Stony Brook University

Functional Characterization of Mustn1 during Skeletal Myogenesis

A Dissertation Presented

by

Cheng Liu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biomedical Engineering

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Cheng Liu

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Michael Hadjiargyrou Ph.D., Dissertation Advisor,
Associate Professor, Department of Biomedical Engineering,
Stony Brook University

Anil Dhundale Ph.D., Chairperson of Defense,
Research Assistant Professor, Department of Biomedical Engineering,
Stony Brook University

Richard A. Clark M.D.,
Professor, Department of Biomedical Engineering,
Dermatology and Medicine, Stony Brook University

David W. Rowe M.D.,
Professor, Department of Genetics and Developmental Biology,
University of Connecticut Health Center

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Functional Characterization of *Mustn1* during Skeletal Myogenesis

by

Cheng Liu

Doctor of Philosophy

in

Biomedical Engineering

Stony Brook University

2007

Mustn1 (Mustang, Musculoskeletal Temporally Activated Novel Gene) was previously identified as a musculoskeletal specific gene based on its restricted expression. Initial characterization of its expression during bone regeneration and embryogenesis revealed robust activity exclusively in cells that are destined to follow a bone/cartilage/skeletal muscle lineage (i.e. preosteoblasts, proliferating condrocytes, myoblasts and mesenchymal cells). Thus, *Mustn1* can serve as a novel marker for studying the musculoskeletal system. However, due to our limited understanding of its transcriptional regulation and function, the role of *Mustn1* remains elusive. Therefore, extended experiments were designed in order to accomplish the following goals: 1) Identify the transcriptional regulation of

Mustn1 by studying its promoter; 2) Characterize its spatiotemporal expression during embryogenesis by studying *Mustn1*^{PRO}-GFP transgenic mice; 3) Probe its function by RNA interference combined with gene expression profiling in an *in vitro* myogenesis model (C2C12 cells myogenic differentiation); and 4) Identify possible interacting proteins using a yeast two-hybrid approach.

The results show: 1) The minimal promoter for *Mustn1* is essentially constituted of one *AP-1* transcription factor binding site, and it is activated by only three members (*c-Fos*, *Fra-2* and *JunD*) of the *AP-1* family. 2) Based on the expression pattern of the *Mustn1* promoter-driven GFP in the transgenic mice, *Mustn1* is primarily activated during the entire myogenenic process as well as intramembranous and endochondral ossification. Additional examination in adult mice showed its activity in periosteum, endosteum and satellite cells of skeletal muscle. Interestingly, *Mustn1* is also expressed in the endothelium of blood vessels. 3) Knock-down of *Mustn1* expression by RNA*i* in C2C12 cells resulted in complete blockage of myofusion, suggesting that *Mustn1* is critical for this event. 4) Due to the nature of *Mustn1* (small protein, lack of representative motifs, possibly needing modification or alkaline environment for interaction), identification of interacting proteins using the yeast two-hybrid system revealed a variety of possible candidates but none were proved to be specific. Taken together, these experiments have expanded our knowledge of *Mustn1*, especially

in skeletal muscle differentiation and development.

Table of Contents

List of Figures……ix

List of Tables……xi

Acknowledgement……xii

Part I: Background and Significance……1
1.1 Identification and characterization of *Mustn1*……1
1.2 Myogenesis……3
1.2.1 The signaling pathway……3
1.2.2 Myogenic cell lineages……6

Part II: Specific Aims, Hypothesis and Rationales……11
2.1 Specific Aim 1……11
2.2 Specific Aim 2……12
2.3 Specific Aim 3……13
2.4 Specific Aim 4……13

Part III: Promoter Identification and Characterization……15
3.1 Specific Aim1……15
3.2 Materials and Methods……15
3.2.1 Materials……15
3.2.2 Cell culture……15
3.2.3 Quantitative real-time RT-PCR (*q*RT-PCR) ……16
3.2.4 Cloning and construction of the *Mustn1* promoter constructs……17
3.2.5 Site-directed mutagenesis……18
3.2.6 Luciferase activity assay……19
3.2.7 Electrophoretic mobility shift assays (EMSA) ……20
3.2.8 Western blotting……21
3.3 Results……22
3.3.1 Identification, cloning and characterization of the *Mustn1* promoter……22
3.3.2 Transcriptional activity of the *Mustn1* promoter in various cell lines……24
3.3.3 Temporal *Mustn1* expression during C2C12 myogenic differentiation……27
3.3.4 Deletion and mutational analysis of *Mustn1* promoter……29
3.3.5 Confirmation of AP-1 binding by EMSA……33

3.3.6 Western blot analysis of the AP-1 family members....36
3.4 Discussion and Conclusion....37

Part IV: *Mustn1*^{PRO}-GFP Transgenic Mice Analysis....43
4.1 Specific Aim 2....43
4.2 Materials and Methods....43
4.2.1 Materials....43
4.2.2 Cloning and transgenic mice generation....43
4.2.3 Genotyping....45
4.2.4 Embryo collection and sections preparation....46
4.2.5 Fluorescent imaging....46
4.2.6 Hematoxylin and eosin staining....47
4.2.7 Safranin O/fast green staining....47
4.2.8 Immunohistochemistry....48
4.3 Results....49
4.3.1 Spatiotemporal activation of *Mustn1* promoter during embryonic myogenesis....49
4.3.2 Spatiotemporal activation of *Mustn1* promoter during embryonic skeletal system development....55
4.3.3 Spatiotemporal activation of *Mustn1* promoter in other GFP positive cells....61
4.3.4 *Mustn1* promoter activity is co-localized with Pax7 in skeletal muscle....63
4.4 Discussion and Conclusion....65

Part V: Functional Perturbation....74
5.1 Specific Aim 3....74
5.2 Materials and Methods....74
5.2.1 Materials....74
5.2.2 Cell culture....74
5.2.3 RNA*i*....75
5.2.4 *q*RT-PCR....78
5.2.5 Myotube quantification....79
5.2.6 Cell proliferation assay....80
5.2.7 Conditioned culture....80
5.2.8 Immunocytochemistry....81
5.2.9 Microarray....82
5.2.10 Microarray data filtering and clustering analysis....82
5.2.11 Statistical analysis....83
5.3 Results....84
5.3.1 Confirmation of *Mustn1* RNA*i*....84
5.3.2 *Mustn1* RNA*i* has no effect on C2C12 proliferation....85

5.3.3 *Mustn1* is required for C2C12 myogenic differentiation.....87
5.3.4 Myofusion is abolished by *Mustn1* RNA*i* in C2C12 cells.....90
5.3.5 Expression profiling of the impact of silencing *Mustn1*.....94
5.3.6 Microarray data validation.....101
5.4 Discussion and Conclusion.....103

Part VI: Protein-protein Interaction.....108
6.1 Specific Aim 4.....108
6.2 Materials and Methods.....108
6.2.1 Materials.....108
6.2.2 The yeast two-hybrid assay.....108
6.2.3 Yeast plasmid isolation.....109
6.2.4 Co-immunoprecipitation (*Co*-IP) 110
6.3 Results.....111
6.3.1 Yeast two-hybrid assay – Round 1.....111
6.3.2 Yeast two-hybrid assay – Round 2.....113
6.3.3 Yeast two-hybrid assay – Round 3.....115
6.3.4 Confirmation of putative interaction by *co*-IP.....117
6.4 Discussion and Conclusion.....118

Part VII: References.....126

List of Figures

Figure 1. Schematic representation of vertebrate embryonic structure after somitic specification.....7

Figure 2. Molecular markers involved in different stages of myogenesis.....9

Figure 3. Genomic organization of the murine *Mustn1*.....23

Figure 4. *Mustn1* expression in various cell lines and comparison of *Mustn1* promoter activity.....26

Figure 5. Temporal expression of *Mustn1* during C2C12 differentiation.....28

Figure 6. Deletion / mutation analyses of the *Mustn1* promoter.....32

Figure 7. Verification of AP-1 binding and identification of specific family members.....35

Figure 8. Western blotting analyses of the AP-1 family members.....36

Figure 9. Map of Topaz/BlueClaPa.6i vector.....44

Figure 10. *Mustn1* promoter activity during embryonic dorsal trunk muscle development.....52

Figure 11. *Mustn1* promoter activity during embryonic cardiac muscle development.....53

Figure 12. *Mustn1* promoter activity during embryonic tongue muscle development.....54

Figure 13. *Mustn1* promoter activity during embryonic vertebrae development.....58

Figure 14. *Mustn1* promoter activity during embryonic intervertebral disc development.....59

Figure 15. *Mustn1* promoter activity during embryonic rib development.....60

Figure 16. *Mustn1* promoter activity during embryonic vasculature development.....62

Figure 17. *Mustn1* promoter activity in adult GFP transgenic mouse.....63

Figure 18. *Mustn1* promoter is active in satellite cells.....64

Figure 19. Lineage tree of somite and speculated *Mustn1* entry point.....71

Figure 20. Validation of *Mustn1* RNA*i*.....84

Figure 21. Proliferation rate comparisons of the C2C12 parental, control and RNA*i* cells.....86

Figure 22. Morphological characterization of C2C12 during myogenic differentiation.....89

Figure 23. Temporal *myogenin* expression *in vitro*.....92

Figure 24. Temporal *MHC* expression *in vitro*.....93

Figure 25. Microarray data clustering by K-mean.....96

Figure 26. Microarray data confirmation by *q*RT-PCR.....102

Figure 27. Co-IP validation of putative interactions.....118

List of Tables

Table 1. Effects of mutations of *MRF*s and other myogenic genes.....10

Table 2. Primers used for *q*RT-PCR and RT-PCR.....17

Table 3. Primers used for Mustn1 promoter deletion construction and site-directed mutagenesis.....19

Table 4. Oligonucleotides for creating *sh*RNA constructs.....76

Table 5. Primers used for *q*RT-PCR.....79

Table 6. Expression of myofusion-related genes.....98

Table 7. Expression of cell adhesion genes in Cluster 7.....100

Table 8. Putative Mustn1-interacting proteins (Round 1).....112

Table 9. Putative Mustn1-interacting proteins (Round 2).....114

Table 10. Putative Mustn1-interacting proteins (Round 3).....116

Table 11. Summary of the yeast two-hybrid screening.....120

Table 12. Summary of putative Mustn1-interacting nuclear proteins.....121

Acknowledgements

This is the prologue of a Ph.D. dissertation, and epilogue of a doctoral study. It is not only a small manuscript that records what has been accomplished, but more of every piece of knowledge that I have learned, every thing that I have thought about, and every step that I have strived for in my daily life. It is the memory of more than six years of living and studying for one objective, and more than six years of devotion that I shall never forget in my lifetime. Delighted I am, I shall never forget the people that offered great help during this quest, without their passionate guidance and selfless assistance, my fate would for sure become totally different.

First of all, I must give the most sincere gratitude to my advisor, Dr. Michael Hadjiargyrou, a man of superb academic quality as well as one beholding very high altitude of morality. Instead of giving instructions and taking controls of everything – which he could, he continuously encouraged me to think independently and bestowed me the steering wheel to explore my fields of interest, whilst helping me to keep on the right track using his insights at all times, and cheering me up when my mood is dampened. On the other hand, his rigorous standard on scientific reasoning, writing, presenting and more importantly, the attitude toward science and education, had profoundly influenced me. Being

a student of such a great teacher and respectful scientist, I feel truly lucky and honored. Yet, there are more I must say. My appreciation also lies in all the help that he had offered in my daily life. As a foreign student, he had showed exceptional care to me. All the suggestions that he had gave to me, from small tips of living in this country to being an upright human in this world, are all of great value. In all, he is a mentor who uses actions to influence, not just relying on words to instruct. And, there is no doubt that his influence will remain for the years to come. In the last, I also want to thank him for being a great Graduate Program Director for the Department of Biomedical Engineering as well as the Associate Vice President for Research of the university. His voluntary contribution only made our community better.

Secondly, I must acknowledge my committee. Dr. Anil Dhundale, a senior professor, one of the founders of our department, and chair of my dissertation committee, has been with us since my arrival in this department. He had showed great concern toward the progress of my research throughout the whole course of my study. His precise yet kind scientific comments had stimulated me and his sense of humor had inspired me. Besides, special thanks for tipping me off at my qualifying examination. I also greatly appreciate his consent of being a collaborator of my project. Without his effort and expertise in the microarray realm, I would never be able to finish my graduate study. Likewise, I must also

thank Dr. David Rowe, for generously allocating his laboratory resources to generate transgenic mice for us. His collaboration not only made my graduate dissertation possible, but more importantly, paved a new path for our research. Finally, I would like to thank Dr. Richard Clark for being serving in my committee and showing great forgiveness for the trouble that I once caused, as I still feel sorry and grateful. Besides, Dr. Clark's knowledge had expanded my horizon and his optimistic attitude had given me confidence. I only wish I could meet him more often.

I feel also compelling to acknowledge other personnel of my department. Dr. Clinton Rubin, for being our chair since the very beginning and striving to make our department a great place to study and live. Dr. Partap Khalsa, Dr. Weiliam Chen, Dr. Stefan Judex and Dr. Yi-Xian Qin, for having been sharing resources with us during the difficult time and showing concern to my study. It is also grateful to the tireless support from Ms. Anne Marie Dusatko, Ms. Rosemary Gaynor and Ms. Colleen Michael. Their work had allowed me to focus on the research without disruption, and made me feel the warmth of being surrounded by a group of nice and caring person.

Next, my appreciation goes to Dr. Mark Kronenberg and Dr. Xi Jiang in the University of Connecticut Health Center, who had spent their time generating and screening transgenic mice for us. Dr. Kronenberg also put tremendous effort in

answering our questions and catering our needs. The extra consulting service supplied by Dr. Bruce Futcher, Dr. Janet Leatherwood and Hong Wang from the Department of Molecular Genetics and Microbiology of our university is also greatly appreciated. Especially, many, many thanks to Dr. Melissa Monaghan, who had analyzed the microarray data, instructed me on how to use the software and trying all sorts of different parameters in order to optimize the result – while she is pregnant!

I have to give my appreciation to all of my colleagues, too. I thank Dr. Frank Lombardo for his pioneering work on *Mustn1*, together with Dr. David Komatsu, Dr. Emily Huang and Dr. Nan Zhong, who had all shared their invaluable experience, and taught me laboratory techniques. Additional thanks to David for being a great friend who also helped to polish my resume. I thank Robert Gersch, my fellow labmate who had always sat in my back. We had worked together so many times to make each other's project rolling. Besides, his great gift of bringing joy to us had made my journey much easier. Many thanks to Hui Pan for the laboratory supplies and information that she had shared with me, as well as all the great food that she had offered when I miss home. I thank Beney Lee for teaching me mouse husbandry skills; Suzanne Ferreri for aiding in my microarray analysis by writing a Matlab program. Thanks to Dr. Jonathan Chiu, Dr. Yi Xia, Dr. Liqin Xie, Dr. Huijuan Liao, Hui Pan, Kim Luu and Ho-Yan Lam for

sharing their reagents and other laboratory resources.

Finally, I feel very fortunate to have a very understanding wife, who chose to marry me while I was still a student, and willing to spend her life taking care of me. I would not be able to through without her encouragement. I also deeply appreciate my parents, who had always been confident in their son, and supported him with everything.

So, here I am.

Part I: Background and Significance

1.1 Identification and Characterization of *Mustn1*

Our laboratory has previously focused on the transcriptional profiling of bone regeneration in exploring the hypothesis that the temporal and spatial expression of specific molecules underlies the essence of the repair process. Bone regeneration is a complex process that essentially replicates embryonic bone development and involves the orchestrated expression, especially re-activation of a multitude of genes [1-3]. We believe that identification of these genes could reveal further insight into the regenerative process at the molecular level, thus leading to the discovery of novel genes, and eventually the development of more effective treatment for bone fractures.

One of these novel genes, *Mustn1* (*Mustang*, Musculoskeletal Temporally Activated Novel Gene), was identified in our profiling experiments via the combined use of suppressive subtractive hybridization and DNA microarray [4]. *Mustn1* encodes for an 82 amino acid nuclear protein with no homology to any known protein families, although, other vertebrate homologues were also identified. Amino acid sequence analyses only predicted N-myristoylation, N-glycosylation, and phosphorylation sites. Further, nuclear localization of *Mustn1* was

demonstrated using a recombinant GFP-Mustn1 construct where green fluorescence was exclusively found in the nucleus (but not in the nucleolus) of transfected pre-osteoblastic MC3T3 cells [5].

Expression studies revealed that this gene was highly up-regulated during fracture repair. Specifically, *Mustn1* was expressed at very low level in the intact bone, whereas in the callus, it was up-regulated in the early phase of the repair process with peak expression coinciding with intramembranous ossification and the beginning of chondrogenesis [5]. This dramatic up-regulation of *Mustn1* and its very low expression in intact bone suggests that *Mustn1* plays an important role during fracture repair. Further tissue expression analyses revealed that *Mustn1* is expressed exclusively in the musculoskeletal system, especially in adult skeletal muscle and tendon [5].

The spatial localization of *Mustn1* was also investigated in order to determine which specific cell types express this gene. Results showed that in the callus, *Mustn1* was expressed in osteoprogenitor cells of the periosteum, as well as osteoblasts and proliferating chondrocytes [5]. Similar analyses with embryo sections revealed that *Mustn1* is expressed mainly in mesenchymal cells [5].

Despite all of these findings, the transcriptional regulation and function of *Mustn1* remains elusive. Based on *Mustn1*'s unique pattern of expression during bone development and regeneration, its restricted expression to the

musculoskeletal system and its nuclear localization, we speculate that Mustn1 may function as a regulatory protein. This idea is supported by the fact that there are only a few major processes (DNA replication, ribosomal RNA synthesis, transcription and post-transcriptional processing) that occur in the nucleus. Since some of these processes are considered "housekeeping" (replication, *r*RNA synthesis, splicing), if *Mustn1* was involved in these, then its expression would not be so restricted or differentially regulated. Additionally, since Mustn1 does not posses any DNA/RNA binding motif, it may not function as a transcription or splicing factor *per se.* Thus, we conclude that Mustn1 may function as part of a large multimeric complex of gene regulatory proteins. As such, it may interact with one or several other components of a transcription complex, either directly or indirectly, to mediate activation or repression of gene expression. However, these potential functions remain unknown.

1.2 Myogenesis

1.2.1 The signaling pathway

Skeletal myogenesis has been extensively studied both *in vivo* and *in vitro* [6-9]. The most widely used *in vitro* model representing myogenic differentiation, is based on the mouse myoblastic cell line C2C12, previously isolated and

immortalized from a C3H mouse [10]. C2C12 is multipotent cell line which naturally differentiates into myocytes upon confluency in a reduced serum environment [11]. However, supplementation of bone morphogenetic protein-2 (BMP-2) inhibits this process and redirects the cells to an osteogenic pathway [12,13]. Additionally, the cells also respond to insulin and dexamethasone and differentiate into adipocytes [14]. Further, C2C12 cells are fibroblast-like in nature when in undifferentiated state. Upon myogenic stimuli, these cells quickly abort their normal proliferative cycle and initiate the myogenic cascade. The spindle-shaped, mono-nucleated cells start to migrate and align to each other, assuming an elongated morphology in preparation for the fusion. The forthcoming myofusion then involves dissociation of cellular membrane at the focal contact points and longitudinal fusion of their cytoplasm, resulting in the formation of multi-nucleated, fiber-like myocytes (or myotubes) [15].

Aside from these cellular level observations, the molecular events involved in myogenesis (*in vivo*) and myogenic differentiation in cultured environment have also been well characterized and reviewed [16-18]. These events are mainly controlled by a group of transcription factors called myogenic regulatory factors (MRFs). These transcription factors, namely *MyoD*, *Myf5*, *myogenin* and *MRF4*, share a common characteristic in that they all contain basic helix-loop-helix (*b*HLH) motifs, which define them as members of the *b*HLH family [19]. Thus,

myogenesis is a process involving orchestrated activation of these *MRF*s, which in turn, leads to the expression of muscle-specific markers, such as *myosin heavy chain* (*MHC*), *desmin,* and *nebulin*, known as terminal differentiation markers, and indicate the finale of this process.

More specifically, myogenic differentiation is heralded by the up-regulation of *MyoD* upon cues conveyed from the environment (*in vitro*: confluency and low serum; *in vivo*: *TGF-β* and *FGF* family growth factors). In response to these signals, muscle precursor cells, called myoblasts, arrest the normal cell cycles and start to express the genes that are needed for the myogenic differentiation program. *Myf5* is another early myogenic differentiation marker that begins to be up-regulated at this stage. Both *MyoD* and *Myf5* are required for myoblast specification since a double-knockout causes complete absence of skeletal muscle in neonatal mice [20]. However, evidence has also shown that *MyoD* and *Myf5* are functionally redundant as mutation of either one failed to abolish skeletal muscle development in embryos [21,22].

Unlike *MyoD* and *Myf5*, *myogenin* is characterized as a late myogenic differentiation marker, and its genomic inactivation results in severe loss of differentiated skeletal muscle but had no effect on undifferentiated myoblasts [23,24]. Another late marker for myogenic differentiation is *MRF4* (*Myf6*, *herculin*) and it is involved in the maintenance of myotubes [20]. It is also able to restore

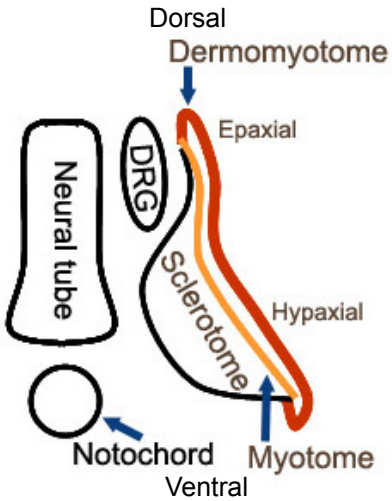
myogenic phenotype in *MyoD*$^{-/-}$:*Myf5*$^{-/-}$ mice [25]. While the expression of *myogenin* and *MRF4* specifies myoblast fate, completion of the differentiation process is indicated by stable expression and massive accumulation of cytoskeletal proteins such as MHC (e.g. myh4) and desmin. On the other hand, it is also known that fibroblasts can take up a myogenic route when *MRFs* are ectopically expressed [26].

1.2.2 Myogenic cell lineages

Several articles have reviewed the process of myogensis *in vivo*, covering both embryonic skeletal muscle formation and adult muscle growth / regeneration [16-18,27,28]. In these reviews, the critical roles of *MRF*s during myogensis are recapitulated since they constitute the essential genes needed for the myogenic progenitor differentiation and myofiber formation. Unfortunately, due to the complicated systemic environment, our understanding of the exact activation, differentiation and self-renewal of these myogenic progenitor cells is still very limited.

One way of unraveling this puzzle is to study the fate map of the skeletal muscle lineages during embryogenesis. It is well established that skeletal muscles share a common somitic origin, except that the head muscles arise from the paraxial head mesoderm and follow a distinct regulatory cascade during later

myogenesis, leading to the formation of epaxial and hypaxial muscle of the neck, laryngopharyngeal muscle as well as tongue [29]. My research is primarily focused on trunk and limb muscle formation, however, documentation on regulation of the latter (head muscle formation) is available [30]. Skeletal myogenesis begins immediately after somitic differentiation and continues throughout the rest of embryonic development as well as postnatal growth. The onset of this event is signified by the formation of the dermomyotome upon signals from the dorsal somite cell, whereas other somite cells migrate toward the ventral side of the embryo and form the sclerotome [31]. These two structures have distinct fates: dermomyotome remains in the dorsal region and gives rise to trunk/limb muscles, endothelia, cartilage, connective tissue and dermis; sclerotome forms the ribs and vertebrae [32]. Embryonic structure at this stage is shown in Figure 1.



http://www.neuro.wustl.edu/

Figure 1. Schematic representation of vertebrate embryonic structure after somitic specification. *Cells located close to the dorsal side of the embryo receive signal from the ectoderm and form dermomyotome. Cells at the ventral side of the somite migrate further ventrally and form sclerotome. Epaxial dermomyotome gives rise to back muscles, whereas hypaxial dermomyotome, also called myotome as the differentiation goes on, gives rise to trunk and limb muscles. Other derivatives of the dermomyotome include cartilage, dermis and connective tissues. Sclerotome is the precursor for ribs and vertebrae. DRG, dorsal root ganglion.*

The signaling cascade involved in skeletal myogenesis is similar to myoblast differentiation *in vitro*. A diagram showing the cascade of the essential genes that are activated during myogenesis is depicted in Figure 2 and their resulting loss-of-function phenotypes are summarized in Table 1. Unlike the cultured myoblasts, initiation of this event is triggered by external stimuli including cytokines (embryonic myogenesis) and mechanical stress (post-natal muscle growth and regeneration) [40,41]. In the embryo, presence of myogenic progenitors can be identified by the expression of molecular markers including *Pax3*, *Pax7* and *Myf5*, all needed to maintain their myogenic potential [16,42]. These progenitor cells continue proliferating during the course of embryogenesis, while a fraction is conserved as a reservoir, the rest enter the myogenic cycle and are committed as myocytes. However, myogenic progenitors diminish toward the end of embryogenesis [42]. Instead, in order to preserve the ability of post-natal muscle growth and repair, a specialized cell type called satellite cells are derived.

Satellite cells are essentially myoblasts that are located between the plasma membrane of the mature myocytes and the basal lamina surrounding each myofiber [43]. They are characterized by the expression of *Pax7* when mitotically quiescent [37]. But, when activated by muscle injury or hypertrophic signals, like the myogenic progenitors, satellite cells will divide to keep a quiescent pool and the rest are programmed to fuse with myofibers [28]. However, there is no

evidence linking satellite cells to the myogenic progenitor lineage and their direct origin is still unknown [28]. Generally it is considered that satellite cells are rooted back into a small fraction of cells in the somite, suggesting all satellite cells are pre-deposited before the end of embryogenesis [44]. However, this is challenged by the discovery that bone marrow cells as well as endothelial cells isolated from the blood vessels can both give rise to satellite cells [45,46]. Interestingly, our data showed *Mustn1* promoter activity in both satellite cells and endothelium, also indicating the existence a secondary source for satellite cells during post-natal development (but remains to be conclusively determined).

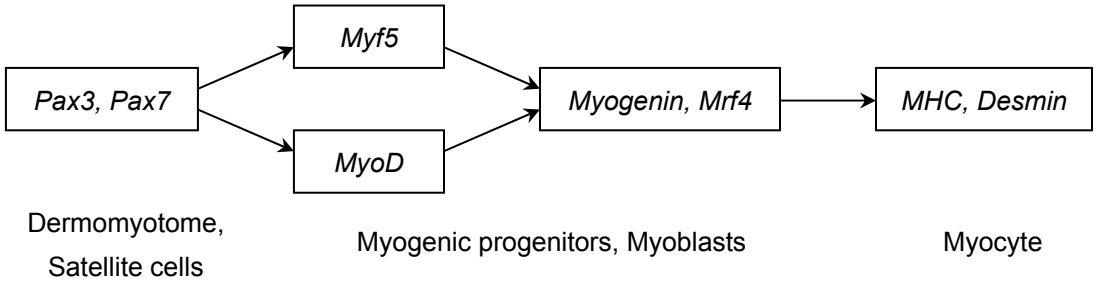


Figure 2. Molecular markers involved in different stages of myogenesis. *Pax3 and Pax7 have similar function in the dermomyotome but not in the satellite cells and their activation is required for the muscle lineage specification. Myf5 and MyoD are early myogenic markers that are needed for the activation of the later markers, myogenin and MRF4. They are expressed at low level in quiescent muscle progenitors and become up-regulated upon stimulation (cells at this stage can take myoblast, osteoblast or adipocyte path in response to different stimuli). The final fate of myocytes is determined by myogenin expression. Terminally differentiated myocytes are characterized by high level of MHC and Desmin expression.*

Table 1: Effects of mutations of *MRF*s and other myogenic genes

Targeted Mutation	*Phenotype*
MyoD$^{-/-}$	*Delayed limb muscle development; increased number of satellite cells; deficient regenerative process [21]*
Myf5$^{-/-}$	*Delayed skeletal muscle development; normal limbs; defective ribs; lethal at birth [22]*
Myogenin$^{-/-}$	*Normal myoblasts; normal MyoD and Myf5 expression; severe loss of skeletal muscle; lethal at birth [23,24]*
MRF4$^{-/-}$	*Normal myoblasts; normal MyoD and Myf5 expression; defective axial myogenesis and rib formation; decreased Myf5 expression [33,34]*
MyoD$^{-/-}$*:Myf5*$^{-/-}$	*Complete absence of myoblasts and skeletal muscle [20]*
Myf5$^{-/-}$*:MRF4*$^{-/-}$	*Delayed skeletal muscle development; defective ribs [34]*
Pax3$^{-/-}$	*Limb muscle absent; normal head muscle [35,36]*
Pax7$^{-/-}$	*Normal embryonic muscle development; satellite cells absent; no post-natal muscle growth [37]*
Pax3$^{-/-}$*:Myf5*$^{-/-}$	*Trunk muscle absent; decreased expression of MyoD [38]*
MyoD$^{-/-}$*:myogenin*$^{-/-}$*:MRF4*$^{-/-}$	*Severe loss of skeletal muscle, including head muscle; lethal at birth [39]*

Part II: Specific Aims, Hypothesis and Rationales

The long term hypothesis for this work is that ***Mustn1 is vital to the musculoskeletal system***. To address this hypothesis, the proposed experiments will be focused on: (1) the transcriptional regulation of the *Mustn1* gene by identification and characterization of its promoter; (2) lineage analysis during embryonic development leading to the formation of the musculoskeletal system using *Mustn1*^{PRO}-GFP transgenic mice; (3) functional perturbation of *Mustn1* with RNA*i* and profiling gene expression with DNA microarray; (4) functional perturbation of Mustn1 by identification of its interacting proteins.

Specific Aim 1: Identify, clone and characterize the *Mustn1* promoter.

Hypothesis*:* Identification and characterization of the *Mustn1* promoter will lead to the identity of the transcription factors responsible for regulating *Mustn1* expression.

Rationale*:* In order to fully understand the significance of *Mustn1* in the musculoskeletal system, it is imperative that we begin by examining its regulation, that is, the identification of sequences required for its expression, and the transcription factors that bind to them. Gene regulation can be investigated by

identifying, isolating and manipulating the *Mustn1* promoter, found upstream of the transcription start site, specifically the promoter *cis*-regulatory elements and transcription factors that bind to them.

Specific Aim 2: Generate *Mustn1*^{PRO}-GFP transgenic mice and analyze skeletal muscle development.

***Hypothesis**: Mustn1*^{PRO}-GFP expression in mice will enable us to map cell lineages involved in musculoskeletal system development.

***Rationale**:* The ability to study specific cell lineages (fate mapping) during development of the mammalian musculoskeletal system is of great importance in enhancing our knowledge of the generation of bone, cartilage, muscle and tendon tissue. Since *Mustn1* is expressed exclusively in the musculoskeletal system (bone, cartilage, muscle, tendon), a worthy and logical approach would be to generate *Mustn1*^{PRO}-GFP reporter transgenic mice, which will give us the ability to *in vivo* visualize, as well as identify and study the birth, migration, proliferation and differentiation (commitment) patterns of musculoskeletal cells and their interactions with each other during embryonic and adult development. This study will focus on bone and skeletal muscle development during embryogenesis. Concomitantly, *Mustn1* promoter activity in adult mice will also be mapped.

Specific Aim 3: Functional perturbation of *Mustn1 in vitro*.

Hypothesis: Silencing of *Mustn1* in myogenic C2C12 cells will affect proliferation and/or differentiation. Gene profiling of this loss-of-function study will enable the discovery of *Mutsn1*-involving biological pathways.

Rationale*:* Since *Mustn1* is expressed in adult skeletal muscle, it is logical to perturb its function during myogenic differentiation. Since the initial discovery includes that *Mustn1* is differentially expressed during the differentiation of myoblasts, the role of *Mustn1* could thus be assessed in muscle precursor cells (C2C12) by determining whether loss (using RNA interference) of *Mustn1* decreases or increases cell proliferation and/or differentiation. These phonotypical alterations, as a result of RNA*i*, are traceable at the *m*RNA level by a microarray experiment. Further analyses of the global expression data should then lead to the elucidation of the biological pathways that enlist *Mustn1* as an essential component.

Specific Aim 4: Identify Mustn1 interacting proteins.

Hypothesis: Mustn1 carries out its function through interacting with other protein(s) which can be identified by a yeast two-hybrid approach *ex vivo*.

Rationale: Understanding a novel gene inevitably includes the knowledge of its functions. *Mustn1* encodes a small protein (82 amino acids) which shows no sign

of canonical motifs that can be used to infer its function. Therefore, it very unlikely carries out important functions on its own. One way of perturbing Mustn1's function is to identify the proteins that interact with it. The yeast two-hybrid system is widely adopted for detecting protein-protein interactions, although limitations exist, it shall still be worthy trying in this scenario.

Part III: Promoter Identification and Characterization

3.1 Specific Aim 1

Identify, clone and characterize the *Mustn1* promoter.

3.2 Materials and Methods

3.2.1 Materials

[γ-^{32}P]-ATP (3000 Ci/mmol) was purchased from Amersham Pharmacia Biotech. Synthetic oligonucleotides were synthesized by Invitrogen Inc. Anti-c-Jun, anti-JunB, anti-JunD, anti-c-Fos, anti-FosB and anti-Fra-2 polyclonal antibodies were purchased from Active Motif. Anti-cyclophilin antibody was purchased from Upstate. Goat-anti-rabbit *Ig*G HRP monoclonal antibody was purchased from Chemicon.

3.2.2 Cell culture

C2C12, NIH3T3, COS-1 and HeLa cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. RCJ3.1C5 cells [47] were maintained in DMEM

supplemented with 15% FBS and 1% penicillin/streptomycin. MC3T3 cells were maintained in minimum essential medium alpha medium (α-MEM) supplemented with 10% FBS and 1% penicillin/streptomycin. All cell lines were cultured at 37°C and 5% CO_2. To induce myogenic differentiation of C2C12 cells, the medium was replaced with DMEM supplemented with 2% horse serum once the cells were confluent.

3.2.3 Quantitative real-time RT-PCR (qRT-PCR)

Proliferating (Day -1) and differentiating (Day 2, 4, 6 and 8) C2C12 cells were collected and total RNA was isolated with RNeasy Mini Kit (QIAGEN) and treated with DNase I (QIAGEN) to remove any traces of DNA. The concentration of each RNA sample was determined by RiboGreen RNA Quantitation Kit (Molecular Probes) following manufacturer's protocol. *q*RT-PCR was carried out with QuantiTect SYBR Green RT-PCR Kit (QIAGEN) on LightCycler system (Roche) as previously described [48]. Temporal expression levels of *Mustn1*, *Myogenin* and *MyoD* were determined by *q*RT-PCR and normalized to that of *GAPDH*. Primer sequences used in these analyses are listed in Table 2. Each experiment was performed in triplicate in order to calculate the standard deviation.

Table 2A: Primers used for qRT-PCR

Target Gene	Accession Number	Primer Sequence	Amplicon Size (bp)	Tm (°C)
GAPDH	*NM_017008*	*Forward: AATGGGGTGATGCTGGTG* *Reverse: GGAAGGGGCGGAGATG*	*119*	*60*
Mustn1	*NM_181390*	*Forward: TGCCCAATGTCCCCAAC* *Reverse: TTCCCTGTCCCACCTCA*	*115*	*60*
Myogenin	*NM_031189*	*Forward: GGAAGTCTGTGTCGGTGGAC* *Reverse: CGCTGCGCAGGATCTCCAC*	*150*	*60*
MyoD	*NM_010866*	*Forward: GCCTGAGCAAAGTGAATGAG* *Reverse: GGTCCAGGTGCGTAGAAGG*	*184*	*60*

Table 2B: Primers used for RT-PCR

Target Gene	Accession Number	Primer Sequence	Amplicon Size (bp)	Tm (°C)
Actin (M,R,H,P)[1]	*M: NM_007393* *R: NM_031144* *H: NM_001101* *P: NM_001009945*	*Forward: AGACCTTCAACACCCCAG* *Reverse: AGGTCCAGACGCAGGATG*	*166*	*60*
Mustn1 (H,P)	*H: NM_205853* *P: NW_104868*	*Forward: AAGAAGAAGCGCCCCCCT* *Reverse: CTTTGGGCTTCTCAAAGAC*	*190*	*60*
Mustn1 (M)	*M: NM_181390*	*Forward: AAGAAGAAGCGGCCCCCT* *Reverse: CTTTGGGCTTCTCAAAGAC*	*190*	*60*
Mustn1 (R)	*R: NM_181368*	*Forward: AAGAAGAAGCGCCCCCCT* *Reverse: GTCTTCGAGAAGCCCAAAG*	*190*	*60*

[1] *M - Mus musculus; R - Rattus norvegicus; H - Homo sapiens; P - Pan troglodytes*

3.2.4 Cloning and construction of the Mustn1 promoter constructs

To clone the *Mustn1* promoter, primers (F(-1447) and R(+65), Table 3A) were designed to amplify the 1,512 bp 5'- *Mustn1*-flanking mouse genomic region using genomic DNA isolated from mouse tail. The PCR amplicon was cloned into pGL3-Basic luciferase vector following manufacturer's instructions (Promega).

Following verification by sequencing, the 1,512 bp sequence was analyzed using Alibaba 2.1 (http://www.gene-regulation.de) in order to identify putative transcription factor binding sites. Based on the regulatory sequences identified by this bioinformatic analysis, we created deleted fragments of the promoter and cloned them into the pGL3-Basic vector. The primers used to generate the various deletion constructs are listed in Table 3A. Finally, all genomic clones were verified by sequencing.

3.2.5 Site-directed mutagenesis

Specific mutations and deletion of the *AP-1* site between -1,151 and -1,161 was accomplished by site-directed mutagenesis using the QuikChange II Site-directed Mutagenesis Kit following the manufacturer's protocol (Stratagene). Specifically, to generate the desired mutation and deletion, a pair of complementary primers was designed for each construct in a way that the sequence of interest is centered and each flanking region contains 13-15 extra oligonucleotides that are identical to the template sequence. Both constructs (L and M) were then amplified / mutated with *PfuUltra* High-Fidelity DNA polymerase using Construct A (Table 3A) as template. Primers used to create these two constructs are listed in Table 3B. Lastly, both mutated and deleted clones were verified by sequencing.

Table 3A: Primers used for the *Mustn1* promoter deletions construction

Target Construct	Primer Name	Primer Sequence	Target Size (bp)	Tm (°C)
A	F(-1447)	Forward: GTCGCTCGAGATGGTGTACTTCCATT	1,512	60
B	F(-1250)	Forward: ATTACTCGAGCCTAGCGTGGTCTA	1,315	60
C	F(-1187)	Forward: ATTACTCGAGCTGGGCATCCCTTATC	1,252	60
D	F(-1132)	Forward: ATTACTCGAGGCATGGCCTGGCCT	1,197	60
E	F(-366)	Forward: TAGTCTCGAGCATCCACCCTTGTTCA	431	60
F	F(-200)	Forward: ATTACTCGAGTAAGCAGCTGTCCCCA	265	60
G	F(-121)	Forward: GGTCCTCGAGAATAAACTCCAGCTAG	186	60
H	F(-60)	Forward: TAATACTCGAGTGACTACCCAGGACG	125	60
I	F(+1)	Forward: ACTACTCGAGATCCTTTCCTGTGGCT	65	60
J	R(-117)	Reverse: CAGTGAATTCTTGGCGATGATGGGCA	388	60
	F(-73)	Forward: TCAGGAATTCCAAAGGAGGGGAGT		
K	R(-286)	Reverse: ACCTGAATTCTGCAAGAACCCATCCC	1381	60
	F(-154)	Forward: TACAGAATTCCTCTCACCAGGGCA		
A-K	R(+65)	Reverse: TTAGCCATGGTGGATGCCAAGCAA	-	60

Table 3B: Primers used for the site-directed mutagenesis

Construct	Application	Primer Sequence
L	AP-1 mutation	Forward: CCTTATCCTTGTCCGCACTAGCCAGCTGTGGG
		Reverse: CCCACAGCTGGCTAGTGCGGACAAGGATAAGG
M	AP-1 deletion	Forward: CATCCCTTATCCTTGTCCTGGGTACTCCTCACAAGG
		Reverse: CCTTGTGAGGAGTACCCAGGACAAGGATAAGGGATG

3.2.6 Luciferase activity assay

Cells used in this study were transfected with each of the *Mustn1* promoter construct, pGL3-Basic vector (promoterless, negative control) and pGL3-Promoter (*SV40* promoter, positive control) (Promega). In a given assay, 8,000 cells from each cell line were plated in designated wells of a 96-well plate in triplicate.

Twenty-four hours later, the cells were transiently transfected using Fugene 6 Reagent (Roche) according to manufacturer's protocol. Forty-eight hours post-transfection, the cells were lysed within the wells and processed using the Steady-Glo Luciferase assay system according to manufacturer's protocol (Promega). Luminescence from each well was measured on a Tropix TR717 Microplate Luminometer (PE Applied Biosystems). For Figure 4, normalized data (Relative Luciferase Activity) is presented as percentage based on the value of the *Mustn1* promoter construct as compared to the value obtained with positive control (pGL3-Promoter). For Figure 6, normalized data (Relative Luciferase Activity) is presented as fold increase based on the value of each construct as compared to the value obtained with the negative control (pGL3-Basic vector).

3.2.7 Electrophoretic mobility shift assays (EMSA)

Nuclear extracts from proliferating (Day -1) and differentiating (Day 6) C2C12 cells were isolated using the Nuclear Extracts Kit (Active Motif) following the manufacturer's protocol. EMSA reactions were prepared by adding the following components: a) ~5-10 µg nuclear extracts from either proliferating (Day -1) or differentiating (Day 6) C2C12 cells, b) binding buffer (10% glycerol, 1 mM $MgCl_2$, 0.5 mM dithiothreitol, 50 mM KCl, 10 mM Tris-HCl, pH 7.4), c) 5 ng/µl poly(*d*I•*d*C), d) 50X wild type unlabeled *AP-1* sequence (*Forward:* 5'-TTGTCCT-

AGTCAGCCAGCTGTG-3', underlined represents wild type *AP-1* site; *Reverse:* 5'-CACAGCTGGCTGACTAGGACAA-3'), e) 50X mutated unlabeled *AP-1* sequence (*Forward:* 5'-TTGTCCTGGTTCGACCGCTGTG-3'; *Reverse:* 5'-CAC-AGCGGTCGAACCAGGACAA-3'), f) individual antibodies specific to AP-1 family members. All reactions were incubated on ice for one hour. Lastly, 1X ^{32}P-ATP labeled wild type probes were added and the reactions were further incubated on ice for 30 min. The protein-DNA complexes were resolved on 4% non-denaturing polyacrylamide gels in 0.25X TBE buffer at 4ºC and visualized by autoradiography. Antibodies used in this assay were included in the Nushift AP-1 Family Kit (Active Motif).

3.2.8 *Western blotting*

Equal amount of protein extracts (same as those used in EMSA analyses) were resolved on 12% SDS-PAGE gels and transferred to nitrocellulose membranes. After blocking with 5% non-fat milk in 1X TBST buffer for 1 hour at room temperature, membranes were incubated with each specific antibody (diluted 1:500 in 1X TBST) overnight at 4°C. The blots were then washed in 1X TBST buffer and probed with goat-anti-rabbit HRP-conjugated IgG secondary antibody (diluted 1:500 in 1X TBST) for 1 hour at room temperature. Antibody binding was visualized with ECL chemiluminescence reagent (Pierce) and then

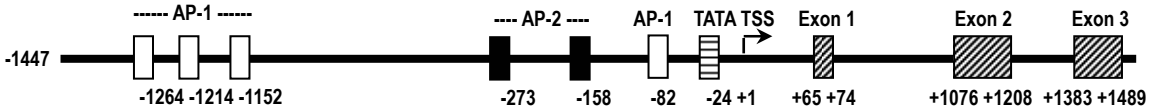
exposed on X-ray film (Kodak). After stripping, the blots were re-probed with anti-cyclophilin antibody (1:1000, Upstate) following the same procedure.

3.3 Results

3.3.1 Identification, cloning and characterization of the Mustn1 promoter

Using the Basic Local Alignment Search Tool (BLAST) in conjunction with the *Mustn1* coding region (NM_181390), we were able to identify *Mustn1's* genomic organization. Alignment of the *Mustn1* coding sequence with the mouse genome database revealed that *Mustn1* resides on chromosome 14. Specifically, *Mustn1's* ORF (open reading frame) is comprised of three exons and two introns (Fig. 3A). A 1,512 bp fragment corresponding to the 5'-flanking region of the gene was selected as the *Mustn1* promoter based on bioinformatic analysis that revealed a translation start codon (ATG), a TATA box, and multiple transcription factor binding sites (Alibaba 2.1, http://www.gene-regulation.de). Notably among them are sequences of activator proteins *AP-1* (4 sites) and *AP-2* (2 sites) (Fig. 3). It is not surprising to find these sites since it is well established that AP-1 and AP-2 factors are key regulators of genes specific to the musculoskeletal system [49-53]. In addition, other transcription factor binding sites were also predicted by the program (not shown).

A



B

```
-1500                                                AGATGGTGTACTTCCATTGTAAACTGGTTTACAGTAAATTTTCACAT

-1400 CTTATTATATAAAAGAAGAAAAAGGAGGAAAAAACCCAGCACTTTTGTGGTCGTTGGTTGTGTGCCCTGTGCCCTGTAACTCACAACCAACTCAACGGGT

-1300 TAATAAACCTGATGGCTTGGCTGCAGTGGGCTCAGAGTCTGCGATGAAAGGAGCCTAGCGTGGTCTAGCCATCATCCTGATGAGCCTAACCCGCTTTTCC
                                  AP-1                                                AP-1

-1200 TCCTGGGTCGTGAGAGCTGGGCATCCCTTATCCTTGTCCTAGTCAGCCAGCTGTGGGTACTCCTCACAAGGCATGGCCTGGCCTCCCCTTGGAATCATGG
                                                 AP-1

-1100 GTATTGCTCAGGTCCCTGCGTATAGACCTGATGGTGCACACACTGACCTACTGGGAGCCAGGCCCAGAACCAAGAATGGGATAGAGTTCGAGGCACAACA

-1000 GCTATGGAATGACATTTAGGATGACTGTTAGTGCCTGGTTGAATTTCATCTGCCTGGTGCCTTCAGGGATAAACACAATGTCCATTGTAGACGGGAGGGA

 -900 GTGGAGAGAAGTCTAAGACAAACCTCCAACCCACATCTACCACAGCCCACCTCAGGCTTAGGGTCCCTGCATTGCCACCAGTGATCTTTCCATGCAAAGA

 -800 TGAAGTTAGTGGTGGGTGCTAAATGATTTTCCAAGGGAGCTCTACTCTTTACTACAGCTTAATAGCTGAAGTGGCATTCACCCCAGCCCTCTAAAGTCCT

 -700 GCCAGCCAGCTTCCAAGTGTCCATCTTAGACAGAAGCCCTCTATAGGCCTTTACTAGCAAGCATGGTCTCCTCCGGGCCTATACTGTACCTTGCTTTGTC

 -600 TCTTTATTAAACTTGTCCCAGGGCTGGCTAGAGGACCAGGACACCAACAAGACTGTGATATGGCAGCTCCACAGGGTCTCGGCCTTCCTGGCTACTCACT

 -500 CTGAGTTGGCTGTGCTCACCATGAGGAAGTGTTTCCAAGCAACAGAAAGGAAGTTTCTCCCCTGCTAATGTGCTGGGATGGAAAATCTATTTCCAGTGGA

 -400 ATCCAGGAGTAGAAGAGAAAGCTAAACAAACTCAGAGCATCCACCCTTGTTCAGGGCCCTCCCTCTCCTCAGAGACTTCCCCTCCCACCCCTGCCACTGG

 -300 GATGGGTTCTTGCAGTCAGCCTGGGCAGGCTGTCTCACACATTCCAGGGTTCACATTCATAAACACAGTCTGTGAACAACAGTGCCAACACCTCCCACCG
                           AP-2

 -200 TCGAGTAAGCAGCTGTCCCCAGAGATGCGGGGTCCCCGCGGCCATCTCCTCTCACCAGGGCAGCTGCCCATCATCGCCAAGAATAAACTCCAGCTAGTCG
                                      AP-2

 -100 CTGCGTGGATCTGAATCATCTGAGAGCTTCCAAAGGAGGGGAGTGACTACCCAGGACGCACAAGCAGACTATATAAGTTCCCAAGGGTTGGGCCTCACGC
            AP-1                                                    TATA Box

   +1 AGATCCTTTCCTGTGGCTACTGCCTGCCAGAGAGCTACCAACAGCGAGCTTGCTTGGCATCCAAC ATG
      TSS
```

Figure 3. Genomic organization of the murine Mustn1. *(A) Genomic organization of the Mustn1 gene and schematic distribution of the transcription factor binding sites of interest. Numbers below each box represent the start position of the site. (B) 5' upstream sequence of the Mustn1 gene and the putative transcriptional factor binding sites. The sequence spans from -1,447 to +65 (1,512 bp in total excluding the translation start codon). Indicated are the AP-1 and AP-2 sites, a TATA box and a transcriptional start site (TSS). The specific sequences corresponding to each transcription factor binding site are all underlined. The TSS is italicized and the translation start codon is in bold.*

3.3.2 Transcriptional activity of the Mustn1 promoter in various cell lines

To study the transcriptional activity of the *Mustn1* promoter, we designed primers (Table 3, Construct A), performed PCR and isolated the 1,512 bp putative promoter sequence from mouse chromosomal DNA. This fragment was subcloned upstream of the firefly luciferase gene in the pGL3-Basic vector (indicated as pGL3-Mus1512 construct) (Fig. 4B). pGL3-Mus1512 construct along with the empty promoterless vector (pGL3-Basic, negative control) and an *SV40*-driven luciferase gene construct (pGL3-Promoter, positive control) were used to transfect the following 6 cell lines: C2C12, RCJ3.1C5, MC3T3, NIH3T3, COS-1 and HeLa. We choose these cell lines because they represent cells where *Mustn1* is known to be expressed (C2C12, myogenic; RCJ3.1C5, chondrogenic; MC3T3, osteogenic and NIH3T3, embryonic fibroblasts) or not (COS-1 and HeLa), as measured by RT-PCR (Fig. 4A). The relative luciferase activity observed from each cell line was normalized to the positive control and presented as a percentage (Fig. 4B). All three lines representing cells of the musculoskeletal system (C2C12, RCJ3.1C5, MC3T3) and NIH3T3 showed high levels of promoter activity (145%, 83%, 28%, and 35%, respectively, as compared to positive control, pGL3-Promoter) consistent with *Mustn1* expression. In contrast, the levels in the other two non-*Mustn1* expressing cell lines (COS-1 and HeLa) were equivalent to those of the negative control (pGL3-Basic) (Fig. 4).

Thus, the promoter activity correlated identically with the expression of *Mustn1*; present in C2C12, RCJ3.1C5, MC3T3 and NIH3T3, but absent in COS-1 and HeLa (Fig. 4).

Further, we observed that the highest levels of luciferase activity in C2C12 cells, corresponded with the higher levels of *Mustn1* expression (Fig. 4), consistent also with our previous results showing that *Mustn1* is highly expressed in adult skeletal muscle [5]. Specifically, luciferase activity driven by the 1,512 bp-*Mustn1* promoter was detected at higher (145%) or almost equal levels (83%) in the myogenic C2C12 and chondrogenic RCJ3.1C5 cells, respectively, as compared to the positive control (pGL3-Promoter), driven by the strong *SV40* viral promoter (Fig. 4B).

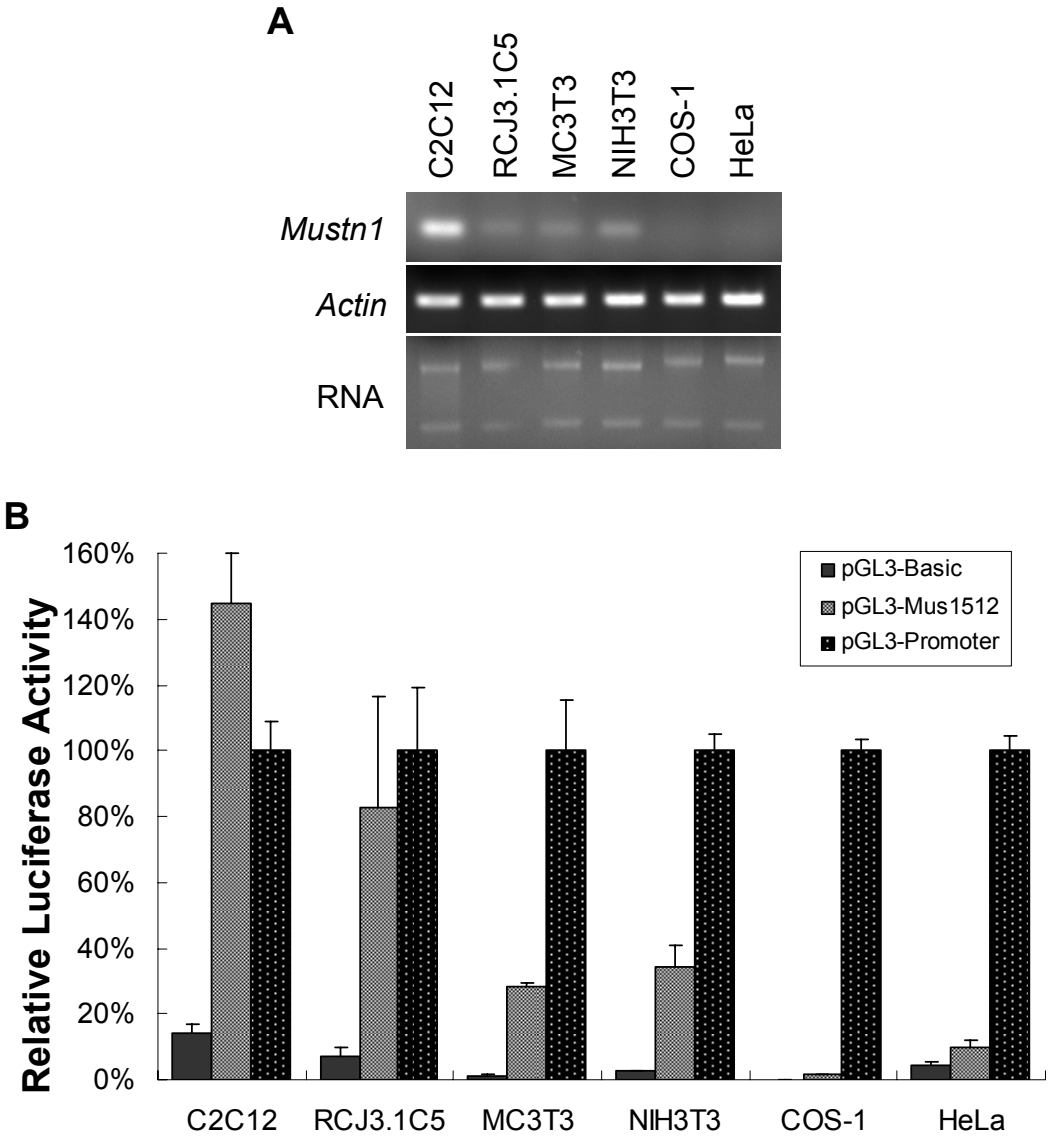


Figure 4. Mustn1 expression in various cell lines and comparison of Mustn1 promoter activity. *Cell lines included in these experiments are: C2C12 (myogenic), RCJ3.1C5 (chondrogenic), MC3T3 (osteogenic), NIH3T3 (embryonic fibroblasts), COS-1 (kidney fibroblasts), HeLa (carcinoma). (A) Mustn1 expression via RT-PCR. Equal amount of RNA was used for each reaction. The cycling was controlled so that all reactions were terminated at the log phase using the same cycle numbers. (B) Lucifierase activity assay of the 1,512 bp Mustn1 promoter-luciferase gene construct in each cell line. pGL3-Basic is the empty vector only (negative control), whereas pGL3-Promoter contains the viral SV40 promoter (positive control). All values were normalized to that of the positive control of each individual cell line and presented as percentage. Error bars indicate standard deviation of triplicate values.*

3.3.3 Temporal Mustn1 expression during C2C12 myogenic differentiation

Based on the aforementioned results, we decided to further characterize the *Mustn1* promoter in C2C12 cells. Thus, we initially analyzed temporal *Mustn1* expression in order to investigate whether there is a direct correlation between its temporal expression and myogenic differentiation. To monitor C2C12 myogenic differentiation we chose to measure *Mustn1* expression levels and those of two myogenesis-specific genes, *MyoD* and *myogenin*, as well as cell morphology. Both *MyoD* and *myogenin* are well established transcription factors and serve as markers for early and late myogenic differentiation, respectively [54,55]. Specifically, using *q*RT-PCR, we analyzed *Mustn1*, *MyoD* and *myogenin* expression using RNA isolated from Day -1 (representing cell proliferation) and Day 2, 4, 6 and 8, (representing various differentiation stages) (Fig. 5A). Results indicate that *Mustn1* expression was temporally regulated with the highest levels detected at the later stages of differentiation (Day 6 and 8) where the cells have formed distinct multinucleated myotubes (Fig. 5BC). In contrast, *MyoD* and *myogenic* expression, although also temporally regulated, peaked at earlier time points, at Day 4 and Day 6, respectively, and then declined (Fig. 5A).

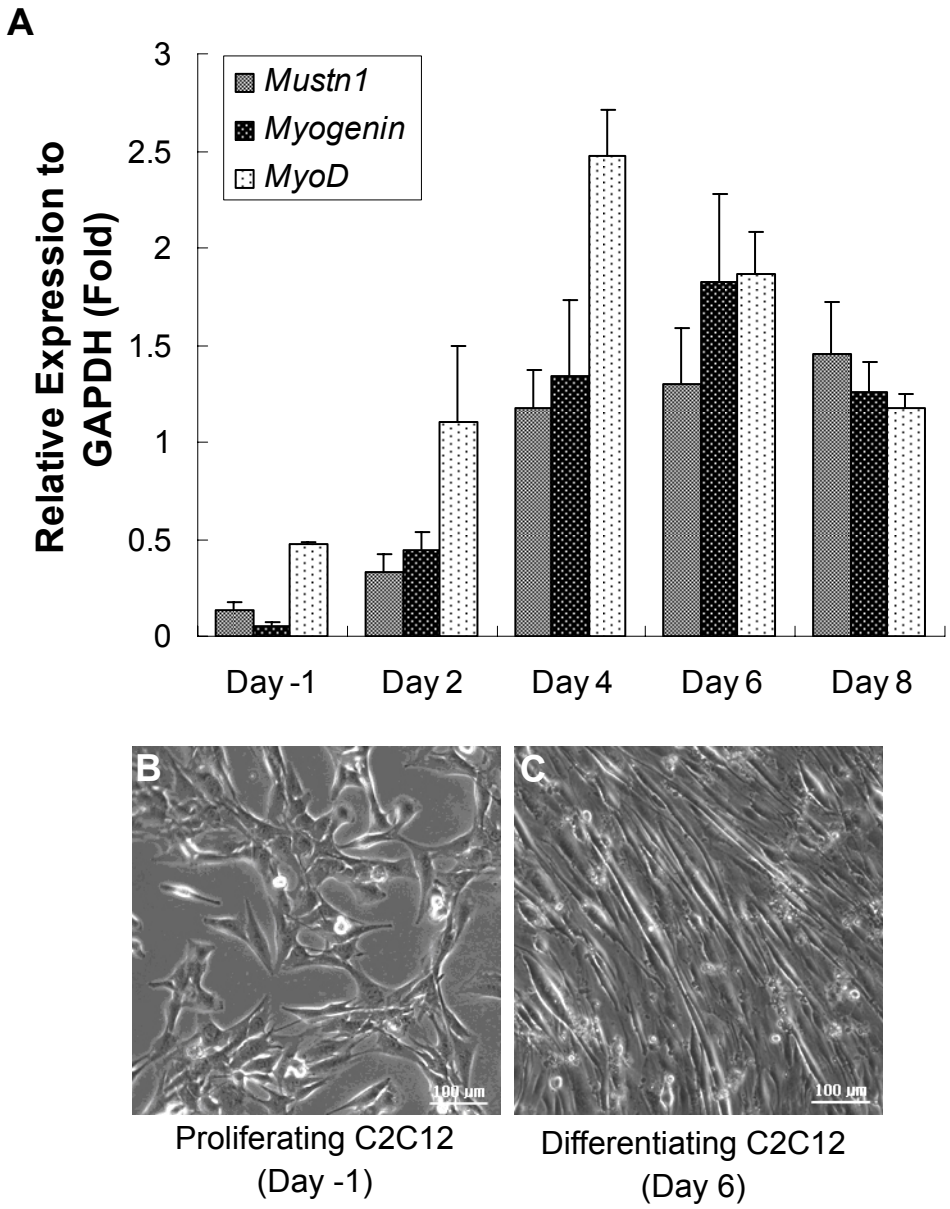


Figure 5. Temporal expression of Mustn1 during C2C12 differentiation. *(A) qRT-PCR of Mustn1, Myogenin and MyoD using RNA from proliferating (Day -1) and differentiating (Day 2, 4, 6 and 8) cells. (B) and (C) actively proliferating (Day -1) and differentiating (day 6) C2C12 cells, respectively. In (C) myotubes, indicative of late differentiation, are also clearly seen.*

3.3.4 Deletion and mutational analysis of Mustn1 promoter

To determine the minimal sequence required for *Mustn1* promoter activity and to define the *cis*-elements responsible for transcriptional activation, we created serially deleted promoter fragments, as well as specific deletion constructs based on the distribution of the *AP-1* and *AP-2* sites within the 1,512 bp promoter region (Fig. 6A). Again, we emphasize that we focused on *AP-1* and *AP-2* because of their involvement in regulating several other musculoskeletal specific genes [49-53]. Eight serially deleted fragments (Construct B-I) and two specific deleted fragments (Construct J and K) were cloned into pGL3-Basic vector (Construct A was the same as pGL3-Mus1512) (Fig. 6). Transient transfection of all constructs followed by luciferase activity assay in C2C12 cells were performed and luciferase activity was measured 48 hours later coinciding with the early stage of myogenic differentiation and *Mustn1*'s up-regulation (see Fig, 3A).

Results revealed maximum luciferase activity with Constructs A-C, indicating that the contribution of the three *AP-1* sites to the transcriptional activation of luciferase (Fig. 6A). In addition, we observe a 40% increase in luciferase activity when the first and second *AP-1* sites are deleted (compare Constructs A-C), suggesting the presence of an inhibitory site within these deleted sequences (Fig. 6A). More importantly, when the third *AP-1* site at -1,151 is deleted (Construct D), luciferase activity decreased by 64.5% (compare Construct

C and D, Fig. 6A). The other *AP-1* and *AP-2* sites showed much lower effect on luciferase activity (Constructs E-G). Further, if all binding sites are deleted (Construct H) as well as the *TATA box* (Construct I), then luciferase activity decreased to the level obtained with the negative control (pGL3-Basic) (Fig. 6A). In addition, we generated two constructs that represent deletions of all *AP-1* (Constructs J) and *AP-2* sites (Construct K). It is clear from these two constructs that deleting both *AP-2* sites only decreases transcriptional activity of the promoter by 12%. In contrast, deleting all four *AP-1* sites (Construct J) reduced the luciferase activity by 73.5%, thus indicating that the *AP-1* sites are the predominant transcription factor binding sites required for maximal promoter activity (Fig. 6A).

Since we observed a dramatic decrease in luciferase activity when the third *AP-1* site at -1,151 along with its adjacent sequences were deleted (Construct D, Fig. 6A), we decided to further analyze this sequence using site-directed mutagenesis (Construct L) and specific deletion (Construct M). Additionally, we wanted to rule out the possibility that there may be other sequences flanking this *AP-1* site that could also contribute to the induction of luciferase. So, the activity of each construct was compared with the wild type *AP-1* (Construct A), as well as those of the negative and positive controls. Results show distinct decreases in luciferase activity with both mutation (40%) and deletion (32%) of the *AP-1* site (Fig.

6B). Combined with the previous serial deletion analyses, these results clearly indicate that AP-1 transcription factors are likely required for the activation of *Mustn1* in differentiating C2C12 myoblasts.

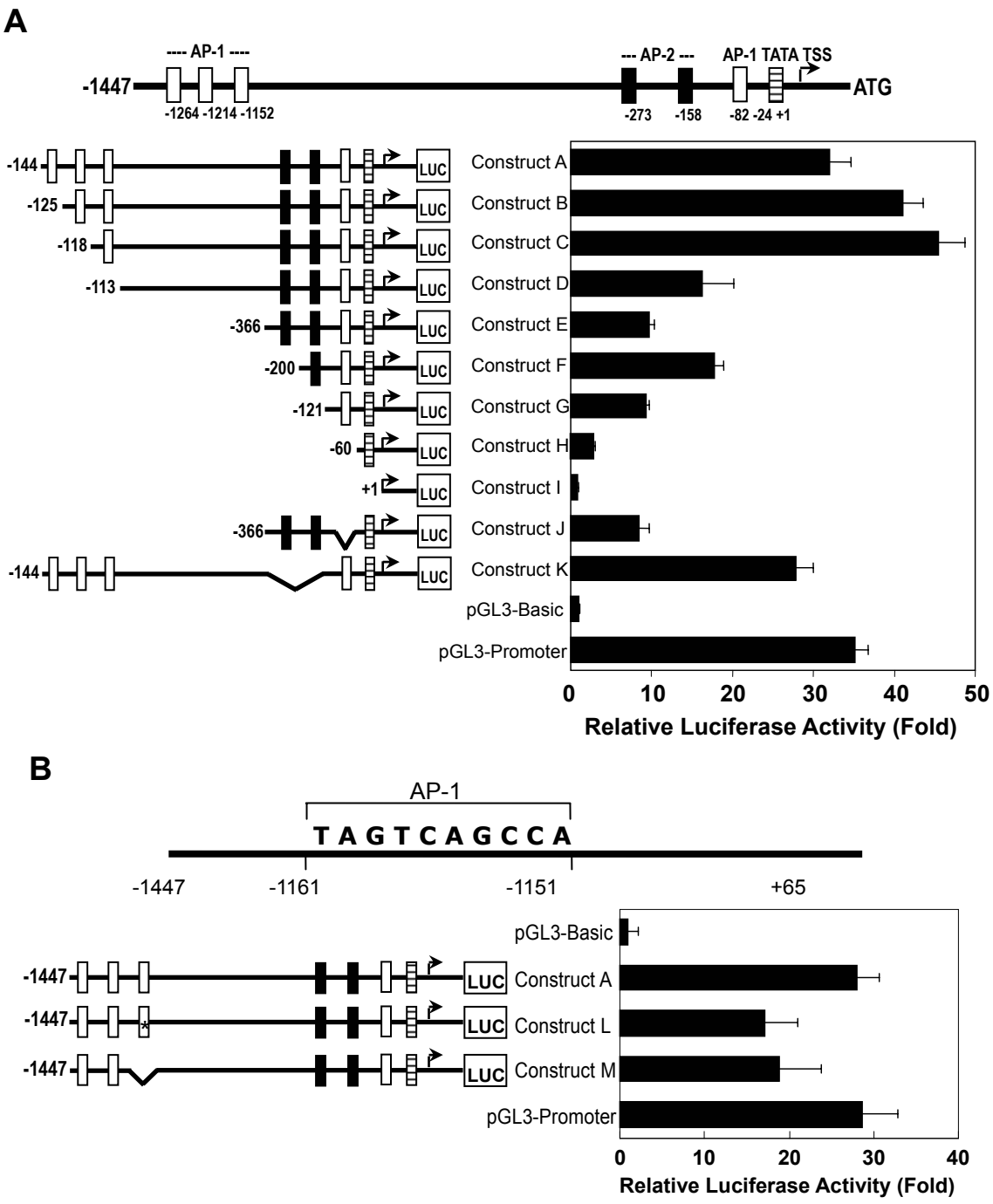


Figure 6. Deletion / mutation analyses of the Mustn1 promoter. *(A) Activity of Mustn1 promoter serially deleted Constructs (A-I) and other specifically deleted Constructs (J: AP-1 deleted; K: AP-2 deleted). C2C12 cells were transfected with equal amount of plasmids containing Constructs A-K. Luciferase activity was measured 2 days after transfection. Promoter activity is reported as fold of each construct over that of the negative control (pGL3-Basic). Error bars indicate standard deviation of triplicate values. (B) Mutation / deletion analysis of the AP-1 site between -1,161 and -1,151. Mutated AP-1 was created by randomly switching the first 5 oligonucleotides to their opposite type (i.e. from purine to pyrimidine or from pyrimidine to purine). Both mutation and deletion was generated by site-directed mutagenesis. Luciferase activity was measured and represented as described for (A).*

3.3.5 Confirmation of AP-1 binding by EMSA

Since the *AP-1* site at -1,151 was identified as the critical regulatory element for *Mustn1* promoter activity, we decided to identify which members of the AP-1 family bind to this site. This was accomplished by using oligonucleotides containing the wild type *AP-1* binding sequence and a mutated sequence in conjunction with nuclear proteins isolated from both proliferating (Day -1) and differentiating (Day 6) C2C12 cells (it has been previously reported that different AP-1 family members are involved in the transition from proliferation to myogenic differentiation [50]), as well as monoclonal antibodies specific to the following AP-1 family members: c-Jun, JunB, JunD, c-Fos, FosB, and Fra-2. Results from these EMSA experiments revealed that AP-1 members bind effectively to the wild type radioactive probes (AP-1 binding site) as indicated in Lane 2 (Fig 5A and B, indicated by band labeled as “shift”). Further, this binding was blocked by the addition of excess (50X) wild type unlabeled probes (Lane 3), while it was not affected by the addition of the same amount of the mutated unlabeled probes (Lane 4). A strong non-specific binding band was also observed, as indicated (Fig. 7AB). Next, in order to identify which AP-1 family members participated in binding to this site, antibodies against c-Jun, JunB, JunD, c-Fos, FosB and Fra-2 were applied. Supershifts (Lane 7, 8 and 10) indicated that JunD, c-Fos and Fra-2 are the only AP-1 members that bind to this AP-1 site (Fig. 7AB). Lastly, no

differences were detected between proliferating and differentiating C2C12 cells (Fig. 7AB).

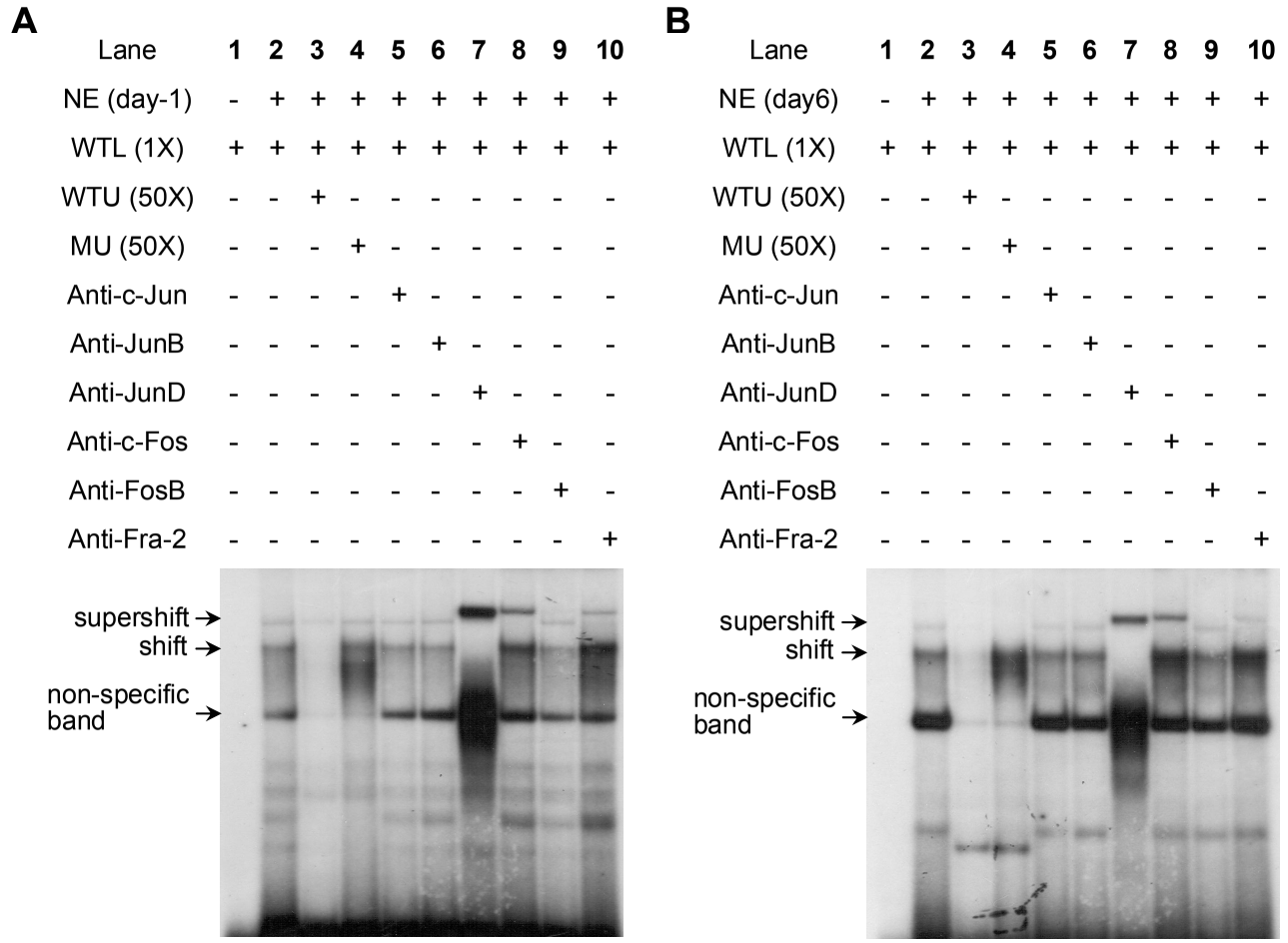


Figure 7. Verification of AP-1 binding and identification of specific family members. *EMSA was performed using C2C12 nuclear extracts (NE) from (A) proliferating cells (Day -1) and (B) differentiating cells (Day 6). Same amount of wild type labeled (WTL) probes were added to each lane. For competition analysis, 50X wild type unlabeled (WTU) probes or mutated unlabeled (MU) probes were also added. Bands that are absent with the addition of excess WTU probe but remain visible with the addition of excess MU probes are shifts caused by AP-1 binding. Supershifts are bands that result from the binding of each specific antibody targeting a particular AP-1 family member. Non-specific bands were also detected.*

3.3.6 Western blot analysis of the AP-1 family members

To further verify the EMSA result, we performed Western blotting analysis to determine the protein expression levels of *c-Jun*, *JunB*, *JunD*, *c-Fos*, *FosB* and *Fra-2* using Day -1 and Day 6 nuclear protein extracts. Results from these analyses showed that *c-Jun*, *JunB*, *JunD*, *c-Fos* and *Fra-2* are expressed during C2C12 proliferation (Day -1) but in differentiating cells only expression of *JunB*, *JunD*, *c-Fos* and *Fra-2* is detected. In contrast, *c-Jun* expression is completely abolished by Day 6 (Fig. 8). No expression was detected for *FosB* at either time point (Fig. 8). These results are also consistent with those obtained from the EMSA analyses in that they show that the three AP-1 members (JunD, c-Fos and Fra-2) that induced supershifts, are also expressed in both proliferating and differentiating cells (Fig. 7AB and Fig. 8). Additionally, two closely-positioned bands were observed when anti-JunD antibody was applied against Day -1 nuclear proteins whereas at Day 6, only one band was detected (Fig. 8).

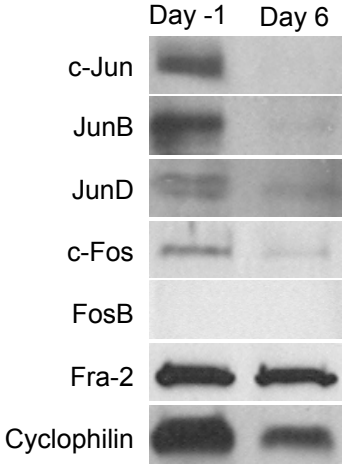


Figure 8. Western blotting analyses of the AP-1 family members. *Western analyses as described in Experimental were used with equal amounts of nuclear extracts from proliferating (Day -1) and differentiating (Day 6) C2C12 cells. For detection of each AP-1 family member, the identical antibodies utilized in EMSA analyses (see Fig. 7) were also used for these western blots. The lower levels of cyclophilin detected is indicative of a general decrease in cellular proliferation, as expected.*

3.4 Discussion and Conclusion

Mustn1 is a novel gene that possesses a unique spatiotemporal expression pattern but whose function remains unknown. Previous analyses showed that its expression was restricted to fracture callus, skeletal muscle, as well as tendon [4]. A temporal analysis of its expression in regenerating fracture calluses indicated that *Mustn1* was up-regulated 55 fold at PF Day 5 when compared to the intact unfractured bone. As the callus proceeded to heal, *Mustn1* expression gradually decreased accordingly and suggested the importance of Mustn1 during the earlier phases of the repair process. More specifically, *in situ* hybridization revealed that *Mustn1* is only strongly expressed in osteoprogenitor cells, young osteoblasts and proliferating chondrocytes of the healing callus. Consistent with the notion that fracture repair recapitulates skeletal development [56], *Mustn1* expression was also detected during embryogenesis, especially in mesenchymal condensations of limbs, vertebral perichondrium, and mesenchymal cells of the intervertebral discs [5]. Provided with these data, we strongly believe that *Mustn1* represents a novel musculoskeletal marker that may play a crucial role in both bone development and regeneration. Thus, identification and characterization of the *Mustn1* promoter will enable us to study the gene's transcriptional regulation.

Based on our bioinformatic analyses, the genomic organization of *Mustn1* reveals that it contains 3 exons separated by 2 introns. A 1,512 bp region

upstream of the 5' end of the first exon was chosen to represent the *Mustn1* promoter due to the presence of classical features, including a transcriptional start site, a *TATA box* and multiple transcription factor binding sites. Some of these sites included *AP-1* sites, which were previously reported to represent regulatory binding sites in many other musculoskeletal specific genes (i.e. *PTHrP*, *myoD*, *beta-catenin*, *osteocalcin* and *collagenase-3*) [57-61]. *AP-2* sites were also found and included in our investigation for their reported activity during embryonic development [53]. Other transcription factor binding sites were also identified through our bioinformatic analyses, however, since our goal was to decipher the link between *Mustn1* and the musculoskeletal system, we chose to focus on the *AP-1* and *AP-2* sites for the reasons mentioned before.

Luciferase activity assays of the full-length *Mustn1* promoter showed that *Mustn1* expression is highly cell type-specific. That is, luciferase activation correlates perfectly with known expression of *Mustn1* in myogenic (C2C12), chondrogenic (RCJ3.1C5), osteogenic (MC3T3) and fibroblastic (NIH3T3) cells. In contrast, in other cell types where *Mustn1* is not expressed (i.e. COS-1 and HeLa), luciferase activation was not detectable. In addition, these luciferase activity assays also indicated that the *Mustn1* promoter represents a very strong promoter, as its ability to activate luciferase in chondrogenic RCJ3.1C5 and myogenic C2C12 cells equal to, or exceeded that of the strong viral *SV40* promoter,

respectively. The strength of the *Mustn1* promoter makes it unique amongst other isolated mammalian promoters, especially in the context of *Mustn1's* highly restricted expression to the musculoskeletal system.

The high level of *Mustn1* expression during C2C12 myogenic differentiation provides an ideal system for studying its regulation. Results from our experiments showed that *Mustn1* underwent prominent and sustained up-regulation during C2C12 differentiation. However, unlike *MyoD* and *myogenin*, two well-known myogenic molecular markers whose expression peaked and then decreased steadily, *Mustn1* expression remains high (at least up to 8 days after cell confluency). Based on these data, as well as the fact that *Mustn1* is highly expressed in terminally differentiated adult skeletal muscle [5], we speculate that *Mustn1* is a significant player in myogenic differentiation and may serve as a late differentiation marker for skeletal muscle cells.

In order to unveil the transcriptional regulation of *Mustn1* expression, we also tried to decipher the information embedded within its promoter. Based on our promoter deletion/mutation data, as well as EMSAs, we determined that the dominant transcriptional regulators for *Mustn1* are specific members of the AP-1 family. This is not surprising since AP-1 has been reported to be a versatile transcription factor that regulates numerous genes involved in a variety of cell types and cellular processes [62]. Interestingly, AP-1 and its family members

(Fos family: c-Fos, FosB, Fra-1 and Fra-2; Jun Family: c-Jun, JunB and JunD) are key regulators of specific genes of the musculoskeletal system and loss-of-function studies showed different degrees of deficiency when specific members were knocked out individually [49].

Selective pairing between the Fos and Jun family members was previously suggested as one source of AP-1's binding versatility [50]. Further, the ATF family proteins which are able to dimerize with certain Fos or Jun proteins added more complexity to this composition-dependant specificity [63]. Another common mechanism of regulating AP-1's specificity is post-translational modification such as phosphorylation and dephosphorylation [64,65]. To date, the full regulatory activity of AP-1 and its family members is still not completely elucidated and our data adds *Mustn1* to the list of AP-1 target genes by showing that *Mustn1* is activated by the binding of c-Fos, JunD and Fra-2 to its promoter.

No compositional change of AP-1 family members was observed between proliferating and differentiating C2C12 cells. It is noteworthy to mention that our Western blots revealed a double band when anti-JunD was applied to the nuclear extracts from proliferating cells, whereas only a single band was detected with nuclear extracts from differentiating cells. This may indicate that during C2C12 differentiation, dephosphorylation of JunD occurs and may have significant functional consequences. Consistent with this idea, a previous study showed

down-regulation of both phosphorylated and dephosphorylated JunD upon C2C12 cell entry in differentiation [50]. So, it is likely that the up-regulation of *Mustn1* expression during the latter phases of C2C12 cell differentiation is connected to JunD dephosphorylation. Further, because Fos family proteins are unable to dimerize with each other unlike the Jun family proteins [50], JunD is then likely to serve as the active and determining component of the AP-1 complex that binds to the *Mustn1* promoter (since JunD was the only member of Jun family detected by EMSA). However, further experiments are required before we can determine conclusively the exact composition of the AP-1 transcriptional complex responsible for regulating *Mustn1* expression during both cell proliferation and differentiation.

While *Mustn1* activation depends on the binding of specific AP-1 family proteins, *MyoD* is negatively regulated by AP-1 through down-regulation of *c-Fos* and *c-Jun* [58]. Further, a dual role of c-Jun (stimulates / represses myoblast differentiation) in regulating myogenin expression has also been reported [66]. So, although evidence has shown that *Mustn1* is tightly linked to myoblast differentiation, its transcriptional regulation could be completely different from that of *MyoD* and *myogenin*. Provided that the process of myogenesis is heavily linked to the *MyoD*-associated signaling pathway, identification of *Mustn1* function could certainly place it within this pathway.

In summary, we have identified and characterized the *Mustn1* promoter. This 1,512 bp DNA sequence contains multiple AP-1 binding sites that activate gene transcription very strongly, especially in cells of the musculoskeletal system. Given that *Mustn1* expression is restricted to the musculoskeletal system, coupled with its high level of expression during development and regeneration, makes its promoter ideal for future studies. Specifically, using *Mustn1* promoter-*GFP* transgenic mice will enable us to characterize the spatiotemporal expression of *Mustn1*, as well as to perform lineage mapping analyses more comprehensively during musculoskeletal development and regeneration.

Part IV: *Mustn1*^{PRO}-GFP Transgenic Mice Analysis

4.1 Specific Aim 2

Generate *Mustn1*^{PRO}-GFP transgenic mice and analyze skeletal muscle development.

4.2 Material and Methods

4.2.1 Materials

The cloning vector (Topaz/BlueClaPa.6i, or Topz/BCP.6i) containing a topaz variant of the green fluorescent protein (GFPtpz) as well the mice (strain *CD-1*) used for generating transgenic mice were provided by Dr. David Rowe at the University of Connecticut Health Center. All animal procedures were performed according to NIH guidelines approved by the Stony Brook University Division of Laboratory Animal Resources (DLAR).

4.2.2 Cloning and transgenic mice generation

The *Mustn1* 1,512 bp promoter sequence (see *Part III*) was amplified by PCR and cloned into the Topz/BCP.6i vector (Fig. 9) using the *Hind III* restriction

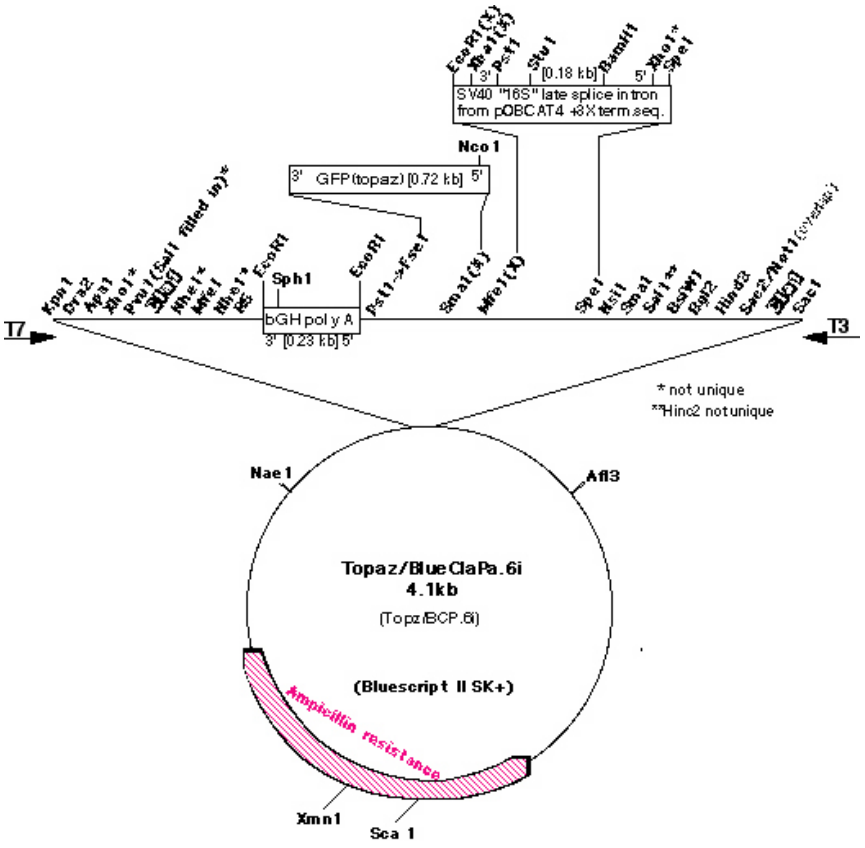


Figure 9. Map of Topaz/BlueClaPa.6i vector. *This 4.1 kb cloning vector encodes a topaz variant of the green fluorescent protein. A spacer sequence is placed between the inserted promoter and GFP (topaz) coding sequence in order to minimize possible interference between these two domains in the recombinant protein. In this experiment, Mustn1 1,512 bp promoter sequence was cloned via the unique restriction site Hind III, with its orientation aligned to that of the GFP (topaz).*

site. Constructed *Mustn1*^{PRO}-GFP plasmid was screened by sequencing for correct promoter insertion. Further, the plasmid was checked for bioactivity by transfecting it into C2C12 cells and looking for GFP expression. Forty-eight hours following transfections, the cells were imaged and GFP expression was clearly visible in the transfected cells (data not shown). Once verified, the construct was linearized and microinjected into mouse oocytes at the UCHC Institutional Gene

Targeting and Transgenic Facility [67]. Positive founders were identified by fluorescent microscopy and confirmed by genotyping.

4.2.3 Genotyping

Genotyping was performed by detecting the presence of GFP sequence in genomic DNA isolated from the founders as well as their *F1* progenies. Genomic DNA was extracted from mouse tail clips according to the following procedure. Initially, each freshly harvested tail clip (~ 1 cm in length) was treated with 600 µl lysis buffer (50 mM Tris, 1.0 mM EDTA, 100 mM NaCl, 0.5% SDS) supplemented with 10 µl proteinase K solution (20 mg/ml proteinase K, 10 mM Tris, 20 mM $CaCl_2$, 50% glycerol) at 55°C overnight. After the tail tissue was digested and cooled down at room temperature for at least 10 minutes, 300 µl 7.5 M ammonium acetate were added into each sample, mixed and centrifuged. The supernatants were then transferred into fresh tubes and well-mixed with 600 µl isopropyl alcohol. DNA pellets were collected by centrifugation and washed with 70% ethanol. The genomic DNA was dissolved in 1X TE buffer for analysis. Genotyping was performed following a common touch-down PCR protocol using annealing temperature of 65°C for the initial 5 cycles followed by 62°C for 30 cycles. The primers used for genotyping were: *forward*: TCATCTGCACCACCG- GCAAGC; *reverse*: AGCAGGACCATGTGATCGCGC, and were designed to amplify a ~250

bp DNA fragment corresponding to the GFP gene.

4.2.4 Embryo collection and sections preparation

In order to characterize the spatiotemporal expression of GFP, positive *F1* progenies were inbred with each other and their embryos were harvested at 12, 15 and 18 days post coitum (dpc). Pregnant females were euthanized with CO_2 and the embryos were surgically removed and immediately fixed in fresh 4% paraformaldehyde for 24 hours. Embryos were then embedded in OCT (Tissue-Tek) and sectioned using a Cryostat (Model: Leica CM3050 S). Multiple adjacent 8 µm-thick sagital sections across the spinal plane were obtained and preserved at -20°C for further analyses. For tail sections, tail snips were harvested from 6 week old *F1* mice (male) and sections were prepared with the same protocol.

4.2.5 Fluorescent imaging

Frozen embryo / tail sections were first blocked with 4% horse serum, the GFP (topaz) signal was then boosted by applying *anti*-GFP antibody conjugated with Alexa Fluro 488 (Invitrogen). A protocol of 1:1000 dilution of the antibody in 4% horse serum and overnight incubation at 4°C was adopted. Following staining, the sections were mounted with cover slips using 50% glycerol in 1X PBS. Green

fluorescence was examined on Axiovert 200M inverted microscope (Carl-Zeiss) under a FITC/Texas Red dual filter set (Chroma, 51006). This filter is capable of differentiating the autofluorescence emitted from skeletal muscle as well as other connective tissue from the GFP signal, which has a very similar absorption-emission spectrum [68]. Under this filter, fluorescent signal is bright green, autofluorescence is brownish-orange, and calcified bone is bright red.

4.2.6 Hematoxylin and eosin staining

Slides containing frozen sections were rinsed in *d*H_2O, followed by staining in hematoxylin for 2 minutes. After rinsing in *d*H_2O again, sections were stained in eosin for 30 seconds. Stained sections were then rinsed in *d*H_2O and dehydrated in ethanol with ascending concentrations (70%, 95%, 100%). Finally, sections were cleared with xylene, and mounted with permanent mounting media (VectaMount, Vector Laboratories).

4.2.7 Safranin O/fast green staining

Safranin O/fast green staining was performed at the Histology Laboratory, School of Medicine, Stony Brook University. Briefly, mouse embryo sections prepared as described above were first stained with hematoxylin for 2 minutes, followed by wash with *d*H_2O. Then the sections were stained with 1% light green

for 3 minutes and rinsed in 1% acetic acid. Lastly, sections were stained with 0.1% safranin O for 5 minutes and rinsed in 1% acetic acid. Stained sections were finally dehydrated in ethanol with ascending concentrations (70%, 95%, 100%) and mounted with permanent mounting media (VectaMount, Vector Laboratories) for observation under light microscope.

4.2.8 Immunohistochemistry

Satellite cells were visualized by staining frozen mouse tail sections (6 weeks old) with *anti*-Pax7 antibody obtained from DHSB (Developmental Studies Hybridoma Bank). First, fixed sections were blocked in 4% horse serum for 1 hour, followed by applying *anti*-Pax7 antibody diluted in 4% horse serum (1:500) overnight at 4°C. Then, appropriate HRP-conjugated secondary antibody diluted in 4% horse serum (1:200) was applied to the sections for 1 hour at room temperature. Finally, all sections were subjected to DAB staining following the manufacturer's protocol (Chemicon). Slides were washed three times in 1X PBS after each step. After the final wash, all slides were mounted with 1X PBS and sealed with nail polish.

4.3 Results

The spatiotemporal distribution of the GFP which represented the activity of the *Mustn1* promoter was studied thoroughly with sagital sections obtained from 12, 15 and 18 dpc (E12, E15 and E18) mouse embryos. GFP signal was detected in a variety of tissues / cells including skeletal muscle and cartilage. Other tissues / cells related to the musculoskeletal system, including tongue muscle and cardiac muscle, however, were not GFP positive. Besides, GFP expression was detected in some non-musculoskeletal tissues / cells, including the endothelial layer of blood vessel and dermis (data not shown). Fig. 10 – 16 show the temporal and spatial localization of GFP in the various tissues / cells.

4.3.1 Spatiotemporal activation of Mustn1 promoter during embryonic myogenesis

Figure 10, 11 and 12 show the temporal activity of *Mustn1* promoter in the dorsal trunk skeletal muscle, cardiac muscle and tongue muscle, respectively. Results show that the GFP expression was robust during embryonic trunk skeletal muscle development, whereas the absence of signal from the cardiac muscle and tongue muscle suggested that *Mustn1* promoter was not activated during the monitored period.

More specifically, during the development of the dorsal trunk muscle,

Mustn1 promoter was highly active at E12 in myogenic progenitors / myofibers, which induced strong GFP expression throughout the cells (Fig. 10D). With the progression of skeletal muscle development at E15, accompanied by the specification and maturation of myofibers, GFP expression was confined to only a few elongated myogenic progenitor-like cells along the nascent myofiber bundles (Fig. 10E). In comparison, at E18, GFP signal was only detectable from the individual oval-shaped cells that were oriented along the mature myofibers, which were completely devoid of the GFP. Additionally, similar cells were observed in the cross sections of ~6 week old mouse tail muscle (Fig. 17A). We suspected that these cells were satellite cells because of their localization between the basal lamina and the cytoplasm of myofibers, and this was proven to be true by a following experiment.

Unlike skeletal muscle, *Mustn1* promoter was inactivated in cardiac muscle at all time points investigated (Fig. 11). Shown in this figure, in E12 embryo, a heart region was specified by the presence of aggregated red blood cells surrounded by a primitive heart envelop (Fig. 11D), which formed thick cardiac muscle mass at E15 (Fig. 11E) and E18 (Fig. 11F). However, close observation of the cardiac muscle showed no sign of GFP, indicating *Mustn1* promoter was inactive during heart development.

The tongue muscle, on the other hand, showed certain resemblance to the

cardiac muscle in that *Mustn1* promoter activity was largely suppressed at all monitored embryonic developmental stages except for a temporary activation at E15 (Fig. 12E). Few GFP-expressing fiber-like myogenic cells were aligned to the matured myocytes at this time, showing an expression pattern that was more similar to E15 developing trunk muscle (Fig. 10E). In contrary, no GFP expression was detected at E12 and E18 (Fig. 12D and F).

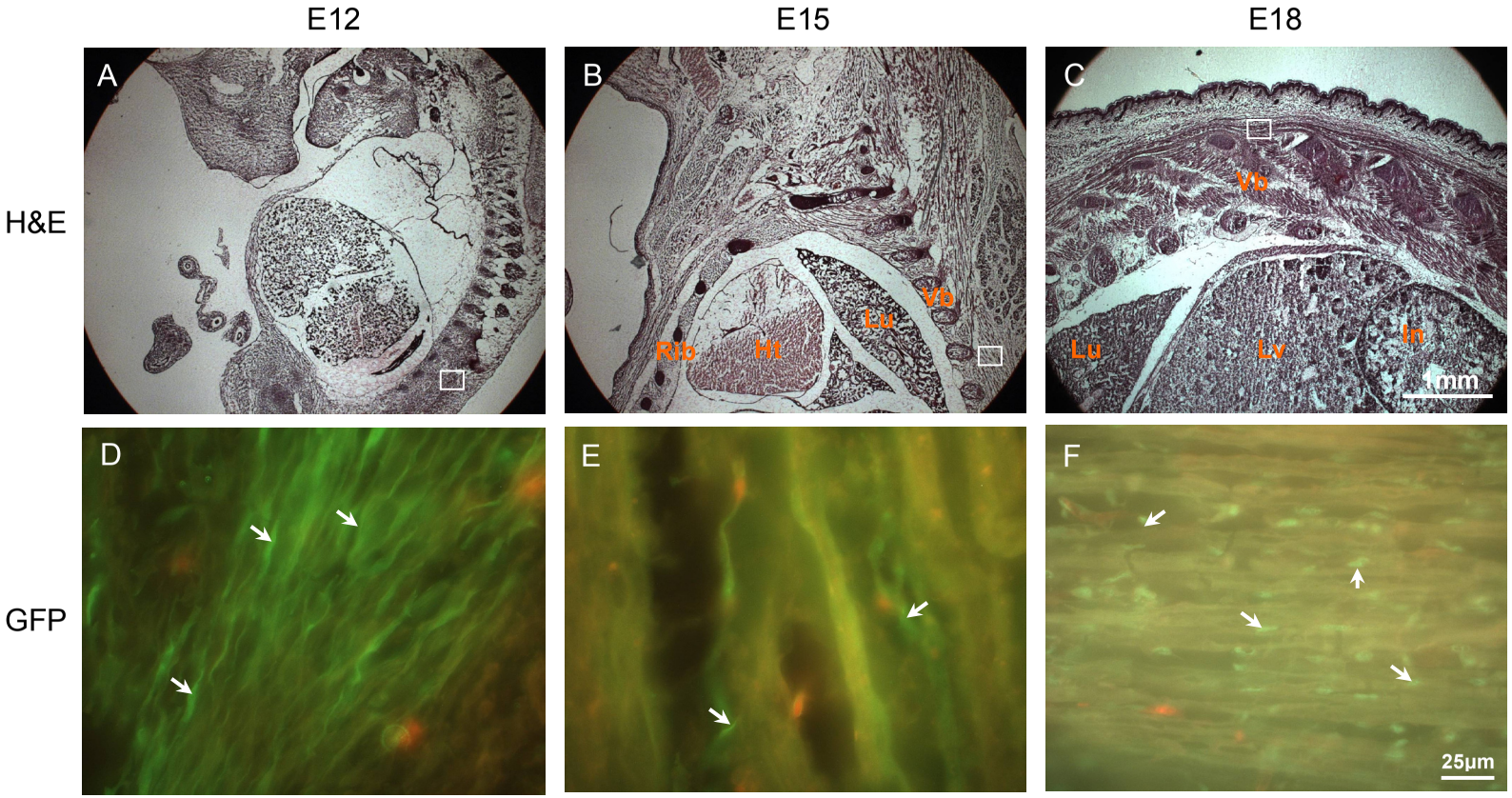


Figure 10. Mustn1 promoter activity during embryonic dorsal trunk muscle development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of H&E stained sections (2.5X) and D, E and F are fluorescence images (63X).White boxes indicate the sources of the images in the lower row, and white arrows indicate GFP-positive cells. Rib, rib; Ht, heart; Lu, lung; Vb, vertebrae; In, intestine.*

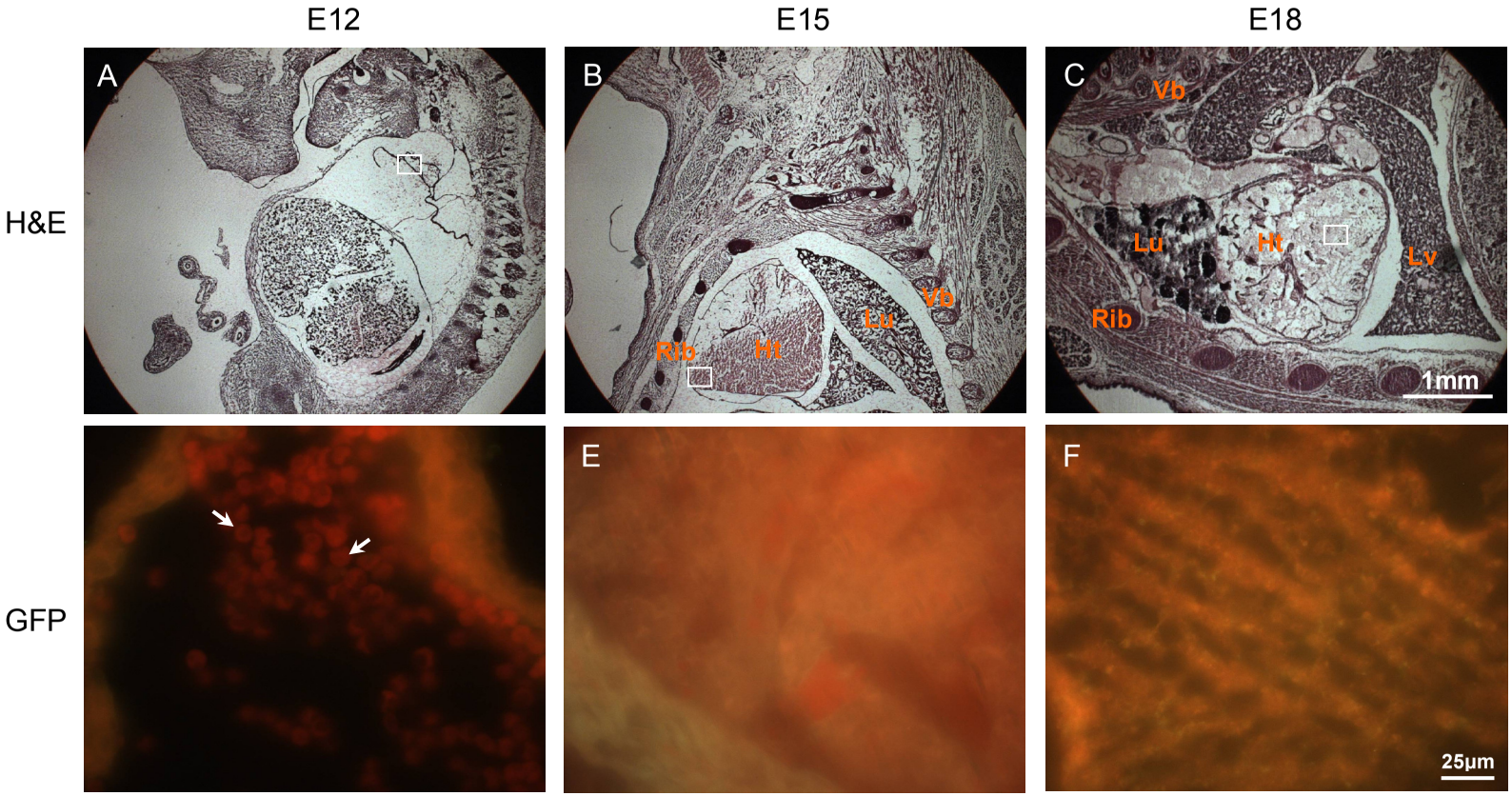


Figure 11. Mustn1 promoter activity during embryonic cardiac muscle development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of H&E stained sections (2.5X) and D, E and F are fluorescence images (63X).White boxes indicate the sources of the images in the lower row and white arrows indicate the red blood cells. Rib, rib; Ht, heart; Lu, lung; LV, liver; Vb, vertebrae.*

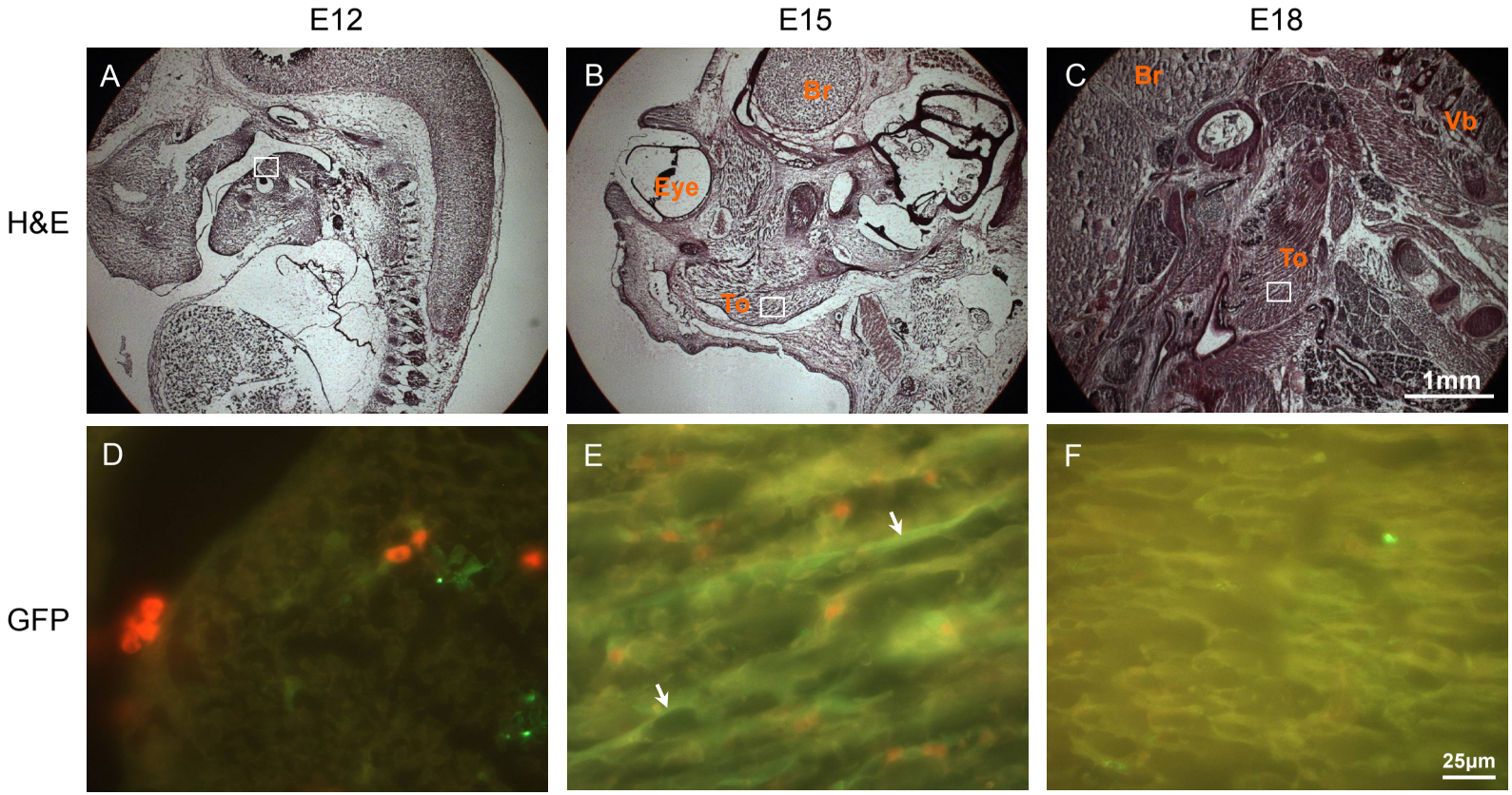


Figure 12. Mustn1 promoter activity during embryonic tongue muscle development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of H&E stained sections (2.5X) and D, E and F are fluorescence images (63X). White boxes indicate the sources of the images in the lower row, and white arrows indicate GFP-positive cells.*

4.3.2 Spatiotemporal activation of Mustn1 promoter during embryonic skeletal system development

In addition to skeletal muscle, a comprehensive examination of the sections revealed that *Mustn1* promoter was also robustly activated during the skeletal development. In fact, results showed that *Mustn1* promoter activity was involved in the entire skeletal system development except cranial skeleton (data not shown). However, only embryonic vertebrae (Fig. 13), intervertebral region (Fig. 14) and the distal end of the ribs (Fig. 15) were selected to demonstrate *Mustn1* promoter activation and GFP expression patterns.

In Fig. 13, the first row of photographs (A, B and C) essentially shows the development of cartilage by safranin O/fast green staining. The presence of cartilage was indicated by red that was primarily found in the center of the cross sections of the nascent bone. As shown, no cartilage was formed at E12 (Fig. 13A). Instead, the longitudinally localized primitive vertebral bodies were clearly visible at this stage and suggested the forthcoming vertebral column. Accordingly, no GFP signal was detected at E12 in the vertebral bodies (Fig. 13D). However, very strong signal was seen between the intervertebral regions which were likely the myogenic progenitors due to their elongated morphology. These cells will go on to form intercostal skeletal muscles and they are shown with greater details in Fig. 14. As the vertebral bodies developed, cartilage was formed at E15,

indicated by the red staining in the vertebrate cavity (Fig. 13B). Fluorescence microscopy showed scattered GFP positive prechondrocytes surrounded by a cortical shell (bright red under the same filter) of the vertebrate (Fig. 13E). Also note the presence of GFP positive cells outside of the cortical shell, which were believed to be skeletal muscle (intecostal) surrounding the vertebrate. The final stage of vertebrate development was signified by the emergence of a large amount of GFP positive cells indicative of proliferating chondrocytes as verified by their red color with safranin O/fast green staining (Fig. 13C).

Fig. 14 shows *Mustn1* promoter activation pattern in the intervertebral region. At E12 the intervertebral region was filled with myogenic progenitors which were aligned to the vertebral bodies (Fig. 14D). However, this region was replaced with GFP negative tissue at E15, leaving only a few GFP positive cells within the intervertebral space (Fig. 14E). Further development revealed a re-activation of the *Mustn1* promoter at E18 as this space was refilled with GFP positive cells, whose identity remains unknown. Note that Fig. 14F also shows a few GFP positive cells (indicated by white arrows) which were different from chondrocytes as they were embedded in the ossified bone, thus likely osteoblasts.

Safranin O/fast green staining showed a very similar pattern of cartilage formation when the distal ribs and vertebrae were compared. Developmentally, formation of the distal end of the ribs was delayed so no structure was visible at

E12 (Fig. 15AD), whereas at E15, rib rudiments were clearly seen at the ventral side of the embryo, in which prechondrocytes are abundant (Fig. 15B) and where *Mustn1* promoter was very active as evident by strong GFP expression (Fig. 15D). Further differentiation was accompanied by the maturation of chondrocytes at E18 in the ribs. At this stage, cross sections of the ribs were shown as red in contrary to the dark brownish color at E15 (Fig. 15C), and bright green fluorescence clearly indicated robust *Mustn1* promoter activity in chondrocytes (Fig. 15F). Notably in both E15 and E18 ribs, cells in the peripheral area were GFP negative (Fig. 15E and F, indicated by yellow arrows), suggesting that *Mustn1* promoter activity declined as the chondrocytes approaching the end of their proliferation / differentiation cycles and became hypertrophic.

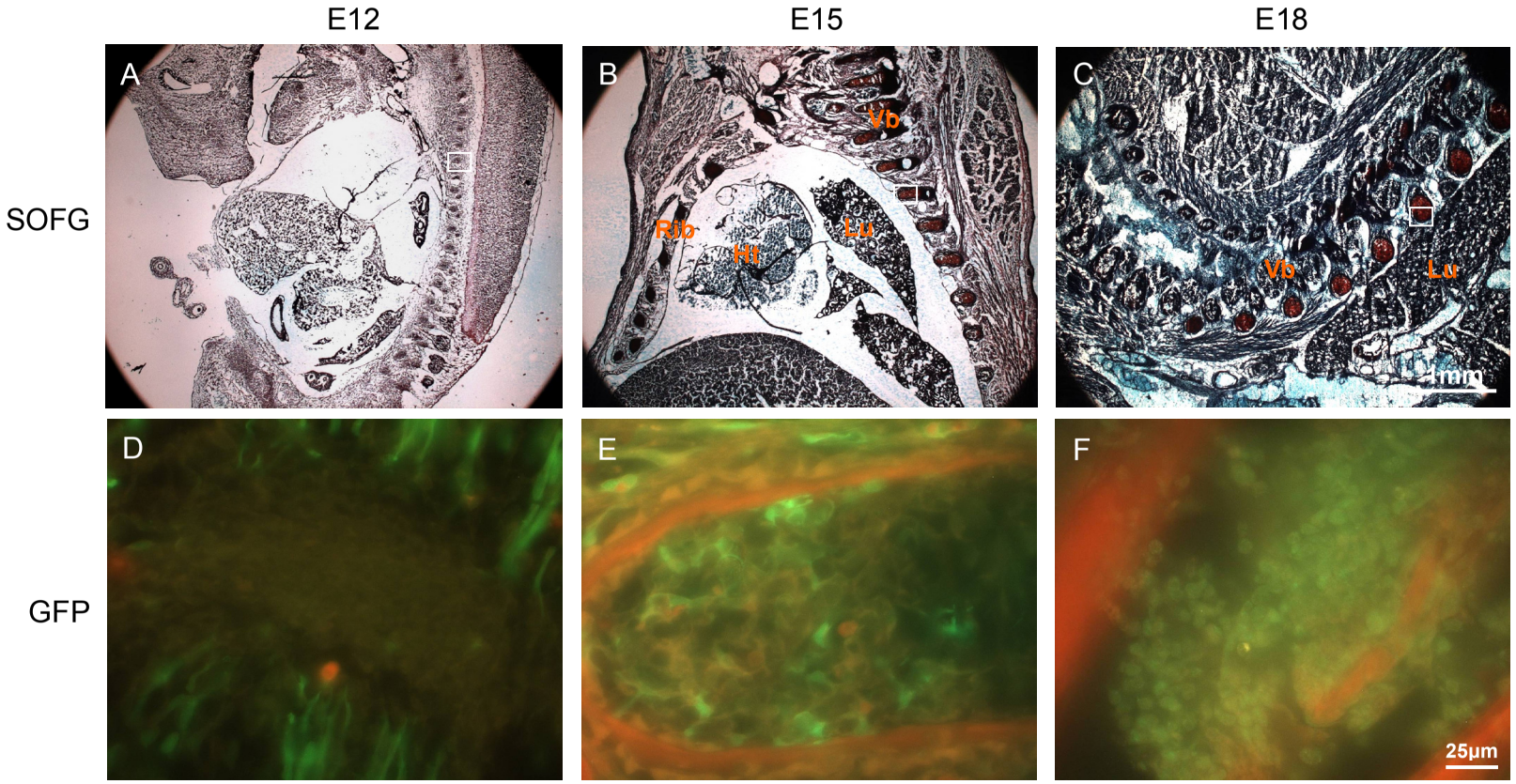


Figure 13. Mustn1 promoter activity during embryonic vertebrae development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of safranin O/fast green stained sections (2.5X) and D, E and F are fluorescence images (63X). White boxes indicate the sources of the images in the lower row. Ht, heart; Lu, lung; Vb, vertebrate; Rib, rib.*

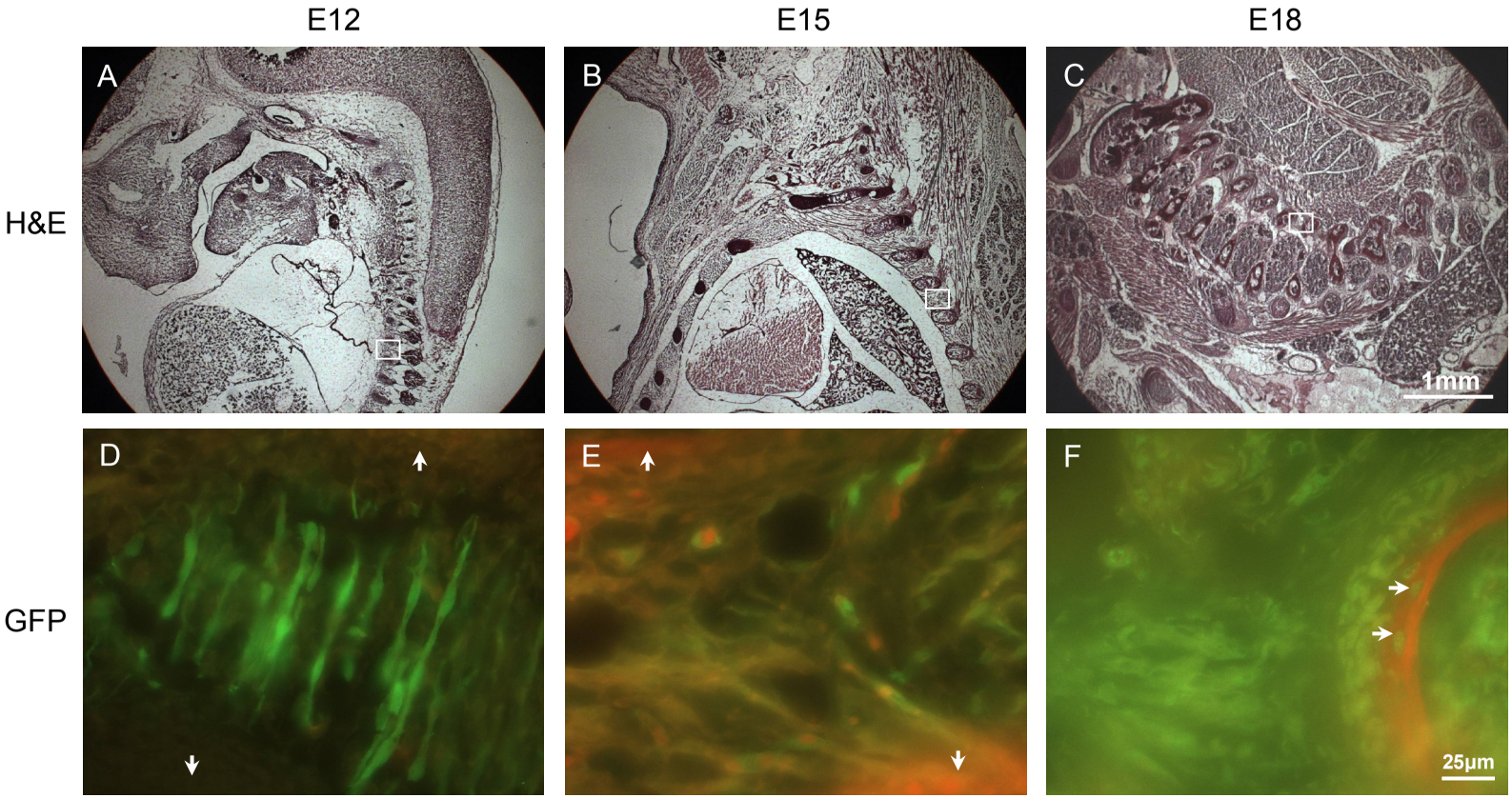


Figure 14. Mustn1 promoter activity during embryonic intervertebral disc development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of H&E stained sections (2.5X) and D, E and F are fluorescence images (63X). White boxes indicate the sources of the images in the lower row. White arrows indicate vertebrate boundary (D) and cortical bone (E and F).*

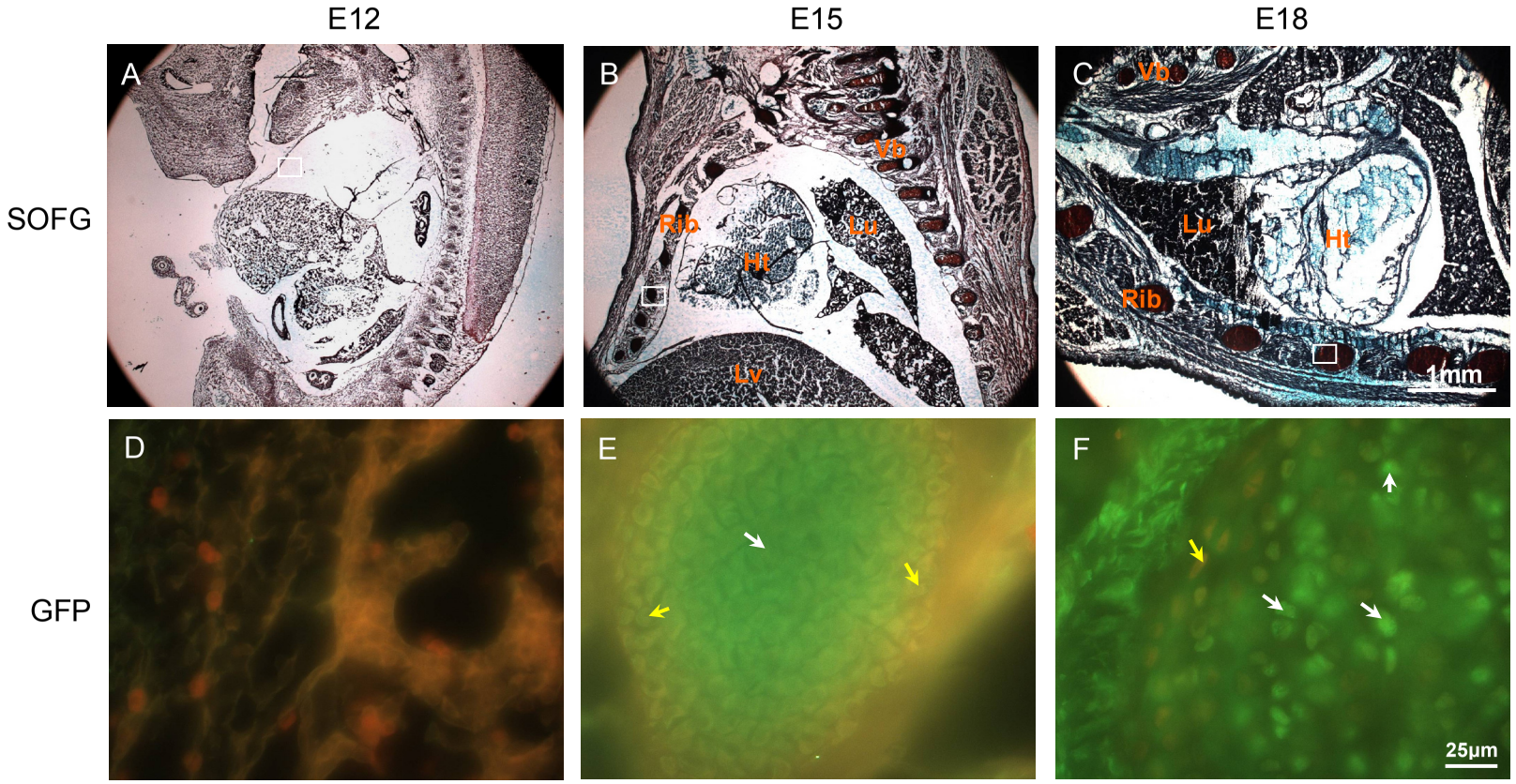


Figure 15. Mustn1 promoter activity during embryonic rib development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of Safranin O/fast green stained sections (2.5X) and D, E and F are fluorescence images (63X). White boxes indicate the sources of the images in the lower row. White arrows indicate proliferating chondrocytes and yellow arrows indicate hypertrophic chondrocytes. Ht, heart; Lu, lung; Lv, liver; Vb, vertebrate; Rib, rib.*

4.3.3 Spatiotemporal activation of Mustn1 promoter in other GFP positive cells

As described in the above sections, *Mustn1* promoter activity was primarily found during the development of the musculoskeletal system of the embryonic trunk. However, GFP expression was not limited to these tissues. We also detected prominent activity of the *Mustn1* promoter in the endothelial cells of blood vessels, as demonstrated in Fig. 16.

Shown in Fig. 16, is expression of GFP that was found consistently at all time points in the inner lining of the blood vessels (i.e. endothelial cells), as indicated by the white arrows (Fig. 16D-F). Moreover, GFP positive endothelial cells were also observed in blood vessels of ~6 week old mouse tail cross sections (Fig. 17B), suggesting that *Mustn1* promoter activity was independent of the age of the animal in these endothelial cells. Besides, weak GFP signal was detected in the outer periphery of the blood vessel (Fig. 16F), mainly smooth muscle and connective tissue. However, further evidence must be obtained to confirm this observation.

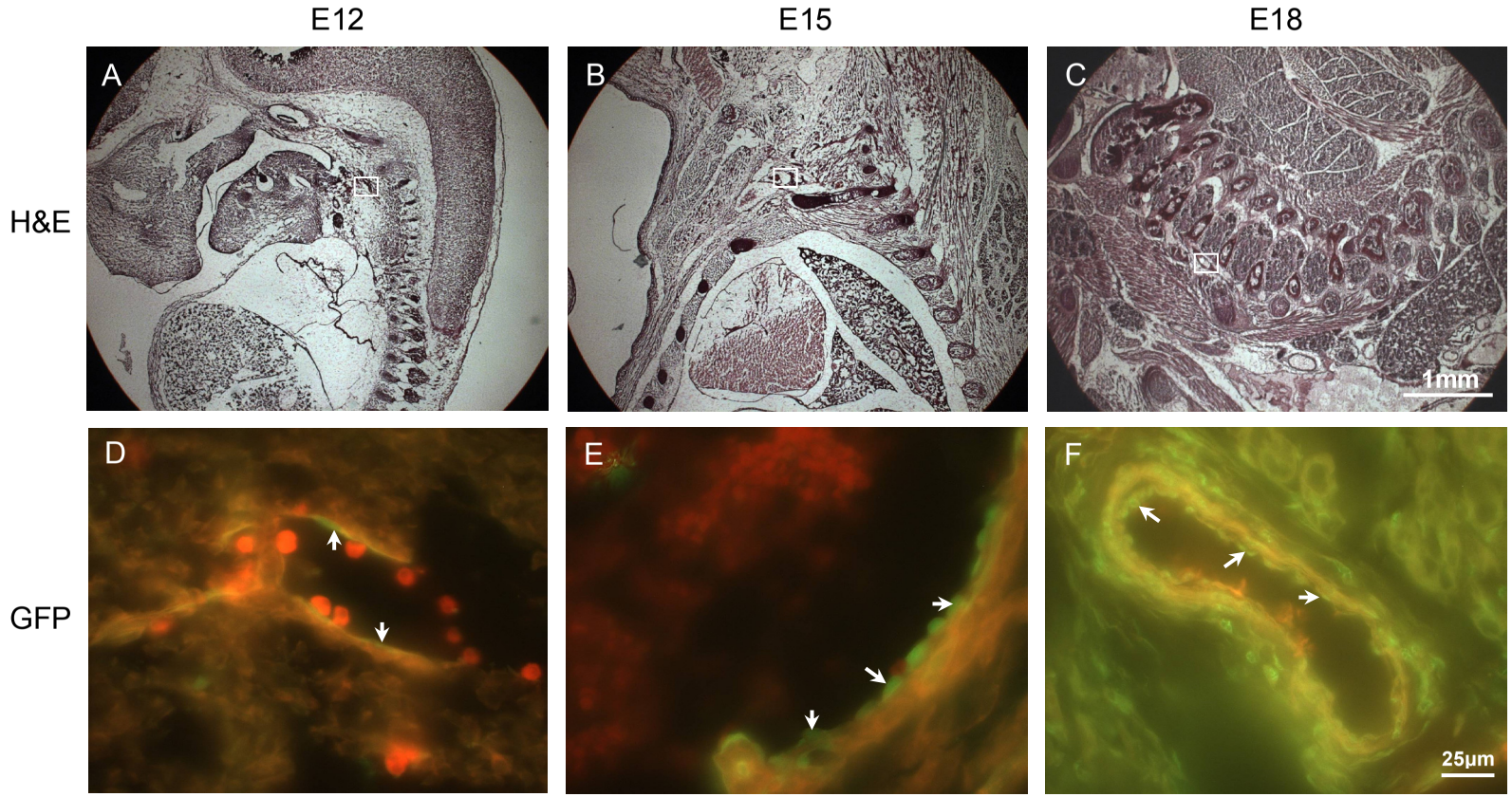


Figure 16. Mustn1 promoter activity during embryonic vasculature development. *Sagittal sections were prepared from E12 (A and D), E15 (B and E) and E18 (C and F) mice embryos. A, B and C are images of H&E stained sections (2.5X) and D, E and F are fluorescence images (63X). White boxes indicate the sources of the images in the lower row.*

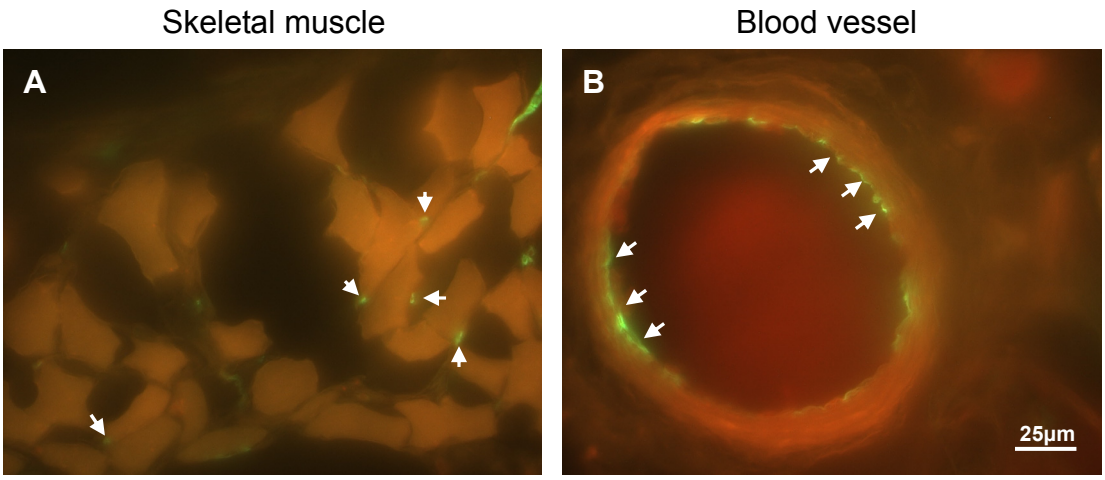


Figure 17. Mustn1 promoter activity in adult GFP transgenic mouse. *Adult mouse (6 weeks old) tail cross sections were stained with anti-GFP antibody. GFP positive cells / tissues are indicated by white arrows. (A) Mustn1 promoter is active in single cells that locate between muscle bundles in the skeletal muscle, likely satellite cells; (B) Mustn1 promoter is active in the endothelial layer of the blood vessel, likely endothelial cells.*

4.3.4 Mustn1 promoter activity is co-localized with Pax7 in skeletal muscle

Since the discovery discussed in *Section 4.3.1* suggested a possible connection between *Mustn1* promoter activity and satellite cells (Fig. 10F), a verification experiment was performed intending to compare the spatial expression pattern of *GFP* and *Pax7*, the latter of which serves as a molecular marker for satellite cells [37,69]. Examination of tail skeletal muscle cross section showed that Pax7 was co-localized with the GFP positive cells (Fig. 18), suggesting that the *Mustn1* promoter was also active in the satellite cells. Specifically, a typical co-localized cell included a dark brown nucleus surrounded by green fluorescent cytoplasm (i.e. Fig. 18, the satellite cell in the center, indicated with an arrow), due to the fact that *Pax7* is a transcription factor while GFP is present throughout the

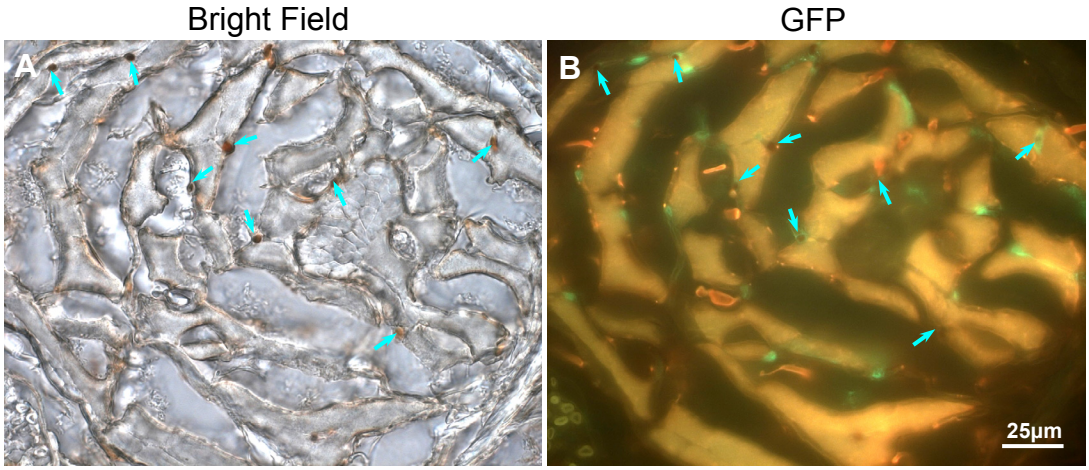


Figure 18. Mustn1 promoter is active in satellite cells. *Adult mouse (6 weeks old) tail cross sections were stained with anti-Pax7 and anti-GFP antibodies simultaneously. Pax7 expression was visualized by DAB staining and observed under bright field microscope (A); GFP expression was observed under fluorescence microscope (B). Pictures of the same area from one of the muscle bundles are shown. (A) Nuclei of the Pax7-expressing cells (satellite cells) are stained in dark brown, as indicated by the cyan arrows. (B) Mustn1 promoter is active in cells showing green fluorescence, and quiescent in orange-colored myofibers. Cyan arrows show the location of the satellite cell nuclei as pointed out in A.*

cytoplasm. Also, since this study was performed on very thin specimens (8 µm thickness), not every satellite cell showed the typical co-localizing characteristics. The majority of the satellite cells showed only one type of staining. In addition, since the probability of sectioning through the cytoplasm was higher than across a nucleus, theoretically more GFP positive cells should be observed than the Pax7 positive cells, which was also the case. Lastly, these results were further verified with *in situ* hybridization experiments where the expression of *Mustn1 m*RNA was also localized to satellite cell in adult skeletal muscle (data not shown).

4.4 Discussion and Conclusion

Although *Mustn1* was initially identified as a novel gene involved in bone regeneration, its pattern of expression in adult tissues was limited to the musculoskeletal system, especially skeletal muscle and tendon [5]. Being dramatically up-regulated in the facture callus, skeletal muscle, tendon, as well as bone and cartilage, it was speculated that *Mustn1* can serve as a musculoskeletal system cell marker. Our primary goal of characterizing the function of *Mustn1* is thus divided to several aspects corresponding to each one of these *Mustn1*-expressing tissues. This study only constitutes part of our laboratory mission by focusing on understanding the role of this gene in the skeletal muscle, because: 1) skeletal muscle is easier to work with when compared to bone tissue and cartilage; 2) *in vitro* and *in vivo* models and protocols of skeletal muscle growth / differentiation / regeneration are well established and easier to work with (based on availability of resources); and 3) the role of *Mustn1* in these tissues, although may be varied, should be largely consistent, thus a study on the skeletal muscle can be vital for the entire quest.

To accomplish the specific aims of this study, the *Mustn1* promoter was first isolated and characterized as reported in *Part III*. Identification of the regulatory mechanism enabled us to further study the gene expression *in vivo* by generating *Mustn1*^{PRO}-GFP transgenic mice. Using these mice as a tool, this study was

specifically designed to depict the spatiotemporal activation pattern of *Mustn1* promoter (via GFP expression) during mouse embryonic development, in the hope that knowing *Mustn1* expression patterns will be helpful in determining its function. Our results showed *Mustn1* promoter activity in a variety of cells, mainly those involved in the development of the musculoskeletal system. However, the focus of this research is regulation and functional characterization of *Mustn1* during myogenesis, thus no in-depth discussion regarding the other *Mustn1*-expressing cells will be covered.

It is generally believed that myogenesis in mouse embryo starts at 8 dpc (E8), during which the somite located between the neural tube and somatopleural mesoderm receives cues from the ectoderm and forms myotome [70]. Further development involves dorsal-ventralization of the myotome under orchestrated stimuli from adjacent tissues (dorsalizing signals: Pax3, Pax7 and Myf5; ventralizing signals: Shh, Noggin and FGFs), resulting in the division between the dermomyotome and sclerotome [71-77]. The sclerotome is located at the ventral side of the structure and gives rise to bone and cartilage, whereas the dorsal-positioned dermomyotome becomes the source of myogenic progenitors for trunk skeletal muscles [17]. Skeletal muscle in the head, on the other hand, takes a different route at the early stage of embryonic development. Head muscle development bifurcates from the rest of the skeletal muscle at the somitic stage:

the most cranial end of the somite receives signal from the embryo anterior and follows head-specific regulatory cascades, which finally lead to formation of head muscles including the tongue [29,78].

In the dermomyotome, which is epithelialized myotome, myogenic progenitors progressively migrate with development [79]. The hypaxial cell migration result in the formation of abdominal muscles as well as the limb muscles, whereas the rest of the cells stay at the epaxial position and differentiate into back muscles [31,80]. Like dorsal-ventralization that occurs in the myotome, this epaxial-hypaxial differentiation is also controlled by antagonization of signals (i.e. *Wnt* and *BMP4*) originated from the corresponding positions [81]. However, unlike the head muscles which adopt a different developmental route from the rest of the skeletal muscles, all trunk muscles undergo similar developmental phases characterized by two waves of muscle fiber formation [82]. The primary muscle fiber formation occurs at E12 in the mouse embryos, during which anlagen of all future muscles is laid down [83]. Based on the primary muscle fibers, the secondary muscle fibers start to form at E15. Secondary muscles are characterized by a much larger scale of formation and they are more specified [82]. Completion of skeletal muscle development is not achieved until the animal is born. As the muscle mass continues to enlarge, satellite cells start to emerge toward the end of embryonic development. Although the origin of which is commonly thought

to be the somite, result of our research suggests that they could be linked to the dermomyotome, a later derivative of the somite.

Specifically, in this study, we analyzed mouse embryos from day E12, E15 and E18 to represent the three important skeletal muscle development stages: primary muscle fiber formation, secondary muscle fiber formation and satellite cell formation. Figure 10 clearly represented these events. In dorsal trunk muscle, *Mustn1* promoter was active in all of the nascent muscle fibers at E12. At E15, as these fibers progressively matured, *Mustn1* promoter became quiescent in most of them, with the exception of few secondary muscle fibers which were formed along the mature primary muscle fibers. As the embryo further developed, all muscle fibers lost GFP expression at E18. Instead, satellite cells were positive and showed prominent activity of the promoter which also remained active in skeletal muscle of adult mouse (Fig. 17A). Tongue muscle, on the other hand, did not show this temporal pattern, probably because its development bifurcated from the rest of the trunk muscle as early as the somitic stage and the signaling events were different [29]. Cardiac muscle, although shares the mesodermal origin with skeletal muscle [84], is independent of the somite and was completely negative of GFP at all time points, suggesting that *Mustn1* promoter activity is restricted to descendents of the somite.

Skeletal muscle and cardiac muscle share some common characteristics

including striation and sarcomeres. However, one of the most distinctive differences between these two types of muscle is that the former is multi-nucleated because of fusion of single-nucleated myoblasts whereas the latter does not [85]. Although we are unable to provide a direct proof that links *Mustn1* function to the muscle formation in this study, given the fact that *Mustn1* promoter was active in the skeletal muscle but not in the cardiac muscle, it is reasonable to attribute *Mustn1* as one of the factors that causes their differences, or, as a result of their differences. However, considering that *Mustn1* is expressed very early during embryogenesis, preceding the maturation of both muscle types, it is more likely the cause rather than the result. Thus, this discovery suggests a possible new direction for our future research, which is sought after and discussed in *Part V*.

On the other hand, the detection of high level of *Mustn1* expression in the cartilage (essentially actively proliferating chondrocytes) aligned well with our previous discovery of *Mustn1* expression in cartilage [5]. *Mustn1* is also likely to play a role in certain cells in bone which are indicated by white arrows in Fig. 14F. These GFP positive cells are only found in E18 embryos where membranous ossification is active [86], and they localize within ossified bone, thus likely to be osteoblasts. These GFP positive sclerotome derivative tissues (bone and cartilage), when combined with the previous discovery that *Mustn1* promoter is very active in the descendants of dermomyotome, strongly suggests that *Mustn1* is

expressed early on during embryogeneis that enters this process with the formation of myotome, but not at the earlier somite stage, as the cranial somitic region failed to show robust *Mustn1* promoter activity. Figure 19 shows a diagram with the putative *Mustn1* entry point in the somite lineage tree (around E9-11).

We have previously proposed that *Mustn1* serves as a musculoskeletal marker based on its spatial expression pattern [5]. This statement is strengthened with the study of *Mustn1*^{PRO}-GFP transgenic mouse embryos. Moreover, our knowledge has expanded in that a few other tissues that are independent of the musculoskeletal system also express *Mustn1* at fairly high level. Endothelial cells residing on the inner layer of blood vessels constitute one of our new discoveries, as *Mustn1* promoter activity was observed throughout the monitored time points as well as in adult blood vessels (Fig. 16 and Fig. 17B). However, this is not surprising since satellite cells and endothelial cells have been shown to co-express myogenic and endothelial markers, putting satellite cells closer to the endothelial cells rather than the seemingly more relevant myogenic progenitors [46,87]. Previous research has also showed that endothelial cells are originated from the dermomyotome [88], so it is possible that satellite cells follow the same developmental route (marked by dotted line in Fig. 19).

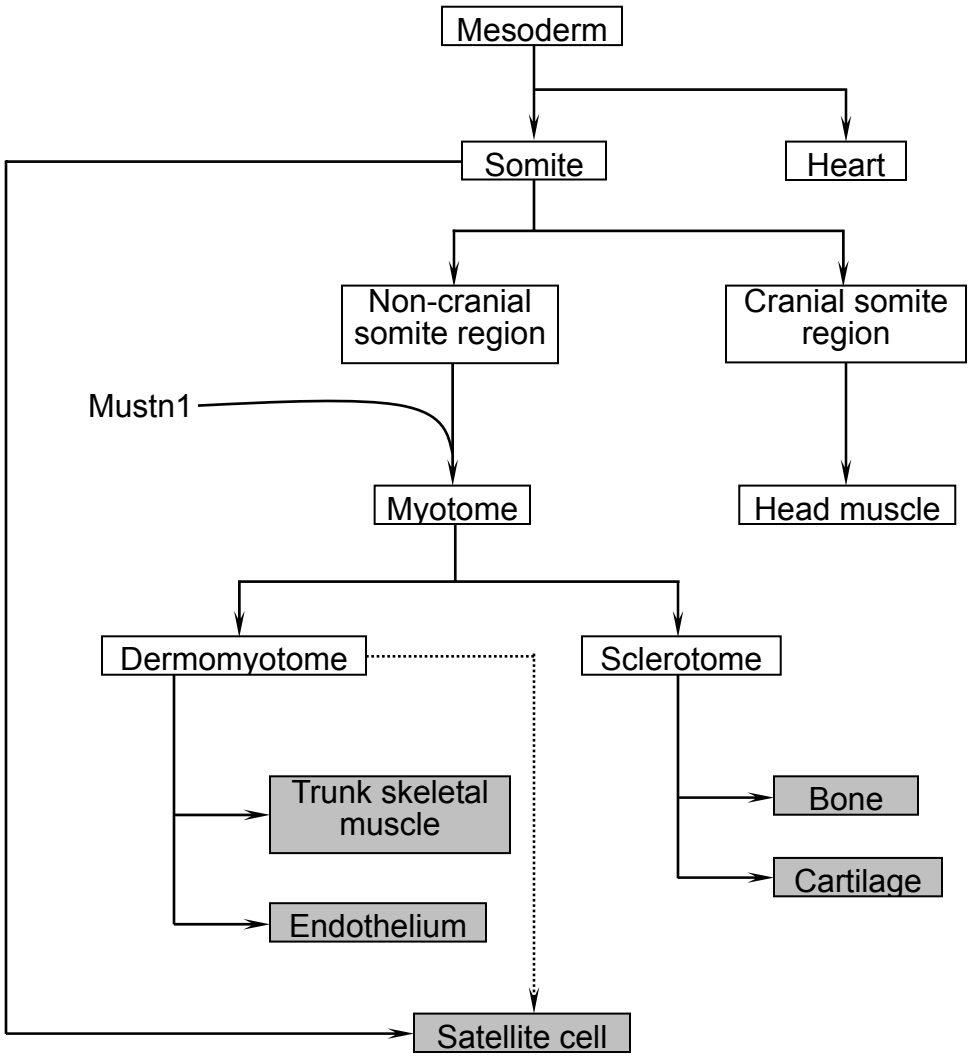


Figure 19. Lineage tree of somite and speculated Mustn1 entry point. *Shaded boxes are tissues showing Mustn1 promoter activity during embryogenesis. All these tissues are derivatives from the myotome, which could also be the first embryonic structure that expresses Mustn1. Dotted line indicates dermomyotome as a possible downstream origin of satellite cells in this tree, instead of the commonly accepted somite.*

Additionally, the significance of this study also includes localizing *Mustn1* activity to the satellite cells. Previous research disclosed skeletal muscle as one of the *Mustn1*-expressing tissues with no further spatial clarification [5]. Examination of the transgenic mouse tail muscle sections revealed GFP

expression in the satellite cells for the first time. Satellite cells are specialized cells that are thought as sources for postnatal muscle growth as well as muscle repair [28,44]. Although they also participate in muscle formation, there is no evidence that links these cells to the myogenic progenitors, except that they both share common somitic origin [89]. To date, the origin of satellite cells still remains elusive with the majority of muscle researchers supporting a somitic root [89]. A few others suggest that satellite cells result from relocated cells from the blood vessel and bone marrow [46,90]. The former is basically a false-proof claim as the somite is a very primitive structure formed during early embryogenesis. Since satellite cells only form at the end of embryogenesis, the spatiotemporal gap between them and the somite makes people to believe the existence of intermediate developmental steps. The latter opinion gives a more specific theory for the source of the satellite cells, which is also supported by our results, because 1) currently there is no evidence showing satellite cells are formed *in situ*; 2) endothelial cells derive from dermomyotome and they emerge earlier than satellite cells during embryogenesis; 3) satellite cells and endothelial cells co-express *Mustn1* as well as a variety of myogenic markers; and 4) physiologically, endothelial cells can circulate with the blood so they are readily available for migrating into skeletal muscles. Thus, it is logical to infer satellite cells as the migrating products of endothelial cells in the skeletal muscle, provided no other

viable theory is available. Nevertheless, further examination is required to verify the exact origin of these cells.

In summary, we have detected *Mustn1* promoter activity in cells that primarily constitute the musculoskeletal system (myogenic progenitors, satellite cells, chondrocytes and osteoblast-like cells) as well as in the close related (endothelial cells). The depiction of the spatiotemporal activity of the *Mustn1* promoter confirms our previous temporal and spatial expression studies and also supports that *Mustn1* is a novel marker for the musculoskeletal system. Moreover, all of these cells while in their early pluripotent stages indicate strong *Mustn1* promoter activity, whereas, as they differentiate *Mustn1* promoter activity also diminishes, further defining *Mustn1* as an indicator of early stem or progenitor cells in the musculoskeletal system. Again, this is consistent with our previous findings that showed *Mustn1 m*RNA expression in periosteoal osteoprogenitor cells and mesenchymal cell condensations [5]. Thus, isolation of these early plutipotent cells by FACS and investigations into their differentiation potential could help to verify their identities and explore *Mustn1* function in them. On the other hand, our research has suggested a possible correlation between *Mustn1* and myofusion, opening a new avenue into its function. Since our goal is to understand *Mustn1* during myogenesis, we chose to study *Mustn1* function by RNA*i* and microarray analysis, as described in the following part.

Part V: Functional Perturbation

5.1 Specific Aim 3

Functional perturbation of *Mustn1 in vitro*.

5.2 Materials and Methods

5.2.1 Materials

Mouse *anti*-myogenin antibody (F5D) and mouse *anti*-MHC antibody (MF20) were obtained from Developmental Studies Hybridoma Bank (DSHB). Goat *anti*-mouse IgG conjugated with Cy2 was obtained from Chemicon.

5.2.2 Cell culture

C2C12 cells and 293 cells were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS). All cells were incubated at 37°C and 5% CO_2. To induce myogenic differentiation of C2C12 cells, the medium was replaced with DMEM supplemented with 2% horse serum once the cells were confluent. To denote the differentiation time line, the time of switching from the growth medium (GM) to the differentiation medium (DM)

is marked as 0h (~90% confluent), and the differentiation process was monitored for up to 96 hours.

5.2.3 RNAi

RNA interference (RNA*i*) was performed to silence *Mustn1* in C2C12 myoblasts. This experiment involved the retroviral delivery of *sh*RNA (short hairpin RNA) into the host cells. Efficacies of four *sh*RNA in silencing *Mustn1* *m*RNA were evaluated relative to a control RNA*i* (sequence selected from GFP*)*, and the one inducing the most suppression was selected for further analyses. All experimental procedures were carried out following instructions from the Retroviral GeneSuppressor System provided by IMGENEX. The detailed procedure is described in the following.

First, in order to prepare the RNA*i* constructs, four different oligonucleotide sequences (19 or 20 bp in length) specific to the mouse *Mustn1* coding region were selected according to the guidelines stated in the kit's manual (available at http://www.imgenex.com). As a control, a sequence (18 bp in length) from the GFP *c*DNA was also selected according to the same guidelines. Next, in order to allow the selected sequences to form functional *sh*RNA in C2C12 cells, a spacer DNA sequence (7-8 bp) was inserted between each original *c*DNA fragment and its complementary sequence. Finally, adaptor sequences were added at both

Table 4: Oligonucleotides for creating *sh*RNA constructs

Construct Name	*Oligonucleotide Sequence*[1,2,3]	*shRNA length*
pSR-C	5′- TCGA**CATACGGAAAACTTACCC**gactcctg**GGGTAAGTTTTCCGTATG**ttttt -3′ 3′- **GTATGCCTTTTGAATGGG**ctgaggac**CCCATTCAAAAGGCATAC**aaaaaGATC -5′	*18 bp*
pSR-I	5′- TCGA**GCCAAGAACCAGGACATCAA**gactactg**TTGATGTCCTGGTTCTTGGC**ttttt -3′ 3′- **CGGTTCTTGGTCCTGTAGTT**ctgatgac**AACTACAGGACCAAGAACCG**aaaaaGATC -5′	*20 bp*
pSR-II	5′- TCGAg**AAGGAAGAAGACCTGAAGG**gactcct**CCTTCAGGTCTTCTTCCTT**ttttt -3′ 3′- c**TTCCTTCTTCTGGACTTCC**ctgagga**GGAAGTCCAGAAGAAGGAA**aaaaaGATC -5′	*19 bp*
pSR-III	5′- TCGAg**CCTGTGAAGGAAGAAGACC**gactcct**GGTCTTCTTCCTTCACAGG**ttttt -3′ 3′- c**GGACACTTCCTTCTTCTGG**ctgagga**CCAGAAGAAGGAAGTGTCC**aaaaaGATC -5′	*19 bp*
pSR-IV	5′- TCGAg**TATTCAGCCGCAACCGCAC**gactcct**GTGCGGTTGCGGCTGAATA**ttttt – 3′ 3′- c**ATAAGTCGGCGTTGGCGTG**ctgagga**CACGCCAACGCCGACTTAT**aaaaaGATC -5′	*19 bp*

[1] *Gene-specific sequences selected for the shRNA constructs are capitalized and shown in bold.*
[2] *Spacer sequences used to allow the formation of shRNA and additional extension sequences are in lower case.*
[3] *5' and 3' overhang sequences used for cloning into pSR vector are capitalized and underscored.*

ends of each construct following the kit's illustration to allow sticky-end cloning into the *p*SuppressorRetro (*p*SR for short) retroviral vector. Table 4 shows the design of all oligonucleotides. All sense and anti-sense oligonucleotides were synthesized by Invitrogen. The paired oligonucleotides were then annealed by incubating at 95°C for 2 minutes and then gradually cooled down to room temperature on a PCR thermo block. Cloning was performed using standard procedures and verified by sequencing at the Stony Brook University Sequencing Facility. Plasmids containing these *sh*RNA constructs are named *p*SR-I, *p*SR-II, *p*SR-III, *p*SR-IV and *p*SR-C (Table 4).

To produce the viruses, all the above plasmids were used individually with the pCL-Eco plasmid (kit provided) to co-transfect 293 cells using the FuGENE 6 Transfection Kit (Roche) following manufacturer's instruction. Viruses were collected from the culture media as instructed and purified by 0.45 μm filters.

Next, viruses carrying the *Mustn1 sh*RNA were applied to proliferating C2C12 cells (~30% confluent) individually in the experimental cell lines to induce RNA*i*. Viruses carrying the control GFP-specific *sh*RNA were also used to infect C2C12 cells to generate the RNA*i* control cell line. Cells were cultured in the presence of 400 μg/ml G418 (Invitrogen), a selection reagent in order to screen for stably infected cells. Single colonies were isolated from the experimental cell lines and maintained separately to identify which one shows: 1) the most dramatic

Mustn1 suppression; and 2) no interferon response, relative to the control cell line, which had all G418(+) cells pooled together. The parental uninfected C2C12 cells, the control cells and the RNA*i* cells showing the greatest knock-down of *Mustn1* were used for all the following analyses.

5.2.4 qRT-PCR

RNA*i* and microarray results were validated with *q*RT-PCR. Total RNA samples were prepared with RNeasy Mini Kit (QIAGEN) and treated with *DNase I* (QIAGEN) to remove residual genomic DNA. RNA quality was determined by gel electrophoresis and concentration measured with NanoDrop ND-1000 (NanoDrop). *q*RT-PCR was carried out with QuantiTect SYBR Green RT-PCR Kit (QIAGEN) on LightCycler system (Roche) following a standard protocol. Genes of interest and their primers were listed in Table 5. All data were normalized to *18S r*RNA. Each experiment was performed in triplicate in order to determine standard deviation.

Table 5: Primers used for *q*RT-PCR

Target Gene	*Accession Number*	*Primer Sequence*	*Amplicon Size (bp)*	*Tm (°C)*
18S	*AY248756*	*Forward: CGCGGTTCTATTTTGTTGGT* *Reverse: AGTCGGCATCGTTTATGGTC*	*219*	*60*
Mustn1	*NM_181390*	*Forward: AAGAAGAAGCGGCCCCCT* *Reverse: CTTTGGGCTTCTCAAAGAC*	*190*	*60*
OAS1	*AF466822*	*Forward: ACCGTCTTGGAACTGGTCAC* *Reverse: CTCCAGTCCTTTGGGTTCAA*	*182*	*60*
MyoD	*NM_010866*	*Forward: GCCTGAGCAAAGTGAATGAG* *Reverse: GGTCCAGGTGCGTAGAAGG*	*184*	*60*
Myf5	*NM_008656*	*Forward: TGAGGGAACAGGTGGAGAAC* *Reverse: AGCTGGACACGGAGCTTTTA*	*198*	*60*
Myogenin	*NM_031189*	*Forward: GGAAGTCTGTGTCGGTGGAC* *Reverse: CGCTGCGCAGGATCTCCAC*	*150*	*60*
Myh4	*NM_010855*	*Forward: CAAGTCATCGGTGTTTGTGG* *Reverse: GGCCATGTCCTCAATCTTGT*	*175*	*60*
Desmin	*NM_010043*	*Forward: GCGGCTAAGAACATCTCTGA* *Reverse: TCCATCATCTCCTGCTTGG*	*116*	*60*
Capn1	*NM_007600*	*Forward: GGTGAAGTGGAGTGGAAAGG* *Reverse: TGCCCTCGTAAAATGTGGTA*	*226*	*60*
Cast	*NM_009817*	*Forward: GAAAGGCAGGAGAAGTGTGG* *Reverse: GGAGGAGTTTGGGATGTGTC*	*236*	*60*
Cav3	*NM_007617*	*Forward: ACGGTGTATGGAAGGTGAGC* *Reverse: TGAGTAGATGTGGCTGATGC*	*203*	*60*
MCadh	*NM_007662*	*Forward: CCCAACTAAGGGGCTCTCTC* *Reverse: ATTCTCCCACCACTCCTGACT*	*150*	*60*

5.2.5 Myotube quantification

Pictures of the parental, control and RNA*i* cell lines were taken at 96h following induction of myogenic differentiation. Five phase contrast views from the center and four corners (located at approximately the middle of the radius) of a 10-cm culture dish were chosen under the microscope for each cell line. Numbers of myotubes and the corresponding standard deviations were calculated. A *t*-test was used to determine the statistical significance of the data.

5.2.6 Cell proliferation assay

Proliferation rate of the parental, control and RNA*i* cell lines were measured by the CellTiter 96 AQueous One Solution Cell Proliferation Assay (MTS) Kit (Promega) as well as the CyQUANT Cell Proliferation Assay Kit (Invitrogen). The same protocol was used with both assays. Briefly, the cells were seeded in triplicates in 24-well plates at a density of 8,000 cells / well. For detection with the MTS protocol, fresh cells were treated with MTS reagent for 30 minutes everyday before diluting the respective conditioned media 1:2 in dH_2O and measuring the absorption at 490 nm on SmartSpec 3000 (Bio-Rad). For detection with the CyQUANT protocol, cells were lysed with the kit-provided cell lysis buffer, followed by applying the CyQUANT GR dye and incubation for 5 minutes in the dark at room temperature. Fluorescence was measured on a CytoFluor Multi-Well Plate Reader Series 4000 (PerSeptive Biosystems) with excitation at 485 nm and emission at 530 nm. For each protocol, measurements were taken every 24 hours starting from 2 hours after cell seeding and ended when at least one of the cell lines was confluent.

5.2.7 Conditioned culture

First, conditioned media were collected at 48h and 96h from the parental, control and RNA*i* cell lines, respectively. Then, media collected from the 48h and

96h parental and control C2C12 cultures were individually applied to the RNA*i* cells at 0h. Likewise, media collected from 48h and 96h RNA*i* cultures were applied to the parental and control C2C12 cells. All cultures were maintained for up to 6 days and the progress of myotube formation was observed daily.

5.2.8 Immunocytochemistry

The parental, control and RNA*i* cell lines were stained with antibodies against myogenin and Myosin Heavy Chain (MHC). To accomplish this, all cell lines were seeded on BD BioCoat™ Collagen I Coated Coverslips (BD Biosciences) at the same density. Coverslips with cells growing on top were harvested at 2h, 48h and 96h after the beginning of myogenic differentiation. Cells were fixed with 4% paraformaldehyde for 15 minutes, washed with 1X PBS; then permeabilized with 0.1% Triton X-100 and washed again with 1X PBS. After blocking with 4% horse serum for 1 hour, primary antibodies (1:1000) were applied and incubated overnight at 4°C, then switched to the Cy2-conjugated secondary antibody (Chemicon, 1:200) for 1 hour. Finally, all coverslips were rinsed with 1X PBS, mounted with permanent mounting media (VectaMount, Vector Laboratories), and observed under a fluorescent microscope (Zeiss Axiovert 200).

5.2.9 Microarray

RNA samples from the control (C) and RNA*i* (E, experimental) cell lines were prepared and treated with *DNase I* to remove residual genomic DNA. Three time points (2, 48, 96 hours) were chosen to represent myogenic differentiation in C2C12 cells. Each RNA sample was labeled according to the cell lines and the time points (i.e. C2, C48, C96, E2, E48 and E96). In addition, an external control RNA sample was prepared from RNA*i* cells that were 48 hours before differentiation (E-48). Microarray was performed at the Stony Book University DNA Microarray Facility. Briefly, all RNA samples were first amplified by Illumina TotalPrep RNA Amplification Kit (Ambion) and labeled with Cy3. The same amount of targets was used to hybridize with custom printed mouse whole genome chips (MEEBO) containing ~30,000 features, which was supplied by the same facility. Hybridized chips were scanned at 532 nm (green channel) and the images were grided with GenePix Pro 5.1 software. Raw data was compensated by normalizing to the total signal intensity of each microarray chip.

5.2.10 Microarray data filtering and clustering analysis

Data calculated by GenePix was subjected to three filters before analyzing. First, all control genes were excluded because their signal was saturated. Second, any gene must have demonstrated a minimum signal to noise ratio of 3.0

in at least two out of the six time points (C2, C48, C96, E2, E48 and E96) in order to be included in further filtering. Third, all of the raw values were normalized to corresponding identities in the E-48 set. Any gene that failed to show an expression level either three fold higher or lower (0.33-3.0 fold) than C2 in at least one of the other five samples was removed from the gene list. After these filtering steps, only the genes that showed dramatic changes passed for further data mining. Data normalization / filtering were performed in GeneSpring GX 7.3. The same software was also used to cluster the remaining genes by K-mean clustering. Subsequently, interesting clusters were identified by their biological relevance to the process of myogenic differentiation. Clusters that best represented elevated expression in the control but low in the RNA*i* C2C12 cells were chosen for in-depth investigation.

5.2.11 Statistical analysis

Statistical analysis was carried by SigmaStat 3.1 software to determine the statistical significance of the myotube numbers. *t*-test was adopted for this analysis and a $p<0.05$ was considered as statistically significant.

5.3 Results

5.3.1 Confirmation of Mustn1 RNAi

Efficacy of RNA*i* in the selected C2C12 colonies was measured by *q*RT-PCR and presented as the percentage of knock-down of *Mustn1* at the *m*RNA level as compared to the control RNA*i*. Our previous study had revealed that *Mustn1* is up-regulated at Day 4 in myogenic differentiation [91]. Thus, Day 4 RNA from both the control and the RNA*i* C2C12 cells were isolated for evaluating RNA*i*, in the belief that *Mustn1* suppression is best represented at this time point. Additionally, the parental unaltered C2C12 cell line was used as an external reference. *18S r*RNA was used as the housekeeping gene. The results showed that 59% of *Mustn1* expression was silenced at Day 4 in the RNA*i* cells comparing to the control cells (Fig. 20). Since virus infection was known to induce nonspecific cellular interferon responses [92-94], a critical gene (*2′,5′-oligoadenylate synthetase 1*, *OAS1*) involved was used to indicate the level of this

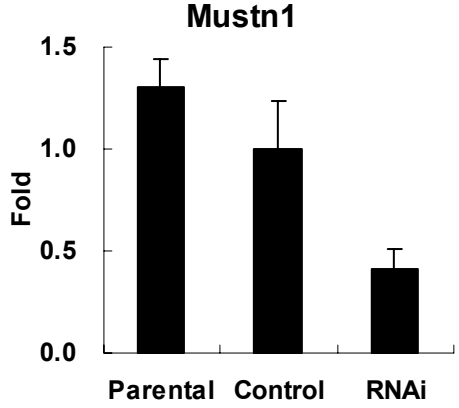


Figure 20. Validation of Mustn1 RNAi. *The expression level of Mustn1 and OAS1 in Day 4 differentiated C2C12 cells were detected by qRT-PCR. All values were normalized to that of the control cell line.Mustn1 is expressed at normal level in the parental and control cell lines, but only 41% of the normal level (indicated by the control) is shown in the RNAi cell line. OAS1 is not expressed at all in all three cell lines (data not shown), showing no virus-induced interferon response was triggered by RNAi.*

effect [92,95]. However, expression of *OAS1* was not detectable in all three cell lines using the standard *q*RT-PCR protocol, nor when the amount of template RNA was increased up to five fold (data not shown), indicating the retrovirus used to deliver the RNA*i* did not induce any interferon response.

5.3.2 Mustn1 RNAi has no effect on C2C12 proliferation

Although *Mustn1* is only up-regulated during myogenic differentiation, a basal level of expression is maintained during proliferation of C2C12 cells [91]. Thus, it was important to find out whether *Mustn1* plays a role during cell proliferation and loss-of-function of *Mustn1* provided an ideal tool. To examine *Mustn1* suppression on the proliferation of C2C12 cells, the growth rate of the parental, control and the selected *Mustn1* RNA*i* cell lines was compared. Proliferation assays were carried out by measuring both cell metabolic activity (MTS method, Fig. 21A) and total amount of nucleic acid (CyQUANT method, Fig. 21B). However, neither assay showed that proliferation rate was affected by silencing of *Mustn1* in C2C12 cells. Additionally, the retroviral infected cells showed no morphological differences when compared to the other two cell lines during proliferation (data not shown). However, since no apparent differences on cell proliferation rates were observed, no further step was carried out to characterize these cell lines at the molecular level.

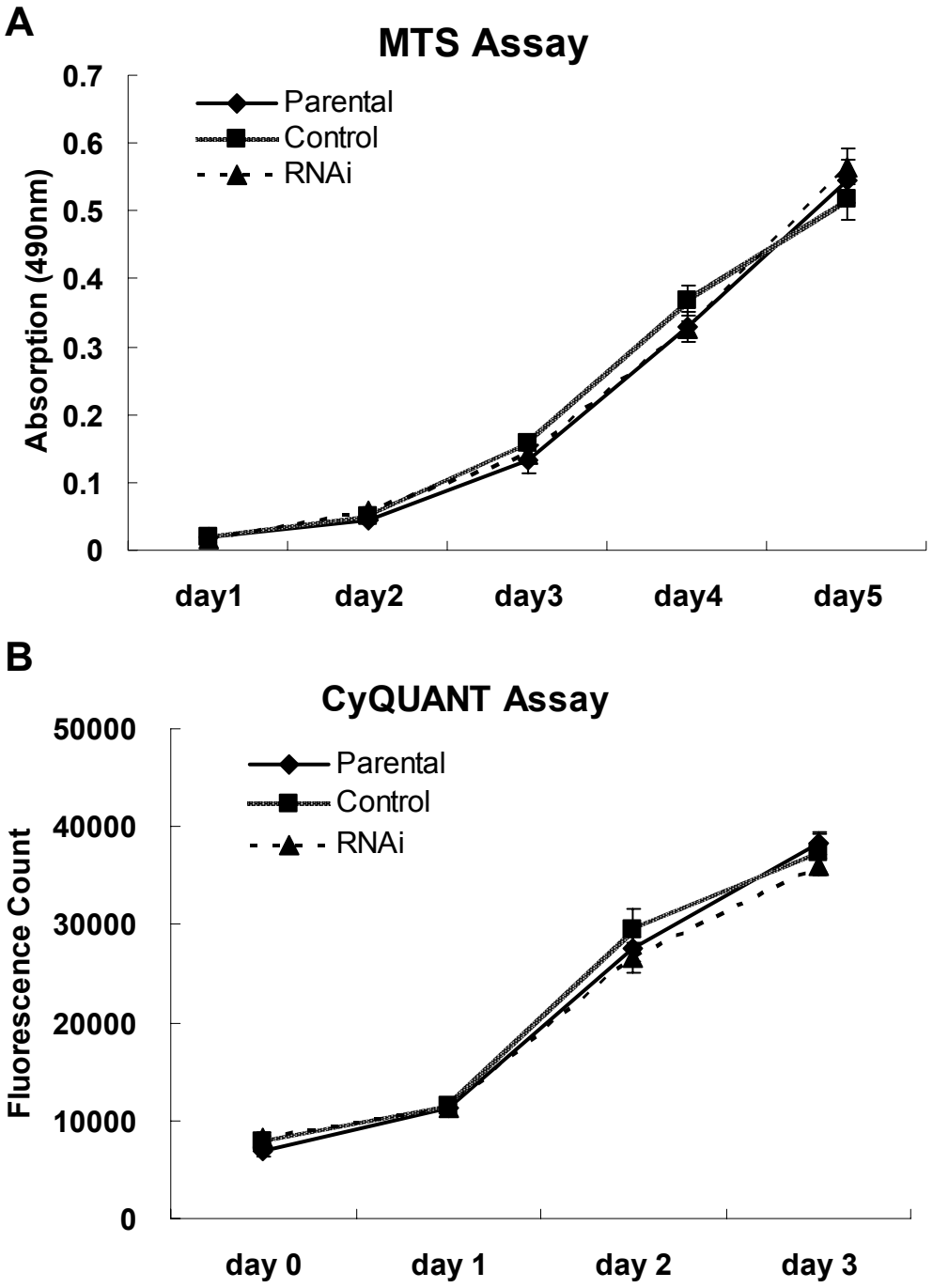


Figure 21. Proliferation rate comparisons of the C2C12 parental, control and RNAi cells. *MTS assay (A) measures proliferation based on cellular metabolic activity, whereas CyQUANT assay (B) measures proliferation based on total quantity of nucleic acid within the cells. However, according to the five-day MTS assay and the three-day CyQUANT assay, knocking down Mustn1 in C2C12 cells is not capable of inducing detectable proliferative alteration.*

5.3.3 Mustn1 is required for C2C12 myogenic differentiation

Similar to our proliferation experiments, C2C12 cell morphology was observed throughout the entire course of myogenic differentiation (Fig. 22). Imaging started when all cell lines were confluent and the time of switching to 2% horse serum culture (0h). As shown in Fig. 22A, all three cell lines showed no morphological difference at the end of proliferation (0h). However, as sporadic elongated myotubes started to form at 48h in the parental and control RNA*i* cells, no myotubes were observed in the *Mustn1* RNA*i* cells except a more crowded cell population (compared to 0h) due to further proliferation. At 96h, myogenic differentiation was completed with the appearance of distinct fully elongated myotubes in the parental and control cell lines. But in the *Mustn1* RNA*i* cells, only very few smaller myotubes were seen at this time point, resembling the early differentiation stages (48h) of the parental and control cells. Quantitative analysis (Fig. 22B) performed at the end of myogenic differentiation (96h) showed that the number of myotubes in the parental (36.8 per view or 100%) and control (32.4 per view or 88%) cells were significantly more than that in the *Mustn1* RNA*i* (9.8 per view or 26.6%) cells.

To test whether the *Mustn1* RNA*i* cells can be rescued in their ability to form myotubes, experiments were performed with conditioned media. For example, treating C2C12 *Mustn1* RNA*i* cells with conditioned media collected from the 48h

and 96h parental and control cells did not rescue the phenotype (i.e. increase myotube formation) caused by *Mustn1* RNA*i*. This experiment was repeated in the reverse fashion, whereby parental and control cells were cultured with conditioned media collected from the *Mustn1* RNA*i* cells but no effects on their differentiation capacity were observed (data not shown).

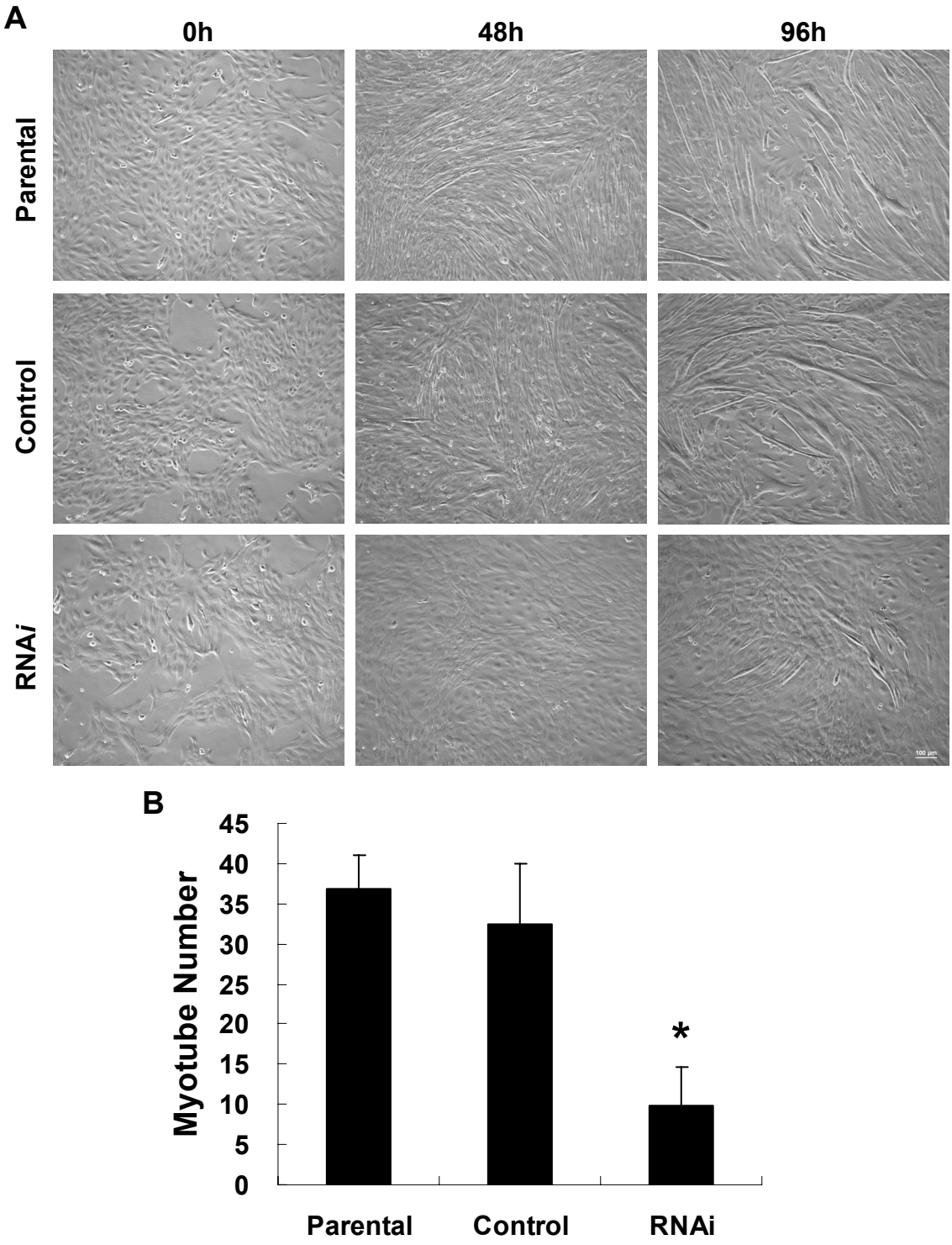


Figure 22. Morphological characterization of C2C12 during myogenic differentiation. *(A) The parental, control and RNAi cells were induced with 2% horse serum at 0h. Cell morphology in culture was recorded at 0h, 48h and 96h. Distinctive myotube formation can be seen in the parental and control culture beginning from 48h. In the contrary, very scarce myotubes were observable until 96h in the RNAi culture. Scale bar 100µm. (B) Average number of myotubes per view (the same size as shown in A) in the parental, control and RNAi cultures at 96h. The RNAi cell line has significantly less myotubes than the other two cell lines ($p<0.05$).*

5.3.4 Myofusion is abolished by Mustn1 RNAi in C2C12 cells

To further investigate the decrease in myotube formation observed with the *Mustn1* RNA*i* cells, parental, control and *Mustn1* RNA*i* C2C12 cells collected at 2h, 48h and 96h after induction with DM and were stained with *anti*-myogenin and *anti*-MHC antibodies. The results showed that silencing of *Mustn1* caused a delay in the expression of myogenin (Fig. 23), and complete abolishment of myofusion (Fig. 24).

More specifically, in Fig. 23, both the parental and control cells differentiated normally with green fluorescence observed in the nuclei beginning from 48h, reflecting myogenin expression. Further differentiation resulted in larger number of myogenin-positive cells in these two cell lines at 96h. In comparison, no myogenin expression was detected in the *Mustn1* RNA*i* cells at 48h, but by 96h, some cells were detected as myogenin-positive, but at much lower numbers. In addition, it is notable that multi-nucleated cells started to appear in the parental and control cells at 48h and became more prominent at 96h, indicating active myofusion during the myogenic differentiation process. However, these multi-nucleated cells were not observed throughout the entire 96h period in the *Mustn1* RNA*i* cells, which do not form any multinucleated myotubes.

Fig. 24 essentially confirmed our observation in Fig. 23 by staining for cytoplasmic MHC. The parental and control C2C12 cells showed normal

myogenic differentiation by expressing *MHC* beginning at 48h. Moreover, longer myofibers were observed at 96h in both cell lines as a result of active myofusion. In contrast, the *Mustn1* RNA*i* cells not only showed much less *MHC*-expressing cells, but also failed to initiate myofusion at 48h (cells elongated but failed to fuse), which was not ameliorated at 96h except a few more single nucleus MHC-positive cells, suggesting complete abolishment of myofusion instead of a possible delay. This was also experimentally verified, as prolonged cultures of up to 8 days failed to show any mature, multinucleated myotubes in (data not shown).

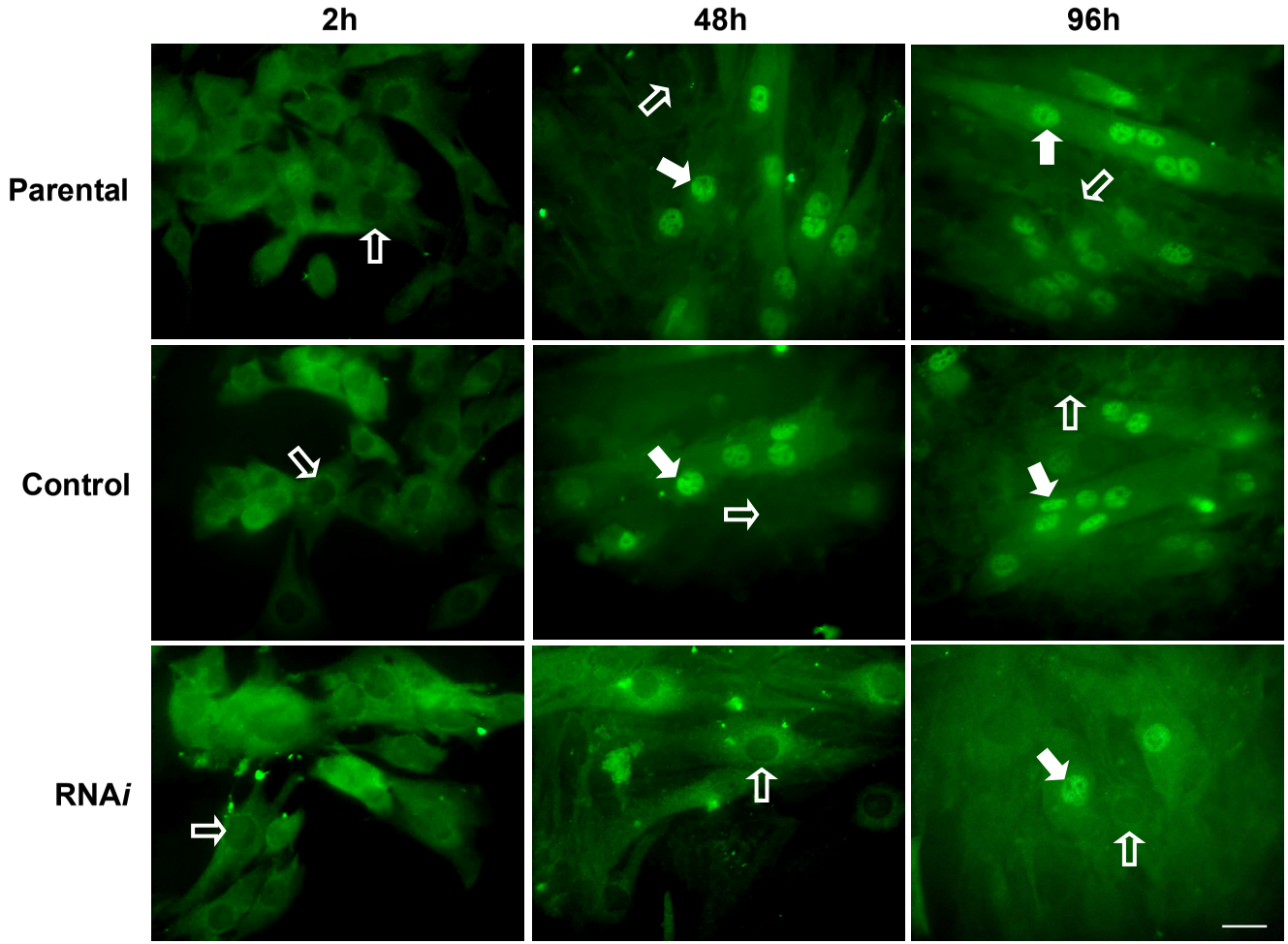


Figure 23. Temporal myogenin expression in vitro. *The parental, control and RNAi C2C12 cells were seeded on cover glasses and harvested at 2h, 48h and 96h. Cells were stained with anti-myogenin antibody. Solid arrows show myogenin expression within the nuclei, and the hollowed arrows show the absence of myogenin expression in the nuclei. Myogenin expression was delayed in the RNAi cells, and occurrence of the myogenin-positive nuclei is much lower in the RNAi cells at 96h. Scale bar 10μm.*

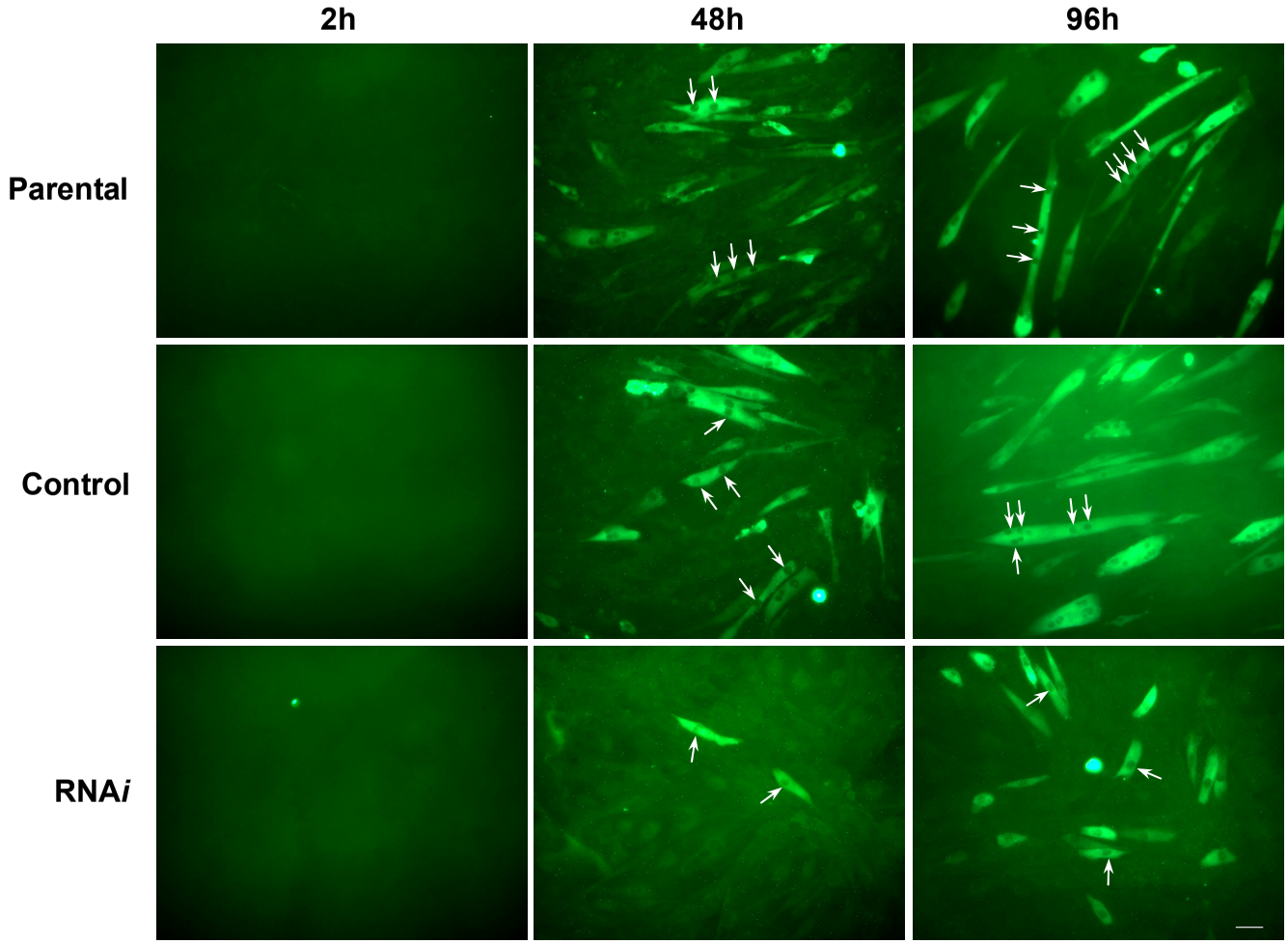


Figure 24. Temporal MHC expression in vitro. *The parental, control and RNAi C2C12 cells were seeded on cover glasses and harvested at 2h, 48h and 96h. Cells were stained with anti-MHC antibody. MHC-expressing cells are shown as bright green. Positions of nuclei are indicated with arrows. Scale bar 50μm.*

5.3.5 Expression profiling of the impact of silencing Mustn1

Since silencing *Mustn1* in C2C12 cells resulted in fundamental alterations in the expression of myogenic markers (i.e. myogenin and MHC) and more importantly, in myotube formation (as described in the previous sections), it was naturally then to search for more global gene expression changes between the *Mustn1* RNA*i* and control cells. Hence, we adopted a global expression profiling approach by performing a microarray experiment. Six samples corresponding to three time points (2h, 48h and 96h) during the myogenic differentiation of both the control (C) and *Mustn1* RNA*i* (E) C2C12 cells were used for the profiling. After data acquisition and processing, 6,039 out of ~30,000 genes which showed more than three fold change in expression were clustered into eight groups by K-mean clustering. Different cluster numbers (6-16) was tried and we determined that eight was the best to show the variety while keeping the cluster number low. Individual cluster profiles are shown in Fig. 25.

As shown, each cluster showed a distinctive expression profile. Cluster 1 (n=975) included the genes that did not show dramatic up- or down-regulation in the control C2C12 cells. However, the expression level of these genes were much lower in the RNA*i* cells, with slight up-regulation at E48, then declined to a lower level by 96h. Cluster 2 (n=829) showed similar trend to the control cells, whereas their expression was maintained at lower level without evident fluctuation. Gene

expression in Cluster 3 (n=459) showed slight up-regulation during myogenic differentiation of the normal cells, however, a number of these genes failed to show detectable level of expression in the RNA*i* cells (considered not expressed). Similarly, Cluster 4 (n=350) also included a myriad of genes that were not expressed in the RNA*i* cells, except that genes in the RNA*i* group showed steady expression (small down-regulation at E48) rather than the generally up-regulation at E48 shown in Cluster 3. Genes in Cluster 5 (n=528) showed trend of up-regulation in both the control and RNA*i* cell, whereas the ones in Cluster 6 (n=866) were opposite. Lastly, in Cluster 7 (n=1003) and Cluster 8 (n=761), genes were up-regulated during normal C2C12 differentiation. However, these genes were down-regulated upon myogenic stimuli at E48, then “caught up” at E96 (Cluster 7) or became completely quiescent (Cluster 8).

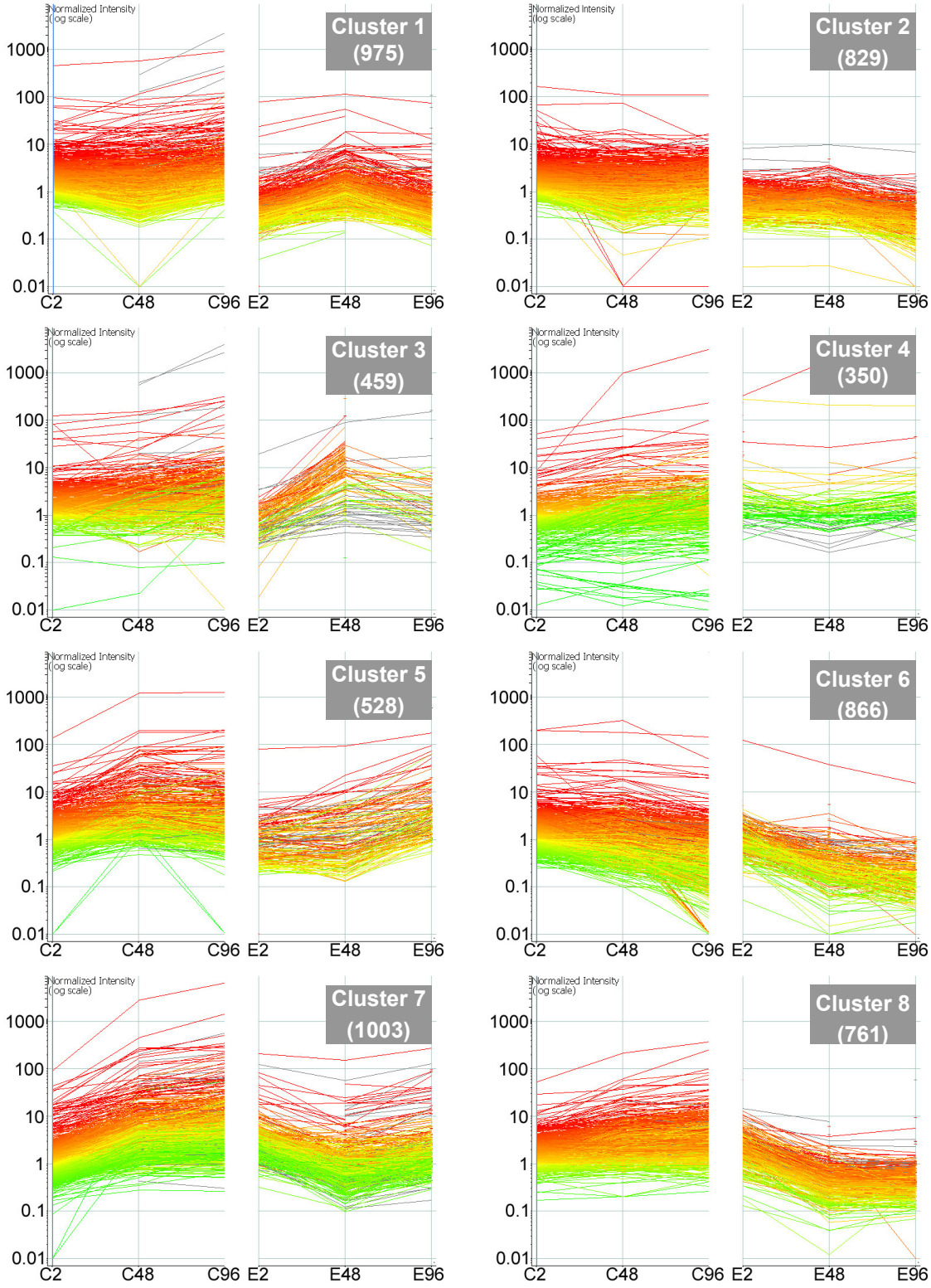


Figure 25. Microarray data clustering by K-mean. *6039 genes that passed all filtering are clustered into eight sets and an unclassified set (contain 269 genes, not shown). The x-axis represents the control (C2, C48, C96) and RNAi (E2, E48, E96) samples. The y-axis represents the fold increase/decrease of each sample after normalized to the E-48 sample (log scale). Number of genes in each cluster is labeled under the corresponding cluster name.*

Since the study in the previous part suggested a possible connection between *Mustn1* and myofusion, we also compiled a list of myofusion-related genes (37 in total) and compared their relative expression profiles (Table 6). As expected, nearly all genes were up-regulated during the normal myogenesis (C2, C48 and C96), except that *Capn2* and *Ccna2* were progressively down-regulated as the myotubes matured. In contrast, the majority of these genes were down-regulated in the *Mustn1* RNA*i* cells (E2, E48 and E96). Their expression levels either remained low across all three time points (e.g. *Myod1* and *Des*), or showed some up-regulation at late time point at E96 (e.g. *Myog*, *Myh4* and *Csrp3*), and even then their levels did not reach those obtained from control cells. In addition, two genes (*Capn6* and *Itgb1*) showed higher than normal expression levels at one or more time points. Since the number of genes that displayed up-regulated expression in the *Mustn1* RNA*i* cells were small, these genes could be very relevant and thus are of interest. Lastly, we also analyzed the clusters that the myofusion genes belonged to. Results showed that the majority of these genes (11 of 37, marked in Table 6) were allocated to Cluster 7, including *Mustn1*. Thus, we decide to focus our further analysis on this cluster, as it is more likely to include genes that pertain to myofusion.

Table 6: Expression of myofusion-related genes

Gene Name	Time Points					
	C2	C48	C96	E2	E48	E96
■ *Atp2a2*	2.42	10.93	27.32	1.69	1.15	1.99
■ *Cacna2d1*	2.12	11.58	22.94	2.44	1.20	2.71
Cacnb1	4.26	23.75	74.89	1.00	1.00	1.00
Cacng1	12.99	29.48	62.81	5.81	3.69	5.57
■ *Calm1*	2.09	6.81	6.86	2.17	1.37	1.59
■ *Capn1*	1.03	3.56	3.24	1.76	0.48	0.88
Capn2	1.14	0.65	0.84	0.41	0.35	0.15
Capn6	0.12	0.48	0.88	0.57	1.60	1.11
Capns1	1.17	2.54	3.54	2.44	0.79	1.40
Cast	3.80	5.51	7.77	5.63	0.70	0.59
Cav3	1.00	6.04	13.32	0.92	1.22	1.03
Ccna2	2.03	1.09	0.19	1.08	0.17	0.12
Ccnd1	0.86	0.93	0.62	1.30	0.15	0.17
Cdh15	5.66	9.19	5.59	1.05	1.59	1.38
Cdkn1a	1.21	1.21	1.21	0.71	0.95	0.99
■ *Chrna1*	6.74	37.68	73.00	9.29	2.21	4.86
Chrnd	2.08	4.09	7.18	1.00	1.00	1.00
■ *Chrng*	25.30	88.16	119.72	0.72	18.15	16.44
Cryab	1.00	0.83	1.00	0.51	1.00	0.56
Csrp3	1.00	554.80	4074.20	1.00	1.00	159.40
Ctsb	1.12	2.26	4.63	0.89	1.05	1.24
Des	4.07	6.00	6.31	2.05	1.83	1.60
Egr3	7.03	7.96	6.63	2.30	2.59	3.33
Itgb1	0.52	1.64	1.02	0.22	0.18	1.96
Mef2a	1.20	2.62	2.47	1.23	0.58	0.65
■ *Mef2c*	1.39	4.98	19.27	1.11	1.00	1.35
■ *Mustn1*	1.00	7.30	18.80	2.51	1.44	3.27
■ *Myf5*	2.55	5.38	5.99	3.99	1.00	1.00
■ *Myh3*	17.43	278.06	277.85	41.76	5.97	72.09
Myh4	320.00	1631.00	1985.00	237.00	351.00	600.00
Myod1	3.40	3.40	4.41	0.80	0.85	1.06
Myog	7.92	55.46	91.10	2.13	3.52	17.06
Pgf	1.00	1.00	1.71	2.45	1.00	2.31
Rb1	1.27	2.09	2.75	0.80	0.84	1.07
■ *Ryr1*	2.70	4.24	34.85	1.00	0.88	1.00
Vcam1	1.88	1.88	1.88	1.88	1.22	1.62
Vcl	1.25	1.25	1.25	1.25	1.22	0.64

A gene ontology analysis (using GeneSpring GX 7.3) was performed on Cluster 7 to identify cell adhesion genes that were differentially regulated, as they are direct regulators of myoblast membrane recognition and fusion. Table 7 shows a list of these genes. All values are after normalization to E-48 and blank means the corresponding raw data failed to pass the aforementioned filtering steps. Similar to myofusion-related genes, all genes listed in Table 7 were down-regulated due to the *Mustn1* RNA*i*. Consequently, the ability of the *Mustn1* RNA*i* cells to recognize and relay fusion signals was impaired, resulting in elongated, single-nucleated cells. In this sub-category of Cluster 7, it is notable that several genes related to calcium ion channels were dramatically down-regulated, including *Capn1*, *Cdh13* and *Clstn1*, indicating the participation of calcium ion mediated signaling pathway.

Table 7: Expression of cell adhesion genes in Cluster 7

Gene Name	Time Points					
	C2	C48	C96	E2	E48	E96
Akt3	1.39	3.54	5.17	2.80		
Bad	0.58	1.59	3.44	3.31	0.69	1.66
Bcl2	3.94	15.44	22.30	6.59		2.19
Birc2	1.02	3.31	5.26	2.17	0.94	1.53
Capn1	1.03	3.56	3.24	1.76	0.48	0.88
Cbll1	1.53	2.74	5.42	3.22		
Cdc42		14.52	12.52	6.19	0.27	0.99
Cdh13	44.40	189.80	352.90			95.00
Clk3	0.47	1.58	1.53	1.89	0.29	1.03
Clstn1			5.92	4.05	1.33	2.85
Col18a1	0.98	4.37	2.48	4.05	0.44	0.70
Col4a5	0.70	2.42	2.38	1.54	1.13	1.95
Ddr1	0.88	1.21	1.26			
Dock1	0.46	1.72	1.62	2.17	0.28	1.01
Dpt	0.69	1.11	1.24			
F11r	4.86	8.01	7.76		6.03	
Glg1	1.55	5.02	5.07	1.88	0.89	1.88
Hapln4	0.94	1.66	2.96		1.18	
Itga10		41.87	63.41		10.79	21.47
Itgb5	0.85	5.00	10.36	7.12	2.29	5.99
Jam3	0.82	3.09	2.90	2.43	0.86	1.48
Lama5	0.58	0.94	1.06	0.90		
Lamb1-1	1.11	4.71	5.98	4.82	0.63	0.89
Limk1	0.68	2.67	1.65	2.07	0.28	0.54
Lpp	0.38	2.58	1.54	2.52	0.41	1.09
Macf2	0.94	2.34	3.12	1.11		0.97
Map2k1	0.98	4.45	4.44	3.29	0.37	0.74
Mybph	11.26	116.51	153.88		18.91	34.98
Ncam1	1.57	7.81	18.19	1.76	1.27	2.69
Nedd9	1.84	7.22	7.62	3.51	1.66	3.71
Neo1		1.67	1.82	2.18	0.25	0.79
Pcdh7	1.05	3.68	4.50	2.09	2.08	2.31
Ptk2	1.07	4.42	3.82	3.22	0.75	0.71
Sdc1	1.43	3.58	3.51			
Shc2	1.81	3.27	3.30			
Sympk	0.78	2.47	3.30	2.83	0.68	
Vegfa	0.65	1.63	3.07	0.83	0.18	0.40

5.3.6 Microarray data validation

To confirm the reliability of the microarray experiment, *q*RT-PCR was carried out on ten genes selected from the myofusion-related gene list, including *Mustn1*, *Myod1* (MyoD), *Myf5* (myogenic factor 5), *Myog* (*Myogenin*), *Des* (*Desmin*), *Myh4* (*myosin heave chain, peptide 4*), *Capn1* (*calcium-activated neutral protease 1*), *Cast* (*calpastatin*), *Cav3* (*caveolin 3*), and *Cdh15* (*cadherin 15*, *M-cadherin*). Data generated by both the *q*RT-PCR and microarray were plotted together in Fig. 26. As shown, seven genes demonstrated similar expression patterns by both assays, whereas the other three (*Myf5*, *Myh4* and *Cast*) showed higher variations between the two assays, probably due to the qualitative nature of the microarray technology. Regardless, with the exception of *Myf5* and *Cast* that show no differences in expression between the control and *Mustn1* RNA*i* samples, the remaining eight all show suppressed expression in *Mustn1* RNA*i* cells, thereby confirming our cell based observations.

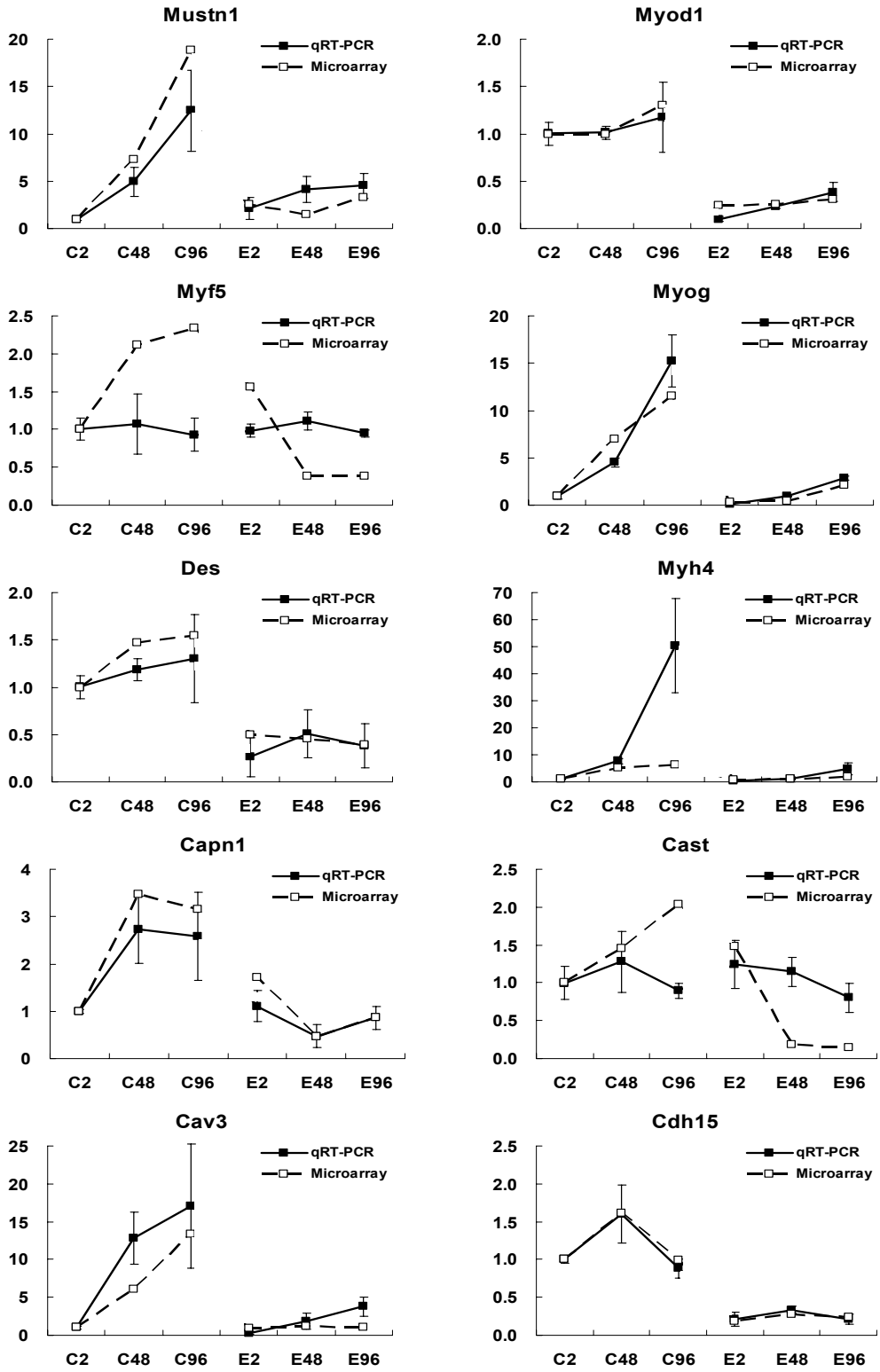


Figure 26. Microarray data confirmation by qRT-PCR. *Fold changes of expression relative to C2 found in qRT-PCR and microarray were plotted for ten genes. Same set of RNA was used for both assays.*

5.4 Discussion and Conclusion

Previous research in our laboratory has identified *Mustn1* as a musculoskeletal-specific gene which a very restricted expression pattern [5]. Our most recent work has revealed new evidence that further supports the connection of this gene to myogenesis (*Part III* and *IV*). To further expand our understanding, in this study, we demonstrated that *Mustn1* is required for myoblast fusion by showing failure of myofusion in *Mustn1* RNA*i*-treated cells.

RNA*i* was introduced as a tool for studying gene function for a number of years and its mechanism covered by a number of recent literatures [96-98]. As a loss-of-function technique, RNA*i* is similar to gene knockout but more widely adopted nowadays because it is faster, less complex and cost-effective. However, pitfalls and caveats such as viral interferon responses and off-target effects are also extensively described [99-101]. In this study, we have successfully silenced *Mustn1* in C2C12 myoblasts. The following measurements were taken in order to ensure the success of this experiment: 1) four *Mustn1*-specific sequences (confirmed by BLAST and presented no homology to other genomic sequence) were selected according to generally accepted guidelines to avoid off-target effect [100]; 2) a control RNA*i* experiment involving a GFP coding sequence was selected according to the same guidelines and performed to assess the delivery vector's impact on gene expression; and 3) since a retrovirus was adopted to

deliver the RNA*i* constructs into the host cells, we titrated the dosage to avoid saturation of the endogenous RNA*i* machinery as well as possible virus-induced interferon responses (monitored by *OAS1* expression to assess the level of such non-specific responses) [92]. As the results show, we were able to knock-down the expression of *Mustn1* by ~60% without triggering an interferon response (*OAS1* level below the detectibility of *q*RT-PCR). A rescue experiment was also attempted subsequently by utilizing the redundancy of the genetic code [102], but we were not able to rescue the *Mustn1* RNA*i* phenotype, possibly due to: 1) the chemical transfection method had limited capacity in inducing over-expression; and 2) endogenous level of *Mustn1* expression in normal C2C12 cells was too high to be remedied by over-expression in RNA*i* cells (data not shown). On the other hand, we also tried to just over-express *Mustn1* in C2C12 cells but that approach was unsuccessful, probably due to the same reasons (data not shown).

The following characterization of the *Mustn1* RNA*i* cells showed that silencing *Mustn1* had no effect on proliferation, which was conceivable because endogenously in C2C12 cells *Mustn1* levels are very low during their proliferative stage and thus it is less likely to carry out any function during this time. However, its impact on the myogenic differentiation was dramatic. The *Mustn1* RNA*i* cells displayed severe deficiency in expressing important myogenic markers, including *Myog* and *Mhy4* (Fig. 23, 24 and 26), the deterministic transcription factor for

myogenesis and vital component of MHC, respectively [103,104]. Phenotypically, since the *Mustn1* RNA*i* cells expressed these markers at much lower level, activation of other myogenic genes was delayed or completely abolished, resulting in impaired myotube formation (Fig. 22). Although the expression of *Myog* and *Myh4* was elevated to a certain level at 96h (due to leaky *Mustn1* expression in the RNA*i* cells), it was still not enough to rescue their phenotype, and thus the cells displayed a complete failure of myotube formation, rather than simply a delay. This was also experimentally verified, as prolonged cultures of up to 8 days failed to show any mature, multinucleated myotubes.

A careful examination of the *Mustn1* RNA*i* cell phenotype revealed that *Mustn1* silencing caused a failure of myoblast fusion, a cellular process that has been studied and reviewed extensively [105-109]. An orchestrated set of steps are involved in this event, of which the beginning is symbolized by protein and calcium dependant cell-cell recognition and adhesion [110,111]. The fact that proteins involved in these processes (*Capn1*, *Cdh13* and *Clstn1*) displayed suppressed expression in the *Mustn1* RNA*i* cells, lead us to believe that it is viable to attribute the failure of myoblast fusion to this very first step.

To further analyze these cell based observations, we also sought to characterize the global expression changes possibly due to *Mustn1* RNA*i* via mouse whole genome microarrays. Cells that were 2h, 24h and 96h in

differentiation culture were harvested to represent the early, middle and late stages of myogenenic differentiation [112]. Data processing with moderate stringency has left 6,039 genes from the total of ~30,000, and then they were clustered into eight distinctive sets. An inspection of these clusters revealed a global change in gene down-regulation due to *Mustn1* RNA*i*. Although the lack of statistical significance prevented us from performing more in-depth analysis, the trend reflected by the microarray analysis suggested that *Mustn1* is an early serum response factor that participates in the regulation of a myriad of myogenic genes (Table 6).

By focusing on the genes that showed the most dramatic changes in expression, which are likely to better indicate true up- or down-regulation, we discovered that *Csrp3* (*cysteine and glycine-rich protein 3*) and *Mef2c* (*myocyte enhancer factor 2c*), both showed similar patterns to *Mustn1*. As both genes are regulators of myogenic differentiation, further exploration of their relationship to *Mustn1* could be worthwhile. Additionally, *Csrp3* is known as a protein containing a LIM-domain which is comprised of two tandemly arranged zinc fingers, thus capable of DNA binding [113]. So, *Mustn1* could conceivably be an interacting protein to *Csrp3* since they are both nuclear proteins. Indirect evidence for this hypothesis comes from a study on bovine skeletal muscle profiling that provides a computer-predicted connection between *Mustn1* and *Crsp3* [114].

Lastly, our analysis of the microarray data, especially on cell adhesion genes in Cluster 7 (Table 7) suggests that the calcium ion signaling pathway maybe a possible component of the *Mustn1*-associated regulation cascade. Since this pathway is documented by other studies of myogenesis [110], the role of *Mustn1* in the pathway will be clarified in future research.

In summary, we were able to perform a loss-of-function study on *Mustn1* in a myogenesis model *in vitro*, followed by global gene expression profiling via microarrays. This study partly revealed the function of *Mustn1* by indicating its involvement during myoblast fusion as well as its importance in myogenic differentiation. However, to obtain more comprehensive knowledge on the function of *Mustn1*, the following study must be considered: 1) identify *Mustn1*-interacting proteins (work of the next part); 2) generating *Mustn1* conditional (in skeletal muscle) knock-out mice (based on the *Mustn1*^{PRO}-GFP transgenic mice described in the previous part) to study this gene systematically *in vivo*; and 3) compare the *Mustn1* knock-out mouse with the *Mustn1*^{PRO}-GFP transgenic mouse to differentiate the true loss-of-function effects from the false ones caused by the leaky *Mustn1* expression.

Part VI: Protein-protein Interaction

6.1 Specific Aim 4

Identify *Mustn1* interacting proteins.

6.2 Materials and Methods

6.2.1 Materials

Yeast strain AH109 and Y187, YPD (yeast extract/peptone/dextrose) medium for normal yeast culture, DDO (-Leu/-Trp double drop-out), TDO (-Leu/-Trp/-His triple drop-out) and QDO (-Leu/-Trp/-His/-Ade quardruple drop-out) media for nutrient-deficient culture were obtained from BD Biosciences. L-Leucine and L-Tryptophan were purchased from Sigma-Aldrich. [^{35}S]-methionine was purchased from Amersham.

6.2.2 The yeast two-hybrid assay

In order to identify *Mustn1* interacting proteins, the MATCHMAKER GAL4 Two-Hybrid System 3 was adopted (Clontech). All experimental procedures were performed according to manufacturer's directions. First, the mouse *Mustn1*

*c*DNA sequence was cloned into the kit-provided vector (*p*GBKT7) containing the sequence encoding the DNA binding domain (BD) following standard protocols. The cloning was in-frame so that *Mustn1* was transcribed and translated correctly. Next, the new *p*GBKT7-*Mustn1* plasmid was amplified and transformed into yeast strain AH109. Then, the cells were mated with a pre-transformed Y187 strain containing the mouse day 17 embryo *c*DNA library (Clontech). Mated colonies were screened on DDO, TDO and QDO media, sequentially. Plasmids within the colonies that survived the most stringent selection were isolated (protocol described in *section 6.2.3*) and sequenced. The identities of the putative *Mustn1* interacting proteins were retrieved by using BLASTn tool against mouse genome in the NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/BLAST) database.

6.2.3 Yeast plasmid isolation

After screening the mated yeast strains (AH109 and Y187) on QDO media, the AD/Library plasmids contained within positive colonies were isolated. Briefly, yeast suspension of the individual colonies were cultured, centrifuged and lysed in 300 µl yeast lysis buffer (2% Triton X-100, 1% SDS, 100mM NaCl, 10mM Tris pH 8.0, 1mM EDTA) and 300 µl phenol-chloroform-isoamyl alcohol (25:24:1) with 300 mg acid-washed glass beads (0.45 – 0.52 mm). The mixture was vortexed

vigorously for 5 minutes to fully lyse the cells. After centrifugation, the plasmid-containing pellet was washed in 70% ethanol and dissolved in 1X TE buffer, then used directly to transform competent *E.coli* cells in order to obtain large quantities of purified AD/Library plasmids for sequencing and identity.

6.2.4 Co-immunoprecipitation (Co-IP)

The AD/Library plasmid isolated from colony 5-3E (*p*GADT7-*mo)* and QS9 (*p*GADT7-*PCNA)*, were chosen to test their interactions with *Mustn1* by *co*-IP. Both *p*GBKT7-*Mustn1* and *p*GADT7-*mo* or *p*GADT7-*PCNA* were transcribed / translated *in vitro* using the T_NT T7 Quick Coupled Transcription / Translation System (Promega) following the manufacturer's instructions. Essentially, this protocol incorporated [^{35}S]-methionine into the translated proteins which enabled the subsequent detection by *co*-IP and SDS-PAGE. *Co*-IP was performed using the MATCHMAKER *Co*-IP kit (BD Biosciences) following the manufacturer's instructions. Briefly, the *in vitro* transcribed / translated proteins was first incubated together to allow putative interactions. Then antibodies corresponding to the epitope tags carried by these proteins were added selectively (as suggested by the manufacture's instructions) to allow a following binding by protein A beads. Washed protein A beads were boiled and the content was subjected to SDS-PAGE. Finally, the protein gel was sufficiently dried and separation of the protein bands

was visualized by X-ray film radiography.

6.3 Results

6.3.1 Yeast two-hybrid assay – Round 1

In the first round of screening, eighteen individual colonies survived on the most stringent QDO medium. Sequencing of all eighteen AD/Library plasmids revealed that they all represent distinct genes (Table 8). Specifically, three (*Cpsf2*, *SRp40* and *NRIP1*) represented nuclear proteins, twelve were cytoplasmic, and the remaining three were unknown genes.

Table 8: Putative Mustn1-interacting proteins (Round 1)

Colony ID	Accession Number	Identity of the Interacting Protein	Cellular Localization
1-A5	BC061012	Mus musculus RIKEN cDNA 2700085E05 gene (2700085E05Rik)	unknown
1-E5	NM_010729	Mus musculus lysyl oxidase-like (Loxl)	cytoplasmic
1-G2	U14172	Mus musculus p162 protein	cytoplasmic
1-G5	NM_030717	Mus musculus lactamase, beta (Lactb)	cytoplasmic
2-G2	NM_023409	Mus musculus Niemann Pick type C2 (Npc2)	cytoplasmic
3-G5	NM_017670	Homo sapiens ubiquitin-specific protease otubain 1 (FLJ20113)	cytoplasmic
3-C7	BC029087	Mus musculus ZW10 interactor	cytoplasmic
4-C7	AK012320	Mus musculus RIKEN cDNA 2700033I16 gene (2700033I16Rik)	unknown
4-D7	NM_017033	Rattus norvegicus Phosphoglucomutase 1 (Pgm1)	cytoplasmic
5-G3	BC038145	Mus musculus, clone IMAGE:1401090	unknown
5-G5	XM_216766	Rattus norvegicus cleavage and polyadenylation specific factor 2 (Cpsf2)	nuclear
8-F4	NM_009159	Mus musculus, splicing factor, arginine/serine-rich 5 (SRp40, HRS)	nuclear
9-A3	AW228963	Mus musculus, Similar to lysozyme	cytoplasmic
10-E1	BC008269	Mus musculus, protein tyrosine phosphatase, non-receptor type 2 (Ptpn2)	cytoplasmic
11-F1	BC030881	Mus musculus aarF domain containing kinase 5	cytoplasmic
12-G6	NM_003489	Homo sapiens nuclear receptor interacting protein 1 (NRIP1)	nuclear
13-B2	BC132169	Mus musculus von Hippel-Lindau binding protein 1 (Vbp1)	cytoplasmic
14-C4	NM_009932	Mus musculus procollagen, type IV, alpha 2 (Col4a2)	cytoplasmic

6.3.2 Yeast two-hybrid assay – Round 2

In the second round of screening, twenty-three individual colonies survived on the most stringent QDO medium. Sequencing of the AD/Library plasmids revealed that they represent twenty-one individual genes (Table 9). Specifically, six represented nuclear proteins, eleven were cytoplasmic, and the remianing four were unknown genes. Noticeably, two *c*DNAs encoding a zinc finger protein 278 (1-3F and 5-4D) and a mitochondrial solute carrier protein (5-6E and 7-2E) appeared twice. In addition, two *c*DNAs, one encoding *Npc2* (2-G2 in Round 1, 5-4E in round 2) and *SRp40* (8-F4 in Round 1, 7-4D in Round 2) were also present in the second screening round.

Table 9: Putative Mustn1-interacting proteins (Round 2)

Colony ID	Accession Number	Identity of the Interacting Protein	Cellular Localization
1-2C	BC048175	*Mus musculus suppressor of K+ transport defect 3*	cytoplasmic
1-3F	BC043035	Mus musculus zinc finger protein 278	nuclear
4-3B	NM_028876	Mus musculus cDNA from various tissues	cytoplasmic
5-2D	AY058327	Drosophila melanogaster GH09688 full length cDNA	cytoplasmic
5-2E	AL627083	Mouse DNA sequence from clone RP23-378E2 on chromosome 4 (genomic)	unknown
5-2G	AK010993	Mus musculus cDNA from various tissues (hemoglobin)	unknown
5-3D	AK020715	Mus musculus adult male hypothalamus cDNA	nuclear
5-3E	AK031964	Mus musculus adult male medulla oblongata cDNA	nuclear
5-3F	U79242	Human clone 23560 mRNA sequence	cytoplasmic
5-3G	AK010980	Mus musculus cDNA from various tissues (ribosomal)	unknown
5-4D	BC043035	Mus musculus zinc finger protein 278	nuclear
5-4E	BC003471	Mus musculus Niemann Pick type C2 (Npc2)	cytoplasmic
5-4F	AL593857	Mouse DNA sequence from clone RP23-439H2 on chromosome 2-genomic	unknown
5-5E	NM_007563	Mus musculus 2,3-bisphosphoglycerate mutase	cytoplasmic
5-6D	NM_023210	Mus musculus acidic (leucine-rich) nuclear phosphoprotein 32 family, member E (Anp32e)	cytoplasmic
5-6E	BC040399	Mus musculus mitochondrial solute carrier protein	cytoplasmic
7-1B	NM_026302	Mus musculus dynactin 4 (Dctn4)	cytoplasmic
7-1D	NM_008298	Mus musculus DnaJ (Hsp40) homolog, subfamily A, member 1 (Dnaja1)	cytoplasmic
7-1E	AF060087	Mus musculus proteasome (prosome, macropain) subunit, alpha type 6 (Psma6)	cytoplasmic
7-1F	BC048176	Mus musculus, Ras-GTPase-activating protein (GAP<120>) SH3-domain binding protein 2	nuclear
7-2E	BC040399	Mus musculus mitochondrial solute carrier protein	cytoplasmic
7-3C	NM_133423	Rattus norvegicus splicing factor YT521-B (YT521)	nuclear
7-4D	NM_009159	Mus musculus, splicing factor, arginine/serine-rich 5 (SRp40, HRS)	nuclear

6.3.3 Yeast two-hybrid assay – Round 3

In the third round of screening, eighteen individual colonies survived on the most stringent QDO medium. Sequencing of the AD/Library plasmids revealed that they all represent distinct genes. Specifically, five were nuclear protein-coding, ten were cytoplasmic protein-coding, and the rest three were not identifiable. A list of their identities is available in Table 10. None of these identities repeated those identified in Round 1 and 2.

Table 10: Putative Mustn1-interacting proteins (Round 3)

Colony ID	Accession Number	Identity of the Interacting Protein	Cellular Localization
QC1	BC030856	*Mus musculus U5 snRNP-specific protein, 200 kDa (DEXH RNA helicase family)*	nuclear
QC3	NM_010181	Mus musculus fibrillin 2 (Fbn2)	cytoplasmic
QC5	NM_133379	Homo sapiens titin (TTN), transcript variant novex-3	cytoplasmic
QS1	BC010315	Mus musculus thymine DNA glycosylase	nuclear
QS2	M32599	Mouse glyceraldehyde-3-phosphate dehydrogenase	cytoplasmic
QS3	U76635	Rattus norvegicus deoxyribonuclease I gene	nuclear
QS4	AK048560	Mus musculus 16 days embryo head cDNA	unknown
QS5	NM_009342	Mus musculus t-complex testis expressed 1 (Tctex1)	cytoplasmic
QS7	U23184	Mus musculus carboxypeptidase E (Cpe)	cytoplasmic
QS8	XM_342653	Rattus norvegicus similar to gamma-filamin (LOC362332)	cytoplasmic
QS9	BC060570	Rattus norvegicus proliferating cell nuclear antigen	nuclear
QS10	BC024394	Mus musculus RNA polymerase 1-3	nuclear
QS11	BC052054	Mus musculus ilvB (bacterial acetolactate synthase)-like	cytoplasmic
QS12	NM_027870	Mus musculus RIKEN cDNA 1200004E24 gene (1200004E24Rik)	unknown
QS14	NM_010180	Mus musculus fibulin 1 (Fbln1)	cytoplasmic
QS15	AF034582	Rattus norvegicus vesicle associated protein (VAP1)	cytoplasmic
QS16	AL627205	Mouse DNA sequence from clone RP23-41E14 on chromosome 11	unknown
QS17	NM_011857	Mus musculus odd Oz/ten-m homolog 3 (Drosophila) (Odz3)	cytoplasmic

6.3.4 Confirmation of putative interactions by co-IP

Following identification of putative interacting proteins, the co-immunoprecipitation assay was adopted as a necessary step to test the authenticity of those interactions. In the MATCHMAKER GAL4 Two-Hybrid System 3, Mustn1 was fused to a *c*-Myc epitope tag, and the mouse *c*DNA library target proteins were fused to a hemagglutinin (HA) epitope tag. Thus, any interaction between Mustn1 and its target protein could be detected by the presence of antibody against either *c*-Myc or HA. If no interaction existed, only one protein could be detected. In both *co*-IP experiments, bands in Lane 1 and Lane 2 should indicate successful *in vitro* transcription / translation and antibody-epitope tag recognition. Lane 3 and Lane 4 should tell the existence of interaction: two bands in both lanes indicated positive interaction; whereas only one band in each indicated negative result. Lane 5 and Lane 6 were negative controls in which swapped antibodies were applied to the epitope tags so no band should be present (Fig. 27). The *co*-IP experiments showed that, no protein-protein interaction was formed between Mustn1 and medulla oblongata (mo) protein or PCNA *in vitro*, although putative interactions were indicated by the yeast two-hybrid assay. It is notable that in both Fig. 27A and B, a faint band which was the same size as the target protein (mo or PCNA) showed up in Lane 3, possibly due to weak interactions. However, same bands in Lane 6 of both

images were also found, suggesting this was caused by residual affinity between anti-*c*-Myc antibody and the HA epitope tag, not real interaction between Mustn1 and mo or PCNA.

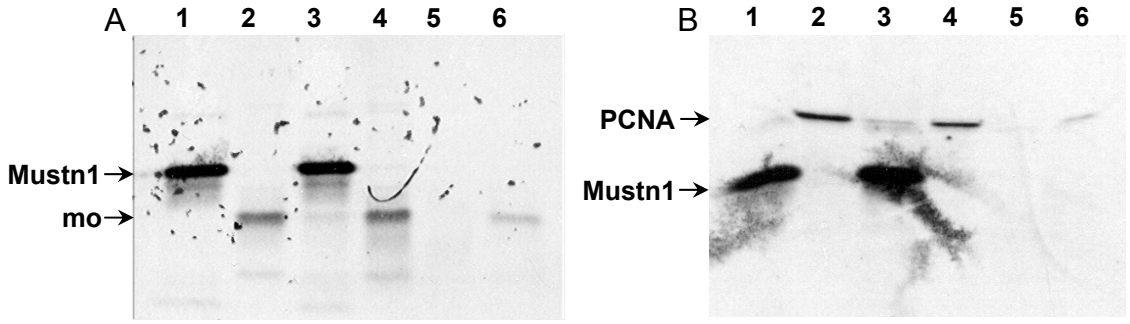


Figure 27. Co-IP validation of putative interactions. *For both figures, Lane 1 shows Mustn1 precipitated by anti-Myc antibody, Lane 2 shows the putative interacting protein precipitated by anti-HA antibody, Lane 3 and 4 shows the co-precipitation by either anti-Myc antibody (Lane 3) or anti-HA antibody (Lane 4), Lane 5 and 6 shows the result when antibody is swapped. Ideally, double bands in both Lane 3 and 4 means there is interaction between Mustn1 and the tested candidate. In these experiments, only one band was shown in Lane 3 and 4 in both figures, suggesting that the interactions detected by yeast two-hybrid were false positives. (A) Co-IP confirmation of interaction between Mustn1 and medulla oblongata protein (5-3E); (B) Co-IP confirmation of interaction between Mustn1 and PCNA (QS9).*

6.4 Discussion and Conclusion

The yeast two-hybrid system is a tool that is widely utilized to identify protein-protein interactions [115,116]. It is based on the fact that the transcriptional activation apparatus of eukaryotic genes is composed of two domains, a DNA binding domain (BD) and an activation domain (AD). Both domains have to be in proximity to each other in order for transcription to start. The yeast two-hybrid system deliberately separates these two domains so that only recognition of interacting proteins (each fused to one of the domains) can

form the BD-AD moiety, thus the activation of a reporter gene, which in the yeast, usually the ones that enable the host to withstand certain nutrient deprivation. This specific interaction can then be screened based upon the survival of the yeast colony on the corresponding nutrient-deficient medium. Finally, the interaction target is identified by isolating and sequencing the AD/Library plasmid that conferred the ability to survive. To confirm the interaction, a *co*-immunoprecipitation (*co*-IP) assay is generally required.

In this study, *Mustn1* was used as the bait to screen against the AD/Library containing the mouse 17 day embryo *c*DNA library, a time point which involves high expression level of *Mustn1* according to our previous unpublished work, hence better chances of detecting Mustn1-interacting proteins. The yeast two-hybrid assay was repeated three times in order to 1) even out the possibility of suboptimal experimental design / technique or incidental false positives / negatives due to round-to-round variations, and 2) enable comparisons between each rounds. A total of fifty-nine genes were identified in these independent assays, with four of them presented in two separate screens (~6.8% of 59), which led to fifty-five distinct interacting candidates. However, none came up more than twice in all three rounds. Table 11 shows a detailed summary of the results. As shown, out of the fifty-five unique identities, thirteen (24%) represented as nuclear protein, whereas thirty-two (58%) were cytoplamic and ten (18%) were unknown.

Table 11: Summary of the yeast two-hybrid screening

	nuclear	*cytoplasmic*	*unknown*	*TOTAL*
Round 1	*3*	*12*	*3*	*18*
Round 2	*7*	*12*	*4*	*23*
Round 3	*5*	*10*	*3*	*18*
TOTAL	*15* [1]	*34* [2]	*10*	*59* [3]

[1,2] Two genes were identified twice each in this category
[3] Actually represent 55 distinct identities

Due to the fact that the entire mouse *c*DNA library was expressed in the two-hybrid system simultaneously, molecules that were spatially or temporally separated might interact with each other in the yeast nuclei, resulting in false positive interactions. In the case of *Mustn1*, since it encoded a differentially expressed nuclear protein [5], it would be reasonable to restrict all further investigation to the nuclear protein-encoding genes only (Table 12). However, although the scope of investigation was narrowed down to only thirteen identities, they still represented a variety of different processes including those that were considered “housekeeping” (four genes as indicated in Table 12), which were unlikely to recruit *Mustn1*, as discussed previously (*Section 1.1*). To confirm the likely interactions, two candidates were selected for further testing using *co*-IP: 5-3E (medulla oblongata *c*DNA) was chosen from the ones with unknown functions which might present as a novel Mustn1-interating protein, and QS9 (*PCNA*) was chosen because it is known to interact with more than one hundred different proteins [117]. However, the results showed that neither was a

Mustn1-interacting protein. Given the poor reproducibility obtained with three rounds of screening, it was unworthy and laborsome to test the authenticity of all these putative interactions by the low-throughput *co*-IP assay. Therefore, no more *co*-IP confirmation was attempted.

Table 12: Summary of putative Mustn1-interacting nuclear proteins

Accession Number	***Identity of the Interacting Protein***
XM_216766	*Rattus norvegicus cleavage and polyadenylation specific factor 2 (Cpsf2)* [2]
NM_009159	*Mus musculus, splicing factor, arginine/serine-rich 5 (SRp40, HRS)* [1]
NM_003489	*Homo sapiens nuclear receptor interacting protein 1 (NRIP1)*
BC043035	*Mus musculus zinc finger protein 278* [1]
AK020715	*Mus musculus adult male hypothalamus cDNA*
AK031964	*Mus musculus adult male medulla oblongata cDNA*
BC048176	*Mus musculus, Ras-GTPase-activating protein (GAP<120>) SH3-domain binding protein 2*
NM_133423	*Rattus norvegicus splicing factor YT521-B (YT521)*
BC030856	*Mus musculus U5 snRNP-specific protein, 200 kDa (DEXH RNA helicase family)*
BC010315	*Mus musculus thymine DNA glycosylase* [2]
U76635	*Rattus norvegicus deoxyribonuclease I gene* [2]
BC060570	*Rattus norvegicus proliferating cell nuclear antigen*
BC024394	*Mus musculus RNA polymerase 1-3* [2]

[1] Found twice during the screenings.
[2] Nuclear “housekeeping” genes.

In summary, instead of an ideal scenario wherein only one or few candidates constitute the majority of the isolated *c*DNAs, identification of as many as fifty-five different genes with very low rate of repeatability (~6.8%) made us questioned the specificity of the interactions. Thus the reliability of these assays, although they did show certain degree of affinity to four of the identified candidates, was extremely low. Failure in confirming the interaction by *co*-IP further dampens

the likelihood of faithful representation of interaction by the yeast two-hybrid system. Since the experiment was repeated three times and the outcomes were not repeatable, it is likely that the yeast two-hybrid system is unsuitable for identifying interacting protein of *Mustn1*, or that special requirements have to be met.

The failure to identify interacting proteins may be due to a number of different factors such as: 1) Structurally, Mustn1 is a small protein of only 82 amino acids which may make Mustn1 unlikely to interact with many proteins. So, the design of this system may physically block the binding motifs or intrinsically prevent its correct folding [115,118], considering Mustn1 is fused to the BD sequence (for the yeast two-hybrid assay) and tagged by *c*-Myc (for the *co*-IP assay); 2) Lack of appropriate post-translational modification in the yeast could also contribute to the inability of identifying true interacting proteins of Mustn1 [115]. N-myristoylation, N-glycosylation, and phosphorylation sites have been previously been predicted from its amino acid sequence, and they could be critical in regulating the activity and conformation of the molecule, provided that no other motif is present. If this is the case, a mammalian two-hybrid system perhaps should be better method to use; 3) It is speculated that Mustn1 is likely to be a co-factor involved in a large transcription activation complex, due to its differential expression pattern and lack of DNA binding motif. Thus, Mustn1 alone may

possess certain capability in activating reporter genes in the yeast, causing random identification of a myriad of false interacting proteins with very low repeatability [119].

Based on all of these, we can conclude that the yeast two-hybrid system is unsuitable for identifying interacting proteins of Mustn1. Therefore, it is imperative to develop other approaches to unravel this puzzle, or, perturb *Mustn1's* function from other angles such as developing Mustn1-specific antibody and performing relevant assays.

Part VII: References

[1] Ferguson, C. M., Miclau, T., Hu, D., Alpern, E. and Helms, J. A. (1998) Common molecular pathways in skeletal morphogenesis and repair. *Ann. N.Y. Acad. Sci.* **857**: 33-42

[2] Ferguson, C. M., Alpern, E., Miclau, T. and Helms, J. A. (1999) Does adult fracture repair recapitulate embryonic skeletal formation? *Mech. Dev.* **87**(1-2): 57-66

[3] Vortkamp, A., Pathi, S., Peretti, G. M., Caruso, E. M., Zaleske, D. J. and Tabin C. J. (1998) Recapitulation of signals regulating embryonic bone formation during postnatal growth and in fracture repair. *Mech. Dev.* **71**(1-2): 65-76

[4] Hadjiargyrou, M., Lombardo, F., Zhao, S., Ahrens, W., Joo, J., Ahn, H., Jurman, M., White, D. W. and Rubin, C. T. (2002) Transcriptional profiling of bone regeneration. Insight into the molecular complexity of wound repair. *J. Biol. Chem.* **277**(33): 30177-30182

[5] Lombardo, F., Komatsu, D. and Hadjiargyrou, M. (2004) Molecular cloning and characterization of Mustang, a novel nuclear protein expressed during skeletal development and regeneration. *FASEB J.* **18**(1): 52-61

[6] Devlin, R. B. and Emerson, C. P. Jr. (1978) Coordinate regulation of contractile protein synthesis during myoblast differentiation. *Cell* **13**(4): 599-611

[7] Miner. J. H. and Wold, B. (1990) Herculin, a fourth member of the MyoD family of myogenic regulatory genes. *Proc. Natl. Acad. Sci. U. S. A.* **87**(3): 1089-1093

[8] Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., et al. (1991) The MyoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**(4995): 761-766

[9] Bergstrom, D. A. and Tapscott, S. J. (2001) Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family. *Mol. Cell Biol*. **21**(7): 2404-2412

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

[11] Andres, V. and Walsh, K. (1996) Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. *J. Cell Biol.* **132**(4): 657-666

[12] Katagiri, T., Yamaguchi, A., Komaki, M., Abe, E., Takahashi, N., Ikeda, T., Rosen, V., Wozney, J. M., Fujisawa-Sehara, A. and Suda, T. (1994) Bone morphogenetic protein-2 converts the differentiation pathway of C2C12 myoblasts into the osteoblast lineage. *J. Cell Biol.* **127**(6 Pt 1): 1755-66

[13] Katagiri, T., Akiyama, S., Namiki, M., Komaki, M., Yamaguchi, A., Rosen, V., Wozney, J. M., Fujisawa-Sehara, A. and Suda, T. (1997) Bone morphogenetic protein-2 inhibits terminal differentiation of myogenic cells by suppressing the transcriptional activity of MyoD and myogenin. *Exp. Cell Res.* **230**(2): 342-351.

[14] Mancini, A., El Bounkari, O., Norrenbrock, A. F., Scherr, M., Schaefer, D., Eder, M., Banham, A. H., Pulford, K., Lyne, L., Whetton, A. D. and Tamura, T. (2007) FMIP controls the adipocyte lineage commitment of C2C12 cells by downmodulation of C/EBP alpha. *Oncogene* **26**(7): 1020-1027

[15] Blau, H. M., Pavlath, G. K., Hardeman, E. C., Chiu, C. P., Silberstein, L., Webster, S. G., Miller, S. C. and Webster, C. (1985) Plasticity of the differentiated state. *Science* **230**(4727): 758-766

[16] Pownall, M. E., Gustafsson, M. K. and Emerson, C. P. (2002) Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu. Rev. Cell Dev. Biol.* **18**: 747-783

[17] Parker, M. H., Seale, P. and Rudnicki, M. A. (2003) Looking back to the embryo: defining transcriptional networks in adult myogenesis. *Nat. Rev. Genet.* **4**(7): 497-507

[18] Charge, S. B. and Rudnicki, M. A. (2004) Cellular and molecular regulation of muscle regeneration. *Physiol. Rev.* **84**(1): 209-238

[19] Olson, E. N. and Klein, W. H. (1994) bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes Dev.* **8**(1994): 1-8

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

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

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

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

[24] Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nonaka, I. and Nabeshima, Y. (1993) Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**(6437): 532-535

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

[26] Duprez, D. (2002) Signals regulating muscle formation in the limb during embryonic development. *Int. J. Dev. Biol.* **46**(7): 915-925

[27] Tajbakhsh, S. (2003) Stem cells to tissue: molecular, cellular and anatomical heterogeneity in skeletal muscle. *Curr. Opin. Genet. Dev.* **13**(4): 413-422

[28] Zammit, P. S., Partridge, T. A. and Yablonka-Reuveni, Z. (2006) The skeletal muscle satellite cell: the stem cell that came in from the cold. *J. Histochem. Cytochem.* **54**(11): 1177-1191

[29] Mootoosamy, R. C. and Dietrich, S. (2002) Distinct regulatory cascades for head and trunk myogenesis. *Development* **129**(3): 573-583

[30] Noden D. M. and Francis-West, P. (2006) The differentiation and morphogenesis of craniofacial muscles. *Dev. Dyn.* **235**(5): 1194-1218

[31] Christ, B. and Brand-Saber,i B. (2002) Limb muscle development. *Int. J. Dev. Biol.* **46**(7): 905-914

[32] Yusuf, F. and Brand-Saberi, B. (2006) The eventful somite: patterning, fate determination and cell division in the somite. *Anat. Embryol. (Berl.)* **211** (Suppl 1): 21-30

[33] Patapoutian, A., Yoon, J. K., Miner, J. H., Wang, S., Stark, K. and Wold, B. (1995) Disruption of the mouse MRF4 gene identifies multiple waves of myogenesis in the myotome. *Development* **121**(10): 3347-3358

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

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

[36] Borycki, A. G., Li, J., Jin, F., Emerson, C. P. and Epstein, J. A. (1999) Pax3 functions in cell survival and in pax7 regulation. *Development* **126**(8): 1665-1674

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

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

[39] Valdez, M. R., Richardson, J. A., Klein, W. H. and Olson, E. N. (2000) Failure of Myf5 to support myogenic differentiation without myogenin, MyoD, and MRF4. *Dev. Biol.* **219**(2): 287-298

[40] Kollias, H. D., Perry, R. L., Miyake, T., Aziz, A. and McDermott, J. C. (2006) Smad7 promotes and enhances skeletal muscle differentiation. *Mol. Cell Biol.* **26**(16): 6248-6260

[41] Wackerhage, H. and Rennie, M. J. (2006) How nutrition and exercise maintain the human musculoskeletal mass. *J. Anat.* **208**(4): 451-458

[42] Buckingham, M., Bajard, L., Daubas, P., Esner, M., Lagha, M., Relaix, F. and Rocancourt, D. (2006) Myogenic progenitor cells in the mouse embryo are marked by the expression of Pax3/7 genes that regulate their survival and myogenic potential. *Anat. Embryol. (Berl.)* **211** (Suppl 1): 51-56

[43] Muir, A. R., Kanji, A. H. and Allbrook, D. (1965) The structure of the satellite cells in skeletal muscle. *J. Anat.* **99**(Pt 3): 435-444

[44] Cossu, G. and Biressi, S. (2005) Satellite cells, myoblasts and other occasional myogenic progenitors: possible origin, phenotypic features and role in muscle regeneration. *Semin. Cell Dev. Biol.* **16**(4-5): 623-631

[45] Bittner, R. E., Schofer, C., Weipoltshammer, K., Ivanova, S., Streubel, B., Hauser, E., Freilinger, M., Hoger, H., Elbe-Burger, A. and Wachtler, F. (1999) Recruitment of bone-marrow-derived cells by skeletal and cardiac muscle in adult dystrophic mdx mice. *Anat. Embryol. (Berl.)* **199**(5): 391-396

[46] De Angelis, L., Berghella, L., Coletta, M., Lattanzi, L., Zanchi, M., Cusella-De Angelis, M. G., Ponzetto, C. and Cossu, G. (1999) Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. *J. Cell Biol.* **147**(4): 869-878

[47] Grigoriadis, A. E., Heersche, J. N. and Aubin, J. E. (1990) Continuously growing bipotential and monopotential myogenic, adipogenic, and chondrogenic subclones isolated from the multipotential RCJ 3.1 clonal cell line. *Dev. Biol.* **142**(2): 313-318

[48] Komatsu, D. E. and Hadjiargyrou, M. (2004) Activation of the transcription factor HIF-1 and its target genes, VEGF, HO-1, iNOS, during fracture repair. *Bone* **34**(4): 680-688

[49] Wagner, E. F. (2002) Functions of AP1 (Fos/Jun) in bone development. *Ann. Rheum. Dis.* **61** (Suppl 2): ii40-42

[50] Andreucci, J. J., Grant, D., Cox, D. M., Tomc, L. K., Prywes, R., Goldhamer, D. J., Rodrigues, N., Bedard, P. A. and McDermott, J. C. (2002) Composition and function of AP-1 transcription complexes during muscle cell differentiation. *J. Biol. Chem.* **277**(19): 16426-16432

[51] Palcy, S., Bolivar, I. and Goltzman, D. (2000) Role of activator protein 1 transcriptional activity in the regulation of gene expression by transforming growth factor beta1 and bone morphogenetic protein 2 in ROS 17/2.8 osteoblast-like cells. *J. Bone Miner. Res.* **15**(12): 2352-2361

[52] Karreth, F., Hoebertz, A., Scheuch, H., Eferl, R. and Wagner, E. F. (2004) The AP1 transcription factor Fra2 is required for efficient cartilage development. *Development* **131**(22): 5717-5725

[53] Eckert, D., Buhl, S., Weber, S., Jager, R. and Schorle, H. (2005) The AP-2 family of transcription factors. *Genome Biol.* **6**(13): 246

[54] Berkes, C. A. and Tapscott, S. J. (2005) MyoD and the transcriptional control of myogenesis. *Semin. Cell Dev. Biol.* **16**(4-5): 585-595

[55] Knapp, J. R., Davie, J. K., Myer, A., Meadows, E., Olson, E. N. and Klein, W. H. (2006) Loss of myogenin in postnatal life leads to normal skeletal muscle but reduced body size. *Development* **133**(4): 601-610

[56] Gersch, R., Lombardo, F., McGovern, S., and Hadjiargyrou, M. (2005) Reactivation of Hox gene expression during bone regeneration. *J. Orthop. Res.* **23**(4): 882-890

[57] Ionescu, A. M., Schwarz, E. M., Vinson, C., Puzas, J. E., Rosier, R., Reynolds, P. R. and O'Keefe, R. J. (2001) PTHrP modulates chondrocyte differentiation through AP-1 and CREB signaling. *J. Biol. Chem.* **276**(15): 11639-11647

[58] Pedraza-Alva, G., Zingg, J. M. and Jost, J. P. (1994) AP-1 binds to a putative cAMP response element of the MyoD1 promoter and negatively modulates MyoD1 expression in dividing myoblasts. *J. Biol. Chem.* **269**(9): 6978-6985

[59] Li, Q., Dashwood, W. M., Zhong, X., Al-Fageeh, M. and Dashwood, R. H. (2004) Cloning of the rat beta-catenin gene (Ctnnb1) promoter and its functional analysis compared with the Catnb and CTNNB1 promoters. *Genomics* **83**(2): 231-342

[60] Goldberg, D., Polly, P., Eisman, J. A. and Morrison, N. A. (1996) Identification of an osteocalcin gene promoter sequence that binds AP1. *J. Cell. Biochem.* **60**(4): 447-57

[61] Winchester, S. K., Selvamurugan, N., D'Alonzo, R. C. and Partridge, N. C. (2000) Developmental regulation of collagenase-3 mRNA in normal, differentiating osteoblasts through the activator protein-1 and the runt domain binding sites. *J. Biol. Chem.* **275**(30): 23310-23318

[62] Hess, J., Angel, P. and Schorpp-Kistner, M. (2004) AP-1 subunits: quarrel and harmony among siblings. *J. Cell. Sci.* **117**(Pt 25): 5965-59673

[63] Bakiri, L., Matsuo, K., Wisniewska, M., Wagner, E. F. and Yaniv, M. (2002) Promoter specificity and biological activity of tethered AP-1 dimers. *Mol. Cell. Biol.* **22**(13): 4952-4964

[64] Au, Y. P., Dobrowolska, G., Morris, D. R. and Clowes, A. W. (1994) Heparin decreases activator protein-1 binding to DNA in part by posttranslational modification of Jun B. *Circ. Res.* **75**(1): 15-22

[65] Le, N. H., van der Bent, P., Huls, G., van de Wetering, M., Loghman-Adham, M., Ong, A. C., Calvet, J. P., Clevers, H., Breuning, M. H., van Dam, H. and Peters, D. J. (2004) Aberrant polycystin-1 expression results in modification

of activator protein-1 activity, whereas Wnt signaling remains unaffected. *J. Biol. Chem.* **279**(26): 27472-27481

[66] Daury, L., Busson, M., Tourkine, N., Casas, F., Cassar-Malek, I., Wrutniak-Cabello, C., Castellazzi, M. and Cabello, G. (2001) Opposing functions of ATF2 and Fos-like transcription factors in c-Jun-mediated myogenin expression and terminal differentiation of avian myoblasts. *Oncogene* **20**(55): 7998-8008

[67] Hogan, B. (1994) Manipulating the mouse embryo: a laboratory manual. *Cold Spring Harbor Laboratory Press*, Plainview (NY)

[68] Jackson, K. A., Snyder, D. S. and Goodell, M. A. (2004) Skeletal muscle fiber-specific green autofluorescence: potential for stem cell engraftment artifacts. *Stem Cells* **22**(2): 180-187

[69] Zammit, P. S., Relaix, F., Nagata, Y., Ruiz, A. P., Collins, C. A., Partridge, T. A. and Beauchamp, J. R. (2006) Pax7 and myogenic progression in skeletal muscle satellite cells. *J. Cell. Sci.* **119**(Pt 9): 1824-1832

[70] Stellabotte, F., Dobbs-McAuliffe, B., Fernandez, D. A., Feng, X. and Devoto, S. H. (2007) Dynamic somite cell rearrangements lead to distinct waves of myotome growth. *Development* **134**(7): 1253-1257

[71] Schmidt, C., Christ, B., Patel, K. and Brand-Saberi, B. (1998) Experimental induction of BMP-4 expression leads to apoptosis in the paraxial and lateral plate mesoderm. *Dev. Biol.* **202**(2): 253-263

[72] Goulding, M. D., Lumsden, A. and Paquett, A. J. (1994) regulation of Pax-3 expression in the dermonmyotome and its role in muscle development. *Development* **120**(4): 957-971

[73] Kiefer, J. C. and Hauschka, S. D. (2001) Myf-5 is transiently expressed in nonmuscle mesoderm and exhibits dynamic regional changes within the presegmented mesoderm and somites I-IV. *Dev. Biol.* **232**(1): 77-90

[74] Fan, C. M. and Tessier-Lavigne, M. (1994) Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell* **79**(7): 1175-1186

[75] Johnson, R. L., Laufer, E., Riddle, R. D. and Tabin, C. (1994) Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. *Cell* **79**(7): 1165-1173

[76] McMahon, J. A., Takada, S., Zimmerman, L. B., Fan, C. M., Harland, R. M. and McMahon, A. P. (1998) Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. *Genes Dev.* **112**(10): 1438-14352

[77] Borycki, A. G. and Emerson, C. P. (2000) Multiple tissue interactions and signal transduction pathways control somite myogenesis. *Curr. Top. Dev. Biol.* **48**:165-224

[78] Huang, R., Zhi, Q., Izpisua-Belmonte, J. C., Christ, B. and Patel, K. (1999) Origin and development of the avian tongue muscles. *Anat. Embryol. (Berl).* **200**(2): 137-152

[79] Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D. and Relaix, F. (2003) The formation of skeletal muscle: from somite to limb. *J. Anat.* **202**(1): 59-68

[80] Roegiers, F. and Jan, Y. N. (2004) Asymmetric cell division. *Curr. Opin. Cell Biol.* **16**(2): 195-205

[81] Dietrich, S., Schubert, F. R., Healy, C., Sharpe, P. T. and Lumsden, A. (1998) Specification of the hypaxial musculature. *Development* **125**(12): 2235-2249

[82] Stockdale, F. E. (1997) Mechanisms of formation of muscle fiber types. *Cell Struct. Funct.* **22**(1): 37-43

[83] Duxson, M. J and Sheard, P. W. (1995) Formation of new myotubes occurs exclusively at the multiple innervation zones of an embryonic large muscle. *Dev. Dyn.* **204**(4): 391-405

[84] Wagner, M, Siddiqui, M. A. (2007) Signal transduction in early heart development (I): cardiogenic induction and heart tube formation. *Exp. Biol. Med. (Maywood).* **232**(7): 852-865

[85] Langer, G. A., Frank, J. S. and Brady, A. J. (1976) The myocardium. *Int. Rev. Physiol.* **9**: 191-237

[86] Chung, U. I., Kawaguchi, H., Takato, T. and Nakamura, K. (2004) Distinct osteogenic mechanisms of bones of distinct origins. *J. Orthop. Sci.* **9**(4): 410-414

[87] Christov, C., Chretien, F., Abou-Khalil, R., Bassez, G., Vallet, G., Authier, F. J., Bassaglia, Y., Shinin, V., Tajbakhsh, S., Chazaud, B. and Gherardi, R. K. (2007) Muscle satellite cells and endothelial cells: close neighbors and privileged partners. *Mol. Biol. Cell* **18**(4): 1397-1409

[88] Brand-Saberi, B. and Christ, B. (2000) Evolution and development of distinct cell lineages derived from somites. *Curr. Top. Dev. Biol.* **48**: 1-42

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

[90] Ferrari, G., Cusella-De Angelis, G., Coletta, M., Paolucci, E., Stornaiuolo, A., Cossu, G. and Mavilio, F. (1998) Muscle regeneration by bone marrow-derived myogenic progenitors. *Science* **279**(5356): 1528-1530

[91] Liu, C. and Hadjiargyrou, M. (2006) Identification and characterization of the Mustang promoter: regulation by AP-1 during myogenic differentiation. *Bone* **39**(4): 815-824

[92] Bridge, A. J., Pebernard, S., Ducraux, A., Nicoulaz, A. L. and Iggo, R. (2003) Induction of an interferon response by RNAi vectors in mammalian cells. *Nat. Genet.* **34**(3): 263-264

[93] Fish, R. J. and Kruithof, E.K. (2004) Short-term cytotoxic effects and long-term instability of RNAi delivered using lentiviral vectors. *BMC Mol Biol.* **5**: 9

[94] Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H. and Williams, B. R. (2003) Activation of the interferon system by short-interfering RNAs. *Nat. Cell Biol.* **5**(9): 834-839

[95] Pebernard, S. and Iggo, R. D. (2004) Determinants of interferon-stimulated gene induction by RNAi vectors. *Differentiation* **72**(2-3): 103-111

[96] Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**(6836): 494-498

[97] Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. and Conklin, D. S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**(8): 948-958

[98] Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**(7006): 343-349

[99] Pei, Y. and Tuschl, T. (2006) On the art of identifying effective and specific siRNAs. *Nat. Methods* **3**(9): 670-676

[100] Cullen, B. R. (2006) Enhancing and confirming the specificity of RNAi experiments. *Nat. Methods* **3**(9): 677-681

[101] Snøve, O., Rossi, J. J. (2006) Expressing short hairpin RNAs in vivo. *Nat. Methods* **3**(9): 689-695

[102] (2003) Whither RNAi? *Nat. Cell Biol.* **5**(6): 489-490

[103] Vinals, F. and Ventura, F. (2004) Myogenin protein stability is decreased by BMP-2 through a mechanism implicating Id1. *J. Biol. Chem.* **279**(44): 45766-45772

[104] Kucera, J., Walro, J. M. and Gorza, L. (1992) Expression of type-specific MHC isoforms in rat intrafusal muscle fibers. *J. Histochem. Cytochem.* **40**(2): 293-307

[105] Pavlath, G. K. and Horsley, V. (2003) Cell fusion in skeletal muscle-central role of NFATC2 in regulating muscle cell size. *Cell Cycle* **2**(5): 420-423

[106] Towler, M. C., Kaufman, S. J. and Brodsky, F. M. (2004) Membrane traffic in skeletal muscle. *Traffic* **5**(3): 129-139

[107] Horsley, V. and Pavlath, G. K. (2004) Forming a multinucleated cell: molecules that regulate myoblast fusion. *Cells Tissues Organs* **176**(1-3): 67-78

[108] Chen, E. H. and Olson, E. N. (2004) Towards a molecular pathway for myoblast fusion in Drosophila. *Trends Cell Biol.* **14**(8): 452-460

[109] Ko, J. A., Kimura, Y., Matsuura, K., Yamamoto, H., Gondo ,T. and Inui, M. (2006) PDZRN3 (LNX3, SEMCAP3) is required for the differentiation of C2C12 myoblasts into myotubes. *J. Cell Sci.* **119**(Pt 24): 5106-5113

[110] Constantin, B., Cognard, C. and Raymond, G. (1996) Myoblast fusion requires cytosolic calcium elevation but not activation of voltage-dependent calcium channels. *Cell Calcium* **19**(5): 365-374

[111] Charrasse, S., Comunale, F., Fortier, M., Portales-Casamar, E., Debant, A. and Gauthier-Rouviere, C. (2007) M-cadherin activates Rac1 GTPase through the Rho-GEF trio during myoblast fusion. *Mol. Biol. Cell.* **18**(5):1734-1743

[112] Tomczak, K. K., Marinescu, V. D., Ramoni, M. F., Sanoudou, D., Montanaro, F., Han, M., Kunkel, L. M., Kohane, I. S. and Beggs, A. H. (2004) Expression profiling and identification of novel genes involved in myogenic differentiation. *FASEB J.* **18**(2):403-405

[113] Kloiber, K., Weiskirchen, R., Kräutler, B., Bister, K. and Konrat, R. (1999) Mutational analysis and NMR spectroscopy of quail cysteine and glycine-rich protein CRP2 reveal an intrinsic segmental flexibility of LIM domains. *J. Mol. Biol.* **292**(4): 893-908

[114] Kostek, M. C., Chen, Y. W., Cuthbertson, D. J., Shi, R., Fedele, M. J., Esser, K. A. and Rennie, M. J. (2007) Gene expression responses over 24h to lengthening and shortening contractions in human muscle: major changes in CSRP3, MUSTN1, SIX1 and FBXO32. *Physiol. Genomics* [Epub ahead of print]

[115] Fields, S. and Sternglanz, R. (1994) The two-hybrid system: an assay for protein-protein interactions. *Trends Genet.* **10**(8): 286-292

[116] Luban, J. and Goff, S. P. (1995) The yeast two-hybrid system for studying protein-protein interactions. *Curr. Opin. Biotechnol.* **6**(1): 59-64

[117] Warbrick, E. (2000) The puzzle of PCNA's many partners. *Bioessays* **22**(11): 997-1006

[118] Van Aelst, L., Barr, M., Marcus, S., Polverino, A. and Wigler, M. (1993) Complex formation between RAS and RAF and other protein kinases. *Proc. Natl. Acad. Sci. USA* **90**(13): 6213-6217

[119] Bartel, P., Chien, C. T., Sternglanz, R. and Fields, S. (1993) Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* **14**(6): 920-924