MUNC, An eRNA Upstream From the MYOD Gene, Induces Myogenic Genes *in trans*, Independently of MyoD

Magdalena Anna Cichewicz

MS Biotechnology, Jagiellonian University, Krakow, 2011

A dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the degree of Doctor of Philosophy

Department of Biochemistry and Molecular Genetics

University of Virginia

July, 2017

Table of Contents

ABSTRACT

Myogenesis occurs during embryonic development while all tissues are being defined, and during regeneration of adult muscle tissue after injury and inflammation. Among myogenesis-specific transcriptional factor, the two main players, MyoD and myogenin, are crucial for the process. In a healthy organism, all myogenic factors work in a strictly time-regulated manner to build new muscle fibers, and a feed-forward mechanism of the molecular cascade has been described. Although we know the main protein factors regulating myogenesis, still we are unable to explain the molecular pathology of various muscular dystrophies.

Since genome wide studies have become widely accessible, researchers now recognize novel transcripts which do not code for proteins but are functional as RNAs. It is established that the majority of transcripts produced by the mammalian genome does not express proteins, but is crucial for proper homeostasis of various biological systems. There are known examples of long noncoding RNAs (lncRNAs) which are tissue-specific, and some of them are cancer-specific. They function as scaffold molecules, chromatin modifiers, and post-transcriptional regulators of mRNAs.

The goal of my project was to describe novel pro-myogenic long noncoding RNAs induced during muscle differentiation, and to investigate their functions and mechanisms. In the beginning of the project, publicly available genome-wide datasets of differentiating murine myoblasts, C2C12 cells, were analyzed. After combining RNA-Seq, RNA PolII ChIP-Seq and H3K4me3 ChIP-Seq, a pool of potential long noncoding transcripts induced during differentiation was established. After confirmation that several of these lncRNAs are indeed induced during differentiation, we decided to focus on one specific transcript,

MyoD Upstream NonCoding Element (MUNC), whose genomic locus partially overlaps with the DRR-enhancer for MyoD. We characterized the sequence of MUNC, and its two isoforms: unspliced and spliced. We found a functional human homolog of the transcript. *In vitro* and *in vivo* depletion experiments established that MUNC is pro-myogenic. By stable overexpression we learned that MUNC does not have one domain responsible for its action, rather at least two fragments of MUNC can independently induce myogenic transcripts. Partial overlap between the MUNC locus and an enhancer for MYOD sequence suggested that the only function of MUNC is activation of MYOD expression as an enhancer RNA (eRNA). To investigate this possibility, we generated a MYOD knockout cell line. By overexpressing MUNC in MYOD deleted cells we discovered that in the absence of MyoD, MUNC is still able to regulate specific genes, mostly skeletal muscle related. We characterized genes regulated by the coordinated action of MyoD and MUNC, and genes regulated by their separate pathways. Our study showed that MUNC does not only work as an enhancer inducing MYOD expression, but also has MyoD-independent functions during myogenesis. Thus it is both an eRNA and an lncRNA.

DEDICATION

To my parents, Jadwiga and Tadeusz Wegrzynscy, and my husband, Karol Cichewicz, for their love and continuous support.

CHAPTER I: INTRODUCTION

Myogenesis- a process of skeletal muscle development and regeneration

Skeletal myogenesis is a process of skeletal muscle development during embryogenesis and during regeneration of adult muscles that enables movement of the whole body. Additionally, muscle tissue is important for energy metabolism because of its high total mass relatively to the whole body. Skeletal muscles are responsible for the majority of glucose uptake in response to insulin. During embryogenesis, paraxial mesoderm undergoes segmentation to build somites. Somites differentiate into dermomyotome, and later to dermatomes and myotomes. During human development first myoblast fusions take place between 10 and 13 weeks of gestation. By 15 weeks myotubes are grouped into bundles and build myofibers with peripheral nuclei. By birth fibers are differentiated into type I fibers (slow-twitch fibers) and type II fibers (fast-twitch fibers)[1]. During adult muscle regeneration, satellite cells which are stem cells for muscle have a crucial role. Under homeostatic conditions –satellite cells are quiescent and they reside between the basal lamina and the sarcolemma of the myofibers. In response to physiologic and pathologic stimuli, such as denervation, muscle inflammation, exercise or injury satellite cells are activated, proliferate and eventually exit the cell cycle to differentiate and fuse to form new myofibers. The pool of satellite cells is maintained by asymmetric divisions of satellite cells during the proliferation stage[2].

Skeletal muscle disorders

Dystrophinopathies are diseases caused by mutations in a gene on X chromosome, coding for a protein dystrophin. The dystrophin gene spanning 2.5 million bp is the largest

gene described in humans. It produces tissue specific isoforms of transcripts and proteins, which are present in muscles, brain, lymphocytes. The protein is associated with the plasma membrane of the muscle, and builds a complex with integral membrane proteins to connect the basal lamina of the extracellular matrix with the inner cytoskeleton. Dystrophin mutations causing disorders include intragenic deletions and rarely duplications. There is no clear correlation between size of the deletion and severity of the disease. Different mutations (according to their location and size) may give similar phenotype of the disease. Commonly, but not without exceptions, mutations with loss of the reading frame are linked to the more severe phenotype of Duchenne muscular dystrophy (DMD), while in-frame mutations/deletions lead to the milder Becker's muscular dystrophy (BMD). Duchenne's dystrophy is the most common form of inherited muscle disease of childhood with both skeletal and cardiac muscles being affected. Progressive muscle necrosis takes place, independent ambulation is lost by early adolescence. Other symptoms are cardiomyopathy and disorder of respiratory function with the patients dying by their 30s. Becker's dystrophy is milder version of the disease that also involves skeletal impairment and dilated cardiomyopathy[3] [4]. Prevalence of DMD is 4.78 per 100,000 boys, while that of BMD is 1.53 in 100,000 boys[5]. Apart from Duchenne and Becker's dystrophies there is a group of highly heterogenous Limb-girdle muscular dystrophies (LGMDs). These disorders are rare, characterized by weakness and wasting of the pelvic and shoulder girdle muscles. They are harder to diagnose because of genetic heterogeneity[4]. There is a need for development and validation of biomarkers which may help to speed up drug development and lead to faster access to new treatments. Biomarkers used now are dystrophin, CK, MMP9[6]. Using meta-analysis comparing healthy patients and DMD patients a list of

deregulated genes in DMD as potential biomarkers for the disease was proposed[7]. A similar study was performed using mice models for various types of muscular dystrophy. In this study authors focused on microRNAs as potential markers[8]. So far, the proposed list of biomarkers contains protein-coding transcripts and miRs. Considering the size of the pool of long noncoding transcripts that are produced from the genome, it is possible that we are still missing a great group of RNA molecules whose quantification could help to diagnose muscular diseases.

Rhabdomyosarcoma (RMS) is a type of cancer of muscle tissue. It is mainly a pediatric disease with about 5% of child cancers being RMS. They tend to develop in head and neck area, or in genital and urinary tracks (embryonal RMS) and in arms, legs, or trunk (alveolar RMS). Main approaches to treat RMS are surgery and chemotherapy. Overall, 5-year survival rate in children is 61%[9]. Approximate incidence of RMS is 4.5 cases per million children[10]. Characterized genes up-regulated in RMS are *FGFR4, NOTCH2, UBE2C, UHRF1, YWHAB*. The disease blocks regular skeletal muscle development, leading to unregulated proliferation. Common mutation in RMS is fusion of PAX-FOXO1 genes, which commit mesenchymal stem cells to a myogenic lineage and skeletal myogenesis by trans activating MYOD and myogenin[11].

Muscle atrophy is a condition of muscle wasting after injury/surgery or when muscles are not used or in elderly population (sarcopenia). It was documented that people by the age of 80 lose 30% of their skeletal muscle fibers. The maintained muscles are not flexible, they are weaker and impaired in their regeneration after injury. Sarcopenia causes a decline in mobility and frequent falls. The main intervention is exercise training, which delays the process of atrophy[12].

Molecular biological processes regulating myogenesis

Embryonic progenitor cells and adult quiescent satellite cells express transcription factors *Pax3* and *Pax7*[13 14]. During active proliferation and differentiation helix-loop helix family transcription factors are activated: MyoD, myogenin, Myf5, Mrf4[15]. MyoD, and Myf5 are engaged in the establishment and the maintenance of myogenic precursor cells, while myogenin and Mrf4 are important for later steps of differentiation[16 17]. Satellite cells are heterogenous in Myf. Myf^{-} cells are capable of asymmetrical divisions, and others that are $Myf5^{+}$ are committed to differentiation[18]. MyoD is induced shortly after satellite cell activation. After cells are committed to differentiation they start producing myogenin and Mrf4 as they undergo multiple rounds of division. Eventually, after exiting cell cycle they express muscle-specific genes, alpha-actin, troponin T, myosin heavy chain and fuse into myotubes to contribute to mature muscle fiber[19]. All four myogenic transcription factors are able to induce trans-differentiation in non-muscle cell lines, but with different potency. It was shown by reporter CAT (chloramphenicol acetylatransferase) assay that out of the four, MyoD is the most potent, even factors belong to the same protein family, have distinct biological roles[20]. The high potential of MyoD to change cell reprogramming to skeletal muscle type was shown by transfecting it into different cell types, liver, melanoma, neuroblastoma, and fibroblasts, where it forced these cells to differentiate into muscle–like cells[21 22 23 24]. Overexpression of MYOD in non-muscle cells results in relaxation of chromatin region in proximity to myogenin and MCK loci. This suggests that when MyoD binds to these sites, it does not only activate their transcription, but also changes local chromatin landscape[25]. $MYOD^{-/-}$ mice are viable and develop healthy muscles but are defective in their regeneration capacity after injury[26 27]. Primary myoblasts derived from

MYOD$^{-/-}$ mice show increased level of Myf5, suggesting a compensatory mechanism between these two proteins. The differentiation process is still delayed in *MYOD*$^{-/-}$ cells compared to WT cells. They show higher proliferation rate than WT[28]. MyoD and Myf5 functions are somewhat redundant. Mice lacking one or another are able to develop skeletal muscle tissue and are viable, but once interbred, double knockout pups do not develop muscle, are immobile and die soon after birth[29]. By comparison between the single mutants embryos, MyoD seems to be more important for the development of hypaxial musculature (limb and abdominal wall), and Myf5 for epaxial musculature (paraspinal and intercostal)[30]. Lineage tracing in mice study showed the presence of two myoblast lineages, Myf5-expressing, and MyoD-expressing. Upon ablation of *MYF5* MyoD-expressing progenitors increase, so myogenesis is recovered, suggesting that the MyoD^{+} population of progenitors is more dominant[31 32]. Molecular, genome-wide experiments show that both factors bind to the same pool of genomic sites, but their bindings have different effects. Myf5 induces histone acetylation marks of bound loci, but does not change gene expression, while MyoD binding robustly induces transcription of its targets[33].

MyoD protein is highly conserved between vertebrates[34]. Like other members of basic helix loop helix (bHLH) family, MyoD consists of four conserved domains: (1) a TAD (transcriptional activation domain) in the amino terminal region, (2) a histidine/cysteine rich domain (H/C domain), (3) a bHLH in the central region, and (4) an amphipathic alpha helix domain in the carboxy terminal region. TAD is important for interactions between MyoD and the histone acetyltransferase p300/CBP, responsible for transactivation by MyoD[35]. The bHLH domain dimerizes with E box proteins E12, binds DNA and changes TAD conformation. MyoD binds to E box consensus sequence

(CANNTG), found in the promoters and enhancers of muscle specific genes[35]. MyoD works in a feed-forward mechanism, activating early- and late-differentiation factors in a specific temporal manner[36 37].

Another family of transcription factors important for muscle differentiation are MEF2A-D (myocyte enhancer factors) proteins. By shRNA screening in C2C12 cells it was shown that functions of these proteins are not fully overlapping and that they regulate distinct groups of myogenic genes. Only MEF2A knockdown impairs C2C12 cell differentiation, and knockdown of other MEFs does not affect the process, suggesting a dominant role of Mef2C in the process[38]. Mice with muscle-specific deletion of MEF2C show myogenesis failure and lethality and mice which survive to adulthood show abnormalities in myofiber types. Lack of MEF2A causes cardiovascular defects without skeletal muscle defects, and MEF2D single deletions do not deregulate muscle development[39]. In a molecular level it was shown that Mef2D interacts with MyoD at myogenic promoters[40]. Interactions between Ashl2 and Mef2D and weaker interactions between Mef2C and Ashl2 were recognized in overexpression experiments. It suggested that Mef2C/D proteins recruit Ashl2 to methylate H3K4me3 on myogenic promoters[39].

MyoD activity is regulated temporally at both transcriptional and post transcriptional level. A proto-oncogene c-Jun inhibits expression of MyoD, and c-Jun and c-Fos interact with the myogenic factor inhibiting its activity[41 42]. MyoD is a short-lived protein with its half-life in proliferating conditions being 30 min. It contains several putative cyclin-dependent kinases phosphorylation sites, of which at least one, S200, is responsible for half-life regulation. S200 is shown to be phosphorylated by Cdk1 and Cdk2.

MyoD is ubiquitylated and degraded by proteasome complex[43] [44]. Id (inhibitor of differentiation) proteins interact with MyoD and E proteins to inhibit their activity. Id proteins are HLH proteins lacking basic region in their sequence. As a result when Id binds to MyoD, it forms a nonfunctional complex since this interaction decreases DNA binding affinity of MyoD[45]. In undifferentiated cells MyoD interacts with HDAC1, which inhibits MyoD activity and maintains deacetylated, repressed chromatin at myogenic promoters and enhancers where MyoD is bound. On the other hand, HDAC1 is diluted away from MyoD by pRb in myotubes[46]. There are studies suggesting that MyoD is recruited to its target myogenic sites (specifically MYOG promoter site) by constitutively bound Pbx1 protein, and later MyoD targets remodeling complex SWI/SNF to change local chromatin accessibility[47]. Activation by MyoD requires several modifiers and remodeling proteins such as SWI/SNF complex, which is an ATP-dependent chromatin remodeling complex. It functions by altering the chromatin structure at myogenic promoters, making it more accessible for transcription factors[48]. MyoD recruits p300/CBP and PCAF to the DNA[49]. Both, p300 and PCAF were shown to acetylate MyoD at conserved Lys (p300- K99, K102, PCAF- K99, K102, K104) within bHLH domain. Acetylation at these residues causes conformational change of the protein and increases its affinity to its DNA target sites[50] [51] [52] [53].

Myogenesis is tightly regulated by the extracellular environment. Transforming growth factor β (TGF-β) suppresses muscle differentiation. It binds to its Ser/Thr kinase receptor, resulting in phosphorylation of Smad2 and Smad3, which form complexes with Smad4. After forming a complex, Smads translocate to the nucleus and change transcription level of target genes. In myogenic cells it was shown that Smad3 is able to

bind to bHLH domain of MyoD and blocks MyoD's binding to its functional partner, E protein[54]. Lack of growth factors in the extracellular environment triggers muscle differentiation[55]. After mitogen withdrawal, P38/MAPK pathway is activated and stimulates activity of MyoD and MEF2 proteins[56]. Autocrine signaling of IGF-II was also shown to positively regulate activity of MyoD and control myogenesis [57].

Distal Regulatory Region - enhancer for *MYOD* and its relevance in muscle differentiation

Gene expression is highly regulated in spatial and in temporal manner. There are various genomic sequences which are significant for this regulation. The most common, working *in cis,* are promoters, located in a proximity to TSS of a gene of interest, and enhancers, located a few kbs upstream from TSS of a gene of interest. Promoters and enhancers are often characterized by reporter assays, showing the minimal sequence of the upstream region required for expression of a gene of interest (reviewed in:[58]). There are three genomic regulatory elements that positively regulate *MYOD* gene expression: Core Enhancer Region (CER), 4kb element located about 23kb upstream from *MYOD* TSS; Distal Regulatory Region (DRR), located 4.7-5.4kb upstream from *MYOD* gene locus, and Proximal Regulatory Region (*MYOD* promoter - PRR), located -275bp- +1 from *MYOD* TSS[59]. Reporter assays performed on chicken primary cells, and mammalian cell lines showed the significance of these specific sequences for induction of a downstream gene. Additionally, by injections of a series of mutated sequences 6kb upstream of *MYOD* into *Xenopus* embryos PRR was found to give very weak reporter gene activation. For high activation of the reporter both DRR and PRR are required. A positive role of DRR on

MYOD activation was shown in myoblasts and 10T1/2 fibroblast cells, both of which convert to myotubes when transfected with *MYOD* [60,61]. Studies using transgenic mice with a fragment from -24kb to the TSS of *MYOD* ligated upstream of a lacZ reporter, showed that both CER and DRR regions are crucial for tissue specific expression of MyoD *in vivo*. β-Gal staining of these mouse embryos mimicked expression pattern of endogenous MyoD in the body[62]. Additional confirmation of the importance of DRR for *MYOD* expression *in vivo* came from targeted mutagenesis of DRR. $DRR^{-/-}$ mice have reduced MyoD expression in E10.5 embryos, as well as in adult leg muscles[63]. DRR sequence is conserved between mouse and human. Comparison between established mouse DRR sequence and human genomic sequence at -4.5kb to -5.2kb upstream from the *MYOD* TSS revealed blocks of high similarity, with the highest homology in the first 445bp – 71%. Conservation of DRR function in both species was confirmed when a reporter gene regulated by human DRR showed a muscle-specific expression pattern in mice[62]. DRR sequence has 3 consensus MyoD binding sites, which suggests a positive autoregulatory loop. MyoD induces *MYOD* expression, and MyoD protein by binding to DRR is able to induce its expression even more[60]. A serum response factor (SRF) and YY1 binding sequence is also seen in the DRR, and needs to be intact for regular MyoD expression. Even single point mutation within this element abolish SRF or YY1 binding and correlates with lower expression of MyoD[64].

Genome-wide studies investigating myogenesis

Scientists interested in myogenesis started employing genome wide techniques in molecular biology as soon as they appeared to better understand the process of myogenesis and the function of crucial factors during muscle differentiation. The main *in vitro* models

used in such studies are C2C12 cells, immortalized murine cell line derived from satellite cell, and primary cultures of murine or human satellite cells. By controlling the culture medium conditions cells can be maintained in an undifferentiated state of proliferating muscle progenitor cells (presence of growth factors), or differentiating cells (low mitogens). Microarray data[65 66 67] and RNA-seq datasets[68] compare total transcriptomes of undifferentiated and differentiated cells, so as to identify groups of genes that are differentially expressed in between the two conditions. Thanks to numerous chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) studies of myogenic transcription factors in C2C12 cell line or in MyoD-transfected fibroblasts[69 67 70 71], we can point genes which are regulated directly by specific factors, including many that are regulated simultaneously by a few transcription factors. Genome-wide studies confirmed that E-boxes are the most common genomic binding motifs for MyoD. Analysis of chromatin marks specific for promoters and enhancers (H3K4me1-3, H3K9Ac, H4K12Ac, H3K18Ac, PolII), gene bodies (H3K36me3), and marks associated with euchromatin and heterochromatin (H3K9me3, H3K27me3) give us great knowledge about different state of chromatin landscape during muscle differentiation[72].

Genome-wide experiments showing a universe of potential noncoding transcripts

The paradigm, that RNA molecules are only an intermediate during protein synthesis pathway needed modification with the discovery of functional noncoding RNA (ncRNA). Characterization of multiple small ncRNAs: micro-RNAs, si-RNAs and piwi-RNAs [73 74 75 76] helped to understand that RNAs *per se* are also crucial for cell biology. Genome-wide techniques such as next generation sequencing and microarray-

hybridization performed by numerous research groups during the last decade have characterized the mammalian protein coding and noncoding transcriptome. This is when it became apparent that there is another, heterogenous group of functional noncoding RNAs, long noncoding RNAs (lncRNAs). LncRNAs are transcripts longer than 200 nt, which do not code for proteins, but are functional as RNA molecules. One of the first genome-wide analysis was RIKEN Mouse Gene Encyclopedia Project. The Project's goal was to determine the total transcriptome of mouse species. Full-length cDNA libraries derived from various tissues and developmental stages were developed and sequenced[77 78]. There were also other groups using similar approaches to build mouse and human transcriptome catalogs[79 80 81]. Chromatin-state maps also help to establish functional transcriptional units. Combination of ChIP-Seq analysis targeting genomic regions occupied by RNA PolII, H3K4me3 (specific for promoters of transcripts), and H3K36me3 (along the length of the transcribed region) define the genomic regions undergoing active transcription[82]. Genome-wide projects utilizing different approaches lead to similar conclusions. All of them show that the mammalian transcriptome contains not only protein-coding transcripts, but also transcripts that have low protein-coding capacity and in fact, the latter are the major component of the transcriptome. To facilitate research on lncRNAs there are catalogs characterizing expression profiles of noncoding transcripts in human and mouse[83] such as lists of lncRNAs showing biological function in eukaryotes and databases containing specification of lncRNAs built using computational approaches[84].

LncRNAs can be categorized according to their genomic position relative to protein-coding genes loci. There are lncRNAs that are sense or antisense with annotated genes, overlapping with exons of genes, derived from introns of known genes, or intergenic

lying in genomic intervals between two genes[85]. LncRNAs use different mechanisms of action. There are nuclear lncRNAs, which affect expression of other genomic sequences *in cis*, some of which are transcribed from enhancers. Enhancers are regulatory genomic sequences that facilitate RNA PolII recruitment to the promoters of target genes sometimes by looping out the intervening sequences[86]. Noncoding RNAs derived from enhancer loci are able to control level of expression for their adjacent and neighboring genes and have been named enhancer RNAs (eRNAs)[87]. LncRNA transcribed from HOXA locus, HOTTIP, regulates its target gene by chromosomal looping[88]. Similarly, ncRNAs are able to activate target gene expression by DNA looping through interactions with Mediator[89]. Another class of lncRNAs are Cis-Natural Antisense Transcripts (cis-NATs). They regulate their antisense gene expression. They work either through transcriptional collision during transcription, or post-transcriptionally as interfering RNA[90]. LncRNAs operate not only *in cis*, but also *in trans*. There are transcripts which are able to diffuse from the coding site and act on distal genes, also at other chromosomes. One of the best characterized lncRNAs is a noncoding transcript Xist, which regulates dosage compensation for X-linked genes between females and males. Xist is a 15kb sequence transcribed from one of the X chromosomes in females. It coats the whole X chromosome to inactivate its transcription[91]. HOTAIR is another lncRNA, which represses transcription of its target gene *in trans* across 40 kb. It is encoded in HOXA, one of the four chromosomal loci comprising the HOX cluster (HOXA-D), and transcribed in antisense direction[92]. This lncRNA binds to protein complexes which deactivate chromosomal landscape and inhibits HOXD expression[93]. Human Alu ncRNA is a transcript which binds to RNA PolII and, inhibits mRNA synthesis after heat shock[94]. There is also genome-wide analysis showing a group of transcripts

which may inhibit expression of different genes *in trans* by interacting with the silencing PRC2 chromatin modification complex. They were identified by RNA immunoprecipitation of the PRC2 complex followed by high throughput sequencing of associated RNA (RIP-Seq)[95]. Another example of a well characterized lncRNA is cytoplasmic H19 RNA, a 2.5 kb noncoding transcript. It is very abundant in the developing mouse embryo but after birth is expressed only in skeletal muscle tissue[96]. In exon 1 of H19 there is an embedded microRNA, miR-675, which is involved in promoting skeletal muscle differentiation[111]. H19 is also famous because it undergoes imprinting and is transcribed only from maternal chromosomal copy[97 98]. There is evidence that lncRNAs are important for the process of development and differentiation. Noncoding transcripts are differentially expressed during differentiation of mouse embryonic stem cells[99]. One specific example is TINCR lncRNA, which acts in somatic epidermal differentiation by binding to and stabilizing mRNAs crucial for the process[100]. Initially lncRNAs were thought to be transcriptional noise resulting from low fidelity of RNA PolII but the evidence for their functionality is increasing. The functional outcome is sometimes hinted by the expression of particular lncRNAs in developmental contexts. Most of already characterized lncRNAs have specific cellular localizations, and for some of them the molecular mechanism of action has been described. Sequence homology analysis shows that lncRNAs have lower sequence conservation between species than protein-coding RNAs, or micro-RNAs[101]. Although the secondary structures of some lncRNAs are conserved between human and mice despite the divergence in sequence. Thus, while all lncRNAs may not be functionally conserved between species, there are definitely subsets of lncRNAs with conserved function between species[102].

Myogenic long noncoding RNAs

Steroid Receptor RNA Activator (SRA) RNA co-regulates transcriptional factors in different tissues and is also present in muscle. The level of SRA during muscle differentiation does not change. The SRA RNA interacts with both RNA helicases p68/72 and MyoD and positively regulates MyoD's function. Knock down of SRA impairs skeletal myoblast differentiation *in vitro*[103]. Interestingly, there is a protein product encoded by the SRA locus called SRAP. The balance between production of RNA and protein is achieved by alternative splicing mechanism. Production of mRNA coding for the protein is a pro-proliferation signal, while increase of the SRA lncRNA lacking ORF is a pro-differentiation signal. The ratio noncoding SRA/coding SRA shows equal levels of both transcripts in samples from myotonic dystrophy muscle. Both, protein and RNA interact together and this interaction block pro-differentiation function of SRA RNA. It shows additional level of MyoD regulation by two products coded in the same locus[104].

Linc-MD1 is an example of competing endogenous RNA (ceRNA). This pro-differentiation factor (shown by knock down and overexpression experiments) is present in the cytoplasm and induced during human and mouse myoblasts differentiation *in vitro*. It sponges microRNAs miR-133 and miR-135 and de-represses their targets, MAML1 and MEF2C, respectively. Both these targets are transcription factors regulating expression of muscle genes, so linc-MD1 functions as a regulator of myogenesis. A strong evidence for its role comes from comparison of human myoblasts derived from healthy individuals and individuals with Duchenne muscle dystrophy. Healthy myoblasts with higher level of linc-MD1 expression differentiated more efficiently compared to Duchenne myoblasts[105].

Additionally, lincMD1 is the parental transcript (primary microRNA) for miR-133b, so the regulation of lincMD1 and miR-133b expression and function is tightly coordinated. Another target of miR-133b, HuR protein is important for the proper balance between miR-133b and linc-MD1. HuR physically associates with linc-MD1protecting the linc from being processed by Drosha to produce miR-133b. Additionally, linc-MD1 sponges existing miR-133b. Thus induction of HuR, decreases the activity of miR-133b, which leads to protection of HuR mRNA from miR-133b building a positive feed-forward loop for HuR production[106].

Gtl2 (Meg3) is a long noncoding nuclear transcript coded within 1Mb long imprinted region *Dlk1-Gtl2*. The locus codes for lncRNAs which are expressed maternally and for protein-coding genes which are expressed paternally. The pattern of expression is responsible for proper organogenesis during development. The imprinting is regulated by a cis-regulatory region IG-DMR located between Dlk1 and Gtl2. Mice carrying paternal deletion of Gtl2 are healthy, whereas mice lacking maternal copy die perinatally. Histological analysis of pups carrying maternal deletion showed abnormalities in all muscle tissue, while other tissues looked healthy[107]. In BT-549 cells (breast cancer cell line) Gtl2 interacts with the PRC2 complex at distal regulatory elements of genes from the TGFβ pathway. Together, Gtl2 and PRC2 maintain H3K27me3 levels of enhancers which results in silencing transcription of their target genes. Genome-wide analysis showed that Gtl2 binds to sequences enriched with GA-rich motifs, forming RNA-DNA triplex structures. Domains of Gtl2 which are required for DNA binding and PRC2 binding are functionally distinct. Mutation of triplex-forming oligonucleotide of Gtl2 disrupts interaction with chromatin, without changing affinity to PRC2. Downregulation of Gtl2

decreases abundance of EZH2 at specific genomic loci suggesting that lncRNA recruits silencing complex to the chromatin[108].

H19, a lncRNA imprinted in mice and humans, is strongly expressed mainly in the skeletal muscle from the maternal allele during embryonic development and after birth. H19 is a cytoplasmic 2.3 kb long transcript. There are conflicting results whether H19 plays pro-proliferation or pro-differentiation role. It was shown that H19 is induced during differentiation, and its knock down causes even more efficient differentiation. The authors suggest that H19 contains functional Let-7 interaction sites that sponge Let-7 from its targets and thus inhibits differentiation[109]. Another study showed a pro-differentiation role of H19 by *in vitro* knock down and also *in vivo* impairment of H19 in skeletal muscles. H19 locus encodes for two miRs: miR-675-3p and miR-675-5p, which are upregulated during differentiation. Skeletal muscle regeneration is impaired in H19 mutant mice, but can be rescued by ectopic overexpression of miR-675-3p and miR-675-5p. The targets for these miRs, Smad1, Smad5, and Cdc6 are pro-proliferative factors. Thus H19 is the primary-microRNA transcript for miR-675-3p and miR-675-5p, which block Smads and Cdc6 and thus promote differentiation[110].

Nctc1 is a lincRNA coded by the region adjacent to Igf2-H19 locus (Figure 1.). It is known that Igf2 and H19 are reciprocally imprinted, Igf2 is transcribed from paternal copy of the locus, while H19 is active from maternal locus. Expression of both genes is dependent on shared regulatory elements. H19ICR is a regulatory region located between Igf2 and H19 genes. H19ICR is bound by CTCF protein on maternal chromosome. CTCF together with cohesins organizes the chromosome into loops which favor H19 expression.

H19ICR knock out causes biallelic expression of both genes, Igf2 and H19. Additional enhancers, EE and CME, important for regulation of H19 and Igf2 expression are located downstream from H19. Nctc1 is a skeletal muscle specific lncRNA expressed through CME enhancer. There are two isoforms of Nctc1, one including exons 1-3, and another with only exons 1-2. While CME loops with Igf2 promoter only on paternal chromosome and loops with H19 promoter only on maternal chromosome, looping between CME and Nctc1 promoter is observed on both types of chromosomes. Quantitative comparison showed that paternal expression of Nctc1 is higher than maternal one, suggesting that two types of loops are not equal with regard to the activation of the Nctc1 transcript[111]. Although Nctc1 is skeletal muscle specific, its specific role in this tissue is not explained yet.

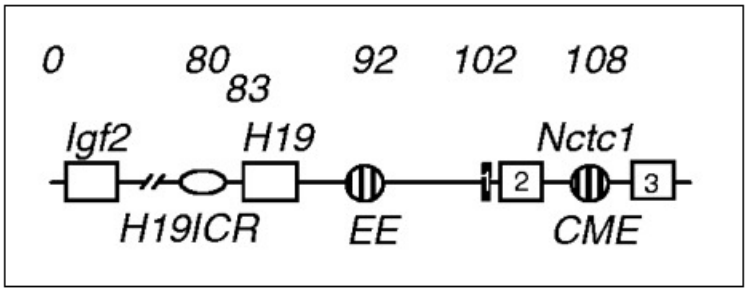


Figure 1. Adapted from Eun B. et al. Genomic locus of Igf2/H19/Nctc1. Numbers describe kbp positions relatively to Igf2 TSS.

LincRNA Yam-1 is a transcript that is positively regulated by Ying Yang 1 (YY1) transcription factor. YY1 is highly expressed in undifferentiated myoblasts, but is gradually downregulated during differentiation. YY1 may play a dual function in myogenesis by recruiting silencing chromatin modifiers (PRC2 complex). but also activating some of its genomic targets. Similarly, Yam-1 is abundant in undifferentiated cells and its negative

role in myogenesis was shown by *in vitro* and *in vivo* knockdown experiments. Differentiation of cells or regeneration of skeletal muscle after cardiotoxin injection was more efficient when Yam-I was depleted. Additional experiments revealed that Yam-1 positively regulates expression of miR-715, which further regulates represses Wnt7b, a positive regulator of differentiation. There are also other myogenic lincRNAs regulated by YY1 such as Yam2-4. They play diverse functions in myogenesis, and their expression patterns during differentiation are different case by case[112].

Another potential role for cytoplasmic lncRNAs is control of their targets by mRNA decay. Rodent M1/2-sbRNAs (B1-B4) are induced during myogenesis. They mediate Staufen1-mediated mRNA decay (SMD) by forming duplexes with SINE-containing 3'UTRs of mRNAs. M1/2-sbRNAs target the mRNAs of pro-proliferative genes such as mRNF168, mCdc6, mTraf6 and thus promote myogenesis. Consistent with this, a decrease in efficiency of SMD inhibits myogenesis[113].

There are lncRNAs, which are abundantly expressed in different tissues under homeostasis or during pathological states. A good example of such a lincRNA is MALAT1, Metastasis-Associated Lung Adenocarcinoma Transcript 1. MALAT1 is a nuclear lincRNA, which promotes proliferation, migration and invasion in various cancers (reviewed in[114]). Additionally, *in vivo* study, where mice were treated with Myostatin, a member of the TGF-β family, a negative regulator of muscle development, showed that MALAT1 is one of the most significantly inhibited transcripts in Myostatin-treated muscles. In differentiating myoblasts, MALAT1 is induced. Si-MALAT1 treatment in undifferentiated cells decreases cell proliferation, while si-MALAT1 treatment in

differentiating cells decreases myogenin level. The latter result suggests a pro-differentiation role of MALAT-1, but more experiments have to be performed to explain the phenotype of MALAT1 knock down in proliferating undifferentiated cells[115]. Custom microarray screening of differentiating myoblasts revealed a group of transcripts induced during differentiation. Two of them are Men ε – a 3.7kb transcript, also known as Neat1, and Menβ – a 23kb transcript, derived from the same locus. They are both localized to nuclear paraspeckles, and probably associate with a known component of paraspeckles, NONO. Knock down of Neat1 in HeLa cells suggested that both transcripts are important for the integrity of nuclear paraspeckles. These transcripts are also present in other tissue, and their specific role in muscle cells is not clear[116].

In muscle tissue, there are examples of lncRNAs derived from the enhancer regions of *MYOD* gene: mostly nuclear CE-RNA and DRR-RNA/MUNC. Both RNAs were shown to be pro-differentiation factors. *In vitro* and *in vivo* knockdown and knock out experiments documented that both transcripts are responsible for upregulating *MYOD* transcription. Their depletion also affects MyoD protein binding to some of its genomic target sites. My work establishes that these enhancer RNAs are not just a side-effect of open chromatin in these enhancer regions, that their function is not simply to open chromatin and promote the expression of *MYOD*. I show that they are functional transcripts with specific roles in muscle differentiation although their mechanism of action is not fully understood[69,117].

Another example of promyogenic lncRNA functioning *in cis* is Dum (developmental pluripotency-associated 2 (Dppa2) Upstream binding muscle RNA). Dppa2 is a factor important for maintaining pluripotent state of embryonic stem cells. Its

knockdown causes differentiation of cells[118]. During muscle differentiation Dum is activated by MyoD. Dum interacts with histone methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Dum recruits them to the Dppa2 locus to silence Dppa2 transcription.[119].

DBE-T, a nuclear lncRNA, is produced selectively in patients with Facioscapulohumeral muscular dystrophy (FSHD). FSHD is associated with the copy number reduction of D4Z4 repeats (CNGCCATNDNND) close to FSHD locus on chromosome 4q35. In healthy individuals, Polycomb group proteins are recruited to D4Z4 and silence this region of chromatin. In pathogenic conditions, due to the decrease in number of D4Z4 repeats, the PRC2 complex is not recruited and H3K27me3 mark is decreased. Instead DBE-T, a transcript derived from a proximal region of FSHD, recruits *in cis* a Trithorax group protein, the transcriptional activator Ash1L to FSHD promoter. This results in activation of the FSHD locus, leading to the production FSHD locus genes such as *ANT1, FRG1, FRG2*[120].

Linc-RAM, a nuclear linc-RNA Activator of Myogenesis, is upregulated during myogenesis. *In vitro* knock down of linc-RAM inhibits differentiation and *in vivo* deletion impairs muscle regeneration. Linc-RAM binds to MyoD and facilitates the building of the complex of MyoD-Baf60c-Brg1, which promotes the transcriptional activity of MyoD. Interestingly, linc-RAM encodes a pro-myogenic micropeptide named MLN, which has its own pro-differentiation function[121] , but it is clear that it works in myogenesis as an lncRNA[122].

Myogenesis-associated lnc-RNA, lnc-mg, is a skeletal muscle specific lncRNA that is induced during myogenesis. Both *in vitro* and *in vivo* studies showed its relevance for

myogenesis. Knock down of lnc-mg in myogenic stem cells impaired their differentiation. Mice lacking lnc-mg have lower muscle mass, while mice overexpressing lnc-mg show muscle hypertrophy. Lnc-mg acts as a molecular sponge for miR-125, inhibiting the latter's repression of insulin-like growth factor 2 (Igf2), a positive regulator of myogenesis[123].

Lnc-MyoD, a conserved promyogenic lncRNA is directly induced by MyoD during muscle differentiation. It is localized to the nucleus. Lnc-MyoD binds to IMP2, which is a pro-proliferative factor. IMP2 binds to mRNAs of *Myc, Igf1r, Igf2, Nras, Rhla* and enhances their stability and eventually their translation. When lnc-MyoD binds to IMP2, it titrates IMP2 from the mRNAs of these proliferation factors mRNAs, which are destabilized, resulting in a decrease in the encoded proteins[124]. Thus lnc-MyoD promotes differentiation.

Sirt-1AS lncRNA is a pro-proliferative lncRNA that is transcribed from the antisense strand of the *SIRT1* locus. Sirt1 is a NAD-dependent class III protein deacetylases and is suggested to promote muscle cells proliferation by inhibiting the expression of cell cycle inhibitor p21 WAF/CIP1. Sirt-1 AS overexpression also induces cell proliferation and expression of cyclins. Sirt-1 AS interacts with 3'UTR of *Sirt1* mRNA and increases its stability by blocking the binding site for miR-34a that otherwise targets *Sirt1* mRNA for degradation. Sirt-1 AS's pro-proliferative role was also shown in *in vivo* by intraperitoneal injections of AAV expressing Sirt1 AS. In these experiments, Sirt-1 AS treated mice had increased muscle fiber size for quadriceps, soleus, and gastrocnemius[125].

Table summarizing known myogenic lncRNAs.

LncRNA name	Role in myogenesis	Mechanism of action	Localization	Coordinates	The name of that locus according to the screen described in Mueller et al.
Steroid Receptor RNA Activator (SRA)	pro-differentiation	positive regulation of MyoD function	nucleus	mm10: chr18:36,667,187-36,670,311	annotated before the study
linc-MD1	pro-differentiation	sponging miRs, primary transcript for miRs	cytoplasm	mm10:chr1:20,669,882-20,682,958	absent
Gtl2/Meg3	pro-differentiation	interaction with PRC2	nucleus	mm10: chr12:109,545,398-109,568,650	annotated before the study
H19	conflicting results	sponging miRs, primary transcript for miRs	cytoplasm	mm10: chr7:142,575,530-142,578,146	annotated before the study
Nctc1	?	?	?	mm10: chr7:142,544,609-142,558,598	annotated before the study
Yam-1	pro-proliferation	sponging miRs	cytoplasm	mm10: chr17: ??	absent (there is no hit on our list from chr17)
M1/2-sbRNAs	?	mRNA decay	cytoplasm	mm10: chr19: 5825000-5845000	--
MALAT1	pro-differentiation	?	nucleus	mm10: chr19:5,795,690-5,802,671	annotated before the study
Neat1	?	?	nucleus	mm10: (chr19:5,824,710-5,845,480)	annotated before the study
CE-RNA	pro-differentiation	activation of genes expression in cis	nucleus	mm9: chr7: 53609000- 53,610,000 (approximate)	--
DRR-RNA/MUNC	pro-differentiation	activation of genes expression in cis, and in trans	nucleus/cytoplasm	mm10: chr7:46371403-46372492	MT_19022-19023
Dum	pro-differentiation	recruitment of histone methylases	nucleus	mm10: chr16: 5kb region around 50726000 (approximate)	MT_6896 (?)
DBE-T	disease marker	activation of genes expression in cis	nucleus	hg38: chr4:190,064,502-190,067,864	lncRNA was described in human cells
linc-RAM	pro-differentiation	induction of MyoD activity	nucleus	mm10: chr10:70,204,664-70,219,711 (approximate)	annotated before the study
lnc-mg	pro-differentiation	sponging miRs	cytoplasm	mm10: chr11:67,224,196-67,237,281	MT_21873; MT_21890; MT_21919; MT_22298 (?)
Lnc-MyoD	pro-differentiation	tittrating pro-proliferative players	nucleus	mm10: chr7:46,373,500 (approximate)	MT_18580 (?)
Sirt1-AS lncRNA	pro-proliferation	increasing stability of Sirt1	cytoplasm	mm10: chr10:63,319,005-63,339,035 (approximate - coordinate for Sirt1 protein-coding gene)	annotated before the study

Role in myogenesis and mechanism of action is not described for all cases of reported lncRNAs.

Similarly, not all studies fully describe location and length of studied transcript. We cannot be confident whether reported lncRNA is a part of transcript, which we found in our initial screening for myogenic transcripts (screening described in CHAPTER II).

Rationale of the study

The main goal of the study presented in this thesis was to build a comprehensive library of lncRNAs induced during myogenesis and important for this process, using C2C12 cell line as an *in vitro* model. Specifically, we concentrated on noncoding transcripts which are induced during differentiation and focused on a few that were pro-myogenic. We are interested in the molecular mechanisms of action of these molecules and their specific protein/RNA/DNA interactors. Ideally, this knowledge will allow us to better understand the regulation of myogenesis. A subsequent analysis of the levels of these lncRNAs in skeletal muscle from muscular dystrophy patients may enable the use of these transcripts as molecular markers of myogenesis and its pathology.

After the initial screening of putative lncRNAs induced during differentiation, we focused on one lncRNA, MUNC. Using *in vitro* and *in vivo* models of differentiation we confirmed its significance in myogenesis (CHAPTER II). Finally, we described MyoD-independent functions of MUNC and identified the genes that are regulated by MUNC (CHAPTER III).

References

1. Romero, N. B., Mezmezian, M. & Fidziańska, A. Main steps of skeletal muscle development in the human. Morphological analysis and ultrastructural characteristics of developing human muscle. *Handb. Clin. Neurol.* **113,** 1299–1310 (2013).

2. Bazgir, B., Fathi, R., Valojerdi, M. R., Mozdziak, P. & Asgari, A. Satellite cells contribution to exercise mediated muscle hypertrophy and repair. *Cell J.* **18,** 473–

484 (2016).

3. Muntoni, F., Torelli, S. & Ferlini, A. Dystrophin and mutations: One gene, several proteins, multiple phenotypes. *Lancet Neurology* **2,** 731–740 (2003).

4. Domingos, J., Sarkozy, A., Scoto, M. & Muntoni, F. Dystrophinopathies and Limb-Girdle Muscular Dystrophies. *Neuropediatrics* (2017). doi:10.1055/s-0037-1601860

5. Mah, J. K. *et al.* A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy. *Neuromuscul. Disord.* **24,** 482–491 (2014).

6. Ferlini, A. *et al.* 204th ENMC International Workshop on Biomarkers in Duchenne Muscular Dystrophy 24-26 January 2014, Naarden, The Netherlands. *Neuromuscul. Disord.* **25,** 184–198 (2015).

7. Kotelnikova, E., Shkrob, M. A., Pyatnitskiy, M. A., Ferlini, A. & Daraselia, N. Novel approach to meta-analysis of microarray datasets reveals muscle remodeling-related drug targets and biomarkers in Duchenne muscular dystrophy. *PLoS Comput. Biol.* **8,** (2012).

8. Vignier, N. *et al.* Distinctive Serum miRNA Profile in Mouse Models of Striated Muscular Pathologies. *PLoS One* **8,** (2013).

9. Egas-Bejar, D. & Huh, W. W. Rhabdomyosarcoma in adolescent and young adult patients: current perspectives. *Adolesc. Health. Med. Ther.* **5,** 115–25 (2014).

10. Ognjanovic, S., Linabery, A. M., Charbonneau, B. & Ross, J. A. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-

2005. *Cancer* **115,** 4218–4226 (2009).

11. El Demellawy, D., McGowan-Jordan, J., de Nanassy, J., Chernetsova, E. & Nasr, A. Update on molecular findings in rhabdomyosarcoma. *Pathology* **49,** 238–246 (2017).

12. Lim, J.-Y. Therapeutic potential of eccentric exercises for age-related muscle atrophy. *Integr. Med. Res.* **5,** 176–181 (2016).

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

14. Jostes, B., Walther, C. & Gruss, P. The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system. *Mech. Dev.* **33,** 27–37 (1990).

15. Almalki, S. G. & Agrawal, D. K. Key transcription factors in the differentiation of mesenchymal stem cells. *Differentiation* 1–11 (2016). doi:10.1016/j.diff.2016.02.005

16. Nabeshima, Y. *et al.* Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364,** 532–5 (1993).

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

18. Kuang, S., Kuroda, K., Le Grand, F. & Rudnicki, M. A. Asymmetric Self-Renewal and Commitment of Satellite Stem Cells in Muscle. *Cell* **129,** 999–1010 (2007).

19. Le Grand, F. & Rudnicki, M. A. Skeletal muscle satellite cells and adult

myogenesis. *Current Opinion in Cell Biology* **19,** 628–633 (2007).

20. Yutzey, K. E., Rhodes, S. J. & Konieczny, S. F. Differential trans activation associated with the muscle regulatory factors MyoD1, myogenin, and MRF4. *Mol. Cell. Biol.* **10,** 3934–44 (1990).

21. Davis, R. L., Cheng, P.-F., Lassar, A. B. & Weintraub, H. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60,** 733–746 (1990).

22. Choi, J. *et al.* MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 7988–92 (1990).

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

24. Pinney, D. F., Pearson-White, S. H., Konieczny, S. F., Latham, K. E. & Emerson, C. P. Myogenic lineage determination and differentiation: Evidence for a regulatory gene pathway. *Cell* **53,** 781–793 (1988).

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

26. Rudnicki, M. A., Braun, T., Hinuma, S. & Jaenisch, R. 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,** 383–390 (1992).

27. Megeney, L. A., Kablar, B., Garrett, K., Anderson, J. E. & Rudnicki, M. A. MyoD is required for myogenic stem cell function in adult skeletal muscle. *Genes Dev.* **10,** 1173–1183 (1996).

28. Sabourin, L. A., Girgis-Gabardo, A., Scale, P., Asakura, A. & Rudnicki, M. A. Reduced differentiation potential of primary MyoD-/- myogenic cells derived from adult skeletal muscle. *J. Cell Biol.* **144,** 631–643 (1999).

29. Rudnicki, M. A. *et al.* MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75,** 1351–1359 (1993).

30. Kablar, B. *et al.* MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. *Development* **124,** 4729–4738 (1997).

31. Haldar, M., Karan, G., Tvrdik, P. & Capecchi, M. R. Two cell lineages, myf5 and myf5-independent, participate in mouse skeletal myogenesis. *Dev. Cell* **14,** 437–445 (2008).

32. Wood, W. M., Etemad, S., Yamamoto, M. & Goldhamer, D. J. MyoD-expressing progenitors are essential for skeletal myogenesis and satellite cell development. *Dev. Biol.* **384,** 114–127 (2013).

33. Conerly, M. L., Yao, Z., Zhong, J. W., Groudine, M. & Tapscott, S. J. Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. *Dev. Cell* **36,** 375–385 (2016).

34. Bergstrom, D. A. & Tapscott, S. J. Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family.

Mol. Cell. Biol. **21,** 2404–12 (2001).

35. Hirai, H., Tani, T. & Kikyo, N. Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *International Journal of Developmental Biology* **54,** 1589–1596 (2011).

36. Bergstrom, D. A. *et al.* Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. *Mol. Cell* **9,** 587–600 (2002).

37. Penn, B. H., Bergstrom, D. A., Dilworth, F. J., Bengal, E. & Tapscott, S. J. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev* **18,** 2348–2353 (2004).

38. Estrella, N. L. *et al.* MEF2 Transcription Factors Regulate Distinct Gene Programs in Mammalian Skeletal Muscle Differentiation. *J. Biol. Chem.* **290,** 1256–68 (2015).

39. Potthoff, M. J. *et al.* Regulation of Skeletal Muscle Sarcomere Integrity and Postnatal Muscle Function by Mef2c. *Mol. Cell. Biol.* **27,** 8143–8151 (2007).

40. Rampalli, S. *et al.* p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation. *Nat. Struct. Mol. Biol.* **14,** 1150–6 (2007).

41. Bengal, E. *et al.* Functional antagonism between c-Jun and MyoD proteins: A direct physical association. *Cell* **68,** 507–519 (1992).

42. Li, L., Chambard, J. C., Karin, M. & Olson, E. N. Fos and Jun repress transcriptional activation by myogeoin and MyoD. *Genes Dev* **6,** 676–689 (1992).

43. Song, A., Wang, Q., Goebl, M. G. & Harrington, M. A. Phosphorylation of nuclear

MyoD is required for its rapid degradation. *Mol. Cell. Biol.* **18,** 4994–9 (1998).

44. Kitzmann, M. *et al.* cdk1- and cdk2-mediated phosphorylation of MyoD Ser200 in growing C2 myoblasts: role in modulating MyoD half-life and myogenic activity. *Mol. Cell. Biol.* **19,** 3167–76 (1999).

45. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & Weintraub, H. The protein Id: A negative regulator of helix-loop-helix DNA binding proteins. *Cell* **61,** 49–59 (1990).

46. Puri, P. L. *et al.* Class I histone deacetylases sequentially interact with MyoD and pRb during skeletal myogenesis. *Mol. Cell* **8,** 885–897 (2001).

47. de la Serna, I. L. *et al.* MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol. Cell. Biol.* **25,** 3997–4009 (2005).

48. de la Serna, I. L., Carlson, K. A. & Imbalzano, A. N. Mammalian SWI/SNF complexes promote MyoD-mediated muscle differentiation. *Nat. Genet.* **27,** 187–190 (2001).

49. Puri, P. L. *et al.* Differential roles of p300 and PCAF acetyltransferases in muscle differentiation. *Mol. Cell* **1,** 35–45 (1997).

50. Sartorelli, V., Huang, J., Hamamori, Y. & Kedes, L. Molecular mechanisms of myogenic coactivation by p300: direct interaction with the activation domain of MyoD and with the MADS box of MEF2C. *Mol. Cell. Biol.* **17,** 1010–26 (1997).

51. Puri, P. L. *et al.* p300 is required for MyoD-dependent cell cycle arrest and muscle- specific gene transcription. *Embo J* **16,** 369–383 (1997).

52. Sartorelli, V. *et al.* Acetylation of MyoD Directed by PCAF Is Necessary for the Execution of the Muscle Program. *Mol. Cell* **4,** 725–734 (1999).

53. Polesskaya, A. *et al.* CREB-binding protein/p300 activates MyoD by acetylation. *J. Biol. Chem.* **275,** 34359–34364 (2000).

54. Liu, D., Black, B. L. & Derynck, R. TGF-?? inhibits muscle differentiation through functional repression of myogenic transcription factors by Smad3. *Genes Dev.* **15,** 2950–2966 (2001).

55. Clegg, C. H., Linkhart, T. A., Olwin, B. B. & Hauschka, S. D. Growth factor control of skeletal muscle differentiation: Commitment to terminal differentiation occurs in G1 phase and is repressed by fibroblast growth factor. *J. Cell Biol.* **105,** 949–956 (1987).

56. Wu, Z. *et al.* p38 and Extracellular Signal-Regulated Kinases Regulate the Myogenic Program at Multiple Steps. *Mol. Cell. Biol.* **20,** 3951–3964 (2000).

57. Wilson, E. M. & Rotwein, P. Control of MyoD function during initiation of muscle differentiation by an autocrine signaling pathway activated by insulin-like growth factor-II. *J. Biol. Chem.* **281,** 29962–29971 (2006).

58. Buffry, A. D., Mendes, C. C. & McGregor, A. P. The Functionality and Evolution of Eukaryotic Transcriptional Enhancers. *Adv. Genet.* **96,** 143–206 (2016).

59. Goldhamer, D. J., Faerman, A., Shani, M. & Emerson, C. P. Regulatory elements that control the lineage-specific expression of myoD. *Science (80-.).* **256,** 538–42 (1992).

60. Tapscott, S. J., Lassar, A. B. & Weintraub, H. A novel myoblast enhancer element

mediates MyoD transcription. *Mol. Cell. Biol.* **12,** 4994–5003 (1992).

61. Asakura, A., Lyons, G. E. & Tapscott, S. J. The regulationof MYOD gene expression - conserved elements mediate expression in embryonic axial muscle. *Dev. Biol.* **171,** 386–398 (1995).

62. Chen, J. C. J., Love, C. M. & Goldhamer, D. J. Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. *Dev. Dyn.* **221,** 274–288 (2001).

63. Chen, J. C. J., Ramachandran, R. & Goldhamer, D. J. Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. *Dev. Biol.* **245,** 213–223 (2002).

64. L'honore, A. *et al.* MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. *Mol. Biol. Cell* **14,** 2151–2162 (2003).

65. Shen, X. *et al.* Genome-wide examination of myoblast cell cycle withdrawal during differentiation. *Dev. Dyn.* **226,** 128–138 (2003).

66. Moran, J. L., Li, Y., Hill, A. A., Mounts, W. M. & Miller, C. P. Gene expression changes during mouse skeletal myoblast differentiation revealed by transcriptional profiling. *Physiol. Genomics* **10,** 103–11 (2002).

67. Blais, A. *et al.* An initial blueprint for myogenic differentiation. *Genes Dev.* **19,** 553–569 (2005).

68. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat.*

Biotechnol. **28,** 511–515 (2010).

69. Mousavi, K. *et al.* ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol. Cell* **51,** 606–617 (2013).

70. Cao, Y. *et al.* Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters. *EMBO J.* **25,** 502–511 (2006).

71. Cao, Y. *et al.* Genome-wide MyoD Binding in Skeletal Muscle Cells: A Potential for Broad Cellular Reprogramming. *Dev. Cell* **18,** 662–674 (2010).

72. Asp, P. *et al.* Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **108,** E149–E158 (2011).

73. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–233 (2009).

74. Czech, B. & Hannon, G. J. Small RNA sorting: matchmaking for Argonautes. *Nat. Rev. Genet.* **12,** 19–31 (2011).

75. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12,** 246–258 (2011).

76. Pink, R. C. *et al.* Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* **17,** 792–798 (2011).

77. Kawai, J. *et al.* Functional annotation of a full-length mouse cDNA collection. *Nature* **409,** 685–690 (2001).

78. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420,** 563–573 (2002).

79. Goldhamer, D. J. *et al.* Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. *Development* **121,** 637–649 (1995).

80. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489,** 101–108 (2012).

81. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10,** 155–159 (2009).

82. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).

83. Dinger, M. E. *et al.* NRED: a database of long noncoding RNA expression. *Nucleic Acids Res.* **37,** D122-126 (2009).

84. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).

85. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136,** 629–641 (2009).

86. Atkinson, S. R., Marguerat, S. & Bähler, J. Exploring long non-coding RNAs through sequencing. *Semin. Cell Dev. Biol.* **23,** 200–205 (2012).

87. Orom, U. A. & Shiekhattar, R. Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet. TIG* **27,** 433–439 (2011).

88. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472,** 120–124 (2011).

89. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin

architecture and transcription. *Nature* **494,** 497–501 (2013).

90. Conley, A. B., Miller, W. J. & Jordan, I. K. Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet. TIG* **24,** 53–56 (2008).

91. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* **36,** 233–278 (2002).

92. Rinn, J. L. *et al.* Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129,** 1311–1323 (2007).

93. Tsai, M.-C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329,** 689–693 (2010).

94. Mariner, P. D. *et al.* Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* **29,** 499–509 (2008).

95. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40,** 939–953 (2010).

96. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse H19 gene. *Nature* **351,** 153–155 (1991).

97. Keniry, A. *et al.* The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat. Cell Biol.* **14,** 659–665 (2012).

98. Kanduri, C. Long noncoding RNAs: Lessons from genomic imprinting. *Biochim. Biophys. Acta* **1859,** 102–111 (2015).

99. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18,** 1433–1445 (2008).

100. Kretz, M. *et al.* Suppression of progenitor differentiation requires the long noncoding RNA ANCR. *Genes Dev.* **26,** 338–343 (2012).

101. Pang, K. C., Frith, M. C. & Mattick, J. S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet. TIG* **22,** 1–5 (2006).

102. Torarinsson, E., Sawera, M., Havgaard, J. H., Fredholm, M. & Gorodkin, J. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res.* **16,** 885–889 (2006).

103. Caretti, G. *et al.* The RNA Helicases p68/p72 and the Noncoding RNA SRA Are Coregulators of MyoD and Skeletal Muscle Differentiation. *Dev. Cell* **11,** 547–560 (2006).

104. Hubé, F., Velasco, G., Rollin, J., Furling, D. & Francastel, C. Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res.* **39,** 513–525 (2011).

105. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147,** 358–369 (2011).

106. Legnini, I., Morlando, M., Mangiavacchi, A., Fatica, A. & Bozzoni, I. A Feedforward Regulatory Loop between HuR and the Long Noncoding RNA linc-MD1 Controls Early Phases of Myogenesis. *Mol. Cell* **53,** 506–514 (2014).

107. Zhou, Y. *et al.* Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. *Development* **137,** 2643–2652 (2010).

108. Mondal, T. *et al.* MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nat. Commun.* **6,** 7743 (2015).

109. Kallen, A. N. *et al.* The Imprinted H19 LncRNA Antagonizes Let-7 MicroRNAs. *Mol. Cell* **52,** 101–112 (2013).

110. Dey, B. K., Pfeifer, K. & Dutta, A. The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and regeneration. *Genes Dev.* **28,** 491–501 (2014).

111. Eun, B., Sampley, M. L., Good, A. L., Gebert, C. M. & Pfeifer, K. Promoter cross-talk via a shared enhancer explains paternally biased expression of Nctc1 at the Igf2/H19/Nctc1 imprinted locus. *Nucleic Acids Res.* **41,** 817–826 (2013).

112. Lu, L. *et al.* Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. *EMBO J.* **32,** 2575–88 (2013).

113. Wang, J., Gong, C. & Maquat, L. E. Control of myogenesis by rodent SINE-containing lncRNAs. *Genes Dev.* **27,** 793–804 (2013).

114. Zhang, R., Xia, L. Q., Lu, W. W., Zhang, J. & Zhu, J. S. LncRNAs and cancer. *Oncol. Lett.* **12,** 1233–1239 (2016).

115. Watts, R., Johnsen, V. L., Shearer, J. & Hittel, D. S. Myostatin-induced inhibition of the long noncoding RNA Malat1 is associated with decreased myogenesis. *Am. J. Physiol. Cell Physiol.* **304,** C995-1001 (2013).

116. Sunwoo, H. *et al.* Men ??/?? nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19,** 347–359 (2009).

117. Mueller, A. C. *et al.* MUNC, a Long Noncoding RNA That Facilitates the Function of MyoD in Skeletal Myogenesis. *Mol. Cell. Biol.* **35,** 498–513 (2015).

118. Du, J., Chen, T., Zou, X., Xiong, B. & Lu, G. Dppa2 knockdown-induced differentiation and repressed proliferation of mouse embryonic stem cells. *J. Biochem.* **147,** 265–271 (2010).

119. Wang, L. *et al.* LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. *Cell Res.* **25,** 335–50 (2015).

120. Cabianca, D. S. *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in fshd muscular dystrophy. *Cell* **149,** 819–831 (2012).

121. Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160,** 595–606 (2015).

122. Yu, X. *et al.* Long non-coding RNA Linc-RAM enhances myogenic differentiation by interacting with MyoD. *Nat. Commun.* **8,** 14016 (2017).

123. Zhu, M. *et al.* Lnc-mg is a long non-coding RNA that promotes myogenesis. *Nat. Commun.* **8,** 14718 (2017).

124. Gong, C. *et al.* A Long Non-coding RNA, LncMyoD, Regulates Skeletal Muscle Differentiation by Blocking IMP2-Mediated mRNA Translation. *Dev. Cell* **34,** 181–191 (2015).

125. Wang, G.-Q. *et al.* Sirt1 AS lncRNA interacts with its mRNA to inhibit muscle formation by attenuating function of miR-34a. *Sci. Rep.* **6,** 21865 (2016).

CHAPTER II

Adapted from: MUNC, a Long Noncoding RNA That Facilitates the Function of MyoD in Skeletal Myogenesis

Adam C. Mueller, Magdalena A. Cichewicz, Bijan K. Dey, Ryan Layer, Brian J. Reon, Jeffrey R. Gagan, Anindya Dutta

Molecular and Cellular Biology, February 2015 Volume 35 Number 3

I am co-first-author on this paper. My contribution to the study:

- Together with Adam Mueller I performed qRT-PCR analysis comparing expression of initial transcript candidates in myoblasts and myotubes (our runs were combined and averaged by Adam) – Fig. 2A
- I took part in PCR-based characterization of 5'end and 3'end of MUNC transcript sequence – Fig. 3 D, E
- Together with Adam Mueller I performed siRNA experiment in C2C12 cells showing that MUNC knock down inhibits myogenesis (I performed 1 biological run with each of the siRNAs, our runs were combined and averaged by Adam) – Fig. 6A-D
- I established murine primary myoblasts cultures: $DRR^{+/-}$ and $DRR^{-/-}$ and analyzed their myogenic markers transcripts levels during differentiation - Fig. 6I-M
- I performed si-MYOD knock down experiment confirming that MUNC expression is MyoD-dependent - Fig. 8 D-F

- I established C2C12 cells stably overexpressing various MUNC isoforms, and measured their efficiency of differentiation – Fig. 9
- Together with Adam Mueller I performed si-human MUNC experiment showing that human MUNC is important for myogenesis in human cells (I combined Adam's and my biological runs and averaged them) - Fig. 10H, I
- I contributed to the first draft of the manuscript and assisted in all the revisions.

MUNC: A lncRNA that induces the expression of pro-myogenic genes in skeletal myogenesis

Adam C. Mueller[a], Magdalena A. Cichewicz[a], Bijan K. Dey, Ryan Layer, Brian J. Reon, Jeffrey R. Gagan, , Anindya Dutta#

Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Jordan Hall 1232, Charlottesville, Virginia 22908

#Corresponding Author: ad8q@virginia.edu

a) Both authors contributed equally to this work.

Abstract

An *in-silico* screen for myogenic long-noncoding RNAs (lncRNAs) revealed nine lncRNAs that are upregulated more than ten-fold in myotubes versus myoblasts. One of these lncRNAs, "MyoD Upstream Non-Coding" (MUNC, also known as DRR^{eRNA}), is encoded 5 kb upstream of the transcription-start-site of *MyoD*, a myogenic transcription factor. MUNC is specifically expressed in skeletal muscle, exists as an unspliced and spliced isoform and its 5' end overlaps with the *cis*-acting Distal Regulatory Region (DRR) of *MyoD*. siRNA of MUNC reduced myoblast differentiation and specifically reduced the association of MyoD to the DRR enhancer and Myogenin promoter, but not to another MyoD-dependent enhancer. Stable overexpression of MUNC from a heterologous promoter increased endogenous *MyoD*, *Myogenin* and *Myh3* (MHC) mRNA, but not the cognate proteins, suggesting that MUNC can act *in trans* to promote gene expression, but that this activity does not require an induction of MyoD protein. MUNC also stimulates the transcription of other genes that are not recognized as MyoD-inducible genes. Knockdown of MUNC *in vivo* impaired murine muscle regeneration, implicating MUNC in primary satellite cell differentiation in the animal. We also discovered a human MUNC, that is induced during differentiation of myoblasts and whose knockdown decreases differentiation, suggesting an evolutionarily conserved role of MUNC lncRNA in myogenesis. Although MUNC overlaps with the DRR enhancer, our results suggest that MUNC is not a classic *cis*-acting e-RNA acting exclusively by stimulating the neighboring *MyoD* gene, but more like a pro-myogenic lncRNA that acts directly or indirectly on multiple promoters to increase myogenic gene expression.

Introduction

Long-noncoding RNAs (lncRNAs) are rapidly becoming recognized as important regulators of gene expression in development and disease (1, 2) . Our genomes undergo widespread transcription (3), and lncRNA genes are in comparable abundance to protein-coding genes (4). While lncRNAs are not as well conserved between species as protein coding genes (5), there have been a number of examples of lncRNAs that confer marked cellular and developmental phenotypes when their expression is altered by experiment or disease (6–9). Recently, a large class of lncRNAs named e-RNAs has been described that are produced from known DNA enhancer elements, and activate transcription of neighboring genes by facilitating enhancer-promoter contacts and/or recruiting core transcription factors such as the mediator complex (reviewed in (10) and (11)).

Skeletal myogenesis is an ordered process where the activation of MyoD in Pax3/Pax7/MyoD expressing skeletal myoblasts results in a cascade of gene expression changes that leads to terminal differentiation into multinucleated myotubes and myofibers (12, 13). The key transcriptional players in skeletal myogenesis are well known, but the mechanisms of their activation are not fully understood. MicroRNAs play a significant role in myogenesis at many key points (14). In addition, it is becoming clear that lncRNAs are involved in the regulation of myogenesis. LncRNAs identified as being upregulated during muscle differentiation often overlap MyoD-binding sites across the genome and are transcribed in a MyoD-dependent manner (15). These lncRNAs are enriched in the enhancer regions of MyoD target genes and may play a role in facilitating expression of a neighboring *cis*-located gene, as reported for other e-RNAs that facilitate chromatin looping and recruitment of Mediator to transcriptional start sites (8).

Categorizing lncRNAs into classes with common roles is not yet possible by sequence analysis. With the advent of widely available public databases of genome-wide histone modification, transcription factor-binding and RNA expression, and the development of lncRNA prediction models (5), it is now possible to predict lncRNAs in a variety of developmental contexts in-silico. Using some of these available data sets, we discovered a number of lncRNAs whose expression is activated during skeletal myogenesis and are expressed in mature skeletal muscle. In this report we focus on one lncRNA located upstream to *MyoD* in the genome, and demonstrate that it acts *in trans* to facilitate MyoD's role in initiating the myogenic cascade *in vitro* in mouse and human myoblasts, as well as *in vivo* during skeletal muscle regeneration in response to injury.

Results

LncRNAs induced during myogenesis. We intersected publicly available data showing where the mouse myoblast and myotube genomes are enriched in Histone H3K4 trimethylation (a mark of actively transcribed chromatin), RNA polymerase II (16) and RNA transcripts (17), to identify transcripts that are enriched in myotubes over myoblasts (**Fig. 1**). From this set we removed all sequences containing known exons and any sequences within 2 kb from known transcriptional start sites (TSS), to predict 29 new potential myotube-specific genes, which could be lncRNAs (**Table 1**). qRT-PCR of RNA from C2C12 myoblasts in growth medium (GM) and from myotubes obtained after 4 days in differentiation medium (DM) confirmed that twenty of these transcripts were induced >2-fold and nine were induced >10-fold in myotubes relative to myoblasts (**Fig. 2A**). Several of these RNAs were located in the genome proximal to protein coding genes of interest in skeletal muscle differentiation. We focused on the two most up-regulated RNAs,

#9 and 13 and on two that were also up-regulated >10 fold and encoded very close to each other, #2 and #3. Trans-differentiation of 10T1/2 fibroblasts by MyoD overexpression was accompanied by the induction of these RNAs (**Fig. 2B**: #2 and #3 show the same pattern, data not shown). The same RNAs were upregulated on the first day of C2C12 differentiation, approaching peak expression between days 2 and 3 of differentiation (**Fig. 2C**) and they were present at a very high level specifically in skeletal muscle (**Fig. 2D**).

***MUNC* lncRNA.** RNAs #2 and #3 were of particular interest because they were located upstream of the TSS of *MyoD*, a master regulator of skeletal muscle differentiation (12), with #2 overlapping with a previously known *MyoD* enhancer element, known as the Distal Regulatory Region (DRR) (**Fig. 2E**). 5' Cap capture RACE PCR and sequencing of the products confirmed #2 and #3 to be part of the same spliced transcript with two exons each corresponding roughly to transcripts #2 and #3 (**Fig. 3B** and **3F**). 3' RACE PCR using oligo dT priming was unable to amplify a product at this locus, suggesting the RNA was not polyadenylated. RT-PCR with a single forward directed primer (1F) and 3' primers located downstream of this locus suggested that the 3' end of the transcript did not extend beyond primer R-B, at the 3' end of RNA #3 (**Fig. 3C**). When we anchored the 3' primer at R-B and used a series of 5' primers, RT-PCR failed to give a product with primers upstream from the indicated TSS (**Fig. 3D**). Interestingly, the 1F and 2R primers revealed that there were two isoforms of the RNA (**Fig. 3C**, lane cDNA, 1) and sequencing the two products showed that the smaller product was from a 518 b spliced RNA composed of the two exons separated by a 577 b intron, while the larger product was a 1095 b genomic length, unspliced RNA (**Fig. 3F**). Primers were designed to distinguish the spliced RNA (primers 1F-2R) and unspliced RNA (primers 1F-1R or 2F-2R) (**Fig. 2E**) and indeed, RT-

PCR with primers 1F-2R selectively detected the spliced RNA (**Fig. 3E**). Q-RT-PCR with the three sets of primers shows that both spliced and unspliced isoforms are induced in primary myotubes relative to myoblasts (**Fig. 2F**).

The Coding Potential Calculator (18) tool predicted the unspliced and spliced transcripts to be non-coding RNAs with low coding potential and low evolutionary conservation in any of their short open reading frames (**Fig. 4A**). Polysome fractionation **(Fig. 4B)** found very little of the spliced or unspliced transcript in the polyribosome containing fraction compared to mRNA for proteins MyoD or GAPDH (**Fig. 4C**). Thus, these transcripts are non-coding and we have named them unspliced and spliced MUNC, for "MyoD Upstream Non-Coding". While this work was in progress, Vittorio Sartorelli and co-workers independently discovered the unspliced MUNC as DRR^{eRNA} (19).

Distribution of MUNC in tissues and cells. lncRNAs MUNC, #9 and #13 levels were high in mouse skeletal muscle, but MUNC had the greatest specificity of expression for skeletal muscle (**Fig. 2D**, **Fig. 5A-B**).

Knockdown of MUNC decreases muscle differentiation. Two independent siRNA duplexes targeting MUNC suppressed the upregulation of spliced and unspliced MUNC over the course of differentiation (**Fig. 6A, B**). MUNC knockdown significantly repressed the mRNAs of myogenic markers upregulated during differentiation. The mRNA for *Myogenin*, an early marker of muscle differentiation and a transcription factor essential for muscle differentiation, and for *Myh3* (Myosin Heavy Chain or MHC) a marker of terminal differentiation, were both decreased by nearly 10X on days 1, 3 and 5 of differentiation (**Fig. 6C, D**). *MyoD* mRNA was repressed 2X on days 1 and 3 and not repressed on day 5

of differentiation (**Fig. 6E**). MHC and Myogenin proteins were significantly repressed upon MUNC knockdown (**Fig. 6H**). MyoD protein was reduced at most two-fold on days 1 and 3 of differentiation, with no repression on day 5 (**Fig. 6H**). Immunofluorescence for MHC showed fewer MHC-positive multinucleated myotubes when MUNC was decreased (**Fig. 6G)**, and the fusion index was repressed by at least 5-fold (**Fig. 6F**). To confirm whether MUNC was equally important for differentiation of primary myoblasts we isolated primary myoblasts from mice which lack MUNC locus (DRR-/-) (20), and compared their differentiation ability (after 6 days of differentiation) *in vitro* with primary myoblasts which were heterozygotes for MUNC locus (DRR+/-). Cells lacking MUNC **(Fig. 6I, 6J)** showed 2X decrease of MyoD RNA level in both growth medium and differentiation medium **(Fig. 6K).** During differentiation Myogenin RNA level was decreased 2X, and Myh3 RNA level was diminished 3X, showing that primary myoblasts lacking MUNC are impaired in differentiation.

A global analysis of gene expression in C2C12 cells during differentiation was carried out by hybridization of cDNAs to Affymetrix microarrays. Unsupervised hierarchical clustering of the gene-expression patterns show two groups of genes, I and II, that are repressed or induced during differentiation, respectively (**Fig. 7A,** compare siControl GM vs. DM3). Gene Ontology enrichment analysis shows that the repressed genes are enriched in genes related to cell proliferation and the induced genes are enriched in genes related to muscle function **(Fig. 7B).** Knockdown of MUNC clearly attenuated the extent of repression of class I genes related to cell proliferation or induction of class II genes related to muscle function (**Fig. 7A**). The vast majority of genes are correlated in expression between cells in GM and DM, but the deviation from a perfect correlation

coefficient of 1 is a measure of the extent of gene expression change during differentiation (**Fig. 7C**). Knockdown of MUNC decreased this deviation from a perfect correlation coefficient from 0.065 to 0.025. Thus there is less change in gene expression when differentiation is induced after MUNC depletion. Consistent with this, genes that are most upregulated or downregulated during differentiation in siControl transfected cells show a significant attenuation in the fold change in the cells where MUNC is knocked down (**Fig. 7D, E**). Thus the global gene expression patterns confirm that MUNC is required for the changes in gene expression that accompany differentiation, probably not because MUNC acts directly on all these genes, but indirectly be affecting a few key factors during initial steps of myogenesis.

We compared in **Fig. 7F** the fold change in expression of genes affected by knockdown of MUNC (Left quadrants: genes repressed by MUNC and so induced by siMUNC; Right quadrants: genes induced by MUNC) with genes that are known to be induced by MyoD (21) (Top quadrants: genes induced by MyoD; Bottom quadrants: genes repressed by MyoD). There was a statistically very significant positive correlation between the effects of MyoD and MUNC, suggesting that MUNC and MyoD stimulate similar pro-myogenic genes. However, there were notable exceptions. In **Fig. 7F** and the Table in **Fig. 7G** we highlight in groups I-III a few genes up-regulated by MyoD that were not similarly regulated by MUNC and one gene repressed by MyoD but induced or not affected by MUNC.

Knockdown of MUNC decreases MyoD binding to the DRR but not as much to the CER. To gain insight into the mechanism by which MUNC facilitates muscle differentiation, we examined by ChIP the binding of MyoD and Myogenin to the following

regulatory sites known to bind the two transcription factors: the Core Enhancer Region at -20 kb relative to *MyoD* TSS (CER), the DRR (at -5 kb relative to *MyoD* TSS), and to the sites in the -0.5 kb region of the *Myogenin Promoter* (**Fig. 8**). These results are expressed as the amount of DNA (relative to input) detected in each ChIP (**Fig. 8A, B, C**). Under control siRNA conditions, after 3 days in differentiation medium, all 3 loci were enriched in the MyoD ChIP over IgG ChIP. With MUNC knockdown however, MyoD binding was decreased to <5% at the DRR, to 40% at the *Myogenin* Promoter and to 60% at the CER (**Fig. 8A-C**). This variability between sites could indicate that knockdown of MUNC does not inhibit MyoD binding to all sites simply by repressing MyoD protein levels. A less likely possibility is that the DRR has such a low affinity for MyoD (compared to the other sites) that MyoD binding to the DRR is more severely affected by the 50% decrease of MyoD protein seen after MUNC knockdown. We consider the latter explanation unlikely given that in control cells >1% of the input DRR locus is precipitated with MyoD, while a much smaller proportion of the two other sites (0.05-0.12%) are precipitated with MyoD, suggesting that the DRR is preferentially occupied by MyoD in control cells.

Myogenin is known to facilitate chromatin remodeling in a MyoD dependent manner (22), and binds to the DRR, CER and its own promoter during myoblast differentiation (23). Knockdown of MUNC decreased the binding of Myogenin at the DRR to 60%, **(Fig. 8A)**, without any effect at the *Myogenin* Promoter or at the CER (**Fig.8B-C**). The milder impairment of Myogenin binding to the DRR compared to MyoD binding suggests that there is some specificity of the action of MUNC on MyoD binding to the DRR that is unlikely to be explained by MUNC expression simply opening the local chromatin at the DRR to make it more accessible to all transcription factors.

MUNC expression is dependent on MyoD presence. MyoD ChIP analysis showed that MUNC is required for MyoD binding to the DRR, which overlaps with the 5' end of MUNC locus. To test whether MUNC induction during differentiation is itself dependent on MyoD, we checked MUNC expression level after MyoD knock down. MyoD depletion in differentiating cells downregulated expression of both unspliced and spliced MUNC **(Fig. 8D-F)**. Given that MUNC is required for MyoD expression and binding to DRR, these results suggest a positive feedback loop between MUNC and MyoD, where both myogenic factors promote the expression of the other.

Overexpression of MUNC stimulates *MyoD, Myogenin* and *Myh3* mRNA levels but not the proteins. To assay whether expression of MUNC from a heterologous site activates *MyoD in trans*, we examined the effect of stable overexpression of exogenous MUNC on C2C12 cells. Overexpression of all three forms of MUNC: MUNC WT (expressing both unspliced and spliced MUNC), Unspliceable MUNC (unspliced MUNC with point mutation in the splice donor sequence (AG→AT), interfering with splicing), and Spliced MUNC. The levels of MUNC, *MyoD, Myogenin* and *Myh3* were measured in GM **(Fig.9 A-E)** and after 3 days in DM **(Fig. 9F-J).**

In GM overexpression of WT unspliced MUNC (10^5 fold increase, **Fig. 9A**) caused 20X increase of *MyoD* RNA **(Fig. 9C)**, and 15X increase of *Myogenin* RNA **(Fig. 9D)**. There was a mild induction of *Myh3* RNA compared to control cells **(Fig. 9E)**. Overexpression of WT unspliced MUNC also induced expression of spliced MUNC (10^4 fold, **Fig. 9B**) but we cannot distinguish whether the spliced MUNC was from the exogenously derived transcript, or from an endogenous MUNC induced by the exogenous unspliced form. In DM3 overexpression of MUNC WT led to a 1000X induction of both unspliced and spliced

isoforms of MUNC (**Fig. 9F, G**). This was accompanied by a 10X increase of MyoD, 40X increase of Myogenin and 200X increase of Myh3 transcripts **(Fig. 9H-J)**. The lower fold induction of MyoD and Myogenin in DM3 is, of course, explained by the induction of the transcripts in the control cells simply by their transfer to differentiation medium

Overexpression of the unspliceable form of MUNC was not very efficient, with only a 10X increase of unspliced MUNC, and no increase of spliced MUNC in GM **(Fig. 9A-B)**. Despite this, there was a 50X induction of MyoD **(Fig. 9C)** but no induction of Myogenin or Myh3 transcripts **(Fig. 9D-E)**. In DM Unspliceable MUNC did not show a marked induction of MUNC RNA **(Fig. 9F)** but still led to significant induction of MyoD RNA compared to the Control **(Fig. 9H)**. Thus the induction of MyoD RNA does not need supra-physiological levels of MUNC.

Overexpression of Spliced MUNC did not increase unspliced MUNC by much, but had approximately the same effects on the RNAs of MyoD, Myogenin and Myh3 as WT MUNC in both GM and DM (**Fig. 9C-9J**).

Although we show results with single pairs of primers, to ensure that the full length MyoD, MyoG and Myh3 transcripts were induced, we used primers distributed all along the lengths of the RNAs and obtained the same results. Collectively, these results lead us to suggest that spliced and unspliced MUNC are both stimulators of MyoD RNA, while Myogenin and Myh3 RNA could be induced more by spliced MUNC than unspliceable MUNC. However, the caveat is that the unspliceable MUNC was not overexpressed sufficiently, perhaps accounting for the failure to stimulate Myogenin and Myh3 RNAs.

A very interesting result was obtained when we examined the MyoD, Mygenin and Myh3 proteins in the cells overexpressing WT or spliced MUNC. Although there was robust induction of the three RNAs after overexpression of MUNC, the levels of the three proteins were not induced in GM **(Fig. 9K)** or in DM **(Fig. 9L)**!

This is a very exciting result on two accounts. First, MUNC overexpression dissociates the induction of MyoD, Myogenin and Myh3 transcripts from the general differentiation program, perhaps because of the absence of parallel signals that emerge during normal differentiation to stimulate the translation or stability of these proteins. Second, the induction of Myogenin and Myh3 RNAs by MUNC, without inducing the MyoD protein, definitively suggests that MUNC's transcription stimulatory function is not secondary to the induction of MyoD protein.

mMUNC is required for skeletal muscle regeneration *in vivo*. Finally, we wanted to determine whether MUNC had a physiologic role *in vivo* during skeletal myogenesis. An early surprising finding in the field of skeletal myogenesis was that *MyoD* is dispensable for embryonic muscle development with *MyoD* knockout mice developing skeletal muscle normally, and reaching adulthood with minimal defects. This is explained by MyoD's close homolog, Myf5, compensating for the loss of MyoD during embryonic development (24, 25). However, the *MyoD*$^{-/-}$ mice are impaired in skeletal muscle regeneration following injury, indicating a critical role of MyoD in skeletal muscle satellite cells that have to proliferate and differentiate for successful regeneration of adult skeletal muscle (26). This role of MyoD is mirrored by the requirement of the DRR DNA locus for MyoD expression in adult satellite cells: the DRR is not required for MyoD expression during embryogenesis, but must be intact for MyoD to be expressed in adult skeletal myoblasts (27).

To examine whether MUNC (which initiates in the DRR locus) is important for adult skeletal muscle regeneration, we knocked down MUNC in the Tibialis Anterior (TA) muscle by injection of siRNA to MUNC in an emulsion of Invivofectamine (Invitrogen). The 1F-1R and 2F-2R pairs of primers measure unspliced MUNC (**Fig. 2E**) while the 1F-2R pair of primers measures spliced MUNC (**Fig. 3E**). The steady state level of unspliced MUNC (1F1R, 2F2R) and spliced MUNC (1F2R) are decreased in adult mouse skeletal muscle after 5 days of siMUNC injection compared to siControl injection (**Fig. 10A**). This was accompanied by a significant decrease in the expression of *MyoD* and *Myogenin* and *Myh3* mRNAs (**Fig. 10B**), suggesting a role of MUNC in maintaining expression of these RNAs in adult skeletal muscle. Note that although Myogenin protein is not seen in adult muscle, the mRNA is normally still detectable at low levels (29).

We next injured the TA muscle with cardiotoxin and followed its regeneration by measuring the appearance of *MyoD*, *Myogenin* and *Myh3* RNAs on day 14 after injury. All three RNAs are normally induced during regeneration (27, 28) . MUNC was depleted during the two weeks of skeletal muscle regeneration by siRNA injection on days 3 and 5 following cardiotoxin injury (**Fig. 10C**). There was a reduction in *MyoD*, *Myogenin*, and *Myh3* mRNAs in the regenerated muscle after MUNC knockdown (**Fig. 10D**). There was a significant reduction in myofiber diameter and increase in inflammatory infiltrates in the regenerating muscle after MUNC knockdown on day 14 after injury. (**Fig. 10E, F**). This suggests that MUNC has a role in facilitating myogenic gene expression and skeletal muscle regeneration in adult skeletal muscle *in vivo*.

MUNC is conserved in humans. LncRNAs are not well conserved in sequence compared to protein-coding RNAs, stimulating us to experimentally determine whether MUNC has

a role in human cells as well. Alignment of human DRR sequence with mouse MUNC revealed striking conservation of sequence with the first exon of MUNC after allowing for a 36 base insertion in human DRR (**Fig. 11**). The sequence conservation extended upstream from the MUNC Transcription Start site (TSS), but as shown in **Fig. 3C**, we did not see any evidence of transcription in mouse myoblasts upstream of the TSS. The conservation also extended to the 5' region of the intron retained in unspliced mouse MUNC. The region in human DRR that matched mouse MUNC exon 1 is transcribed in human LHCN myoblasts and the transcript (human MUNC) is significantly up-regulated during differentiation (**Fig. 10G**). siRNA mediated knockdown of the human MUNC transcript in LHCN cells was not very efficient (**Fig. 10H**), but despite this, the expression of *MyoD*, *Myogenin* and *Myh3* mRNAs were repressed upon differentiation (**Fig. 10I**). Thus, human myoblasts contain human MUNC RNA similar to mouse MUNC and with similar pro-myogenic function.

Discussion

We identified several RNAs whose expression is upregulated during myogenesis. Within this set we discovered a noncoding RNA MUNC produced from a previously characterized enhancer upstream from the *MyoD* gene, required for *MyoD* expression in adult skeletal muscle satellite cells(20, 26). We fully characterized its transcriptional isoforms: unspliced and spliced and showed that both forms can promote muscle differentiation. Previous studies found that this enhancer element facilitates muscle specific transcription (30). Our findings suggest that this is likely due to the fact that this enhancer encodes a lncRNA which facilitates MyoD binding to the DRR and to a lesser

extent to the *Myogenin* promoter, thereby promoting the expression of MyoD targets: *Myogenin* as well as both *MyoD* and MUNC itself.

There has been much recent focus on the rapidly expanding roles of lncRNAs in mammalian genomes. Noncoding RNAs have been implicated in regulation of gene expression by facilitating gene and chromosome silencing through the recruitment of PRC2 complexes to chromatin (31). Numerous examples have been discovered suggesting lncRNAs play a common role in gene silencing in higher eukaryotes (32–34) (1, 21, 32). There have also been discoveries of lncRNAs activating gene expression (11). Most enhancer-RNAs (e-RNAs) facilitate expression of a gene neighboring the enhancer *in cis*, by recruiting the transcriptional protein Mediator to chromatin (8). Some e-RNAs can function by facilitating looping of chromatin, bringing regulatory elements on the same chromosome into proximity of each other. MUNC differs from a classic e-RNA because (a) it is required for MyoD binding to chromatin at the Myogenin locus (which is not *in cis* with MUNC or *MyoD*) and (b) it stimulates the expression of *MyoD* and *Myogenin* even when expressed from a heterologous locus *in trans*.

Sartorelli and co-workers recently reported ncRNA products of the MyoD DRR and CER loci, finding them to facilitate myogenesis (19). In contrast to our study, knockdown of MUNC (which they call DRR^{eRNA}) in their paper had no effect on MyoD expression, but like our study, there was a repression of Myogenin. We find that MUNC is required for MyoD localization to the DRR of the *MyoD* promoter and to the *Myogenin* promoter, while they report that it is required for directing MyoD only to the *Myogenin* promoter. Additionally, we notice that there is differential requirement of MUNC for MyoD binding to different genomic targets, Core Enhancer Region binding is not affected

so much by siMUNC. In addition we saw only slight effect in binding of Myogenin to the DRR and no effect in binding to Myogenin Promoter or CER. The differential effect of MUNC on MyoD and Myogenin binding to the DRR suggests that MUNC is unlikely to act at this locus by opening up the chromatin to make it more accessible to all transcription factors. In the future we plan to do MyoD ChIP-Seq, MyoG ChIP-Seq after knockdown of MUNC in differentiating C2C12 cells and compare with published data sets (16) (19) to find more binding sites for MyoD and MyoG that are affected by MUNC.

In addition, we find that MUNC expressed from a heterologous artificial locus *in trans* stimulates both *MyoD* and *Myogenin* promoters, and not *Myogenin* alone. Despite these small differences, we agree that such *trans* stimulation of transcription by MUNC from a heterologous locus eliminates models where transcription stimulation involves specific 3-D interactions bringing the MUNC locus in proximity to a neighboring target promoter. We also suggest that an important function of MUNC is to stimulate the auto-activation of the *MyoD* promoter by facilitating the binding of MyoD to the DRR.

We provide the first evidence of MUNC function *in vivo*. Depletion of MUNC in regenerating mouse skeletal muscle *in vivo* impaired expression of myogenic markers and impeded regeneration, suggesting its function is important for repair of damage to skeletal muscle in adults. Here, too, we saw a decrease in MyoD expression when MUNC was depleted.

One mechanism by which e-RNAs stimulate genes *in cis* is that the act of transcription of the e-RNA promotes the local opening of chromatin *in cis*. Although MUNC induces *MyoD* transcript, its role in myogenesis does not appear to be limited to

MyoD RNA induction alone. MUNC knockdown reduced MyoD binding to target sites in the *MyoD* and *Myogenin* promoters. The reduction in MyoD binding at the DRR was much greater than the reduction in expression of MyoD itself, and was not uniform at all MyoD binding sites, suggesting that MUNC facilitates MyoD binding to specific target sites, including to *MyoD*'s auto-stimulatory DRR enhancer. Most strikingly, overexpression of MUNC stimulated *Myogenin* and *Myh3* transcripts without inducing MyoD protein. Thus, while MUNC is required for *MyoD* transcription, increasing the level of MyoD protein is not necessary for MUNC to stimulate the transcription of *Myogenin* or *Myh3*, clearly suggesting that the pro-myogenic function of MUNC is not simply on account of its role as an e-RNA that induces *MyoD*. The global analysis of gene expression changes after MUNC knockdown reinforces this message by highlighting that (a) not all MUNC regulated genes are necessarily MyoD dependent, and (b) not all MyoD induced genes require MUNC.

Many uncertainties remain about which e-RNAs are conserved across species. Sequence conservation of lncRNAs is often quite poor between species. Not enough RNA secondary structures have been solved, so that it is also difficult to identify lncRNAs conserved by structure alone. Hence we are excited to show that a MUNC-like transcript is preserved in humans and that it may be equally important for differentiation. Future experiments will test whether human MUNC can complement the loss of mouse MUNC. The overexpression of unspliceable MUNC hints that spliced and unspliced MUNC may have different target specificities, but this conclusion needs to be corroborated by future experiments where we carefully regulate the levels of expression of spliced and unspliceable MUNC so that they are expressed at approximately equal levels.

Our results are in broad agreement with that of Sartorelli and co-workers (19) and we conclude by summarizing the new findings in this report. (1) MUNC knockdown downregulates MyoD transcript and MyoD protein. (2) MUNC is required for expression of muscle differentiation marker genes in primary myoblasts. (3) An unexpected spliced isoform of MUNC suggests that MUNC is an RNA Polymerase II driven transcript, and that not all the transcribed MUNC is simply left on chromatin like an e-RNA. (4) The differential requirement of MUNC for MyoD binding to different E-boxes suggests a function in addition to regulating MyoD expression. (5) The differential requirement of MUNC for binding of MyoD and Myogenin to the DRR site, suggests that the role of MUNC is more specific than simply opening up the chromatin at the DRR to give access to all transcriptional factors. (6) MUNC is required for skeletal muscle regeneration in adult mice. (7) A human homolog of MUNC suggests evolutionary conservation of MUNC function. (8) A positive feedback loop between MyoD and MUNC suggests how the two could be involved in turning on a switch towards differentiation. (9) Induction of *MyoD*, *Myogenin* or *Myh3* transcripts by overexpressed MUNC is not sufficient to induce the corresponding proteins, thus dissociating the transcriptional induction program from other aspects of differentiation. (10) The induction of *Myogenin* and *Myh3* RNA by MUNC does not require an induction of MyoD protein, suggesting that MUNC has actions beyond simply acting as an e-RNA that induces *MyoD*. (11) The global analysis of gene expression changes following MUNC knockdown shows that although many MUNC-induced genes are also induced by MyoD, there are clear examples of genes induced by MyoD alone or MUNC alone.

Materials and Methods

Cell culture

Under growth conditions C2C12 cells were cultured in DMEM media with 10% FBS, when differentiating, media was switched to DMEM containing 2%FBS. C3H 10T1/2 Cells were grown in Eagle's Basal medium with 2-mM L-glutamine, 1.5g/L sodium bicarbonate and Earle's BSS, with 10% FBS for growth conditions and 2% FBS for differentiation conditions. LHCN cells were cultured on gelatin-coated plates, in DMEM/M199 media (4:1, vol/vol), supplemented with 15% FBS, zinc sulfate (0.03µg/ml), vit.B12 (1.4µg/ml), dexamethasone (0.055µg/ml), HGF (2.5ng/ml), bFGF (10ng/ml), HEPES (0.02M). Differentiation was performed in serum-free DMEM/M199 media (4:1, vol/vol), supplemented with HEPES (0.02M), zinc sulfate (0.03µg/ml), vit.B12 (1.4µg/ml), insulin (10 µg/ml).

lncRNA screen

Myoblast sites were defined as the regions identified by both PollII Chip-seq (GSM721286), RNA-seq (GSM521256), H3K4me3 Chip-seq (GSM721292) experiments in myoblast cells. Myotube sites were defined by similiar experiments in myotube cells (GSM721287, GSM521258, GSM721293). Both sites were filtered for previously-known transcripts, including: protein-coding and non-protein-coding genes, known miRNAs, and transcription start sites. All of these filters were retrieved for the mm9 assembly on 2012-06-12 from the UCSC Genome browser. Finally, myoblast sites were filtered for any common to myotubes, and myotube sites were filtered for any sites common to myoblasts (**Fig. 1**).

siRNA transfection of cells

Cells were transfected with Life Technologies "Silencer Select" siRNAs targeting MUNC, of target sequence ggaugagcugugugcuucu or cgaccaaugggagagagca, or commercial negative control silencer select siRNAs. The "Silencer Select" RNAs have proprietary modifications that allow them to efficiently target nuclear RNAs. siRNA was transfected at a final concentration of 30nM in growth medium, with 3uL/mL total media volume of Lipofectamine RNAimax. siRNA and Lipofectamine was mixed in 2mL/10mL total media volume Optimem for 25 minutes prior to addition to cells seeded onto plates at 25% confluency.

Stable overexpression of MUNC in C2C12 cells

Using amplified sequences of unspliced MUNC (PCR with C2C12 genomic DNA as a template) insert was cloned into pLPCX vector via in fusion method (Clontech). The construct was linearized and introduced to the cells (XtremeGENE transfection reagent Roche). After 24hr, pools of stably transfected cells were selected with puromycin.

RNA expression by qRT-PCR

RNA was isolated by Trizol extraction, and cDNA synthesis was performed using Life Technologies Superscript III RT cDNA synthesis kits, using random hexamer priming. Prior to cDNA synthesis, RNA samples were treated with RQ1 RNAse free DNAse to eliminate potential DNA contamination of samples. qPCR was performed with BioRad iCycler, using iQ SYBR Green Supermix. Primers for the ncRNA screen were designed using BatchPrimer3. All primers used in this study are listed in **Table 2 and Table 3**.

Microarray analysis

RNA was hybridized to Affymetrix Mouse Exon ST arrays and analyzed for gene expression using Affymetrix Expression Console software and Microsoft Excel. For designing the heat map of genes expression the microarray data was analyzed using Bioconductor. The top 400 genes, which demonstrated the most variance between samples, were used for unsupervised hierarchical clustering. Gene ontology analysis of the top gene clusters was performed using DAVID Bioinformatics Resources (http://david.abcc.ncifcrf.gov/home.jsp) (35).

Western Blotting

Cells were lysed in IPH buffer and run on 10% polyacrylamide SDS-PAGE gels and transferred to nitrocellulose membranes. Membranes were blocked for 30 minutes in 5% Milk containing PBS-T, and incubated overnight with primary antibody in 3% BSA. Secondary antibody incubation was carried out for 1 hour after washing, and at a 1:4000 dilution before washing and incubation with Millipore Immobilon HRP substrate. Chemiluminescent images were captured on a G:BOX geldoc system (Syngene). Antibodies and dilutions were used as follows, MyoD C-20 1:250 (Santa Cruz), Myogenin 1:250 (Santa Cruz), MHC 1:250 (Millipore), Tubulin 1:3000(Sigma).

ChIP studies

Cells were washed in PBS and crosslinked with 1% formaldehyde for 10 minutes, and then treated with 0.15M glycine for 15 minutes. Cells were then PBS washed twice and lysed

in lysis buffer containing 1%SDS with 50mM Tris 10mM EDTA, protease and RNAse inhibitors. Chromatin was sonicated to an average size of 300bp, and incubated overnight with Protein G Dynabead/Antibody complexes, with 2ug antibody per 1.5x10^6 cells. After overnight incubation, Dynabeads were washed with RIPA containing 150 mM NaCl followed by 500mM NaCl, RIPA containing 250mM LiCl, and twice with Tris-EDTA. Beads were then de-crosslinked overnight at 65°C and treated with Proteinase K, RNAse A, RNAse H. DNA was then isolated by phenol:chloroform extraction and analysed by qPCR.

Immunofluorescence

Cells were plated on glass cover slips in the presence of 30nM siRNA. Cells were collected in growth medium or after 24, 72 or 120 hours in differentiation medium. After 60 minutes, the cover slips were fixed with 4% formaldehyde in PBS, and permeabilized in 0.5% Triton X-100 in PBS. Coverslips were blocked in 5% goat serum. Coverslips were incubated at room temperature with primary antibody for 1 hour, and Alexa 488 or 549 conjugated secondary antibody for 1 hour, with three TBS washes following each antibody incubation. Coverslips were then mounted with Vectashield mounting solution (Vector Laboratories). Antibodies used were anti-MyoD C-20 antibody (Santa Cruz Laboratories), and anti-Myosin Heavy Chain A4.1025 antibody (Millipore). Antibodies were diluted 1:200 in 5% goat serum containing PBS.

Microscopy

Images were captured using a Nikon Microphot SA upright microscope equipped with a Nikon NFX35 camera using SPOT imaging software (Diagnostic Instruments Inc.)

and a Nikon PlanApo 60x oil objective lens. Fluorescence images were acquired on the same day using the same exposure times, gamma, and gain between samples. Images were enhanced for brightness and contrast to the same extent within Photoshop.

Polyribosome fractionation and qRT-PCR

Polysome fractionation assay was performed as described (36). The total RNAs from monoribosome and polyribosome fractionations were extracted separately, and subjected to qRT-PCR analysis.

Isolation and growth of primary myoblasts

Mice were genotyped and sacrified at Day 9. The procedure was performed according to the protocol (37).

Mouse skeletal muscle regeneration following cardiotoxin injury

Cardiotoxin regeneration expreriments were performed as described in (38). Invitrogen in-Vivo silencers targeting the MUNC-1 sequence were injected on day 2 and 5 of regeneration using Invivofectamine (Invitrogen). Tissue samples from regenerating TA muscle were collected on day 14 following cardiotoxin injection and analyzed by qRT-PCR and immunofluorescence. The use of animals in all of the studies was done following protocols approved by the Animal Care and Use Committee (ACUC) of University of Virginia.

Acknowledgments

We thank Woodring E. Wright, M.D., Ph.D, UT Southwestern Medical Center for providing us LHCN-M2 cell line.

We thank David J. Goldhamer, PhD, University of Connecticut for providing us o C57Bl/6J DRR loxP/loxP mice.

The study was supported by R01 AR053948 and P01 CA104106.

References

1. **Wang KC, Chang HY**. 2011. Molecular Mechanisms of Long Noncoding RNAs. Mol. Cell.

2. **Wapinski O, Chang HY**. 2011. Long noncoding RNAs and human disease. Trends Cell Biol.

3. **Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, Ponting CP, Stadler PF, Morris K V., Morillon A, Rozowsky JS, Gerstein MB, Wahlestedt C, Hayashizaki Y, Carninci P, Gingeras TR, Mattick JS**. 2011. The reality of pervasive transcription. PLoS Biol. **9**.

4. **Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, Brown JB, Lipovich L, Gonzalez JM, Thomas M, Davis CA, Shiekhattar R, Gingeras TR, Hubbard TJ, Notredame C, Harrow J, Guigó R**. 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. **22**:1775–1789.

5. **Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A**. 2010. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. **28**:503–510.

6. **Bernard D, Prasanth K V, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F, Triller A, Spector DL, Bessis A**. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J. **29**:3082–3093.

7. **Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL, Wang Y, Brzoska P, Kong B, Li R, West RB, van de Vijver MJ, Sukumar S, Chang HY**. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature **464**:1071–1076.

8. **Lai F, Orom U a, Cesaroni M, Beringer M, Taatjes DJ, Blobel G a, Shiekhattar R**. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature **494**:497–501.

9. **Maass PG, Rump A, Schulz H, Stricker S, Schulze L, Platzer K, Aydin A, Tinschert S, Goldring MB, Luft FC, Bähring S**. 2012. A misplaced lncRNA causes brachydactyly in humans. J. Clin. Invest. **122**:3990–4002.

10. **Mousavi K, Zare H, Koulnis M, Sartorelli V**. 2014. The emerging roles of eRNAs in transcriptional regulatory networks. RNA Biol. **11**:106–10.

11. **Orom UA, Shiekhattar R**. 2013. XLong noncoding RNAs usher in a new era in the biology of enhancers. Cell.

12. **Berkes CA, Tapscott SJ**. 2005. MyoD and the transcriptional control of myogenesis. Semin. Cell Dev. Biol.

13. **Yokoyama S, Asahara H**. 2011. The myogenic transcriptional network. Cell. Mol. Life Sci.

14. **Gagan J, Dey BK, Dutta A**. 2012. MicroRNAs regulate and provide robustness to the myogenic transcriptional network. Curr. Opin. Pharmacol. **12**:383–388.

15. **Blum R, Vethantham V, Bowman C, Rudnicki M, Dynlacht BD**. 2012. Genome-wide identification of enhancers in skeletal muscle: The role of MyoD1. Genes Dev. **26**:2763–2779.

16. **Asp P, Blum R, Vethantham V, Parisi F, Micsinai M, Cheng J, Bowman C, Kluger Y, Dynlacht BD**. 2011. Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. Proc. Natl. Acad. Sci. U. S. A. **108**:E149–E158.

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

18. **Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G**. 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. **35**.

19. **Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager G, Sartorelli V**. 2013. ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. Mol. Cell **51**:606–617.

20. **Chen JCJ, Ramachandran R, Goldhamer DJ**. 2002. Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. Dev. Biol. **245**:213–223.

21. **Bergstrom DA, Penn BH, Strand A, Perry RLS, Rudnicki MA, Tapscott SJ**. 2002. Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. Mol. Cell **9**:587–600.

22. **Du C, Jin YQ, Qi JJ, Ji ZX, Li SY, An GS, Jia HT, Ni JH**. 2012. Effects of myogenin on expression of late muscle genes through myoD-Dependent chromatin remodeling ability of myogenin. Mol. Cells **34**:133–142.

23. **Johnson DS, Mortazavi A, Myers RM, Wold B**. 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science **316**:1497–1502.

24. **Rudnicki MA, Braun T, Hinuma S, Jaenisch R**. 1992. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. Cell **71**:383–390.

25. **Rudnicki MA, Schnegelsberg PN, Stead RH, Braun T, Arnold HH, Jaenisch R**. 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. Cell **75**:1351–1359.

26. **Megeney LA, Kablar B, Garrett K, Anderson JE, Rudnicki MA**. 1996. MyoD is required for myogenic stem cell function in adult skeletal muscle. Genes Dev. **10**:1173–1183.

27. **Chen JCJ, Love CM, Goldhamer DJ**. 2001. Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. Dev. Dyn. **221**:274–288.

28. **Zhao P, Hoffman EP**. 2004. Embryonic Myogenesis Pathways in Muscle Regeneration. Dev. Dyn. **229**:380–392.

29. **Yan Z, Choi S, Liu X, Zhang M, Schageman JJ, Lee SY, Hart R, Lin L, Thurmond FA, Williams RS**. 2003. Highly coordinated gene regulation in mouse skeletal muscle regeneration. J. Biol. Chem. **278**:8826–8836.

30. **Tapscott SJ, Lassar AB, Weintraub H**. 1992. A novel myoblast enhancer element mediates MyoD transcription. Mol. Cell. Biol. **12**:4994–5003.

31. **Brockdorff N**. 2013. Noncoding RNA and Polycomb recruitment. RNA **19**:429–442.

32. **Aguilo F, Zhou MM, Walsh MJ**. 2011. Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. Cancer Res.

33. **Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, Segura V, Huang L, Moreno I, Navarro A, Monzó M, García-Foncillas J, Rinn JL, Guo S, Huarte M**. 2013. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. Genome Biol. **14**:R104.

34. **Wu L, Murat P, Matak-Vinkovic D, Murrell A, Balasubramanian S**. 2013. Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. Biochemistry **52**:9519–9527.

35. **Huang DW, Sherman BT, Lempicki RA**. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. **4**:44–57.

36. **Li Y, Bor Y-C, Misawa Y, Xue Y, Rekosh D, Hammarskjöld M-L**. 2006. An intron with a constitutive transport element is retained in a Tap messenger RNA. Nature **443**:234–237.

37. **Springer ML, Rando TA, Blau HM**. 2002. Gene delivery to muscle. Curr. Protoc. Hum. Genet. **Chapter 13**:Unit13.4.

38. **Dey BK, Gagan J, Yan Z, Dutta A**. 2012. miR-26a is required for skeletal muscle differentiation and regeneration in mice. Genes Dev **26**:2180–2191.

Table 1.

Chromosome	RNAseq start coordinate	RNAseq stop coordinate	RNAseqID	NCRNA_ID
chr7	30766640	30766824	MT_18580	1
chr7	53626843	53627031	MT_19022	2
chr7	53627330	53627626	MT_19023	3
chr7	53629487	53629561	MT_19024	4
chr7	53629872	53629954	MT_19025	5
chr14	22566952	22567040	MT_141	6
chr14	32026105	32026292	MT_248	7
chr19	3765132	3765533	MT_20418	8
chr8	13201970	13202444	MT_1230	9
chr8	13202768	13202922	MT_1231	10
chr8	13203007	13203263	MT_1232	11
chr8	126419999	126420073	MT_2710	12
chr1	20612312	20612490	MT_2953	13
chr6	29381128	29381504	MT_24494	14
chr6	88850233	88850446	MT_25164	15
chr6	149263245	149263339	MT_25901	16
chr11	46206645	46206735	MT_21873	17
chr11	48687223	48687425	MT_21890	18
chr11	50026771	50027017	MT_21919	19
chr11	58952620	58952746	MT_22298	20
chr16	23989822	23989993	MT_6896	21
chr3	14530455	14530643	MT_8317	22
chr15	27958322	27958509	MT_11215	23
chr4	119962964	119963170	MT_26893	24
chr2	91789193	91789586	MT_13961	25
chr9	21891831	21892015	MT_28975	26
chr13	75846117	75846436	MT_16126	27
chr13	75846651	75847469	MT_16127	28
chr5	31912782	31912875	MT_30734	29

Table 2.

ncRNA ID	Direction of primer	Sequence
1	F	TTCGTGAGAGTATCCCACAGG
	R	TGTGAAGAGGAGATGTCCAGAA
2	F	AGCCTCAGGATGAGCTGTGT
	R	CTCAATGCAGGGCCTCTTAG
3	F	TTCCAAAAAGGAGGAAGCAA
	R	ATGGATGTGGGGTTCATCAT
4	F	CCAATGCTAAACAACCATCTGA
	R	ATCCATTTGGAGGGCACTG
5	F	CAGGACCTTTGCACATGTTT
	R	GGATGAAGGGAGACAGAAGC
6	F	TAAGGGTAAAGGCGGAGCTA
	R	TTGCAGACTCCGCTCAGTAA
7	F	TCGACATACCCTGTCTGCAA
	R	CTTCCCATCTCCCAGTGTTG
8	F	CCATGTGCAAGAACTCCAAA
	R	TGGTACCCCTTCTCCAAATG
9	F	GACCTTGACCTTTCCCCAGT
	R	TTCCAGCTCTGTGTGGTCAG
10	F	CACATGGATCCCTGGAGTG
	R	AGAGCATGCCTTCATTCTCAA
11	F	CAGCAGAGGTTGGTCCTCTT
	R	GGAGGTGGGTATGCAGTGAG
12	F	ACCATGGAGCCATTCACTTT
	R	AGCTATTTTGGGAGCGCTTA
13	F	GCTTGGTGTCCCTCAGTGAT
	R	GTGCTCTCAGCCACACAGAA
14	F	CCCGACTGGAGATCCTCATA
	R	AGTAGGGGTTTGGGCAGAGT
15	F	GACATAGGGAGGGTCCCAGT
	R	AGGTAGTGTTCCTGGCTTGC
16	F	TAGCGCCAGTCTTCTTCAGG
	R	GTTAGAGCCAGGGCCTCAAT
17	F	ATCTGACCTGCCAGGAAGC
	R	CGTCTTTTCCTGTTCTCTTCCA
18	F	CCCACAGGGACAGAGATAGG
	R	TCTCTGTGACAGCTGGAGGA
19	F	AGTCAGACCAGGCATCTTGC
	R	ACAAGCCTTTCCCTTTCCTC
20	F	GGGCACAGATGGTGAGTTG
	R	CTGGAGTGTGGGCTGCTG
21	F	AAATGTGTGTGTGGGTACGTG
	R	GGGGGAATGTTCAAGACCTTA
22	F	GGGGTTGGAACAGTGAAGAA
	R	CATTAGCTCCAGCAGGCATT
23	F	GCCTAGATGGTTGGCATTGT
	R	TGAGTGGGTAAGGCACACAG
24	F	TGTGCTTGCCCATACAACTC
	R	TTGGGACACTGTGTGGGATA

25	F	GCCACCCATCTACTTTTCCA
	R	TCAGGTGCTTTCTGTGCATC
26	F	CTGCAGGAAGTGCTGCTCTA
	R	CAAGCACAGTGGCACAAGAT
27	F	CGAAAGTGGACATGTTGTCG
	R	AATCCTGTGGGGTGTAGCTG
28	F	TCTCAGAGGCTCCCAAAGAA
	R	GGCTTCCCCTTAATCTCCAC
29	F	TGAGCTCTGGGGAGTCTCTG
	R	GGTGGGAAAGAACAGCACAG

Table 3.

Name of the primer	Direction of primer	Sequence
qGAPDH F	F	GCACAGTCAAGGCCGAGAAT
qGAPDH R	R	GCCTTCTCCATGGTGGTGAA
qMHC F	F	TCCAAACCGTCTCTGCACTGTT
qMHC R	R	AGCGTACAAAGTGTGGGTGTGT
qMYOD F	F	CATCCGCTACATCGAAGGTC
qMYOD R	R	GTGGAGATGCGCTCCACTAT
qMYOGENIN F	F	AGCGCAGGCTCAAGAAAGTGAATG
qMYOGENIN R	R	CTGTAGGCGCTCAATGTACTGGAT
q human ACTIN β F	F	GGCACCAGATCATGTTTGAG
q human ACTIN β R	R	GAGTCCATCACGATGCCAGT
q human MHC F	F	CTTCCCTGCACCAGATTCTC
q human MHC R	R	GTATAAGCCCGAGGTGGTGA
q human MYOD F	F	GGGGCTAGGTTCAGCTTTCT
q human MYOD R	R	GCTCTGGCAAAGCAACTCTT
q human MyOGENIN F	F	GCCAGACTATCCCCTTCCTC
q human MYOGENIN R	R	GAGGCCGCGTTATGATAAAA
q human DRR F	F	CTGGGCAGAGCAGCCAAGGGAGCTG
q human DRR R	R	GAGGGGCTCATTTGGTGGGGAGTGGG
CER ChIP F	F	GGGCATTTATGGGTCTTCCT
CER ChIP R	R	CTCATGCCTGGTGTTTAGGG
DRR ChIP F	F	TCAGGCCAGGACCATGTCT
DRR ChIP R	R	CTGGACCTGTGGCCTCTTAC
Myogenin Promoter ChIP F	F	GAATCACATGTAATCCACTGGA
Myogenin Promoter ChIP R	R	ACGCCAACTGCTGGGTGCCA
1F	F	AGCCTCAGGATGAGCTGTGT
1R	R	CTCAATGCAGGGCCTCTTAG
2F	F	TTCCAAAAAGGAGGAAGCAA
2R	R	ATGGATGTGGGGTTCATCAT
nested 5'RACE PRIMER outer	R	TCTCTCCCATTGGTCGGTTG
nested 5'RACE PRIMER inner (R-B')	R	GTTATTCACCGAGGGACACG
F-A	F	TAGCCAAGGGAGCTGAAATG
R-B	R	GTTATTCACCGAGGGACACG
siMUNC-1		GGAUGAGCUGUGUGCUUCUTT
siMUNC-2		CGACCAAUGGGAGAGAGCATT
si HUMAN MUNC		AGUCCUCCCCUCUCAGCUCCCtt
si-MYOD		CCAAUGCGAUUUAUCAGGUGCUUUG
pLPCX-MUNC F (in fusion cloning primer)		GCCTCGGCCAAACATCGATA
pLPCX-MUNC R (in fusion cloning primer)		GAGTCCGGTAGCGCTAGC
MUNC GT110AG F (MUNC splicing mutant cloning primer)	F	GAGCTGTGTGCTTCTCCAGAGCAGTGGGCCTACAGCCTAAG
MUNC GT110AG R (MUNC splicing mutant cloning primer)	R	CTTAGGCTGTAGGCCCACTGCTCTGGAGAAGCACACAGCTC

A.

	Excluding transcription start sites <5kb	Excluding known exons	Overlapping RNA PolII + RNA Seq peaks	Overlapping peaks from myotube specific H3K4me3 ChIP-Seq	Excluding alternate splicing variants of known coding genes
Myotubes specific RNA PolII ChIP-Seq peaks	22,239	3,361	75	42	29
Myotubes specific RNA Seq peaks	37,999	10,360			

B.

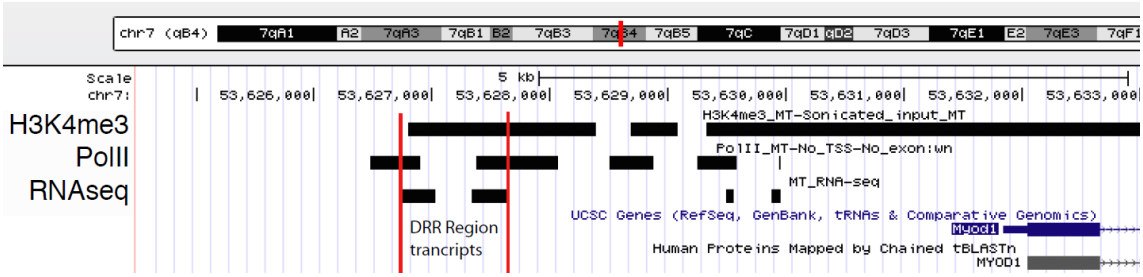


Figure 1. A. Workflow (Left to Right) of the computational screen that identified potential long noncoding RNA induced during muscle differentiation. **B.** UCSC Genome Browser screenshot showing the locations of PolII ChIP Seq, H3K4me3 ChIP-Seq and RNA-Seq signal.

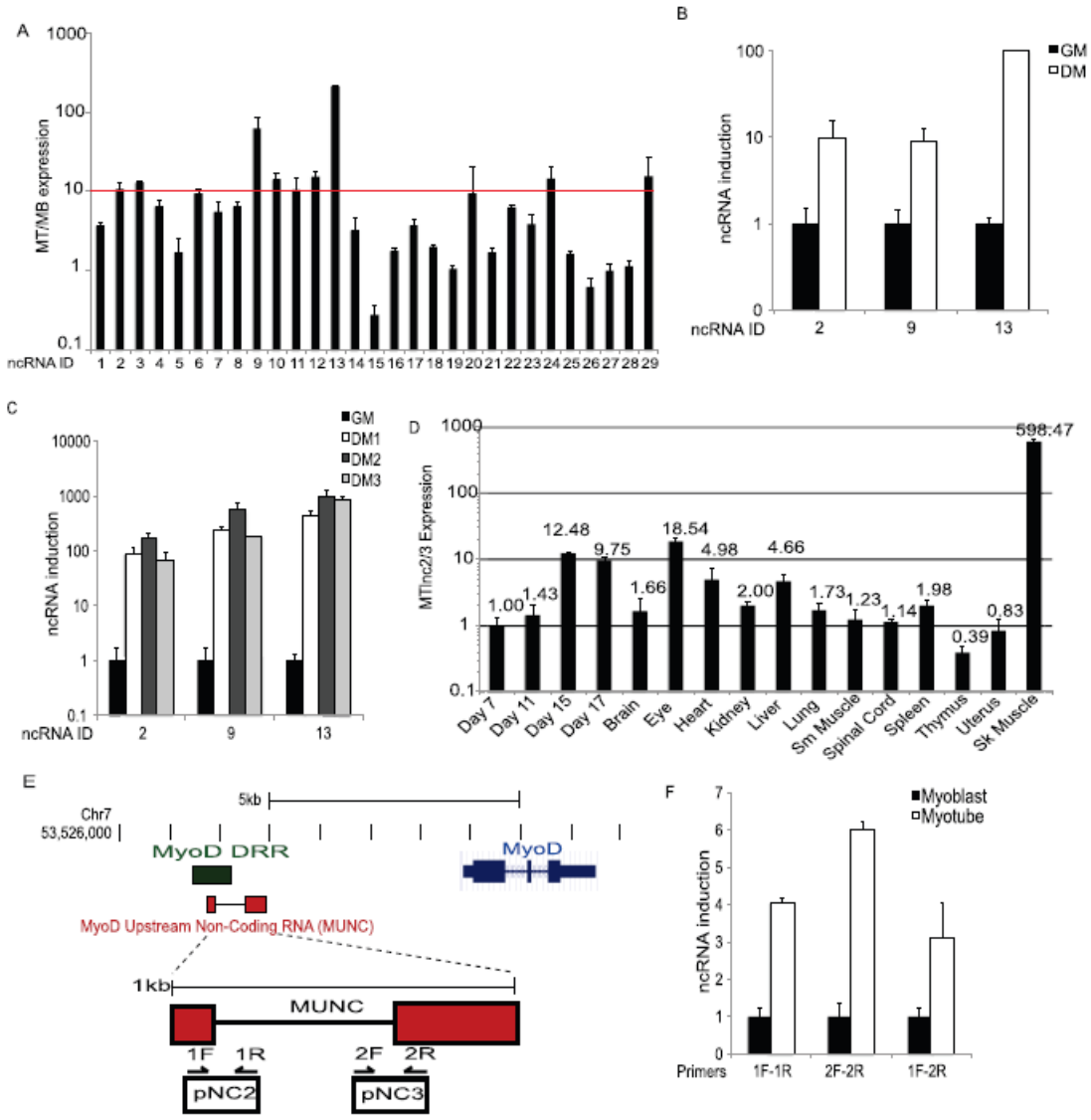


Figure 2. Predicted muscle specific non-coding RNAs are upregulated in myotubes

A. qPCR confirmation of predicted non-coding RNAs induced in myotubes versus myoblasts. 9 predicted RNAseq fragments at 7 independent genomic loci were >10X induced in myotubes versus myoblasts. **B.** qPCR analysis of MyoD transfected trans-differentiating 10T1/2 fibroblasts for several predicted myogenically regulated lncRNAs when transferred to low serum differentiation medium (DM) versus growth medium (GM). **C.** RT-qPCR of lncRNAs from screen in differentiating C2C12 Myoblasts on indicated days after switch to low serum differentiation media. **D.** RT-qPCR of indicated lncRNAs 2/3 (or MUNC) from mouse embryos

(embryonic days 7, 11, 15, 17) and murine tissues showing high expression in skeletal muscle. **E.** Schematic of MUNC genomic region upstream of the MyoD1 locus, MUNC overlaps previously characterized Distal Regulatory Region (DRR) enhancer and putative noncoding transcripts 2 and 3 (pNC2 and pNC3). **F.** RT-qPCR analysis of MUNC expression in primary myoblasts and myotubes. MUNC 1F-1R and 2F-2R primers measure the primary unspliced MUNC while 1F-2R primers measure spliced MUNC (see **Fig. 3E**). Data represent mean +/- SEM, n=3.

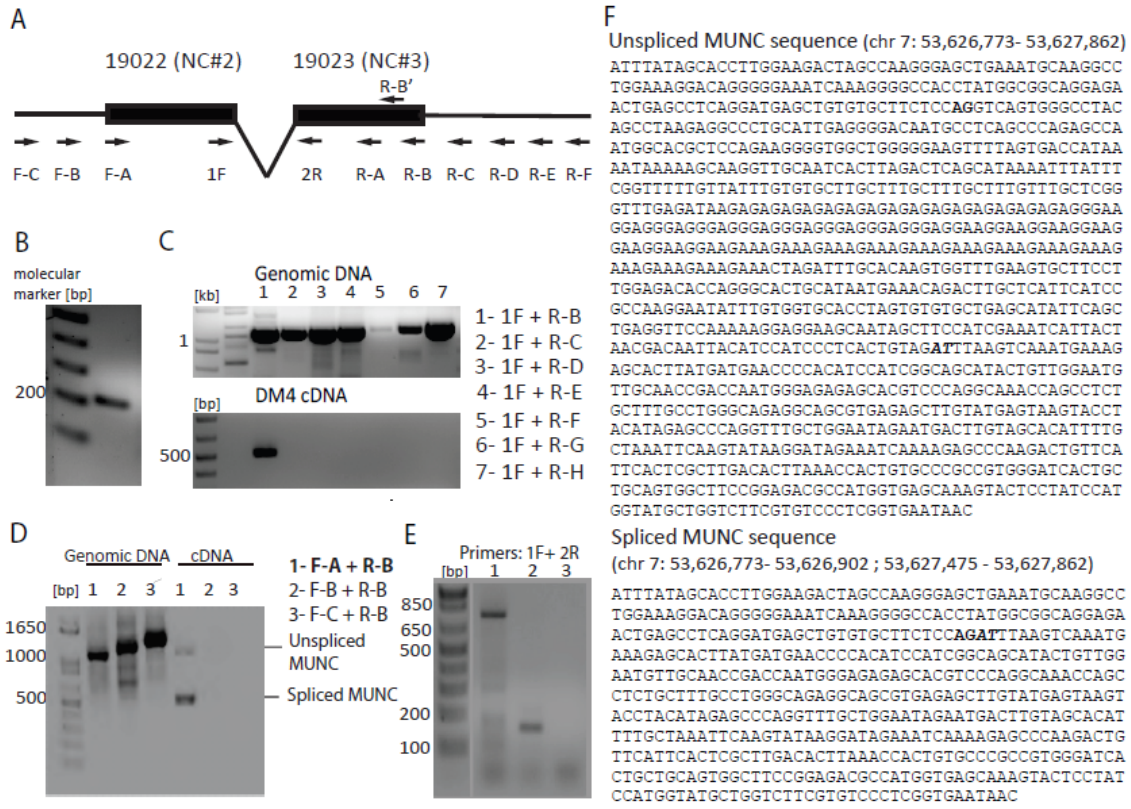


Figure 3. Characterization of MyoD Upstream NonCoding transcript.

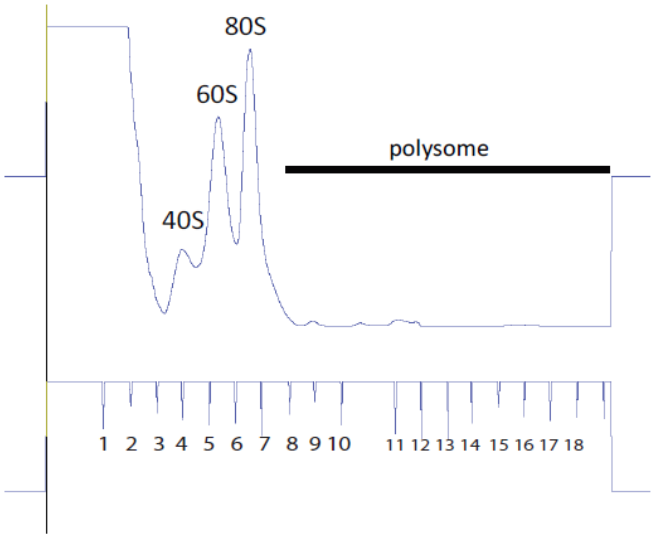
A. Schematic of primers used for 5'RACE PCR and PCR primer walking to determine ends of MUNC transcript. **B**. 5' end mapping of MUNC: PCR product generated by 5'RACE PCR on cap-captured DM4 C2C12 RNA with R-B' (nested) primer. **C.** 3' end mapping of MUNC: PCR products with 1F and indicated R primers on genomic DNA (positive control) and DM4 C2C12 cDNA. Only R-B gave a product on the cDNA putting it at the 3' end of MUNC. **D.** PCR on genomic DNA (positive control) and cDNA from DM4 C2C12 cells confirm the 5' end of MUNC and the presence of unspliced and spliced isoforms. F-A + R-B produced two products on cDNA: genomic length unspliced (~1000 bases) and spliced (~500 bases). **E.** PCR products amplified by 1F-2R primers. Lane 1: genomic DNA from C2C12 with extension time 60 sec. Lane 2: cDNA from DM3 C2C12, extension time 20 sec. Lane 3 negative control (no DNA), extension time 20 sec. **F,** Sequence of Unspliced and Spliced MUNC amplified by F-A and R-B primers, **bold**

represents 5’splice site and ***bold italic*** represents 3’ splice site. Coordinates according to the UCSC Genome Browser (Assembly July 2007).

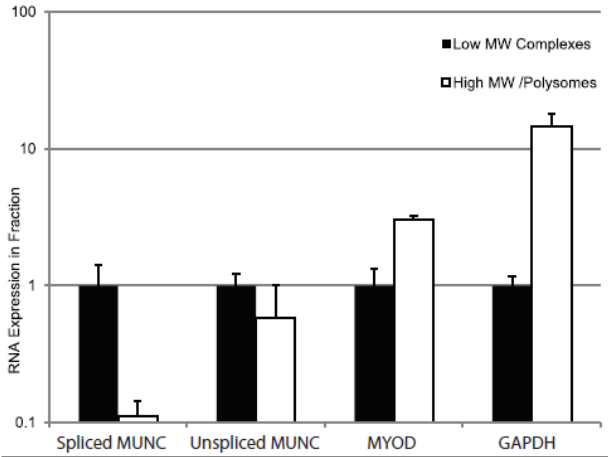
A

FORM OF MUNC		GENOMIC MUNC	SPLICED MUNC
HOMOLOGY FEATURES	HIT NUM	0	0
	HIT SCORE	0.0	0.0
	FRAME SCORE	0.0	0.0
ORF_FRAMEFINDER	COVERAGE	18.80%	23.17%
	LOG-ODDS SCORE	36.89	25.83
	TYPE	PARTIAL	PARTIAL

B



C



	Spliced MUNC	Primary MUNC	MyoD	GAPDH
Low MW Complexes	1.00	1.00	1.00	1.00
High MW /Polysomes	0.11	0.59	3.06	14.60

Figure 4. Unspliced and spliced MUNC predicted to have low coding potential and is not associated with polysomes during differentiation.

A, Analysis obtained from the Coding Potential Calculator based on evolutionary conservation and ORF attributes (http://cpc.cbi.pku.edu.cn/). Both forms of MUNC are likely to be non-coding transcripts. **B,** Polysome fractionation profile of differentiating C2C12 cells. Monosomes: Fractions 3-8. Polysomes: Fractions 9-18. **C,** qRT-PCR of polysome fractions and monosome fractions. Spliced and unspliced MUNC are depleted from the polysome fraction while the mRNAs for MyoD and GAPDH are enriched in the polysome fraction. Data represent mean +/- SEM, n=3.

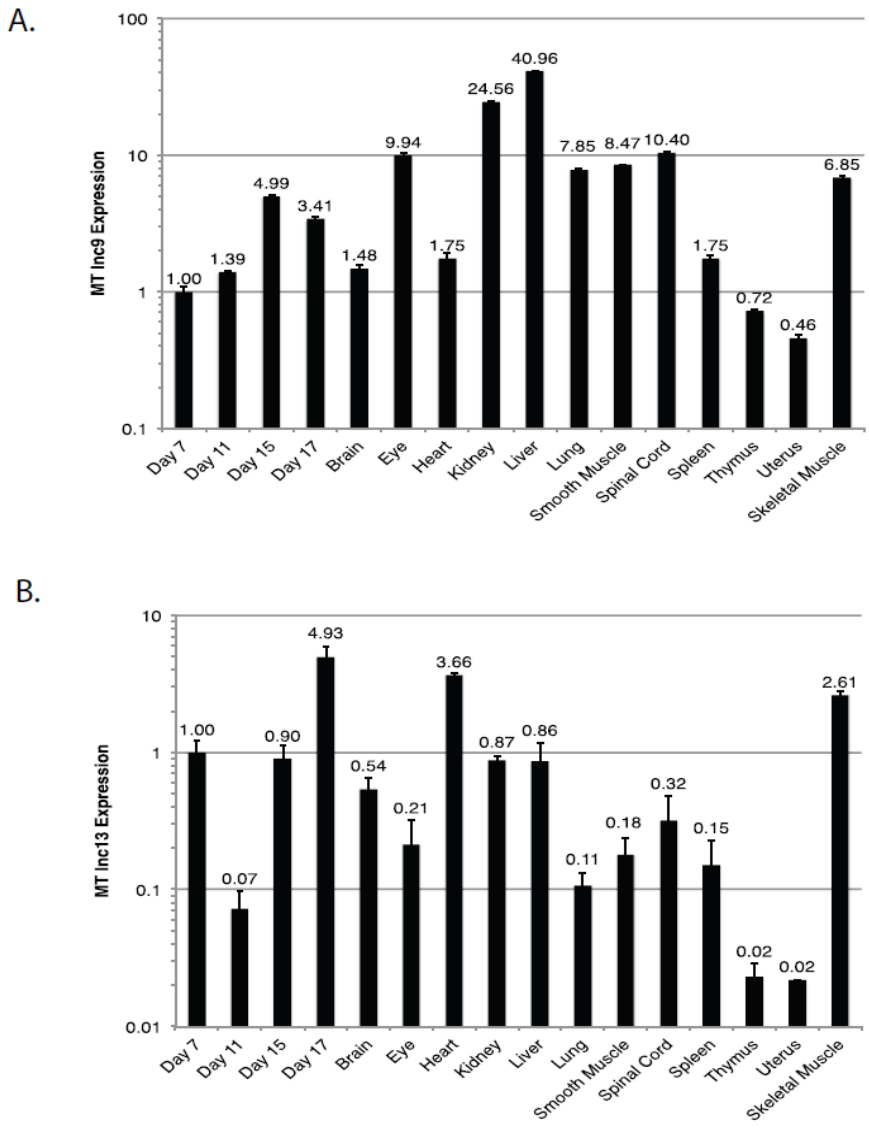


Figure 5. Tissue expression of two other myogenically upregulated long noncoding RNA transcripts, #9 and #13. A-B, qRT-PCR of indicated transcripts across a panel of embryonic and adult mouse tissue samples. Values are normalized to expression of RPS13 a housekeeping gene with low tissue variability and plotted relative to expression in Day 7 embryonic tissue. Data represents mean +/- SEM, n=3.

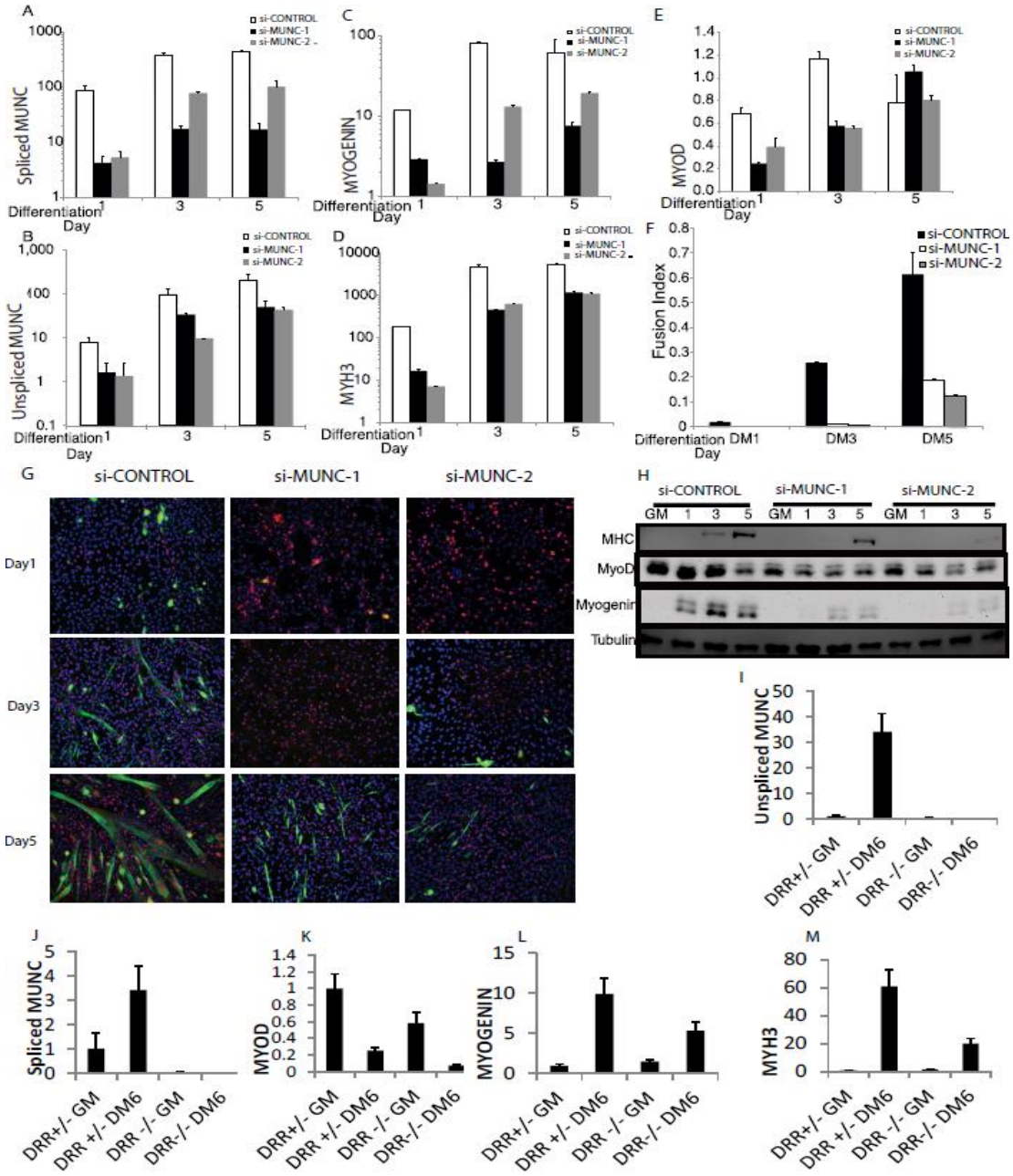
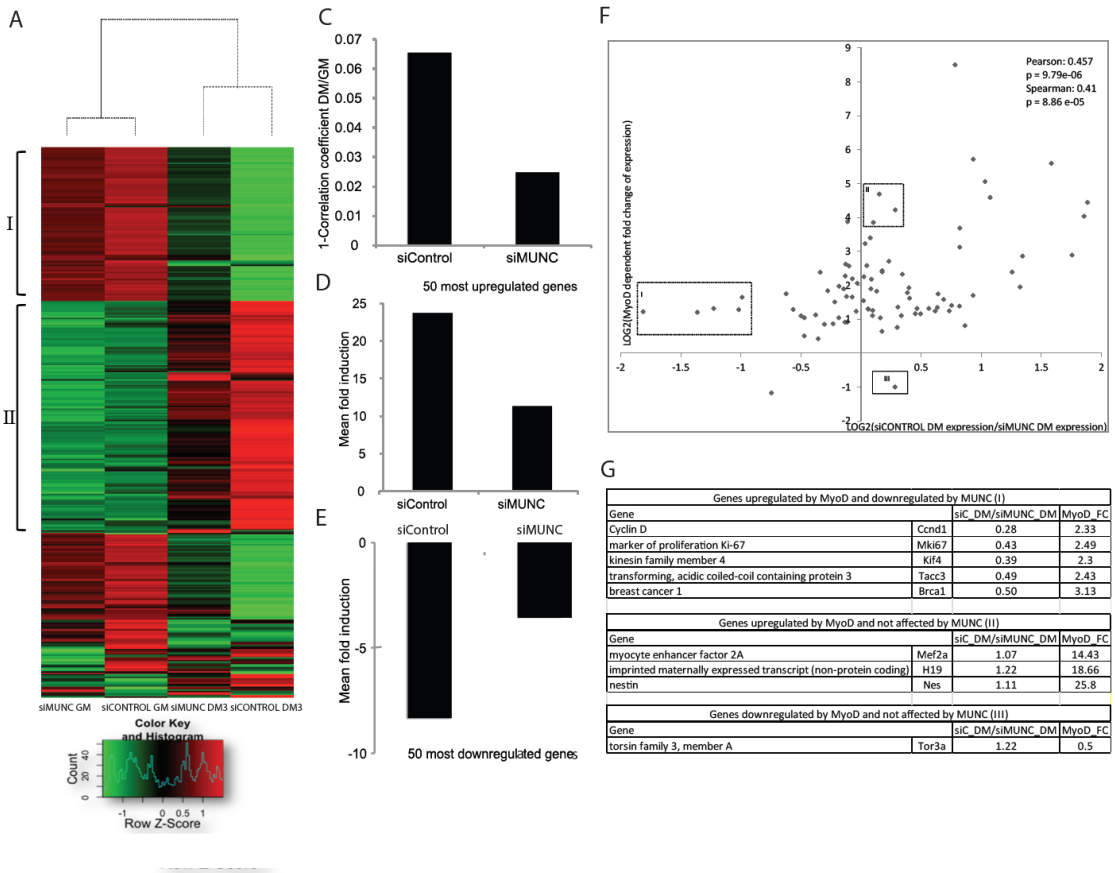


Figure 6. MUNC knockdown represses the induction of myogenic differentiation markers and impairs myotube formation in culture. **A-E.** qRT-PCR measuring induction of unspliced and spliced MUNC and myogenic markers MyoD, Myogenin and Myh3 (MHC) during differentiation of C2C12 cells incubated with either control siRNA or siRNA targeting the 5' or 3' end of MUNC. MUNC levels normalized to GAPDH. Data represent mean +/- SEM, n=3. Note log-scale of Y-

axis in A-D. **F.** Fusion index of differentiating C2C12 cells shows MUNC knockdown impairs myotube formation. Fusion index calculated by dividing number of nuclei contained within multinucleated cells by number of total nuclei in a field. **G.** Immunofluorescence of MHC (green), and MyoD (blue) in differentiating C2C12 cells. C2C12 incubated with control siRNA show much greater formation of MHC positive, multinucleated cells on differentiation day 3 and 5 than cells incubated with siRNA targeting MUNC. Data represent mean +/- SEM, n=3. **H.** Western blot analysis of MHC, Myogenin and MyoD in differentiating C2C12 cells. Independent siRNA targeting MUNC reduce expression of these myogenic proteins. **(I-J)** qRT-PCR measuring induction of unspliced and spliced form of MUNC and myogenic markers MyoD, Myogenin and Myh3 during differentiation of primary murine myoblasts derived from *DRR*$^{+/-}$ and *DRR*$^{-/-}$ mice. Expression levels normalized to GAPDH. Data represent mean +/- SEM, n=3.



G

Genes upregulated by MyoD and downregulated by MUNC (I)			
Gene		siC_DM/siMUNC_DM	MyoD_FC
Cyclin D	Ccnd1	0.28	2.33
marker of proliferation Ki-67	Mki67	0.43	2.49
kinesin family member 4	Kif4	0.39	2.3
transforming, acidic coiled-coil containing protein 3	Tacc3	0.49	2.43
breast cancer 1	Brca1	0.50	3.13

Genes upregulated by MyoD and not affected by MUNC (II)			
Gene		siC_DM/siMUNC_DM	MyoD_FC
myocyte enhancer factor 2A	Mef2a	1.07	14.43
imprinted maternally expressed transcript (non-protein coding)	H19	1.22	18.66
nestin	Nes	1.11	25.8

Genes downregulated by MyoD and not affected by MUNC (III)			
Gene		siC_DM/siMUNC_DM	MyoD_FC
torsin family 3, member A	Tor3a	1.22	0.5

B

	Biological Process	p-Value	FDR
Cluster I	Cell Cycle	4.42E-27	6.90E-24
	Cell Cycle Process	4.71E-24	7.36E-21
	Mitosis	2.86E-23	4.46E-20
	Nuclear Division	2.86E-23	4.46E-20
	Organelle Fission	8.24E-23	1.29E-19
	Cell Division	1.08E-22	1.69E-19
	Chromosome Segregation	7.63E-09	1.19E-05
	DNA Metabolic Process	2.11E-08	3.29E-05
	DNA Replication	5.71E-07	8.91E-04
	DNA Metabolic Process	2.11E-08	3.29E-05
Cluster II	Sarcomere	2.91E-09	3.25E-06
	Contractile Fiber Part	5.21E-09	5.82E-06
	Myofibril	8.19E-09	9.16E-06
	Contractile Fiber	1.16E-08	1.29E-05
	Z Disc	2.13E-06	2.40E-03
	I Band	4.36E-06	4.80E-03

Figure 7. Global gene expression changes that occur during skeletal myogenesis are inhibited by MUNC depletion. A. Hierarchical clustering of the 400 genes that varied the most upon

differentiation of C2C12 cells. The cells were either treated with MUNC or control silencer RNAs in both GM and DM. Signals are scaled to *Z*-scores of the rows. **B.** I and II, denote gene clusters that are enriched in GO terms associated with cell cycle and growth, and muscle specific processes, respectively. **C.** 1 minus correlation coefficient of gene expression profiles of GM versus DM3 cells with either siControl or siMUNC as measured by Affymetrix exon array analysis. **D-E**. Mean expression fold change of the 50 most upregulated (**D**) or downregulated (**E**) genes in control differentiating cells. This fold change is suppressed in siMUNC cells. Mean +/- SEM, N=5. **F.** Y-axis: fold change of genes known to be induced by MyoD (21) (Top: induced by MyoD; Bottom: repressed by MyoD). X-axis: Fold change of same genes after knockdown of MUNC (Left: repressed by MUNC; Right: induced by MUNC). Box I: induced by MyoD but downregulated by MUNC. Box II: induced by MyoD but not regulated by MUNC. Box III: downregulated by MyoD, but not affected by MUNC. **G.** Genes in the three sub-classes (Boxes I-III) identified in (F).

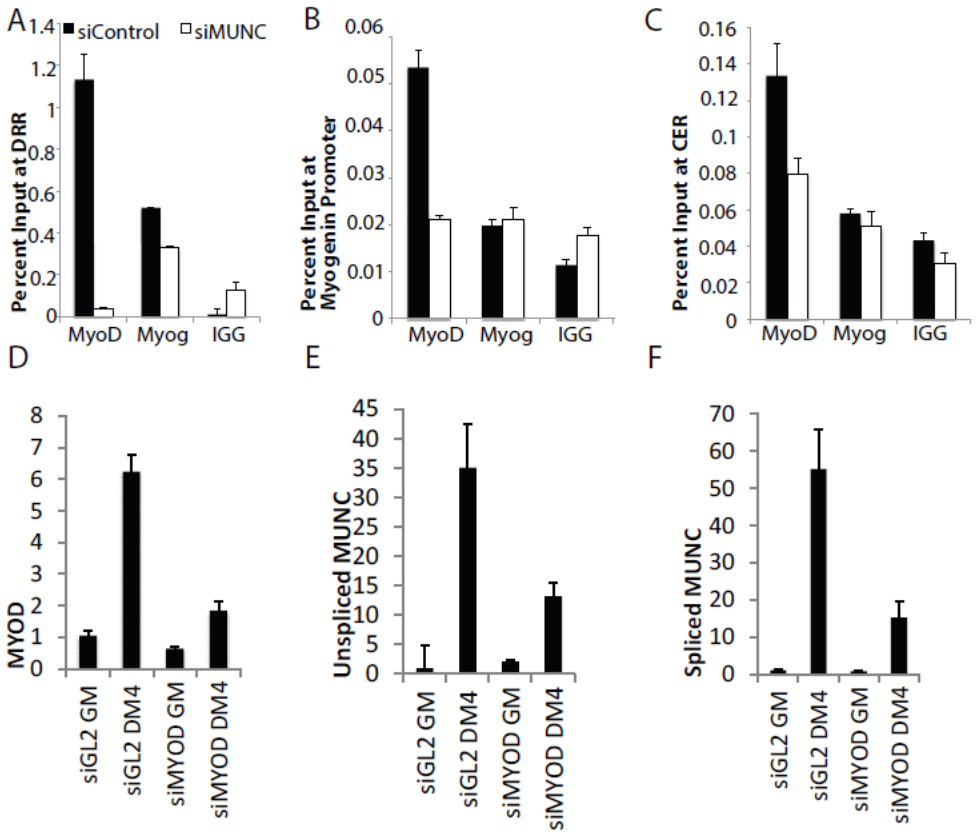


Figure 8. MUNC is required for MyoD binding at certain target sites and not others.

MUNC expression is dependent on MyoD. MyoD and Myogenin (MyoG) ChIP at the *MyoD* Distal Regulatory Region **(A),** *Myogenin* Promoter Region **(B)** and *MyoD* Core Enhancer Region **(C).** Cells were treated with either control siRNA or siMUNC and incubated in Differentiation Medium for 72 hours. Data represent mean +/- SEM, n=3. **(D-F)** MUNC expression level in proliferating C2C12 cells (GM) and in differentiating cells (Differentiation Day 4) in control conditions (siGL2) and after MyoD knockdown (siMyoD). Cells were transfected with siRNA and harvested 48hrs later (GM) or re-transfected (at DM0 and DM2) and incubated in Differentiation Medium for 4 days (DM4). Data represent mean +/- SEM, n=3.

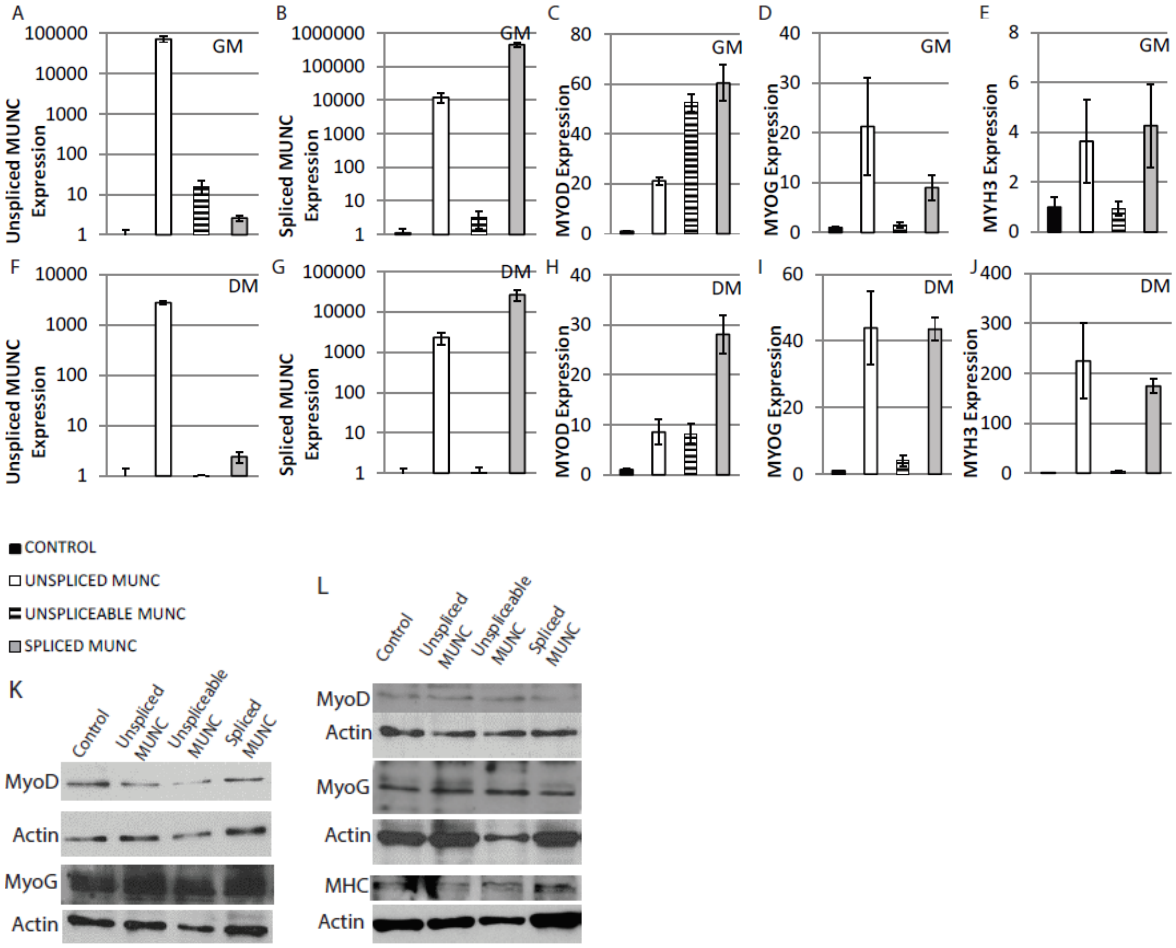


Figure 9. Stable overexpression of MUNC enhances RNA of myogenic markers, but not the corresponding proteins. qRT-PCR expression of MUNC isoforms and myogenic markers following C2C12 transfection with linearized vectors coding **1)** WT unspliced form of MUNC, **2)** Unspliceable form of MUNC with point mutation preventing RNA splicing, **3)** Spliced form of MUNC. Measurements were performed on proliferating cells (GM) and differentiating cells after 3 days in differentiation medium (DM). Expression of Unspliced MUNC **(A, F)**, Spliced MUNC **(B, G)**, MyoD **(C, H)**, Myogenin **(D, I)**, Myh3 RNAs **(E, J)**. Data normalized to GAPDH expression and then normalized again in each panel to the level in vector-transfected cells in GM or DM. Data represent mean +/- SEM,N=3. **(K)** Western blot analysis showing level of MyoD and MyoG proteins in C2C12 cells overexpressing MUNC in GM. Actin was used as a loading control. **(L)** Same as in (K) except in cells are in DM for 3 days

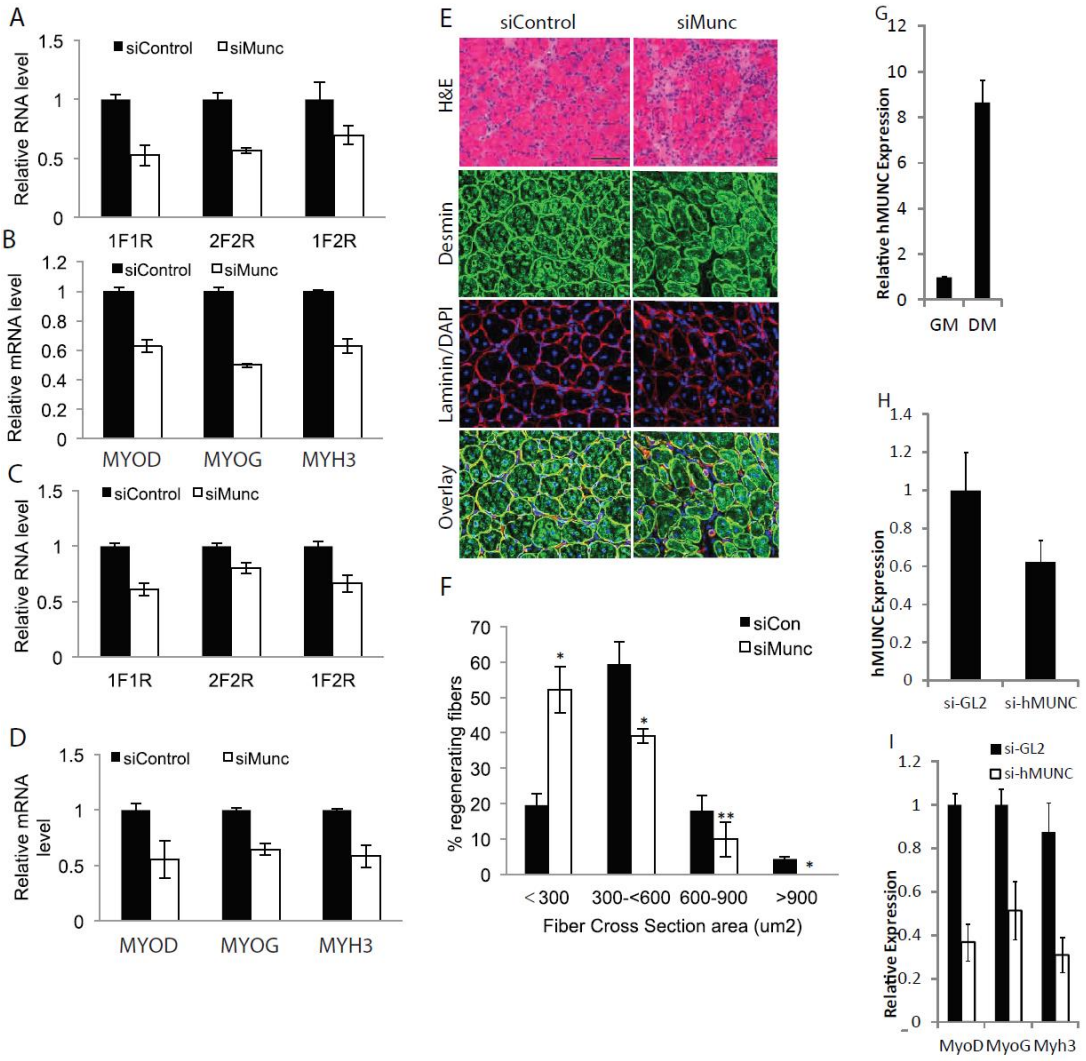


Figure 10. MUNC knockdown reduces myogenic marker expression during skeletal muscle regeneration in adult mice and impairs regeneration. MUNC is conserved between humans and mice. A-B, qRT-PCR expression of **(A)** MUNC isoforms and **(B)** myogenic markers following knockdown of MUNC in three adult mouse TA muscles. Mice were injected twice with siMUNC (Invitrogen in vivo silencer/invivofectamine complexes) and RNA was harvested 5 days following the first injection. Control siRNA was injected into the contralateral leg. 1F-1R, and 2F-2R measure the 5' and 3' regions of the unspliced MUNC while 1F-2R measure spliced MUNC. Data represent mean +/- SEM. N=3. **C-D.** qRT-PCR expression of **(C)** MUNC isoforms and **(D)** myogenic markers after knockdown in adult mice TA muscle 14 days after injury with

cardiotoxin. Mice were injected with cardiotoxin, then siMUNC (Invitrogen in vivo silencer/invivofectamine complexes) twice, on day 2 and 5 following injury. RNA was harvested 14 days following the injury. Control siRNA was injected into the contralateral leg. Data represent mean +/- SEM. N=4. **E.** Representative H&E, desmin, and laminin stained sections of regenerating mouse TA muscle 14 days after cardiotoxin injection and control or MUNC in-vivo siRNA knockdown. **F.** Quantitation of myofiber cross-sectional area. Data represents mean +/- SEM, N=4. For statistical analysis Student's T test was used. * $p < 0.05$ (significant); ** $p > 0.05$ (not significant). **G**. Induction of expression of human MUNC RNA during differentiation of LHCN cells. **H-I.** qRT-PCR analysis of (**H**) human MUNC RNA and (**I**) myogenic markers in LHCN human myoblasts after transfection with siRNA targeting human MUNC and 7 days of differentiation in low serum. Data normalized to beta-actin expression. Data represent mean +/- SEM N=3.


human GATCT-----------------ATACCCATG-TGGTGGGTGGATTTTGGCTTTCCCAGGG 42
mouse GGTCAGCCTAAATTGGCCAGATCTACACTTGGTGGCAGGTAGTTTCAGGCTTTCTGGG-A 59
human AGCAAGTTTGTCAGGGGACAGAGGGAGGCACTCAGGTTGGACCCAGGAACAGGAAGGGAA 102
mouse AGCAAAACTGGCAGAGAACAGAG---------CAGGAT---CCTTGAGTTGGGAAAGGAA 107
human AGGCTGGGGACAGAGAGGGGACCTGGAGCTGGCCCTGCCCCACCAGGCCCACTCATGCTT 162
mouse AGTCTAGGGCCAGAGACTGAACCTGGGGCTGGTCCTGTTCCACCTGTCCTCCCCGTGGTT 167
human TTACCTTCTGGCCCTTTGGCGCCCCCCACCTCCCGGCCAGATACGCAGCCTGTGTCAGCC 222
mouse TCATCCTCCAGTCCTTC--AGCCCCCTAGACCCAAGCCAGCCATGCAGCCCGCAGTAG-C 224
human CCAGTGCAGAGCCACAGGCCCAGCCTGGGCAGGGGCAGGGTGCGTGAAGACTGGGGCAGG 282
mouse AAAGTAAGAGGCCACAGGTCCAGACTGGGTAGGGCAGAGGTGCCTGAGGCTTGGGGCAGG 284
human TGCAGGCTGGATTGGGTTTCCAGAGGCTATATATATATA--AAGGCTGCCGGGAGCCCCAGG 340
mouse TGCTAGTTGGATCCGGTTTCCAGAGGCTATATATATATATAAAGGCTGCTG-TTTCCCCATG 343
human GCCGCTCCCTGAGGGCACAACACTGTGGGGGCCCAGCCAGGCCCACATTCCTTTCCAGAG 400
mouse --------------GTGCAACACCCCAGAGGCCTAGCCAGACCAACATTCCTGCCCAAAA 389
human GCCAGCTCTCCATTTATAGCCCCTGGGCAGAGCAGCCAAGGGAGCTGAGAGGGGAGGACT 460
mouse GCCAGCTCTCCATTTATAGCACCTTGGAAGACTAGCCAAGGGAGCTGAAATGCAAGGCCT 449
human GGAAAGGGCAGAGGGAGAAGGGGCAGCCCAGGCAGCACTCCCTCCCCACTCCCCACCAAA 520
mouse GGAAAGGACAGGGGGAAAT---CAAA 472
human TGAGCCCCTCATCATGAAGACAGCAGAAGCCAGGCCCAGGGCGAGGTGTGCACATGCCCC 580
mouse GGGGCCACCTAT---GGCGGCAGGAGAACTGAGCCTCAGGATGAGCTGTGTGCTT-CTCC 528
human CAAGCACAGAGCCTACCATTCTGGTCAGACCTGCGTTGAGGGGTGAGGGGGCTGCCAGGG 640
mouse AGGTCAGTGGGCCTACAG-CCTAAGAGGCCCTGCATTGAGGGGtgtagattt570
human GATCCCCTCAAAGTCCCTCAGCCCATTGCTAGTGGCCCCTCACAGAACAAGTCCAGCACC 700
mouse -------ACAATG--CCTCAGCCCAGAGCCAATGGCACG----------CTCCAGAAGG 610
human TGTGGACAAAGGGCACCCTTGACTAGACTCTGCAGTATAAGAGTTTGAATGTTTTCAGCT 760
mouse GGTG-GCTGGGGGAAGTTTT--AGTGACCATAAAATAAAAAG----------------- 649
human TCCAAACTTGGTATCCTTTTTCCCTCCGCCCCCAACCCAGCACTGGGACTAAAAGGACAA 820
mouse --CAAGGTTG-CAATCACTTAGACTCAGCATAAAATTTATTTCGG----------TT 693
human CATGTCCCAGGTTGGACATACTTCTCCCTGCTCTGTGGGCAGCAGGGAAGAGATGATGGT 880
mouse TTTGTTATTTG-TGTGCTTGCTTTGCTTTGCTTTGTTTGCT-CGGGTTTGAGATAAGAGA 751

```
human          GT-TGACAAACCTCTCTCCAAAGAGGAGACGCAACCAGAAGGGTGATTCCAGGCAGGTGT 939
mouse          GAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGAAGGAGGGAGGGAGGGAGGGAGGGAGGGA 811
               *: :** *.* . . . . *.****.... *.*. ..*.****:*. :  ..* ..* *:

human          GGATGCCAGGCATGGAG-AGGTCTGAAATGGTCACCGAGTTCAGTGAGTTCCAATCTTTT 998
mouse          GGGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGAAAGA---AAGAAAGAAAGAAAGAAAG 868
               **.:* .***.* *.** *.* .:*.** *.: *..**   .**:.**::. **: :::

human          TTTGAGCAACGGAAGCCTGGTAGCAAACAAAAATCCCACTTGAAAGCCTAATATAAAAAT 1058
mouse          AAAGAAAGAAAGAAAGAAAGAAAGAAACTAGATTTGCACAAGTGG----TTTGAAGTGCT 924
               :::**...*..***. .:.*:*. ****:*.*:*  ***::*:..    ::*.:*.:..*

human          GGCATTTTACCACTAGAATGTCTGTGTGCTTTAAAACAGCGCTTCCTAATTATGGAAGAA 1118
mouse          TCCTTGGAGACACCAG---GGCACTGCATAATGAAACAGACTTGCTCATTCATCCGCCAA 981
                 *:*  :..*** **   * *: ** . ::*.******.  * *  *:* **  .. **

human          GATGTAGCTGCAAATCAAGCTTAAAACTGTCAAAGCAGTTTAGATTTATAAGCCATAAGT 1178
mouse          GGAATATTTGTGGTGCAC-CTAGTGTGTG-CTGAGCATATTCAGCTGAGGTTCCAAAAAG 1039
               *.:.**  ** ..: **. **:.:.: ** *:.**** :**... * * .: ***:**.

human          GATAAAATATTAAATGTGTTTGGTAAGTTCAAACATATAACATTTACTTATTTATTGTAA 1238
mouse          GAGGAAGCAATAGCTTCCATCG--AAATCATTACTAACGACAATTACATC---------- 1087
               ** .**. *:**..*   :* *  **.* .::**::* .***:****:*.

human          AGGCAACTTGATGACAGCCCTGAGGAAGTTTTTAGAAACACAAAGCACAAAAAGCAAAGT 1298
mouse          ---CATC-------CCTCACTGTAGATTTAAGTCAAATGAAAGAGCACTTATGATGAACC 1137
                  **:*       *. *.***:.**: *:: *..**: *.*.*****::*:.. .**

human          TGTATTCACTTGTCTCAGCATCCAATTTATTTTGTAGTTTCTTGCTTATTCAGATTTGGG 1358
mouse          CCACATCCATCG--GCAGCATACTG-TTGGAATGTTGCAACCGACCAAT--------GGG 1186
                 :.:**..* *   ******.*:. **. ::***:* ::*  .* :**        ***

human          GAAAATCTAGATTTGCATAGATAAGTGGTTTGAATAGCTCACCTGGGAATCTCAGAGTAC 1418
mouse          AGAGAGCACG------------------------------TCCCAGGCAAACCAGCCTCT 1216
               ..*.* *:.*                             :** .**.*:. ***. *.

human          TCTTTAATTAAGTAGACTCATTCATTCATTTGCCCAAGAAATATTTATTGAGTGCCTACT 1478
mouse          GCTTTGCCTGGGCAGAGGCAG-----CGTGAGAGCTTGTATGAG----TAAGTACCTACA 1267
                ****.. *..* ***  **      *.* :*. *::*:*: *     *.***.*****:

human          ATGCACCTGGCTTTCTGCTGAGCCTTGAGCAAGTAGGAAGAAGTATTTGTTTCTAAATCA 1538
mouse          TAGAGCCCAGGTT--TGCTGG----------AATAGAATGAC----TTG---------- 1300
               ::*..** .* **  *****.          *.***.*:**.    ***

human          TCAATACAACAATTACTTACATTCCTCATTGTGAGTATAGTGAAAAAACAATAAGAAGAC 1598
mouse          ----TAGCACATTTTGCTAAATTCA--------AGTATAAGGATAGAAATCAAAAGAGCC 1348
                   ** .***:**:  **.****.        ******. **:*.**.:.:**..**.*

human          ACATCGAGATGCCAAATCTCTCTATAGTATACCACTACAATATTGCAGCATGCCATGTAT 1658
mouse          CAA-------GACTGTTCATTCACTCGCTTGACACTTAAACCACTGTGCCCGCCGTG--- 1398
               ..*       *.*:.:**: **:.*.* :*..****:.** .:   :**. ***.**

human          AATGCCCAGTTCAAGAGAACATGCCCCAGGCAGATGGGGTCTGAGCCTTTCCTGGGAAGA 1718
mouse          --GGATCACTGCTGCAGTGGCTTCCGGAGACG------------------CCATGGTGAG 1438
                  *. ** * *:. **:. .* **  **.*.                  **: **:...

human          GCAAGTGTAATAGAACATGGATGTCTAATCATGTATGTGACTTCCCAGTTTTGAAGAAAT 1778
mouse          CAAAGTACTCCTATCCATGGTATGCTGGTCTT---CGTG---TCCCT----CGGTGAATA 1488
                .****. :. :.:.*****::  **..**:*    ***   ****:     *.:***::
```

Figure 11. Alignment between human DRR sequence plus downstream 1000 bp (chr11:17714232-17716026) with mouse MUNC locus, from 400bp upstream of TSS to end of MUNC (chr7:46371003-46372492). Blue in human sequence indicates DRR, red in mouse sequence indicates MUNC exon1 and MUNC exon2.

CHAPTER III

Adapted from: MUNC, an eRNA upstream from the MYOD gene, induces a subgroup of myogenic transcripts *in trans,* independently of MyoD

Magdalena A. Cichewicz, Manjari Kiran, Róża K. Przanowska, Anindya Dutta

under review

My contribution to the paper:

- I established C2C12 cells stably overexpressing different parts of MUNC sequences and analyzed them – Fig. 1
- I established MYOD KO and MUNC KO cells and compared their differentiation efficiency to the WT cells – Fig. 2
- I established MYOD KO cells stably overexpressing MUNC, overexpressed MYOD in them, and analyzed – Fig. 3, 4
- I confirmed regulation of a few gene candidates regulated by MUNC, found in genome-wide analysis (Dr. Manjari Kiran) – Fig. 6

MUNC, an eRNA upstream from the MYOD gene, induces a subgroup of myogenic transcripts *in trans*, independently of MyoD

Magdalena A. Cichewicz*, Manjari Kiran*, Roza K. Przanowska, Anindya Dutta#

Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Pinn Hall 1232, Charlottesville, Virginia 22908

#Corresponding Author: ad8q@virginia.edu

*M.A.C. and M.K. contributed equally to this study

Abstract

MyoD Upstream Noncoding RNA (MUNC), initiates in the Distal Regulatory Region enhancer of *MYOD* (DRR), and is formally classified as an enhancer RNA (DRR^{eRNA}). MUNC is required for optimal myogenic differentiation, induces specific myogenic transcripts *in trans* (*MYOD*, *MYOGENIN*, and *MYH3*) and has a functional human homolog. The vast majority of eRNAs are believed to act *in cis* primarily on their neighboring genes (1, 2), making it likely that MUNC action is dependent on the induction of *MYOD* RNA. Surprisingly, MUNC overexpression in $MYOD^{-/-}$ C2C12 cells induces myogenic transcripts in the complete absence of MyoD protein. Genome wide analysis shows that while some of the MUNC-regulated gene expression is dependent on MyoD, there is a core set of genes that are regulated by MUNC, both upwards and downwards, independent of MyoD. MUNC and MyoD even appear to act antagonistically on certain transcripts. Consistent with the idea that MUNC acts more like a lncRNA than a classic eRNA, there are at least two independent functional sites on the MUNC lncRNA, exon 1 being more active than exon 2, with very little activity from the intron. These results show that although MUNC is an eRNA, it regulates expression of many genes as a *trans*-acting lncRNA.

Introduction

Myogenesis is a process of skeletal muscle differentiation occurring during vertebrate embryo development, and in the adult during regeneration of muscle fibers after injury. During embryonic development muscles derive from the mesoderm, where myoblast – embryonic stem cells give rise to muscle fibers (3). Myogenesis requires a

network of muscle-specific transcription factors composed of four muscle regulatory factors (MRFs) from the basic helix-loop-helix (bHLH) family of transcription factors (myogenic factor 5- Myf5, myoblast determination protein- MyoD, myogenin and muscle specific regulatory factor 4- MRF4). When myogenesis is activated, MyoD with E-proteins bind to E-box sequence in promoters of genes, driving their transcription and setting off a transcriptional cascade (4). This activation leads to the expression of several muscle specific target genes such as *MYOGENIN*, *M-CADHERIN*, myosin heavy and light chains (such as *MYH3)*, and muscle creatine kinase (5).

In vitro studies identified three *MYOD* regulatory elements in mice: Proximal Regulatory Region (PRR) that is adjacent to the transcription start site (TSS) of *MYOD*, a 720 bp long Distal Regulatory Region (DRR) located ~5 kb upstream from *MYOD* TSS, and a Core Enhancer Region (CER) located ~23 kb upstream from *MYOD* TSS (6) (7) (8). The DRR sequence is functionally conserved between mouse and human, sharing blocks of sequence identity over a 445 bp region between the two species. DRR deletion reduces MyoD RNA and protein level in adult muscle (9) (10). DRR contains consensus binding sites for MyoD, MEF-2 and SRF (10) (11), explaining how it positively regulates MYOD expression like a classic enhancer. DRR is essential for skeletal muscle cells development as an enhancer, but it also serves as the initiation site of a myogenic enhancer RNA, MUNC or DRR^{eRNA}, which plays a positive regulatory role during muscle development (12) (13).

Long non-coding RNAs (lncRNAs) form a diverse family of RNA transcripts longer than 200 nt, which do not encode proteins but have different functions in the cell as RNA molecules (reviewed in (14)). Global analysis of the mouse transcriptome suggests that lncRNAs are major functional components of the genome (15). Mainly transcribed by

RNA polymerase II (PolII), lncRNA can be intergenic, multiexonic, antisense to known genes or from regulatory elements located distally from known TSS (14). High-throughput RNA sequencing identified many novel lncRNAs specifically expressed during skeletal muscle differentiation (16). Their mechanisms of actions are heterogenous and they are localized differently in cells (reviewed: (17) (18)). Nuclear lncRNAs can mediate epigenetic changes by recruiting chromatin remodeling complexes to specific genomic loci. Muscle-specific Steroid Receptor RNA Activator (SRA), promotes muscle differentiation through its interactions with RNA helicase coregulators p68, p72, and MyoD (19). Another example of promyogenic lncRNA functioning *in cis* is Dum (developmental pluripotency-associated 2 (Dppa2) Upstream binding muscle RNA), which silences its neighboring gene, Dppa2, by recruiting Dnmts to its locus (20). DBE-T, a lncRNA produced selectively in patients with Facioscapulohumeral muscular dystrophy (FSHD), binds to the chromatin and recruits transcriptional activator Ash1L, to de-repress the FSHD locus (21).

An important group of nuclear lncRNAs may work as enhancer RNAs (eRNAs) stimulating transcription of adjacent genes (1). A recent study of 12 mouse lncRNAs identified 5 that act as eRNAs in that they stimulate the transcription of the adjoining gene *in cis* by a process that involves the transcription and splicing of the eRNA, but is not dependent on the sequence of the actual RNA transcript (2). Myogenic eRNAs include DRR^{eRNA} or MUNC and CER^{eRNA} and consistent with current models of eRNA function, stimulate expression of the adjoining *MYOD* gene *in cis* by increasing chromatin accessibility for transcriptional factors. DRR^{eRNA} or MUNC is already a little atypical as

an eRNA because it can induce expression not only of the *MYOD* gene located *in cis*, but also of *MYOGENIN* and *MYH3*, which are located on different chromosomes (12) (13).

In this study, we show that MUNC has a function independent of its action as an eRNA that stimulates expression of *MYOD*. Specifically, MUNC has a MyoD-independent pro-myogenic function during skeletal muscle differentiation, has multiple separate functional regions, can act *in trans* on multiple genes on different chromosomes. This raises the possibility that although many eRNAs act as classic enhancer RNAs that stimulate transcription of adjoining genes merely from the act of transcription and splicing, some of them may have additional roles as *trans*-acting lncRNA where the sequence of the RNA matters for its function.

Results

MUNC as a lncRNA has multiple domains important for its function.

In the previous study we showed that stable overexpression of MUNC from a heterologous site in C2C12 cells increases levels of myogenic RNAs, *MYOD*, *MYOGENIN*, and *MYH3* (13). This in itself is at odds with the prevailing model that the act of transcription and splicing at the endogenous eRNA locus is important for the action of the eRNA. We therefore decided to investigate the second tenet of the eRNA hypothesis: is the specific sequence of the MUNC transcript irrelevant for stimulating the myogenic transcripts? Fragments of MUNC containing different parts of the RNA were stably overexpressed in C2C12 cells (**Fig. 1A**). The overexpression was confirmed both in proliferating myoblasts **(Fig. 1C-E)** and in differentiating myotubes **(Fig. 1F-H)**. For comparison, we used C2C12 cells stably transfected with the spliced isoform of MUNC

and with the genomic sequence of MUNC (overexpressing both spliced and unspliced isoforms). We compared the expression levels of *MYOD*, *MYOGENIN*, and *MYH3* RNAs in proliferating conditions (GM) to see whether MUNC is able to induce myogenic factors when cells still proliferate, and after 3 days of differentiation (DM3) to see whether overexpression of MUNC is still able to change myogenic RNA levels when other myogenic factors are already induced (**Fig. 1B**). In differentiating cells the overexpression of MUNC induced *MYOGENIN* and *MYH3* to much higher levels than in proliferating cells, suggesting that differentiating cells may express additional factors that facilitate MUNC's action. Second, *MYOD* induction in DM3 was much lower than that of *MYOGENIN* and *MYH3*, suggesting that there may not be a quantitative correlation between *MYOD* induction and that of *MYOGENIN* and *MYH3*, as would have been expected if *MYOD* is an obligate intermediary connecting MUNC with *MYOGENIN* or *MYH3*. This lack of correlation is consistent with our earlier observation that MUNC overexpression induced *MYOGENIN* and *MYH3* mRNA, without inducing MyoD protein (despite the induction of *MYOD* mRNA) (13). Finally, exon 1 was more potent than exon 2 at inducing these transcripts, while the intron was virtually ineffective. Consistent with the structure of the RNA being important, we observed that although the intron was ineffective by itself, it inhibited the activity of exon 1 and stimulated the activity of exon 2 when fused to the exons. Therefore, the effects of MUNC overexpression on these three genes are not coordinated, and MUNC function depends on its sequence and secondary structure, predicting the importance of its interaction with functional partner(s) to turn on these myogenic transcripts.

There have been a few reports of lncRNAs encoding micropeptides with biological function (22). Spliced MUNC transcript could code for three such micropeptides unrelated in sequence to each other (three red lines in **Fig. 1A**). The structure function analysis reported above rules out the possibility that the induction of the three genes is due to any of these micropeptides.

The MUNC genomic fragment expressed both spliced and unspliced MUNC. Spliced MUNC and genomic MUNC had variable patterns of induction on the three genes. In GM spliced MUNC induced *MYOD, MYH3* and *MYOGENIN* >4x more than genomic MUNC. In DM3, however, spliced MUNC induced *MYOD* 5x more than genomic MUNC but the two forms of MUNC induced *MYOGENIN* and *MYH3* equally. This difference in action on individual transcripts is difficult to explain if all genes are stimulated by the same mechanism, namely the induction of *MYOD*. For simplicity's sake, most of the subsequent analysis was done exclusively with cells stably overexpressing spliced MUNC because it is more potent than genomic MUNC at inducing *MYOD*.

***MYOD1* knock out diminishes muscle differentiation *in vitro*.**

A crucial role of MyoD during skeletal muscle differentiation was established both *in vitro* as well as *in vivo*. Skeletal muscles of *MYOD*$^{-/-}$ mice displayed reduced capacity of regeneration following injury (23), and *in vitro* knock down of MYOD in differentiating C2C12 cells decreased the efficiency of differentiation (24) (13). It is also known that knock-down of MUNC decreases expression of *MYOD*, and negatively affects other downstream effectors of muscle differentiation(13). To investigate whether the role of MUNC during muscle differentiation is primarily through the induction of MyoD, or whether MUNC has activities independent of MyoD, we engineered a *MYOD1*$^{-/-}$ C2C12

cells. Using CRISPR-Cas9 technology(25) both alleles of *MYOD1* were knocked out by deletion of 150bp of *MYOD1* exon1 (corresponding to amino acids P7-L57 of MyoD protein and throwing the rest of the protein out of frame) **(Fig. 2A)**. The deletion was confirmed by PCR on the genomic DNA **(Fig. 2B)** and by Sanger sequencing of the PCR products (data not shown). The homozygous deletion was associated with the complete absence of MyoD protein in the cells, confirmed by antibodies recognizing an epitope in the C-terminus of the protein **(Fig. 2C)**. To check how absence of MyoD protein affects C2C12 differentiation, we differentiated WT and *MYOD*$^{-/-}$ cells for 3 days and compared the induction of myogenic markers. WT cells showed the expected induction of specific myogenic transcripts after differentiation: *MYOD* was induced 5X, **(Fig. 2D)**, *MUNC* by 30X **(Fig. 2E)** and *MYOGENIN* and *MYH3* by 1000X and 10,000X respectively **(Fig.2F, G)**. In contrast *MYOD*$^{-/-}$ cells lacking the deleted part of the *MYOD* transcript **(Fig. 2D)** had low expression of *MUNC* **(Fig. 2E)**, and only a 10X induction of *MYOGENIN* or *MYH3* RNAs **(Fig. 2F, G)**. In addition, myosin heavy chain (MHC) detected by immunostaining in DM3 cells was lower in *MYOD*$^{-/-}$ differentiating cells than in WT cells **(Fig. 2H)**. These results agree with previous reports that *MYOD1* is essential for myogenesis *in vitro* and confirm that we have successfully deleted *MYOD1* in the C2C12 cells.

MUNC knockout causes a greater disruption of myogenesis than _MYOD1_ knockout.

In parallel we generated *MUNC*$^{-/-}$ C2C12 clones **(Fig. 2I)** and evaluated them after >20 passages, hypothesizing that if MUNC works entirely in a MyoD dependent manner, we will see exactly the same phenotype for both MUNC and *MYOD1* knock-outs. We compared the differentiation efficiency of *MUNC*$^{-/-}$ clone at high passage (>20 passages)

to that of *MYOD1*$^{-/-}$ C2C12 cells at low passage (<10) and high passage (>20). The deletion of MUNC by CRISPR-Cas9 engineering was confirmed by PCR **(Fig. 2J)** and Sanger sequencing of the PCR products (data not shown). The early passage *MYOD1*$^{-/-}$ cells and the *MUNC*$^{-/-}$ cells were equally disabled in differentiation: *MYOGENIN* or *MYH3* RNAs were decreased at least 10X compared to WT cells (**Fig. 2L, M**). Surprisingly the late passage *MYOD*$^{-/-}$ cells differentiated as efficiently as WT cells (**Fig. 2L, M**), even though *MYOD1* RNA was still deleted (**Fig. 2K**). Thus, upon passage, C2C12 cells without MyoD protein activated some back-up pathway to substitute for MyoD. In contrast, lack of the *MUNC/DRR* locus gave a more severe phenotype where the cells could not recover their ability to differentiate even after additional passages. Therefore, lack of *DRR* or the MUNC transcript caused a greater disruption of differentiation *in vitro* than produced by the simple decrease of MyoD protein.

Stable overexpression of MUNC in *MYOD*$^{-/-}$ cells induces *MYOGENIN* and *MYH3* transcripts levels independently of MyoD

After seeing a clear difference between *MUNC*$^{-/-}$ and *MYOD*$^{-/-}$ cells we hypothesized that MUNC has an additional MyoD-independent role during skeletal muscle differentiation. To test this, we stably overexpressed spliced MUNC in *MYOD*$^{-/-}$ C2C12 cells. Cells overexpressing MUNC **(Fig. 3A)** showed higher expression of *MYOGENIN* in both GM (100X induction) and DM (10X induction) than control cells not overexpressing MUNC **(Fig. 3B)**. *MYH3* RNA was also increased by 10X in both GM and DM **(Fig. 3C)**. This observation shows that lncRNA MUNC is able to induce *MYOGENIN* and *MYH3* in the complete absence of MyoD protein.

Transient expression of MyoD does not further stimulate the induction of *MYOGENIN* or *MYH3* by MUNC

The 10X induction of *MYOGENIN* and *MYH3* by MUNC overexpression in DM *MYOD*$^{-/-}$ C2C12 cells is lower than the induction of these RNAs by MUNC in WT C2C12 cells (40X and 180X, respectively) (13) leading us to test whether transient expression of MyoD protein will stimulate the induction of these genes by MUNC. *MYOD*$^{-/-}$ cells overexpressing MUNC were transiently transfected with a vector expressing *MYOD* **(Fig. 4A, B)**. Relative to control cells, *MYOGENIN* RNA was induced 3X by MyoD alone, 4X by MUNC alone and 6X by both MyoD and MUNC **(Fig. 4D)**. *MYH3* RNA was similarly induced 3X by MyoD alone, 4.5X by MUNC alone and 6X by both MyoD and MUNC **(Fig. 4E)**. The differences between MyoD or MUNC alone and MUNC+MyoD expression do not reach statistical significance. Thus, although MUNC acts in a MyoD-independent manner on *MYOGENIN* and *MYH3*, and although the induction is 4-18 fold more in WT C2C12 cells than in *MYOD*$^{-/-}$ cells, transient restitution of MyoD protein is insufficient to add to the stimulation of these two RNAs by MUNC.

MUNC overexpression regulates many cellular genes in the complete absence of MyoD protein

MUNC induces *MYOGENIN* and *MYH3* independent of MyoD. To estimate the number of MUNC-induced RNAs which are dependent and independent of MyoD we checked global RNA changes produced by MUNC overexpression in WT cells (WT) and *MYOD*$^{-/-}$ cells after 3 days of differentiation (DM3) **(Fig. 5A and C)**. The Venn diagram in **Fig. 5A** shows that as expected, we identified many (3678) genes which were induced by MUNC only in the WT cells, but not in the *MYOD*$^{-/-}$ cells, suggesting that there is a

large fraction of genes that are induced by MUNC in a MyoD dependent manner. There were 35 genes similar to *MYOGENIN* and *MYH3* that are induced by MUNC in the presence or absence of MyoD and 157 genes that were induced by MUNC in *MYOD*$^{-/-}$ cells, but not in WT cells. These last two groups clearly show that MUNC can regulate the expression of many genes independent of MyoD protein.

The scatter plots in **Fig. 5A** examine how individual genes in each of these three groups behave upon MUNC overexpression in WT cells and in *MYOD*$^{-/-}$ cells. Genes that were induced in both types of cells (top-right plot), were less induced in the absence of MyoD, similar to the behavior of *MYOGENIN* and *MYH3* described above. The genes that were exclusively induced by MUNC in WT cells (lower-left plot) were mostly unaffected in the *MYOD*$^{-/-}$ cells (log2 fold change from 0.2 to -0.2), though there were a few that were repressed by MUNC in *MYOD*$^{-/-}$ cells. Surprisingly, of the 157 genes that were induced by MUNC exclusively in the *MYOD*$^{-/-}$ cells (lower-right plot), a large number were *repressed* by MUNC in WT cells, suggesting that the presence of MyoD in the latter reverses the direction of change induced by MUNC. This stimulated us to examine whether MyoD might actually repress some of these 157 genes (**Fig. 5B**), by comparing their expression in DM3 in WT cells against *MYOD*$^{-/-}$ cells (without overexpression of MUNC). Out of these 157 genes, 88 are differentially regulated (P value < 0.05) when comparing WT DM cells to *MYOD*$^{-/-}$ DM cells. Out of these 88 genes, 45 are higher and 43 are lower in WT cells compared to *MYOD*$^{-/-}$ cells, suggesting that 45 are positively regulated by MyoD, and 43 are repressed by MyoD. Thus, of the 157 genes induced exclusively by MUNC in the absence of MyoD, 43 could be repressed by MUNC in WT cells indirectly through the induction of MyoD. However, in the absence of MyoD, MUNC regulates them

in the opposite direction (induction), providing further support to the hypothesis that MUNC regulates many genes completely independent of MyoD protein. However, 45 genes are induced (directly or indirectly) by MyoD and by MUNC in the absence of MyoD, and yet, when MUNC is overexpressed in the presence of MyoD in WT cells, the genes are repressed, suggesting that their regulation in WT cells cannot be explained by postulating additive regulation by MUNC and MyoD.

Another interesting category of genes were those that are repressed by MUNC in differentiating conditions **(Fig. 5C)**. Here, too, there were many genes (4021) that were repressed by MUNC only in the presence of MyoD, suggesting that MUNC represses these genes perhaps indirectly through the action of MyoD. 26 genes were repressed by MUNC in the presence or absence of MyoD and 173 genes are repressed by MUNC only in the absence of MyoD, again showing evidence of MUNC activity independent of MyoD protein. The scatter plots again suggest that the last category of 173 genes includes many genes that are paradoxically upregulated by MUNC in WT cells. Of these, the plot in **Fig. 5D** identifies 6 genes that are induced by the presence of MyoD, and so might be induced by MUNC in WT cells through the induction of MyoD. However, in the absence of MyoD, MUNC independently acts on these same genes and represses them. Here again, there are 88 genes that are repressed by MyoD (in the comparison of WT and *MYOD*$^{-/-}$ cells) and by MUNC in the absence of MyoD, yet overexpression of MUNC in WT cells (in the presence of MyoD), does not lead to their repression, suggesting that MyoD and MUNC do not act additively on all promoter.

Collectively these results suggest that although MUNC and MyoD co-operate to regulate many genes, clearly there is a subset of genes that are regulated by MUNC in the

complete absence of MyoD protein, consistent with our hypothesis that MUNC is not merely an eRNA whose only role is to induce *MYOD1* transcription. Additionally, we observed a group of genes which are regulated by MyoD and MUNC in opposite directions, which suggests that the two factors may work in some pathways as antagonists.

A co-factor for MUNC present in WT cells but not in *MYOD*$^{-/-}$ cells

Since the RNA-Seq analysis identified 35 genes besides *MYOGENIN* and *MYH3* that are induced by MUNC in WT and *MYOD*$^{-/-}$ C2C12 cells (**Fig. 5A**), we again tested by qRT-PCR whether the induction of these genes is significantly greater in WT cells (**Fig. 6A**) than in *MYOD*$^{-/-}$ cells (**Fig. 6B**). We primarily focused on genes whose products are functionally and structurally connected to skeletal muscle function. The genes were: *Tmem8c*, a gene coding for Myomaker, a protein essential for fusion of embryonic and adult myoblasts (26), *Acta1*, one of the main structural proteins in the sarcomere unit (27); *Mylpf*, a gene coding for the regulatory light chain of striated muscle (28); *Ablim3*, encoding a protein strongly binding to F-actin suggesting its role as a scaffold for actin cytoskeleton signaling (29); *Tnnc1*, a gene coding for Troponin C a part of troponin complex, a structural complex responsible for muscle contraction (27). As for *MYOGENIN* and *MYH3*, we discovered that MUNC stimulates three of the genes much more in WT cells than in *MYOD*$^{-/-}$ cells: *Tmem8c* (10,000X vs 6X), *Acta1* (40X vs. 6X) and *Mylpf* (7X vs. 1.8X). On the other hand, *Ablim3* and *Tnnc1* were stimulated to the same extent in WT cells as in *MYOD*$^{-/-}$ cells, suggesting that the results with previous three genes was not due to a global defect of MUNC expressed in the *MYOD*$^{-/-}$ cells. Thus, it is likely that WT cells possess a co-factor of MUNC for certain genes that is absent in the *MYOD*$^{-/-}$ cells,

though at present we cannot say (based on the results in **Fig. 4**) that the factor is MyoD protein.

MUNC regulates muscle related genes in *MYOD*$^{-/-}$ cells

To demonstrate the reproducibility of regulation of gene expression by MUNC overexpression independent of MyoD, we performed hierarchical clustering of the differentially expressed genes in *MYOD*$^{-/-}$ C2C12 cells both in GM and DM3 **(Fig. 7A).** In both GM and DM3 MUNC overexpression induces as well as represses genes in absence of MyoD, and the pattern of this regulation is preserved in two independent experiments. Gene Ontology terms that are enriched among the genes regulated by MUNC in DM3 in the absence of MyoD indicate that many of them are associated with skeletal muscle development and muscle structure **(Fig. 7B).** A similar analysis of GO terms enriched among genes regulated by MUNC in GM, in *MYOD*$^{-/-}$ cells, shows less enrichment of genes involved in skeletal muscle development and structure (data not shown). This is consistent with the idea that even in the absence of MyoD, there are other factors present in DM3 conditions, but not in GM conditions, that co-operate with MUNC to regulate myogenic genes.

To determine the most significant molecular pathway regulated by MUNC in the absence of MyoD, we performed a gene set enrichment analysis (GSEA) on the genes differentially regulated upon MUNC overexpression in MYOD$^{-/-}$ cells in DM3. The plot shows significant enrichment of genes involved in muscle contraction among the genes induced by MUNC. The table below the plot shows a list of the top 10 genes contributing to the enrichment score for muscle contraction GO term, which are mainly muscle structure proteins coding genes **(Fig. 7C)**.

We wanted to confirm whether the global change in gene expression induced by MUNC in WT C2C12 cells in GM is similar to that seen when the same cells undergo differentiation in DM. 1982 genes were induced and 1733 genes were repressed by MUNC in WT cells growing in GM. When these genes are compared with the genes that are induced or repressed upon differentiation of WT C2C12 cells a highly significant number of genes are found to overlap (**Fig. 7D**). This result suggests that MUNC overexpression alone in GM is able to push C2C12 cells in the direction of myogenic differentiation, although of course, MUNC overexpression alone is not as potent as the differentiation induced by moving cells from GM to DM.

Discussion

The first question, this paper answers is whether MUNC is an lncRNA or just an eRNA. Recent literature suggests that long noncoding RNAs deriving from enhancer loci directly regulate the expression level of neighboring genes by a *cis*-acting mechanism (2). P53-bound enhancer regions produce eRNAs which regulate transcription of adjacent genes, as shown by reporter assays and RNAPolII ChIP assay (30). Additional examples are activating ncRNAs, ncRNA-a3 and ncRNA-a7, whose depletion decreases RNAPolII abundance at adjacent genes, as well as the recruitment of Mediator to the adjoining promoter. Chromatin looping between these enhancers and target genes is decreased when the eRNA is depleted, which is accompanied by decreased expression of the target gene (31). ERα- inducible enhancer RNAs are functionally important for expression of their target genes and are crucial for proper chromatin looping between enhancers loci and target gene bodies, which facilitates interactions between chromatin modifiers and transcription machinery (32). It was suggested that MUNC, coded by DRR genomic sequence, also acts

as enhancer RNA (12), inducing transcription of *MYOD*, and that this is MUNC's only function. However, we present data showing that MUNC positively regulates different myogenic genes, not only *MYOD*, and it has many target genes that are deregulated by MUNC overexpression in the complete absence of MyoD protein. The fact that the sequence of the RNA is necessary for the induction of *MYOD*, *MYOGENIN* or *MYH3* argues that the mere act of transcription or splicing of MUNC is not sufficient for its activity, as has been suggested for eRNAs (2). In addition, the structure-function studies show that even in WT cells, different parts of MUNC stimulate *MYOD*, *MYOGENIN* and *MYH3* RNAs to different extents that are not correlated with each other, something that would have been expected if *MYOD* was an obligate intermediate to the induction of *MYOGENIN* and *MYH3*. These results suggest that MUNC is both a classical eRNA that induces transcription of the adjoining *MYOD* RNA, and also an lncRNA that has actions independent of *MYOD* induction.

This result opens the possibility that there could be other eRNAs that also act as lncRNAs. So far, reports suggest that eRNAs are not being spliced, that transcription from the enhancer region is bidirectional and that transcriptionally active enhancers are tagged with H3K4me1 rather than H3K4me3 marks. Enhancer RNAs are also usually much shorter than lncRNAs (33). We know from this report and our previous study (13) that MUNC is being spliced, that the predominant stable transcript at the DRR locus is in the direction of MUNC and that the DRR genomic locus during muscle differentiation shows H3K4me3 marks. We hypothesize that eRNAs with similar features may have a dual action as an eRNA (enhancing the transcription of the adjoining gene) and as an lncRNA, that executes functions independent of its nearby neighbor.

The next question is whether MUNC lncRNA acts through the expression of an encoded micropeptide. There is growing literature that some lncRNAs code for functional micropeptides of even 30 amino acids. The most recent examples are micropeptides described by Olson and colleagues, which by interaction with SERCA regulate calcium signaling in muscle (34) (35) and nonmuscle cells (36). Additionally, it was shown that one genomic locus may produce both: functional micropeptide MLN and functional lncRNA linc-RAM, working exclusively from each other (22). Spliced MUNC transcript could code for three such micropeptides unrelated in sequence to each other (three red lines in **Fig. 1A**). However, the structure-function studies on MUNC revealed two important points. Two exons of MUNC have different potency of inducing myogenic transcripts, and addition of intronic sequence to each of them has different effects. The intron inhibits the activity of exon 1 and stimulates the activity of exon 2 when fused to the exons. Collectively, these results rule out the possibility that MUNC's lncRNA-like function could be due to any of the three putative micro-ORFs in MUNC**.**

Both MUNC and MyoD are pro-myogenic factors, raising the question whether they are additive with each other and whether they ever act in opposite directions. Our results suggest that although MUNC and MyoD co-operate to regulate many genes, clearly there is a subset of genes that are regulated by MUNC in the complete absence of MyoD protein. Additionally, we observed a group of genes which are regulated by MyoD and MUNC in opposite directions, which suggests that the two factors may work in some pathways as antagonists. A very interesting result emerged when we compared the differentiation of early and late passage *MYOD*$^{-/-}$ cells and *MUNC*$^{-/-}$ cells. The lack of MUNC produces a more severe disability in differentiation than that obtained from the lack

of MyoD protein. In fact, it is quite interesting that the *MYOD*$^{-/-}$ cells adapt, with sufficient passage, to the absence of MyoD and differentiate normally, but the *MUNC*$^{-/-}$ cells do not. This is yet another result that suggests that MUNC has additional functions besides the stimulation of MyoD expression, functions that may allow cells to adapt to the absence of MyoD.

In our first paper (13) overexpression of MUNC induced the expression of three genes, and so we had focused on MUNC as a positive factor for gene expression. The genome-wide analysis of genes regulated by MUNC in WT cells and in *MYOD*$^{-/-}$ cells presents a more complicated picture where in both types of cell MUNC induces and represses a large number of genes. MyoD, similarly, was initially thought to be a transcriptional factor that positively regulated expression of its target genes. However, it has since been recognized that MyoD also plays a role as a repressor of transcription in cooperation with HDAC1. For example, during proliferation MyoD together with HDAC1 binds to the promoter region of *MYOGENIN* and suppresses its transcription. After serum withdrawal, MyoD changes its interaction partners to P/CAF and activates transcription of *MYOGENIN* (37). It is possible that similarly MyoD is bound to specific genes promoters to repress their transcription during differentiation. It has also been proposed that MyoD can interact with chromatin looping proteins such as CTCF to disrupt repressive loops and thus inducing transcription from specific genomic regions (38). Thus, there are different, independent mechanisms by which MyoD regulates its targets. Similarly, we propose that MUNC interacts with different cellular factors, to induce or repress different targets, and that the induction and repressive functions are sometimes MyoD-dependent and sometimes not.

Although one important conclusion of this paper is that MUNC can act independent of MyoD and sometimes in the opposite direction as MyoD, it is clear that there are many functional interactions between the two pro-myogenic factors. For one, MyoD promotes the transcription of MUNC (as evidenced by the decrease of MUNC in the *MYOD*$^{-/-}$ cells) and the ChIP of MyoD protein to the DRR just upstream of the transcription start site of MUNC (**Fig. 8**). However, MUNC also promotes the expression of MyoD. In addition, there are many genes that are regulated in the same direction by MUNC (in the presence or absence of MyoD) and by MyoD. Our future goal is to describe how MUNC and MyoD co-operate together on the genes that they both induce or repress. The first possibility is that both molecules interact physically, making MyoD more efficient. Although till now we have failed to detect any direct physical interaction between MyoD and MUNC, we cannot yet rule out this possibility. Transient and weak interactions between MyoD and MUNC may be functional but difficult to show. MyoD interacts with numerous proteins to build whole complexes that regulate the expression of target genes. MUNC may interact with and activate another protein from such complex or may function as a scaffold, helping to maintain stability of interaction between transcriptional factors and chromatin remodelers.

A related goal is to describe how MUNC acts on many genes independent of MyoD (**Fig. 8**). We have to identify cellular proteins that interact with MUNC independent of MyoD. The MUNC-overexpressing *MYOD*$^{-/-}$ cells will be very important for such a search. As a nuclear transcript, MUNC may interact with chromatin modifiers, transcription factors or repressors on the chromatin. Thus, another goal will be to examine whether we can identify

specific genomic sites at which MUNC interacts with the chromatin and/or alters the chromatin landscape, without actually being found stably associated with the genome site.

Materials and Methods

Cell culture

C2C12 cells were cultured in DMEM-high glucose medium (GE Healthcare Life Sciences co.) with 10% FBS (Life Technologies co.), when differentiating, serum was switched to 2% horse serum (GE Healthcare Life Sciences co.)

Knock out strategy

CRISPR protocol with minor changes was followed to achieve deletion of a part of the *MYOD1* gene(25). Briefly, sgRNAs were designed using CRISPR DESIGN tool: http://crispr.mit.edu/. Cells were co-transfected with vectors coding for Cas9 (vector obtained from Addgene (#41815)), sgRNAs cloned into gRNA_GFP-T2 (vector obtained from Addgene #41820), and a spiking vector coding for a resistance gene. After 24-48 hrs cells were treated with puromycin (C=2ug/ml), and resistant cells were seeded to 96-well plates using single cell dilution method. Growing clones were examined for desired deletion by PCR on extracted genomic DNA (Quick Extract DNA Extraction Solution, Epicentre co.), and candidates with complete loss of the WT PCR product (homozygous deletion) were screened by immunoblotting for MyoD protein.

Stable overexpression of MUNC in C2C12 cells

PCR amplified sequence of genomic MUNC (PCR using C2C12 genomic DNA) or of spliced MUNC (PCR using cDNA from C2C12 cells DM3) was cloned into pLPCX vector

by ligation. The constructs were linearized and introduced to the C2C12 cells (XtremeGENE transfection reagent Roche). After 24hr, pools of stably transfected cells were selected with puromycin (C=2ug/ml). Vectors coding for mutant forms of MUNC were generated similarly using genomic DNA or DM3 cDNA as necessary.

To generate reagents for MUNC overexpression in MYOD-/-cells the insert was cloned into pLHCX vector by ligation method. The construct was linearized and introduced to the cells (XtremeGENE transfection reagent Roche). After 48 hr, pools of stably transfected cells were selected with hygromycin (C=300ug/ml).

Transient overexpression of MYOD in C2C12 cells

Cells were seeded on 6 well plate, and after 12 hours, were transfected with vector coding for MyoD. The medium was changed 12 hours post-transfection to differentiation medium, and cells were harvested 2 days later.

RNA analysis by qRT-PCR

RNA was isolated by Trizol extraction, RNA samples were treated with RQ1 RNAse free DNAse (Promega co.) to eliminate potential DNA contamination of samples. cDNA synthesis was performed using Superscript III RT cDNA synthesis kit (Life Technologies co.) with random hexamer priming. After cDNA synthesis, qPCR was performed with Applied Biosystems 7500 Real-Time PCR Systems using Power SYBR Green Master Mix (ThermoFisher Scientific). All primers used in this study are listed in **Table 1.**

Western Blotting

Cells were lysed in IPH buffer (50mM Tris-Cl, 150mM NaCl, 0.5%NP-40, 50mM EDTA) and run on 10% polyacrylamide SDS-PAGE gels and transferred to nitrocellulose membranes. Membranes were blocked for 30 minutes in 5% milk containing PBS-T, and incubated overnight with primary antibody in 1% milk. Secondary antibody incubation was carried out for 1 hour after washing, and at a 1:4000 dilution before washing and incubation with Millipore Immobilon HRP substrate. Antibodies and dilutions were used as follows, MyoD sc-304 1:500 (Santa Cruz co.), Tubulin 1:3000 (sc5286, Santa Cruz co.).

Immunofluorescence

Cells were plated on glass cover slips and collected in growth medium or after 3 days of differentiation medium. Cover slips were fixed with 4% formaldehyde in PBS for 10 min, permeabilized in 0.5% Triton X-100 in PBS and blocked in 5% goat serum. Coverslips were incubated at room temperature with primary antibody for 1 hour, and Alexa 488 or 549 conjugated secondary antibody for 1 hour, with three PBS washes following each antibody incubation. Coverslips were then mounted with Vectashield mounting solution (Vector Laboratories). Antibodies used were anti-MyoD C-20 antibody (Santa Cruz Laboratories), and anti-Myosin Heavy Chain A4.1025 antibody (Millipore). Antibodies were diluted 1:200 in 5% goat serum containing PBS.

Microscopy

Images were captured using a Nikon Microphot SA upright microscope equipped with a Nikon NFX35 camera using SPOT imaging software (Diagnostic Instruments Inc.) and a Nikon PlanApo 60x oil objective lens. Fluorescence images were acquired on the same day

using the same exposure times, gamma, and gain between samples. Images were enhanced for brightness and contrast to the same extent within Photoshop.

RNA-Seq libraries preparation

RNA samples were isolated from proliferating or differentiating cells using RNAeasy Mini kit (Qiagen co.). 1 ug of RNA was enriched for poly(A) tailed mRNA molecules using NEBNext Poly(A) mRNA Magnetic Isolation Module and RNA-Seq libraries were made using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB co.) according to manufacturer protocol. Pooled libraries were sequenced using paired-end protocol on Illumina platform, using NextSeq500 instrument, in the Biomolecular Analysis Facility, University of Virginia School of Medicine.

RNA-seq Analysis

We obtained >=40 million paired-end 75 bp long reads for wildtype (WT) and MYOD knockout ($MYOD^{-/-}$) conditions. Both WT and $MYOD^{-/-}$ conditions are grown in growth medium (GM) and differentiating condition (DM). Paired end reads were obtained from the two biological replicates with EV and MUNC over-expression in both GM and DM medium in WT and $MYOD^{-/-}$ C2C12 cell lines. Transcripts for mm10 refseq genes were downloaded from UCSC table browser (http://genome.ucsc.edu). We used default settings of Kallisto (39) to build an index for the downloaded 35818 transcript sequences and then quantified abundance of each transcripts from the paired end reads(39). We used DESeq2 package in R for differential expression analysis of the quantified data obtained from Kallisto (40). Gene Trail (41) and GSEA (42) were used for functional gene ontology term enrichment analysis and gene set enrichment analysis, respectively.

Acknowledgments

We thank all members of the Dutta lab for helpful discussion, especially Dr. Etsuko Shibata for technical guidance regarding CRISPR/Cas9 method. This work was supported by a grant from the NIH, R01 AR067712 to A.D. M.A.C. was partly supported by a Wagner Fellowship from the University of Virginia during this study. M.K. was partly supported by a DOD award PC151085.

References

1. Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, Guigo R, Shiekhattar R. 2010. Long noncoding RNAs with enhancer-like function in human cells. Cell 143:46–58.

2. Engreitz AJM, Haines JE, Munson G, Chen J, Elizabeth M, Kane M, Mcdonel PE, Guttman M, Lander ES. 2016. Neighborhood regulation by lncRNA promoters, transcription, and splicing. Nature 539:1–15.

3. Braun T, Gautel M. 2011. Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. Nat Rev Mol Cell Biol 12:349–361.

4. Yokoyama S, Asahara H. 2011. The myogenic transcriptional network. Cell Mol Life Sci.

5. Berkes CA, Tapscott SJ. 2005. MyoD and the transcriptional control of myogenesis. Semin Cell Dev Biol.

6. Tapscott SJ, Lassar AB, Weintraub H. 1992. A novel myoblast enhancer element mediates MyoD transcription. Mol Cell Biol 12:4994–5003.

7. ASAKURA A, LYONS GE, TAPSCOTT SJ. 1995. THE REGULATION OF MYOD GENE-EXPRESSION - CONSERVED ELEMENTS MEDIATE EXPRESSION IN EMBRYONIC AXIAL MUSCLE. Dev Biol 171:386–398.

8. Goldhamer DJ, Faerman A, Shani M, Emerson CP. 1992. Regulatory elements that control the lineage-specific expression of myoD. Science (80-) 256:538–42.

9. Chen JCJ, Ramachandran R, Goldhamer DJ. 2002. Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. Dev Biol 245:213–223.

10. L'honore A, Lamb NJ, Vandromme M, Turowski P, Carnac G, Fernandez A. 2003. MyoD distal regulatory region contains an SRF binding CArG element required for MyoD expression in skeletal myoblasts and during muscle regeneration. Mol Biol Cell 14:2151–2162.

11. Chen JCJ, Love CM, Goldhamer DJ. 2001. Two upstream enhancers collaborate to regulate the spatial patterning and timing of MyoD transcription during mouse development. Dev Dyn 221:274–288.

12. Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager G, Sartorelli V. 2013. ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. Mol Cell 51:606–617.

13. Mueller AC, Cichewicz MA, Dey BK, Layer R, Reon BJ, Gagan JR, Dutta A. 2015. MUNC, a Long Noncoding RNA That Facilitates the Function of MyoD in Skeletal Myogenesis. Mol Cell Biol 35:498–513.

14. Schonrock N, Harvey RP, Mattick JS. 2012. Long noncoding RNAs in cardiac development and pathophysiology. Circ Res.

15. Okazaki Y, Furuno M, Waterston R, Lander ES, Rogers J, Birney E, Hayashizaki Y, Consortium F, I RGERGP, Team II. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 420:563–573.

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

17. Ballarino M, Morlando M, Fatica A, Bozzoni I. 2016. Non-coding RNAs in muscle differentiation and musculoskeletal disease. J Clin Invest 126:2021–2030.

18. Neguembor MV, Jothi M, Gabellini D. 2014. Long noncoding RNAs, emerging players in muscle differentiation and disease. Skelet Muscle 4:8.

19. Colley SM, Leedman PJ. 2011. Steroid Receptor RNA Activator - A nuclear receptor coregulator with multiple partners: Insights and challenges. Biochimie.

20. Wang L, Zhao Y, Bao X, Zhu X, Kwok YK-Y, Sun K, Chen X, Huang Y, Jauch R, Esteban MA, Sun H, Wang H. 2015. LncRNA Dum interacts with Dnmts to regulate Dppa2 expression during myogenic differentiation and muscle regeneration. Cell Res 25:335–50.

21. Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D. 2012. A long ncRNA links copy number variation to a polycomb/trithorax

epigenetic switch in fshd muscular dystrophy. Cell 149:819–831.

22. Yu X, Zhang Y, Li T, Ma Z, Jia H, Chen Q, Zhao Y, Zhai L, Zhong R, Li C, Zou X, Meng J, Chen AK, Puri PL, Chen M, Zhu D. 2017. Long non-coding RNA Linc-RAM enhances myogenic differentiation by interacting with MyoD. Nat Commun 8:14016.

23. Megeney LA, Kablar B, Garrett K, Anderson JE, Rudnicki MA. 1996. MyoD is required for myogenic stem cell function in adult skeletal muscle. Genes Dev 10:1173–1183.

24. Rajan S, Dang HCP, Djambazian H, Zuzan H, Fedyshyn Y, Ketela T, Moffat J, Hudson TJ, Sladek R. 2012. Analysis of early C2C12 myogenesis identifies stably and differentially expressed transcriptional regulators whose knock-down inhibits myoblast differentiation. Physiol Genomics 44:183–197.

25. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308.

26. Millay DP, O'Rourke JR, Sutherland LB, Bezprozvannaya S, Shelton JM, Bassel-Duby R, Olson EN. 2013. Myomaker is a membrane activator of myoblast fusion and muscle formation. (Supp). Nature 499:301–5.

27. Perry S V. 1985. Properties of the muscle proteins--a comparative approach. J Exp Biol 115:31–42.

28. Wang Y, Szczesna-Cordary D, Craig R, Diaz-Perez Z, Guzman G, Miller T, Potter JD. 2007. Fast skeletal muscle regulatory light chain is required for fast and slow

skeletal muscle development. FASEB J Off Publ Fed Am Soc Exp Biol 21:2205–2214.

29. Matsuda M, Yamashita JK, Tsukita S, Furuse M. 2010. AbLIM3 is a novel component of adherens junctions with actin-binding activity. Eur J Cell Biol 89:807–816.

30. Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Vrielink JAFO, Elkon R, Melo SA, Léveillé N, Kalluri R, de Laat W, Agami R. 2013. ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. Mol Cell 49:524–535.

31. Lai F, Orom U a, Cesaroni M, Beringer M, Taatjes DJ, Blobel G a, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494:497–501.

32. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen Y, Merkurjev D, Zhang J, Ohgi K, Song X, Kim H, Glass CK, Rosenfeld MG. 2013. Functional Importance of eRNAs for Estrogen-dependent Transcriptional Activation Events. Nature 498:516–520.

33. Kim TK, Hemberg M, Gray JM. 2015. Enhancer RNAs: A class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb Perspect Biol 7.

34. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. 2015. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160:595–606.

35. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F, Reese AL, McAnally JR, Chen X, Kavalali ET, Cannon SC, Houser SR, Bassel-Duby R, Olson EN. 2016. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science (80-) 351:271–275.

36. Anderson DM, Makarewich CA, Anderson KM, Shelton JM, Bezprozvannaya S, Bassel-Duby R, Olson EN. 2016. Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. Sci Signal 9.

37. Mal A, Harter ML. 2003. MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis. Proc Natl Acad Sci U S A 100:1735–1739.

38. Battistelli C, Busanello A, Maione R. 2014. Functional interplay between MyoD and CTCF in regulating long-range chromatin interactions during differentiation. J Cell Sci 3757–3767.

39. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–7.

40. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

41. Keller A, Backes C, Al-Awadhi M, Gerasch A, Küntzer J, Kohlbacher O, Kaufmann M, Lenhof H-P. 2008. GeneTrailExpress: a web-based pipeline for the statistical evaluation of microarray experiments. BMC Bioinformatics 9:552.

42. Subramanian A, Subramanian A, Tamayo P, Tamayo P, Mootha VK, Mootha VK,

Mukherjee S, Mukherjee S, Ebert BL, Ebert BL, Gillette M a, Gillette M a, Paulovich A, Paulovich A, Pomeroy SL, Pomeroy SL, Golub TR, Golub TR, Lander ES, Lander ES, Mesirov JP, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–50.

qGAPDH F	GCACAGTCAAGGCCGAGAAT
qGAPDH R	GCCTTCTCCATGGTGGTGAA
qMYOD F	CATCCGCTACATCGAAGGTC
qMYOD R	GTGGAGATGCGCTCCACTAT
qMYOGENIN F	AGCGCAGGCTCAAGAAAGTGAATG
qMYOGENIN R	CTGTAGGCGCTCAATGTACTGGAT
qMYH3 F	TCCAAACCGTCTCTGCACTGTT
qMYH3 R	AGCGTACAAAGTGTGGGTGTGT
qMUNC F	AGCCTCAGGATGAGCTGTGT
qMUNC R	ATGGATGTGGGGTTCATCAT
MUNC exon1 F	TAGCCAAGGGAGCTGAAATG
MUNC exon1 R	AGTTCTCCTGCCGCCATAG
MUNC intron F	GGTTTGAAGTGCTTCCTTGG
MUNC intron R	GAGGGATGGATGTAATTGTCG
MUNC exon2 F	TATGATGAACCCCACATCCA
MUNC exon2 R	GGACGTGCTCTCTCCCATT
MUNC_HindIIIF (cloning into pLHCX)	TAAGCAAAGCTTATAGCACCTTGGAAGACTAGCCA
MUNC_HpaIR (cloning into pLHCX)	TGCTTAGTTAACTTATTCACCGAGGGACACGAAG
MUNC BglII F (cloning into pLPCX)	CTTAGATCGCAGATCTAGACTAGCCAAGGGAGCTGAA
MUNC NotI R (cloning into pLPCX)	CCGAGCTCTTGCGGCCGCTCAGTTATTCACCGAGGGACA
MUNCex1 NotI R (cloning into pLPCX)	CCGAGCTCTTGCGGCCGCACTGACCTGGAGAAGCACACAG
MUNCex2 BglII F (cloning into pLPCX)	CTTAGATCGCAGATCTTCAAATGAAAGAGCACTTATGATGA
MUNC intronic BglII F (cloning into pLPCX)	CTTAGATCGCAGATCTGTCAGTGGGCCTACAGCCTA
MUNC intronic NotI R (cloning into pLPCX)	CCGAGCTCTTGCGGCCGCACAGTGAGGGATGGATGTAATTGT
sgMYOD1	AGCTTCTATCGCCGCCACTCCGG
sgMYOD2	TGTAGCGGATGGCGTTGCGCAGG
MYODcrisprKO_F	CGAAGCTATGGAGCTTCTATCGCCGCCA
MYODcrisprKO_R	CCTTACCATGCCATCAGAGCAGTTGGAG
sgMUNC_1	CACCTTGGAAGACTAGCCAAGGG
sgMUNC_2	GCATACCATGGATAGGAGTATGG
MUNCcrisprKO_F	CTTGAGTTGGGAAAGGAAAGTCTAGGG
MUNCcrisprKO_R	GTCTCAGATCTCAACTCCAAAGTCATTTTT

Table 1. Primers used in the study.

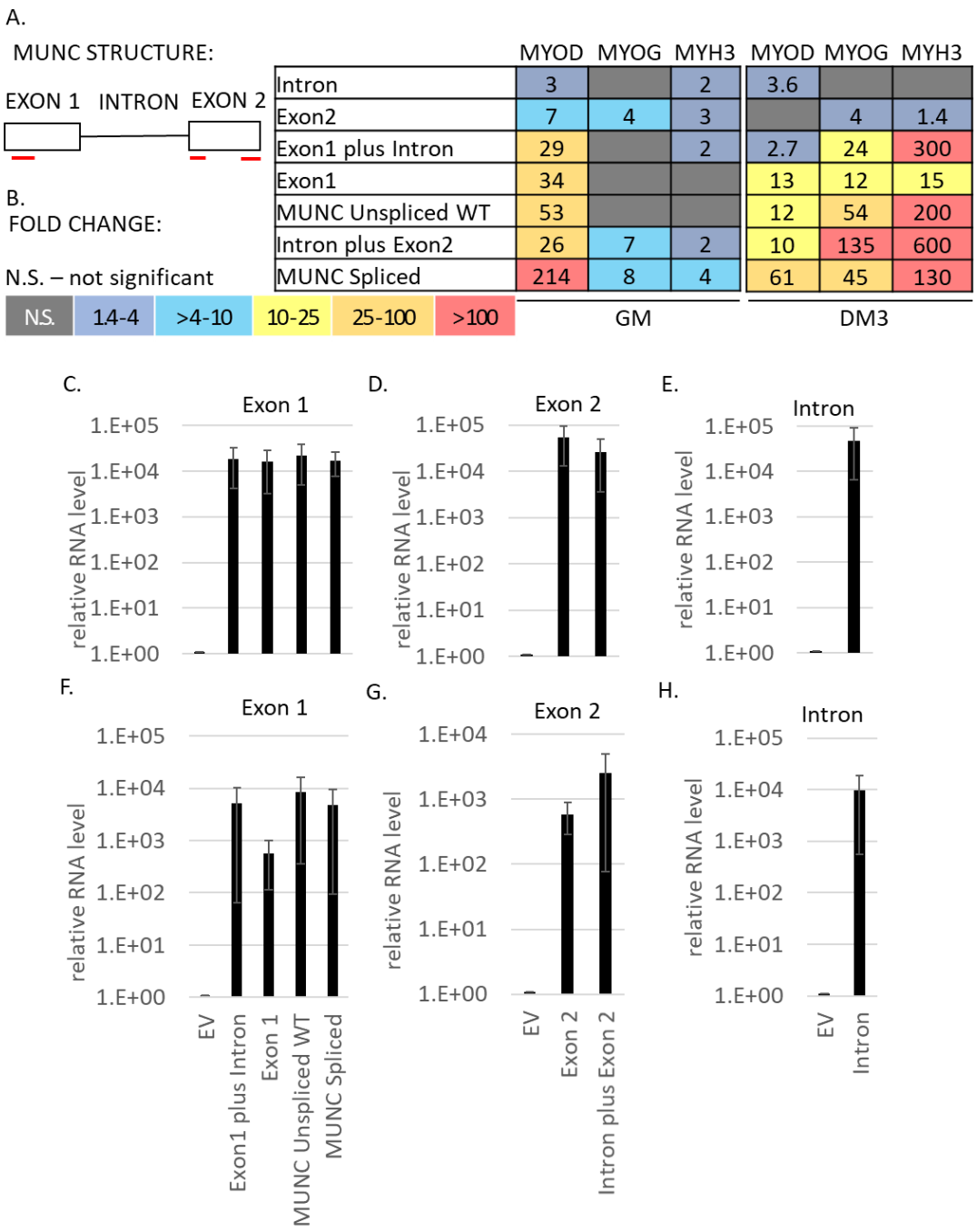


Figure 1. Legend

A. A schematic illustrating MUNC structure. Red lines indicate three potential micropeptides coded by MUNC spliced sequence: two of 20 amino acids, and one of 60 amino acids. Micropeptides were defined using translation tool (http://web.expasy.org/translate/).

B. Heat maps showing summaries of qRT-PCR analysis of C2C12 mutant cells stably overexpressing different truncated MUNC sequences. Levels of myogenic factor transcripts were measured in three biological runs, normalized to GAPDH level, and to control cells in each condition, mean values were calculated. Colors used in heat maps correspond to fold changes according to the legend. Analysis of proliferating cells and differentiating cells.

C-H. QRT-PCR analysis of mutant cells overexpressing truncated MUNC sequences showing levels of different parts of the transcript (Exon1, Intron, Exon2) in GM (C-E) and in DM3 (F-H). Data were normalized to GAPDH, and to control cells (EV). Values represent three biological replicates, presented as mean +/− SEM.

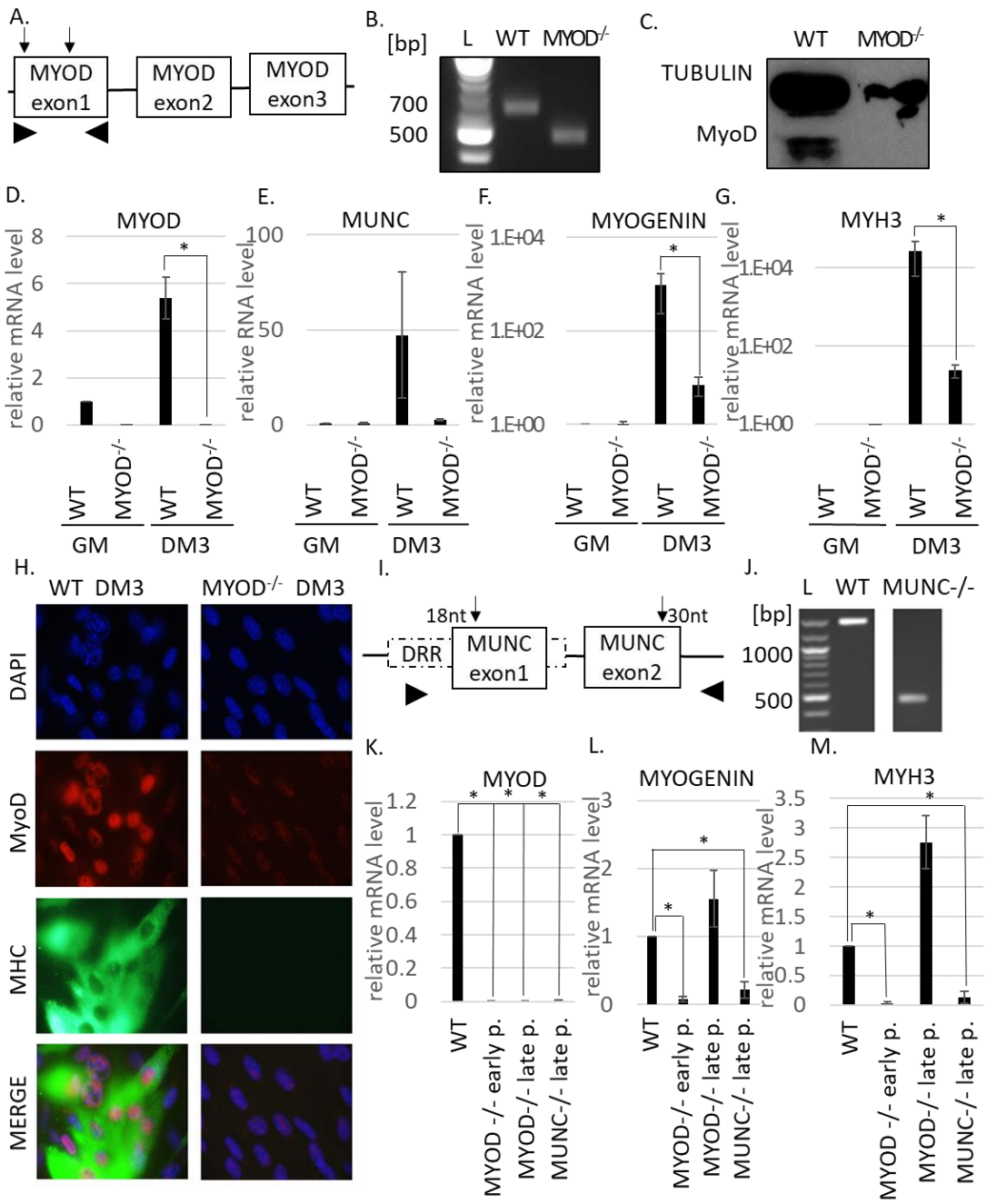


Figure 2. Legend

A. An illustration showing the deletion of MYOD1 genomic sequence causing MyoD protein deletion. Arrows indicate sites targeted by sgRNAs. Triangles indicate used genotyping primers target sites.

B. PCR products with the genotyping primers on genomic DNA confirming MYOD1 sequence deletion in $MYOD^{-/-}$ cells.

C. Western blot analysis confirming absence of MyoD protein in $MYOD^{-/-}$ cells. Tubulin serves as a loading control.

D-G. qRT-PCR analysis of proliferating (GM) and differentiating (after 3 days of differentiation, DM3) cells that are wild type for MYOD (WT) or $MYOD^{-/-}$ cells. Levels of expression were measured for MYOD (D.), MUNC (E.), MYOGENIN (F.) and MYH3 (G.) mRNAs. Data is normalized to GAPDH expression level and shown as relative to proliferating WT cells (WT GM). Values represent three biological replicates, presented as mean +/− SEM.

Statistical significance was calculated using Wilcoxon-Mann-Whitney Test. (*) indicates p-value =< 0.05.

H. Immunofluorescence analysis of fixed cells 3 days after differentiation (DM3). Cells were immunostained with antibodies against MyoD, MHC. DAPI was used to visualize nuclei.

I. An illustration showing the segment from MUNC genomic sequence that was deleted. Arrows indicate sites targeted by sgRNAs. Triangles indicate target sites of genotyping primers.

J. PCR products genotyping MUNC in WT and $MUNC^{-/-}$ cells. The lanes are from the same gel.

K-M. qRT-PCR analysis of differentiating (after 3 days of differentiation, DM3) cells that are wild type for MYOD and MUNC (WT), $MYOD^{-/-}$ early passage cells (passage number<10), $MYOD^{-/-}$ late passage cells (passage number>20), $MUNC^{-/-}$ late passage cells (passage number>20). Levels of expression were measured for MYOD (K.), MYOGENIN (L.) and MYH3 (M.) mRNAs. Data normalized to GAPDH expression level and shown as relative to WT cells. Values represent three biological replicates, presented as mean +/− SEM. Statistical significance was calculated using Wilcoxon-Mann-Whitney Test. (*) indicates p-value =< 0.05.

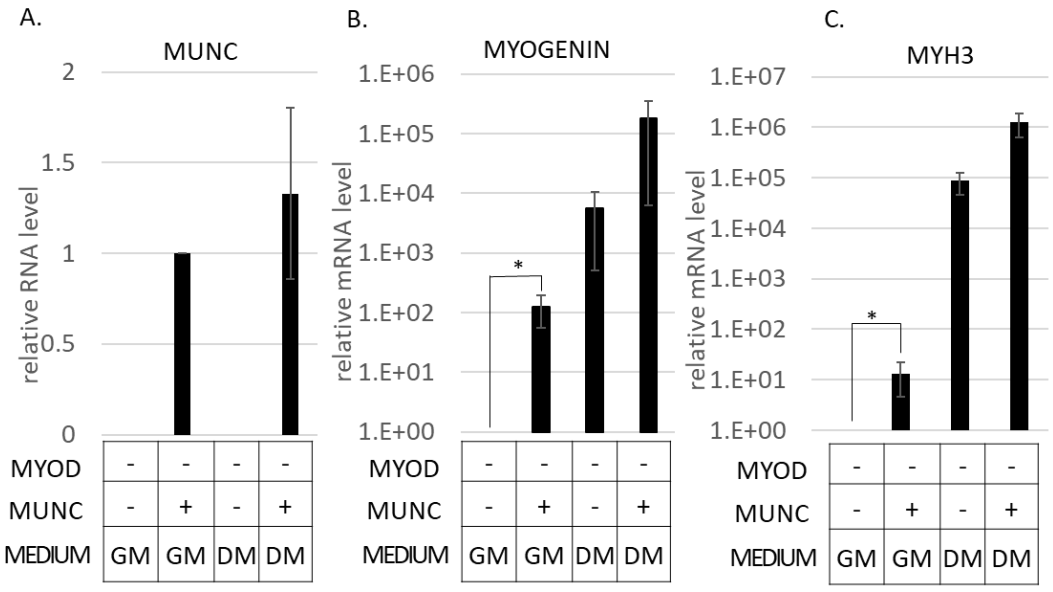


Figure 3. Legend.

QRT-PCR analysis of RNA from proliferating (GM) and differentiating (after 3 days of differentiation, DM3) MYOD$^{-/-}$ cells stably transfected with vector expressing MUNC. Levels of expression were measured for MUNC (A), MYOGENIN (B) and MYH3 (C) mRNAs. Data normalized to GAPDH expression level and shown as relative to proliferating cells (GM). Values represent three biological replicates, presented as mean +/− SEM. Statistical significance was calculated using Wilcoxon-Mann-Whitney Test. (*) indicates p-value =< 0.05.

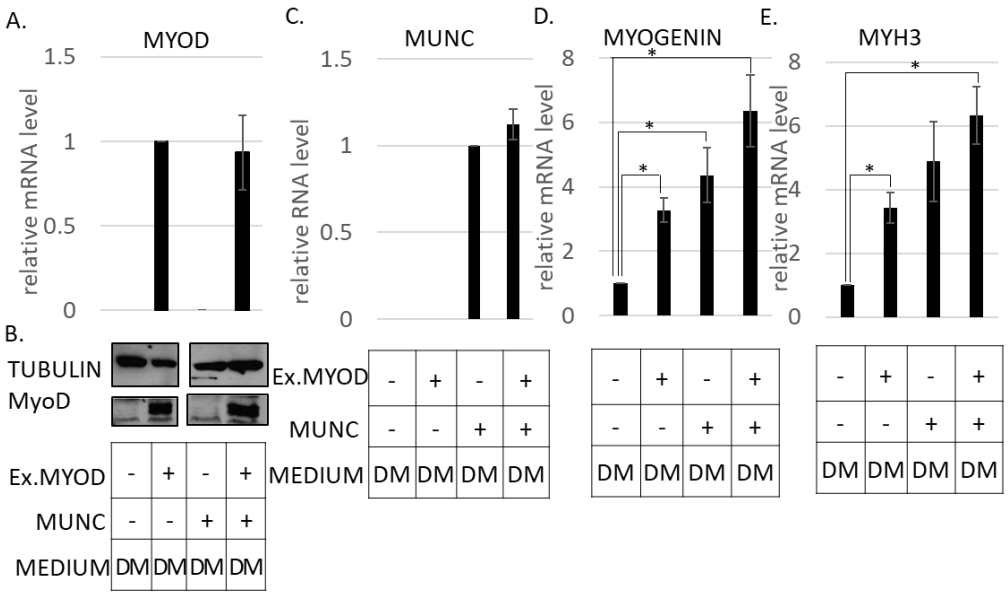


Figure 4. Legend.

(A, C-E). QRT-PCR analysis of MYOD$^{-/-}$ cells stably overexpressing MUNC, transiently transfected with vector coding cDNA for MYOD, and differentiated for 2 days. Levels of expression were measured for MYOD (A), MUNC (C), MYOGENIN (D) and MYH3 (E) mRNAs. Data normalized to GAPDH expression level and shown as relative to control cells. Values represent three biological replicates, presented as mean +/− SEM. Statistical significance was calculated using Wilcoxon-Mann-Whitney Test. (*) indicates p-value =< 0.05. B. Western blot analysis showing MyoD protein level in cells transfected with cDNA for MYOD (Ex. MYOD), TUBULIN signal serves as loading control.

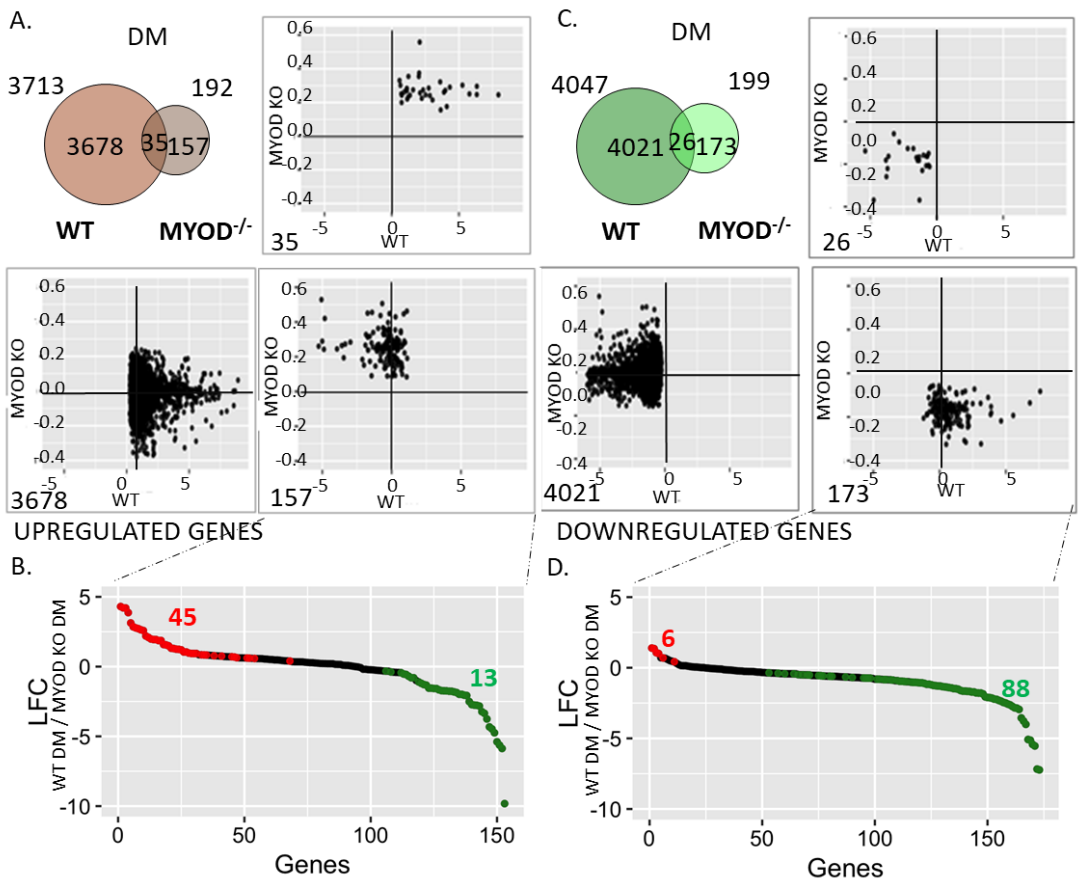


Fig.5. Legend.

A. Venn diagram representing overlap of genes which are upregulated upon MUNC overexpression in WT or *MYOD*$^{-/-}$ cells on day 3 of differentiation (DM3). The scatter plots examine how MUNC overexpression regulates (log2 fold change ovMUNC/EV) the three classes of genes in WT cells and *MYOD*$^{-/-}$ cells.

B. The 157 genes upregulated by MUNC only in *MYOD*$^{-/-}$ cells are examined to see if they are induced or repressed by MyoD. Plots represent log2 fold changes of genes on DM3 in WT vs *MYOD*$^{-/-}$ cells. The red and green dots are genes that are induced or repressed in WT cells (induced or repressed by the presence of MyoD protein) at $p<0.05$.

C. Same as **(A)**, except for genes downregulated upon MUNC overexpression in WT or *MYOD*$^{-/-}$ cells.

D. Same as **(B)**, except for 173 genes from **(C)** that are downregulated by MUNC in only in *MYOD*$^{-/-}$ cells.

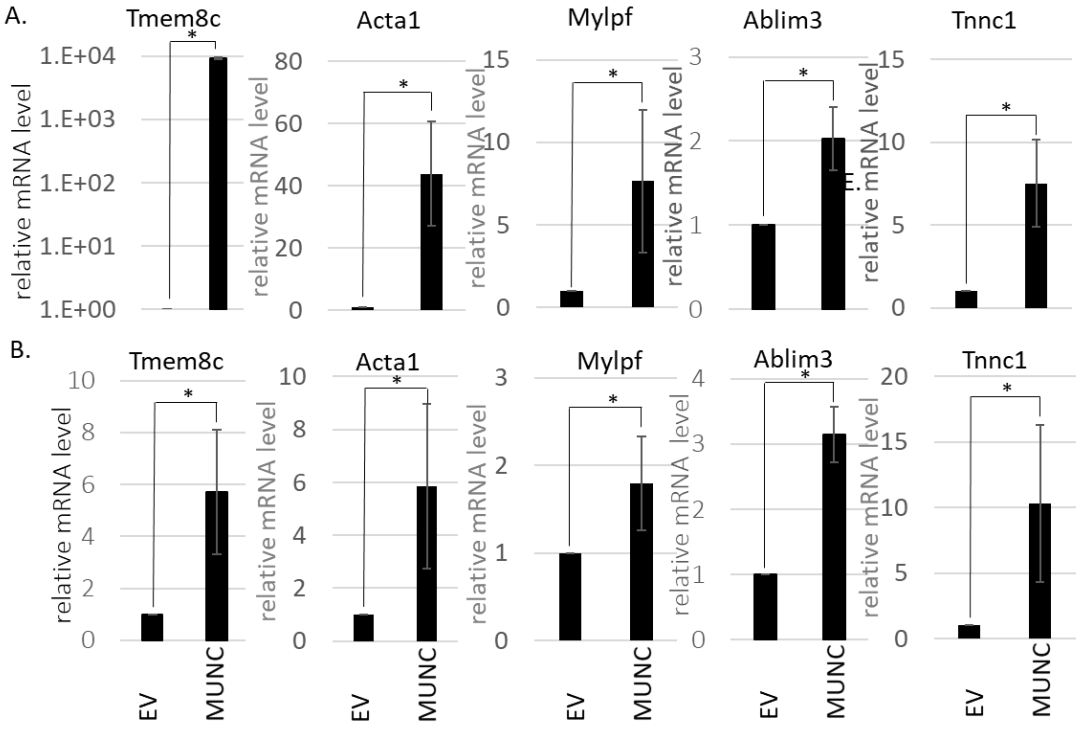
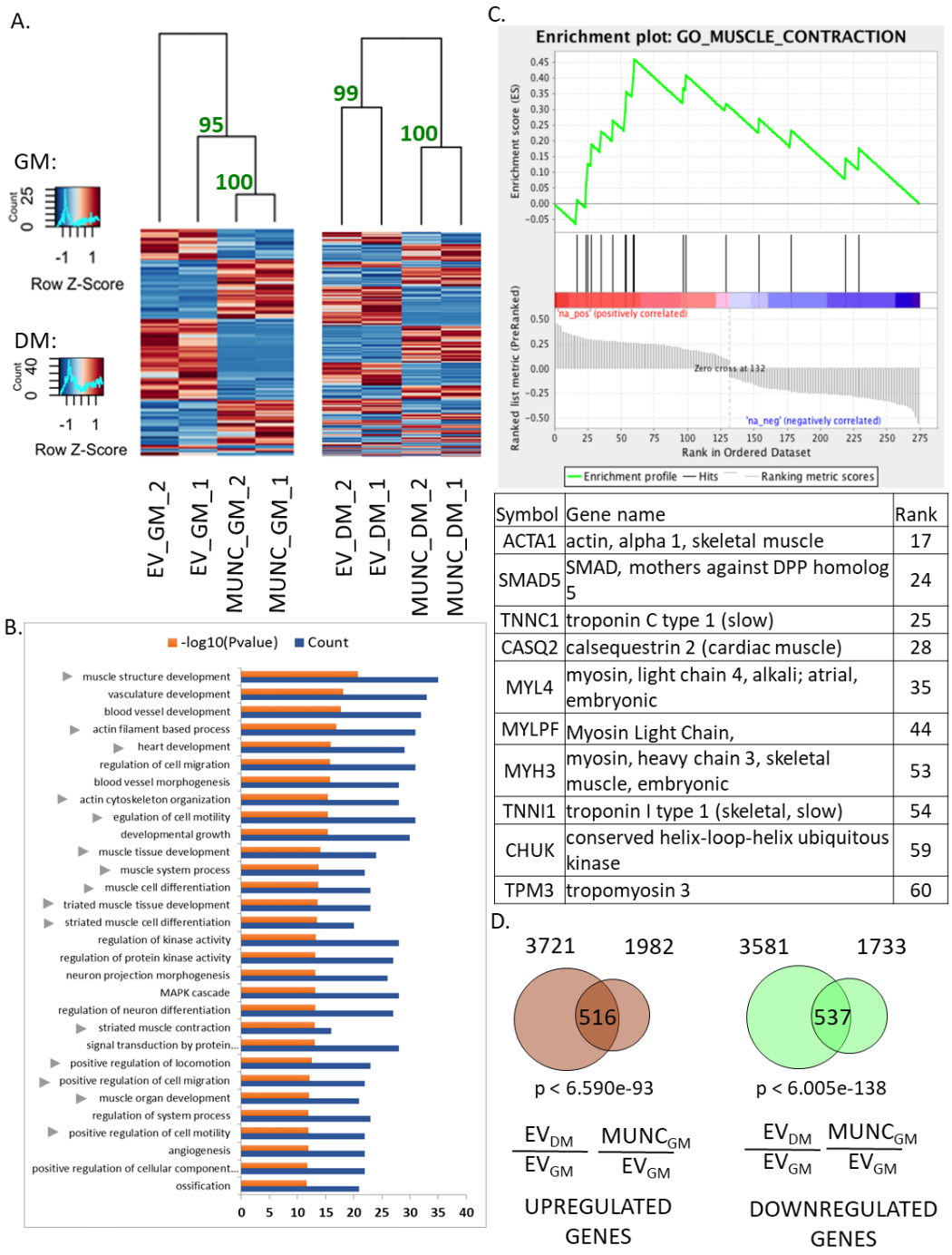


Figure 6. Legend.

Q-RT-PCR confirmation of genes upregulated upon MUNC overexpression. Analysis for WT cells (A) and for MYOD$^{-/-}$ cells (B) in differentiating conditions. Data is normalized to GAPDH expression level and shown as relative to control cells (EV). Values represent three biological replicates, presented as mean +/− SEM. Statistical significance was calculated using Wilcoxon-Mann-Whitney Test. (*) indicates p-value =< 0.05.



Symbol	Gene name	Rank
ACTA1	actin, alpha 1, skeletal muscle	17
SMAD5	SMAD, mothers against DPP homolog 5	24
TNNC1	troponin C type 1 (slow)	25
CASQ2	calsequestrin 2 (cardiac muscle)	28
MYL4	myosin, light chain 4, alkali; atrial, embryonic	35
MYLPF	Myosin Light Chain,	44
MYH3	myosin, heavy chain 3, skeletal muscle, embryonic	53
TNNI1	troponin I type 1 (skeletal, slow)	54
CHUK	conserved helix-loop-helix ubiquitous kinase	59
TPM3	tropomyosin 3	60

Fig.7. Legend.

A. Heatmaps showing clustering of conditions based on differentially regulated genes upon MUNC overexpression in proliferating conditions- GM (left) and in differentiating conditions- DM (right) in MYOD$^{-/-}$ cells. Two biological repeats for each condition were shown. Bootstrap

values based on 1000 repetition are shown near to the corresponding branches. B. Top 30 significant Gene Ontology terms enriched in differentially expressed genes in DM upon MUNC overexpression in $MYOD^{-/-}$ cells. Arrows show gene terms related to skeletal muscle development and regeneration. C. Enrichment plot from gene set enrichment analysis (GSEA) showing the gene set involved in muscle contraction is enriched among differentially regulated genes upon MUNC overexpression in $MYOD^{-/-}$ cells in DM ($P < 0.01$). The table shows list of top 10 genes contributing to enrichment score for muscle contraction GO term. D. Venn diagrams representing overlap between differentially expressed genes upon differentiation of control cells (EVDM/EVGM) versus differentially expressed genes upon MUNC overexpression in proliferating conditions (MUNCGM/EVGM) in WT cells.

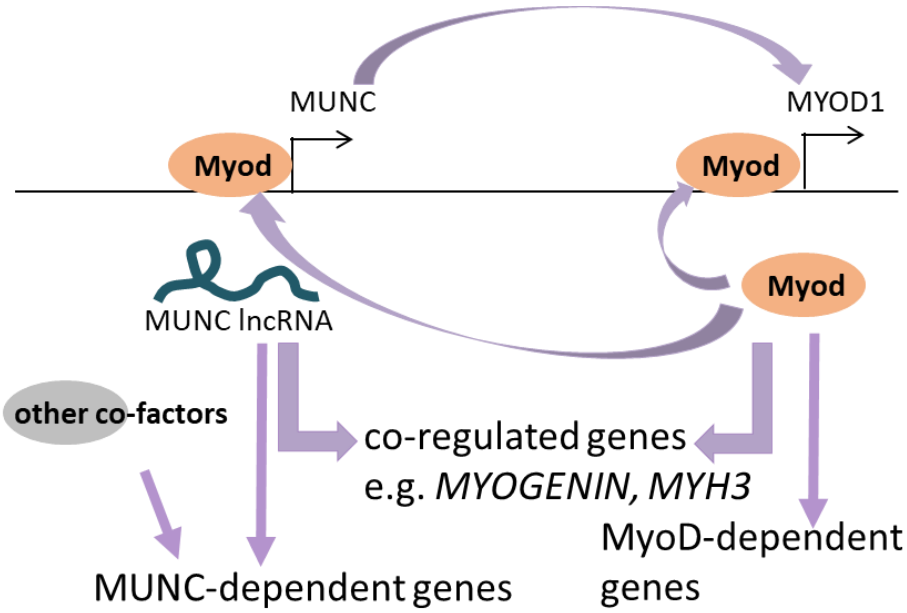


Fig.8. Legend.

Schematic showing that MUNC and MYOD1 positively regulate each other, and co-regulate many genes, but also regulate many genes independent of each other.

CHAPTER IV: UNPUBLISHED RESULTS AND DISCUSSION

Introduction

The goal of the project was to characterize the role of lncRNA MUNC in myogenesis. We successfully identified the full sequence of the transcript and its biological function during differentiation. Furthermore, we showed its MyoD-independent signaling pathway during the process. Yet the mechanism of MUNC action still remains to be addressed. One hypothesis is that MUNC acts through its protein interactor; depending on the cellular fraction where specific molecules of MUNC reside, there could be different interacting proteins. Candidates for such MUNC interacting proteins could be chromatin modifiers, such as histone acetyltransferases. MUNC may be a scaffold of the whole protein complexes and may enhance their interactions with chromatin at myogenic promoters or enhancers, which makes chromatin more accessible for transcription machinery. To test this hypothesis, we set out to identify the genomic loci that are directly affected by the transcript, and the ones that are downstream from these direct effectors of MUNC. Initial trials and future plans to characterize specific genomic sites which interact with MUNC will be discussed later in this CHAPTER.

Genome-wide screens used to find novel or functional long noncoding RNAs

One goal of the project was to build a list of putative lncRNAs upregulated during muscle differentiation. Genome-wide datasets of RNA-seq, RNA PolII ChIP-seq, and H3K4me3 ChIP-seq[1,2] were analyzed (analysis performed by Dr. Ryan Layer) to find transcripts upregulated in differentiating cells. Transcripts located within 2 kb from known exons were excluded to decrease the probability of getting products from alternative

splicing variants of known protein-coding genes. Such broad criteria exclusively focused on active transcription marks are commonly used to screen for novel long noncoding RNAs. Additionally, transcriptome-derived results (like our analysis) are sometimes filtered with ChIP-seq data for transcription factors. For example, Mousavi et al. focused on sites which were bound and presumably regulated by two critical transcription factors for myogenesis, MyoD and myogenin, to identify myogenic lncRNAs[3]. Another method to filter on a genome-wide scale for novel lncRNAs is GRO-seq. This technique recognizes nascent transcripts and so includes lncRNAs, although not all transcripts identified by GRO-seq are stable. GRO-seq is also directional and helps to describe divergent transcription occurring at Transcription Start Sites (TSS) for protein-coding genes, which presumably identifies lncRNAs that are divergently transcribed from known genes and gives an additional level of transcription regulation[4]. The above filters do not guarantee the putative lncRNA candidates do not code for proteins or micro-peptides. Before a specific transcript is investigated in details, its Coding Potential should be checked to calculate its probability of being translated into proteins[5 6]. Mass spectrometry identified peptide databases are also screened to ensure that a small peptide is not expressed from a putative lncRNA.

Functional genome-wide screens correlating transcript levels with various phenotypes have also be applied to look for functional lncRNAs. Cancer biology researchers investigate lncRNAs which are regulated by tumor suppressors, such as p53[7], by estrogens in breast cancers[8] or androgens in prostate cancers[9 10]. There are studies that identify lncRNAs whose expression level in a tumor correlates with the patient prognosis. These studies hope to find lncRNAs which may be prognosis markers during diagnosis,

and also help prioritize the lncRNAs that should be studied because of the high likelihood that they have significant biological function[11 12].

Other lncRNAs candidates from initial screening for pro-myogenic transcripts induced during differentiation

After the genome-wide screening for transcripts induced during muscle differentiation, we decided to focus on one transcript, MUNC (discussed in CHAPTER II, and CHAPTER III). Besides MUNC, I also tested the role of two other transcripts in myogenesis: MT_141 (number 6 from CHAPTER II, Figure 2A) and MT_2953 (number 13 from CHAPTER II, Figure 2A).

MT_141 is a transcript induced 10X during muscle differentiation. It is coded by a locus located within an intron of Dusp13 gene **(Fig. 1A)**. The gene encodes different Dusp13 proteins through the alternate splicing. Dusps13 are tyrosine phosphatases, which cooperate with protein kinases to regulate cell proliferation and differentiation. Expression of the distinct proteins from various mRNA isoforms of this gene has been found to be tissue specific[13]. By comparing MT_141 expression level in different murine tissues, I found that the transcript is highly upregulated in skeletal muscle, stomach and small intestine **(Fig. 1B)**. Knock down of MT_141 on differentiating C2C12 cells did not change levels of main myogenic markers (*MYOD, MYOG, MYH3*- data not shown). Because cells treated with si-MT_141 lost their adherent properties, I decided to measure the expression level of genes coding for adherent proteins. Spondin-2, a gene coding for an adhesion protein[14] was the only gene downregulated in MT_141 depleted cells **(Fig. 1C)**. Although

this observation was very interesting, there was high possibility of MT_141 being a part of an unannotated splicing variant of Dusp13 gene. The project was discontinued.

During my lab work I was fortunate to work with an undergraduate student (Myles Anderson) and a high school student (Caitlin Dutta), who with my guidance performed experiments to show that MT_2953 is relevant for differentiation. MT_2953 is induced during differentiation >100 fold, and its genomic locus is 46 kb upstream from linc-MD1 **(Fig. 2A)**, another promyogenic lncRNA, reported by the Bozzoni group[15]. I hypothesized that linc-MD1 and MT_2953 may work together to regulate myogenesis, or co-regulate each other's transcription. Using differentiating C2C12 cells **(Fig. 2B-C)** and differentiating primary myoblasts **(Fig. 2D-E),** we were able to confirm that MT_2953 is highly induced during myogenesis. Knock down of MT_2953 in C2C12 cells did not change the level of *MYOD* expression significantly, but decreased *MYOGENIN* and *MYH3* **(Fig. 2F-I)**. These preliminary data suggest that MT_2953 could be another example of a long noncoding RNA that positively regulates skeletal muscle differentiation. I really hope that one of the new students in the lab will appreciate these results and will decide to work on MT_2953.

Human homolog of MUNC

While studying the role of MUNC in myogenesis, I mainly used murine C2C12 cell line as an *in vitro* model. In addition, I also used human myoblasts to characterize human homolog of MUNC (CHAPTER II). Published results showed that human myoblasts during differentiation express a transcript from syntenic region of MUNC, upstream from the human *MYOD* locus. Knock down experiment confirmed that the transcript is relevant

for human myogenic transcripts expression. Additionally, analysis of genomic regions upstream from *MYOD* between different species showed that MUNC sequence is conserved throughout different organisms (analysis performed by Dr. Yoshiyuki Shibata) **(Fig. 3)**. I performed an experiment in which I stably overexpressed the human MUNC (1.5 kb sequence of syntenic region) in murine cells. By qRT-PCR I showed that overexpression of human MUNC sequence correlated with higher levels of murine myogenic transcripts **(Fig. 4)**. This result clearly showed that MUNC trans-complements between species. Further work is needed to characterize the 5'/ 3'end, splicing-pattern and polyA tail (if any) of the human MUNC.

RNA Pol II ChIP on the loci of myogenic genes in MUNC-overexpressing cells

From published results (CHAPTER II and III) we know that MUNC overexpression from a heterologous locus induces *MYOD, MYOGENIN and MYH3* RNA. I tested whether the MUNC mediated induction of myogenic factors is caused by direct upregulation of the transcription machinery. Using ChIP, I measured the abundance of RNA PolII molecules at the gene bodies of these three factors. RNA PolII ChIP in control cells (transfected with empty vector EV) was compared to that in cells stably overexpressing spliced MUNC (ovMUNC) in differentiation medium (DM3) to show a 3 fold increase of RNA PolII abundance at the *MYOD* gene locus upon MUNC overexpression **(Fig. 5A)**. This result suggests that even when MUNC is expressed from a heterologous locus *in trans*, it increases the transcription of *MYOD* by an unknown mechanism, different from eRNAs acting *in cis*. The abundance of RNA PolII at *MYOGENIN* **(Fig. 5B)** and *MYH3* **(Fig. 5C)** loci was not stimulated by MUNC

overexpression. Currently I am working on a nuclear run on-based assay to directly test whether MUNC controls transcription at these three loci.

Actinomycin D and DRB treatment of cells overexpressing MUNC

Another explanation for increased level of myogenic transcripts upon MUNC overexpression would be that MUNC stabilizes the RNAs of *MYOD, MYOG,* and *MYH3*. To test this I treated differentiating cells with chemicals Actinomycin D and DRB, which block the transcription machinery and I measured the abundance of the transcripts at different time points. From the representative results **(Fig. 6)** I learned that MUNC does not change stability of *MYOD, MYOGENIN* and *MYH3* transcripts produced during myogenesis.

Methods used to find protein interactors for long noncoding RNAs

One important question is to find the protein interactors for MUNC, which will help explain the mechanism of MUNC action. However, this is a common bottleneck in the lncRNA field due to the low abundance of lncRNAs. The low yield causes detection problems for associated proteins upon pull down of the RNA. People are trying to overcome this issue by using systems overexpressing transcripts of interest, but then false-positive interactors become more abundant among the potential results. Another struggle is that long noncoding RNA – protein interactions may be transient and weak so capturing them without any crosslinking methods is difficult. Still, I performed a series of screening experiments to capture potential MUNC-interacting protein candidates. Firstly, I pulled down MUNC RNA produced *in vitro* and labelled with biotin-UTP or with BrUTP and incubated the beads with cell lysate (whole cell lysate or nuclear fraction of cells). The

associated proteins were then analyzed using mass spectrometry[16]. I optimized the time of incubation of the transcript with the lysate and stringency of the washes. None of the runs gave me results worth pursuing further. The identified proteins were the most abundant proteins in the cell: mitochondrial and ribosomal proteins. After experiencing problems with high background of proteins associated *in vitro* with transcribed RNA immobilized on beads I turned to techniques of pulling down endogenously transcribed RNA. To increase pull-down efficiency, such protocols use chemical fixatives or UV light to cross link specific interactions. I decided to pursue a protocol called ChIRP-MS[17], which uses formaldehyde as a cross linker and a pool of biotinylated oligonucleotides specifically antisense to a transcript of interest to pull the lncRNA down. I tried this method because of my experience in using ChIRP to capture MUNC-associated genomic sites (experiment discussed in another paragraph of the discussion). Pull down of endogenously expressed MUNC gave me very few peptides, so the results were very difficult to interpret. My experience suggests that we should stably overexpress MUNC in cells to find its protein interactors. I would suggest overexpression of S1 aptamer tagged MUNC - a tag recognized by streptavidin, and so captured by biotinylated beads[18]. I would prefer adding a tag to the MUNC RNA instead of pulling down the RNA using antisense oligonucleotides, because we do not know whether MUNC interacts with its partner using its secondary structure or using sequence specific interaction. Antisense oligonucleotides may interfere with such interactions during the oligonucleotide hybridization step. Because of known issues with high background, I would focus on establishing multiple negative controls, such as overexpression of S1 aptamer fused to a nonmyogenic lncRNA, and optimize the fixation (cross-linking) times.

Chromatin Isolation by RNA Purification (ChIRP) for MUNC

From cellular fractionation on differentiating C2C12 cells we know that MUNC is mostly present in the cell nucleus[19]. We hypothesized that the transcript interacts with specific unknown genomic sites and changes their genomic landscape to activate or to silence expression of specific genes. To investigate this, I performed ChIRP experiment in which glutaraldehyde-fixed differentiated C2C12 cell lysate is incubated with biotinylated 20-mers antisense to the MUNC lncRNA. After hybridization and pull down, genomic DNA is recovered and sequenced. This approach will identify specific genomic sites which interact with the transcript, both directly and indirectly through additional proteins[20]. In the pilot experiment, I used differentiated WT C2C12 cells, which after lysis were incubated with MUNC-specific oligos (positive sample) or with lacZ-specific oligos (negative control). The experiment was performed twice to isolate DNA and prepare ChIRP-seq libraries. Libraries were sequenced on Mi-Seq Illumina platform (in total 4 libraries). After analysis of DNA reads (performed by Dr. Manjari Kiran), a few thousand peaks were enriched at least 4x for each biological run in MUNC-specific pull down compared to nonspecific (lacZ) pull-down , but the overlap of enriched peaks between two independent biological runs was low. We assumed that read depth of Mi-Seq platform is not enough to reach saturation level of ChIRP-seq. Re-analysis was performed, where all reads were pooled together (2 lacZ libraries were treated as one sample and 2 MUNC specific libraries were treated as one sample) **(Fig. 7)** and the top hits were confirmed by ChIRP-qPCR **(Fig. 8A, B)**. An additional problem appeared when I used MUNC knock-out cells (described in CHAPTER III) to perform the same pull-down experiment. Genomic DNA sites retrieved by MUNC specific oligos were also present in pull down samples from MUNC knock out

cells. This suggests ChIRP may have a high rate of false positive results, probably because of the irreversible crosslinking with glutaraldehyde **(Fig. 8C)**. Other groups working with this technique used additional steps to narrow down genomic sites that truly interact with the transcript of interest: (1) use of additional sets of nonspecific oligos, (2) use of two separate pools of specific oligos (odd and even) to focus only on the DNA sequences which are enriched in both pools, (3) additional pull down samples with a cell line that does not express the transcript of interest (in our case MUNC knock out cells), (4) ChIRP results are compared with complementary techniques, such as CHART or RAP (discussed below) and only common DNA sequences are followed up[21]. In the future, I recommend we perform ChIRP-seq on C2C12 WT cells and C2C12 MUNC KO cell simultaneously, with two separate pools of MUNC specific oligos, for higher stringency of defining positive genomic sites which interact with MUNC. I would also like to sequence DNA libraries on Hi-Seq platform to achieve better depth of reads, so the analysis would be statistically more reliable. Taking into account that MUNC yield in cells is low, it is also a good idea to work with higher number of cells or to use cells stably overexpressing MUNC.

Recently developed methods to describe structure and function of lncRNAs

Capture Hybridization Analysis of RNA Targets (CHART) is a method similar to ChIRP technique. The goal of CHART protocol is the same, which is to find genomic sites and proteins that specifically interact with the lncRNA of interest. Like ChIRP, CHART uses biotinylated antisense oligonucleotides which hybridize to the transcript and pull down all interactors of the specific RNA. This method uses a different cross linker, formaldehyde. To initially test whether an antisense oligonucleotide binds to its target, an RNAse H sensitivity screen could be performed. RNA retrieval is compared between

samples pulled down by a given oligonucleotide followed by with or without RNase H treatment. RNAse H hydrolyzes RNA molecules in RNA:DNA hybrids. Accessible parts of the transcript are hybridized to the oligonucleotides and recognized by RNase H. As a result, the transcript's signal is high in the pull down sample without RNAse H, while the signal is very low in the sample after RNAse H treatment. Only the antisense oligonucleotides that produce the greatest sensitivity to RNase H will be used in the following pull down steps. Fragments of the transcript which are not accessible to the oligonucleotides, are probably parts of the transcript that are protected by interacting proteins or are annealed to other parts of the transcript to form double-stranded RNA. CHART is an improvement over ChIRP because it uses oligonucleotides that are known to bind to the RNA, and like ChIRP helps to describe whether given lncRNA functions *in cis* close to its locus, or *in trans* on multiple genomic sites [22] [23] [24].

I described earlier my negative results with ChIRP-MS to identify proteins associated with MUNC RNA. A newer protocol, RNA antisense purification followed by mass spectrometry (RAP-MS), is another method developed to describe protein interactors of an lncRNA. Cells are grown in heavy or light SILAC medium, and UV crosslinked to "freeze" direct RNA-protein interactions. The cell lysate is then hybridized with a 90 nt biotinylated DNA probe antisense to the specific RNA with say the samples from the heavy medium. After streptavidin capture of DNA-RNA-protein complex, the eluate is saved. As negative control the cells grown in the light medium are processed similarly, except that the antisense oligonucleotide pull down a negative control RNA. The samples are mixed and analyzed by MS, with the SILAC protocol allowing us to detect peptides that are

significantly enriched in the heavy peptides relative to the light peptides[25]. These are likely to be from proteins that associate specifically to the test lncRNA.

The structure of MUNC lncRNA may also help to understand its function. SHAPE-MaP (selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling) identifies RNA functional motifs. The protocol uses 2'-hydroxyl-selective reagents which form a covalent 2'O-adduct with nucleotides at conformationally flexible RNA regions where the RNA is in a single-stranded form. The reverse transcriptase misreads SHAPE-modified nucleotides and incorporates mutations, which are mapped by high throughput sequencing. Sequences are aligned to the reference sequence to identify mutations, and SHAPE reactivity map is built. The map is used to model secondary structure of the transcript with the understanding that heavily mutated sites are in more single-stranded conformation, while the protected sites are either in double-strand form or protected because of their interactions with proteins. The SHAPE-MaP profile obtained from naked RNA compared to that of RNA obtained from a cell lysate with associated proteins will help distinguish sites that become protected when cellular proteins are present. The method was first used to model the structure of the HIV genome, but it is also useful for lncRNA structure modeling[26 27].

These recently developed methods will allow us to investigate the molecular mechanism of MUNC and other lncRNAs in detail. Researchers can use these complementary techniques to obtain additional confidence about the results. LncRNAs were once believed to be transcriptional noise at the time of their discovery by high throughput sequencing a few years ago. Now with technical advances and efforts from various labs, we are finally beginning to understand their functional significance.

Summary

We characterized a pro-myogenic lncRNA, MUNC, which is induced during muscle differentiation by direct interaction of MyoD to its genomic locus. Initially it was proposed that MUNC is an eRNA for MyoD, and its only role during differentiation is to induce MyoD expression. Experiments performed by us suggest that this is not a case. MUNC must work through a different mechanism rather than simple induction of its adjacent gene, MYOD. We hypothesize that MUNC targets multiple genomic sites to change their transcriptional status. To collect evidence for this, we need to perform ChIRP-seq experiment, ideally combined with CHART method to characterize MUNC-specific genomic sites. When we learn which genomic sites are directly affected by MUNC, we may start looking for transcription factors, or histone marks specific for these sites. We will try to characterize common binding motif of these sites, which may be recognized by MUNC itself or by its protein partner. We need to keep working on pull down experiments to find a specific protein partner for MUNC. After establishing MUNC overexpression system it may be easier to specifically pull it down and characterize its interactors.

It is known that MyoD, as a master transcription factor for myogenesis, works in a feed-forward mechanism. It activates other myogenic factors which bind to multiple myogenic sites to activate them. MyoD also binds to its own promoter to enhance its transcription. MUNC is also activated during the process, and probably interacts with one of the myogenic factors that are recruited to its target sites. Although MyoD starts the whole myogenic transcriptional cascade, cells are able to adapt and differentiate upon MyoD deletion, probably because of the redundancy between MyoD and Myf5. Interestingly, we did not observe such redundancy while deleting MUNC. This suggests that unlike MyoD

and other myogenic transcription factors, MUNC does not have other "MUNC-like lncRNAs" family members.

MUNC and MyoD are coded by adjacent loci and both positively regulate myogenesis. They also co-induce each other. MUNC depletion decreases MyoD level, and si-MYOD treatment decreases MUNC transcript level. These features could suggest that both genes form an eukaryotic operon-like structure[28]. In bacteria, the genes in an operon are transcribed as a common pre-mature transcript is processed to form the RNAs of the individual genes. In Eukaryotes, perhaps the only aspect of the operon that is retained is that the adjoining genes have a common function. The mature lncRNA MUNC and mRNA for MyoD protein are adjacent to each other. From Mousavi et al. study[19] we also know that Core Enhancer Region for MYOD, CER, located -20kb from MYOD TSS, produces another pro-myogenic lncRNA. Another report described a lnc-MyoD, another pro-differentiation noncoding transcript transcribed from the locus -30kb from MYOD TSS[29]. Hypothetically, all four transcripts lncMyoD, CE-RNA, MUNC, and MYOD could be part of a eukaryotic operon, simply because they are adjacent to each other and are all positive regulators of myogenesis. My PCR-based analysis showed that there is no common transcript encoding MUNC and MYOD. Also from RNA-PolII ChIP seq data we see separate peaks of RNA PolII at the four loci. Finally, from RNA-seq analysis we know that absolute abundance of protein-coding MyoD transcript is higher than the abundance of the noncoding RNAs which would suggest that lncMyoD, CE-RNA, MUNC and MYOD are transcribed as separate transcriptional units. With the present data, we cannot say that lncMyoD, CE-RNA, MUNC, and MYOD are transcribed as one transcription unit, but we know that all four transcripts are co-regulated, are closely located next to each other and

that they all function as pro-myogenic factors. It is tempting to suggest that they have co-evolved as one unit, although they are not exactly the same as a prokaryotic operon.

References

1. Asp, P. *et al.* Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc. Natl. Acad. Sci.* **108,** E149–E158 (2011).

2. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–5 (2010).

3. Mousavi, K. *et al.* ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol. Cell* **51,** 606–617 (2013).

4. Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA / mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2876–81 (2013).

5. Kong, L. *et al.* CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35,** (2007).

6. Wang, L. *et al.* CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41,** (2013).

7. Baldassarre, A. & Masotti, A. Long non-coding RNAs and p53 regulation. *International Journal of Molecular Sciences* **13,** 16708–16717 (2012).

8. Jonsson, P. *et al.* Single-Molecule Sequencing Reveals Estrogen-Regulated Clinically Relevant lncRNAs in Breast Cancer. *Mol. Endocrinol.* **29,** 1634–1645 (2015).

9. Misawa, A., Takayama, K., Urano, T. & Inoue, S. Androgen-induced Long Noncoding RNA (lncRNA) SOCS2-AS1 Promotes Cell Growth and Inhibits Apoptosis in Prostate Cancer Cells. *J Biol Chem* **291,** 17861–17880 (2016).

10. Sakurai, K., Reon, B. J., Anaya, J. & Dutta, A. The lncRNA DRAIC/PCAT29 Locus Constitutes a Tumor-Suppressive Nexus. *Mol. Cancer Res.* **13,** 828–838 (2015).

11. Reon, B. J. *et al.* Expression of lncRNAs in Low-Grade Gliomas and Glioblastoma Multiforme: An In Silico Analysis. *PLoS Med.* **13,** (2016).

12. Shukla, S. *et al.* Development of a RNA-seq based prognostic signature in lung adenocarcinoma. *J. Natl. Cancer Inst.* **109,** (2017).

13. Patterson, K. I., Brummer, T., O'brien, P. M. & Daly, R. J. Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochem. J.* **418,** 475–489 (2009).

14. Han, X. H., Jin, Y. R., Seto, M. & Yoon, J. K. A WNT/β-catenin signaling activator, R-spondin, plays positive regulatory roles during skeletal myogenesis. *J. Biol. Chem.* **286,** 10649–10659 (2011).

15. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147,** 358–369 (2011).

16. Carmichael, G. G. Regulatory non-coding RNAs: Methods and protocols. *Regul. Non-Coding RNAs Methods Protoc.* **1206,** 1–178 (2014).

17. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161,** 404–

416 (2015).

18. Leppek, K. & Stoecklin, G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* **42,** (2014).

19. Mousavi, K. *et al.* ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol. Cell* **51,** 606–617 (2013).

20. Chu, C., Quinn, J. & Chang, H. Y. Chromatin Isolation by RNA Purification (ChIRP). *Journal of Visualized Experiments* (2012). doi:10.3791/3912

21. Hu, T. *et al.* Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function. *Nucleic Acids Res.* **45,** gkx055 (2017).

22. Simon, M. D. in *Current Protocols in Molecular Biology* (2013). doi:10.1002/0471142727.mb2125s101

23. Vance, K. W. *et al.* The long non-coding RNA paupar regulates the expression of both local and distal genes. *EMBO J.* **33,** 296–311 (2014).

24. Chalei, V. *et al.* The long non-coding RNA *Dali* is an epigenetic regulator of neural differentiation. *Elife* **3,** e04530 (2014).

25. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521,** 232–236 (2015).

26. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*

11, 959–965 (2014).

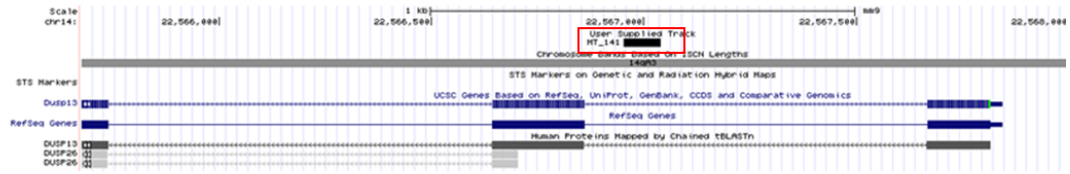
27. Smola, M. J. *et al.* SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U. S. A.* 201600008 (2016). doi:10.1073/pnas.1600008113

28. Nannapaneni, K. *et al.* Computational identification of operon-like transcriptional loci in eukaryotes. *Comput. Biol. Med.* **43,** 738–743 (2013).

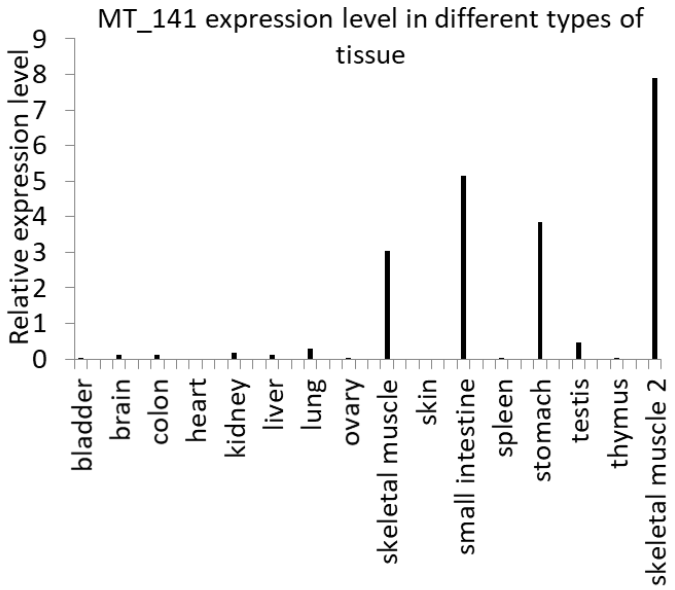
29. Gong, C. *et al.* A Long Non-coding RNA, LncMyoD, Regulates Skeletal Muscle Differentiation by Blocking IMP2-Mediated mRNA Translation. *Dev. Cell* **34,** 181–191 (2015).

30. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 167–177 (2015).

A.



B.



C.

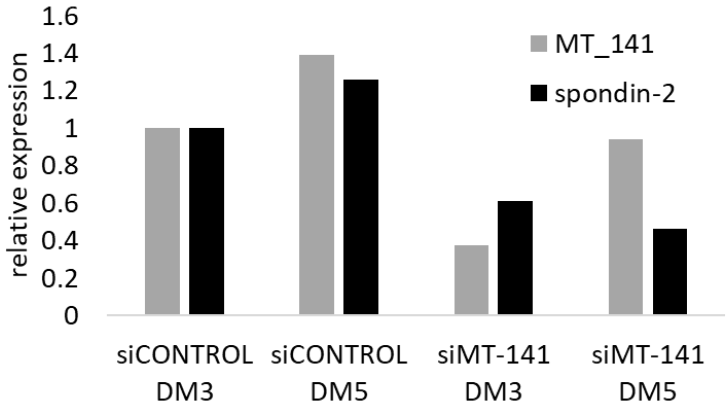


Figure 1. Legend. A. Genomic location of MT_141 locus. B. QRT-PCR analysis of MT_141 expression level in various murine tissues. Representative results. C. MT_141 knock down experiment of C2C12 cells differentiating for 3 days and for 5 days. Analysis of expression of MT_141 and spondin-2. Representative results.

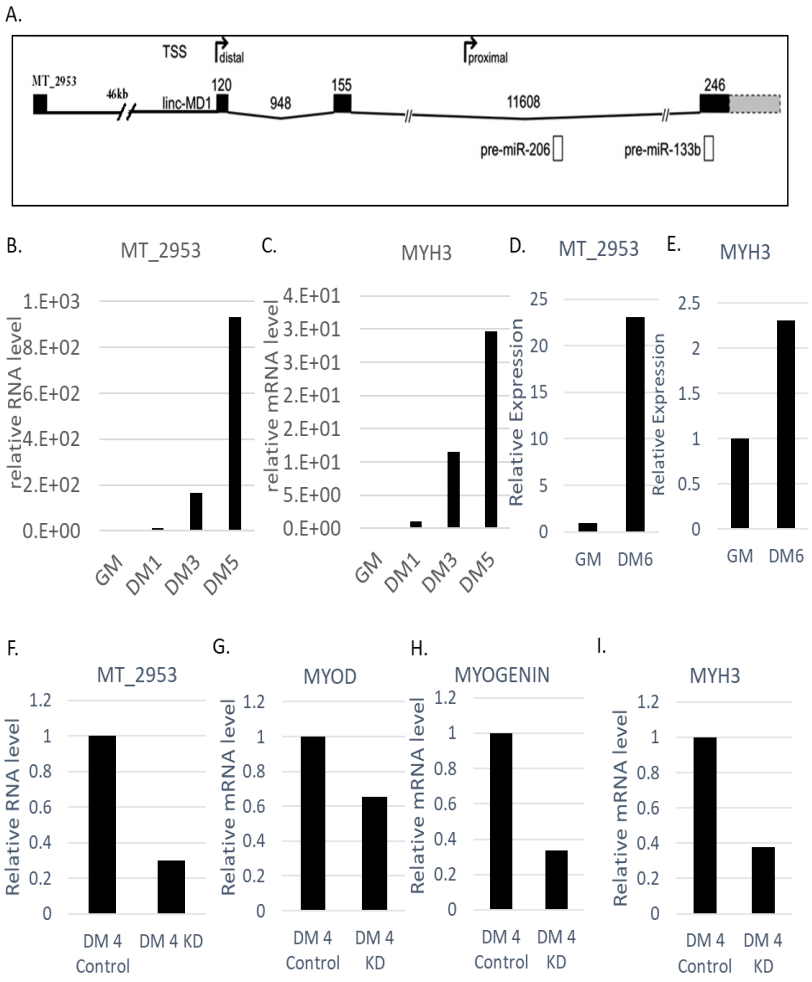


Figure 2. Legend. A. Genomic location of MT_2953 locus and its distance from another pro-myogenic lncRNA, linc-MD1. QRT-PCR analysis of differentiating C2C12 for 1, 3, 5 days showing expression of MT_2953 (B.) and expression of MYH3 (C.) Representative results (Caitlin Dutta)

QRT-PCR analysis of differentiating primary myoblasts for 6 days showing expression of MT_2953 (D.) and expression of MYH3 (E.) Representative results (Myles Anderson). MT_2953 knock down experiment of C2C12 cells differentiating for 4 days. Analysis of expression of MT_2953 (F.), MYOD (G.), MYOGENIN (H), MYH3 (I.) Representative results (Myles Anderson).

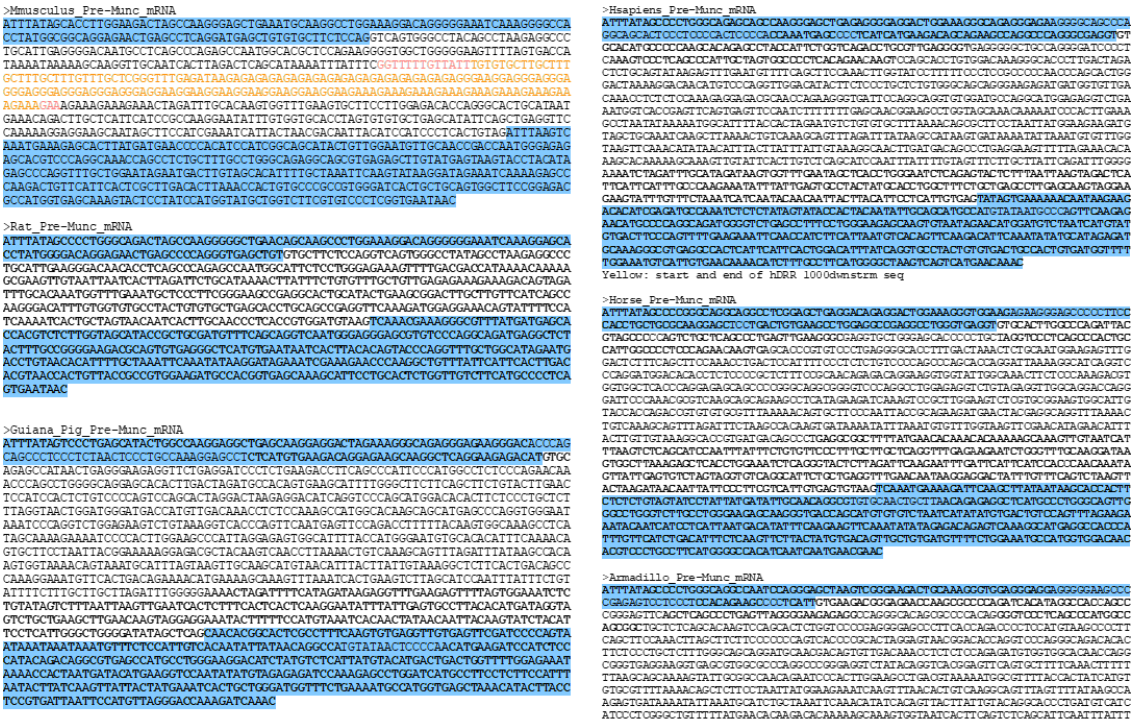


Figure 3. Legend.

Syntenic upstream from MYOD TSS region sequences from different species: mouse, rat, Guiana pig, human, horse, armadillo. Blue color highlights exons of MUNC described in mouse, and putative exons in other species. Orange color highlights highly repetitive region in intronic region of murine MUNC. Bold font shows highly conserved stretches of sequences between species (analysis performed by Dr. Yoshiyuki Shibata).

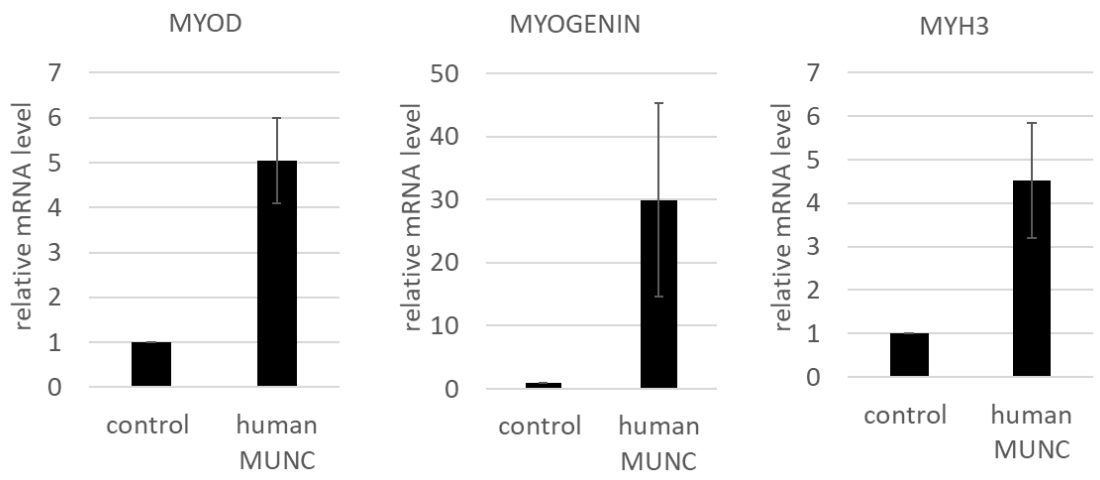


Figure 4. Legend.

Q-RT-PCR analysis of myogenic genes expression in C2C12 cells upon human MUNC overexpression under proliferating conditions. Data is normalized to GAPDH expression level and shown as relative to control cells. Values represent three biological replicates, presented as mean +/− SEM.

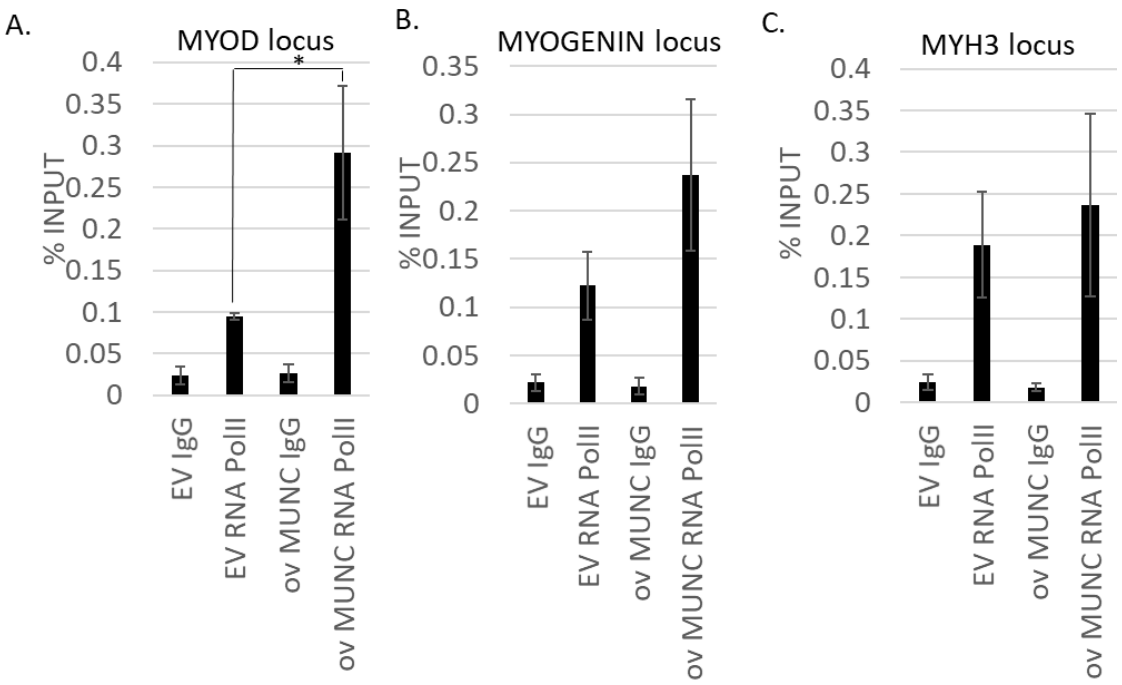


Figure 5. Legend

A-C. RNA PolII ChIP analysis in control cells (EV) at day 3 of differentiation (DM3) and cells overexpressing MUNC DM3, showing abundance of RNA PolII at MYOD genomic locus (A.), MYOGENIN genomic locus (B.), and MYH3 genomic locus (C.) Values represent three biological replicates, presented as mean +/− SEM. Statistical significance was calculated using Wilcoxon-Mann-Whitney Test. (*) indicates p-value =< 0.05.

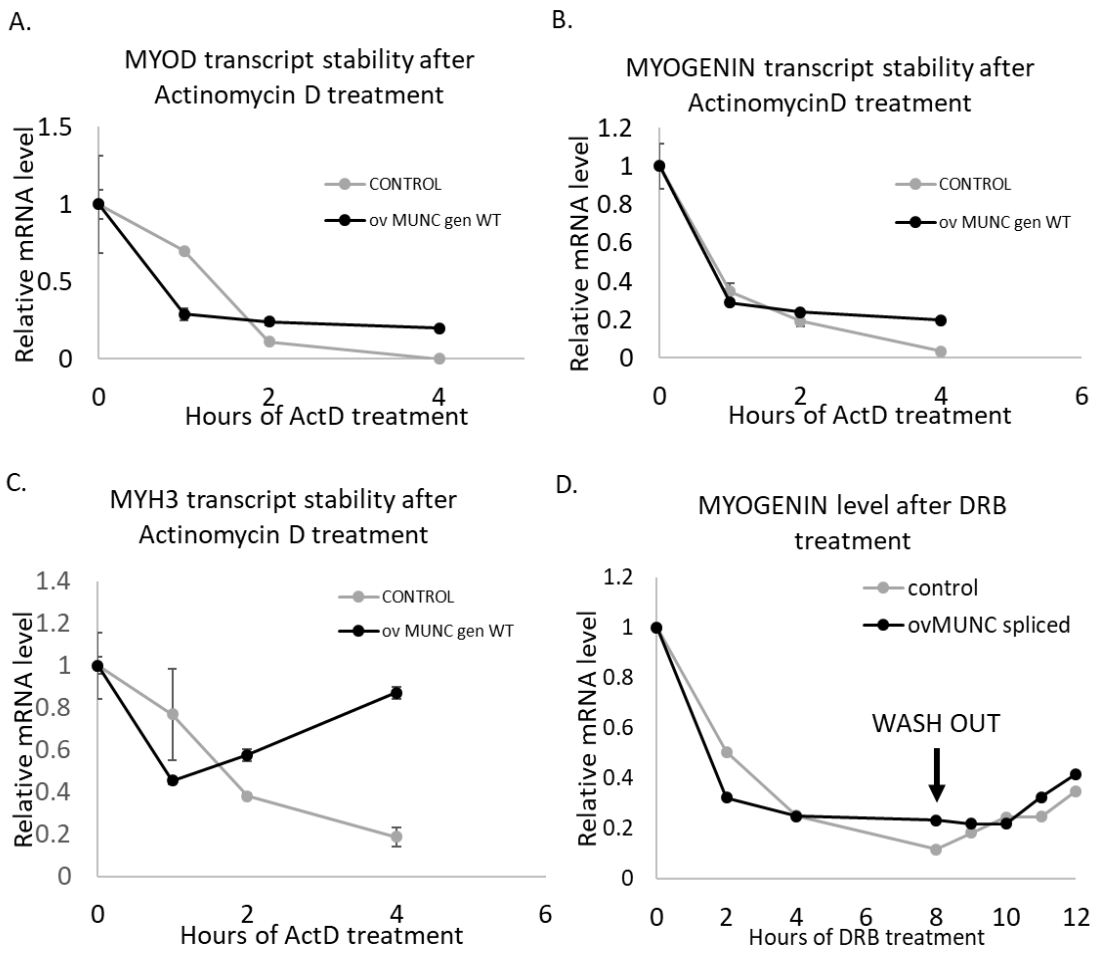
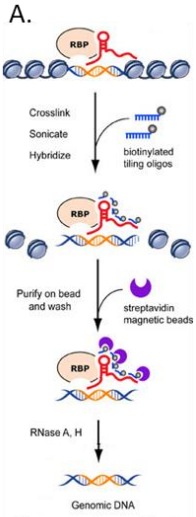


Figure 6. Legend. Representative results of relative transcript levels after Actinomycin D treatment. Contreol cells (control) and cells overexpressing MUNC unspliced form (ovMUNC gen WT) were differentiating for 3 days, and then they were treated with the drug (C=5ng/ul). Samples were collected after 1, 2, and 4 hrs of treatment. QRT-PCR was performed to measure RNA level for MYOD (A.), MYOGENIN (B.), and MYH3 (C.). D. Representative results of relative transcript level of MYOGENIN after DRB treatment. Control cells (control) and cells overexpressing MUCN spliced (ovMUNC spliced) were differentiating for 3 days. The drug was added for 8 hrs (C=25uM), and after 8 hrs cells were washed out. Samples were collected after 2, 4, 8 hrs of treatment and 1, 2, 3, 4hrs after wash out. QRT-PCR was performed to measure RNA level.

A.



Chromatin Isolation by RNA Purification

C. Chu et al. 2012

B.

coordinate:MUNC (I)	PeakScore	coordinate:MUNC (II)	GeneName	PeakScore	Average PeakScore
chr16:80250352-80250596	4732.2	chr16:80250355-80250529	Ncam2	242.8	2487.5
chr5:140173726-140173970	3830.8	chr5:140173719-140173893	Mad1l1	936.6	2383.7
chr9:32131427-32131671	4386.7	chr9:32131380-32131554	Arhgap32	138.8	2262.75
chr10:90304946-90305190	4206.4	chr10:90305048-90305222	Anks1b	138.8	2172.6
chr16:11143957-11144201	1862.8	chr16:11144008-11144182	Txndc11	1561.1	1711.95
chr3:76071452-76071696	2944.5	chr3:76071499-76071673	Fstl5	242.8	1593.65
chr7:46372053-46372297	1126.7	chr7:46372228-46372402	Myod1	919.3	1023
chr13:6038620-6038864	1337	chr13:6038641-6038815	Klf6	451	894
chr1:54957322-54957566	1292	chr1:54957217-54957391	Ankrd44	208.1	750.05
chr1:57936459-57936703	1322	chr1:57936506-57936680	Kctd18	121.4	721.7
chr15:97268592-97268836	946.4	chr15:97268753-97268927	Amigo2	190.8	568.6
chr8:17696598-17696842	886.3	chr8:17696731-17696905	Csmd1	208.1	547.2
chr9:85577379-85577623	691	chr9:85577408-85577582	Ibtk	381.6	536.3
chr18:42217526-42217770	811.2	chr18:42217595-42217769	Plac8l1	190.8	501
chr4:150739675-150739919	736.1	chr4:150739684-150739858	Slc45a1	156.1	446.1
chr8:114471089-114471333	661	chr8:114471083-114471257	Wwox	156.1	408.55
chr11:98370092-98370336	555.8	chr11:98370260-98370434	Stard3	156.1	355.95
chr6:98683431-98683675	450.7	chr6:98683414-98683588	Gm765	190.8	320.75
chr14:56473911-56474155	450.7	chr14:56473887-56474061	Rnf17	121.4	286.05
chrX:169123779-169124023	390.6	chrX:169123817-169123991	Arhgap6	156.1	273.35
chr8:128394662-128394906	315.5	chr8:128394684-128394858	Mir1903	173.5	244.5
chr9:109310015-109310259	240.4	chr9:109310037-109310211	Fbxw14	208.1	224.25
chr10:114088416-114088660	240.4	chr10:114088545-114088719	Mir669h	173.5	206.95
chr11:38927430-38927674	225.3	chr11:38927442-38927616	Gm12130	138.8	182.05
chr9:123654402-123654646	240.4	chr9:123654428-123654602	Slc6a20b	121.4	180.9
chr3:41713423-41713667	195.3	chr3:41713479-41713653	Sclt1	121.4	158.35

Figure 7. Legend.

A. Workflow of ChIRP method. B. Top hits of genomic sites interacting with MUNC according to the initial ChIRP-seq screen. Genomic sites are called by Gene Name- the adjacent gene name for specific genomic site, and their specific genomic coordinates. The table presents peak scores for each of biological runs (enrichment of MUNC specific read over lacZ reads) separately, and as an average.

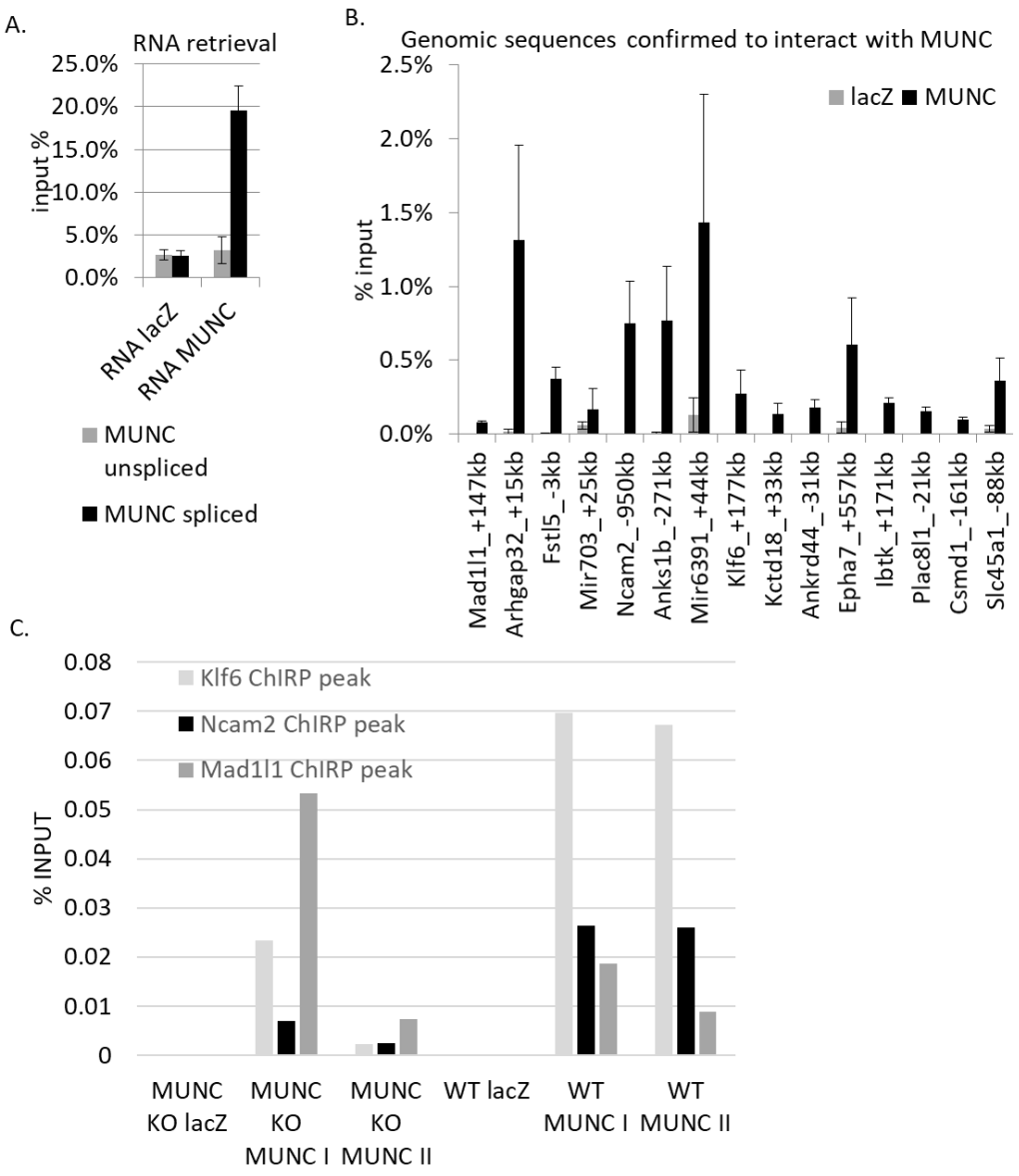


Figure 8. Legend. A. Representative results of retrieval level of MUNC after ChIRP-qRT-PCR, presented are both isoforms: unspliced MUNC, and spliced MUNC. B. ChIRP-qPCR confirmation for top hits. Peaks are called by their adjacent gene name and relative distance between a peak and TSS for a gene (- peak is located upstream from the TSS, + peaks is located downstream from the TSS. Results show 3 biological runs, with error bars representing +/- SEM. C. Representative results of ChIRP-qPCR performed on C2C12 MUNC knock out cells and on C2C12 WT cells, using two sets of MUNC specific oligos (I and II).

Appendix

Scientific contribution to another study

In addition to my main project I made a contribution to the study listed below:

Regulation of several androgen-induced genes through the repression of the miR-99a/let-7c/miR-125b-2 miRNA cluster in prostate cancer cells.

Sun D, Layer R, Mueller AC, **Cichewicz MA**, Negishi M, Paschal BM, Dutta A.

Oncogene. 2014 Mar 13;33(11):1448-57. doi: 10.1038/onc.2013.77.

Abstract

The androgen receptor (AR) stimulates and represses gene expression to promote the initiation and progression of prostate cancer. Here, we report that androgen represses the miR-99a/let7c/125b-2 cluster through AR and anti-androgen drugs block the androgen-repression of the miRNA cluster. AR directly binds to the host gene of the miR-99a/let7c/125b-2 cluster, LINC00478. Expression of the cluster is repressed or activated by chromatin remodelers EZH2 or JMJD3 in the presence or absence of androgen, respectively. Bioinformatics analysis reveals a significant enrichment of targets of miR-99a, let-7c and miR-125b in androgen-induced gene sets, suggesting that downregulation of the miR-99a/let7c/125b-2 cluster by androgen protects many of their target mRNAs from degradation and indirectly assists in the gene induction. We validated the hypothesis with 12 potential targets of the miR-99a/let7c/125b-2 cluster induced by androgen: 9 out of the 12 mRNAs are downregulated by the microRNA cluster. To ascertain the biological significance of this hypothesis, we focused on IGF1R, a known prostate cancer growth

factor that is induced by androgen and directly targeted by the miR-99a/let7c/125b-2 cluster. The androgen-induced cell proliferation is ameliorated to a similar extent as anti-androgen drugs by preventing the repression of the microRNAs or induction of IGF1R in androgen-dependent prostate cancer cells. Expression of a microRNA-resistant form of IGF1R protects these cells from inhibition by the miR-99a/let7c/125b-2 cluster. These results indicate that a thorough understanding of how androgen stimulates prostate cancer growth requires not only an understanding of genes directly induced/repressed by AR, but also of genes indirectly induced by AR through the repression of key microRNAs.

My contribution (all experiments were performed under close supervision of a senior graduate student at that time, Dandan Sun):

- I analyzed pri-miR99a/let7c/125b cluster level in LNCaP cells treated with androgen, and after addition of androgen receptor inhibitors: Bicalutamide and Flutamide to confirm that listed miRs are inhibited specifically by androgen receptor activation – Fig.2. E.
- I performed knock down of JMJD3 (H3K27 demethylase) in cells which were not treated or which were treated with androgen to discover that JMJD3 is required for therepression of the three pri-miRs by androgen– Fig.3. D.
- I performed proliferation assay on Du145 cells, after addition of androgen and miR-99a/let7c/125b to show that this miRs cluster did not suppress cell growth in androgen-independent cell line- Du145 – Fig S4. D.