12-1-2015

Transcriptional regulation of muscle development in Drosophila melanogaster

Tonya Brunetti

Tonya M. Brunetti

Candidate

Biology

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Richard M. Cripps, Chairperson

Stephen A. Stricker

Christopher A. Johnston

Melanie E. Moses

TRANSCRIPTIONAL REGULATION OF MUSCLE DEVELOPMENT IN DROSOPHILA MELANOGASTER

by

TONYA M. BRUNETTI

B.A., Molecular and Cell Biology, University of Colorado, 2009
M.S., Computer Science, University of New Mexico, 2015

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Biology

The University of New Mexico
Albuquerque, New Mexico

December, 2015

DEDICATION

To my loving family, including our doggy Skyler, I could not have done any of this without your support and guidance. I am so blessed to have you all in my life and I cannot begin to express how grateful I am for all the unconditional support.

-Tonya

ACKNOWLEDGEMENTS

I would like to thank Dr. Richard Cripps, my advisor and mentor, for believing that I could do this PhD. I appreciate all the advice, guidance, and support you have given me through the past five years. Additionally, I cannot begin to thank you for your support and understanding with the additional time constraint with the added degree. Thank you so much, I certainly could not have done this without you and I could not have asked for a better mentor.

Thank you to my committee members, Dr. Christopher Johnston, Dr. Stephen Sticker, and Dr. Melanie Moses for directing me through this process. I know it has not always been easy, but I am thankful to have had such a knowledgeable and talented group of professors guide me through this degree and encourage me to dig deeper in intellectual knowledge.

I most certainly cannot even begin to thank my lab family, including Evan Dewey! You guys have been great and I consider you my family away from home. A set of special thank yous to Sandy Oas, Antonio Banuelos, Ginny Morriss, TyAnna Lovato, Lizzie Clarke, and Marilyn Cisneros. Sandy Oas, oh what to say, too many things. For keeping me sane, letting me vent my frustrations, and brightening my tough days with chocolate, you know me too well! Thank you to Antonio…for everything, I cannot even begin to list the things you have done for me. Ginny Morriss, for welcoming me into this lab and answering all my annoying lab questions.TyAnna, for pretty much being the lab superhero, need I say more. Lizzie and Evanare all the encouragement and support, and many after-work stress beers. Last but not least, MARY-LYN! All the support, the coffee runs, and the silly remarks helped me make it through graduate school. You all are truly remarkable people, you will always be in my heart, and I know I will always have a family here with Cripps lab.

Finally, I would like to thank my family, this means you too Miss Kimmel! To Mom and Dad: the pouring encouragement and support you guys have given me is priceless. I am so incredibly lucky to have you as my parents, you have supported me through the hardest of times and through the happiest of times. Thank you for calling me every week to make sure I was doing well and for showering me with love and support. There is nothing I would trade in the entire world for that, and I love you guys with all my heart and I WILL come back to Colorado! Thank you to my brothers Michael and Ricky for visiting me and providing moral support through the years, I am lucky to have such caring siblings. Also, thank you to Rachel Kimmel, my sister and best friend. You have really inspired me to achieve my goals and have provided me so much valuable advice since we were 12 years old. I can't imagine life without you, thank you for everything!

TRANSCRIPTIONAL REGULATION OF MUSCLE DEVELOPMENT IN DROSOPHILA MELANOGASTER

by

TONYA M. BRUNETTI

B.A., Molecular and Cell Biology, University of Colorado, 2009
M.S., Computer Science, University of New Mexico, 2015
PhD., Biology, University of New Mexico, 2015

ABSTRACT

The transcriptional regulation of muscle development involves several complex processes that must work together in order to form functional, syncytial muscle cells. However, when transcription is mis-regulated, muscle development is often times negatively affected and can lead to muscle diseases such as muscular dystrophy and cardiac myopathies. In order to gain more insight into how transcription is regulated, I use *Drosophila melanogaster* as a model for understanding muscle development. In chapter one, I use a traditional genetic screen to phenotypically and molecularly identify two Hox co-factors, *extradenticle* and *homothorax* that have the ability to change muscle identity. Additionally, in chapter two, through the identification of a mechanism, I identify a gene critical in adult myoblast fusion and is directly regulated by the transcription factor, Myocyte Enhancer Factor-2 (MEF2). Lastly, in chapter three a computation approach is used to discover new potential co-factor binding sites that may work in conjunction with MEF2 in transcriptional muscle regulation. Together, these results provide new

information into how muscle is transcriptionally regulated during different stages of development.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Muscle development is the process from which single, undifferentiated cells,take on a muscle precursor fate to become myoblasts that fuse together to ultimately become fully, functional, syncytial muscles. However, in order for this process to ensue, a diverse population of genes must be expressed at the proper times in development. The developmental stages in which specific genes are either turned on or off are strictly and carefully regulated by a group of proteins called transcription factors. Therefore, it is critical to gain a comprehensive understanding of how the process of transcription is regulated in order to gain insight into the mechanisms of normal muscular gene expression. Understanding normal gene expression also provides insight into the problems that arise due to the misregulation of gene expression which can lead to congenital muscle defects and other myopathies.

Many muscle disorders that arise are genetic in nature and arisefrom mutations within structural genes and the loss of transcriptional regulation. These include muscular dystrophies, congenital myopathies, and metabolic muscle diseases (Kennedy Krieger Institute, 2012). For example, one of the most severe forms of muscular dystrophy isDuchenne's muscular dystrophy which is caused by a mutation in the gene *dystrophin* (Hoffman et al., 1987). Other congenital myopathies such as actin myopathy and nemaline myopathy are the result of mutations in the *actin* gene (Nowak et al., 1999). It is critical to study the genetics of these diseases so possible therapeutic remedies can be found to aid in the treatment and cure of these disorders. Current therapeutic research has investigated the use of stem cells derived from skeletal muscle in mice as a possible remedy for muscular dystrophy (Lee et al., 2000). Lee et al., 2000 showed that stem cells extracted

from muscle and injected into mice with induced muscular dystrophy had myofibers that were able to regenerate and showed they were positive for Dystrophin. Similarly, another study illustrated that intravenously injecting stems cells derived from haematopoietic stem cells or a population of muscle derived stem cells were able to partially rescue Dystrophin in mice induced with muscular dystrophy (Gussoni et al., 1999). Other therapeutic techniques involve the use of gene therapy using the adeno-associated virus (Fisher et al., 1997). Fisher at al., 1997 have shown that the recombination adeno-associated virus injected to the muscle of mice were able to successfully incorporate the viral genome into the nucleus, illustrating the potential for site-directed gene therapy. Nevertheless, to more effectively control the treatment of muscle disease, it is necessary to gain a greater understanding of the regulatory processes that orchestrate normal muscle development.

Although a seemingly unlikely candidate, the common fruit fly, known as *Drosophila melanogaster*, is a good model organism for studying the transcriptional regulation of muscle development. *Drosophila* have a relatively fast life cycle which make them an optimal organism for studying genetics. At room temperature, *Drosophila* goes from embryo to adulthood in a matter of 9 days and their development can be sped up to a week by increasing the temperature to 29ºC (Bloomington Drosophila Stock Center).

Another benefit is the lack of redundancy in the *Drosophila* genome. Unlike vertebrates, *Drosophila* typically have a single ortholog of many regulatory genes, which means when a gene is knocked down or knocked out, there does not exist another copy of the gene elsewhere that may rescue the defect. An example of this is the myogenic transcription factor, Myocyte Enhancer Factor-2 (MEF2). While vertebrates possess four copies of this gene, MEF2A-D, *Drosophila* only have a single gene (Lilly et al., 1994), and

whereas multiple mammalian Mef2 genes need to be knocked out in order to uncover a role in muscle development (Potthoff et al., 2007), knockouts of the single Drosophila *Mef2* show severe defects in muscle differentiation (Lilly et al 1995). Furthermore, there is a strong similarity between the alleles that cause human disease phenotypes with those found in *Drosophila*. Roughly, 77% of disease causing alleles in humans can be mapped to a highly similar allele within the *Drosophila* genome (Reiter et al., 2001). In particular, it is a well-documented organism in the study of muscle diseases. For example, the human gene, *MLP* has a comparable homolog in *Drosophila* known as *mlp84B*, which aids in proper cardiac function (Mery et al., 2007).*Drosophila* also possesses a similar gene to the vertebrate *dystrophin* also called *Dystrophin* or *DLP2* (Roberts and Bobrow, 1998). Similar to vertebrates *Drosophila* forms a Dystroglycan-Dystrophin complex which is often times the interaction that is affected in muscular dystrophies (Shcherbata et al., 2007). In addition to *Dystrophin*, *Drosophila* has genes that are similar to *Dystrophin* such as *MSP-300*, making it a good model for studying muscular dystrophy (Rosenberg-Hasson et al., 1996).

Since *Drosophila* has contributed to the study of genetics for the last 100 years as a model organism, researchers have been able to fully sequence and annotate the genome (Adams et al., 2000). This has allowed researchers to relatively easily find genes of interest and the possible sequence regions of transcriptional regulation (modMine, 2012). Additionally, the comprehensive information about gene annotation has provided researchers with a database of possible gene functions, expression levels during various developmental time points and expression levels in specific tissues. This has guided molecular research in the way of finding potential gene and protein interactions.

Computationally, this has provided computer scientists with a large data base of sequence information that has allowed researchers to find patterns of sequences similarity and functionality in both the genome and proteome, and make predictions about gene and protein interactions, conservation, and functionality (Xenarios et al., 2002; Letovsky and Kasif, 2003). The annotation information has allowed computational researchers to build clusters and networks of genes and proteins from the entire genome that share similar attributes. In combination, this has provided a symbiotic relationship with the field of molecular biology and computational biology (Hasty et al., 2001; Kitano, 2002). By computational approaches providing biologists with new information regarding possible genes, proteins, and genetic networks with which a particular gene or protein may interact gives biologists new insight into a gene or protein's molecular mechanism and functionality (Friedman et al., 2004; Eissing et al., 2011). In a similar manner, the research and information provided by biologists help computational biologists validate and modify their existing algorithms to make better predictive measures of gene and protein interactions (Datta, 2006; Martelotto et al., 2014). Therefore, taken in combination, this provides biologists and computer scientist a cyclical and mutualistic process to further both areas of research andinvestigators of*Drosophila*biology have been able to exploit this process due to its well-annotated and fully sequenced genome.

Muscle development in *Drosophila*shares many of the similar transcriptional regulators and pathways as in vertebrates and is amenable to both genetic and computational analysis. *Drosophila* muscle development begins with the specification of the mesoderm during the embryonic stage of development. The mesoderm is specified by the expression and inhibition of specific genes. In particular, the expression of the gene

decapentaplegic (*dpp*) aids in the regulation and expression of other mesodermal specification genes (Staehling-Hamptonet al., 1994). The expression of *dpp* activates transcription factors such as *tinman* (*tin*) and *bagpipe* (*bap*) which specifies the dorsal mesoderm to eventually become the *Drosophila* cardiac muscle (Azpiazu and Frasch, 1993). At this point, the mesoderm also expresses transcription factors such as*twist* (*twi*) and *Myocyte enhancer factor-2* (*Mef2*) (Thisse et al., 1988; Lilly et al., 1995). Competing pathways pattern the mesoderm through the expression of the gene *even-skipped* (*eve*) and *sloppy-paired* (*slp*) through the hedgehog pathway and the wingless pathway, respectively (Riechmann et al., 1997). The visceral mesodermal fate is determined by the presence of *eve*in conjunction with low levels of *twi* expression while*slp* serves as a cardiac specifier by *eve* repression of *slp* and the maintenance of high levels of *twi*expression (Baylies and Bate, 1996 ;Riechmann et al., 1997). This maintenance of high levels of *twi*expression is also critical in the development of what will ultimately differentiate into the skeletal muscle in concert with *Mef2* expression (Baylies and Bate, 1996).In vertebrates this myogenic fate requires additional transcription factors in the protein family known as myogenic regulatory factors (MRFs). Some examples include MyoD and Myf-5, whose roles are to aid in the specification of myoblasts (Rudnicki et al., 1993; Rawls and Olson, 1997). Loss of MyoD and Myf-5 result in the lack of specification for myoblast precursors. Additionally, the Pax family of transcription factors, Pax-3 and -7, in vertebrates also aid in specifying a precursor myoblast fate upstream of myogenic regulators (Grand and Rudnicki, 2007).

Once the mesoderm has been established and myogenic cell fates are specified, the process of myoblast fusion must occur. This process is conserved between *Drosophila* and

vertebrates (Srinivas et al., 2007).Myoblast fusion begins with the specification of two distinct populations of myoblasts: the founder cell (FC) versus those that are destined to become fusion competent myoblasts (FCM). Founder cells begin to express FC specific genes such as *dumbfounded* (*duf*) or Kirrel in vertebrates and *roughest* (*rst*) also a Kirrel family gene, which is translated into a critical protein required for attracting FCMs to fuse with FCs (Ruiz-Gómez et al., 2000; Strunkelnberg et al., 2001; Srinivas et al., 2007). Additionally, the gene *antisocial* (*ants)* is also expressed specifically in the FC population (Chen and Olson, 2001). *Ants* interacts with Duf to rearrange the cytoskeleton of FCs (Chen and Olson, 2001). Vertebrates possess a homolog of this gene known as *tanc1,* which similarly interacts with the Kirrel family protein in muscle formation (Rochlin et al., 2009).

Conversely, fusion competent cells tend to differ more genetically from each other unlike the founder cell population of myoblasts (Taylor, 2002). Although the genes of the FCMs are required for fusion, many of them are not expressed at the same dosage within FCMs (Taylor, 2002). Fusion competent myoblasts express the gene *lameduck* (*lmd*) and *sticks and stones* (*sns*) also referred to as Nephrin in vertebrates (Duan et al., 2001; Bour et al., 2000; Sohn et al., 2009).*lmd*is critical for FCMs since *Mef2* and *sns* are dependent upon *lmd* expression for differentiation in the FCMs (Duan et al., 2001). Additionally, the gene *hibris* (*hbs*), which is also Nephrin in vertrebrates, is expressed specifically in FCMs.Hbs co-localizes with Sns works in a manner that is dose-dependent depending on the FCM (Artero et al., 2001, Dworak et al., 2001).

Once the specification of the two distinct myoblast populations has occurred, FCs attract FCMs to come into close proximity with each other, so fusion can physically occur.

The membrane proteins of FCs, Duf and Rst, recognize and adhere to the FCM membrane proteins Sns and Hbs (Rochlin et al., 2009). Cytoskeletal changes begin to occur in each cell and at this point during myoblast fusion, the formation of the pre-fusion complex occurs. Genes such as *kette*,*blownfuse* (*blow*), and *singles bar* (*sing*) work together to make the pre-fusion complex and aid in membrane breakdown between fusing myoblasts (Richardson et al., 2008). *blow*and *kette*interact together by establishing electron dense plaques that are required for fusing cells to progress past the pre-fusion complex, while *sing* works in vesicle transport between cells (Schröter et al., 2004; Estrada et al., 2007).At the point of membrane breakdown, the FC becomes one with the FCM to form a multinucleated myotube and subsequent rounds of fusion are able to proceed. Clearly, myoblast fusion is a complex process requiring the actions of a number of structural genes, however there is relatively little informationavailable as to how the fusion genes are transcriptionally regulated.

Post myoblast fusion, naïve myotubes still need to be directed to specific muscle identities. Muscle identity is regulated by a process called differential gene expression through the actions of transcriptional regulators. This is a very important time in muscle development in that myotubes begin to take on different somatic muscle identities depending on how they need to function. In the *Drosophila*pupa, somatic muscle can become fibrillar muscle such as that of the indirect flight muscle (IFM), or tubular muscle such as the tergal depressor of trochanter (TDT or "jump") muscle, both of which express a different set of structural genes. Muscle fated to become fibrillar begins expressing structural genes such as *Actin 88F, TpnC4,* and *flightin* which help enable the muscle to function as highly oxidative muscle for high endurance and low force (Bryantsev et

al.,2012). By contrast, the *Drosophila* jump muscle is tubular and is used for high force and rapid contraction, therefore tires quickly. In order to accommodate for this functionality it expresses structural genes such *Actin79B* and *TpnC41C* (Bryantsev et al., 2012).The expression of structural genes is mediated by a large collection of transcription factors and co-factors. In *Drosophila*the genes *twi* and *Mef2* play a crucial role regulating the expression of structural genes through the regulation of the Notch pathway (Tapanes-Castillo and Baylies, 2004).

Since the onset of the genomic era, traditional genetic and regulatory analysis of muscle development has been complemented by high throughput "genomic" approaches. Most prominent amongst these are the chromatin immunoprecipitation- sequencing assays pioneered by the Furlong laboratory, RNA-seq data from the modENCODE project, and a new technology termed Enhancer-FACS-seq developed by theBulyk laboratory (Sandmann et al., 2007; modMine, 2012; Gisselbrecht, et al., 2013). In these studies, transcriptional targets of several key factors controlling mesoderm development have been identified at the genomic level. In particular, MEF2 was experimentally shown to bind 670 sites at the embryonic stage of development (Sandmann et al., 2007). Of these sites, roughly 32% had the ability to bind MEF2 only at later stages of embryonic development (Sandmann et al., 2007). Studies such as this identify a wealth of transcriptional targets to analyze for their roles in muscle formation; in addition, these studies also provide a great deal of genomic sequence information that can be mined to further understand the genomic context of sequences and genes that are regulated by MEF2 and related transcription factors.

For my thesis work, I have worked to combine both genetic/molecular approaches and computational studies to understand how transcription factors control muscle fate in the *Drosophila* system. The approaches I have used in the lab to address this research are: to characterize phenotypes of knock down candidate genes thought to be involved in muscle development, to identify mechanisms by which genes are regulated, and finally to use computational approaches as a segue to finding new candidate genes to genetically and molecularly test in the wet laboratory. Therefore, I present three chapters that illustrate each of these approaches in an attempt to further understand the transcriptional regulation behind muscle development.

In chapter one, I characterize two Hox co-factors, Extradenticle and Homothorax, as part of a large genetic screen for factors that phenotypically and molecularly switch muscle identity. I show that Exd and Hth co-localize to transcriptionally regulate structural gene expression. The loss of Exd/Hth illustrate the ability for the co-factors to change the expression of structural muscle genes in different muscle types.

In chapter two, I define a mechanism for the transcriptional regulation for the myoblast fusion gene, *singles bar* by showing that it is a MEF2 target and that *sing* functions both at the embryonic stage of myoblast fusion as well as the adult stage. This has defined a novel role for MEF2 in the fusion process of adults.

The focus of chapter three is to use computational biology to find other co-factors that may work with MEF2 in the transcriptional regulation of muscle development. From some preliminary cell culture work performed in chapter two, I knew that MEF2 was incapable of activating *sing* expression by itself. Through the use of a genetic algorithm I

was able to identify regions of sequence regularity in a MEF2 ChIP-chip dataset that may serve as candidate binding sites for MEF2 co-factors.

Overall these studies define new mechanisms for how muscle forms in the developing animal, that can be translated to the vertebrate system based upon the strongly conserved developmental regulatory mechanisms shared between these diverse groups of animals. In addition, my new computational approach can be used to identify new regulatory sequences controlling muscle development. This approach can be applied both to systems in higher animals and to other transcriptional regulatory processes for which genome-wide binding data have been generated.

Chapter 1: Extradenticle and Homothorax Control Adult Muscle Fiber Identity in *Drosophila*

ABSTRACT

Muscle diseases such as muscular dystrophy are often due to genetic mutations. It is therefore important to study how muscle develops and how its development is regulated in order to gain insight into mechanisms that govern myopathies. Critical to muscle development in vertebrates are two homeobox co-factors, Meis1 and Pbx1. Meis1 is responsible for translocating Pbx1 into the nucleus thereby initiating transcriptional regulation of particular genes in the nucleus. *Drosophila* have homologous genes, *homothorax*(*hth*)and *extradenticle*(*exd*), respectively, whose roles had previously been associated primarily with limb development and patterning. Through a genetic screen we identified that *hth* and *exd* may also have critical roles in muscle differentiation. Knockdowns of *extradenticle* and *homothorax* results in the indirect flight muscles to begin expressing genes that characterize the jump muscle. Additionally, overexpression of these co-factors cause the jump muscle to express genes that define the indirect flight muscles. We characterize a novel role for the homeodomain co-factors, *extradenticle* and *homothorax* in the ability to change muscle identity in *Drosophila*.

Keywords: *Drosophila*, homothorax, extradenticle, muscle identity, Hox

INTRODUCTION

Understanding muscle development and the differences between different muscle types is critical for understanding various problems that arise due to the improper development of muscles. These myopathies are often attributed to genetic disorders where the mis-expression of a muscle gene due to the lack of transcriptional regulation often times result in truncated proteins that are non-functional or lack the protein altogether. Myopathies such as muscular dystrophy, result in the degradation of skeletal muscle and the inability to restore the damaged muscles. Therefore, it is important to understand the transcriptional regulation of skeletal muscle gene expression in order to gain a broader insight into how these myopathies result and how mew muscle tissue might be generated

In mammals,skeletal muscle fibers are classified into four types based on the abundance of specific myosin heavy chain (MyHC) isoforms (Schiaffino and Reggiani, 2011). The four muscle types, type I, type IIa, type IIb, and type IIx, can be further classified by their rate of metabolic properties such as the oxidative and glycolytic capabilities (Greising et al., 2012).Type I muscle, known as the slow-twitch myofibers, are oxidative and therefore utilize aerobic respiration, unlike their counterpart myofibers of type II. Type II muscle fibers are categorized as fast-twitch myofiberswhich have a high glycolytic rate as they endure the process of anaerobic respiration (Smerdu et al., 1994; Schiaffino and Reggiani, 2011). Type IIa, type IIb, and type IIx are all subcategories of fast-twitch muscles,although less is known about type IIx(Smerdu et al., 1994). Within the type II muscle category, type IIa, IIb, and IIx increase in the rate at which they fatigue, type IIx experiencing fatigue the fastest.

The heterogeneity of mammalian skeletal muscle fibers as described above have led to the difference in muscle functionality between fiber types. Type I fibers are utilized in muscles when there is a demand for sustained activity, whereas, type II fibers are stimulated when there is a quick need for short periods of activity (Grifone et al., 2004). Therefore, it is not surprising that these fibers are composed of different isoforms of myosin. Muscle type has a profound effect on the aging population due to sarcopenia. Research has shown that type II muscle is substantially lost in elderly individuals with a significant decline in the size of type II muscle fibers as well (Verdijk et al., 2007; Nilwik et al., 2013). This is opposed to the type I fibers which tend to be unaffected with age (Larsson et al., 1978). Thus, it is important to determine the differences between type I and type II muscle fibers in order to understand the mechanisms that underlie sarcopenia.

In *Drosophila*, the adult skeletal muscles are comprised of two distinct skeletal muscle types: the indirect flight muscles (IFM) and the tergal depressor of the trochanter (TDT). Each of these muscle types are functionally and molecularly distinct, similar to their mammalian counterparts, despite their broad characterization as skeletal muscle. Therefore, understanding how different Drosophila fiber types are specified will provide insight into how muscle development is transcriptionally regulated at later stages in development.

The IFMs of *Drosophila* can be sub-categorized into two groups: the dorsal longitudinal muscle (DLM) and the dorsoventral muscle (DVM) (Dutta and VijayRaghavan, 2006).Structurally, the indirect flight muscles are fibrillar. These muscles function asynchronously by stretch activation for sustained flight(Josephson et al., 2000). The stretch activation of these muscles can be attributed to the functionality of the DLMs

and DVMs of the adult indirect flight muscles. The DLMs cause the thorax of the fly to contract and tighten, resulting in the lowering of the wings. On the contrary, the movement of the DVMs cause the thorax to relax and lead to the opposing function, the lifting of the wing(Dutta and VijayRaghavean, 2006). The requirement for sustained flight causes these muscles to generally be highly oxidative.

In contrast to the IFM, the tergal depressor of the trochanter (TDT) also known as the jump muscle, is synchronous and highly innervated (Josephson et al., 2000). Whereas the IFMS are made of fibrillarmyofibrils from the DLMs and DVMs, the TDT is composed of roughly 32 tubular myofibrils arranged in a rosette pattern (Peckham et al., 1990). As the name suggests, they aid in jumping and serve as the "fight or flight" response in *Drosophila*. The lack of sustained use in these muscles makes them metabolically glycolytic and more similar to fast-twitch muscles.

Much research has focused upon the regulatory and developmental events that control the formation of the adult muscles.It has been shown that the absence of the transcription factor *twist* (*twi*) with the presence of the transcription factor, Myocyte Enhancer Factor-2 (MEF2) during muscle development enables myoblasts to undergo differentiation (Hebrok et al., 1994). This process is regulated through the Notch signaling pathway that causes the down-regulation of Twi and the up-regulation of MEF2. Once differentiation is initiated, specification of muscle identity also takes place. The expression of different structural genes begins to give rise to different muscle types. In *Drosophila* the structural genes, *Actin88F*and *TpnC4*, are expressed specifically in the indirect flight muscles, whereas*Actin79B*and *TpnC41C* are genes that structurallydefine the TDT (Hiromi and Hotta, 1985; Courchesne-Smith and Tobin, 1989; Herranz et al., 2004).

In order to understand how these structural genes are specified for each muscle type, our laboratory performed an RNAi genetic screen of transcription factors. Using a flight muscle specific driverwe were able to determine if particular levels of a beta-galactosidase reporter were either significantly reduced or increased in the IFMs. Two genes, *homothorax*(*hth*) and *extradenticle*(*exd*) showed significant decreases in expression of *lacZ*in the IFMs of *fln-lacZ* adults when they were knocked down (Bryantsev et al., 2012).

Exd and Hthare homeotic co-factors that exhibit high conservation across different animals. They belong to a lineage of proteins called the three-amino-acid-loop-extension homeodomain proteins (TALE), which typically function as transcriptional regulators by dimerizing and trimerizing with other proteins from the TALE family (Burglin 1997; Liu et al., 2010). In *C. elegans,* CEH-20 is the Exd homolog and UNC-62 is the Hthhomolog (Jiang et al., 2009).Furthermore, both have homologs in mice. The vertebrate homolog of Exd being Pbx1 and Hth corresponding to Meis1 (Burglin, 1997). Functionally, the proteins are similar across animals as well.Hth/Meis and Exd/Pbx1 have roles pertaining to the proximal-distal patterning of the limbs (Mercader, et al., 1999). In all three homologs, it has been shown that Meis1/UNC-62/Hth and Pbx1/CEH-20/Exd must interact together in order to translocate Pbx1/CEH-20/Exd into the nucleus (Jaw et al., 1999, Jiang et al, 2009; Berthelsen et al., 1999).

In this manuscript, I identify novel roles for the two homeodomain genes, *extradenticle*(*exd*) and *homothorax* (*hth*), in promoting muscle fiber specification via their abilities to transform IFM into a TDT fate and vice versa. By cloning and generating overexpression constructs of *exd* and *hth* I show that ectopic expression of these

homeodomain genes results in a switch of the jump muscle to an IFM-like identity and that a loss of these proteins switches flight muscle to a TDT-like tubular muscle identity (Bryantsev et al., 2012).

MATERIALS AND METHODS

Drosophila Stocks and Crosses

Overexpression *Drosophila* stocks of *hth* and *exd* were generated using RT-PCR and Gateway cloning technology (see Transgenic Lines section). Knockdown stocks of *hth*and *exd*were obtained from the Bloomington *Drosophila* Stock Center and the Vienna *Drosophila* RNAi Center. Knockdown and overexpression lines were crossed at 29 degrees C with an *1151-Gal4; fln* driver.

Transgenic Lines

The following PCR primers were used to generate *homothorax* and *extradenticle* overexpression lines via RT-PCR:

hth_F1_attB1-

5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTAGTGGCACAAATCGGGTTAG-3’

hth_R1_attB2-

5’-GGGGACCATTTGTACAAGAAAGCTGGGTTCCAGTTTCATTTCCGGTTC-3’

exd-F1_attB1-

5’-GGGACAAGTTTGTACAAAAAAGCAGGCTTGGATTTGTAGCTTGCTTTGTG-3’

exd_R1_attB2-

5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCCGCAACTGTATGAGGGATT-3’

RNA for RT-PCR was extracted from adult pupal wildtype flies. Gateway technology was used to clone both the *hth* and *exd* constructs into a pUAST-attBvector.Constructs for generating transgenic lines were injected into *Drosophila* embryos according to the protocol published by Rubin and Spradling (1982).

Cell Culture

Gateway technology was used to clone *hth* and *exd* constructs into a pDONOR221 vector and then into a pAW vector for use in cell culture. *Drosophila* S2 cell lines were used for cell culture experiments, and experiments were repeated in duplicate. Samples were prepared according to the TransIT-2020 transfection reagent protocol (Mirus Bio) with sterile round 12mm coverslip placed on the bottom of a 24-well plate. Transfected cells were stained with a rabbit anti-*hth* antibody and guinea pig anti-*hth*antibody (Dr. Richard Mann) and a mouse anti-*exd* antibody at a concentrations of 1:20 (University of Iowa Developmental Studies Hybridoma Bank). Secondary Alexa-Fluor antibodies, goat anti-guinea pig 568 and goat anti-mouse 488 were used at 1:500 dilutions in PBTxN.

Histochemical Staining

Frozen horizontal sections of wildtype*hth/exd* knockdowns and *exd/hth* overexpression adults were stained to visualize the metabolic properties of the flight and jump muscles. Samples were collected at the pharate stage (96 hours after puparium formation (apf)and onward) and embedded inTissue-Tek OCT Compound (Sakura) before

being frozen in liquid nitrogen and stored at -80 degrees C. Tissue staining for succinate dehydrogenase and acetylcholinesterase were carried out according to the protocol from Deak, 1977.

RESULTS

Design and Construction of Clones

Full length protein sequences were made for *extradenticle* and *homothorax* to be used in cell culture and for use in the construction of transgenic flies. This was made by designing primers to target the transcriptional start site and stop of both *exd* and *hth*. The primers had an additional modification which was an added attB1 and attB2 site to the forward and reverse primers, respectively. The attB sites allow for the use of Gateway recombination technology into another Gateway plasmid.

The primers were used on cDNA generated from adult pupal RNA using an Invitrogen First Strand Synthesis kit. To verify that the primers targeted the correct sequence without mutation, the PCR product was sequenced and then translated to ensure the proper sequence of amino acids was generated.

This product was recombined into two different plasmids using Gateway technology of the attB sites. One plasmid was a pUAST-attB plasmid to be used for injection into flies, to produce overexpression transgenic flies of *exd* and *hth*. The second plasmid was a pAW-attB plasmid used in cell culture which has an *Actin5C*promoter so it can be constitutively active in s2 cells.

Generated cell culture constructs extradenticle and homothorax co-localize in the nucleus

To verify the generated cell culture plasmids function correctly and to verify the requirement for *hth* in order for *exd* to enter the nucleus, cell culture of *Drosophila* S2 cells were used to visualize localization of each homeobox gene.Transfection of cells with only *exd* and cells transfected only with *hth* shows that neither homeobox co-factor localizes in the nucleus (Figure 1A, B). However, when cells are transfected with both *hth* and *exd*, Hth and Exd were able to co-localize in the nucleus as illustrated in yellow (Figure 1C). This result demonstrated that the constructs that I had generated produced proteins, that are appropriate for in vivo analysis.

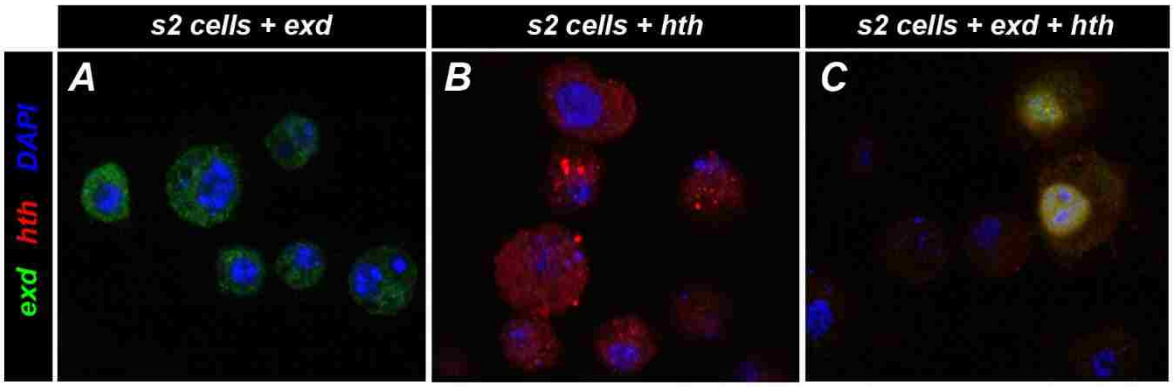


Figure 1: The homeobox co-factor *exd*requires *hth* in order to localize to the nucleus. (A) *Drosophila* s2 cells transfected with *exd*(B) cells transfected with *hth* (C) cells transfected with both *exd*and *hth*. Yellow illustrates the co-localization of *exd* and *hth* in the nucleus. Anti-*exd*is visualized in green, anti-*hth* is visualized in red, and blue is DAPI to denote the nucleus.

Knockdown and overexpression of homothorax and extradenticle change muscle biochemistry

Next, we wanted to determine biochemically how*hth* knockdowns affect the muscle of both the IFMs and TDT muscles. To identify if muscles retained their biochemical nature I performed a succinate dehydrogenase assay on cryosections of adult flies. Succinate dehydrogenase is an enzyme that is prevalent during the Krebs cycle of cellular

respiration and functions to oxidize succinate to fumarate to aid in the production of NADH, therefore it is used as a marker for highly oxidative muscles (Ackrell et al., 1992; Rustin et al., 2002). In wildtype sections, only the IFMs stained positive for SDH as expected (Figure 2A). However in *hth* knockdowns, the stain is significantly diminished in the IFM, indicating *hth* may have a vital role in maintaining the biochemical characteristics in the IFMs (Figure 2B). To further verify this result, we overexpressed *exd* and *hth* in the adult flies using the *UAS-exd* and *UAS-hth* constructs that I have generated above and performed the SDH assay again. Interestingly, SDH staining was expanded into the TDT muscle, suggesting the TDT had changed muscle identity from a neuronal activated muscle to that more similar to oxidative muscle (Figure 2C).

To further illustratehow muscle identity has changed in these crosses, I performed a histochemical assay to test for the presence of acetylcholinesterase in the neuronally activated TDT muscles. Cryosections of wildtype adult IFMs and TDT muscle confirmed the presence of acetylcholine inonly the TDT (Figure 2D). When *hth* was knocked down, acetylcholinesterase presence had increased into the IFMs, suggesting a biochemical transformation of the IFM to a TDT identity (Figure 2E). Furthermore, overexpression of *exd*and *hth* resulted in a complete loss of acetylcholinesterase in not only the IFMs but in the TDT muscle as well (Figure 2F). This result, in combination with the SDH stains, provided evidence that altering the expression of homeobox co-factors *hth* and *exd* can lead to a switch in muscle identity.

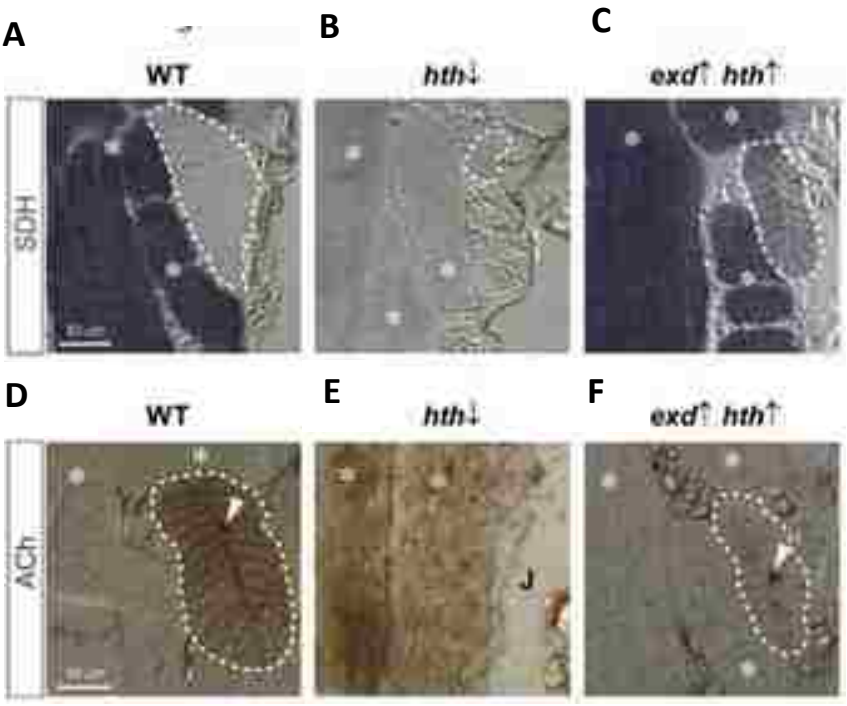


Figure 2: Homothorax and extradenticle are able to change muscle identity in adult flies.(A-C) A succinate dehydrogenase (SDH) stain of sections of wildtype, hth knockdowns, and exd/hth overexpression lines of adult flies of both the IFM and TDT. (D-F) An acetyl-cholinesterase (Ach) stain of cryosections of wiltype, *hth* knockdowns, and *exd*/*hth* overexpression lines of both the IFM and TDT. White dotted lines denote the TDT, asterisks indicate the locations of the IFMs and the white arrowheads are pointing to the nerve innervating the TDT.

DISCUSSION

The co-localization of Exd/Hth in the nucleus is dependent on the presence of Hth in order to translocate Exd into the nucleus (Jaw et al., 1999). As we have shown here, when we made our constructs for cell culture, Exd/Hth co-localized in the nucleus only when both were present, which is consistent with the literature. Interestingly, there is some evidence that this nuclear localization is due to competing signals that act upon one of the three domains of Exd. Exd has a nuclear localization signal, a nuclear export signal, and a specific Hth mediated region for nuclear localization in the presence of Hth binding (Abu-

Shaar et al., 1999). Upon Hth binding, Exd is localized in the nucleus, however, in the absence of Hth a nuclear export signal is activated and Exd is localized in the cytoplasm of the cell (Abu-Shaar et al., 1999).

Additionally, we have identified a new role for Exd/Hth contributing to muscle identity. Traditionally, Exd/Hth are found to play a role in the proximal-distal patterning of the limb, however, we illustrate that they also have a role in specifying specific muscle types as seen in Figure 2. The *C. elegans*homologs support this find as the homologs CEH-20 and UNC-62 function in mesodermal development (Jiang et al., 2009). Jiang et al., 2009 illustrates not only do CEH-20 and UNC-62 interact and co-localize similar to Exd/Hth and Meis1/Pbx1, but loss of UNC-62 results in the loss of cell fate and differentiation in the M lineage.

This work represents a portion of a larger work that was published in *Developmental Cell* in 2012 by Bryantsev et al. In this publication, we discovered *exd*/*hth*expression throughout adult muscle development was localized to the IFMs. In conjunction with Figure 2 from above, sections of frozen IFM and TDT muscles were stained for F-actin in both the wildtype and *hth* knockdowns. The sections showed that the IFMs of knockdown animals had presented a structure most similar to the TDT, which is tubular rather than fibrillar. On the contrary, the TDT took on the morphology of the IFM in *hth*plus exd ectopic expression. This was further verified through the use of electron microscopy images.

The phenotypic switch was confirmed molecularly by performing RT-PCR of wildtype and *hth*knockdown muscles of the IFM and TDT. Knockdown *hth*adult fly IFMs

began expressing jump muscle mRNA such as *actin79B* and *TpnC41C*, and the TDT muscle showed expression of IFM mRNA, *actin88F*, *flightin*, and *TpnC4*.

More interesting is the role of Pbx and Meis in controlling muscle fiber fate in vertebrates. As shown above, Pbx and Meis have homologs in *Drosophila*Exd/Hth, respectively, that control muscle fate and identity (Burglin, 1997; Bryantsev et al., 2012). Research in Pbx/Meis interaction has demonstrated the critical role it has in muscle identity through the formation of a complex with MyoD in order to bind the E box of the *myogenin* promoter (Heidt et al., 2007). More specifically, it has been shown that the loss of Pbx results in the repression of genes required for the specification of fast-twitch muscle, type II (Maves et al., 2007). Pbx/Meis also works with other myogenic regulatory factors early in development to aid in early skeletal muscle pattering and specification (Braun and Gautel, 2011). Despite its role in skeletal muscle differentiation, research has shown that Pbx also promotes cardiac muscle differentiation through interactions with hand in zebrafish (Maves et al., 2009).Taking our results together with those from the zebrafish studies indicates the possible existence of an evolutionarily conserved pathway to specify different muscle fiber types in animals, and provides an opportunity for this process to be genetically dissected in the *Drosophila* system.

ACKNOWLEDGMENTS

Thank you to Dr. Richard Mann for providing the lab with the anti-*hth* antibody.

Chapter 2: Identification of singles bar as a direct transcriptional target of Drosophila Myocyte enhancer factor-2 and a regulator of adult myoblast fusion

ABSTRACT

In *Drosophila*, myoblast fusion is a conserved process in which founder cells (FCs) and fusion competent myoblasts (FCMs) fuse to form a syncytial muscle fiber. Mutants for the myogenic regulator Myocyte enhancer factor-2 (MEF2) show a failure of myoblast fusion, indicating that MEF2 regulates the fusion process. Indeed, chromatin immunoprecipitation studies show that several genes involved in myoblast fusion are bound by MEF2 during embryogenesis. Of these, the MARVEL domain gene *singles bar* (*sing*), is down-regulated in MEF2 knockdown pupae, and has five consensus MEF2 binding sites within a 9000-bp region. To determine if MEF2 is an essential and direct regulator of sing during pupal muscle development, we identified a 315-bp myoblast enhancer of *sing*. This enhancer was active during myoblast fusion, and mutation of two MEF2 sites significantly decreased enhancer activity. We show that lack of *sing* expression resulted in adult lethality and muscle loss, due to a failure of fusion during the pupal stage. Additionally, we sought to determine if sing was required in either FCs or FCMs to support fusion. Interestingly, knockdown of *sing* in either population did not significantly affect fusion, however, knockdown in both FCs and FCMs resulted in muscles with significantly reduced nuclei numbers, provisionally indicating that *sing* function is required in either cell type, but not both. Finally, we found that MEF2 regulated sing expression at the embryonic stage through the same 315-bp enhancer, indicating that *sing* is a MEF2 target at both critical stages of myoblast fusion. Our studies define for the first time how MEF2 directly controls fusion at multiple stages of the life cycle, and provide further evidence that the

mechanisms of fusion characterized in *Drosophila* embryos is also used in the formation of the more complex adult muscles.

Keywords: *Drosophila*; Myoblast fusion; MEF2; Transcriptional regulation; MARVEL domain; *singles bar*

INTRODUCTION

Myoblast fusion is a conserved and critical process in the formation of mature, functional muscle fibers. Mammals and invertebrates share several of the key steps and components of the fusion process, from coalescence of myoblasts at sites of fusion, to membrane breakdown to generate the muscle syncytium (Richardson et al., 2008). In *Drosophila*, myoblast fusion begins by the designation of a founder cell (FC) and fusion competent myoblasts (FCMs). FCs differentially express a subset of genes, that function to attract FCMs, and fusion of the FC and the initial FCMs to form an early multi-nucleated muscle cell constitutes the initial round of fusion (Chen and Olson, 2004). Subsequent fusion of further FCMs to the nascent myotube complete myoblast fusion (Schroter et al., 2004). Several of the genes involved in each step of the fusion process are conserved between *Drosophila* and vertebrates: for example, myoblast adhesion can be partially attributed to the *Drosophila* protein Sticks and stones, for which Nephrin is the vertebrate ortholog (Rochlin et al., 2010); and the *Drosophila* protein Myoblast city is required during cytoskeletal rearrangement within fusing myoblasts (Erickson et al., 1997), as are the vertebrate orthologs, Dock1/Dock2 (Rochlin et al., 2010). Clearly, understanding the molecular mechanisms that regulate myoblast fusion in *Drosophila* can provide insight into the fusion process in vertebrates.

While numerous studies have identified genes required for embryonic myoblast fusion in *Drosophila* (Paululat et al., 1999, Chen and Olson, 2004 and Abmayr and Pavlath, 2012), less is known about the genes involved in the phase of fusion that occurs in the development of the adult muscles. For the adult thoracic muscles, the fusion process begins with the migration of adepithelial cells originating from the imaginal discs into the

developing thorax. While most adult muscles arise from de novo fusion of pupal FCs and FCMs (Dutta et al., 2004), the dorsal longitudinal muscles (DLMs) develop upon larval muscle templates (Fernandes et al., 1991), where the larval muscles function as FCs (Dutta et al., 2004). Of the few published studies on adult myoblast fusion, WASp, an actin nucleator required for embryonic myoblast fusion (Massarwa et al., 2007 and Schafer et al., 2007), is required at the time of adult myoblast fusion prior to pre-fusion complex formation (Mukherjee et al., 2011). The lack of WASp results in a complete hindrance of fusion in adult muscles (Mukherjee et al., 2011). More recently Gildor et al. (2012) showed that *sticks and stones*/*hibris* and *dumbfounded*/*roughest* have redundant functions in fusion of adult myoblasts. Thus, there are at least some commonalities in the mechanisms of myoblast fusion between embryos and pupae.

The transcriptional regulation of factors participating in adult myoblast fusion has not been investigated in detail. One candidate regulator is Myocyte enhancer factor-2 (MEF2). MEF2 is a conserved myogenic transcription factor that is critical for muscle differentiation in both skeletal and cardiac muscles (Potthoff and Olson, 2007). There are four orthologs of MEF2 in mammals while *Drosophila* has a single MEF2 gene, but for which the encoded protein shares the conserved A/T rich binding domain and function as a regulator of muscle differentiation (Lilly et al., 1995 and Bour et al., 1995). However, the genetic redundancy of MEF2 genes in vertebrates makes it difficult to study the context of MEF2 solely in relation to myoblast fusion events. In *Drosophila*, studies have indicated that MEF2 has an essential role in embryonic myoblast fusion, since mutation of *Mef2* resulted in unfused myoblasts in ß3-Tubulin-stained embryos (Bour et al., 1995). Expression in *Drosophila* of *Mef2* RNAi lines results in a lack of adult muscle formation

and the accumulation of unfused myoblasts in Mef2 knockdown pupae, also indicating a requirement for MEF2 in the fusion of adult myoblasts (Bryantsev et al., 2012 and Soler et al., 2012).

Embryonic chromatin immunoprecipitation-microarray (ChIP-chip) studies in *Drosophila* support the hypothesis that MEF2 is a direct regulator of fusion gene transcription (Sandmann et al., 2006). The fusion genes *blown fuse* (*blow*) and *lameduck* (*lmd*) are bound by MEF2 during embryonic muscle development, and loss of MEF2 results in loss of their expression (Chen and Olson, 2004 and Sandmann et al., 2006). Similarly, *roughest* (*rst*) is required for myoblast fusion (Strünkelnberg et al., 2001) and responds to MEF2 activity in the embryo (Apitz et al. 2005). Nevertheless, although ChIP-chip data suggests a critical role for MEF2 in the regulation of many fusion genes, binding data is not sufficient to determine if MEF2 is essential for fusion gene expression: the fusion gene *sticks and stones* (*sns*), an immunoglobulin family gene expressed in FCMs, has MEF2 binding sites both upstream and downstream of the gene, as determined by ChIP-chip analysis (Sandmann et al. 2006); however, *sns* expression in embryos is not MEF2 dependent (Bour et al., 2000), suggesting that although MEF2 binds to the region, it is not necessary for *sn*s gene expression. Instead other factors, or factors functioning redundantly with MEF2, must control *sns*transcription. In addition to *sns*, *blown fuse* expression is not affected in MEF2 mutants, indicating that MEF2 may not directly regulate fusion gene transcription despite the presence of MEF2 binding sites (Schroter et al., 2006).

There is some evidence that fusion genes may also be regulated by MEF2 in the pupal stages of myoblast fusion. We recently demonstrated that knockdown

of Mef2 function during pupal development resulted in a failure of adult myogenesis, including a complete lack of myoblast fusion. By using RT-PCR of RNA collected from control and Mef2 knockdown pupal myoblasts, the embryonic fusion gene *singles bar* (*sing*) was down-regulated in MEF2 knockdown samples (Bryantsev et al., 2012). Estrada et al. (2007)previously identified *sing* as encoding a protein with a conserved transmembrane protein known as a MARVEL domain. This domain is believed to function in junction formation between cells and vesicle trafficking in vertebrates (Sánchez-Pulido et al., 2002) suggesting that *sing* may be involved in the formation of the pre-fusion complex. The findings from Bryantsev et al. (2012) suggested firstly that MEF2 may be a direct and essential regulator of *sing* during myogenesis, and secondly that *sing* functions in myoblast fusion at both embryonic and pupal stages.

To test these hypotheses, we identify in this manuscript a 315-bp enhancer for sing expression that functions at both adult and embryonic stages of myoblast fusion. We show that *sing* expression is directly regulated by MEF2 via two conserved binding sites in the enhancer, and that the knockdown of *sing* during adult myoblast fusion results in lethality and drastically reduced muscle formation arising from a failure of myoblast fusion. We also demonstrate that, whereas *sing* expression is observed in FCs and FCMs in embryos, *sing* knockdown in both cell types is necessary for defects in fusion to be observed. Overall, our results identify a regulatory role for MEF2 in myoblast fusion at multiple stages of development, and identify *sing* as a fusion gene that functions during both the embryonic and pupal stages.

MATERIALS AND METHODS

Drosophila stocks and crosses

Stocks were maintained on Jazz-Mix Drosophila Fly Food (Fisher Scientific). *rp298-gal4* driver has been previously described (Nose et al., 1998 and Ruiz-Gomez et al., 2000). *Mef2-gal4* was from Dr. Aaron Johnson (University of Colorado at Denver), *sns-gal4* was from Dr. Elizabeth Chen (Johns Hopkins University Medical School), and *1151-gal4* was from Dr L.S. Shashidara (Anant et al., 1998). The *UAS-sing RNAi* lines, *P{GD3396}v12203* and *P{GD3396}v12202/TM3* were obtained from Vienna Drosophila RNAi Center. The *Mef2* knockdown line, UAS-*dcr; UAS-Mef2 RNAi(15550)* was described in Bryantsev et al. (2012). The *Mef2* null allele, *P544*, was balanced over a C*yO, wg-lacZ* balancer chromosome to enable visualization of homozygous mutant embryos.

Transgenic lines and mutagenesis

The following PCR primers were used to generate the *sing* enhancer using genomic DNA as a template:

Sing315-attB1:

5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCTTCCGCATAGACA-3’

Sing315-attB2:

5’-GGGGACCACTTTGTACAAGAAAGCTGGGTACAGAACGAACCCGAAATTG-3′

Gateway technology was used to clone the construct into pDONOR-lacZ-attB vector. Mutagenesis of the MEF2 sites in the *sing315-lacZ* construct was made by Gene

SOE-ing site directed mutagenesis (Horton, 1993). The following primers were used to mutate the MEF2 sites within the enhancer:

Sing315-1_Mef2_mutation_forward:

5′-AATTGCTGTTATGGTACCTACTGGAGATTG-3′

Sing315-1_Mef2_mutation_reverse:

5′-CAATCTCCAGTAGGTACCATAACAGCAATT-3′

Sing315-2_Mef2_mutation_forward:

5′-AATTGCTGTTATGGTACCTACTGGAGATTG-3′

Sing315-2_Mef2_mutation_reverse:

5’-ACCAGGTTTAGTACCATCTGCCGATAC-3′

Constructs for generating transgenic lines were injected into Drosophila embryos according to the protocol published by Rubin and Spradling (1982).

In situ hybridization

Embryos were collected on agar-grape juice plates at 25 °C and fixed according to standard protocols (Patel, 1994). In situ hybridization experiments were modified from a previously described method by the Berkeley Drosophila Genome Project (Weiszmann et al., 2009). RNA probes were made by amplification of *sing* from embryonic RNA using the following primers:

Sing_forward_with_HindIII:5’-AAGCTTATCAGTTGCAATCAGACC-3′

Sing_reverse_with_XhoI:5’-CTCGAGTGCTTTTGTCTGGCCG-3′

The resulting PCR product was cloned into pGEM-T Easy Vector (Promega) and linearized using restriction enzymes HindIII (New England BioLabs) and XhoI (New England BioLabs) for generation of sense and antisense probes, respectively.

Cryosectioning and immunostaining

Frozen sections of pupal samples were stained as described by Morriss et al. (2011). Briefly, pupae collected at 16, 18, 24, 30, and 48 h after puparium formation (APF), and those collected just prior to eclosion, had pupal casings removed prior to being submerged in Tissue-Tek OCT Compound (Sakura). Samples were frozen in liquid nitrogen and stored at −80 °C until ready for sectioning. Samples were horizontally sectioned at a thickness of 10–12 μm, and sections collected on a slide. Sections were fixed for eight minutes on a rotator in a 1:10 solution of 37% (v/v) formaldehyde and PBS. Slides were washed in PBTx [0.2%(w/v) BSA, 0.05% (v/v) Triton X-100, PBS] before incubation in Triton-X/PBTx solution for 30 min. Slides were incubated in primary antibody (anti-MEF2 diluted 1:1000, anti-Beta-galactosidase (Promega) diluted 1:1000, anti-Phospho-histone H3 (Thermo Scientific) diluted 1:400, and anti-Lamin (University of Iowa Development Studies Hybridoma Bank) diluted 1:10) in a humid chamber overnight before PBTx washing. Alexa Fluor secondary antibodies (Life Technologies) were diluted 1:300 in PBTx and incubated with sections in the dark at room temperature for 2 h. Rabbit anti-MEF2 was from Dr Bruce Paterson.

Fluorescence and confocal microscopy

Stained sections of pharate adults were imaged using an Olympus BX51 fluorescent microscope. High resolution images for nuclei counts were taken using a 20×, 0.8 NA objective lens on a Zeiss LSM710 confocal microscope, and images were captured using Zen software.

Nuclei counts

Nuclei counts from confocal images of stained adult muscle sections were recorded using the ITCN plugin for ImageJ (Rasband and ImageJ, 2014). All images were taken at 200× magnification on the confocal microscope. The threshold for detection was set to 0.8, nuclei width was set at 16 pixel, and nuclei distance was set to 8 pixel. Criteria for region specification for counting were based upon the largest continuous area of myoblasts or indirect flight muscle. Counts were normalized by determining the area of the region observed, and converting the nuclei counts from counts per square pixel to counts per 10,000 μm^2. A Dunnett–Tukey–Kramer pairwise multiple comparison test was used to determine significance between genotypic groups at p=0.05 level. Statistics and graphs were generated and programmed in R using the DTK package (Lau, 2013; R Core Team, 2013).

Electrophoretic mobility shift assay

MEF2 protein was generated using the TNT Coupled Reticulocyte Lysate System (Promega) using the pSK-MEF2 plasmid (Lilly et al., 1994). Details of binding conditions were as described in Gossett et al. (1989). The MEF2 site from *Act57B* was used as a

positive control (Kelly et al., 2002). Wild-type and mutant probe sequences were as follows (top strand shown only):

Sing315-1 5'-GGAATTGCTGTTCTAAATTTAGCTGGAGATTG-3′

Sing315-2 5'-GGGTATCGGCAGCTATTTATAGAACCTGGTTG-3′

Sing315-1 mut 5'-GGAATTGCTGTTATGGTACCTACTGGAGATTG-3′

Sing315-2 mut 5'-GGGTATCGGCAGATGGTACCTAAACCTGGTTG-3′

RESULTS

A 315bp enhancer upstream of sing containing two conserved MEF2 binding sites is active in adult myoblasts

To test the hypothesis that *sing* is a direct transcriptional target of MEF2 during pupal muscle development, we first sought to identify sequences that control *sing* expression. Sandmann et al. (2006) demonstrated, using ChIP-chip, that MEF2 bound to a ~4-kb region upstream of the *sing* transcription start site during embryogenesis (Fig. 1A). Additionally, there are five consensus MEF2 binding sites in the region of the*sing* gene (asterisks on Fig. 1A). We used these data as a starting point and amplified several fragments of genomic DNA to test for enhancer activity (Fig. 1A).

To determine if the DNA fragments had enhancer activity in pupal myoblasts, we fused them to *lacZ*reporter genes and generated transgenic animals carrying the *sing-lacZ* constructs. Homozygotes for the transgenic constructs were aged to 24 h after puparium formation (APF), and then frozen for cryosectioning and immunofluorescence. We chose 24 h APF as the time point, since this is the period during pupal development when myoblast fusion is occurring (Atreya and Fernandes, 2008). Moreover, high-

throughput RNA sequencing of *Drosophila* at different stages of development indicates that 24 h APF is the time at which peak pupal expression of *sing* is observed (St. Pierre et al., 2014).

In order to visualize the location of *sing-lacZ* activity relative to the swarming myoblasts, cryosections of transgenic pupae were stained with DAPI, and with antibodies against ß-Galactosidase (ßGal) and MEF2. We found that there was strong reporter expression in the myoblasts for only one construct, a 315-bp region that we termed sing315 (outlined in orange in Fig. 1A and B), demonstrating that the fragment of *sing*used in our assays had myoblast enhancer activity. Together with the observations from RNA sequencing analyses showing *sing* expression at this pupal time point (St. Pierre et al., 2014), plus the detection of *sing*transcripts in pupal myoblasts (Bryantsev et al., 2012), our data support the hypothesis that the 315-bp DNA fragment being tested is an enhancer for pupal myoblast expression of *sing*. Since none of the other fragments tested showed enhancer activity at adult nor embryonic stages (not shown), we conclude that sing315 is the predominant *cis*-regulatory region for *sing*.

To guide us in identifying important regulatory sequences within sing315, we next compared its sequence in *Drosophila melanogaster* with the equivalent sequences in four other Drosophila species. We observed strong sequence similarity close to the transcriptional start site, as well as several areas of conservation elsewhere in the enhancer. Notably, the two consensus MEF2 binding sites, YTA$(A/T)_4$TAR (Andres et al., 1995), were 100% conserved across the five species tested in our alignments (Fig. 1C), supporting the hypothesis that the MEF2 sites are important to *sing* expression. In more distantly-

related *Drosophila* species, the enhancer is less well conserved, however the most promoter-proximal MEF2 site is always conserved (not shown).

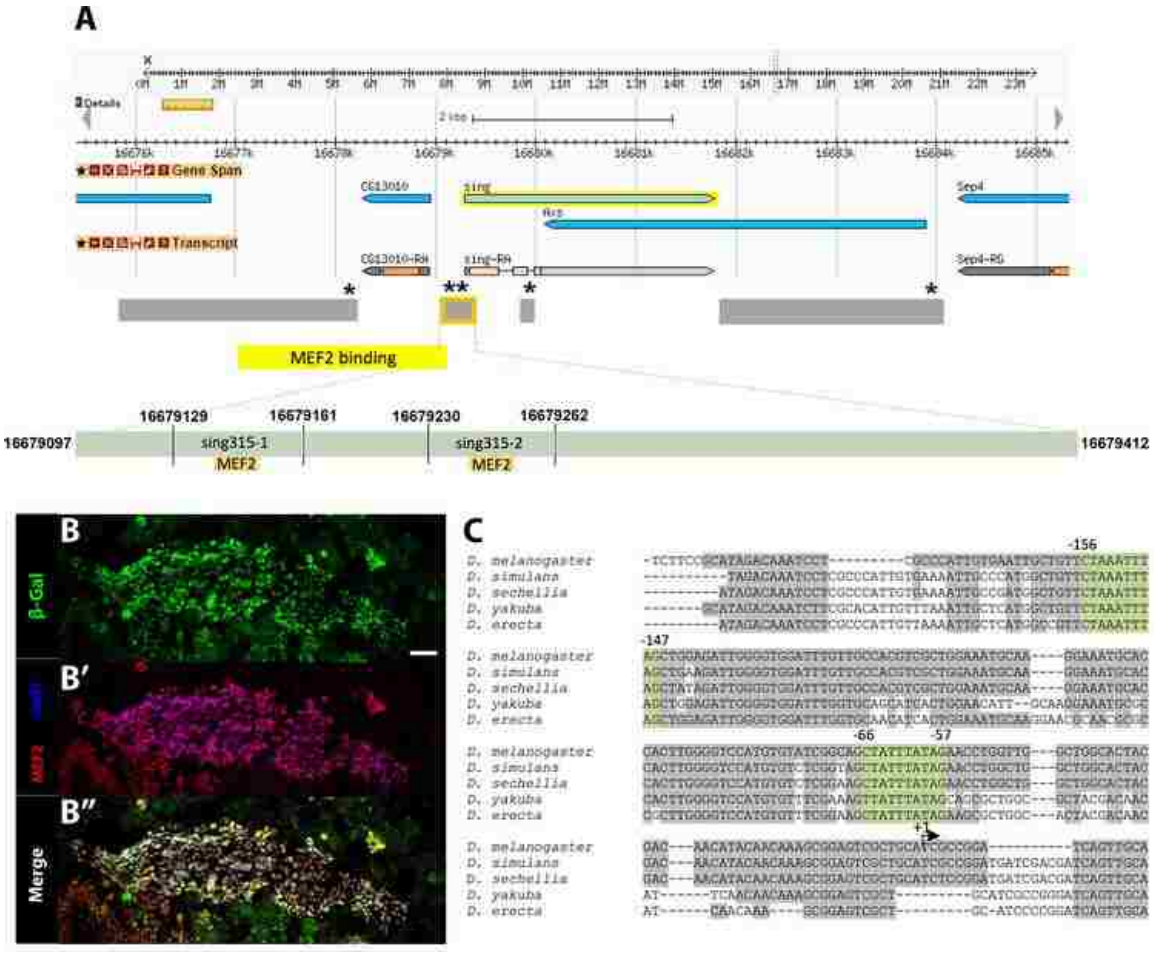


Figure 1: A 315-bp enhancer of *sing* containing two conserved MEF2 binding sites is active in adult myoblasts. (A) Diagram from Flybase.org of the genomic region surrounding *sing*, based upon Release 6 of the Drosophila genome. The regions tested for enhancer activity are shown in gray, and the genome region shown to bind MEF2 in ChIP-chip assays (Sandmann et al., 2006) is shown in yellow. The 315-bp enhancer is outlined in orange. Asterisks indicate the approximate locations of consensus MEF2 binding sites. A more detailed view of the enhancer region is shown below, with the two MEF2 binding sites highlighted in orange. Coordinates above the putative MEF2 sites indicate the sizes of probes used in DNA binding assays. (B-B″) Horizontal section of 24 h APF transgenic pupae carrying the *sing315-lacZ* reporter. A large area of cells was positive for ßGal (green), which corresponded to swarming myoblasts positive for MEF2 (red). Scale bar, 20µm. (C) The *Drosophila melanogaster* sing315 enhancer has two conserved MEF2 binding sites (highlighted in green) when compared to four other species of *Drosophila*.

Mutation of MEF2 sites in vitro and in vivo results in lack of MEF2 binding and diminished sing315 activity in adult myoblasts

To determine if MEF2 is capable of binding to either of the MEF2 sites, MEF2 protein was generated in vitro and used for electrophoretic mobility shift assays (EMSA). Double-stranded DNA corresponding to a known MEF2 site from *Drosophila Act57B* (Kelly et al., 2002), and to the two sites sing315-1 and sing315-2, were radioactively labeled with ^{32}P and then used in EMSA (Fig. 1A). *Actin57B* exhibited strong binding to MEF2 (Fig. 2A, lane 2). The addition of unlabeled *Actin57B* at a 100 fold greater concentration resulted in a decrease in the intensity of the shifted band (Fig. 2A, lane 3). When MEF2 was added to labeled sing315-1, strong binding to MEF2 was observed (Fig. 2A, lane 5). MEF2 also showed robust binding with labeled sing315-2 probe (Fig. 2A, lane 9). This confirmed that MEF2 is able to bind to both of the conserved MEF2 binding sites within the *sing* enhancer region.

To confirm that this binding was sequence-specific, we competed the MEF2-*sing* binding reactions with unlabeled wild type and mutant competitors, each at 100-fold greater concentration than the labeled probe. Both sing315-1 and sing315-2 showed almost a complete loss of MEF2 binding with the addition of the wild type competitor probe (Fig. 2A, lanes 6 and 10). When the MEF2 binding sites were mutated in the mutant competitor, nearly all binding expression was recovered in both sing315-1 and sing315-2 (Fig. 2A, lanes 7 and 11 respectively). This confirmed that MEF2 binding to both sites in sing315 was specific, and therefore supported our hypothesis that *MEF2* is a regulator of *sing* expression.

Next, we wanted to determine if MEF2 was a regulator of *sing315* expression in vivo. A construct of sing315 was generated in which both MEF2 binding sites were mutated. This construct was fused with a *lacZ*reporter, and inserted into the genome. Transgenic animals carrying the wild type *sing315-lacZ* construct, as well as those carrying the mutated *sing-lacZ* construct, were collected at 24 h APF. Samples were sectioned and stained in parallel, to assess the relative *lacZ* expression levels controlled by the wild-type and mutant enhancers. In both sections, myoblasts could be observed based upon co-localization of MEF2 and DAPI (Fig. 2B′ and C′). However, when accumulation of βGal was visualized, there was a significant reduction in reporter activity in the MEF2 mutated version of *sing315-lacZ* compared to the non-mutated*sing315-lacZ* (Fig. 2B″ and C″).

We also generated animals carrying the *sing315-lacZ* reporter and in which *Mef2* expression had been reduced using RNAi. We found that when MEF2 levels were strongly reduced, β-gal expression was diminished (Fig. 2D and D″). These results paralleled our prior observations that expression of endogenous *sing* was dependent upon MEF2 (Bryantsev et al., 2012), and therefore provided further support that MEF2 is a direct transcriptional regulator of *sing* expression during adult myogenesis.

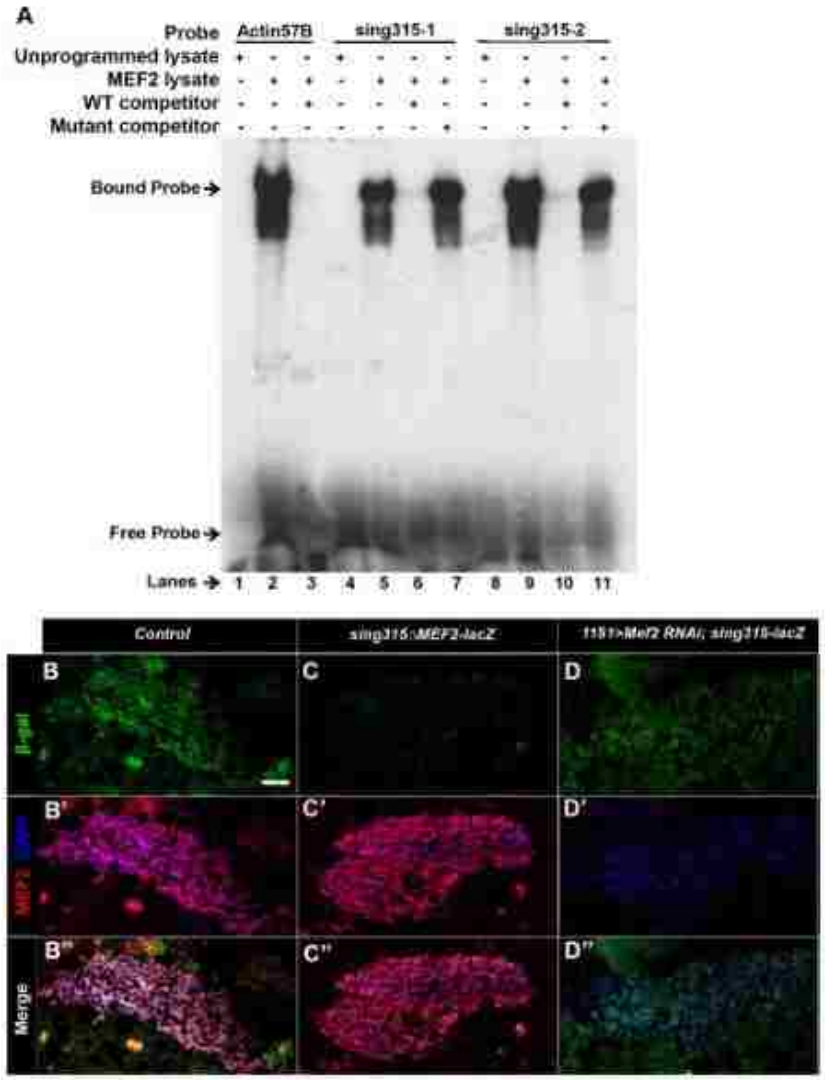


Figure 2: MEF2 binds to the*sing*enhancer, and the MEF2 sites are required for enhancer activity. (A) Electrophoretic mobility shift assay of MEF2 interacting with three different probes: *Actin57B* control (lanes 1-3), sing315-1 (lanes 4-7), and sing 315-2 (lanes 8-11). Wild type competitor was used in lanes 3, 6, and 10, and mutant competitor was used in lanes 7 and 11. MEF2 bound to the two sites in the *sing* enhancer and this interaction was sequence-specific, since wild-type sequences competed the interaction, whereas mutant sequences did not compete the interaction. The smear below the shifted band probably represents a minor modified or breakdown isoform of MEF2 interacting with the DNA. (B–B″) Horizontal section of 24 h APF *sing-lacZ* animals stained to visualize MEF2, DAPI, and βGal in adult myoblasts. Note the accumulation of the ßGal reporter in myoblasts. (C–C″) Horizontal section of 24 h APF transgenic animals carrying *sing-lacZ* with both MEF2 binding sites mutated. Sections were stained as in B. Note the absence of ßGal staining. (D–D″) Horizontal section of 24 h APF *1151>dcr+Mef2-RNAi* animals carrying *sing315*-lacZ.ßGal staining was diminished in the absence of MEF2. Scale bar, 20 μm.

Knockdown of sing during adult myoblast fusion results in reduced muscle formation and lethality at the pharate adult stage

Although MEF2 may be regulating other genes involved in the fusion process, the requirement of MEF2 for*sing* expression in pupal myoblasts provided one potential mechanism for the failure of myoblast fusion in*Mef2* knockdown pupae. In this model, MEF2 activates *sing* expression, which in turn is required for adult myoblast fusion.

To determine if *sing* is critical to adult myoblast fusion, we knocked down *sing* expression using a *Mef2-gal4* driver crossed to *UAS-sing RNAi*. In an initial experiment, we allowed control and *sing* knockdown pupae to develop to the pharate adult stage, after which control animals eclosed from the pupal case, but*sing* knockdowns were lethal. We used a *sing RNAi* line for this experiment in which the RNAi is not predicted to have any off-target effects, providing evidence that the phenotypes we observed were due to loss of *sing* expression, and not due to effects upon other genes. When knockdown adults were sectioned and stained with Phalloidin, anti-Lamin, and DAPI, we observed a considerable reduction in muscle mass in the *sing* knockdowns compared to wild type (Fig. 3A and B). Interestingly, the *sing* knockdowns still partially developed DLMs, although these muscles were smaller than normal. In the absence of significant fusion of myoblasts to the muscle templates, we propose that the muscles nevertheless grow and attempt to fulfill a role as DLMs. No other skeletal muscles were consistently observed in the *sing* knockdowns, indicating that *sing* function is essential for adult muscle development. In addition, the nuclei in the *sing*knockdown muscles were often clustered together, and always fewer in number compared to the homogenously dispersed nuclei in the wild type muscles (Fig. 3A' and B').

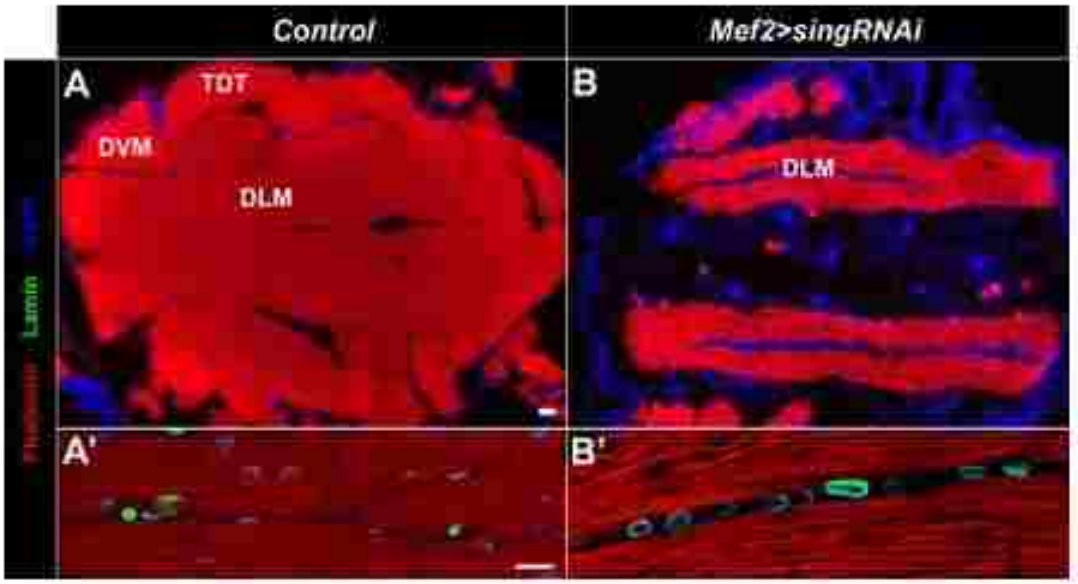


Figure 3:Knockdown of*sing*results in a failure of adult muscle formation. (A) Horizontal section of wild type flies at the pharate adult stage. The muscles, stained for accumulation of F-actin, are large and contain numerous nuclei. (B) Horizontal sections of *sing*knockdown flies at the pharate adult stage show there is a significant failure of muscle formation in the knockdowns. (A′) Higher magnification of control sample at the pharate adult stage showed robust muscle formation with numerous nuclei per muscle fiber. (B′) Higher magnification in *sing* knockdown animals. The residual muscles that do form are the DLMs, which are smaller than their control counterparts, and only have sparse nuclei. In all panels Phalloidin (red) was used to visualize F-actin, and DAPI (blue) was used to visualize nuclei. Lamin (green) was detected to outline nuclei in A′ and B′ panels. DLM, Dorsal longitudinal muscle; DVM, Dorsoventral muscle; TDT, tergal depressor of the trochanter (jump muscle). Scale bar, 20 μm for A, B; 10 μm for A', B'.

To determine if the phenotype observed in knockdown adults was a result of a fusion defect, we performed a time course analysis of muscle formation in control and sing knockdown pupae. At 16 h APF, we sectioned and stained samples with DAPI and anti-PH3 to determine myoblast number and myoblast proliferation rates. We found that myoblast number was slightly reduced in the sing knockdowns, but this difference was minimal and was not significant. In addition, myoblast proliferation rates were not significantly different between control and knockdown (Fig. 4A and B). These results

indicated that the lack of muscle seen in the sing knockdowns could not be attributed to a smaller starting pool of myoblasts, nor was it due to a slower myoblast proliferation rate.

We next assessed the formation of F-actin foci, a hallmark of fusing myoblasts. Knockdowns of *sing* at 18 h APF compared to control had normal formation of actin foci on the developing templates (arrows, Fig. 4C). Thus, the *sing* knockdown phenotype was not due to a failure of the FCMs to migrate to founder templates, nor due to a failure to initiate the process of fusion. Although actin foci formation appeared normal in the knockdown samples, a failure of fusion was evident at this time point because the developing templates contained founder cell nuclei (arrowheads, Fig. 4C) that were surrounded by few myoblast nuclei within the templates. This indicated that FCMs had not fused to the templates. To determine if the lack of fusion at 18 h APF was due to a failure of fusion, or simply due to a delay in fusion, we also studied samples at 24 h APF. At this later stage, the control templates had increased in size due to extensive fusion of FCMs with the templates, and by this stage F-actin foci were less evident in controls. In the *sing* knockdown, the templates were smaller, the F-actin foci were still apparent, and there was still little evidence of fusion (Fig. 4C). This result indicated that lack of *sing* expression caused a failure of myoblast fusion at the stage following the formation of F-actin foci. In addition, it suggested that when foci formed they remained stable when not resolved into a fusion event. Examination of stained sections staged to 30 h APF and 48 h APF revealed that the *sing* knockdown animals failed to form robust muscle compared to controls. In controls, the samples showed muscle forming at 30 h APF due to the accumulation of dense F-actin, and at later stages the formation of the adult jump muscle (TDT) and indirect flight muscles (DLM, DVM) could be observed. In

the *sing* knockdowns, it was difficult to discern any muscle formation based upon F-actin accumulation, other than a rudimentary DLM that must have arisen from the persistent larval templates. The defects in the knockdown animals arise presumably due to the lack of fusion in these samples. These results collectively suggested that the *sing* knockdown phenotype we characterized is indeed attributed to a fusion defect.

Taking all of this together, our data indicate that at least a part of the failure of fusion in *Mef2* knockdowns arises from a failure of MEF2 to activate *sing* expression. In this model, it would be predicted that *Mef2*knockdown myoblasts should not proceed past the formation of F-actin foci. To investigate this model, we sectioned and stained control, *sing* knockdown, and *Mef2* knockdown animals at 24 h APF, and determined if the *Mef2* knockdown myoblasts were capable of forming F-actin foci. We observed foci outlining the template in the MEF2 knockdown samples compared to the controls (arrows, Fig. 5A and B). This *Mef2* knockdown phenotype was similar to that for *sing* knockdown (Fig. 4 and Fig. 5C), consistent with the model described above.

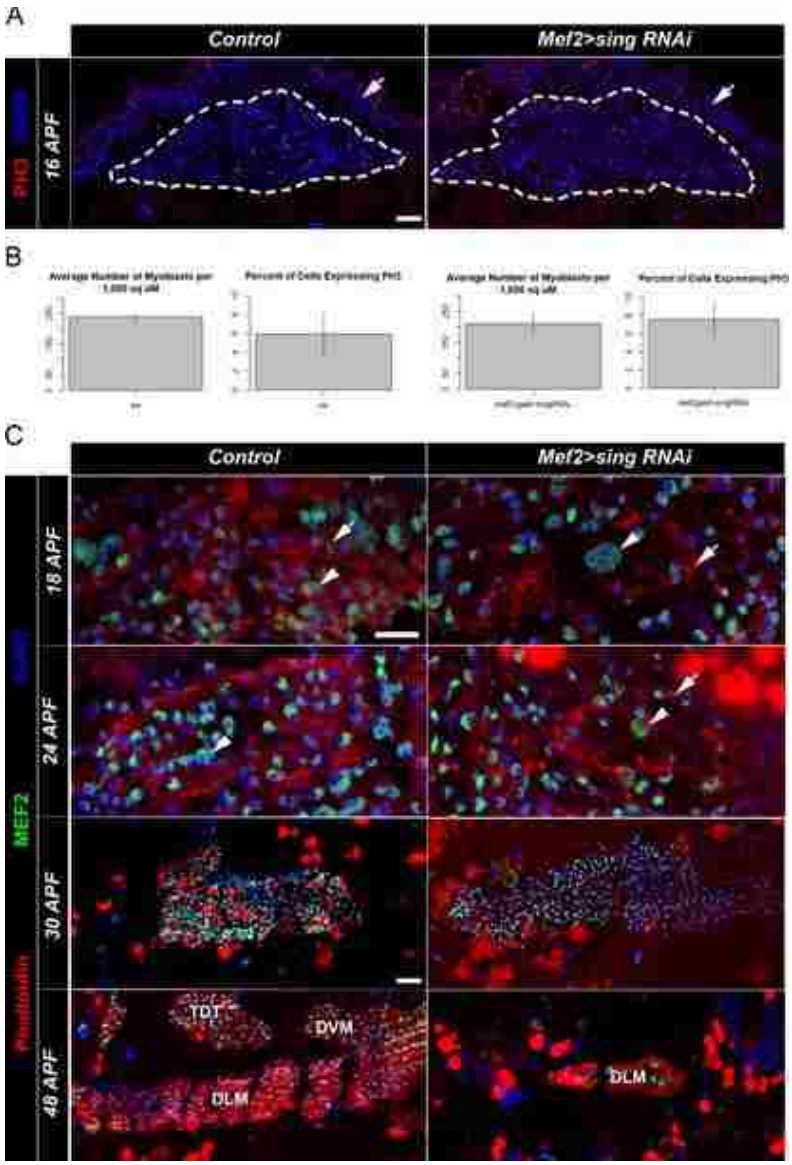


Fig. 4. The adult*sing*knockdown phenotype results from a failure of myoblast fusion. (A) Horizontal sections of control and *sing*knockdown animals at 16 APF respectively, stained for location of nuclei with DAPI (Blue), and for proliferating cells using anti-phospho-Histone3 (PH3, Red). Dotted lines indicate the pool of myoblasts, and arrowheads indicate the cuticle. Scale bar, 20 μm. (B) Quantification of myoblast density and proliferation in control and *sing* knockdown animals. There is no statistical significance between the control and *sing* knockdown, $p>0.05$. (C) Time course of developing adult thoracic muscles through adult myoblast fusion, comparing Control and *sing* knockdown samples. Larger FC nuclei are often apparent (arrowheads). In the *sing* knockdown animals the FCs have few closely-apposed nuclei, indicating that myoblast fusion is not occurring. Note that F-actin foci (arrows) are apparent in both wild type and *sing* knockdown samples. DLM, dorsal longitudinal muscle; DVM, dorsoventral muscle; TDT, tergal depressor of the trochanter, or jump muscle. Scale bar, 10 μm for 18 APF, 24 APF; 20 μm for 30APF, 48APF.

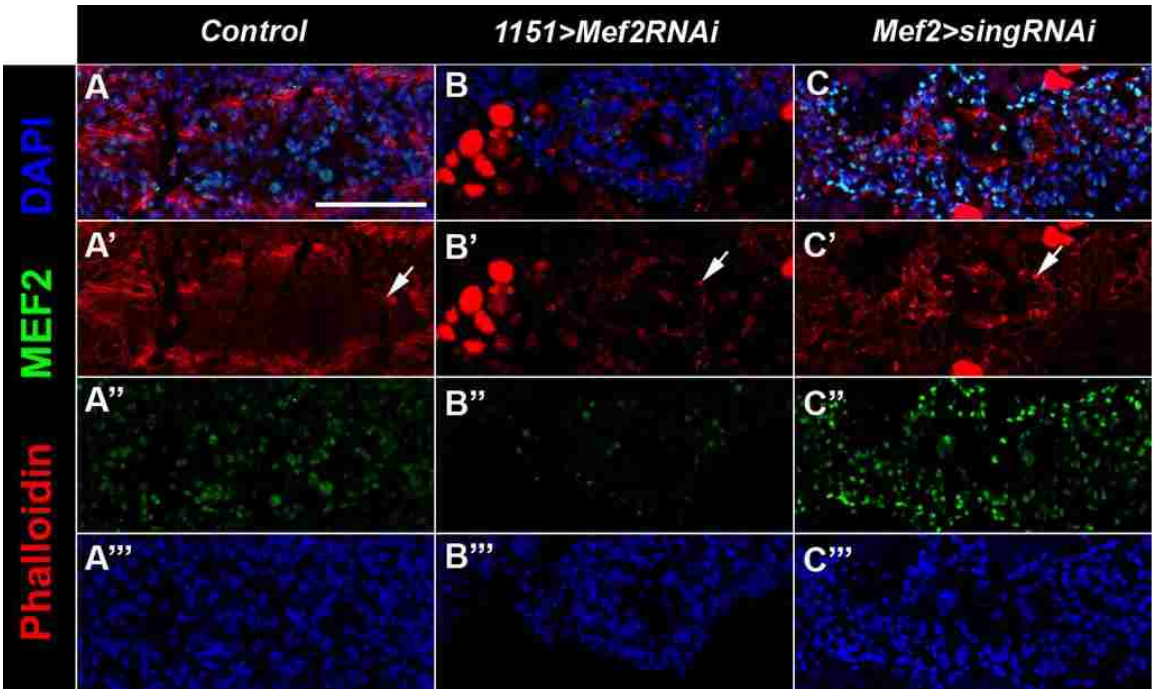


Figure 5:F-actin foci are detected in*sing*and*Mef2*knockdowns. (A–C) Horizontal sections of samples aged to 24 h APF stained with Phalloidin (red) to visualize F-actin, anti-MEF2 (green), and DAPI (blue) to visualize nuclei. Arrows mark F-actin foci. (A) Wild type control shows normal fusion of myoblasts to the larval templates. (B) *Mef2 RNAi* show smaller templates with fewer nuclei and pronounced actin foci. (C) *sing RNAi* shows actin foci at the periphery of the template. Scale bar, 50 µm.

Knockdown of sing in FCs and FCMs results in lethality and reduction in nuclei numbers

To test whether *sing* function is required in both the FCs and FCMs, or in just one cell type, we sought to knockdown *sing* expression individually in the FCs or the FCMs. To achieve this we used cell-specific Gal4 drivers for the FCs and FCMs. To assess the activities of the drivers, we first crossed each to U*AS-lacZ* and assessed reporter activity in pupae. As previously described, *Mef2-gal4* was active in both the FCs and the FCMs (Fig. 6A, left panel; Ranganayakulu et al., 1998), and *rp298-gal4* was active in the FC but not the FCMs (Fig. 6A, center panel; Nose et al., 1998). An FCM driver, *sns-gal4*, directed *lacZ* expression in FCMs immediately surrounding the FCs, but more

distantly-located FCMs did not show reporter activity (Fig. 6A, right panel; Stute et al., 2006). We interpret this result to indicate that *sns-gal4* becomes active in FCMs shortly prior to fusion. This activity mirrors expression of the endogenous *sns* gene in adult muscle development (Gildor et al., 2012). We also observed reporter activity in the FCs of *sns>lacZ* samples that we propose to arise from fusion of ßGal-positive FCMs to the FC templates.

We next used the cell-specific drivers to determine if we could uncover a role for *sing* in either the FCs of the FCMs. Using *rp298-gal4,* we expressed *sing RNAi* in just the FCs. The resulting progeny were 100% viable. When pharate adults were sectioned and stained for F-actin and MEF2, muscle formation was similar to that seen in wild type animals from the same stage (Fig. 6B and C). This result suggested that*sing* knockdown in the FCs was not enough to halt adult myoblast fusion. Similarly, when *sing* expression was knocked down in only FCMs, using *sns-gal4*, the progeny were 100% viable and muscles formed normally (Fig. 6D). These data suggested that *sing* might be required in either cell type, but that its presence is not essential in both FCs and FCMs.

To test this model, we also crossed flies in order to knock down *sing* simultaneously in FCs and FCMs. These progeny were lethal and died as pharate adults. Upon cryosectioning, whilst the muscles appeared robust, there was a clear reduction in the number of nuclei per muscle, and in many cases these nuclei appeared smaller than in other crosses (Fig. 6E). To determine whether the number of nuclei present in the double-driver knockdown was significantly different from the other samples, the number of nuclei per 10,000 square microns was calculated from confocal images of control and knockdown muscles. ImageJ was used to count nuclei, and the results of each group were plotted on a

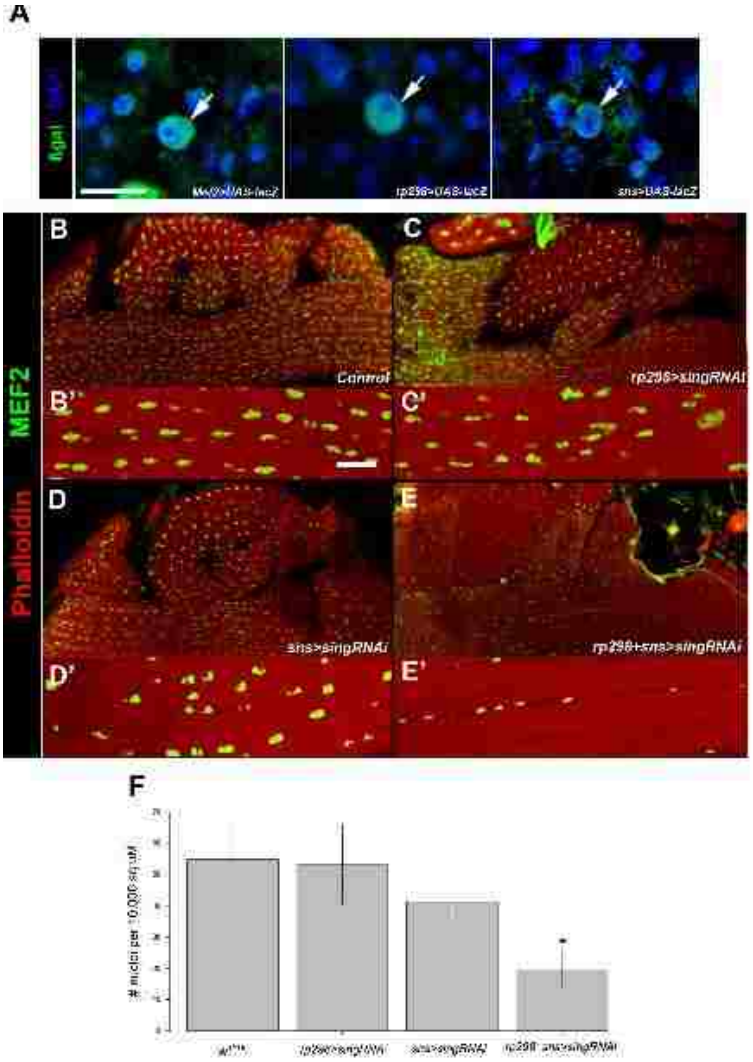


Figure 6: Knockdown of*sing*in both FCs and FCMs results in lethality and reduction in muscle nuclei numbers. (A) Horizontal sections of 13 h APF animals to show activities of *gal4* drivers used in FCs (arrows) and FCCs. *UAS-lacZ* was crossed to each driver, and samples were stained with anti-β-galactosidase (green) and DAPI (blue) to visualize nuclei. *Mef2-gal4* is active in all myoblasts;*rp298-gal4* is active in FCs; and *sns-gal4* is active in FCMs close to the template. ßGal accumulation in founder cell nuclei of*sns>lacZ* samples probably arises from fusion of ßGal-positive FCMs to the template. (B–E) Horizontal sections of pharate adults stained for accumulation of F-actin (Phalloidin) and MEF2. (B′–E′) Higher magnification views of muscle fibers and MEF2-positive nuclei (B–B′) Wild type; (C–C′) *sing* knockdown in founder cells; (D–D′) *sing* knockdown in fusion competent myoblasts. (E–E′) *sing*knockdown in founder cells plus fusion competent myoblasts. Note that muscle formation appears normal in all genotypes, but the size and number of nuclei is reduced in E and E′. (F) Quantification of average nuclei counts per unit area. Samples from the double driver are the only group that shows a significant reduction in nuclei number ($p>0.05$). Scale bar, 20 µm.

bar graph (Fig. 6F). A pairwise analysis of each group showed that the numbers of nuclei were significantly different in the double-driver group compared to each of the other samples; there was no significant difference seen between each of the other groups (Fig. 6F). Since we previously showed myoblast proliferation rate and myoblast numbers remained unaffected in *sing* knockdowns, we hypothesize the lowered nuclei counts in the *rp298+sns>sing RNAi* samples resulted from reduced myoblast fusion occurring. This suggests that *sing* expression is required in either the FCMs or FCs, but not both. Additionally, the lack of *sing* in both cell types results in lethality and lowered nuclei counts.

sing315 is active during embryonic myoblast fusion and is regulated by MEF2

Given that *sing* function was first characterized in the embryo (Estrada et al., 2007), we determined if the enhancer for adult myoblasts also functioned at the embryonic stage. Using in situ hybridization, we first confirmed that *sing* was expressed in myoblasts at stage 13, as previously demonstrated by Estrada et al. (2007) (Fig. 7A). Under the same conditions, we performed *sing* in situ hybridization in *Mef2* mutants, to determine if *sing* expression depended upon *Mef2* function. We saw diminished *sing* transcript levels in homozygous *Mef2* mutant embryos (Fig. 7B), consistent with our data in adults demonstrating that *sing* is genetically downstream of *Mef2*.

To determine if the embryonic expression of *sing* arises from the sing315 enhancer, we carried out immunofluorescent staining of the *sing-lacZ* embryos. We observed mesoderm-specific expression of the*lacZ* reporter at stage 13, based upon co-localization of β-galactosidase and MEF2 (Fig. 7C). In transgenic embryos carrying the *sing-lacZ* with

both MEF2 sites mutated, there was no expression of the *lacZ* reporter at any stage of embryonic development (Fig. 7D), indicating a direct role for MEF2 in activating *sing* at the embryonic stage as well as the pupal stage.

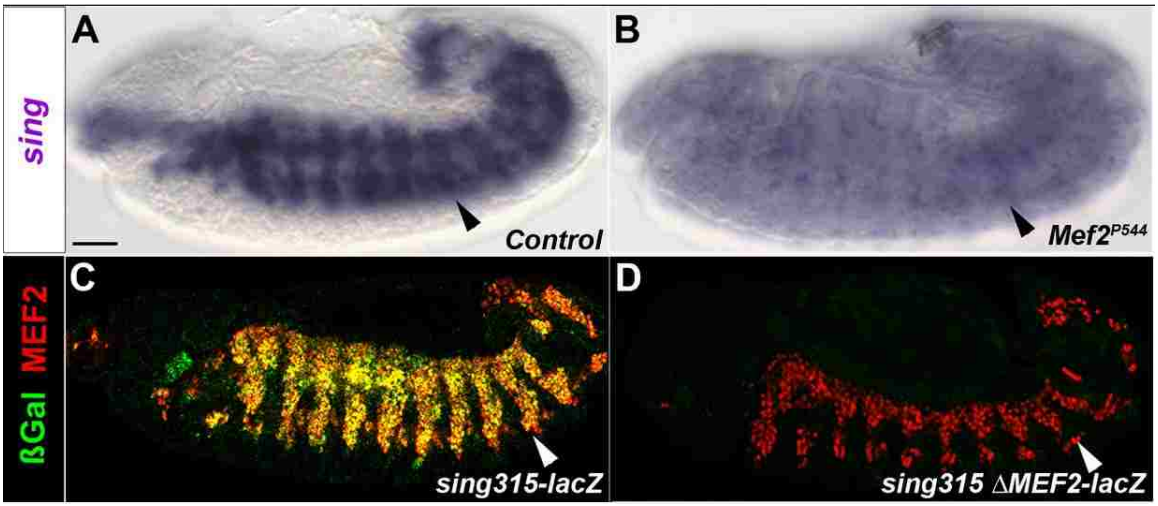


Figure 7: *sing*is directly regulated by MEF2 at the embryonic stage. (A–D) Stage 13 embryos. (A) In situ hybridization to detect *sing*transcripts in control embryo, with *sing* transcripts observed in myoblasts (arrowhead). (B) *sing* expression in *Mef2* null embryo is strongly diminished. (C) Immunofluorescent stain of *sing-lacZ* embryos displays co-localization of MEF2 and β-gal. (D) Immunofluorescent stain of *sing-lacZ* with mutated MEF2 sites lacks ßGal accumulation in embryos. Scale bar, 50 μm.

DISCUSSION

In this paper, we demonstrate that MEF2 is a transcriptional regulator of adult myoblast fusion, through direct activation of the fusion gene *singles bar*. We identify a 315-bp enhancer for *sing*, and show that mutation of conserved MEF2 sites in the enhancer results in a lack of enhancer activity during adult myoblast fusion. We also show that the knockdown of *sing* during adult muscle development results in pupal lethality and a strong reduction in muscle formation, and that this arises from a failure of fusion. Additionally we demonstrate that the 315 bp *sing-lacZ* enhancer is functional during embryonic

myoblast fusion and directly regulated by MEF2. Together our results show a direct role for MEF2 in myoblast fusion through the activation of *sing*.

Transcriptional control of myoblast fusion

The transcriptional regulation of myoblast fusion genes has received relatively little attention. MEF2 is thought to be a major activator of fusion gene expression, based upon both its requirement for fusion at embryonic and pupal stages (Bour et al., 1995 and Bryantsev et al., 2012), and its direct interaction with a number of fusion genes during embryogenesis (Sandmann et al., 2006). Here, we support these observations by demonstrating a direct and essential role for MEF2 in controlling *sing* expression and by indicating a requirement for *sing* in adult myoblast fusion. Together, our data and that previously published, provide a direct mechanistic link between MEF2 and myoblast fusion. While there are likely to be a number of additional MEF2 target genes that function in adult myoblast fusion, *sing* is the first such gene that has been demonstrated to be both required for adult myoblast fusion and that is directly regulated by MEF2.

Nevertheless, there are clearly a number of fusion genes whose expression is not absolutely dependent upon MEF2, either because their expression persists in *Mef2* null embryos such as *sns* (Bour et al., 2000), or because the fusion genes are not bound by MEF2 in embryonic ChIP-chip assays such as *rost* and *mbc* (Sandmann et al., 2006). Moreover, adult myoblasts can at least proceed to the F-actin foci stage of myoblast fusion in the absence of MEF2 function, indicating that genes controlling earlier steps of fusion might be expressed independently of MEF2. Identification of additional transcription factors that regulate fusion, and their target genes, will provide a more detailed mechanistic

insight into this process, and will also determine if a transcriptional network for fusion differs between FCs and FCMs.

We note that additional regulators of *sing* expression might still remain to be characterized. In addition to the MEF2 sites, other regions of the sing315 enhancer are evolutionarily conserved, including an E-box located between the two MEF2 sites. The E-box might be a target of activation by Twist, particularly sinceSandmann et al. (2007) identified *sing* as a target of Twist using ChIP-chip assays. On the other hand this E-box is not as well conserved in more divergent species (not shown), suggesting either that the E-box is not critical to *sing* activation, or that differing mechanisms for *sing* transcriptional activation might be used in more divergent species.

sing function is required for adult myoblast fusion

Our studies also show a requirement for *sing* in adult myoblast fusion, with the *sing* knockdown showing a failure of fusion, muscle loss, and pupal lethality. Close examination of the persistent DLM muscles reveals that a limited amount of fusion has occurred. This may indicate that our *sing* knockdown is not a fully effective knockdown, and that a small quantity of *sing* transcript is enough for cells to pass the pre-fusion complex. Nevertheless, it is clear that there is a major requirement for Sing in the formation of the adult muscles.

The persistence of the DLMs can be accounted for by the observation that DLMs form from larval muscle templates, rather than from de novo fusion of myoblasts to newly-specified FCs (Fernandes et al., 1991). It is interesting to note that large muscles can still be formed from the larval templates when there is little fusion, suggesting that relatively

small numbers of nuclei can support the formation of a larger muscle fiber. Interestingly, in WASp pupal knockdowns where there was a failure of fusion, there was no overt formation of the DLM (Mukherjee et al., 2011), which differs from our observations for the DLM. The differences in our observations may either result from some residual fusion taking place in the *sing* knockdowns; or from an additional requirement for WASp function at subsequent stages of muscle formation.

The function of sing in FCs and FCMs

sing is expressed in both the FCs and FCMs of the developing embryonic myoblasts (Estrada et al., 2007). Our studies show that the knockdown of *sing* in both the FCs and FCMs, using either *Mef2-Gal4* or a combination of *rp298-Gal4* and *sns-Gal4*, resulted in adult lethality and lowered number of nuclei in the muscles. Nevertheless, the phenotype was much stronger using *Mef2-Gal4*, suggesting that this driver more effectively silenced *sing* expression, probably by the *Mef2-Gal4* driver being active at a higher transcriptional level.

This conclusion impacts our interpretation of cell-specific knockdown studies, where we showed that knockdown of *sing* using drivers for FCs or FCMs did not significantly affect fusion, but that knockdown using the combined drivers affected fusion and muscle function. We interpret these results to mean that*sing* must be present in only one cell type for fusion to occur. Nevertheless we note that an alternative interpretation is that, only when the drivers were combined, was there sufficient RNAi produced to down-regulate *sing* expression. A resolution to these alternative explanations must await cell-specific drivers that are active at higher levels, or a more detailed molecular understanding of how Sing impacts myoblast fusion.

ACKNOWLEDGEMENTS

We thank Drs Susan Abmayr, Elizabeth Chen, Aaron Johnson, L.S. Shashidhara, and Bruce Paterson for reagents. This work was supported by GM061738 awarded by the NIH/NIGMS to RMC, and a March of Dimes Birth Defects Foundation grant 1-FY14-323 awarded to RMC. Brayon Freminwas supported by the Initiatives to Maximize Student Diversity program, NIGMS grant R25 GM060201. We acknowledge technical support from the Molecular Biology Facility at the Department of Biology, UNM, supported by NIH grant P20 GM103452 from the Institute Development Award (IDeA) Program of NIGMS.

Chapter 3: Identification of novel transcription factor binding sites near known MEF2 binding sites

ABSTRACT

Transcriptional regulation is a complicated process that requires the unwinding of the proper sites of DNA and the recruitment of several proteins to these exposed DNA sites to work together to signal RNA polymerase to start transcribing a gene of interest. These sites within the DNA that aid in the recruitment of RNA polymerase are bound by proteins such as transcription factors and co-factors. Although there are databases such as JASPAR and TRANSFAC that store a vast amount of known transcription factor binding site data, there are still many transcription factors binding sites that remain unknown. With the outpouring of data from next-generation sequencing techniques, genome data is abundant, which allows for data scientists to mine these nucleotide sequences for possible regions of regularity that may be conserved binding sites within the DNA. In this paper, we focus on Myocyte Enhancer Factor-2 (MEF2), a transcription factor that, as shown in Chapter 2, plays a critical role in muscle development. Using previously published MEF2-ChIP data, we aim to develop a new genetic algorithm that utilizes this data to find consensus sequences from a position weight matrix (PWM) using Shannon entropy as a measure of PWM fitness. These sequences may work in conjunction with MEF2 binding sites to bind factors to aid in muscle development in *Drosophila melanogaster*. We perform a series of computational most-frequent k-mer experiments to help validate the algorithm and to illustrate that the most conducive use of this algorithm is to pair with the k-mer experiments to get the most diverse population of conserved sequence candidates. In addition to computationally validating the algorithm, we identify these binding sequences predicted

by the GA and the k-mer data and perform an electrophoretic mobility shift assay (EMSA) with nuclear extract to confirm proteins can bind to these computationally identified sequences as molecular validation.

Keywords: MEF2, genetic algorithm, position weight matrix (PWM), transcription factor, Shannon entropy, k-mer

INTRODUCTION

Many factors contribute to the complex, yet elegant transcriptional regulation of muscle development. Myocyte enhancer factor-2 (MEF2) is a well-characterized example of one of the many transcription factors that regulate muscle development. MEF2 knockdowns fail to develop muscle in adult flies and flies null for MEF2 are not viable (Ranganayakulu et al., 1995; Bryantsev et al., 2012). Mutations in MEF2 have produced effects such as cell death in myoblasts and the lack of certain isoforms of myosin required in the dorsal vessel (Ranganayakulu et al., 1995). Although it has a pertinent role in this conserved process, MEF2 does not act alone to mediate muscle formation. Other known transcription factors in *Drosophila* such as Twist (Twi) are known to work in cooperation with MEF2 to aid in proper muscle development (Cripps et al., 1998).

However, other transcriptional regulators and co-factors may work with MEF2 in addition to Twist, underscored by the identification of MEF2 co-factors identified in vertebrates (Black and Cripps, 2010). The task of identifying conserved binding sites is one that is particularly difficult and problematic due to the large variation in "conserved" sites from wobble bases and the ability for sites to be positioned within a large window of nucleotides from a particular point of reference, such as a transcription factor start site or another transcription factor binding site.It is nearly impossible to identify these unknown regions by hand without the time and monetary expenses of performing numerous traditional wet laboratory experiments.

Within the last decade, bioinformatics and computational and systems biology has emerged as a field in order to address these time and monetary limitations. Additionally, the influx of large datasets due to next-generation sequencing technology and high-

throughput methods of obtaining data has made it possible to have access to lots of information in a short period of time. This has made it possible and necessary for data to be computationally analyzed. Currently, many algorithms and computational approaches for predicting possible transcription factor binding sites exist, however, many of them have limitations and many require complex probabilistic calculations in order to obtain an accurate prediction.

The traditional computational approach for detecting transcription factor binding site conservation is to use a position weight matrix (PWM). Sequences are aligned into a matrix, each row corresponding to a single sequence. The frequency of each nucleotide is then calculated into a probability of observing that nucleotide at the given position. The underlying assumption is that each nucleotide position in the matrix is probabilistically independent from the previous position (Ben-Gal, 2008). The nucleotide with the highest probability at each position is picked for the prediction of the overall consensus motif (Vavouri and Elgar, 2005; Compeau and Pevzner, 2014). This method works well when a motif can be localized to small region, however, it lacks sophistication and accuracy when searching for motifs de novo in large areas of sequence.

To combat this problem, researchers have either paired the PWM with numerous algorithms or modified the PWM to provide a more accurate prediction of transcriptional regulatory regions. Previous research has shown that consecutive nucleotides are not probabilistically independent of each other as assumed by PWMs (Bulyk et al., 2002). Therefore to address the interdependencies of consecutive nucleotides the PWM has been modified to make a dinucleotide weight matrix (DWM) increasing the accuracy of predicting transcription factor binding sites.Rather than looking at the positions of four

single nucleotides, the matrix considers 16 different pairs of nucleotides to make a prediction.Siddharthan (2010) has validated this using the yeast genome with 40 known transcription factor binding sites and observed a statistically significant increase in accuracy of predicting transcription factor binding sites. Software packages such as MOtif Occurrence Detection Suite (MOODS), which encompasses a wide range of algorithms to scan sequences data for the best PWM by finding subsets within a set of sequences that score the highest probability due to most matches and occurrences (Korhonen et al., 2009). This results in a fast and efficient way to find a general PWM over a large range of sequences.

Other approaches have used Bayesian statistics joined with hidden Markov models (HMM) to add quantitative measures for predicting motif conservation. In the hidden Markov model approach, positions in a DWM are used to define states in the Markov model. The probabilities are calculated using a forward-backward algorithm, such as the Baum-Welch algorithm, and the probabilities are assigned as the transition states using a learned maximum likelihood estimationfrom the data (Durbin et al., 1998; Mathelier and Wasserman, 2013). The combination of the DWM and the HMM have resulted in improved transcription factor binding site predictability and improved accuracy as validated through ChIP on chip data. In addition, this hybrid approach accounts for the interdependencies that occur between nucleotides while assigning a quantitative predictive value (Mathelier and Wasserman, 2013).Other hybrid algorithms involving HMMs such as Site Tracking and Recognition (SiTaR) developed by Fazius et al., 2011, have equally shown similar results. While this approach is more flexible and accurate, it can be time consuming and difficult to implement.

Other methods for transcription factor binding site discovery have involved the use of genetic algorithms. Genetic algorithms (GA) arise from the biological concept of evolution. A population of individual sequences each represent a chromosome in the GA. The chromosomes are assessed for fitness by a defined fitness function established based upon the convergence criteria (Beasley et al., 1993; Mitchell, 2009). After determining which population of chromosomes are deemed the most fit, they are placed back into the next generation of the algorithm and those deemed less fit have the opportunity to undergo mutation and crossover. This process is repeated over a set number of generations or until the convergence criteria is reached.

Genetic algorithms for discovering sequence conservation is not necessarily a new approach for finding new binding motifs. Many of the genetic algorithms currently focus on the optimization of PWMs. The flexibility of using a GA in the TFBS prediction due to the infinite measures for establishing a fitness function have resulted in several different GAs that aim to attempt to find the best solution to a similar problem by PWM optimization. Li et al., 2007 have introduced a GA they call GAPWM, which optimizes the PWM so a better consensus sequence can be derived. This method relies on starting PWM in order to find a local optimum, which may be problematic if a starting PWM is unknown. Additionally, the fitness criteria is a more complex function that accounts for the interdependencies between nucleotide interactions.SiteGA is an algorithm developed by Levitsky et al, 2007 in which the fitness measure for a PWM involves re-sampling the random shuffling of a set of sequences. By performing this operation, the algorithm is able to assign a probability to which local dinucleotides may have the greatest interactions, and therefore, derive a sequence that has the highest probability of occurring.

Despite the wide range of current algorithms for mining genomic data sequences, I have developed a novel genetic algorithm that can simplistically and quickly output regions of regularity in the genome. By predicting consensus sequences derived from a position weight matrix predicted by the GA using Shannon entropy as a measure of fitness, we have been able to bind nuclear extract to several of the sequences resulting from the GA. Using previously published ChIP data for MEF2 binding sites (Sandmann et al., 2007), we have been able to predict the known MEF2 binding sequence from the data as validation in addition to the prediction of new binding sites near these known MEF2 regions. Additionally, we show the best approach to finding sequence conservation is to pair this GA with a most-frequent k-mers approach in order to capture sequence diversity for conserved sequences. Pairing these techniques together, we have identified some potential candidate sequences that are located near MEF2 binding sites and that bind nuclear proteins that may cooperatively work with MEF2 in transcriptional regulation.

MATERIALS AND METHODS

MEF2 ChIPBinding Data

MEF2 ChIP binding data was obtained from the Furlong laboratory from their ChIP-on-chip data download on their website (Zinzen et al., 2009). Sequences of cytological chromosomal regions were identified using GBrowse from Flybase.org. All sequences were scanned to find the location of the consensus MEF2 binding sites. Sequences in which a MEF2 binding site was found, were then trimmed to add 200 base pairs upstream of the binding site and 200 base pairs downstream of the binding site for a

final sequence length of 410 base pairs. Any sequences that were found as duplicates in the data set were removed.

Entropy and Combinatorics Data

Shannon entropy was calculated based on the regularity of each position of each nucleotide in the entire population of sequences. Entropy was measured using the following equation:

A list of all possible 6-mers with their relative expected and observed probabilities were generated using a simple script coded in Python. Probabilities were obtained by the calculating the frequency of each nucleotide in the Furlong data set and assuming each nucleotide in the k-mer acts independently from each other.

Genetic Algorithm

The initial population is represented by a set of k-mers from the set of 410 nucleotide data set from above. The fitness of the population is determined by taking the sum of the entropies of ten consecutive nucleotide positions; the lowest entropy sum is considered the best fit population. Since not every position in the array must be occupied there is the possibility that a position is left empty and the entropy is 0. To deter the algorithm from recognizing these regions as one of low entropy, a 2.0 penalty is added to the sum for each empty position, the maximum entropy for any k-mer in a DNA sequence. Elitism is implemented by picking all the sequences that fall within the window of lowest entropy and keeping those sequences in their current alignment in the array for the next

generation. Sequences that do not lie within the best fit group qualify for the opportunity to undergo mutation. Mutation is represented as a shift in the alignment position of the sequence either left or right up to three positions. Any sequence that is not in the best fit population or is not chosen for mutation is either randomly realigned or stays in its current position. This process is repeated for 3000 generations.

Parameter Optimization

Optimization of array size and k-mer size were determined by the ability for the genetic algorithm to correctly identify an already know MEF2 binding site from the ChIP data. Array widths ranging from 50 characters up to 300 characters in length in conjunction with sequences of k-mer length 15, were tested using 10 independent GA runs of 3000 generations each. Array size of 100 nucleotides and 200 nucleotides scored equally high, therefore, the array size that returned the most MEF2 binding sites one mutation away was picked.

Electrophoretic mobility shift assay

Shifts used with embryonic lysate were derived from the nuclear extraction of 1 gram of wildtype embryos. Extraction was carried out as described in Pazin, 2000. Adult lysate was used from the nuclear extraction of 50 pharateadult wildtype flies ranging from 72-96APF. Nuclear extract was prepared as described by Schreiber et al., 1989.The EMSA was carried out according to the protocol described in Gossett et al. (1989). The MEF2 site from *Act57B* was used as a positive control (Kelly et al., 2002). The probes used

from the output of the GA were ordered through Sigma Aldrich and the sequences were as follows (top strand shown only):

0-105-01 5’-GGATTAATATATATATTTATTATTTA-3’

0-105-02 5’-GGAAAAATAATAAAATATTAATTATA-3’

0-105-08 5’-GGTTTAAATATATATATTTAAAAAAT-3’

50-156-01 5’-GGTTTTATATATATTTTAAATTATTG-3’

50-156-06 5’-GGTTTTTTATAAAATATTTTTTTATT-3’

50-156-08 5’-GGAAAAAATATTAAATTTTAATATGT-3’

200-306-01 5’-GGATATTATTTATTAAAAAATAATAA-3’

200-306-06 5’-GGTAATATTTTTATAATATTATTTTA-3’

200-306-09 5’-GGAATATTTTTATTTAATATTATTTT-3’

250-356-04 5’-GGTTAATAAATATAATAAAAAATATA-3’

250-356-08 5’-GGAGAATATATATATTTATTTTATTT-3’

250-356-10 5’-GGAATAATAAAAATTAAATAATATTA-3’

300-405-02 5’-GGAAATATTTTATTTTATATTTAAAG-3’

300-405-03 5’-GGAAATATTTTAATATTTTTAATTTC-3’

300-405-09 5’-GGAATATAATAATTTTAAATTATAAA-3’

The probes used to check 6-mers (underlined) that had the greatest fold increase in frequency are as follows (top strand shown only):

5’-GGGAGGAGAAGCGGCAGCAGAAAAATCC-3’

5’-GGTTCAACTGCGGCGGCGACTAGAATAA-3’

5’-GGTTAGTTGGTGGGCGGGGGTAGTGAGG-3’

5’-GGATGATACTGTGGCGGCGATATAAAAT-3’

5’-GGATTAACAGCGGCAGCAGAAAGAGTAA-3’

5’-GGAGTAGATGCAGCAGCGGCAGCATACA-3’

5’-GGCATAAGTATTGCTGGCTAAATATAAA-3’

5’-GGTGTGTGTGTGTGTGTGTGTGTGTGTG-3’

The probes are presented in descending order of greatest fold increase comparing observed to expected frequency, to lower fold increase from observed to expected frequency. The 6-mers of conservation are located in the middle of the sequence. The flanking regions were determined by a finding all the occurrences of the 6-mer, aligning them to each other, and creating a PWM to find the most frequent nucleotides at each position. Ten nucleotides upstream of the 6-mer sequence and 10 nucleotides downstream of the 6-mer sequence were used for a total probe length of 26 nucleotides before the 5’-GG tag.

RESULTS

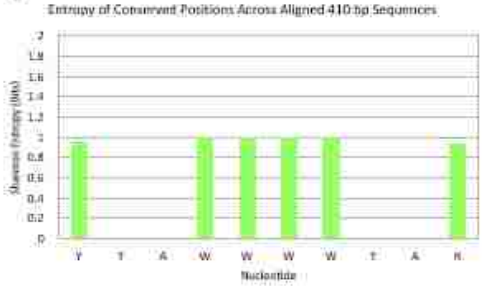
Alignment of 410-base pair sequences shows regularity only in known MEF2 sites

To test our hypothesis that there are conserved regions near known MEF2 binding sites, I initially took a traditional approach by perfectly aligning each nucleotide position of each of the possible 385 sequences and calculated the entropy at each position (Figure 1). To ensure we did not miss any significant regularity at each position, we lowered our stringency to pick out any positions that have a value greater than or equal to an entropy reading of 1.5 bits, since the maximum Shannon entropy at any given position is 2.0 bits. This alsoallowed us to validate our data set to ensure that the positions in which MEF2 is

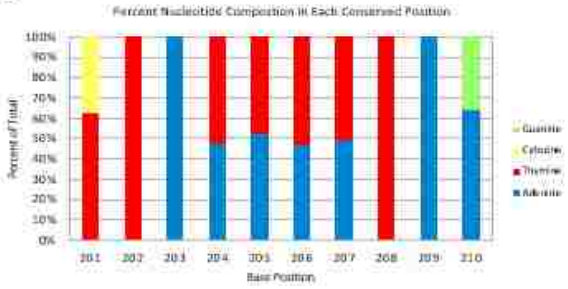
A

Adenine	Thymine	Cytosine	Guanine	Entropy
0	241	144	0	0.95371326
0	385	0	0	0
385	0	0	0	0
183	202	0	0	0.99824245
201	184	0	0	0.9985931
179	206	0	0	0.99644936
189	196	0	0	0.99976152
0	385	0	0	0
385	0	0	0	0
246	0	0	139	0.94354223

B



C



[redacted]

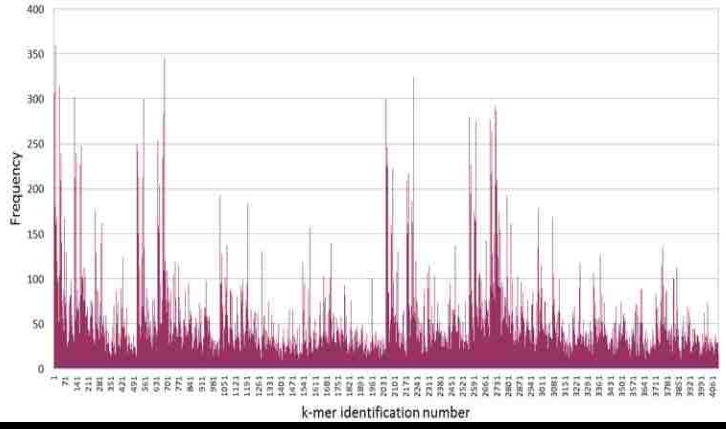
located are extracted as expected. This procedure revealed regions of perfect regularity with an entropy of 0, and regions that have a considerable amount of regularity with entropies reading around 1.0 (Figure 1A, B). Despite having the program output regions with an entropy value of 1.5 bits or lower, the only conserved regions were positioned at base pair numbers 201-210 of each of the 385 possible sequences tested (Figure 1C). These positions correspond exactly to the known MEF2 sites of each sequence.

However, I did not find any additional regions of sequence showing low entropy. This does not mean that there are zero additional conserved regions in the enhancers being analyzed. Rather, this result indicates that any conserved sequences that exist are not

spaced a specific distance from the MEF2 sites. They instead are spaced by varying distances.

Most frequent 6-mers are those that have high A-T nucleotide content and most of them belong to the MEF2 binding site

A



B

Sequences occurring >=300	
AAAATA	359
ATTTTT	346
TATTTT	325
AAATAA	314
AAAAAT	307
AATAAA	302
TAAAAA	300
ATATTT	300

[redacted]

Since using the PWM methodology did not find any new regions of regularity, I decided to determine if there were sequences of nucleotides that are seen more frequently in the MEF2 ChIP dataset than others. Upon finding all possible combinations of 6-mers, I wrote a Python script that counted the frequencies of each 6-mer that occurred in the MEF2 ChIP data set. Although the average length of a transcription factor binding site is roughly 10 nucleotides, there are several transcription factor binding sites that fall below

this length (Stewart et al., 2012). Therefore, in order to prevent the over-shadowing of smaller transcription factor binding sites, I decided to make k-mers of 6 characters in length.

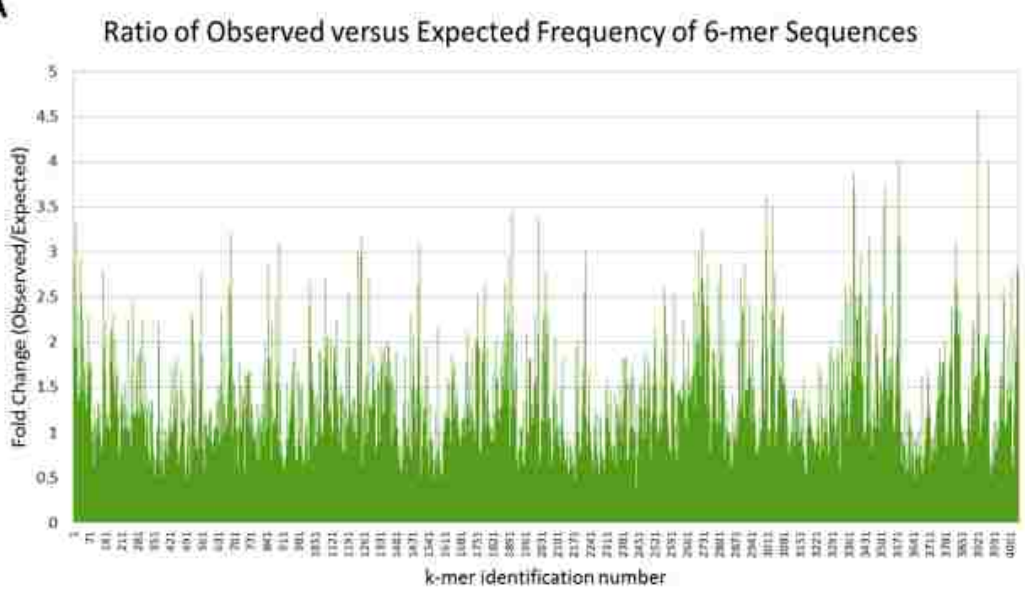
I noticed overall, there were certainly 6-mers that were represented more frequently in the data than others (Figure 2A). In order to narrow down the number of 6-mers, I divided the 6-mers into 4 categories: those appearing at least 300 times, those appearing at least 250 times, those appearing at least 200 times, and those appearing at least 150 times. I observed that there were 85 sequences that appeared at least 150 times (data not shown), 46 sequences that appeared at least 200 times (data not shown), 21 sequences that appeared at least 250 times (data not shown), and 8 sequences that appeared at least 300 times (Figure 2B). Of those sequences that appeared at least 300 times, all but one, AATAAA, are part the MEF2 consensus binding site (Figure 2B). Interestingly, of the 6-mers that appeared at least 150 times, most were also A-T rich (data not shown). Only 21 of the 85 sequences had a G or T in the 6-mer and only one of the 85 sequences had a total of two cytosines in the 6-mer which had a frequency of 156 occurrences in the ChIP data (data not shown). Therefore, this provides a certain level of validation in that the MEF2 binding site appears as someof the most frequent 6-mers.

Highest fold increase in change from expected are those that are G-C rich or have sites conserved within the MEF2 binding site

In order to determine if there are regions of regularity that are masked by the A-T "rich" nature of the data set, I determined the expected frequency of each 6-mer relative to the observed frequency in the data set. I calculated the expected frequency by making the

assumption that nucleotides appear independently of each other. I calculated the total frequency of each nucleotide in the MEF2 ChIP data set as 30% each of adenine and thymine, and 20% each of cytosine and guanine, which is consistent with the published literature on the Drosophila genome composition (Hastings and Kirby, 1965). When comparing the ratio of observed frequency to

A



B

Fold Change >=3.0	Observed Frequency	Expected Probability	Expected Frequency	Obs/Exp Freq
GGCAGC	65	9.60E-005	14.2296	4.56794288
GCGGCG	38	6.40E-005	9.4864	4.005734525
GGCGGG	38	6.40E-005	9.4864	4.005734525
GGCGGC	37	6.40E-005	9.4864	3.900320459
GCAGCA	83	0.000144	21.3444	3.888607785
GCAGCG	53	9.60E-005	14.2296	3.724630348
GCTGGC	53	9.60E-005	14.2296	3.724630348
TGTGTG	116	0.000216	32.0166	3.623120506
TGGCCA	75	0.000144	21.3444	3.513802215
GCTGCC	50	9.60E-005	14.2296	3.513802215
CGCTGC	49	9.60E-005	14.2296	3.443526171
CGGCGG	32	6.40E-005	9.4864	3.373250127
AAAATA	359	0.000729	108.056025	3.322350605
TTTTCC	156	0.000324	48.0249	3.248314937
* ATTTTT	346	0.000729	108.056025	3.202042644
CAGCGG	45	9.60E-005	14.2296	3.162421994
GCCGCT	45	9.60E-005	14.2296	3.162421994
GCGGCA	45	9.60E-005	14.2296	3.162421994
GCGGGC	30	6.40E-005	9.4864	3.162421994
GTGTGT	100	0.000216	32.0166	3.123379747
AGCGGC	44	9.60E-005	14.2296	3.092145949
CCGCTG	44	9.60E-005	14.2296	3.092145949
TGTGCG	65	0.000144	21.3444	3.045295253
* TATTTT	325	0.000729	108.056025	3.007699015

[redacted]

expected frequency I noticed there were some 6-mers that appeared more frequently than expected (Figure 3A). Of those sequences where the ratio fold-change was at least 3.0, several of the sequences had a high GC content (Figure 3B). Although they appear with higher frequency than expected, they are much less frequent in their expected probabilities than their AT rich counterparts (Figure 3B).

The genetic algorithm is consistent with data represented by the most frequent 6-mers

Next, we wanted to determine if we could use the MEF2 ChIP data directly to identify conserved regions near the MEF2 binding sites. Since transcription factor binding sites can be within a window of position variability from an already known transcription factor site, we decided to design a genetic algorithm in an attempt to address the flexibility in position.

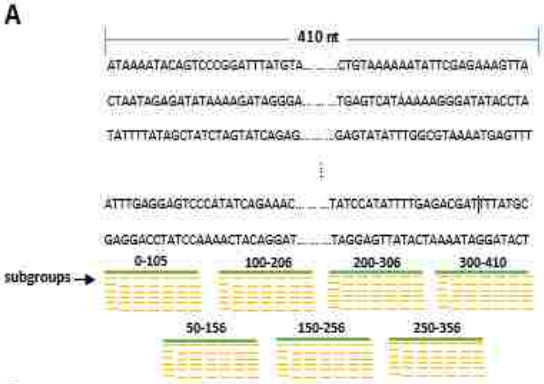
The data set is first broken down and tiled into segments of roughly 105 nucleotide subsets (Figure 4A, green lines). Each subset is then further segmented in 15 nucleotide k-mers consecutively aligned, all of which is done via a Python program (Figure 4A, yellow lines). The population size in each subset ranges from 2695to 3080 different 15-mers. These parameters ensure the GA is able to run in a timely manner by starting with a smaller population size, and additionally, it ensures the lack of redundancy and over-representation of AT rich regions by preventing k-mer overlap.

The GA was run over 3000 generations for each subgroup population (Figure 4B). The GA works by initialization of the population by reading in a text file of 15-mers generated by a Python program I had written to modify and mutate the ChIP data. Upon initialization of the population, each 15-mer was randomly aligned to an array of size 100

nucleotides long before being exposed to the GA for 3000 iterations (Figure 4B). Once aligned, the entropy of each of the 100 positions in the array was calculated and the sum of every possible 10 consecutive position entropy values was stored in memory. The window with the smallest sum was chosen as the best fit window in the matrix. Any sequences that were located within the window remained in their current position for the next generation. The remaining 15-mers were then randomly chosen for mutation. In this GA, a mutation does not refer to a bit switch or character change in a string. Instead, those k-mers selected for mutation were given the option to shift their current positions to the right or left within the area by a maximum of 3 positions. Any 15-mers not selected for mutation maintained their current position. The fitness of the population was reassessed by calculating the entropy at each position and once again taking the sum of every 10 consecutive positions and picking the smallest window as best fit. This process is repeated for 3000 generations (Figure 4B).

After 3000 generations, the sequences that had the overall lowest entropy sum were selected for the final consensus string output by the GA (Figure 4B). This was repeated ten times for each subgroup. The top three strings in each subgroup with their corresponding window entropy selected from the output of the GA are shown in Figure 4C. All the sequences primarily consist of adenine and thymine nucleotides however, importantly,they do not possess the typical YTAWWWWTAR MEF2 binding sequence and might correspond to binding sites for novel nuclear proteins.

A



B

```
Input: A set of k-mers from one subgroup

0.  matrix=100;
1.  for all k-mers in population
2.      randomly select a number for position of 1st character of k-mer
3.      Place 1st character in the matrix of randomly picked number
4.
5.  def fitness():
6.      for every column in matrix
7.          Calculate position entropy
8.      for every position in matrix – 10
9.          Sum consecutive 10 position entropies
10.     currentBestFit=min(sum consecutive 10 position entropies)
11.
12. def mutation():
13.     if k-mer is in overallBestFit
14.         continue
15.     else
16.         randomly pick 0 or 1
17.         if pick 1
18.             randomly select whole number between [-3, 3] but not 0
19.             shift k-mer position in matrix by number selected
20.         else
21.             k-mer keeps current alignment in matrix
22.
23. for 3000 generations:
24.     fitness();
25.     mutation();
26.     if currentBestFit<overallBestFit:
27.         overallBestFit=currentBestFit
28.
29. return overallBestFit
```

C

String ID	Entropy	String Output
0-105-01	18.1818438073	ATTAATATATATATTTATTATTTA
0-105-02	18.1792797289	AAAAATAATAAAATATTAATTATA
0-105-08	18.1682158093	TTTAAATATATATATTTAAAAAAT
50-156-01	18.1892048265	TTTTATATATATTTTAAATTATTG
50-156-06	17.9191576606	TTTTTTATAAAATATTTTTTTATT
50-156-08	18.2704441793	AAAAAATATTAAATTTTAATATGT
200-306-01	17.6114872589	ATATTATTTATTAAAAAATAATAA
200-306-06	17.1095308187	TAATATTTTTATAATATTATTTTA
200-306-09	17.0978393544	AATATTTTTATTTAATATTATTTT
250-356-04	17.9466810388	TTAATAAATATAATAAAAAATATA
250-356-08	18.1238145313	AGAATATATATATTTATTTTATTT
250-356-10	17.9951246634	AATAATAAAAATTAAATAATATTA
300-405-02	18.0841598528	AAATATTTTATTTTATATTTAAAG
300-405-03	18.1950925202	AAATATTTTAATATTTTTAATTTC
300-405-09	18.1761941154	AATATAATAATTTTAAATTATAAA

Figure 4: Top three results of each subgroup from GA show regions of conservation in AT-rich sequences. (A) Schematic showing MEF2 ChIP data breakdown for the GA (B) Flow diagram illustrating the process the algorithm follows in order to build a string with the lowest entropy in each subgroup (C)The resulting strings from the GA from each subgroup reveals regions that have the lowest entropy tend to be those that are high in AT content. The top 3 strings with the lowest entropy that do not have the consensus MEF2 site were chosen for each subgroup. The first two numbers in the string ID correspond to the subgroup data that was used and the last number corresponds to the experiment ID.

Several sequences from GA bind proteins from nuclear extract that are not MEF2 protein

In order to determine if the sequences predicted by the GA are regions of importance for transcription factor binding, I performed an electrophoretic mobility shift assay (EMSA) with nuclear extract.Each of the sequences in Figure 4C was annealed to

the corresponding anti-sense strand to generate double-stranded probeslabeled with radioactive ^{32}P. In addition to the GA probes, a double-stranded sequence upstream of *Actin57B* that is known to bind MEF2 was generated and used as a positive control (Kelly et al., 2002). Nuclear extract from 1 gram of wildtype embryos was collected and extracted for use in the EMSA. The positive control, *Act57B*, bound a protein in the extract that was previously shown to correspond to MEF2 (Figure 5A, lane 2). Additionally, two other shiftsthat differed from MEF2 in their mobility were observed (Figure 5A, lanes 3-6, 8, 13, 16). The shifts running just below the MEF2 shift corresponded to the following GA sequences: 0-105-01, 0-105-08, 50-156-01, and 250-356-08 (Figure 5A, lanes 3, 5, 6, 13). The shifts that are located farther down the gel were: 0-105-02, 50-156-08, and 300-405-03 (Figure 5A, lanes 4, 8, 16).

Next, I determined whether these sequences were able to bind to nuclear extract at later stages in development, possibly indicating the sequence is critical at both the early and late stages of development. To address this question, I extracted nuclear extract from adult flies 72-96h APF, nearing the time of eclosion. Similar to the embryonic shift, I observed that the same two shift locations were seen in the adult nuclear extract (Figure 5B, lanes 3-6, 8, 13, 16).By visual inspection, the adult shifts correspond to the same sequences that shifted with the embryonic extract, indicating that the protein may be required at both the embryonic and adult stages of muscle development.

To confirm that the shifts we observed were specific and not a result of non-specific protein binding to the probe DNA, I added competitors to the binding reactions and ran them on an EMSA. Due to the large set of probes that shifted in Figure 5, we decided to run one probe from each set of the probes that shifted at the same size. I selected 0-105-

01 for the shift seen just below MEF2 and 50-156-08 for the shift located father away from MEF2 to run on the EMSA. Once again, I used *Act57B* as a positive control. All competitors were not radioactively labeled and added at a 100 times greater concentration than the labeled probe. With the addition of a wildtype competitor, the intensity of the shift band was significantly reduced, indicative of the higher probability of the same protein binding the non-radioactive competitor over the radioactiveprobe (Figure 6, lanes 4, 5 and 8, 9). Since we did not know the exact binding sequence, we decided to make the mutant competitor one of the probes that presented a shift on the EMSA that was of a different size. Therefore, the mutant competitor for 0-105-01 was 50-156-08 and vice versa. Addition of the mutant competitor for both the 0-105-01 probe and the 50-156-08

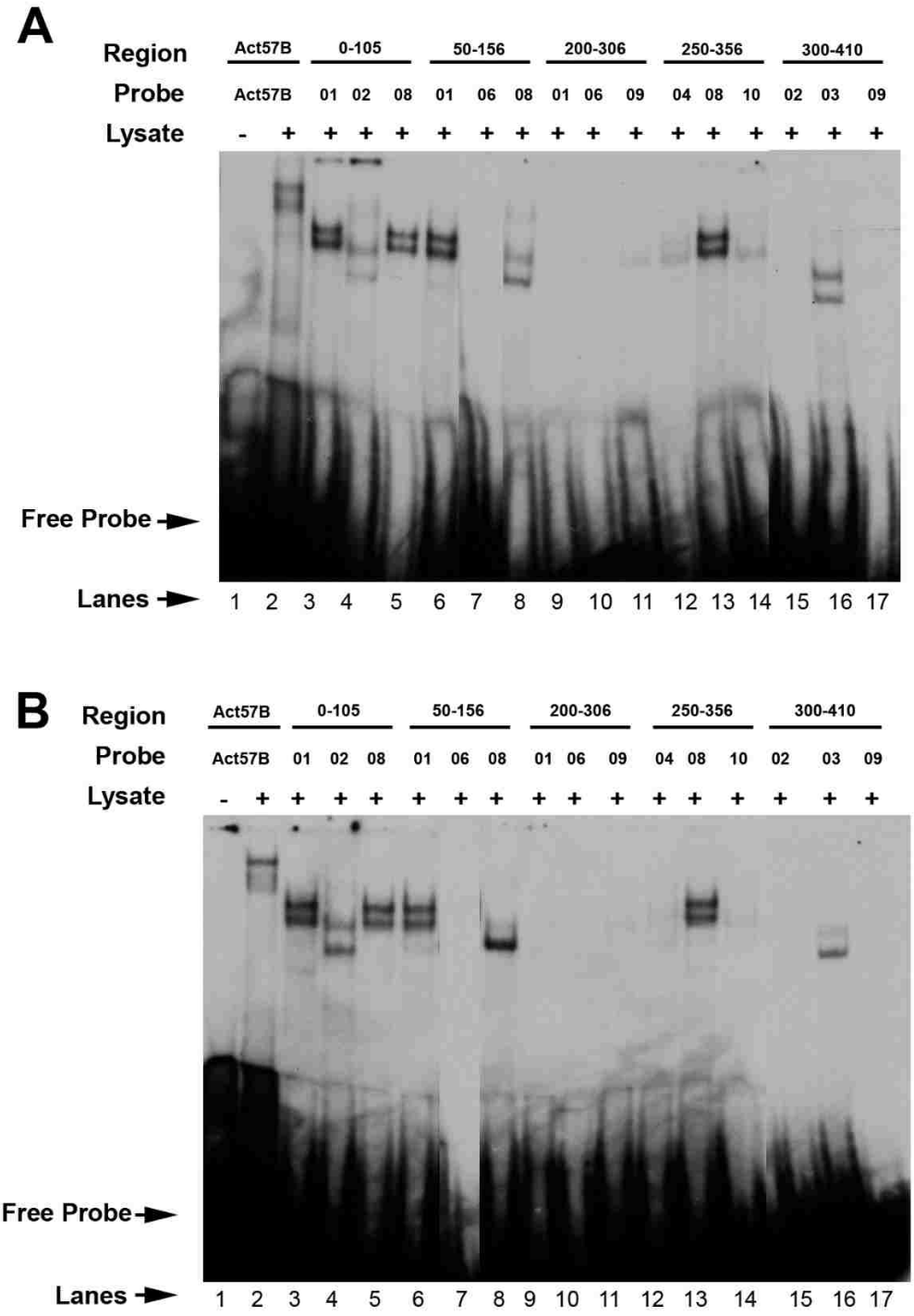


Figure 5: Probes predicted by the GA bind nuclear extractfrom *Drosophila* (A) EMSA using nuclear extract derived from embryos (B) EMSA using nuclear extract derived from pharate adults. Region refers to the subgroup the probes were predicted from and probe refers to the experiment ID from the string ID in Figure 4.

probes resulted in the reappearance of the shift band (Figure 6, lanes 6 and 10, respectively). From this data, we were able to conclude that the shifts seen in the EMSA are due to specific protein binding.

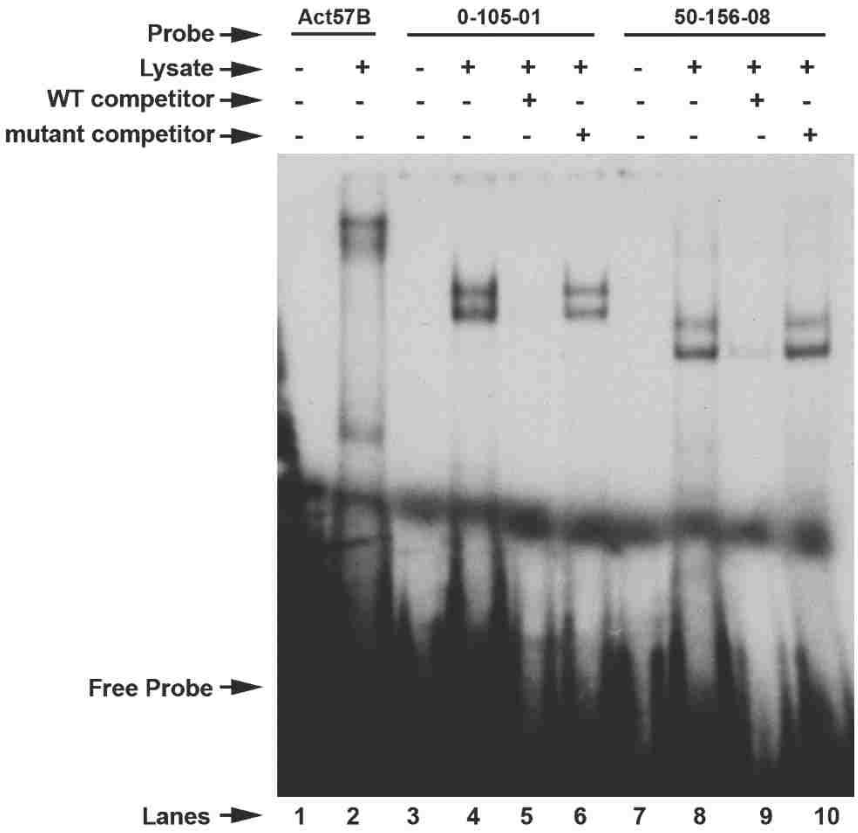


Figure 6: Nuclear extract from embryos specifically bind probe sequences. Binding of nuclear extract to probes 0-105-01 and 50-156-08 is specific due to the ability for each to out-compete itself with the WT competitor, lanes 5 and 9, respectively. Lanes 6 and 10 illustrate the reappearance the corresponding shifts, indicating specificity.*Act57B* is used as a positive control.

Conservation within sequences predicted by GA methods and fold change frequencies

Next, we wanted to determine if there any similarities in binding between the probes from the GA and the 6-mers that exhibited a higher frequency than expected from Figure 3. We took the top eight 6-mers, those sequences that had a fold change greater than 3.51, and made probes of 28 base pairs in length (see EMSA methods). The addition of each probe with embryonic lysate yielded shifts in some of the probes that differed from shifts seen in the GA EMSA (Figure 7). In lanes 5 and 10 of the EMSA, we observed stronger binding affinity whereas there were some weaker shifts seen in lane 9. The higher shift observed in lane 9 results from a MEF2 sequence in one of the flanking regions of the probe (Figure 7, denoted by asterisk). Overall, these results indicated that while the

different approaches identified different binding sequences, they each identified sequences that are recognized by proteins in Drosophila nuclear extracts.

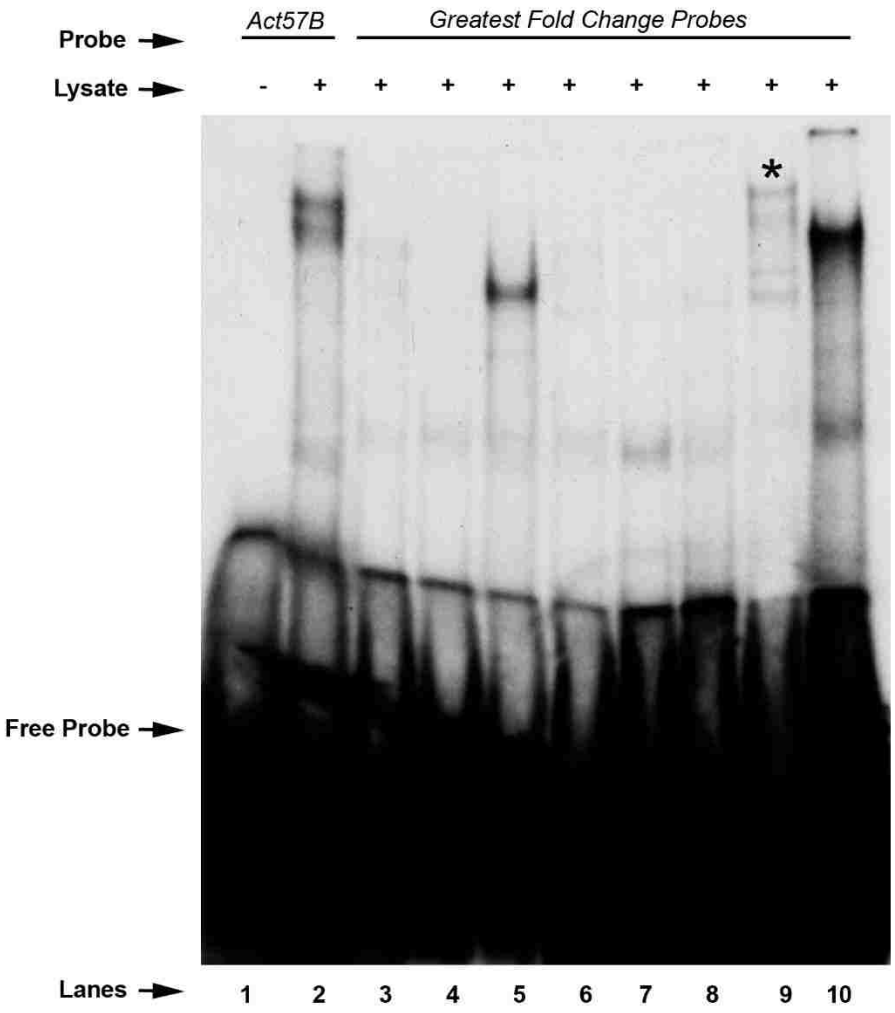


Figure 7: Embryonic nuclear extract is able to bind probes generated from the greatest fold change that are different from those predicted by the GA. Lanes 5, 9, and 10 show evidence of a shift differing from those seen in the GA. (*) denotes MEF2 binding a flanking region of the probe.

DISCUSSION

The best technique for predicting transcription factor binding sites is to use multiple methods

Finding potential transcription factor binding sites is a difficult task. From our results, we observed that the best way to find likely candidates is to use a variety of computational methods in order to get the most diverse group of sequences and validate them through wet laboratory methods (Figures 5, 6, 7).

The benefit of using the genetic algorithm was in that it was able to predict binding sites more regularly between probes, which is convenient for predicting which parts of the probes are most conserved. The shifts were similar for probes 0-105-01, 0-105-08, 50-156-01, and 250-356-08 for higher shift, and 0-105-02, 50-156-08, and 300-410-03 were the same shift size for the lower shift (Figure 5). Therefore it has an advantage for finding a consensus sequence.

However, the benefit of making probes based upon the expected frequencies of 6-mers seen in the data set, was in the ability to bind nuclear extract that was different from that seen in the GA (Figure 7). The sequences of the probes that bound protein are wildly different in nucleotide composition than those predicted that the GA, although these results should be used as a stepping stone to drive further research rather than taking them as definitive binding sequences (Lehman and Stanley, 2011). I noticed the sequences that were captured by the fold-change increase data were able to catch the subtleties in the GC composition that lacked in the more AT rich sequences predicted by the GA. Due to this difference, it would seem as though the best method for the prediction of binding sites computationally, may be to use a genetic algorithm paired with a most-frequent k-mer approach or multiple computational methods.

Additionally, it was critical to test these predicted sequences through traditional molecular methods such as the EMSA experiments shown in this paper. Not only was the

EMSA able to validate that protein was able to bind the predicted sequences, but it was also able to show that there were also predicted sequences that evaluated as false positives. This illustrates that my computational methods alone are not enough to verify a true binding site. Computational hypotheses need to be validated experimentally. The collection of these results show that the most accurate approach to locating transcription factor binding sites would be to use a collection of computational methods in conjunction with wet laboratory experiments.

Entropy as a fitness measure for predicting PWMs for sequence conservation

The prediction of PWMs has been a critical area of research in determining *de novo* conserved transcription factor binding site sequences. Although several algorithms currently exist that attempt to optimize the PWM, many of them require complex fitness functions with probability estimations that may be difficult to derive or are challenging to implement.

In this paper, we illustrate that by simply using Shannon entropy as a measure of fitness in determining a PWM, the GA can output a PWM with that has a lower entropy compared to background sequence indicating there are regions of regularity in the data set.

Although we have shown that the sum entropy within a 10 nucleotide window in lower than background sequence, we still need to continue to optimize on the sensitivity of the GA. Since the Drosophila genome is approximately 60% adenines and thymines, it puts the GA at a disadvantage for identifying regions that are guanine and cytosine rich.

As we have shown above, assuming nucleotide independence, we observed there were 6-mers that were expressed as high as a four-fold increase above expected. However, since these changes were changes that were expected to differ in frequency from 9-30 appearances of the 6-mers to 38-116 actual observed sequences, they tend to become masked by the fact that the 6-mers with only adenines and thymines are seen on an upwards of 359 times (Figure 2B, 3B). This may be due to the slight bias in the *Drosophila* genome which has an overall AT percentage of 60%.This makes the algorithm less sensitive to picking out these fold-change increases since the basic principle behind a PWM relies on frequency of a nucleotide at a given position. To account for this bias, a bootstrap method may be implemented to randomly pick out a subset of samples from each subgroup rather than using the entire population from each subgroup. This method may reduce the some of the AT rich samples in the GA to determine if the GA can predict sequences with higher GC content.

Discovery of new transcriptional co-factors that may work with MEF2

Transcriptional regulation is a complex process and more often than not, a single transcription factor is not enough to actively recruit RNA polymerase to transcribe a gene, thereby actively turning on gene expression. As mentioned previously, MEF2 cooperatively works with Twist to initiate muscle differentiation ranging from the embryonic stages to the adult stages of muscle development in *Drosophila*. For example, cell culture of the *sing* enhancer with an added MEF2 activator is not enough to activate *sing* expression alone, therefore, we know other co-factors and proteins are required for transcriptional activation (data not shown).

We have presented evidence that there are other regions near known MEF2 binding sites that exhibit protein binding which may interact with MEF2 during transcriptional regulation. By using MEF2-ChIP data by Sandmann et al., 2006, we illustrate that a genetic algorithm has the capability to predict regions of regularity that bind to both embryonic and adult nuclear extract as verified through an EMSA. We also show that using a frequency probability approach yields other regions in the data set that are also able to bind nuclear extract that differ from those sequences predicted by the GA.

In order to confirm that these binding regions can interact with MEF2, the exact sequence must be identified in which the protein is binding to in the EMSA as well as which protein is binding the DNA sequence. This will require a co-immunoprecipitation assay, where the EMSA probe is used as bait to bind nuclear extract and elution of the protein to be submitted for mass spectrometry analysis. Furthermore, upon determination of the consensus sequence, the sequence can be used to find this pattern in other enhancer regions where MEF2 is known to bind and be mutated to determine if activation of the enhancer can still persist without the wildtype site intact.

Additionally, the genetic algorithm can be optimized for balancing out adenine and thymine biases in the MEF2 ChIP data set. By using a bootstrap method of selecting a subset of 15-mers from each subgroup as an initial population for the GA, this may result in less over-shadowing of the guanine and cytosine content and a more accurate representation of potential binding sites.

CONCLUSION

The development of muscle through the various stages in the *Drosophila* life cycle requires many different transcription factors and regulatory factors. These factors are required in order to regulate transcription and therefore gene expression at the proper times. The mis-regulation of a muscle gene can cause detrimental effects such as cardiac myopathies and genetically inherited diseases such as muscular dystrophy. Therefore, it is critical to research and understand how muscle formation is regulated in order to better understand disease-causing mutations.

In this manuscript, I have illustrated the importance of understanding the transcriptional regulation of muscle development by using various methods of analysis in order to identify transcription factors that are involved in this process. One method of analysis is through the identification of a phenotype. In chapter one, I illustrate the significance of two Hox co-factors, *extradenticle* and *homothorax* primarily through the characterization of a phenotype. Through the use of RNAi and histochemical staining, I showed that the knockdown of *homothorax* in adult flies resulted in a change in muscle identity from a jump muscle to that of an indirect flight muscle. When both *homothorax* and *extradenticle* were over-expressed however, the indirect flight muscle had an identity more similar to that of the jump muscle. These results indicated that *homothorax* and *extradenticle* have important roles in muscle identity.

In chapter two, I take a mechanistic approach to show that the transcription factor, MEF2, is a direct regulator of the myoblast fusion gene, *singles bar* (*sing*) and that both have an important role in adult myoblast fusion. To illustrate this, I identified a 315 base pair region upstream of *sing* that contained two MEF2 binding sites. *In vitro* experiments

using an electrophoretic mobility shift assay in addition to *in vivo* experiments indicated this region is the enhancer for *sing* expression and that MEF2 is mechanistically upstream of *sing*. Additionally, I defined a requirement for *sing* during adult myoblast fusion. This was shown by knock downs of *sing* expression in adults over a time course prior to adult myoblast fusion through post myoblast fusion in which *sing* knockdowns show lack of proper the proper muscle skeleton due to lack of fusion of actin.

Lastly, in chapter three, I take a computational approach by using a genetic algorithm (GA) to search for regions in the genome that may have possible transcription factor binding sites. Through the use of a previously published MEF2 ChIP data set, I developed a genetic algorithm using entropy as a measure of fitness to find potential candidate binding sites in the data set. In order to validate the GA, I used an electrophoretic mobility shift assay (EMSA) to determine if nuclear extract from *Drosophila* was able to bind any of the potential candidate probes output from the GA. Although several of the probes were able to bind nuclear extract, there were also probes that returned false positives with no protein shifts seen in the EMSA, illustrating the importance of validating computational hypotheses with molecular techniques.

Together, these results illustrate some of the transcription factors that regulate muscle development in *Drosophila* as well as their involvement in potential mechanisms and conserved processes such as myoblast fusion. Additionally, I demonstrate various methods that can be used to elucidate the genes involved in transcriptional regulation of muscles.

REFERENCES

Abmayr, S., Pavlath, G., 2012. Myoblast fusion: lessons from flies and mice. Development 139, 641–656.

Abu-Shaar, M., Ryoo, H.D., Mann, R., 1999. Control of the nuclear localization of Extradenticle by competing nuclear import and export signals. Genes and Development 13, 935-945.

Ackrell, BAC., Johnson, MK., Gunsalus, RP., Cecchini, G., 1992. Structure and function of succinate dehydrogenase and fumarate reductase. In Muller F (ed) Chemistry and Biochemistry of Flavoproteins, Boca Raton, FL: CRC Press, vol 3, 229-297.

Adams, M., et al., 2000.The genome sequence of Drosophila melanogaster. Science 287 (5461), 2185-2195.

Anant, S., Roy, S., VijayRaghavan, K., 1998. Twist and Notch negatively regulate adult muscle differentiation in Drosophila. Dev. Biol. 125 (8), 1361–1369.

Andres, V., Cervera, M., Mahdavi, V., 1995. Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints. J. Biol. Chem. 270 (40), 23246–23249.

Apitz, H., Strunkelnberg, M., de Couet, H.G., Fischbach, K.-F., 2005. Single-minded, Dmef2, Pointed, and Su(H) act on identified regulatory sequences of the roughest gene in Drosophila melanogaster. Dev. Genes Evol 215, 460–469.

Artero, R., Castanon, I., Baylies, M., 2001. The immunoglobulin-like protein Hibris functions as a dose-dependent regulator of myoblast fusion and is differentially controlled by Ras and Notch signaling. Development 128, 4251-4264.

Atreya, K., Fernandes, J., 2008. Founder cells regulate fiber number but not fiber

formation during adult myogenesis in Drosophila. Dev. Biol. 321 (1), 123–140.

Azpiazu, N., and Frasch, M., 1993. Tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm. Genes and Development 7 (7B), 1325-1340.

Baylies, M., and Bate, M. 1996. twist: a myogenic switch in Drosophila. Science 272, 1481-1484.

Beasley, D., Bull, D., Martin, R., 1993. An overview of genetic algorithms: part 1, fundamentals. University Computing 15 (2), 58-69.

Ben-Gal, I., 2008.Bayesian Networks. Encyclopedia of Statistics in Quality and Reliability. I.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F., Zappavigna, V., 1999. The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. Genes and Dev. 13, 946-953.

Black, B., and Cripps, R., 2010. Myocyte enhancer factor 2 transcription factors in heart development and disease. In Heart Development and Regeneration. Oxford Academic Press, 673-699.

Bloomington Drosophila Stock Center, 2007.Basic methods of culturing Drosophila. Retrieved from http://flystocks.bio.indiana.edu/Fly_Work/culturing.htm

Bour, B., Chakravarti, M., West, J., Abmayr, S., 2000. Drosophila SNS, a member of the immunoglobulin superfamily that is essential for myoblast fusion. Genes Dev. 14, 1498–1511.

Bour, B.A., O'Brien, M.A., Lockwood, W.L., Goldstein, E.S., Bodmer, R., Taghert, P.H.,

Abmayr, S.M., Nguyen, H.T., 1995. Drosophila MEF2, a transcription factor that is essential for myogenesis. Genes Dev 9, 730–741.

Braun, T., and Gautel, M., 2011. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. Nature Reviews Molecular Cell Biology 12, 349-361.

Bryantsev, A., Baker, P., Lovato, T., Jaramillo, M., Cripps, R.M., 2012. Differential requirements for Myocyte Enhancer Factor-2 during adult myogenesis in Drosophila. Dev. Biol. 362 (2), 191–207.

Bryantsev, A., Duong, S., Brunetti, T.M., Chechenova, M.B., Lovato, T.L., Nelson, C., Shaw, E., Uhl J.D., Gebelein, B., Cripps, R.M., 2012. Extradenticle and homothorax control muscle fiber identity in Drosophila. Dev. Cell. 23 (3), 664-673.

Bulyk, M., Johnson, P., Church, G., 2002. Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors. Nucleic Acids Research. 30 (5), 1255-1261.

Burglin, T., 1997. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TFIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. 25, 4173-4180/

Chen, E., Olson, E., 2001. Antisocial, an intracellular adaptor protein, is required for myoblast fusion in Drosophila. Dev Cell 1 (5), 705-715.

Chen, E., Olson, E., 2004. Towards a molecular pathway for myoblast fusion is Drosophila. Trends Cell Biol. 14 (8), 452–460.

Compeau, P., and Pevzner, P., 2014. Bioinformatics Algorithms: an active approach,

Active Learning Publishers, La Jolla, CA.

Courchesne-Smith, C., and Tobin S., 1989. Tissue-specific expression of the 79B actin gene during Drosophila development. Developmental Biology, 133 (2), 313-321.

Cripps, R., Black, B., Zhao., B., Lien, C.L., Schulz, R., Olson, E., 1998. The regulatory geneMef2 is a direct target for transcriptional activation by Twist during Drosophila myogenesis. Genes Dev. 12(3), 422-44.

Datta, S., 2006. Validation measures for clustering algorithms incorporating biological information. Computer and Computational Sciences, 2006. IMSCCS '06 First International Multi-Symposiums on June 20-24 1, 131-135.

Deak, II. A histochemical study of muscles of Drosophila melanogaster. J. Morphol.1977;153:307–316.

Duan, H., Skeath, J., Nguyen, H., 2001. Drosophila Lame duck, a novel member of the Gli superfamily, acts as a key regulator of myogenesis by controlling fusion-competent myoblast development. Development 128, 4487-4500.

Durbin, R., Eddy, S., Krogh, A., Mitchison, G., 1998. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge, UK: Cambridge University Press

Dutta, D., Anant, S., Ruiz-Gomez, M., Bate, M., 2004. Founder myoblasts and fibre number during adult myogenesis in Drosophila. Development 131 (15), 3761–3772.

Dutta, D., and VijayRaghavan, K., 2006. Metamorphosis and the formation of the adult musculature. In H. Sink, Muscle Development in Drosophila (125-142). Texas: Landes Bioscience.

Dworak, H., Charles, M., Pellerano, L., Sink, H., 2001. Characterization of Drosophila hibris, a gene related to human nephrin. Development 128, 4265-4276.

Eissing, T., et al., 2011. A computational systems biology software platform for multiscale modeling and simulation: integrating whole-body physiology, disease biology, and molecular reaction networks. Front. Physiol. 2 (4), 1-10.

Erickson, M., Galletta, B., Abmayr, S., 1997. Drosophila myoblast city encodes a conserved protein that is essential for myoblast fusion, dorsal closure, and cytoskeletal organization. J. Cell Biol. 138 (3), 589–603.

Estrada, B., Maeland, A., Gisselbrecht, S., Bloor, J., Brown, N., Michelson, A., 2007. The MARVEL domain protein, Singles Bar, is required for progression past the pre-fusion complex stage of myoblast fusion. Dev. Biol. 307 (2), 328–339.

Fazius, E., Shelest, V., Shelest, E., 2011.SiTaR: a novel tool for transcription factor binding site prediction. Bioinformatics 27 (20), 2806-2811.

Fernandes, J., Bate, M., VijayRaghavan, K., 1991. Development of the indirect flight muscles of Drosophila. Development 113, 67–77.

Fisher, K., Jooss, K., Alston, J., Yang, Y., Haecker, S., High, K., Pathak, R., Raper, S., Wilson, J., 1997. Recominant adeno-associated virus for muscle directed gene therapy. Nature Medicine 3, 306-312.

Friedman, N., Linial, M., Nachman, I., Pe'er, D., 2004. Using Bayesian networks to analyze expression data. Journal of Computational Biology 7 (3-4), 601-620.

Gildor, B., Schejter, E.D., Shilo, B.Z., 2012. Bidirectional Notch represses fusion competence in swarming adult Drosophila myoblasts. Development 139 (21), 4040–4050.

Gisselbrecht, S., et al., 2013. Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos. Nature Methods 10, 774-780.

Gossett, L.A., Kelvin, D.J., Sternberg, E.A., Olson, E.N., 1989. A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. Mol. Cell. Biol. 9, 5022–5033.

Grand, F., and Rudnicki, M., 2007.Skeletal muscle satellite cells and adult myogenesis. Current Opinion in Cell Biology, 19 (6), 628-633.

Greising, S., Gransee, H., Mantilla, C., Sieck, G., 2012. Systems biology of skeletal muscles: fiber type as an organizing principle. WIREs Sys Biol Med, 4, 457-473.

Grifone, R., Laclef, C., Spitz, F., Lopez, S., Demignon, J., Guidotti, J.E., Kawakami, K., Xu, P.X., Kelly, R., Petrof, B.J., Daegelen, D., Concordet, J.P., Maire, P., Six1 and Eya1 expression can reprogram adult muscle from slow-twitch phenotype into the fast-twitch phenotype. Mol. Cell. Biol., 24 (14), 6253-6267.

Gussoni, E., Soneoka, Y., Strickland, C., Buzney, E., Khan, M., Flint, A., Kunkel, L., Mulligan, R., 1999.Dystrophin expression in the mdx mouse restored by stem cell transplantation. Nature 401, 390-394.

Hastings, J.R.B., and Kirby, K.S., 1965. The nucleic acids of Drosophila melanogaster. Biochem. J. 100, 532-539.

Hasty, J., McMillen, D., Isaacs, F., Collins, J., 2001. Computational studies of gene regulatory networks: in numero molecular biology. Nature Reviews Genetics 2, 268-279.

Hebrok, M., Wertz, K., Fuchtbauer, E.M., 1994. M-twist is an inhibitor of muscle differentiation. Developmental Biology, 165, 537-544.

Heidt, A., Rojas, A., Harris, I., Black, B., 2007. Determinants of myogenic specificity withinMyoD are required for noncanonical E box binding. Mol. Cell. Biol. 27 (16), 5910-5920.

Herranz, R., Díaz-Castillo, C., Nguyen, T., Lovato, T., Cripps, R., Marco, R., 2004. Expression patterns of the whole troponin C gene repertoire during Drosophila development. Gene Expression Patterns, 4 (2), 183-190.

Hiromi, Y., and Hotta, Y., 1985. Actin gene mutations in Drosophila; heat shock activation of the indirect flight muscles. EMBO, 4 (7), 1681-1687.

Hoffman, E., Brown Jr, R., Kunkel, L., 1987. Dystrophin: The protein product of the duchenne muscular dystrophy locus. Cell, 51 (6), 919-928.

Horton, R.M., 1993. In vitro recombination and mutagenesis of DNA. In: White, B.A. (Ed.), In: PCR Protocols: Current Methods and Applications, 15. Humana Press, Totowa, NJ.

Jaw, T., You, L., Knoepfler, P., Yao, L, Pai, C., Tang, C., Chang, L., Berthelsen, J., Blasi, F., Kamps, M., Sun, H., 1999. Direct interaction of two homeoproteins, Homothorax and Extradenticle, is essential for EXD nuclear localization and function. Mechanisms of Development 91 (2000), 279-291.

Jiang, Y., Shi, H., Liu, J., 2009. Two Hox co-factos, the Mesi/Hth homolog UNC-62 and thePbx/Exd homolog CEH-20, function together during C. elegans postembryonic mesodermal developemt. Developmental Biology, 335 (2), 535-546.

Josephson, R.K., Malamud, J.G., Stokes, D.R., 2000. Asynchronous muscle: a primer. The Journal of Experimental Biology, 203, 2713-2722.

Kelly, K., Meadows, S., Cripps, R.M., 2002. Drosophila MEF2 is a direct regulator of

Actin57B transcription in cardiac, skeletal, and visceral muscle lineages. Mech. Dev. 110 (1-2), 39–50.

Kennedy Krieger Institute. 2012. Center for Genetic Muscle Disorders. Retrieved from http://www.kennedykrieger.org/patient-care/patient-care-centers/center-genetic-muscle-disorders

Kitano, H., 2002. Computational systems biology. Nature 420, 206-210.

Korhonen, J., Martinmäki, P., Pizzi, C., Rastas, P., Ukkonen, E., 2009. MOODS: fast search for position weight matrix matches in DNA sequences. Bioinformatics 25 (23), 3181-3182.

Larsson, B., Sjodin, B., Karlsson, J., 1978. Histochemical and biochemical changes in human skeletal muscle with age in sedentary males, age 22-65 years. Acta. Physiol. Scand. 103, 31-39.

Lau, M.K., 2013. DTK: Dunnett–Tukey–Kramer Pairwise Multiple Comparison Test adjusted for unequal variances and unequal sample sizes. R Package Version 3, 5.

Lee, J., Qu-Petersen, Z., Cao, B., Kimura, S., Jankowski, R., Cummins, J., Usas, A., Gates, C., Robbins, P., Wernig, A., Huard, J., 2000. Clonal isolation of muscle-derived cells capable of enhancing muscle regeneration and bone healing. The Journal of Cell Biology 150 (5), 1085-1099.

Lehman, J., and Stanley, K., 2011. Abandoning objectives: Evolution through the search for novelty alone. Journal of Evolutionary Computation 19 (2), 189-223.

Levistsky, S., and Kasif, S., 2003. Predicting protein function from protein/protein interaction data: a probabilistic approach. Bioinformatics 19, i197-i204.

Levitsky, V., Ignatieva, E., Ananko, E., Turnaev, I., Merkulova, T., Kolchanov, N.,

Hodgman, TC., 2007. Effective transcription factor binding site prediction using a combination of optimization, a genetic algorithm and discriminant analysis to capture distant interactions. BMC Bioinformatics 8 (481), 1-20.

Li, L., Liang, Y., Bass, R., 2007. GAPWM: a genetic algorithm method for optimizing a position weight matrix. Bioinformatics 23 (10), 1188-1194.

Liu, J., Wang, Y., Birnbaum M., Stoffers, D., 2010. Three-amino-acid-loop-extension homeodomain factor Meis3 regulates cell survival via PDK1. Proc Natl AcadSci, 107 (47), 20494-20499.

Lilly, B., Galewsky, S., Firulli, A., Schulz, R., Olson, E., 1994. D-MEF2: A MADS box transcription factor expressed in differentiating mesoderm and muscle cell lineages during Drosophila embryogenesis. Proc. Natl. Acad. Sci. 91, 5662–5666.

Lilly, B., Zhao, B., Ranganayakulu, G., Paterson, B., Schulz, R., Olson, E.N., 1995. Requirement of MADS domain transcription factor D-MEF2 for muscle formation in Drosophila. Science 267, 688–693.

Mathelier, A., and Wasserman, W., 2013. The next generation of transcription factor binding site prediction. PLoSCompuational Biology 9.

Martelotto, L., Ng, C., Fe Filippo, M., Zhang, Y, Piscuoglio, S., Lim, R., Shen, R., Norton, L., Reis-Filho, J., Weigelt, B., 2014. Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations. Genome Biology 15 (10), 484-504.

Massarwa, R., Carmon, S., Shilo, B.Z., Schejter, E.D., 2007. WIP/WASp-based actin polymerization machinery is essential for myoblast fusion in Drosophila. Dev. Cell. 12, 557–569.

Maves, L., Tyler, A., Moens, C., Tapscott, S., 2009.Pbx acts with Hand2 in early myocardial differentiation. Developmental Biology 333 (2), 409-418.

Maves, L., Waskiewicz, A., Paul, B., Cao, Y., Tyler, A., Moens, C., Tapscott, S., 2007. Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. Development 134, 3371-3382.

Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martínez, C., Torres, M., 1999.Conserved regulation of proimodistal limb axis development by Meis1/Hth. Nature 402 (6760), 425-429.

Mery, A., Taghli-Lamallem, O., Clark, K., Beckerle, M., Wu, X., Ocorr, K., Bodmer, R., 2007. The Drosophila muscle LIM protein, Mlp84B, is essential for cardiac function. The Journal of Experimental Biology, 211, 15-23.

Mitchell, M., 2009. Complexity: a guided tour. Oxford University Press, New York.

modMine: flexible access to modENCODE data. 2012. Nucleic Acids Res 40 (D1), D1082-D1088.

Morriss, G.R., Bryantsev, A.L., Chechenova, M., LaBeau, E.M., Lovato, T.L., Ryan, K.M., Cripps, R.M., 2011. Analysis of skeletal muscle development in Drosophila. In: DiMario, J. (Ed.), Myogenesis: Methods and Protocols. Springer, New York.

Mukherjee, P., Gildor, B., Shilo, B.Z., VijayRaghavan, K., Schejter, E., 2011. The actin nucleatorWASp is required for myoblast fusion during adult Drosophila myogenesis. Development 138 (11), 2347–2357.

Nilwik, R., Snijders, T., Leenders, M., Groen, B., van Kranenburg, J., Verdijk, L., von

Loon, L., 2013. The decline in skeletal muscle mass with aging is mainly attributed to a reduction in type II muscle fiber size. Experimental Gerontology 48 (5), 492-498.

Nose, A. Isshiki, Takako, Takeichi, Masatoshi, 1998. Regional specification of muscle progenitors in Drosophila: the role of the mshHomeobox gene. Development 125, 215–223.

Nowak, K., Wattanasirichiaigoon, D., Goebel, H., Wilce, M., Pelin, K., Donner, K., Jacob, R., Hübner, C., Oexle, K., Anderson, J., Verity, C., North, K., Iannaccone, S., Müller, C., Nürnberg, P., Muntoni, F., Sewry, C., Hughes, I., Sutphen, R., Lacson, A., Swoboda, K., Vigneron, J., Wallgren-Pettersson, C., Beggs, A., Laing, N., 1999. Mutations in the skeletal muscle α-actin gene in patients with actin myopathy and nemaline myopathy. Nature Genetics, 23, 208-212.

Paululat, A., Holz, A., Renkawitz-Pohl, R., 1999. Essential genes for myoblast fusion in Drosophila embryogenesis. Mech. Dev. 83 (1-2), 17–26.

Pazin, M., 2007. Preparation of nuclear extracts from Drosophila embryos. Cold Spring Harbor Protocols

Peckham, M., Molloy, J.E., White, D.C., 1990. Physiological properties of the dorsal longitudinal flight muscle and the tergal depressor of the trochanter muscle of Drosophila melanogaster. Journal of Muscle Research and Cell Motility 11 (3), 203-215.

Potthoff, M., Olson, E., 2007. MEF2: a central regulator of diverse developmental programs. Development 134 (23), 4131–4140.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R

Foundation for Statistical Computing, Vienna, Austria, URL.

Ranganayakulu, G., Elliott, D.A., Harvey, R.P., Olson, E.N., 1998. Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice.Development 125, 3037–3048.

Ranganayakulu, G., Zhao, B., Dokidis, A., Molkentin, J., Olson E., Schulz, R., 1995. A series of mutations in the D-MEF2 transcription factor reveal multiple function in larval and adult myogenesis in Drosophila. Developmental Biology 171 (1), 169-181.

Rasband, W.S., ImageJ, U.S., 2014.National Institutes of Health, Bethesda. Maryland, USA 1997.

Rawls, A., and Olson, E., 1997. MyoD meets its maker. Cell 89, 5-8.

Reiter, L., Potocki, L., Chien, S., Gribskov, M., Bier, E., 2001. A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. Genome Res. 11 (6), 1114-1125.

Richardson, B., Nowak, S., Baylies, M., 2008. Myoblast Fusion in Fly and Vertebrates: New Genes, New Processes and New Perspectives. Traffic 9 (7), 1050–1059.

Riechmann, V., Irion, U., Wilson, R., Grosskortenhaus, R., Leptin, M., 1997. Control of cellfaste and segmentation in the Drosophila mesoderm. Development 124, 2915-2922.

Roberts, R., and Bobrow, M., 1998.Dystrophins in vertebrates and invertebrates. Human Molecular Genetics 7 (4), 589-595,

Rochlin, K., Shannon, Y., Roy, S., Baylies, M., 2009. Myoblast fusion: where it takes more to make one. Dev. Biol. 341 (1), 66–83.

Rosenberg-Hasson, Y., Renert-Pasca, M., Volk, T., 1996. A Drosophila Dystrophin-related protein, MSP-300, is required for embryonic muscle morphogenesis. Mechanisms of Development, 65, 83-94.

Rubin, G.M., Spradling, A.C., 1982. Genetic transformation of Drosophila with transposable element vectors. Science 218, 348–353.

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

Ruiz-Gomez, M., Coutts, N., Price, A., Taylor, M., Bate, M., 2000. Drosophila dumbfounded: a myoblast attractant essential for fusion. Cell 102, 189–198.

Rustin, P., Munnich, A., and Rötig, A., 2002. Succinate dehydrogenase and human diseases: new insights into a well-known enzyme. European Journal of Human Genetics 10 (5), 289-291.

Sánchez-Pulido, L., Martín-Belmonte, F., Valencia, A., Alonso, M.A., 2002. MARVEL: a conserved domain involved in membrane apposition events. Trends Biochem. Sci. 27 (12), 599–601.

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

Sandmann, T., Jensen, L., Jakobsen, J., Karzynski, M., Eichenlaub, M., Bork, P., Furlong, E., 2006. A temporal map of transcription factor activity: Mef2 directly regulates target genes at all stages of muscle development. Dev. Cell. 10 (6), 797–807.

Schafer, G., Weber, S., Holz, A., Bogdan, S., Schumacher, S., Muller, A., Renkawitz-Pohl, R., Onel, S.F., 2007. The Wiskott-Aldrich syndrome protein (WASP) is essential for myoblast fusion in Drosophila. Dev. Biol. 304, 664–674.

Schiaffino, S., and Reggiani, C, 2011. Fiber types in mammalian skeletal muscles. Physiological Reviews 91 (4), 1447-1531.

Schreiber, E., Matthias, P., Müller, M., Schaffner, W., 1989. Rapid detection of octamer binding proteins with mini-extracts prepared from a small number of cells. Nucleic Acids Research 17 (15), 6419.

Schroter, R., Buttgereit, D., Beck, L., Holz, A., Renkawitz-Pohl, R., 2006. Blown fuse regulates stretching and outgrowth but not myoblast fusion of the circular visceral muscles in Drosophila. Differentiation 74 (9-10), 608–621.

Schroter, R., Lier, S., Holz, A., Bogdan, S., Klambt, C., Beck, L., Renkawitz-Pohl, R., 2004. Kette and blown fuse interact genetically during the second fusion step of myogenesis in Drosophila. Development 131, 4501–4509.

Shcherbata, H., Yatsenko, A., Patterson, L., Sood, V., Nudel, U., Yaffe, D., Baker, D., Ruohola-Baker, H., 2007.Dissecting muscle and neuronal disorders in a Drosophila model of muscular dystrophy. The EMBO Journal 26, 481-493.

Siddharthan, R., 2010. Dinucleotide weight matrices for predicting trnascroption factor binding sites: generalizing the position weight matrix. PLOS One 5(3)

Smerdu, V., Karsch-Mizrachi, I., Campione, M., Leinwand, L., Schiaffino, S., 1994. Type IIx myosin heavy chain transcripts are expressed in type IIb fibers of human skeletal muscle. American Journal of Physiology-Cell Physiology 267 (6), C1723-C1728.

Sohn, R., Huang, P., Kawahara, G., Mitchell, M., Guyon, J., Kalluri, R., Kunkel, L., Gussoni, E., 2009. A role for nephrin, a renal protein, in vertebrate skeletal muscle cell fusion.Proc. Natl. Acad. Sci 106, 9274-9279.

Soler, C., Han, J., Taylor, M., 2012. The conserved transcription factor Mef2 has multiple roles in adult Drosophila musculature formation. Development 139 (7), 1270–1275.

Srinivas, B., Woo, J., Leong, W., Roy, S., 2007. A conserved molecular pathway mediates myoblast fusion in insects and vertebrates. Nature Genetics 39 (6), 781-786.

St. Pierre, S.E., Ponting, L., Stefancsik, R., McQuilton, P., the FlyBase Consortium, 2014. FlyBase 102 advanced approaches to interrogating FlyBase. Nucleic Acids Res. 42 (D1), D780 D788. http://dx.doi.org/10.1093/nar/gkt1092 [FBrf0223749].

Staehling-Hampton, K., Hoffmann, M., Baylies, M., Rushtont, E., Bate, M., 1994. dpp induces mesodermal gene expression in Drosophila. Nature 372, 783-786.

Stewart, A., Hannenhalli, S., Plotkin, J., 2012. Why transcription factor binding sites are ten nucleotides long. Genetics 192, 973-985.

Strünkelnberg, M., Bonengel, B., Moda, L.M., Hertenstein, A., de Couet, H.G., Ramos, R.G., Fischbach, K.F., 2001. rst and its paralogue kirre act redundantly during embryonic muscle development in Drosophila. Development 128, 4229–4239.

Stute, C., Kesper, D., Holz, A., Buttgerei, D., Renkawitz-Pohl, R., 2006. Establishment of cell type specific Gal-4 driver lines for the mesoderm of Drosophila. D.I.S 89, 111–115.

Tapanes-Castillo, A., and Baylies, M., 2004. Notch signaling patterns Drosophila

mesodermal segments by regulating the bHLH transcription factor twist. Development 131, 2359-2372.

Taylor, M., 2002. Muscle differentiation: how two cells become one. Current Biology 12 (6), R224-R228.

Thisse, B., Stoetzel, C., Gorostiza-Thisse, C., Perrin-Schmitt, F., 1988.Seqeunce of the twist gene and nuclear localization of its protein in endomesodermal cells of early Drosophila embryos. EMBO J. 7, 2175-2183.

Vavouri, T., and Elgar, G., 2005. Predication of cis-regulatory elements using binding sitematrices-the successes, the failures and the reasons for both. Current Opinion in Genetics and Development 15, 395-402.

Verdijk, L., Koopman, R., Schaart, G., Meijer, K., Savelberg, H., van Loon, L., 2007. Satellite cell content is specifically reduced in type II skeletal muscle fibers in the elderly. Am. J. Physiol. Endocrinol. Metab.292, E151-E157.

Weiszmann, R., Hammonds, A., Celniker, S., 2009. Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. Nat. Prot. 4, 605–618.

Xenarios, I., Salwínski, L., Duan, X., Higney, P., Kim, S., Eisenberg, D., 2002. DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Research 30 (1), 303-305.

Zinzen, R., Girardot., C., Gagneur, J., Braun, M., Furlong, E., 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature,* 452, 65-70.