uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Yubing Liu

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biochemistry)

GRADE / DEGREE

Department of Biochemistry, Microbiology and Immunology

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Genomic Characterization of Six 1 Function During Skeletal Muscle Differentiation

TITRE DE LA THÈSE / TITLE OF THESIS

Alexandre Blais

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Jocelyn Coté

Ilona Skerjanc

Gary W. Slater

Le Doyen de la Faculté des études superieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Genomic characterization of Six1 function during skeletal muscle differentiation

A Thesis Submitted to the

Faculty of Graduate and Postdoctoral Studies

University of Ottawa

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

By

Yubing Liu

Your file Votre référence
ISBN: 978-0-494-73785-9
Our file Notre référence
ISBN: 978-0-494-73785-9

Table of Contents

List of Tables

List of Figures

Abstract

The Six family of transcription factors counts six members in vertebrates (from Six1 to Six6). Six1 is important for skeletal muscle development: *Six1*$^{-/-}$ mouse neonates die at birth because of severe muscle hypoplasia. However, the molecular targets of Six1 are poorly characterized systematically. The Myogenic Regulatory Factors (MRFs) are critical transcription factors in skeletal muscle development, and comprise MyoD, Myf5, Myogenin and Myf6. Based on previous bioinformatics predictions, it was proposed that Six1 had a genome-wide function and cooperated with MRFs by co-regulating gene expression. A functional genomics approach (ChIP-on-Chip) was used to identify targets directly bound by Six1 in the C2C12 mouse myoblast cell line at different stages of differentiation. I found that Six1 significantly targeted genes involved in several functional categories, including muscle development and muscle function, through the MEF3 DNA sequence motif. In addition, I confirmed the co-occupation of SIX1 and MRFs by bioinformatics analysis and conventional ChIP assay followed by q-PCR.

Acknowledgement

First, I thank my supervisor, Dr. Alexandre Blais, for his continuous support in my Master program. Alex was always available to listen and to discuss. He instructed me with his patience and knowledge during the entire project. He taught me how to design the experiments, how to interpret the results and how to deal with the failed trials. He showed me that there were many ways to address a question and with a good idea the technique was never a problem. He encouraged me when I was down and never hesitated to praise me when I made progress. He has open mind and also careful thinking. He is a good teacher and a nice friend. Without him, I could not have completed the project and the thesis.

I also thank my co-supervisor, Dr. Daniel Figeys, and two committee meeting members, Dr. Alain Stintzi and Dr. Mads Kaern, for their critical comments and instructive suggestions during the TAC meetings.

Let me say a special thank to Dr. Ilona Skerjanc. She is an inspiring teacher in the courses and is an amazing mentor during the weekly lab meetings. She is always showing the positive attitude to the research and exemplifies what a scientist should be.

I would like to thank my lab members. Thanks to Iman Chakroun and Alphonse Chu for their insightful discussion and technical support. They were always there when I met problems and made their best to help me to solve the troubles. Alphonse is a nice lab partner and helps me to solve many software problems. Iman and I have gone through some tough problems in the project and it was very grateful to work with her and her lovely daughter, Maria. Thanks

to the former lab members, Uzma Islam, Galen Chen and Genevieve Pelletier, with whom I spent very happy time, for their support in this project.

I thank all our collaborators. Thanks to Dr. Pascal Maire for his advice on Gel Shift. He is always there to help me. Thanks to Dr. Jeffrey Dilworth for his kind gift of recombinant SIX1 protein. Thanks to our neighboring labs for all the reagents and equipments I borrowed.

During this research, I was supported by University of Ottawa for the admission/excellent scholarship and Ontario Graduate Scholarship. This project was supported by Muscle Dystrophy Association.

Last, but not least, I thank my family. Thanks to my parents, Guizhen Jiang and Naijin Liu, for giving me life, for unconditional support of my education, and for encouraging me to pursue my interest. Thanks to my brother, Zhao Liu, who is always a listener to my complaints. I owe a big thanks to my husband, Yang Song, for his mental support when I doubted myself, for his inviting me to Canada, for his yummy daily food, and for his long-lasting believing in me.

Introduction

1. Mammalian Myogenesis

The formation of skeletal muscle in vertebrates starts in the early embryogenesis. During vertebrate embryo development, there are three primary germ cell layers: ectoderm, endoderm and mesoderm. The paraxial mesoderm, the mesodermal tissue on both sides of the neural tube, separates into periodic blocks of compact coil-shaped segments, named somites. Somites develop in an anterior-posterior order and appear in pairs next to the neural tube at regular intervals. A transient epithelial structure termed the dermomyotome, derived from somites, will give rise to the myotome. The myotome is the first place where skeletal muscle forms. In myotome, muscle progenitor cells in the region closest to the neural tube develop into epaxial skeletal muscles (the deep back muscles) and those in ventro-lateral region form hypaxial muscles (the body wall, diaphragm, limbs and tongue). The head musculature comes from the cranial paraxial mesoderm. The limb skeletal muscle originates from the hypaxial dermomyotome of the limb-level somites where muscle precursor cells migrate to the corresponding regions of limb buds. Myoblasts proliferate and then differentiate into multinucleated myotubes when cells exit the cell cycle to form the limb musculature (Bryson-Richardson & Currie, 2008; Le Grand & Rudnicki, 2007).

In general, the embryogenesis processes in vertebrates are thought to be very similar. In fact, most genes involved in embryogenesis are highly conserved among mammals, especially between humans and mouse. Most of our knowledge about skeletal muscle development is based on mouse studies. In the mouse, embryogenic skeletal myogenesis occurs in three main phases: 1) the formation of the primary myotome, which consists of differentiated

muscle cells delaminating from the epithelial lips of the dermomyotome; the primary myotome is the scaffold for subsequent myogenesis to take place; 2) the development of the second myogenic phase, which is driven by a number of progenitor cells with high proliferative potential; those progenitor cells give rise to most embryonic muscles; 3) the production of satellite cells, which are muscle progenitor cells accounting for regulating post-natal muscle growth and regeneration. Satellite cells and progenitor cells in the second myogenic phase are derived from the same central region of the dermomyotome (Bryson-Richardson & Currie, 2008; Gros, Manceau, Thome, & Marcelle, 2005; Relaix, Rocancourt, Mansouri, & Buckingham, 2005).

At the post-natal stage, satellite cells reside between the basement membrane and sarcolemma of adult myofibers. Satellite cells are mono-nucleated and mitotically quiescent myogenic cells and they are responsible for growth and regeneration of muscles after injury or physical exercise (Montarras et al., 2005). Satellite cells derive from Pax3/Pax7-positive (paired box proteins 3 and 7) cells in the central region of the dermomyotome, where they transit from epithelial to mesenchymal cells (Relaix et al., 2005). Satellite cells function as stem cells in their niche, where they can precisely balance between self-renewal and proliferation when needed (Collins et al., 2005; Le Grand & Rudnicki, 2007).

As a complex process, skeletal muscle development, including the embryonic stage and post-natal stage, is under the control of a certain number of transcription factors (Fig. 1). They temporally and spatially govern the expression of their target genes along the orchestrated myogenesis (Bryson-Richardson & Currie, 2008; Buckingham, 2006). In general, it is the Pax (paired box protein) family members that specify muscle precursor cells and it is the myogenic regulatory factors (MRFs) family members that determine myogenic fate,

including myogenin (Myog), myogenic differentiation 1 (MyoD), myogenic factor 5 (Myf5) and myogenic factor 6 (Myf6, also termed Mrf4) (Fig. 1). Six1 and Six4, the sine oculis-related homeobox 1 and 4 homolog (*Drosophila*) belonging to the Six family, are crucial for skeletal muscle development at the embryonic stage and are still functional at the adult stage. Both Six1 and Six4 functionally interact with Pax family and MRF family members.

Figure 1 the genetic network in embryonic skeletal myogenesis

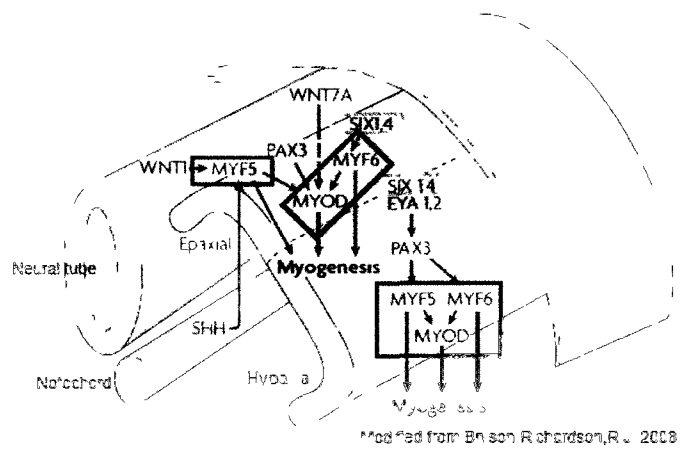


Fig.1| A regulatory network of the signaling pathways and transcription factors involved in murine embryonic myogenesis is shown here, which was summarized in the review by Bryson-Richardson and Currie (2008). Pax3, MRFs, and Six1/4 are exemplified as the regulatory network of myogenesis. The arrows mean the induction relationship. In epaxial muscle, WNT signaling, PAX3, MYF5 and MYF6 induce the expression of Myod independently (Borello et al., 2006; Tajbakhsh, Rocancourt, Cossu, & Buckingham, 1997; Tajbakhsh et al., 1998). SIX1/4 and SHH signaling indirectly induce the expression of Myod (Borycki et al., 1999; Grifone et al., 2005). In hypaxial muscle, Myod is under the indirect control of PAX3 (Bajard et al., 2006). SIX1/4 and EYA1/2 regulate the expression of Pax3

(Bryson-Richardson & Currie, 2008; Bryson-Richardson & Currie, 2008; Grifone et al., 2007).

2. Pax family transcription factors: Pax3 and Pax7

The Pax family encodes transcription factors which can bind sequence-specific DNA through a paired domain (Fig. 2). Pax3 and Pax7 are a subgroup of the Pax family. Pax3 and Pax7 play an important role in muscle development at the embryonic stage and in muscle regeneration at the adult stage (Buckingham & Relaix, 2007; Relaix et al., 2005). Pax3 is transcribed in paraxial mesoderm just before segmentation into somites and is consistently expressed throughout the dorsal dermomyotome, including the epaxial and hypaxial extremities (Relaix, Rocancourt, Mansouri, & Buckingham, 2004). Pax3 is required for specification and migration of mesodermal muscle precursor cells from the dermomyotome to the limb buds (Bober, Franz, Arnold, Gruss, & Tremblay, 1994; Daston, Lamar, Olivier, & Goulding, 1996; Williams & Ordahl, 1994).

Pax3 genetically regulates MRFs' expression. In *Pax3* mutant mice, MRFs transcripts (Myod, Myog, and Myf5) are not detectable in the limbs. Pax3 can directly bind the regulatory element of *Myf5* and affect its expression in hypaxial myotome, and MYF5 activates Myod expression consequently (Bajard et al., 2006). In the epaxial muscle, PAX3 and MYF5 can induce the expression of Myod independently (Tajbakhsh et al., 1997). After migration of myoblasts to the regions in limb buds, Pax3 can induce the expression of Myf5 and Myod directly, which govern the irreversible commitment of precursor cells into the myogenic lineage (Bajard et al., 2006).

Figure 2 the structure and domains of PAX proteins

Modified from Lang et al. 2007

Pax family group	Pax protein structure Paired octapeptide homeodoma n	Pax Family Member	Embryonic expression
Group I		PAX*	Skeleton, thymus, 3rd/ 4th pharyngeal pouch
		PAX9	Skeleton, teeth thymus
Group II		PAX2	Kidney, CNS
		PAX5	B-cell, CNS
		PAX8	Kidney, thyroid, CNS
Group III		PAX3	**Somites/muscle, neural crest, CNS**
		PAX7	
Group IV		PAX4	Pancreas, gut
		PAX6	Pancreas, gut, CNS, eye

Fig.2| the Pax family has 9 members. The Pax family plays a crucial role in the development of many organs. Among 9 members, Pax3 and Pax7 are expressed in the somites, neural crest and central nervous system (CNS). Pax3 and Pax7 are critical for muscle development reviewed in (Lang, Powell, Plummer, Young, & Ruggeri, 2007).

By contrast, Pax7 is mainly expressed in the central region of the dermomyotome, from which the satellite cells originate. Pax7 is a key regulator of satellite cell specification and the maintenance of satellite cell characteristics (Daston et al., 1996; Seale et al., 2000). *Pax7* mutant mice showed a severe reduction of the satellite cell number in the adult, due to their progressive loss by apoptosis (Seale et al., 2000). Over-expression of Pax7 promotes cell cycle exit and may play a crucial role in reverting activated satellite cells (proliferating

and/or cells with differentiating potential) to the quiescent and undifferentiated cells (Olguin & Olwin, 2004). PAX7 was reported to reciprocally inhibit the expression of MRFs. PAX7 can negatively affect the expression of Myod, while its own expression was down-regulated by MYOG in vivo (Olguin, Yang, Tapscott, & Olwin, 2007).

Normally, satellite cells are quiescent in their niche, expressing Pax3 and Pax7. After activation, satellite cells can divide asymmetrically: MRF positive cells replenish the injured muscle (Myod directs myoblast differentiation, and Myf5 controls myoblast proliferation and homeostasis), and Pax positive cells maintain their self-renewing characteristics (Fig. 3) (Kuang, Kuroda, Le Grand, & Rudnicki, 2007).

Figure 3 the embryonic and adult skeletal myogenesis

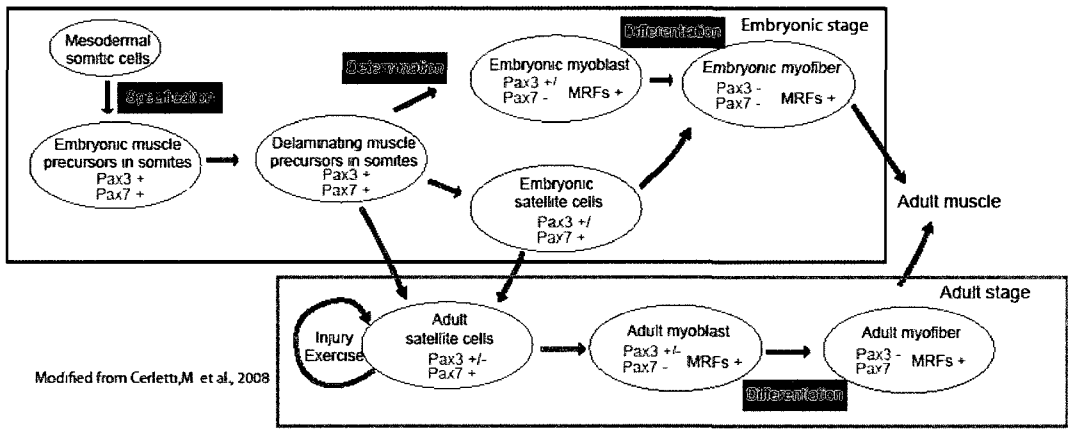


Fig.13| Pax3 is crucial for the specification of muscle precursor cells. Pax7 is critical for the maintenance of satellite cells. MRFs (Myod, Myf6 and Myf5) play a pivotal role in the determination of muscle lineage. MRFs (Myf6 and Myog) also play a substantial role in muscle differentiation (Cerletti, Shadrach, Jurga, Sherwood, & Wagers, 2008).

Accumulating evidence supports that both Pax3 and Pax7 are important for embryonic muscle development and muscle regeneration. However, their essential role in adult muscle regeneration has been challenged recently, implicating that Pax7 is only required for the maintenance of satellite cells in the early post-natal life (Lepper, Conway, & Fan, 2009; Relaix et al., 2005). The maintenance of satellite cells in adult may depend on other transcription factors, such as Six1 and Six4, or depend on the intrinsic change of chromatin. When Myod or Myf5 is continuously expressed, satellite cells will be activated to proliferate and finally differentiate into myotubes with the expression of Myog and Myf6. However, the factors driving satellite cells to repair injured muscles are still poorly understood.

3. MRF family transcription factors

The MRF family also plays a critical role in muscle development, including MyoD, Myf5, Myog, and Myf6. MRFs are basic helix-loop-helix (bHLH) transcription factors. Each member of the MRF family has a bHLH domain and an E-box binding domain (recognizing the core consensus sequence, CANNTG). The MRF family, belonging to class B bHLH factors, is specifically expressed in muscle. Class A bHLH factors, such as E proteins, are ubiquitously expressed among tissues. MRF family members can homodimerize or heterodimerize with ubiquitous class A bHLH factors (E protein) to bind paired E-box DNA sequence sites (Hamamori, Wu, Sartorelli, & Kedes, 1997; Weintraub, Davis, Lockshon, & Lassar, 1990).

Myod was discovered based on its ability to induce the myogenic conversion of 10T1/2 fibroblasts (Davis, Weintraub, & Lassar, 1987). Later on, Myod was reported to be capable of converting a wide range of other cell types to skeletal muscle (Weintraub et al., 1989).

The following studies revealed that all four MRFs could convert many non-muscle cells into skeletal muscle in vitro cell culture, though with different conversion efficiency (Aurade, Pinset, Chafey, Gros, & Montarras, 1994; Braun, Buschhausen-Denker, Bober, Tannich, & Arnold, 1989; Braun, Bober, Winter, Rosenthal, & Arnold, 1990; Weintraub et al., 1989; Wright, Sassoon, & Lin, 1989).

MyoD, Mrf4 and Myf5 are key factors for commitment of multi-potent somite cells to the myogenic lineage. Myod and Myf5 have redundant roles in the development of skeletal muscle since loss-of-function of either one has no major morphological abnormalities in skeletal muscle in newborn mice (Braun, Rudnicki, Arnold, & Jaenisch, 1992; Rudnicki, Braun, Hinuma, & Jaenisch, 1992; Rudnicki et al., 1993). Muscle progenitors expressing Myf5 give rise to epaxial muscles (in the back and body wall) and those expressing Myod contribute to hypaxial muscles (in limb buds, tongue, diaphragm and head) (Kablar et al., 1998). Myf6 has potential roles in both muscle determination and differentiation (Montarras et al., 1991). Myf6 acts upstream of Myod and activates Myod expression in *Myf5* null mice (Kassar-Duchossoy et al., 2004). Myog does not have overlapping functions with Myod and Myf5. Myog is critical for the terminal muscle differentiation by activating the expression of a series of muscle-specific genes such as Des (Desmin) and Chrna1 [cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)] (Blais et al., 2005; H. Li & Capetanaki, 1993). *Myog*-null mice have a severe deficiency of terminal muscle differentiation with few mature myofibers. However, Myod can convert fibroblasts to muscles in the absence of functional Myog, which suggests a Myog-independent differentiation process that may be due to Myf6 (Hasty et al., 1993; Nabeshima et al., 1993; Rawls et al., 1995). The four MRFs appear to have distinct but overlapping transcriptional targets, which is also supported by the

identification of their binding sites using genomic approaches (Blais et al., 2005; Cao et al., 2006; H. Li & Capetanaki, 1993).

To activate the expression of MRFs, signals from different upstream regulators are integrated in the myotome, reviewed in (Tapscott, 2005). Many important developmental signaling pathways from the surrounding tissues of myotome have been shown to regulate the expression of MRFs in the somite (Fig. 1). Sonic hedgehog secreted from the notochord has a regulatory effect on the development of epaxial muscle through the induction of Myf5 (Borycki et al., 1999; Kruger et al., 2001). Wnt signaling from the dorsal neural tube regulates myogenesis through the activation of MRFs: Myf5 and Myod (Borello et al., 2006; Tajbakhsh et al., 1998). MRFs are the reported targets of PAX3 in the somites (Bajard et al., 2006). Myod is capable of inducing its own expression and is downstream of MYF5 (Tajbakhsh et al., 1997; Thayer et al., 1989). Myog is a genetic target of Myod and Myf5. In the epaxial and hypaxial myotome, MRFs are regulated by members of the Six family, which are also critical for muscle development (Fig. 1).

4. Six Family transcription factors

Extensive embryology studies indicate that the Six family plays a pivotal role in embryonic development. The Six gene family is the murine sequence homolog of *sine oculis* (*so*) related homeobox in *Drosophila*. In *Drosophila*, *so* functions within a network to regulate the compound eye formation, and is sufficient to induce ectopic eye formation (Oliver et al., 1995; Weasner, Salzer, & Kumar, 2007). The network consists of *eyeless* (the Pax family in mammalian), *eyes absent* (the Eya family in mammalian), *dachshund* (the Dach family in mammalian) and *so* (Fig. 4). In mouse, the Six family has 6 members (Six1-Six6), of which

all are believed to evolve from a common ancestor (Kumar, 2009). SIX proteins are characterized by a Six-type homeobox nucleic acid recognition domain (SHD, 60 amino acids in length) and a Six-domain (SD, 146-150 amino acids in length), both of which are evolutionarily conserved (Fig. 5).

Figure 4 the gene network of the eye development in *Drosophila*

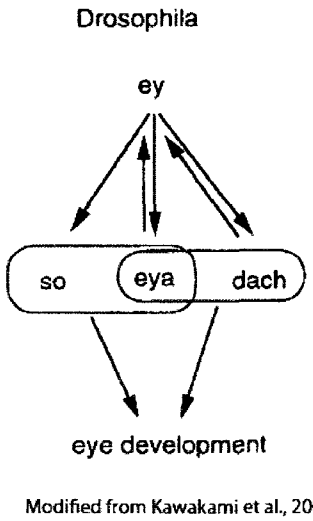


Fig.4| The gene network in *Drosophila*: *ey-so-eya-dach* (Kawakami, Sato, Ozaki, & Ikeda, 2000). *Drosophila ey* is a homologue of vertebrate Pax6. *Drosophila so* is a homologue of vertebrate Six1. *Drosophila eya* is a homologue of Eya vertebrate. *Drosophila dach* is a homologue of vertebrate Dach. Arrows show induction pathways. Six and Eya or Dach and Eya can directly interact with each other to form a complex and have synergic function on its targets (Ikeda et al., 2007; X. Li et al., 2003).

Figure 5 the structure and domains of Six proteins

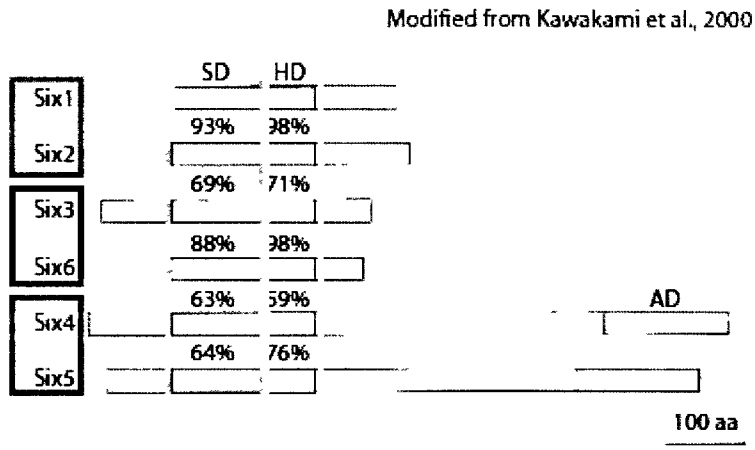


Fig.5| The structure and domains of Six family transcription factors (Kawakami et al., 2000). The Six family is the *sine oculis*-related homeobox homologues (*Drosophila*). The Six family is involved in the development of many organs. Among the 6 members, Six1 and Six4 are crucial for muscle development.

Comparing the similarity of the amino acids in SHDs and SDs, Six genes were classified into three subfamilies, *Six1/2*, *Six3/6*, and *Six4/5*. Only the subgroup of *Six4/5* encodes a specific transactivation domain (TD) in the C-terminus, which is downstream of SHD and SD, reviewed in (Kawakami et al., 2000; Kumar, 2009). SHD, contributing to DNA binding, is an atypical homeodomain and diverges from the universal homeodomain, which allows members of the Six family to have a specific DNA binding activity other than the typical homeodomain recognizing sequence (the "TAAT" core motif) (Noyes et al., 2008). Six1/2 and Six4/5 have the similar binding motif while Six3 subgroup may bind to a different DNA motif containing an ATTA core motif (Kumar, 2009; Zhu et al., 2002). Both the SD and SHD seem critical for protein-protein interaction (Ohto et al., 1999; Ruf et al., 2004).

Six members have distinct roles in different developmental contexts, depending primarily on their binding partners. The Six family is a node in the Pax-Eya-Six-Dach regulatory network, which plays an important role in the development of many organs. Pax3 is required for the induction of Six1 and Eya2 in P19 embryonic carcinoma cells (Ridgeway & Skerjanc, 2001). SIX can physically interact with EYA protein to regulate the expression of their downstream genes (Ikeda, Watanabe, Ohto, & Kawakami, 2002; X. Li et al., 2003; Zhang & Stavnezer, 2009). Besides the direct interaction between these two families, Li X. et al found that the phosphatase activity of EYA can help to switch the repressive function of Six to an active function (X. Li et al., 2003). Six proteins can help the nuclear translocation of Eya protein (Kawakami et al., 2000). Over-expression of Six and Eya has a synergic activation on myogenic genes, such as *Myog* (X. Li et al., 2003). Six family members can interact with transcriptional repressors, such as the Groucho/TLE family and the DACH family (Ikeda et al., 2002; Lopez-Rios, Tessmar, Loosli, Wittbrodt, & Bovolenta, 2003). Dach1 and Dach2, which can physically interact with nuclear receptor co-repressor (N-CoR) and histone deacetylase 3 (HDAC3), can enhance the repressive effect of Six6 on its target genes (X. Li, Perissi, Liu, Rose, & Rosenfeld, 2002; Zhu et al., 2002).

Six1 expression is first detected at E8.2-8.5 in the mouse embryo. Six1 is present in somites, head mesenchyme, dermomyotome, branchial arches, nasal placode, otic vesicle, nephrogenic chord and limb buds (Laclef et al., 2003; Laclef, Souil, Demignon, & Maire, 2003; Oliver et al., 1995). After E13.5, Six1 expression becomes weak in head and trunk muscles until E16.0 (Oliver et al., 1995). However, the expression of Six1 in muscles is persistent from embryonic myogenesis to the adult period (Laclef, Souil et al., 2003). *Six1*-mutant mice died at birth, which is due to their inability to breathe: in those mutant mice, the

diaphragm muscle was absent and the thoracic cage was deformed. *Six1*-mutant mice had severe muscle hypoplasia affecting especially hypaxial muscles (Laclef et al., 2003). Loss of function of Six1 in a mouse model revealed that Six1 was required for the development of murine kidney, thymus, diaphragm, parotid gland, lacrimal gland, hypaxial muscle, and sensor organs (inner ear, nose cavity, tongue and otic vesicle) (Laclef, Souil et al., 2003; Xu et al., 2003). Taken together, these findings reveal that Six1 is critical for embryonic myogenesis and is important for the development of several other organs.

The expression pattern of Six4 highly overlaps with that of Six1 at the embryonic stage (Ozaki et al., 2001). However, *Six4*$^{-/-}$ mice developed without major detectable defect, which implies compensation by other family members (Grifone et al., 2005; Ozaki et al., 2001). *Six1/4* double mutant mice had a more severe phenotype than *Six1* null alone. In *Six1/4* double null mice, Pax3 expression was impaired and MRFs expression was extensively reduced, while they were not or moderately affected in *Six1*-null mice (Grifone et al., 2005). It is hypothesized that Six4 may have a complementary role to Six1.

Six1 function could be regulated by post-translation modifications. As an evolutionarily conserved gene, SIX1 in human is specifically hyperphosphorylated in mitosis, possibly by casein kinase II (CK2), which is a ubiquitous protein kinase. As a consequence, reduced DNA binding affinity of SIX1 was observed (Ford et al., 2000). Eya proteins have phosphatase activity, which is implicated in the conversion of SIX/DACH complex from repression to activation through the recruitment of co-activators (X. Li et al., 2003). SIX1 protein is reported to be modulated by APC^{cdh1}-mediated degradation during the cell cycle, which suggests a potential role of Six1 in cell cycle regulation (Christensen, Brennan, Aldridge, & Ford, 2007).

Previous studies reported that mutations in *Six1* can cause abnormality and disease in the human and the mouse. In the human, *Eya1* mutations cause branchio-oto-renal (BOR) syndrome, while *Six1* mutations are linked to similar symptoms in BOR patients (Abdelhak et al., 1997; Chang et al., 2004; Ruf et al., 2004). BOR is an autosomal dominant genetic disorder, featured by abnormal ears, neck and kidneys (Kochhar, Fischer, Kimberling, & Smith, 2007). In the mouse, mutations in the SD and SHD of *Six1* disrupt both its interaction with Eya1 and its binding to DNA (Kochhar et al., 2007; Ruf et al., 2004). A missense mutation in the SHD of *Six1* was found in the catweasel (Cwe) phenotype mice, which affects sensory systems and resembles the human BOR syndrome. The BOR syndrome is less likely to be caused directly through Six1 since there is no detectable defect in muscle tissues. However, it is more likely due to its failure to recruit Eya protein to the corresponding DNA loci. However, a direct role of Six1 in BOR cannot be completely excluded. The mutations in *Six1* abolish the synergistic activation of the *Myog* promoter by SIX1 and EYA1 in vitro (Ruf et al., 2004). Other SIX members may compensate for Six1 function when *Six1* is mutated. Six4 and Six5 can interact with EYA proteins as well (Ikeda et al., 2002; Ohto et al., 1999).

Moreover, accumulating evidence showed that Six1 was implicated in cancer initiation and metastasis. Six1 over-expression was observed in numerous human cancers, especially in metastatic lesions (Ford et al., 2000). Six1 and Ezrin were over-expressed in highly metastatic rhabdomyosarcoma (RMS) cell lines, and Six1-induced metastasis was through Ezrin, which was important for cell surface structure adhesion and migration (Yu et al., 2004; Yu, Davicioni, Triche, & Merlino, 2006). Over-expression of Six1 in mammary carcinoma cells can positively regulate the expression of Ccna1 (cyclin A1) and promote the

G2/M transition, and thus increase cell proliferation (Coletta et al., 2008). SIX1 can also attenuate the G1/S check point by directly regulating the expression of Myc (myelocytomatosis oncogene, also known as c-Myc) and Ccnd1 (Cyclin D1) (X. Li et al., 2002; Yu et al., 2006). Six1 was capable of inducing epithelial-mesenchymal transition (EMT) and then increasing the invasive capacity of tumor cells (Coletta et al., 2008). The activation of cyclin D1 promoter was repressed by DACH1 and was induced by SIX1 in HEK293T, a human embryonic kidney cell line (Wu et al., 2006). Taken together, these observations showed that Six1 was not only involved in embryonic development, but was also implicated in cell proliferation in cancers which often mimic the embryonic stage.

Using classical biological approaches, some of Six1 targets were discovered. In Six1/4 double mutant mice, a list of genes were found to be down-regulated in embryonic axial tissues in a gene expression profiling study (Niro et al., 2010). This list included structural muscle genes, such as *Tnnc2* (troponin C2, fast) and *Tnnt3* (troponin T3, skeletal, fast). Some genes encoding the membrane channel subunits were also among the list, such as *Atp2a1* (ATPase, Ca++ transporting, cardiac muscle, fast twitch 1), and *Kcne1l* (potassium voltage-gated channel, Isk-related family, member 1-like). In an earlier study by the same group, *Six1/4* mutant mice showed reduced or absent expression of transcription factors in somites and limb buds, such as MRFs, Lbx1 [ladybird homeobox homolog 1 (Drosophila)], Met (met proto-oncogene) and Pax3, which are important for muscle development (Grifone et al., 2005). A study from another group using *Six1/4* double mutant mice showed that genes involved in kidney development were down-regulated in metanephric mesenchyme, including *Pax2*, *Pax8*, and *Gdnf* (glial cell line derived neurotrophic factor) (Kobayashi, Kawakami, Asashima, & Nishinakamura, 2007). Reduced expression of Myf5 was observed

in Six1 mutant mice and the direct binding of Six1 on the regulatory region of Myf5 was confirmed in C2C12 (a murine myoblast cell line) (Giordani et al., 2007). Six1 was reported to be able to directly regulate the expression of Ccna1 in MCF7 cells (a human breast cancer cell line). The increasing list of genes affected by Six1 covers genes in many functional categories. Although the binding of SIX1 to a few targets was confirmed by in vivo and/or in vitro techniques, it remains inconclusive whether the regulatory role of SIX1 (especially in muscle development) is through a direct regulation of gene transcription or it is through an indirect regulation of gene transcription via Six1 down-stream targets, such as MRFs. Therefore, a genome-wide mapping of Six1 binding sites becomes necessary to understand its role.

Although the function of Six1 was intensively studied in embryonic myogenesis, its role in muscle regeneration at the adult stage is still poorly characterized. Besides its clear critical role in embryonic myogenesis, Six1 was still expressed and functional in adult skeletal muscles to establish the fast-twitch muscles (Grifone et al., 2004). The muscle regeneration program, the differentiation of myoblast to myotubes, is critical for muscle repair after birth. Six1 was reported to be able to regulate the expression of structural and functional muscle genes at the embryonic stage (Niro et al., 2010). It is possible that the regulatory function of Six1 on these muscle genes persists until the adult stage. However, to this date, the molecular function of Six1 in adult muscle is not well studied.

To explore the role of Six1 in adult muscle development and function, C2C12, a well-established mouse myoblast cell line, was chosen as the muscle regeneration model. C2C12 cells were originally isolated by Yaffe and Saxel from the injured thigh muscle and were subcloned by Blau et al. by its strong myogenic potential to differentiate to myotubes (Blau,

Chiu, & Webster, 1983; Yaffe & Saxel, 1977). These cells can proliferate as myoblasts and can be induced to differentiate into myotubes by growth factor withdrawal. C2C12 cells have been extensively studied and have a well-defined time course of differentiation. The ease with which these cells can be propagated makes it possible to collect enough chromatin for large scale experiments at different stages of their differentiation program.

5. ChIP-on-Chip

In order to study the function of Six1 at the genome-wide level, the main approach proposed is to identify its direct target genes in a global manner. To achieve this, ChIP-on-Chip was used in this study (Fig. 6). ChIP-on-Chip is a DNA tiling array analysis of protein-DNA interaction. It was first developed by the group of Dr. Young at the Whitehead Institute (Ren et al., 2000). They combined the ChIP (chromatin immunoprecipitation) with DNA chips (microarrays) to study protein-DNA interactions in the yeast. Also known as the genome-wide location analysis, this approach can provide a systematic view of the binding events of a certain protein to DNA, usually a transcription factor, in order to identify its target genes and eventually help to establish a comprehensive global regulatory network.

ChIP-on-Chip allows mapping of binding events to discover sets of DNA sequences bound by a specific transcription factor. The binding of a transcription factor to DNA is restricted spatially and temporally, which may be due to either the state of genomic DNA (chromatin condensation, histone acetylation, and/or methylation), or the nuclear abundance of the factor itself, or its post-translational modification (phosphorylation, acetylation, SUMOylation, and ubiquitylation). In combination with certain bioinformatics tools, ChIP-on-Chip can help identify the common and unique targets at different developmental stages and in different

tissues, which reflects the modulated binding of a transcription factor to the genome. Gene expression profiling combined with loss of function or ectopic expression, offers the functional impact of a transcription factor on target genes. Integrating ChIP-on-Chip and gene expression profiling data together provides an effective way to validate predicative function and to discover new features of the factor under study.

Figure 6 the workflow of the ChIP-on-Chip procedure

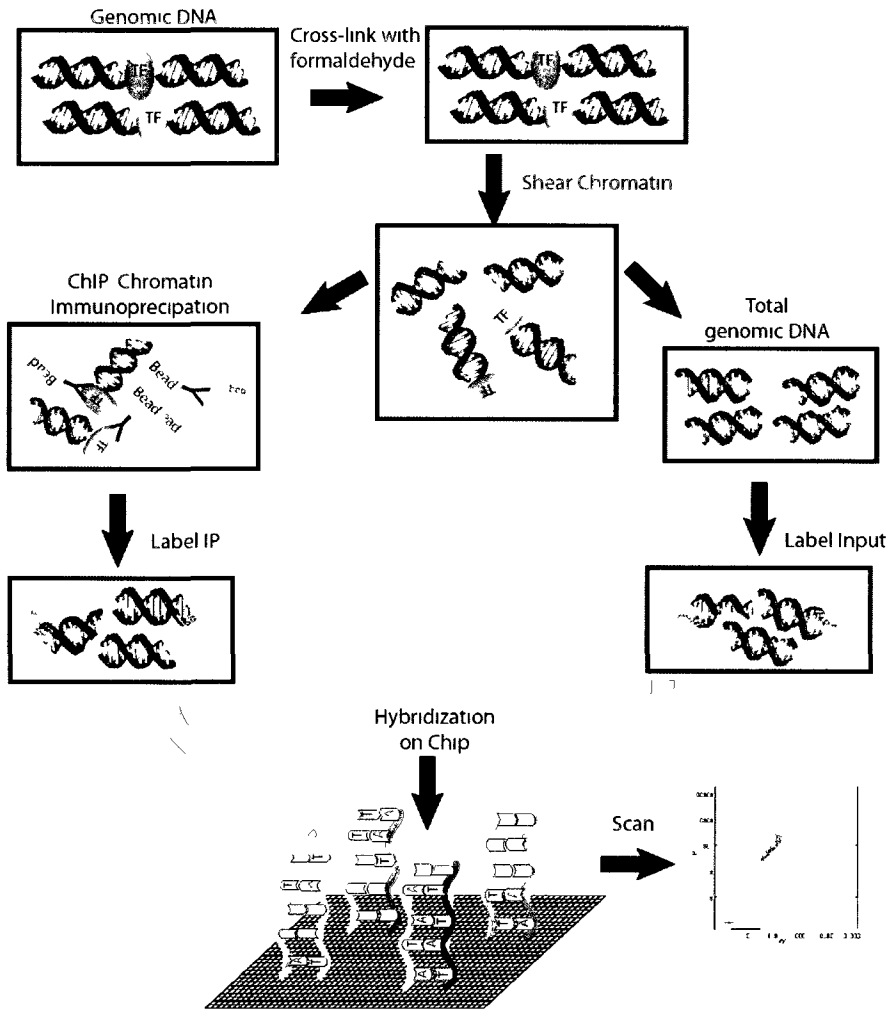


Fig.6| The interaction of protein and DNA is cross-linked along the genome. Genomic chromatin is sheared into small fragment. A proportion of fragmented chromatin is immunoprecipitated by a specific antibody (IP), and another proportion is used as the background control (Input). The DNA in both IP and Input is purified, amplified and then labeled with different fluorescent molecules. Labeled IP and Input are co-hybridized onto the same microarray slides. The fluorescent signals are detected and analyzed. Several bioinformatics tools are used to analyze the data.

Hypothesis and aims

Skeletal muscle development is a complex and precisely regulated process, involving the orchestration of many transcription factors. Six1 plays a crucial role in muscle development and several types of organ development during embryogenesis. In adult muscle, Six1 is still present and functional. MRF family transcription factors are crucial for skeletal muscle development at both embryonic and adult stages. The previous studies suggest that Six1 may function in combination with MRF family members. Several studies have discovered the genomic binding sites of MRFs. However, Six1 function is poorly characterized at the molecular level and at the genome-wide scale in adult skeletal muscle.

In this study, I hypothesize that:

1. Six1 is functional in adult muscle development.
2. Six1 functionally binds common targets with MRF family transcription factors.

Specific aims:

1. Genomic mapping of Six1 binding sites using ChIP-on-Chip in C2C12 cells.
2. Annotate Six1 binding sequences to related genes.
3. Perform gene function analysis for annotated Six1 target genes
4. Identify enriched DNA sequence motifs among Six1 binding sequences
5. Validate some of Six1 targets and other findings using ChIP assays and other approaches
6. Comparison of Six1 genomic binding data with other genomic data,

Material and Methods

1. Plasmid Constructions

To generate SIX1 protein to immunize a rabbit, the mouse Six1 coding sequence (encoding full-length protein) was fused to the N-terminal hexa-histidine plasmid (pHIS2 plasmid) (a kind gift from Dr. J.-F. Couture, U. of Ottawa), which allowed the bacterial production of protein bearing an N-terminal 6*HIS tag followed by the TEV protease cleavage site.

In order to generate recombinant SIX1 protein to purify the anti-SIX1 antibody, the mouse Six1 coding sequence (encoding full-length protein) was subcloned into pGST2, creating a fusion protein of the glutathione-S-transferase (GST) followed by the TEV cleavage site and the amino acid sequence of mouse Six1.

To over-express SIX1 in C2C12, the mouse SIX1 coding sequence cloned into the p3xFlag-myc plasmid (Sigma) was a kind gift from G. Merlino (Yu et al., 2006). The EGFP coding sequence was cloned to the p3xFlag-myc plasmid as the negative control. Using the retrovirus strategy, C2C12 cells were stably transfected with either FLAG-SIX1 or FLAG-EGFP after selection using 400μg/ml Puromycin. Plasmids were sequenced by the TCAG sequencing facility (Toronto) to ensure that no mutations were introduced.

2. Recombinant Protein Expression

BL21 (DE3) pLysS strain of *Escherichia coli* (*E. coli*) cells were used to produce recombinant HIS-SIX1 protein. BL21 *E. coli* cells were induced by adding 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1 hour at 37 °C. *E. coli* cells were collected by centrifugation at 5000 *g* and were then treated with one freeze-thaw cycle in liquid nitrogen. Cell pellets were lysed in lysis buffer, containing 500 mM NaCl, 6 M Urea,

40 mM imidazole, and 1 mM beta- mercaptoethanol (BME) and phosphate buffer solution (PBS, 137 mM NaCl, 2.7 mM KCl, 1.76 mM KH_2PO_4 and 10 mM Na_2HPO_4, pH 7.4). Cell lysates were sonicated in ice-cold water-bath at moderate intensity with 1 second on and 4 seconds off until the cell suspension had cleared up. The soluble supernatant was collected after centrifugation at 15000 *g* at 4 °C and was then filtered with 0.45 μm filter. The supernatant was applied to Nickel Sepharose affinity Gel (GE Healthcare). The column was washed with lysis buffer and was then eluted with lysis buffer containing 500 mM imidazole. The eluate was dialyzed against 1L of dialysis buffer (25 mM $NaHCO_3$ at pH 11.0, 1 M Urea, 150 mM NaCl, and 1 mM EDTA) for 4 times at 4 °C and each time for 1 hour. After dialysis, precipitates were removed by centrifugation at 2500 *g* for 10 min and the supernatant were concentrated on Centricon 10K at 4 °C. The concentration of soluble supernatant was measured by bicinchoninic acid assay (BCA) method following manufacture's manual (Pierce Chemicals).

BL21 strain of *E. coli* cells were used to produce recombinant GST-SIX1 protein. *E. coli* cells were induced by adding 0.05 μM of IPTG at 16 °C for 16 h. *E. coli* cells were collected by centrifugation at 5000 *g* and were then resuspended in lysis buffer [5 mM Dithiothreitol (DTT), 1 mM phenylmethanesulfonylfluoride (PMSF) and PBS]. The cell suspension was treated with one freeze-thaw cycle in liquid nitrogen. 1% of Triton was added to the suspension and two more freeze-thaw cycles were repeated. Cell lysates were sonicated in ice-cold water-bath at moderate intensity with 1 second on and 4 seconds off until the cell suspension had cleared up. Soluble supernatant were collected after centrifugation at 12,000 *g* at 4 °C for 15min.

Immobilized glutathione agarose beads were resuspended in GST-SIX1 supernatant, rocking at 4 °C for 16 h and then at room temperature (RT) for 1 h (Pierce Chemicals). The column was washed with lysis buffer and then washed with PBS containing 1 mM PMSF. Disuccinimidyl Suberate (DSS) solution was applied to GST-SIX1 column to cross-link protein-protein interaction. The column was rocked at RT for 1 h and was then blocked with 2 ml of 0.1 M Tris (pH 8.0), rocking at RT for 15 min. The column was washed with washing buffer [50 mM Tris (pH 8.0), 150 mM NaCl, and 1% Triton] and was then eluted with washing buffer containing 20 mM Glutathione (GSH).

3. Antibody Production

Purified 6*HIS-SIX1 protein was used for immunizing each rabbit at Spring Valley (Open Biosystems, Woodbine, MD). Rabbit sera were collected for purification. Tris buffered saline (TBS, 20 mM Tris pH 8.0 and 150 mM NaCl) was added to the sera (the final concentration of TBS is 20 mM, pH8.0 and 150 mM NaCl). The sera were filtered with 0.22 μm filter and applied to GST-SIX1 coupled immobilized glutathione agarose beads (Pierce Chemicals). The mixture of sera and agarose beads was incubated at 4 °C for 16 h. The column was washed with TBS. Anti HIS-SIX1 antibody was released by adding 0.1 M Glycine (pH 2.8). The eluate was neutralized immediately by adding 0.1 M Tris (pH 8.0) and was then dialyzed against 2L of PBS at 4 °C twice: for 16 h and then for 4 h. The soluble eluate after dialysis was collected by centrifugation at 2500 *g* and further concentrated with Millipore-Microcon (100kD) (Millipore) following the manufacture's protocol. BCA (Pierce Chemicals) was performed to determine the protein concentration. The crude antibody and affinity purified antibody were tested on C2C12 cell lysates by western blotting.

4. Cell culture

C2C12 myoblasts were acquired from the American Type Culture Collection (ATCC) and were grown in growth medium (GM), containing DMEM High Glucose (Hyclone), 10% characterized Fetal Bovine Serum (Hyclone), 2 mM L-Glutamine (Hyclone) and 100 unit / L Penicillin / Streptomycin (Hyclone). To induce differentiation, cells were cultured in differentiation medium (DM), consisting of DMEM High Glucose, 2% Horse Serum (Invitrogen), 2 mM L-Glutamine and 100 unit / L Penicillin / Streptomycin (Blais et al., 2005). Of note, not all myoblasts can be differentiated into myotubes. To get pure myotubes, 1:20 diluted Trypsin in PBS (Invitrogen, final concentration 0.0025%) was used to separate myotubes from the undifferentiated reserve cells.

5. Nuclear extracts

C2C12 cells were grown to confluence in GM, induced to differentiate in DM and harvested 24 h after differentiation. Cells were scraped in ice-cold PBS. Packed cell volume (PCV) was determined after centrifugation at 1000 *g* at 4 °C for 10 min. Cells were resuspended in 5 PCV volume of buffer A, which contains10 mM HEPES/K^+ pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, and 1x PIC {2 mM PMSF or 1 mM AEBSF [4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride], 2 μg/ml Aprotinin, 1 μg/ml Leupeptin, 1 μg/ml Pepstatin}. Cells were kept on ice for 10 min. Swollen cells were spun at 1000 *g* at 4 °C for 10 min. Swollen cells were resuspended in 2 PCV volume of buffer A and were lysed with 10 strokes of a type B pestle in a Dounce-homogenizer. Nuclear pellet was collected by centrifugation at 15,000 *g* at 4 °C, and nuclear pellet volume (NPV) was estimated. Nuclei pellet was resuspended in 2 NPV volume of buffer C [20 mM HEPES/K^+ pH 7.6, 1.5 mM $MgCl_2$, 650 mM KCl, 0.2 mM EDTA, 25% Glycerol (v/v), 1x PIC, 2 mM DTT and 0.2 mM

PMSF]. Nuclear pellet was lysed with 10 strokes of the type B pestle and kept at 4 °C for 60 min, shaking gently. Then nuclei were collected at 15,000 *g* at 4 °C for 30 min. The supernatant was dialyzed against 250 ml of buffer D [20 mM HEPES/K^+ pH 7.6, 1.5 mM $MgCl_2$, 100 mM KCl, 0.2 mM EDTA, 20% Glycerol (v/v), 1x PIC, 2 mM DTT and 0.2 mM PMSF] for 4 times, and each for 1h. The soluble supernatant was collected after centrifugation at 15,000 *g* at 4 °C for 15 min. BCA was performed to determine the protein concentration.

6. Antibodies

Anti-SIX1 antibody was prepared by immune-affinity purification. Anti-SIX4 antibody was prepared by other lab members in a manner analogous to the purification of anti-SIX1. Other antibodies include anti-myosin heavy chain (MHC, MF20clone, obtained from the Developmental Studies Hybridoma Bank, DSHB), anti-Myog (clone F5D, DSHB, contributed by W. Wright), anti-β-tubulin (clone E7, DSHB), and anti-FLAG (mouse monoclonal M2, Sigma). Secondary antibodies against mouse or rabbit IgG, coupled to horseradish peroxidase (Pierce Chemicals), were used in western blotting experiments following the manufacture's instruction.

Anti-MYOG antibody was purified through immune-affinity column. Biotin conjugated sheep anti-mouse IgG (Jackson Immuno Research) was coupled to immobilized Avdin resin (Pierce). The supernatant of F5D hybridoma was filtered in 0.22 μm filter, and was then applied to the column twice. The column was washed with PBS, and was then eluted with 0.1 M glycine (pH 2.8). The eluate was neutralized with 0.1 M Tris (pH 7.5) immediately and was dialyzed against 2L of PBS twice: for 4 h and for 16 h. Soluble supernatant was

concentrated with Millipore-Microcon (100kD) (Millipore) following the manufacture's protocol. BCA was performed to determine the protein concentration.

7. Chromatin preparation

Cross-linking buffer (1% Formaldehyde, 9.1 mM NaCl, 91 μM EDTA, 45 μM EGTA, and 4.55 mM Hepes pH 8.0) was added to the culture medium of the cell monolayer at RT, shaking for 30 min. A final concentration of 0.125 M Glycine was added to quench the reaction at RT, shaking for 5 min. Cells were rinsed with ice-cold PBS and were then resuspended in lysis buffer 1 [LB1, 50 mM Hepes-KOH pH7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol (v/v), 0.75% NONIDET P-40 (NP-40) (v/v), and 1xPIC], rocking at 4 °C for 10 min. The nuclei were collected by spinning at 1000 *g* at 4 °C for 5 min and were then resuspended in lysis buffer 2 (LB2, 200 mM NaCl, 1.0 mM EDTA, 0.5 mM EGTA, 10 mM Tris, and 1xPIC) at RT, rocking for 10 min. The nuclei were resuspended in lysis buffer 3 [LB3, 1.0 mM EDTA, 0.5 mM EGTA, 10 mM Tris, 0.5% sodium sarkosyl (w/v), and 1xPIC] and were then sonicated in a ice-cold water-bath at moderate intensity for 100 s (ChIP-on-Chip) or for 6 min (ChIP). Soluble supernatant was collected after spinning at 16,000 *g* at 4 °C for 10 min. To quantify the DNA in each sample, the chromatin was reverse-cross-linked with incubation at 65 °C for 16 h, and was then digested with 0.1 mg/ml Rnase A at 37 °C for 1 h and with 0.2 mg/ml Proteinase K at 65 °C for 30 min. The DNA fragment was then cleaned up with Qiagen PCR purification kit. The size of the DNA fragment was checked on 1% Agarose Gel.

8. Chromatin Immunoprecipitation (ChIP)

The chromatin sample containing 25 μg of DNA was pre-cleared with 20 μl of BSA blocked Sepharose Protein A beads (GE Healthcare, 50% slurry) in IP buffer [1% Triton (v/v), 0.1% Na Deoxycholate (Na-Doc) (w/v), and 1xPIC in LB3] for 2 h. Beads were removed from the chromatin solution by spinning at 500 *g* at 4 °C for 1 min. The chromatin was then incubated with 2 μg of antibody, rocking at 4 °C for 16 h. Normal rabbit IgG (Pierce Chemicals) was used as a negative control antibody. 20 μl of blocked beads were added to the mixture of chromatin and antibody, rocking at 4 °C for 2 h. The supernatant was then discarded after centrifugation and the beads were washed with RIPA buffer [50 mM Hepes-KOH pH7.5, 500 mM LiCl, 1 mM EDTA, 0.7% Na-Doc (w/v), 1% NP-40 (v/v), and 1xPIC] for 8 times rocking 4 °C and each time for 5 min. Beads were eluted with elution buffer (50 mM Tris pH8.0, 10 mM EDTA, 1% SDS) at 65 °C for 15 min. The eluate was cleaned up and was then purified by phenol: chloroform: isoamyl-alcohol extraction. The immunoprecipitated DNA was resuspended in ddH_2O (double distilled water) and was quantified by quantitative PCR (qPCR) employing the SYBR green chemistry on an MX3005p instrument (Stratagene). Each sample was used in duplicate PCR reactions with three distinct chromatin preparations (three biological replicates). For each time point and each replicate, an aliquot of the "input" starting chromatin material was used to create a serial dilution curve in order to calculate target DNA abundance in the IP samples as a percentage of that in the starting material (percent of input values). The *Hoxd10* promoter was used as a negative control locus, with no recruitment of the transcription factors being studied.

9. ChIP-on-chip

For ChIP-on-chip experiments, a pilot microarray was designed to optimize the protocol, which represents non-repetitive sequences of mouse chromosome 7 and includes some putative Six1 targets and negative loci as controls. According to pilot microarray analysis, a genome-wide microarray set (counting 3 arrays and counting approximately 2.9 million probes) was designed, covering approximately 17% of the mouse genome (manufactured by Agilent Technologies). These probes typically tile genomic regions with a density of 5 probes per kilo base. ChIP-on-Chip was performed at three time points during myoblast differentiation and each time point was performed in three independent biological replicates.

Chromatin was immunoprecipitated, amplified and labeled following the Agilent mammalian ChIP-on-chip protocol. Magnetic Protein A beads (Invitrogen, 50% slurry) was blocked by BSA and was then incubated with 10 μg of anti-SIX1 antibody at 4 °C for 4h. The chromatin sample containing 250 μg of DNA was pre-cleared with blocked magnetic beads in IP buffer [1% Triton (v/v), 0.1% Na-Doc (w/v) and 1xPIC in LB3] for 2 h. The chromatin was then incubated with antibody-coupled magnetic beads, rocking at 4 °C for 16 h. The beads were washed with RIPA buffer for 8 times rocking 4 °C and each time for 5 min. Beads were eluted with elution buffer at 65 °C for 15 min. The eluate was cleaned up and was then purified with phenol: chloroform: isoamyl-alcohol extraction. The immunoprecipitated DNA (IP) was resuspended in ddH_2O. A proportion of starting chromatin (Input) was cleaned up as mentioned in IP sample.

Input and IP samples were amplified by ligation-mediated polymerase chain reaction (LM-PCR) and were then labeled with different fluorescent molecules (Cyanine 5-dUTP for IP

and Cyanine 3-dUTP for Input). Labeled IP and Input samples were co-hybridized along with 50 μg of mouse Cot-1 DNA at 20 rpm (Agilent Microarray Hybridization Chamber) at 65 °C for 40 h. For the pilot microarray, 5 μg labeled DNA each for the IP and the control Input samples were used. Considering that genome-wide microarrays have denser probes than the pilot microarray, 12.5 μg labeled DNA each for the IP and the control Input samples were used. After washing, the slides were immediately scanned with an Agilent DNA microarray scanner at 2 μm resolution. The raw image data were extracted using Agilent Feature Extraction 10.1 (Agilent). Median fluorescence signals (raw data, Cy3 labeled Input and Cy5 labeled IP) for each replicate were regrouped and analyzed using the CisGenome program (Ji, Vokes, & Wong, 2006). First using a perl script, FE2Cisgenome.perl, median signal was extracted from the Feature Extraction output files. For the genome-wide microarrays, the median signals of three microarray slides consisting of the entire microarray set were combined together as one output file using another perl script: platform combiner.perl. At last, three independent biological data for each experiment were normalized by quantiles in CisGenome after log (2) transformation.

Typically, the workflow includes: normalization of raw data using the quantile method, calculation of log2 (IP/Input) ratio, detection of binding regions on log2 ratio using TileMap, annotation of nearby genes using a designed algorithm, and searching de novo motifs using Gibbs Motif Sampler.

To map binding regions, moving average (MA) was calculated for each probe. For the pilot microarray, IP enrichment values were not log2 transformed. The half Window size was set at 2, so IP enrichment values of 5 probes (the center probe, and 2 probes on each side) will be picked up together to compute MA statistic of the central probe. MAs passing a cut-off of

3 will be collected to form binding regions. The half window size was also limited to 300 base pairs [the entire window was 601 bp (2*300+1)], in order to exclude any probe which was too far from its neighbors. To be considered as a binding site, the region has to be more than 500 bp in case of dense probe cluster. If 3 consecutive MAs passed the cut-off, the false discovery rate (FDR) of this region will be calculated as a threshold. If the FDR was ≤10%, this peak will be regarded as a high confidence target. Contiguous peaks with a distance less than 1000 bp or less than 5 probes were considered to be generated by a single binding event and were merged as a single peak.

For the genome-wide microarrays, considering the 1M microarrays are different from the pilot microarray, the parameters of CisGenome peak finding program were further optimized. To calculate the moving average for each probe, half window size was set at 300bp, half window probes are 2 probes and the cut-off of MA was set at 2 [since the raw data are log (2) transformed, this cut-off was 4 for the raw data]. A peak has to contain more than 5 probes passing the cut-off of 2 and its length was longer than 100bp. Two peaks with the distance less than 1000 bp or less than 5 probes would be combined as one peak. This cut-off was stringent to generate fewer negative positives causing a negative E-O distance (E-O distance: between empirical threshold and the optimal threshold along the ROC-like curve, (Johnson et al., 2008). An arbitrary FDR threshold for each peak at 0.1 was set as good FDR (Ji et al., 2008). It is a considerably stringent verge which may compromise the false negative rate.

For combinatory location analysis of experiments at different time points (Mb, T24, Mt), a chromosome position sorting was applied to all the binding events. The chromosomes were

fragmented to 5kb bins. For different time points, any binding sites falling into the same bin were considered as one binding event.

In order to assign a binding event to a regulated gene, a heuristic approach was used which was similar to the one described by Sandman et al. for their Mef2 ChIP-on-chip experiments (Sandmann et al., 2006). Importantly, we do not assign binding events to target genes solely on the basis of their distance to neighboring genes. Instead, for each binding event, all genes, whose TSSs were located within 200 kb upstream/downstream the center of the bound region, were considered as candidates. For genes falling into this 400 kb window, genes whose expression were modulated by 2-fold or more after the knock-down of *Six1* were deemed to be "regulated" by Six1 and took precedence over genes with TSS closer to the peak center but not affected by Six1 knock-down (gene expression profiling data are unpublished and performed by A. Chu).

10. DNA sequence motif searches

For DNA sequence motif discovery on each time point of ChIP-on-chip data, the sequences of the 500 targets sites with highest binding confidence (lowest estimated FDR) were collected. A second group of sequences, with globally similar genomic distributions as the "bound" group, were selected randomly, to constitute a "background" sequence group. Background sequences were selected using CisGenome, according to a strategy which minimized bias in this process (Ji et al., 2006). The background sequences had the same length but were 20 times as numerous as the "bound" set of sequences. The Amadeus *de novo* motif discovery program (Linhart, Halperin, & Shamir, 2008) was then used to identify DNA sequence motifs more abundant among the "bound" sequences than among the

"background" set. Those motifs were given by Amadeus as position-weight matrices (PWMs). The PWMs occurrence rates in "bound" and "background" sequences were calculated using CisGenome. The cumulative hypergeometric probability was calculated for each motif: the population size was set to the DNA length surveyed (bound and background sets together, in base pairs); the sample size was set to the length of the bound sequences; the sample successes were the number of PWM "hits" in the bound sequences; and the population successes were the total number of PWM hits in bound and background sets.

For known motif exploration, PWMs of vertebrate transcription factor were obtained from TRANSFAC (Matys et al., 2003) and a bioinformatics study (Berger et al., 2008). Their abundance in Six1 bound and "background" regions were assessed as described above, using CisGenome.

11. Electrophoretic mobility shift assay (EMSA)

Double-stranded oligonucleotide probes corresponding to Myod-CER (core enhancer region) region were prepared by annealing complementary oligonucleotides that were synthesized by Operon Biofins. Each of the top strand harbors an extra "GGG" 5' overhang, which allows the annealed fragments to be filled in by Klenow fragment using biotin-labeled dCTP (Invitrogen) (Blais, Monte, Pouliot, & Labrie, 2002). Free dCTP was removed by Qiagen PCR purification kit (Qiagen). Binding reactions [34 mM KCl, 25 mM Hepes pH 7.6, 5 mM $MgCl_2$, 10% Glycerol (v/v), 1 mM DTT, 1 μg of dI: dC, 20 μg of IgG-free BSA, and 20 fmol labeled probe] were incubate with either 50 ng of purified recombinant SIX1 (a kind gift from Dr. J. Dilworth, Ottawa Hospital Research Institute), or 5 μg of nuclear extract at RT for 15 min (Grifone et al., 2007). For competition experiment, 1 pmol mutant probe

competitor (50 molar fold excess) was added to each reaction before the addition of the labeled probe. 0.25*TBE (22.25 mM Tris base, 22.25 mM boric acid, 0.5 mM EDTA, pH 8.0) was used to prepare and run the gel (5% native poly-acrylamide gel). Poly-acrylamide gel was pre-run at 100V at 4 ℃ for 1 h (8*10cm, Biorad). After loading the samples, the gel was run until the bromophenol blue dye front reached the bottom of the gel (~90 min). The biotin-labeled DNA oligonucleotides were transferred to Biodyne® Nylon membrane at 100V at 4 ℃ in 0.5* TBE (44.5 mM Tris base, 44.5 mM boric acid, 1 mM EDTA, pH 8.0) for 30 min. DNA was cross-linked to the membrane using UV light (120mJ/cm^2) for 45 s and was detected with detection kit (Pierce) according to the recommended protocol. The signals on the membrane were exposed to X-ray films (Thermo). Probe and competitor sequences were given in Table 1.

Table 1 Oligonucleotide sequences

Gene Name	Strand	Sequence	Usage
Chrng	Forward	AGAGGTGTAGGACCGAAGCA	ChIP
	Reverse	GCTGGGCTCAGATTACATGG	ChIP
Myog	Forward	AGAGGGAAGGGGAATCACAT	ChIP
	Reverse	TGGCATGAACCCAGAGATAA	ChIP
Myod1-CER	Forward	TGCTTCTTTCGGCCAAGTAT	ChIP
	Reverse	CCAACTGGCTGTGTTGTGAG	ChIP
Tnnc1	Forward	GCATACCAAAGTCCCCAGAA	ChIP
	Reverse	AGCTGTCCCTCAGGTTGAGA	ChIP
Hoxd10	Forward	GAGAAATCGGACTCACCTTCC	ChIP
	Reverse	CACATACCCAGGCAGAACG	ChIP
Six4	Forward	ATCTGGCCGATCAGGTTTC	ChIP
	Reverse	AAGAATAGCTGGGAGTGAATGC	ChIP
Ldb3	Forward	CCTCCCCAGGGACTATATTAGC	ChIP
	Reverse	ACCGGAGGAGTCACGTTG	ChIP
Myod1-DRR	Forward	ACAGGTCCAGACTGGGTAGG	ChIP
	Reverse	GGAATGTTGGTCTGGCTAGG	ChIP
Chrna1	Forward	GACAAGCCTCTGACTCATGATCTATGT	ChIP
	Reverse	GCTGCCGGTCCTACTCCACCCTGGCT	ChIP
Pax3	Forward	GATGCAGGAAGATGACATGC	ChIP
	Reverse	TCGTAAGGGGCTCTGGTTAC	ChIP
Myf5	Forward	TGAACCATTTCAAGGCAACC	ChIP
	Reverse	GAGTTTGAGGTTTTTCGTCATTC	ChIP
BDNF	Forward	AGTGTCTATTTCGAGGCAGAGG	ChIP
	Reverse	TCCTCGGTGAATGGGAAAG	ChIP
Pbx1	Forward	GAATGGCTGTGGAACAAACC	ChIP
	Reverse	GCTGGCAGGTTTTCAGTCC	ChIP
Cacng1	Forward	CCTTGGGGCTGAGAGACAC	ChIP
	Reverse	AAACCTGAGCCGAAGGAATC	ChIP
Six5	Forward	CTGTCGCTTGTCTTTGTTGTG	ChIP
	Reverse	GGCGTCAGAGACCAGAGC	ChIP
Six1	Forward	GGCTCCTTTGAGCTTTTGTG	ChIP
	Reverse	GGGGATAAAAACCCTGGATG	ChIP
Ankrd1	Forward	GCTTGTCATCTCCCTCTTGG	ChIP
	Reverse	GTGGATGAACCCCTGGAAC	ChIP
Chrna1_exon8	Forward	CACAATGAAAAGACCATCCAGA	ChIP
	Reverse	ATGTACTTCACGCCCTCGAT	ChIP
Myod_CER_Mef 3_WT	Forward	GGGTAAACTTCTGAGACAGTAATTTTATCCTGCTTCTTTCGGCC	EMSA
	Reverse	GGGGGCCGAAAGAAGCAGGATAAAATTACTGTCTCAGAAGTTTA	EMSA
Myod_CER_Mef 3_MT	Forward	GGGTAAACTTCTGAGACAGTGGTTGGGTCCTGCTTCTTTCGGCC	EMSA
	Reverse	GGGGGCCGAAAGAAGCAGGACCCAACCACTGTCTCAGAAGTTTA	EMSA
Myod_CER_Ap 1_WT	Forward	GGGGCTTAAAACCCGTGACTCACAACACAGCCAG	EMSA
	Reverse	CCCCTGGCTGTGTTGTGAGTCACGGGTTTTAAGC	EMSA
Myod_CER_Ap 1_MT	Forward	gggGCTTAAAACCCGT**atata**ACAACACAGCCAG	EMSA
	Reverse	cccCTGGCTGTGTTGT**tatat**ACGGGTTTTAAGC	EMSA

Table 1| Oligo sequences designed based on mm9 on the UCSC Genome Browser in this study.

Results

1. Purified rabbit anti mouse SIX1 antibody is specific.

Chromatin Immunoprecipitation requires the use of an antibody recognizing a certain protein of interest, which is usually coupled to immobilized matrix, to pull down this protein and its binding DNA sequences. Therefore, a pure and specific antibody is a critical factor for ChIP success. Considering a large amount of antibody to be used, customized antibody was produced. An anti-SIX1 antibody was raised in rabbits immunized with recombinant 6*HIS tagged full length murine SIX1 protein that was purified from *E. coli*. The anti-SIX1 antibody was purified with GST-SIX1 coupled immobilized glutathione agarose beads. Before and after purification, the anti-SIX1 antibody was tested for specificity by western blot on C2C12 cell lysates.

Figure 7 Specific anti-SIX1 antibody purified from rabbit serum

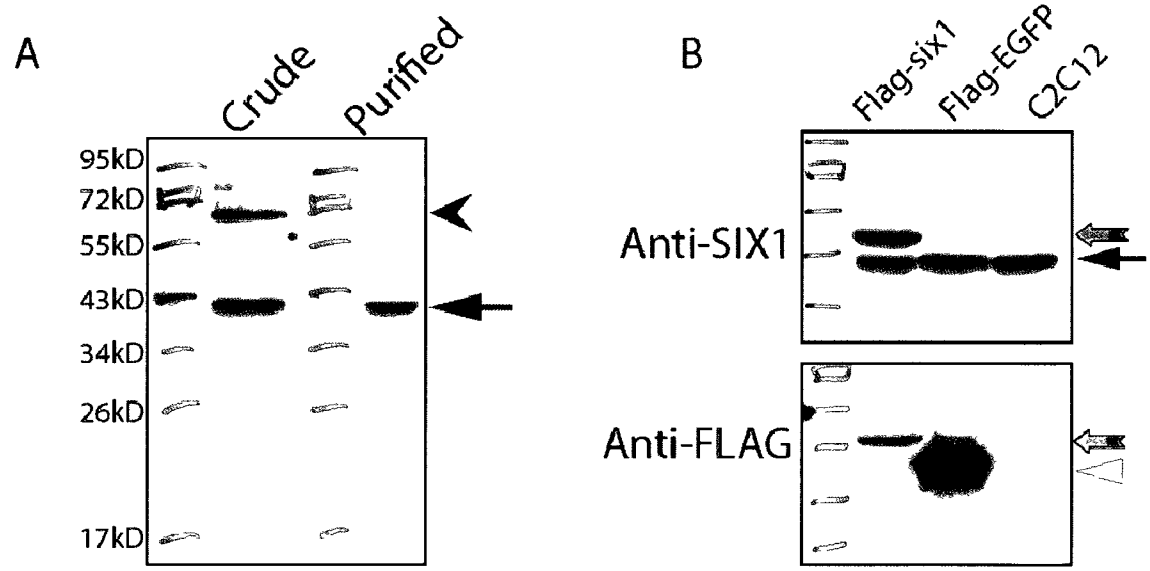


Fig. 7| The specificity and efficiency of anti-SIX1 antibody were tested by western blot. A, total cell lysate of C2C12 at T24 (myoblasts induced to differentiate for 24 hours) was used.

Membranes were probed for SIX1with crude rabbit serum or purified antibody. The blue arrow indicates SIX1 protein. The black arrow-head points the major cross-reacting band. B, total cell lysate of C2C12 at T24 was used. Membranes were probed for SIX1 with purified antibody and for FLAG tag with commercial antibody. C2C12 were stably transfected with expression plasmids of FLAG-Six1, or FLAG-EGFP. The purple arrow shows FLAG-SIX1 protein. The red triangle shows FLAG-EGFP protein.

A dominant band was shown around 34kD on western blot after purification (Fig. 7A, arrow), while crude serum had high background (Fig. 7A, arrow head). The efficiency of the anti-SIX1 antibody was also tested on C2C12 myoblasts stably over-expressing FLAG-SIX1. Endogenous SIX1 (Fig. 7B, arrow) was comparable in abundance to exogenous FLAG-SIX1 (Fig. 7B, arrow head). These results show that the anti-SIX1 antibody is specific.

2. Six1 is expressed and functional in C2C12

In previous studies, Six1 has been shown to be a key regulator of mammalian embryonic myogenesis (Grifone et al., 2005; Laclef, Souil et al., 2003; Niro et al., 2010). In murine embryos, Six1 is expressed in many tissues, including epaxial and hypaxial muscles (Laclef et al., 2003; Xu et al., 2003). Six1 is still expressed and functional in adult skeletal muscles (Grifone et al., 2004).

The C2C12 mouse myoblast cell line is a well-established muscle regeneration model. These cells can proliferate as myoblasts and can be induced to differentiate into myotubes by serum withdrawal in vitro (Fig. 8A). Here the expression of Six1, Six4, Myog and Mhc was examined by western blot (Fig. 8B). Five time points were chosen, including 1) myoblast

(Mb), 2) T0: confluent C2C12 cells before differentiation, 3) T24: 24 hours after differentiation by serum withdrawal, 4) T48: 48 hours after differentiation, and 5) myotubes (Mt): differentially trypsinized cells after 96 hours of differentiation (Fig. 8A). As expected, the expression of Myog started at T0 and became highly induced after T24. Six1 and Six4 were expressed consistently from myoblasts to myotubes. MHC was detected only in myotubes.

Figure 8 Six1 is expressed and functional in C2C12 myoblasts

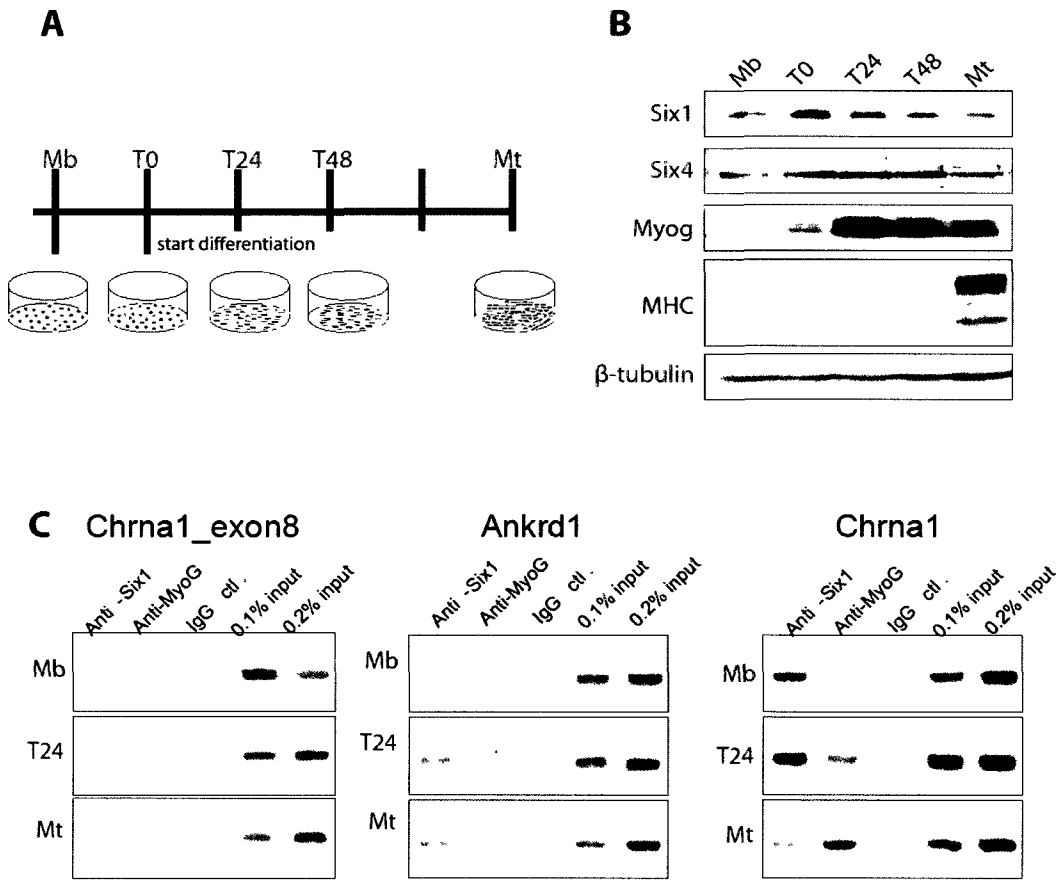


Fig. 8| A, the C2C12 differentiation model. Myoblast, Mb. Confluent cells are termed T0. Cells induced to differentiate for 24 hours are called T24, for 48 hours named T48, and for 96 hours termed Myotubes (Mt). B, Western blot detection of Six1 expression during a

C2C12 differentiation time course. Total cell lysates from Mb, T0, T24, T48 and Mt were used and membranes were probed for SIX1, SIX4, MYOG, MHC and β-tubulin. β-tubulin was used as the loading control. C, ChIP assays showed co-binding of SIX1 and MYOG on two DNA loci. Chromatin extracted from Mb, T24 and Mt was immunoprecipitated with anti-SIX1, anti-MYOG or control IgG. Primer pairs amplifying the promoter regions of *Ankrd1* and *Chrna1* were used. A primer pair amplifying the exon 8 of *Chrna1*, which is 15 kb away from its transcription start site (TSS), was used as the negative control. 0.1% and 0.2% of Input DNA were used as the indicators to measure the relative enrichment of IPs.

Myog plays an important role in terminal differentiation, which agrees with its induction at an early stage of differentiation. As a component of myosin, MHC is important for muscle contraction and is detected only in mature myotubes. The consistent expression of Six1 and Six4 reflects their involvement in muscle development and suggests that they are functional during muscle differentiation. Despite lower expression in myotubes, it was shown that the protein of Six1 and Six4 was preferentially accumulated in myotube nuclei (Spitz et al., 1998).

A study, using a small scale promoter study on the mouse genome to map MRF binding sites, showed that Six binding DNA sequence motif (MEF3 motif) was enriched among those MRF targets that were most induced during the differentiation of C2C12 myoblasts (Blais et al., 2005). Six1 and Myog may have a cooperative function by binding similar regulatory regions. ChIP was performed in C2C12 cells with anti-SIX1 antibody, anti-MYOG antibody, and control rabbit IgG at different time points: Mb, T24 and MT. The regulatory regions of *Ankrd1* [ankyrin repeat domain 1 (cardiac muscle)] and *Chrna1* were

co-occupied by SIX1 and MYOG (Fig. 8C). A DNA region located 15kb away from the *Chrna1* TSS was not bound by either SIX1 or MYOG (Fig. 8C). Ankrd1 may be involved in the switch of muscle fibers to fast-type (Laure et al., 2009). Chrna1 is an essential component of the nicotinic cholinergic receptor, which is required for muscle contraction upon nerve stimulation (Michalk et al., 2008). Six1 can directly bind the regulatory regions of these genes, indicating that it is functional in C2C12. Of note, the binding of SIX1 and MYOG on *Chrna1* promoter region was modulated during differentiation. SIX1 binding is consistent while MYOG binding increases during C2C12 differentiation.

Based on these results, three representative time points, Mb, T24 and Mt, were chosen to map global Six1 binding. Mb cells are actively proliferating cells, T24 cells start to differentiate with dramatic re-programming, and Mt cells are fully differentiated cells with irreversible cell cycle exit and massive cell fusion. To better address issues studied here, a microarray platform covering the whole mouse genome was aimed at. Before designing the genome-wide microarray, a pilot study is necessary to optimize the protocol and is helpful to design the global microarray platform.

3. Pilot DNA microarray provides the basis to design genome-wide microarrays

As highlighted in the introduction, Six1 plays an important role for the formation of many tissues during embryogenesis, and has also been shown to be involved in certain aspects of cancer biology. This suggests that Six1 has a broad function at both embryonic and adult stages.

Most of the knowledge of Six1 function was from the mouse embryo studies. To address the

role of Six1 in the adult muscle, a functional genomics tool, ChIP-on-Chip (DNA microarray), was chosen as the principal technique (Fig. 6). A pilot microarray platform was designed to tile the non-repetitive regions of the whole chromosome 7 and several control genes. Two biological replicates of Mb were examined from independent experiments. Six1 binding regions were calculated by CisGenome using median signal normalized by quantile method, termed as "peak detection". The parameters of the peak detection algorithm were optimized to obtain reasonable peak numbers and FDR according to the microarray design that we were working with.

Figure 9 the distribution of Six1 binding sites from the closest TSS in pilot microarrays: on chromosome 7 only

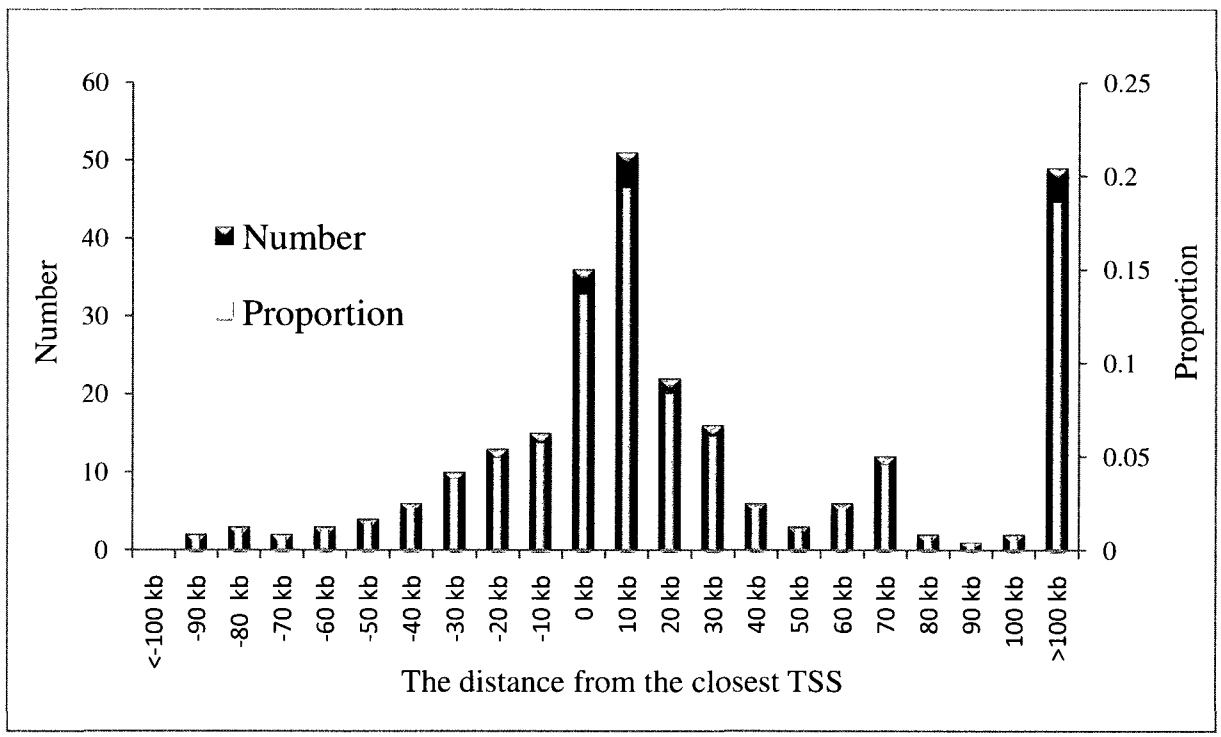


Fig. 9| The distance from the center of the binding site to the closest TSS was calculated. The number and percentage of peaks within each 10kb bin were accounted. Negative numbers

represent binding sites upstream of TSS and positive numbers represent downstream of TSS.

FDR of peaks was calculated by CisGenome peak detection program to indicate how good the peaks were. The peaks with smaller FDR have higher confidence to be real binding sites. Peaks with FDR less than 0.1 were regarded as "good peaks" (Ji et al., 2008). 363 good peaks (out of 376 peaks) were found in Mb using the pilot microarray. 264 peaks out of 376 were in Chromosome 7. The median peak length was 686bp and on average each peak had 6.13 probes. Most of Six1 binding sites were close to Transcription Start Sites (TSS) of genes but the sites were not restricted to the promoter regions (Fig. 9). The gene which was closest to the center of the peak was assigned to each peak. Enriched gene categories were computed on David Bioinformatics website using Gene Ontology (GO) analysis. System development and muscle development were significantly enriched ($p < 10^{-6}$ and 10^{-5}, hypergeometric distribution), which supported the hypothesis that Six1 was important for muscle development but also had a global function (Table 2).

Table 2 A few examples of enriched gene functional categories in pilot microarrays

Term	Gene Category	Count	%	PValue	List Total	Pop Hits	Pop Total
GO:0048731	system development	57	20.00	5.5E-05	224	2027	13451
GO:0010468	regulation of gene expression	66	23.16	6.1E-05	210	2429	12272
GO:0048568	embryonic organ development	14	4.91	1.9E-04	224	241	13451
GO:0007399	nervous system development	30	10.53	2.0E-04	220	854	13138
GO:0007517	muscle organ development	11	3.86	7.4E-04	220	176	13138
GO:0060538	skeletal muscle organ development	7	2.46	1.6E-03	210	74	12272
GO:0007423	sensory organ development	12	4.21	4.6E-03	210	257	12272
GO:0042127	regulation of cell proliferation	18	6.32	8.3E-03	224	538	13451
GO:0000087	M phase of mitotic cell cycle	9	3.16	1.6E-02	220	194	13138

Table 2| several gene categories were significantly enriched among Six1 target genes in the pilot microarray study. The program was the Gene Ontology analysis of functional

annotation at David Bioinformatics Resource website (Huang da, Sherman, & Lempicki, 2009) (http://david.abcc.ncifcrf.gov/home.jsp). Six1 binding in myoblasts was mapped in duplicates, using pilot microarray (covering chromosome 7 and some control DNA loci). The binding data were analyzed using CisGenome program and were annotated with the closest gene for each binding event. List total is the number of Six1 target genes. Count is the number of Six1 target genes in the corresponding gene category. Pop total is the number of genes in the whole genome. Pop hit is the number of genes in the corresponding gene category in the whole genome. % is the number of genes / the total number of genes in one category. The *p* value is calculated with a modified Fisher Exact test.

Next, several new DNA loci bound by Six1 were validated by ChIP, followed by PCR (Fig. 10), such as the regulatory regions of *Hist2h4* (histone cluster 2, H4), *Angptl* (angiopoietin 1), *Six1*, *Gas2* (growth arrest specific 2), and *Mos* (moloney sarcoma oncogene). Some of the genes were also targeted by Myog. Hist2h4 encodes one of the histone proteins composing the nucleosome. Angpt1 plays an important role in vascular development and angiogenesis (Gale & Yancopoulos, 1999). Gas2 is a caspase substrate and is involved in the cell shape changes during apoptosis (Porter, Ng, & Janicke, 1997). Mos is implicated in the mitogen-activated protein kinase (MAPK) cascade and plays a critical role in vertebrate oocyte meiosis (Gebauer & Richter, 1997). Of note, Six1 can bind its own promoter region, suggesting an auto-regulatory loop. Taken together, these findings suggest that Six1 has spreading DNA binding loci and a broad function, including system development and muscle development.

Figure 10 Validation of a few Six1 targets found in pilot microarrays

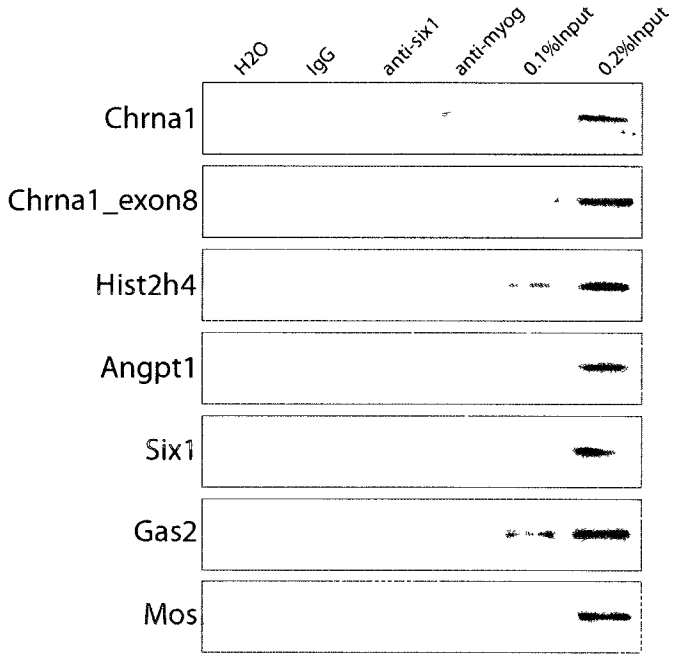


Fig. 10| Several Six1 binding loci were validated with ChIP assays. Chromatin extracted from T24 was immunoprecipitated with anti-SIX1, anti-MYOG or control IgG. Primer pairs amplifying Six1 binding sites, which are in the regulatory regions of *Hist2h4*, *Angpt1*, *Six1*, *Gas2*, and *Mos*, were used. Primer pairs amplifying the promoter region of *Chrna1* and the exon 8 region of *Chrna1* were used as positive and negative controls, respectively. 0.1% and 0.2% of Input DNA were used as the indicators to measure the relative enrichment of IPs.

4. Mapping of Six1 binding sites on genome-wide DNA microarray

Based on the distribution of Six1 binding sites, the pilot ChIP-on-Chip study indicates that, by extrapolating to the entire genome, approximately 79% of Six1 target genes in C2C12 cells will be identified if a microarrays covers: A) transcription regulatory regions: from 7.5 kb upstream to 7.5 kb downstream of each TSS, according to UCSC genes in UCSC genome

browser, B) phylogenetically conserved regions of extended regions on each side of TSS (from -50 kb to -7.5 kb and from +7.5 kb to +50 kb), and C) mammalian highly conserved regions outside the TSS-centered 100 kb window (Visel et al., 2009). Conserved regions selected with UCSC genome browser (regions at least 11bp long with the conservation score above 11) are tiled and extended 1.5 kb on each side (Bejerano, Siepel, Kent, & Haussler, 2005). On this basis, we designed a microarray set (counting 3 arrays and numerating a total of 2.9 million probes), covering 17% of the mouse genome (Agilent Technologies). These probes typically tile genomic regions with a density of 5 probes per kilo base. Of note, the genome-wide microarrays do not cover all the non-repetitive regions as the pilot microarray does, which may produce some gaps in some regions and may consequently cause missing of some binding events

Figure 11 Demonstration of SIX1 binding sites in genome-wide microarrays

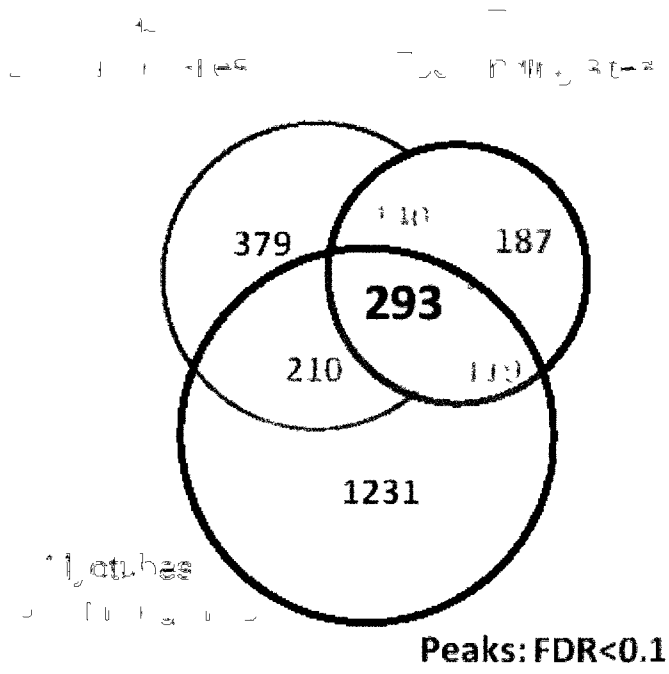


Fig. 11| Venn diagram of SIX1 binding sites in myoblasts, T24 and myotubes. 1022 binding sites were found in myoblasts. 739 binding sites were found at T24. 1853 binding sites were found in myotubes. Myoblasts shared 433 binding sites with T24. Myoblasts shared 503 binding sites with myotubes. T24 shared 412 binding sites with myotubes. There were 293 binding sites shared by all three time points.

Biological triplicates were performed for each of three time points. Based on optimized parameters (explained in details in the materials and methods section), 1022 binding sites with FDR < 0.1 were identified in Mb, 739 in T24 and 1853 in Mt. In total, 2559 individual SIX1 binding sites were discovered (Fig. 11).

Similar binding patterns were found among different time points. Around a certain promoter region, overlapping or very close binding events were often found in Mb, T24 and Mt (Fig. 12). To better annotate related genes, chromosome position sorting was applied to the genome-wide data in order to comprehensively understand SIX1 binding during biological processes. The chromosomes were fragmented to 5kb bins. No matter how many binding events out of three fall into a certain chromosome bin, it would be regarded as one binding site to avoid repetitive annotation for one regulatory region during the analysis of the three datasets.

For gene annotation, genes within a range of +/- 200 kb from the center of the peak were searched and were then sorted according to their importance. Two criteria were evaluated in order of priority, 1) their expression change after the knock-down of Six1, 2) their absolute distance from TSS. Taking this algorithm, 754 genes in Mb, 408 genes at T24, and 1134 genes in Mt were annotated among SIX1 binding sites, for a total of 1640 genes.

Figure 12 A few examples of Six1 bound DNA loci in the genome-wide microarray study

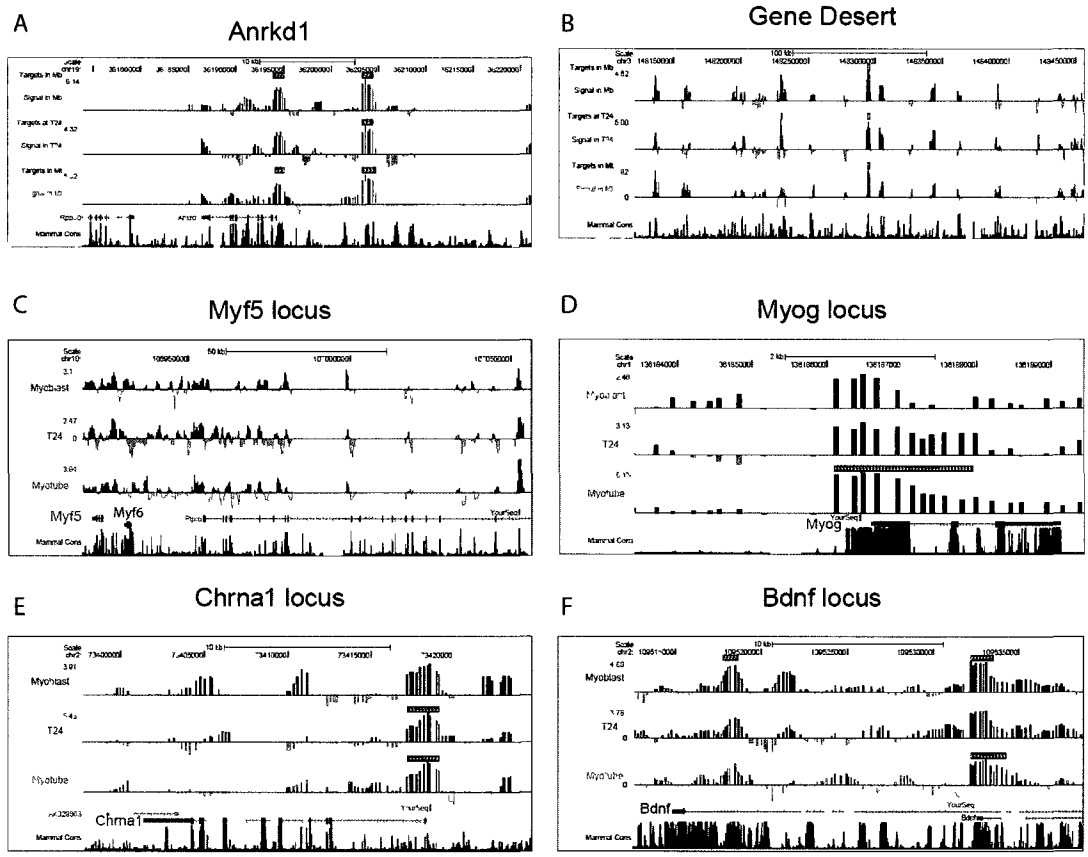


Fig. 12| Binding data were loaded into the UCSC Genome Browser (mm9) (http://genome.ucsc.edu/index.html?org=Mouse&db=mm9&hgsid=159045136). There were binding data in three time points, myoblast (on the top), T24 (in the middle) and Myotubes (at the bottom). At each time point, there were two types of information: the peak calling on the top (black rectangles) and the signal intensity of each probe at the bottom (colored histograms). A, two binding sites are upstream of *Ankrd1*. B, a binding site has no upstream or downstream gene within 200 kb. C, a binding site in myotubes is at 130 kb upstream of *Myf5* and at 121 kb upstream of *Myf6*. D, a binding site in myotubes is in the promoter region of *Myog*. E, a binding site in T24 and myotubes is in the promoter region of *Chrna1*.

F, a binding site is in the promoter region of the short isoform of *Bdnf*. Another binding site is in the intron of the long isoform of *Bdnf*.

The decreased number of annotated genes compared to that of binding sites could be explained, at least partially, by two reason, 1) for a certain gene, there are more than one SIX1 binding sites nearby which are annotated to the same gene (Fig. 12A), and 2) for a certain binding event, there is not any known transcribed region within the searching range (Fig. 12B).

Gene Ontology study found that some gene categories were significantly enriched among those targets, such as system development, muscle development, muscle contraction, neurogenesis, ureteric budding, angiogenesis and regulation of cell cycle (Table 3). Using a myoblast cell line as the model, it was not surprising to find genes involved in muscle development (MyoD1, Myog, Myf6, Myf5 and Pax3), which was consistent with a number of embryonic development studies (Giordani et al., 2007; Grifone et al., 2005; Laclef et al., 2003). In addition, muscle contraction genes were also significantly enriched, such as Titin, troponin, myosin isoforms, and ATP-dependent sodium-potassium channels, which implies SIX1 is important not only for early myogenesis but also for muscle function.

Kidney development and ureteric bud development were also significantly enriched among Six1 targets. Six1 was shown to be able to regulate embryonic nephrogenesis. Kidneys were not formed correctly in *Six1*$^{-/-}$ mice (Laclef, Souil et al., 2003; Xu et al., 2003). However, a significant number of genes involved in nephrogenesis were still bound by SIX1 in C2C12 cells.

Table 3 Gene Ontology study in genome-wide microarrays

Term	Gene Category	Count	%	PValue	List Total	Pop Hits	Pop Total
GO:0048731	system development	308	18.16	1.31E-22	1221	2027	13451
GO:0001944	vasculature development	65	3.83	6.48E-15	1186	250	13138
GO:0048514	blood vessel morphogenesis	54	3.18	2.07E-13	1186	198	13138
GO:0042692	muscle cell differentiation	39	2.30	8.49E-13	1186	117	13138
GO:0007399	nervous system development	135	7.96	5.94E-11	1186	854	13138
GO:0001657	ureteric bud development	17	1.00	3.26E-07	1221	42	13451
GO:0042127	regulation of cell proliferation	85	5.01	3.51E-07	1186	538	13138
GO:0001525	angiogenesis	31	1.83	1.93E-06	1186	133	13138
GO:0001822	kidney development	27	1.59	2.12E-06	1186	107	13138
GO:0035108	limb morphogenesis	27	1.59	8.78E-06	1186	115	13138
GO:0048870	cell motility	43	2.54	9.38E-04	1186	284	13138
GO:0051726	regulation of cell cycle	34	2.00	1.58E-03	1186	214	13138

Table 3| several gene categories were significantly enriched among Six1 target genes in the genome-wide microarray study. The program is the Gene Ontology analysis of functional annotation at David Bioinformatics Resource website (Huang da et al., 2009) (http://david.abcc.ncifcrf.gov/home.jsp). Six1 binding were mapped at three time points (myoblast, T24 and myotubes) and were mapped in triplicates, using genome-wide microarrays. The binding data were analyzed by CisGenome program and were annotated with the customized algorithm. List total is the number of Six1 target genes. Count is the number of Six1 target genes in the corresponding gene category. Pop total is the number of genes in the whole genome. Pop hit is the number of genes in the corresponding gene category in the whole genome. % is the number of genes / the total number of genes in one category. The *p* value was calculated with a modified Fisher Exact test.

Figure 13 A variant MEF3 motif discovered in genome-wide binding data is specific and sensitive

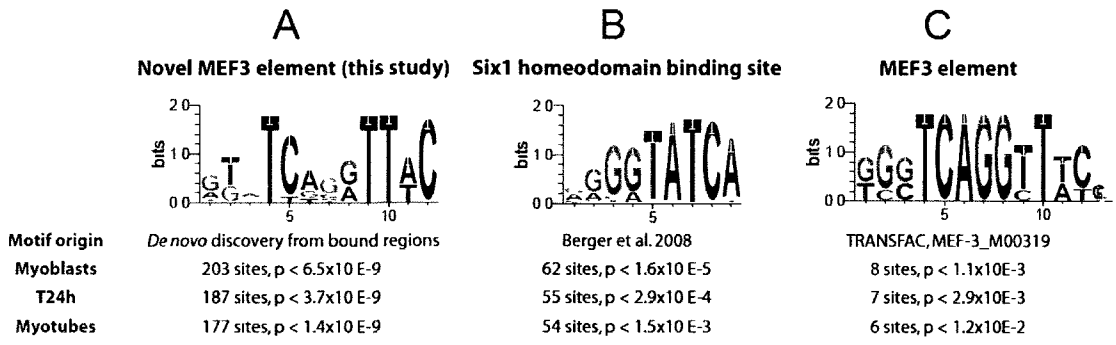


Fig. 13| A variant MEF3 motif was discovered by using *de novo* motif finding among SIX1-bound sequences using the Amadeus motif discovery platform. Very similar motifs were detected in Mb, T24 and Mt. Here, it shows the MEF3 motif found in myotubes. The numbers of good matches to the position weight matrix (PWM) were calculated in the regions bound by Six1 in Mb, T24 and Mt. The *p* values were computed for the cumulative hypergeometric probability, comparing the numbers of good matches in total targets at a certain time point, to those in a random set of sequences along the genome. The canonical MEF3 motif in the TRANSFAC database (C) and the MEF3 motif reported by Berger et al. (B) were shown as comparison.

The MEF3 motif, the DNA binding element for Six1, Six2, Six4, and Six5, has been characterized (Berger et al., 2008; Kawakami et al., 2000; Kawakami et al., 2000). However, the MEF3 motif was derived either based on analysis of a few of skeletal specific enhancer sequences (Hidaka, Yamamoto, Arai, & Mukai, 1993; Parmacek et al., 1994; Salminen, Lopez, Maire, Kahn, & Daegelen, 1996), or according to a protein binding microarray (PBM) study with synthetic DNA and recombinant proteins (Berger et al., 2008; Noyes et al., 2008). In this study, the Six1 binding motif was searched in a representative and unbiased

way and in a native chromatin context. The amadeus *de novo* motif discovery program was performed on SIX1 binding sequences from Mb, T24 and Mt microarray data respectively. Similar or identical MEF3 motifs were found to be significantly enriched in the three datasets. A representative motif from the myotube experiment was shown (Fig. 13 A). This MEF3 motif has a flexible core and two extended tails, which provides better specificity and better sensitivity than the known MEF3 motifs (Fig. 13 B, C). To compare this MEF3 motif to the reported ones, the enrichment of each MEF3 motif (two reported ones and the one in this study) in the binding sequences was searched for each dataset using CisGenome known motif mapping program. Using the MEF3 motif in this study, more sites were discovered and a better *p* value was calculated. To test if this variant MEF3 motif can reflect the SIX1 binding flexibility, a divergent MEF3 motif in CER of MyoD was tested with EMSA and it can be bound by recombinant SIX1 protein (Fig. 14).

Figure 14 Recombinant SIX1 can bind a divergent MEF3 motif

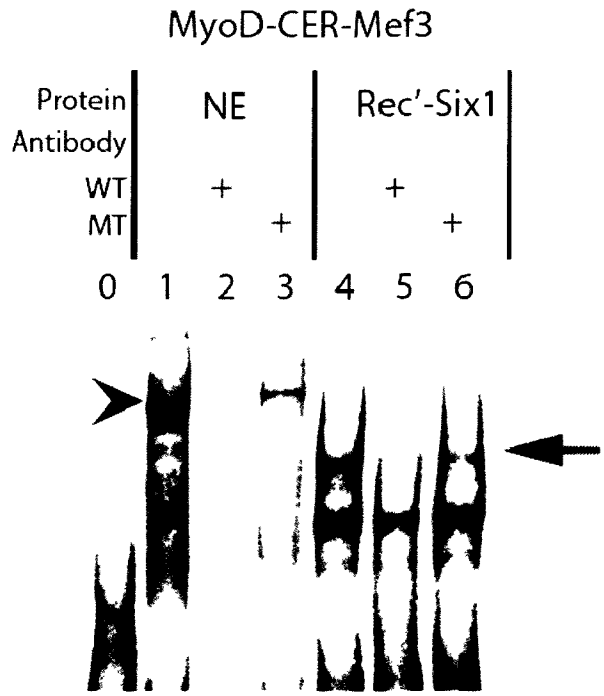


Fig. 14| EMSA showed that SIX1 can bind a divergent version of MEF3 motif. A DNA probe derived from CER of the *Myod* gene, with the sequence 5'-ggatAaaattac-3' (the underline position is a C in the canonical MEF3 motif), was used to bind either C2C12 nuclear extract at T24 or recombinant SIX1 (a kind gift from Dr. Dilworth). The arrow shows the Six1 binding. The arrow head shows the binding in nuclear extract. NE: nuclear extract. WT: wild type competitor. MT: mutant competitor. Rec'-Six1: recombinant SIX1.

Figure 15 E-box and bZIP/AP1 DNA sequence motifs were significantly enriched among SIX1-bound sequences

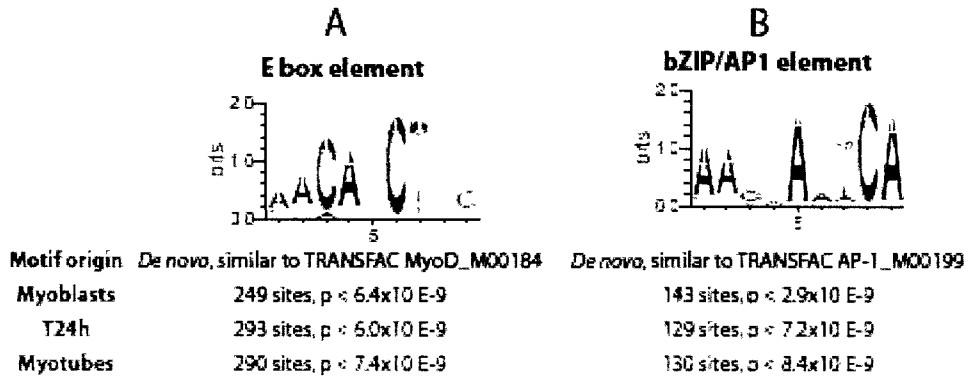


Fig. 15| Using Amadeus *de novo* motif discovery platform, E-box and bZIP/AP1 DNA sequence motifs were discovered to be significantly enriched among SIX1 binding sequences. A, E-box element, bound by MRFs. B, bZIP/AP1 element, bound by AP1 and other transcription factors containing basic leucine zipper domain. Very similar motifs were detected in Mb, T24 and Mt. The numbers of good matches to the position weight matrix (PWM) were calculated in the regions bound by Six1 in Mb, T24 and Mt. The *p* values were computed for the cumulative hypergeometric probability, comparing the numbers of good matches in total targets at a certain time point, to those in a random set of sequences along the genome.

Interestingly, E-box, the binding element of the MRF transcription factors, was also significantly enriched among Six1 targets (Fig. 15A). This finding is consistent with the previous study and strongly supports the concurrent binding of SIX1 and MRFs. bZIP/AP1 element, which is the DNA sequence motif bound by AP1 (activator protein 1) and other transcription factors containing basic leucine zipper domain, was also enriched among Six1 binding sites (Fig. 15B). Since the enrichment is significant, the presence of bZIP/AP1 element is not a random event. There are two possibilities to explain its presence: 1) SIX1 may directly bind this element, which may be a "variant" unknown Six1 binding sequence. This was ruled out according to the EMSA (Fig. 16); 2) SIX1 may bind bZIP/AP1 element through a protein complex, while SIX1 may or may not physically interact with bZIP/AP1-binding proteins. With DNA probes containing bZIP/AP1 element, an EMSA study, using anti-SIX1 to supershift the protein complex in nuclear extracts, will help to test this probability.

Figure 16 Recombinant SIX1 cannot bind bZIP/AP1element

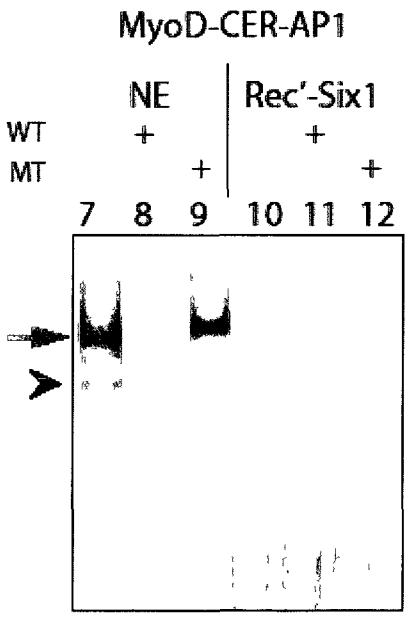


Fig. 16| Electrophoretic mobility shift assay (EMSA) shows that recombinant SIX1 cannot bind bZIP/AP1element. A DNA probe derived from CER of the *Myod* gene, with the sequence 5'-TGAC/GTCA-3' (AP1 motif), was used to bind either C2C12 nuclear extract at T24 or recombinant SIX1 (a kind gift from Dr. Dilworth). The arrow shows the shift by a protein or a protein complex. The arrow head shows non-specific binding.

In summary, besides its role in muscle development, Six1 is also globally involved in system development, targeting a list of genes involved in the genesis of multiple organs. These genome-wide datasets also allowed identifying a MEF3 motif recognized by the SIX1 protein in the context of native chromatin binding in vivo.

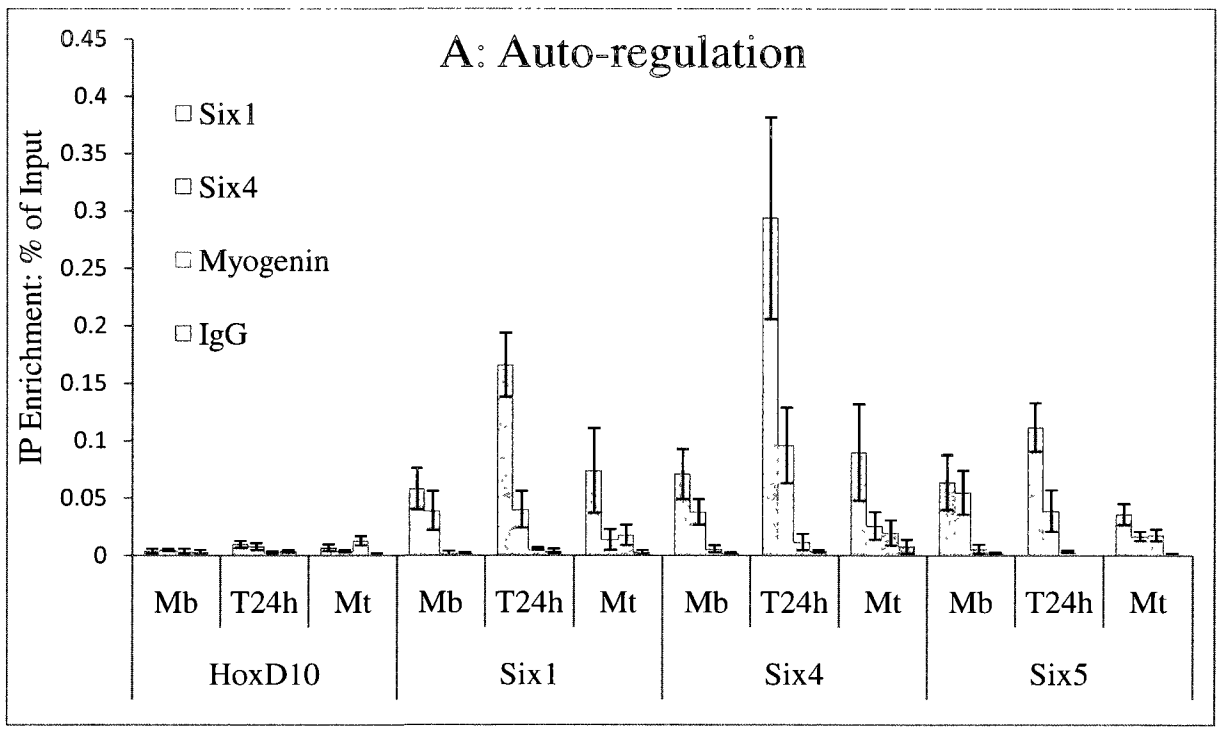
5. Validation of novel Six1 targets

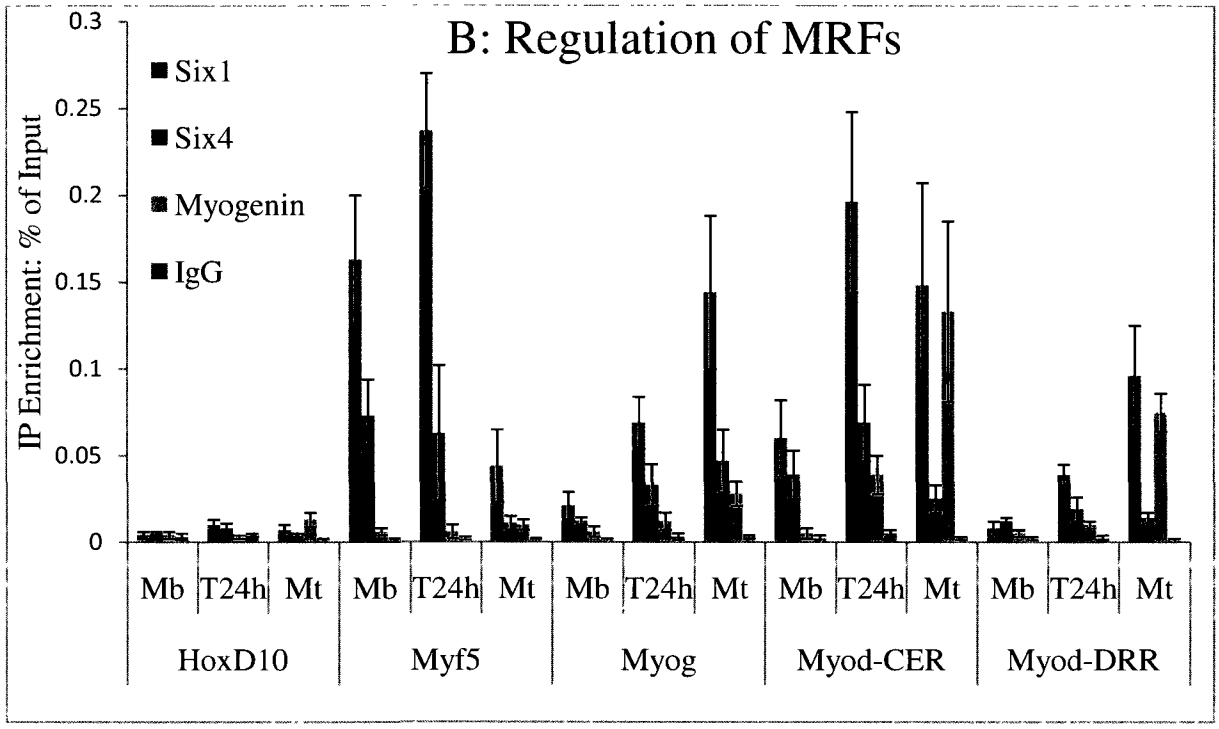
ChIP-on-Chip is a high throughput technology, generating a large amount of data. However, further validation is necessary and important. Here, a few interesting targets of SIX1 were selected for further validation. ChIP assay followed by quantitative real time PCR (q-ChIP) was performed to test several putative SIX1 binding DNA loci. These Six1 binding regions were annotated with interesting genes falling into several functional categories (Fig. 17, Table 3). First of all, the promoter regions of *Six4*, *Six5* and *Six1* were bound by SIX1, which suggests the existence of the regulatory loop among Six family members (Fig. 17 A).

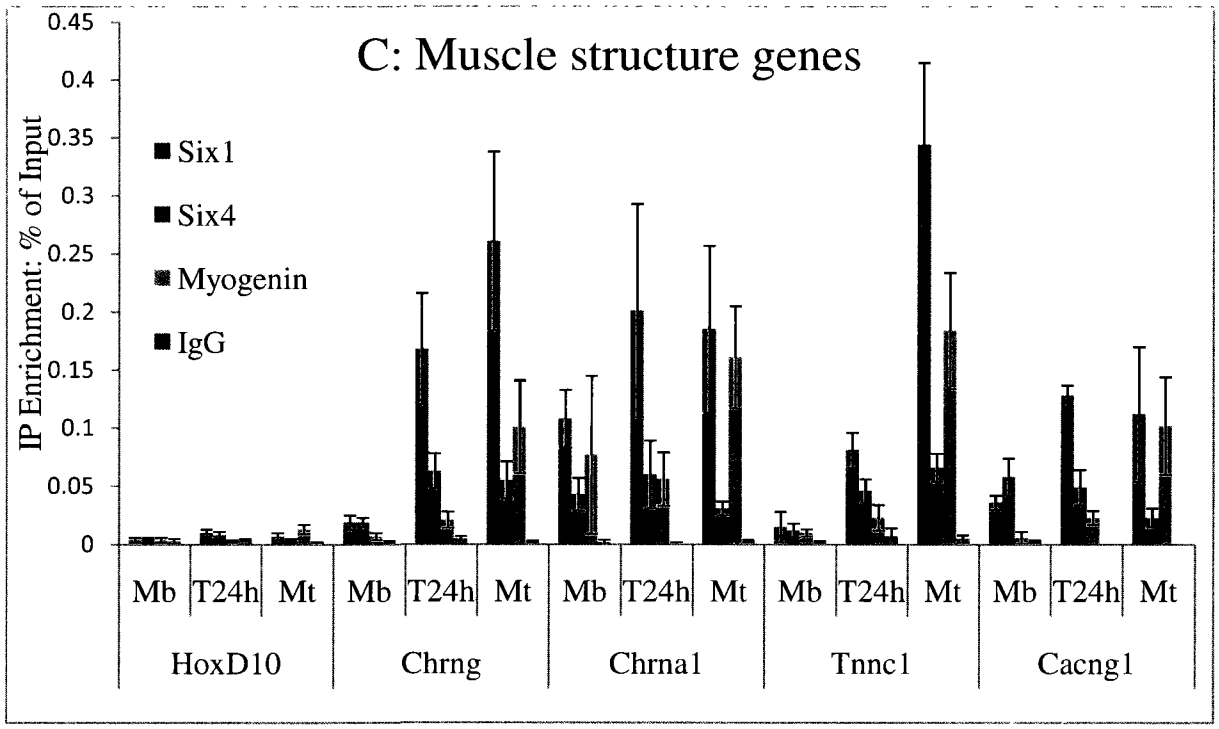
Some of the MRF members were shown to be downstream of SIX1 (directly or not) during myogenesis (Giordani et al., 2007; Grifone et al., 2005; Spitz et al., 1998). Here, the direct binding of SIX1 on the regulatory DNA loci of MRFs in C2C12 was confirmed by q-ChIP (Fig. 17B). Myf5, Myod1 and Myog are all expressed and are regulated by Six1 in C2C12 (Liu, Chu, Chakroun, Islam, & Blais, 2010). The binding of SIX1 on *Myf5* increases during

muscle differentiation, while the expression of Myf5 gradually decreases (unpublished data, A.Chu). This suggests a negative regulation of Six1 on Myf5 expression. Six1 also increasingly binds *Myog* promoter region along the time course, while the expression of Myog increases at the beginning of differentiation. It suggests that Six1 may upregulate Myog expression. Six1 binds two enhancers of *Myod*: distal regulatory region (DRR) and CER, while the expression of Myod is consistent in C2C12. Taken together, all evidence suggests that Six1 may have dual effect on the expression of MRFs as both activator and repressor during muscle differentiation.

Figure 17 Validation of a few Six1 targets found in the genome-wide microarrays




B: Regulation of MRFs
IP Enrichment: % of Input
0.3
0.25
0.2
0.15
0.1
0.05
0
Six1
Six4
Myogenin
IgG
Mb T24h Mt Mb T24h Mt Mb T24h Mt Mb T24h Mt Mb T24h Mt
HoxD10
Myf5
Myog
Myod-CER
Myod-DRR


C: Muscle structure genes
IP Enrichment: % of Input
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Six1
Six4
Myogenin
IgG
Mb T24h Mt Mb T24h Mt Mb T24h Mt Mb T24h Mt Mb T24h Mt
HoxD10
Chrng
Chrna1
Tnnc1
Cacng1

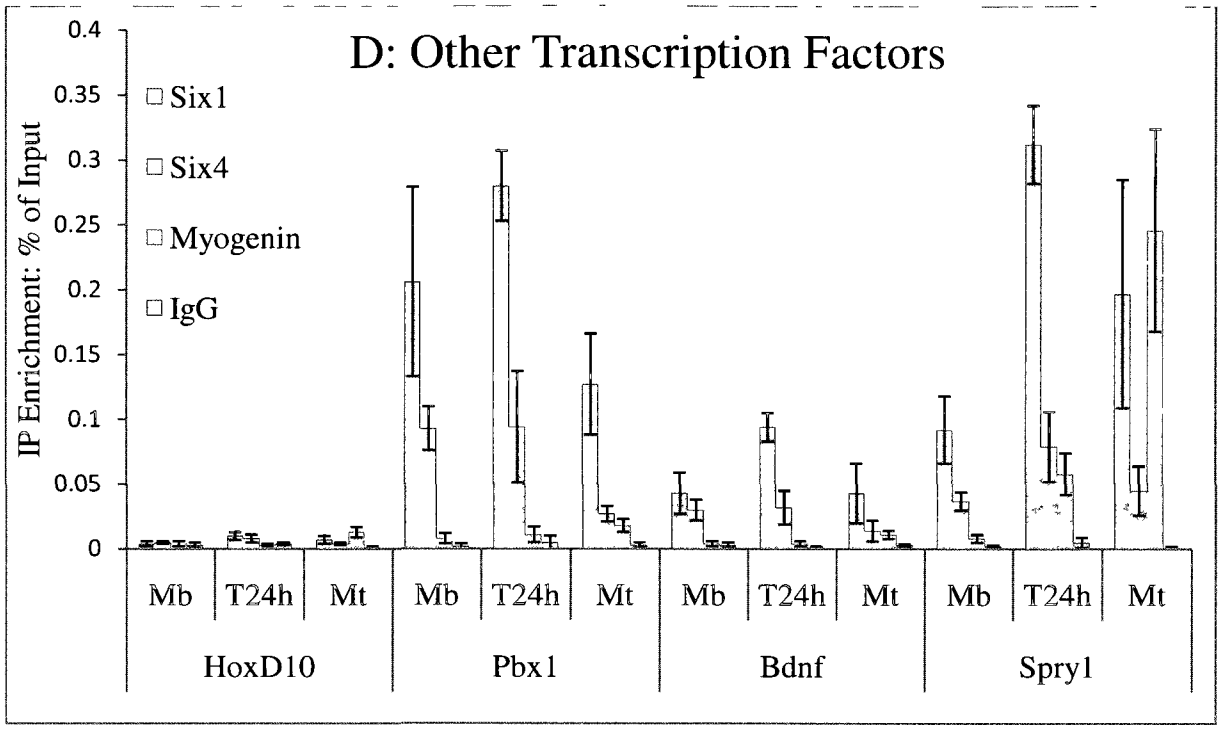


Fig. 17| q-ChIP was performed in Mb, T24 and Mt. The IP enrichment was shown as percentage of input DNA. Primer pairs amplifying Six1 binding sites in the regulatory regions of *Six1*, *Six4*, *Six5*, *Chrna1*, *Chrng*, *Cacng1*, *Tnnc1*, *Myf5*, *Myog*, *Myod* (CER and DRR), *Pbx1*, *Bdnf*, *Ldb3* and *Spry1* were used. The primer sequences are shown in Table 1. *Hoxd10* is the negative control locus. SEM is the standard error of the mean. At least three replicates were performed.

SIX1 binds a regulatory region at 130 kb upstream of *Myf5*. The -130 kb regulatory element is different from the reported SIX1 binding region at -57.5 kb of *Myf5*, while the -57.5 kb element was shown to be required for Myf5 expression in the limb buds and mature somites during embryonic development (Giordani et al., 2007). This is consistent with the spreading regulatory regions of Myf5. Different regulatory regions may be selectively required for expression at distinct developmental stages or anatomical sites.

In addition, Six1 was implicated in regulating the expression of genes encoding muscle structure proteins (Fig. 17C). As the components of nicotinic cholinergic receptor in neuromuscular junction, *Chrna1* and *Chrng* (cholinergic receptor, nicotinic, gamma polypeptide) are increasingly bound by Six1 during C2C12 differentiation, which suggests a positive role of Six1 in muscle maturation. *Cacng1* (Calcium channel, voltage-dependent, gamma subunit 1) is also targeted by SIX1. This gene is induced during differentiation and is important for muscle contraction upon neuronal stimulation. *Tnnc1* (Troponin C, cardiac/slow skeletal) is an essential component for muscle contraction and is specifically expressed in slow-twitch muscle. *Ldb3* (LIM domain binding 3) encodes a muscle structural protein, reported to interact with α-actin (Frey & Olson, 2002). The direct binding of SIX1 on structural muscle genes suggests that Six1 is functional in adult muscle development.

Several transcription factors involved in other types of organogenesis were also bound by SIX1 (Fig. 17 D). Pbx1 (Pre B-cell leukemia transcription factor 1) is bound by Six1 most extensively at T24. In early developmental stages, Pbx1 acts as the "pioneer" factor by marking some silent DNA loci in order to guide and enhance the binding of Myod on silent loci, such as the Myog promoter region (Gerber, Klesert, Bergstrom, & Tapscott, 1997). The consistent expression of Pbx1 drops in myotubes, when the myogenic programming is close to the end. The promoter region of *Six1* is bound by PBX1 in C2C12, which suggests a potential regulatory loop (Berkes et al., 2004). Bdnf (brain derived neurotrophic factor) is involved in the development of the nervous system (Ernfors, Lee, & Jaenisch, 1994; Jones, Farinas, Backus, & Reichardt, 1994) and was reported to be critical for satellite cell function and regeneration in vivo (Clow & Jasmin, 2010). Spry1 [(Sprouty homolog 1 (*Drosophila*)] is implicated in the development of the kidney (Basson et al., 2005; Basson et al., 2006). The

binding of SIX1 on the regulatory region of *Bdnf* and *Spry1* was weak in Mb and Mt, and was strong at T24.

Table 4 Validation of Six1 binding on several DNA loci using q-PCR

IP		Six1					Six4					Myogenin					Control IgG		
Gene	Time	Avg %Input	SEM	T test IgG	Fold over IgG	T test Ctrl	Avg %Input	SEM	T-test IgG	Fold over IgG	T-test Ctrl	Avg %Input	SEM	T-test IgG	Fold over IgG	T test Ctrl	Avg. %Input	SEM	T-test Ctrl
HoxD10	Mb	0 004	0 002	0 108			0 005	0 001	0 03			0 004	0 002	0 127			0 003	0 002	
	T24h	0 01	0 003	0 032			0 008	0 003	0 111			0 003	0 001	0 432			0 004	0 001	
	Mt	0 007	0 003	0 037			0 004	0 001	0 024			0 013	0 004	0 025			0 001	0 001	
Six4	Mb	**0.071**	0 022	0 014	30 8	0 008	**0.038**	0 011	0 016	16 59	0 01	0 006	0 003	0 107	2 6	0 303	0 002	0 001	0 293
	T24h	**0.294**	0 088	0 015	77 1	0 006	**0.096**	0 033	0 026	25 17	0 016	0 012	0 007	0 133	3 03	0 407	0 004	0 001	0 063
	Mt	**0.09**	0 042	0 043	10 7	0 04	0 026	0 012	0 038	3 08	0 081	0 02	0 011	0 061	2 38	0 148	0 008	0 006	0 414
Chrng	Mb	**0.019**	0 006	0 022	10 8	0 019	**0.019**	0 004	0 009	11 36	0 007	0 007	0 003	0 052	3 98	0 247	0 002	0 001	0 22
	T24h	**0.168**	0 048	0 012	36 4	0 005	**0.063**	0 015	0 008	13 69	0 005	0 021	0 007	0 021	4 61	0 086	0 005	0 002	0 1
	Mt	**0.261**	0 077	0 014	105	0 006	**0.055**	0 016	0 013	22 16	0 008	**0.101**	0 04	0 032	40 76	0 023	0 002	0 001	0 093
Myf5	Mb	**0.163**	0 037	0 011	201 1	0 001	**0.073**	0 021	0 02	89 72	0 004	0 006	0 002	0 036	7 13	0 298	0 001	0 001	0 159
	T24h	**0.237**	0 033	0 003	125 4	0	0 063	0 039	0 111	33 25	0 084	0 006	0 004	0 16	3 14	0 252	0 002	0 001	0 041
	Mt	0 044	0 021	0 065	29 7	0 047	0 011	0 004	0 039	7 35	0 231	0 01	0 003	0 034	6 84	0 223	0 001	0 001	0 076
Pbx1	Mb	**0.206**	0 073	0 035	118 9	0 008	**0.093**	0 017	0 006	53 55	0	0 008	0 004	0 064	4 75	0 177	0 002	0 002	0 257
	T24h	**0.28**	0 027	0 001	52 6	0	0 094	0 043	0 064	17 64	0 031	0 011	0 006	0 038	2 11	0 417	0 005	0 005	0 242
	Mt	**0.127**	0 039	0 024	37 5	0 005	0 027	0 006	0 021	7 88	0 009	0 018	0 005	0 037	5 26	0 042	0 003	0 002	0 19
Tnnc1	Mb	0 015	0 013	0 206	8 7	0 161	0 012	0 006	0 134	7 1	0 09	0 01	0 003	0 038	5 96	0 077	0 002	0 001	0 272
	T24h	**0.081**	0 015	0 039	11	0	0 046	0 01	0 077	6 22	0 003	0 023	0 011	0 056	3 17	0 089	0 007	0 007	0 375
	Mt	**0.344**	0 071	0 019	70 4	0	**0.066**	0 012	0 009	13 51	0	**0.184**	0 05	0 033	37 68	0 002	0 005	0 003	0 333
Bdnf	Mb	**0.043**	0 016	0 034	14 7	0 016	**0.03**	0 008	0 036	10 14	0 006	0 004	0 002	0 332	1 41	0 006	0 003	0 002	0 397
	T24h	**0.094**	0 011	0 001	108 5	0	0 032	0 013	0 053	36 45	0 056	0 004	0 002	0 053	5 02	0 056	0 001	0 001	0 026
	Mt	0 043	0 023	0 077	16 1	0 056	0 014	0 008	0 11	5 14	0 207	0 011	0 003	0 014	4 16	0 207	0 003	0 001	0 127
Chrna1	Mb	**0.108**	0 025	0 012	45 1	0 001	**0.043**	0 014	0 039	17 76	0 009	0 077	0 068	0 18	31 94	0 131	0 002	0 002	0 332
	T24h	0 201	0 092	0 059	164 3	0 026	0 06	0 029	0 063	49 16	0 044	**0.056**	0 023	0 047	45 61	0 031	0 001	0 001	0 03
	Mt	**0.185**	0 072	0 042	72 8	0 013	**0.031**	0 006	0 006	12 3	0 003	**0.161**	0 044	0 018	63 4	0 003	0 003	0 001	0 12
Ldb3	Mb	**0.092**	0 029	0 019	13 2	0 006	**0.122**	0 035	0 015	17 51	0 003	0 013	0 009	0 106	1 93	0 135	0 007	0 005	0 294
	T24h	**0.179**	0 026	0 016	81 6	0	0 098	0 026	0 051	44 56	0 003	0 026	0 02	0 099	12 01	0 196	0 002	0 001	0 069
	Mt	**0.203**	0 062	0 021	22	0 004	0 076	0 039	0 063	8 22	0 041	0 025	0 011	0 131	2 67	0 063	0 009	0 007	0 381
Pax3	Mb	**0.092**	0 026	0 018	52 4	0 003	**0.037**	0 007	0 01	21 19	0 001	0 008	0 003	0 041	4 61	0 154	0 002	0 001	0 253
	T24h	**0.312**	0 03	0 001	68 5	0	**0.079**	0 027	0 037	17 27	0 012	**0.058**	0 016	0 023	12 83	0 006	0 005	0 004	0 161
	Mt	0 197	0 088	0 057	177 8	0 022	**0.045**	0 019	0 049	40 83	0 027	**0.246**	0 078	0 026	221 67	0 005	0 001	0 001	0 06
Cacng1	Mb	**0.036**	0 006	0 017	12 9	0 001	**0.058**	0 016	0 041	20 96	0 002	0 006	0 005	0 308	2 2	0 34	0 003	0 001	0 375
	T24h	**0.128**	0 009	0 003	245 8	0	**0.049**	0 015	0 04	94 6	0 007	**0.023**	0 006	0 029	45 13	0 031	0 001	0	0 04
	Mt	0 112	0 058	0 097	3051 9	0 024	**0.023**	0 008	0 048	620 89	0 028	0 102	0 042	0 067	2779 3	0 011	0	0	0 06
Six5	Mb	**0.064**	0 024	0 035	35	0 013	**0.055**	0 019	0 04	30 33	0 01	0 006	0 004	0 081	3 47	0 298	0 002	0 001	0 263
	T24h	**0.112**	0 021	0 007	100	0 001	0 039	0 018	0 064	34 38	0 06	0 004	0 001	0 01	3 53	0 085	0 001	0	0 027
	Mt	**0.036**	0 009	0 014	41 6	0 005	**0.017**	0 004	0 013	19 05	0 039	**0.018**	0 005	0 025	20 35	0 043	0 001	0 001	0 053
Six1	Mb	**0.058**	0 018	0 025	36 1	0 005	0 039	0 017	0 057	24 37	0 028	0 002	0 002	0 489	0 95	0 239	0 002	0 001	0 244
	T24h	**0.166**	0 028	0 004	46 4	0	0 04	0 016	0 06	11 12	0 039	0 006	0 001	0 249	1 6	0 169	0 004	0 002	0 091
	Mt	0 074	0 037	0 073	22 2	0 038	0 014	0 009	0 138	4 32	0 194	0 018	0 009	0 119	5 31	0 117	0 003	0 002	0 171
Myog	Mb	**0.021**	0 008	0 036	20	0 038	**0.012**	0 002	0 01	11 02	0 029	0 006	0 003	0 071	5 86	0 268	0 001	0 001	0 15
	T24h	**0.069**	0 015	0 007	23 6	0 003	**0.033**	0 012	0 035	11 22	0 048	0 012	0 005	0 027	4	0 375	0 003	0 002	0 05
	Mt	**0.144**	0 044	0 015	54 8	0 008	**0.047**	0 018	0 039	17 96	0 032	**0.028**	0 007	0 008	10 7	0 012	0 003	0 001	0 108
Myod-CLR	Mb	**0.06**	0 022	0 022	26 5	0 017	**0.039**	0 014	0 031	17 36	0 017	0 005	0 003	0 022	2 12	0 399	0 002	0 002	0 313
	T24h	**0.196**	0 052	0 011	38 2	0 004	**0.069**	0 022	0 025	13 43	0 014	0 039	0 011	0 019	7 53	0 017	0 005	0 002	0 142
	Mt	**0.148**	0 059	0 034	62 6	0 022	**0.025**	0 008	0 019	10 71	0 029	**0.133**	0 052	0 033	55 88	0 022	0 002	0 001	0 086
Myod-DRR	Mb	0 008	0 004	0 11	5 3	0 171	0 012	0 002	0 003	7 74	0 02	0 005	0 002	0 096	3 18	0 374	0 002	0 001	0 221
	T24h	**0.039**	0 006	0 01	16 8	0 002	0 019	0 007	0 057	8 03	0 133	0 01	0 002	0 081	4 14	0 486	0 002	0 002	0 06
	Mt	**0.096**	0 029	0 021	70 1	0 005	0 014	0 003	0 025	10 31	0 065	**0.075**	0 011	0 004	54 7	0	0 001	0 001	0 075

Table 4| Several Six1 binding sites in the genome-wide study were validated using q-ChIP. For each DNA locus, at least triplicates were performed with biologically independent samples for each time point. *p* value was calculated by T-test for comparing: the testing locus with the negative control locus (*Hoxd10*) (t-test ctrl), and the testing transcription factor with the negative control antibody (IgG) (t-test IgG). SEM is the standard error of the mean. Cells in yellow indicate the corresponding loci that meet all three significance tests (t-test IgG < 0.05, t-test ctrl < 0.05, and Fold over IgG > 10).

6. MRFs can co-occupy some loci with Six1

The genome-wide binding location analysis of Myod and Myog in a mouse gene promoter study revealed that the MEF3 motif, Six1 binding element, was significantly enriched among MRF targets which were induced during muscle differentiation (Blais et al., 2005). Reciprocally, E-box was also significantly enriched among Six1 binding sequences in this study (Fig. 15). A ChIP sequencing study (ChIP-seq), which is an alternative technology to map protein-DNA interactions, was performed with antibodies directed against MYOD by Dr. Tapscott's group(Cao et al., 2010). A bioinformatics analysis, the "tables" function on the genome browser website, was used to examine the overlapping binding sites between the Six1 ChIP-on-Chip data and the Myod ChIP-seq data, according to DNA sequence alignment. I found that approximately 40% of genomic regions that were bound by SIX1 overlapped with Myod bindings sites (Fig. 18). This striking overlap between Myod and Six1 binding strongly suggests that the two factors have the potential to regulate the expression of common target genes together.

Figure 18 SIX1 and MYOD co-bind some DNA loci

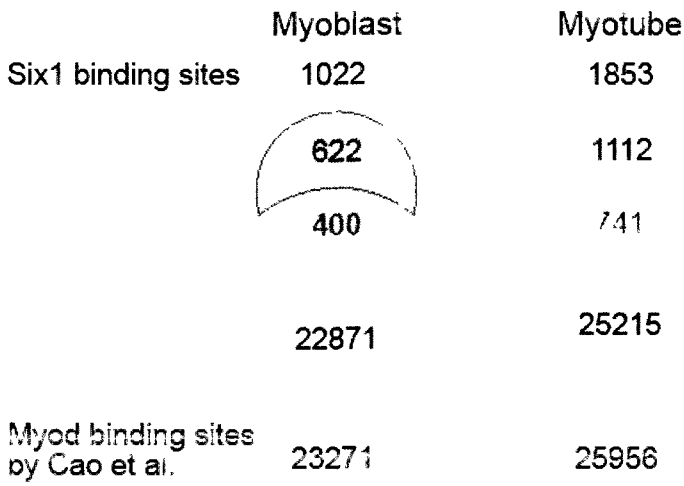


Fig. 18| Venn diagram of the binding sites of SIX1 and MYOD, showing roughly 40% of SIX1 binding sites have concurrent MYOD binding nearby. 1022 SIX1 binding sites and 23271 MYOD binding sites were present in myoblast. 400 of SIX1 binding sites were bound by MYOD in myoblast. 1853 SIX1 binding sites and 25956 MYOD binding sites were present in myotubes. 741 of SIX1 binding sites were bound by MYOD in myotubes.

q-ChIP assay also showed the co-binding of SIX1 and MYOG on several DNA loci, such as the regulatory regions of *Spry1*, *Myod* (CER and DRR), *Myog*, *Chrna1*, *Chrng*, *Tnnc1*, and *Cacng1* (Fig. 17, Table 4). This further substantiates the claim of co-regulation between SIX1 and the MRFs.

7. Six4 can co-occupy some loci with Six1

Six1 and *Six4* have similar expression patterns during mouse embryonic development. *Six1*$^{-/-}$ mutant mice die at birth because of severe muscle defects (Laclef et al., 2003). *Six4*$^{-/-}$ mutant mice survive without detectable muscle defects, a result possibly due to the compensatory

function of Six1 or Six5 (Ozaki et al., 2001). The SIX4 protein is expected to recognize the same MEF3 DNA sequence element as SIX1. Here, qChIP assays confirmed that SIX4 was often recruited to DNA loci targeted by SIX1, such as the regulatory regions of *Chrna1*, *Chrng*, *Tnnc1*, *Cacng1*, *Ldb3*, *Myf5*, *Myog*, *Myod*, *Pbx1* and *Spry1* (Fig. 17). The binding enrichment of SIX4 at target genes was lower than that of SIX1, but this could be due to the lower affinity of the anti-SIX4 antibody. The data do not reveal whether the binding of SIX1 indeed compensates for the absence of Six4 in *Six4*$^{-/-}$ embryos, but nevertheless confirm that these two factors most likely have overlapping functions by binding to identical genomic loci.

Discussion

The first functional genome-wide mapping of SIX1 binding sites was performed using ChIP-on-Chip. Until now, the systematic molecular basis of Six1 function remained poorly characterized. Furthermore, the role of Six1 in muscle differentiation and muscle repair after stimuli is not completely understood. In this project, the global Six1 function was addressed by functional genomics analysis following ChIP-on-Chip.

1. ChIP-on-Chip is an efficient platform to study the interaction between transcription factors and DNA

Here, the first mapping of Six1 global binding sites was described using the Agilent ChIP-on-Chip platform. It was reported that the choice of amplification methods, platforms and detecting algorithms had little effect on the most enriched targets (Johnson et al., 2008). All microarray platforms are very sensitive at high enrichment levels, while at extremely low enrichment levels, platforms with long oligonucleotides were more sensitive (Johnson et al., 2008). Agilent has comparable detecting sensitivity and specificity, and compares advantageously to other platforms. The CisGenome program used to detect binding events is a well-established method, which has been widely used (Ji et al., 2008).

ChIP-sequencing (ChIP-seq) is an alternative technology to study protein-DNA interaction, which combines ChIP with following massive parallel sequencing. In ChIP-seq, all immunoprecipitated DNA fragments will be sequenced and aligned to the genome sequence or origin, which provides more precise binding information and minimizes information loss at repetitive regions (which cannot be unambiguously resolved using microarrays). Ji et al. reported that compared to ChIP-seq, array-specific peaks may represent technical noise (Ji et al., 2008), but this remains hypothetical at the moment.

Figure 19 the demonstration of the microarray design

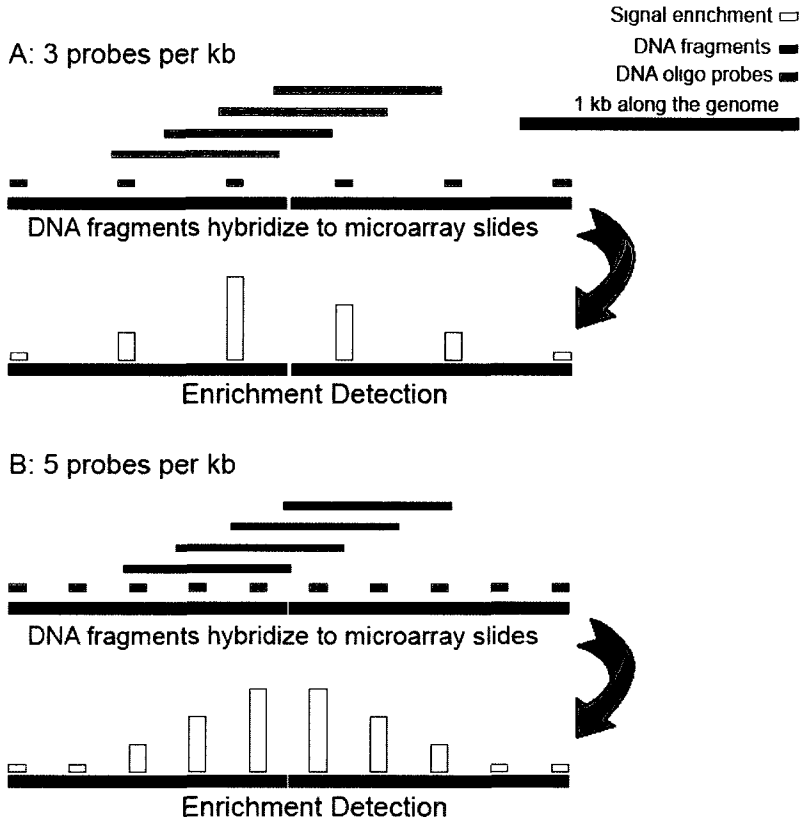


Fig. 19| The demonstration of the microarray design: the red bars represent the typical 60 bp oligo probes. With the same DNA fragments, microarrays with denser probes will give rise to narrower and higher signal detection.

As a DNA tiling assay, ChIP-on-Chip has its own limitation. It can only cover non-repetitive genome sequences. The spacing between neighboring probes is crucial for the signal resolution (Fig. 19). Smaller spacing helps to discover the binding events more sensitively. On the other hand, the number of probes is limited on each microarray. With unceasing improvement of probe density, ChIP-on-Chip has smaller spacing between probes, resulting

more precise detection. ChIP-on-Chip has been broadly and efficiently used and manifested itself as a good choice to study protein-DNA interactions in large scale.

Starting with a massive amount of location data, the method of analysis is very important to ensure a proper interpretation. It is a general question how to annotate a related gene for a binding site. The regulated genes can be simply predicted based on the proximity. However, it may be inaccurate for a particular locus: it may affect a distant gene other than the gene nearby (Fig. 12C). Therefore a customized algorithm may be more representative for specific data. Considering the importance of Six1 in muscle development, the genes modulated due to the absence of *Six1* (caused by RNA interference, in our case) would be the most likely regulated genes. If no gene in the proximity is modulated, the closest gene would be chosen as the potential target. The selection criteria may contribute a slight bias for genes whose expression is affected by Six1 knock-down, whether they are directly targeted by Six1 or not. But we believe that correct annotations are more numerous than if binding sites were annotated on the sole basis of their proximity to genes. Our annotation scheme will miss putative instances where one binding event regulates two genes concurrently.

2. Six1 targets regulating muscle development and muscle function

Taking C2C12 as a muscle regeneration model, a set of Six1 target genes was discovered in this study, during and after differentiation of myoblasts in a genome-wide manner. This study substantially broadens its previously known target pool. Several functional categories are significantly abundant among all the genes potentially regulated by Six1. Some genes in those enriched functional categories were reported to be regulated by Six1 in previous embryonic studies, such as genes related to fast-twitch muscle (Niro et al., 2010). For the

vast majority of Six1 target genes identified here, it is the first report of SIX1 binding on their regulatory regions.

The most enriched gene categories among Six1 genes may reflect the general biological function themes regulated by Six1. Using a myoblast cell line, as expected, muscle development is among the most enriched categories. Pax3 is crucial for embryonic myogenesis as the main regulator of the delamination and specification of muscle. Pax3 and Six1 seem to have mutual regulatory loop. Based on the mutant mouse model, Six1/4 were required for the proper expression of Pax3 in the dermomyotome. At the same time, ectopic expression of Pax3 in P19 embryonic carcinoma cells, which have myogenic potential, can induce expression of Six1. Here a putative regulatory binding site for Six1 was found at 150 kb upstream of the TSS of *Pax3*. In addition, as members of MRFs, Myf5, Myod and Myog are the key factors modulating muscle development from determination to differentiation. *Myf5* and *Myog* were reported to be directly regulated by Six1 during embryogenesis (Giordani et al., 2007; Spitz et al., 1998). In this study, Six1 binding to the regulatory regions of *Myf5* and *Myog* was confirmed in C2C12 cells through ChIP-on-Chip and q-ChIP, and in the case of *Myod* was also confirmed by EMSA. SIX1 can bind two enhancers of Myod: CER and DRR. DRR and CER of *Myod* were shown to be important for the proper expression of Myod during embryonic myogenesis (Chen, Ramachandran, & Goldhamer, 2002; Kucharczuk, Love, Dougherty, & Goldhamer, 1999). Considering that the location of *Myf6* is close to that of *Myf5* (8.6 kb away), it is not clear whether *Myf5* and *Myf6* share the same regulatory region. This evidence suggests that Six1 may be broadly involved in regulating the expression of all MRFs. As a target of Six1, Pitx2 (paired-like homeodomain transcription factor 2) is expressed in embryonic and adult muscles and is required for the

development of hind limb buds (Marcil, Dumontier, Chamberland, Camper, & Drouin, 2003; Shih, Gross, & Kioussi, 2007). Taken together, all evidence suggests that Six1 takes part in the regulatory network of myogenesis through regulation of the expression of transcription factors at different developmental stages. A rescue experiment, over-expressing one of Six1 downstream genes that encode important transcription factors (such as Myog) in Six1 knock-out or knock-down cells, would help to validate this hypothesis.

Besides its participation in muscle development through regulating other transcription factors, SIX1 also directly targets genes encoding muscle structure proteins. Tnnc1, Tnnt1 (troponin T1, skeletal, slow), Tnnt3 (troponin T3, skeletal, fast), Tnni1 (troponin I, skeletal, slow 1), Tnni2 (troponin I, skeletal, fast 2), and Tnni3 (troponin I, cardiac 3) are the essential components for muscle contraction. Tnnc1, Tnnt1, Tnni1 are specifically expressed in slow-twitch muscle, while Tnnt3 and Tnni2 are specifically expressed in fast-twitch muscle. Although Six1 was reported to be able to establish a fast-twitch type of muscle, here it seems that Six1 can bind and may regulate genes involved in establishing both types of myofibers (Grifone et al., 2004).

Moreover, Six1 target genes encoding the subunits of channels, such as ATPase-dependent channels and voltage-dependent calcium channel, which are essential for muscle excitation. Atp1a1 (ATPase, Na+/K+ transporting, alpha 1 polypeptide), Atp1a2 (ATPase, Na+/K+ transporting, alpha 2 polypeptide), Atp2a1, and Atp2b1 (ATPase, Ca++ transporting, plasma membrane 1) were bound by SIX1. These proteins are important for muscle to maintain its contractive function. Some genes encoding components of the voltage-dependent calcium channel on the postsynaptic muscle plate are also targeted by Six1, such as Cacna1s (calcium channel, voltage-dependent, L type, alpha 1S subunit), Cacnb1 (calcium channel, voltage-

dependent, beta 1 subunit), Cacna2d1 (calcium channel, and voltage-dependent, alpha2/delta subunit 1), Cacnb2 (calcium channel, voltage-dependent, beta 2 subunit), and Cacng1. These subunits are essential for muscle contraction. Taken together, these results suggest that Six1 is a key regulator of muscle development and function in many layers: from the embryonic stage to adult stage; from regulation of other transcription factors to regulation of essential muscle structural genes. Of note, some of those genes were reported to be MRF targets, which suggests that Six1 can not only regulate the expression of MRFs but also function in parallel with MRFs (Fig. 20).

Figure 20 Six1 may function in cascade and in parallel with MRFs in adult skeletal muscle

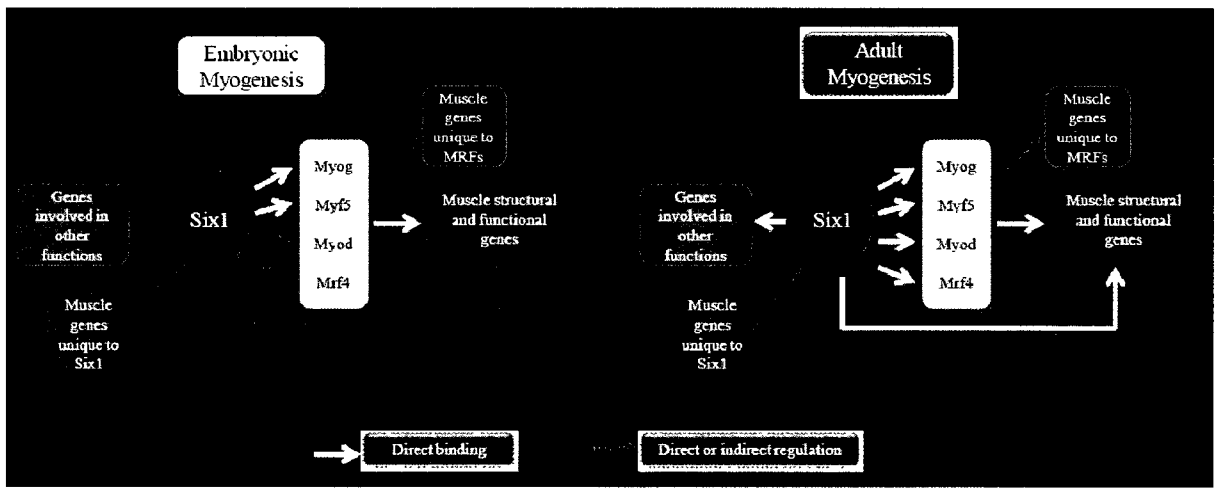


Fig. 20| A) In embryonic myogenesis, Six1 was shown to be able to directly regulate the expression of Myog and Myf5 (Giordani et al., 2007; Spitz et al., 1998). Six1 may directly or not regulate the expression of Myod and Myf6 (Grifone et al., 2005; Laclef et al., 2003). All four members of MRFs were able to induce structural and functional muscle genes, may be or may not be under the control of Six1(Aurade et al., 1994; Braun et al., 1989; Braun et al., 1990; Weintraub et al., 1989; Wright et al., 1989). Some structural and functional muscle

genes may be unique to Six1. Six1 is required for controlling the development of other organs (Ikeda et al., 2007; Xu et al., 2003). B) During adult myogenesis, Six1 putatively binds the regulatory regions of all four MRF family genes. Six1 can directly target genes involved the development of other organs and muscle structure and function.

Six1 and MRFs (Myod and Myog) have common targets, including some structural and functional muscle genes. Furthermore, Six1 seems to have a role in regulating muscle response to nervous system stimuli. Some of Six1 targets encode subunits of postsynaptic transmembrane receptors which receive the signal of neurotransmitters, such as acetylcholine receptor. Chrng, Chrna1 and Chrna9 (cholinergic receptor, nicotinic, alpha polypeptide 9) are the subunits of an acetylcholine receptor. The nicotinic acetylcholine receptors are activated by acetylcholine signals from the motor neuron and are critical to trigger an action potential for muscle contraction. This suggests a potential role of Six1 in establishing the skeletal muscle plate in the neuromuscular junction. A conditional knock-out of *Six1* in skeletal muscle will help to validate this speculation.

3. Myogenic regulatory network: possible cooperation among different transcription factor family

The precise control of gene expression in a specific tissue usually depends on a regulatory network including several transcription factors. They may function as a cascade, or work in a combinatorial manner. A good example is the regulation of the expression of Myog. *Myog* promoter region recruited PBX1, MYOD, SIX1, and MEF2C (myocyte enhancer factor 2C) (Tapscott, 2005) and is under the control of the integrated signal in myogenic cells. Myog expression was reduced due to the loss of SIX1, which suggests that SIX1 acts functionally upstream of Myog [(Laclef et al., 2003) and unpublished data by A., Chu]. Besides the

regulation cascade, Six1 may function in parallel to the MRFs. Previous analysis of promoter sequences bound by MYOG revealed that MEF2 DNA element (DNA sequence bound by MEF2) and MEF3 DNA element were significantly enriched in Myog targets which were most induced during differentiation. The co-occupation of MEF2 and MYOG, as well as MYOG and SIX1, was confirmed by ChIP in previous and this study respectively (Blais et al., 2005). MEF2 and MYOG were reported to be able to synergistically activate the *Myf6* promoter during myogenesis (Naidu, Ludolph, To, Hinterberger, & Konieczny, 1995). Furthermore, in this study, MRF binding element (E-box) is significantly enriched among the global binding sequences of SIX1. Indeed many Six1 targets are also Myog targets (Table 4). All evidence supports the co-occupation of Six1 and Myog.

Recently, Dr. Tapscott's group reported the global binding sequences of MYOD using ChIP-seq (Cao et al., 2010). Approximate 40% of Six1 binding sites in myoblast (400 out of 1022) and in myotubes (741 out of 1853) overlap with Myod binding sites. It strongly supports that MYOD and SIX1 have combinatorial function in a genome-wide fashion by co-binding the neighboring DNA loci.

Among Six1 binding sequences, E-box DNA motif is significantly enriched. However, taking the MYOD binding data from the ChIP-seq study (Cao et al., 2010), the MEF3 motif (discovered in this study) is not significantly enriched using the known motif mapping program. There are two possibilities explaining this divergence. First, Six1 may target a small proportion of Myod targets, so its binding motif cannot stand out as being significantly enriched. MYOD has much more spreading binding than SIX1 does. Second, some binding sites of SIX1 may not be discovered by the Myod peak calling (perhaps very close to MYOD binding sites), since the platforms (ChIP-on-Chip *vs.* ChIP-seq) and the peak calling

algorithms were different between these two experiments. Consequently, some or most important (i.e. functional) MEF3 elements may be missed among Myod binding sequences.

Interestingly, there may be auto-regulation among Six family members. SIX1 and SIX4 can bind the promoter regions of *Six1*, *Six4* and *Six5* (Fig. 17A, table 4). However, it is contradictory to the report in a mouse embryonic study that Six1 did not regulate its own transcription and did not affect the expression of Six4 and Six5 (Laclef et al., 2003). Several possibilities may explain this apparent inconsistency. First, the binding of Six1 and Six4 may be required but not essential to induce their expression. Second, in vivo (in the mouse embryo) and in vitro (in the C2C12 cell line) systems may be characterized by some intrinsic chromatin differences. Thirdly, it may be due to the different developmental stages. In embryonic development, other transcription factors, such as Pax3, may compensate to regulate the expression of Six genes. In C2C12, such transcription factors may be expressed at low level which cannot allow compensation to occur.

Six4 is another interesting muscle-relevant Six family member. *Six4*$^{-/-}$ mutant mice survived without major developmental defects, which was believed to be due to the compensation function of Six1 or Six5 (Ozaki et al., 2001). The expression of Six1 and Six4 have similar pattern in mouse embryos. Six4 and Six1 were reported to be important for the similar organogenesis affecting many organs (Grifone et al., 2005; Kobayashi et al., 2007; Niro et al., 2009). Several common targets were identified in C2C12 model at T24 (Fig.17). When Six4 global binding data become available, it could help to determine whether Six1 can substitute for the absence of Six4 in *Six4*$^{-/-}$ mice.

4. Several other gene functional categories are significantly targeted by SIX1

In addition to skeletal muscle, many tissues were affected, to various extents, by loss of function of Six1 during embryogenesis, such as sensory tissues and kidney (Laclef, Souil et al., 2003; Xu et al., 2003). In those previous studies, a small set of Six1 targets were identified by classical biological methods. Six1 seems to have broad function in controlling development. The wider population of Six1 targets identified here will help to interpret how Six1 is involved in these processes.

Despite the fact that this study was performed with a myoblast cell line model, genes involved in kidney development and neurogenesis were enriched among Six1 targets. Some key factors controlling the function or development of other tissues were discovered to be Six1 targets as well. Pbx1 played a crucial role in mesenchymal function during renal morphogenesis (Schnabel, Godin, & Cleary, 2003). Gdnf was described as a key regulator of the formation of ureteric bud and ureteric branching (Pichel et al., 1996). Spry1 was shown to be able to negatively regulate the number and size of ureteric buds. The balanced opposing function of Gdnf and Spry1 seems to be important for the proper ureteric branching morphogenesis. The binding of SIX1 on the regulatory regions of both *Gdnf* and *Spry1* may explain the crucial role of SIX1 in kidney development during embryogenesis.

Genes involved in neurogenesis were over-represented in Six1 targets as well. NeuroD6 (neurogenic differentiation 6) was required for the formation of several cell types in the nervous system, including inner ear sensory neurons and retinal photoreceptor cells (Kim et al., 2001; Pennesi et al., 2003). Bdnf was required for the survival of many cell types in the nervous system, especially affecting the neurons derived from the ectodermal placodes (Ernfors et al., 1994; Jones et al., 1994). Hes6 [Hairy and enhancer of split 6 (*Drosophila*)] can promote the neuronal fate during cortical development (Jhas et al., 2006). Several

previous studies showed that Six1 was required for sensor organ development in embryogenesis (Brugmann, Pandur, Kenyon, Pignoni, & Moody, 2004; Christophorou, Bailey, Hanson, & Streit, 2009; Zou, Silvius, Fritzsch, & Xu, 2004). All evidence suggests that Six1 can regulate neurogenesis through transcriptional modulation.

However, it is still unexpected to have such significantly enriched gene categories in Six1 targets. It may be due to several possibilities. First, genes in those categories may have some function in muscle development as well, which may not have been fully characterized yet. As demonstrated before, some transcription factors were involved in both muscle and other developmental processes. Besides its role in neurogenesis, Hes6 was shown to be able to promote proliferation and repress differentiation of C2C12 myoblasts (Cossins, Vernon, Zhang, Philpott, & Jones, 2002). Mef2c has multiple roles in development: Mef2c was implicated in the differentiation of cardiomyoblasts into cardiomyocytes (Karamboulas et al., 2006); Mef2c was also required for neural cell differentiation (H. Li et al., 2008); Mef2c manifested its role in regulating the cell fate between lymphoid and myeloid in multipotent progenitors (Stehling-Sun, Dade, Nutt, DeKoter, & Camargo, 2009); Mef2c was also involved in maintaining the proper muscle structure (Potthoff et al., 2007). Pbx1 was reported to have a role in both myogenesis and nephrogenesis (Berkes et al., 2004; Schnabel et al., 2003). Among Six1 targets, some genes, like *Mef2c* and *Pbx1*, may contribute to divergent enriched functional categories. Furthermore, it is very likely that some genes may have some previously uncharacterized effect on muscle development. It can partially explain how such genes were bound by SIX1 in a myoblast cell line. Basically, Six1 can regulate muscle development through modulating the expression of those muscle-related genes.

Second, Six1 may bind genes with no or little expression in muscle, acting as a component of the repressor complex to reduce the expression of these genes, which may be required to keep the characteristics of the muscle lineage. Combined analysis with gene expression profiling data in a time course of C2C12 cells (the work was done by Alphonse Chu in our lab), only a small proportion of Six1 targets (251 out of 2242) have undetectable expression in C2C12. These 251 genes were preferentially expressed in nervous system (brain, neural stem cell, and spinal cord) according to the tissue expression analysis using the David bioinformatics resource (Huang da et al., 2009). This finding suggests that the expression of this set of genes is probably repressed by a mechanism that involves SIX1, in cells committed to the muscle lineage.

Thirdly, Six1 may bind some genes as a scaffold of the transcriptional complex, lacking its tissue-specific co-activators in C2C12. These co-activators may be critical for the induction of these genes in other tissues, but would be missing in C2C12 cells. Pbx1 non-preferentially binds M*yog* promoter and marks this locus for Myod binding in myogenic cells (Berkes et al., 2004; Maves et al., 2007). Six1 lacks a recognizable transactivation domain. At early developmental stage, Six1 may occupy these loci and recruit its co-activators to initiate the expression of the targeted genes in multi-potent cells. Later in the committed stage, the functional level of Six1 co-activators is reduced, and therefore activation of these genes cannot be initiated even with the persistent Six1 binding. One example could be Robo2 [roundabout homolog 2 (*Drosophila*)]. Robo2 had a role in guiding olfactory sensory neuron axons in the olfactory bulb (Cho, Lepine, Andrews, Parnavelas, & Cloutier, 2007). Loss of function study showed that Six1 was critical for early neurogenesis during the development of olfactory epithelium (Ikeda et al., 2007). Although it was never reported that Six1 directly

regulates the expression of Robo2 in olfactory epithelium, it is shown here that Six1 is present in the proximity of *Robo2*. It is likely that Six1 plays a role in Robo2 expression in multi-potent cells through cooperation with its co-activators, such as Eya1, to promote neurogenesis and Six1 retains its binding in another cell fate (Zou et al., 2004). In fact, most of Six1 targets are more or less expressed during C2C12 differentiation, which may be due to the basal expression level of these genes without further functional activation by other co-regulators. This phenomenon may reflect multiple roles of Six1 in the early developmental stage.

5. The involvement of Six1 in signaling pathways

Several intracellular signaling pathways were significantly enriched among Six1 targets, such as MAPK signaling pathway, TGF-beta signaling pathway, p53 signaling pathway, apoptosis and Wnt signaling pathway. Some previous studies provided evidence supporting the involvement of Six1 in these processes. However, a more comprehensive understanding of the potential role of Six1 in each of these pathways will require further investigation.

Among these signaling pathways, Wnt signaling pathway has been studied extensively. The Wnt signaling pathway played an important role in controlling cell growth and differentiation during both normal development and tumorigenesis (Huang & He, 2008; Petersen & Reddien, 2009). More specifically, Wnt signalling pathway was also involved in myogenesis. Wnt1 and Wnt7A signaling can activate the expression of Myf5 and Myod respectively (Borello et al., 2006; Tajbakhsh et al., 1998). β-catenin was a direct regulator of *Myf5* expression (Borello et al., 2006).

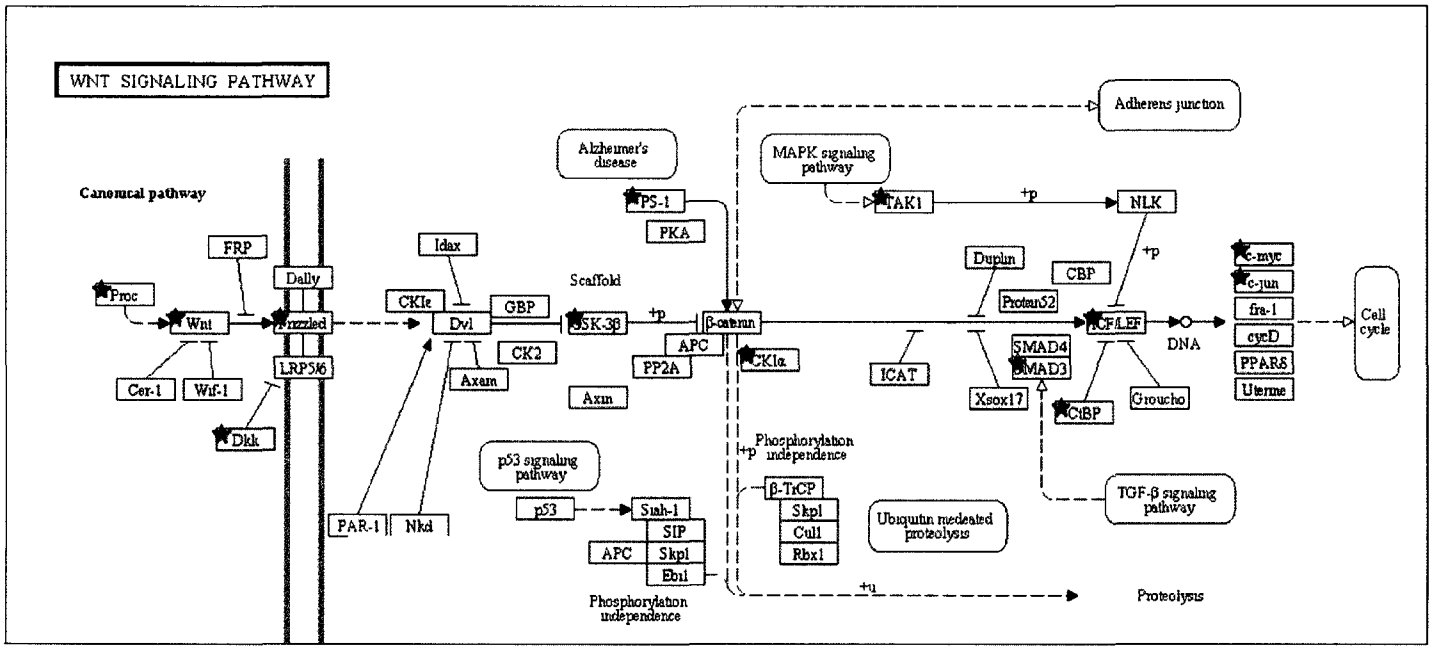
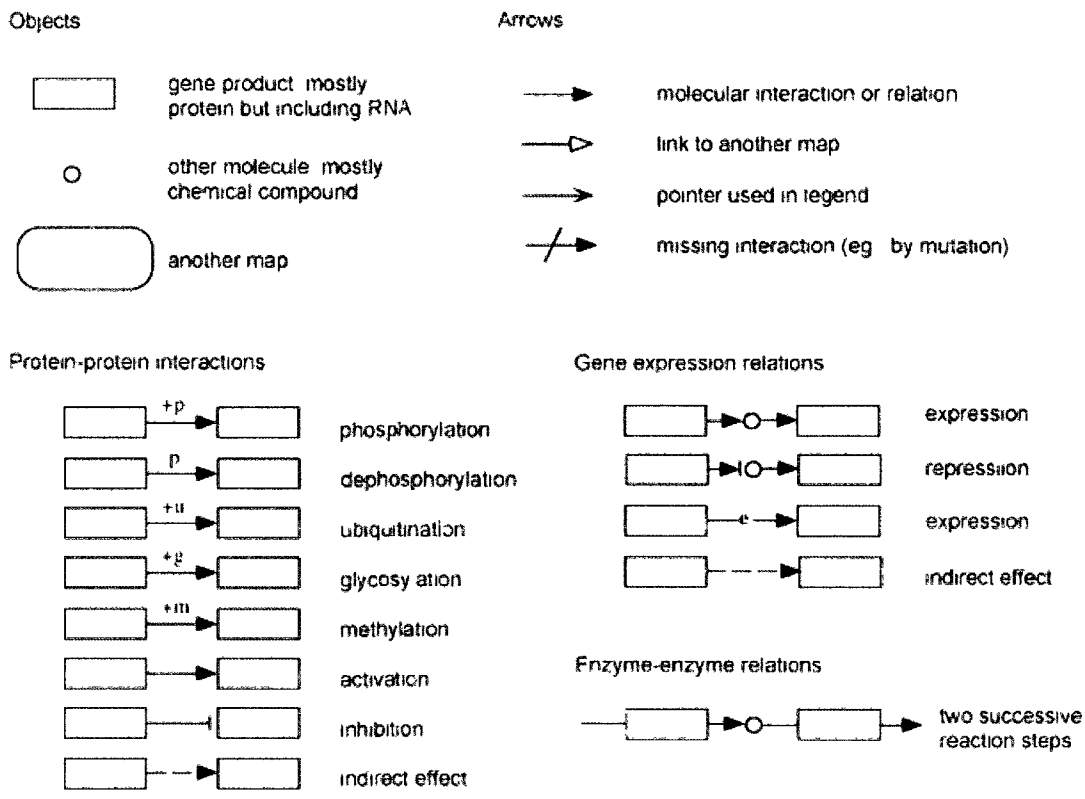


Fig. 21| A few genes encoding proteins involved in Wnt signalling pathway were discovered to be targeted by SIX1 in C2C12, using David Bioinformatics program (Huang da et al. 2009). The notation is illustrated as the following image. The red-starred components are targeted by Six1.



The notion is for Figure 21 (Huang da et al., 2009).

As a transcription factor expressed early in the mesoderm, Six1 may be implicated in Wnt signaling pathway in early embryogenesis. Six1 bound some DNA loci encoding Wnt signaling pathway components, including *Wnt6*, *Fzd4*, *Fzd7*, *Lef1*, *Tcf4* and *Tcf3* (Fig. 21). Wnt6 (wingless-related MMTV integration site 6) encodes a secreted ligand for members of the frizzled family of seven transmembrane receptors. Wnt6 is expressed in the limb ectoderm during embryogenesis (Witte, Dokas, Neuendorf, Mundlos, & Stricker, 2009). Fzd4 [Frizzled homolog 4 (Drosophila)] and Fzd7 [frizzled homolog 7 (Drosophila)] encode frizzled seven-transmembrane domain proteins that are receptors for Wnt signalling proteins. Six1 may have effect on the expression of both ligands and the receptors, which may consequently influence the level of β-catenin. Six1 can also target *Lef1* (lymphoid enhancer

binding factor 1), *Tcf3* (transcription factor 3), and *Tcf4* (transcription factor 4), which encode transcription factors bound by active β-catenin. Without active β-catenin, LEF1 and TCF3/4 interact with co-repressor protein, Groucho family, and repress the expression of their target genes. Six1 may regulate the expression of Lef1 and Tcf3/4 to interfere with Wnt/β-catenin pathway. Mcaf1 (activating transcription factor 7 interacting protein, also known as *Atf7ip*), which had multiple roles in regulating Wnt signaling pathway (Huang & He, 2008), was also targeted by Six1. Six1 seems to have a role in embryonic patterning through modulating the Wnt signaling pathway. How it does so would be a very interesting issue to explore in the future.

6. The implication of Six1 in cancer

Over-expression of human SIX1 was found in numerous tumor samples, such as breast cancer and rhabdomyosarcoma (a soft-tissue cancer with a muscle-phenotype) (Christensen, Patrick, McCoy, & Ford, 2008; Reichenberger, Coletta, Schulte, Varella-Garcia, & Ford, 2005; Yu et al., 2006). Evidence supports that Six1 is involved in cancer initiation step: uncontrolled cell cycle and EMT that is characterized by increased cell mobility and is important for metastasis) (Coletta et al., 2010; Kumar, 2009). Suppression of Six1 expression had therapeutic effect for hepatocellular carcinoma through decreasing cell proliferation rate (Ng et al., 2009). In fact, the functional category "cell motility" was significantly enriched among Six1 target genes in this study (Table 3).

Six1 was reported to be involved in cell cycle regulation (Christensen et al., 2008; Kumar, 2009). Genes involved in cell cycle were also over-represented in Six1 targets. The pathways implicated in cancer development and progression are also targeted by Six1.

Indeed, some studies revealed that SIX1 had an impact on cell cycle regulation. Six1 can target *Myc*, *Ccnd1* and *Ccna1* to have influence on G1/S and G2/M check point, and thus regulates the proliferation (Coletta et al., 2004; X. Li et al., 2002; Yu et al., 2006). Wee1 [WEE 1 homolog 1 (S. pombe)], Rb1 (retinoblastoma 1), E2f7 (E2F transcription factor 7), and Cdkn1a [cyclin-dependent kinase inhibitor 1A (p21)] are the key regulators of cell cycle and they were all targeted by Six1 in this study. Six1 may have multiple roles in regulating cell cycle. What is the precise role of SIX1 in regulating cell cycle is worthy of exploring, given the fact that accumulating studies revealed the importance of SIX1 in cancer development.

7. Enriched DNA sequence motifs in Six1 targets

The MEF3 motif discovered in this study is unbiased from the analysis of genome-wide binding data. This newly-defined MEF3 motif can better represent the real binding DNA preference of Six1, probably Six4/2/5 as well, since it was shown before that these members of the Six family were able to bind similar sequences (Ohto et al., 1998; Spitz et al., 1998). The MEF3 motif has been studied extensively on the promoter regions of myogenic genes during past decades, which provided the classic yet rigid MEF3 motif (Hidaka et al., 1993; Parmacek et al., 1994; Salminen et al., 1996). Therefore, it is necessary and is informative to investigate a motif in a general context, such as the global binding data (Ji et al., 2008). In addition, in contrast to some studies using in vitro purified protein and naked DNA oligo sequences, ChIP-on-Chip explores the DNA-protein interaction in vivo, which reveals native binding events in an unbiased way. The MEF3 motif discovered in this study was specific and sensitive (Fig. 13).

Although Six2, Six4 and Six5 were also shown to be able to bind a similar MEF3 motif (Ohto et al., 1998; Spitz et al., 1998), they are likely to have their own specific sets of target genes since they have overlapping but distinct expression patterns (Klesert et al., 2000; Ohto et al., 1998; Ozaki et al., 2001). A slightly different nucleic acid preference may exist among these family members, which would affect the similarities/differences of their targets at the molecular level. Gathering more global binding data of other family members should provide a more comprehensive analysis about how they prefer different targets, if in fact they do.

Several other DNA motifs were discovered among Six1 binding sequences as well, such as the bZIP/AP1, which is bound by AP1 and other bZIP proteins. The direct binding of Six1 to the AP1 motif (TGAG/CTCA) was ruled out by EMSA: recombinant SIX1 cannot bind the probe containing AP1 element. However, the probe containing AP1 element can be shifted by the nuclear extract from C2C12 cells at T24. It brings up the probability that DNA regions containing AP1 motif are bound by SIX1 indirectly through a transcriptional complex residing in proximity of regulatory regions. Some evidence supports this assumption. c-JUN physically interacted with MYOD and DACH (Bengal et al., 1992; Wu et al., 2007). DACH1 and SIX1 physically interacted with each other and they were in the same network to regulate myogenesis (X. Li et al., 2003) . MYOD and SIX1 were co-recruited to 40% of Six1 target genes. Taken together, these observations suggest that these transcription factors could hypothetically bind DNA as a protein complex, contributing to the enrichment of the AP1 DNA motif among Six1 targets. It is possible that the proper spatiotemporal function of Six1 is modulated by its cooperation with different co-factors. The functional interaction between SIX1 and other transcription factors, whose DNA binding motifs were enriched in Six1 targets, will be an interesting theme to explore in the future.

Besides the potential functional interaction between Six1 and AP1, Six1 may directly regulate the expression of AP1 components. The regulatory regions of *Fos* (FBJ osteosarcoma oncogene, also known as *c-Fos*), *Jdp2* (Jun dimerization protein 2), *Jun* (Jun oncogene, also known as *c-Jun*), *Atf3* (activating transcription factor 3), *Atf6* (activating transcription factor 6), and *Fosl2* (fos-like antigen 2, also known as *Fra-2*) were bound by Six1. Although the direct regulation of Six1 on these genes is not validated yet, it provides a very interesting theme to further explore in the future by functional assays, such as promoter reporter assays.

8. The transcriptional activation by Six1

It is not clear yet how, as a transcription factor, Six1 may function to stimulate the transcriptional machinery to transcribe its target genes. Some previous studies supported that the co-operation with other TFs can recruit, in a synergistic fashion, the essential units for the transcription machinery. The MRFs, which were found often present in proximity to genomic regions bound by Six1, were able to recruit co-activators to initiate the transcription. MYOD can access to inactive DNA loci through chromatin remodeling by recruiting p300, which enrolls p300/CBP-associated factor (pCAF), and SWI/SNF complex, both of which are required for fully activation of MYOD (Dilworth, Seaver, Fishburn, Htet, & Tapscott, 2004). Eya3 can physically interact with Six1 in vivo. Eya3 was required for the recruitment of co-activator CBP and RNA polymerase II (X. Li et al., 2003). Although no evidence shows Six1 can recruit components of the basic transcriptional machinery directly, it is probable that Six1 functions through an indirect interaction mediated by co-regulators.

Conclusion

This study provides a substantial proof that Six1 is functional in adult skeletal muscle by regulating the expression of structural and functional muscle genes. Bioinformatics studies strongly suggest the existence of a functional cooperation between Six1 and MRFs.

A large number of putative Six1 targets were identified, which would extensively enlarge the known population of Six1 targets. With a great correlation with embryonic studies, this important resource will help us to study the function of Six1 both in vitro and in vivo.

It also brings up an intriguing question: what determines the spatial and temporal binding of a transcription factor on its target genes during development? In this study, a significant number of genes involved in the genesis of other organs were found in a committed muscle cell line. Why does it bind such genes in the "seemingly inappropriate" tissue (committed cells) and in the "seemingly inappropriate" time (the adult stage)? Is this unexpected binding phenomenon a general process? Hopefully, these questions could be answered in the near future.

Abbreviations

AEMSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
Angptl	angiopoietin 1
Ankrd1	ankyrin repeat domain 1 (cardiac muscle)
AP1	activator protein 1
APC	Anaphase-Promoting Complex
APC^{cdh1}	CDH1 dependent Anaphase-Promoting Complex
ATCC	American Type Culture Collection
Atf3	activating transcription factor 3
Atf6	activating transcription factor 6
Atp1a1	ATPase, Na+/K+ transporting, alpha 1 polypeptide
Atp1a2	ATPase, Na+/K+ transporting, alpha 2 polypeptide
Atp2a1	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1
Atp2b1	ATPase, Ca++ transporting, plasma membrane 1
BCA assay	bicinchoninic acid assay
Bdnf	brain derived neurotrophic factor
β-ME	β-mercaptoethanol
BOR	branchio-oto-renal
Cacna1s	calcium channel, voltage-dependent, L type, alpha 1S subunit
Cacna2d1	calcium channel, and voltage-dependent, alpha2/delta subunit 1
Cacnb1	calcium channel, voltage-dependent, beta 1 subunit
Cacnb2	calcium channel, voltage-dependent, beta 2 subunit
Cacng1	calcium channel, voltage-dependent, gamma subunit 1

Ccna1	cyclin A1
Ccnd1	Cyclin D1
Cdkn1a	cyclin-dependent kinase inhibitor 1A (p21)
ChIP	chromatin immunoprecipitation
Chrna1	cholinergic receptor, nicotinic, alpha polypeptide 1 (muscle)
Chrna9	cholinergic receptor, nicotinic, alpha polypeptide 9
Chrng	cholinergic receptor, nicotinic, gamma polypeptide
CK2	casein kinase II
dach	dachshund
Dach	dachshund (Drosophila)
Dach1	dachshund 1 (Drosophila)
Dach2	dachshund 2 (Drosophila)
Des	Desmin
DSS	Disuccinimidyl Suberate
DTT	Dithiothreitol
E2f7	E2F transcription factor 7
EMT	epithelial-mesenchymal transition
Eya1	eyes absent 1 homolog (Drosophila)
Eya2	eyes absent 2 homolog (Drosophila)
Eya3	eyes absent 3 homolog (Drosophila)
Eya4	eyes absent 4 homolog (Drosophila)
Ezr	Ezrin
FDR	false discovery rate

Fos (c-Fos)	FBJ osteosarcoma oncogene
Fosl2 (Fra-2)	fos-like antigen 2
Fzd4	Frizzled homolog 4 (Drosophila)
Fzd7	frizzled homolog 7 (Drosophila)
Gas2	growth arrest specific 2
Gli2	GLI-Kruppel family member GLI2
GO	Gene Ontology
GSH	Glutathione
HCM	hypertrophic cardiomyopathy
HDAC3	histone deacetylase 3
Hes6	hairy and enhancer of split 6 (Drosophila)
Hist2h4	histone cluster 2, H4
Hoxd10	homeobox D10
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Jdp2	Jun dimerization protein 2
Jun(c-Jun)	Jun oncogene
Kcne1l	potassium voltage-gated channel, Isk-related family, member 1-like
Lbx1	ladybird homeobox homolog 1 (Drosophila)
Ldb3	LIM domain binding 3
Lef1	lymphoid enhancer binding factor 1
MA	moving average
Mb	Myoblast
Mcaf1(Atf7ip)	activating transcription factor 7 interacting protein

Mef2c	myocyte enhancer factor 2C
MEF3 motif	Six1 binding DNA sequence motif
Mhc	Myosin heavy chain
Mos	Moloney sarcoma oncogene
MRFs	myogenic regulatory factors
Mt	Myotube
Myc	myelocytomatosis oncogene
Myf5	myogenic factor 5
Myf5 (Mrf4)	myogenic regulatory factor 4/myogenic factor 6
Myh7	myosin, heavy polypeptide 7, cardiac muscle, beta
Myod1	myogenic differentiation 1
Myod-CER	the core enhance region of myogenic differentiation 1
Myod-DRR	the distal regulatory region of myogenic differentiation 1
Myog	myogenin
N-CoR	nuclear receptor co-repressor
Neurod6	neurogenic differentiation 6
NP-40	NONIDET P-40
Pax3	paired box gene 3
Pax6	paired box gene 6
Pax7	paired box gene 7
Pax8	paired box gene 8
PBS	Phosphate buffered saline
Pbx1	pre B-cell leukemia transcription factor 1

pCAF	p300/CBP-associated factor
PMSF	phenylmethanesulfonylfluoride
q-ChIP	chromatin immunoprecipitation followed by quantitative polymerase chain reaction
q-PCR	quantitative polymerase chain reaction
Rb1	retinoblastoma 1
RMS	rhabdomyosarcoma
RT	Room temperature
SD	Six domain
SHD	Six-type homeobox nucleic acid recognition domain
Six1	sine oculis-related homeobox 1 homolog (Drosophila)
Six2	sine oculis-related homeobox 2 homolog (Drosophila)
Six3	sine oculis-related homeobox 3 homolog (Drosophila)
Six4	sine oculis-related homeobox 4 homolog (Drosophila)
Six5	sine oculis-related homeobox 5 homolog (Drosophila)
Six6	sine oculis-related homeobox 6 homolog (Drosophila)
Spry1	sprouty homolog 1 (Drosophila)
T24	C2C12, 24 hours after differentiation by serum withdrawal
Tcf3	transcription factor 3
Tcf4	transcription factor 4
TD	transactivation domain
Tnnc1	troponin C, cardiac/slow skeletal
Tnni1	troponin I, skeletal, slow 1

Tnni2	troponin I, skeletal, fast 2
Tnni3	troponin I, cardiac 3
Tnnt1	troponin T1, skeletal, slow
Tnnt2	troponin C2, fast
Tnnt3	troponin T3, skeletal, fast
TSS	Transcription Start Sites
Wee1	WEE 1 homolog 1 (S. pombe)
Wnt6	Wingless-related MMTV integration site 6

References

Abdelhak, S., Kalatzis, V., Heilig, R., Compain, S., Samson, D., Vincent, C., Levi-Acobas, F., Cruaud, C., Le Merrer, M., Mathieu, M., Konig, R., Vigneron, J., Weissenbach, J., Petit, C., & Weil, D. (1997). Clustering of mutations responsible for branchio-oto-renal (BOR) syndrome in the eyes absent homologous region (eyaHR) of EYA1. *Human Molecular Genetics, 6*(13), 2247-2255.

Aurade, F., Pinset, C., Chafey, P., Gros, F., & Montarras, D. (1994). Myf5, MyoD, myogenin and MRF4 myogenic derivatives of the embryonic mesenchymal cell line C3H10T1/2 exhibit the same adult muscle phenotype. *Differentiation; Research in Biological Diversity, 55*(3), 185-192.

Bajard, L., Relaix, F., Lagha, M., Rocancourt, D., Daubas, P., & Buckingham, M. E. (2006). A novel genetic hierarchy functions during hypaxial myogenesis: Pax3 directly activates Myf5 in muscle progenitor cells in the limb. *Genes & Development, 20*(17), 2450-2464. doi:10.1101/gad.382806

Basson, M. A., Akbulut, S., Watson-Johnson, J., Simon, R., Carroll, T. J., Shakya, R., Gross, I., Martin, G. R., Lufkin, T., McMahon, A. P., Wilson, P. D., Costantini, F. D., Mason, I. J., & Licht, J. D. (2005). Sprouty1 is a critical regulator of GDNF/RET-mediated kidney induction. *Developmental Cell, 8*(2), 229-239. doi:10.1016/j.devcel.2004.12.004

Basson, M. A., Watson-Johnson, J., Shakya, R., Akbulut, S., Hyink, D., Costantini, F. D., Wilson, P. D., Mason, I. J., & Licht, J. D. (2006). Branching morphogenesis of the ureteric epithelium during kidney development is coordinated by the opposing functions

of GDNF and Sprouty1. *Developmental Biology, 299*(2), 466-477. doi:10.1016/j.ydbio.2006.08.051

Bejerano, G., Siepel, A. C., Kent, W. J., & Haussler, D. (2005). Computational screening of conserved genomic DNA in search of functional noncoding elements. *Nature Methods, 2*(7), 535-545. doi:10.1038/nmeth0705-535

Bengal, E., Ransone, L., Scharfmann, R., Dwarki, V. J., Tapscott, S. J., Weintraub, H., & Verma, I. M. (1992). Functional antagonism between c-jun and MyoD proteins: A direct physical association. *Cell, 68*(3), 507-519.

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., Khalid, F., Zhang, W., Newburger, D., Jaeger, S. A., Morris, Q. D., Bulyk, M. L., & Hughes, T. R. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell, 133*(7), 1266-1276. doi:10.1016/j.cell.2008.05.024

Berkes, C. A., Bergstrom, D. A., Penn, B. H., Seaver, K. J., Knoepfler, P. S., & Tapscott, S. J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Molecular Cell, 14*(4), 465-477.

Blais, A., Monte, D., Pouliot, F., & Labrie, C. (2002). Regulation of the human cyclin-dependent kinase inhibitor p18INK4c by the transcription factors E2F1 and Sp1. *The Journal of Biological Chemistry, 277*(35), 31679-31693. doi:10.1074/jbc.M204554200

Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., & Dynlacht, B. D. (2005). An initial blueprint for myogenic differentiation. *Genes & Development, 19*(5), 553-569. doi:10.1101/gad.1281105

Blau, H. M., Chiu, C. P., & Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell, 32*(4), 1171-1180.

Bober, E., Franz, T., Arnold, H. H., Gruss, P., & Tremblay, P. (1994). Pax-3 is required for the development of limb muscles: A possible role for the migration of dermomyotomal muscle progenitor cells. *Development (Cambridge, England), 120*(3), 603-612.

Borello, U., Berarducci, B., Murphy, P., Bajard, L., Buffa, V., Piccolo, S., Buckingham, M., & Cossu, G. (2006). The Wnt/beta-catenin pathway regulates gli-mediated Myf5 expression during somitogenesis. *Development (Cambridge, England), 133*(18), 3723-3732. doi:10.1242/dev.02517

Borycki, A. G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C., & Emerson, C. P.,Jr. (1999). Sonic hedgehog controls epaxial muscle determination through Myf5 activation. *Development (Cambridge, England), 126*(18), 4053-4063.

Braun, T., Bober, E., Winter, B., Rosenthal, N., & Arnold, H. H. (1990). Myf-6, a new member of the human gene family of myogenic determination factors: Evidence for a gene cluster on chromosome 12. *The EMBO Journal, 9*(3), 821-831.

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

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

Brugmann, S. A., Pandur, P. D., Kenyon, K. L., Pignoni, F., & Moody, S. A. (2004). Six1 promotes a placodal fate within the lateral neurogenic ectoderm by functioning as both a transcriptional activator and repressor. *Development (Cambridge, England), 131*(23), 5871-5881. doi:10.1242/dev.01516

Bryson-Richardson, R. J., & Currie, P. D. (2008). The genetics of vertebrate myogenesis. *Nature Reviews.Genetics, 9*(8), 632-646. doi:10.1038/nrg2369

Buckingham, M. (2006). Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Current Opinion in Genetics & Development, 16*(5), 525-532. doi:10.1016/j.gde.2006.08.008

Buckingham, M., & Relaix, F. (2007). The role of pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annual Review of Cell and Developmental Biology, 23*, 645-673. doi:10.1146/annurev.cellbio.23.090506.123438

Cao, Y., Kumar, R. M., Penn, B. H., Berkes, C. A., Kooperberg, C., Boyer, L. A., Young, R. A., & Tapscott, S. J. (2006). Global and gene-specific analyses show distinct roles for myod and myog at a common set of promoters. *The EMBO Journal, 25*(3), 502-511. doi:10.1038/sj.emboj.7600958

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G. J., Parker, M. H., MacQuarrie, K. L., Davison, J., Morgan, M. T., Ruzzo, W. L., Gentleman, R. C., & Tapscott, S. J. (2010). Genome-wide MyoD binding in skeletal muscle cells: A potential for broad cellular reprogramming. *Developmental Cell, 18*(4), 662-674. doi:10.1016/j.devcel.2010.02.014

Cerletti, M., Shadrach, J. L., Jurga, S., Sherwood, R., & Wagers, A. J. (2008). Regulation and function of skeletal muscle stem cells. *Cold Spring Harbor Symposia on Quantitative Biology, 73*, 317-322. doi:10.1101/sqb.2008.73.054

Chang, E. H., Menezes, M., Meyer, N. C., Cucci, R. A., Vervoort, V. S., Schwartz, C. E., & Smith, R. J. (2004). Branchio-oto-renal syndrome: The mutation spectrum in EYA1 and its phenotypic consequences. *Human Mutation, 23*(6), 582-589. doi:10.1002/humu.20048

Chen, J. C., Ramachandran, R., & Goldhamer, D. J. (2002). Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. *Developmental Biology, 245*(1), 213-223. doi:10.1006/dbio.2002.0638

Cho, J. H., Lepine, M., Andrews, W., Parnavelas, J., & Cloutier, J. F. (2007). Requirement for slit-1 and robo-2 in zonal segregation of olfactory sensory neuron axons in the main olfactory bulb. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 27*(34), 9094-9104. doi:10.1523/JNEUROSCI.2217-07.2007

Christensen, K. L., Brennan, J. D., Aldridge, C. S., & Ford, H. L. (2007). Cell cycle regulation of the human Six1 homeoprotein is mediated by APC(Cdh1). *Oncogene, 26*(23), 3406-3414. doi:10.1038/sj.onc.1210122

Christensen, K. L., Patrick, A. N., McCoy, E. L., & Ford, H. L. (2008). The six family of homeobox genes in development and cancer. *Advances in Cancer Research, 101*, 93-126. doi:10.1016/S0065-230X(08)00405-3

Christophorou, N. A., Bailey, A. P., Hanson, S., & Streit, A. (2009). Activation of Six1 target genes is required for sensory placode formation. *Developmental Biology,* doi:10.1016/j.ydbio.2009.09.025

Clow, C., & Jasmin, B. J. (2010). Skeletal muscle-derived BDNF regulates satellite cell differentiation and muscle regeneration. *Molecular Biology of the Cell,* doi:10.1091/mbc.E10-02-0154

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., & Ford, H. L. (2004). The Six1 homeoprotein stimulates tumorigenesis by reactivation of cyclin A1. *Proceedings of the National Academy of Sciences of the United States of America, 101*(17), 6478-6483. doi:10.1073/pnas.0401139101

Coletta, R. D., Christensen, K. L., Micalizzi, D. S., Jedlicka, P., Varella-Garcia, M., & Ford, H. L. (2008). Six1 overexpression in mammary cells induces genomic instability and is sufficient for malignant transformation. *Cancer Research, 68*(7), 2204-2213. doi:10.1158/0008-5472.CAN-07-3141

Coletta, R. D., McCoy, E. L., Burns, V., Kawakami, K., McManaman, J. L., Wysolmerski, J. J., & Ford, H. L. (2010). Characterization of the Six1 homeobox gene in normal

mammary gland morphogenesis. *BMC Developmental Biology, 10*(1), 4. doi:10.1186/1471-213X-10-4

Collins, C. A., Olsen, I., Zammit, P. S., Heslop, L., Petrie, A., Partridge, T. A., & Morgan, J. E. (2005). Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. *Cell, 122*(2), 289-301. doi:10.1016/j.cell.2005.05.010

Cossins, J., Vernon, A. E., Zhang, Y., Philpott, A., & Jones, P. H. (2002). Hes6 regulates myogenic differentiation. *Development (Cambridge, England), 129*(9), 2195-2207.

Daston, G., Lamar, E., Olivier, M., & Goulding, M. (1996). Pax-3 is necessary for migration but not differentiation of limb muscle precursors in the mouse. *Development (Cambridge, England), 122*(3), 1017-1027.

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

Dilworth, F. J., Seaver, K. J., Fishburn, A. L., Htet, S. L., & Tapscott, S. J. (2004). In vitro transcription system delineates the distinct roles of the coactivators pCAF and p300 during MyoD/E47-dependent transactivation. *Proceedings of the National Academy of Sciences of the United States of America, 101*(32), 11593-11598. doi:10.1073/pnas.0404192101

Ernfors, P., Lee, K. F., & Jaenisch, R. (1994). Mice lacking brain-derived neurotrophic factor develop with sensory deficits. *Nature, 368*(6467), 147-150. doi:10.1038/368147a0

Ford, H. L., Landesman-Bollag, E., Dacwag, C. S., Stukenberg, P. T., Pardee, A. B., & Seldin, D. C. (2000). Cell cycle-regulated phosphorylation of the human SIX1 homeodomain protein. *The Journal of Biological Chemistry, 275*(29), 22245-22254. doi:10.1074/jbc.M002446200

Frey, N., & Olson, E. N. (2002). Calsarcin-3, a novel skeletal muscle-specific member of the calsarcin family, interacts with multiple Z-disc proteins. *The Journal of Biological Chemistry, 277*(16), 13998-14004. doi:10.1074/jbc.M200712200

Gale, N. W., & Yancopoulos, G. D. (1999). Growth factors acting via endothelial cell-specific receptor tyrosine kinases: VEGFs, angiopoietins, and ephrins in vascular development. *Genes & Development, 13*(9), 1055-1066.

Gebauer, F., & Richter, J. D. (1997). Synthesis and function of mos: The control switch of vertebrate oocyte meiosis. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 19*(1), 23-28. doi:10.1002/bies.950190106

Gerber, A. N., Klesert, T. R., Bergstrom, D. A., & Tapscott, S. J. (1997). Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: A mechanism for lineage determination in myogenesis. *Genes & Development, 11*(4), 436-450.

Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M., & Maire, P. (2007). Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. *Proceedings of the National Academy of Sciences of the United States of America, 104*(27), 11310-11315. doi:10.1073/pnas.0611299104

Grifone, R., Demignon, J., Giordani, J., Niro, C., Souil, E., Bertin, F., Laclef, C., Xu, P. X., & Maire, P. (2007). Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. *Developmental Biology, 302*(2), 602-616. doi:10.1016/j.ydbio.2006.08.059

Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M. J., Hamard, G., & Maire, P. (2005). Six1 and Six4 homeoproteins are required for Pax3 and mrf expression during myogenesis in the mouse embryo. *Development (Cambridge, England), 132*(9), 2235-2249. doi:10.1242/dev.01773

Grifone, R., Laclef, C., Spitz, F., Lopez, S., Demignon, J., Guidotti, J. E., Kawakami, K., Xu, P. X., Kelly, R., Petrof, B. J., Daegelen, D., Concordet, J. P., & Maire, P. (2004). Six1 and Eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype. *Molecular and Cellular Biology, 24*(14), 6253-6267. doi:10.1128/MCB.24.14.6253-6267.2004

Gros, J., Manceau, M., Thome, V., & Marcelle, C. (2005). A common somitic origin for embryonic muscle progenitors and satellite cells. *Nature, 435*(7044), 954-958. doi:10.1038/nature03572

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

Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N., & Klein, W. H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature, 364*(6437), 501-506. doi:10.1038/364501a0

Hidaka, K., Yamamoto, I., Arai, Y., & Mukai, T. (1993). The MEF-3 motif is required for MEF-2-mediated skeletal muscle-specific induction of the rat aldolase A gene. *Molecular and Cellular Biology, 13*(10), 6469-6478.

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols, 4*(1), 44-57. doi:10.1038/nprot.2008.211

Huang, H., & He, X. (2008). Wnt/beta-catenin signaling: New (and old) players and new insights. *Current Opinion in Cell Biology, 20*(2), 119-125. doi:10.1016/j.ceb.2008.01.009

Ikeda, K., Ookawara, S., Sato, S., Ando, Z., Kageyama, R., & Kawakami, K. (2007). Six1 is essential for early neurogenesis in the development of olfactory epithelium. *Developmental Biology, 311*(1), 53-68. doi:10.1016/j.ydbio.2007.08.020

Ikeda, K., Watanabe, Y., Ohto, H., & Kawakami, K. (2002). Molecular interaction and synergistic activation of a promoter by six, eya, and dach proteins mediated through CREB binding protein. *Molecular and Cellular Biology, 22*(19), 6759-6766.

Jhas, S., Ciura, S., Belanger-Jasmin, S., Dong, Z., Llamosas, E., Theriault, F. M., Joachim, K., Tang, Y., Liu, L., Liu, J., & Stifani, S. (2006). Hes6 inhibits astrocyte differentiation and promotes neurogenesis through different mechanisms. *The Journal of Neuroscience :*

The Official Journal of the Society for Neuroscience, 26(43), 11061-11071. doi:10.1523/JNEUROSCI.1358-06.2006

Ji, H., Jiang, H., Ma, W., Johnson, D. S., Myers, R. M., & Wong, W. H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nature Biotechnology, 26*(11), 1293-1300. doi:10.1038/nbt.1505

Ji, H., Vokes, S. A., & Wong, W. H. (2006). A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. *Nucleic Acids Research, 34*(21), e146. doi:10.1093/nar/gkl803

Johnson, D. S., Li, W., Gordon, D. B., Bhattacharjee, A., Curry, B., Ghosh, J., Brizuela, L., Carroll, J. S., Brown, M., Flicek, P., Koch, C. M., Dunham, I., Bieda, M., Xu, X., Farnham, P. J., Kapranov, P., Nix, D. A., Gingeras, T. R., Zhang, X., Holster, H., Jiang, N., Green, R. D., Song, J. S., McCuine, S. A., Anton, E., Nguyen, L., Trinklein, N. D., Ye, Z., Ching, K., Hawkins, D., Ren, B., Scacheri, P. C., Rozowsky, J., Karpikov, A., Euskirchen, G., Weissman, S., Gerstein, M., Snyder, M., Yang, A., Moqtaderi, Z., Hirsch, H., Shulha, H. P., Fu, Y., Weng, Z., Struhl, K., Myers, R. M., Lieb, J. D., & Liu, X. S. (2008). Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Research, 18*(3), 393-403. doi:10.1101/gr.7080508

Jones, K. R., Farinas, I., Backus, C., & Reichardt, L. F. (1994). Targeted disruption of the BDNF gene perturbs brain and sensory neuron development but not motor neuron development. *Cell, 76*(6), 989-999.

Kablar, B., Asakura, A., Krastel, K., Ying, C., May, L. L., Goldhamer, D. J., & Rudnicki, M. A. (1998). MyoD and myf-5 define the specification of musculature of distinct embryonic origin. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire, 76*(6), 1079-1091.

Karamboulas, C., Dakubo, G. D., Liu, J., De Repentigny, Y., Yutzey, K., Wallace, V. A., Kothary, R., & Skerjanc, I. S. (2006). Disruption of MEF2 activity in cardiomyoblasts inhibits cardiomyogenesis. *Journal of Cell Science, 119*(Pt 20), 4315-4321. doi:10.1242/jcs.03186

Kassar-Duchossoy, L., Gayraud-Morel, B., Gomes, D., Rocancourt, D., Buckingham, M., Shinin, V., & Tajbakhsh, S. (2004). Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature, 431*(7007), 466-471. doi:10.1038/nature02876

Kawakami, K., Sato, S., Ozaki, H., & Ikeda, K. (2000). Six family genes--structure and function as transcription factors and their roles in development. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 22*(7), 616-626. doi:2-R

Kim, W. Y., Fritzsch, B., Serls, A., Bakel, L. A., Huang, E. J., Reichardt, L. F., Barth, D. S., & Lee, J. E. (2001). NeuroD-null mice are deaf due to a severe loss of the inner ear sensory neurons during development. *Development (Cambridge, England), 128*(3), 417-426.

Klesert, T. R., Cho, D. H., Clark, J. I., Maylie, J., Adelman, J., Snider, L., Yuen, E. C., Soriano, P., & Tapscott, S. J. (2000). Mice deficient in Six5 develop cataracts:

Implications for myotonic dystrophy. *Nature Genetics, 25*(1), 105-109. doi:10.1038/75490

Kobayashi, H., Kawakami, K., Asashima, M., & Nishinakamura, R. (2007). Six1 and Six4 are essential for gdnf expression in the metanephric mesenchyme and ureteric bud formation, while Six1 deficiency alone causes mesonephric-tubule defects. *Mechanisms of Development, 124*(4), 290-303. doi:10.1016/j.mod.2007.01.002

Kochhar, A., Fischer, S. M., Kimberling, W. J., & Smith, R. J. (2007). Branchio-oto-renal syndrome. *American Journal of Medical Genetics.Part A, 143A*(14), 1671-1678. doi:10.1002/ajmg.a.31561

Kruger, M., Mennerich, D., Fees, S., Schafer, R., Mundlos, S., & Braun, T. (2001). Sonic hedgehog is a survival factor for hypaxial muscles during mouse development. *Development (Cambridge, England), 128*(5), 743-752.

Kuang, S., Kuroda, K., Le Grand, F., & Rudnicki, M. A. (2007). Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell, 129*(5), 999-1010. doi:10.1016/j.cell.2007.03.044

Kucharczuk, K. L., Love, C. M., Dougherty, N. M., & Goldhamer, D. J. (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 (Cambridge, England), 126*(9), 1957-1965.

Kumar, J. P. (2009). The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. *Cellular and Molecular Life Sciences : CMLS, 66*(4), 565-583. doi:10.1007/s00018-008-8335-4

Laclef, C., Hamard, G., Demignon, J., Souil, E., Houbron, C., & Maire, P. (2003). Altered myogenesis in Six1-deficient mice. *Development (Cambridge, England), 130*(10), 2239-2252.

Laclef, C., Souil, E., Demignon, J., & Maire, P. (2003). Thymus, kidney and craniofacial abnormalities in six 1 deficient mice. *Mechanisms of Development, 120*(6), 669-679.

Lang, D., Powell, S. K., Plummer, R. S., Young, K. P., & Ruggeri, B. A. (2007). PAX genes: Roles in development, pathophysiology, and cancer. *Biochemical Pharmacology, 73*(1), 1-14. doi:10.1016/j.bcp.2006.06.024

Laure, L., Suel, L., Roudaut, C., Bourg, N., Ouali, A., Bartoli, M., Richard, I., & Daniele, N. (2009). Cardiac ankyrin repeat protein is a marker of skeletal muscle pathological remodelling. *The FEBS Journal, 276*(3), 669-684. doi:10.1111/j.1742-4658.2008.06814.x

Le Grand, F., & Rudnicki, M. A. (2007). Skeletal muscle satellite cells and adult myogenesis. *Current Opinion in Cell Biology, 19*(6), 628-633. doi:10.1016/j.ceb.2007.09.012

Lepper, C., Conway, S. J., & Fan, C. M. (2009). Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements. *Nature, 460*(7255), 627-631. doi:10.1038/nature08209

Li, H., & Capetanaki, Y. (1993). Regulation of the mouse desmin gene: Transactivated by MyoD, myogenin, MRF4 and Myf5. *Nucleic Acids Research, 21*(2), 335-343.

Li, H., Radford, J. C., Ragusa, M. J., Shea, K. L., McKercher, S. R., Zaremba, J. D., Soussou, W., Nie, Z., Kang, Y. J., Nakanishi, N., Okamoto, S., Roberts, A. J., Schwarz, J. J., & Lipton, S. A. (2008). Transcription factor MEF2C influences neural stem/progenitor cell differentiation and maturation in vivo. *Proceedings of the National Academy of Sciences of the United States of America, 105*(27), 9397-9402. doi:10.1073/pnas.0802876105

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., & Rosenfeld, M. G. (2003). Eya protein phosphatase activity regulates Six1-dach-eya transcriptional effects in mammalian organogenesis. *Nature, 426*(6964), 247-254. doi:10.1038/nature02083

Li, X., Perissi, V., Liu, F., Rose, D. W., & Rosenfeld, M. G. (2002). Tissue-specific regulation of retinal and pituitary precursor cell proliferation. *Science (New York, N.Y.), 297*(5584), 1180-1183. doi:10.1126/science.1073263

Linhart, C., Halperin, Y., & Shamir, R. (2008). Transcription factor and microRNA motif discovery: The amadeus platform and a compendium of metazoan target sets. *Genome Research, 18*(7), 1180-1189. doi:10.1101/gr.076117.108

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 Research,* doi:10.1093/nar/gkq585

Lopez-Rios, J., Tessmar, K., Loosli, F., Wittbrodt, J., & Bovolenta, P. (2003). Six3 and Six6 activity is modulated by members of the groucho family. *Development (Cambridge, England), 130*(1), 185-195.

Marcil, A., Dumontier, E., Chamberland, M., Camper, S. A., & Drouin, J. (2003). Pitx1 and Pitx2 are required for development of hindlimb buds. *Development (Cambridge, England), 130*(1), 45-55.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V., Kloos, D. U., Land, S., Lewicki-Potapov, B., Michael, H., Munch, R., Reuter, I., Rotert, S., Saxel, H., Scheer, M., Thiele, S., & Wingender, E. (2003). TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Research, 31*(1), 374-378.

Maves, L., Waskiewicz, A. J., Paul, B., Cao, Y., Tyler, A., Moens, C. B., & Tapscott, S. J. (2007). Pbx homeodomain proteins direct myod activity to promote fast-muscle differentiation. *Development (Cambridge, England), 134*(18), 3371-3382. doi:10.1242/dev.003905

Michalk, A., Stricker, S., Becker, J., Rupps, R., Pantzar, T., Miertus, J., Botta, G., Naretto, V. G., Janetzki, C., Yaqoob, N., Ott, C. E., Seelow, D., Wieczorek, D., Fiebig, B., Wirth, B., Hoopmann, M., Walther, M., Korber, F., Blankenburg, M., Mundlos, S., Heller, R., & Hoffmann, K. (2008). Acetylcholine receptor pathway mutations explain various fetal akinesia deformation sequence disorders. *American Journal of Human Genetics, 82*(2), 464-476. doi:10.1016/j.ajhg.2007.11.006

Montarras, D., Chelly, J., Bober, E., Arnold, H., Ott, M. O., Gros, F., & Pinset, C. (1991). Developmental patterns in the expression of Myf5, MyoD, myogenin, and MRF4 during myogenesis. *The New Biologist, 3*(6), 592-600.

Montarras, D., Morgan, J., Collins, C., Relaix, F., Zaffran, S., Cumano, A., Partridge, T., & Buckingham, M. (2005). Direct isolation of satellite cells for skeletal muscle regeneration. *Science (New York, N.Y.), 309*(5743), 2064-2067. doi:10.1126/science.1114758

Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nonaka, I., & Nabeshima, Y. (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature, 364*(6437), 532-535. doi:10.1038/364532a0

Naidu, P. S., Ludolph, D. C., To, R. Q., Hinterberger, T. J., & Konieczny, S. F. (1995). Myogenin and MEF2 function synergistically to activate the MRF4 promoter during myogenesis. *Molecular and Cellular Biology, 15*(5), 2707-2718.

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., & Man, K. (2009). Suppression of tumorigenesis and metastasis of hepatocellular carcinoma by shRNA interference targeting on homeoprotein Six1. *International Journal of Cancer.Journal International Du Cancer,* doi:10.1002/ijc.25105

Niro, C., Demignon, J., Vincent, S., Liu, Y., Giordani, J., Sgarioto, N., Favier, M., Guillet-Deniau, I., Blais, A., & Maire, P. (2009). Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. *Developmental Biology,* doi:10.1016/j.ydbio.2009.11.031

Niro, C., Demignon, J., Vincent, S., Liu, Y., Giordani, J., Sgarioto, N., Favier, M., Guillet-Deniau, I., Blais, A., & Maire, P. (2010). Six1 and Six4 gene expression is necessary to activate the fast-type muscle gene program in the mouse primary myotome. *Developmental Biology, 338*(2), 168-182. doi:10.1016/j.ydbio.2009.11.031

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., & Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell, 133*(7), 1277-1289. doi:10.1016/j.cell.2008.05.023

Ohto, H., Kamada, S., Tago, K., Tominaga, S. I., Ozaki, H., Sato, S., & Kawakami, K. (1999). Cooperation of six and eya in activation of their target genes through nuclear translocation of eya. *Molecular and Cellular Biology, 19*(10), 6815-6824.

Ohto, H., Takizawa, T., Saito, T., Kobayashi, M., Ikeda, K., & Kawakami, K. (1998). Tissue and developmental distribution of six family gene products. *The International Journal of Developmental Biology, 42*(2), 141-148.

Olguin, H. C., & Olwin, B. B. (2004). Pax-7 up-regulation inhibits myogenesis and cell cycle progression in satellite cells: A potential mechanism for self-renewal. *Developmental Biology, 275*(2), 375-388. doi:10.1016/j.ydbio.2004.08.015

Olguin, H. C., Yang, Z., Tapscott, S. J., & Olwin, B. B. (2007). Reciprocal inhibition between Pax7 and muscle regulatory factors modulates myogenic cell fate determination. *The Journal of Cell Biology, 177*(5), 769-779. doi:10.1083/jcb.200608122

Oliver, G., Wehr, R., Jenkins, N. A., Copeland, N. G., Cheyette, B. N., Hartenstein, V., Zipursky, S. L., & Gruss, P. (1995). Homeobox genes and connective tissue patterning. *Development (Cambridge, England), 121*(3), 693-705.

Ozaki, H., Watanabe, Y., Takahashi, K., Kitamura, K., Tanaka, A., Urase, K., Momoi, T., Sudo, K., Sakagami, J., Asano, M., Iwakura, Y., & Kawakami, K. (2001). Six4, a putative myogenin gene regulator, is not essential for mouse embryonal development. *Molecular and Cellular Biology, 21*(10), 3343-3350. doi:10.1128/MCB.21.10.3343-3350.2001

Parmacek, M. S., Ip, H. S., Jung, F., Shen, T., Martin, J. F., Vora, A. J., Olson, E. N., & Leiden, J. M. (1994). A novel myogenic regulatory circuit controls slow/cardiac troponin C gene transcription in skeletal muscle. *Molecular and Cellular Biology, 14*(3), 1870-1885.

Pennesi, M. E., Cho, J. H., Yang, Z., Wu, S. H., Zhang, J., Wu, S. M., & Tsai, M. J. (2003). BETA2/NeuroD1 null mice: A new model for transcription factor-dependent photoreceptor degeneration. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 23*(2), 453-461.

Petersen, C. P., & Reddien, P. W. (2009). Wnt signaling and the polarity of the primary body axis. *Cell, 139*(6), 1056-1068. doi:10.1016/j.cell.2009.11.035

Pichel, J. G., Shen, L., Sheng, H. Z., Granholm, A. C., Drago, J., Grinberg, A., Lee, E. J., Huang, S. P., Saarma, M., Hoffer, B. J., Sariola, H., & Westphal, H. (1996). Defects in enteric innervation and kidney development in mice lacking GDNF. *Nature, 382*(6586), 73-76. doi:10.1038/382073a0

Porter, A. G., Ng, P., & Janicke, R. U. (1997). Death substrates come alive. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 19*(6), 501-507. doi:10.1002/bies.950190609

Potthoff, M. J., Arnold, M. A., McAnally, J., Richardson, J. A., Bassel-Duby, R., & Olson, E. N. (2007). Regulation of skeletal muscle sarcomere integrity and postnatal muscle function by Mef2c. *Molecular and Cellular Biology, 27*(23), 8143-8151. doi:10.1128/MCB.01187-07

Rawls, A., Morris, J. H., Rudnicki, M., Braun, T., Arnold, H. H., Klein, W. H., & Olson, E. N. (1995). Myogenin's functions do not overlap with those of MyoD or myf-5 during mouse embryogenesis. *Developmental Biology, 172*(1), 37-50. doi:10.1006/dbio.1995.0004

Reichenberger, K. J., Coletta, R. D., Schulte, A. P., Varella-Garcia, M., & Ford, H. L. (2005). Gene amplification is a mechanism of Six1 overexpression in breast cancer. *Cancer Research, 65*(7), 2668-2675. doi:10.1158/0008-5472.CAN-04-4286

Relaix, F., Rocancourt, D., Mansouri, A., & Buckingham, M. (2004). Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes & Development, 18*(9), 1088-1105. doi:10.1101/gad.301004

Relaix, F., Rocancourt, D., Mansouri, A., & Buckingham, M. (2005). A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature, 435*(7044), 948-953. doi:10.1038/nature03594

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., & Young, R. A. (2000). Genome-wide location and function of DNA binding proteins. *Science (New York, N.Y.), 290*(5500), 2306-2309. doi:10.1126/science.290.5500.2306

Ridgeway, A. G., & Skerjanc, I. S. (2001). Pax3 is essential for skeletal myogenesis and the expression of Six1 and Eya2. *The Journal of Biological Chemistry, 276*(22), 19033-19039. doi:10.1074/jbc.M011491200

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

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

Ruf, R. G., Xu, P. X., Silvius, D., Otto, E. A., Beekmann, F., Muerb, U. T., Kumar, S., Neuhaus, T. J., Kemper, M. J., Raymond, R. M.,Jr, Brophy, P. D., Berkman, J., Gattas, M., Hyland, V., Ruf, E. M., Schwartz, C., Chang, E. H., Smith, R. J., Stratakis, C. A., Weil, D., Petit, C., & Hildebrandt, F. (2004). SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America, 101*(21), 8090-8095. doi:10.1073/pnas.0308475101

Salminen, M., Lopez, S., Maire, P., Kahn, A., & 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. *Molecular and Cellular Biology, 16*(1), 76-85.

Sandmann, T., Jensen, L. J., Jakobsen, J. S., Karzynski, M. M., Eichenlaub, M. P., Bork, P., & Furlong, E. E. (2006). A temporal map of transcription factor activity: Mef2 directly regulates target genes at all stages of muscle development. *Developmental Cell, 10*(6), 797-807. doi:10.1016/j.devcel.2006.04.009

Schnabel, C. A., Godin, R. E., & Cleary, M. L. (2003). Pbx1 regulates nephrogenesis and ureteric branching in the developing kidney. *Developmental Biology, 254*(2), 262-276.

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

Shih, H. P., Gross, M. K., & Kioussi, C. (2007). Expression pattern of the homeodomain transcription factor Pitx2 during muscle development. *Gene Expression Patterns : GEP, 7*(4), 441-451. doi:10.1016/j.modgep.2006.11.004

Spitz, F., Demignon, J., Porteu, A., Kahn, A., Concordet, J. P., Daegelen, D., & Maire, P. (1998). Expression of myogenin during embryogenesis is controlled by Six/sine oculis homeoproteins through a conserved MEF3 binding site. *Proceedings of the National Academy of Sciences of the United States of America, 95*(24), 14220-14225.

Stehling-Sun, S., Dade, J., Nutt, S. L., DeKoter, R. P., & Camargo, F. D. (2009). Regulation of lymphoid versus myeloid fate 'choice' by the transcription factor Mef2c. *Nature Immunology, 10*(3), 289-296. doi:10.1038/ni.1694

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

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

Tapscott, S. J. (2005). The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development (Cambridge, England), 132*(12), 2685-2695. doi:10.1242/dev.01874

Thayer, M. J., Tapscott, S. J., Davis, R. L., Wright, W. E., Lassar, A. B., & Weintraub, H. (1989). Positive autoregulation of the myogenic determination gene MyoD1. *Cell, 58*(2), 241-248.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature, 457*(7231), 854-858. doi:10.1038/nature07730

Weasner, B., Salzer, C., & Kumar, J. P. (2007). Sine oculis, a member of the SIX family of transcription factors, directs eye formation. *Developmental Biology, 303*(2), 756-771. doi:10.1016/j.ydbio.2006.10.040

Weintraub, H., Davis, R., Lockshon, D., & Lassar, A. (1990). MyoD binds cooperatively to two sites in a target enhancer sequence: Occupancy of two sites is required for activation. *Proceedings of the National Academy of Sciences of the United States of America, 87*(15), 5623-5627.

Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proceedings of the National Academy of Sciences of the United States of America, 86*(14), 5434-5438.

Williams, B. A., & Ordahl, C. P. (1994). Pax-3 expression in segmental mesoderm marks early stages in myogenic cell specification. *Development (Cambridge, England), 120*(4), 785-796.

Witte, F., Dokas, J., Neuendorf, F., Mundlos, S., & Stricker, S. (2009). Comprehensive expression analysis of all wnt genes and their major secreted antagonists during mouse limb development and cartilage differentiation. *Gene Expression Patterns : GEP, 9*(4), 215-223. doi:10.1016/j.gep.2008.12.009

Wright, W. E., Sassoon, D. A., & Lin, V. K. (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell, 56*(4), 607-617.

Wu, K., Li, A., Rao, M., Liu, M., Dailey, V., Yang, Y., Di Vizio, D., Wang, C., Lisanti, M. P., Sauter, G., Russell, R. G., Cvekl, A., & Pestell, R. G. (2006). DACH1 is a cell fate determination factor that inhibits cyclin D1 and breast tumor growth. *Molecular and Cellular Biology, 26*(19), 7116-7129. doi:10.1128/MCB.00268-06

Wu, K., Liu, M., Li, A., Donninger, H., Rao, M., Jiao, X., Lisanti, M. P., Cvekl, A., Birrer, M., & Pestell, R. G. (2007). Cell fate determination factor DACH1 inhibits c-jun-induced contact-independent growth. *Molecular Biology of the Cell, 18*(3), 755-767. doi:10.1091/mbc.E06-09-0793

Xu, P. X., Zheng, W., Huang, L., Maire, P., Laclef, C., & Silvius, D. (2003). Six1 is required for the early organogenesis of mammalian kidney. *Development (Cambridge, England), 130*(14), 3085-3094.

Yaffe, D., & Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature, 270*(5639), 725-727.

Yu, Y., Davicioni, E., Triche, T. J., & Merlino, G. (2006). The homeoprotein six1 transcriptionally activates multiple protumorigenic genes but requires ezrin to promote metastasis. *Cancer Research, 66*(4), 1982-1989. doi:10.1158/0008-5472.CAN-05-2360

Yu, Y., Khan, J., Khanna, C., Helman, L., Meltzer, P. S., & Merlino, G. (2004). Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein six-1 as key metastatic regulators. *Nature Medicine, 10*(2), 175-181. doi:10.1038/nm966

Zhang, H., & Stavnezer, E. (2009). Ski regulates muscle terminal differentiation by transcriptional activation of myog in a complex with Six1 and Eya3. *The Journal of Biological Chemistry, 284*(5), 2867-2879. doi:10.1074/jbc.M807526200

Zhu, C. C., Dyer, M. A., Uchikawa, M., Kondoh, H., Lagutin, O. V., & Oliver, G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the groucho-related family of co-repressors. *Development (Cambridge, England), 129*(12), 2835-2849.

Zou, D., Silvius, D., Fritzsch, B., & Xu, P. X. (2004). Eya1 and Six1 are essential for early steps of sensory neurogenesis in mammalian cranial placodes. *Development (Cambridge, England), 131*(22), 5561-5572. doi:10.1242/dev.01437