SAPIENZA
Università di Roma
Facoltà di Scienze Matematiche Fisiche e Naturali

DOTTORATO DI RICERCA
IN GENETICA E BIOLOGIA MOLECOLARE

XXVIII Ciclo
(A.A. 2014/2015)

# Circular RNAs Expression and Function in Myogenesis

Dottorando
Ivano Legnini

Docente guida
Prof. Irene Bozzoni

Tutore
Prof. Anna Tramontano

Coordinatore
Prof. Silvia Bonaccorsi

# Index

# Preface

In this work I will present the results obtained in the last three years of my career during my post-graduate studies in Genetics and Molecular Biology, with a main focus on two projects that I ran in the Prof. Irene Bozzoni's lab at Sapienza University of Rome.

The first project that I will discuss is about a long non-coding RNA that was discovered in the lab while I was working at my master's degree thesis (Cesana et al., 2011), named *linc-MD1*. We found that this molecule is able to induce differentiation of myoblasts by acting as a decoy for two microRNAs, miR-133 and miR-135, thereby releasing the expression of their natural targets. During the ending of my master's and the beginning of my PhD, I worked on a follow-up project regarding linc-MD1 biosynthesis and found that one of the targets of miR-133 is the RNA binding protein HuR, which in turn sustains linc-MD1 production by inhibiting its cleavage by the RNase Drosha (Legnini et al., 2014). These results, as they have been published two years ago and were part of a project that I started before my post-graduate education, will be included in a dedicated section of the introduction chapter, after briefly reviewing the minimal knowledge about the biology of muscle differentiation and the problem of non-coding RNA needed to appropriately understand the frame of this work. A brief part about different approaches to quantitative modelling of how microRNA decoys can work will follow (a field I was not directly involved in, but that I actively followed and discussed about with some of the authors of the papers that I will present). I will finally introduce the second main project I've been working on, that is the primary object of this thesis, about a new and enigmatic class of RNA molecules named *circular RNAs*. These molecules originate as covalently closed ribonucleic chains after a particular splicing reaction called *back-splicing*, they are present across the whole

eukaryotic domain and they are particularly abundant in mammals. Although some key features of their biosynthesis have been recently revealed, almost nothing is known about their function. In this work I tried to understand the basis of circular RNA expression dynamics in muscle differentiation and found that they seem to exert important functions in such biological process. After reeling off the little but fundamental knowledge accumulated about circular RNAs over the past couple of years, I will discuss the aim of this thesis and start describing the results I obtained. Discussion will be included in the same chapter, as each of the main findings will be described and commented at the same time. Some general concluding remarks are reported in the following chapter. Finally, after a detailed description of the techniques used in this work and a brief glossary for the non-expert reader, I will list the bibliography cited in the thesis and the publications I've been involved in as an author, together with a synopsis of the findings of each paper and my contribution to them.

# Summary

Skeletal myogenesis is a well-characterized biological process driving the formation of the adult muscle tissue from mesodermal progenitors. The key molecular and cellular steps of such process can be replicated in vitro, by culturing muscle precursor cells called *myoblasts*, which can be induced to differentiate into mature *myotubes*. We characterized the role of a long non-coding RNA, linc-MD1, in regulating in vitro differentiation of murine myoblasts by acting as a molecular decoy for two microRNAs, miR-133 and miR-135, and we further dissected the regulation of linc-MD1 biosynthesis by the RNA binding protein HuR. Searching for new non-canonical RNA species involved in myogenesis, we directed our attention to circular RNAs (circRNAs). Circular RNAs are produced by the spliceosome via a particular reaction called *back-splicing*, which links a donor splice site to an upstream acceptor site, thereby generating a covalently closed RNA molecule. We performed high-throughput expression profiling of circRNAs in murine and human cells and studied the principles of their regulation during myogenesis. We then applied a high-content functional genomic screen of conserved circRNAs and found that these molecules are actively involved in the control of myoblasts differentiation. Among them, we further characterized *circ-ZNF609* and found that it is able to control myoblast proliferation, providing the first example of a circular RNA involved in a relevant biological process. Moreover, sequence and biochemical analyses suggested that circ-ZNF609 might be translated in a functional protein, representing a possible molecular mechanism for circRNA function.

# Introduction

## 1. Muscle differentiation from embryology to the new frontiers of genomic science.

### 1.1 The core myogenic gene expression program relies on a transcriptional activation cascade.

The development of the skeletal muscle system of vertebrates occurs through a complex series of morphogenetic events, globally called *myogenesis*. During embryo development, muscles of the trunk and limbs originate from progenitor cells located in the somites which are able to respond to inductive signals, mainly coming from the notochord and the neural tube, with the activation of a muscle-specific gene expression program (Asakura et al., 2002). Those proliferating cells, also called *myoblasts*, are able to delaminate and migrate, eventually entering the differentiation stage that leads to the formation of mature *myotubes* (figure 1). Differentiation of myoblats into myotubes involves three key steps: cell cycle arrest, orientation and cell fusion. Myogenesis occurs also in the adult life to allow muscle growth and maintenance or after exercise and injury, and it is usually initiated by a population of stem cells located at the periphery of muscle fibers, known as *satellite cells.* The whole process of differentiation is directed by a complex gene expression program initiated with the contribution of a specific group of transcription factors called *myogenic regulatory factors* (MRFs).

MRFs expression is tightly regulated in time and space, restricted to the muscle tissue and leads to the activation of the core gene expression program required for muscle differentiation. They all contain a basic helix-loop-helix domain which is able to recognise

the consensus binding motif CANNTG, known as *E-box*, in enhancer and promoters of muscle-specific genes (Tapscott, 2005).

Certain MRFs are required for the initial specification of muscle progenitor cells into myoblasts and sustain their proliferative state, whereas other factors, subsequently activated by the early ones, control the terminal differentiation stage of myoblasts. The master genes of myoblast specification are *Myogenic factor 5* (Myf5) and *Myogenic Differentiation 1* (MyoD). Although they're not expressed exactly at the same time during embryogenesis, only knock-out mice for both factors have complete deficiency of muscle-tissue, while single knock-outs do not, indicating that these two MRFs act in a redundant fashion, thereby providing robustness to the myogenic program (Rudnicki et al., 1993). Early MRFs are regulated by the paired domain and homeobox-containing proteins *Pax3* and *Pax7*. While the first is more active in the embryo (in turn activated by *Six* and *Eya* family proteins (Bajard et al., 2006), the second is the key activator of adult myogenesis, being a common marker of activated satellite cells (Seale et al., 2003). Terminal differentiation of myoblasts is driven by a second activation cascade of gene expression led by later MRFs, including *Myogenin* and *Mrf4* (also called *Myf6*). While Mrf4 plays a role in progenitor cells specification with MyoD and Myf5 (Hasty et al., 1993 and Patapoutian et al., 1995), Myogenin is considered the typical late myogenic factor, whose action is required for transcription of the genes necessary for the architecture and function of mature muscle fibers (e.g. myosins, muscle creatine kinase, dystrophin etc,). MRFs act together with the *Serum Response Factor* (SRF) and another family of transcription factors, known as *Myocyte Enhancing Factor 2 family* (MEF2) which contain a MADS box domain and are able to bind DNA carrying a CC[A/T]$_6$GG motif (CarG-box) (Dodou et al., 2003). It has also been shown that the protein *Mastermind-like 1* (Maml1), known for participating in in the Notch signaling pathway, has also a Notch-independent function in the muscle lineage, being able to

form a transcriptional activator complex with the Mef2 family member Mef2C (Wilson-Rawls et al. 1999).

Myoblasts differentiation can be replicated *in vitro* by culturing myoblasts or satellite cells in growth factors-containing media that keep the cells proliferating, until these are confluent enough to touch one another and eventually are exposed to serum starvation. In these conditions, cultured cells are able to exit the cell cycle, orientate and eventually fuse into multi-nucleated syncytia as happening in naturally occurring differentiation (Yaffe et al., 1965, figure 2).

## Figure 1

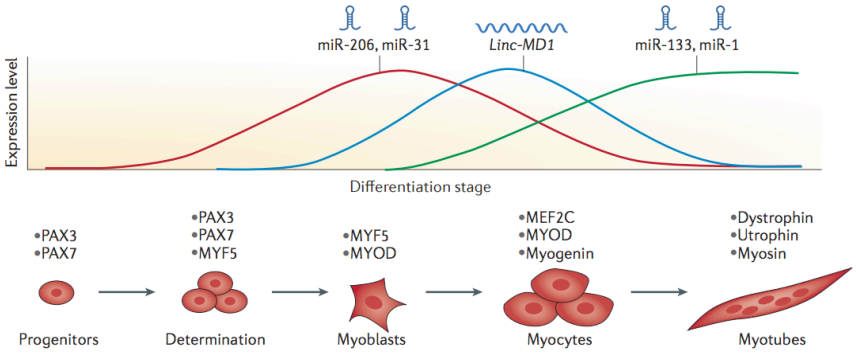



*Figure 1. Schematic representation of the differentiation stages from progenitor muscle cells to terminally differentiated fibres is shown. The cells are labelled with the characteristic proteins that are expressed at each stage. These include master transcription factors that regulate the switch from one stage to the following one - such as paired box protein Pax-3 (PAX3), PAX7, myogenic factor 5 (MYF5), myoblast determination protein (MYOD), myocyte enhancer factor 2C (MEF2C) and myogenin - as well as the late myogenic proteins dystrophin, utrophin and myosin123. The graph shows the corresponding temporal expression patterns of selected non-coding RNAs. Adapted from Fatica et al., (2015), reproduced under licence N° 3786520528882.*

**Figure 2**

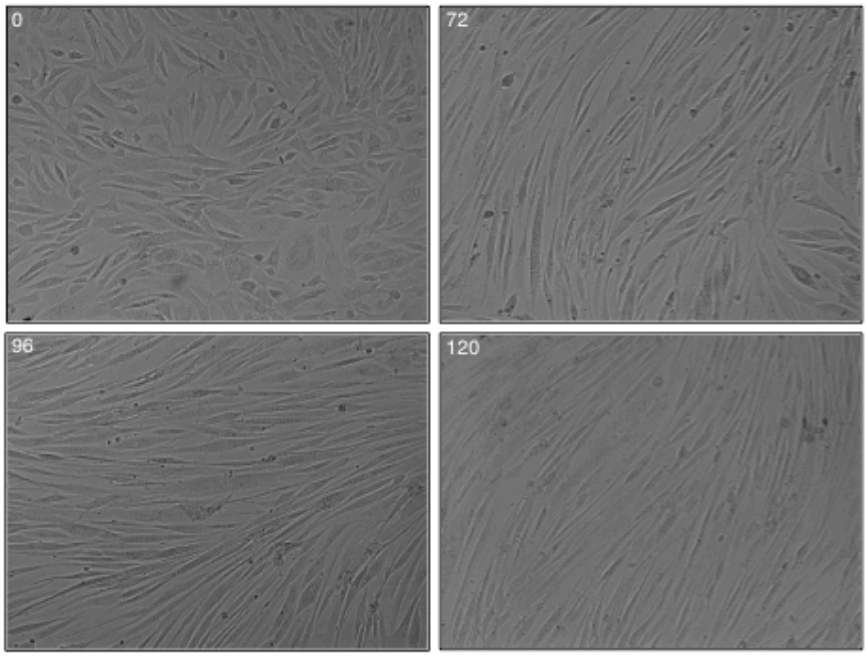



*Figure 2. Bright-field pictures of in vitro cultured human primary myoblasts at different times of differentiation (hours). Cells are initially observed as single disordered objects, then they elongate and become aligned, eventually they fuse to form myotubes.*

## 1.2 The study of non-coding RNAs and system biology approaches add new information on myogenesis.

Not only transcription factors participate to defining the myogenic gene expression program: more recently, also non-coding RNAs have been linked to the regulation of myogenesis (figure 1). Recent studies revealed that myoblasts proliferation, differentiation and muscle physiology are also regulated by a set of muscle-specific microRNAs, whose expression is directly induced by myogenic transcription factors. The best widely studied muscle-specific microRNAs belong to the *miR-1* and *miR-133* families. These microRNAs are produced from three bi-cistronic clusters: miR-1-1/miR-133a-2, miR-1-2/miR-133a-1 and miR-206/miR-133b. Another group of muscle-specific microRNAs is the *miR-208a/b* family, whose sequence is embedded in the MYH6 and MYH7 myosin genes, and whose expression is restricted to cardiomyocytes, where they play a relevant role in cardiac development and stress-dependent remodelling (van Rooij et al., 2007). The pivotal role of microRNAs in muscle biology was demonstrated by analysing Dicer-null mice, which have severe cardiac defects (da Costa Martins et al., 2007). Moreover, transcriptional activation of the miR-1/miR-133a loci relies on the binding of MyoD, MEF2 and SRF, thus implying a possible part of these molecules in the MyoD-dependent control of differentiation (Zhao et al., 2005 and Rao et al., 2006). Several studies demonstrated a central role of these microRNAs in myoblasts proliferation and differentiation, a role fulfilled by blocking the expression of genes like CyclinD2, Cdc42, IRX5, SRF, Connexin43, HDAC4, G6PD and many others (Chen et al., 2006 and Cacchiarelli et al. 2010). MiR-206 is likely to have an independent function, as it is able to control the activation of satellite cells by repressing Pax7 (Cacchiarelli et al. 2010), besides a common function in myoblasts where it targets multiple genes responsible for proliferation (Anderson et al., 2006 and Gagan et al., 2012). Another important role is played by *miR-31*, which is a ubiquitous microRNA, very abundant in proliferating myoblasts

and down-regulated after differentiation. It is able to target Dystrophin (Cacchiarelli et al., 2011), a tardive marker of mature myotubes, but also Myf5, a target that may result paradoxical for the expression pattern of miR-31. On the contrary, in quiescent satellite cells, which are primed for myogenesis but are poised until the proper external signals occur, miR-31 activity is crucial, as the targeting of Myf5 prevents the cells entering the active proliferative state that foreshadows their differentiation. Such block is conveniently transient, since the Myf5 mRNA is confined to mRNP granules where it is stored until activation of quiescent cells occurs (Crist et al., 2012).

Although that complex gene expression regulatory network, made of transcription factors and microRNAs, has been deeply dissected in each of its single components, a system biology approach to understand the general principles behind the regulation of myogenesis in time and space is still lacking. A few efforts for capturing a more global picture of the regulatory events occurring in myoblast differentiation have been made in the recent past.

Particularly, the rise of single cell sequencing allowed dissecting the trajectory of the myogenic gene expression program by using an approach called *pseudo-temporal ordering*. One of main obstacles to the exact understanding of the temporal evolution of the transcriptome in cell populations is their asynchronicity. Sequencing RNA from many single cells within a population instead of the bulk itself allows dramatically increasing the information regarding the multiple states where those cells lie during their differentiation. By minimizing the distance in terms of gene expression between those states and maximizing the distance between the first and the last state, an algorithm called *Monocle* is able to create a pseudo-temporal trajectory that may correspond to the actual temporal evolution of a given cell during differentiation (Trapnell et al., 2014).

Another example is a ChIP-seq study of MyoD and Myogenin, which revealed that such transcription factors, in addition to well-

known muscle-specific genes, extensively bind to extra-genic regions enriched for H3K4 monomethylation, a landmark of enhancer activity (Mousavi et al., 2013). Those regions were also enriched for H3K27 acetylation and PolII and, as expected, peaks of RNA synthesis were mapped to the same sites. Such enhancer RNAs (eRNAs) were found also in previously known MyoD activating elements, playing an active role in chromatin remodelling of its own locus.

A third important contribution to a wider understanding of myogenesis from the point of view of the transcriptome consisted of the profiling of long non-coding RNAs during in-vitro differentiation of murine myobalsts (Ballarino et al., 2015). In the past ten years, an unprecedented attention was directed towards long non-coding RNAs (lncRNAs) or large intervening non-coding RNAs (lincRNAs). Particularly, international projects such as Phantom and Encode revealed with dozens of genomic and transcriptomic reports that a large portion of the mammalian genomes is transcribed, producing a huge number of non-coding RNAs in a tissue-specific and regulated fashion. Many efforts were therefore focused on the understanding of the possible functions of long non-coding RNAs, finding that actually many of them can be involved in various biological processes (reviewed in Fatica et al., 2013). In Ballarino et al. (2015), 29 new long non-coding RNAs whose expression was regulated during differentiation were annotated and experimentally validated. Among them, the long non-coding RNA *lnc-31*, which is the host transcript of miR-31, was identified as a crucial regulator of myoblast proliferation. This work followed the identification of the first long non-coding RNA involved in myogenesis, *linc Muscle Differentiation 1*, or linc-MD1, to which the next paragraph will be dedicated.

# 2. The role of linc-MD1 and HuR in myogenesis.

**Note:** some of the figures included in this section are part of two published papers whose copyright is owned by Elsevier (doi: 10.1016/j.cell.2011.09.028 and 10.1016/j.molcel.2013.12.012). According to Elsevier copyright policy, their inclusion in this thesis does not constitute a copyright infringement (license N° 3785820559384).

## 2.1. The miR-206/133 locus produces a stable cytoplasmic long non coding RNA.

MicroRNAs are usually processed from longer precursors in two steps of endonucleolytic cleavage. The first one is realized by *Drosha*, an component of the multi-protein complex called *Microprocessor*, which is loaded onto the nascent primary transcript (pri-miR), then recognises and cuts the common short hairpin structure that distinguishes almost all microRNA precursors. Pri-miRs can be protein-coding genes (carrying the short hairpin in introns or exons) but also non-coding transcripts, which were thought to exhaust their function by giving rise to the mature form of the embedded microRNA. The study of linc-MD1 revealed that exactly like mRNAs, the non-coding product of a microRNA-containing gene can be functional on its own (Cesana et al., 2011). As described previously in this thesis, miR-206 and miR-133b are two fundamental myogenic factors and their expression is linked to the same genomic locus. While studying their biosynthesis, we discovered by 5' RACE, ChIP and conservation analysis that such locus contains two main promoters (called *Dist* and *Prox* for "distal" and "proximal" to miR-206/133b, figure 3), both bound by MyoD. While the second seems to produce mainly the microRNA precursor transcript, the distal regulatory element sustains the synthesis of a multi-exonic, polyadenylated stable RNA, linc-MD1. Such long non-coding

RNA is exported to the cytoplasm, where it seems crucial for the proper differentiation of murine myoblasts.

### 2.2. Linc-MD1 acts as a decoy for miR-133 and miR-135.

Given the fact that linc-MD1 is accumulated during in vitro myogenesis (figure 3), we speculated that it might be involved in such process. Two complementary approaches were used to test the supposition: siRNAs-mediated knock-down and cDNA overexpression from a plasmid vector, followed by western blot analysis of Myogenin and Myosin Heavy Chain, an early and late marker of differentiation respectively. As hypothesized, both of the markers were decreased as consequence of linc-MD1 down-regulation and increased after up-regulation (figure 4). At this, point, we sought to understand the mechanism underlying such impact on myoblast differentiation. One idea came from some reports that had been published shortly before (Poliseno et al., 2010 and Wang et al., 2010), stating that non-coding RNAs could act as "sponges" for microRNAs, thus regulating the expression of mRNAs that share a similar pool of microRNA target sites. The whole model relies on the idea that if two transcripts are targeted by the same microRNAs, a perturbation in the concentration of one of the two targets is transferred to the other because it impacts on the concentration of free (non-bound) microRNAs (Salmena et al., 2011). We therefore looked for microRNA binding sites and their possible targets that could explain the pro-myogenic activity of linc-MD1. Bioinformatic predictions, confirmed by luciferase assays, revealed that two microRNAs, miR-133 and miR-135, are able to bind linc-MD1 sequence. We then screened possible targets of these two microRNAs for identifying genes that could be responsible for the phenotype associated to linc-MD1 modulation, and selected for further validation Maml1 and Mef2C. As stated before in this chapter, these two proteins are able to interact and form a transcriptional activator complex that sustains myogenesis. Again, knock-down and overexpression of linc-MD1 in the

presence and absence of the two microRNAs, as well as luciferase assays, were compatible with the model of the microRNA sponge, suggesting that linc-MD1 sustains Maml1 and Mef2C expression by sequestrating miR-133 and miR-135 (figure 5).

But at this point a few questions were still open. First, how could linc-MD1 escape Drosha processing and accumulate in the cytoplasm of differentiating myoblasts? Second, what's the functional meaning of a locus producing a microRNA and a decoy of the same microRNA? And third, what are the quantitative and qualitative requirements of linc-MD1 for being able to impact on microRNAs availability? I will try to answer those questions in the following paragraphs.

## Figure 3

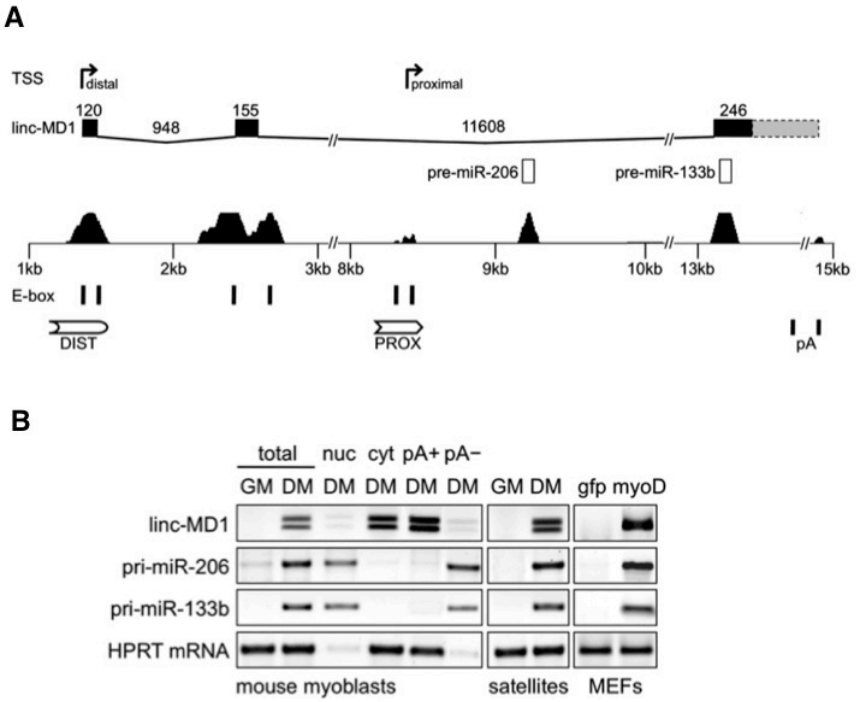



*Figure 3. The miR-206/133b locus produces a long non coding RNA. A. Genomic organization of the miR-206/133b locus with linc-MD1 exons (in black, an alternative 3' end in grey), the two pre-microRNAs (in white), the conservation score in the middle and the regulatory elements at the bottom. B. RT-PCR analysis of linc-MD1, pri-miR-206 and 133b and HPRT is a loading control from undifferentiated and differentiated myoblasts (GM and DM), nuclear and cytoplasmic fractions, polyadenylated and non-polyadenylated fractions, as well as from undifferentiated and differentiated satellite cells and mouse embryo fibroblasts (MEF). Adapted from Cesana et al. (2011), reproduced under license N° 3785820559384.*

## Figure 4

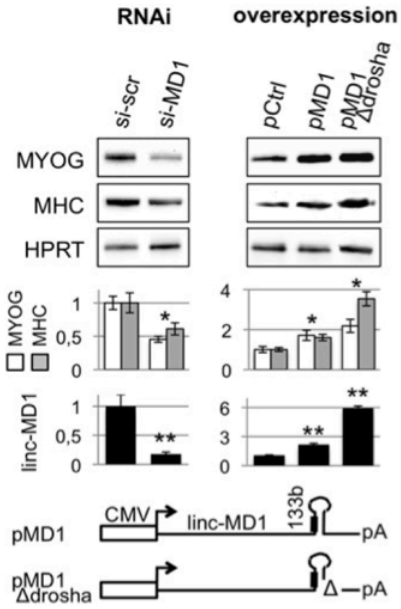



*Figure 4. Knock-down and overexpression of linc-MD1 affects myoblasts differentiation. In the upper panels representative western blots for Myogenin and MHC in control and linc-MD1 knock-down (left) or overexpression (right) are shown. Below is shown a bar plot for western blot densitometric quantification of three independent experiments (white bars for Myogenin, grey bars for MHC). Black bar plots show linc-MD1 levels upon RNAi and overexpression. In the bottom panel a schematic representation of the vectors used for linc-MD1 overexpression is shown. Adapted from Cesana et al. (2011), reproduced under license N° 3785820559384.*

**Figure 5**

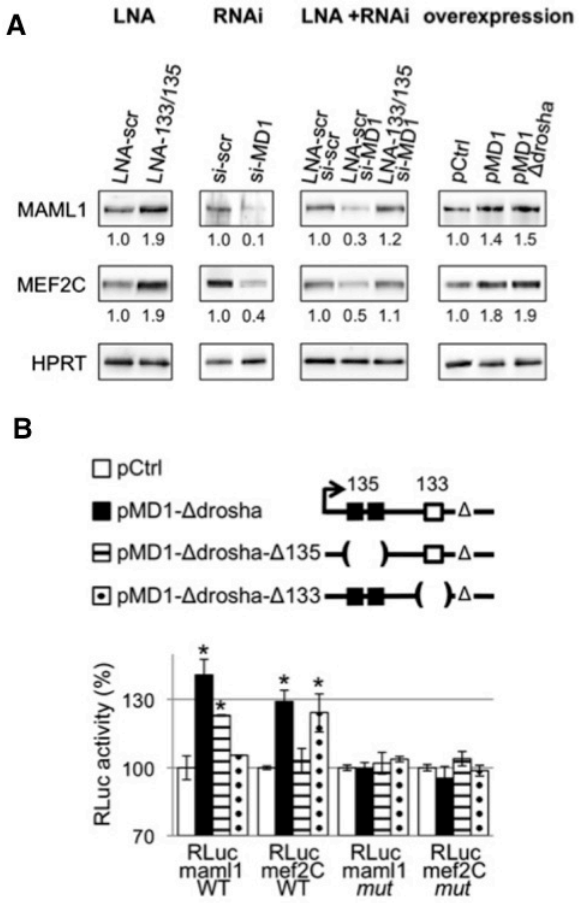



*Figure 5. A. linc-MD1 regulates Maml1 and Mef2C in a microRNA-dependent manner. Western blot for Maml1 and Mef2C in control and linc-MD1 knock-down (left blots) or overexpression (right blots) in the presence or absence of miR-133/135-blocking LNAs. HPRT is used as loading control. B. Upper panel shows the constructs tested in the lower panels. Lower panel: luciferase analysis of the Maml-1 and Mef2C 3'UTRs (WT or mutated in the miR-133/miR-135 binding sites) upon overexpression of linc-MD1 (WT or mutated in miR-133/135 binding sites. Adapted from Cesana et al. (2011), reproduced under license N° 3785820559384.*

## 2.3 A feed-forward regulatory loop between linc-MD1 and HuR regulates the timing of in vitro myoblasts differentiation.

Very often in the history of molecular biology the observation of a gene producing a phenotype was explained by the postulation of a linear pathway, made of a series of elements connected one to each other hierarchically, tested by the observation of epistatic mutants or with complementation experiments. This view has been radically revised with the advent of system biology and *-omic* technologies, which try to capture the global picture of a biological phenomenon, confirming existing linear pathways but finding that those are usually embedded in much more intricate networks. In the case of linc-MD1, we were aware that not only two microRNA targets out of potential hundreds more could explain its function. We therefore looked for additional putative targets of miR-133 and miR-135 that could result interesting for understanding more about this gene. Among the predicted miR-133 targets, we found HuR (Legnini et al., 2014), a gene coding an RNA binding protein that is implicated in muscle differentiation, as it is able to confer stability to pro-myogenic mRNAs such as MyoD, Myogenin and p21 (Figueroa et al., 2003 and von Roretz et al., 2011). This protein interacts with a large number of RNAs (Lebedeva et al., 2011 and Mukherjee et al., 2011) and the functional meaning of those interactions has been object of extensive study and speculation. One of the intriguing jobs of HuR is its ability to control microRNA biogenesis and microRNA target recognition. Particularly, it is able to block miR-7 biogenesis by direct binding to its precursor together with the RNA binding protein Musashi 2 (Choudhury et al., 2013), and it can also compete or cooperate with the binding of specific microRNAs to their targets (Kim et al., 2009 and Kundu et al., 2012). After validating HuR as a miR-133 target and a linc-MD1-regulated gene (figure 6), we started investigating a possible physical interaction between the two.

Immunoprecipitation of HuR and RNA pull-down of linc-MD1 confirmed the existence of such interaction, opening new fascinating scenarios (figure 7).

We already knew the effect of linc-MD1 knock-down on HuR (as for other miR-133 targets) and we knew that their expression was tightly correlated during myoblast differentiation (figure 8A), but we missed information about HuR impact on linc-MD1. SiRNA-mediated knock-down of HuR caused strong down-regulation of linc-MD1 (figure 8B and C), compatibly with an effect on biogenesis or on stability. We excluded the second hypothesis by RNA turnover analysis (performed with actinomycin D treatment followed by qRT-PCR) and found that HuR down-regulation caused Drosha to cleave more efficiently the miR-133b precursor sequence (figure 8C), as observed with the measurement of 5'cut-off molecules and miR-133b after knock-down. Moreover, luciferase analysis revealed that HuR is able to cooperate with miR-135 binding to the linc-MD1 sequence (not shown). In order to elaborate a meaningful model of the linc-MD1/HuR feed-forward regulatory loop and unravel the apparent paradox of the linc-MD1 locus producing miR-133b and its decoy alternatively, a couple of facts need to be mentioned. First, miR-133b is functionally identical to its paralogues miR-133a-1 and miR-133a-2, coming from two unrelated loci and having much more Drosha-compatible pri-miR structures. Second, the timing of expression of all those actors during myogenesis: while HuR and linc-MD1 are strongly but transiently up-regulated at the beginning of differentiation, the whole miR-133 pool is accumulated more lately, in mature myotubes (figure 8A). Therefore, although all those actors are to be considered pro-myogenic, they act in distinct time windows. At the beginning of differentiation, HuR and linc-MD1 are activated (linc-MD1 at the transcriptional level, HuR we don't know) and sustain each other expression at the post-transcriptional level. During this window of time lasting 48-72 hours, miR-133 activity is blocked by linc-MD1. Later on during differentiation, miR-133 is accumulated, mainly from the miR-

133a loci and its concentration increases to a level that cannot be buffered by linc-MD1 anymore, resulting in HuR down-regulation and consequently on linc-MD1 biogenesis shutdown. At this stage, miR-133 is able to repress its targets, with the important consequence that such repression follows an on/off behaviour, rather than a continuous ascent, thanks to linc-MD1 activity in the previous phase (figure 9). Globally, those observations suggest that microRNA decoys can sharpen the temporal (and possibly spatial) frame of microRNA activity and can represent powerful switches to trigger the entry or exit of proteins and RNAs to or from feedback loops.

**Figure 6**

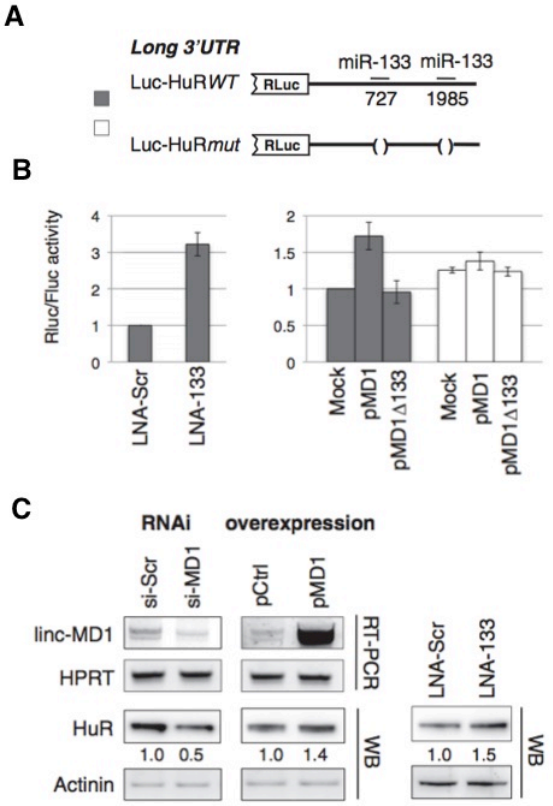



*Figure 6. Linc-MD1 regulates HuR expression. A. Schematic representation of the contructs used below (HuR 3'UTR cloned downstream to the luciferase ORF). B. Luciferase assay of the aforementioned constructs with miR-133 inhibition mediated by LNAs and with linc-MD1 overexpression in the WT and mutated form (by deleting miR-133 binding sites). C. RT-PCR and Western blot analyses showing linc-MD1 and HuR expression upon knock-down and overexpression of linc-MD1 and upon miR-133 inhibition by LNAs. Adapted from Legnini et al., 2014, reproduced under the license N° 3785820559384.*

**Figure 7**

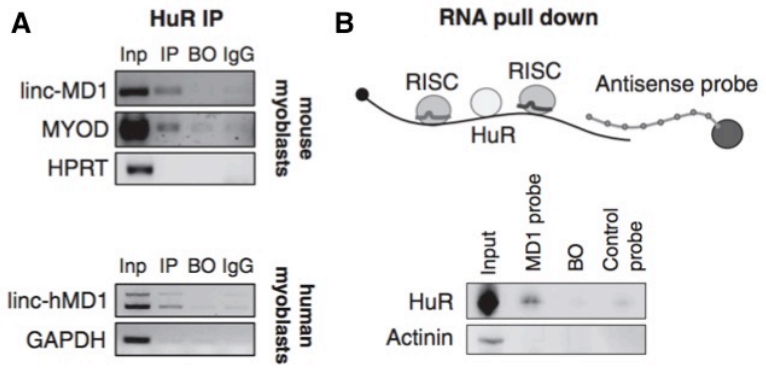



*Figure 7. linc-MD1 is bound by HuR. A. HuR immunoprecipitation in mouse (upper panel) and human (lower panel myoblasts). RT-PCR for linc-MD1 is shown together with a positive control (MYOD) and a negative one (HPRT). B. RNA pulldown for linc-MD1 obtained wth antisense biotinylated probes. Western blot analysis for HuR and the negative control Actinin is shown. Adapted from Legnini et al., 2014, reproduced under the license N° 3785820559384.*

## Figure 8

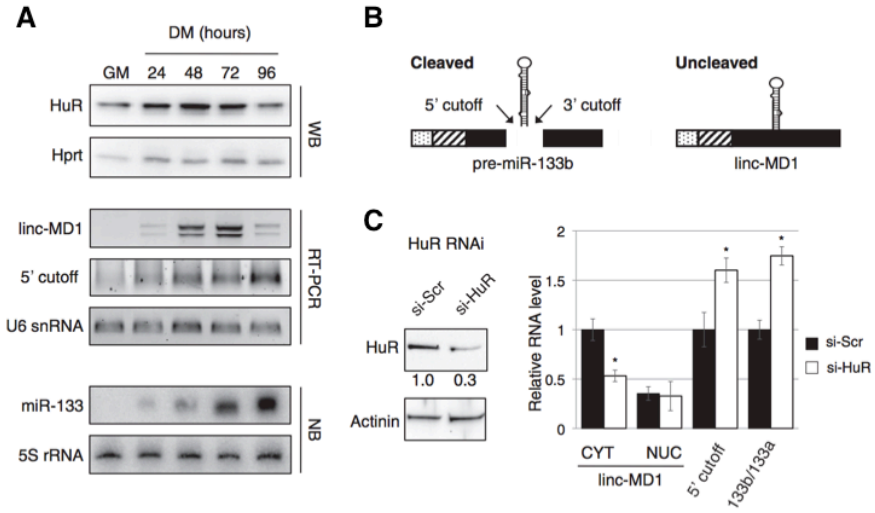



*Figure 8. linc-MD1 and HuR feedback regulation. A. HuR, linc-MD1 plus its 5' cut-off product and miR-133 expression levels during in vitro myoblasts differentiation measured by Western Blot, RT-PCR and Northern blot respectively (Hprt, U6 snRNA and 5S rRNA are used as loading controls). B. Schematic representation of the alternative fates of linc-MD1: it can be cleaved by Drosha, producing two cut-pff products and a pre-miR-133b (left) or it can be exported to the cytoplasm as linc-MD1 (right). C. Western blot of HuR after HuR knock-down (Actinin as loading control, left) and RT-PCR for linc-MD1 in nuclear and cytoplasmic fractions, for its 5' cut-off product and for miR-133b after HuR knock-down. Adapted from Legnini et al., 2014, reproduced under the license N° 3785820559384.*

## Figure 9

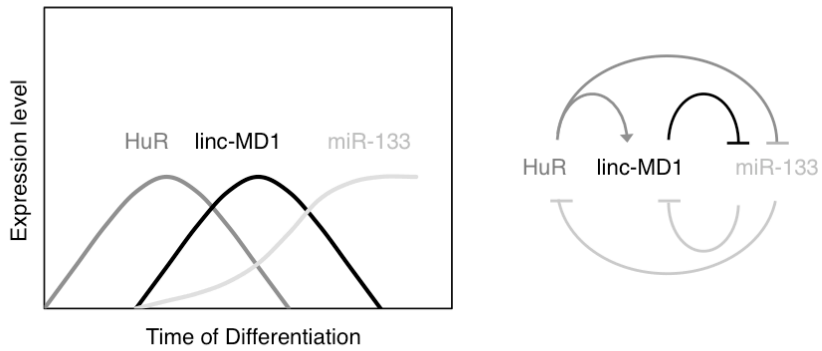



*Figure 9. Possible model for HuR and linc-MD1 activity during myoblast differentiation. HuR is activated after induction of differentiation and it sustains linc-MD1 synthesis, which in turn prevents HuR to be degraded by miR-133. MiR-133 accumulates and once it exceeds a certain threshold it triggers both HuR and linc-MD1 degradation. This provides a temporal window for miR-133 activity, which is restricted to terminal differentiation.*

## 2.4. Contradictions and possibilities arise from quantitative modelling of competition for microRNA binding.

Besides our observations, several papers contributed to the formulation of the so-called *ceRNA hypothesis* (Tay et al., 2011, Karreth et al., 2011 and many others). According to such model, *competing endogenous RNAs*, or *ceRNAs*, are transcripts that share multiple microRNAs binding sites that represent a sort of language through which ceRNAs can cross-talk one to each other. This means that a perturbation in a given ceRNA (e.g. transcriptional activation) is transferred to any other ceRNA that shares sufficient microRNA target sites, because of a change in the concentration of free-to-bind microRNAs. This hypothesis has been strongly criticised on the basis of quantitative assessments indicating that physiological concentrations of mRNAs and microRNAs within living cells are not compatible with ceRNA-like effects (Denzler et al., 2014). Specifically, titration of target abundance of miR-122 in hepatocytes and liver by overexpression produced a release of other targets repression in a threshold-like manner only after adding more than $1.5 \times 10^5$ additional target molecules per cell, a value that easily exceeds the quantity of any endogenous transcript.

A few theoretical analyses tried to reconcile the experimental evidence of ceRNA-like effects with quantitative models describing the dynamics and steady state of a system comprised of $N$ ceRNA molecules in competition for binding to a given microRNA (Figliuzzi et al., 2013 and 2014, Ala et al., 2013, Bosia et al., 2013 and Jens et al., 2015). A theoretical approach is necessary for defining the limits within a ceRNA effect is possible regardless the contingent experimental conditions that characterize laboratory manipulations (such as the concentration of a given microRNA and its targets). Therefore, it is possible to investigate the dependency on parameters of a simplified model, which may determine the ideal conditions required for observing ceRNA-like effects. A couple of studies (Figliuzzi et al., 2013 and 2014 and Bosia et al., 2013) faced the problem in depth and reported

convergent conclusions. First of all, it is predicted that competition can happen in a threshold-like manner, around a concentration of bound ceRNA approximately equal to half of the theoretical maximum, a regime that allows fast changes of ceRNA-microRNA complexes concentration in response to relatively small changes in free microRNA concentration. Second, such effect is observed at steady state only when introducing a stoichiometric component of the decay of ceRNA-microRNA complexes. This means that the rate of ceRNA-microRNA complex decay is nonzero or, in other words, that when a microRNA binds to its target they can be degraded together. It is not easy to predict a biological meaning for such parameter, but not impossible: one may suggest that a reliable ceRNA effect can be observed after sequestration of microRNA-target complexes in P-bodies or other stable foci able to bring the microRNA irreversibly out of solution. On the other hand, target-mediated degradation of microRNAs has been also observed, but only in specific and limited cases (Ameres et al., 2010 and Cazalla et al., 2010). Another important aspect emerging from those quantitative studies is that strong and specific ceRNA effects can be detected out of equilibrium. This means that after inducing a perturbation in the system, e.g. a change in the transcription rate of a ceRNA, an effect on the levels of other ceRNAs is observed also in the absence of stoichiometric decay but only before a full steady state relaxation. It therefore seems that a fully competitive system cannot account for ceRNA cross-talk, which instead can take place in a subset of specific conditions (stoichiometric decay, non equilibrium). The reasons why such conditions are not likely to exist in average biological systems are fully described in Jens et al. (2015), where the authors used a simpler but elegant model for predicting the impact of a perturbation in a microRNA target on the occupancy of the rest of its binding sites. Such model is based on the assumption that steady state is a good approximation of biologically meaningful scenarios and it results in the observation that ceRNA-like effects are achievable only when the amount of target sites added to the system is overwhelmingly high compared

to that of average mRNAs (roughly equal to the sum of all endogenous targets of the microRNA). It has to be said that both this study and the previous ones rely on parameters that are either estimated or measured in vitro. As an example, the number of total binding sites of a given microRNA (on which ceRNA-like effects strongly depend) can be predicted computationally or estimated through CLIP-based experiments, which anyway provide qualitative information on a bulk and not on single cells. On the other hand, the impact of dynamic changes in gene expression might be more important than thought, since many differentiation processes (e.g. myogenesis and linc-MD1 activation) proceed through fast and robust changes in the transcriptome of a cell population. In a similar scenario, a non-equilibrium effect might play a critical role that could be even resilient to approaching the steady state if feedback loops are taken into account (e.g. a sudden change in the concentration of a transcription factor may have strong and stable effects, regardless the transient nature of the up-regulation of its mRNA).

Regarding linc-MD1, I performed quantitative assessments of its concentration and of miR-133 and miR-135 concentrations and found interesting results (unpublished data). When inducing differentiation of $C_2C_{12}$ myoblasts, linc-MD1 abundance turns from roughly 0 to $10^3$ molecules per cell, while miR-135 (linc-MD1 has two binding sites for it) is present in circa $1.5 \times 10^3$ molecules per cell and miR-133 in $5 \times 10^3$ molecules per cell. Such quantities, although do not fit with a simple ceRNA-like effect, are indicative of a possible decoy mechanism in particular conditions (e.g. transient effect or local effect), as suggested by experiments in Cesana et al. (2011) and Legnini et al. (2014).

Finally, an important feature of the ceRNA hypothesis (that anyway does not apply to linc-MD1 and other cases) still waits to be tested: its dependency on the number and diversity of microRNA binding sites. Even in a steady state, non-stoichiometric scenario, an additive behaviour of several binding events of different microRNAs may contribute to detectable ceRNA effects.

To what extent such behaviour is possible depends on several factors, including the intrinsic nature of the microRNA degradation pathway, but also just simple quantitative parameters. Anyway, to study such possibility is beyond the aim of this work.

In the end, it can be concluded that quantitative models do not support a generalized ceRNA theory, although they allow ceRNA effects in certain conditions. Nevertheless, some aspects of such models have to be faced more cautiously, and the biological relevance of transient effects, local concentrations and physical constraints to molecular diffusion should be more carefully taken into account when interpreting data on microRNA-target interactions. Conversely, when attributing an observed change in a given mRNA to a ceRNA effect, this has to be validated cautiously with quantitative approaches, and the dependency of such effect purely on microRNA binding has to be studied in detail.

# 3. Circular RNAs are a new, large class of potentially functional molecules in Eukaryotes.

## 3.1. High throughput total RNA sequencing and computational methods revealed an entire new class of RNA.

In search of new potential actors in spatio-temporal regulation of myogenesis, we focused on a promising new class of RNAs: circular RNAs (circRNAs). In Eukaryotes, RNA molecules are normally linear polynucleotides which are necessary for the completion of the three main phases of the genetic information flow: bringing the coding information into the cytoplasm (mRNAs), converting it into proteins (rRNAs, tRNAs), and regulating the former two processes in time and space (miRNAs, lncRNAs etc.). Protein-coding genes and class II genes in general are typically transcribed into multi-exonic RNA species with 5' and

3' termini protected by a 5-methyl-guanosyne cap and a poly-adenosine tail respectively. Although these two features provide mRNAs certain stability, coding transcripts typically have short half lives compared to other RNAs (e.g. ribosomal RNAs) or different biomolecules (e.g. proteins), and this is thought to be a key feature for achieving fast and robust regulation of gene expression.

Recently, the availability of deep cDNA sequencing data obtained from ribosomal-depleted total RNA samples, combined with new computational approaches for aligning spliced sequencing reads, allowed the identification of several splicing products that were thought to be mistakes or unusual events before. Among them, an entire new class of RNA molecules lacking 5' and 3' termini have been discovered (Salzman et al., 2012, Jeck et al., 2013 and Memczak et al., 2013). According to such reports, the splicing apparatus can join a 5' splice site with the 3' splice site of an upstream intron, thereby producing a topologically closed RNA molecule, or circular RNA (figure 10). CircRNAs had been previously observed in a variety of model systems (Sanger et al., 1976, Grabowski et al., 1981 and Danan et al., 2012), but the fact that these molecules are commonly produced by thousands of eukaryotic genes was ignored before the aforementioned studies. CircRNAs are capable of unprecedented stability due to the lack of 5' and 3' ends, which are the substrates of the common system for RNA decay in Eukaryotes, exonucleases. Although this biochemical property, combined with other common RNA features, suggests a plethora of possible biological functions, it is still not clear whether they are capable of impacting biological processes. On the other hand, many efforts have been spent on dissecting the mechanism(s) of circRNA biosynthesis, showing that these molecules are broadly produced by the spliceosome in some specific contexts.

**Figure 10**

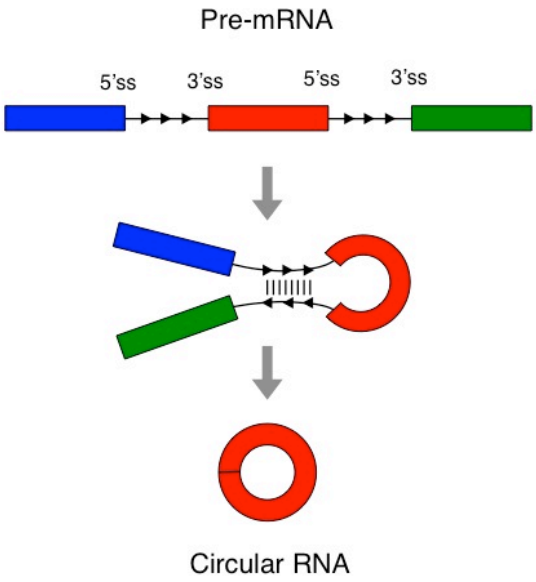



*Figure 10. Schematic representation of circular RNA production. A pre-mRNA comprised of N exons (here in blue, red and green) can produce a circular RNA from an internal exon (red) as consequence of a non-canonical splicing event, often driven by base pairing of the flanking introns.*

### 3.2. Synthesis of circular RNAs is not (entirely) stochastic.

In the aforementioned studies, thousands of circular RNAs have been detected from sequencing reads and their expression levels seem to shape in a random distribution with an impressive number of circRNAs with very low reads, in addition to an expected tail on the right of the curve representing a few, very highly expressed molecules. Although this might suggest a stochastic model for circRNA production, circRNA expression profiles form different cell types, tissues and organisms, in addition to the active circularization mechanisms that will be discussed in the next paragraph, suggest that among a noisy background of randomly produced circRNAs, a population of evolutionary conserved and possibly functional molecules exists.

A first hint comes from the observation that the major product of hundreds of genes that can produce alternatively a messenger RNA or a circular RNA is the second one (Salzman et al., 2012). Therefore, circularization might be a way of getting rid of a functional transcript in certain conditions, or also a way of producing a molecule that has a biological function *per se*. The idea of functional circular RNAs was confirmed for the first time by two studies (Memczak et al., 2013 and Hansen et al., 2013) reporting that the circRNA CDR1as (or CiRS-7) is able to function as a sponge for miR-7 in brain. This circular RNA cannot be targeted by Ago2-dependent decay because of the lack of 5' and 3' termini, therefore is a perfect microRNA decoy, being able to bind to miR-7 with its > 70 sites and block this microRNA until it gets degraded in other ways. This mechanism of action seems to be restricted to CDR1as and perhaps a few other circRNAs (Hansen et al., 2013, Guo et al., 2014), indicating that if other members of this class of molecules exert any function, this must be a different one. A remarkable observation comes from several reports describing the relative abundance of circRNAs in different tissues or cell states. Especially for neuronal differentiation and brain development, it has been shown that circRNA production can be

dramatically regulated (You et al., 2015, Rybak-Wolf et al., 2015 and Venø et al., 2015). Circular RNA sub-cellular localization is also regulated, being known that they can be transported into the nucleus and that they can be extremely enriched at synapses (Hansen et al., 2011 and You et al., 2015). Dynamic expression, regulation, relative abundance, sequence and positional conservation (Memczak et al., 2013 and Rybak-Wolf et al., 2015) are important features that make us speculate that a subset of circRNAs may be indeed functional molecules, but, given the exception of few microRNA sponges, this has still to be proven.

### 3.3. RNA circularization in Eukaryotes proceeds via a back-splicing reaction.

Circular RNAs made up of covalently closed exonic sequences in Eukaryotes were discovered more than 20 years ago (Nigro et al. 1991, Capel et al. 1993; Cocquerelle et al. 1993 and Zaphiropoulos 1997), and their synthesis was ascribed to the splicing machinery, that could join the splicing donor of an intron to the acceptor of another intron upstream in the pre-mRNA, a phenomenon that was considered a mistake by many (*mis-splicing, exon shuffling, exon scrambling*).

When those molecules were recently re-discovered, the first question that people in the field tried to address was about their biosynthesis. The first information we had was that exon circularization or back-splicing was more likely to happen in the presence of very long introns and was associated to the occurrence of *Alu* repeats in the flanking intronic sequences (Jeck et al., 2013). Such observation was confirmed and extended by others: while human cells have Alus, in mice and other species circular RNA production is associated to repetitive elements with a different origin (Liang et al., 2014, Zhang et al., 2014 and Ivanov et al., 2015). The use of minigenes carrying exons that are naturally circularized with different portions of their flanking introns revealed that most of the information needed for back-splicing is

contained in introns (Ashwal-Fluss et al., 2014). Therefore, circularization occurs via base-pairing of short sequences embedded in intronic regions, thus bringing in physical proximity the donor and acceptor sites that will be joined to form a circRNA. This is confirmed by the high efficiency of circRNA overexpression obtained with plasmids where the upstream intronic region of a given circRNA is inverted and cloned again downstream to it (Hansen et al., 2013). Also A-to-I RNA editing can play a role in the process, as ADAR1 was shown to inhibit circularization by blocking base pairing of Alu elements through its catalytic activity (Ivanov et al., 2015). The whole process is thought to occur co-transcriptionally, as it was observed that circular RNAs are present in chromatin-bound nascent RNA in flies (Ashwal-Fluss et al., 2014). On the other hand, in human cells, circularization can be linked to 3' end formation, as the presence of any kind of terminator is required for efficient back-splicing (Liang et al., 2014 and Kramer et al., 2015). This seems to be dependent on the length of the flanking repetitive elements: when a couple of short inverted repeats is substituted with longer ones, indeed circularization happens also without a correct 3' end formation in human cells (Kramer et al., 2015). This observation suggests that circularization can occur with different dynamics: strong base pairing can induce fast and efficient back-splicing, while circRNAs with a less affinity of the flanking introns may be produced through a slower reaction, possibly favoured by additional factors. This is also strongly suggested by the observation that back-splicing happens also in S. Pombe, where repetitive elements are absent and the back-splicing reaction proceeds via an intermediate made of an exon-containing lariat (Barrett et al., 2015). Being known that the canonical splicing machinery is responsible for back-splicing (Starke et al., 2015), additional elements that may contribute to circularization are obviously RNA binding proteins (RBPs). The first one that has been identified is *Muscleblind*, that it is able to control its own concentration by inducing circularization of the second exon in its

pre-mRNA (and also other pre-mRNAs, Ashwal-Fluss et al., 2014). More recently, an RNAi-based sceening of RNA binding proteins modulated during epithelial-to-mesenchimal transition identified *Quaking* as the responsible factor of the global increase of circRNA production during such process (Conn et al., 2015). In both cases the direct binding of the RBP is required for circularization, suggesting the idea that this binding may facilitate base pairing of the flanking introns or bring the splice sites in close proximity as result of protein oligomerization. In general, RBPs, SR proteins and hnRNPs seem to play a role in circularization (Kramer et al., 2015), again indicating that such process may be not simply a splicing mistake, but a way of regulating gene expression or even producing functional molecules.

# Aim of the thesis

While concluding my work on linc-MD1 and post-transcriptional control of myogenesis, the field of circular RNAs exploded with unprecedented and fascinating findings. Although it seemed clear that something important resided behind the broad production of those new molecules, most of the focus was directed towards the understanding of their synthesis, and not their function. This was mostly due to the fact that circular RNAs are a family of biochemically-related molecules and apparently not a functional class. Therefore, understanding their role in biological phenomena, if there's any, is extremely challenging. One finding came with the characterization of CDR1as/CiRS-7 as a microRNA decoy, but it was clear at the time that such mechanism of action was an exception and not the rule. Although the biochemical features of circular RNAs, especially their exceptional stability, suggest a myriad of captivating biological functions, we still don't have a clue about their evolutionary meaning. The study of muscle differentiation has been a powerful source of paradigms in molecular biology, especially in the field of non-coding RNA. The molecular mechanisms that trigger cell differentiation are clear and the ability to reproduce such process in vitro with cells coming from different species is a potent tool for studying gene functions from an evolutionary point of view. Moreover, the transition from proliferating myoblasts to differentiated myotubes through a series of distinct and recognizable stages enables to infer possible functions from gene expression patterns (e.g. a gene activated during terminal differentiation is more likely to play a role in mature myotubes rather than in proliferating cells). We therefore decided to approach the circular RNA field and try to understand if these molecules are capable of biological function in myogenesis. The idea was to analyse circRNA expression dynamics in both

human and murine myoblasts differentiation, then select a subset of highly expressed, regulated and conserved molecules, knock them down and screen for phenotypes associated to their down-regulation. After identifying circRNA-associated phenotypes, we wanted to understand possible general mechanisms of action, focusing on one main candidate that will be described in the following section. All the steps that I just mentioned were almost completely unexplored at the time this work was started, we therefore had to invent and set up most of the computational and experimental procedures that will be described in the next chapters.

# Results and Discussion

**Note:** some of the results and figures included in this section are part of a manuscript submitted for publication. In case of acceptance, all the publication rights will be ceased to the journal publisher, regardless the timing of publication of this thesis.

## 1. CircRNAs expression dynamics in myogenesis.

### 1.1. CircRNAs are abundant, conserved and highly expressed.

With the aim of identifying circular RNAs expressed during muscle differentiation, we collected total RNA from 2 biological replicates of human (9208) and mouse ($C_2C_{12}$) myoblasts cultured in growth medium (GM) or induced to differentiate into myotubes (DM). We performed paired-end ribominus RNA sequencing and adapted the FindCirc computational pipeline (Memczak et al., 2013) in order to detect back-splicing events. We first processed RNAseq data for gene expression analysis: we mapped reads with TopHat and reconstituted the transcriptome with Cufflinks (data summarizes in table 1). Differential gene expression analysis was performed in order to check the quality of the in vitro differentiation experiments. While FPKM of genes in technical replicates were almost perfectly correlated (> 98% Pearson correlation, figure 11, upper panels), gene expression in myotubes compared to myoblasts was highly diverse with > 3000 genes significantly up-regulated or down-regulated in one of the two conditions, in both human and mouse systems (figure 11, lower panels). When looking at the enrichment of gene products up-regulated in myotubes versus myoblasts, in terms of gene ontology

(GO) biological process keywords, we found a consistent and significant enrichment for genes related to muscle differentiation and function (figure 12). Moreover, we assessed the expression of well-known specific markers of muscle differentiation (MyoD, Myosin Heavy Chain, Muscle Creatine Kinase and Dystrophin) by qRT-PCR, which confirmed the data obtained by RNAseq (figure 13). At this point we focused on circular RNAs, which can be identified through reads spanning the so-called *head-to-tail* splice junctions resulting from back-splicing events. RNAseq reads were re-aligned to the reference genomes, the ones contiguously aligned were discarded and the remaining reads were used as input for FindCirc (figure 14A). Among the thousands of such events that we identified, we applied some filters (> 2 unique reads spanning the splice junction, high quality mapping and low distance between donor and acceptor sites) for discarding possible artefacts. As described in figure 14B, thousands of *bona fide* circular splicing events were found in both human and mouse samples, many of which expressed at low levels, with a tail on the right of the distribution constituted by a fraction of a few highly expressed molecules (figure 15). Almost 90% of them derive from internal exons of protein coding genes, while a few are expressed from non-coding genes or intergenic regions, possibly due to their average lower expression level (figure 14C). Almost 10% of circRNAs were expressed at similar or higher levels with respect to the linear counterpart (by comparing the number of reads mapping on circular versus linear splice junctions). Interestingly, we found more than 500 positionally conserved circRNAs between human and mouse (figure 14D). Our criterion for conservation required the genomic location in human to overlap with the synthenic region in mouse. Considering that low abundance species might have been missed by this analysis, the actual number of conserved circRNAs may be even higher, as also indicated by the observation of a similar overlap between two replicates of the same sample.

## Table 1

| 9808 (Human) Genes | | |
|---|---|---|
| | GM | DM |
| FPKM>1 | 15004 | 15488 |
| FPKM>10 | 6032 | 6009 |
| FPKM>100 | 803 | 732 |
| Diff. Expr. | 3686 | 3303 |
| FC>10 | 258 | 333 |

| C2C12 (Mouse) Genes | | |
|---|---|---|
| | GM | DM |
| FPKM>1 | 12682 | 13019 |
| FPKM>10 | 6244 | 6365 |
| FPKM>100 | 907 | 879 |
| Diff. Expr. | 3848 | 3975 |
| FC>10 | 359 | 404 |

*Table1. Summary of gene expression analysis performed with TopHat/ Cufflinks/ Cuffdiff. Genes expressed with more than 1, 10 and 100 FPKM are shown, together with the number of differentially expressed in GM versus DM and among them the ones with a fold change > 10.*

**Figure 11**

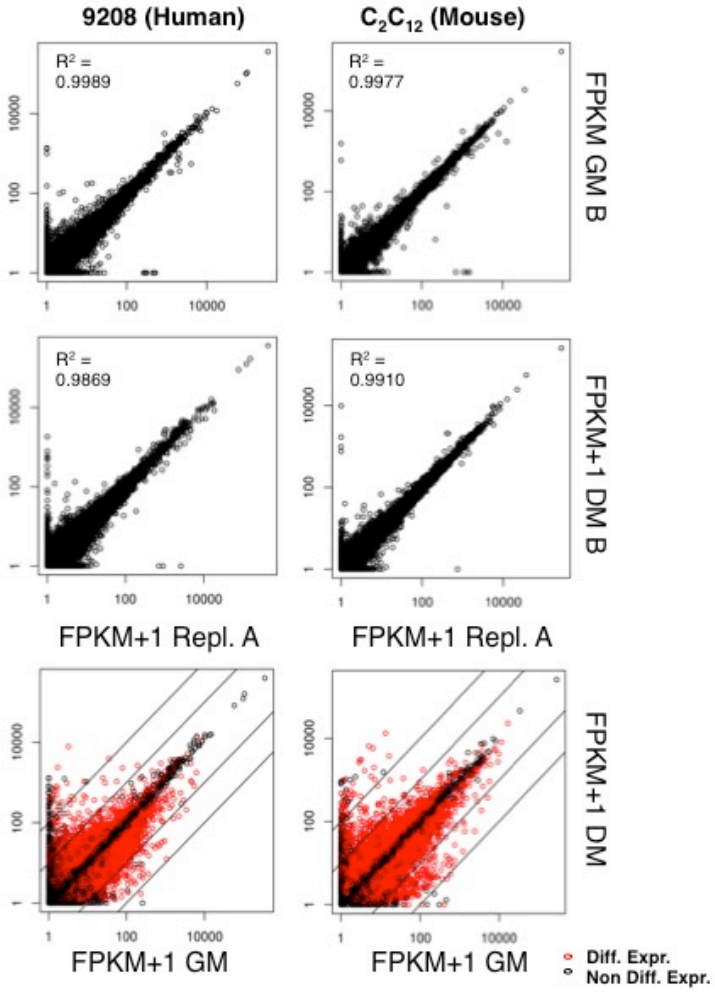



*Figure 11. Scatter plots for biological replicates of FPKM of genes in human (left) and mouse (right) for GM (up) and DM (middle) samples. On the bottom are shown scatterplots for mean FPKM in GM versus DM. In red are highlighted genes differentially expressed according to Cuffdiff. Straight lines delimiting genes with a fold change > 10 and > 100 are drawn in black.*

## Figure 12


*Figure 12. Wordcloud of Gene Onthology biology process keywords for genes differentially expressed in DM versus GM for human (left) and mouse (right). The dimensions of the words correlate with their enrichment.*

**Figure 13**

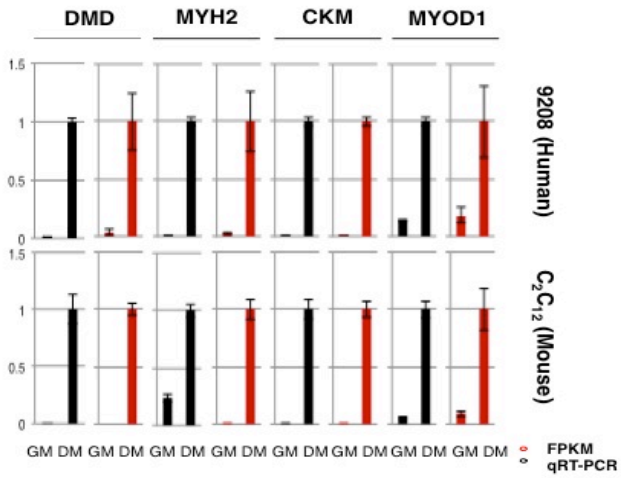
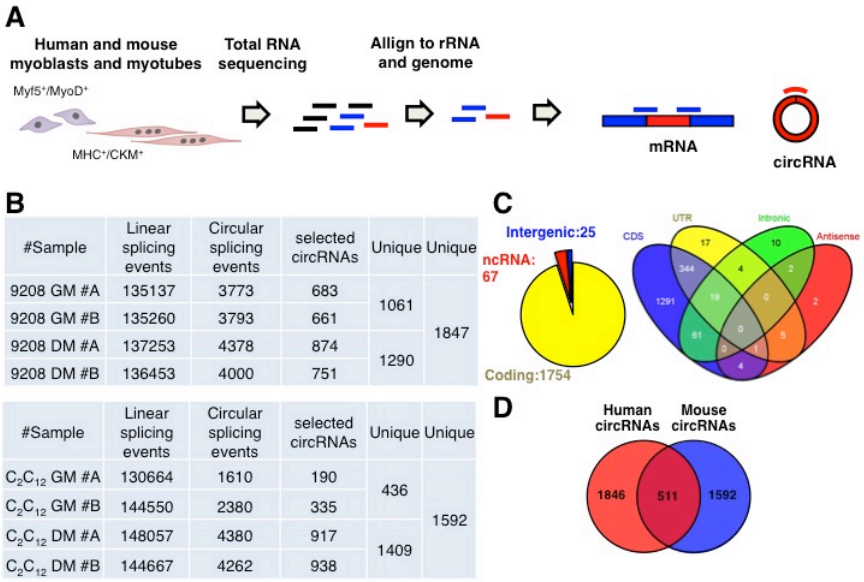



*Figure 13. qRT-PCR validation of known markers of muscle differentiation (dystrophin or DMD, myosin or MYH2, Creatine Kinase or CKM, MyoD) in human (up) and mouse cells (down) in GM and DM. The expression level measured by qRT-PCR (black) is shown together with the FPKM (red).*

## Figure 14




| #Sample | Linear splicing events | Circular splicing events | selected circRNAs | Unique | Unique |
|---|---|---|---|---|---|
| 9208 GM #A | 135137 | 3773 | 683 | 1061 | 1847 |
| 9208 GM #B | 135260 | 3793 | 661 | | |
| 9208 DM #A | 137253 | 4378 | 874 | 1290 | |
| 9208 DM #B | 136453 | 4000 | 751 | | |

| #Sample | Linear splicing events | Circular splicing events | selected circRNAs | Unique | Unique |
|---|---|---|---|---|---|
| $C_2C_{12}$ GM #A | 130664 | 1610 | 190 | 436 | 1592 |
| $C_2C_{12}$ GM #B | 144550 | 2380 | 335 | | |
| $C_2C_{12}$ DM #A | 148057 | 4380 | 917 | 1409 | |
| $C_2C_{12}$ DM #B | 144667 | 4262 | 938 | | |

*Figure 14. Identification of circRNAs in mammalian myoblast differentiation. A. Experimental and computational pipelines for identification of circRNAs from human and mouse myoblasts (expressing MyoD and Myf5) and myotubes (expressing Myosin Heavy Chain, MHC and Creatin Kinase Muscle, CKM). B. Number of detected linear and circular splicing events per sample together with bona fide circRNAs passing selection filters. Samples are human (9208) and mouse ($C_2C_{12}$) myoblasts (GM) and myotubes (DM) in two replicates (#A and #B). C. Left panel: genomic annotation of circRNAs. Right panel: structural annotation of circRNAs mapping to coding regions. D. Overlap of unique circRNAs in human and mouse samples.*

## Figure 15

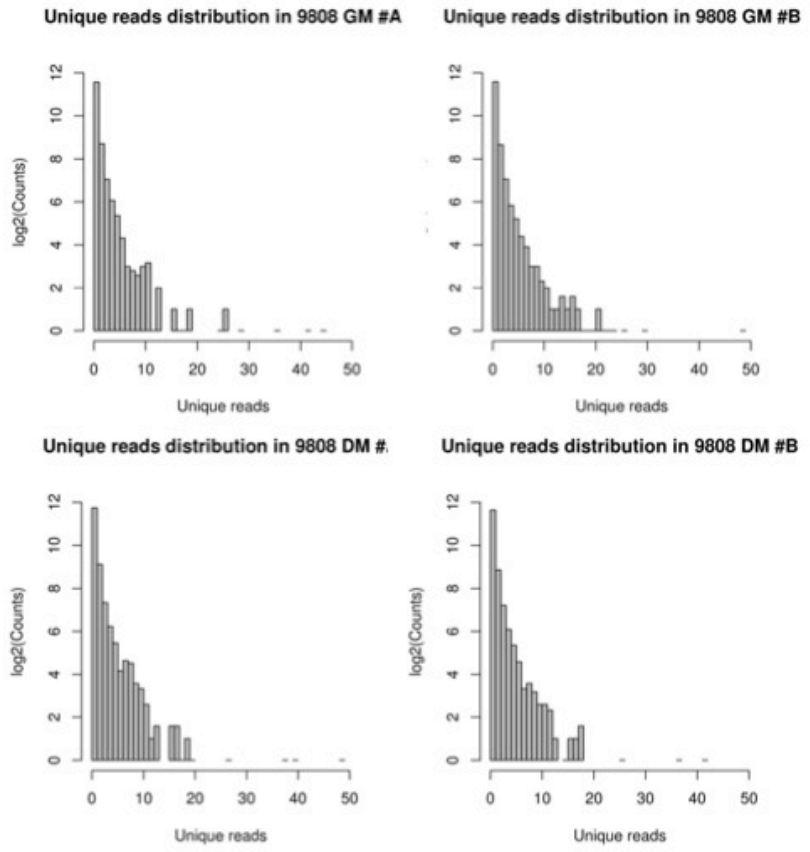



*Figure 15. Histogram showing the number of circRNAs detected in relation with the number of unique reads spanning the head-to-tail junction (only human samples are shown).*

### 1.2. CircRNA expression is modulated in muscle differentiation and disease.

In order to perform quantitative analyses on circRNA expression and regulation, we took human samples and further reduced our list to a set of highly expressed molecules by requiring at least 5 unique reads mapping to the head-to-tail junction in at least one sample. With this threshold, while the technical replicates had an overall good correlation (Fig. 16), the comparison between myoblasts and myotubes revealed a global change in circRNA expression (~45% circRNAs having at least 2-fold variation in one of the two conditions, figure 17), meaning that the expression of these molecules is actively modulated in response to cell differentiation. We also analysed RNAseq data from human myoblasts derived from Duchenne Muscular Dystrophy (DMD) patients and found substantial changes of circRNA expression with respect to normal myoblasts. Hierarchical clustering analysis of normal and dystrophic myoblasts and myotubes revealed that indeed DMD cells have a unique signature in terms of circRNA expression levels (figure 17): specific subsets of transcripts were differently abundant in DMD with respect to controls both in GM and in DM conditions. This is in line with the notion that DMD cultured cells have a slower progression into the differentiation process: they have a higher proportion of non-fusing myoblasts, fusion itself is delayed and differentiation markers are down-regulated (Delaporte et al., 1984, Cacchiarelli et al., 2010 and Cazzella et al., 2012). Notably, circRNA abundance tends in general to increase during differentiation (figure 14), as has been described for neuronal differentiation and in epithelial-to-mesenchimal transition. The circular/linear ratio also followed this trend (figure 18). This is possibly related to the high stability of these molecules: during differentiation, induction of circRNAs at the transcriptional level, combined with a slow turnover, could lead to their accumulation over time. This fact has been also observed in vivo (Westholm et al., 2014) and might represent an important

feature of circRNAs, as they could be involved in ageing-related phenomena or in epigenetic inheritance.

We then tried to address the question of whether modulation of circRNAs expression in myogenesis was due to transcriptional regulation of their host genes or to post-transcriptional events, such as competitive biogenesis between the linear and the circular isoforms within the same gene. Therefore, we analysed the relationship between the fold change ratios of circular versus linear expression level in GM versus DM and found a positive correlation (figure 18C). Additionally, the abundance of circRNAs produced from up-regulated, stable or down-regulated genes, according to FPKMs calculated with Cuffdiff, indicated that the induction of circRNAs coming from up-regulated genes is significantly higher than that of stably expressed genes, which in turn is higher than that of down-regulated genes (Fig. 2F). However, exceptions to this trend were detected, such as the BNC2 gene that in growth conditions produces mainly the linear mRNA, while in differentiated cells the circRNA. In figure 19 the BNC2 locus coverage and circRNA abundance are shown together to those of circ-CDYL, which represents a paradigm of the most common scenario of (possibly) transcriptional activation of a locus followed by accumulation of an embedded circRNA. BNC2 and circ-BNC2 expression pattern is conserved in mouse differentiation and may represent an interesting fact, which will be discussed more in detail in the next section.

## Figure 16

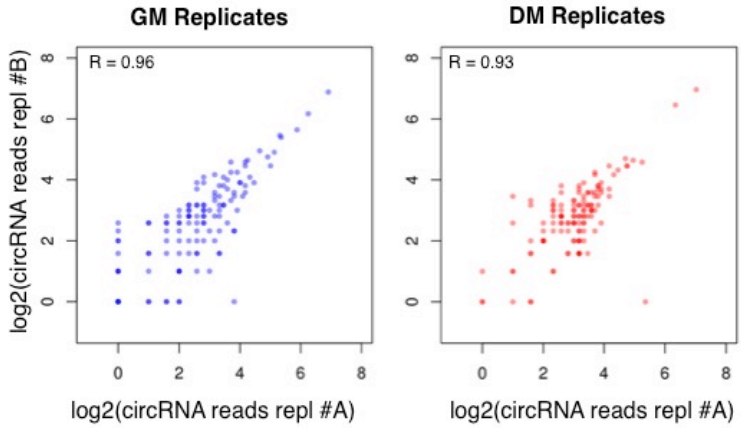



*Figure 16. Correlation of circRNA reads in replicates. Scatterplots showing reads mapping to highly expressed circRNA junctions in two myoblast replicates (GM, #A and #B, left) and myotube replicates (DM, #A and #B, right).*

**Figure 17**

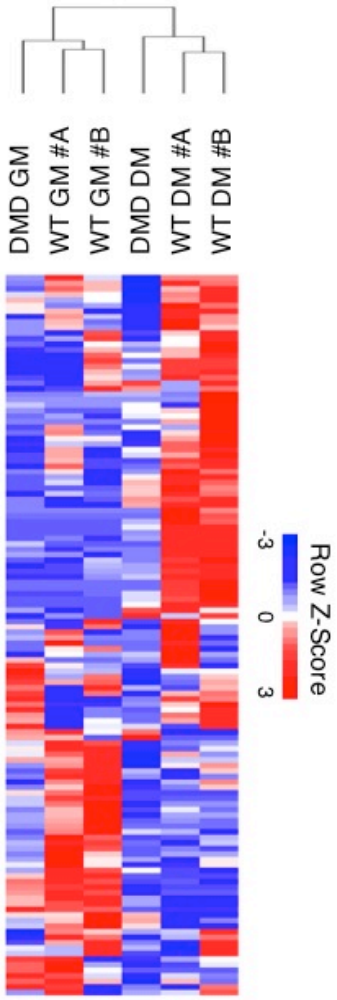



*Figure 17. Heat map showing selected circRNA expression levels in WT and DMD myoblasts (GM, #A and #B) and myotubes (DM, #A and #B). Values are normalized as row Z-scores.*

## Figure 18

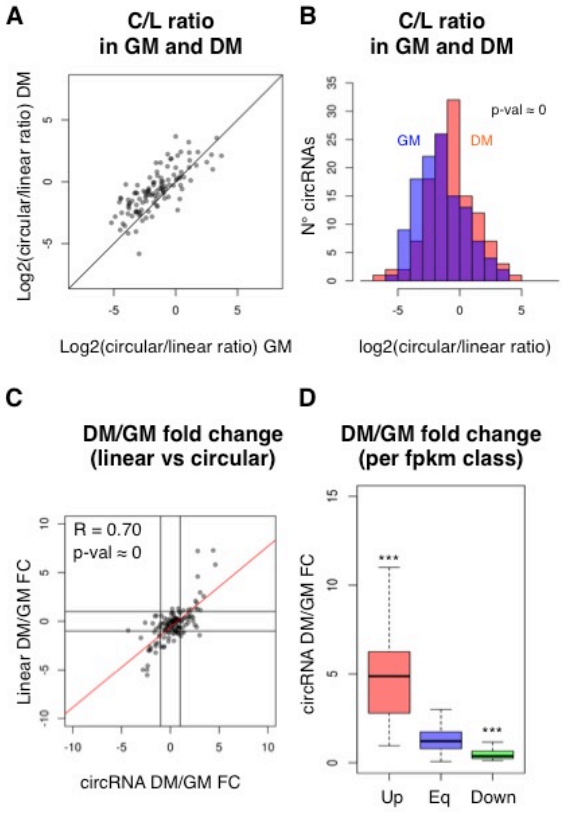



*Figure 18. A. Scatterplot showing circular/linear ("C/L") ratio in GM versus DM, calculated as log2 of the number of reads mapping to a circular junction divided by the mean of reads mapping linearly at the same genomic coordinates. B. Histogram of C/L ratio: in blue distribution in GM, in red in DM. C. Scatterplot of fold change of reads mapping to circRNA junctions versus reads mapping to linear junctions at the same coordinates. Linear regression is shown in red. Black lines correspond to a fold change of 2. D. Bar plot representing fold change of circRNA expression in myotubes versus myoblasts (DM/GM) grouped by differential expression of host gene according to Cuffdiff FPKMs. "Up" are up-regulated genes, "Eq" are non differentially expressed genes, "Down" are down-regulated genes (fold change > 2). Three asterisks indicate a Wilcoxon-Mann-Whitney test-derived p-value below 0.01.*

## Figure 19

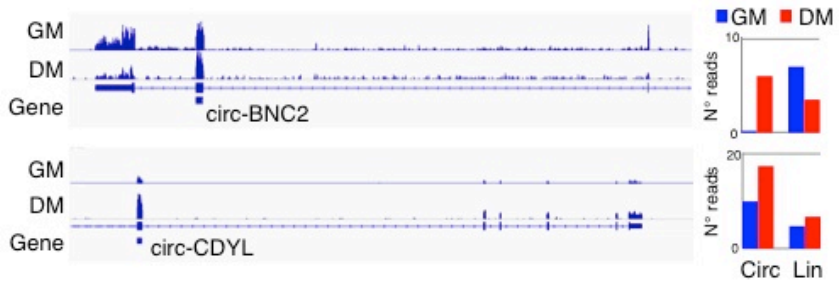



*Figure 19. Left panels: Coverage plot of genomic loci of circ-BNC2 and circ-CDYL. Right panels: bar plot showing reads mapping to circular (Circ) and linear (Lin) splice sites for circ-BNC2 and circ-CDYL in GM (blue bars) and DM (red bars).*

# 2. CircRNAs are functional.

## 2.1. Experimental validation of circular RNAs.

At this point, we wanted to understand if circular RNAs could impact on the process of myoblast differentiation. We therefore selected specific criteria in order to restrict the number of candidates for further characterization: I) conservation, II) expression level, III) differential expression upon differentiation, and IV) circular/linear ratio. This filtering yielded 31 circRNAs, which are shown in table 2. We proceeded with experimental validation of the selected candidates in order to confirm their expression, and we decided to use RT-PCR and RNase R treatment. By using a couple of divergent primers in the PCR reaction, we could obtain an amplification signal only in case of a truly circular RNA. Moreover, the presence of concatemers generated by rolling-circle retro-transcription would confirm such conformational status. We therefore designed a couple of divergent primers for all the 31 candidates, accompanied by a third primer in an external exon for amplifying the corresponding linear mRNA. As summarized in table 2 (full results in figures 20 and 21), out of 31 candidates, only one, circ-MYL4, was not detected as a band of the expected size in human samples. Circ-TTTY16 was instead not detected in mouse samples, because it is located in the Y chromosome that is not present in $C_2C_{12}$ cells. Almost all circRNAs showed, besides the predominant band of the expected size, additional slower migrating bands, probably corresponding to concatemers generated by rolling circle retro-transcription. We cloned and sequenced the amplicons and those concatemers for a subset of circRNAs (figure 22), confirming our hypothesis. As a second important validation test, we used RNase R treatment of total RNA samples. RNase R is an exonuclease able to degrade linear RNA molecules but not circular ones (Suzuki, 2010). Out of 30 putative circRNAs, 29 were confirmed as non-accessible to the

exonuclease. Only one, circ-NEB, was digested by RNaseR as efficiently as its linear counterpart (summarized in table 2, full results in figures 23 and 24). Circ-NEB is predicted to come from the Nebulin gene, which contains dozens of repetitive exons and it is highly expressed in muscle tissues. To less extent, this is true also for MYL4. Mapping of sequencing error-containing reads can easily result in the annotation of false circular RNAs. In our case, RT-PCR and RNase R resistance revealed those false positives, returning a list of 29 *bona fide* conserved circular RNAs.

We then performed nuclear/cytoplasmic fractionation of human and mouse cells in order to obtain a hint on circRNA biogenesis and localization. Results indicated that all circular species were almost exclusively located in the cytoplasm both in mouse and human cells (figures 25 and 26). This was already observed in a few other studies and constitutes an intriguing fact: although the Exon Junction Complex (EJC, which is presumably deposited on circRNAs) is known to play a role in RNA export, it is known that besides specific exportins (e.g. Exp-t and Exp-5 for tRNAs and pri-microRNAs respectively), nuclear export depends on the presence of 5m-guanosyne cap- and polyA-binding proteins. We can imagine a cap-independent mechanism of circular RNA export depending on specific structural motifs, which anyway have not been identified so far, or we also can hypothesize a non-active mechanism of slow accumulation of circRNAs in the cytoplasm as a consequence of mitotic nuclear envelope break down. Such hypothesis could be easily tested in non-dividing cells, were circular RNA synthesis should be followed by nuclear accumulation.

## Table 2

*(continues on next page)*

| Human circRNAs (9808 cells) | | | | | | | |
|---|---|---|---|---|---|---|---|
| **circRNA** | **Chromosome** | **Start** | **End** | **Unique reads** | **RT-PCR** | **Rnase-R** | **Cyt/Nuc** |
| Circ-ACVR2A | chr2 | 148653869 | 148657467 | 0,3,9,6 | yes | yes | cyt |
| Circ-ADAMTS6 | chr5 | 64747301 | 64769779 | 19,18,0,4 | yes | yes | cyt |
| Circ-ANKIB1 | chr7 | 91924202 | 91957214 | 0,3,4,3 | yes | yes | cyt |
| circ-ASH1L | chr1 | 155408117 | 155408859 | 5,3,0,0 | yes | yes | cyt |
| circ-ASPH | chr8 | 62531536 | 62566219 | 3,0,2,0 | yes | yes | cyt |
| circ-BACH1 | chr21 | 30693541 | 30702014 | 6,8,7,3 | yes | yes | cyt |
| circ-BNC2 | chr9 | 16435552 | 16437522 | 0,0,7,5 | yes | yes | cyt |
| circ-CAMSAP1 | chr9 | 138773478 | 138774924 | 10,9,10,13 | yes | yes | cyt |
| circ-CDYL | chr6 | 4891946 | 4892613 | 10,10,17,18 | yes | yes | cyt |
| circ-CRKL | chr22 | 21288066 | 21288532 | 0,2,9,8 | yes | yes | cyt |
| circ-DCBLD2 | chr3 | 98568304 | 98600611 | 17,21,3,3 | yes | yes | cyt |
| circ-HIPK3 | chr11 | 33307958 | 33309057 | 45,30,38,37 | yes | yes | cyt |
| circ-MED13L | chr12 | 116668337 | 116675510 | 6,6,5,10 | yes | yes | cyt |
| circ-MEF2A | chr15 | 100211527 | 100215663 | 0,0,2,0 | yes | yes | cyt |
| circ-MYL4 | chr17 | 45297314 | 45297419 | 0,0,3,7 | no | Na | Na |
| circ-NEB | chr2 | 152544139 | 152550950 | 0,0,17,11 | yes | no | Na |
| circ-PMS1 | chr2 | 190656515 | 190682906 | 3,0,6,3 | yes | yes | cyt |
| circ-PTP4A2 | chr1 | 32381495 | 32385259 | 0,0,7,9 | yes | yes | cyt |
| circ-QKI | chr6 | 163876310 | 163956157 | 0,0,7,7 | yes | yes | cyt |
| circ-RTN4 | chr2 | 55252221 | 55255356 | 0,0,4,5 | yes | yes | cyt |
| circ-RTN4bis | chr2 | 55209650 | 55214834 | 9,10,16,17 | yes | yes | cyt |
| Circ-RUNX1 | chr21 | 36206706 | 36231875 | 0,2,8,5 | yes | yes | cyt |
| circ-SEPT9 | chr17 | 75398140 | 75398785 | 3,0,4,2 | yes | yes | cyt |
| Circ-SLC8A1 | chr2 | 40655612 | 40657444 | 36,22,10,9 | yes | yes | cyt |
| circ-SPECC1 | chr17 | 20107645 | 20109225 | 3,5,11,6 | yes | yes | cyt |
| Circ-TMEFF1 | chr9 | 103261046 | 103279053 | 3,0,5,4 | yes | yes | cyt |
| circ-TTN | chr2 | 179557223 | 179558735 | 0,0,2,3 | yes | yes | cyt |
| circ-ZFHX3 | chr16 | 72984367 | 72994093 | 3,2,2,0 | yes | yes | cyt |
| circ-ZNF292 | chr6 | 87925620 | 87928449 | 5,9,9,4 | yes | yes | cyt |
| circ-ZNF609 | chr15 | 64791491 | 64792365 | 11,4,10,6 | yes | yes | cyt |
| Circ-TTTY16 | chrY | 7565120 | 7567528 | 0,0,2,3 | yes | yes | cyt |

## Table 2

*(continues from previous page)*

| Mouse circRNAs ($C_2C_{12}$ cells) | | | | | | | |
|---|---|---|---|---|---|---|---|
| **circRNA** | **Chromosome** | **Start** | **End** | **Unique reads** | **RT-PCR** | **Rnase-R** | **Cyt/Nuc** |
| circ-Acvr2a | chr2 | 48870288 | 48873473 | 0,0,4,6 | yes | yes | cyt |
| circ-Adamts6 | chr13 | 104294446 | 104314383 | 0,2,2,5 | yes | yes | cyt |
| circ-Ankib1 | chr5 | 3747020 | 3772787 | 0,4,14,13 | yes | yes | cyt |
| circ-Ash1l | chr3 | 89001772 | 89007877 | 0,0,5,3 | yes | yes | cyt |
| circ-Asph | chr4 | 9583811 | 9610938 | 0,0,5,4 | yes | yes | cyt |
| circ-Bach1 | chr16 | 87715330 | 87722608 | 2,0,0,2 | yes | yes | cyt |
| circ-Bnc2 | chr4 | 84291692 | 84293662 | 2,2,12,19 | yes | yes | cyt |
| circ-Camsap1 | chr2 | 25965662 | 25966948 | 0,0,8,7 | yes | yes | cyt |
| circ-Cdyl | chr13 | 35815908 | 35816575 | 7,5,19,25 | yes | yes | cyt |
| circ-Crkl | chr16 | 17468892 | 17469358 | 0,0,1,0 | yes | yes | cyt |
| circ_Dcbld2 | chr16 | 58424560 | 58433463 | 5,10,2,6 | yes | yes | cyt |
| circ-Hipk3 | chr2 | 104470748 | 104471847 | 5,7,43,42 | yes | yes | cyt |
| circ-Med13l | chr5 | 118593332 | 118593570 | 6,4,17,27 | yes | yes | cyt |
| circ-Mef2a | chr7 | 67264864 | 67268400 | 0,0,8,3 | yes | yes | cyt |
| circ-Myl4 | chr11 | 104584022 | 104584127 | 0,0,6,3 | yes | yes | cyt |
| circ-Neb | chr2 | 52299476 | 52305391 | 0,0,39,37 | yes | no | Na |
| circ-Pms1 | chr1 | 53256628 | 53282092 | 0,4,6,8 | yes | yes | cyt |
| circ-Ptp4a2 | chr4 | 129838977 | 129846539 | 2,0,0,0 | yes | yes | cyt |
| circ-Qk | chr17 | 10238892 | 10282981 | 0,0,2,0 | yes | yes | cyt |
| circ-Rtn4 | chr11 | 29605534 | 29608770 | 2,3,6,13 | yes | yes | cyt |
| circ-RTN4bis | chr11 | 29733632 | 29736498 | 0,0,0,2 | yes | yes | cyt |
| circ-Runx1 | chr16 | 92644153 | 92668812 | 0,0,5,6 | yes | yes | cyt |
| circ-Sept9 | chr11 | 29705534 | 29708770 | 2,2,9,9 | yes | yes | cyt |
| circ-Slc8a1 | chr17 | 81647808 | 81649638 | 0,0,12,16 | yes | yes | cyt |
| circ-Specc1 | chr11 | 62117699 | 62119282 | 0,2,0,0 | yes | yes | cyt |
| circ-Tmeff1 | chr4 | 48604580 | 48617331 | 0,2,6,5 | yes | yes | cyt |
| circ-Ttn | chr2 | 76766305 | 76791371 | 0,0,3,7 | yes | yes | cyt |
| circ-Zfhx3 | chr8 | 108792201 | 108800691 | 2,3,4,3 | yes | yes | cyt |
| circ-Zfp292 | chr4 | 34839331 | 34839986 | 0,0,12,14 | yes | yes | cyt |
| circ-Zfp609 | chr9 | 65794621 | 65795495 | 0,0,10,7 | yes | yes | cyt |
| circ-Ttty16 | Na | Na | Na | Na | Na | Na | Na |

*Table 2. List of the selected 31 circular RNAs and summary of their experimental validation. Columns contain: gene name, genomic coordinates (hg19 for human, mm10 for mouse) of the head-to-tail splice junction, unique reads mapped to the splice junction in GM#A, GM#B, DM#A and DM#B, RT-PCR validation (yes if the expected band was obtained, no if not), RNase R resistance (yes if the recovery after RNase R treatment of the circular was higher than for the linear, no if not) and localization (cyt for cyoplasm).*

## Figure 20

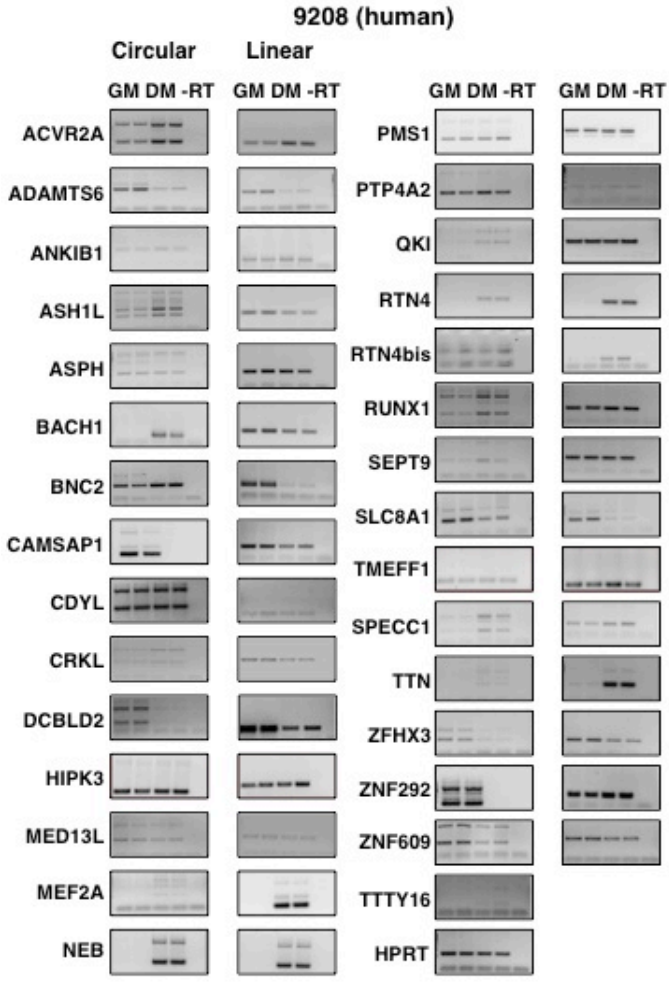



*Figure 20. RT-PCR validation of human circRNAs. CircRNAs and the corresponding linear transcripts were amplified from cDNA obtained from myoblasts (GM) and myotubes (DM) in biological duplicates with divergent and convergent primers*

**Figure 21**

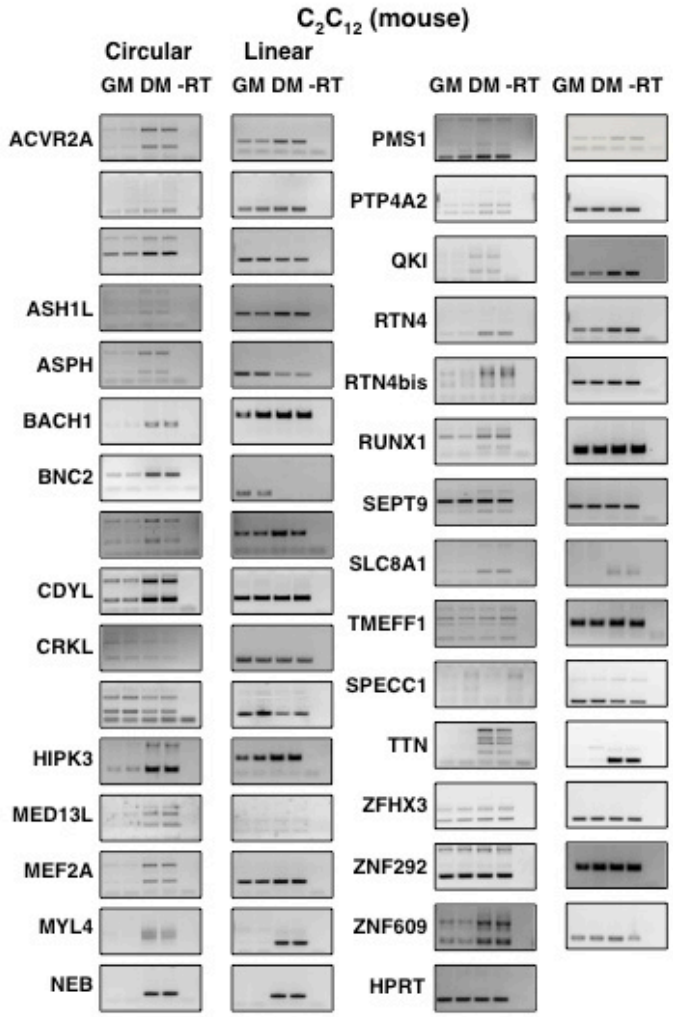



*Figure 21. RT-PCR validation of mouse circRNAs. CircRNAs and the corresponding linear transcripts were amplified from cDNA obtained from myoblasts (GM) and myotubes (DM) in biological duplicates with divergent and convergent primers.*

## Figure 22

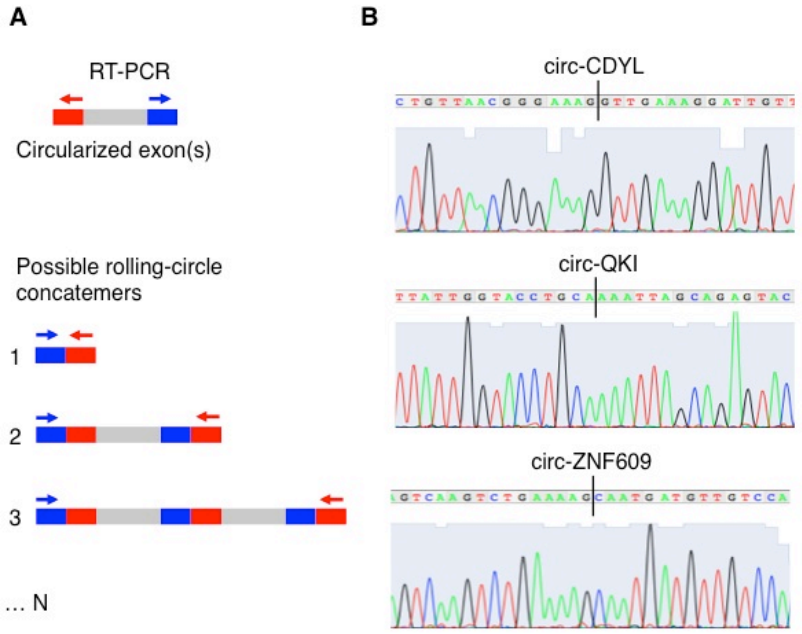



*Figure 22. A. Schematic representation of the amplicons obtained by RT-PCR on circular RNAs. B. Electropherograms showing the Sanger sequencing results of the head-to-tail junction present in the amplicons obtained by RT-PCR for circ-CDYL, QKI and ZNF609.*

**Figure 23**

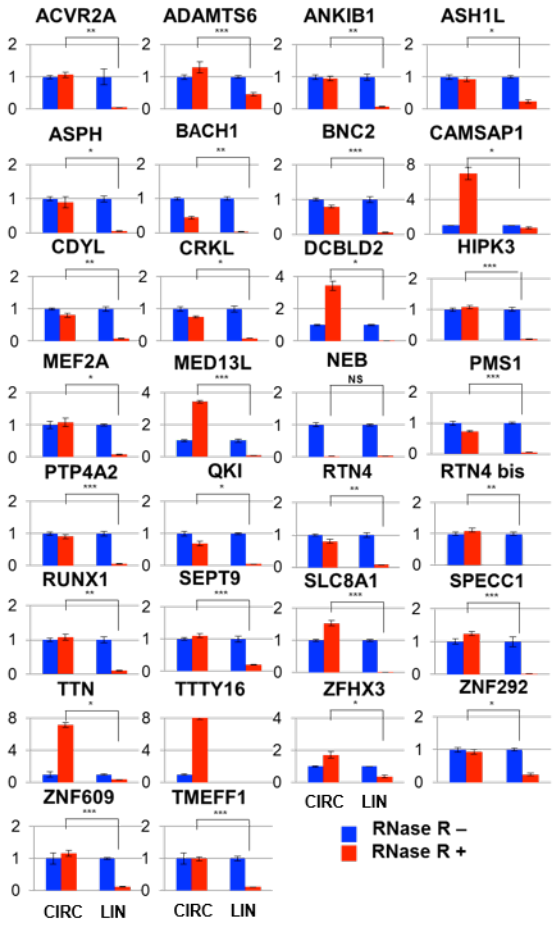



*Figure 23. RNase R validation of human circRNAs. CircRNAs that were amplified by qRT-PCR from cDNA prepared from RNA treated or non treated with RNase R, together with their corresponding linear RNAs. The ratio of treated versus non-treated, of circRNA versus corresponding linear RNAs, was tested by two tailed Student's t test. Red bars: RNase R +; Blue bars: RNase R -.*

## Figure 24

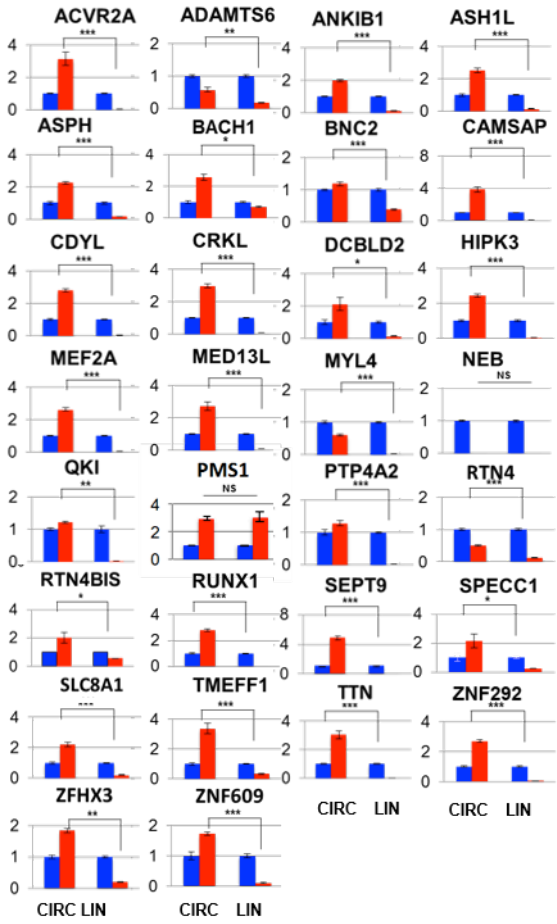



*Figure 24. RNase R validation of mouse circRNAs. CircRNAs that were amplified by qRT-PCR from cDNA prepared from RNA treated or non treated with RNase R, together with their corresponding linear RNAs. The ratio of treated versus non-treated, of circRNA versus corresponding linear RNAs, was tested by two tailed Student's t test. Red bars: RNase R +; Blue bars: RNase R -.*

**Figure 25**

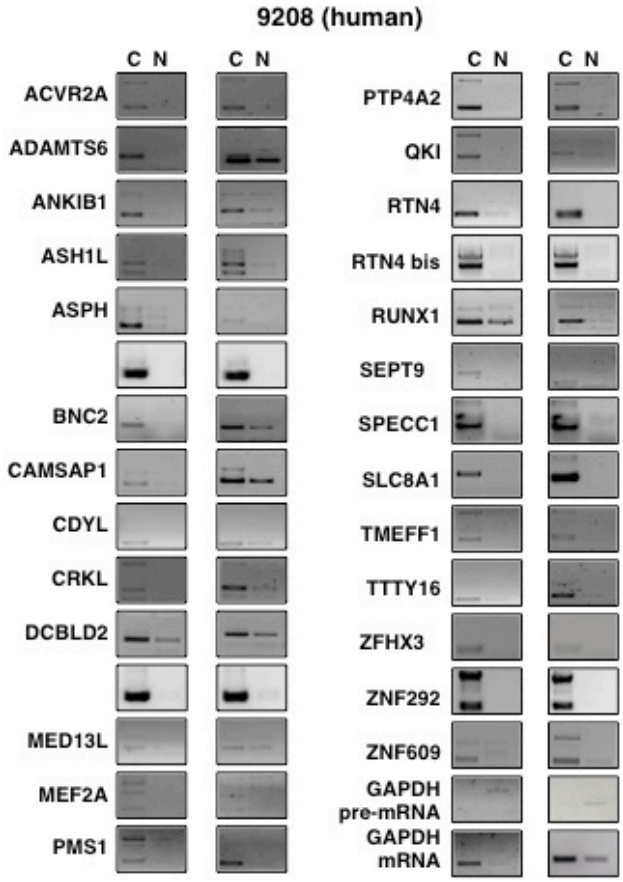



*Figure 25. Human circRNAs that passed the previous validation steps were screened for subcellular localization by RT-PCR amplification from cDNA obtained by retrotranscribing nuclear and cytoplasmic RNA separately. Fractionation was performed in duplicate.*

## Figure 26

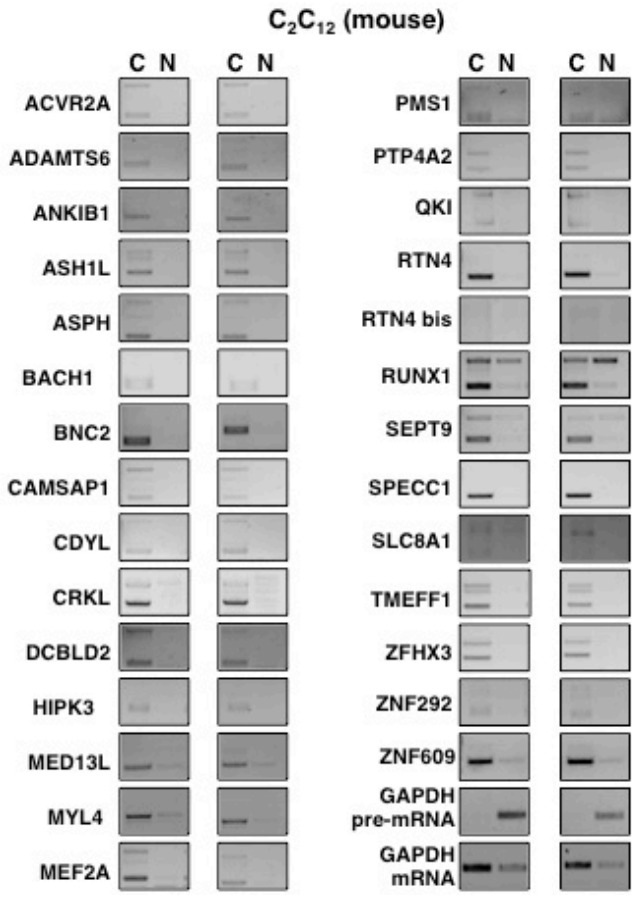



*Figure 26. Mouse circRNAs that passed the previous validation steps were screened for subcellular localization by RT-PCR amplification from cDNA obtained by retrotranscribing nuclear and cytoplasmic RNA separately. Fractionation was performed in duplicate.*

## 2.2. An RNAi-based functional screening reveals circular RNAs actively involved in myogenesis.

At this point, having set a list of experimentally validated circular RNAs, all expressed at high levels and conserved between human and mouse myoblasts, we sought to understand whether among them there were any functional ones. Defining "functional" a certain molecule is trickier than one could expect. As a general rule of thumb, functionality in classical genetic studies was defined by the insurgence of a phenotype in association to a certain genotype. If such mutation was mapped in a gene, then that gene was thought to be responsible for the phenotype, it therefore had a "function". With the advent of genetic engineering, a different definition of functionality arose. Overexpression and knock-down of gene products became possible and many studies focused on cellular or molecular phenotypes, namely quantitative or qualitative features of cultured cells or cellular extracts specifically achieved after the experimental perturbation (e.g. the proliferation rate or the transcriptome). The theoretical space of the possible phenotypes is then virtually infinite, as any feature observed in any model system after a perturbation in the concentration of a bio-molecule could be interpreted as sympthomatic of a "function". To discuss the meaning of a phenotype observed in cell culture systems and its relationship with in vivo model systems is beyond the aim of this thesis. Anyway, I'll use the word "phenotype" as indicative of "function" from the molecular point of view. The aim of this work, indeed, is mainly to understand if circular RNAs can impact biological processes, it doesn't matter if in an artificial system or in living organisms. To do so, most of the focus is centred on the design and perfection of an experimental set up that is useful to univocally attribute a phenotype to the function of a given circular RNA, a task that is sufficiently difficult itself.

The first issue to be solved was to define the phenotype to observe. Since the working hypothesis was that circular RNAs might be involved in myogenesis, we wanted to set a list of quantitative

parameters describing the “myogenic state” of cultured cells. We opted for leading an in vitro differentiation experiment of myoblasts after knock-down of each circRNA candidate, followed by fixation and staining of cells at two stages: I) after 48h of differentiation (72 post-knock-down) and II) after 96h of differentiation (120 post-knock-down). At those time points, we stained cells by immunofluorescence for Myogenin (Myog), a marker of early differentiation, and Myosin Heavy Chain (MHC), a marker of terminal differentiation, in addition to nuclear DAPI staining (set-up experiment shown in figure 27). We performed the experiment in a 96-well format in triplicate and took a picture of each treatment with a 4X objective in an automatized system for capturing an area corresponding to almost the entire well. We ended up with a total of more than 200 pictures in 3 channels (Myog/MHC, DAPI, DIC) that were then analysed with a script that we wrote especially for this aim. We included the definition of a set of parameters used for quantitate the “myogenic state” of the treated cells. First of all, for cells at 48h of differentiation, we measured the total number of cells, the fraction of Myog-positive cells and the intensity of the Myog signal. For cells at 96h of differentiation, we measured the number of nuclei, the Fusion Index (number of nuclei embedded in multi-nucleated fibers / total number of nuclei), the average number of nuclei per fiber, the Maturation Index, that we defined as the number of nuclei embedded in multi-nucleated fibers / the nuclei in MHC-positive cells, the dimensions of MHC-positive cells and MHC-positive fibers, the fraction of MHC-positive nuclei and the fraction of MHC-positive fibers. All those parameters are indicative of the differentiation state of the cells and not necessarily correlated, providing enough information to detect substantial changes in the differentiation process in response to knock-down experiments. The quantity of such parameters was then called a “phenotype” if statistically significant as defined by Students t-test calculation followed by Bonferroni correction with a 0.1 null-hypothesis rejection threshold for each triplicate with respect to the bulk of the

experiment. The sum of significant parameters for each treatment contributed to the estimate of the intensity of the phenotype, therefore providing a tool for choosing a few candidates to further characterize. In addition to this analysis of differentiation, we decided to chose a subset of circRNAs that were highly expressed in proliferating cells for an additional phenotype screening: following knock-down, we replaced the culture medium with bromodeoxyuridine (BrdU)-containing medium, which is used for labeling of the replicating DNA with BrdU-specific antibodies, therefore allowing to monitor the proliferation rate of the cells.

A second, important issue to solve was how to achieve efficient and specific circular RNA knock-down. The molecular tool that we decided to use is RNA interference, as it is fast, cheap and usually efficient. SiRNAs had never been used for circular RNAs, we therefore had to design an appropriate strategy, including the sequence specificity, the chemistry, the delivery. To achieve circRNA-specific knock-down, one has to consider that circular RNAs usually (in our case, always) are produced from loci that also make linear transcripts containing the same sequence of the circle. Therefore, the only sequence that is really specific for the circRNA product is the head-to-tail junction, where we decided to design our siRNAs. Moreover, siRNAs can work as microRNAs if targeting with their seed sequence any given RNA, therefore they have to be designed in a way that excludes off- targets. The most obvious off-targets are the linear counterparts of circular RNAs, since part of the head-to-tail junction is embedded in their linear sequence. Where possible (overcoming technical restrictions to siRNA design), we opted for positioning the seed sequence on the head-to-tail junction. Since this wasn't always achievable, the chemistry of siRNAs was changed, including a patent-protected modification (*On-target plus*, Dharmacon) in the seed sequence that prevents microRNA-like effects when the siRNA binds only with its seed, without extensive 3' complementarity. To further reduce the risk of off-targeting effects and to increase the chance to obtain efficient knock-down, we tried to design two different

siRNAs for each target. Following these rules, it was possible to design two siRNAs for 20 circRNAs, one for 5 circRNAs and none for the 4 remaining ones (see list of siRNAs in Materials and Methods).

Regarding the delivery of siRNAs, we had first to decide for an appropriate cell line, and subsequently set up the optimal experimental procedures. We decided to use human primary myoblasts (9208, Telethon Biobank), because they resemble molecular and morphological features of myogenesis better than commercial cell lines and are relatively easy to culture and differentiate in vitro (figure 2 and 27). Due to their poor transfectability, we tried several concentrations of different liposome-based reagents and developed a reverse transfection protocol that allows high transfection efficiency of siRNAs in these cells (see Materials and Methods).

Having set up the whole experiment, we proceeded with transfection, differentiation and analysis of the high-content genomic screen (figure 28). From the same samples, the relative abundance of the circular versus linear form was tested in order to confirm the specific down-regulation of the circRNA. This revealed that at least one siRNA for 17 different genes specifically down-regulated the circular form without affecting the linear mRNA (figure 29).

As depicted in figure 30, many phenotypes were associated to the circRNAs knock-down. Among them, we focused on two species (circ-QKI and circ-BNC2) that displayed strong altered and opposite effects (representative images in figure 31). Of the two siRNAs used for circ-QKI, only one was specific for the circ-RNA (siRNA QKI#1), while the second induced a strong repression also of the linear mRNA (figure 32A). Both siRNAs inhibited differentiation, thus indicating that both circular and linear species may cooperate in the same biological process (figure 30). As an additional control, we used siRNAs that specifically target QKI mRNA and repeated our morphological analysis together with new

circ-QKI knock-down experiments (figure 32B). The phenotype associated to the circular RNA knock-down was confirmed and a similar phenotype was associated with the mRNA knock-down (figure 32B). Western blot for Myosin Heavy Chain confirmed these results, suggesting that both circ-QKI and QKI mRNA are required for proper differentiation of human myoblasts (figure 32C). Of the two siRNAs used for circ-BNC2, one strongly repressed the circRNA and had a mild effect on BNC2 mRNA (siRNA circ-BNC2#1), while another resulted in poor knock-down of the circRNA (figure 32A). We focused on the first siRNA and, again, compared its effect on the myogenic phenotype and by Western Blot to the result of specific BNC2 mRNA knock-down (figure 32B and C). In this case, mRNA down-regulation produced a stronger pro-myogenic phenotype, indicating one of two alternative scenarios: a) as for circ-QKI and QKI mRNA, the effects of circular and linear RNAs are cooperative or b) the effect on differentiation observed in both cases is due BNC2 mRNA knock-down and not to circ-BNC2. This second scenario is also intriguing, given the fact that the BNC2 locus produces alternatively the linear mRNA in myoblasts and the circular RNA in differentiated myotubes (figure 19). Therefore, circularization of exon 6 of the BNC2 pre-mRNA mediated by some specific factor might be a way of getting rid of the anti-myogenic BNC2 mRNA when differentiation is stimulated.

Lastly, both circ-QKI and circ-BNC2, up-regulated during in vitro differentiation of control myoblasts, are down-regulated in DMD conditions (figure 33); this compares well to the notion that dystrophic cells have a much slower progression into the differentiation process and confirms that those two genes might play a role in correct myogenesis.

## Figure 27

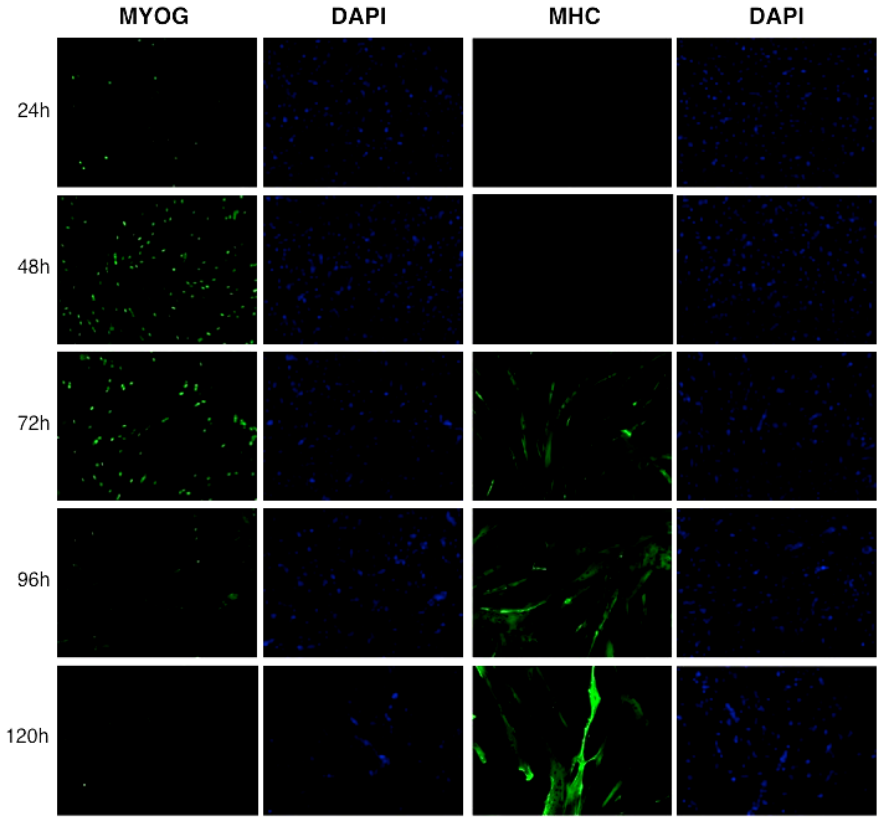



*Figure 27. Immunofluorescence analysis of Myogenin and Myosin Heavy Chain (MHC) during differentiation of human primary myoblasts (from 24h to 120h post-differentiation). DAPI staining is shown next to each picture.*

## Figure 28

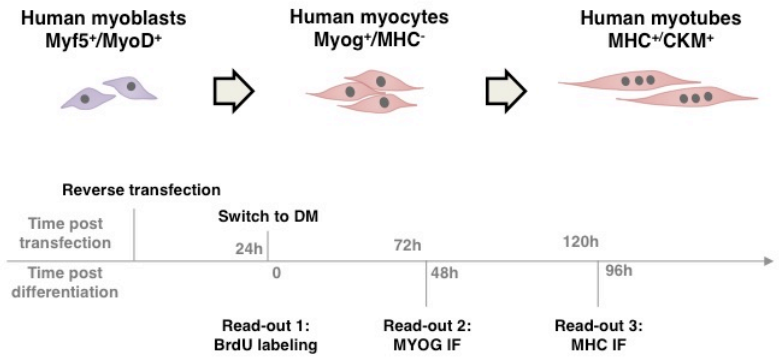



*Figure 28. Schematic representation of the RNAi screening setup. Myoblasts are cultured for 24 hours after reverse transfection of siRNAs, then switch to differentiation medium and fixed for Myogenin and MHC immunofluorescence after 48 and 96 hours of differentiation.*

## Figure 29

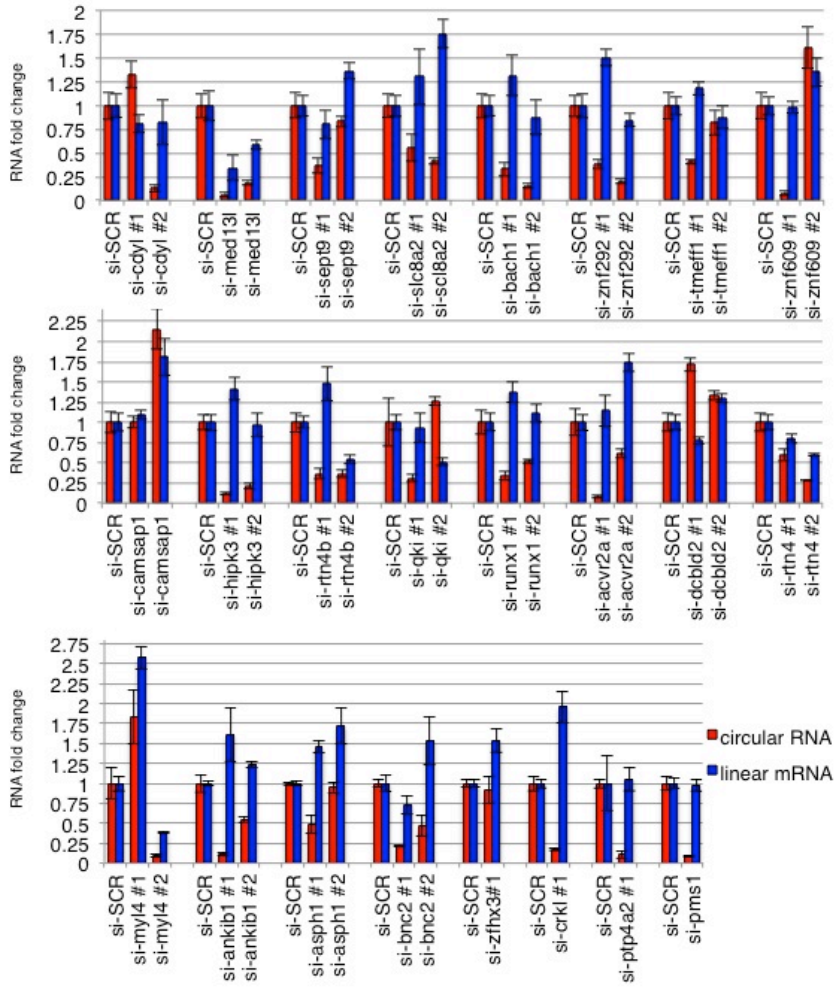



*Figure 29. Knock-down efficiency of circRNAs and linear mRNAs for each siRNA, as measured in the high-throughput screening pilot experiment. All circRNAs/linear mRNAs level in control scramble siRNA (si-ctr) are set at 1. RNA levels in siRNA-treated cells are shown as fold change with respect to control and normalized to HPRT mRNA, measured by qRT-PCR. Levels of circRNA and linear mRNA are shown in red and blue, respectively.*

**Figure 30**

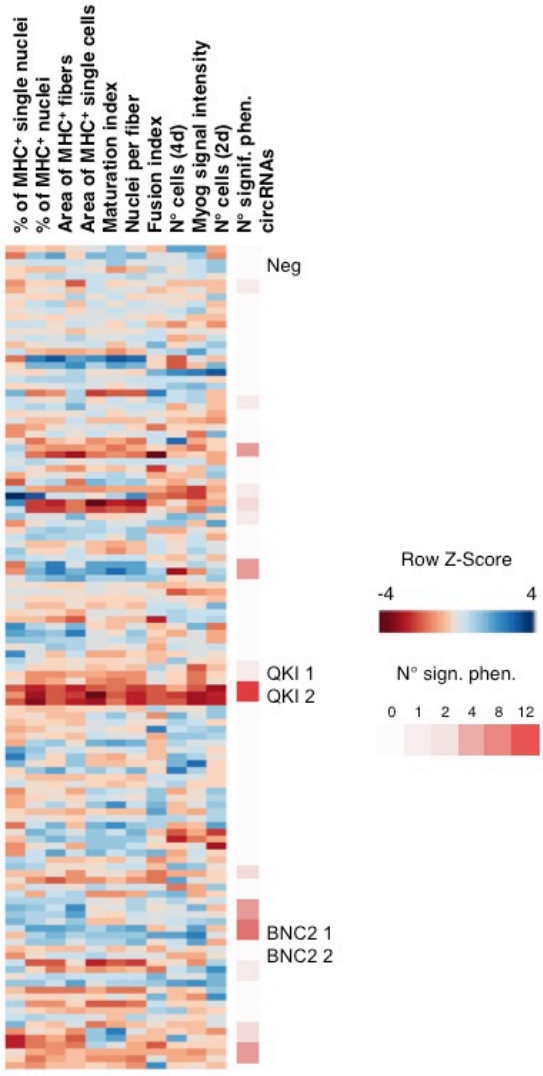



*Figure 30. Heat map showing the analysed "phenotypes" in color scale after Z-score normalization. High values are shown in blue, while low values in red. Each column indicates one phenotype, each row one sample. The number of detected significant phenotypes is shown in red scale on the right.*

**Figure 31**

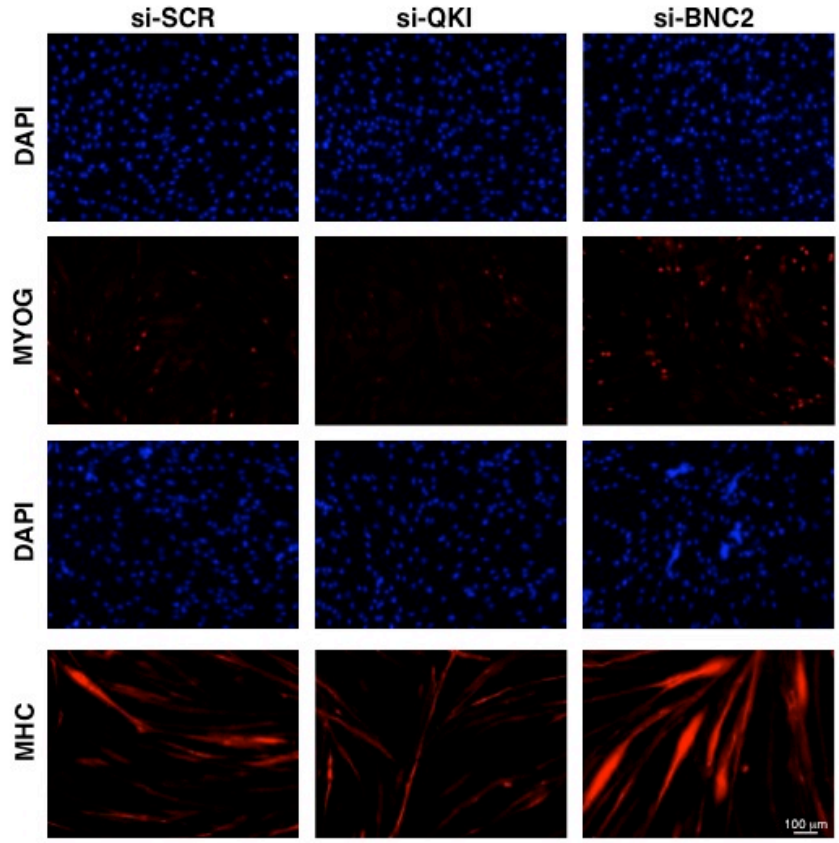



*Figure 31. Human primary myoblasts were treated with either a control scramble siRNA (si-SCR) or with siRNAs against circ-QKI (si-QKI, oligo #1) and circ-BNC2 (si-BNC2, oligo #1) and induced to differentiate for 48 and 96 h. A representative immunofluorescence for MYOG and MHC together with DAPI staining is shown.*

## Figure 32

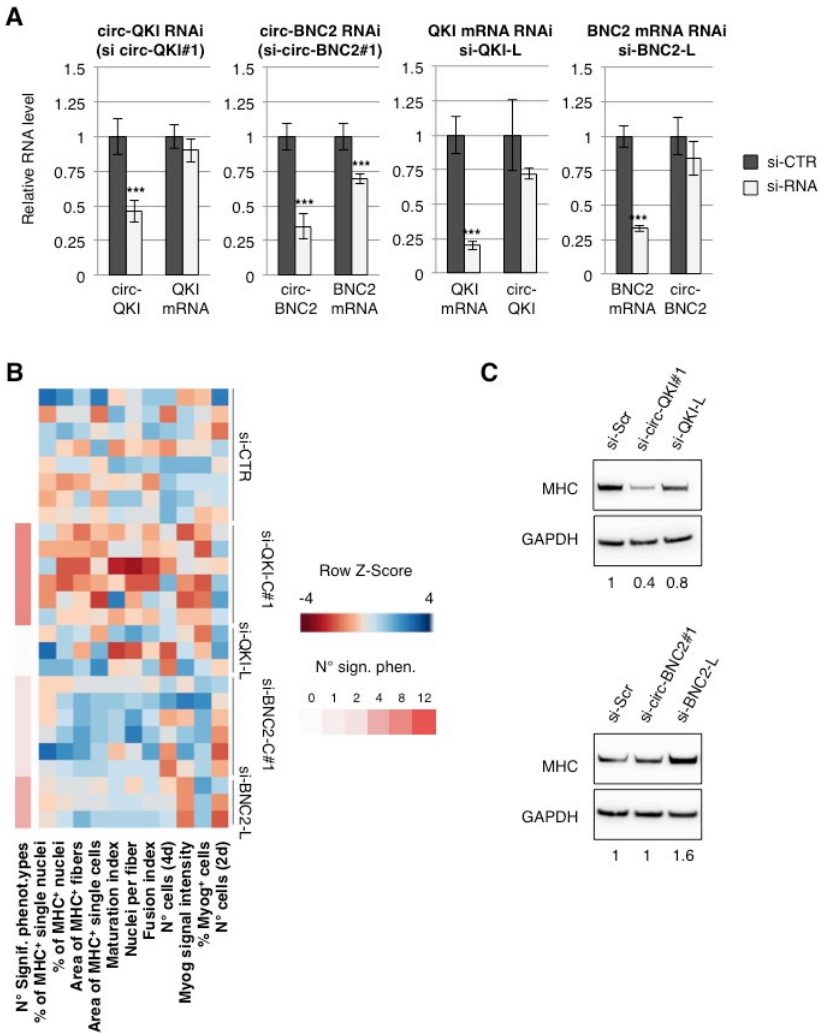



*Figure 32. A. Knock-down of QKI and BNC2 circular and linear isoforms with the indicated siRNAs. Dark grey bars represent RNA level in negative control (si-CTR) and are set at 1. Light grey bars represent RNA level after siRNA treatment. All RNA levels were measured by qRT-PCR in two independent experiments and normalized to Hprt mRNA. B. Heat map showing selected "phenotypes" in color scale after Z-score normalization. High values are shown*

*in blue, while low values in red. Each column indicates one phenotype, each row one sample, indicated on the right. The number of detected significant phenotypes is shown in red scale on the left. C. Western blot analysis of Myosin Heavy Chain (MHC) and GAPDH after knock-down of QKI (upper panel) and BNC2 (lower panel) circRNA and mRNA. A representative blot is shown on top of densitometric measurement of MHC/GAPDH in two independent experiments.*

## Figure 33

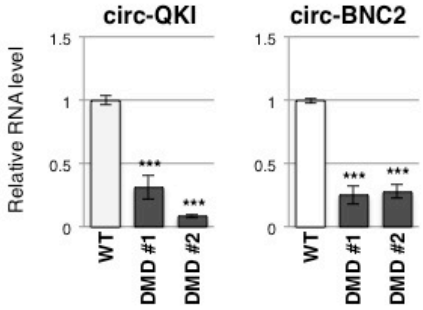



*Figure 33. RNA levels measured by qRT-PCR of circ-QKI and circ-BNC2 in human wild type (WT) and Duchenne (DMD) differentiated cells. RNA measurements were performed from biological duplicates of 1 WT cell line (light grey bars) and 2 DMD cell lines (dark grey bars).*

## 2.3. Circ-ZNF609 regulates myoblasts proliferation.

As previously introduced, to investigate the role of circRNAs in myoblasts proliferation, we performed an independent analysis using bromodeoxyuridine (BrdU) labeling of nascent DNA (figure 34). Among the circRNAs tested, we found that circ-ZNF609 knockdown strongly reduced BrdU incorporation, suggesting that it is required for myoblast proliferation. This initial screening was validated by adding 3 new biological replicates. The results showed that circ-ZNF609 down-regulation reduced BrdU incorporation by 70% (figure 35A). Moreover, markers of proliferation such as CCDN1 and CDC25 were also significantly reduced (figure 35B). Confirming the selectivity of the siRNA, circ-ZNF609 levels were reduced by 90% of the control conditions, whereas the linear form was completely unaffected (figure 35B). To further confirm the specificity of this phenotype, we proceeded with different approaches that may define a standard of circRNA functional studies. In addition to circRNA-specific siRNAs (in this case a single siRNA, because the second didn't work), we used linear mRNA-specific siRNAs, we performed the experiment also in mouse myoblasts and we performed complementation experiments. Knock-down of the linear mRNA did not induce any change in BrdU incorporation, suggesting that the circRNA itself is responsible for the phenotype (figure 35C). Knock-down of the circRNA in mouse myoblasts achieved results similar to the human system, although at lesser extent (not shown). This was anyway expected because we used a murine cell line ($C_2C_{12}$) having a proliferation rate that exceeds the one of any primary cell culture system, possibly due to mutations in the cell division control pathway. Lastly, and most importantly, the proliferation phenotype due to the knock-down of circ-ZNF609 was complemented by transfection of an in-vitro synthesized circRNA, confirming its specificity beyond reasonable doubts (figure 35D).

Notably, circ-ZNF609, which is down-regulated during myogenesis in control myoblasts, is instead found at elevated levels in DMD conditions (figure 36A). This is consistent with the delayed differentiation phenotype of dystrophic primary myoblasts, paralleled by persistence of a high proportion of proliferating, non-fusing, myoblasts. Moreover, consistently with circ-ZNF609 involvement in cell proliferation, this molecule is also up-regulated in cell lines derived from patients with Rhabdomyosarcoma, a rare tumour deriving from muscle progenitor cells (figure 36B).

**Figure 34**

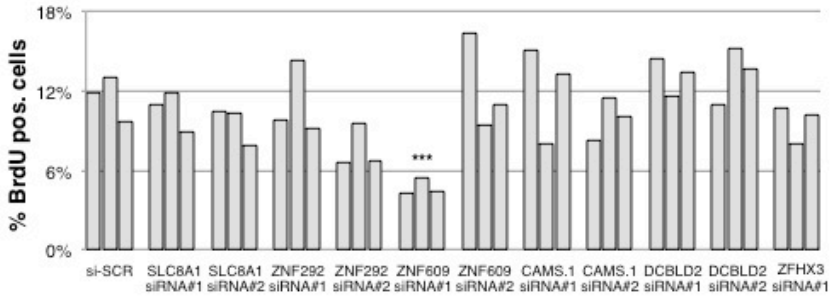



*Figure 34. BrdU incorporation screening. Primary myoblasts were treated with either a control scramble siRNA (si-SCR) or with siRNAs against the circRNAs indicated and maintained in growth conditions. Bar plots represent the percentage of BrdU positive cells. The values derive from three biological independent experiments. 3 asterisks indicate a p-value < 0.01 estimated with Student's T test between an individual triplicate and the bulk experiment.*

## Figure 35.

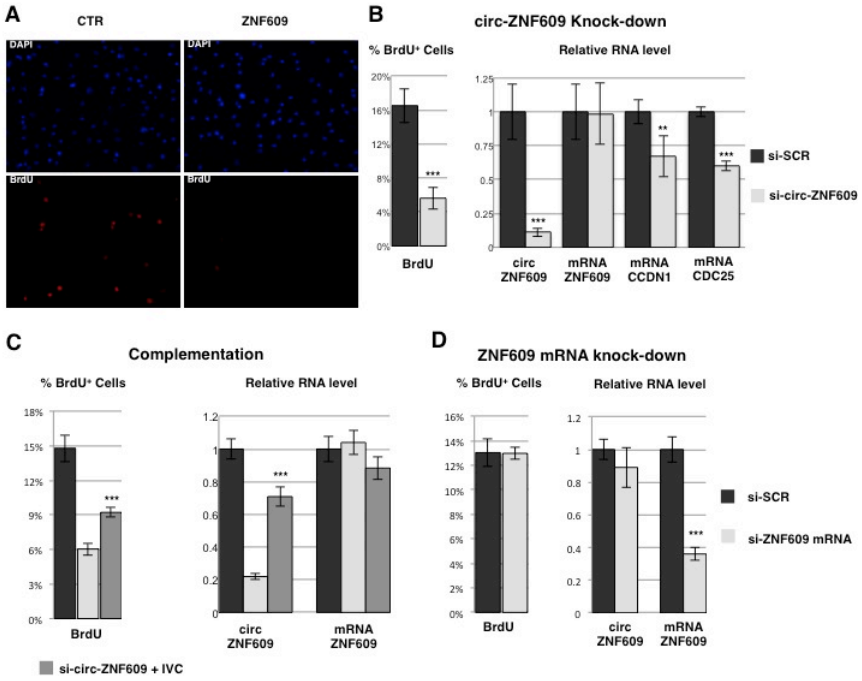



*Figure 35. circ-ZNF609 specifically controls myoblasts proliferation. A. Left panel: primary myoblasts were treated with either a control scramble siRNA (si-SCR) or with siRNAs against circ-ZNF609 (si-circ-ZNF609, oligo #1) and maintained in growth conditions. A representative BrdU labeling assay is shown together with DAPI staining. Right panel: bar plot showing BrdU incorporation levels in cells treated as in the left panel (dark grey bars, si-SCR; light grey bars, si-circ-ZNF609). B. RNA quantification by qRT-PCR of circ-ZNF609 and ZNF609, CCND1 and CDC25 mRNAs upon siRNA transfection (dark grey bars, si-Scr; light grey bars, si-ZNF609). C. As panel B with the addition of a sample treated with siRNA for circ-ZNF609 plus complementation with synthetic circ-ZNF609 (IVC, intermediate grey shadow). D. Same as panel B, but with interference of linear ZNF609 instead of the circular form (si-ZNF609 mRNA). The values derive from three biological independent experiments: 1 asterisk indicates a p-value < 0.05, 2 asterisks indicate a p-value < 0.02 and 3 asterisks a p-value < 0.01 measured with Student's t test.*

## Figure 36

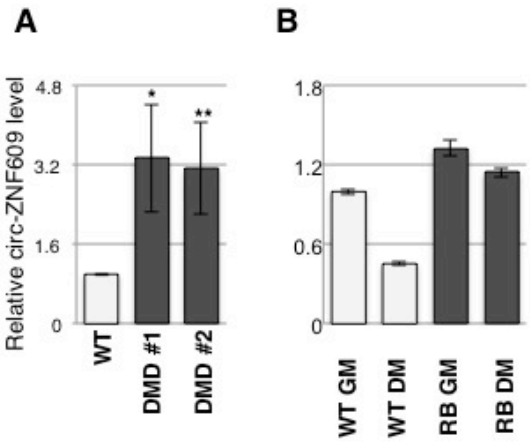



*Figure 36. circ-ZNF609 is mis-regulated in pathological conditions. A. qRT-PCR level measured in WT and DMD cells measured by qRT-PCR (mean and standard deviation from biological duplicates). B. Preliminary results of circ-ZNF609 in WT versus Rhabdomyosarcoma cell lines (RB) in GM and DM. 1 asterisk indicate a p-value < 0.05, 2 asterisks indicate a p-value < 0.02 and 3 asterisks a p-value < 0.01 measured with Student's t test.*

## 2.4. Biochemical analyses suggest protein-coding capacity of circ-ZNF609.

Having univocally attributed a role in myoblast proliferation to circ-ZNF609, we sought to identify the molecular mechanism behind its function. One suggestion came from the analysis of its sequence and genomic position. It originates from the circularization of the second exon of the ZNF609 gene. Interestingly, a 753-nt open reading frame is present in circ-ZNF609, spanning from the AUG of the host gene to a stop codon created three nucleotides after the splice junction (figure 37). Translation of natural circular RNA templates has been observed in Archea (Kjems et al., 1988), but never in Eukaryotes. On the other hand, the chance of circRNA coding potential is indicated by some observations of artificial circularized mRNAs being translated in cultured cells (Abe et al., 2015). In order to determine whether this ORF was indeed translated, we first used sucrose gradient fractionation of human cell lysates to test circ-ZNF609 association with polysomes. Figure 38 shows that a significant portion of circ-ZNF609 sediments with heavy polysome fractions, while the rest remains in the non-bound fraction. The circular conformation of the polysome-bound circ-ZNF609 RNA was confirmed by RNaseR resistance (figure 39). In puromycin-treated myoblasts, circ-ZNF609 shifted to lighter polysomes, similarly to the behaviour of the linear ZNF609 and HPRT mRNAs, suggesting that its sedimentation pattern was due to active translation. A similar observation came from fractionation of mouse cell lysates (not shown). On the other hand, the fact that only a portion of the total circ-ZNF609 migrates with polysomes suggests that, besides the chance of a protein coding function, also a merely RNA-dependent mechanism is possible.

Since antibodies against the N-terminal and C-terminal portions of ZNF609 protein were available, we tested the presence of a circ-ZNF609 encoded-peptide in human and mouse myoblasts. As shown in figure 40A, besides a band corresponding to the full-

length ZNF609 recognized by both antibodies, the antibody targeting the ZNF609 N-terminus revealed a band compatible with a circRNA-encoded peptide (predicted weight 26 kDa, band observed between 25 and 30 kDa). Moreover, such band was down-regulated in response to circ-ZNF609 knock-down in both human and mouse cells (figure 40B). A protein-synthesis template without a 5-methyl-guanosyne cap is not a conventional mRNA. Internal ribosome entry sites (IRES) have been observed for viral and cellular transcripts, and if a natural circular RNA is translated, it has to contain an IRES by definition, as it lacks a 5' end. Interestingly, the circ-ZNF609 sequence is significantly more conserved than the rest of the ZNF609 exons, mainly thanks to the contribution of the 5' UTR region, which is roughly twice conserved among vertebrates with respect to the rest of ZNF609 5' and 3' UTRs (figure 41). We decided to clone the sequence spanning the head-to-tail splice junction of circ-ZNF609 until the AUG start codon in an IRES reporter vector that we produced. It is made of the juxtaposition of two luciferase-encoding ORFs (Firefly and Renilla), and a sequence that is to be tested for IRES activity cloned between those two ORFs. The luciferase activity of the two enzymes can be easily measured form cell lysates, therefore allowing fast and sensitive measurement of the two proteins concentration. The ratio Firefly/Renilla (F/R) can be then used as an estimate of the IRES activity of the sequence cloned in between, as the Renilla luciferase can be translated only if the ribosome is engaged after the Firefly stop codon. We compared the F/R ratio of 5 constructs: an empty vector (p-Luc), a vector containing the ZNF609 putative IRES sequence (p-Luc-UTR), a vector containing the same sequence with the addition of an artificial intron for reconstituting the splice junction of circZNF609 (p-Luc-intr.-UTR), a splicing-mutant version of the latter (p-Luc-intr.-UTR-Δ-spl.) and a positive control containing an IRES derived from the encephalomyocarditis virus (p-Luc-EMCV-IRES). As shown in figure 42, the vector that produces the circ-ZNF609 sequence after splicing makes a bicistronic mRNA

capable of potent IRES activity, even higher than the viral IRES that we tested. Interestingly, such activity was completely abolished after mutating the donor splice site, indicating that internal translation entirely depends on splicing. It is interesting to note the exact same RNA sequence, which has not experienced the splicing event, is not able to trigger translation of the Renilla luciferase. Therefore, if circ-ZNF609 is translated, it is likely that it depends on the Exon Junction Complex or any other modification that is deposited on the RNA as a consequence of splicing.

Although these data altogether strongly indicate that circ-ZNF609 can encode a functional protein, additional experiments are required for proving it unequivocally. Currently, we are producing plasmids that are able to encode stable artificial circular RNAs and we will use them to produce flag tag-containing circular ORFs, which should produce a flag tagged protein only if circularization is achieved. On the other hand, we are producing stable mouse embryonic stem cell clones carrying a flag tag-coding sequence immediately before the predicted stop codon of circ-ZNF609, inserted through the CRISPR/Cas9 technology. Finally, targeted mass spectrometry could confirm if the protein encoded by circ-ZNF609 exists in a natural context.

## Figure 37

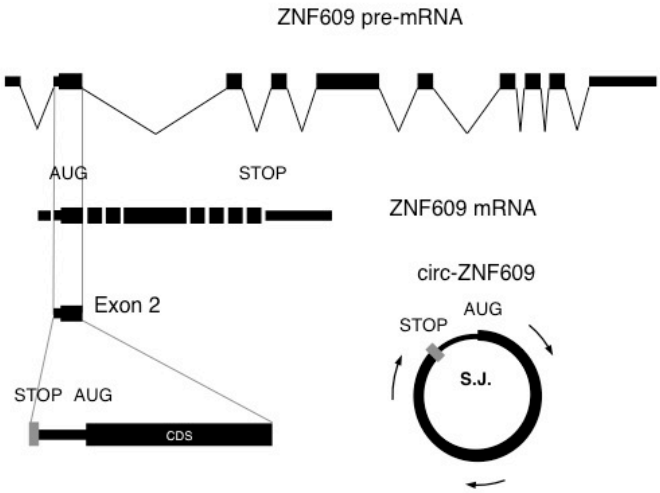



*Figure 37. Schematic representation of the ZNF609 pre-mRNA (top), mRNA (middle) and exon 2 that can be circularized (bottom). The AUG and STOP codon are indicated.*

**Figure 38**

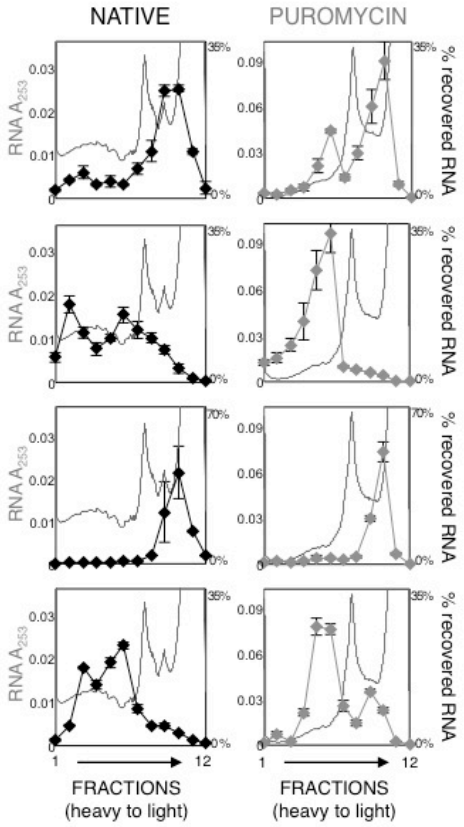



*Figure 38. Cytoplasmic extracts from proliferating primary human myoblasts, either untreated (native, black lines) or treated with puromycin (grey lines), were loaded on 15-50% sucrose gradients and fractions measured by absorbance at 253 nm (light grey behind each plot). Fraction density decreases from left to right; the panels show one representative profile out of three independent biological replicates. Individual fractions were analysed by qRT-PCR and represented as percentage of RNA level in each of the 12 fractions. Profiles are shown for circ-ZNF609 and ZNF609 mRNA. circ-PMS1 was used as a non translatable RNA, while HPRT is used as a positive control.*

## Figure 39

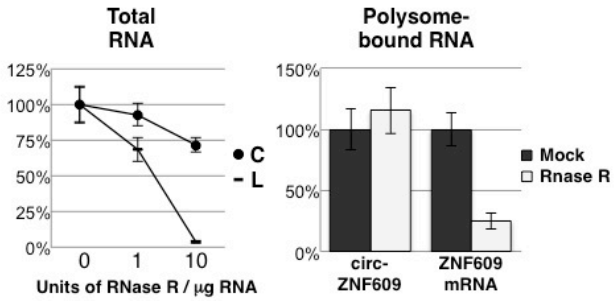



*Figure 39. RNaseR treatment of total RNA and polysome-bound RNA. Left panel: set-up of the treatment on total RNA with different amounts of enzyme. The percentage of the RNA, recovered after treatment and measured by qRT-PCR, is shown. Right panel: RNase R treatment of polysome-bound RNA (obtained by sucrose fractionation): the percentage of recovery of circular and linear ZNF609 in control and RNase R treated samples is shown.*

**Figure 40**

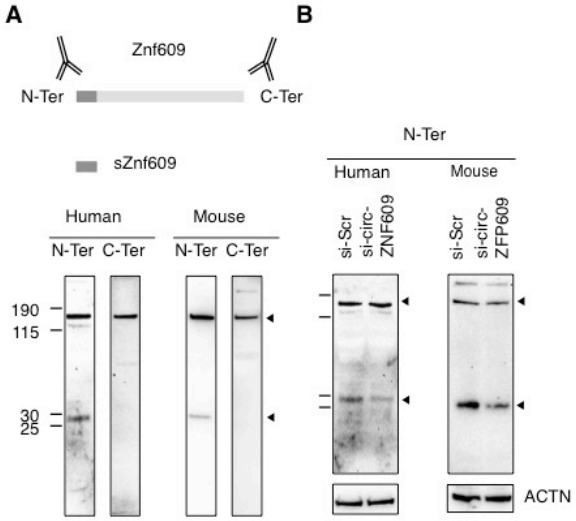



*Figure 40. Western Blot analysis of ZNF609-encoded proteins. A. Upper panel: schematic representation of the ZNF609 protein with the antibodies used in this work. Lower panels: Western blot analysis on human and mouse myoblast protein extracts with N-ter and C-ter targeting antibodies. B. Western blot analysis of human and mouse myoblasts treated with control and circ-ZNF609 siRNAs. The black arrows show the predicted size of the full-length and the putative circ-ZNF609-encoded proteins.*

## Figure 41

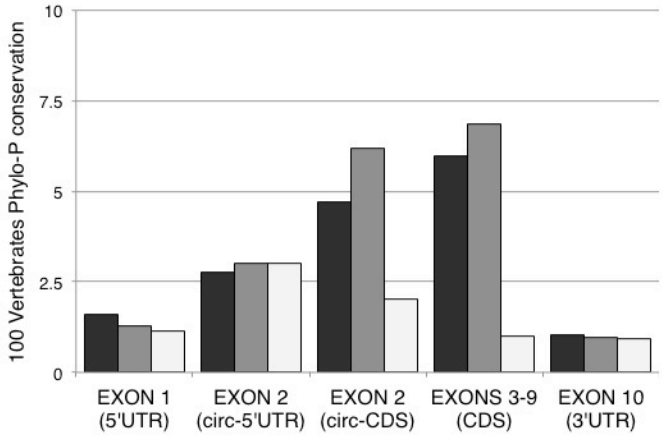



*Figure 41. Conservation analysis of circ-ZNF609. Phylo-P-derived conservation score for 100 Vertebrates are shown in the bar plot, divided by the reading frame (aligned to the one of the ZNF609 mRNA ORF, +1 in dark, +2 in middle and +3 in light grey). Exon 2 (that produces circ-ZNF609) has been divided into the predicted 5' UTR and the coding sequence (CDS).*

## Figure 42

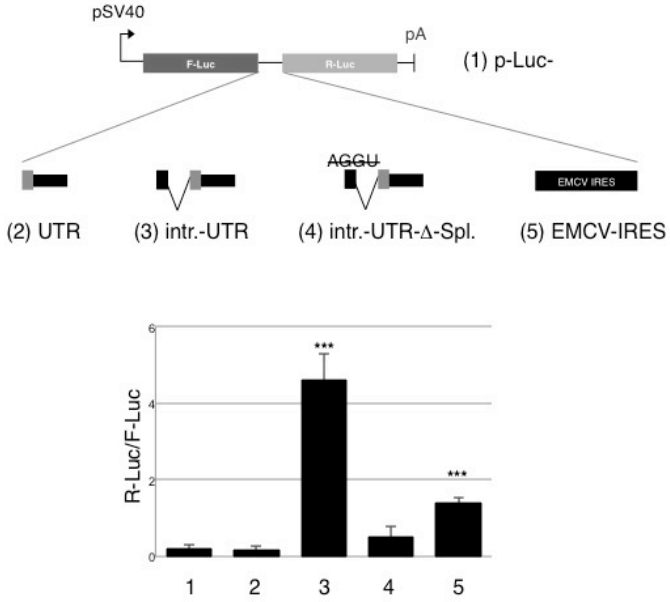



*Figure 42. Luciferase analysis of 5 different constructs (represented in the upper panel) performed in mouse myoblasts in 5 biological replicates. Mean and standard deviation are shown in the bar plot. 3 asterisks correspond to a p-value < 0.01 derived though Student's t test with respect to the negative control (p-Luc).*

# Concluding Remarks

As stated before in this thesis, the main aim of my work was to try to answer one simple question arising from several observations about circular RNAs expression in Eukaryotes: do these molecules do something?

One example of functional circular RNA was previously identified (CDR1as/CiRS-7, Memczak et al., 2013 and Hansen et al., 2013), but it was and it is clear that it represents an exceptional molecule, with biosynthetic and molecular features significantly different from the rest of circular RNAs. Moreover, it is still not known whether this molecule impacts on a relevant biological process.

The implications of knowing about circular RNAs functional capacities are many and important: these molecules are thousands or tens of thousands in human cells, they can be extremely abundant and their expression has already been observed to vary in response to physiological changes. They may be involved in development, disease, inheritance, ageing, signaling and whatever RNA-related phenomenon we can think about. Moreover, they can be turned into useful tools for biology and medicine, thanks to their exceptional stability and the easiness they can be produced with.

In this work I tried to add a bit of information to the field, and achieved some specific and general pieces of knowledge that may result useful in the future.

First of all, I studied and described the expression pattern of circular RNAs in muscle differentiation and found several features that may be generalized, in agreement with a few similar reports in different model systems. We now know that circular RNAs are globally regulated in differentiation and disease, such regulation occurs primarily at the transcriptional level and in second instance

at the splicing level. They are stable and can be accumulated over time. Their evolutionary conservation and conserved pattern of expression suggest that their production may be an evolutionary-relevant event and they might be even functional on their own. Testing such functionality can be achieved by means of dedicated knock-down strategies and read-out tools, which indeed demonstrate that molecular phenotypes are associated with them. Finally, the specific case of circ-ZNF609 reveals for the first time that a circular RNA is definitely able to control relevant cellular activities - in this case cell division - and this is possibly due to protein-coding activity, a fact that may disclose an entire new fraction of the eukaryotic proteome that has not been explored so far.

# Materials and Methods

*Cell culture and treatments.*

Mouse myoblasts ($C_2C_{12}$, ATCC) were cultured in Growth Medium (GM, DMEM supplemented with 20% FBS, L-Glutammine 2 mM and Penicillin/Streptomycin) and induced to differentiate in Differentiation Medium (same as GM, but FBS reduced to 0.5%). Human myoblasts (9208 (wt) and 9981 (DMD), from Telethon Biobank) were cultured in Growth Medium (GM, DMEM supplemented with 10% FBS, L-Glutammine 2 mM, insulin 50 mg/ml, FGFb 25 ng/ml, EGGF 1 ng/ml, Penicillin/Streptomycin) and induced to differentiate by culturing in Human Skeletal Muscle Differentiation Medium (DM, Promocell).

Reverse transfection of human myoblasts was performed as follows: 96-well format optical plates (Corning) were treated overnight with 350mg of Collagen Type I in water and 0.5% acetic acid, then washed with PBS and UV-sterilized. Twenty microliters of siRNA transfection mix (150 nM siRNA, 0.2 µl Dharmafect lipid reagent from Dharmacon, 20ml transfection medium: DMEM with L-glutammine 2mM and insulin 50 mg /ml) was seeded in each well, left at RT for 20 min and then diluted with 30 µl of transfection medium. Cells (12000 per well) were added in each well after resuspension in double complete medium (GM with double concentration of serum, antibiotics, FGFb and EGF). Cells were switched to DM after 24h.

*RNA isolation, treatments and analysis.*

Total RNA in this study was extracted with Qiazol reagent according to the manufacturer's protocol (Qiagen). RNA from 96-

well plates and nuclear/cytoplasmic fractions was purified with miRNEasy spin columns according to the manufacturer's specifications (Qiagen).

RNase R treatment was performed as follows: 1 or 6 µg of total RNA was diluted in 20 µl of water with 1 µl of RNase R and 2 µl of RNase R buffer (Epicentre), then incubated 15 min at 37°C and purified by phenol chloroform extraction. Ten picograms of a DNA spike-in molecule were added to each reaction for qPCR normalization.

Retrotranscription of RNA in this study was achieved with VILO Superscript enzyme (Life Technologies): up to 2 µg of RNA was retrotranscribed in a 10-ml reaction mix with 1 µl of enzyme and 2 µl of buffer, then incubated 10 min at 25°C, 60 min at 42°C and 5 min at 85°C.

qRT-PCR analyses in this study were performed as follows: cDNA (2-20 ng equivalent) was added to a reaction composed of 7.5 µl 2X SYBR Mastermix (Qiagen), 1.5ml of 5mM primers and water to a final volume of 15 µl. DNA amplification (40 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by melting curve analysis) was monitored with an ABI 7500 Fast qPCR instrument.

RT-PCR for circRNA detection was performed with 0.2 µl of Mytaq DNA polymerase (Bioline) in 25 µl water with 5 µl reaction buffer, 15 ng of cDNA, 2.5µl 5mM primers. Reaction was carried out for up to 32 cycles at 95°C for 25 s, 55°C for 25 s, 72°C for 25 s. Five microliters of final products were run on 2% agarose gels. All primers used in this study are reported in Supplementary Table 1. RT-PCR for circular and linear RNA from the same locus was carried out for the same number of cycles.

*Western blot.*

Cells were harvested with 50-100 µl of Protein Extraction Buffer (100 mM Tris pH7.5, EDTA 1 mM, SDS 2%, PIC1X (Complete, EDTA free, Roche) and incubated 20 min on ice , then centrifuged at 15000 x g for 15 min at 4°C. 15 µg of proteins were loaded on 4-12 % bis-tris-acrylamide gel (Life technologies) and transferred to a nitrocellulose membrane. The membrane was blocked in 5 % milk and then hybridized with: 1.5 mg / ml anti-ZNF609 antibody (Sigma Aldrich, AV3178) for 2 hours at room temperature. Anti-rabbit secondary antibody was used for detection after one-hour incubation at room temperature.

*RNAseq and circular RNA detection.*

Illumina TruSeq library preparation was performed by Istituto di Genomica Applicata (Udine, Italy) from 1µg of total RNA, depleted of ribosomal RNA with Ribominus technology (Life Technologies). For each sample, 20-40 million 100bp long paired-end reads were sequenced (raw data available at GEO, accession number GSE70389), then trimmed and analysed as discussed in the text. Briefly, for standard gene expression analysis, reads were mapped with TopHat, transcriptomes defined by Cufflinks and differential expression called by Cuffdiff. For circRNA detection, paired-end reads were split and used separately as input to the following pipeline: reads were aligned to rRNA with Bowtie2, then remaining ones were mapped to hg19 or mm10 genome assemblies with Bowtie2. Reads remaining after these alignments were used as input to FindCirc (Memczak et al., 2013) for finding circRNAs. For quantitative analyses of circRNA regulation in human muscle differentiation, circRNAs with at least 5 reads in one sample were used. The total number of reads mapping to the head-to-tail splice junction was used for measuring circRNA expression level, while the mean of total reads mapping linearly to the same splice junction was used for measuring the abundance of the linear isoform. For comparing WT and DMD samples, circRNA expression level was normalized by dividing the reads mapping to

the head-to-tail junction by the total number of reads used for the analysis.

After the circRNA detection process, all human circRNAs were annotated using bedtools to the Ensembl GRCh37 gene set.

More details and all scripts are available upon request.

*Cell fractionation.*

Nuclear/cytoplasmic fractionation of human and mouse cells was performed according to Legnini et al. (34) as follows: cells (1 x 106) were washed with PBS and immediately lysed on ice with 500 ml of cytoplasmic lysis buffer (140 mM NaCl, 1.5 mM MgCl2, 10 mM Tris pH 7.5, 0.5% NP40) for 5 min. After scraping the cells and placing them in a 2-ml collection tube, 500 µl of a sucrose cushion (same as lysis buffer with 12% Sucrose) was gently pipetted to the bottom of the tube. Sample was centrifuged 10 min at 12000 x g. Supernatant and pellet were then separately resuspended in 1 ml of Qiazol, while sucrose was discarded. RNA extraction was carried out as described above.

Cytoplasm fractionations on sucrose gradients were performed as follows: 5x106 cells were lysed 5 min on ice with 500 ml of TNM (10 mM Tris pH 7.5, 10mM NaCl, 10 mM MgCl2) lysis buffer supplemented with 10 mg/ml cycloheximide, 1X PIC (Complete, EDTA free, Roche) and 1X RNase guard (Thermo Scientific), then spinned 2 min at 15000 x g for discarding nuclear pellets. Samples were centrifuged on 9 ml 15%-45% sucrose gradient at 38000 rpm with a SW41 rotor (Beckman) for 1 h 30 min at 4°C. Fractions were collected with a Bio-logic LP (Biorad), 0.001 ng of a spike-in DNA as internal control was added to each fraction, together with 10 mg of glycogen, 100 mg of Proteinase K (Roche) and 40 ml of SDS. Samples were left 1 h at 37°C and RNA was purified by PCA extraction and precipitated with isopropanol. For puromycin treatments, medium was replaced 3 hours before lysis with 1 mM puromycin containing-medium.

*Immunofluorescence and image analysis.*

Cells were washed twice in 100 ml PBS, fixed with 100 µl of 4% paraformaldehyde for 15 min at room temperature, washed twice with 100 µl of PBS, permeabilized and blocked with 100 µl of PBS with 0.2% Triton X-100 and 1% goat serum, incubated with 50 µl of in-house made anti-myogenin/anti-myosin antibody (available upon request), washed twice in PBS, incubated with 50 µl anti-mouse Cy3/AlexaFluor488 secondary antibody diluted 1:300 in PBS-DAPI and finally washed twice in PBS. Acquisition for image-based phenotypic analysis was performed with a widefield Nikon TiE Microscope equipped with a Lumencor SpectraX LED light source and PerfectFocus 3; 4× and 10× objectives were used (Nikon Instruments).

The acquired images were analysed using two custom image analysis algorithms (available upon request), developed and executed in the Acapella software development/run-time environment (Perkin Elmer). Both algorithms used a Perkin Elmer proprietary algorithm on the DAPI staining for nuclei detection.

DAPI and MYOG staining at 48 h were used to extract (for each siRNA treatment) the total cell number, the percentage of MYOG+ cells, and the average MYOG signal intensity. The DAPI&MYOG algorithm workflow can be summarized as follows: 1) nuclei detection (on DAPI channel); 2) MYOG signal intensity estimation (from Cy3 channel) and 3) MYOG +/- nuclei classification. In the absence of a cytoplasmic marker, following the segmentation on DAPI, Voronoi tessellations were used to define local regions for each detected nucleus. Thus, foreground and background masks were defined to represent the nuclear area and its local background. These masks were further used to extract average DAPI and MYOG intensities for every nucleus. The standard deviations of MYOG intensity in the nuclei local background were estimated. The DAPI/MYOG signal intensity of every nucleus was defined as the difference between the DAPI/MYOG intensities of its local foreground and background. This local approach compensates for

both intra- and inter-image spatial variability of illumination and minimizes possible errors in estimating DAPI/MYOG signal intensities over different wells (and fields). Following a similar local approach, and this time compensating for different inter-image noise level/content, the MYOG-positive and -negative classification was developed to use a local image adaptive threshold to apply on the MYOG signal intensity. For each analysed image, the threshold was defined as 1.5*median of all standard deviations of MYOG signal intensities of the local backgrounds of all nuclei detected in the current image. Then, each nucleus was classified as MYOG+ or MYOG- based on the comparison of its MYOG signal intensity and the current local image threshold (see the example Figure showing the MYOG-positive classification).

DAPI and MHC staining at 96 h were used to extract (for each siRNA treatment) the total number of cells, the percentage of MHC-positive nuclei, and additional morphological parameters like the fusion index and the number of nuclei per fiber. The algorithm workflow can be summarized as follows: 1) nuclei detection (on DAPI channel); 2) nuclei count correction procedure; 3) cytoplasm segmentation (on Red channel); 4) MHC signal intensity estimation (from the Red channel); 5) MHC-positive and -negative nuclei classification; 6) Red Cell/Fiber MHC+ nuclei classification. Prior to all analyses, a circular mask was applied on the images of all channels in order to remove vignetting.

Following DAPI segmentation of the nuclei, we applied a nuclei count correction procedure. This was required to overcome the issue of contiguous, ‘syncytial-like’ nuclei that were present in myotubes. The nuclei count correction procedure was developed based on a combination of some morphological features of the detected nuclei: the normalized nuclear area (with respect to the median of all nuclei area of the current image) and the nuclear axial length ratio (nucleus width/length). We could distinguish three classes of nuclei. First, all nuclei having a normalized area smaller than 1, which count as 1 nucleus each. A second class was

defined by taking the integer of values derived after applying the floor function to nuclei with a normalized area greater than 1 (for example, nuclei with values between 2.0 and 2.99 were scored as 2, while those between 3.0 and 3.99 were scored as 3). Third, we defined a subclass with all nuclei having a normalized area between 1.5 and 2 and with an axial length < 0.5 were counting as 2 nuclei each.

We used a Perkin Elmer proprietary algorithm to define the cytoplasmic area for each nucleus. The cytoplasmic area is assigned to the nuclei based on the signal in the Red channel. This was associated with two problems. First, because of the peculiar morphology of cells and fibers (cells have elongated shape, fibers appear stitched/clumpy and are multinucleated), individual cells/fibers were erroneously subdivided. Second, in some cases, the method attempted to assign cytoplasm to cells that, on visual inspection, had no signal in the Red channel.

Thus, we defined a population of 'cell' objects including a nuclear region (with one or more nuclei) and a cytoplasmic region. All the detected 'cell' objects that had a cytoplasmic/nuclear area smaller than one were removed from the cytoplasmic segmentation result, since they were potentially MHC-negative.

From the remaining population of cells, we defined a cellular mask (cytoplasm and nuclear areas) on which MHC-positive/negative nuclei classification was performed. All nuclei with at least one quarter of their area overlapping the cellular mask were classified as MHC-positive. All other nuclei were labelled as MHC-negative.

The Red Cell/Fiber classification was implemented using a work-around of the cytoplasm segmentation. The definition of a fiber was as follows: stitched/clumpy cells with multiple and short distanced nuclei (clusters). Red Cells were defined as elongated and mono-nucleate. We used the presence of nuclei clusters to differentiate between Red Cells and Fibers. Specifically, nuclei that were within 100 μm of each other were defined as being part of the same cluster. Thus, we used the morphological dilation

operator (circular structuring element of 30 pixels diameter) on the nuclear mask given by all MHC-positive nuclei. The resulting 'dilated' mask was then intersected ('AND' operator) with the cellular mask to separate nuclei of neighbouring Red Cells. This resulting binary image was used to define nuclei clusters by using connected component labelling (D. H. Ballard and C. M. Brown, "Computer Vision," Prentice-Hall, 1982.) and selecting only those components containing more than one MHC-positive nucleus. Finally, all MHC-positive nuclei belonging to nuclei clusters were labelled as 'Fiber nuclei'. Similarly, all those MHC-positive nuclei that were not part of a nuclear cluster are labelled as Red Cell nuclei. The result of this classification was transferred to the nuclei-associated 'cell' objects, and thus labeled as Fibers and Red Cells.

For each well, this algorithm reports a set of 8 parameters. The first four parameters are based on nuclei, namely, total number of (detected) nuclei, percentage of MHC-positive nuclei, percentage of MHC-positive Red-Cell nuclei, average number of nuclei per Fiber (estimated by the average number of nuclei per nuclei cluster). The following two parameters are cell area descriptors: the fraction of Red Cell area and the fraction of Fibers area. These fractions were estimated as the total area of Red Cells/Fibers with respect to the entire area of the analysed image. Two additional parameters were implemented to better describe cellular phenotypes. The fusion index was defined as the fraction of total number of Fiber nuclei to the total number of detected nuclei. The maturation index was defined as the fraction of the total number of Fiber nuclei to the total number of MHC-positive nuclei. All raw images, segmentation results and scripts are available upon request.

For BrdU staining, we used 5-Bromo-2'-deoxy-uridine Labeling and Detection kit I (Roche). Labeling reagent was added to the cell culture medium (diluted 1:1000) 3 hours before fixation (human primary myoblasts) or 15' before fixation (C2C12), then cells were fixed with a solution made of 70% ethanol and 30% 50mM glycine

at pH 2.0 for 10 minutes at -20°C. Cells were then washed twice with PBS and incubated for 1 hour at room temperature with 30 µl of antibody anti-BrdU diluted 1:10 in PBS. Cells were again washed twice and incubated for 30 minutes with anti-mouse Cy3/AlexaFluor488 secondary antibody diluted 1:300 in PBS-DAPI and finally washed twice in PBS. We took 4 pictures of each well with a Zeiss AXIO Observer A.1 with a 10X objective, both in the DAPI and Cy3 channels. These images were analysed manually by using ImageJ, by setting a fixed threshold for both DAPI and BrdU signal intensity and detecting nuclei with the built-in particle analysis plug-in.

*Statistical Analyses.*

All experiments were repeated for at least 3 times; arithmetic means together with standard error are reported in all bar plots shown in this study. Statistical significance for comparisons of means was assessed by Student's T or Wilcoxon-Mann-Whitney U test depending on the data. P-values below 0.05 were marked by 1 asterisk, while 2 asterisks indicate a p-value < 0.02 and 3 asterisks a p-value < 0.01. When multiple comparisons were performed, p-values were scaled according to the Bonferroni correction and then filtered for corrected p-value < 0.1.

*PCR primers used in this study*

For each circular RNA was used a couple of divergent oligos named F1 and R1. For each linear RNA, an oligo in an external exon was coupled with one of the two divergent oligos (either R2 with F1 or F2 with R1).

HUMAN circular

>hsaAcvr2aR1:CCAGGTAGCAAAACAATGCCGCC

>hsaAcvr2aF1:GATGGCCTACCCTCCTGTACTTG

>hsaAnkib1R1:GATGAGTGCTTTACGGAATTTGGTGG
>hsaAnkib1F1:GGCTCCTCTCTTTACTGCTGAAG
>hsaAsh1lF1:GGTTGGCTATTGGAAGAACAGACCAG
>hsaAsh1lR1:CTATCCACAGAGTCATCCCCAGAAG
>hsaAsphF1:CAGATGATCCAGAACAAAAAGCAAAAG
>hsaAsphR1:GCCTCCACAGGAACCTGCTCCTC
>hsaMef2aR1:GTCCTCCGAGAGTGGACTGTGC
>hsaMef2aF1:CTCACAGTGCCAAATGGAGCTGG
>hsaMyl4R1:GCACTGGCCGTAGGTGATCTTC
>hsaMyl4F1:GGGGATGTACTGCGGGCCCTG
>hsaNebF1:GTGGACGCCATTCCCCTGTTG
>hsaNebR1:GGAGTGCAGCATCTTGGGATCG
>hsaPms1R1:CTTGAAAGGAGTCGAACTGTTGCC
>hsaPms1F1:GATCTCCTCATGAGCTTTGGTATCC
>hsaPtp4a2R1:GGAGTGACGACTTTGGTTCGAG
>hsaPtp4a2F1:CTCGGTGTCCAGGAGTCTTC
>hsaRtn4F1:CACTCAAGCAGAGATAGAGAGCATAG
>hsaRtn4R1:GCAGCAGGAAGAGCAAAAAGGGTC
>hsaRunx1F1:CTGCCTTTAACCCTCAGCCTCAG
>hsaRunx1R1:GGCGACTTGCGGTGGGTTTGTG
>hsaSeptR1:GTGGAGTTGGGTGTCTCGACC
>hsaSeptF1:GAGCCTGCCCCTGTGTCTCAGC
>hsaSlc8a1F1:CTGCCAGAGGTGGAGGGGAGG
>hsaSlc8a1R1:GGTGGGTGAAAGACTTAATCGCCG
>hsaTmeff1R1:GGCATGCACATTTCAAACCATCTCC

>hsaTmeff1F1:GCTGAGTGTGATGAAGATGCAG
>hsaTtnF1:GAAGTGGCCCCTGTTCCTATCC
>hsaTtnR1:CCTTATACCTGTGACTGACACCTCC
>hsaCdylR1:AGGCTTAGCTGTTAACGGG
>hsaCdylF1:CTGTTCCGGCTCCCAAGTGT
>hsaZfp609F1:AAACCGGAGCCAGAGGAAGG
>hsaZfp609R1:CAGCTATGTTCTCAGACCTGC
>hsaBnc2F1:GTTCACTTGAAAGAGATGCACG
>hsaBnc2R1:CTGAAGGGTGATGATTTCCTC
>hsaMEd13lF1:CTTAACAGAATTGGATCATCTTGGGC
>hsaMed13lR1:AATTAATCCCTGACTGAATTGGAAAGG
>hsaCamsap1F1:CCAGTGTCAAGCGCTTCTCAAC
>hsaCamsap1R1:GCTTGGACAGGAGAAGCTTGATAAC
>hsaZfhx3R1:CGAGTCACAGCCTTCCATGGTAAGG
>hsaZfhx3F1:CTACTACACCAACAGCCTGGAGAAG
>hsaDcbld2R1:CGCGGATTTTTGGCCTCATACTC
>hsaDcbld2F1:GGATGTAAGGGTTCCACTCTCAG
>hsaAdamts6R1:CCCACCAAAGTGATGGAAACACC
>hsaAdamts6F1:GGATGTTCCACTGTTTAAGAGCCAC
>Convergent_circCRKL_fwd/hs:CTACCTGCAGTTTCCGGTTC
>Convergent_circCRKL_rev/hs:GGTTGGGTGCTGAGACAGAT
>hsaBach1F1:CAG AAC AGC TGG ATT GTA TCC
>hsaBach1R1:GTTGTCGGGAAGTTCAGTGG
>hsaRtn4bisF1:GTC TCT CCA GTA CAG GAG GTC
>hsaRtn4bisR1:GAGGAGTTGGTTCAGAAGTAC

>hsaHipk3F1:CGGCCAGTCATGTATCAAAG
>hsaHipk3R1:CCTGGAATACACAACTGCTTGG
>hsaSpecc1F1:CTGTCTTGCAATGAGCTCAG
>hsaSpecc1R1:TACTGAAAATTCCCGTGGGG
>hsaZnf292F1:CCC AGG AAC CAT TGG ATA AG
>hsaZnf292R1:CAGAGGTAAGATAAGGTCGG
>hsaYYY-1F:ATGACCTGACATTGCTACTCC
>hsaYYY-1R:CACAACAAAGGTGCTCCCAG

linear-specific oligo (coupled with circular F1)
>hsaAcvr2aR2:GACACAACCAAATCTTCCCCTTGC
>hsaAnkib1R2:CAGTTCTCCGGGTTGGACATCC
>hsaAsh1lR2:GCTGGGGTTTCAGAAGGACTGG
>hsaAsphR2:CTGAGGGTATTTGCGTACTAGTTC
>hsaMef2aR2:GACTTTGCCTAAGCTATTTGCACC
>hsaMyl4R2:GGTGCCCTGCTCCTTGTTGC
>hsaNebR2:GGAGTGCAGCATCTTTGGATCG
>hsaPms1R2:GACTAGTGAAAGAGTGGTCTGC
>hsaPtp4a2R2:GAGCGTTCAATTCCAAACAGC
>hsaRtn4R2:GTCTCTCCAGTACAGGAGGTCAAC
>hsaRunx1R2:GTAGGACTGATCGTAGGACCAC
>hsaSeptR2:CTCTGGGGGTGTCGGCCATGTC
>hsaSlc8a1R2:GGTTCCTCAAGCACAAGGGAG
>hsaTmeff1R2:GCTTCTCGAACAAAACAGGGATTG
>hsaTtnR2:CTTTTGGGGGAGCAGCAGGTTCC

>hsaCdylR2:CGATTGCGTTTCAGCGTGAG
>hsaZfp609R2:CCATGATCTGCTCACAAGGAGA
>hsaBnc2R2:CCATGTCATCGAGTTCTTTGG
>hsaMed13lR2:TTTGAAGAGCAGCGTCCTACATTC
>hsaCamsap1R2:TGAGCTGGACTTTCCAATAACTGTTG
>hsaZfhx3R2:TTTCCCGTTGTTGCGCATGGCG
>hsaDcbld2R2:CAGACAACCAGCTGGGCAGTAC
>hsaAdamts6R2:GCTGCAGCTCCTTTCAGGCTC
>Divergent_circCRKL_fwd/hs:GGTGTCCGAGAACTCGCGG
>hsaBach1F2:CTGAGCTGGATTTAGCGAAG
>hsaRtn4bisR2:GTAGGTTTGTGCAGTTACAGC
>hsaHipk3R2:CTGATCATACTCCAAGGCTC
>hsaSpecc1R2:AGCTGTTCGGCTTGGTGATC
>hsaZnf292R2:CTGGATGGTCAGAACACAGC

MOUSE circular
>mmuAcvr2aF1:GGGTGTACAGACATCACAAGATGG
>mmuAnkib1F1:GGCCCCAGGATCTTCGTAGG
>mmuAnkib1R1:GTGCTTTGCGGAATTTGGTGGTTG
>mmuAsh1lR1:GTGCTTTCTGATGATCGCTGGG
>mmuAsh1lF1:GAACATAAGAAGGGGGTGAAGAGG
>mmuAsphF1:GAACAGCAAGACACACCACCAG
>mmuAsphR1:GAGTCTCTGCCTCCTCCTCCGGTG
>mmuMef2aF1:GACCTCACGGTGCCAAATGGAGC
>mmuMef2aR1:CCTCCGAGAGTGGACTGTGCTC

>mmuMyl4R1:CTGCCCGTAGGTGATCTTCATC
>mmuMyl4F1:GTGGGGACGTGCTGCGGGCC
>mmuNebF1:TACAAAGCAAAGCATGAAGGTGAG
>mmuNebR1:CTTCTCCCAGTCTTGTTTGTAGAC
>mmuPms1F1:CCTGAAACCTGATGTGAGGATTAC
>mmuPms1R1:GAACTGGACAGGAGGCGAAC
>mmuPtp4a2R1:CAACGTATATTCCAACGAAAAACC
>mmuPtp4a2F1:GCTAGTTGCGCTTGCATTGATTG
>mmuRtn4R1:GCAGGAAGAGCAAAAAGGGTCTC
>mmuRtn4F1:GGGCAACATAGTTAAACCCAAAG
>mmuRunx1F1:CTCAGCCTCAAAGTCAGATGCAG
>mmuRunx1R1:GATGGCTCTATGGTAGGTGGC
>mmuSeptR1:CAATCTCCTCGACTTCAAACGATCTC
>mmuSeptF1:CCCAACTGTCCGCTCCCGTC
>mmuSlc8a1F1:CACCTGTGGAGAGCTCGAATTCC
>mmuSlc8a1R1:CCAGAGCTACCAGACGAAATCCC
>mmuTmeff1R1:CATGCGCATTTCAAACCATCTCC
>mmuTmeff1F1:GTGACGAGGATGCGGAGAACG
>mmuTtnF1:CAGAAGAAGTGGTCCCTGTTCCC
>mmuTtnR1:CCT GAG ACA GAC ACC TCT TCC TC
>mmuCdylR1:GCACCAGATATTCTGTCTTCCC
>mmuCdylF1:GTGACTGCTGCCATGGCCAC
>mmuZfp609F1:GAAGGGGAGAATGAGTGCCG
>mmuZfp609R1:GTC AAC GTC CCA CCT CAA GGT TC
>mmuBnc2F1:GCTTCACTACAGGAACGTTCAC

>mmuBnc2R1:GCG GTC CAG TAC CTT GCC AG

>mmuMEd13lR1:CTGAAATTATGGGTCCACAGTCC

>mmuMed13lR1:CTGTGTGTATGGCGTCGTGATG

>mmu1aCamsap1F1:CCTGTCCAGTGAGCTGTATTGCC

>mmu2Camsap1R1:GAT GTA AAA CGG GTC TCG GAG G

>mmuZfhx3R1:CATTGTCCTTCCCCGAGACGAC

>mmuZfhx3F1:GTGCAACGCTTGTGACTACTAC

>mmuDcbld2R1:GTAAGGGTTCCACTCTCAGG

>mmuDcbld2F1:CAGTGCTGTTCATGAGTGGAAC

>mmuAdamts6R1:GCCAAGCTCAGAGTGATAGG

>mmuAdamts6F1:CCATTCTCTCCCACCAAAGTG

>Convergent_circCRKL_fwd/mmu:CCTGAAGAGCAGTGGTG GAG

>Convergent_circCRKL_rev/mmu:GATGGCAAAGGGTTGATC GC

>mmuBach1R1:TATGCACAGAGGACTCGTAG

>mmuBach1F1:AGT CGG AAA TCG AGA AGC TG

>mmuRtn4bisF1:GAGGCGTCTCTTCTTAGTTGATG

>mmuRtn4bisR1:GCCGTTACACTGACAATGCTG

>mmuHipk3F1:GGATCGGCCAGTCATGTATC

>mmuHipk3R1:CCGCTTGGCTCTACTTTGAG

>mmuSpecc1F1:GTCTTGCACAGAGCTCAGACAAG

>mmuSpecc1R1:GATAGTTACTGAGAGCTCCC

>mmuZnf292F1:CTCCACTATCCTTTCCCAGG

>mmuZnf292R1:CACTCAGAGGTAAGGTATGG

linear-specific oligo (coupled with circular F1)

>mmuAnkibR2:CCGCTGGCAGCAGTTCTCTG

>mmuAshlR2:CACAGGTTGCAAGGTATTTTCAGG

>mmuAsphR2:CTTCAAATGCGTTCACTGCTTCC

>mmuMef2aR2:CCTGGTGGCGGGGGAGACTTTG

>mmuMyl4R2:CATCTCGAAGTCCAGTGTCTTGG

>mmuNebR2:CACCTTGATGTCGAACTTCTTAGC

>mmuPms1R2:CATTAGAGCCATCCTGTGATCTGG

>mmuPtp4a2R2:CTCTGAAGCGTAACCGCATC

>mmuRtn4R2:CTCTCCAGTACAGGAGGTCAAC

>mmuRunxR2:GAAATGGGTGTCGCTGGGTG

>mmuSeptR2:CAGCTGGGAGTCACCTCGGAG

>mmuSlc8aR2:CTTTCTTCTCACTCATCTCCACC

>mmuTmeffR2:GCTTTATACACGACGCTTCTCGG

>mmuTtnR2:GGGCAGCTGGCTCTTTCTTCTTG

>mmuCdylR2:CAC TGA AAC GCA ACC GCT TGT C

>mmuZfp609R2:TGA CCA CTG GCA CTA ACA GG

>mmuBnc2R2:CGATGCAGGTTTATATTGGCAC

>mmuMed13lR2:GGC CAT TTT CCC AGA GTC CTT C

>mmuCamsap1R2:GTGCTCTCGCCGATAACGGAC

>mmuZfhx3R2:CGCCACTCTCATGCTGCTGC

>mmuDcbld2R2:GCAAAAGGCAAAAGACAGCCAGC

>mmuAdamts6R2:CACACATTCCAGCCACAGAGG

>Divergent_circCRKL_fwd/mmu:TCATGCGTATGCTCAACCTC

>Divergent_circCRKL_rev/mmu:TGCTGAGACAGAACCCACTG

>mmuBach1F2:GC GCA GAG GGA GTG AGT CAC

>mmuRtn4bisR2:CAGGAATACTGAAGAGTGAGATC

>mmuHipk3R2:CCAGTGACCACATGTCTATG

>mmuSpecc1R2:CTGCTCAGCTGTTCTGCTTG

>mmuZnf292R2:CAGCTTTGCTAGAGCGGTTG

*siRNAs used in this work*

circ-Acvr2a dplx 1:UGUUCCAACUCAAGUGCUAUU

circ-Acvr2a dplx 2:UUCCAACUCAAGUGCUAUAUU

circ-Ankib1 dplx 1:ACUUCGAGCUCAUGAAUGUUU

circ-Ankib1 dplx 2:UCAUGAAUGUGAAAGAUGUUU

circ-ASPH dplx 1:GCAAAAGGACUUAAAGAGAUU

circ-ASPH dplx 2:AAAAGGACUUAAAGAGAGAUU

circ-BACH1 dplx 1:GAAGCUGGUUGAUGAUAAUUU

circ-BACH1 dplx 2:AGAAGCUGGUUGAUGAUAAUU

circ-BNC2 dplx 1:GACAGGAUGCUGCUGGCAAUU

circ-BNC2 dplx 2:GCCGAAGCCGAGACAGGAUUU

circ-CAMSAP1 dplx 1:UGGAUCAACAAGAUAACAUUU

circ-CAMSAP1 dplx 2:GGAUCAACAAGAUAACAUCUU

circ-CDYL dplx 1:UGUUAACGGGAAAGGUUGAUU

circ-CDYL dplx 2:CGGGAAAGGUUGAAAGGAUUU

circ-CRKL dplx 1:UGGCAUUAGAGGUAUCCAAUU
circ-DCBLD2 dplx 1:AGAUAAACAAGGUGAUGGAUU
circ-DCBLD2 dplx 2:AUAAACAAGGUGAUGGAUGUU
circ-HIPK3 dplx 1:GGUACUACAGGUAUGGCCUUU
circ-HIPK3 dplx 2:AAUCUCGGUACUACAGGUAUU
circ-MED13L dplx 1:GAGCUGAACUCACGGGAAUUU
circ-MED13L dplx 2:AGAGCUGAACUCACGGGAAUU
circ-MYL4 dplx 1:CCUGAAGAGAUGAAGAUCAUU
circ-MYL4 dplx 2:GCCCAAGCCUGAAGAGAUGAA
circ-PMS1 dplx 1:GUACAUAACAAGCUGCUCUUU
circ-PTP4A2 dplx 1:GAAUCCACGUUCUAGUUUUUU
circ-QKI dplx 1:UCUCUGAAGUUGUUGACCUUU
circ-QKI dplx 2:AUUAUUGGUACCUGCAAAAUU
circ-RTN4 dplx 1:GAGUAAAACUUCAGAUGAGUU
circ-RTN4 dplx 2:AGUAAAACUUCAGAUGAGAUU
circ-RTN4bis dplx 1:UCUCUGAAGUUGUUGACCUUU
circ-RTN4bis dplx 2:UCUCUGAAGUUGUUGACCUUU
circ-Runx1 dplx 1:AGAGUCAGACUGAGGGGAAUU
circ-Runx1 dplx 2:AGUCAGAUGCAGGGGAAAAUU
circ-SEPT9 dplx 1:CAGGAGGCCUUGAAAAGAUUU
circ-SEPT9 dplx 2:AGCCAGGAGGCCUUGAAAAUU
circ-Slc8a1 dplx 1:UGAUGAAAUUGUUAGGUUGUU
circ-Slc8a1 dplx 2:AUUGUUAGGUUGUGACAGUUU
circ-Tmeff1 dplx 1:GCAGAAAAUGUUGGAAUUAUU
circ-Tmeff1 dplx 2:UGGAAUUAAAUGUGAGGGAUU

circ-TTN dplx 1:AAAAGUACCAGCUGUGCAUUU
circ-ZFHX3 dplx 1:CUGAAGUUGUACAAGGUCCUU
circ-ZNF292 dplx 1:GGAUAAAGACACUCCUAGAUU
circ-ZNF292 dplx 2:UUGGAUAAGGAUAAAGACAUU
circ-ZNF609 dplx 1:AGUCAAGUCUGAAAAGCAAUU
circ-ZNF609 dplx 2:UGAAAAGCAAUGAUGUUGUUU

# Glossary

**9208**: human primary myoblasts deriving from the Telethon Biobank.

**Back-splicing**: a particular splicing reaction generating head-to-tail junctions. The result is a circular RNA. It can be triggered by base-pairing of the flanking introns or by RNA binding proteins.

**Bowtie**: a program for fast mapping of ultra-high-throughput short sequencing reads.

**$C_2C_{12}$**: mouse myoblasts cell line widely used in molecular and cellular studies.

**ceRNA**: competing endogenous RNA; a transcript able to regulate other transcripts by binding to a similar pool of microRNAs. **circRNA**: circular RNA; covalently closed circular RNA molecule produced from linear precursors by the spliceosome in a reaction called back-splicing, in which the donor splice site is joined to an acceptor splice site located upstream of the donor.

**CRISPR/Cas9**: Clustered regularly-interspaced short palindromic repeats (CRISPR) are short sequences in prokaryotic DNA that produce small RNAs able to function as guides for the protein Cas9, a DNA endonuclease that is responsible for a form of acquired immunity towards plasmids and phages. This bacterial immune system has been engineered very recently for realizing fast and efficient specific double strand breaks in eukaryotic DNA, which can be followed by homologous recombination in order to achieve precise gene editing (Jinek et al., eLife 2013).

**Cuffdiff**: a program for identifying differentially expressed genes between different RNA-seq samples. It used TopHat alignment results.

**Cufflinks**: a program for reconstituting and quantitating the transcriptome in a RNA sample using the mapping results of TopHat.

**DM**: differentiation medium; a cell culture medium used for inducing myoblast differentiation (see materials and methods).

Head-to-tail junction: a splice junction generated by the process of back-splicing. It characterizes circular RNAs.

**DMD**: Duchenne Muscular Dystrophy; a genetic disorder caused by the loss of dystrophin.

**Findcirc**: a pipeline (and more stringently a script used in the pipeline) used for identifying head-to-tail splice junction from RNA-seq data.

**FPKM**: Fragment per Kilobase of exon per Million fragments mapped. It is a unit for measuring transcript abundance, commonly calculated by Cufflinks and Cuffdiff. It corresponds to the number of reads mapped to a transcript, normalized for the length of the transcript and for the total reads mapped in the experiment. It can be corrected for multiple factors.

**GM**: growth medium; a cell culture medium used for keeping myoblasts undifferentiated (see materials and methods).

**lincRNA**: large intergenic non coding RNA; an RNA molecule larger than 200 nt, usually capped and polyadenylated, whose expression depends on an independent promoter recognised by RNA Pol II.

**microRNA**: small RNA produced from a larger transcript by two subsequent endonucleolytic reactions catalyzed by Drosha and Dicer; it is then loaded onto the RNA induced silencing complex (RISC) binding to the proteins of the Argonaute (Ago) family; they function as a guide (due to partial Watson:Crick complementarity) for RISC to find its targets and repress their expression by either RNA degradation or trnaslation inhibition..

**MRF**: myogenic regulatory factors; historically attributed to a group of transcription factors able to trigger and drive myogenesis. It could be attributed to microRNAs or other non-coding RNAs shown to play a role similar to those transcription factors.

**Myoblast**: committed muscle progenitor cell able to proliferate and enter the differentiation program in response to external stimuli.

**Myogenesis**: the process of muscle differentiation, usually referred to skeletal muscle.

**Myotube**: syncytium of differentiated fused myoblasts.

**Paired-end**: a feature of certain sequencing experiments in which cDNA fragments are sequenced from both termini.

**Ribominus/Ribozero**: a group of methods (or the feature of RNA itself) for purifying RNA from ribosomal species, patent-protected by Thermofisher and Illumina respectively.

**RNA-seq**: RNA sequencing. It denotes an ensemble of techniques used for high-throughput sequencing of cDNA libraries for the purpose of RNA quantification. It is often used referring to the Illumina system, a particular RNA-seq technology that is currently the most widely used worldwide.

**Sequencing read**: the sequence of a fragment obtained by RNA-seq technologies. It's usually associated to an unique ID, a sequence, a quality score for each letter of the sequence and depending on the technology a set of coordinates identifying the position of the cDNA fragment in the support used for sequencing.
**siRNA**: small interfering RNA; chemically identical to a microRNA, it is able to trigger its target degradation due to extensive (perfect) complementarity.

**TopHat**: a program for mapping spliced reads. It uses the short read aligner Bowtie and then analyses the mapping results for retrieving splice junctions.

# Bibliography

Abe, Naoko et al. "Rolling Circle Translation Of Circular RNA In Living Human Cells". Sci. Rep. 5 (2015): 16435.

Ala, U. et al. "Integrated Transcriptional And Competitive Endogenous RNA Networks Are Cross-Regulated In Permissive Molecular Environments". Proceedings of the National Academy of Sciences 110.18 (2013): 7154-7159.

Ameres, S. L. et al. "Target RNA-Directed Trimming And Tailing Of Small Silencing Rnas". Science 328.5985 (2010): 1534-1539.

Anderson, C., H. Catoe, and R. Werner. "MIR-206 Regulates Connexin43 Expression During Skeletal Muscle Development". Nucleic Acids Research 34.20 (2006): 5863-5871.

Asakura, A. "Myogenic Specification Of Side Population Cells In Skeletal Muscle". The Journal of Cell Biology 159.1 (2002): 123-134.

Ashwal-Fluss, Reut et al. "Circrna Biogenesis Competes With Pre-Mrna Splicing". Molecular Cell 56.1 (2014): 55-66.

Bajard, L. "A Novel Genetic Hierarchy Functions During Hypaxial Myogenesis: Pax3 Directly Activates Myf5 In Muscle Progenitor Cells In The Limb". Genes & Development 20.17 (2006): 2450-2464.

Ballarino, M et al. "Novel Long Noncoding Rnas (Lncrnas) In Myogenesis: A Mir-31 Overlapping Lncrna Transcript Controls Myoblast Differentiation". Molecular Cell Biology 35.4 (2015): 728-736.

Barrett, Steven P, Peter L Wang, and Julia Salzman. "Circular RNA Biogenesis Can Proceed Through An Exon-Containing Lariat Precursor". eLife 4 (2015): n. pag.

Bosia, Carla, Andrea Pagnani, and Riccardo Zecchina. "Modelling Competing Endogenous RNA Networks". PLoS ONE 8.6 (2013): e66609.

Cacchiarelli, Davide et al. "Micrornas Involved In Molecular Circuitries Relevant For The Duchenne Muscular Dystrophy Pathogenesis Are Controlled By The Dystrophin/Nnos Pathway". Cell Metabolism 12.4 (2010): 341-351.

Cacchiarelli, Davide et al. "Mir-31 Modulates Dystrophin Expression: New Implications For Duchenne Muscular Dystrophy Therapy". EMBO Rep 12.2 (2011): 136-141.

Cazalla, D., T. Yario, and J. A. Steitz. "Down-Regulation Of A Host Microrna By A Herpesvirus Saimiri Noncoding RNA". Science 328.5985 (2010): 1563-1566.

Cazzella, Valentina et al. "Exon 45 Skipping Through U1-Snrna Antisense Molecules Recovers The Dys-Nnos Pathway And Muscle Differentiation In Human DMD Myoblasts". Mol Ther 20.11 (2012): 2134-2142.

Cesana, Marcella et al. "A Long Noncoding RNA Controls Muscle Differentiation By Functioning As A Competing Endogenous RNA". Cell 147.4 (2011): 947.

Chen, J F et al. "The Role Of Microrna-1 And Microrna-133 In Skeletal Muscle Proliferation And Differentiation". Nature Genetics 38.2 (2006): 228-233.

Choudhury, N. R. et al. "Tissue-Specific Control Of Brain-Enriched Mir-7 Biogenesis". Genes & Development 27.1 (2013): 24-38.

Cocquerelle, C et al. "Mis-Splicing Yelds Circular RNA Molecules". Faseb Journal 7.1 (1993): 155-160.

Conn, Simon J. et al. "The RNA Binding Protein Quaking Regulates Formation Of Circrnas". Cell 160.6 (2015): 1125-1134.

Crist, Colin G., Didier Montarras, and Margaret 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 11.1 (2012): 118-126.

da Costa Martins, P. A. et al. "Conditional Dicer Gene Deletion In The Postnatal Myocardium Provokes Spontaneous Cardiac Remodeling". Circulation 118.15 (2008): 1567-1576.

Danan, M. et al. "Transcriptome-Wide Discovery Of Circular Rnas In Archaea". Nucleic Acids Research 40.7 (2011): 3131-3142.

Delaporte, Christine, MichÃ¨le Dehaupas, and Michel Fardeau. "Comparison Between The Growth Pattern Of Cell Cultures From Normal And Duchenne Dystrophy Muscle". Journal of the Neurological Sciences 64.2 (1984): 149-160.

Denzler, Racmy et al. "Assessing The Cerna Hypothesis With Quantitative Measurements Of Mirna And Target Abundance". Molecular Cell 54.5 (2014): 766-776.

Dodou, Evdokia, Shan-Mei Xu, and Brian L. Black. "Mef2c Is Activated Directly By Myogenic Basic Helix-Loop-Helix Proteins During Skeletal Muscle Development In Vivo". Mechanisms of Development 120.9 (2003): 1021-1032.

Fatica, Alessandro, and Irene Bozzoni. "Long Non-Coding Rnas: New Players In Cell Differentiation And Development". Nat Rev Genet 15.1 (2013): 7-21.

Figliuzzi, Matteo, Andrea De Martino, and Enzo Marinari. "RNA-Based Regulation: Dynamics And Response To Perturbations Of Competing Rnas". Biophysical Journal 107.4 (2014): 1011-1022.

Figliuzzi, Matteo, Enzo Marinari, and Andrea De Martino. "Micrornas As A Selective Channel Of Communication Between

Competing Rnas: A Steady-State Theory". Biophysical Journal 104.5 (2013): 1203-1213. 12 Jan. 2016.

Figueroa, A. et al. "Role Of Hur In Skeletal Myogenesis Through Coordinate Regulation Of Muscle Differentiation Genes". Molecular and Cellular Biology 23.14 (2003): 4991-5004.

Gagan, J. et al. "Notch3 And Mef2c Proteins Are Mutually Antagonistic Via Mkp1 Protein And Mir-1/206 Micrornas In Differentiating Myoblasts". Journal of Biological Chemistry 287.48 (2012): 40360-40370.

Grabowski, Paula J., Arthur J. Zaug, and Thomas R. Cech. "The Intervening Sequence Of The Ribosomal RNA Precursor Is Converted To A Circular RNA In Isolated Nuclei Of Tetrahymena". Cell 23.2 (1981): 467-476.

Guo, Junjie U et al. "Expanded Identification And Characterization Of Mammalian Circular Rnas". Genome Biol 15.7 (2014): n. pag.

Hansen, Thomas B et al. "Mirna-Dependent Gene Silencing Involving Ago2-Mediated Cleavage Of A Circular Antisense RNA". The EMBO Journal 30.21 (2011): 4414-4422.

Hansen, Thomas B. et al. "Natural RNA Circles Function As Efficient Microrna Sponges". Nature 495.7441 (2013): 384-388.

Hasty, Paul et al. "Muscle Deficiency And Neonatal Death In Mice With A Targeted Mutation In The Myogenin Gene". Nature 364.6437 (1993): 501-506.

Ivanov, Andranik et al. "Analysis Of Intron Sequences Reveals Hallmarks Of Circular RNA Biogenesis In Animals". Cell Reports 10.2 (2015): 170-177.

Jeck, W. R. et al. "Circular Rnas Are Abundant, Conserved, And Associated With ALU Repeats". RNA 19.2 (2012): 141-157.

Jens, Marvin, and Nikolaus Rajewsky. "Competition Between Target Sites Of Regulators Shapes Post-Transcriptional Gene Regulation". Nat Rev Genet 16.2 (2014): 113-126.

Jinek, Martin et al. "RNA-Programmed Genome Editing In Human Cells". eLife 2 (2013): n. pag.

Karreth, Florian A. et al. "In Vivo Identification Of Tumor-Suppressive PTEN Cernas In An Oncogenic BRAF-Induced Mouse Model Of Melanoma". Cell 147.2 (2011): 382-395.

Kim, H. H. et al. "Hur Recruits Let-7/RISC To Repress C-Myc Expression". Genes & Development 23.15 (2009): 1743-1748.

Kjems, JÃ¸rgen, and Roger A. Garrett. "Novel Splicing Mechanism For The Ribosomal RNA Intron In The Archaebacterium Desulfurococcus Mobilis". Cell 54.5 (1988): 693-703.

Kramer, Marianne C. et al. "Combinatorial Control Of Drosophila Circular RNA Expression By Intronic Repeats, Hnrnps, And SR Proteins". Genes & Development 29.20 (2015): 2168-2182.

Kundu, P. et al. "Hur Protein Attenuates Mirna-Mediated Repression By Promoting Mirisc Dissociation From The Target RNA". Nucleic Acids Research 40.11 (2012): 5088-5100.

Lebedeva, Svetlana et al. "Transcriptome-Wide Analysis Of Regulatory Interactions Of The RNA-Binding Protein Hur". Molecular Cell 43.3 (2011): 340-352.

Legnini, Ivano et al. "A Feedforward Regulatory Loop Between Hur And The Long Noncoding RNA Linc-MD1 Controls Early Phases Of Myogenesis". Molecular Cell 53.3 (2014): 506-514.

Liang, D, and J E Wilusz. "Short Intronic Repeat Sequences Facilitate Circular RNA Production". Genes and Development 28.20 (2014): 2233-2247.

Memczak, Sebastian et al. "Circular Rnas Are A Large Class Of Animal Rnas With Regulatory Potency". Nature 495.7441 (2013): 333-338.

Mousavi, Kambiz et al. "Ernas Promote Transcription By Establishing Chromatin Accessibility At Defined Genomic Loci". Molecular Cell 51.5 (2013): 606-617.

Mukherjee, Neelanjan et al. "Integrative Regulatory Mapping Indicates That The RNA-Binding Protein Hur Couples Pre-Mrna Processing And Mrna Stability". Molecular Cell 43.3 (2011): 327-339.

Nigro, Janice M. et al. "Scrambled Exons". Cell 64.3 (1991): 607-613.

Patapoutian, A et al. "Disruption Of The Mouse MRF4 Gene Identifies Multiple Waves Of Myogenesis In The Myotome". Development 121.10 (1995): 3347-3358.

Poliseno, Laura et al. "A Coding-Independent Function Of Gene And Pseudogene Mrnas Regulates Tumour Biology". Nature 465.7301 (2010): 1033-1038.

Rao, P. K. et al. "Myogenic Factors That Regulate Expression Of Muscle-Specific Micrornas". Proceedings of the National Academy of Sciences 103.23 (2006): 8721-8726.

Rudnicki, Michael A. et al. "Myod Or Myf-5 Is Required For The Formation Of Skeletal Muscle". Cell 75.7 (1993): 1351-1359.

Rybak-Wolf, Agnieszka et al. "Circular Rnas In The Mammalian Brain Are Highly Abundant, Conserved, And Dynamically Expressed". Molecular Cell 58.5 (2015): 870-885.

Salmena, Leonardo et al. "A Cerna Hypothesis: The Rosetta Stone Of A Hidden RNA Language?". Cell 146.3 (2011): 353-358.

Salzman, Julia et al. "Circular Rnas Are The Predominant Transcript Isoform From Hundreds Of Human Genes In Diverse Cell Types". PLoS ONE 7.2 (2012): e30733.

Sanger, H. L. et al. "Viroids Are Single-Stranded Covalently Closed Circular RNA Molecules Existing As Highly Base-Paired

Rod-Like Structures.". Proceedings of the National Academy of Sciences 73.11 (1976): 3852-3856.

Seale, Patrick et al. "Pax7 Is Required For The Specification Of Myogenic Satellite Cells". Cell 102.6 (2000): 777-786.

Starke, Stefan et al. "Exon Circularization Requires Canonical Splice Signals". Cell Reports 10.1 (2015): 103-111.

Suzuki, H. "Characterization Of Rnase R-Digested Cellular RNA Source That Consists Of Lariat And Circular Rnas From Pre-Mrna Splicing". Nucleic Acids Research 34.8 (2006): e63-e63.

Tapscott, S. J. "The Circuitry Of A Master Switch: Myod And The Regulation Of Skeletal Muscle Gene Transcription". Development 132.12 (2005): 2685-2695.

Tay, Yvonne et al. "Coding-Independent Regulation Of The Tumor Suppressor PTEN By Competing Endogenous Mrnas". Cell 147.2 (2011): 344-357. 12 Jan. 2016.

Trapnell, Cole et al. "The Dynamics And Regulators Of Cell Fate Decisions Are Revealed By Pseudotemporal Ordering Of Single Cells". Nat Biotechnol 32.4 (2014): 381-386.

Twayana, Shyam et al. "Biogenesis And Function Of Non-Coding Rnas In Muscle Differentiation And In Duchenne Muscular Dystrophy". Biochm. Soc. Trans. 41.4 (2013): 844-849.

van Rooij, E., and E. N. Olson. "Micrornas Put Their Signatures On The Heart". Physiological Genomics 31.3 (2007): 365-366.

Ven A,, Morten T. et al. "Spatio-Temporal Regulation Of Circular RNA Expression During Porcine Embryonic Brain Development". Genome Biol 16.1 (2015): n. pag.

von Roretz, Christopher et al. "Hur And Myogenesis: Being In The Right Place At The Right Time". Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1813.9 (2011): 1663-1667.

Wang, J. et al. "CREB Up-Regulates Long Non-Coding RNA, HULC Expression Through Interaction With Microrna-372 In Liver Cancer". Nucleic Acids Research 38.16 (2010): 5366-5383.

Westholm, Jakub O. et al. "Genome-Wide Analysis Of Drosophila Circular Rnas Reveals Their Structural And Sequence Properties And Age-Dependent Neural Accumulation". Cell Reports 9.5 (2014): 1966-1980.

Wilson-Rawls, Jeanne et al. "Activated Notch Inhibits Myogenic Activity Of The MADS-Box Transcription Factor Myocyte Enhancer Factor 2C". Molecular and Cellular Biology 19.4 (1999): 2853-2862.

Yaffe, David, and Michael Feldman. "The Formation Of Hybrid Multinucleated Muscle Fibers From Myoblasts Of Different Genetic Origin". Developmental Biology 11.2 (1965): 300-317.

You, Xintian et al. "Neural Circular Rnas Are Derived From Synaptic Genes And Regulated By Development And Plasticity". Nature Neuroscience 18.4 (2015): 603-610.

Zaphiropoulos, PG. "Exon Skipping And Circular RNA Formation In Transcripts Of The Human Cytochrome P-450 2C18 Gene In Epidermis And Of The Rat Androgen Binding Protein Gene In Testis". Molecular Cell Biology 17.6 (1997): 2985-2993.

Zhang, Xiao-Ou et al. "Complementary Sequence-Mediated Exon Circularization". Cell 159.1 (2014): 134-147.

Zhao, Yong, Eva Samal, and Deepak Srivastava. "Serum Response Factor Regulates A Muscle-Specific Microrna That Targets Hand2 During Cardiogenesis". Nature 436.7048 (2005): 214-220.

# List of Publications

1. Martone J, Briganti F, **Legnini I**, Morlando M, Picillo E, Sthandier O, Politano L, Bozzoni I

"Lack of the Celf2a splicing factor is able to convert a Duchenne genotype into a Becker phenotype."

Nature Communications, 2016 (under press)

*This paper describes a case report of a patient with Duchenne Muscular Dystrophy with abnormally mild symptoms. We found that the lack of a protein, called Celf2A, causes such fortunate physiopathological state, as this splicing factor is involved in dystrophin splicing and particularly in exon 45 inclusion. Since the patient has a mutation in exon 44, non-inclusion of exon 45 in the mature transcript due to the lack of Celf2A restores the correct reading frame of the dystrophin mRNA. This may represent a new pharmacological target for treating Duchenne dystrophy patients with deletions of exon 45. I contributed to this work both hands on the bench and with bioinformatic analyses. I mapped the mutation of the patient, studied the promoter of Celf2A, produced and analysed RNAseq data of cells obtained from a muscle biopsy of the patient as well as of both healthy and Duchenne controls in order to characterize the impact of Celf2A absence on the transcriptome.*

2. Hughes JM, **Legnini I**, Salvatori B, Masciarelli S, Marchioni M, Fazi F, Morlando M, Bozzoni I, Fatica A.

"C/EBPα-p30 protein induces expression of the oncogenic long non-coding RNA UCA1 in acute myeloid leukemia."

Oncotarget. 2015 Jul 30;6(21):18534-44.

*In this paper we used erythroid cells in order to study the targets of a transcription factor involved in myeloid differentiation and in leukemya, C/EBPα. We produced stable lines expressing one of the two isoforms of C/EBPα, p42 and p30, the first an important differentiation factor while the second reported to be involved in blood tumours insurgence. We found that p30 activates transcription of UCA1, a long non-coding RNA that is indeed up-regulated in acute myeloid leukemia and directly involved in cell proliferation. In this work I performed bioinformatic analyses of RNA sequencing data of cells expressing p30 and p42 isoforms of C/EBPα, identifying the genes that are specifically activated by p30 such as UCA1.*

3. **Legnini I**, Morlando M, Mangiavacchi A, Fatica A, Bozzoni I.

"A feedforward regulatory loop between HuR and the long noncoding RNA linc-MD1 controls early phases of myogenesis."

Mol Cell. 2014 Feb 6;53(3):506-14. doi: 10.1016/ j.molcel.2013.12.012.

*This paper is fully described in the introduction of this thesis. I designed all of the experiments together with the last author and the co-first author; I performed many of the experiments and wrote the paper together with the last and the co-first authors.*

4. Twayana S, **Legnini I**, Cesana M, Cacchiarelli D, Morlando M, Bozzoni I.

"Biogenesis and function of non-coding RNAs in muscle differentiation and in Duchenne muscular dystrophy."

Biochem Soc Trans. 2013 Aug;41(4):844-9. doi: 10.1042/BST20120353. Review.

*In this review we describe the role of non-coding RNAs, particularly microRNAs and lncRNAs, in muscle differentiation and Duchenne dystrophy. We also report unpublished data about the human linc-MD1 gene. I wrote part of the paper and supervised the experiments performed by the first author.*

5. Cesana M, Cacchiarelli D, **Legnini I**, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I.

"A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA."

Cell. 2011 Oct 14;147(2):358-69. doi: 10.1016/j.cell.2011.09.028.

*This paper is fully described in the introduction chapter of this thesis. I contributed by performing some of the key experiments reported in the work.*

6. Cacchiarelli D, **Legnini I**, Martone J, Cazzella V, D'Amico A, Bertini E, Bozzoni I.

"miRNAs as serum biomarkers for Duchenne muscular dystrophy."

EMBO Mol Med. 2011 May;3(5):258-65. doi: 10.1002/emmm.201100133.

*In this paper we report that some muscle-specific microRNAs can be used as biomarkers of Duchenne Muscular Dystrophy, as they are highly enriched in the serum of Duchenne patients with respect*

*to healthy controls. Moreover, in murine models of Duchenne dystrophy, levels of these microRNAs are reduced after treatment with exon-skipping. I performed some of the key experiments and analysed all the data.*

7. 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. doi: 10.1038/mt.2010.123.

*In this work we studied the efficacy of several molecules in inducing skipping of exon 51 as treatment for certain Duchenne dystrophy mutations. Here I set-up a new approach for measuring dystrophin exon 51 skipping with the use of luciferase reporter minigenes. The coding sequence of the Firefly luciferase is split in two exons and dystrophin exon 51 (or other exons to be tested) is cloned in between. Inclusion of such exon disrupts the luciferase open reading frame resulting into the lack of this enzyme, whose concentration can be easily inferred by measuring its enzymatic activity.*

# Acknowledgments

I thank Irene Bozzoni for mentoring and supervising my work with trust and respect, for providing me with great opportunities of professional growth and always letting me pursue my objectives.

I would like to acknowledge the authors of the manuscripts that I presented in the introduction of this thesis: Marcella Cesana, Davide Cacchiarelli, Tiziana Santini, Olga Sthandier, Mauro Chinappi, Anna Tramontano, Mariangela Morlando, Arianna Mangiavacchi and Alessandro Fatica. Regarding the unpublished results presented in this work, they would not have been achieved without the fundamental contribution of Gaia Di Timoteo, Francesca Rossi, Francesca Briganti, Alessandro Fatica, Mariangela Morlando and Olga Sthandier.

The RNAi screening was analysed with Adrian Andronache and Mark Wade, member and head of the Screening Unit at the Istituto Italiano di Tecnologia in Milan respectively.

I acquired most of the computational and bioinformatic skills used in this work during a short period spent with an EMBO fellowship at the Max Delbrück Center for Molecular Medicine of Berlin, in the lab of Nikolaus Rajewsky, who also gave fundamental advice and help on this work. I thank Marta Rodriguez-Orejuela, Panagiotis Papavasileiou, Christin Stottmeister, Sebastian Memczak and Marvin Jens from the Rajewsky Lab for useful help and discussions about circular RNAs.

I thank Marcella Marchioni, Massimo Arceci and Tiziana Santini for technical advice and assistance.

I thank Fabrizio Loreni and Silvia D'Amico from his lab at the University of Tor Vergata for help with sucrose gradients.

I thank Francesca Meggiorni of the Policlinico Umberto I for providing Rhabdomyosarcoma cell lines.

I also thank all the present and former members of the Bozzoni Lab for sharing results and protocols, for scientific and trivial discussions, humanity and friendship.

Finally, I thank my family and girlfriend, my friends in Rome and out of Rome for giving me the necessary support for surviving a PhD without severe physical and mental consequences.