Bioinformatic and functional analyses of muscle cell diversification in Drosophila melanogaster

Preethi Poovathumkadavil

Year 2021

Doctoral School of Life, Health, Agronomy, Environmental Sciences (SVSAE)

University of Clermont Auvergne

Thesis

INSERM U1103, CNRS UMR6293, University of Clermont Auvergne

Genetics, Reproduction and Development institute (iGReD)

Preethi Poovathumkadavil

Bioinformatic and functional analyses of muscle cell diversification in *Drosophila melanogaster*

Specialty Bioinformatics and Biology - Health

Thesis publicly defended on April 23, 2021

Jury members

Dr. Athanassia Sotiropoulos	President of the jury
Dr. Michael V. Taylor	Reviewer
Dr. Maria Spletter	Reviewer
Dr. Laurence Dubois	Examiner
Dr. Maria Eugenia Gallego	Examiner
Dr. Krzysztof Jagla	PhD supervisor

A PhD brings out the best in you. It brings out the worst in you. It is so important to know both sides of oneself. Having been associated with the corporate IT world for a long time, this was another glimpse into the human psyche in an academic setting. This level of education is aptly titled ‘Doctorate of Philosophy’. I would like to thank Krzysztof Jagla, iGReD and the UCA for this opportunity. As someone who is interested in multiple domains, my PhD opened the door to the whole new world of biological experimentation. It reinforced my belief in the importance of cross domain interactions in any field. It again helped me recognize and appreciate the contribution of each domain, and how each of them is indispensable to a project, be it in IT or the Biological Sciences.

It was great to have my autonomy during these three years. I learnt Science as well as some more life lessons, which is what education is really about for me. I would also like to thank everyone who came into my life, even for a period as short as a few seconds, because every interaction taught me something.

“Anything worth doing in life has risks… In my case, I decided a long time ago that exploring the universe was worth taking a risk for.”

- Chris Hadfield, astronaut, former commander of the International Space Station.

The biggest blunder of our complex times might be our tendency to underestimate the power of simplicity.

- Personal thought.

Globalization is the new dynamite. It can give us people like Malala Yousafzai, Bill Gates and Greta Thunberg who are trying to find global solutions to global problems while staying true to their identities. It can also cause the global spread of COVID-19. I hope I learn from these exceptional people to think globally because a problem anywhere causes ripples everywhere. And because Science is for humanity.

- Personal thought.

ABSTRACT

The voluntary skeletal muscles in vertebrates are the main effectors of locomotion. Processes and genes implicated in human myogenesis are of immense interest to better understand the deregulations caused in muscular and neuromuscular disorders and to find therapeutic targets. The body wall or somatic muscles of the fruit fly, *Drosophila melanogaster*, are similar to vertebrate skeletal muscles. As is the case for vertebrate skeletal muscles, each *Drosophila* embryonic somatic muscle possesses its specific identity that clearly distinguishes from its immediate neighbors. In *Drosophila*, some muscle identity transcription factors (iTFs) have been identified, but others remain elusive.

In order to dissect mechanisms regulating the diversification of committed muscle cells to attain their final identity, the team had previously generated transcriptomics data for mRNA under translation in the Lms+ lateral transverse (LT) and Slou+ muscle subsets as well as the Duf+ global muscle set over three time windows of development. My analyses of this data helped identify the evolutionarily conserved gene that is part of the conserved Wnt enhanceosome, *Sequence-specific single-stranded DNA-binding protein* (*Ssdp*) as a determinant of final muscle identity. Its vertebrate homologue *Single stranded DNA binding protein 3* (*SSBP3*) is downregulated and mis-spliced in human myotonic dystrophies, but its role in myogenesis has not been studied. My study reveals a role for Ssdp in embryonic myogenesis for the first time. A temporally regulated, isoform-specific expression of *Ssdp* was identified. Further analyses showed that the initial muscle identity program proceeds normally for the most part in the absence of zygotic Ssdp, but muscles fail to establish their final identity due to the deregulation of iTFs and identity processes that establish muscle morphology, innervation and attachment. Comparative analyses revealed that specific *Ssdp* mutant phenotypes overlap subsets of phenotypes observed in the context of loss of function of a *Drosophila* Wnt, Wg and dTCF, an effector of the canonical Wnt pathway, suggesting specific interactions between these factors. Potential genetic interactions between the LT iTF, Ap (a Lhx2 orthologue) with Mid and the Ssdp partner, Chi (a LDB1 orthologue, also part of the Wnt enhanceosome) were unveiled.

In addition, my *in silico* analysis identified other potential candidates implicated in muscle identity such as the TFs D, Sox14 and Sox21b for LT muscles and Stat92E for Slou+ muscles, with Nf-YB acting as a potential upstream regulator. Muscle subset specific enrichments of CT-rich motifs in the LT subset and GATA motifs in the Slou subset were also identified.

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

In this first chapter, I present details to introduce the context of my project. I start off with a short evolutionary view of why muscles arose. I follow this with a detailed description of myogenesis in vertebrates and highlight the necessity for simpler model systems. Then, I introduce *Drosophila* as a model system and how understanding muscle development in simpler models would help us better understand pathological manifestations.

1.1. Staying grounded and standing up to gravity

Unlike plants that evolved roots to anchor them and thick stems to withstand the gravitational force, the animal kingdom adopted a different strategy. Eukaryotic protozoans, organisms that possess a nucleus enclosed within a nuclear envelope and are unicellular, developed simple locomotory mechanisms based on pseudopodia or cilia as seen in amoebae or paramecia respectively. These locomotory units are based on the actomyosin machinery (Mirvis, Stearns, et James Nelson 2018; Tekle et Williams 2016). During the Cambrian explosion some 540 million years ago, a combination of biotic factors like genetics (Holland 2015) and abiotic factors like oxygen levels (Sperling et al. 2013) led to the rapid evolution of immensely diverse forms of terrestrial multicellular metazoans characterized by muscular, nervous and digestive systems (Briggs 2015). Muscles consisting of a sophisticated contractile machinery around the actomyosin core originated in aquatic ecological systems and were key to seeding sustainable life on land, where it was essential to counter the forces of gravity pulling the large animal body mass down towards the earth. Unicellular organisms possess non-muscle myosin that performs contractile functions during processes such as cell motility and cytokinesis (Shutova et al. 2012). The first metazoans, the sponges can contract their bodies despite the lack of a muscular system. This has been attributed to the presence of proteins associated with the contractile apparatus such as non-muscle myosin and Myosin Heavy Chain type II (MHCII) that is characteristic of bilaterian (animals with bilateral or left-right symmetry such as humans) striated muscles. MHCII is localized at specific regions distinct from non-muscle myosin. These proteins, thus, appear to have functionally diversified before the emergence of muscles (Steinmetz et al. 2012). The sea anemone, *Nematostella vectensis* has a strong expression of

this protein in its fast-contracting smooth muscles lacking the characteristic striations generated by the contractile sarcomere units observed in human skeletal muscles. It was the appearance of limbs that permitted the transition from aquatic to terrestrial life. Studies of fossil records and the extant lobe-finned coelacanth fish, *Latimeria chalumnae* have shown that fins are the ancestors of tetrapod limbs (Miyake et al. 2016; Hirasawa et Kuratani 2018).

The evolution of the neuromuscular system provided the ability to perform voluntary movements to actively explore new ecological niches and spread out all over the earth. Insects are the most abundant species of the animal kingdom to have successfully evolved in all environmental niches. Apart from limb muscles, flying insects such as the fruit fly, *Drosophila* have special adaptations for flight in their flight muscles. We, the *Homo sapiens*, started exploring outer space that provided further insights into the specificity of and necessity for muscle evolution. Space missions have lent attention to so-called 'antigravity muscles', the postural muscles that hold us upright against gravity such as the soleus leg muscle (Sandonà et al. 2012; Wuehr et al. 2014) because they atrophy in the absence of gravity during extended lengths of time in space despite regular exercise (Fitts, Riley, et Widrick 2001).

The study of processes involved in muscle development is an active field of immense interest in the scientific community, not only to understand how to prevent muscle atrophy in space and improve the performance of athletes, but also to develop therapies for muscular and neuromuscular disorders such as muscular dystrophies, Parkinson's disease and multiple sclerosis. Apart from the effect of gravity on specific muscle subsets, different muscular dystrophies affect different muscle subsets (Figure 1).

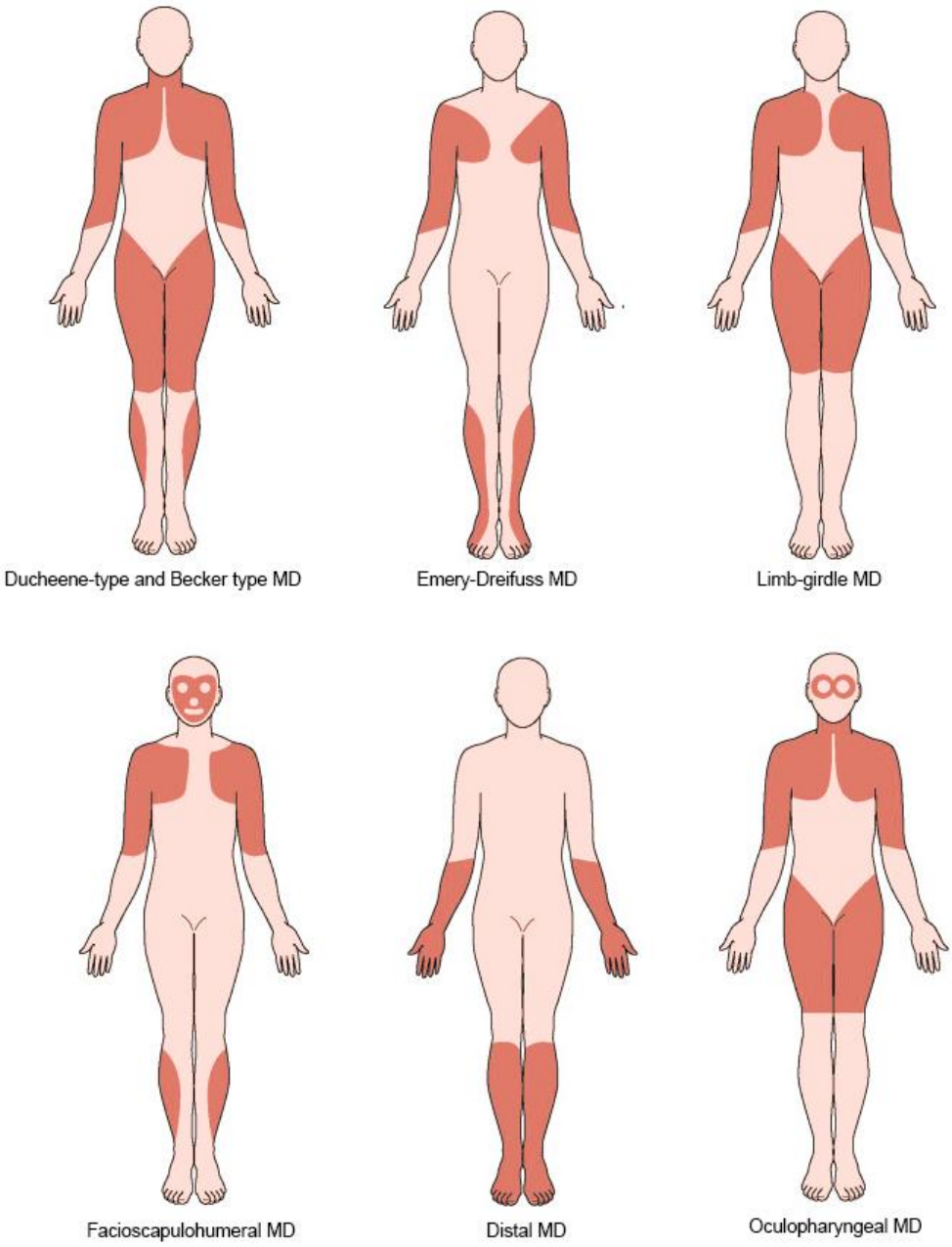


Figure 1. Muscle subsets affected in muscular dystrophies (MD).
Multiple forms of muscular dystrophies (myogenic disorders characterized by muscle wasting or atrophy and weakness) affect specific subsets of muscles highlighted in red (*adapted from Emery 2002*).

1.2. Vertebrate muscle development – how do these incredible structures form?

1.2.1. Types of vertebrate muscles

40% of the human body weight is made up of muscles (Janssen et al. 2000). They provide heat, balance us and stabilize our joints. There are 3 types of muscles, skeletal, cardiac and smooth muscles (Figure 2). Unlike skeletal muscles that receive signals from the somatic nervous system to control our voluntary movements, cardiac and smooth muscles receive signals from the autonomic nervous system to control involuntary movements such as heartbeat and the peristaltic movements of the intestine respectively. Skeletal and cardiac muscles are striated unlike smooth muscles that line the intestine or vasculature to promote involuntary peristaltic movements of food or blood. I focus on skeletal muscles. Each skeletal muscle is made up of uniquely patterned slow and fast muscle fibers characterized by the expression of specific isoforms of contractile proteins such as myosin heavy chain (MyHC) in humans (Bottinelli et Reggiani 2000; Talbot et Maves 2016).

1.2.2. Development of vertebrate skeletal muscles

1.2.2.1. Gastrulation and the specification of the mesoderm

The central dogma of patterning states that morphogen gradients hold positional information that leads to compartmentalization into domains, which then triggers identity acquisition in each domain by the expression of 'selector' genes and this finally leads to cross-tissue communication and the establishment of new gradients (Lawrence et Struhl 1996). This holds true for vertebrates and invertebrates and is the driver of all processes from the embryo into adult life. Therefore, I believe that an understanding of the very early developmental processes is essential to our understanding of later processes and process deregulations. So, I begin at gastrulation. Gastrulation is the process of formation of the three germ layers, the ectoderm, mesoderm and endoderm that give rise to all the tissues and organs. This process is characterized by an evolutionarily conserved series of cell movements (Solnica-Krezel 2005) that result in body axis definition and tissue primordia in vertebrates. It is initiated via the

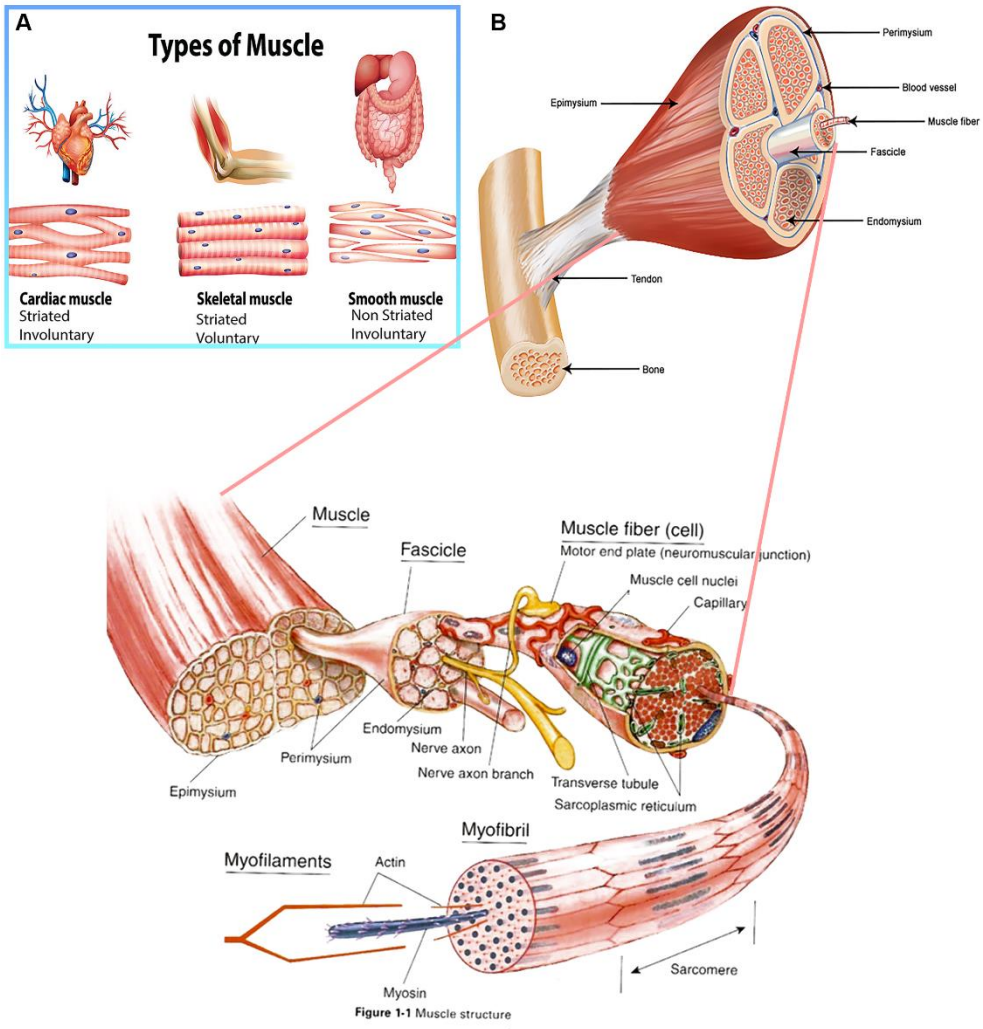


Figure 2. The types of vertebrate muscles and the structure of skeletal muscles.
(A) Humans possess three types of muscles. Cardiac muscles line the myocardium of the heart, have a striated appearance and are involuntary. Skeletal muscles assist with voluntary movements, are striated and attach to the skeletal system via tendons. Smooth muscles are involuntary and non-striated and line the gut and vascular system (*image courtesy: https://nursecepts.com/*). **(B)** Skeletal muscles are ensheathed by a protective epimysium. Each muscle is made up of multiple fascicles, each covered by a perimycium. Each fascicle consists of individual muscle cells known as muscle fibers ensheathed by an endomysium. A zoom on the muscle fiber shows its intricate structure with a syncytium formed of multiple muscle nuclei (myonuclei) and organelles such as the sarcoplasmic reticulum and transverse tubules forming an organized network, and its association with the nervous system at the neuromuscular junction as well as capillaries to supply oxygen. Each muscle fiber has multiple myofibrils made up of myofilaments that are composed of contractile units called sarcomeres. Actin and myosin form integral components of the sarcomere (*images courtesy: https://training.seer.cancer.gov/anatomy/ and https://functionalanatomyblog.com/2014/06/12/muscle/*).

invagination or internalization of cells through an opening in the blastula known as the blastopore in general among all vertebrates. In amniotes such as reptiles and mammals, it is always referred to using the term primitive streak.

In mammals, the embryo initially comprises a single layer of cells lining the amniotic sac known as the epiblast and another layer lining the blastocoel known as the hypoblast. The epiblast gives rise to all germ layers. Gastrulation is preceded by signals from the lip of the developing blastopore from a group of cells capable of inducing and specifying the future germ layers known as the node in mammals. It holds the key to the organism's fate map that decides cell fates (Iain Martyn, Siggia, et Brivanlou 2019). The pluripotent epiblast first forms a furrow at one end of the embryo by apical constriction of a few cells that change shape into wedge shaped cells triggered by molecular cues (Sawyer et al. 2010). This displaces other cells and the furrow deepens into the midline to give rise to the primitive streak. It is patterned into germ layers by BMP4->Wnt->NODAL signaling (Ben-Haim et al. 2006; I Martyn et al. 2018). As it grows, the cells in the center undergo an epithelial to mesenchymal transition (EMT) becoming motile and migrate down into the embryo to form a primitive groove. The Sry-related HMG box (Sox) transcription factor (TF), Sox2 is responsible for holding neural fated cells in a partial EMT state and preventing them from acquiring mesodermal fates by full EMT (Kinney et al., 2020). The cells that complete EMT divide and move laterally and radially to again circle towards the primitive streak (Figure 3A). Some of the cells settle between the epiblast and hypoblast to form the mesoderm while others move in between and displace hypoblast cells towards the edges to give rise to the endoderm (Figure 3B). The timing and position of cells defines which germ layer they are destined to be part of (Ferretti and Hadjantonakis 2019). As gastrulation progresses, the primitive streak recedes posteriorly.

1.2.2.2. Mesoderm diversification and somitogenesis

A tube-like mesoderm-derived notochord runs along the length of the embryo. It is situated below the developing ectodermal neural tube that later forms the brain and spinal chord. By gestational day 17 in humans, the mesoderm flanking the notochord diversifies into the paraxial, intermediate and lateral plate mesoderm triggered by morphogen gradients (Ferretti et Hadjantonakis 2019) (Figure 3C). The paraxial mesoderm gives rise to progenitors that contribute to the skeletal muscles and axial skeleton (Chal et Pourquié 2017). The intermediate

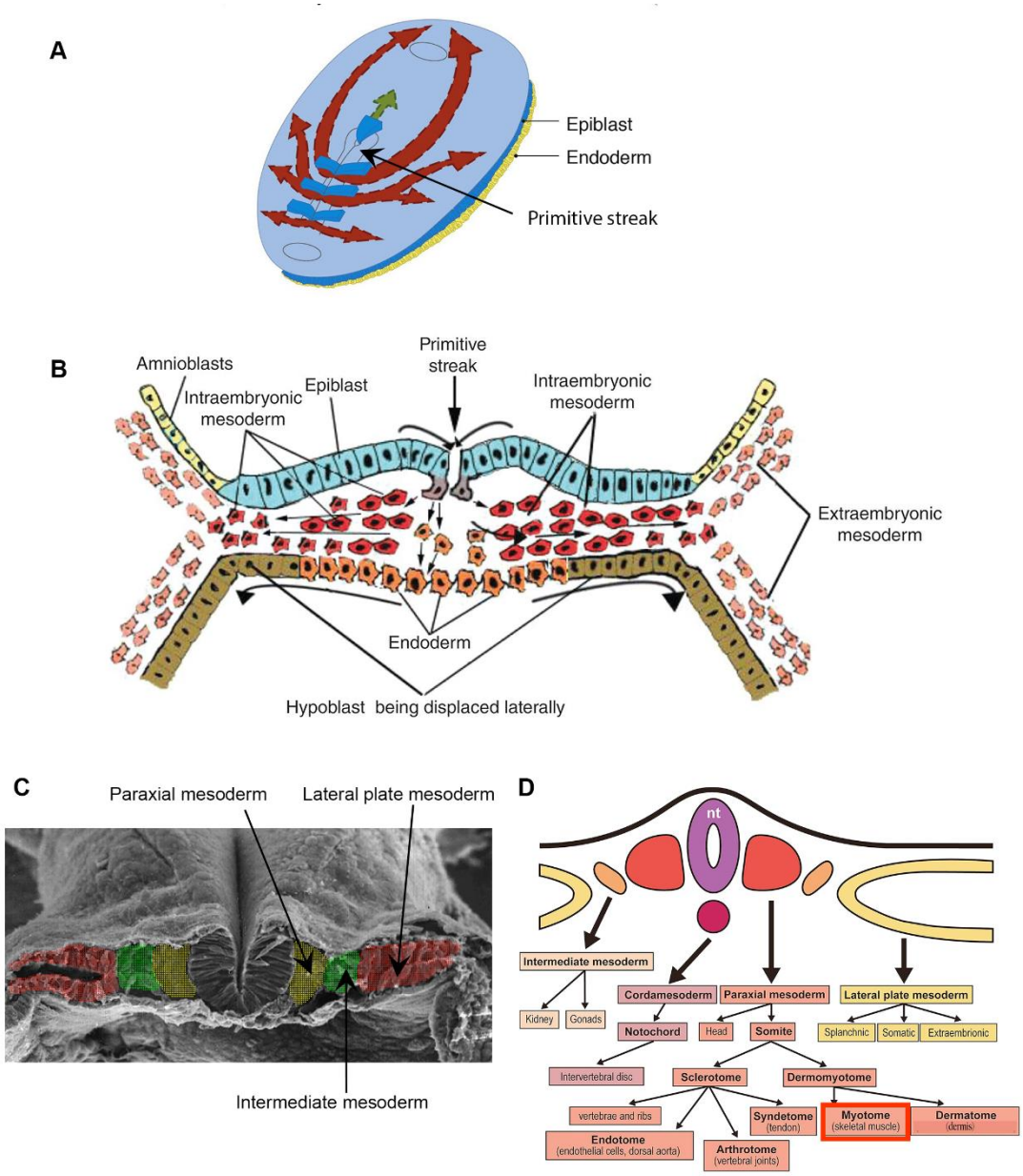


Figure 3. Gastrulation in vertebrate embryos.
(A) Cells in the primitive streak undergo epithelial to mesenchymal transition (EMT) and move laterally and radially again towards the primitive streak (*adapted from DeSesso 2017*). **(B)** Mammalian embryos initially consist of two layers, the epiblast lining the amniotic sac and the hypoblast lining the blastocoel. After undergoing EMT, some cells settle under the epiblast to give rise to the myogenic mesoderm while others displace the hypoblast to form the endoderm (*image courtesy: https://pocketdentistry.com/2-development-of-the-head-face-and-mouth/*). **(C)** The mesoderm then diversifies and is specified into the paraxial, intermediate and lateral plate mesoderm (*image courtesy: https://syllabus.med.unc.edu/courseware/embryo_images/*). **(D)** The different types of mesoderm give rise to specific tissues. The skeletal muscles (red rectangle) develop from periodically generated somites arising from the paraxial mesoderm (*adapted from Tani et al. 2020*).

mesoderm between the paraxial and lateral plate mesoderm gives rise to the urogenital system. The lateral plate mesoderm forms the heart and cardiovascular system, smooth muscles and limb skeleton (Prummel, Nieuwenhuize, et Mosimann 2020) (Figure 3D). The early paraxial mesoderm constitutes bilateral strips of mesenchymal cells flanking the notochord and is known as the presomitic mesoderm. It is specified due to the inhibition of BMPs by Noggin produced by the notochord and high Wnt/FGF signaling (McMahon et al. 1998). The winged helix TFs Foxc1 and Foxc2 determine the paraxial mesoderm fate versus intermediate mesoderm fate (Wilm et al. 2004). It also expresses downstream Wnt signaling effectors such as Brachyury (T), Tbx6, and Mesogenin1 (MSGN1). Tbx6 suppresses neural progenitor fates by repressing Sox2 (Sadahiro et al. 2018). In the anterior region where some cells need to adopt a neural fate, retinoic acid (RA) represses FGF8 (S. Kumar et Duester 2014).

In the head region, the paraxial mesoderm remains an unsegmented mesenchymal population and contributes to the heart, some skull bones and the skeletal muscles of the head and neck. In the trunk region, it undergoes bilateral somitogenesis from the head to tail end to generate a series of somites in a segmental pattern in a periodic fashion rostro-caudally (Chal et Pourquié 2017). While somites are forming anteriorly, the presomitic mesoderm grows caudally in coordination with embryonic growth by the addition of new tissue. Each somite is a block of epithelial cells formed by mesenchymal to epithelial transition (MET) of the presomitic mesoderm. The bHLH TF, Paraxis regulates this epithelialization mediated by ectodermal Wnt signaling (Burgess et al. 1996). Its expression is Foxc1 and Foxc2 dependent (Kume et al. 2001). The periodicity of somite generation and somite boundaries are determined by molecular oscillators in a 'clock and wavefront' model where the clock defines the periodicity while the wavefront determines the segment boundary (Pourquié 2011).

Periodic oscillations of expression of genes involved in FGF/Wnt and Notch signaling set up the segmentation clock whereas morphogen gradients of Wnt, FGF and RA set up the wavefront (Figure 4). The segmentation clock is comprised of a Hes7 dependent negative feedback loop that causes oscillations in FGF/Wnt and Notch signaling (Bessho et al. 2003; Niwa et al. 2007). Notch and FGF/Wnt show alternate oscillations (Aulehla and Hermann, 2004). FGF/Wnt signaling initiates the clock (Niwa et al. 2007; Wahl et al. 2007) causing the expression of their targets such as Sprouty4 and Axin2 that are FGF and Wnt repressors respectively. This causes a negative feedback and concomitant Notch signaling activation that promotes somite formation. Notch signaling activates the expression of its target genes including Hes7 that is a

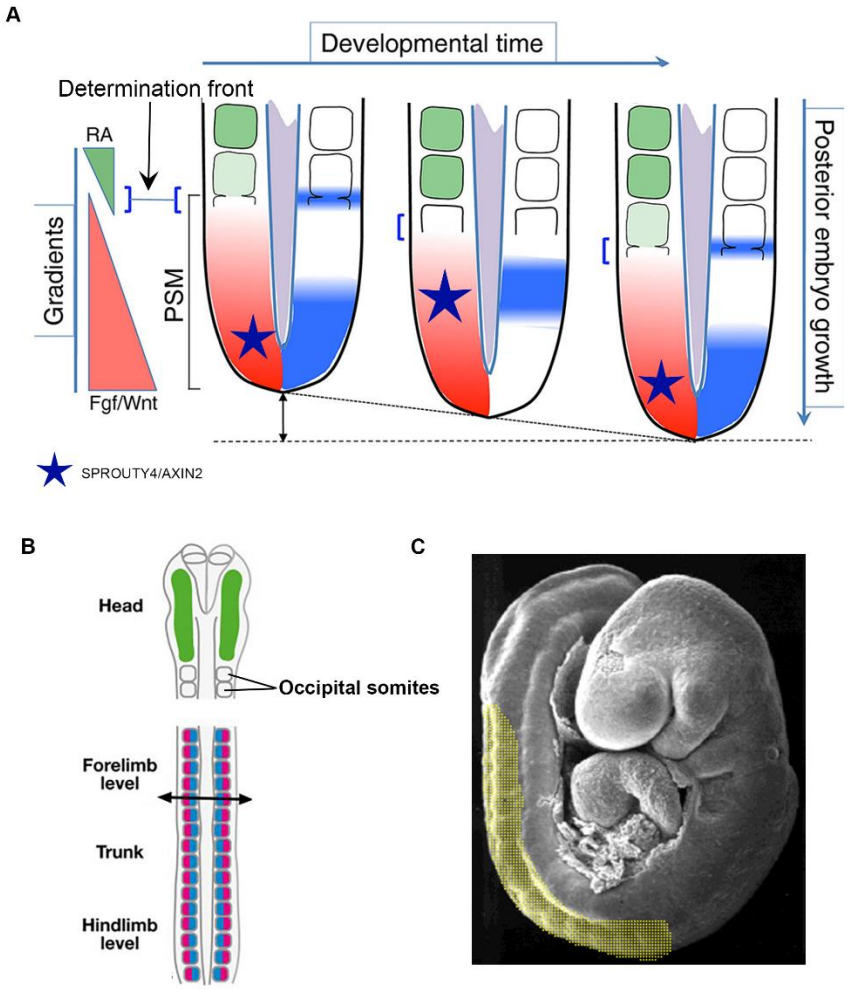


Figure 4. Somite generation by the 'clock and wavefront' model of cyclic gene activation.
(A) FGF/Wnt set up a posterior to anterior morphogen gradient (red) and retinoic acid (RA) sets up an anterior to posterior countergradient (green). FGF and Wnt inhibit somitogenesis whereas Notch (blue) promotes it (*adapted from Mallo 2016*). Their mRNA is transcribed at the tail bud and decays as the embryos grows posteriorly, thus causing their gradients. The 'segmentation clock' determines the periodicity of somitogenesis. FGF/Wnt activate the expression of genes including their inhibitors, *Sprouty4*/*Axin2*. Their repression permits the activation of Notch expression that in turn activates the expression of its repressor, *Hes7*. This causes cyclic FGF/Wnt and Notch signaling oscillations. These oscillations are absent at the determination front (blue brackets) where the FGF/Wnt levels fall below a threshold. Here, RA represses FGF and promotes epithelial condensation of the paraxial mesoderm known as presomitic mesoderm on each Notch wave to generate somites. **(B)** A view of the early mouse embryo after somitgenesis. The cranial region (green) lacks any overt segmentation. The first few somites called occipital somites give rise to some craniofacial muscles. In the trunk, there is a clear, segmented pattern of somites. **(C)** A view of a mouse embryo at day 9 of gestation reveals the segmental somites (yellow) (*image courtesy: https://syllabus.med.unc.edu/courseware/embryo_images/*).

Notch signaling inhibitor and downregulates Notch targets. This signal alternation leads to alternate oscillations in Notch and FGF/Wnt signaling. Hes7 protein is highly unstable and also inhibits its own transcription, thus maintaining oscillatory expression. Since the inactivation of FGF signaling abolishes Hes7 transcription, it has been proposed that it initiates Hes7 mRNA expression that is maintained by Notch signaling (Niwa et al. 2007; Wahl et al. 2007). Another gene that modulates Notch signaling cyclically is the Notch repressor Lunatic fringe (LFNG) (Dale et al. 2003; Okubo et al. 2012).

Wnt and FGF also establish a posterior to anterior gradient whereas retinoic acid (RA) establishes an anterior to posterior countergradient forming the 'wavefront' to determine somite boundaries. The Wnt/FGF gradient is generated due to embryonic growth since their mRNA is transcribed at the tail bud and then starts decaying as the embryo grows posteriorly, thus establishing a gradient that keeps moving posteriorly. The RA gradient is established by high levels of RA synthesizing enzymes in the anterior region and RA degrading enzymes in the posterior region (Aulehla et Pourquié 2010). High Wnt/FGF represses somitogenesis by maintaining cells in a mesenchymal state (Naiche, Holder, et Lewandoski 2011). At a certain point in the anterior region called the determination front, the threshold of these proteins falls below their inactivation potential and cyclic oscillations of genes is absent. Here, RA activates epithelial condensation by repressing FGF signaling and thus favors somite formation on the next Notch wave (Aulehla et Pourquié 2010; S. Kumar et Duester 2014). The RA gradient also determines somite polarity. MSGN1 is downregulated while MESP2 is upregulated at the onset on somitogenesis. MESP2 establishes somite boundaries by repressing Notch signaling activity (Saga 2007). It was shown in chick embryos that the progressive collinear activation of homeobox (HOX) genes that are arranged in clusters reflecting their advent of expression determines the termination of body axis elongation that in turn dictates the number of somites formed (Denans, Iimura, et Pourquié 2015). More posterior HOX genes repress Wnt signaling more strongly, resulting in the depletion of stem cell populations until no more stem cells are available for somite formation.

1.2.2.3. Somite specification

Each somite is a transient structure divided into an anterior TBX positive and posterior UNCX positive region. Soon after its generation, it compartmentalizes to give rise to a dorsal, epithelial dermomyotome and a ventral, mesenchymal sclerotome driven by cues from surrounding cells

(Figure 5). High Wnt levels from the roof plate of the neural tube and surface ectoderm induces the dermomyotome fate (Ikeya et Takada 1998) while Shh signaling from the notochord and the floor plate of the neural tube combined with low levels of Wnt and BMP signaling determines the sclerotome fate. The dermomyotome gives rise to the dermis as well as the skeletal muscles of the trunk, limbs and some head muscles while the sclerotome gives rise to the axial skeleton, tendons and cartilage to which muscles attach (Bentzinger, Wang, et Rudnicki 2012; S. Tajbakhsh et Cossu 1997).

In mice, the concerted action of the homeobox genes Meox1 and Meox2 is required for dermomyotome and sclerotome patterning and differentiation (Mankoo et al. 2003), but can partially compensate for each other's loss. Meox genes are upstream of paired box Pax TFs. Mice Meox1 mutants predominantly display defects of sclerotome derivatives while Meox2 mutants display limb muscle defects. Pax3 is initially expressed in the neural tube and paraxial mesoderm, then the entire somite, but subsequently becomes restricted to the dermomyotome along with Pax7 (Bentzinger, Wang, et Rudnicki 2012). Pax3 is essential for cell survival in the somite as evidenced from extensive apoptosis in studies conducted on Pax3 deficient mouse presomitic mesoderm explant cultures (Borycki et al. 1999). The dorsomedial and ventrolateral lips of the dermomyotome have high Pax3 expression while the central dermomyotome has high Pax7 expression (Galli et al. 2008; Kassar-Duchossoy et al. 2005). Pax7 positive cells are retained as a muscle progenitor pool by Notch signaling, are capable of self-renewal and play a significant role in muscle maintenance during adult stages (Relaix et al. 2005; Schuster-Gossler, Cordes, et Gossler 2007; Seale et al. 2000). Inhibition of Notch signaling leads to a depletion of this progenitor population and in muscle deficits during later stages of development.

1.2.2.4. Vertebrate myogenesis

Myogenesis or muscle development requires close cooperation and communication among different tissues (Deries et Thorsteinsdóttir 2016). Four stages of myogenesis can be distinguished during amniote muscle development: embryonic, fetal, neonatal (P0-P21 in mouse limbs, P is post-natal day) and adult myogenesis (after P21 in mouse limbs) (Murphy et Kardon 2011) (Figure 6). During these stages, myogenesis proceeds sequentially by first generating muscle progenitors competent to generate muscles. They are then specified to myoblasts that differentiate into myocytes. During fetal and neonatal stages, myocytes

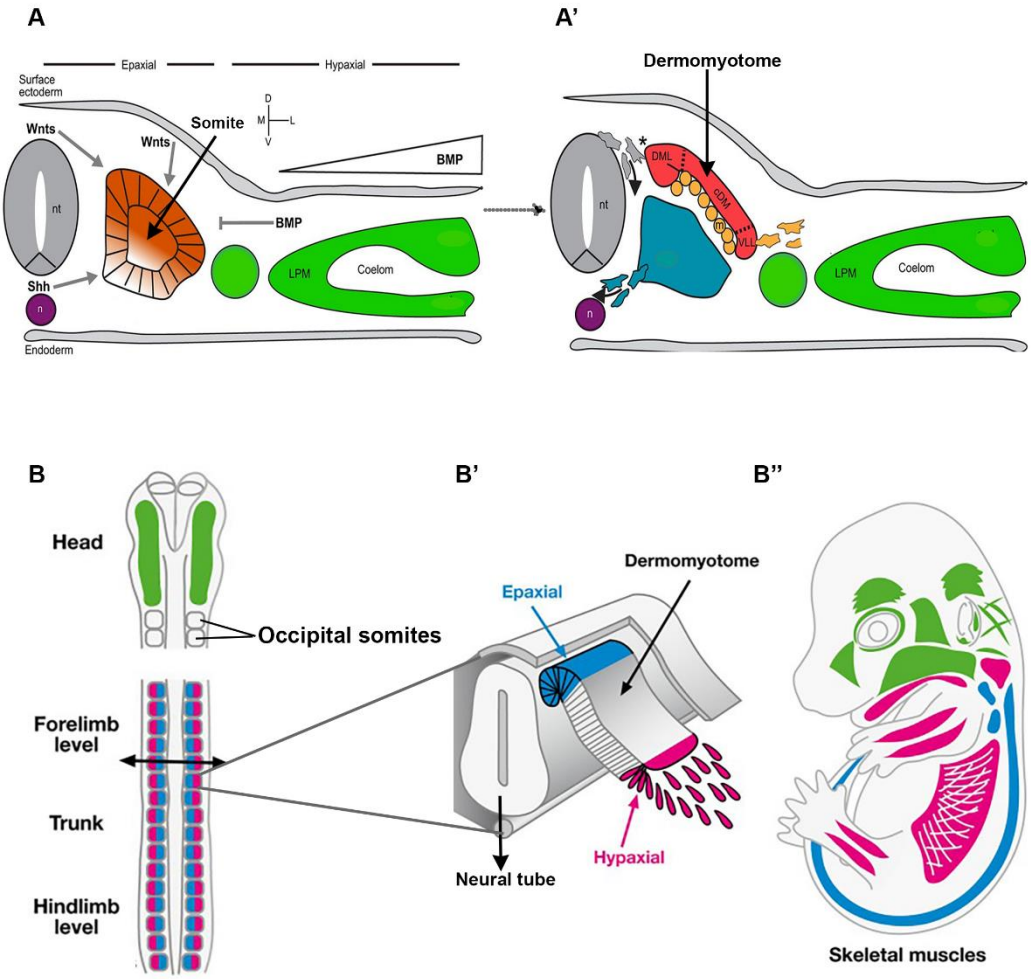


Figure 5. Somite specification

(A-A') Soon after a somite is formed, it is specified into different regions based on signaling cues from adjacent tissues. High Wnt levels from the roof plate of the neural tube (nt) and surface ectoderm (A) induce the dermomyotome fate (red in A'). Shh signaling from the notochord (n) and floor plate of the neural tube and low Wnt and BMP levels (A) induce the sclerotome fate (blue in A'). Muscle progenitor cells (MPCs) then delaminate from the dorsomedial lip (DML) and central dermomyotome (cDM) to form the myotome (m) under it. A few MPCs from the ventrolateral lip (VLL) migrate laterally towards the lateral plate mesoderm (LPM) to enter the limb buds (*adapted from Chal et Pourquié 2017*). **(B-B'')** The unsegmented, mesenchymal paraxial mesoderm in the embryonic head (green in B) gives rise to craniofacial muscles in the adult (green in B''). MPCs adjacent to the neural tube form the epaxial myotome (blue in B') that gives rise to epaxial or back muscles (blue in B'') while the MPCs adjacent to the surface ectoderm form the hypaxial mytome (pink in B') that give rises to hypaxial muscles such as the ventral body wall, limb and tongue muscles (pink in B'') (*adapted from Nassari, Duprez, et Fournier-Thibault 2017*).

subsequently form primary myofibers by fusing with myoblasts and then secondary myofibers by the fusion of myoblasts either with themselves or to the primary myofibers. During adult stages, muscles grow by hypertrophy (increase in muscle size) rather than *de novo* generation. The basic muscle pattern is established during embryonic myogenesis, the muscle grows and matures during fetal myogenesis and there is further growth and repair during adult stages. A muscle stem cell (MuSC) population is maintained during adult stages for muscle maintenance and regeneration after injury.

1.2.2.4.1. Prenatal myogenesis

1.2.2.4.1.1. Development of trunk muscles

Prenatal myogenesis can be divided into a primary embryonic phase (E10.5-E12.5 in mouse, E = embryonic day in mouse) and secondary fetal phase (E14.5-E17.5) (Biressi, Molinaro, et Cossu 2007; Chal et Pourquié 2017). During the primary phase, post mitotic Pax3 and Pax7 positive muscle progenitor cells (MPCs) from the dorsomedial and ventrolateral lips of the dermomyotome delaminate to mature and commit to the muscle fate and form the myotome under the dermomyotome, the earliest stage of skeletal muscles in amniotes (Gros et al. 2005). At this stage, their identity is specified by expressing high levels of the basic helix-loop-helix (bHLH) TFs Myf5, Mrf4 and MyoD and they are termed myoblasts (Kassar-Duchossoy et al. 2004). In mice, paracrine factors from adjoining tissues pattern the myotome. It was shown that in mice paraxial mesoderm cultured with adjoining tissues, in the dermomyotome, medial muscle progenitors adjacent to the neural tube and notochord receive cues from these tissues and form the epaxial myotome that gives rise to epaxial or back muscles. These Wnt1 cues activate Myf5 expression as early as immediately after somite formation. Muscle progenitors adjacent to the surface ectoderm receive cues initiated by ectodermal Wnt7A that activate MyoD expression and give rise to the hypaxial myotome that forms ventral body wall muscles, limbs, diaphragm and tongue (S. Tajbakhsh et al. 1996; G. Cossu et al. 1996; S. Tajbakhsh et al. 1998).

The muscular system displays a high degree of plasticity as evidenced by the redundancy between Myf5 and MyoD. Although these TFs are expressed in different parts of the myotome, they can partially compensate for each other's loss (Rudnicki et al. 1993). Pax3 and Myf5 play multiple roles over the course of muscle development. Myf5 is first expressed in the epaxial myotome and subsequently in all skeletal muscles. Pax3 cooperates with Six4 to regulate the

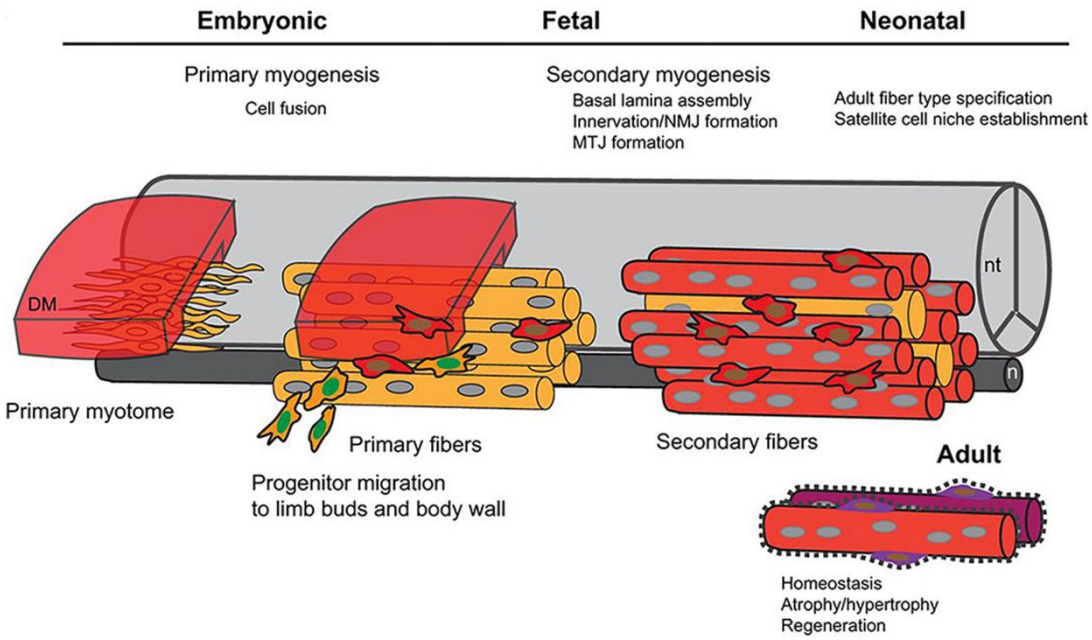


Figure 6. Embryonic, fetal and neonatal myogenesis.
During primary embryonic myogenesis, muscle progenitor cells (MPCs) delaminate from the dermomyotome (DM) to give rise to the myotome underlying it. They differentiate into elongated, mononucleated myocytes by expressing terminal muscle differentiation factors such as Myog. They then fuse with embryonic myoblasts to form syncytial primary myofibers. MPCs destined to form limbs delaminate from the ventrolateral lip of the dermomyotome and migrate long distances to limb buds instead of the short distance to integrate with the myotome as in the case of body wall muscles. During fetal development, muscles undergo secondary myogenesis during which the primary myofibers fuse with fetal myoblasts with characteristics distinct from embryonic myoblasts to give rise to secondary myofibers. These are initially attached to primary fibers and subsequently separate by forming a basal lamina. Muscles are innervated at the neuromuscular junctions (NMJ) and attach to the skeleton via tendons at the myotendinous junctions (MTJ). Satellite cells (SCs) start to occupy their positions between the sarcolemma and basal membrane. Once the final muscle identity is established, the SC niche environment is established and muscles acquire individual muscle fiber properties with slow/fast fibers. Postnatally, muscles grow in size by increasing their volume by fusing with SCs. SCs also contribute to adult muscle homeostasis and regeneration (Chal et Pourquié 2017).

spatiotemporal expression of *Myf5* (Daubas et Buckingham 2013; Giordani et al. 2007). Apart from Pax3 and Six binding sites, the *Myf5* enhancer has binding sites for TCF/LEF that is involved in transducing Wnt signals via the canonical TCF-LEF/ß-catenin pathway as well as Gli that transduces Shh signals. Both these binding sites are essential in mice to drive full early *Myf5* expression in the epaxial myotome (Borello et al. 2006). In mice, myogenic progenitors also express the bHLH TF, M-Twist (Füchtbauer 1995). Shh appears to be important for lineage progression of the myotome as shown in studies in the anamniote, zebrafish where the absence of Shh signaling increased Pax3/7 positive cells, but prevented subsequent differentiation (Hammond et al. 2007). Conversely, Shh overexpression repressed Pax3/7. ChIP-Seq studies in mice embryoid body cultures in conjunction with ATAC-Seq and TF knockdown have revealed that Pax3 is an initiator of the myogenic program by modifying chromatic accessibility, which renders enhancers accessible to TFs such as Six4 and Tead2 (Magli et al. 2019). *Pax3* mutant homozygous (*Splotch*) mice that possess a spontaneous *Pax3* mutation exhibit severe deformations and loss of musculature derived from the hypaxial myotome as well as deformations in the skeleton. These mutants exhibit a high level of apoptosis and upregulated Pax7 expression in the somites. Pax3 has also been shown to regulate *Pax7* expression and cell survival (Borycki et al. 1999). Myf5 and MyoD expressing myoblasts differentiate into mononucleated muscle fibers called myocytes by expressing differentiation factors such as Myogenin (Myog) and Mrf4. Mrf4 has a biphasic expression in mice where it is first expressed transiently at the onset of myogenesis and can commit cells to a myogenic fate, then switches off and is again expressed during fetal stages and lingers after birth (Bober et al. 1991; Moretti et al. 2016). *Mrf4* (also known as *MYF-6*) is linked to *Myf5* in the same gene locus (Braun et al. 1990).

The bHLH factors Myf5, MyoD, Mrf4 and Myog are evolutionarily conserved vertebrate skeletal muscle TFs termed as Myogenic Regulatory Factors (MRFs) (Figure 7). MRFs act in concert with the MADS-box Mef2 family of TFs comprising Mef2A-D. Different Mef2 isoforms are detected early on in somites and limb buds, but its role during this stage is unknown. It is capable of inducing myogenesis in fibroblasts when coexpressed with MyoD or Myog (Molkentin et al. 1995). In mice, alternative splicing generates a muscle-specific *Mef2d* isoform (J. F. Martin et al. 1994). During mouse embryogenesis, *Mef2c* expression is detected almost a day after *Myf5* expression, whereas *Mef2a* and *Mef2d* are expressed only later and are expressed in myogenic cells distinct from *Mefc* expressing cells (Edmondson et al. 1994).

Mef2c has been shown to be essential for cardiac myogenesis and are embryonic lethal (Lin et al. 1997). The *Mrf4* and *Myog* promoters carry Mef2 binding sites.

The myocytes elongate and differentiate, then fuse with embryonic myoblasts forming primary myofibers that organize themselves rostro-caudally to form the primary myotome (Gros, Scaal, et Marcelle 2004; Murphy et Kardon 2011) that lays out the primitive muscle morphology and position. Many signaling molecules and proteins involved in cell migration and recognition are known in vertebrates, but very little is known about the mechanisms of membrane fusion to generate syncytial myotubes. It was only recently shown that two genes, *Mymk* and *Mymx* coding for Myomaker and Myomerger/Minion/Myomixer respectively were capable of inducing non fusogenic cells to fuse and that they control distinct steps during fusion (Bi et al. 2018; Leikina et al. 2018; Lehka et Rędowicz 2020). The myotome grows rostro-caudally while still maintaining its segmented nature without any muscle innervation (Deries, Collins, et Duxson 2008). The primary myotome expresses high levels of the slow MyHC, MyHC-I and and the embryonic MyHC, MyHC-emb. Sox6 promotes *MyHC-I* transcription by promoting Mef2c phosphorylation and activation (Taglietti et al. 2016).

During the second phase of fetal prenatal myogenesis, muscle progenitors in the central lip of the dermomyotome undergo EMT losing epithelial markers such as ß-catenin and intercalating between the primary myotome (Delfini et al. 2009) in chick embryos. This is dependent on FGF cues received by the deromomyotome from the primary myotome. They also undergo proliferation and fetal myoblasts with characteristics distinct from embryonic myoblasts (Giulio Cossu et Biressi 2005) fuse with the primary myofibers to form secondary myofibers of this secondary myotome (Gros et al. 2005). There remains a reserve myoblast population of proliferating $Pax3^{+}/Pax7^{+}$ progenitors at this stage. Canonical Wnt signaling increases the number of fetal $Pax7^{+}$ progenitors, but not embryonic progenitors (Hutcheson et al. 2009). The secondary myofibers initially remain in close contact with primary myofibers. Subsequently, a basal lamina is formed around each myofiber and $Pax7^{+}$ reserve progenitors called satellite cells (SCs) situate themselves between it and the plasma membrane of the myofiber (Seale et al. 2000; Giulio Cossu et Biressi 2005). In $Pax3^{-}/Pax7^{-}$ mice mutants, several trunk muscles are severely compromised, although the initial myotome does form (Relaix et al. 2005). During fetal myogenesis, secondary myofibers lose the expression of *MyHC-I*, whose transcription is

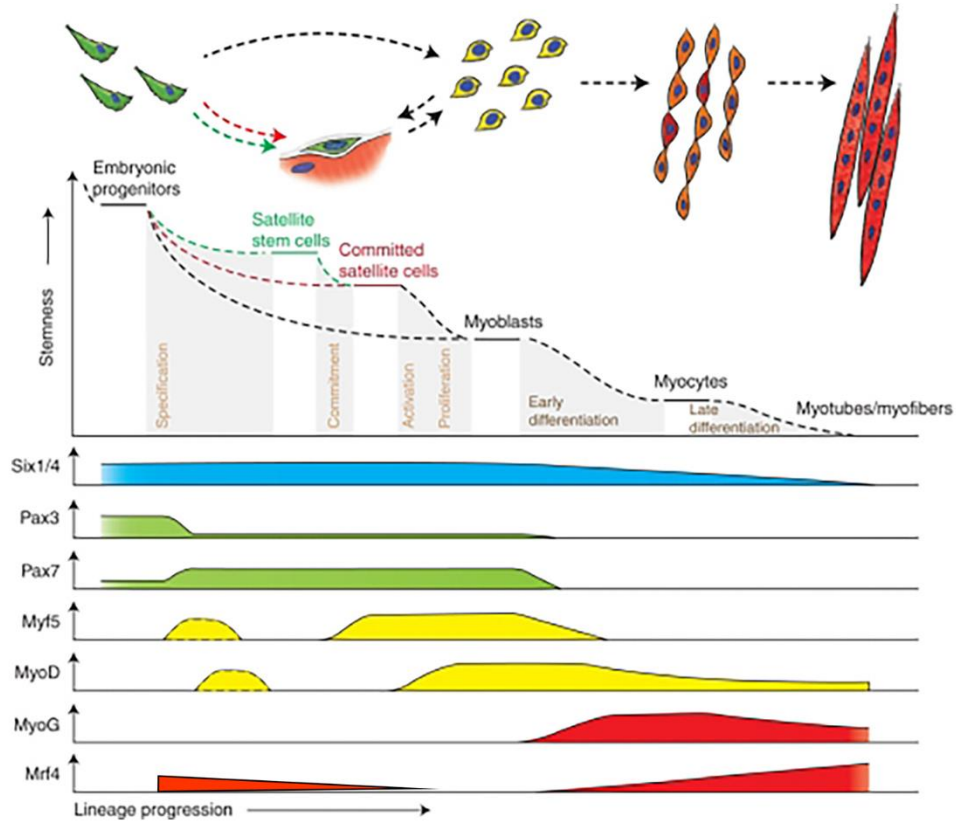


Figure 7. Expression of myogenic regulatory factors (MRFs) and other myogenic genes that regulate lineage progression of muscle cells.
In multipotent embryonic muscle progenitor cells (MPCs) in the trunk region, Six1/4 cooperate with Pax3 to initiate the myogenic program and activate *Pax7* expression among others. Depending on the region of the dermomyotome that the MPCs delaminate from, they activate the expression of genes such as *Mrf4*, *MyoD* and/or *Myf5* and commit to the myogenic fate as myoblasts. *Mrf4* has a biphasic expression with an initial transient expression along with *Myf5* and *MyoD* that tapers off and then reinitiates during fetal myogenesis. *Myog* expression is initiated when myoblasts begin terminal differentiation. A *Pax7+* reserve progenitor cell population gives rise to satellite cells that maintain adult muscles and participate in muscle regeneration upon injury. The basic helix-loop-helix TFs *Mrf4*, *MyoD*, *Myf5* and *Myog* are known as myogenic regulatory factors (MRFs) and are highly conserved among vertebrates (*adapted from Bentzinger, Wang, et Rudnicki 2012*).

repressed by Sox6 contrary to its role in the activation of *MyHC-I* transcription along with Mef2c during embryonic myogenesis (Taglietti et al. 2016).

1.2.2.4.1.2. Development of limb muscles

Unlike the trunk muscles that develop within the myotome, limb muscle progenitors are $Pax3^{+}/Pax7^{-}$ and undergo EMT to migrate into limb buds generated by FGF signaling (Cohn et al. 1995; Hutcheson et al. 2009; McQueen et Towers 2020) after delaminating from the lateral dermomyotome in response to cues from the lateral plate mesoderm (Christ et Brand-Saberi 2002; Francis-West, Antoni, et Anakwe 2003). This delamination requires ß-catenin (Hutcheson et al. 2009). These migratory cells are generated only from somites at the level of limb buds and express C-met and the homeodomain TF Ladybird homeobox 1 (Lbx1), both being downstream of *Pax3* (Bladt et al. 1995; Epstein et al. 1996; Mennerich, Schäfer, et Braun 1998; Brohmann, Jagla, et Birchmeier 2000). Lbx1 phosphorylation at its C-terminal activates progenitor migration in chick embryos (Masselink et al. 2017). Pax3, C-met and Lbx1 are necessary for progenitor migration. Their expression is reduced and migratory cells undergo apoptosis in $Six1^{-/-}/Six4^{-/-}$ double mutant mouse embryos (Grifone et al. 2005). The homeodomain factors Msx1 and Meox2 act as MRF transcription repressors to prevent premature differentiation of migrating progenitors (Daubas et Buckingham 2013). So does the bHLH-Pas (Per-Arnt-Sim) TF, Single-minded 2 (Sim2) (Havis et al. 2012) (Figure 8A, C, D). The migrating progenitors remain $Pax7^{-}$ until they migrate to the limb bud. Pax7 downregulates differentiation markers such as *MyoD* and *Myog*.

In the developing limb buds of the chick embryo, Myostatin, a TGF-ß family member, was shown to regulate the proportion of progenitor and differentiated muscle cell populations. An earlier study in the chick embryo observed reduced differentiation markers in the chick limb bud upon Myostatin overexpression. A subsequent study found that Myostatin induces the expression of differentiation markers such as MyoD in progenitors, thus depleting the progenitor population and reducing muscle mass (Amthor et al. 2006; Manceau et al. 2008). Others studies in chick embryos helped provide more insights. Myogenic differentiation proceeds proximo-distally based on the proximity of progenitors to an area of thickened ectoderm at the tip of the limb bud known as the apical ectodermal ridge (AER). The AER and its underlying mesenchyme secrete multiple factors that inhibit myogenic differentiation including many FGFs (FGF2, 4) (Robson et Hughes 1996) and scatter factor/hepatocyte growth

factor (SF/HGF) (Scaal et al. 1999). These inhibitory signals help maintain a progenitor pool by the expression of the transcriptional repressor Msx1 (Bendall et al. 1999). The Wnt antagonist, Sfrp2 is expressed in migrating progenitors, but not in differentiating myoblasts suggesting a role in the differentiation of limb progenitors, although it appears that different Wnts have different stage-specific roles (Anakwe et al. 2003; Geetha-Loganathan et al. 2005; Hutcheson et al. 2009).

In aquatic vertebrates such as fish and tadpoles, myotomes form immediately after somite formation without a distinguishable dermomyotome stage and differentiate into slow and fast muscles that are rapidly innervated to allow them to swim (Devoto et al. 1996; Hollway et Currie 2003). Very few muscles form by the migration of muscle progenitors.

1.2.2.4.1.3. Development of craniofacial muscles

The craniofacial skeletal muscles are comprised of extraocular, branchial, laryngoglossal and axial muscles. Extraocular muscles rotate the eyes and protect the cornea, branchial muscles include muscles of the jaw among others, laryngoglossal muscles include the tongue and laryngeal muscles and axial muscles include muscles that rotate or elevate the skull. These muscles are derived from the non-somitic prechordal (forms extraocular muscles), the cranial paraxial mesoderm that remains mesenchymal without overt somatic segmentation (forms branchiomeric muscles including the masticatory, pharyngeal and laryngeal muscles) as well as the first few somites termed as occipital somites (form tongue and ventral neck muscles) (Noden et Francis-West 2006). This craniofacial myogenic mesoderm is surrounded by tissues distinct from that found in the trunk and thus generates a distinct set of signals to generate muscles. Similar to limb buds, most craniofacial muscle primordia migrate remotely. Pax3 or Pax7 are not detected. Pax3/Myf5 double mutants have no trunk muscles, but head muscles appear normal (Shahragim Tajbakhsh et al. 1997). On the other hand, in mutants for the T-Box TF Tbx1, branchiomeric myogenesis is severely affected (Kelly, Jerome-Majewska, et Papaioannou 2004). The development of some pharyngeal muscles was shown to be Tbx1 and Myf5 dependent, but Mrf4 independent (Figure 8A, B).

Cranial muscles arising from the first branchial arch are missing in mutants for the bicoid-related homeobox gene, Pitx2 (Shih, Gross, et Kioussi 2007). Pitx2 was found to bind the *Tbx1* enhancer and is necessary for the expression of the branchiomeric progenitor markers, *Tcf21*

(*Capsulin*) and *MSC* (*MyoR*) (Lu et al. 2002; Shih, Gross, et Kioussi 2007). Either Myf5 or Mrf4 were shown to be required for the activation of the differentiation program in mice extraocular muscles upstream of MyoD (Sambasivan et al. 2009) (Figure 8A, B). However, this study observed a small number of MyoD/Myog positive cells. A subsequent study showed that Pitx2 was an upstream activator of extraocular muscle development in mice by directly binding to *Myf5* and *MyoD* promoters (Zacharias et al. 2011). Twist1 genetically interacts with IRF6 during craniofacial development in mice and both genes are associated with human craniofacial disorders (Fakhouri et al. 2017). In chick embryos, Iselet1 (Isl1) contributes to distinct SCs associated with jaw muscles and represses muscle differentiation.

1.2.2.4.2. Postnatal and adult myogenesis

During peri and postnatal development, muscles grow by hypertrophy (increasing the muscle size) rather than by hyperplasia (*de novo* muscle generation) and undergo massive growth. $Pax7^{+}$ SCs divide slowly and contribute to this growth by adding new muscle nuclei. Postnatal myofibers undergo fiber specialization. During the embryonic to fetal/postnatal switch, Mef2a cooperates with NFIX (Imbriano et Molinari 2018). After postnatal growth and the generation of a mature muscle, post mitotic SCs enter a quiescent state. During this, Notch induced expression of *miR-708* maintains quiescence and represses *Tns3*, which leads to the retention of SCs in their niche between the basal lamina and sarcolemma instead of activating migration (Baghdadi et al. 2018). These SCs become activated in the adult muscle on injury and contribute to muscle regeneration.

1.2.2.4.3. Muscle attachment and innervation

Muscle maturation occurs when it is associated with connective tissues. They attach to tendons via the myotendinous junction (MTJ) and to motor neurons via the neuromuscular junction (NMJ). Skeletal muscles attach to the skeletal system via the MTJ formed by muscles and tendons that are fibrous connective tissues derived from the mesoderm. They are also innervated by motor neurons at NMJs in order to be able to receive signals from the central nervous system (CNS) to initiate contractions. These processes are only just beginning to be understood in vertebrates. Scleraxis (Scx) positive tendon progenitors arise in the sclerotome localized ventral to the dermomyotome in somites in a region called the syndetome. Scx expression is absent in $MyoD^{-/-}/Myf5^{-/-}$ mice mutants lacking myotome differentiation (Brent, Braun, et Tabin 2005). Tendon progenitor induction requires FGF signals from the myotome (Brent et Tabin 2004) and TGFß signaling (Pryce et al. 2009). Unlike these somitic tendons,

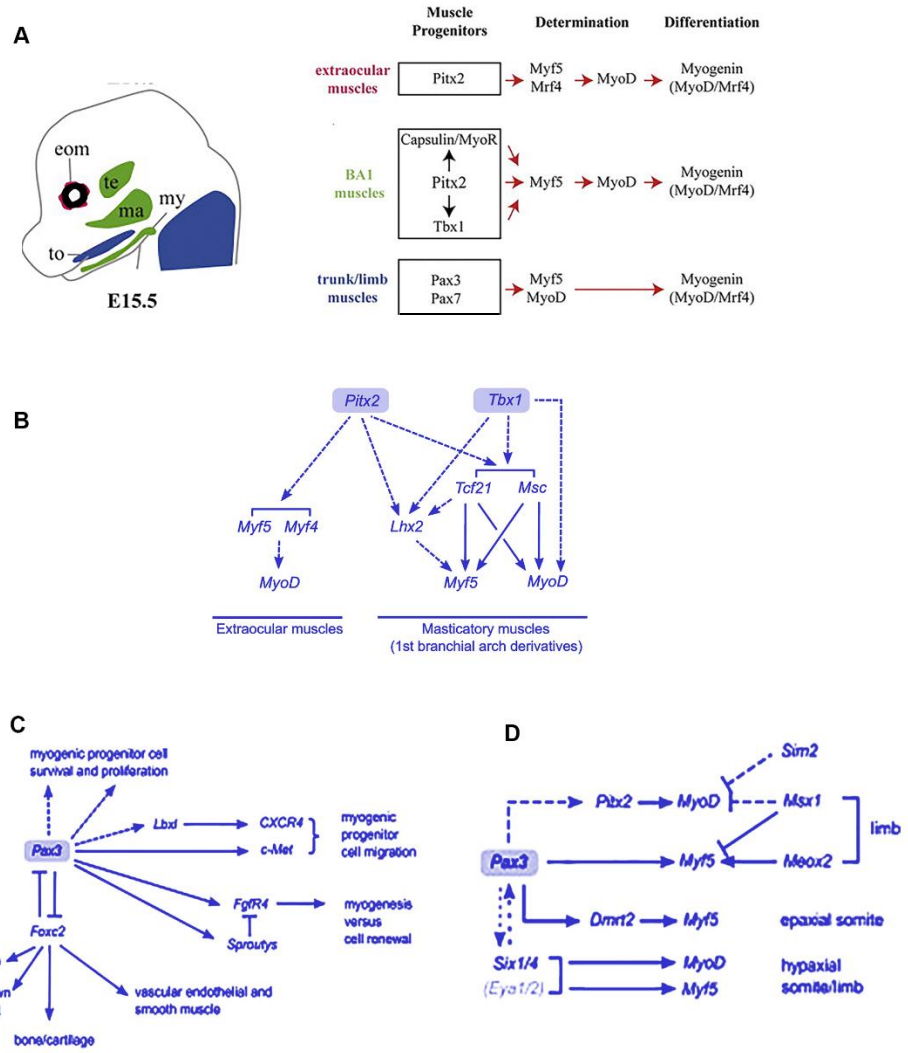


Figure 8. Regulatory networks dictating skeletal muscle identity.
(A) The major upstream regulators of the identity of specific muscle subsets are displayed. Extraocular muscles that originate from the prechordal mesoderm (pink, eom) are regulated by Pitx2 that activates *Myf5* and *Mrf4* expression. Either Myf5 or Mrf4 is sufficient for the normal development of these muscles. Muscles derived from the first branchial arch (BA1) (ma: masseter, te: temporalis, my: mylohyoid in green) are also specified by Pitx2 along with Tbx1 and Tcf21 (also known as Capsulin) and Msc (also known as MyoR) that activate MRF gene expression. The trunk and limb muscles are Pax3 and Pax7 dependent *(adapted from Buckingham et Vincent 2009*). **(B)** A more elaborate TF network implicated in craniofacial muscle development highlights the role of Lhx2 in specifying BA1 muscle identity (*adapted from Buckingham 2017*). **(C)** Pax3 plays multiple roles during myogenesis. It acts as a determinant of the myogenic program by repressing *Foxc2* expression. It promotes cell renewal by binding to the promoter and activating *FGR4* that is coexpressed with Pax7 to maintain stemness. In addition, it activates the migratory behavior of limb muscle progenitors by activating *Lbx1* expression. **(D)** *Pax3* is itself positively regulated by Six1/4 and it in turn genetically interacts with them to regulate MRF gene expression. It also genetically interacts with Pitx2 during limb muscle development to promote differentiation. During the migratory phase, differentiation is inhibited by Sim2 and Msx1.

limb tendons are induced in the limb bud mesenchyme by BMP signals from the overlying ectoderm (Schweitzer et al. 2001). While the tendon attaches to the muscle on one end, the other end is attached to the bone forming an enthesis (Shaw et Benjamin 2007). Innervation also exhibits differences between myotome-derived and limb muscles. Deries et al. (Deries, Collins, et Duxson 2008) showed that the myotome differentiates in the absence of innervation in rat. Using the muscle differentiation marker MHC along with Pax3 to identify differentiated muscles and neurofilament+synaptophysin to visualize motor axons, they studied the innervation of developing myotomal epaxial muscles at the limb level so as to visualize developing limb muscles in parallel. Their study showed that while the motor neurons that innervate epaxial muscles are in the vicinity of the myotome, they pause parallel to the myotome until they differentiate before innervating them. Limb muscles, on the other hand, differentiate and grow in contact with their innervating motor neuron. A subsequent study showed that Pax3+ muscle progenitors invade the limb bud before it is invaded by nerves whereas Pax7+ progenitors appear only in the close vicinity of the invading nerves (Hurren et al. 2015).

The myogenic processes lead to the formation of mature, attached, innervated muscles with distinct muscle identities.

1.3. The need for simpler model systems

Studies using human tissues are constrained by ethical considerations. This also applies to other mammalian models such as monkeys, rats and mice making them difficult and expensive to procure. In addition, they have relatively long gestation periods and life cycles necessitating long waits for studies in developmental biology. Mice, for example, have a 21-day gestation period and 50-60 day life cycle. Other vertebrate models such as zebrafish with an around 12-week life cycle are popular. Vertebrate systems, however, are highly complex and processes are generally regulated by large gene families instead of one or few genes. Given that many processes and genes are conserved across evolution, this led to the search for simpler models including invertebrate models. Muscles are present in the last common ancestor of invertebrates, animals that lack a vertebral column, as well as vertebrates such as *Homo sapiens* (Hooper et Thuma 2005).

The fruit fly, *Drosophila* emerged as a model system during the beginning of the 20th century when Thomas Hunt Morgan's laboratory popularized it (Hales et al., 2015). The *Drosophila* genus is part of the Arthropod phylum that is the largest phylum in the animal kingdom (G. W. C. Thomas et al. 2020). The *Drosophila melanogaster* species has a worldwide distribution and thus provides incredible opportunities as a model system. It has served as a valuable model system since the 20th century due to multiple factors. It has a short life cycle (Figure 9) that allows the generation of adult progeny in as few as 10 days. There are a multitude of genetic manipulation strategies available that have been generated and refined over the decades of research using it as a model system (Botas 2007; Hales et al. 2015; Markow 2015). *Drosophila* also has well characterized reference genomes with curated, up-to-date information available in the expansive Flybase database dedicated to it (http://flybase.org/; Marygold et al., 2016). A large number of genes in fruit flies have vertebrate homologues (Fouchécourt et al. 2019). Large gene families of vertebrate genes are often represented by one *Drosophila* homologue, the muscle differentiation gene *Myocyte enhancer factor 2* (*Mef2*), for example. This correlates with observations of less severe phenotypes in vertebrates when a single gene member of a family is mutated while mutations in the single *Drosophila* homologue lead to drastic phenotypes. For example, in *Drosophila* mutant flies that fail to complement a genetic deficiency or removal of the *Mef2* gene in trans due to the absence of the Mef2 protein, muscle

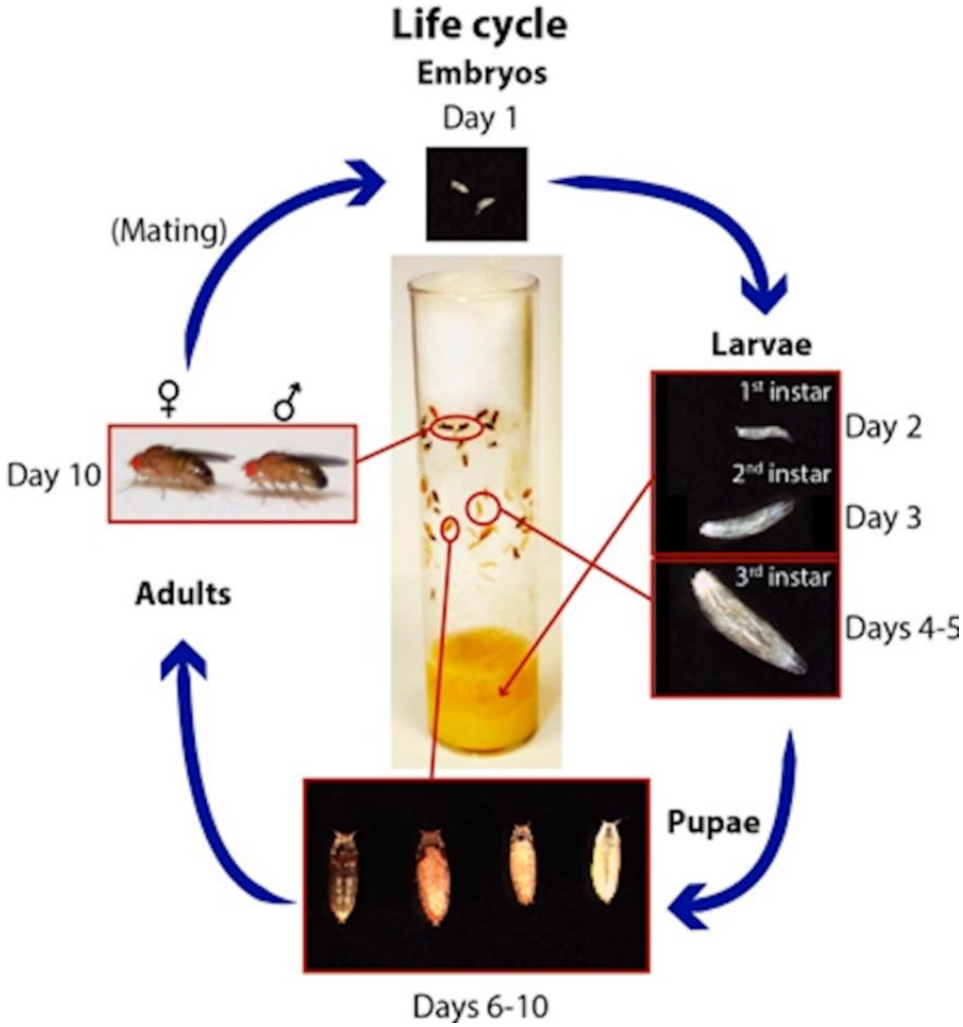


Figure 9. The life cycle of *Drosophila melanogaster*.
Drosophila is a holometabolous insect with life stages that include the egg, larva, pupa and adult. *Drosophila melanogaster* is one of the *Drosophila* species. Following egg laying, embryonic stages are complete in a span of a few hours at 25°C. The eggs hatch into larvae within a day, thanks to somatic muscles that develop during embryonic stages. They undergo three larval instars by molting. The larval stage lasts for 4-5 days following which they undergo pupation. Adults emerge in as few as 10 days after egg laying depending on the temperature. High temperatures promote faster growth (G. W. C. Thomas et al. 2020).

cells known as founder cells form, but completely fail to differentiate into muscles (Bour et al. 1995). The vertebrate *Mef2* gene family consists of four genes, *Mef2a*, *Mef2b*, *Mef2c* and *Mef2d*. Among these, mice with an early conditional deletion of *Mef2c* are characterized by a severe disorganization of skeletal muscle fibers and neonatal lethality. In contrast, in early *Mef2a* and *Mef2d* mutants, skeletal muscles develop normally (Potthoff et al. 2007).

Considering that it is an invertebrate, holometabolous insect with egg, larval, pupal and adult stages and lacks placental development like mammals, there are, however, limitations to its extension to vertebrate studies. It serves as a simpler model to study conserved genes and processes as well as to understand genetic and epigenetic control mechanisms of various processes. Although a large number of genes have fly orthologues, they might not be functionally similar. Flight muscles in flies, for example, have special flight adaptations not seen in human skeletal muscles.

Drosophila is a popular model to study myogenesis. The *Drosophila* embryonic body wall or somatic muscles as well as adult abdominal, flight and leg muscles present structural and functional similarities to vertebrate skeletal muscles (Taylor 2006; Piccirillo et al. 2014). It presents a simpler system to study conserved factors implicated in the acquisition of identity of individual muscles. Although a few key players in the acquisition of identity of muscle subsets are known in vertebrates, all the factors that distinguish individual muscles from their neighbors are yet to be identified.

1.4. Muscle development in *Drosophila melanogaster*

1.4.1. Gastrulation and the formation of the mesoderm

Similar to vertebrates, muscles are generated by the mesoderm germ layer in *Drosophila*. The embryo initially consists of a single layer of epithelial cells. Concentration gradients of maternal genes establish fate maps. The Dorsal (Dl) protein (an NF-kB homolog) gradient plays a crucial role in determining dorso-ventral polarity. It localizes to the nuclei only in ventral cells by Toll activated nuclear transport (Ganguly, Jiang, et Ip 2005) that is repressed by WntD in other regions. Cells with high nuclear Dl levels activate the expression of the bHLH transcriptional activator, Twist (Twi) that is also a myogenic factor. Twi autoregulates its own expression and

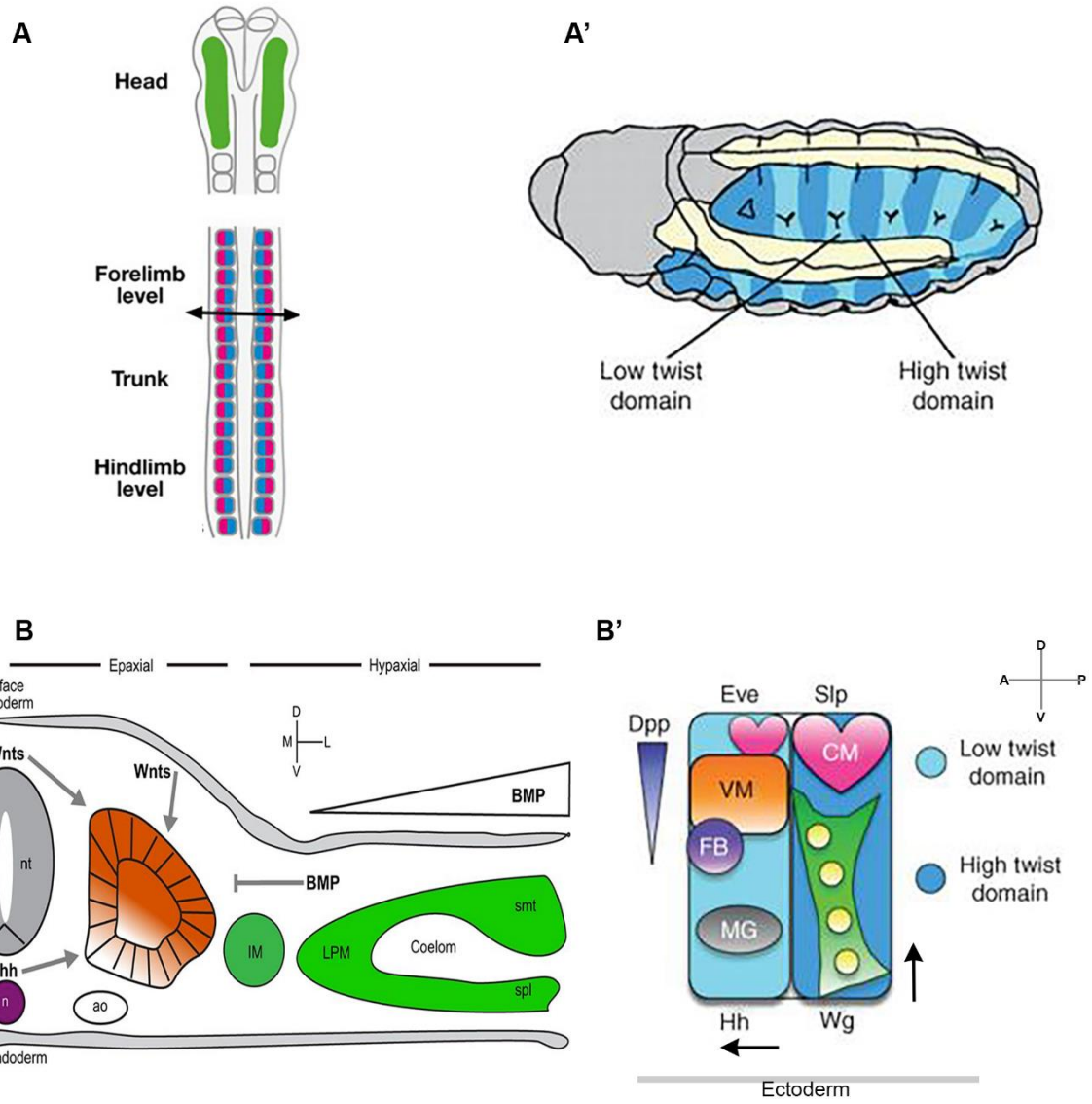


Figure 10. Comparison of mesoderm segmentation and specification in vertebrates and *Drosophila melanogaster*.

(A-A') Vertebrates undergo periodic segmentation of the paraxial mesoderm to generate somites (blue/pink) that give rise to skeletal muscles (A). *Drosophila* undergoes simultaneous segmentation of the mesoderm germ layer that gives rise to somatic or body wall muscles (A'). Each segment (blue) is divided into high and low Twist (Twi) domains. **(B-B')** In vertebrates, each somite has low BMP levels and the domain closest to the neural tube and surface ectoderm receiving high Wnt levels (dark brown) specifies the myogenic dermomyotome (B). In *Drosophila*, the somatic muscle domain is specified in the high Sloppy-paired (Slp) domain that has high Wg (a Wnt orthologue) signaling and activates *twi* (*Drosophila images adapted from Dobi, Schulman, et Baylies, 2015*).

activates the expression of mesodermal genes such as *Pox meso* (*Poxm*), *Actin57B* (*Act57B*) and *ß3-Tubulin* (M. Leptin 1991). A ChIP-on-chip study by Sandmann et al. also identified Twi binding sites in the enhancers of the fly homologue of *BMP2/4*, *Decapentaplegic* (*Dpp*) containing high affinity Dl binding sites that repress *Dpp*, although its significance is yet to be determined (Sandmann et al. 2007). Twi and Dl then coactivate the expression of the transcriptional repressor *snail* (*sna*) (A. C. Martin 2020). Twi and Sna targets include the FGF receptor *heartless* (*htl* or *DFR1*) and *Drosophila Mef2* (Cripps et al. 1998). Similar to vertebrate gastrulation, the first event during *Drosophila* gastrulation in the embryo is apical constriction that leads to the formation of a ventral furrow in the single layer of epithelial cells (Leptin et Grunewald, 1990.). This displaces surrounding cells and the prospective mesodermal cells undergo EMT and invaginate due to gene repression by Sna (Hemavathy et al. 2004). They then elongate with the overlying ectoderm along the posterior end in a process called germ band extension, at the end of which the mesoderm germ layer is formed under the ectoderm.

After the formation of the mesoderm, the embryo undergoes simultaneous segmentation, unlike the periodic somitogenesis in vertebrates. Each segment is demarcated by epidermal segmental grooves and is subsequently patterned by the segmentation genes *even-skipped* (*eve*) and *sloppy-paired* (*slp*) along with *Dpp, Hedgehog* (*Hh) and* one of the *Drosophila* Wnt genes, *wingless* (*wg)*. *twi* mRNA decays in *sloppy paired* (*slp)* mutants and reduces in *wg* mutants (Riechmann et al. 1997). The high Slp, Wg, Twi domain receiving high levels of Dpp dorsally gives rise to cardiac progenitors while somatic muscle progenitors arise below this domain (Figure 10).

1.4.2. The development of embryonic, larval and adult somatic muscles

For further details on *Drosophila* myogenesis, I refer the reader to the following review that provides an overview of myogenesis. In addition, it provides a comparison of vertebrate and *Drosophila* muscle components, highlights the structure and function of the highly conserved contractile units known as sarcomeres and presents an overview of muscle homeostasis:

cells

MDPI

Review

Genetic Control of Muscle Diversification and Homeostasis: Insights from *Drosophila*

Preethi Poovathumkadavil * and Krzysztof Jagla

Institute of Genetics Reproduction and Development, iGReD, INSERM U1103, CNRS UMR6293, University of Clermont Auvergne, 28 Place Henri Dunant, 63000 Clermont-Ferrand, France; christophe.jagla@uca.fr

* Correspondence: preethi.poovathumkadavil@uca.fr

Received: 30 May 2020; Accepted: 23 June 2020; Published: date: 25 June 2020

Abstract: In the fruit fly, *Drosophila melanogaster*, the larval somatic muscles or the adult thoracic flight and leg muscles are the major voluntary locomotory organs. They share several developmental and structural similarities with vertebrate skeletal muscles. To ensure appropriate activity levels for their functions such as hatching in the embryo, crawling in the larva, and jumping and flying in adult flies all muscle components need to be maintained in a functionally stable or homeostatic state despite constant strain. This requires that the muscles develop in a coordinated manner with appropriate connections to other cell types they communicate with. Various signaling pathways as well as extrinsic and intrinsic factors are known to play a role during *Drosophila* muscle development, diversification, and homeostasis. In this review, we discuss genetic control mechanisms of muscle contraction, development, and homeostasis with particular emphasis on the contractile unit of the muscle, the sarcomere.

Keywords: *Drosophila*; muscle; genetic control; muscle diversification; muscle homeostasis

1. Introduction

1.1. General Overview

Drosophila melanogaster, a holometabolic insect with a short lifespan, has served as a simple model to study myogenesis [1,2] and contractile proteins [3] for decades. Myogenesis in *Drosophila* occurs in two waves, one during the embryonic stage that gives rise to the larval body wall or somatic muscles and the second during pupal development that gives rise to adult flight, leg, and abdominal muscles [4]. All these muscles are voluntary, syncytial (multinucleate), and striated making them similar to vertebrate skeletal muscles [5]. Multiple signaling pathways, genes, and processes are conserved from *Drosophila* to vertebrates [6,7]. Muscles provide force to ensure various locomotory behaviors such as crawling, walking, jumping, and flying in *Drosophila*. Thus, they need to carry high levels of a mechanical load and are subject to constant strains, which can potentially disrupt homeostasis. Muscle movements need to be precise and coordinated, where communication with other tissues such as the nervous system provides critical inputs [8]. Muscles are the major reservoir for amino acids in the body that contribute to

muscle mass and protein homeostasis [9]. All muscle functionalities require that they are correctly formed in the first place to attain a homeostatic state in which they are physiologically active and stable. Muscle intrinsic signaling as well as signaling from external organs contribute to muscle homeostasis. Muscles display a high degree of plasticity or flexibility at the signaling, metabolic, myonuclear, mitochondrial, and stem cell levels.

This review is divided into three parts. The first part presents an overview of the mechanisms of muscle contraction in *Drosophila*. The second part focuses on the development of the larval and adult muscles. In the third part, we discuss the maintenance of muscle homeostasis in normal conditions and the adverse effects of the loss of this homeostasis in pathological conditions. Throughout the review, the focus is on sarcomeres, which are the basic contractile units of the muscle.

1.2. Major Structural Components of the Drosophila Muscle and Their Vertebrate Counterparts

In *Drosophila*, muscle function is coordinated by sensory, excitatory, and mechanical inputs by its connection to the nervous system via neuromuscular junctions and to the epidermis via myotendinous junctions akin to vertebrate systems though they present differences, some of which are outlined below.

1.2.1. Sarcomeres

Sarcomeres are the basic contractile units of the muscle and provide the force for contraction during movements (Figure 1). They are repetitively arranged in a regular pattern that gives a striated appearance under the microscope to vertebrate skeletal muscles as well as *Drosophila* somatic, flight, and leg muscles [10,11]. Sarcomeric length, functional domains, and many component proteins are conserved between invertebrates and vertebrates, although studies also point to interesting differences among species, which appear to be adaptations to individual muscle function [12–15]. Despite structural differences in *Drosophila* sarcomeric proteins in comparison to vertebrate counterparts, they have similar functional interactions and possess conserved functional domains; for example, the PEVK domain of the *Drosophila* titin, Sallimus (Sls) confers elasticity similar to vertebrates [16]. Thus, the sarcomere provides an example of nature reusing and repurposing components across evolution.

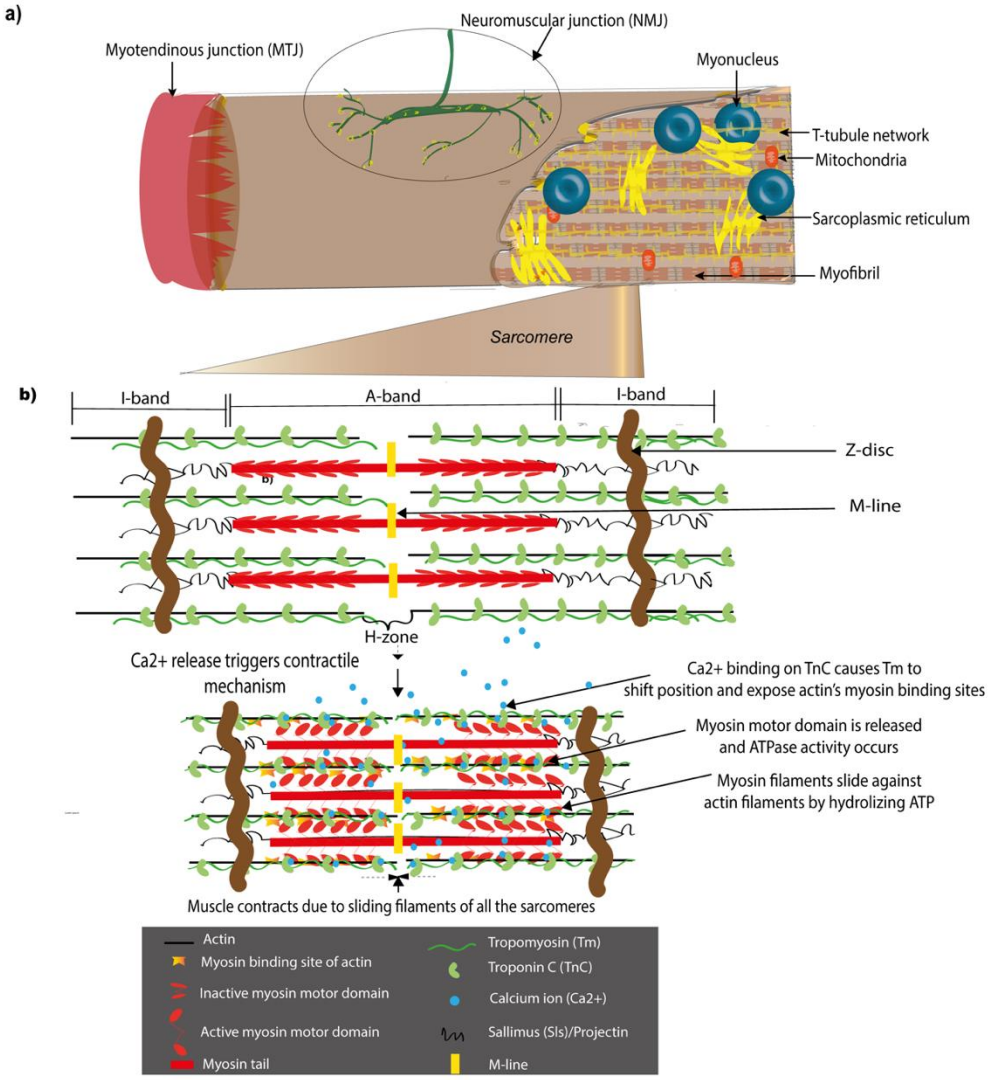


Figure 1. Schematic representation of the larval body wall or somatic muscle structure and the sliding filament theory of muscle contraction. (**a**) Muscle structure with myofibrils and the network of myonuclei, sarcoplasmic reticulum (SR), T-tubules, and mitochondria. The muscle is connected to the nervous system via the neuromuscular junction (NMJ) and to the epidermis via the myotendinous junction (MTJ). Myofibrils are formed of repetitive contractile units, the sarcomeres. (**b**) The structure of a sarcomere and the mechanism of contraction proposed by the sliding filament theory. Ca^{2+} ions released upon neurotransmitter signaling from the NMJ launch a cascade by binding to TroponinC (TnC) on the thin filaments of sarcomeres. This Ca^{2+} binding causes a conformation change in Tropomyosin (Tm) bound to actin, exposing actin's myosin binding sites. This permits the activated myosin motor domain to bind to actin and slide against it by utilizing the energy stored in Adenosine Triphosphate (ATP).

1.2.2. Myotendinous Junctions (MTJs)

In *Drosophila*, the MTJ is an attachment formed between the muscle and specialized groups of tendon-like cells of ectodermal origin called tendon cells, also known as apodemes (Figure 1a). Unlike vertebrates, *Drosophila* does not have an internal skeleton and tendon cells help anchor the muscles firmly to the cuticular exoskeleton instead, which helps transmit the contractile forces to the body to generate motion. This makes

them functionally similar to vertebrate tendons despite their distinct embryological origins, mesodermal for vertebrates and ectodermal for *Drosophila* [17,18]. The formation and maintenance of the MTJ is mediated through the ECM by specific integrin heterodimers on the muscle and tendon ends in *Drosophila* similar to vertebrates [19–22].

1.2.3. Neuromuscular Junctions (NMJs)

The NMJ is the point of contact between the motor neurons of the nervous system and the muscle, which enables environmental inputs to be transmitted via synapses to the muscle (Figure 1a). The *Drosophila* larval NMJ is an established model for NMJ formation and function. This NMJ is glutamatergic and responds to the neurotransmitter glutamate unlike vertebrate NMJs that are cholinergic and respond to acetylcholine. However, they are of particular interest owing to their similarity to mammalian brain glutamatergic synapses that express multiple genes orthologous to *Drosophila* genes and the ease with which NMJ assembly can be studied in this model [23–25]. It continues to be an active field of study with focus equally shifting to adult motor neurons formed after metamorphosis [26,27].

2. The Sarcomere and Molecular Mechanisms of Muscle Contraction

Voluntary muscle contraction is a highly coordinated process that depends on cooperative signaling from sensory neurons via interneurons and motor neurons to the NMJ of the muscle [28–30]. Given that the principal muscle function is to generate movements by contracting, the sarcomeric contractile units are indispensable for muscle function and their maintenance is crucial. The *Drosophila* adult indirect flight muscle (IFM) is established as a model to study sarcomere assembly and the functions of its components [31]. IFMs are built of multiple myofibers and have a stereotypic pattern of sarcomeric proteins forming highly ordered myofibrils similar to human skeletal muscles allowing the study of sarcomere malformations under mutant conditions. The IFM is also a model to study stretch activation (SA) [32]. During SA, there is a high frequency of contraction although the nervous system input frequency is much lower. This is possible due to the delayed increase in tension following muscle stretching. SA is a mechanism found in all muscles though it has particular significance in certain muscle types with rhythmic activity such as human cardiac muscles and the fruit fly flight muscles. In contrast to the multi-fiber IFM muscles of the adult, the somatic muscles in the *Drosophila* embryo and larvae are built of only one muscle fiber per muscle and present a much simpler model to study myofibers.

A sarcomere is a specialized structure adapted for muscle contraction (Figure 1). During myofibrillogenesis, newly formed sarcomeres align in repeating units along the length of a muscle to form a myofibril and multiple myofibrils covered by the plasma membrane form a myofiber. A sarcomere is built of thin-actin and thick-myosin filaments with associated proteins facilitating contraction-relaxation cycles. The thick filaments consist of myosin polymers with each myosin consisting of a myosin tail and two myosin heads, which are capable of attaching to actin during muscle contraction. The two ends of a sarcomere are demarcated by a Z-disc, a huge protein complex that anchors the thin filaments that form I-bands on either side of a sarcomere, while the thick filaments form

an A-band in the center (Figure 1). In between the two I-bands is an H-zone lacking myosin heads and in the center of the H-zone is an M-line that corresponds to another large protein complex that anchors the thick filaments [33].

Sarcomere function is intricately linked to other organelles such as the mitochondria [34], myonuclei [35], sarcoplasmic reticulum (SR), and T-tubules [10,36]. The efficient function of sarcomeres is closely coupled with the periodic arrangement of the SR and T-tubules around them [10,36–38]. T-tubules are regular tubular invaginations of the plasma membrane at each sarcomere. The membrane organelle SR is linked to the myonuclei and T-tubules to facilitate the exchange of proteins and ions. The SR is the major intracellular reservoir of calcium (Ca^{2+}) ions in the muscle, which are essential for muscle contraction. The T-tubule and SR form a specialized triad/dyad structure, which is indispensable for correct muscle functioning by excitation-contraction (EC) coupling. This EC coupling enables the transmission of excitation potentials from the NMJ to the SR, which triggers Ca^{2+} release from the SR that in turn initiates sarcomeric sliding movements leading to muscle contraction. Apart from Ca^{2+}, other ions contribute to muscle contraction [39]. The Na^+K^+-ATPase is a Na^+-K^+ pump that can pump Na^+ out of and K^+ into the cells against their normal concentration gradients. In muscles, the concentration of these ions fine-tunes the force of contraction [40]. In *Drosophila*, muscles are one of the major organs that express the Na^+K^+-ATPase α subunit [41]. One form of the Na^+K^+-ATPase β subunit, Nrv1 interacts with Dystroglycan (Dg), which is part of a complex that helps transmit forces into the muscle cell [42].

The mechanism of muscle contraction is explained by the sliding filament theory [43,44], reviewed by Hugh Huxley [33]. This theory proposes that the myosin head domain acts as a motor and slides against the actin filament powered by the energy stored in ATP. This sliding of the central myosin along the thin filaments causes the two I bands on either side to come closer to each other. During contraction, environmental inputs are transmitted by the nervous system to the NMJ leading to Ca^{2+} binding to the Troponin C (TnC) subunit of the Troponin (Tn) complex. This leads to the Troponin T (TnT) subunit that binds to the actin binding protein Tropomyosin (Tm) triggering a conformational change in Tm, thus shifting its position on actin and exposing the myosin binding site of actin [45–47]. Myosin that is turned 'on' by a myosin regulatory light chain (Rlc) phosphorylation [48] liberates the motor domains in the myosin head that were folded onto the myosin tail, thus facilitating its binding to actin. Subsequent ATP hydrolysis and energy release, thanks to its ATPase activity, permits it to move along the thin filament to contract the muscle. For the muscle to relax, the Troponin I (TnI) troponin subunit inhibits the actomyosin interaction [49] so that Tm covers the myosin binding site of actin and the myosin is switched 'off' and folded back onto the myosin tail [50,51]. This coordinated key muscle function highlights the importance of ionic and sarcomeric component homeostasis in muscles, which implies the supply and maintenance of the right quantities of the right ions and sarcomeric components at the right time to ensure muscle functionality.

During contraction, the MTJ helps anchor the myofibrils and transmits forces [19,52]. Tight interactions between sarcomeric components ensure myofibrillar integrity and prevent disintegration due to contractile forces. CapZ binds to the actin barbed end and links it to the Z-disc [13] while Z-disc proteins such as the filamin Cher [53], Zasp, and α-

actinin anchor the thin filaments [54]. Similarly, the M-line protein Obscurin that associates with the thick filament [55], Muscle LIM protein at 84B (Mlp84B) that cooperates with Sallimus (Sls) known as the *Drosophila* titin [56], integrins [57], and other proteins ensure muscle integrity. Sarcomeres are subject to constant mechanical strain due to the thin and thick filament friction and need to be consistently replenished to ensure their function over a lifetime. Since these muscles are voluntary, they also need to be able to stop contracting at will and go back to their natural state. Defective sarcomeric formation, maintenance, and homeostasis are associated with muscular diseases [15,58].

3. Muscle Diversification—On the Road to Muscle Homeostasis

Muscle development is a finely orchestrated, synchronized process that occurs in spatial and temporal coordination with the development of other communicating tissues to finally form a homeostatic muscle. There are similarities as well as differences between *Drosophila* and vertebrate myogenesis [59]. During development, each muscle diversifies to attain an identity tailored to its specific functional requirements. The study of muscle diversification during development is of interest in the context of homeostasis for two primary reasons:

1. Events similar to those occurring during development need to be reinitiated to repair and regenerate an injured muscle and reestablish muscle homeostasis [60]. This is a new field of study in *Drosophila* stemming from the recent discovery of muscle satellite cells in adult flies [61].
2. The two waves of myogenesis in *Drosophila* result in two homeostatic states, one in the larva and one in the adult. The larval homeostatic states are highly dynamic given the large growth spurt that occurs over the three larval instars. This might provide insights into mechanisms of muscle atrophy and hypertrophy. Forkhead box sub-group O (Foxo), for example, has been shown to inhibit larval muscle growth by repressing *diminutive* (*myc*) [62]. In mice, excess c-Myc has been shown to induce cardiac hypertrophy [63].

3.1. Embryonic Myogenesis of Larval Muscles

Embryonic myogenesis gives rise to monofiber larval somatic muscles whose main function is to aid in hatching and the peristaltic, crawling movements of the larvae. The embryonic and larval somatic musculature consists of a stereotypical pattern of muscles in each segment, with 30 muscles in most abdominal hemisegments (A2–A6) (figure in Table 1). There are fewer muscles in the posterior and first abdominal hemisegment and a slightly different set of muscles in the three thoracic hemisegments (T1–T3). Embryonic muscles arise from the mesoderm germ layer and their development requires intrinsic mesodermal cues and extrinsic cues from the adjacent epidermal and neural cells. Thus, they develop in synchrony with the development of muscle-interactors such as tendon cells and motor neurons and need to 'speak a common language' to communicate for coordinated development and maintenance.

Somatic muscle specification and differentiation have been reviewed extensively in the past [1,2,7,60–62] and this review presents complementary as well as new information that emphasizes the role of developmental factors in future muscle homeostasis.

3.1.1. Muscle Diversification by the Specification of Muscle Founder Cells Expressing Identity Transcription Factors (iTFs)

The embryo undergoes gastrulation by invagination [64], which brings the three germ layers, the ectoderm, the somatic muscle forming mesoderm, and endoderm in juxtaposition with each other. This helps provide extrinsic signals to the developing mesoderm. Following this juxtaposition, the mesoderm is divided into domains by morphogenic signaling [65] giving rise to a somatic muscle domain in which the transcription factor (TF) Twist (Twi) provides a myogenic switch [66]. Subsequently, equivalence or promuscular cell clusters expressing the neurogenic gene *lethal of scute* (*l'sc*) form and one muscle progenitor cell is singled out from each cluster by lateral inhibition involving Notch and Ras/MAPK signaling [67,68]. The remaining Notch activated cells in the equivalence groups become fusion competent myoblasts (FCMs). This process is reminiscent and coincides temporally with the specification of neural lineages from the neurectoderm [69,70], which occurs during embryonic stages 8–11, while muscle cell identity specification occurs during stages 9–11.

The singled-out muscle progenitors divide asymmetrically to give rise to founder cells (FCs), which are believed to carry all the information necessary to give rise to the diversity of muscle types. Asymmetric divisions of progenitors can give rise to two FCs, an FC and a Numb negative adult muscle precursor (AMP) or an FC and a cardiac progenitor, which subsequently migrate away from each other [67,71,72]. Each FC contains the information to establish one muscle's identity since it can form correct attachments and be correctly innervated even in the absence of myoblast fusion with surrounding FCMs [73,74]. It expresses its characteristic code of TFs known as muscle identity transcription factors (iTFs) (Figure 2). The expression of a combinatorial code of iTFs in distinct progenitors is the result of their spatial positioning as well as tissue specific convergence of multiple signaling cascades [75]. For example, Wg signaling from the adjacent developing central nervous system (CNS) is implicated in the specification of Slouch (Slou) positive FCs [76] highlighting the importance of coordinated tissue development.

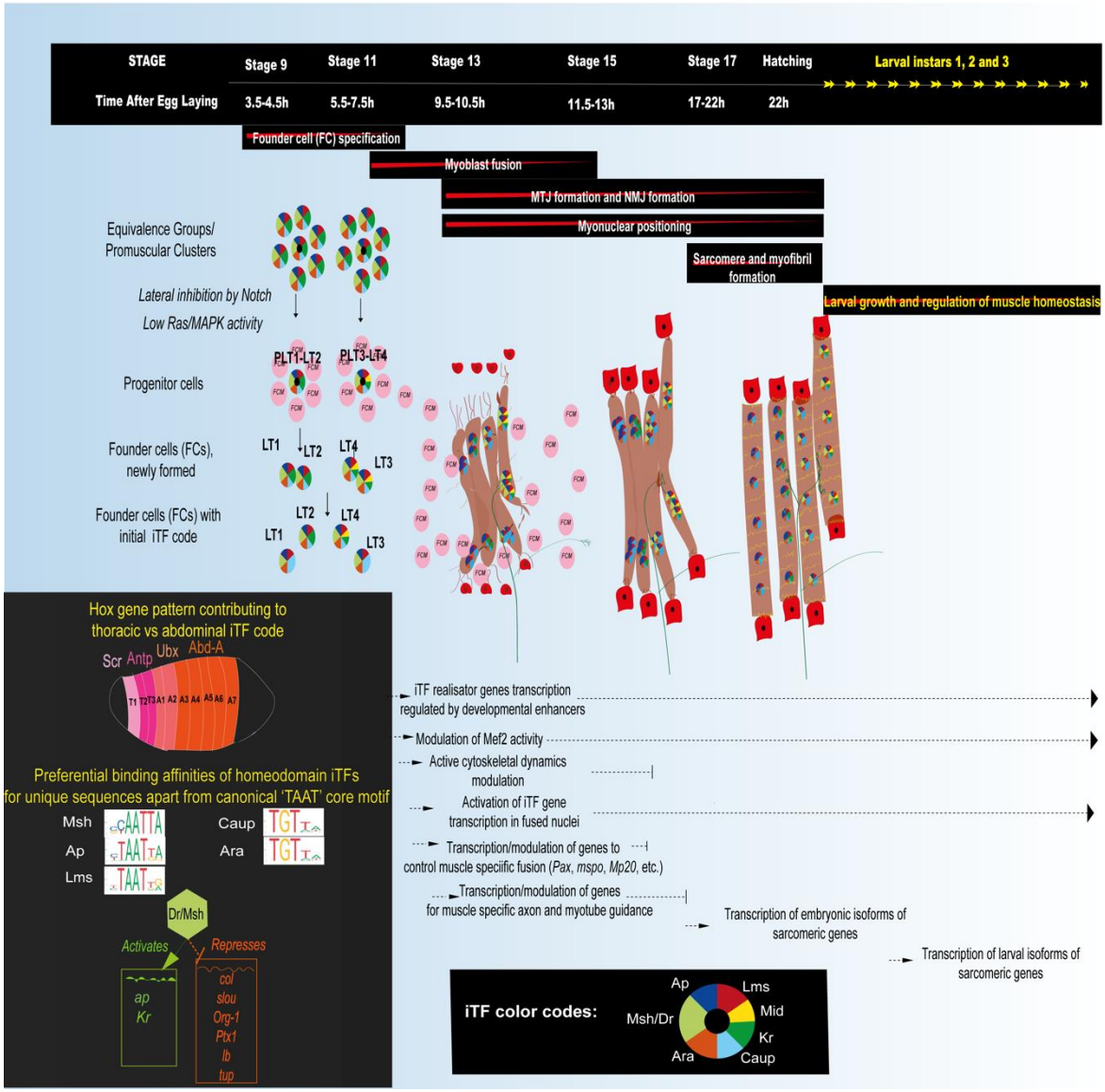


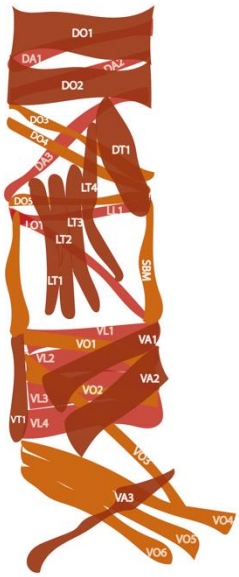
Figure 2. Spatial and temporal expression muscle identity transcription factors (iTFs) of the larval lateral transverse (LT) muscles. Sizes are not up to scale. Following the specification of progenitor cells by a lateral inhibition by Notch and low Ras/MAPK activity, founder cells (FCs) expressing muscle specific iTFs are specified for each LT muscle, LT1, LT2, LT3, and LT4 with a contribution from homeobox (Hox) genes to specify thoracic versus abdominal identities. Each iTF has preferential binding abilities to certain enhancers. The iTF expression is followed by the regulation of transcription and modulation of expression of their realisator genes which establish muscle identity over the course of development. The spatial and temporal expression of iTFs coupled with their modulation of realisator genes, which include generic muscle genes, in collaboration with Mef2 begs the question about their contribution to muscle homeostasis. Abbreviations: FCM: Fusion competent myoblasts; FC: Founder cells; LT: Lateral transverse muscles; iTF: Identity transcription factor.

3.1.2. The Role of iTFs

After the initial discovery of distinct Slou expressing FCs [77], many other TFs expressed in discrete subsets of FCs were subsequently identified and collectively named muscle identity transcription factors or iTFs (Table 1). A loss or gain of iTF function can cause muscle loss [78,79] or transformation of one muscle to another muscle fate [80,81]

and impede muscle development [82] thus disrupting muscle patterns. The iTFs such as Ap, Slou, Eve, Kr, Lb, and Lms are also expressed in the CNS [78]. Many identified iTFs such as Dr/Msh, Lms, Ap, Ara, Caup, Lb, Slou, Eve, Ptx1, and Tup are homeodomain TFs that are known to recognize similar canonical TAAT containing binding motifs, but they could have preferential high affinity binding motifs (Figure 2), as has been shown for Slou [83] and Caup [84]. The iTFs from other TF families are Twi, Nau, Kr, Kn/Col, Mid, Six4, Poxm, Org-1, and Vg. Newly identified iTFs for a subset of dorsal muscles are Sine occulis (So), No ocelli (Noc), and the cofactor ETS-domain lacking (Edl) [85], which act sequentially with their cofactors. The iTF Vg also acts with a cofactor, Sd [86].

Table 1. The iTF expression patterns in embryonic somatic muscle founder cells.

iTF	Human Orthologs	FCs Expressing iTF [1]	References	Embryonic Somatic Muscle Pattern
Apterous (Ap)	LHX	LT1, LT2, LT3, LT4, VA2, VA3	[78]	External muscles are represented in dark brown, intermediate muscles in a medium shade of brown, and internal muscles in fuchsia. 
Araucan (Ara)	IRX	LT1, LT2, LT3, LT4, SBM, DT1-*DO3*	[87]	
Caupolican (Caup)	IRX	LT1, LT2, LT3, LT4, SBM, DT1-*DO3*	[87]	
Collier (Col)/Knot (Kn)	EBF	*DA2*, DA3-*DO5*, *DT1-DO3*, *LL1-DO4*	[82,85,88]	
Drop (Dr)/Muscle segment homeobox (Msh)	MSX	DO1, DO2, LT1-LT2, LT3-LT4, VA2, VA3	[89,90]	
Even-skipped (Eve)	EVX	DA1, *DO2*	[91,92]	
Krüppel (Kr)	KLF	DA1, DO1, *LT1*-LT2, *LT3*-LT4, LL1, *VA1*-VA2, *DO2*, VL3, VO2, VO5	[87,92,93]	
Ladybird (Lb)	LBX	SBM	[94]	
Lateral muscles scarcer (Lms)	-	LT1-LT2, LT3-LT4	[95]	
Midline (Mid)	TBX20	*LT3*-LT4, LO1, *VA1*-VA2	[96]	
Nautilus (Nau)	MYOD	DO1, DA2, DA3-DO5, DO3, LL1-DO4, LO1, VA1	[79,85,88,97]	
Optomotor-blind-related-1 (Org-1)	TBX1	LO1, VT1, SBM	[98]	

Pox meso (Poxm)	PAX	DT1-*DO3*, VA1-VA2, VA3	[99]
Ptx1	PITX	Ventral muscles	[100]
Runt		DO2, VA3, VO4	[92,101]
Slouch (Slou)/S59	NKX1	DT1-*DO3*, *VA1*-VA2, VA3, VT1, *LO1*	[77,80,87]
Scalloped (Sd)	TEF-1	*All FCs transiently*, maintained in VL1, VL2, VL3, VL4	[86]
Vestigial (Vg)	VGLL	DA1-DA2, DA3, LL1, VL1, VL2, VL3, VL4	[86]
Tailup (Tup)	ISL	DA1, DA2, DO1, DO2	[81]
Eyes absent (Eya)		Differential temporal expression in multiple FCs	[85,102]
Six4	SIX	Differential temporal expression in multiple FCs	[102,103]
Sine occulis (So)	SIX	*DA2*, DA3-DO5, *LL1*-*DO4*	[85]
No ocelli (Noc)	ZNF	DA3-DO5	[85]
ETS-domain lacking (Edl)	-	*DA2*, DA3	[85]

[1] In the 'FCs Expressing iTF' column, each FC name is shown in the colour corresponding to the muscle it generates as depicted in the figure in the column on the extreme right. FCs known to be generated from an asymmetric division of the same progenitor cell are hyphenated. FCs with transient expression are shown in italics.

The iTF code can be hierarchic and activate other iTFs as has been shown for Org-1 that activates Slou and Lb [98]. In addition to hierarchy, there seems to be isoform specificity in iTF expression [85]. Certain iTFs confer identity by repressing other iTFs. Dr, for example, represses Lb that is normally active only in the SBM muscle and Eve that is normally continually expressed in only the DA1 muscle [104], while Tup represses Col in DA2 [81]. The expression levels of one isoform of the chromatin remodeling factor Sin3A is implicated in modulating the response to iTFs by acting on the *slou* enhancer [105].

The same iTFs can be expressed in different muscles, but with different co-iTFs. The identity code for one specific muscle subset, the lateral transverse or LT muscles, which comprises the four muscles LT1-4, for example, is known to be set up by a combinatorial expression of Dr [89], Ap [78], Kr [93], Lms [95], and the Ara/Caup complex [87] (Figure 2). Dr appears to directly or indirectly activate the transcription of many LT iTFs such as *Kr*, *ap,* and itself while repressing non-LT iTFs such as *col*, *slou*, *Org-1*, *Ptx1*, *lb,* and *tup* [83] and its expression is lost by mid embryonic stages. Only Lms is specific to all four LT muscles while others are also expressed in other muscle subsets, although not in

combination with the same co-iTFs. This seems to be the way the iTF code is set up, where they are repurposed in different combinations to define the identity of different muscles [81,85,88]. Even amongst the LT muscles, each muscle has a specific combination of these iTFs (Figure 2). Some iTFs such as Lms are persistently expressed while others such as Ap and Kr are transient. A characteristic feature of Kr is that it is transiently expressed and subsequently lost from one of the two sibling FCs that arise from the progenitor that expresses it [87,93].

The thoracic LT muscles have slightly different characteristics such as a different number of myonuclei, which might depend on the iTF code along with individual iTF dynamics in conjunction with other TFs [106]. Homeotic or Hox genes such as *Antennapedia* (*Antp), Abdominal-A (Abd-A), Abdominal-B (Abd-B), and Ultrabithorax (Ubx)* control the muscle pattern along the anterior-posterior axis and are thus part of the iTF code [106–108] (Figure 2). The mechanism of Hox gene regulation in muscles could be by repressing genes specifying alternative fates by altering the epigenetic landscape in a tissue specific manner [109]. Hox genes could also be involved in the coordination of the proper innervation of muscles [110].

Once an FC initiates its diversification with a specific identity determined by an iTF code, it starts differentiating by activating realisator genes acting downstream of iTFs. Some muscle identity realisator genes have been identified. They include several muscular differentiation genes such as *sallimus* (*sls*), *Paxillin* (*Pax*), *Muscle protein 20* (*Mp20*), and *M-spondin* (*mspo*), which are differentially expressed in muscle subsets to control the acquisition of specific muscle properties such as the number of myoblast fusion events or the specific attachment to tendon cells. [111,112]. Thus, an iTF code and a downstream realisator gene code are both essential to generate a diversity of muscle types with specific functions and to set the foundation for muscle homeostasis.

3.1.3. Mef2, a Key Muscle Differentiation Factor and Its Interactions with iTFs

The *Drosophila* Myocyte enhancer factor 2 (Mef2) acts along with iTFs and their realisator genes to cause the muscle to differentiate. Mef2, similar to its vertebrate ortholog MEF2, is indispensable for muscle differentiation [113–115]. It has an equally important role in fully differentiated muscles and the control of its expression and activity is dynamic. Though it is expressed in the mesoderm during all stages, its loss of function does not prevent initial muscle specification and FC generation, but completely blocks subsequent differentiation so that muscle cells undergo apoptosis in later embryonic stages [116,117]. Mef2 activity levels change over time and appear to be adapted to varying target gene expression requirements during different developmental stages [86,118]. It regulates a vast array of muscle specific genes [119,120], sometimes in cooperation with other TFs such as Cf2 [121,122]. It is itself regulated by various mechanisms including autoregulation [123], signal and TF integration at its specific cis regulatory modules (CRMs) [124], or post transcriptionally by highly conserved miRNAs such as miR-92b [125]. TFs such as Twi and Lameduck (Lmd) [119,126] acting on muscle specific CRMs and Akirin-bearing chromatin remodeling complexes [127] are known regulators of Mef2 transcriptional activity. The RNA modifying enzyme Ten-eleven-translocation family protein (Tet) shows a strong overlap with Mef2 expression in somatic muscles and its

depletion in muscle precursors leads to larval locomotion defects [128] though the relationship between the two factors is unclear.

The iTFs Vg and Sd physically interact with each other and with Mef2 either alone or in combination [129]. Each of them has a spatially and temporally controlled expression pattern and altering their expression levels severely affects the development of specific ventral muscles during late stages by affecting the levels of realisator genes. Thus, iTFs could play a key role in the modulation of Mef2 interactions. Given the central role of Mef2 in muscle development, disrupted Mef2 expression can have deleterious consequences at all stages of muscle development and maintenance.

3.1.4. Myoblast Fusion and Myonuclear Positioning

In order to form a differentiated muscle, in the mid-stage embryo, a specific number of neighboring FCMs fuse with the FC to form a syncytium (Figure 2). The formation of syncytial fibers by myoblast fusion is complete by the end of stage 15 [130–132]. As fusion proceeds, the round-shaped FC becomes a myotube that elongates, becomes polarized, and locally sends out filopodia in the presumptive area of MTJ and NMJ formation. Fusion involves complementary cell adhesion molecules (CAMs) such as Dumbfounded (Duf) or its paralogue Roughest (Rst) expressed on FCs [133,134] and Stick and Stones (Sns) or its paralogue Hibris (Hbs) expressed on FCMs [135,136], respectively. They trigger a signaling cascade, thereby modulating cytoskeleton dynamics to form a fusogenic synapse that helps integrate the FCM nucleus into the FC/myotube. In *Drosophila*, the iTF code dictates the number of fusion events by controlling the expression level of fusion genes encoding actin cytoskeleton modulators such as *Muscle Protein 20* (*Mp20*) and *Paxillin* (*Pax*) or the ECM component *m-spondin* (*mspo*) [112]. A recent study provides insights into the FCM-FC transcription dynamics in a syncytial myotube [137]. FCMs appear to be naïve and respond to the local environment that recruits them for fusion. Upon fusion, the FCM adopts an FC transcriptional program triggered by the transcription of certain muscle specific iTFs. However, once fusion is complete, differences in gene transcription among myonuclei within the same muscle are observed. For example, not all myonuclei transcribe the iTFs at a given timepoint, which could help maintain an mRNA-protein balance. Evidence from this study suggests that even after fusion is complete the FC nucleus that seeded the muscle retains a transcriptional program that is distinct from other myonuclei.

As fusion proceeds, at around stage 14, the nuclei of newly fused FCMs start exhibiting characteristic movements until they are positioned peripherally to maximize the internuclear distances. This process, also observed in vertebrate muscles [138,139], has been extensively studied in the LT muscles in the *Drosophila* embryo. In these muscles the new myonuclei initially cluster into two groups, unlike in vertebrates where nuclei cluster in the center of the myotube [140], then disperse and are finally arranged along the periphery of the myotube. Correct myonuclear positioning is dependent on the LINC complex [141] that links the inner nuclear membrane (INM) to the outer nuclear membrane (ONM) and the ONM to the microtubules (MT) and the actin cytoskeleton. Mispositioned myonuclei in *Drosophila* larvae cause locomotion defects, and in humans, are associated with various diseases [142]. This is not surprising considering the close association of myonuclei with muscle structural components such as the NMJ, MTJ, actin

cytoskeleton, microtubule, SR, Golgi complex, and T-tubules. During the larval growth spurt following hatching, myonuclei increase in size along with the increasing muscle size by Myc dependent endoreplication to adapt transcription to muscle functionality requirements [62].

3.1.5. Myotendinous Junction (MTJ) Formation

MTJ formation has been previously reviewed [17,143–145]. Once FCs are specified, they migrate towards the ectoderm while tendon precursor cells are specified in the ectoderm in parallel in a muscle independent fashion by the induction of expression of the early growth response factor (Egr)-like zinc finger TF, Stripe (Sr). Interestingly, tendon progenitor cells in mice express the Sr orthologs, the early growth response TFs EGR1 and EGR2 [146]. The StripeB (SrB) isoform is induced during the precursor stage to maintain the tendon cells in a non-differentiated state until later when they differentiate following signals from the approaching muscles. These signals lead to an increase in the expression of the StripeA (SrA) isoform in an integrin dependent manner by promoting *stripe* splicing by the short isoform How(S) of the splice factor interactor How [147]. SrA induces the expression of tendon differentiation markers such as *short stop* (*shot*), *delilah* (*dei*), and *β1-tubulin* (*β1-tub*). At stage 14, tendon cells guide myotubes to their final attachment sites. The targeting of muscles to tendon cells at stage 15 is facilitated by muscle type dependent and generic CAMs as well as signaling molecules. These include Slit-Robo [148] in some ventral muscles, Derailed (Drl) [149] in LT muscles, Kon-tiki (Kon), Glutamate receptor interacting protein (Grip), and Echinoid (Ed), probably involving integrin complexes [149–152]. Once muscles target their tendon cells, integrin complexes assemble on the muscle and tendon cells facilitated by the αPS2-βPS integrin heterodimer on the muscle end and αPS1-βPS on the tendon cell end to stabilize the attachments [153]. Each attachment site is muscle type specific and the iTF code could potentially modulate the expression of genes such as *kon* [137].

MTJ formation is complete by the end of stage 16 and is then further refined to withstand contractile forces. Talin phosphorylation contributes to MTJ refinement [154]. This is followed by myofibril maturation and attachment to the MTJ. Once muscles start contraction, mechanical forces stabilize the MTJ by reducing integrin turnover [155]. The MTJ grows along with the massive larval growth spurt following hatching.

3.1.6. Sarcomere Assembly and Myofibrillogenesis

Sarcomere assembly has been extensively studied in the *Drosophila* indirect flight muscles (IFM) [31] and other invertebrate models as well as in vertebrate models and cultured human cells [156,157]. These studies point to similarities as well as differences in vertebrate and insect muscles. The premyofibril theory that is widely accepted for vertebrate sarcomere assembly proposes the formation of premyofibrils along the cell periphery containing non muscle myosin, which then incorporate muscle myosin to form nascent myofibrils that subsequently form mature myofibrils [158,159]. In the early stages, distinct Mhc positive fibrils and I-Z-I complexes containing thin filaments protruding from α-actinin positive central Z bodies are seen in invertebrates as well as vertebrates [160,161]. In *Drosophila*, it has been proposed that the individual components of the

sarcomere are assembled separately as latent complexes and are then assembled into sarcomeres without assembling into premyofibrils [162–164]. Most studies on *Drosophila* sarcomere assembly have been using the IFM as a model and not much attention has been given to embryonic sarcomere development.

In the *Drosophila* embryo, sarcomere assembly is initiated at stage 17. Individual sarcomere constituents are first assembled and then integrated into a mature sarcomere by integrin dependent interdigitation [162]. The precise stage at which each sarcomere component is added is currently not known. Certain sarcomeric proteins such as actin [165] and myosin [166] express sarcomere specific as well as generic cytoplasmic isoforms with roles in other muscle components such as the MTJ. TFs such as Mef2, Chorion factor 2 (Cf2), and E2F transcription factor 1 (E2f1) have been shown to regulate the expression of sarcomeric genes [122,167]. *Drosophila* has six actin isoforms including Act57B and Act87E that are muscle specific and incorporate into larval sarcomeres [165]. Thin filament formation and elongation requires actin binding factors such as the *Drosophila* formin Dishevelled Associated Activator of Morphogenesis (DAAM) [168] and Sarcomere length short (Sals) [169], which localize to the growing thin filament pointed ends. Once the thin actin filament attains its final length, it is capped by a short embryonic isoform of Tmod [169,170]. While non-muscle myosin is a component of premyofibrils in vertebrates, this does not seem to be the case in *Drosophila*, which has only one non-muscle myosin, Zipper (Zip). During stage 16, it colocalizes with PS2 integrin at muscle attachment sites and at stage 17 when sarcomeres form, it also colocalizes to Z-discs and is essential for myofibril formation [162,171]. PS2 integrin follows a similar expression pattern in culture with initial occurrence at contact sites, then at Z-discs [172]. The observation of Zip association with PS2 before sarcomere assembly is significant because myofibrils attach to the MTJ via integrin complexes with Zip acting downstream of PS2 signaling [52]. Therefore, it would appear that the embryo is getting individual components ready for future integration into myofibrils.

By stage 17 of embryogenesis, several sarcomere proteins localize to Z-discs and thin and thick filament organization and myofibril structures are seen. A knockdown of Z-disc proteins Zip, Zasp, and α-actinin at this stage disrupts sarcomerogenesis [162], though Zasp mutant sarcomeres disintegrate after initial correct formation. The myoblast fusion protein Rolling pebbles (Rols7) also colocalizes to the Z-discs during sarcomerogenesis [173], but its function remains to be elucidated. Integrins are distributed along the width of the muscle and align with Z-discs during embryonic sarcomerogenesis. Their loss results in clumping, where I-Z-I body components stay distinct from Mhc containing components. In addition, integrins associate with the ECM and mutant larvae for the ECM type IV collagen Col4a1 present abnormalities in thin-thick filament interdigitation and the degeneration of body wall muscles [162,172,174]. They are also present at epidermal muscle attachment sites along with several Z-disc proteins. Mature sarcomeres align themselves to form myofibrils that attach to the MTJ via the terminal Z-disc to be able to sustain muscle contractions [162,173].

Auld and Folker showed that myonuclear movements are intricately linked to sarcomere and myofibril formation [35]. Their study showed that the Z-disc protein Zasp66, one of the *Drosophila* Zasp family of proteins, localizes as puncta to the cytoplasmic face of the nuclei along with F-actin during initial stages of sarcomerogenesis.

At later stages, puncta were observed throughout the muscle. They showed that LINC complex components such as Klarsicht (Klar) and Klaroid (Koi) coordinate initial colocalization of puncta around the nucleus. However, Z-disc-like structures still formed and aligned into myofibrils in LINC component depleted muscles, although they had altered morphology suggesting a specialized role for myonuclei-associated Zasp66 puncta. *sals* mutants display clustered myonuclei at muscle ends as well as myofibrils with numerous shorter sarcomeres suggesting a role for correct myonuclear positioning in myofibril organization [169].

Embryonic myofibrillogenesis within the egg is complete by late stage 17. Asynchronous, episodic contractions occur during the process of myofibril assembly, but coordinated contractions only occur later after mature NMJ formation results in adequate motor inputs [175,176]. Following hatching, during larval stages when the muscles rapidly grow in size, new sarcomeres are generated and organized into myofibrils during an approximately five-day period [62]. In fully mature larval muscles, T-tubules and the SR organize themselves around each sarcomere for excitation-contraction coupling and this organization is Amphiphysin (Amph) dependent [37]. The iTF code could play a role in modulating the muscle specific expression of sarcomeric genes, as has been shown for Vg and Sd that form a complex with Mef2 to modulate Mef2 targets involved in sarcomerogenesis including *Act57B* and *Mhc* [129].

3.1.7. Innervation and Neuromuscular Junction (NMJ) Formation

The development of the NMJ of larval somatic muscles has been previously reviewed [24,25,177–179] and represents another example of intricate communication between two different tissues. After neuroblasts differentiate into motor neurons (MNs) in parallel with FC specification [180–182], their dendrites in the CNS are organized in a 'myotopic map' reflecting the innervation pattern of their target muscles and MNs can reach target locations even in the absence of muscles [182]. Each neuroblast expresses a characteristic code of TFs that defines its identity as is the case for muscle FCs expressing iTFs [26]. By stage 12, MN axons fasciculate in each hemisegment within three peripheral nerves, the intersegmental nerve (ISN), segmental nerve (SN), and transverse nerve (TN) that extend towards specific target muscles from the ventral nerve chord (VNC). The ISN, SN, and TN branch stereotypically as they extend growth cones towards muscles to form MN nerve branches that further defasciculate into axons to innervate muscles. The SN nerve, for example, branches into SNa, SNb, SNc, and SNd with a subset of MNs from the SNa innervating a muscle subset including LT muscles [181].

At around stage 14, each MN extends numerous filopodia from axon growth cones towards muscles to explore their target muscles. Muscles in turn extend myopodia that cluster together on axon growth cone arrival and intermingle with growth cone filopodia. Muscles also form lamellipodia during innervation [183]. Target muscle recognition and contact are facilitated by muscle and MN specific CAMs, Cell Surface and Secreted (CSS) proteins, and other proteins [184]. Certain guidance molecules such as the homophilic Connectin (Con) are expressed in the SNa MN as well as the LT muscles it innervates [185]. Con is also expressed in the DT1 muscle and its expression is potentially modulated by the iTF code [137]. Some MNs and the muscles they innervate express the same iTF, as is

the case for the Eve expressing DA1 muscle and its innervating aCC MN in the ISNb [91,181,182,186]. Eve indirectly modulates the MN expression of the Netrin repulsive presynaptic receptor Unc-5 in the ISNb [187] to guide MN axons. Upon MN contact, muscles start to accumulate Glutamate Receptor (GluR) at synaptic zones mediated by Disks large (Dlg) to form primitive synapses in an innervation dependent fashion [188,189]. By the end of stage 17, non-target synapses are pruned and mature synapses form, which exhibit a stereotyped morphology of boutons with active zones for vesicle release on the presynaptic end and novel synthesis and clustering of more GluR on the postsynaptic end [190]. Once NMJ formation is complete, muscles are ready to contract in a coordinated manner.

During larval stages, some MNs are remodeled and this is reflected in the larval CNS myotopic map [191]. Until the third larval instar, the NMJ grows by arborization and addition of boutons, a process that requires the gene *miles to go* (*mtgo*), which is an ortholog of mammalian *FNDC3* genes [192], and integrins [193]. There is also an activity dependent refinement of the synapse mediated by Ca^{2+} [194]. Tenurins, a conserved family of transmembrane proteins enriched in the vertebrate brain that possesses glutamatergic synapses are implicated in *Drosophila* axon guidance as well as synaptic organization and signaling with muscle specific expression [195].

As muscles form, abdominal adult muscle precursors (AMPs) arrange themselves in niches between specific peripheral nerves and muscles. They form an interconnected network connecting to each other and to the peripheral nerves by extending filopodia [196–198]. All embryonic muscle development processes finally lead to the formation of functional larval body wall muscles that closely communicate with the epidermis via the MTJ and with the nervous system via the NMJ to ensure larval locomotion.

3.2. Pupal Myogenesis of Adult Muscles

Adult muscles are generated during a second wave of myogenesis during pupal metamorphosis where most larval muscles are histolyzed. Metamorphosis marks the end of larval muscle homeostatic states. Adult flies have a pair of wings in the thoracic segment T2 and three pairs of legs in thoracic segments T1–T3, which are powered by specialized thoracic flight and appendicular muscles, respectively (figure in Table 2). Adult myogenesis has been reviewed recently in [199,200]. All adult thoracic muscles including the indirect flight muscles (IFMs), direct flight muscles (DFMs), and leg muscles possess a multi-fiber structure similar to vertebrate skeletal muscles. However, unlike heterogenous mammalian skeletal muscles with one muscle composed of slow and fast fiber types, each *Drosophila* muscle appears to have a single fiber type. IFMs are constituted of the dorsoventral muscles (DVMs) and the dorsal longitudinal muscles (DLMs), which facilitate upward and downward wing strokes respectively during flight. The muscle fibers that build the IFM and leg muscles differ in organization and morphology to adapt to different functionalities. The IFMs are fibrillar, asynchronous muscles while the tergal depressor of the trochanter (TDT) or leg jump muscles and DFM are tubular, synchronous muscles [201–203]. Similar to mammals, individual fiber types in the adult fly differ in component constitution such as expressing specific myosin heavy chain isoforms

[203,204]. The generation of adult muscles is initiated by a series of coordinated processes again requiring close communication between tissues.

3.2.1. Myoblast Pool Generation by Adult Muscle Precursors (AMPs) during Larval Stages

Embryonic myogenesis sets the foundation for adult muscle development since the asymmetric divisions of embryonic muscle progenitor cells give rise to adult muscle precursors (AMPs) in addition to the embryonic muscle FCs. AMPs are Notch positive, Numb negative cells that remain quiescent with persistent Twist (Twi) expression until initial pupal stages when they get reactivated [205] and contribute to adult muscle development. Abdominal AMPs are closely associated with the larval muscles and with the peripheral nervous system (PNS) enabling crosstalk and providing positional cues to the AMPs that give rise to adult abdominal muscles [196–198]. In the thoracic segments, AMPs associate with wing and leg imaginal discs, which are epidermal cell clusters set aside in the embryo and larva and act as precursors for the future generation of adult wings and legs, respectively [206]. During the first and second instar larval stages, these AMPs undergo symmetric divisions giving rise to an imaginal disc associated monolayer of Twi and Notch positive adepithelial cells. In the abdominal segments, they proliferate while remaining associated with their muscle fibers similar to vertebrate satellite cells [207,208]. During the third larval instar, due to the activation of Wg signaling from the imaginal discs, AMPs undergo asymmetric divisions forming one stem cell and one Numb positive post-mitotic myoblast where Notch signaling is inhibited [209,210]. Thus, a large pool of myoblasts is primed for metamorphosis.

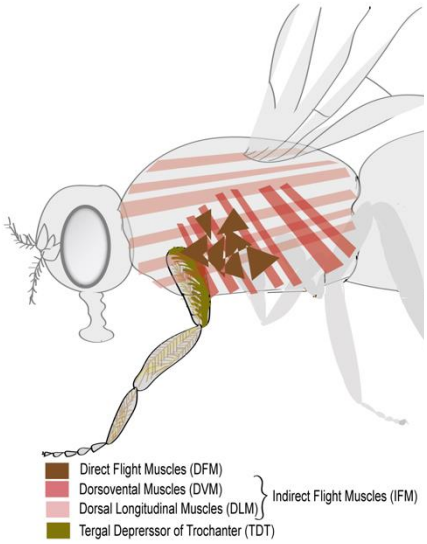
The myoblasts primed to form IFM express high levels of Vestigial (Vg), which represses Notch and promotes IFM differentiation [211], and low levels of the TF Cut (Ct) while DFM myoblasts express high Ct levels, with the levels being governed extrinsically by the ectoderm [212]. DFM myoblasts also express Lms [95]. The myoblasts associated with the leg imaginal disc on the other hand express Ladybird (Lb) similar to vertebrate limb bud myoblasts that express the Lb orthologue LBX1 [213,214], which represses Vg. Mutant *vg*, *ct*, *lms*, and *lb* flies have severely disrupted muscle pattern or function and they thus contribute to the adult muscle iTF code (Table 2). Vg, Lms, and Lb also act as embryonic somatic muscle iTFs expressed in a subset of embryonic muscles [86,94,95] (Table 1). Among embryonic myogenic factors, it was noticed that Apterous (Ap) expression defines all prospective flight muscle epidermal muscle attachment sites in the wing disc [215]. Similar to embryonic stages, Duf positive adult FC specification takes place by the third larval instar, but in contrast to embryos it is driven by Heartless (Htl) mediated Fibroblast growth factor (Fgf) signaling and Hox genes [212,214,216,217].

Table 2. The iTF expression patterns in myoblasts of adult muscles.

Adult iTF	Human Orthologs	Adult Myoblast Expression	Embryonic iTF Function [1]	References	Adult Flight and Leg Muscle Pattern
Vestigial (Vg)	VGLL	IFM	DA1-DA2, DA3, LL1, VL1, VL2, VL3, VL4	[211]	Indirect flight muscles (IFM) are shown in shades of red and the direct flight muscles (DFM) in dark brown. Among the leg muscles, only the

Extradenticle (Exd)	PBX	IFM		[218]
Homeothorax (Hth)	MEIS	IFM		[218]
Spalt major (Salm)	SALL	IFM		[219]
Erect wing (Ewg)	NRF1	IFM		[220]
Cut (Ct)		DFM		[213,214]
Lateral muscles scarcer (Lms)	--	DFM	LT1-LT2, LT3-LT4	[95]
Apterous (Ap)	LHX	DFM	LT1, LT2, LT3, LT4, VA2, VA3	[215]
Ladybird (Lb)	LBX	Leg muscles	SBM	[214]

tergal depressor of trochanter (TDT) muscles are highlighted in olive green. Other leg muscles are in a light shade of green.



[1] In the 'Adult myoblast expression' column, names are shown in the colour corresponding to the muscles they generate as depicted in the figure in the column on the extreme right. Embryonic FCs known to be generated from an asymmetric division of the same progenitor cell are hyphenated.

3.2.2. Histolysis of Larval Muscles, Adult iTF Code Refinement, and the Contribution of AMPs

During pupal stages, most of the larval muscles are histolyzed in the thoracic as well as abdominal hemisegments [221,222]. Myoblasts generated from AMPs either fuse with non-histolyzed larval muscle scaffolds to which they associate or give rise to adult muscles *de novo* [221]. In the T2 mesothoracic segment, three larval dorsal oblique muscles, DO1, DO2, and DO3 escape histolysis and serve as templates for the formation of the DLMs while the DVMs and leg muscles are generated de novo. At the end of the third larval instar, the myoblasts start expressing the muscle differentiation factor Mef2 in an ecdysone dependent manner [123]. As with embryonic myogenesis, adult muscle formation is seeded by FCs, with the number of FCs generated corresponding to the number of muscles they will seed [217]. The DLMs are an exception where the three remnant larval muscles serve as FCs and express the marker Duf. Nevertheless, if the larval muscles giving rise to adult DLMs are ablated they still form muscles *de novo* by an innervation dependent process, although with aberrations [223].

During early pupal stages myoblasts start migrating. MNs play a significant role in initial adult myogenesis by regulating myoblast proliferation during the second larval instar and subsequent myoblast migration during pupal stages. In denervated flies, DVM muscle formation is severely compromised and it leads to the reduction in DLM size when using larval muscles as templates whereas if larval templates are ablated, their *de novo* formation is abolished [223]. In the abdominal segments, myoblasts migrate and associate with nerves to form adult muscles [224]. In the thoracic segments, the wing and leg discs evaginate and myoblasts migrate along them to reach their destinations where adult

muscles are generated. The myoblasts either fuse with FCs or with larval templates using a similar machinery to embryonic myoblast fusion to form fully differentiated adult muscles by 36 h after puparium formation (APF). Muscles extend as they fuse and attach to the MTJ on either end [221,225,226].

Apart from Vg, Ct, and Lb that act as adult muscle iTFs (Table 2) to confer myoblast identity in the imaginal discs during larval stages, the expression of the embryonic iTF Ap is initiated during pupal stages in myoblasts that will give rise to the DFM but not IFM in addition to epidermal attachment sites [215]. Unlike the embryonic FCs, it is expressed in adult FCMs instead of adult FCs, but similar to the embryonic FCs they contribute to the same muscle's iTF code along with Lms. This hints at specific muscle patterning information derived from these iTFs. Ap is necessary for the correct formation of DFMs and continues to be expressed in fully formed DFMs. It is also necessary for IFM attachment by regulating Stripe (Sr) expression which, similar to the embryo, is essential for adult muscle attachment. In *lms* mutants, the wing disc Vg domain is expanded and although muscles seem normal, the adult wings exhibit a held-out phenotype suggesting contraction abnormalities [95]. As fusion begins, the IFM FCs also express the adult iTFs Extradenticle (Exd), Homeothorax (Hth), and Spalt major (Salm), which genetically interact to specify a fibrillar versus tubular fate by regulating fiber specific gene expression and splicing regulated by Arrest (Aret) [201,218,219]. The iTF Erect wing (Ewg) also significantly contributes to IFM identity [220].

3.2.3. MTJ Formation

The wing and leg imaginal discs generate Sr positive tendon-like precursor cell clusters starting from the third larval instar until the beginning of pupation. Sr expression is initiated by Notch signaling [227,228]. Leg muscles attach to the internal tendons on one end and the tendon cells in the exoskeleton on the other end. At about 3 h APF, the leg disc Sr positive tendon precursor cells invaginate into an evaginating leg disc and are closely associated with myoblasts that give rise to leg muscles [18]. Disrupting tendon precursors also disrupts myoblast localization. The epidermal tendon precursor cells' shape changes to form tubular structures during invagination giving rise to internal tendons to which each leg muscle attaches on one end with tendon specificity. DLM muscles that form by the splitting of remnant larval muscles extend filopodia on either end as they grow and split. Still in the process of splitting, their filopodia interdigitate with those of their target tendon cells and initiate MTJ formation that requires Kon, integrins, Tsp, and Talin similar to embryos. DLM filopodia disappear after a mature MTJ forms by 30 h APF and tendon cells elongate due to tension [163]. In the abdomen, MTJ maturation follows a similar process but is complete only by 40 h APF [229].

3.2.4. Sarcomere Assembly

Similar to embryos, premyofibrils are absent in DLM muscles. Mhc positive complexes are observed throughout the muscle by 26 h APF and assemble rapidly and synchronously across the entire muscle into myofibrils at 30 h APF immediately following tension generated by MTJ maturation [163,230]. This myofibril assembly fails in the absence of muscle attachment. The terminal Z-disc attaches to the MTJ mediated by

integrins and IAPs [52]. Myofibrils are refined to regular arrays of sarcomeres over the next several hours where more sarcomeres are added. DLM myofibrils are flanked by MT arrays during initial stages of assembly that are dissembled by the end of pupation. The myofibril length then increases without other structural changes to reach its final length shortly after eclosion [231]. In the IFM, distinct transcriptional dynamics are associated with different stages of myofibrillogenesis, with the iTF Salm contributing to the transition after 30 h APF and its expression is maintained to establish IFM fate [230,232]. A similar sequence of myofibrillogenesis occurs in abdominal muscles that form mature MTJ by 50 h APF when myofibril assembly synchronously starts and is refined further to form the transversely aligned sarcomeres seen in abdominal muscles. Thin and thick filament complexes appear separately, then start interdigitating to form immature myofibrils by 46 h when muscles have stably attached to MTJ and exhibit spontaneous contractions. They subsequently assemble into ordered myofibrils by 50 h APF and are refined over the next several hours to begin coordinated contraction [229].

During IFM sarcomere assembly, thin filaments elongate from their pointed ends as is the case during embryonic myogenesis [170]. They initially form a dispersed pattern by the polymerization of actin into nascent thin filaments which become regularly patterned after 30 h APF. At this time, active incorporation of actin at both ends of the thin filament and further refinement and growth occurs by new actin monomer incorporation at the pointed ends of thin filaments and the formation of new thin filaments at the sarcomere periphery. Tmod and Sals that are located to pointed ends are necessary for thin filament length control [170,233]. The nebulin repeat containing protein Lasp regulates thin filament length by regulating its stability [234]. The *Drosophila* formin Fhos mediates thin filament assembly by initially regulating actin monomer incorporation into thin filaments during mid pupal stages and then localizes near Z-discs to facilitate radial growth of thin filament arrays to increase myofibril diameter [233]. In *Drosophila*, IFM thick filaments are associated with many insect-specific proteins such as myofilin [235], arthrin which is a ubiquitinated actin [236], paramyosin [237], minipramyosin [238], and flightin [239,240] not found in vertebrates, which could represent proteins adapted for flight [241]. The insect and IFM specific protein flightin is implicated in regulating the thick filament length by associating with myosin filaments as they grow [242,243]. Z-disc formation fails in the IFM if actin lacks its α-actinin binding domain showing the importance of sarcomere component interdigitation [244]. A downregulation of Sls results in smaller Z-discs around which a normal thick filament assembly occurs with abnormally long thick filaments at the periphery lacking the Z-disc [164,245]. As myofibrils grow, the Z-disc protein Zasp controls the final myofibril diameter by switching to growth restricting isoforms [246]. After complete myofibril growth, coordinated contractions can be initiated after mature NMJ formation.

3.2.5. Innervation and NMJ Formation

Embryonic neuroblast lineages undergo a second larval wave of neurogenesis where embryonic neuroblasts are re-specified to give rise to adult MN lineages whose dendrites are organized in a 'myotopic map' within the CNS that reflects the innervation pattern of their target adult muscles similar to embryonic/larval stages [247–250]. MNs innervate

adult muscles in a stereotypical pattern. For DLMs generated from larval templates, the primary larval ISN branch remains while secondary branches are initially retracted, and extensive new branching is generated as the muscles fuse with adult myoblasts and then split. Initial nerve arrival is muscle independent, but subsequent nerve branching occurs only in the presence of the target muscle [251]. Among DVMs, DVM I and DVM II are innervated by new branches arising from the larval ISN while the larval SN innervates DVM III [251,252]. The 14 leg muscles are innervated by around 50 MNs arising from specific neuroblasts in the CNS in a stereotypical pattern [250]. Following initial innervation, the NMJ is formed by extensive branching and synapse formation. The glial cells at the IFM NMJ express the glutamate *Drosophila* Excitatory Amino Acid Transporter 1 (dEAAT1) unlike during other stages for efficient neurotransmission [253]. Muscle iTFs contribute to correct innervation since malformed muscles cause MN branching aberrations as has been shown for Ewg [220].

In the end, a stereotypical muscle pattern along with stereotypical innervation generates fully functional adult muscles.

3.2.6. Programmed Cell Death Following Eclosion of New Adults

Some larval abdominal muscles persist through metamorphosis and are used for the eclosion of new adults. These muscles degenerate after eclosion along with associated nerves [254].

4. The Maintenance of Muscle Homeostasis

4.1. Muscle Homeostasis under Normal Conditions

Functional larval somatic muscles and adult muscles represent two different homeostatic states during the fly lifetime. The embryonic wave of myogenesis takes only one day leading to the formation of functional larval muscles, which undergo continuous growth and refinement during the larval stages spanning five days. Larval muscle homeostasis needs to be coordinated with larval growth during the three larval instars until metamorphosis to ensure functional stability. Following metamorphosis and the pupal wave of myogenesis over a period of five days, adult flies eclose from their pupae and adult muscle homeostasis needs to be maintained during the fly lifespan of several weeks.

The stereotypical muscle pattern is associated with iTFs and their realisator genes that also exhibit tightly controlled spatial and temporal expression patterns in larval and adult muscles. Therefore, some of the iTFs can play a key role in the maintenance of muscle specific homeostasis by regulating the levels of key myogenic factors such as Mef2 as well as the expression of realisator genes [111,112,129,137] (Figure 2). The control of the level of activity of the key differentiation TF Mef2 is quintessential throughout the fly lifetime since this in turn controls the muscle specific levels of its vast array of target genes [118,129]. In the embryo, various genes were shown to require different Mef2 activity, with early expressing genes such as *Act57B* requiring lower levels compared to late expressing genes such as *Mhc* [118]. In the adult, the development and maintenance of the adult DLM muscles have been observed to be sensitive to the levels of Mef2 as well as its antagonist Holes in muscles (Him). Tubular adult muscles such as the TDT and DVM muscles seem

to require lower Mef2 activity than the fibrillar DLM muscles since RNAi lines affect these muscles differently [255,256]. TFs such as Cf2 and E2f1 acting along with Mef2 could also contribute to setting the muscle homeostatic state (García-Zaragoza et al. 2008; Zappia et Frolov 2016). A study identified putative Cf2 and Mef2 binding site clusters for multiple sarcomeric genes including *Mhc*, *Tm1*, *Tm2*, *up*, *wupA* (or *TnI*), and *paramyosin* (*Prm*) [122]. On Cf2 depletion, the stoichiometry of proteins such as TnT, TnI, and Prm was found to be altered and this imbalance worsened over the course of development. Another study detected E2f binding site enrichment upstream of myogenic genes such as *how*, *sals*, *Tm1*, *Mef2*, etc. This study also showed that E2f1 depletion altered the gene expression levels of *Tm2*, *Act88F*, *Mlc2*, *how*, and *Mef2* [167].

One hallmark of muscle homeostasis in *Drosophila* larval and adult muscles is the expression of fiber specific protein isoforms. Many sarcomeric genes switch between embryonic, larval, and/or adult isoforms during development, with different muscle types also exhibiting isoform specificity. Isoform switching usually occurs by switching to a predominant isoform. Embryonic *Mhc* transcripts contain exon 19, which is spliced out of adult versions and results in a different carboxy terminal [242,257]. Embryonic isoforms lack the functionality for the high ATPase rate and sliding velocity required for adult muscles [258]. The IFM muscles initially express an Mhc isoform containing exon 19 and switch to the adult exon 18 containing isoform during late stages of myofibril assembly [242]. A shorter embryonic/larval isoform of the pointed end capping protein Tmod is associated with actin during pupal sarcomere assembly and there is a switch to a longer Tmod isoform in eclosed adults [170]. Adult *Drosophila* muscles express fiber specific actins, with Act88F being expressed in the IFM and Act79B in the TDT, for example [202]. Two IFM specific Tm1 isoforms are expressed in adult flies [259,260]. Kettin is the predominant *Drosophila* titin isoform in embryos and the IFM muscles switch to the IFM specific predominant long Sls(700) isoform [245]. Zasp52 and other Zasp proteins also switch to adult isoforms [246,261], with Zasp52 expressing an exon 8 containing isoform absent in embryos, but present in the IFM and TDT. Obscurin expresses a single larval isoform and two IFM isoforms [262].

Isoform switches are potentially associated with cis regulatory modules (CRMs) that seem to be arranged in sequential modules mirroring developmental expression and regulation by different TFs and cofactors. Marin et al. identified an upstream regulatory element (URE) and an intronic regulatory element (IRE) in intron 1 of the *wupA* (or *TnI*) gene that acted synergistically and was capable of driving LacZ tagged TnI expression. Mas et al. identified similar elements in the *up* (or *TnT*) gene [263]. They showed that these elements synergistically interact in larval muscles, whereas the contribution of the IRE is higher in adult muscles. In addition, they showed that there was decreasing IRE contribution from the IFM to the jump muscles to the visceral muscles [264]. Garcia-Zaragoza et al. followed up on this study and identified the URE and potential IRE elements of *Tm1*, *Tm2*, and *Mhc*. *Tm1* was previously shown to be coordinately regulated by two intronic enhancers in cooperation with Mef2 and its interactor PAR domain protein 1 (Pdp1) [265–267]. Mature muscles need to ensure the activation and maintenance of the correct protein isoforms [268] since aberrant isoform expression impedes muscle function. For example, transient overexpression of a shorter Tmod isoform during mid-to late IFM assembly leads to normal length thin filaments at the periphery of the myofibrils that are

correctly capped by the long Tmod isoform. However, they exhibit shorter core thin filaments within the myofibril caused by the permanent association of the shorter Tmod at their pointed ends, which cannot be dynamically uncapped to permit thin filament elongation. Therefore, this prevents its elongation causing defective sarcomeres that interfere with flight during adult stages [170]. The embryonic Mhc isoform fails to substitute for the IFM isoform due to different physiological properties [258,269].

Post transcriptional mechanisms such as phosphorylation could potentially contribute to muscle homeostasis. Thin and thick filament disruptions, for example, are associated with concomitant flightin phosphorylation deregulations [239]. Tm1 IFM isoforms are phosphorylated only in adult flies, which could have functional implications [260]. Impaired Talin phosphorylation leads to severe muscle detachment at late embryonic stages [154]. This means the right CRM regulatory mechanisms as well as post translational mechanisms such as phosphorylation [48,270] need to be dynamically maintained since specific protein domains are necessary for muscle specific functionality [269,271,272].

The accumulation of insoluble protein aggregates in the muscle is associated with protein aggregate myopathies (PAM) and in *Drosophila*, p38b deficiency leads to the deposition of polyubiquitinated protein aggregates in adult thoracic muscles and to locomotor defects [273]. Loss of components of the proteasome, which mediate protein turnover were shown to cause protein aggregates and progressive muscle atrophy in larval muscles [274]. Ubiquitin protein ligases such as Mind bomb 2 (Mib2) and Ubiquitin protein ligase E3A (Ube3A), which tag proteins for proteasomal degradation, have been associated with muscle defects. The loss of function of *mib2* was shown to trigger embryonic muscle apoptosis [275] and the over or under expression of *Ube3a* alters larval NMJ neurotransmission with associated altered number of active zones [276]. Proteostasis is thus integral to muscle maintenance.

Muscle contraction is associated with multiple biochemical and morphological changes as well as large mechanical strains. This necessitates efficient mechanisms to withstand these forces to prevent muscle disintegration during contraction and to reinstate the stable muscle state (Figure 3). Protein stoichiometry is integral to sarcomere integrity since varying the expression levels of one protein has a cascading effect on the levels of other sarcomeric proteins leading to altered muscle functionality [277,278]. Sarcomeric integrity during contractions is maintained by components such as Mlp84B, Cher, small heat shock proteins (sHsps) such as dCryAB and Hsp67Bc and integrin-mediated adhesions. Mlp84B localizes to the Z-disc and genetically interacts with Sls. Mlp84B-Sls transheterozygotes exacerbate individual mutant phenotypes disrupting myofibrillar integrity [56]. Cher also interacts with Sls in addition to actin stably anchoring them to each other [53]. In addition, Cher interacts physically with dCryAB and a disruption of this interaction affects sarcomeric integrity [279]. The chaperone Hsp67Bc also colocalizes to the Z-disc although its function is unknown [280]. Integrin mediated adhesions maintain sarcomeric integrity and reduced adhesion results in the progressive age-dependent loss of sarcomeric cytoarchitecture [57]. Integrin and IAP stoichiometries at the MTJ are important to respond to different types of forces [166]. The myonuclear LINC complex and associated components such as Msp300 and Spectraplakin, which regulate MT organization, play a role in myonuclear maintenance by providing elasticity

to resist contractile forces with the help of the MT network that surrounds it [281–283]. In addition, Msp300 associates with the Z-disc and keeps the mitochondria and SR anchored to the Z-disc during contractions [284]. Its presence around myonuclei near the larval NMJ also regulates glutamate receptor density to control locomotion [285].

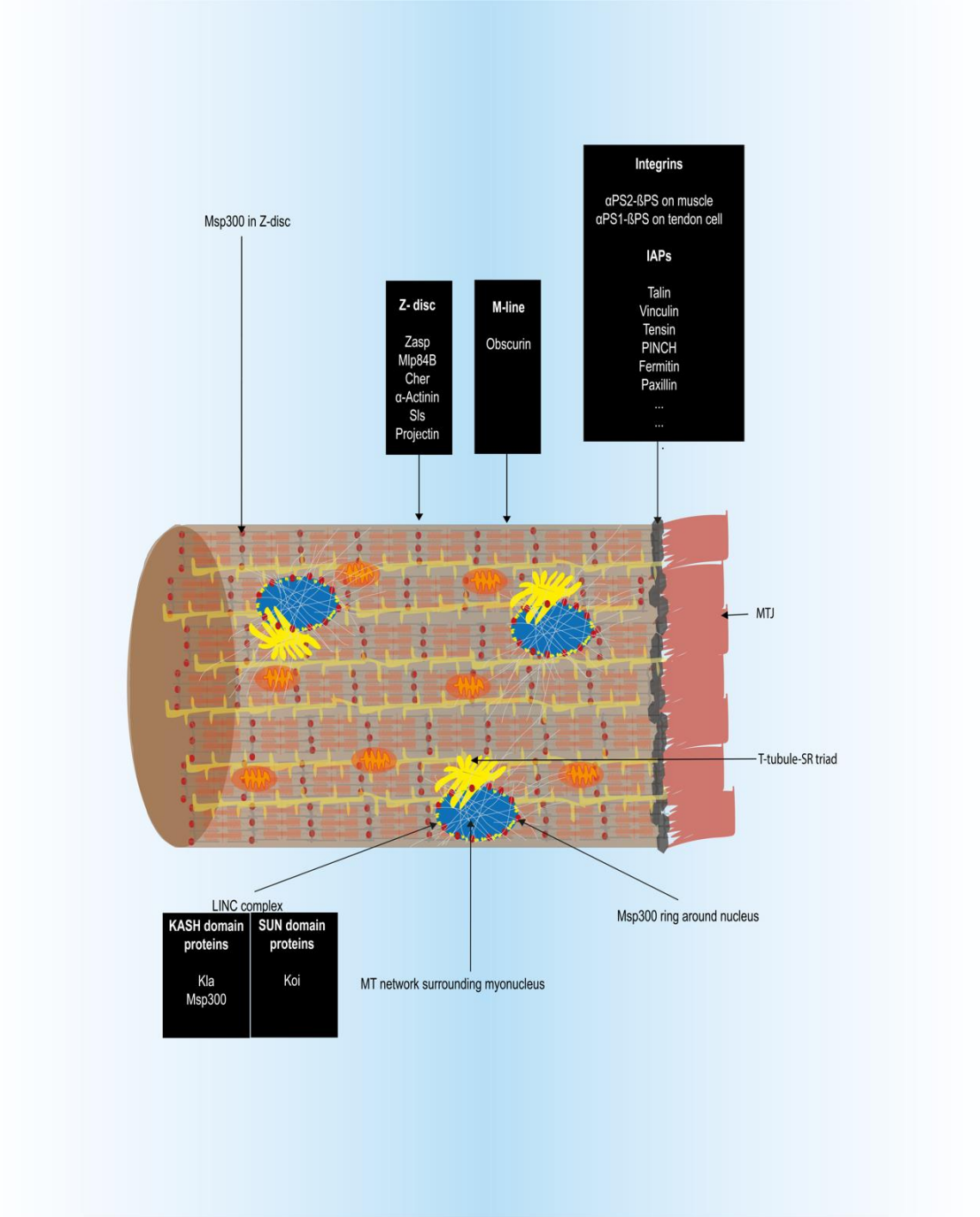


Figure 3. Maintenance of myofibril integrity and homeostasis. The integrin complex links the myofibrils to the MTJ via the extracellular matrix (ECM) and senses the forces transmitted by the MTJ. Integrins and Integrin Associated Proteins (IAPs) constitute the integrin complex. Integrin complex turnover and constitution are adapted to the forces sensed during contraction. The dense microtubule (MT) network anchored to the myonuclei by the Msp300 ring associated with the LINC complex on the nuclear envelope provides myonuclear elasticity during contractions to prevent disintegration of myonuclei and dissociation of the

myofibril network. Msp300 in the Z-disc ensure regular spacing of organelles such as mitochondria and the SR for contractions. Z-disc and M-line components provide anchorage and elasticity to ensure sarcomeric integrity.

NMJ activity perturbations lead to homeostatic synaptic plasticity, which enables compensatory modulations of the NMJ synaptic strength to resist these perturbations and stabilize synaptic activity. Lifelong synaptic plasticity ensures efficient neurotransmission of signals at the NMJ. The NMJ adapts various homeostatic mechanisms to maintain appropriate muscle function levels [286–289]. Mutants for *endophilin* (*endo*) exhibit tremendous synaptic overgrowth, but the overall synaptic strength is stabilized by reducing the active zone number in synaptic buttons, which modulates neurotransmitter release [286]. The NMJ adapts a homeostatic scaling mechanism called presynaptic homeostatic potentiation (PHP), where there is a compensatory increase in neurotransmitter release to maintain muscle excitation in response to abnormally reduced GluR on the postsynaptic end. This compensation appears to be associated with an uncharacteristic multilayer ring of electron dense T-bars in active zones to increase the neurotransmitter release [289]. The PHP maintenance has been shown to require inositol triphosphate (IP_3) directed signaling [290]. During the larval growth spurt, NMJ homeostasis needs to be maintained even though the presynaptic end grows slower than the muscle surface that tends to accumulate GluRs. Ziegler et al. showed that the amino acid transporter, Juvenile hormone Inducible-21 (JhI-21) is a gene that coevolved with GluRs, is expressed at presynaptic ends and plays a role in suppressing excess GluR accumulation [288].

The close association of the mesoderm with other germ layers right from the embryonic stage and the continued association with epidermal and nervous tissues over the fly lifetime highlights the importance of coordinated intrinsic and extrinsic signaling for homeostasis.

4.2. Re-Establishment of Muscle Homeostasis Following Muscle Injury

Muscle regeneration has not been described in the larva. However, the larval stem cell-like AMPs that are capable of differentiating and giving rise to adult muscles were noted to have similarities to vertebrate muscle stem cells (MuSCs), also known as satellite cells. Similar to MuSCs, the Notch pathway [209,291] and zinc-finger homeodomain 1 (Zfh1), the *Drosophila* homolog of the vertebrate ZEB1/ZEB2 [292,293], maintain the AMPs in an undifferentiated state and they are capable of self-renewal by asymmetric divisions [209,294]. In addition, they are capable of fusion with existing larval muscle remnants during the formation of DLM muscles, which is reminiscent of muscle repair. It was initially thought that all muscle stem cell-like cells or AMPs are depleted during adult muscle formation and thus adult muscles were believed to lack regenerative capacity. Recently, Chaturvedi et al. identified a population of Zfh1 positive adult stem cells closely apposed to the adult muscle, which appear to possess the ability to proliferate and contribute to muscle regeneration upon injury [61] similar to vertebrate MuSCs [293]. Boukhatmi and Bray subsequently showed that Notch directly regulates Zfh1 to antagonize the differentiation of these cells by expressing a short Zfh1 isoform transcribed from an alternate promoter that is not subject to regulation by the conserved micro RNA,

miR-8 [292]. Using the G-TRACE method for cell lineage analysis, their study showed that these cells, which they termed population of progenitors that persist in adults or pMPs, were mitotically active and incorporated into adult muscles even under normal conditions. Thus, they reiterated that these cells contributed to adult muscle homeostasis. Since this is a recent discovery, further studies could provide insights into the extent of repair in *Drosophila* adult muscles and the mechanisms involved in re-establishing and maintaining muscle identity and homeostasis.

4.3. Muscle Homeostasis under Pathological Conditions

Multiple studies in *Drosophila* models have reproduced defects observed in human pathological conditions and could provide important insights into disruptions of muscle homeostasis under pathological conditions. Many myopathies and neuromuscular disorders are associated with or even caused by myonuclear defects and others are associated with sarcomeric defects leading to muscle dysfunction, wasting, and/or degeneration. *Drosophila* models exist for multisystemic disorders such as Myotonic Dystrophy Type 1 (DM1) that is caused by CTG expansions in the *Dystrophia Myotonica Protein Kinase* (*DMPK*) gene leading to the sequestration of RNA binding proteins such as MBNL1 in nuclear foci [295,296]. This causes a disruption of muscle homeostasis as indicated by the progressive muscle degeneration observed in the IFM muscles in a model expressing 480 CTG repeats. The Dystrophin (Dys)-Dystroglycan (Dg) transmembrane complex at the plasma membrane acts as a crucial signaling mediator by relaying information to and from muscles to interacting tissues via the ECM. Mutations in genes constituting this complex or their interactors thus cause a disruption of homeostasis, thereby causing diseases such as Duchenne Muscular Dystrophy (DMD) where there is muscle wasting. In *Drosophila*, *Dg* was shown to be under miRNA regulation by miR-9a to ensure correct MTJ formation [297]. Large scale genetic and interactome screens in *Drosophila* have identified factors affecting muscle integrity such as stress response components [298] and components of the Hippo signaling pathway [42].

Laminopathies are disorders caused by mutations in the human *LMNA* genes which code for lamins present in the INM providing structural support and regulating gene expression. One *Drosophila* model revealed increased reductive stress due to the nuclear translocation of Nrf2, which is normally sequestered in the cytoplasm and released only during oxidative stress [299]. Chandran et al. observed a loss of muscle proteostasis in a *Drosophila* model of laminopathies and corroborated this by RNA-seq analyses of human muscle biopsy tissues. Interestingly, they were able to rescue the muscular phenotypes by the modulation of the AMPK pathway which could present future therapeutic directions [300]. Apart from laminopathies, other myopathies such as Centronuclear Myopathies (CNM) are associated with myonuclear positioning defects [301]. Muscle development studies in *Drosophila* are beginning to unveil mechanisms for myonuclear positioning and factors that disrupt this [141,284,302,303].

Other studies in *Drosophila* are providing insights into pathological features caused by disruptions in sarcomeric components. Muscular phenotypes caused by a mutation in the *Tm2* gene was found to be rescued by a suppressor mutation in the *wupA* gene coding for TnI [304]. *Drosophila* models exist for myosin myopathies such as Inclusion Body

Myopathy Type 3 and Laing Distal Myopathy (LDM). A study has shown that the formation of large aggregates in muscles similar to those seen in human patients with *ZASP* mutations is caused by an imbalance in the levels of Zasp isoforms [246]. Dahl-Halvarsson et al. showed that the overexpression of the Thin protein, a homolog of the human TRIM family of proteins that is implicated in maintaining sarcomeric integrity, could alleviate LDM-like phenotypes [305].

5. Discussion

In vertebrates, the loss of skeletal muscle homeostasis is the cause of various muscular disorders. Studies in vertebrate systems are complicated by the presence of large gene families for multiple genes. *Drosophila* is a simple model organism with various conserved pathways and genes to study muscle homeostasis while at the same time mostly having one to a few genes orthologous to large vertebrate gene families that perform functions similar to vertebrate genes. Thus, it appears that genes are reused/repurposed over the course of evolution instead of 'reinventing the wheel'. *Drosophila* muscle development has been studied for decades. The embryonic somatic muscles being uni-fiber muscles present a simple model to study development since all muscles have been well characterized along with their specific attachment sites and innervating MNs [143,180]. The IFM muscles have been equally well characterized [31,221,252]. In addition, a large number of tools are available in *Drosophila* to study in vivo mechanisms [306].

A better understanding of developmental and post developmental processes would help us gain a better understanding of the mechanisms of maintenance and disruption of homeostasis. The short life cycle of the fruit fly facilitates the quick and detailed study of processes making it a valuable model for the study of factors that initiate, maintain, and disrupt muscle homeostasis. The study of muscle regeneration following muscle injury, where developmental processes need to be re-initiated, represents an example of how the *Drosophila* model could help understand the mechanisms of muscle homeostasis. Some potential therapeutic targets have been unveiled by studies in *Drosophila* models of myopathies [300,305]. The recent discovery of stem cell-like cells associated with adult muscles is an exciting new direction of research to study muscle regeneration and homeostasis [61]. In vertebrates, aging is related to a depletion of the MuSC population leading to sarcopenia or age-related gradual loss of muscle mass and function [307] that is also characteristic of pathological conditions such as DMD [308]. The short life span of the *Drosophila* model presents a huge advantage to study homeostatic disruptions during aging.

Large gaps exist in our understanding of pathological mechanisms and simpler models could provide valuable insights and therapeutic directions. In *Drosophila*, although a lot of attention has been given to the major muscle components including the sarcomeres, MTJ and NMJ, muscle organelles that play an equally central role such as the SR, T-tubules, golgi complex, and transport vesicles have received lesser attention, although myonuclei are beginning to be studied in detail. Given the detailed characterization and tools available for this established model system that has already helped advance research [309,310], it would continue to serve as an important backbone for research into various physiological processes including muscle development and homeostasis.

Author Contributions: Conceptualization, P.P.; writing—original draft preparation, P.P.; writing—review and editing, P.P. and K.J.; visualization, P.P.; supervision, K.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the AFM-Telethon, grant number 21182 to the MyoNeurAlp Alliance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bate, M. The embryonic development of larval muscles in Drosophila. *Development* **1990**, *110*, 791.
2. Abmayr, S.M.; Erickson, M.S.; Bour, B.A. Embryonic development of the larval body wall musculature of Drosophila melanogaster. *Trends Genet.* **1995**, *11*, 153–159, doi:10.1016/S0168-9525(00)89030-7.
3. Fyrberg, E. Study of contractile and cytoskeletal proteins using Drosophila genetics. *Cell Motil. Cytoskelet.* **1989**, *14*, 118–127, doi:10.1002/cm.970140121.
4. Dobi, K.C.; Schulman, V.K.; Baylies, M.K. Specification of the somatic musculature in Drosophila. *Wiley Interdiscip. Rev. Dev. Biol.* **2015**, *4*, 357–375, doi:10.1002/wdev.182.
5. Mukund, K.; Subramaniam, S. Skeletal muscle: A review of molecular structure and function, in health and disease. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2020**, *12*, e1462, doi:10.1002/wsbm.1462.
6. Taylor, M. Comparison of muscle development in Drosophila and vertebrates. In *Muscle development in Drosophila*; Sink, H., Ed.; Landes Bioscience/Springer: Georgetown, TX, USA; New York, NY, USA, 2006; pp. 169–203.
7. Piccirillo, R.; Demontis, F.; Perrimon, N.; Goldberg, A.L. Mechanisms of muscle growth and atrophy in mammals and Drosophila. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **2014**, *243*, 201–215, doi:10.1002/dvdy.24036.
8. Kohsaka, H.; Guertin, P.A.; Nose, A. Neural Circuits Underlying Fly Larval Locomotion. *Curr. Pharm. Des.* **2017**, *23*, 1722–1733, doi:10.2174/1381612822666161208120835.
9. Wolfe, R.R. The underappreciated role of muscle in health and disease. *Am. J. Clin. Nutr.* **2006**, *84*, 475–482, doi:10.1093/ajcn/84.3.475.
10. Veratti, E. Investigations on the fine structure of striated muscle fiber read before the Reale Istituto Lombardo, 13 March 1902. *J. Biophys. Biochem. Cytol.* **1961**, *10*, 1–59, doi:10.1083/jcb.10.4.1.
11. Hanson, J.; Huxley, H.E. Structural Basis of the Cross-Striations in Muscle. *Nature* **1953**, *172*, 530–532, doi:10.1038/172530b0.
12. Royuela, M.; Fraile, B.; Arenas, M.I.; Paniagua, R. Characterization of several invertebrate muscle cell types: A comparison with vertebrate muscles. *Microsc. Res. Tech.* **2000**, *48*, 107–115, doi:10.1002/(SICI)1097-0029(20000115)48:2<107::AID-JEMT6>3.0.CO;2-U.
13. Littlefield, R.S.; Fowler, V.M. Thin filament length regulation in striated muscle sarcomeres: Pointed-end dynamics go beyond a nebulin ruler. *Semin. Cell Dev. Biol.* **2008**, *19*, 511–519, doi:10.1016/j.semcdb.2008.08.009.
14. Lemke, S.B.; Schnorrer, F. Mechanical forces during muscle development. *Mech. Dev.* **2017**, *144*, 92–101, doi:10.1016/j.mod.2016.11.003.
15. Wang, L.; Geist, J.; Grogan, A.; Hu, L.-Y.R.; Kontrogianni-Konstantopoulos, A. Thick Filament Protein Network, Functions, and Disease Association. *Compr. Physiol.* **2018**, *8*, 631–709, doi:10.1002/cphy.c170023.
16. Hooper, S.L.; Thuma, J.B. Invertebrate Muscles: Muscle Specific Genes and Proteins. *Physiol. Rev.* **2005**, *85*, 1001–1060, doi:10.1152/physrev.00019.2004.
17. Schweitzer, R.; Zelzer, E.; Volk, T. Connecting muscles to tendons: Tendons and musculoskeletal development in flies and vertebrates. *Dev. Camb. Engl.* **2010**, *137*, 2807–2817, doi:10.1242/dev.047498.
18. Soler, C.; Laddada, L.; Jagla, K. Coordinated Development of Muscles and Tendon-Like Structures: Early Interactions in the Drosophila Leg. *Front. Physiol.* **2016**, *7*, 22–22, doi:10.3389/fphys.2016.00022.
19. Lemke, S.B.; Weidemann, T.; Cost, A.-L.; Grashoff, C.; Schnorrer, F. A small proportion of Talin molecules transmit forces at developing muscle attachments in vivo. *PLoS Biol.* **2019**, *17*, e3000057, doi:10.1371/journal.pbio.3000057.

20. Richier, B.; Inoue, Y.; Dobramysl, U.; Friedlander, J.; Brown, N.H.; Gallop, J.L. Integrin signaling downregulates filopodia during muscle-tendon attachment. *J. Cell Sci.* **2018**, *131*, jcs217133, doi:10.1242/jcs.217133.
21. Nawrotzki, R.; Willem, M.; Miosge, N.; Brinkmeier, H.; Mayer, U. Defective integrin switch and matrix composition at alpha 7-deficient myotendinous junctions precede the onset of muscular dystrophy in mice. *Hum. Mol. Genet.* **2003**, *12*, 483–495, doi:10.1093/hmg/ddg047.
22. Marshall, J.L.; Chou, E.; Oh, J.; Kwok, A.; Burkin, D.J.; Crosbie-Watson, R.H. Dystrophin and utrophin expression require sarcospan: Loss of α7 integrin exacerbates a newly discovered muscle phenotype in sarcospan-null mice. *Hum. Mol. Genet.* **2012**, *21*, 4378–4393, doi:10.1093/hmg/dds271.
23. Broadie, K.; Bate, M. The Drosophila NMJ: A genetic model system for synapse formation and function. *Semin. Dev. Biol.* **1995**, *6*, 221–231, doi:10.1016/S1044-5781(06)80031-9.
24. Menon, K.P.; Carrillo, R.A.; Zinn, K. Development and plasticity of the Drosophila larval neuromuscular junction. *WIREs Dev. Biol.* **2013**, *2*, 647–670, doi:10.1002/wdev.108.
25. Harris, K.P.; Littleton, J.T. Transmission, Development, and Plasticity of Synapses. *Genetics* **2015**, *201*, 345–375, doi:10.1534/genetics.115.176529.
26. Lacin, H.; Truman, J.W. Lineage mapping identifies molecular and architectural similarities between the larval and adult Drosophila central nervous system. *eLife* **2016**, *5*, e13399, doi:10.7554/eLife.13399.
27. Pérez-Moreno, J.J.; O'Kane, C.J. GAL4 Drivers Specific for Type Ib and Type Is Motor Neurons in Drosophila. *G3 GenesGenomesGenetics* **2019**, *9*, 453, doi:10.1534/g3.118.200809.
28. Dasari, S.; Cooper, R.L. Modulation of sensory–CNS–motor circuits by serotonin, octopamine, and dopamine in semi-intact Drosophila larva. *Neurosci. Res.* **2004**, *48*, 221–227, doi:10.1016/j.neures.2003.10.005.
29. Kohsaka, H.; Takasu, E.; Morimoto, T.; Nose, A. A Group of Segmental Premotor Interneurons Regulates the Speed of Axial Locomotion in Drosophila Larvae. *Curr. Biol.* **2014**, *24*, 2632–2642, doi:10.1016/j.cub.2014.09.026.
30. Babski, H.; Jovanic, T.; Surel, C.; Yoshikawa, S.; Zwart, M.F.; Valmier, J.; Thomas, J.B.; Enriquez, J.; Carroll, P.; Garcès, A. A GABAergic Maf-expressing interneuron subset regulates the speed of locomotion in Drosophila. *Nat. Commun.* **2019**, *10*, 4796–4796, doi:10.1038/s41467-019-12693-6.
31. Vigoreaux, J.O. Genetics of the Drosophila flight muscle myofibril: A window into the biology of complex systems. *BioEssays* **2001**, *23*, 1047–1063, doi:10.1002/bies.1150.
32. Campbell, K.B.; Chandra, M. Functions of stretch activation in heart muscle. *J. Gen. Physiol.* **2006**, *127*, 89–94, doi:10.1085/jgp.200509483.
33. Huxley, H.E. Fifty years of muscle and the sliding filament hypothesis. *Eur. J. Biochem.* **2004**, *271*, 1403–1415, doi:10.1111/j.1432-1033.2004.04044.x.
34. Wang, Z.-H.; Clark, C.; Geisbrecht, E.R. Analysis of mitochondrial structure and function in the Drosophila larval musculature. *Mitochondrion* **2016**, *26*, 33–42, doi:10.1016/j.mito.2015.11.005.
35. Auld, A.L.; Folker, E.S. Nucleus-dependent sarcomere assembly is mediated by the LINC complex. *Mol. Biol. Cell* **2016**, *27*, 2351–2359, doi:10.1091/mbc.E16-01-0021.
36. Al-Qusairi, L.; Laporte, J. T-tubule biogenesis and triad formation in skeletal muscle and implication in human diseases. *Skelet. Muscle* **2011**, *1*, 26–26, doi:10.1186/2044-5040-1-26.
37. Razzaq, A.; Robinson, I.M.; McMahon, H.T.; Skepper, J.N.; Su, Y.; Zelhof, A.C.; Jackson, A.P.; Gay, N.J.; O'Kane, C.J. Amphiphysin is necessary for organization of the excitation-contraction coupling machinery of muscles, but not for synaptic vesicle endocytosis in Drosophila. *Genes Dev.* **2001**, *15*, 2967–2979, doi:10.1101/gad.207801.
38. Ackermann, M.A.; Ziman, A.P.; Strong, J.; Zhang, Y.; Hartford, A.K.; Ward, C.W.; Randall, W.R.; Kontrogianni-Konstantopoulos, A.; Bloch, R.J. Integrity of the network sarcoplasmic reticulum in skeletal muscle requires small ankyrin 1. *J. Cell Sci.* **2011**, *124*, 3619–3630, doi:10.1242/jcs.085159.
39. Maughan, D.W.; Godt, R.E. Equilibrium distribution of ions in a muscle fiber. *Biophys. J.* **1989**, *56*, 717–722, doi:10.1016/S0006-3495(89)82719-5.
40. Clausen, T. Na+-K+ Pump Stimulation Improves Contractility in Damaged Muscle Fibers. *Ann. N. Y. Acad. Sci.* **2006**, *1066*, 286–294, doi:10.1196/annals.1363.021.
41. Lebovitz, R.M.; Takeyasu, K.; Fambrough, D.M. Molecular characterization and expression of the (Na+ + K+)-ATPase alpha-subunit in Drosophila melanogaster. *EMBO J.* **1989**, *8*, 193–202.

42. Yatsenko, A.S.; Kucherenko, M.M.; Xie, Y.; Aweida, D.; Urlaub, H.; Scheibe, R.J.; Cohen, S.; Shcherbata, H.R. Profiling of the muscle-specific dystroglycan interactome reveals the role of Hippo signaling in muscular dystrophy and age-dependent muscle atrophy. *BMC Med.* **2020**, *18*, 8, doi:10.1186/s12916-019-1478-3.
43. Huxley, H.; Hanson, J. Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation. *Nature* **1954**, *173*, 973–976, doi:10.1038/173973a0.
44. Huxley, A.F.; Niedergerke, R. Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres. *Nature* **1954**, *173*, 971–973, doi:10.1038/173971a0.
45. Potter, J.D.; Sheng, Z.; Pan, B.-S.; Zhao, J. A Direct Regulatory Role for Troponin T and a Dual Role for Troponin C in the Ca2+ Regulation of Muscle Contraction. *J. Biol. Chem.* **1995**, *270*, 2557–2562, doi:10.1074/jbc.270.6.2557.
46. Qiu, F.; Lakey, A.; Agianian, B.; Hutchings, A.; Butcher, G.W.; Labeit, S.; Leonard, K.; Bullard, B. Troponin C in different insect muscle types: Identification of two isoforms in Lethocerus, Drosophila and Anopheles that are specific to asynchronous flight muscle in the adult insect. *Biochem. J.* **2003**, *371*, 811–821, doi:10.1042/BJ20021814.
47. Vibert, P.; Craig, R.; Lehman, W. Steric-model for activation of muscle thin filaments 1 1 Edited by P.E. Wright. *J. Mol. Biol.* **1997**, *266*, 8–14, doi:10.1006/jmbi.1996.0800.
48. Farman, G.P.; Miller, M.S.; Reedy, M.C.; Soto-Adames, F.N.; Vigoreaux, J.O.; Maughan, D.W.; Irving, T.C. Phosphorylation and the N-terminal extension of the regulatory light chain help orient and align the myosin heads in Drosophila flight muscle. *J. Struct. Biol.* **2009**, *168*, 240–249, doi:10.1016/j.jsb.2009.07.020.
49. Beall, C.J.; Fyrberg, E. Muscle abnormalities in Drosophila melanogaster heldup mutants are caused by missing or aberrant troponin-I isoforms. *J. Cell Biol.* **1991**, *114*, 941–951, doi:10.1083/jcb.114.5.941.
50. Lee, K.H.; Sulbarán, G.; Yang, S.; Mun, J.Y.; Alamo, L.; Pinto, A.; Sato, O.; Ikebe, M.; Liu, X.; Korn, E.D.; et al. Interacting-heads motif has been conserved as a mechanism of myosin II inhibition since before the origin of animals. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1991–E2000, doi:10.1073/pnas.1715247115.
51. Jung, H.S.; Komatsu, S.; Ikebe, M.; Craig, R. Head-head and head-tail interaction: A general mechanism for switching off myosin II activity in cells. *Mol. Biol. Cell* **2008**, *19*, 3234–3242, doi:10.1091/mbc.e08-02-0206.
52. Green, H.J.; Griffiths, A.G.; Ylänne, J.; Brown, N.H. Novel functions for integrin-associated proteins revealed by analysis of myofibril attachment in Drosophila. *eLife* **2018**, *7*, e35783, doi:10.7554/eLife.35783.
53. González-Morales, N.; Holenka, T.K.; Schöck, F. Filamin actin-binding and titin-binding fulfill distinct functions in Z-disc cohesion. *PLoS Genet.* **2017**, *13*, e1006880, doi:10.1371/journal.pgen.1006880.
54. Liao, K.A.; González-Morales, N.; Schöck, F. Zasp52, a Core Z-disc Protein in Drosophila Indirect Flight Muscles, Interacts with α-Actinin via an Extended PDZ Domain. *PLoS Genet.* **2016**, *12*, e1006400, doi:10.1371/journal.pgen.1006400.
55. Katzemich, A.; West, R.J.H.; Fukuzawa, A.; Sweeney, S.T.; Gautel, M.; Sparrow, J.; Bullard, B. Binding partners of the kinase domains in Drosophila obscurin and their effect on the structure of the flight muscle. *J. Cell Sci.* **2015**, *128*, 3386–3397, doi:10.1242/jcs.170639.
56. Clark, K.A.; Bland, J.M.; Beckerle, M.C. The Drosophila muscle LIM protein, Mlp84B, cooperates with D-titin to maintain muscle structural integrity. *J. Cell Sci.* **2007**, *120*, 2066, doi:10.1242/jcs.000695.
57. Perkins, A.D.; Ellis, S.J.; Asghari, P.; Shamsian, A.; Moore, E.D.W.; Tanentzapf, G. Integrin-mediated adhesion maintains sarcomeric integrity. *Dev. Biol.* **2010**, *338*, 15–27, doi:10.1016/j.ydbio.2009.10.034.
58. Prill, K.; Dawson, J.F. Assembly and Maintenance of Sarcomere Thin Filaments and Associated Diseases. *Int. J. Mol. Sci.* **2020**, *21*, 542, doi:10.3390/ijms21020542.
59. Ciglar, L.; Furlong, E.E. Conservation and divergence in developmental networks: A view from Drosophila myogenesis. *Cell Differ. Cell Div. Growth Death* **2009**, *21*, 754–760, doi:10.1016/j.ceb.2009.10.001.
60. Karalaki, M.; Fili, S.; Philippou, A.; Koutsilieris, M. Muscle regeneration: Cellular and molecular events. *Vivo Athens Greece* **2009**, *23*, 779–796.
61. Chaturvedi, D.; Reichert, H.; Gunage, R.D.; VijayRaghavan, K. Identification and functional characterization of muscle satellite cells in Drosophila. *eLife* **2017**, *6*, e30107, doi:10.7554/eLife.30107.

62. Demontis, F.; Perrimon, N. Integration of Insulin receptor/Foxo signaling and dMyc activity during muscle growth regulates body size in Drosophila. *Dev. Camb. Engl.* **2009**, *136*, 983–993, doi:10.1242/dev.027466.
63. Xiao, G.; Mao, S.; Baumgarten, G.; Serrano, J.; Jordan, M.C.; Roos, K.P.; Fishbein, M.C.; MacLellan, W.R. Inducible activation of c-Myc in adult myocardium in vivo provokes cardiac myocyte hypertrophy and reactivation of DNA synthesis. *Circ. Res.* **2001**, *89*, 1122–1129, doi:10.1161/hh2401.100742.
64. Martin, A.C. The Physical Mechanisms of Drosophila Gastrulation: Mesoderm and Endoderm Invagination. *Genetics* **2020**, *214*, 543, doi:10.1534/genetics.119.301292.
65. Azpiazu, N.; Lawrence, P.A.; Vincent, J.P.; Frasch, M. Segmentation and specification of the Drosophila mesoderm. *Genes Dev.* **1996**, *10*, 3183–3194, doi:10.1101/gad.10.24.3183.
66. Baylies, M.K.; Bate, M. twist: A Myogenic Switch in Drosophila. *Science* **1996**, *272*, 1481, doi:10.1126/science.272.5267.1481.
67. Carmena, A.; Bate, M.; Jiménez, F. Lethal of scute, a proneural gene, participates in the specification of muscle progenitors during Drosophila embryogenesis. *Genes Dev.* **1995**, *9*, 2373–2383, doi:10.1101/gad.9.19.2373.
68. Carmena, A.; Buff, E.; Halfon, M.S.; Gisselbrecht, S.; Jiménez, F.; Baylies, M.K.; Michelson, A.M. Reciprocal Regulatory Interactions between the Notch and Ras Signaling Pathways in the Drosophila Embryonic Mesoderm. *Dev. Biol.* **2002**, *244*, 226–242, doi:10.1006/dbio.2002.0606.
69. Doe, C.Q.; Skeath, J.B. Neurogenesis in the insect central nervous system. *Curr. Opin. Neurobiol.* **1996**, *6*, 18–24, doi:10.1016/S0959-4388(96)80004-3.
70. Crews, S.T. Drosophila Embryonic CNS Development: Neurogenesis, Gliogenesis, Cell Fate, and Differentiation. *Genetics* **2019**, *213*, 1111, doi:10.1534/genetics.119.300974.
71. Gomez Ruiz, M.; Bate, M. Segregation of myogenic lineages in Drosophila requires numb. *Development* **1997**, *124*, 4857.
72. Liu, J.; Qian, L.; Wessells, R.J.; Bidet, Y.; Jagla, K.; Bodmer, R. Hedgehog and RAS pathways cooperate in the anterior–posterior specification and positioning of cardiac progenitor cells. *Dev. Biol.* **2006**, *290*, 373–385, doi:10.1016/j.ydbio.2005.11.033.
73. Rushton, E.; Drysdale, R.; Abmayr, S.M.; Michelson, A.M.; Bate, M. Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for Drosophila muscle development. *Development* **1995**, *121*, 1979.
74. Prokop, A.; Landgraf, M.; Rushton, E.; Broadie, K.; Bate, M. Presynaptic Development at the Drosophila Neuromuscular Junction: Assembly and Localization of Presynaptic Active Zones. *Neuron* **1996**, *17*, 617–626, doi:10.1016/S0896-6273(00)80195-6.
75. Halfon, M.S.; Carmena, A.; Gisselbrecht, S.; Sackerson, C.M.; Jiménez, F.; Baylies, M.K.; Michelson, A.M. Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors. *Cell* **2000**, *103*, 63–74, doi:10.1016/S0092-8674(00)00105-7.
76. Cox, V.T.; Beckett, K.; Baylies, M.K. Delivery of wingless to the ventral mesoderm by the developing central nervous system ensures proper patterning of individual slouch-positive muscle progenitors. *Dev. Biol.* **2005**, *287*, 403–415, doi:10.1016/j.ydbio.2005.09.013.
77. Dohrmann, C.; Azpiazu, N.; Frasch, M. A new Drosophila homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. *Genes Dev.* **1990**, *4*, 2098–2111, doi:10.1101/gad.4.12a.2098.
78. Bourgouin, C.; Lundgren, S.E.; Thomas, J.B. apterous is a drosophila LIM domain gene required for the development of a subset of embryonic muscles. *Neuron* **1992**, *9*, 549–561, doi:10.1016/0896-6273(92)90192-G.
79. Keller, C.A.; Grill, M.A.; Abmayr, S.M. A Role for nautilus in the Differentiation of Muscle Precursors. *Dev. Biol.* **1998**, *202*, 157–171, doi:10.1006/dbio.1998.9009.
80. Knirr, S.; Azpiazu, N.; Frasch, M. The role of the NK-homeobox gene slouch (S59) in somatic muscle patterning. *Development* **1999**, *126*, 4525.
81. Boukhatmi, H.; Frendo, J.L.; Enriquez, J.; Crozatier, M.; Dubois, L.; Vincent, A. Tup/Islet1 integrates time and position to specify muscle identity in Drosophila. *Development* **2012**, *139*, 3572, doi:10.1242/dev.083410.

82. Crozatier, M.; Vincent, A. Requirement for the Drosophila COE transcription factor Collier in formation of an embryonic muscle: Transcriptional response to notch signalling. *Development* **1999**, *126*, 1495.
83. Busser, B.W.; Shokri, L.; Jaeger, S.A.; Gisselbrecht, S.S.; Singhania, A.; Berger, M.F.; Zhou, B.; Bulyk, M.L.; Michelson, A.M. Molecular mechanism underlying the regulatory specificity of a Drosophila homeodomain protein that specifies myoblast identity. *Dev. Camb. Engl.* **2012**, *139*, 1164–1174, doi:10.1242/dev.077362.
84. Busser, B.W.; Gisselbrecht, S.S.; Shokri, L.; Tansey, T.R.; Gamble, C.E.; Bulyk, M.L.; Michelson, A.M. Contribution of distinct homeodomain DNA binding specificities to Drosophila embryonic mesodermal cell-specific gene expression programs. *PLoS ONE* **2013**, *8*, e69385, doi:10.1371/journal.pone.0069385.
85. Dubois, L.; Frendo, J.-L.; Chanut-Delalande, H.; Crozatier, M.; Vincent, A. Genetic dissection of the Transcription Factor code controlling serial specification of muscle identities in Drosophila. *eLife* **2016**, *5*, e14979, doi:10.7554/eLife.14979.
86. Deng, H.; Bell, J.B.; Simmonds, A.J. Vestigial is required during late-stage muscle differentiation in Drosophila melanogaster embryos. *Mol. Biol. Cell* **2010**, *21*, 3304–3316, doi:10.1091/mbc.E10-04-0364.
87. Carrasco-Rando, M.; Tutor, A.S.; Prieto-Sánchez, S.; González-Pérez, E.; Barrios, N.; Letizia, A.; Martín, P.; Campuzano, S.; Ruiz-Gómez, M. Drosophila araucan and caupolican integrate intrinsic and signalling inputs for the acquisition by muscle progenitors of the lateral transverse fate. *PLoS Genet.* **2011**, *7*, e1002186, doi:10.1371/journal.pgen.1002186.
88. Enriquez, J.; de Taffin, M.; Crozatier, M.; Vincent, A.; Dubois, L. Combinatorial coding of Drosophila muscle shape by Collier and Nautilus. *Dev. Biol.* **2012**, *363*, 27–39, doi:10.1016/j.ydbio.2011.12.018.
89. Nose, A.; Isshiki, T.; Takeichi, M. Regional specification of muscle progenitors in Drosophila: The role of the msh homeobox gene. *Development* **1998**, *125*, 215.
90. Lord, P.C.W.; Lin, M.-H.; Hales, K.H.; Storti, R.V. Normal Expression and the Effects of Ectopic Expression of the Drosophila muscle segment homeobox (msh) Gene Suggest a Role in Differentiation and Patterning of Embryonic Muscles. *Dev. Biol.* **1995**, *171*, 627–640, doi:10.1006/dbio.1995.1310.
91. Knirr, S.; Frasch, M. Molecular Integration of Inductive and Mesoderm-Intrinsic Inputs Governs even-skipped Enhancer Activity in a Subset of Pericardial and Dorsal Muscle Progenitors. *Dev. Biol.* **2001**, *238*, 13–26, doi:10.1006/dbio.2001.0397.
92. Fujioka, M.; Wessells, R.J.; Han, Z.; Liu, J.; Fitzgerald, K.; Yusibova, G.L.; Zamora, M.; Ruiz-Lozano, P.; Bodmer, R.; Jaynes, J.B. Embryonic even skipped-dependent muscle and heart cell fates are required for normal adult activity, heart function, and lifespan. *Circ. Res.* **2005**, *97*, 1108–1114, doi:10.1161/01.RES.0000191546.08532.B2.
93. Ruiz-Gomez, M.; Romani, S.; Hartmann, C.; Jackle, H.; Bate, M. Specific muscle identities are regulated by Kruppel during Drosophila embryogenesis. *Development* **1997**, *124*, 3407.
94. Jagla, T.; Bellard, F.; Lutz, Y.; Dretzen, G.; Bellard, M.; Jagla, K. ladybird determines cell fate decisions during diversification of Drosophila somatic muscles. *Development* **1998**, *125*, 3699.
95. Müller, D.; Jagla, T.; Bodart, L.M.; Jährling, N.; Dodt, H.-U.; Jagla, K.; Frasch, M. Regulation and functions of the lms homeobox gene during development of embryonic lateral transverse muscles and direct flight muscles in Drosophila. *PLoS ONE* **2010**, *5*, e14323, doi:10.1371/journal.pone.0014323.
96. Kumar, R.P.; Dobi, K.C.; Baylies, M.K.; Abmayr, S.M. Muscle cell fate choice requires the T-box transcription factor midline in Drosophila. *Genetics* **2015**, *199*, 777–791, doi:10.1534/genetics.115.174300.
97. Corbin, V.; Michelson, A.M.; Abmayr, S.M.; Neel, V.; Alcamo, E.; Maniatis, T.; Young, M.W. A role for the Drosophila neurogenic genes in mesoderm differentiation. *Cell* **1991**, *67*, 311–323, doi:10.1016/0092-8674(91)90183-Y.
98. Schaub, C.; Nagaso, H.; Jin, H.; Frasch, M. Org-1, the Drosophila ortholog of Tbx1, is a direct activator of known identity genes during muscle specification. *Development* **2012**, *139*, 1001, doi:10.1242/dev.073890.
99. Duan, H.; Zhang, C.; Chen, J.; Sink, H.; Frei, E.; Noll, M. A key role of Pox meso in somatic myogenesis of Drosophila. *Development* **2007**, *134*, 3985, doi:10.1242/dev.008821.
100. Vorbrüggen, G.; Constien, R.; Zilian, O.; Wimmer, E.A.; Dowe, G.; Taubert, H.; Noll, M.; Jäckle, H. Embryonic expression and characterization of a Ptx1 homolog in Drosophila. *Mech. Dev.* **1997**, *68*, 139–147, doi:10.1016/S0925-4773(97)00139-1.

101. Drysdale, R.; Rushton, E.; Bate, M. Genes required for embryonic muscle development in Drosophila melanogaster A survey of the X chromosome. *Rouxs Arch. Dev. Biol. Off. Organ EDBO* **1993**, *202*, 276–295, doi:10.1007/BF00363217.
102. Liu, Y.-H.; Jakobsen, J.S.; Valentin, G.; Amarantos, I.; Gilmour, D.T.; Furlong, E.E.M. A Systematic Analysis of Tinman Function Reveals Eya and JAK-STAT Signaling as Essential Regulators of Muscle Development. *Dev. Cell* **2009**, *16*, 280–291, doi:10.1016/j.devcel.2009.01.006.
103. Clark, I.B.N.; Boyd, J.; Hamilton, G.; Finnegan, D.J.; Jarman, A.P. D-six4 plays a key role in patterning cell identities deriving from the Drosophila mesoderm. *Dev. Biol.* **2006**, *294*, 220–231, doi:10.1016/j.ydbio.2006.02.044.
104. Jagla, T.; Bidet, Y.; Da Ponte, J.P.; Dastugue, B.; Jagla, K. Cross-repressive interactions of identity genes are essential for proper specification of cardiac and muscular fates in Drosophila. *Development* **2002**, *129*, 1037.
105. Dobi, K.C.; Halfon, M.S.; Baylies, M.K. Whole-Genome Analysis of Muscle Founder Cells Implicates the Chromatin Regulator Sin3A in Muscle Identity. *Cell Rep.* **2014**, *8*, 858–870, doi:10.1016/j.celrep.2014.07.005.
106. Capovilla, M.; Kambris, Z.; Botas, J. Direct regulation of the muscle-identity gene apterous by a Hox protein in the somatic mesoderm. *Development* **2001**, *128*, 1221.
107. Michelson, A.M. Muscle pattern diversification in Drosophila is determined by the autonomous function of homeotic genes in the embryonic mesoderm. *Development* **1994**, *120*, 755.
108. Enriquez, J.; Boukhatmi, H.; Dubois, L.; Philippakis, A.A.; Bulyk, M.L.; Michelson, A.M.; Crozatier, M.; Vincent, A. Multi-step control of muscle diversity by Hox proteins in the Drosophila embryo. *Dev. Camb. Engl.* **2010**, *137*, 457–466, doi:10.1242/dev.045286.
109. Domsch, K.; Carnesecchi, J.; Disela, V.; Friedrich, J.; Trost, N.; Ermakova, O.; Polychronidou, M.; Lohmann, I. The Hox transcription factor Ubx stabilizes lineage commitment by suppressing cellular plasticity in Drosophila. *eLife* **2019**, *8*, e42675, doi:10.7554/eLife.42675.
110. Hessinger, C.; Technau, G.M.; Rogulja-Ortmann, A. The Drosophila Hox gene Ultrabithorax acts in both muscles and motoneurons to orchestrate formation of specific neuromuscular connections. *Development* **2017**, *144*, 139, doi:10.1242/dev.143875.
111. Junion, G.; Bataillé, L.; Jagla, T.; Da Ponte, J.P.; Tapin, R.; Jagla, K. Genome-wide view of cell fate specification: Ladybird acts at multiple levels during diversification of muscle and heart precursors. *Genes Dev.* **2007**, *21*, 3163–3180, doi:10.1101/gad.437307.
112. Bataillé, L.; Delon, I.; Da Ponte, J.P.; Brown, N.H.; Jagla, K. Downstream of Identity Genes: Muscle-Type-Specific Regulation of the Fusion Process. *Dev. Cell* **2010**, *19*, 317–328, doi:10.1016/j.devcel.2010.07.008.
113. Black, B.L.; Olson, E.N. Transcriptional control of muscle development by myocyte enhancer factor-2 (Mef2) proteins. *Annu. Rev. Cell Dev. Biol.* **1998**, *14*, 167–196, doi:10.1146/annurev.cellbio.14.1.167.
114. Pon, J.R.; Marra, M.A. MEF2 transcription factors: Developmental regulators and emerging cancer genes. *Oncotarget* **2016**, *7*, 2297–2312, doi:10.18632/oncotarget.6223.
115. Taylor, M.V.; Hughes, S.M. Mef2 and the skeletal muscle differentiation program. *Skelet. Muscle Dev. 30th Anniv. MyoD* **2017**, *72*, 33–44, doi:10.1016/j.semcdb.2017.11.020.
116. Bour, B.A.; O'Brien, M.A.; Lockwood, W.L.; Goldstein, E.S.; Bodmer, R.; Taghert, P.H.; Abmayr, S.M.; Nguyen, H.T. Drosophila MEF2, a transcription factor that is essential for myogenesis. *Genes Dev.* **1995**, *9*, 730–741, doi:10.1101/gad.9.6.730.
117. Ranganayakulu, G.; Zhao, B.; Dokidis, A.; Molkentin, J.D.; Olson, E.N.; Schulz, R.A. A Series of Mutations in the D-MEF2 Transcription Factor Reveal Multiple Functions in Larval and Adult Myogenesis in Drosophila. *Dev. Biol.* **1995**, *171*, 169–181, doi:10.1006/dbio.1995.1269.
118. Elgar, S.J.; Han, J.; Taylor, M.V. mef2 activity levels differentially affect gene expression during Drosophila muscle development. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 918–923, doi:10.1073/pnas.0711255105.
119. Cunha, P.M.F.; Sandmann, T.; Gustafson, E.H.; Ciglar, L.; Eichenlaub, M.P.; Furlong, E.E.M. Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity. *PLoS Genet.* **2010**, *6*, e1001014, doi:10.1371/journal.pgen.1001014.

120. Junion, G.; Jagla, T.; Duplant, S.; Tapin, R.; Da Ponte, J.-P.; Jagla, K. Mapping Dmef2-binding regulatory modules by using a ChIP-enriched in silico targets approach. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18479, doi:10.1073/pnas.0507030102.
121. Tanaka, K.K.K.; Bryantsev, A.L.; Cripps, R.M. Myocyte Enhancer Factor 2 and Chorion Factor 2 Collaborate in Activation of the Myogenic Program in Drosophila. *Mol. Cell. Biol.* **2008**, *28*, 1616, doi:10.1128/MCB.01169-07.
122. García-Zaragoza, E.; Mas, J.A.; Vivar, J.; Arredondo, J.J.; Cervera, M. CF2 activity and enhancer integration are required for proper muscle gene expression in Drosophila. *Mech. Dev.* **2008**, *125*, 617–630, doi:10.1016/j.mod.2008.03.003.
123. Cripps, R.M.; Lovato, T.L.; Olson, E.N. Positive autoregulation of the Myocyte enhancer factor-2 myogenic control gene during somatic muscle development in Drosophila. *Dev. Biol.* **2004**, *267*, 536–547, doi:10.1016/j.ydbio.2003.12.004.
124. Nguyen, H.T.; Xu, X. Drosophila mef2 Expression during Mesoderm Development Is Controlled by a Complex Array ofcis-Acting Regulatory Modules. *Dev. Biol.* **1998**, *204*, 550–566, doi:10.1006/dbio.1998.9081.
125. Chen, Z.; Liang, S.; Zhao, Y.; Han, Z. miR-92b regulates Mef2 levels through a negative-feedback circuit during Drosophila muscle development. *Dev. Camb. Engl.* **2012**, *139*, 3543–3552, doi:10.1242/dev.082719.
126. Cripps, R.M.; Black, B.L.; Zhao, B.; Lien, C.L.; Schulz, R.A.; Olson, E.N. The myogenic regulatory gene Mef2 is a direct target for transcriptional activation by Twist during Drosophila myogenesis. *Genes Dev.* **1998**, *12*, 422–434, doi:10.1101/gad.12.3.422.
127. Nowak, S.J.; Aihara, H.; Gonzalez, K.; Nibu, Y.; Baylies, M.K. Akirin links twist-regulated transcription with the Brahma chromatin remodeling complex during embryogenesis. *PLoS Genet.* **2012**, *8*, e1002547, doi:10.1371/journal.pgen.1002547.
128. Wang, F.; Minakhina, S.; Tran, H.; Changela, N.; Kramer, J.; Steward, R. Tet protein function during Drosophila development. *PLoS ONE* **2018**, *13*, e0190367, doi:10.1371/journal.pone.0190367.
129. Deng, H.; Hughes, S.C.; Bell, J.B.; Simmonds, A.J. Alternative requirements for Vestigial, Scalloped, and Dmef2 during muscle differentiation in Drosophila melanogaster. *Mol. Biol. Cell* **2009**, *20*, 256–269, doi:10.1091/mbc.e08-03-0288.
130. Haralalka, S.; Abmayr, S.M. Myoblast fusion in Drosophila. *Exp. Cell Res.* **2010**, *316*, 3007–3013, doi:10.1016/j.yexcr.2010.05.018.
131. Deng, S.; Azevedo, M.; Baylies, M. Acting on identity: Myoblast fusion and the formation of the syncytial muscle fiber. *Semin. Cell Dev. Biol.* **2017**, *72*, 45–55, doi:10.1016/j.semcdb.2017.10.033.
132. Lee, D.M.; Chen, E.H. Drosophila Myoblast Fusion: Invasion and Resistance for the Ultimate Union. *Annu. Rev. Genet.* **2019**, *53*, 67–91, doi:10.1146/annurev-genet-120116-024603.
133. Ruiz-Gómez, M.; Coutts, N.; Price, A.; Taylor, M.V.; Bate, M. Drosophila Dumbfounded: A Myoblast Attractant Essential for Fusion. *Cell* **2000**, *102*, 189–198, doi:10.1016/S0092-8674(00)00024-6.
134. Strünkelnberg, M.; Bonengel, B.; Moda, L.M.; Hertenstein, A.; de Couet, H.G.; Ramos, R.G.P.; Fischbach, K.-F. rst and its paralogue kirre act redundantly during embryonic muscle development in Drosophila. *Development* **2001**, *128*, 4229.
135. Bour, B.A.; Chakravarti, M.; West, J.M.; Abmayr, S.M. Drosophila SNS, a member of the immunoglobulin superfamily that is essential for myoblast fusion. *Genes Dev.* **2000**, *14*, 1498–1511.
136. Artero, R.D.; Castanon, I.; Baylies, M.K. The immunoglobulin-like protein Hibris functions as a dose-dependent regulator of myoblast fusion and is differentially controlled by Ras and Notch signaling. *Development* **2001**, *128*, 4251.
137. Bataillé, L.; Boukhatmi, H.; Frendo, J.-L.; Vincent, A. Dynamics of transcriptional (re)-programming of syncytial nuclei in developing muscles. *BMC Biol.* **2017**, *15*, doi:10.1186/s12915-017-0386-2.
138. Azevedo, M.; Baylies, M.K. Getting into Position: Nuclear Movement in Muscle Cells. *Trends Cell Biol.* **2020**, *30*, 303–316, doi:10.1016/j.tcb.2020.01.002.
139. Roman, W.; Gomes, E.R. Nuclear positioning in skeletal muscle. *SI Nucl. Position.* **2018**, *82*, 51–56, doi:10.1016/j.semcdb.2017.11.005.
140. Cadot, B.; Gache, V.; Gomes, E.R. Moving and positioning the nucleus in skeletal muscle - one step at a time. *Nucl. Austin Tex* **2015**, *6*, 373–381, doi:10.1080/19491034.2015.1090073.

141. Starr, D.A.; Fridolfsson, H.N. Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 421–444, doi:10.1146/annurev-cellbio-100109-104037.
142. Folker, E.S.; Baylies, M.K. Nuclear positioning in muscle development and disease. *Front. Physiol.* **2013**, *4*, 363–363, doi:10.3389/fphys.2013.00363.
143. Volk, T. Singling out Drosophila tendon cells: A dialogue between two distinct cell types. *Trends Genet.* **1999**, *15*, 448–453, doi:10.1016/S0168-9525(99)01862-4.
144. Schnorrer, F.; Dickson, B.J. Muscle Building: Mechanisms of Myotube Guidance and Attachment Site Selection. *Dev. Cell* **2004**, *7*, 9–20, doi:10.1016/j.devcel.2004.06.010.
145. Subramanian, A.; Schilling, T.F. Tendon development and musculoskeletal assembly: Emerging roles for the extracellular matrix. *Dev. Camb. Engl.* **2015**, *142*, 4191–4204, doi:10.1242/dev.114777.
146. Lejard, V.; Blais, F.; Guerquin, M.-J.; Bonnet, A.; Bonnin, M.-A.; Havis, E.; Malbouyres, M.; Bidaud, C.B.; Maro, G.; Gilardi-Hebenstreit, P.; et al. EGR1 and EGR2 involvement in vertebrate tendon differentiation. *J. Biol. Chem.* **2011**, *286*, 5855–5867, doi:10.1074/jbc.M110.153106.
147. Volohonsky, G.; Edenfeld, G.; Klämbt, C.; Volk, T. Muscle-dependent maturation of tendon cells is induced by post-transcriptional regulation of stripeA. *Development* **2007**, *134*, 347, doi:10.1242/dev.02735.
148. Kramer, S.G.; Kidd, T.; Simpson, J.H.; Goodman, C.S. Switching Repulsion to Attraction: Changing Responses to Slit During Transition in Mesoderm Migration. *Science* **2001**, *292*, 737, doi:10.1126/science.1058766.
149. Callahan, C.A.; Bonkovsky, J.L.; Scully, A.L.; Thomas, J.B. derailed is required for muscle attachment site selection in Drosophila. *Development* **1996**, *122*, 2761.
150. Schnorrer, F.; Kalchhauser, I.; Dickson, B.J. The Transmembrane Protein Kon-tiki Couples to Dgrip to Mediate Myotube Targeting in Drosophila. *Dev. Cell* **2007**, *12*, 751–766, doi:10.1016/j.devcel.2007.02.017.
151. Estrada, B.; Gisselbrecht, S.S.; Michelson, A.M. The transmembrane protein Perdido interacts with Grip and integrins to mediate myotube projection and attachment in the Drosophila embryo. *Development* **2007**, *134*, 4469, doi:10.1242/dev.014027.
152. Swan, L.E.; Schmidt, M.; Schwarz, T.; Ponimaskin, E.; Prange, U.; Boeckers, T.; Thomas, U.; Sigrist, S.J. Complex interaction of Drosophila GRIP PDZ domains and Echinoid during muscle morphogenesis. *EMBO J.* **2006**, *25*, 3640–3651, doi:10.1038/sj.emboj.7601216.
153. Chanana, B.; Graf, R.; Koledachkina, T.; Pflanz, R.; Vorbrüggen, G. αPS2 integrin-mediated muscle attachment in Drosophila requires the ECM protein Thrombospondin. *Mech. Dev.* **2007**, *124*, 463–475, doi:10.1016/j.mod.2007.03.005.
154. Katzemich, A.; Long, J.Y.; Panneton, V.; Fisher, L.A.B.; Hipfner, D.; Schöck, F. Slik phosphorylation of Talin T152 is crucial for proper Talin recruitment and maintenance of muscle attachment in Drosophila. *Development* **2019**, *146*, dev176339, doi:10.1242/dev.176339.
155. Pines, M.; Das, R.; Ellis, S.J.; Morin, A.; Czerniecki, S.; Yuan, L.; Klose, M.; Coombs, D.; Tanentzapf, G. Mechanical force regulates integrin turnover in Drosophila in vivo. *Nat. Cell Biol.* **2012**, *14*, 935–943, doi:10.1038/ncb2555.
156. Sparrow, J.C.; Schöck, F. The initial steps of myofibril assembly: Integrins pave the way. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 293–298, doi:10.1038/nrm2634.
157. Gautel, M.; Djinović-Carugo, K. The sarcomeric cytoskeleton: From molecules to motion. *J. Exp. Biol.* **2016**, *219*, 135, doi:10.1242/jeb.124941.
158. Rhee, D.; Sanger, J.M.; Sanger, J.W. The premyofibril: Evidence for its role in myofibrillogenesis. *Cell Motil. Cytoskeleton* **1994**, *28*, 1–24, doi:10.1002/cm.970280102.
159. Jirka, C.; Pak, J.H.; Grosgogeat, C.A.; Marchetii, M.M.; Gupta, V.A. Dysregulation of NRAP degradation by KLHL41 contributes to pathophysiology in nemaline myopathy. *Hum. Mol. Genet.* **2019**, *28*, 2549–2560, doi:10.1093/hmg/ddz078.
160. Antin, P.B.; Tokunaka, S.; Nachmias, V.T.; Holtzer, H. Role of stress fiber-like structures in assembling nascent myofibrils in myosheets recovering from exposure to ethyl methanesulfonate. *J. Cell Biol.* **1986**, *102*, 1464–1479, doi:10.1083/jcb.102.4.1464.
161. Epstein, H.; Fischman, D. Molecular analysis of protein assembly in muscle development. *Science* **1991**, *251*, 1039–1044, doi:10.1126/science.1998120.

162. Rui, Y.; Bai, J.; Perrimon, N. Sarcomere formation occurs by the assembly of multiple latent protein complexes. *PLoS Genet.* **2010**, *6*, e1001208, doi:10.1371/journal.pgen.1001208.
163. Weitkunat, M.; Kaya-Çopur, A.; Grill, S.W.; Schnorrer, F. Tension and Force-Resistant Attachment Are Essential for Myofibrillogenesis in Drosophila Flight Muscle. *Curr. Biol.* **2014**, *24*, 705–716, doi:10.1016/j.cub.2014.02.032.
164. Orfanos, Z.; Leonard, K.; Elliott, C.; Katzemich, A.; Bullard, B.; Sparrow, J. Sallimus and the Dynamics of Sarcomere Assembly in Drosophila Flight Muscles. *J. Mol. Biol.* **2015**, *427*, 2151–2158, doi:10.1016/j.jmb.2015.04.003.
165. Röper, K.; Mao, Y.; Brown, N.H. Contribution of sequence variation in Drosophila actins to their incorporation into actin-based structures in vivo. *J. Cell Sci.* **2005**, *118*, 3937, doi:10.1242/jcs.02517.
166. Bulgakova, N.A.; Wellmann, J.; Brown, N.H. Diverse integrin adhesion stoichiometries caused by varied actomyosin activity. *Open Biol.* **2017**, *7*, 160250, doi:10.1098/rsob.160250.
167. Zappia, M.P.; Frolov, M.V. E2F function in muscle growth is necessary and sufficient for viability in Drosophila. *Nat. Commun.* **2016**, *7*, 10509, doi:10.1038/ncomms10509.
168. Molnár, I.; Migh, E.; Szikora, S.; Kalmár, T.; Végh, A.G.; Deák, F.; Barkó, S.; Bugyi, B.; Orfanos, Z.; Kovács, J.; et al. DAAM is required for thin filament formation and Sarcomerogenesis during muscle development in Drosophila. *PLoS Genet.* **2014**, *10*, e1004166, doi:10.1371/journal.pgen.1004166.
169. Bai, J.; Hartwig, J.H.; Perrimon, N. SALS, a WH2-Domain-Containing Protein, Promotes Sarcomeric Actin Filament Elongation from Pointed Ends during Drosophila Muscle Growth. *Dev. Cell* **2007**, *13*, 828–842, doi:10.1016/j.devcel.2007.10.003.
170. Mardahl-Dumesnil, M.; Fowler, V.M. Thin filaments elongate from their pointed ends during myofibril assembly in Drosophila indirect flight muscle. *J. Cell Biol.* **2001**, *155*, 1043–1053, doi:10.1083/jcb.200108026.
171. Bloor, J.W.; Kiehart, D.P. zipper Nonmuscle Myosin-II Functions Downstream of PS2 Integrin in Drosophila Myogenesis and Is Necessary for Myofibril Formation. *Dev. Biol.* **2001**, *239*, 215–228, doi:10.1006/dbio.2001.0452.
172. Volk, T.; Fessler, L.I.; Fessler, J.H. A role for integrin in the formation of sarcomeric cytoarchitecture. *Cell* **1990**, *63*, 525–536, doi:10.1016/0092-8674(90)90449-O.
173. Kreiskôther, N.; Reichert, N.; Buttgereit, D.; Hertenstein, A.; Fischbach, K.-F.; Renkawitz-Pohl, R. Drosophila Rolling pebbles colocalises and putatively interacts with alpha-Actinin and the Sls isoform Zormin in the Z-discs of the sarcomere and with Dumbfounded/Kirre, alpha-Actinin and Zormin in the terminal Z-discs. *J. Muscle Res. Cell Motil.* **2006**, *27*, 93, doi:10.1007/s10974-006-9060-y.
174. Kelemen-Valkony, I.; Kiss, M.; Csiha, J.; Kiss, A.; Bircher, U.; Szidonya, J.; Maróy, P.; Juhász, G.; Komonyi, O.; Csiszár, K.; et al. Drosophila basement membrane collagen col4a1 mutations cause severe myopathy. *Matrix Biol.* **2012**, *31*, 29–37, doi:10.1016/j.matbio.2011.09.004.
175. Katzemich, A.; Liao, K.A.; Czerniecki, S.; Schöck, F. Alp/Enigma family proteins cooperate in Z-disc formation and myofibril assembly. *PLoS Genet.* **2013**, *9*, e1003342, doi:10.1371/journal.pgen.1003342.
176. Crisp, S.J.; Evers, J.F.; Bate, M. Endogenous patterns of activity are required for the maturation of a motor network. *J. Neurosci. Off. J. Soc. Neurosci.* **2011**, *31*, 10445–10450, doi:10.1523/JNEUROSCI.0346-11.2011.
177. Broadie, K.; Bate, M. Development of the embryonic neuromuscular synapse of Drosophila melanogaster. *J. Neurosci.* **1993**, *13*, 144, doi:10.1523/JNEUROSCI.13-01-00144.1993.
178. Keshishian, H.; Broadie, K.; Chiba, A.; Bate, M. The Drosophila Neuromuscular Junction: A Model System for Studying Synaptic Development and Function. *Annu. Rev. Neurosci.* **1996**, *19*, 545–575, doi:10.1146/annurev.ne.19.030196.002553.
179. Ruiz-Cañada, C.; Budnik, V. Introduction on The Use of The Drosophila Embryonic/Larval Neuromuscular Junction as A Model System to Study Synapse Development and Function, and A Brief Summary of Pathfinding and Target Recognition. In *International Review of Neurobiology*; Academic Press: New York, NY, USA, 2006; pp. 1–31 ISBN 0074-7742.
180. Landgraf, M.; Bossing, T.; Technau, G.M.; Bate, M. The origin, location, and projections of the embryonic abdominal motorneurons of Drosophila. *J. Neurosci. Off. J. Soc. Neurosci.* **1997**, *17*, 9642–9655, doi:10.1523/JNEUROSCI.17-24-09642.1997.
181. Schmid, A.; Chiba, A.; Doe, C.Q. Clonal analysis of Drosophila embryonic neuroblasts: Neural cell types, axon projections and muscle targets. *Development* **1999**, *126*, 4653.

182. Landgraf, M.; Jeffrey, V.; Fujioka, M.; Jaynes, J.B.; Bate, M. Embryonic Origins of a Motor System: Motor Dendrites Form a Myotopic Map in Drosophila. *PLOS Biol.* **2003**, *1*, e41, doi:10.1371/journal.pbio.0000041.
183. Ritzenthaler, S.; Suzuki, E.; Chiba, A. Postsynaptic filopodia in muscle cells interact with innervating motoneuron axons. *Nat. Neurosci.* **2000**, *3*, 1012–1017, doi:10.1038/79833.
184. Nose, A. Generation of neuromuscular specificity in Drosophila: Novel mechanisms revealed by new technologies. *Front. Mol. Neurosci.* **2012**, *5*, 62, doi:10.3389/fnmol.2012.00062.
185. Nose, A.; Umeda, T.; Takeichi, M. Neuromuscular target recognition by a homophilic interaction of connectin cell adhesion molecules in Drosophila. *Development* **1997**, *124*, 1433.
186. Patel, N.H.; Ball, E.E.; Goodman, C.S. Changing role of even-skipped during the evolution of insect pattern formation. *Nature* **1992**, *357*, 339–342, doi:10.1038/357339a0.
187. Labrador, J.P.; O'Keefe, D.; Yoshikawa, S.; McKinnon, R.D.; Thomas, J.B.; Bashaw, G.J. The Homeobox Transcription Factor Even-skipped Regulates Netrin-Receptor Expression to Control Dorsal Motor-Axon Projections in Drosophila. *Curr. Biol.* **2005**, *15*, 1413–1419, doi:10.1016/j.cub.2005.06.058.
188. Chen, K.; Featherstone, D.E. Discs-large (DLG) is clustered by presynaptic innervation and regulates postsynaptic glutamate receptor subunit composition in Drosophila. *BMC Biol.* **2005**, *3*, 1–13, doi:10.1186/1741-7007-3-1.
189. Bachmann, A.; Kobler, O.; Kittel, R.J.; Wichmann, C.; Sierralta, J.; Sigrist, S.J.; Gundelfinger, E.D.; Knust, E.; Thomas, U. A perisynaptic ménage à trois between Dlg, DLin-7, and Metro controls proper organization of Drosophila synaptic junctions. *J. Neurosci. Off. J. Soc. Neurosci.* **2010**, *30*, 5811–5824, doi:10.1523/JNEUROSCI.0778-10.2010.
190. Broadie, K.; Bate, M. Innervation directs receptor synthesis and localization in Drosophila embryo synaptogenesis. *Nature* **1993**, *361*, 350–353, doi:10.1038/361350a0.
191. Kim, M.D.; Wen, Y.; Jan, Y.-N. Patterning and organization of motor neuron dendrites in the Drosophila larva. *Dev. Biol.* **2009**, *336*, 213–221, doi:10.1016/j.ydbio.2009.09.041.
192. Syed, A.; Lukacsovich, T.; Pomeroy, M.; Bardwell, A.J.; Decker, G.T.; Waymire, K.G.; Purcell, J.; Huang, W.; Gui, J.; Padilla, E.M.; et al. Miles to go (mtgo) encodes FNDC3 proteins that interact with the chaperonin subunit CCT3 and are required for NMJ branching and growth in Drosophila. *Dev. Biol.* **2019**, *445*, 37–53, doi:10.1016/j.ydbio.2018.10.016.
193. Beumer, K.J.; Rohrbough, J.; Prokop, A.; Broadie, K. A role for PS integrins in morphological growth and synaptic function at the postembryonic neuromuscular junction of Drosophila. *Development* **1999**, *126*, 5833.
194. Vonhoff, F.; Keshishian, H. In Vivo Calcium Signaling during Synaptic Refinement at the Drosophila Neuromuscular Junction. *J. Neurosci. Off. J. Soc. Neurosci.* **2017**, *37*, 5511–5526, doi:10.1523/JNEUROSCI.2922-16.2017.
195. DePew, A.T.; Aimino, M.A.; Mosca, T.J. The Tenets of Teneurin: Conserved Mechanisms Regulate Diverse Developmental Processes in the Drosophila Nervous System. *Front. Neurosci.* **2019**, *13*, 27, doi:10.3389/fnins.2019.00027.
196. Bate, M.; Rushton, E.; Currie, D.A. Cells with persistent twist expression are the embryonic precursors of adult muscles in Drosophila. *Development* **1991**, *113*, 79.
197. Figeac, N.; Jagla, T.; Aradhya, R.; Da Ponte, J.P.; Jagla, K. Drosophila adult muscle precursors form a network of interconnected cells and are specified by the rhomboid-triggered EGF pathway. *Development* **2010**, *137*, 1965, doi:10.1242/dev.049080.
198. Lavergne, G.; Zmojdzian, M.; Da Ponte, J.P.; Junion, G.; Jagla, K. Drosophila adult muscle precursor cells contribute to motor axon pathfinding and proper innervation of embryonic muscles. *Development* **2020**, *147*, dev183004, doi:10.1242/dev.183004.
199. Gunage, R.D.; Dhanyasi, N.; Reichert, H.; VijayRaghavan, K. Drosophila adult muscle development and regeneration. *Skelet. Muscle Dev. 30th Anniv. MyoD* **2017**, *72*, 56–66, doi:10.1016/j.semcdb.2017.11.017.
200. Laurichesse, Q.; Soler, C. Muscle development : A view from adult myogenesis in Drosophila. *Semin. Cell Dev. Biol.* **2020**, doi:10.1016/j.semcdb.2020.02.009.
201. Oas, S.T.; Bryantsev, A.L.; Cripps, R.M. Arrest is a regulator of fiber-specific alternative splicing in the indirect flight muscles of Drosophila. *J. Cell Biol.* **2014**, *206*, 895–908, doi:10.1083/jcb.201405058.

202. DeAguero, A.A.; Castillo, L.; Oas, S.T.; Kiani, K.; Bryantsev, A.L.; Cripps, R.M. Regulation of fiber-specific actin expression by the Drosophila SRF ortholog Blistered. *Dev. Camb. Engl.* **2019**, *146*, dev164129, doi:10.1242/dev.164129.
203. Schiaffino, S.; Reggiani, C. Fiber Types in Mammalian Skeletal Muscles. *Physiol. Rev.* **2011**, *91*, 1447–1531, doi:10.1152/physrev.00031.2010.
204. Hastings, G.A.; Emerson, C.P., Jr Myosin functional domains encoded by alternative exons are expressed in specific thoracic muscles of Drosophila. *J. Cell Biol.* **1991**, *114*, 263–276, doi:10.1083/jcb.114.2.263.
205. Figeac, N.; Daczewska, M.; Marcelle, C.; Jagla, K. Muscle stem cells and model systems for their investigation. *Dev. Dyn.* **2007**, *236*, 3332–3342, doi:10.1002/dvdy.21345.
206. Beira, J.V.; Paro, R. The legacy of Drosophila imaginal discs. *Chromosoma* **2016**, *125*, 573–592, doi:10.1007/s00412-016-0595-4.
207. Aradhya, R.; Zmojdzian, M.; Da Ponte, J.P.; Jagla, K. Muscle niche-driven Insulin-Notch-Myc cascade reactivates dormant Adult Muscle Precursors in Drosophila. *eLife* **2015**, *4*, e08497, doi:10.7554/eLife.08497.
208. Tavi, P.; Korhonen, T.; Hänninen, S.L.; Bruton, J.D.; Lööf, S.; Simon, A.; Westerblad, H. Myogenic skeletal muscle satellite cells communicate by tunnelling nanotubes. *J. Cell. Physiol.* **2010**, *223*, 376–383, doi:10.1002/jcp.22044.
209. Gunage, R.D.; Reichert, H.; VijayRaghavan, K. Identification of a new stem cell population that generates Drosophila flight muscles. *eLife* **2014**, *3*, e03126, doi:10.7554/eLife.03126.
210. Vishal, K.; Brooks, D.S.; Bawa, S.; Gameros, S.; Stetsiv, M.; Geisbrecht, E.R. Adult Muscle Formation Requires Drosophila Moleskin for Proliferation of Wing Disc-Associated Muscle Precursors. *Genetics* **2017**, *206*, 199, doi:10.1534/genetics.116.193813.
211. Bernard, F.; Dutriaux, A.; Silber, J.; Lalouette, A. Notch pathway repression by vestigial is required to promote indirect flight muscle differentiation in Drosophila melanogaster. *Dev. Biol.* **2006**, *295*, 164–177, doi:10.1016/j.ydbio.2006.03.022.
212. Dutta, D.; Umashankar, M.; Lewis, Edward, B.; Rodrigues, V.; VijayRaghavan, K. Hox Genes Regulate Muscle Founder Cell Pattern Autonomously and Regulate Morphogenesis Through Motor Neurons. *J. Neurogenet.* **2010**, *24*, 95–108, doi:10.3109/01677063.2010.494317.
213. Sudarsan, V.; Anant, S.; Guptan, P.; VijayRaghavan, K.; Skaer, H. Myoblast Diversification and Ectodermal Signaling in Drosophila. *Dev. Cell* **2001**, *1*, 829–839, doi:10.1016/S1534-5807(01)00089-2.
214. Maqbool, T.; Soler, C.; Jagla, T.; Daczewska, M.; Lodha, N.; Palliyil, S.; VijayRaghavan, K.; Jagla, K. Shaping leg muscles in Drosophila: Role of ladybird, a conserved regulator of appendicular myogenesis. *PLoS ONE* **2006**, *1*, e122, doi:10.1371/journal.pone.0000122.
215. Ghazi, A.; Anant, S.; Vijay Raghavan, K. Apterous mediates development of direct flight muscles autonomously and indirect flight muscles through epidermal cues. *Development* **2000**, *127*, 5309.
216. Dutta, D.; Shaw, S.; Maqbool, T.; Pandya, H.; VijayRaghavan, K. Drosophila Heartless Acts with Heartbroken/Dof in Muscle Founder Differentiation. *PLOS Biol.* **2005**, *3*, e337, doi:10.1371/journal.pbio.0030337.
217. Dutta, D.; Anant, S.; Ruiz-Gomez, M.; Bate, M.; VijayRaghavan, K. Founder myoblasts and fibre number during adult myogenesis in Drosophila. *Development* **2004**, *131*, 3761, doi:10.1242/dev.01249.
218. Bryantsev, A.L.; Duong, S.; Brunetti, T.M.; Chechenova, M.B.; Lovato, T.L.; Nelson, C.; Shaw, E.; Uhl, J.D.; Gebelein, B.; Cripps, R.M. Extradenticle and homothorax control adult muscle fiber identity in Drosophila. *Dev. Cell* **2012**, *23*, 664–673, doi:10.1016/j.devcel.2012.08.004.
219. Schönbauer, C.; Distler, J.; Jährling, N.; Radolf, M.; Dodt, H.-U.; Frasch, M.; Schnorrer, F. Spalt mediates an evolutionarily conserved switch to fibrillar muscle fate in insects. *Nature* **2011**, *479*, 406–409, doi:10.1038/nature10559.
220. DeSimone, S.; Coelho, C.; Roy, S.; VijayRaghavan, K.; White, K. ERECT WING, the Drosophila member of a family of DNA binding proteins is required in imaginal myoblasts for flight muscle development. *Development* **1996**, *122*, 31.
221. Fernandes, J.; Bate, M.; Vijayraghavan, K. Development of the indirect flight muscles of Drosophila. *Development* **1991**, *113*, 67.
222. Kuleesha, Y.; Puah, W.C.; Wasser, M. Live imaging of muscle histolysis in Drosophila metamorphosis. *BMC Dev. Biol.* **2016**, *16*, 12–12, doi:10.1186/s12861-016-0113-1.

223. Fernandes, J.J.; Keshishian, H. Motoneurons regulate myoblast proliferation and patterning in Drosophila. *Dev. Biol.* **2005**, *277*, 493–505, doi:10.1016/j.ydbio.2004.09.038.
224. Currie, D.A.; Bate, M. The development of adult abdominal muscles in Drosophila: Myoblasts express twist and are associated with nerves. *Development* **1991**, *113*, 91.
225. Rivlin, P.K.; Schneiderman, A.M.; Booker, R. Imaginal Pioneers Prefigure the Formation of Adult Thoracic Muscles in Drosophila melanogaster. *Dev. Biol.* **2000**, *222*, 450–459, doi:10.1006/dbio.2000.9676.
226. Mukherjee, P.; Gildor, B.; Shilo, B.-Z.; VijayRaghavan, K.; Schejter, E.D. The actin nucleator WASp is required for myoblast fusion during adult Drosophila myogenesis. *Dev. Camb. Engl.* **2011**, *138*, 2347–2357, doi:10.1242/dev.055012.
227. Ghazi, A.; Paul, L.; VijayRaghavan, K. Prepattern genes and signaling molecules regulate stripe expression to specify Drosophila flight muscle attachment sites. *Mech. Dev.* **2003**, *120*, 519–528, doi:10.1016/S0925-4773(03)00042-X.
228. Laddada, L.; Jagla, K.; Soler, C. Odd-skipped and Stripe act downstream of Notch to promote the morphogenesis of long appendicular tendons in Drosophila. *Biol. Open* **2019**, *8*, bio038760, doi:10.1242/bio.038760.
229. Weitkunat, M.; Brasse, M.; Bausch, A.R.; Schnorrer, F. Mechanical tension and spontaneous muscle twitching precede the formation of cross-striated muscle in vivo. *Dev. Camb. Engl.* **2017**, *144*, 1261–1272, doi:10.1242/dev.140723.
230. Spletter, M.L.; Schnorrer, F. Transcriptional regulation and alternative splicing cooperate in muscle fiber-type specification in flies and mammals. *Dev. Biol.* **2014**, *321*, 90–98, doi:10.1016/j.yexcr.2013.10.007.
231. Reedy, M.C.; Beall, C. Ultrastructure of Developing Flight Muscle in Drosophila. I. Assembly of Myofibrils. *Dev. Biol.* **1993**, *160*, 443–465, doi:10.1006/dbio.1993.1320.
232. Spletter, M.L.; Barz, C.; Yeroslaviz, A.; Zhang, X.; Lemke, S.B.; Bonnard, A.; Brunner, E.; Cardone, G.; Basler, K.; Habermann, B.H.; et al. A transcriptomics resource reveals a transcriptional transition during ordered sarcomere morphogenesis in flight muscle. *eLife* **2018**, *7*, e34058, doi:10.7554/eLife.34058.
233. Shwartz, A.; Dhanyasi, N.; Schejter, E.D.; Shilo, B.-Z. The Drosophila formin Fhos is a primary mediator of sarcomeric thin-filament array assembly. *eLife* **2016**, *5*, e16540, doi:10.7554/eLife.16540.
234. Fernandes, I.; Schöck, F. The nebulin repeat protein Lasp regulates I-band architecture and filament spacing in myofibrils. *J. Cell Biol.* **2014**, *206*, 559–572, doi:10.1083/jcb.201401094.
235. Qiu, F.; Brendel, S.; Cunha, P.M.F.; Astola, N.; Song, B.; Furlong, E.E.M.; Leonard, K.R.; Bullard, B. Myofilin, a protein in the thick filaments of insect muscle. *J. Cell Sci.* **2005**, *118*, 1527, doi:10.1242/jcs.02281.
236. Ball, E.; Karlik, C.C.; Beall, C.J.; Saville, D.L.; Sparrow, J.C.; Bullard, B.; Fyrberg, E.A. Arthrin, a myofibrillar protein of insect flight muscle, is an actin-ubiquitin conjugate. *Cell* **1987**, *51*, 221–228, doi:10.1016/0092-8674(87)90149-8.
237. Becker, K.D.; O'Donnell, P.T.; Heitz, J.M.; Vito, M.; Bernstein, S.I. Analysis of Drosophila paramyosin: Identification of a novel isoform which is restricted to a subset of adult muscles. *J. Cell Biol.* **1992**, *116*, 669–681, doi:10.1083/jcb.116.3.669.
238. Maroto, M.; Arredondo, J.J.; Román, M.S.; Marco, R.; Cervera, M. Analysis of the Paramyosin/Miniparamyosin Gene: MINIPARAMYOSIN IS AN INDEPENDENTLY TRANSCRIBED, DISTINCT PARAMYOSIN ISOFORM, WIDELY DISTRIBUTED IN INVERTEBRATES. *J. Biol. Chem.* **1995**, *270*, 4375–4382, doi:10.1074/jbc.270.9.4375.
239. Vigoreaux, J.O. Alterations in flightin phosphorylation inDrosophila flight muscles are associated with myofibrillar defects engendered by actin and myosin heavy-chain mutant alleles. *Biochem. Genet.* **1994**, *32*, 301–314, doi:10.1007/BF00555832.
240. Reedy, M.C.; Bullard, B.; Vigoreaux, J.O. Flightin is essential for thick filament assembly and sarcomere stability in Drosophila flight muscles. *J. Cell Biol.* **2000**, *151*, 1483–1500, doi:10.1083/jcb.151.7.1483.
241. Craig, R.; Woodhead, J.L. Structure and function of myosin filaments. *Curr. Opin. Struct. Biol.* **2006**, *16*, 204–212, doi:10.1016/j.sbi.2006.03.006.
242. Orfanos, Z.; Sparrow, J.C. Myosin isoform switching during assembly of the Drosophila flight muscle thick filament lattice. *J. Cell Sci.* **2013**, *126*, 139, doi:10.1242/jcs.110361.
243. Contompasis, J.L.; Nyland, L.R.; Maughan, D.W.; Vigoreaux, J.O. Flightin Is Necessary for Length Determination, Structural Integrity, and Large Bending Stiffness of Insect Flight Muscle Thick Filaments. *J. Mol. Biol.* **2010**, *395*, 340–348, doi:10.1016/j.jmb.2009.11.021.

244. Sparrow, J.; Reedy, M.; Ball, E.; Kyrtatas, V.; Molloy, J.; Durston, J.; Hennessey, E.; White, D. Functional and ultrastructural effects of a missense mutation in the indirect flight muscle-specific actin gene of Drosophila melanogaster. *J. Mol. Biol.* **1991**, *222*, 963–982, doi:10.1016/0022-2836(91)90588-W.
245. Burkart, C.; Qiu, F.; Brendel, S.; Benes, V.; Hååg, P.; Labeit, S.; Leonard, K.; Bullard, B. Modular Proteins from the Drosophila sallimus (sls) Gene and their Expression in Muscles with Different Extensibility. *J. Mol. Biol.* **2007**, *367*, 953–969, doi:10.1016/j.jmb.2007.01.059.
246. González-Morales, N.; Xiao, Y.S.; Schilling, M.A.; Marescal, O.; Liao, K.A.; Schöck, F. Myofibril diameter is set by a finely tuned mechanism of protein oligomerization in Drosophila. *eLife* **2019**, *8*, e50496, doi:10.7554/eLife.50496.
247. Truman, J.W.; Schuppe, H.; Shepherd, D.; Williams, D.W. Developmental architecture of adult-specific lineages in the ventral CNS of Drosophila. *Development* **2004**, *131*, 5167, doi:10.1242/dev.01371.
248. Brierley, D.J.; Blanc, E.; Reddy, O.V.; Vijayraghavan, K.; Williams, D.W. Dendritic targeting in the leg neuropil of Drosophila: The role of midline signalling molecules in generating a myotopic map. *PLoS Biol.* **2009**, *7*, e1000199, doi:10.1371/journal.pbio.1000199.
249. Enriquez, J.; Venkatasubramanian, L.; Baek, M.; Peterson, M.; Aghayeva, U.; Mann, R.S. Specification of Individual Adult Motor Neuron Morphologies by Combinatorial Transcription Factor Codes. *Neuron* **2015**, *86*, 955–970, doi:10.1016/j.neuron.2015.04.011.
250. Baek, M.; Mann, R.S. Lineage and Birth Date Specify Motor Neuron Targeting and Dendritic Architecture in Adult Drosophila. *J. Neurosci.* **2009**, *29*, 6904, doi:10.1523/JNEUROSCI.1585-09.2009.
251. Fernandes, J.J.; Keshishian, H. Nerve-muscle interactions during flight muscle development in Drosophila. *Development* **1998**, *125*, 1769.
252. Fernandes, J.; VijayRaghavan, K. The development of indirect flight muscle innervation in Drosophila melanogaster. *Development* **1993**, *118*, 215.
253. Rival, T.; Soustelle, L.; Cattaert, D.; Strambi, C.; Iché, M.; Birman, S. Physiological requirement for the glutamate transporter dEAAT1 at the adult Drosophila neuromuscular junction. *J. Neurobiol.* **2006**, *66*, 1061–1074, doi:10.1002/neu.20270.
254. Kimura, K.I.; Truman, J.W. Postmetamorphic cell death in the nervous and muscular systems of Drosophila melanogaster. *J. Neurosci. Off. J. Soc. Neurosci.* **1990**, *10*, 403–411, doi:10.1523/JNEUROSCI.10-02-00403.1990.
255. Soler, C.; Taylor, M.V. The Him gene inhibits the development of Drosophila flight muscles during metamorphosis. *Mech. Dev.* **2009**, *126*, 595–603, doi:10.1016/j.mod.2009.03.003.
256. Soler, C.; Han, J.; Taylor, M.V. The conserved transcription factor Mef2 has multiple roles in adult Drosophila musculature formation. *Development* **2012**, *139*, 1270, doi:10.1242/dev.077875.
257. Zhang, S.; Bernstein, S.I. Spatially and temporally regulated expression of myosin heavy chain alternative exons during Drosophila embryogenesis. *Mech. Dev.* **2001**, *101*, 35–45, doi:10.1016/S0925-4773(00)00549-9.
258. Swank, D.M.; Bartoo, M.L.; Knowles, A.F.; Iliffe, C.; Bernstein, S.I.; Molloy, J.E.; Sparrow, J.C. Alternative exon-encoded regions of Drosophila myosin heavy chain modulate ATPase rates and actin sliding velocity. *J. Biol. Chem.* **2001**, *276*, 15117–15124, doi:10.1074/jbc.M008379200.
259. Karlik, C.C.; Fyrberg, E.A. Two Drosophila melanogaster tropomyosin genes: Structural and functional aspects. *Mol. Cell. Biol.* **1986**, *6*, 1965, doi:10.1128/MCB.6.6.1965.
260. Mateos, J.; Herranz, R.; Domingo, A.; Sparrow, J.; Marco, R. The structural role of high molecular weight tropomyosins in dipteran indirect flight muscle and the effect of phosphorylation. *J. Muscle Res. Cell Motil.* **2006**, *27*, 189–201, doi:10.1007/s10974-005-9044-3.
261. Katzemich, A.; Long, J.Y.; Jani, K.; Lee, B.R.; Schöck, F. Muscle type-specific expression of Zasp52 isoforms in Drosophila. *Gene Expr. Patterns* **2011**, *11*, 484–490, doi:10.1016/j.gep.2011.08.004.
262. Katzemich, A.; Kreisköther, N.; Alexandrovich, A.; Elliott, C.; Schöck, F.; Leonard, K.; Sparrow, J.; Bullard, B. The function of the M-line protein obscurin in controlling the symmetry of the sarcomere in the flight muscle of Drosophila. *J. Cell Sci.* **2012**, *125*, 3367, doi:10.1242/jcs.097345.
263. Marín, M.-C.; Rodríguez, J.-R.; Ferrús, A. Transcription of Drosophila Troponin I Gene Is Regulated by Two Conserved, Functionally Identical, Synergistic Elements. *Mol. Biol. Cell* **2004**, *15*, 1185–1196, doi:10.1091/mbc.e03-09-0663.

264. Mas, J.-A.; García-Zaragoza, E.; Cervera, M. Two Functionally Identical Modular Enhancers in Drosophila Troponin T Gene Establish the Correct Protein Levels in Different Muscle Types. *Mol. Biol. Cell* **2004**, *15*, 1931–1945, doi:10.1091/mbc.e03-10-0729.
265. Gremke, L.; Lord, P.C.W.; Sabacan, L.; Lin, S.-C.; Wohlwill, A.; Storti, R.V. Coordinate Regulation of Drosophila Tropomyosin Gene Expression Is Controlled by Multiple Muscle-Type-Specific Positive and Negative Enhancer Elements. *Dev. Biol.* **1993**, *159*, 513–527, doi:10.1006/dbio.1993.1260.
266. Lin, M.H.; Nguyen, H.T.; Dybala, C.; Storti, R.V. Myocyte-specific enhancer factor 2 acts cooperatively with a muscle activator region to regulate Drosophila tropomyosin gene muscle expression. *Proc. Natl. Acad. Sci.* USA **1996**, *93*, 4623, doi:10.1073/pnas.93.10.4623.
267. Reddy, K.L.; Wohlwill, A.; Dzitoeva, S.; Lin, M.-H.; Holbrook, S.; Storti, R.V. The Drosophila PAR Domain Protein 1 (Pdp1) Gene Encodes Multiple Differentially Expressed mRNAs and Proteins through the Use of Multiple Enhancers and Promoters. *Dev. Biol.* **2000**, *224*, 401–414, doi:10.1006/dbio.2000.9797.
268. Vigoreaux, J.O.; Saide, J.D.; Pardue, M.L. Structurally different Drosophila striated muscles utilize distinct variants of Z-band-associated proteins. *J. Muscle Res. Cell Motil.* **1991**, *12*, 340–354, doi:10.1007/BF01738589.
269. Zhao, C.; Swank, D.M. The Drosophila indirect flight muscle myosin heavy chain isoform is insufficient to transform the jump muscle into a highly stretch-activated muscle type. *Am. J. Physiol. Cell Physiol.* **2017**, *312*, C111–C118, doi:10.1152/ajpcell.00284.2016.
270. Vigoreaux, J.O.; Perry, L.M. Multiple isoelectric variants of flightin in Drosophila stretch-activated muscles are generated by temporally regulated phosphorylations. *J. Muscle Res. Cell Motil.* **1994**, *15*, 607–616, doi:10.1007/BF00121068.
271. Daley, J.; Southgate, R.; Ayme-Southgate, A. Structure of the Drosophila projectin protein: Isoforms and implication for projectin filament assembly11Edited by M. F. Moody. *J. Mol. Biol.* **1998**, *279*, 201–210, doi:10.1006/jmbi.1998.1756.
272. Glasheen, B.M.; Ramanath, S.; Patel, M.; Sheppard, D.; Puthawala, J.T.; Riley, L.A.; Swank, D.M. Five Alternative Myosin Converter Domains Influence Muscle Power, Stretch Activation, and Kinetics. *Biophys. J.* **2018**, *114*, 1142–1152, doi:10.1016/j.bpj.2017.12.045.
273. Belozerov, V.E.; Ratkovic, S.; McNeill, H.; Hilliker, A.J.; McDermott, J.C. In vivo interaction proteomics reveal a novel p38 mitogen-activated protein kinase/Rack1 pathway regulating proteostasis in Drosophila muscle. *Mol. Cell. Biol.* **2014**, *34*, 474–484, doi:10.1128/MCB.00824-13.
274. Haas, K.F.; Woodruff, E., 3rd; Broadie, K. Proteasome function is required to maintain muscle cellular architecture. *Biol. Cell* **2007**, *99*, 615–626, doi:10.1042/BC20070019.
275. Nguyen, H.T.; Voza, F.; Ezzeddine, N.; Frasch, M. Drosophila mind bomb2 is required for maintaining muscle integrity and survival. *J. Cell Biol.* **2007**, *179*, 219–227, doi:10.1083/jcb.200708135.
276. Valdez, C.; Scroggs, R.; Chassen, R.; Reiter, L.T. Variation in Dube3a expression affects neurotransmission at the Drosophila neuromuscular junction. *Biol. Open* **2015**, *4*, 776–782, doi:10.1242/bio.20148045.
277. Marco-Ferreres, R.; Arredondo, J.J.; Fraile, B.; Cervera, M. Overexpression of troponin T in Drosophila muscles causes a decrease in the levels of thin-filament proteins. *Biochem. J.* **2005**, *386*, 145–152, doi:10.1042/BJ20041240.
278. Firdaus, H.; Mohan, J.; Naz, S.; Arathi, P.; Ramesh, S.R.; Nongthomba, U. A cis-regulatory mutation in troponin-I of Drosophila reveals the importance of proper stoichiometry of structural proteins during muscle assembly. *Genetics* **2015**, *200*, 149–165, doi:10.1534/genetics.115.175604.
279. Wójtowicz, I.; Jabłońska, J.; Zmojdzian, M.; Taghli-Lamallem, O.; Renaud, Y.; Junion, G.; Daczewska, M.; Huelsmann, S.; Jagla, K.; Jagla, T. Drosophila small heat shock protein CryAB ensures structural integrity of developing muscles, and proper muscle and heart performance. *Development* **2015**, *142*, 994, doi:10.1242/dev.115352.
280. Jabłońska, J.; Dubińska-Magiera, M.; Jagla, T.; Jagla, K.; Daczewska, M. Drosophila Hsp67Bc hot-spot variants alter muscle structure and function. *Cell. Mol. Life Sci. CMLS* **2018**, *75*, 4341–4356, doi:10.1007/s00018-018-2875-z.
281. Wang, S.; Reuveny, A.; Volk, T. Nesprin provides elastic properties to muscle nuclei by cooperating with spectraplakin and EB1. *J. Cell Biol.* **2015**, *209*, 529–538, doi:10.1083/jcb.201408098.

282. Wang, S.; Volk, T. Composite biopolymer scaffolds shape muscle nucleus: Insights and perspectives from Drosophila. *Bioarchitecture* **2015**, *5*, 35–43, doi:10.1080/19490992.2015.1106061.
283. Lorber, D.; Rotkopf, R.; Volk, T. In vivo imaging of myonuclei during spontaneous muscle contraction reveals non-uniform nuclear mechanical dynamics in Nesprin/klar mutants. *bioRxiv* **2019**, 643015, doi:10.1101/643015.
284. Elhanany-Tamir, H.; Yu, Y.V.; Shnayder, M.; Jain, A.; Welte, M.; Volk, T. Organelle positioning in muscles requires cooperation between two KASH proteins and microtubules. *J. Cell Biol.* **2012**, *198*, 833–846, doi:10.1083/jcb.201204102.
285. Morel, V.; Lepicard, S.N.; Rey, A.; Parmentier, M.-L.; Schaeffer, L. Drosophila Nesprin-1 controls glutamate receptor density at neuromuscular junctions. *Cell. Mol. Life Sci.* **2014**, *71*, 3363–3379, doi:10.1007/s00018-014-1566-7.
286. Goel, P.; Dufour Bergeron, D.; Böhme, M.A.; Nunnelly, L.; Lehmann, M.; Buser, C.; Walter, A.M.; Sigrist, S.J.; Dickman, D. Homeostatic scaling of active zone scaffolds maintains global synaptic strength. *J. Cell Biol.* **2019**, *218*, 1706–1724, doi:10.1083/jcb.201807165.
287. Goel, P.; Dickman, D. Distinct homeostatic modulations stabilize reduced postsynaptic receptivity in response to presynaptic DLK signaling. *Nat. Commun.* **2018**, *9*, 1856, doi:10.1038/s41467-018-04270-0.
288. Ziegler, A.B.; Augustin, H.; Clark, N.L.; Berthelot-Grosjean, M.; Simonnet, M.M.; Steinert, J.R.; Geillon, F.; Manière, G.; Featherstone, D.E.; Grosjean, Y. The Amino Acid Transporter JhI-21 Coevolves with Glutamate Receptors, Impacts NMJ Physiology, and Influences Locomotor Activity in Drosophila Larvae. *Sci. Rep.* **2016**, *6*, 19692, doi:10.1038/srep19692.
289. Hong, H.; Zhao, K.; Huang, S.; Huang, S.; Yao, A.; Jiang, Y.; Sigrist, S.; Zhao, L.; Zhang, Y.Q. Structural Remodeling of Active Zones Is Associated with Synaptic Homeostasis. *J. Neurosci.* **2020**, *40*, 2817, doi:10.1523/JNEUROSCI.2002-19.2020.
290. James, T.D.; Zwiefelhofer, D.J.; Frank, C.A. Maintenance of homeostatic plasticity at the Drosophila neuromuscular synapse requires continuous IP3-directed signaling. *eLife* **2019**, *8*, e39643, doi:10.7554/eLife.39643.
291. Mourikis, P.; Gopalakrishnan, S.; Sambasivan, R.; Tajbakhsh, S. Cell-autonomous Notch activity maintains the temporal specification potential of skeletal muscle stem cells. *Dev. Camb. Engl.* **2012**, *139*, 4536–4548, doi:10.1242/dev.084756.
292. Boukhatmi, H.; Bray, S. A population of adult satellite-like cells in Drosophila is maintained through a switch in RNA-isoforms. *eLife* **2018**, *7*, e35954, doi:10.7554/eLife.35954.
293. Siles, L.; Ninfali, C.; Cortés, M.; Darling, D.S.; Postigo, A. ZEB1 protects skeletal muscle from damage and is required for its regeneration. *Nat. Commun.* **2019**, *10*, 1364, doi:10.1038/s41467-019-08983-8.
294. Kuang, S.; Kuroda, K.; Le Grand, F.; Rudnicki, M.A. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **2007**, *129*, 999–1010, doi:10.1016/j.cell.2007.03.044.
295. de Haro, M.; Al-Ramahi, I.; De Gouyon, B.; Ukani, L.; Rosa, A.; Faustino, N.A.; Ashizawa, T.; Cooper, T.A.; Botas, J. MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Drosophila model of myotonic dystrophy type 1. *Hum. Mol. Genet.* **2006**, *15*, 2138–2145, doi:10.1093/hmg/ddl137.
296. Picchio, L.; Plantie, E.; Renaud, Y.; Poovthumkadavil, P.; Jagla, K. Novel Drosophila model of myotonic dystrophy type 1: Phenotypic characterization and genome-wide view of altered gene expression. *Hum. Mol. Genet.* **2013**, *22*, 2795–2810, doi:10.1093/hmg/ddt127.
297. Yatsenko, A.S.; Shcherbata, H.R. Drosophila miR-9a Targets the ECM Receptor Dystroglycan to Canalize Myotendinous Junction Formation. *Dev. Cell* **2014**, *28*, 335–348, doi:10.1016/j.devcel.2014.01.004.
298. Kucherenko, M.M.; Marrone, A.K.; Rishko, V.M.; Magliarelli, H. de F.; Shcherbata, H.R. Stress and muscular dystrophy: A genetic screen for Dystroglycan and Dystrophin interactors in Drosophila identifies cellular stress response components. *Dev. Biol.* **2011**, *352*, 228–242, doi:10.1016/j.ydbio.2011.01.013.
299. Dialynas, G.; Shrestha, O.K.; Ponce, J.M.; Zwerger, M.; Thiemann, D.A.; Young, G.H.; Moore, S.A.; Yu, L.; Lammerding, J.; Wallrath, L.L. Myopathic lamin mutations cause reductive stress and activate the nrf2/keap-1 pathway. *PLoS Genet.* **2015**, *11*, e1005231, doi:10.1371/journal.pgen.1005231.
300. Chandran, S.; Suggs, J.A.; Wang, B.J.; Han, A.; Bhide, S.; Cryderman, D.E.; Moore, S.A.; Bernstein, S.I.; Wallrath, L.L.; Melkani, G.C. Suppression of myopathic lamin mutations by muscle-specific activation

of AMPK and modulation of downstream signaling. *Hum. Mol. Genet.* **2019**, *28*, 351–371, doi:10.1093/hmg/ddy332.
301. Jungbluth, H.; Gautel, M. Pathogenic mechanisms in centronuclear myopathies. *Front. Aging Neurosci.* **2014**, *6*, 339, doi:10.3389/fnagi.2014.00339.
302. Schulman, V.K.; Folker, E.S.; Rosen, J.N.; Baylies, M.K. Syd/JIP3 and JNK signaling are required for myonuclear positioning and muscle function. *PLoS Genet.* **2014**, *10*, e1004880, doi:10.1371/journal.pgen.1004880.
303. Rosen, J.N.; Azevedo, M.; Soffar, D.B.; Boyko, V.P.; Brendel, M.B.; Schulman, V.K.; Baylies, M.K. The Drosophila Ninein homologue Bsg25D cooperates with Ensconsin in myonuclear positioning. *J. Cell Biol.* **2019**, *218*, 524–540, doi:10.1083/jcb.201808176.
304. Naimi, B.; Harrison, A.; Cummins, M.; Nongthomba, U.; Clark, S.; Canal, I.; Ferrus, A.; Sparrow, J.C. A Tropomyosin-2 Mutation Suppresses a Troponin I Myopathy inDrosophila. *Mol. Biol. Cell* **2001**, *12*, 1529–1539, doi:10.1091/mbc.12.5.1529.
305. Dahl-Halvarsson, M.; Olive, M.; Pokrzywa, M.; Ejeskär, K.; Palmer, R.H.; Uv, A.E.; Tajsharghi, H. Drosophila model of myosin myopathy rescued by overexpression of a TRIM-protein family member. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E6566–E6575, doi:10.1073/pnas.1800727115.
306. Ugur, B.; Chen, K.; Bellen, H.J. Drosophila tools and assays for the study of human diseases. *Dis. Model. Mech.* **2016**, *9*, 235–244, doi:10.1242/dmm.023762.
307. Day, K.; Shefer, G.; Shearer, A.; Yablonka-Reuveni, Z. The depletion of skeletal muscle satellite cells with age is concomitant with reduced capacity of single progenitors to produce reserve progeny. *Dev. Biol.* **2010**, *340*, 330–343, doi:10.1016/j.ydbio.2010.01.006.
308. Jiang, C.; Wen, Y.; Kuroda, K.; Hannon, K.; Rudnicki, M.A.; Kuang, S. Notch signaling deficiency underlies age-dependent depletion of satellite cells in muscular dystrophy. *Dis. Models Amp Mech.* **2014**, *7*, 997, doi:10.1242/dmm.015917.
309. Letsou, A.; Bohmann, D. Small flies—Big discoveries: Nearly a century of Drosophila genetics and development. *Dev. Dyn.* **2005**, *232*, 526–528, doi:10.1002/dvdy.20307.
310. Bellen, H.J.; Tong, C.; Tsuda, H. 100 years of Drosophila research and its impact on vertebrate neuroscience: A history lesson for the future. *Nat. Rev. Neurosci.* **2010**, *11*, 514–522, doi:10.1038/nrn2839.

1.4.3. Muscle diversification and the mystery and complexity of muscle identity transcription factors (iTFs)

Current knowledge in vertebrate myogenesis suggests that coordination between signaling pathways and transcription factors contributes to the identity of muscle subsets. It can be gathered from what is known in *Drosophila* myogenesis that a specific code of transcription factors known as muscle identity transcription factors (iTFs) contributes to the individual muscle identity of muscles within subsets that express some iTFs in common (Table 1 in the review article in Section 1.4.2). A gene is designated an iTF if its loss results in muscle subset or individual muscle phenotypes such as missing, deformed or supernumerary muscles as well as the transformation of one muscle into another muscle type. To summarize the muscle diversification process, muscle progenitor cells are formed within promuscular clusters in specific locations in each hemisegment. Each progenitor cell expresses a characteristic set of TFs. It undergoes asymmetric division to generate individual muscle founder cells (FCs). Immediately after formation, the initial TF expression pattern in the progenitor cell diversifies into an individual iTF code. Each FC is thought to contain the identity information for a single muscle. Abdominal hemisegments A2-A6 in the *Drosophila* embryo consist of a stereotypical pattern of 30 distinct muscles whose characteristic features defining their identity such as morphology, attachment and innervation pattern are well-known. This, thus, provides an excellent system to study individual muscle diversification. The NK-like homeobox TFs, Lateral muscles scarcer (Lms) and Slouch (Slou) or S59 are two iTFs expressed in non-overlapping muscle subsets that can specifically be targeted for studies. Lms is expressed in four lateral transverse (LT) muscles, LT1-4, whereas Slou is expressed in the VT1, VA2 and DT1 muscles. A large number of iTFs expressed in individual muscles have been identified, but they fail to explain all phenotypes observed, except in rare cases. I illustrate the complexity in deciphering the iTF code with what is currently known about the LT and Slou+ muscle subset iTFs below.

1.4.3.1. The LT muscle iTF code

Current understanding of the diversification of the expression pattern of LT muscle iTFs is highlighted in Figure 2 in the review article in Section 1.4.2. The TFs, Drop (Dr), Apterous (Ap), Krüppel (Kr), Midline (Mid), Araucan/Caupilican (Ara/Caup) and Lms are currently

known to contribute to LT identity. However, none of their mutants display 100% penetrant phenotypes indicating that there are unknown factors contributing to individual muscle identity.

1.4.3.1.1. Drop (Dr)

Drop (Dr) or Msh is a NK-like homeobox TF. Nose et al. (Nose, Isshiki, et Takeichi 1998) showed that it was expressed in the progenitors of the DO1, DO2, LT, VA2 and VA3 muscles. $msh^{\Delta 68}$ mutants with a P-element insertion giving rise to a deletion in the 5' end of *Dr* resulted in a partial, but substantial loss of LT muscles (also refered to as muscles 21-24) whereas an ectopic expression or misexpression using the early expressing 24B-Gal4 driver resulted in a dose-dependent phenotype. LT muscles were present, although a high-expressor *UASm25-m1* line led to more morphological defects than a low-expressor *UASm29-m1* line. Ectopic expression as well as mutations led to much more severe phenotypes in the DA1/2 (also referred to as muscles 9/10) muscles that also express it. An earlier study by Lord et al. (Lord et al. 1995) observed an absence of Slou expression on ectopic expression of Dr suggesting its role as a Slou repressor and providing a glimpse into the process of regulation of the muscle iTF code. Slou is an iTF playing a role in the identity of non-LT muscles DT1, VA2 and VT1. Thus, Dr appears to play a role in the diversification of LT FCs in a dose-dependent manner and cannot by itself regulate all identity characteristics of LT muscles.

1.4.3.1.2. Lateral muscles scarcer (Lms)

Lateral muscles scarcer (Lms) is a NK-like homebox TF. Müller et al. (Müller et al. 2010) showed that its mRNA was expressed in an LT-specific manner from stages where promuscular clusters are specified. Its expression is severely reduced in $msh^{\Delta 68}$ mutants lacking Dr function. On the contrary, in mutants for the gene *ladybird* (*lb*) that is normally expressed in the SBM muscle, its expression domain is expanded. lms^{S95} null mutants lacking the complete coding sequence display partially penetrant phenotypes of missing LT muscles. Thus, it appears to be downstream of Dr and is not singularly responsible for LT muscle identity.

1.4.3.1.3. Apterous (Ap)

Apterous (Ap) is a LIM-homeodomain TF. Bourgouin et al. (Bourgouin, Lundgren, et Thomas 1992) observed that *ap* mRNA as well as a LacZ enhancer trap reporter were expressed in LT muscle promuscular clusters, progenitors arising from them and developing LT muscles. They also noted expression in the VA2 and VA3 muscles (also referred to as muscles 27 and 29), but its role in these muscles is unknown. An *ap-P44* mutant line with a 5' end deletion led to

partially penetrant phenotypes of missing LT muscles. Misexpression using a heat-shock inductible Ap line, on the other hand, led to supernumerary LT muscles. Capovilla et al. (Capovilla, Kambris, et Botas 2001) identified a 680bp *ap* enhancer that they termed *apME680* in a long *ap* intron that drove LT-muscle specific expression in the promuscular clusters and continued into late developmental stages. They showed that the expression of an *apME680-LacZ* reporter was lost in the thoracic T2 and T3 segments in mutants for the *Hox* gene, *Antennapedia* (*Antp*). These hemisegments have a much stronger expression of the *LacZ* reporter compared to abdominal hemisegments. Müller et al. (Müller et al. 2010) observed that *ap* transcript expression appeared reduced in $msh^{\Delta 68}$ mutants lacking *Dr*. *lms* expression was relatively normal in ap^{UG035} null mutants. ap^{UG035}/lms^{S95} double mutants display partially penetrant missing muscle phenotypes similar to *lms* mutants. Ap, thus, appears to play a role in the specification of LT muscle FCs regulated by an LT-specific enhancer and its expression is potentially at least in part regulated by Dr. Thus, Ap, like Lms and Dr, is not singularly responsible for all LT muscle identity features.

1.4.3.1.4. Krüppel (Kr)

Krüppel (Kr) is a C2H2 zinc-finger TF. Ruiz-Gomez et al. (Ruiz-Gomez et al. 1997) noted the expression of *Kr* mRNA in specific promuscular clusters and that its expression was maintained in only one of the daughter FCs arising from progenitors. They observed its expression in the LT2 and LT4 muscles in stage 14 embryos. It was also expressed in other muscles including the Slou+ VA2 muscle. A $Kr^{l}Kr^{CD+}$ deficiency mutant that has a Kr^{l} lack-of-function allele in trans to a *Kr* minigene lacking enhancer regions necessary for late *Kr* expression caused the partial loss of LT muscles as well as what appear to be LT4-LT3 transformations resulting in four LT muscles at the same level instead of LT4 being located dorsal to LT1-3. It also resulted in transformations of a subset of dorsal muscles such as DA1 that expressed Kr into other muscles. Fujika et al. (Fujioka et al. 2005) noted a loss of Kr expression in dorsal muscles on the loss of mesodermal Even-skipped (Eve) expression. Eve is an iTF expressed in a subset of dorsal muscles. Thus, Kr dictates the identity of LT2 and LT4 muscles and its loss of function results in the partial loss of these muscles, again illustrating an incapability to regulate all identity characteristics on its own.

1.4.3.1.5. Midline (Mid)

Midline (Mid) is a T-box TF. Kumar et al. (R. P. Kumar et al. 2015) noted Mid expression in the progenitor of LT3/4 as well as the VA1/2 muscles along with Kr. Its expression was maintained only in LT4 and VA2. Mid is partially redundant with H15 and in *Df(2L)x528/Df(2L)GpdhA* mutants that lack both genes, a Kr+ LT4 FC could not be detected, but led to only a partial loss of LT3/4 muscles. Similarly, mid^{1} null mutants resulting from the insertion of a stop codon in its coding sequence led to the partial loss of LT3/4 muscles. Misexpressing Mid in the mesoderm with a 24B-Gal4 or Mef2-Gal4 driver led to an increased number of Kr+ cells. This indicates a positive regulatory effect of Mid on Kr and a role in specifying the identity of the LT3 and Kr+ LT4 muscles, but that is not the sole iTF for these muscles.

1.4.3.1.6. Araucan/Caupolican (Ara/Caup)

Araucan (Ara) and Caupolican (Caup) encode TALE homeobox TFs and are members of the conserved Iroquois complex. Carrasco-Rando et al. (Carrasco-Rando et al. 2011) observed that Caup was expressed in the LT muscles as well as Slou+ muscles such as DT1, VT1 and VA3 and the SBM muscle. They observed an earlier expression in LT muscles starting in promuscular clusters contrary to later expression in other muscles. In $Df(3L)iro^{DFM3}$ and $Df(3L)iro^{EGP6}$ larval lethal mutants caused by imprecise P-element excisions that delete both *ara* and *caup* genes present in the *iro* locus, they noted an almost 96% LT muscle loss and a duplication of the VA1 and VA2 muscles, but other Ara/Caup expressing muscles remained unaffected. They discovered that LT FCs segregated normally from progenitors, but were mis-specified with Slou expression appearing in Kr+ LT2 and LT4 FCs, but not in LT1 or LT3 FCs mimicking the VA1/2 FC specification where Kr is maintained with Slou only in VA2. Their study also showed that in WT embryos, there was a direct repression of *slou* by Caup in the LT promuscular cluster. This repression was blocked by the ectopic expression of an activated form of Ras, ras^{V12} leading them to conclude that *slou* repression was mediated by the Ras/MAPK pathway in LT muscle precursors. They hypothesized that the de-repression of *slou* led to the supernumerary VA1/2 muscles observed in *iro* mutants due to the transformation of LT progenitors into VA1/2 progenitors. No phenotypes were observed in single mutants for either *ara* or *caup* suggesting a redundant function between these factors. *lms* expression was lost in around 92% of the embryos and there was a gain of *slou* expression in 75% of the embryos. Ara/Caup, thus appear to be indispensable for defining the LT muscle identity at the promuscular cluster stage.

1.4.3.2. The Slou positive somatic muscle iTF code

Slouch (Slou) positive muscles include LO1 (muscle 5), VT1 (muscle 25), VA1/2 muscles, DT1 (muscle 18) and VA3 (muscle 29). LO1 and VA1/VA3 FCs express Slou transiently. Many LT muscle iTFs are also expressed in subsets of Slou+ muscles, although in-depth studies have not been realized in most cases. Dr is expressed in the VA2 and VA3 ventral muscles, although its role remains to be elucidated (Nose, Isshiki, et Takeichi 1998). Mid is expressed in the VA1/2 progenitor, maintained in VA2 and its loss of function leads to severe morphological defects in many ventral muscles (R. P. Kumar et al. 2015). Ap is expressed in the VA2 and VA3 muscles (Bourgouin, Lundgren, et Thomas 1992).

The non-LT dHLH TF and vertebrate MyoD homologue, Nautilus (Nau) is expressed in many progenitors (Wei, Rong, et Paterson 2007) including the VA1/2 progenitors and is retained in VA2 (Paterson et al. 1991).

1.4.3.2.1. Slouch (Slou)

Slouch (Slou) or S59 is a NK-like homeobox TF. It was the first identified muscle iTF (Dohrmann, Azpiazu, et Frasch 1990). Knirr et al. (Knirr, Azpiazu, et Frasch 1999) observed that Slou was expressed in three progenitors that give rise to the LO1, VT1, VA1/2, DT1 and VA3 muscles. A *slou*^{286} null mutant resulting from an ethyl methanesulfonate (EMS) mutation resulting in the insertion of a premature stop codon that generates a truncated protein led to the loss of LO1, VA3 and VT1, aberrant VA1/2 morphologies and a duplicated SBM (muscle 8). It is unclear if these phenotypes are fully penetrant. The authors mention a loss of *slou* mRNA in 85% of the mutants, 10% of which survive to adulthood, but are weak fliers and are short-lived. *slou*^{286} mutants were shown to display normal Kr expression patterns suggesting that it is probably upstream of Slou. Due to the duplication of the SBM muscle that normally expresses the iTF, Ladybird (Lb) (Jagla et al. 1998), the authors looked at the Lb expression pattern in *slou* mutants. They followed the fate of WT Slou+ progenitors in *slou* mutants using a *slou-LacZ* enhancer trap and noted that Lb was coexpressed with LacZ in FCs normally fated to become VT1 and LO1 muscles leading to a fate transformation to a SBM muscle instead. Overexpression of the SBM-specific Lb in turn has been shown to repress Slou expression (Junion et al. 2007). So, Slou appears indispensable for the specification of identity of specific muscle subsets with varying roles in distinct muscles judging from the phenotypes. It is required for the repression of Lb expression, but not for Kr expression.

1.4.3.2.2. Krüppel (Kr)

The LT iTF, Kr is expressed in the Slou+ VA2 muscle. Ruiz-Gomez et al. (Ruiz-Gomez et al. 1997) noted the expression of *Kr* mRNA in the promuscular cluster that gives rise to the progenitor for the VA1/2 muscles and that its expression was maintained in only one of the daughter FCs, VA2. It was coexpressed with Slou in VA1/2 progenitors. After the generation of VA1 and VA2 FCs, its expression declined from VA1 and was maintained in VA2. In VA2, it was co-expressed with Slou. The authors noted that in *Kr* mutants, *slou* expression was not maintained in the VA2 FC and was lost in 100% of embryos. They noticed a VA2 to VA1 fate transformation. Thus, Kr is required for the maintenance of *slou* expression to confer a VA2 identity in VA2 FCs.

1.4.3.2.3. Araucan/Caupolican (Ara/Caup)

Carrasco-Rando et al. (Carrasco-Rando et al. 2011) observed that the LT muscle iTFs, Ara and Caup were also expressed in the DT1 and SBM muscles. Unlike their expression in earlier stages in the LT promuscular clusters, they are expressed in the DT1 progenitor and SBM FC. In *iro* mutants, duplicated VA2 muscles that coexpress Slou and Kr were observed. This is probably due to LT fate transformation due to the gain of *slou* expression as mentioned in Section 1.4.3.1.6.

1.4.3.2.4. Optomotor-blind-related-gene-1 (Org-1)

Optomotor-blind-related-gene-1 (Org-1) is a vertebrate Tbx1 homologue and a T-box TF. Schaub et al. (Schaub et al. 2012) observed that it was expressed in the progenitors of muscles LO1/VT1 and SBM. It was initially expressed in all FCs for these muscles, but was subsequently retained only in VT1. An *org-1*^{OJ487} null mutant resulting from an imprecise P-element excision leading to the deletion of the first six exons led to missing or blob-like SBM and LO1 whereas VT1 presented morphological abnormalities. A pan-mesodermal misexpression led to a severely disrupted muscle pattern. *slou* and *lb* enhancer reporter expression were absent in *org-1* mutant contexts. Thus Org-1 acts upstream of *slou* to activate its expression and the determine VT1 fate.

1.4.3.2.5. Pox meso (Poxm)

Pox-meso (Poxm) is a PAX family Paired domain TF. Duan et al. (Duan et al. 2007) observed that it was expressed in the ventral and lateral mesoderm during early stages. Its expression then became restricted to the progenitors for DO3/DT1, VA1/VA2 and VA3 that coexpressed

slou. Its expression was subsequently lost from DO3. A $Poxm^{R361}$ null allele induced by an EMS point mutation resulting in the introduction of a premature stop codon and truncated protein led to a disrupted somatic muscle pattern with missing, duplicated and malformed muscles. A 24B-Gal4 driven pan mesodermal misexpression also led to an altered somatic muscle pattern with supernumerary muscles in the dorsal region that normally lacks Poxm expression, although the defects were less pronounced compared to *org-1* mutants. The authors observed that Poxm expression was driven by early and late enhancers, with late enhancers being responsible for the identity of the DT1 and VA1-3 muscles. Absence of late Poxm led to missing DT1 in around 68% of embryos. In *Poxm* mutants, *slou* expression was still present in the Poxm negative VT1 muscles, but was lost from VA2 and DT1. Dubois et al. noted that *Poxm* expression was lost from DO3/DT1 progenitors in *collier* (*col*) mutants. Thus, Poxm determines the fate of DT1 and VA1-3 muscles along with Slou and positively regulates *slou* expression.

1.4.3.3. The iTF code is complex and individual muscle identity codes are yet to be determined

Just looking at the two muscle subsets above highlights the complexity of the iTF code (Figure 11). These muscle subsets share common iTFs such as Kr, Dr and Ara/Caup, but they are not always expressed in all muscles in each subset. While multiple factors involved in specifying the identity of muscles subsets are known, most mutant phenotypes are only partially penetrant in muscle subsets expressing them (such as the *ap* mutant phenotypes) whereas others are fully penetrant (such as the VA2 phenotype in *Kr* mutants). In addition, phenotypes are not restricted to single muscles. *Kr* mutants, for example, affect the LT muscles as well as VA2 whereas *slou* mutants affect different Slou-expressing muscles to varying degrees. In addition, a gain or loss of iTFs also affects neighboring muscles in some cases, as seen in the case of loss of *slou* that causes an SBM duplication. Thus, additional factors contributing to individual muscle identity instead of muscle subsets still remain to be identified.

Some iTFs repress others to determine muscle identity, *slou* repression by Ara/Caup and Lb for example. There also appears to be a regulatory hierarchy in the iTF code with Poxm, for example, being upstream of *slou* in DT1 whereas org-1 activates *slou* in VT1 while Kr maintains *slou* expression in VA2. iTFs such as Kr and Slou are initially expressed in many FCs, but become restricted to a smaller subset at later stages. iTFs such as Poxm are regulated by early and late enhancers, which could be a common feature of many iTFs.

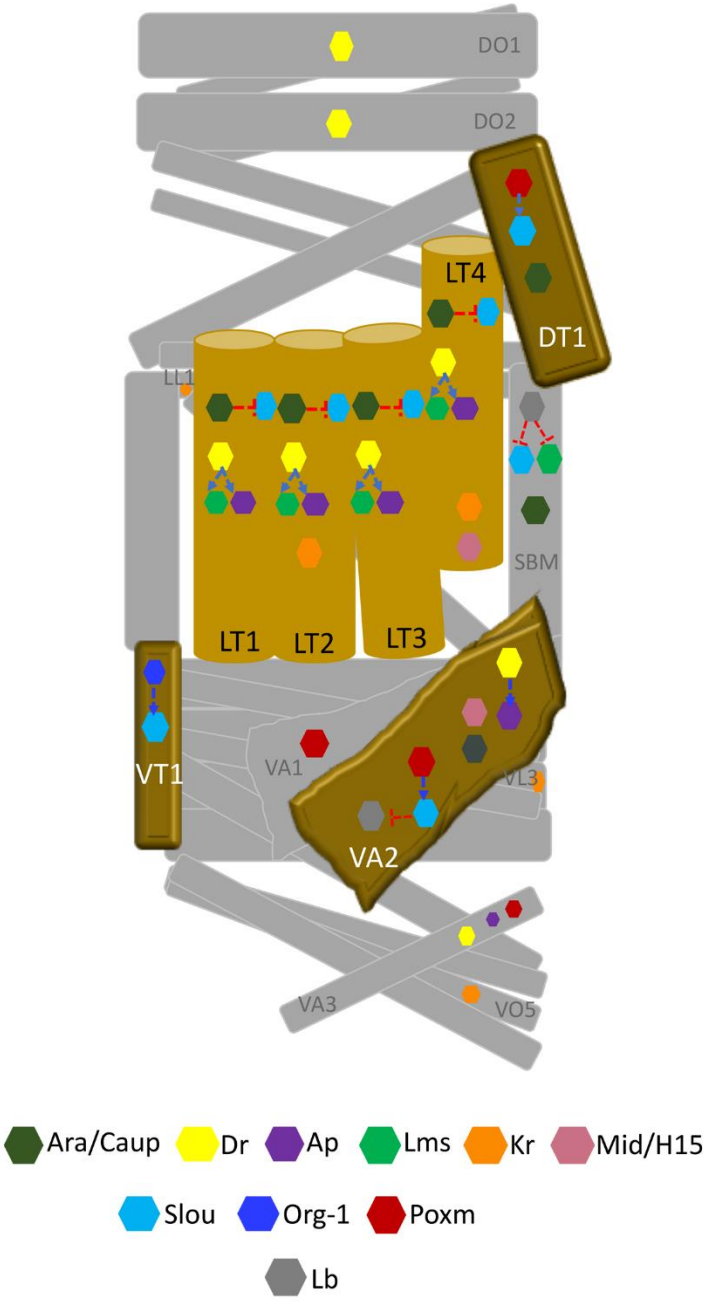


Figure 11. The iTF code for the Lms positive lateral transverse (LT) muscles and Slou positive VA2, VT1 and DT1 muscles.
Only iTFs whose expression is maintained during later developmental stages are displayed. The known genetic interactions are shown here. Gene repression is displayed using red arrows and activation using blue arrows. Non-LT and non-Slou+ muscles are greyed out. As seen in the figure, some LT iTFs are also expressed in Slou+ muscles as well as other muscles, as is the case for Slou+ muscle iTFs. It is also evident that there are iTF cross-regulatory mechanisms in muscles to prevent the acquisition of incorrect muscle identity. For example, Slou is repressed by the LT iTFs Ara/Caup in LT muscles (Carrasco-Rando et al. 2011) whereas in the neighboring SBM muscle, it is repressed by Lb (Junion et al. 2007) that also represses the LT iTF, Lms (Müller et al. 2010).

Signaling pathways could play a role in coordinating spatial and temporal gene regulation. So far, only the Ras/MAPK pathway has been identified to coordinate *slou* repression by Ara/Caup in LT muscles. Other spatial and temporal signaling pathways remain to be discovered.

The partially penetrant phenotypes indicate a potential disturbance in protein stoichiometry, thus, hinting at cofactors and signaling factors involved in the iTF regulatory pathway. Factors upstream of iTFs that directly regulate their expression are unknown at present. Temporal roles of iTFs also remain to be elaborated on. These lacunae in our current understanding of the specification and establishment of identity can be filled by further detailed studies into the various aspects of muscle identity.

CHAPTER 2 - Project objective and methodology

2.1. Project objective

The team had previously generated transcriptomics data for mRNA undergoing translation using the TRAP-rc method (Translating Ribosome Affinity Purification from rare cell populations) (Bertin et al. 2015). This data was generated for the Duf-Gal4 driven global muscle population as well as the Slou-Gal4 and Lms-Gal4 driven muscle subsets. Slou is a muscle identity transcription factor (iTF) expressed in the muscle subset comprising VA2, VT1 and DT1 (and transiently in muscles VA1, LO1 and VA3) while the iTF Lms is expressed in four lateral transverse (LT) muscles, LT1-4 (Figure 12). A dataset was also obtained for global embryonic mRNA. To summarize the TRAP-rc procedure, the UAS/Gal4 system (Brand et Perrimon 1993) was used to drive the expression of an eGFP tagged version of the 60S ribosomal subunit, RpL10a in a specific manner driven either by Duf-Gal4, Slou-Gal4 or Lms-Gal4. Embryos were homogenized and mRNA bound to the tagged ribosome was immunoprecipitated using magnetic beads coupled to anti GFP antibodies. mRNA was then extracted from the bound mixture and hybridized onto Agilent one color microarrays that contained probes mapping to multiple gene transcripts.

The aim of my project was to identify new factors involved in the specification, establishment and maintenance of muscle identity. This might include new iTFs, transcription cofactors, downstream realisator genes or signaling factors.

2.2. General methodology

I present details of the material and methods used in the ‘Results’ chapter (Chapter 3) and also provide details in the preprint of the article in Section 3.2.3.2. So, I will summarize the general methodology I applied for data analysis and validation of candidate genes here. The team had earlier validated that this transcriptomics microarray data was pertinent and had performed an

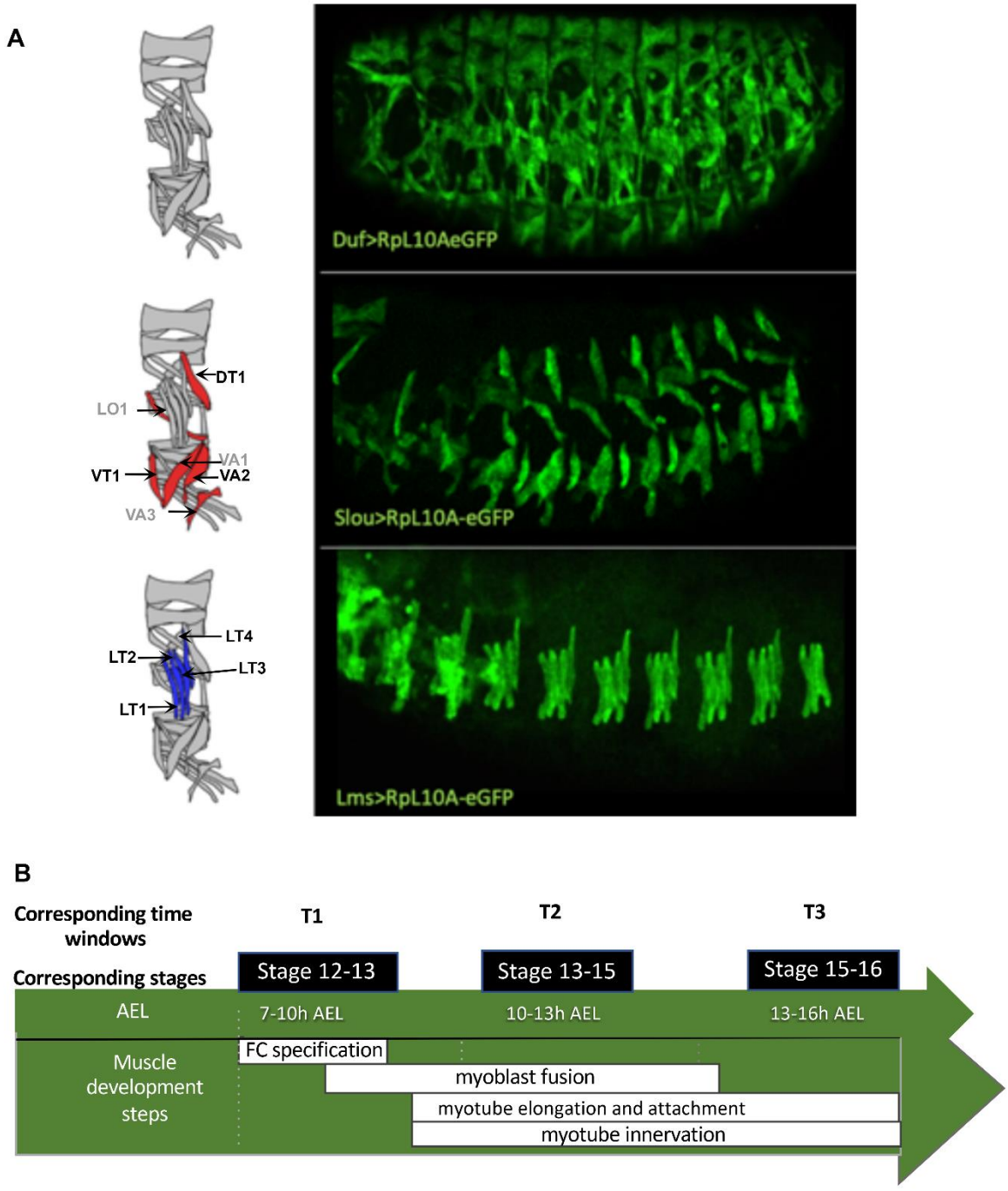


Figure 12. The muscle populations and myogenic processes during the time windows targeted for TRAP-rc.

(A) The 60S ribosomal subunit, RpL10a was expressed in a population-specific manner driven either by Duf-Gal4 in the global muscle population or Slou-Gal4 or Lms-Gal4 in muscle subsets (Bertin et al. 2015). The names of muscles expressing GFP are indicated in the figure. Muscles with transient expression are marked using grey lettering. **(B)** The TRAP-rc data was generated for three time windows of development at 7-10 hours AEL (After Egg Laying), 10-13 hours AEL and 13-16 hours AEL. The developmental stages corresponding to these time windows are indicated in black boxes above. For subsequent analyses, these time windows are referred to as T1, T2 and T3. The muscle developmental processes occurring during these time windows are indicated (*Figure adapted from Bertin, 2017*).

initial data analysis (Bertin et al., 2015, 2017). During this, data quality was assessed biologically by confirming the expression levels of a few known muscular and non-muscular genes in muscular versus non-muscular datasets by qPCR. Differential gene expression was determined using MS Excel. No in-depth analysis was performed for cis regulatory regions or gene interactions. I reanalyzed this data by using an integrative approach outlined here to identify the most interesting candidate genes. My goal for the bionformatic analysis was to gather as much information as possible without any bias at each step that could then be consolidated at the end to drive candidate selection.

I first ran a biological quality check of the data by verifying the expression pattern of GFP in the fly lines used to generate the transcriptomic data in order to determine if there was any discernable expression in non-muscular tissues to be able to interpret results logically. The expression was strongly muscular in all cases with minor non muscular expression. In order to determine new candidate genes that participate in the definition and establishment of muscle identity, I reanalyzed the TRAP-rc transcriptomics data bioinformatically. The basic analysis for new methods such as RNA-Seq follows the same principles as microarray data analysis. The limma package in R was initially developed for microarray data analysis and was extended to include RNA-Seq analysis (Ritchie et al. 2015).

I eliminated probes with low signal intensities using R to prevent downstream statistical noise generated by them. Replicate grouping into expected clusters and global gene expression patterns in the muscular versus non-muscular datasets was verified by principal component analysis (PCA) that permits the detection of key factors/genes that explain differences among datasets (Raychaudhuri, Stuart, et Altman 2000), again using R. Once the data was determined to be of good quality, I proceeded to perform a detailed analysis. This microarray had transcript probes annotated using the UCSC dm3 annotation of the genome that corresponds to Flybase/BDGP/GenBank Release 5. I remapped the microarray probes to the current UCSC dm6 or Flybase/BDGP/GenBank Release 6 genome annotation using Flybase and R. In order to do this, I first mapped dm3 coordinates to dm6 coordinates with Flybase, then mapped them to the corresponding dm6 gene and the precise gene region (UTR/exon/intron) using the dm6 genome annotation GTF file downloaded from Flybase. I then performed an analysis of temporal expression signatures of data that was normalized and fitted using limma to verify if there were identifiable gene clusters that moved together over time without being differentially expressed. This was done with the help of the Mfuzz package in R (L. Kumar et E Futschik

2007). I then determined differentially expressed genes using the R limma package (Ritchie et al. 2015). Heatmaps were generated using R to determine clusters of genes with interesting signatures among differentially expressed genes in different datasets. Gene ontology (GO) enrichment analyses were performed to determine potential enrichment of GO terms in gene lists using R, HOMER ((Heinz et al. 2010), http://homer.ucsd.edu/homer/), DAVID (Huang et al. 2007) and/or Flymine ((Lyne et al. 2007), https://www.flymine.org). Differentially regulated TFs and cofactors were identified using Flymine and then examined bibliographically as well as for enriched protein domains and GO and KEGG terms.

I also compared our Duf+ dataset to the only publicly available GEO RNA-Seq dataset that was comparable to ours because the study analyzed developmental timepoints very close to ours (Gaertner et al. 2012). This represented all muscular mRNA contrary to our dataset that represented RNA under translation. Pertinent SRA archives were downloaded from the GEO database, converted to raw fastq data, aligned to the dm6 genome assembly using STAR aligner (Dobin et al. 2013) and read counts were determined using featureCounts (Liao, Smyth, et Shi 2014). The count data was then analyzed using R. It was normalized across datasets by determining the counts per million (CPM). Data was corrected for composition bias to prevent oversampling in libraries where more reads were present due to larger library sizes and not due to a higher transcript expression. Significant differential expression was determined from counts after filtering out low intensity genes using the DESeq2 package in R (Love, Huber, et Anders 2014) with a log2FC cutoff of 0.58 (upregulated) or -1 (downregulated) and a p-value cutoff of 0.05. This differential expression was subsequently compared with the Duf+ TRAP population to assess data correlation. This acts as another data quality check for our datasets by determining if the same muscular genes are differentially expressed as expected and also gives an indication of differences at the transcriptional and translational levels.

I then compared differentially expressed genes in various muscle populations with data available in the BDGP database as well as data mined from publications. In order to do this, data from these different sources were populated into MySQL databases using Python and then queried using R to assess them for enrichment (Figure 13 presents an example of the structure of the sarcomere_genes database I created). This analysis was followed by a search for enrichment in cis regulatory elements using various tools such as HOMER ((Heinz et al. 2010), http://homer.ucsd.edu/homer/), i-cisTarget (Herrmann et al. 2012; Imrichová et al. 2015) and the MEME Suite (Bailey et al. 2015). i-cisTarget gave the most pertinent information.

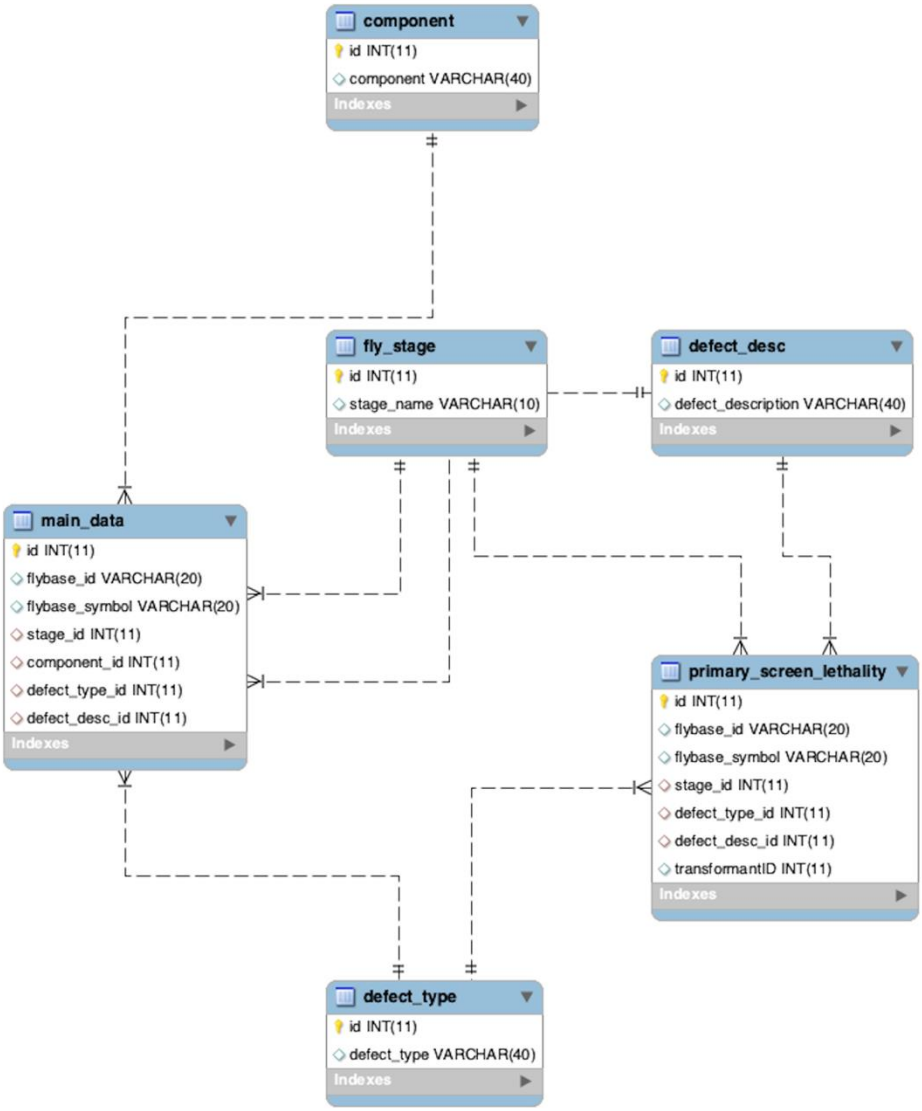


Figure 13. The entity relationship (ER) diagram representing the structure of the sarcomere_genes database.
Data pertaining to sarcomere defects resulting from a high throughput RNAi screen by (Schnorrer et al. 2010) was populated into the normalized tables in this database.

Protein/gene interaction analysis among sets of significantly differentially expressed TFs and cofactors resulting from the above analyses was performed using the String online database (Szklarczyk et al. 2019). For interesting TFs, potential downstream direct targets were identified using publicly available ChIP-Seq data from the modERN consortium for *Drosophila* TFs (Kudron et al. 2018). I downloaded pertinent BED files where they were available for embryonic developmental stages and mapped them to peaks using PAVIS. Candidate genes were selected for biological validation by consolidating information from all the above analyses.

After the identification of candidate genes, I performed a functional analysis of candidates. I performed an expression pattern analysis to identify genes that would be of most interest. After choosing the best candidates, *in situ* analysis was performed to visualize the mRNA expression pattern more clearly. Mutant lines of candidate genes were analyzed along with known and potential cofactors to identify their implication in muscle diversification. I followed this up with cis regulatory element analysis by generating GFP tagged enhancer reporter lines to identify putative enhancer regions on candidate genes. These steps are detailed in Chapter 3. The methods used are detailed in the 'Materials and Methods' section of the publication preprint as well as in other sections pertaining to biological analysis in Section 3.2 wherever they are not detailed in the publication preprint.

CHAPTER 3 - RESULTS

In this chapter, I present my results, starting off with the results from the initial bioinformatic analysis of the datasets culminating in candidate gene selection. I then present the results for the biological validation of these candidates.

3.1. Bioinformatic analysis of transcriptomics data

3.1.1. Data quality check

Datasets were subject to quality checks in order to ensure that sample replicates clustered together as expected and were amenable to statistical analysis before proceeding to retain them for downstream analyses.

3.1.1.1. Elimination of low intensity genes

I first background corrected the data to remove background noise (Silver, Ritchie, et Smyth 2009). This was followed by normalization between arrays by quantile normalization (Y. Zhao, Wong, et Goh 2020) (Figure 14A, A'). I eliminated probes with low intensity values across all datasets to prevent their interference with downstream data analysis (Calza et al. 2007). The Agilent one color microarrays contained negative control probes designed to represent background intensities. I verified that the mean values of background intensities were uniform across arrays. The mean background intensity was around 2.8. I looked at the distribution of the mean background intensity values for each probe across all arrays and given that this peaks at a value of 3, I used an arbitrary cutoff of 3 to eliminate low intensity genes across all samples analyzed (Figure 14B, B').

3.1.1.2. Verification of sample clustering

A PCA of muscle samples clusters them as expected. Principal component 1 (PC1) explains most of the variance observed (29%) while PC2 accounts for 11% (Figure 14C, C'). Samples cluster on PC1 based on timepoint in a sequential temporal fashion while PC2 clusters them based on sample grouping, clustering biological replicates together as expected. This pattern of clustering shows that time accounts for most of the variance between groups and that there are

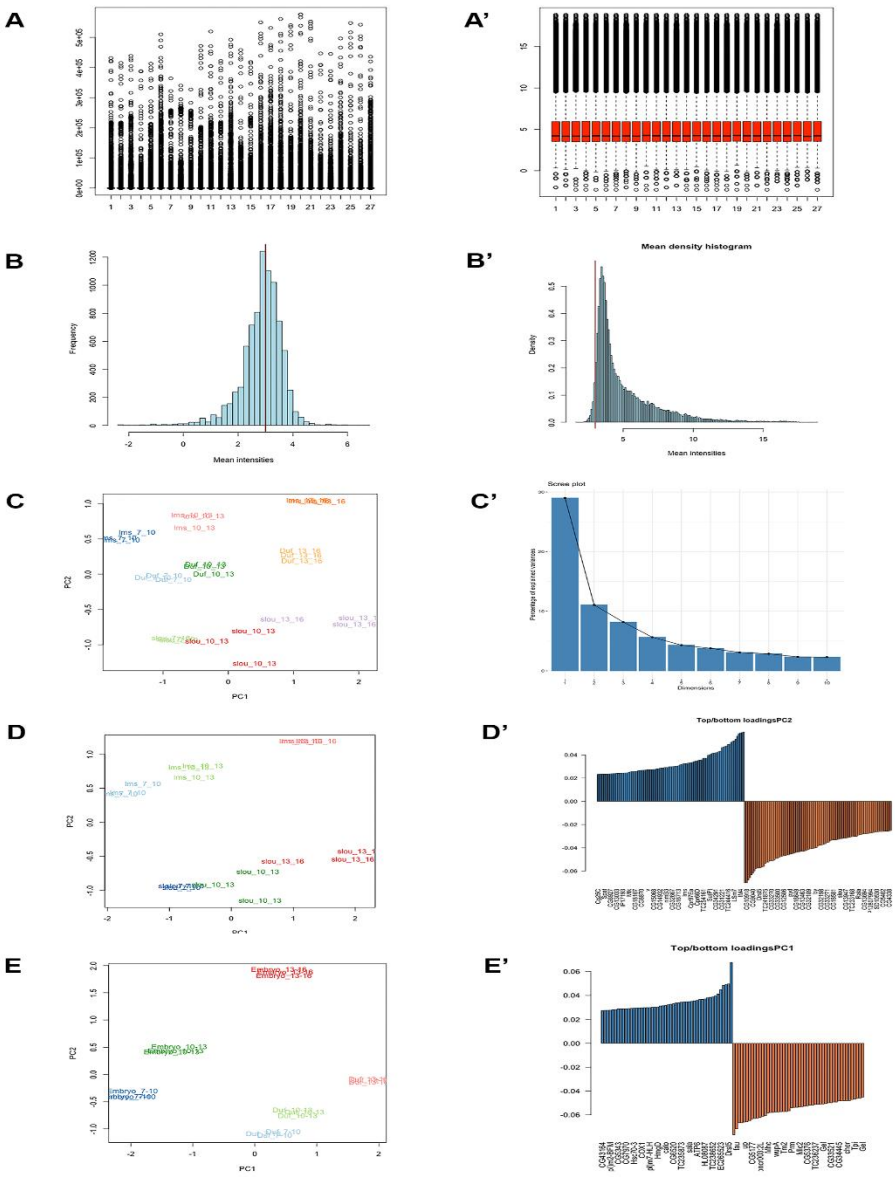


Figure 14. Initial data processing and visualization.
(A-A') Before normalization, (A) probe intensities are unevenly distributed across datasets reflecting technical variability among samples and are difficult to compare. Once data is quantile normalized (A'), intensities gain a similar distribution around a common mean permitting cross comparisons. **(B-B')** The mean probe intensity distribution for background probes on the microarrays peaks at ~3 (B). So, this cutoff was used to eliminate low intensity probes across all datasets (B') to minimize interference in downstream analysis. **(C-C')** A samples PCA clusters replicates together as expected (C) and separates them along two major principal components (PCs) (C'). PC1 explains the largest proportion of variance and clusters them based on timepoint. PC2 clusters them based on sample grouping. **(D-D')** A PCA of only the Lms and Slou subsets reproduces this grouping (D). PC2 that separates them based on subsets (D') has *lms* and *slou* among the top 250 genes as expected. **(E-E')** A PCA of the Duf TRAP samples against global embryonic mRNA samples clusters them into expected groups (E). In contrast to the PCA for muscle samples, PC1 separates them based on sample grouping indicating that the major variation is explained by the difference between groups. Among the top genes contributing to the variation on this component are a large number of known muscular genes (E') such as *Mhc*, *up*, *wupA*, *Tm2*, *Prm* and *cher*.

subtler differences that contribute to differences between sample groups within the same timepoint. A PCA of the Lms and Slou muscle populations (Figure 14D) without considering the global Duf population shows that among the top 250 genes that account for the variance between groups along PC2 are the key iTF genes *lms* and *slou* (Figure 14D'). This list of genes also has the LT iTF Drop (Dr) (Lord et al. 1995; Nose, Isshiki, et Takeichi 1998) as well as Dichaete (D) and Sox21b that have been shown to physically interact with Lms in a large-scale yeast two hybrid assay by the FlyBi consortium (http://flybi.hms.harvard.edu) among top loading genes in the Lms subset. Other genes include *Gel* that was earlier identified as an LT muscle specific gene by our team (Bertin et al. 2021) and *Ssdp* that has been shown to complex with LIM homeobox TFs such as the LT muscle iTF Apterous (Ap) (Bourgouin, Lundgren, et Thomas 1992; van Meyel, Thomas, et Agulnick 2003).

A PCA of each muscle sample versus the embryonic transcriptomic population also clusters them into expected groups (Figure 14E). In this case, samples cluster on the first PC, PC1 based on their grouping and then on PC2 based on their temporal profiles indicating that most of the variance can be attributed to tissue specificity. For example, in the Duf+ transcriptomics population, on PC1, multiple known myogenic genes such as *Mhc*, *up*, *Mlc2*, *wupA*, *Prm* and *Tm2* are among the genes contributing to variance along this PC (Figure 14E').

3.1.2. Data analysis

I present the stepwise integrative data analysis in this section. I first collected as many bits of information as possible without any selection bias. I then consolidated and integrated these bits and pieces of disjoint or linked information from various analyses detailed below to finally select candidate genes for biological validation.

3.1.2.1. Mapping dm3 probes to dm6

Since the mRNA was hybridized onto an older Agilent microarray, it had probes that mapped to UCSC genome assembly dm3 (Flybase/BDGP/GenBank Release 5). The latest release is dm6 (Flybase/BDGP/GenBank Release 6). So, I first mapped the dm3 probes to dm6 coordinates (Figure 15). Replicate spots for each probe were then averaged. Since these were probes targeting gene transcripts, I mapped them to specific transcripts and specific regions within transcripts such as the 5' UTR, exon or 3' UTR. I removed duplicates where a probe

ProbeName	dm3_GeneName	Coordinates_dm3	Coordinates_dm6	chr	start	end	dm3_chr	dm3_start	dm3_end	dm3_strand	dm6_ckr	feature	dm6_start	dm6_end	dm6_strand	dm6_geneid	dm6_genesymbol	dm6_transcriptid
A_09_P032094	14-3-3epsilon	chr3R:14,074,131..14,074,190	3R:18,248,409..18,248,468	3R	18248409	18248468	chr3R	14074131	14074190	+	3R	3UTR	18248289	18250550	+	FBgn0020238	14-3-3epsilon	FBtr0083568
A_09_P222670	TC240803	chr3R:14,074,168..14,074,227	3R:18,248,446..18,248,505	3R	18248446	18248505	chr3R	14074168	14074227	+	3R	3UTR	18248289	18250550	+	FBgn0020238	14-3-3epsilon	FBtr0083568
A_09_P145745	TC250174	chr3R:14,074,206..14,074,265	3R:18,248,484..18,248,543	3R	18248484	18248543	chr3R	14074206	14074265	+	3R	3UTR	18248289	18250550	+	FBgn0020238	14-3-3epsilon	FBtr0083568
A_09_P123580	14-3-3epsilon	chr3R:14,074,441..14,074,500	3R:18,248,719..18,248,778	3R	18248719	18248778	chr3R	14074441	14074500	+	3R	3UTR	18248289	18250550	+	FBgn0020238	14-3-3epsilon	FBtr0083568
A_09_P185585	TC256870	chr3R:14,075,888..14,075,947	3R:18,250,166..18,250,225	3R	18250166	18250225	chr3R	14075888	14075947	+	3R	3UTR	18248289	18250550	+	FBgn0020238	14-3-3epsilon	FBtr0083568
A_09_P224480	14-3-3zeta	chr2R:5,991,841..5,991,900	2R:10,104,336..10,104,395	2R	10104336	10104395	chr2R	5991841	5991900	+	2R	exon	10104256	10104409	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P051716	K03418	chr2R:5,993,253..5,993,312	2R:10,105,748..10,105,807	2R	10105748	10105807	chr2R	5993253	5993312	+	2R	CDS	10105702	10105825	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P215875	SD12734	chr2R:5,993,808..5,993,867	2R:10,106,303..10,106,362	2R	10106303	10106362	chr2R	5993808	5993867	+	2R	mRNA	10104115	10108474	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P011321	14-3-3zeta	chr2R:5,993,914..5,994,987	2R:10,106,409..10,107,482	2R	10106409	10107482	chr2R	5993914	5994987	+	2R	mRNA	10104115	10108474	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P178010	14-3-3zeta	chr2R:5,994,256..5,994,984	2R:10,106,751..10,107,479	2R	10106751	10107479	chr2R	5994256	5994984	+	2R	mRNA	10104115	10108474	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P011316	14-3-3zeta	chr2R:5,995,632..5,995,691	2R:10,108,127..10,108,186	2R	10108127	10108186	chr2R	5995632	5995691	+	2R	3UTR	10107623	10108474	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P148605	14-3-3zeta	chr2R:5,995,850..5,995,909	2R:10,108,345..10,108,404	2R	10108345	10108404	chr2R	5995850	5995909	+	2R	3UTR	10107623	10108474	+	FBgn0004907	14-3-3zeta	FBtr0332917
A_09_P011311	14-3-3zeta	chr2R:5,997,058..5,997,117	2R:10,109,553..10,109,612	2R	10109553	10109612	chr2R	5997058	5997117	+	2R	3UTR	10107623	10109687	+	FBgn0004907	14-3-3zeta	FBtr0332916
A_09_P012261	140up	chr3R:9,950,241..9,950,182	3R:14,124,460..14,124,519	3R	14124460	14124519	chr3R	9950241	9950182	-	3R	exon	14124163	14124862	-	FBgn0010340	140up	FBtr0337095
A_09_P145625	CR40712	chrU:002,951,421..002,951,362	X:23,282,357..23,282,416	X	23282357	23282416	chrU	2951421	2951362	-	X	exon	23280998	23282931	-	FBgn0267521	3SrRNA-Psi:CR458	FBtr0346900
A_09_P010356	18w	chr2R:16,004,245..16,004,304	2R:20,116,740..20,116,799	2R	20116740	20116799	chr2R	16004245	16004304	+	2R	3UTR	20116043	20116932	+	FBgn0004364	18w	FBtr0086309
A_09_P162015	26-29-p	chr3L:13,987,439..13,987,498	3L:13,994,339..13,994,398	3L	13994339	13994398	chr3L	13987439	13987498	+	3L	exon	13994231	13994400	+	FBgn0250848	26-29-p	FBtr0075766
A_09_P066231	26-29-p	chr3L:13,989,607..13,989,666	3L:13,996,507..13,996,566	3L	13996507	13996566	chr3L	13989607	13989666	+	3L	3UTR	13996471	13996571	+	FBgn0250848	26-29-p	FBtr0075766
A_09_P005446	JV208113	chrU:004,570,839..004,570,780	X:23,262,562..23,262,621	X	23262562	23262621	chrU	4570839	4570780	-	X	exon	23261772	23262666	-	FBgn0085819	3SrRNA-Psi:CR416	FBtr0346904
A_09_P196360	2mit	chr3R:8,948,700..8,948,641	3R:13,122,919..13,122,978	3R	13122919	13122978	chr3R	8948700	8948641	-	3R	3UTR	13121951	13126068	-	FBgn0260793	2mit	FBtr0330143
A_09_P074621	2mit	chr3R:8,952,490..8,952,431	3R:13,126,709..13,126,768	3R	13126709	13126768	chr3R	8952490	8952431	-	3R	CDS	13126114	13128567	-	FBgn0260793	2mit	FBtr0344730
A_09_P074626	TC231741	chr3R:8,957,130..8,956,961	3R:13,131,239..13,131,408	3R	13131239	13131408	chr3R	8957130	8956961	-	3R	mRNA	13124219	13132416	-	FBgn0260793	2mit	FBtr0344730
A_09_P160070	2mit	chr3R:8,958,204..8,958,145	3R:13,132,423..13,132,482	3R	13132423	13132482	chr3R	8958204	8958145	-	3R	exon	13131367	13133093	-	FBgn0260793	2mit	FBtr0330143
A_09_P197665	TC234601	chr3R:8,969,609..8,969,549	3R:13,143,827..13,143,887	3R	13143827	13143887	chr3R	8969609	8969549	-	3R	mRNA	13121951	13155687	-	FBgn0260793	2mit	FBtr0330143
A_09_P014281	CG32016	chr4:972,486..972,427	4:951,801..951,860	4	951801	951860	chr4	972486	972427	-	4	exon	951542	952077	-	FBgn0052016	4E-T	FBtr0334626
A_09_P222225	CG32016	chr4:972,944..972,885	4:952,259..952,318	4	952259	952318	chr4	972944	972885	-	4	CDS	952138	952461	-	FBgn0052016	4E-T	FBtr0334626
A_09_P205150	TC241428	chr4:974,975..974,916	4:954,290..954,349	4	954290	954349	chr4	974975	974916	-	4	mRNA	951542	958170	-	FBgn0052016	4E-T	FBtr0334626
A_09_P199245	5-HT1A	chr2R:14,978,430..14,978,489	2R:19,090,925..19,090,984	2R	19090925	19090984	chr2R	14978430	14978489	+	2R	CDS	19090757	19091458	+	FBgn0004168	5-HT1A	FBtr0474126
A_09_P010201	5-HT1A	chr2R:15,009,350..15,009,409	2R:19,121,845..19,121,904	2R	19121845	19121904	chr2R	15009350	15009409	+	2R	3UTR	19120537	19121958	+	FBgn0004168	5-HT1A	FBtr0474126
A_09_P105630	5-HT1B	chr2R:14,911,149..14,911,090	2R:19,023,585..19,023,644	2R	19023585	19023644	chr2R	14911149	14911090	-	2R	CDS	19023552	19023683	-	FBgn0263116	5-HT1B	FBtr0330056
A_09_P010716	5-HT1B	chr2R:14,911,176..14,911,117	2R:19,023,612..19,023,671	2R	19023612	19023671	chr2R	14911176	14911117	-	2R	CDS	19023552	19023683	-	FBgn0263116	5-HT1B	FBtr0330056

Figure 15. Mapping dm3 probes to dm6 genes.
Genomic RNA on the Agilent one color microarrays were represented by UCSC genome assembly dm3 (Flybase/BDGP/GenBank Release 5) probes for gene transcripts as well as non-coding RNA. Since the latest genomic assembly release is UCSC dm6 (Flybase/BDGP/GenBank Release 6), I mapped dm3 probe coordinates to corresponding dm6 coordinates and to specific features such as exons, 5'UTR or 3'UTR.

mapped to multiple transcripts and retained one of the matches to facilitate smoother downstream analysis at this point.

3.1.2.2. Temporal expression signature analysis

The microarrays contained some positive control probes that did not map to any transcript, which I eliminated. Data was fit to a linear model to simplify downstream analysis. Linear models allow us to statistically model and simplify complex datasets such as largescale gene expression/probe intensity data so that they can be represented in a mathematical form that can be used for computations. This allows us to determine the comparative degree of change of probe intensities using bioinformatic tools (G. Smyth 2005). Temporal clustering was performed on each group using the R Mfuzz package to study gene signatures over time and to discover genes that move together even though they are not significantly differentially expressed. I used the coefficients of linear fitted data, which represent the degree of change in intensity for each probe per dataset, as input to analyze the temporal expression signatures of these datasets over time during the three time windows of development. I refer to the time windows analyzed for each group as T1 (7-10h or stage 12-13), T2 (10-13h or stage 13-15) and T3 (13-16h or stage 15-16). 10 iterations were performed to verify cluster stability. A cluster was designated as stable if more than 65% of high membership genes in the cluster, with a membership value >=0.7, were retained in the same cluster or a similar cluster during all 10 iterations (Figure 16A, A'). I looked for significant enrichments in gene ontology (GO) terms among these clusters.

3.1.2.2.1. Duf+ global muscle population

In the global Duf+ datasets, one group of known iTF genes are part of high membership genes (*ap*, *lms*, *slou*, *Ptx1*, *eve*, *org-1*) that move between two clusters where gene expression is upregulated between T1 and T2 along with *wnt4* and then gets downregulated between T2 and T3 (Figure 16A, clusters 3 and 9). This group is significantly enriched for molecular functions (MF) related to transmembrane signaling (Figure 16B). Another group with known iTFs has genes whose expression is downregulated over time (*mid*, *caup*, *ara*, *Kr*) between T2 and T3 along with *wg* (Figure 16A, cluster 7). This cluster is significantly enriched for biological processes (BP) largely associated with negative regulation including transcription and mRNA metabolic processes (Figure 16C). Genes upregulated over time move between two clusters (Figure 16A, clusters 4 and 5) during different iterations and are significantly enriched for KEGG pathways such as the Wnt signaling pathway and neuroreceptor-ligand interactions

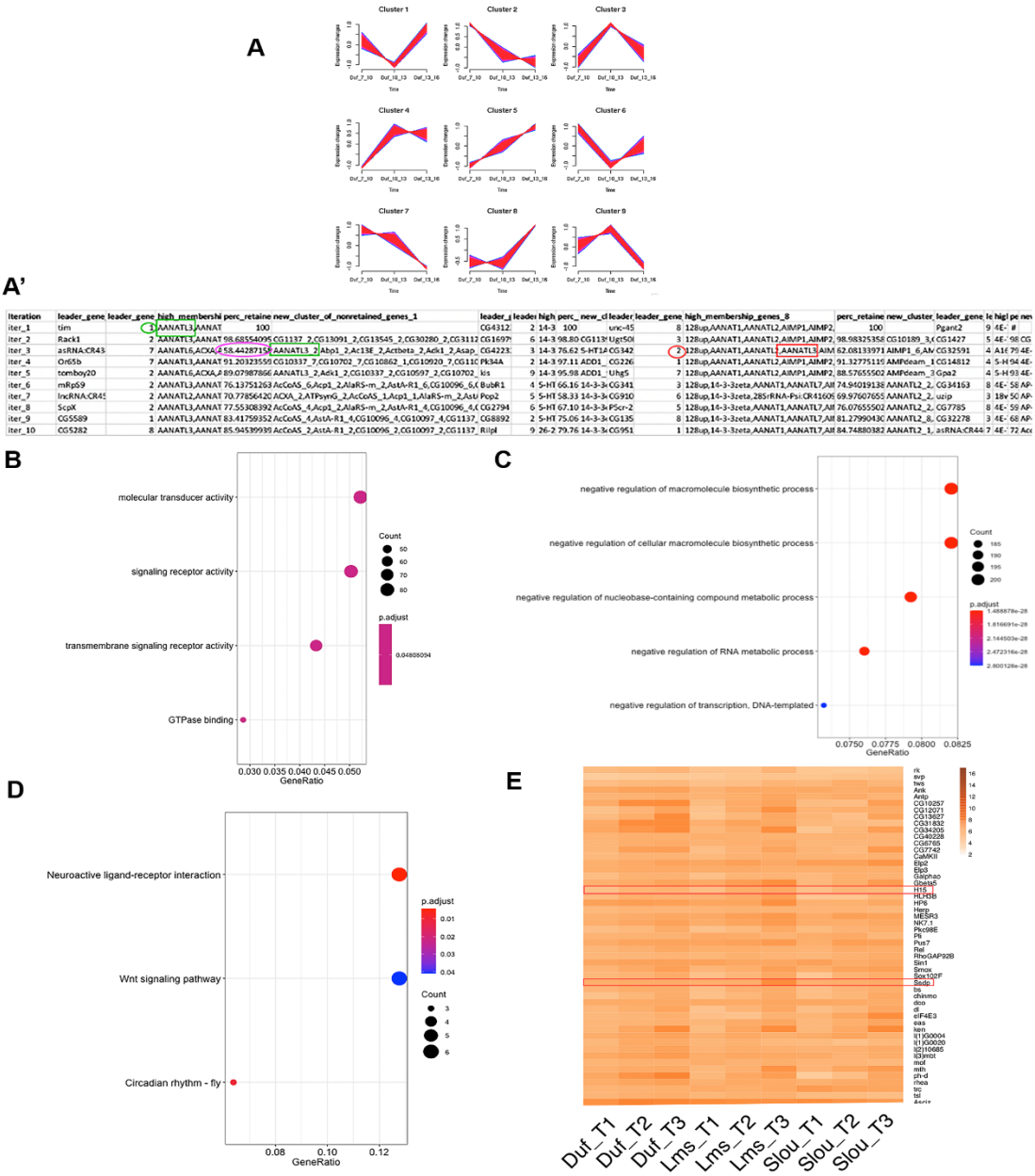


Figure 16. Temporal expression profiling of the Duf+ global muscle population.
(A-A') Temporal expression signature analysis was performed using a fuzzy clustering algorithm to analyze gene signatures and genes that move together over time in different muscle groups (A). Ten iterations (iter) were performed. Clusters were designated as stable if more than 65% of the high membership genes (HMG) of a cluster in iter1 were retained in the same cluster or moved to a similar cluster during all iters (A'). An example for AANATL3 is presented. It belongs to cluster 1 in (A) during iter1. It remains in cluster 1 during iter2, but moves to cluster 2 during iter3 (that corresponds to cluster 8 during iter1 since it contains the majority of HMGs in the iter1 cluster 8) when only around 58% of HMG genes are retained in this cluster. Other cluster 1 HMGs also move to cluster 8 during this iter. Clusters 1 and 8 have similar profiles with genes downregulated between T1 and T2 and upregulated between T2 and T3. So, cluster 1 is designated as a stable cluster. **(B)** iTF genes *ap*, *lms*, *slou*, *Ptx1*, *eve*, and *org-1* as well as *wnt4* move between clusters 3 and 9. The combined cluster is enriched for transmembrane signaling. **(C)** Cluster 7 with LT muscle iTF genes *mid*, *caup*, *ara* and *Kr* and *wg* is enriched for processes associated with negative regulation including transcription and mRNA metabolic processes. **(D)** Genes such as *Mef2*, *H15*, *Ssdp* and *wnt10* move between clusters 4 and 5 that are enriched for the KEGG Wnt signaling pathway among others. **(E)** A heatmap (kmeans clustering, n=8) for TFs and cofactors of coefficients of normalized data fitted to a linear model clusters *Ssdp* with *H15* that is partially redundant with the LT iTF Mid and homeobox TFs such as *HLH3B*, *Antp* and *NK7.1*.

(Figure 16D). This combined cluster has the myogenic differentiation factor *Mef2*, *wnt10* as well as *Ssdp* that is implicated in the translation of canonical Wnt signaling into gene transcription (Fiedler et al. 2015). A heatmap of TFs, TF cofactors and signaling factors (Figure 16E) clusters together *Ssdp* with *H15* that is partially redundant with the LT muscle iTF gene *mid* (R. P. Kumar et al. 2015) as well as homeobox TF genes such as *NK7.1*, *Antp* and *HLH3B*. No roles for mesodermal Wnts have been identified by studies as yet. It is currently known that ectodermal Wg cues are essential for myogenesis (Bejsovec et Martinez Arias 1991; Cox et Baylies 2005).

3.1.2.2.2. Lms+ muscle subset

Lms is part of a temporal cluster where most genes are upregulated between T1 and T2 and slightly downregulated between T2 and T3 (Figure 17A, cluster 5). Since TFs, cofactors and signaling pathways control downstream events, and thus gene signatures, I decided to select genes already implicated in these categories to check if there was a particular clustering pattern. I first identified genes in these categories, then generated heatmaps to compare the movement of these genes between the Lms and Slou muscle subsets (Figure 17B). This helped reveal that a few genes are highly upregulated during T2 and T3. The upregulation of *lms* temporally correlates with a boost in the expression of *salvador* (*sav*), a gene encoding a scaffolding protein implicated in the conserved Hippo signaling pathway that is involved in organ size control (Tapon et al. 2002; B. Zhao, Tumaneng, et Guan 2011). The Slou subset lacks this expression boost. In addition, it reveals subtle temporal differences in the expression of known iTFs and other TFs among the two muscle subsets. *Hand*, a gene implicated in cardiac development is more upregulated in the Slou subset. The iTF genes *ap* and *ara* are upregulated between T1 and T2 and then downregulated in the Lms subset. *vg* is expressed at slightly higher levels only in the Slou subset while *Ptx1* expression rises only in the Slou subset during T3 as expected.

One cluster has a consistently rising temporal expression signature (Figure 17A, cluster 9). It is interesting to note that the BP, adenylate cyclase-activating G protein-coupled receptor signaling pathway is enriched in this cluster during T3 as evidenced by Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005) (Figure 17C). A large number of genes localize to the cellular component (CC) mitochondria. Various KEGG signaling pathways are enriched in this cluster including Hippo, mTOR, FoxO and MAPK signaling among others (Figure 17D). Among the TFs in it is Ssdp that is known to complex with the LT muscle iTF Ap, a LIM

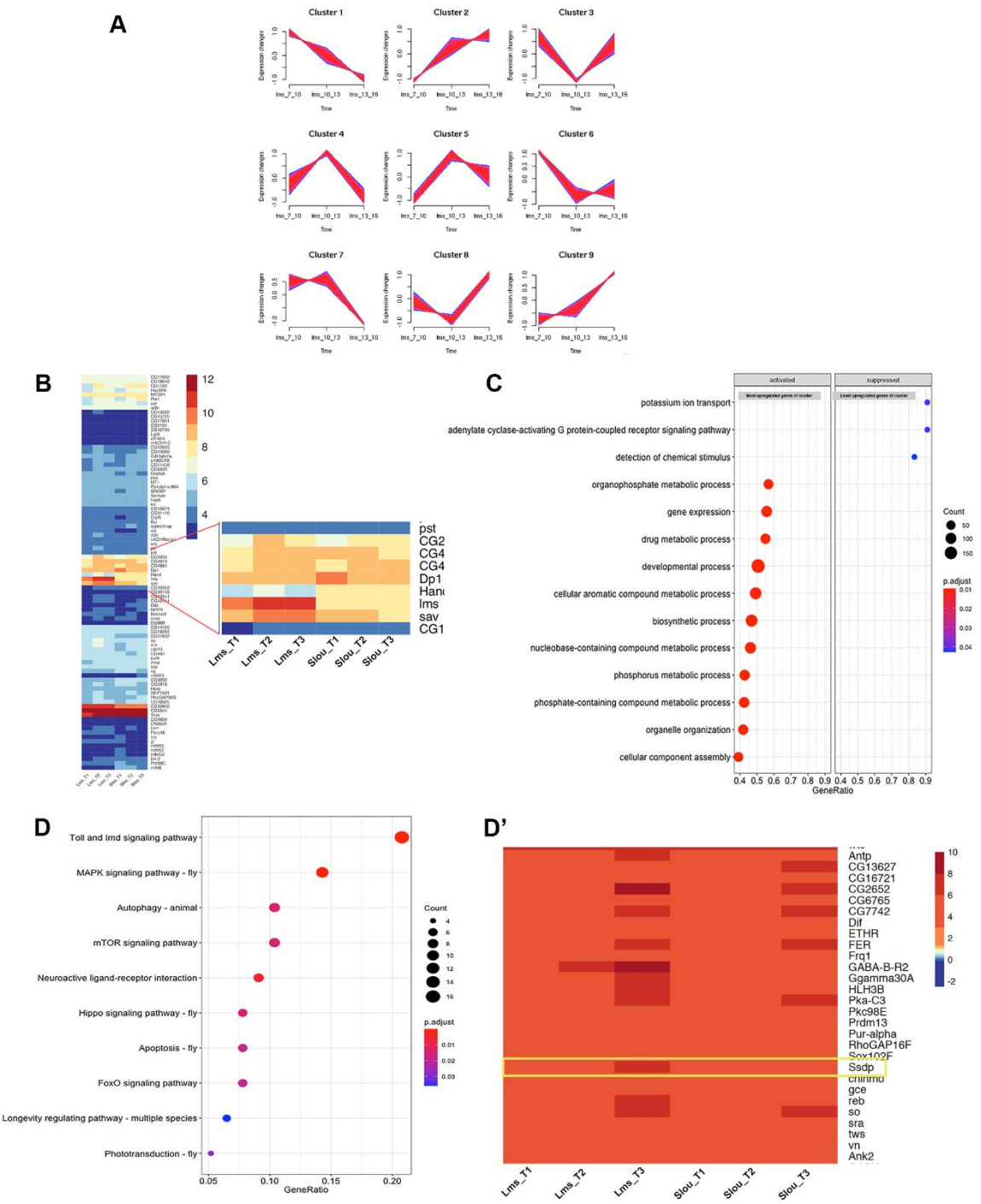


Figure 17. Temporal expression profiling of the Lms+ muscle subset.
(A) *lms* is part of stable cluster 5 where most of the genes are upregulated between T1 and T2 and are then slightly downregulated**. (B)** A heatmap kmeans clustering (n=15) of TFs and cofactors in this cluster reveals that *lms* upregulation correlates with an expression boost in *salvador* (*sav*) that acts as a scaffolding protein in the conserved Hippo signaling pathway. It unveils subtle differences among muscle subsets, with iTFs such as Ap and Ara being upregulated between T1 and T2 in the Lms subset and TFs such as Ptx1 and Vg expressing Slou specific temporal patterns. **(C)** A GSEA (Gene Set Enrichment Analysis) analysis of cluster 9 that has genes that are mostly upregulated over time reveals that the biological process (BP) associated with adenylate cyclase-activating G protein-coupled receptor signaling is enriched in this cluster during T3. **(D)** Cluster 9 is also enriched for various KEGG signaling pathways including mTOR, Hippo and FoxO signaling pathways. **(D')** Among the TFs in cluster 9 is Ssdp whose expression correlates with an increase in the expression of factors involved in the G protein-coupled receptor signaling pathway such as GABA-B-R2 and Ggamma30A.

homeodomain TF, to regulate transcription during wing disc development (van Meyel, Thomas, et Agulnick 2003). *Ssdp* is temporally upregulated between T2 and T3 only in the Lms population. A clustering of TFs, cofactors and signaling factors in this cluster shows that the increase in *Ssdp* expression correlates with an increase in the expression of factors involved in the G protein-coupled receptor signaling pathway such as GABA-B-R2 and Ggamma30A (Figure 17D').

3.1.2.2.3. Slou+ muscle subset

Slou is part of a cluster that is enriched for CCs associated with the mitochondria and the neuromuscular junction. Among the enriched BPs is neuromuscular synaptic transmission. A clustering of TFs, cofactors and signaling pathway components identified in this cluster reveals that *slou* and *Signal-transducer and activator of transcription protein at 92E* (*Stat92E*), an effector of the JAK/STAT signaling pathway move together (Figure 18A, B). The exact role of Stat92E during myogenesis is unknown. One study showed that the loss of maternal Stat92E in the mesoderm led to severe embryonic dorsal and lateral muscle phenotypes while the loss of maternal as well as zygotic Stat92E led to much more severe phenotypes (Y.-H. Liu et al. 2009). This study also showed that a knockdown of the zebrafish homologue, *stat5.1* led to defects in myotome organization. Another study revealed a genetic interaction with Retinoblastoma-family protein (Rbf) that interacts with E2f1, the key regulator of adult myogenesis (Zappia et al. 2019). Using a mesoderm specific Mef-Gal4 driver, the authors showed that a loss of Stat92E reduced the viability of adults under Rbf knockdown conditions. So, the significance of this differential expression between muscle subsets is unclear.

A cluster of genes that is upregulated over the course of time (Figure 18A, cluster 9) is significantly enriched for CCs localizing to the mitochondrion similar to the Lms cluster with a similar temporal expression profile. Apart from the MAPK and Toll signaling pathways that are also significantly enriched in the corresponding Lms subset cluster, the inositol phosphate metabolism pathway is significantly enriched in this Slou cluster (Figure 18C, D).

3.1.2.3. Differential expression discovery

eBayes smoothing of standard errors was performed in order to minimize gene-wise variances to permit stable statistical inference (Phipson et al. 2016; G. K. Smyth 2004). Differential gene expression analysis was performed on this processed data where each probe mapped to a gene transcript. Among muscle populations, differential expression signatures using a log fold

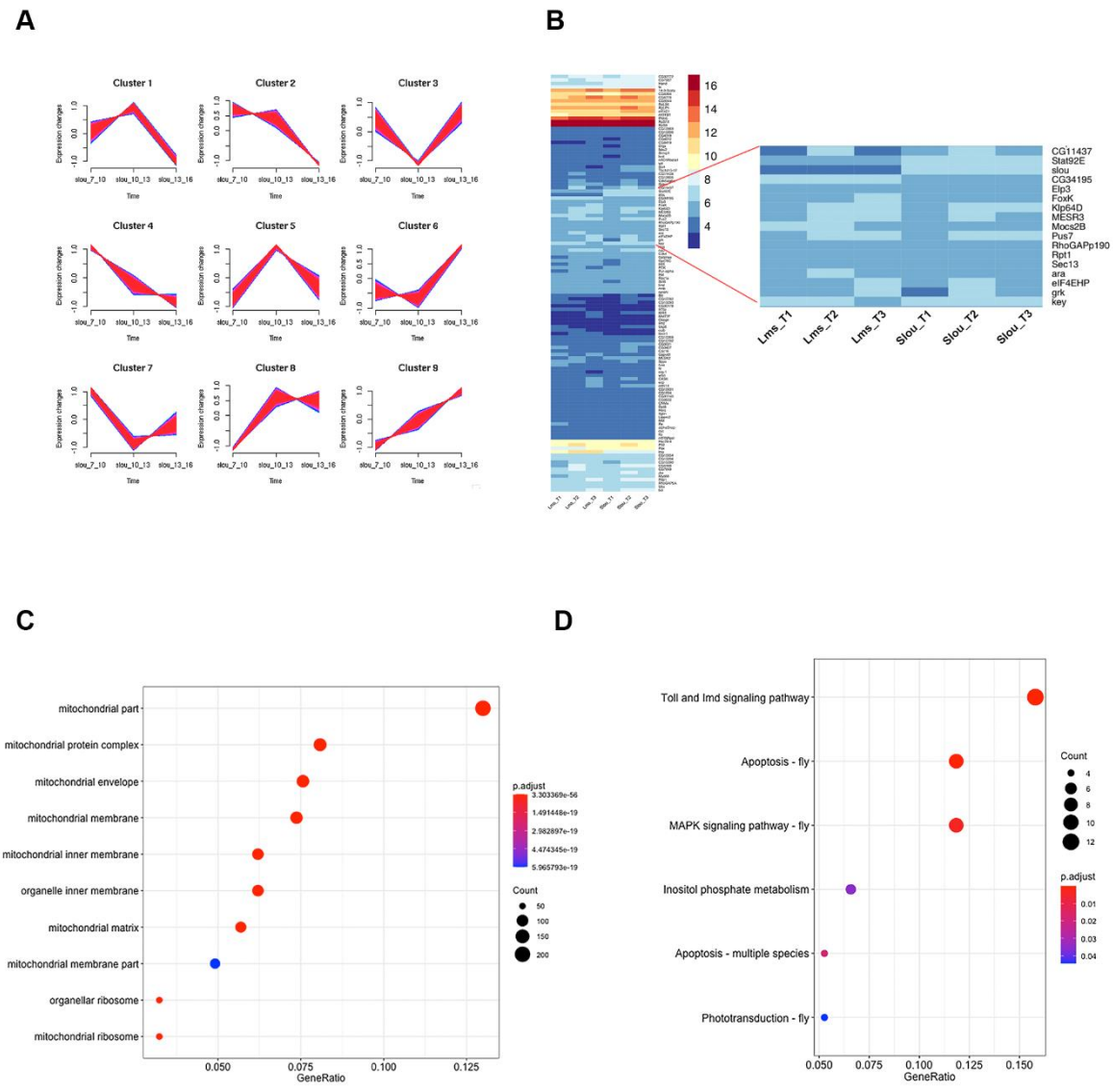


Figure 18. Temporal expression profiling of the Slou+ muscle subset.
(A) Slou is part of stable cluster 8. **(B)** A clustering of TFs, cofactors and signaling pathway components identified in cluster 8 that is enriched for muscle differentiation BPs reveals that *slou* and *Signal-transducer and activator of transcription protein at 92E* (*Stat92E*), an effector of the JAK/STAT signaling pathway move together. **(C)** Cluster 9, where most genes are consistently upregulated over the three time windows, is significantly enriched for mitochondrial cellular components (CC). **(D)** Cluster 9 also has a significant enrichment for the inositol phosphate metabolism pathway apart from other pathways that are also significantly enriched in the Lms cluster with a similar upward temporal expression profile.

change (logFC) cutoff of 0.58 or -1 (corresponds to a fold change of 1.7 or 0.5) and a p-value of 0.05 resulted in a manageable number of differentially expressed probes. At this point, I performed a gene level analysis. In instances where the expression of different probes mapping to the same transcript did not correlate, I considered only probes that were differentially expressed. While some probes on the microarray mapped to multiple transcripts and could thus reflect transcript level differences, many genes with multiple isoforms contained one only probe. So, this would make an analysis at the transcript level biased, although probes mapping to multiple transcripts could provide valuable information.

3.1.2.3.1. Comparison of the muscle populations with the global embryonic transcriptomic expression

The Duf+ TRAP dataset was compared with the whole embryonic mRNA dataset with a logFC cutoff of 1 or -1 and a p-value cutoff of 0.05. Differential gene expression was visualized by MD (Mean-difference) plots in order to get an idea of differential gene expression (Figure 19A-A'). The number of genes that are significantly upregulated increases over the course of development, as does the number of downregulated genes (Figure 19B, B'). Comparing the Lms subset to embryonic mRNA results in a similar trend, but there are a remarkable number of up and down regulated genes during timepoint T3, around twice that in the Duf population (Figure 19C, C'). The Slou population follows a trend where the number of upregulated genes increases over time while the number of downregulated genes is highest during T1, moderate during T2 and increases again during T3 (Figure 19D, D'). Fewer genes are differentially regulated in the Slou subset compared to the Duf population during T3. So, it appears that Lms had a tendency towards the upregulation of genes whereas Slou has a tendency towards gene upregulation during T1 and downregulation during T3. This indicates that the global muscle population masks differences between muscle subsets.

3.1.2.3.2. Comparison of muscle subsets with the global muscle transcriptomic expression

I first performed a comparison of all differentially expressed genes where at least one probe for a transcript was differentially regulated and followed this up with a comparison of differentially expressed genes considering only cases where all probes mapping to the same transcript were differentially regulated. A lower logFC cutoff of 0.58 with a p-value cutoff of 0.05 was applied to identify upregulated genes since the preceding analyses revealed that the differences between muscle subsets were subtler. A logFC cutoff of -1 was retained for downregulated genes.

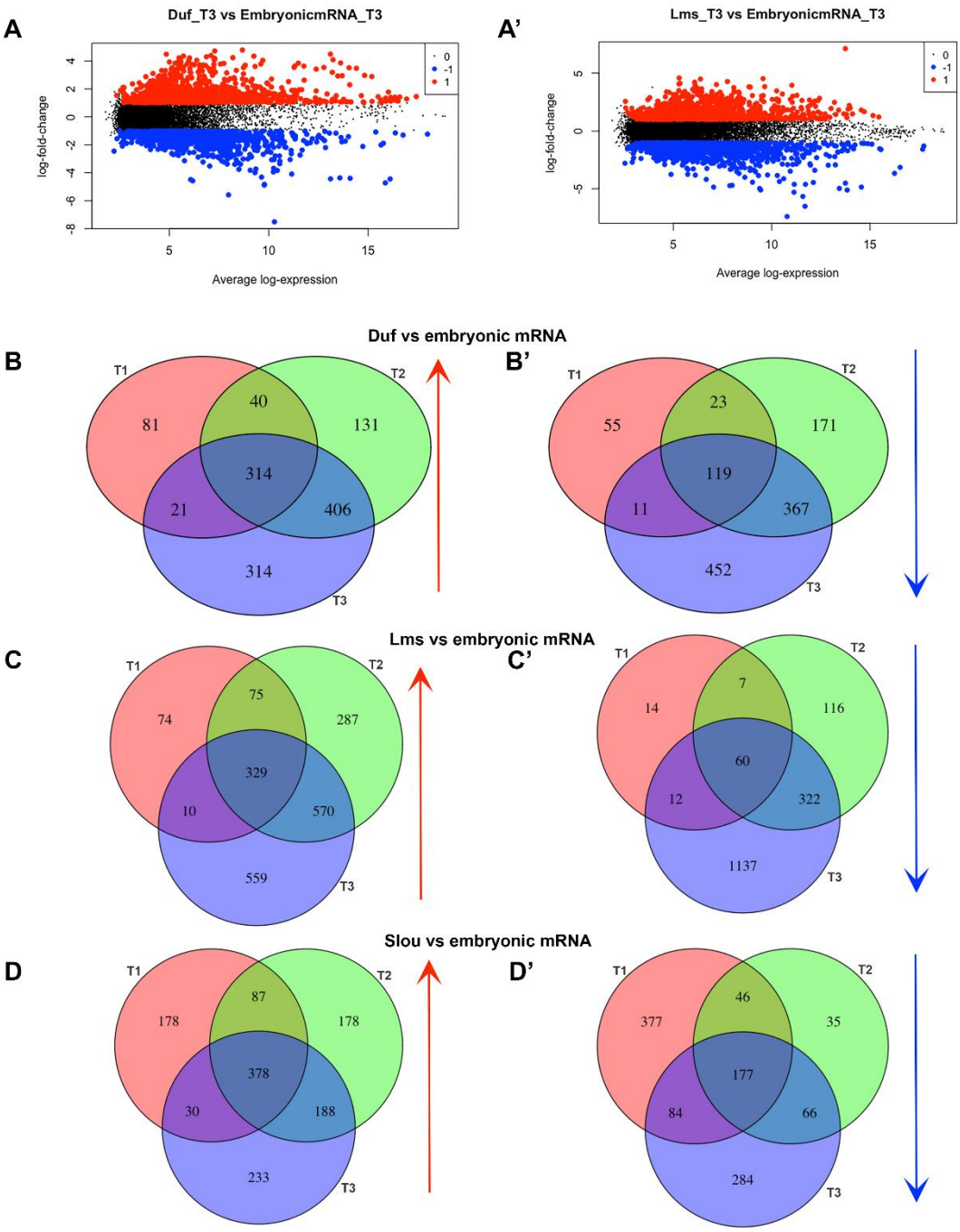


Figure 19. Differential expression discovery in the TRAP datasets for muscle populations versus global embryonic mRNA.
(A-A') Examples of Mean-difference (MD) plots for two different group comparisons. They show the number of differentially regulated genes in the datasets compared. **(B-B')** With a logFC cutoff of 1 (upregulated) or -1 (downregulated), the number of significantly differentially regulated genes in the Duf+ global muscle population versus the global embryonic mRNA increases over the three time windows of development in the upward (B) and downward (B') directions. **(C-C')** The Lms+ muscle subset displays a different pattern of significant differential expression signatures with around two times more significantly differentially expressed genes during T3 compared to the Duf muscle population represented in (B, B'). This indicates that clubbing all muscle subsets in the global muscle population masks differences between them. **(D-D')** The Slou muscle subset signature presents yet another pattern where there are around double the number of significantly differentially regulated genes compared to the Duf population during T1 and around half the number of significantly differentially regulated genes during T3.

A comparison of the Lms population with the Duf population reveals that the number of upregulated and downregulated genes increases over time. Including only genes where all probes mapping to a transcript are differentially regulated substantially decreased the number of significantly differentially regulated genes (Figure 20A-A'''). However, since this is an older microarray, it is possible that a large number of true positives were eliminated from the analysis in this case since there was a discrepancy in probe distribution on the microarray with some gene transcripts having multiple probes on the array and others having a single probe. In addition, some probes mapped to multiple transcripts and the uncorrelated probe intensities could be a result of differences at the isoform level. So, in the end I decided to retain all probes that were significantly differentially expressed and consider their means for further analysis. Comparing the Slou muscle subset against the Duf population reveals that more genes are up and down regulated during T2 than any other stage (Figure 20B, B').

3.1.2.3.3. Comparison of the muscle subsets with each other

Comparing the Lms subset against the Slou subset reveals that there are more significantly differentially upregulated genes during T1 and T3 than T2 while significantly downregulated genes increase over time (Figure 20C, D). A comparison in the other direction, Slou vs Lms, shows an inverse trend with an increase in significantly upregulated genes over time whereas more genes are downregulated during T1 and T3.

3.1.2.3.4. Comparison of our microarray dataset with publicly available RNA-Seq data (Gaertner et al., 2012)

I looked at the availability of datasets in the GEO database, especially RNA-Seq data, for comparative purposes. I discovered that our global Duf+ dataset could be compared with the RNA-Seq dataset from Gaertner et al. (Gaertner et al. 2012). To elaborate on the roles of paused RNA polymerase II during the embryonic development of *Drosophila melanogaster*, along with a ChIP-Seq analysis, this team performed a timeseries mRNA-Seq analysis of FACSed Mef2-Gal4 driven, membrane-labeled muscle cells at 4 timepoints, 6-8h AEL, 8-10h AEL, 10-12h AEL and 14-17h AEL (GEO accession numbers GSM846272-GSM846279 from the GEO dataset GSE34304). The last three timepoints correspond closely to the timepoints in our TRAP analysis. Since entire muscle cells were subject to FACS for the RNA-Seq analysis, it is possible that they were subject to high amounts of stress (Machado et al. 2017). However, organisms are subject to varying amounts of stress under normal developmental conditions and are programmed to adapt. It is not entirely clear how drastically this affects the expression

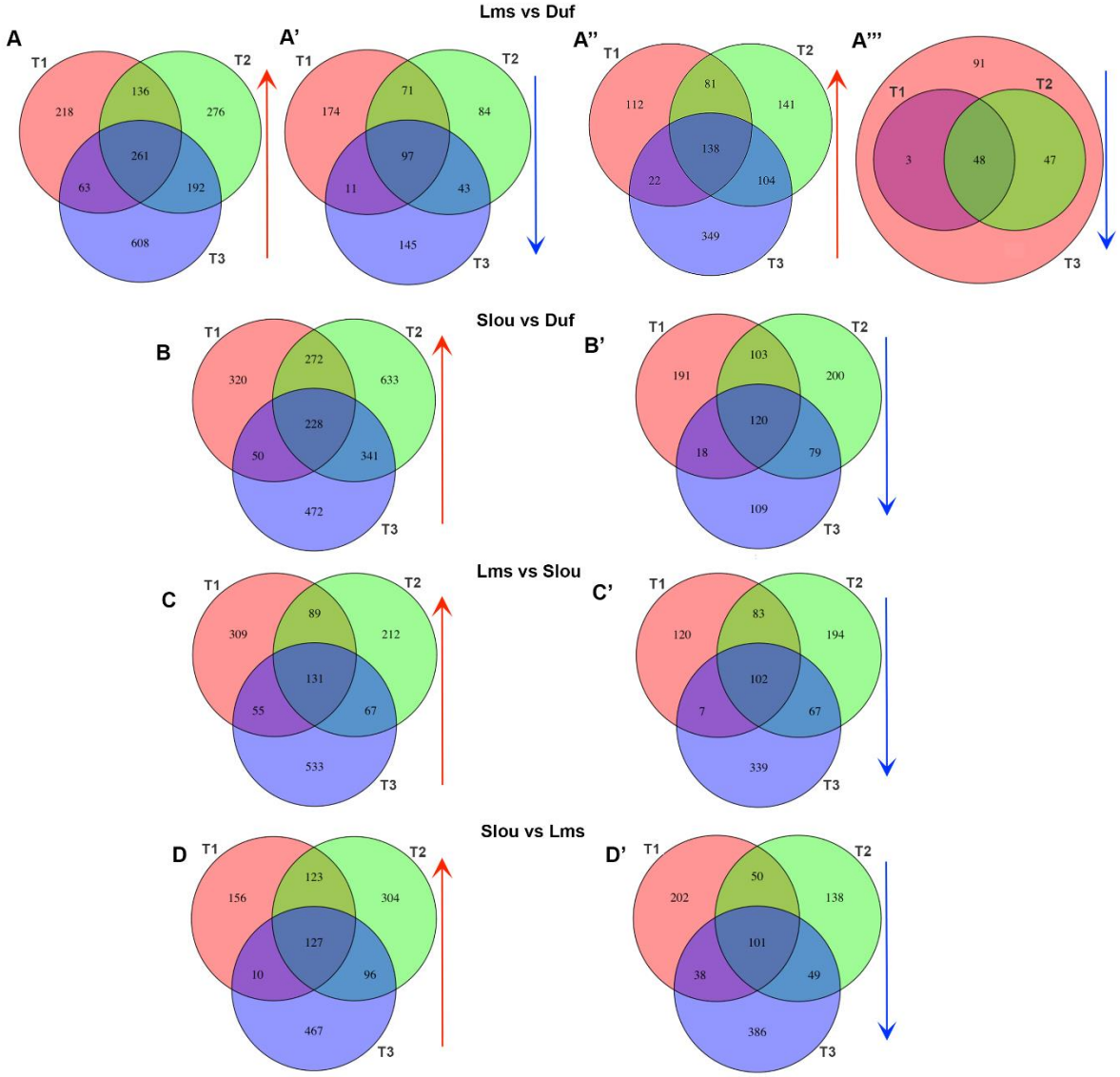


Figure 20. Differential expression discovery in the TRAP datasets among muscle populations. **(A-A''')** Applying a logFC cutoff of 0.58 (up) or -1 (down) and p-value cutoff of 0.05, the number of significantly differentially expressed genes increases over time in the Lms subset when compared against the global Duf+ population. There is more than twice the number of differentially regulated genes when the average of all significantly differentially expressed probes for each gene is considered in both the up (A) and down (A') directions compared to when the average of only those transcripts for each gene for which all mapping probes are similarly differentially expressed are considered either in the up (A'') or down (A''') direction. Only averages of all significantly differentially expressed probes for each gene were considered for further downstream analysis since there is unequal probe distribution between genes as well as between transcripts of a gene. **(B-B')** Comparing the Slou muscle subset against the Duf population reveals that more genes are significantly up and down regulated during T2 than any other timepoint. **(C-C')** Comparing the Lms subset against the Slou subset shows an increase in significantly upregulated genes during T1 and T3 and in downregulated genes over time. **(D-D')** A comparison in the other direction, Slou vs Lms, shows an increase in significantly upregulated genes over time whereas more genes are downregulated during T1 and T3.

patterns of genes that are tissue specific. I analyzed these datasets to verify the amount of correspondence. While the mRNA-Seq data would reveal mRNA that are differentially expressed over time, the TRAP dataset corresponds to mRNA under translation. In addition, microarray hybridization differs greatly from RNA-Seq data generation. So, there would not be a complete correlation between datasets. The idea is to find genes that are consistently differentially regulated in these datasets at the level of transcription and translation. This also acts as a data quality check. I refer to the mRNA-Seq time windows as mRNA-SeqT1 (6-8h AEL), mRNA-SeqT2 (8-10h AEL), mRNA-SeqT3 (10-12h AEL) and mRNA-SeqT4 (14-17h AEL). mRNA-SeqT4 corresponds closely to T3 in our dataset, mRNA-SeqT3 to T2 and mRNA-SeqT2 to T1. I first looked at the correlation of significantly upregulated genes between our Duf+ timepoints and the mRNA-Seq timepoints (Figure 21A-B''). Looking at the significantly upregulated genes between DufT2 vs DufT1 (10-13h vs 7-10h), it is evident that the correlation of gene expression is strongest with mRNA-SeqT3vsT2 (10-12h vs 8-10h) whereas for DufT3 vs DufT2 (13-16h vs 10-13h), there is an almost 50% correlation with mRNA-SeqT4vsT3 (14-17h vs 10-12h). Considering that these timepoints are very close, this amount of correlation between these very different datasets could roughly indicate the degree of correlation between transcription and translation.

A comparison of GO enrichment for BPs across all datasets shows that genes for muscle development and differentiation are indeed overrepresented in all groups (Figure 21C). What this also reveals is that the DufT2 vs DufT1 enrichment represents a combination of mRNA-SeqT3vsT2 and mRNA-SeqT2vsT1. I looked at genes that are part of the BP cluster annotated with 'striated muscle cell differentiation' and there are commonly upregulated genes between DufT3 vs DufT2 and mRNA-SeqT3vsT2 as well as mRNA-SeqT4vsT3 (Figure 21D). There is excellent correspondence between BP enrichment in DufT3 vs DufT2 and mRNA-SeqT4vsT3. This adds an additional layer of validation of the pattern of gene expression during embryonic muscle development and quality check of our dataset.

3.1.2.4. Analysis of differentially expressed genes

After identifying differentially expressed genes, I analyzed them using various strategies detailed below including a comparison with published/curated information, clustering analysis to identify genes that showed similar expression signatures, a search for enrichment of GO or

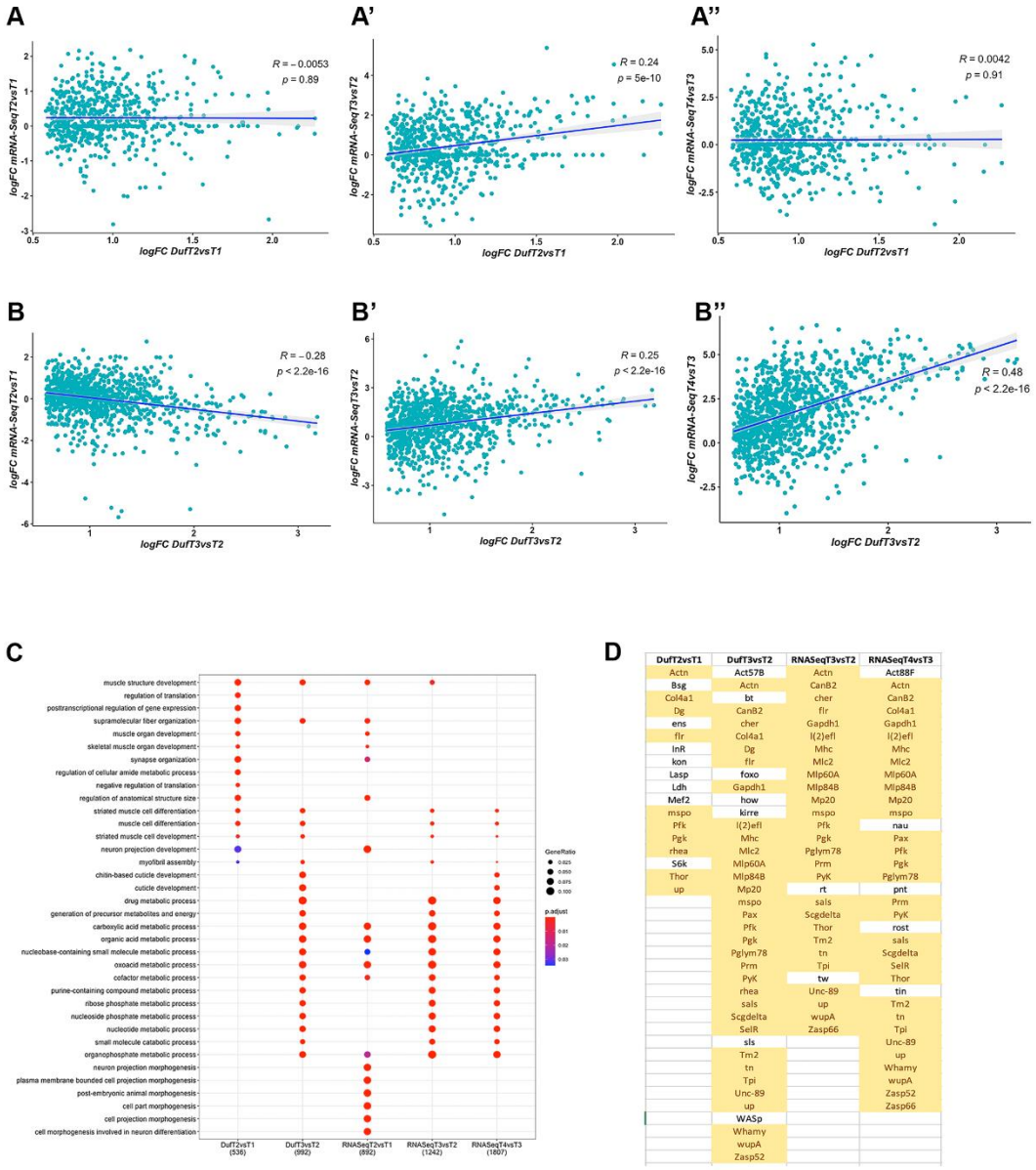


Figure 21. Comparison of the Duf+ global muscle subset with the GEO mRNA-Seq dataset from Gaertner et al. 2012.

(A-A'') Comparing significantly upregulated gene expression patterns between DufT2 vs T1 and the mRNA-Seq data, it is evident that there is a significant correlation between DufT2 vs T1 (10-13h vs 7-10h) and mRNA-SeqT3 vs T2 (10-12h vs 8-10h) that represent the timepoints closest to the Duf groups compared. It results in a Pearson correlation coefficient of 0.24 or a 24% correlation (R = correlation coefficient, p = p-value). **(B-B'')** Similarly, DufT3 vs T2 (13-16h vs 10-13h) is significantly correlated to the expression pattern of genes in mRNA-Seq T4 vs T3 (14-17h vs 10-12h) producing a correlation close to 50%. (B'') **(C)** A comparison of significantly enriched GO terms for the term BP, on the other hand, reveals that significant GO terms in DufT2 vs T1 are distributed between mRNA-SeqT2 vs T1 and mRNA-SeqT3 vs T2. DufT3 vs T2 terms are enriched similarly in mRNA-SeqT3 vs T2 as well as mRNA-SeqT3 vs T4. Muscle development related terms are significantly enriched. **(D)** A further analysis of the genes annotated to the BP 'striated muscle cell differentiation' reflects this pattern with genes similarly distributed between mRNA-SeqT3 vs T2 and mRNA-SeqT4 vs T3.

KEGG terms or known and *de novo* cis regulatory motifs as well as gene interaction analysis in order to gather information to guide me in candidate gene selection.

3.1.2.4.1. Comparison of differentially expressed genes with published and/or curated information

As an additional data quality check measure and to streamline the choice of candidate genes, I compared our datasets with the following publicly available data in published articles or curated databases by populating this data into local MySQL databases I created and queried them:

1. BDGP database: a publicly accessible database that consolidates data for high throughput *in situ* hybridization during embryogenesis (https://insitu.fruitfly.org/cgi-bin/ex/insitu.pl). I subsequently queried for muscle and CNS annotated genes only because many CNS genes have been attributed muscular functions.
2. (Schnorrer et al. 2010): a high throughput RNAi screen using RNAi mediated knockdown of genes driven by the mesodermal driver Mef2-Gal4. The lines were then analyzed for locomotion defects or lethality followed by the analysis of sarcomere defects in larval and adult muscles.
3. (Sandmann et al. 2006): a high throughput analysis of Mef2 targets during different stages of embryogenesis using a combined approach with a ChIP-on-chip from early to mid developmental stages and Mef2 mutant transcriptomic analysis spanning early to late embryogenesis.
4. (S. Deng, Azevedo, et Baylies 2017): a review that consolidates information about all currently known myoblast fusion genes in *Drosophila*.

In general, the number of differentially expressed, previously identified muscular genes generally increases over time in both the upward (Figure 22) and downward (Annexe 1) directions in all groups. There are more genes annotated to muscular processes among the significantly differentially expressed genes in the Duf versus embryonic mRNA comparison, fewer in the comparison of muscles subsets to the Duf population and still fewer when comparing muscle subsets with each other reiterating the presence of subtle differences among muscle subsets suggested by the initial samples PCA. A clustering of genes that have been annotated as having muscular RNA expression in BDGP and are upregulated in various groups shows that some genes move in a similar direction in the general Duf+ muscle population as well as the muscle subsets whereas some genes are differentially expressed in muscle subsets

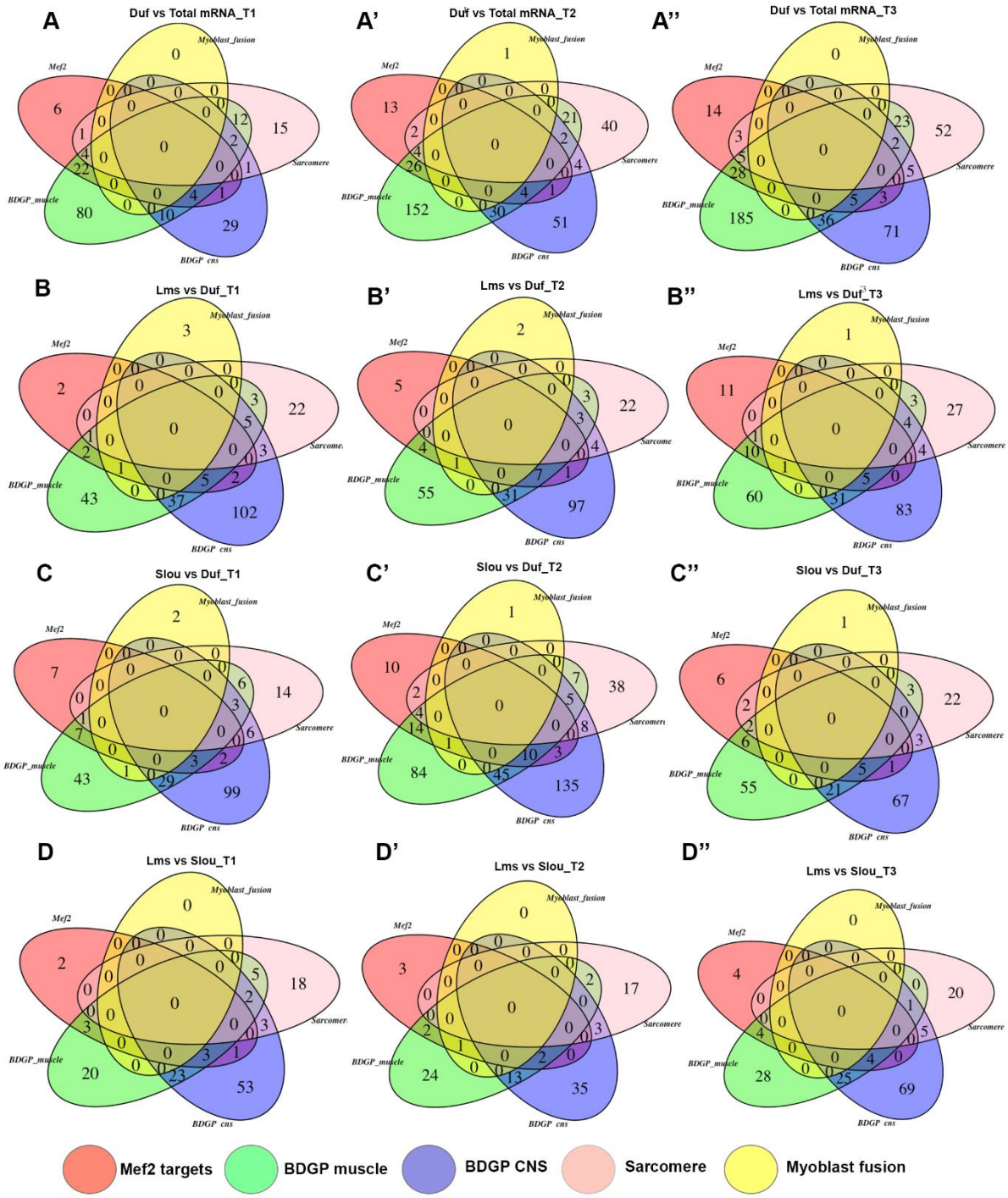


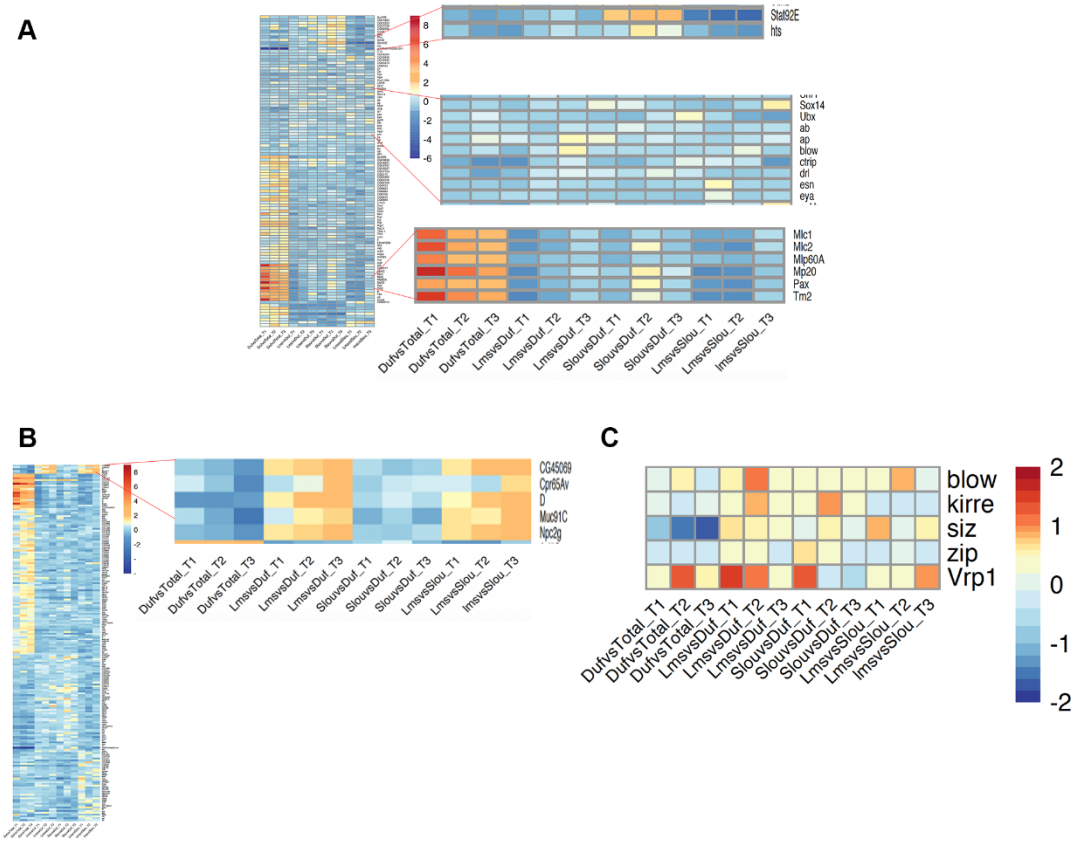
Figure 22: Comparison of significantly upregulated genes with public data from articles (Sandmann et al. 2006 for Mef2 targets; Schnorrer et al. 2010 for sarcomeric genes; S. Deng, Azevedo, et Baylies 2017 for myoblast fusion genes) and curated databases (BDGP for genes with expression in the somatic muscles and/or CNS).

(A-A'') The number of genes already annotated to be expressed in muscles or identified to have muscle expression increases over the three timepoints in the Duf TRAP vs embryonic mRNA comparison as do genes annotated to be expressed in the CNS. **(B-B'')** Comparing the Lms muscle subset with Duf, the number of upregulated muscle genes increases while the number of CNS genes decreases over time. **(C-C'')** Comparing Slou with Duf, the number of annotated significantly upregulated muscle as well as CNS genes is highest during T2. **(D-D'')** Comparing Lms with Slou, the number of annotated significantly upregulated muscle genes increases over time while the number of CNS genes decreases between T1 and T2 and then increases.

with respect to the general muscle population. *Mlc2*, *Tm2*, and *Mp20*, for example, are significantly upregulated in the Slou subset versus the Duf population, but downregulated in the Lms subset. On the other hand, genes such as *Tis11*, *sbb*, *Dr* and *eya* are significantly upregulated in the Lms subset and downregulated in the Slou subset. *Dr*, for example, is significantly upregulated in the Lms subset versus the Slou and Duf subsets during all three timepoints while *Tis11* is significantly upregulated during T3 and *eya* during T1 only versus the Slou subset. One *sbb* probe is significantly upregulated during T3 in the Lms population versus the Duf and Slou subsets. It shows a consistently rising temporal profile only in the Lms subset for all probes. *Ssdp* is annotated with muscular expression during very early developmental stages, but it consistently shows up as being significantly upregulated in the Lms population with respect to the Duf as well as Slou subsets during T3.

Among Mef2 targets, subtle differences in the levels of expression of its target genes such as *ap* and *Sox14* that are significantly upregulated only in the Lms subset are evident (Figure 23A). A similar analysis of sarcomere related genes from Schnorrer et al. reveals that *D* is consistently specific to the Lms subset (Figure 23B). In their study, RNAi against *D* resulted in actin blob formation in myofibrils. As seems to generally be the case, myoblast fusion genes are differentially regulated in the muscle subsets with respect to the Duf population and with respect to each other. *kirre*, *siz* and *zip*, for example, are significantly upregulated in the muscle subsets whereas they are not in the general muscle population. *Vrp1* and *blow* are significantly upregulated in each muscle subset at distinct stages (Figure 23C).

I also performed a comparison with a list of genes that were identified in a screen for motor neuron axon targeting (Kurusu et al. 2008). Kurusu et al. performed a large scale misexpression analysis of cell surface and secreted (CSS) proteins using UAS EP lines inserted 5' of CSS proteins driven by 24B-Gal4 in somatic muscles to identify candidates causing axon mistargeting and synaptic bouton defects. There is some overlap in the gene sets at various timepoints and in various groups (Figure 23D). Some genes appear in significantly up and downregulated lists either because of potential differential transcript regulation or because their mapping is unclear. An example is *mspo*, a large gene spanning more than 100KB with many smaller genes getting transcribed in the same direction from its intronic regions to which probes map. Some subtle gene dynamics are revealed by a comparative analysis. *Gel*, for example, is significantly upregulated in the Lms subset with respect to the Duf population as well as the Slou subset and is significantly downregulated in the Slou subset with respect to Duf. *Mur2B*



D

Up

LmsvsDuf_T2_up	LmsvsDuf_T3_up	SlouvsDuf_T2_up	SlouvsDuf_T3_up	LmsvsSlou_T2_up	LmsvsSlou_T3_up	SlouvsLms_T2_up	SlouvsLms_T3_up
caps	CG12290	caps	CG32373	Gel	Clrl	CG31431	CG32373
CG31635	CG32373	CG31431	ImpE1	Mur2B	Gasp	CG5597	fred
Gli	CG32432	CG32373	mgl	NA	Gel	Clrl	Gasp
obst-B	CG42709	CG45263	mspo	NA	Pvf3	Gasp	mgl
NA	daw	CG5597	Ten-a	NA	shf	Wnt2	mspo
NA	Ddr	Clrl	NA	NA	uif	NA	Mur2B
NA	Fur1	Gasp	NA	NA	NA	NA	NA
NA	galectin	Ibk	NA	NA	NA	NA	NA
NA	Gel	mgl	NA	NA	NA	NA	NA
NA	ImpE1	Ten-a	NA	NA	NA	NA	NA
NA	obst-B	Vang	NA	NA	NA	NA	NA
NA	uif	NA	NA	NA	NA	NA	NA

Down

LmsvsDuf_T2_down	LmsvsDuf_T3_down	SlouvsDuf_T2_down	SlouvsDuf_T3_down	LmsvsSlou_T2_down	LmsvsSlou_T3_down	SlouvsLms_T2_down	SlouvsLms_T3_down
mspo	CG32373	CG32373	Gasp	CG5597	CG32373	Gel	Clrl
NA	mspo	CG45263	Gel	Clrl	fred	Mur2B	Gasp
NA	NA	CG9550	haf	Gasp	mgl	NA	Gel
NA	NA	Gel	mspo	NA	mspo	NA	Pvf3
NA	NA	glec	NA	NA	NA	NA	shf
NA	NA	mspo	NA	NA	NA	NA	NA
NA	NA	tutl	NA	NA	NA	NA	NA
NA	NA	zormin	NA	NA	NA	NA	NA

Figure 23. Analysis of intersects between significantly differentially regulated genes in any group and published and/or curated information.
(A) A heatmap for significantly differentially expressed genes intersecting predicted ChIP-on-chip Mef2 targets (Sandmann et al. 2006) reveals that *Stat92E* consistently appears in the list of differentially regulated genes in the Slou subset with respect to the Lms subset. *ap*, *Sox14*, *blow* and *eya* are differentially regulated in the Lms subset with respect to the Slou subset at different stages. Muscular genes such as *Mlc2*, *Mp20* and *Tm2* are upregulated in the Slou subset during T2. **(B)** A heatmap analysis of genes intersecting those associated with sarcomeric phenotypes (Schnorrer et al. 2010) reveals *Dichaete* (*D*) coding for a TF as being Lms specific. **(C)** Intersecting the lists with known myoblast fusion genes (S. Deng, Azevedo, et Baylies 2017) reveals specific differences in the dynamics of gene expression among the Lms and Slou subsets. **(D)** An intersection of differentially expressed genes with cell surface and secreted proteins associated with axon targeting and bouton defects reveals that *Gel* and *Mur2B* are significantly upregulated specifically in the Lms subset with respect to the Duf and Slou subsets. *fred*, on the other hand, is significantly upregulated in the Slou subset with respect to Lms, but not versus Duf while *mgl* is significantly upregulated in the Slou subset with respect to both populations.

is significantly upregulated in the Lms subset versus the Slou subset in which it is significantly downregulated. *mgl* is significantly upregulated in the Slou subset with respect to Duf and Lms in which it is significantly downregulated. *fred* is significantly upregulated in the Slou subset only with respect to the Lms subset. Many chitin development genes such as *Mur2B*, *fred* and *mgl* have been observed to affect axon targeting and synapse development in this study.

3.1.2.4.2. Clustering and GO enrichment analysis

I performed a kmeans clustering analysis based on log fold changes to identify genes with similar expression patterns across different populations. For interesting clusters or lists of genes identified, I performed an enrichment analysis for GO and KEGG terms to help identify known processes and pathways that are enriched in specific sets or during specific timepoints.

3.1.2.4.2.1. Analysis of the Duf+ global muscle population

Muscle genes that are normally present in all muscles such as *Mlc1* and *wupA* are consistently significantly upregulated in the Duf population over all three time points (Figure 24). A heatmap clustering of genes that are similarly differentially regulated during all time windows reveals one set of genes that are consistently significantly downregulated in the muscles including the ribonuclease *RNaseMRP:RNA*, the acyl-CoA dehydrogenase *CG3902*, the small *nuclear RNA snRNA:U2:34ABb* and the small nucleolar RNA *snoRNA:Pst28S-2566* (Figure 24A). Genes that transition differently across time windows cluster into distinct groups. Among these genes are two coding for Like Sm proteins that are part of the spliceosome complex, *LSm3* and *LSm4* (Herold et al. 2009). *LSm3* is significantly upregulated during T2 while *LSm4* expression increases between T1 and T2 and then again between T2 and T3 (Figure 24B). The *LSm3* cluster also contains the Wnt signaling factor *dsh* and the LT iTF *ap*. For the global muscle population versus global embryonic genes, a GSEA enrichment analysis (Subramanian et al. 2005) combining significantly up and down regulated genes during T3 shows an enrichment for CCs related to the sarcomere and cytoskeleton (Figure 24C). Many of these genes were also identified as contributing to the variance on PC1 in the Duf+ population in the initial PCA analysis (Figure 14E') that made no *a priori* assumptions and only looked at probe intensities, unlike the GSEA analysis that compares data to gene sets known to be implicated in various processes. In the T3 time window, genes that are significantly upregulated show an enrichment for CCs related to the sarcomere.

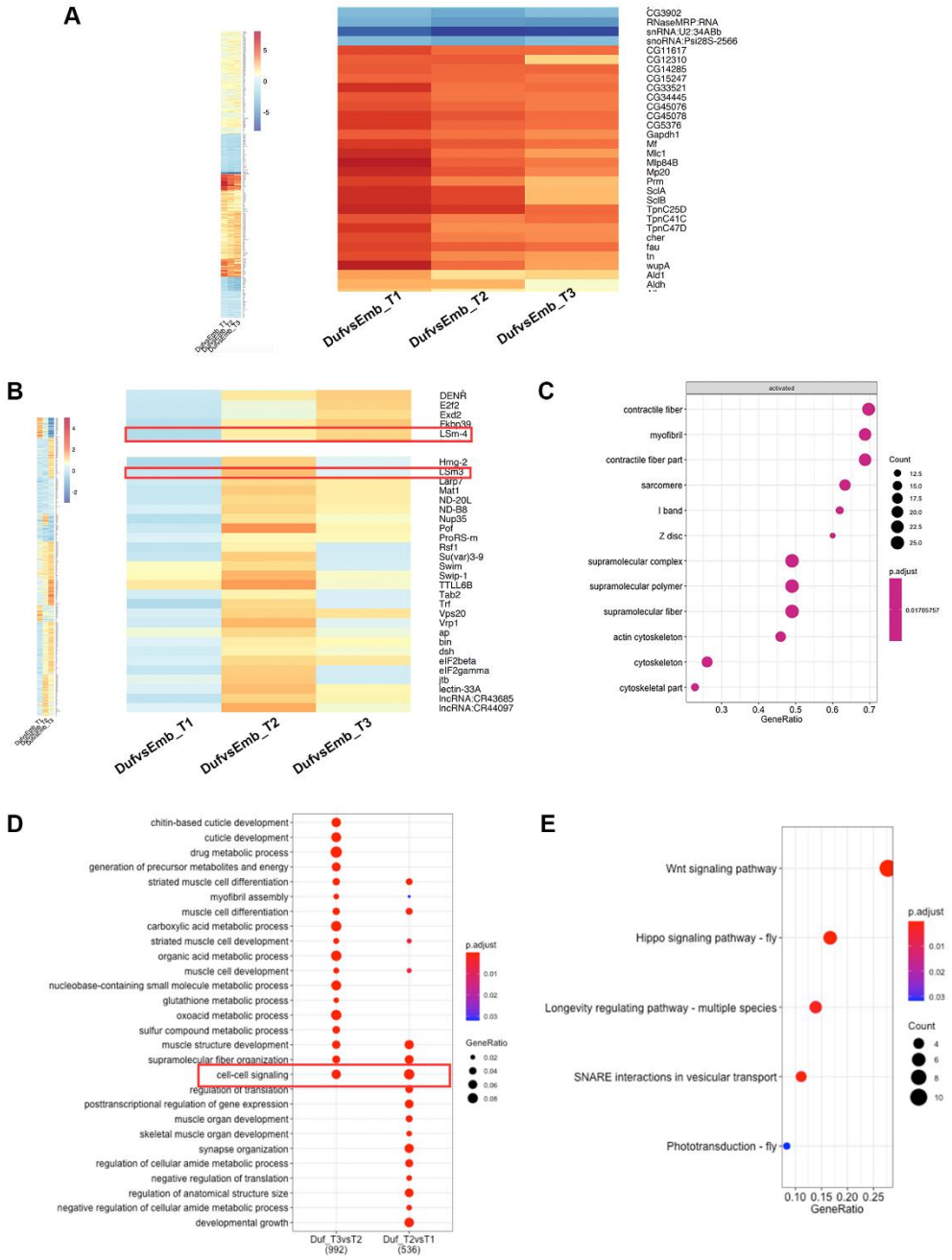


Figure 24. Clustering and GO enrichment analysis of the Duf population versus global embryonic mRNA at different timepoints.

(A) Certain muscle specific genes are significantly upregulated across all time windows in the Duf population (reds on the heatmap) while genes such as *RNaseMRP:RNA*, the acyl-CoA dehydrogenase *CG3902*, the small *nuclear RNA snRNA:U2:34ABb* and the small nucleolar RNA *snoRNA:Pst28S-2566* are significantly downregulated. **(B)** Among genes that have a specific temporal expression signature in muscles are the spliceosome components *LSm3* and *LSm4*. *LSm3* is only significantly upregulated during T2 whereas *LSm4* is upregulated between T1 to T2 to T3. **(C)** At T3, there is a significant enrichment for CCs associated with the cytoskeleton and sarcomere among significantly differentially regulated genes. **(D)** It becomes apparent when comparing the Duf timepoints against each other that there is a temporal pattern of gene expression related to specific BPs. While translation related BPs are significantly enriched between T1 and T2, muscle differentiation processes are enriched between T2 and T3. Among processes enriched between T2 and T3 is cell-cell signaling. **(E)** Differentially expressed genes that are part of the cell-cell signaling cluster are enriched for various KEGG signaling pathways including Wnt and Hippo signaling as well as for SNARE mediated vesicular transport.

I compared the Duf T3 dataset with T2 and T2 with T1 to verify genes that change between stages. Interestingly, between T1 and T2 there is a significant enrichment of genes associated with the regulation of translation. Between T2 and T3, there is an enrichment of genes associated with muscle cell differentiation (Figure 24D). Between these stages, multiple genes involved in cell-cell signaling including *Ssdp* (not all probes) are significantly upregulated. A search for KEGG enrichment among these genes reveals a significant enrichment for the Wnt and Hippo signaling pathways as well as SNARE mediated vesicular transport (Figure 24E). A comparison of the Lms and Slou muscle subsets with the global embryonic mRNA also reveals an enrichment for muscle differentiation related GOs. Given that these datasets are enriched for muscular genes as expected when compared to the total embryonic mRNA, thus validating them, I decided to compare the muscle subsets with the global muscle population.

3.1.2.4.2.2. Analysis of the Lms+ muscle subset versus the Duf population

A kmeans clustering of genes that are significantly upregulated during T3 in the Lms subset versus the Duf population reveals distinct clusters, some of which are specific to the Lms subset. *lms* is part of a cluster that is highly upregulated with respect to Duf unlike in the Slou subset (Figure 25A). This cluster includes the G protein-coupled receptor signaling factor *GABA-B-R2*, *Dr* and *kohtalo* (*kto*) (or *Med12*) while a sister cluster where genes are even more upregulated across all timepoints includes *su(Hw)* and *Su(P)*. Another Lms specific cluster where genes are significantly upregulated only during T3 includes another G protein-coupled receptor signaling factor *Ggamma30A* as well as *Ssdp* (all probes) that is implicated in the positive regulation of transcription via Wnt canonical signaling mediated by *Pygopus* (*Pygo*) along with *kto* (Carrera et al. 2008; Fiedler et al. 2015).

A comparison of GO enrichment between T2 and T3 for the GO term 'BP' among significantly upregulated genes reveals an enrichment for multiple genes involved in the negative regulation of transcription and biosynthetic processes during T2 (Figure 25B). For the GO term 'CC', there is an enrichment for localizations to proteasome complexes and transcriptional repressor complexes during T2 and for genes localizing to the mitochondria during T3 (Figure 25C). An analysis for the GO term 'MF' shows that during T3, actin binding genes as well as genes implicated in electron transfer and ATP synthesis are upregulated (Figure 25D). Actin binding genes include multiple genes of the ADF-H/gelsolin-like domain superfamily such as *Gel*, *Svil*, *tsr* and *qua*, the filamin gene *cher* and the fascin gene *sn* (Figure 25D'). Among the significantly downregulated genes in Lms during T2 and T3, almost all of them are similarly regulated in the

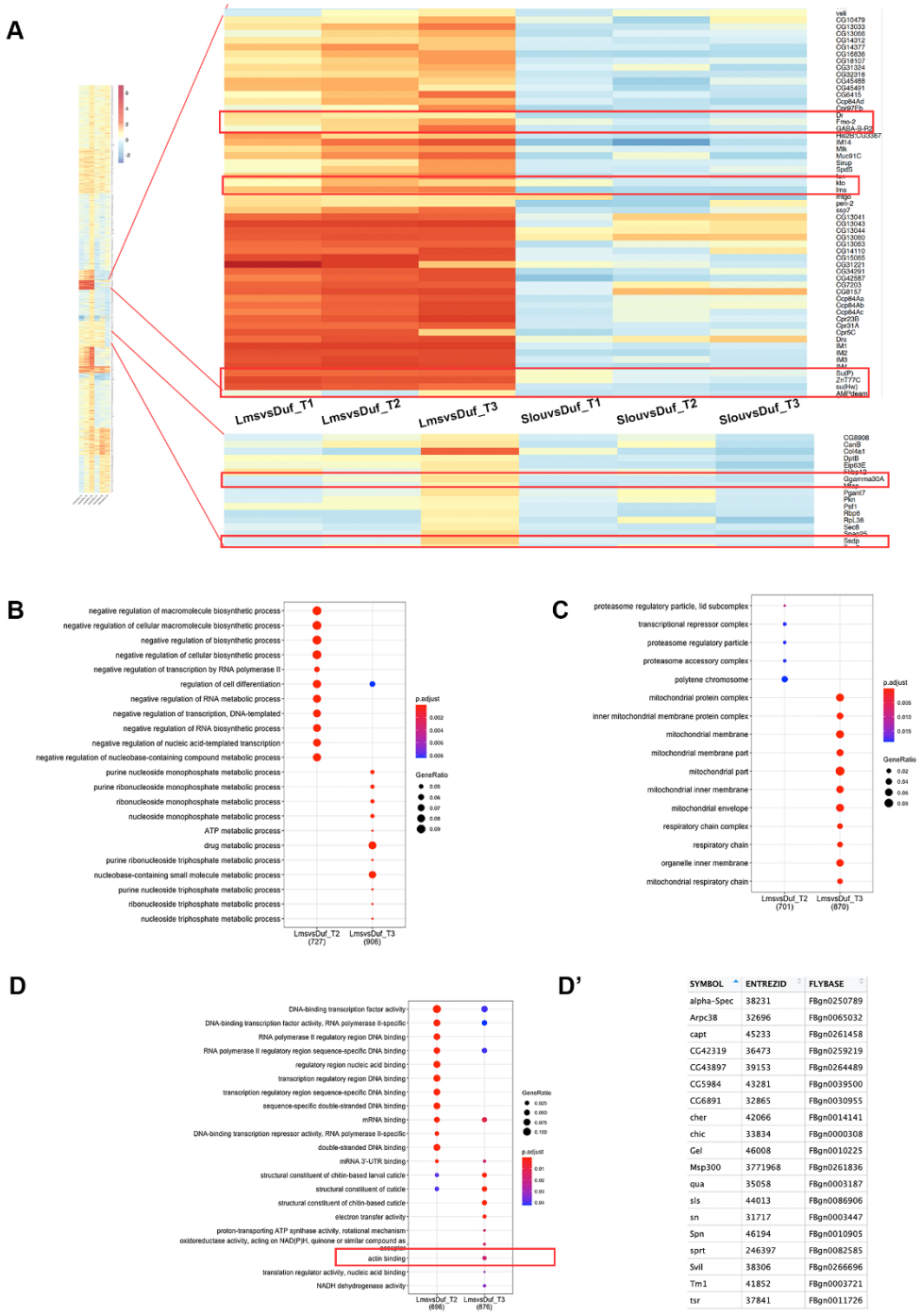


Figure 25. Clustering and GO analysis of the Lms subset versus the Duf population.
(A) Genes such as *lms*, the canonical Wnt effector *kto*, the G protein-coupled receptor signaling factor *GABA-B-R2*, *Su(P)*, *su(Hw)* and the LT muscle iTF gene *Dr*, are consistently significantly upregulated in the Lms subset during all three timepoints. Others, such as another G protein-coupled receptor signaling factor *Ggamma30A* and *Ssdp* are significantly upregulated only during T3 in the Lms subset. **(B)** Comparing the T2 timepoint with T3 reveals temporal differences in BPs. There is a significant enrichment of genes related to the negative regulation of processes during T2. **(C)** For the GO term 'CC', there is an enrichment for localizations to proteasomal complexes during T2 and mitochondria during T3. **(D)** For the GO term 'MF', genes involved in transcriptional regulation are very active during T2 while terms such as 'actin binding' are among those enriched during T3. **(D')** The list of genes upregulated in the Lms subset with respect to Duf during T3 annotated with the GO term 'actin binding' includes multiple proteins of the ADF-H/gelsolin-like domain superfamily such as *Gel*, *qua*, *Svil* and *tsr*.

Slou subset, except for one set of genes that is significantly upregulated in the Slou subset with respect to Duf.

3.1.2.4.2.3. Analysis of the Slou+ muscle subset versus the Duf population

In the Slou subset, there are more genes significantly differentially expressed during T2 contrary to T3 in the Lms population (Figure 20). Genes upregulated during T2 and T3 cluster into several distinct Slou specific clusters. Genes implicated in proliferation and differentiation are specifically significantly upregulated during T2. At T3, there is an increase in the number of genes implicated in BPs associated with transport, secretion and localization (Figure 26A). As a consequence, enriched genes localize to organelles such as the ER and Golgi complex during T3 (Figure 26B). Similar to the Lms population, there is an enrichment for mRNA binding genes during T2. In addition, there is an enrichment for enzyme binding genes.

3.1.2.4.2.4. Analysis of the Lms+ muscle subset versus the Slou+ muscle subset

A PCA of normalized, fitted data for genes in the Lms and Slou subsets reveals one principal component that contributes to around 20% of the variance that contains both Lms and Slou. A heatmap clustering of these genes shows that they form distinct groups with genes specifically significantly upregulated or downregulated in the Lms population (Figure 27A, B). Among the significantly upregulated genes in the Lms population are *lms* and *Gel* apart from the homeodomain TF genes *Dr*, *D* whose protein physically interacts with Lms, *HLH3B* and *Hmx*. This also includes the spliceosome component *LSm7* and the extracellular matrix (ECM) gene *fondue* (*fon*) implicated in muscle attachment (Green et al. 2016) (Figure 27C). Among the genes significantly downregulated in the Lms subset with respect to the Slou subset are the known Slou muscle iTF genes *slou* and *Six4* apart from *Stat92E*, *Hand*, *velo* and others. There is only an enrichment for the Toll signaling pathway and cuticle formation among these genes. This could, however, mean that additional functions of most of the genes are yet to be identified.

A side-by-side comparison of the significantly upregulated genes in the Lms and Slou subsets with fold changes against Duf reveals a temporal difference in the enrichment of processes as well as a specific enrichment of transport related processes in the Slou subset (Figure 27D). It also reveals that factors that localize to the mitochondria are enriched only during T3 in the Lms subset whereas in the Slou subset, they are enriched starting T2. Proteasomal localizations are enriched only in the Lms subset during T2 (Figure 27E). This might indicate either a

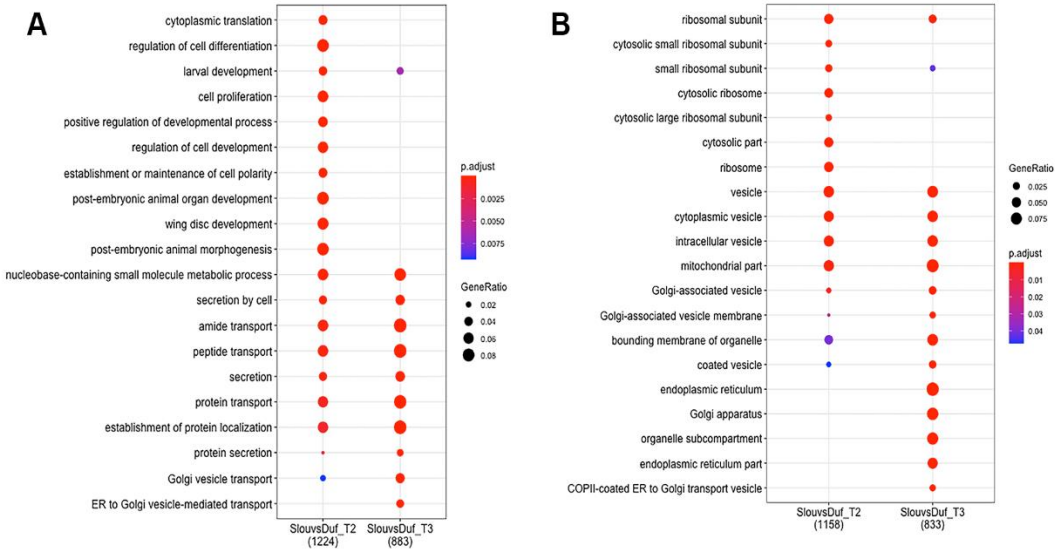


Figure 26. Clustering and GO analysis of the Slou subset versus the Duf population.
(A) Comparing significantly upregulated genes in SlouvsDuf T2 with SlouvsDuf T3 reveals an enrichment of biological processes related to secretion and transport during T3. **(B)** In line with this observation, there is an enrichment of localizations to the cellular components ER and Golgi complex during this time window.

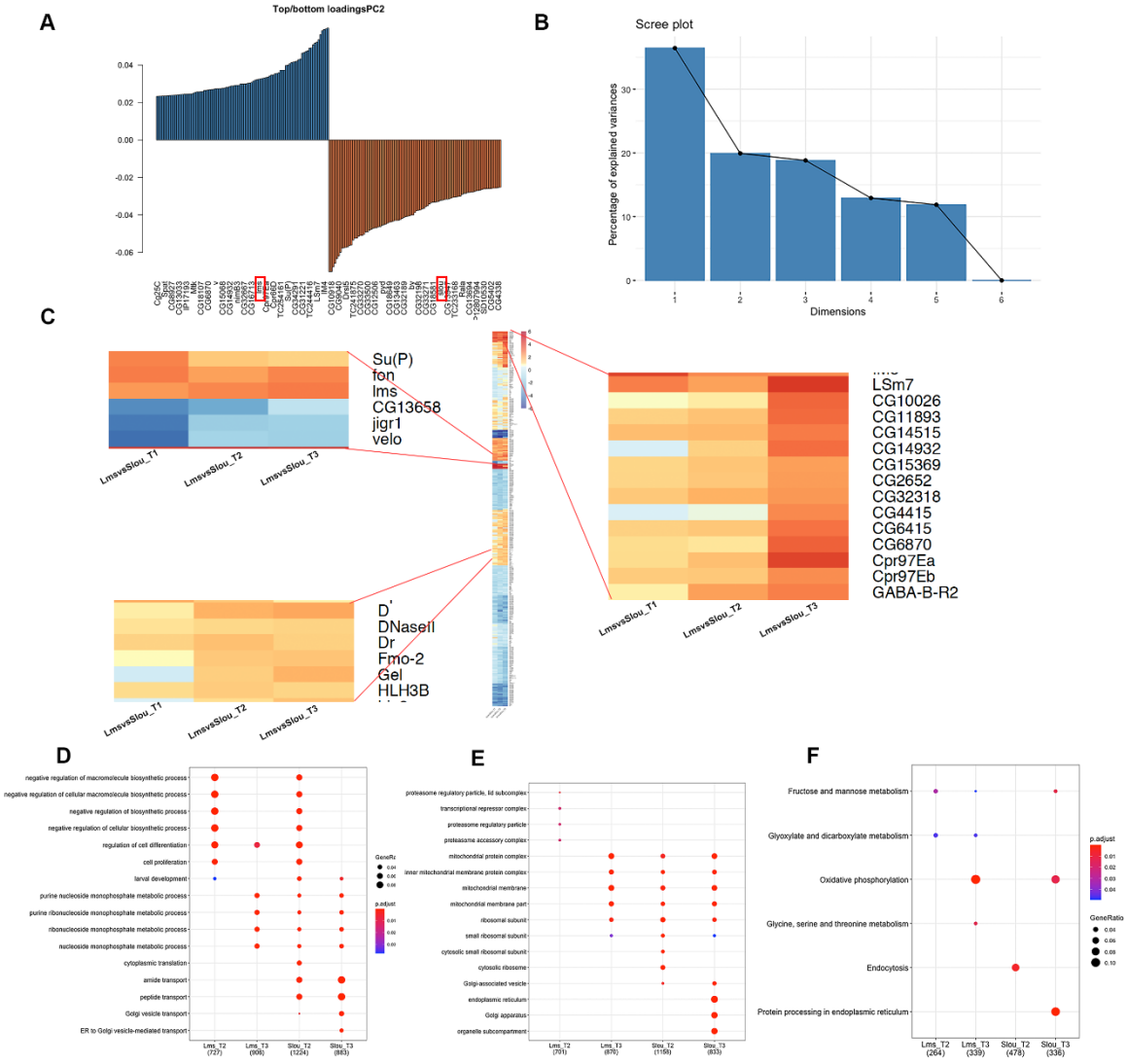


Figure 27. Clustering and GO analysis of the Lms subset versus the Slou subset.
(A) A PCA of normalized, fitted Lms and Slou datasets has these iTFs among the top loading genes on PC2 that explains around 20% of the variance. **(B)** The percentage of variance explained by each PC is displayed in the scree plot. **(C)** A kmeans clustering of the top loading genes on PC2 generates distinct clusters with some Lms specific clusters such as one containing *lms*, one containing *Gel*, *Dr and D* and another with *LSm7* and *GABA-B-R2*. A cluster containing *velo* is downregulated in the Lms subset. The heatmap displays logFCs at T1, T2 and T3 for Lms vs Slou. **(D)** BP terms related to transport are specifically enriched among significantly upregulated genes in the Slou subset. LogFCs in both subsets are against Duf. **(E)** In line with the data in (D), CCs associated with the ER and Golgi complex are significantly enriched. **(F)** Among the KEGG pathways significantly enriched is the glycine, serine and threonine metabolism pathway among significantly upregulated genes in the Lms subset during T3.

developmental lag between the two populations or a true difference defining muscle identity. The Lms T3 dataset is also enriched for the KEGG pathway glycine, serine and threonine metabolism (Figure 27F).

Among differentially expressed TF genes in the Lms subset with respect to the Slou subset, during T3, there is an enrichment of genes implicated in cuticle development among significantly upregulated genes while significantly downregulated genes are enriched for serine peptidase and hydrolase activities (Figure 28A). In conjunction with the significant enrichment of glycine, serine and threonine metabolism that is seen in the upregulated genes in the Lms subset, this suggests that an upregulation in serine metabolism correlates with a downregulation in serine peptidase activity. A comparison of GO enrichment among significantly up and down regulated genes at different timepoints shows a significant enrichment of cell fate specification genes at T1 in the Lms subset followed by mRNA splicing and transcription during T2 that continues into T3 (Figure 28B). The HOMER software detects a significant enrichment of genes with the PFAM 'Homeobox' protein domain among significantly upregulated genes in the Lms subset during T3 (Figure 28C). It is interesting to note the differential expression of various known muscular genes that are associated with actin dynamics as well as sarcomere formation in the two subsets at different timepoints. Actin binding genes such as *Gel*, *qua*, *cher* and *fln* are significantly upregulated at T3 in the Lms subset. Genes linked to sarcomere formation such as *wupA, up, sals, Mlc1, Mlc2, sls, zasp52, zasp66, Tm1* and *Tm2* are very slightly upregulated in the Lms subset with respect to Slou (Figure 28D). Again, this might either indicate temporal differences in the regulation of genes or a necessity for differential expression levels of these genes to define muscle subset identity. The TF Ssdp is significantly upregulated only during T3 in the Lms subset. A KEGG pathway enrichment analysis reveals an enrichment for RNA degradation as well as spliceosomal pathways during T2 in the Lms subset (Figure 28E).

A reverse comparison of the Slou subset versus the Lms subset reveals an enrichment in muscular BPs among significantly upregulated genes at early and mid stages while late stages are enriched for ER and Golgi transport. BPs related to the negative regulation of localization, RNA metabolism and transcription are enriched among significantly downregulated genes at T1 while mRNA and neurotransmitter metabolism are enriched at T2 and T3 (Figure 28F). Considering all significantly downregulated TFs during all timepoints, there is a significant downregulation for various KEGG signaling pathways including the mTOR pathway that is upregulated in the Lms subset in the temporal profiling.

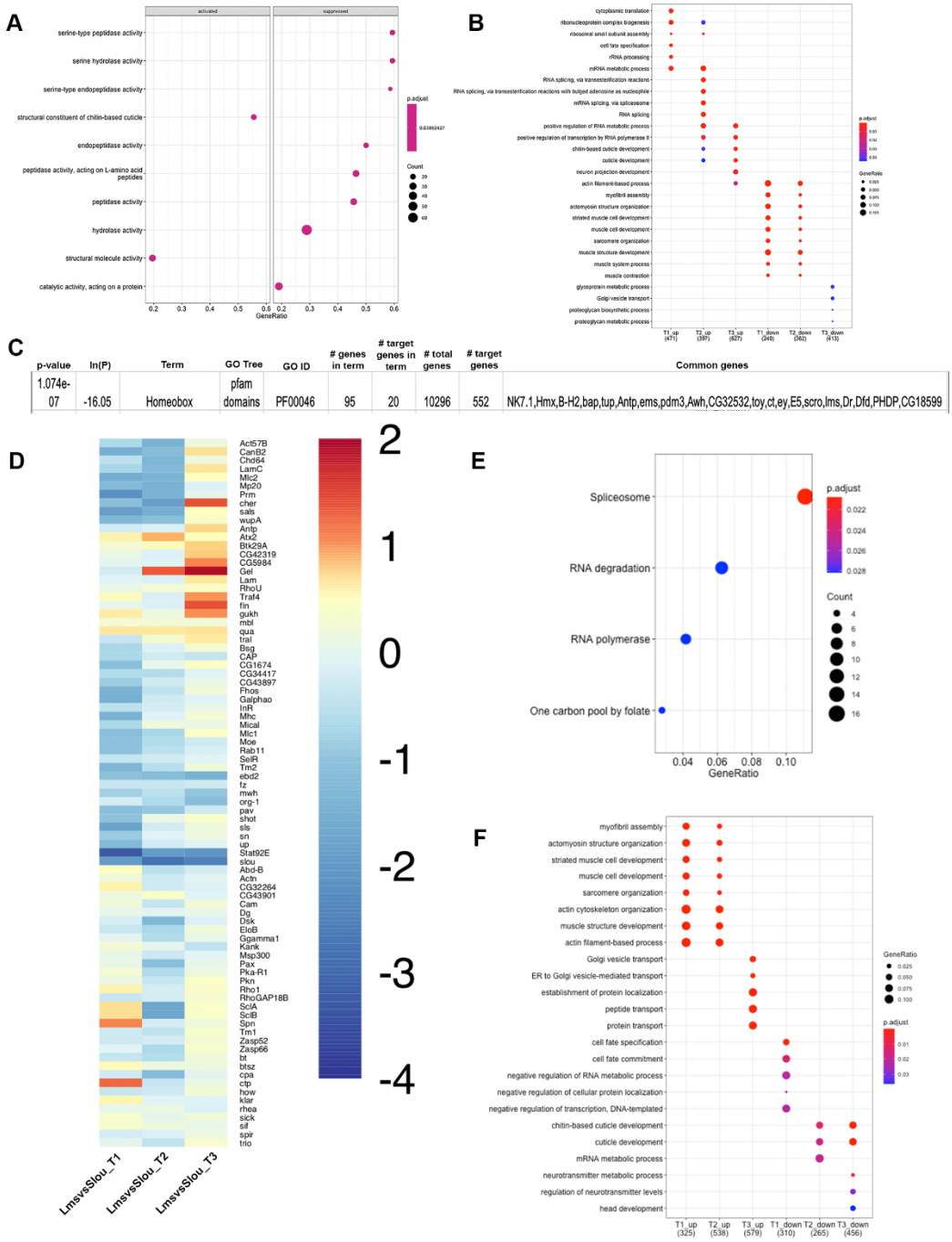


Figure 28. Comparison of the Lms and Slou muscle subsets against each other without considering the Duf population.

(A) Considering significantly differentially expressed TFs in Lms versus Slou, serine peptidase and hydrolase activities are significantly enriched among the downregulated genes. **(B)** Among all significantly differentially expressed genes, multiple BP terms associated with muscle development and differentiation are downregulated in the Lms subset during T1 and T2 whereas mRNA splicing and metabolism are upregulated. Golgi vesicle transport is downregulated during T3. **(C)** The HOMER software detects an enrichment for 'Homeobox' PFAM protein domain among upregulated genes during T3 in the Lms subset. **(D)** Differentially expressed genes annotated to muscles present subtle expression dynamics among the two subsets. For example, actin binding genes such as *Gel*, *qua*, *fln* and *cher* are specifically significantly upregulated in the Lms subset. **(E)** During T2, among upregulated genes, there is a significant enrichment for KEGG pathways associated with the spliceosome and RNA degradation in the Lms subset. **(F)** A Slou versus Lms comparison results in an enrichment of Golgi transport related processes during T3 among upregulated genes.

3.1.2.4.3. *de novo* and known cis regulatory element enrichment analysis

In order to determine if there was an enrichment of specific motifs among the significantly differentially regulated genes, I performed an analysis using i-cisTarget (Herrmann et al. 2012; Imrichová et al. 2015) that searches for enrichment of *de novo* and known TF motifs and histone modifications 5KB upstream of genes as well as within transcribed regions. I also performed a *de novo* motif search using HOMER and the MEME Suite (Bailey et al. 2015). Such an enrichment would indicate coregulation of significantly differentially expressed genes by specific TFs or cofactors.

3.1.2.4.3.1. Duf+ muscle population

Among the significantly upregulated genes during T1, T2 and T3 in the Duf population compared to total embryonic mRNA, the motifs most significantly enriched are those of the orthologues of the muscle differentiation TF, Myocyte enhancer factor (Mef2) from various species including yeast, humans and mice (Figure 29). During T1, there is also a significant enrichment for the vertebrate JUND motif, a TF that is orthologous to the *Drosophila* Jun-related antigen (Jra) as well as vertebrate ATF7 that binds cAMP responsive elements on DNA (Persengiev, Devireddy, et Green 2002) (Figure 29A). During T2 is a significant enrichment for motifs for the myogenic factor Chorionic factor 2 (Cf2) and vertebrate TFAP4 that is orthologous to the myogenic factor Cropped (Crp) in *Drosophila* (Dobi, Halfon, et Baylies 2014). modERN targets for Diminutive (Dm) (also known as Myc) are also significantly enriched (Figure 29B). During T3, motifs for Cf2 and the chromatin remodeling factor DREF along with Dm modERN (Kudron et al. 2018) predicted targets are among the most significantly enriched (Figure 29C). A *de novo* motif enrichment analysis by HOMER predicts two enriched *de novo* motifs in the Duf versus embryonic mRNA set during T3, one of which closely resembles the Ultraspiracle (Usp) motif (Figure 29D). This TF has been implicated in adult flight muscle development (Zappia et al. 2019).

Putative modERN targets for neurogenic TFs such as Pros and Jim as well as the vertebrate ZNF384 motif orthologous to *Drosophila* Squeeze (Sqz) are significantly enriched among significantly downregulated genes at all timepoints (Figure 29E).

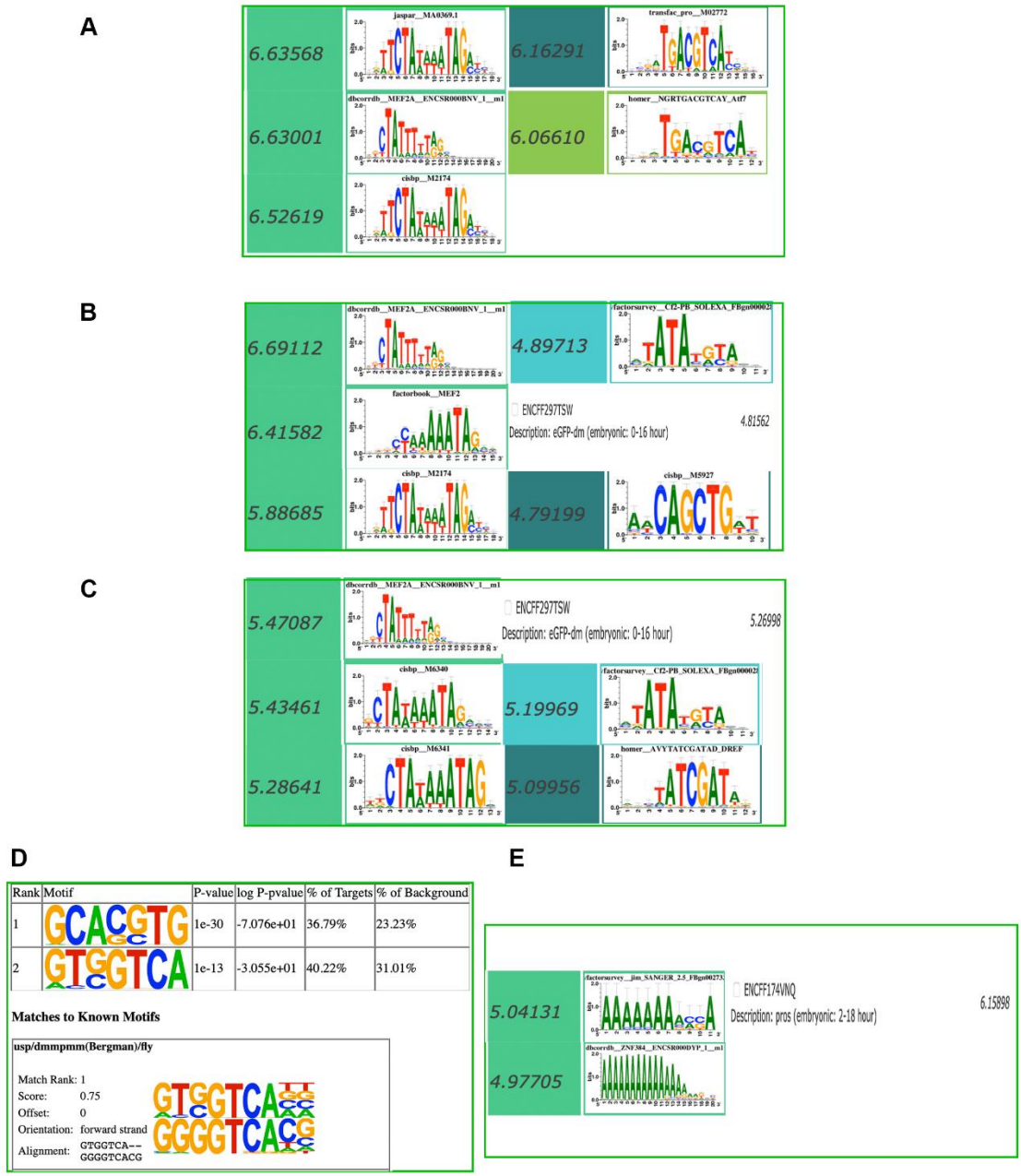


Figure 29. Cis regulatory element enrichment analysis in the Duf population versus global mRNA. For analysis performed using i-cisTarget, a normalized enrichment score (NES) threshold of 3 was applied. The numbers on the left of each enriched entity represent the NES. Vertebrate and invertebrate motifs for the evolutionarily conserved muscle differentiation factor Mef2 are highly enriched during T1, T2 and T3 among significantly upregulated genes (motifs on the left-hand side in A, B and C). **(A)** During T1, there is also an enrichment for the vertebrate JUND motif, orthologous to the *Drosophila* Jun-related antigen (Jra), and AFT7 motifs. **(B)** During T2, there is a significant enrichment for motifs for the myogenic TF, Cf2 and vertebrate TFAP4, orthologous to *Drosophila* Cropped (Crp), and for putative modERN ChIP-Seq targets for Diminutive (Dm), also known as Myc**. (C)** During T3, there is significant enrichment for motifs for Chorionic factor 2 (Cf2) and the chromatin remodeling factor DREF along with putative modERN Dm targets. **(D)** During T3, HOMER predicts two significantly enriched *de novo* motifs, with one closely resembling the Ultraspiracle (Usp) motif. **(E)** A-rich motifs and putative modERN targets for neurogenic TFs such as Pros, Jim and the vertebrate ZNF384, orthologous to the *Drosophila* Squeeze (Sqz), are among the most significantly enriched cis regulatory elements among significantly downregulated genes during all time windows. T3 is shown in the figure.

3.1.2.4.3.2. Lms+ and Slou+ muscle subsets

Among genes significantly upregulated in the Lms subset with respect to the Slou subset, Misexpression suppressor of ras4 (MESR4) and Ets at 97D (Ets97D) modERN peaks are most significantly enriched during T1. During this timepoint, motifs for Cf2 and vertebrate SIX6 and FUBP1 that binds the far upstream elements (FUSE) of genes such as c-Myc (Elman et al. 2019) are most enriched among significantly downregulated genes (Figure 30A, B).

During T2, in the Lms subset in comparison against the Slou subset, there is an enrichment for vertebrate motifs that bind different RXR-ALPHA heterodimers, which regulate retinoic acid triggered transcription (Figure 30A'). The *Drosophila* RXR orthologue, Usp complexes with ecdysone receptor (EcR) and has been shown to be capable of heterodimerizing with vertebrate RXR to regulate transcription (H. E. Thomas, Stunnenberg, et Stewart 1993; Yao et al. 1992). EcR peaks predicted by modERN are also enriched. The motif for the vertebrate Zinc-finger ZZ-type Containing 3 (ZZZ3) that is orthologous to Ada2a-containing complex component 1 (Atac1) is also enriched. It is part of the highly conserved Ada2-containing (ATAC) complex involved in histone acetylation (Guelman et al. 2006). Vertebrate CREB motifs are highly enriched among the significantly downregulated genes (Figure 30B'). *CrebB* is among the genes significantly upregulated in the Slou versus Lms comparison at this timepoint, although it is not significantly downregulated in Lms versus Slou indicating a fold change less than 2 in Slou versus Lms. Putative targets include 62.68% of genes significantly upregulated in Slou versus Lms during T2. These include muscle genes such as *wupA*, *Prm*, *Mf*, *Pax*, *sls*, *Tm1* and *Tm2*. Therefore, this could be part of the defining factors of muscle identity of the two subsets.

During T2, comparing Slou versus Lms, among significantly downregulated genes is an enrichment for Nf-YB motifs (Figure 30C). Schnorrer et al. (Schnorrer et al. 2010) identified one RNAi line for this TF as producing weak flier phenotypes. Among the significantly downregulated genes containing putative binding sites for this TF are the LT muscle iTF genes *lms*, *Dr* and *ap* along with other TFs significantly upregulated in the Lms population such as *D*, *SoxN*, *Sox21b*, *E5*, *hkb* and *bap*. Nf-YB could, thus, potentially be upstream of these LT muscle specific TFs. This gene was shown to activate the transcription of tissue specific catalase gene in culture (Luo et Rando 2003). I downloaded the modERN ChIP-Seq peak file and mapped them to putative targets in order to verify how significant the peak calls were for these genes. All of these TFs have putative binding sites mostly in the upstream region close to the TSS, in introns in some cases and rarely in the 5' UTR. Among the top 100 peaks are regions in TF

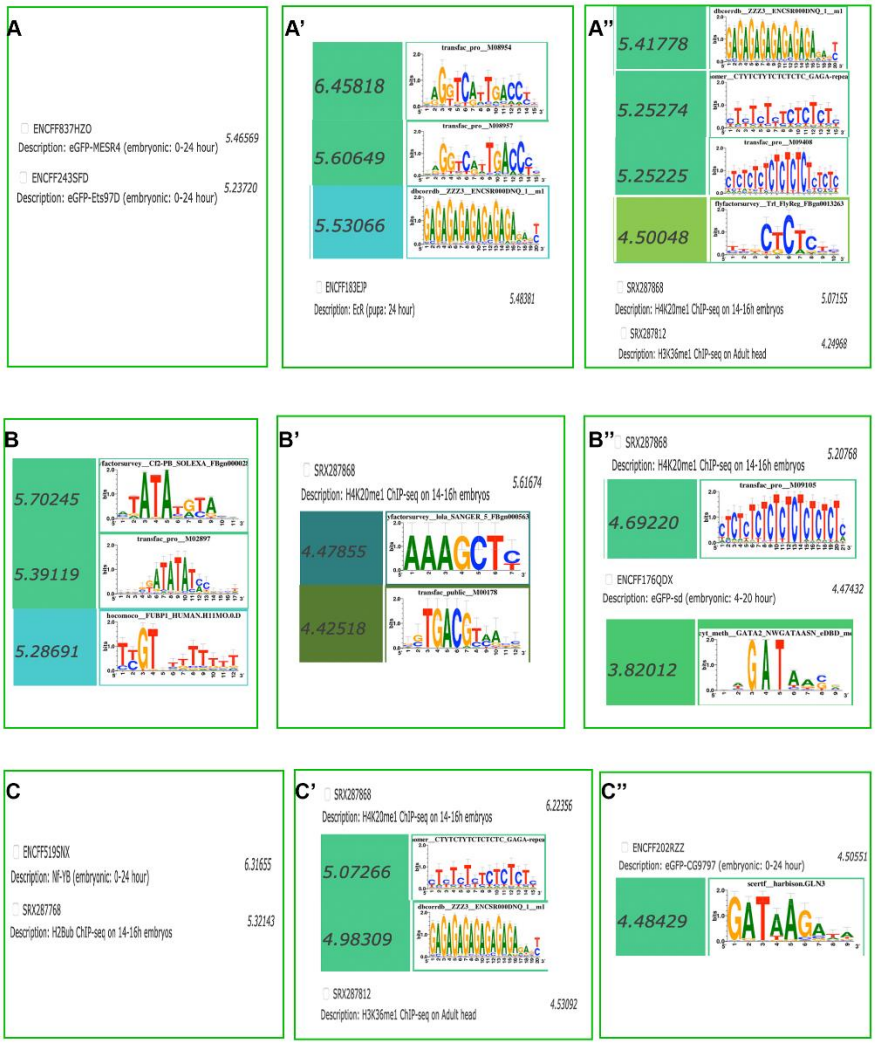


Figure 30. Cis regulatory element enrichment analysis in the Lms and Slou muscle subsets. **(A-A'')** During T1 (A), in the Lms subset versus the Slou subset, modERN ChIP-Seq targets for MESR4 and Ets97D are significantly enriched among significantly upregulated genes. During T2 (A'), a significant enrichment for two vertebrate motifs binding RXR-ALPHA heterodimers is observed along with the vertebrate motif for ZZZ3, orthologous to Ada2a-containing complex component 1 (Atac1). modERN predicted EcR targets are also significantly enriched. GA-rich vertebrate ZZZ3 motifs as well as complementary CT-rich plant and *de novo* motifs are the most enriched during T3 (A''). One known CT-rich motif for Trl is also enriched. Apart from this, there is an enrichment in H4K20 and H3K36 methylation histone modifications. **(B-B'')** Among genes significantly downregulated in the Lms subset versus Slou during T1, there is significant enrichment for Cf2 and vertebrate SIX6 (Six4 is a Slou specific iTF) and FUBP1 motifs. FUBP1 binds far upstream elements (FUSE) of c-Myc (B). Among downregulated genes in the Lms subset during T2, there is significant enrichment for H4K20 methylation as well as Lola and CrebB motifs (B'). Among significantly downregulated genes during T3, H4K20 methylation and CT-rich motifs are significantly enriched among this list. modERN targets of the ventral iTF cofactor Scalloped (Sd) are also significantly enriched. (B''). Apart from this, there is significant enrichment for vertebrate GATA motifs. **(C-C'')** An inverse comparison of the Slou subset versus the Lms subset reveals a significant enrichment for Nf-YB motifs during T2 (C). In the Slou versus Lms comparison, CT-rich and GA-rich motifs are significantly enriched among genes significantly downregulated during T3 (C'). Among significantly upregulated genes in Slou versus Lms, putative modERN targets for Motif 1 Binding Protein (M1BP, CG9797) are the most enriched. GATA motifs are also highly enriched contrary to the enrichment of this motif among significantly downregulated genes in the Lms subset.

genes such as *ap*, *ara*, *caup*, *D*, *E5*, *ey*, *lbe*, *lbl*, *pnr*, *Sox102F*, *sr*, and *SoxN*. Among the top 200 are the gene encoding the Ap competitor Bx (Bronstein et al. 2010), *Dfd*, *Dr*, *emc*, *ems*, *eya*, *HLH54F*, *kn*, *mid*, *H15*, *odd*, *slou*, *Sox15*, *Sox21b*, *twi*, *zfh1* and *zfh2*. *lms* is among the top 850 with a predicted binding site 177bp from the TSS. So, Nf-YB appears to be potentially upstream of many iTFs including non-LT iTFs.

Among CRM motifs significantly enriched during T3 among significantly upregulated genes in the Lms versus Slou comparison are motifs for vertebrate ZZZ3 (Figure 30A''). The gene coding for Ada2a that is part of the ATAC complex is among the genes significantly upregulated during T3. The ZZZ3 motif is rich in GA content. Complementary CT-rich known plant motifs as well as *de novo* motifs are also highly enriched in this list. Among the significantly upregulated genes containing the ZZZ3 motif is the evolutionarily conserved *Ssdp* that was originally identified as a protein that binds CT-rich tracts *in vitro* (Bayarsaihan, Soto, et Lukens 1998). Apart from this is an enrichment for H4K20 and H3K36 methylation and the motif for Trithorax like (Trl) that is implicated in chromatin modification and potentially the rate of paused RNA polymerase II that plays a key role in development by *Hox* gene interaction (Shimojima et al. 2003; Tsai et al. 2016). Among the motifs enriched among significantly downregulated genes during T3 in the Lms population are modERN peaks for Scalloped (Sd), implicated in ventral muscle identity, as well as vertebrate GATA motifs (Figure 30B''). The CT-rich motif and H4K20 methylation are also among the enriched motifs in this list as well. Conversely, among significantly upregulated genes during T3 in Slou versus Lms, the GATA motif is enriched. Interestingly, Ada2a is significantly downregulated in this subset. On the other hand, Motif 1 Binding Protein (M1BP, CG9797) putative modERN targets are the most enriched (Figure 30C''). This protein is involved in transcriptional pausing and *Hox* gene interaction. The human cyclic AMP-dependent Transcription Factor (ATF2) motif is also significantly enriched. The GA-rich and CT-rich motifs enriched among significantly upregulated Lms T3 genes are enriched among significantly downregulated genes during T3 in this population. So are H4K20 and H3K36 histone methylations (Figure 30C').

3.1.2.4.4. Analysis of transcription factors and cofactors

3.1.2.4.4.1. Protein and gene interaction analysis

Given the enrichment of cis regulatory elements, the T3 timepoint seems extremely interesting in terms of final muscle identity establishment. An analysis of gene interactions among TFs and

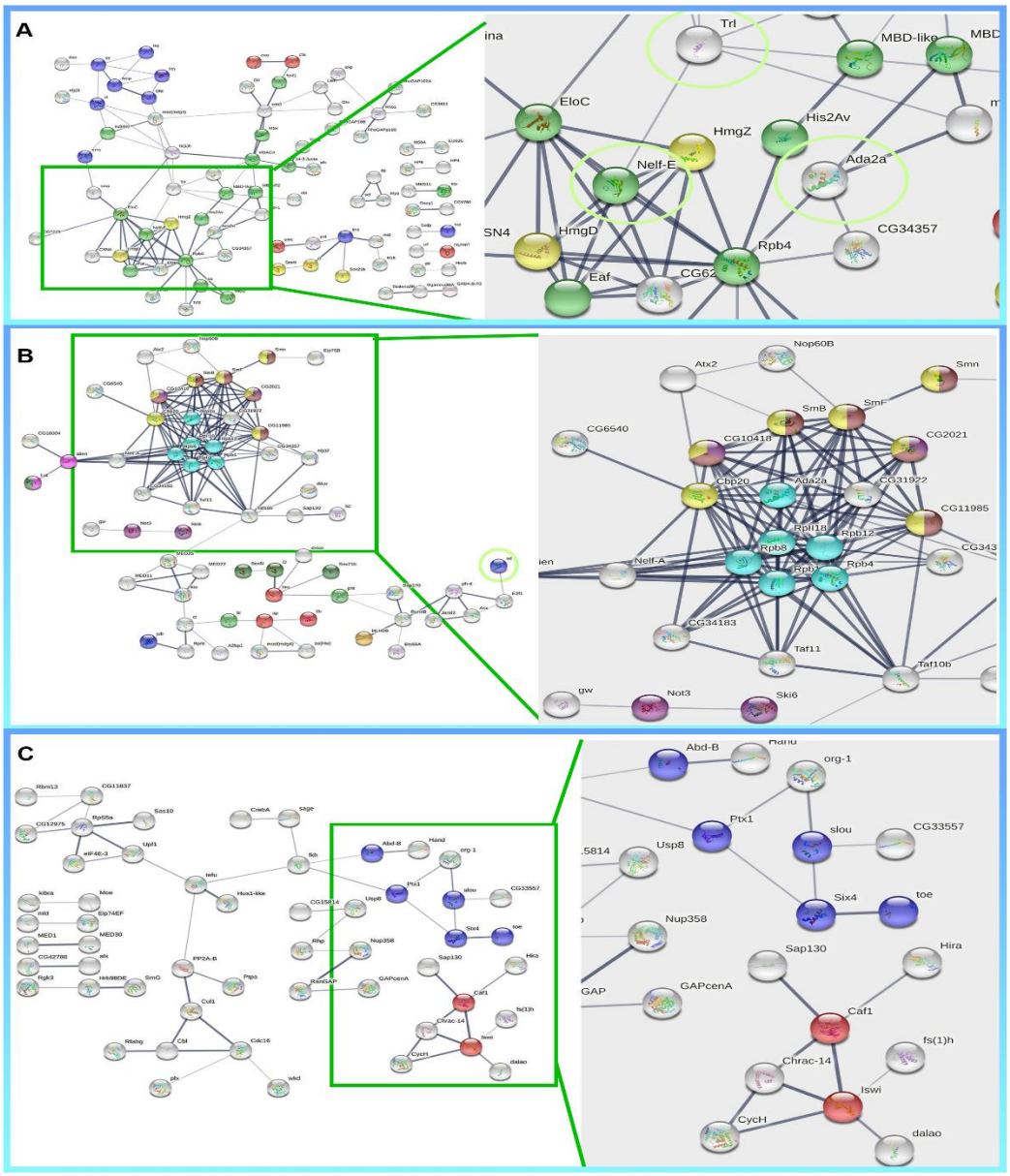


Figure 31. Protein/gene interactions identified using the String database among TFs significantly upregulated in the Lms subset versus the Slou subset.

Disconnected nodes are hidden. A zoomed view is presented for parts of each network. **(A)** During T3, TF and cofactor analysis reveals a strongly interconnected network of proteins comprising Nelf-E that indicates a protein complex. This complex connects to Ada2a via Rpb4. Ada2a and Rpb4 genetically interact. Trl whose binding motif is enriched in this subset also genetically interacts with Ada2a. Proteins involved in the 'gene expression' KEGG pathway (light green circles), HMG-box domain proteins (olive green), Homeobox domain proteins (dark blue) and helix-loop-helix domain proteins (red) are highlighted**. (B)** During T2, there is a complex comprising components of the RNA polymerase KEGG pathway (light blue) and those regulating RNA splicing (yellow) and spliceosome components (brown). Proteins involved in the RNA degradation pathway (purple) are also enriched. In dark blue are proteins involved in the Hippo signaling KEGG pathway. It is interesting to note that non-LT iTF, *sd* is significantly upregulated during T2 (green circle). HMG proteins are highlighted in green, helix-loop-helix domain proteins in orangish yellow and homeodomain proteins in red. In pink are components enriched for the 'CC' term ecdysone receptor holocomplex**. (C)** Among genes significantly upregulated in the Slou subset versus the Lms subset at T3, are Slou specific iTFs such as Slou, Six4, Ptx1 and Org-1. Selected significantly enriched terms are highlighted; homeodomain TFs are highlighted in blue and components of the NURF complex including Iswi in red.

cofactors confirms the cis regulatory element enrichments observed (Figure 31). During T3 in the Lms subset, it becomes clear that *Negative elongation factor E* (*Nelf-E*) is also significantly upregulated and forms a strongly interlinked structure indicating a complex. This complex also links to Ada2a and Trl whose binding motif is enriched in this subset (Figure 31A). Nelf-E genetically interacts with Trl (Chopra et al. 2009; Tsai et al. 2016). During T2, there is a complex of components implicated in splicing (Figure 31B). Among the significantly downregulated genes are multiple ventral muscle iTFs such as Slou, Six4 and Org-1. In the Slou subset, during T3, there is a complex centered around the chromatin remodeling factor Imitation SWI (Iswi) (Badenhorst et al. 2002) (Figure 31C). So, there appear to be differences in chromatin modification factors between these muscle subsets.

Looking at the interactions among significantly downregulated TFs and cofactors, distinct differences between the two subsets become obvious. Known Slou iTF genes such as *slou*, *Six4* and *org-1* are significantly downregulated in the Lms subset along with the Pax TF *twin of eyg* (*toe*), the actin-binding protein gene *Moesin* (*Moe*) and the positive regulator of the Hippo signaling pathway *kibra*. In the Slou subset on the other hand, at T3, these genes are significantly upregulated along with *Ptx1* whereas known LT iTFs such as *lms* and *mid* are significantly downregulated along with other genes that are significantly upregulated in the Lms subset such as *Mtor*, *Ssdp*, *tup*, *D*, *ct*, *ey*, *ems*, *SoxN* and *Sox21b*.

3.1.2.4.4.2. modERN data analysis for differentially expressed TFs

Since our data is based on GFP-tagged enhancer trap lines, I looked at TFs available in the modERN database, mapped peaks to genes and compared this to the list to genes significantly differentially expressed in muscle subsets to look for potential targets of these TFs among co-expressed genes (Figure 32). I started with TFs significantly upregulated during T3 in the Lms subset. Some putative target genes are common to multiple significantly upregulated TFs. Caup putative targets include a peak 125bp from the TSS of *Trl* (top 450) and upstream of *Sox14* (top 100) and *elav* (top 100) (Figure 32A). Among Tup putative targets are *Trl* (top 350), *Sox14* (top 100), *elav* (top 200) and *Nelf-E* (top 250). They also include *org-1*, which is significantly downregulated in the Lms subset (Figure 32B). *Ssdp* is among the top 35 modERN ChIP-Seq putative targets for Trl at 8-16h AEL with a predicted binding site in an intron. *Hmx* and *E5* are among the top 650 (Figure 32C). D putative targets include itself, *elav* and *Sox14* in the top 100 peaks. This might indicate a regulatory hierarchy in gene expression (Figure 32D).

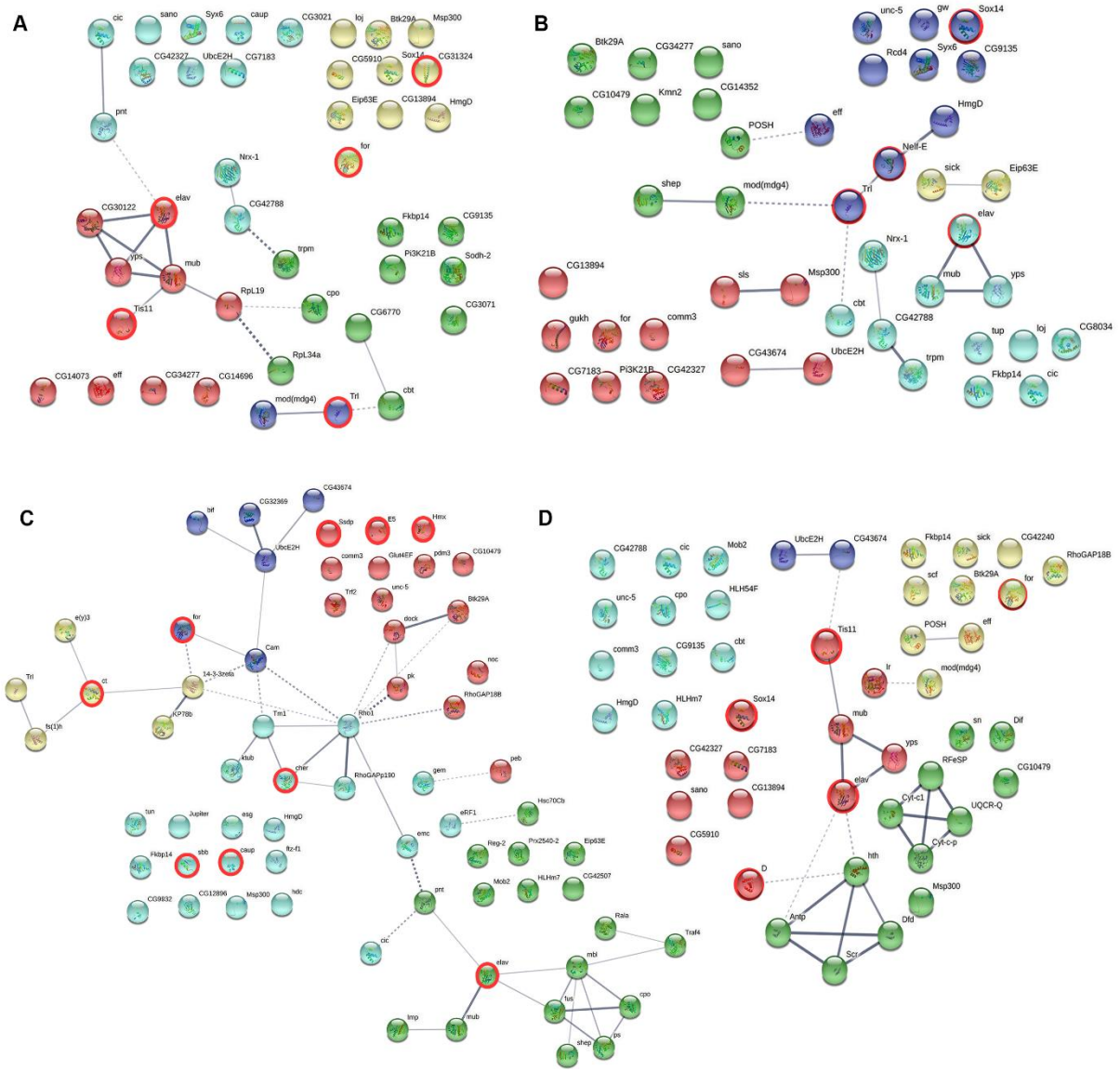


Figure 32. Protein/gene interaction networks of genes common between modERN putative targets of TFs significantly differentially expressed and significantly upregulated genes during T3 in the Lms subset.

(A) A putative Caup binding site is detected 125bp upstream of the TSS of *Trl*. Other putative targets include *Tis11*, *for*, *elav* and *Sox14* (outlined by red circles). **(B)** Among the top 350 Tup putative targets are *Trl*, *Sox14*, *elav* and *Nelf-E*. **(C)** *Ssdp* is among the top 35 modERN ChIP-Seq putative targets for Trl at 8-16h AEL with a predicted binding site in an intron. Other putative targets include *Hmx*, *caup*, *elav sbb*, *for*, *ct* and *E5* along with *cher*. **(D)** D putative targets include itself, *elav* and *Sox14* in the top 100 peaks apart from *Tis11* and *for*.

org-1 is identified as a putative Slou target among genes that are significantly upregulated in the Slou subset (Figure 33A). Interestingly, there are 26 Slou putative targets that are common among significantly downregulated genes in the Slou subset during T3, all of which are significantly upregulated in the Lms subset including *elav*, *Sox21b*, the gene coding for SNARE binding protein Syntaxin 6 (Syx6), *Tis11* and the gene coding for the chromatin organizing factor HmgD (Figure 33B). Same is the case for Stat92E whose putative targets include *Gel* (top 200) and *Ssdp* with a binding site predicted 619bp upstream (top 350). This suggests that these TFs might function as transcriptional repressors of these Lms specific genes. Another possibility is that the Lms subset has an expression boost of certain genes unlike the Slou subset.

3.1.2.4.5. Selection of candidate genes

I wanted to select candidates that were not the obvious choice, but could play logical roles in muscle identity acquisition. With this aim, I looked at annotated gene functions for genes, preferably with corresponding vertebrate orthologues. There are multiple recurring genes significantly differentially expressed among the muscle subsets with vertebrate homologues or known orthologues including *Ssdp* (*SSBP2/3/4*), *Tis11* (*ZFP36*) and *qua* (*AVIL*/*VIL1*/*VILL*) that are significantly upregulated in the Lms subset and *velo* (*SENP6/7*) that is significantly downregulated in the Lms subset and upregulated in the Slou subset. *tsr* (*CFL2*) is significantly upregulated in the Lms subset versus Duf only during T3 along with *Gel* and *qua* whereas it is significantly upregulated during T2 and T3 in the Slou subset. *sbb* (*ZNF608/609*) is a gene that displays a slightly different temporal transition profile in the Lms subset compared to the Duf and Slou subsets. Its expression falls between T1 and T2 and increases slightly during T3 in the Lms subset whereas it consistently falls in the others. From the preceding analyses, it appears that minor differences in expression levels might contribute to muscle identity. Lms specific homeobox domain TFs NK7.1 and Hmx as well as the bHLH TF emc are interesting given their role as TFs. Nf-YB is an interesting candidate to analyze the genes upstream of iTFs, but was not analyzed at this juncture. iTFs themselves are interesting candidates to study genetic interactions between them. So, I also chose to analyze LT muscle iTFs such as Ap, Lms and Mid.

3.1.2.4.6. Analysis of interactions of candidate genes

I performed an initial interaction analysis with the help of the String database (Figure 34). I perused available literature associated with these genes in String before choosing to analyze

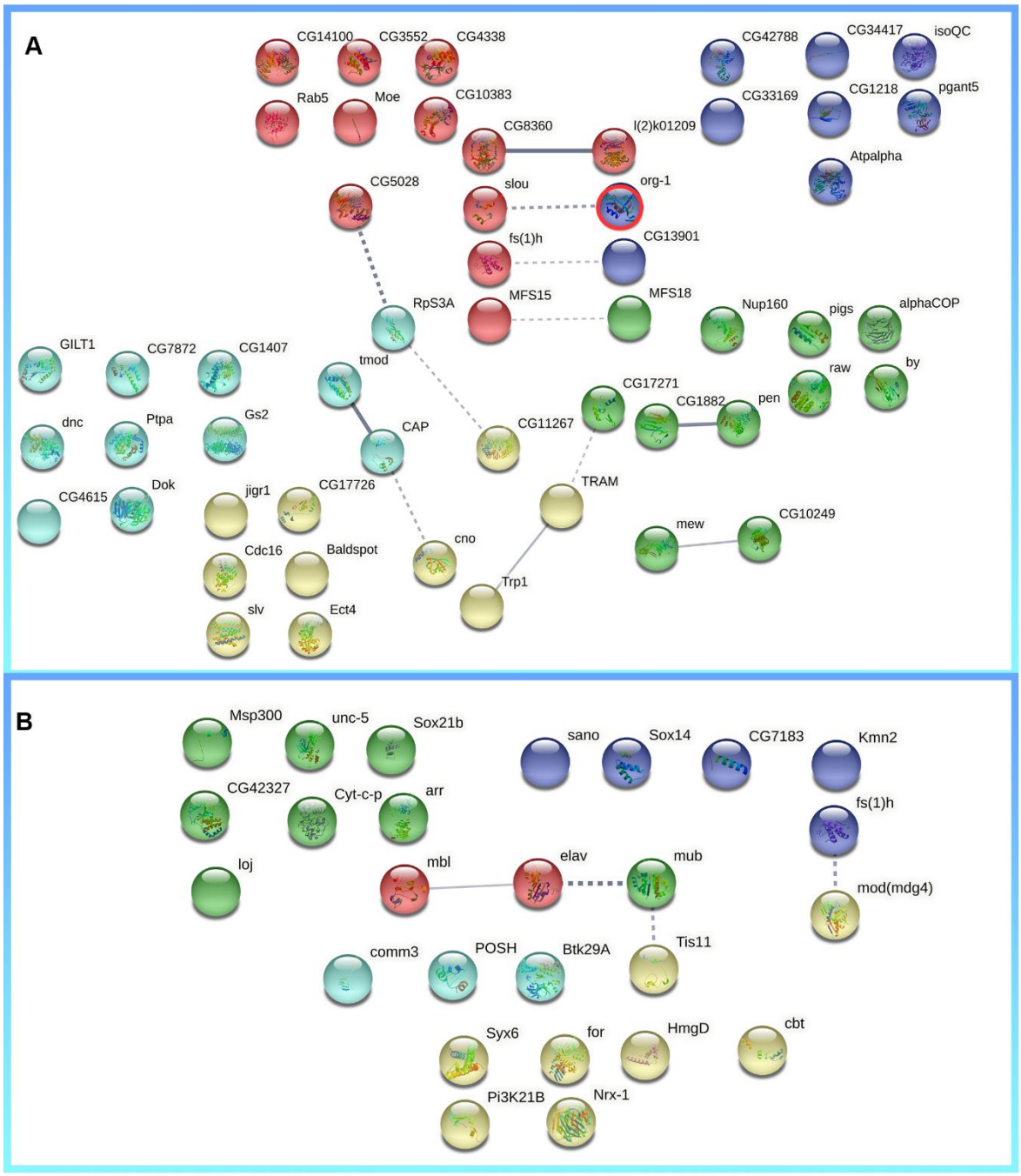


Figure 33. Protein/gene interactions between common genes that are significantly differentially expressed in the Slou subset with respect to the Lms subset and are putative Slou targets.
(A) There is nothing immediately striking about known protein interactions identified by the String database among putative Slou targets that are significantly upregulated during T3, except that the Slou+ muscle iTF gene *org-1* is a potential Slou target (circled in red). **(B)** The putative Slou targets that are significantly downregulated during T3 are extremely interesting as 26 of them shown in the figure are significantly upregulated in the Lms subset during T3.

them further. *Ssdp* is significantly upregulated in the Duf population between T2 and T3. It is also significantly upregulated in the Lms subset with respect to Duf as well as Slou during this time window. It is a highly conserved protein that is part of the ChiLS (Chip/LDB-Ssdp) complex that includes Chip (Chi), but its transcriptional targets and the processes it regulates are largely unknown (Fiedler et al. 2015). Various Ssdp interactors are significantly upregulated in the Lms subset at different timepoints (Figure 34A). *Chi* is upregulated during T1 in the Lms subset. It physically interacts with Ssdp and with multiple homeodomain TFs including Tailup (Tup) whose transcript is significantly upregulated during T3 in the Lms subset (Torigoi et al. 2000; Biryukova et Heitzler 2005).

Sbb is a transcriptional coregulator and has been implicated in photoreceptor axon target choice (Senti et al. 2000). It interacts with Serum Response Factor (SRF) that has been shown to regulate the expression of specific actin isoforms in specific adult flight muscle fibers (DeAguero et al. 2019) (Figure 34B). It exhibits a temporal transition profile where its expression goes up only in the Lms subset with respect to the Duf and Slou subsets. It physically interacts with various transcriptional coregulators such as Groucho (Gro), Mediator complex subunit 19 (MED19) and Grunge (Gug) and represses Hedgehog (Hh) expression in the wing disc (Bejarano et al. 2007).

Tis11 is significantly upregulated at T3 in the Lms subset and downregulated in the Slou subset. It is a member of the tristetraprolin family of proteins that bind AU-rich mRNA elements and regulate their stability (Yeh et al. 2012). Among its known targets is *eya* that is significantly upregulated only during T1 in the Lms subset with respect to the Slou subset, but not with respect to the Duf subset (Figure 34C).

Twinstar (Tsr) is the *Drosophila* cofilin, an actin binding protein, required for establishing planar cell polarity among other functions (Blair et al. 2006). The *tsr* gene is significantly upregulated with respect to Duf in the Lms subset during T3 and during T2 and T3 in the Slou subset. This might be significant in conjunction with the significant upregulation of *Gel* and *qua* during T3 that is Lms specific. Tsr genetically interacts with another ADF-H/Gelsolin-like superfamily member Twinfilin (Twf) (Figure 34D). It aggravates Twf homozygous phenotypes in the adult eye (Wahlström et al. 2001). Other members of the ADF-H/Gelsolin-like family such as Gel and Qua are significantly upregulated in the Lms subset. Qua is a Villin-like F-actin

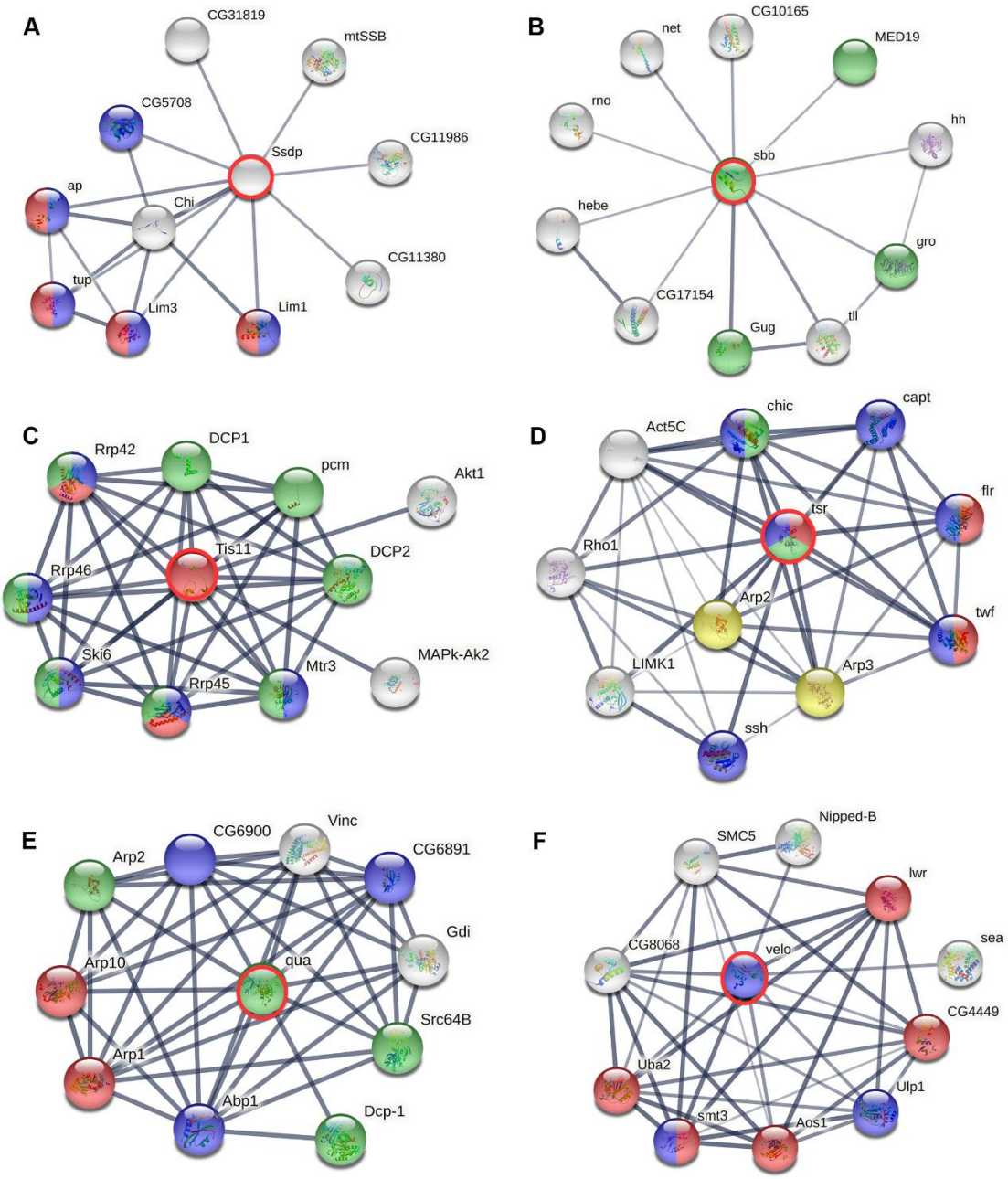


Figure 34. Protein/gene interaction analysis of individual candidate genes using the String database.

(A) Ssdp along with Chi is part of the evolutionarily conserved ChiLS complex. It interacts with multiple Homeodomain TFs (blue) in complex with Chi. A majority of known partners are LIM homeodomain proteins (red). **(B)** Sbb physically interacts with various transcriptional coregulators (green). **(C)** Tis11 interacts genetically with other AU-rich element binding proteins (red), U4 snRNA 3'-end processing proteins (blue) and proteins involved in RNA degradation (green). **(D)** Tsr genetically interacts with multiple actin binding proteins (blue), some of which are involved in actin polymerization (green) or depolymerization (red) as well as Arp2/3 protein complex components (yellow). **(E)** Qua and its homologues in vertebrates genetically interact with multiple proteins involved in actin filament organization (green), components of the dynactin complex (red) or Cofilin-tropomyosin-type actin-binding domain proteins (blue). **(F)** Velo is a SUMO protease that is involved in the post translational protein modification process called desumoylation by coordinating with other SUMO proteases (blue) as well as protein sumoylation factors (red).

crosslinking protein that has been implicated in filopodial dynamics by recruiting Daam1 (Jaiswal et al. 2013) (Figure 34E).

Velo is a SUMO protease involved in post translational protein modification and olfactory axon guidance that appears in the list of significantly upregulated genes specific to the Slou subset during T1, T2 and T3 as well as significantly downregulated genes in this subset during T2 (Figure 34F). While probes mapping to velo-RD are significantly upregulated during all timepoints in the Slou subset, velo-RC is significantly downregulated during T2. The trend is the exact opposite in the Lms subset. This gene was not analyzed at present, although it appears to be a very interesting candidate considering that its probes mapping to 2 different transcripts are differentially regulated in the Lms and Slou subsets.

3.1.3. New insights gained from the reanalysis of TRAP datasets

As mentioned previously, these datasets were initially analyzed when they were first generated (Bertin et al. 2015; Bertin 2017). The methods used during this analysis were different from mine. Data normalization was performed using the Solo package for microarrays (http://www-microarrays.u-strasbg.fr/Solo/index.html) followed by differential expression discovery using MS Excel. Following differential expression discovery, spatial profiling of differential gene expression between muscle subsets and temporal expression profiling of difference in gene expression between timepoints were performed based on the fold changes. All analyses were based on fold changes and performed on dm3 genes. This prior analysis did not compare data with other public RNA-Seq datasets or consolidated data from multiple published/curated information sources, nor did it focus on in-depth CRM analysis or signaling pathways. This analysis resulted in the identification of the cytoskeletal genes *Gelsolin* (*Gel*) and *lethal (2) essential for life* (*l(2)efl*) as being specific to the LT muscle subset (Bertin et al. 2021) and as potential downstream realisator genes and focused on them.

I reanalyzed these datasets using R and identified new candidate genes, TFs and signaling pathways potentially implicated in muscle diversification (summarized in Table 1). I mapped the dm3 probes to genes as well as specific transcript regions (exons, 3'UTR, etc.) in the latest dm6 genome assembly before analyzing them. My approach was to use an integrative analysis strategy by consolidating information gleaned from PCA, temporal expression profiling, comparison with public mRNA-Seq datasets and published/curated information, clustering and

GO analysis, CRM analysis as well as gene interaction analysis. I performed an initial analysis based on probe intensities instead of fold changes to identify differential patterns of gene expression among subsets by PCA and temporal expression signature analysis to identify gene clusters that move in the same direction without necessarily being differentially regulated. After identifying differentially expressed genes, I compared the Duf datasets to a publicly available mRNA-Seq dataset close to our timepoints to find the possible rate of correlation between transcription and translation. This revealed a 48% correlation of DufT3vsDufT2 with the closest mRNA-Seq timepoints. Comparing significantly differentially expressed genes in various populations with published/curated data populated into MySQL databases to query and integrate information from these different sources helped highlight specific genes annotated with known muscular and CNS functions as being differentially expressed in specific muscle subsets. It also helped identify common genes that were identified by multiple sources and their known expression patterns.

This integrative approach helped identify potential new iTFs and TF hierarchies along with potential downstream realisator genes and signaling pathways that could contribute to differences in muscle identity. Table 1 summarizes the significant results from this analysis along with the candidate genes that were finally subject to biological validation.

Table 1: Significant results from the bioinformatic analysis. Genes that were biologically analyzed are highlighted in red.

Method (tools used)	Identified component type	Component name(s)	Enriched in dataset	Enriched with respect to dataset	Timepoint(s)
PCA (R)	Spliceosome component	*LSm7*	Lms	Slou	Up T1, T2, T3
Temporal expression signature (Mfuzz)	KEGG signaling pathway	mTOR, FOXO, Hippo, G protein-coupled receptor signaling	Lms	-	Up between T1 to T2 to T3
	KEGG pathway	Inositol phosphate metabolism pathway	Slou	-	Up between T1 to T2 to T3
	Gene	*sbb*	Lms	-	Up between T2 and T3
Comparison with published/curated information (Python, MySQL, R)	Sarcomere development	*D*	Lms	Slou	Up at T1, T2, T3
	Mef2 target	*Sox14*	Lms	Slou	Up T3
		Mlc2, Tm2, Mp20	Slou	Lms	Up T1, T2
	Myoblast fusion gene	*siz*	Lms	Slou	Up T1, T3
		blow	Lms	Slou	Up T2
		Vrp1	Lms	Slou	Up T3
	Axon targeting	*Gel*	Lms	Slou	Up T2, T3
		Mur2B	Lms	Slou	Up T2
		fred, mgl	Slou	Lms	Up T3
	BDGP CNS gene	*Ssdp*	Lms	Slou	Up T3
	BDGP Muscle gene (early stages)	*Ssdp*	Lms	Slou	Up T3
	BDGP Muscle gene (early-late stages)	*Tis11*	Lms	Slou	Up T3
	Mef2 target	*Stat92E*	Slou	Lms	Up T1, T2, T3
Clustering and GO analysis (R)	KEGG pathway	Glycine, serine, threonine metabolism	Lms	Slou	Up T3

continued…

Method (tools used)	Identified component type	Component name(s)	Enriched in dataset	Enriched with respect to dataset	Timepoint(s)
	Biological process	Golgi transport	Slou	Lms	Up T3
	Non-coding RNA	*RNaseMRP:RNA, snRNA:U2:34ABb , snoRNA:Psi28S-2566*	Duf	Embryonic mRNA	Down T1, T2, T3
	Spliceosomal genes	*LSm3*	Duf	Embryonic mRNA	Up T2
		LSm4	Duf	Embryonic mRNA	Up T2, T3
	Canonical Wnt signaling genes	*kto*	Lms	Slou	Up T1, T2, T3
		Ssdp	Lms	Slou	Up T3
	G protein-coupled receptor signaling factors	*GABA-B-R2*	Lms	Slou	Up T2, T3
		Ggamma30A	Lms	Slou	Up T3
	Actin-binding genes	*tsr*	Lms Slou	Duf Duf	Up T3 Up T2, T3
		Svil	Lms Slou	Duf Duf	Up T1, T3 Up T2
		Gel	Lms	Slou	Up T2, T3
		qua	Lms	Slou	Up T3
		cher	Lms	Slou	Up T3
	Differential transcript expression	*velo:* *velo-RD* *velo-RC*	Lms Lms	Slou Slou	Down T1, T2, T3 Up T2
	Homeodomain TFs	*Hmx*	Lms	Slou	Up T3
		NK7.1	Lms	Slou	Up T3
	bHLH TFs	*emc*	Lms	Slou	Up T3
	RNA processing	*Tis11*	Lms	Slou	Up T3
CRM analysis (i-cisTarget, HOMER)	Nf-YB motif		Slou	Lms	Down T2
	CT-rich motifs		Lms	Slou	Up T3
	GATA motif		Slou	Lms	Up T3
	Trl motif		Lms	Slou	Up T3

continued …

Method (tools used)	Identified component type	Component name(s)	Enriched in dataset	Enriched with respect to dataset	Timepoint(s)
Analysis of putative TF targets (modERN, PAVIS)	Trl-putative targets	*Ssdp*	Lms	Slou	Up T3
	D putative targets	*D*	Lms	Slou	Up T1, T2, T3
	Slou putative targets	*Sox21b*	Lms	Slou	Up T2, T3
		Tis11	Lms	Slou	Up T3
	Stat92E putative targets	*Gel*	Lms	Slou	Up T2, T3
		Ssdp	Lms	Slou	Up T3
Protein/gene interaction analysis (String database)	Lms interactors	D	Lms	Slou	Up T1, T2, T3
		Sox21b	Lms	Slou	Up T2, T3
		Mid	Lms	Slou	Up T3
	RNA stability complexes	Tis11	Lms	Slou	Up T3

3.2. Biological analysis – functional analysis of candidate genes

Following the in-depth bioinformatic analysis of transcriptomics data and candidate selection, I proceeded to perform functional analyses of the chosen candidates to validate their roles in muscle identity.

3.2.1. Analyses of candidate gene mRNA and/or protein expression patterns

I first verified muscular expression of candidate genes. I performed an *in situ* hybridization to verify if candidate genes were expressed in somatic muscles, and if yes, if there was any evidence of muscle specific expression (Figure 35). *Ssdp* and *sbb* have high levels of expression in somatic muscles as well as the CNS from mid to late stages, although their expression is not restricted to these tissues (Figure 35A-B''). Their expression mirrors each other, except for *sbb* having much stronger expression in the brain (not shown) and in the muscles during late stages when muscles start separating from each other and *Ssdp* expression in somatic muscles declines. I will elaborate on the *Ssdp* expression pattern later (Section 3.2.3.2). *Tis11* displays weak expression in the somatic muscles (Figure 35C-C'').

Actin binding genes are of particular interest owing to their muscle specific roles. *tsr* has a strong muscle specific expression (Figure 35D-D''). *qua* displays a patterned expression in each segment. The thoracic segments have a widely distributed expression over the somatic muscles (not shown) while in the abdominal segments, in addition to expression in the muscles, *qua* transcripts are concentrated in cells that could be part of the peripheral nervous system (PNS) and/or the overlying epidermis. The bioinformatic analysis revealed that *qua* mRNA is significantly upregulated only during T3 in the LT muscles and not during T2 as observed in Figure 35E-E''. This could be due to the fact that *in situ* hybridization detects all transcripts of a given gene while TRAP detects only a subset of them associated with ribosomes, which are under translation.

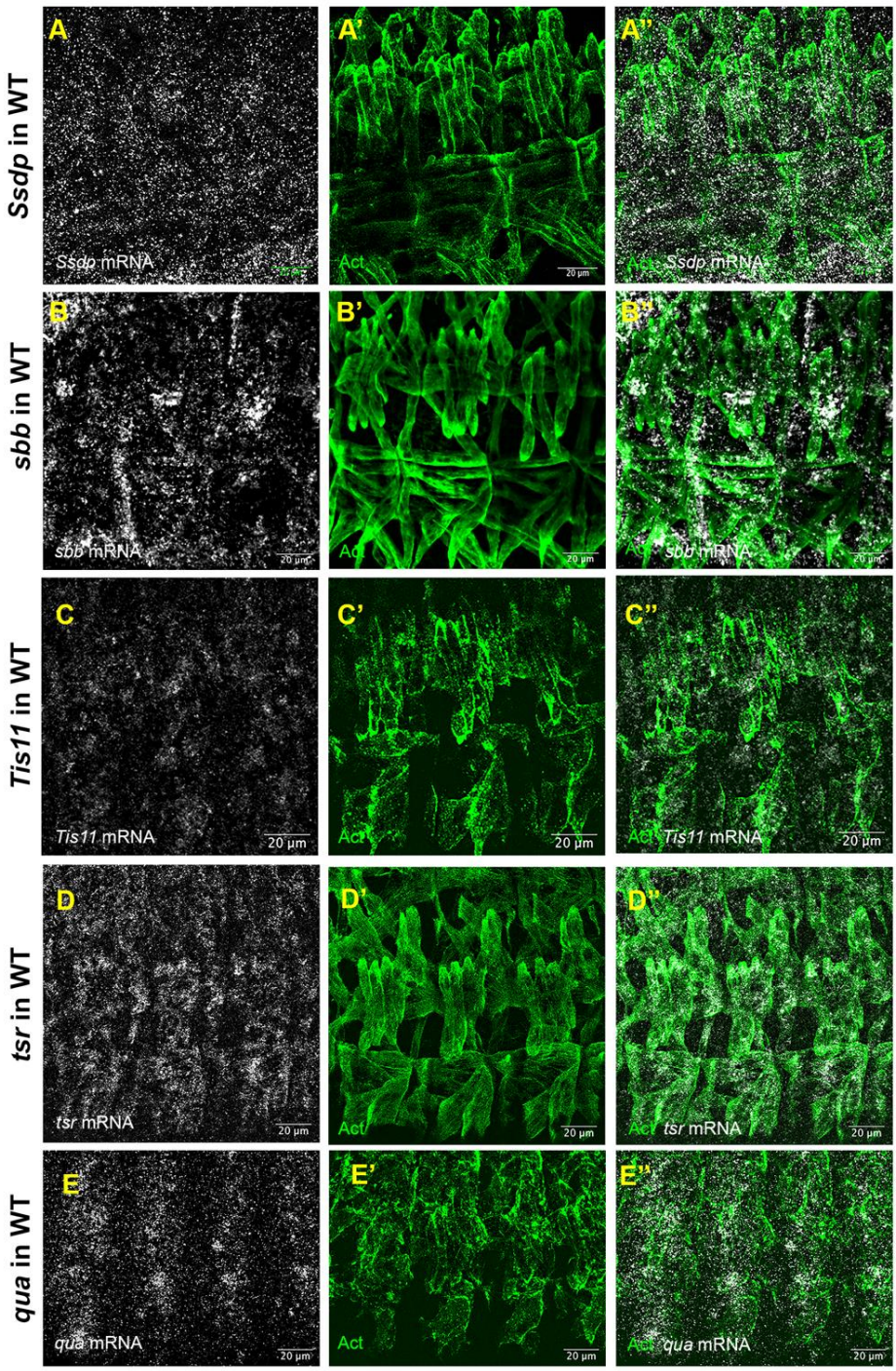


Figure 35. *in situ* hybridization of candidate genes selected after bioinformatic analysis of transcriptomics data.

(A-A'') *Ssdp* is strongly expressed in the somatic muscles at stage 16. **(B-B'')** *sbb* displays a strong expression pattern in the CNS and muscles similar to *Ssdp*, but its expression persists into very late stage 16 as shown here. It is also expressed in the propioceptory, sub-epidermal chordotonal organ. Only muscular expression is highlighted in both cases. **(C-C'')** *Tis11* displays a weaker, but clear expression in stage 16 somatic muscles. **(D-D'')** *tsr* is very strongly expressed in somatic muscles at stage 16. **(E-E'')** In the abdominal segments of stage 14 embryos, *qua* expression is detected in muscles and is also concentrated in a group of cells between LT2 and LT3 where they fork off at the location where the dorsal branch of the SNa defasciculates. These are presumably epidermal cells.

I also tested the expression patterns of GFP-tagged TFs generated by the modERN consortium using P(acman) BAC libraries (Kudron et al. 2018). Among the genes coding for homeodomain proteins that are significantly differentially expressed in the Lms subset during T3 are the bHLH TF, *extra macrochaetae* (*emc*) and the NK-like homebox TFs, *H6-like-homeobox* (*Hmx*) and *NK7.1* (Figure 36). In the case of Emc and Hmx, the GFP lines do not show any detectable muscle expression, but label CNS and PNS cells, respectively. This could be due to the partial expression of the TF-GFP BAC construct and does not exclude muscular expression of endogenous *emc* and *Hmx* genes. In contrast, NK7.1-GFP labels somatic muscles of the abdominal hemisegments with a strong dorsal and weaker ventral expression. The Ssdp partner Chip (Chi) is ubiquitously expressed as evidenced by immunohistochemistry with an anti-Chi antibody.

3.2.2. Phenotypic analysis of mutants for TFs and cofactors involved in LT muscle identity

Among the candidates, Ssdp seems to be the most intriguing given its interaction partners such as Chi and Ap and as yet unknown muscle function. To test its potential role in LT muscle identity, I analyzed the impact of the loss of function of *Ssdp* in detail along with its partner *Chi* and the genes encoding LT muscle iTFs Ap, Lms and Mid in order to analyze similarities and differences in mutant phenotypes. This was followed by an analysis of heterozygotes and transheterozygous mutant contexts in order to determine possible genetic interactions among these TFs and cofactors. I initially did a thorough visual inspection of phenotypes of all analyzed transgenic lines before proceeding to perform a quantitative, comparative study between homozygotes versus heterozygotes versus transheterozygotes since my goal was to identify potential genetic interactions among these iTFs and cofactors.

3.2.2.1. Phenotypic characterization of individual mutants during embryonic stages

lms^{S95} mutants have been previously characterized by our team along with ap^{UG035} mutants (Müller et al. 2010). I tested the effect of their loss of function on LT muscles in comparison with other iTFs and cofactors to get an idea about the similarities and differences between phenotypes.

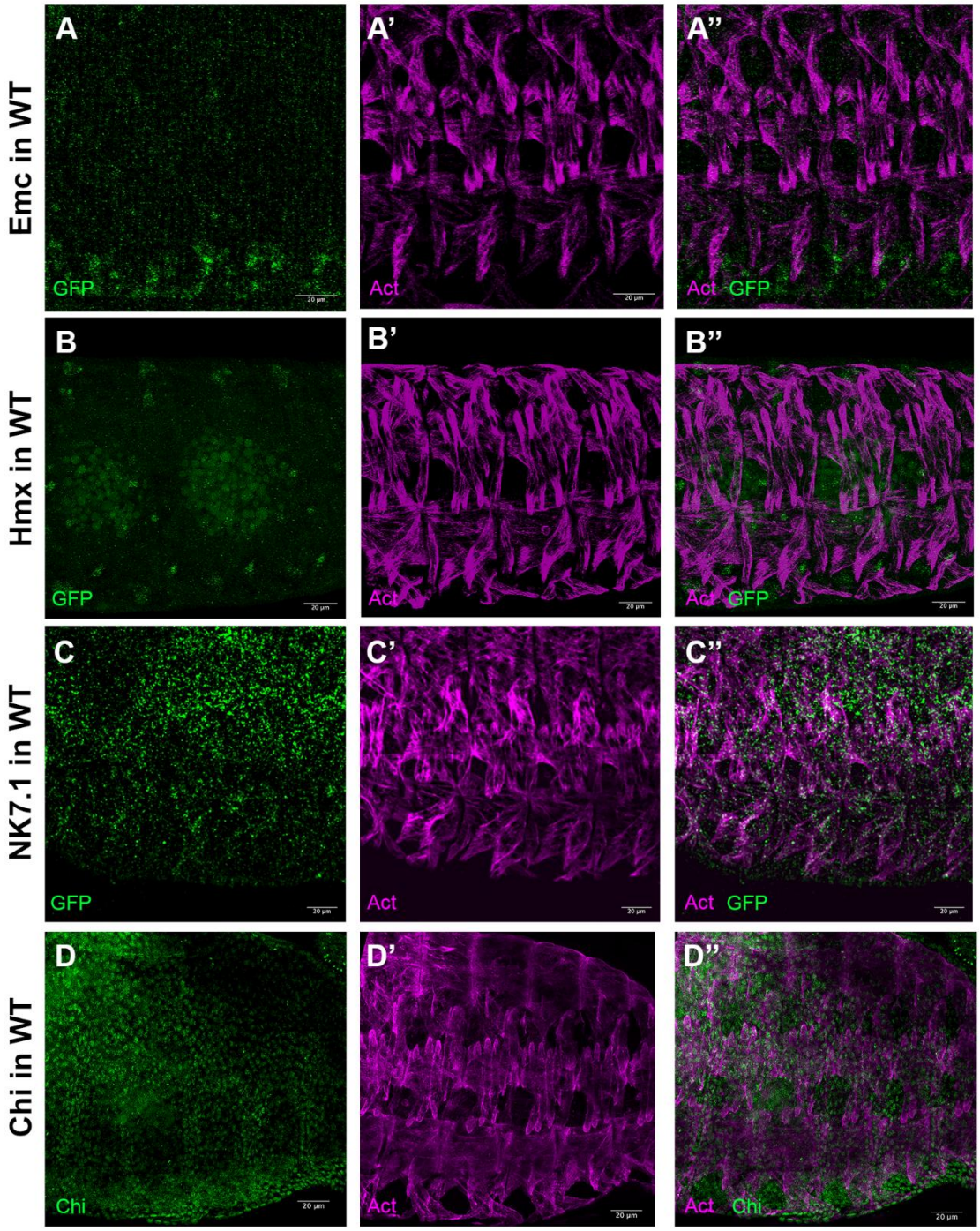


Figure 36. Analysis of protein expression patterns of candidates. Stage 16 embryos are shown. (A-A'', B-B'', C-C'') GFP-tagged proteins driven by TF enhancers used in the large-scale ChIP-Seq modERN project to unveil TF binding profiles for Emc (A-A''), Hmx (B-B'') and NK7.1 (C-C'') are displayed. The *emc* and *Hmx* lines do not display any expression in the somatic muscles. NK7.1 is expressed strongly in the dorsal abdominal somatic muscles and expression decreases dorso-ventrally. **(D-D'')** An antibody against Chi shows that it is ubiquitously expressed.

3.2.2.1.1. *Ssdp* mutants display severe defects in somatic muscles

A detailed characterization is presented in section 3.2.3.2. I present a summary here. $Ssdp^{L5}$ and $Ssdp^{L7}$ somatic mutants lacking zygotic Ssdp were analyzed. $Ssdp^{L7}$ is a deletion of exon 2 that contains the complete Ssdp protein coding sequence whereas $Ssdp^{L5}$ is a partial deletion of the exon, both generated by imprecise P-element excisions (van Meyel, Thomas, et Agulnick 2003). Both are, thus, null alleles. Homozygotes for both are lethal by the second instar larvae. Both mutants display similar muscle and innervation defects (Figure 37F shows a summary figure for a $Ssdp^{L5}$ embryo). All somatic muscles are affected, but not to the same degree. LT muscles and external ventral muscles such as the VA1/2 and VO4-6 are the most affected morphologically whereas VA3 is absent. 100% of embryos display these LT and ventral muscle morphology phenotypes. Muscles appear at their characteristic positions in each hemisegment, but several of them present defective morphological characteristics. Innervation defects are also observed in these mutants.

3.2.2.1.2. *Chi* mutants present duplications of LT muscles

I analyzed $Chi^{e5.5}$ mutants in order to test if they mimic *Ssdp* mutant phenotypes since they partner with each other as part of the ChiLS complex. $Chi^{e5.5}$ mutants are homozygous lethal and die by the third instar larvae (Morcillo, Rosen, et Dorsett 1996; Morcillo et al. 1997), permitting the analysis of embryonic muscles. The mutant was generated by an ethyl methanesulfonate (EMS) induced frame shift mutation resulting in a null allele. $Chi^{e5.5}$ mutants display much milder phenotypes compared to $Ssdp^{L7}$ mutants, with the dominant phenotype being LT muscle duplication detected in 9% of mutant embryos. Duplications of these muscles are extremely rare in WT embryos (at 1%) and difficult to detect in *Ssdp* mutants in which LTs always have an aberrant morphology making it difficult to assess the number of muscles. This phenotype is evident from mid to late embryonic stages in $Chi^{e5.5}$ mutants. In instances where duplications occur, the dorsally placed LT4 muscle is present, but there are one or more additional LT muscles anterior and more ventral to it. Other muscles are similar to the WT. So, this is an LT muscle specific defect. The dorsal branch of the SNa that innervates LT muscles defasciculates normally in these embryos (Figure 37B-B''). These milder phenotypes are in line with Bronstein et al.'s study where they analyzed genetic interactions by the overexpression of putative Ssdp targets in *Ssdp* and *Chi* mutants and found striking differences (Bronstein et al. 2010). Some putative targets enhance/suppress only *Ssdp* mutant phenotypes while some suppress only $Chie^{5.5}$ mutant phenotypes. In summary, $Chi^{e5.5}$ mutant phenotypes do not

reproduce the severe muscle defects observed in a *Ssdp* loss of function context suggesting a *Chi*-independent role of *Ssdp*.

3.2.2.1.3. *ap* mutants present severe LT muscle defects

ap is an orthologue of human *Lhx2*, both of which are expressed in the limb and expressing human *Lhx2* in flies is capable of activating Ap target gene expression in the wings (Rincón-Limas et al. 1999). ap^{UG035} mutants were generated by an imprecise P element excision leading to the deletion of ~6KB that removes the first *ap* protein coding exon (Cohen et al. 1992). The severity of ap^{UG035} mutant phenotypes is reminiscent of that of ap^{UG035} in trans with a complete deletion of the *ap* locus indicating that it is a null allele. Homozygous ap^{UG035} flies develop to the adult stage, although they are short lived. The posterior LT3 and LT4 muscles are missing in many hemisegments of these mutant embryos, with LT4 being highly affected. Interestingly, contrary to this observation, a study using a P44 5' end deletion *ap* mutant resulting from the imprecise excision of a P-element noted that the LT1/2 muscles were mostly missing (Bourgouin, Lundgren, et Thomas 1992). This raises the very interesting possibility that it could be due to the deletion of a LT1/2 specific enhancer. Muscles are reduced to blobs in 41.8% of hemisegments A2-A6 analyzed (n=55 hemisegments) (Figure 37C-C''). This is at 0% in the WT embryos analyzed. The dorsal branch of the SNa is affected in some hemisegments similar to $Ssdp^{L5}$ mutants. Muscle 'transformations' (a transformation of LT4 into LT3) are also observed. However, this term is mentioned in quotations since these could also be potential LT duplications with a missing LT4.

3.2.2.1.4. *lms* loss of function leads to affected posterior LTs

lms^{S95} is a null mutant generated by the imprecise excision of a P element deleting its entire coding sequence. Homozygous lms^{S95} individuals develop into viable adults. Similar to ap^{UG035} mutants, the posterior LT muscles are frequently missing in these mutants, with LT4 being the most affected. In severely affected embryos, muscles are not distinctively separated from each other unlike in WT embryos. The dorsal branch of the SNa motor neuron fails to defasciculate in these cases (Figure 37D-D'').

3.2.2.1.5. Loss of *mid* affects posterior LTs and ventral muscles

Midline (Mid) is homologous to human Tbx20 (R. P. Kumar et al. 2015). mid^{1} is a null allele resulting from a point mutation that introduces a stop codon in its coding sequence. It has been previously characterized (R. P. Kumar et al. 2015) and homozygotes are embryonic lethal

(Tripathy et al. 2014). Given the observation that these mutants lack the Ara and Kr expressing LT4 muscle founder cell (FC), Kumar et al. surmised that this either meant that the loss of Mid led to the loss of this FC or the loss of expression of the Ara and Kr iTFs in the FCs. These mutants frequently, but not always, lack the LT4 muscle. In a large number of hemisegments, LT4 to LT3 'transformations' are observed. In addition, I observe much more severe ventral muscle phenotypes including some missing muscles (Figure 37E).

3.2.2.2. Phenotypic characterization of transheterozygotes

I then looked at transheterozygote combinations of mutants in order to determine possible genetic interactions. Preliminary observations show that $Ssdp^{L5}$/+;ap^{UG035}/+ transheterozygotes exhibit WT phenotypes and they need to be analyzed further. On the contrary, ap^{UG035}/+;$Chi^{e5.5}$/+ (Figure 38A) and ap^{UG035}/+;mid^{1}/+ (Figure 38B) exhibit LT muscle duplications in a few hemisegments similar to $Chi^{e5.5}/Chi^{e5.5}$ where LT4 is present, but there are supernumerary LT muscles ventral to it. The rate of duplications ventral to the single LT4 muscle in individual ap^{UG035} and mid^{1} homozygotes, on the other hand, is insignificant.

3.2.2.3. Comparative analysis of mutant phenotypes reveals the role of Chi and Ap in the duplication of LT muscles

I performed a quantitative analysis to get a better understanding of these phenotypes. To this end, I analyzed hemisegments A2-A5 in lms^{S95} homozygotes and T1-T3 and A1-A6 in ap^{UG035}, $Chi^{e5.5}$ and mid^{1} homozygotes. T1 was analyzed only to test if there were T1 to T2 transformations. This is observed in 10% of mid^{1} homozygote embryos. No phenotype is 100% penetrant in homozygotes except for a full penetrance of the $Ssdp^{L5}$ phenotype. Missing muscles are the predominant phenotype in homozygotes for ap^{UG035}, lms^{S95} and mid^{1} as expected given their role as LT muscle iTFs. At least one LT muscle was missing in slightly more than 50% of the hemisegments analyzed in lms^{S95} homozygotes (Figure 38C). All four LT muscles are missing in around half of these cases (Figure 38C'). This percentage is at 40% in ap^{UG035} and 36% in mid^{1} homozygotes. In both cases there is a preferential loss of LT4, that is very evident in the case of ap^{UG035}, followed by LT3. This phenotype is evident from early stages in all three cases indicating impaired FC specification. The percentages for lms^{S95} and ap^{UG035} homozygote phenotypes closely correlate with those found by a prior study in our team (Müller et al. 2010).

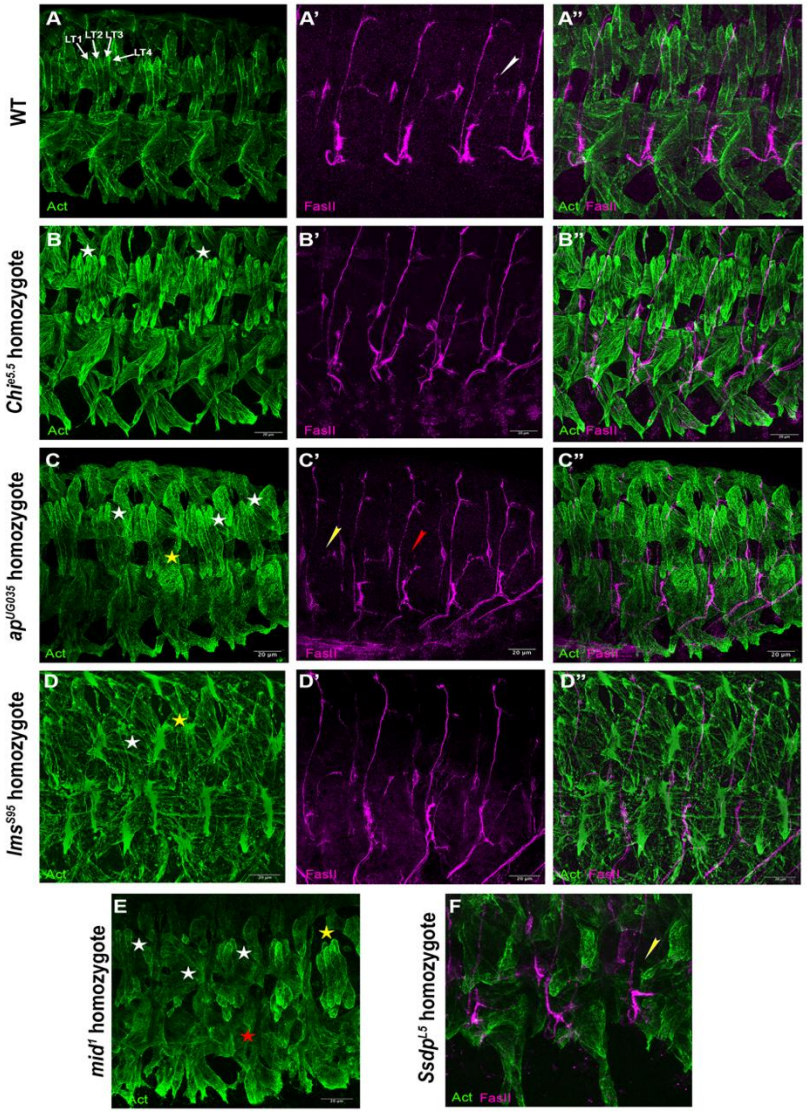


Figure 37. Characterization of iTF mutant phenotypes. Stage 16 embryos are shown.
(A-A'') A WT stage 16 embryo possesses 4 lateral transverse (LT) muscles in each hemisegment, LT1-LT4 indicated in (A). During this stage, they are innervated by the dorsal branch of the SNa motor neuron (arrowhead in A'). Innervation follows a segmental pattern with anterior segments being innervated first. **(B-B'')** Homozygous *Chi*$^{e5.5}$ mutant embryos exhibit LT muscle duplications in a few hemisegments, with up to 6 LT muscles in certain instances (white asterisks in B). The dorsal branch of SNa successfully defasciculates in all of these embryos. **(C-C'')** The posterior LT3 and LT4 muscles are missing in most hemisegments in homozygous *ap*^{UG035} mutants. LT4 is the most affected (white asterisks in C). Sometimes muscles are reduced to blobs (yellow asterisk in C). Innervation defects are evident in these mutants. The SNa either fails to defasciculate in multiple segments or there are multiple dorsal branches in some cases (arrowheads in C'). **(D-D'')** In *lms*^{S95} homozygotes, some LT muscles are missing (white asterisk in D) in a few hemisegments. The LT muscles that are usually well separated and distinct from each other appear fused together in some embryos (yellow asterisk in D). The dorsal branch of the SNa fails to defasciculate in these homozygotes (D'). **(E)** *mid*^{1} homozygotes exhibit phenotypes of missing muscles (white asterisks in E), most frequently LT4 and transformations of LT4 to LT3 (yellow asterisk in E). This phenotype is not 100% penetrant similar to the phenotypes for the other homozygote mutants analyzed. All the ventral muscles exhibit severe defects (red asterisk in E). **(F)** *Ssdp*^{L5} homozygotes exhibit the most severe somatic muscle defects. The external lateral and ventral muscles are the most affected morphologically. The dorsal branch of the SNa fails to defasciculate in certain segments (yellow arrowhead) with LT muscles that appear fused or are reduced to blobs.

I also analyzed mutant heterozygotes to test if there was a similar profile when one copy of the gene was lost. I used it as a negative control to ensure that phenotypes observed in transheterozygotes are indeed a result of genetic interaction and not due to the loss of one copy of a gene since the percentages observed are low (Figure 38D). In ap^{UG035} heterozygotes, there is no significant muscle duplication. There is a very low, probably insignificant, percentage of muscle loss similar to the percentage of muscle 'transformations' in these heterozygotes compared to homozygotes. $Chi^{e5.5}$ heterozygotes, on the other hand, present around half the percentage of muscle duplications seen in homozygotes. mid^{1} heterozygotes display a much lower percentage of missing muscles and no 'transformations'. On the other hand, they display a higher percentage of duplications compared to homozygotes. $Ssdp^{L5}$ heterozygotes display none of the phenotypes studied here.

The observation of a high number of missing muscles in individual lms^{S95}, ap^{UG035} and mid^{1} homozygotes that are evident even during early stages suggests that this is due to a lack of correct specification of FCs that generate these muscles (Figure 38C). The incomplete penetrance of this phenotype suggests that there are multiple factors at play that might cause enough modifications in protein stoichiometry to affect these phenotypes. The preference for LT4 muscle loss followed by LT3 in ap^{UG035} and mid^{1} homozygotes indicates that these iTFs play a significant role in establishing the identity of these muscles (Figure 38C'). A notable percentage of presumable LT4 to LT3 'transformations' that are observed in ap^{UG035} and much more significantly in mid^{1} homozygotes suggests that either the LT4 FC was mis-specified as an LT3 FC or that LT4 is missing and another LT muscle is duplicated. This duplication could be because they have a role in the control of the number of rounds of myoblast fusion that determines the final number of myonuclei in each muscle. An overexpression the FC protein, Duf that is implicated in myoblast fusion in muscles (Ruiz-Gómez et al. 2000) induces extra fusion events that could lead to muscle splitting and duplicated muscles (Bertin et al. 2021). The fact that the 'transformation' percentage remains almost the same in ap^{UG035} homozygotes and heterozygotes and an observation of this phenotype only during late embryonic stages (shown in Figure 37, normal early stages not shown) suggests a potential requirement for high levels of *ap* in the control of the number of rounds of myoblast fusion. It is interesting that the 'transformation' phenotype is absent in mid^{1} heterozygotes whereas there is an increase in the percentage of muscle duplication. These could be indicators of the consequences of minor stoichiometric disturbances.

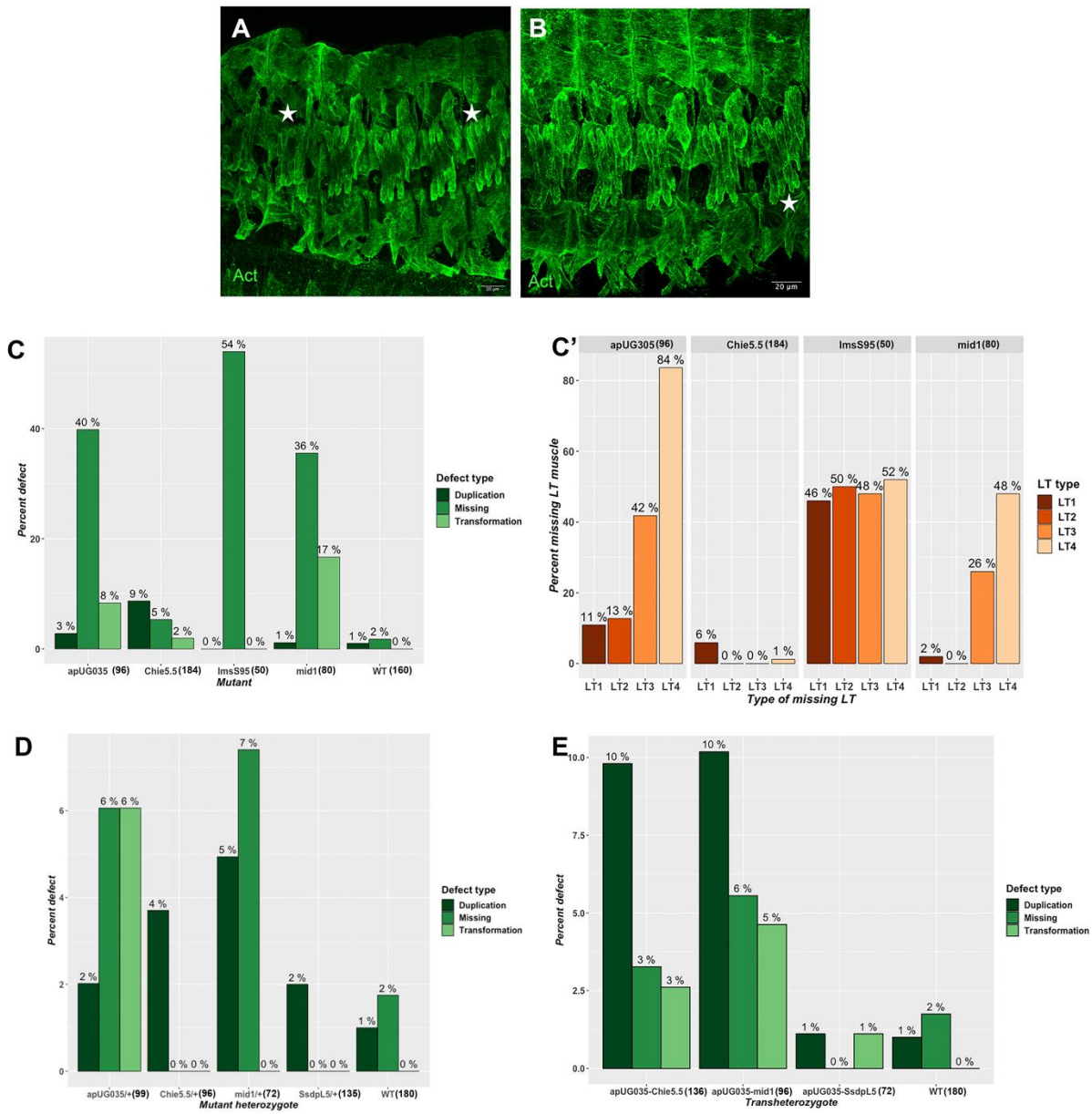


Figure 38. Analysis of homozygotes, heterozygotes and transheterozygotes for LT iTFs along with *Ssdp^{L5}* and *Chi$^{e5.5}$*.

(A) *ap^{UG035}*/+;*Chi$^{e5.5}$*/+ transheterozygotes exhibit LT muscle duplications similar to *Chi$^{e5.5}$* homozygotes (asterisks). **(B)** *ap^{UG035}*/+;*mid^{1}*/+ transheterozygotes display similar phenotypes (asterisk). **(C)** A quantification of the number of duplicated, missing and LT4 to LT3 transformed muscles in different homozygous mutant contexts is shown. Duplications are significantly higher than WT embryos in *Chi$^{e5.5}$* homozygotes. Missing muscles are prevalent in *lms^{S95}*, *ap^{UG035}* and *mid^{1}* homozygotes with mildly increasing severity of the phenotype: *lms^{S95}* > *ap^{UG035}* > *mid^{1}*. On the contrary, in *Chi$^{e5.5}$*/*Chi$^{e5.5}$*, there are supernumerary muscles instead of missing muscles. **(C')** A quantification of the type of LT muscle missing in each homozygote mutant condition reveals that among the missing muscles in *ap^{UG035}*/*ap^{UG035}*, 84% are LT4 and 42% LT3. A similar LT4 and LT3 preference is observed in *mid^{1}*/*mid^{1}*. In *lms^{S95}*/*lms^{S95}*, on the other hand, all LT muscles are equally affected. **(D)** An analysis of mutant heterozygotes that serve as a negative control for transheterozygote phenotypes reveals that somatic muscles in *Ssdp^{L5}*/+ embryos do not display these phenotypes. **(E)** The rate of duplication in *ap^{UG035}*/+;*Chi$^{e5.5}$*/+ and *ap^{UG035}*/+;*mid^{1}*/+ transheterozygotes is similar to that in *Chi$^{e5.5}$*/*Chi$^{e5.5}$* homozygotes, not observed in *ap^{UG035}*/+ heterozygotes, but is present in half the percentage of embryos in *Chi$^{e5.5}$*/+ and *mid^{1}*/+ heterozygotes suggesting that a genetic interaction between *Chi* or *mid* with *ap* aggravates the phenotype. ('n' represents the number of T1-A6 hemisegments analyzed).

Interestingly, in both ap^{UG035}/+;$Chi^{e5.5}$/+ and ap^{UG035}/+;mid^{1}/+ transheterozygotes, the LT4 muscle is present and the balance shifts towards a higher percentage of muscle duplications not observed in ap^{UG035} heterozygotes (Figure 38D, E). ap^{UG035}/+;mid^{1}/+ transheterozygotes, however, do present a percentage of missing muscles and 'transformations' higher than in the WT. The percentage of presumptive LT4 to LT3 'transformations' is comparable in ap^{UG035} homozygotes and heterozygotes as well as ap^{UG035}/+;mid^{1}/+ transheterozygotes whereas it is not observed in mid^{1}/+ heterozygotes. So, the loss of one copy of *ap* potentially influences this phenotype whereas the loss of one copy of *mid* does not, but the loss of both copies of *mid* aggravates the phenotype. A lack of availability of good antibodies that mark individual LT muscles makes it difficult to analyze this in detail. Since it cannot be conclusively stated that these cases represent the same type of defect, a definitive conclusion about these observations is deferred. The low percentage of defects potentially reflects phenotypes resulting from minor disturbances in protein stoichiometry and potentially reflect interaction instead of just background noise in comparison to WT embryos.

The percentage of duplication phenotypes observed for non-LT4 muscles is comparable in $Chi^{e5.5}$ homozygotes as well as ap^{UG035}/+;$Chi^{e5.5}$/+ and ap^{UG035}/+;mid^{1}/+ transheterozygotes and twice the percentage observed in mid^{1} heterozygotes. From these observations, it can be concluded that the loss of one copy of *mid* or *Chi* along with *ap* has a dominant effect on this phenotype that replicates the loss of both copies of *Chi* at late developmental stages whereas the loss of both copies of *ap* or both copies of *mid* leads to LT muscle loss during early stages. This suggests a requirement for Ap and Mid for correct FC specification during early stages when their absence leads to the absence of muscles as has already been demonstrated (Bourgouin, Lundgren, et Thomas 1992; R. P. Kumar et al. 2015) and the maintenance of their expression during late stages in order to ensure the correct number of fusion events and, thus, avoid supernumerary muscles. It can also be concluded that during late stages, Ap and Chi as well as Ap and Mid genetically interact to influence this event. The observation that ap^{UG035}/+;$Chi^{e5.5}$/+ leads to a similar LT duplication level as $Chi^{e5.5}$/$Chi^{e5.5}$ suggests a physical interaction between Chi and Ap rather than an influence of Chi on *ap* transcription, in which case one would expect a significant loss of LT muscles early during development, unless Chi levels affect only late *ap* transcription. The observation that ap^{UG035}/+;$Ssdp^{L5}$/+ do not display any phenotypes suggests that the ChiLS complex might not influence this phenotype and Chi potentially forms an alternate complex with Ap. It could also mean that one *Ssdp* copy provides

sufficient levels of protein. These transheterozygotes need to be reanalyzed to confirm a lack of phenotypes.

3.2.3. Ssdp – a new actor in muscle identity

3.2.3.1. Ssdp: the gene family, the protein structure, the ChiLS complex and functions

Sequence-specific single-stranded DNA-binding protein (*Ssdp*) is an evolutionarily conserved gene. *Drosophila* possesses a single *Ssdp* gene (van Meyel, Thomas, et Agulnick 2003). Multiple *Ssdp* homologues have been identified in humans including *SSBP2*, *SSBP3* and *SSBP4* (Figure 39A, B, C). Another gene, *SSBP1* codes for a mitochondrial single-stranded DNA-binding protein. Despite its name, it shares no similarity with SSBP2-4, but is orthologous to the *Drosophila mitochondrial single-stranded DNA-binding protein* (*mtSSB*) gene. This is evident from a sequence alignment and phylogenetic analysis (Figure 39B, C). *Ssdp* homologues have been identified across the animal kingdom in diverse species from chicken (Bayarsaihan, Soto, et Lukens 1998) to mouse (Chen et al. 2002) to zebrafish (Zhong et al. 2011). This gene family is of interest for the role of its members in human cancers since they are mis-regulated in multiple forms of cancer (J.-W. Liu et al. 2008; Poitras et al. 2008; Y. Wang et al. 2010). Human single-stranded DNA binding proteins and their orthologues are involved in a wide range of processes including mRNA metabolism, DNA repair (Li et al. 2009; Lawson et al. 2020; Ashton et al. 2016), skeletogenesis (Feldhahn et al. 2012), cancerogenesis (J.-W. Liu et al. 2008; Y. Wang et al. 2010) and differentiation (J. Liu et al. 2016). The functions of SSBP family members appear to be distinct. The N-terminal of Ssdp is highly conserved and contains a LUFS (LUG/LHS, Flo8, single-stranded DNA-binding protein) domain that was first identified in the LEUNIG protein in *Arabidopsis thaliana* (Conner et Liu 2000). This domain contains a Lis homology (LisH) motif, which is found in a large number of eukaryotic proteins (Kim et al. 2004) and is involved in microtubule dynamics (Emes et Ponting 2001), followed by the P-X-GFL-XX-WW-X-VFWD motif within the first 100 amino acids (Figure 39D). It was recently shown that this domain can lead to the formation of SSBP2 tetramers (H. Wang et al. 2019). Tetramers were also observed for *Drosophila* Ssdp (Fiedler et al. 2015; Renko et al. 2019) (Figure 39E, F).

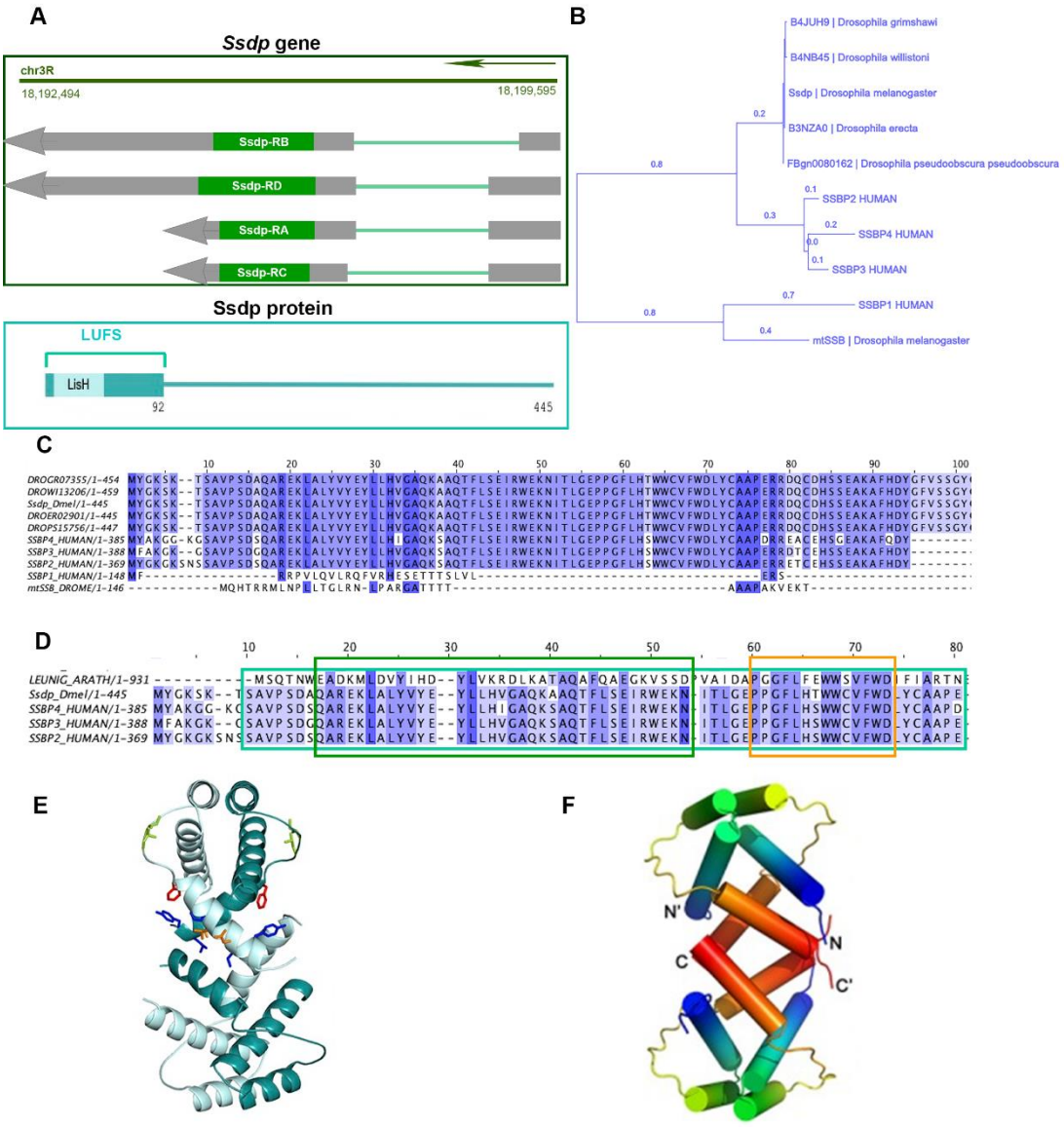


Figure 39. Ssdp protein structure and conservation.
(A) The *Drosophila melanogaster (Dmel) Ssdp* gene is located on chromosome 3R and is transcribed in the 3' to 5' direction. 4 isoforms have been identified to date with two longer isoforms containing longer 3' UTRs. The Ssdp protein contains an evolutionarily conserved LUG/LHS, Flo8, single-stranded DNA-binding protein (LUFS) domain at its N-terminal end. This domain harbors a LisH motif. **(B)** The Ssdp protein sequence is highly conserved among distantly related *Drosophila* species and shares high similarity with three members of the human *Single Stranded DNA-Binding Protein* (*SSBP*) gene family, SSBP2, SSBP3 and SSBP4. Another member, SSBP1 is mitochondrial and shares almost no similarity with these members, but shares high similarity with the *Drosophila* mitochondrial single stranded DNA-binding protein (mtSSB). **(C)** The conservation in the first 100 amino acids at the highly conserved N-terminal of Ssdp among distantly related *Drosophila* species and the human SSBP proteins is shown. It is immediately apparent that SSBP1 is an outlier. **(D)** Sequence alignment of Ssdp and human SSBPs with the LEUNIG protein in *Arabidopsis* where the LUFS domain (cyan rectangle) was first identified. The LisH motif found in many eukaryotes is highlighted by a dark green rectangle and the conserved P-X-GFL-XX-WW-X-VFWD motif C-terminal to it in orange. The extent of conservation becomes apparent. **(E)** Renko et al. (Renko et al. 2019) showed that the N-terminal of Ssdp is capable of forming stable tetramers in solution. Each cyan helix represents one Ssdp N-terminal. **(F)** Wang et al. (H. Wang et al. 2019) had earlier shown the crystal structure for this tetramerization for vertebrate SSBP2. Each SSBP2 subunit is represented in blue to red gradients.

Ssdp forms a complex with its conserved transcriptional cofactor Chip (Chi), known as LIM domain-binding protein 1 (LDB1) in humans (Agulnick et al. 1996; Becker et al. 2002; H. Wang et al. 2020). The Chip/LDB-Ssdp (ChiLS) complex is part of the Wnt enhanceosome (Figure 40F) that is known to translate Wnt signals into gene transcription along with Pygopus (Pygo) and Pangolin (Pan, also commonly known as dTCF) (Fiedler et al. 2015). Ssdp does not possess a nuclear localization sequence (NLS) domain. van Meyel et al. (van Meyel, Thomas, et Agulnick 2003) showed that Chi was required for the nuclear localization of Ssdp. Ssdp in turn protects Chi from proteasomal degradation (Güngör et al. 2007; Y. Wang et al. 2010; Bronstein et al. 2010). Chi possesses three highly conserved domains, the dimerization domain (DD) at the N-terminal, the LDB1/Chip conserved domain (LCCD) and a LIM interaction domain (LID) at its C-terminal (Figure 40A-C''). The LID can bind LIM homeodomain (LIM-HD) TFs or LIM only proteins.

Renko et al. detailed the structure of the Ssdp-Chi complex in *Drosophila* (Renko et al. 2019). They discovered that the LCCD domain of Chi alone or the LCCD-DD domains can complex with a Ssdp dimer and that in the presence of the LCCD domain, Ssdp binds this domain instead of itself. Their model for the Chi-Ssdp complex predicted a core Chi dimer that bound four Ssdp proteins (Figure 40D). The crystal structure of the interaction of human SSBP2 with the LCCD and DD domains of LDB1 was recently elucidated in a subsequent study by Wang et al. (H. Wang et al. 2020) (Figure 40E). This study found that SSBP2 residues 1-94 in the N-terminal bound LDB1 more strongly than the LUFS domain alone (residues 10-77). These SSBP2 residues bound dramatically more strongly to LDB1 DD-LCCD than to LDB1 LCCD alone. They also stated that this complex is formed by a core LDB1 dimer containing the DD domains flanked by LCCD domains. Each LDB1 monomer binds an SSBP2 dimer. This confirms previous observations by Renko et al. in *Drosophila*.

Introduction of excess mRNA of the LIM-HD TF Xlim1 along with human LDB1 or its *Xenopus* orthologue XLdb1 into *Xenopus* embryos led to a synergistic induction of partial secondary axes and ectopic muscle formation whereas the introduction of any one of them had no effect. Castro et al. (Castro et al. 2002) observed that introducing high levels of mouse or *Drosophila Ssdp* mRNA in *Xenopus* embryos along with low levels of mRNA of the *Xenopus* LIM homeodomain protein Xlim1 and *Xenopus* XLdb1 induced secondary body axis formation

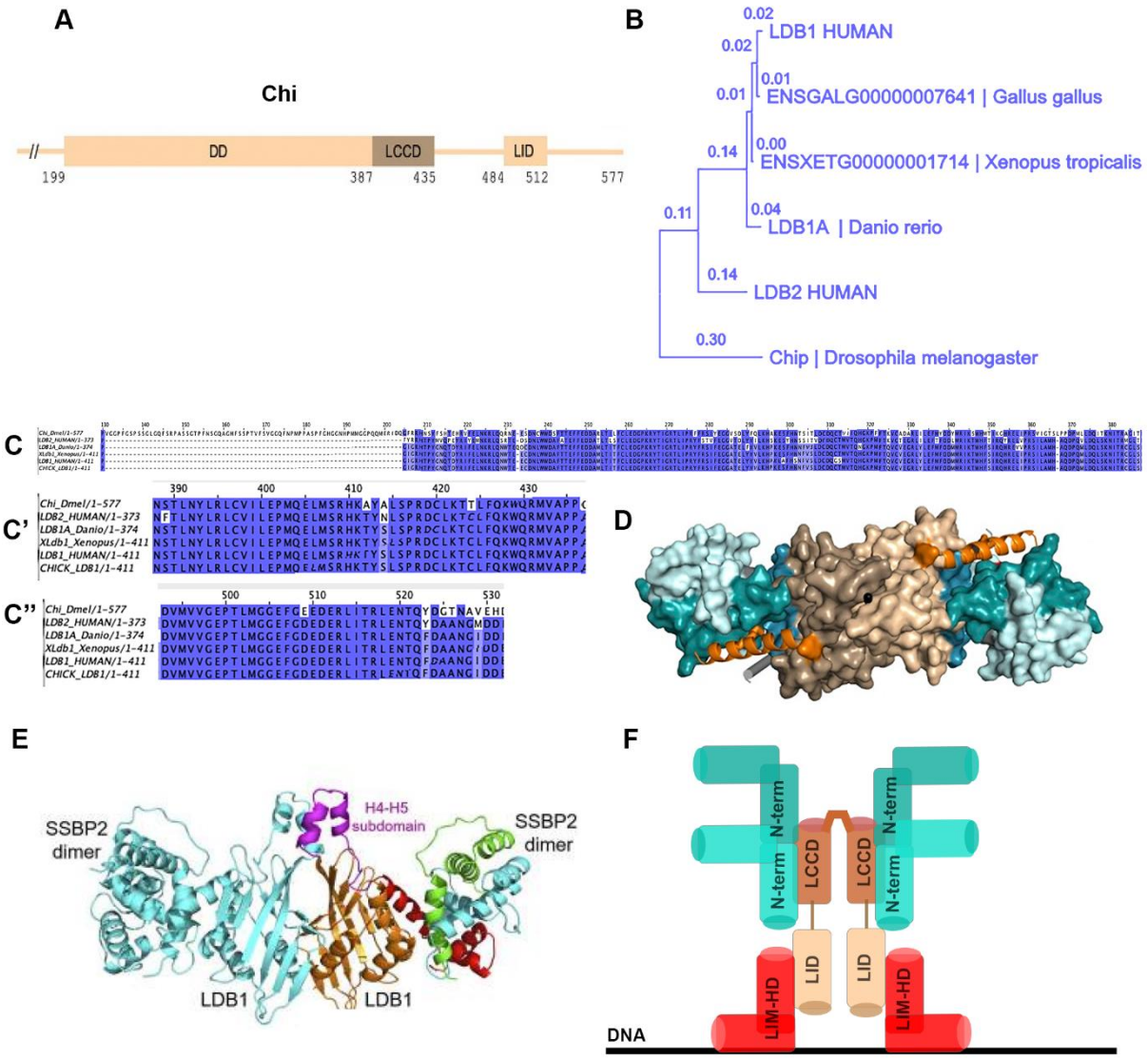


Figure 40. The structure and conservation of the Chi protein and the ChiLS (Chip/LDB-Ssdp) complex.

(A) Chi possesses three evolutionarily conserved domains at its N-terminal. The dimerization domain (DD) facilitates the formation of Chi-Chi dimers, the LDB1/Chip conserved domain (LCCD) facilitates interaction with Ssdp and the LIM interaction domain (LID) interacts with LIM domain proteins. **(B)** Chi shows evolutionary conservation with vertebrate species. Humans possess two paralogs of the orthologous LIM domain binding protein (LDB), LDB1 and LDB2 that display very high protein sequence similarity. **(C-C'')** The degree of conservation of the DD among species is shown in C. This alignment shows that *Drosophila* Chi has a slightly longer N-terminal end. C' displays a similar conservation image for the LCCD and C'' for the LID. **(D)** Renko et al. (Renko et al. 2019) predicted the structure of the Chi-Ssdp complex from their analysis with a core Chi dimer, each Chi monomer binding a Ssdp dimer. Each Chi is displayed in shades of wheatish brown and each Ssdp in shades of cyan. **(E)** Wang et al. (H. Wang et al. 2020) subsequently confirmed the correctness of this prediction in vertebrates with the crystal structure of the LDB1/SSBP2 complex. SSBP2 and LDB1 subunits are labeled. **(F)** A schematic representation of the ChiLS-LIM-HD (Chip/LDB-Ssdp-LIM-Homeodomain) complex. The ChiLS complex interacts with LIM-HD TFs such as Ap to regulate gene transcription. A Chi dimer (shades of brown) binds to two Ssdp dimers (shades of cyan) via the LCCD domains and to a LIM-HD protein dimer (red) via the LID domains.

whereas low levels had no effect. SSDP1 and its cofactors are implicated in neural patterning and differentiation of specific axonal projections in zebrafish embryos (Zhong et al. 2011). These studies indicate that LDB1 and its partners participate in patterning. Correct axis formation is central to the normal positioning and development of organs and is dependent on morphogen gradients such as Wg and Dpp in *Drosophila* (Lecuit et Cohen 1997; Strigini et Cohen 1997). They first direct primary axis formation by patterning the anterio-posterior (A/P) and dorso-ventral (D/V) axes. Subsequently, secondary axes for tissues such as vertebrate limbs or insect legs and wings are established thanks to their effectors. The LT muscle iTF Ap, for example, patterns the wing disc during larval stages to generate a fate map for the future adult wing. Wg restricts Ap expression at the D/V boundary (Williams, Paddock, et Carroll 1993). Ap complexes with Chi to regulate wing disc patterning (Fernández-Fúnez et al. 1998). LIM-only proteins act as competitors to LIM-HD proteins for LID binding (Agulnick et al. 1996). The LIM-only protein Beadex (Bx) acts as an Ap antagonist and interferes with its binding to Chi (Milán, Diaz-Benjumea, et Cohen 1998). These observations along with the finding that ChiLS is part of the Wnt enhanceosome indicate that the *Drosophila* embryonic muscle iTFs are potential effectors of segmental somatic muscle patterning in response to morphogen gradients.

Bronstein et al. (Bronstein et al. 2010) performed a study of Ssdp interactors by comparing the expression of genes in WT versus *Ssdp* mutant embryos with the help of a high throughput microarray analysis. Their observations reveal that Ssdp-Chi dynamics are much more complex in that among the Ssdp potential targets identified in this study, some enhance or suppress $Chi^{e5.5}$ phenotypes only whereas others enhance or suppress *Ssdp* mutant phenotypes only and a minority interact with both. For example, the gene *M(2)21AB* enhances the *Ssdp* mutant phenotype and suppresses the *Chi* mutant phenotype. This indicates that they are possibly not the only interaction partners of each other. Chi physically interacts with multiple homeodomain TFs including Tailup (Tup) whose transcript is upregulated during T3 in the Lms subset (Torigoi et al. 2000; Biryukova et Heitzler 2005). An analysis of *tup* mosaics induced by various clones in the wing disc showed that Tup genetically interacts with LT iTFs such as Ara/Caup and Dr (de Navascués et Modolell 2007) by modifying their expression patterns.

Bronstein et al. identified 189 putative targets of Ssdp in the wing disc. I cross-verified the list of genes identified from this study with genes significantly differentially expressed during T3

in the Lms subset, during which time point *Ssdp* is significantly upregulated with respect to the Slou subset, and did not find a significant representation of putative targets identified in the wing disc. I found an enrichment for identified putative targets in the Lms cluster generated by temporal expression profiling that contained *Ssdp*. There are 35 genes in common or 18.5%. However, 24 of these are upregulated in *Ssdp* mutants indicating a negative regulation in the wing disc while they belong to a cluster of genes whose expression consistently increases in the Lms subset. This cluster is enriched for CCs related to mitochondria. The mitochondrial *mtSSB* has strong membership to a cluster whose expression is downregulated between T2 and T3, contrary to *Ssdp*. These observations highlight the complexity of Ssdp interactions and the necessity for a fine-tuned, detailed analysis to better understand its roles.

3.2.3.2. Publication preprint: Ssdp influences Wg expression and embryonic somatic muscle identity in *Drosophila melanogaster*.

The preprint of the following article is now available on bioRxiv:

Poovathumkadavil, Preethi, Jean-Philippe Da Ponte, et Krzysztof Jagla. « Ssdp influences Wg expression and embryonic somatic muscle identity in Drosophila melanogaster ». bioRxiv, 8 June 2021. https://doi.org/10.1101/2021.06.08.447509.

Ssdp influences Wg expression and embryonic somatic muscle identity in *Drosophila melanogaster*

Preethi Poovathumkadavil, Jean-Philippe Da Ponte and Krzysztof Jagla*

Institute of Genetics Reproduction and Development, iGReD, INSERM U1103, CNRS UMR6293, University of Clermont Auvergne, 28 Place Henri Dunant, 63000 Clermont-Ferrand, France

* Correspondence: christophe.jagla@uca.fr

Abstract:

The somatic muscles of the *Drosophila* embryo and larvae share structural and functional similarities with vertebrate skeletal muscles and serve as a powerful model for studying muscle development. Here we show that the evolutionarily conserved Ssdp protein is required for the correct patterning of somatic muscles. Ssdp is part of the conserved Chi/LDB-Ssdp (ChiLS) complex that is a core component of the conserved Wg/Wnt enhanceosome, which responds to Wg signals to regulate gene transcription. *Ssdp* shows isoform specific expression in developing somatic muscles and its loss of function leads to an aberrant somatic muscle pattern due to a deregulated muscle identity program. *Ssdp* mutant embryos fail to maintain adequate expression levels of muscle identity transcription factors and this results in aberrant muscle morphology, innervation, attachment and fusion. We also show that the epidermal expression of Wg is downregulated in *Ssdp* mutants and that Ssdp interacts with Wg to regulate the properties of a subset of ventral muscles. Thus, our data unveil the dual contribution of Ssdp contribution to muscle diversification by regulating the expression of muscle-intrinsic identity genes and by interacting with the extrinsic factor, Wg. The knowledge gained here about Ssdp and its interaction with Wg could be relevant to vertebrate muscle development.

Keywords: Sequence-specific single-stranded DNA-binding protein (Ssdp); muscle identity transcription factors (iTFs); Wingless (Wg)

1. Introduction

Muscle development is a finely orchestrated process in vertebrates as well as invertebrates involving intrinsic myogenic factors and various signaling molecules that are transduced by downstream effectors into specific gene transcription. An imbalance in any of the proteins involved in this muscle developmental symphony can result in a compensatory mechanism by other players (Mankoo et al. 2003; Relaix et al. 2005; Rudnicki et al. 1993; Kumar et al. 2015) or in case of key factors, might trigger a cascade of deregulation resulting in the disruption of the developmental process (Borello et al. 1999; Lee et Frasch 2000). Vertebrates possess gene families that are often represented by a single orthologue in Drosophila (Potthoff et Olson 2007). This correlates with observations of less severe phenotypes in vertebrates when one

family member is mutated while mutations in the single Drosophila orthologue lead to drastic phenotypes. Patterning by morphogens plays an important role during development. The central dogma of patterning states that morphogen gradients hold positional information that leads to compartmentalization into domains, which then triggers identity acquisition in each domain by the expression of 'selector' genes that finally leads to cross-tissue communication and the establishment of new gradients (Lawrence et Struhl 1996). This holds true for vertebrates and invertebrates. In mammals, the evolutionarily conserved morphogen family, Wnt is one of the principal conductors of the developmental processes. It comprises 19 members while Drosophila has 7 Wnt homologues including Wingless (Wg). During vertebrate myogenesis, Wnt family members perform non-redundant functions (Münsterberg et al. 1995; Tajbakhsh et al. 1998; Sweetman et al. 2008).

During embryonic skeletal muscle myogenesis in vertebrates, Wnt signaling is among the factors directing the formation of periodically generated somites that form trunk and limb muscles. Soon after formation, ectodermal cues pattern each somite into domains including the high Wnt dermomyotome domain that gives rise to skeletal muscles (Ikeya et Takada 1998; Wagner et al. 2000). Post-mitotic myogenic $Pax3^{+}$ progenitors then delaminate from the dermomyotome to form the myotome by initiating the expression of muscle differentiation genes. In mice, muscle progenitors also express M-Twist (Füchtbauer 1995). These muscle progenitors express specific transcription factors (TFs) depending on their position within the dermomyotome by receiving specific Wnt cues from adjoining tissues. Wnt1 activates MYF5 to form the epaxial myotome that gives rise to back muscles whereas Wnt7a activates MYOD to form the hypaxial myotome that gives rise to muscles of the limb by migrating to limb buds as well as muscles of the ventral body wall, diaphragm and tongue (Tajbakhsh et al. 1998). The initial primary myotome differentiates and elongates anterio-posteriorly in a Wnt11 dependent fashion (Gros, Serralbo, et Marcelle 2009), then undergoes primary myogenesis by fusing with embryonic myoblasts to form primary fibers by acquiring a specific muscle identity that determines its morphology, innervation, attachment and function. The factors that define the identity of individual muscles that distinguish them from their neighbors are yet to be determined in vertebrates.

Drosophila somatic or body wall muscles are similar to vertebrate skeletal muscles since they are syncytial, striated and voluntary. *Drosophila* has 30 larval somatic muscles per hemisegment arranged in a stereotypical pattern generated during embryonic myogenesis. *Drosophila* embryos undergo simultaneous and synchronous segmentation of the germ layers giving rise to parasegments in a Wingless (Wg) dependent fashion (Bejsovec et Martinez Arias 1991). They are subsequently divided into domains including the high Wg (epidermal cue), high Twist (Twi) mesodermal domain that generates larval somatic muscles (Lee et
Frasch 2000). Muscle progenitors are subsequently generated that then divide asymmetrically to give rise to founder cells (FCs), each FC containing the information

for one specific muscle. Each FC expresses a muscle identity transcription factor (iTF) code that dictates its identity. Ectodermal Wg cues are implicated in the specification of some Slou^{+} ventral muscle progenitors (Cox, Beckett, et Baylies 2005). It is not known if Wnt signals play a role in regulating gene expression in specific muscles at later stages.

Although a broad requirement for Wnt signaling at various stages of muscle development has been identified, various factors involved in transducing this signal and translating it into gene transcription at each timepoint as well as those regulating its own expression remain elusive. *Sequence-specific single-stranded DNA-binding protein* (*Ssdp*) is an evolutionarily conserved gene with homologues across the animal kingdom (Bayarsaihan, Soto, et Lukens 1998; Castro et al. 2002; Chen et al. 2002; van Meyel, Thomas, et Agulnick 2003). It is part of the conserved Chip/LDB-Ssdp (ChiLS) complex along with the transcriptional cofactor known as Chip (Chi) in *Drosophila* and LIM domain-binding protein 1 (LDB1) in humans (H. Wang et al. 2019, 2020). LDB1 is required for the nuclear localization of Ssdp (van Meyel, Thomas, et Agulnick 2003) and binds LIM homeodomain TFs with high affinity while Ssdp protects LDB1 from proteasomal degradation (Güngör et al. 2007; Y. Wang et al. 2010; Bronstein et al. 2010). ChiLS has recently been shown to constitute a core component of the Wnt enhanceosome that translates the conserved canonical Wnt signaling mediated by ß-catenin into gene transcription (Fiedler et al. 2015; Renko et al. 2019). Humans possess 3 *Ssdp* homologues (*SSBP2*, *SSBP3* and *SSBP4*) while *Drosophila* has only one. This gene family is of interest for the role of its members in human cancers since they are mis-regulated in multiple forms of cancer (Liu et al. 2008; Poitras et al. 2008; Y. Wang et al. 2010), as is the case for Wnt signaling components (Delgado-Deida, Alula, et Theiss 2020).

The role of Ssdp during myogenesis has not yet been investigated, neither have its spatial and temporal roles. Here, we identify *Ssdp* as a significantly differentially regulated gene using a muscle-subset-specific Translating Ribosome Affinity Purification (TRAP) approach and aim to dissect the consequences of its loss of function on *Drosophila* embryonic muscle development. We show that *Ssdp* expression is enriched in somatic muscles during mid to late stages of muscle development. Its loss of function affects the levels of expression of muscle iTFs and influences the acquisition of muscle identity with most severe defects in the ventral and lateral muscles. Considering the known role of Ssdp in Wg-dependent gene regulation and the partial overlap of *Ssdp* and *wg* loss of function muscle phenotypes, we propose that in addition to its intrinsic role in maintaining iTF expression, Ssdp also ensures late Wg function in muscles.

2. Materials and Methods

Drosophila strains

All stocks except the temperature sensitive wg^{ts} were grown at a temperature of 25°C. The following stocks were gifted to us: $Ssdp^{L5}$ and $Ssdp^{L7}$ (gifts from Donald J van Meyel, McGill Centre for Research in Neuroscience, Montreal, Canada (van Meyel, Thomas, et Agulnick 2003)), S59-Gal4 (gift from Manfred Frasch, Erlangen, Germany). The following strains were obtained from Bloomington *Drosophila* Stock Center: wg^{ts} is a heat sensitive amorphic allele (wg[I-12]bw[1]/Cyo) that was rebalanced on a CyO(Act-GFP) balancer to distinguish homozygotes,
24B-Gal4 (w[*]; P{w[+mW.hs]=GawB}how[24B]),
w[1118]; PBac{w[+mC]=IT.GAL4}Ssdp[2082-G4]/TM6B Tb[1],
lms-Gal4 (w[1118]; P{y[+t7.7] w[+mC]=GMR88F08-GAL4}attP2),
UAS-LAGFP (y[1] w[*]; P{y[+t*] w[+mC]=UAS-Lifeact-GFP}VIE-260B) and
UAS-RpL10a-GFP (w[*]; P{w[+mC]=UAS-GFP-RpL10Ab}BF2b).
The following stock was obtained from Kyoto Stock Center: w[*]; P{UAS-Act5C.T:GFP}10-2

We used the following genotypes: 24B>Gal4;UAS>$dTCF^{DN}$, S59>Gal4/MKRS;UAS>RpL10a-GFP and lms>Gal4; UAS>LAGFP.

Temperature shift experiments

Temperature shift experiments were conducted as follows: To determine the role of Wg during early stages of muscle development, wg^{ts} flies were grown at a permissive temperature of 18°C for 12 hours (mid stage 11) after which the apple juice plates with embryos were shifted to a non-permissive temperature of 28°C for 9 more hours before fixing them. To determine the role of Wg at slightly later stages, embryos were staged by letting flies lay eggs for 3 hours at 18°C after which the apple juice plates were collected and let to develop at 18°C for 14 more hours and subsequently shifted to 28°C for 5 hours.

Immunofluorescent staining

The following primary antibodies were used: rat anti-Actin (1/500, MAC 237; Babraham Bioscience Technologies), rabbit anti-β3 Tubulin (1:5000; R. Renkawitz-Pohl, Philipps University, Marburg, Germany), rat anti-Tropomyosin (1:200, ab50567, Abcam), mouse anti-FasII (1:500, 1D4, Developmental Studies Hybridoma Bank (DHSB)), anti-GFP (1:1000, DHSB), mouse anti-βPS integrin (1:200 DSHB), mouse anti-Col (1:50, from Alain Vincent, Center for Integrative Biology, Toulouse, France), rabbit anti-Mef2 (1:500, from Eileen Furlong, EMBL, Germany), rabbit anti-Slou (1:300, from Manfred Frasch, Erlangen, Germany) and mouse anti-Wg (1:500, 4D4, DHSB). Fluorescence conjugated secondary antibodies from Jackson ImmunoResearch produced in donkey conjugated with Alexa 488, Cy3 or Cy5 were used at a concentration of 1:300. For immunostainings using primary antibodies

produced in rat and mouse, minimal cross secondary antibodies against both species were used.

RNA FISH and in situ hybridization

For RNA FISH experiments (Raj et al. 2008), 29 Quasar 570 conjugated Stellaris probes from LGC Biosearch Technologies (Orjalo, Johansson, et Ruth 2011) targeting the long isoforms of Ssdp were used. Fixed embryos were used to hybridize the probes using the standard Stellaris hybridization procedure. This was followed by antibody staining against actin to visualize all muscles and GFP to visualize muscle subsets.

For *in situ* hybridization, we used the following primer pairs:
Ssdp: 5′-TGTACGAATATCTGCTGCACG-3′ and
5′-GCATCGTCGAGTTAGGGAAG-3′
Probe generation and *in situ* hybridization were performed following the standard procedure (Legendre et al. 2013). Reverse primers with a T7 promoter sequence prefix were used. PCR fragments were amplified using the genomic DNA as template, purified and reverse transcribed using the Roche SP6/T7 transcription kit with Dig-UTP to generate Digoxygenin labelled mRNA probes. Fixed embryos were then hybridized *in situ* with the probes followed by TSA amplification and antibody staining against actin.

Image acquisition, processing and statistical analysis of images

All images were acquired on a Leica SP8 microscope using a 40X objective at a resolution of 1024x1024 or 2048x2048. They were analyzed and processed using ImageJ. Statistical tests and graph generation were performed in R. For CTCF quantification of fluorescence intensities of Col and Slou in WT versus *Ssdp* mutants, 25 stacks were acquired for each embryo analyzed. An equal number of images at 1024x1024 and 2048x2048 were included in each group to be compared against. ROIs were manually selected on maximum projections of each image and the mean of three areas close to each ROI was used as the background fluorescence. To quantify the number of Mef2+ nuclei, nuclei were manually counted in each DT1 muscle in abdominal hemisegments A2-A5 for each embryo. Similarly, the number of Eve+ pericardial cells were manually counted in hemisegments A2-A5 of each embryo analyzed.

Statistical analysis of transcriptomic data and cis regulatory motif analysis

Differential gene expression of the transcriptomic microarray data was determined in R using the limma package. Enrichment for cis regulatory motifs (CRMs) was determined using i-cisTarget (Herrmann et al. 2012; Imrichová et al. 2015). A normalized enrichment score (NES) threshold of 3 was used. Graphs were generated using R.

3. Results

3.1. Ssdp mRNA under translation is differentially expressed in muscle subsets

We used TRAP data generated earlier for two somatic muscle subsets expressing distinct iTFs, one expressing Slouch (Slou/S59) (S. Knirr, Azpiazu, et Frasch 1999) and the other expressing Lateral muscles scarcer (Lms) (Müller et al. 2010; Bertin et al. 2015, 2021) as well for the global muscle population expressing Duf. We also generated transcriptomic data for the entire embryo. Data analysis revealed that *Ssdp* was among the genes that were significantly upregulated in the Lms subset during late stages (13-16 hours after egg laying or AEL that we refer to as time window 3 or T3 here) (Figure 1A). We also observed a significant enrichment for CT-rich as well as complementary GA-rich cis regulatory motifs among the genes significantly upregulated during this time window (Figure 1B). Chicken SSDP was initially identified as a protein capable of binding CT-rich tracts in the α2(I) collagen gene. These tracts were subsequently shown to be capable of binding the fly Ssdp protein (Bayarsaihan, Soto, et Lukens 1998; Bronstein et al. 2010). The expression of *Ssdp* mRNA under translation mapping to all transcripts showed an upward trend in both muscle subsets with respect to the global embryonic mRNA. When compared with translating mRNA in the global muscle population, it is significantly upregulated in the Lms subset with respect to the Slou subset (Figure 1C-E).

Since Ssdp is an evolutionarily conserved protein, we examined its alignment with the canonical isoforms of human SSBP proteins. Apart from the already recognized conservation in the LUFS domain, we observed three other smaller blocks with highly conserved amino acid residues (Supplementary figure 1). One of the conserved regions is part of a proline-rich region whose deletion led to a headless phenotype in mice (Enkhmandakh, Makeyev, et Bayarsaihan 2006) due to the loss of the fore and midbrain. No function, if any, has been attributed to the other conserved regions yet.

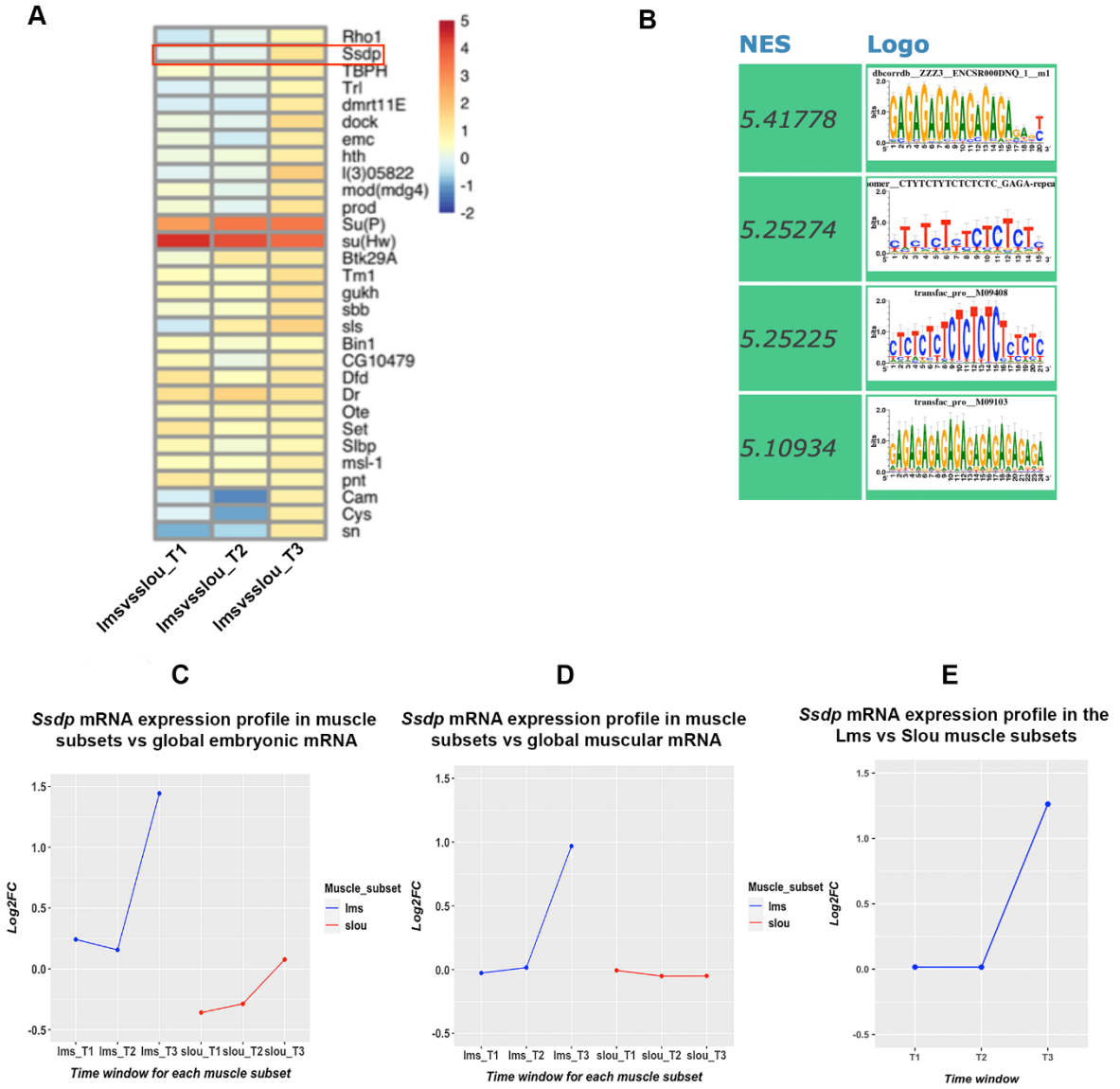


Figure 1.

***Ssdp* mRNA undergoing translation display significant differential expression in muscle subsets at late embryonic stages.**

(A) *Ssdp* is among the genes that are significantly upregulated in the Lms muscle subset during time window T3 that represents late embryonic stages. **(B)** An i-cisTarget analysis of cis regulatory regions among genes significantly upregulated in the Lms subset versus the Slou subset during T3 reveals a significant enrichment for CT-rich and complementary GA-rich motifs (*NES = Normalized Enrichment Score calculated by i-cisTarget above a threshold value of 3*). **(C-E)** The expression profile of *Ssdp* mRNA undergoing translation derived from TRAP data shows that it displays an upward trend in the Lms as well as Slou muscle subsets with respect to the global embryonic mRNA (C) and is significantly upregulated in the Lms population with respect to the global Duf+ muscle population (D). This is confirmed by a comparison of the muscle subsets amongst themselves (E).

3.2. Ssdp mutants display severe somatic muscle phenotypes

Given the differential expression observed in our TRAP datasets, we were interested to see if this gene played specific roles in individual muscles during muscle diversification. To this end, we analyzed $Ssdp^{L5}$ and $Ssdp^{L7}$, both considered to be *Ssdp* null mutants. $Ssdp^{L7}$ is a deletion of exon 2 that contains the complete Ssdp protein coding sequence while $Ssdp^{L5}$ is a partial deletion of this exon (van Meyel, Thomas, et Agulnick 2003). We observe that both mutants display similar severe muscle phenotypes as well as concomitant muscle innervation defects (Figure 2).

In the late stage WT embryo, somatic muscles are arranged in a stereotypical pattern. In both $Ssdp^{L5}$ and $Ssdp^{L7}$ mutants, however, although muscles are generally present, individual muscle fibers appear compacted and/or of aberrant morphology (Figure 2A, B, C). This is particularly obvious in lateral and ventral regions. Aberrations in the innervation of lateral and ventral muscles by their specific motor neurons are also apparent. The dorsal branch of the SNa motor neuron that innervates the WT lateral transverse (LT) muscles is missing in some segments in stage 16 $Ssdp^{L5}$ and $Ssdp^{L7}$ mutants (Figure 2A', B', C'). Similarly, the SNb and SNc branches normally innervating ventral muscles are severely affected with irregular morphologies and non-uniform innervation patterns in different hemisegments while the ISN branch targeting dorsal muscles shows only minor trajectory defects. The percentage of hemisegments where the dorsal branch of the SNa fails to defasciculate is slightly more pronounced in $Ssdp^{L7}$ mutants compared to $Ssdp^{L5}$ mutants (Figure 2D).

These observations reveal that Ssdp is required for proper patterning and innervation of somatic muscles with a major impact on ventral and lateral muscles. Our subsequent analyses were performed on $Ssdp^{L5}$ flies since both mutants display similar phenotypes and this line is easier to amplify. Henceforth, we will refer to the homozygous $Ssdp^{L5}$ embryos as '*Ssdp* mutants' and explicitly refer to $Ssdp^{L7}$ where applicable.

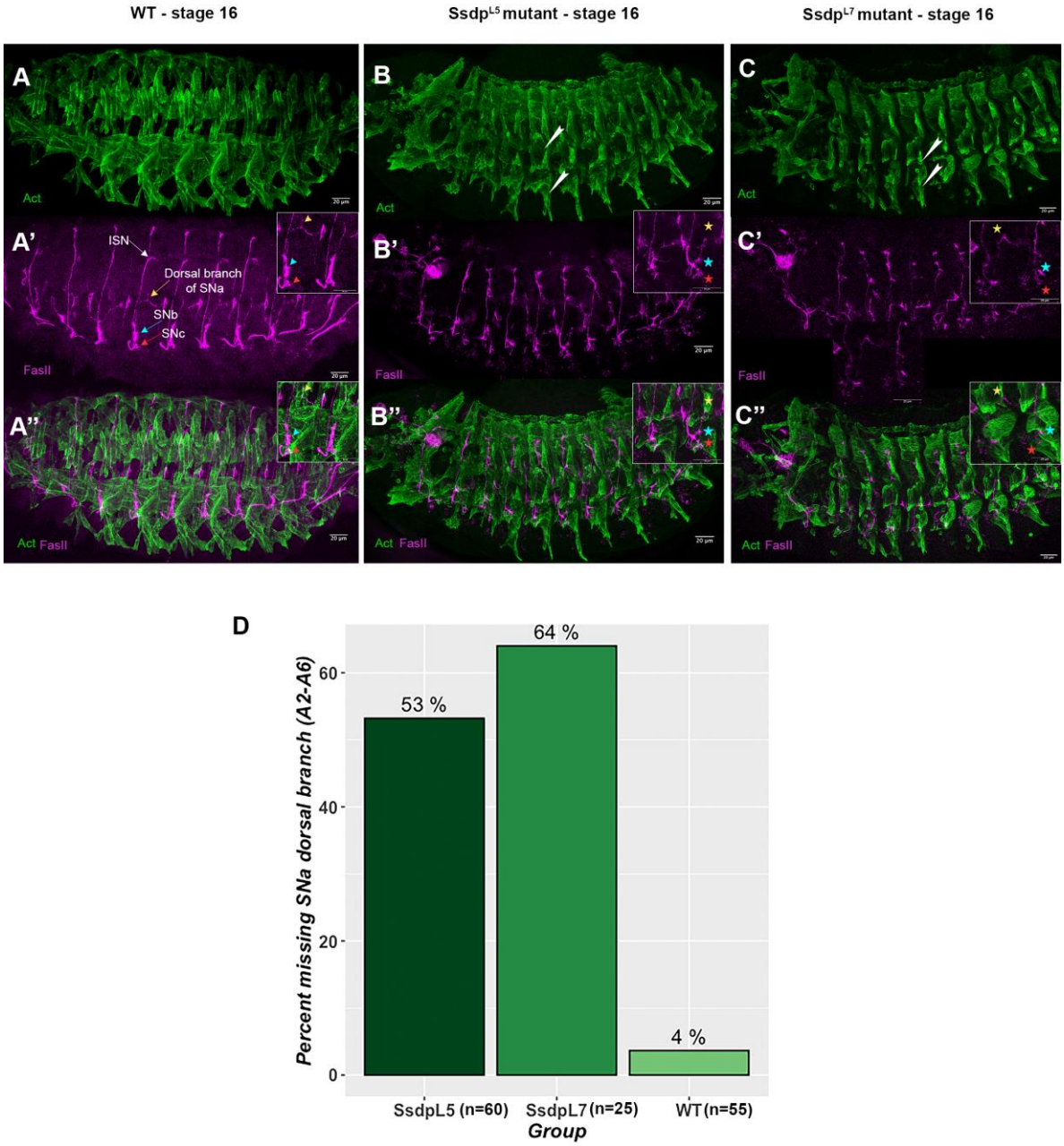


Figure 2.

Loss of function of *Ssdp* leads to severe defects in the somatic muscles.

(A-A'') The muscle (A) and innervation (A') pattern in WT stage 16 embryos. Different motor neurons are indicated by different colored arrows. Insets in A' and A'' show the stereotypical innervation and muscle patterns in each hemisegment. **(B-B'')** Both the muscle pattern (arrowheads in B) and innervation (asterisks in the inset in B') are severely defective in *Ssdp^{L5}* mutants, which presumably have a partial deletion in the *Ssdp* gene that contains a single protein coding exon. The lateral and ventral muscles are severely affected as indicated by arrowheads in B. The insets in B' and B'' display zoomed views highlighting innervation defects such as morphologically defective SNb and SNc motor neurons and a missing dorsal branch of the SNa. Compare asterisks to similarly colored arrowheads in the WT inset in A'. **(C-C'')** *Ssdp^{L7}* mutants lacking the entire length of the *Ssdp* gene display phenotypes resembling *Ssdp^{L5}* mutants with muscle morphology (arrowheads in C) and innervation (asterisks in the inset in C', C'') defects. **(D)** The percentage of hemisegments (A2-A6) missing the dorsal branch of the SNa is slightly higher in *Ssdp^{L7}* mutants compared to *Ssdp^{L5}* mutants, although it is significantly higher than in WT embryos in both mutants.

3.3. Ssdp mRNA is expressed in somatic muscles

In order to analyze embryonic expression patterns of *Ssdp*, we used *in situ* hybridization to reveal its transcripts and a transgenic *Ssdp* enhancer trap line in which GAL4 expression is driven by *Ssdp* regulatory sequences, thus giving an indication of the tissues in which Ssdp is expressed and its expression levels. We first performed RNA FISH (Raj et al. 2008) with probes that map to the 3' UTR of the two long *Ssdp* isoforms (Figure 3). *Ssdp* transcripts for the long isoforms cannot be detected in the somatic mesoderm during the specification of muscle precursors/founders (stage 11-12). However, later in development (stage 15), *Ssdp* mRNAs accumulate in all somatic muscles. In stage 15 and early stage 16 embryos, *Ssdp* transcripts are uniformly distributed in muscles as well as the ventral nerve chord (VNC) (Figure 3B-D'). Thus, using RNA FISH, we do not detect a particular enrichment of *Ssdp* transcripts in any muscle subset. However, this analysis represents whole muscular mRNA as opposed to the TRAP transcriptomic data that aimed to discover mRNA under translation. Also, the RNA FISH probes map to the two long *Ssdp* isoforms with long 3'UTR regions, and could thus present only a partial picture of the *Ssdp* expression pattern.

To determine the expression pattern of all isoforms, we performed a classic *in situ* hybridization with a *Ssdp* probe targeting a region of the protein coding exon of *Ssdp* present in all transcripts. In addition to the pattern described above with *Ssdp* expression in the VNC and developing muscles, we also detect *Ssdp* transcripts in epidermal and mesodermal layers at stage 12 (Figure 3E-H''). This also reveals epidermal expression during early to late stages followed by a segmentally repeated expression in ventral epidermal stripes during late stage 16 that is not observed for the long isoforms. Thus, in contrast to the TRAP profiles, neither the probes targeting long *Ssdp* isoforms nor those targeting all its isoforms detect an enrichment in particular muscle subsets. This suggests that the TRAP dataset could reveal a muscle-subset specific regulation of *Ssdp* at the level of mRNA undergoing translation. Due to the unavailability of a Ssdp antibody, we are unable to confirm this possibility. A Ssdp-Gal4 driven expression of GFP tagged actin5C reveals a muscular expression pattern that increases over time with marked GFP accumulation at stage 15 in Lms-positive LT muscles and Slou-positive DT1 (Supplementary Figure 2). This developmentally regulated expression suggests stage and isoform specific roles for Ssdp in somatic muscles.

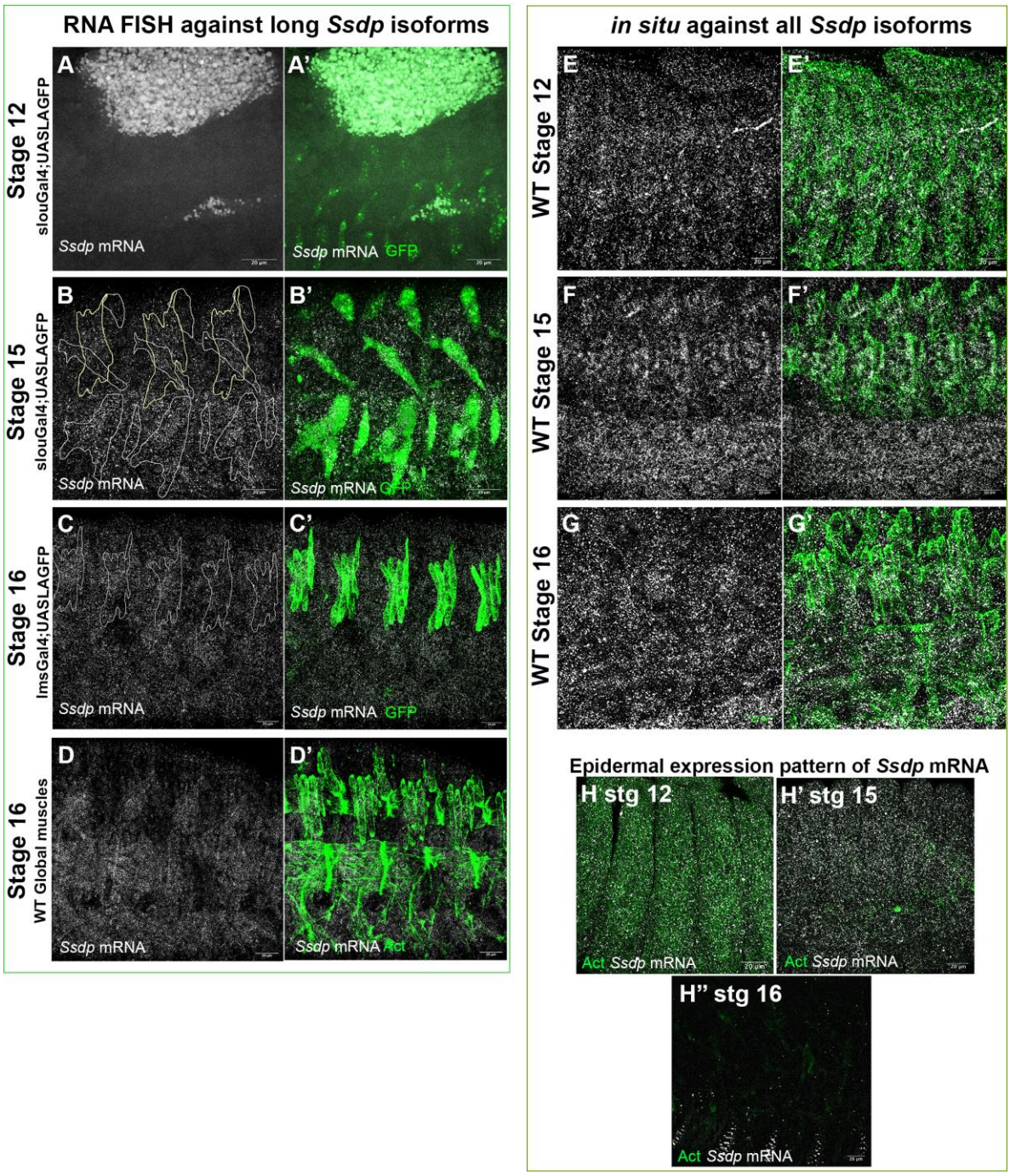


Figure 3.

RNA FISH against the long *Ssdp* isoforms reveals muscular expression at late stages whereas an *in situ* hybridization targeting all isoforms also reveals expression at early and very late stages.

(A-A') No remarkable transcript expression is detected during stage 12. The Slou+ muscle subset is visualized by an anti-GFP antibody as revealed by the Slou-Gal4 driven expression of RpL10aGFP. **(B-B')** Stage 15 embryos show high muscular expression levels of *Ssdp* mRNA. The Slou and Lms muscle subsets are outlined in white and yellow respectively in (B) as revealed by an antibody against actin (not shown). **(C-C')** *Ssdp* expression persists in stage 16 embryos. The Lms+ lateral transverse (LT) muscles are outlined in (C) as revealed by LifeActGFP driven by the lms-Gal4 driver. **(D-D')** *Ssdp* expression in global somatic muscles in stage 16 embryos as revealed by an antibody against actin. **(E-E')** *in situ* hybridization targeting all *Ssdp* isoforms reveals mesodermal expression at stage 12 that is absent for long isoforms. **(F-G')** The somatic muscles and ventral nerve chord (VNC) express high levels of *Ssdp* at stage 15 (F-F') and stage 16 (G-G'). **(H-H'')** *in situ* reveals epidermal expression that is not present for long isoforms. While this expression is throughout the epidermis during stages 12-16,

by very late stage 16, expression becomes restricted to a characteristic segmentally repeated pattern of anterio-ventral epidermal stripes.

3.4. Cytoskeletal muscle components are disorganized in the absence of Ssdp

Actin is a key muscle protein that is an integral component of sarcomeres, the contractile units of the muscle, apart from playing its generic role in the actin cytoskeleton (A. F. Huxley et Niedergerke 1954; H. Huxley et Hanson 1954). Actin dynamics during developmental stages 12-15 are involved in the formation of the fusogenic synapse that permits fusion of fusion competent myoblasts (FCMs) with the myotube (Sens et al. 2010; J. H. Kim et al. 2015). During muscle attachment, it plays a crucial role in extending filopodia to sense correct attachment sites (Schnorrer et Dickson 2004; Richier et al. 2018). During muscle innervation, muscles extend myopodia to communicate and connect with the right presynaptic filopodia (Ritzenthaler, Suzuki, et Chiba 2000). Thus, any disruptions to actin and its cofactors are potentially detrimental to muscle development. We examined the expression pattern of actin as well as its binding partner and muscle differentiation marker Tropomyosin (Tm). Both proteins are expressed in *Ssdp* mutants. Actin presents a highly organized arrangement in WT stage 16 embryos. Notably, it is cortically enriched outlining lateral transverse (LT) muscle shapes. This actin distribution is lost in *Ssdp* mutants in which actin-stained individual LT muscles are difficult to distinguish (Figure 4A, B). Similarly, the muscle differentiation marker Tm2 that is implicated in myotube elongation by co-localizing with F-actin (Williams et al. 2015) and enriched in LT muscle termini in WT embryos, displays a fuzzy, irregular pattern in *Ssdp* mutant LTs that fail to fully elongate (Figure 4C, D).

Another essential cytoskeletal muscle component is the microtubule (MT) network. MT associated proteins such as dynein play a crucial role in determining muscle length and myonuclear positioning in the LT muscles (Folker, Schulman, et Baylies 2012). WT LT muscles extend longitudinally in both directions with ß3-Tubulin accumulating at LT extremities. In *Ssdp* mutants, ß3-Tubulin enrichment at LT ends cannot be detected and LT muscles extend over much shorter distances compared to the WT with muscle ends bending towards each other (Figure 4E, F). We observe that the LT1 muscle extends over around half the distance compared to the WT (Figure 4G).

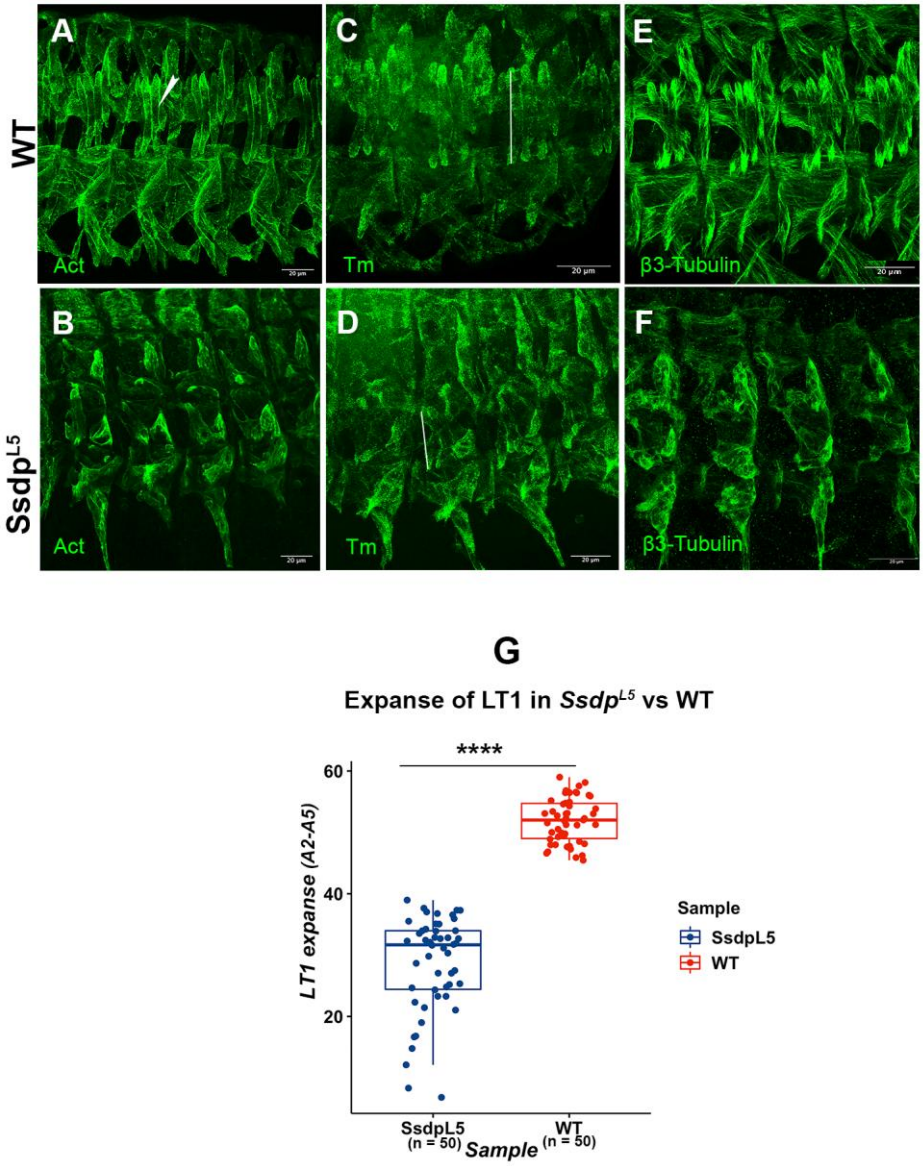


Figure 4.

***Ssdp* mutants exhibit severe disorganization of the actin cytoskeleton and microtubules at stage 16.** **(A-B)** In WT embryos (A), a staining for actin using an anti-actin antibody reveals a structured organization of the actin cytoskeleton whereas this organization is severely disrupted in *Ssdp* mutants with a disorganized concentration of actin at muscle ends (B). **(C-D)** A disorganization similar to the actin network is observed for Tm2, an actin binding protein and muscle differentiation marker, where WT embryos have an organized arrangement (C) as opposed to *Ssdp* mutants (D). **(E-F)** The microtubule network as visualized by an antibody against ß3-Tubulin shows that this network is equally disorganized in *Ssdp* mutants (F) in comparison to WT embryos (E). **(G)** The expanse of the distance to which the LT1 muscles extend in each hemisegment is significantly lower in *Ssdp* mutants as indicated by a t-test. **** = p-value < 0.0001 at a 95% confidence interval. The expanse measured for each LT1 is indicated by a line in (C) and (D).

Since the muscle differentiation marker, Tm2 is expressed in somatic muscles and muscles are arranged more or less in their WT pattern although they are severely affected, this indicates that the muscles initiate the differentiation program, but fail to establish their identity and the identity program is possibly deregulated.

3.5. iTF expression is downregulated in the absence of Ssdp

Muscle identity acquisition is regulated by iTFs and their downstream realisators. We previously demonstrated that the attenuated expression of one iTF, Ladybird (Lb) (Junion et al. 2007) causes perturbations in the identity acquisition of the Lb-dependent segment border muscle (SBM). The affected muscle pattern and innervation observed in *Ssdp* mutants prompted us to test whether the loss of *Ssdp* could have an impact on the expression of iTFs and in turn the acquisition of muscle identity. We chose to test two iTFs, Collier (Col) determining dorsal DA3 muscle identity (Crozatier et Vincent 1999) and Slou involved in the identity of several ventral and lateral muscles including the ventral acute VA2/3, dorso-lateral DT1 and ventral transverse VT1 muscles. The expression of Col as well as Slou is attenuated in a *Ssdp* loss of function context (Figure 5).

The reduced levels of iTF expression in *Ssdp* mutants is significant in the context of the acquisition of muscle identity since iTFs maintain muscle-specific levels of transcription of realisator genes that are downstream of iTFs in individual muscles to generate appropriate levels of protein for each muscle to attain its specific identity (Bataillé et al. 2017).

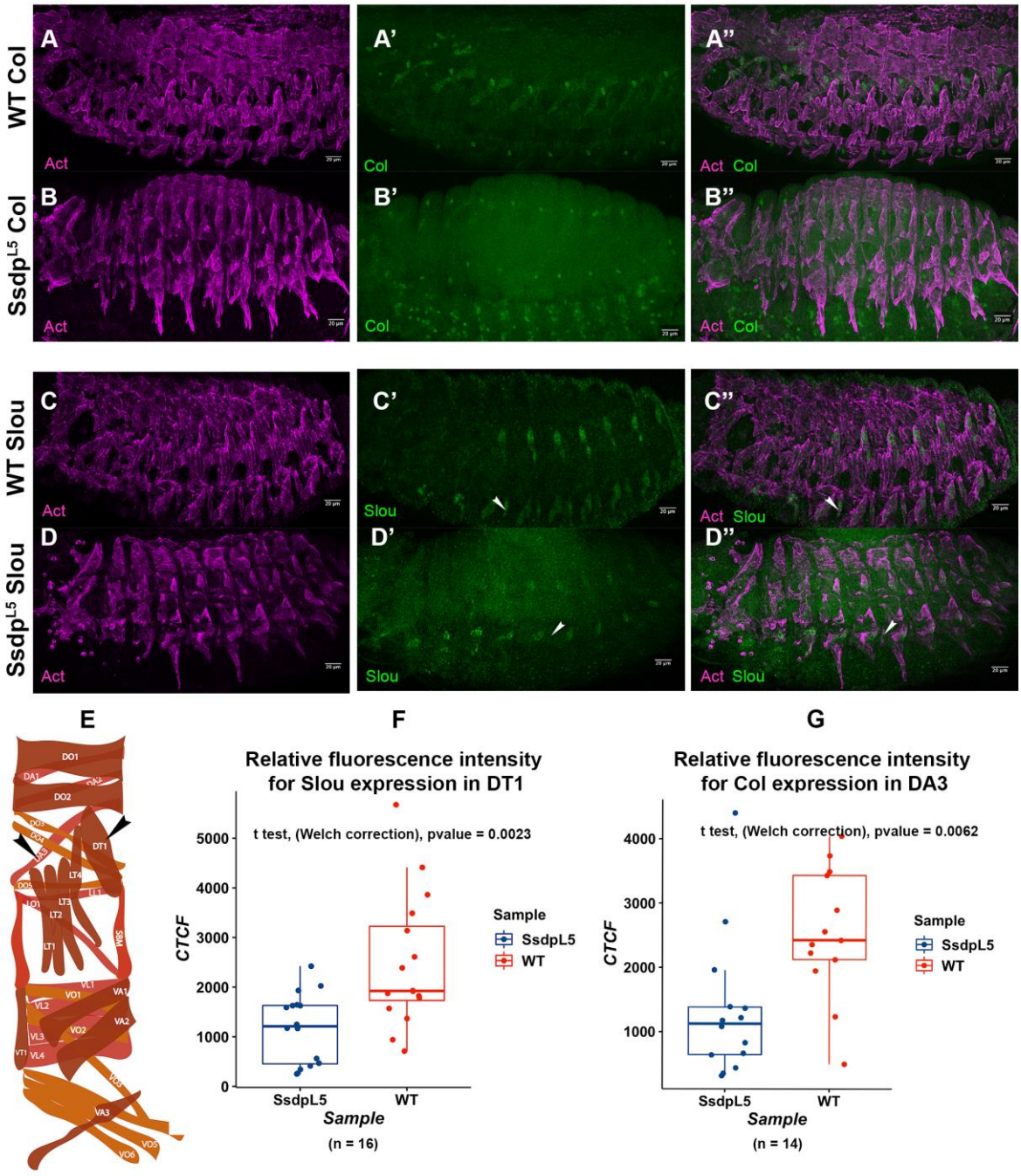


Figure 5.

Muscle identity transcription factor (iTF) expression is downregulated in *Ssdp* mutants.

(A-A″) Col expression as visualized by staining using an anti-Col antibody in the DA3 muscle in WT stage 16 embryos. **(B-B″)** Col expression is downregulated in *Ssdp* mutant stage 16 embryos. **(C-C″)** Slou expression as visualized by immunostaining using an anti-Slou antibody in stage 16 WT embryos. **(D-D″)** Slou expression is downregulated in *Ssdp* mutants. Expression is almost negligible in the VT1 muscles (arrowheads in D′, D′′) compared to the WT muscles (arrowheads in (C′, C′′). **(E)** The WT embryonic somatic muscle pattern. The DA3 and DT1 muscles are highlighted by arrowheads. **(F-G)** A quantification of intensities of Slou (F) and Col (G) by CTCF (Corrected Total Cell Fluorescence) using ImageJ shows a significant reduction in fluorescence intensities in *Ssdp* mutants with respect to WT.

3.6. Ssdp mutant muscles differentiate, but fail to acquire their final identity

Given the downregulation of iTFs and aberrant expression of the muscle differentiation marker Tm2, we wanted to clearly distinguish between a differentiation defect versus a defect in the acquisition of muscle identity in *Ssdp* mutants. To this end, we assessed whether Ssdp is required for the acquisition of two major iTF regulated properties of individual muscles, their attachment and fusion program apart from the morphological and innervation identity phenotypes observed.

3.6.1. *Ssdp* loss results in muscle-specific attachment phenotypes

ßPS-integrin localizes to the tips of LT muscles at locations where they attach to their intrasegmental attachment sites as well as to the termini of muscles that attach to intersegmental attachment sites. The LT tip-associated ßPS expression is absent in *Ssdp* mutants and its localization at the intersegmental attachment sites of the severely affected ventral muscles is much less expansive in *Ssdp* mutants compared to the WT (Figure 6A-B'').

A lack of muscle extension to reach their attachment sites and misdirected ventrally extending VO4-6 have been observed in conditions where the ventral muscle iTF *vestigial* (*vg*) and its interacting partner *scalloped* (*sd*) are ectopically expressed in all somatic muscles driven by Mef2-Gal4 (Deng et al. 2009). This phenotype is also observed in *stripe* (*sr*) mutants and on ectopic expression of the *sr-b* isoform of *sr* in the ventral midline (Frommer et al. 1996; Vorbrüggen et Jäckle 1997). The ventral most VO4-6 muscles are severely affected in conditions of *Ssdp* loss of function and appear fused and indistinguishable from each other. In WT embryos, they traverse into the adjacent segment to find their attachment sites and attach to them. In *Ssdp* mutants, they either remain in the same segment and travel straight ventrally instead or do not extend at all (Figure 6A', B'). Thus, the loss of Ssdp and the ectopic expression of Vg result in redundant ventral muscle phenotypes indicating a Ssdp-induced imbalance in yet another iTF, Vg and/or deregulated ectodermal Sr expression.

Muscles assuming a rounded appearance are observed in *sr* (Frommer et al. 1996) mutants that fail to attach and *myospheroid* (*mys* coding for ßPS-integrin) mutants that detach after initial attachment (Wright 1960; Leptin et al. 1989). The rounded muscle phenotype is frequently seen in *Ssdp* mutants. In 15% of embryos, almost all muscles present a rounded appearance with the lateral region being the most affected. In embryos where only a small proportion of muscles are rounded, up to 4 external muscles are rounded per embryo in hemisegments A2-A5 (Figure 6C, C').

3.6.2. Myoblast fusion is defective under the loss of *Ssdp*

Myocyte enhancer factor 2 (Mef2), the *Drosophila* homolog of vertebrate MEF2, is a MADS box transcription factor in the absence of which muscle FCs fail to differentiate

after they are correctly specified (Ranganayakulu et al. 1995). It regulates the expression of a vast array of genes (Junion et al. 2005; Sandmann et al. 2006) in a dose dependent manner (Elgar, Han, et Taylor 2008). It regulates muscle identity in concert with iTFs such as Ladybird (Lb) and Vestigial (Vg) (Junion et al. 2007; Deng et al. 2009) by differentially regulating the levels of muscle genes based on its interactors. Muscle size is determined by the number of rounds of fusion, that is in turn dictated by the iTF code that regulates the muscle-specific expression levels of identity realisators including cytoskeletal modulator genes such as *Muscle Protein 20* (*Mp20*) and *Paxillin* (*Pax*) to regulate fusion (Bataillé et al. 2010). These genes start to get expressed from stage 13, which coincides with the time when we first observe detectable expression of the long isoforms of *Ssdp* in muscles.

We observe Mef2 expression in the nuclei of developing myotubes in *Ssdp* mutants similar to WT myonuclei (Figures 6D-E''). However, once the differentiation program is correctly initiated by the fusion of FCs with FCMs in *Ssdp* mutants, there are aberrations in the execution of muscle identity dependent fusion programs. In the WT, the Slou+ DT1 muscle, for example, has 10-11 myonuclei after fusion in the A2-A5 abdominal hemisegments with very little variation among embryos. *Ssdp* mutant DT1 muscles exhibit huge variations in the number of myonuclei among different embryos as well as within the same embryo ranging from a missing DT1 muscle to presenting up to 17 myonuclei (Figure 6F). In addition, DT1 presents an aberrant, elongated morphology with centralized nuclei as opposed to nuclei that localize more towards muscle ends in WT DT1 muscles (Figure 6D'', E'').

These morphology, innervation, attachment and fusion defects observed along with the attenuated, but correctly patterned expression of iTFs such as Slou and Col in expected muscles in parallel with the unaffected expression of the key muscle differentiation factor Mef2 support the view that the muscle identity program is initiated on differentiation, but the maintenance and establishment of a muscle-specific identity program is hindered in embryos lacking *Ssdp*.

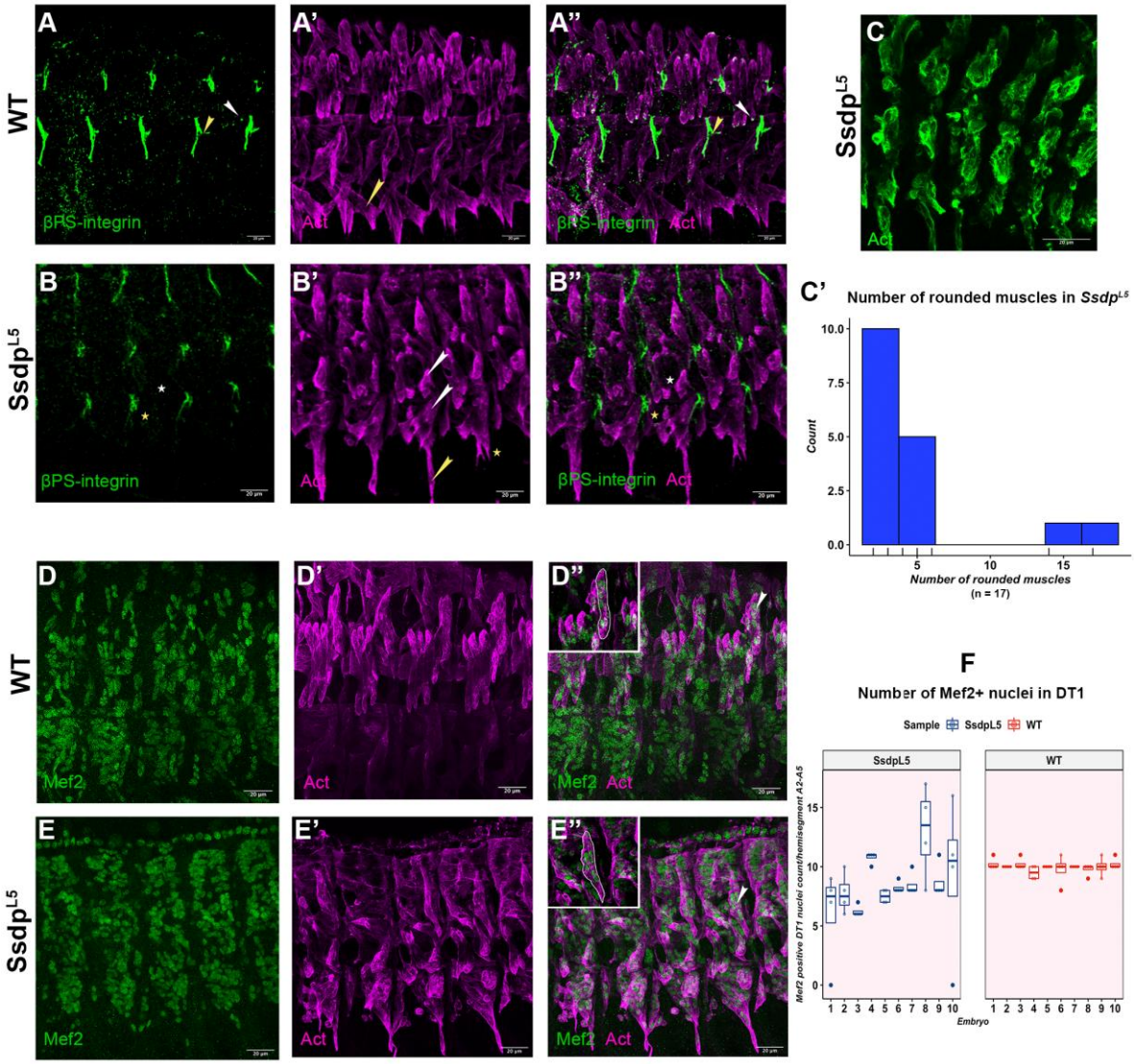


Figure 6.

The somatic muscles of *Ssdp* mutant embryos exhibit aberrant attachment and fusion.

(A-A'') In stage 16 WT embryos, ßPS-integrin localizes to the tips of LT muscles where they attach to intrasegmental attachment sites (white arrowhead in A, A'') and at the intersegmental attachment sites of ventral muscles (yellow arrowhead in A, A''). **(B-B'')** *Ssdp* mutants lack ßPS-integrin localization at the tips of LTs (white asterisk in B, B'') and there is much less accumulation at ventral muscle attachment sites (yellow asterisk in B, B''). Muscles that attach to intrasegmental sites such as the LTs and VA2 muscles are frequently reduced to globs (white arrowheads in B'). Unlike WT embryos where the ventral VO4-VO6 muscles traverse into the next hemisegment for attachment (arrowhead in A'), they navigate down ventrally (yellow arrowhead in B') or fail to extend in search of their attachment sites (asterisk in B') in *Ssdp* mutants. **(C)** In *Ssdp* mutants, around 15% of the embryos (3 out of 20) present with almost all muscles being rounded. **(C')** Among embryos where only a portion of the muscles are rounded, the majority of embryos have 2-4 rounded muscles in hemisegments A2-A5. **(D-D'')** In the DT1 of WT stage 16 embryos, Mef2+ nuclei are localized mostly towards muscle ends (arrowhead and inset in D''). **(E-E'')** Mef2+ nuclei are clearly discernible in *Ssdp* mutants. In DT1, however, they are localized centrally (arrowhead and inset in E''). **(F)** The number of Mef2+ nuclei in the DT1 muscle in hemisegments A2-A5 displays a high degree of variation in *Ssdp* mutants, between embryos as well as within the same embryo. WT DT1 muscles have 10-11 nuclei with very little variation.

3.7. The loss of Ssdp affects Wg expression

It was recently shown that Ssdp is involved in the transduction of Wnt/Wg signaling as part of Wnt enhanceosome complexes (Fiedler et al. 2015; Renko et al. 2019), although no function has been attributed to it during embryonic development yet. We thus asked whether Ssdp-Wg interactions could at least in part explain the complex muscle phenotypes of *Ssdp* mutants. We first tested whether embryonic Wg expression is maintained in a *Ssdp* loss of function context. In WT embryos, Wg displays a stage-specific, patterned and segmental epidermal expression (Ohlmeyer et Kalderon 1997). In *Ssdp* mutants, epidermal Wg expression is reduced during mid stages of development and is undetectable in late stage embryos (Figure 7A-D'').

3.8. The loss of Wg and Ssdp impact Eve expression similarly

Two studies showed that the muscle and heart iTF Even-skipped (Eve) is responsive to Wg signaling. Reducing Wg signaling by expressing a dominant negative form of dTCF in the mesoderm or mutating dTCF sites in a mesodermal specific *eve* enhancer resulted in one Eve+ pericardial cell instead of the normal two per hemisegment by stage 13 (Halfon et al. 2000; Stefan Knirr et Frasch 2001). Interestingly, in *Ssdp* mutants, we observe an average of one Eve+ pericardial cell per hemisegment similar to these studies (Figure 7E-G).

These observations suggest that Ssdp is required either: i) for the maintenance of late Wg expression in the epidermis (in line with our observation of reduced epidermal Wg in *Ssdp* mutants and the epidermal expression of *Ssdp* mRNA), ii) for the transduction of Wg signals to the mesoderm as a component of the Wg enhanceosome (in line with Ssdp expression in the developing muscles) and/or iii) for both these functions.

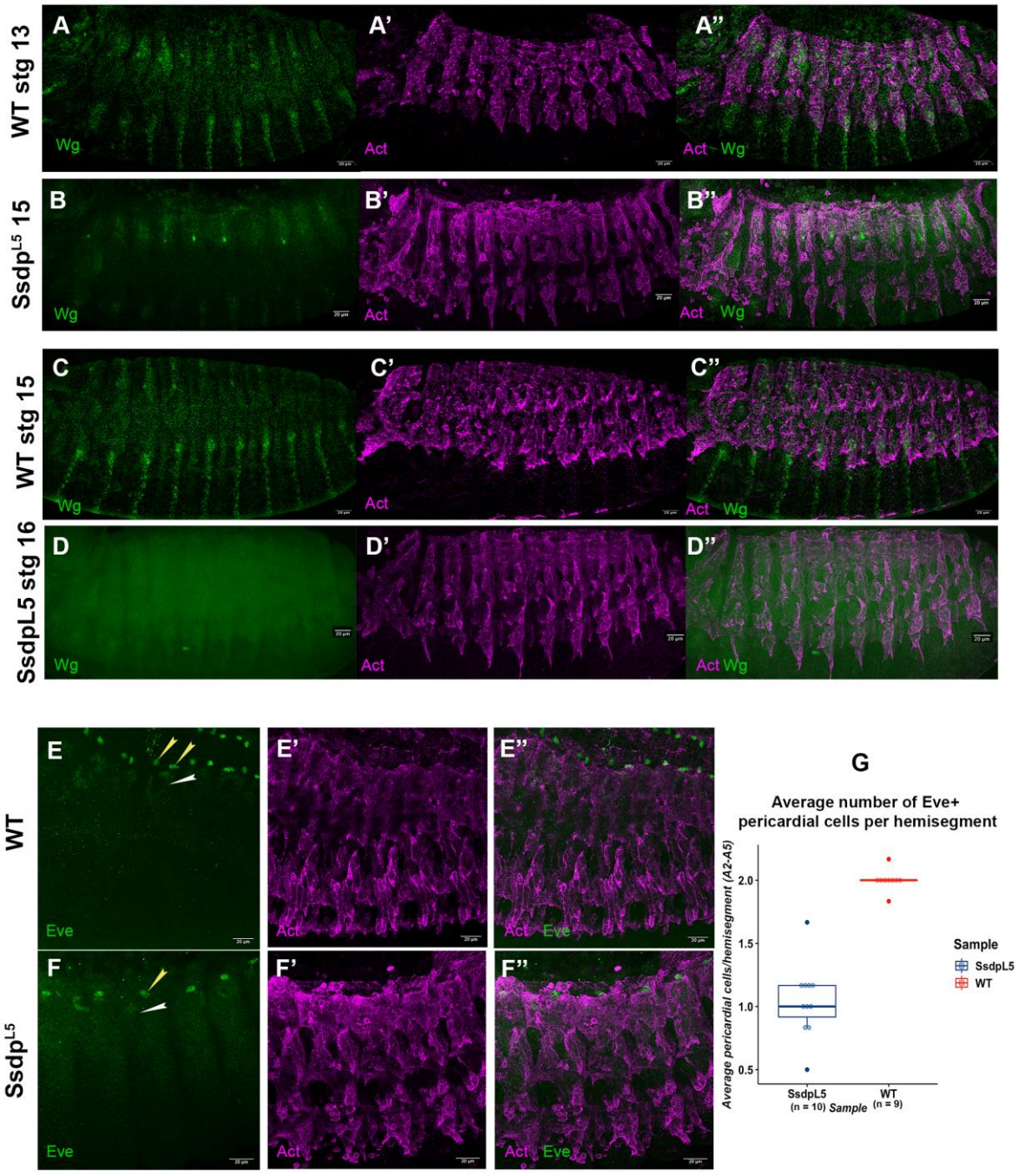


Figure 7.

The loss of function of *Ssdp* influences Wg expression and the number of Eve+ pericardial cells. **(A-A'')** In stage 13 WT embryos, a staining with an anti-Wg antibody reveals patterned dorsal and ventral expression. **(B-B'')** In *Ssdp* mutants, Wg expression is highly reduced during mid embryonic stages where a high number of fusion foci are evident. **(C-C'')** Wg expression is maintained in stage 15 WT embryos. Its expression is more expansive ventrally. **(D-D'')** Wg expression is undetectable at later stages in *Ssdp* mutants. **(E-E'')** In WT embryos, the Eve protein is expressed in two pericardial cells per hemisegment (yellow arrowheads in E) and in the DA1 muscle (white arrowhead in E) by stage 13. **(F-F'')** In *Ssdp* mutants, we observe a single pericardial cell per hemisegment (yellow arrowhead in F). Eve expression in the DA1 muscle is highly reduced (white arrowhead in F). **(G)** A quantification of Eve+ pericardial cells in *Ssdp* mutants confirms the pericardial cell phenotype.

3.9. Ssdp regulates a subset of Wg driven muscle characteristics and Wg plays stage-specific roles during embryonic myogenesis

In light of the similar impact of reduced *wg* and loss of *Ssdp* on the Eve iTF that is also expressed in the cardiac mesoderm, we sought to verify whether muscle defects induced by the loss of *wg* or its effector dTCF (or Pangolin (Pan)) are reminiscent of those in *Ssdp* mutants. It has previously been shown that Wg signaling mediated by dTCF is required for the proper specification of ventral muscle progenitors that give rise to the VA1, VA2 and VA3 muscles (Cox et Baylies 2005; Cox, Beckett, et Baylies 2005). Intriguingly, these muscles are severely affected in *Ssdp* mutants. VA1 and VA2 are malformed and VA3 is missing in *Ssdp* mutants (Figure 8A-B). Driving a dominant negative form of dTCF (dTCF^{DN}) in all muscles using the early mesodermal 24B-Gal4 leads to partial phenotypes of missing VA1, VA2 and VA3 indicating that dTCF^{DN} is not fully penetrant in embryonic muscles. This leads to a heterogeneity in phenotypes. However, some of the muscle phenotypes including an aberrant, elongated shape or loss of DT1 as well as missing VA3 muscles are common to dTCF^{DN} and *Ssdp* mutant embryos (Figure 8B-D). Therefore, dTCF^{DN} mutant phenotypes only partially overlap that of *Ssdp* mutants.

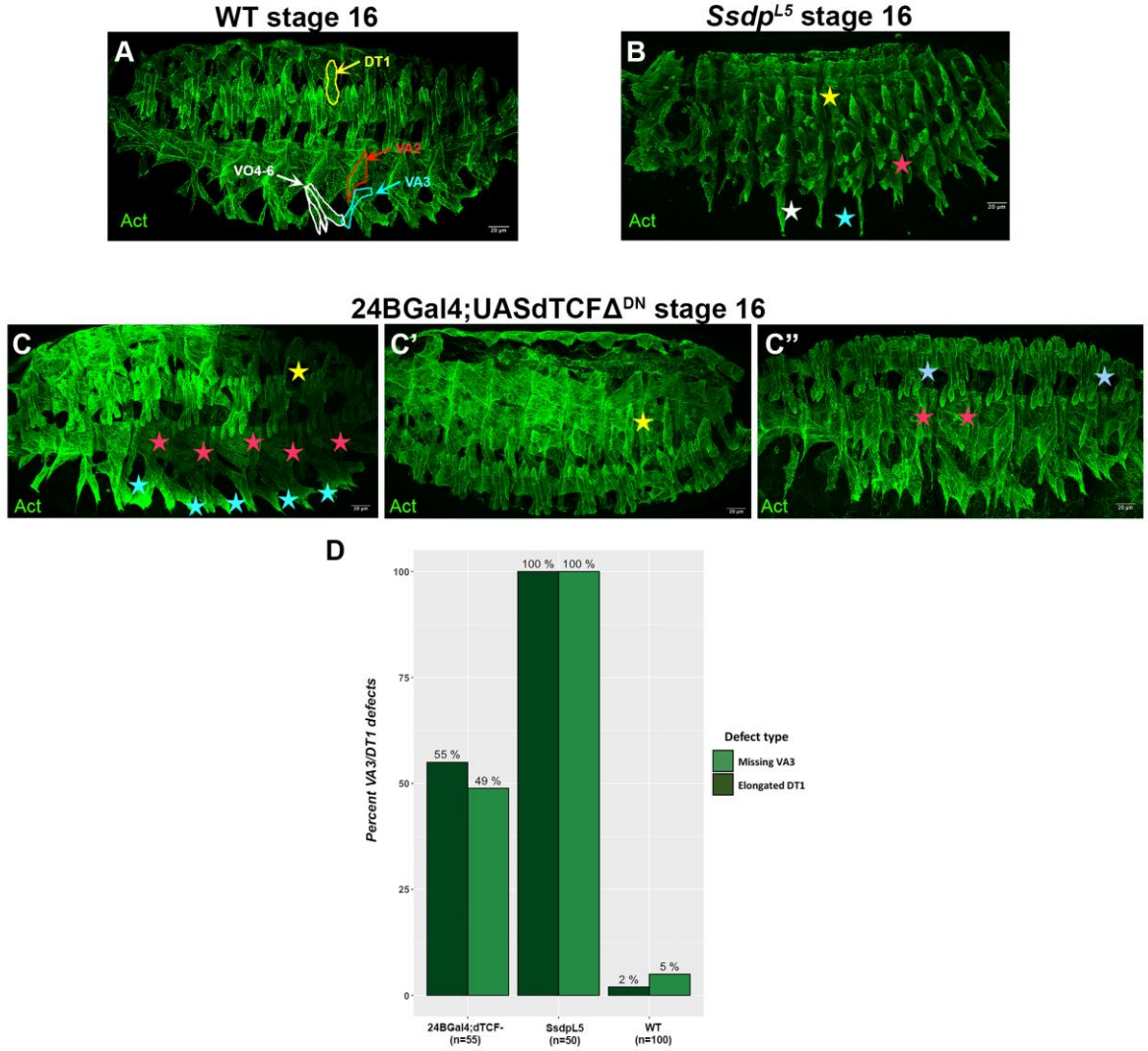


Figure 8.

***Ssdp* mutant phenotypes are more pronounced than dTCF dominant negative mutants.**

(A) The somatic muscle pattern in stage 16 WT embryos. The DT1, VA2, VA3 and VO4-6 muscles are highlighted in yellow, red, white and cyan respectively. **(B)** The somatic muscle pattern in stage 16 *Ssdp* mutant embryos in which all of these muscles exhibit morphological defects (highlighted by similar colored asterisks as muscles in (A)). **(C)** Expressing a dominant negative form of dTCF ($dTCF^{DN}$) using an early muscle specific driver, 24B-Gal4 reveals a heterogeneity in phenotypes depending on the penetrance of $dTCF^{DN}$. The ventral VA1, VA2 and VA3 muscles are specified in a Wg dependent fashion and these muscles are missing in all hemisegments in C (red and cyan asterisks). In this embryo, the Slou+ DT1 muscles are missing or malformed (yellow asterisk). **(C'')** Highlighting the heterogeneity of phenotypes, this embryo has missing VA1 and/or VA2 (red asterisks) in a few hemisegments associated with LT duplications (light blue asterisks). **(C'')** In some embryos, DT1 presents an abnormal, elongated morphology (yellow asterisk) or is absent similar to *Ssdp* mutants (B). **(D)** A quantification of the missing VA3 and elongated DT1 phenotypes in $dTCF^{DN}$ versus *Ssdp* mutants reveals that these phenotypes are present in close to 50% of hemisegments in the partially penetrant $dTCF^{DN}$ mutants.

We also analyzed *wg* temperature sensitive mutants (wg^{ts}) by inhibiting *wg* at different developmental stages to test Wg requirements in developing somatic muscles. When flies were allowed to develop normally at 18°C until around stage 9-11 and the eggs were then shifted to a non-permissive temperature of 28°C to inhibit *wg* expression until stage 16, we observe 2 distinct phenotypes that might correspond to different stages when *wg* was switched off in the embryo (Figure 9B, B'). The first phenotype is a complete disruption of muscle development. This is expected since Wg is required for mesoderm specification in the embryo (Azpiazu et al. 1996). The second phenotype is highly deregulated somatic muscle development with the ventral VO4-6 muscles being directed straight down ventrally instead of traversing into the adjacent hemisegment similar to *Ssdp* mutants. When *wg* is switched off between stage 15 and 16, we observe an additional phenotype where most muscles are specified correctly, but VA3 is undetectable as in the case of *Ssdp* mutants, which raises the question of whether Wg is necessary for the maintenance of this muscle (Figure 9B''). Muscles display severe attachment defects with the LT muscles extending too far dorsally to attach and ventral muscles associating with incorrect attachment sites. VA1 and VA2 present aberrant morphologies that could be the result of them attaching to ectopic sites. Thus, Wg plays stage-specific roles during embryonic somatic muscle development.

We analyzed precisely staged wg^{ts} embryos to verify if there were specific phenotypes that were distinguishable. When flies were allowed to lay eggs for 3 hours at the permissive temperature and the eggs were allowed to develop normally until around stage 12-13, which is right after FC specification, before being shifted to the non-permissive temperature of 28°C, we observe distinct phenotypes (Figure 9C). LT muscles extend to much greater distances than WT embryos. The VO4-6 muscles

project ventrally in search of attachment sites as observed in *Ssdp* mutants in many cases. The VA2 and VA3 muscles are smaller and present aberrant morphologies. Given the temporal specificity of phenotypes observed, we examined similarly staged wg^{ts}/+;$Ssdp^{L5}$/+ transheterozygotes to determine possible genetic interactions between Ssdp and Wg (Figure 9C'). These embryos display heterogenous actin staining indicating a disorganized actin cytoskeleton as seen in *Ssdp* mutants. They exhibit severe morphological and attachment defects in VO4-6 in the anterior-most hemisegments. Somatic muscles appear relatively normal in some embryos, but ventral muscles exhibit defects in morphology and attachment. VA2 and VA3 are absent or smaller with aberrant morphology also observed in similarly staged wg^{ts} mutants. A quantification of these phenotypes in staged embryos (Figure 9C'') reveals that ventral defects are more pronounced in wg^{ts} mutants compared to wg^{ts}/+;$Ssdp^{L5}$/+ transheterozygotes with VA3 being the most affected, followed by VA2.

These observations suggest that all *Ssdp* loss of function phenotypes cannot be explained by muscular transduction of Wg signals by dTCF. However, one has to take into consideration that $dTCF^{DN}$ is not fully penetrant in embryonic muscles. wg^{ts} mutants, on the other hand, present more severe phenotypes. There is distinct overlap between phenotypes observed in *Ssdp* mutants versus $dTCF^{DN}$ mutants and wg^{ts} mutants. The analysis of staged wg^{ts}/+;$Ssdp^{L5}$/+ transheterozygotes indicates a genetic interaction between Wg and Ssdp at mid developmental stages to regulate actin cytoskeletal dynamics and the establishment of muscle identity of a specific subset of ventral muscles given that heterozygotes for each of them do not display these phenotypes. These observations suggest a stage dependent requirement and specific interaction between Wg signaling and Ssdp in the regulation of the identity of the VA2, VA3 and VO4-6 muscles during embryonic somatic muscle development.

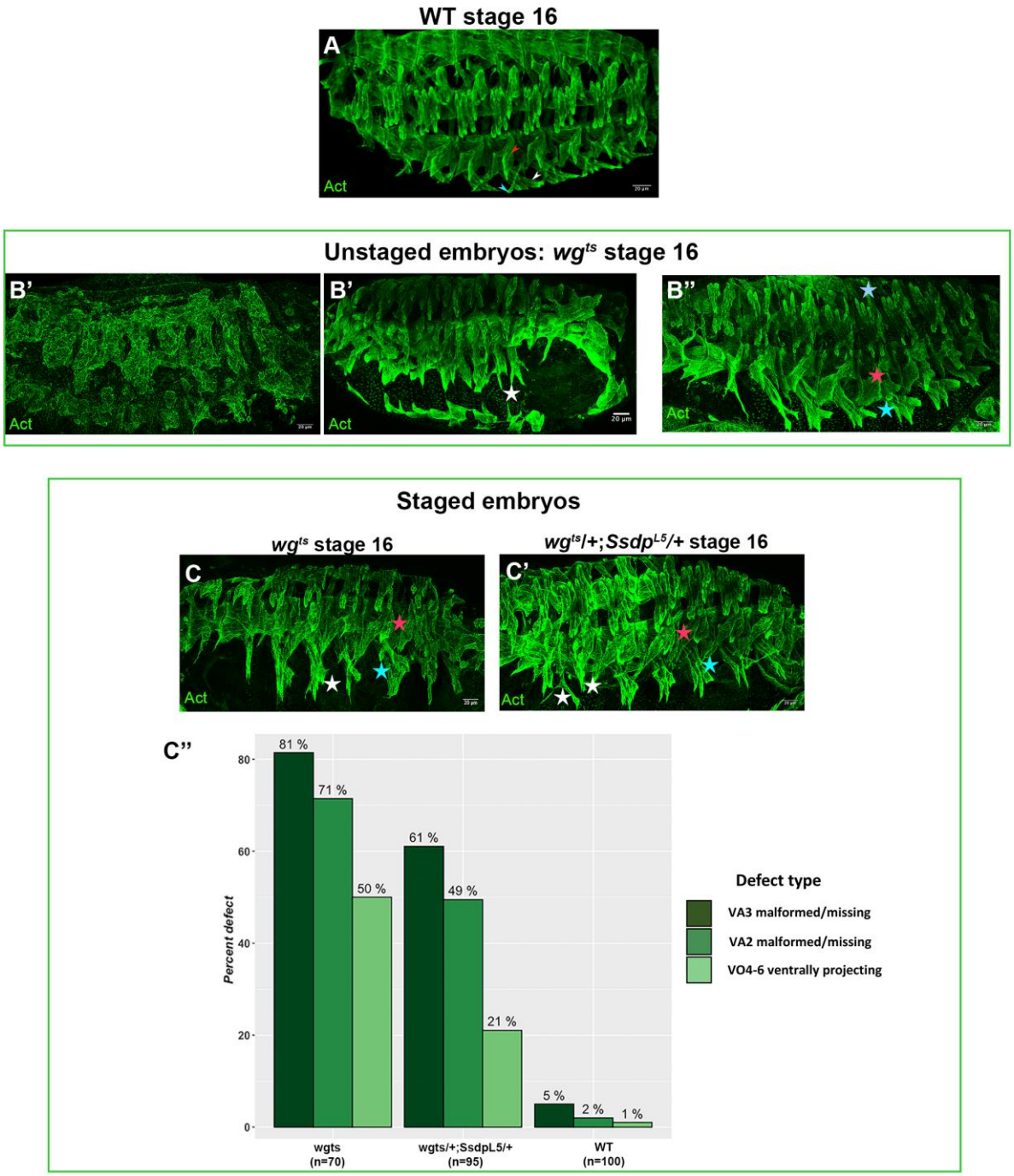


Figure 9.

Wg plays stage-specific roles during embryonic somatic muscle development and genetically interacts with Ssdp during mid stages to regulate VA2 and VA3 identity.

(A) The somatic muscle pattern in stage 16 WT embryos. VA2, VA3 and VO4 are highlighted (red, cyan and white arrowheads respectively). **(B-B'')** Unstaged wg^{ts} mutant embryos with 0-12 hours of normal development with continuous egg-laying at 18°C (~stage 9-11) before shifting to the non-permissive 28°C to deactivate *wg* until stage 16 are displayed in (B) and (B'). They display multiple phenotypes including completely disrupted myogenesis (B) and deregulated myogenesis with ventrally projecting VO4-6 (B') also observed in *Ssdp* mutants. When allowed to develop normally for longer (0-26 hours or ~stage 15-16) before being shifted to 28°C (B''), additional severe attachment phenotypes are observed in VA2 and LTs (red and light blue asterisks in B'' respectively). VA3 is undetectable (cyan asterisk). **(C)** Staged wg^{ts} mutant embryos with 3 hours of egg laying at 18°C followed by normal development for 13-17 hours (~stage 12-13) before shifting to 28°C exhibit specific phenotypes. VA2 and VA3 are malformed/missing while VO4-6 are indistinguishable from each other

and project ventrally similar to *Ssdp* mutants. **(C')** Similarly staged wg^{ts}-/+;$Ssdp^{L5}$-/+ transheterozygotes display a disorganized actin cytoskeleton. They display attachment defects in the ventral-muscles in anterior hemisegments (white asterisks). Some phenotypes such as missing or malformed VA2 and VA3 muscles (red and cyan asterisks respectively) overlap those observed in wg^{ts} mutants. **(C'')** A quantification of ventral muscle defects in staged wg^{ts} mutants and wg^{ts}/+;$Ssdp^{L5}$/+ transheterozygotes reveals that the percentages of these defects are significantly higher compared to the WT.

4. Discussion

Ssdp and isoform specificity during somatic myogenesis

In line with the differential expression of *Ssdp* in muscles in our transcriptomic data, our analysis of *Ssdp* mutants shows that this gene plays a critical role during muscle development. The differential expression of *Ssdp* mRNA under translation among different muscle subsets in our transcriptomic data reveals a potential muscle-subset-specific role for this gene. This might be due to its requirement to form specific complexes with LIM homeodomain factors such as the LT iTF, Ap (Bronstein et al. 2010). The observation of an enrichment for CT-rich motifs in the LT subset is also in line with the significant upregulation of *Ssdp* under translation in our dataset. No direct DNA-binding has been proven *in vivo* for Ssdp. It has been shown to bind DNA by interacting with its cofactors as part of the ChiLS and Wnt enhanceosome complexes. So, it is possible that this enrichment represents Ssdp's interaction with other cofactors. Trl (Trithorax-like) is a potential interactor given that it is known to bind complementary GA-rich motifs and is significantly upregulated along with Ssdp during T3 in our data.

In humans, *SSBP3* downregulation and mis-splicing with the retention of exon 6 has been observed in the skeletal muscles of myotonic dystrophy type 1 and 2 patients as well as patients with neuromuscular diseases (NMD) such as Duchenne muscular dystrophy (Bachinski et al. 2014), although it has received no attention and no role during myogenesis has been attributed to it yet. The differences in the expression patterns of the *in situ* probes targeting the long isoforms versus all isoforms indicate the presence of predominant isoforms with a potential isoform switch at different stages of development, which could be related to isoform specific functionality already hinted at in human muscular dystrophies.

Ssdp and its influence on muscle identity properties

Identified methods by which iTFs direct identity acquisition include controlling the muscle-specific number of rounds of myoblast fusion (Bataillé et al. 2010), reprogramming newly fused nuclei to adopt a muscle-specific program (Bataillé et al. 2017) and correct attachment site selection (Carayon et al. 2020). Given this, our observations of downregulated iTF expression as well as deregulated myoblast fusion and attachment in specific muscles suggest a requirement for Ssdp during mid myogenesis after the muscle identity program and muscle differentiation have been initiated. This is when we detect expression of the long *Ssdp* isoforms in somatic

muscles, which is potentially related to muscle and stage-specific functions for these isoforms. However, this does not rule out a requirement for Ssdp during early stages of development when it is ubiquitously expressed and the zygotic null mutants used here might hamper the observation of this early requirement.

The roles of Ssdp and Wg on muscle phenotypes

The extremely low Wg expression in *Ssdp* mutants during mid stages and lack of Wg observed at later stages indicate a role for Ssdp in the maintenance of Wg expression at these stages when the long isoforms are highly expressed. Since the Wg dependent Slou+ cluster is largely specified except for VA3, this indicates sufficient Wg levels at early stages of muscle development when VA1/2 progenitors are specified, but a dependence on Ssdp for VA3 specification. It is unclear if the low levels of iTF expression observed at later stages is the consequence of downregulated Wg. Given that Ssdp is part of the evolutionarily conserved canonical Wnt enhanceosome, it is not surprising that it in turn affects Wg expression as this study unveils for the first time. Since the vertebrate Wnt genes play very specific roles during myogenesis, it remains to be seen if Ssdp also influences the expression of other *Drosophila* Wnts and if they have roles during myogenesis. The analysis of wg^{ts}/+;$Ssdp^{L5}$/+ transheterozygotes confirms stage and muscle specific interactions of Ssdp to regulate Wg signaling, especially in establishing the identity of VA2 and VA3. Whether this is via the canonical pathway is an open question.

This study also reveals stage-specific roles for Wg during all stages of embryonic somatic myogenesis for the first time. Although Wg is known to be required for the specification of the mesoderm and the specification of some FCs during early stages, no role has been attributed to it during mid-late stages. We identify a role for it in regulating the correct attachment of muscles during later stages as evidenced by the severe attachment phenotypes observed when *wg* is deactivated during late stages.

Singular or converging roles of Ssdp could play a part in somatic muscle development

As a component of the Wnt enhanceosome and because Wnt signaling is a central component of the developmental symphony, any deregulations in components of this pathway can have deleterious cascading effects. Since humans have 3 *SSBP* family members and *Drosophila* only one, it could mean that the single gene carries out all functions associated with vertebrate counterparts or that *Drosophila* requires only a subset of its vertebrate counterpart family's functionality.

The stark variability in the number of myonuclei in the DT1 muscle suggests a stochasticity due to deregulations in protein stoichiometry. So, Ssdp might play a role in providing a context-dependent boost of gene transcription. Given the incomplete overlap of phenotypes in various contexts in our study, it could be the known role of Ssdp as part of ChiLS that triggers the defects observed or a combination and convergence of this with unknown Ssdp roles that have a cumulative effect on myogenesis. Whether Ssdp plays cytoplasmic roles apart from forming nuclear

complexes is unknown. The LisH motif (van Meyel, Thomas, et Agulnick 2003) present in Ssdp has been implicated in controlling MT dynamics by homodimerizing or dimerizing in trans with other LisH proteins as well as in regulating protein localization (Emes et Ponting 2001; M. H. Kim et al. 2004; Gerlitz et al. 2005; Kannan et al. 2017). Analyzing Ssdp cofactors and conserved regions in the Ssdp protein other than the LUFS domain would help elaborate on Ssdp functions.

Author Contributions: Study conceptualization, P.P. and K.J.; methodology, P.P. and K.J.; software, P.P.; validation, P.P..; formal analysis, P.P.; investigation, P.P.; resources, P.P. and K.J. with help from J.P.; data curation, P.P. and K.J.; writing—original draft preparation, P.P.; writing—review and editing, P.P. and K.J.; supervision, K.J.; project administration, K.J.; funding acquisition, K.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the AFM-Telethon grant number 21182 to the MyoNeurAlp Alliance.

Acknowledgments: We would like to thank Benjamin Bertin for generating the TRAP datasets used for bioinformatic analysis in this study.

References

1. Azpiazu, N, P A Lawrence, J P Vincent, et M Frasch. 1996. « Segmentation and specification of the Drosophila mesoderm. » *Genes & Development* 10(24): 3183-94.
2. Bachinski, Linda L et al. 2014. « Most Expression and Splicing Changes in Myotonic Dystrophy Type 1 and Type 2 Skeletal Muscle Are Shared with Other Muscular Dystrophies ». *Neuromuscular disorders : NMD* 24(3): 227-40.
3. Bataillé, Laetitia et al. 2010. « Downstream of Identity Genes: Muscle-Type-Specific Regulation of the Fusion Process ». *Developmental Cell* 19(2): 317-28.
4. Bataillé, Laetitia, Hadi Boukhatmi, Jean-Louis Frendo, et Alain Vincent. 2017. « Dynamics of transcriptional (re)-programming of syncytial nuclei in developing muscles ». *BMC Biology* 15.
5. Bayarsaihan, Dashzeveg, Ricardo J. Soto, et Lewis N. Lukens. 1998. « Cloning and characterization of a novel sequence-specific single-stranded-DNA-binding protein ». *Biochemical Journal* 331(2): 447-52.
6. Bejsovec, A., et A. Martinez Arias. 1991. « Roles of Wingless in Patterning the Larval Epidermis of Drosophila. » *Development (Cambridge, England)* 113(2): 471-85.
7. Bertin, Benjamin et al. 2015. « TRAP-Rc, Translating Ribosome Affinity Purification from Rare Cell Populations of Drosophila Embryos ». *Journal of visualized experiments : JoVE* (103): 52985.
8. Bertin, Benjamin et al. 2021. « Gelsolin and dCryAB act downstream of muscle identity genes and contribute to preventing muscle splitting and branching in Drosophila ». *bioRxiv*: 784546.
9. Borello, U. et al. 1999. « Transplacental delivery of the Wnt antagonist Frzb1 inhibits development of caudal paraxial mesoderm and skeletal myogenesis in mouse embryos ». *Development* 126(19): 4247.
10. Bronstein, Revital et al. 2010. « Transcriptional Regulation by CHIP/LDB Complexes ». *PLoS genetics* 6(8): e1001063-e1001063.

11. Carayon, Alexandre et al. 2020. « Intrinsic Control of Muscle Attachment Sites Matching. » *eLife* 9.
12. Castro, Patricia, Hong Liang, Jan C Liang, et Lalitha Nagarajan. 2002. « A Novel, Evolutionarily Conserved Gene Family with Putative Sequence-Specific Single-Stranded DNA-Binding Activity ». *Genomics* 80(1): 78-85.
13. Chen, Lan et al. 2002. « Ssdp Proteins Interact with the LIM-Domain-Binding Protein Ldb1 to Regulate Development ». *Proceedings of the National Academy of Sciences of the United States of America* 99(22): 14320-25.
14. Cox, Virginia T., et Mary K. Baylies. 2005. « Specification of individual Slouch muscle progenitors in Drosophila requires sequential Wingless signaling ». *Development* 132(4): 713.
15. Cox, Virginia T., Karen Beckett, et Mary K. Baylies. 2005. « Delivery of wingless to the ventral mesoderm by the developing central nervous system ensures proper patterning of individual slouch-positive muscle progenitors ». *Developmental Biology* 287(2): 403-15.
16. Crozatier, M., et A. Vincent. 1999. « Requirement for the Drosophila COE transcription factor Collier in formation of an embryonic muscle: transcriptional response to notch signalling ». *Development* 126(7): 1495.
17. Delgado-Deida, Yaritza, Kibrom M Alula, et Arianne L Theiss. 2020. « The influence of mitochondrial-directed regulation of Wnt signaling on tumorigenesis ». *Gastroenterology Report* 8(3): 215-23.
18. Deng, Hua, Sarah C. Hughes, John B. Bell, et Andrew J. Simmonds. 2009. « Alternative Requirements for Vestigial, Scalloped, and Dmef2 during Muscle Differentiation in Drosophila Melanogaster. » *Molecular biology of the cell* 20(1): 256-69.
19. Elgar, Stuart J, Jun Han, et Michael V Taylor. 2008. « Mef2 Activity Levels Differentially Affect Gene Expression during Drosophila Muscle Development ». *Proceedings of the National Academy of Sciences of the United States of America* 105(3): 918-23.
20. Emes, Richard D., et Chris P. Ponting. 2001. « A new sequence motif linking lissencephaly, Treacher Collins and oral–facial–digital type 1 syndromes, microtubule dynamics and cell migration ». *Human Molecular Genetics* 10(24): 2813-20.
21. Enkhmandakh, Badam, Alexandr V Makeyev, et Dashzeveg Bayarsaihan. 2006. « The Role of the Proline-Rich Domain of Ssdp1 in the Modular Architecture of the Vertebrate Head Organizer ». *Proceedings of the National Academy of Sciences of the United States of America* 103(31): 11631-36.
22. Fiedler, Marc et al. 2015. « An Ancient Pygo-Dependent Wnt Enhanceosome Integrated by Chip/LDB-SSDP ». *eLife* 4: e09073.
23. Folker, Eric S, Victoria K Schulman, et Mary K Baylies. 2012. « Muscle Length and Myonuclear Position Are Independently Regulated by Distinct Dynein Pathways ». *Development (Cambridge, England)* 139(20): 3827-37.
24. Frommer, G et al. 1996. « Epidermal Egr-like Zinc Finger Protein of Drosophila Participates in Myotube Guidance ». *The EMBO journal* 15(7): 1642-49.

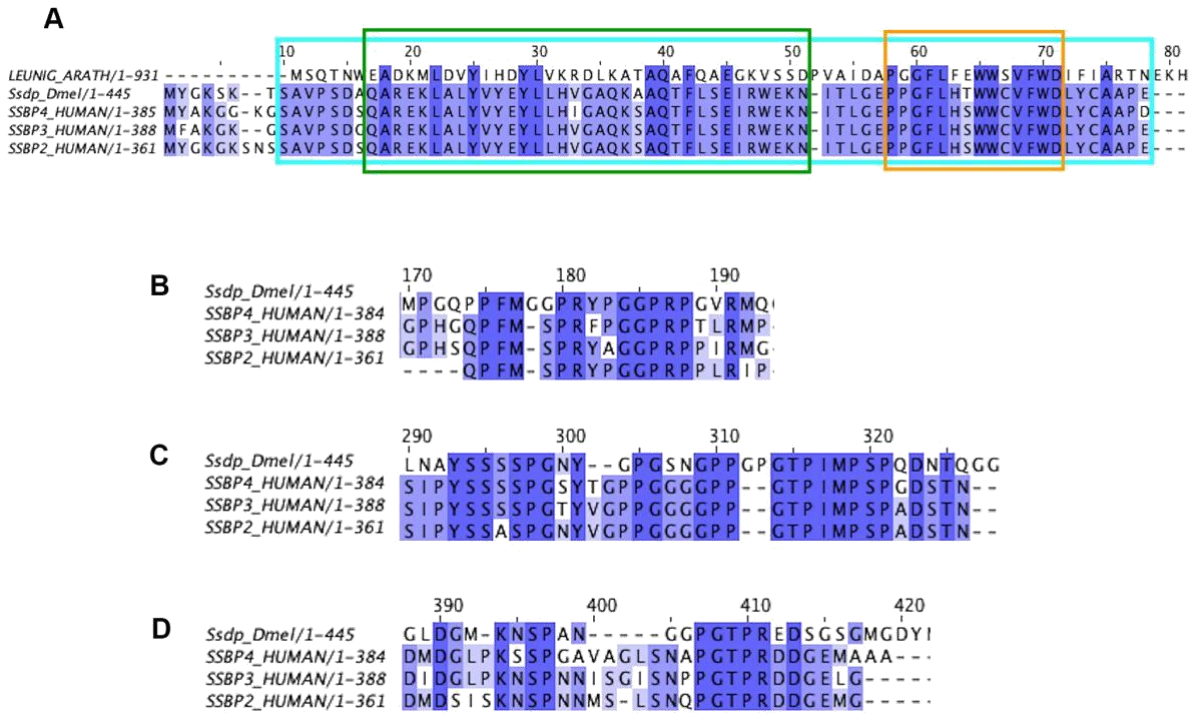
25. Füchtbauer, E. M. 1995. « Expression of M-Twist during Postimplantation Development of the Mouse. » *Developmental dynamics : an official publication of the American Association of Anatomists* 204(3): 316-22.
26. Gerlitz, Gabi et al. 2005. « Novel Functional Features of the Lis-H Domain: Role in Protein Dimerization, Half-Life and Cellular Localization. » *Cell cycle (Georgetown, Tex.)* 4(11): 1632-40.
27. Gros, Jérôme, Olivier Serralbo, et Christophe Marcelle. 2009. « WNT11 acts as a directional cue to organize the elongation of early muscle fibres ». *Nature* 457(7229): 589-93.
28. Güngör, Cenap et al. 2007. « Proteasomal Selection of Multiprotein Complexes Recruited by LIM Homeodomain Transcription Factors. » *Proceedings of the National Academy of Sciences of the United States of America* 104(38): 15000-5.
29. Halfon, Marc S et al. 2000. « Ras Pathway Specificity Is Determined by the Integration of Multiple Signal-Activated and Tissue-Restricted Transcription Factors ». *Cell* 103(1): 63-74.
30. Herrmann, Carl, Bram Van de Sande, Delphine Potier, et Stein Aerts. 2012. « I-CisTarget: An Integrative Genomics Method for the Prediction of Regulatory Features and Cis-Regulatory Modules ». *Nucleic acids research* 40(15): e114-e114.
31. Huxley, A. F., et R. Niedergerke. 1954. « Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres ». *Nature* 173(4412): 971-73.
32. Huxley, Hugh, et Jean Hanson. 1954. « Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation ». *Nature* 173(4412): 973-76.
33. Ikeya, M., et S. Takada. 1998. « Wnt signaling from the dorsal neural tube is required for the formation of the medial dermomyotome ». *Development* 125(24): 4969.
34. Imrichová, Hana et al. 2015. « I-CisTarget 2015 Update: Generalized Cis-Regulatory Enrichment Analysis in Human, Mouse and Fly ». *Nucleic acids research* 43(W1): W57-64.
35. Junion, Guillaume et al. 2005. « Mapping Dmef2-binding regulatory modules by using a ChIP-enriched in silico targets approach ». *Proceedings of the National Academy of Sciences of the United States of America* 102(51): 18479.
36. Junion, Guillaume et al. 2007. « Genome-wide view of cell fate specification: ladybird acts at multiple levels during diversification of muscle and heart precursors ». *Genes & Development* 21(23): 3163-80.
37. Kannan, Meghna et al. 2017. « WD40-Repeat 47, a Microtubule-Associated Protein, Is Essential for Brain Development and Autophagy. » *Proceedings of the National Academy of Sciences of the United States of America* 114(44): E9308-17.
38. Kim, Ji Hoon et al. 2015. « Mechanical Tension Drives Cell Membrane Fusion ». *Developmental Cell* 32(5): 561-73.
39. Kim, Myung Hee et al. 2004. « The Structure of the N-Terminal Domain of the Product of the Lissencephaly Gene Lis1 and Its Functional Implications. » *Structure (London, England : 1993)* 12(6): 987-98.

40. Knirr, S., N. Azpiazu, et M. Frasch. 1999. « The Role of the NK-Homeobox Gene Slouch (S59) in Somatic Muscle Patterning. » *Development (Cambridge, England)* 126(20): 4525-35.
41. Knirr, Stefan, et Manfred Frasch. 2001. « Molecular Integration of Inductive and Mesoderm-Intrinsic Inputs Governs Even-Skipped Enhancer Activity in a Subset of Pericardial and Dorsal Muscle Progenitors ». *Developmental Biology* 238(1): 13-26.
42. Kumar, Ram P, Krista C Dobi, Mary K Baylies, et Susan M Abmayr. 2015. « Muscle Cell Fate Choice Requires the T-Box Transcription Factor Midline in Drosophila ». *Genetics* 199(3): 777-91.
43. Lawrence, Peter A, et Gary Struhl. 1996. « Morphogens, Compartments, and Pattern: Lessons from Drosophila? » *Cell* 85(7): 951-61.
44. Lee, H. H., et M. Frasch. 2000. « Wingless Effects Mesoderm Patterning and Ectoderm Segmentation Events via Induction of Its Downstream Target Sloppy Paired. » *Development (Cambridge, England)* 127(24): 5497-5508.
45. Legendre, Félix et al. 2013. « Whole Mount RNA Fluorescent in Situ Hybridization of Drosophila Embryos ». *Journal of visualized experiments : JoVE* (71): e50057-e50057.
46. Leptin, Maria, Thierry Bogaert, Ruth Lehmann, et Michael Wilcox. 1989. « The function of PS integrins during Drosophila embryogenesis ». *Cell* 56(3): 401-8.
47. Liu, Jun-Wei et al. 2008. « SsDNA-Binding Protein 2 Is Frequently Hypermethylated and Suppresses Cell Growth in Human Prostate Cancer. » *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(12): 3754-60.
48. Mankoo, Baljinder S. et al. 2003. « A ». *Development (Cambridge, England)* 130(19): 4655-64.
49. van Meyel, Donald J., John B. Thomas, et Alan D. Agulnick. 2003. « Ssdp Proteins Bind to LIM-Interacting Co-Factors and Regulate the Activity of LIM-Homeodomain Protein Complexes in Vivo. » *Development (Cambridge, England)* 130(9): 1915-25.
50. Müller, Dominik et al. 2010. « Regulation and Functions of the Lms Homeobox Gene during Development of Embryonic Lateral Transverse Muscles and Direct Flight Muscles in Drosophila ». *PloS one* 5(12): e14323-e14323.
51. Münsterberg, A E et al. 1995. « Combinatorial signaling by Sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. » *Genes & Development* 9(23): 2911-22.
52. Ohlmeyer, Johanna Talavera, et Daniel Kalderon. 1997. « Dual pathways for induction of wingless expression by protein kinase A and Hedgehog inDrosophila embryos ». *Genes & Development* 11(17): 2250-58.
53. Orjalo, Arturo, Hans E Johansson, et Jerry L Ruth. 2011. « Stellaris™ fluorescence in situ hybridization (FISH) probes: a powerful tool for mRNA detection ». *Nature Methods* 8(10): i-ii.
54. Poitras, Jennifer L. et al. 2008. « Novel SSBP2-JAK2 fusion gene resulting from a t(5;9)(q14.1;p24.1) in pre-B acute lymphocytic leukemia ». *Genes, Chromosomes*

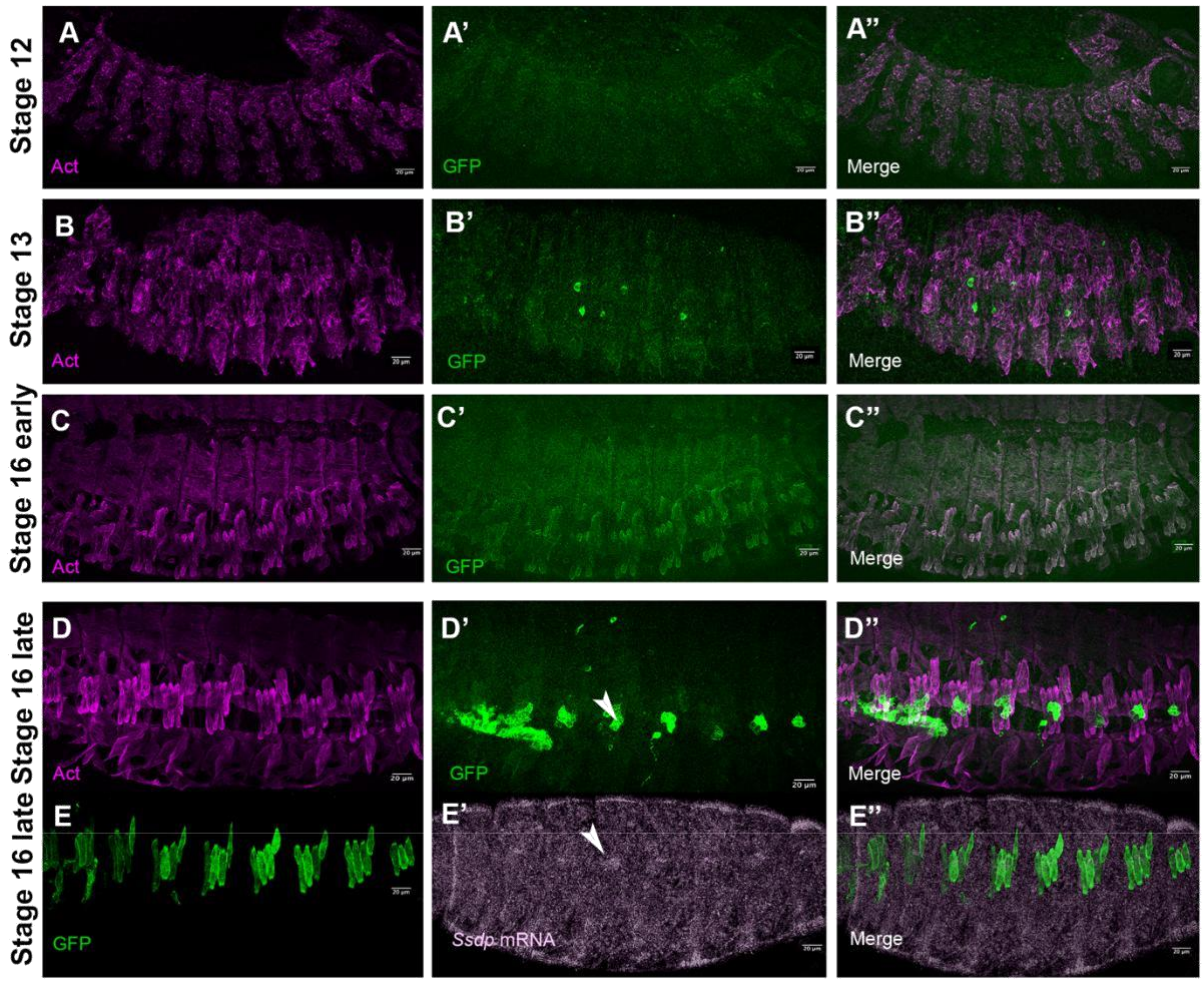
and Cancer 47(10): 884-89.
55. Potthoff, Matthew J., et Eric N. Olson. 2007. « MEF2: a central regulator of diverse developmental programs ». *Development* 134(23): 4131.
56. Raj, Arjun et al. 2008. « Imaging Individual MRNA Molecules Using Multiple Singly Labeled Probes ». *Nature methods* 5(10): 877-79.
57. Ranganayakulu, Gogineni et al. 1995. « A Series of Mutations in the D-MEF2 Transcription Factor Reveal Multiple Functions in Larval and Adult Myogenesis in Drosophila ». *Developmental Biology* 171(1): 169-81.
58. Relaix, Frédéric, Didier Rocancourt, Ahmed Mansouri, et Margaret Buckingham. 2005. « A Pax3/Pax7-dependent population of skeletal muscle progenitor cells ». *Nature* 435(7044): 948-53.
59. Renko, Miha et al. 2019. « Rotational Symmetry of the Structured Chip/LDB-SSDP Core Module of the Wnt Enhanceosome ». *Proceedings of the National Academy of Sciences of the United States of America* 116(42): 20977-83.
60. Richier, Benjamin et al. 2018. « Integrin Signaling Downregulates Filopodia during Muscle-Tendon Attachment ». *Journal of cell science* 131(16): jcs217133.
61. Ritzenthaler, Sarah, Emiko Suzuki, et Akira Chiba. 2000. « Postsynaptic filopodia in muscle cells interact with innervating motoneuron axons ». *Nature Neuroscience* 3(10): 1012-17.
62. Rudnicki, Michael A. et al. 1993. « MyoD or Myf-5 is required for the formation of skeletal muscle ». *Cell* 75(7): 1351-59.
63. Sandmann, Thomas et al. 2006. « A Temporal Map of Transcription Factor Activity: Mef2 Directly Regulates Target Genes at All Stages of Muscle Development ». *Developmental Cell* 10(6): 797-807.
64. Schnorrer, Frank, et Barry J Dickson. 2004. « Muscle Building: Mechanisms of Myotube Guidance and Attachment Site Selection ». *Developmental Cell* 7(1): 9-20.
65. Sens, Kristin L et al. 2010. « An Invasive Podosome-like Structure Promotes Fusion Pore Formation during Myoblast Fusion ». *The Journal of cell biology* 191(5): 1013-27.
66. Sweetman, Dylan et al. 2008. « The Migration of Paraxial and Lateral Plate Mesoderm Cells Emerging from the Late Primitive Streak Is Controlled by Different Wnt Signals ». *BMC developmental biology* 8: 63-63.
67. Tajbakhsh, S. et al. 1998. « Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5 ». *Development* 125(21): 4155.
68. Vorbrüggen, G, et H Jäckle. 1997. « Epidermal Muscle Attachment Site-Specific Target Gene Expression and Interference with Myotube Guidance in Response to Ectopic Stripe Expression in the Developing Drosophila Epidermis ». *Proceedings of the National Academy of Sciences of the United States of America* 94(16): 8606-11.
69. Wagner, J., C. Schmidt, W. Jr Nikowits, et B. Christ. 2000. « Compartmentalization of the Somite and Myogenesis in Chick Embryos Are Influenced by Wnt Expression. » *Developmental biology* 228(1): 86-94.

70. Wang, Hongyang et al. 2019. « Crystal Structure of the LUFS Domain of Human Single-Stranded DNA Binding Protein 2 (SSBP2). » *Protein science : a publication of the Protein Society* 28(4): 788-93.
71. Wang, Hongyang et al. 2020. « Crystal Structure of Human LDB1 in Complex with SSBP2 ». *Proceedings of the National Academy of Sciences of the United States of America* 117(2): 1042-48.
72. Wang, Y. et al. 2010. « SSBP2 Is an in Vivo Tumor Suppressor and Regulator of LDB1 Stability. » *Oncogene* 29(21): 3044-53.
73. Williams, Jessica, Nathan G. Boin, Juliana M. Valera, et Aaron N. Johnson. 2015. « Noncanonical roles for Tropomyosin during myogenesis ». *Development* 142(19): 3440.
74. Wright, Theodore R. F. 1960. « The phenogenetics of the embryonic mutant lethal myospheroid, in Drosophila melanogaster ». *Journal of Experimental Zoology* 143(1): 77-99.

SUPPLEMENTARY DATA:



Supplementary figure 1.
Ssdp has four regions with highly conserved amino acid residues when aligned against human SSBPs.
(A) The first 80 residues of the Ssdp protein aligned with the conserved LUFS domain of the LEUNIG protein in *Arabidopsis thaliana* and canonical isoforms of the human homologues, SSBP2, SSBP3 and SSBP4. The LUFS domain that is highlighted with a cyan rectangle permits the formation of a complex with Ssdp partner, Chi (LDB1 in humans). Within it is a LisH domain (green rectangle) and another conserved domain (orange rectangle) of unknown function. **(B-D)** Apart from the already identified LUFS domain, the alignment unveils three other regions with blocks of highly conserved amino acid residues. (B) is part of a proline-rich stretch that was identified as being essential for fore and midbrain development in mice.



Supplementary figure 2.
Ssdp protein expression as revealed by a Ssdp-Gal4 driven UASAct5CGFP.
(A-A'') Low levels of GFP can be detected in the dorsal muscles by stage 12. **(B-B'')** All muscles express GFP by stage 13. **(C-C'')** Expression is much stronger at early stage 16 with marked GFP expression in the Lms+ LT and Slou+ DT1 muscles. **(D-D'')** By late stage 16, a high GFP signal is detected in the sub-epidermal chordotonal organ (arrowhead) that acts as a proprioceptor and is situated just above the LT muscles. **(E-E'')** RNA FISH of *Ssdp* transcripts reveals similar high expression levels of mRNA in this region (arrowhead) above the LT muscles.

3.2.3.3. Additional observations in *Ssdp* mutants

Twinstar (Tsr) is an actin binding protein involved in actin depolymerization. It is required for sarcomerogenesis (Balakrishnan et al., 2020). In WT embryos, its expression starts much before the initiation of sarcomere formation similar to other sarcomeric proteins such as Tm2. It could thus either have unknown non sarcomeric roles or could be involved in preparing muscle cells for sarcomerogenesis. I examined the expression of its mRNA by *in situ* hybridization in *Ssdp* mutants to test if it is deregulated. I used the same methods for probe generation and hybridization as mentioned in the 'Materials and Methods' section of the publication preprint. The following primers were used:

tsr: 5'-AAGGCTTCTGGTGTAACTGTG-3' and 5'-CGGGACACCACGACATAAGG-3'

Its mRNA localizes to punctae distributed along the length of LT muscles in WT embryos (Figure 41A-A''). *tsr* mRNA is expressed in somatic muscles in *Ssdp* mutants, but is highly reduced (Figure 41B-B'') and appears to be concentrated at irregular locations, especially at muscle extremities unlike the lengthwise distribution observed in WT LT muscles. Some punctae are evident in muscles reduced to blobs. It is unclear if this deregulated *tsr* distribution is due to the affected cytoskeletal and MT elements in these mutants or vice versa.

3.2.4. Analysis of potential Wg dependent CRMs in putative enhancers of iTFs

Since the downstream Wnt effector, dTCF appears to play specific roles similar to Ssdp (Section 3.2.3.2), I performed an *in silico* analysis to identify potential enhancer regions. The aim was to validate these putative enhancers by cloning them to generate GFP-tagged enhancer reporter lines and test if mutating dTCF sites affected their expression patterns.

3.2.4.1. Bioinfomatic analysis of putative enhancers

dTCF binding sites were previously identified in early stage embryos by genome-wide ChIP-on-chip analysis (Junion et al. 2012). In addition, dTCF is known to bind a short CT-rich sequence (van de Wetering et al. 1997) and consensus motifs are available in the public JASPAR database. Peaks with high scores were recovered from (Junion et al. 2012) and

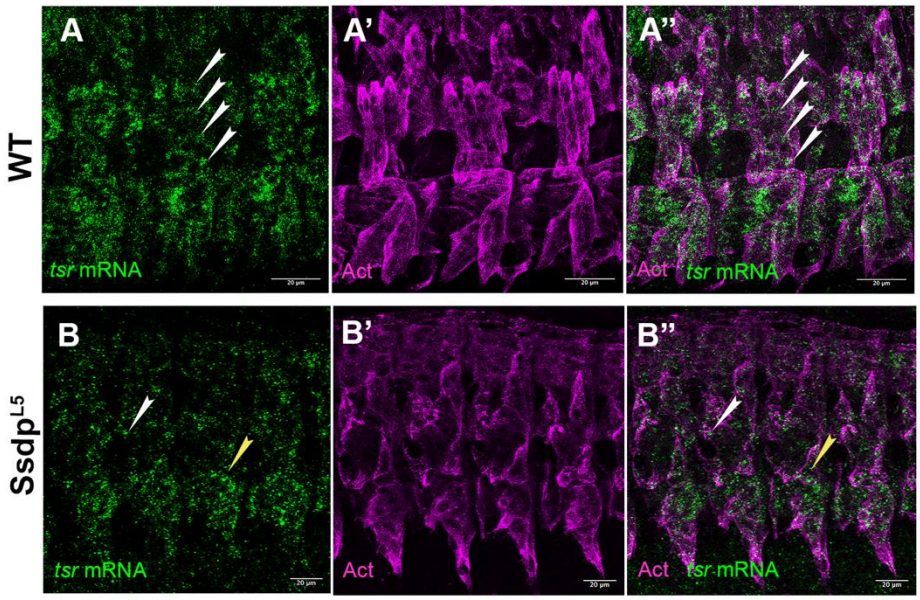


Figure 41. *tsr* expression is deregulated in stage 16 *Ssdp* mutant embryos.
(A-A") *in situ* hybridization against the mRNA of the actin binding partner Tsr reveals muscle expression in stage 16 WT embryos. Expression is concentrated in punctae along the length of the lateral transverse (LT) muscles (arrowheads in A, A''). **(B-B")** In $Ssdp^{L5}$ mutants, *tsr* mRNA is concentrated in fewer punctae, especially at muscle ends when they grow longitudinally (white arrowhead in B, B''). It displays a one-sided polarized expression in muscles reduced to blobs (yellow arrowhead in B, B'').

mapped to nearest genes. This study looked at early to mid stages of embryonic development (0-8h) and it is, thus, highly possible that dTCF site occupancy is not the same during other developmental stages. I also identified putative dTCF targets by providing the dTCF consensus motifs as input to the FIMO software in the MEME suite. I noticed, however, that the longer consensus motif in JASPAR was generating many hits due to nucleobase redundancy at multiple positions that varied largely from the shorter motif and could result in more false positives. So, I used the more precise short motif defined by (van de Wetering et al. 1997), CCTTTGA/TA/T, that matches the binding motif of the mammalian dTCF orthologues, TCF-1 and LEF-1 (van de Wetering et al. 1993). In addition, I performed a transcription factor binding site search using the Genomatix Software Suite with a core similarity of >= 0.85 using an optimized matrix threshold to reduce false positives (Cartharius et al. 2005).

A large number of putative dTCF targets were identified by this combined analysis. So, I tried to select those that were highly conserved. I decided to focus on the two LT muscle iTFs, Ap and Mid since multiple tools detected putative dTCF sites either upstream of their TSS or in introns or the 5'UTR. *apME680* is a previously identified *ap* enhancer present in a long intron that is specifically expressed in the LT muscles (Capovilla, Kambris, et Botas 2001). Genomatix detected multiple putative dTCF sites in this region, one of which was also detected using the precise sequence provided to FIMO (Figure 42A). The study by Junion et al. in early-to-mid stage embryos detected peaks only in the 5'UTR of *ap*. Since *apME680* is a known muscle enhancer, I decided to analyze this further. The putative site identified by FIMO and Genomatix (TGATTTGATGTTG) is highly conserved among *Drosophila* species (Figure 42B). Another sequence identified by Genomatix (CCCTTTGATCGAT) closely matches the 'canonical' dTCF binding site identified by van de Wetering et al. Both are on the template strand in the *ap* intron. This second site, however, is not detected as a conserved region by the UCSC phastCons algorithm. This region also harbors a conserved Runt binding site (L. Wang, Brugge, et Janes 2011). Runt is a conserved TF and an ortholog of the vertebrate RUNX1. The Wnt enhanceosome complex has been shown to be composed of other cofactors including Runt (Fiedler et al. 2015).

During embryonic and wing development, Wg has been shown to be negatively regulated by the T-box transcription factor Midline (Mid) (Buescher et al. 2004; Fu et al. 2016), but it is unknown if Mid is regulated by the Wg pathway in specific tissues or specific muscles. This is

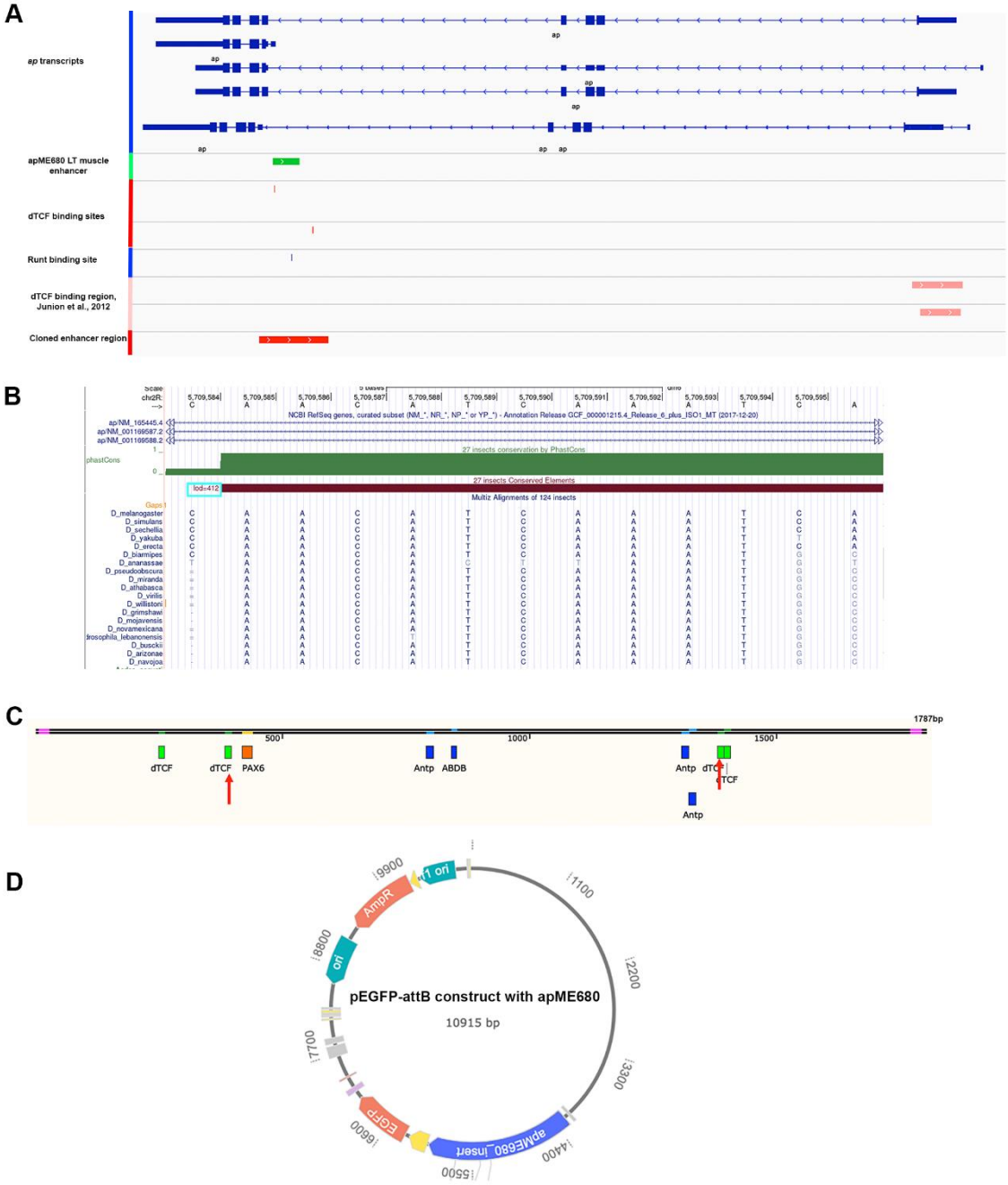


Figure 42. *in silico* identification of dTCF binding sites and cloning of a putative *ap* enhancer. **(A)** An LT muscle specific *ap* enhancer, *apME680* (green rectangle), has been previously identified in a long *ap* intron (Capovilla, Kambris, et Botas 2001). A bioinformatic analysis identified a putative dTCF binding site in this enhancer. Another is present 5' to this enhancer (red bars). A conserved Runt binding site was also identified in the *apME680* enhancer (blue bar). A genome wide ChIP-on-chip analysis of 0-8h embryos, on the other hand, did not detect significant peaks in this region, but detected high scoring peaks in the 5'UTR (peach rectangles). A 1787bp region harboring the entire *apME680* region along with the two dTCF sites was chosen for cloning. **(B)** The dTCF site within *apME680* is highly conserved among *Drosophila* species as evidenced using the UCSC Genome Browser. **(C)** The MatInspector software of the Genomatix Software Suite identified putative binding sites for Hox genes, Pax6 or Eyeless (Ey) and dTCF in this chosen putative enhancer. Putative dTCF sites that were mutated are indicated by red arrows. **(D)** The structure of the vector construct with the putative enhancer designed using the GenSmart Design web service is displayed.

a possibility considering its dual role as a cardiac and muscle TF and signaling pathways could very well be defining factors. I performed an *in silico* analysis of a 5KB region upstream of the *mid* TSS (Figure 43). The ChIP-on-chip analysis by Junion et al. identified two dTCF peaks in this region. Genomatix identified an approximately 2KB region upstream of *mid* that is rich in putative binding sites for Hox genes, somatic muscle TFs such as Twist (Twi) and Chorion factor 2 (Cf2) as well as many LT muscle iTFs including Ap, Kr and Caup. This region contains a dTCF peak identified by Junion et al. It also contains putative binding sites for the cardiac TF, Tinman (Tin). It has blocks of DNA sequences that are highly conserved among *Drosophila* species. One putative dTCF site (TCGTTTGACTTTC) on the coding strand is highly conserved among *Drosophila* species. (Figure 43B). Genomatix identified a second region upstream of the first region that contains multiple Slou binding sites. The UCSC Genome Browser identifies blocks of conserved DNA sequences within this region. Junion et al.'s study identified one dTCF peak within this region and Genomatix identified a putative dTCF site within this peak (CGCTTTGATAAAT) that is highly conserved among *Drosophila* species (Figure 44A, B). A highly conserved putative Runt binding site is present between this and another dTCF ChIP-on-chip peak.

3.2.4.2. Generation of putative Wg-dependent iTF enhancer-GFP reporter lines

After the identification of putative enhancers and putative dTCF and Runt binding sites that they harbor, I identified regions in each enhancer that could be cloned. The criterion was to be able to mutate at least two dTCF/Runt sites. I identified two sites within each chosen region that could be mutated such that overlapping PCR fragments could be generated using mutated primers that could subsequently be joined by Gibson Assembly. Following PCR amplification using the Takara PrimeStar GXL DNA polymerase, I cloned DNA fragments into the p-EGFP-attB vector (1424, Drosophila Genomics Resource Center) by Gibson assembly (NEBuilder ® HiFi DNA Assembly kit) and cloning was validated by sequencing. The constructs were injected into y1 w67c23;P{Cary}attP2 flies (BestGene Inc.) to generate transgenic lines by PhiC31 integrase-mediated transgenesis carrying the GFP tagged enhancer on chromosome 3.

I chose to mutate the 'TTT' that is recurring and highly conserved in dTCF motifs. For *ap*, I chose a 1787bp region that covers the *apME680* enhancer (Figure 42A) with two dTCF sites

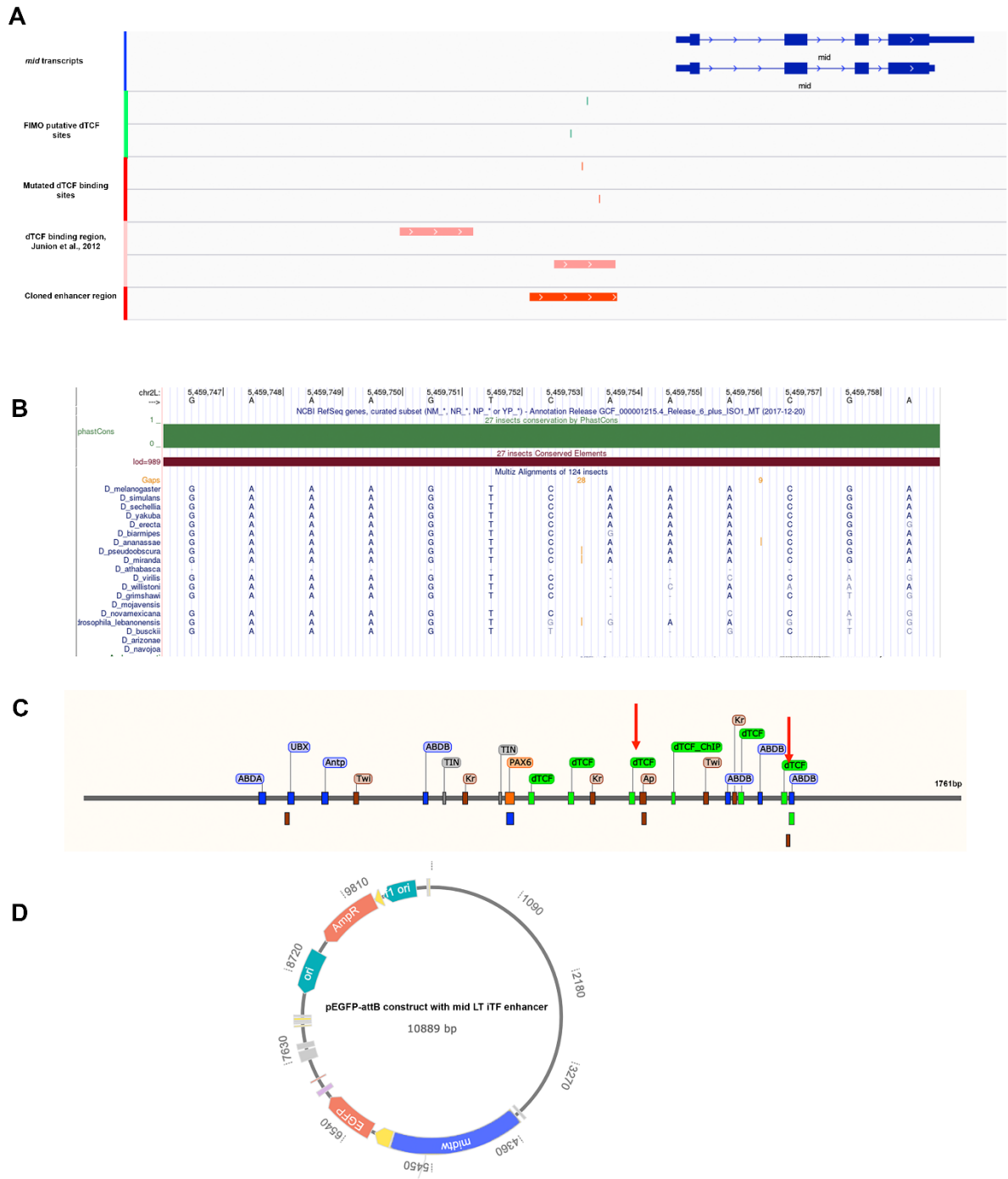


Figure 43. *in silico* identification of dTCF binding sites and cloning of a putative *mid* enhancer harboring potential binding sites for LT iTFs.
(A) Multiple putative dTCF binding sites were identified by FIMO and Genomatix (green and red bars respectively) within a ChIP-on-chip peak identified by Junion et al. (Junion et al. 2012) (peach rectangles). I chose a 1761bp region spanning this peak for cloning. **(B)** The dTCF site within the ChIP-on-chip peak is highly conserved among *Drosophila* species as revealed by UCSC Genome Browser. **(C)** The MatInspector software identified putative binding sites for multiple LT iTFs including Ap, Kr and Caup as well as Hox genes, Pax6, Twi and the cardiac TF, Tin in this chosen putative enhancer. Putative dTCF sites chosen for mutation are indicated by red arrows. **(D)** The structure of the vector construct with the putative enhancer designed using the GenSmart Design web service is displayed.

that could be mutated, one is the conserved sequence in *apME680* and the other is the site that closely resembles the 'canonical' dTCF site (Figure 42C). I designed a vector construct with this insert using the pEGFP-attB vector (Figure 42D). I performed a Gibson Assembly with the full non-mutated fragment as well as the mutated fragments. I subsequently verified correct incorporation of the non-mutated and mutated inserts into the vector by sequencing using forward and reverse primers contained in the vector a short distance from the insert. GFP reporter lines of this putative enhancer were generated.

For the mid fragment with Slou binding sites, I chose to clone a 2099bp fragment with a dTCF peak identified by Junion et al. harboring a putative Genomatix dTCF site. I chose to mutate this site along with the conserved Runt site (Figure 44C, D). GFP reporter lines of this putative enhancer were generated. These WT and mutant enhancer lines remain to be analyzed along with the *ap* enhancer lines.

For the *mid* enhancer containing putative LT iTF binding sites, I chose a 1761bp fragment. I chose the conserved dTCF site and had to choose a second non-conserved site to mutate due to the difficulty finding appropriate PCR primers in this fragment (Figure 43A, C and Figure 45C). The following PCR primers were used to generate enhancer fragments using a genomic DNA template:

midEn^{LTiTF}enhancer sequence: 5'-GCGTTCTCAGTGCAAACAACTG-3' and 5'-GCGACAAGGAAACTCGAAACAAC-3'.

The midEn^{LTiTF}dTCF^{-} enhancer sequence was split into 3 fragments and PCR amplified.

Fragment 1: 5'-GCGTTCTCAGTGCAAACAACTG-3' and 5'-GTCCTCGATAGACTTTCTCCTTG-3'

Fragment 2: 5'-CAAGGAGAAAGTCTATCGAGGAC-3' and 5'-CATTAACATATTCATAAGCATAACGGACG-3'

Fragment 3: 5'-CGTCCGTTATGCTTATGAATATGTTAATG-3' and 5'-GCGACAAGGAAACTCGAAACAAC-3'

I designed a vector construct (Figure 43D), cloned it and verified the non-mutated and mutated constructs by sequencing. GFP reporter lines were then generated as mentioned previously. I analyzed the non-mutated and mutated GFP reporter lines for this enhancer. Flies successfully incorporated the full length as well as mutated constructs that I will refer to as midEn^{LTiTFs} and

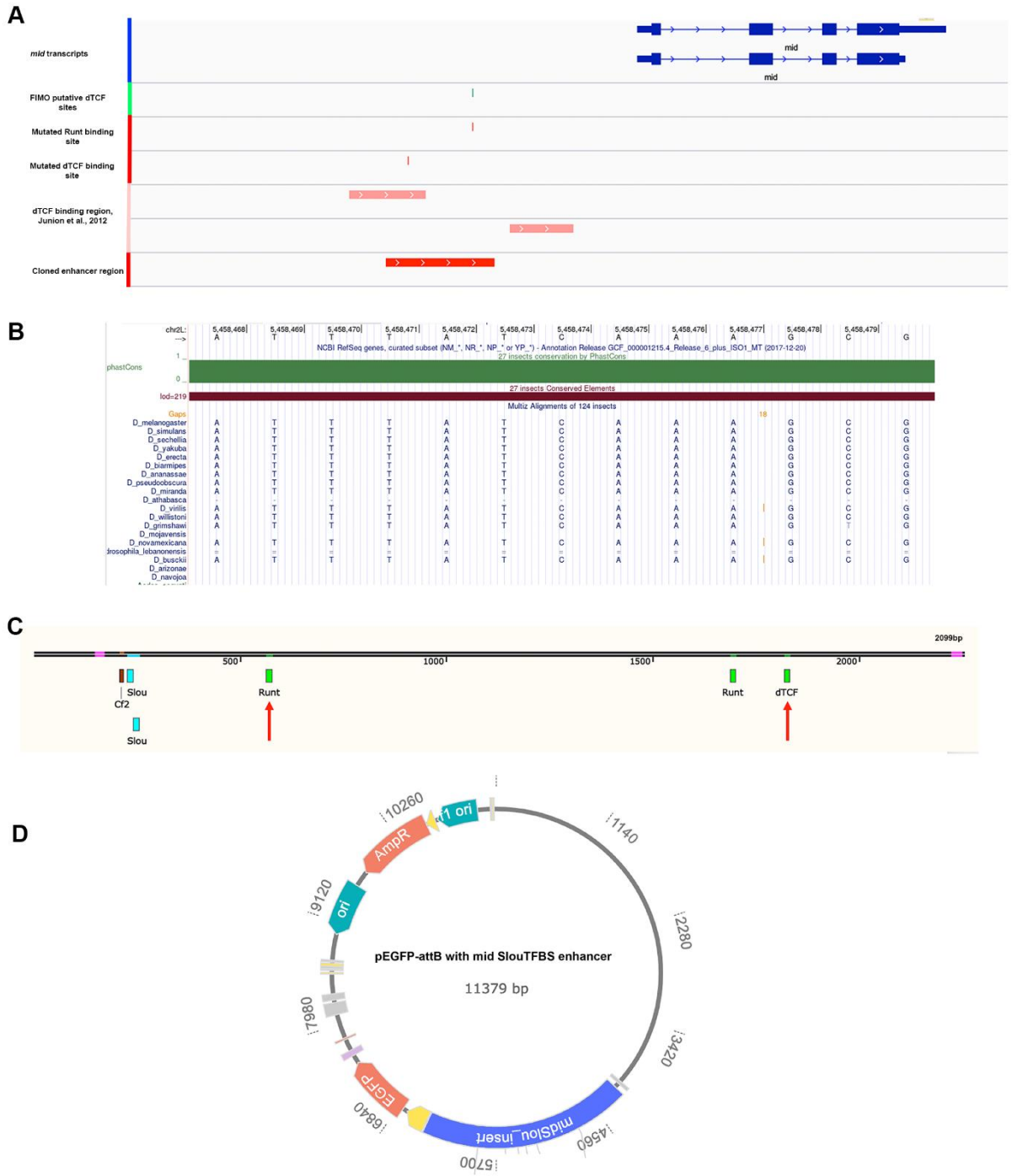


Figure 44. *in silico* identification of dTCF binding sites and cloning of a putative *mid* enhancer harboring potential binding sites for Slou.
(A) A putative dTCF binding site was identified by FIMO close to a Runt binding site (green and red bars respectively) between two ChIP-on-chip dTCF peaks identified by Junion et al. (Junion et al. 2012) (peach rectangles). I chose a 2099bp region spanning one of the ChIP-on-chip peaks with a putative dTCF site identified by Genomatix for cloning. **(B)** The dTCF site within the ChIP-on-chip peak is highly conserved among *Drosophila* species. **(C)** The MatInspector software identified multiple putative binding sites for Slou, dTCF and Runt in this chosen putative enhancer. Putative sites that were mutated are indicated by red arrows. **(D)** The structure of the vector construct with the putative enhancer designed using the GenSmart Design web service is displayed.

midEn^{LTiTFs}dTCF^{-} respectively (Figure 45). The Mid protein displays a low, diffused expression in cardioblasts in *Ssdp* mutants (Figure 45A-C'). The full-length GFP-enhancer is expressed only in cardioblasts similar to the Mid protein as is midEn^{LTiTFs}dTCF^{-}. While midEn^{LTiTFs} displays a strong and even expression during stages 12 and 16 (Figure 4E-E'', G-G''), the midEn^{LTiTFs}dTCF^{-} with mutated dTCF binding sites clearly displays a much weaker diffused expression at stage 12 (Figure 45F-F''). Expression appears normal at stage 16, although it is not as strong as in the WT (Figure 45H-H''). This indicates a dependence on the mutated dTCF sites during stage 12 and not during stage 16 for this enhancer.

It remains to be seen if the other putative *mid* and *ap* enhancers display differences in expression patterns of their respective non-mutated and mutated GFP reporters.

3.2.5. Complementary analyses

3.2.5.1. Development of a protocol for the extraction of nuclei from a muscle subset for 10X genomics snRNA-Seq analysis

The TRAP analysis helped identify differences between the Lms and Slou muscle subsets. However, the Lms subset consists of 4 LT muscles, each with its distinct identity. In all probability, there are extremely subtle differences between these muscles. A single cell or single nucleus RNA-Seq analysis has the potential to help distinguish these subtle differences unless the differences are miniscule. scRNA-Seq experiments on muscle cells are complicated due to the large size of each muscle fiber leading to difficulty in FACSing them. snRNA-Seq provides an alternative approach. 10X Genomics Chromium is a widely used scRNA-Seq method that has proven to be successful for snRNA-Seq experiments on frozen samples (Wu et al. 2019) and can be performed on as few as 500 good quality nuclei. It is far cheaper and less time-consuming than Smart-seq2 and can use a 3' end sequencing approach instead of the Smart-seq2 full-length approach. It also permits all analysis to be performed inside a single tube per sample. 10X Genomics Chromium uses a droplet-based technology where each droplet or Gel Beads-in-emulsion (GEM) is optimized to contain a single cell/nucleus. Each GEM is dissolved followed by lysis and cDNA generation and amplification. During this, all poly(A) RNA of a cell/nucleus is uniquely barcoded to distinguish between cells/nuclei and a Unique Molecular Identifier (UMI) is appended to each poly(A) RNA in a cell/nucleus in order to distinguish it

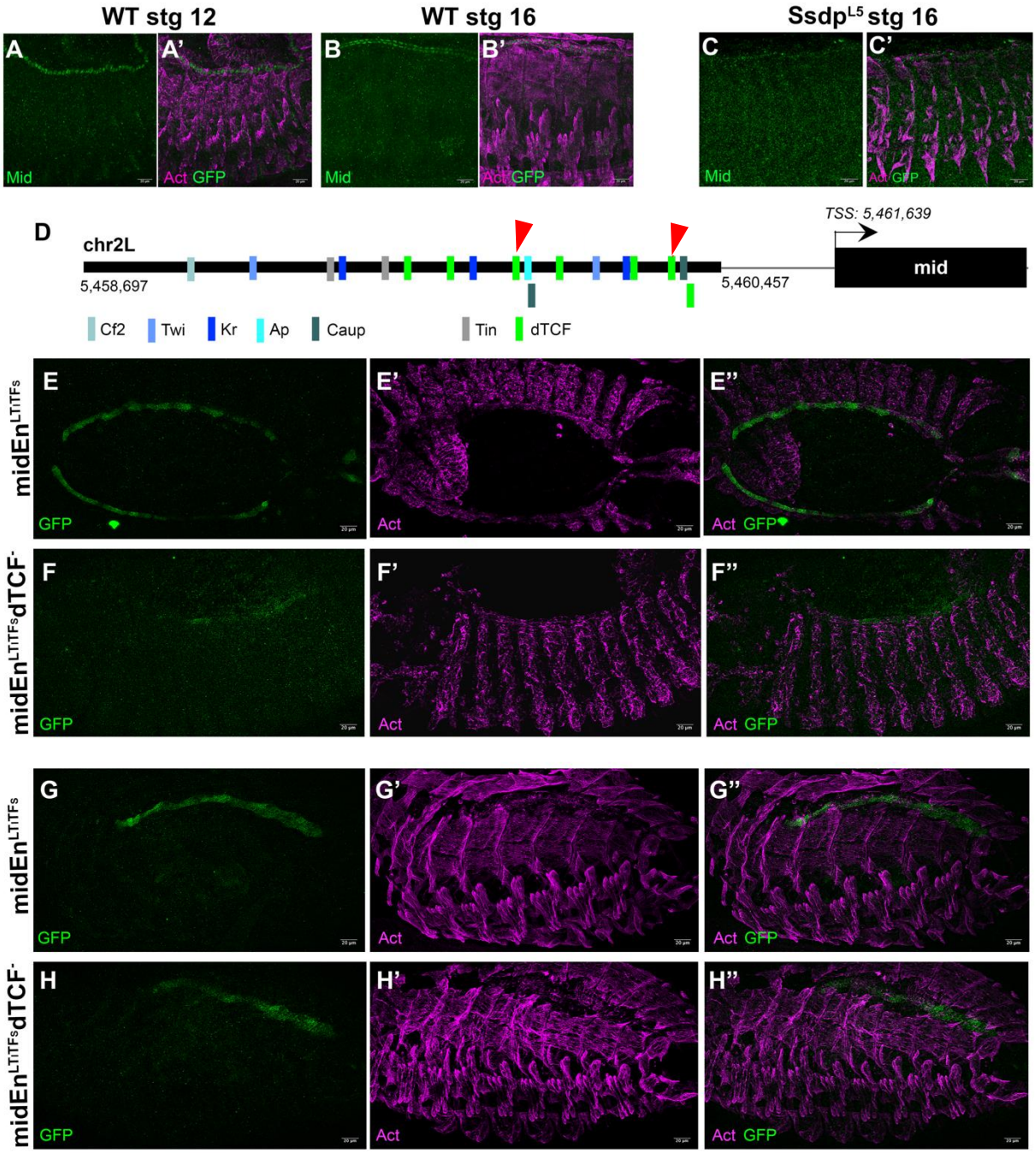


Figure 45. Analysis of the expression patterns of the non-mutated and mutated *mid* enhancer GFP reporters containing LT iTF binding sites.
(A-A') An antibody staining against Mid in stage 12 (A-A') WT embryos reveals a high expression in cardioblasts. (**B-B')** Similar high levels of expression are maintained in the cardioblasts of stage 16 WT embryos as revealed by an anti-Mid antibody. **(C-C')** An antibody staining against Mid in stage 16 *Ssdp* mutants reveals a low, diffused expression. **(D)** Schematic diagram of the *mid* enhancer with positional information along with the *mid* TSS, TF binding sites and mutated dTCF sites (red arrowheads). **(E-E'')** Antibody staining against GFP reveals that the midEn^{LTiTFs} enhancer displays a strong, uniform expression pattern at stage 12. **(F-F'')** The mutated midEn^{LTiTFs}dTCF^{-} enhancer displays a weak, diffused expression at stage 12 resembling late stage *Ssdp* mutants in (C). **(G-G'')** The midEn^{LTiTFs} enhancer continues to display high, evenly distributed expression at stage 16. **(H-H'')** The mutated midEn^{LTiTFs}dTCF^{-} enhancer displays a slightly diffused expression, but this is not as remarkably low as that observed during stage 12.

from others. cDNA libraries are generated by appending sequencing primers and sequenced. This method detects mRNAs and lncRNAs with poly(A) tails (X. Wang et al. 2021).

The apME-NLS::dsRed line marks Ap^{+} LT muscle nuclei (Folker, Schulman, et Baylies 2014). It also marks a few cells in the VNC (Figure 46A). I took advantage of its largely LT-specific expression to develop a protocol to extract good quality myonuclei from the LT muscle subset. This protocol is inspired from the 10X Genomics protocol (https://support.10xgenomics.com/single-cell-gene-expression/sample-prep/doc/demonstrated-protocol-isolation-of-nuclei-for-single-cell-rna-sequencing) that has been demonstrated to yield successful results for the extraction of myonuclei from mouse muscles (Dos Santos et al. 2020) and from the BiTS-ChIP protocol to isolate nuclei for chromatin immunoprecipitation in *Drosophila* (Bonn et al. 2012). Some buffers and procedures in these protocols are retained while others have been removed or modified as necessary.

3.2.5.1.1. Ingredients

Buffers should be freshly prepared before each experiment.
1. 0.26% bleach.
2. Lysis buffer: to prepare the lysis buffer, mix 10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3mM MgCl2, 0.1% Nonidet P40 in nuclease free water.
4. Wash buffer: to prepare the wash buffer, add 2% (wt/vol) BSA to 1ml PBS and 0,2U/µl RNase inhibitor (Roche). Allow it to settle by placing at 4 degrees for 10 minutes. Filter the wash buffer using a 40µm filter before use to avoid any BSA aggregates in the nuclei resuspension solution that is passed for FACS.

3.2.5.1.2. Cryopreservation procedure:

1. Dechorionate *Drosophila* embryos in bleach for 3 minutes.
2. Allow them to dry completely.
3. Flash freeze them in liquid nitrogen for a duration of 10 seconds.
4. Remove immediately and store them at -80°C.

3.2.6.1.3. Procedure for the extraction of nuclei

1. Always ensure that centrifugation and incubation steps are performed at 4°C on ice. Place all buffers on ice.

2. Thaw approximately 0.5g of previously cryopreserved embryos briefly in a 37°C water bath for about 9 seconds.
3. Transfer the embryos into a previously chilled 2ml Dounce homogenizer on ice.
4. Dounce about 15 times by using delicate, but firm strokes of a chilled loose pestle followed by 10 times with a chilled tight pestle until a homogenous solution is formed.
5. Pre-rinse a 70µm filter with 0.5ml chilled wash buffer.
6. Filter the lysate with the pre-rinsed 70µm filter, then a 40µm filter.
7. Filter the lysate through 20µm Sefar Nitex membrane filter.
8. Spin at 4°C at 5500 rpm (or around 3380g) for 5 min to pellet the nuclei and carefully discard the murky supernatant ensuring not to discard the pellet at the bottom.
9. Resuspend the pellet in 1 ml of wash buffer in an Eppendorf tube.
10. Pellet the nuclei again at 4°C at 5500 rpm for 5 min.
11. Resuspend the nuclei in 1 ml of wash buffer and transfer them into chilled FACS tubes by filtering through 40µm filters using a pipette.
12. Pass them to a FACS machine set to retrieve nuclei into PBS solution at 4°C.

Ensure that the nuclei are FACS sorted as rapidly as possible within a few minutes to ensure good quality. Following FACS, the nuclei should be loaded immediately onto a 10X Genomics chip. Performing the procedure in a dark room at 4°C might yield better results by preserving fluorescence, although this has not been verified.

3.2.5.2. Verification of the specificity and quality of extracted nuclei

I used WT nuclei extracted from the whole embryo and not carrying a fluorescent marker as a negative control to verify if it was possible to FACS nuclei extracted using the above protocol and to retrieve a sufficient quantity of viable myonuclei. Stephanie Maupetit from our FACS platform helped me get trained on our FACS machine to be able to perform the experiment autonomously. We set up the parameters to FACS nuclei of the correct size and verified that it detected fluorescence only in nuclei extracted from the apME-NLS::dsRed population and not in the WT population (Figure 46B, B'). In order to quickly verify the quality of FACS sorted nuclei, I directly FACS sorted nuclei onto slides coated with a mounting medium containing DAPI (Figure 46C, C'). I quickly verified the quality of nuclei under the microscope and they appeared round, healthy and viable (Figure 46D-D''). I then FACS sorted dsRed nuclei into tubes containing around 100µl of PBS for 1 hour. I fixed WT nuclei in 4% PFA and verified

them under the microscope since there were too few dsRed+ nuclei to perform this verification. The quality of nuclei did not appear to be as good as a short-term sort onto slides, but some viable-looking nuclei were observed. This yielded around 65000 dsRed+ nuclei (Figure 46E, E'). I tried a half hour FACS that yielded roughly half as many nuclei.

This protocol needs to be further optimized to reduce time and tested to check if the nuclei are viable enough for an snRNA-Seq. Since individual LT muscle types probably represent rare cell populations, analyzing a sufficient number of nuclei is important. In addition, an snRNA-Seq on the entire muscle population has the potential to provide much more interesting information regarding differential gene expression between muscle subsets. Given that this is nuclear RNA, it has its limitations. A large proportion will be pre-mRNA. An analysis of cytoplasmic mRNA might provide much more insights. On the other hand, this has the potential to provide interesting insights into non-coding RNA expression patterns.

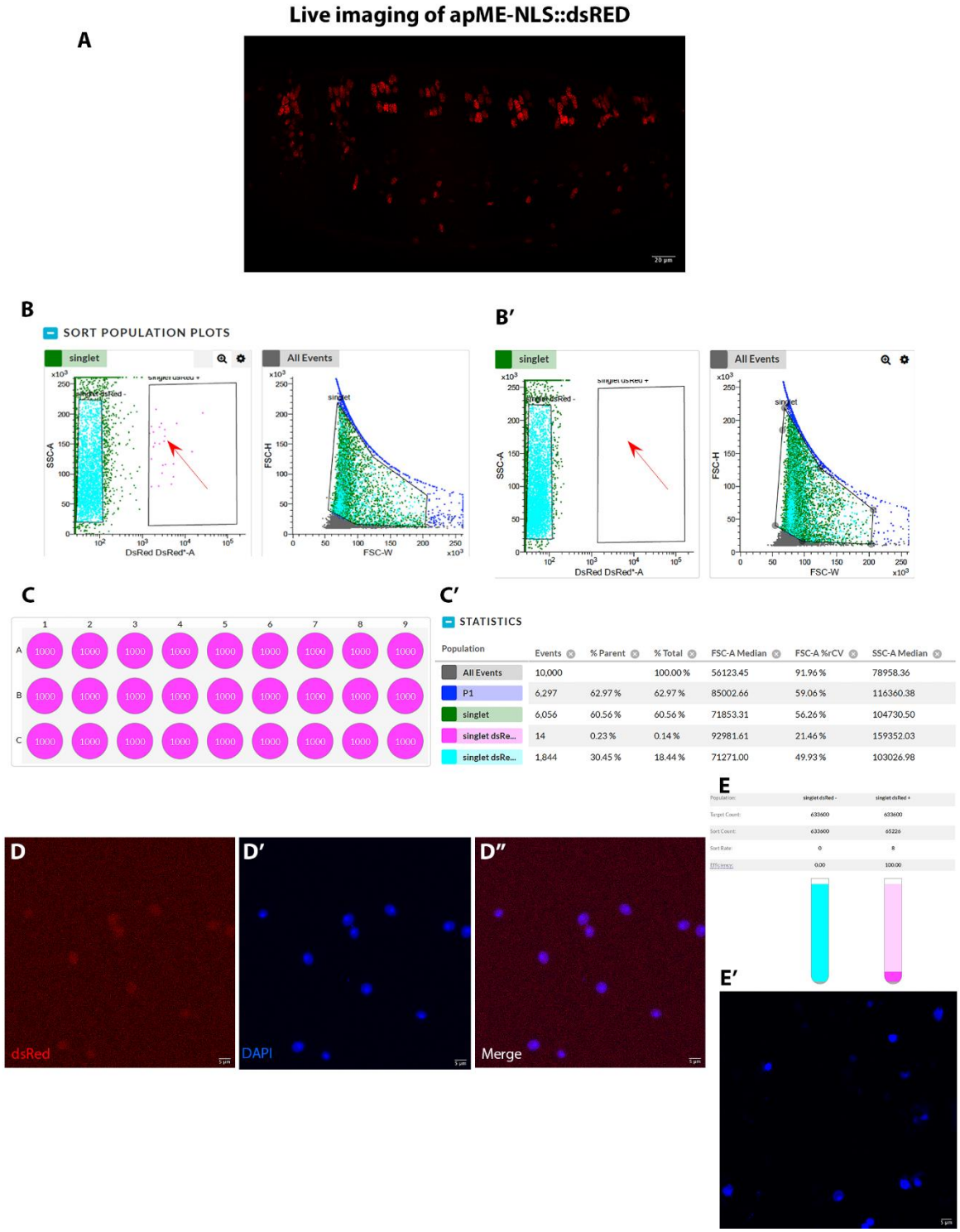


Figure 46. Extraction and FACS sorting of dsRed+ LT myonuclei.
(A) A live imaging of apME-NLS::dsRed shows a strong signal in LT muscles and in select cells of the VNC. **(B)** A FACS sort of nuclei extracted from apME-NLS::dsRed embryos successfully detects dsRed florescence (red arrow). **(B')** A FACS sort of nuclei extracted from WT embryos as a negative control using the same FACS parameters does not detect fluorescence and almost no nuclei are sorted in the dsRed bin as expected. **(C)** A quick sort of dsRed+ nuclei directly onto a slide coated with a mounting medium containing DAPI. **(C')** The percentage of dsRed+ nuclei is extremely insignificant compared to dsRed- embryos as expected. **(D-D'')** A microscopic view of the slide-sorted nuclei reveals well shaped, viable looking nuclei. **(E)** A FACS sort for a duration of one hour results in around 65,000 dsRed+ nuclei. **(E')** The WT nuclei that were sorted into the dsRed- tube were precipitated and fixed in 4% PFA. There still appear to be some rounded, viable looking nuclei, but the quality appears reduced.

CHAPTER 4 - Discussion

4.1. Ssdp and its role in the establishment of muscle identity

This study identifies several new genes whose mRNA under translation are differentially expressed in *Drosophila* somatic muscles including *Ssdp*. A central role for Ssdp in *Drosophila* myogenesis was identified by bioinformatic analysis and subsequently validated for the first time. It has gained attention in recent years as a component of the evolutionarily conserved ChiLS complex involved in transducing Wnt signaling via the canonical pathway. In humans, *SSBP3* mis-splicing with the retention of exon 6 in conjunction with its downregulation has been observed in the skeletal muscles of myotonic dystrophy type 1 and 2 as well as neuromuscular disorder (NMD) patients (Bachinski et al. 2014), but its role in myogenesis has not received any attention. The stage-specific expression of the long and short *Ssdp* isoforms noted in my study indicates a role for the long *Ssdp* isoforms during mid to late developmental stages in embryonic somatic muscles. A BLAST of the cDNA clone (GM14473) (Rubin et al. 2000) used to generate *in situ* probes in the BDGP database maps it to *Ssdp-RC*, a short isoform, with the highest score and a contiguous, almost perfect match except for 9bp. BDGP does not note any *Ssdp* expression in the somatic muscles during mid to late stages with their probes. This again reinforces my observation of significant expression of the long *Ssdp* isoforms during these stages. Given that $Ssdp^{L5}$ and $Ssdp^{L7}$ are zygotic mutants, the generation of RNAi or somatic CRISPR lines targeting different isoforms would help elaborate on the consequences of this differential *Ssdp* isoform expression.

In vertebrates, *Ssdp* has been shown to play a role in tissue differentiation (Liang, Samanta, et Nagarajan 2005; J. Liu et al. 2016) including the differentiation of specific axons in zebrafish (Zhong et al. 2011). My study shows that a loss of *Ssdp* expression leads to severe defects in the establishment of the muscle identity program. The expression of iTFs in muscles at expected locations indicates a largely successful FC specification program followed by the initiation of muscle differentiation. However, the low levels of iTF expression as well as myoblast fusion and attachment defects observed suggest an impaired muscle identity establishment for most muscles. The observation of ventrally projecting VO4-6 muscles begs the question of iTF mis-

expression in these muscles given that this phenotype has earlier been observed for ectopic Vg/Sd expression (H. Deng et al. 2009).

The downregulation of muscle iTFs is probably a result of the deregulation of signaling processes and tissue cross talk as underlined by the downregulation of epidermal Wg. Thus, Ssdp either directly or indirectly influences Wg expression. An analysis of potential dTCF and CT-rich elements upstream of Wg as well as in its introns and 5'UTR could provide clues about potential direct binding by Ssdp and its cofactors. It is difficult to distinguish between WT and *Ssdp* zygotic mutant embryos at very early stages due to a difficulty in visualizing the GFP-tagged balancer in order to distinguish GFP^- *Ssdp* homozygotes during early stages, which would have helped detect early Wg expression patterns. However, since Wg is essential for initial mesoderm specification and the specification of VA1/2 and since these processes are unaffected in *Ssdp* mutants, it appears that the zygotic expression of Ssdp is not necessary for initial Wg expression, but for Wg maintenance at later stages. The partial overlap of *Ssdp* mutant phenotypes with $dTCF^{DN}$, wg^{ts} mutants and $Ssdp^{L5}$/+;wg^{ts}/+ transheterozygotes suggests stage and muscle subset specific interactions between the Wg pathway and Ssdp. A more detailed analysis of wg^{ts} mutants and $Ssdp^{L5}$/+;wg^{ts}/+ transheterozygotes at different timepoints could help shed more light on the Ssdp-Wg dynamics. If $Ssdp^{L5}$/+;wg^{ts}/+ transheterozygotes reproduce *Ssdp* or wg^{ts} mutant phenotypes, it would mean that they are the result of the loss of genetic interaction between these two factors at the specific stage analyzed and would help identify specific phenotypic properties that are affected by this interaction.

While its role in myogenesis has been confirmed, the significance of its upregulation in the Lms subset versus the Slou subset as well as versus the Duf population remains to be elucidated. This study stopped short of quantifying mRNA expression in the two muscle subsets due to the inability to correctly segment Slou positive muscles using the Slou-Gal4;UASRpL10a line. Segmentation is the process of partitioning a stacked image into identifiable regions that permits data analysis restricted to these regions. This failure to segment appropriately was due to the presence of GFP signals of varying intensities within the image. The Lms-Gal4;UASLifeActGFP line was segmentable due to a uniform intensity distribution that helped isolate these muscles so that *Ssdp* RNA FISH probe dots in the LT muscles alone could be quantified. A Slou-Gal4;UASLifeActGFP line would have to be generated to be able to compare the number of detected RNA spots in Lms+ versus Slou+ muscles, although this might not necessarily indicate differences at the translational level like the TRAP data.

4.2. Ssdp and its cofactors and interactors

Preliminary observations of *Ssdp*^{L5}/+;*ap*^{UG035}/+ transheterozygotes present no phenotypes, suggesting that a single WT copy of each gene is sufficient for WT phenotypes in this transheterozygous context or that the duplication phenotypes observed in *ap*^{UG035}/+;*Chi*$^{e5.5}$/+ are not dependent on the ChiLS complex. Although Ssdp lacks an NLS and although its Chi-binding domain is necessary for its strong nuclear localization, there appears to be very weak nuclear localization in a few myonuclei expressing a form of Ssdp lacking its Chi-binding domain (Annexe 2). So, it is also possible that it can localize to nuclei by itself or is assisted by other cofactors. A study found that the localization of endogenous vertebrate SSBP2 to nuclei is probably mediated by LDB1-independent tyrosine phosphorylation in 293T cells (Fleisig et al. 2007). If the ChiLS complex is involved, these results concur with studies in *Xenopus* embryos where only extremely high or low expression of one factor with respect to other cofactors resulted in the disruption of stoichiometry and strong phenotypes (Agulnick et al. 1996; Castro et al. 2002). The significance of the upregulation of *Ssdp* in the LT subset with respect to the Slou subset needs to be analyzed further since *ap*^{UG035}/+;*Ssdp*^{L5}/+ transheterozygotes do not appear to present any phenotypes.

The switch towards a muscle duplication phenotype instead of missing muscles in *ap*^{UG035}/+;*Chi*$^{e5.5}$/+ and *ap*^{UG035}/+;*mid*^{1}/+ suggests a genetic interaction between *ap*, *mid* and *Chi*. These observations are in line with a parallel study in the team where mutants for the gene *Gel* that is significantly upregulated in the LT subset exhibit muscle duplications. Ectopic Ap (an Lhx2 orthologue) expression was sufficient to induce *Gel* expression (Bertin et al. 2021). Mid (a Tbx20 homologue) is possibly an additional factor influencing this phenotype by regulating fusion. The FlyBi consortium identified physical interactions of Lms with Mid, D and Sox21b. Since *D* and *Sox21b* are consistently among the significantly upregulated genes specific to the Lms population and given that Schnorrer et al.'s RNAi screen (Schnorrer et al. 2010) revealed sarcomere defects for RNAi against *D*, they are potential new LT iTFs. These observations help elaborate on the iTF code defining LT muscle identity (Figure 47).

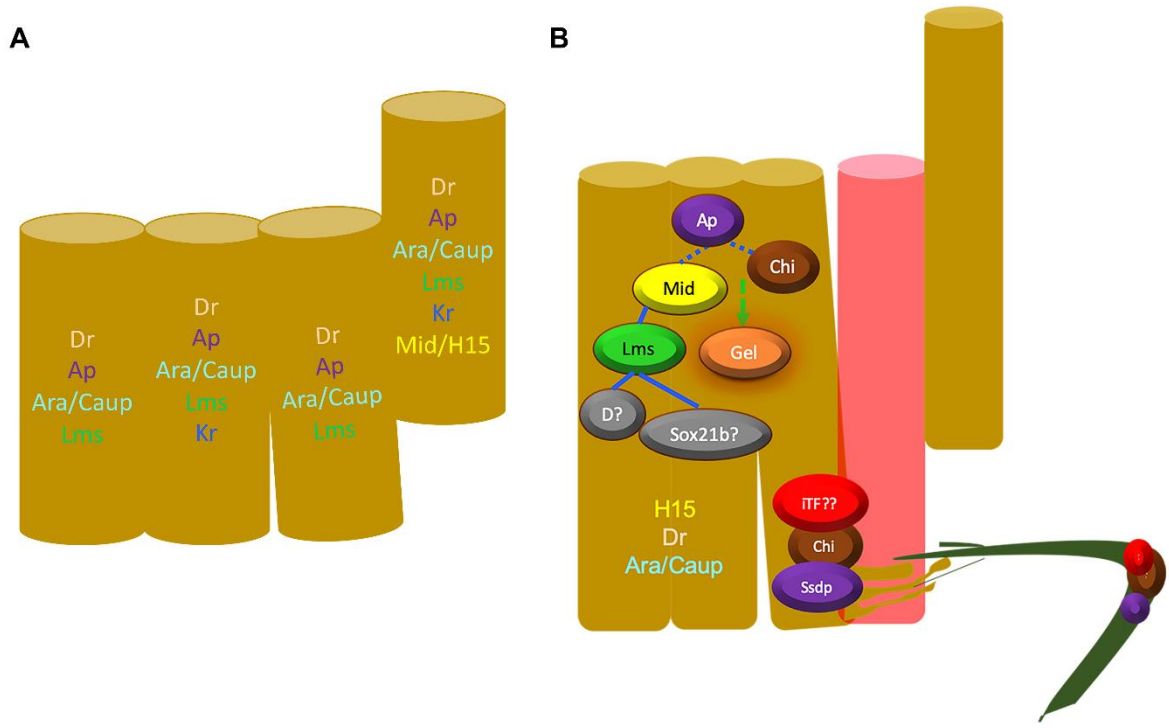


Figure 47. Identification of new players regulating LT muscle identity.
(A) The currently known iTFs contributing to the iTF code for individual LT muscles. **(B)** New factors regulating general LT muscle subset identity. The switch towards a muscle duplication phenotype (indicated by a red muscle) instead of missing muscles in ap^{UG035}/+;$Chi^{e5.5}$/+ and ap^{UG035}/+;mid^{1}/+ suggests a genetic interaction between *ap*, *mid* and *Chi*. This duplication phenotype is also observed in *Gel* mutants, which indicates that this is a potential downstream target of Ap and/or Mid. Thus, Chi interacts with Ap that in turn genetically interacts with Mid, potentially to regulate the number of fusion events by regulating the expression of realisator genes such as *Gel*. The relevance of the significant upregulation of *Ssdp* in this muscle subset with respect to the Slou subset remains to be elucidated. Its significant upregulation in the Lms subset indicates that it is a potential effector of LT muscle identity, but given that preliminary observations of ap^{UG035}/+;$Ssdp^{L5}$/+; display no phenotypes, this is probably either not via the ChiLS complex or one copy of each gene suffices for normal function via ChiLS. The FlyBi consortium identified a physical interaction of Lms with Mid, D and Sox21b. Since these are also among the significantly upregulated Lms specific genes in the TRAP dataset, it would be interesting to verify if D and Sox21b play a role in LT muscle identity.

Given these observations, it would be interesting to study the effects of overexpression of *Ssdp* in a *Chi* mutant context and vice versa. It would also be interesting to generate an anti-Ssdp antibody to visualize its endogenous expression pattern in WT and mutant conditions. This would potentially also open avenues to mass spectrometric (MS) analysis of its chromatin complexes to identify cofactors by RIME (Rapid immunoprecipitation mass spectrometry of endogenous proteins) (Mohammed et al. 2016). A parallel RIME using an anti-Chi antibody would help distinguish ChiLS from non-ChiLS complexes. The striking actin cytoskeleton and MT defects along with the involvement of the LisH domain present in Ssdp in regulating MT dynamics makes an analysis of Ssdp interactors, including non-chromatin complex interactors, by MS extremely pertinent. It would also be interesting to verify if there are LisH domain proteins among differentially regulated genes, although this need not necessarily be the case since it could be the availability and levels of Ssdp that control the dynamics of proteins that are not necessarily differentially expressed.

4.3. Ssdp and its targets

Liang et al. (Liang, Samanta, et Nagarajan 2005) observed a downregulation of c-Myc on inducible ectopic *SSBP2* expression in human Acute myelogenous leukemia (AML) cell lines that normally lack *SSBP2* expression, highlighting *SSBP2*'s role in differentiation. In contrast, my bioinformatic analysis detected an enrichment for putative modERN targets for the evolutionarily conserved Dm (or Myc) during T2 and T3 in the Duf+ global muscle population versus total embryonic mRNA among significantly upregulated genes when *Ssdp* is also upregulated. The vertebrate FUBP1 motif that binds far upstream elements (FUSE) of c-MYC is enriched among downregulated genes in the Lms versus Slou comparison only during T1. This seems at odds with Myc's well-known role as an oncogene promoting tissue proliferation (Dang 2012; Grifoni et Bellosta 2015) and the implication of *SSBP*s in differentiation. However, one study identified a role for Myc in depleting the neural progenitor pool and promoting differentiation in the developing chick neural tube (Zinin et al. 2014) in an embryonic stage dependent fashion. This study showed that the pro-differentiation role of Myc was dependent on its DNA-binding activity and expressing a dominant negative form led to reduced differentiation. Other studies in mice reported similar results of this gene promoting differentiation in the epidermis in a context-dependent manner (Watt, Frye, et Benitah 2008). In addition, the expression of c-Myc has been shown to be driven by a CT-rich element with mirror repeats closely resembling the CT-rich element that was thought to bind chicken SSDP

in the chicken a2(I) collagen gene promoter (Takimoto et al. 1993). Since no direct binding to DNA has been proven for Ssdp *in vivo*, it could bind DNA via its cofactors. So, it would be interesting to verify if there are differences in *Dm* expression in WT versus *Ssdp* mutant somatic muscles given the conserved nature of both genes. This might help elaborate on the muscle subset dependent overlap in phenotypes between *Ssdp*, dTCF^{DN} and *wg^{ts}* mutants. The generation of an anti-Ssdp antibody to use for a ChIP-Seq analysis would help shed more light on direct versus indirect Ssdp targets. As an alternative to ChIP-Seq, a CUT&Tag (Cleavage Under Targets and Tagmentation) (Kaya-Okur et al. 2020) analysis could help map Ssdp targets with high confidence even for small sample sizes, such as specific muscle subsets.

4.4. Stage specific and muscle subset specific roles for Ssdp and the Wnt signaling pathway

Another finding from this study is the stage and muscle specific phenotypes under conditions of loss of Wg and Wg signaling via the canonical pathway involving dTCF. The ventrally projecting VO4-6 muscles that were previously observed under conditions of ectopic Vg/Sd expression that are also observed when Wg is deactivated during very early or mid development indicates a role for Wg signaling in the manifestation of this phenotype. The observation of this same phenotype in *Ssdp* mutants points towards the involvement of the ChiLS complex. The presence of a single Eve+ pericardial cell in *Ssdp* mutants similar to that observed in conditions of low Wg or mutated dTCF binding sites on the *eve* enhancer and the diffuse expression of the midEn^{LTiTFs}dTCF^{-} mutated enhancer at stage 12 resembling Mid expression in *Ssdp* mutants supports the role of Ssdp in regulating these cardiac-cum-muscle TFs via the canonical Wnt pathway in cardiac cells. The overlapping phenotypes in specific ventral muscles between *Ssdp^{L5}*, dTCF^{DN} and *wg^{ts}* mutants suggest a muscular role for this pathway. These observations indicate that it is the availability of Ssdp as well as that of its regulators and interactors that would dictate cell fates and differentiation. If and when a Ssdp antibody becomes available, it would be interesting to verify its expression at various stages in *wg^{ts}* mutants to verify if there is some kind of feedback loop regulating the expression of Wg and Ssdp since both Wg and Ssdp appear to play stage and muscle subset specific roles.

4.5. Elaboration of genes involved in myogenesis

My bioinformatic analysis identified potential candidates involved in myogenesis. This study identified a downregulation of non protein-coding mRNA such as *RNaseMRP:RNA,* the small nuclear RNA *snRNA:U2:34ABb* and the small nucleolar RNA *snoRNA:Pst28S-2566* in muscles. A temporally regulated differential expression of specific spliceosomal components such as *LSm3* and *LSm4* in all muscles with respect to whole embryonic mRNA was identified. The bioinformatic analysis revealed a significant differential expression of genes such as *qua* and *Tis11* as well as an upward temporal profile for *sbb* in the Lms subset with respect to the Slou subset. It also identified a significant upregulation of *tsr* during T3 only (along with *Gel* and *qua*) in the Lms subset with respect to Duf and during T2 and T3 in the Slou subset. A classic *in situ* hybridization confirmed a muscular expression for *qua*, *Tis11* and *sbb* that indicate potential roles for them during myogenesis. Tis11 is an mRNA binding protein that governs their stability (Yeh et al. 2012). So, different levels of the protein in different subsets might be necessary for muscle subset specific protein stoichiometry.

4.6. Genes and pathways that dictate the muscle identity code

Multiple studies in *Drosophila* have unveiled TFs and cofactors that comprise the muscle identity code. Yet, they do not explain all phenotypes observed and much still remains to be discovered. My bioinformatic analysis revealed potential new contributors to muscle identity including signaling pathways, potential components of the iTF code and downstream realisator genes. Temporal profiling shows a steady upregulation of genes implicated in the FOXO, Hippo and mTOR signaling in the Lms population versus genes implicated in inositol phosphate metabolism in the Slou population. This could potentially be the defining factor for the differential expression of TFs and cofactors in the Lms subset versus Slou subset. The analysis identified new TFs that are differentially expressed in the Lms and Slou subsets. While *D*, *Sox14*, *Sox21b*, *NK7.1*, *Hmx* and *Ssdp* are upregulated in the Lms subset, *Stat92E* is upregulated in the Slou subset. This differential expression of TFs could potentially subsequently regulate differential gene transcription in these two muscle subsets. A prior study revealed that muscular genes are not expressed at the same levels in all muscles (Bataillé et al. 2017). My bioinformatic analysis of the transcriptomics data also reveals the same trend for known Mef2

targets as well as genes implicated in sarcomerogenesis and myoblast fusion. Apart from this, the spliceosomal component *LSm7* and the histone acetyltransferase Ada2a were identified as being Lms specific. The *Drosophila* villin, *qua* is significantly differentially expressed in the LT subset during T3 and displays a noticeable patterned, localized expression in cells located between the LT2 and LT3 muscles where the dorsal branch of the SNa defasciculates. Qua is an actin bundling protein that has been implicated in the formation of filopodia-like actin cables (Huelsmann, Ylänne, et Brown 2013). Given its characteristic positioning during the time when innervation processes are put in place and its significant upregulation with other actin binding proteins such as Tsr and Gel, it would be interesting to verify if there are innervation defects in *qua* mutants.

4.7. Cis regulatory regions and the muscle identity code

Apart from the identification of differentially regulated genes, my cis regulatory element enrichment analysis identified potential upstream regulators of muscle subset specific genes. GA and CT-rich motifs are highly enriched in the Lms subset when compared to the Slou subset and could indicate muscle specific regulation. Multiple observations corroborate this:

1. An enrichment of the GA-rich vertebrate Zinc-finger ZZ-type Containing 3 (ZZZ3) motif orthologous to Ada2a-containing complex component 1 (Atac1) in the Lms subset during T2 and T3.

2. The significant upregulation of Ada2a in this subset and its downregulation in the Slou subset might indicate muscle subset specific histone acetylation.

3. *Ssdp* is among putative target genes containing the CT-rich motif. Ssdp itself was initially identified as a TF that binds CT-rich motifs and is known to act in complex with other proteins.

4. Trl that is involved in chromatin modification and binds another CT-rich motif is significantly upregulated during T3 and its motif is enriched among significantly upregulated genes during this timepoint in the Lms subset. Among the top ChIP-Seq modERN peaks for Trl is Ssdp. Since no direct DNA-binding has been proven for Ssdp, it probably binds DNA via its cofactors.

5. The Ada2a interactor Nelf-E is significantly upregulated during T3. Trl and Nelf-E are implicated in RNA polymerase II pausing to regulate stage dependent transcription (Tsai et al. 2016; Yamaguchi et al. 1999).

6. Ada2a is significantly downregulated in the Slou subset.

On the other hand, in the Slou subset, modERN putative targets for another TF implicated in RNA polymerase II pausing, M1BP and GATA motifs as well as vertebrate CrebB motifs are enriched among significantly upregulated genes. The identification of multiple putative modERN Slou targets that are significantly downregulated in the Slou subset and upregulated in the Lms subset indicates that it might act as a transcriptional repressor. Given that the Nf-YB motif is enriched among genes significantly downregulated during T2 in the Slou population and that putative modERN targets of this TF include multiple iTFs, this TF could potentially act upstream of muscle iTFs. The observation that many LT muscle specific iTFs that are putative Nf-YB targets are significantly downregulated in the Slou subset reinforces the potential role of Slou as a transcriptional repressor.

Interestingly, Sd, a ventral muscle iTF is upregulated in the LT subset during T2 while during T3, there is significant enrichment for its modERN targets among significantly downregulated genes. This indicates a temporal regulation of Sd in the LT subset.

Figure 48 represents the potential TF hierarchy and interactions summarizing all the observations, with Nf-YB as a gene upstream of significantly upregulated Lms genes such as the LT iTF, Caup and the potential new LT iTF, D. These in turn regulate the expression of other LT specific genes. Caup, for example, positively regulates *Trl* expression while D autoregulates itself. Another TF that is upregulated in the Lms subset is Tup, a homeodomain TF that physically interacts with Chi (Biryukova et Heitzler 2005; Torigoi et al. 2000) and potentially with Ssdp, that also positively regulates Trl that in turn regulates *Ssdp* expression. *slou* is another Nf-YB putative target that represses the expression of multiple genes significantly upregulated in the LT subset including *Sox14*, *Tis11* and *Sox21b*. The biological analysis of *Ssdp* mutants, on the other hand, revealed that Ssdp influences the levels of expression of the iTFs Slou, Col and Mid as well as Wg. It also unveiled both Ap and Mid as potential upstream regulators of *Gel* expression. Observations from all the analyses can be summarized in the potential muscle identity regulatory network shown in Figure 48.

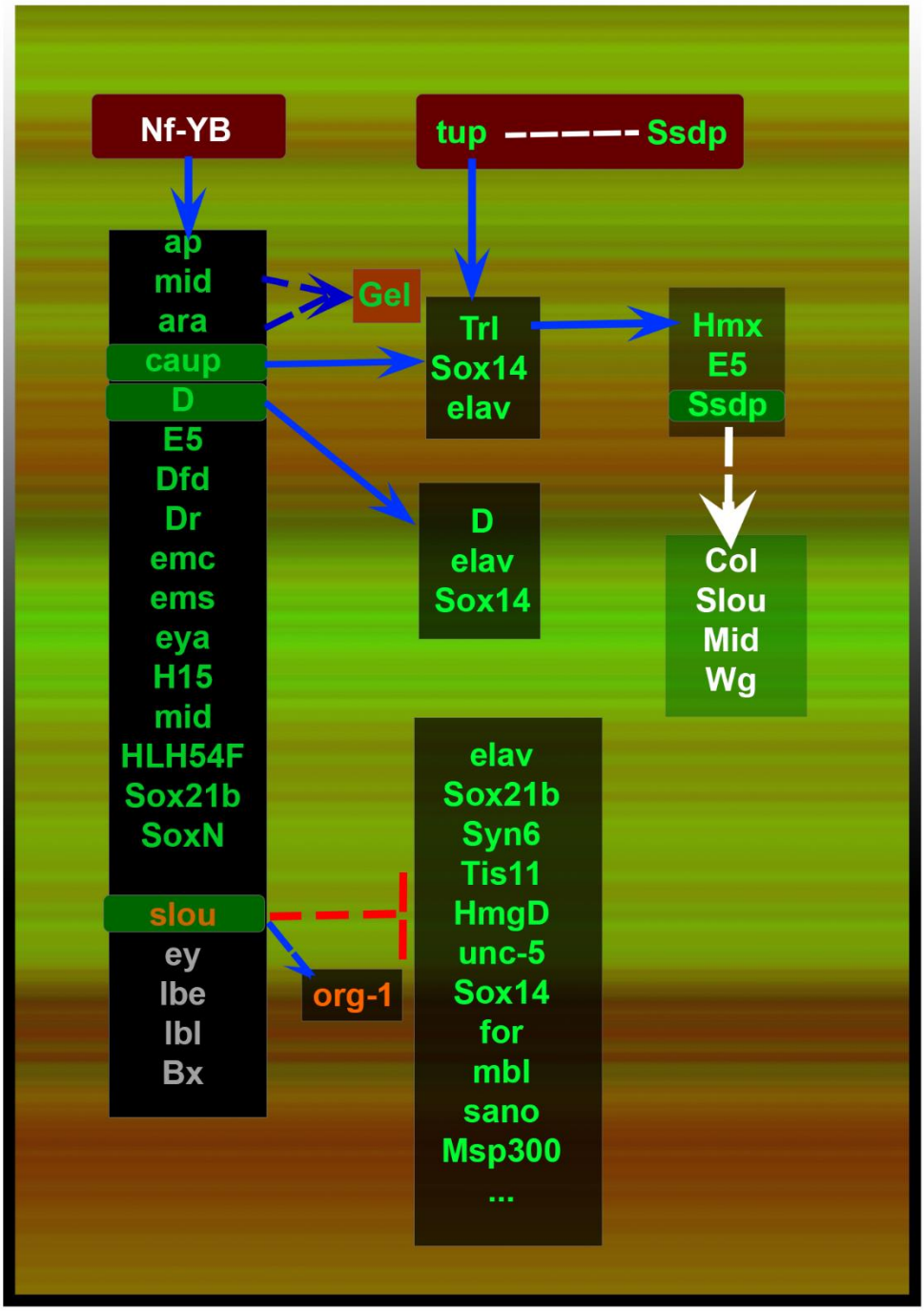
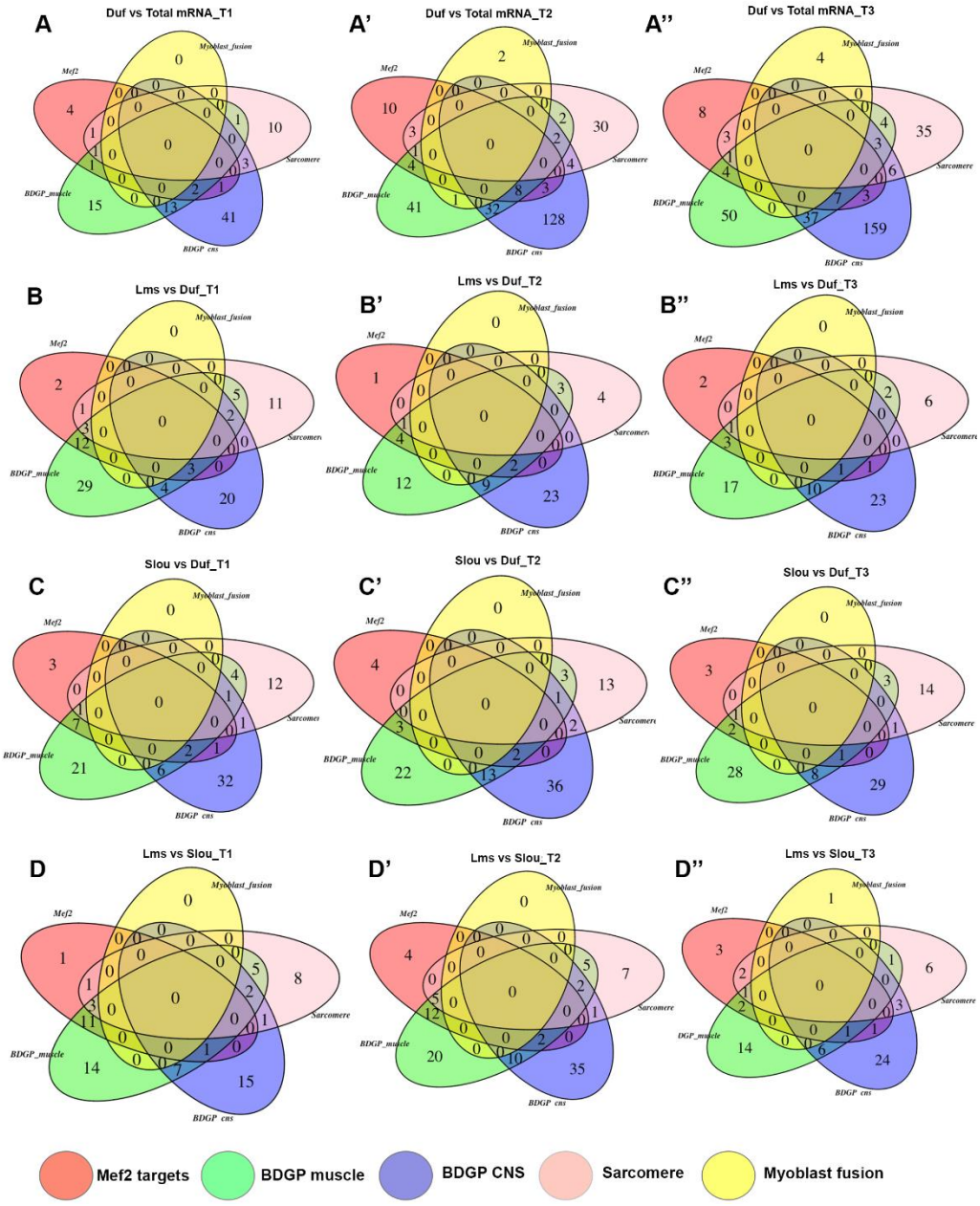


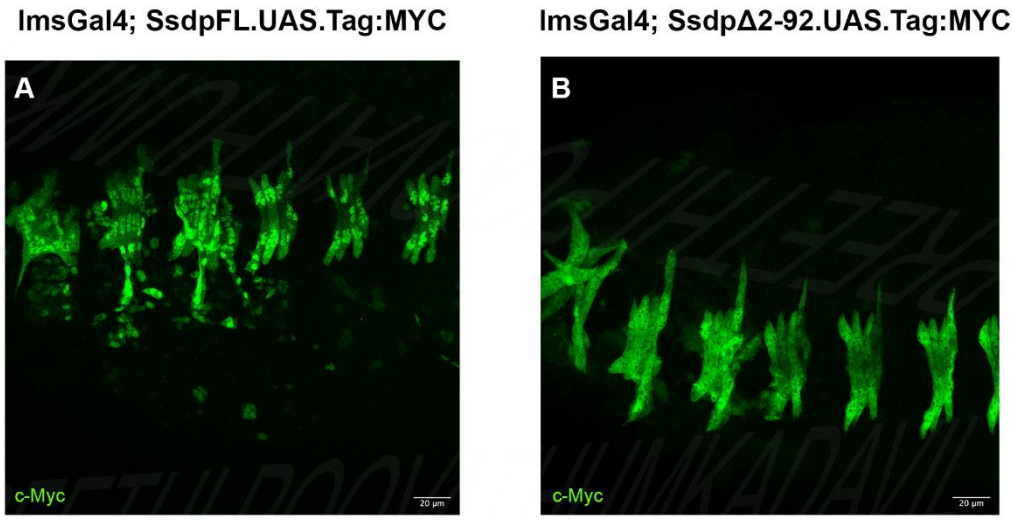
Figure 48. Summary of the potential muscle identity network identified from bioinformatic and biological analyses. In green are genes significantly upregulated in the Lms subset at any timepoint.

Nf-YB potentially acts upstream to positively regulate the expression of multiple identity genes including multiple LT iTF genes such as *ap*, *ara*, *caup*, *mid* and *Dr* as well as the Slou subset identity gene, *slou*. A modERN analysis revealed the Slou subset iTF, *org-1* as a potential Slou target apart from multiple genes upregulated in the Lms subset including *Sox21b* and *Tis11*. Caup positively regulates *Trl* expression in parallel with Tup. Tup in turn positively regulates *Ssdp* expression. It is a homeodomain TF and a known interactor of Chi and potentially also interacts with Ssdp as part of the ChiLS complex. From the analysis of *Ssdp* mutants, one can conclude that Ssdp influences the protein levels of iTFs such as Col, Slou and Mid as well as that of Wg. The analysis of transheterozygotes taken together with the parallel study of *Gel* in the team reveals both Ap and Mid as potential upstream regulators of *Gel*.



Annexe 1: Comparison of significantly downregulated genes with public data from articles for Mef2 targets, sarcomeric genes and myoblast fusion genes and the curated BDGP database for genes with expression in the somatic muscles and/or CNS.

(A-A'') The number of genes already annotated to muscles or identified to have muscle expression increases over the three timepoints in the Duf TRAP vs embryonic mRNA comparison as do genes annotated to be expressed in the CNS. **(B-B'')** Comparing the Lms subset versus the Duf population, the number of significantly downregulated muscle genes decreases between T1 and T2. **(C-C'')** Comparing the Slou subset versus the Duf population, the number of annotated significantly downregulated muscle genes slightly increases over time. **(D-D'')** Comparing the Lms subset versus the Slou subset, the number of annotated significantly downregulated muscle and CNS genes is highest during T2.



Annexe 2. Nuclear localization of Ssdp.
(A) Expression of Myc tagged full length Ssdp protein driven by Lms-Gal4 results in a strong and full nuclear localization of Ssdp**. (B)** Expression of a mutated version of Ssdp lacking its Chi-binding domain results in very weak nuclear localization in a few LT nuclei while most of the expression is distributed in the cytoplasm.

REFERENCES

1. Agulnick, Alan D. et al. 1996. « Interactions of the LIM-domain-binding factor Ldbl with LIM homeodomain proteins ». *Nature* 384(6606): 270-72.
2. Amthor, Helge et al. 2006. « Myostatin imposes reversible quiescence on embryonic muscle precursors ». *Developmental Dynamics* 235(3): 672-80.
3. Anakwe, Kelly et al. 2003. « Wnt Signalling Regulates Myogenic Differentiation in the Developing Avian Wing. » *Development (Cambridge, England)* 130(15): 3503-14.
4. Ashton, Nicholas W. et al. 2016. « Novel insight into the composition of human single-stranded DNA-binding protein 1 (hSSB1)-containing protein complexes ». *BMC Molecular Biology* 17(1): 24.
5. Aulehla, Alexander, et Olivier Pourquié. 2010. « Signaling Gradients during Paraxial Mesoderm Development ». *Cold Spring Harbor perspectives in biology* 2(2): a000869-a000869.
6. Bachinski, Linda L et al. 2014. « Most Expression and Splicing Changes in Myotonic Dystrophy Type 1 and Type 2 Skeletal Muscle Are Shared with Other Muscular Dystrophies ». *Neuromuscular disorders : NMD* 24(3): 227-40.
7. Badenhorst, Paul, Matthew Voas, Ilaria Rebay, et Carl Wu. 2002. « Biological Functions of the ISWI Chromatin Remodeling Complex NURF ». *Genes & development* 16(24): 3186-98.
8. Baghdadi, Meryem B. et al. 2018. « Notch-Induced MiR-708 Antagonizes Satellite Cell Migration and Maintains Quiescence. » *Cell stem cell* 23(6): 859-868.e5.
9. Bailey, Timothy L., James Johnson, Charles E. Grant, et William S. Noble. 2015. « The MEME Suite. » *Nucleic acids research* 43(W1): W39-49.
10. Bataillé, Laetitia, Hadi Boukhatmi, Jean-Louis Frendo, et Alain Vincent. 2017. « Dynamics of transcriptional (re)-programming of syncytial nuclei in developing muscles ». *BMC Biology* 15.
11. Balakrishnan, M.; Yu, S.F.; Chin, S.M.; Soffar, D.B.; Windner, S.E.; Goode, B.L.; Baylies, M.K. Cofilin Loss in Drosophila Muscles Contributes to Muscle Weakness through Defective Sarcomerogenesis during Muscle Growth. *Cell Rep* **2020**, *32*, 107893, doi:10.1016/j.celrep.2020.107893.
12. Bayarsaihan, Dashzeveg, Ricardo J. Soto, et Lewis N. Lukens. 1998. « Cloning and characterization of a novel sequence-specific single-stranded-DNA-binding protein ».

Biochemical Journal 331(2): 447-52.

13. Becker, Thomas et al. 2002. « Multiple functions of LIM domain-binding CLIM/NLI/Ldb cofactors during zebrafish development ». *Mechanisms of Development* 117(1): 75-85.
14. Bejarano, Fernando et al. 2007. « Hedgehog Restricts Its Expression Domain in the Drosophila Wing ». *EMBO reports* 8(8): 778-83.
15. Bejsovec, A., et A. Martinez Arias. 1991. « Roles of Wingless in Patterning the Larval Epidermis of Drosophila. » *Development (Cambridge, England)* 113(2): 471-85.
16. Bendall, A. J. et al. 1999. « Msx1 Antagonizes the Myogenic Activity of Pax3 in Migrating Limb Muscle Precursors. » *Development (Cambridge, England)* 126(22): 4965-76.
17. Ben-Haim, Nadav et al. 2006. « The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4 ». *Developmental Cell* 11(3): 313-23.
18. Bentzinger, C Florian, Yu Xin Wang, et Michael A Rudnicki. 2012. « Building Muscle: Molecular Regulation of Myogenesis ». *Cold Spring Harbor perspectives in biology* 4(2): a008342.
19. Bertin, Benjamin et al. 2015. « TRAP-Rc, Translating Ribosome Affinity Purification from Rare Cell Populations of Drosophila Embryos ». *Journal of visualized experiments : JoVE* (103): 52985.
20. Bertin, Benjamin. 2017. « Analyses Moléculaires Et Fonctionnelles De La Diversification Musculaire Par De Nouvelles Approches Génomiques Cellules-Spécifiques Chez La Drosophile. » Université Clermont Auvergne. https://www.theses.fr/2017CLFAS002.
21. Bertin, Benjamin et al. 2021. « Gelsolin and dCryAB act downstream of muscle identity genes and contribute to preventing muscle splitting and branching in Drosophila ». *bioRxiv*: 784546.
22. Bessho, Yasumasa, Hiromi Hirata, Yoshito Masamizu, et Ryoichiro Kageyama. 2003. « Periodic Repression by the BHLH Factor Hes7 Is an Essential Mechanism for the Somite Segmentation Clock. » *Genes & development* 17(12): 1451-56.
23. Bi, Pengpeng et al. 2018. « Fusogenic micropeptide Myomixer is essential for satellite cell fusion and muscle regeneration ». *Proceedings of the National Academy of Sciences* 115(15): 3864.
24. Biressi, Stefano, Mario Molinaro, et Giulio Cossu. 2007. « Cellular heterogeneity during vertebrate skeletal muscle development ». *Developmental Biology* 308(2): 281-93.
25. Biryukova, Inna, et Pascal Heitzler. 2005. « The Drosophila LIM-homeodomain protein Islet antagonizes proneural cell specification in the peripheral nervous system ».

Developmental Biology 288(2): 559-70.

26. Bladt, Friedhelm et al. 1995. « Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud ». *Nature* 376(6543): 768-71.
27. Blair, Adrienne et al. 2006. « Twinstar, the Drosophila homolog of cofilin/ADF, is required for planar cell polarity patterning ». *Development* 133(9): 1789.
28. Bober, E. et al. 1991. « The Muscle Regulatory Gene, Myf-6, Has a Biphasic Pattern of Expression during Early Mouse Development. » *The Journal of cell biology* 113(6): 1255-65.
29. Bonn, Stefan et al. 2012. « Cell type–specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP ». *Nature Protocols* 7(5): 978-94.
30. Borello, Ugo et al. 2006. « The Wnt/β-catenin pathway regulates Gli-mediated Myf5 expression during somitogenesis ». *Development* 133(18): 3723.
31. Borycki, A.G. et al. 1999. « Pax3 functions in cell survival and in pax7 regulation ». *Development* 126(8): 1665.
32. Botas, Juan. 2007. « Drosophila researchers focus on human disease ». *Nature Genetics* 39(5): 589-589.
33. Bottinelli, R, et C Reggiani. 2000. « Human skeletal muscle fibres: molecular and functional diversity ». *Progress in Biophysics and Molecular Biology* 73(2): 195-262.
34. Bour, B A et al. 1995. « Drosophila MEF2, a Transcription Factor That Is Essential for Myogenesis. » *Genes & Development* 9(6): 730-41.
35. Bourgouin, Catherine, Scott E. Lundgren, et John B. Thomas. 1992. « apterous is a drosophila LIM domain gene required for the development of a subset of embryonic muscles ». *Neuron* 9(3): 549-61.
36. Brand, A. H., et N. Perrimon. 1993. « Targeted Gene Expression as a Means of Altering Cell Fates and Generating Dominant Phenotypes. » *Development (Cambridge, England)* 118(2): 401-15.
37. Braun, T. et al. 1990. « Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. » *The EMBO Journal* 9(3): 821-31.
38. Brent, Ava E., Thomas Braun, et Clifford J. Tabin. 2005. « Genetic Analysis of Interactions between the Somitic Muscle, Cartilage and Tendon Cell Lineages during Mouse Development. » *Development (Cambridge, England)* 132(3): 515-28.
39. Brent, Ava E., et Clifford J. Tabin. 2004. « FGF Acts Directly on the Somitic Tendon Progenitors through the Ets Transcription Factors Pea3 and Erm to Regulate Scleraxis Expression. » *Development (Cambridge, England)* 131(16): 3885-96.

40. Briggs, Derek E.G. 2015. « The Cambrian explosion ». *Current Biology* 25(19): R864-68.
41. Brohmann, H., K. Jagla, et C. Birchmeier. 2000. « The Role of Lbx1 in Migration of Muscle Precursor Cells. » *Development (Cambridge, England)* 127(2): 437-45.
42. Bronstein, Revital et al. 2010. « Transcriptional Regulation by CHIP/LDB Complexes ». *PLoS genetics* 6(8): e1001063-e1001063.
43. Buckingham, Margaret. 2017. « Gene regulatory networks and cell lineages that underlie the formation of skeletal muscle ». *Proceedings of the National Academy of Sciences* 114(23): 5830.
44. Buckingham, Margaret, et Stéphane D Vincent. 2009. « Distinct and dynamic myogenic populations in the vertebrate embryo ». *Differentation and gene regulation* 19(5): 444-53.
45. Buescher, Marita et al. 2004. « Drosophila T Box Proteins Break the Symmetry of Hedgehog-Dependent Activation of Wingless. » *Current biology : CB* 14(19): 1694-1702.
46. Burgess, Rob et al. 1996. « Requirement of the paraxis gene for somite formation and musculoskeletal patterning ». *Nature* 384(6609): 570-73.
47. Calza, Stefano et al. 2007. « Filtering Genes to Improve Sensitivity in Oligonucleotide Microarray Data Analysis. » *Nucleic acids research* 35(16): e102.
48. Capovilla, M., Z. Kambris, et J. Botas. 2001. « Direct regulation of the muscle-identity gene apterous by a Hox protein in the somatic mesoderm ». *Development* 128(8): 1221.
49. Carrasco-Rando, Marta et al. 2011. « Drosophila Araucan and Caupolican Integrate Intrinsic and Signalling Inputs for the Acquisition by Muscle Progenitors of the Lateral Transverse Fate ». *PLoS genetics* 7(7): e1002186-e1002186.
50. Carrera, Inés et al. 2008. « Pygopus Activates Wingless Target Gene Transcription through the Mediator Complex Subunits Med12 and Med13 ». *Proceedings of the National Academy of Sciences of the United States of America* 105(18): 6644-49.
51. Cartharius, K. et al. 2005. « MatInspector and beyond: promoter analysis based on transcription factor binding sites ». *Bioinformatics* 21(13): 2933-42.
52. Castro, Patricia, Hong Liang, Jan C Liang, et Lalitha Nagarajan. 2002. « A Novel, Evolutionarily Conserved Gene Family with Putative Sequence-Specific Single-Stranded DNA-Binding Activity ». *Genomics* 80(1): 78-85.
53. Chal, Jérome, et Olivier Pourquié. 2017. « Making muscle: skeletal myogenesis in vivo and in vitro ». *Development* 144(12): 2104.
54. Chen, Lan et al. 2002. « Ssdp Proteins Interact with the LIM-Domain-Binding Protein Ldb1 to Regulate Development ». *Proceedings of the National Academy of Sciences of the United States of America* 99(22): 14320-25.

55. Chopra, Vivek S, Jessica Cande, Joung-Woo Hong, et Michael Levine. 2009. « Stalled Hox Promoters as Chromosomal Boundaries ». *Genes & development* 23(13): 1505-9.
56. Christ, Bodo, et Beate Brand-Saberi. 2002. « Limb Muscle Development. » *The International journal of developmental biology* 46(7): 905-14.
57. Cohen, B. et al. 1992. « Apterous, a Gene Required for Imaginal Disc Development in Drosophila Encodes a Member of the LIM Family of Developmental Regulatory Proteins. » *Genes & development* 6(5): 715-29.
58. Cohn, M. J. et al. 1995. « Fibroblast Growth Factors Induce Additional Limb Development from the Flank of Chick Embryos. » *Cell* 80(5): 739-46.
59. Conner, J, et Z Liu. 2000. « LEUNIG, a Putative Transcriptional Corepressor That Regulates AGAMOUS Expression during Flower Development ». *Proceedings of the National Academy of Sciences of the United States of America* 97(23): 12902-7.
60. Cossu, G. et al. 1996. « Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm ». *Development* 122(2): 429.
61. Cossu, Giulio, et Stefano Biressi. 2005. « Satellite cells, myoblasts and other occasional myogenic progenitors: Possible origin, phenotypic features and role in muscle regeneration ». *Biology of Hypoxia and Myogenesis and Muscle Disease* 16(4): 623-31.
62. Cox, Virginia T., et Mary K. Baylies. 2005. « Specification of individual Slouch muscle progenitors in Drosophila requires sequential Wingless signaling ». *Development* 132(4): 713.
63. Cripps, R M et al. 1998. « The Myogenic Regulatory Gene Mef2 Is a Direct Target for Transcriptional Activation by Twist during Drosophila Myogenesis ». *Genes & development* 12(3): 422-34.
64. Dale, J. K. et al. 2003. « Periodic Notch Inhibition by Lunatic Fringe Underlies the Chick Segmentation Clock. » *Nature* 421(6920): 275-78.
65. Dang, Chi V. 2012. « MYC on the Path to Cancer ». *Cell* 149(1): 22-35.
66. Daubas, Philippe, et Margaret E. Buckingham. 2013. « Direct molecular regulation of the myogenic determination gene Myf5 by Pax3, with modulation by Six1/4 factors, is exemplified by the −111kb-Myf5 enhancer ». *Developmental Biology* 376(2): 236-44.
67. DeAguero, Ashley A et al. 2019. « Regulation of Fiber-Specific Actin Expression by the Drosophila SRF Ortholog Blistered ». *Development (Cambridge, England)* 146(7): dev164129.
68. Delfini, Marie-Claire et al. 2009. « The Timing of Emergence of Muscle Progenitors Is Controlled by an FGF/ERK/SNAIL1 Pathway. » *Developmental biology* 333(2): 229-37.

69. Denans, Nicolas, Tadahiro Iimura, et Olivier Pourquié. 2015. « Hox genes control vertebrate body elongation by collinear Wnt repression » éd. Marianne E Bronner. *eLife* 4: e04379.
70. Deng, Hua, Sarah C Hughes, John B Bell, et Andrew J Simmonds. 2009. « Alternative Requirements for Vestigial, Scalloped, and Dmef2 during Muscle Differentiation in Drosophila Melanogaster ». *Molecular biology of the cell* 20(1): 256-69.
71. Deng, Su, Mafalda Azevedo, et Mary Baylies. 2017. « Acting on Identity: Myoblast Fusion and the Formation of the Syncytial Muscle Fiber ». *Seminars in cell & developmental biology* 72: 45-55.
72. Deries, Marianne, Jennifer J. P. Collins, et Marilyn J. Duxson. 2008. « The Mammalian Myotome: A Muscle with No Innervation. » *Evolution & development* 10(6): 746-55.
73. Deries, Marianne, et Sólveig Thorsteinsdóttir. 2016. « Axial and limb muscle development: dialogue with the neighbourhood ». *Cellular and Molecular Life Sciences* 73(23): 4415-31.
74. DeSesso, John M. 2017. « Vascular ontogeny within selected thoracoabdominal organs and the limbs ». *Developmental Angiogenesis* 70: 3-20.
75. Devoto, S.H., E. Melancon, J.S. Eisen, et M. Westerfield. 1996. « Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation ». *Development* 122(11): 3371.
76. Dobi, Krista C, Victoria K Schulman, et Mary K Baylies. 2015. « Specification of the Somatic Musculature in Drosophila ». *Wiley interdisciplinary reviews. Developmental biology* 4(4): 357-75.
77. Dobi, Krista C., Marc S. Halfon, et Mary K. Baylies. 2014. « Whole-Genome Analysis of Muscle Founder Cells Implicates the Chromatin Regulator Sin3A in Muscle Identity ». *Cell Reports* 8(3): 858-70.
78. Dobin, Alexander et al. 2013. « STAR: Ultrafast Universal RNA-Seq Aligner. » *Bioinformatics (Oxford, England)* 29(1): 15-21.
79. Dohrmann, C, N Azpiazu, et M Frasch. 1990. « A new Drosophila homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. » *Genes & Development* 4(12a): 2098-2111.
80. Dos Santos, Matthieu et al. 2020. « Single-Nucleus RNA-Seq and FISH Identify Coordinated Transcriptional Activity in Mammalian Myofibers ». *Nature communications* 11(1): 5102-5102.
81. Duan, Hong et al. 2007. « A key role of Pox meso in somatic myogenesis of Drosophila ». *Development* 134(22): 3985.

82. Edmondson, D. G., G. E. Lyons, J. F. Martin, et E. N. Olson. 1994. « Mef2 Gene Expression Marks the Cardiac and Skeletal Muscle Lineages during Mouse Embryogenesis. » *Development (Cambridge, England)* 120(5): 1251-63.
83. Elman, Jessica S. et al. 2019. « Identification of FUBP1 as a Long Tail Cancer Driver and Widespread Regulator of Tumor Suppressor and Oncogene Alternative Splicing. » *Cell reports* 28(13): 3435-3449.e5.
84. Emery, Alan EH. 2002. « The muscular dystrophies ». *The Lancet* 359(9307): 687-95.
85. Emes, Richard D., et Chris P. Ponting. 2001. « A new sequence motif linking lissencephaly, Treacher Collins and oral–facial–digital type 1 syndromes, microtubule dynamics and cell migration ». *Human Molecular Genetics* 10(24): 2813-20.
86. Epstein, J A et al. 1996. « Pax3 Modulates Expression of the C-Met Receptor during Limb Muscle Development ». *Proceedings of the National Academy of Sciences of the United States of America* 93(9): 4213-18.
87. Fakhouri, Walid D. et al. 2017. « Intercellular Genetic Interaction Between Irf6 and Twist1 during Craniofacial Development. » *Scientific reports* 7(1): 7129.
88. Feldhahn, Niklas et al. 2012. « The HSSB1 Orthologue Obfc2b Is Essential for Skeletogenesis but Dispensable for the DNA Damage Response in Vivo. » *The EMBO journal* 31(20): 4045-56.
89. Fernández-Fúnez, P. et al. 1998. « The Relative Expression Amounts of Apterous and Its Co-Factor DLdb/Chip Are Critical for Dorso-Ventral Compartmentalization in the Drosophila Wing. » *The EMBO journal* 17(23): 6846-53.
90. Ferretti, Elisabetta, et Anna-Katerina Hadjantonakis. 2019. « Mesoderm specification and diversification: from single cells to emergent tissues ». *Differentiation and disease* 61: 110-16.
91. Fiedler, Marc et al. 2015. « An Ancient Pygo-Dependent Wnt Enhanceosome Integrated by Chip/LDB-SSDP ». *eLife* 4: e09073.
92. Fitts, Robert H., Danny R. Riley, et Jeffrey J. Widrick. 2001. « Functional and structural adaptations of skeletal muscle to microgravity ». *Journal of Experimental Biology* 204(18): 3201.
93. Fleisig, H. B. et al. 2007. « Adenoviral E1B55K Oncoprotein Sequesters Candidate Leukemia Suppressor Sequence-Specific Single-Stranded DNA-Binding Protein 2 into Aggresomes. » *Oncogene* 26(33): 4797-4805.
94. Folker, Eric S, Victoria K Schulman, et Mary K Baylies. 2014. « Translocating Myonuclei Have Distinct Leading and Lagging Edges That Require Kinesin and Dynein ». *Development (Cambridge, England)* 141(2): 355-66.

95. Fouchécourt, Sophie et al. 2019. « An evolutionary approach to recover genes predominantly expressed in the testes of the zebrafish, chicken and mouse ». *BMC Evolutionary Biology* 19(1): 137.
96. Francis-West, Philippa H, Laurent Antoni, et Kelly Anakwe. 2003. « Regulation of Myogenic Differentiation in the Developing Limb Bud ». *Journal of anatomy* 202(1): 69-81.
97. Fu, Chong-Lei et al. 2016. « The T-Box Transcription Factor Midline Regulates Wing Development by Repressing Wingless and Hedgehog in Drosophila ». *Scientific reports* 6: 27981-27981.
98. Füchtbauer, Ernst-martin. 1995. « Expression of M-twist during postimplantation development of the mouse ». *Developmental Dynamics* 204(3): 316-22.
99. Fujioka, Miki et al. 2005. « Embryonic Even Skipped-Dependent Muscle and Heart Cell Fates Are Required for Normal Adult Activity, Heart Function, and Lifespan ». *Circulation research* 97(11): 1108-14.
100. Gaertner, Bjoern et al. 2012. « Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression ». *Cell reports* 2(6): 1670-83.
101. Galli, Lisa M et al. 2008. « Identification and Characterization of Subpopulations of Pax3 and Pax7 Expressing Cells in Developing Chick Somites and Limb Buds ». *Developmental dynamics : an official publication of the American Association of Anatomists* 237(7): 1862-74.
102. Ganguly, Atish, Jin Jiang, et Y. Tony Ip. 2005. « Drosophila WntD is a target and an inhibitor of the Dorsal/Twist/Snail network in the gastrulating embryo ». *Development* 132(15): 3419.
103. García-Zaragoza, Elena et al. 2008. « CF2 activity and enhancer integration are required for proper muscle gene expression in Drosophila ». *Mechanisms of Development* 125(7): 617-30.
104. Geetha-Loganathan, Poongodi et al. 2005. « Ectodermal Wnt-6 Promotes Myf5-Dependent Avian Limb Myogenesis. » *Developmental biology* 288(1): 221-33.
105. Giordani, Julien et al. 2007. « Six Proteins Regulate the Activation of Myf5 Expression in Embryonic Mouse Limbs ». *Proceedings of the National Academy of Sciences of the United States of America* 104(27): 11310-15.
106. Green, Nicole et al. 2016. « A Common Suite of Coagulation Proteins Function in Drosophila Muscle Attachment ». *Genetics* 204(3): 1075-87.
107. Grifone, Raphaelle et al. 2005. « Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo ». *Development* 132(9): 2235.

108. Grifoni, Daniela, et Paola Bellosta. 2015. « Drosophila Myc: A Master Regulator of Cellular Performance ». *Biochimica et biophysica acta* 1849(5): 570-81.
109. Gros, Jérôme, Marie Manceau, Virginie Thomé, et Christophe Marcelle. 2005. « A common somitic origin for embryonic muscle progenitors and satellite cells ». *Nature* 435(7044): 954-58.
110. Gros, Jérôme, Martin Scaal, et Christophe Marcelle. 2004. « A Two-Step Mechanism for Myotome Formation in Chick ». *Developmental Cell* 6(6): 875-82.
111. Guelman, Sebastián et al. 2006. « Host Cell Factor and an Uncharacterized SANT Domain Protein Are Stable Components of ATAC, a Novel DAda2A/DGcn5-Containing Histone Acetyltransferase Complex in Drosophila ». *Molecular and cellular biology* 26(3): 871-82.
112. Güngör, Cenap et al. 2007. « Proteasomal Selection of Multiprotein Complexes Recruited by LIM Homeodomain Transcription Factors. » *Proceedings of the National Academy of Sciences of the United States of America* 104(38): 15000-5.
113. Hales, Karen G, Christopher A Korey, Amanda M Larracuente, et David M Roberts. 2015. « Genetics on the Fly: A Primer on the Drosophila Model System ». *Genetics* 201(3): 815-42.
114. Hammond, Christina L et al. 2007. « Signals and Myogenic Regulatory Factors Restrict Pax3 and Pax7 Expression to Dermomyotome-like Tissue in Zebrafish ». *Developmental biology* 302(2): 504-21.
115. Havis, Emmanuelle et al. 2012. « Sim2 Prevents Entry into the Myogenic Program by Repressing MyoD Transcription during Limb Embryonic Myogenesis ». *Development (Cambridge, England)* 139(11): 1910-20.
116. Heinz, Sven et al. 2010. « Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. » *Molecular cell* 38(4): 576-89.
117. Hemavathy, Kirugaval et al. 2004. « The repressor function of Snail is required for Drosophila gastrulation and is not replaceable by Escargot or Worniu ». *Developmental Biology* 269(2): 411-20.
118. Herold, Nadine et al. 2009. « Conservation of the Protein Composition and Electron Microscopy Structure of Drosophila Melanogaster and Human Spliceosomal Complexes ». *Molecular and cellular biology* 29(1): 281-301.
119. Herrmann, Carl, Bram Van de Sande, Delphine Potier, et Stein Aerts. 2012. « I-CisTarget: An Integrative Genomics Method for the Prediction of Regulatory Features and Cis-Regulatory Modules ». *Nucleic acids research* 40(15): e114-e114.

120. Hirasawa, Tatsuya, et Shigeru Kuratani. 2018. « Evolution of the Muscular System in Tetrapod Limbs ». *Zoological letters* 4: 27-27.
121. Holland, Peter W. H. 2015. « Did homeobox gene duplications contribute to the Cambrian explosion? » *Zoological Letters* 1(1): 1.
122. Hollway, Georgina E, et Peter D Currie. 2003. « Myotome Meanderings. Cellular Morphogenesis and the Making of Muscle ». *EMBO reports* 4(9): 855-60.
123. Hooper, Scott L., et Jeffrey B. Thuma. 2005. « Invertebrate Muscles: Muscle Specific Genes and Proteins ». *Physiological Reviews* 85(3): 1001-60.
124. Huang, Da Wei et al. 2007. « DAVID Bioinformatics Resources: Expanded Annotation Database and Novel Algorithms to Better Extract Biology from Large Gene Lists ». *Nucleic acids research* 35(Web Server issue): W169-75.
125. Huelsmann, Sven, Jari Ylänne, et Nicholas H. Brown. 2013. « Filopodia-like Actin Cables Position Nuclei in Association with Perinuclear Actin in Drosophila Nurse Cells ». *Developmental Cell* 26(6): 604-15.
126. Hurren, Bradley, Jennifer J P Collins, Marilyn J Duxson, et Marianne Deries. 2015. « First Neuromuscular Contact Correlates with Onset of Primary Myogenesis in Rat and Mouse Limb Muscles ». *PloS one* 10(7): e0133811-e0133811.
127. Hutcheson, David A et al. 2009. « Embryonic and Fetal Limb Myogenic Cells Are Derived from Developmentally Distinct Progenitors and Have Different Requirements for Beta-Catenin ». *Genes & development* 23(8): 997-1013.
128. Ikeya, M., et S. Takada. 1998. « Wnt signaling from the dorsal neural tube is required for the formation of the medial dermomyotome ». *Development* 125(24): 4969.
129. Imbriano, Carol, et Susanna Molinari. 2018. « Alternative Splicing of Transcription Factors Genes in Muscle Physiology and Pathology ». *Genes* 9(2): 107.
130. Imrichová, Hana et al. 2015. « I-CisTarget 2015 Update: Generalized Cis-Regulatory Enrichment Analysis in Human, Mouse and Fly ». *Nucleic acids research* 43(W1): W57-64.
131. Jagla, T. et al. 1998. « ladybird determines cell fate decisions during diversification of Drosophila somatic muscles ». *Development* 125(18): 3699.
132. Jaiswal, Richa et al. 2013. « The Formin Daam1 and Fascin Directly Collaborate to Promote Filopodia Formation ». *Current biology : CB* 23(14): 1373-79.
133. Janssen, Ian, Steven B. Heymsfield, ZiMian Wang, et Robert Ross. 2000. « Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr ». *Journal of Applied Physiology* 89(1): 81-88.
134. Junion, Guillaume et al. 2007. « Genome-wide view of cell fate specification: ladybird

acts at multiple levels during diversification of muscle and heart precursors ». *Genes & Development* 21(23): 3163-80.

135. Junion, Guillaume et al. 2012. « A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History ». *Cell* 148(3): 473-86.
136. Kassar-Duchossoy, Lina et al. 2004. « Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice ». *Nature* 431(7007): 466-71.
137. Kassar-Duchossoy, Lina et al. 2005. « Pax3/Pax7 Mark a Novel Population of Primitive Myogenic Cells during Development ». *Genes & development* 19(12): 1426-31.
138. Kaya-Okur, Hatice S. et al. 2020. « Efficient low-cost chromatin profiling with CUT&Tag ». *Nature Protocols* 15(10): 3264-83.
139. Kelly, Robert G., Loydie A. Jerome-Majewska, et Virginia E. Papaioannou. 2004. « The Del22q11.2 Candidate Gene Tbx1 Regulates Branchiomeric Myogenesis. » *Human molecular genetics* 13(22): 2829-40.
140. Kim, Myung Hee et al. 2004. « The Structure of the N-Terminal Domain of the Product of the Lissencephaly Gene Lis1 and Its Functional Implications. » *Structure (London, England : 1993)* 12(6): 987-98.
141. Knirr, S., N. Azpiazu, et M. Frasch. 1999. « The Role of the NK-Homeobox Gene Slouch (S59) in Somatic Muscle Patterning. » *Development (Cambridge, England)* 126(20): 4525-35.
142. Kudron, Michelle M et al. 2018. « The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis Elegans Transcription Factors ». *Genetics* 208(3): 937-49.
143. Kumar, Lokesh, et Matthias E Futschik. 2007. « Mfuzz: A Software Package for Soft Clustering of Microarray Data ». *Bioinformation* 2(1): 5-7.
144. Kumar, Ram P, Krista C Dobi, Mary K Baylies, et Susan M Abmayr. 2015. « Muscle Cell Fate Choice Requires the T-Box Transcription Factor Midline in Drosophila ». *Genetics* 199(3): 777-91.
145. Kumar, Sandeep, et Gregg Duester. 2014. « Retinoic Acid Controls Body Axis Extension by Directly Repressing Fgf8 Transcription. » *Development (Cambridge, England)* 141(15): 2972-77.
146. Kume, T., H. Jiang, J. M. Topczewska, et B. L. Hogan. 2001. « The Murine Winged Helix Transcription Factors, Foxc1 and Foxc2, Are Both Required for Cardiovascular Development and Somitogenesis. » *Genes & development* 15(18): 2470-82.
147. Kurusu, Mitsuhiko et al. 2008. « A Screen of Cell-Surface Molecules Identifies Leucine-Rich Repeat Proteins as Key Mediators of Synaptic Target Selection ». *Neuron* 59(6):

972-85.

148. Lawrence, Peter A, et Gary Struhl. 1996. « Morphogens, Compartments, and Pattern: Lessons from Drosophila? » *Cell* 85(7): 951-61.
149. Lawson, Teegan et al. 2020. « The structural details of the interaction of single-stranded DNA binding protein hSSB2 (NABP1/OBFC2A) with UV-damaged DNA ». *Proteins: Structure, Function, and Bioinformatics* 88(2): 319-26.
150. Lecuit, Thomas, et Stephen M. Cohen. 1997. « Proximal–distal axis formation in the Drosophila leg ». *Nature* 388(6638): 139-45.
151. Lehka, Lilya, et Maria Jolanta Rędowicz. 2020. « Mechanisms regulating myoblast fusion: A multilevel interplay ». *Differentiation of skeletal muscles* 104: 81-92.
152. Leikina, Evgenia et al. 2018. « Myomaker and Myomerger Work Independently to Control Distinct Steps of Membrane Remodeling during Myoblast Fusion ». *Developmental Cell* 46(6): 767-780.e7.
153. Leptin, M. 1991. « Twist and Snail as Positive and Negative Regulators during Drosophila Mesoderm Development. » *Genes & development* 5(9): 1568-76.
154. Leptin, Maria, et Barbara Grunewald. « Cell Shape Changes during Gastrulation in Drosophila ». : 14.
155. Li, Yongjiang et al. 2009. « HSSB1 and HSSB2 Form Similar Multiprotein Complexes That Participate in DNA Damage Response. » *The Journal of biological chemistry* 284(35): 23525-31.
156. Liang, Hong, Susmita Samanta, et Lalitha Nagarajan. 2005. « SSBP2, a candidate tumor suppressor gene, induces growth arrest and differentiation of myeloid leukemia cells ». *Oncogene* 24(16): 2625-34.
157. Liao, Yang, Gordon K. Smyth, et Wei Shi. 2014. « featureCounts: an efficient general purpose program for assigning sequence reads to genomic features ». *Bioinformatics* 30(7): 923-30.
158. Lin, Q, J Schwarz, C Bucana, et E N Olson. 1997. « Control of Mouse Cardiac Morphogenesis and Myogenesis by Transcription Factor MEF2C ». *Science (New York, N.Y.)* 276(5317): 1404-7.
159. Liu, Jifeng et al. 2016. « Single-stranded DNA binding protein Ssbp3 induces differentiation of mouse embryonic stem cells into trophoblast-like cells ». *Stem Cell Research & Therapy* 7(1): 79.
160. Liu, Jun-Wei et al. 2008. « SsDNA-Binding Protein 2 Is Frequently Hypermethylated and Suppresses Cell Growth in Human Prostate Cancer. » *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(12): 3754-60.

161. Liu, Ya-Hsin et al. 2009. « A Systematic Analysis of Tinman Function Reveals Eya and JAK-STAT Signaling as Essential Regulators of Muscle Development ». *Developmental Cell* 16(2): 280-91.
162. Lord, Pat C.W., Meei-Hua Lin, Karen H. Hales, et Robert V. Storti. 1995. « Normal Expression and the Effects of Ectopic Expression of the Drosophila muscle segment homeobox (msh) Gene Suggest a Role in Differentiation and Patterning of Embryonic Muscles ». *Developmental Biology* 171(2): 627-40.
163. Love, Michael I, Wolfgang Huber, et Simon Anders. 2014. « Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2 ». *Genome biology* 15(12): 550-550.
164. Lu, Jian-rong et al. 2002. « Control of Facial Muscle Development by MyoR and Capsulin ». *Science* 298(5602): 2378.
165. Luo, Dan, et Thomas A. Rando. 2003. « The Regulation of Catalase Gene Expression in Mouse Muscle Cells Is Dependent on the CCAAT-Binding Factor NF-Y. » *Biochemical and biophysical research communications* 303(2): 609-18.
166. Lyne, Rachel et al. 2007. « FlyMine: An Integrated Database for Drosophila and Anopheles Genomics. » *Genome biology* 8(7): R129.
167. Machado, Léo et al. 2017. « In Situ Fixation Redefines Quiescence and Early Activation of Skeletal Muscle Stem Cells ». *Cell Reports* 21(7): 1982-93.
168. Magli, Alessandro et al. 2019. « Time-Dependent Pax3-Mediated Chromatin Remodeling and Cooperation with Six4 and Tead2 Specify the Skeletal Myogenic Lineage in Developing Mesoderm ». *PLoS biology* 17(2): e3000153-e3000153.
169. Mallo, Moisés. 2016. « Revisiting the involvement of signaling gradients in somitogenesis ». *The FEBS Journal* 283(8): 1430-37.
170. Manceau, Marie et al. 2008. « Myostatin Promotes the Terminal Differentiation of Embryonic Muscle Progenitors. » *Genes & development* 22(5): 668-81.
171. Mankoo, Baljinder S. et al. 2003. « The Concerted Action of Meox Homeobox Genes Is Required Upstream of Genetic Pathways Essential for the Formation, Patterning and Differentiation of Somites. » *Development (Cambridge, England)* 130(19): 4655-64.
172. Markow, Therese Ann. 2015. « The secret lives of Drosophila flies ». *eLife* 4: e06793.
173. Martin, Adam C. 2020. « The Physical Mechanisms of Drosophila Gastrulation: Mesoderm and Endoderm Invagination ». *Genetics* 214(3): 543-60.
174. Martin, J F et al. 1994. « A Mef2 Gene That Generates a Muscle-Specific Isoform via Alternative MRNA Splicing ». *Molecular and cellular biology* 14(3): 1647-56.
175. Martyn, I et al. 2018. « Self-Organization of a Human Organizer by Combined Wnt and

Nodal Signalling ». *Nature* 558(7708): 132-35.

176. Martyn, Iain, Eric D Siggia, et Ali H Brivanlou. 2019. « Mapping Cell Migrations and Fates in a Gastruloid Model to the Human Primitive Streak ». *Development (Cambridge, England)* 146(17): dev179564.
177. Masselink, Wouter et al. 2017. « Phosphorylation of Lbx1 controls lateral myoblast migration into the limb ». *Developmental Biology* 430(2): 302-9.
178. McMahon, J. A. et al. 1998. « Noggin-Mediated Antagonism of BMP Signaling Is Required for Growth and Patterning of the Neural Tube and Somite. » *Genes & development* 12(10): 1438-52.
179. McQueen, Caitlin, et Matthew Towers. 2020. « Establishing the pattern of the vertebrate limb ». *Development* 147(17): dev177956.
180. Mennerich, D., K. Schäfer, et T. Braun. 1998. « Pax-3 Is Necessary but Not Sufficient for Lbx1 Expression in Myogenic Precursor Cells of the Limb. » *Mechanisms of development* 73(2): 147-58.
181. van Meyel, Donald J., John B. Thomas, et Alan D. Agulnick. 2003. « Ssdp Proteins Bind to LIM-Interacting Co-Factors and Regulate the Activity of LIM-Homeodomain Protein Complexes in Vivo. » *Development (Cambridge, England)* 130(9): 1915-25.
182. Milán, M, F J Diaz-Benjumea, et S M Cohen. 1998. « Beadex Encodes an LMO Protein That Regulates Apterous LIM-Homeodomain Activity in Drosophila Wing Development: A Model for LMO Oncogene Function ». *Genes & development* 12(18): 2912-20.
183. Mirvis, Mary, Tim Stearns, et W James Nelson. 2018. « Cilium Structure, Assembly, and Disassembly Regulated by the Cytoskeleton ». *The Biochemical journal* 475(14): 2329-53.
184. Miyake, Tsutomu et al. 2016. « The pectoral fin muscles of the coelacanth Latimeria chalumnae: Functional and evolutionary implications for the fin-to-limb transition and subsequent evolution of tetrapods ». *The Anatomical Record* 299(9): 1203-23.
185. Mohammed, Hisham et al. 2016. « Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes ». *Nature Protocols* 11(2): 316-26.
186. Molkentin, J. D., B. L. Black, J. F. Martin, et E. N. Olson. 1995. « Cooperative Activation of Muscle Gene Expression by MEF2 and Myogenic BHLH Proteins. » *Cell* 83(7): 1125-36.
187. Morcillo, Patrick, Christina Rosen, Mary K. Baylies, et Dale Dorsett. 1997. « Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila ». *Genes & Development* 11(20): 2729-40.

188. Morcillo, Patrick, Christina Rosen, et Dale Dorsett. 1996. « Genes Regulating the Remote Wing Margin Enhancer in the Drosophila cut Locus ». *Genetics* 144(3): 1143.
189. Moretti, Irene et al. 2016. « MRF4 negatively regulates adult skeletal muscle growth by repressing MEF2 activity ». *Nature Communications* 7(1): 12397.
190. Müller, Dominik et al. 2010. « Regulation and Functions of the Lms Homeobox Gene during Development of Embryonic Lateral Transverse Muscles and Direct Flight Muscles in Drosophila ». *PloS one* 5(12): e14323-e14323.
191. Murphy, Malea, et Gabrielle Kardon. 2011. « Origin of Vertebrate Limb Muscle: The Role of Progenitor and Myoblast Populations ». *Current topics in developmental biology* 96: 1-32.
192. Naiche, L A, Nakisha Holder, et Mark Lewandoski. 2011. « FGF4 and FGF8 Comprise the Wavefront Activity That Controls Somitogenesis ». *Proceedings of the National Academy of Sciences of the United States of America* 108(10): 4018-23.
193. Nassari, Sonya, Delphine Duprez, et Claire Fournier-Thibault. 2017. « Non-myogenic Contribution to Muscle Development and Homeostasis: The Role of Connective Tissues ». *Frontiers in Cell and Developmental Biology* 5: 22.
194. de Navascués, Joaquín, et Juan Modolell. 2007. « Tailup, a LIM-HD Gene, and Iro-C Cooperate in Drosophila Dorsal Mesothorax Specification. » *Development (Cambridge, England)* 134(9): 1779-88.
195. Niwa, Yasutaka et al. 2007. « The Initiation and Propagation of Hes7 Oscillation Are Cooperatively Regulated by Fgf and Notch Signaling in the Somite Segmentation Clock. » *Developmental cell* 13(2): 298-304.
196. Noden, Drew M., et Philippa Francis-West. 2006. « The differentiation and morphogenesis of craniofacial muscles ». *Developmental Dynamics* 235(5): 1194-1218.
197. Nose, A., T. Isshiki, et M. Takeichi. 1998. « Regional specification of muscle progenitors in Drosophila: the role of the msh homeobox gene ». *Development* 125(2): 215.
198. Okubo, Yusuke et al. 2012. « Lfng regulates the synchronized oscillation of the mouse segmentation clock via trans -repression of Notch signalling ». *Nature Communications* 3(1): 1141.
199. Paterson, B M et al. 1991. « The Drosophila Homologue of Vertebrate Myogenic-Determination Genes Encodes a Transiently Expressed Nuclear Protein Marking Primary Myogenic Cells ». *Proceedings of the National Academy of Sciences of the United States of America* 88(9): 3782-86.
200. Persengiev, Stephan P, Laxminarayana R Devireddy, et Michael R Green. 2002. « Inhibition of Apoptosis by ATFx: A Novel Role for a Member of the ATF/CREB Family

of Mammalian BZIP Transcription Factors ». *Genes & development* 16(14): 1806-14.

201. Phipson, Belinda et al. 2016. « Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression ». *Ann. Appl. Stat.* 10(2): 946-63.
202. Piccirillo, Rosanna, Fabio Demontis, Norbert Perrimon, et Alfred L Goldberg. 2014. « Mechanisms of Muscle Growth and Atrophy in Mammals and Drosophila ». *Developmental dynamics : an official publication of the American Association of Anatomists* 243(2): 201-15.
203. Poitras, Jennifer L. et al. 2008. « Novel SSBP2-JAK2 fusion gene resulting from a t(5;9)(q14.1;p24.1) in pre-B acute lymphocytic leukemia ». *Genes, Chromosomes and Cancer* 47(10): 884-89.
204. Potthoff, Matthew J et al. 2007. « Regulation of Skeletal Muscle Sarcomere Integrity and Postnatal Muscle Function by Mef2c ». *Molecular and cellular biology* 27(23): 8143-51.
205. Pourquié, Olivier. 2011. « Vertebrate Segmentation: From Cyclic Gene Networks to Scoliosis ». *Cell* 145(5): 650-63.
206. Prummel, Karin D., Susan Nieuwenhuize, et Christian Mosimann. 2020. « The lateral plate mesoderm ». *Development* 147(12): dev175059.
207. Pryce, Brian A et al. 2009. « Recruitment and Maintenance of Tendon Progenitors by TGFbeta Signaling Are Essential for Tendon Formation ». *Development (Cambridge, England)* 136(8): 1351-61.
208. Raychaudhuri, S, J M Stuart, et R B Altman. 2000. « Principal Components Analysis to Summarize Microarray Experiments: Application to Sporulation Time Series ». *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*: 455-66.
209. Relaix, Frédéric, Didier Rocancourt, Ahmed Mansouri, et Margaret Buckingham. 2005. « A Pax3/Pax7-dependent population of skeletal muscle progenitor cells ». *Nature* 435(7044): 948-53.
210. Renko, Miha et al. 2019. « Rotational Symmetry of the Structured Chip/LDB-SSDP Core Module of the Wnt Enhanceosome ». *Proceedings of the National Academy of Sciences of the United States of America* 116(42): 20977-83.
211. Riechmann, V. et al. 1997. « Control of cell fates and segmentation in the Drosophila mesoderm ». *Development* 124(15): 2915.
212. Rincón-Limas, D E et al. 1999. « Conservation of the Expression and Function of Apterous Orthologs in Drosophila and Mammals ». *Proceedings of the National Academy of Sciences of the United States of America* 96(5): 2165-70.
213. Ritchie, Matthew E et al. 2015. « Limma Powers Differential Expression Analyses for

RNA-Sequencing and Microarray Studies ». *Nucleic acids research* 43(7): e47-e47.

214. Robson, L. G., et S. M. Hughes. 1996. « The Distal Limb Environment Regulates MyoD Accumulation and Muscle Differentiation in Mouse-Chick Chimaeric Limbs. » *Development (Cambridge, England)* 122(12): 3899-3910.
215. Rubin, Gerald M. et al. 2000. « A Drosophila Complementary DNA Resource ». *Science* 287(5461): 2222.
216. Rudnicki, Michael A. et al. 1993. « MyoD or Myf-5 is required for the formation of skeletal muscle ». *Cell* 75(7): 1351-59.
217. Ruiz-Gomez, M. et al. 1997. « Specific muscle identities are regulated by Kruppel during Drosophila embryogenesis ». *Development* 124(17): 3407.
218. Ruiz-Gómez, Mar et al. 2000. « Drosophila Dumbfounded: A Myoblast Attractant Essential for Fusion ». *Cell* 102(2): 189-98.
219. Sadahiro, Taketaro et al. 2018. « Tbx6 Induces Nascent Mesoderm from Pluripotent Stem Cells and Temporally Controls Cardiac versus Somite Lineage Diversification ». *Cell stem cell* 23(3): 382-395.e5.
220. Saga, Yumiko. 2007. « Segmental border is defined by the key transcription factor Mesp2, by means of the suppression of notch activity ». *Developmental Dynamics* 236(6): 1450-55.
221. Sambasivan, Ramkumar et al. 2009. « Distinct Regulatory Cascades Govern Extraocular and Pharyngeal Arch Muscle Progenitor Cell Fates ». *Developmental Cell* 16(6): 810-21.
222. Sandmann, Thomas et al. 2006. « A Temporal Map of Transcription Factor Activity: Mef2 Directly Regulates Target Genes at All Stages of Muscle Development ». *Developmental Cell* 10(6): 797-807.
223. Sandmann, Thomas et al. 2007. « A Core Transcriptional Network for Early Mesoderm Development in Drosophila Melanogaster ». *Genes & development* 21(4): 436-49.
224. Sandonà, Dorianna et al. 2012. « Adaptation of Mouse Skeletal Muscle to Long-Term Microgravity in the MDS Mission ». *PloS one* 7(3): e33232-e33232.
225. Sawyer, Jacob M et al. 2010. « Apical Constriction: A Cell Shape Change That Can Drive Morphogenesis ». *Developmental biology* 341(1): 5-19.
226. Scaal, M. et al. 1999. « SF/HGF Is a Mediator between Limb Patterning and Muscle Development. » *Development (Cambridge, England)* 126(21): 4885-93.
227. Schaub, Christoph, Hideyuki Nagaso, Hong Jin, et Manfred Frasch. 2012. « Org-1, the Drosophila ortholog of Tbx1, is a direct activator of known identity genes during muscle specification ». *Development* 139(5): 1001.
228. Schnorrer, Frank et al. 2010. « Systematic Genetic Analysis of Muscle Morphogenesis

and Function in Drosophila. » *Nature* 464(7286): 287-91.

229. Schuster-Gossler, Karin, Ralf Cordes, et Achim Gossler. 2007. « Premature Myogenic Differentiation and Depletion of Progenitor Cells Cause Severe Muscle Hypotrophy in Delta1 Mutants. » *Proceedings of the National Academy of Sciences of the United States of America* 104(2): 537-42.
230. Schweitzer, Ronen et al. 2001. « Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments ». *Development* 128(19): 3855.
231. Seale, Patrick et al. 2000. « Pax7 Is Required for the Specification of Myogenic Satellite Cells ». *Cell* 102(6): 777-86.
232. Senti, K., K. Keleman, F. Eisenhaber, et B. J. Dickson. 2000. « Brakeless Is Required for Lamina Targeting of R1-R6 Axons in the Drosophila Visual System. » *Development (Cambridge, England)* 127(11): 2291-2301.
233. Shaw, H. M., et M. Benjamin. 2007. « Structure-Function Relationships of Entheses in Relation to Mechanical Load and Exercise. » *Scandinavian journal of medicine & science in sports* 17(4): 303-15.
234. Shih, Hung Ping, Michael K. Gross, et Chrissa Kioussi. 2007. « Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch ». *Proceedings of the National Academy of Sciences* 104(14): 5907.
235. Shimojima, Tsukasa et al. 2003. « Drosophila FACT Contributes to Hox Gene Expression through Physical and Functional Interactions with GAGA Factor ». *Genes & development* 17(13): 1605-16.
236. Shutova, Maria, Changsong Yang, Jury M Vasiliev, et Tatyana Svitkina. 2012. « Functions of Nonmuscle Myosin II in Assembly of the Cellular Contractile System ». *PloS one* 7(7): e40814-e40814.
237. Silver, Jeremy D, Matthew E Ritchie, et Gordon K Smyth. 2009. « Microarray Background Correction: Maximum Likelihood Estimation for the Normal-Exponential Convolution ». *Biostatistics (Oxford, England)* 10(2): 352-63.
238. Smyth, G. 2005. « LIMMA: Linear models for microarray data ». In *Bioinformatics and Computational Biology Solutions using R and Bioconductor, ,* 397-420.
239. Smyth, Gordon K. 2004. « Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. » *Statistical applications in genetics and molecular biology* 3: Article3.
240. Solnica-Krezel, Lilianna. 2005. « Conserved Patterns of Cell Movements during Vertebrate Gastrulation. » *Current biology : CB* 15(6): R213-228.
241. Sperling, Erik A et al. 2013. « Oxygen, Ecology, and the Cambrian Radiation of

Animals ». *Proceedings of the National Academy of Sciences of the United States of America* 110(33): 13446-51.

242. Steinmetz, Patrick R H et al. 2012. « Independent Evolution of Striated Muscles in Cnidarians and Bilaterians ». *Nature* 487(7406): 231-34.
243. Strigini, M., et S. M. Cohen. 1997. « A Hedgehog Activity Gradient Contributes to AP Axial Patterning of the Drosophila Wing. » *Development (Cambridge, England)* 124(22): 4697-4705.
244. Subramanian, Aravind et al. 2005. « Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles ». *Proceedings of the National Academy of Sciences* 102(43): 15545.
245. Szklarczyk, Damian et al. 2019. « STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets ». *Nucleic acids research* 47(D1): D607-13.
246. Taglietti, Valentina et al. 2016. « Nfix Induces a Switch in Sox6 Transcriptional Activity to Regulate MyHC-I Expression in Fetal Muscle ». *Cell reports* 17(9): 2354-66.
247. Tajbakhsh, S. et al. 1996. « Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle ». *Developmental Dynamics* 206(3): 291-300.
248. Tajbakhsh, S. et al. 1998. « Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5 ». *Development* 125(21): 4155.
249. Tajbakhsh, S., et G. Cossu. 1997. « Establishing Myogenic Identity during Somitogenesis. » *Current opinion in genetics & development* 7(5): 634-41.
250. Tajbakhsh, Shahragim, Didier Rocancourt, Giulio Cossu, et Margaret Buckingham. 1997. « Redefining the Genetic Hierarchies Controlling Skeletal Myogenesis: Pax-3 and Myf-5 Act Upstream of MyoD ». *Cell* 89(1): 127-38.
251. Takimoto, M. et al. 1993. « Specific Binding of Heterogeneous Ribonucleoprotein Particle Protein K to the Human C-Myc Promoter, in Vitro. » *The Journal of biological chemistry* 268(24): 18249-58.
252. Talbot, Jared, et Lisa Maves. 2016. « Skeletal Muscle Fiber Type: Using Insights from Muscle Developmental Biology to Dissect Targets for Susceptibility and Resistance to Muscle Disease ». *Wiley interdisciplinary reviews. Developmental biology* 5(4): 518-34.
253. Tani, Shoichiro, Ung-il Chung, Shinsuke Ohba, et Hironori Hojo. 2020. « Understanding paraxial mesoderm development and sclerotome specification for skeletal repair ». *Experimental & Molecular Medicine* 52(8): 1166-77.

254. Tapon, Nicolas et al. 2002. « salvador Promotes Both Cell Cycle Exit and Apoptosis in Drosophila and Is Mutated in Human Cancer Cell Lines ». *Cell* 110(4): 467-78.
255. Taylor, MV. 2006. « Comparison of muscle development in Drosophila and vertebrates. » In *Muscle development in Drosophila*, éd. Helen Sink. Georgetown, TX /New York, NY: Landes Bioscience/Springer, 169-203.
256. Tekle, Yonas I., et Jessica R. Williams. 2016. « Cytoskeletal architecture and its evolutionary significance in amoeboid eukaryotes and their mode of locomotion ». *Royal Society Open Science* 3(9): 160283.
257. Thomas, Gregg W C et al. 2020. « Gene Content Evolution in the Arthropods ». *Genome biology* 21(1): 15-15.
258. Thomas, Helen E., Hendrik G. Stunnenberg, et A. Francis Stewart. 1993. « Heterodimerization of the Drosophila ecdysone receptor with retinoid X receptor and ultraspiracle ». *Nature* 362(6419): 471-75.
259. Torigoi, Emi et al. 2000. « Chip interacts with diverse homeodomain proteins and potentiates Bicoid activity in vivo ». *Proceedings of the National Academy of Sciences* 97(6): 2686.
260. Tripathy, Ratna, Prabhat S. Kunwar, Hiroko Sano, et Andrew D. Renault. 2014. « Transcriptional Regulation of Drosophila Gonad Formation. » *Developmental biology* 392(2): 193-208.
261. Tsai, Shih-Ying et al. 2016. « GAGA Factor, a Positive Regulator of Global Gene Expression, Modulates Transcriptional Pausing and Organization of Upstream Nucleosomes ». *Epigenetics & chromatin* 9: 32-32.
262. Wahl, Matthias B., Chuxia Deng, Mark Lewandoski, et Olivier Pourquié. 2007. « FGF signaling acts upstream of the NOTCH and WNT signaling pathways to control segmentation clock oscillations in mouse somitogenesis ». *Development* 134(22): 4033.
263. Wahlström, G et al. 2001. « Twinfilin Is Required for Actin-Dependent Developmental Processes in Drosophila ». *The Journal of cell biology* 155(5): 787-96.
264. Wang, Hongyang et al. 2019. « Crystal Structure of the LUFS Domain of Human Single-Stranded DNA Binding Protein 2 (SSBP2). » *Protein science : a publication of the Protein Society* 28(4): 788-93.
265. Wang, Hongyang et al. 2020. « Crystal Structure of Human LDB1 in Complex with SSBP2 ». *Proceedings of the National Academy of Sciences of the United States of America* 117(2): 1042-48.
266. Wang, Lixin, Joan S Brugge, et Kevin A Janes. 2011. « Intersection of FOXO- and RUNX1-Mediated Gene Expression Programs in Single Breast Epithelial Cells during

Morphogenesis and Tumor Progression ». *Proceedings of the National Academy of Sciences of the United States of America* 108(40): E803-12.

267. Wang, Xiliang et al. 2021. « Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2 ». *Genomics, Proteomics & Bioinformatics*. https://www.sciencedirect.com/science/article/pii/S1672022921000486.
268. Wang, Y. et al. 2010. « SSBP2 Is an in Vivo Tumor Suppressor and Regulator of LDB1 Stability. » *Oncogene* 29(21): 3044-53.
269. Watt, Fiona M, Michaela Frye, et Salvador Aznar Benitah. 2008. « MYC in Mammalian Epidermis: How Can an Oncogene Stimulate Differentiation? » *Nature reviews. Cancer* 8(3): 234-42.
270. Wei, Qin, Yikang Rong, et Bruce M. Paterson. 2007. « Stereotypic founder cell patterning and embryonic muscle formation in Drosophila require nautilus (MyoD) gene function ». *Proceedings of the National Academy of Sciences* 104(13): 5461.
271. van de Wetering, M. et al. 1997. « Armadillo Coactivates Transcription Driven by the Product of the Drosophila Segment Polarity Gene DTCF. » *Cell* 88(6): 789-99.
272. van de Wetering, M., M. Oosterwegel, K. van Norren, et H. Clevers. 1993. « Sox-4, an Sry-like HMG Box Protein, Is a Transcriptional Activator in Lymphocytes. » *The EMBO journal* 12(10): 3847-54.
273. Williams, J.A., S.W. Paddock, et S.B. Carroll. 1993. « Pattern formation in a secondary field: a hierarchy of regulatory genes subdivides the developing Drosophila wing disc into discrete subregions ». *Development* 117(2): 571.
274. Wilm, Bettina, Richard G. James, Thomas M. Schultheiss, et Brigid L. M. Hogan. 2004. « The Forkhead Genes, Foxc1 and Foxc2, Regulate Paraxial versus Intermediate Mesoderm Cell Fate. » *Developmental biology* 271(1): 176-89.
275. Wu, Haojia, Yuhei Kirita, Erinn L Donnelly, et Benjamin D Humphreys. 2019. « Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis ». *Journal of the American Society of Nephrology : JASN* 30(1): 23-32.
276. Wuehr, Max et al. 2014. « Balance Control and Anti-Gravity Muscle Activity during the Experience of Fear at Heights ». *Physiological reports* 2(2): e00232-e00232.
277. Yamaguchi, Y. et al. 1999. « NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. » *Cell* 97(1): 41-51.
278. Yao, Tso-Pang et al. 1992. « Drosophila ultraspiracle modulates ecdysone receptor function via heterodimer formation ». *Cell* 71(1): 63-72.
279. Yeh, Po-An et al. 2012. « Drosophila Eyes Absent Is a Novel MRNA Target of the

Tristetraprolin (TTP) Protein DTIS11 ». *International journal of biological sciences* 8(5): 606-19.

280. Zacharias, Amanda L, Mark Lewandoski, Michael A Rudnicki, et Philip J Gage. 2011. « Pitx2 Is an Upstream Activator of Extraocular Myogenesis and Survival ». *Developmental biology* 349(2): 395-405.
281. Zappia, Maria Paula, et Maxim V Frolov. 2016. « E2F Function in Muscle Growth Is Necessary and Sufficient for Viability in Drosophila ». *Nature communications* 7: 10509-10509.
282. Zappia, Maria Paula, Alice Rogers, Abul B M M K Islam, et Maxim V Frolov. 2019. « Rbf Activates the Myogenic Transcriptional Program to Promote Skeletal Muscle Differentiation ». *Cell reports* 26(3): 702-719.e6.
283. Zhao, Bin, Karen Tumaneng, et Kun-Liang Guan. 2011. « The Hippo Pathway in Organ Size Control, Tissue Regeneration and Stem Cell Self-Renewal ». *Nature cell biology* 13(8): 877-83.
284. Zhao, Yaxing, Limsoon Wong, et Wilson Wen Bin Goh. 2020. « How to do quantile normalization correctly for gene expression data analyses ». *Scientific Reports* 10(1): 15534.
285. Zhong, Zhen et al. 2011. « SSDP Cofactors Regulate Neural Patterning and Differentiation of Specific Axonal Projections ». *Developmental biology* 349(2): 213-24.
286. Zinin, Nikolay et al. 2014. « MYC Proteins Promote Neuronal Differentiation by Controlling the Mode of Progenitor Cell Division. » *EMBO reports* 15(4): 383-91.