SAPIENZA
Università di Roma
Facoltà di Scienze Matematiche Fisiche e Naturali

DOTTORATO DI RICERCA
IN GENETICA E BIOLOGIA MOLECOLARE

XXVI Ciclo
(A.A. 2012/2013)

Role of noncoding RNAs in muscle differentiation.

Dottorando
Chiara Pinnarò

Docente guida
Prof.ssa Irene Bozzoni

Tutor
Prof. Antonio Musarò

Coordinatore
Prof. Irene Bozzoni

Alla mia famiglia e a Pino, che presto ne farà ufficialmente parte.

INDEX

ABSTRACT

The process of generating muscle fibers, also called myogenesis, is a perfectly orchestrated and continuous mechanism that leads to the formation of multinucleated myofibres capable of contraction. Myogenesis is a highly regulated process involving the expression of muscle specific transcription factors inducing the production of appropriate developmental stage-specific transcripts.
In the last years our group identified several non-coding RNAs (ncRNAs) with a specific role in both muscle differentiation and muscle degenerative pathologies. Among them we studied a microRNA, miR-31, expressed at high levels in muscle affected by Duchenne muscular dystrophy (DMD). We were able to demonstrate the importance of this miRNA in the differentiation process and its ability to repress, by targeting the mRNA 3' UTR, the expression of Dystrophin, a protein that has a key role in muscle fibres contraction and integrity.
Recently, our group also discovered a long non-coding RNA (lncRNA), named linc-MD1, cross-regulating specific mRNAs by competing for miRNA binding *via* their miRNA recognition motifs. These studies opened the intriguing possibility that lncRNA-based mechanisms might influence different sets of transcripts during the execution of crucial metabolic pathways, as cellular differentiation programs.
Based on these assumptions my research project has its main focus on the contribution of non-coding RNAs (ncRNAs) in muscle differentiation. By combining advanced high-throughput RNA sequencing with cellular and molecular biology techniques I have worked to get a more comprehensive catalogue of muscle specific lncRNAs and to decipher how these molecules regulate gene expression and chromatin dynamics. Once identified these new player in muscle physiology we will move towards understanding their functional implication in muscle differentiation and pathologies.

Studying non-coding RNAs in murine muscles, I also identified two lncRNAs deriving from the miR-31 locus, located on the fourth chromosome (linc-31D and linc-31P).
These two lncRNAs have a different structure and a different localization in the intracellular compartments. Linc-31P is localized in the nucleus, associated with the chromatin while linc-31D is found in the cytoplasm. RNAi and over-expression experiments allowed me to demonstrate that linc-31D has a role in promoting the switch between proliferation and differentiation.
In the human genome the miR-31 locus is located on chromosome 9 and I found, once again, two lncRNAs transcribed from two different TSS. I measured their levels in WT and DMD primary myoblasts reaching the conclusion that they are expressed, as in mouse, at high levels in proliferating conditions. Similarly to mouse, the induction of the differentiation parallels a decrease in the two isoforms' levels in WT cells.
Viceversa, the lncRNA levels remain quite high in DMD conditions where the myogenetic process is delayed, corroborating, once again, our hypothesis of a role for these lncRNAs in promoting the switch between proliferation and differentiation.

.

INTRODUCTION

Skeletal muscle development

The process of generating muscle fibers, also called myogenesis, is a perfectly orchestrated and continuous mechanism that leads to the formation of a multinucleated myofibres capable of contraction (**Figure 1**).
Myogenesis can be divided into several phases, highly regulated by the coordinated expression of regulatory RNAs, chromatin-remodeling elements and a family of transcription factors known as myogenic regulatory factors (MRFs) that are able to control the production of appropriate developmental stage-specific transcripts (Tajbakhsh 2009) (figure 1).
Proteins in the MRF family share a homologous basic Helix-Loop-Helix (bHLH) domain that is able to bind DNA to the consensus E-box sequence CANNTG, that is found in the promoter sequence of many muscle-specific genes and to form heterodimers with the E-protein family of transcription factors (Tapscott 2005)
During embryogenesis, skeletal myogenic progenitors are derived from the somites originating from the paraxial mesoderm (Aulehla and Pourquié 2006) and their specification to a muscular lineage is guided by the combinatorial action of extrinsic factors, either positive and negative signals, emanated from adjacent cells, and MRFs inside the nucleus (Parker et al. 2003); in particular, MyoD (Ishibashi et al. 2005, Tapscott 2005, Cao et al. 2006), Myf5 (Ishibashi et al. 2005), Myogenin (Cao et al. 2006) and Mrf4 (Kassar-Duchossoy et al. 2004) that have a well-characterized role in directing the process.
MyoD, Myf5 and Mrf4 are responsible of determining skeletal muscle cell identity and, in particular, it has been demonstrated that either MyoD or myf5 is required for the formation of skeletal muscle, and, even if mice lacking both of them were born alive, they were immobile and died soon after birth (Rudnicki et al. 1993).

In double-mutant mice, progenitor cells remain multi-potent and can change their fate (Tajbakhsh et al. 1996, Kablar et al. 1999) but in their absence MRF4 is able to determine skeletal muscle identity, that, on the contrary, is completely lost when, in addition to MyoD and myf5, MRF4 is absent too (Kassar-Duchossoy et al. 2004).
MRFs control over the specification process is so strong that their forced expression into a variety of non-muscle cells is able to activate the myogenesis (Braun et al. 1989, Edmondson and Olson 1989, Rhodes and Konieczny 1989), and, in some cases, to lead to the formation of terminally differentiate muscle cells (Davis et al. 1987).
Proliferating cells expressing MyoD and Myf5 are called myoblasts and are set to exit the cell cycle and to start the differentiation process undergoing morphological changes, at this point cells begin to express MRFs and myogenin that are more directly involved in this phase of the process and trigger the expression of late differentiation myogenic marker such as myosin heavy chain (MHC) and muscle creatine kinase (MCK).
Myogenin, as the other MRFs has a crucial role in the muscle development and mice that are lacking this factor, are able to survive the fetal development but die shortly after birth showing a severe reduction of skeletal muscles (Hasty et al. 1993).
In the final stages of differentiation, cells expressing structural proteins start to fuse forming a multinucleated syncytium and finally generating a mature fiber capable of contraction. This results, at the end of the embryonic development, in the formation of innervated, vascularized and contractile muscle tissues.

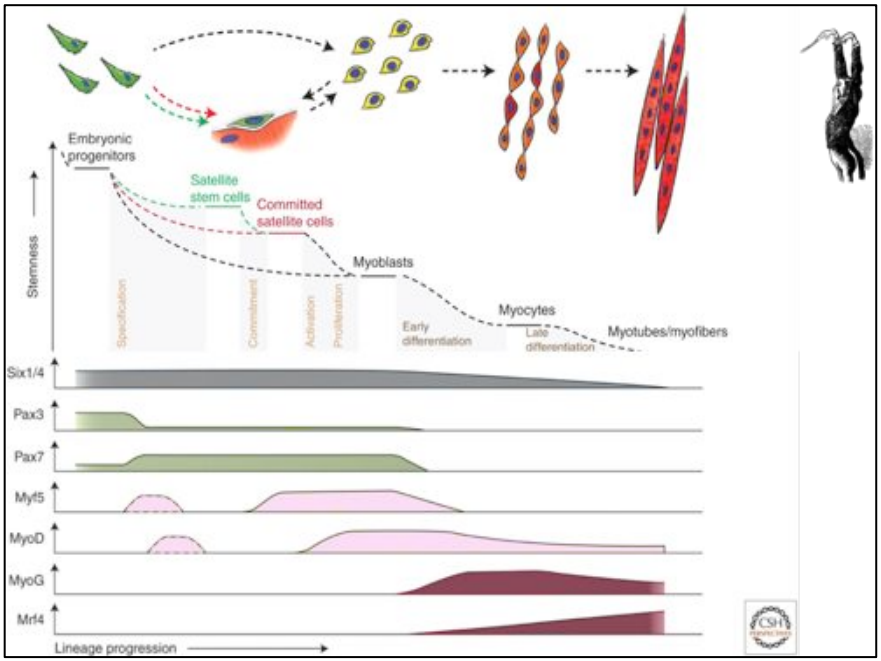


Figure 1 – Hierarchy of transcription factors orchestrating the progression through the myogenic lineage. *Muscle progenitors that are involved in embryonic muscle differentiation skip the quiescent satellite cell stage and directly become myoblasts. Some progenitors remain as satellite cells in postnatal muscle and form a heterogeneous population of stem and committed cells. Activated committed satellite cells (Myoblasts) can eventually return to the quiescent state. Six1/4 and Pax3/7 are master regulators of early lineage specification, whereas Myf5 and MyoD commit cells to the myogenic program. Adapted from (Bentzinger et al. 2012).*

Satellite cells and adult myogenesis.

During the muscular tissue development, a pool of cells, named satellite cells for their sub-laminar location, fails to enter the differentiation process, remaining associated to the periphery of skeletal muscle myofiber, in a quiescent state (Mauro 1961).
Satellite cells are considered to be the main players in the regenerative processes of the muscular tissue. Skeletal muscle has, in fact, the incredible ability to respond to damages and to repair injuries, reforming healty tissue: this regeneration is largely due to interplay between the satellite cells and their niche in the muscle.
Satellite cells are quite recognizable, thanks to a characterized series of transcription factors that are expressed at high levels.
In adult skeletal muscle, in fact, most of satellite cells are found to express the paired domain transcription factors Pax7 (Seale et al. 2000) and Pax3 (Buckingham et al. 2003), and the already described MRF, myf5 (Cornelison and Wold 1997).
Once activated, in response to the deterioration of the basal lamina, a pool of cells exit from their quiescent state and enter the cell cycle. Proliferating satellite cells follow two different destinies, a group return to quiescence to maintain the satellite cell pool reservoir, which is critical for the long-term health of the tissue, and another group is formed by cells, originating from symmetric division, that are set to become myogenic precursor cells (Collins et al. 2005) (figure 2). This newly originated myoblasts start to express MyoD and to down-regulate Pax7 entering in a path that closely resemble the one followed by myoblast originated during the embryonic development (Yin et al. 2013).

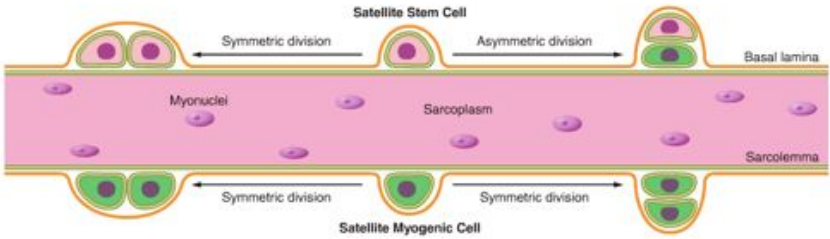


Figure 2 – Proliferation and commitment of the satellite cell population. (Yin et al. 2013)

Duchenne Muscular Distrophy

Genetic diseases affect the population worldwide in all stages of life with different consequences and clinical outcomes; among them Muscular Dystrophies form the most common class of single-gene disorders affecting millions of people and causing progressive wasting and weakness of skeletal muscle.
Duchenne Muscular Distrophy (DMD) is the most severe form of muscular dystrophy and the most common, affecting one in 3500 male births. This fatal disease, is an X-linked recessive disorder caused by mutations in the dystrophin gene (*dmd*) that encodes for a 427 kDa protein that has the very delicate and important role of connecting the actin in the cytoskeleton to the extracellular matrix in muscle fibers by forming interactions with a protein complex named the dystrophin-associated protein complex (DAPC) (Matsumura and Campbell 1994)(figure 3). The DAPC has a major role in protecting the muscle fiber and maintaining the integrity of the sarcolemma (Ervasti and Sonnemann 2008) and moreover its correct localization influences the intracellular nitric oxide (NO) pathway (Brenman et al. 1995).
When the dystrophin level in the cell is reduced or the protein is completely absent, the entire complex is delocalized causing severe sarcolemma fragility and leading to breakages of the muscle membrane in response to contractions.
In addition to that, the disruption of the association between the sarcolemmal neuronal Nitric Oxide Synthase (nNOS) and DAPC leads to impaired NO production in dystrophic muscles; this cause the alteration of the HDAC2 S-nitrosylation and its chromatin association, deregulating the expression of a specific subset of microRNA genes crucial in DMD physiopathology (Cacchiarelli et al. 2010).
In DMD patients the damages in the tissue activate the quiescent pool of satellite cells and injuries are, at first, repaired, thanks to the activity of cells that fuse to existing myotubes but the regenerative process is, unfortunately, largely inefficient compared to great extent of the damages and the satellite pool soon loose its potential and the muscle fiber is subject to severe degeneration causing

chronic inflammation, susceptibility of the tissue to oxidative stress and finally necrosis. The necrotic tissue is slowly replaced by adipose and fibrotic infiltrations making the damages permanent. DMD patients soon loose the ability to walk or simply stand and are forced to a wheel chair by the age of 12.
The most affected skeletal muscle is the diaphragm where the tissue degeneration leads to the death of the patient due to respiratory failure but the pathology also affects severely the cardiac muscle, and more than 90% of DMD patients develop cardiomyopathies (Nigro et al. 1990).

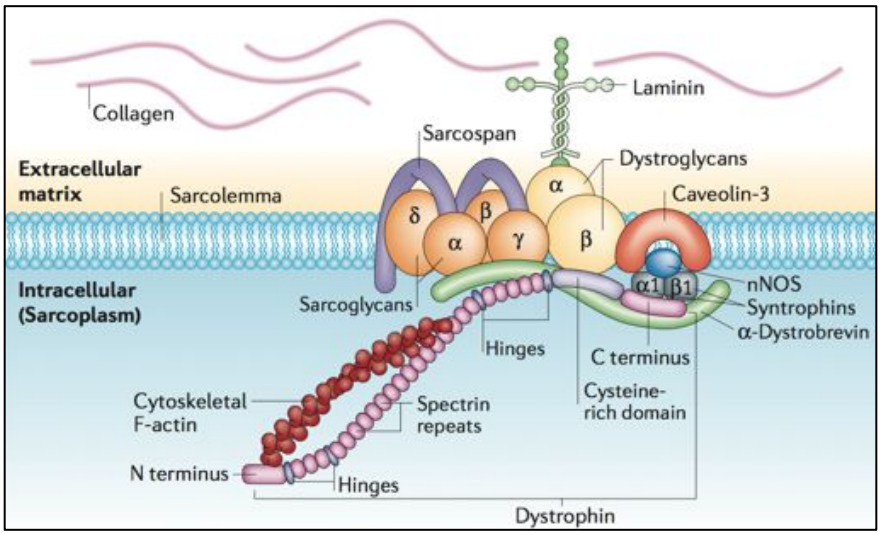


Figure 3 - Dystrophin binds to the DAPC at the sarcolemma. *Dystrophin, which is localized at the sarcolemma, has a long central rod domain made up of spectrin repeats, which are interspersed with hinge regions. The C-terminus is preceded by a cysteine-rich domain and binds to the dystrophin-associated protein complex (DAPC). The DAPC is comprised of sarcoplasmic proteins (α-dystrobrevin, syntrophins and neuronal nitric oxide synthase (nNOS)), transmembrane proteins (β-dystroglycan, the sarcoglycans, caveolin-3 and sarcospan) and extracellular proteins (α-dystroglycan and laminin).* (Davies and Nowak 2006)

Exon skipping approaches

DMD, is an ideal candidate for gene therapy, as it derives from single-gene mutations affecting the *dmd* gene; the majority of patients, carry mutations that disrupt the reading frame of the dystrophin gene, generating a truncated unstable dystrophins lacking the C-terminal bridging domain responsible for the connection to the DAPC (Aartsma-Rus et al. 2006).
Interestingly, there is another form of muscular dystrophy caused by mutations in the *dmd* gene: the Becker muscular dystrophy (BMD) (Koenig et al. 1989), a milder pathology that is frequently due to mutations that interest the gene maintaining its correct reading frame. In BMD there is the production of an internally truncated dystrophin protein that only lacks part of the central rod domain and maintain a partial functionality (Hoffman et al. 1988).
These observations have led to the formulation of the reading-frame rule to explain the phenotypic differences between BMD and DMD patients (Monaco et al. 1988), and to the idea that it is possible to realize, using an exon-skipping strategy, a dystrophin protein with a partially deleted flexible central rod domain. This therapeutic approach is based on the use of antisense oligonucleotides (AONs) capable to interfere with the splicing mechanism targeting specific splice sites or splicing regulatory regions (enhancers) on the pre-mRNA, hiding them from the splicing machinery and causing their exclusion from the mature mRNA (figure 4A). Many studies showed a successful dystrophin reading-frame restoration in the *mdx* model mouse, a naturally occurring dystrophin-deficient mutant that was first described in 1984 in a colony of C57BL/10 mice (C57BL/10ScSnJ) (Willmann et al. 2009) and that carrying a premature stop codon in its exon 23, has been widely used to study the DMD pathology (Dunckley et al. 1998, Wilton et al. 1999, Mann et al. 2001). Exon skipping approaches also proved to be effective on DMD patient derived muscle cells (van Deutekom et al. 2001). Thanks to the successes obtained both *in vitro* and *in vivo*, AONs-based therapies, aimed to the skipping of human exon 51, have been tested in clinical trials based both on intramuscular (van Deutekom et al. 2007, Kinali et al. 2009) and systemic (Cirak et al.

2011, Goemans et al. 2011) administration of molecules chemically modified to improve their stability. These studies demonstrated that exon skipping for DMD is a safe approach and that is successful in increasing dystrophin levels but also showed that there is a significant obstacle to overcome: these modified AONs are cleared rapidly from the circulation and have some difficulties in reaching all the muscular districts in the body. A strategy, aimed to the in-situ production of AONs, was developed to avoid reiterating administrations and to improve therapeutic efficiency; this approach is based on the use of chimeric small nuclear RNAs (snRNAs) that have been designed to shuttle the antisense molecule (figure 4B). Good results, in term of construct delivery and dystrophin rescue, have been obtained, using viral vector-mediated U1snRNA; the approach was, in fact, successful both in DMD cultured cells (De Angelis et al. 2002, Incitti et al. 2010, Cazzella et al. 2012) and in the *mdx* mouse (Denti et al. 2006) that has also been used to test the long term benefits of the treatment (Denti et al. 2008).

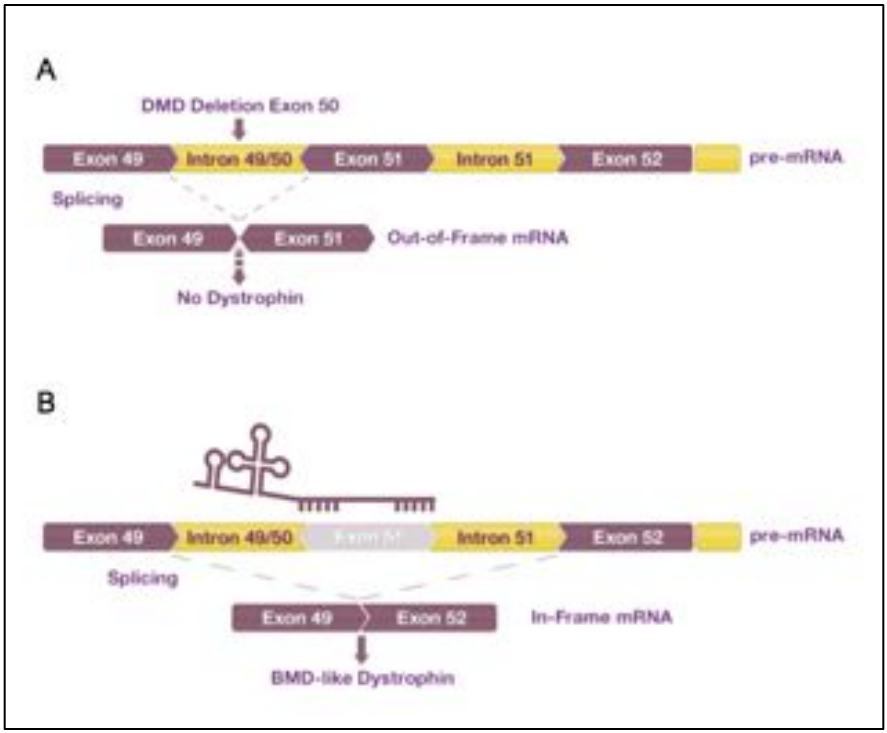


Figure 4 Example of exon-skipping strategy.

A) *Deletion of exon 50 in DMD gene results in an out-of-frame mRNA transcript and a prematurely aborted dystrophin synthesis*. **B)** *Employing of antisense molecules against exon 51 splice sites produce an in frame mRNA transcript resulting in a shortened BMD-like dystrophin protein*.

Non-coding RNAs

One of the great paradoxes in molecular biology is the relationship between the biological complexity across eukaryotes and the number of protein-coding genes and their extent. During the evolution, in fact, the number of protein coding genes has remained quite stable but, on the contrary, a larger amount of non-coding DNA is found in the genome of complex organism.
These sequences have been considered, for years, to be genetically inert fragments and have been called "junk DNA" that was considered as "tolerated" more easily by complex organisms. With the genome sequencing era and the technological advances that allowed a deep analysis of the trascriptome, came the realization that the vast majority of the complex organisms genome is transcribed to produce a large number of ncRNAs that have been found to be antisense, intergenic or overlapping with protein-coding genes and whom transcription is regulated trough evolution.
Rapid progress in characterizing these ncRNAs has lead to the identification of many subclasses that, even sharing the common function of regulating gene expression, vary widely in size, sequence and mechanism-of-action. It is now very clear that they play a major role in development of organisms and in many cellular pathways and that if mis-regulated can have a significant role in many diseases.
The first characterized ncRNAs, ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) account for more than 95% of cellular RNA content. Other ncRNAs are broadly classified on the basis of length. Molecules shorter than 200 nucleotides are classified as short ncRNAs and include microRNAs, small nucleolar and nuclear RNAs, piwi-interacting RNAs, and small interfering RNAs.
Molecules longer than 200 nucleotides have been arbitrary grouped together and called long non-coding RNAs (lncRNAs).

Micro-RNAs

Micro-RNAs (miRNAs) are small, evolutionarily conserved noncoding RNAs, of ≈22nt in length, that are transcribed, with few exceptions, by RNA polymerase II (Cai et al. 2004, Lee et al. 2004). This class of transcripts shows a peculiar genomic localization as they can be exonic, intronic or intergenic and be in found monocistronic or polycistronic units (Morlando et al. 2008, Ballarino et al. 2009).
They are initially transcribed in a hairpin-shaped molecule called pri-miRNA that undergoes a first endonucleolitic cleavage by the microprocessor complex, formed by Drosha and DGCR8, which produces a 70 nucleotides hairpin RNA called pre-miRNA, which can, thanks to the exportin 5, be exported to the cytoplasm and processed by Dicer to yield a duplex RNA of 22-23 nucleotides. Only one filament of the duplex, will then be chosen to be loaded in the RNA-induced silencing complex (RISC) as mature miRNA, while the other strand, the miRNA* will be degradated (Bartel and Chen 2004) (figure 5).
miRNAs have the ability of regulating gene expression through a sequence-specific interaction with a target mRNA, usually at its 3' untranslated region (UTR), resulting in translational repression or mRNA destabilization (He and Hannon 2004).
Some miRNAs exert their effects through a strong repression of a relatively small number of targets but it is more frequent to find miRNAs targeting hundreds of mRNA realizing a complex network of interactions that achieve a fine-tuning in gene expression.
Recent studies have described new types of regulations mediated by this class of ncRNAs such as translation up-regulation (Vasudevan et al. 2007) and heterochromatin formation (Kim et al. 2008); in addition to that, small RNAs have been described to be directly involved in transcription processes through sequence-specific interactions with promoter elements of target genes (Schwartz et al. 2008) adding another layer of complexity to the cell regulatory mechanisms.

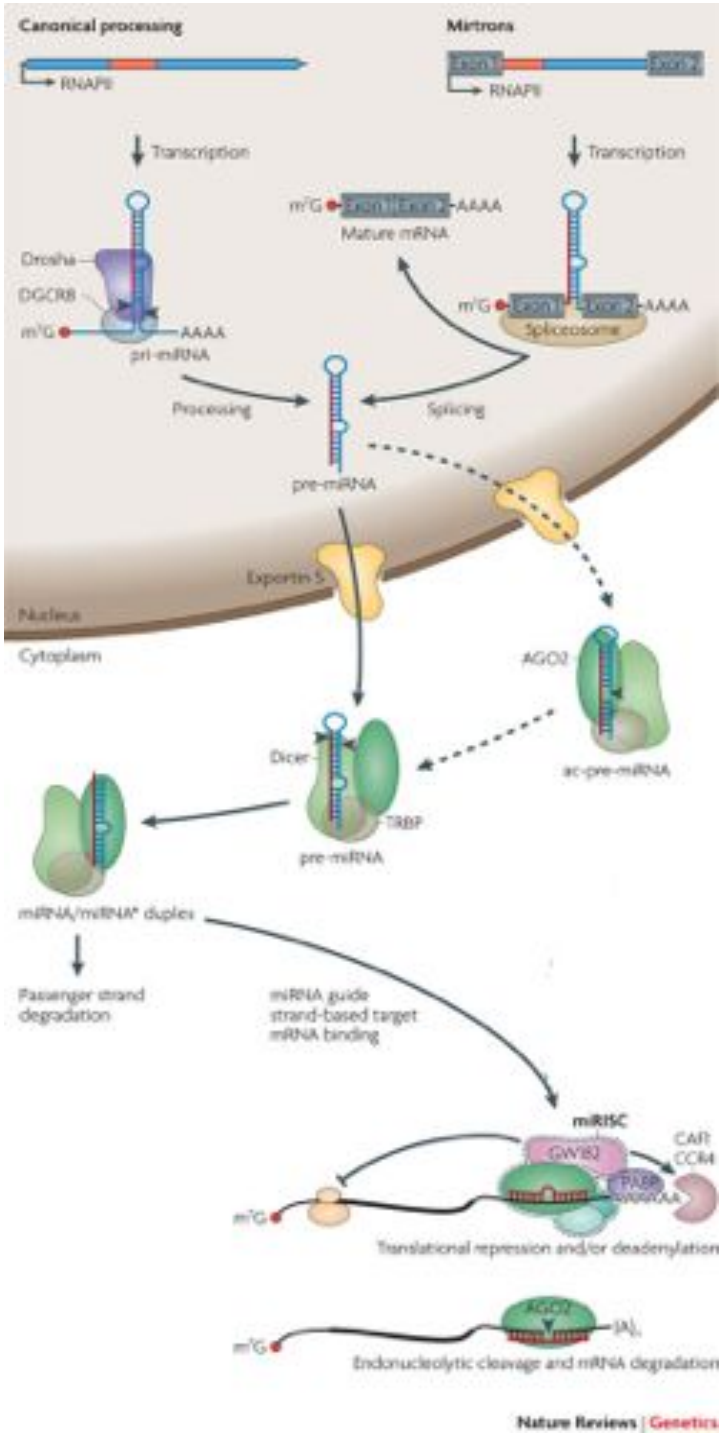

Canonical processing
Mirtrons
RNAPII
RNAPII
Transcription
Transcription
m7G
AAAA
Mature mRNA
Drosha
DGCR8
m7G
AAAA
pri-miRNA
m7G
AAAA
Spliceosome
Processing
Splicing
pre-miRNA
Exportin 5
Nucleus
Cytoplasm
AGO2
Dicer
TRBP
pre-miRNA
ac-pre-miRNA
miRNA/miRNA* duplex
Passenger strand degradation
miRNA guide strand-based target mRNA binding
miRISC
GW182
CAF1
CCR4
PABP
AAAAA
m7G
Translational repression and/or deadenylation
AGO2
m7G
An
Endonucleolytic cleavage and mRNA degradation
Nature Reviews | Genetics

Figure 5- miRNA biogenesis

MicroRNAs are processed from RNA polymerase II (RNAPII)-specific transcripts of independent genes or from introns of protein-coding genes.
In the canonical pathway, primary precursor (pri-miRNA) processing occurs in two steps, catalyzed by two members of the RNase III family of enzymes, Drosha and Dicer, operating in complexes with dsRNA-binding proteins (dsRBPs). (Krol et al. 2010)

Long non-coding RNAs

LncRNAs are molecules with little to none protein encoding capacity that have been grouped together even if they show a great heterogeneity in size (from 200 nucleotides to beyond 10 kilobases), functions localization and role in the cell.
These transcripts, normally expressed at lower levels compared with protein-coding RNAs, are generally transcribed by conventional promoters by RNA polymerase II, often spliced and can be either polyadenylated or not, and can be localized both in the nuclear compartment or in the cytoplasm (Ponting et al. 2009) exerting different functions in the cell.
It is estimated that thousands of lncRNAs are encoded in the human genome, expressed mostly in tissue-specific patterns, they are versatile molecules that are able to interact physically and functionally with DNA, other RNAs and proteins through direct base pairing or through functional domains, created thanks to the many possibilities offered by their folding ability. These properties endow lncRNAs with an incredible range of capabilities and mechanisms of action (Paralkar and Weiss 2013) (figure 6).
One of the first characterized lncRNAs, Xist, is a perfect example of the role these molecules can play in epigenetic processes; once transcribed from one of the X chromosome in female's XX cells, Xist coats that chromosome itself and recruits repressive complexes to condense the chromatin and silence it in a process termed Lyonization (Heard and Disteche 2006). Another example of epigenetic regulation mediated by lncRNAs is represented by HOTAIR (Rinn et al. 2007) that, transcribed from the HOXC locus, recruits repressive complexes to silence genes in the HOXD locus and is the first lncRNA described acting *in trans* on another chromosome.
LncRNAs have also been shown to be involved in mRNA stability and translation (Gong and Maquat 2011), in the organization of nuclear architecture (Mao et al. 2011) and to interfere with protein–DNA binding (Kino et al. 2010) and directly alter protein function (Willingham et al. 2005). Recent studies have first suggested and then demonstrated that RNAs are able to influence each other levels

by competing for the limited pool of miRNAs (Seitz 2009, Poliseno et al. 2010) and open the door to a novel interpretation of the conventional RNA logic. According to the competing endogenous RNA (ceRNA) theory (Salmena et al. 2011), RNA molecules are able to sequester miRNAs protecting their targets from repression. LncRNAs have indeed shown, in accord with the ceRNA theory, the ability to modulate mRNA levels by competing for microRNA binding (Cesana et al. 2011).
Such array of functions enhanced regulatory networks in the cell increasing the complexity of the system and providing an important evolutionary advantage to higher eukaryotes(Guttman and Rinn 2012, Rinn and Chang 2012)

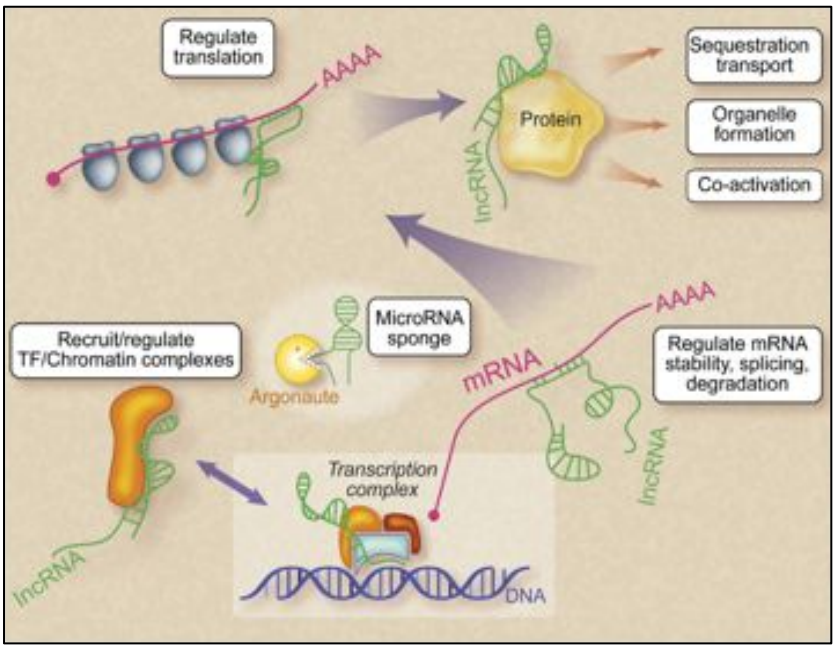


Figure 6: Mechanisms of LncRNA action.

lncRNAs (indicated in green) have been shown to regulate gene expression at multiple levels: chromatin,transcription, mRNA, translation, and protein. Hematopoietic lncRNAs may act at any of these levels. "MicroRNA sponge" refers to the ability of lncRNAs to sequester cellular microRNAs and prevent them from binding mRNA target (Paralkar and Weiss 2013).

Non-coding RNAs in muscle biology

Muscle tissue development is, as already described, a highly regulated process orchestrated by an evolutionary conserved network of transcription factors responsible of controlling the delicate switch between proliferation and differentiation processes of muscle cells. In addition to the coordinated regulation of muscle specific genes, the myogenic development involves the expression of a collection of ncRNAs, which modulate, at many levels, muscle development and homeostasis.
The essential role of miRNA in muscle development has been, initially, demonstrated generating Dicer knock-out mutants in mice; this approach resulted in severe muscle defects due to the loss of miRNAs expressed in cardiac and skeletal muscles (O'Rourke et al. 2007).
Several miRNAs are, in fact, found to be specifically expressed or highly enriched in skeletal and/or cardiac muscles (McCarthy 2008). Among them the most widely characterized are the members of the miR1/206 and miR133a/133b families, that have different expression profiles (figure 7).

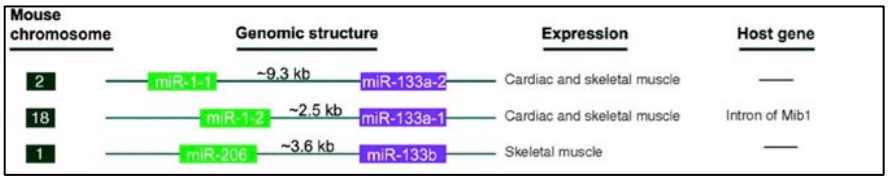


Figure 7: Genomic structures of muscle-specific miRNAs and their sequence homologies. *The genomic locations of muscle-specific miRNA genes, miR-1-1/miR-133a-2, miR-1-2/miR-133a-1, miR-206/miR-133b on mouse chromosomes. The expression of these miRNAs and the host genes in which they reside are also indicated . adapted from (Chen et al. 2009)*

- ***miR-1-1/133a-2*** and ***miR-1-2/133a-1*** clusters, show a cardiac and skeletal muscle specific transcription, that is controlled by two separate enhancers, one upstream and the other intergenic, where the coordinated action of myogenic transcription factors SRF, MEF2, and MyoD has been reported (Zhao et al. 2005, Rao et al. 2006, Liu et al. 2007).
- The cluster encoding ***miR-206*** and ***miR-133b*** is transcribed exclusively in skeletal muscle (Chen et al. 2006) and is characterized by a complex architecture in terms of transcriptional control: *miR-206* is expressed autonomously from its own proximal promoter and *miR-133b* instead, is transcribed from a 13Kb distal promoter in conjunction with *Linc-MD1 (*long non-coding RNA, muscle differentiation-1). This two miRNAs also show a different timing of expression during the differentiation process; miR-206 is, in fact, already expressed in proliferating myoblasts, whereas miR-133b transcription is activated only upon differentiation (Cesana et al. 2011).

All these miRNAs cooperate with transcription factors to orchestrate the precise temporal expression pattern of muscle genes. An example of the complex circuitry formed by them, their targets, MRFs and other factors involved in skeletal muscle and cardiac development is shown in figure 8.

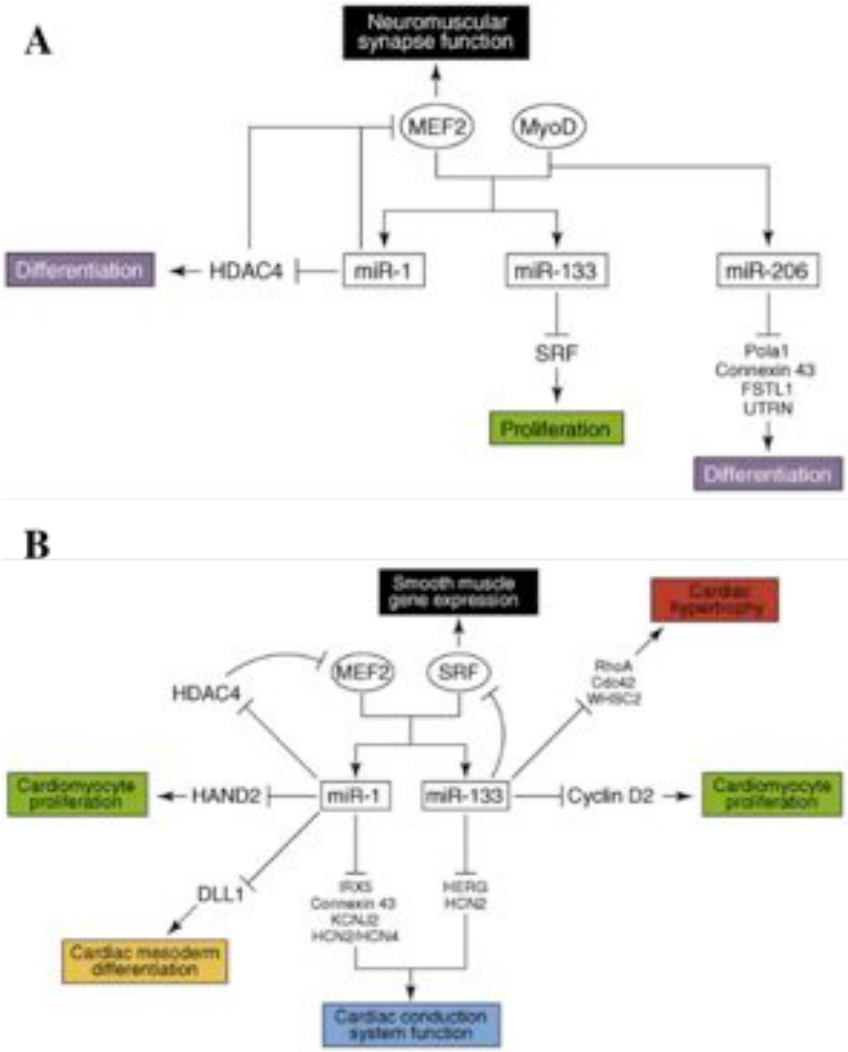


Figure 8 – miRNAs circuitries during skeletal muscle (A) and cardiac (B) development.

In addition to these muscle-specific miRNAs (also called myoMirs) a key role in the myogenesis is played by ***miR-31***, a microRNA well known for his involvement in tumor methastatic progression (Valastyan and Weinberg 2010).
miR-31 has another key role in muscle development; it expressed in muscle satellite cells and is involved in the process of regeneration, targeting Myf5 mRNA and preventing its accumulation in the quiescent pool (Crist et al. 2012).
Our group also demonstrated that miR-31 has another important target, it is in fact able to repress dystrophin synthesis (Cacchiarelli et al. 2011).
This discovery was possible thanks to the observation that miR-31 is highly enriched in dystrophic condition, both in *mdx* mice and human DMD myoblasts; in particular its expression has been found localized in regenerating myoblasts of dystrophic muscles indicating that its high levels of are due to the intensive regeneration program which is mediated by the activation of satellite cells (Cacchiarelli et al. 2011). Interestingly, in dystrophic myoblasts and satellite cells the lack of dystrophin correlated with a delay of the maturation process of the cells. In the same work was also shown that in dystrophic conditions, when dystrophin synthesis is rescued through the exon skipping strategy, the inhibition of miR-31 activity increased dystrophin production.
Since in a compromised muscle the contribution to dystrophin production by regenerating fibers is quite relevant, miR-31 repression in this compartment can represent an improvement to current therapeutic treatments aimed to increase the levels of dystrophin synthesis.

-Linc-MD1 is one of the first muscular lncRNAs identified, conserved in human and mouse, with a crucial role in myogenesis (Cesana et al. 2011).
Expressed with a specific timing during *in-vitro* differentiation of mouse myoblasts, *Linc-MD1* controls the progression from early to late phases of muscle differentiation by functioning as a ceRNA; competing for the binding of miR-133 and miR-135, it regulates the expression of mastermind-like protein 1 (MAML1) and myocyte-

specific enhancer factor 2C (MEF2C), transcription factors with a key role in activating muscle genes involved in the late stages of muscle differentiation (figure 8c).
This LncRNA has another important characteristic; its expression is found strongly reduced in DMD patients' myoblasts. In these cells, the rescue of linc-MD1, produced the recovery of both MAML1 and MEF2C synthesis and partial rescue of the correct timing of the differentiation program, suggesting a relevant conserved role in the control of muscle differentiation.

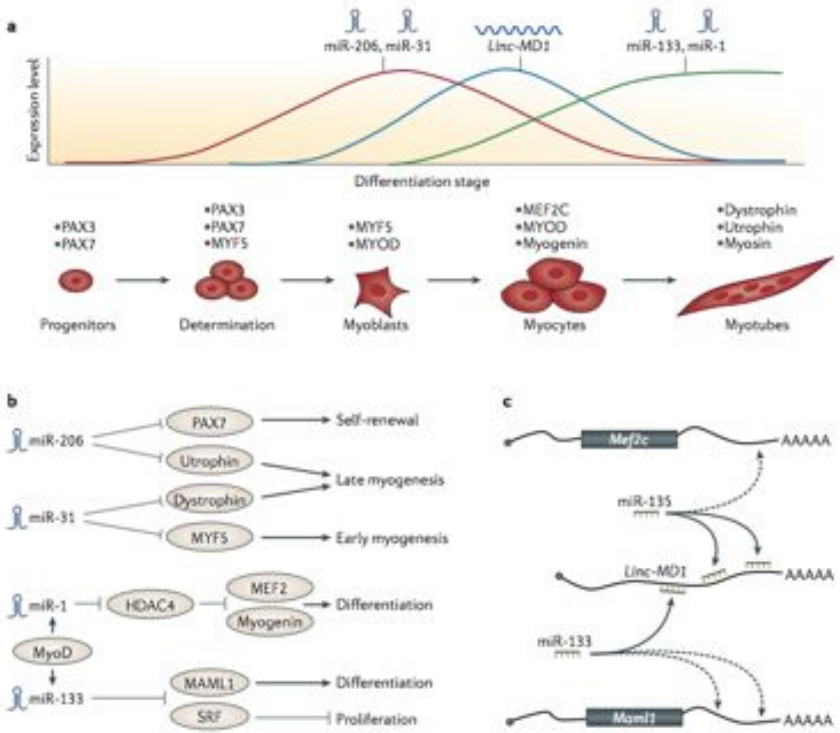


Figure 9 - ncRNAs in muscle differentiation.
A) *Schematic representation of the differentiation stages of muscle cells and of the protein factors involved in each phase. The graph shows the corresponding temporal expression patterns of selected non-coding RNAs (ncRNAs).*
B) *MicroRNAs (miRNAs) cooperating with transcription factors to sharpen their temporal expression pattern. miRNA that prevent the early activation of late myogenic proteins, such as utrophin and dystrophin and late myogenic miRNAs that reinforce late differentiation stages are indicated.*
C) *Linc-MD1 reinforces the switch from early to late differentiation gene expression by acting as a 'sponge' to limit the repressive effect of miR-133 on Maml) and of miR-135 on Mef2c. SRF, serum response factor. (Fatica and Bozzoni 2014)*

AIMS

In the last two decades genome-sequencing projects in conjunction with transcriptomic analysis have shown that the majority of the genome in animals and plants is transcribed in a developmentally regulated manner to produce large numbers of non–protein-coding RNAs (ncRNAs). The number of these molecules increases with evolution, suggesting that RNA-based regulatory mechanisms may have a relevant role in the increase of developmental complexity in eukaryotes; It has been now widely demonstrated, that many of these transcripts are functional, and have important roles in regulating gene expression at many levels in the cell. During this intense RNA-era it has been discovered that, even in the muscle tissue, in addition to the coordinated regulation of MRFs, the differentiation process involves the expression of different kind of ncRNAs, able to modulate, at many levels, muscle development.
My PhD project was dedicated to get a more comprehensive catalogue of muscle specific ncRNAs and to decipher how these molecules regulate gene expression and chromatin dynamics in this tissue development and homeostasis. Thanks to advanced high-throughput RNA sequencing and cellular and molecular biology techniques I have found a set of polyadenilated lncRNAs expressed in muscle and I defined their tissue specificity, subcellular localization and timing of expression.
Among them I identified two lncRNAs deriving from the locus containing miR-31, a well characterized, microRNA that has an important role in the muscle differentiation process and is able to repress the expression of dystrophin, a protein with a key role in muscle fibres integrity.
I demonstrated that these two lncRNAs, expressed at high levels in proliferating cells, have a different structure and a different localization in the intracellular compartments. Modulation of their expression, trough different experimental approaches, suggests that they could have a role in promoting the switch between proliferation and differentiation.
I also examined miR-31 locus in the human genome and I found

expessed, as in mouse, two different lncRNAs. I finally confronted their levels in WT and DMD conditions, both in human cells and in the *mdx* mouse, the murine model for the pathology.

RESULTS

Having a comprehensive view of the transcriptome, is essential to reveal the RNA components of cells and tissues, and also for understanding the molecular mechanisms involved in development and disease. In particular an accurate transcriptomic analysis is necessary to catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs expressed in a model system at different timing during the differentiation process.
A powerfull tool to obtain that information is represented by RNA-Seq, an approach that, taking advantages of the recently developed deep-sequencing technologies, allows to sequence the entire population of RNA (total or fractionated, such as poly(A)+) extracted from cultured cells or tissues. Following sequencing, the resulting reads are either aligned to a reference genome or reference transcripts, or assembled de novo without the genomic sequence to produce a genome-scale transcription map that consists of both the transcriptional structure and/or level of expression for each gene (Wang et al. 2009).
In collaboration with components of the research group where I worked for this thesis, I realized a RNA-Seq experiment, using Illumina TruSeq™ Technology, on the poli(A)$^+$ RNA fraction extracted form C2C12, an immortalized cell line of murine myoblasts, derived from satellite cells (Yaffe and Saxel 1977, Blau et al. 1985), that are commonly used as an *in-vitro* model of skeletal muscle development.
These cells are, in fact, able to differentiate into myocytes and to form multinuclated myotubes under appropriate culture conditions (see matherial and methods).
The experiment was realized on samples obtained from cells collected during proliferation and after 1, 3, and 5 days from the induction of to the differentiation process, with the aim of underlining changes and modulation occurring in the trascriptome during the myogenesis with particular attention to the non-coding portion of the genome.
Datas obtained from the RNAseq were analysed by collaborators of prof. Anna Tramontano and, thanks to their bioinformatics analysis

I obtained a list of putative, not previously annotated long non-coding RNAs that showed a sufficient level of expression and were modulated during differentiation.
We selected from the list two groups of lncRNAs, one formed by putative lncRNAs that appeared to be up-regulated and the other formed by putative lncRNAs that appeared to be down-regulated after the induction of differentiation .

Characterization of putative lncRNAs expression in murine myoblasts

To validate data obtained from the sequencing and to verify the pattern of expression during the differentiation process of the selcted putative lncRNAs I collected RNA samples from cultured proliferating C2C12, in growth medium (GM) and after 5 days from the induction of the differentiation process (DM).
I used the obtained samples to perform a semi-quantitave Reverse Trascription Polimerase Chain Reaction (sqRT-PCR) experiment that allowed us to detect the RNA expression levels.
I designed and used for the experiment, different pairs of primers that were designed, accordingly to the RNA-seq data, to specifically amplify the putative ncRNAs and results of the experiment are shown in **Figure 10.**
Figure 10A shows lncRNAs whose expression, according to the sequencing experiment, is induced or strongly up-regulated after the switch to differentiation media. Linc-MD1, that is absent in growth conditions (GM) and activated upon shift to differentiation (DM) of mouse myoblasts (Cesana et al. 2011), has been used as a positive control and GAPDH as a normalization control.
It is interesting to notice that, in accordance with the RNA-seq data the expression of many of these lncRNAs is strongly up-regulated at day 5 of differentiation (DM). An exception is represented by the levels of *lnc-182*, *lnc-082* and *lnc-058,* that seem to remain quite constant in proliferating and differentiated cells.
I also found that two of the putative lncRNAs, *lnc-405* and l*nc-149*

are present in two isoforms with different expression patterns. *lnc-405* locus, in particular, seems to produce a specific isoform in proliferating condition (GM) and another different one in differentiated cells (DM). To ascertain the specificity of the amplification, PCR products have been sequenced and the different isoforms produced from the amplification of *lnc-405* and *lnc-149* resulted deriving from the correct locus.

In **Figure 10B** are grouped lncRNAs whose expression should decrease, according to sequencing data, after the switch to differentiation media.

In some cases, as for *lnc-456* and *lnc-686*, the level of the ncRNA seems to remain constant in proliferating (GM) condition and in differentiated cells (DM); in other cases, as for *lnc-254* and *lnc-793*, the level of expression is found slightly decreased in differentiated C2C12 (DM).

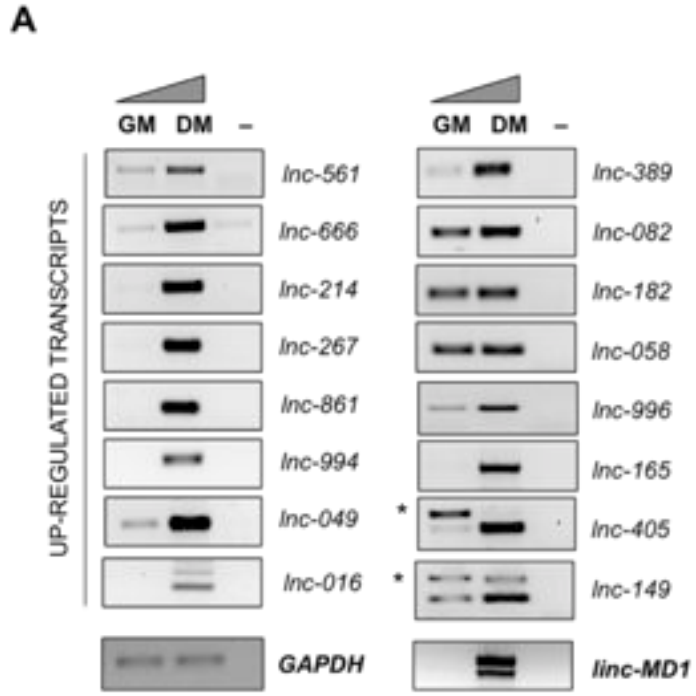


Figure 10:lncRNAs expression in C2C12 myoblasts

sqRT-PCR performed on RNA exctracted from mouse myoblast in proliferating conditions (GM) and after 5 days of differentiation (DM.) GAPDH is used as normalization control.

(A) *lncRNAs induced or up-regulated during the differentiation process. Stars are used to indicate different isoforms.*

(B) *lncRNAs whose expression decreases or doesn't vary during the differentiation process*

lnc-RNAs subcellular localization.

It is known that lncRNAs can be localized in different subcellular compartments wherethey exert different functions (Ponting et al. 2009).
To further characterize the identified lncRNAs I investigated their subcellular localization in proliferating or differentiated C2C12 myoblasts depending on their timing of expression.
Using a cell fractionation procedure I was able to obtain the cytoplasmic, nucleoplasmic and chromatin extracts from cell collected in proliferating conditions (GM) or after 5 days of differentiation (DM).
RNA was extracted and analysed by qRT-PCR with the same primers used in the characterization of their expression (Figure 10).
To verify the correct nuceus/cytoplasmic separation I used GAPDH mRNA (GAPDH) as a cytoplasmic marker (C) and its pre-mRNA (pre-GAPDH) as a nuclear control (Chr).
Figure 11A shows lncRNAs whose expression proved to remain stable while **Figure 11B** displays the sub-cellular localization of lncRNAs whose expression was induced or up-regulated upon differentiation. In the latter case cells were collected after 5 days from the induction of differentiation (DM).
lnc-405 and *lnc-149,* that have two isoforms with different expression patterns were analysed both in proliferating and differentiated cells (**Figure 11A and B)** .
Notably, the *lnc-149* isoforms are localized in different subcellular compartments (Cytoplasm and Nucleus) maintaining the same pattern in both proliferating (GM) and differentiated cells (DM).
From *the lnc-405* locus are produced two isoforms, one specifically expressed in proliferating condition (GM- Figure 10) that is found localized in the nucleus of C2C12 cells (GM-chr **figure 11A)** and another one specifically expressed in differentiated cells (DM-Figure 10) that appears to be present both in the cytoplasm (C) and associated to the chromatin (chr) (**Figure 11B**) with a stronger signal obtained in the latter.

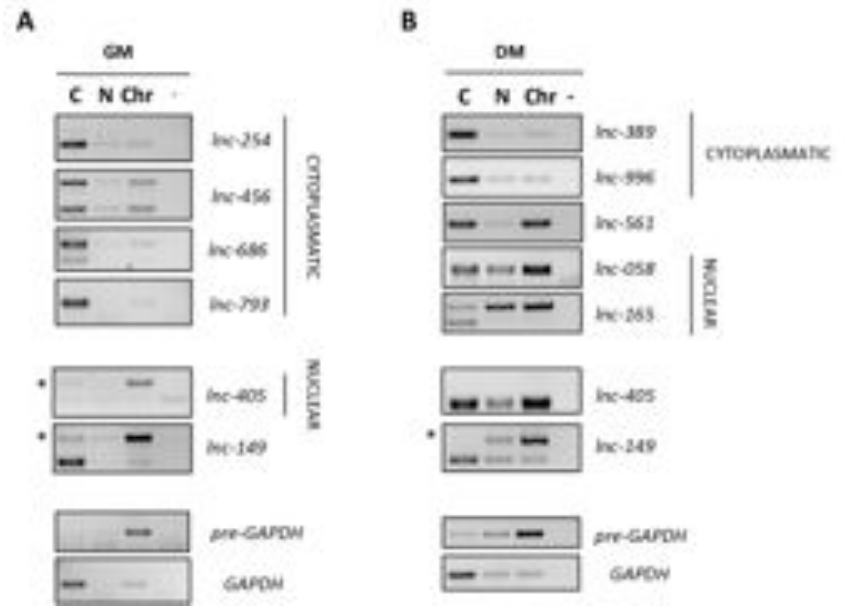


Figure 11: lncRNAs subcellular localization in C2C12 myoblasts.
*(**A**) sqRT-PCR performed on RNA exctracted from cyroplasmic (C), nucleoplasmic (N) and Chromatinic (Chr) fraction of mouse myoblast in proliferating conditions (GM). GAPDH mRNA (GAPDH) and pre mRNA (pre-GAPDH) were used as endogenous controls.*
*(**B**) sqRT-PCR performed on RNA exctracted from cyroplasmic (C), nucleoplasmic (N) and Chromatinic (Chr) fraction of differentiated mouse myoblast (DM). GAPDH mRNA (GAPDH) and pre mRNA (pre-GAPDH) were used as endogenous controls.*
Stars are used to indicate different isoforms

lnc-RNAs tissue specificity in WT and *mdx* mouse.

The *mdx* mouse is a dystrophin-deficient mutant that has been widely used to study the DMD pathology.
These mice have a shorter life span as compared to wild-type controls and in their muscle, show the typical degeneration processes associated with the pathology.
Between the second and the sixth week of life, marked waves of degeneration and regeneration interest the mice muscle tissue, which results in an increase in the number of newly differentiating myofibers characterized by centralized nuclei and an increased heterogeneity in myofiber size. Subsequently, loss of muscle tissue is slowed and general muscle weakness is not evident until later in the life of the animals (Willmann et al. 2009).
I chose to analyse the expression of our lncRNAs in different muscular tissues obtained form 6 weeks old *mdx* and wt mice; I chose that age because, in that period, in *mdx* mice the degeneration and regeneration process are already occurring and it is possible to observe a great number of newly differentiating myofibers characterized by centralized nuclei and an increased heterogeneity in myofiber size. Moreover I wanted to assess if our lncRNAs were expressed only in muscle, showing tissue specificity, or if their expression could be detected in other district of the mice body.
In order to do that, *Mdx* mice were sacrificed in parallel with wild-type (WT) isogenic/aged matched animals. Different muscular districts (gastrocnemius, heart and tibialis) were dissected alongside with non-muscular one (Brain, Cerebellum and Lung); total RNA was extracted from powered tissues (see material and methods) and used for a sqRT-PCR experiment performed with the same primer pairs used for the characterization of the lncRNAs expression in C2C12 cells (**Figure 10**)
The results are shown in **figure 12.**
Several lncRNAs showed a muscle specific expression; in particular *lnc-267*, *lnc-994* are expressed at high levels in skeletal muscle, slightly up-regulated in *mdx* condition. In contrast *lnc-996, lnc-049*, and both isoforms of *lnc-149* are only found to be expressed in *mdx*

skeletal muscle resulting almost absent in WT mice and in heart, brain, cerebellum and lung of both *mdx* and WT animals.
Once again, the expression profile of *lnc-405* appears to be very interesting; the two isoforms seem to have a very precise expression pattern. One of them has a muscle-specific expression and is found both in skeletal muscle and in the heart; the other one is found in non-muscular tissues, brain, cerebellum and lung. Neither of them seems to vary between wt and *mdx* conditions.
In addition to the already mentioned lncRNAs, I found other lncRNAs that do not have a tissue specific expression but that are, instead, deregulated in *mdx* conditions; an example of that is represented by *lnc*-666 that in WT skeletal muscle is not expressed while it is present in the same tissues derived from the *mdx* mice.
Finally there are lncRNAs that have a very low level of expression in adult mice: *lnc-023* and *lnc-793* are good examples. These data confirm the decrement in their expression revealed in the sequencing experiment and confirmed our sqRT-PCR in C2C12 *in-vitro* differentiation.

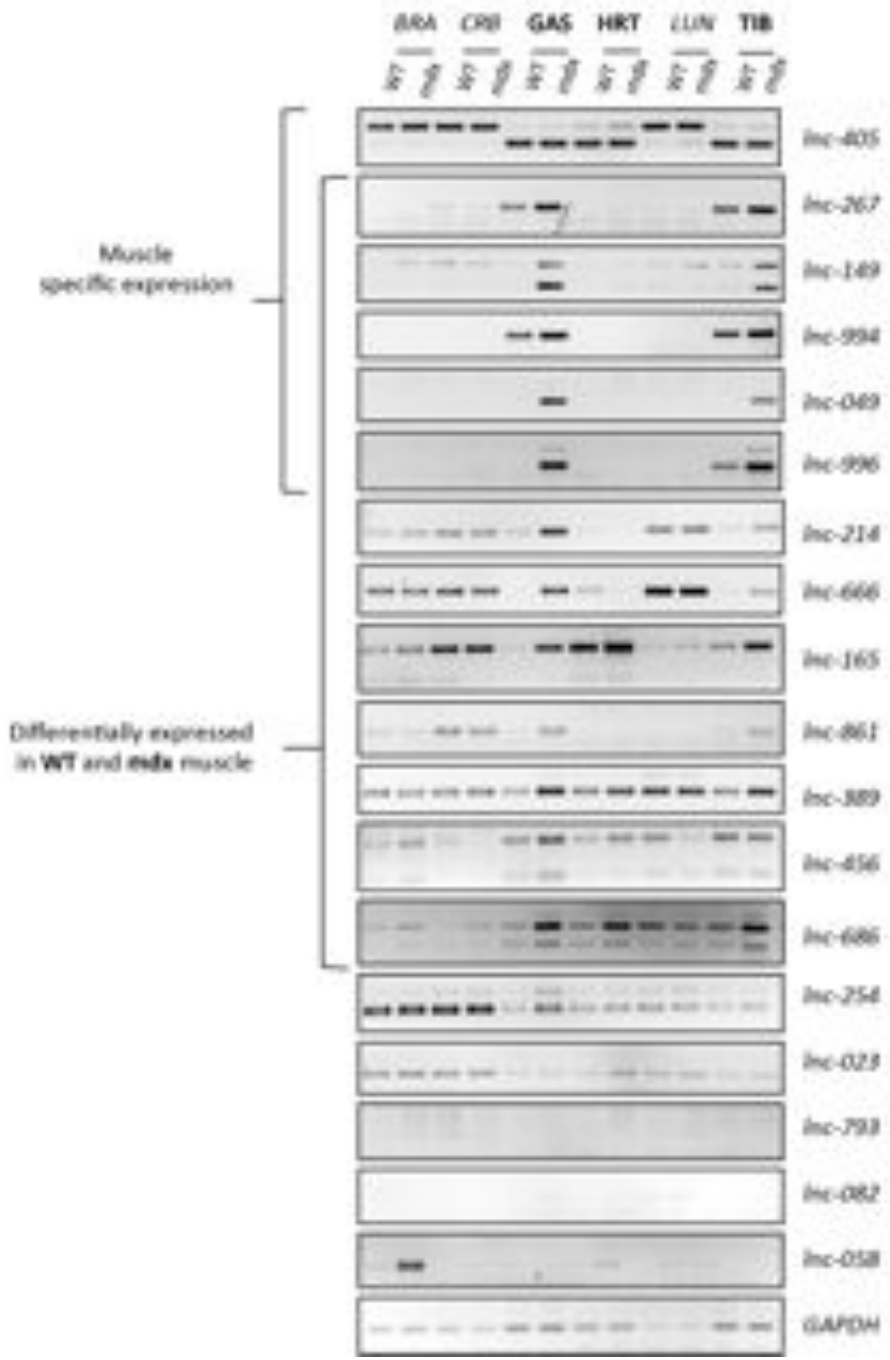

BRA
CRB
GAS
HRT
LUN
TIB
lnc-405
lnc-267
lnc-149
lnc-994
lnc-049
lnc-996
lnc-214
lnc-666
lnc-165
lnc-861
lnc-389
lnc-856
lnc-686
lnc-254
lnc-023
lnc-793
lnc-082
lnc-058
GAPDH
Muscle specific expression
Differentially expressed in WT and mdx muscle

Figure 12: lnc-RNAs tissue specificity in WT and *mdx* mouse.
sqRT-PCR performed on total RNA exctracted from 6 weeks old WT and mdx mice Brain (BRA), Cerebellum (CRB), Gastrocnemius (GAS), Heart (HRT), Lung (LUN) and tibialis (TIB). Lnc-RNAs are grouped according to their expression pattern.
GAPDH mRNA (GAPDH) was used as endogenous controls.

Characterization of a new lncRNA arising from miR-31 locus in mouse.

Data obtained from the RNAseq indicated a putative ncRNA transcript arising from miR-31 genetic locus. This transcript, even if didn't seemed to be expressed at high level, caught our attention; miR-31, in fact, expressed from the same locus, has a key role in muscle cells since it targets Myf5 and dystrophin mRNA (Cacchiarelli et al. 2011, Crist et al. 2012) and his levels have been found to be higher in dystrophic conditions, in both mouse and human biopsies.
Previous studies by Hong and colleagues (Sun et al. 2009) showed that in mouse miR-31 is transcribed from a single locus in the 4th chromosome. This locus (chr4:88,556,461-88,556,566) is formed by 3 non-coding exons and the miRNA is hosted in the third one.
To confirm our RNA-seq data and this genomic organization we performed an advanced 5' rapid amplification of cDNA end (5'RACE) assay able to select for 5'end capped RNAs using RNA extracted form proliferating C2C12 murine myoblasts.
Our analysis revealed the presence of two different transcripts.
The first one has a Trascriptional Start Site (TSS), indicated by an arrow in **figure 13**, located ≅30Kb upstream pre-miR-31 and is formed by the three predicted exons.
The second isoform previously not annotated, has a TSS, indicated by an arrow, that maps 750bp upstream pre-miR-31 **figure 13.**
When performing the RACE experiment, first strand cDNA was synthesized from RNA using oligo-dT and considering this and also according to the RNA-seq data it is possible to conclude that both transcripts are polyadenilated. The two isoforms were named linc-31D (Distal) and linc-31P (Proximal) based on TSS position from pre-miR-31. Both linc-31D and linc-31P contain the pre-miR-31 sequence.

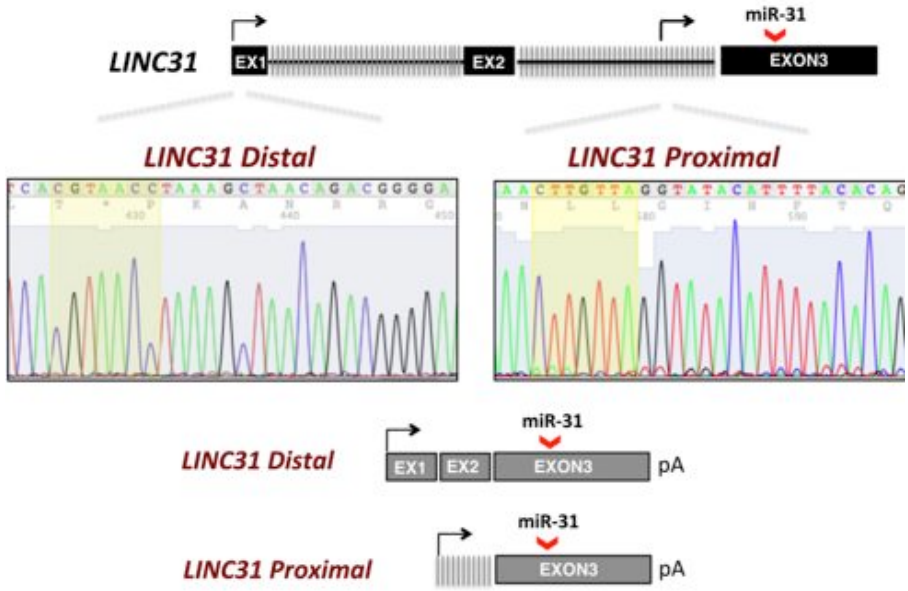


Figure 13: linc31 genomic structure.
Upper panel shows the schematic representation of the murine miR-31 genomic locus and the transcriptional start sites (TSS) found by '5 RACE analysis indicated by arrows. The electropherograms highlight the 5' sequences of the two linc31 isoforms (LINC31 Distal and LINC31 Proximal).
Lower panel shows the schematic representation of LINC31 Distal and LINC31 Proximal transcripts containing miR-31 sequence.

linc-31 expression is modulated during *in vitro* muscle differentiation.

Our group has already demonstrated that miR-31 level is high in proliferating myoblasts and then decrease upon the induction of differentiation (Cacchiarelli et al. 2011). I wanted to analyse the expression of the two transcripts and to compare their levels in C2C12 myoblasts in proliferation and during the differentiation process. RNA was obtained from cultured cells collected in growth medium (GM) and after 1, 3 and 5 days from the induction of the differentiation process. sqRT-PCR was performed using specific primers that are able to discriminate between the two Linc-31 isoforms. In particular I used forward common primer, annealing to exon-1 of both isoforms and two specific reverse primers anneling or on the Distal isoform or on the Proximal one. I designed these primers to obtain two different-sized specific bands that allow us to co-amplify and discriminate between the two isoforms.
figure 14 highlights that both linc-31D and linc-31P are expressed in proliferating cells and even if they are still expressed one day after the induction of the differentiation process, their levels decrease with the progression of the myogenesis. Their expression parallels that of miR-31's one, that was analysed by Northern blot (**figure 14-**lower panel), to confirm the expected expression pattern, in the same cells (Cacchiarelli et al. 2011). Similarly to linc-31D and linc-31P, miR-31 levels are high in proliferating conditions (GM) and gradually decrease after the induction of the *in-vitro* differentiation process.

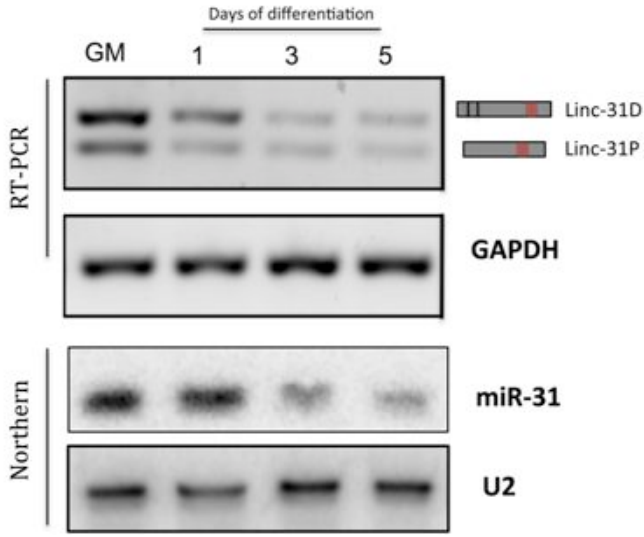


Figure 14: Linc-31 expression profile during an *in-vitro* muscle differentiation.
RNA samples collected from myoblast in proliferating conditions (GM) and after 1, 3 or 5 days from the induction of the differentiation process were used to analyse linc-31 a miR-31 levels.
The expression pattern of the two transcripts (linc-31D and linc-31P) originating from miR-31 locus has been analized by sqRT-PCR, GAPDH mRNA (GAPDH) was used as normalization control (upper panel).
miR-31 expression level in the same cells, has been analysed by Northern Blot. U2 snRNA is used as endogenous control (lower panel).

linc-31P and linc-31D subcellular localization

As for our collection of ncRNAs we wanted to investigate the subcellular localization of linc-31D and linc-31P isoforms. Using a cell fractionation procedure I was able to obtain the cytoplasmic, nucleplasmic and chromatin exctracts from C2C12 cells collected while proliferating. I proceed to RNA extraction and to sqRT-PCR with the same primers used for the characterization of the lncRNAs isoforms (**figure 10).**

The panel in **figure 15** shows that the two isoforms have a different localization: the proximal transcript (linc-31P) is not exported in the cytoplasm but is instead localized in the nucleus and, in particular, it resulted associated with the chromatin (CHR).

Linc-31D isoform instead, seems to be exported and localized in the cytoplasm (CYT).

To confirm these data, Fluorescence *In Situ* Hybridization (FISH) was then performed in proliferating C2C12 myoblasts.

Digoxigenin (DIG)-conjugated Locked Nucleic Acid (LNA) probes designed to specifically recognize either Linc-31D or Linc-31P were utilized.

Linc-31D fluorescence appeared diffusely and weakly staining the cytoplasm while the linc-31P isoform instead showed multiple and strong foci localized in the nucleus of proliferating cells while being absent after differentiation (**Figure 16)**

Therefore, the FISH experiment confirmed the data obtained by sqRT-PCR.

Background signals were obtained using a scramble sequenced probe (scramble) and another control LNA probe targeting a region upstream linc-31P TSS (INTRON).

A probe targeting U6 small nuclear RNA (snRNA) has been used to obtain a signal that represents our positive control.

Background signals were obtained using a scramble sequenced probe (scramble) and another control LNA probe targeting a region upstream linc-31P TSS (INTRON).

A probe targeting U6 small nuclear RNA (snRNA) has been used to obtain a signal that represents our positive control.

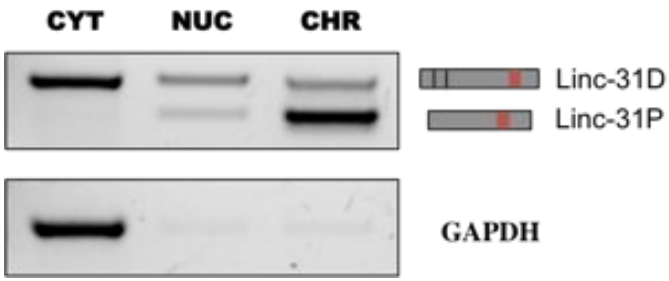


Figure 15: subcellular localization of linc-31 isoforms in C2C12 myoblasts.
sqRT-PCR performed on RNA exctracted from cyroplasmic (C), nucleoplasmic (N) and Chromatinic (Chr) fraction of mouse myoblast in proliferating conditions (GM) to analyze linc-31D and linc-31P subcellular localization. GAPDH mRNA (GAPDH) was used as endogenous controls.

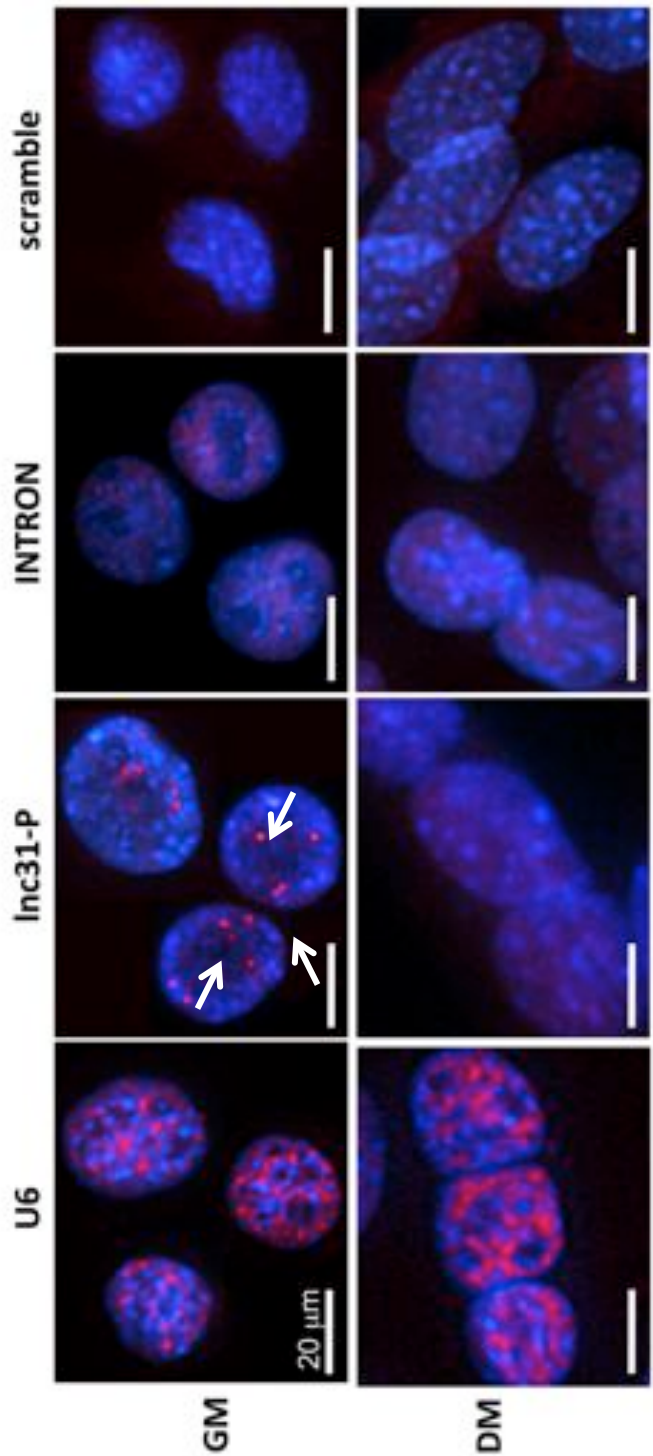

U6
lnc31-P
INTRON
scramble
GM
DM
20 µm

Figure 16: Linc-31P Subcellular localization

Fluoresce in-situ hibridization (FISH) analysis on mouse myoblast in proliferating conditions (GM) and after 5 days of differentiation (DM) performed with Digossigenin labelled LNA probes for linc-31P **(Linc31-p).** *White arrows underlines the presence of linc-31P foci in the nucleus.*
A scramble sequenced LNA probe (**scramble**) *and a LNA-probe targeting a region upstream linc-31P TSS* (**INTRON**) *have been used to detect background signals.*
A *probe targeting U6 small-nuclear RNA (**U6**) has been used to obtain a nuclear signal as positive control.*
Original magnification 40x, scale bar = 20μm

Modulation of linc-31 affects *in-vitro* C2C12 differentiation.

Linc-31 levels are high in proliferating cells and start to decrease upon the induction of the differentiation process (**figure 10**). Observing that behaviour that parallels miR-31 expression, we hypothesized a role for our lncRNA in the transition from a proliferation to a differentiation stage.
In order to confirm our hypothesis we decided to modulate its expression through RNA interference and through overexpression experiments in the C2C12 model system. To focus our attention on the transition, both approaches were realized collecting samples from cells in proliferating condition (GM) and after 1, 2 and 3 days from the induction of the differentiation process.
Figure 17 shows the expression profile of two myogenic proteins, myogenin and MyoD (MYOG and Myod) during *in-vitro* differentiation of C2C12 cells. The expression of these proteins was selected to test the effects of linc-31 modulation on the progression of myogenesis.
To obtain the ectopic expression of linc-31D, the cytoplasmic isoform, I realized a construct (**P**linc31-D-ΔDrosha **figure 18A)** mutated in the miR-31 flanking region to prevent the Drosha cleavage of our isoform and miR-31 release. This construct ensures that the effects that we could observe in the myogenesis are due only to the lncRNA overexpression and not to high levels of miR-31 that could be produced by the processing of our construct.
We verified our construct effectiveness transfecting in C2C12 cells both our mutated construct (**P**linc31-D-ΔDrosha) and a non-mutated control construct (Plinc31-D) expressing linc-31 (**Figure 18A).**
As shown in **Figure 18B** sqRT-PCR analysis revealed that both constructs are able to produce linc-31D while the levels of miR-31produced from **P**linc31-D-ΔDrosha are significantly lower compared to the control construct.

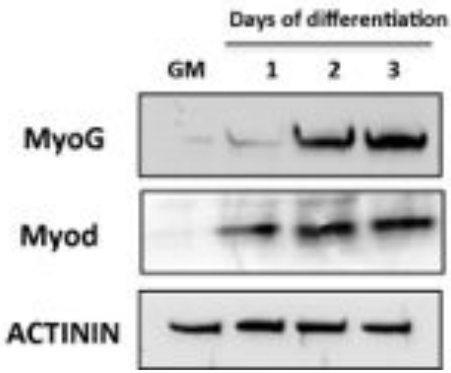


Figure 17: Analysis of muscle differentiation markers in C2C12 cells.
Protein samples were extracted from myoblasts in proliferation (GM) and after shift to differentiation medium for the indicated times. The panel shows a western blot analysis for myogenin (MyoG), myoD, and Actinin as a loading control.

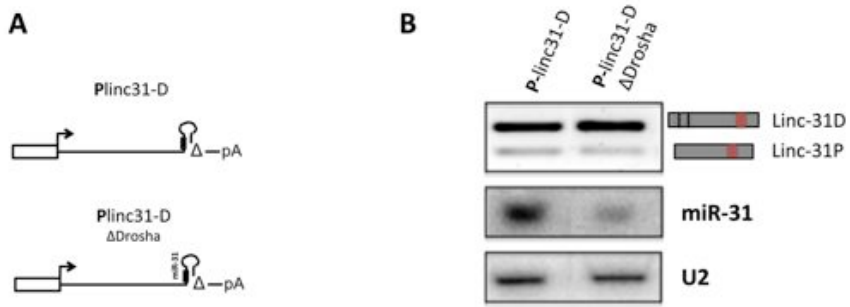


Figure 18: Linc-31P overexpression construct analysis.
A) *Schematic representation of the over-expression constructs.*
B) *RNA samples were collected from myoblast transfected with Plinc31-D and Plinc31-D-ΔDrosha. The upper panel shows linc-31D overexpression analized by sqRT-PCR. GAPDH mRNA (GAPDH) was used as normalization control. Lower panel shows miR-31 expression level in the same cells, analysed by Northern Blot. U2 snRNA is used as endogenous control*

Once verified that the effects that we could observe in the myogenesis are due only to the lncRNA overexpression, we transfected C2C12 cells and collected RNA and protein samples from proliferating cells (GM) (36h post-transfection) and after 1, 2, and 3 from the induction of the differentiation process.
Figure 19A shows the relative quantification of linc-31D produced in C2C12 after the transfection of Plinc31-D-ΔDrosha (**Lnc31-DΔ**) or an empty vector (**ctrl).**
Figure 19B indicates the levels of myogenin and myoD both in control cells and cell overexpressing linc-31D (**Lnc31-DΔ**). It can be visualized that linc-31D over-dosage has no effect on myogenin and MyoD expression at day 1 of differentiation when their level is still low (**figure 19B day 1**). However at day 3 and 5, linc-31D overexpressing cells (**Lnc31-DΔ),** show a 30% decreased levels of myogenic markers of compared to control cells (**ctrl**).
To perform the RNA interference experiment I used a siRNA that specifically targets the linc-31D isoform (**si-lnc31D**) and, as control, a scrambled siRNA (**ctrl**). Even in this case we transfected C2C12 cells and collected RNA and protein samples from proliferating cells (GM) (36h post-transfection) and after 1, 2, and 3 from the induction of the differentiation process.
The relative quantification of linc-31D levels is presented in **figure 20A** showing that a reduction of more than 60% in linc31-D levels in grow condition (**si-lnc31D - GM**) was obtained.
The left panel (**figure 20B**) shows that, the reduction of linc-31D levels by RNA interference (**si-lnc31D**) upon differentiation anticipates the appearance of myogenin and MyoD markers at day 1 and increases their accumulation at later stages with respect to control cells (**ctrl**).
Overall, these results are consistent with the hypothesis that linc-31 might have a role in regulating the transition from proliferation to differentiation.

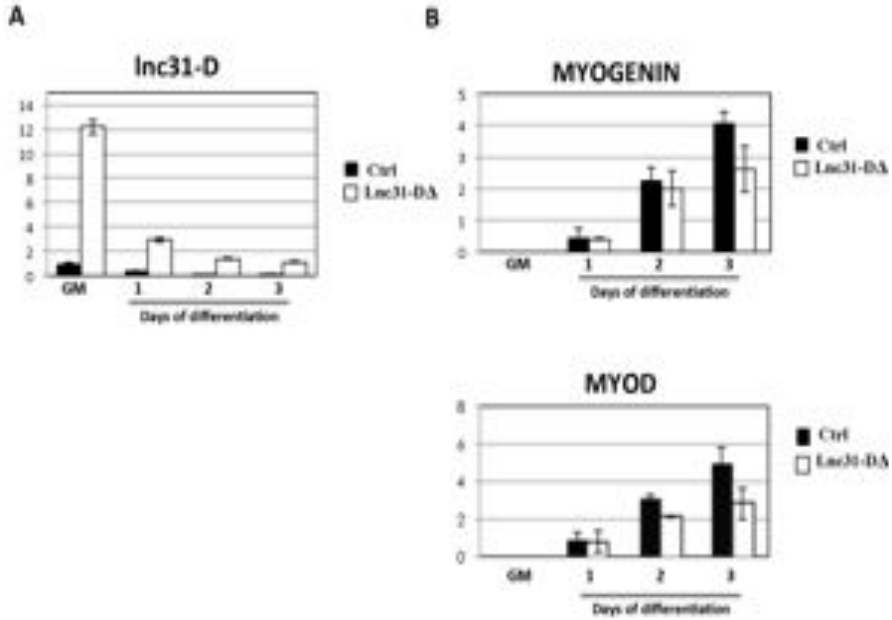


Figure 19: Modulation of Linc-31 affects myogenesis

linc-31D overexpression was obtained by transfection of Plinc31-D-ΔDrosha (plinc31-DΔ) construct together with an empty control vector (Ctrl). Protein and RNA samples were extracted from myoblasts in proliferation (GM) and after shift to differentiation medium for the indicated times.

A) *The graph displays the levels of linc31-D, normalized for GAPDH, measured by qRT-PCR. Data are shown with respect to control experiments in grow medium set to a value of 1. Data are shown as mean ± SD*

B) *The graphs displays the values derived from densitometric analysis of western blot experiments to detect myogenin (upper graph) and MyoD (lower graph) levels.*

Values are normalized for Actinin and are reported with respect to the control sample in grow medium set to a value of 1. Data are shown as mean ± SD

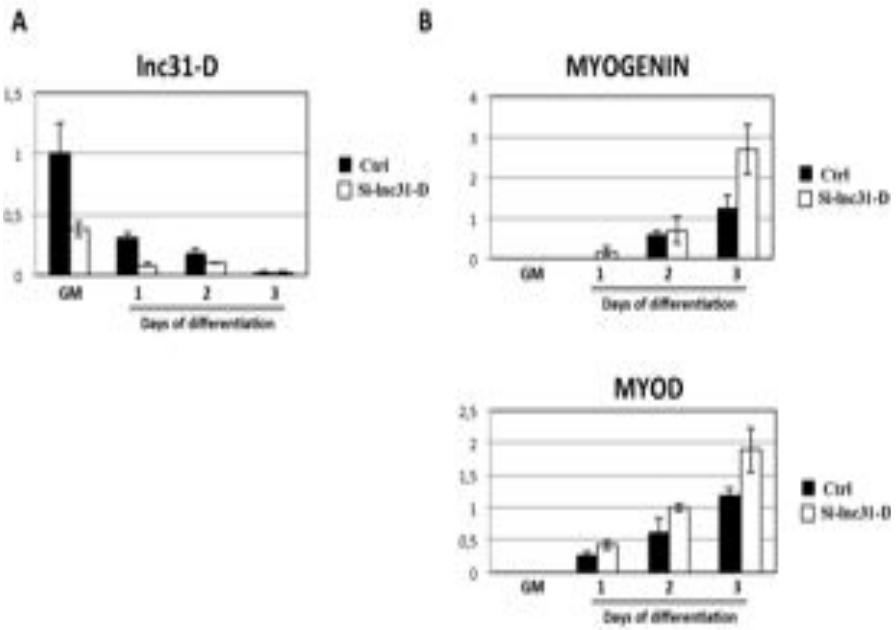


Figure 20: Modulation of Linc-31 affects myogenesis

linc-31D depletion was obtained by transfection in C2C12 cells of a siRNA, specifc for the distal isoform, (si-linc31-D) together with an scramble sequenced control siRNA (Ctrl). Protein and RNA samples were extracted from myoblasts in proliferation (GM) and after shift to differentiation medium for the indicated times.

A) *The graph displays the levels of linc31-D, normalized for GAPDH, measured by qRT-PCR. Data are shown with respect to control experiments in grow medium set to a value of 1. Data are shown as mean ± SD*

B) *The graphs displays the values derived from densitometric analysis of western blot experiments to detect myogenin (upper graph) and MyoD (lower graph) levels.*

Values are normalized for Actinin and are reported with respect to the control sample in grow medium set to a value of 1. Data are shown as mean ± SD

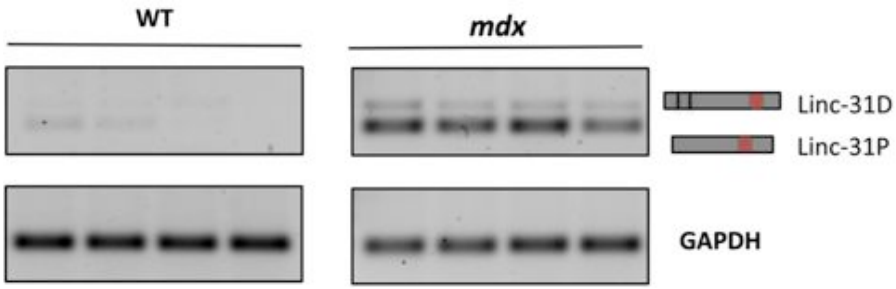
Linc-31 expression in WT and *mdx* mouse muscle tissue

Afterward I analysed the expression of linc-31 in the *mdx* the dystrophin-deficient model mouse to understand if, as many other ncRNAs and miR31(Cacchiarelli et al. 2011), its expression is enriched in DMD pathology. I chose to compare its expression in mdx and WT mice, sacrificed at 6 weeks of age. Four *mdx* mice were sacrificed in parallel with wild type (WT) isogenic/aged matched animals.

Different muscular districts were dissected and we collected RNA samples from gastrocnemii derived from WT and *mdx* animals.

A sqRT-PCR experiment was performed on the samples using specific primers that are able to discriminate between the two Linc-31 isoforms. **Figure 21A** indicates clearly that there is a strong enrichment of this RNA in *mdx* muscle, and that alterated levels of expression can be observed for both isoforms. **Figure 21B** shows the mean level of expression of Linc31P and Linc31D in WT and *mdx,* obtained by densitometric analysis, normalized for GAPDH. Linc31D resulted almost 3 times more expressed in the mdx tissue compared to the WT control; Linc31P resulted even more enriched and its level resulted 7 times higher in the *mdx* mice. Considering the high number of regenerating fibers present at 6 weeks of life in the *mdx* mouse model, the high levels of linc-31 are consistent with the hypothesis that it might have a role in regulating the transition from proliferation to differentiation in myotubes.

A


WT
mdx
Linc-31D
Linc-31P
GAPDH

B

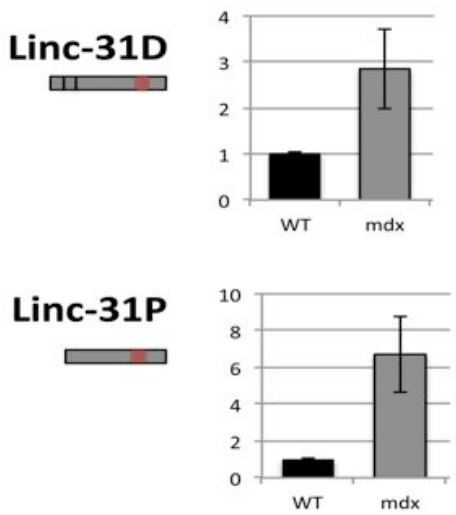

Linc-31D
4
3
2
1
0
WT
mdx
Linc-31P
10
8
6
4
2
0
WT
mdx

Figure 21: Linc-31 expression in WT and *mdx* mouse muscle tissue.

***A**) linc-31D and linc-31P levels were analysed by sqRT-PCR performed on total RNA exctracted from 6 weeks old WT and mdx Gastrocnemius. GAPDH mRNA (GAPDH) was used as normalization control.*

B) *The graphs displays the mean values derived from densitometric analysis of sqRT-PCR in (A) Values are normalized for GAPDH and are reported with respect to the mean value of the WT samples set to a value of 1.*

Linc-31 expression in human myoblasts

A 2009 study has predicted miR-31 to be transcribed from within the first intron of a host gene, LOC554202, on human chromosome 9 (Corcoran et al. 2009). In-silico analyses have confirmed these findings and suggest that LOC554202, which is formed by 4 non-coding exons, could be transcribed into a long non-coding RNA and may have two different transcriptional start sites (Augoff et al. 2012) (**Figure 22A**). The architecture of the locus is not conserved in structure between human and mouse and in the human genome, the miRNA is localized in an intronic portion of the host gene. We have nevertheless demonstrated that there is correspondence in miR-31 expression pattern in murine and human myoblast *in vitro* differentiation (Cacchiarelli et al. 2011), hence I decided to analyse the expression profile of LOC554202 looking for its role in myogenesis.

I analysed the levels of expression by sqRT-PCR on RNA obtained from *in vitro* cultured myoblasts derived from healthy donors (WT) and DMD patients looking for the two putative transcripts. I collected the samples from proliferating cells (GM) and after 1,3 and 5 days from induction of differentiation. Interestingly I identified two different isoforms (**figure 22B)** named Proximal and Distal, depending on their 5' distance from the pri-miR31 sequence. Both are expressed, as in the mouse, at high levels in WT and DMD myoblasts in proliferating conditions (GM) (**figure 23A).**

Upon induction of differentiation linc31-D expression slowly decreases in WT cells but remains quite high in DMD conditions. Linc31-P levels in WT cells, start to decrease at day 1 and almost disappear after 5 days from the induction of differentiation.

In DMD conditions linc-31P levels don't vary at day 1 and remain higher than in control cells with only a slight decrement between day 3 and 5 (**figure 23A)**. Notably, this expression pattern parallels miR-31 expression in the same cells (**figure 23B)**

A

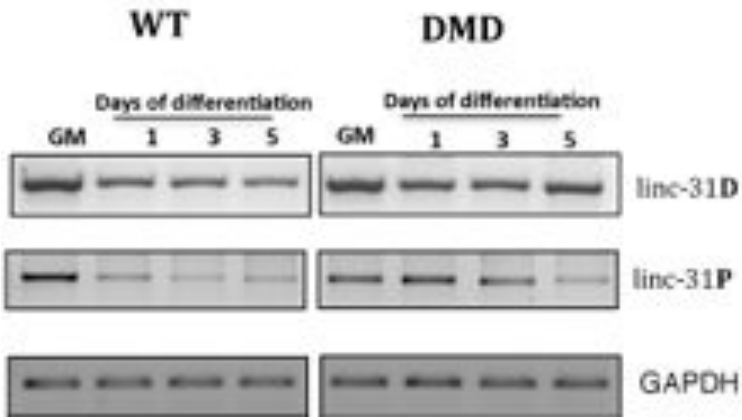
B

Figure 22: linc31 genomic structure in human.

A) *Schematic representation of the human miR-31 genomic locus. The putative transcriptional start sites (TSS) are indicated by arrows*

B) *Schematic representation of LINC31 Proximal (containing miR-31 sequence) and LINC31 Distal transcripts in human cells*

A


WT
DMD
Days of differentiation
Days of differentiation
GM
1
3
5
GM
1
3
5
linc-31D
linc-31P
GAPDH

B

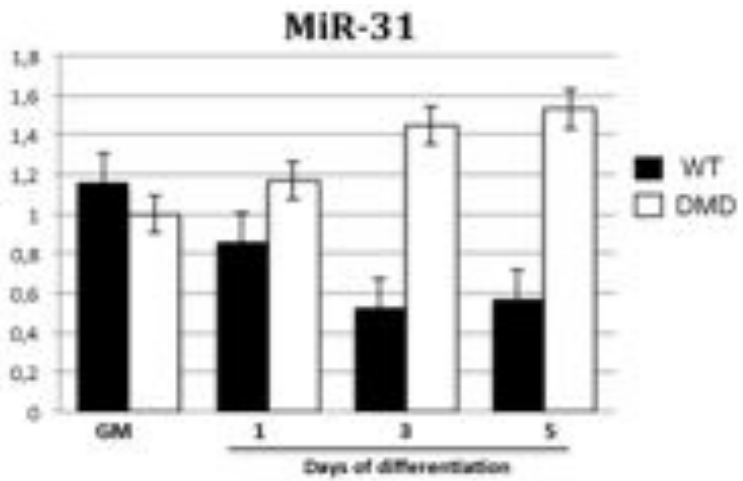

MiR-31
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
WT
DMD
GM
1
3
5
Days of differentiation

Figure 23: Linc-31 expression in human myoblasts

A) *linc-31D and linc-31P levels were analysed by sqRT-PCR performed on total RNA exctracted from WT and DMD myoblasts in proliferation (GM) and after shift to differentiation medium for the indicated times. GAPDH mRNA (GAPDH) was used as normalization control*

B) *qRT-PCR of miR-31 relative expression in human primary myoblasts from healthy donors (WT, black bars) or DMD patients (DMD, White bars) in growth medium (GM) and at the indicated time points after shift to differentiation medium.*

DISCUSSION

The process of generating muscle fibers, also called myogenesis, is a perfectly orchestrated and continuous mechanism that leads to the formation of a multinucleated myofibers capable of contraction.
Myogenesis can be divided into several phases characterized by the expression of muscle specific transcription factors known as myogenic regulatory factors (MRFs) that control the production of developmental stage-specific transcripts. For many years, the attention has been focused only on protein factors, especially MyoD (Ishibashi et al. 2005, Tapscott 2005, Cao et al. 2006), Myf5 (Ishibashi et al. 2005), Myogenin (Cao et al. 2006) and Mrf4 (Kassar-Duchossoy et al. 2004), for their ability to control the conversion of precursor cells into terminally differentiated muscle fibers (Braun et al. 1989, Edmondson and Olson 1989, Rhodes and Konieczny 1989).
Recently, the scientific community focused its attention on the RNA world; genome-sequencing projects in conjunction with transcriptomic analysis have shown that the majority of the genome in animals and plants is transcribed in a developmentally regulated manner to produce large numbers of non–protein-coding RNAs (ncRNAs). The number of these molecules increases with developmental complexity, suggesting that RNA-based regulatory mechanisms may have a relevant role in the evolution of developmental complexity in eukaryotes (Mattick 2011, Nagano and Fraser 2011). It has been now widely demonstrated, that these transcripts could be functional, and have important roles in regulating gene expression at many levels in the cell.
During this intense new RNA-era it has been discovered that, in addition to the coordinated regulation of MRFs, myogenesis also involves the expression of a collection of ncRNAs, able to modulate, at many levels, muscle development and homeostasis.
In particular our group identified several non-coding RNAs (ncRNAs) with a specific role in both muscle differentiation and muscle degenerative pathologies. Among them we studied a micro-

RNA, miR-31, that we found expressed at higher levels in muscle affected by Duchenne muscular dystrophy (DMD).
It is also known that this microRNA is expressed in muscle satellite cells targeting myf5 (Crist et al. 2012) but we were able to demonstrate the importance of this miRNA in the differentiation process and its ability to repress, by targeting its 3' untranslated region, the expression of Dystrophin, a protein that has a key role in muscle fibres contraction and integrity (Cacchiarelli et al. 2011). Moreover we found that miR-31 could be considered a new therapeutic target in DMD because we demonstrated that its modulation enhances our U1snRNA-based exon skipping approach.
Recently, our group also discovered a long non-coding RNA (lncRNA), named linc-MD1, acting as a ceRNA in muscle cells, cross-regulating specific mRNAs by competing for miRNA binding *via* their miRNA recognition motifs (Cesana et al. 2011).
Except few examples, as linc-MD1, there aren't many annotated lncRNAs identified as myogenic regulators so far, and I decided to focus my efforts to get a more comprehensive catalogue of muscle specific lncRNAs and to decipher how these molecules regulate gene expression and chromatin dynamics in this tissue development and homeostasis.
I profiled C2C12 mouse myoblast transcriptome by NGS during proliferation and at different stages of differentiation and I was able to identify, thanks to bioinformatics analysis, a set of non-annotated, polyadenylated putative long non coding RNAs varying their expression during myoblast differentiation, with most of them being up-regulated in myotubes.
I confirmed their existence in C2C12 myoblasts, and studied their expression pattern both in proliferating and differentiated muscle cells and discovered, for some of them, as *lnc-405* different transcriptional isoforms. Since it is known that lncRNAs can be localized in different cellular compartment and that, depending on that localization, they can exert different functions in several molecular mechanisms in the cell (Ponting et al. 2009), before starting a phenotypic analysis resulting from their modulation I investigated their localization inside the cell. I identified many lncRNAs with a specific nuclear localization; we can hypothesize

for them a function, as per many nuclear long noncoding RNAs, in guiding chromatin modifiers to specific genomic loci. There are, in fact evidence, that these molecule are able to recruit DNA and histones methyltransferase and histone modifiers, to mediate transcriptional repression through the formation of silent heterochromatine (Rinn et al. 2007, Nagano et al. 2008, Zhao et al. 2008). On the other hand they can mediate transcriptional activation recruiting chromatin-modifying complexes able to change the three-dimensional architecture of the chromatin (Wang et al. 2011).
To further increase the complexity, it is know that these noncoding transcripts can execute their function on the genome acting both in *cis* and in *trans* (Rinn and Chang 2012).
Finally lncRNAs can also have indirect regulatory effects on many different genes, by acting as decoys that sequester transcription factors (Sun et al. 2013) in the nucleus.
In addition to nuclear lncRNAs I found many other molecules localized in the cytoplasm where they could have a wide array of possible roles. lncRNAs have been described to modulate mRNA stability both increasing (Kretz et al. 2013) or decreasing their stability (Gong and Maquat 2011).
They can also regulate translation by direct base-pairing, in some described cases enhancing it (Carrieri et al. 2012), in other blocking it (Yoon et al. 2012). Finally they could act as ceRNAs (Cesana et al. 2011, Salmena et al. 2011) binding to and sequestering specific miRNAs, to protect their target mRNAs from repression.
I also wanted to asses if the lncRNAs characterized were expressed only in muscle cells and whether their expression changed in WT versus Duchenne conditions; I found that many of them had a muscle specific expression, (*lnc-267*, *lnc-994)* and that others (*lnc-996, lnc-049*, both isoforms of *lnc-149*) are expressed in *mdx* skeletal muscles and almost absent in WT mice.
In addition to this new collection of lncRNA, data obtained from the RNAseq showed a putative ncRNA transcript, that, even if not expressed at high levels, caught our attention, being it transcribed from miR-31 genetic locus.

I showed the existence of two different transcript originating form this locus and I named them linc-31D (Distal) and linc-31P (Proximal) based on TSS position from pre-miR-31.
Both linc-31D and linc-31P contain the pre-miR-31 sequence. Furthermore, both of them resulted upregulated in proliferating muscle cells and decreased with the progression of the myogenesis, paralleling miR-31 expression.
The two isoforms have a different localization: the proximal transcript (linc-31P) is not exported in the cytoplasm but is instead localized in the nucleus and, in particular, it resulted associated with the chromatin, linc-31D isoform instead, is localized in the cytoplasm.
As far as the function of linc-31D is concerned, we showed that its modulation impinged on myogenesis. linc-31D overexpression in mouse myoblasts, in fact, produced a decrease in the accumulation of myogenic markers, Myogenin and MyoD, while its RNAi-dependent downregulation led to their increase.
Overall these findings suggest that linc-31D might have a role in regulating the transition from proliferation to differentiation .
I also analysed the expression of linc-31 in the *mdx* mouse skeletal muscles, to understand if, as many other ncRNAs and miR31 (Cacchiarelli et al. 2011), its expression is enriched in DMD pathology. I confirmed that both isoforms are enriched in DMD conditions and, considering the high number of regenerating fibers present in the *mdx* mouse model, the high levels of linc-31 are consistent with our hypothesis that it might have a role in regulating the exit from the proliferative state.
The architecture of the locus is not conserved in structure between human and mouse and in the human genome, miR-31 is localized in an intronic portion of a host gene with 4 predicted non-coding exons but I demonstrated that as in mouse, in human myoblasts derived from healthy donors (WT) and DMD patients there are two different non-coding isoforms transcribed form the locus and I also found them expressed, as in mouse, at high levels in proliferating conditions both in WT and DMD myoblasts. Upon the induction of the differentiation process I detected a decrease in the two isoforms' levels in WT cells. The lncRNAs levels remains quite high in DMD

conditions corroborating, once again, our hypothesis of their role in regulating the exit from the proliferative state.

MATERIALS AND METHODS

Cell Cultures and Treatments.
C2C12 myoblasts were cultured in growth medium (DMEM high glucose, 2x glutamine, 1X Penicillin/Streptomycin 20% fetal bovine serum). Myogenic differentiation is initiated upon reaching confluence by switching the cells to differentiation medium (DMEM high glucose, 2x glutamine, 1X Penicillin/Streptomycin, 2% horse serum)
Cells were transfected with plasmid DNA using lipofectamine-2000 (Invitrogen).
siRNA molecules designed against linc-31D were transfected using HiPerfect (QIAGEN). All transfections were performed in grow medium according to manufacturer's specifications.
Control and Duchenne primary myoblasts carrying exon 44 deletion (WT-9808 and Δ44-9981) were obtained from the Telethon Neuromuscular Biobank and were first pre-plated in order to separate fibroblasts from the primary line, then seeded in Human Skeletal Muscle Growth Medium (PromoCell, Haidelberg, Germany).
All cell culture are grown in a humidified incubator, at 5% CO2 and 37°C.

Overexpression constructs.
Constructs for the over-expression of linc-31D was obtained by cloning linc-31D cDNA in pCDNA3.1- plasmid (Invitrogen) and the mutant linc-31D-ΔDrosha was obtained by inverse PCR on the first construct.

RACE analysis:
5' RACE analyses were performed using 5' RACE System for Rapid Amplification of cDNA Ends (invitrogen) choosing reverse primers surrounding pre-miRNA sequences.
cDNA synthesis, PCR and nested-PCR were performed according to manufacturer's specifications.

Western blot analysis:
Western blot analyses were carried out as previously described (Incitti et al. 2010). Primary antibodies: anti-myogenin (MyoG sc-12732, 1:1,000 in TBST); anti-MyoD (DAKO, Glostrup, Denmark, 1:500 in 3% milk); anti-actinin (ACTN sc-15335, 1:1000 in TBST). Secondary antibodies: ImmunoPure Goat Anti-Rabbit IgG Peroxidase-Conjugated (Pierce, Rockford, IL, 1:5,000 in 5% milk); ImmunoPure®Goat Anti-Mouse IgG Peroxidase Conjugated (Pierce, 1:10,000 in 5% milk); donkey anti-goat IgG-HRP (sc-2020, diluted 1:5,000 in 3% milk).

RNA preparation and analysis:
Total RNA was prepared from Cells harvested with and liquid nitrogen powdered tissues homogenized in1 ml of QIAzol Lysis Reagent (Qiagen). RNAs were extracted by miRNeasy (Qiagen), following manufacturer's specifications; concentration was assessed with Nanodrop ND-1000 Spectrophotometer (CELBIO; Pero, Milan, Italy).
qRT-PCR were performed using miScript System (Qiagen™). **sqRT-PCR** was performed using SuperScript® III First-Strand Synthesis System (Invitrogen™) and MyTaq™ Hot-Start DNA Polymerase (Bioline).
Primers sequences are available on request.
Northern blots for miRNAs were performed according to Cacchiarelli et al. (2010) using LNA detection probes (Exiqon).

In situ hybridization to interphase nuclei
C2C12 cells were seeded onto polylisinated coverslips and fixed in 2% PFA in PBS for 20 min at 4°C. Cells were firstable permeabilized in 0,5% Triton X-100/2 mM VRC/PBS on ice for 10 min, and then rinsed in 2X SSC prior to hybridization. Hybridization was carried out using custom DIG conjugated-LNA probe (Exiqon) in moist chambre at 53°C for 1 h.
In situ hybridization on DIG-labelled probe was performed with fluorescent monoclonal anti-Digossigenin antibody (JacksonImmunoResearch).
Coverslips were then washed, mounted with Vectashield (Vector

lab) mounting medium.

The sequence of the probe utilized are:
U6: CACGAATTTGCGTGTCATCCTT
INTRON: TAACAGTGCAACAGAGCTACA
LINC31: AATGCAGTGGTCCTTAGAGTGT
SCRAMBLE: GTGTAACACGTCTATACGCCCA

Image acquisition and analysis:
For examination was used Zeiss AxioObserver A1 inverted fluorescence microscope equipped with Axiocam MRM R camera and Plan-Neofluar EC 10X/0,3 M27 and LD 40X/0,6 M27 objectives. The images were acquired with AxioVision Rel.4.8 imaging software.

Statistical analyses:
Each data shown in histograms is the result of at least three independent experiments performed on at least three different samples/animals. Data are shown as mean ± standard deviation. Unless specifically stated, statistical significance of differences between means was assessed by two-tailed t-test and a $p < 0.05$ was considered significant.

ACKNOWLEDGMENTS

I would like to thank the following people who collaborated in different parts of the project: Dr. Cazzella Valentina, Dr. Ballarino Monica, Dr. Martone Julie, Dr. Santini Tiziana, Dr. Majello Sara, Bisceglie Lavinia, Cipriano Andrea, Dr. Morlando Mariangela, Dr. Nicoletti Carmine, Dr. Santini Tiziana, Dr. Sthandier Olga, Dr. Marchioni Marcella, Dr. D'Andrea Daniel, Dr. Grassi Luigi, Prof Tramontano Anna and Prof. Bozzoni Irene.
Dr. M. Mora and the Telethon Neuromuscular Biobank for providing material.

BIBLIOGRAPHY

Aartsma-Rus, A., J. Van Deutekom, I. Fokkema, G.-J. B. Van Ommen and J. Den Dunnen. "Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule." *Muscle & nerve* (2006) **34**(2): 135-144.

Augoff, K., B. McCue, E. F. Plow and K. Sossey-Alaoui. "miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer." *Mol Cancer* (2012) **11**: 5.

Aulehla, A. and O. Pourquié. "On periodicity and directionality of somitogenesis." *Anatomy and embryology* (2006) **211 Suppl 1**: 3-8.

Ballarino, M., F. Pagano, E. Girardi, M. Morlando, D. Cacchiarelli, M. Marchioni, N. J. Proudfoot and I. Bozzoni. "Coupled RNA processing and transcription of intergenic primary microRNAs." *Mol Cell Biol* (2009) **29**(20): 5632-5638.

Bartel, D. and C.-Z. Chen. "Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs." *Nature reviews. Genetics* (2004) **5**(5): 396-400.

Benedetti, S., H. Hoshiya and F. Tedesco. "Repair or replace? Exploiting novel gene and cell therapy strategies for muscular dystrophies." *The FEBS journal* (2013) **280**(17): 4263-4280.

Bentzinger, C., Y. Wang and M. Rudnicki. "Building muscle: molecular regulation of myogenesis." *Cold Spring Harbor perspectives in biology* (2012) **4**(2).

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

Braun, T., G. Buschhausen-Denker, E. Bober, E. Tannich and H. Arnold. "A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts." *The EMBO journal* (1989) **8**(3): 701-709.

Brenman, J., D. Chao, H. Xia, K. Aldape and D. Bredt. "Nitric oxide synthase complexed with dystrophin and absent from skeletal muscle sarcolemma in Duchenne muscular dystrophy." *Cell* (1995) **82**(5): 743-752.

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

Cacchiarelli, D., T. Incitti, J. Martone, M. Cesana, V. Cazzella, T. Santini, O. Sthandier and I. Bozzoni. "miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy." *EMBO reports* (2011) **12**(2): 136-141.

Cacchiarelli, D., J. Martone, E. Girardi, M. Cesana, T. Incitti, M. Morlando, C. Nicoletti, T. Santini, O. Sthandier, L. Barberi, A. Auricchio, A. Musarò and I. Bozzoni. "MicroRNAs involved in molecular circuitries relevant for the Duchenne muscular dystrophy pathogenesis are controlled by the dystrophin/nNOS pathway." *Cell metabolism* (2010) **12**(4): 341-351.

Cai, X., C. H. Hagedorn and B. R. Cullen. "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." *RNA* (2004) **10**(12): 1957-1966.

Cao, Y., R. Kumar, B. Penn, C. Berkes, C. Kooperberg, L. Boyer, R. Young and S. Tapscott. "Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters." *The EMBO journal* (2006) **25**(3): 502-511.

Carrieri, C., L. Cimatti, M. Biagioli, A. Beugnet, S. Zucchelli, S. Fedele, E. Pesce, I. Ferrer, L. Collavin, C. Santoro, A. R. Forrest, P.

Carninci, S. Biffo, E. Stupka and S. Gustincich. "Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat." *Nature* (2012) **491**(7424): 454-457.

Cazzella, V., J. Martone, C. Pinnarò, T. Santini, S. Twayana, O. Sthandier, A. D'Amico, V. Ricotti, E. Bertini, F. Muntoni and I. Bozzoni. "Exon 45 skipping through U1-snRNA antisense molecules recovers the Dys-nNOS pathway and muscle differentiation in human DMD myoblasts." *Molecular therapy : the journal of the American Society of Gene Therapy* (2012) **20**(11): 2134-2142.

Cesana, M., D. Cacchiarelli, I. Legnini, T. Santini, O. Sthandier, M. Chinappi, A. Tramontano and I. Bozzoni. "A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA." *Cell* (2011) **147**(2): 358-369.

Chen, J.-F., T. Callis and D.-Z. Wang. "microRNAs and muscle disorders." *Journal of cell science* (2009) **122**(Pt 1): 13-20.

Chen, J. F., E. M. Mandel, J. M. Thomson, Q. Wu, T. E. Callis, S. M. Hammond, F. L. Conlon and D. Z. Wang. "The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation." *Nat Genet* (2006) **38**(2): 228-233.

Cirak, S., V. Arechavala-Gomeza, M. Guglieri, L. Feng, S. Torelli, K. Anthony, A. Stephen, G. Maria Elena, B. John, J. W. Dominic, D. George, J. A. W. Matthew, D. W. Steve, S. Volker, K. Ryszard, B. S. Stephen, S. Caroline, E. M. Jennifer, B. Kate and F. Muntoni. "Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study." *The Lancet* (2011) **378**.

Collins, C., I. Olsen, P. Zammit, L. Heslop, A. Petrie, T. Partridge and J. Morgan. "Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche." *Cell* (2005) **122**(2): 289-301.

Corcoran, D. L., K. V. Pandit, B. Gordon, A. Bhattacharjee, N. Kaminski and P. V. Benos. "Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data." *PLoS One* (2009) **4**(4): e5279.

Cornelison, D. and B. Wold. "Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells." *Developmental biology* (1997) **191**(2): 270-283.

Crist, C. G., D. Montarras and M. Buckingham. "Muscle Satellite Cells Are Primed for Myogenesis but Maintain Quiescence with Sequestration of Myf5 mRNA Targeted by microRNA-31 in mRNP Granules." *Cell stem cell* (2012) **11**(1): 118-126.

Davies, K. and K. Nowak. "Molecular mechanisms of muscular dystrophies: old and new players." *Nature reviews. Molecular cell biology* (2006) **7**(10): 762-773.

Davis, R. L., H. Weintraub and A. B. Lassar. "Expression of a single transfected cDNA converts fibroblasts to myoblasts." *Cell* (1987) **51**(6): 987-1000.

De Angelis, F., O. Sthandier, B. Berarducci, S. Toso, G. Galluzzi, E. Ricci, G. Cossu and I. Bozzoni. "Chimeric snRNA molecules carrying antisense sequences against the splice junctions of exon 51 of the dystrophin pre-mRNA induce exon skipping and restoration of a dystrophin synthesis in Delta 48-50 DMD cells." *Proceedings of the National Academy of Sciences of the United States of America* (2002) **99**(14): 9456-9461.

Denti, M. A., T. Incitti, O. Sthandier, C. Nicoletti, F. G. De Angelis, E. Rizzuto, A. Auricchio, A. Musaro and I. Bozzoni. "Long-term benefit of adeno-associated virus/antisense-mediated exon skipping in dystrophic mice." *Hum Gene Ther* (2008) **19**(6): 601-608.

Denti, M. A., A. Rosa, G. D'Antona, O. Sthandier, F. G. De Angelis, C. Nicoletti, M. Allocca, O. Pansarasa, V. Parente, A.

Musaro, A. Auricchio, R. Bottinelli and I. Bozzoni. "Chimeric adeno-associated virus/antisense U1 small nuclear RNA effectively rescues dystrophin synthesis and muscle function by local treatment of mdx mice." *Hum Gene Ther* (2006) **17**(5): 565-574.

Dunckley, M. G., M. Manoharan, P. Villiet, I. C. Eperon and G. Dickson. "Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides." *Hum Mol Genet* (1998) **7**(7): 1083-1090.

Edmondson, D. G. and E. N. Olson. "A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program." *Genes & Development* (1989) **3**.

Ervasti, J. and K. Sonnemann. "Biology of the striated muscle dystrophin-glycoprotein complex." *International review of cytology* (2008) **265**: 191-225.

Fatica, A. and I. Bozzoni. "Long non-coding RNAs: new players in cell differentiation and development." *Nature reviews. Genetics* (2014) **15**(1): 7-21.

Goemans, N., M. Tulinius, J. van den Akker, B. Burm, P. Ekhart, N. Heuvelmans, T. Holling, A. Janson, G. Platenburg, J. Sipkens, J. Sitsen, A. Aartsma-Rus, G.-J. B. van Ommen, G. Buyse, N. Darin, J. Verschuuren, G. Campion, S. de Kimpe and J. van Deutekom. "Systemic administration of PRO051 in Duchenne's muscular dystrophy." *The New England journal of medicine* (2011) **364**(16): 1513-1522.

Gong, C. and L. E. Maquat. "lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements." *Nature* (2011) **470**(7333): 284-288.

Guttman, M. and J. Rinn. "Modular regulatory principles of large non-coding RNAs." *Nature* (2012) **482**(7385): 339-346.

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

He, L. and G. J. Hannon. "MicroRNAs: small RNAs with a big role in gene regulation." *Nat Rev Genet* (2004) **5**(7): 522-531.

Heard, E. and C. M. Disteche. "Dosage compensation in mammals: fine-tuning the expression of the X chromosome." *Genes Dev* (2006) **20**(14): 1848-1867.

Hoffman, E. P., K. H. Fischbeck, R. H. Brown, M. Johnson, R. Medori, J. D. Loike, J. B. Harris, R. Waterston, M. Brooke, L. Specht and et al. "Characterization of dystrophin in muscle-biopsy specimens from patients with Duchenne's or Becker's muscular dystrophy." *N Engl J Med* (1988) **318**(21): 1363-1368.

Hu, W., J. Alvarez-Dominguez and H. Lodish. "Regulation of mammalian cell differentiation by long non-coding RNAs." *EMBO reports* (2012) **13**(11): 971-983.

Incitti, T., F. De Angelis, V. Cazzella, O. Sthandier, C. Pinnarò, I. Legnini and I. Bozzoni. "Exon skipping and duchenne muscular dystrophy therapy: selection of the most active U1 snRNA antisense able to induce dystrophin exon 51 skipping." *Molecular therapy : the journal of the American Society of Gene Therapy* (2010) **18**(9): 1675-1682.

Ishibashi, J., R. Perry, A. Asakura and M. Rudnicki. "MyoD induces myogenic differentiation through cooperation of its NH2- and COOH-terminal regions." *The Journal of cell biology* (2005) **171**(3): 471-482.

Kablar, B., K. Krastel, C. Ying, S. Tapscott, D. Goldhamer and M. Rudnicki. "Myogenic determination occurs independently in somites and limb buds." *Developmental biology* (1999) **206**(2): 219-231.

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

Kim, D., P. Saetrom, O. Snøve and J. Rossi. "MicroRNA-directed transcriptional gene silencing in mammalian cells." *Proceedings of the National Academy of Sciences of the United States of America* (2008) **105**(42): 16230-16235.

Kinali, M., V. Arechavala-Gomeza, L. Feng, S. Cirak, D. Hunt, C. Adkin, M. Guglieri, E. Ashton, S. Abbs, P. Nihoyannopoulos, G. Maria Elena, R. Mary, M. Caroline, P. Linda, R. G. Ian, D. George, J. A. W. Matthew, J. W. Dominic, D. W. Steve, K. Ryszard, S. Volker, B. Kate, S. Caroline, E. M. Jennifer and F. Muntoni. "Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study." *The Lancet Neurology* (2009) **8**.

Kino, T., D. E. Hurt, T. Ichijo, N. Nader and G. P. Chrousos. "Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor." *Sci Signal* (2010) **3**(107): ra8.

Koenig, M., A. Beggs, M. Moyer, S. Scherpf, K. Heindrich, T. Bettecken, G. Meng, C. Müller, M. Lindlöf and H. Kaariainen. "The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion." *American journal of human genetics* (1989) **45**(4): 498-506.

Konieczny, P., K. Swiderski and J. Chamberlain. "Gene and cell-mediated therapies for muscular dystrophy." *Muscle & nerve* (2013) **47**(5): 649-663.

Kretz, M., Z. Siprashvili, C. Chu, D. E. Webster, A. Zehnder, K. Qu, C. S. Lee, R. J. Flockhart, A. F. Groff, J. Chow, D. Johnston, G. E. Kim, R. C. Spitale, R. A. Flynn, G. X. Zheng, S. Aiyer, A.

Raj, J. L. Rinn, H. Y. Chang and P. A. Khavari. "Control of somatic tissue differentiation by the long non-coding RNA TINCR." *Nature* (2013) **493**(7431): 231-235.

Krol, J., I. Loedige and W. Filipowicz. "The widespread regulation of microRNA biogenesis, function and decay." *Nature reviews. Genetics* (2010) **11**(9): 597-610.

Lee, Y., M. Kim, J. Han, K. H. Yeom, S. Lee, S. H. Baek and V. N. Kim. "MicroRNA genes are transcribed by RNA polymerase II." *EMBO J* (2004) **23**(20): 4051-4060.

Liu, N., A. Williams, Y. Kim, J. McAnally, S. Bezprozvannaya, L. Sutherland, J. Richardson, R. Bassel-Duby and E. Olson. "An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133." *Proceedings of the National Academy of Sciences of the United States of America* (2007) **104**(52): 20844-20849.

Mann, C., K. Honeyman, A. Cheng, T. Ly, F. Lloyd, S. Fletcher, J. Morgan, T. Partridge and S. Wilton. "Antisense-induced exon skipping and synthesis of dystrophin in the mdx mouse." *Proceedings of the National Academy of Sciences of the United States of America* (2001) **98**(1): 42-47.

Mao, Y. S., H. Sunwoo, B. Zhang and D. L. Spector. "Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs." *Nat Cell Biol* (2011) **13**(1): 95-101.

Matsumura, K. and K. P. Campbell. "Dystrophin-glycoprotein complex: its role in the molecular pathogenesis of muscular dystrophies." *Muscle Nerve* (1994) **17**(1): 2-15.

Mattick, J. "The genetic signatures of noncoding RNAs." *PLoS genetics* (2009) **5**(4).

Mattick, J. S. "Long noncoding RNAs in cell and developmental biology." *Semin Cell Dev Biol* (2011) **22**(4): 327.

Mauro, A. "Satellite cell of skeletal muscle fibers." *The Journal of biophysical and biochemical cytology* (1961) **9**: 493-495.

McCarthy, J. J. "MicroRNA-206: the skeletal muscle-specific myomiR." *Biochim Biophys Acta* (2008) **1779**(11): 682-691.

Monaco, A. P., C. J. Bertelson, S. Liechti-Gallati, H. Moser and L. M. Kunkel. "An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus." *Genomics* (1988) **2**(1): 90-95.

Morlando, M., M. Ballarino, N. Gromak, F. Pagano, I. Bozzoni and N. J. Proudfoot. "Primary microRNA transcripts are processed co-transcriptionally." *Nat Struct Mol Biol* (2008) **15**(9): 902-909.

Nagano, T. and P. Fraser. "No-nonsense functions for long noncoding RNAs." *Cell* (2011) **145**(2): 178-181.

Nagano, T., J. A. Mitchell, L. A. Sanz, F. M. Pauler, A. C. Ferguson-Smith, R. Feil and P. Fraser. "The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin." *Science* (2008) **322**(5908): 1717-1720.

Nigro, G., L. I. Comi, L. Politano and R. J. Bain. "The incidence and evolution of cardiomyopathy in Duchenne muscular dystrophy." *Int J Cardiol* (1990) **26**(3): 271-277.

O'Rourke, J., S. Georges, H. Seay, S. Tapscott, M. McManus, D. Goldhamer, M. Swanson and B. Harfe. "Essential role for Dicer during skeletal muscle development." *Developmental biology* (2007) **311**(2): 359-368.

Paralkar, V. and M. Weiss. "Long noncoding RNAs in biology and hematopoiesis." *Blood* (2013) **121**(24): 4842-4846.

Parker, M., P. Seale and M. Rudnicki. "Looking back to the embryo: defining transcriptional networks in adult myogenesis." *Nature reviews. Genetics* (2003) **4**(7): 497-507.

Poliseno, L., L. Salmena, J. Zhang, B. Carver, W. J. Haveman and P. P. Pandolfi. "A coding-independent function of gene and pseudogene mRNAs regulates tumour biology." *Nature* (2010) **465**(7301): 1033-1038.

Ponting, C., P. Oliver and W. Reik. "Evolution and functions of long noncoding RNAs." *Cell* (2009) **136**(4): 629-641.

Rao, P., R. Kumar, M. Farkhondeh, S. Baskerville and H. Lodish. "Myogenic factors that regulate expression of muscle-specific microRNAs." *Proceedings of the National Academy of Sciences of the United States of America* (2006) **103**(23): 8721-8726.

Rhodes, S. and S. Konieczny. "Identification of MRF4: a new member of the muscle regulatory factor gene family." *Genes & development* (1989) **3**(12B): 2050-2061.

Rinn, J. and H. Chang. "Genome regulation by long noncoding RNAs." *Annual review of biochemistry* (2012) **81**: 145-166.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang. "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs." *Cell* (2007) **129**(7): 1311-1323.

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

Salmena, L., L. Poliseno, Y. Tay, L. Kats and P. Pandolfi. "A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?" *Cell* (2011) **146**(3): 353-358.

Schwartz, J. C., S. T. Younger, N. B. Nguyen, D. B. Hardy, B. P. Monia, D. R. Corey and B. A. Janowski. "Antisense transcripts are targets for activating small RNAs." *Nat Struct Mol Biol* (2008) **15**(8): 842-848.

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

Seitz, H. "Redefining microRNA targets." *Curr Biol* (2009) **19**(10): 870-873.

Sun, F., J. Wang, Q. Pan, Y. Yu, Y. Zhang, Y. Wan, J. Wang, X. Li and A. Hong. "Characterization of function and regulation of miR-24-1 and miR-31." *Biochem Biophys Res Commun* (2009) **380**(3): 660-665.

Sun, S., B. C. Del Rosario, A. Szanto, Y. Ogawa, Y. Jeon and J. T. Lee. "Jpx RNA activates Xist by evicting CTCF." *Cell* (2013) **153**(7): 1537-1551.

Taft, R., M. Pheasant and J. Mattick. "The relationship between non-protein-coding DNA and eukaryotic complexity." *BioEssays : news and reviews in molecular, cellular and developmental biology* (2007) **29**(3): 288-299.

Tajbakhsh, S. "Skeletal muscle stem cells in developmental versus regenerative myogenesis." *Journal of internal medicine* (2009) **266**(4): 372-389.

Tajbakhsh, S., D. Rocancourt and M. Buckingham. "Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice." *Nature* (1996) **384**(6606): 266-270.

Tapscott, S. "The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription." *Development (Cambridge, England)* (2005) **132**(12): 2685-2695.

Valastyan, S. and R. A. Weinberg. "miR-31: a crucial overseer of tumor metastasis and other emerging roles." *Cell Cycle* (2010) **9**(11): 2124-2129.

van Deutekom, J., M. Bremmer-Bout, A. Janson, I. Ginjaar, F. Baas, J. den Dunnen and G. van Ommen. "Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells." *Human molecular genetics* (2001) **10**(15): 1547-1554.

van Deutekom, J., A. Janson, I. Ginjaar, W. Frankhuizen, A. Aartsma-Rus, M. Bremmer-Bout, J. den Dunnen, K. Koop, A. van der Kooi, N. Goemans, S. de Kimpe, P. Ekhart, E. Venneker, G. Platenburg, J. Verschuuren and G.-J. B. van Ommen. "Local dystrophin restoration with antisense oligonucleotide PRO051." *The New England journal of medicine* (2007) **357**(26): 2677-2686.

Vasudevan, S., Y. Tong and J. Steitz. "Switching from repression to activation: microRNAs can up-regulate translation." *Science (New York, N.Y.)* (2007) **318**(5858): 1931-1934.

Wang, K. C., Y. W. Yang, B. Liu, A. Sanyal, R. Corces-Zimmerman, Y. Chen, B. R. Lajoie, A. Protacio, R. A. Flynn, R. A. Gupta, J. Wysocka, M. Lei, J. Dekker, J. A. Helms and H. Y. Chang. "A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression." *Nature* (2011) **472**(7341): 120-124.

Wang, Z., M. Gerstein and M. Snyder. "RNA-Seq: a revolutionary tool for transcriptomics." *Nature reviews. Genetics* (2009) **10**(1): 57-63.

Willingham, A. T., A. P. Orth, S. Batalov, E. C. Peters, B. G. Wen, P. Aza-Blanc, J. B. Hogenesch and P. G. Schultz. "A strategy for probing the function of noncoding RNAs finds a repressor of NFAT." *Science* (2005) **309**(5740): 1570-1573.

Willmann, R., S. Possekel, J. Dubach-Powell, T. Meier and M. Ruegg. "Mammalian animal models for Duchenne muscular dystrophy." *Neuromuscular Disorders* (2009) **19**.

Wilton, S., F. Lloyd, K. Carville, S. Fletcher, K. Honeyman, S. Agrawal and R. Kole. "Specific removal of the nonsense mutation from the mdx dystrophin mRNA using antisense oligonucleotides." *Neuromuscular disorders : NMD* (1999) **9**(5): 330-338.

Yaffe, D. and O. Saxel. "A Myogenic Cell Line with Altered Serum Requirements for Differentiation." *Differentiation* (1977) **7**(1-3): 159-166.

Yin, H., F. Price and M. Rudnicki. "Satellite cells and the muscle stem cell niche." *Physiological reviews* (2013) **93**(1): 23-67.

Yoon, J. H., K. Abdelmohsen, S. Srikantan, X. Yang, J. L. Martindale, S. De, M. Huarte, M. Zhan, K. G. Becker and M. Gorospe. "LincRNA-p21 suppresses target mRNA translation." *Mol Cell* (2012) **47**(4): 648-655.

Zhao, J., B. K. Sun, J. A. Erwin, J. J. Song and J. T. Lee. "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." *Science* (2008) **322**(5902): 750-756.

Zhao, Y., E. Samal and D. Srivastava. "Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis." *Nature* (2005) **436**(7048): 214-220.

LIST OF PUBLICATIONS:

Incitti T, De Angelis FG, Cazzella V, Sthandier O, **Pinnarò C**, Legnini I, Bozzoni I.
Exon skipping and duchenne muscular dystrophy therapy: selection of the most active U1 snRNA antisense able to induce dystrophin exon 51 skipping.
Mol Ther. 2010 Sep;18(9):1675-82.

Cazzella V, Martone J, **Pinnarò C**, Santini T, Twayana S, Sthandier O, D'Amico A, Ricotti V, Bertini E, Muntoni F, Bozzoni I.
Exon 45 skipping through U1-snRNA antisense molecules recovers the Dys-nNOS
pathway and muscle differentiation in human DMD myoblasts.
Mol Ther. 2012 Nov;20(11):2134-42