Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXVI

Settore Concorsuale di afferenza: 05/I1

Settore Scientifico disciplinare: BIO/18

TITOLO TESI

Transcriptional Dynamics of the Human *DMD* Locus

Presentata da: Erriquez Daniela

Coordinatore Dottorato **Relatore**

Vincenzo Scarlato **Giovanni Perini**

Esame finale: Bologna, Aprile 2014

Abstract

The human *DMD* locus encodes dystrophin protein. Absence or reduced levels of dystrophin (DMD or BMD phenotype, respectively) lead to progressive muscle wasting. Little is known about the complex coordination of dystrophin expression and its transcriptional regulation is a field of intense interest. In this work we found that *DMD* locus harbours multiple long non coding RNAs which orchestrate and control transcription of muscle dystrophin mRNA isoforms. These lncRNAs are tissue-specific and highly expressed during myogenesis, suggesting a possible role in tissue-specific expression of *DMD* gene isoforms. Their forced ectopic expression in human muscle and neuronal cells leads to a specific and negative regulation of endogenous dystrophin full lenght isoforms. An intriguing aspect regarding the transcription of the *DMD* locus is the gene size (2.4Mb). The mechanism that ensures the complete synthesis of the primary transcript and the coordinated splicing of 79 exons is still completely unknown. By ChIP-on-chip analyses, we discovered novel regions never been involved before in the transcription regulation of the *DMD* locus. Specifically, we observed enrichments for Pol II, P-Ser2, P-Ser5, Ac-H3 and 2Me-H3K4 in an intronic region of 3Kb (approximately 21Kb) downstream of the end of DMD exon 52 and in a region of 4Kb spanning the DMD exon 62. Interestingly, this latter region and the TSS of Dp71 are strongly marked by 3Me-H3K36, an histone modification associated with the regulation of splicing process. Furthermore, we also observed strong presence of open chromatin marks (Ac-H3 and 2Me-H3K4) around intron 34 and the exon 45 without presence of RNA pol II. We speculate that these two regions may exert an enhancer-like function on Dp427m promoter, although further investigations are necessary. Finally, we investigated the nuclear-cytoplasmic compartmentalization of the muscular dystrophin mRNA and, specifically, we verified whether the exon skipping therapy could influence its cellular distribution.

Introduction

Duchenne and Becker Muscular Dystrophies: an Overview

There are more than 30 different types of inherited dystrophies that are characterized by muscle wasting and weakness of variable distribution and severity, manifesting at any age from birth to middle years, resulting in mild to severe disability and even short life expectancy in the worse cases. Clinical and pathological features are generally the parameters to classify the most common type of Muscular Dystrophies (MDs). The broad spectrum of MDs arises from many different genetic mutations that reflect defects not only in structural proteins, but also in signaling molecules and enzymes. Dystrophin was the first mutant structural protein shown to cause MD. The *DMD* gene codes for dystrophin and mutations in the *DMD* gene range from single-nucleotide changes to chromosomal abnormalities (http://www.dmd.nl/). The two more common type of dystrophy, both defined as X-linked muscle disorder, are the severe Duchenne muscular dystrophy (DMD OMIM 300677) due to out-of-frame mutations, and the milder Becker muscular dystrophy (BMD OMIM 300376) caused by in-frame mutations. Some "exceptions to the reading frame rule" are associated with intermediate phenotypes. These involve both patients with BMD who carry frame-shift deletions/duplications or DMD with in-frame deletions/duplications. Moreover an increasing number of mutation in the *DMD* gene have been shown to result in severe dilated cardiomyopathy with no apparent skeletal muscle pathology. The term X-linked dilated cardiomyopathy (XLDC OMIM #302045) has been assigned to this clinical phenotype (Davies and Nowak, 2006; Ferlini et al., 2013). The incidence of DMD is estimated at 1 in 3500 male newborns with a prevalence of 6 in 100,000 males (Emery, 1991). DMD is characterized by weakness of leg, pelvic and shoulder girdle muscles starting in early childhood. By the early teens, the heart and respiratory muscles also are affected. The average life expectancy for patients afflicted with DMD is around 25. BMD is a milder variant than DMD with a better prognosis. Incidence of BMD is approximately 1 in 18.450 males and prevalence is 2.4 per 100,000 in the general population (Bushby et al., 1991) and the course is slower and less predictable than that of DMD. Its onset is between the age of 3 and 21 years with a mean age of onset at 11 years. Age at death is 21–89 years with an average age of about 45 years (Bushby and Gardner-Medwin, 1993; Hermans et al., 2010).

The majority of female carriers of DMD mutations are asymptomatic, nevertheless, a certain number, defined as "manifesting" or "symptomatic", develop symptoms of the disease, which vary from a mild muscle weakness to a DMD-like clinical course. Despite intensively explored, the

pathogenic mechanism underlying clinical manifestation in DMD female carriers still remains a controversial issue (Brioschi et al., 2012).

The dystrophin is a large protein that acts as elastic bridge between the cell cytoskeleton and the extracellular matrix to stabilize the sarcolemma. Dystrophin belongs to the protein system involved in force transduction in muscle membrane. It is part of a multimeric protein complex called dystrophin–glycoprotein complex (DGC). Via syntrophin, a member of DGC, the enzyme neuronal Nitric Oxide Syntase (nNOS) is also localized to the membrane of muscle fibers and it regulates a sub-set of muscular gene-targets by modulation of histone deacetylases (HDACs) activity. Absence of the dystrophin protein (DMD) and reduced levels or abnormal configuration of dystrophin (BMD) leads to membrane fragility making muscle cells susceptible to damage from contraction (Petrof et al., 1993). Secondary, increase in free radicals and the impairment of nNOS-signalling are thought to further contribute to muscle degeneration. Moreover, as muscle disease advances, muscle repair and regeneration cannot adequately compensate for damage, leading to degeneration and necrosis of skeletal myofibers and cardiomyocites and gradual replacement by fibrofatty tissue (Durbeej and Campbell, 2002; Ervasti and Sonnemann, 2008; Wallace and McNally, 2009).

For this and other reasons, from over a decade, several attempts have been made in order to find a cure for DMD, but the dystrophin gene complexity, its fine regulation, the structural function of the protein as well as the large body muscle mass to be treated, including the heart, have raised many difficulties in offering an efficacious and safe treatment.

Exon Skipping: a Promising Therapeutic Approach for DMD

In the DMD patients, large deletions of one or more exons are the most common mutational event (accounting for the 65% of the dystrophin mutations) that lead to a disrupted reading frame (see Leiden Muscular Dystrophy Pages website), thus precluding the production of the muscular dystrophin protein and leading to progressive muscle weakness, cardiomyopathy and respiratory failure. Deletions are typically clustered in a hot-spot region between exons 45 and 55, and less commonly in a second deletion hot-spot towards the 5′ end of the gene (between exons 3 and 7). At present there is no effective therapy to stop the lethal progression of the disease, although several promising experimental strategies are currently under investigation (Ferlini et al., 2013).

The observation that the milder allelic variant BMD caused by in-frame mutations allows the translation of a smaller but partially functional dystrophin provided the strong rationale for the application of the exon skipping strategy to DMD. The final goal of this therapeutic strategy is overcoming an out-frame mutation in the muscular dystrophin transcript to convert the severe DMD

phenotype into the milder BMD form. Patients with BMD are able to retain ambulation into late adulthood and have a normal lifespan.

Although many different antisense oligonucleotide (AON) chemistries exist, so far, two classes are under clinical experimentation: 2′O-methylphosphorothioate oligoribonucleotides (2′OMe) and phosphorodiamidate morpholino oligomers (PMOs). Both chemistries, by direct sequence-specific steric block via hybridization to pre-mRNA, target specific exons to alter pre-mRNA splicing and causing their skipping during the splicing process. Given that most of the critical functional domains within the dystrophin protein are at the amino and carboxyl terminals and in the cysteine-rich domain, the resulting shortened dystrophin protein product retain still its structural activity, because partly deleted of internal sequences that encode spectrin-like rod repeat domains (Fairclough et al., 2013).

AONs have been extensively and successfully tested *in vitro* (Dunckley et al., 1998) and, more importantly, *in vivo* (Aoki et al., 2012; Goyenvalle et al., 2010; Yokota et al., 2012). Proof of principle has been established in Phase I and Phase II clinical trials led by groups in the Netherlands (Leiden University Medical Centre in collaboration with Prosensa–GlaxoSmithKline) and the United Kingdom (MDEX Consortium in collaboration with Sarepta Therapeutics), targeting *DMD* exon 51, an approach that is applicable to ~13% of patients. The 2′-*O*-methyl-phosphorothioate (2′OMe) and morpholino phosphorodiamidate oligonucleotide (PMO) AONs used in these trials were well-tolerated and restored dystrophin protein to variable degrees (Cirak et al., 2011; Goemans et al., 2011; Kinali et al., 2009; Mercuri and Muntoni, 2013; van Deutekom et al., 2007).

Despite the encouraging results, several hurdles limiting the therapeutic approach have to be overcome, such as the poor cellular uptake of AONs, either when delivered via intramuscular or systemic routes, and their relative rapid clearance from circulation, which means repeated administrations and probably a lifelong treatment. Another important matter to consider is the high variability in exon-skipping efficiency among different muscle types. Furthermore, the use of AON-mediated exon skipping therapy is still limited by the fact that they cannot be utilized for a significant number of DMD patients, in particular those with large deletions or with mutations in regulatory or N-/C-terminal regions of dystrophin (Aartsma-Rus et al., 2009; Benedetti et al., 2013).

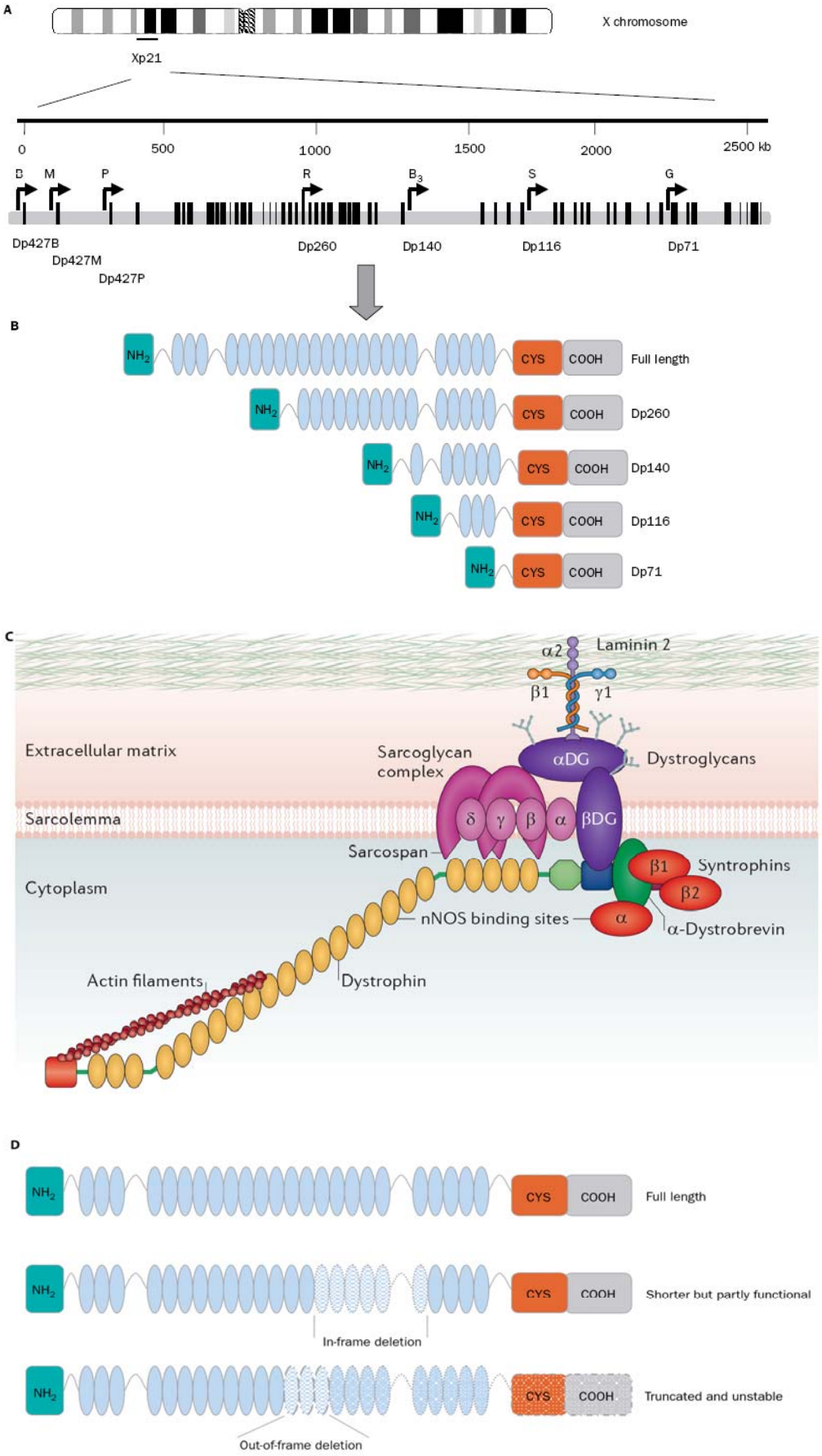


Figure 1. A) Human *DMD* Locus; B) Dystrophin isoforms; C) dystrophin-glycoprotein complex; D) Out-of -frame and In-frame *DMD* mutation (Fairclough et al., 2013; Muntoni et al., 2003)

The Human DMD Locus

The Overall Organization of the Human DMD locus

The gene encoding dystrophin (*DMD*) is the largest gene in the human genome and accounts for approximately 0.1% of the entire human DNA sequence; it is 2.4 Mb long and it lies on the short arm of the X chromosome (Xp21.2). It consists of 79 exons, 78 introns and at least 7 recognized promoters that give rise to 7 different dystrophin isoforms, each displaying a tissue- and temporal-specific pattern of expression. Three main promoters, the ancient Brain (B), the Muscular (M) and the Purkinje (P) promoter, drive the expression of the three full length isoforms, each contain unique first exon, spliced with a common set of 78 exons. The 14kb messenger RNA encode a protein with a molecular weight of 427 kDa. The skeletal and cardiac muscles are the main tissues where the Dp427m isoform is expressed, whereas the Dp427b (also named Dp247c) isoform is predominantly expressed in the brain (hypothalamus and cortex) and also, at low levels, in striated muscles. The Dp427p isoform is expressed in Purkinje cerebellar neurons, and, at very low concentrations, in skeletal muscle. Four internal promoters, named retinal (R), brain-3 (B3), Schwann cell (S), e general (G), give rise to shorter dystrophin proteins. Each of these promoter use a unique first exon (exons 29, 44, 55 and 62) that splices into exons 30, 45, 56, and 63 to generate protein products of 260 kDa (Dp260), 140 kDa (Dp140), 116 kDa (Dp116), and 71 kDa (Dp71), respectively. Dp260 is expressed in high concentrations in the retina, where it coexists with the full-length brain and muscle isoforms (D'Souza et al., 1995; Pillers et al., 1993). Dp140 is expressed in brain, retina, and kidney tissues (Lidov et al., 1995). Dp116 is only expressed in adult peripheral nerves (Byers et al., 1993). Dp71 is detected in most non-muscle tissues including brain, retina, kidney, liver, and lung and is present in cardiac but not in fully differentiated skeletal muscle. It is denoted as ubiquitous dystrophin isoform (Bar et al., 1990; Rapaport et al., 1992; Sadoulet-Puccio and Kunkel, 1996).

In addition to these isoforms, the dystrophin gene produces many isoforms generated through alternative splicing events. These splice variants are formed both through the exclusion of some exons from the primary transcript (exon skipping) and by subversion of the reciprocal order of exons (exon scrambling) (Sadoulet-Puccio and Kunkel, 1996; Surono et al., 1999). These events, which commonly occur in a tissue-specific way, generate further protein diversity and account for the complex expression regulation of the tissue-specific dystrophin functions (Muntoni et al., 2003).

The Introns of DMD Locus

The full sequencing of dystrophin introns has revealed that very large and in some cases huge introns are very common in the architecture of this gene, in particular the introns close to alternative transcription starting sites (as introns 1-Muscle, intron 1-Brain and intron 1-Purkinje, as well as intron 44) . Indeed, the intronic sequences account for more than 99% of the *DMD* locus length and this intriguingly aspect suggest that such genomic portion conserves sequences or structural features performing essential function or containing important regulatory domains. The DMD Leiden pages (http://www.dmd.nl/) report all the known intronic sequences and their relative accession numbers in the HGMP GeneBank. Furthermore, the detailed analysis of the large introns, emerging from the full Genome Project sequencing, have revealed interesting characteristics and focused attention on possible roles and functions they may play in gene regulation, especially transcription and splicing (McNaughton et al., 1998; McNaughton et al., 1997).

Another controversial matter regarding the *DMD* locus is the high occurrence of alternative splicing events occurring in this gene that raised the question why such large introns have been maintained during evolution. Their unusually large size has been asserted as one of the major causes of the high mutation rate known to occur in some regions of the gene, which give rise to two well known mutation hot-spot regions (exons 3-7 and exons 44-53). The genomic breakpoint of such hot-spots frequently lies within introns 2 and 7 or interest predominantly the intron 44 (Muntoni et al., 2003; Nobile et al., 1995). In particular, although the introns 7 and 44 experience the highest recombination rates of the dystrophin gene, they seem to be sites of positive directional selection (Nachman and Crowell, 2000). This support the idea that possible regulatory motifs are contained within. Intron 7 is adjacent to a region of "exceptions to the rule" of the Monaco open-reading-frame theory (Monaco et al., 1988); in fact a restarting dystrophin ATG located within exon 6 has been postulated as mechanism to rescue dystrophin translation in mutations located within exon 2–6 (Gurvich et al., 2009). These mutations cause a BMD phenotype despite of being out-of-frame, pointing out again that this region might be involved in regulatory processes.

To date, the co-transcriptional splicing mechanism of human huge introns is yet a controversial question. In a recent study, it has been proposed a 'nested splicing hypothesis' where many splice site-like sequences within the large intron could play roles in splicing events of nested introns, bringing the distant authentic splice sites into close proximity to facilitate the final splicing (Suzuki et al., 2013).

Notably, introns are the segments involved in the DNA duplication process and duplication forks formation, fact that is mechanistically linked to the gene disruption due to non-allelic homologous

recombination (NAHR), non-homologous end joining (NHEJ), and microhomology-mediated replication-dependent recombination (MMRDR) mechanisms, known to explain DNA rearrangements associated with genomic disorders (Ferlini et al., 2013). In the human dystrophin gene an unusual extended area of DNA association with nuclear matrix was been characterized and defined as an extended DNA loop anchorage region (LAR) spanning approximately 200 Kbp and covering a part of the intron 43, exon 44, and most of large intron 44. The extended LAR identified harbors the major recombination hot-spot of the dystrophin gene and also a replication origin. Because of the mechanisms determining positions of deletion hot-spots in *DMD* locus is not yet understood the author of this paper proposed a model where DNA topoisomerase II-mediated cleavage at the nuclear matrix may enhance recombination events within this extended LAR (Iarovaia et al., 2006).

Taken together, all these points of view enhance the idea that behind the large *DMD* introns are hidden several more information of that we think.

Transcriptional Dynamics of the DMD Locus

The Transcription Rate of DMD Locus

Monitoring the transcript accumulation from four different sites within the gene, it is estimated that RNA polymerase II (Pol II) takes 16 hours to transcribe the entire *DMD* locus at an average elongation rate of 2.4 Kb min^{-1}.The rate of transcript accumulation seems to be reduced at the 3' end of the gene relative to the 5' end. This could be due, at least in part, either to the exceptional length of the dystrophin gene or the nature of internal DNA promoter sequences difficult to transcribe, likely because bound by transcriptional repressors that might interfere with the extraordinary journey of RNA polymerase II along the entire locus.

The mechanism that ensures a coordinated splicing of 79 exons is another important matter regarding the full length dystrophin transcript . Although this aspect is still unclear, lines of evidence showed that both splice site selection and splicing occur in a orderly manner, in a 5' to 3' direction, on the nascent dystrophin transcript and before transcription is complete (co-transcriptional splicing)(Tennyson et al., 1995). Measurements by quantitative RT–PCR indicated that mature dystrophin mRNA is 5–10 per nucleus in adult skeletal muscle tissue. This result is consistent with other studies indicating the dystrophin is a low abundance transcript and is comparable with the level of transcript expressed in myogenic cultures (Tennyson et al., 1996).

The precursor mRNA and mature mRNA are the main target of the most promising molecular therapies aimed to rescue a muscular full-length protein production in DMD patients. In the first case, as said above, the exon skipping approach aims to reframe the DMD transcripts redirecting pre-mRNA splicing of dystrophin. In the second one, the read-through of stop codons strategy uses compounds that force the cell to ignore premature stop codons during translation process. Both the approaches rely on the quantification of muscle dystrophin expression level and, overall, on the stability of the transcript to monitor the therapeutic potential success.

Since DMD transcripts can generally picked up in patient-derived muscles, many efforts have been made to correlate the type of mutation to amount of muscle dystrophin and the related phenotype. Mostly by microarray-analysis, several reports have attempted to quantify the DMD transcripts, concluding that the *DMD* gene is less expressed in DMD patients compared with controls (Chelly et al., 1990). Recently, a very appealing technology, a Custom Micro-Fluidic Exome Array (named FluiDMD), allow to profile which dystrophin isoforms are expressed in a dystrophic sample and establish changes in mRNA decay, among several pathogenic effects caused by mutations in the dystrophin gene. Both features are linked to the transcriptional dynamics of the *DMD* locus (Bovolenta et al., 2012b). However, latest lines of evidence focused the attention on the 5'-3' imbalance of the muscle dystrophin transcript rather than the transcriptional rate, suggesting that this is an important phenomenon that determines the dystrophin protein levels. The authors of this paper demonstrated that mutated transcript in the *mdx* mouse model shows a stronger 5' to 3' imbalance compared with that of its wild-type counterpart and reading frame restoration via antisense-mediated exon skipping does not correct this event. They also report significant transcript instability in human BMD samples, supporting the theory that transcript imbalance is not caused by premature nonsense mutations (Spitali et al., 2013).

Nonetheless, the export of Dp427m transcript into the cytoplasm is another important process that determine the final amount of dystrophin protein in muscle cells. The mRNA can be either committed for export to the cytoplasm or accumulates in the nucleus where it may be degraded. Formation of an export-competent messenger ribonucleoproteins (mRNP) begins at transcription. Progressively, mRNAs are channeled into the specific export pathway coordinately with their processing and assembly into mRNPs. Among the factors bound to the pre-mRNAs are also export adaptors that serve to establish a physical bridge between the mRNA molecule and its export receptors (Carmody and Wente, 2009; Köhler and Hurt, 2007). Importantly, several studies demonstrated that the factors involved in mRNA processing specify the fate of a transcript. Thus, if a transcript is not properly processed, it can be recognized by the nuclear surveillance machinery, retained in the nucleus and degraded by the nuclear exosome. This has been documented in elegant

studies of mutants that are defective in mRNA splicing, export and polyadenylation (Brodsky and Silver, 2000; Carmody and Wente, 2009; Hilleren et al., 2001; Lei and Silver, 2002; Libri et al., 2002; Zenklusen et al., 2002) For these reasons the export process should be taken into account in a dystrophic context. The availability of dystrophin mature transcripts is a key factor to determine the protein abundance and thus will influence the outcome of mRNA-targeting therapies.

The Transcriptional Regulation of Muscle Dystrophin Isoform

In terms of tissue-specific expression, the muscular dystrophin isoform is transcribed mainly in skeletal and cardiac muscle tissue and to a lesser extent in smooth muscle, kidney and brain. It accumulates as normal myoblasts differentiate in multinucleated myotubes (Chelly et al., 1988).
Regulation of *DMD* gene expression in muscle is complex and appear to require several *cis*-regulatory elements and *trans*-acting factors that drive its transcription in a tissue- and temporal-specific pattern.
In early molecular and functional studies about muscle-specific promoter region, some of these *cis*-acting sequences and *trans*-acting factors have been defined as involved in myogenic regulation of *DMD* gene transcription. The upstream region of muscle dystrophin exon 1 was compared to that of other muscle-specific gene promoters. Sequences analysis indicated that in addition to an ATA-rich region (at position -24), thought to bind RNA polymerase II, and the GC box (at position -61), this region contains most relevant conserved domains implicated in the regulation of other muscle-specific genes: a CArG box at -91 bp; myocyte-specific enhancer-binding nuclear factor 1 (MEF-1) binding site homologies at -58, -535, and -583 bp and a muscle-CAAT (MCAT) consensus sequence at -394 bp relative to the cap site. The 850 bp of 5'-flanking the transcriptional start site of muscle dystrophin isoform is capable to drive the *DMD* gene transcription in a cell- and developmental stage-specific manner. More specifically, the 149 bp-promoter fragment upstream of muscle exon 1 contains the *cis*-acting sequences required for a preferential muscle-specific activation of *DMD* transcription (Klamut et al., 1990). Although also a well-conserved putative E-box element has been found at -49 position in the minimal promoter fragment, this motif is not involved in the dystrophin muscle-specific transcription, because it was non-responsive to trans-activation by the known binding myogenic master regulator, MyoD (Myogenic Dyfferentiation 1 factor) (Gilgenkrantz et al., 1992; Weintraub, 1993).
In independent studies the CArG box motif was shown to be an essential functional regulatory element. This element revealed to have different trans-acting factor binding properties depending on the combined activity of several regulatory protein, which determine different transcriptional

outcome. Immediately downstream of CArG box motif, the partially overlapping sequence GAAACC seems necessary for a tissue-specific promoter activity. The first one is recognized by the positive serum response factor (SRF), which in turn requires the activity of the dystrophin promoter bending factor (DPBF) bound to second one. Since SRF alone is not sufficient to drive muscle-specific transcription, it has been postulated that the nuclear factor DPBF functions as an architectural component which induces a conformational change in the dystrophin promoter, likely allowing SRF to interact with other muscle-component of the transcriptional complex (Galvagni et al., 1997). Dystrophin CArG box element is also recognized by another bending factor, YY1, a zinc finger protein that, in a muscular cellular context, acts as a negative regulator of the dystrophin promoter. DPBF and YY1 compete each other to regulate the promoter activity and exert an opposite effect of bending on double helix. These results suggest that an alternative spatial organization of the dystrophin promoter could be relevant for its transcriptional regulation. Moreover, YY1 seems regulate the promoter activity in a developmental specific manner. Indeed, an up-regulation of promoter correlates with a down-regulation of the factor YY1 during muscular cell differentiation (Galvagni et al., 1998).

However, further transgenic mouse studies demonstrated that the muscle dystrophin promoter alone drives the *lacZ* reporter gene expression only in the right ventricle of heart. No promoter activity was reported in mature skeletal muscle, thus providing indirect evidence that additional muscle-specific regulatory control elements are necessary to target the rest of heart as well as the entire skeletal muscle tissue (Kimura et al., 1997).

Functional analysis of a 36 kb region surrounding the muscle transcription start site has identified a muscle-specific enhancer within intron one of the dystrophin gene (DME1; dystrophin muscle enhancer one) positioned 6.5 kb downstream of muscle exon one (Klamut et al., 1990). This 5kb fragment exhibits properties consistent with a muscle-specific transcriptional enhancer since it was shown to be inactive in fibroblasts and functionally independent of position and orientation.

Although this regulatory element was shown to have a positive influence on the transcriptional activity of the dystrophin muscle promoter in both immature and mature skeletal muscle, as well as in transgenic mice engineered to express *lacZ* reporter gene under control of mouse dystrophin muscle promoter/enhancer sequences, it seems to have higher activity in cardiac muscle-derived cells as compared to skeletal muscle-derived cells. To date the exact role of DME1 in the regulation of endogenous dystrophin gene transcription in skeletal and cardiac muscle remains unclear (Bastianutto et al., 2001; Klamut et al., 1996; Marshall et al., 2002).

The intron 1 enhancer activity and the muscle dystrophin promoter are not sufficient to fully explain the transcriptional dynamics of the *DMD* locus in terms of developmental and spatial regulation and

point to the existence of both additional *cis*-acting enhancer elements in other regions of the *DMD* gene and several other *trans*-acting factors/molecules that drive its transcriptional regulation in the skeletal and cardiac muscle compartments. However, since the enhancers can be located up to hundreds of kilobases from the promoters that they control, their identification become challenging. In fact, they are able to establish long range interactions with the promoters of regulated genes and act independently of orientation of transcription.
Without doubt, the general mechanism underlying the fine spatio-temporal transcriptional reglutation of muscle dystrophin needs many further efforts to be deeply understood.

Non coding RNAs in Muscle Biology

Muscle is a dynamic tissue that goes through many recurrent phases of degeneration and regeneration throughout an individual's lifetime. During normal muscle development, specific molecular circuitries and signaling pathways control several events in different cell types such as activation of satellite cell proliferation, progenitor cell maintenance, myoblast differentiation, muscle cell homeostasis and immune cell recruitment. It is therefore not surprising that their deregulation heavily contributes to the degeneration of dystrophic muscles and they are the object of intense research (Marrone and Shcherbata, 2011).
There is a continuous flow of scientific reports that underpin functional links between non-coding RNA molecules (ncRNAs) and skeletal muscle biology, suggesting that these ones can play a crucial function both in physiological muscle development and in pathological muscle disorders.
Transcription of the eukaryotic genome yields only 1–2% of protein coding transcripts and the remainder is classified as non-coding RNAs (ncRNAs). In other words, non-coding RNAs are the main output of the global transcription process, highlighting the idea that such an intense cellular effort cannot be just simple noise. Rather, it is reasonable to speculate that this underscored transcriptome possesses specific vital functions (Carninci et al., 2005; Kapranov et al., 2007; Mattick and Makunin, 2006).
In general, non-coding RNAs are divided into structural and regulatory RNAs. The first ones include ribosomal, transfer, small nuclear and small nucleolar RNAs (rRNAs, tRNAs, snRNAs and snoRNAs, respectively), which have been deeply characterized at the functional level. The second ones are a very broad class of RNAs whose main categorization essentially relies on their length.
Small ncRNAs are defined as transcripts shorter than 200 nucleotides. The most functionally characterized are microRNAs (miRNAs), piwi-interacting RNAs (piRNAs) and small interfering RNAs (siRNAs), which are critical for the assembly and the activity of the RNA

interference machinery. RNAs longer than 200 nucleotides are named long non-coding RNAs (lncRNAs) and are a very heterogeneous group of molecules. Because there is not an official way to classify them, they can be placed in one or more categories depending on their genome localization and/or on their orientation (sense, antisense, bidirectional, intronic or intergeniclncRNAs) (Ponting et al., 2009; Wang et al., 2011).

The emerging studies about this intriguing category of molecules are revealing that ncRNAs are tightly interconnected with the main fundamental aspects of muscular tissue, both in physiological and in pathological contexts, revealing that they are important players in processes such as cellular lineage commitment, growth and differentiation of skeletal muscle. Since muscle differentiation and regeneration are key features that require to be considered when designing novel therapies, addressing the role of ncRNAs in MDs is of high clinical relevance. Furthermore, regulatory RNAs may serve as biomarkers, providing information on disease course, disease severity and response to therapies. Aberrant expression levels of non-coding RNAs can result in novel types of defects that cause remarkable changes in processes such as mRNA maturation, translation, signaling pathways or gene regulation. To date, it is clear that there is involvement of several miRNAs in the muscular dystrophies, on the contrary, very little is known about the role of long ncRNAs (Erriquez et al., 2013).

microRNAs control the stability and/or the translational efficiency of target messenger RNAs, thus causing post-transcriptional gene silencing. Mammalian miRNAs are transcribed as long primary transcripts (pri-miRNAs) and encode one or more miRNAs. Then, they are further processed to yield ~22 bp mature transcripts. miRNAs actively take part in the proliferation and differentiation of skeletal muscle cells as an integral component of genetic regulatory circuitries. miR-1, miR-133a/b and miR-206 are largely studied and defined "muscle-specific" miRNAs (myomiRs). They are regulated in muscular transcriptional networks via myogenic regulatory factors (MRFs) and via others key-regulators of the myogenic program, MEF2 (myocyte enhancer factor 2) and SRFs (serum response factors). Recently, a new regulatory pathway, the mechanistic target of rapamycin (mTOR) signaling was seen to regulate miR-1 expression and was also found responsible for MyoD stability (Liu et al., 2007; Rao et al., 2006; Rosenberg et al., 2006; Sun et al., 2010; Zhao et al., 2005). It is possible to functionally define miR-133 as enhancer of myoblast proliferation while miR-1 and miR-206 as enhancers of skeletal muscle differentiation (Chen et al., 2009; Eisenberg et al., 2009; Ge and Chen, 2011; van Rooij et al., 2008).

Interestingly, many miRNAs are defined as “non-muscle specific” (or also ubiquitously expressed), because essentially they are not exclusively expressed in muscular tissue. It has been demonstrated that they also play key-roles in modulating important pathways involved in the regulation of

muscular metabolism and cellular commitment. Some of these miRNAs counteract the differentiation process since their activity is aimed to positively regulate the proliferation phase during muscular development. In contrast to this set of miRNAs, many other "ubiquitous" miRNAs exert an active role in muscle differentiation through different mechanisms. An up-to date list of miRNAs involved in muscle biology are reported in two recent reviews providing for each miRNA the context in which they were studied and highlighting their muscular pathways/targets (Erriquez et al., 2013; Ge and Chen, 2011).

It is therefore not surprising that their deregulation heavily contributes to the degeneration of dystrophic muscles. For example muscle specific myomiR miR-1 and miR-133 and the ubiquitous miR-29c and miR-30c are down-regulated in *mdx* mice. It is possible to rescue wild-type levels of these miRNAs by treating animals with an exon-skipping approach. The same results are confirmed also in human DMD samples. These findings corroborate the direct correlation between miRNAs levels and dystrophin protein levels. An interesting target of muscle specific miR-133b and miR-206 is Utrophin (Utrn), a Dystrophin protein homolog, involved in a compensatory mechanism in DMD pathology (Basu et al., 2011; Cacchiarelli et al., 2010; Rosenberg et al., 2006). Notably, the non-muscle specific miR-31 exert its repressing activity directly targeting the 3'UTR of dystrophin transcript to regulate muscle terminal differentiation. miR-31, as well as miR-206, has a preferential localization in regenerating myoblasts, and is highly expressed in Duchenne muscles, probably due to an intensified activation of satellite cells. In both human and murine wild-type conditions its expression is detected in early phases of myoblast differentiation, supporting the idea that it contributes to avoid early expression of late differentiation markers. Its de-regulation is thought linked to the delay in the maturation program occurring in the DMD pathological context (Cacchiarelli et al., 2011a; Durbeej and Campbell, 2002).

Although the lncRNAs category is less explored than miRNAs one, in general, the biological relevance of the lncRNAs is supported by the fact that they are regulated during development and involved in almost all levels of gene expression and cellular functions, including chromosomal dosage compensation, chromatin modification, cell cycle regulation, control of imprinting, alternative splicing, intracellular trafficking, cellular differentiation, and reprogramming of stem cells (Li et al., 2013). Recently, lines of evidence linking lncRNAs to muscle are emerging both in physiological and pathological context, as well.

Key features of dystrophic muscle include central nuclei, small regenerating fibers and accumulation of connective tissue and fatty tissue. Muscle differentiation *in vitro* is a useful system to investigate the activity of long non-coding RNAs that show muscular specific pattern of expression. A new regulatory network involving cross-talk between several ncRNAs has been

identified by Cesana and colleagues. Relying on ability of myomiRs to orchestrate muscular proliferation and differentiation, the genomic region of miR-206/-133b has been analyzed in detail. Thus a novel muscle specific long non coding transcript has been identified. Because of its non-coding potential and its activated expression upon myoblast differentiation it was termed linc-MD1. More specifically, linc-MD1 is expressed in newly regenerating fibers and is abundant in dystrophic condition, however no expression is detected in mature differentiated fibers. linc-MD1 is localized in the cytoplasm and is a polyadenylated transcript. Through a series of functional studies it has been defined as competing endogenous RNA (ceRNA). linc-MD1 acts as a natural decoy for miR-133 and -135, thus interfering with miRNA repressing activity on the important targets involved in myogenic differentiation MAML1(Mastermind-like 1) and MEF2, respectively (Cesana et al., 2011).

The deeply studied lncRNA Malat1 (Metastasis associated lung adenocarcinoma transcript 1) is another example belonging to this new class of transcripts and linked to muscle biology. Metastasis associated lung adenocarcinoma transcript 1 (Malat1) is a highly conserved 8.7 kb non-coding transcript that is abundantly expressed in cancer cells and a strong predictor of metastasis . Malat1 has been proposed to regulate alternative splicing (Tripathi et al., 2010; Wilusz et al., 2012), transcriptional activation and the expression of nearby genes (Gutschner et al., 2013). Numerous experimental examples support its functional role in the regulation of cell growth, but the exact mechanism of action of Malat1 in different physiological and pathological conditions still needs to be elucidated. By a microarray data analysis obtained using skeletal muscle of mice (gastrocnemius muscle) treated with recombinant myostatin it was observed that the Malat1 expression levels are significantly decreased. Myostatin is a potent negative regulator of myogenesis that inhibits myoblast proliferation and differentiation (Langley et al., 2002; Zhang et al., 2012). Further expression analysis confirmed a persistent up-regulation of Malat1 during the differentiation of myoblasts into myotubes in C2C12 cells as well as in primary human skeletal muscle cells. Conversely, targeted knockdown of Malat1 using siRNA suppressed myoblast proliferation by arresting cell growth in the G0/G1 phase. These results reveal Malat1 as a novel downstream target of myostatin with a considerable ability to regulate myogenesis. Although Malat1 appears largely dispensable for normal mouse development, it is plausible that Malat1 has a role in the transition from the proliferative phase to differentiation in skeletal myogenesis, as well as in the commitment to muscle differentiation (Eißmann et al., 2012; Watts et al., 2013).

Many other lncRNAs have been discovered but not yet fully characterized, as for example Men ε/β lncRNAs. To date it is known that two long non-coding isoforms (Men ε/β lncRNAs) which are expressed in several human tissues, including muscle, arise from the Multiple Endocrine Neoplasia

I locus (*MEN1*). Experimental lines of evidence show their up-regulation upon differentiation of C2C12 myoblats, although their biological role in muscular development is not yet clear. Men ε (also known as NEAT1) and Men β are transcribed from the same RNA polymerase II promoter and are both retained in the nucleus. Suwoo and colleagues formally demonstrated that Men ε/β transcripts are critical structural/organizational components of paraspeckles, organelles localized in the nucleoplasm close to nuclear speckles, where RNA-binding proteins and *Cat2*-transcribed nuclear RNA (CTN-RNA) are stored. Moreover, large-scale analysis revealed that many other lncRNAs are differentially expressed in C2C12 cells upon myoblast differentiation into myotubes, although their biological functions have not been investigated (Clemson et al., 2009; Hubé et al., 2011; Sasaki et al., 2009).

Among many functions ascribed to lncRNAs, there are examples of lncRNAs modulating the activity of transcriptional activators or co-activators, directly or through the regulation of their sub-cellular localization (Li et al., 2013). Two of these have been seen also in a muscular context. The steroid receptor RNA activator (SRA) RNA is a very peculiar transcript that exists as both non-coding and coding RNA (yielding SRA ncRNA and protein SRAP respectively). The SRA ncRNA is highly expressed in skeletal muscle and works as a co-activator of MYOD transcription factor, a master regulator of skeletal myogenesis. To address the significance of the enigmatic bi-functional property of this transcript, Hube and colleagues performed an exhaustive analysis clarifying the opposite function of non-protein coding SRA versus ORF-containing transcripts. The balance between coding and non-coding SRA isoforms changes during myogenic differentiation in primary human cells. In particular it is shown that an increased expression of SRA ncRNA and a parallel decrease of protein SRAP occurs during myogenic differentiation in healthy muscle satellite cells. This does not happen in cells isolated from DM1 patients (Myotonic dystrophy type 1), probably because of a delay in the differentiation program. Remarkably, only the ncRNA species enhances MYOD transcriptional activity. The protein SRAP prevents this SRA RNA-dependent co-activation through interaction with its RNA counterpart. However how this is achieved is not known (Caretti et al., 2007; Caretti et al., 2006).

Non-coding repressor of NFAT (NRON) is another case of lncRNA that shows a regulatory activity on a transcription factor. NRON is not highly expressed but it has a distinct tissue specific expression. It has been found enriched in placenta, muscle, and lymphoid tissues. NFAT is a transcription factor responsive to local changes in calcium signals. It is essential for the T cell receptor–mediated immune response and plays a critical role in the development of heart and vasculature, musculature, and nervous tissue. The first study about the role of NRON showed that it regulates NFAT's subcellular localization rather than its transcriptional activity. Sharma and

coworkers confirmed these data demonstrating that NRON takes part in a large cytoplasmic RNA-protein complex that acts as a scaffold for NFAT to modulate its nuclear trafficking and thus its response activity (Sharma et al., 2011; Willingham et al., 2005).

DBT-E is an example of lncRNA recently discovered within a pathological muscular context and that evidences how this type of non-coding RNAs might gain a functional role when an altered chromatin status exists. Among the many lncRNAs interacting with chromatin remodeling enzymes the most famous are Xist and HOTAIR, both acting as a negative regulators of gene expression by recruitment of PRC2 (Polycomb Repressive Complex 2) on PcG target genes (Lee, 2009; Rinn et al., 2007). Polycomb (PcG) and Trithorax (TrxG) group proteins antagonistically act in the epigenetic regulation of gene expression. Typically, TrxG counteracts PcG-mediated epigenetic gene silencing. Cabianca *et al*. were the first to discover an lncRNA interacting with the TrxG in the Facioscapulohumeral muscular dystrophy (FSHD, omim 158900). FSHD is an autosomal-dominant disease characterized by progressive wasting of facial, upper arm, and shoulder girdle muscles. In up to 95% of cases, the genetic defect is mapped to the subtelomeric region of chromosome 4q35 containing a macrosatellite tandem array of 3.3 Kb long D4Z4 repeats. FSHD is caused by deletions reducing copy number of D4Z4 below 11 units rather than a classical mutation in a coding-protein gene. D4Z4 deletion is associated with a loss of repressive epigenetic marks that switches from a heterochromatic/close state to a more euchromatic/open conformation of the chromatin structure. A novel long non-coding RNA, named DBT-E is produced selectively in FSHD patients. DBT-E is transcribed from D4Z4 repeats and it is a chromatin-associated lncRNA that coordinates de-repression of genes located in the 4q35 region. DBT-E recruits the Trithorax group protein Ash1L to the FSHD locus driving histone H3 lysine dimethylation and thus chromatin remodeling (Cabianca et al., 2012).

According to these lines of evidence, it would be reasonable to think that the complex regulation of dystrophin expression is carried out by a combined activity of transcription factors and lncRNAs (Erriquez et al., 2013).

Dynamics of Transcription and Co-Transcriptional Processes: an Overview

To transcribe the entire 14 Kb muscle dystrophin mRNA, Pol II acts throughout at least four internal promoter regions where many other regulative factors are bound to. In addition, Pol II elongation can be impeded in response to repressive chromatin structures (Saint-André et al., 2011; Shukla et al., 2011). As a further complication, Pol II elongation does not occur on a barrier free linear template (Larson et al., 2011; Selth et al., 2010). To achieve a efficient elongation,

nucleosomes are displaced in front of elongating Pol II and reformed in its wake in a process that is dependent on elongation rate (Bintu et al., 2011; Kristjuhan and Svejstrup, 2004; Schwabish and Struhl, 2004).

The mechanism to accomplish such remarkable task along the *DMD* locus is still completely unknown. Many open questions are arisen regarding this matter such as the processivity of the enzyme, the biochemical composition of Pol II along with associated auxiliary factors involved in the transcriptional regulation of the *DMD* locus, and how it coordinates mRNA synthesis with the splicing machinery and with the mRNA-export. Additionally, never has been investigated the chromatin context in which Pol II works to transcribe a gene containing 79 exons spread over 2.4 Mb of DNA. It would be very important to define how Pol II interacts with chromatin components and the chromatin regulatory enzymes over a journey 16 hours long.

A plethora of studies put in evidence that the transcription elongation rate and specific chromatin modifications contribute to splice site recognition during transcript processing. Rather than operating independently, these processes are tightly integrated to promote co-transcriptional pre-mRNA splicing. Surprisingly, pre-mRNA splicing reciprocally influence Pol II elongation rate and chromatin structure, as well. The idea that arise is that these mechanisms fine-regulate each other and for this reason they should be examined together to provide a more detailed status on transcriptional regulation of a specific gene of interest (Shukla and Oberdoerffer, 2012).

Chromatin modifications either facilitate or hinder access of the DNA to regulate transcription expression. The basic unit of chromatin consists of an octamer of four core histones (H2A, H2B, H3 and H4) wrapped around 146 bp of DNA. Particularly histones H3 and H4 are subjected to post-translational modifications including methylation and acetylation and many of these characteristic patterns of modification are associated with distinct transcription states. Two important markers of open chromatin, histone H3-acetylation (H3ac) and methylation of lysine 4 on histone H3 (H3K4me), could be fundamental features to be considered within a so huge DNA locus like that *DMD* gene. Deletion mutations either could influence the overall chromatin epigenetic marks of the surrounding regions or remove important chromatin regions that preserve a status of DNA permissive to transcription (Eberharter and Becker, 2002; Yan and Boyd, 2006).

H3K36 methylation is considered a hallmark of transcribed DNA as well, although it has been associated with several significance and its role nowadays is not completely clear.

Cotranscriptional methylation of H3K36 by the SET2 family of methyltransferases has been implicated in control of transcription elongation, alternative splicing, and mRNA export. Several investigations have implicated H3K36me3 in alternative splice site choice. Notably, in a "splicing-affects-chromatin model" it seems that the splicing is necessary for establishment and/or

maintenance of normal patterns of H3K36 trimethylation. Intriguingly, intronless genes show lower levels of H3K36me3 as compared to intron-containing genes, irrespective of their expression status (de Almeida et al., 2011; Kim et al., 2011; Sims and Reinberg, 2009).

Variation in patterns of methylations and acetylation of histone tails reflects and modulates chromatin structure and function. Importantly, the post-translational, covalent modifications (PTMs) of histones are tightly coupled to Pol II and, particularly, to its major carboxil-terminal modifications. Rpb1, the largest Pol II subunit, contains a highly flexible structure at its C-terminus. The carboxil-terminal domain (CTD) consists of tandem repeats with a consensus sequences of 7 amino acids (Tyr-Ser-Pro-Thr-Ser-Pro-Ser). The number of repeats varies from 26 (all consensus) repeats in yeast to 52 (21 consensus and 31 non-consensus) repeats in the mammalian CTD. CTD modification plays a direct role in coupling transcription with co-transcriptional nuclear processes such as chromatin modification, mRNA splicing and mRNA export. CTD modification undergoes changes during transcription to recruit factors at the appropriate point of transcription cycle. Dynamic phosphorylation of serine residues on CTD heptad repeats is associated with the stages of Pol II elongation. Phospho-Ser 5 peaks near promoters and declines when the Pol II moves along the gene. In contrast, Phospho-Ser 2 gradually increases with the distance from the promoter.

The carboxil-terminal tail of Pol II is thought as a platform that ensures the correct processing factors recruitment at the appropriate phase of transcription process to coordinate the chromatin remodeling during mRNA synthesis as well as the processing of the nascent transcript. The "CTD code" guarantees the specificity of interactions between the factors involved in all these events, as by direct than by indirect mechanisms (Egloff et al., 2012).

It was demonstrated that phosphorylation of Pol II CTD on serine 5 results in recruitment of the yeast histone methyltransferase Set1, which subsequently directs methylation of H3K4 (Ng et al., 2003). Trimethylation of H3K4 peaks at promoters, whereas dimethylation extends into the 5' region of coding regions and monomethylation persists throughout the gene. The pattern of H3K36 methylation shows a reverse gradient relative to K4, wherein trimethylation increases towards the 3' ends of genes (Barski et al., 2007; Heintzman et al., 2007). According to these lines of evidence about the distribution of chromatin markers, the histone 3 K36 methyltransferase, Set2, binds to Pol II phosphorylated on serine 2 (Li et al., 2003).

The Neugebauer laboratory isolated chromatin associated RNA to show that Pol II pauses within terminal exons allowing sufficient time for intron excision prior to transcript release (Carrillo Oesterreich et al., 2010). Similarly, the Beggs laboratory utilized a high-resolution splicing reporter system to demonstrate splicing dependent Pol II pausing at the 3' ends of introns coincident with splicing factor recruitment (Alexander et al., 2010). Based on these experimental data, intriguingly,

both the H3K36me3 and phospho Ser-2 marks seem linked to a kinetic regulation of co-transcriptional splicing, at least in yeast. Furthermore, these studies raise the question whether Pol II pausing represents a splicing "checkpoint" and, since H3K36me3 mark is strongly associated with splicing, whether H3K36me3 is correlated to this phenomena.

Results

DMD-GEx Microarray Data Analysis

To explore the possibility that long non coding RNAs might be transcribed within *DMD* locus the entire gene was been interrogated in both orientation (sense and antisense) through the design of two novel customised DMD specific gene-expression tiling array (DMD-Gex).

Two independent hybridisation experiments were performed with both the DMD-GEx Sense and Antisense microarrays, using poly A+ RNA from human normal brain, heart, skeletal muscle and skin. To attest the reliability of the hybridization reaction a control of probe set (Table S1) were assayed as well and the chosen control genes displayed an expression profile in agreement with the expected orientation and tissue distribution.

Data were normalized in accordance with the Agilent Quality Controls probes (Spike-in) (see Methods and as reported in the link http://www.chem.agilent.com/library/usermanuals/public/g4140-90041_one-color_tecan.pdf).

Values higher than the fluorescence intensity corresponding to 90% of all ranked probes (90th percentile) on the array were considered as positive hybridization signals. Analysis of both DMD-GEx Sense and Antisense arrays data allowed to identify as statistically significant a total of 14 poly-adenylated transcripts which were named according to their intron of origin and orientation relative to *DMD* gene transcriptional direction (Figure 1B**)**.

As it will be described below none of them contained ORFs encoding protein longer than 100 amino acids and thereby they were all considered non-coding RNAs (ncRNAs). Twelve of these ncRNAs, originating from introns 1 M (2 transcripts), 1P, 2, 29, 32, 37, 44 (2 transcripts), 51, 55 and 67, were found to be transcribed in the same orientation as the *DMD* gene. One ncRNA was found to correspond to the terminal exon and the 3'UTR of the Dp40 isoform (NCBI: NM_004019.2), a known coding dystrophin isoform. The remaining two ncRNAs, originating from intron 55 (ncINT55as) and from 3' UTR (nc3UTRas), were found to be transcribed in antisense orientation to the *DMD* gene. A significant proportion of the identified ncRNAs arose from introns harboring dystrophin isoform promoters or flanking isoform-specific first exon. In particular, three ncRNAs (ncINT1Ms, ncINT1Ms2 and ncINT1Ps) originated from intron 1 of the Dp427m and Dp427p full-length isoforms, one (ncINT29s) from intron 29 (Dp260), two (ncINT44s and ncINT44s2) from intron 44 (Dp140) and two (ncINT55s and ncINT55as) from intron 55 (Dp116). All identified transcripts were strongly expressed in at least one of the three tissues known to

express large amounts of dystrophin. A significant proportion of them were shown to be present in either the skeletal muscle (85.71%) or the heart (71.43%), and nine transcripts were identified in both tissues. Five ncRNAs were found in brain tissue, one of which (ncINT37s) was uniquely expressed in this district (Table 1) (Bovolenta et al., 2012a).

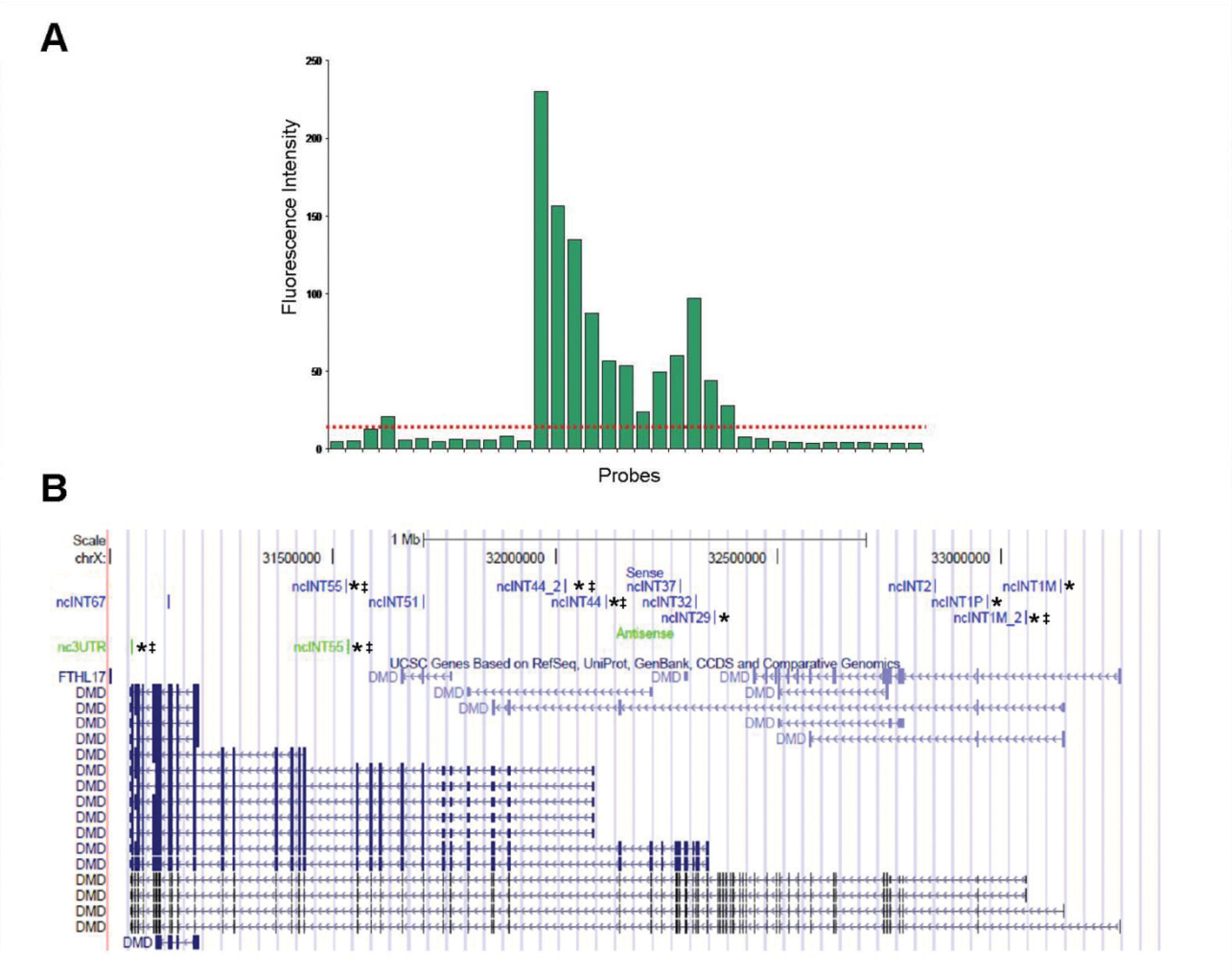


Figure 1. Example of transcripts (ncINT1Ms2) identified by the DMD-GEx array. A) a series of consecutive probes in the genome with fluorescence intensities that rank above the 90th percentile over all probes on the array (indicated with a dashed red line) and mapping within intron 1 M (chrX:33057364-33059742). B) location of the transcripts identified from poly A+ RNA with respect to the DMD gene isoforms. Sense transcripts are represented by blue bars, whereas antisense transcripts are indicated by green bars. The transcripts marked with an asterisk and a double cross were characterized by Northern blotting, and RACE PCR. (doi:10.1371/journal.pone.0045328.g001)

Transcript Validation by Northern Blot Analysis

To confirm and determine in a more detailed manner the pattern of expression of the identified transcripts as well as their precise size and relative amount, northern blotting analyses were performed. ncRNAs whose location within the *DMD* gene suggested a possible regulatory role in the expression of the many *DMD* gene isoforms were preferentially chosen. Nine poly A+ sense

ncRNAs mapping in proximity to the *DMD* gene promoters or nearby isoform first exons, and two antisense transcripts, mapping at the 3' end of the *DMD* locus were selected.

The probes identifying transcripts ncINT1Ms1, ncINT1Ps and ncINT29s revealed complex patterns of hybridisation, consisting of multiple bands or a smeared signal, in several tissue. The difficulty to obtain defined and clear bands is the reason for which we decided to not longer study these ncRNAs.

ncRNA	hg 18	hg 19	size	Heart	Brain	SkM	Skin
ncINT1Ms	chrX:33131738-33131978	chrX:33221817-33222057	241 bp	-	-	√	-
ncINT1Ms2	chrX:33057364-33059742	chrX:33147443-33149821	2379 bp	√	-	√	-
ncINT1Ps	chrX:32969451-32969742	chrX:33059530-33059821	292 bp	√	√	√	√
ncINT2s	chrX:32853214-32853517	chrX:32943293-32943596	304 bp	-	-	√	-
ncINT29s	chrX:32357565-32357851	chrX:32447644-32447930	287 bp	√	-	√	-
ncINT32s	chrX:32316388-32317177	chrX:32406467-32407256	790 bp	-	-	√	-
ncINT37s	chrX:32282002-32282905	chrX:32372081-32372984	904 bp	-	√	-	-
ncINT44s	chrX:32112871-32115470	chrX:32202950-32205549	2600 bp	√	-	√	-
ncINT44S2	chrX:32019551-32022268	chrX:32109630-32112347	2718 bp	√	-	√	-
ncINT51s	chrX:31701831-31702235	chrX:31791910-31792314	405 bp	√	-	√	-
ncINT55as	chrX:31492064-31533929	chrX:31582143-31624008	41866 bp	√	-	-	-
ncINT55s	chrX:31528840-31531268	chrX:31618919-31621347	2429 bp	√	√	√	-
ncINT67s	chrX:31129547-31129738	chrX:31219626-31219817	192 bp	√	√	√	√
nc3UTRas	chrX:31047530-31049401	chrX:31137609-31139480	1872 bp	√	√	√	√
			Total	10	5	12	3
			Percentage	71.4	35.7	85.7	21.4

Table 1: Genomic location, length and tissue representation of the human transcripts identified. (doi:10.1371/journal.pone.0045328.t002)

On the contrary, Northern blotting results obtained for other four sense transcripts (ncINT1Ms2, ncINT44s, ncINT44s2 and ncINT55s) and two antisense transcripts (ncINT55as and nc3UTRas) revealed several distinct transcripts ranging from 1.4 to 4 kb in length and confirm their expression in at least one of the three tissues in which dystrophin is normally expressed (skeletal muscle (SkM), heart and brain) reflecting that observed in the array (Figure 2). (ncINT1Ms2, ncINT44s, ncINT44s2 and ncINT55as: SkM and heart; ncINT55s and nc3UTRas: SkM, heart and brain).

Notably, the ncINT1Ms2, ncINT44s2, ncINT55as and nc3UTRas were highly represented in the heart tissue, whereas the ncINT44s and ncINT55s were found to be more intensively transcribed in skeletal muscle. Furthermore, most transcripts were also detected in the liver, placenta and/or kidney (Figure 2). Different tissue-specific isoforms were detected in four transcripts: ncINT1Ms2 presented at least four isoforms, ranging from about 1.8 to 5 Kb in size. A 2.4 Kb form is prevalent in the heart and liver, whereas it is poorly represented in the skeletal muscle and kidney; two ncINT44s2 isoforms were found in the skeletal muscle tissue, while only the larger one (about 2.7 kb) was detectable in the heart; ncINT55s was detectable in the heart and SkM in at least three

different isoforms (from 2.4 to 3 Kb), showing a specular pattern of expression in the two tissues. The larger of these isoforms was also found to be present in the kidney and placenta. An additional isoform of about 5 Kb is present exclusively in the brain. nc3UTRas show at least two isoforms (1.5 kb and 1.8 Kb). The larger form is widely represented in all tissues analysed, with the exception of the thymus and colon while both isoforms show a preferential expression in placenta, liver and heart (Figure 2). Only one form of transcript has been detected for ncINT44s and ncINT55as. ncINT44s is roughly 2.4 Kb in size and is expressed almost exclusively in the heart and skeletal muscle, being particularly abundant in the latter. In contrast, the ncINT55as transcript is about 2.4 Kb in length and displayed widespread distribution, being present in all analysed tissues except the brain and muscle (Figure 2) (Bovolenta et al., 2012a).

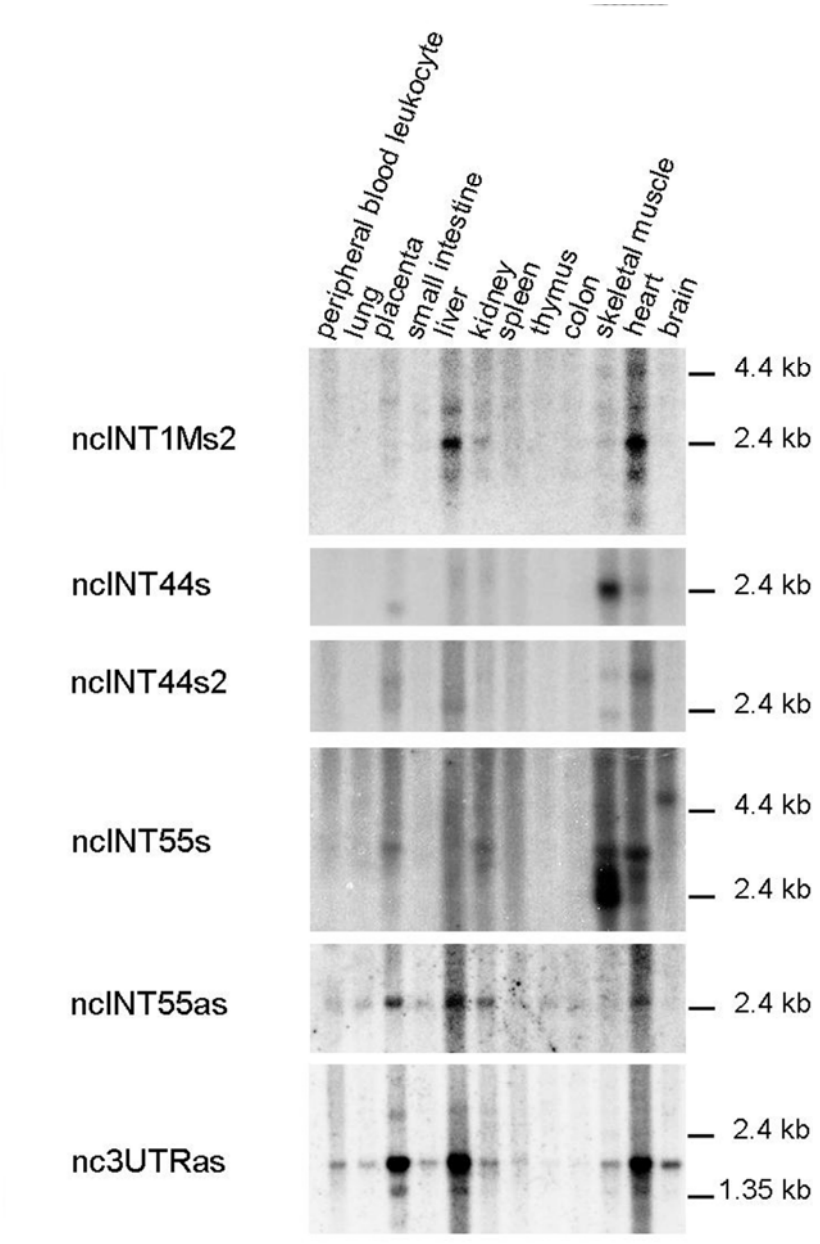


Figure 2. Northern blotting analyses on a 12-lane human poly A+ RNA filter using probes designed on ncRNAs originating near the first exons of DMD isoforms and antisense transcripts. Transcript ncINT1Ms2 is located near full-length DMD gene isoform Dp427p, whereas ncINT44s and

ncINT44s2 surround isoform Dp140. Transcripts ncINT55s and ncINT55as are located upstream the Dp116 isoform. Nc3UTRas overlaps with 39UTR in antisense direction with respect to the DMD gene. All transcripts were expressed in at least one tissue in which DMD isoforms were also expressed, but also in the liver, kidney, spleen and placenta. ncINT1Ms2, ncINT44s2, ncINT55s and nc3UTRas were found to be expressed in multiple isoforms, while one single isoform was detected for ncINT44s and ncINT55as.(doi:10.1371/journal.pone.0045328.g002)

Full-length Transcript Characterization by RACE PCR

To precisely define the size of the six DMD ncRNAs validated by Northern blotting, 3' and 5'-RACE analyses were performed on poly A+ RNA.

ncINT1Ms2: the products obtained from 3'-RACE, transcript walking and 5'-RACE were sequenced and combined, generating a full-length sequence of 2379 bp. The ncRNA was sense-transcribed with dystrophin mRNA and was found to be colinear with the dystrophin intron 1 sequence (Figure 3 A, B and C). The 3' end of the ncINT1Ms2 transcript is located 897 bp upstream of the 5'UTR Dp427p full-length dystrophin isoform.

ncINT44s and ncINT44s2: for ncINT44s, analysis of transcript walking, 3'- and 5'-RACE products yielded a full-length 2.6 Kb transcript entirely transcribed from intron 44 that did not undergo splicing (Figure 3 B and C). For ncINT44s2, the product obtained from 3'-RACE was combined with the two products identified by 5'-RACE, which defined two alternative transcriptional start sites (2452-bp and 2718-bp sequences) corresponding to the two full-length isoforms found in the Northern blotting experiment. Sequencing analysis revealed that these RNAs were entirely transcribed from intron 44, without undergoing splicing events (Figure 3 B and C). The two ncRNAs identified within intron 44 are located 29 Kb upstream (ncINT44s) and 61 Kb downstream (ncINT44s2) of the Dp140 dystrophin isoform promoter.

ncINT55s and ncINT55as: for ncINT55s, 3'-RACE identified two different 3' ends. After transcript walking, 5'-RACE gave rise to two distinct PCR products corresponding to different transcriptional start sites (Figure 3 A). Nucleotide sequence analysis revealed that these transcripts originated entirely from intron 55, without splicing events (Figure 3 A, B and C). Combination of these start and polyadenylation sites potentially generates at least four transcripts of 2805, 2720, 2513 and 2428 bp, which is consistent with the Northern blotting data. The ncINT55as is transcribed in an antisense direction with respect to transcription of the dystrophin gene. 3'-RACE identified two different 3' ends, and 5'-RACE detected three isoforms with a common 5' end (Figure 3 A). Interestingly, these transcripts are alternatively spliced, generating RNAs of different sizes. A shorter isoform carrying only the common 3' and 5' ends was also identified (Figure 3 B and C). The predicted molecular weight of the three different isoforms ranges from 2642 bp to 2732 bp, being consistent with a single product detectable by Northern Blotting analysis. Notably, ncINT55s

residues within that DNA region is a part of the largest "intron" of the mature ncINT55as RNA isoforms. All of these different transcripts identified within intron 55 are located roughly 55 Kb upstream of the Dp116 promoter.

nc3UTRas: sequence analysis of the 3' and 5'-RACE products gave a full-length 1872-bp transcript transcribed in an antisense direction from the 3'UTR region of the *DMD* gene. Since splicing events were not observed (Figure 3 A, B and C), this transcript overlaps (in an antisense orientation) roughly 2 Kb of the 3'UTR region of all full-length and 3'UTR region of all shorter dystrophin isoforms (Bovolenta et al., 2012a).

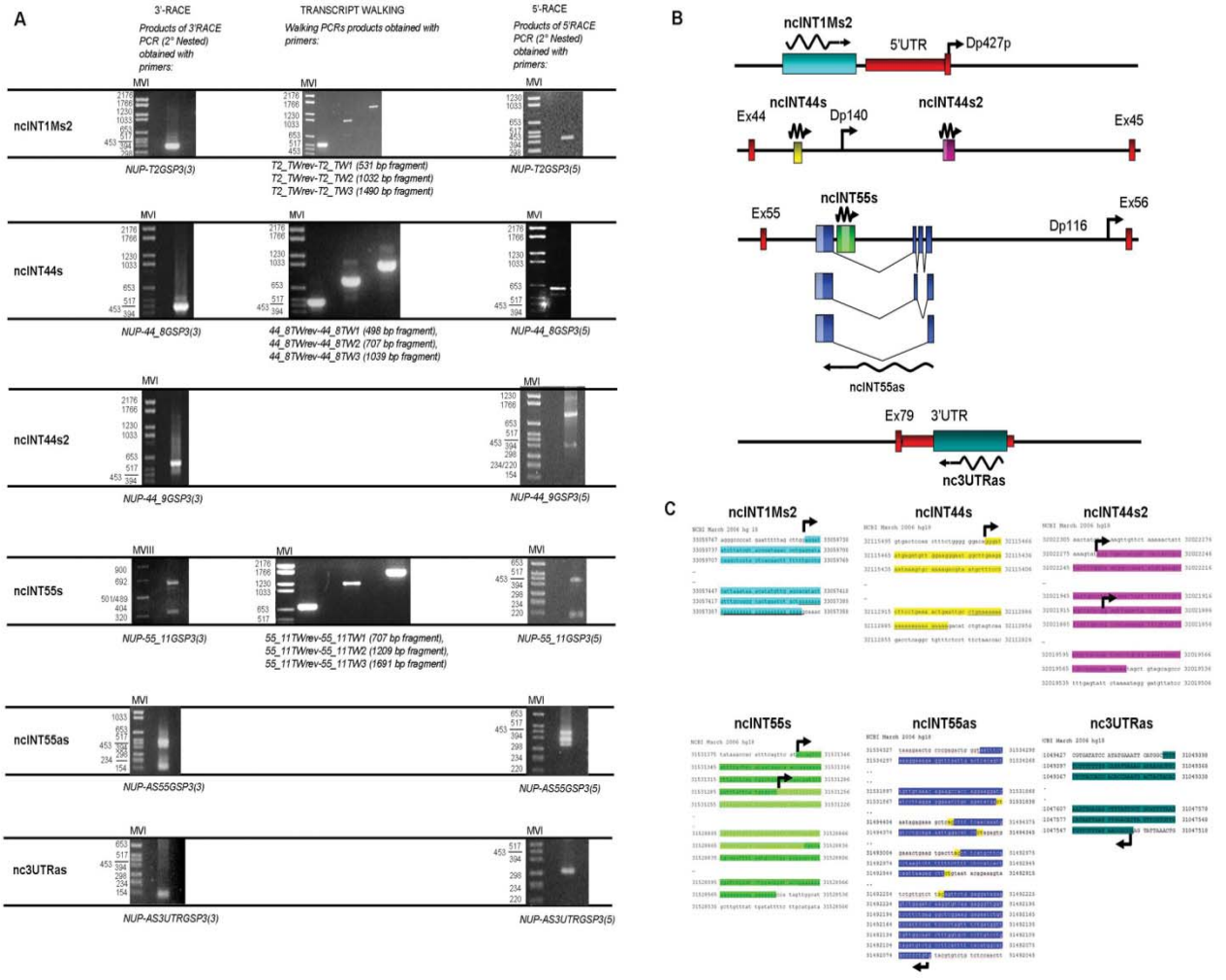


Figure 3. Full-length transcript characterisation by RACE PCR. A) For transcript ncINT1Ms2, a single product of 370 bp was obtained with 3'-RACE. After transcript walking, 5'-RACE was performed and identified a 396-bp PCR product. ncINT44s 3'-RACE generated a single product of 423 bp. After transcript walking, PCR, 5'-end was identified by 5'-RACE, and gave rise to a 599-bp PCR product. For transcript ncINT44s2, a single product of 552 bp was obtained by performing 3'-RACE, whereas 5'-RACE originated two distinct PCR products of 328 bp and 689 bp, corresponding to different transcriptional start sites. 3'-RACE of transcript ncINT55s revealed two products, of 150 bp and 442 bp, corresponding to different 3' ends of the transcript whereas 5'-RACE gave rise to two distinct PCR products of 673 bp and 352 bp. NcINT55as analysis by 3'-RACE identified two PCR products of 99 and 431 bp, corresponding to different 3' ends. 5'-RACE identified three PCR products of 307, 343 and 398 bp. Both 3' and 5'-RACE PCR products for transcripts nc3UTRas showed unique bands of 99 bp and 286 bp, respectively. Relevant pairs of primers used

are listed below each image. B) Schematic representations of the RACE PCR results for transcripts ncINT1Ms2, ncINT44s, ncINT44s2, ncINT55s, ncINT55as and nc3UTRas. Transcript orientation is indicated by the zigzag arrows. Their position with respect to adjacent dystrophin exons and isoform promoters is shown. DMD gene exons are shown as vertical red boxes, and 5' and 3' UTRs as horizontal red boxes. Lighter colours within the transcripts ncINT44s2, ncINT55s and ncINT55as boxes represent alternative starting or polyadenylation sites. The three alternative spliced isoforms of transcript ncINT55as are represented. C) Sequences of the identified transcripts are shown with reference to human genome build 18 (hg18, March 2006). Donor and acceptor splice sites are highlighted in yellow for transcript ncINT55as. Polyadenylation sites are underlined in the sequences. Curved arrows show the starting site for transcription of non-coding transcripts. (doi:10.1371/journal.pone.0045328.g003)

Bioinformatics Analysis of Coding Potential and Secondary Structure

Using bioinformatic tools the six novel transcripts were tested for coding potential, ability to fold into secondary structures and for their degree of conservation.

To confirm that the ncRNAs did not possess the potential to encode for polypeptides, ORF-Finder software was employed to predict putative ORFs with a minimal length of 100 amino acids. As shown in Figure 4A, none of the three sense-frames contained any ORF fitting those criteria. The same results were obtained using the CPC tool, which identified a weak coding potential (one ORF of 51 amino acids) for the ncINT1Ms2 (Figure S2). CPC also analysed homology with UTR regions, and found 37 and 35 hits for ncINT1Ms2 and ncINT55as, respectively.

Using Mfold software, we also determined whether these transcripts could fold into higher secondary structures typical of non-coding RNAs. Interestingly all six ncRNAs can fold into secondary structures with a ΔG below -50 Kcals (Figure 4 B), thereby suggesting that these structure may be extremely stable and likely to serve as domains for interaction with other cellular components.

Conservation analysis performed using the UCSC Genome Browser (UCSC Genome Browser website. Available: http://genome.ucsc.edu/) showed that all identified transcripts are poorly conserved, with the exception of nc3UTRas, which is transcribed from the 3'UTR of the *DMD* gene, and transcript ncINT1Ms2, which presents a highly conserved sequence of roughly 130 bp. The UCSC Genome Browser was also employed to check for the presence of any known ESTs or mRNAs overlapping DMD ncRNAs. For some of the ncRNAs identified, predominantly small tags intersecting their sequences were detected: ncINT1Ms was found to overlap with part of a wider mRNA found in the uterus (Ota et al., 2004); ncINT1Ms2 contained two smaller ESTs found by Robertson et al. (Robertson et al., 1994) in 16-22-week foetal cochleas; ncINT1P overlapped several ESTs previously described (NCBI-CGAP website. Availble: http://www.ncbi.nlm.nih.gov/ncicgap.; NIH-MGC website. Available: http://mgc.nci.nih.gov/.) and corresponds to the TBCAP1 pseudogene (ENST00000436520); ncINT32s was found to contain a 9 bp *Gallus Gallus* DMD exon; and ncINT44s and ncINT55s, respectively, overlapped two smaller

ESTs identified in neuroblastoma tissues (Ohira et al., 2003). No tRNAs were detected within the sequences of the ncRNAs identified (data not shown) (Bovolenta et al., 2012a).

ncRNA Expression Profiles Mirror those of Full-length Muscle and Brain Dystrophin Isoforms

To understand whether the ncRNA expression might be functionally associated with that of dystrophin, we compared the expression of the validated ncRNAs and dystrophin isoforms in six different human cell lines. Three cell lines HEK-293, HeLa and p493 were chosen for the absence of expression of the muscle and brain dystrophin isoforms, although they express low levels of the Purkinje isoform and moderate-good levels of the Dp71 ubiquitous isoform, (Fig. 5A left). In contrast, the SHSY5Y neuroblastoma cell line and, in particular, the two rhabdomyosarcoma cell lines (RD and SJCRH30) were chosen for their good expression of the muscle and brain dystrophin isoforms (Fig. 5A, right). Expression of five ncRNAs was monitored in the cell lines by qRT-PCR. Instead, nc3UTRas was not included in this part of the study since analysis could have been complicated by the fact that it overlaps the 3'UTR region of all full-length and all shorter 3'UTR dystrophin isoforms. Expression of each ncRNA was calculated as the ratio between expression of the ncRNA in a given cell line and that in HeLa cells, which were found to have the lowest level of transcription of all tested ncRNAs. As shown in Fig. 5B, expression of the five ncRNAs is several orders of magnitude higher in cells displaying good levels of full-length muscle and brain dystrophin isoforms than in the other cell lines, thereby suggesting that ncRNAs and dystrophin expression may follow similar dynamics. To further support this idea, we determined the expression levels of our ncRNAs in human normal fibroblasts and human fibroblasts previously transformed with MyoD, a well recognised master transcription regulator that can induce fibroblasts to undergo myogenesis. Results show that MyoD-transformed fibroblasts present strong expression of all tested ncRNAs, as compared to untransformed fibroblasts, which parallels the expression of full-length muscle and brain DMD isoforms (Bovolenta et al., 2012a).

DMD lncRNAs are Confined to the Nuclear Compartment

To further explore whether analysed ncRNAs may be functionally linked to the expression dynamics of the *DMD* locus, their intracellular distribution was determined. GAPDH mRNA was used as a fractionation control, since it is known to be restricted to the cytoplasm. Results show that

all tested ncRNAs are essentially nuclear (Fig. 6), suggesting that expression of the ncRNAs may be important for the transcriptional architecture of the *DMD* locus (Bovolenta et al., 2012a).

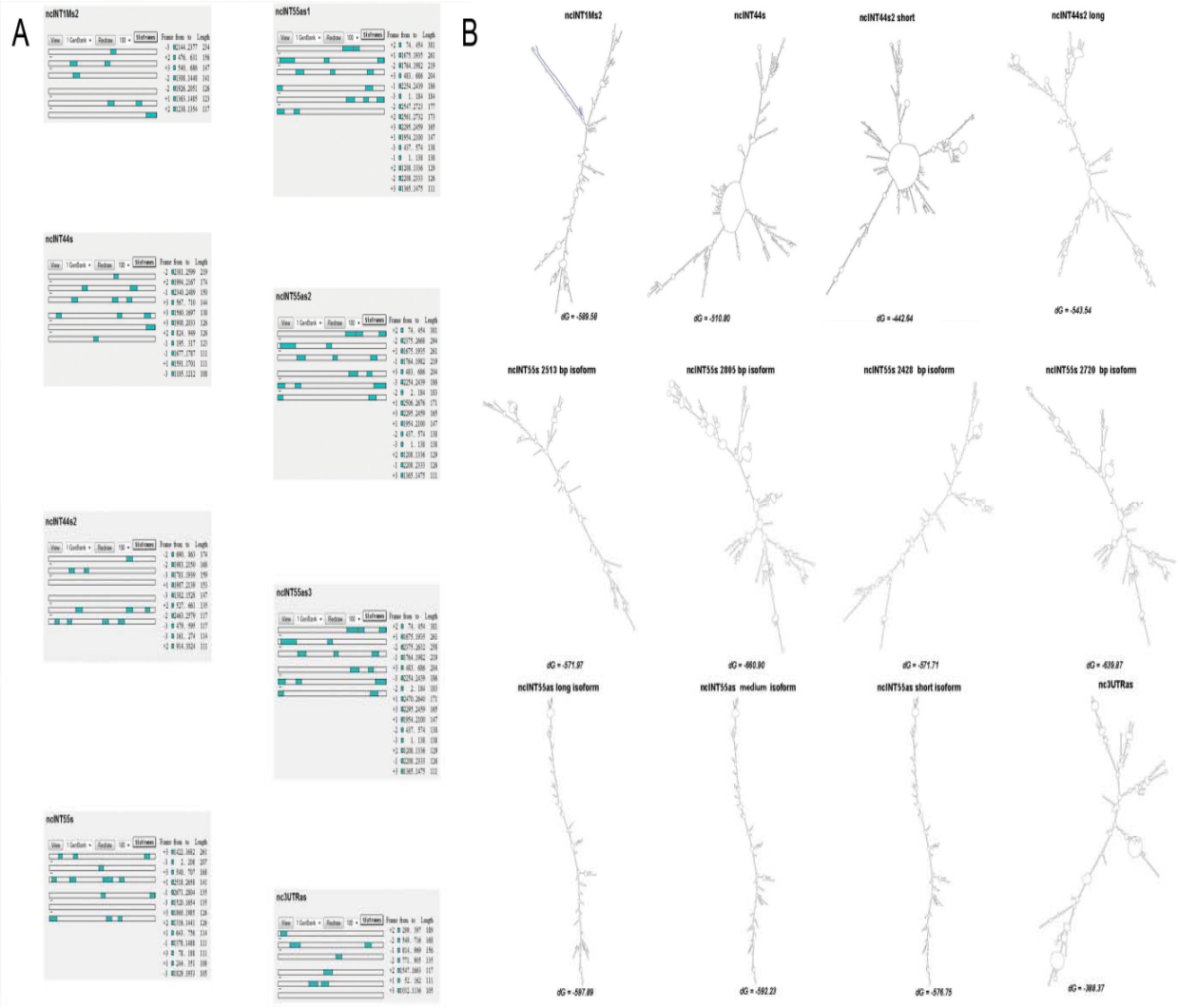


Figure 4. Coding potential and secondary structure bioinformatics analysis. A) ORF prediction. The on-line tool ORF Finder was used to detect open reading frames in our lncRNAs. ORFs identified by the programme are shown in green. Each grey line represents the reading frame, divided into sense (first three lines) and antisense (fourth to sixth lines) strands. The length, reading frame, nucleotide start and end position of each ORF are listed on the right of each transcript; B) Mfold analysis of the secondary structure of the lncRNAs characterised: Mfold software showed that all six lncRNAs were able fold into secondary structures with negative DG below 250 kcals, indicating the great stability of these structures.(doi:10.1371/journal.pone.0045328.g004)

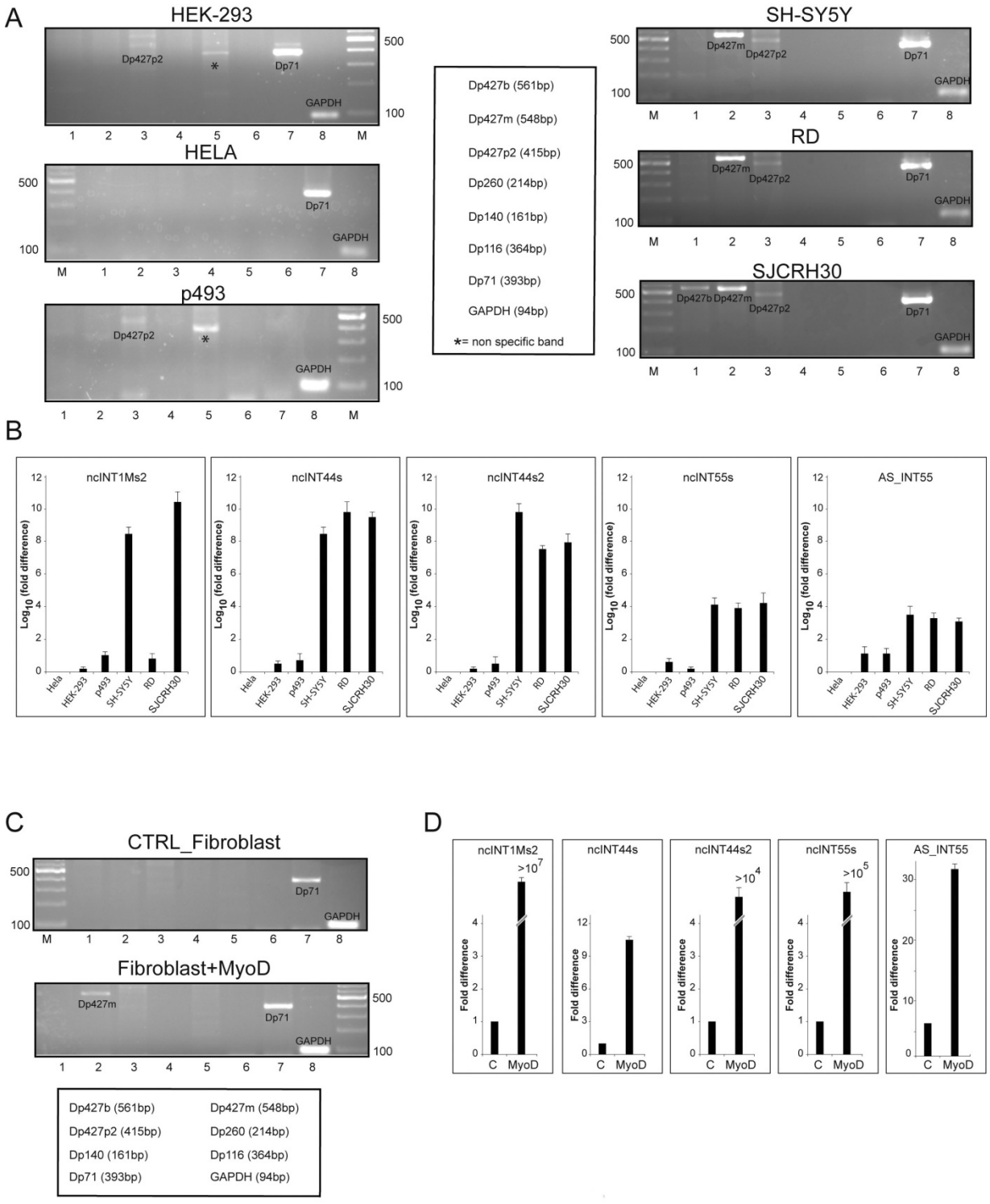
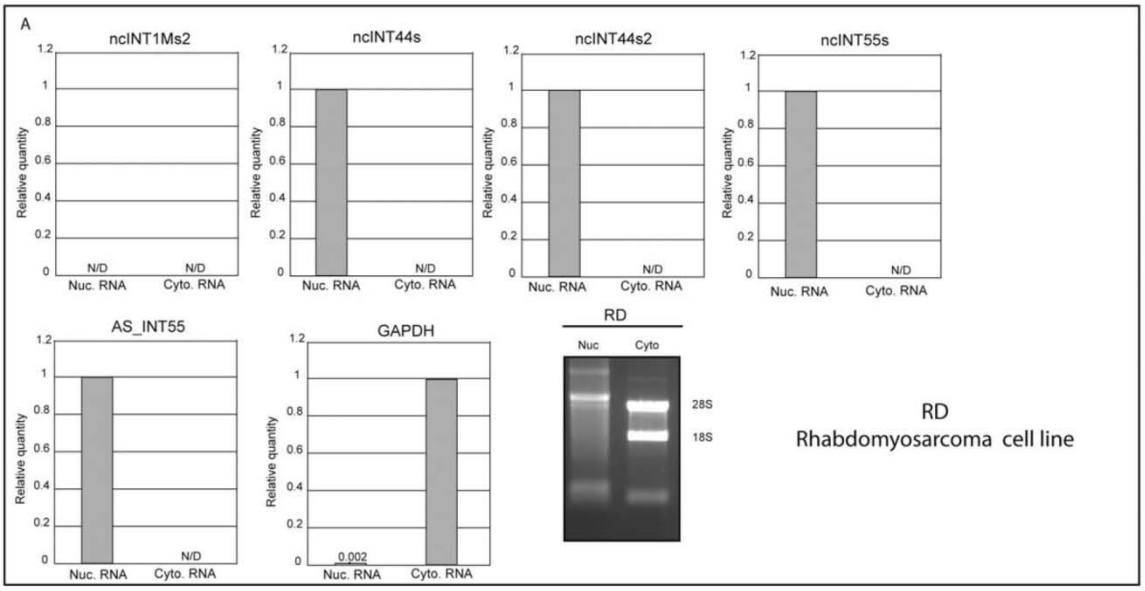
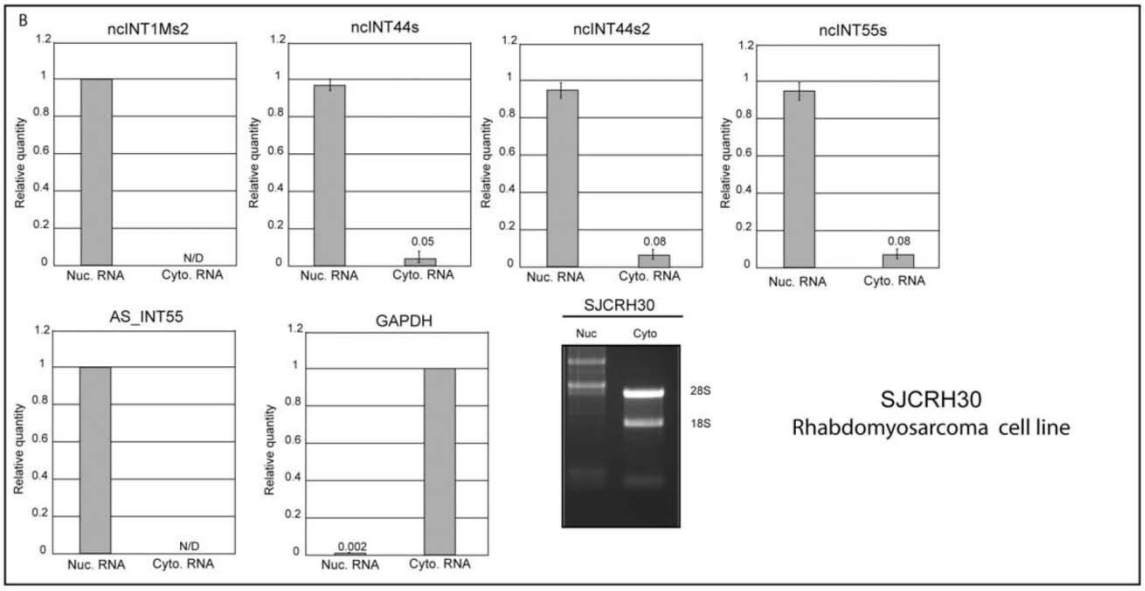


Figure 5. Expression of lncRNA correlates with that of full-length dystrophin isoforms. A) Standard RT-PCR of various dystrophin isoforms in six different human cell lines: HEK-293 (embryo kidney), HeLa (epithelial cervical cancer) p493 (lymphoblastoid), SH-SY5Y (neuroblastoma), RD (rhabdomyosarcoma) SJCRH30 (rhabdomyosarcoma), B) Expression of ncRNA in different cell lines determined by qRT-PCR. HeLa cells were chosen as reference cells since expression of lncRNAs in those cells was the lowest detected. Therefore, expression of lncRNAs in other cell lines was compared to that of HeLa cells, which was set to 1. Fold difference is expressed as Log10 of the ratio. Data represent the average of three independent experiments performed in triplicate. Standard error deviation is shown. C) Expression of dystrophin isoforms in human normal fibroblast (CTRL) and fibroblasts transformed with MyoD as determined by standard RT-PCR. D) Relative expression of lncRNA in MyoD fibroblasts (MyoD), as compared to that of untransformed fibroblasts (control, CTRL), which was set to 1. Data are the average of three independent experiments performed in triplicate. Standard error deviation is shown. (doi:10.1371/journal.pone.0045328.g005)


A
ncINT1Ms2
ncINT44s
ncINT44s2
ncINT55s
AS_INT55
GAPDH
RD
Nuc
Cyto
28S
18S
Relative quantity
Nuc. RNA
Cyto. RNA
N/D
0.002
RD
Rhabdomyosarcoma cell line


B
ncINT1Ms2
ncINT44s
ncINT44s2
ncINT55s
AS_INT55
GAPDH
SJCRH30
Nuc
Cyto
28S
18S
Relative quantity
Nuc. RNA
Cyto. RNA
N/D
0.05
0.08
0.002
SJCRH30
Rhabdomyosarcoma cell line

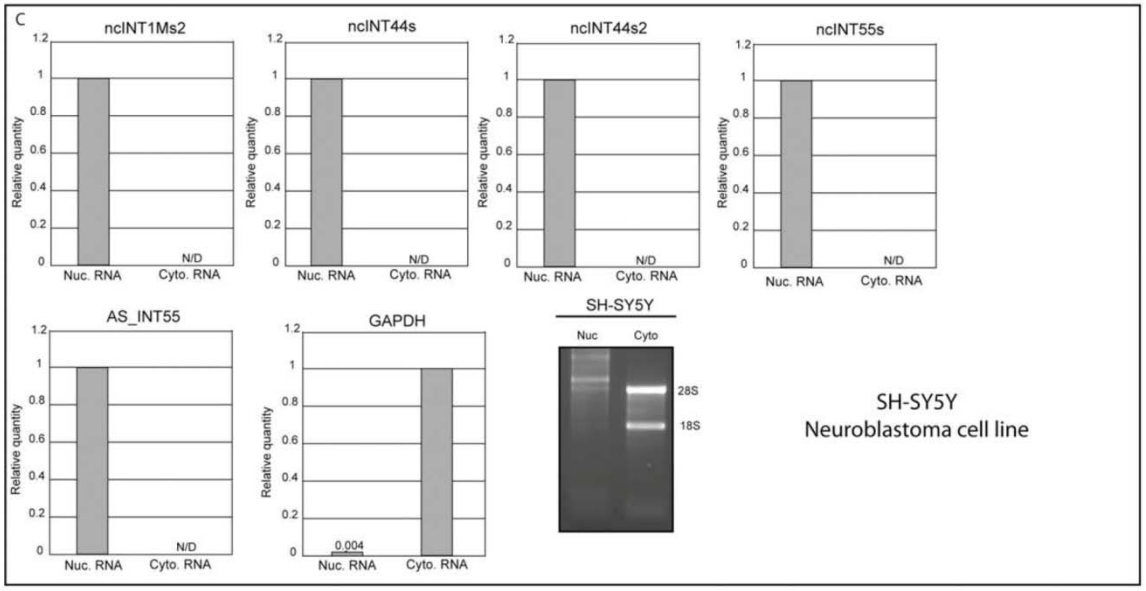

C
ncINT1Ms2
ncINT44s
ncINT44s2
ncINT55s
AS_INT55
GAPDH
SH-SY5Y
Nuc
Cyto
28S
18S
Relative quantity
Nuc. RNA
Cyto. RNA
N/D
0.004
SH-SY5Y
Neuroblastoma cell line

Figure 6. Intracellular localization of lncRNAs in three distinct human cell lines. Cytoplasmic and nuclear fractions of cellular RNA were analysed for lncRNA expression by qRT-PCR. Cyt = Cytoplasm, Nuc = Nuclear. qRT-PCR was performed for more than 50 cycles in order to detect even extremely low transcript levels. When expression was detected just in one compartment, the relative quantity of a given lncRNA was set to 1, indicating that 100% of the transcript was observed only in that compartment. In contrast, when both compartments expressed some levels of the lncRNA, distribution of the transcript between the two compartments was quantified as a relative percentage. In the specific case, standard error deviation was calculated. Quality of extracted RNAs was evaluated by running 1 ug of nuclear and cytoplasmic RNAs on a denaturing agarose gel. 28S and 18S ribosomal RNA are indicated as molecular weight markers. (doi:10.1371/journal.pone.0045328.g006)

Sense ncRNAs Negatively Modulate Dp427m, Dp427b and Dp427p Isoforms but not Dp71

To investigate the function of our long ncRNAs, we reasoned that similarly to other long ncRNAs function, DMD ncRNAs may act to regulate the expression of the dystrophin transcription. To test this hypothesis, we initially used RNA interference to deplete the expression of a ncRNAs. However, this approach failed, most likely because siRNAs cannot efficiently target nuclear transcripts (data not shown). We decided then to adopt an alternative approach through which we wished to perturb the endogenous dystrophin expression by exogenous overexpression of the DMD lncRNAs. To this purpose ncINT44s, ncINT44s2, ncINT55s and nc3UTRas sequences were cloned into the pcDNA3.1(+) expression vector, under the transcriptional control of the CMV promoter. Rhabdomyosarcoma (SJCRH30) or neuroblastoma (SH-SY-5Y) cells were individually transfected with each ncRNA. As a negative control, cells were also transfected with the pcDNA3.1(+) empty vector. Total RNAs were prepared and tested by qRT-PCR for expression levels of endogenous dystrophin isoform mRNAs. GraphPad Prism 5 software and one-way ANOVA with Dunnett's test, were used to statistically analyse the mean of RT-PCR results obtained from triplicates of three independent experiments performed in each tested cell line. Figure 7 shows that, with the sole exception of the nc3UTRas, overexpression of each long non coding transcript results in a significant reduction of Dp427m and Dp427b in both cell lines when compared to the control sample. Although modest, we could also observe some reduction on the Purkinje dystrophin isoform (Dp427p). Importantly, no variation was observed with regard to the Dp71 transcript, suggesting that the effect of sense lncRNAs on muscle and brain dystrophin isoforms is specific. Overall our finding support a model in which specific DMD lncRNAs can control muscle and brain dystrophin isoforms by down-modulating their transcription level. Finally to understand how lncRNAs can affect dytrophin transcription we have analysed their effect on a luciferase reporter gene whose expression is driven by the Dp427m promoter. A SV40-luc reporter was also used as a negative control. The analysis was performed both in SH-SY5Y human cells and in mouse C2C12 undifferentiated myoblasts. Figure 8 shows that all three lncRNAs can significantly down-regulate

expression of the Dp427m-Luc reporter but not that of the SV40-luc reporter, thus supporting the model that DMD lncRNAs can negatively control muscle specific dystrophin isoform expression by targeting the promoter (Bovolenta et al., 2012a).

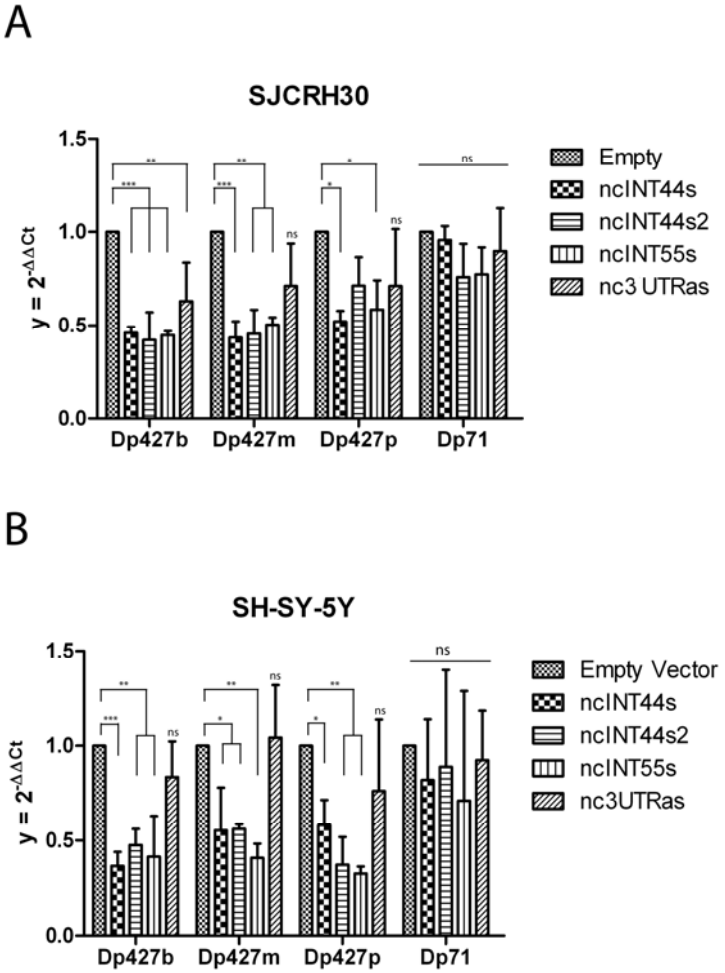


Figure 7. Overexpression of individual lncRNAs downregulates muscle and brain dystrophin mRNA isoform expression. SJCRH30 (panel A) and SH-SY-5Y (panel B) cells were separately transfected with vectors expressing ncINT44s, ncINT44s2, ncINT55s, nc3UTRas respectively. Endogenous expression of Dp427b, Dp427m, Dp427p and Dp71 dystrophin mRNA isoforms was determined by qRT-PCR with specific TaqMan systems. The pcDNA3.1(+) empty vector was used as a negative control. Amount of each dystrophin isoform was normalized to ß-actin mRNA, and was expressed as 2-DDCt. Dystrophin expression levels determined for each ncRNA transfection were compared to the condition with the empty vector which was set to 1. Statistical analyses were performed using one-way ANOVA with Dunnett's test based on results of three indepedendent transfections in which each transfection point was performed in triplicate. The error bars and asterisks represent the standard deviations and p-values respectively. (*) = p,0.05; (**) p,= 0.01; (***) = p,0.001, ns = non significant. doi:10.1371/journal.pone.0045328.g007

Relative Expression of lncRNAs and Dystrophin Isoforms in Muscles of DMD Female Carriers

To support the idea that a negative relationship between lncRNAs and dystrophyn mRNA levels may exist, we analysed expression of lncRNAs in muscle samples of 9 DMD female carriers either healthy or mildly affected (Table 2). Since the relative amount of dystrophin mRNA can be

significantly different among samples, the relative amount of each lncRNAs was compared with the ratio between dystrophin mRNA and lncRNAs amounts of the same sample. Through this mathematical device we should be able to observe whether the two variables (ncRNA and dystrophin mRNA) are somehow causally linked or completely independent. As predicted we found an inverse correlation between ncRNA expression and dystrophin expression (Fig. 9), thus supporting the model of a negative regulation of lncRNAs on specific muscle dystrophin transcripts (Bovolenta et al., 2012a).

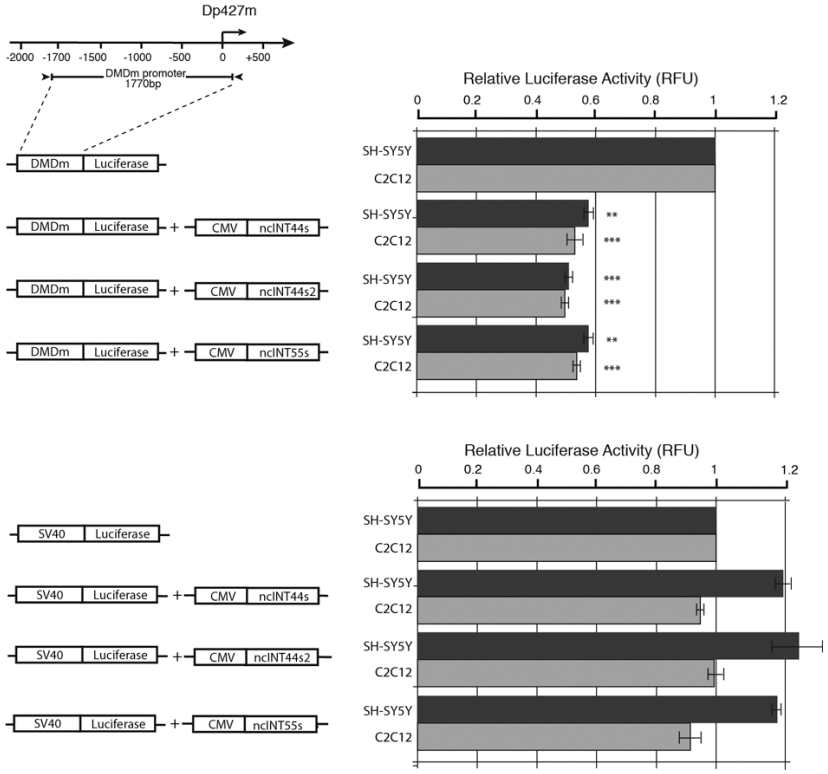


Figure 8. lncRNAs down-regulate expression of a luciferase reporter driven by the Dp427m promoter. Cartoons describes the Lucreporter construct and lncRNAs expression vectors. Expression of the muscle dystrophin reporter was determined as a function of co-transfection of the ncNT44s, ncNT442s and ncNT55s RNAs and of the expression vector alone used as the negative control whose activity was arbitrarily set to 1. An SV40-luc reporter was used as a control of the specificity of the lncRNA effect. Transfections were performed in human neuroblastoma SH-SY5Y cells (drak grey bars) as well as in mouse C2C12 myoblasts (light grey bars). Results are the average of four indipendent trasfection experiments in which each point was analysed in triplicate. Statistical significance is indicated by either two or three asterisks representing p-values ,0.01 and 0.001 respectively. (doi:10.1371/journal.pone.0045328.g008)

ID	Classification	DMD mutation
C1	Symptomatic	t(X;9)(p21.1;p22.1)
C3	Symptomatic	Dup exons 5–7 c.265-?_649+?dup (out of frame)
C4	Symptomatic	Del exons 8–9 c.650-?_960+?del (out of frame)
C8	Asymptomatic	Dup 1P-7 and dup 13–42 chrX:g.(33,068,711_33,068,771)_(32,684,693_32,684,750)dup g.(32,523,766_32,523,826)_(32,228,415_32,228,475)dup
C12	Asymptomatic	Del exons 46–51 c.6615-?_7542+?del (out of frame)
C13	Asymptomatic	Del exons 49–50 c.7099-?_7309+?del (out of frame)
C17	Asymptomatic	c.1615C>T (ex 14) p.R539X

Table 2. Classification and mutations of the Female Carriers. (doi:10.1371/journal.pone.0045328.t001)

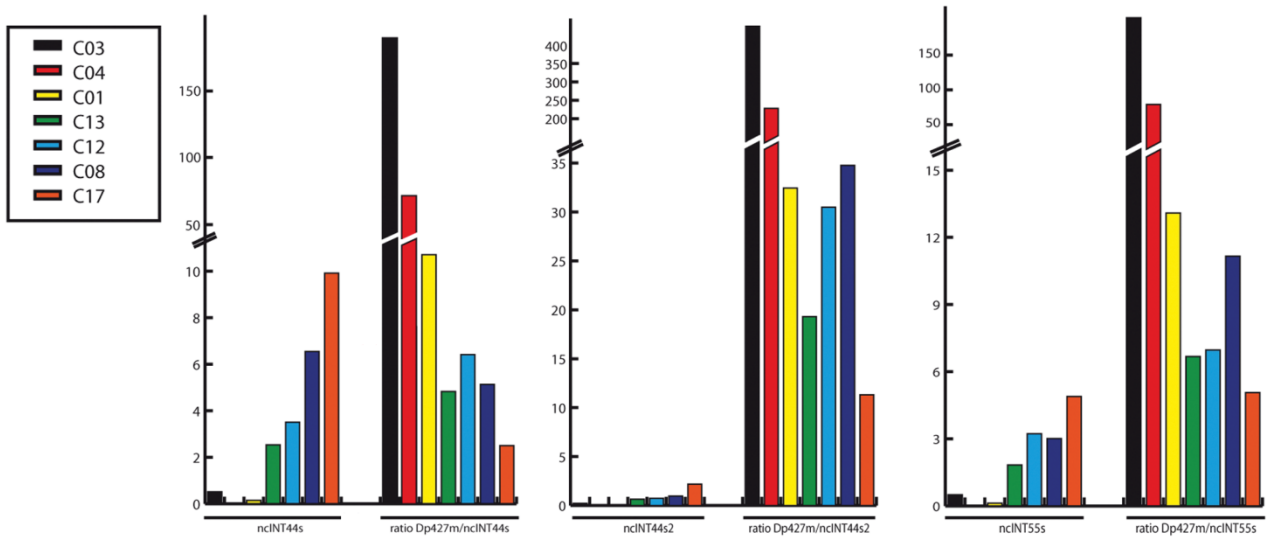


Figure 9. Correlation between the expression of different dystrophin mRNA isoforms and expression of ncRNAs. Expression of individual ncRNAs as well as of distinct dystrophin isoforms was determined in muscle samples obtained from DMD female carriers by qRT-PCR. Relative amount of each lncRNA was correlated to the the ratio between dystrophin mRNA amount and lncRNAs amount. In the graph female carriers were ordered based on their relative expression of lncRNAs ranging from that with lowest to that with highest lncRNA levels. The same order was maintained for dystrophin/lncRNA ratios. (doi:10.1371/journal.pone.0045328.g009)

New Regulative Regions Within Human DMD Locus

An intriguing aspect regarding the transcription of the *DMD* locus is due to the gene size (2.4 Mb). In fact, RNA polymerase II takes about sixteen hours to complete synthesis of the primary transcript and how the polymerase can accomplish such a extraordinary task has never been investigated.

It has been demonstrated that the phosphorylation of specific serine residues on Pol II CTD heptads' repeats reflects each specific phase of the transcription cycle. At the initiation of transcription, the serine at position 5 (P-Ser5) becomes phosphorylated to permit promoter clearance. As Pol II elongates toward the 3' end of the gene Ser5 phosphorylation declines and there is increased phosphorylation of serine at position 2 (P-Ser2) which promotes recruitment of the 3' end processing factors.

Importantly, Pol II CTD modifications are functional coupled to processing of pre-mRNA as well as to chromatin modification. Trimethylation of H3K4 peaks at promoters, whereas dimethylation (2Me-H3K4) extends into the 5’ region of coding regions and monomethylation persists throughout the gene. The pattern of H3K36 methylation shows a reverse gradient relative to K4. The trimethylation of H3K36 increases towards the 3’ ends of genes and it is functionally associated with actively transcribed genes and in particular with splicing process. H3K36me3 exon marking provides a dynamic link between transcription and splicing. Finally, H3 acetylation (Ac-H3)

represent an important hallmark that define chromatin regions permissive for gene expression and in particular localized histone acetylation is observed in promoter and enhancer elelments.

To explore the entire *DMD* locus while it is actively transcribed, we performed ChIP-on-chip analysis. Specifically, we analyzed the distribution of RNA polymerase II (Pol II), its CTD functional modification, phosphorylation on serine 2 and serine 5 (P-Ser2, P-Ser5) and the major histone modifications (Ac-H3, 2Me-H3K4, 3Me-H3K36) associated with the transcription along the 2.4 Mb spanning the gene. To identify new regulative regions within the *DMD* locus, ChIP-on-chip experiments were carried out in two specific cellular models: the SJCRH30 rhabdomyosarcoma cell line, deriving from a male striated muscle tumor, which expresses the two most relevant full-length dystrophin isoforms, Dp427b and Dp427m and the Dp71 isoform; the HeLa cells, which instead expresses only the ubiquitous Dp71 isoform.

The immunoprecipitated DNA was amplified by a whole genome PCR method, labeled and hybridized onto custom made chip DNA arrays carrying probes corresponding to the single copy portion of the *DMD* locus. Raw data have been processed for statistical significance using the Whitehead Per-Array Neighborhood algorithm.

ChIP-on-chip results obtained from SJCRH30 cells have shown enrichment for those DNA regions in close proximity to transcriptional start site (TSS) of Dp427b and Dp427m, as expected. Pol II, P-Ser2 and P-Ser5 are enriched through the first exon of each isoform while P-Ser2 is enriched along the most distal regions. Their active transcriptional status was demonstrated by showing that the some regions were enriched also for the two markers of open chromatin, Ac-H3 and 2Me-H3K4 as well. These findings (Fig. 10A) corroborate the accepted model in which the CTD modifications of Pol II are coupled to the stages of transcriptional cycle as well as to post-translation histone modifications. When Pol II is hyperphosphorylated only in P-Ser5 it is stalled on the transcriptional start site (producing abortive RNAs), whereas, when it became phosphorylated in P-Ser2, Pol II switch on to a processive status and catalyzes polymerization of transcripts.

Furthermore, as we expected, slight enrichment for P-Ser2 and P-Ser5 were been obtained in the chromosomal region coding for Dp71 isoform, the ubiquitous expressed dystrophin (Fig.10A). None enrichment, instead, were found on Dp427p TSS.

Surprisingly, the same pattern of enrichments (for Pol II, P-Ser2, P-Ser5, Ac-H3 and 2Me-H3K4) were been found in regions that have never been associated before with a transcriptional activity within the *DMD* locus: an intronic region of 3Kb, approximately 21Kb downstream of the end of *DMD* exon 52, and a region of 4Kb, spanning the *DMD* exon 62 (Fig. 1B).

The presence of the 3Me-H3K36 mark downstream of the exon 62 region suggests that, here, the RNA polymerase could coordinate two important transcriptional processes that are occurring: the

recognition of exon 62/63 junction during transcription of two full-length dystrophin isoforms (where the Pol II pauses to regulate the splicing events) and the TSS of the Dp71 isoform transcription, which is contained within the intron 62.

Interestingly, we found strong presence of open chromatin marks (Ac-H3 and 2Me-H3K4) around intron 34 and the exon 45 of the *DMD* locus, notably without presence of RNA polymerase II. These findings suggest that these DNA regions may exert a regulatory function, possibly acting as enhancer-like elements on specific dystrophin promoters (Fig 10C).

To verify whether Pol II distribution and the pattern of histone H3 methylation and acetylation found in rhabdomyosarcoma cell line could be either correlated to transcriptional dynamics of the *DMD* locus or unrelated to this, the ChIP-on-chip analysis was also performed in HeLa cells. Moreover, as we said before, since the *DMD* locus is regulated in a different manner in the two cell lines, comparing the profile of Pol II distribution and chromatin modifications obtained in both cellular contexts allowed us to discriminate which putative regulative regions are involved in muscle and brain dystrophin isoforms expression.

In HeLa cells, the ChIP-on-chip data analysis showed no enrichments of Pol II around the Dp427b, Dp427m and Dp427p TSS, while revealed an actively transcribed chromatin around the Dp71 TSS region (Pol II, P-Ser2, P-Ser5, Ac-H3 and 2Me-H3K4), as we expected. This result underlines that ChIP-on-chip methods is a powerful and reliable experimental approach to disclose the transcriptional status of a gene of interest (Fig 10D). As shown in Figure 10E, RNA polymerase II and pan-acetylation of histone 3 signals were not detected from the regions around the intron 52 and exon 62. Nonetheless, the 2Me-H3K4 mark was still present on the intron 52 and exon 62. Interestingly, distribution of this mark on exon 62 was, however, different from that observed in rhabdomyosarcoma cells. Together, these data suggest either that the new possible transcriptional start sites could be inactive or that these chromatin regions play other functions for the *DMD* locus in HeLa cells (Fig. 10E). Finally, contrary to that we found in SJCRH30 cells, the marks of transcriptionally active chromatin such as pan-acetylation and 2Me-H3K4 were completely absent at the in intron 34 and exon 45 regions (Fig.10F), supporting the hypothesis that these regions are specifically involved in muscle and brain dystrophin isoforms expression.

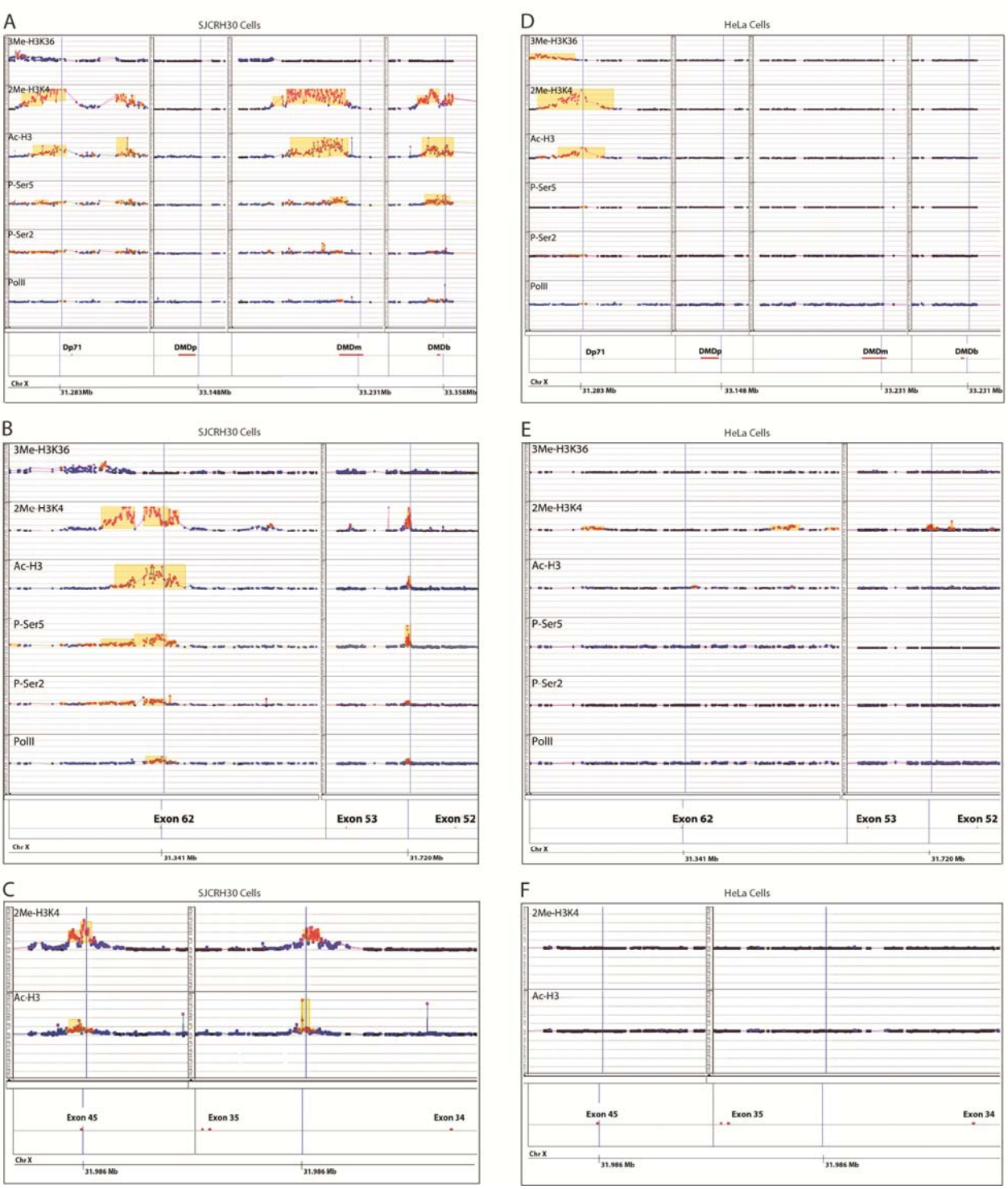


Figure 10. Distribution of RNA polymerase II (Pol II), its CTD functional modification, phosphorylation on serine 2 and serine 5 (P-Ser2, P-Ser5), and the major histone modifications associated with the transcription and splicing (Ac-H3, 2Me-H3K4, 3Me-H3K36) along the human *DMD* locus. The enrichment signals found by ChIP-on-chip analyses were displayed. In SJCRH30 cells: A) Dp427b, Dp27m and Dp71 TSS regions showed enrichments for Pol II, P-Ser5, P-Ser2, Ac-H3, 2Me-H3K4 and 3Me-H3K36; B) regions surrounding intron 52 and exon 62 were enriched for Pol II, P-Ser5, P-Ser2, Ac-H3, 2Me-H3K4. The 3Me-H3K36 mark was found only downstream of exon 62; C) the regions around intron 34 and exon 45 showed enrichments for the Ac-H3, 2Me-H3K4; In HeLa cells: D) the Dp71 TSS region showed enrichments for Pol II, P-Ser5, P-Ser2, Ac-H3, 2Me-H3K4 and 3Me-H3K36; E) the 2Me-H3K4 mark was found around intron 52 and exon 62 regions but it showed a different distribution from that observed in rhabdomyosarcoma cells F) no enrichments for Ac-H3, 2Me-H3K4 were found within intron 34 and exon 45 regions.

Further Analysis of DMD Intron 52 and Exon 62 Regions

To validate the distribution of RNA polymerase II and its major CTD modifications on the previously identified DMD intron 52 and exon 62 regions, qChIP experiments were performed using antibodies against Pol II, P-Ser2 and P-Ser5. Several pairs of primers encompassing the DNA regions found through the ChIP-on-chip method were designed to define more finely where the enzyme shows a stronger signal of occupancy (Fig. 11). Moreover, to discriminate whether the Pol II-enriched intron 52 and exon 62 regions are new transcriptional start sites or pausing sites, we performed qChIP on SJCRH30 cells after treatment with 5,6-DichloroBenzimidazole Riboside (DBR), an RNA polymerase inhibitor which prevents Pol II to switch from the stalled to the processive conformation. The rationale is that, after treatment, only the transcriptional start sites will be enriched for RNA Polymerase II, while an hypothetical pausing site will be depleted of a processive polymerase, in particular of the signal relative to the Pol II modified on CTD serine 2 residues. As shown in figure 11, we observed a depletion of RNA Polymerase II occupancy on the intron 52, suggesting that this region may represent a pausing site for the polymerase during the transcription of Dp427m and/or Dp427b mRNAs. Regarding the exon 62 region, instead, the enrichments for Pol II and P-Ser5 were still preserved after DRB treatment, but that for the P-Ser2 was strongly reduced, suggesting that the region may bear a new possible transcriptional start site (Fig. 11) on which the polymerase is sitting, however without being processive. This implies that, under certain circumstances this region may generate novel RNAs not yet described. As a control experiment, we also analyzed the region encompassing the known TSS of GAPDH gene. Again, after DRB treatment we can observe a dramatic reduction of the P-Ser2 signal (Fig. 11).

Further Analysis of DMD Intron 34 and Exon 45 Regions

To investigate what kind of regulatory function intron 34 and exon 45 regions might exert on the brain and muscle dystrophin transcription (i.e. enhancer-like or silencer-like), each DNA region was cloned, in both orientation (direct and inverse in respect with transcriptional direction), downstream of luciferase gene reporter (Luc) driven by either Dp427b or Dp427m promoter. The two regions were positioned downstream of the reporter gene in order to reflect their original position with respect to the dystrophin promoters in the *DMD* locus. Each construct was transiently transfected in the HeLa and SJCRH30 and luciferase activity was monitored at 24 hours from transfection. Results showed no significant variation in Dp427b promoter activity for both regions (data not shown).

Surprisingly, the intron 34 and exon 45 regions seem modulate Dp427m promoter (Fig. 12) in both tested cell lines. Regarding the first one, it showed the ability to stimulate the Dp427m promoter transcription when cloned in both orientation, but preferentially in the inverse orientation with respect to transcriptional direction (Fig. 12A). By sequential deletions of the cloned intron 34 region it was possible to identify the minimal region required for transcriptional stimulation of Dp427m promoter. Such region, deleted in Int 34Δ1 construct, corresponds to chromosomal coordinate ChrX: 32389220-32389513 (GRCh37/hg19).

The region spanning the DMD exon 45 shows a direction dependent behavior; indeed it can stimulate the Dp427m promoter only when it is cloned in opposite manner with respect to transcriptional direction (Fig. 12B).To identify the minimal region of exon 45 required for Dp427m promoter stimulation we proceeded to sequential deletion of the cloned region. Stimulating activity was still present in Ex45 InvΔ1 construct while it was lost in Ex45 InvΔ2 plasmid, suggesting that stimulating region is uniquely deleted in Ex45 InvΔ2 and corresponds to chromosomal coordinate ChrX: 31986208-31982621 (GRCh37/hg19). Although the luciferase reporter technology does not take into account the chromosomal distances between the Dp427m promoter and either intron 34 or exon 45 regions, it revealed precious insights about the molecular significance of such DNA elements within *DMD* locus.

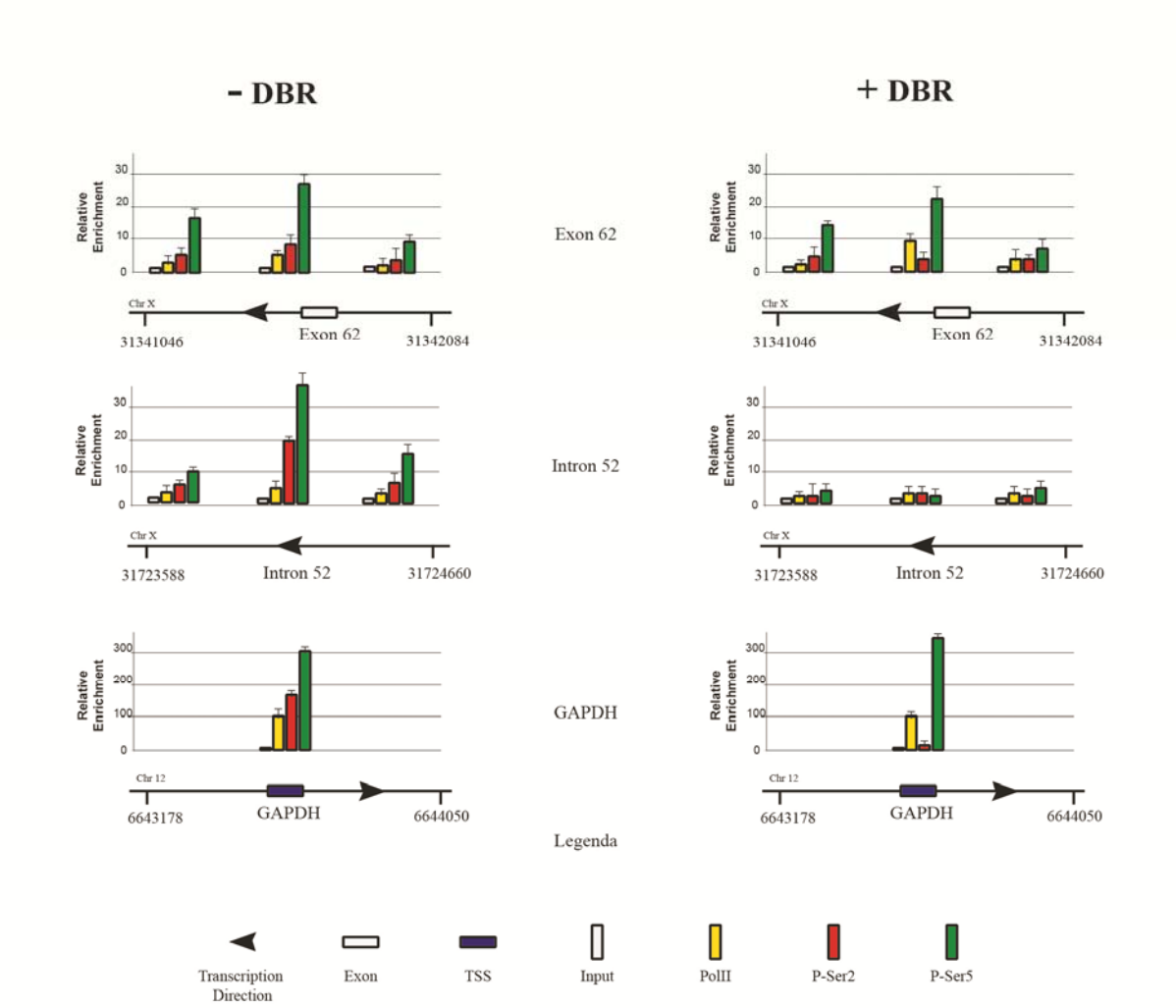


Figure 11. Validation of DMD intron52 and exon 62 regions were carried out by qChIP experiments, using antibodies against Pol II, P-Ser2 and P-Ser5 (graphs on the left specified as -DRB). qChIP experiments were performed on the same regions after the treatment with 5,6-DichloroBenzimidazole Riboside (DBR), an RNA Polymerase inhibitor that prevents Pol II switch from the stalled to the processive conformation (graphs on the right specified as +DRB).

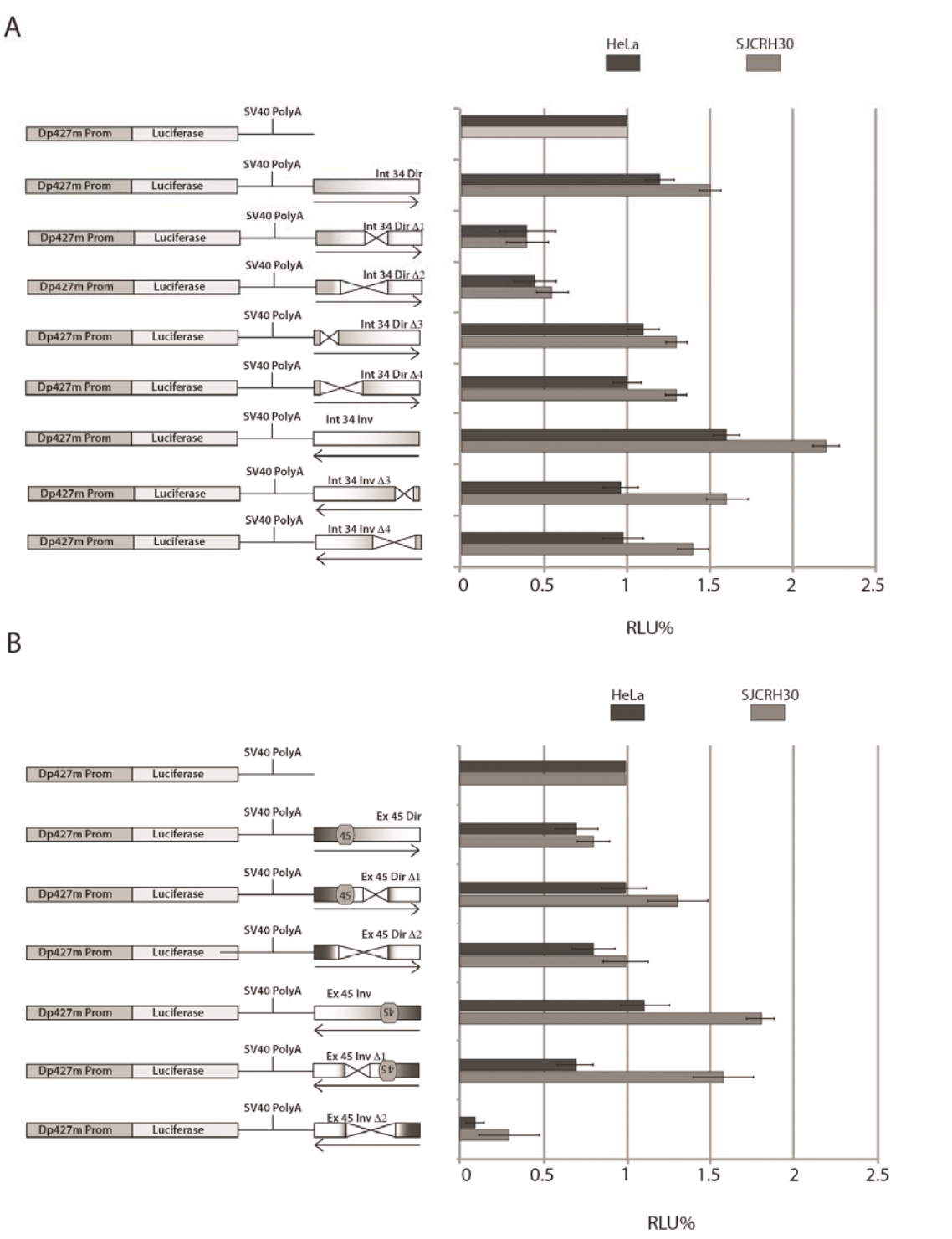


Figure 12 . DMD intron 34 and exon 45 regions seem exert an enhancer-like activity on luciferase reporter gene expression driven by the Dp427m promoter. A) Cartoons describe Luc reporter construct carrying only Dp427m promoter, used as negative control (Dp427m Prom) and the constructs containing the DMD intron 34 region (Int 34) cloned downstream of Dp427m promoter. B) Cartoons describe Luc reporter construct carrying only Dp427m promoter, used as negative control (Dp427m Prom) and the constructs containing the DMD exon 45 region (Ex 45) cloned of Dp427m promoter. For both regions: the orientation with respect to transcriptional direction are indicated as direct (Dir) or inverse (Inv). The full-length identified region or its sequential deletions (specified as delta Δ) were tested by transient transfection after 24 hours. Transfections were performed in HeLa cell line (drak grey bars) and in Rhabdomyosarcoma SJCRH30 cell line (light grey bars). The negative control activity was arbitrarily set to 1. Results are the average of three independent transfection experiments in which each point was analysed in triplicate.

Compartmentalization Dynamics of DMD Coding and Non-Coding Transcripts after Exon Skipping Treatments

Duchenne muscular dystrophy (DMD) is a genetic disorder caused by mutations in the dystrophin gene. With a few exceptions, all DMD mutations (substitutions, deletions and duplications) lead to frameshifts, resulting in the formation of premature termination codons that impair dystrophin translation. Dystrophin is a cytoskeletal protein located at the inner surface of the plasma membrane; in muscle the N-terminus binds F-actin while its cysteine-rich domain, located near the C-terminus, interacts with a large oligomeric complex of sarcolemmal proteins and glycoproteins, the dystrophin-glycoprotein complex (DGC). This provides a link between the intracellular contractile apparatus and the extracellular matrix environment. Immunohistochemical studies have shown that in differentiating human skeletal muscle cultures, the sarcolemmal localisation of dystophin occurs after myoblasts fuse into myotubes (Chevron et al., 1994; Miranda et al., 1988; Torelli et al., 1999). The exon skipping is a promising therapeutic approach for the DMD. It based on the skipping of specific mutated exons. This can be obtained by use of chemically modified antisense oligonucleotides (AONs) that, pairing with specific splice junctions or exonic splicing enhancers, prevent exon recognition by the splicing machinery. This allows to restore the frame of muscular dystrophin transcript, thus resulting in an internally partially truncated but still functional dystrophin. Despite the promising potential regarding the exon skipping therapeutic strategy, the skipped dystrophin amount is generally lower than the hoped result.

Both the export of Dp427m transcript from nucleus into the cytoplasm and its availability to be translated in this compartment belong to an important aspect regarding the transcriptional dynamics of *DMD* locus. The packaging of mRNA in mature ribonucleoprotein (RNP) complexes and the overcoming the nuclear membrane are processes tightly linked to RNA transcription, mRNA splicing, correct 3' end formation and mRNA surveillance mechanisms. To understand how these phenomena determine the proper amount of dystrophin in a muscle cell, it is important to evaluate the general cellular distribution of Dp427m mRNA.

To verify the main cellular localization of muscle dystrophin transcript, a compartmentalization study was performed in differentiated muscular cells (immortalized human myotubes) by qPCR expression analysis. To evaluate the transcriptional distribution, at the steady state, of either transcripts starting from regions near the Dp427m promoter or a transcript yielded in a distant point from this region, within the *DMD* locus, the other two full length distrophin isoforms (Dp427c and Dp427p) and the Dp71 mRNAs were checked, as well.

More importantly, to verify whether the exon skipping therapy could perturb such equilibrium, the same *in-vitro* assay was also carried out after AONs treatment. Essentially, the ultimate aim of this compartmentalization analysis was to observe whether exists a link between the export of muscular transcript and the low efficiency of exon skipping therapeutic approach in term of final protein amount. Since the new identified lncRNAs, ncINT44s, ncINT44s2 and ncINT55s, are involved, as *in vitro* as *in vivo*, in the transcriptional regulation of muscle dystrophin isoform, they were also included in the investigation to study whether they could be correlated, in some way, to an induced re-framing process as well as to the mRNA export.

Immortalized human myoblasts deriving either from a wild-type individual or from a DMD patient (carrying a deletion spanning the exons 48-50) were induced to differentiate into myotubes for 7 days and then, both fractionated to obtain nuclear and cytoplasmic RNA extracts, before and after AON51 treatment (27 hours). In the wild-type myotubes, the AON51 treatment should simulate the mRNAs distribution of dystrophic myotubes, since here the skipping of exon 51 generates a non-in-frame Dp427m transcript. As further control, the wilt-type sample was treated with a second out-framing AON (AON44), based on the analogous rationale. As a complementary experiment, the wilt-type myotubes were also treated with an in-framing AON (AON16). In this case the skipping of exon 16 yields a functional dystrophin mRNAs. These latter two treatments were not performed in DMD cells because of their extremely reduced potential to differentiate.

Two types of expression analysis were performed. The first one showed, for each transcript, the fold differences intra-compartments ($2^{-\Delta\Delta Ct}$ expressed through percentage values) between the AON-treated myotubes and the only-PEI treated ones (the transfection agent that we used as control condition). In other words, we evaluated the AON/PEI ratio within each nuclear or each cytoplasmic RNA fraction. The second one describes the fold differences of each transcript between the two compartments ($2^{-\Delta\Delta Ct}$ expressed as values of nuclear enrichment) or in other words the ratio between nucleus versus cytoplasm. The latter kind of analysis was also carried out for the lncRNAs, although they are essentially localized into the nucleus.

At first, by running 1ug of total, cytoplasmic and nuclear RNAs on a denaturing agarose gel, we evaluated the quality of fractionation technique. The ribosomal RNA (rRNA) is the most abundant type of RNA in a eukaryotic cell (it is roughly the 95% of total RNA amount) and it is the only RNA specie visible by an elettrophoretic run. The rRNA molecules are differentially distributed between the nucleus and the cytoplasm, based on the their status of processing. Specifically, rRNA 28S and 18S are preferentially present in the cytoplasm, while their unprocessed rRNA precursors, at higher molecular weight, are still retained in the nucleus. Comparing the cytoplasmic and nuclear pattern of bands to the total RNA extracted, we could observe that the RNA fractions

resulted in a stronger enrichment of one class of molecules than its counterpart (enrichment of higher molecular weight rRNAs in nucleus and enrichment of lower molecular weight rRNAs in cytoplasm) (Figures 13D, 13E, 13F). At second, to estimate the cellular partitioning we also considered the differences of RNA concentration between the two RNA fractions. The cytoplasmic and nuclear RNA samples were eluted in the same volume at the final step of each fractionation and their concentration was measured by nanodrop. Then, we calculated the ratio between the two concentration values. It is known that the cytoplasmic RNA amount is higher than its nuclear counterpart, although the relative proportion Cyto:Nucleus changes depending on the cellular type. For our differentiated myotubes, the average ratio value of RNA concentration between cytoplasm and nucleus was 5:1, confirming this theory and the good partition seen by the electrophoresis analysis.

Analyzing the fold differences of expression between AON- versus PEI-treated nuclear fractions (the first qPCR analysis describe above) of the DMD myotubes, we observed that the level of all full-lenght isoform transcripts (Dp427c, Dp427m and Dp427p) increase after AON51 treatment. Notably, the rates of increase regarding the muscle and purkinje dystrophin isoforms were roughly comparable each other (roughly 50% and 40% respectively), while the nuclear accumulation of brain isoform is considerable higher (roughly 150% higher than nuclear amount found in PEI-treated nuclei). The fold differences of expression between AON and PEI conditions in the DMD cytoplasmic fractions showed an opposite behavior, as expected. We observed that the AON treatment induce decreased levels of all full-length dystrophin transcripts in the cytoplasm in a consistent manner. These results suggest that there is a nuclear accumulation of the transcripts. The Dp71 transcript distribution parallels to the other dystrophin isoforms (Fig. 14, graphs on the left panel).

Regarding the M isoform in wild type myotubes, the observed fold differences between AON-treated nuclei and control PEI-treated nuclei is not dependent on the final effect of exon skipping treatment (in frame or not in frame resulting muscular dystrophin transcript). The Dp427m mRNA does not accumulate in the nuclei treated with AON44 and AON16 and its cytoplasmic distribution is not affected at all by the treatment, suggesting that the axis transcription-export-translation is not altered. The skipping of exon 51 in wilt-type myotubes seems to affect very slightly this equilibrium (Fig. 14, graphs on the left panel).

The Dp427c and Dp71 mRNA are still accumulated in the AON44- and AON16-treated nuclei, but there is not a correspondent reduction of the transcripts in the cytoplasm. Actually, their levels increase in the cytoplasm and in particular the Dp71 transcript accumulate in all AON-treated wild-type cytoplasms, suggesting that the export of this dystrophin isoform is enhanced. Also the

purkinje dystrophin isoform seems improperly exported in the cytoplasm, while its nuclear transcriptional steady-state is not perturbed (Fig. 14, graphs on the left panel).

As we said before, by the same qPCR analysis, we also evaluated the nuclear and cytoplasmic ratio between AON- and PEI-treated myotubes (wild type as well as dystrophic) relative to the regulatory lncRNAs transcribed within the *DMD* locus (ncINT44s, ncINT44s2 and ncINT55s) . We could observe a strong accumulation of all non-coding transcripts in DMD nuclei after exon skipping (DMD AON51) (Fig. 15, graphs on the left panel). Moreover, comparing these values to wilt-type nuclear AON/PEI ratio counterparts, we could notice that the first ones are considerably higher. Importantly, taking into account the DMD nuclear and cytoplasmic fractions, the lncRNAs cellular distribution after AON51 treatment mirror that seen for the other full-lenght dystrophins, supporting their role as transcriptional regulators of the *DMD* locus in the dystrophic context, as well.

In wilt-type samples, the lncRNAs cytoplasmic enrichments found after AON treatments remain a controversial question. It is important to underline that they were found as exiguous transcripts in the cytoplasmic fractions (late Ct values in qPCR raw data), so these results could be a false cytoplasmic enrichments of lncRNAs, due to nuclear contamination of cytoplasmic RNA extracts. On the other hand, however, it is not possible to exclude that, in wilt-type myotubes, the AONs might induce an improper export of transcripts, coding as well as non-coding. In parallel, as we supposed before, an enhanced export seems occur also for the Dp427c, Dp427p and Dp71 mRNAs, in wilt-type samples. Importantly, the muscular dystrophin transcript represents the exception to this "rule".

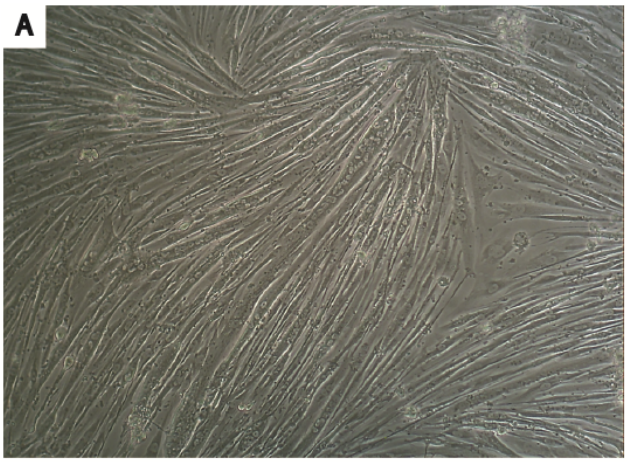
By second type of qPCR analysis we evaluated the fold difference of each transcript between nuclear and cytoplasmic compartment. For each experimental condition (untreated, PEI- or AON-treated myotubes), we pointed to 1 the level of expression of each gene in the cytoplasm and in the figure we represented each nuclear fold enrichments ($2^{-\Delta\Delta Ct}$), respectively. In this way, we could answer the our first question: "how Dp427m mRNA is distributed in a muscle differentiated cell?". The transcript levels of muscular dystrophin are remarkably enriched in the nucleus in all checked experimental conditions. The fold difference is roughly 40:1 between the two compartments in the untreated myotubes (denoted as non treated, NT, in the figures). We observed that the other two full length dystrophin mRNAs are also nuclear enriched transcripts (Fig. 14, graphs on the right panel). Furthermore, we confirmed again the essentially nuclear localization of lncRNAs ncINT44s, ncINT44s2 and ncINT55s in muscular differentiated cells, as well. In all experiments (named WT AON44, WT AON16, WT AON51 and DMD AON51 in the figures) and for each experimental condition (denoted as NT, PEI or AON), the lncRNAs are 200 fold enriched, at least, but the ratio nucleus:cytoplasm reached values of roughly 500:1, as well (Fig. 15, graphs on the right panel). The

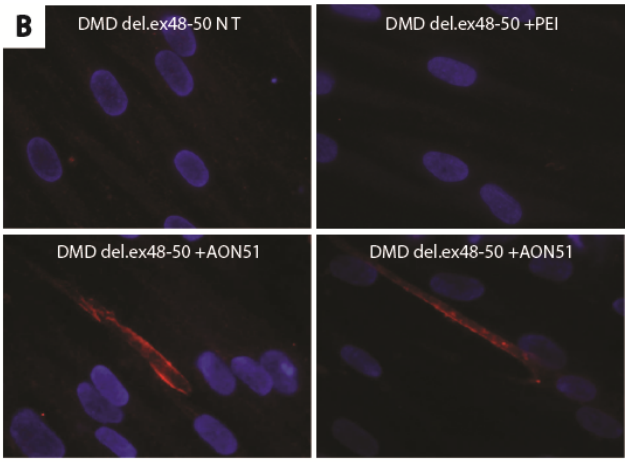
Dp71 transcript, instead, is equally distributed between the two compartments in differentiated muscular cells, showing an average nuclear enrichiment of 2 fold in the wild-type myotubes.
Overall, in the WT experiments the exon skipping seems not alter the cellular distribution of full-length dystrophin mRNAs as well as the compartmentalization at steady-state of the ubiquitous dystrophin isoform transcript. Instead, comparing the ratio nucleus:cytoplasm relative to three conditions (NT, PEI, AON) obtained for the experiments performed in wild type cells (WT AON44, WT AON16, WT AON51) we can see that the lncRNAs (ncINT44s, ncINT44s2 and ncINT55s) are less abundant in the nuclear fractions after AON44 and AON16 treatments. Likely, this might be due either to nuclear contamination of the cytoplasmic fractions or the improper export of non-coding RNAs occurring in a wild-type cells, as we speculated before. The AON51 instead did not significantly perturb the equilibrium between the two compartments (Fig. 14 and Fig.15, graphs on the right panel).
Comparing the fold differences between the two compartments (the nuclear enrichment, indicated as N/C in y-axes of the graphs) obtained in AON51-treated and control treated (PEI) dystrophic myotubes (DMD AON51 experiment), we observed that all full-length dystrophin mRNAs as well as all three lncRNAs are strongly compartmentalized in the nuclei after the AON treatment, indicating that the transcription-splicing-export axis could be affected by presence of antisense oligonucleotides in a dystrophic cellular context. Notably, the Dp71 transcript compartmentalization seems be not particularly affected, suggesting that this accumulation does not occurs for transcript starting downstream of exon 51 (the first exon of Dp71 is located within intron 62 of full-length dystrophin isoform) (Fig. 14 and Fig.15, graphs on the right panel).

Compartmentalization of Skipped Muscular mRNA

To assess whether the exon skipping treatment occurred, we performed immunofluorescence assay on DMD myotubes (Figure 13B). We can see that some dystrophic myotubes was able to restore the synthesis of muscular dystrophin protein after AON51 treatment. To quantify the amount of skipped muscular dystrophin mRNAs in each nuclear and cytoplasmic fraction, we checked the proportion (expressed as percentage) between skipped and non-skipped Dp427m transcript in RNA samples obtained from wild-type and dystrophic myotubes, before and after AONs treatment. Specific pairs of primers surrounding the region of interest was designed based on the AON used. The cDNA obtained from each RNA sample was amplified by PCR and then quantified by bioanalyzer (Agilent High Sensitivity DNA chip). The exon skipping efficiency was also evaluated in RNA total samples to estimate the overall efficiency of each type of antisense oligonucleotide

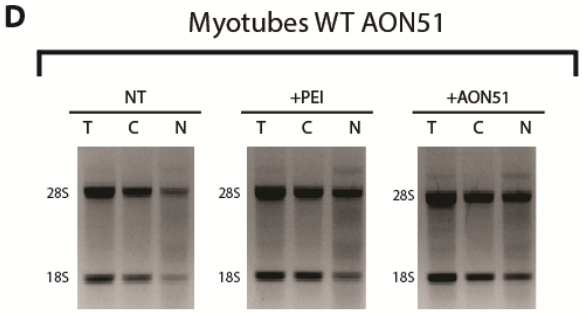
used, since it is known that, although every exon targeted *in vitro* should be removed from the dystrophin mRNA, some exons are more efficiently excluded than others (Wilton et al., 2007). In the total RNA samples, the efficiency of exon skipping was roughly the 5% (quantity of skipped versus non-skipped PCR product) for each type of AON treatment (AON16, AON44, AON51). Comparing the amount of skipped mRNA between nuclear fractions, we found that it was 5% in dystrophic cells while it represented the 8% for the wild type cells treated with AON44 and AON51. The AON16, instead, did not work at the same efficiency (only 1% of skipped Dp427m transcript relative to nuclear non-skipped mRNA). Interestingly, we could note that the percentage of skipped muscular dystrophin mRNA in the cytoplasmic fraction of DMD myotubes was the highest (11%), as compared to that obtained in the wilt-type cytoplasmic counterparts. This result displays how, in a dystrophic cell, functional Dp427m transcripts are efficiently recognized among a myriad of not-functional muscular distrophin mRNAs and exported into the cytoplasm to be translated in protein (Fig. 13C).

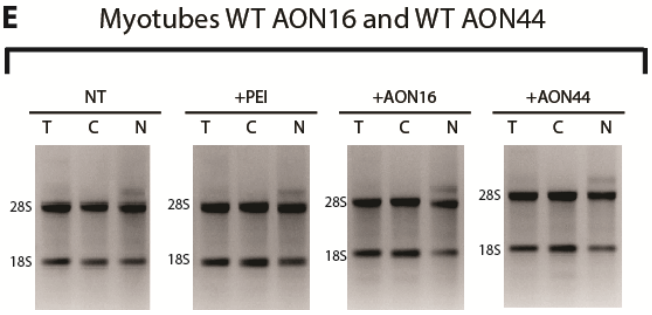




C

	% Dp427m Skipped RNA			
	WT 16	**WT 44**	**WT 51**	**DMD 51**
Tot. NT	-	1%	-	-
Cyto. NT	-	-	-	-
Nucl. NT	-	1,5%	-	-
Tot. PEI	-	1%	-	-
Cyto. PEI	-	-	-	-
Nucl. PEI	-	1%	-	-
Tot. AON	**6%**	**5%**	**4%**	**5%**
Cyto. AON	1%	4%	-	**11%**
Nucl. AON	1%	8%	8%	5%





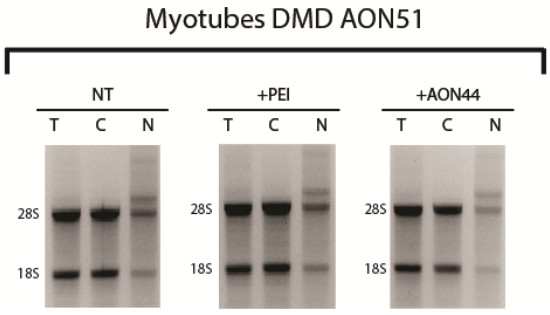
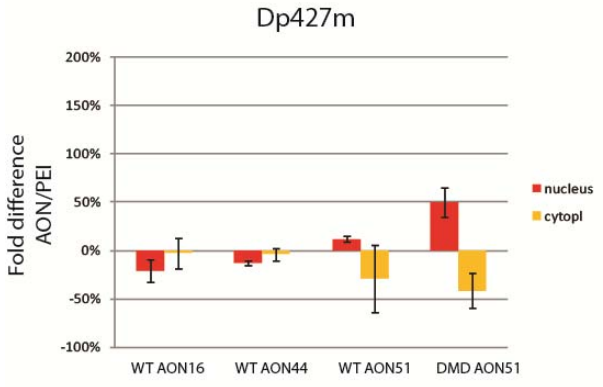
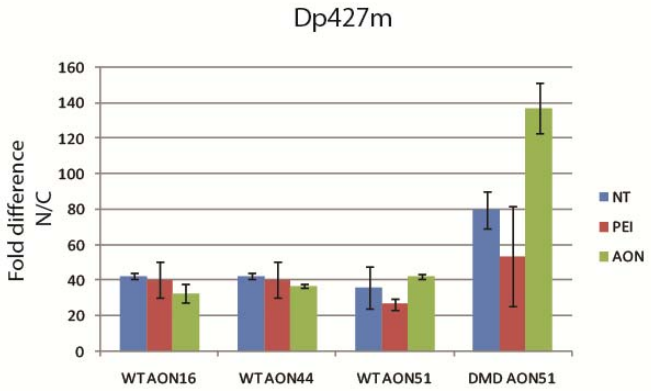
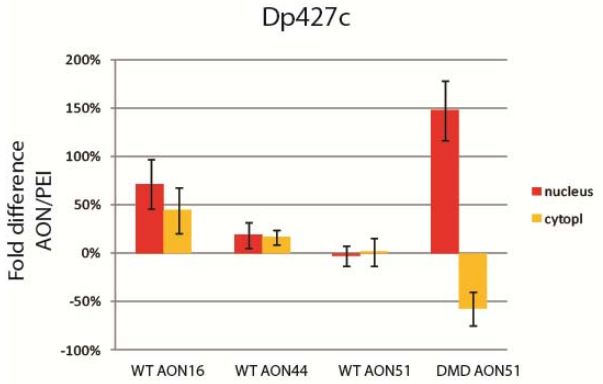
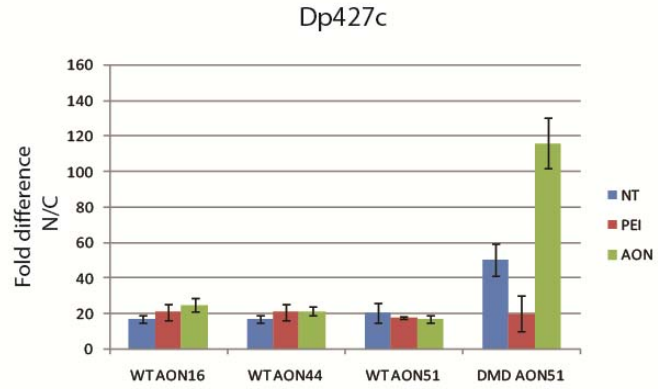


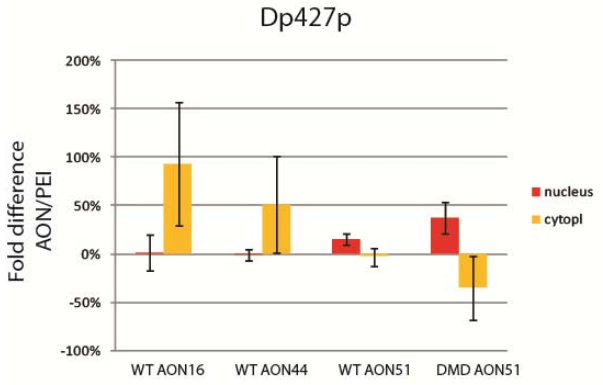
Figure 13. A) Wild type differentiated myotubes cultured after 7 days with growth medium DMEM+2%HS. B) Immunofluorescence staining of DMD myotubes carrying a deletion spanning exons 48-50. Upper left panel shows DMD myotubes untreated (NT); upper right panel shows DMD myotubes treated with PEI trasfecting agent only; lower panels represent two fields of DMD myotubes treated with AON51 that express dystrophin protein (marked in red). C) Table shows the proportion (expressed as percentage) between skipped and non-skipped Dp427m transcript in RNA samples obtained from wild-type and dystrophic myotubes, before and after AONs treatment. Specific pairs of primers surrounding the region of interest was designed based on the AON used. The cDNA obtained from each RNA sample (Tot., Cyto., or Nucl.) was amplified by PCR and then quantified by bioanalyzer technologies (Agilent High Sensitivity DNA chip). Tot.= total RNA; Cyto= cytoplasmic RNA; Nucl.= nuclear RNA; NT= non treated myotubes; PEI= only PEI-treated myotubes; AON= AON-treated myotubes; WT 16, WT 44, WT 51 indicate the experiments performed in wild-type myotubes treated or not with AON16, AON44 and AON51, respectively; DMD 51 indicates the experiment performed in dystrophic myotubes treated or not with AON51. D) E) F) Quality of extracted RNAs evaluated by running 1 ug of nuclear and cytoplasmic RNAs on a denaturing agarose gel. 28S and 18S ribosomal RNA are indicated as molecular weight markers; T=total RNA, C=cytoplasmic RNA, N=nuclear RNA.

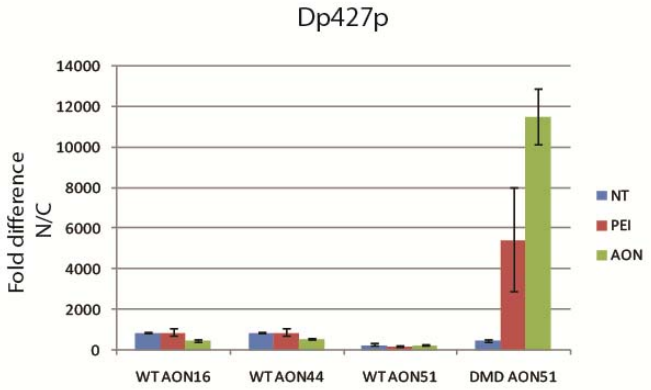

Dp427m
Fold difference AON/PEI
200%
150%
100%
50%
0%
-50%
-100%
WT AON16
WT AON44
WT AON51
DMD AON51
nucleus
cytopl

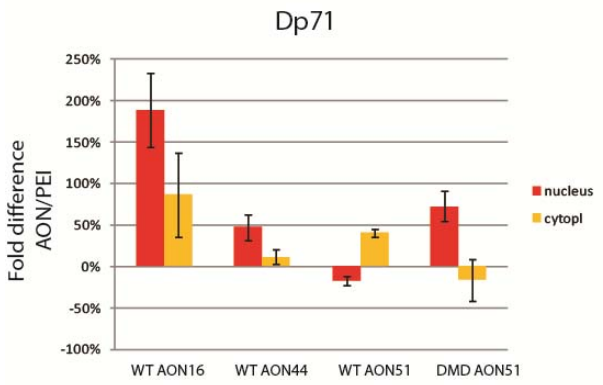

Dp427m
Fold difference N/C
160
140
120
100
80
60
40
20
0
WT AON16
WT AON44
WT AON51
DMD AON51
NT
PEI
AON


Dp427c
Fold difference AON/PEI
200%
150%
100%
50%
0%
-50%
-100%
WT AON16
WT AON44
WT AON51
DMD AON51
nucleus
cytopl


Dp427c
Fold difference N/C
160
140
120
100
80
60
40
20
0
WT AON16
WT AON44
WT AON51
DMD AON51
NT
PEI
AON


Dp427p
Fold difference AON/PEI
200%
150%
100%
50%
0%
-50%
-100%
WT AON16
WT AON44
WT AON51
DMD AON51
nucleus
cytopl


Dp427p
Fold difference N/C
14000
12000
10000
8000
6000
4000
2000
0
WT AON16
WT AON44
WT AON51
DMD AON51
NT
PEI
AON


Dp71
Fold difference AON/PEI
250%
200%
150%
100%
50%
0%
-50%
-100%
WT AON16
WT AON44
WT AON51
DMD AON51
nucleus
cytopl

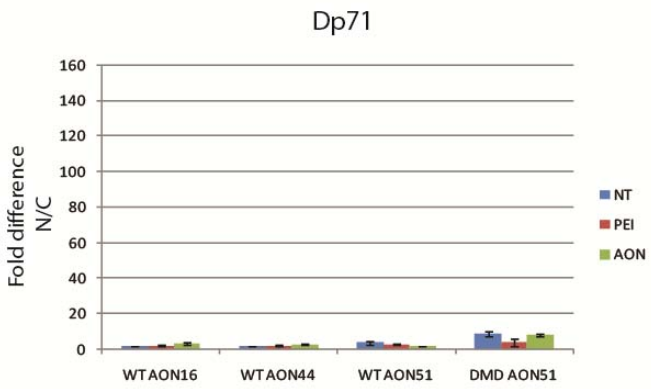

Dp71
Fold difference N/C
160
140
120
100
80
60
40
20
0
WT AON16
WT AON44
WT AON51
DMD AON51
NT
PEI
AON

Figure 14. Compartmentalization analysis of Dp427m, Dp427c, Dp427p and Dp71 mRNA isoforms. The graphs represent qRT-PCR expression analyses (2-ΔΔCt). GAPDH was used as housekeeping gene. Endogenous expression of Dp427b, Dp427m, Dp427p and Dp71 dystrophin mRNA isoforms was determined by qRT-PCR with specific TaqMan systems. Nuclear and cytoplasmic RNA fractions obtained from the experiments performed in wild-type myotubes treated or not with AON16, AON44 and AON51 are indicated as WT AON16, WT AON44, WT AON51, respectively, on the x-axis. The experiment performed in dystrophic myotubes treated or not with AON51 is indicated as DMD AON51 on the x-axis. Graphs on the left: the expression analysis shows, for each transcript, the fold differences intra-compartments (2-ΔΔCt expressed through percentage values) between AON condition and PEI condition (the transfection agent used as control condition). On the y-axis are reported AON/PEI ratio values (Fold difference AON/PEI) obtained for each nuclear (red bars) or each cytoplasmic (yellow bars) RNA fraction. Each relative PEI condition was pointed to 0% and it does not show in the graphs. Graphs on the right: the expression analysis describes the fold differences of each transcript between two compartments (2-ΔΔCt expressed as values of nuclear enrichment). On the y-axis are reported the ratio values between nucleus versus cytoplasm (Fold difference N/C) for each experimental condition (NT=non treated; PEI= only PEI treated; AON= AON treated myotubes). Each relative cytoplasmic value was pointed to 1 and it does not show in the graphs.

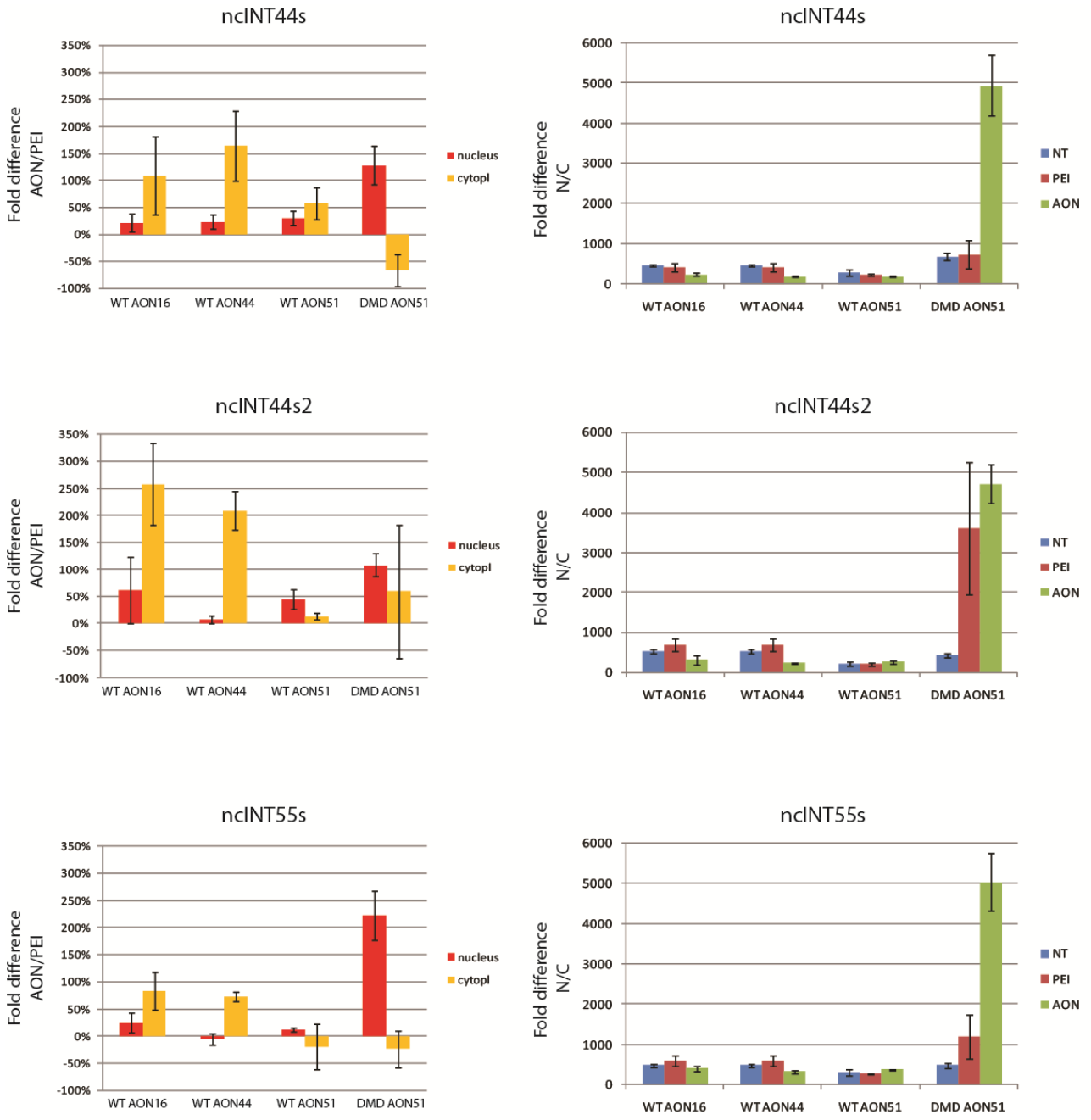


Figure 15. Compartmentalization analysis of lncRNAs, ncINT44s, ncINT44s2 and ncINT55s, transcribed within the DMD locus. The graphs represent qRT-PCR expression analyses (2-ΔΔCt). GAPDH was used as housekeeping gene. Endogenous expression of ncINT44s, ncINT44s2 and ncINT55s lncRNAs was determined by qRT-PCR with specific TaqMan systems. Nuclear and cytoplasmic RNA fractions obtained from the experiments performed in wild-type myotubes, treated or not with AON16, AON44 and AON51, are indicated as WT AON16, WT AON44, WT AON51, respectively, on the x-axis. The experiment performed in dystrophic myotubes, treated or not with AON51, is indicated as DMD AON51 on the x-axis. Graphs on the left: the expression analysis shows, for each transcript, the fold differences intra-compartments (2-ΔΔCt expressed through percentage values) between AON condition and PEI condition (the transfection agent used as control condition). On the y-axis are reported AON/PEI ratio values (Fold difference AON/PEI) obtained for each nuclear (red bars) or each cytoplasmic (yellow bars) RNA fraction. Each relative PEI condition was pointed to 0% and it does not show in the graphs. Graphs on the right: the expression analysis describes the fold differences of each transcript between two compartments (2-ΔΔCt expressed as values of nuclear enrichment). On the y-axis are reported the ratio values between nucleus versus cytoplasm (Fold difference N/C) for each experimental condition (NT=non treated; PEI= only PEI treated; AON= AON treated myotubes). Each relative cytoplasmic value was pointed to 1 and it does not show in the graphs.

Discussion

The DMD Locus Harbours Multiple Long Non-Coding RNAs

The dystrophin gene is an extraordinary source of information for unraveling molecular mechanisms underlining transcription regulation in humans, as demonstrated by the extreme fine-tuning of dystrophin transcription, its correlation with specific phenotypes and basic mechanisms, and the recent molecular therapy approach based on splicing modulation (Wood et al., 2010). Taking all this into account, exploration of the occurrence of regulatory RNAs within this gene represents an appealing and obvious target, and their full characterization will surely have great impact on our understanding of the genetic programming of the *DMD* locus. In this context, the development of our gene-specific Gene Expression microarray has led for the first time to the definition of the contribution of the *DMD* gene to non-coding RNA production. In fact, we could identify 14 non-coding poly A+ transcripts in human adult skeletal muscle, brain, heart and skin tissues transcribed mostly within introns and sense-oriented like dystrophin transcripts.

Tiling arrays of the entire human genome have expanded upon these analyses, detecting messages in the liver that map 1,529 and 1,566 novel intronic transcriptionally-active regions, arising, respectively, from the antisense or sense strands of the corresponding genes (Nakaya et al., 2007).

Although several transcribed regions have previously been identified in whole chromosomes, including the X chromosome, in a similar fashion, the *DMD* gene has never before been examined in such a detail in human skeletal muscle, brain or heart tissues. By doing so, our approach, has allowed the discovery of completely new transcripts for this genomic region that were not identified by previous array experiments, thereby demonstrating the sensitivity of our method. Further investigations demonstrated that the identified transcripts are mostly tissue-specific, highly expressed during myogenesis, and compartmentalised into the nucleus (Bovolenta et al., 2012a).

Architecture, Compartmentalization and Conservation of Long Non-coding RNAs in the DMD Gene

ncRNAs are currently categorized on the basis of their structure, and the transcripts we identified show several structural features common to many previously characterized lncRNAs. In general, as reported for other non-coding transcripts (Pang et al., 2006), our transcripts are poorly conserved. Sequence conservation could be relevant for lncRNAs likely to function in trans, where secondary

structure is a requirement for the recognition of RNA binding protein targets, and in order to exert the specific cellular functions. In fact, lncRNA expression seems to be subject to diverse levels of evolutionary constraint in mammals. In general, small ncRNAs such as miRNAs are highly conserved, while longer transcripts are less prone to conservation than typical exons. Xist, responsible for guiding X chromosome inactivation, and Air, involved in mouse-imprinted gene silencing at the Igf2r locus, are examples of poorly conserved lncRNAs. Basing on these observations, as well as on our functional data (see below), our DMD lncRNAs seem to belong to this latter class of ncRNAs, therefore their regulatory role is expected to be *in cis*. The length of our characterised transcripts ranged from about 1800 to 2800 nt, which is well above the typical length reported for exons of protein-coding genes (mean 141 nt). As for several other lncRNAs (Cheng et al., 2005; Ginger et al., 2006; Wagner et al., 2007), two of our characterized lncRNAs also exist in different isoforms, which are differentially expressed in the tissues studied, presumably reflecting tissue-specific functions. Moreover, with the exception of ncINT55as, most identified ncRNAs are unspliced, similar to Kcnq1ot1 (Redrup et al., 2009) and MEN epsilon/beta isoforms (Sasaki et al., 2009). Unbiased analyses have already indicated that a significant proportion of un-annotated ncRNAs are exclusively detected in nuclear or cytoplasmic cellular extracts (Cheng et al., 2005; Wu et al., 2008). Intronic ncRNA expression seems to be predominantly nuclear, although some subsets have primarily been detected in the cytoplasm, and only a few seem to be equally expressed in both compartments (Kapranov et al., 2007). Our results are consistent with these observations, since all analysed transcripts were found to be mainly, if not exclusively, located in the nuclear compartment (Bovolenta et al., 2012a).

DMD lncRNAs Function

lncRNAs are known to be involved in three main functional areas: i) chromatin modification and epigenetic regulation; ii) subcellular and structural organisation of transcripts; and iii) regulation of expression of neighbouring genes either *in cis* or *in trans* (a phenomenon known as transvection) (Mercer et al., 2009). Our transcripts were observed to have preferential distribution in proximity to the unique first exons of several *DMD* gene isoforms, as reported for other ncRNAs (Kapranov et al., 2007; Nakaya et al., 2007). Interestingly, considering the full repertoire of human lncRNAs, a higher proportion of transcripts are mapped to the first introns of protein-coding genes, a region relatively close to the promoters. This intense transcription activity of ncRNAs around promoter regions appears to relax the chromatin structure, making it more accessible to the transcription machinery (Bevilacqua et al., 2003). Nevertheless, a contrasting phenomenon also appears to hold

some truth. For instance, a partially intronic ncRNA produced from the genome locus that encodes dihydrofolate reductase (DHFR) directly interacts with the DHFR major promoter to reduce the expression of the protein-coding RNA (Martianov et al., 2007). Along with the latter line of evidence stand our data. Indeed we have shown that forced expression of specific lncRNAs in human cells can down-regulate expression of muscle and brain dystrophin isoforms. Apparently this negative modulation is achieved through some kind of interaction between the lncRNA and the dystrophin promoter region. Although additional work is clearly required to characterize this mechanism and to identify the cellular components involved in this phenomenon, nonetheless our analyses have revealed that some DMD lncRNAs can specifically target the dystrophin promoter and somehow dictate the dystrophin transcription rate. This is also corroborated by the analyses performed on muscle samples of DMD female carriers and from indirect observations that, for instance, a lncRNA like ncINT1Ms2, whose 3' end overlaps the promoter region of the Dp427p full-length dystrophin isoform, is expressed in the heart and in the skeletal muscle where Dp427p is not expressed, whereas it is undetectable in the brain where Dp427p is produced. Despite their relevance for the biology of the dystrophin gene, our data become particularly significant for a more comprehensive understanding of dystrophinopathies and eventually for the better design of the therapeutic approaches. For instance a similar scenario to that unraveled by this study has been recently highlighted in FSHD, in which ncRNAs were demonstrated to act as repressors of the binding copy number variations within the FSHD D4Z4 repeats with implications for the transcriptional control of putatively *in cis* pathogenic genes such as FRG1(Cabianca et al., 2012). Moreover, in another study it has been discovered that a particular lncRNA interacting with two miRNAs (one of which was enriched in the serum of DMD patients and could be correlated with the clinical assessment of the pathology (Cacchiarelli et al., 2011b) can regulate muscle differentiation (Cesana et al., 2011), with a plausible impact on the pathogenesis of DMD (Bovolenta et al., 2012a).

ChIP-on-chip Analyses of DMD Locus Unravel Novel Unexplored DMD Regions

The strictly association between RNA pol II and almost all molecular processes linked to transcription is supported by an inexorable flow of experimental data. Particularly, the dynamic phosphorylation of serine residues on the carboxy-terminal domain (CTD) heptad repeats of RNA pol II functions as "landing pad" for the recruitment (either directly or indirectly) of protein factors and enzymes that coordinate the entire cycle of transcription: from initiation of transcription to elongation, capping, splicing, 3'-end cleavage, polyadenilation, mRNA release and finally its export

into the cytoplasm. Moreover, all players that contribute to yield a mature and functional mRNA, as well as the chromatin-modifying enzymes promoting the transcription, are intricately intertwined and they influence each others. Thus, a combinatorial study of more determinants becomes a precious tool to shed light on the transcriptional dynamics of a gene, especially when the locus of interest is so complex as that encodes dystrophin. Following the extraordinary journey of RNA polymerase II along the *DMD* locus and analyzing the chromatin context within which it works allowed us to discover and focus on novel interesting regions, never been considered before.

DMD Intron 52 and Exon 62 Regions

Several efforts were been made to understand the genomic organization of human *DMD* locus and to accurately define the exon-intron boundaries (Coffey et al., 1992; Klamut et al., 1990; Koenig et al., 1987; Monaco et al., 1992; Nobile et al., 1995; Roberts et al., 1992). The *DMD* gene consists of 79 exons and at least 7 promoters that drives as many different dystrophin isoforms. Their transcription is finely regulated in a tissue- and developmental-specific manner. Actually, further level of complexity arise from the fact that the dystrophin gene produces many other isoforms generated through alternative splicing events, either excluding some exons from the primary transcript (exon skipping) or subverting the reciprocal order of exons (exon scrambling) (Sadoulet-Puccio and Kunkel, 1996; Surono et al., 1999). These events generate a further level of protein diversity and account for the complex expression regulation to which the dystrophin gene undergoes. For this reasons, it is not surprisingly that a so intricate gene could contain additional TSS, either of coding or non-coding transcripts. By ChIP-on-chip analyses we identified two regions never been associated before with transcriptional activity within the *DMD* locus. They are located around intron 52 and exon 62 regions and displaying the same pattern of enrichments found in the well known TSS of muscular, brain and ubiquitous dystrophin isoforms: the typical distribution of RNA polymerase II and its major CTD modifications associated with a promoter region (Pol II, P-Ser2, P-Ser5) and strong presence of two markers characterizing an actively transcribed chromatin (Ac-H3 and 2Me-H3K4). However, the treatment with the reversible inhibitor of new transcription, DRB, allowed to discriminate between the two regions, suggesting that the exon 62 can be still considered a putative TSS, while the intron 52 region could be an important pausing site where the pol II sit-in to coordinate the multiplicity of events occurring in the neighborhood of this portion of the dystrophin gene.

The strong enrichment for the 3Me-H3K36 mark downstream of exon 62 region and downstream of Dp71 TSS still remain a controversial matter. H3K36me3 modification is found to peak within

the body of active genes. Several studies in yeast demonstrated that the levels of H3K36me3 rise and fall intermittently within transcribed regions. Notably, exonic regions are more enriched for H3K36me3, as compared to intronic segments within the same gene. Moreover, alternatively spliced exons showed reduced H3K36me3 levels when compared to exons that are constitutively expressed, suggesting that H3K36me3 levels demark which exons are incorporated or removed by the splicing machinery. The H3K36me3 pattern of exon marking is evolutionary conserved also in mice and human (Kolasinska-Zwierz et al., 2009). Since the rate of transcript elongation seems to have an important role in determining specific exon inclusion, it is attractive to speculate that the H3K36me3 recognition by chromatin remodeling complexes might influence the rate of RNA pol II passage, thus modulating the splicing process (de la Mata et al., 2003; Howe et al., 2003). Based on these lines of evidence, the intragenic 3Me-H3K36 signature around the DMD exon 62 region and Dp71 TSS, might be necessary to allow RNA polymerase II to coordinate the splicing events together a new event of transcriptional start. Two important processes are occurring within the same temporal window: the recognition of exon/intron splicing junctions during the transcription of the full-length brain and muscular dystrophin isoforms (where likely the Pol II pauses) and the recognition of a promoter region. However, all these hypothesis remain to be confirmed. Additionally, another theory could be hazarded to explain the mean of 3Me-H3K36 mark found in these two regions, based on experimental evidence correlating the histone acetylation/deacetylation cycle to this type of histone modification. The H3K36me3 is a general hallmark of transcribed DNA. It increases toward the 3′ end of genes and then falls sharply in 3′ flanking regions (Barski et al., 2007). Similar to histone modifications at promoters, intragenic histones can be acetylated, methylated or ubiquitylated with distinct outcomes on gene expression. Histone acetyltransferase (HAT) and histone deacetylase (HDAC) complexes are enriched throughout the coding regions of actively transcribed genes, suggesting significant co-transcriptional histone remodelling (Selth et al., 2010). Intriguingly, Pol II colocalizes with HATs and HDACs in transcribed sequences during elongation cycle. In yeast the Eaf3 subunit of the Rpd3s HDAC complex is assembled at transcribed genes through Pol II coupled to H3K36 methylation (Carrozza et al., 2005; Krogan et al., 2003; Li et al., 2007; Li et al., 2003; Xiao et al., 2003). Deletion of Rpd3 resulted in spurious transcription from cryptic promoters, suggesting HDACs are required to reestablish repressive chromatin in Pol II's wake (Carrozza et al., 2005). H3K36 di- and tri-methylation recruits the Rpd3s HDAC complex, and thereby plays a critical role in repressing cryptic transcription (Carrozza et al., 2005; Shukla and Oberdoerffer, 2012). It sounds reasonable speculate that, downstream of an internal TSS, H3K36me3 might act as signal to recruit specific HDACs, thus avoiding spurious transcription. Also this hypothesis remains to be determined.

DMD Intron 34 and Exon 45 Regions

The promoter region is the main responsible DNA sequence that drive the gene expression, but not less important is the contribution of additional *cis*-regulatory DNA sequences, such as enhancers or silencers. These regulatory elements act independently of their orientation and can significantly influence the transcription in a tissue- or developmental-specific manner, although they are located at a variable distance from the transcription start site (TSS) of the gene they regulate. Thanks to the recent technological advances to map histone modifications and chromatin-associated factors, several studies have begun to characterize chromatin signatures of active enhancers, providing a more detailed picture of the basic enhancer organizations and functions (Bulger and Groudine, 2011; Ong and Corces, 2011).

In our pioneering study, the combination of chromatin immunoprecipitation with microarrays (ChIP-on-chip) has proved to be a powerful technique to explore the *DMD* locus and allows us to precisely identify new regions characterized by histone modifications associated with a enhancer element. By this experimental strategy it has been possible to restrict the research field and focus the attention on specific *DMD* potential region of interest. The trimethylation of lysin 4 on histone H3 (H3K4me3) is a strong marker found at promoters and enhancers, and this latter is bound by the transcription factor p300, a general transcriptional co-activator (Heintzman et al., 2009; Heintzman et al., 2007). The laboratory of Keji Zhao has actually shown that putative enhancer regions are enriched in several histone marks, including the three levels of H3K4 methylation, H3K27 acetylation (H3K27ac) and the histone variant H2A. Z. (Barski et al., 2007; Wang et al., 2008). In particular, an active enhancers were found to be generally associated with the presence of both H3K4me2 and H3K4me3 and, gain or loss of H3K4me2/3 at distal genomic regions correlated with, respectively, the induction or the repression of associated genes during T-cell development (Pekowska et al., 2011).

Notably, the nucleosomes represent the major physical barrier for the access of TFs to their target sites *in vivo*. The positioning of nucleosomes at enhancer elements and the different patterns of histone modification might mirror a cell lineage- and developmental-specific regulation of gene expression (He et al., 2010; Hu et al., 2011; Schones et al., 2008).

Based on these several line of evidences, we found that DMD intron 34 and exon 45 regions were marked by pan-acetylation and 2Me-K4 of histone H3 in SJCRH30 cell line, while such modifications were completely absent in HeLa cells. These findings strongly support the idea that these regions might play a role in the transcriptional regulation of Dp427m and Dp427b mRNAs as gene-distal *cis*-regulatory elements.

By Dual Luciferase Reporter assays, for each region, the potential enhancer-like regulatory activity on brain and muscular promoter was tested in both orientation with respect to direction of transcription. Interestingly, the results showed no significant variation in Dp427b promoter activity for both regions, while intron 34 as well as exon 45 regions seem positively modulate Dp427m promoter (Fig. 12) in both tested cell lines, especially in the rhabdomyosarcoma cells. However, the obtained results highlight some questions that have to be solved. The intron 34 region seems exert a positive effect in both orientations, although it contributes in a modest manner to the M promoter activity. It is possible that the alone co-presence of Dp427m promoter and the putative enhancer region is not sufficient to simulate the physiological genomic interaction between the two DNA elements. Notably, the cloned *cis*-regulatory element is missing of those histone marks that might be important to recruit specific transcription factors. On the other hand, the exon 45 stimulates the muscular promoter in a preferential orientation, thus deviating from the classical definition of "enhancer" functionally independent of position and orientation, but the same reasons discussed above might be the keystone explaining such behavior.

Nuclear Compartmentalization of Muscular Dystrophin mRNA

In an early study it was observed that mRNAs were accumulated in the nuclear compartment and the mature transcripts were the predominantly species, compared to their primary or partially spliced counterparts, providing an unexpected view of the nuclei of mammalian cells. The authors suggested that the export from the nucleus rather than splicing appears to be the process limiting the expression of a large number of genes. The nuclear retention might be an important determinant of the nuclear mRNA metabolism (Gondran et al., 1999). By compartmentalization analysis performed on wilt-type and dystrophic myotubes, we could observe that the Dp427m mRNA and the other two full-length dystrophin transcript (Dp427c and Dp427p) are confined preferentially in the nuclear compartment. Although in our analysis we did not discriminate between processed and non-processed transcripts, it is reasonable to speculate that their nuclear accumulation might be a mechanisms to control the production of dystrophin protein. On the other hand, it is possible that full-length transcripts undergo to a slow and finely regulated mRNA processing intimately coordinated with the mRNA export pathway. According to this latter hypothesis, the shorter Dp71 isoform, in fact, is more equally distributed between nucleus and cytoplasm at the steady state, indicating a further mRNA metabolism.

The mRNA can be either committed for export to the cytoplasm or accumulated in the nucleus, where it will be degraded. The components of the mRNA transcriptional machinery "communicate"

each other to monitor the nascent transcript, thus determining its release from transcription site and its further export into the cytoplasm, where it will be translated into protein. Practically every step of mRNA biogenesis involves rigorous proofreading mechanisms that safeguard the quality of the messenger transcripts and avoid the erroneous export of non functional mRNAs into the synthesis compartment. Several studies in yeast demonstrated that both the factors involved in mRNA export and the mRNA processing factors take part to and are essential for the development of the mature transcript. Many works carried out in different model organisms revealed that the mRNA export pathway is conserved from yeast to humans, although every model organism has its own peculiarities (Cullen, 2003; Reed and Cheng, 2005; Reed and Hurt, 2002; Stutz and Izaurralde, 2003). There are four main processing events that occur during the formation of a mature mRNA: 5'-capping, splicing, 3'-end cleavage and polyadenylation. Each of these modifications impacts export. Importantly, 5'-capping and mRNA processing events serve as trigger to recruit protein factors that are necessary for export (Carmody and Wente, 2009).

Reed and coworkers suggested a close link between splicing and mRNA export in higher eukaryotes, demonstrating that ALY/REF and UAP56 play a key role in this coupling (Luo et al., 2001). Subsequent work from several laboratories reported that ALY/REF, UAP56 and TAP–p15 can associate with the exon-junction complex (EJC). The metazoan TAP-p15 complex, also known as NXF1–NXT1, and its homologous Mex67–Mtr2 complex in yeast, functions as general mRNA export receptors to transport mRNPs through the Nuclear Pore Complexes (NPCs). (Grüter et al., 1998; Segref et al., 1997). The ALY/REF adaptor (Yra1 in yeast) associates directly with the general mRNA exporter and physically interacts with another conserved export fator, UAP56 (Sub2 in yeast). UAP56 is a RNA helicases that can associate with several complexes involved in messenger rinonucleoprotein (mRNP) biogenesis. Finally, the exon-junction complex (EJC) is a complex of proteins that is deposited onto mRNA during pre-mRNA splicing (~20–24 nucleotides upstream of exon–exon junctions). The EJC remains bound to the mRNA during nuclear export and influences surveillance, translation and localization of mature mRNAs in the cytoplasm (Köhler and Hurt, 2007; Le Hir et al., 2001). The SR (Ser/Arg-rich) proteins have essential roles in splicing but also function as adaptors and regulators of multiple steps of mRNA metabolism including mRNA export, stability and translation (Huang and Steitz, 2005). SR proteins are abundant, evolutionarily conserved phosphoproteins. After splicing, several SR proteins remain bound to the spliced transcript and follow it through the cytoplasm. Here, they dissociate from the transcript and are re-imported. The phosphorylation status of SR proteins allows them to interact with splicing machinery or the general mRNA export receptor TAP–p15 (Huang et al., 2003).

The X-linked dystrophin gene (*DMD*) is the largest gene in the human genome. Its 79 exons cover 2.5 million base pairs (bp). This large size makes the gene prone to rearrangement and recombination events that cause mutations. The main common type of mutations is deletion of one or more exons. It occurs in the 60%-70% of the cases. However, duplications, translocations and point mutations have also been found. In general, mutations that disrupt the reading frame of the dystrophin transcript lead to prematurely aborted dystrophin synthesis and cause the severe Duchenne muscular dystrophy (DMD). DMD is the most frequent lethal heritable childhood disease. At present, limited and poorly effective therapies exist to treat DMD affected males and these are aimed to mitigate the secondary symptoms rather than correct the primary genetic cause. However, the relative new exon skipping strategy is genetic therapeutic approach that is providing promising chances to restore the protein expression in muscle. The mechanism of action is based on the use of antisense oligonucleotides (AONs) that, by pairing in a sequence specific manner, target a splice site or an exonic splicing enhancer to direct exon skipping during pre-mRNA splicing. This allows to produce a shorter, but still functional dystrophin protein, such as patients affected by the milder Becker muscular dystrophy form (BMD). Despite the encouraging expectations regarding the exon skipping therapy, the skipped dystrophin amount is generally lower than the hoped result. For this reason it is need to understand the molecular limit that impede to achieve higher levels of dystrophin protein. Dystrophin is an important structural element in muscle cells that anchors proteins from the internal cytoskeleton to those in the fiber membrane. Improving the dystrophin synthesis up to 30%, at least, allows to confer significant protection from muscle fiber degeneration (Neri et al., 2007).

To address this issue, we performed compartmentalization analysis on a dystrophic human cellular model carrying a deletion of the exons 48-50 in the *DMD* gene. Specifically, we monitored the distribution of Dp427m mRNA between the nuclear and cytoplasmic compartments before and after exon skipping treatment with AON51. Through this strategy we could evaluate whether a therapy affecting the splicing process could perturb also the mRNA export process, since they are strictly linked. To extrapolate a general mechanisms we took in account also the cellular distribution of the other two full-length dystrophin mRNAs (coding for brain and purkinje dystrophin isoforms) and the Dp71 transcript (coding for ubiquitous dystrophin). Furthermore, since the lncRNAs have been demonstrated to play a variety of roles, being involved in control of gene expression (through epigenetic, splicing and transcription mechanisms) and disease modulation, we evaluated whether the identified regulatory long non coding RNAs (ncINT44s, ncINT44s2 and ncINT55s) transcribed within DMD could be indirectly involved in a splicing-related therapeutic approach.

We observed that, in dystrophic myotubes, all full-length dystrophin mRNAs as well as all three lncRNAs are strongly compartmentalized in the nuclei after the AON51 treatment, indicating that the transcription-splicing-export axis is affected by presence of antisense oligonucleotides. This phenomenon does not occur in a wilt-type muscular context. Intriguingly, the overall compartmentalization of the Dp71 transcript seems be not particularly affected. It is reasonable to hypothesize that, within the *DMD* locus, the transcriptional machinery continues to work in a undisturbed manner for transcripts that begin downstream of exon 51 region (the Dp71 starts to be transcribed from exon 63) and the linked co-transcriptional processes (including mRNA splicing and export) follow their regular flow of events.

Intriguingly, the lncRNAs cellular distribution after AON51 treatment mirrors the distribution of full-lenght dystrophin mRNAs. This support their role as transcriptional regulators of the *DMD* locus in the dystrophic context, as well. In particular, this result might mean that they are involved in the compartmentalization regulation of full-length dystrophin transcripts, although it remain to be formally demonstrated.

Finally, regarding the exon skipping therapy, the important results obtained by quantifying the amount of skipped M transcript in the DMD cytoplasmic fractions (11%) underlines two essential questions: first, in a dystrophic cell, functional Dp427m mRNA molecules are efficiently recognized among a myriad of not-functional muscular distrophin transcripts and, second, they are exported into the cytoplasm to be translated in protein. Based on these lines of evidence it possible to speculate that the events that determine a correct splicing and the relative mRNA export are slowed down, likely because the AONs treatment represent an additional event of surveillance that occurs in the "defective transcriptional process" of a mutated *DMD* locus.

In the future, a detailed analysis dissecting the several factors involved in the splicing and mRNA export regulation will be necessary to explain the reason of such increased accumulation of the muscular dystrophin mRNA in the DMD context after AONs treatment. This could be the missing key to ameliorate the rate of dystrophin synthesis.

Materials and Methods

Gene Expression Microarray Design

We tiled the entire DMD gene, in both sense and antisense directions, using the web-based Agilent eArray database, Version 4.5 (Agilent Technologies), with 60-mer oligos every 66 bp of repeat-masked genome sequence. We defined probe sets for both orientations, encompassing the DMD exons, promoters, introns, predicted miRNA (identified by PromiRII in PromiRII website. Available: http://cbit.snu.ac.kr/˜ProMiR2/. Accessed 2007 Feb.) and conserved non-coding sequences (CNSs) identified within dystrophin introns using the VISTA programme (VISTA website. Available: http://genome.lbl.gov/vista/index.shtml). Two specific sets of probes were designed to cover, in both directions, the cDNA sequences of a group of control genes (Table S1) identified in the Gene Expression Omnibus (GEO) database (GEO website. Available: http://www.ncbi.nlm.nih.gov/geo/) and expressed equally in both normal and dystrophic muscles. Each probe set was opportunely distributed and replicated several times in order to obtain two 4×44k microarrays, referred to as DMD GEx Sense (GEO Platform number: GPL13120) and DMD GEx Antisense (GEO Platform number: GPL13121), respectively, able to detect transcripts in the same and opposite directions as that of DMD gene transcription (Table S2). All the data obtained from these platforms are MIAME compliant and can be accessed from the GEO database by searching GSE27068 in the GEO accession box.

Sample Processing

Three commercial poly A+ RNAs from normal human brain, heart and skeletal muscle tissues were utilised (Ambion). Skin poly A+ RNA was isolated from total Skin RNA (Stratagene) using the Qiagen Oligotex kit. All RNA samples were checked for purity using a ND-1000 spectrophotometer (NanoDrop Technologies), and for integrity by electrophoresis on a 2100 BioAnalyzer (Agilent Technologies). Sample labelling and hybridisation were performed according to the protocols provided by Agilent (One-Color Microarray-Based Gene Expression Analysis version 5.0.1). The array was analysed using the Agilent scanner and Feature Extraction software (v9.1). Reliability of results was verified via Agilent Quality Controls (Spike-in), consisting of a mixture of 10 in vitro-synthesized, polyadenylated transcripts derived from the Adenovirus E1A gene, premixed at concentrations spanning six logs and differing by one-log or half-log increments.

Data Analysis

We considered aggregate transcription consisting of at least three consecutive probes, whose genomic coordinates lay within a 250-nt window, exhibiting fluorescence intensities in the top 90th intensity percentile, as described by Bertone et al., 2004 (Bertone et al., 2004). In brief, fluorescence intensities for each probe designed within the DMD gene sequence were ranked from the lowest to the highest. Values higher than the fluorescence intensity corresponding to 90% of all ranked probes on the array were considered as positive hybridisation signals. Subsequently, the probes were ordered by their genomic coordinates, and transcripts were identified when at least three consecutive probes in a set of 250 base pairs were above the 90th percentile value (Figure 1 A). Identified transcripts were named according to their intron of origin, with the prefix 'nc' for non-coding and a suffix corresponding to their sense of transcription ('s' for sense and 'as' for antisense; e.g., ncINT44s is a transcript localized in intron 44 and transcribed in the same direction as the DMD gene).

Northern Blotting

Sense transcripts originating from the intronic regions harbouring the DMD gene isoform promoters and antisense transcripts were validated by Northern blotting, using a 12-lane human poly A+ RNA filter (Clontech). Riboprobes for ncINT1Ms1, ncINT1Ms2, ncINT1Ps, ncINT29s, ncINT44s1, ncINT44s2, ncINT55s, ncINT55as and nc3UTRas were generated, labelled and hybridized to the RNA filter. Hybridisation and washing were performed according to the manufacturer's instructions (Clontech), in the presence of herring sperm DNA (10 ug/ul).

RACE PCR

Rapid Amplification of cDNA Ends (RACE) was performed by using a SMART RACE cDNA Amplification Kit (Clontech), according to adapted manufacturer's instructions. For both 3′-RACE and 5′-RACE, 1 µg of poly A+ RNA (Ambion) from skeletal muscle (transcripts ncINT44s1, ncINT44s2 and ncINT55s) or heart tissue (transcripts ncINT1Ms2, ncINT55as and nc3UTRas) were reverse-transcribed using PrimeScript reverse transcriptase (TaKaRa). For 3′ RACE, cDNAs obtained from oligo(dT) extension were amplified in three subsequent PCRs in order to increase the degree of reaction specificity. The first amplification was followed by 2 rounds of nested PCRs, performed using Nested Universal Primer (NUP) and a gene-specific primer (GSP2 in the first round and GSP3 for the second). All PCR amplifications were performed using LA Taq (TaKaRa).

For 5′-RACE, first-strand cDNA synthesis was primed using a gene-specific antisense oligonucleotide (GSP-RT), and cDNA was tailed with SMARTII oligonucleotide. 5′-RACE was performed using the same PCR conditions as for 3′RACE.

For three transcripts (ncINT44s1, ncINT55s and ncINT1Ms2), a transcript-walking strategy was adopted, using the sequence that hybridised on the array as an anchor. For both 3′-RACE and 5′-RACE, PCR products were purified and sequenced both directly and after cloning into pCRII vectors using a TA cloning kit (Invitrogen). All primer sequences used for RACE PCR are listed in Table S3.

Bioinformatics Prediction

The ncRNAs characterized by RACE PCR were tested for secondary structure with mFOLD 3.2 (mFOLD 3.2 website. Available: http://mfold.rna.albany.edu/?q=mfold/RNAFolding-Form), presence of ORF using NCBI ORF Finder (NCBI ORF FINDER website. Available: http://www.ncbi.nlm.nih.gov/projects/gorf/) and the Coding Potential Calculator (Coding Potential Calculator website. Available: http://cpc.cbi.pku.edu.cn/) and presence of tRNA by TRNAscan-SE 1.21 (TRNAscan website. Available: http://lowelab.ucsc.edu/tRNAscan-SE/). To determine whether our transcripts overlapped any previously annotated ESTs or mRNA, we employed the UCSC Genome Browser to inspect our ncRNAs and identify any spliced and unspliced human ESTs and mRNAs.

lncRNAs Expression and Compartmentalization Analysis

Total RNA was purified from human RD and SJCRH30 rhabdomyosarcoma cell lines, SH-SY5Y neuroblastoma cell line and both normal and MyoD-induced fibroblasts (transformed with an Ad5-derived, EA1-deleted adenoviral vector carrying the MyoD gene as previously described in (Spitali et al., 2009) using TRi-REAGENT (Sigma Aldrich) and treating with DNAse I (Ambion). Reverse transcription (RT) was performed using random hexanucleotide primers and Superscript III enzyme (Invitrogen). Real-time PCR was performed in triplicate in a 96-well plate; each 25 µl reaction consisted of 1X Taqman Master Mix (Applied Biosystems), 300 nM forward and reverse primers, 100 nM Taqman probes and 40 ng of cDNA.

To characterise the intracellular compartmentalisation of the ncRNAs, nucleic and cytoplasmic RNA fractions were separated from 1×10^7 cells (SJCRH30, RD, SH-SY5Y) by glucose gradient centrifugation as previously reported (Carneiro and Schibler, 1984), and purified using TRi-REAGENT (Sigma Aldrich). RNA was treated with DNAse I (Ambion), and reverse transcribed

using random exanucleotide primers and Superscript III enzyme (Invitrogen). Real-Time PCR was performed as previously described. The relative quantity of the target sequence (non-coding RNA) is expressed as the relative distribution of each single sequence in the nucleic or cytoplasmic fraction. Sequences of Taqman probes and amplifying primers are listed in Table S4.

Expression of the different dystrophin isoforms was determined by standard qRT-PCR. 250 ng of cDNA derived from total RNA was amplified by PCR using Taq DNA polymerase recombinant (Invitrogen) according to the manufacturer's instructions. Primer sequences are listed in Table S5.

To determine the inclusion of the DMD ncRNAs within the full-length dystrophin transcript, we amplified by RT-PCR the exons adjacent to the introns in which we identified our ncRNAs. The regions amplified were exons 1–3; 28–30; 31–33; 36–38; 43–45; 50–52; 54–56 and 66–68. Primers sequences are available upon request.

lncRNA Expression Vectors and Transient Transfections in Cultured Human Cells

The expression plasmids pcDNA3.1(+)-ncINT44s, pcDNA3.1(+)-ncINT44s2, pcDNA3.1(+)-ncINT55s and pcDNA3.1(+)-nc3UTRas were generated by cloning the PCR product of each ncRNAs into the pcDNA3.1(+) vector (Invitrogen). The template used for PCR amplification was genomic DNA extracted from a healthy human male. Primers listed in Table S6 contain the appropriate restriction enzyme at 5′-end to facilitate cloning.

SJCRH30 Rhabdomyosarcoma and SH-SY-5Y Neuroblastoma cell lines were transiently transfected using Lipofectamine2000 according to manufacturer's instructions (Invitrogen). Cell cultures were harvested after 48 hours and RNA extracted. Each transfection was repeated three time and each point of RT-PCR was performed in triplicate. Statistical analyses were performed using one-way ANOVA with Dunnett's test.

DMD Gene Micro Fluidic Card (FluiDMD v1.1) Analysis

We explored all transcripts originating from the DMD locus, including the lncRNAs identified by our array as well as the whole exome of the dystrophin gene, using a slightly modified design of our FluiDMD card already reported (Bovolenta et al., 2012b). We replaced the replicated junction systems with specific systems able to detect the following ncRNAs: ncINT1Ms1, ncINT1Ms2, ncINT1Ps, ncINT44s1, ncINT44s2, ncINT55s, ncINT55as and nc3UTRas. The fluidic card protocol was performed as described in (Bovolenta et al., 2012b). We used these novel card (FluiDMD-ncRNAs) both in muscle biopsies and cells from dystrophinopathic patients. We analysed seven DMD females' muscle biopsies. The females were previously selected and known

being either symptomatic or asymptomatic carriers (Table 2). Total RNA was obtained from muscle biopsy of the seven females, carriers of mutations in the DMD gene as previously described (Brioschi et al., 2012).

Reporter vectors and luciferase assay

The pGL3-basic, pGL3-Sv40promoter and Renilla-TK vectors were purchased from Promega. Promoter of selected Dp427m isoform was obtained using PCR and cloned into the pGL3-basic vector. The template used for PCR amplification was genomic DNA extracted from a healthy human male. pGL3-DMDm and pGL3-Sv40promoter were co-transfected with Renilla-TK vector, lncRNA expression vectors or pcDNA3.1(+) empty vector in SH-SY-5Y Neuroblastoma cells and C2C12 myoblast cells using Lipofectamine2000 (LifeTechnologies) according to manufacturer instructions'. The activity of firefly or Renilla luciferase was measured 24h after transfection with a dual luciferase assay kit (Promega) according to the instructions. The assay was repeated three times. Statistical analyses were performed using one-way ANOVA with Dunnett's test.

The DMD intron34 and exon 45 regions were cloned in pGL3-DMDm Promoter and the relative deleted constructs were obtained by Whole Around PCR. Each plasmid (the pGL3-DMDm Promoter alone, the pGL3-DMDm Promoter_intron34 or relative deleted constructs, the pGL3-DMDm Promoter_exon45 and relative deleted constructs) were co-transfected with Renilla-TK vector in HeLa cells and Rhabdomyosarcoma cell line (SJCRH30) using Lipofectamine2000 (LifeTechnologies) according to manufacturer instructions'. The activity of firefly or Renilla luciferase was measured 24h after transfection with a dual luciferase assay kit (Promega) according to the instructions. The assay was repeated three times. All Primers used for cloning are listed in Table S7.

Cell culture and treatments

Human rhabdomyosarcoma SJCHR30 cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 50 mg/ml gentamycin and grown at 37 ºC and 7% CO2. Human HeLa cells were cultured in Dulbecco's modified Eagle's medium containing 10% heat-inactivated fetal bovine serum and 50 mg/ml gentamycin and grown at 37 ºC and 5% CO2. 5,6-DICHLOROBENZIMIDAZOLE RIBOSIDE (DBR) were purchased from Sigma-Aldrich and cells were treated for 6h at 100uM before performing qChIP

qChIP

ChIP was performed as previously described (Weinmann and Farnham, 2002).The antibodies employed in this study were: Anti-RNA polymerase II Abcam (ab76123), Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) Abcam (ab5095),Anti-RNA olymerase II CTD repeat YSPTSPS (phospho S5) Abcam (ab5131), Anti-AcethylHistone H3 Upstate (06-599), Anti-DimethylHistone H3 (Lys4) Upstate (07030), Anti-TrimethylHistone H3 (Lys 36) Abcam (ab9050). Specific pairs of primers used for quantitative ChIP are listed in Table S8.

ChIP-on-chip

Immunoprecipitated DNA from ChIP was amplified with GenomePlex® Whole GenomeAmplification (WGA) Kit according to manufacturer's instructions. Labelling and hybridisation were performed following the protocols provided by Agilent (Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis protocol v5.0). The array was analysed with the Agilent scanner and the Feature Extraction software (v9.1). A graphical overview and analysis of the data were obtained using the Agilent Genomic Workbench software (v3.5).

Immortalized human myoblasts

The wild-type (any genetic defect in dystrophin gene) and DMD (deletion of exons 48-50 in dystrophin gene) immortalized human myoblasts were obtained as described in (Mamchaoui et al., 2011) and kindly provided to Section of Medical Genetics, University of Ferrara. $5x10^4$ cells were plated in a six well. Five six wells were used for each experimental condition (untreated, PEI-treated and AON-treated).

The cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 50 mg/ml gentamycin and grown at 37 ºC and 5% CO_2. Differentiation was induced at confluence by replacing the growth medium with DMEM supplemented with 100 µg/ml transferrin, 10 µg/ml insulin 50 µg/ml of gentamycin and 2% Horse Serum (Sigma-Aldrich) for 7 days. The myotubes were treated with PEI transfection agent alone or with antisense oligonucleotides for 4 hours at 37 ºC and 5% CO_2. The myotubes were washed and cultured for 27 hours after treatment under differentiation conditions. Nucleic and cytoplasmic RNA fractions were separated by glucose gradient centrifugation as previously reported (Carneiro and Schibler, 1984), and purified using Tri-REAGENT (Sigma Aldrich). RNA was treated with DNAse I (Ambion), and reverse transcribed using oligo dT and random primers and iScript Reverse Transcription Supermix (Bio-Rad). Sequences of Taqman probes and amplifying primers are listed in Table S4.

Byoanalyzer analyses

To determine the skipped and non-skipped full-length M dystrophin transcript, we amplified by PCR the region surrounding the exon of interest. Specific pairs of primers was designed based on the AON used. The cDNA obtained from each RNA sample was amplified by PCR (Platinum®Taq DNA Polymerase, Life Technologies) and then quantified by bioanalyzer (Agilent High Sensitivity DNA chip). The amplified regions spanning: exons 46-52 to obtain a product of 840bp (non-skipped) and 610bp (skipped) in wild-type cDNA; exons 46-52 to obtain a product of 540bp (non-skipped) and 310bp (skipped) in DMD cDNA; exons 43-45 to obtain a product of 320bp (non-skipped) and 170bp (skipped); exons 12-17 to obtain a product of 700bp (non-skipped) and 520bp (skipped);

Table S1

Accession numbers and names of genes used as controls in the custom-designed DMD-GEx microarrays. (doi:10.1371/journal.pone.0045328.s005)

NM_004748.3	Cell cycle progression 1 (CCPG1)
NM_003047.2	Solute carrier family 9 (sodium/hydrogen exchanger)(SLC9A1)
NM_004045.3	ATX1 antioxidantprotein 1 homologue (ATOX1)
NM_002766.1	Pyrophosphate synthetase-associated protein 1 (PRPSAP1)
NM_001428.2	Enolase 1, (alpha) (ENO1)
NM_003095.2	Small nuclear ribonucleoprotein polypeptide F (SNRPF)
NM_001914.2	Cytochrome b5 type A (microsomal) (CYB5A)
NM_000426.3	Laminin, alpha 2 (merosin, congenital muscular dystrophy) (LAMA2)
NM_003279.2	Troponin C type 2 (fast) (TNNC2)
NM_003280.1	Troponin C type 1 (slow) (TNNC1)
NM_003281.3	Troponin I type 1 (skeletal, slow) (TNNI1)
NM_001101.2	Actin, beta (ACTB),
NM_002046.3	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
NM_006275.4	Splicing factor, arginine/serine-rich 6 (SFRS6)
NM_005626.3	Splicing factor, arginine/serine-rich 4 (SFRS4)

Table S2

Name, number of probes and reiteration of each probe set within the 4x44k sense and antisense DMD gene expression microarrays. (doi:10.1371/journal.pone.0045328.s006)

Probe set name	**N° of probes**	**Replicated in DMD GEx sense**	**Replicated in DMD GEx antisense**
Sense DMD	14928	2	0
Sense DMD exons	174	8	8

Sense DMD CNS	2411	3	0
Sense DMD MiR	51	9	0
Sense DMD promoters	34	8	0
Sense Controls	295	8	5
Antisense DMD	14914	0	2
Antisense DMD CNS	2421	0	3
Antisense DMD exons	169	4	3
Antisense DMD promoters	53	0	8
Antisense DMD MiR	51	0	8
Antisense Controls	311	5	8
	Total	**43803**	**43785**

Table S3

Name and nucleotide sequence of the primers used for 5′ and 3′ RACE
(doi:10.1371/journal.pone.0045328.s007)

Primers used for 3’ RACE
Primer Nucleotide sequence (5’-3’)
T2_GSP1 cgccctcgagctaggacatggaac
T2_GSP2 tcacatttaagagggaaggagaaacgcc
T2_GSP3 (3’) cccaccaacacctcctgttctctcc
44_8GSP1 aaatccatattcctgggggcgatgc
44_8GSP2 cgaaggtgcagccagactgggagtt
44_8GSP3 (3’) tgtgggactgcctcttgctcctgaca
44_9GSP1 cccagaaggggccttgtctgaagtg
44_9GSP2 ccagcggattacggtgtgggtgaat
44_9GSP3 (3’) tcagggatgctgctgctctctaggc
55_11GSP1 ggttatgcttcccgtctcttcacagagc
55_11GSP2 cccacgatctggaacagactggcggata
55_11GSP3 (3’) ccgcaataactctgtgaagtgctg
AS55_GSP1 aggacatgaatggatgataatttgtggg
AS55_GSP2 gaaaagcatgccataagctgttcttagc
AS55_GSP3 (3’) tggagcagtgaaaagtagattttggtgtcg
AS3UTR_GSP1 cagggatgggctgggaatccatag
AS3UTR _GSP2 ggcattgctagcagcaggaagctg
AS3UTR _GSP3 (3’) ctgccccactcagctgacagttctc

Primers used for 5’ RACE
Primer Nucleotide sequence (5’-3’)
T2_GSPRT gatgcggaaatttcatttgg
T2_TWrev ttagcatgccctttcaatcag
T2_TW1 ccagagctaatgaggccaag
T2_TW2 tgagctccaatctcctcacc
T2_TW3 tcagttgattctgatgagcacc
T2_GSP1 caggtgaggagattggagctcagagagg
T2_GSP2 gcccctagacccgtggttctccaac
T2_GSP3 (5’) aacccagtgcccaaactgcacttca
44_8GSPRT ccctcttttctaccttctaggc
44_8TWrev ggacaaaagggttgcacagt

44_8TW1	aggtagtccattttgcatcagg
44_8TW2	aacaaaacctccccaatgc
44_8TW3	tcattctgagaaccacatctgc
44_8GSP1	cagcttgaaatatgtgcattggggagg
44_8GSP2	ttgagtggggcaatttctggcagg
44_8GSP3 (5')	tggcaggtggggaagagagtaaagaag
44_9GSPRT	cagttccattctccacagg
44_9GSP1	ttcatccactgatggacgtttaggttg
44_9GSP2	tagtgggattgctggatcatctggtagc
44_9GSP3 (5')	aaagcaagacaggtcaacaaaggccag
55_11GSPRT	gttcagactatgtccatacacagg
55_11TWrev	tctgtgaagagacgggaagc
55_11TW1	gggtcacaaacgtgaaggtc
55_11TW2	ccttccagattcagcatgtg
55_11TW3	tatctctaggccgtcggttc
55_11GSP1	tcctgctttctcttccttgggttgc
55_11GSP2	cagccatggcctcacataggtgg
55_11GSP3 (5')	cccgagtagctgatgtcccttcagc
AS55_GSPRT	agctacaagctgctagacagg
AS55_GSP1	caaataagctgttggtggcaggaggtg
AS55_GSP2	tcagcatttcccaggaagggtgatg
AS55_GSP3 (5')	gggtcaagaatgcgaaggtcaaggag
AS3UTR_GSPRT	tctctgcgagtagttccacac
AS3UTR_GSP1	gcttttggagagtgggctgacatcaagtg
AS3UTR _GSP2	tgcacacctgagttcacagcttcacc
AS3UTR _GSP3 (5')	tgacgctggacctttctttacccaagg

Table S4

Name and sequence of the TaqmanRealTime systems used for the ncRNAs compartmentalisation study and dystrophin transcripts expression levels. (doi:10.1371/journal.pone.0045328.s008)

Primer Name and Sequence	
ncINT1Ms2_F	TTTGGATGCAGTACTCTTTGTAGACA
ncINT1Ms 2_R	GGAGGTGTTGGTGGGAAAAA
ncINT1Ms 2_Probe	6-FAM-AATACGGTACGATTAATG
ncINT44s_F	GGCTCCCTCATTCAATGAATCTA
ncINT44s_R	CAGCAAAAGATGTCAGGAGCAA
ncINT44s_Probe	6-FAM-ACTATGTGGGACTGCCT
ncINT44s2_F	ACCAGGAGCTCTGCTTGCAT
ncINT44s2_R	TTGTGCATGATAATGTGCCTCAA
ncINT44s2_Probe	6-FAM-CTGGGAGTTGCCCATG
ncINT55s_F	TGATAACTTTCATGCCCATTAACATAG
ncINT55s_R	AACAGGACACAAATTCAGCACTTC
ncINT55s_Probe	6-FAM-ACCGCAATAACTCTG
ncINT55as_F	CATTTAGAGCAAGAGATACAGGCATT
ncINT55as_R	GGGCACAAAAAAGAATAATTTGCTA
ncINT55as_Probe	6-FAM-AATGTATTACACTGCTACTAAAG
Dp427b_F	TTGATTTGTTACAGCAGCCAACTT
Dp427b_R	CTTCCATGCCAGCTGTTTTTC
Dp427b_Probe	6-FAM TGGCATGATGGAGTGACA

Dp427m_F	GAAGAACTTTTACCAGGTTTTTTTTAT
Dp427m_R	CTTCTTCCCACCAAAGCATTTT
Dp427m_Probe	6-FAMTGCCTTGATATACACTTTT
Dp427p_F	CATAGAATGTGTAAGAGAAAAGTACCAACA
Dp427p_R	GCTGGCTACACACCTTCATAGGA
Dp427p_Probe	6-FAM AAATCAGCAAAAAGC
Dp71_F	TGCAGCCATGAGGGAACAG
Dp71_R	GGATGGTCCCAGCAAGTTGT
Dp71_Probe	6-FAM TCAAAGGCCACGAGACT

Human ACTB	4352035E (Life Technologies, Catalog Number)
Human GAPDH	4333764F (Life Technologies, Catalog Number)

Table S5

Name and sequence of the primers used to amplify the DMD gene isoforms in cDNA samples.
(doi:10.1371/journal.pone.0045328.s009)

Primer Name and Sequence	
Dp427bFw	CGTATCAGATAGTCAGAGTGGTTAC
Dp427mFw	CCTGGCATCAGTTACTGTGTTGAC
Dp427p2Fw	CCTATGAAGGTGTGTAGCCAGCC
Dp427Rv	CCATCTACGATGTCAGTACTTCC
Dp260Fw	AGGAACATTCGACCTGAGAAAG
Dp260Rv	TCCACCTTGTCTGCAATATAAGC
Dp140Fw	ATTGCTGGCTGCTCTGAACTAA
Dp140Rv	CATCTGTTTTTGAGGATTGCTG
Dp116Fw	GGGTTTTCTCAGGATTGCTATG
Dp116Rv	CCGGCTTAATTCATCATCTTTC
Dp71Fw	GAAGCTCACTCCTCCACTCGTA
Dp71Rv	AGCCAGTTCAGACACATATCCAC
GAPDH Fw	CATGGGGAAGGTGAAGGTC
GAPDH Rv	AACAATATCCACTTTACCAGAGTT

Table S6

Name and sequence of primers used for DMD ncRNAs cloning into pcDNA3.1(+)
(doi:10.1371/journal.pone.0045328.s010)

Primer Name and Sequence	
ncINT44s(EcoRI)_F	AGAATTCGGGATATGAGATGTTGGAAG
ncINT44s(EcoRI)_R	AGAATTCGTCTTGACTACAGATGTCTTTT
ncINT44s2(BamHI)_F	AGGATCCGCCACTAATTCTTATTGCCATTTC
ncINT44s2(BamHI)_R	AGGATCCAAAGGGCTGCTACAGCTATTTT
ncINT55s(EcoRI)_F	AGAATTCCATTTTGCTGCACAATAAACAACC
ncINT55s(EcoRI)_R	AGAATTCCAAGGCATTAAAATCTGCATGATG
nc3UTRas(EcoRI)_F	AGAATTCGGTGGTTATAAAGAACACAACACG
nc3UTRas(EcoRI)_R	AGAATTCGGCGTGATATCCATATGAAATTCAT

Table S7

Primer Name and Sequence used for cloning:

Dp427m Promoter in pGL3-basic; DMD intron34 or its relative deleted sequence downstream of the luciferase sequence driven by Dp427m Promoter in pGL3-basic; DMD exon45 or its relative deleted sequence downstream of the luciferase sequence driven by Dp427m Promoter in pGL3-basic

DMDm Prom.(KpnI)_ F	gatggtaccGGTGCTTTAGACATTACCCAGGACA
DMDm Prom.(BglII)_ R	tacagatctGGGGGAAAGTGAGTGATCCCAACA
DMD ex34/35 (BamHI)_F	catggatccTGACAAAGCAGTTCTTGAGAGAAGAG
DMD ex34/35 (BamHI)_R	gacggatccAGTGACTCCAATGGCATGTAACAGA
DMD ex45 (BamHI)_F	gacggatccAACATGGAACATCCTTGTGGGGACA
DMD ex45cl (BamHI)_R	catggatccAGGCTATAATTCTTTAACTTTGGCAAGGG
Int34 Δ1&Δ2 For	CTTTCTGTTTGCCTGTGGTTTTC
Int34 Δ1 Rev	ACAGGCGGTAGGATGAACATAC
Int34 Δ2 Rev	CAGACTGTAGTAGGTGATAGATAC
Ex45 Δ1&Δ2 For	CCAGTTGTTGATGTTAATGTGTCTTG
Ex45 Δ1 Rev	GAAGATATTCACCTTTAAGCAATC
Ex45 Δ2 Rev	TCCTGGAGTTCCTGTAAGATACC

Table S8

qChIP primers

Int 53 A For	GATGAAACTAATGGACTGTGACC
Int 53 A Rev	AGCCTGTATTGTGGAAGAAGC
Int 53 B For	GGTAAAGAACAAGGGCAAAGC
Int 53 B Rev	ATCAGTGACCACCCAAGAGG
Int 53 C For	TGGAAAACGACTGAACTTGAAG
Int 53 C Rev	AGAGATTATGAAAGAGAGGAGAGG
Ex62 A For	GTATCGTAATCACTCAATATCCTC
Ex62 A Rev	CACTCAGTATTTGTCACCTAAC
Ex62 B For	TTCTTAAAAGTCGTTCCCCATTG
Ex62 B Rev	AGAGCATTTATTGTGACTAACCTG
Ex62 C For	ATGAAGACAGGATGGATTAACC
Ex62 C Rev	GGGAGACACATACAAGAAACC
GAPDH TSS For	AAGACCTTGGGCTGGGACT
GAPDH TSS Rev	GCTGCGGGCTCAATTTATAG

References

Aartsma-Rus, A., I. Fokkema, J. Verschuuren, I. Ginjaar, J. van Deutekom, G. J. van Ommen, and J. T. den Dunnen, 2009, Theoretic applicability of antisense-mediated exon skipping for Duchenne muscular dystrophy mutations: Hum Mutat, v. 30, p. 293-9.

Alexander, R. D., S. A. Innocente, J. D. Barrass, and J. D. Beggs, 2010, Splicing-dependent RNA polymerase pausing in yeast: Molecular cell, v. 40, p. 582-93.

Aoki, Y., T. Yokota, T. Nagata, A. Nakamura, J. Tanihata, T. Saito, S. M. Duguez, K. Nagaraju, E. P. Hoffman, T. Partridge, and S. Takeda, 2012, Bodywide skipping of exons 45-55 in dystrophic mdx52 mice by systemic antisense delivery: Proc Natl Acad Sci U S A, v. 109, p. 13763-8.

Bar, S., E. Barnea, Z. Levy, S. Neuman, D. Yaffe, and U. Nudel, 1990, A novel product of the Duchenne muscular dystrophy gene which greatly differs from the known isoforms in its structure and tissue distribution: Biochem J, v. 272, p. 557-60.

Barski, A., S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao, 2007, High-resolution profiling of histone methylations in the human genome: Cell, v. 129, p. 823-37.

Bastianutto, C., J. A. Bestard, K. Lahnakoski, D. Broere, M. De Visser, M. Zaccolo, T. Pozzan, A. Ferlini, F. Muntoni, T. Patarnello, and H. J. Klamut, 2001, Dystrophin muscle enhancer 1 is implicated in the activation of non-muscle isoforms in the skeletal muscle of patients with X-linked dilated cardiomyopathy: Hum Mol Genet, v. 10, p. 2627-35.

Basu, U., O. Lozynska, C. Moorwood, G. Patel, S. D. Wilton, and T. S. Khurana, 2011, Translational regulation of utrophin by miRNAs: PLoS One, v. 6, p. e29376.

Benedetti, S., H. Hoshiya, and F. S. Tedesco, 2013, Repair or replace? Exploiting novel gene and cell therapy strategies for muscular dystrophies: FEBS J, v. 280, p. 4263-80.

Bertone, P., V. Stolc, T. E. Royce, J. S. Rozowsky, A. E. Urban, X. Zhu, J. L. Rinn, W. Tongprasit, M. Samanta, S. Weissman, M. Gerstein, and M. Snyder, 2004, Global identification of human transcribed sequences with genome tiling arrays: Science, v. 306, p. 2242-6.

Bevilacqua, A., M. C. Ceriani, S. Capaccioli, and A. Nicolin, 2003, Post-transcriptional regulation of gene expression by degradation of messenger RNAs: J Cell Physiol, v. 195, p. 356-72.

Bintu, L., M. Kopaczynska, C. Hodges, L. Lubkowska, M. Kashlev, and C. Bustamante, 2011, The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes: Nat Struct Mol Biol, v. 18, p. 1394-9.

Bovolenta, M., D. Erriquez, E. Valli, S. Brioschi, C. Scotton, M. Neri, M. S. Falzarano, S. Gherardi, M. Fabris, P. Rimessi, F. Gualandi, G. Perini, and A. Ferlini, 2012a, The DMD locus harbours multiple long non-coding RNAs which orchestrate and control transcription of muscle dystrophin mRNA isoforms: PLoS One, v. 7, p. e45328.

Bovolenta, M., C. Scotton, M. S. Falzarano, F. Gualandi, and A. Ferlini, 2012b, Rapid, comprehensive analysis of the dystrophin transcript by a custom micro-fluidic exome array: Hum Mutat, v. 33, p. 572-81.

Brioschi, S., F. Gualandi, C. Scotton, A. Armaroli, M. Bovolenta, M. S. Falzarano, P. Sabatelli, R. Selvatici, A. D'Amico, M. Pane, G. Ricci, G. Siciliano, S. Tedeschi, A. Pini, L. Vercelli, D. De Grandis, E. Mercuri, E. Bertini, L. Merlini, T. Mongini, and A. Ferlini, 2012, Genetic characterization in symptomatic female DMD carriers: lack of relationship between X-inactivation, transcriptional DMD allele balancing and phenotype: BMC Med Genet, v. 13, p. 73.

Brodsky, A. S., and P. A. Silver, 2000, Pre-mRNA processing factors are required for nuclear export: RNA, v. 6, p. 1737-49.

Bulger, M., and M. Groudine, 2011, Functional and mechanistic diversity of distal transcription enhancers: Cell, v. 144, p. 327-39.

Bushby, K. M., and D. Gardner-Medwin, 1993, The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy. I. Natural history: J Neurol, v. 240, p. 98-104.

Bushby, K. M., M. Thambyayah, and D. Gardner-Medwin, 1991, Prevalence and incidence of Becker muscular dystrophy: Lancet, v. 337, p. 1022-4.

Byers, T. J., H. G. Lidov, and L. M. Kunkel, 1993, An alternative dystrophin transcript specific to peripheral nerve: Nat Genet, v. 4, p. 77-81.

Cabianca, D. S., V. Casa, B. Bodega, A. Xynos, E. Ginelli, Y. Tanaka, and D. Gabellini, 2012, A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy: Cell, v. 149, p. 819-31.

Cacchiarelli, D., T. Incitti, J. Martone, M. Cesana, V. Cazzella, T. Santini, O. Sthandier, and I. Bozzoni, 2011a, miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy: EMBO Rep, v. 12, p. 136-41.

Cacchiarelli, D., I. Legnini, J. Martone, V. Cazzella, A. D'Amico, E. Bertini, and I. Bozzoni, 2011b, miRNAs as serum biomarkers for Duchenne muscular dystrophy: EMBO Mol Med, v. 3, p. 258-65.

Cacchiarelli, D., J. Martone, E. Girardi, M. Cesana, T. Incitti, M. Morlando, C. Nicoletti, T. Santini, O. Sthandier, L. Barberi, A. Auricchio, A. Musarò, and I. Bozzoni, 2010, MicroRNAs involved in molecular circuitries relevant for the Duchenne muscular dystrophy pathogenesis are controlled by the dystrophin/nNOS pathway: Cell Metab, v. 12, p. 341-51.

Caretti, G., E. P. Lei, and V. Sartorelli, 2007, The DEAD-box p68/p72 proteins and the noncoding RNA steroid receptor activator SRA: eclectic regulators of disparate biological functions: Cell Cycle, v. 6, p. 1172-6.

Caretti, G., R. L. Schiltz, F. J. Dilworth, M. Di Padova, P. Zhao, V. Ogryzko, F. V. Fuller-Pace, E. P. Hoffman, S. J. Tapscott, and V. Sartorelli, 2006, The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation: Dev Cell, v. 11, p. 547-60.

Carmody, S. R., and S. R. Wente, 2009, mRNA nuclear export at a glance: J Cell Sci, v. 122, p. 1933-7.

Carneiro, M., and U. Schibler, 1984, Accumulation of rare and moderately abundant mRNAs in mouse L-cells is mainly post-transcriptionally regulated: J Mol Biol, v. 178, p. 869-80.

Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V. B. Bajic, S. E. Brenner, S. Batalov, A. R. Forrest, M. Zavolan, M. J. Davis, L. G. Wilming, V. Aidinis, J. E. Allen, A. Ambesi-Impiombato, R. Apweiler, R. N. Aturaliya, T. L. Bailey, M. Bansal, L. Baxter, K. W. Beisel, T. Bersano, H. Bono, A. M. Chalk, K. P. Chiu, V. Choudhary, A. Christoffels, D. R. Clutterbuck, M. L. Crowe, E. Dalla, B. P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C. F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T. R. Gingeras, T. Gojobori, R. E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T. K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S. P. Krishnan, A. Kruger, S. K. Kummerfeld, I. V. Kurochkin, L. F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, et al., 2005, The transcriptional landscape of the mammalian genome: Science, v. 309, p. 1559-63.

Carrillo Oesterreich, F., S. Preibisch, and K. M. Neugebauer, 2010, Global analysis of nascent RNA reveals transcriptional pausing in terminal exons: Mol Cell, v. 40, p. 571-81.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman, 2005, Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription: Cell, v. 123, p. 581-92.

Cesana, M., D. Cacchiarelli, I. Legnini, T. Santini, O. Sthandier, M. Chinappi, A. Tramontano, and I. Bozzoni, 2011, A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA: Cell, v. 147, p. 358-69.

Chelly, J., H. Gilgenkrantz, M. Lambert, G. Hamard, P. Chafey, D. Récan, P. Katz, A. de la Chapelle, M. Koenig, and I. B. Ginjaar, 1990, Effect of dystrophin gene deletions on mRNA levels and processing in Duchenne and Becker muscular dystrophies: Cell, v. 63, p. 1239-48.

Chelly, J., J. C. Kaplan, P. Maire, S. Gautron, and A. Kahn, 1988, Transcription of the dystrophin gene in human muscle and non-muscle tissue: Nature, v. 333, p. 858-60.

Chen, J. F., T. E. Callis, and D. Z. Wang, 2009, microRNAs and muscle disorders: J Cell Sci, v. 122, p. 13-20.

Cheng, J., P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, J. Long, D. Stern, H. Tammana, G. Helt, V. Sementchenko, A. Piccolboni, S. Bekiranov, D. K. Bailey, M. Ganesh, S. Ghosh, I. Bell, D. S. Gerhard, and T. R. Gingeras, 2005, Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution: Science, v. 308, p. 1149-54.

Chevron, M. P., F. Girard, M. Claustres, and J. Demaille, 1994, Expression and subcellular localization of dystrophin in skeletal, cardiac and smooth muscles during the human development: Neuromuscul Disord, v. 4, p. 419-32.

Cirak, S., V. Arechavala-Gomeza, M. Guglieri, L. Feng, S. Torelli, K. Anthony, S. Abbs, M. E. Garralda, J. Bourke, D. J. Wells, G. Dickson, M. J. Wood, S. D. Wilton, V. Straub, R. Kole, S. B. Shrewsbury, C. Sewry, J. E. Morgan, K. Bushby, and F. Muntoni, 2011, Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study: Lancet, v. 378, p. 595-605.

Clemson, C. M., J. N. Hutchinson, S. A. Sara, A. W. Ensminger, A. H. Fox, A. Chess, and J. B. Lawrence, 2009, An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles: Mol Cell, v. 33, p. 717-26.

Coffey, A. J., R. G. Roberts, E. D. Green, C. G. Cole, R. Butler, R. Anand, F. Giannelli, and D. R. Bentley, 1992, Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach: Genomics, v. 12, p. 474-84.

Cullen, B. R., 2003, Nuclear mRNA export: insights from virology: Trends Biochem Sci, v. 28, p. 419-24.

D'Souza, V. N., T. M. Nguyen, G. E. Morris, W. Karges, D. A. Pillers, and P. N. Ray, 1995, A novel dystrophin isoform is required for normal retinal electrophysiology: Hum Mol Genet, v. 4, p. 837-42.

Davies, K. E., and K. J. Nowak, 2006, Molecular mechanisms of muscular dystrophies: old and new players: Nat Rev Mol Cell Biol, v. 7, p. 762-73.

de Almeida, S. F., A. R. Grosso, F. Koch, R. Fenouil, S. Carvalho, J. Andrade, H. Levezinho, M. Gut, D. Eick, I. Gut, J. C. Andrau, P. Ferrier, and M. Carmo-Fonseca, 2011, Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36: Nat Struct Mol Biol, v. 18, p. 977-83.

de la Mata, M., C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley, and A. R. Kornblihtt, 2003, A slow RNA polymerase II affects alternative splicing in vivo: Mol Cell, v. 12, p. 525-32.

Dunckley, M. G., M. Manoharan, P. Villiet, I. C. Eperon, and G. Dickson, 1998, Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides: Hum Mol Genet, v. 7, p. 1083-90.

Durbeej, M., and K. P. Campbell, 2002, Muscular dystrophies involving the dystrophin-glycoprotein complex: an overview of current mouse models: Curr Opin Genet Dev, v. 12, p. 349-61.

Eberharter, A., and P. B. Becker, 2002, Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics: EMBO Rep, v. 3, p. 224-9.

Egloff, S., M. Dienstbier, and S. Murphy, 2012, Updating the RNA polymerase CTD code: adding gene-specific layers: Trends Genet, v. 28, p. 333-41.

Eisenberg, I., M. S. Alexander, and L. M. Kunkel, 2009, miRNAS in normal and diseased skeletal muscle: J Cell Mol Med, v. 13, p. 2-11.

Eißmann, M., T. Gutschner, M. Hämmerle, S. Günther, M. Caudron-Herger, M. Groß, P. Schirmacher, K. Rippe, T. Braun, M. Zörnig, and S. Diederichs, 2012, Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development: RNA Biol, v. 9, p. 1076-87.

Emery, A. E., 1991, Population frequencies of inherited neuromuscular diseases--a world survey: Neuromuscul Disord, v. 1, p. 19-29.

Erriquez, D., G. Perini, and A. Ferlini, 2013, Non-coding RNAs in muscle dystrophies: Int J Mol Sci, v. 14, p. 19681-704.

Ervasti, J. M., and K. J. Sonnemann, 2008, Biology of the striated muscle dystrophin-glycoprotein complex: Int Rev Cytol, v. 265, p. 191-225.

Fairclough, R. J., M. J. Wood, and K. E. Davies, 2013, Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches: Nat Rev Genet, v. 14, p. 373-8.

Ferlini, A., M. Neri, and F. Gualandi, 2013, The medical genetics of dystrophinopathies: molecular genetic diagnosis and its impact on clinical practice: Neuromuscul Disord, v. 23, p. 4-14.
Galvagni, F., E. Cartocci, and S. Oliviero, 1998, The dystrophin promoter is negatively regulated by YY1 in undifferentiated muscle cells: J Biol Chem, v. 273, p. 33708-13.
Galvagni, F., M. Lestingi, E. Cartocci, and S. Oliviero, 1997, Serum response factor and protein-mediated DNA bending contribute to transcription of the dystrophin muscle-specific promoter: Mol Cell Biol, v. 17, p. 1731-43.
Ge, Y., and J. Chen, 2011, MicroRNAs in skeletal myogenesis: Cell Cycle, v. 10, p. 441-8.
Gilgenkrantz, H., J. P. Hugnot, M. Lambert, P. Chafey, J. C. Kaplan, and A. Kahn, 1992, Positive and negative regulatory DNA elements including a CCArGG box are involved in the cell type-specific expression of the human muscle dystrophin gene: J Biol Chem, v. 267, p. 10823-30.
Ginger, M. R., A. N. Shore, A. Contreras, M. Rijnkels, J. Miller, M. F. Gonzalez-Rimbau, and J. M. Rosen, 2006, A noncoding RNA is a potential marker of cell fate during mammary gland development: Proc Natl Acad Sci U S A, v. 103, p. 5781-6.
Goemans, N. M., M. Tulinius, J. T. van den Akker, B. E. Burm, P. F. Ekhart, N. Heuvelmans, T. Holling, A. A. Janson, G. J. Platenburg, J. A. Sipkens, J. M. Sitsen, A. Aartsma-Rus, G. J. van Ommen, G. Buyse, N. Darin, J. J. Verschuuren, G. V. Campion, S. J. de Kimpe, and J. C. van Deutekom, 2011, Systemic administration of PRO051 in Duchenne's muscular dystrophy: N Engl J Med, v. 364, p. 1513-22.
Gondran, P., F. Amiot, D. Weil, and F. Dautry, 1999, Accumulation of mature mRNA in the nuclear fraction of mammalian cells: FEBS Lett, v. 458, p. 324-8.
Goyenvalle, A., A. Babbs, D. Powell, R. Kole, S. Fletcher, S. D. Wilton, and K. E. Davies, 2010, Prevention of dystrophic pathology in severely affected dystrophin/utrophin-deficient mice by morpholino-oligomer-mediated exon-skipping: Mol Ther, v. 18, p. 198-205.
Grüter, P., C. Tabernero, C. von Kobbe, C. Schmitt, C. Saavedra, A. Bachi, M. Wilm, B. K. Felber, and E. Izaurralde, 1998, TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus: Mol Cell, v. 1, p. 649-59.
Gurvich, O. L., B. Maiti, R. B. Weiss, G. Aggarwal, M. T. Howard, and K. M. Flanigan, 2009, DMD exon 1 truncating point mutations: amelioration of phenotype by alternative translation initiation in exon 6: Hum Mutat, v. 30, p. 633-40.
Gutschner, T., M. Hämmerle, and S. Diederichs, 2013, MALAT1 -- a paradigm for long noncoding RNA function in cancer: J Mol Med (Berl), v. 91, p. 791-801.
He, H. H., C. A. Meyer, H. Shin, S. T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien, P. Mieczkowski, J. D. Lieb, K. Zhao, M. Brown, and X. S. Liu, 2010, Nucleosome dynamics define transcriptional enhancers: Nat Genet, v. 42, p. 343-7.
Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis, and B. Ren, 2009, Histone modifications at human enhancers reflect global cell-type-specific gene expression: Nature, v. 459, p. 108-12.
Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren, 2007, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome: Nat Genet, v. 39, p. 311-8.
Hermans, M. C., Y. M. Pinto, I. S. Merkies, C. E. de Die-Smulders, H. J. Crijns, and C. G. Faber, 2010, Hereditary muscular dystrophies and the heart: Neuromuscul Disord, v. 20, p. 479-92.
Hilleren, P., T. McCarthy, M. Rosbash, R. Parker, and T. H. Jensen, 2001, Quality control of mRNA 3'-end processing is linked to the nuclear exosome: Nature, v. 413, p. 538-42.
Howe, K. J., C. M. Kane, and M. Ares, 2003, Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae: RNA, v. 9, p. 993-1006.
Hu, G., D. E. Schones, K. Cui, R. Ybarra, D. Northrup, Q. Tang, L. Gattinoni, N. P. Restifo, S. Huang, and K. Zhao, 2011, Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1: Genome Res, v. 21, p. 1650-8.

Huang, Y., R. Gattoni, J. Stévenin, and J. A. Steitz, 2003, SR splicing factors serve as adapter proteins for TAP-dependent mRNA export: Mol Cell, v. 11, p. 837-43.
Huang, Y., and J. A. Steitz, 2005, SRprises along a messenger's journey: Mol Cell, v. 17, p. 613-5.
Hubé, F., G. Velasco, J. Rollin, D. Furling, and C. Francastel, 2011, Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation: Nucleic Acids Res, v. 39, p. 513-25.
Iarovaia, O. V., V. Borounova, Y. S. Vassetzky, and S. V. Razin, 2006, An unusual extended DNA loop attachment region is located in the human dystrophin gene: J Cell Physiol, v. 209, p. 515-21.
Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, I. Bell, E. Cheung, J. Drenkow, E. Dumais, S. Patel, G. Helt, M. Ganesh, S. Ghosh, A. Piccolboni, V. Sementchenko, H. Tammana, and T. R. Gingeras, 2007, RNA maps reveal new RNA classes and a possible function for pervasive transcription: Science, v. 316, p. 1484-8.
Kim, S., H. Kim, N. Fong, B. Erickson, and D. L. Bentley, 2011, Pre-mRNA splicing is a determinant of histone H3K36 methylation: Proc Natl Acad Sci U S A, v. 108, p. 13564-9.
Kimura, S., K. Abe, M. Suzuki, M. Ogawa, K. Yoshioka, T. Kaname, T. Miike, and K. Yamamura, 1997, A 900 bp genomic region from the mouse dystrophin promoter directs lacZ reporter expression only to the right heart of transgenic mice: Dev Growth Differ, v. 39, p. 257-65.
Kinali, M., V. Arechavala-Gomeza, L. Feng, S. Cirak, D. Hunt, C. Adkin, M. Guglieri, E. Ashton, S. Abbs, P. Nihoyannopoulos, M. E. Garralda, M. Rutherford, C. McCulley, L. Popplewell, I. R. Graham, G. Dickson, M. J. Wood, D. J. Wells, S. D. Wilton, R. Kole, V. Straub, K. Bushby, C. Sewry, J. E. Morgan, and F. Muntoni, 2009, Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study: Lancet Neurol, v. 8, p. 918-28.
Klamut, H. J., L. O. Bosnoyan-Collins, R. G. Worton, P. N. Ray, and H. L. Davis, 1996, Identification of a transcriptional enhancer within muscle intron 1 of the human dystrophin gene: Hum Mol Genet, v. 5, p. 1599-606.
Klamut, H. J., S. B. Gangopadhyay, R. G. Worton, and P. N. Ray, 1990, Molecular and functional analysis of the muscle-specific promoter region of the Duchenne muscular dystrophy gene: Mol Cell Biol, v. 10, p. 193-205.
Koenig, M., E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener, and L. M. Kunkel, 1987, Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals: Cell, v. 50, p. 509-17.
Kolasinska-Zwierz, P., T. Down, I. Latorre, T. Liu, X. S. Liu, and J. Ahringer, 2009, Differential chromatin marking of introns and expressed exons by H3K36me3: Nat Genet, v. 41, p. 376-81.
Kristjuhan, A., and J. Q. Svejstrup, 2004, Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo: EMBO J, v. 23, p. 4243-52.
Krogan, N. J., M. Kim, A. Tong, A. Golshani, G. Cagney, V. Canadien, D. P. Richards, B. K. Beattie, A. Emili, C. Boone, A. Shilatifard, S. Buratowski, and J. Greenblatt, 2003, Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II: Mol Cell Biol, v. 23, p. 4207-18.
Köhler, A., and E. Hurt, 2007, Exporting RNA from the nucleus to the cytoplasm: Nat Rev Mol Cell Biol, v. 8, p. 761-73.
Langley, B., M. Thomas, A. Bishop, M. Sharma, S. Gilmour, and R. Kambadur, 2002, Myostatin inhibits myoblast differentiation by down-regulating MyoD expression: J Biol Chem, v. 277, p. 49831-40.
Larson, M. H., R. Landick, and S. M. Block, 2011, Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes: Mol Cell, v. 41, p. 249-62.
Le Hir, H., D. Gatfield, E. Izaurralde, and M. J. Moore, 2001, The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay: EMBO J, v. 20, p. 4987-97.
Lee, J. T., 2009, Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome: Genes Dev, v. 23, p. 1831-42.

Lei, E. P., and P. A. Silver, 2002, Intron status and 3'-end formation control cotranscriptional export of mRNA: Genes Dev, v. 16, p. 2761-6.
Li, B., M. Gogol, M. Carey, D. Lee, C. Seidel, and J. L. Workman, 2007, Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin: Science, v. 316, p. 1050-4.
Li, B., L. Howe, S. Anderson, J. R. Yates, and J. L. Workman, 2003, The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II: J Biol Chem, v. 278, p. 8897-903.
Li, X., Z. Wu, X. Fu, and W. Han, 2013, Long Noncoding RNAs: Insights from Biological Features and Functions to Diseases: Med Res Rev, v. 33, p. 517-53.
Libri, D., K. Dower, J. Boulay, R. Thomsen, M. Rosbash, and T. H. Jensen, 2002, Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation: Mol Cell Biol, v. 22, p. 8254-66.
Lidov, H. G., S. Selig, and L. M. Kunkel, 1995, Dp140: a novel 140 kDa CNS transcript from the dystrophin locus: Hum Mol Genet, v. 4, p. 329-35.
Liu, N., A. H. Williams, Y. Kim, J. McAnally, S. Bezprozvannaya, L. B. Sutherland, J. A. Richardson, R. Bassel-Duby, and E. N. Olson, 2007, An intragenic MEF2-dependent enhancer directs muscle-specific expression of microRNAs 1 and 133: Proc Natl Acad Sci U S A, v. 104, p. 20844-9.
Luo, M. L., Z. Zhou, K. Magni, C. Christoforides, J. Rappsilber, M. Mann, and R. Reed, 2001, Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly: Nature, v. 413, p. 644-7.
Mamchaoui, K., C. Trollet, A. Bigot, E. Negroni, S. Chaouch, A. Wolff, P. K. Kandalla, S. Marie, J. Di Santo, J. L. St Guily, F. Muntoni, J. Kim, S. Philippi, S. Spuler, N. Levy, S. C. Blumen, T. Voit, W. E. Wright, A. Aamiri, G. Butler-Browne, and V. Mouly, 2011, Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders: Skelet Muscle, v. 1, p. 34.
Marrone, A. K., and H. R. Shcherbata, 2011, Dystrophin Orchestrates the Epigenetic Profile of Muscle Cells Via miRNAs: Front Genet, v. 2, p. 64.
Marshall, P., N. Chartrand, Y. De Repentigny, R. Kothary, L. Pelletier, R. Mueller, and R. G. Worton, 2002, Mouse dystrophin enhancer preferentially targets lacZ expression in skeletal and cardiac muscle: Dev Dyn, v. 224, p. 30-8.
Martianov, I., A. Ramadass, A. Serra Barros, N. Chow, and A. Akoulitchev, 2007, Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript: Nature, v. 445, p. 666-70.
Mattick, J. S., and I. V. Makunin, 2006, Non-coding RNA: Hum Mol Genet, v. 15 Spec No 1, p. R17-29.
McNaughton, J. C., D. J. Cockburn, G. Hughes, W. A. Jones, N. G. Laing, P. N. Ray, P. A. Stockwell, and G. B. Petersen, 1998, Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene: Gene, v. 222, p. 41-51.
McNaughton, J. C., G. Hughes, W. A. Jones, P. A. Stockwell, H. J. Klamut, and G. B. Petersen, 1997, The evolution of an intron: analysis of a long, deletion-prone intron in the human dystrophin gene: Genomics, v. 40, p. 294-304.
Mercer, T. R., M. E. Dinger, and J. S. Mattick, 2009, Long non-coding RNAs: insights into functions: Nat Rev Genet, v. 10, p. 155-9.
Mercuri, E., and F. Muntoni, 2013, Muscular dystrophies: Lancet, v. 381, p. 845-60.
Miranda, A. F., E. Bonilla, G. Martucci, C. T. Moraes, A. P. Hays, and S. Dimauro, 1988, Immunocytochemical study of dystrophin in muscle cultures from patients with Duchenne muscular dystrophy and unaffected control patients: Am J Pathol, v. 132, p. 410-6.
Monaco, A. P., C. J. Bertelson, S. Liechti-Gallati, H. Moser, and L. M. Kunkel, 1988, An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus: Genomics, v. 2, p. 90-5.
Monaco, A. P., A. P. Walker, I. Millwood, Z. Larin, and H. Lehrach, 1992, A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene: Genomics, v. 12, p. 465-73.
Muntoni, F., S. Torelli, and A. Ferlini, 2003, Dystrophin and mutations: one gene, several proteins, multiple phenotypes: Lancet Neurol, v. 2, p. 731-40.

Nachman, M. W., and S. L. Crowell, 2000, Contrasting evolutionary histories of two introns of the duchenne muscular dystrophy gene, Dmd, in humans: Genetics, v. 155, p. 1855-64.

Nakaya, H. I., P. P. Amaral, R. Louro, A. Lopes, A. A. Fachel, Y. B. Moreira, T. A. El-Jundi, A. M. da Silva, E. M. Reis, and S. Verjovski-Almeida, 2007, Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription: Genome Biol, v. 8, p. R43.

Neri, M., S. Torelli, S. Brown, I. Ugo, P. Sabatelli, L. Merlini, P. Spitali, P. Rimessi, F. Gualandi, C. Sewry, A. Ferlini, and F. Muntoni, 2007, Dystrophin levels as low as 30% are sufficient to avoid muscular dystrophy in the human: Neuromuscul Disord, v. 17, p. 913-8.

Ng, H. H., F. Robert, R. A. Young, and K. Struhl, 2003, Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity: Mol Cell, v. 11, p. 709-19.

Nobile, C., F. Galvagni, J. Marchi, R. Roberts, and L. Vitiello, 1995, Genomic organization of the human dystrophin gene across the major deletion hot spot and the 3' region: Genomics, v. 28, p. 97-100.

Ohira, M., A. Morohashi, H. Inuzuka, T. Shishikura, T. Kawamoto, H. Kageyama, Y. Nakamura, E. Isogai, H. Takayasu, S. Sakiyama, Y. Suzuki, S. Sugano, T. Goto, S. Sato, and A. Nakagawara, 2003, Expression profiling and characterization of 4200 genes cloned from primary neuroblastomas: identification of 305 genes differentially expressed between favorable and unfavorable subsets: Oncogene, v. 22, p. 5525-36.

Ong, C. T., and V. G. Corces, 2011, Enhancer function: new insights into the regulation of tissue-specific gene expression: Nat Rev Genet, v. 12, p. 283-93.

Ota, T., Y. Suzuki, T. Nishikawa, T. Otsuki, T. Sugiyama, R. Irie, A. Wakamatsu, K. Hayashi, H. Sato, K. Nagai, K. Kimura, H. Makita, M. Sekine, M. Obayashi, T. Nishi, T. Shibahara, T. Tanaka, S. Ishii, J. Yamamoto, K. Saito, Y. Kawai, Y. Isono, Y. Nakamura, K. Nagahari, K. Murakami, T. Yasuda, T. Iwayanagi, M. Wagatsuma, A. Shiratori, H. Sudo, T. Hosoiri, Y. Kaku, H. Kodaira, H. Kondo, M. Sugawara, M. Takahashi, K. Kanda, T. Yokoi, T. Furuya, E. Kikkawa, Y. Omura, K. Abe, K. Kamihara, N. Katsuta, K. Sato, M. Tanikawa, M. Yamazaki, K. Ninomiya, T. Ishibashi, H. Yamashita, K. Murakawa, K. Fujimori, H. Tanai, M. Kimata, M. Watanabe, S. Hiraoka, Y. Chiba, S. Ishida, Y. Ono, S. Takiguchi, S. Watanabe, M. Yosida, T. Hotuta, J. Kusano, K. Kanehori, A. Takahashi-Fujii, H. Hara, T. O. Tanase, Y. Nomura, S. Togiya, F. Komai, R. Hara, K. Takeuchi, M. Arita, N. Imose, K. Musashino, H. Yuuki, A. Oshima, N. Sasaki, S. Aotsuka, Y. Yoshikawa, H. Matsunawa, T. Ichihara, N. Shiohata, S. Sano, S. Moriya, H. Momiyama, N. Satoh, S. Takami, Y. Terashima, O. Suzuki, S. Nakagawa, A. Senoh, H. Mizoguchi, Y. Goto, F. Shimizu, H. Wakebe, H. Hishigaki, T. Watanabe, A. Sugiyama, et al., 2004, Complete sequencing and characterization of 21,243 full-length human cDNAs: Nat Genet, v. 36, p. 40-5.

Pang, K. C., M. C. Frith, and J. S. Mattick, 2006, Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function: Trends Genet, v. 22, p. 1-5.

Pekowska, A., T. Benoukraf, J. Zacarias-Cabeza, M. Belhocine, F. Koch, H. Holota, J. Imbert, J. C. Andrau, P. Ferrier, and S. Spicuglia, 2011, H3K4 tri-methylation provides an epigenetic signature of active enhancers: EMBO J, v. 30, p. 4198-210.

Petrof, B. J., J. B. Shrager, H. H. Stedman, A. M. Kelly, and H. L. Sweeney, 1993, Dystrophin protects the sarcolemma from stresses developed during muscle contraction: Proc Natl Acad Sci U S A, v. 90, p. 3710-4.

Pillers, D. A., D. E. Bulman, R. G. Weleber, D. A. Sigesmund, M. A. Musarella, B. R. Powell, W. H. Murphey, C. Westall, C. Panton, and L. E. Becker, 1993, Dystrophin expression in the human retina is required for normal function as defined by electroretinography: Nat Genet, v. 4, p. 82-6.

Ponting, C. P., P. L. Oliver, and W. Reik, 2009, Evolution and functions of long noncoding RNAs: Cell, v. 136, p. 629-41.

Rao, P. K., R. M. Kumar, M. Farkhondeh, S. Baskerville, and H. F. Lodish, 2006, Myogenic factors that regulate expression of muscle-specific microRNAs: Proc Natl Acad Sci U S A, v. 103, p. 8721-6.

Rapaport, D., D. Lederfein, J. T. den Dunnen, P. M. Grootscholten, G. J. Van Ommen, O. Fuchs, U. Nudel, and D. Yaffe, 1992, Characterization and cell type distribution of a novel, major transcript of the Duchenne muscular dystrophy gene: Differentiation, v. 49, p. 187-93.

Redrup, L., M. R. Branco, E. R. Perdeaux, C. Krueger, A. Lewis, F. Santos, T. Nagano, B. S. Cobb, P. Fraser, and W. Reik, 2009, The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing: Development, v. 136, p. 525-30.

Reed, R., and H. Cheng, 2005, TREX, SR proteins and export of mRNA: Curr Opin Cell Biol, v. 17, p. 269-73.

Reed, R., and E. Hurt, 2002, A conserved mRNA export machinery coupled to pre-mRNA splicing: Cell, v. 108, p. 523-31.

Rinn, J. L., M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang, 2007, Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs: Cell, v. 129, p. 1311-23.

Roberts, R. G., A. J. Coffey, M. Bobrow, and D. R. Bentley, 1992, Determination of the exon structure of the distal portion of the dystrophin gene by vectorette PCR: Genomics, v. 13, p. 942-50.

Robertson, N. G., U. Khetarpal, G. A. Gutiérrez-Espeleta, F. R. Bieber, and C. C. Morton, 1994, Isolation of novel and known genes from a human fetal cochlear cDNA library using subtractive hybridization and differential screening: Genomics, v. 23, p. 42-50.

Rosenberg, M. I., S. A. Georges, A. Asawachaicharn, E. Analau, and S. J. Tapscott, 2006, MyoD inhibits Fstl1 and Utrn expression by inducing transcription of miR-206: J Cell Biol, v. 175, p. 77-85.

Sadoulet-Puccio, H. M., and L. M. Kunkel, 1996, Dystrophin and its isoforms: Brain Pathol, v. 6, p. 25-35.

Saint-André, V., E. Batsché, C. Rachez, and C. Muchardt, 2011, Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons: Nat Struct Mol Biol, v. 18, p. 337-44.

Sasaki, Y. T., T. Ideue, M. Sano, T. Mituyama, and T. Hirose, 2009, MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles: Proc Natl Acad Sci U S A, v. 106, p. 2525-30.

Schones, D. E., K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao, 2008, Dynamic regulation of nucleosome positioning in the human genome: Cell, v. 132, p. 887-98.

Schwabish, M. A., and K. Struhl, 2004, Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II: Mol Cell Biol, v. 24, p. 10111-7.

Segref, A., K. Sharma, V. Doye, A. Hellwig, J. Huber, R. Lührmann, and E. Hurt, 1997, Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores: EMBO J, v. 16, p. 3256-71.

Selth, L. A., S. Sigurdsson, and J. Q. Svejstrup, 2010, Transcript Elongation by RNA Polymerase II: Annu Rev Biochem, v. 79, p. 271-93.

Sharma, S., G. M. Findlay, H. S. Bandukwala, S. Oberdoerffer, B. Baust, Z. Li, V. Schmidt, P. G. Hogan, D. B. Sacks, and A. Rao, 2011, Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex: Proc Natl Acad Sci U S A, v. 108, p. 11381-6.

Shukla, S., E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, and S. Oberdoerffer, 2011, CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing: Nature, v. 479, p. 74-9.

Shukla, S., and S. Oberdoerffer, 2012, Co-transcriptional regulation of alternative pre-mRNA splicing: Biochim Biophys Acta, v. 1819, p. 673-83.

Sims, R. J., and D. Reinberg, 2009, Processing the H3K36me3 signature: Nat Genet, v. 41, p. 270-1.

Spitali, P., P. Rimessi, M. Fabris, D. Perrone, S. Falzarano, M. Bovolenta, C. Trabanelli, L. Mari, E. Bassi, S. Tuffery, F. Gualandi, N. M. Maraldi, P. Sabatelli-Giraud, A. Medici, L. Merlini, and A. Ferlini, 2009, Exon skipping-mediated dystrophin reading frame restoration for small mutations: Hum Mutat, v. 30, p. 1527-34.

Spitali, P., J. C. van den Bergen, I. E. Verhaart, B. Wokke, A. A. Janson, R. van den Eijnde, J. T. den Dunnen, J. F. Laros, J. J. Verschuuren, P. A. 't Hoen, and A. Aartsma-Rus, 2013, DMD transcript imbalance determines dystrophin levels: FASEB J, v. 27, p. 4909-16.

Stutz, F., and E. Izaurralde, 2003, The interplay of nuclear mRNP assembly, mRNA surveillance and export: Trends Cell Biol, v. 13, p. 319-27.

Sun, Y., Y. Ge, J. Drnevich, Y. Zhao, M. Band, and J. Chen, 2010, Mammalian target of rapamycin regulates miRNA-1 and follistatin in skeletal myogenesis: J Cell Biol, v. 189, p. 1157-69.
Surono, A., Y. Takeshima, T. Wibawa, M. Ikezawa, I. Nonaka, and M. Matsuo, 1999, Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing: Hum Mol Genet, v. 8, p. 493-500.
Suzuki, H., T. Kameyama, K. Ohe, T. Tsukahara, and A. Mayeda, 2013, Nested introns in an intron: evidence of multi-step splicing in a large intron of the human dystrophin pre-mRNA: FEBS Lett, v. 587, p. 555-61.
Tennyson, C. N., H. J. Klamut, and R. G. Worton, 1995, The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced: Nat Genet, v. 9, p. 184-90.
Tennyson, C. N., Q. Shi, and R. G. Worton, 1996, Stability of the human dystrophin transcript in muscle: Nucleic Acids Res, v. 24, p. 3059-64.
Torelli, S., A. Ferlini, L. Obici, C. Sewry, and F. Muntoni, 1999, Expression, regulation and localisation of dystrophin isoforms in human foetal skeletal and cardiac muscle: Neuromuscul Disord, v. 9, p. 541-51.
Tripathi, V., J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth, and K. V. Prasanth, 2010, The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation: Mol Cell, v. 39, p. 925-38.
van Deutekom, J. C., A. A. Janson, I. B. Ginjaar, W. S. Frankhuizen, A. Aartsma-Rus, M. Bremmer-Bout, J. T. den Dunnen, K. Koop, A. J. van der Kooi, N. M. Goemans, S. J. de Kimpe, P. F. Ekhart, E. H. Venneker, G. J. Platenburg, J. J. Verschuuren, and G. J. van Ommen, 2007, Local dystrophin restoration with antisense oligonucleotide PRO051: N Engl J Med, v. 357, p. 2677-86.
van Rooij, E., N. Liu, and E. N. Olson, 2008, MicroRNAs flex their muscles: Trends Genet, v. 24, p. 159-66.
Wagner, L. A., C. J. Christensen, D. M. Dunn, G. J. Spangrude, A. Georgelas, L. Kelley, M. S. Esplin, R. B. Weiss, and G. J. Gleich, 2007, EGO, a novel, noncoding RNA gene, regulates eosinophil granule protein transcript expression: Blood, v. 109, p. 5191-8.
Wallace, G. Q., and E. M. McNally, 2009, Mechanisms of muscle degeneration, regeneration, and repair in the muscular dystrophies: Annu Rev Physiol, v. 71, p. 37-57.
Wang, X. Q., J. L. Crutchley, and J. Dostie, 2011, Shaping the Genome with Non-Coding RNAs: Curr Genomics, v. 12, p. 307-21.
Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao, 2008, Combinatorial patterns of histone acetylations and methylations in the human genome: Nat Genet, v. 40, p. 897-903.
Watts, R., V. L. Johnsen, J. Shearer, and D. S. Hittel, 2013, Myostatin-induced inhibition of the long noncoding RNA Malat1 is associated with decreased myogenesis: Am J Physiol Cell Physiol, v. 304, p. C995-1001.
Weinmann, A. S., and P. J. Farnham, 2002, Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation: Methods, v. 26, p. 37-47.
Weintraub, H., 1993, The MyoD family and myogenesis: redundancy, networks, and thresholds: Cell, v. 75, p. 1241-4.
Willingham, A. T., A. P. Orth, S. Batalov, E. C. Peters, B. G. Wen, P. Aza-Blanc, J. B. Hogenesch, and P. G. Schultz, 2005, A strategy for probing the function of noncoding RNAs finds a repressor of NFAT: Science, v. 309, p. 1570-3.
Wilton, S. D., A. M. Fall, P. L. Harding, G. McClorey, C. Coleman, and S. Fletcher, 2007, Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript: Mol Ther, v. 15, p. 1288-96.
Wilusz, J. E., C. K. JnBaptiste, L. Y. Lu, C. D. Kuhn, L. Joshua-Tor, and P. A. Sharp, 2012, A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails: Genes Dev, v. 26, p. 2392-407.
Wood, M. J., M. J. Gait, and H. Yin, 2010, RNA-targeted splice-correction therapy for neuromuscular disease: Brain, v. 133, p. 957-72.

Wu, Q., Y. C. Kim, J. Lu, Z. Xuan, J. Chen, Y. Zheng, T. Zhou, M. Q. Zhang, C. I. Wu, and S. M. Wang, 2008, Poly A- transcripts expressed in HeLa cells: PLoS One, v. 3, p. e2803.

Xiao, T., H. Hall, K. O. Kizer, Y. Shibata, M. C. Hall, C. H. Borchers, and B. D. Strahl, 2003, Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast: Genes Dev, v. 17, p. 654-63.

Yan, C., and D. D. Boyd, 2006, Histone H3 acetylation and H3 K4 methylation define distinct chromatin regions permissive for transgene expression: Mol Cell Biol, v. 26, p. 6357-71.

Yokota, T., A. Nakamura, T. Nagata, T. Saito, M. Kobayashi, Y. Aoki, Y. Echigoya, T. Partridge, E. P. Hoffman, and S. Takeda, 2012, Extensive and prolonged restoration of dystrophin expression with vivo-morpholino-mediated multiple exon skipping in dystrophic dogs: Nucleic Acid Ther, v. 22, p. 306-15.

Zenklusen, D., P. Vinciguerra, J. C. Wyss, and F. Stutz, 2002, Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p: Mol Cell Biol, v. 22, p. 8241-53.

Zhang, B., G. Arun, Y. S. Mao, Z. Lazar, G. Hung, G. Bhattacharjee, X. Xiao, C. J. Booth, J. Wu, C. Zhang, and D. L. Spector, 2012, The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult: Cell Rep, v. 2, p. 111-23.

Zhao, Y., E. Samal, and D. Srivastava, 2005, Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis: Nature, v. 436, p. 214-20.

Acknowledgements

I wish to thank:

Prof. Giovanni Perini and All Perini-Bernardoni lab's Members, notably Dr. Samuele Gherardi

The collaborators of the Section of Medical Genetics of the University of Ferrara, notably Prof. Alessandra Ferlini, Dr. Matteo Bovolenta, Dr. Francesca Gualandi and Dr. Maria Sofia Falzarano

This work was supported by

Telethon

Industria Chimica Emiliana

BIO-NMD EU-Network

Association Française contre les Myopathies

CIRI-Scienze della vita e Tecnologie della salute – UNIBO

UNIFE