THE ROLE OF TWIST2 IN PHYSIOLOGIC AND PATHOLOGIC MYOGENSIS

APPROVED BY SUPERVISORY COMMITTEE

First Name Last Name, credentials

Eric Olson, Ph.D.

James Amatruda, M.D., Ph.D.

Helen Hobbs, M.D.

Sean Morrison, Ph.D.

DEDICATION

I would like to dedicate this thesis to my friends, family, and mentors for providing me guidance and support throughout the years.

ACKNOWLEDGEMENTS

The work presented in this dissertation would not be possible without the help of a large cast of supporters. First and foremost, I wish to express my sincerest gratitude towards my mentor, Eric Olson, for taking a chance on me and welcoming me into his lab. Along the course of my training path, Eric has been a scientific inspiration in many regards. His curiosity and love for science is highly contagious and signifies that science is not just a career, it is a passion. This is best shown by his remarkable ability to convey his scientific work in a meaningful, clear, and inspiring manner. This skill is often pushed to the sidelines in the academic world, but being a strong scientific communicator is vital for scientific success and is one of the traits I admire most. In addition to serving as an inspiration, Eric has also enabled me to freely pursue my interests within the lab. I was able to take an existing project, explore multiple avenues of future direction, and come up with new hypotheses for my own thesis work. During this time, Eric has provided me with the motivation, intellectual support, and resources to grow and succeed as a scientist. Without his mentorship, I would not be where I am today. He will continue to serve as my scientific role model for the rest of my career.

I would also like to thank my mentor Rhonda Bassel-Duby for all of the help she has provided me during my training. On a daily basis, Rhonda helps to facilitate collaborations, resource attainment, and lab well-being. Additionally, Rhonda provided enormous amounts of help in writing and editing grants, papers, presentations. When I first joined the lab, it was

Rhonda who helped guide me on making and giving a clear presentation. This skill will be one that I value and carry with me for the rest of my career. Not only has she been vital for my success in the lab, she also played an instrumental role in my acceptance to the MSTP program. Unbeknownst to me, Rhonda had vouched for me during my application process and coincidentally, I ended up joining the Olson lab after arriving at UT Southwestern. Perhaps it was fate that I ended up completing my dissertation here.

My growth and development as a basic scientist would not be possible without my day-to-day mentor Ning Liu. Ning has been available to me every day for guidance and advice on scientific technique and experiments. When I initially joined the lab, I was able to jump onto her project exploring the biology of a novel Twist2-dependent muscle progenitor. As time progressed, she gave me more and more independent control of the direction of this project. Her trust gave me the confidence to take control of the scientific reigns of my thesis. Ning has also helped me enormously in developing my scientific writing and presentation skills. Most importantly the knowledge she imparted to me will enable me to succeed as an independent investigator in the future. I will miss all of the time we have spent together brainstorming ideas.

In addition to my mentors, I also wish to thank the other members of the Olson lab. The Olson lab is really just a large tight-knit family who are there to help and support each other during our growth as scientists. From all the happy hours and parties we have attended, I have truly come to enjoy my time within the lab because I get to see all of my friends on a daily basis. In particular I also want to thank the graduate students of the Olson lab who have

been there and looked out for me: Bercin, Yi-Li, Andres, Yu, Zhaoning, Tony, Akansha, and Dileep. You have been the best friends and lab-mates anyone could have.

I also want to thank the members of my committee: Dr. James Amatruda, Dr. Helen Hobbs, and Dr. Sean Morrison. Your scientific guidance and advice has been invaluable to me along my training path as a physician scientist. I remember distinctly during my first committee meeting that all three of you told me to avoid spreading myself thin and stick to the Twist project. This advice has been incredibly important to me and is the reason why I am able to finish and present this complete body of work for my dissertation.

Lastly, I would like to thank my family for all of the support they have given me over the years. To my mom, dad, and brother, thank you for being there and cultivating my curiosity as I grew up. Thank you for also believing in me and trusting my ability to carve my own career path. Thank you for fostering and providing me the resources to achieve my dreams. To my fiancée (and future wife!) Simone, thank you for all of the support you give to me on a daily basis. Science can often times be stressful and emotionally draining when experiments or hypotheses fail. Thank you for being there when I come home to help cheer me up and keep me emotionally grounded. The long and winding road of the physician-scientist training path becomes less daunting when you have your life-long partner and best friend by your side to illuminate the way.

THE ROLE OF TWIST2 IN MYOGENESIS

by

STEPHEN LI

DISSERTATION / THESIS

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2019

THE ROLE OF TWIST2 IN PHYSIOLOGIC AND PATHOLOGIC MYOGENESIS

Publication No._____

Stephen Li, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2019

Supervising Professor: Eric N. Olson, Ph.D.

Skeletal muscle is a highly regenerative tissue required for vertebrate life. It composes a significant portion of body mass and enables the physiologic processes of movement and breathing. Given its importance, skeletal muscle is also highly susceptible to aging and diseases such as cancer. Aging-related muscle atrophy (sarcopenia) is a process that affects nearly every person, contributing to debilitations and reductions in quality of life. As people age, fast-twitch muscle fibers selectively atrophy resulting in weakness. Normally, muscle regenerates through a population of stem cells called satellite cells, which differentiate and non-selectively fuse to existing myofibers in order to repair the damaged muscle tissue. However, several recent studies have suggested that the contribution of

satellite cells to muscle during homeostasis and aging is minimal. Thus, it's possible that loss of an alternative muscle precursor that fuses specifically with fast-twitch fibers may be a mechanism by which aging-related muscle atrophy occurs. Through the technique of lineage-tracing (fate-mapping) of the transcription factor Twist2, we have identified a novel muscle progenitor that fuses specifically to type IIb/x (fast-twitch, glycolytic) muscle fibers during both aging and homeostasis. Additionally, loss of Twist2+ cells result in specific atrophy of fast-twitch myofibers. I show that Twist2 plays a role in regulating fiber-type specificity through upregulation of the membrane receptor Nrp1. Additionally, the Nrp1 chemo-repulsive ligand, Sema3a, is expressed by both type I and IIa fibers. This Sema3a-Nrp1 signaling mechanisms prevents Twist2+ cells from fusing to type I and IIa fibers, and exogenous overexpression of Sema3a in type IIb fibers impairs the contribution of Twist2+ cells to these fibers. I also found that Twist2 is highly amplified in rhabdomyosarcoma (RMS), a pediatric soft tissue sarcoma expressing hallmarks of the skeletal muscle lineage. Twist2 overexpression was capable of reversibly repressing myogenesis and promoting dedifferentiation of myotubes. Through integrated genomic analyses, I show that Twist2 epigenetically remodels the chromatin landscape to redirect MyoD binding from myogenic loci to oncogenic loci, enabling MyoD to adopt novel functions. Our findings identify the previously unknown roles of Twist2 in regulating mammalian muscle biology.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

1. **Li, S**., Karri, D., Sanchez-Ortiz, E., Jaichander, P., et al. Sema3a-Nrp1 signaling mediates fast-twitch myofiber specificity of $Tw2^{+}$ cells. (2019) Submitted
2. **Li, S.**, Chen, C., Zhang, Y., Barnes, S., et al. (in press). Amplification of Twist2 in rhabdomyosarcoma represses myogenesis and promotes oncogenesis by redirecting MyoD DNA binding. *Genes Dev*. (2019)
3. Liu, N., Garry, G. A., **Li, S.**, Bezprozvannaya, S., Sanchez-Ortiz, E., Chen, B., et al. (2017). A Twist2-dependent progenitor cell contributes to adult skeletal muscle. *Nature Cell Biology*, *19*(3), 202–213. http://doi.org/10.1038/ncb3477
4. Zhu, J., **Li, S.**, Wangler, C., Wangler, B., Bruce Lennox, R., & Schirrmacher, R. (2015). Synthesis of 3-chloro-6-((4-(di-tert-butyl[18F]fluorosilyl)-benzyl)oxy)-1,2,4,5-tetrazine ([18F]SiFA-OTz) for rapid tetrazine-based 18F-radiolabeling. *Chem. Commun.*, *51*, 12415–12418. http://doi.org/10.1039/C5CC03623B
5. Moustafa-Kamal M, Gamache I, Lu Y, **Li S**, Teodoro JG. BimEL is phosphorylated at mitosis by Aurora A and targeted for degradation by *β*TrCP1.*Cell Death and Differentiation*. 2013;20(10):1393-1403. http://doi:10.1038/cdd.2013.93

LIST OF FIGURES

LIST OF TABLES

LIST OF DEFINITIONS

bFGF – basic fibroblast growth factor

Bp – base pair

ChIP – chromatin immunoprecipitation

CSA – cross-sectional area

CTX – cardiotoxin

DM – differentiation media

DTA – diphtheria toxin

EMT – epithelial-mesenchymal transition

FACS – fluorescence activated cell sorting

FAP – fibroadipogenic progenitor

FBS – fetal bovine serum

FNRMS – fusion-negative rhabdomyosarcoma

FPRMS – fusion-positive rhabdomyosarcoma

GFP – green fluorescent protein

GM – growth media

GP – gastrocnemius and plantaris

IP – intraperitoneal

iTwist2 – inducible Twist2

MCK – muscle creatine kinase

MEFs – mouse embryonic fibroblasts

MSCs – muscle stem cells

mT/mG – membrane tdTomato/membrane GFP

My32 – fast myosin

ORF – open reading frame

RMS – rhabdomyosarcoma

SC – satellite cell

T7E1 – T7 endonuclease I

TA – tibialis anterior

tdTO – tdTomato

TMX – tamoxifen

TOPO-TA – topoisomerase-based thymidine to adenosine cloning

Tw1 – Twist1

Tw2 – Twist2

WGA – wheat-germ agglutinin

CHAPTER ONE
Introduction

SKELETAL MUSCLE

Overview

Function

Skeletal muscle is a highly regenerative tissue that is vital for all vertebrate life. It accounts for 30-40% of body mass and its proper function enables us to move, breathe, and interact with our surroundings and environment. Impaired skeletal muscle function through genetic or acquired means can be devastating, resulting in significant debilitations or even death. Therefore, the body dedicates significant resources to maintain the structure and function of skeletal muscle. These processes ultimately result in a tightly controlled and coordinated signaling network that enables muscles to adapt and grow in response to injury and stress.

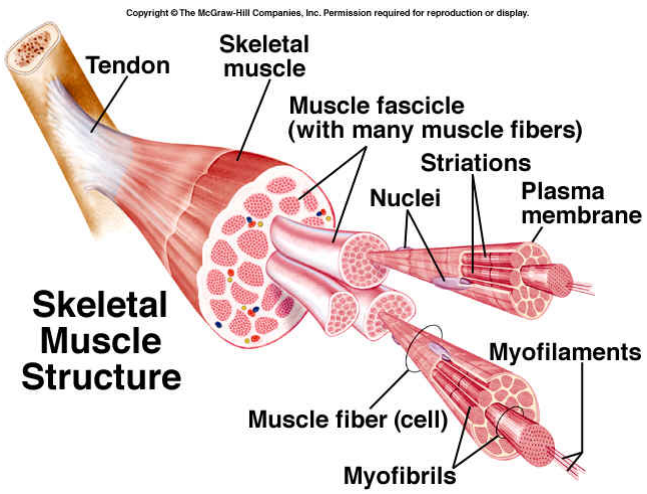


Figure 1-1. Skeletal muscle anatomy

Structure

A key component of skeletal muscle is its highly organized gross and molecular structure which enables the muscle to contract. On a gross scale, muscle is composed of bundles of parallel myofibers (Figure 1-1). Each myofiber is a single, multinucleated syncytium composed of contractile units called sarcomeres. These sarcomeres consist of actin thin filaments and myosin thick filaments, which slide across each other to enable the coordinated contraction required for force generation. At the periphery of these contractile units are the muscle nuclei (myonuclei). Myonuclei are positioned throughout the cytoplasm to enable optimal protein synthesis and maintenance of the muscle sarcomere. These "myonuclear domains" are responsible for local protein synthesis and protein degradation. The balance between these two processes are a key aspect of muscle homeostasis, hypertrophy, and atrophy.

Fiber-type

While the overall functions of mammalian myofibers are the same, there exist sub-types of myofibers that have unique contractile and metabolic properties. These fibers are classified as type I, IIa, IIb, and IIx based on the isoform of myosin heavy chain that is expressed. Type I fibers are slow-twitch fibers and typically found in postural muscles (Schiaffino and Reggiani 2011). Type II fibers are fast-twitch fibers and typically found in muscles required for strength and power generation (Schiaffino and Reggiani 2011). At a metabolic level, Type I and IIa fibers are oxidative fibers and contain significant amounts of mitochondria (Schiaffino and Reggiani 2011). Due to this property, muscles containing these

fibers are preferentially used during long endurance exercises where a consistent amount of energy output is required (Schiaffino and Reggiani 2011). On the other hand, Type IIb and IIx fibers have a glycolytic profile and are responsible for exercises that require a burst of force and action (Schiaffino and Reggiani 2011). Their large stores of glycogen enable very rapid and powerful contractions, however long-term usage results in lactic acid buildup and subsequent muscle fatigue (Schiaffino and Reggiani 2011).

Development

Muscle development in mice occurs through a series of spatiotemporally controlled events beginning with the segmentation of the paraxial mesoderm around embryonic day 8 (Chal and Pourquié 2017). These segmentations result in the formation of paired somites that flank the developing neural tube (Shi and Garry 2006). Through additional signals, somites further differentiate into the sclerotome and dermomyotome around embryonic day 9 (Chal and Pourquié 2017). The sclerotome will later form the skeletal system, while progenitors from the dermomyotome migrate to form the myotome to generate the skeletal muscle lineage (Shi and Garry 2006). During this process, a population of progenitors expressing the transcription factors Pax3 and Pax7 arise (Chal and Pourquié 2017). These progenitors undergo two stages of myogenesis to form embryonic muscle. During primary myogenesis, interstitial $Pax3^{+}$ progenitors differentiate to form primary myofibers (Hutcheson et al. 2009**)**. A subset of these progenitors will downregulate Pax3 and upregulate Pax7, initiating secondary myogenesis (Murphy and Kardon 2011). During this process, the progenitors fuse to primary myofibers or with each other to form secondary myofibers **(**Murphy and Kardon

2011). After birth, these myogenic progenitors adopt a new anatomical location beneath the muscle basement becoming the muscle satellite cells required for muscle regeneration (Murphy and Kardon 2011).

Regeneration

Skeletal muscle is a remarkably regenerative tissue, capable of completely repairing

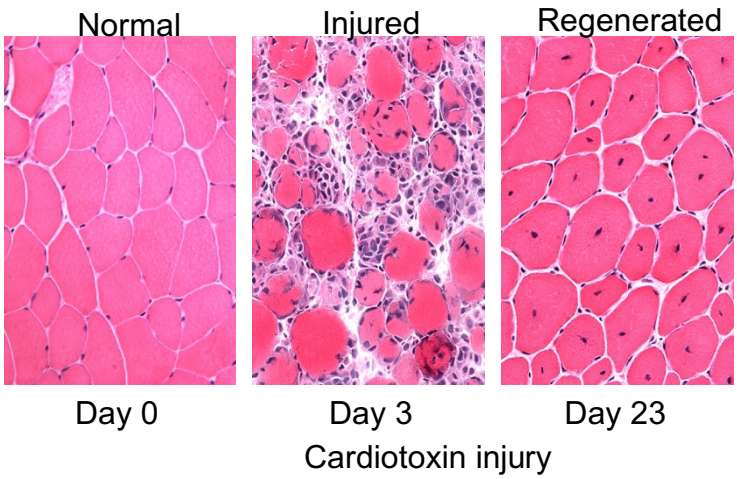


Figure 1-2. Muscle regeneration after injury

itself within a few weeks of injury (Liu et al. 2012). This healing capacity is due to the presence of resident stem cells called satellite cells (Lepper et al. 2011). During the injury process, a coordinated remodeling event occurs that enables activation of satellite cells. Upon activation, satellite cells can repair muscle fibers through fusion with existing fibers, or with each other to form new myofibers (Lepper et al. 2011). The end result is the contribution of additional myonuclei to meet the increased demand for protein synthesis required by a regenerating myofiber (Lepper et al. 2011). Many other cell types within the muscle also play

important roles in regulating the repair process. These cells will be discussed later in this chapter.

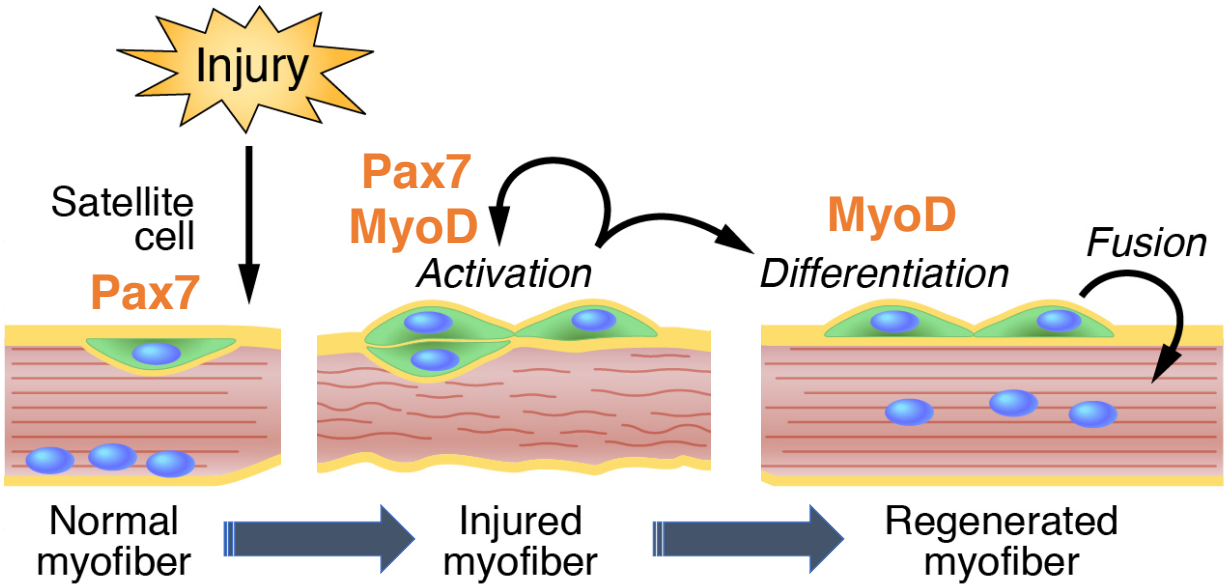


Figure 1-3. Satellite cell activation during regeneration

Myf5, MyoD expression enables satellite cells to proliferate and marks their differentiation into myoblasts (Shi and Garry 2006). During the late stage of myoblast differentiation, MyoD and Myf5 expression subsides while transcription factors such as Myogenin and Mrf4 act in conjunction with myocyte enhancer factor 2 (MEF2) to maintain the differentiation program **(**Shi and Garry 2006**)**. The ultimate result is the upregulation of muscle fusogens such as Myomaker (Mymk) and Myomixer (Mymx), which enable satellite cells to fuse and form terminally differentiated myofibers (Millay et al. 2013; Bi et al. 2017).

Resident Stem Cells

Satellite Cell Identification

Satellite cells were first identified in 1961 by Alexander Mauro through the examination of Xenopus muscle using electron microscopy (Mauro 1961). He observed the presence of cells closely associated with muscle fibers, orbiting them like satellites (Mauro 1961). Since their initial identification, satellite cells have become one of the most well-studied model systems for stem cell and progenitor biology. A fundamental aspect for their scientific appeal is their unique expression of lineage factors like Pax7 and surface markers such as Sdc4, Itga7, CD34, and VCAM1 (Maesner et al. 2016). The restricted expression of these factors enables laboratory isolation of satellite cells from skeletal muscle and their subsequent expansion in vitro (Maesner et al. 2016). Additionally, cultured satellite cells exhibit a remarkably similar differentiation scheme to their in vivo activation (Danoviz and Yablonka-Reuveni 2012).

Satellite Cell Niche

A key aspect of satellite cell function and signaling arises from their unique anatomical niche. Unlike the majority of mononuclear cells which reside within the interstitium of skeletal muscle, satellite cells reside within the myofiber compartment between the basement membrane and the myofiber itself (Yin et al. 2013). This intimate association enables the rapid response and repair observed upon muscle injury (Yin et al. 2013). Additionally, its niche enables satellite cells to receive signals from blood vessels, interstitial cells, the muscle fiber, and the basement membrane (Yin et al. 2013).

Quiescence and Self-Renewal

For a long period of time, it was debated whether or not satellite cells were true stem cells or merely myogenic progenitors. However, recent studies have found that satellite cells do indeed exhibit the key hallmarks of stem cells: maintenance of a quiescent state and the ability self-renew through asymmetric division (Kuang et al. 2007). Pax7 and other epigenetic regulators play important roles in regulating the quiescent state of satellite cells (Kuang et al. 2006). These factors are down-regulated in response to muscle injury, allowing satellite cells to activate and undergo a round of asymmetric division, followed by multiple rounds of symmetric division (Kuang et al. 2007). During asymmetric division, satellite cells divide perpendicularly with respect to the basement membrane (Kuang et al. 2007). The daughter cell that will re-enter quiescence remains attached to the basement membrane, while the daughter cell that will continue differentiating associates with the myofiber (Kuang et al. 2007). These distinct cell fates are specified by Notch3, which is found highly expressed in the self-renewing quiescent daughter cell (Kuang et al. 2007).

In addition to asymmetric division, satellite cells will also undergo multiple rounds of symmetric division (Kuang et al. 2007). Typically, this occurs after a single round of asymmetric division, whereby differentiated satellite cells can expand to generate the required cell number to repair muscle (Kuang et al. 2007). This coordinated sequence of events enables both the rapid expansion of a differentiating cell population as well as maintenance of the initial source of quiescent stem cells.

Interstitial Cell Populations

Fibroadipogenic Progenitors

While satellite cells are key architects of muscle development and regeneration, many interstitial cell-types also contribute vital supporting roles in mediating the myogenic process. Amongst the most well studied interstitial cell-types are fibroadipogenic progenitors (FAPs) (Joe et al. 2010; Uezumi et al. 2010). These mesenchymal cells are responsible for the accumulation of fibrotic and adipose tissue within muscle during aging, injury, and disease (Wosczyna and Rando 2018). While heterogeneous in nature, the predominant FAP population can be isolated from muscle through markers such as PDGFRa and Sca1 (Wosczyna and Rando 2018). In culture these cells show remarkable lineage plasticity, with the ability to differentiate towards adipogenic, fibrogenic, chondrogenic, and osteogenic lineages (Wosczyna and Rando 2018). During the injury response, FAPs rapidly expand and secrete growth factors such as IGF-1, IL-6, and Wnt proteins that can enhance satellite cell myogenic potential (Wosczyna and Rando 2018). Additionally, FAPs can crosstalk with infiltrating immune cells as well as remodel the extracellular matrix (Wosczyna and Rando 2018).

Endothelial Cells

In order for circulating growth factors to reach muscle, vascular remodeling is required. Endothelial cells play a key role in regulating this process through two steps. Initially upon muscle injury, there is a dramatic decrease in oxygen content due to degeneration of capillaries (Wosczyna and Rando 2018). This process, mediated by HIF2a, triggers a temporary hypoxia which helps to inhibit myogenic differentiation and promote satellite cell self-renewal (Xie et al. 2018). During this time, there is rapid expansion of

satellite cells (Xie et al. 2018). Afterwards, endothelial cells undergo rapid angiogenesis in order to revascularize the muscle (Scholz et al. 2003). During this time, the increased oxygen content enables the expanded satellite cell pool to undergo differentiation and subsequent muscle repair(Yin et al. 2013) After the injury process, endothelial cell proximity is maintained to enable quiescence through VEGFA and Notch signaling (Verma et al. 2018).

Pericytes

Residing adjacent to endothelial cells are pericytes expressing markers such as NG2 proteoglycan and alkaline phosphatase (Dellavalle et al. 2007). These cells have been shown to play both supporting and myogenic roles during muscle regeneration (Dellavalle et al. 2007; 2011). Initially, it was shown that pericytes were capable of myogenic potential in vitro as well as during transplantation (Dellavalle et al. 2007). However recent studies showing genetic ablation of this cell population found decreased myofiber diameter, suggesting a role for pericytes in regulating satellite cell quiescence and differentiation (Kostallari et al. 2015).

Interstitial Muscle Progenitors

Since the early 2000s, several interstitial cells have been characterized as having myogenic potential. Some of these cell types include $NG2^{+}$ pericytes, PW1 interstitial cells, muscle side population cells, muscle derived stem cells, and myoendothelial cells (Dellavalle et al. 2007; Mitchell et al. 2010; Majka et al. 2003; Qu-Petersen et al. 2002; Zheng et al. 2007). The majority of these cells exhibit myogenic potential in vitro and show some

capacity to regenerate muscle through transplantation (Wosczyna and Rando 2018). Additionally, they are typically Pax7^{-} suggesting a lineage distinct from satellite cells (Wosczyna and Rando 2018). Despite these findings, mice harboring genetic ablation of satellite cells are unable to regenerate injured muscle, suggesting that myogenic potential from interstitial progenitors are unable to compensate for loss of satellite cells (Lepper et al. 2011; Sambasivan et al. 2011). How these cells contribute during physiologic and pathologic processes remains to be determined.

Aging and Disease

Sarcopenia

Sarcopenia is the loss of muscle mass during the normal aging process (Morley et al. 2001). It is not well understood how or why this process occurs; however many hypotheses have been proposed (Morley et al. 2001). Some of these include the loss of anabolic hormones, reduced circulating growth factors, upregulation of proteolysis, increased inflammatory response, and enhanced autophagy (Jones et al. 2009). Additionally, satellite cell dysfunction through basement membrane and extracellular matrix remodeling can also play a significant role in regulating the muscle environment during the aging process (Lee et al. 2016). One unique aspect of sarcopenia is fiber-type specific atrophy (Morley et al. 2001). Aging-related muscle wasting preferentially affects fast-twitch fibers, particularly type IIb and IIx fibers (Wang and Pessin 2013; Ciciliot et al. 2013). Hormonal and neurologic mechanisms have been proposed for these fiber-type specific changes (Wang and Pessin

2013), however the role of muscle progenitors and other interstitial cells in this process is unknown.

Cachexia

In addition to physiologic muscle wasting during aging, muscle wasting can also occur under pathologic conditions such as cancer (Fearon et al. 2011). This process, known as cachexia, is typically found in late stage cancer patients and ultimately results in death (Fearon et al. 2011). The exact causes and mechanisms of cachexia are also unknown; however recent studies have begun to shed light on both the muscle extrinsic as well as intrinsic mechanisms of wasting (Fearon et al. 2011). One proposed initiating event of cachexia includes release of circulating factors such as activin A that impair muscle hypertrophy and promote atrophy through activation of the myostatin pathway (Tisdale 2010). Additionally, upregulation of inflammatory cytokines due to the cancer can also potentially work through this pathway (Guttridge et al. 2000). Nutrient starvation may also play a role in mediating muscle-wasting seen with tumors of the gastrointestinal tract (Michaelis et al. 2017). Interestingly, a zinc transporter has been recently identified as a mediator of cachexia suggesting a larger role of this ion in regulating muscle mass (Wang et al. 2018). The majority of these upstream cachexia signals converge onto a core pathway of muscle atrophy and proteolysis (Fearon et al. 2011). The key proteins that mediate this process of muscle breakdown include the E3-ligases Atrogin-1 and Murf-1, as well as the FOXO family of transcription factors (Bodine and Baehr 2014).

Rhabdomyosarcoma

While skeletal muscle can be subjected to wasting, it can also serve as a home to cancer. Rhabdomyosarcoma (RMS) is an aggressive pediatric soft-tissue sarcoma expressing hallmarks of the muscle lineage, including MyoD, Myf5, and Desmin (Saab et al. 2011). These tumors typically arise in young patients between the ages of 6-12 and can be subdivided into two categories based on their molecular characteristics (Skapek et al. 2019). Fusion-positive RMS (FPRMS) represent tumors formed due to a balanced chromosomal translocation that generates a PAX3-FOXO1 fusion protein (Skapek et al. 2019). Fusion-negative RMS (FNRMS) encompass tumors that do not contain this fusion-protein (Skapek et al. 2019). Despite the difference in oncogenic drivers, the key similarity between these tumors is the inability of tumor cells to undergo normal myogenic differentiation, even in the presence of muscle regulatory factors (Skapek et al. 2019). Studies have showed that normal MyoD function is impaired in RMS cells, potentially through dysregulated interactions with RMS-specific transcription factors (MacQuarrie et al. 2013b; Tapscott et al. 1993). Identification of the factors which impair MyoD function and restoration of differentiation ability may be key to unlocking novel therapeutics for RMS treatment.

TWIST-FAMILY TRANSCRIPTION FACTORS

History and Identification

Twist was first identified in drosophila as a gene required for dorso-ventral patterning and formation of the mesoderm **(**Franco et al. 2011; Simpson 1983**)**. Mammals contain two orthologs of drosophila Twist: Twist1 and Twist2 **(**Franco et al. 2011). Twist1 was first

discovered in 1991 based on its high amino acid sequence similarity to drosophila Twist (Wolf et al. 1991). On the other hand, Twist2 was initially identified in 1996 in our lab through a yeast-two-hybrid screen for binding partners of E12 (Li et al. 1995). Initially referred to as Dermo1 because of its high expression in the developing dermis, it was later renamed Twist2 due to its high sequence similarity and expression pattern to Twist1 (Lee et al. 2000).

Overview of Basic-Helix-Loop-Helix Transcription Factors

Twist proteins belong to the basic-helix-loop-helix (bHLH) family of transcription factors which include proteins like MyoD, Myc, and NeuroD (Jones 2004). These transcription factors play universally important roles in regulating development and lineage-specification, and their dysregulation can often result in diseases like cancer (Jones 2004). bHLH transcription factors typically function through homo or heterodimerization with other bHLH members, which enables binding to conserved CANNTG motifs termed E-boxes (Jones 2004). Through recruitment of additional transcription factors and epigenetic remodelers, bHLH proteins can act as both activators and repressors of transcription (Jones 2004; Franco et al. 2011).

Role of Twist in Development

Twist proteins are expressed during gastrulation and play important roles in patterning the mesoderm (Füchtbauer 1995). In mice, homozygous loss of Twist1 expression results in embryonic lethality due to failure of neural tube closure (Chen and Behringer

1995). Loss of Twist2 also appears to be lethal, however the phenotype varies depending on the background of mice (Šošić et al. 2003). These results, along with the high sequence identity between these two proteins suggest that they may be functionally redundant (Franco et al. 2011**)**. Twist is also expressed in the developing dermomyotome and marks cells that eventually migrate to become the dermis **(**Barnes and Firulli 2009**)**. In the adult, Twist expression is mostly restricted to adipose, dermal, mesenchymal, and hematopoietic cells (Tabula Muris Consortium et al. 2018).

Role of Twist in Cancer

While Twist1 and 2 expression gradually declines throughout embryonic development, they are often reactivated during cancer (Yang et al. 2004; Qin et al. 2012; Li et al. 2012). Similar to their role in development, Twist proteins serve as master regulators of epithelial-mesenchymal transition (Yang et al. 2004**)**. This process may play a role in regulating metastasis of tumor cells to new tissues and organs **(**Yang et al. 2006**)**. Additionally, Twist proteins can also negatively regulate cell death and may impart drug resistance to various tumors resulting in enhanced survivability (Maestro et al. 1999**)**.

Role of Twist in Differentiation

Immune system

The role of Twist proteins in regulating differentiation has been characterized in several cell lineages. In the immune system, Twist proteins have been shown to impair the expression of NF-kB and pro-inflammatory cytokines **(**Šošić et al. 2003**)**. Additionally, they

also serve as negative regulators of myeloid development (Merindol et al. 2014). Twist2 appears to play a more significant role in this process due to the elevated levels of inflammatory cytokines which appear in Twist2-null mice (Šošić et al. 2003).

Osteogenesis

In the context of the skeletal system, Twist proteins play roles as both positive and negative regulators of osteoblast differentiation depending on the formation of homo- or heterodimers (Franco et al. 2011). These functions largely occur through modulation of Runx2 expression, a master regulator of bone development (Yousfi et al. 2002). Twist1 appears to play a more significant role in osteoblast differentiation given that Twist1-heterozygous mice and humans have defects in bone formation and manifest with premature fusion of cranial sutures (Ghouzzi et al. 1997).

Myogenesis

Lastly, Twist has been shown to negatively regulate myoblast differentiation (Baylies and Bate 1996; Hebrok et al. 1994). In drosophila, Twist expression marks adult muscle precursors where it acts to repress myogenesis (Tapanes-Castillo 2004). Twist expression is down-regulated in response to decreased Notch signaling, resulting in subsequent myoblast differentiation (Tapanes-Castillo 2004). In the mammalian system, the role of Twist in muscle is poorly characterized and has previously been unknown. In vitro studies have shown that Twist1 and Twist2 can both inhibit differentiation of C2C12 myoblasts, potentially through interactions with MyoD and MEF2 (Hebrok et al. 1994; Spicer et al.

1996; Gong and Li 2002). Whether or not mammalian Twist may play a parallel role to drosophila Twist in skeletal muscle progenitors is unknown.

CHAPTER TWO

A TWIST2-DEPENDENT PROGENITOR CELL CONTRIBUTES TO ADULT SKELETAL MUSCLE

Acknowledgement

Parts of this chapter, including figures, have been reproduced, with or without modifications, from our previously published work (Liu et al. 2017).

Abstract

Skeletal muscle possesses remarkable regenerative potential due to satellite cells, an injury-responsive stem cell population located beneath the muscle basal lamina that expresses Pax7. By lineage tracing of progenitor cells expressing the Twist2 (Tw2) transcription factor in mice, we discovered a myogenic lineage that resides outside the basal lamina of adult skeletal muscle. Tw2+ progenitors are molecularly and anatomically distinct from satellite cells, are highly myogenic in vitro, and can fuse with themselves and with satellite cells. Tw2+ progenitors contribute specifically to type IIb/x myofibres during adulthood and muscle regeneration, and their genetic ablation causes wasting of type IIb myofibres. We show that Tw2 expression maintains progenitor cells in an undifferentiated state that is poised to initiate myogenesis in response to appropriate cues that extinguish Tw2 expression. Tw2-expressing myogenic progenitors represent a previously unrecognized, fibre-type-specific stem cell involved in postnatal muscle growth and regeneration.

Introduction

Skeletal muscle is among the most regenerative adult tissues. Its remarkable regenerative capacity originates from a population of resident stem cells, termed satellite cells (SCs), located beneath the muscle basal lamina (Chang and Rudnicki 2014). SCs are marked by expression of Pax7, a transcription factor critical for muscle regeneration (Chang and Rudnicki 2014). In response to injury and disease, SCs become activated and undergo self-renewal and differentiation to form new myofibers(Chang and Rudnicki 2014; Brack and Rando 2012; Sacco et al. 2008). While SCs are essential for muscle regeneration, their genetic ablation in adult mice does not accelerate sarcopenia (Fry et al. 2015; Keefe et al. 2015; McCarthy et al. 2011). Thus, additional mechanisms or cell types might contribute to maintenance of muscle mass during aging.

Skeletal muscle is composed of heterogeneous myofiber types that differ in contractile and metabolic properties and expression of distinctive myosin isoforms. Four major fiber types are present in rodent muscles: one type of slow-twitch fiber (type I) and 3 types of fast-twitch fibers (type IIa, IIx/d, and IIb). While type I and type IIa fibers exhibit oxidative metabolism and high endurance; type IIx and IIb fibers are glycolytic and display low endurance (Schiaffino and Reggiani 2011). Slow and fast twitch fibers also differ in their responses to hypertrophic or atrophic stimuli. For example, type IIb and IIx myofibers are more susceptible than slow twitch fibers to a variety of atrophic signals such as denervation, nutrient deprivation, cancer cachexia, and chronic heart failure (Wang and Pessin 2013; Tonkin et al. 2012; Arany 2008). While SCs can fuse into all myofiber types in injured

muscle (Pawlikowski et al. 2015), it remains unknown whether fiber-type specific myogenic progenitors might also exist.

The Drosophila basic helix-loop-helix transcription factor Twist is expressed in muscle progenitors during embryogenesis and is essential for the formation of mesoderm and muscle (Barnes and Firulli 2009; Baylies and Bate 1996; Cripps et al. 1998). Within the adult musculature of Drosophila, Twist expression is restricted to muscle precursors that are normally quiescent but are activated by extracellular cues to regenerate the adult musculature during metamorphosis (Bate et al. 1991; Cripps and Olson 1998; Currie and Bate 1991). Two mammalian Twist genes, Twist1 (Tw1) and Twist2 (Tw2), are expressed in various mesenchymal cell types, but not in differentiated myofibers (Füchtbauer 1995; Li et al. 1995). Tw1 and Tw2 have been shown to block myogenesis in vitro (Li et al. 1995; Rohwedel et al. 1995; Hebrok et al. 1994; Spicer et al. 1996; Hjiantoniou et al. 2008), but their potential roles in muscle formation or regeneration in mammals have not been explored.

Here, we traced the fate of Tw2-dependent cell lineages in mice and discovered that Tw2 expression marks a previously unrecognized interstitial myogenic progenitor cell that forms type IIb/x myofibers in adult muscle. Tw2-expressing progenitors represent a population of myogenic progenitor cells that contributes to specific fiber types during muscle homeostasis and regeneration, highlighting the ancestral functions of Twist as a regulator of muscle formation.

Results

Twist Expression in Interstitial Cells Within Adult Skeletal Muscle

In adult muscle, Tw2 transcript is barely detectable in whole G/P muscle at 1, 2 and 4 months of age by RNA-seq analysis, in contrast to MyoD and Myh4 that are readily detected (Figure 2-1a). Real-time RT-PCR revealed that Tw2 was highly enriched in mononuclear non-myofiber cells compared to whole quadriceps muscle (Figure 2-1b). Immunostaining of transverse sections of gastrocnemius muscle from 3 months old wild-type (WT) mice revealed Tw2 protein in interstitial cells outside of the myofibers, but not within myofibers (Figure 2-2a). Furthermore, Tw2 protein was not co-localized with Pax7, which was restricted to SCs beneath the basal lamina (Figure 2-2a and b). Similar mutual exclusivity of expression of Tw2 and Pax7 was observed in muscles of 12 month-old mice (Figure 2-1c). We conclude that Tw2 is expressed in the myofiber interstitium and not in mature myofibers or SCs in adult muscle.

To analyze Tw2 expression during muscle regeneration, we performed cardiotoxin (CTX) injury on tibialis anterior (TA) muscle of WT mice, and harvested muscles on 4, 7, and 14 days post-CTX injury. Tw2+ and Pax7+ cells were detected by immunostaining. We did not observe cells that expressed both Pax7 and Tw2 proteins at any time point examined (Figure 2-2c and d). Tw2+ cells declined immediately following CTX injury but the number of Tw2+ cells was rapidly restored by day 7 post-CTX.

To further compare the contributions of Tw2+ and Pax7+ cells to adult muscle, we performed lineage tracing for Pax7 by breeding Pax7-CreERT2 mice (Lepper et al. 2009) with ROSA26-loxp-stop-loxp-tdTomato (R26-tdTO) reporter mice. At 3 months of age, Pax7-CreERT2; R26-tdTO mice were treated with tamoxifen (TMX) on 3 alternate days, and TA muscles were injected with CTX 1 week later. Immunostaining revealed that

upon CTX injury, tdTO+ labeling (representing the Pax7+ lineage) marked regenerating myofibers but did not co-localize with Tw2+ cells (Figure 2-2e and f). There was also no overlap between tdTO signal and Tw2 protein expression in uninjured contralateral muscle (Figure 2-1d). Together, these results confirm that Tw2+ cells and Pax7+ SCs represent distinct cell lineages.

Tw2+ Cells Contribute to a Subset of Type II Myofibers

To identify and trace the fates of Tw2-expressing cells in vivo, we introduced a TMX-inducible Cre expression cassette (CreERT2) into the mouse Tw2 gene by homologous recombination and bred mice harboring this allele to the R26-tdTO mice (Figure 2-3a, b). Following treatment of Tw2-CreERT2; R26-tdTO mice with TMX on 3 alternate days at 8 weeks of age, we monitored tdTO expression as a marker for the Tw2 lineage in skeletal muscle (Figure 2-7a). Ten days after TMX treatment, we observed tdTO labeling of individual cells within the muscle interstitium, outside of the basal lamina, but no labeling of myofibers (Figure 2-7b, left and middle panels). The anatomical location of the tdTO+ interstitial cells differed from that of Pax7+ SCs, which reside beneath the basal lamina. Immunostaining confirmed that tdTO+ cells do not express Pax7 protein (Figure 2-7b, right panel). The tdTO+ interstitial cells often exhibited a distinctive spindle-shaped appearance with elongated processes that was also clearly distinct from the morphology of SCs (Figure 2-7b, middle panel).

By 3 weeks following TMX treatment, strong tdTO staining was observed in the G/P muscles, and the number of tdTO+ myofibers continued to increase up to 8 weeks post-TMX

treatment, at which time 58% of myofibers were labeled by tdTO (Figure 2-7c, d). No tdTO+ myofibers were observed in the absence of TMX (Figure 2-3c), validating the specificity of the lineage tracing method. Strong tdTO labeling was observed in all hindlimb muscles, diaphragm, and masseter muscles by 5 months post-TMX with the exception of the tongue (Figure 2-3d, e). Instead, interstitial cells within the tongue and adjacent epithelial cells along the surface of the tongue were labeled with tdTO (Figure 2-3e). Longitudinal sections of G/P muscle at 4 months post-TMX also revealed tdTO expression throughout the entire length of myofibers (Figure 2-3f). Tw2+ cells did not give rise to other cell types such as endothelial cells (marked by CD31) or fibroblasts (marked by vimentin) in adult muscle (Figure 2-3g).

To determine whether the Tw2+ lineage contributed to specific myofiber types, we analyzed tdTO labeling in G/P and soleus muscles, which are comprised of different proportions of type I and type II myofibers. There was no overlap between type I myofibers and tdTO+ myofibers in the soleus, but a subset of type II fibers within G/P that also expressed tdTO (Figure 2-7e). Further staining revealed tdTO labeling of type IIb but not type IIa myofibers in G/P muscle (Figure 2-7e). Some tdTO+ fibers were negative for IIb staining (Figure 2-7e). By exclusion, we interpret these to be type IIx fibers. Therefore, we conclude that Tw2 cells specifically label type IIb/x fibers. The fiber type-specific tdTO labeling remained even 18 months after the TMX pulse (Figure 2-4a, b). To determine whether the fiber-type specificity also remained in aged mice, we performed lineage tracing on 7-month old Tw2-CreERT2; R26-tdTO mice, and observed robust contribution of tdTO+ cells to myofibers and the same type II fiber-specific labeling at 3 months post-TMX (Figure 2-4c, d).

To further assess the type IIb/x myofiber specificity of the Tw2 lineage, we compared the contribution of SCs to adult myofibers by treating Pax7-CreERT2;R26-tdTO mice with TMX at 8 weeks of age. Pax7+ SCs contributed to myofibers in all muscles including tongue at 8 weeks post-TMX (Figure 2-5a), consistent with prior studies (Pawlikowski et al. 2015; Keefe et al. 2015). SCs also labeled all types of myofibers (Figure 2-5b, c). This finding is consistent with a recent study showing that Pax7+ SCs fuse into myofibers in uninjured skeletal muscle and that a fiber-type preference does not exist for Pax7+ SCs during fusion (Pawlikowski et al. 2015). Thus, in contrast to the broad muscular contributions of SCs, the Tw2-expressing lineage contributes specifically to adult type IIb/x myofibers

To determine whether Tw2+ cells contribute to embryonic myogenesis during development, we performed lineage tracing of Tw2+ cells by breeding mice harboring a constitutively active Cre expression cassette inserted in the Tw2 locus with R26-tdTO mice (Šošić et al. 2003; Yu et al. 2003). At embryonic day (E) 10.5, Tw2+/tdTO+ cells were present in regions surrounding the somites, but not within somitic muscle (Figure 2-6). At E15.5 and P1, tdTO was expressed in interstitial cells within various muscle groups but not within myofibers (Figure 2-6). Together, these findings demonstrate that Tw2+ cells do not contribute to primary or secondary myogenesis in embryos but are dedicated to the formation of muscle postnatally.

Ablation of the Tw2+ Lineage Causes Atrophy of Type IIb Myofibers

We genetically ablated Tw2+ cells in adult mice by breeding the Tw2-CreERT2 mice to mice harboring a diphtheria toxin (DTA) expression cassette in the Rosa26 locus (R26-

DTA mice) (Voehringer et al. 2008). Tw2-CreERT2; R26-DTA/+ (named Cre+;DTA) mice and control R26-DTA/+ (named DTA) mice were injected with TMX and maintained on TMX-containing diet beginning at 4 weeks of age (Figure 2-8a) and analyzed at 13 months of age. G/P and quadriceps muscles were significantly smaller in Cre+;DTA mice compared to DTA control mice, whereas other tissues such as heart, kidney and liver were approximately comparable in size (Figure 2-8b and Figure 2-9). Histological analysis of liver, WAT and BAT also revealed no significant difference between DTA and Cre+; DTA mice (Figure 2-9c). Intriguingly, in Cre+; DTA mice, we observed a significant decrease in the mean cross-sectional area (CSA) of type IIb fibers in G/P, quadriceps and masseter muscles (Figure 2-8c,d). The mean CSA of other fiber types remained unchanged (Figure 2-8d). The number of type IIb fibers per field as well as the total fiber numbers were also increased in quadriceps and masseter but not G/P muscles (Figure 2-9d). No centralized nuclei were observed in Cre+ DTA mice. These results indicate that ablation of Tw2+ cells caused type IIb fiber-specific atrophy, supporting the notion that Tw2+ cells are important for maintenance of type IIb myofiber size during adulthood.

The Tw2 Lineage Contributes to Muscle Regeneration

To determine whether the Tw2 lineage forms new myofibers in response to injury, we performed CTX injury on TA muscle (Figure 2-10a). Seven days after CTX injection, we observed tdTO labeling of newly regenerated myofibers, marked by centralized nuclei, and the labeling became more robust by day 14 post-CTX when tdTO+ myofibers represented ~27.4% of newly regenerated myofibers (Figure 2-10b). Myosin staining revealed that 55%

of the regenerated tdTO+ myofibers were type IIb myofibers (Figure 2-10c, d and Figure 2-11a).

To determine whether activated Tw2+ cells differentiate autonomously or fuse with new myofibers, we immunostained muscle sections for desmin, which is highly expressed in immature muscle fibers during fetal life and regeneration (Goebel 1995; Helliwell 1988) (Figure 2-10e). Although most of the activated Tw2+ cells were negative for desmin expression (Figure 2-11b), we occasionally observed small desmin+/tdTO+ myofibers indicating that Tw2+ cells differentiated into desmin-expressing myofibers (Figure 2-10e). Activated Tw2+ cells did not express Pax7, as demonstrated by immunostaining (Figure 2-11c, d).

To further determine whether Tw2 cells contribute to regeneration by initiating myogenesis autonomously or by fusing with existing myofibers, we bred Tw2-CreERT2 mice to R26-mT/mG mice, which constitutively express a membrane-targeted tdTO protein (mT) from the *Rosa26* locus (Muzumdar et al. 2007) (Figure 2-12a). Upon Cre-activation, tdTO fluorescence is lost and membrane-targeted eGFP (mG) becomes expressed. If Tw2 cells fuse with existing myofibers, both signals will be present, but a *de novo* myofiber formed from Tw2+ cells will only be green. We analyzed muscles from Tw2-CreERT2; R26-mTmG/+ mice at 4 months post-TMX. In the absence of injury, almost all GFP+ myofibers retained tdTO expression (Figure 2-12b), indicating that Tw2+ cells contribute to myofibers via fusion with existing fibers during homeostasis, as reported for satellite cells (Keefe et al. 2015). Upon CTX-injury, although the majority (93.8%) of mG+ new myofibers also expressed mT, 6.2% of new myofibers expressed mG and lost mT expression (Figure 2-10b,

c). These findings demonstrate that Tw2+ cells can initiate myogenesis autonomously during regeneration, although the majority of these cells fuse with newly regenerated myofibers.

For comparison, we performed CTX injury on Pax7-Cre-ERT2; R26-tdTO mice following the same treatment regimen as in (Figure 2-10a). The Pax7+ lineage labeled all regenerating myofibers (Figure 2-9n). We conclude that following injury Tw2+ cells contribute to a subset of new type IIb myofibers that are also derived from the Pax7 lineage.

To test whether Tw2+ cells also possess the capacity to engraft into injured muscle, we isolated Tw2+ (tdTO+) cells by FACS sorting at 10 days post-TMX and injected them into the TA muscles of *mdx* mice, which display extensive muscle degeneration and regeneration. Four weeks post-transplantation, tdTO+ myofibers were observed in the TA muscle of *mdx* mice, revealing the ability of exogenous Tw2+ cells to engraft and form new muscle (Figure 2-12c). However, the engraftment efficiency was limited with an average of only 5.5 myofibers per field (Figure 2-12d), which is lower than the engraftment capacity of SCs.

Freshly Isolated Tw2+ Cells are Distinct from Pax7+ SCs

To further characterize Tw2+ cells, we examined expression of cell-surface markers on freshly isolated Tw2+ cells by FACS analysis from 8-week old Tw2-CreERT2; R26-tdTO/+ mice 10 days post-TMX (Table 1, and Figure 2-13a, b). Only 0.46% of Tw2+ cells expressed the SC-specific marker α7-integrin. Similarly, other SC markers, including CD34, Vcam1 (CD106), and CXCR4 (CD184), were expressed in only a small fraction of Tw2+ cells. The majority of tdTO+ cells also did not express markers for endothelial cells (CD31)

or hematopoietic cells (CD45) (Figure 2-13a-c). Approximately 98.8% of tdTO+ cells were positive for CD29 (β1-integrin), a widely expressed marker of mesenchymal stem cells (MSCs), lymphocytes, hematopoietic stem cells (HSC), and other cell types. In addition, tdTO+ cells positive for Sca1 were further separated as Sca1-High (68.8%) and Sca1-medium (8.6%) populations. Other MSC markers, including CD73 and CD105, were expressed in only a small fraction of tdTO+ cells (Table 1). Together, these findings suggest that the Tw2 lineage represents a population of muscle precursors clearly distinct from SCs and most closely resembling cells of a mesenchymal lineage. Expression of cell surface markers remained largely unchanged in Tw2+ cells upon CTX injury (Figure 2-13c).

We performed RNA-seq analysis on freshly sorted Tw2+ and Tw2− cells from Tw2-CreERT2; R26-tdTO mice 10 days post-TMX treatment at 8 weeks of age. For comparison, freshly sorted Pax7+ and Pax7− cells from Pax7-CreERT2; R26-tdTO mice were also analyzed. Freshly isolated Tw2+ and Pax7+ cells showed distinct transcriptome signatures, whereas Tw2− and Pax7− cells displayed similar expression profiles (Figure 2-14a). SC-enriched transcripts such as Pax7, Cadh15 (M-cad), Fgfr4, Sdc4, Met, Itga7 and Vcam1 were all strongly enriched in Pax7+ cells. Strikingly, none of these mRNAs were enriched in Tw2+ cells (Figure 2-14b, c). MyoD transcript was barely detectable in freshly isolated Tw2+ cells in contrast to Pax7+ cells (Figure 2-14b, c). Similarly, Peg3/PW1, a marker for PICs, was not enriched in the Tw2+ cells relative to either Tw2− cells or Pax7+ cells (Figure 2-14b, c). On the contrary, genes highly enriched in Tw2+ cells (relative to Tw2− cells) were all repressed in Pax7+ cells, again confirming that Tw2+ and Pax7+ cells have distinct transcriptome signatures (Figure 2-14d). Furthermore, platelet derived growth factor receptor

alpha (PDGFRα), a marker for human and mouse MSCs, was highly enriched in Tw2+ cells but barely detected in Pax7+ cells (Figure 2-14e). PDGFRβ, a marker for pericytes and vascular smooth muscle cells, was also highly enriched in Tw2+ but not Pax7+ cells (Figure 2-14e). Ingenuity Pathway Analysis (IPA) revealed that mRNAs enriched in Tw2+ cells relative to Pax7+ cells are involved in cancer metastasis, interleukin/NFAT/endothelin signaling, and extra-cellular matrix (ECM) remodeling (Figure 2-14f).

Tw2 Cells Transition Through a Pre-Myogenic Pax7+ State

Freshly isolated Tw2+ cells proliferated rapidly in growth medium (GM) and remained tdTO-positive (Figure 2-15a). When switched to differentiation medium (DM), 81% of Tw2+ cells differentiated into multinucleated myosin-positive myotubes (Figure 2-16a, b and Figure 2-15a). Muscle genes such as *Myogenin*, *Ckm*, and *Myh4* were up-regulated in Tw2+ cells after 2 days in DM (Figure 2-16c). The proliferative activity and efficiency of differentiation of Tw2+ cells are comparable to Pax7+ SCs and exceed those of other previously identified interstitial myogenic progenitors (Mitchell et al. 2010; Dellavalle et al. 2011; Doyle et al. 2011).

Although Tw1 and Tw2 transcripts were highly expressed in freshly isolated Tw2+ cells, expression was extinguished after maintaining the cells in culture (Figure 2-16d and Figure 2-15b). In contrast, cultured Tw2+ cells quickly became Pax7-positive and MyoD-positive in GM (Figure 2-16e, f), Thus, whereas Tw2+ cells do not express Pax7 in their native interstitial location in vivo, they transition through a Pax7+ state en route to differentiation. In addition, the full transcriptome profiles between Tw2+ cells and Pax7+

cells in GM and DM were largely similar (Figure 2-15c). Furthermore, when co-cultured at equal numbers in GM and exposed to DM, Tw2+ cells (tdTO+) and Pax7+ SCs (GFP+) fused with each other to form chimeric myotubes that were positive for both GFP and tdTO (Figure 2-15d, e). We conclude that although Tw2+ cells are distinct from Pax7+ SCs, when removed from their native environment and cultured in vitro, Tw2+ cells acquire a Pax7+ cell fate and display similar myogenic potential to Pax7+ cells.

FACS analysis revealed that a small fraction (2.6%) of Tw2+ cells (tdTO+) expressed CD34, a marker commonly used to isolate SCs (Table 1). To exclude the possibility that this small fraction of CD34+ cells might be responsible for the myogenic capacity of Tw2+ cells in vitro, we isolated the tdTO+/CD34- population by FACS. Freshly sorted tdTO+/CD34− cells proliferated rapidly within 48 hours post isolation (Figure 2-17a). After 4 days in DM, 50% of tdTO+/CD34− cells formed multi-nucleated myotubes, and after 10 days in DM, the fusion index increased to 66%, with many myotubes contracting spontaneously (Figure 2-18a, b). These results exclude the possibility that the myogenic potential of Tw2+ cells comes from a rare population of tdTO+/CD34+ cells.

To examine whether the tdTO+/CD34− cells can form other cell lineages in addition to muscle cells, we exposed them to different conditions that support the formation of adipocytes and osteoblasts, respectively. After 10 days in adipogenic medium, the majority of tdTO+/CD34− cells formed multinucleated myotubes and only 1.5% of the cells differentiated into adipocytes (positive for Oil Red O staining), among more than 5,000 nuclei analyzed (Figure 2-18c, d). Thus tdTO+/CD34− cells are not adipogenic in culture. Under osteogenic conditions, the majority of tdTO+/CD34− cells died within the first 24 hrs

and after 10 days in culture, the few surviving cells proliferated and differentiated into osteoblasts, which stained for alkaline phosphatase (Figure 2-18e). No My32-positive myotubes were observed under osteoblastogenic conditions. Together, these results indicate that the Tw2 cells that did not express CD34 are myogenic and osteogenic in vitro.

To further confirm the myogenic capacity of Tw2+ cells, we performed clonal analysis on tdTO+/CD34− cells. Briefly, single tdTO+/CD34− clones were seeded by FACS sorting onto each well of 96-well plates that were pre-coated with inactivated mouse embryonic fibroblasts (MEFs) as feeder cells. Single clones were grown in GM with bFGF for 1 week before being switched to DM. After 1 week in DM, 29.8% of surviving clones formed My32+ myotubes, while the others did not express My32 and had different morphologies (Figure 2-18f, g). These results confirmed the intrinsic myogenic capacity of Tw2+ cells.

Tw2 Blocks Myogenesis in vitro

The finding that Tw2+ cells do not express Pax7 or MyoD in vivo, but quickly lose Tw2 gene expression and up-regulate Pax7 and MyoD expression in vitro suggested that Tw2 maintains cells in a progenitor cell state, preventing them from entering the muscle differentiation pathway. To test this hypothesis, we over-expressed Tw2 by retrovirus infection in Tw2+ cells and analyzed myotube formation. While control GFP retrovirus-infected cells formed multinucleated myotubes, Tw2-infected cells in DM underwent a dramatic change in morphology from spindle-shaped myoblasts to more flattened fibroblast-

like shapes and did not form myotubes or express myosin (Figure 2-19a). The few myosin-positive cells observed in Tw2-infected cultures were not infected by Tw2 retrovirus.

The gene expression profiles in GM were very similar between GFP-infected and Tw2-infected cells (Figure 2-19b). However, after 5 days in DM, muscle-specific genes that were up-regulated in GFP-DM cells were not up-regulated in Tw2-DM cells, and conversely, genes strongly up-regulated in Tw2-DM cells were all down-regulated in GFP-DM (Figure 2-19c). IPA pathway analysis revealed that genes involved in cell-cycle regulation, cancer metastasis, and EMT were specifically up-regulated in Tw2-DM cells (Figure 2-19d). These findings confirm that Tw2-overexpression in Tw2+ cells inhibits myogenesis and induces EMT-like programs. Similarly, over-expression of Tw2 in Pax7+ cells inhibited myotube formation and caused morphological changes in DM (Figure 2-19e). EdU labeling showed that Tw2-infected Pax7+ cells continued to proliferate in DM compared to GFP-infected cells, further demonstrating the ability of Tw2 to maintain cells in an undifferentiated state (Figure 2-19f). Tw2 thus functions as a repressor of the myogenic program and its down-regulation triggers the onset of myogenesis in Tw2 progenitors (Figure 2-20).

Discussion

Our results reveal a previously unrecognized population of skeletal muscle progenitor cells marked by expression of Tw2. Tw2+ cells display a unique cell-surface marker expression profile and contribute specifically to type IIb/x fibers. These cells are distinct from Pax7+ SCs and appear to represent a subset of MSCs. In recent years, various non-SC muscle progenitors have been identified, including bone marrow-derived circulating stem

cells, pericytes, PW1+ interstitial cells (PICs), muscle side population (SP) cells muscle-derived stem cells and mesoangioblasts (Mitchell et al. 2010; Dellavalle et al. 2011; Doyle et al. 2011; LaBarge and Blau 2002; Minasi et al. 2002; Qu-Petersen et al. 2002; Dellavalle et al. 2007). Tw2+ cells appear to be distinct from these cell lineages because they display a unique cell-surface marker profile, which does not resemble that of any known non-SC progenitors (Mitchell et al. 2010; Dellavalle et al. 2011; Joe et al. 2010; Kuang et al. 2006). Tw2+ cells share overlapping cell surface marker profiles (e.g., Sca1 and PDGFRa) with muscle-resident fibroadipogenic progenitors (FAPs), but appear functionally distinct with respect to their differentiation potential (Joe et al. 2010). In addition to their robust myogenic potential, Tw2+ cells can form osteoblasts under appropriate conditions.

SCs contribute to all myofiber types in adult mice (Keefe et al. 2015; Pawlikowski et al. 2015), whereas Tw2+ cells contribute specifically to adult type IIb/x myofibers. This specificity remained even in mice that were pulse-chased for 1.5 years. Remarkably, even when Tw2+ cells were in direct contact with type I myofibers, they remained completely distinct throughout life. Genetic ablation experiments revealed that Tw2+ cells are required for the maintenance of Type IIb/x fiber size during homeostasis without affecting muscle regeneration. In contrast, it has been reported that ablation of Pax7+ SCs completely abolished muscle regeneration, but its effect on myofiber size was minimal during aging (Fry et al. 2015; Keefe et al. 2015; McCarthy et al. 2011; Lepper et al. 2011; Murphy et al. 2011; Sambasivan et al. 2011). The distinct phenotypes resulting from depleting Pax7+ versus Tw2+ cells suggested that unlike Pax7+ cells, which are essential for regeneration, Tw2+ cells play more important roles in normal muscle growth. Our results demonstrating that

Tw2+ cells also contribute to muscle regeneration might seem contradictory to the absence of regeneration in Pax7-depleted muscle. An interpretation of these results is that although Tw2+ cells can initiate de novo myogenesis in vivo, the majority of them likely fuse with Pax7+ cells to form IIb/x myofibers. In addition, ablation of Tw2+ cells does not affect muscle regeneration, suggesting that Tw2+ cells are not a reservoir for rapid repair. The Tw2+ cells appear important for skeletal muscle maintenance and are insufficient to mount a rapid and comprehensive response to an acute injury.

Our results show that within their native interstitial locations in adult skeletal muscle, Tw2+ cells are distinct from SCs. However, following isolation they rapidly down-regulate Tw1 and Tw2 expression and activate Pax7 and MyoD expression to enter the myogenic pathway. Expression of Tw2 maintains these progenitors in a stem-like state, and the rapid down-regulation of Tw2 in isolated cells allows the transition to a Pax7+ state and acquisition of a muscle phenotype. Furthermore, forced expression of Tw2 in either Tw2+ or Pax7+ progenitors in culture prevented differentiation and maintained the cells in a proliferative state characterized by expression of genes involved in cell signaling, EMT and oncogenesis. Interestingly, Twist proteins can maintain an immature “stem-like” state and repress terminal differentiation of human MSCs and cancer cells (Beck et al. 2015; Isenmann et al. 2009; Schmidt et al. 2015; Yang et al. 2004). Such functions would be consistent with the role of Twist that we uncovered in skeletal muscle progenitors.

It is curious that Tw2+ cells are specifically excluded from the tongue musculature. Tongue and laryngeal muscles are derived from occipital somites, located at the boundary of the trunk and head, whereas limb, trunk and diaphragm muscles are derived from trunk

somites (Michailovici et al. 2015). The intrinsic difference in initiation of the myogenic program between tongue and limb muscles could offer a clue into why Tw2+ cells are excluded from the tongue musculature. It will be interesting to investigate the potential contributions of Tw2+ cells in different muscle lineages during development.

Tw2+ progenitors are, to our knowledge, the first example of a fiber type specific myogenic progenitor population. In this regard, type IIb fibers are the most abundant fibers and are most susceptible to injury and disease in mice (Webster et al. 1988). Therefore, it is essential to maintain the size and integrity of type IIb fibers during aging, and the contribution of Tw2+ cells to type IIb fibers may represent such a mechanism. Humans do not have type IIb fibers in skeletal muscle; instead the predominant fiber type is equivalent to mouse IIx/d fibers (Scott et al. 2001). It will be interesting to study whether Tw2+ cells exist in human skeletal muscle and whether they contribute to specific fiber types.

Materials and Methods

Generation of Tw2-CreERT2 mice

The targeting strategy for generation of Tw2-CreERT2 mice is shown in Figure 2-3a. A 7.0 kb EcoRI-NotI genomic fragment upstream of the mouse *Twist2* coding region, and a BspEI-HindIII 1.5 kb genomic fragment downstream of the *Twist2* coding region, were used as a long and short arm, respectively, in the construction of the targeting vector. A CreERT2-Frt-Neo-Frt expression cassette was inserted between the two regions, resulting in a vector designed to delete the entire coding region of *Twist2*. The targeting vector was linearized and electroporated into 129SvEv-derived embryonic stem cells. Out of 500 embryonic stem cell

clones, 10% were correctly targeted. Three clones with a properly targeted allele were injected into 3.5-d C57BL/6 blastocysts, and high-percentage chimaeric male mice were crossed to C57BL/6 females to achieve germline transmission of the targeted allele.

Mouse models and treatment

Tw2-Cre (stock no. 008712), R26-tdTO ((*ROSA)26Sor*$^{tm14(CAG-tdTomato)}$, stock no. 007914), R26-DTA ((*ROSA)26Sor*$^{tm1(DTA)Lky}$, stock no. 009669), CAG-eGFP (*C57BL/6-Tg(CAG-EGFP)1Osb/J*, stock no. 003291), R26-mT/mG mice and mdx (C57BL/10ScSn-*Dmd*^{mdx}*/J*, stock no. 001801) mice were obtained from JAX laboratory. The Pax7-Cre-ERT2 mice were provided by C.-M. Fan (Carnegie Institution for Science, Baltimore, USA). Both male and female adult mice were used in the studies. Mice were maintained on a mixed genetic background. All animal procedures were approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center.

Tamoxifen and cardiotoxin treatment

Tamoxifen (TMX, Sigma-Aldrich) was dissolved at 10 mg ml^{-1} in a mixture of sesame oil and ethanol (9:1) and 1 mg of TMX was administered by intraperitoneal injection to 8-week-old mice as schematized in the figures. Tw2-CreERT2; R26-DTA/+ mice and R26-DTA/+ mice were placed on a TMX-containing diet (250 mg kg^{-1}) (Harlan Laboratories) immediately following TMX injection. Cardiotoxin (CTX, from *Naja mossambica mossambica*, Sigma-Aldrich) was dissolved in sterile saline at 10 μM concentration. For CTX injury, 50 μl CTX (10 μM) was delivered by intramuscular injection

to TA muscle, as described previously (Liu et al. 2012). TA muscles were harvested at 3, 7 and 14 days post-CTX injury.

Immunohistochemistry

Skeletal muscles were harvested at desired ages, and fixed in 4% paraformaldehyde at 4 °C for 1 h. Tissues were then switched to 10% sucrose/PBS overnight and in 18% sucrose/PBS at 4 °C overnight, before they were frozen embedded and sectioned as previously described (Liu et al. 2012). For immunohistochemistry, the following antibodies were used: Twist2 (Abcam, no. ab66031, 1:200), Pax7, myosin IIa (SC-71), myosin IIb (BF-F3) (all from Developmental Studies Hybridoma Bank, 1:10), sarcomeric α-actinin (Sigma-Aldrich, no. A7811, 1:100), laminin (Sigma-Aldrich, no. L9393, 1:500), desmin (DAKO, clone D33, 1:100), Slow myosin (Sigma-Aldrich, clone NOQ7.5, 1:250), Fast myosin (Sigma-Aldrich, clone My32, 1:250), vimentin (Sigma-Aldrich, no. V2258 1:100), CD31 (BD Pharmingen, 553370; 1:200). Alexa Fluor secondary antibodies were used according to the manufacturer's instructions. TdTO signals were detected by direct fluorescent imaging.

Immunohistochemistry of frozen sections was performed as previously described (Liu et al. 2012). Wheat germ agglutinin staining was performed on both frozen and paraffin-embedded sections, using WGA-Alexa Fluor 555 (W32464) or WGA-Alexa 647 conjugate (W32466) (Life Technologies, 50 mg ml^{-1}). Images were taken on a Zeiss LSM700 confocal microscope. Muscle cross-sectional area and fibre numbers were quantified by ImageJ.

Fluorescence-activated cell sorting analysis

Following isolation, mononuclear cells were resuspended in PBS/2% BSA at 1 × 10^6 cells per 50 µl and aliquoted into 100 µl per tube, and incubated on ice for 1 h with Fc blocking agent and one of the following fluorophore-conjugated antibodies: APC-CD45 (1:50), APC-CD31 (1:50), APC-CD90.2 (1:50), APC-Cy7 Sca1 (1:50), Alexa Fluor 647-CD106 (1:50) (all from BD Biosciences-Pharmingen), APC-CD29 (1:50) (eBioscience), Alexa Fluor 488-CD34 (1:50) (AbD Serotec), FITC-CD184 (1:50), FITC-Integrin-α7 (1:50). Isotype-specific controls were also performed on these cells. Cells isolated from wild-type mice were used as stained negative controls using Alexa Fluor 488-CD34, Alexa Fluor 647-CD106, FITC-Integrin-α7, APC-Cy7 Sca1 and APC-CD29. Unstained cells isolated from Cre-positive and Cre-negative mice were used as unstained controls. Cells were analysed on FACSCalibur (BD) flow cytometer and FACS data were analysed using FlowJo Software (TreeStar). Values of FACS analysis were averaged from three independent experiments.

Engraftment of $Tw2^+$ cells into mdx mice

TA muscles of adult mdx mice were injected with 25 µl of CTX (10 µM) 1 day before transplantation. $Tw2^+$ cells were isolated by FACS sorting from Tw2-CreERT2; R26-tdTO mice 10 days post-TMX. Freshly isolated cells were resuspended in sterile saline at a concentration of 2,400 cells μl^{-1}. For transplantation, 60,000 cells (25 µl) were injected into TA muscle of mdx mice with a Hamilton gastight syringe and 30G standard hypodermic needle (COVIDIEN). Four weeks post-transplantation, TA muscles were harvested, fixed, sucrose protected, frozen embedded and sectioned. Engraftment experiments were performed three times, each time with three mice.

Cell culture and myogenesis, adipogenesis and osteogenesis in vitro

$Tw2^+$ cells and $Pax7^+$ cells were cultured as previously described (Liu et al. 2012). Satellite cells from CAG-eGFP mice were isolated by FACS sorting as the integrin-α7-positive and CD45/CD31/Sca1-negative population. For mixing experiments, satellite cells from CAG-eGFP mice were co-cultured with $Tw2^+$ cells at a 1:1 ratio in GM and then switched to DM to induce myotube formation.

Freshly sorted $tdTO^+/CD34^-$ cells were grown in GM for 2 days and subsequently differentiated towards myotubes, adipocytes or osteoblasts. For myotube differentiation, the growth medium was replaced with DM for 4–10 days. Cells were then immunostained for MyHC to visualize differentiated myotubes. For adipocyte differentiation, the growth medium was replaced with DMEM high glucose, 10% fetal bovine serum, Pen/Strep, 1.72 μM insulin, 625 nM dexamethasone, 2 μM rosiglitazone and 0.5 mM isobutyl-methyl-xanthine for 10 days. Cells were fixed for 15 min, stained with Oil Red O for 30 min, and then washed three times with PBS. Cells were then visualized with light microscopy. Cells were also stained with My32 and Hoechst to visualize myotubes and nuclei. For osteoblast differentiation, the growth medium was replaced with DMEM high glucose, 10% fetal bovine serum, Pen/Strep, 50 μg ml^{-1} ascorbic acid, 10 mM β-glycerol phosphate, and 100 nM dexamethasone. Cells were fixed and then stained for alkaline phosphatase using SIGMAFAST BCIP/NBT. Cells were then washed and visualized with light microscopy. For quantification of Oil Red O-positive cells, more than 500 nuclei per field and a total of 10 fields were counted and averaged.

Clonal analysis

Mouse embryonic fibroblasts (MEFS) were used as feeder cells for clonal analysis. One day before FACS sorting, confluent MEFS were treated with 8 μg ml^{-1} of mitomycin C for 2.5 h at 37 °C to inhibit cell cycling. Cells were then washed twice with PBS, followed by trypsinization. MEFS were then seeded into pre-Matrigel-coated 96-well plates at a density of 2 × 10^5 cells per well in DMEM/10%FBS/Pen/Strep and allowed to attach overnight at 37 °C. Two hours prior to sorting, the medium was replaced with 100 μl of satellite cell growth media (SCGM) comprising Ham's F10, 20% FBS, 0.2% Primocin and 2.5 ng μl^{-1} bFGF. $Tw2^+$ cells isolated from hind limb muscle were isolated by sorting for the $tdTO^+/CD34^-$ population. Single $Tw2^+$ cells were sorted directly into each well of 96-well plates and were allowed to attach overnight. To prevent cell detachment, 50 μl of fresh SCGM containing the appropriate concentration of bFGF was added on top of existing media on a daily basis. When the maximum volume capacity of each individual well was reached (300 μl), 200 μl of SCGM was removed followed by replacement of 100 μl of fresh GM containing bFGF. Cells were cultured for 1 week in GM, followed by four days of differentiation in differentiation medium (DMEM, 2% horse serum and 0.2% Primocin). Differentiated cells were fixed and immunostained with My32 antibody.

Immunostaining of cultured cells.

No commercial cell lines were used in this study. For immunostaining, cells grown on plates or chamber slides were fixed with 4% paraformaldehyde at room temperature for

15 min and subsequently permeabilized with 0.3% Triton X-100/PBS. Cells were blocked in 10% goat serum diluted in 0.1% Triton X-100/PBS at room temperature for 1 h. Primary antibodies were diluted in 1% goat serum/0.1% Triton X-100 and incubated at room temperature for 2 h. Cells were washed and secondary antibodies were diluted in 1% goat serum/0.1% Triton X-100 and incubated at room temperature for 1 h. Cells were stained with Hoechst dye (1:2,000 in PBS) at room temperature for 10 min. Primary antibodies include: fast myosin (Sigma-Aldrich, My32, 1:250), MyoD (Santa Cruz Biotechnology, sc-377460, 1:100). Fusion index was calculated as a ratio of the number of $tdTO^+$ nuclei within a multinucleated myotube to the total number of $tdTO^+$ nuclei. A minimum of 10 independent microscopic fields was used for each group over three independent differentiation experiments at 1 and 2 days in DM. For EdU labelling, EdU was used at 10 μM on $Pax7^+$ cells for 24 h in DM. EdU detection was performed according to the Click-iT EdU imaging kit (Invitrogen). Images were taken using the Nikon Eclipse Ti microscope.

Retroviral infection of $Tw2^+$ and $Pax7^+$ cells.

The following retroviral plasmid constructs expressing either eGFP or myc-Tw2-IRES-eGFP were cloned into the retroviral vector pMX. Ten micrograms of retroviral plasmid DNA was transfected using FuGENE 6 (Roche) into Platinum E cells (Cell Biolabs), which were plated on a 10-cm tissue culture dish at a density of 3×10^6 cells per dish, 24 h before transfection. After 48 h of transfection, viral medium was harvested and filtered through a 0.45 μm cellulose filter. The viral supernatant was mixed with Polybrene (Sigma) to a final concentration of 6 μg ml^{-1}.

Tw2$^+$ and Pax7$^+$ cells were plated at a density of 200,000 cells per 35-mm plate in growth medium. After 24 h, the growth medium was replaced with freshly made viral mixture containing Polybrene and bFGF (5 ng ml^{-1}). Twenty-four hours later, viral medium was replaced with growth medium with bFGF. Twenty-four hours later, cells were switched to differentiation medium. Cells were either fixed and stained or harvested for RNA isolation at certain time points.

Real-time RT–PCR analysis.

Total RNA was extracted from sorted cells with Trizol (Invitrogen) following the manufacturer's instructions. From this RNA, cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen). Selected gene expression was analysed by real-time RT–PCR using Taqman probes. Probes are twist2 (Mm00492147_m1), twist1 (Mm04208233_m1), myog (Mm00446195_g1), ckm (Mm00432556_m1), myh4 (Mm01332518_m1).

RNA isolation and RNA-seq analysis.

Cells were resuspended in 1 ml of Trizol and homogenized using 20G needles. Following chloroform extraction, the supernatant was mixed with an equal volume of 75% ethanol and loaded on RNeasy mini-columns (Qiagen). Total RNA was isolated according to the manufacturer's instructions (Qiagen). RNA quality was verified by the Agilent 2100 Bioanalyzer and RNA-seq was performed using Illumina HiSeq 2500 by UTSW Genomics and Microarray Core Facility. Quality assessment of the RNA-seq data was done using NGS-

QC-Toolkit (Patel and Jain 2012). Reads with more than 30% nucleotide with phred quality scores less than 20 were removed from further analysis. Quality-filtered reads were then aligned to the mouse reference genome GRCm38 (mm10) using the Tophat2 (v 2.0.0) aligner (Kim et al. 2013) using default settings except for –library-type = fr-firststrand. Aligned reads were counted using featurecount(Liao et al. 2014) (v1.4.6) per gene ID. Differential gene expression analysis was done using the R package edgeR (Robinson et al. 2010) (v 3.8.6). For each comparison, genes were required to have 10 cpm (counts per million) in at least half of the total number of samples to be considered as expressed. They were used for normalization factor calculation. Gene differential expression analysis was done using the GLM approach following edgeR official documentation. (https://www.bioconductor.org/packages/3.3/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf). Cutoff values of fold change greater than 2 were then used to select for differentially expressed genes between sample group comparisons. Normalized gene cpm values were averaged within groups for heat map generation.

Pathway enrichment analysis.

Significant pathway enrichment analysis was performed using Ingenuity Pathways Analysis (Ingenuity Systems). Differentially expressed genes from the RNA expression data are associated with a biological function supported by at least one publication in the Ingenuity Pathways Knowledge Base. Fisher's exact test was used to calculate the *P* value and determine the probability that each biological function was enriched in the data set due to

chance alone. Statistically significant biological pathways were then identified by selection for pathways with *P* values less than 0.05.

Statistics and reproducibility.

All statistical analyses were performed using GraphPad Prism 7 (GraphPadSoftware). Data are presented as mean ± s.e.m. Differences between groups were tested for statistical significance by using the two-sample *t*-test. $P < 0.05$ was considered significant. The number of biological (non-technical) replicates for each experiment is indicated in the figure legends. Two independent sets of RNA samples from $Tw2^+$ cells and $Tw2^-$ cells, and one set of RNA samples from $Pax7^+$ and $Pax7^-$ cells were analysed by RNA-seq. For comparison between DTA and Cre^+; DTA mice, $n = 5$ age- and gender-matched littermates were analysed to provide statistical significance of the study. Randomization was not used in most of the animal studies. However, images of type IIb/laminin staining (Figure 2-8c) were taken by an investigator who was blinded to the group allocation. All immunofluorescence images are representative of at least three independent experiments or mice of the same genotypes. Western blots were representative of three independent experiments. No statistical method was used to predetermine sample size. Experiments described here were not randomized.

Data availability.

RNA-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE84377, GSE84378, GSE84379, GSE84380.

Figures

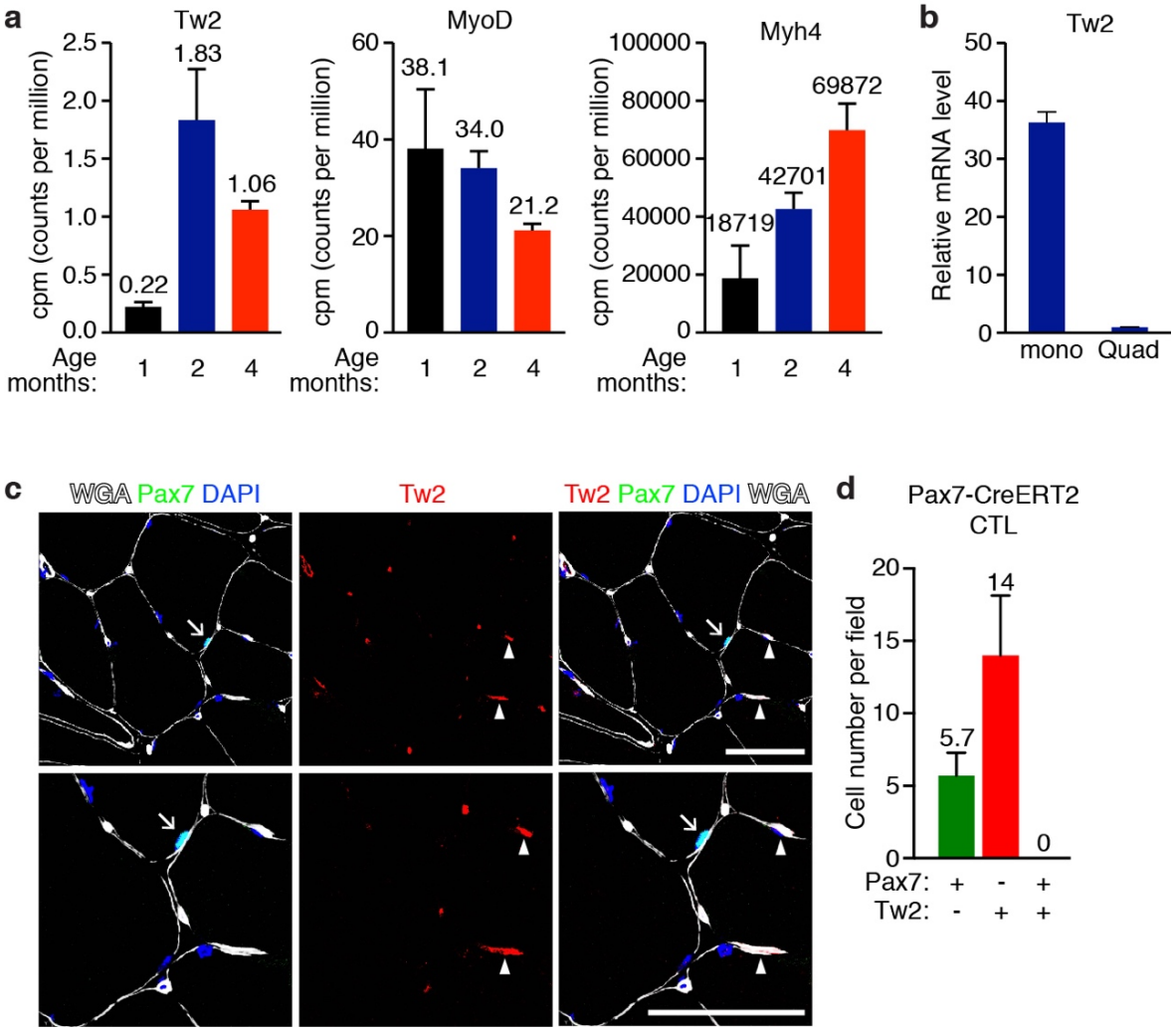


Figure 2-1. (a) CPM (counts per million) of Tw2, MyoD and Myh4 expression by RNA-seq in G/P muscles of WT mice at 1, 2 and 4 months of age. Data are mean ± S.E.M. N=3 mice for each time point. (b) Real-time RT-PCR demonstrates Tw2 mRNA is enriched in mononuclear cells of the adult muscle (mono) compared to whole quadriceps muscle (quad). Data are mean ± S.E.M; N=3 independent experiments. (c) Immunostaining of Tw2 (red) and Pax7 (green) on transverse sections of G/P muscle of 12-month old WT mice.

Myofibers were co-stained with wheat germ agglutinin (white) and DAPI (blue). Arrows indicate Pax7+ cells and arrowheads indicate Tw2+ cells. Scale bar: 50 um. (d) Quantification of the number of Pax7+ (tdTO+), Tw2+, and Pax7+/Tw2+ double positive cells per field in CTL TA muscles of Pax7-CreERT2; R26-tdTO mice. For each muscle section, at least 6 different fields were quantified and averaged. Data are mean ± S.E.M; N=3 mice.

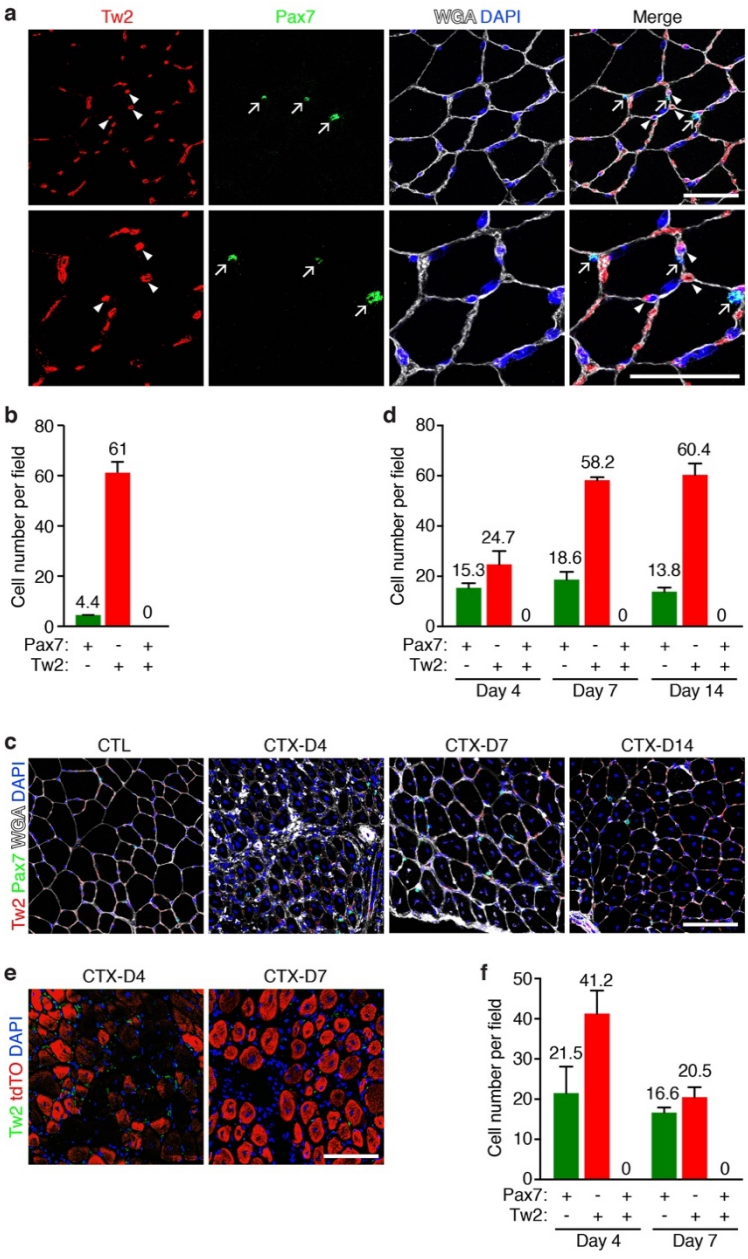


Figure 2-2. (**a**) Immunostaining of Tw2 (red) and Pax7 (green) on transverse sections of gastrocnemius muscle of 3-month-old C57Bl6 WT mice. Myofibres were co-stained with wheat germ agglutinin (WGA; white) and DAPI (blue). The arrows indicate $Pax7^{+}$ cells and the arrowheads indicate $Tw2^{+}$ cells. Scale bars, 50 µm. (**b**) Quantification of the

number of $Pax7^+$, $Tw2^+$, and $Pax7^+/Tw2^+$ double-positive cells per field in 3-month-old C57Bl6 WT mice. Quantification was performed on the top panel images in **a**. For each muscle section, at least 6 different fields were quantified and averaged. Data are mean ± s.e.m. $N = 6$ mice. (**c**) TA muscle of C57Bl6 WT mice was injured by CTX injection and harvested on days 4, 7 and 14 post injury. Muscle sections were stained for Pax7 (green), Tw2 (red), DAPI (blue) and WGA (white). CTL, contralateral TA muscle. Scale bar, 100 μm. (**d**) Quantification of the number of $Pax7^+$, $Tw2^+$, and $Pax7^+/Tw2^+$ double-positive cells per field in C57Bl6 WT mice on days 4, 7 and 14 post-CTX injury. For each muscle section, at least 6 different fields were quantified and averaged. Data are mean ± s.e.m. $N = 3$ mice. (**e**) Tw2 expression in TA muscle of Pax7-CreERT2; R26-tdTO mice after CTX injury. Adult Pax7-CreERT2; R26-tdTO mice were treated with TMX for 3 alternating days. One week after the first dose of TMX, CTX injury was performed on TA muscle. Muscles were harvested at days 4 and 7 post injury and immunostained for Tw2 (green) and DAPI (blue). Scale bar, 100 μm. (**f**) Quantification of the number of $Pax7^+$ ($tdTO^+$), $Tw2^+$, and $Pax7^+/Tw2^+$ double-positive cells per field in Pax7-CreERT2; R26-tdTO on days 4 and 7 post-CTX injury. $Pax7^+$ cells were detected by $tdTO^+$ signal. For each muscle section, at least 6 different fields were quantified and averaged. Data are mean ± s.e.m. $N = 3$ mice.

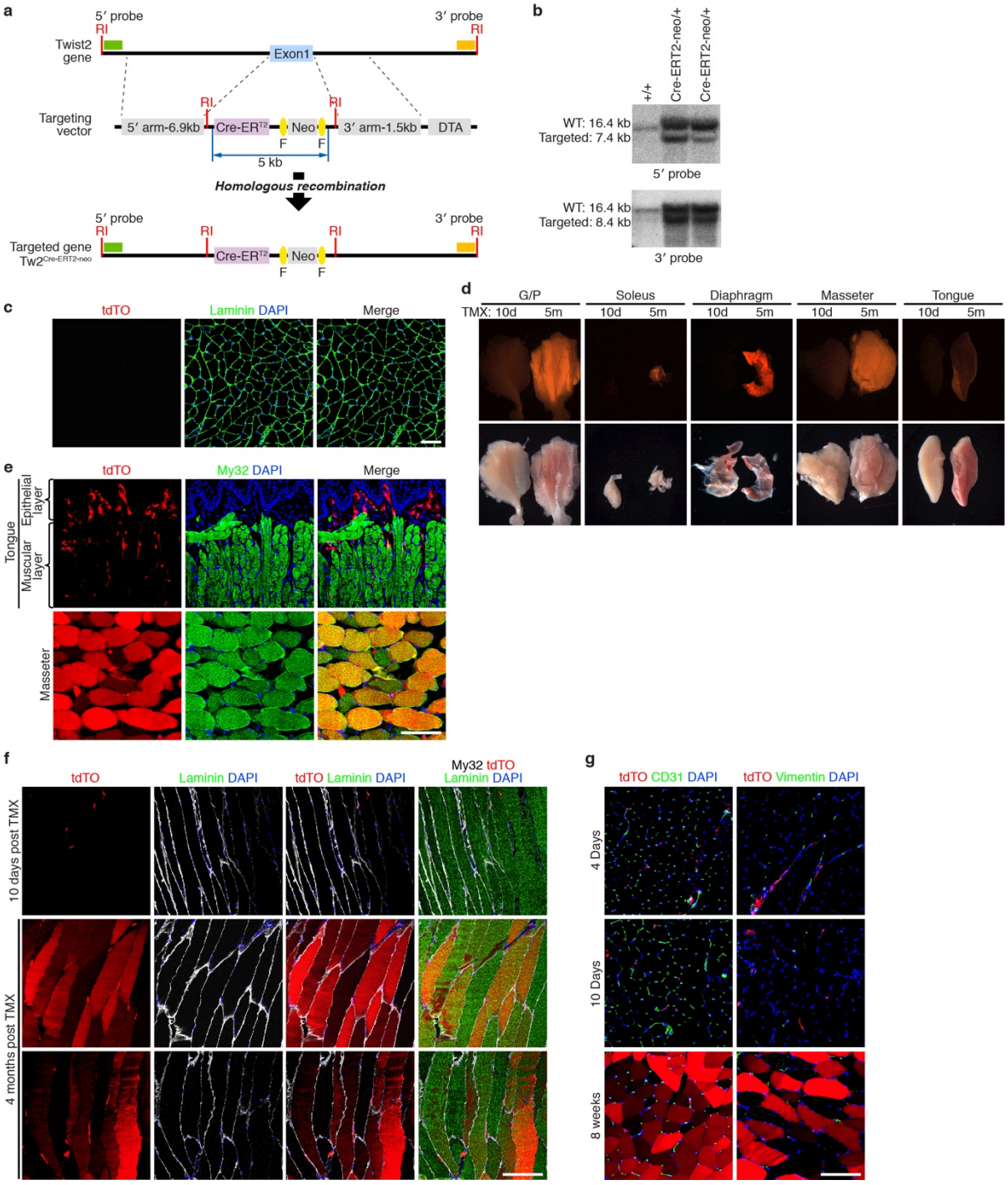


Figure 2-3. (a) Targeting strategy of the Tw2-CreERT2 allele. Coding region of the Tw2 gene was replaced by a CreERT2-Frt-Neo-Frt cassette by homologous recombination in ES

cells. Out of 500 ES cell clones 10% were correctly targeted. (b) Southern blot analysis of ES cells using 5' probe and 3' probes to demonstrate correct targeting. (c) In the absence of TMX treatment, tdTO is not detected in muscles of Tw2-CreERT2; R26-tdTO mice at 9 months of age. Scale bar: 100 um. (d) Whole mount images showing tdTO+ muscles at 10 days (left) and 5 months (right) post-TMX. At 5 months, all muscles examined showed strong tdTO signals except for the tongue muscle. (e) Transverse-sections of tongue and masseter muscles from Tw2-CreERT2; R26-tdTO mice at 4 months post-TMX were co-stained with My32 (green) and DAPI (blue). Scale bar: 100 um. (f) Longitudinal-sections of TA muscle from Tw2-CreERT2; R26-tdTO mice at 10 days and 4 months post-TMX were co-stained with My32 (green), Laminin (white) and DAPI (blue). tdTO signal expands through the entire myofibers. Scale bar: 100 um. (g) Transverse sections of Tw2-CreERT2;R26-tdTO mice at indicated days post TMX were co-stained with CD31 (green; left panel), an endothelial cell marker and vimentin (green; right panel), a fibroblast marker. Tw2+ cells do not give rise to endothelial cells or fibroblasts in adult skeletal muscle. Scale bar: 100 um.

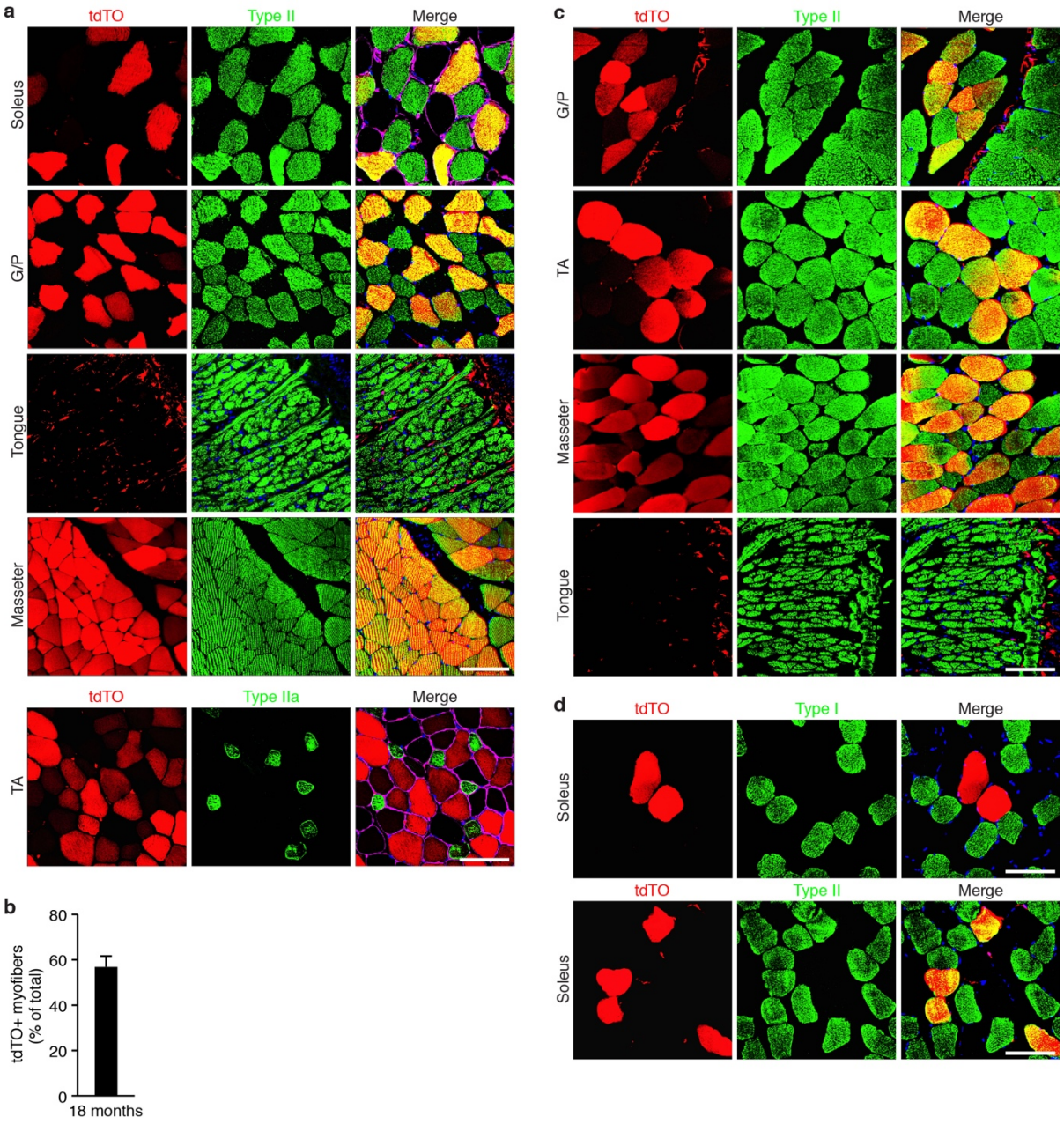


Figure 2-4. (a) Type II myofiber specificity after prolonged labeling of Tw2-CreERT2; R26-tdTO mice. Transverse sections of indicated muscles of Tw2-CreERT2; R26-tdTO mice were obtained at 18 months post-TMX. Myosin staining using a type II specific antibody (My32, green) showed only a subset of Type II fibers were labeled by tdTO. Myofibers in tongue

muscle were excluded from tdTO expression. Bottom panels show co-staining for type IIa. myofibers. Scale bar: 100 um. (**b**) Quantification of the percentage of tdTO+ myofibers among all myofibers in each field in G/P muscle at 18 months post-TMX. Data are mean ± S.E.M; N=3 mice. (**c**) Seven-month old Tw2-CreERT2;R26- tdTO/+ mice were injected with 3 doses of TMX as described in Figure 2-7a. Three months later, muscles were harvested and stained for type II myofibers. Tw2+ cells strongly labeled a subset of type II fibers in G/P, TA, masseter muscles, but not in tongue. Scale bar: 100 um. (**d**) Soleus muscle of the same mice described in Figure 2-3b were co-stained with antibodies against type I and type II myofibers. A subset of type II myofibers but not type I myofibers are labeled by tdTO. Scale bar: 100 um.

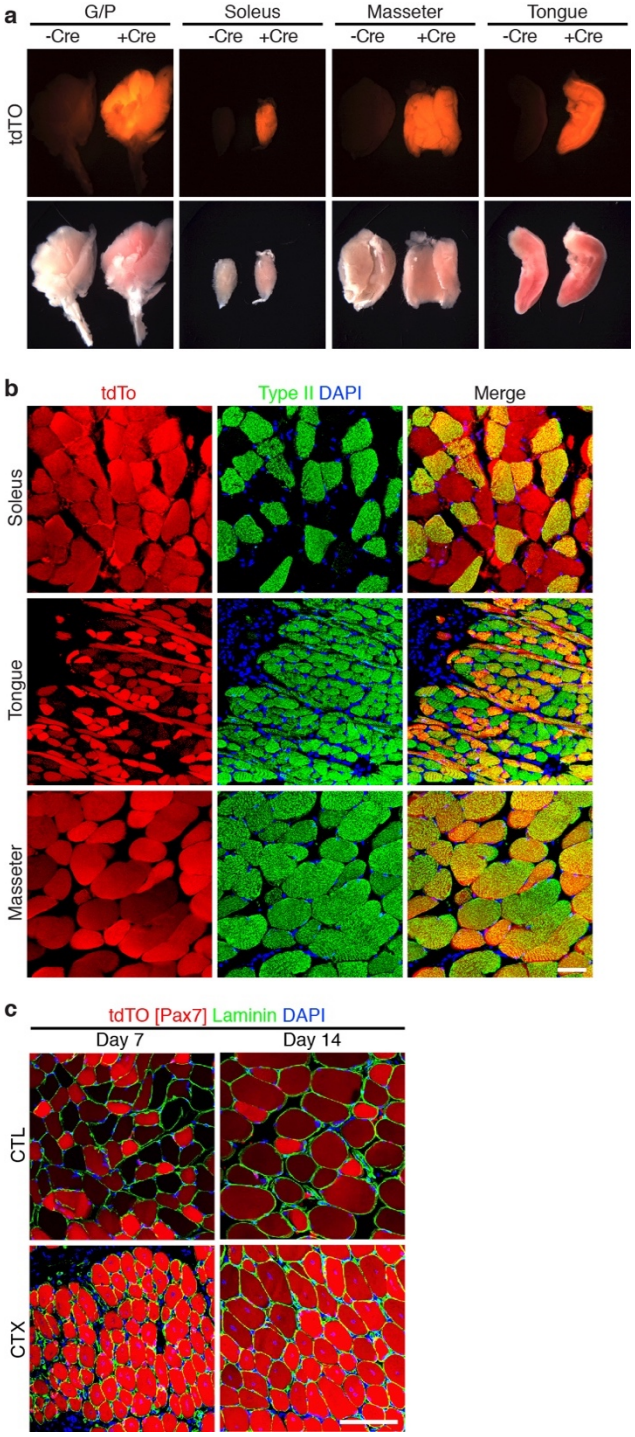


Figure 2-5. (**a**) Whole mount images showed intense tdTO signals in G/P, soleus, masseter and tongue of Pax7-CreERT2; R26-tdTO mice (right) compared to muscles of R26-tdTO

mice (left) 8 weeks post-TMX. Mice were treated with the same regimen as shown in Figure 2-7a. (**b**) Myosin staining using a type II specific antibody (My32, green) showed tdTO signals in the majority of myofibers of soleus, tongue and masseter muscle from Pax7-CreERT2; R26-tdTO mice at 8 weeks post-TMX. Scale bar: 100 um. (**c**) Pax7-CreERT2; R26-tdTO mice were subjected to CTX injury and transverse sections of TA muscles were analyzed 7 and 14 days later. Contralateral uninjured muscle served as control (CTL). The results showed that Pax7+ cells contribute to all regenerating myofibers (indicated by centralized nuclei) on days 7 and 14 after CTX injury. Sections were co-stained with Laminin (green) and DAPI (blue). Scale bar: 100 um.

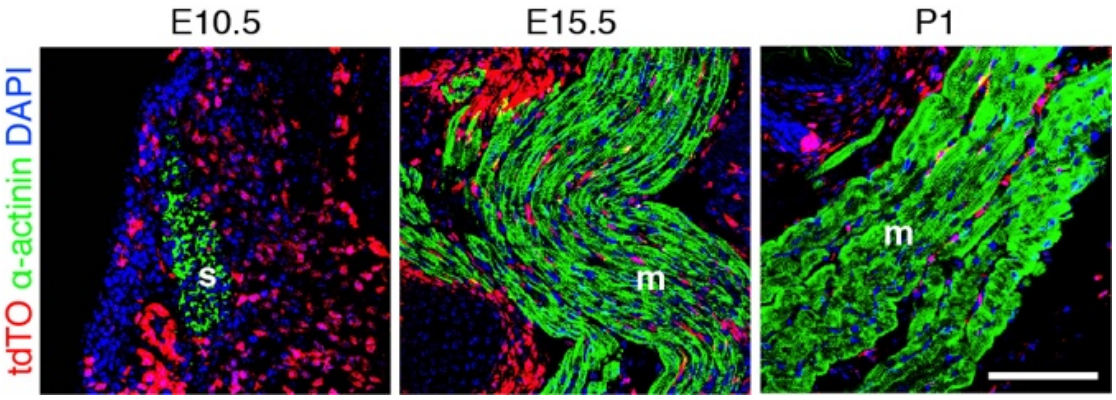


Figure 2-6. Tw2+ cells do not contribute to embryonic myogenesis during development. Sections of Tw2-Cre; R26-tdTO embryos at E10.5, E15.5 and P1 were stained with α-actinin to detect somites (s) and muscle cells (m). Body wall muscles were shown for E15.5 and P1. Scale bar: 100 um.

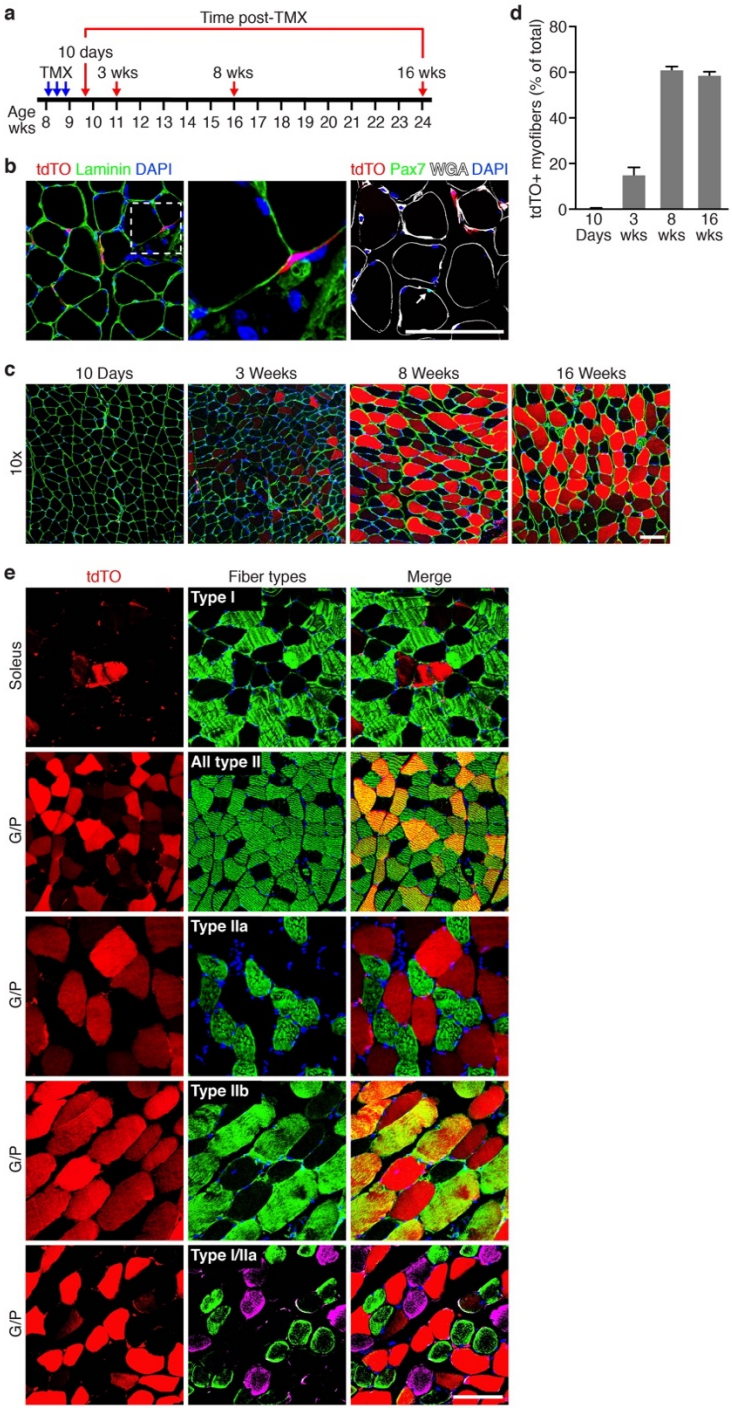


Figure 2-7. (**a**) Schematic of TMX treatment. Tw2-CreERT2; R26-tdTO mice on a mixed genetic background were injected with TMX at 8 weeks of age on 3 alternating days. Mice were analyzed at various time points following initial TMX injection. (**b**) At 10 days post-

TMX, $Tw2^+$cells marked by $tdTO^+$ (red) were located outside the basal lamina (green) (left panel). The middle panel is an enlarged image of the left panel showing the morphology of $Tw2^+$ cells. In contrast, $Pax7^+$ cells (stained with Pax7 antibody, green) were beneath the basal lamina (right panel). The arrow indicates a $Pax7^+$ cell. Scale bar, 100 μm. (**c**) Progressive tdTO labelling of myofibres of G/P muscle at the indicated times following TMX treatment. Transverse sections of G/P muscle were co-stained with laminin (green) and DAPI (blue). Scale bar, 100 μm. (**d**) Quantification of the percentage of $tdTO^+$ myofibres among all myofibres in each field in G/P muscle. Data are mean ± s.e.m. For each time point, $N = 3$ mice were analysed. For each mouse, between 1,000 and 2,000 myofibres of G/P muscle were quantified. (**e**) Transverse sections of different muscle groups obtained from Tw2-CreERT2; R26-tdTO mice at 4 months post-TMX were stained with various fibre-type-specific antibodies (green as indicated). Scale bar, 100 μm.

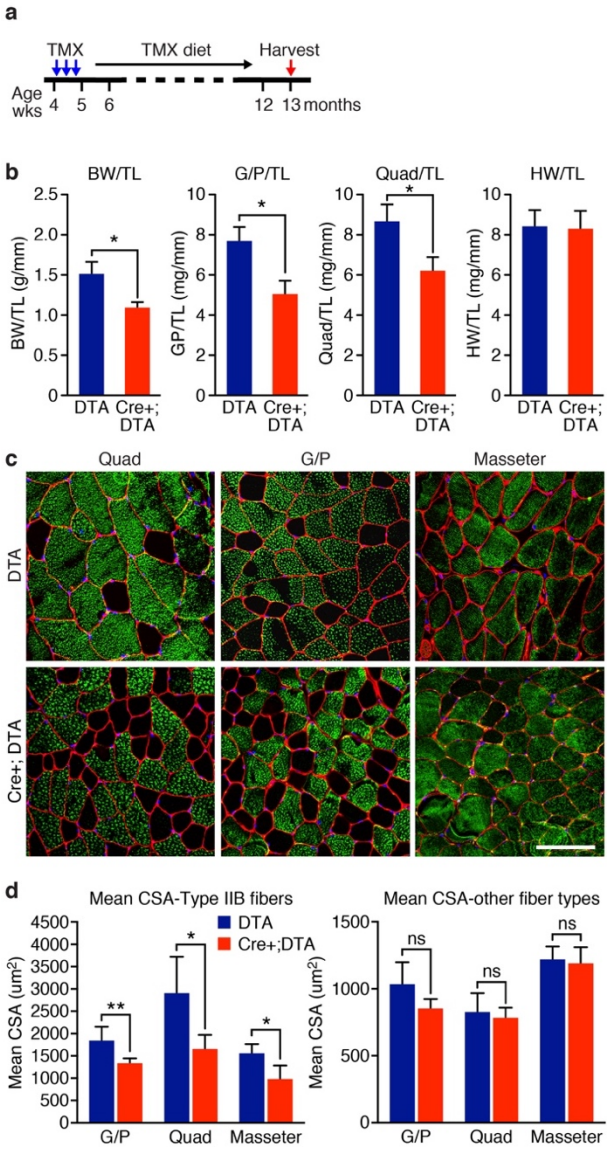


Figure 2-8. (**a**) Schematic of TMX treatment. Tw2-CreERT2; R26-DTA/+ (Cre^{+}; DTA) and R26-DTA/+ (DTA) mice on a mixed genetic background were injected with TMX at 4 weeks of age on 3 alternating days. Mice were kept on a TMX-containing diet until the time of analysis. (**b**) Measurement of body weight (BW), heart weight (HW) and muscle mass normalized to tibia length (TL) of mice. Data are mean ± s.e.m.; two-sample *t*-test;*: $P < 0.05$. $N = 5$ male mice for each genotype. (**c**) Type IIb (green) and laminin (red)

immunostaining of transverse sections of quad, G/P and masseter muscles of Cre^{+}; DTA and DTA mice. Scale bar, 100 μm. (**d**) Measurement of mean myofibre cross-sectional area (CSA) of type IIb fibres (top) and other fibres (bottom). Data are mean ± s.e.m.; two-sample *t*-test; $*P < 0.05$. $**P < 0.01$. NS, not significant. $N = 5$ mice for each genotype. For each muscle, more than 300 fibres were measured and averaged.

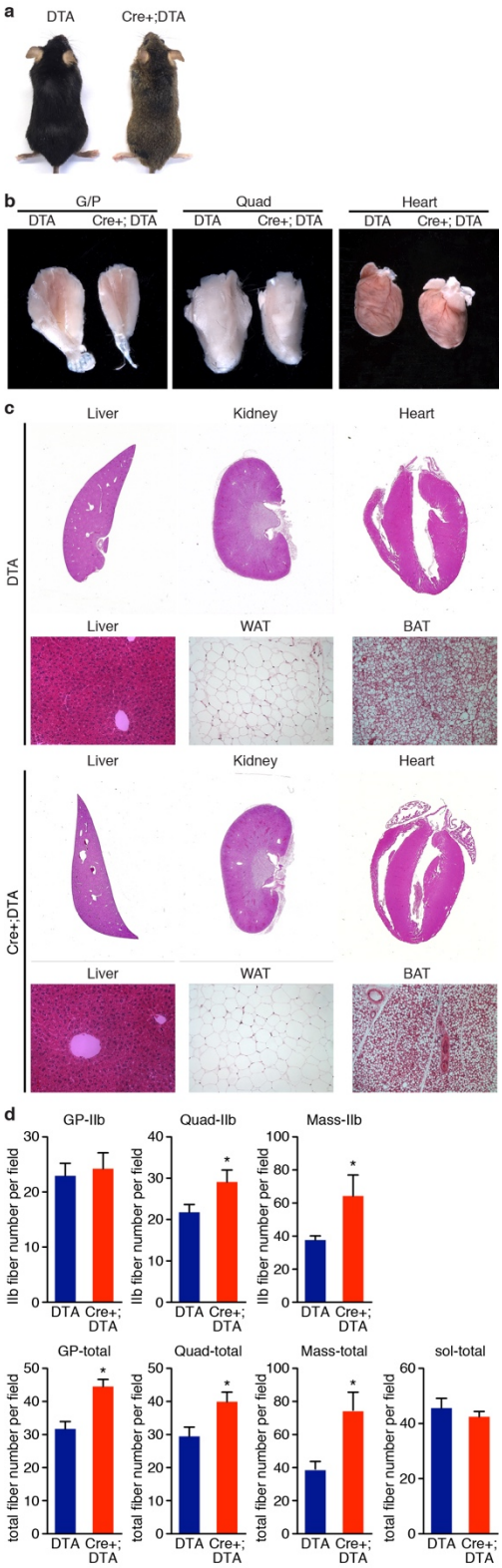


Figure 2-9. (a) Cre+;DTA mice were smaller than control DTA mice at 9 months post-TMX. The difference in coat color is due to the mixed genetic background. (b) Whole mount images of G/P, Quad and heart of Cre+;DTA and control DTA mice at 9 months post-TMX. (c) Hematoxylin and eosin staining of liver, kidney, heart, white adipose tissue (WAT) and brown adipose tissue (BAT) from DTA and Cre+; DTA mice at 9 months post TMX. Black scale bar: 1 mm; White scale bar: 100 um. (d) Quantification of type IIb myofibers per field and total myofibers per field in G/P, quad, mass and soleus muscles of DTA and Cre+; DTA mice. Data are mean ± S.E.M; N =5 mice for each genotype. Two sample t-test; *: $P < 0.05$.

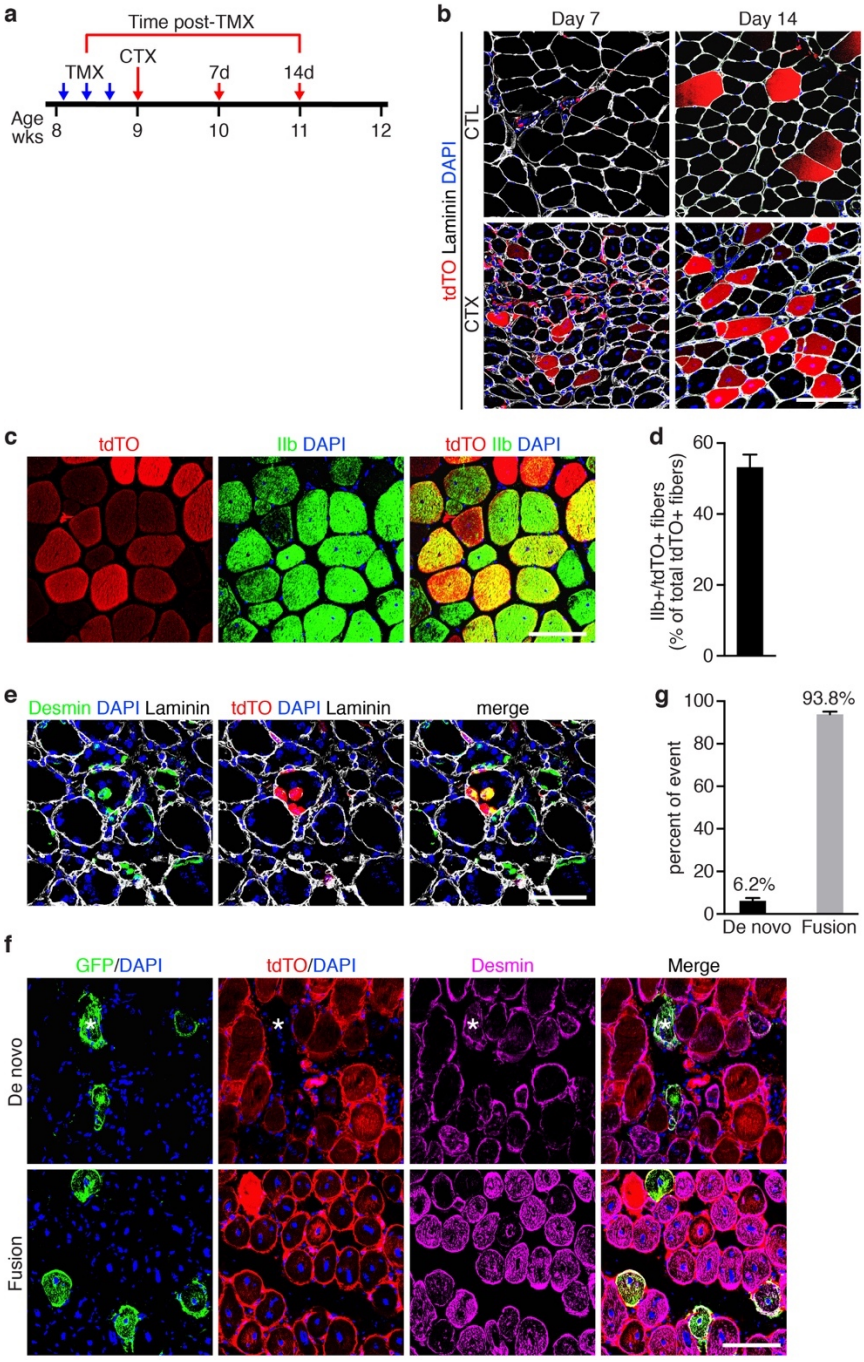


Figure 2-10. (**a**) Schematic of TMX and CTX treatment on adult Tw2-CreERT2; R26-tdTO mice on a mixed genetic background. CTX was injected into the TA muscle at 1 week after the first dose of TMX. Both contralateral (CTL) and CTX-injured TA muscles were

harvested at 7 and 14 days post-CTX. (**b**) Immunostaining of transverse sections of TA muscles revealed that $Tw2^+$ cells are activated and contribute to regenerating myofibres (indicated by centralized nuclei) on days 7 and 14 after CTX injury. Sections were co-stained with laminin (white) and DAPI (blue). Scale bar, 100 μm. (**c**) Co-staining of IIb myosin (green) with newly regenerated $tdTO^+$myofibres on day 14 post-CTX. Scale bar, 100 μm. (**d**) Quantification of the percentage of $tdTO^+$ myofibres that are type-IIb-positive among all $tdTO^+$ myofibres on day 14 post-CTX. Data are mean ± s.e.m.; $N = 5$ mice. (**e**) Three days post-CTX, occasional small $tdTO^+$ cells expressed desmin (green), indicating that these cells differentiated into desmin-positive myoblasts. Scale bar, 100 μm. (**f**) Transverse sections of TA muscle on day 7 post-CTX from Tw2-CreERT2; R26-mT/mG mice on a mixed genetic background at 4 months post-TMX. The asterisk represents a GFP^+ regenerated new myofibre that lost mT (tdTO) expression. Scale bar, 100 μm. (**g**) Percentage of *de novo* versus fusion events in Tw2-CreERT2; R26-mT/mG mice on day 7 post-CTX. Data are mean ± s.e.m.; $N = 3$ mice were analysed.

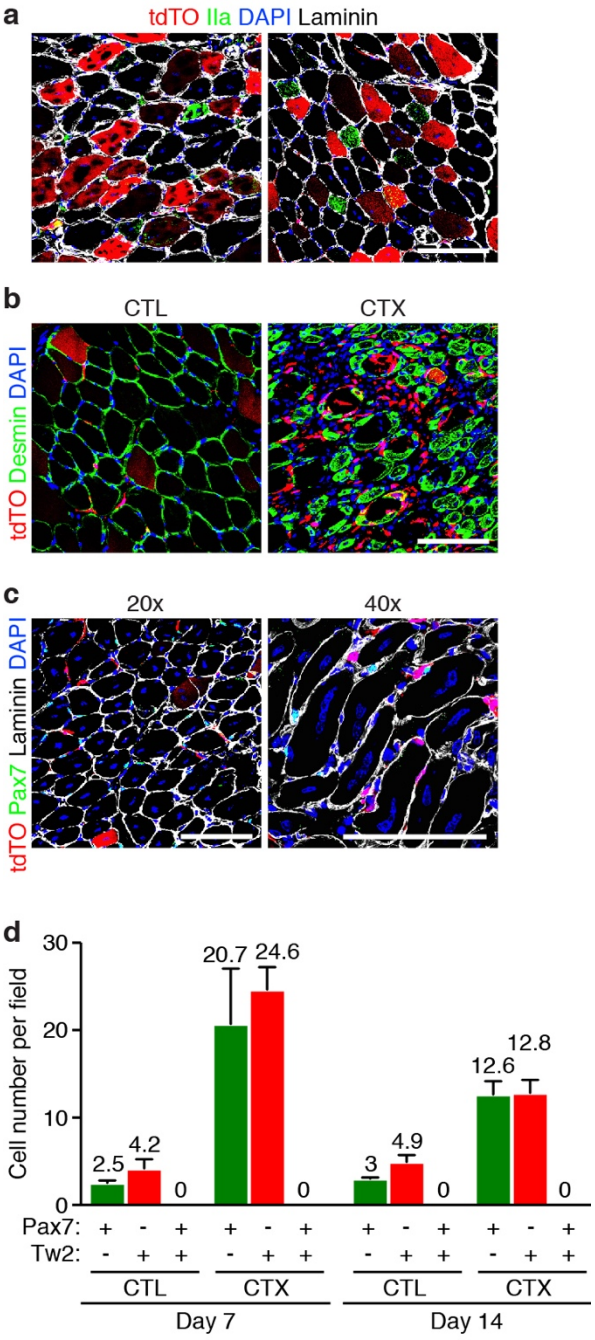


Figure 2-11. (**a**) Co-staining for type IIa myosin revealed regenerated tdTO+ myofibers are not type IIa myofibers. Scale bar: 100 um. (**b**) Sections of TA muscle from Tw2-CreERT2;R26-tdTO mice were co-stained with desmin (green) on day 3 after CTX. The majority of tdTO+ cells are negative for desmin. Scale bar: 100 um. (**c**) Sections of TA muscle from Tw2-CreERT2;R26-tdTO mice were co-stained with Pax7 (green) on day 7 after CTX. tdTO+ cells are negative for Pax7. Scale bar: 100 um. (**d**) Quantification of the number of Pax7+, Tw2+ (tdTO+) and Pax7+/Tw2+ double positive cells per field in Tw2-CreERT2; R26-tdTO on days 7 and 14 post-CTX injury. CTL: contralateral TA muscle. For each muscle section, at least 6 different fields were quantified and averaged. Data are mean ± S.E.M; N=3 mice.

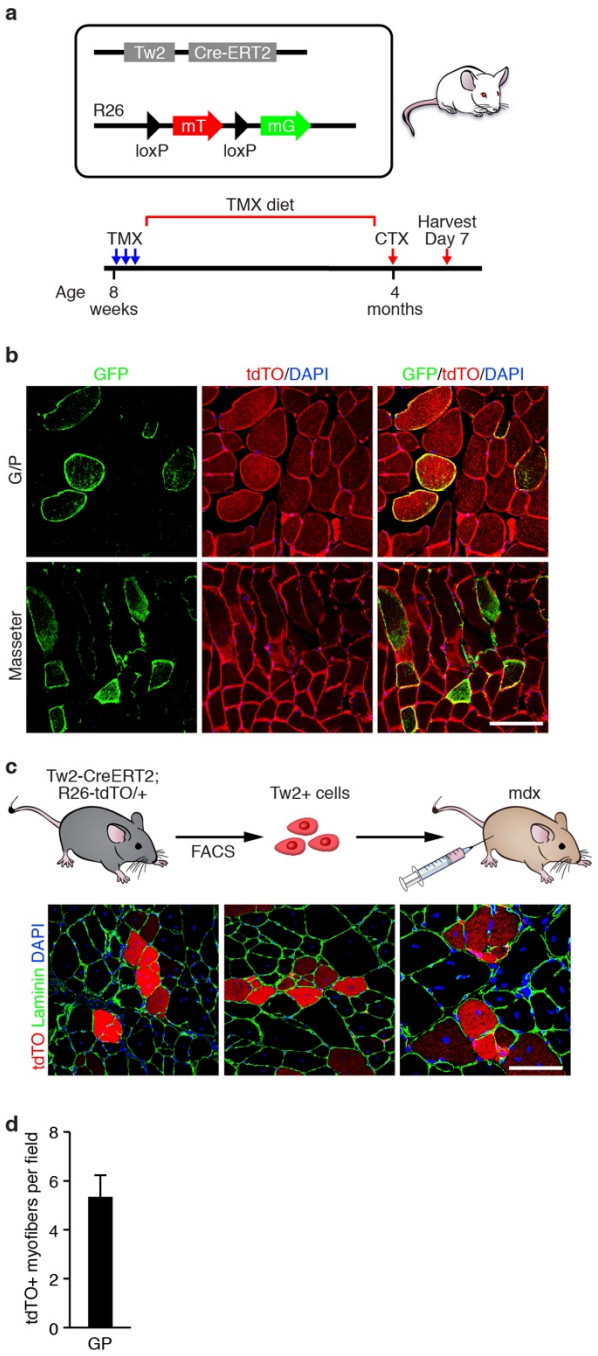
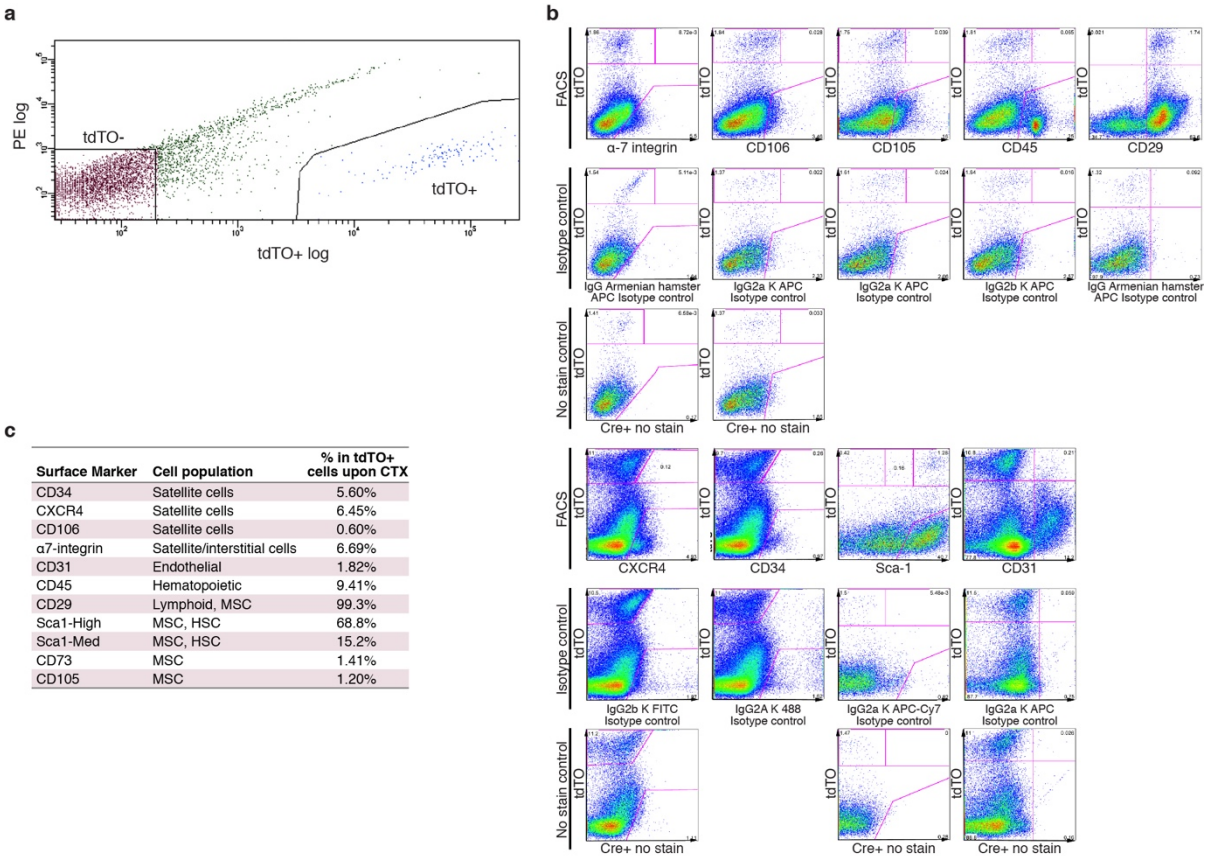


Figure 2-12. (**a**) Schematic of lineage tracing with Tw2-CreERT2; R26- mT/mG/+ mice. (**b**) Transverse-sections of G/P and masseter muscles from Tw2-CreERT2; R26-mT/mG mice at 4 months post-TMX. Tw2+ cells are labeled by GFP expression. All GFP+ myofibers remained TdTO+ in G/P and master, indicating Tw2+ cells fuse with existing myofibers. Scale bar: 100 um. (**c**) Freshly isolated Tw2+ cells can engraft and form myofibers when transplanted into the TA muscle of mdx mice. 60,000 freshly isolated Tw2+ cells from Tw2-CreERT2; R26-tdTO mice at 10 days post-TMX were injected into TA muscle of 4 month-old mdx mice, which were injected with CTX 1 day prior to engraftment. TA muscles were harvested at 4 weeks post-injection and stained with laminin (green) and DAPI (blue) to visualize engrafted tdTO+ myofibers. Scale bar: 100 um. (**d**) Quantification of the number of tdTO+ myofibers in the engraftment experiments. For each mouse, at least 6 fields of G/P muscle sections were quantified and averaged. Data are mean ± S.E.M; N=3 mice.



c

Surface Marker	Cell population	% in tdTO+ cells upon CTX
CD34	Satellite cells	5.60%
CXCR4	Satellite cells	6.45%
CD106	Satellite cells	0.60%
α7-integrin	Satellite/interstitial cells	6.69%
CD31	Endothelial	1.82%
CD45	Hematopoietic	9.41%
CD29	Lymphoid, MSC	99.3%
Sca1-High	MSC, HSC	68.8%
Sca1-Med	MSC, HSC	15.2%
CD73	MSC	1.41%
CD105	MSC	1.20%

Figure 2-13. (**a**) Representative FACS plots of Tw2+ (tdTO+) cells. Mononuclear cells from Tw2- CreERT2;R26-tdTO mice at 10 days post-TMX were sorted based on expression of tdTO. Approximately 3.3% of all mononuclear cells were

positive for tdTO+. Sorting gates were drawn as indicated for both tdTO+ and tdTO- cell populations. (**b**) FACS plots of cell surface marker expression of tdTO+ cells from the Tw2-CreERT2; R26-tdTO mice at 10 days post-TMX. (**c**) Expression of cell-surface markers in $Tw2^{+}$ cells following CTX injury.

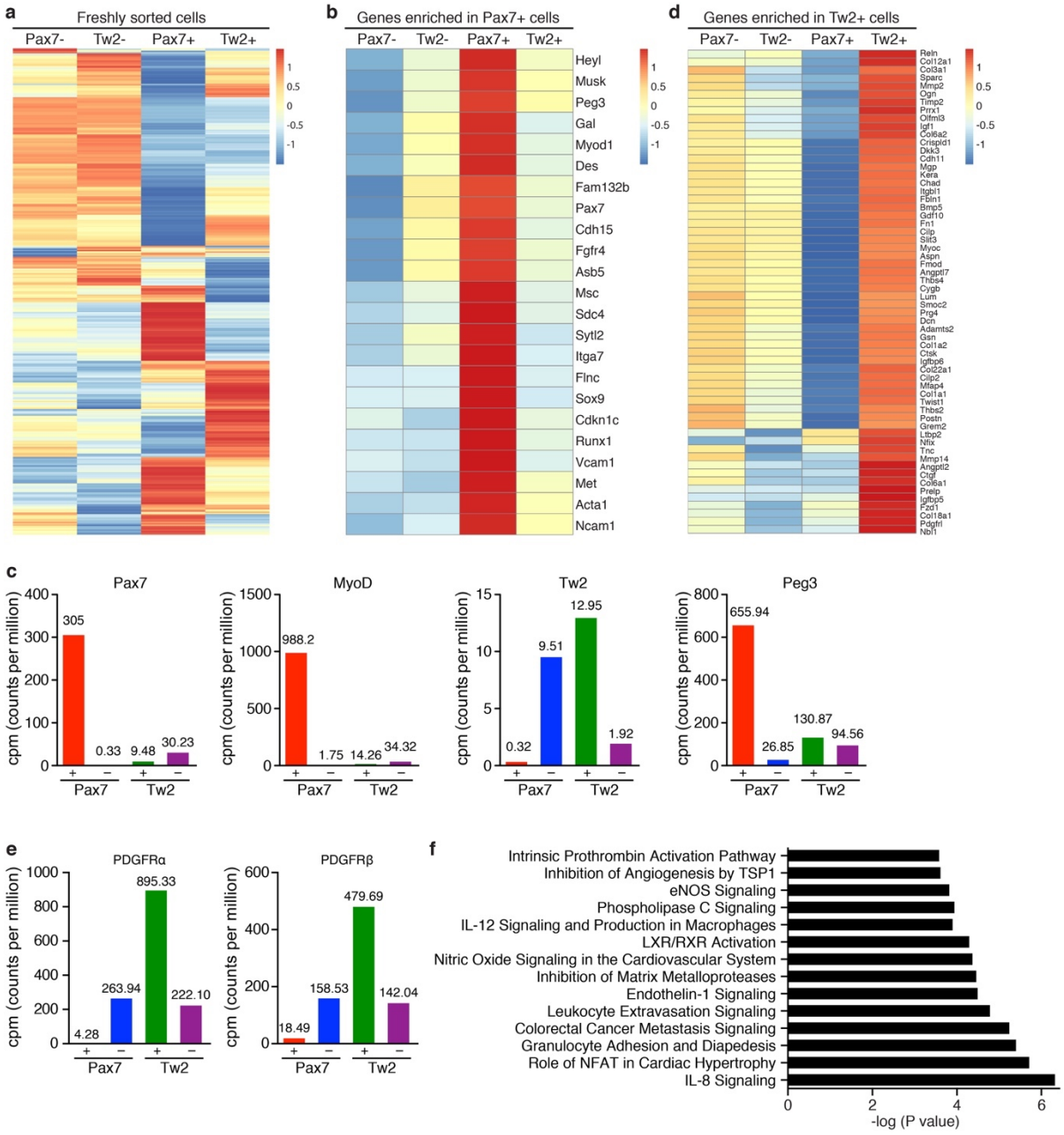


Figure 2-14. (a) Heat map of 4,980 genes expressed in freshly sorted Tw2^{+}, Tw2^{-} Pax7^{+} and Pax7^{-} cells identified by RNA-seq. Cells were isolated by FACS sorting from Tw2-CreERT2; R26-tdTO mice and Pax7-CreERT2; R26-tdTO mice at 10 days post-TMX, respectively. (b) Heat map of the top 23 genes enriched in Pax7^{+} cells compared with

Pax7^{-} cells. (c) Counts per million (cpm) of Pax7, MyoD, Tw2 and Peg3 expression by RNA-seq in Tw2^{+}, Tw2^{-}, Pax7^{+} and Pax7^{-}cells. (d) Heat map of the top 60 genes enriched in Tw2^{+} cells relative to Tw2^{-} cells. (e) Counts per million of PDGFRα and PDGFRβ expression by RNA-seq in Tw2^{+}, Tw2^{-}, Pax7^{+} and Pax7^{-} cells. (f) Pathways enriched in freshly sorted Tw2^{+} cells relative to Tw2^{-} cells identified by Ingenuity Pathway Analysis.

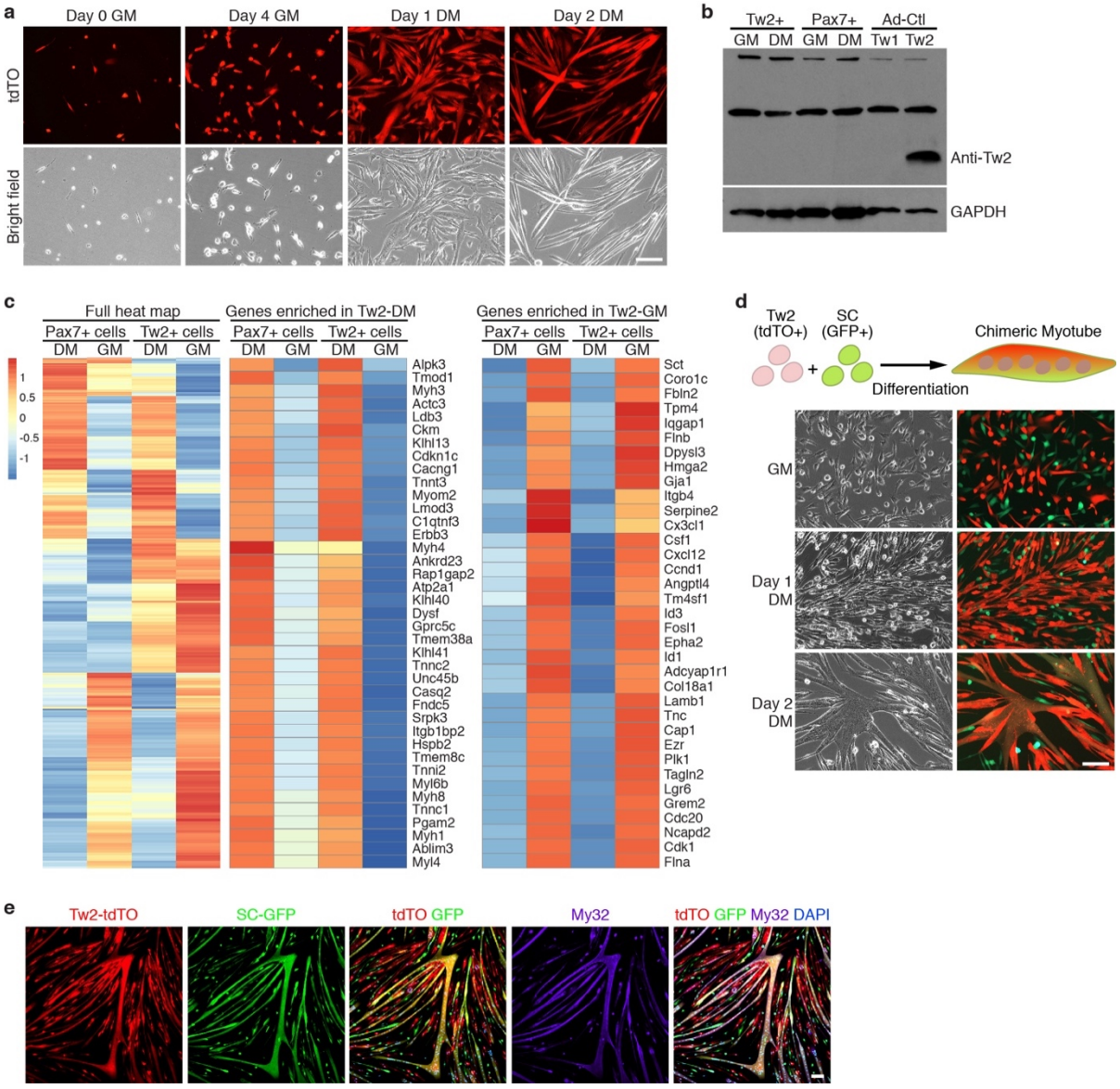


Figure 2-15. (**a**) Tw2+ cells isolated by FACS sorting proliferated efficiently in growth medium (GM) and differentiated into multinucleated myotubes in differentiation medium (DM). These cells remain tdTO+ in GM and DM. Bright field images are presented for comparison. Scale bar: 20 um. (**b**) Western blotting analysis for Tw2 protein of Tw2+ cells and Pax7+ cells in GM and DM. Ad-Ctl represents protein samples from neonatal rat

cardiomyocytes infected with adenoviruses expressing either Twist1 (Tw1) or Twist2 (Tw2). The upper band present in all samples represents a non-specific band. GAPDH protein is detected as loading control. (**c**) Heat map of genes expressed in Pax7+ cells and Tw2+ cells in GM and DM identified by RNA-seq analysis (left panel). Heat map of the top 39 genes enriched in Pax7-DM vs. Pax7-GM are shown in the middle panel, and the heat map of top 35 genes enriched in Pax7-GM vs. Pax7-DM are list on the right panel. Importantly, these genes showed the same trend of enrichment and repression in Tw2-DM vs. Tw2-GM samples. (**d**) Tw2+ cells and SCs can fuse with each other to form multinucleated myotubes. Tw2+ cells were isolated by FACS sorting from adult Tw2-CreERT2; R26-tdTO mice 10 days post-TMX, which are labeled by tdTO expression. SCs, which are labeled by GFP expression, were isolated by FACS sorting from CAG-eGFP mice. Equal numbers of Tw2+ cells and SCs were mixed and grown in GM, followed by differentiation in DM. Cells were visualized by direct fluorescence. (**e**) Myosin immunostaining (My32) revealed formation of multi-nucleated myotubes that express both GFP and tdTO. Scale bar: 100 um.

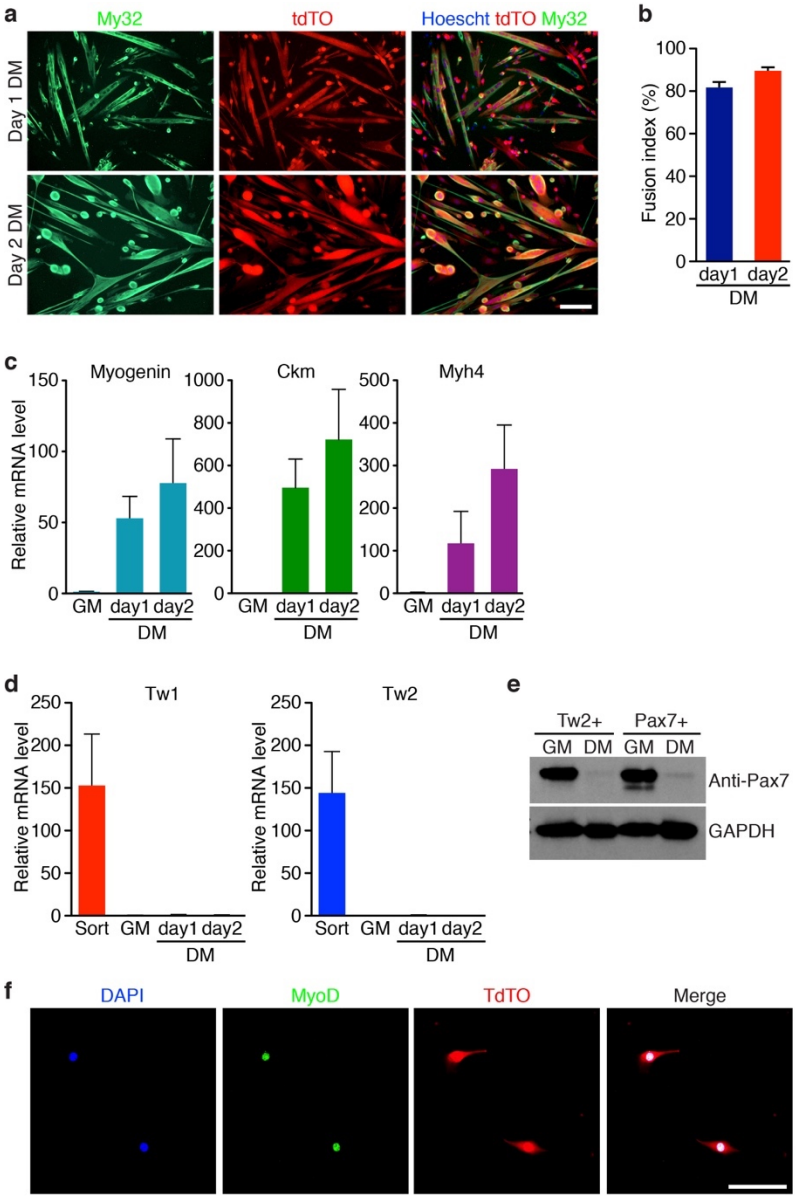


Figure 2-16. (**a**) Myosin staining (My32, green) revealed that the majority of Tw2$^+$ (tdTO$^+$) cells expressed myosin and formed multinucleated myotubes starting at 1 and 2 days in DM. Scale bar, 20 μm. (**b**) Fusion index was calculated as the percentage of tdTO$^+$ nuclei within multinucleated myotubes compared with the total number of tdTO$^+$ nuclei in the culture. Data are mean ± s.e.m.; $N = 3$ independent experiments. (**c**) Real-time RT–PCR revealed that

muscle genes such as *Myog*, *Ckm* and *Myh4* were strongly activated when $Tw2^+$ cells were cultured in DM. Values were normalized to those of the GM sample, which is set as 1. Data are mean ± s.e.m. $N = 4$ independent experiments. (**d**) Real-time RT–PCR revealed that *Tw1* and *Tw2* mRNAs were enriched in freshly sorted $Tw2^+$ cells (Sort). Expression was extinguished once cells were plated in culture. Values were normalized to those of the GM sample, which is set as 1. Data are mean ± s.e.m. $N = 4$ independent experiments. (**e**) Western blot analysis for Pax7 protein in $Tw2^+$ and $Pax7^+$ cells. Pax7 protein is expressed in both $Tw2^+$ and $Pax7^+$ cells in GM, but not in cells cultured for 2 days in DM. GAPDH protein was detected as loading control. (**f**) MyoD staining of $Tw2^+$ ($tdTO^+$) cells in GM revealed MyoD expression. Scale bar, 100 μm.

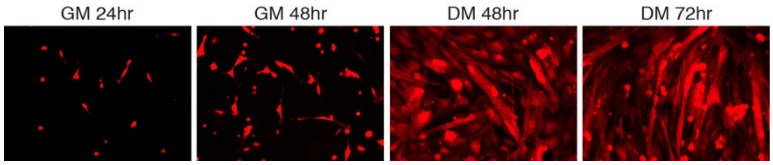


Figure 2-17. Growth and myogenesis of tdTO+/CD34- cells in culture. Freshly sorted tdTO+/CD34- cells were grown in GM for 48 hours before being switched to DM to induce myogenesis. Scale bar: 20 um.

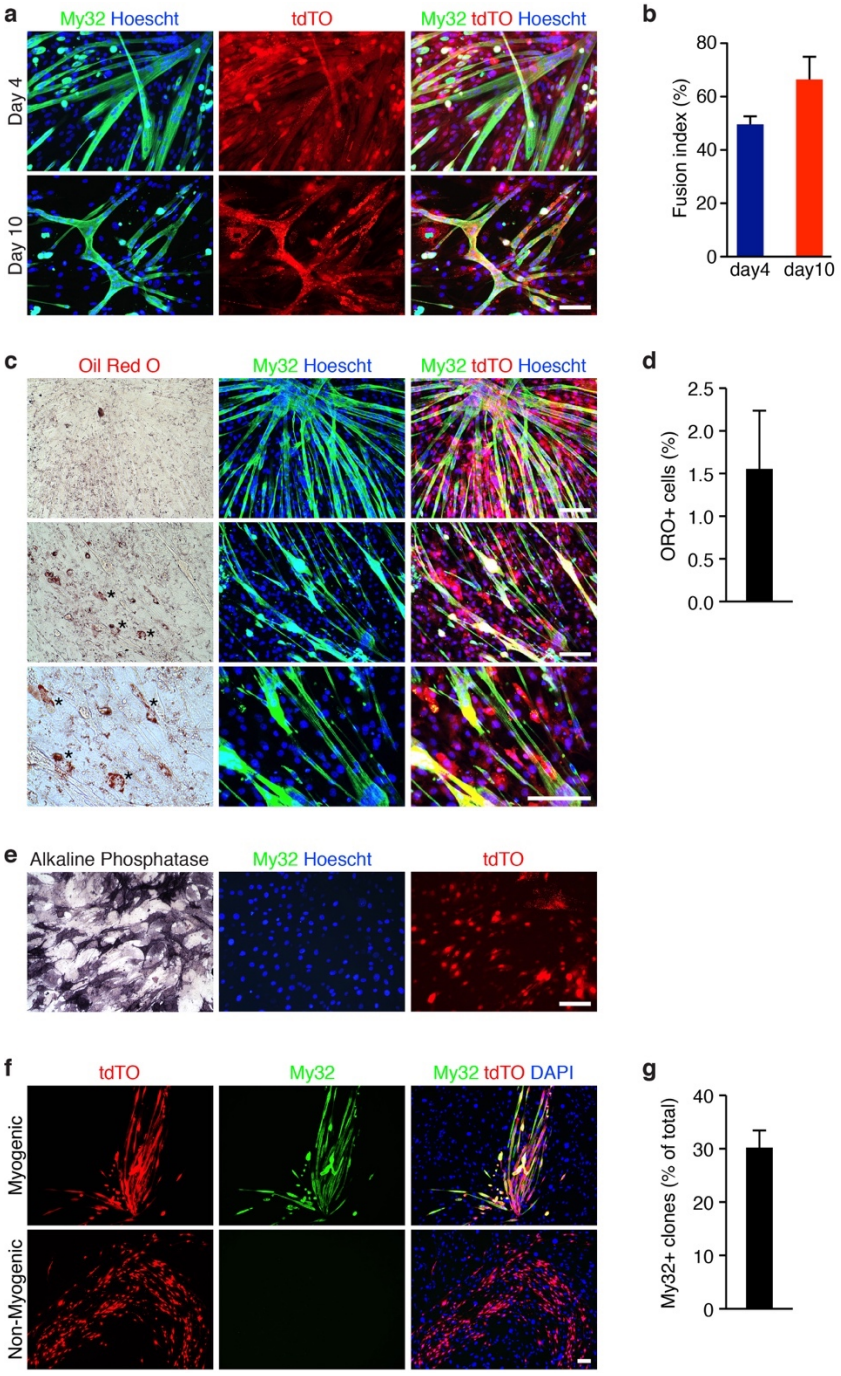


Figure 2-18. (**a**) Myosin staining (My32, green) revealed that tdTO^{+}/CD34^{-} cells expressed myosin and formed multinucleated myotubes after 4 and 10 days in DM. Freshly sorted tdTO^{+}/CD34^{-}cells were cultured in GM for 48 h before induction of myogenesis. Scale bar,

20 μm. (**b**) Fusion index was calculated as the percentage of $tdTO^+$ nuclei within multinucleated myotubes compared with the total number of $tdTO^+$ nuclei in the culture. Data are mean ± s.e.m. $N = 3$ independent experiments. (**c**) Oil Red O staining revealed that only very few $tdTO^+$ cells differentiated into adipocytes. Freshly sorted $tdTO^+/CD34^-$ cells were cultured in GM for 48 h before induction of adipogenesis for 10 days. Cells were stained with My32 and Hoechst before Oil Red O staining. Scale bars, 20 μm. (**d**) Percentage of Oil Red O-positive cells per field. Data are mean ± s.e.m. $N = 3$ independent experiments. For each experiment, more than 500 nuclei per field and a total of 3 fields were counted and averaged. (**e**) Alkaline phosphatase staining revealed that $tdTO^+/CD34^-$ cells can form osteoblasts when exposed to osteogenic medium. These cells did not express myosin. Scale bar, 20 μm. (**f**) Clonal analysis of $tdTO^+/CD34^-$ cells for myogenic potential. Single $tdTO^+/CD34^-$ clones were grown on inactivated MEF feeder layers and induced for myogenesis. Myogenic clones were identified by My32 staining. DAPI staining identified both $Tw2^+$ cells ($tdTO^+/CD34^-$) and MEFs. Scale bar, 100 μm. (**g**) Percentage of myogenic clones ($My32^+$) of the total surviving clones in the clonal analysis. Data are mean ± s.e.m. $N = 5$ independent experiments.

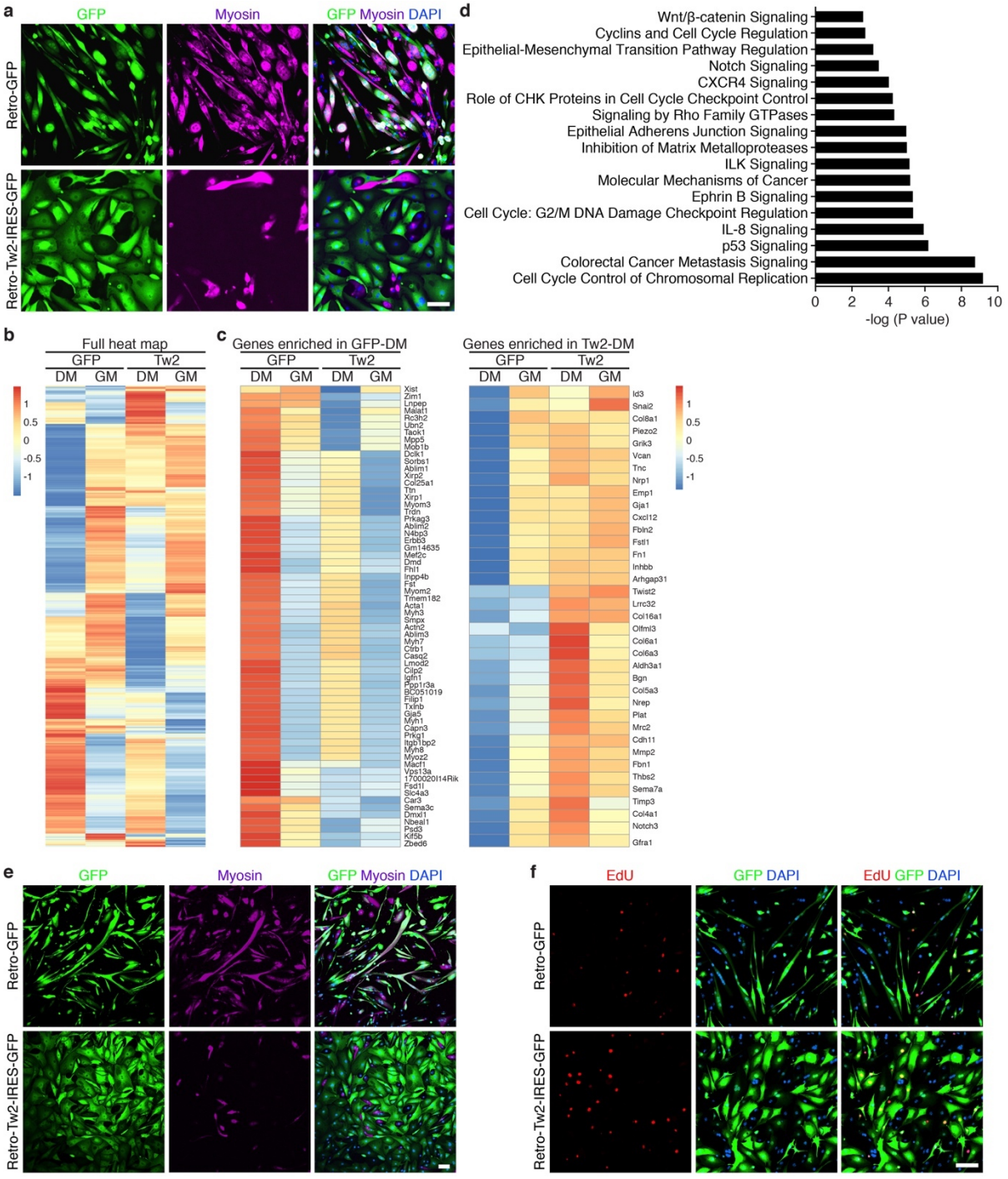


Figure 2-19. (**a**) Tw2^{+} cells were infected with retroviruses expressing either GFP (Retro-GFP) or Tw2 (Retro-Tw2-IRES-GFP) for 24 h before switching to DM. After 5 days in DM, My32 staining (purple) was performed to detect myotubes. Scale bar, 100 μm. (**b**) Heat map

of genes expressed in GFP- or Tw2-infected $Tw2^+$ cells in GM and DM identified by RNA-seq. (**c**) Heat maps of the top genes enriched and repressed in GFP-DM versus Tw2-DM. (**d**) Pathways of genes enriched in Tw2-DM relative to GFP-DM identified by Ingenuity Pathway Analysis. (**e**) Overexpression of Tw2 inhibits myotube formation of $Pax7^+$ cells. $Pax7^+$ cells were infected with Retro-GFP or Retro-Tw2-IRES-GFP for 24 h before switching to DM. After 5 days in DM, My32 staining (purple) was performed to detect myotubes. Scale bar, 100 μm. (**f**) EdU labelling revealed increased DNA synthesis in Tw2-infected $Pax7^+$ cells in DM. Infected $Pax7^+$cells were treated with EdU (10 μM) for 24 h in DM before staining. Scale bar, 100 μm.

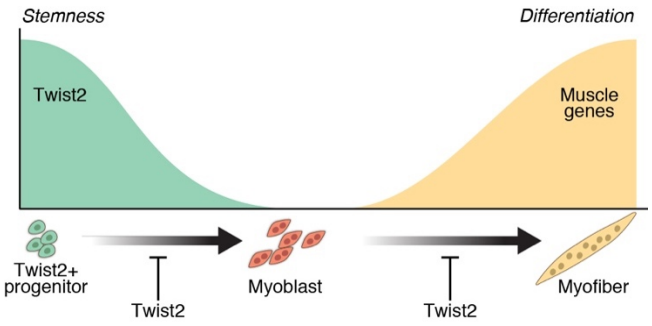


Figure 2-20. Model of Tw2 maintains stemness and blocks myogenesis. Pax7 expression is not detectable in Tw2+ cells in vivo. However, when removed from their native milieu, Tw2+ cells rapidly down-regulate Twist expression and enter a Pax7+ state enroute to a myogenic pathway.

Tables

Surface Marker	Cell population	% in tdTO+ cells
CD34	Satellite cells	2.61%
CXCR4	Satellite cells	1.20%
CD106	Satellite cells	1.67%
α7-integrin	Satellite/interstitial cells	0.46%
CD31	Endothelial	1.90%
CD45	Hematopoietic	3.46%
CD29	Lymphoid, MSC	98.80%
Sca1-High	MSC, HSC	68.80%
Sca1-Med	MSC, HSC	8.60%
CD73	MSC	4.23%
CD105	MSC	2.17%

Table 1. Enrichment of surface markers on isolated $Tw2^{+}$ progenitors from hind limb muscle.

CHAPTER THREE

NRP1-SEMA3A SIGNALING MEDIATES TYPE IIB MYOFIBER SPECIFICITY OF TWIST2^{+} CELLS

Abstract

Skeletal myofibers are classified into distinct subtypes based on their expression of specific myosin heavy chain proteins (type I, IIa, IIb, and IId/x), metabolic profile (oxidative or glycolytic), and speed of contraction (slow- or fast- twitch). We previously identified a unique population of interstitial muscle progenitors, marked by expression of Twist2, which fuse specifically to type IIb fast-twitch myofibers. Twist2^{+} progenitors are distinct from satellite cells, a muscle progenitor that expresses Pax7 and contributes to all myofiber types. To explore the mechanistic basis of the type IIb fiber type specificity of Twist2^{+} cells, we compared the gene expression profiles of Twist2^{+} and Pax7^{+} myogenic progenitors and identified the membrane receptor Neuropilin1 (Nrp1) as a marker of Twist2^{+} cells. We also found that Sema3a, a chemorepellent ligand for Nrp1, is expressed by type I and IIa myofibers, but not IIb myofibers. Using a stripe migration assay, we show that Twist2^{+} cells preferentially avoid Sema3a stripes and that Nrp1 is both necessary and sufficient for this effect. In a chimeric cell-cell fusion assay, Nrp1-Sema3a signaling impaired fusion of Twist2^{+} cells to Sema3a-expressing fibers. Similarly, forced expression of Sema3a in type IIb myofibers in mice prevented fusion with Twist2^{+} cells. Our findings reveal a previously

unrecognized signaling mechanism whereby a cell surface receptor for a chemorepellent confers specificity of intercellular fusion of a specific muscle progenitor with its target tissue.

Introduction

Skeletal muscles display a range of metabolic and contractile phenotypes that enable the diversity of muscle functions required for life. Slow-twitch or type I myofibers exhibit oxidative metabolism and are fatigue-resistant, whereas fast-twitch or type II myofibers are more glycolytic, display rapid bursts of contraction, and fatigue rapidly (Schiaffino and Reggiani 2011). Myofibers can be further classified as type I, IIa, IIx/d or IIb depending on the myosin isoforms they express (Schiaffino and Reggiani 2011). Fiber type specificity is initially established during fetal life, but can be modulated after birth by activity, hormonal influences and disease (Schiaffino and Reggiani 2011).

In response to injury and aging, skeletal muscle is among the most regenerative tissues, owing to the presence of satellite cells, a resident stem cell population that resides beneath the muscle basal lamina (Sambasivan et al. 2011; Lepper et al. 2009). Satellite cells contribute to all myofiber types without fiber type specificity (Pawlikowski et al. 2015; Liu et al. 2017). Recently, through lineage tracing of cells expressing the basic-Helix-Loop-Helix (bHLH) transcription factor Twist2, we discovered a distinct skeletal muscle progenitor (Liu et al. 2017). Twist2^{+} progenitors reside in the muscle interstitium and are anatomically and transcriptionally distinct from Pax7^{+} satellite cells, which reside beneath the muscle basal lamina(Liu et al. 2017). Moreover, in contrast to satellite cells, Twist2^{+} cells fuse only with type IIb/x myofibers and genetic ablation of Twist2^{+} cells in mice results in specific atrophy

of type IIb myofibers (Liu et al. 2017). Twist2^{+} cells represent the only known fiber-type specific muscle progenitor (Liu et al. 2017). Twist2 expression maintains these cells in an undifferentiated, mesenchymal state and, when cultured in vitro, Twist2^{+} cells rapidly down-regulate Twist2 expression and differentiate into multi-nucleated myotubes (Liu et al. 2017). Ectopic expression of Twist2 globally represses the myogenic program and activates genes involved in cellular migration and matrix degradation(Liu et al. 2017). The findings highlight the unique roles of Twist2 in regulating myogenesis and invasiveness.

In this study, we investigated the mechanistic basis of the specificity of Twist2^{+} progenitors in the formation of type IIb myofibers. Through RNA-Sequencing and ChIP-sequencing, we identified the Neuropilin1 (Nrp1) gene as a direct transcriptional target gene of Twist2 and its expression is enriched in Twist2^{+} cells compared to Pax7^{+} satellite cells. Neuropilins act as receptors for Semaphorins, a class of membrane-bound and secreted proteins that regulate cellular migration, differentiation, and function (Raper 2000; Sharma et al. 2012). There are 5 classes of semaphorins: classes 4-7 are membrane-bound, while class 3 semaphorins are secreted. Semaphorin3a (Sema3a) belongs to the class 3 semaphorin family and has well-known roles in cancer where it regulates angiogenesis by inhibiting integrin activity (Raper 2000; Sharma et al. 2012; Alto and Terman 2017; Tamagnone and Comoglio 2000). The interaction of Sema3a with Nrp1 and a Plexin co-receptor triggers a G-protein coupled cascade, resulting in activation of cytoskeletal modifying proteins such as Rac and collapsin response mediator proteins (CRMP) (Koncina et al. 2007; Nasarre et al. 2014). This results in destabilization and depolymerization of microtubule networks and chemorepulsion from the Sema3a source (Kumanogoh and Kikutani 2013). We show that Sema3a is

expressed only on type I and IIa, but not type IIb myofibers. In vitro assays reveal selective avoidance of Sema3a by Twist2$^+$ cells but not Pax7$^+$ satellite cells, which is due to the differential expression of Nrp1. Employing a chimeric fusion assay, we show that the Sema3a/Nrp1 interaction is sufficient to prevent cell-cell fusion between cells expressing the Sema3a ligand and the Nrp1 receptor, respectively. Finally, over-expression of Sema3a in type II myofibers in mice is sufficient to prevent Twist2$^+$ cells from fusing to type IIb myofibers in vivo. These findings uncover an intercellular signaling mechanism whereby Twist2 controls expression of the Nrp1 receptor in Twist2$^+$ cells, which confers myofiber specificity of Twist2$^+$ cell fusion by sensing the presence or absence of its ligand Sema3a.

Results

Nrp1 is enriched in Twist2$^+$ cells and is a Twist2 target gene.

To explore the mechanism underlying the fiber-type specificity of Twist2$^+$ cells, we compared gene expression profiles of Twist2$^+$ and Pax7$^+$ cells freshly isolated from adult mouse skeletal muscle(Liu et al. 2017). Among various genes that were differentially expressed between these two cell populations, we found that Nrp1 mRNA was highly enriched in Twist2$^+$ cells compared to Pax7$^+$ satellite cells (Figure 3-1A). Analysis of single-cell RNA sequencing data from the Tabula Muris Consortium also revealed that Nrp1 was expressed by Twist2$^+$ cells but not by satellite cells (Figure 3-2A-D)(Tabula Muris Consortium et al. 2018). We also confirmed by western blot that Nrp1 protein was enriched in Twist2$^+$ cells compared to Pax7-derived myoblasts (Figure 3-1B).

Overexpression of Twist2 in Twist2$^+$ cells resulted in upregulation of Nrp1 expression, suggesting that *Nrp1* may be a transcriptional target of Twist2 (Figure 3-1C) Consistent with this notion, ChIP-Seq analysis of Twist2 binding in Twist2$^+$ cells showed enrichment at the *Nrp1* promoter in both growth media (GM) and differentiation media (DM) (Figure 3-1D).

To determine the localization of Nrp1-expressing cells within muscle tissue, we performed immunofluorescent staining of mouse gastrocnemius muscle with an anti-Nrp1 antibody. Co-staining of Nrp1 with wheat germ agglutinin (WGA) revealed that the majority of Nrp1-expressing cells are interstitial cells (Figure 3-2E). This is similar to the location of Twist2$^+$ cells, but not that of Pax7$^+$ cells, which are located beneath the muscle basal lamina. Co-staining of Nrp1 and Twist2 revealed near complete co-localization (Figure 3-1E). Together, these results show that Nrp1 is preferentially expressed in Twist2$^+$ cells but not Pax7 cells, likely due to direct regulation by Twist2.

*Sema3a is enriched in Type I and IIa myofibers***.**

Given that Twist2$^+$ cells are enriched for Nrp1, we sought to determine where its chemorepellent ligand, Sema3a, was expressed in muscle. Co-immunostaining of soleus and gastrocnemius muscle sections for Sema3a and markers of various muscle fiber types revealed Sema3a expression exclusively on type I and IIa myofibers, but not type IIb myofibers (Figure 3-3A). Additionally, Sema3a expression and tdTO expression labeled by Twist2-lineage tracing are mutually exclusive in gastrocnemius muscle, with no co-localization (Figure 3-3B). These findings suggest that Nrp1-expressing Twist2$^+$ cells are

repelled from Sema3a-expressing type I and IIa fibers. Interestingly, Twist2$^+$ progenitors do not fuse to tongue muscle, despite the presence of type IIb fibers[5]. Co-staining of type IIb or fast myosin (My32) with Sema3a in the tongue revealed that type IIb fibers of tongue muscle express Sema3a, which may explain the lack of Twist2$^+$ cell fusion (Figure 3-3C; Figure 3-4). Together, these results suggest that the Nrp1 and Sema3a signaling axis may confer type IIb fiber-type specificity to Twist2$^+$ cells.

Twist2$^+$ cells and Pax7$^+$ cells differentially respond to Sema3a.

In order to validate that Nrp1 expression in Twist2$^+$ cells is capable of mediating repulsion from Sema3a, we performed Sema3a stripe migration assays. We generated stripes of recombinant Sema3a on tissue culture dishes and seeded various cell types onto the stripes overnight before fixing and imaging (Figure 3-5A)(Yamagishi et al. 2016). When we seeded Twist2$^+$ cells onto the stripes, we found robust avoidance of Sema3a stripes (Figure 3-5B and C). However, Pax7$^+$ satellite cells showed no avoidance of Sema3a stripes, indicating that Twist2$^+$ cells and Pax7$^+$ cells differentially respond to Sema3a (Figure 3-5B and C). We also performed the stripe assay with other cell types, including C2C12 myoblasts, 10T1/2 fibroblasts, mouse embryonic fibroblasts (MEFs), and Cos-7 cells, and found no avoidance of these cells from Sema3a stripes (Figure 3-6A and B).

In order to determine if the repulsion of Twist2$^+$ cells by Sema3a was sufficient to alter myotube formation, we differentiated Twist2$^+$ cells on the Sema3a stripes (Figure 3-5D). We found that Twist2$^+$ cells preferred to form myofibers away from the Sema3a-stripes

(Figure 3-5D). These results suggest that Twist2$^+$ cells are highly responsive to Sema3a-mediated chemorepulsion in both growth and differentiation conditions.

Nrp1 is necessary and sufficient for repulsion from Sema3a stripes.

In order to validate the role of Nrp1 in mediating cellular avoidance of Sema3a stripes, we altered levels of Nrp1 expression in Twist2$^+$ and Pax7$^+$ cells. Using a retroviral construct, we overexpressed Nrp1 in Twist2$^+$ cells and seeded those cells onto Sema3a stripes (Figure 3-7A and Figure 3-8). We found that overexpression of Nrp1 in Twist2$^+$ cells resulted in an ~10% increase of Sema3a avoidance compared to control infected Twist2$^+$ cells (Figure 3-7A and B). When we overexpressed Nrp1 in Pax7$^+$ cells, we found that Pax7$^+$ cells acquired repulsiveness to the Sema3a stripes to a similar extent to Twist2$^+$ cells (Figure 3-7C and D). This suggests that Nrp1 expression is sufficient to drive the repulsion from Sema3a.

Next, we investigated whether Nrp1 was required for Twist2$^+$ cells to avoid Sema3a stripes by reducing Nrp1 expression through shRNA-mediated knockdown and CRISPR/Cas9-mediated knockout. We first tested knockdown of Nrp1 mRNA in Twist2$^+$ cells using 4 different retroviral shRNA constructs. Two candidates, shNrp1-1 and shNrp1-2, yielded the greatest degree of knockdown and were used for the stripe assay (Figure 3-9A). We infected Twist2$^+$ cells with retroviral vectors containing shNrp1-1, shNrp1-2, or a control shRNA and seeded them onto Sema3a stripes (Figure 3-10A). We found that knockdown of Nrp1 significantly decreased Sema3a avoidance, suggesting that Nrp1 is necessary for Sema3a avoidance in Twist2$^+$ cells (Figure 3-10A and B).

In order to corroborate the knockdown data, we deleted the Nrp1 gene in Twist2$^+$ cells using CRISPR/Cas9. To target the Nrp1 gene, we two short guide RNAs (sgRNAs) targeting exon 2 (Figure 3-9B). We packaged these guides in lentivirus and infected Twist2$^+$ cells in order to knockout Nrp1. We validated the cutting efficiency of the sgRNAs in Twist2$^+$ cells using a T7E1 assay (Figure 3-9C). Correct targeting by sgRNAs was confirmed by TOPO cloning and sequencing (Figure 3-9D). For the majority of the guides, we identified indels resulting in frameshift mutations (Figure 3-9D). Western blot of Nrp1 revealed that the majority of Nrp1 protein was eliminated, confirming efficient deletion of Nrp1 by CRISPR/Cas9 (Figure 3-10C).

Next, we seeded sgNrp1-infected Twist2$^+$ cells onto Sema3a stripes and performed a stripe migration assay (Figure 3-10D). We found that single guide-mediated deletion of Nrp1 in Twist2$^+$ cells was sufficient to abolish the majority of Sema3a avoidance, providing further evidence that Nrp1 is required for Twist2$^+$ cells to avoid Sema3a (Figure 3-10E).

The Sema3a/Nrp1 signaling axis represses chimeric myotube formation.

We previously showed that Twist2$^+$ cells could fuse to Pax7$^+$ cells to form chimeric myotubes (Liu et al. 2017). To explore whether the Sema3a/Nrp1 signaling pathway could keep Twist2$^+$ cells from fusing to Sema3a-expressing fibers, we performed a chimeric fusion assay between Twist2$^+$ cells and primary myoblasts (Figure 3-11A). Twist2$^+$ cells were infected with either Nrp1, shNrp1-2, or a control empty vector, and primary myoblasts were infected with a Sema3a construct containing an EGFP reporter (Figure 3-11A and Figure 3-12A). We first differentiated these cells individually and found that alteration of Nrp1 or

Sema3a levels did not impair myotube formation (Figure 3-12B and C). We then mixed each population of Twist2^{+} cells with Sema3a-overexpressing primary myoblasts and allowed them to differentiate for 7 days (Figure 3-11A). Our results showed that overexpression of Nrp1 significantly decreased the number of nuclei within chimeric myofibers, whereas knockdown of Nrp1 significantly increased the number of chimeric nuclei between Twist2^{+} and primary myoblast myotubes (Figure 3-11B and C; Figure 3-12D). These data suggest that Sema3a-Nrp1 signaling is sufficient to impair Twist2^{+} cell fusion into Sema3a-expressing myofibers.

Overexpression of Sema3a in muscle prevents Twist2^{+} cell contribution to type IIb fibers.

In order to study the contribution of Nrp1-Sema3a signaling to Twist2^{+} cell fiber type specificity in vivo, we generated a Sema3a transgenic mouse using the muscle creatine kinase (MCK) promoter (Figure 3-13A). This promoter is preferentially active in fast-twitch myofibers relative to slow-twitch fibers, enabling the expression of Sema3a in fibers where it is normally absent (Dunant et al. 2003; Tai et al. 2011; Yamashita and Yoshioka 1991). We verified Sema3a overexpression by immunostaining of gastrocnemius muscle and real-time PCR (Figure 3-13B; Figure 3-14A). We then crossed the MCK-Sema3a transgenic line with a Tw2-CreERT2; tdTO mouse to generate MCK-Sema3a; Tw2-CreERT2; tdTO mice (Figure 3-13A). These mice were injected with 3 doses of tamoxifen starting at 8 weeks of age to begin lineage tracing of Twist2^{+} cells (Figure 3-13A). Eight weeks after the first injection, we observed a significant decrease in the number of tdTO^{+} type IIb myofibers in the MCK-

Sema3a transgenic mice compared to controls, indicating that Sema3a can prevent Twist2$^+$ cells from fusing to type IIb myofibers in vivo (Figure 3-13C and D).

Discussion

We previously identified a population of Twist2$^+$ progenitors in the muscle interstitium and found that they fuse exclusively to type IIb/x myofibers (Liu et al. 2017). The results of the present study reveal a unique mechanism for fiber-type specificity of these Twist2$^+$ progenitors. Through transcriptional profiling, we found that the membrane receptor Nrp1 is differentially enriched in Twist2$^+$ cells compared to Pax7$^+$ satellite cells, reflecting the direct activation of the Nrp1 gene by Twist2. Sema3a, the chemorepulsive ligand for the Nrp1 receptor, is localized exclusively to type I and IIa myofibers, which do not fuse with Twist2$^+$ cells, and is absent from type IIb fibers, which recruit Twist2$^+$ cells. These findings suggest that Twist2$^+$ cells are repelled from the type I and IIa myofibers that are enriched for Sema3a expression, but are able to fuse to type IIb fibers lacking Sema3a (Figure 3-15). Since Pax7$^+$ satellite cells are not enriched for Nrp1, they are not repelled from type I and IIa fibers and therefore show no fiber-type specificity.

In vitro stripe migration assays confirmed that manipulating Nrp1 expression altered the ability of Twist2$^+$ cells and Pax7$^+$ cells to respond to Sema3a. Increasing Nrp1 expression enhanced Sema3a avoidance of both Twist2$^+$ and Pax7$^+$ cells. Additionally, ablation of Nrp1 expression in Twist2$^+$ cells in vitro by either RNAi-mediated knockdown or CRISPR-mediated knockout resulted in loss of Sema3a avoidance. These findings suggest that Nrp1 is both necessary and sufficient for this interaction. We noticed in vitro that Twist2$^+$ cells

overexpressing Nrp1 were often more compact and smaller in the presence of Sema3a-expressing satellite cells. This is likely a manifestation of a Sema3a signaling gradient surrounding Twist2^{+} cells causing cytoskeletal reorganization on all sides. In vivo, the formation of a muscle secreted-Sema3a gradient may explain why Twist2^{+} cells are present in slow-twitch muscles such as the soleus but are unable to contribute to the myofibers. Whether Sema3a secreted from type I and IIa myofibers traverses across the muscle basement membrane or whether Twist2^{+} cells must first reach the muscle compartment before being repelled remains to be answered. However, responsiveness of myofibers and satellite cells to vascular cues suggests that Sema3a is likely capable of diffusing across the basal lamina (Wosczyna and Rando 2018). The repulsive effects of Sema3a on Nrp1-expressing Twist2^{+} cells are reminiscent of the role of the Sema3a-Nrp1 signaling axis in the control of motor neuron migration during development and endothelial cell signaling during angiogenesis (Saller et al. 2016; Acevedo et al. 2008).

The Nrp1 family member Nrp2 is also highly enriched in Twist2^{+} cells and upregulated upon Twist2-overexpression. Nrp1 and 2 confer specificity of Sema3a binding and signaling with Plexin co-receptors. Binding of Sema3a to this co-receptor complex triggers a G-protein coupled cascade, resulting in cytoskeletal and growth cone collapse (Tamagnone and Comoglio 2000). While Neuropilins are the primary determinants of Sema3a recognition, other Nrp1 co-receptors, such Lcam1, Chl1, and Robo1 have also been shown to interact with Nrp1 to mediate downstream cytoskeletal reorganization (Alto and Terman 2017). Which of these co-receptors are key for Nrp1 signaling in Twist2^{+} cells and

whether their differential expression may have important biological function in muscle remains to be explored.

The specification of different muscle fiber types and specialization of function is an important aspect of vertebrate evolution. The specific atrophy of type IIb fibers during aging and other disease suggests differential signaling mechanisms maintain structure and function of individual fiber-types (Lee et al. 2016). Our identification of Twist2^{+} cells and their unique ability to fuse with type IIb/x fibers positions them as potential regulators of fast-twitch fiber homeostasis (Liu et al. 2017). In particular, the larger size of type IIb fibers suggests additional mechanisms are required to help maintain their cross-sectional area throughout life (Schiaffino and Reggiani 2011). During aging, atrophy of fast-twitch fibers could be partially due to impaired Twist2^{+} cell function or a loss in Twist2^{+} cell number. This is supported by our finding that type IIb fiber-specific atrophy is seen after ablation of Twist2^{+} cells (Liu et al. 2017). Whether Twist2^{+} cell number or function are impaired over time is an area of future study. Additionally, it is interesting to consider why a mechanism of fiber-type specificity revolving around the repulsion of Twist2^{+} cell fusion would be evolutionarily maintained. It would be logical to conclude that additional fusion of myogenic cells to type I and IIa fibers would actually be beneficial. A likely explanation is that Sema3a secreted by type I and IIa fibers may play additional important roles in regulating fiber-type specific motor neuron innervation or vascular remodeling (Anderson et al. 2016). In this context, the gain of Twist2^{+} cell fusion may not compensate for other benefits imparted by Sema3a signaling. It is also possible that Twist2^{+} cell fusion may actually alter the

transcriptional or metabolic profile of fibers. The evolution of Twist2^{+} cells and their fiber-type specificity will be an exciting area of future study.

Overall, our study not only reiterates our findings of Twist2^{+} cells as fiber-type specific muscle progenitors, but also provides a novel mechanistic basis for this fiber-type specificity. The idea of muscle progenitor cells adopting pathways traditionally involved in the nervous and vascular systems offers a new class of signaling molecules to analyze in the context of muscle progenitor biology. Given the shared developmental origin of muscle and nervous tissues, it is unsurprising that shared pathways exist that regulate key aspects of their function. Recent studies have shown that Sema3a secreted from satellite cells may regulate fiber-type specification during muscle injury[22-26]. Additionally, other guidance molecules such as Ephrins have also been implicated in muscle fiber identity, motor neuron innervation, and satellite cell fusion (Stark et al. 2015; 2011). It will be interesting to see whether these systems are differentially expressed within satellite cells or other muscle progenitors where they may mark additional fiber-type specific populations or sub-populations.

Materials and Methods

RNA and ChIP sequencing

Bulk RNA-seq data was obtained from our previously published study and is available in the Gene Expression Omnibus (GEO) under accession codes GSE84377, GSE84378, GSE84379, and GSE 84380(Liu et al. 2017).

Single Cell RNA-Sequencing data was pulled from Tabula Muris (https://tabula-muris.ds.czbiohub.org).

ChIP-Sequencing data was obtained from our previously published study and is available in the Gene Expression Omnibus (GEO) under accession codes (GSE127988).

Western Blot

Twist2^{+} and Pax7^{+} -derived myoblasts were harvested in RIPA buffer (Sigma) supplemented with cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma) phosphatase inhibitor. Western blot was performed as previously described(Makarewich et al. 2018). Primary antibodies used were: Nrp1 (Abcam, ab81321, 1:1000) and GAPDH (Millipore, CB1001, 1:10,000). Western blots were washed in TBST, incubated with HRP-conjugated secondary antibodies (Bio-Rad), and then developed using a ChemiDoc MP Imaging System (Bio-Rad).

Immunohistochemistry

Immunohistochemistry of skeletal muscle was performed as previously described(Liu et al. 2017). For immunohistochemistry of fixed frozen sections, the following antibodies were used: Twist2 (Abcam, ab66031, 1:200), Sema3a (Abcam, ab23393, 1:100), Nrp1 (Abcam, ab81321, 1:100), and Myh4 (Invitrogen, 14-6503-82, 1:100). For immunohistochemistry of raw embedded frozen sections, the following antibodies were used: Twist2 (Abcam, ab66031, 1:200), Sema3a (Abcam, ab23393, 1:100), Nrp1 (Abcam, ab81321, 1:100), myosin IIa (Developmental Studies Hybridoma Bank, SC-71, 1:10), myosin IIb (Developmental Studies Hybridoma Bank, BF-F3, 1:10), Slow myosin (Sigma-Aldrich, NOQ7.5D, 1:250), and Fast myosin (Sigma-Aldrich, clone My32, 1:250). Alexa

Fluor secondary antibodies were used according to the manufacturer's instructions. TdTO signals were detected by direct fluorescent imaging. Wheat germ agglutinin staining was performed on both frozen and paraffin-embedded sections, using WGA-Alexa Fluor 555 (W32464) or WGA-Alexa 647 conjugate (W32466) (Life Technologies, 50 mg ml^{-1}). Images were taken on a Zeiss LSM700 confocal microscope. Muscle fiber numbers were quantified by ImageJ.

Immunostaining of cultured cells

Immunostaining cultured cells was performed as previously described(Liu et al. 2017). Primary antibodies include: fast myosin (Sigma-Aldrich, My32, 1:250). Cells were counterstained with Hoechst (ThermoFisher, Hoechst, 3342, 1:1000). Alexa Fluor secondary antibodies were used according to the manufacturer's instructions.

Stripe Migration Assay

The protocol used for the *in vitro* stripe migration assay was adapted from a previous study(Yamagishi et al. 2016). In summary, 2.5 µg of recombinant Sema3a-Fc (R&D Systems, 5926-S3-025) was mixed with 7.5 µg of goat anti-mouse IgG Alexa 488 secondary antibody (Abcam, ab150113) in 50 µl PBS and incubated at room temperature for 30 minutes. The mixture was then injected into silicon matrices (Purchased from Martin Bastemeyer) attached to tissue culture dish using an insulin needle. The dishes were then incubated at 37 ^{o}C in a cell culture incubator. The stripes were then washed three times with PBS followed by addition of 10 µg of control Fc fragments (Invitrogen, 31205) in 200 µl of

PBS on top of the stripes as a counter coating for 30 mins at 37 ^{o}C. The dish was washed three times with PBS and then coated with Matrigel (Corning, 356234) for 1 hr at 37 ^{o}C. The dish was then washed three times with PBS followed by seeding of cells on top. After 24 hrs, dishes were fixed and counter-stained with Hoechst in order to count the number of cells on and off the stripes.

Cell Culture

Cell culture was performed as previously described(Liu et al. 2017). In brief, Twist2^{+} cells, Pax7^{+} cells, and primary myoblasts were cultured on Matrigel-coated plates in satellite cell growth medium (SCGM; F10 Ham's, 20% FBS, 0.2% Primocin, and 2.5 ng µl^{-1} basic Fibroblast Growth Factor (Gibco)). For myoblast differentiation, we used differentiation media (DM) containing: DMEM, 2% horse serum, 0.2% Primocin. For the culture of all other cells, we used standard growth media (GM) containing: DMEM, 10% FBS, Pen/Strep.

Real-time RT–PCR analysis

Total RNA was extracted from cultured cells and muscle tissue with Trizol (Invitrogen, 15596026) following the manufacturer's instructions. From this RNA, cDNA was synthesized using iScript Reverse Reverse Transcriptase Supermix (Bio-Rad, 1708840). Selected gene expression was analysed by real-time RT–PCR. For Nrp1 expression, Taqman probe was used (Mm00435379_m1). For Sema3a expression SYBR green was used with the following primers: Fwd – 5' -GAAGAGCCCTTATGATCCCAAAC-3'; Rev – 5'-AGATAGCGAAGTCCCGTCCC-3'.

Retroviral Infection

The open reading frame of Nrp1 was cloned into pBabe-mPlum. The open reading frame of Sema3a were cloned into pBabe-EGFP (addgene, 36999). shNrp1 plasmids were purchased from Origene (OriGene, TG513573A). 15 μg of retroviral plasmid DNA was transfected using FuGENE HD (Roche, E2311) into Platinum E cells (Cell Biolabs), which were plated on a 10-cm tissue culture dish at a density of 5×10^6 cells per dish, 24 h before transfection. After 48 h of transfection, viral medium was harvested and filtered through a 0.45 μm cellulose filter. The viral supernatant was concentrated overnight using Retro-X concentrator (Clontech, 631456) according to manufacturer's protocol. The virus was then resuspended in growth medium mixed with Polybrene (Sigma) to a final concentration of 6 μg ml^{-1}.

$Twist2^+$ and $Pax7^+$ cells were plated at a density of 200,000 cells per 35-mm plate in growth medium. After 24 h, the growth medium was replaced with freshly made viral mixture containing Polybrene and bFGF (5 ng ml^{-1}). 24 hrs later, viral medium was replaced with growth medium with bFGF for recovery.

CRISPR/Cas9 Deletion of Nrp1

In order to delete Nrp1 from $Twist2^+$ cells, we selected two different Nrp1 short guide RNAs (sgRNA) and cloned them into a pLentiV2 vector for lentiviral infection. Lenti-X 293T cells (Clontech, 632180) were transfected with pLentiV2-sgNrp1, psPax2 (addgene, 12260), and pMD2.G (addgene, 12259) using FUGENE HD reagent (Promega, E2311). 48

hours after transfection virus was harvested and concentrated using Lenti-X concentrator (Clontech, 631232) according to manufacturer's protocol. Twist2^{+} cells were then infected with the lentivirus. Infected Twist2^{+} cells were sorted and subsequently harvested for DNA. The sgRNA targeted region was PCR amplified using the following primers: Fwd – 5'-GAGGGTTTATGGGGGACACT-3', Rev – 5'-CAGGACATCTGGGGCTACAT-3'. Amplified PCR product was denatured and re-hybridized, and subsequently subjected to treatment with or without T7E1 endonuclease for 15 minutes at room temperature. The treated products were run on a 1.5% agarose gel alongside T7E1 untreated control samples. The T7E1 untreated bands were gel-isolated, cloned into pCRII-TOPO vector (ThermoFisher, 45-0640), and individual colonies were sequenced to reveal the indels generated by CRISPR/Cas9.

Chimeric Fusion Assay

Twist2^{+} cells were infected with retroviruses expressing shNrp1, control empty vector, or Nrp1 and mixed with primary myoblasts overexpressing Sema3a-EGFP or EGFP control. The next day, cells were placed into DM and differentiated for 7 days. Cells were then fixed and counterstained with Hoechst (ThermoFisher, Hoechst, 3342, 1:1000). The percent of chimeric fusion was calculated as the number of nuclei within chimeric myofibers vs the number of total nuclei.

Generation of MCK-Sema3a Transgenic; Tw2-CreERT2; R26-tdTomato mice

The open reading frame of Sema3a was cloned downstream of the muscle creatine kinase (MCK) promoter(Shield et al. 1996). Transgenic mice were generated as previously described(Papizan et al. 2017),(Naya et al. 2000). Mice were genotyped using primers specific to the MCK promoter and the proximal coding region of Sema3a: Fwd – 5'-CCTGTAGGCTCCTCTATATAACC-3'; Rev – 5'- AGATAGCGAAGTCCCGTC-3'.

Mck-Sema3a transgenic mice were crossed with Twist2-CreERT2; R26-tdTO(Liu et al. 2017) mice to obtain MCK-Sema3a; Twist2-CreERT2; R26-tdTO mice. Both male and female adult mice were used in the studies. Mice were maintained on a mixed genetic background. All animal procedures were approved by the Institutional Animal Care and Use Committee at the University of Texas Southwestern Medical Center.

Tamoxifen Treatment

Tamoxifen treatment was performed as previously described(Liu et al. 2017). Tamoxifen was administered by intraperitoneal injection to 8-week-old MCK-Sema3a;Twist2-CreERT2; R26-tdTO/+ and Twist2-CreERT2; R26-tdTO/+ mice as schematized in Fig. 7A. Hind limb muscles were harvested at 8 weeks after the first injection.

Statistical analysis

All statistical analyses were performed using GraphPad Prism 7 (GraphPadSoftware). Data are presented as mean ± s.e.m. Differences between groups were tested for statistical significance by using the two-sample t-test. P-value definitions and number of biological replicates for each experiment is indicated in the figure legends. Randomization was not used

in most of the animal studies. However, images of type IIb and Sema3a staining were taken by an investigator who was blinded to the group allocation. All immunofluorescence images are representative of at least three independent experiments or mice of the same genotypes. No statistical method was used to predetermine sample size.

Figures

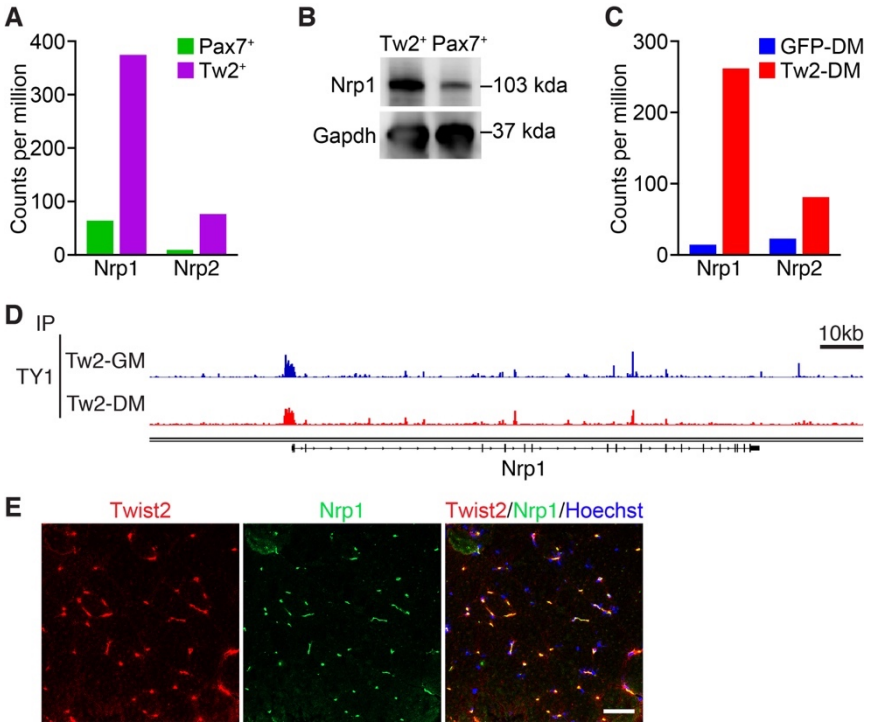


Figure 3-1. (**A)** Levels of Nrp1 and Nrp2 mRNA expression in freshly isolated Twist2$^+$ cells vs Pax7$^+$ cells, as determined by RNA-seq. (**B**) Western blot of Nrp1 protein levels in cultured Twist2$^+$ and Pax7$^+$ myoblasts. GAPDH serves as a loading control. (**C**) Levels of of Nrp1 and Nrp2 mRNA expression in Twist2-overexpressing Twist2$^+$ cells (Tw2-DM) vs GFP-infected Twist2$^+$ cells (GFP-DM) after 4 days in differentiation medium, as determined by RNA-seq. (**D**) Genome browser shot of Twist2 binding at the Nrp1 promoter in both growth media (GM) and differentiation media (DM). (**E**) Co-immunostaining of Nrp1 (green), Twist2 (red), and HOECHST (blue) in adult mouse transverse section of quadriceps muscle. Scale: 50 µm.

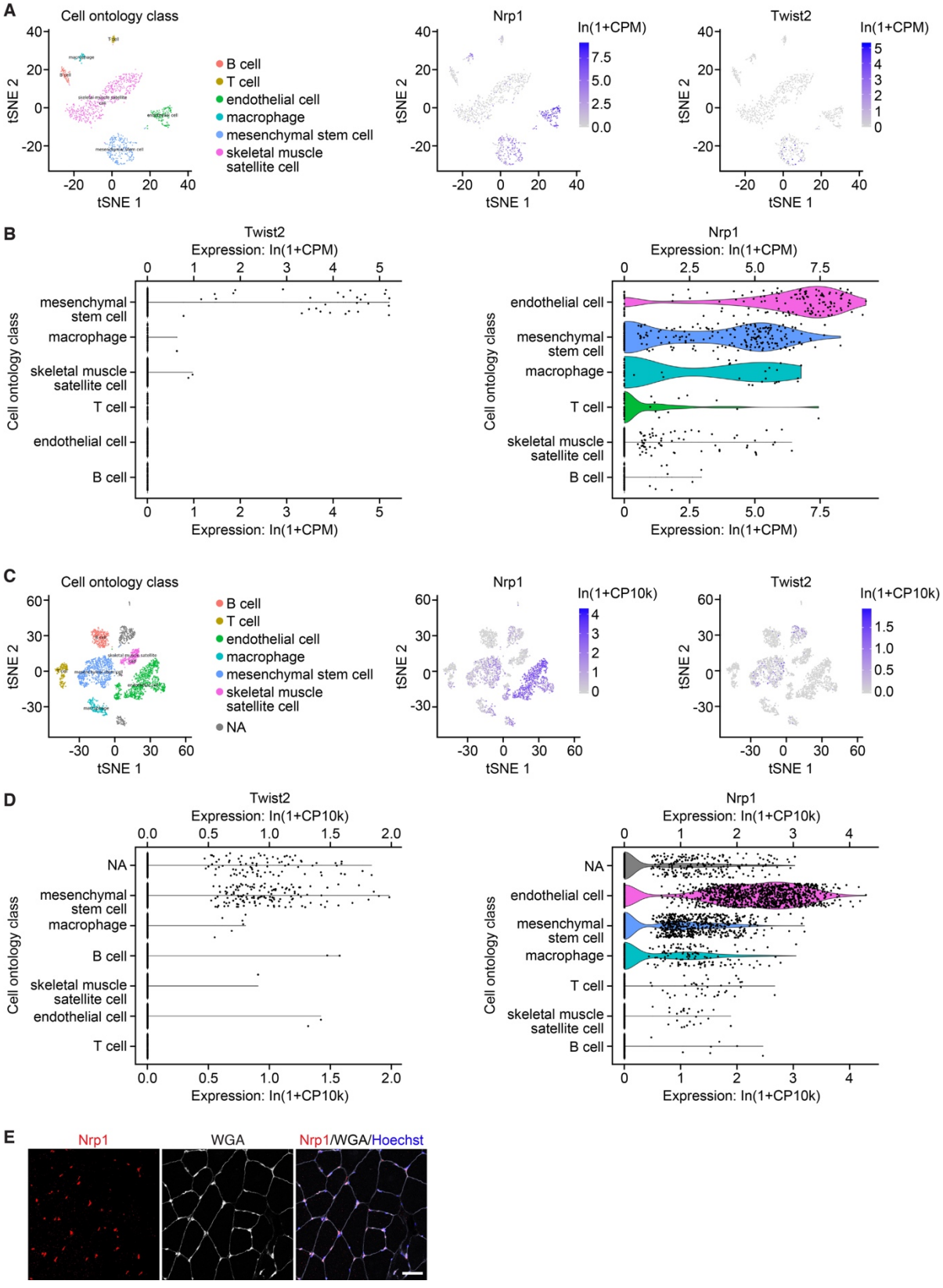

A
Cell ontology class
B cell
T cell
endothelial cell
macrophage
mesenchymal stem cell
skeletal muscle satellite cell
tSNE 1
tSNE 2
Nrp1
ln(1+CPM)
Twist2
ln(1+CPM)
B
Twist2
Expression: ln(1+CPM)
Nrp1
Expression: ln(1+CPM)
Cell ontology class
C
Cell ontology class
NA
Nrp1
ln(1+CP10k)
Twist2
ln(1+CP10k)
D
Twist2
Expression: ln(1+CP10k)
Nrp1
Expression: ln(1+CP10k)
E
Nrp1
WGA
Nrp1/WGA/Hoechst

Figure 3-2. (**A**) tSNE and (**B**) violin plots of FACS-mediated single cell RNA sequencing of limb muscle from Tabula Muris depicting Nrp1 and Twist2 expression. (**C**) tSNE and (**D**) violin plots of droplet-based single cell RNA sequencing of limb muscle from Tabula Muris depicting Nrp1 and Twist2 expression. (**E**) Immunostaining of Nrp1 (red), wheat germ agglutinin (WGA) (white), and Hoechst (blue) on transverse sections of gastrocnemius muscle from 3-month-old C57Bl6 wild-type mice.

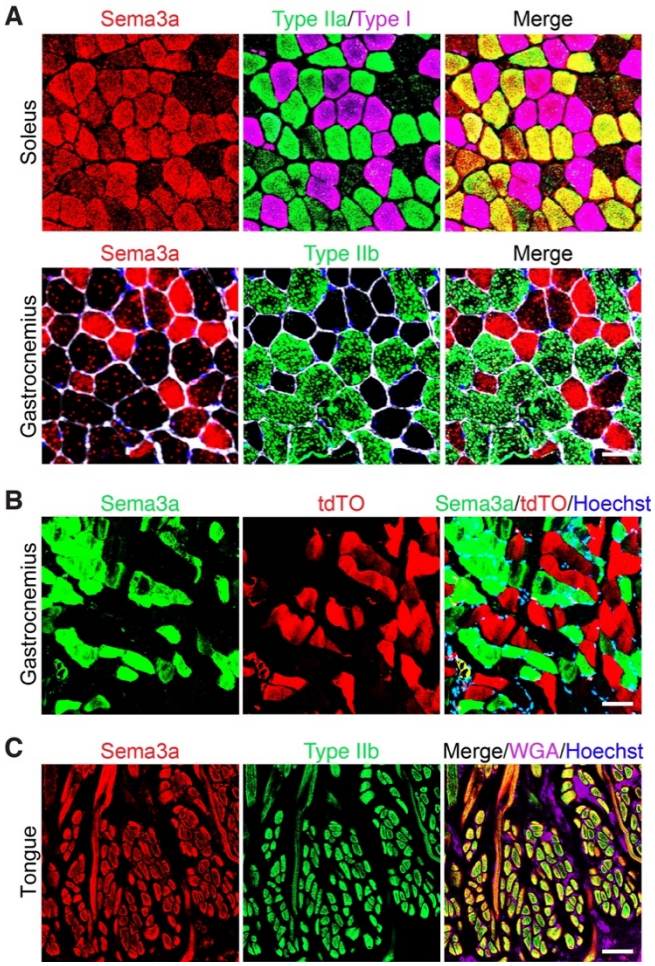


Figure 3-3. (**A**) (Top) Transverse sections of soleus muscle from 3-month-old C57Bl6 mice were co-immunostained for Sema3a (red), type IIa myosin (green), and type I myosin (magenta). (Bottom) Transverse sections of gastrocnemius muscles were co-immunostained for Sema3a (red), Type IIb myosin (green), wheat-germ agglutinin (white), and Hoechst (blue). Scale bar: 50 µm. (**B**) Transverse gastrocnemius muscle sections of Tw2-CreERT2; R26-tdTO mice at 4 months post-TMX were fixed and immunostained for Sema3a (green). Sema3a staining was overlayed with tdTO signal from myofibers receiving Twist2^{+} cell contribution (red) and counterstained with Hoechst (blue). Scale bar: 50 µm. (C) Transverse sections of tongue muscle from 3-month-old C57Bl6 mice were co-immunostained for

Sema3a (red), type IIb myosin (green), wheat-germ agglutinin (magenta), and Hoechst (blue). Scale: 50 μm.

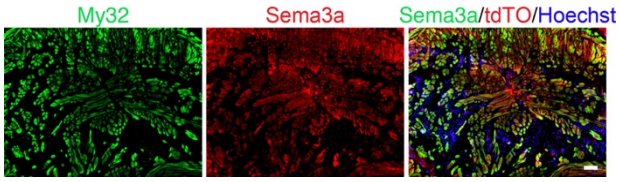


Figure 3-4. Transverse sections of tongue muscle from 3-month-old C57Bl6 WT mice were co-immunostained with My32 (green), Sema3a (red), and Hoechst (blue).

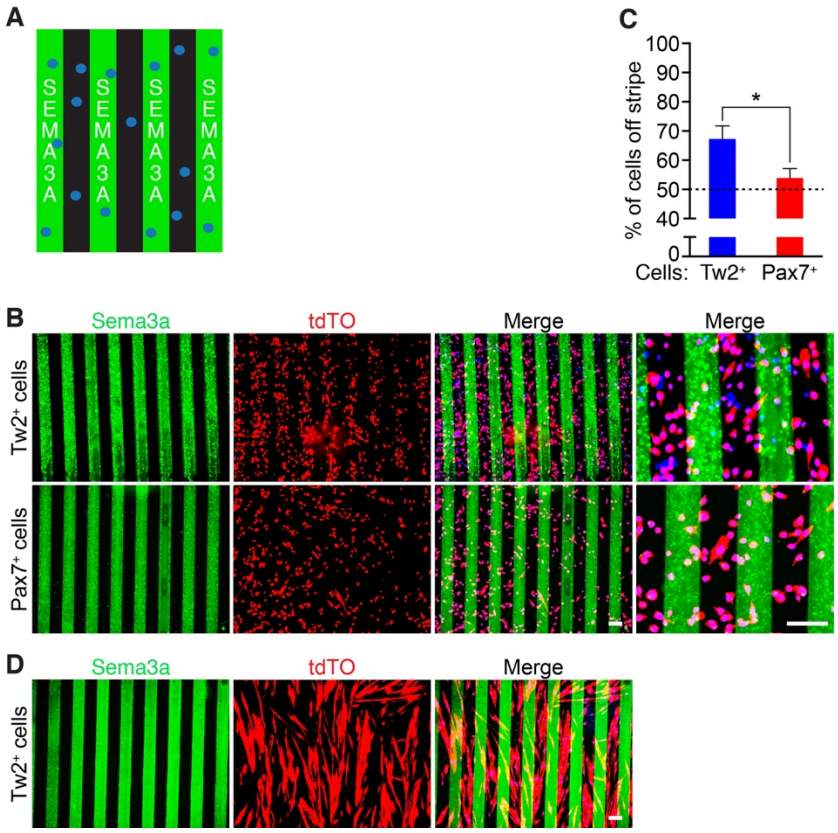


Figure 3-5. (**A**) Schematic of Sema3a stripe assay. (**B**) (Top) Twist2$^+$ cells (tdTO$^+$) were seeded onto Sema3a stripes (green) and analyzed one day after seeding. (Bottom) Pax7$^+$ satellite cells (tdTO$^+$) were seeded onto on Sema3a stripes (green) and analyzed one day after seeding. Cells were co-stained with Hoechst (blue). Scale bar: 100 μm. (**C**) Quantification of Sema3a avoidance as the percent of cells residing off the stripe over total cells in the field. The dashed line represents the baseline for cells unresponsive to Sema3a. Three separate fields were quantified for each sample with a total of 3 samples per cell type. *: $p < 0.05$. (**D**) Twist2$^+$ cells (tdTO$^+$) were differentiated 1 day after seeding onto Sema3a stripes (green). Cells were fixed after 4 days in DM and co-stained with Hoechst (blue).

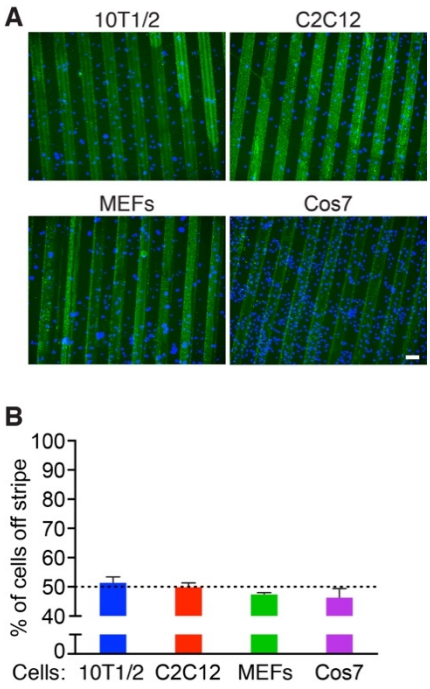


Figure 3-6. (**A**) 10T1/2 fibroblasts, C2C12 myoblasts, mouse embryonic fibroblasts (MEFs), and Cos7 cells were fixed and stained with Hoechst (blue) 1 day after seeding onto Sema3a stripes (green). (**B**) Quantification of Sema3a avoidance for 10T1/2 fibroblasts, C2C12 myoblasts, mouse embryonic fibroblasts, and Cos7 cells. The dashed line represents the baseline for cells unresponsive to Sema3a. Three fields per sample per experiment were quantified.

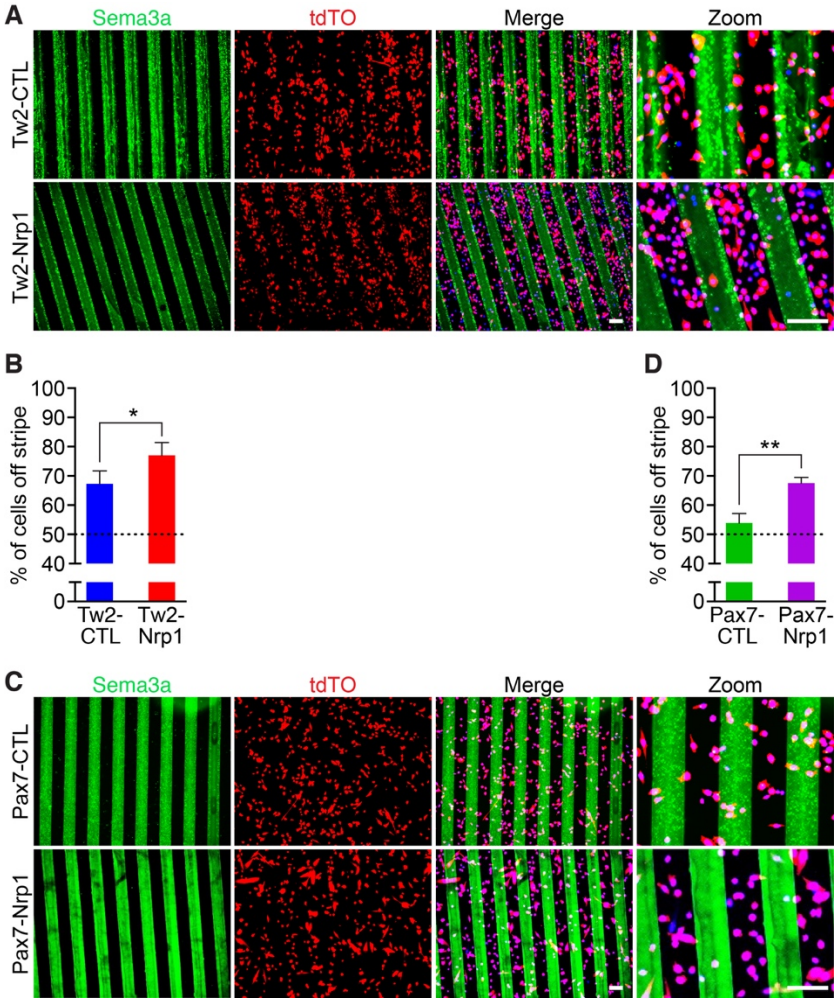


Figure 3-7. (**A**) Overexpression of Nrp1 increases repulsion of Twist2$^+$ cells from Sema3a stripes. (Top) Control Twist2$^+$ cells (tdTO$^+$) were seeded on Sema3a stripes (green) and analyzed 1 day after seeding. (Bottom) Nrp1-overexpressing Twist2$^+$ cells (red) were seeded on Sema3a stripes (green) and analyzed 1 day after seeding. Cells were co-stained with Hoechst (blue). Scale bar: 100 μm (**B**) Quantification of Sema3a avoidance as the percent of cells residing off the stripe in Fig. 4A. The dashed line represents the baseline for cells unresponsive to Sema3a. Three separate fields were quantified for each sample with a total of 3 samples per cell type. *: $p < 0.05$. (**C**) Overexpression of Nrp1 resulted in repulsion of

$Pax7^+$ cells from Sema3a stripes. (Top) Control-infected $Pax7^+$ cells ($tdTO^+$) were seeded on Sema3a stripes (green) and analyzed 1 day after seeding. (Bottom) Nrp1-overexpresssing $Pax7^+$ cells ($tdTO^+$) were seeded on Sema3a stripes (green). Cells were analyzed 1 day after seeding and co-stained with Hoechst (blue). Scale bar: 100 μm. (**D**) Quantification of Sema3a avoidance as the percent of cells residing off the stripe in Fig. 4C. The dashed line represents the baseline for cells unresponsive to Sema3a. Three separate fields were quantified for each sample with a total of 3 samples per cell type. **: $p < 0.005$.

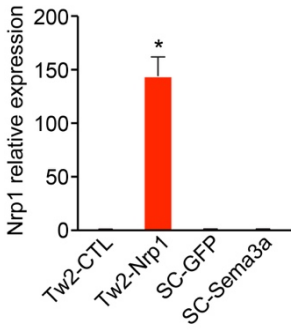


Figure 3-8. Quantitative real-time PCR confirming over-expression of Nrp1 in $Twist2^+$ cells infected with Nrp1 expressing retrovirus. *: $p < 0.05$.

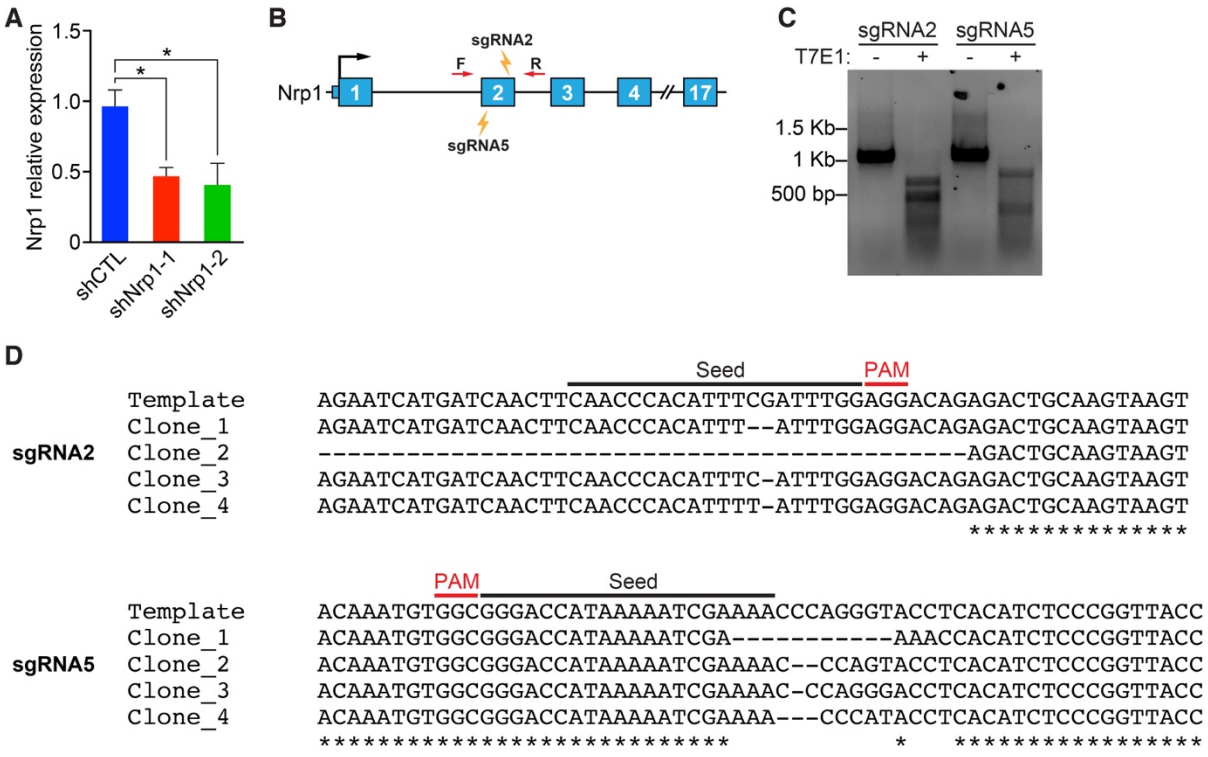


Figure 3-9. (**A**) Quantitative real-time PCR confirming knockdown of Nrp1 using two separate shRNAs. 3 separate experiments were performed. *$p < 0.05$. (**B**) Schematic showing the Nrp1 genomic locus and the locations of representative sgRNA's targeting exon 2. Locations of primers for T7E1 assay are depicted in red. (**C**) Nrp1 sgRNAs 2 and 5 were cloned into cas9-containing lentiviral vectors in order to knockout Nrp1 in Twist2^{+} cells. T7E1 assay was used to validate cutting efficiency. (**D**) TOPO TA cloning and sequencing of Nrp1 sgRNA infected Twist2^{+} cells. Representative clones are shown. Cas9-sgRNA targeting sites are depicted on the template strand.

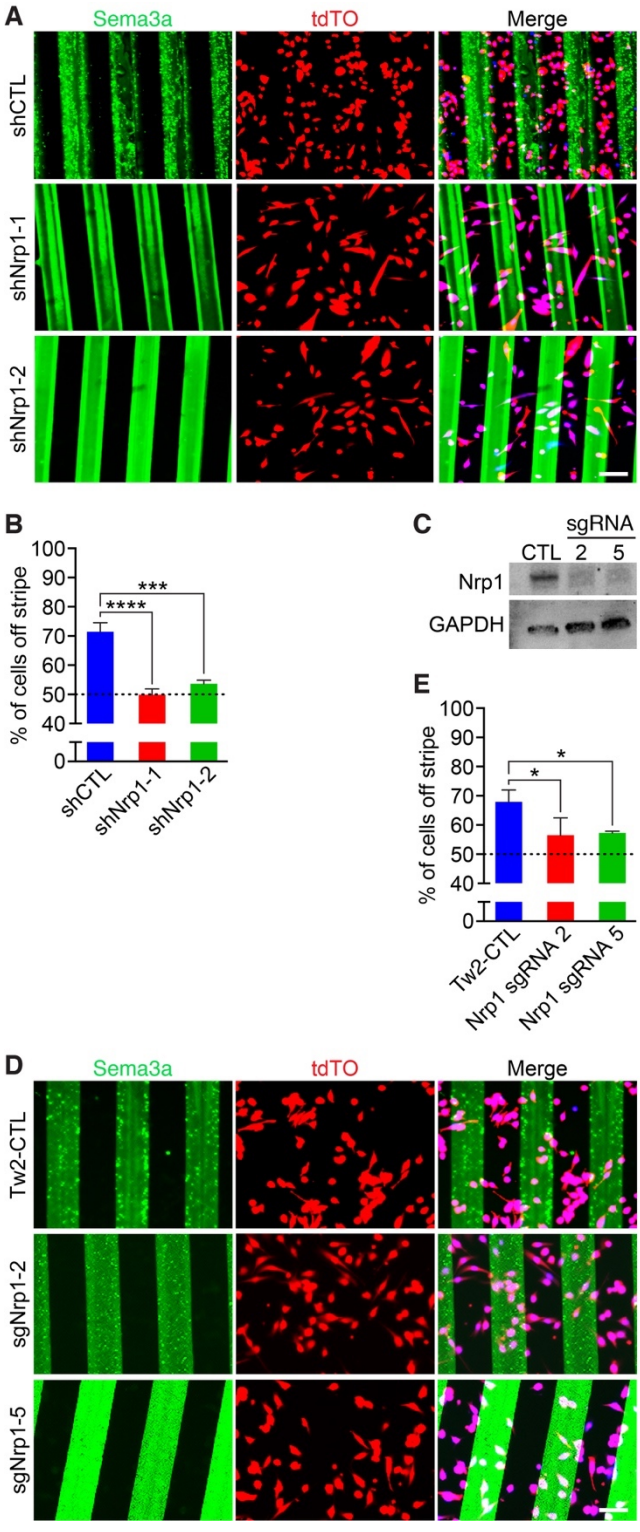

A
Sema3a
tdTO
Merge
shCTL
shNrp1-1
shNrp1-2
B
% of cells off stripe
100
90
80
70
60
50
40
0
***
****
shCTL
shNrp1-1
shNrp1-2
C
sgRNA
CTL
2
5
Nrp1
GAPDH
E
% of cells off stripe
100
90
80
70
60
50
40
0
*
*
Tw2-CTL
Nrp1 sgRNA 2
Nrp1 sgRNA 5
D
Sema3a
tdTO
Merge
Tw2-CTL
sgNrp1-2
sgNrp1-5

Figure 3-10. (**A**) Knockdown of Nrp1 by shRNA in Twist2$^+$ cells (tdTO$^+$) abolished Sema3a avoidance. (Top) Control shRNA (shCtrl) infected Twist2$^+$ cells 1 day after seeding on Sema3a stripes (green). (Middle, Bottom) Twist2$^+$ cells overexpressing 2 separate shNrp1 (shNrp1-1 and shNrp1-2) 1 day after seeding on Sema3a stripes (green). Cells were co-stained with Hoechst (blue). Scale bar: 100 μm. (**B**) Quantification of Sema3a avoidance as the percent of cells residing off the stripe in Fig. 5A. The dashed line represents the baseline for cells unresponsive to Sema3a. Three separate fields were quantified for each sample with a total of 3 samples per cell type. ***: $p < 0.0005$, ****: $p < 0.00005$. (**C**) Western blot showing loss of Nrp1 protein in Twist2$^+$ cells infected with sgRNAs targeting Nrp1. GAPDH was used as a loading control. (**D**) (Top) Control pLentiCrisprV2-infected Twist2$^+$ cells (red) 1 day after seeding on Sema3a stripes (green). (Middle, Bottom) Two separate Nrp1 sgRNA-infected Twist2$^+$ cells (sgNrp1-2 and sgNrp1-5) 1 day after seeding on Sema3a stripes (green). Cells were co-stained with Hoechst (blue). Scale bar: 100 μm. (**E**) Quantification of Sema3a avoidance as the percent of cells residing off the stripe in Fig. 5E. The dashed line represents the baseline for cells unresponsive to Sema3a. Three separate fields were quantified for each sample with a total of 3 samples per cell type. *: $p < 0.05$.

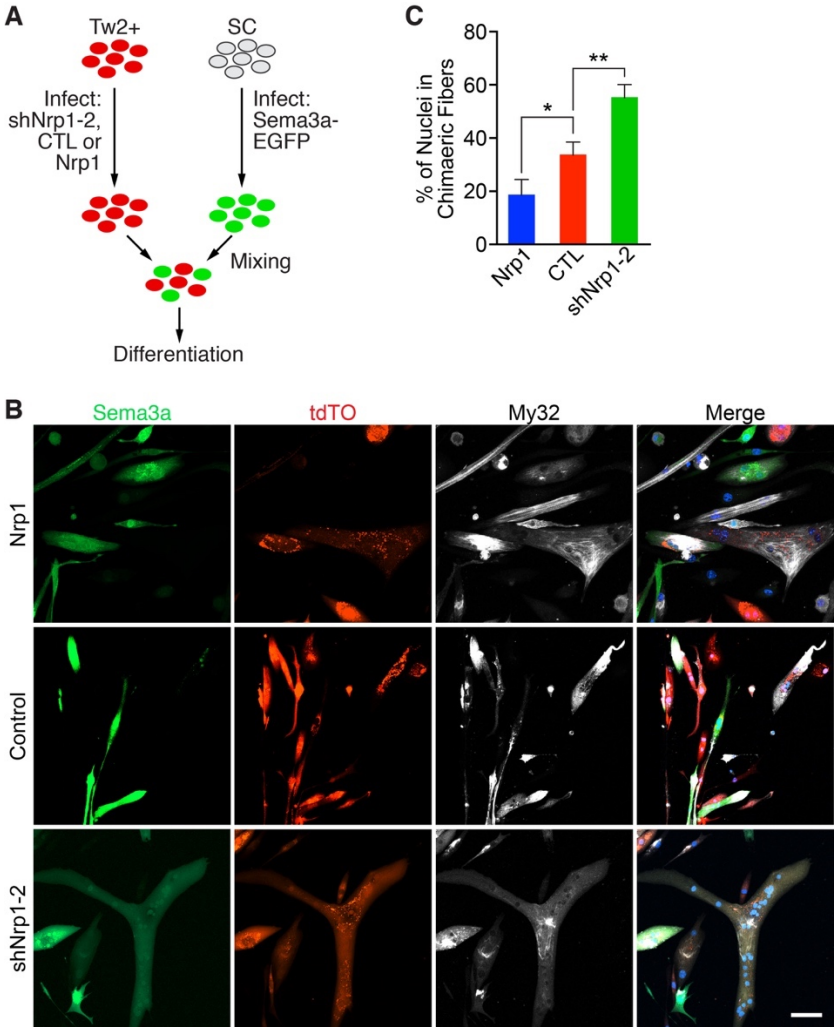


Figure 3-11. (**A**) Experimental scheme for chimeric fusion assay. Twist2^{+} cells were retrovirally infected with shNrp1-2, control empty vector, or Nrp1, and mixed with primary myoblasts (SCs) overexpressing Sema3a-EGFP in equal numbers. Cells were then differentiated for 7 days. (**B**) Primary myoblasts over-expressing Sema3a (green) were mixed with Twist2^{+} cells (tdTO^{+}) infected with Nrp1 (top), control empty vector (middle), or shNrp1-2 (bottom) and differentiated for 7 days. Cells were fixed and stained with Hoechst (blue) and an antibody recognizing fast myosin (MY32; white). Scale bar: 50 μm. (**C**) Quantification of percent of nuclei in chimeric fibers for Figure 6B. Percent of nuclei in

chimeric fibers was calculated as the percent of the number of nuclei in chimeric fibers over the total number of nuclei (red-only, green-only, and chimeric myofibers). Three fields per sample per experiment were quantified. Three separate experiments were performed. *: $p < 0.05$, **: $p < 0.005$.

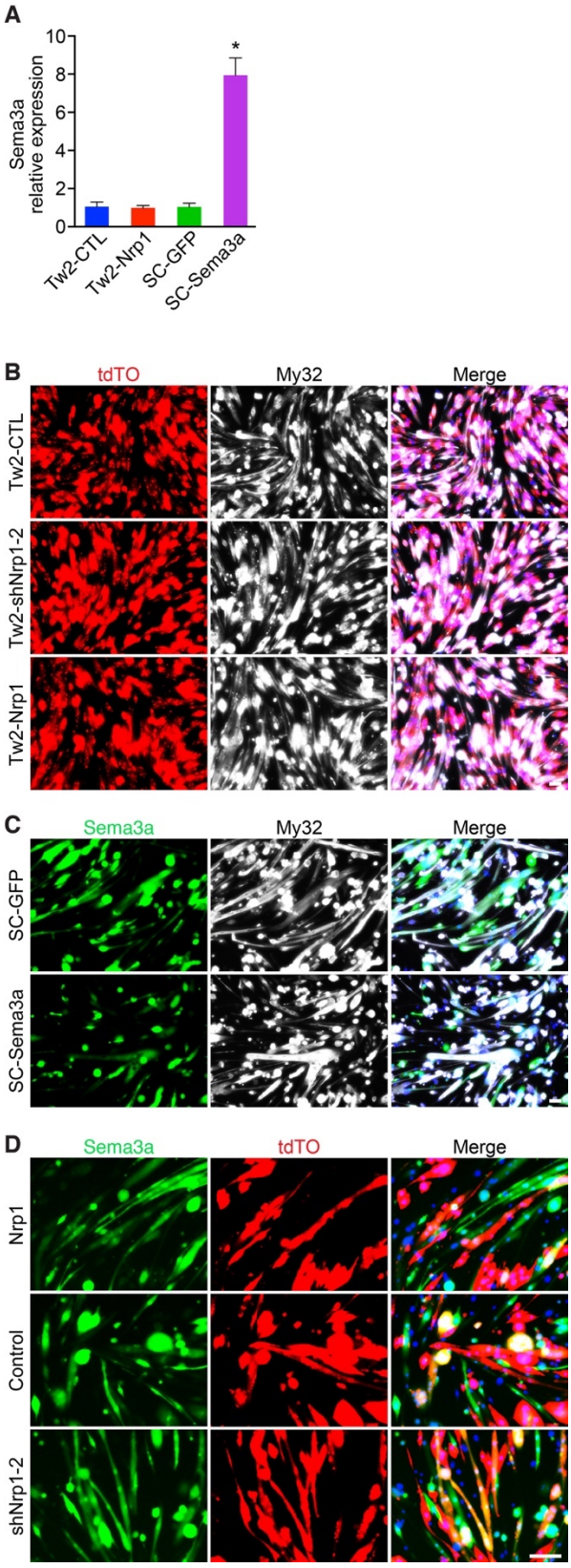

A
Sema3a relative expression
10
8
6
4
2
0
*
Tw2-CTL
Tw2-Nrp1
SC-GFP
SC-Sema3a
B
tdTO
My32
Merge
Tw2-CTL
Tw2-shNrp1-2
Tw2-Nrp1
C
Sema3a
My32
Merge
SC-GFP
SC-Sema3a
D
Sema3a
tdTO
Merge
Nrp1
Control
shNrp1-2

Figure 3-12. (**A**) Quantitative real-time PCR confirms overexpression of Sema3a in primary myoblasts infected with Sema3a. *: $p < 0.05$. (**B**) Twist2^{+} cells (red) infected with control empty vector, shNrp1-2, or Nrp1 were differentiated for 7 days before fixing and staining fast myosin (white) and Hoechst (blue). Scale bar: 100 μm. (**C**) Primary myoblasts infected with EGFP control (green) or Sema3a-EGFP (green) were differentiated for 7 days before fixing and staining for fast myosin (white) and Hoechst (blue). Scale bar: 100 μm. (**D**) Primary myoblasts over-expressing Sema3a (green) were mixed with Twist2^{+} cells (red) infected with Nrp1 (top), control empty vector (middle), or shNrp1-2 (bottom) and differentiated for 7 days. Cells were fixed and stained with Hoechst (blue). Scale bar: 100 μm.

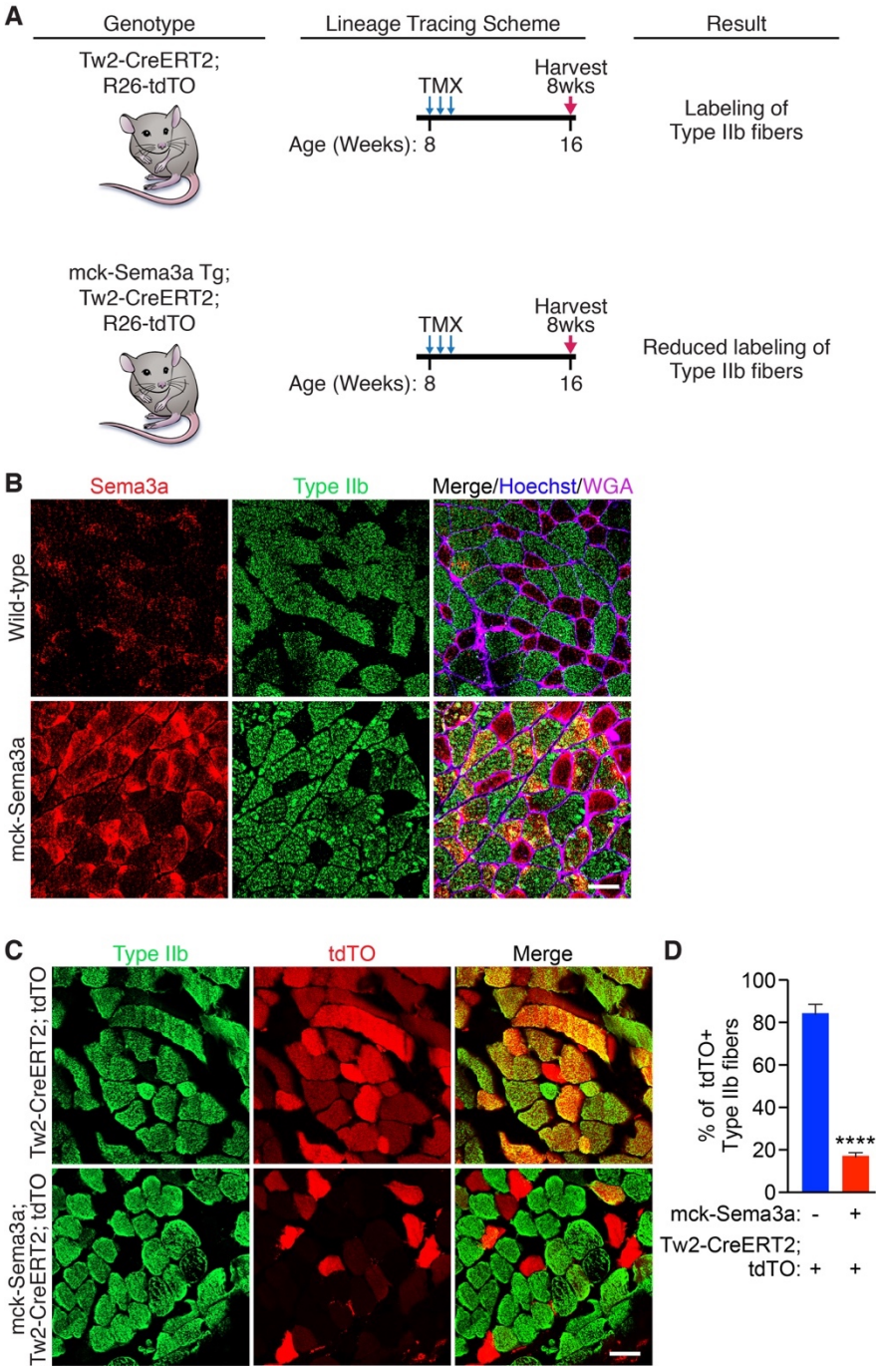


Figure 3-13. (**A**) Schematic of Twist2^{+} cell lineage tracing in MCK-Sema3a transgenic mice. Twist2-CreERT2; tdTO and MCK-Sema3a Tg; Twist2-CreERT2; R26-tdTO mice were injected with 3 doses of TMX on 3 alternating days starting at 8 weeks of age. Muscles from

these mice were then harvested 8 weeks later. (**B**) Transverse gastrocnemius muscle sections of wild-type C57Bl6 and mck-Sema3a transgenic littermates were immunostained for Sem3a (red), type IIb myofibers (green), wheat-germ agglutinin (magenta), and Hoechst (blue). (**C**) Immunostaining of gastrocnemius muscle from Tw2-CreERT2; tdTO mice and MCK-Sema3a; Tw2-CreERT2; tdTO mice for type IIb fibers (green) and tdTO (red). Scale : 50µm. (**D**) Quantification of the percent of $tdTO^+$ type IIb fibers over total type IIb fibers. 500 type IIb myofibers were quantified per mouse (Tw2-CreERT2; tdTO, n = 3 mice; mck-Sema3a; Tw2-CreERT2; tdTO, n = 4 mice). ****: $p < 0.00005$.

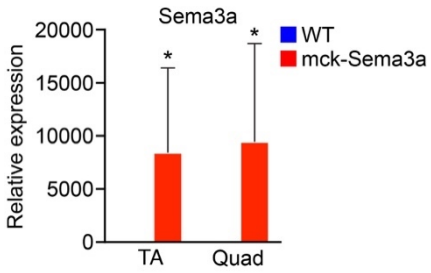


Figure 3-14. Quantitative real-time PCR of tibialis and quadriceps muscle depicting over-expression of Sema3a in MCK-Sema3a transgenic mice. n = 4 for each genotype; *: $p < 0.05$.

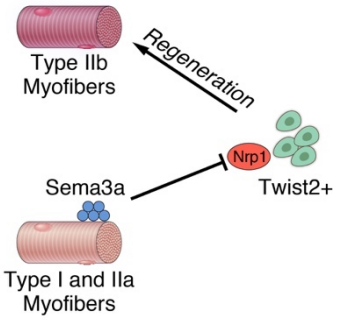


Figure 3-15. Model of Sema3a and Nrp1-mediated Twist2^{+} cell fiber-type specificity in muscle.

CHAPTER FOUR

AMPLIFICATION OF TWIST2 IN RHABDOMYOSARCOMA REPRESSES MYOGENESIS AND PROMOTE ONCOGENESIS THROUGH REDIRECTION OF MYOD DNA BINDING

Acknowledgement

Parts of this chapter, including figures, have been reproduced, with or without modifications, from our previously published work (Li et al. 2019).

Abstract

Rhabdomyosarcoma (RMS) is an aggressive pediatric cancer composed of myoblast-like cells. Recently, we discovered a unique muscle progenitor marked by the expression of the TWIST2 transcription factor. Genomic analyses of 258 RMS patient tumors uncovered prevalent copy number amplification events and increased expression of *TWIST2* in fusion-negative RMS. Knockdown of *TWIST2* in RMS cells results in upregulation of *MYOGENIN* and decrease in proliferation, implicating TWIST2 as an oncogene in RMS. Through an inducible-Twist2 expression system, we identify Twist2 as a reversible inhibitor of myogenic differentiation with the remarkable ability to promote myotube dedifferentiation in vitro. Integrated analysis of genome-wide ChIP-Seq and RNA-Seq data revealed the first dynamic chromatin and transcriptional landscape of Twist2 binding during myogenic differentiation. During differentiation, Twist2 competes with MyoD at shared DNA motifs to direct global

gene transcription and repression of the myogenic program. Additionally, TWIST2 shapes the epigenetic landscape to drive chromatin opening at oncogenic loci and chromatin closing at myogenic loci. These epigenetic changes redirect MyoD binding from myogenic genes towards oncogenic, metabolic, and growth genes. Our study reveals the dynamic interplay between two opposing transcriptional regulators that control the fate of RMS and provides insight into the molecular etiology of this aggressive form of cancer.

Introduction

Rhabdomyosarcoma (RMS) is an aggressive pediatric soft tissue tumor expressing hallmarks of the skeletal muscle lineage, such as MyoD and Myf5 (Hettmer and Wagers 2010; Saab et al. 2011; Skapek et al. 2019). A key characteristic of RMS is the inability of the tumor cells to undergo myogenic differentiation, even in the presence of elevated levels of muscle master regulators such as MyoD and Myf5 (Tapscott and Weintraub 1991; Saab et al. 2011; Tapscott et al. 1993; MacQuarrie et al. 2013; Skapek et al. 2019). Traditionally, RMS was classified into two main histological subtypes: alveolar and embryonal (Skapek et al. 2019). However recent advances in molecular diagnostics have enabled a more refined classification of RMS into fusion-positive and fusion-negative RMS, based on the presence or absence of key chromosomal translocations (Tsokos 1994; Skapek et al. 2019). Fusion-positive RMS (FPRMS) is commonly driven by the balanced translocation of chromosomes 2 and 13 to generate an oncogenic Pax3-Foxo1 fusion protein (Saab et al. 2011; Skapek et al. 2019). This fusion protein acts as an overactive transcription factor at Pax3 binding sites co-localized with E-box motifs to organize super enhancers and drive RMS pathogenesis

(Gryder et al. 2017; Cao et al. 2010a). The mechanism of pathogenesis for fusion-negative RMS (FNRMS) is less understood. Compared to FPRMS, FNRMS exhibits significantly greater genomic instability (Xu et al. 2018). Traditional oncogenic regulators such as KRAS, NRAS, and TP53 are only mutated in less than 10% of RMS patient cohorts, respectively, suggesting other mechanisms (Shern et al. 2014). Bioinformatic analyses of RMS genomic data have revealed copy number alterations of oncogenes and tumor suppressors that play a key role in driving RMS development (Xu et al. 2018; Preussner et al. 2018). While targeting growth is a common strategy for many cancers, targeting myogenic differentiation may represent a unique and exploitable vulnerability in RMS, given that terminally differentiated myofibers are permanently post-mitotic (Saab et al., 2011; Tremblay et al., 2014; Xu et al., 2018; Skapek et al. 2019).

Skeletal muscle is a highly regenerative tissue required for animal life. Muscle regeneration occurs through a population of resident stem cells called satellite cells, which are marked by expression of the transcription factor Pax7 (Chang and Rudnicki 2014; Shi and Garry 2006). Upon muscle injury, satellite cells up-regulate the muscle master regulator MyoD to proliferate and repair injured muscle (Chang and Rudnicki 2014; Shi and Garry 2006). Our lab recently discovered a unique muscle progenitor through fate-mapping of the transcription factor Twist2 (Tw2+ cells) (Liu et al. 2017). These Tw2+ cells are distinct from satellite cells and do not express Pax7 in vivo (Liu et al. 2017). In culture, Tw2+ cells downregulate Twist2 expression and upregulate MyoD expression, allowing them to form terminally differentiated myotubes (Liu et al. 2017). We found that overexpression of Twist2

in vitro was sufficient to drive transcriptional and phenotypic repression of the myogenic program (Liu et al. 2017). Mammalian Twist2, along with its paralog Twist1, differ in their temporal and spatial expression patterns and are important regulators of mesoderm development, epithelial-mesenchymal transition, and cellular differentiation (Franco et al., 2011; Merindol et al., 2014; Zhang et al., 2008; (Li et al. 1995; Gong and Li 2002). Additionally, in vitro studies have also shown that both Twist proteins can inhibit MyoD-induced transdifferentiation of fibroblasts into myoblasts (Gong and Li, 2002; Spicer et al., 1996). The functional roles of Twist1 and Twist2 are likely imparted by their DNA binding activity, which is mediated by the basic-helix-loop-helix (bHLH) domain. Like other bHLH transcription factors, both Twist1 and Twist2 are capable of binding canonical CANNTG E-box motifs (Jones 2004; Chang et al. 2015). While Twist2 de novo motif analysis has not been previously reported, de novo motif analysis of Twist1 reveals a preference for CAGATG E-boxes as well as a secondary preference for double E-box motifs (Chang et al. 2015). Other bHLH proteins such as MyoD and NeuroD prefer different variations of the middle two nucleotides (Fong et al. 2012; 2015). Along with tissue expression pattern, the exact binding sequence likely defines the target genes and functions of different bHLH transcription factors (Fong et al. 2012; 2015). Indeed, swapping of the bHLH domains between MyoD and NeuroD converts NeuroD into a muscle master regulator and MyoD into a master neurogenic regulator (Fong et al. 2015).

Traditionally, bHLH transcription factors were thought to act as direct transcriptional activators of gene expression. The advent of ChIP-Seq has revealed that bHLH factors like

N-myc and MyoD are capable of invading enhancer regions throughout the genome to regulate the global chromatin architecture (Zeid et al. 2018; la Serna et al. 2005). These chromatin modifications are often associated with histone acetylation, enabling the activation of tightly regulated transcriptional networks (Zeid et al. 2018; la Serna et al. 2005). In the case of N-myc, epigenetic alterations of occupied enhancers involve coordination and co-occupancy of additional bHLH transcription factors such as TWIST1 to drive activation of the neuroblastoma gene program (Zeid et al. 2018). In the case of muscle differentiation, binding of MyoD to DNA associates with histone acetylation around myogenic loci to promote muscle gene expression (Blum et al. 2012; la Serna et al. 2005). Recent advances in genome-wide analysis of transcription factor binding and epigenetic remodeling provide a powerful tool for refining existing knowledge of bHLH transcription factor function in a global context. In particular, TWIST2 epigenetic and transcriptional regulation during myogenic differentiation is multi-faceted and unexplored.

In this study, we analyzed genomic and transcriptomic sequencing data from RMS tumors of 258 patients and identified frequent copy number amplification of the TWIST2 loci, which in turn drives overexpression of TWIST2 in RMS. We show that knockdown of *TWIST2* in RMS cell lines results in upregulation of Myogenin, the essential transcriptional switch for myogenic differentiation. Using ChIP-Seq and RNA-Seq, we uncovered the previously unknown interplay between Twist2 and MyoD during myogenic differentiation and its effect on transcriptional output and the chromatin landscape. Our findings indicate that during differentiation, Twist2 globally shifts MyoD activity from myogenic loci to EMT

and tumor growth loci through competition and epigenetic modifications at promoters and enhancers, highlighting the central role of TWIST2 in RMS pathogenesis.

Results

TWIST2 and TWIST1 genes are highly amplified in fusion-negative embryonal RMS rhabdomyosarcoma

Previously, our lab identified a previously unrecognized progenitor through lineage tracing of the transcription factor *Twist2* (Liu et al. 2017). We showed that Tw2+ cells downregulate Twist2 expression in culture and that overexpression of Twist2 was capable of repressing myogenesis (Liu et al. 2017), providing evidence for a physiological role of Twist2 in regulating myoblast differentiation. To explore the potential role of Twist2 in myogenic regulation within pathological contexts, we analyzed SNP array data from 258 RMS patients (Xu et al. 2018) for copy number gain of both *TWIST2* and *TWIST1* loci (Figure 4-1A). *TWIST* amplification was not significantly associated with FPRMS cases, however both *TWIST2* and *TWIST1* were significantly amplified in FNRMS cases. Coordinately, we observed increased expression of *TWIST2* and *TWIST1* in FNRMS patients but not FPRMS patients, implicating *TWIST* genes as previously unknown oncogenes for FNRMS patients (Figure 4-1B). Strikingly, 72% (114 out of 158 cases) of FNRMS patients contain copy-number amplification events on loci of either *TWIST2*, *TWIST1*, or both (Figure 4-1C). In contrast, somatic mutations in known cancer driver genes such as KRAS, NRAS, and TP53 occur in less than 10% of FNRMS patient cohorts. We also found that *TWIST* expression levels decreased during differentiation of human myoblast (Figure 4-1D),

suggesting that decreased *TWIST2* gene expression may play a role during normal muscle differentiation.

In order to validate the pathological role of *TWIST2* amplification in blocking myogenic differentiation of RMS, we performed siRNA knockdown of *TWIST2* and *TWIST1*. We first screened various RMS cell lines to identify which ones exhibited the highest level of *TWIST2* expression. We found that RD and RH18 cells showed higher levels of *TWIST2* expression (Figure 4-2A). We then performed siRNA knockdown of *TWIST2* and *TWIST1* and validated knockdown by real-time PCR (Figure 4-2B, C). Loss of either *TWIST2* or *TWIST1* expression resulted in a significant increase in the expression of *MYOG* with protein being detectable in RD cells treated with siTWIST1 (Figure 4-1E; Figure 4-2D). Loss of either TWIST2 or TWIST1 also resulted in upregulation of fast myosin light chain, MYL1, in RD but not RH18 cells, likely due to differing genetic backgrounds (Figure 4-2E, F). We also found that knockdown of *TWIST2 and TWIST1* reduced RD and RH18 cell accumulation (Figure 4-1F). In accordance, knockdown of TWIST2 and TWIST1 in RD cells resulted in decreased EdU labeling (Figure 4-2G). These results suggest that amplification of *TWIST* in primary RMS tumors may impair differentiation and drive oncogenesis.

Twist2 is a reversible inhibitor of myogenic differentiation

A hallmark of RMS is the inability of tumor cells to undergo normal myogenic differentiation even in the presence of MyoD and Myf5 (Saab et al. 2011). To explore the mechanism by which TWIST2 represses myogenic differentiation, we first generated a

doxycycline (Dox)-inducible *Twist2-IRES-GFP* (iTwist2) stable Tw2+ cell line (Figure 4-3A) and tested whether Twist2-mediated myogenic repression was a permanent or reversible event. iTwist2 cells were provided 10ug/ml of Dox in growth media (GM) on day 1 to induce expression of Twist2 and a GFP reporter (Figure 4-3B). On the subsequent day, iTwist2 cells were switched to differentiation medium (DM) supplemented with Dox (Figure 4-3B). After 3 days in DM, iTwist2 cells were either maintained on Dox or switched to vehicle (Figure 4-3B). iTwist2 cells receiving continuous Dox treatment were unable to differentiate, as shown by lack of myosin heavy chain (Myosin) co-staining with GFP (Figure 4-3B). However, iTwist2 cells removed from Dox were able to resume myogenic differentiation, as shown by formation of GFP-positive myotubes (Figure 4-3B). This result revealed that Twist2 represses myogenesis in a reversible manner.

We next sought to determine whether Twist2 expression in differentiated myotubes could drive dedifferentiation into mononuclear cells. We first modified our iTwist2 system to enable lineage-tracing with a dsRed reporter (iTwist2-tracer), which allowed us to distinguish mononuclear dedifferentiated myoblasts from mononuclear cells that never differentiated (Figure 4-3C; Figure 4-4A-B). Since Tw2+ cells are tdTomato^{+} and indistinguishable from the dsRed reporter, we generated iTwist2-tracer primary myoblasts for dedifferentiation experiments. Following previously published dedifferentiation schemes (Wang et al. 2015), we allowed iTwist2-tracer cells to differentiate for 4 days before adding Dox (Figure 4-3C, D). Cells were kept in DM + Dox for 2 additional days, before being switched to GM + Dox. iTwist2-tracer cells that did not receive Dox, maintained their differentiated state as shown

by immunofluorescence and western blotting for Myosin and GFP (Figure 4-3D; Figure 4-4C). However, iTwist2-tracer cells that received Dox on day 6 were able to undergo dedifferentiation back into mononuclear cells within 2 days of switching to GM, as evidenced by the presence of significantly increased $dsRed^{+}$ mononuclear cells and downregulation of Myosin expression (Figure 4-3D, E; Figure 4-4C). We also used live cell imaging to capture the dedifferentiation event of a myofiber after induction of *Twist2* expression (Supplemental Movie S1, 2). Interestingly, while dedifferentiated cells did not re-enter cell cycle (data not shown), they were able to re-differentiate when placed back into DM, suggesting Twist2-mediated dedifferentiation was also reversible (Figure 4-4D). Taken together, these data suggest that Twist2 is a potent but reversible inhibitor of myogenesis. We believe these results are generalizable to Twist1 as well since Twist1 and Twist2 both impair myogenic differentiation (Figure 4-4E). Additionally, both Twist1 and Twist2 overexpressing myoblasts closely resemble the refractile morphology of RD cells, a well-known RMS cell line (Figure 4-4F), providing additional in vitro evidence for the role of TWIST proteins in RMS pathogenesis.

Twist2 recognizes a conserved E-box and double-E-box motif

Our previously published data suggested that Twist2 represses the expression of MyoD target genes, but not that of MyoD itself (Liu et al. 2017). To explore Twist2 binding dynamics in the context of differentiation, we performed genome-wide ChIP-Seq for 3xTy1-Twist2 and MyoD in Tw2+ myoblasts in GM and DM conditions (Figure 4-5A). We used a *Twist2*-overexpression system to model the role of *TWIST2* copy-number amplification and

over-expression in RMS patients. Previously published data suggest that endogenous and exogenous bHLH transcription factors bind similar sites in vitro (Yao et al. 2013). Peak calling quality was validated by principle component analysis and correlation matrix of samples (Supplemental Figure 4-5A, B). Using MEME suite, we performed de novo motif analysis on Twist2 occupied peaks to identify the Twist2 consensus motif. We found that the most enriched and significant primary motif in both GM and DM was a canonical CANNTG E-box motif with a preference for GA or GC as the third and fourth nucleotides (Figure 4-5B). We also identified a double E-box motif as an enriched binding sequence for Twist2 (Figure 4-5C). The double E-box consisted of the primary Twist2 binding motif followed by a secondary E-box 5 bp upstream or downstream (Figure 4-5C, D). This 5 bp-separated double E-box motif has previously been identified in both human and *Drosophila* cells as an evolutionarily conserved motif recognized by the WR domain of Twist1 (Chang et al. 2015). The Twist2 WR domain is mostly identical with that of Twist1, which may explain why Twist2 also recognizes double E-boxes (Figure 4-6C). To explore the potential role of double-E-box motifs, we performed gene ontology analysis on Twist2 target genes with and without 5 bp separated double-E-box motifs. While single E-box containing genes were enriched for developmental pathways, genes associated with double E-boxes were enriched for pathways regulating cellular migration, adhesion, and muscle development (Figure 4-5E; Figure 4-6D).

Next, we explored where Twist2 binds throughout the genome. The majority of Twist2 peaks in both GM and DM occurred in intronic and intergenic regions, which is

reflective of many bHLH transcription factors (Figure 4-5F). This is likely due to the prevalence of E-boxes scattered across the genome but could also suggest a role for Twist2 in chromatin remodeling at distal regulatory regions. Additionally, a significant fraction of Twist2 peaks are found at promoter regions, which likely indicate direct Twist2 transcriptional targets.

Twist2 is a direct transcriptional activator of EMT and a direct repressor of myogenesis

Overexpression of Twist2 in myoblasts caused drastic morphological changes in DM but not in GM, suggesting differential Twist2-binding in GM and DM conditions (Supplemental Figure 4-3D, E) (Liu et al. 2017). Through our ChIP-Seq analysis, we identified both up- and down- regulated Twist2 peaks in DM compared to GM (Figure 4-7A). The up-regulated peaks are associated with genes involved in developmental, metabolic, and muscle pathways, as well as extracellular matrix (ECM), cell-cell junction, and cytoskeletal compartments (Figure 4-7B, C). In the past, Twist2 has been viewed as a transcriptional repressor, however more recent studies have suggested roles for TWIST2 in activating gene transcription, particularly in pathways involving EMT (Shi et al. 2014; Franco et al. 2011).

To determine whether Twist2 functions as an activator or repressor in muscle cells, we intersected our Twist2 ChIP-Seq data set with our previously published *Twist2*-overexpression RNA-Seq data set (Liu et al. 2017); Figure 4-7D). We identified Twist2 promoter peaks from Twist2 ChIP-Seq and mapped them to the nearest genes to identify

putative Twist2 direct targets (Figure 4-7D). We then analyzed our RNA-Seq data to determine whether these genes were up- or down-regulated upon Twist2 over-expression (Figure 4-8A). We found that Twist2 acts as both a direct activator and repressor (Figure 4-8A). Through gene ontology analysis, we found that Twist2 directly activates genes involved in developmental, cellular adhesion, and ECM remodeling pathways, while repressing those involved in muscle development and function (Figure 4-8B, C).

In addition to its roles in development, Twist2 is also a master regulator of EMT. Previously, we showed that Twist2 overexpression upregulated Snai2 (also known as Slug) (Liu et al. 2017). Snai2 is a known regulator of RMS pathogenesis (Xu et al. 2018), mediator of TWIST1 induced EMT (Casas et al. 2011; Rinon et al. 2011), and inhibitor of MYOD (Soleimani et al. 2012). We found that the *Snai2* locus contains multiple Twist2 binding sites and is highly upregulated upon Twist2 over-expression (Figure 4-8D). A hallmark of EMT is the ability of cells to invade through a basement membrane. Using a transwell migration assay, we found that over-expression of Twist2 was capable of enhancing the cellular invasiveness of myoblasts to levels comparable with RD cells (Figure 4-8E). We also examined muscle gene loci, where Twist2 presumably acts to inhibit differentiation. Analysis of the *Myogenin* locus revealed ample Twist2 binding and subsequent downregulation of its expression (Figure 4-8F). Taken together, these data confirm Twist2 as a direct activator of development and EMT, and a repressor of myogenesis.

Twist2 drives global redirection of MyoD DNA binding

Given that Twist2 binds to MyoD target genes and represses their transcription, we investigated whether Twist2 impairs the binding of MyoD to these genomic sites. We first analyzed MyoD peaks in Tw2+ myoblasts by ChIP-Seq. Similar to Twist2, we found that MyoD exhibited preferences for intronic and intergenic regions (Figure 4-9A). De novo motif analysis revealed that MyoD preferred its canonical GC and CC containing E-boxes (Figure 4-10A;(Cao et al. 2010). Unlike Twist2, MyoD peaks did not enrich for any significant secondary motifs, including the double E-box motif found with Twist1 and Twist2 (Figure 4-10A). Protein sequence alignment of MyoD with the WR domains of Twist1 and Twist2 shows little to no identity, indicating that double E-box binding may be a function unique to the Twist family transcription factors (Figure 4-6B). Previous studies have suggested that bHLH transcription factors bind both private and shared E-box sequences (Fong et al., 2012; 2015). Comparison of our Twist2 and MyoD motifs reveals that both proteins bind to a common GC E-box (Figure 4-10B), however, MyoD also binds to a private CC E-box, while Twist2 binds to a private GA E-box (Figure 4-10B). While the GA E-box likely recruits Twist2 to genes involved in EMT and ECM remodeling, the overlap of Twist2 and MyoD at the GC E-box suggests that Twist2 and MyoD may directly compete for binding sites that are important for myogenic regulation (Cao et al. 2010; Shield et al. 1996). Indeed, decreased MYOD binding to GC E-boxes relative to CC E-boxes was observed in RMS cells (MacQuarrie et al. 2013a).

Next we analyzed our global MyoD peak changes in the presence or absence of Twist2 to determine the impact of Twist2 expression on MyoD binding (Figure 4-10C). In

both GM and DM, a significant number of MyoD peaks are differentially regulated in the presence of Twist2 overexpression (Figure 4-10C, D; Figure 4-9C). Down-regulated MyoD peaks are co-occupied by Twist2, suggesting that Twist2 directly competes with MyoD (Figure 4-10D). Examples of this are shown at the *Klhl40* and *Mymk* loci where in the absence of Twist2, MyoD binds strongly to the promoters (Figure 4-10E; Figure 4-9D). However, in the presence of Twist2, MyoD binding is significantly reduced while Twist2 now occupies the same binding sites (Figure 4-10E; Figure 4-9D). On the global scale, Twist2 displaces MyoD from genes involved in cytoskeletal reorganization and muscle differentiation (Figure 4-10F). We also found that Twist2 expression enabled MyoD to bind to a significant number of previously inaccessible sites (Figure 4-10C, D). These new sites were associated with genes involved in developmental, metabolic, growth, and cancer pathways (Figure 4-10G). Interestingly, motif analysis of down-regulated MyoD peaks identified the presence of TGIF2/1, MEF2C, and TEAD2 motifs, while motif analysis of up-regulated MYOD peaks revealed the presence of NF1, BATF, RUNX2, and HOXB3 motifs (Figure 4-9E, F). These coincident motifs are similar to those identified as differentially bound MyoD sites in RD cells versus normal myotubes (MacQuarrie et al. 2013a), providing further evidence for Twist2-mediated redirection of MyoD binding.

The bHLH domain of Twist2 is critical for repressing myogenesis

The results so far support the hypothesis that DNA binding specificity is a primary determinant of Twist2 and MyoD activities (Fong et al. 2015). Since binding sequence preference is largely specified by the bHLH domain (Fong et al. 2015), we swapped the

bHLH domains of Twist2 and MyoD to generate chimeric proteins (Figure 4-11A, B). We tagged these chimeric constructs with a myc-tag and infected them into Tw2+ primary myoblasts to determine how they impacted differentiation (Figure 4-11C). After a 4-day course of differentiation, we found that cells infected with the chimera containing the MyoD bHLH region within Twist2, referred to as Twist2(M), were able to undergo normal differentiation when compared with MyoD- and GFP-infected controls (Figure 4-11D, E; Figure 4-12). However, cells infected with the chimera containing the Twist2 bHLH within MyoD, referred to as MyoD(T), failed to differentiate (Figure 4-11D, E). Additionally, the cells infected with the latter construct resembled the enlarged and fibroblastic morphology of *Twist2*-infected cells (Figure 4-11D), suggesting that the Twist2 bHLH was also sufficient to repress myogenesis and activate an EMT and ECM remodeling profile. These data show that the major determinant of Twist2 function arises from its unique binding signature, which allows it to interfere with MyoD binding at shared GC E-boxes.

Twist2 dynamically regulates global chromatin organization during myogenesis

Recent studies have identified a key role for MyoD in driving RMS growth and initiation (Tenente et al. 2017; Gryder et al. 2017). During our analysis, we noticed a significant number of MyoD peaks that were gained only in the presence of Twist2 over-expression (Figure 4-10C, D). Additionally, these peaks were enriched for the canonical MyoD GC E-box motif and were associated with genes involved in development, metabolism, growth, and cancer pathways (Figure 4-10G; Figure 4-9F). These observations suggested that Twist2 may play an additional role in chromatin remodeling to alter MyoD

accessibility to target genes. To study how Twist2 affects global chromatin changes, we analyzed the active and repressive histone modifications, H3K27ac and H3K27me3, respectively, by ChIP-Seq in cells maintained in GM and DM, as well as in the presence or absence of Twist2 (Figure 4-13A). We first validated peak calling quality through principle component analysis and matrix correlation (Figure 4-13B, C). On a global scale, we observed changes in both H3K27ac and H3K27me3 that coincided with Twist2 binding in GM and DM (Figure 4-14A). During differentiation, upregulated Twist2 peaks were associated with increased deposition of H3K27ac, while downregulated Twist2 peaks were associated with a loss of H3K27ac (Figure 4-14A). Additionally, downregulated Twist2 peaks were also associated with a slight gain of H3K27me3 marks (Figure 4-14A). Interestingly, upregulated Twist2 peaks during differentiation exhibit both H3K27ac and H3K27me3 marks, suggesting chromatin bivalency (Figure 4-14A). This finding may explain the reversibility of myogenic repression imparted by Twist2.

We also observed broad changes in chromatin accessibility during differentiation in the presence and absence of Twist2 (Figure 4-14B-D). At myogenic loci where Twist2 competes with MyoD, we observed a significant decrease in H3K27ac and a corresponding increase in H3K27me3 (Figure 4-14C, D). An example of this occurs at the *Mymk* locus where a sharp increase in H3K27ac signal was observed during normal differentiation (Figure 4-14E). In presence of Twist2, this increase was significantly reduced, suggesting that Twist2 prevents the deposition of active H3K27ac marks at myogenic loci (Figure 4-14E). Additionally, repressive H3K27me3 marks were normally eliminated at the *Mymk*

locus during differentiation but were maintained in the presence of Twist2 (Figure 4-14E). On a genome-wide level, Twist2 impairs H3K27ac deposition primarily at myogenic loci (Figure 4-14F).

In addition to competition, Twist2-overexpression also induces new MyoD peaks at metabolic and developmental loci (Figure 4-10G). These peaks are associated with Twist2 binding as well as increased H3K27ac deposition, suggesting that Twist2 opens chromatin loci of cancer related genes to enable MyoD binding (Figure 4-10G; Figure 4-14C). An example of this occurs at the *Notch3* locus (Figure 4-14G). NOTCH is a regulator of muscle stem cell quiescence (Raimondi et al. 2012), myogenesis (Figure 4-13D), and RMS differentiation (Kuang et al. 2007). In the absence of Twist2, there is little H3K27ac within the *Notch3* locus, corresponding with the absence of MyoD binding (Figure 4-14G). Upon Twist2 over-expression, there is a sharp increase in H3K27ac that coincides with Twist2 binding. Additionally, the opening of chromatin now allows MyoD to bind a previously inaccessible site (Figure 4-14G). At a genome-wide level, Twist2 induces H3K27ac at EMT and ECM remodeling loci (Figure 4-14H). In contrast to H3K27ac, H3K27me3 marks at cancer-related loci are unchanged in the presence or absence of Twist2 (Figure 4-14D). This pattern is similar to Twist2 epigenetic regulation during differentiation, where we also observed chromatin bivalency (Figure 4-14A).

In order to capture potential transcription factors that might cooperate with Twist2 to remodel the chromatin, we performed motif analysis on Twist2 peaks within up- and down-

regulated H3K27ac peaks. Those motifs were then associated with known transcription factors. Unique transcription factors with motifs found only in the lost H3K27ac peaks include MYOG, HAND1, and MEF2, while those found in gained H3K27ac peaks include TWIST1 and NFIA (Figure 4-13E, F). Shared motifs include those for AP-1, TEAD2, and RUNX (Figure 4-13E, F). Some of these motifs are also found within differentially regulated MyoD binding sites in RD cells compared to C2C12 myoblasts (MacQuarrie et al. 2013a), further implicating Twist2 as a key mediator of altered MyoD binding in RMS. Our motif analyses also allow us to speculate on the mechanism by which Twist2 induces MyoD binding. Twist2 binding at GC E-boxes associated with downregulation of H3K27ac peaks, while Twist2 binding at GA E-boxes associated with upregulation of H3K27ac (Figure 4-13E, F). However, the predominant MyoD motif in both up and down-regulated regions was a GC E-box (Figure 4-9E, F), suggesting that Twist2 binding to GA E-boxes results in opening of chromatin, which may uncover nearby GC E-boxes for MyoD to bind. Taken together, these results suggest that in addition to its roles in direct transcriptional regulation, Twist2 functions to block the opening of chromatin at myogenic gene loci while simultaneously promoting opening of chromatin at developmental and EMT loci. Additionally, these epigenetic changes can shape the binding profile of MyoD allowing it to subvert its normal physiological function in favor of new pathological functions.

Discussion

In the current study, we identified TWIST2 as a key oncogene that is amplified and highly expressed in fusion-negative rhabdomyosarcoma. Additionally, Twist2 acts as a

reversible inhibitor of myogenic differentiation with the ability to promote myotube dedifferentiation. To explore the mechanistic basis of Twist2-mediated myogenic repression, we employed ChIP-Seq and RNA-Seq techniques to generate the first comprehensive functional genomics analysis of Twist2 and its interplay with MyoD during differentiation. Our findings demonstrate that Twist2 regulates gene expression through both direct promoter/enhancer binding and histone modification. Additionally, Twist2-mediated epigenetic modifications are sufficient to redirect MyoD binding from myogenic towards oncogenic loci. This function is not only vital for maintaining an undifferentiated state in myogenic progenitors, but also plays a significant role in RMS pathogenesis.

We focused our study on Twist2 due to our discovery of a Twist2-expressing lineage of muscle progenitors (Liu et al. 2017). Our iTwist2 system showed that Twist2 is capable of reversibly repressing myogenesis and promoting myotube dedifferentiation. It is interesting to note that previously characterized regulators of myotube dedifferentiation such as Twist1, Msx1, and a variety of small molecules all play roles in cytoskeletal processes (Mastroyiannopoulos et al. 2013; Odelberg et al. 2000; Wang et al. 2015; Paliwal and Conboy 2011; Yang et al. 2014). Given the well-established role of Twist transcription factors in regulating EMT and cellular invasion, it appears that a key aspect of dedifferentiation involves cytoskeletal remodeling that is incompatible with the organization of mature sarcomeres. Interestingly, we observed that iTwist2 de-differentiated myotubes do not begin proliferating, similar to previous observations (Wang et al. 2015; Duckmanton et al. 2005). Recent studies have shown that part of the dedifferentiation process may be

mediated by apoptotic proteins and that generation of full proliferative myoblasts requires impairment of p53 and cell cycle inhibitors (Wang et al. 2015; Paliwal and Conboy 2011).

While the ability of Twist transcription factors to block MyoD-mediated differentiation has previously been shown using reporter constructs and in vitro assays (Spicer et al. 1996; Gong and Li 2002), the exact mechanism has not been explored. In particular, the increasingly evident role of epigenetics in transcription factor biology has not been investigated in the context of Twist-mediated repression of myogenesis. Our study employed ChIP-Seq and RNA-Seq to generate a comprehensive genomics profile of Twist2 function on MyoD binding and chromatin remodeling during myogenic differentiation. Our results suggest that Twist2-mediated myogenic repression occurs at multiple levels across the genome. First, we identify direct competition between Twist2 and MyoD at shared GC E-boxes, which are associated with myogenic genes (MacQuarrie et al. 2013a). This interaction results in loss of MyoD and subsequent replacement by Twist2. Second, we also identify Twist2 as a regulator of chromatin accessibility. Twist2 binding at myogenic loci was associated with loss of H3K27ac and gain of H3K27me3. Despite these global changes, Twist2-mediated myogenic repression is reversible. A likely explanation is the presence of bivalent chromatin in regions with strong Twist2 binding during differentiation. Chromatin bivalency typically enables cellular plasticity and is often associated with developmental genes (Bernstein et al. 2006). The reversibility of Twist2 is of key importance in biological contexts. In *Drosophila* adult muscle precursors and mammalian Tw2+ cells, initial Twist

expression is important for maintaining an undifferentiated state, however subsequent downregulation of Twist is necessary prior to myogenic differentiation.

Twist2 is not only a repressor of myogenesis, but an activator of EMT, cellular invasion, development, and other oncogenic pathways (Yang et al. 2006; Jung and Yang 2015; Yang et al. 2004; Franco et al. 2011; Eckert et al. 2011; Li et al. 2012). Our study corroborates this knowledge by identifying Twist2 as a direct and epigenetic regulator of these pathways through binding of its private GA E-box and double E-box motifs. We have also identified a potential role of Twist proteins in mediating RMS etiology. A prior study found that 8/15 RMS samples had high level expression of *TWIST1* (Maestro et al. 1999). Through comprehensive genomic analysis, we discovered that *TWIST* transcription factors are highly amplified and over-expressed in the majority of FNRMS patients. Additionally, knockdown of TWIST2 and TWIST1 results in significant upregulation of MYOGENIN expression and decrease in proliferation. While most cancers are conventionally targeted through exploitation of pathways regulating cell proliferation, RMS is a unique case in which tumor differentiation exists as a possible treatment strategy (Saab et al. 2011). The reversibility of Twist2-mediated myogenic repression suggests that targeting of Twist2 and other amplified oncogenes may enable RMS cells to resume normal differentiation (Xu et al. 2018). Given the difficult nature of targeting transcription factors for therapeutic treatment, additional efforts to generate comprehensive upstream and downstream TWIST2 regulatory networks may reveal additional druggable targets. Since fusion-negative RMS patients may

experience copy number amplifications of multiple drivers (Xu et al. 2018), the regulatory networks of these drivers could be intersected to identify unique targetable nodes (Gong et al. 2015).

TWIST2-induced changes in chromatin structure prevent MYOD from accessing its physiologic target genes, but also enable MYOD to bind new target genes. Recent studies have suggested that MYOD and MYF5 adopt key roles in regulating RMS growth and initiation (Tenente et al. 2017; Olguín et al. 2011). Through ChIP-Seq of ERMS cell lines, Tenente et al. observed that both factors regulate genes involved in muscle development, growth factor signaling, smooth muscle and endothelial cell biology, and TGF-beta signaling (Tenente et al. 2017). Knockdown of either MYOD or MYF5 resulted in a near complete loss of RMS proliferative potential (Tenente et al. 2017). These novel findings suggest that in ERMS, MYOD and MYF5 shift their functions from muscle differentiation towards cancer growth. In support of these findings, we found that Twist2 overexpression remodels the chromatin to shift MyoD binding towards genes involved in EMT, metabolic, smooth muscle, FGFR signaling, and growth pathways. Previous attempts to differentiate RMS by over-expressing MYOD have been unsuccessful (Tapscott et al. 1993). Our data suggest that additional MYOD is likely re-directed towards oncogenic genes in the context of RMS. Mef2c-associated sites were previously found to have poor MyoD binding in RD cells (MacQuarrie et al. 2013a). Interestingly, TWIST2-repressed H3K27ac and MYOD peaks were enriched for MEF2 family motifs. We also noted that many of the TWIST2-induced H3K27ac and MYOD peaks contain binding motifs for developmental and cancer regulators

such as RUNX, AP-1, and NFI. These motifs were also associated with differentially bound MYOD sites in RD cells relative to normal myotubes (MacQuarrie et al. 2013a). These findings implicate TWIST2 as a potential driver of epigenetic changes seen in FNRMS tumorigenesis. In FPRMS, Pax3- and Pax7-Foxo1 may redirect myogenic transcription factors to organize autoregulatory super enhancers that contribute to its unique gene signature (Gryder et al. 2017; Olguín et al. 2011). The similarity of Twist2 function in this context suggests that oncogene amplification and fusion-protein generation represent two parallel pathways in RMS etiology.

This study provides intriguing insights regarding the function of MYOD in RMS. It is tantalizing to hypothesize that MYOD and other transcription factors may adopt novel, non-physiologic roles in the context of RMS through epigenetic remodeling by oncogenic drivers like Twist2 and Pax3-Foxo1. How these novel functions contribute to RMS pathogenesis and whether they expose new vulnerabilities for therapeutic treatment will be an exciting topic for future studies. Finally, it is becoming increasingly apparent that cancer progression can occur through lineage plasticity and subversion of tissue-specific developmental pathways (Mu et al. 2017). RMS appears to be such a case where epigenetic remodeling by developmental transcription factors is vital for its formation (Gryder et al. 2017; Tenente et al. 2017; MacQuarrie et al. 2013a). Given the depth of knowledge of myogenic regulatory mechanisms and the central role of Twist in these processes, targeting of Twist-regulated pathways may be a key aspect of future RMS therapy.

Materials and Methods

Genomic analysis of RMS patient cohort

Genomic data from 258 specimens, collected from 258 patients and de-identified before use, were from sources that were detailed in a previous publication from our lab (Xu, et al. 2018). Genomics analysis of archived patient samples was approved by the UT Southwestern Medical Center (UTSW) Institutional Review Board, and genomic data used in this study has been deposited to dbGaP database under accession number phs000720. Detailed methods are available in supplemental methods.

Copy number analysis of 258 RMS patient tumors

To explore the copy-number alterations of 258 RMS patients, SNP array datasets were processed by SNP-FASST segmentation algorithm implemented in Nexus BioDiscovery software (BioDiscovery, El Segundo, CA). Significantly altered CNVs were examined by the Genomic Identification of Significant Targets in Cancer 2.0 (GISTIC 2.0) method using a default q-value of 0.25 to define statistical significance as previously described (Mermel et al. 2011). For gene expression data, RNAseq was performed using the TruSeq Stranded Total RNA LT (Illumina, San Diego, CA), with RNA integrity was determined using the Agilent 2100 Bioanalyzer. Libraries were generated and sequenced using the HiSeq 2000, producing 100 bp paired-end reads, aligned to the human reference genome (hg19) using TopHat2 (Chudnovsky et al. 2014). Transcript assembly and abundance estimation were performed by Cufflinks (Trapnell et al. 2010).

RNAi knockdown of TWIST1 and TWIST2

RD and RH18 cells were cultured in GM composed of DMEM, 10% FBS, and Pen/Strep and transfected using RNAiMAX lipofectamine (ThermoFisher) with siRNAs targeting either TWIST1 (ThermoFisher, s14523), TWIST2 (ThermoFisher, s532272), or a negative control (ThermoFisher, 4390843) at a concentration of 10 nM per well. Transfection protocol used was standard for the RNAiMax manufacturer's protocol. RNA was harvested from cells using Trizol Reagent (Invitrogen) 48 hours post-transfection. cDNA was synthesized using iScript Reverse Transcriptase (Bio-Rad). qPCR was performed to detect the levels of TWIST1 (F: 5'-AGTCTTACGAGGAGCTGCAG-3'; R: 5'-ATCTTGCTCAGCTTGTCCGA-3'), TWIST2 (F: 5'-CAGAGCGACGAGATGGACAAT-3'; R: 5'-TAGTGGGAGGCGGACATGGA-3'), MYOGENIN (F: 5'-CAGCTCCCTCAACCAGGAG-3'; R: 5'-GCTGTGAGAGCTGCATTCG-3'), and MYL1 (F: 5'-CTCTTCAGCTTCCCTGCTGT -3'; R: 5'-CAGGAGAAATGCCTCCTTGA -3').

Cell Accumulation Assay

RD (ATCC) and RH18 (P. Houghton, St. Jude Children's Research Hospital) cells were seeded into 96-well plates at a density of 5000 cells/well. The next day, cells were transfected with siRNA targeting TWIST1 (Sigma, SASI_Hs01_00048450) and TWIST2 (ThermoFisher, s532272). Cell number was measured using CyQUANT Cell Proliferation Assay Kit (ThermoFisher, C7026) on days 0 and 6.

EDU labeling

Cell proliferation of RD cells treated with siCTRL, siTWIST1, and siTWIST2 were measured using the Click-iT EdU Flow Cytometry Assay Kit (ThermoFischer, C10419) according to manufacturer's protocol.

Generation of a Dox-inducible Twist2 stable cell line.

Twist2-IRES-GFP was Gateway cloned into pCW57.1 (a gift from David Root, Addgene plasmid 41393). For lentiviral production, iTwist2 plasmid was transfected using FuGENE 6 (Roche) into Lenti-X 293T cells (Clontech). 48 hours after transfection, media were harvested and filtered through a 0.45 μm cellulose filter, and then concentrated using Lenti-X concentrator (Clontech). The lentivirus was then resuspended in 10 ml of growth medium (GM) comprising of Ham's F10, 20% Fetal Bovine Serum, 0.2% Primocin (Invivogen), and 2.5 ng μl^{-1} basic Fibroblast Growth Factor (Gibco). Tw2+ cells were infected with resuspended lentivirus containing polybrene (6 $\mu g\ ml^{-1}$) 24-hrs after seeding. 24-hrs after infection, the viral media was removed and replaced with GM for an additional 24-hrs. To generate a stable iTwist2 cell-line, 2 $\mu g\ ml^{-1}$ of puromycin was added to GM daily on infected Tw2+ cells for five days.

Generation of iTwist2-tracer system

For dedifferentiation study, we generated a tracing system using primary myoblasts. We first generated a stable iTwist2 primary myoblast line as described above. We then cloned Lox-Stop-Lox-dsRed (LSL-dsRed) into pMX-puro. For retroviral production and packaging, 15 μg of LSL-dsRed and pMSCVhygro-Cre (Addgene #34565) plasmids were

transfected separately using FuGENE 6 (Roche) into PE cells on a 10-cm tissue culture dish at a density of 10 x 10^6 cells per dish. 48 hours after transfection, the retroviral media was harvested and filtered through a 0.45 μm cellulose filter, and then concentrated using Retro-X concentrator (Clontech) according to manufacturer's protocol. The retrovirus was then resuspended in 10 ml of satellite cell growth medium (GM) and used to infect iTwist2 primary myoblasts as described above. To generate stable LNLdsRed and pMSCVhygro-Cre cell-lines, infected LSL-dsRed and pMSCVhygro-Cre cells were selected in selection media containing GM and 3 mg ml^{-1} of either G418 (Invivogen, #ant-gn-1) or hygromycin (Invivogen #ant-hg-1) 400 μg ml^{-1}, respectively. Selection media was replaced on a daily basis for 8 days to select for LSL-dsRed cells, and 6 days for pMSCV-Cre cells.

Reversible differentiation and dedifferentiation of iTwist2 cells

iTwist2 cells were cultured in GM containing 10 μg ml^{-1} of doxycycline (GM + Dox) for 24 hours. GM + Dox was then replaced with differentiation media comprised of DMEM, 2% Horse Serum, and 0.2% Primocin, and containing 10 μg ml^{-1} of doxycycline (DM + Dox). After 3 days of DM + Dox treatment, half the iTwist2 cells continued DM + Dox treatment and the other half were changed to DM only.

For dedifferentiation tracing experiments, LSLdsRed and pMSCVhygro-Cre primary myoblasts were co-cultured in GM at a 1:1 ratio on day 0. On day 2, GM was replaced with DM to begin differentiation. On day 6, cells were either kept in DM or switched to DM + Dox to begin dedifferentiation. On day 8, DM was replaced with GM while DM + Dox was replaced with GM + Dox to enhance dedifferentiation as previously described (Wang et al.

2015; Paliwal and Conboy 2011). Cells were then fixed and stained on day 10 or harvested for western blot.

Immunostaining of cultured cells

Immunostaining of cultured cells was performed as previously described (Liu et al. 2017). Primary antibodies include: fast myosin (Sigma-Aldrich, My32, 1:250). Cells were counterstained with Hoechst (ThermoFisher, Hoechst, 3342, 1:1000). Alexa Fluor secondary antibodies were used according to the manufacturer's instructions. Differentiation index was quantified as the percentage of $Myosin^{+}$ nuclei over total nuclei. Dedifferentiation was calculated as a ratio of the number of nuclei within $dsRed^{+}GFP^{+}$ monuclear cells to total nuclei at day 10. Multiple representative fields were counted from 3 biological replicates.

Live cell imaging for dedifferentiation

Primary myoblasts were transduced with the iTwist2 expression vector and either pCAG-Cre or lox-neo-stop-lox-dsRed. The two population of cells were mixed at a 1:1 ratio and differentiated for five days. Cells were subsequently induced to dedifferentiate with 10 ug/ml of doxycycline. Cells were then imaged in 6-well plates on a modified Andor Spinning Disk Microscope with Andor Neo sCMOS camera. Images were captured every 10 minutes. At 24 hours post induction, the DM was replaced with tet-free GM with and without doxycycline. Finalized timelapse images were analyzed and assembled into a movie on ImageJ.

Western Blot

All cells were harvested in RIPA buffer (Sigma) supplemented with cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail (Roche) and PhosSTOP (Sigma) phosphatase inhibitor. Western blot was performed as previously described (Makarewich et al. 2018). Primary antibodies used were: fast myosin (Sigma-Aldrich, M4276, 1:1000), GFP (Invitrogen, A-11122, 1:1000), GAPDH (Millipore, CB1001, 1:10,000), beta-tubulin (Abcam, Ab6046, 1:1000), Myogenin (SantaCruz, sc12732, 1:1000) and Myc (Invitrogen, R950-25, 1:1000). Western blots were washed in TBST, incubated with HRP-conjugated secondary antibodies (Bio-Rad), and then developed using a ChemiDoc MP Imaging System (Bio-Rad).

Dedifferentiation-Redifferentiation

Dedifferentiated iTwist2 primary myoblasts were placed into GM or DM for 4 days. Cells were then fixed and stained.

Transwell migration assay

Transwell migration assay was performed according to manufactures protocol. Ctrl-Tw2+, Twist2-Tw2+, and RD cells were first serum starved for 24 hrs before the assay. Number of cells migrated was measured using Calcein AM (ThermoFisher) and a plate reader.

Generating Twist2 and MyoD bHLH domain chimaeras

Gene block fragments (IDT) were synthesized with the swapped bHLH domains and Infusion cloned (Clontech) into pBabe-puro (Addgene, #1764) for retroviral transduction. Retrovirus packaging for GFP, myc-Twist2, myc-MyoD, myc-Twist2(M), and myc-MyoD(T) constructs was performed as previously described (Liu et al. 2017). Infected-Tw2+ cells were cultured for 24 hours in GM then differentiated in DM for 4 days. Cells were then fixed and stained for myosin.

ChIP-Seq for transcription factors and histone marks

Chromatin immunoprecipitation was performed as described in the ChIP-IT High Sensitivity Kit (Active Motif). In brief, Tw2+ cells were infected with 3xTy1-Twist2 or a GFP control and kept in GM or switched to DM 24 hours after infection. After 3 days in GM or 4 days in DM, we harvested cells and performed ChIP using antibodies recognizing Ty1 (Diagenode), MyoD (Santa Cruz, sc-32758), H3K27ac (Diagenode, C15410196), and H3K27me3 (Diagenode, C15410195). Sequencing was performed in the McDermott Sequencing Core at UT Southwestern. Data is available in the Gene Expression Omnibus under accession GSE127998. Detailed ChIP-Seq data analysis is provided in supplemental methods.

ChIP-Seq data analysis

The raw reads were aligned to the mouse reference genome (GRCh38/mm10) using default parameters in BWA v0.7.12 (Li and Durbin 2009). The aligned reads were subsequently filtered for quality and uniquely mappable reads were retained for further

analysis using Samtools version 1.3 (Li et al. 2009) and Sambamba version 0.6.6 (Tarasov et al. 2015). Library complexity was measured using BEDTools version 2.26.0 (Quinlan and Hall 2010) and meets ENCODE data quality standards (Landt et al. 2012). Relaxed peaks were called using MACS version 2.1.0 (Feng et al. 2012) with a p-value of $1x10^{-2}$ for H3K27ac and H3K27me3 and SPP version 1.4 (Kharchenko et al. 2008) with an IDR = 0.01 for each other sample. Data has been deposited into the Gene Expression Omnibus under accession GSE127998.

Peaks were called using findpeaks command from HOMER software package version 4.7. For transcription factor MyoD and Twist2 ChIP-Seq data parameter '–style factor' was used, and the FDR threshold (for poisson p-value cutoff) was set to 10^{-7}, similar to a previous study (Cao et al. 2010). For histone marker ChIP-Seq data, peaks were called by findpeaks command with parameter '–style histone' and other parameters set as default. To identify differential peaks between two samples, called peaks were merged from each sample and annotatePeaks.pl command was used to generate raw tag count matrix. Then differential peaks were identified using R package DEseq version 3.8. Peaks with > 2-fold change and padj <0.05 were designated as DE peaks.

To identify the primary motif present in peaks, we performed *de novo* motif discovery using MEME version 4.11.1 with the 'meme-chip' command (Bailey et al. 2009). Motif enrichment histograms were generated using the 'annotatePeaks.pl' script in HOMER and custom double E-box configuration motifs with variable nucleotide spacing were generated using HOMER's 'seq2profile.pl' script ((Heinz et al. 2010; 2015).

To perform principle component analysis of peaks, raw count matrix was log normalized and supplied as input for R function prcomp.

For heatmaps and profiles of ChIP-Seq intensities around interested genomic region, we used ngsplot version 2.6 command ngs.plot.r to generate read abundance from ChIP-Seq datasets around peak center (± 1 kb). These matrices were then used to create heatmaps and profiles with ngs.plot.r command from ngsplot package. The Up and Down DE peaks were randomized separately and combined together in a .bed file as input for ngs.plot.r command.

To identify potential TF motifs enriched in interested peaks regions, we used findMotifsGenome.pl command from HOMER software package, using peak region .bed file as input, with parameter '–size 200' to search for motif enrichment in 200bp window surrounding peak center.

To analyze the functional significance of bind peaks, Genomic Regions Enrichment of Annotations Tool (GREAT) was used with mm10 as the background genome and other parameters set as default.

For the GO term analysis of DE genes identified from RNA-Seq data, we used the Metascape online service.

RNA-Seq

RNA-Seq data was obtained from our previously published study and is available in the Gene Expression Omnibus (GEO) under accession codes GSE84377, GSE84378, GSE84379, and GSE 84380 (Liu et al. 2017).

Statistical Analysis

All quantitative data are displayed as mean ± SEM. Unpaired two-tailed t test was performed for all analyses: * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.00005$.

Figures

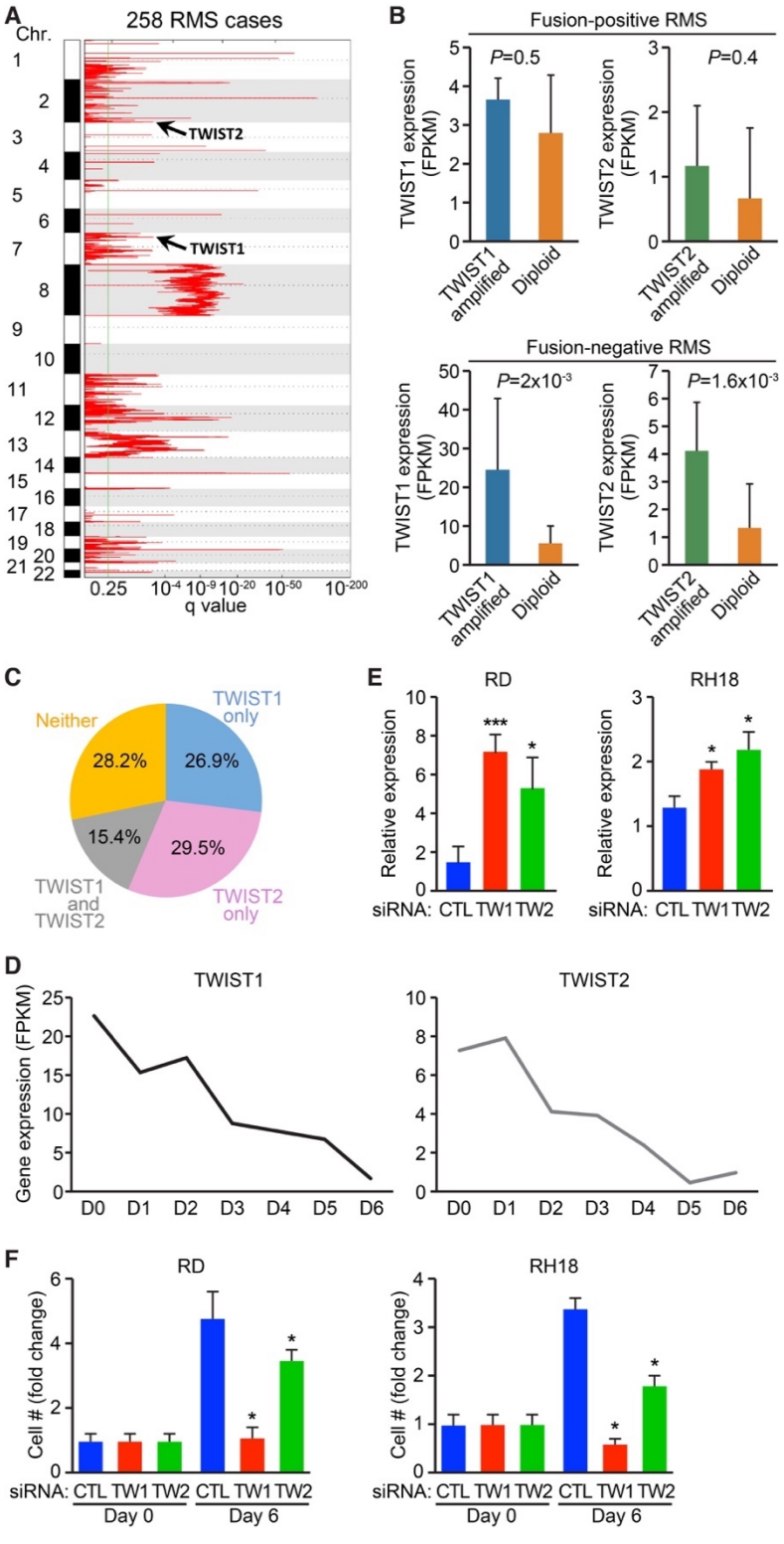

A
258 RMS cases
Chr.
TWIST2
TWIST1
q value
B
Fusion-positive RMS
P=0.5
P=0.4
TWIST1 expression (FPKM)
TWIST2 expression (FPKM)
TWIST1 amplified
Diploid
TWIST2 amplified
Fusion-negative RMS
P=2x10-3
P=1.6x10-3
C
Neither
TWIST1 only
28.2%
26.9%
15.4%
29.5%
TWIST1 and TWIST2
TWIST2 only
D
TWIST1
TWIST2
Gene expression (FPKM)
D0 D1 D2 D3 D4 D5 D6
E
RD
RH18
Relative expression
siRNA: CTL TW1 TW2
F
RD
RH18
Cell # (fold change)
siRNA: CTL TW1 TW2 CTL TW1 TW2
Day 0
Day 6

Figure 4-1. (**A**) Charts show genomic regions with statistically significant copy number gains identified by Genomic Identification of Significant Targets in Cancer 2.0 (GISTIC2) algorithm in 258 RMS cases. Additional amplified loci can be found in (Xu, et al. 2018) (**B**) Separate integrated analysis on copy-number alterations and gene expression for TWIST1 and TWIST2 genes in fusion-positive (upper panel) and fusion-negative (lower panel) RMS. TWIST expression in patient tumors containing TWIST amplification were compared to those without TWIST amplification (diploid). (**C**) Venn diagrams display the percentage of RMS cases with copy-number amplification events on TWIST1 (blue), or TWIST2 (pink), or both (grey). (**D**) Charts show gene expression for TWIST1 and TWIST2 in human myoblasts cultured in growth medium (D0) or differentiation medium for the indicated number of days (D1-D6). Measurement of MYOGENIN expression in RD cells (left) or RH18 cells (right) **(E)** or cell proliferation in RD cells (left) or RH18 cells (right) **(F)** upon siRNA-mediated knockdown of *TWIST2* or *TWIST1*. Unpaired two-tailed t test: * $p < 0.05$; *** $p < 0.0005$.

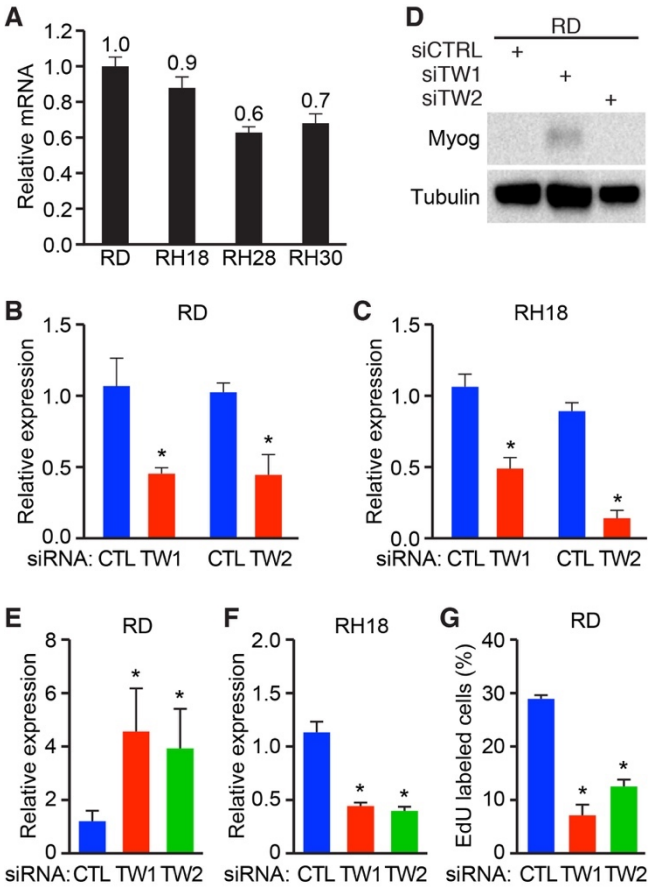


Figure 4-2. (**A**) qRT-PCR of *TWIST2* expression in various RMS cell lines. Validation of siRNA-mediated *TWIST1 and TWIST2* knockdown by qRT-PCR in **(B)** RD cells or (**C**) RH18 cells. (**D**) Western blot depicting MYOGENIN protein levels in RD cells after TWIST1 and TWIST2 knockdown. qRT-PCR of *MYL1* expression upon *TWIST1* and *TWIST2* knockdown in (**E**) RD cells or (**F**) RH18 cells. (**G**) EdU labeling of RD cells after knocking down *TWIST1* or *TWIST2*. (how long is the EdU labeling?)

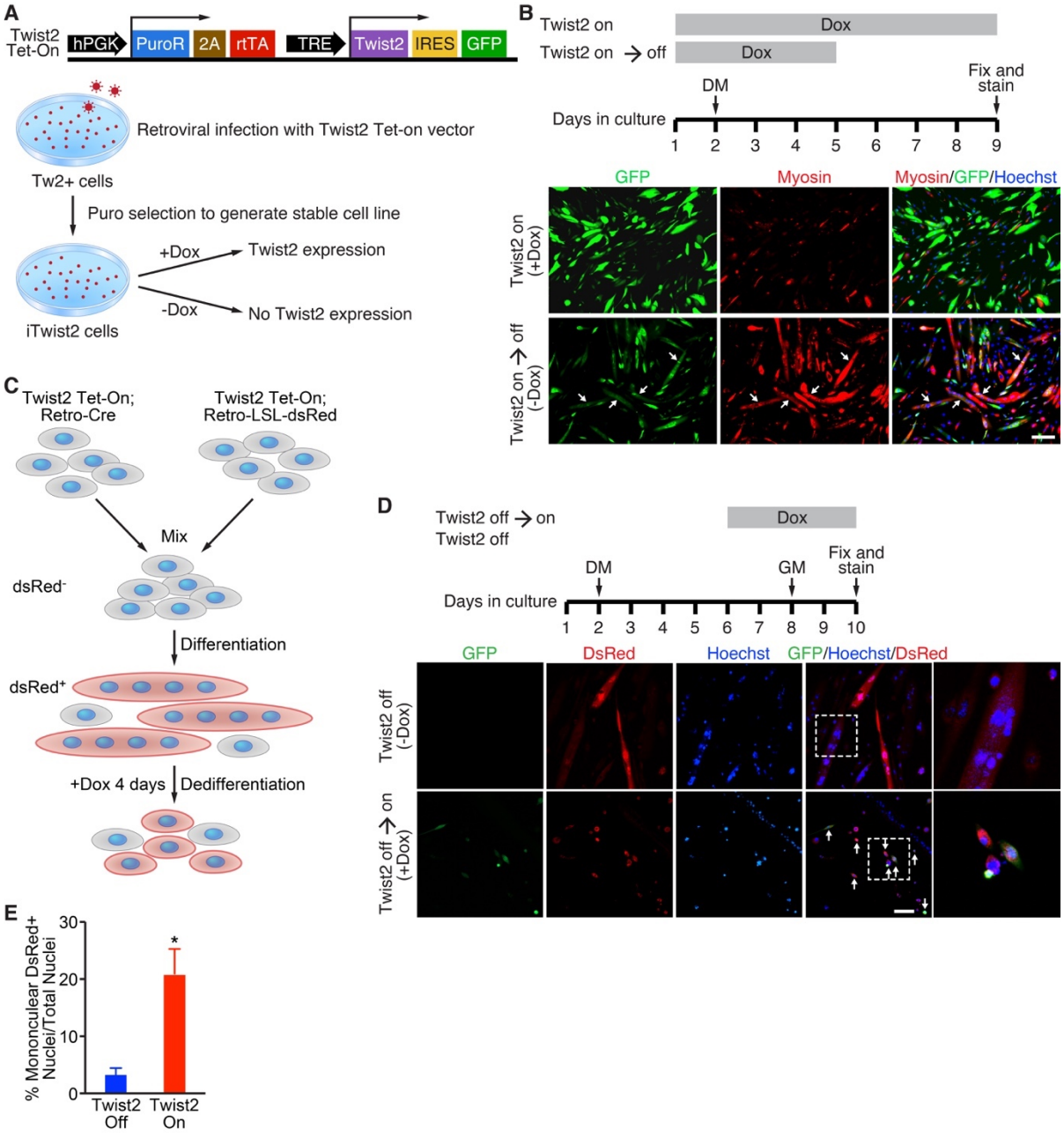


Figure 4-3. (**A**) (Top) Schematic of Twist2 Tet-On vector. (Bottom) Generation of iTwist2 stable cell line in Twist2+ myoblasts. (**B**) Experimental scheme of Twist2 reversibly inhibiting differentiation. Dox was added to iTwist2 cells to induce Twist2-IRES-GFP expression. On day 2, growth media (GM) was replaced with differentiation media (DM). On

day 5, Dox was removed to shut off Twist2 expression. Cells were fixed and stained on day 9 for fast myosin (red), GFP (green), and nuclei (blue). Arrows refer to differentiated myotubes that previously expressed Twist2. Scale bar = 100 μm (**C**) Schematic of dedifferentiation tracing system using iTwist2 primary myoblasts. (**D**) Time-line of dedifferentiation tracing experiment. Cells were mixed on day 1 and differentiated on day 2. On day 6 of differentiation, Dox was added to induce Twist2 expression. On day 8 cells were switched back to GM to enhance dedifferentiation. Cells were fixed and stained on day 10 for fast myosin (red), GFP (green), and nuclei (blue). Arrows indicate $dsRed^{+}$ dedifferentiated mononuclear cells. Scale bar = 50 μm. (**E)** Quantification of myotube dedifferentiation. The number of nuclei within mononuclear $dsRed^{+}$ cells were quantified as a percentage of total nuclei. For each condition, 2-3 fields were quantified for each of three biological samples. Unpaired two-tailed t test: * $p < 0.05$.

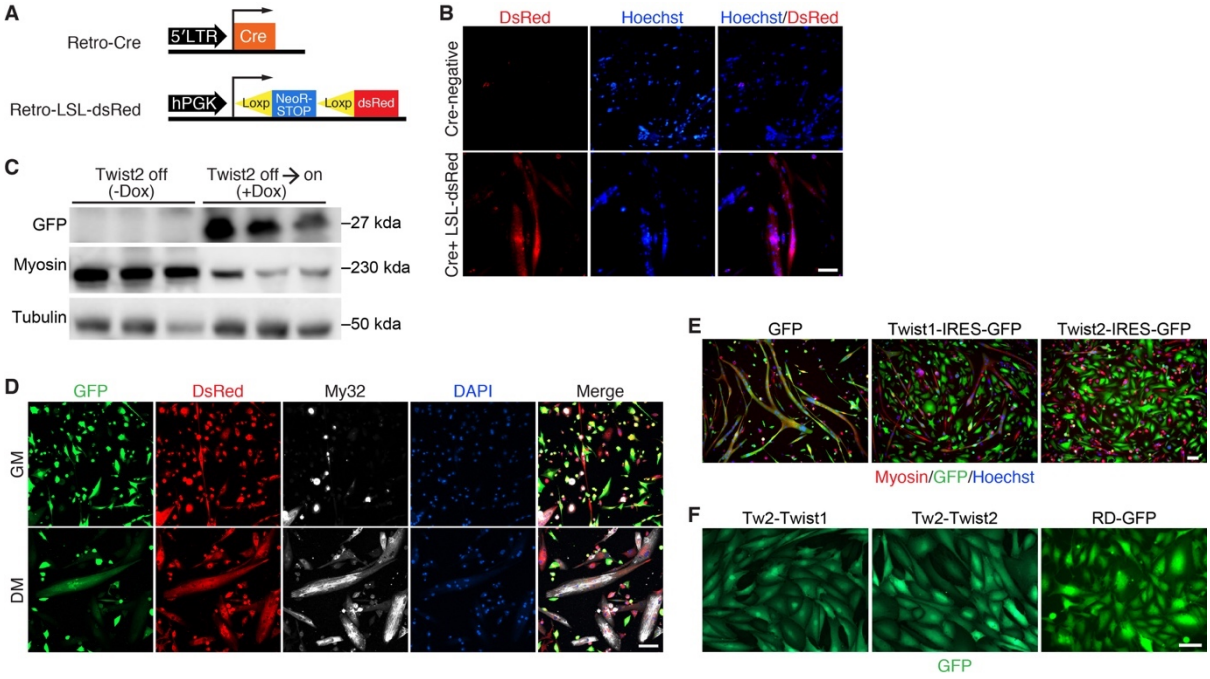


Figure 4-4. (A) Schematic of retroviral constructs used for lineage-tracing of myoblast dedifferentiation. **(B)** iTwist2 myoblasts stably expressing LSL-dsRed were either differentiated in DM alone or mixed with iTwist2 myoblasts stably expressing Cre and then differentiated. DsRed was only expressed in cells expressing both Cre and LSL-dsRed, not in cells expressing LSL-dsRed alone, confirming lack of Cre leakage. Scale bar = 50 µm. **(C)** Western blot of dedifferentiated myoblasts. Protein lysates from iTwist2 cells cultured in the presence or absence of Dox were collected and blotted with antibodies for GFP, Myosin, and Tubulin. GFP served as a marker of *Twist2* induction. Myosin was used as a marker for myogenic differentiation. Tubulin was used as a loading control. (**D**) Dedifferentiated iTwist2 primary myoblasts were incubated in GM or DM for 4 days to induce redifferentiate. Cells were then fixed and stained for GFP, Myosin, and DAPI. Scale bar = 50 µm. **(E)** Primary myoblasts were infected with GFP, Twist1-IRES-GFP, or Twist2-IRES-GFP and

differentiated for 4 days. Cells were then fixed and stained for fast myosin (red), GFP (green), or Hoechst (blue). Scale bar = 100 μm. **(F)** Tw2+ myoblasts infected with either Twist1-IRES-GFP (left) or Twist2-IRES-GFP (center), and RD cells infected with GFP (right) were differentiated for 4 days. Cells were subsequently fixed and stained for GFP (green). Scale bar = 100 μm.

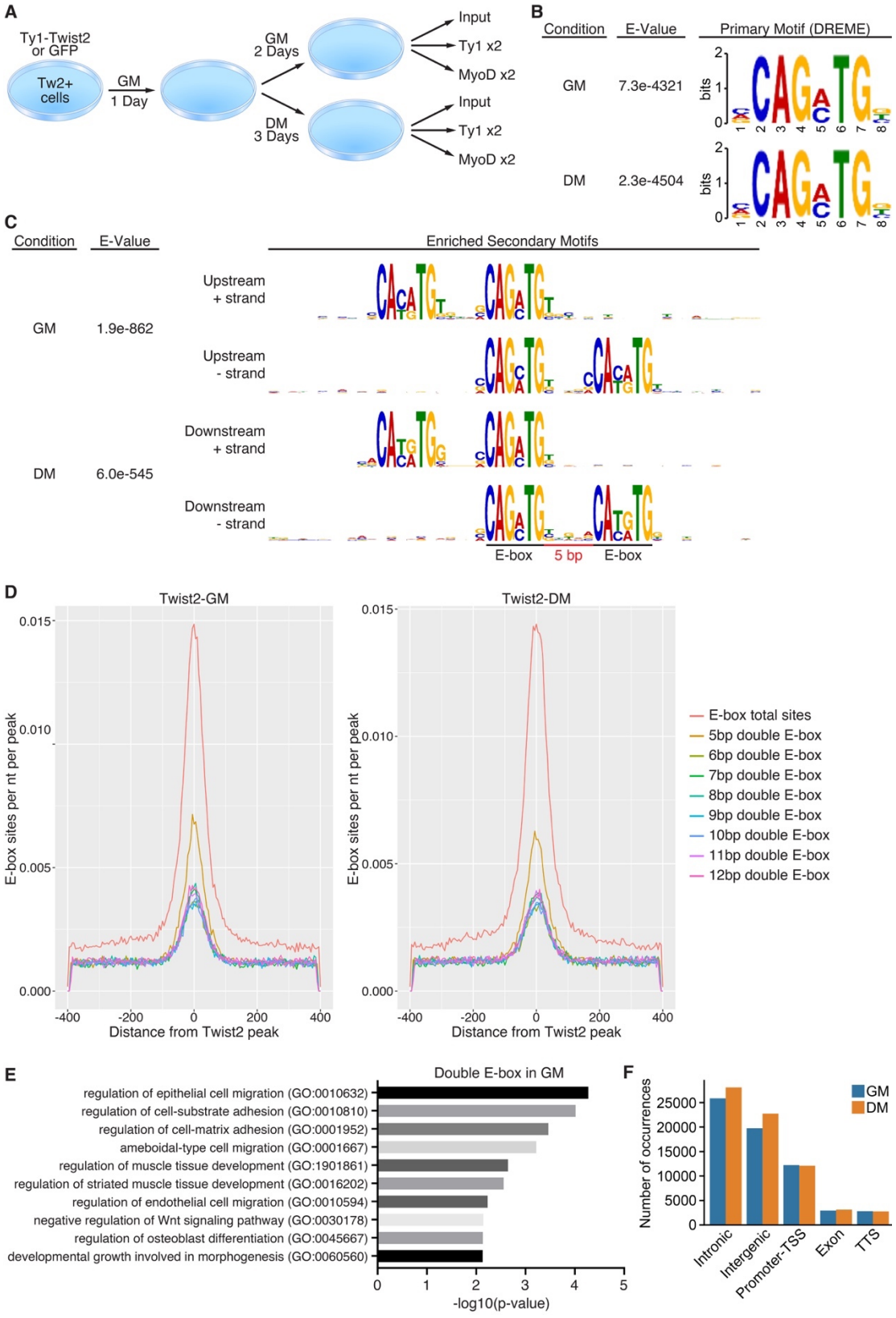

A
Ty1-Twist2 or GFP
Tw2+ cells
GM 1 Day
GM 2 Days
DM 3 Days
Input
Ty1 x2
MyoD x2
Input
Ty1 x2
MyoD x2
B
Condition
E-Value
Primary Motif (DREME)
GM
7.3e-4321
DM
2.3e-4504
bits
C
Condition
E-Value
Enriched Secondary Motifs
GM
1.9e-862
DM
6.0e-545
Upstream + strand
Upstream - strand
Downstream + strand
Downstream - strand
E-box
5 bp
E-box
D
Twist2-GM
Twist2-DM
E-box sites per nt per peak
Distance from Twist2 peak
E-box total sites
5bp double E-box
6bp double E-box
7bp double E-box
8bp double E-box
9bp double E-box
10bp double E-box
11bp double E-box
12bp double E-box
E
Double E-box in GM
regulation of epithelial cell migration (GO:0010632)
regulation of cell-substrate adhesion (GO:0010810)
regulation of cell-matrix adhesion (GO:0001952)
ameboidal-type cell migration (GO:0001667)
regulation of muscle tissue development (GO:1901861)
regulation of striated muscle tissue development (GO:0016202)
regulation of endothelial cell migration (GO:0010594)
negative regulation of Wnt signaling pathway (GO:0030178)
regulation of osteoblast differentiation (GO:0045667)
developmental growth involved in morphogenesis (GO:0060560)
-log10(p-value)
F
Number of occurrences
GM
DM
Intronic
Intergenic
Promoter-TSS
Exon
TTS

Figure 4-5. (A) Schematic of ChIP-Seq for 3xTy1-Twist2 and MyoD in Tw2+ cells in GM or DM. **(B)** De novo motif analysis of the primary Twist2 binding motif in GM and DM performed by MEME-ChIP. **(C)** De novo motif analysis of a secondary double E-box Twist2 binding motif in GM and DM performed by MEME-ChIP. **(D)** Relative incidence of Twist2 bound to single versus double E-boxes of various spacing. **(E)** Panther pathway analysis of Twist2-bound genes associated with 5 bp double E-box motifs. **(F)** Twist2 binding distribution across the genome. TSS: Transcription start site.

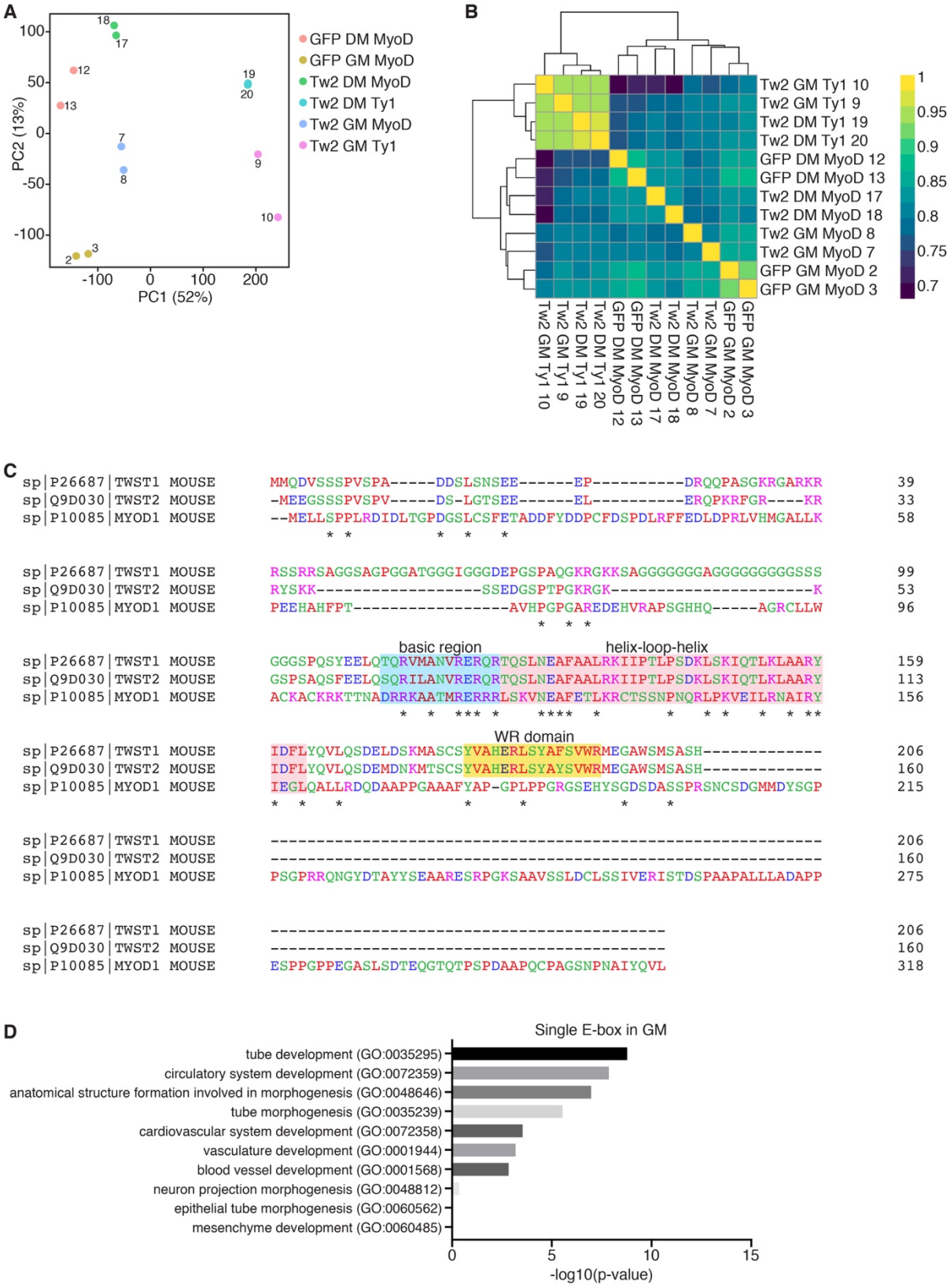

A
PC2 (13%)
PC1 (52%)
GFP DM MyoD
GFP GM MyoD
Tw2 DM MyoD
Tw2 DM Ty1
Tw2 GM MyoD
Tw2 GM Ty1
B
Tw2 GM Ty1 10
Tw2 GM Ty1 9
Tw2 DM Ty1 19
Tw2 DM Ty1 20
GFP DM MyoD 12
GFP DM MyoD 13
Tw2 DM MyoD 17
Tw2 DM MyoD 18
Tw2 GM MyoD 8
Tw2 GM MyoD 7
GFP GM MyoD 2
GFP GM MyoD 3
C
basic region
helix-loop-helix
WR domain
sp|P26687|TWST1_MOUSE
sp|Q9D030|TWST2_MOUSE
sp|P10085|MYOD1_MOUSE
D
Single E-box in GM
tube development (GO:0035295)
circulatory system development (GO:0072359)
anatomical structure formation involved in morphogenesis (GO:0048646)
tube morphogenesis (GO:0035239)
cardiovascular system development (GO:0072358)
vasculature development (GO:0001944)
blood vessel development (GO:0001568)
neuron projection morphogenesis (GO:0048812)
epithelial tube morphogenesis (GO:0060562)
mesenchyme development (GO:0060485)
-log10(p-value)

Figure 4-6. (A) Principle component analysis (PCA) plot of MyoD and Twist2 ChIP-Seq data using the first 2 Principle Components. Biological replicates are closer to each other than with other conditions. **(B)** Correlation coefficient matrix of MyoD and Twist2 ChIP-Seq data. Biological replicates are clustered together. **(C)** Protein sequence alignment of MyoD, Twist1, and Twist2. The basic helix-loop-helix regions and the Twist1 and Twist2 WR domains are highlighted. **(D)** Panther pathway analysis of genes associated with Twist2 peaks containing a single E-box.

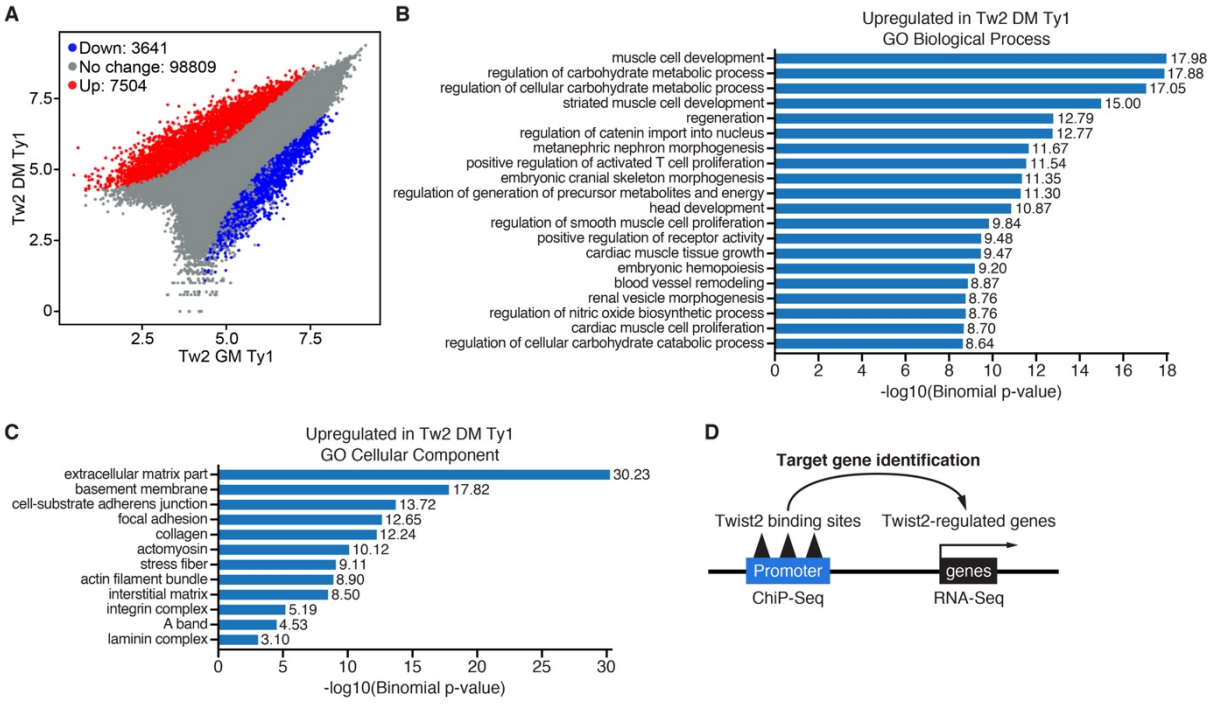


Figure 4-7 (A) Differential binding analysis of Twist2 peaks in DM versus GM conditions. Cut-off for differentially bound peaks was set at fold change greater than or equal to 2 and a $p_{adj} < 0.05$. **(B)** GO biological process terms enriched for peaks with increased Twist2 binding during differentiation. **(C)** GO cellular component terms enriched for peaks with increased Twist2 binding during differentiation. **(D)** Schematic showing the mapping of direct Twist2 target genes.

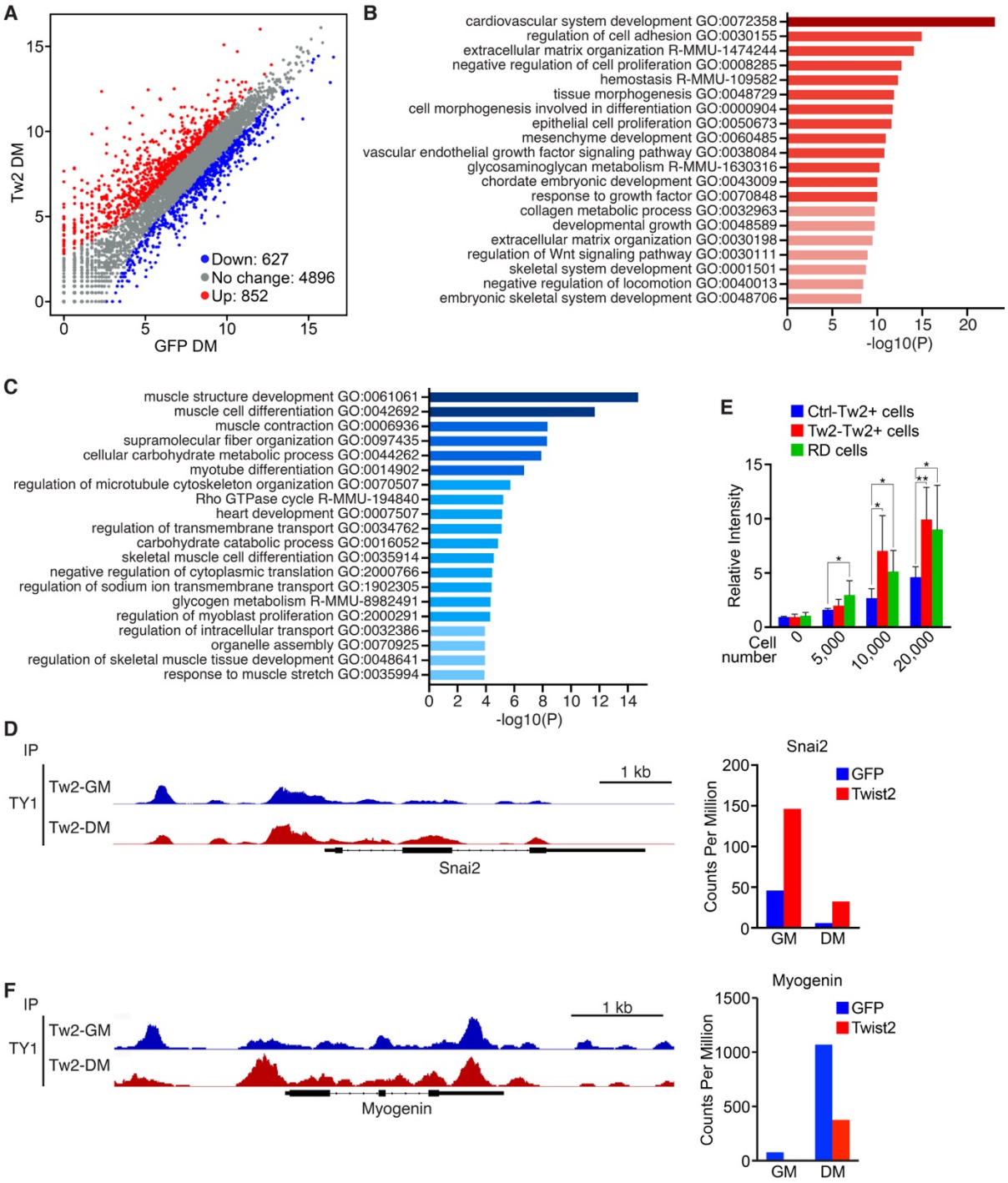
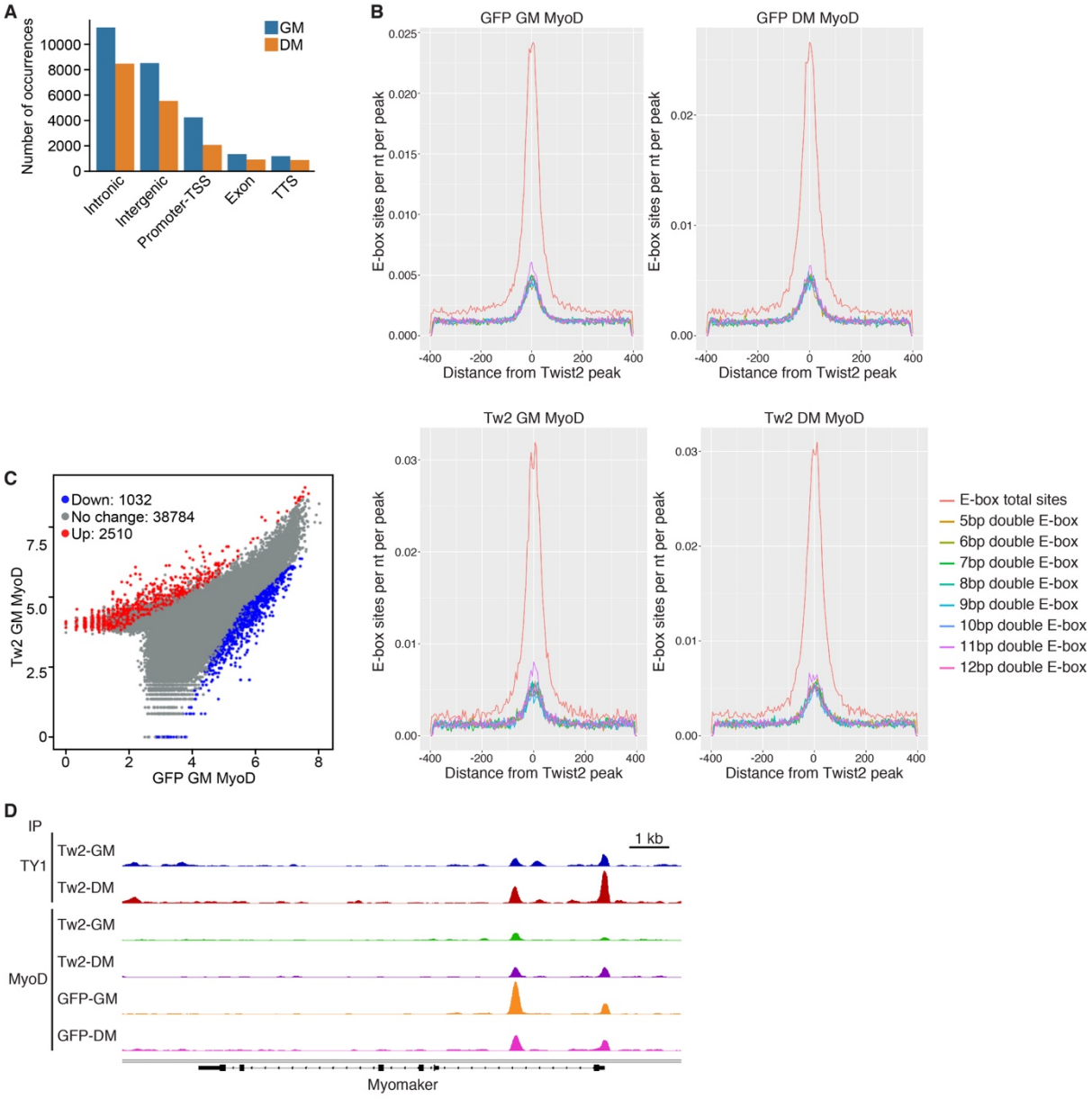


Figure 4-8. Twist2 is a direct transcriptional activator of EMT and a direct repressor of myogenesis.

(A) Scatter plot of Twist2 direct target gene expression identified by ChIP-Seq and RNA-Seq intersection in DM. The cut off was ≥ 2-fold change and p_{adj} <0.05. **(B)** GO (Gene Ontology) enrichment analysis of target genes activated by Twist2 as shown in A. **(C**) GO enrichment analysis of genes repressed by Twist2 as shown in A. **(D)** (Left) Genome browser shot displaying Twist2 binding within the *Snai2* locus. (Right) Upregulation of *Snai2* expression induced by Twist2. **(E)** Transwell migration assay. Ctrl-infected Tw2+ cells and RD cells were used as negative and positive controls, respectively. The number of invading cells was quantified as a relative intensity compared to 0 cells 24 hours after seeding using Calcein AM. Unpaired two-tailed t test: * $p < 0.05$; ** $p < 0.005$ **(F)** (Left) Genome browser shot displaying Twist2 binding within the *Myog* locus. (Right) Downregulation of *Myog* expression induced by Twist2.



E

Motif	Best match	% of targets	p-value
	MyoG(bHLH)	88.69%	$1e^{-1980}$
	Tgif2 (homeobox)	55.13%	$1e^{-220}$
	Tgif1	25.01%	$1e^{-154}$
	Mef2c(Mads)	19.48%	$1e^{-90}$
	TEAD2	14.27%	$1e^{-50}$

F

Motif	Best match	% of targets	p-value
	MyoG(bHLH)	54.17%	$1e^{-590}$
	BATF(bZIP)	12.51%	$1e^{-82}$
	NF1	30.34%	$1e^{-43}$
	RUNX2(Runt)	14.52%	$1e^{-34}$
	HOXB3	18.23%	$1e^{-50}$

Figure 4-9. (A) Distribution of MyoD peaks across the genome. TSS: Transcription start site. **(B)** Incidence of MyoD bound single versus double E-boxes of various spacing in TW2-GM, TW2-DM, GFP-GM, and GFP-DM conditions. **(C)** Scatter plot depicting differential binding of MyoD peaks in Tw2-GM versus GFP-GM conditions. Cut-off for differentially bound peaks was set at fold change greater than or equal to 2 and a $p_{adj} < 0.05$. **(D)** Genome browser shot depicting Twist2 and MyoD competitive binding at the *Mymk* locus. **(E)** Motif analysis of MyoD peaks down-regulated in presence of Twist2. Motifs were matched with previously known motifs. **(F)** Motif analysis of MyoD peaks up-regulated in presence of Twist2. Motifs were matched with previously known motifs.

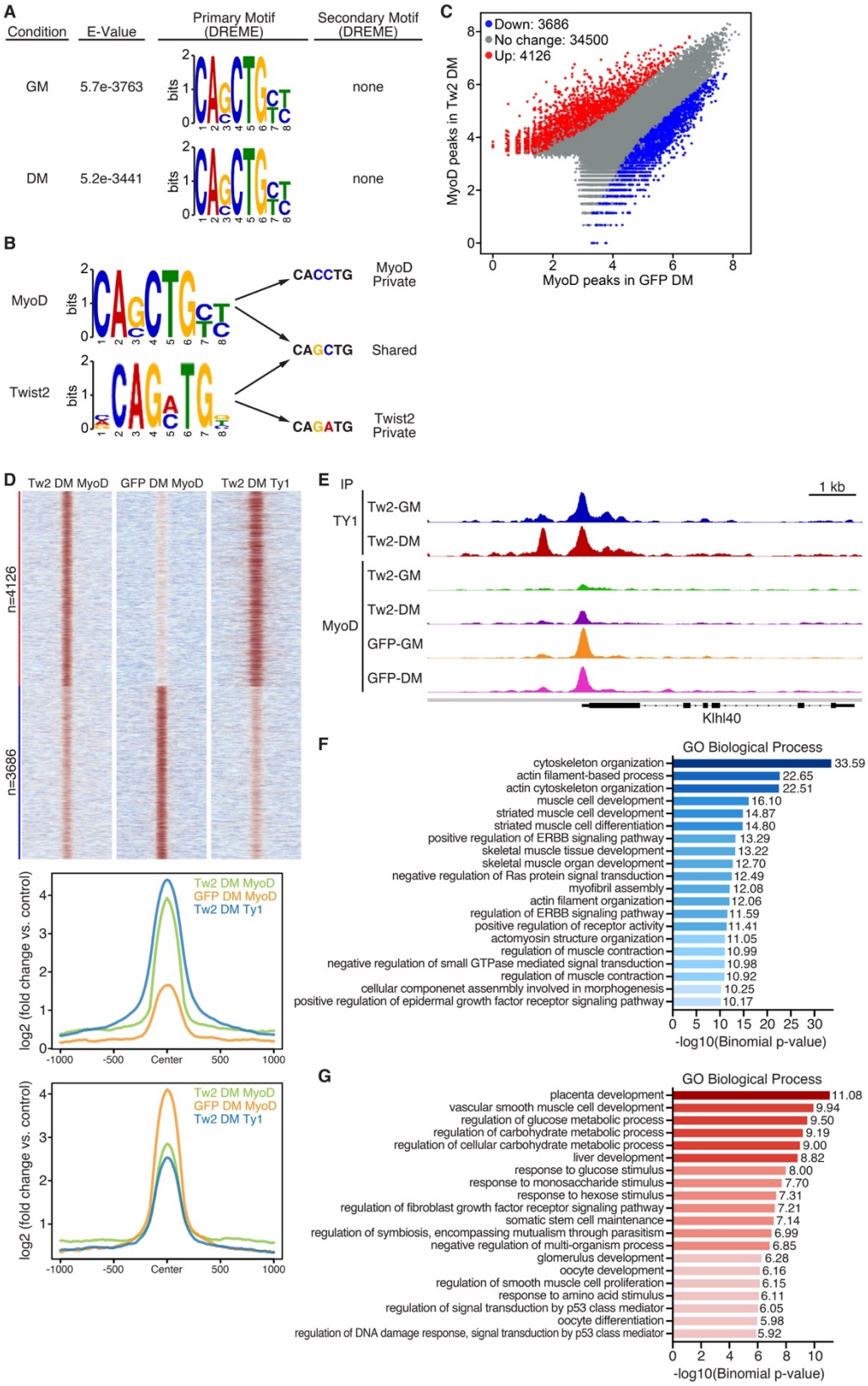

A
Condition
E-Value
Primary Motif (DREME)
Secondary Motif (DREME)
GM
5.7e-3763
none
DM
5.2e-3441
none
bits
B
MyoD
Twist2
CACCTG
MyoD Private
CAGCTG
Shared
CAGATG
Twist2 Private
C
Down: 3686
No change: 34500
Up: 4126
MyoD peaks in Tw2 DM
MyoD peaks in GFP DM
D
Tw2 DM MyoD
GFP DM MyoD
Tw2 DM Ty1
n=4126
n=3686
log2 (fold change vs. control)
-1000
-500
Center
500
1000
E
IP
1 kb
TY1
Tw2-GM
Tw2-DM
MyoD
GFP-GM
GFP-DM
Klhl40
F
GO Biological Process
cytoskeleton organization 33.59
actin filament-based process 22.65
actin cytoskeleton organization 22.51
muscle cell development 16.10
striated muscle cell development 14.87
striated muscle cell differentiation 14.80
positive regulation of ERBB signaling pathway 13.29
skeletal muscle tissue development 13.22
skeletal muscle organ development 12.70
negative regulation of Ras protein signal transduction 12.49
myofibril assembly 12.08
actin filament organization 12.06
regulation of ERBB signaling pathway 11.59
positive regulation of receptor activity 11.41
actomyosin structure organization 11.05
regulation of muscle contraction 10.99
negative regulation of small GTPase mediated signal transduction 10.98
regulation of muscle contraction 10.92
cellular componenet assenmbly involved in morphogenesis 10.25
positive regulation of epidermal growth factor receptor signaling pathway 10.17
-log10(Binomial p-value)
G
GO Biological Process
placenta development 11.08
vascular smooth muscle cell development 9.94
regulation of glucose metabolic process 9.50
regulation of carbohydrate metabolic process 9.19
regulation of cellular carbohydrate metabolic process 9.00
liver development 8.82
response to glucose stimulus 8.00
response to monosaccharide stimulus 7.70
response to hexose stimulus 7.31
regulation of fibroblast growth factor receptor signaling pathway 7.21
somatic stem cell maintenance 7.14
regulation of symbiosis, encompassing mutualism through parasitism 6.99
negative regulation of multi-organism process 6.85
glomerulus development 6.28
oocyte development 6.16
regulation of smooth muscle cell proliferation 6.15
response to amino acid stimulus 6.11
regulation of signal transduction by p53 class mediator 6.05
oocyte differentiation 5.98
regulation of DNA damage response, signal transduction by p53 class mediator 5.92
-log10(Binomial p-value)

Figure 4-10. (A) De novo motif analysis of the MyoD binding motif in GM and DM performed by MEME-ChIP. MyoD has no preference for secondary motifs. **(B)** Comparison of Twist2 and MyoD private and shared E-box variants. **(C)** Scatterplot of differential MyoD peaks in the presence and absence of Twist2 in DM. Differential peak cut-off was defined as signal ≥ 2-fold change and p_{adj} <0.05. **(D)** (Top) Heatmap depicting differential binding of MyoD peaks in the presence or absence of Twist2 in DM. (Bottom) ChIP signal distribution plot of differentially bound MyoD peaks in the presence or absence of Twist2. **(E)** Genome browser shot depicting Twist2 and MyoD competitive binding at the *Klhl40* locus. **(F)** GREAT (Genomic Region Enrichment of Annotations Tool) analysis of MyoD peaks downregulated MyoD in TW2-DM vs GFP-DM. **(G)** GREAT analysis of MyoD peaks upregulated in TW2-DM vs GFP-DM.

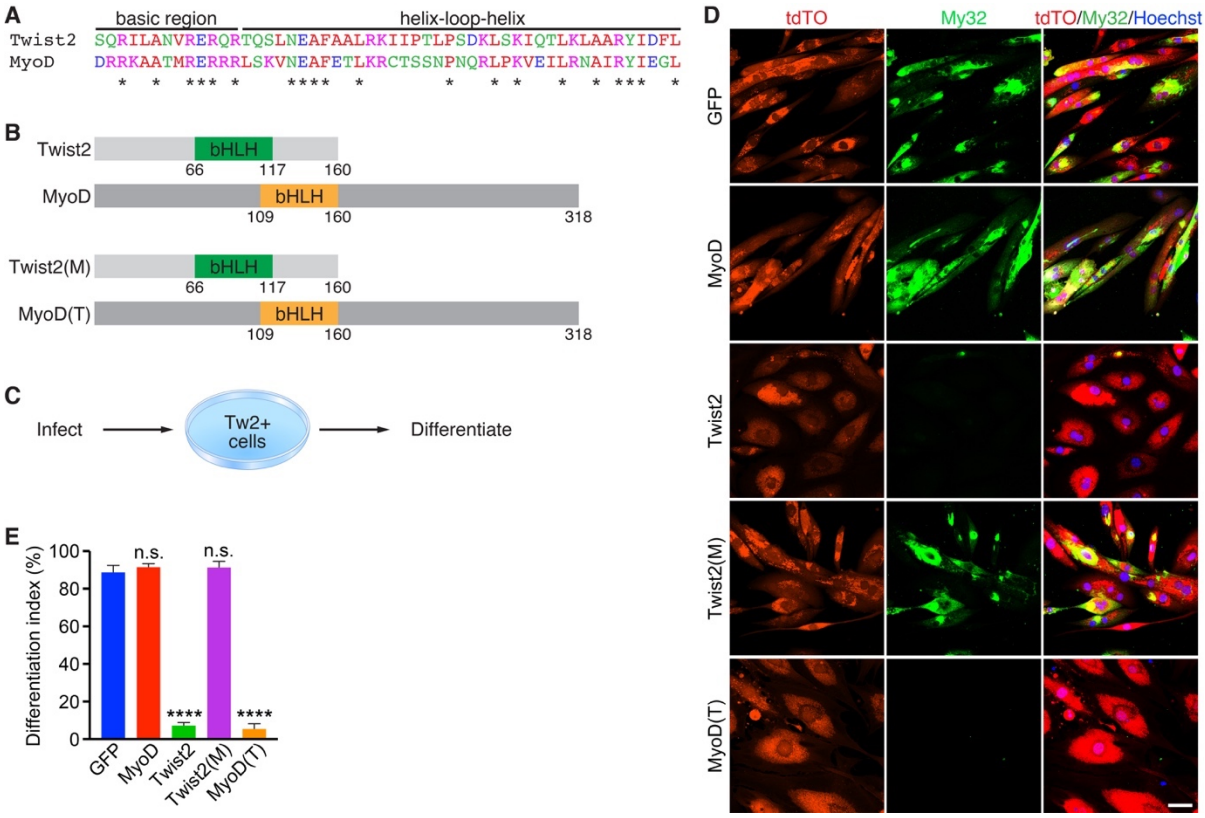


Figure 4-11. (A) Sequence alignment of Twist2 and MyoD bHLH domains. * represents identical amino acids. **(B)** Chimeric Twist2 and MyoD bHLH constructs. Numbers represent amino acid number. **(C)** Experimental scheme to test the effect of Twist2 and MyoD chimeras on differentiation. **(D)** Tw2+ myoblasts infected with GFP, MyoD, Twist2, Twist2(M) or MyoD(T) were differentiated for 4 days and subsequently fixed and stained for My32 (green), tdTomato (red), and nuclei (blue). Scale bar = 50 μm **(E)** Number of differentiated cells were quantified as a differentiation index based on the percent of nuclei within $My32^{+}$ cells compared to total nuclei. Unpaired two-tailed t test: **** $p < 0.00005$.

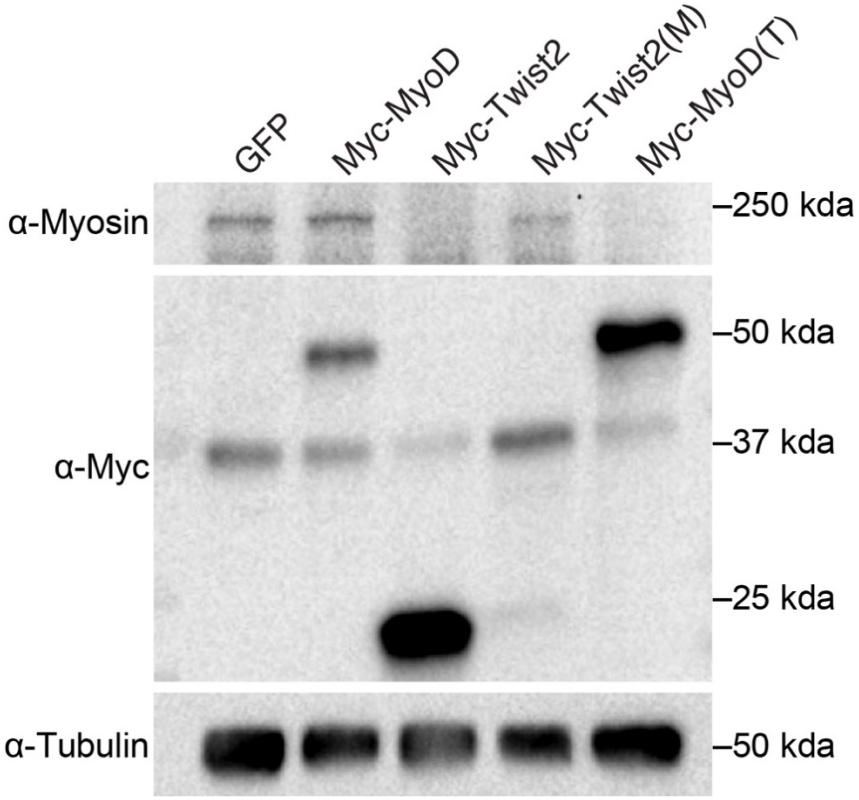


Figure 4-12. Tw2+ myoblasts infected with GFP, myc-MyoD, myc-Twist2, myc-Twist2(M) or myc-MyoD(T) were differentiated for 4 days. Cells were then harvested for western blot analysis. Cell lysates were blotted for myosin, myc, and tubulin.

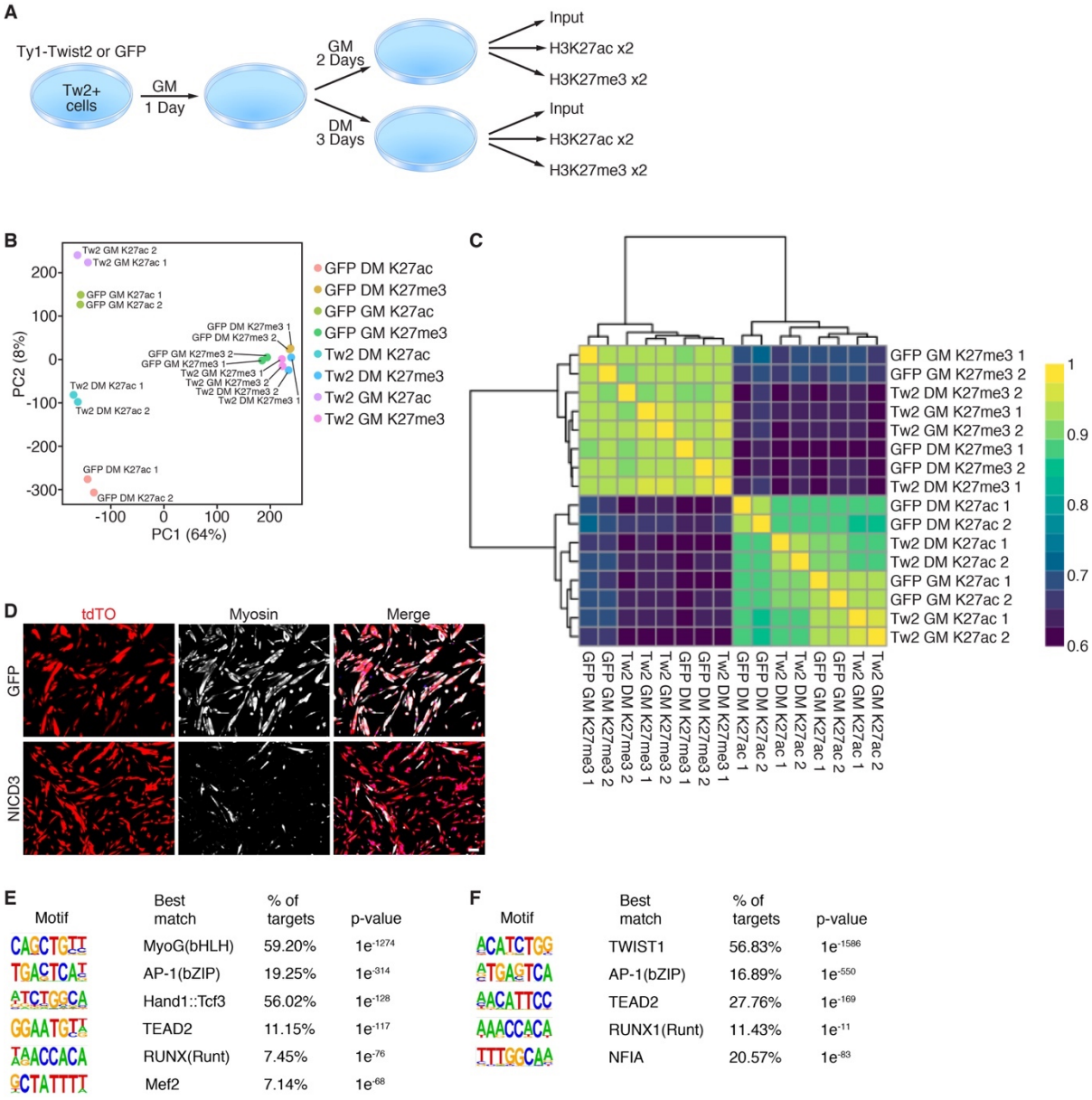


Figure 4-13. **(A)** Schematic of ChIP-seq for H3K27ac and H3K27me3 in Tw2-GM, Tw2-DM, GFP-GM and GFP-DM conditions. **(B)** Principle component analysis (PCA) plot of H4K27ac and H4K27me3 ChIP-Seq data using first 2 Principle Components. Biological replicates are closer to each other than with other conditions. **(C)** Correlation coefficient

matrix of H4K27ac and H4K27me3 ChIP-Seq data. Biological replicates are clustered together. **(D)** Overexpression of Notch3 intracellular domain (NICD3) in Tw2+ myoblasts. Tw2+ myoblasts were infected with retroviral constructs containing NICD3 or GFP and subsequently differentiated for 4 days. Cells were the fixed and stained for fast myosin (white) and tdTomato (red). Scale bar = 100 μm. **(E)** Motif analysis of Twist2 peaks within down-regulated H3K27ac peaks in Tw2-DM vs GFP-DM. Motifs were then associated with known motifs of transcription factors obtained through ChIP-Seq experiments. **(F)** Motif analysis of Twist2 peaks within up-regulated H3K27ac peaks in Tw2-DM vs GFP-DM. Motifs were matched with previously known motifs.

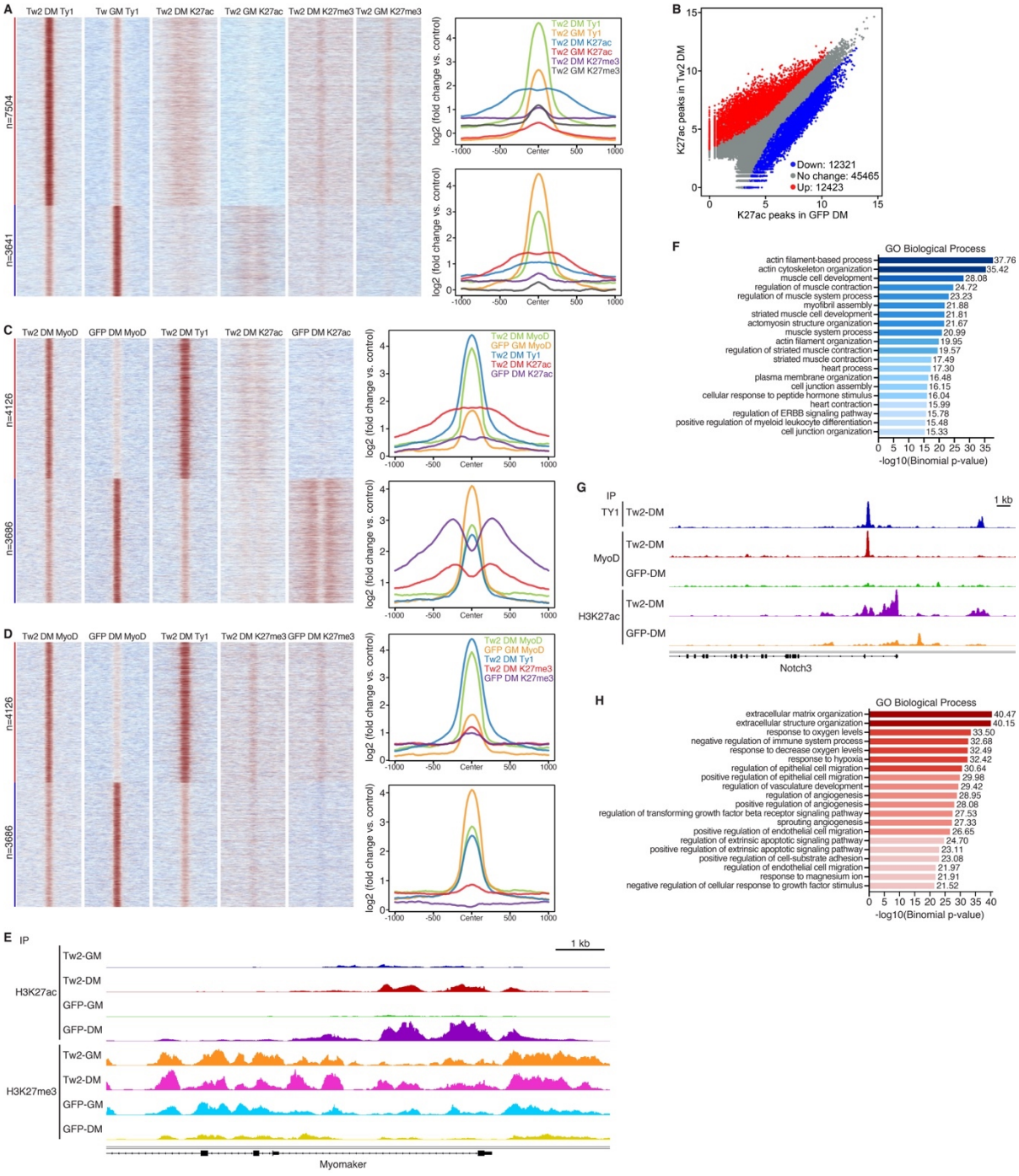


Figure 4-14. **(A)** (Left) Heatmap depicting differential binding of TWIST2 in GM vs DM conditions and the effect on H3K27ac and H3K27me3. (Right) ChIP signal distribution plot of TWIST2, H3K27ac, and H3K27me3 peaks associated with up- and down-regulated

TWIST2 peaks in DM versus GM conditions. **(B)** Scatterplot depicting differential binding of H3K27ac in the presence and absence of TWIST2 in DM. Differential binding cut-off was set at a fold change ≥ 2 and p_{adj} <0.05. **(C)** (Left) Heatmap depicting the effect of TWIST2 on differential binding of MYOD and H3K27ac. (Right) ChIP signal distribution plot of TWIST2, MYOD, and H3K27ac peaks associated with up- and down-regulated MYOD peaks in TW2-DM versus GFP-DM. **(D)** (Left) Heatmap depicting the effect of TWIST2 on differential binding of MYOD and H3K27me3. (Right) ChIP signal distribution plot of TWIST2, MYOD, and H3K27me3 peaks associated with up- and down-regulated MYOD peaks in TW2-DM versus GFP-DM. **(E)** Genome browser shot depicting H3K27ac and H3K27me3 changes at the *Mymk* locus during differentiation in the presence and absence of TWIST2. **(F)** GREAT analysis of H3K27ac peaks downregulated in TW2-DM vs GFP-DM. **(G)** Genome browser shot depicting gain of H3K27ac and MYOD peaks in the *Notch3* locus upon TWIST2 binding. **(H)** GREAT analysis of H3K27ac peaks upregulated in TW2-DM vs GFP-DM.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

Function of Twist2$^+$ Cells in Disease and Aging

From cellular identity to transcription regulation in homeostasis and disease, the work presented in this dissertation provides a comprehensive and multifaceted exploration of Twist2 in muscle biology. Our discovery of a novel fiber-type specific muscle progenitor unlocked a previously unexplored area of within muscle biology. While the role and significance of these Tw2$^+$ cells are yet to be understood, the unique biology of these Tw2$^+$ cells provides clues to address this question. In particular, preferential atrophy and loss of fast-twitch fibers in aging and disease is a phenomenon that is actively explored yet remains poorly understood (Lee et al. 2016). Despite numerous studies analyzing the role of hormonal and neuronal changes in fast-twitch fiber atrophy, our study provides evidence for cellular fusion as a novel mechanism of fast-twitch fiber maintenance (Schiaffino and Reggiani 2011; Liu et al. 2017). This hypothesis is supported by our data showing that ablation of Tw2$^+$ cells results in type IIb myofiber atrophy (Liu et al. 2017). Given that Tw2$^+$ cells are myogenic and fuse to myofibers, it is an appealing hypothesis that atrophy during aging could be a result of Tw2$^+$ cell functional or numerical decline. This hypothesis is made more appealing given that satellite cells, which fuse to all fiber-types, do not appear to play a significant role during the aging process in regard to muscle fusion and contribution (McCarthy et al. 2011). Instead they appear to adopt a more supportive role through remodeling of extracellular matrix (Fry et al. 2015). Even if their fusion to muscle played a

significant role during the aging process, it would still not explain why some fibers are preferentially atrophied. To address this question, future studies should be focused on exploring $Tw2^+$ cell number in young and aged muscle. Additionally, functional analysis can also be carried out young and old mice to analyze myogenic capacity or potential secreted factors that may lead to fiber-type specific changes.

In parallel to muscle atrophy, myofiber size is also regulated in a fiber-type specific manner (Schiaffino and Reggiani 2011). Type IIb fibers have the largest cross-sectional areas among myofiber types, suggesting that mechanisms exist which enable selective contribution towards these fibers (Schiaffino and Reggiani 2011). In this position, $Tw2^+$ cells may serve a prime role in promoting and maintaining type IIb fiber size during normal homeostasis. Additionally, type IIb fibers are highly responsive to weight bearing exercises, which suggests that $Tw2^+$ cells may play a role in mediating exercise related hypertrophy (Fry 2004). In order to address the role of $Tw2^+$ cells in growth, it will be interesting to examine their function and numbers in sedentary and exercised mice.

The mechanism of fiber-type specificity is also a major question addressed by my thesis work. My identification of a Sema3a-Nrp1 signaling mechanism suggests that there are intrinsic fiber-type specific signaling mechanisms that can result in differential cross-talk among cells in the muscle interstitium. During aging and disease, these signaling mechanisms could also be altered resulting in non-cell autonomous effects on interstitial muscle cells like the $Tw2^+$ cells. Perhaps a mechanism for type IIb fiber atrophy could be the upregulation of Sema3a, which results in the inability of $Tw2^+$ cells to fuse. Targeting of Sema3a may therefore be a therapeutic strategy to restore $Tw2^+$ cell fusion. A broader

implication of this work may be the role of Sema3a in other aspects of muscle biology. Similar to ephrins and ephrin receptors, differential expression of Sema3a by fast-twitch myofibers could also result in differential responses to other Nrp1 expressing cell populations in the muscle. Such populations include neuronal and endothelial cells, both of which use chemoattractive and chemorepulsive signaling process to regulate their migration and branching (Acevedo et al. 2008). A fundamental difference between slow- and fast-twitch fibers is their unique metabolic and contractile properties (Schiaffino and Reggiani 2011). Both of these properties can be downstream of both vascularization and innervation, suggesting that Sema3a may even play a major role in the identity of the fiber-itself (Schiaffino and Reggiani 2011). This is supported by a recent finding that Sema3a expression by satellite cells during injury is important for maintaining a slow fiber identity (Tatsumi et al. 2017). Whether $Tw2^{+}$ cells are merely an evolutionary remnant, or are vital players in physiologic and pathologic process will be an exciting topic for future exploration. These questions will require a thorough understanding and identification of $Tw2^{+}$ cell identity.

Resolving the Muscle Interstitial Soup

The identification of $Tw2^{+}$ cells sparked a significant level of controversy. Among the most prominent attitudes was a disbelief in not having identified these cells before. In this section, I hope to resolve some of these questions and controversies, and tie together the seemingly contradictory findings of the muscle field in regard to muscle interstitial cell populations.

Myogenic interstitial populations have traditionally been identified in two manners. The first is through flow cytometry using a variety of cell surface markers like ITGA7, VCAM1, and PDGFRa. Cells like myoendothelial cells (Zheng et al. 2007), $CD133^+$ cells (Torrente et al. 2004), PW1 interstitial cells (Mitchell et al. 2010), muscle-derived stem cells (Qu-Petersen et al. 2002), and muscle side population cells (Majka et al. 2003) were initially identified in this manner. The second method of identification revolved around lineage tracing of specific genes using Cre-loxP recombination system. $NG2^+$ pericytes(Dellavalle et al. 2007) and $Tw2^+$ cells (Liu et al. 2017) were both identified in this manner. All of these populations were shown to have myogenic potential in vitro and in vivo, as well as to be distinct from satellite cells. Could it be that the muscle interstitium is really this heterogeneous, or is there something else at play?

They key to addressing this question lies in the nature of the method used in identification. In all the above studies, one or a small handful of unique markers were used to separate the population of interest from satellite cells. Therefore, it is possible that all of these myogenic populations are actually the same population, identified through multiple methods. If this is the case, then why do transcriptional analyses of these different populations not completely intersect? Using $Tw2^+$ cells as a case-study, we can begin to understand why this may be the case. Our identification of this cell population revolved around lineage tracing of the transcription factor Twist2. Through flow cytometry and transcriptomics analysis, we identified $Tw2^+$ cells as a transcriptionally distinct population from satellite cells, endothelial cells, fibroblasts, and other muscle interstitial cells (Liu et al. 2017). However through clonal analysis of the $Tw2^+$ cell population, we found that in fact,

only a minority of $Tw2^+$ cells harbored myogenic potential (Liu et al. 2017). This result suggested that our Twist2-CreERT2 lineage tracing system was labeling a heterogenous population of cells.

One can easily envision that all previous identification of myogenic and multipotent muscle progenitor populations also run into the same issue of cellular heterogeneity. The recent advent of single cell-RNA sequencing has been revolutionary in the identification of heterogenous cell populations. In particular, published data from the Tabula Muris consortium and Fabien Le Grande's group as well as our own unpublished data are finally able to shed light on the conflicting reports of interstitial muscle progenitors and their identities. First, the Tabula Muris single cell data of muscle reveals the entire interstitial cell population of mouse hind limb muscle (Tabula Muris Consortium et al. 2018). In contrast to the numerous reports of unique interstitial muscle progenitors, only two unidentified populations exist (Tabula Muris Consortium et al. 2018). In support of our own published data, both populations appear to be Twist2-expressing, suggesting that at least some of the $Tw2^+$ lineage may be contained within these unidentified cell populations. Additional clarity can be drawn from a single cell study of muscle hindlimb from the Le Grande group (Giordani et al. 2019), which sought to functionally characterize the unknown cell populations. They subsequently classified these two unknown populations into interstitial tenocyte cells and smooth muscle-mesenchymal cells (Giordani et al. 2019). While the tenocytes do not exhibit myogenic potential, they do show an enrichment for osteogenic genes. This fits well with our own data that a subpopulation of $Tw2^+$ cells are capable of differentiating into the osteogenic lineage. On the other hand, smooth muscle-mesenchymal

cells harbor intrinsic myogenic potential (Giordani et al. 2019). Surprisingly, this cell population also preferentially forms type IId/x myofibers in vitro (Giordani et al. 2019). This fiber type is one that $Tw2^{+}$ cells robustly fuse with (Liu et al. 2017). The data presented here suggest that $Tw2^{+}$ cells may actually be comprised of both muscle tenocyte cells as well as this smooth muscle-mesenchymal cell. In particular, the smooth muscle-mesenchymal cell may actually be the specific population that is type IIb fiber-specific. In support of this, our unpublished single-cell RNA sequencing of $Tw2^{+}$ cells reveals the presence of both of these populations in the Twist2 lineage. In addition, we also find a subpopulation of endothelial cells and fibroadipogenic progenitors that are also present in the Twist2-lineage.

Taken together, these results begin to finally paint a clear picture of what is happening in the muscle interstitium. It is possible that all previous studies reporting the discovery of interstitial muscle progenitors are actually referring to the same cell: the "smooth-muscle mesenchymal cell". Additionally, since we were the only group that performed lineage tracing of this population, we were the first to identify its unique fiber-type specific properties. It will be important to identify unique markers of smooth-muscle mesenchymal cells in order to perform lineage tracing to observe if there are any fiber-type specific preferences for fusion. The resolution and clarification of these data will aid immensely in both clearing up residual confusion and controversy regarding interstitial cell types, as well as enabling groups to begin studying the biological importance of these cells.

Integrating Big Data to Tackle Rhabdomyosarcoma

Beyond the role of Twist2 in muscle physiology, it will be important to explore its role in pathology. I showed that both Twist1 and Twist2 are highly amplified in the context of fusion-negative RMS, where their role is to redirect MyoD binding to impair myogenic differentiation and promote oncogenesis (Li et al. 2019). Currently, the RMS field lacks clear and consistent knowledge on genetic alterations that may explain pathogenesis and guide therapy of FNRMS (Shern et al. 2014). The high incidence of Twist amplification suggests that Twist-regulated pathways may be key targets for future therapies (Li et al. 2019). However, the recent surge in large scale genomic sequencing enables an even greater degree of precision when it comes to identifying therapeutic targets. As I showed during my thesis work, integrated analysis of RNA-seq and ChIP-seq data enables the identification of direct transcriptional targets of amplified oncogenes such as Twist2 (Li et al. 2019). These data allow us to separate the immediate effects of Twist2 amplification from its secondary effects. Beyond analysis of just Twist2, similar experiments can be performed with all of the commonly amplified oncogenes in FNRMS. Additionally, intersection of these pathways will enable the identification of common pathways that may be vital for tumor growth. Exploiting these shared pathways may be critical for developing a multi-faceted therapy for RMS treatment.

In addition to mechanism of pathogenesis, the cell of origin for RMS is still highly debated. This is an additional area where single-cell RNA sequencing may finally shed light onto when and where RMS originates. By performing single-cell RNA sequencing on human tumors, we can generate a comprehensive map of the tumor environment that contains a diverse and heterogenous population of tumor and stromal cells. In particular, the

heterogeneity of RMS cell mutations may enable us to generate a pseudo-time map of RMS clonogenicity and mutational evolution. By tracing backwards along this pseudo-time trajectory, we may be able to pinpoint the cell of origin as well as tumor initiating mutations. These results can also be compared to single-cell RNA sequencing of RMS derived from various mouse models to identify which models are the most representative of RMS pathogenesis. Additionally, data obtained from patient RMS sequencing can also be used to guide the generation of new mouse models which accurately model the clinical presentation.

Conclusions and Future Perspectives

The development of these new technologies represents a turning point in our ability to address fundamental questions in muscle biology. They serve as powerful tools to resolve existing controversies to a greater resolution and to generate novel areas of scientific exploration. The use of these tools will enable us to revisit past topics and may revitalize the fields of developmental and cell biology. In fact, this resurgence is already happening as numerous labs turn towards high resolution sequencing to study development. It is an exciting time to be a developmental biologist and I recommend any prospective students deciding on areas of academic pursuit to give developmental and molecular biology, and the Olson lab a strong consideration. The sandbox of tools empowers curious students with the ability to make significant discoveries. However, given the rapid generation of large-scale data, it will be of utmost importance for students to think deeply about interesting and unique biological hypotheses. At the end of the day, it is important to remember that these new technologies are merely tools and that tools always need a purpose.

Bibliography

Acevedo LM, Barillas S, Weis SM, Göthert JR, Cheresh DA. 2008. Semaphorin 3A suppresses VEGF-mediated angiogenesis yet acts as a vascular permeability factor. *Blood* **111**: 2674–2680.

Alto LT, Terman JR. 2017. Semaphorins and their Signaling Mechanisms. *Methods Mol Biol* **1493**: 1–25.

Anderson JE, Do M-KQ, Daneshvar N, Suzuki T, Dort J, Mizunoya W, Tatsumi R. 2016. The role of semaphorin3A in myogenic regeneration and the formation of functional neuromuscular junctions on new fibres. *Biol Rev Camb Philos Soc*.

Arany, Z. PGC-1 coactivators and skeletal muscle adaptations in health and disease. Curr. Opin. Genet. Dev. 18, 426–434 (2008).

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–8.

Barnes RM, Firulli AB. 2009. A twist of insight - the role of Twist-family bHLH factors in development. *Int J Dev Biol* **53**: 909–924.

Baylies MK, Bate M. 1996. twist: a myogenic switch in Drosophila. *Science* **272**: 1481–1484.

Beck, B. et al. Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. Cell Stem Cell 16, 67–79 (2015).

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315–326.

Bi P, Ramirez-Martinez A, Li H, Cannavino J, McAnally JR, Shelton JM, Sanchez-Ortiz E, Bassel-Duby R, Olson EN. 2017. Control of muscle formation by the fusogenic micropeptide myomixer. *Science* **356**: 323–327.

Bodine SC, Baehr LM. 2014. Skeletal muscle atrophy and the E3 ubiquitin ligases MuRF1 and MAFbx/atrogin-1. *Am J Physiol Endocrinol Metab* **307**: E469–84.

Brack, A. S. & Rando, T. A. Tissue-specific stem cells: lessons from the skeletal muscle satellite cell. Cell Stem Cell 10, 504–514 (2012).

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, et al. 2010. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* **18**: 662–674.

Chal J, Pourquié O. 2017. Making muscle: skeletal myogenesis in vivo and in vitro. *Development* **144**: 2104–2122.

Chang AT, Liu Y, Ayyanathan K, Benner C, Jiang Y, Prokop JW, Paz H, Wang D, Li H-R, Fu X-D, et al. 2015. An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors. *Genes Dev* **29**: 603–616.

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

Chen ZF, Behringer RR. 1995. twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev* **9**: 686–699.

Chudnovsky Y, Kim D, Zheng S, Whyte WA, Bansal M, Bray M-A, Gopal S, Theisen MA, Bilodeau S, Thiru P, et al. 2014. ZFHX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell Rep* **6**: 313–324.

Ciciliot S, Rossi AC, Dyar KA, Blaauw B, Schiaffino S. 2013. Muscle type and fiber type specificity in muscle wasting. *Int J Biochem Cell Biol* **45**: 2191–2199.

Cripps, R. M. & Olson, E. N. Twist is required for muscle template splitting during adult Drosophila myogenesis. Dev. Biol. 203, 106–115 (1998).

Currie, D. A. & Bate, M. The development of adult abdominal muscles in Drosophila: myoblasts express twist and are associated with nerves. Development 113, 91–102 (1991).

Danoviz ME, Yablonka-Reuveni Z. 2012. Skeletal muscle satellite cells: background and methods for isolation and analysis in a primary culture system. *Methods Mol Biol* **798**: 21–52.

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

Dellavalle A, Sampaolesi M, Tonlorenzi R, Tagliafico E, Sacchetti B, Perani L, Innocenzi A, Galvez BG, Messina G, Morosetti R, et al. 2007. Pericytes of human skeletal muscle are myogenic precursors distinct from satellite cells. *Nat Cell Biol* **9**: 255–267.

Doyle, M. J. et al. Abcg2 labels multiple cell types in skeletal muscle and participates in muscle regeneration. J. Cell Biol. 195, 147–163 (2011).

Duckmanton A, Kumar A, Chang Y-T, Brockes JP. 2005. A single-cell analysis of myogenic dedifferentiation induced by small molecules. *Chem Biol* **12**: 1117–1126.

Dunant P, Larochelle N, Thirion C, Stucka R, Ursu D, Petrof BJ, Wolf E, Lochmüller H. 2003. Expression of dystrophin driven by the 1.35-kb MCK promoter ameliorates muscular dystrophy in fast, but not in slow muscles of transgenic mdx mice. *Mol Ther* **8**: 80–89.

Eckert MA, Lwin TM, Chang AT, Kim J, Danis E, Ohno-Machado L, Yang J. 2011. Twist1-induced invadopodia formation promotes tumor metastasis. *Cancer Cell* **19**: 372–386.

Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, Jatoi A, Loprinzi C, MacDonald N, Mantovani G, et al. 2011. Definition and classification of cancer cachexia: an international consensus. *Lancet Oncol* **12**: 489–495.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**: 1728–1740.

Fong AP, Yao Z, Zhong JW, Cao Y, Ruzzo WL, Gentleman RC, Tapscott SJ. 2012. Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev Cell* **22**: 721–735.

Fong AP, Yao Z, Zhong JW, Johnson NM, Farr GH, Maves L, Tapscott SJ. 2015. Conversion of MyoD to a neurogenic factor: binding site specificity determines lineage. *Cell Rep* **10**: 1937–1946.

Franco HL, Casasnovas J, Rodríguez-Medina JR, Cadilla CL. 2011. Redundant or separate entities?--roles of Twist1 and Twist2 as molecular switches during gene transcription. *Nucleic Acids Res* **39**: 1177–1186.

Fry AC. 2004. The role of resistance exercise intensity on muscle fibre adaptations. *Sports Med* **34**: 663–679.

Fry CS, Lee JD, Mula J, Kirby TJ, Jackson JR, Liu F, Yang L, Mendias CL, Dupont-Versteegden EE, McCarthy JJ, et al. 2015. Inducible depletion of satellite cells in adult, sedentary mice impairs muscle regenerative capacity without affecting sarcopenia. *Nat Med* **21**: 76–80.

Füchtbauer EM. 1995. Expression of M-twist during postimplantation development of the mouse. *Dev Dyn* **204**: 316–322.

Ghouzzi el V, Le Merrer M, Perrin-Schmitt F, Lajeunie E, Benit P, Renier D, Bourgeois P, Bolcato-Bellemin AL, Munnich A, Bonaventure J. 1997. Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat Genet* **15**: 42–46.

Giordani L, He GJ, Negroni E, Sakai H, Law JYC, Siu MM, Wan R, Corneau A, Tajbakhsh S, Cheung TH, et al. 2019. High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations. *Mol Cell*.

Goebel, H. H. Desmin-related neuromuscular disorders. Muscle Nerve 18, 1306–1320 (1995).

Gong W, Koyano-Nakagawa N, Li T, Garry DJ. 2015. Inferring dynamic gene regulatory networks in cardiac differentiation through the integration of multi-dimensional data. *BMC Bioinformatics* **16**: 74.

Gong XQ, Li L. 2002. Dermo-1, a multifunctional basic helix-loop-helix protein, represses MyoD transactivation via the HLH domain, MEF2 interaction, and chromatin deacetylation. *J Biol Chem* **277**: 12310–12317.

Gryder BE, Yohe ME, Chou H-C, Zhang X, Marques J, Wachtel M, Schaefer B, Sen N, Song Y, Gualtieri A, et al. 2017. PAX3–FOXO1 Establishes Myogenic Super Enhancers and Confers BET Bromodomain Vulnerability. *Cancer Discov* **7**: 884–899.

Guttridge DC, Mayo MW, Madrid LV, Wang CY, Baldwin AS. 2000. NF-kappaB-induced loss of MyoD messenger RNA: possible role in muscle decay and cachexia. *Science* **289**: 2363–2366.

Hebrok M, Wertz K, Füchtbauer EM. 1994. M-twist is an inhibitor of muscle differentiation. *Dev Biol* **165**: 537–544.

Helliwell, T. R. Lectin binding and desmin staining during bupivicaine-induced necrosis and regeneration in rat skeletal muscle. J. Pathol. 155, 317–326 (1988).

Hjiantoniou, E. et al. Twist induces reversal of myotube formation. Differentiation 76, 182–192 (2008).

Hutcheson DA, Zhao J, Merrell A, Haldar M, Kardon G. 2009. Embryonic and fetal limb myogenic cells are derived from developmentally distinct progenitors and have different requirements for beta-catenin. *Genes Dev* **23**: 997–1013.

Isenmann, S. et al. TWIST family of basic helix-loop-helix transcription factors mediate human mesenchymal stem cell growth and commitment. Stem Cells 27, 2457–2468 (2009).

Joe AWB, Yi L, Natarajan A, Le Grand F, So L, Wang J, Rudnicki MA, Rossi FMV. 2010. Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. *Nat Cell Biol* **12**: 153–163.

Jones S. 2004. An overview of the basic helix-loop-helix proteins. *Genome Biol* **5**: 226.

Jones TE, Stephenson KW, King JG, Knight KR, Marshall TL, Scott WB. 2009. Sarcopenia--mechanisms and treatments. *J Geriatr Phys Ther* **32**: 83–89.

Jung H-Y, Yang J. 2015. Unraveling the TWIST between EMT and cancer stemness. *Cell Stem Cell* **16**: 1–2.

Keefe, A. C. et al. Muscle stem cells contribute to myofibres in sedentary adult mice. Nat. Commun. 6, 7087 (2015).

Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351–1359.

Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).

Koncina E, Roth L, Gonthier B, Bagnard D. 2007. Role of Semaphorins during Axon Growth and Guidance. In *Axon Growth and Guidance*, Vol. 621 of *Advances in Experimental Medicine and Biology*, pp. 50–64, Springer New York, New York, NY.

Kostallari E, Baba-Amer Y, Alonso-Martin S, Ngoh P, Relaix F, Lafuste P, Gherardi RK. 2015. Pericytes in the myovascular niche promote post-natal myofiber growth and satellite cell quiescence. *Development* **142**: 1242–1253.

Kuang S, Chargé SB, Seale P, Huh M, Rudnicki MA. 2006. Distinct roles for Pax7 and Pax3 in adult regenerative myogenesis. *J Cell Biol* **172**: 103–113.

Kuang S, Kuroda K, Le Grand F, Rudnicki MA. 2007. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**: 999–1010.

Kumanogoh A, Kikutani H. 2013. Immunological functions of the neuropilins and plexins as receptors for semaphorins. *Nat Rev Immunol* **13**: 802–814.

la Serna de IL, Ohkawa Y, Berkes CA, Bergstrom DA, Dacwag CS, Tapscott SJ, Imbalzano AN. 2005. MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol Cell Biol* **25**: 3997–4009.

LaBarge, M. A. & Blau, H. M. Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. Cell 111, 589–601 (2002).

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**: 1813–1831.

Lee JD, Fry CS, Mula J, Kirby TJ, Jackson JR, Liu F, Yang L, Dupont-Versteegden EE, McCarthy JJ, Peterson CA. 2016. Aged Muscle Demonstrates Fiber-Type Adaptations in Response to Mechanical Overload, in the Absence of Myofiber Hypertrophy, Independent of Satellite Cell Abundance. *J Gerontol A Biol Sci Med Sci* **71**: 461–467.

Lee MS, Lowe G, Flanagan S, Kuchler K, Glackin CA. 2000. Human Dermo-1 has attributes similar to twist in early bone development. *Bone* **27**: 591–602.

Lepper C, Conway SJ, Fan C-M. 2009. Adult satellite cells and embryonic muscle progenitors have distinct genetic requirements. *Nature* **460**: 627–631.

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

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li L, Cserjesi P, Olson EN. 1995. Dermo-1: a novel twist-related bHLH protein expressed in the developing dermis. *Dev Biol* **172**: 280–292.

Li S, Zhang Y, Chen K, Barnes SD, Jaichander P, Zheng Y, Hassan M, Malladi VS, Skapek SX, Xu L, et al. 2019. (in press). Amplification of Twist2 in Rhabdomyosarcoma represses myogenesis and promote oncogenesis through redirection of Myod DNA binding. *Genes Dev*.

Li Y, Wang W, Wang W, Yang R, Wang T, Su T, Weng D, Tao T, Li W, Ma D, et al. 2012. Correlation of TWIST2 up-regulation and epithelial–mesenchymal transition during tumorigenesis and progression of cervical carcinoma. *Gynecologic Oncology* **124**: 112–118.

Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).

Liu N, Garry GA, Li S, Bezprozvannaya S, Sanchez-Ortiz E, Chen B, Shelton JM, Jaichander P, Bassel-Duby R, Olson EN. 2017. A Twist2-dependent progenitor cell contributes to adult skeletal muscle. *Nat Cell Biol* **19**: 202–213.

Liu N, Williams AH, Maxeiner JM, Bezprozvannaya S, Shelton JM, Richardson JA, Bassel-Duby R, Olson EN. 2012. microRNA-206 promotes skeletal muscle regeneration and delays progression of Duchenne muscular dystrophy in mice. *J Clin Invest* **122**: 2054–2065.

MacQuarrie KL, Yao Z, Fong AP, Diede SJ, Rudzinski ER, Hawkins DS, Tapscott SJ. 2013a. Comparison of genome-wide binding of MyoD in normal human myogenic

cells and rhabdomyosarcomas identifies regional and local suppression of promyogenic transcription factors. *Mol Cell Biol* **33**: 773–784.

MacQuarrie KL, Yao Z, Fong AP, Tapscott SJ. 2013b. Genome-wide binding of the basic helix-loop-helix myogenic inhibitor musculin has substantial overlap with MyoD: implications for buffering activity. *Skelet Muscle* **3**: 26.

Maesner CC, Almada AE, Wagers AJ. 2016. Established cell surface markers efficiently isolate highly overlapping populations of skeletal muscle satellite cells by fluorescence-activated cell sorting. *Skelet Muscle* **6**: 35.

Maestro R, Dei Tos AP, Hamamori Y, Krasnokutsky S, Sartorelli V, Kedes L, Doglioni C, Beach DH, Hannon GJ. 1999. Twist is a potential oncogene that inhibits apoptosis. *Genes Dev* **13**: 2207–2217.

Majka SM, Jackson KA, Kienstra KA, Majesky MW, Goodell MA, Hirschi KK. 2003. Distinct progenitor populations in skeletal muscle are bone marrow derived and exhibit different cell fates during vascular regeneration. *J Clin Invest* **111**: 71–79.

Makarewich CA, Baskin KK, Munir AZ, Bezprozvannaya S, Sharma G, Khemtong C, Shah AM, McAnally JR, Malloy CR, Szweda LI, et al. 2018. MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. *Cell Rep* **23**: 3701–3709.

Mastroyiannopoulos NP, Antoniou AA, Koutsoulidou A, Uney JB, Phylactou LA. 2013. Twist reverses muscle cell differentiation through transcriptional down-regulation of myogenin. *Biosci Rep* **33**: 903–911.

MAURO A. 1961. Satellite cell of skeletal muscle fibers. *J Biophys Biochem Cytol* **9**: 493–495.

McCarthy JJ, Mula J, Miyazaki M, Erfani R, Garrison K, Farooqui AB, Srikuea R, Lawson BA, Grimes B, Keller C, et al. 2011. Effective fiber hypertrophy in satellite cell-depleted skeletal muscle. *Development* **138**: 3657–3666.

Merindol N, Riquet A, Szablewski V, Eliaou J-F, Puisieux A, Bonnefoy N. 2014. The emerging role of Twist proteins in hematopoietic cells and hematological malignancies. *Blood Cancer J* **4**: e206–e206.

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**: R41.

Michaelis KA, Zhu X, Burfeind KG, Krasnow SM, Levasseur PR, Morgan TK, Marks DL. 2017. Establishment and characterization of a novel murine model of pancreatic cancer cachexia. *J Cachexia Sarcopenia Muscle* **8**: 824–838.

Michailovici, I., Eigler, T. & Tzahor, E. Craniofacial muscle development. Curr. Top. Dev. Biol. 115, 3–30 (2015).

Millay DP, O'Rourke JR, Sutherland LB, Bezprozvannaya S, Shelton JM, Bassel-Duby R, Olson EN. 2013. Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* **499**: 301–305.

Minasi, M. G. et al. The meso-angioblast: a multipotent, self-renewing cell that originates from the dorsal aorta and differentiates into most mesodermal tissues. Development 129, 2773–2783 (2002).

Mitchell KJ, Pannérec A, Cadot B, Parlakian A, Besson V, Gomes ER, Marazzi G, Sassoon DA. 2010. Identification and characterization of a non-satellite cell muscle resident progenitor during postnatal development. *Nat Cell Biol* **12**: 257–266.

Morley JE, Baumgartner RN, Roubenoff R, Mayer J, Nair KS. 2001. Sarcopenia. *Journal of Laboratory and Clinical Medicine* **137**: 231–243.

Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen C-C, Wongvipat J, Ku S-Y, Gao D, Cao Z, et al. 2017. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**: 84–88.

Murphy M, Kardon G. 2011. Origin of vertebrate limb muscle: the role of progenitor and myoblast populations. *Curr Top Dev Biol* **96**: 1–32.

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

Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. Genesis 45, 593–605 (2007).

Nasarre P, Gemmill RM, Drabkin HA. 2014. The emerging role of class-3 semaphorins and their neuropilin receptors in oncology. *Onco Targets Ther* **7**: 1663–1687.

Naya FJ, Mercer B, Shelton J, Richardson JA, Williams RS, Olson EN. 2000. Stimulation of slow skeletal muscle fiber gene expression by calcineurin in vivo. *J Biol Chem* **275**: 4545–4548.

Odelberg SJ, Kollhoff A, Keating MT. 2000. Dedifferentiation of mammalian myotubes induced by msx1. *Cell* **103**: 1099–1109.

Paliwal P, Conboy IM. 2011. Inhibitors of tyrosine phosphatases and apoptosis reprogram lineage-marked differentiated muscle to myogenic progenitor cells. *Chem Biol* **18**: 1153–1166.

Papizan JB, Garry GA, Brezprozvannaya S, McAnally JR, Bassel-Duby R, Liu N, Olson EN. 2017. Deficiency in Kelch protein Klhl31 causes congenital myopathy in mice. *J Clin Invest* **127**: 3730–3740.

Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7, e30619 (2012).

Pawlikowski B, Pulliam C, Betta ND, Kardon G, Olwin BB. 2015. Pervasive satellite cell contribution to uninjured adult muscle fibers. *Skelet Muscle* **5**: 42.

Preussner J, Zhong J, Sreenivasan K, Günther S, Engleitner T, Künne C, Glatzel M, Rad R, Looso M, Braun T, et al. 2018. Oncogenic Amplification of Zygotic Dux Factors in Regenerating p53-Deficient Muscle Stem Cells Defines a Molecular Cancer Subtype. *Cell Stem Cell*.

Qin Q, Xu Y, He T, Qin C, Xu J. 2012. Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. *Cell Res* **22**: 90–106.

Qu-Petersen Z, Deasy B, Jankowski R, Ikezawa M, Cummins J, Pruchnic R, Mytinger J, Cao B, Gates C, Wernig A, et al. 2002. Identification of a novel population of muscle stem cells in mice: potential for muscle regeneration. *J Cell Biol* **157**: 851–864.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Raimondi L, Ciarapica R, De Salvo M, Verginelli F, Gueguen M, Martini C, De Sio L, Cortese G, Locatelli M, Dang TP, et al. 2012. Inhibition of Notch3 signalling induces rhabdomyosarcoma cell differentiation promoting p38 phosphorylation and p21(Cip1) expression and hampers tumour cell growth in vitro and in vivo. *Cell Death Differ* **19**: 871–881.

Raper JA. 2000. Semaphorins and their receptors in vertebrates and invertebrates. *Curr Opin Neurobiol* **10**: 88–94.

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140 (2010).

Rohwedel, J., Horak, V., Hebrok, M., Fuchtbauer, E. M. & Wobus, A. M. M-twist expression inhibits mouse embryonic stem cell-derived myogenic differentiation in vitro. Exp. Cell Res. 220, 92–100 (1995).

Saab R, Spunt SL, Skapek SX. 2011. Myogenesis and rhabdomyosarcoma the Jekyll and Hyde of skeletal muscle. *Curr Top Dev Biol* **94**: 197–234.

Sacco, A., Doyonnas, R., Kraft, P., Vitorovic, S. & Blau, H. M. Self-renewal and expansion of single transplanted muscle stem cells. Nature 456, 502–506 (2008).

Saller MM, Huettl R-E, Hanuschick P, Amend A-L, Alberton P, Aszodi A, Huber AB. 2016. The role of Sema3-Npn-1 signaling during diaphragm innervation and muscle development. *J Cell Sci* **129**: 3295–3308.

Sambasivan R, Yao R, Kissenpfennig A, Van Wittenberghe L, Paldi A, Gayraud-Morel B, Guenou H, Malissen B, Tajbakhsh S, Galy A. 2011. Pax7-expressing satellite cells are indispensable for adult skeletal muscle regeneration. *Development* **138**: 3647–3656.

Schiaffino S, Reggiani C. 2011. Fiber types in mammalian skeletal muscles. *Physiol Rev* **91**: 1447–1531.

Schmidt, J. M. et al. Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation. Cell Rep. 10, 131–139 (2015).

Scholz D, Thomas S, Sass S, Podzuweit T. 2003. Angiogenesis and myogenesis as two facets of inflammatory post-ischemic tissue regeneration. *Mol Cell Biochem* **246**: 57–67.

Scott, W., Stevens, J. & Binder-Macleod, S. A. Human skeletal muscle fiber type classifications. Phys. Ther. 81, 1810–1816 (2001).

Sharma A, Verhaagen J, Harvey AR. 2012. Receptor complexes for each of the Class 3 Semaphorins. *Front Cell Neurosci* **6**: 28.

Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, Ambrogio L, Auclair D, Wang J, Song YK, et al. 2014. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. *Cancer Discov* **4**: 216–231.

Shi J, Wang Y, Zeng L, Wu Y, Deng J, Zhang Q, Lin Y, Li J, Kang T, Tao M, et al. 2014. Disrupting the Interaction of BRD4 with Diacetylated Twist Suppresses Tumorigenesis in Basal-like Breast Cancer. *Cancer Cell* **25**: 210–225.

Shi X, Garry DJ. 2006. Muscle stem cells in development, regeneration, and disease. *Genes Dev* **20**: 1692–1708.

Shield MA, Haugen HS, Clegg CH, Hauschka SD. 1996. E-box sites and a proximal regulatory region of the muscle creatine kinase gene differentially regulate expression in diverse skeletal muscles and cardiac muscle of transgenic mice. *Mol Cell Biol* **16**: 5058–5068.

Simpson P. 1983. Maternal-Zygotic Gene Interactions during Formation of the Dorsoventral Pattern in Drosophila Embryos. *Genetics* **105**: 615–632.

Skapek SX, Ferrari A, Gupta AA, Lupo PJ, Butler E, Shipley J, Barr FG, Hawkins DS. 2019. Rhabdomyosarcoma. *Nat Rev Dis Primers* **5**: 1.

Soleimani VD, Yin H, Jahani-Asl A, Ming H, Kockx CEM, van Ijcken WFJ, Grosveld F, Rudnicki MA. 2012. Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. *Mol Cell* **47**: 457–468.

Spicer DB, Rhee J, Cheung WL, Lassar AB. 1996. Inhibition of myogenic bHLH and MEF2 transcription factors by the bHLH protein Twist. *Science* **272**: 1476–1480.

Stark DA, Coffey NJ, Pancoast HR, Arnold LL, Walker JPD, Vallée J, Robitaille R, Garcia ML, Cornelison DDW. 2015. Ephrin-A3 promotes and maintains slow muscle fiber identity during postnatal development and reinnervation. *J Cell Biol* **211**: 1077–1091.

Stark DA, Karvas RM, Siegel AL, Cornelison DDW. 2011. Eph/ephrin interactions modulate muscle satellite cell motility and patterning. *Development* **138**: 5279–5289.

Šošić D, Richardson JA, Yu K, Ornitz DM, Olson EN. 2003. Twist regulates cytokine gene expression through a negative feedback loop that represses NF-kappaB activity. *Cell* **112**: 169–180.

Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**: 367–372.

Tai PW, Fisher-Aylor KI, Himeda CL, Smith CL, Mackenzie AP, Helterline DL, Angello JC, Welikson RE, Wold BJ, Hauschka SD. 2011. Differentiation and fiber type-specific activity of a muscle creatine kinase intronic enhancer. *Skelet Muscle* **1**: 25.

Tamagnone L, Comoglio PM. 2000. Signalling by semaphorin receptors: cell guidance and beyond. *Trends Cell Biol* **10**: 377–383.

Tapanes-Castillo A. 2004. Notch signaling patterns Drosophila mesodermal segments by regulating the bHLH transcription factor twist. *Development* **131**: 2359–2372.

Tapscott SJ, Thayer MJ, Weintraub H. 1993. Deficiency in rhabdomyosarcomas of a factor required for MyoD activity and myogenesis. *Science* **259**: 1450–1453.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**: 2032–2034.

Tatsumi R, Suzuki T, Do M-KQ, Ohya Y, Anderson JE, Shibata A, Kawaguchi M, Ohya S, Ohtsubo H, Mizunoya W, et al. 2017. Slow-Myofiber Commitment by Semaphorin 3A Secreted from Myogenic Stem Cells. *Stem Cells* **35**: 1815–1834.

Tenente IM, Hayes MN, Ignatius MS, McCarthy K, Yohe M, Sindiri S, Gryder B, Oliveira ML, Ramakrishnan A, Tang Q, et al. 2017. Myogenic regulatory transcription factors regulate growth in rhabdomyosarcoma. *Elife* **6**: 35.

Tisdale MJ. 2010. Reversing cachexia. *Cell* **142**: 511–512.

Tonkin, J., Villarroya, F., Puri, P. L. & Vinciguerra, M. SIRT1 signaling as potential modulator of skeletal muscle diseases. Curr. Opin. Pharmacol. 12, 372–376 (2012).

Torrente Y, Belicchi M, Sampaolesi M, Pisati F, Meregalli M, D'Antona G, Tonlorenzi R, Porretti L, Gavina M, Mamchaoui K, et al. 2004. Human circulating AC133(+) stem cells restore dystrophin expression and ameliorate function in dystrophic skeletal muscle. *J Clin Invest* **114**: 182–195.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.

Uezumi A, Fukada S-I, Yamamoto N, Takeda S, Tsuchida K. 2010. Mesenchymal progenitors distinct from satellite cells contribute to ectopic fat cell formation in skeletal muscle. *Nat Cell Biol* **12**: 143–152.

Verma M, Asakura Y, Murakonda BSR, Pengo T, Latroche C, Chazaud B, McLoon LK, Asakura A. 2018. Muscle Satellite Cell Cross-Talk with a Vascular Niche Maintains Quiescence via VEGF and Notch Signaling. *Cell Stem Cell* **23**: 530–543.e9.

Voehringer, D., Liang, H. E. & Locksley, R. M. Homeostasis and effector function of lymphopenia-induced 'memory-like' T cells in constitutively T cell-depleted mice. J. Immunol. 180, 4742–4753 (2008).

Wang G, Biswas AK, Ma W, Kandpal M, Coker C, Grandgenett PM, Hollingsworth MA, Jain R, Tanji K, López-Pintado S, et al. 2018. Metastatic cancers promote cachexia through ZIP14 upregulation in skeletal muscle. *Nat Med* **24**: 770–781.

Wang H, Lööf S, Borg P, Nader GA, Blau HM, Simon A. 2015. Turning terminally differentiated skeletal muscle cells into regenerative progenitors. *Nat Commun* **6**: 858.

Wang Y, Pessin JE. 2013. Mechanisms for fiber-type specificity of skeletal muscle atrophy. *Current Opinion in Clinical Nutrition and Metabolic Care* **16**: 243–250.

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

Wolf C, Thisse C, Stoetzel C, Thisse B, Gerlinger P, Perrin-Schmitt F. 1991. The M-twist gene of Mus is expressed in subsets of mesodermal cells and is closely related to the Xenopus X-twi and the Drosophila twist genes. *Dev Biol* **143**: 363–373.

Wosczyna MN, Rando TA. 2018. A Muscle Stem Cell Support Group: Coordinated Cellular Responses in Muscle Regeneration. *Dev Cell* **46**: 135–143.

Xie L, Yin A, Nichenko AS, Beedle AM, Call JA, Yin H. 2018. Transient HIF2A inhibition promotes satellite cell proliferation and muscle regeneration. *J Clin Invest* **128**: 2339–2355.

Xu L, Zheng Y, Liu J, Rakheja D, Singleterry S, Laetsch TW, Shern JF, Khan J, Triche TJ, Hawkins DS, et al. 2018. Integrative Bayesian Analysis Identifies Rhabdomyosarcoma Disease Genes. *Cell Rep* **24**: 238–251.

Yamagishi S, Kesavamoorthy G, Bastmeyer M, Sato K. 2016. Stripe Assay to Study the Attractive or Repulsive Activity of a Protein Substrate Using Dissociated Hippocampal Neurons. *J Vis Exp*.

Yamashita K, Yoshioka T. 1991. Profiles of creatine kinase isoenzyme compositions in single muscle fibres of different types. *J Muscle Res Cell Motil* **12**: 37–44.

Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA. 2004. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**: 927–939.

Yang J, Mani SA, Weinberg RA. 2006. Exploring a new twist on tumor metastasis. *Cancer Res* **66**: 4549–4552.

Yang Z, Liu Q, Mannix RJ, Xu X, Li H, Ma Z, Ingber DE, Allen PD, Wang Y. 2014. Mononuclear cells from dedifferentiation of mouse myotubes display remarkable regenerative capability. *Stem Cells* **32**: 2492–2501.

Yao Z, Fong AP, Cao Y, Ruzzo WL, Gentleman RC, Tapscott SJ. 2013. Comparison of endogenous and overexpressed MyoD shows enhanced binding of physiologically bound sites. *Skelet Muscle* **3**: 8.

Yin H, Price F, Rudnicki MA. 2013. Satellite cells and the muscle stem cell niche. *Physiol Rev* **93**: 23–67.

Yousfi M, Lasmoles F, Marie PJ. 2002. TWIST inactivation reduces CBFA1/RUNX2 expression and DNA binding to the osteocalcin promoter in osteoblasts. *Biochem Biophys Res Commun* **297**: 641–644.

Yu, K. et al. Conditional inactivation of FGF receptor 2 reveals an essential role for FGF signaling in the regulation of osteoblast function and bone growth. Development 130, 3063–3074 (2003).

Zeid R, Lawlor MA, Poon E, Reyes JM, Fulciniti M, Lopez MA, Scott TG, Nabet B, Erb MA, Winter GE, et al. 2018. Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat Genet* **50**: 515–523.

Zheng B, Cao B, Crisan M, Sun B, Li G, Logar A, Yap S, Pollett JB, Drowley L, Cassino T, et al. 2007. Prospective identification of myogenic endothelial cells in human skeletal muscle. *Nat Biotechnol* **25**: 1025–1034.