UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biologia

SCUOLA DI DOTTORATO DI RICERCA IN: Bioscienze e Biotecnologie

INDIRIZZO: Genetica e Biologia Molecolare dello Sviluppo

CICLO XXVIII

THE ROLE OF LONG NON-CODING RNAS IN PATHOPHYSIOLOGICAL CONDITIONS OF SKELETAL MUSCLE

Direttore della Scuola: Ch.mo Prof. Giuseppe Zanotti

Coordinatore d'indirizzo: Ch.mo Prof. (nome e cognome)

Supervisore: Ch.mo Prof. Gerolamo Lanfranchi

Co-supervisore: Dott. Stefano Cagnin

Dottorando: Enrico Alessio

Table of Contents

Abstract (English)

Background. Long noncoding RNAs (lncRNAs) form an abundant class of transcripts that is acquiring an increasing importance due to the roles that some of these molecules have been demonstrated to exert in various biological processes. Unfortunately, the function of the majority of them remains elusive. LncRNA are involved in different epigenetic mechanisms of gene regulation, from chromatin remodelling to post-transcriptional regulation. Since their function is related to their subcellular localization and expression, a genome-wide approach to determine the localization of lncRNAs could greatly improve our understanding of their role in pathophysiology of skeletal muscle.

Results. To investigate the subcellular localization of lncRNAs expressed in single skeletal muscle fibers, we studied their compartmentalization by coupling microarray technique with differential extraction and analysis of nuclear and cytoplasmic RNA. We found that 481 lncRNAs preferentially localize in the nucleus, 655 in the cytoplasm and 297 show an alternate localization depending on muscle type and/or lncRNA isoform. Microarray based lncRNA subcellular localization was validated by Fluorescent In Situ Hybridization (FISH) on 6 lncRNAs. All FISH tests confirmed the subcellular localization previously evidenced by the microarray approach. We then chose 32 lncRNAs to investigate their expression profile among different tissues. Of this group, 24 appear to be preferentially expressed in skeletal muscle, 5 in the heart, and only 2 have a high expression in the liver. Skeletal muscle can be approximately classified according contractile capacity as fast and slow, reflecting fiber composition and metabolism. For this reason, we tested the group of 32 lncRNAs for expression in fast or slow myofibers using single-cell analysis, evidencing that 9 lncRNAs have a specific fiber expression probably affecting metabolic traits. To identify lncRNAs involved in skeletal muscle atrophy we also investigated their expression in skeletal muscle during denervation and starvation as well as in a mouse model for Amyotrophic Lateral Sclerosis. We then analysed the expression of the same lncRNAs in proliferating and differentiating C2C12 myoblasts both in purified myonuclei and cytoplasm. 20 lncRNAs presented with differential expression during myogenic differentiation.

Conclusions. We localized several lncRNAs inside the skeletal muscle evidencing their preferential expression in this tissue and also in specific myofibers. Moreover, we evidenced that many lncRNAs expressed in the muscle are involved in skeletal muscle atrophy induced by different models and in its differentiation. Finally, we evidenced that lncRNA Pvt1 is able to shuttle between cytoplasm and nucleus during myoblast differentiation *in-vitro*.

Abstract (Italiano)

Background. La classe di trascritti composta dai long non coding RNA (*lncRNA*) sta acquisendo una crescente importanza a causa dei molteplici ruoli che alcune di queste molecole hanno dimostrato di avere in vari processi biologici. Sfortunatamente, la funzione della maggior parte di questi RNA rimane al momento sconosciuta. I *lncRNA* sono coinvolti in diversi meccanismi epigenetici di regolazione genica, dal rimodellamento della cromatina alla regolazione post-trascrizionale. Dal momento che la loro funzione è legata alla loro espressione e localizzazione subcellulare, usare un approccio *genome-wide* per determinare la localizzazione di *lncRNA* potrebbe migliorare notevolmente la nostra comprensione del loro ruolo nella fisiopatologia del muscolo scheletrico.

Risultati. Per studiare la localizzazione subcellulare dei *lncRNA* in singole fibre del muscolo scheletrico abbiamo estratto RNA nucleare e citoplasmatico per analizzarlo mediante la tecnica microarray. Abbiamo identificato 481 *lncRNA* che localizzano preferenzialmente nel nucleo, 655 nel citoplasma e 297 che localizzano in diversi compartimenti a seconda dell'isoforma scelta o del tipo di muscolo analizzato. La localizzazione subcellulare identificata via microarray è stata validata, per 6 *lncRNA*, mediante *Fluorescent In Situ Hybridization* (FISH). Questi esperimenti hanno confermato la localizzazione subcellulare precedentemente identificata con l'analisi *microarray*. Successivamente abbiamo scelto 32 lncRNA di cui abbiamo studiato il profilo di espressione in diversi tessuti. Di questo gruppo, 24 risultano essere espressi preferenzialmente nel muscolo scheletrico, 5 nel cuore, e 2 nel fegato. Il muscolo scheletrico può essere caratterizzato come veloce o lento a seconda della velocità di contrazione, questa caratteristica è dovuta alla presenza di fibre muscolari che sfruttano un metabolismo ossidativo (contrazione lenta) o glicolitico (contrazione veloce). Per questo motivo, abbiamo analizzato l'espressione degli stessi 32 *lncRNA* in singole miofibre veloci e lente. Grazie a questa analisi abbiamo evidenziato 9 lncRNA che hanno un'espressione fibra-specifica e sono possibilmente legati al metabolismo. Per identificare lncRNA coinvolti nell'atrofia del muscolo scheletrico abbiamo studiato la loro espressione nel durante denervazione e denutrizione. In aggiunta abbiamo analizzato la loro espressione in un modello murino di sclerosi laterale amiotrofica (ALS). Abbiamo poi analizzato l'espressione degli stessi *lncRNA* in colture di mioblasti C2C12, in RNA estratto differenzialmente da nucleo e citoplasma. Abbiamo individuato 20 *lncRNA* differenzialmente espressi durante il differenziamento miogenico.

Conclusioni. Abbiamo individuato la localizzazione subcellulare di diversi *lncRNA* nel muscolo scheletrico, evidenziando la loro espressione preferenziale in questo tessuto e individuando in che tipo di fibra sono più espressi. Inoltre, abbiamo evidenziato che l'espressione di molti *lncRNA* muscolari è alterata durante il differenziamento dei mioblasti e durante l'atrofia indotta da diversi modelli. Infine, abbiamo evidenziato che la localizzazione del *lncRNA* Pvt1 cambia da citoplasma a nucleo durante il differenziamento dei mioblasti in vitro.

Chapter 1 – Introduction

1.1 Long non-coding RNAs (LncRNAs)

Large-scale sequencing projects and genomic tilling arrays have provided new unexpected and intriguing information about the organization and function of the mammalian genomes. Transcription in mammals is pervasive throughout the genome and not limited to protein-coding regions, giving rise to a complex network of transcripts, most of which have little or no coding capacity. This class of transcripts is referred to as non-coding RNAs (ncRNAs)[1].

The organism complexity among species better correlates with the portion of genome that is transcribed into non-coding RNAs than with the number of protein-coding genes, even when diversification by both alternative splicing and post-translational regulation are taken into account. This suggests that RNA-based regulatory mechanisms played a relevant role in the evolution of the complex developmental mechanisms in eukaryotes, highlighting the importance of RNA-based control in the evolution of multicellular organisms[2]. RNA is in fact, a preferred substrate for epigenetic dynamics because it can rapidly shift between multiple stable conformations and undergo allosteric transitions, acting as a responsive switch[3].

Non-coding RNAs below 200 bp are defined Small ncRNA and include transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNAs (snRNA), small nucleolar RNAs (snoRNA), microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs).

Non-coding transcripts longer than 200 bp are instead defined as long non coding RNAs[4].

Transcriptional regulation of loci includes both coding and non-coding transcripts with lncRNAs representing a sizable expansion of the transcriptome. It is difficult to gauge an exact number of lncRNAs. However, whereas the number of known protein-coding genes has remained stable over recent years, the number of known lncRNAs continues to increase.As a result of these studies, lncRNAs are being increasingly accepted as a new gene class[3].

Taken together, these findings suggest that non protein-coding sequences house a set of information-rich instructions, many of which are likely to be regulatory in nature and have contributed to the increase of biological complexity[5].

1.1.1 Evolution of a non-coding world

Most non-coding RNA genes are subjected to a low degree of evolutionary constraints and are thus diverging rapidly. It is therefore difficult to depict the evolutionary scenario that led to the emergence of a functional non-coding gene. Different hypothesis can explain the underlying mechanisms.

The first scenario is the emergence of a non-coding RNA gene from a protein-coding gene. This conversion involves the loss of protein-coding capacity and the acquisition of a newly function associated with the presence of the RNA in the cell. This scenario is possible taken into account the great number of pseudogenes present in eukaryotic genomes. Many of them can be transcribed and play functional roles.

Another possibility is the emergence of non-coding RNAs following the insertion of transposable elements[1]. The tissue specificity of long non-coding RNAs has been attributed to the presence of transposable elements that are embedded in the vicinity of lncRNA transcription start sites. This relevance is supported by the discovery that, in vertebrates, transposable elements are found in more than two-thirds of mature lncRNAs, whereas they are rarely found in protein-coding transcripts[2].

A third mechanism is represented by the *de novo* emergence of non-coding RNA genes, for example, from the transcriptional activation of a large intergenic region containing pre-existing cryptic splice sites[1].

As a final hypothesis, the evolution of non-coding genes could result from a combination of different mechanisms, as illustrated by the evolutionary history of the X-inactivation center[5].

1.1.2 Long non-coding RNAs at every stage of gene expression

The role of lncRNAs in genomic imprinting is the most characterized. Genomic imprinting in mammals results in a monoallelic expression pattern of a subset of genes due to the parental origin. Most of these genes are located in clusters that are regulated through long non-coding RNAs (a well-known example is *H19/Igf2*). Imprinted gene expression is conferred through lncRNAs. This is supported by the evidence that the majority of imprinted clusters contain at least one lncRNA[6].

LncRNAs have emerged as key regulators of gene expression at multiple levels controlling different aspects of cell fate[4]. Moreover, lncRNAs are involved in stem cell maintenance, environment-responsive programs[2], and also contribute to

disease[4]. LncRNAs are novel "fine tuners" of gene regulation[2]: their expression pattern is developmental stage- and cell type- dependent[4] and their spatial-temporal expression is specific across various stages of differentiation[2]. Moreover, lncRNAs help to orchestrate the epigenetic transcriptional state[7] at almost every stage of gene expression[3]. These findings reinforce their biological role and highlight them as central regulators in a very plastic tissue such as skeletal muscle that undergoes continuous adaptive responses.

1.1.3 Biogenesis of long non-coding RNAs

Derrien and colleagues compared lncRNAs features to those of protein-coding[8] RNAs, to evidence similarities and discrepancies with most known coding genes.

They evidenced that lncRNAs are expressed at lower level than protein-coding transcripts and their pattern of expression is more developmental stage and cell type specific[4].

Analyses on lncRNAs maturation indicate that they are generated through pathways similar to that of protein-coding genes, but as independent transcriptional units, even if lncRNAs loci are often in close association with protein coding genes. LncRNAs are defined as transcript with little or no coding capacity (Figure 1) and some lncRNAs are preferentially processed into small RNAs such as miRNAs or snoRNAs[8]. It was evidenced that small polypeptides (micropeptides) can origin from lncRNAs impacting on muscle performance[9]. Therefore, a more reasonable approach may be to use a definition of non-coding that focuses on a coding-independent functional role of the RNA transcript; a lncRNA must function as a RNA transcript, whether or not it may also code for polypeptide[10]. For example, it could maintain the ability of coding micropeptides in some circumstances and in other to function as decoy for miRNAs or interact with other proteins.

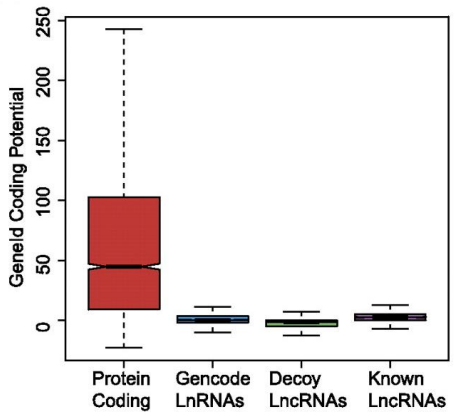


Figure 1. **Most lncRNAs lack any protein-coding potential.**

To test whether lncRNAs are able to code for a protein and to find the longest ORF in lncRNAs it was used GeneID algorithm. Results were compared with known protein-coding transcripts (red), experimentally validated lncRNAs such as XIST and H19 (blue), and a set of ''decoy'' lncRNAs (purple), obtained by mapping lncRNA gene structures randomly onto the genome. LncRNAs have no coding potential. Image modified from Derrien *et al.*[8].

Most lncRNAs are transcribed by RNA polymerase II and many share mRNA-like features such as 5'cap, poly-A tail and splicing sites; alternatively, non-polyadenylated lncRNAs are likely generated by RNA polymerase III[4].

In GENCODE database[11], that gets information from the ENCODE project[12] and ENSEMBL database[13], an unusual exonic structure is described for lncRNAs. A lot of lncRNAs are spliced (98%) using similar splicing signals used for coding transcripts, but lncRNAs show a striking tendency to have only few exons.

Even if lncRNA exons are slightly longer than those of protein-coding RNAs, overall lncRNA transcripts are shorter than protein-coding because there are a minor number of exons.

Analyses for histone methylation and acetylation patterns around TSSs (transcription start sites) of lncRNAs and coding transcripts show that histone profiles are similar for several active histone marks (H3K4me2, H3K4me3, H3K9ac, H3K27ac), but have slightly excess levels of other marks associated with both chromatin silencing (H3K27me3) and activation (H3K36me3) in TSSs of lncRNA. This information is important because histone modifications are key players in the regulation of gene expression.

1.1.4 LncRNA structure and biochemistry

A peculiar feature of lncRNAs is the propensity to fold into thermodynamically stable secondary and high-ordered structures. RNA has then the capacity to form double helices, hairpin loops, bulges and pseudoknots that are connected in high order tertiary interactions primarily mediated by non-canonical Watson-Crick base-pairing[3]. RNAs have flexible and modular scaffold nature that creates a complex structural landscape in lncRNAs and, in many cases, the secondary structures dictate their function because of the presence of functional domains.

The secondary structures generally fold initially and independently, before subsequent tertiary interactions occur, resulting in the hierarchical assembly of RNA structure. It is important to understand the biochemical features of RNA from which secondary structures derive[3].

As shown in Figure 2, four functional domains can be present in lncRNAs:

1. *RNA binding domains*. Thanks to the ability of pairing with other RNAs, lncRNAs can act as highly specific sensors of mRNAs, miRNAs and other lncRNAs. Small and often imperfect regions of nucleotide complementarity

are sufficient for specific interactions. This imperfect complementarity has the advantage of allowing multiple RNA agonists to compete for binding, allowing lncRNAs to sense different RNA expression signals in the cells.

2. *Protein binding domains*. Proteins are the major partners of lncRNAs, with complexed ribonucleoprotein (RNP) particles acting as chaperones, transport aids or effectors. Proteins-RNA interactions tend to result in conformational changes of protein structure, RNA structure or both. The structural variety of RNA in combination with the abundance of RNA-binding proteins provides a broad interface for communication between the proteome and the transcriptome.

3. *DNA binding domains*. There are few evidences for direct interaction between lncRNAs and DNA. RNA-DNA hybrids or triplex structure can allow single strands of RNA to interact with DNA duplexes by base-pair interactions. These direct RNA-DNA interactions could efficiently and selectively target RNA signals to genomic loci. Rather than being involved in direct complementarity, RNA folds may create a DNA binding pocket as a protein transcription factor or facilitate protein localization in specific genomic domains. An example of this property is the CRISPR RNA (which can recruit Cas9 endonuclease to target specific sequence in the genome.

4. *Conformational switch*. LncRNAs can act as regulatory devices by allosterically coupling binding domains and thereby activate or suppress linked functional domains. Examples of these functional domains can be A- and C-repeats deriving from the insertion and remodeling of repetitive elements.

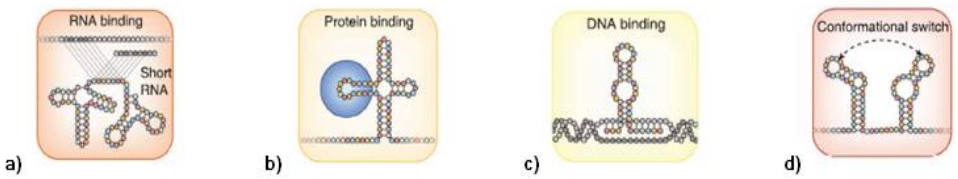


Figure 2. **Domain architecture of lncRNAs.**

LncRNAs contain structural domains that can sense or bind **a)** other RNAs via complementary base pair interactions, **b)** proteins and **c)** DNA that can induce **d)** allosteric conformational changes to other structures in the lncRNA. Image modified from Mercer *et al.*[3].

1.1.5 Classification of lncRNAs

As seen before, non protein-coding RNAs (ncRNAs) are divided into small and long non-coding. LncRNAs can be further classified based on their genomic localization with respect to protein-coding genes in *genic* and *intergenic* lncRNAs[8,14]:

- Genic lncRNAs, are divided in
  a) *Exonic lncRNAs:* lncRNAs that overlap exonic regions of a protein-coding gene;
  b) *Intronic lncRNAs:* lncRNAs that overlap intronic regions of a protein-coding gene, but do not intersect any exons;
  c) *Overlapping lncRNAs:* lncRNAs that overlap a protein-coding gene in the sense or antisense direction.

- Intergenic lncRNAs (*lincRNAs)* are lncRNAs located between annotated protein-coding genes, generally in their close proximity, but do not overlap in any way with annotated transcriptional units.

Figure 3 shows how the GENCODE consortium has defined different classes of lncRNAs.

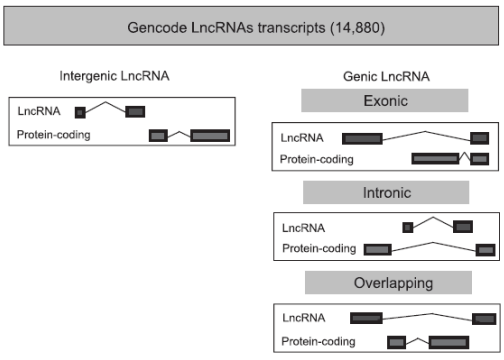


Figure 3. **Classification of lncRNAs.**

Classification of lncRNAs based on intersection with protein-coding genes by GENCODE consortium. Image modified from Derrien *et al.*[8].

LncRNAs can epigenetically regulate gene expression by recruiting specific methylases or demethylases, acethylases or deacethylases to chromatin. For this reason, it would be expected that lncRNAs should be preferentially localized in the nucleus, in contrast to protein-coding mRNAs that are trafficked to the cytoplasm for translation[8]. This is not always true, since a substantial portion of lncRNAs reside within, or are dynamically shuttled to the cytoplasm, where they regulate protein localization, mRNA translation and stability[3]. Based on their site of action, lncRNAs can be also divided into two functional categories: *nuclear* and *cytoplasmic* lncRNAs[4].

Nuclear lncRNAs have their site of action in the nucleus. They can be further subdivided in: a) cis-*acting RNAs* and b) *trans-acting RNAs* with respect to the target site.

Cis-acting RNAs work in proximity to their site of transcription and can sometimes spread their effect to long distances on the same chromosome[2,4] (Figure 4).

Trans-acting RNAs, instead, operate at distant loci[2,4] (Figure 4).

Both can activate or repress transcription, but the targeting mechanism is still far from being understood. Different hypothesis can explain lncRNA roles in the nucleus.

A lncRNA can recognize its target directly, a) by the recruitment of bridging proteins, b) by the formation of an RNA-DNA triplex or c) through DNA recognition by RNA structures[2]. Alternatively a lncRNA indirectly can, a) act as a decoy to sequester transcription factors, b) allosterically modulate regulatory proteins, or c) alter nuclear domains and long-range three-dimensional chromosome structure[2].

LncRNAs can also guide chromatin remodels and modifiers to specific genomic loci. For example, the formation of heterochromatin to induce transcriptional repression occurs with the recruitment of DNA methyltransferase 3 (DNM3) and Histone H3 lysine 9 (H3K9) methyltransferases. Transcription activation indeed occurs by activation of MILL1 complex (a multiprotein complex that mediates both histone H3 trimethylation at lysine 4 (H3K4me3) and histone H4 acetylation at lysine 16 (H4K16ac) and by activation of specific enhancer regions through changes in the three-dimensional chromatin conformation[2].

It is known that lncRNAs are able to recruit or prevent the binding of transcriptional machinery and transcription factors, thus modifying gene transcription. An example is represented by enhancer RNAs (eRNAs), encoded by extragenic enhancer regions that promote transcription of surrounding genes.

LncRNAs also participate in co- and post-transcriptional regulation: they can interact with the splicing machinery or directly with a nascent mRNA to guide particular splicing events.

Finally, lncRNAs can shape the subnuclear architecture in different ways: a) regulating chromosome looping, favoring or disrupting chromosomal interactions and b) acting as structural scaffold for the formation and regulation of nuclear compartments such as speckles, paraspeckles and Polycomb bodies[4].

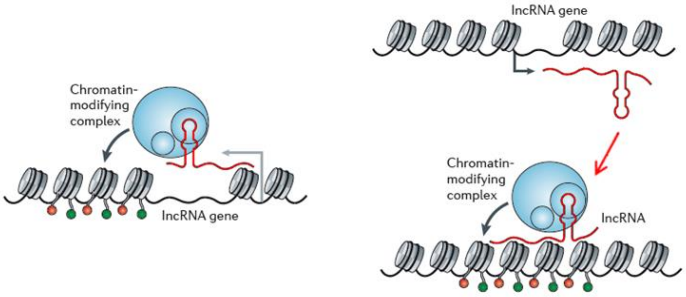


Figure 4. **Mechanisms of action of lncRNAs that guides chromatin modifying complexes.**

On the left, a cis-acting lncRNA, on the right, a trans-acting lncRNA. Image modified from Fatica *et al.*[2].

Cytoplasmic lncRNAs have their site of action in the cytoplasm. LncRNAs that work in this compartment, recognize their target by base pairing thanks to sequence complementarity[2]. Cytoplasmic lncRNAs influence translational output in two main ways. *Firstly*, they can modulate the translational rate by regulating polysome loading to an mRNA molecule or through the control of internal ribosomal entry sites (IRES); *secondly*, they can regulate gene expression by reducing or stimulating mRNA decay[4]. An example of the first case, is the rapid response to environmental changes, mediated by UCHL1 expression without de novo RNA synthesis.

The expression of UCHL1 protein cannot take place by its normal cap-dependent manner because of stress-induced inhibition of mTOR activity. In this case, the lncRNA *Uchl1-as1* (antisense transcript to the neuron specific *Uchl1* gene) is exported in the cytoplasm.

Through a base pairing with the Uchl1, *Uchl1-as1* stimulates Uchl1 cap-independent translation thanks to an inverted SINEB2 element in *Uchl1-as1* (Figure 5)[2].

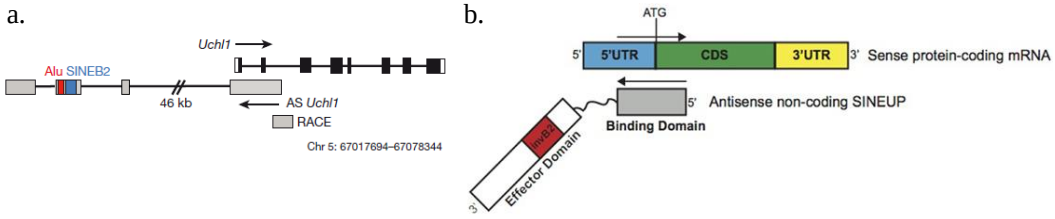


Figure 5. **Uchl1/antisense (AS) Uchl1 genomic organization and mechanism of action.**

a. *Uchl1* exons are in black; 3' and 5' UTRs are in white; antisense *Uchl1* exons are grey; repetitive elements are in red (Alu) and blue (SINEB2). Introns are indicated as lines. b. The interaction between protein coding RNA and antisense non-coding RNA presenting an inverted SINEB2 element allows the translation of protein coding, augmenting the protein content without affecting mRNA abundance. Image modified from Carrieri *et al.*[15].

An example of mRNA decay influenced by lncRNAs is represented by Staufen-mediated mRNA dacay (SMD). When a SINE element in 3' UTR of a protein-coding RNA forms intermolecular base pairing with a partially complementary SINE of one or more lncRNAs, the resulting double-stranded RNA can be recognized by Staufen 1 and Staufen 2 and degraded through SMD (Figure 6).

Another interesting mechanism of action is the hybridization to miRNAs. Thus lncRNAs can act as competing endogenous RNAs (ceRNA). Since they can act as molecular sponges for miRNAs, they are able to modulate the repressive activity of miRNA on their mRNA targets[4] creating a regulatory circuit where coding and non-coding RNAs cross-talk to each other (Figure 6)[2].

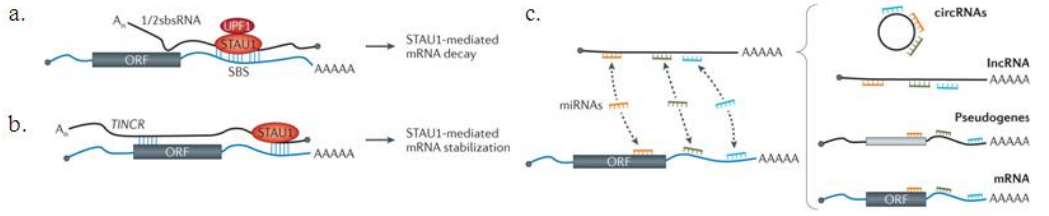


Figure 6. **Staufen mediated decay (SMD) and action of competing endogenous RNAs (ceRNAs).**

a. SMD is induced when intermolecular base pairing is formed between an Alu element (or short interspersed element (SINE) in mice) in the 3′ UTR of the mRNA and an Alu/SINE element within a long half-STAU1-binding site RNA (1/2sbsRNA). This mRNA decay mechanism also involves the RNA helicase up-frameshift 1 (UPF1). b. By contrast, STAU1-mediated mRNA stabilization has been described in the case of tissue differentiation-inducing non-protein coding RNA (*TINCR*). c. Antisense recognition has been shown to also control translation. LncRNAs can act as miRNA sponges, thus favouring the expression of the mRNAs targeted by the sequestered miRNA. As lncRNA also circular RNAs (circRNAs) can function as miRNA sponges. Image modified from Fatica *et al.*[2].

1.2 Muscular system

The muscular system is not only devoted to support, protect and move the body, but it also has an important role in body metabolismwith its capacity of adapting to different conditions like changes in mechanical demands, or food intake.

1.2.1 Three main types of muscles

Muscle cells differentiate along one of three main pathways to form skeletal, cardiac or smooth muscles.

Skeletal muscle forms the bulk of the muscular tissue of the body and consists of parallel bundles of long, multinucleated fibers. It is capable of powerful contraction because of the regular organization of contractile proteins that occupy the cytoplasm of muscle cells. Skeletal muscle is referred to as voluntary muscle, because their contraction is often initiated under conscious control. However this is a misleading term: skeletal muscle is involved in many movements, such as breathing, mainly or exclusively driven at an unconscious level.

Cardiac muscle is found only in the heart and in the walls of large veins where they enter the heart. It consists of a branching network of individual cells that are linked electrically and mechanically to function as a unit. It is much less powerful but more resistant to fatigue compared to skeletal muscle. Cardiac muscle differs structurally and functionally from skeletal muscle: it is intrinsically capable of rhythmic contraction, with a rate and strength that are responsive to hormonal and autonomic nervous control.

Smooth muscle is found in many different systems of the body: gastrointestinal, respiratory, urinary, reproductive tracts, and in the dermis. The elongated cells are tapered at the ends and are smaller than those of striated muscle. They are capable of slow but sustained contractions and, although the contraction is less powerful, the amount of shortening can be much greater. Smooth muscle is involved in the regulation of the size of lumen and the consequent movement of the luminal content. This type of muscles is excited by autonomic nerve fibers, or a blood-borne neuro-hormone; none is under conscious control, and for this reason it is referred as involuntary muscle.

Both skeletal and cardiac muscles may be called striated muscles, because myosin and actin filaments are organized into regular, repeating elements (the sarcomeres) which give to the cells a finely cross-striated appearance when observed at the microscope. Smooth muscle, in contrast, lacks such repeating elements and thus present no striations.

1.2.2 Skeletal muscle

Skeletal muscles are composed by an heterogeneous collection of cell types: myofibers, fibroblasts, endothelia cells, blood cells, nerves and many other. All together these different cells allow for the wide variety of capabilities that skeletal muscles of different body districts display[16].

1.2.2.1 Anatomy of skeletal muscle

1.2.2.1.1 Macroscopic Structure

Cellular units of skeletal muscle are enormous, multinucleated muscle fibers, that are cylindrical structures ranging from 10 to 100 µm in diameter and from millimeters to many centimeters in length. Along an entire fiber, nuclei can be several hundred. The cytoplasm of each fiber, the sarcoplasm, is surrounded by a plasma membrane called sarcolemma. Between the sarcolemma and the surrounding basal lamina, myogenic satellite cells (muscle stem cells) reside. Each fiber consists in an ensemble of myofibrils, long narrow structure (1-2 µm in diameter) forming the bulk of the sarcoplasm; myofibrils develop in turn by the fusion of individual myoblasts. When a lot of muscle fibers join together they form a single fascicle, and the union of multiple fascicles forms a muscle (Figure 7).

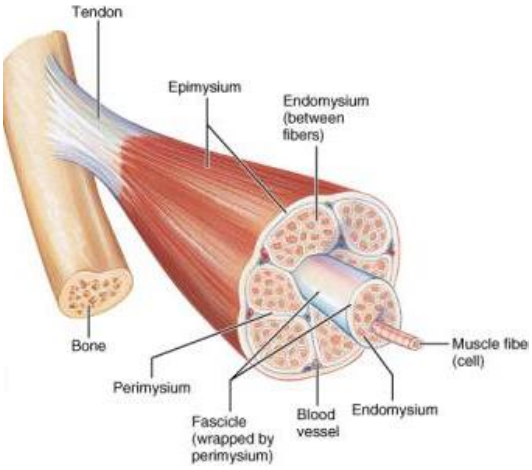


Figure 7. **Organizational levels of skeletal muscle.**

Different levels of organization in skeletal muscle are represented. From inside to out: muscle fiber, fasciculo, entire muscle. For each level is also shown the respective sheaths of connective tissue that surround it: endomysium, perimysium, epimysium. Image from[17].

Epimysium, perimysium and endomysium (Figure 7) are sheaths of connective tissue prevalently made of collagen that respectively surround an entire muscle, a single fascicle, and a single fiber. They also connect muscle to adjacent structures such as tendons. In particular, endomysium is a delicate network of connective tissue that forms the immediate external environment of a muscle fiber. It is the site of metabolic exchange between muscle and blood, and contains capillaries and bundles of small nerve fibers.

1.2.2.1.2 Microscopic Structure

Skeletal muscle, as stated before, is a striated muscle because of its protein organization. Z-discs (*Zwischenscheiben* discs), streaks of proteins, divide the myofibril into a linear series of repeating contractile units, the sarcomeres, each of which is typically 2.2 μm long in resting muscle. Sarcomeres consist of two types of filaments, thick and thin, organized into regular arrays. Thick filaments are composed of myosin while thin filaments are mainly composed of actin. The array of thick and thin filaments form a partially overlapping structure that under an electron microscope appears as a pattern of alternate dark and light bands: the banded appearance of individual myofibrils is a function of the regular alternation of the thick and thin filament array as shown in Figure 8. The Z-discs are contained in the I band (isotropic at the polarized light) made mainly of actin filaments. The other band is the dark A band (anisotropic at the polarized light) made of both actin and myosin filaments interposed one another. At the center of the A band there is a smaller band called H band represented by the myosin filaments. The M line, made by proteins that interconnect myosin filaments, is present at the center of the H band.

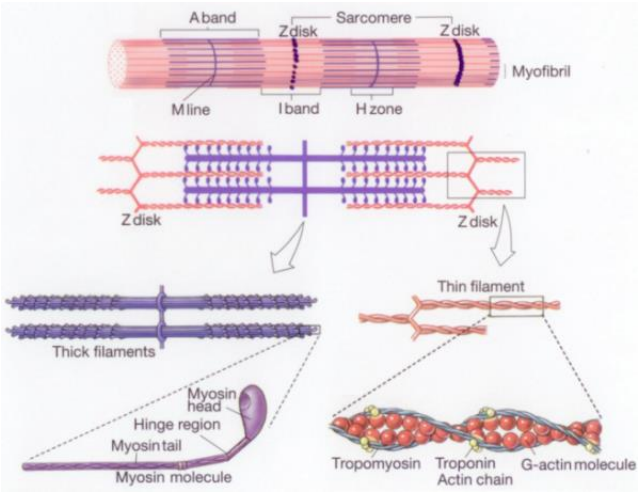


Figure 8. **Protein composition of the sarcomere.**

Different bands are present in the sarcomere: A-band constituted by myosin and actin filaments, the I-band constituted by actin filaments only and the H-band constituted by myosin filaments only. The sarcomere is delimited by Z-discs at both ends. Image modified from[17].

1.2.2.1.3 Proteins and supporting structures

As already discussed, sarcomeres consist of two types of filaments, thick and thin: the interaction of these elements allows muscles to contract[16]. The myosin protein in the thick filaments constitutes 60% of the total myofibrillar protein content. This protein is composed of six polypeptides: two heavy and four light chains. The heavy chains contain the myosin heads that interact with actin and allow muscle to contract. The myosin heavy chain also contains the ATP binding sites: they provide energy for muscle contraction[16]. Other regulatory proteins of

myosin are the tropomyosin and the troponin proteins that play an important role in the control of contraction. Actin is the next most abundant contractile protein and constitutes 20% of the total myofibrillar proteins. In its filamentous form F-actin is the principal protein of the thin filaments. A number of proteins which are neither contractile nor regulatory are responsible for the structural integrity of the myofibrils, and for this reason, for fibers contraction capacity. Another important protein found in the muscle is dystrofin, it binds to actin intracellularly and is also associated with a large oligomeric complex of glycoproteins, the dystroglycan/sarcoglycan complex that stabilizes the muscle fiber and transmits forces to the extracellular matrix.

Dystrophin gene is affected in Duchenne muscular dystrophy (DMD), a fatal disorder that develops from the absence of this protein. The gene is one of the largest encoded and accounts for the high mutation rate of DMD (35% of cases are new mutations).

Essential for muscle fiber functions are organelles such as ribosomes, Golgi apparatus and mitochondria. Most of them are located around the nuclei, between myofibrils and sarcolemma. Mitochondria, lipid droplets and glycogen provide the metabolic support needed by active muscle. The number of mitochondria in an adult muscle fiber is not fixed, but can increase or decrease quite readily in response to sustained changes in muscle activity. Lipid droplets represent a rich source of energy that can be used only by the oxidative metabolic pathway in fibers that have a high mitochondrial content and good capillary supply. During burst of activity, glycogen granules provide an important source of anaerobic energy that is independent from blood flow to the muscle fiber. Finally, the sarcoplasmic reticulum (SR) is a specialized form of smooth endoplasmic reticulum forming cisternae that fills much of the space between myofibrils. In the cisternae calcium is stored at a high concentration retained by calsequestrin, and can be released during muscle contraction.

1.2.2.1.4 Muscle contraction

Voluntary movements mediated by motor neurons determine the activity of skeletal muscle. An action potential at the neuromuscular junction causes acetylcholine to be released into the synaptic cleft. Acetylcholine triggers an instantaneous increase in permeability and conductance of postsynaptic membrane. The local depolarization initiates an action potential in the surrounding sarcolemma, an excitable membrane in which action potential propagates rapidly, ensuring that all parts of the muscle fibers are activated rapidly and synchronously. The action potential leads to the release of calcium

ions from the sarcoplasmic reticulum that causes an increase of calcium concentration in the cytosol.

Calcium mediates the excitation-contraction coupling thanks to calcium-sensitive switches in the thin filaments (troponin-tropomyosin complex). Contraction leads to sarcomere shortening, but the length of the thick and thin filaments does not change during muscle contraction (Figure 9). The sarcomere shortens thanks to the filaments sliding one between the other. Myosin head binds to actin and pulls the actin filament along the myosin filament, allowing the sarcomere to shorten during contraction (Figure 9). ATP provides energy for the attachment and detachment of myosin heads to actin filaments. At the end of excitation, calcium is actively transported in the sarcoplasmic reticulum by calcium ATPase pumps, and the muscle relaxes.

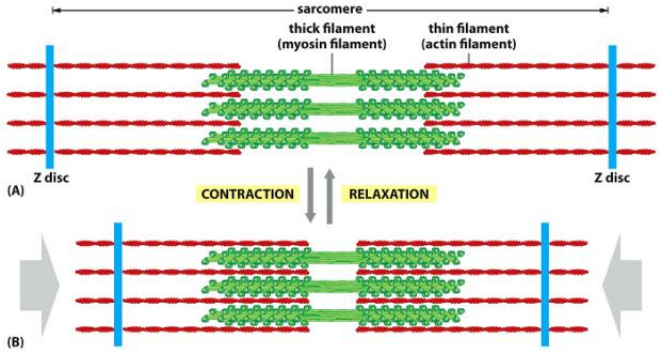


Figure 9. **Contraction and relaxation of a sarcomere during muscle contraction.**

During contraction, the sarcomere shortens by the sliding of thick and thin filaments one another; in relaxation the sarcomere revert to initial conformation. Image from[18].

1.2.2.2 Muscle fibers

Over the past decades, the number of techniques available for classifying muscle fibers has increased, resulting in several classification systems[16]. The first classification of muscles was based on the colour of muscles themselves. The division into white and red muscle depends on the amounts of myoglobin and on the capillary content. This division was also in agreement with the next classification of muscles into fast and slow made on the speed of contraction.

Today, muscle fiber types can be described using three main approaches: 1) the histochemical staining for myosin ATPase, 2) the identification of myosin heavy chain isoforms and 3) the metabolic assay.

1. *Histochemical staining for myosin ATPase*. During the cross-bridge cycle, the myosin molecule hydrolyses ATP for force generation therefore measures of ATPase activity can be interpreted in terms of contraction speed. Using the myosin ATPase staining two main categories can be described: (a) type I fibers, slow contracting, and (b) type II, fast contracting.

2. *Myosin heavy chain isoforms identification*. Identification of different myosin heavy chain isoforms (*MCH*) allows for fiber type classification. The single-fiber electrophoretic separation (SDS-PAGE technique) is used to identify the pattern of myosin heavy chain isoforms expressed in a given muscle. Although the human genome contains at least ten genes for myosin heavy chains, only three are expressed in the adult. From the slowest to fastest, these isoforms are *MCHI*, *MCHIIa* and *MCHIIx*. A further fast isoform is present in mice: the *MCHIIb*.

3. *Metabolic assay*. On the basis of metabolic differences, individual fibers can be classified as predominantly oxidative (slow-twitching "red" fibers), or glycolitic (fast-twitching "white" fibers). Oxidative/aerobic fibers obtain their energy by aerobic oxidation of substrates, particularly of fats and fatty acids. They have large numbers of mitochondria, contain myoglobin and are supported by a well-developed network of capillaries. At the other end of the spectrum, glycolitic/anaerobic fibers have few mitochondria, little myoglobin content, and sparse capillary network, these fibers store energy as cytoplasmic glycogen granules. Their immediate energy requirements are met through anaerobic glycolysis: they are capable of brief burst of intense activity but less sustainable than oxidative metabolism.

This classification leads to three fiber types: fast-glycolytic (FG), fast-oxidative (FOG), and slow oxidative (SO) (Table 1).

Type I fibers are generally oxidative (slow oxidative) and resistant to fatigue, type IIA are moderately oxidative (fast oxidative) and fatigue resistant, type IIB largely rely on glycolytic metabolism (fast glycolytic) and are easily fatigued. In mouse, type IIX fibers have an intermediate phenotype (oxidative/glycolic), but they are usually described as fast fibers.

Table 1. **Skeletal muscle fiber types.**

Muscles contain fibers of different types; in general more than one myosin isoform is expressed in a single fiber, a condition far from the simple pure fiber type. The IIX isoform is expressed only in mouse, even though the gene is also found in the human genome.

Fiber Type	Myosin gene	Myosin Protein	Contractile speed	Metabolism
I	*MYH7*	MyHC-I	Slow	Oxidative
IIA	*MYH2*	MyHC-IIA	Fast	Oxidative
IIX	*MYH1*	MyHC-IIX	Fast	Oxidative/Glycolitic
IIB	*MYH4*	MyHC-IIB	Fast	Glycolitic

Not only structural protein expression and metabolism can be used to classify muscle fibers but also the association with motor neurons. Based on the contractile and fatigue characteristics of the muscle fibers, motor units are first classified as slow-twitching (S) or fast-twitching (F), and then as fast-twitching fatigue-resistant (FR), fast-twitching fatigue-intermediate (Fint), and fast-twitching fatigable (FF)[16].

1.2.2.2.1 Fiber type transformation

Fiber composition of a muscle is determined in part by genetic factors. However, myofibers are not fixed units but instead are capable of changing their phenotype profile in terms of both size and type[16]. This is associated with muscle plasticity. If fast muscles are stimulated with an electrical pattern similar to that normally experienced by slow muscles (several weeks at 10Hz), they develop slow contractile characteristics and acquire a red appearance with a resistance to fatigue even greater than that of slow muscles. If this stimulation is removed, the sequence of events is reversed and the muscle regains, over a period of weeks, the original characteristics. This plasticity in contractile and metabolic properties occurs in response to demand, muscle diseases and endurance exercise. The muscle undergoes changes in a) the metabolic pathways responsible for the generation of ATP and, b) in the dependence on anaerobic/aerobic metabolism[19].

1.2.2.2.2 Muscle classification

When a certain type of fiber prevails, the muscle in its totality shows characteristics that reflect that type of fiber.

In literature a lot of articles examine the fiber composition of the leg muscles such as soleus, extensor digitorum longus (EDL), tibialis anterior (TA), and gastrocnemius[20]. The soleus is composed mainly by type I fibers, while EDL and TA are composed mainly by type II fibers (both IIA and IIB)[20–22]. Therefore, soleus is often described as a slow oxidative muscle while EDL and TA as fast glycolytic muscle

Muscle classification according to fiber type composition was important to understand if some lncRNAs have a different subcellular location in different muscle types (slow/oxidative vs fast/glycolytic).

1.2.2.3 Myogenesis

Myogenesis is the process by which progenitor cells give rise to myoblasts that fuse onto contractile multi-nucleated myofibers. The myogenic gene expression program is orchestrated by a hierarchy of transcription factors, including the myogenic regulatory factors (MyoD, Myf5, Myogenin, MRF4) and the families of myocyte enhancer factor (MEF2A-D) (Figure 10).

These factors act in coordination with other transcriptional regulators to execute the muscle differentiation program[4]. These myogenic transcription factors belong to a family called bHLH (basic helix-loop-helix) proteins (sometimes also referred to as myogenic regulatory factors, or MRFs). Basic helix-loop-helix proteins bind similar sites on the DNA (E-boxes) to activate muscle-specific genes. Any cell expressing a myogenic bHLH transcription factor becomes committed to forming a muscle cell. The initial cascade starts with the expression of *Pax3* that activate the *MyoD* and *Myf5* genes. In the formation of skeletal muscles, MyoD establishes a temporal cascade of gene activation: it binds directly regulatory regions to activate gene expression of muscle specific genes (*Mef2*), then it activates other genes whose products act as cofactors for MyoD, activating a second set of muscle-specific genes. MyoD also directly activates its own gene. During myogenesis myoblasts align and fuse to form multinucleated myotubes. Muscle cell fusion begins when the myoblasts leaves the cell cycle, stop dividing, secrete fibronectin onto their extracellular matrix and bind to it by integrin proteins. The second step is the alignment of the myoblasts into chains and the following fusion. Fusion is mediated by metalloproteinases called meltrins. As the myoblasts become capable of fusion, another myogenic bHLH protein, myogenin, becomes active. Myogenin binds to the regulatory regions of several muscle-specific genes

and activate their expression. Instead, myostatin, a member of the TGF-B family of paracrine factors[23] is a negative regulator of myogenesis, is.

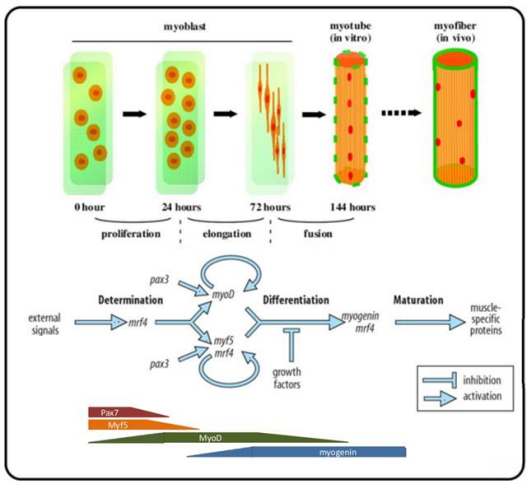


Figure 10. **Myogenesis.**

Myogenesis occurs following a precise scheme. Stem cell maintenance is mediated by the prolonged expression of *Pax3* and *Pax7* then the cell starts to express myogenic transcription factors like MyoD and Myf5. These signals cooperate to induce muscle differentiation, because they promote the activation of myogenin. When muscle specific proteins are expressed myoblasts fuse and myotubes are formed. Image modified from[24].

1.2.2.4 Muscle cell maintenance: satellite cells

The mechanisms responsible for the maintenance and regeneration of skeletal muscle were unclear until the discovery that a population of satellite cells, so-called because of their position on the edge of the fibers, existed between the basal lamina and its sarcolemma of the muscle fiber. Although satellite cells constitute 2-5% of the nuclei enclosed by the basal lamina, they have a critical role in muscle regeneration that was described only recently[19]. Satellite cells are not a homogeneous population; they do not fuse but instead remain potentially available throughout adult life. In satellite cells, the combination of *Pax3* and *Pax7* expression does not allow activation of bHLH transcription factors. *MyoD* is not expressed and *Pax7* protects the satellite cells against apoptosis. Satellite cells that express *Pax7* but not *Myf5* (*Pax7*$^{+}$/*Myf5*$^{-}$) appear to be able to divide asynchronously and produce two types of cells: *Pax7*$^{+}$/*Myf5*$^{-}$ stem cells and *Pax7*$^{+}$/*Myf5*$^{+}$ satellite cells which can differentiate into muscle. Injury or exercise can causes satellite cells to enter the cell cycle, produce bHLH transcription factors, and fuse with the existing muscle fibers in three or four days. Therefore satellite cells are both necessary and sufficient for effective regeneration of damaged skeletal muscle[23].

1.3 LncRNAs in skeletal muscle

The functions of many lncRNAs are still poorly understood, however, there are many evidences that they take part of the muscle regulatory network and are critical regulators of muscle differentiation[4].

H19 lncRNA was the first lncRNA described in mammalian cells. It is transcribed from the maternal allele of *H19/Igf2* locus and produces a lncRNA highly expressed in developing embryo and adult muscle, both in human and mouse. *H19* contains several binding sites for the miRNA *let-7 family*, and may act as a miRNA "sponge" for the members of this family[4]. The depletion of *H19* causes precocious muscle differentiation, a phenotype that is recapitulated by *let-7* overexpression; high *let-7* levels are in fact associated with increased cellular differentiation, so probably *H19* inhibits *let-7* activity, thereby preventing precocious differentiation[2].

^{DRR}RNA and ^{CE}RNA lncRNAs. Many nuclear lncRNAs regulate the activity of enhancers, regulatory elements involved in proper temporal and tissue specific expression of protein-coding genes. The finding that eRNAs may have a key role in the regulation of muscle differentiation is recent: for instance, they are involved in the regulation of *MyoD*[4]. In two enhancer regions of the *MyoD* gene, the Distal Regulatory Regions (DRR) and the Core Enhancer (CE), two enhancer-associated lncRNAs, *^{DRR}RNA*, and *^{CE}RNA*, were identified.

^{CE}RNAs facilitate the occupancy of RNA Polymerase II *in cis* by increasing chromatin accessibility (Figure 11), stimulating the expression of *MyoD*, while *^{DRR}RNAs* function *in trans* to promote the expression of *Myogenin*. These eRNAs seem provide a positive feedback mechanism to reinforce the myogenic differentiation commitment upon satellite cell activation[14].

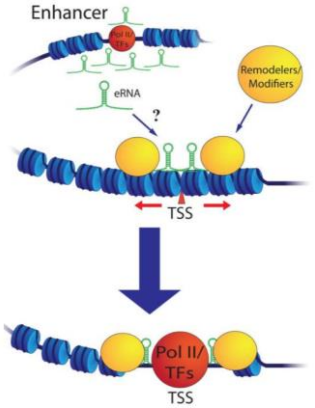


Figure 11. **Mechanism of action of eRNAs.**

Enhancer RNAs can allow polymerase accession in a promoter region to favour gene transcription. Image modified from[25].

SRA lncRNA. LncRNAs control gene expression not only modifying chromatin accessibility to RNA polymerase, but also by directly modulating the activity of sequence-specific transcription factors. This is the case for *Steroid receptor RNA activator* (*SRA*) lncRNA, the first example of lncRNA regulating myogenesis. It acts as a scaffold (Figure 2b), bringing together multiple factors that modulate gene expression, including *MyoD*. In muscle, a complex composed by SRA, MyoD and the RNA helicases p68 and p72 is present during myogenesis. The two helicases are co-activators required for transcription of *MyoD* targets genes and for muscle differentiation. The *SRA* gene produces multiple transcripts through alternative splicing; the retention of the intron 1 generates the long non-coding *SRA*, instead the spliced form generates an open reading frame that code for the SRA protein (SRAP). The ratio SRAP/*SRA* varies during muscle differentiation: SRAP mRNA is more abundant in myoblasts, while *SRA* lncRNA is more abundant in myotubes. SRAP is an RNA binding protein that specifically binds *SRA* lncRNA, thus preventing SRA-mediated regulation of MyoD transcriptional activity. The correct balance between coding and non-coding SRA molecules is important for normal muscle differentiation (e.g. aberrant *SRA* splicing occurs in myotonic dystrophy patients)[4].

Glt2/Meg3 lncRNA. The *Dlk1-Dio3* region is a very complex imprinted locus involved in tissue growth regulation and human cancer[4]. The locus contains genes for protein-coding RNAs, lncRNAs, miRNAs and snoRNAs both in mouse and human. *Glt2/Meg3* is a lncRNA originating from this locus that acts *in cis*. It binds and recruits Polycomb Repressive Complex 2 (PRC2) to repress transcription of the protein coding gene *Dlk1* and of *Glt2-as*, another lncRNA encoded by the same locus. The knockout of *Glt2/Meg3* lncRNA results in perinatal lethality with defects in skeletal muscle development[14].

Yams lncRNAs. Yin Yang 1 (YY1) is a multifunctional transcription factor whose gene is located in the vicinity of *Glt2/Meg3* lncRNA and that regulates various processes of development and differentiation. It is highly expressed in proliferating myoblasts and gradually downregulated when differentiation starts. YY1 is well known to play a dual role since it can act as repressor or activator of transcription depending on the recruited cofactors[26]. YY1 regulates a family of lncRNA known as *Yams* (YY1-associated muscle lncRNAs) that show distinct expression patterns during myogenesis and affect muscle differentiation. While *Yam-2* and *Yam-3* promote muscle differentiation, *Yam-1* and *Yam-4* are negative regulators of myogenesis. Interestingly, *Yam-1* displays a *cis* effect on the expression of neighbouring genes like those encoding for *miR-715*. This miRNA targets *Wnt7b* in skeletal muscle[14].

LincRNA MD1. One of the first lncRNA identified with a role in myogenesis was *linc-MD1* (long non-coding RNA, muscle differentiation 1). It is expressed in a specific temporal window and was shown to activate late stages of the myogenic program by functioning as a ceRNA (Figure 6). It acts as a molecular sponge for *miR-133* and *miR-135*. Through this mechanism it promotes the expression of MEF2C and mastermind-like protein 1 (MAML1), two transcription factors with important role in muscle differentiation[2,4].

LncRNAs containing SINE elements. LncRNAs containing SINE elements regulate the stability of several mRNAs encoding for protein with a role in muscle differentiation. During myogenesis, the efficiency of nonsense-mediated mRNA decay (NMD) decreases, while the efficiency of Staufen-mediated mRNA decay (SMD) increases. These are in fact competitive pathways playing both an important role in muscle differentiation. Myogenin and PAX3 are differentially targeted by these two pathways of degradation and this different susceptibility contributes to their relative abundance during differentiation. *PAX3* mRNA is an SMD target and its increased decay promotes myogenesis, whereas decreased degradation of the NMD target myogenin is required for myogenesis.

Malat1 lncRNA. Metastasis-associated lung adenocarcinoma transcript1 (*Malat1*) was one of the first cancer-associated lncRNAs discovered. *Malat1* localizes to bodies known as nuclear speckles, where it regulates alternative splicing and gene expression at the molecular level[10] interacting with the epigenetic repressor Polycomb protein Cbx4 and with the SR family of splicing factors. *Malat1* has a role in the transition from the proliferative phase to the

permanent cell cycle, as well as in the commitment to differentiation. Moreover, recently it was found as a novel downstream target of myostatin.

1.4 LncRNA associated to skeletal muscle diseases

Considering the wide range of roles that lncRNAs play in cellular networks, it is not surprising that non-coding RNAs have been implicated in disease. Genome-wide association studies have revealed that a great number of diseases or traits associated to diseases reside outside of protein coding genes. A direct involvement of lncRNAs is demonstrated nowadays for some pathologies.

Skeletal muscle mass can be markedly reduced through a process called *atrophy*. Atrophy is characterized by the reduction of fiber volume, a decrease in protein content, fiber diameter, force production, and fatigue resistance. During atrophy, enhanced protein breakdown is the primary cause of the rapid loss of muscle proteins. Atrophy is a complex modification occurring in skeletal muscles as a result of a variety of causes, from mutations in genes related to muscle function, to others diseases that have repercussions on muscle activity. Alterations in the activity of the IGF-1/Akt/FoxO and myostatin pathways seem to play key roles in skeletal muscle growth, causing atrophy or hypertrophy. Neurodegenerative diseases are hereditary and sporadic conditions characterized by progressive nervous system dysfunction. These disorders are often associated with atrophy of the affected central or peripheral structures of the nervous system. They include diseases such as Alzheimer's disease and other dementias, brain cancer, degenerative nerve diseases, encephalitis, epilepsy, genetic brain disorders, Parkinson's disease, multiple Sclerosis, Amyotrophic Lateral Sclerosis (ALS), Huntington's disease, Prion diseases, and others.

Atrophy induced by denervation. Denervation consists of a loss of communication between nerve and skeletal muscle that cause a strong reduction of muscle fiber dimensions. This can happen as a result of muscle/nerve injury or can be related to muscle/nerve pathologies. Skeletal muscle denervation causes muscle atrophy via complex molecular mechanisms that are not completely understood. Muscle fibers that lose innervation show a characteristic shrivelled and angular shape.

Denervation atrophy progresses rapidly during the initial stages of disease, and myofibers result as an ensemble of cellular nuclei without sarcoplasma.

Duchenne muscular dystrophy (DMD). DMD is the most common and severe myopathy. It is an X-linked recessive inherited pathology and is characterized by severe muscle wasting from early childhood that usually arises in leg and pelvic muscles, and later extends to the trunk of the body, compromising the heart and respiratory muscles. DMD is caused by mutations in the Dystrophin gene. Recently, lncRNAs were found to be transcribed from intronic sequences of Dystrophin gene, thus demonstrating a concomitant expression of lncRNAs with dystrophin[4].

Facioscapulohumeral muscular dystrophy (FSHD). FSHD is the third most common type of muscular dystrophy. It results in a more restricted pattern of muscular weakness compared with DMD, mainly confined to the facial mimic and shoulder girdle muscles but extending to abdominal and leg muscles in the most severe cases. The genetic cause affects the copy number of the 3.3Kb macrosatellite D4Z4 mapping at the subtelomeric region of chromosome 4. FSHD patients carry strong deletions of repeated units causing profound change in the epigenetic status. Recently, a lncRNA was identified that, selectively synthesized by FSHD patients, promotes the depletion of the locus associated with FSHD contributing with pathogenesis[4].

Amyotrophic lateral sclerosis (ALS). Amyotrophic lateral sclerosis is a term used to cover the spectrum of syndromes characterized by progressive degeneration of motor neurons. It causes progressive and cumulative physical disabilities, and leads to death due to respiratory muscle failure. In ALS, both the upper motor neurons and the lower motor neurons degenerate or die, ceasing to send messages to muscles. Unable to function, the muscles gradually weaken, waste away, and twitch. ALS is heterogeneous in its presentation, course, and progression. We do not yet fully understand the causes of the disease, nor the mechanisms for its progression. Unfortunately, there is no definitive diagnostic test for ALS; often disease progression is a pre requisite[27]. LncRNAs have been associated with ALS, most notably *Neat1* was shown to modulate the functions of ALS-associated RNA-binding proteins during the early phase of ALS[28].

Recognition of the roles of lncRNAs in human disease has unveiled new diagnostic and therapeutic opportunities. Advances in lncRNA genetic study will probably led to new therapeutic approaches.

1.5 miRNAs involved in muscle physiopathology

Several miRNAs have been identified to directly influence muscle development. Most miRNAs modulate the status of quiescent satellite cells and/or myogenic activation by directly targeting *Pax3/7*, the master transcription factors that maintain satellite cell quiescence. For example, *miR-31* in satellite cells maintains the quiescence state by sequestering *Myf5* mRNA in the nucleus.

The *miR-1/206* family is capable of promoting myogenesis, in part by inhibiting *Pax3/7* in embryonic muscle precursors and satellite cells, instead *miR-186* targets myogenin and thus inhibit terminal muscle differentiation. *MiR-675-3p* is another recently identified miRNA that promotes muscle differentiation by targeting *Smad1* and *Smad5*. This miRNA is encoded in the exon of *H19* lncRNA, which also plays a critical role in myogenesis. *MyomiRs* are defined as muscle specific miRNAs and particularly interesting miRNAs in this category are those codified from intronic regions of the MyHC genes (*Myh6*, *Myh7* and *Myh7b*). Those that play important roles in regulating muscle myosin isoforms and therefore muscle fiber type are *miR-208a*, *-208b*, and *-499*. *MiR-208b*, and *-499* repress fast myofiber genes while activates slow muscle specific genes. The effect is mediated by targeting transcriptional repressors of slow muscle fiber genes, such as *Sox6*[14]. *MiR-133* family (*miR-133a-1*, *-133a-2* and *-133a-3*) is also critical for both skeletal and cardiac muscles development.

Chapter 2 – Aims of the thesis

The aim of this thesis was to identify the subcellular localization of all long non-coding RNAs (lncRNAs) expressed in fast and slow myofibers and then relate them with the pathophysiological models of skeletal muscle. LncRNAs are emerging regulators of many biological processes; they are prevalently studied in cancer and only a small group of lncRNAs has been studied in skeletal muscle. Since lncRNA functions are related to their subcellular localization and expression, our analysis were focused on defining both these aspects at the single cell level.

In previous microarray experiments we collected lncRNA expression data associated with differential RNA extraction from purified nuclei and cytoplasm; in this thesis we analysed the expression data and showed that lncRNAs can localize specifically in nuclei or cytoplasm of myofibers isolated from either Soleus, Tibialis Anterior (TA) or Extensor Digitorum Longus (EDL). In order to validate this analysis, we performed Fluorescence In Situ Hybridization (FISH) experiments. This allowed the analysis of subcellular localization of specific lncRNAs in skeletal muscle tissue sections also evidencing their fine regionalization inside the nuclei.

Skeletal muscle, together with liver and earth, is one of the most important tissue regulating mammalian metabolism; we analysed the expression of a group of lncRNA in these tissues in order to identify those lncRNA that were preferentially expressed in skeletal muscle.

Different types of skeletal muscle can present different types of metabolism; they can be divided in slow-contracting muscles (oxidative metabolism) or fast-contracting muscles (glycolytic metabolism) depending on their different myofiber composition. For this reason we also analysed the expression of lncRNAs in single myofibers and compared the expression in slow-twitching single myofibers with the expression in fast-twitching myofibers.

Since it is tricky to induce lncRNA down- or up-regulation *in-vivo* it is important to have *in-vitro* systems where studied lncRNA have the same subcellular location of *in-vivo*. We used the model cell line C2C12, routinely used to mimic muscle development, and analysed the expression of lncRNAs in C2C12 cells differentiation.

Specific roles for lncRNAs have been documented in different diseases, but no data are available on skeletal muscle atrophy. Skeletal muscle atrophy is a severe

disease affecting a large amount of people in the world, its development can be linked with many different factors like cancer, bed rest, and ageing and the identification of potential regulators could improve the development of new drugs able to counteract effects of this invalidating aspects of life. We evaluated the expression of lncRNAs in different atrophy models: denervation or starvation induced atrophy and amyotrophic lateral sclerosis (ALS) to identify common or specific induced/repressed lncRNAs.

Chapter 3 – Materials and Methods

3.1 Genome-wide approach for long non-coding RNAs localization

LncRNAs have different functions in relation to their subcellular location. As discussed in Chapter 1, nuclear lncRNAs can influence gene expression modulating chromatin architecture or interacting with transcription factors, while cytoplasmic lncRNAs can function as sponges for miRNAs. It is important to study subcellular location of as many lncRNAs as possible in order to understand their role and function. For this reason we adopted a genome wide approach associating specific nuclear or cytoplasmic RNA extraction with microarray technique.

3.1.1 Statistical analysis: Significance Analysis of Microarrays (SAM)

In order to define subcellular localization of lncRNAs we identified differentially expressed lncRNAs between nuclear and cytoplasmic RNA preparations derived from the same fiber pool. Differentially expressed lncRNAs were identified according to Significance Analysis of Microarrays (SAM) analysis[29] implemented in T-MeV suite[30] also used to produce heat maps. 1% of false significant genes were considered in this analysis. LncRNAs most expressed in the nucleus were considered as nuclear, while those most expressed in the cytoplasm were considered as cytoplasmic.

3.1.2 Validation of the genome wide lncRNA subcellular localization

To validate the genome-wide approach we selected about 30 lncRNAs to analyse their localization with two alternative techniques: the qRT-PCR associated with nuclear and cytoplasm fractionation and the Fluorescent In Situ Hybridization (FISH). We selected lncRNAs which were identified as being localized in the nucleus, in the cytoplasm or that didn’t show a specific localization.

3.2 C2C12 Cell Culture

C2C12 myoblast cell line was chosen as in vitro model of skeletal muscle differentiation. Cultures were used for both RNA extraction and FISH analisys.

3.2.1 Thawing

Frozen aliquots containing 10^6 C2C12 myoblasts were quickly thawed in a 37° C water bath. Cells were diluted slowly with pre-warmed (37° C) proliferative medium (DMEM supplemented with 10% FBS and 1% Penicillin; ThermoFisher Scientific) and then centrifuged at 300 g for 5 min.

Pellet containing the thawed C2C12 myoblasts was resuspended in 10 ml of proliferative medium and plated on a 10 cm diameter Tissue Culture petri dish.

3.2.2 Propagation

Plated C2C12 cell cultures were maintained proliferating in an incubator at 37° C with high humidity and 5% CO_2. To avoid spontaneous differentiation, the cells were detached every 2-3 days, upon reaching ~80% confluence, using Trypsin-EDTA 0.25%. Detached cells were plated with a 1:10 dilution.

3.2.3 Differentiation

Differentiation of the C2C12 myoblasts into myotubes was induced switching the growth medium from proliferative medium to differentiative medium (DMEM with 2% HS and 1% Penicillin; ThermoFisher Scientific) when the cell culture reached confluence. Once switched to differentiative medium the C2C12 were monitored daily and the medium was changed every day. The first myotubes were evident after 3 days in differentiative medium.

3.2.4 Cryopreservation

C2C12 were cryopreserved in order to stock the culture. Cells were detached using Trypsin-EDTA and centrifuged to remove all proliferative medium; the pellet was then resuspended in FBS with 10% DMSO and stored in aliquots of 10^6 cells at -80° C.

3.2.5 Nuclei purification from C2C12 cultures

To separate the nuclei from the cytoplasm, C2C12 cell cultures were washed two times with ice-cold PBS buffer and then lysed with ~550 µl of RLN buffer prepared as follow: 50 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5% (v/v) Triton X-100 (1.06 g/ml).

Just before use, it was added with 1,000 U/ml of RNase inhibitor. In order to assure complete lysis, RLN solution containing the cell suspensions were passed repeatedly through a syringe needle (0.45 mm of diameter) and the result was checked at the microscope with the addition of trypan blue and DAPI.

Nuclei were then pelleted via low speed centrifugation (3 minutes at 6,000 g) and cytoplasmic fractions (surnatants) were moved to new microcentrifuge tubes to be concentrated with a vacuum-pump centrifuge.

After volume reduction of cytoplasmic fractions, 1 ml of Trizol (ThermoFisher Scientific) was added both to purified nuclei and cytoplasmic fractions and solution were stored at -80° C until needed for the RNA extraction.

3.3 RNA Extraction from single myofibers

3.3.1 Enzymatic dissociation of myofibers from whole muscle

We modified published methods for long fibers isolation[31–33] in order to keep the digestion time as short as possible and avoid activation of stress response genes. Muscles from both hind limbs of the same mouse were immediately removed by microdissection, taking care to handle them only by their tendons to minimize mechanical damage to the fibers. Digestion proceeded for 40 – 45 min. at 37° C in 1 ml high-glucose Dulbecco's modified Eagle medium (DMEM; ThermoFisher Scientific) containing 10 mg type I collagenase (220 U/mg; Sigma). The collagenase-treated muscles were sequentially rinsed for 2 min. in 3 ml of DMEM, 3 ml DMEM supplemented with 10% fetal bovine serum (FBS) and 3 ml of DMEM and finally transferred into 50 mm x 18 mm well containing 3 ml of DMEM with 10% FBS. Single myofibers were liberated by gentle physical trituration with a wide-mouth plastic Pasteur pipette (about 4 mm diameter). The triturating process was repeated until several intact fibers were obtained. After each physical trituration, the muscles were transferred in a new well, to get rid of collagen wisps and hyper-contracted fibers. Quickly, intact and well-isolated

fibers were picked under stereo-microscope and washed first in DMEM and then in PBS. All samples were collected within 45 min. from the last trituration step.

3.3.2 Total RNA extraction from single myofibers

Total RNA was extracted from fibers using the silica membrane technology of RNeasy Micro Kit (Qiagen). Single fibers were disrupted by adding 500 µl of Trizol (ThermoFisher Scientific) and lysate was homogenized by vortexing for 5 min. The protocol used was essentially the one suggested by the manufacturer, with the following modification: RNA elution was performed with 14 µl RNase-free water pre-heated at 37° C and repeated a second time to avoid loss of RNA in the column. Due to the dead volume of the column, we recovered about 20 – 24 µl. We estimated that the amount of total RNA purified from a single fiber is in the range of one to few nanograms.

3.3.3 Total RNA extraction from tissue samples and cell coltures

All tissue samples were homogenized in sufficient Trizol (ThermoFisher Scientific) volume to avoid RNA degradation. Pelleted C2C12 nuclei and cytoplasm were prepared as previously described, while 30 mg of each tissue sample were homogenized in 1 ml of Trizol (ThermoFisher Scientific). To each tube were added 200 µl of chloroform per ml of Trizol (ThermoFisher Scientific) and mixed well by inverting the tube.

After an incubation of 20 min. on ice, the samples were centrifuged at 13,000 g for 20 min. at 4° C. The resulting aqueous phase (~ 400 µl per ml of Trizol) was transferred to a clean microcentrifuge tube paying attention do not recover material from the lower phase. An equal volume of isopropanol was added and the resulting mix was stored at -20° C overnight. To precipitate all RNA present in the solution it was centrifuged at 13,000 g for 30 min. at 4° C. Supernatant was removed and RNA pellet was washed with 500 µl of ice-cold 80% Ethanol in water. Following another centrifugation (13,000 g at 4° C for 10 min.) pellet was air dried, resuspended in RNase free water and stored at -80° C until needed.

3.4 RNA quality control

RNA extracted from single fibers, C2C12 nuclei and cytoplasm fractions and tissues was quantized using nanodrop (Celbio) spectrophotometer and analysed using the RNA 6000 Pico or Nano LabChip on a 2100 Bioanalyzer (Aligent). The sample (1 µl) was separated electrophoretically as described by the manufacturer and data was displayed as a gel-like image and/or an electropherogram. All poor quality RNA samples were discarded.

3.5 RNA retro-transcription

RNA extracted as previously described was reverse transcribed into its complementary DNA (cDNA) in a retro-transcription reaction using SuperScript II Reverse Transcriptase (ThermoFisher Scientific). We followed the manufacturer's protocol but used both oligo dT and random hexamers as primers for the reaction. This allows the retro-transcription of RNAs do not present poly(A) tail. cDNA obtained was precipitated adding 1/10 volumes of sodium acetate and 2.8 volumes of absolute ethanol, centrifuged at 13,000 g for 20 min. at 4° C and resuspended in nuclease free water.

3.6 Primers for qRT-PCR and FISH analyses

Primers were designed using the web tool Primer3 from http://biotools.umassmed.edu/bioapps/primer3_www.cgi. qRT-PCR primers were chosen to amplify a small region (between 60 and 130 base pairs) while primers used to prepare FISH probes were chosen to produce a bigger amplicon (between 500 and 700 base pairs). In both cases, primers were designed to amplify the same isoforms the microarray probes identified as differentially localized between cytoplasm and nucleus (see paragraph 3.6). An example of isoform specificity for qRT-PCR, FISH primers and microarray probes for lncRNA *A330023F24Rik* is shown in Figure 12. For this reason we used transcript sequences downloaded from Ensembl database.

In order to check the stability of secondary structures as well as the stability of primer dimers, prier sequences were analysed using the online tool Oligo Analizer, from http://eu.idtdna.com/calc/analyzer. The list of primers used is included in *Supplementary Table 1*.

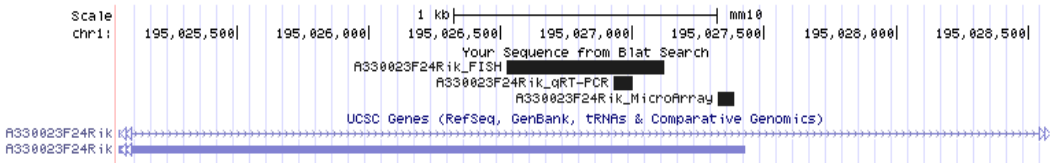


Figure 12. **Genome location of probes for FISH, microarray and qRT-PCR analyses.** Comparison between the region targeted by microarray probe and the regions amplified using qRT-PCR or FISH primers.

3.7 Real Time Analysis

Real-Time analysis was performed in a 7500 Real-Time PCR System (Applied Biosystem) using HOT FIREPol EvaGreen qPCR Supermix (Solis BioDyne). Real-Time analysis was performed following the manufacturer's instructions and using 10 ng/reaction of cDNA as template. PCR cycle was 95°C 12 min.; 95°C 15 sec, 60°C for 20 sec, 72°C 35 sec for 40 cycles and finally the dissociation step. Delta Ct method was used to identify differentially expressed coding and non-coding RNAs. As reference genes were used *Tbp*, *Tnx1* and *B2m*. The list of primers used is included in *Supplementary Table 1*.

3.8 PCR amplification - synthesis of probes for FISH analysis

For each target transcript, a region of about 600 nucleotides was cloned via TA ligation into a pCR™II-TOPO® vector (ThermoFisher Scientific), which is a high copy number *E. coli* plasmid with SP6 and T7 promoters on 5' and 3' end of clonal site. TA ligation is based on the terminal transferase activity of Taq Polymerase that add at the end of amplicons preferentially adenines. PCR amplification cycle used to produce amplicons was: 95°C 10 min., 95°C 20 sec., 59°C 30 sec., 72°C 1 min., for 40 cycles and 72° C 12 min.

No purification of PCR was performed before cloning. The orientation and sequence of the insert was checked via Sanger sequencing (BMR genomics). After linearization of the plasmids (see supplementary Table 2), probes were synthetized via in-vitro transcription using Amino Allyl UTP (in addition to conventional NTPs). This allowed us to couple the resulting 600 nucleotides probes with either modified Cy3 or Cy5 fluorophores. To remove unincorporated reagents the aRNA solution was purified using MEGAclear™ Transcription Clean-Up Kit

(ThermoFisher Scientific). Purified probes and incorporated fluorophores were quantized using nanodrop (Celbio) spectrophotometer.

Fluorescent probes were then chemically fragmented using the RNA Fragmentation kit (ThermoFisher Scientific) to obtain a final size for the probes comprised between 50 and 150 nucleotides. Probes complementary to the target RNA sequence were used to detect the presence of the transcript in the samples while probes with the sense sequence were used as negative controls.

In the *Supplementary Table 2* can be found a complete list of the enzymes used in each linearization reaction for each target gene.

3.9 Fluorescence in situ hybridization (FISH)

FISH experiments were performed on C2C12 proliferating and differentiating myoblasts and on tibialis anterior muscle (TA) transversal sections. C2C12 myoblasts were seeded on 18 x 18 mm coverslips treated with 70% EtOH (Sigma-Aldrich), fixed with 4% PFA in PBS and then dehydrated using increasing concentrations of ethanol (80%, 90%, 100%). Coverslips were then stored at -80°C. Each muscle, after mouse sacrifice, was quickly collected and frozen in liquid nitrogen. 20 µm thick sections were cut using a Leica CM 1850 cryostat at -21°C, divided in groups of three laids on SuperFrost Plus slides (Thermo Scientific) and stored at -80° C. Prior to a FISH experiment, muscle sections were thawed at room temperature and fixed with 4% PFA.

Both C2C12 cell cultures and muscle sections were permeabilized in 70% ethanol at -20° C prior incubation with the FISH probe, for at least 2 hours (C2C12) or overnight (muscle sections). 1 µg of probe was used to treat each C2C12 culture coverslip or muscle section. Hybridization step was performed overnight in a humid environment at 30° C. Samples were then washed and treated with RNaseA to remove excess of probe. Nuclei were stained using DAPI and samples where mounted using Fluoromount mounting medium (Sigma-Aldrich) to preserve the fluorescence of the probes. Images were acquired with Leica TCS SP5 confocal laser microscope.

3.10 Functional analysis of mRNA correlated with lncRNAs

To better understand *Pvt1* activity in skeletal muscle, coding and lncRNA *Pvt1* were correlated according to their expression. Pearson correlation was used and only nuclear abundant mRNAs were considered in this analysis. Nuclear abundant mRNAs were identified using the same analysis performed for lncRNAs. Enrichment analysis of *Pvt1* correlated mRNA were performed using the web tool DAVID[34].

3.11 Animals

Wild-type CD1 mice (Charles River) were housed in a normal environment provided with food and water. Adult females were killed by rapid cervical dislocation, to minimize suffering, at three months age (weight: 33-35 g). Fasting was performed as previously described, by removing chow with free access to water[35]. Denervation experiments were performed by cutting the sciatic nerve of one limb while the other was used as control. Mice were sacrificed after 3, 7 or 14 days after denervation. Mice expressing mutants SOD1G93A (glycine substituted to alanine at position 93) develop motor neuron disease that resembles features of amyotrophic lateral sclerosis (ALS). We used this mouse model to mimic atrophy derived from ALS.

3.12 Ethics statement

All aspects of animal care and experimentation were performed in accordance with the Guide for the Care and Use of Laboratory Animals published by the National Institutes of Health (NIH Publication No. 85-23, Revised 1996) and Italian regulations (DL 116/92) concerning the care and use of laboratory animals. Experimental procedures were approved by the local Ethical Committee of the University of Padova.

Chapter 4 - Results and Discussion

4.1 Genome-wide analysis

To define the subcellular localization and fiber specificity of long non-coding RNAs (lncRNAs), different microarray expression data sets were analysed using the Significance Analysis of Microarrays (SAM) algorithm.

4.1.1 LncRNA subcellular localization

A total of 1,433 lncRNAs were found to be preferentially localized in one subcellular compartment (Figure 13). 655 are more expressed in the cytoplasm, and 481 in the nucleus. A third category, represented by 297 lncRNAs, shows a different localization in relation to the type of muscle analysed (EDL, SOL or TA) or in relation to the different lncRNA isoform. A relatively large number of lncRNA (45%) was found to preferentially localize in the cytosol, in accordance with the previous observation made by *Van Heesch, S. et al* in human colon cancer cells[36]. Interestingly, skeletal muscle specificity of lncRNA expression is lesser than sharing. This could be in accordance with the fact that a small number of lncRNAs is needed to specify for different metabolism, or that most function of different muscles are related and therefore most lncRNAs share the same localization.

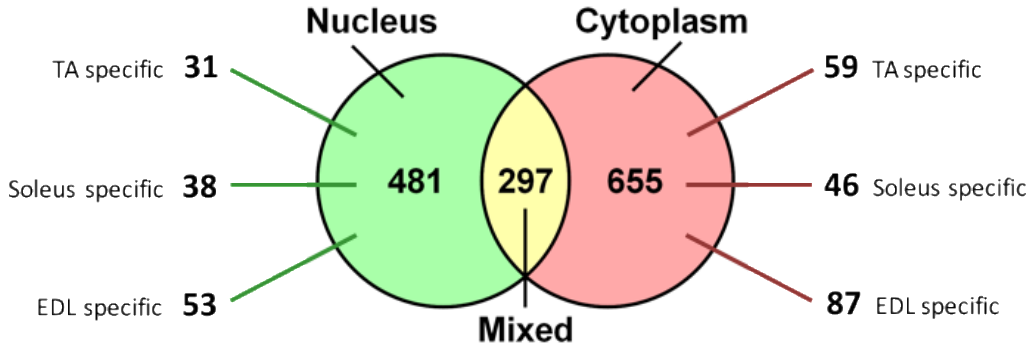


Figure 13. **Classification of nuclear and cytoplasmic lncRNAs.**

481 lncRNAs localize prevalently in the nucleus, while 655 in the cytoplasm of soleus, EDL and TA derived myofibers. 297 lncRNAs have a mixed localization. Most lncRNAs share the same localization albeit derived from different myofibers, while a small number evidence a fiber specification.

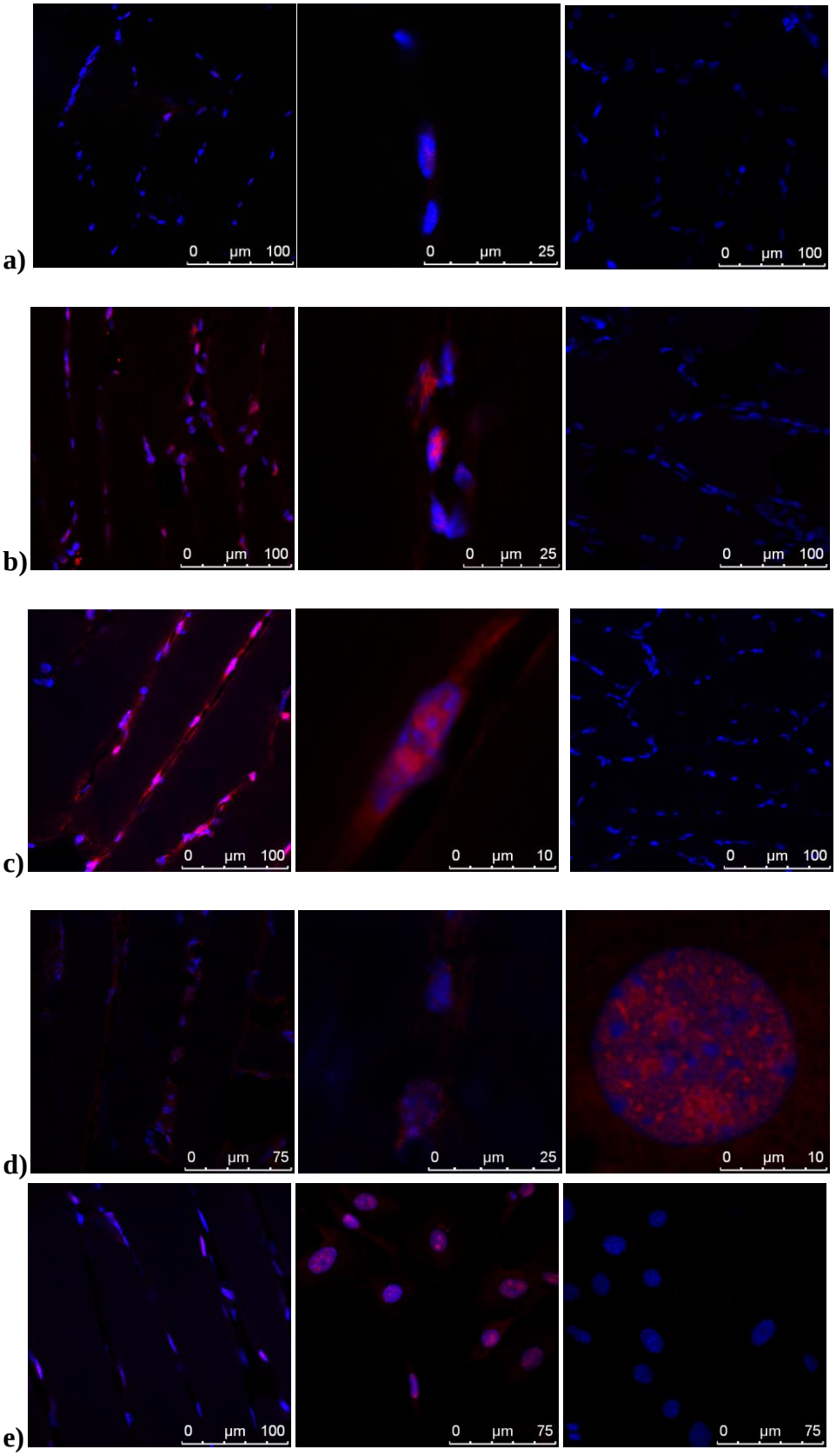
4.2 Validation of the genomewide analysis

In order to validate the results obtained with our genome wide approach, we analysed the localization of lncRNAs via Fluorescent In Situ Hybridization in both skeletal muscle (TA) tissue sections and cell cultures (C2C12).

4.2.1 Fluorescent In Situ Hibridization

Fluorescence In Situ Hybridization (FISH) is a technique that allows to target a specific sequence with a fluorescent probe.. It consists in the hybridization of a DNA or RNA probe to its complementary sequence on tissue slides or fixed cell cultures. and it can be used to detect both DNA and RNA targets., We used FISH to target 7 lncRNAs transcripts chosen among those showing differential nucleus-cytoplasm localization[37].

Using FISH analysis we were able to confirm what we observed in the previous microarray analysis. This result confirms the possibility of using cell compartment fractionation in association with genome-wide analyses, such as microarray, to infer lncRNA localization. Through FISH analysis we were also able to obtain a more accurate localization picture, discriminating between positive and negative nuclei in the same image; in many cases we were able to observe also a finer *subnuclear* localization. Not all nuclear compartments are equivalent. In fact, nuclei host transcriptional active or inactive DNA regions. The association of lncRNAs with one of these regions is useful to infer their possible positive or negative action on gene expression. Moreover, FISH on skeletal muscle evidenced a prevalent expression of the analysed lncRNA in myofibers and not in other cell type composing the muscle. This is in accordance with the microarray analyses using single myofibers that allows the removal of muscle non-contractile cells. Through this way we produced a more accurate localization and expression eliminating a background due to different cell types in muscles. The panels included in Figure 14 show the localization of lncRNAs with enlargments of some nuclear positive signals.


a)
0 µm 100
0 µm 25
0 µm 100
b)
0 µm 100
0 µm 25
0 µm 100
c)
0 µm 100
0 µm 10
0 µm 100
d)
0 µm 75
0 µm 25
0 µm 10
e)
0 µm 100
0 µm 75
0 µm 75

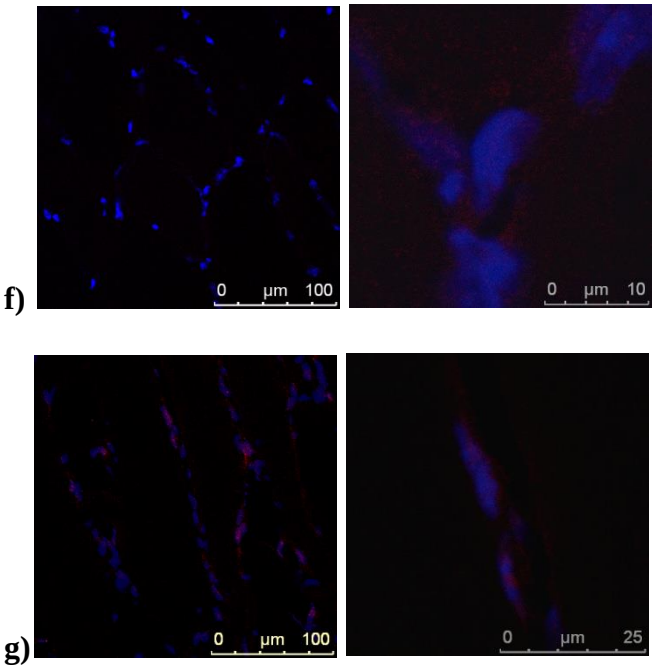


Figure 14. ***FISH on skeletal muscle sections and C2C12 myoblasts.***

From left to right are represented probe for lncRNA, magnifications of nuclear signals and negative controls (a-d). **a)** FISH for *Gt(ROSA)26Sor* that is low expressed and localizes in nuclei; **b)** FISH for *Neat1* that localizes in nuclei; **c)** FISH for *H19* that localizes both in nuclei and in cytoplasm; **d)** FISH for *Airn* that localizes both in nuclei and in cytoplasm. **e)** FISH for *Mir143hg*: left image represents FISH on tissue slides, the central the FISH on C2C12 cells, and the right the FISH with the negative control probe on C2C12. This lncRNA shows a nuclear localization; **f)** *Mir22hg*: FISH on tissue slides, magnification of nuclei. Also this lncRNA has a low expression **g)** *Nctc1*-RI: FISH on tissue slides, magnification of nuclei.

FISH analysis confirmed the localization predicted by the genome-wide analysis for all targeted lncRNA, in skeletal muscle (TA) tissue sections, cell cultures (C2C12) or both. *Gt(ROSA)26Sor*, *Neat1*, *Mir143hg* and *Nctc1-RI* have been identified as nuclear long non-coding RNAs while *H19* and *Airn* are found in both nucleus and cytoplasm. *Mir22hg* is prevalently expressed in the cytoplasm or in the perinuclear compartment.

4.3 *In-vitro* analysis of lncRNA subcellular localization

Results derived from the analysis of skeletal muscle or single myofibers are important to understand the role of lncRNAs *in-vivo*, but, at the same time, it is important to have an *in-vitro* tool that can be used to characterise lncRNA functions. C2C12 cell cultures are a well-established *in-vitro* model to study muscle differentiation. Using these cell cultures, we studied lncRNA subcellular localization during muscle differentiation.

4.3.1 Differential RNA extraction from differentiating C2C12 cells

Microarray and qRT-PCR data production require the extraction of RNA independently from nuclei and cytoplasm. We started from the isolation of nuclei and tested the quality of the purification before and after the RNA extraction. During the isolation of nuclei from muscle fibers or C2C12 cells, we checked if purified nuclei were contaminated by fiber fragments or cell debris evidencing that no fragments/debris were present (Figure 15).

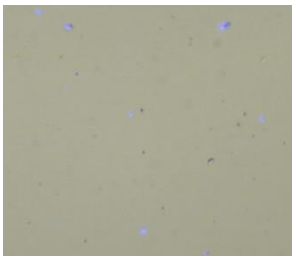


Figure 15. **Nuclei isolated from the fibers.**

To confirm that nuclei used for RNA extraction were not contamined from other cellular debris, some images were cought at the microscope. Image shows a clear preparation of purified nuclei.

After RNA was extracted from both nuclei and cytoplasm, a further analysis was performed to confirm the quality of the extraction. Using a 2100 Bioanalyzer (Agilent) we were able to discriminate between nuclear and cytoplasmic RNA according to the presence of specific peaks[38].

We evidenced that extracted RNA extracted was not degraded (Figure 16) and more importantly, that there was no contamination of cytoplasmic RNA in the nuclear RNA samples (Figure 16) since specific peaks after the 28S band (~ 4,000 nt) and at 200 nt were found, as expected, in the nuclear RNA and not in the cytoplasmic RNA (Figure 16).

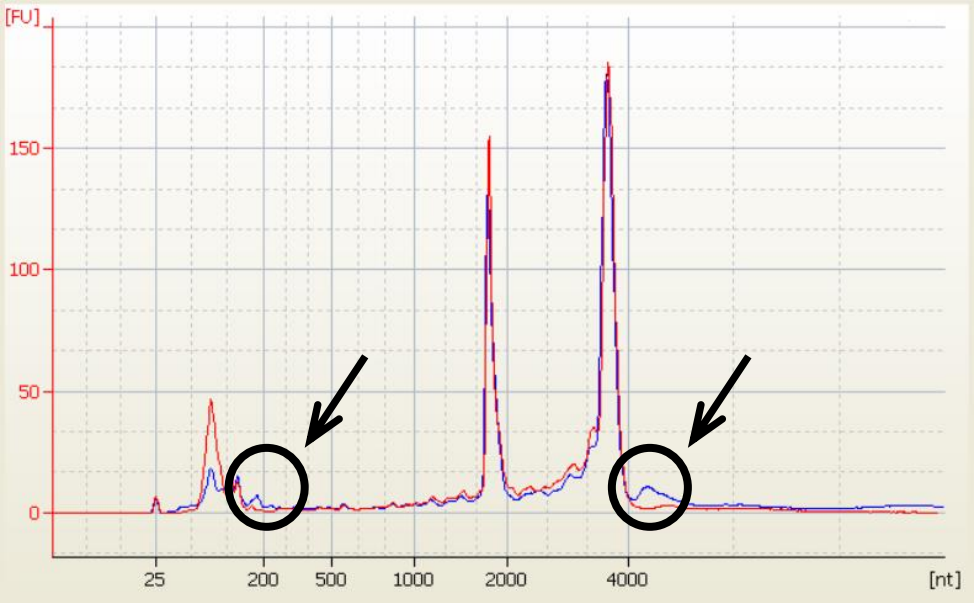


Figure 16. **RNA quality control.**

Good quality of RNA confirmed by 2100 Byoanalyzer. In red is represented cytoplasmic RNA and in blue nuclear RNA. Results are in agreement with Trask, H. W. *et al.*[38].

4.3.2 Quantitative Real-Time PCR

In order to further validate microarray data, the expression of 41 genes (31 lncRNAs and 10 coding RNAs) was analysed via quantitative Real-Time PCR (qRT-PCRs) in different conditions. Coding genes lying in a genomic location near one of the analysed lncRNAs were also included (e.g. lncRNA is in antisense with coding RNA, lncRNA is codified from an intron of coding RNA).

4.3.3 Comparison between FISH and qRT-PCRs

LncRNAs expression was analysed separately for nuclei and cytoplasm during C2C12 differentiation at five different time points (0, 1, 3, 7, 14 days) in order to have a complete picture of their expression and localization during the differentiation process. Heat map in Figure 17 shows a) if the lncRNAs tested are present only in the nucleus or only in the cytoplasm, or in both compartments, and b) if the lncRNAs changes localization during myoblasts differentiation.

Nuclear lncRNAs are *9530059O14Rik*, *A330026F24Rik*, *Dancr*, *Dnm3os*, *Gas5*, *Gm11766*, *Gm11767*, *Mir143hg*, *Snhg1* and *Snhg7*.

Among protein-coding genes, *Sox6* and *Cd46* are nuclear. The reason why the transcript for a membrane protein (Cd46) that is part of the regulatory

complement system is found in the nucleus is not clear. Interestingly, genes for *Cd46* and lncRNA *A330026F24Rik* are found in the same genomic region and code for transcripts that show the same pattern of localization (99.3% correlation). Both transcripts are found in the nucleus in accordance with microarray experiments on pools of myofibers from EDL, soleus and TA. Only *H19*, *Igf2os*, and *Snhg6* show a cytoplasmic localization during terminal differentiation. Moreover, in agreement with the genome-wide analysis, *2310065F04Rik* is found in both subcellular compartments.

Different lncRNAs increase their expression during C2C12 differentiation: (*1110020A21Rik*, *2310065F04Rik*, *Dlx6os*, *Gm16062*, *Nctc1-lnc*, *Nctc1-RI*) as well as the protein-coding gene *Dio3*. Probably these lncRNAs could be involved in cell cycle withdrawn, fusion of cells and formation of myotubes.

The protein-coding gene *Myh3* shows a spike of expression at seven days of differentiation in both subcellular compartments. Similarly, *Neat1* is up-regulated in both nucleus and cytoplasm at days 1 and 14 days of differentiation.

Vice versa, *Snhg6* progressively increases in cytoplasm. *Snhg6* was classified as nuclear lncRNA in the genome-wide analysis but here seems that it is present on both compartments, diffusing from nucleus to cytoplasm.

As stated before for lncRNAs that increase their expression during C2C12 differentiation, also for the long non-coding and the protein coding genes with a precise activation we can speculate that they can have a precise spatio-temporal function.

Interestingly, we evidenced that the localization of lncRNA *Pvt1* changes from cytoplasmic during cell proliferation (day 0) to nuclear during myoblast differentiation.

This result was later confirmed by FISH analysis as well (Figure 18) and it is interesting because *Pvt1*, a lncRNA well studied in tumours[39], codify for different miRNAs and is able to modulate Myc function[39]. In the cytoplasm it could function as a sponge for different miRNAs or is located when it is needed for the synthesis of specific miRNAs; while in the nucleus it can participate to transcriptional and cell cycle regulation during C2C12 differentiation.

§

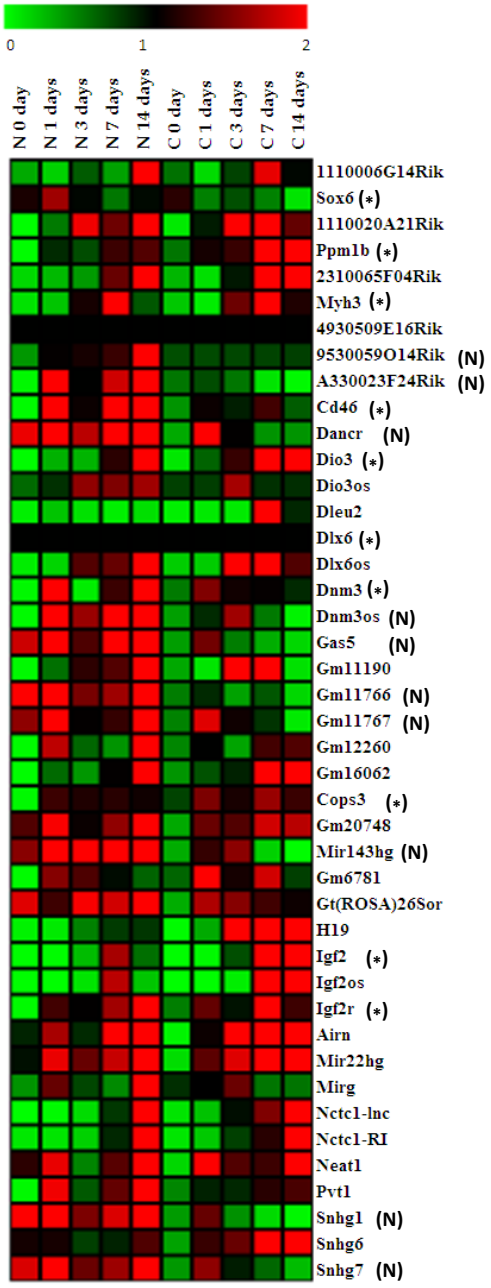


Figure 17. **Expression of lncRNAs during C2C12 differentiation considering RNA in nuclei and cytoplasm separately.**

Heat map reports the results of qRT-PCRs from RNA extracted from C2C12 nuclei and cytoplasm. Red squares indicate high expression level instead green squares indicate low expression level. Column names represent the different sources of RNA:
Columns named **"N (0, 1, 3, 7, 14) days"** are relative to RNA extracted from Nuclei at 0, 1, 3, 7 and 14 days of differentiation;
Columns named **"C (0, 1, 3, 7, 14) days"** are relative to RNA extracted from Cytoplasm at 0, 1, 3, 7 and 14 days of differentiation.
Gene names are marked with **(N)** if the lncRNA resulted to have a nuclear subcellular localization. Protein coding genes are identified with an asterisk **(*)**.

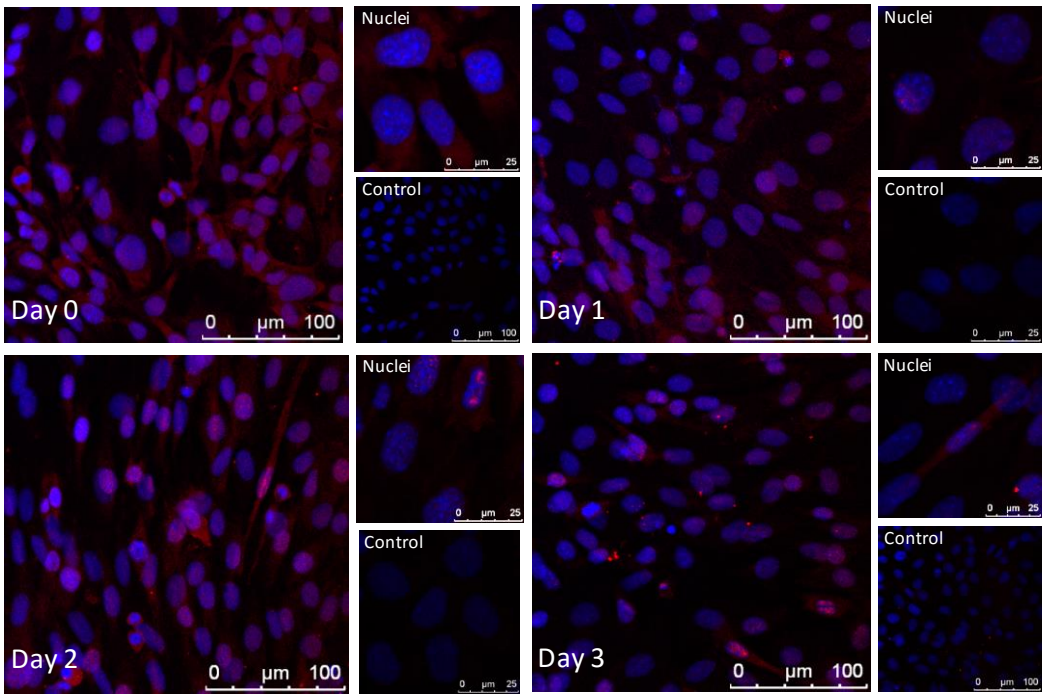


Figure 18. **FISH of *Pvt1* during C2C12 differentiation.**

Pvt1 moves from cytoplasm to nuclei during C2C12 differentiation.

4.4 Whole tissue qRT-PCR analysis

To analyse the expression of lncRNAs in different tissues we extracted total RNA from mouse heart and liver samples and from two different types of muscle (extensor digitorum longus and soleus) and used it in qRT-PCR analysis of previously selected lncRNAs for C2C12 differentiation analysis. Samples were collected from CD1 wild-type mice (Figure 19).

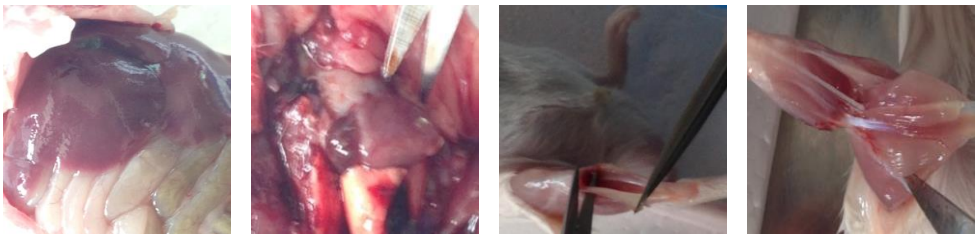


Figure 19. **Tissue collection from CD1 mice.**

From left to right: liver, heart, extensor digitorum longus, and soleus muscles.

We evidenced 19 lncRNAs (*1110006G14Rik, 2310065F04Rik, 4930509E16Rik, 9530059O14Rik, A330023F24Rik, Dleu2, Dlx6os, Gas5, Gm11190, Gm16062, Gt(ROSA)26Sor, H19, Mir22hg, Mirg, Nctc1-Retained Intron, Nctc1, Neat1, Snhg6, Snhg7*) that are more expressed in skeletal muscle compared to the other tissues (Figure 20). Among these lncRNAs, *Dlx6os, Gas5, Gm11190,* and *H19* were found to be more expressed in soleus muscle (Figure 20), suggesting that their potential role could be specific of slow fibers. Only *Gas5* was confirmed to be more expressed in isolated slow myofibers (see paragraph 4.5.2).

Skeletal muscle is a heterogeneous tissue composed by a number of different cell types. LncRNA levels calculated by using RNA extracted from whole muscle can be substantially different from those found using RNA specifically extracted form single or pools of myofibers. In fact, using whole muscle we have an averaged expression influenced by blood, neuronal, and endothelial cells, while, using single myofibers, this aspect is avoided and the calculated expression is due only to the specific analysed cell type.

1110006G14Rik, 2310065F04Rik, and *Snhg6* resulted expressed only in EDL muscle (Figure 20). All these lncRNAs resulted prevalently expressed in fast myofibers which are typical of EDL muscle (see paragraph 4.5.2).

Some lncRNAs were found to have a higher expression in striated muscles (EDL, Soleus and Heart) rather than liver. This group includes *Airn*, *Dancr*, *Gm12260*, *Neat1*, *Igf2os,* and *Snhg7* (Figure 20).

Heart specific lncRNAs are *Dnm3os, Mir143hg, Gm11766* and its antisense *Gm11767*.

Finally, lncRNAs exhibiting elevated expression in liver are *Pvt1*, and *1110020A21Rik.* We found that *Pvt1* has a fast-specific expression in myofibers (see paragraph 4.5.2). In this analysis, the higher expression in EDL (fast muscle) is masked by the greater expression in the liver. *Pvt1* association with hepatocellular carcinoma was already described by *Ding et al.* in liver so the higher expression in that tissue was expected[40].

Interestingly, lncRNA *1110006G14Rik* which maps inside one of *Sox6* introns shows a specific expression in EDL muscle, just like *Sox6* does. The connection between *1110006G14Rik* and *Sox6* will be further discussed in paragraph 4.5.2.

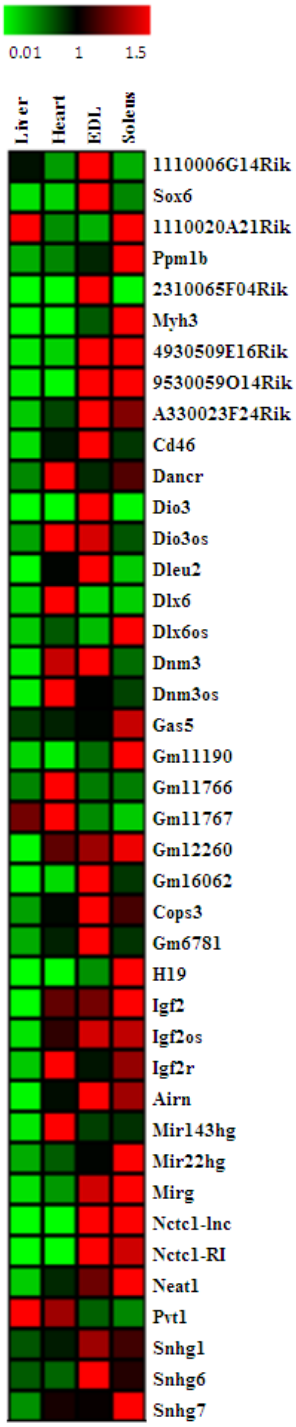


Figure 20. **Heat map evidencing the expression of several lncRNAs and associated coding RNAs in different tissues.**

The heat map shows the tissue specificity of lncRNAs and associated protein-coding genes. Red squares indicate high expression level, instead green squares indicate low expression level of the genes tested in the four tissues indicated on the top of the image.

A similar situation is represented by the lncRNA *1110020A21Rik* and the protein-coding gene *Ppm1b*. Even if in this case the lncRNA is coded in the opposite strand of the gene, both are highly expressed in soleus, a muscle which is prevalently composed by slow oxidative myofibers. *Ppm1b* gene is known to be involved in the control of Pax genes expression (*Pax2*)[41]. Pax proteins induce specific pathways of activation or repression during development and are also

involved in myogenesis and muscle cells maintenance. The repressive function of Pax proteins is mediated by the recruitment of members of the Groucho co-repressor family where *Ppm1b* is an essential component[41]. *Ppm1b* can inhibit H3K4 methylation and therefore gene activation. SINE BASE (http://sines.eimb.ru/) analysis of lncRNA *1110020A21Rik* predicts the presence of SINE sequences; this means that the lncRNA could function as described for *Uchl1-as*.

In contrast with the previous examples, the protein-coding/lncRNA couple Myh3/*2310065F04Rik* presents a discordant expression pattern. The lncRNA is prevalently expressed in EDL, while the associated *Myh3* gene is prevalently expressed in soleus muscle (Figure 20). This could result from an inhibitory function of the lncRNA on the gene.

Among protein-coding genes, *Cops3*, *Igf2* and *Igfr* are found to be more expressed in striated muscles rather than in liver.

Insulin growth factor 2 and its receptor are important for cell growth and for the development of skeletal muscle and heart tissues. The lncRNAs found in antisense to these protein coding genes (*Igf2os* and *Airn*) follow their same expression pattern suggesting that they could share functions or be part of a common regulatory loop.

Summarizing the whole analysis, we can say that the most analysed lncRNA have an elevated expression in the muscle tissue except for two of them (*Pvt1* and *1110020A21Rik*). This could be in accordance with the fact that their selection derive from high expression in the analysis of single myofibers and enhance their importance in the skeletal muscle biology.

4.5 Single Fibers

Skeletal muscle is a heterogeneous tissue composed by a variety of different cell types, but only myofibers are responsible for the contractile capability.

Analysing the expression of lncRNAs in single myofibers allows the removal of non-contractile muscle cells (fibroblasts, endothelial, nervous, blood cells). Also, results can be matched with the type of the single fiber, giving a much more precise characterization of lncRNAs expression in myofibers in relation with metabolisms.

4.5.1 Classification of muscle fibers

A group of isolated myofibers were prepared from EDL and Soleus muscles, as reported in paragraph 3.3, and total RNA was extracted, retro-transcribed and analysed with qRT-PCR (see chapter 3). In order to classify these fibers, we analysed the expression of myosin heavy chain isoforms (*Myh1*, *Myh2*, *Myh4* and *Myh7*). Myofibers showing high expression levels of *Myh7* were defined as "slow myofibers" (Type 1) while those with high expression levels of *Myh4* were defined as "fast myofibers" (Type 2b). After classification, four fibers of each type were selected to be used as biological replicates (Figure 22) to analyse lncRNA fiber specificity. The expression of myosin heavy chain genes in the selected myofibers is showed in Figure 21. All fast myofibers present high expression levels of *Myh4* low expression of any other *Myh* gene; the same profile is found for fibers with high expression of *Myh7* (slow myofibers). These results are in agreement with previous data about muscle fiber classification[16,20–22].

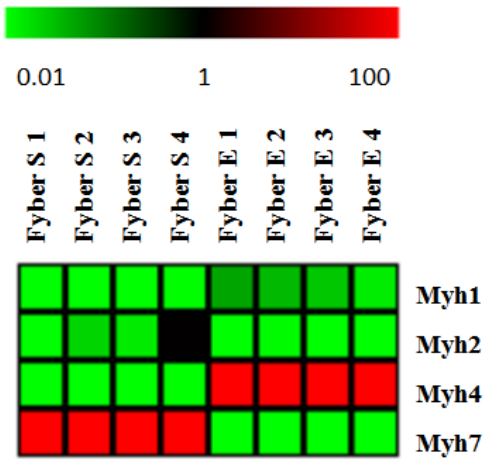


Figure 21. **Myofibers Classification.**

The heat map reports the expression of specific myosin heavy chain genes used to classify myofibers in slow or fast type. For each fiber, the expression of *Myh1*, *Myh2*, *Myh4* and *Myh7* genes was tested. Red squares indicate high expression level, instead green squares indicate low expression level. Fiber S1: fiber slow one; Fiber S2: fiber slow two; Fiber S3: fiber slow three; Fiber S4: fiber slow four; Fiber F1: fiber fast one; Fiber F2: fiber fast two; Fiber F3: fiber fast three; Fiber F4: fiber fast four.

4.5.2 Fiber specificity of lncRNAs

Expression levels of lncRNAs were analysed in the four slow and four fast myofibers chosen as described (see paragraph 4.5.1). *1110020A21Rik*, *Dancr*, and *Gas5* lncRNAs showed higher expression in slow myofibers, while *1110006G14Rik*, *2310065F04Rik*, *Gm16062*, *Pvt1*, and *Snhg6* lncRNAs showed higher expression in fast myofibers (Figure 22). This result is in accordance with the analysis performed on whole muscle (EDL and soleus) for all lncRNAs

(paragraph 4.4). Three protein-coding genes *Dio3*, *Cops3* and *Sox6* were also found to have a higher expression in fast myofibers, while only one, *Myh3*, showed higher expression in slow myofibers (Figure 22). The expression profile of the complete set of all lncRNAs and protein coding genes in single myofibers is shown in Figure 22 using a heatmap, or with histograms for specific lncRNA in *Supplementary Figure 1*. Interestingly, not all fibers respond in the same way about the lncRNA expression. This could be due to cell specificity.

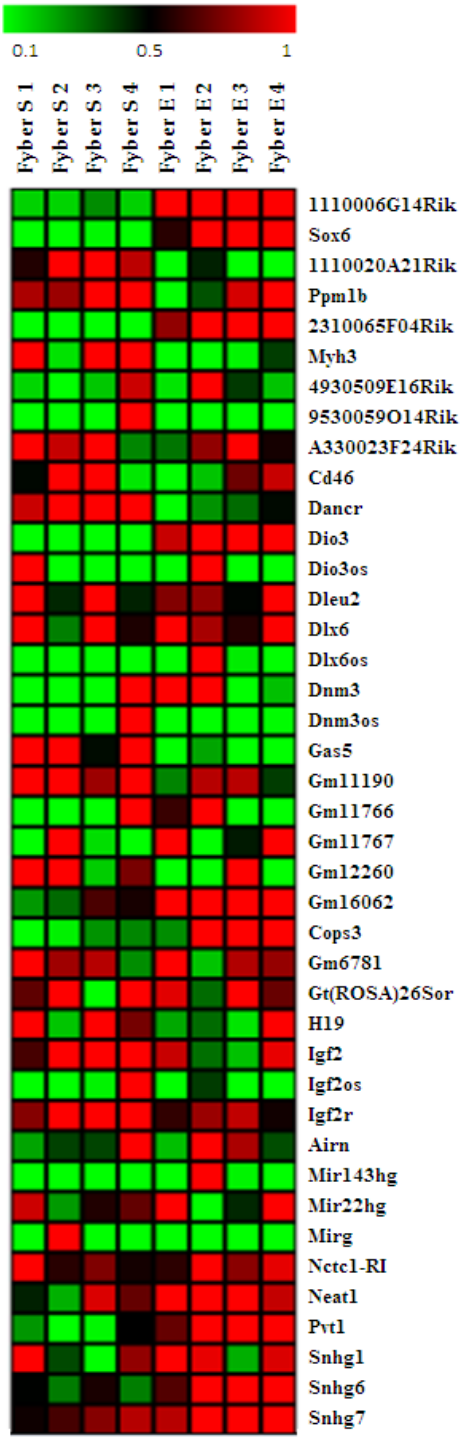


Figure 22. **Fiber specificity of lncRNAs and related coding RNAs.**

The heat map shows the fiber type specificity for lncRNAs and protein-coding genes. Red squares indicate high expression level, instead green squares indicate low expression level. Fiber S1: fiber one, type slow; Fiber S2: fiber two, type slow; Fiber S3: fiber three, type slow; Fiber S4: fiber four, type slow; Fiber F1: fiber one, type fast; Fiber F2: fiber two, type fast; Fiber F3: fiber three, type fast; Fiber F4: fiber four, type fast. The names of genes are reported on the right.

LncRNA *Gas5*, which is found to be prevalent in slow myofibers, is a member of the 5' terminal oligo-pyrimidine gene class. Part of its secondary structure mimics the glucocorticoid response element (GRE): it binds the DNA binding domain of the glucocorticoid receptor, blocking it[42]. Other receptors such as androgen, progesterone and mineralocorticoid receptors can be regulated through *Gas5* binding to GRE. *Gas5* is also known to be involved in the control of proliferation and cell survival of T-cells through mTOR pathway, an important pathway sustaining skeletal muscle hypertrophy. Glucocorticoid receptors are control genes with important functions in development, metabolism and immune response.

As previously described, the lncRNA *1110006G14Rik* gene is coded from one of *Sox6* introns. Both are found highly expressed in fast myofibers, suggesting that the lncRNA may have a role in the maintenance of fast phenotype just like Sox6. Sox6 is known to be involved in the repression of genes that maintain the slow phenotype. *Sox6* is a member of the Sox family of transcription factors that can act either as transcriptional activators or repressors. The transcription factor regulates postnatal muscle fiber-type differentiation, muscle performance and metabolism. Sox6 acts as direct repressor of slow myofibers genes by binding specific sites proximal to loci that encode for myosin heavy and light chains, plus the slow fiber-specific isoform of the sarcoplasmic reticulum Ca^{2+} uptake channel SERCA (ATP2a2 or SERCA2)[42].

Another example of lncRNA/protein-coding gene couple is composed by lncRNA *Gm16062* whose genomic sequence is antisense of protein coding gene *Cops3*. Both genes are expressed in fast fibers, suggesting a possible connection between their transcription. Also, this result is in accordance with previous literature data that describe *Cops3* as predominantly expressed in cardiac and skeletal muscle (Figure 20). COP signalosome complex (CSN) is involved in various cellular and developmental processes. The CSN complex is an essential regulator of the ubiquitin (Ubl) conjugation pathway, important in atrophy. For this reason further analyses will be needed to evaluate the expression of *Cops3* and related lncRNA to better explain their function.

Highly expressed in fast fibers is also *Pvt1*. This gene has been identified as a candidate oncogene. In fact, the over-expression in humans was linked to the development of different types of cancer like, for example, breast and ovarian cancer or acute myeloid leukemia[43].

LncRNA *Pvt1* is involved in the regulation of the proto-oncogene *Myc* to promote tumorigenesis. Its gene is found in a region near *Myc* on chromosome 15 in mouse (chr 8 in *Homo sapiens*) that has numerous physiological functions (e.g. apoptosis regulator, oncogene).

To understand the activity of *Pvt1* we correlated its expression with coding genes (see paragraph 4.7).

4.6 Skeletal muscle atrophy

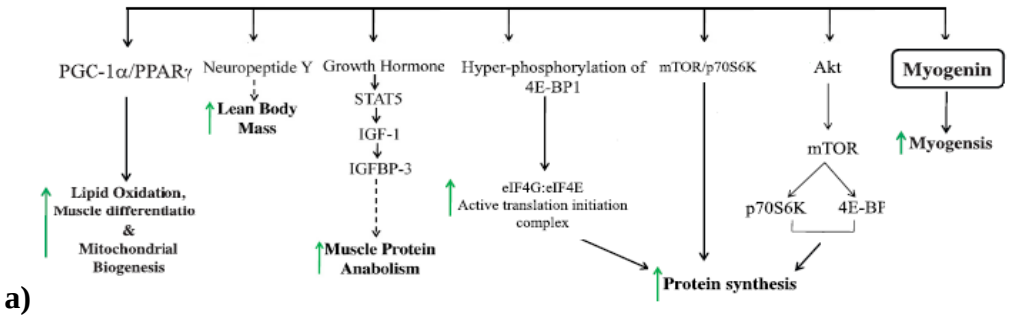
Skeletal muscle atrophy (or wasting) is known to accompany various chronic illnesses as well as the aging process reducing muscle mass and function. In order to study lncRNAs role in muscular atrophy in vivo we chose three different mouse models of atrophy: denervation, Amyotrophic Lateral Sclerosis (ALS) and starvation.

4.6.1 Denervation

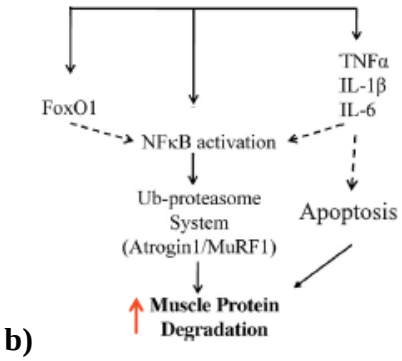
Atrophy of skeletal muscle can arise as a consequence of denervation due to the lack of communication between skeletal muscle and nerve. This process appears to be associated to aging and ALS and can be induced surgically by cutting the sciatic nerve of one limb.

During normal conditions, muscles experiment a constant protein turnover that is altered during atrophy; proteins are not correctly re-established and as a result muscle loses mass and strength. Atrophied muscles are characterized by substantial decrease in muscle fiber cross-section area and number, protein content and overall contractile strength. Increased fatigability and insulin resistance are also observed as a consequence of atrophy.

Analysing the expression of lncRNAs during muscle atrophy could help to identify lncRNAs that can regulate different pathways involved in this condition (Figure 23) and can allow the identification of novel target for more potent drugs contrasting skeletal muscle atrophy.



a)



b)

Figure 23. **Principal pathways governing muscle stability and muscle atrophy.**

a) PGC-1α/PPARy, Growth hormones, mTOR, Akt are the principal pathways by which muscle growths; **b)** when FoxO1, NFkB, TNFα, IL-1 and IL-6 are activated, muscle atrophy begins and a protein degradation event is unrestrainable. Image modified from[44].

Denervation conditions were obtained as described in paragraph 3.11, cutting the sciatic nerve of one limb and using the other as negative control.

Total RNA was extracted from whole muscle samples and analysed via qRT-PCRs to investigate if pathologic conditions can affect lncRNAs expression and how this effect varies during the progression of muscle atrophy (Figure 24).

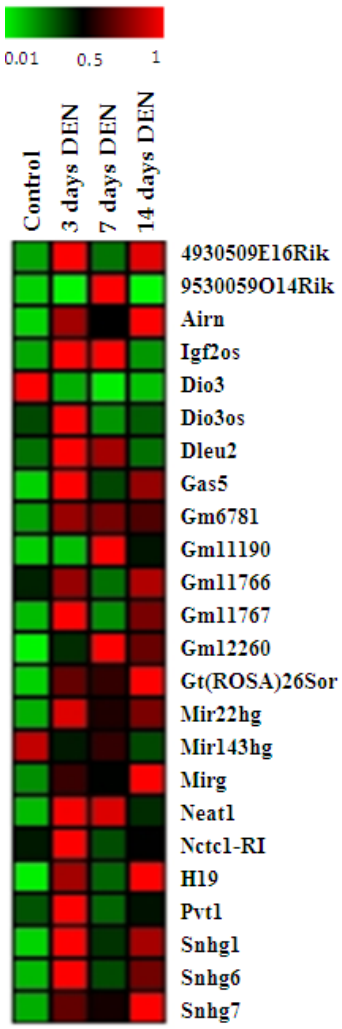


Figure 24. **LncRNA expression during skeletal muscle denervation**.

In the heat map are reported which lncRNAs change their expression pattern during muscle denervation. Red squares indicate high expression level, instead green squares indicate low expression level of the genes tested during atrophy caused by denervation. Control: control, 3 days DEN: RNA purified from muscle after three days of denervation, 7 days DEN: RNA purified from muscle after seven days of denervation, 14 days DEN: RNA purified from muscle after 14 days of denervation. Gene names are on the right, protein coding genes are identified with an asterisk (*).

This analysis demonstrates that lncRNAs can be divided in different groups. All the analysed genes show an alteration of expression after denervation. In fact we evidenced that most lncRNAs are upregulated during the initial phase of denervation (Figure 24). This could indicate that they are involved in the initial process of atrophy when a huge reorganization of the muscle is commenced with the activation and inhibition of many genes.

An interesting lncRNA is *Mirg* because it is highly expressed during denervation and previously showed with skeletal muscle specificity (Figure 20). Similarly, protein-coding gene *Dio3* and lncRNA *Mir143hg* expression is altered after denervation; in this case we evidenced a diminished expression from three to fourteen days after denervation meaning that these genes could be targets of some atrophy related pathway.

When *Dio3* is active it stops growth by blocking thyroid hormones; in denervation *Dio3* is not expressed, probably to induce growth freeing thyroid hormones from

Dio3. Figure 23 shows the effects of growth hormones in the activation of pathways controlling protein anabolism. We evidenced that *Dio3* is expressed in fast fibers and in EDL muscles, so probably its effects on metabolism through thyroid hormones are more important for fast muscles; this could be due to the higher energy requirements of fast myofibers compared to slow myofiber.

As previously evidenced, many genes were found to reach a peak of expression after three days of denervation, like for example lncRNAs *Igf2os*, *Dio3os*, *Dleu2*, *Neat1*, *Nctc1-RI*, *H19* and *Pvt1*. LncRNAs *Igf2os*, *Nctc1-RI* and H19 are found in the same imprinted locus of the mouse genome suggesting a critical role for this locus in the control of skeletal muscle growth and pathophysiology.

4.6.2 Amyotrophic Lateral Sclerosis (ALS)

As previously described, Amyotrophic Lateral Sclerosis, is a detrimental disease caused by the degeneration of lower and upper motor neurons. Recently, it was supposed that this could be a retrograde effect that starts from the loss of Neuro Muscular Junction (NMJ)[45]. Loss of NMJ is also associated with aging. We used SOD1G93A transgenic mice a model for ALS (see paragraph 3.11) and, after total RNA isolation from skeletal muscle samples, we analysed the expression of lncRNAs and protein coding genes via qRT-PCR. We compared the expression of these genes in muscles extracted from one, three and four months aged mice. One month old mice do not manifest any muscular defect due to ALS, while at four months the disease is always visible. As control we used wild-type mice of the same age. The heat map in Figure 25 shows the expression level measured via qRT-PCRs in both transgenic and control mice.

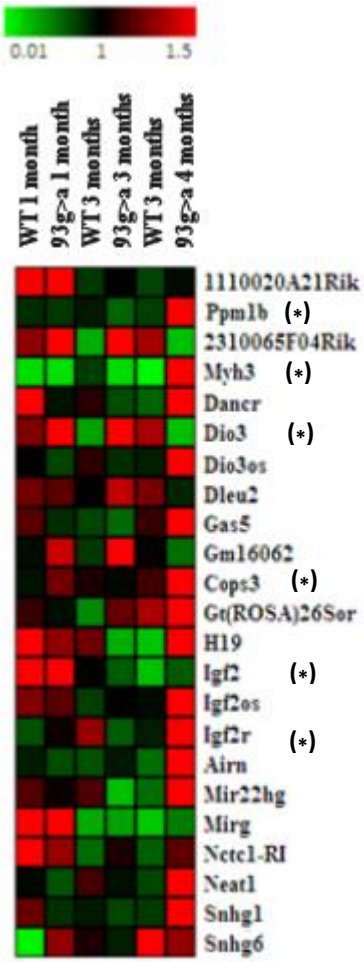


Figure 25. **LncRNAs expression in SOD1G93A transgenic mice.**

Heat map shows the expression of 12 lncRNAs in ALS mice. Red squares indicate higher expression levels while green squares indicate lower expression levels. Each column represents a different sample; WT 1 month: wilde-type mouse at one month of life (control), 93g>a 1 month: mutant mouse one month old (tested mouse), WT 3 months: wilde-type mouse at three months of life (control), 93g>a 3 months: mutant mouse three months old (tested mouse), WT 4 months: wilde-type mouse at four months of life (control), 93g>a 4 months: mutant mouse four months old (tested mouse). Gene names are listed on the right, protein coding genes are marked with an asterisk (*).

We found that the great majority of lncRNAs are more expressed during the latest phase of the disease when the pathologic phenotype is clearly established (Figure 25). This is in accordance with the expression of lncRNAs during denervation. When the muscle is wasted its expression changes radically. As seen in paragraph 4.5.2, part of the secondary structure of *Gas5* mimics the glucocorticoid response element, blocking it. Many pathological conditions characterized by muscle atrophy are often associated with increased levels of circulating glucocorticoids, suggesting that these hormones could play a role in muscle atrophy. Glucocorticoids have an effect on protein synthesis through an anti-anabolic action ultimately preventing the intake of amino acids into the muscle, and a catabolic action leading to the activation of the ubiquitin-proteasome system (UPS) to induce proteolysis[46]. *Gas5* mimics a glucocorticoid molecule, triggering atrophy by promoting muscle proteins degradation. We found that *Gas5* expression increased in G93A transgenic mice compared to control mice (Figure 25), suggesting a possible role for *Gas5* in atrophy (Figures 26 and 27) and possibly in ALS.

Similarly to what already observed in atrophy induced by denervation, the expression of genes found in the Igf2 locus is altered in ALS (Figure 25). The

opposite expression pattern of *Igf2* and *Igf2os* already described earlier is also found in ALS: when *Igf2* is expressed in normal conditions, *Igf2os* is not expressed. This result suggests that the presence of lncRNA *Igf2os* may prevent the expression of *Igf2* transcript. Transcripts of the lncRNA/protein-coding couple *Airn*/*Igf2r* are down-regulated in wild-type mice and over-expressed in mutant mice. The correlation in the expression of these genes was found also in single myofibers and differentiating myoblasts. It is known that in the *Igf2r* imprinted cluster, the paternally expressed *Airn* lncRNA silences the paternal alleles of *Igf2r* to ensure the expression of *Igf2r* from the maternal chromosome.

In this study we also evidenced an increased expression of *Neat1* lncRNA at later stages of ALS. *Neat1* is a nuclear lncRNA that forms the core structural component of paraspeckles and it is also a transcriptional regulator for numerous genes. *Nishimoto et al.*[28], demonstrated that *Neat1* is predominantly expressed in spinal motor neurons in the early phase of ALS pathological process. *Neat1* function was not completely characterised but Authors propose that *Neat1* may act as a scaffold for RNAs and RNA binding proteins in the nuclei of ALS motor neurons, modulating the functions of ALS-associated RNA-binding proteins during the early phase of ALS.

4.6.3 Starvation

Skeletal muscle atrophy can also be induced by nutrient deprivation. Starvation causes decreased protein synthesis and increased protein degradation in both types of skeletal muscle but slow muscles, such as the soleus, are less sensitive to starvation compared to fast muscles. This response is likely linked to the difference in sensitivity between fast and slow muscles to corticosteroids.

Similarly to what already described previously for denervation and ALS, the majority of lncRNAs change their expression in starved mice.

The expression of lncRNA/protein-coding couple Airn/Igf2r is found to correlate during starvation, both genes are found to be increasingly upregulated at 24 and 48 hours of starvation.

On the contrary, the anticorrelation between Igf2os and Igf2 is not confirmed in starvation; conversely of what happens in denervation, we evidenced a coupled increase of Igf2os and Igf2 expression during at 24 and 48 hours of starvation.

This could suggest a different role for lncRNA Igf2os during starvation rather than during denervation induced atrophy and/or ALS.

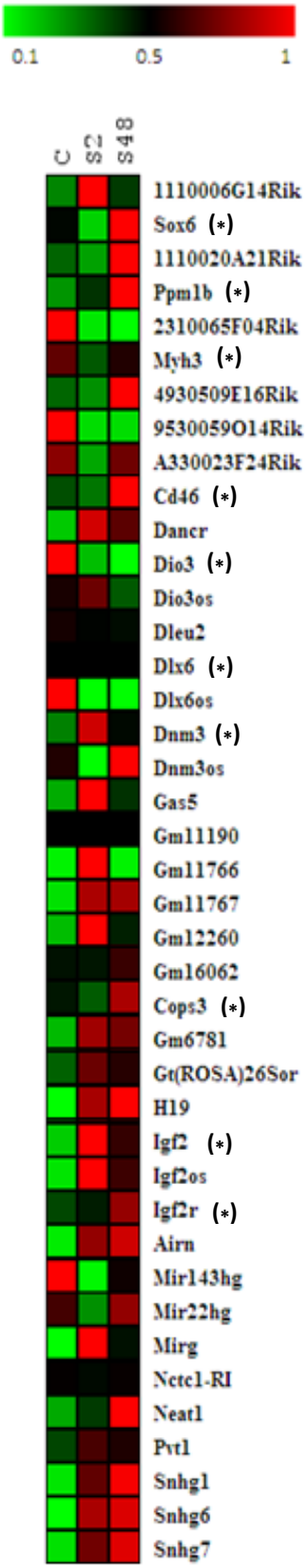


Figure 26. **LncRNA expression during starvation in skeletal muscle**.

In the heat map is reported the expression of lncRNAs in skeletal muscle during starvation. A red square indicates high expression level while a green squares indicate low expression level of the analysed genes. Sources of RNA are are listed on each column, symbols indicate respectively: control RNA extracted from untreated muscle (C), RNA extracted after 24 hours of starvation (S24) and RNA extracted from skeletal muscle after 48 hours of starvation (S48) Gene names are listed on the right, protein coding genes are identified with an asterisk (*).

4.6.4 Myofiber type and skeletal muscle atrophy

Denervation induced atrophy and ALS are known to affect more strongly fast myofibers rather than slow myofibers. We analysed the expression levels of the various myosin heavy chain isoforms to confirm that this pattern was present in our experiments.

As can be seen in Figure 27 the expression of fast specific myosin isoform (Myh4) decreases after denervation. On the contrary, during starvation the expression of this gene does not change.

The different effect on Myh4 is caused by the different processes that lead to atrophy in starvation rather than denervation; in starvation the neuromuscular junction is not impaired, while in denervation the loss of innervation causes fibers to lose their fast phenotype in favour of a more mixed/slow phenotype. These results confirm that denervation induced atrophy disrupts the expression balance of myosin heavy chain isoforms leading to a change in fiber type composition for affected muscles.

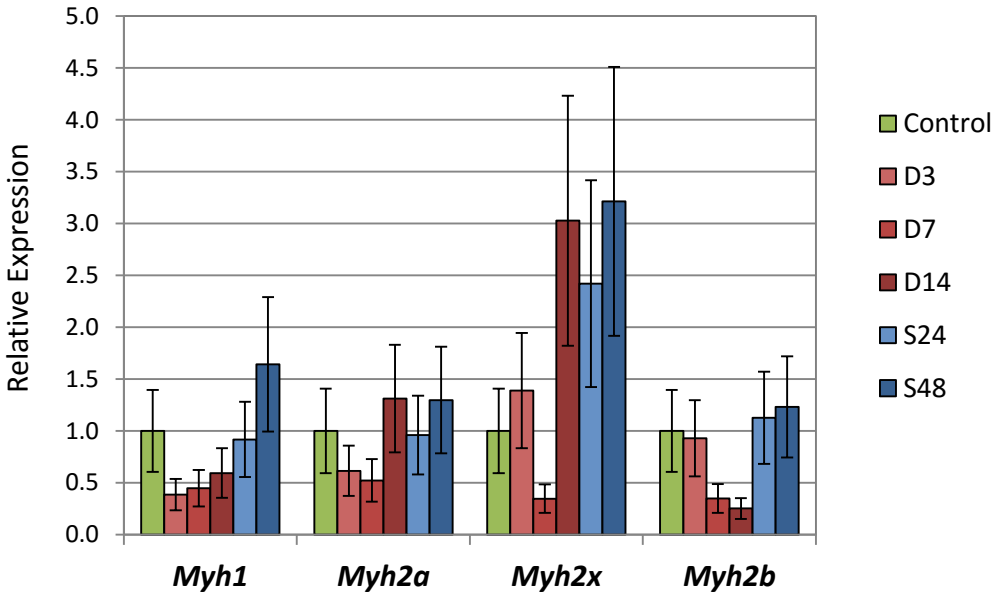


Figure 27. **Myh expression during denervation and starvation**

Histograms represent the expression of myosin heavy chain isoforms during denervation and starvation. Analysed samples are: RNA from muscles extracted 3, 7 and 14 days after denervation (D3, D7 and D14 respectively) and RNA from muscles extracted at 24 and 48 hours of starvation (S24 and S48 respectively). RNA extracted from an untreated muscle was used as reference (Control). Expression levels for each gene are relative to the expression of that gene in the reference sample.

4.7 The function of lncRNAs

Long non-coding RNAs (lncRNAs) are emerging as key regulators of diverse biological processes and diseases. They represent a heterogeneous group of RNAs regulating the expression at the epigenetic, transcriptional or post-transcriptional level[47,48]. However, the function of these molecules are still poorly understood. LncRNAs bind specific DNA regions not only to prevent the binding of transcriptional activators, but also to recruit chromatin remodelling proteins to silence gene expression. LncRNAs can also interact with chromatin modifying proteins *in trans* to epigenetically silence genes at distant loci. LncRNA transcription may open the chromatin structure making it accessible to the transcriptional machinery and enhancing the expression of neighboring protein-coding genes. Similarly, transcription of lncRNAs lying near or inside protein-coding loci can repress transcription by physically preventing the transcriptional machinery to reach protein-coding genes. Since genes with similar expression patterns across multiple conditions may share similar functions[49], or be involved in related biological pathways[50], identifying protein-coding genes that are co-expressed with lncRNAs may help to assign functions to the lncRNA counterpart. By analysing lncRNA-mRNA co-expression pattern, *Guttman et al.* identified several sets of mouse lncRNAs associated with protein-coding gene sets of distinct GO functional categories[51]. In addition, *Guo et al.* and *Liao et al.* independently in two recent studies, constructed a mouse co-expressed lncRNA-mRNA network using mouse microarray data and assigned functions to 340 and 1,625 mouse lncRNAs[52,53].

Here we analysed co-expressed genes of all lncRNA identified and, as particularly interesting examples, we report results for *Pvt1* and *1110006G14Rik*. As previously discussed (paragraph 4.5.2), *Pvt1* is a well-studied lncRNA in tumors because of its genome location nerby to *Myc* and its ability to interact with the oncogene. It was evidenced the ability of *Pvt1* to influence cell cycle, tissue morphogenesis and development[54]. We evidenced that genes having a correlated expression with *Pvt1* were involved in microtubule organization, cell division, M phase progress and morphogenesis (*Supplementary table 4*). Our data support a *Pvt1* function in cell cycle regulation also in skeletal muscle. It is an important aspect of skeletal muscle development and repair because both satellite cells, involved in muscle repair, and C2C12 cells, a model for muscle development, should withdraw from cell cycle to fuse into myotubes. Regarding *1110006G14Rik*, that is up regulated in fast muscles, we evidenced that its expression is related to genes involved in carbohydrate methabolism and in particular to specific muscle genes (*Supplementary table 5*). We evidenced that this lncRNA could be related to Sox6 function because of the genomic colocalization of those genes and a correlated expression pattern. Sox6 specify for fast myofibers by repressing genes of slow phenotype[42]. We evidenced that

1110006G14Rik expression also correlates with the expression of multiple protein coding genes involved in carbohydrate metabolism, which is specific of fast myofibers. This correlation could be the result of either an action of *1110006G14Rik* on those genes or the regulation of *1110006G14Rik* and the correlated genes by the same transcription factors. Moreover, the *Myh4* gene, specific for the fast myofibers, has a correlated expression with *1110006G14Rik*. Other genes, specific for fast myofibers exhibited a related expression with *1110006G14Rik* (*Supplementary table 6*). Among muscle specific genes there are parvalbumin, a small, stable protein containing EF-hand type calcium binding sites, that is localised in fast-contracting muscles[55] and calsequestrin 1 also prevalently expressed in EDL[56] indicating that calcium regulation is important in muscle fibre specification[55].

Chapter 5 – Conclusions

Whole genome technologies allow the identification of non-coding RNAs as important regulators of gene expression. Long non-coding RNAs can participate in the regulation of different cellular processes such as stem cell pluripotency, immune response, neural processes, and cell cycle regulation.

We investigated the role of lncRNAs in skeletal muscle using a genome wide approach to describe their subcellular localization and specific gene expression to describe their tissue specificity, fiber type specificity and involvement in C2C12 differentiation. We found that most analysed lncRNAs are expressed specifically in striated muscles and that many showed a fiber specific expression.

Genome wide approach allowed the identification of 655 cytoplasmic and 481 nuclear lncRNAs, confirming the unexpected abundance of cytoplasmic lncRNAs already described by *van Heesch, S. et al*[36] in human colon cancer cells. Moreover, we evidenced that most of them are expressed in the same subcellular compartment in myofibers derived from Soleus, EDL and Tibialis Anterior with a small number of lncRNAs that are instead specific for each muscle.

Genome-wide lncRNA subcellular localization was then confirmed using Fluorescence In Situ Hybridization (FISH) to target specific lncRNAs. Notably, *Gt(ROSA)26Sor*, *Neat1*, *Mir143hg* and *Nctc1-RI* evidenced nuclear localization in skeletal muscle while *H19* and *Airn* were found in both compartments, as predicted from microarray analysis. This result can sustain the validity of our approach to dissect lncRNA subcellular localization in a genome wide fashion.

QRT-PCR test not only showed that most of the target lncRNAs were specifically expressed in striated muscle, but also identify which lncRNA are specifically expressed in extensor digitorum longus (fast glycolytic muscle) and in soleus (slow oxidative muscle).

To further confirm lncRNA relation with muscle metabolism we analyzed RNA extracted from both fast and slow single myofibers. In fact, RNA extracted from whole muscle is a mixture of RNA from conctractile cells and from epithelial, neuronal, blood cells present in the muscle. The analysis of single myofibers showed that *1110020A21Rik*, *Dancr*, and *Gas5* are over expressed in slow myofibers, while *1110006G14Rik*, *2310065F04Rik*, *Gm16062*, *Pvt1*, and *Snhg6* are over expressed in fast myofibers. We also identified three protein-coding genes specifically expressed in fast myofibers (*Dio3*, *Cops3* and *Sox6*) and one in slow myofibers (*Myh3*).

The identification that *Sox6* is specific for fast fibers and EDL can be considered as a validating data since it is known to be enriched in fast myofiber. It, in fact, exert its action repressing slow muscle gene[41].

Interestingly, lncRNAs lying in a common genomic position with coding RNAs have, in some cases, a correlated expression that could result from a functional relationship or from the sharing of a common regulatory expression pattern. An example of this behavior is the lncRNA/protein-coding couple *1110006G14Rik/Sox6;* in this case the lncRNA is located within an intron of the protein-coding gene, and the two transcripts show the same expression profile both in tissues and single myofibers. Since *Sox6* is involved in muscle fiber specification, it could be very interesting to analyze the involvement of this lncRNA in the same process. There is little evidence of lncRNA involvement in fiber specification and therefore in metabolism regulation. We correlated the expression of whole coding expression genes with *1110006G14Rik* evidencing that correlated genes are involved in carbohydrate metabolism, that is specific for fast myofibers, and peculiar fast muscle specific genes. These results could sustain the involvement of this lncRNA in fiber specification.

We then used qRT-PCR to measure lncRNA expression and subcellular localization in differentiating myoblasts.

Several lncRNA analysed in cell cultures were found to have the same localization previously identified in skeletal muscle myofibers. This supports the use of C2C12 cells to study the function of lncRNA *in-vitro* altering their expression, supposing that the observed effects compare the function of lncRNA *in-vivo*.

Notably, lncRNA *Pvt1* that is known to be involved in cell cycle regulation, was found to switch its localization from cytoplasmic to nuclear after the induction of myoblast differentiation, leading us to believe that it could possibly have a role in muscle development and/or regeneration. *Pvt1* could function as miRNA sponge or miRNA precursor when localized in the cytoplasm and as chromatin regulator when localized in nuclei. A correlation analysis with coding genes demonstrated that *Pvt1* correlates with genes involved in cell cycle regulation supporting the idea that it could regulate cell cycle during differentiation. In fact, it is important that cells withdraw from cell cycle to allow cell fusion and the formation of myotubes.

Since evidences of lncRNA involvement in different diseases are already clear, we decided to monitor their expressions during atrophy. We chose three different models of atrophy: denervation and starvation induced atrophy plus the most commonly used transgenic mouse model for ALS.

Most lncRNAs resulted activated in all atrophy models with *Pvt1* showing an early activation during denervation. Moreover, Myh4 expression which is typical of fast myofibers was found to decrease during denervation, confirming the stronger effect of denervation on fast myofibers.

References

1. Romito, A. & Rougeulle, C. Origin and evolution of the long non-coding genes in the X-inactivation center. *Biochimie* **93,** 1935–1942 (2011).
2. Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15,** 7–21 (2013).
3. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **20,** 300–307 (2013).
4. Neguembor, M., Jothi, M. & Gabellini, D. Long noncoding RNAs, emerging players in muscle differentiation and disease. *Skelet. Muscle* **4,** 8 (2014).
5. Liu, G., Mattick, J. & Taft, R. J. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle* **12,** 2061–2072 (2013).
6. Barlow, D. P. & Bartolomei, M. S. Genomic Imprinting in Mammals. *Cold Spring Harb. Perspect. Biol.* **6,** a018382–a018382 (2014).
7. Maclary, E., Hinten, M., Harris, C. & Kalantry, S. Long nonoding RNAs in the X-inactivation center. *Chromosome Res.* **21,** 601–614 (2013).
8. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22,** 1775–1789 (2012).
9. Anderson, D. M. *et al.* A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* **160,** 595–606 (2015).
10. Sun, M. & Kraus, W. L. From Discovery to Function: The Expanding Roles of Long NonCoding RNAs in Physiology and Disease. *Endocr. Rev.* **36,** 25–64 (2015).

11. GENCODE - Home page. at <http://www.gencodegenes.org/>
12. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
13. Ensembl genome browser 83. at <http://www.ensembl.org/index.html>
14. Nie, M., Deng, Z.-L., Liu, J. & Wang, D.-Z. Noncoding RNAs, Emerging Regulators of Skeletal Muscle Development and Diseases. *BioMed Res. Int.* **2015,** 1–17 (2015).
15. Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491,** 454–457 (2012).
16. Scott, W., Stevens, J. & Binder-Macleod, S. A. Human skeletal muscle fiber type classifications. *Phys. Ther.* **81,** 1810–1816 (2001).
17. *Pearson Education Inc., publishing as Pearson Benjamin Cummings.*
18. *Essential cell biology*. (Garland Science, 2009).
19. Standring, S. New focus on anatomy for surgical trainees. *ANZ J. Surg.* **79,** 114–117 (2009).
20. Augusto, V., Padovani, C. R. & Campos, G. E. R. Skeletal muscule fiber types in C57BL6J mice. *Braz J Morphol Sci* **21,** 89–94 (2004).
21. Kammoun, M., Cassar-Malek, I., Meunier, B. & Picard, B. A simplified immunohistochemical classification of skeletal muscle fibres in mouse. *Eur. J. Histochem.* **58,** (2014).
22. Agbulut, O., Noirez, P., Beaumont, F. & Butler-Browne, G. Myosin heavy chain isoforms in postnatal muscle development of mice. *Biol. Cell Auspices Eur. Cell Biol. Organ.* **95,** 399–406 (2003).
23. Gilbert, S. F. *Developmental biology*. (Sinauer Associates, 2003).

24. Ohtake, Y. Multifunctional roles of MT1-MMP in myofiber formation and morphostatic maintenance of skeletal muscle. *J. Cell Sci.* **119,** 3822–3832 (2006).
25. Mousavi, K. *et al.* eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol. Cell* **51,** 606–617 (2013).
26. Lu, L. *et al.* Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. *EMBO J.* **32,** 2575–2588 (2013).
27. Orsini, M. *et al.* Amyotrophic lateral sclerosis: new perpectives and update. *Neurol. Int.* **7,** (2015).
28. Nishimoto, Y. *et al.* The long non-coding RNA nuclear-enriched abundant transcript 1_2 induces paraspeckle formation in the motor neuron during the early phase of amyotrophic lateral sclerosis. *Mol. Brain* **6,** 31 (2013).
29. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci.* **98,** 5116–5121 (2001).
30. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34,** 374–378 (2003).
31. Rosenblatt, J. D., Lunt, A. I., Parry, D. J. & Partridge, T. A. Culturing satellite cells from living single muscle fiber explants. *Vitro Cell. Dev. Biol. - Anim.* **31,** 773–779 (1995).
32. Shefer, G., Van de Mark, D. P., Richardson, J. B. & Yablonka-Reuveni, Z. Satellite-cell pool size does matter: Defining the myogenic potency of aging skeletal muscle. *Dev. Biol.* **294,** 50–66 (2006).
33. Calderón, J. C., Bolaños, P. & Caputo, C. Myosin heavy chain isoform composition and Ca(2+) transients in fibres from enzymatically dissociated

murine soleus and extensor digitorum longus muscles. *J. Physiol.* **588,** 267–279 (2010).

34. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).
35. Soares, R. J. *et al.* Involvement of MicroRNAs in the Regulation of Muscle Wasting during Catabolic Conditions. *J. Biol. Chem.* **289,** 21909–21925 (2014).
36. van Heesch, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **15,** R6 (2014).
37. Volpi, E. & Bridger, J. FISH glossary: an overview of the fluorescence in situ hybridization technique. *BioTechniques* **45,** 385–409 (2008).
38. Trask, H. W. *et al.* Microarray analysis of cytoplasmic versus whole cell RNA reveals a considerable number of missed and false positive mRNAs. *RNA* **15,** 1917–1928 (2009).
39. The Oncogenic Potential of MYC Requires a Long Noncoding RNA. *Cancer Discov.* (2014). doi:10.1158/2159-8290.CD-RW2014-142
40. Ding, C. *et al.* Long non- coding RNA PVT1 is associated with tumor progression and predicts recurrence in hepatocellular carcinoma patients. *Oncol. Lett.* (2014). doi:10.3892/ol.2014.2730
41. Abraham, S. *et al.* The Groucho-associated Phosphatase PPM1B Displaces Pax Transactivation Domain Interacting Protein (PTIP) to Switch the Transcription Factor Pax2 from a Transcriptional Activator to a Repressor. *J. Biol. Chem.* **290,** 7185–7194 (2015).

42. Quiat, D. *et al.* Concerted regulation of myofiber-specific gene expression and muscle performance by the transcriptional repressor Sox6. *Proc. Natl. Acad. Sci.* **108,** 10196–10201 (2011).
43. Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA Gas5 Is a Growth Arrest- and Starvation-Associated Repressor of the Glucocorticoid Receptor. *Sci. Signal.* **3,** ra8–ra8 (2010).
44. Dutt, V., Gupta, S., Dabur, R., Injeti, E. & Mittal, A. Skeletal muscle atrophy: Potential therapeutic agents and their mechanisms of action. *Pharmacol. Res.* **99,** 86–100 (2015).
45. Dupuis, L. & Loeffler, J.-P. Neuromuscular junction destruction during amyotrophic lateral sclerosis: insights from transgenic models. *Curr. Opin. Pharmacol.* **9,** 341–346 (2009).
46. Schakman, O., Gilson, H. & Thissen, J. P. Mechanisms of glucocorticoid-induced myopathy. *J. Endocrinol.* **197,** 1–10 (2008).
47. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44,** 667–678 (2011).
48. Simon, M. D. *et al.* The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20497–20502 (2011).
49. Lee, H. K., Hsu, A. K., Sajdak, J., Qin, J. & Pavlidis, P. Coexpression analysis of human genes across many microarray data sets. *Genome Res.* **14,** 1085–1094 (2004).
50. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95,** 14863–14868 (1998).

51. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458,** 223–227 (2009).
52. Guo, X. *et al.* Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res.* **41,** e35 (2013).
53. Liao, Q. *et al.* Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.* **39,** 3864–3878 (2011).
54. Colombo, T., Farina, L., Macino, G. & Paci, P. PVT1: A Rising Star among Oncogenic Long Noncoding RNAs. *BioMed Res. Int.* **2015,** 1–10 (2015).
55. Naya, F. J. *et al.* Stimulation of slow skeletal muscle fiber gene expression by calcineurin in vivo. *J. Biol. Chem.* **275,** 4545–4548 (2000).
56. Murphy, R. M., Larkins, N. T., Mollica, J. P., Beard, N. A. & Lamb, G. D. Calsequestrin content and SERCA determine normal and maximal Ca2+ storage levels in sarcoplasmic reticulum of fast- and slow-twitch fibres of rat. *J. Physiol.* **587,** 443–460 (2009).

Supplementary materials

Supplementary Table 1. **Primers for qRT-PCR and construction of FISH probes**

Gene name	Primers	qRT-PCR	FISH
Aim	F	CTCCAGAAACTAGACACTACAGACCA	GCTCCAATTTCCCATGTCTC
	R	TGTAAAGTCTTCAAGGATCTAGGGC	TGCCTCTTTCCCACTTCTGT
Cd46	F	CCCTATTTCAGATGCTGGTTG	-
	R	TGCCAAATGAAGGGTCTTGT	-
Cd46	F	TGCATTAGAAAGCAGCAGTAGC	-
	R	ATTCCTTCACGGGGACTAGG	-
Cops3	F	ATTCAACCAACAACCCGTCT	-
	R	AGTGAAGGTCTCGCTGTGCT	-
Dancr	F	CCCTTCTTCATGTCCCACTG	TGACTGAATGGCTCCAAGG
	R	AAACGGCGAGCATGTCATAG	AGAATTGACACAGGAAGCCTTTA
Dio3	F	CATCCGCAAGCATTTCC	CAAATGCTCCAAAGAAAGTCAA
	R	GCATCTCCTCGCCTTCAC	CGAAGTCCATCCCTTACCA
Dio3os	F	TCCTTCTGGTCCTGCTCTATG	CTGGGGGCAAGGGAGTAA
	R	TTTCCCAAGTGTCCCATATCA	ATCCTGAGGTGTGGGGTGTA
Dleu2	F	GATGTTGGGGCGGAGAG	-
	R	TCAGCGACGGAGGAAGAC	-
Dlx6	F	CTTAGGACTGACACAAACACAGGT	CACCATCACCACCACCAG
	R	GGTCACTCTCGTGTGGGTTACTAC	AGGCAGAAACGTCCCACA
Dlx6os	F	GCCCTCTGCTTCACTTCC	GCCCTCTGCTTCACTTCCA
	R	GACCTGGCTCCCTCTCTCT	CCCGCTCACTTCCTCCA
Dnm3	F	CCAAACCAGAAACCCTACCA	CGAAGTCAGCAACACAATGAA
	R	CGGGGAGAGAAGAGGAATACA	CTTGAGCAGTGGAATGAGGA
Dnm3os	F	GGCACACTTGAGATTTATCCTTC	ACACCTTTGCTGCCTCCT
	R	CGACTACTACTGTTTGTTGTTTCCA	CCCTGACCACTTCTCCAGTT
Gas5	F	CTGGCTTGCTTGGGTAAGA	-
	R	GCATGTTCAAAAGCTAAATGTTATG	-
Gm6781	F	GGAGGTGTGGGGCAGATAC	-
	R	GCTGTAGGGCGTTGGTTATG	-
Gm11190	F	CCCTTCCTCATTTCCTCCAG	-
	R	ATCTGCTCCCGATTCACTCC	-
Gm11766	F	CTGAACTCGGGCTATGGTCA	-
	R	GTTGGAAGGGGACAGGAGAG	-
Gm11767	F	GGTGTTTTTCCTGCCTTCCT	-
	R	TCCTGCCTAAGTGTCTGAGCA	-
Gm12260	F	GAGATCGCCCAGGACTTCA	-
	R	GCACAGGTTCGTGTCTTCG	-
Gm16062	F	CCCTGTCCCTACCGAAGAG	-
	R	GTCCCAAGAACTCCCAGAATC	-
Gt(ROSA)26or	F	TTTGTATGTGAGGATAAAGGTGTTTG	CAGGACAGTGCTTGTTTAAGG
	R	TTTGGAACTTTGGGGAAATGT	CTCGCACCAACACAAAAGT
H19	F	GAGACCACCACCCACATCAT	CCTCCCCCTACCTTGAACC
	R	GGAGGAAGAAGAAAAAGACAGGA	CAGACGGAGATGGACGACA
Igf2	F	CGGTGTGTGTCAGCCAAG	AGGTGGAAGAATCAGGAGGA
	R	TGAGTTTCTGTCCAATGTTCCA	CCCCAAATGCTCAGAAGG

(continues inthe next page)

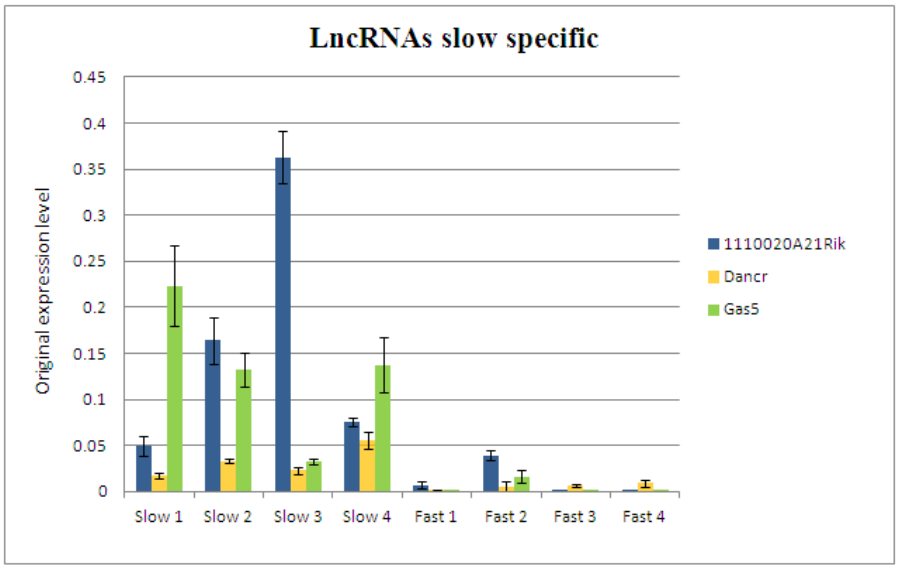
(continues from the previous page)

Gene name	Primers	qRT-PCR	FISH
Igf2os	F	GCTGGAAGGGGCTGAGTC	TCCCAGTTGTCCTTTTGTCC
	R	CAGTGGTGGAGAGCAGAAGC	ACCCTCTAATGCCCCCATAC
Igf2r	F	GTTGGTGTAGGGCCAGTGTT	AAACCCTTGCTGCTTTAGTCC
	R	AAATTCTGCGGGGTACTTTG	TTTCTCTGACACCTCAACTCCA
Mir22hg	F	ATGGCTCTGCTGTCCTCATC	CGCTGGGAAGAGACAGAG
	R	GGTCCTCCACTGTCTTGTGC	ACCTACCAACTGAGCTACAACC
Mir143hg	F	CAGTCACCACGAAGCAAAGG	AAGCAATAACACCCACAACC
	R	ACCCCCAAACCACCTACTCA	TAGAACCTGCCGATGACTTT
Mirg	F	CCACCCCAGAGCCTTGTATT	-
	R	AAGAGCAGAAACCCCTCCTTC	-
Myh3	F	CGCCTGAAGAAGAAGATGG	-
	R	CGGAGGTGCTTGATGGT	-
Neat1	F	CATTGTGGGTTTGGCTTGA	TCCAATGCTGCTATCTAAAGG
	R	GACAGTTTAACAGCTTCCCCTCT	GTAAAGGGGAGGAAAATGGT
Nctc1-lnc	F	TGAGAAGGAAGCAGTGACAAGTA	CAACCACCACACCAGGACTA
	R	AAGGCAGAGTTGGGCTATCC	GGCATTGGAGAGGAGCATAGA
Nctc1-RI	F	CAGCCCAGAGTTTTGTAAGCC	CTTTGAGGGTCTGGGAAGGT
	R	CAAATAACACCTGGAGAGGAAGG	TGCCTGGTGTGTGTTCTCTC
Ppm1b	F	AGGGTTAGAGATTACTGGTTGGA	-
	R	GGTATGAGGATGGAAAATAAGCA	-
Pvt1	F	CTTGTATGGGTGGTGGCTTT	CTTAGTGAATGCTGGCTTGTG
	R	CCAGGGAGAGAGTGGTGTG	TGGCTACTGAAAGAAGGAAGGT
Snhg1	F	AGGGATCATTTTTGTGCGTAGA	AGAACAAAGGAGCCGTACCA
	R	GCTCATTCTTTTCCTCAGACCTG	TTGCTAGAACAAACACAGAGTACAAG
Snhg6	F	ATGTAGGTGGCTGTAGTGGATG	-
	R	CAATAAAGGGTTTTTAGACATCACAG	-
Snhg7	F	GATTGGCTTGGGTTTGACTCT	TGGCGGCTCTATGTTTGTATC
	R	GGGTCCTCCATCCAGTGTTT	CCAGTTATCACCACCCAACTC
Sox6	F	CTTTGTCCCCATCCTGACC	-
	R	CTTTTCCATCCCTCCCAGAA	-
A330023F24Rik	F	TTGATGAGTGATGTTGAGGTTTG	ATCCTTGCTGGTTCCTTCC
	R	TTAGCCCAAGGGTGAGGAG	AAATAACATCTGCCACACATCC
1110006G14Rik	F	TCGGTTTGTGTACTGCTGAGA	-
	R	GGGACTCCAGGATCTTCGTA	-
1110020A21Rik	F	GGTAGAGTCCTCCGTCAGTCA	-
	R	GCCATTTGAACCAATCAGAAA	-
2310065F04Rik	F	GTGAGCGTGAGATGTTCGTG	ACCCCTCTCCCCTATTCTCC
	R	AAAGTGCATCCCCTCTGCTT	CACGAACATCTCACGCTCAC
4930509E16Rik	F	AGGACAACTGAAGAGACTACCCAAT	-
	R	TCTGACAAGACTGAAAACAACAAGA	-
9530059O14Rik	F	CCCATTCAAAATCAATACACTTACA	CCTGACTTGTATATTCTGATTCCTTG
	R	TTCTTCTCCTCCACCACACC	TGTTCCTAGTCCAGTTCTTTGTTG
B2m	F	CCGTCTACTGGGATCGACAC	-
	R	GCTATTTCTTTCTGCGTGCAT	-
Tbp	F	TGCTGTTGGTGATTGTTGGT	-
	R	AACTGGCTTGTGTGGGAAAG	-
Tnrc1	F	TCCAATGTGGTGTTCCTTGA	-
	R	GGCTTCAAGCTTTTCCTTGTT	-

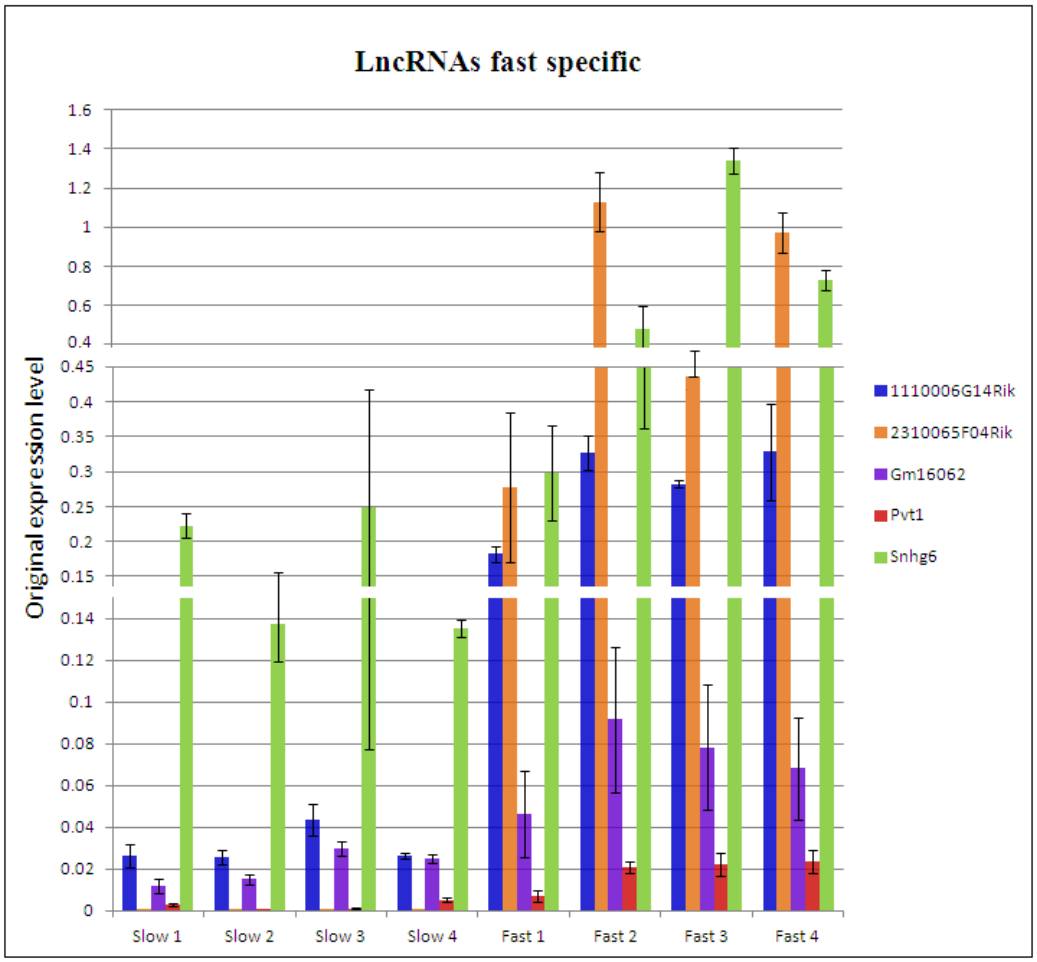
Supplementary Table 2. **Restriction Enzymes**

		Enzyme	Conc.[U/ml]	Sequence	Cut site	Overhang
Neat1	control	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT
	probe	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
Igf2os	control	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT
	probe	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
Airn	control	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
	probe	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT
H19	control	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
	probe	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT
Mir143hg	control	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
	probe	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT
Gt(ROSA)26Sor	control	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT
	probe	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
***Nctc1*-RI**	control	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
	probe	Spe1	10U/ml	ACTAGT	A/CTAGT TGATC/A	5′ - CTAG
Mir22hg	control	Spe1	10U/ml	ACTAGT	A/CTAGT TGATC/A	5′ - CTAG
	probe	Xba1	20U/ml	TCTAGA	T/CTAGA AGATC/T	5′ - CTAG
Pvt1	control	Noti	10U/ml	GCGGCCGC	GC/GGCCGC CGCCGG/CG	5′- GGCC
	probe	HindIII	20U/ml	AAGCTT	A/AGCTT TTCGA/A	5′ - AGCT

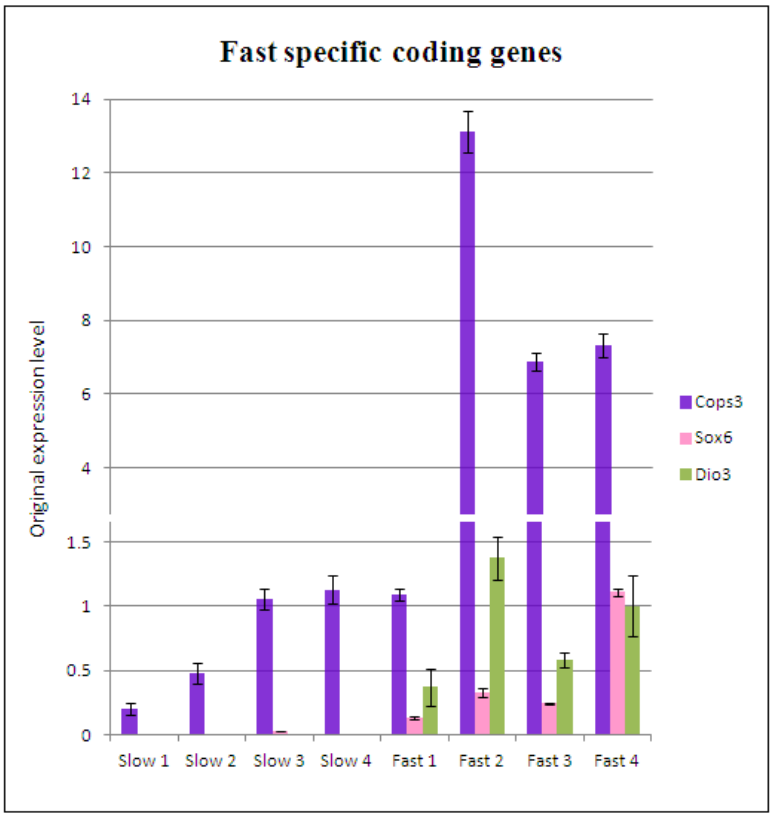
Supplementary Figure 1. **Histogram of slow specific lncRNAs.**



Supplementary Figure 2. **Histogram of fast specific lncRNAs.**



Supplementary Figure 3. **Histogram of fast protein-coding genes.**



Supplementary Figure 4. **Genes having correlated expression with Pvt1** *(pvalue < 0.05)*

Annotation Cluster	**Enrichment Score: 0.9785600266815854**
Term	**Genes**
GO:0005856 cytoskeleton	SS18, FMN2, FERMT2, GM5593, KIF19A, TGM3, MYO19, TTLL7, CDC27, GABARAP, ARHGAP26, SNTA1
GO:0015630 microtubule cytoskeleton	SS18, GM5593, KIF19A, TTLL7, CDC27, GABARAP
GO:0007017 microtubule-based process	SS18, FMN2, KIF19A, GABARAP
GO:0005874 microtubule	SS18, KIF19A, CDC27, GABARAP
GO:0044430 cytoskeletal part	SS18, FERMT2, GM5593, KIF19A, MYO19, TTLL7, CDC27, GABARAP
GO:0005815 microtubule organizing center	GM5593, TTLL7, CDC27

Annotation Cluster	Enrichment Score: 0.7676543517241499
Term	**Genes**
GO:0001763 morphogenesis of a branching structure	FGFR1, GCM1, CD44, PHB2
GO:0022612 gland morphogenesis	FGFR1, CD44, PHB2

Annotation Cluster	Enrichment Score: 0.6069845696013708
Term	**Genes**
GO:0007017 microtubule-based process	SS18, FMN2, KIF19A, GABARAP
GO:0000226 microtubule cytoskeleton organization	SS18, FMN2, GABARAP
GO:0007010 cytoskeleton organization	SS18, FMN2, GABARAP

Annotation Cluster	Enrichment Score: 0.5412778122684933
Term	**Genes**
GO:0051301 cell division	FMN2, DSN1, GM5593, PELO, CDC27
cell division	DSN1, GM5593, PELO, CDC27
GO:0007049 cell cycle	FMN2, GM13337, DSN1, GM5593, PELO, GM3852, CDC27
GO:0022402 cell cycle process	FMN2, GM13337, DSN1, GM5593, CDC27
GO:0000279 M phase	FMN2, DSN1, GM5593, CDC27
GO:0000280 nuclear division	DSN1, GM5593, CDC27
GO:0007067 mitosis	DSN1, GM5593, CDC27
GO:0022403 cell cycle phase	FMN2, DSN1, GM5593, CDC27
GO:0000087 M phase of mitotic cell cycle	DSN1, GM5593, CDC27
GO:0048285 organelle fission	DSN1, GM5593, CDC27
GO:0000278 mitotic cell cycle	DSN1, GM5593, CDC27
cell cycle	DSN1, GM5593, PELO, GM3852

Supplementary Figure 6. **Genes that correlate with 1110006G14Rik** *(pvalue < 0.05)*

Annotation Cluster	**Enrichment Score: 6.146094352993051**
Term	**Genes**
GO:0006006 glucose metabolic process	ALDOA, GM14148, LDHA, GM4217, GM5787, GM2308, PHKA1, GM15191, GPI1, GM8318, TPI1, GM7784, ENO3, GM3839, GM8174, GM9034, AGL, GM7251, GPD1, GM7611, GM10313, GM6498, PGM2, GM5732, GM9568, GM11953, GM3809, GM7286, GM9061
glycolysis	ALDOA, GM14148, LDHA, GM7611, GM10313, GM4217, GM5787, GM2308, GM15191, GM6498, GPI1, GM8318, GM5732, GM7784, TPI1, GM11953, GM8174, GM9568, GM3839, ENO3, GM3809, GM7286, GM9034, GM9061
GO:0044275 cellular carbohydrate catabolic process	ALDOA, GM14148, LDHA, GM4217, GM5787, GM2308, GM15191, GPI1, GM8318, TPI1, GM7784, ENO3, GM3839, GM8174, GM9034, GPD1, GM7611, GM10313, GM6498, GM5732, GM9568, GM11953, GM3809, GM7286, GM9061
GO:0006096 glycolysis	ALDOA, GM14148, LDHA, GM7611, GM10313, GM4217, GM5787, GM2308, GM15191, GM6498, GPI1, GM8318, GM5732, GM7784, TPI1, GM11953, GM8174, GM9568, GM3839, ENO3, GM3809, GM7286, GM9034, GM9061
GO:0016052 carbohydrate catabolic process	ALDOA, GM14148, LDHA, GM4217, GM5787, GM2308, GM15191, GPI1, GM8318, TPI1, GM7784, ENO3, GM3839, GM8174, GM9034, GPD1, GM7611, GM10313, GM6498, GM5732, GM9568, GM11953, GM3809, GM7286, GM9061

Annotation Cluster	**Enrichment Score: 3.587680656719181**
Term	**Genes**
GO:0044275 cellular carbohydrate catabolic process	ALDOA, GM14148, LDHA, GM4217, GM5787, GM2308, GM15191, GPI1, GM8318, TPI1, GM7784, ENO3, GM3839, GM8174, GM9034, GPD1, GM7611, GM10313, GM6498, GM5732, GM9568, GM11953, GM3809, GM7286, GM9061
GO:0016052 carbohydrate catabolic process	ALDOA, GM14148, LDHA, GM4217, GM5787, GM2308, GM15191, GPI1, GM8318, TPI1, GM7784, ENO3, GM3839, GM8174, GM9034, GPD1, GM7611, GM10313, GM6498, GM5732, GM9568, GM11953, GM3809, GM7286, GM9061
GO:0034637 cellular carbohydrate biosynthetic process	GPD1, TPI1, GPI1, AGL
GO:0016051 carbohydrate biosynthetic process	GPD1, TPI1, GPI1, AGL

Annotation Cluster	**Enrichment Score: 2.0873684198886866**
Term	**Genes**
muscle protein	ACTC1, MYBPC2, MYH1, PVALB, PHKA1, MYH4, CASQ1, TPM1, MYH8, TNNI2
GO:0006936 muscle contraction	MYBPC2, MYOM2, MYH1, MYH4, TPM1, MYH8
thick filament	MYBPC2, MYH1, MYH4, MYH8
GO:0003012 muscle system process	MYBPC2, MYOM2, MYH1, MYH4, TPM1, MYH8
GO:0032982 myosin filament	MYBPC2, MYH1, MYH4, MYH8
GO:0006941 striated muscle contraction	MYH1, MYH4, TPM1, MYH8
GO:0005200 structural constituent of cytoskeleton	MYBPC2, MYOM2, GM5620, TPM1
IPR015650:Heavy chain of Myosin	MYH1, MYH4, MYH8
GO:0005863 striated muscle thick filament	MYH1, MYH4, MYH8
GO:0030017 sarcomere	ACTC1, MYH1, MYH4, TPM1, MYH8
GO:0044449 contractile fiber part	ACTC1, MYH1, MYH4, TPM1, MYH8
IPR004009:Myosin, N-terminal, SH3-like	MYH1, MYH4, MYH8
GO:0005859 muscle myosin complex	MYH1, MYH4, MYH8
GO:0030016 myofibril	ACTC1, MYH1, MYH4, TPM1, MYH8
IPR002928:Myosin tail	MYH1, MYH4, MYH8
GO:0043292 contractile fiber	ACTC1, MYH1, MYH4, TPM1, MYH8
GO:0016460 myosin II complex	MYH1, MYH4, MYH8
actin-binding	MYBPC2, MYH1, MYH4, TPM1, MYH8, TNNI2
GO:0016459 myosin complex	MYBPC2, MYH1, MYH4, MYH8
GO:0015629 actin cytoskeleton	MYBPC2, MYH1, MYH4, MYOZ1, TPM1, MYH8
region of interest:Actin-binding	MYH1, MYH4, MYH8
domain:Myosin head-like	MYH1, MYH4, MYH8
IPR001609:Myosin head, motor region	MYH1, MYH4, MYH8

Term	Genes
GO:0003779 actin binding	MYBPC2, MYH1, MYH4, TPM1, MYH8, TNNI2
myosin	MYH1, MYH4, MYH8
GO:0051015 actin filament binding	MYH1, MYH4, MYH8
GO:0008092 cytoskeletal protein binding	MYBPC2, MYH1, MYH4, TPM1, MYH8, TNNI2
motor protein	MYH1, MYH4, MYH8
mmu04530:Tight junction	GNAI2, MYH1, MYH4, MYH8
GO:0003774 motor activity	MYH1, MYH4, MYH8
GO:0005856 cytoskeleton	ACTC1, MYBPC2, MYOM2, MYH1, GM5620, MYH4, MYOZ1, TPM1, MYH8
GO:0044430 cytoskeletal part	MYBPC2, MYH1, GM5620, MYH4, TPM1, MYH8

Supplementary Figure 6. **Genes that correlate with 1110006G14Rik**

ID	Gene Name	Species
Actc1	actin, alpha, cardiac muscle 1; similar to alpha-actin (AA 27-375)	Mus musculus
Casq1	calsequestrin 1	Mus musculus
Mybpc2	myosin binding protein C, fast-type	Mus musculus
Myh1	myosin, heavy polypeptide 2, skeletal muscle, adult; myosin, heavy polypeptide 1, skeletal muscle, adult	Mus musculus
Myh4	myosin, heavy polypeptide 4, skeletal muscle	Mus musculus
Myh8	myosin, heavy polypeptide 8, skeletal muscle, perinatal	Mus musculus
Pvalb	parvalbumin	Mus musculus
Phka1	phosphorylase kinase alpha 1	Mus musculus
Tpm1	tropomyosin 1, alpha	Mus musculus
Tnni2	troponin I, skeletal, fast 2	Mus musculus

Aknowledgments

I would like to thank Prof. Gerolamo Lanfranchi and Dr. Stefano Cagnin for their continuous support of my Ph.D. study and their help in all the time of research and writing of this thesis.

I would also like to thank Benny, Caterina, Francesco, Lisa, Lucia, Matteo, Paolo, and all the other people that worked with me during these three years, without their support it would have been a much different experience.