood 116, 2356–2364.

324. Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626.

325. 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. Proc. Natl. Acad. Sci. U. S. A. 95, 14220–14225.

326. Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F., and Soler, E. (2012a). Transcription regulation by distal enhancers. Transcription 3, 181–186.

327. Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R.-J., Bryne, J.C., van den Heuvel, A., Stevens, M., de Boer, E., Kockx, C., van der Sloot, A., et al. (2012b). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. EMBO J. 31, 986–999.

328. Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature 403, 41–45.

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

330. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

331. Tao, Y., Zheng, W., Jiang, Y., Ding, G., Hou, X., Tang, Y., Li, Y., Gao, S., Chang, G., Zhang, X., et al. (2014). Nucleosome organizations in induced pluripotent stem cells reprogrammed from somatic cells belonging to three different germ layers. BMC Biol. 12, 109.

332. Tapscott, S.J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. Dev. Camb. Engl. 132, 2685–2695.

333. Teves, S.S., An, L., Hansen, A.S., Xie, L., Darzacq, X., and Tjian, R. (2016). A dynamic mode of mitotic bookmarking by transcription factors. eLife 5.

334. Thomas, K., Engler, A.J., and Meyer, G.A. (2015). Extracellular matrix regulation in the muscle satellite cell niche. Connect. Tissue Res. 56, 1–8.

335. Tremblay, P., Dietrich, S., Mericskay, M., Schubert, F.R., Li, Z., and Paulin, D. (1998). A crucial role for Pax3 in the development of the hypaxial musculature and the long-range migration of muscle precursors. Dev. Biol. 203, 49–61.

336. Tropberger, P., Pott, S., Keller, C., Kamieniarz-Gdula, K., Caron, M., Richter, F., Li, G., Mittler, G., Liu, E.T., Bühler, M., et al. (2013). Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell 152, 859–872.

337. Urciuolo, A., Quarta, M., Morbidoni, V., Gattazzo, F., Molon, S., Grumati, P., Montemurro, F., Tedesco, F.S., Blaauw, B., Cossu, G., et al. (2013). Collagen VI regulates satellite cell self-renewal and muscle regeneration. Nat. Commun. 4, 1964.

338. Vierbuchen, T., and Wernig, M. (2011). Direct Lineage Conversions: Unnatural but useful? Nat. Biotechnol. 29, 892–907.

339. Vignali, M., Hassan, A.H., Neely, K.E., and Workman, J.L. (2000). ATP-Dependent Chromatin-Remodeling Complexes. Mol. Cell. Biol. 20, 1899–1910.

340. Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854–858.

341. Wallis, D.E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E.H., Rommens, J., and Muenke, M. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. Nat. Genet. 22, 196–198.

342. Walz, A.L., Ooms, A., Gadd, S., Gerhard, D.S., Smith, M.A., Guidry Auvil, J.M., Guidry Auvil, J.M., Meerzaman, D., Chen, Q.-R., Hsu, C.H., et al. (2015). Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. Cancer Cell 27, 286–297.

343. Wang, L., Tang, Y., Cole, P.A., and Marmorstein, R. (2008). Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. Curr. Opin. Struct. Biol. 18, 741–747.

344. Watanabe, A., Yamada, Y., and Yamanaka, S. (2013). Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. Phil Trans R Soc B 368, 20120292.

345. Wegert, J., Ishaque, N., Vardapour, R., Geörg, C., Gu, Z., Bieg, M., Ziegler, B., Bausenwein, S., Nourkami, N., Ludwig, N., et al. (2015). Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA Microprocessor Complex Underlie High-Risk Blastemal Type Wilms Tumors. Cancer Cell 27, 298–311.

346. Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc. Natl. Acad. Sci. U. S. A. 86, 5434–5438.

347. 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.

348. Weintraub, H., Dwarki, V.J., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, A., and Tapscott, S.J. (1991). Muscle-specific transcriptional activation by MyoD. Genes Dev. 5, 1377–1386.

349. Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448, 318–324.

350. Wong, J., Mehta, V., Voronova, A., Coutu, J., Ryan, T., Shelton, M., and Skerjanc, I.S. (2013). β-catenin is essential for efficient in vitro premyogenic mesoderm formation but can be partially compensated by retinoic acid signalling. PloS One 8, e57501.

351. 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.

352. Wu, K., Yang, Y., Wang, C., Davoli, M.A., D'Amico, M., Li, A., Cveklova, K., Kozmik, Z., Lisanti, M.P., Russell, R.G., et al. (2003). DACH1 inhibits transforming growth factor-beta signaling through binding Smad4. J. Biol. Chem. 278, 51673–51684.

353. Wu, K., Liu, M., Li, A., Donninger, H., Rao, M., Jiao, X., Lisanti, M.P., Cvekl, A., Birrer, M., and Pestell, R.G. (2007). Cell fate determination factor DACH1 inhibits c-Jun-induced contact-independent growth. Mol. Biol. Cell 18, 755–767.

354. Wu, K., Li, Z., Cai, S., Tian, L., Chen, K., Wang, J., Hu, J., Sun, Y., Li, X., Ertel, A., et al. (2013a). EYA1 phosphatase function is essential to drive breast cancer cell proliferation through cyclin D1. Cancer Res. 73, 4488–4499.

355. Wu, S.-Y., Lee, A.-Y., Lai, H.-T., Zhang, H., and Chiang, C.-M. (2013b). Phospho switch triggers Brd4 chromatin binding and activator recruitment for gene-specific targeting. Mol. Cell 49, 843–857.

356. Wu, W., Ren, Z., Liu, H., Wang, L., Huang, R., Chen, J., Zhang, L., Li, P., and Xiong, Y. (2013c). Core promoter analysis of porcine Six1 gene and its regulation of the promoter activity by CpG methylation. Gene 529, 238–244.

357. Xie, H., Hoffmann, H.M., Meadows, J.D., Mayo, S.L., Trang, C., Leming, S.S., Maruggi, C., Davis, S.W., Larder, R., and Mellon, P.L. (2015). Homeodomain Proteins SIX3 and SIX6 Regulate Gonadotrope-specific Genes During Pituitary Development. Mol. Endocrinol. Baltim. Md 29, 842–855.

358. Xu, P.-X. (2013). The EYA-SO/SIX complex in development and disease. Pediatr. Nephrol. Berl. Ger. 28, 843–854.

359. Xu, J., Wong, E.Y.M., Cheng, C., Li, J., Sharkar, M.T.K., Xu, C.Y., Chen, B., Sun, J., Jing, D., and Xu, P.-X. (2014). Eya1 interacts with Six2 and Myc to regulate expansion of the nephron progenitor pool during nephrogenesis. Dev. Cell 31, 434–447.

360. Xu, P.X., Woo, I., Her, H., Beier, D.R., and Maas, R.L. (1997). Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. Dev. Camb. Engl. 124, 219–231.

361. Xu, P.X., Zheng, W., Huang, L., Maire, P., Laclef, C., and Silvius, D. (2003). Six1 is required for the early organogenesis of mammalian kidney. Dev. Camb. Engl. 130, 3085–3094.

362. Yajima, H., Motohashi, N., Ono, Y., Sato, S., Ikeda, K., Masuda, S., Yada, E., Kanesaki, H., Miyagoe-Suzuki, Y., Takeda, S., et al. (2010). Six family genes control the proliferation and differentiation of muscle satellite cells. Exp. Cell Res. 316, 2932–2944.

363. Yalley, A., Schill, D., Hatta, M., Johnson, N., and Cirillo, L.A. (2016). Loss of Interdependent Binding by the FoxO1 and FoxA1/A2 Forkhead Transcription Factors

Culminates in Perturbation of Active Chromatin Marks and Binding of Transcriptional Regulators at Insulin-sensitive Genes. J. Biol. Chem. 291, 8848–8861.

364. Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., and Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol. Cell 21, 227–237.

365. Yan, B., Neilson, K.M., Ranganathan, R., Maynard, T., Streit, A., and Moody, S.A. (2015). Microarray identification of novel genes downstream of Six1, a critical factor in cranial placode, somite and kidney development. Dev. Dyn. Off. Publ. Am. Assoc. Anat. 244, 181–210.

366. Yang, Z., Yik, J.H.N., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell 19, 535–545.

367. Yin, H., Price, F., and Rudnicki, M.A. (2013). Satellite cells and the muscle stem cell niche. Physiol. Rev. 93, 23–67.

368. Yu, Y., Khan, J., Khanna, C., Helman, L., Meltzer, P.S., and Merlino, G. (2004). Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein Six-1 as key metastatic regulators. Nat. Med. 10, 175–181.

369. Yu, Y., Davicioni, E., Triche, T.J., and Merlino, G. (2006). The homeoprotein six1 transcriptionally activates multiple protumorigenic genes but requires ezrin to promote metastasis. Cancer Res. 66, 1982–1989.

370. Zammit, P.S., Relaix, F., Nagata, Y., Ruiz, A.P., Collins, C.A., Partridge, T.A., and Beauchamp, J.R. (2006). Pax7 and myogenic progression in skeletal muscle satellite cells. J. Cell Sci. 119, 1824–1832.

371. Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. 37, 76–81.

372. Zaret, K.S., Caravaca, J.M., Tulin, A., and Sekiya, T. (2010). Nuclear mobility and mitotic chromosome binding: similarities between pioneer transcription factor FoxA and linker histone H1. Cold Spring Harb. Symp. Quant. Biol. 75, 219–226.

373. Zhang, W., Behringer, R.R., and Olson, E.N. (1995). Inactivation of the myogenic bHLH gene MRF4 results in up-regulation of myogenin and rib anomalies. Genes Dev. 9, 1388–1399.

374. Zhang, X., Patel, S.P., McCarthy, J.J., Rabchevsky, A.G., Goldhamer, D.J., and Esser, K.A. (2012). A non-canonical E-box within the MyoD core enhancer is necessary for circadian expression in skeletal muscle. Nucleic Acids Res. 40, 3419–3430.

375. Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLOS ONE 9, e78644.

376. Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu. Rev. Biochem. 81, 119–143.

377. Zhu, C.C., Dyer, M.A., Uchikawa, M., Kondoh, H., Lagutin, O.V., and Oliver, G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. Dev. Camb. Engl. 129, 2835–2849.

378. Zimmerman, J.E., Bui, Q.T., Steingrímsson, E., Nagle, D.L., Fu, W., Genin, A., Spinner, N.B., Copeland, N.G., Jenkins, N.A., Bucan, M., et al. (1997). Cloning and characterization of two vertebrate homologs of the Drosophila eyes absent gene. Genome Res. 7, 128–141.

379. Zingg, J.M., Pedraza-Alva, G., and Jost, J.P. (1994). MyoD1 promoter autoregulation is mediated by two proximal E-boxes. Nucleic Acids Res. 22, 2234–2241.

380. Zou, D., Silvius, D., Fritzsch, B., and Xu, P.X. (2004). Eya1 and Six1 are essential for early steps of sensory neurogenesis in mammalian cranial placodes. Dev. Camb. Engl. 131, 5561–5572.

381. Zou, Y., Zhang, R.-Z., Sabatelli, P., Chu, M.-L., and Bönnemann, C.G. (2008). Muscle interstitial fibroblasts are the main source of collagen VI synthesis in skeletal muscle: implications for congenital muscular dystrophy types Ullrich and Bethlem. J. Neuropathol. Exp. Neurol. 67, 144–154.

CONTRIBUTION OF COLLABORATORS

List of affiliations

1. Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada
2. Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada
3. Ottawa Hospital Research Institute, Ottawa, ON, Canada
4. Departments of Medicine, and Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada

List of Contributors

1. Alexandre Blais[1,2] (*All Manuscripts*)

Dr. Blais is my PhD thesis supervisor that vastly contributed to the writing, editing and finalizing of manuscripts one and two. Dr. Blais also contributed to the critical reading and editing of the manuscript three as well as my thesis. Dr. Blais has performed bioinformatics studies and contributed to finalize figures (Figures 2.1, 2.2, 3.5A) and tables (Tables 2.1 and 2.2) in the manuscript one. He also purified recombinant Six1 protein for all the electrophoresis mobility shift assays (EMSA) in the manuscript two. In addition, he performed luciferase assays for Figure 3.5B. He also helped with Figure 4.4C using the R program in the manuscript three.

2 and 3. Dr. Soumyadeep Nandi[1,2] and Ilya Ioshikhes[1,2] (*Manuscript One*)

Dr. Nandi was a postdoctoral fellow in Dr. Ilya Ioshikhes' lab. He optimized the MEF3 DNA motif using the bioinformatics approach for Figure 2.5A. Dr. Ioshikhes is currently a professor in our institute.

4. **Andre Martel[2] (*Manuscript One*)**

Andre was a 2010-2011 fourth year Honors student in our lab. He helped me to optimize the experimental conditions to purify recombinant Six1.

5. **Alen Antoun[2] (*Manuscript One*)**

Alen was my 2011-2012 fourth year Honors student in our lab. He performed EMSA experiments and contributed to Figure 2.4.

6. **Imane Chakroun[1,2] (*Manuscripts Two and Three*)**

Imane was a PhD student in our lab. She performed ChIP assays for Figures 3.5F and 3.5G. She also contributed to Six1 ChIP-seq data.

7. **Dabo Yang[1,2] (*Manuscript Two*)**

Dabo is our lab technician who performed immunostaining on frozen muscle sections for Figure 3.1 and contributed to the following counting. He also provided technical support for the mouse work.

8. **Ellias Horner[2] (*Manuscript Two*)**

Ellias was a 2012-2013 fourth year Honors student in our lab who performed luciferase assays in primary myoblasts for Figure 2.4D.

9. **Jieyi Liang[2] (*Manuscript Two*)**

Jieyi was a 2012 summer student who contributed to the counting of immunostaining for Figure 3.1.

10 And 11. Arif Aziz[3,4] and F. Jeffrey Dilworth[3,4] (*Manuscript Two*)

Arif was a postdoctoral fellow in Dr. F. Jeffrey Dilworth's lab, who established the C2C12 stable cell line that has inducible Myod over-expression. Dr. Dilworth is currently a professor in the Ottawa Hospital Research Institute, who also contributed to critical reading of the manuscript.

12. Alphonse Chu[1,2] (*Manuscript Two*)

Alphonse was a Master's student in our lab. He performed western blots for Figure 3.2.

13. And 14. Yves De Repentigny[3,4] and Rashmi Kothary[3,4] (*Manuscript Two*)

Yves is a lab technician in Dr. Rashmi Kothary's lab. He generated transgenic mice that carry the CER+PRR reporter plasmid for Figure 3.4B. Dr. Kothary is currently a professor in the Ottawa Hospital Research Institute.

15. Elisa Bergamin (*Discussion*)

Dr. Bergamin was a postdoctoral fellow in Dr. Jean-François Couture's lab. She contributed to artificial nucleosome assembly for Figure 5.X.

COPYRIGHT PERMISSIONS

The two papers in Chapter 2 and Chapter 3 are published in open access journals.

- Chapter 2 is published in *Nucleic Acids Research*.

Liu, Y., Nandi, S., Martel, A., Antoun, A., Ioshikhes, I., & Blais, A. (2012). Discovery, optimization and validation of an optimal DNA-binding sequence for the Six1 homeodomain transcription factor. *Nucleic Acids Research*, *40*(17), 8227–8239.

- Chapter 3 is published in *PLOS ONE*.

Liu, Y., Chakroun, I., Yang, D., Horner, E., Liang, J., Aziz, A., … Blais, A. (2013). Six1 regulates MyoD expression in adult muscle progenitor cells. *PloS One*, *8*(6), e67762.