University of Milan

Faculty of Medicine and Surgery

Department of Medical Biotechnology and Translational Medicine

Ph.D. course in
Experimental Medicine and Medical Biotechnology
(XXIX cycle)

DOCTORAL THESIS

FUNCTIONAL RELEVANCE OF D4Z4-DERIVED ANTISENSE AND CODING SENSE TRANSCRIPTS ON DYSREGULATION OF MYOGENESIS IN FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY

Doctoral Candidate:
Raniero CHIMIENTI
R10547

Tutor:
prof. Anna MAROZZI

Academic Year 2015/2016

Table of contents

Chapter I - General introduction **6**

1.1 Skeletal muscle biology **6**

1.1.1 Embrionyc myogenesis 6

1.1.2 Transcriptional regulation of myogenesis 8

1.1.3 Adult myogenesis stem cells 9

1.1.4 Myoblast fusion 10

1.1.5 Signaling mechanisms in myoblast fusion 10

1.1.6 Mediators of muscular atrophy and hypertrophy 13

1.1.7 mTOR signaling pathway in muscle differentiation 14

1.2 Noncoding RNAs **17**

1.2.1 Long noncoding RNAs 17

1.2.2 Natural antisense transcripts 19

1.2.3 MicroRNAs 21

1.2.4 LncRNAs and miRNAs in muscle differentiation 23

1.3 Facioscapulohumeral dystrophy **25**

1.3.1 Clinical characteristics 25

1.3.2 Genotype-phenotype correlation 25

1.3.3 Penetrance and anticipation 26

1.3.4 Molecular basis of FSHD 27

1.3.5 DUX4 gene 29

1.3.6 DUX4c 31

1.3.7 DUX4-like genes 32

1.3.8 DBE-T 34

1.3.9 Other noncoding transcripts from D4Z4 macrosatellite 34

Chapter II - Materials and Methods 36

2.1 Cell lines 36
2.2 AZA-TSA treatment 37
2.3 RNA fractionation 38
2.4 RNA extraction 38
2.5 3’ Rapid Amplification of cDNA Ends (3’RACE) 39
2.6 Deep sequencing 40
2.7 One-step strand-specific RT-PCR and nested PCR 41
2.8 Cloning and Sanger sequencing 42
2.9 Quantitative Real-Time PCR 42
2.10 MiRNA retrotranscription and quantification 44
2.11 In vitro transcription 44
2.12 5’end capping and 3’end biotinylation 45
2.13 RNA-protein pull down assay 46
2.14 Long non-coding RNA, siRNA and miRNA transfection 46
2.15 SDS-PAGE and immunoblotting 47
2.16 Immunofluorescence 47
2.17 Flow citometry 48
2.18 Target prediction, gene ontology and pathway analysis 48
2.19 Sequence homology-based conserved domain analysis 49
2.20 Secondary structure analysis 49
2.21 De novo modeling and structure validation 50
2.22 Molecular dynamics simulation 50
2.23 Statistical analysis 50

Chapter III - Results

3.1 Characterization of FSHD myogenesis 51

3.1.1 FSHD myoblasts exhibit early commitment to myogenesis, fusing into disorganized myotubes, when compared to control muscle cells 51

3.1.2 DUX4 full-length isoforms are expressed in FSHD and control myoblasts and increase during myogenic differentiation 57
3.1.3 Use of knockout serum replacement (KOSR) as alternative to HS accelerates myogenic differentiation in FSHD and control myoblasts by enhancing DUX4 expression 58

3.2 Characterization of D4Z4-derived miRNAs 60
3.2.1 Analysis of the transcription start site distribution within D4Z4 macrosatellite and DUX4-like genes 60
3.2.2 Evidences of alternative TSSs for DUX4-AS2 (4q35.2) and its putative homologous 66
3.2.3 Experiments of 3'RACE reveal that DUX4-AS2, as well as its putative homologous, is non-polyadenylated 68
3.2.4 Analysis of sequencing suggests the presence of at least three different antisense transcripts from DUX4 gene family 71
3.2.5 Strand-specific RT-PCR corroborates existence of three antisense transcripts from chromosome 3, 4 and Y 78
3.2.6 Antisense transcript expression profiling during muscle differentiation highlights putative role of DUX4-AS2 in FSHD 80
3.2.7 DUX4-AS2 is localized into the nucleus, where interacts with DROSHA/DGCR8 complex 82
3.2.8 Secondary structure and predictive analysis report several miRNAs originating from DUX4-AS2 85
3.2.9 Six D4Z4-derived miRNAs are expressed in FSHD and/or healthy control muscle cells 88
3.2.10 Validated miRNAs are differentially expressed during FSHD myogenesis, when compared to control 90
3.2.11 Target prediction suggests involvement of as-miR-276-5p and as-miR-320-3p in myoblast fusion and early differentiation 92
3.2.12 Overexpression of as-miR-276-5p and as-miR-320-3p affects members of mTOR signaling pathway 96

3.3 Structural and functional analysis of DUX4-fl 99
3.3.1 Ab initio modeling of DUX4-fl protein 99
3.3.2 Conformational and functional evaluation of DUX4-fl C-terminus domain reveals homology with ubiquitin-binding CUE domains 105
3.3.3 DUX4-fl translocates into cytoplasm during early muscle differentiation and its silencing slows myoblast fusion rate 111
3.3.4 DUX4-fl co-localizes with M-cadherins, resulting in their altered distribution along plasma membrane in FSHD cells 114

Chapter IV - Discussion 115

Bibliography 122

Chapter I - General introduction

1.1 Skeletal muscle biology

Skeletal muscle is a contractile tissue managing voluntary movement and involved in maintaining of body's posture and temperature. The development of musculoskeletal apparatus is controlled by a complex system of extrinsic and intrinsic regulatory mechanisms, starting after gastrulation, with somitogenesis, and continuing throughout life.

1.1.1 Embryonic myogenesis

During embryogenesis, first skeletal muscle fibers of trunk and limbs originate from paraxial mesoderm-derived structures, called somites, and additional myofibers are generated by subsequent waves of muscle progenitor cells that differentiate in rostro-caudal and medio-lateral direction.

Spatiotemporal somitogenesis represents the earliest morphogenetic event of embryonic development and proceeds through the following stages: periodicity and separation of somites, epithelialization, specification, and differentiation.

Paraxial mesoderm produces somites at cyclic species-specific intervals by expression of "oscillating genes" involved directly or indirectly in Notch and Wnt signaling pathways, and by gradient of Wnt, FGF8 and retinoic acid (RA) providing a "wave" motion (Betzinger et al, 2012).

Somites separation proceeds in cranio-caudal direction and epithelialization begins before their detachment from paraxial mesoderm by deposition of extracellular matrix, expression of adhesion proteins (i.e. N-cadherin) and organization of a basal lamina between somites themselves and adjacent structures such as neural tube, notochord, dorsal ectoderm and lateral mesoderm.

Although specification of somites into different structures along the antero-posterior axis of embryo is defined at early stages of somitogenesis by fine regulation of Hox genes, differentiation of somitic compartments into sclerotome, dermatome and myotome depends by local signals from the tegumentary ectoderm, notochord, dorsal laminae and floor plate of neural tube (Musumeci et al, 2015).

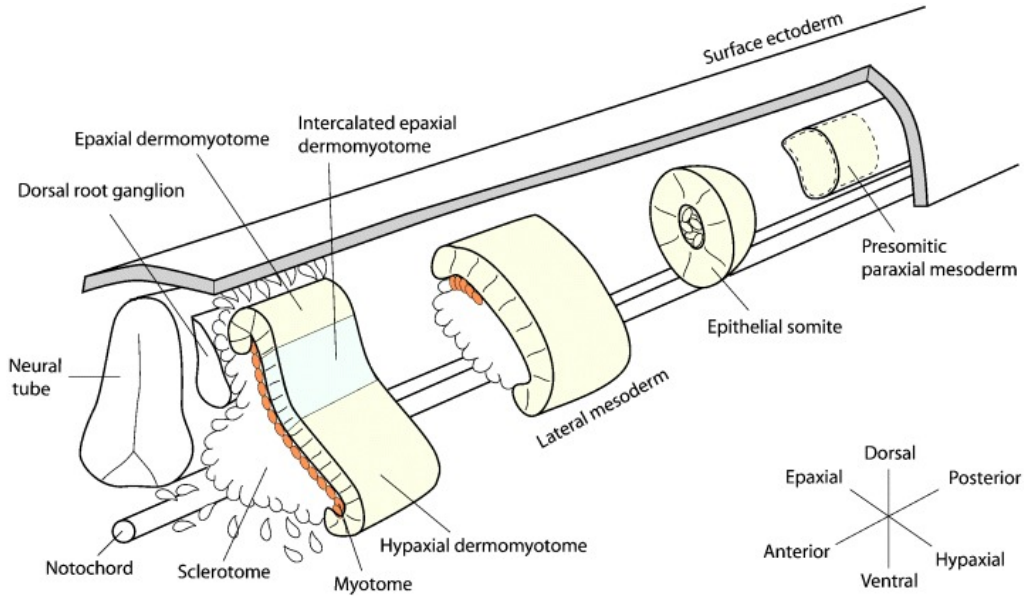


Fig. 1 - Schematic representation of somitogenesis.

Initially, toward the caudal region of the paraxial mesoderm, gradient of Wnt and FGF8 maintains cells to undifferentiated mesenchymal state and controls cyclic expression of “oscillating genes”. When cells move to cranial part, periodic activity of Notch pathway stops and increasing levels of RA polarize somites in dorso-ventral direction (Hofmann et al, 2004).

The most ventral compartment forms the mesenchymal sclerotome, containing precursors that will generate cartilage and bone. The dorsal epithelial compartment constitutes the dermomyotome, where myogenesis occurs. Lateral portions of dermomyotome constitute the myotome, whose cells separate, migrate and generate myoblasts. According to their distance from neural tube, myoblasts differentiate in different kind of myofibers: myoblasts closest to the neural tube produce epaxial and extensor muscles, whereas those in the most distant region from above structure are precursors of hypaxial and flexor muscles (Musumeci et al, 2015).

Myoblasts differentiation is orchestrated by expression of myogenic regulatory factors Myf5 and MyoD in response to extrinsic factors, such as Wnt1 and Wnt3, secreted from the dorsal neural tube, and Wnt4, Wnt6 and Wnt7, produced from the surface ectoderm. Sonic hedgehog (Shh), released from the notochord and floor plate of the neural tube, is also involved in cellular commitment in the somite, by down-regulation of Pax3/7 e subsequent increasing of Myf5 levels (Cossu et al, 1996; Dietrich et al, 1998).

The formation of head musculature, instead, differs significantly from trunk and limbs muscles specification; it originates from the cranial paraxial mesoderm (CMP) and from lateral splanchnic mesoderm (SpM), and is strongly influenced by cranial neural crest cells.

Furthermore, while most of myoblasts differentiate to originate muscles, some of them remain in undifferentiated state and surround mature myofibers. These undifferentiated cells become satellite cells (SCs), responsible for postnatal muscle growth and repair (Betzinger et al, 2012).

1.1.2 Transcriptional regulation of myogenesis

Overall, commitment and terminal differentiation of muscle cells undergo transcriptional control of four myogenic regulatory factors (MRFs): Myf5, MRF4 (also known as Myf6), MyoD and myogenin (MyoG).

MRFs are transcription factors sharing homologous basic helix-loop-helix (bHLH) domain that is required for DNA binding and dimerization with E-protein family. Both MRF monomer and MRF-E protein heterodimer bind E-box consensus sequence CANNTG in the promoters of downstream muscle-specific genes, leading to muscle cell differentiation (Heidt et al, 2007).

Critical transcription factors of muscle cell determination are Myf5 and MyoD and their specific knockout completely abrogates skeletal muscle formation. Moreover, MyoG is essential for terminal differentiation, as well as MRF4, regulating fusion and fiber maintenance. In particular, MFR4 plays a dual role, since it is also expressed in undifferentiated proliferating cells as determination gene (Londhe et al, 2011).

1.1.3 Adult myogenesis stem cells

Adult muscle growth and regeneration is guaranteed by Pax3/7-positive myogenic stem cells, named satellite cells (SCs), and located between basal lamina and sarcolemma of associated myofibers. These cells can both replicate themselves (self-renew) and, after activation trigger, escape from quiescent state and give rise proliferating myoblasts by re-entering cell cycle.

Most of quiescent SCs are characterized by expression of both Pax3/7 and CD34 (also known as sialomucin, an anti-adhesive molecule that act to aid migration); all CD34-positive quiescent SCs show activity in Myf5 locus, suggesting they are committed to myogenic lineage (Beauchamp et al, 2000).

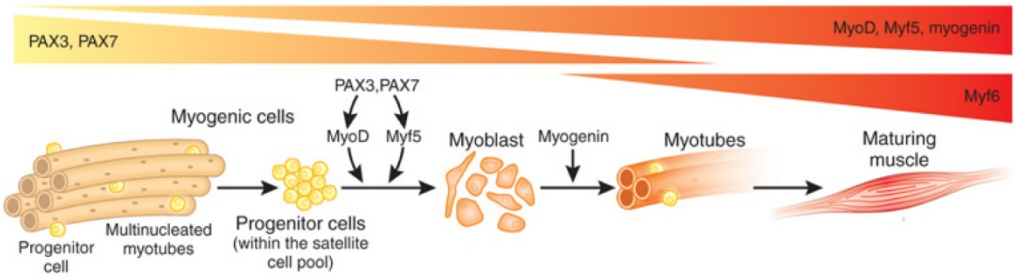


Fig. 2 - Muscle differentiation stages and temporal expression of their regulatory genes.

Unlike myogenic precursor cells, quiescent SCs don't express the transcription factor MyoD until they switch to proliferating state, becoming active. Co-expression of MyoD, desmin and MyoG in SCs was observed to be associated with a subpopulation of committed SCs that are totally prone to differentiate (Smith et al, 1994; Rantanen et al,1995).

Both Myf5 and MyoD then play a specific and distinct role in SCs biology and adult myogenesis compared to that performed during embryonic development. Indeed, MyoD is required for differentiation potential of SCs-derived skeletal myoblasts (Sabourin et al, 1999; Cornelison et al; 2000), whereas Myf5 regulates their proliferation rate and homeostasis (Gayraud-Morel et al, 2007; Ustanina et al, 2007).

After expression of MyoD, a hierarchical gene expression circuitry is activated and leads to irreversible and sequential cascade of events regulating myoblast fusion and differentiation (Penn et al, 2004; Berkes and Tapscott, 2005).

1.1.4 Myoblast fusion

During postnatal life, fusion of myogenic precursors is required to allow the muscle growth and replacement, and regeneration of myofibers after injury. The fusion process consists of two main phases characterized by the alignment of the myoblasts into parallel arrays and membrane fusion following rearrangement of actin cytoskeleton at contact sites. First step of myoblast fusion (which is referred to as "primary fusion") is required for the formation of nascent myofibers (or primary myotubes). In the second phase, the recruitment of mononucleated myogenic cells to the primary myotubes complete the myofiber growth and nuclear accretion by a secondary fusion wave (Rochlin et al, 2010; Abmayr and Pavlath, 2012).
Several studies conducted on myoblast cell culture models have been allowed to identify an extensive array of cell signaling pathways playing a critical role in myoblast fusion: some of these are activate subsequently to recruitment and contact of cell-surface proteins between fusion partners; other are integral part of the myogenic differentiation program, but contributing equally to fusion process.

1.1.5 Signaling mechanisms in myoblast fusion

Primary myoblast fusion and sarcomere assembly are mediated by fine organization on the plasma membrane of integrins, caveolin and others cell-surface receptors, such as neogenin. Integrins, whose subunit β1 is an essential component of myogenic cells, link extracellular matrix ligands to several intracellular proteins. Binding of integrins to external molecules, in particular, induces integrin clustering and the recruitment and autophosphorylation of the non-receptor protein tyrosine kinase FAK (Schwander et al, 2003; Schaller, 2010). Signaling triggered by FAK affects molecules involved in regulation of actin cytoskeleton, focal adhesion, mitogen-activated protein kinase (MAPK), and insulin signaling pathway (Quach et al, 2009). As consequence of FAK phosphorylation, which recruitment is orchestrated by protein kinase c (PKC) isoform PKCθ, the levels of caveolin-3 and β1D-integrin increase, enhancing

the extent of myoblast fusion (Madaro et al, 2011). Other components that promote fusion by interaction with FAK are neogenin and its ligands, the netrins. Specifically, Netrin2 determines FAK activation in a neogenin-dependent manner in cultured primary myoblasts (Bae et al, 2009).

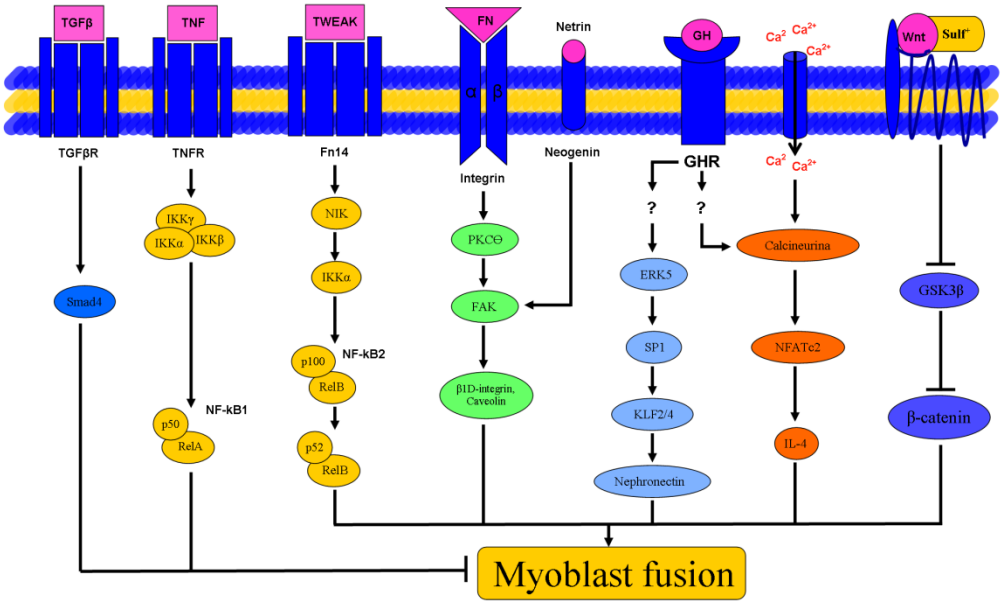


Fig. 3 - Signaling pathways involved in myoblast fusion.

Rho family of guanosine triphosphatase (GTPase) plays a role in cytoskeletal re-organization through modulating activity of several downstream kinases and transcription factors.

Time-dependent activation of RhoA protein is required for the induction of skeletal myogenesis in response to serum deprivation (Wei et al, 1998), but it necessitate to be deactivate before fusion occurs. RhoA inhibition during early stages of differentiation is mediated by RhoE (Fortier et al, 2008) and Rho GTPase-activating protein GRAF1 (GTPase regulator associated with focal adhesion kinase-1) (Doherty et al, 2011). RhoA activity, indeed, reduces stability and alters localization of M-cadherin, a cell adhesion molecule that is fundamental for myoblast fusion (Charrasse et al, 2006).

In contrast to RhoA inhibitory role, other GTPases, such as Rac GTPases, positively regulate fusion process along its stages, governing activity of c-Jun N-terminal kinases (JNKs), serum response factor (SRF) and p38 MAPK (Meriane et al, 2000; Charrasse et al, 2002; Bryan et al, 2005).

Involvement in muscle differentiation of well-characterized subfamilies of MAPK (such as p38, ERK1/2, ERK5 and JNK) and of transforming growth factor beta (TGFβ)-activated kinase-1 is supported by several studies conducted on different myoblast cell models (Gredinger et al, 1998; Wu et al, 2000; Mariane et al, 2000; Jones et al, 2001; Bhatnagar et al, 2010). The p38γ isoform is expressed preferentially in muscle tissue and it increases during differentiation (Lechner et al, 1996; Li et al, 1996). It stimulates the activation of MyoD and MEF2C, and synergistically increases the trans-activation potential of MyoD (Cuenda and Cohen, 1999). Expression of caveolin-3, a protein localized to the sarcolemma where it gives rise to complex with dystrophin, G-protein, Src-like kinases and other glycoproteins (Song et al, 1996), requires p38 activation (Galbiati et al, 1999). Disruption of caveolin-3-associated complexes has been observed in autosomal dominant limb girdle muscular dystrophy (LGMD), suggesting the importance of caveolins in the pathogenesis of muscular dystrophy (Minetti et al, 1998). In this regard, MAPK play a pivotal role in myotube survival and maintenance of differentiated condition.

On the other hand, p38 intervenes also in the TAK-1/p38/nNFkB inhibitor signaling of myoblast differentiation (Trendelenburg et al, 2012), and ERK1/2 appears strongly required for myoblast proliferation, but dispensable in promoting of muscle-specific gene expression and cell fusion (Jones et al, 2001). In contrast, ERK5 drives myoblast differentiation through activation of transcription factors SP1 (specificity protein 1) which, in turn, promotes expression of transcription factors Klf2/4 (Kruppel-like factor 2/4) and nephronectin (Npnt), an extracellular matrix protein involved in cell-matrix adhesion (Sunadome et al, 2011).

The outcome of fusion process into multinucleated myotubes is also regulated by calcium-dependent signaling covering several downstream targets including NFAT transcription factor family (Horsley et al, 2002, 2003; Wu et al, 2007). In particular, calcineurin-mediated NFATc2 activation plays a critical role in secondary myoblast fusion, regulation of muscle mass and myofibers diameter by increasing of interleukin-4 (Horsley et al, 2003) and MyoF (Demonbreun et al, 2010).

In adult myogenesis, NF-kB family proteins play an ambivalent role. In fact, activation of NF-kB occurs through canonical and non-canonical signaling pathways (Hayden and Grosh, 2004), that mediate inhibition (Li and Kumar, 2008; Bakkar et al, 2008) or promote muscle differentiation (Enwere et al, 2012) respectively.

Wnt signaling pathway, as well as it is involved in embryonic differentiation, is equally crucial for myoblast fusion in adulthood. Specifically, inhibition of GSK-3β, a component of classical Wnt signaling network, enhances nuclear accumulation of β-catenin and promotes insulin-induced muscle differentiation (Rochat et al, 2004; von Maltzahn et al, 2012). Similarly, myoblast fusion is stimulated and accelerated by Wnt1 and Wnt3a ligands (Rochat et al, 2004; Pansters et al, 2011). In contrast, non-canonical Wnt pathway, that is activated in response to muscle injury in Sulf1/2 knockdown muscle models, inhibits fusion of skeletal progenitor cells by Wnt7a-mediated delocalization of FAK (Tran et al, 2012). An antagonist function is also played by canonical TGF-β signaling pathway, that, through activation of SMAD4, inhibits both myoblast proliferation and differentiation (Olson et al, 1986).

1.1.6 Mediators of muscular atrophy and hypertrophy

Muscle mass is controlled by the fine balance of anabolic (leading to protein synthesis and hypertrophy) and catabolic (enhancing protein degradation and inducing atrophy) processes (Mitch and Goldberg, 1996; Hasselgren, 1999; Jagoe and Goldberg, 2001). Increasing of muscle size and hypertrophy depends on insulin-like growth factor-1 (IGF-1) that mediates activation of PI3K/AKT pathway, resulting in phosphorylation of the FOXO proteins.

AKT-mediated phosphorylation of transcription factor family FOXO prevents its translocation into the nucleus, hence the beginning of atrophy program (Brunet et al, 1999). Additionally, hypertrophy elicits activation of p70 and PHAS-1, two downstream effectors of mTOR signaling pathway involved in protein synthesis (Rommel et al, 2001; Reynolds et al, 2002).

During atrophy, instead, there is an increase of ubiquitin ligases MAFbx (muscle atrophy F box, also called atrogin-1) and MuRF1 (muscle RING finger

1) expression, ubiquitin-protein conjugates and ubiquitin degradation pathway activity (Hasselgren, 1999; Bodine et al, 2001b; Gomes et al, 2001).
Also myostatin (a TGF-β family member) signaling pathway plays a critical role as negative regulator of muscle mass and in regulating of skeletal fiber homeostasis, preventing myogenic program. Its antagonist, follistatin, as part of inhibin-activin-follistatin axis, has been shown to prevent muscle atrophy, enhancing muscle growth and strength.

1.1.7 mTOR signaling pathway in muscle differentiation

Mammalian target of rapamycin (mTOR) is involved in control of several cellular processes, such as cell growth, proliferation, differentiation, survival, autophagy and metabolism (Laplante and Sabatini, 2012). It forms two distinct complexes, mTORC1, mainly composed by GβL (or MLST8) and RAPTOR, and mTORC2, defined by presence of GβL and RICTOR. The rapamycin-sensitive complex mTORC1 has two well-documented substrates involved in protein synthesis: S6K1 and 4E-BP1 (eIF4E-binding protein 1) (Ma and Blenis, 2009). Moreover, mTORC1 integrates signals from growth factor and other stressors regulating cell growth by mediation of TSC1/TSC2 complex, acting as GTPase-activating protein for Rheb (Manning and Cantley, 2003).
On the other hand, the rapamycin-insensitive mTORC2 activity provides phosphorylation of AKT, PKCα and other kinases regulating cell survival, actin cytoskeleton organization and, in particular, myogenesis (Ge and Chen, 2012).
Preferentially through mTORC2, mTOR intervenes in primary and secondary myoblast fusion, and fiber maturation, in a kinase-independent and kinase-dependent manners, respectively. Key mediator of kinase-independent pathway is IGF2, which triggers signaling cascade along the IRS1/PI3K/AKT axis. Furthermore, mTOR has been shown to suppress negative regulation of YY1 on IGF2/AKT signaling components (Blatter et al, 2012).
Kinase-dependent pathway under control of mTOR is necessary for secondary myocyte fusion and maturation of nascent myofibers. It starts with direct regulation by mTOR of MyoD and the subsequent activation of miR-1, which

suppresses HDAC4 and stabilizes MyoD itself, mediating over-expression of follistatin. Because of this, it has been hypothesized that mTOR might also manage many MyoD-regulated myogenic programs, in addition to evidences of a link between mTOR pathway and miRNAs involved in muscle differentiation (Sun et al, 2010).

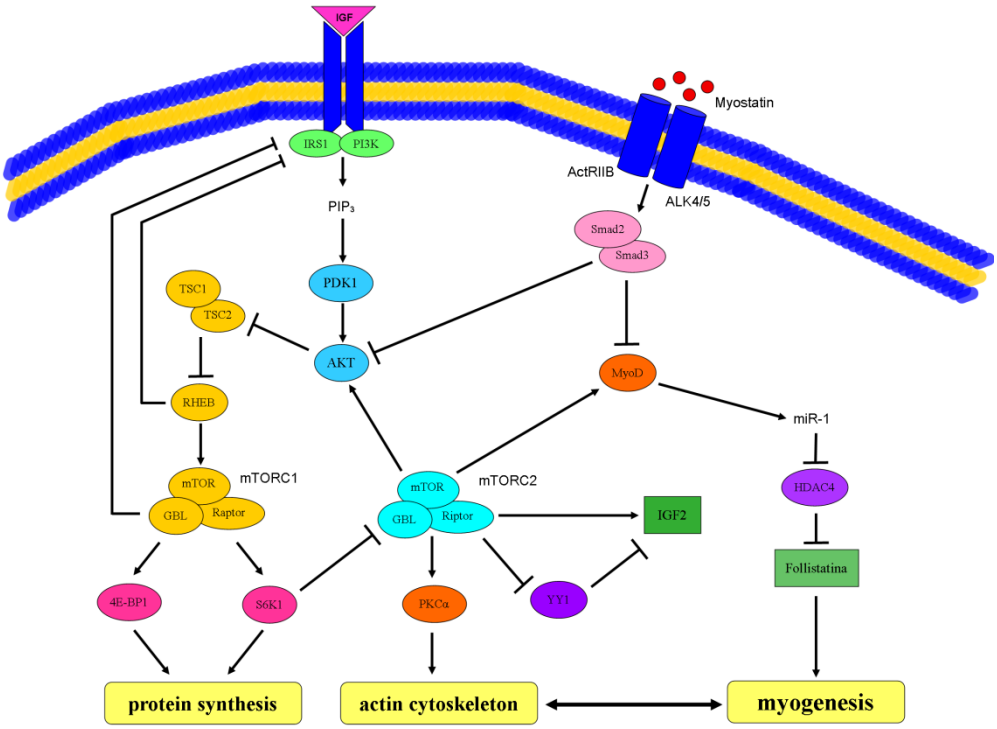


Fig. 4 - mTOR signaling pathway.

A positive role of mTOR during myogenesis is attributable to mTORC2 and its obligate component RICTOR, whereas mTORC1 acts a suppressive function, shifting the balance toward proliferation. Indeed, RAPTOR inhibits myoblast fusion by destabilization of IRS1 and consequent suppression of PI3K/AKT signaling (Ge et al, 2011). Also Rheb perturbs IRS1 protein, whereas S6K1 mediates inhibition of RICTOR (Dibble et al, 2009; Julien et al, 2010; Ge et al, 2011).

Myogenic mTOR signaling intersects numerous pathways, some of which have been reported to directly or indirectly contrast with its activity. For example, FoxO1 induces degradation of mTORC2 components via proteasome, but is inhibited by AKT (Wu et al, 2008). Myostatin, that has been seen to negatively

regulate myoblast fusion, suppresses both AKT and MyoD in a SMAD2/3-dependent manner (Langley et al, 2002; Trendelenburg et al, 2009).

1.2 Noncoding RNAs

The noncoding RNAs (ncRNAs) fall within the major group of mammalian RNAs, accomplishing a remarkable number of biological functions. Indeed, they mediate DNA synthesis or genome rearrangement; participate to regulation of gene expression, intervening at the level of transcription, RNA maturation and translation; most of them operate as RNA-protein complexes (Cech and Steitz, 2014). In the last years, some classes of noncoding RNAs, such as long noncoding RNAs (lncRNAs) and microRNAs (miRNAs), have gained widespread attention as a potentially crucial layer of regulations implicated in several developmental processes and diseases, although the knowledge of their finely regulated expression and processing, and their mechanisms of action is still surprisingly limited.

1.2.1 Long noncoding RNAs

By definition, lncRNAs are non-protein coding transcripts longer than 200 nt. Their classification can be dealt with according to genomic context or a functional perspective, keeping always in mind that the current categorization, on the basis of these parameters, is considered to be mutually nonexclusive.

LncRNAs may be transcribed as stand-alone units, which often referred to as lincRNAs for large intergenic (or intervening) noncoding RNAs (Guttman et al, 2009; Cabili et al, 2011; Ulitsky et al, 2011); or alternately transcribed from enhancers (eRNAs) (Kim et al, 2010; Wang et al, 2011a), promoters (TSSa-RNAs, uaRNAs, pasRNAs and PROMPTs) (Core et al, 2008; He et al, 2008; Seila et al, 2008; Kanhere et al, 2010) or introns (Louro et al,2009; Rearick et al, 2011).

Moreover, many lncRNAs have been identified as *pseudogenic*. Pseudogenes are evolutionarily “dead” genes due to nonsense, frameshift or other mutational events (Balakirev and Ayala, 2003; Pink et al, 2011). Transcribed pseudogenes were found to have acquired a new function and to regulate gene expression by

epigenetic or post-transcriptional mechanisms. In example, Xist transcript is hypothesized to be derived from pseudogenization, following the integration of several transposon-derived repeat elements, of the Lnx3 protein coding gene (Duret et al, 2006; Elisaphenko et al, 2008).

In a strictly functional point of view, lncRNAs are involved in epigenetics, as recruiters, tethers and scaffolds; in both transcriptional and post-transcriptional regulation, and in the organization of nuclear compartments.

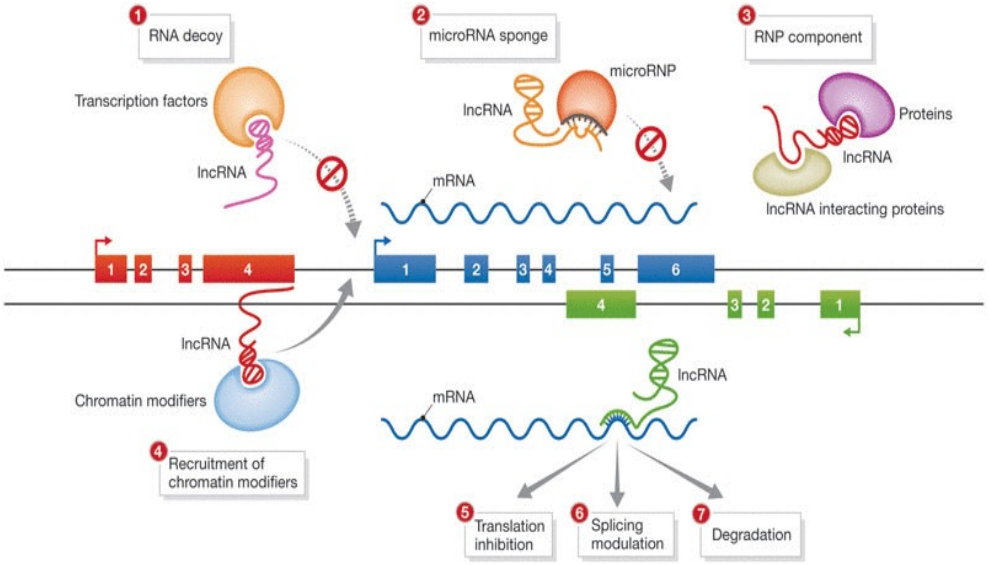


Fig. 5 - Mechanisms of lncRNA function.

As regulators of chromatin states, lncRNAs have the ability to interact with chromatin-modifying complexes, such as PRC2 (Khalil et al, 2009; Kanhere et al, 2010; Zhao et al, 2010). In fact, Polycomb complexes, including several factors that bind or modify chromatin marks, have been seen to be recruited to specific genomic loci in mammals by lncRNAs, such as HOTAIR (Rinn et al,2007), Kcnq1ot1 (Pandey et al, 2008) and ANRIL (Yap et al, 2010; Kotake et al, 2011).

Furthermore, lncRNAs can directly affect the process of transcription, acting as transcriptional co-regulators, such as SRA, a co-activator for a number of nuclear steroid receptors (Lanz et al, 1999, 2002); or acting as decoys for transcriptional factors, such as PANDA and NRON lncRNAs (Willingham et al, 2005; Hung et al, 2011) or inhibitors of polymerase II activity, such as

lncRNAs produced from SINEs (Espinoza et al, 2004; Mariner et al, 2008) or from DHFR promoter (Schnell et al, 2004).

In post-transcriptional regulation, lncRNAs mediate mRNA processing or influence its stability and translation efficiency.

MALAT1 has been found to affect alternative splicing; analogously, Gomafu/MIAT lncRNA may hide spliceosome formation, altering splicing of a subset of messengers (Sone et al, 2007; Tsuiji et al, 2011). Moreover, MALAT1, as well as NEAT2, has been shown to localize splicing factors to nuclear compartments called speckles (Bernard et al, 2010), suggesting that lncRNAs are simultaneously regulators of nuclear organization. Both Xist and Kcnq1ot1 have been also observed to target their interactors to perinucleolar compartment to maintain silencing (Zhang et al, 2007; Pandey et al, 2008).

Interestingly, lncRNAs are emerging as both sources and sinks of small RNAs. Rather than competing for miRNA-binding sites, lncRNAs can compete for the miRNAs themselves. Several mammalian pseudogenes-derived lncRNAs, such as PTEN1 and KRASP1, have miRNA-binding sites in their 3'-UTR, acting as sponges to sequester miRNAs away from their targets (Poliseno et al, 2010). On the other hand, lncRNAs can themselves generate miRNAs and other small RNA species, such as H19, that hosts miR-675 (Cai and Cullen, 2007; Keniry et al, 2012), or GTL2, anti-RTL1 and Mirg lncRNAs, that give rise up to 50 miRNAs and at least 40 snoRNAs (da Rocha et al, 2008).

1.2.2 Natural antisense transcripts

Around 70% of annotated sense transcriptional units have reported antisense counterparts. Natural antisense transcripts (NATs) are considered a subgroup of lncRNAs that is poorly conserved among species (Johnsson et al, 2014). Generally, they affect their partner coding genes by suppression or activation, or act in stabilization of long-range chromosomal interactions. One of the well-documented sense antisense (SAS) pairs is Xist/Tsix, involved in control of X chromosome inactivation (Lee et al, 1999a). Moreover, other SAS pairs have been reported in imprinted region, such as Kcnq1/Kcnq1ot1 (Kanduri et al, 2006) and Igf2r/Air (Lyle et al, 2000).

NATs can be spliced, polyadenylated or not-polyadenilated, and have been classified basing on their relative localization to coding gene sequences: head-to-head, tail-to-tail and fully overlapping. Often, NATs have been reported as physically discrete transcriptional units, very close to sense coding gene, and are referred to as in nearby head-to-head and nearby tail-to-tail orientations (Osato et al, 2007).

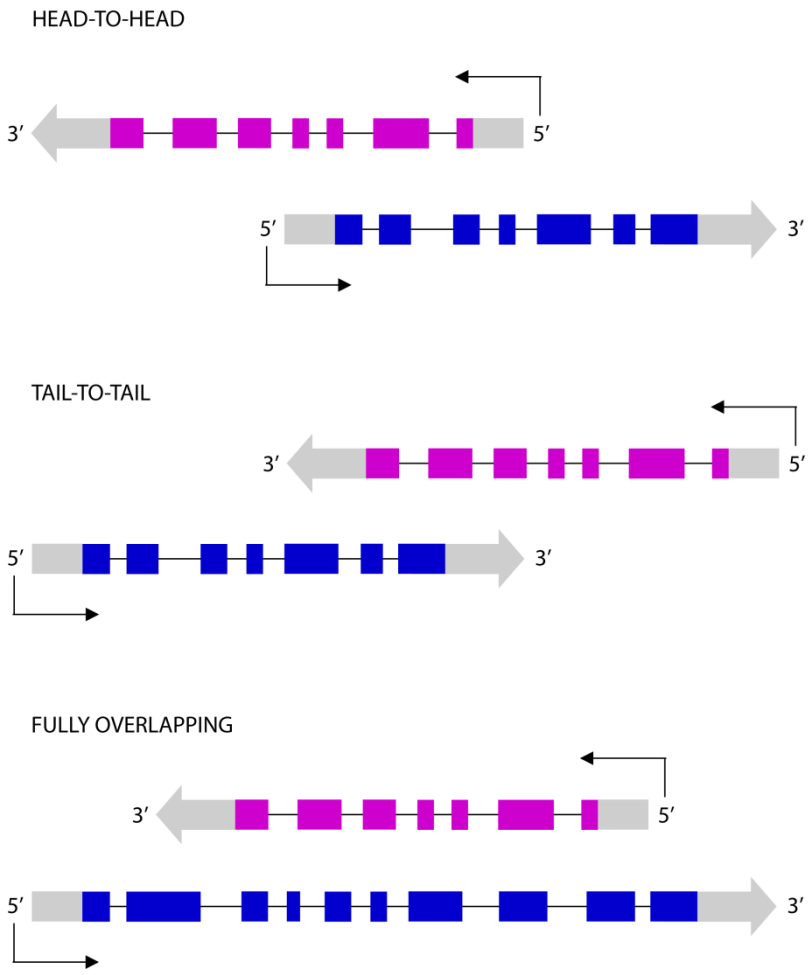


Fig. 6 - Transcriptional orientation of NATs.

For most of NATs, as well as other lncRNAs, biological functions proposed contemplate modulation of mRNA stability and decoy, scaffolding and tethering mechanisms. Recently has been suggested that NATs may serve as precursor for endogenous siRNA and miRNA generation. This hypothesis is supported by the importance of SAS duplex in siRNA biogenesis and by evidence of interaction between NATs and argonaute (AGO) family proteins (Polikepahad and Corry, 2013; Werner et al, 2014).

1.2.3 MicroRNAs

MicroRNAs (miRNAs or miRs) are small RNAs of 21-25 nucleotide in length, produced by two RNase III proteins, DROSHA and DICER, and able to negatively regulate the gene expression at post-transcriptional level (silencing), mediating the mRNA degradation or inhibiting its translation.

MicroRNA-coding genes are transcribed by RNA polymerase II as primary miRNA (pri-miRNA), typically long over 1 kb and containing local stem-loop structures that undergoes several step of maturation. Its processing begins in the nucleus by action of Microprocessor complex, of which the core elements are the nuclear RNase III DROSHA and its essential co-factor DGCR8.

The pri-miRNA stem-loop recognition is operated by the two double strand RNA binding domains (dsRBDs) of DGCR8, whereas its C-terminus interacts with DROSHA, recruiting it to the pri-miRNA cutting site.

DROSHA cleaves the hairpin approximately 11 bp away from the “basal” junction between the single strand RNA and dsRNA of the stem-loop, and 22 bp away from the “apical” junction linked to the terminal loop (Zeng et al, 2005; Han et al, 2006). DROSHA’s cleavage generates a short dsRNA (~65 nt) called pre-miRNA, which is exported to cytoplasm, where maturation can be completed. Transferring of pre-miRNA from nucleus to cytoplasm is mediated by RAN-GTP/exportin 5 (EXP5) complex; hydrolysis of GTP bonded to nuclear protein RAN results in disassembly of the complex and in the release of pre-miRNA into the cytosol. Here the pre-miRNA is processed by DICER, an RNase III endonuclease recognizing the 3’ overhangs leaved by DROSHA.

DICER’s processing generates a small RNA duplex (miRNA:miRNA*) that is subsequently loaded onto AGO protein in the RNA-induced silencing complex (RISC) (Hammond et al, 2001; Mourelatos et al, 2002).

After loading into RISC, RNA duplex is unwound and the strand selection occurs. As strand selection is not completely strict, the more instable strand can be retained with varying frequency. Moreover, the less abundant passenger strand (miRNA*) is active in silencing, albeit usually less potently than the more abundant guide strand (miRNA). Alternative strand selection (also known as arm switching) has been observed comparing miRNA isoforms (Chiang et al, 2010).

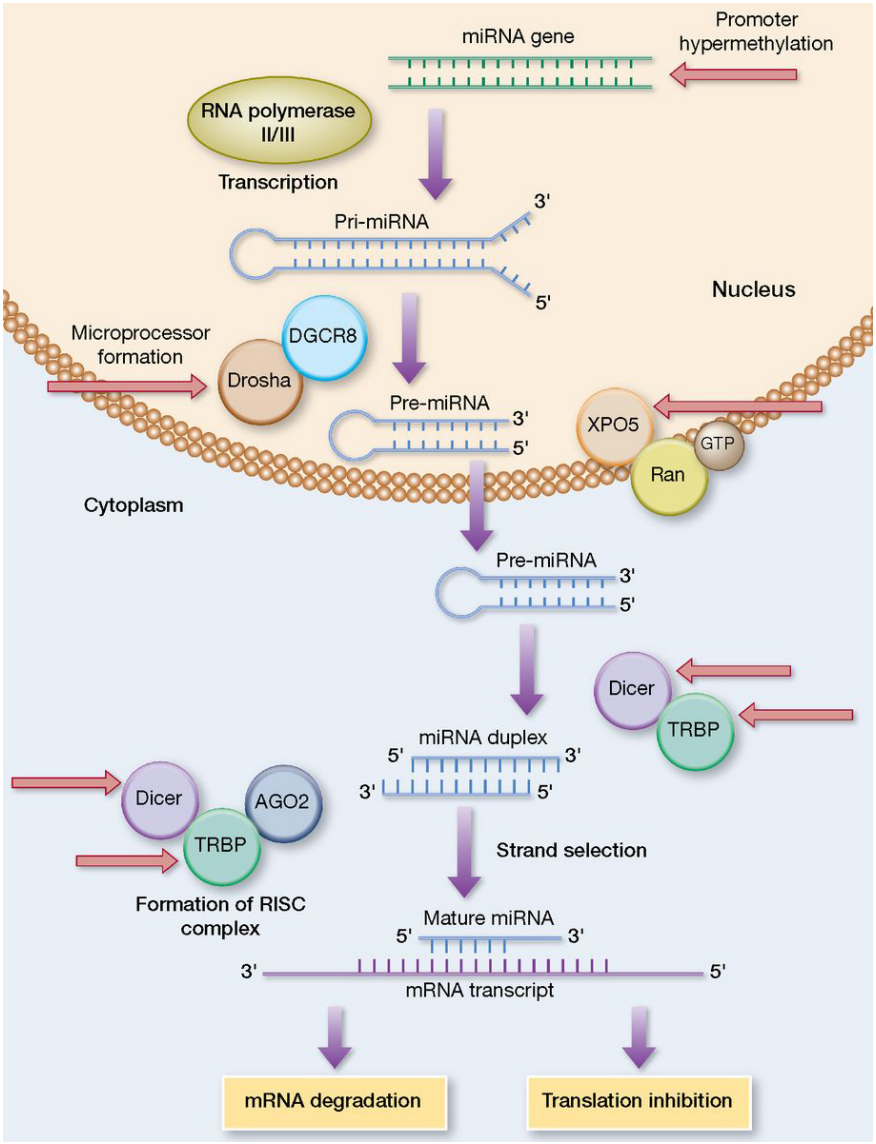


Fig. 7 - Canonical pathway of miRNA biogenesis.

That described above is considered the canonical miRNA biogenesis pathway. However, various alternative mechanisms generating miRNAs or miRNA-like short RNAs have been observed. For example, small RNA precursors may be generated through mRNA splicing in a DROSHA-independent manner (Flynt et al, 2010); analogously, miRNAs may also be produced by a DICER-independent AGO2-mediated (Cheloufi et al, 2010; Yang et al, 2010) or by terminal uridylyl transferase (TUTase)-dependent DICER-mediated cleavage (Heo et al, 2012).

Control of gene expression attended to by mature miRNAs happens at posttranscriptional level, mainly by messenger degradation or translation

inhibition. Generally, interaction between miRNA and mRNA is based on complementarity of a 6-8 nt sequence known as *seed* (Zynoviev et al, 2010).
The seed sequence complementarity and its length establish the specificity of action and the repression strength of miRNA, respectively, although several evidences demonstrate the possibility of nonseed-dependent target recognition (Brodersen et al, 2009).
Commonly, miRNAs regulate their target by binding to 3'UTR. Multiple binding sites for the same miRNA in 3'UTR strongly enhance mRNA suppression. Overall, miRNA binding sites were found also in 5'UTR and coding region; interestingly, genes containing miRNA binding sites in both coding sequence and 3'UTR are more significantly regulated than those targeted in 3'UTR only (Fang and Rajewsky, 2011).

1.2.4 LncRNAs and miRNAs in muscle differentiation

As lncRNAs have emerged to be able to regulate gene expression at various levels, demonstrating to be a class of functionally versatile transcripts, they have been studied in relation to development, cell identity maintenance and differentiation. In the last few years, as expected, lncRNAs appeared to play a crucial role also in skeletal myogenesis.
Two members (CE and DRR) of a group of lncRNAs called enhancer-derived RNAs (eRNAs), originating upstream the regulatory region of MyoD, modulate muscle differentiation by directing chromatin-remodeling events to promote expression of MyoD and MyoG. Another lncRNA, Yams (YY1-associated muscle lincRNAs), recently identified, displays function of transcriptional activator (Lu et al, 2013), as well as SRA, that besides functioning as scaffold, forms complex with MyoD and p68/p72, acting as co-activator of transcription of a subset of MyoD target genes (Caretti et al, 2006).
Fascinatingly, MALAT1 has been revealed as downstream target of myostatin (Watts et al, 2013), suggesting a role in control of muscle mass and atrophy.
Also H19 works at different levels: binding Polycomb complex, it mediates Igf2 repression (Ripoche et al, 1997); containing several binding sites for let-7 microRNA family, it may act as RNA sponge (Kallen et al, 2013). Moreover,

H19 exon 1 encodes miR-675-5p and miR-675-3p (Cai and Cullen, 2007), that targeting SMAD1, SMAD5 and Cdc6, play an important role in skeletal muscle differentiation and regeneration (Dey et al, 2014).

Linc-MD1, strongly reduced in Duchenne dystrophy, acts as sponge of miR-133 and miR-135 to regulate expression of MAML1 and MEF2C. Its up-regulation mediates anticipation of myogenesis program (Cesana et al, 2011).

The boundary between lncRNAs and miRNAs results to be particularly labile. Indeed, a newly discovered noncoding transcript (lnc-31) overlapping coding region of miR-31 has been shown to counteract differentiation (Ballarino et al, 2015), as well as miR-31 controls proliferation by regulating of Myf5 (Crist et al, 2012) and dystrophin (Cacchiarelli et al, 2011).

Unlike lncRNAs, miRNAs involved in skeletal muscle differentiation are much more characterized. In muscle cells, SRF and MEF2 family transcription factors cooperate with MyoD and MyoG to activate muscle-specific miRNAs (myomiRs): miR-1-1 and miR-133a-2 (clustered on chromosome 20), miR-1-2 and miR-133a-1 (clustered on chromosome 18) and miR-206 and miR-133b (clustered on chromosome 6) (Goljanek et al, 2012). As miR-133 is involved in proliferation maintenance, miR-1 and miR-206 increase during myogenesis (Cacchiarelli et al, 2010) and promote differentiation, blocking several genes such as Pax3 (Goljanek et al, 2011) and Notch3 (Gagan et al, 2012).

Several other nonmuscle-specific miRNAs are equally involved in myoblast expansion and fiber maturation: miR-489 rapidly decrease after SCs activation, making Dek able to promote proliferation; miR-22, miR-34, miR-145, miR-365 and miR-486 are instead induced during differentiation (Marzi et al, 2012). In particular, miR-486 regulates Pax7 and atrophy-related genes, such as PTEN and FoxO1 (Alexander et al, 2011), preserving muscle mass reduction, as well as miR-23a, which suppresses both MAFbx/atrogin-1 and MuRF1 (Wada et al, 2011), and has been predicted to target myostatin and YY1 (Lin et al, 2009).

1.3 Facioscapulohumeral dystrophy

1.3.1 Clinical characteristics

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common autosomal dominant muscular dystrophy after dystrophinopathies and myotonic dystrophy, affecting approximately 1 in 15,000 individuals worldwide (Flanigan et al, 2001).

Clinical symptoms usually appear during the second decade and are mainly characterized by asymmetrical and progressive muscle weakness and atrophy, initially of the facial, scapular and humeral muscles, involving the abdominal muscles and the musculature of the lower limbs and feet in the late stages of disease. Deltoids remain minimally affected, whereas biceps and triceps are particularly involved, resulting in sparing of the forearm musculature (Lemmers et al, 2014).

Several non-muscle tissues are also frequently affected in FSHD; indeed, an high frequency in hearing loss (75%) and retinal telangiectasia (60%) has been reported (Brouwer et al, 1991; Padberg et al, 1995). Central nervous system defects may also occur, with learning difficulties and epilepsy, particularly evident in some severely affected individuals (Di Lazzaro et al, 2004). Other, less frequent clinical manifestations include respiratory insufficiency and cardiac conduction alterations (Kilmer et al, 1995; Galetta et al, 2005; Trevisan et al, 2006).

1.3.2 Genotype-phenotype correlations

Genotype-phenotype association studies on FSHD affected patients have highlighted two clinical forms of FSH muscular dystrophies (which are referred to as type I and type II respectively) that are indistinguishable from a

symptomatological point of view, although they are linked with two markedly different genotypes.

FSHD type I (also called FSHMD1A) represents more than 95% of clinical cases and is associated with the contraction of tandemly repeated 3.3 kb units of the D4Z4 macrosatellite at the subtelomeric q35 region on chromosome 4 (4q35). In normal individuals, the array varies between 11 and 110 units, whereas FSHD type I affected subjects carry a contraction to 1-10 repeats. Besides that at least one macrosatellite unit is required to develop FSHD, the degree of D4Z4 contraction would seem to correlate with disease severity and a more rapid progression of musculature involvement (Bindoff et al, 2006; Hobson-Webb and Caress, 2006; Klinge et al, 2006).

FSHD type II, instead, is characterized by normal-sized D4Z4 repeat array and reported linked to mutation in SMCHD1 gene, on chromosome 18, encoding a protein containing a hinge domain involved in the structural maintenance of chromatin in a more heterochromatic state. Loss of function mutation leads to decreasing of the levels of SMCHD1 protein and, subsequently, to hypomethylation of D4Z4 macrosatellite (Lemmers et al, 2012). An euchromatic-like D4Z4 repeat array represents the only common feature between FSHD type I and II.

1.3.3 Penetrance and anticipation

Penetrance of FSHD depends on age and gender, resulting in 83% for 30 years old individuals and significantly greater for males (95%) than females (69%) (Zatz et al, 1998; Tonini et al, 2004). Gender differences in penetrance remain poorly understood; analogously, putative mechanism for anticipation is still controversial. Originally suggested by Zatz and colleagues, on observation of multigenerational families, anticipation phenomenon in FSHD appears to have been subsequently excluded by more extensive epidemiologic studies (Flanigan et al, 2001).

1.3.4 Molecular basis of FSHD

Over the years, FSHD has been proved to be a genetically heterogeneous disorder involving both genetic and epigenetic alterations. A detailed genomic characterization of 4q35 region revealed existence of three 4q subtelomere alleles (named A, B and C), fifteen single nucleotide polymorphisms (SNPs) in D4F104S1 (p13E-11) region, and ten simple sequence length polymorphism (SSLP) variants localized 3.5 kb proximal to D4Z4 repeat array. Considering these features, 4q alleles was grouped in eighteen haplotypes (Lemmers et al, 2010a). The 4qA allele differs from the other two by the presence of a 6.2 kb β-satellite sequence (van Geel et al, 2002). FSHD is exclusively associated with 4qA variant (Lemmers et al, 2002; Thomas et al, 2007); furthermore, D4Z4 contraction gives rise to dystrophic phenotype only in three specific haplotype backgrounds: 4qA159, 4qA161 and 4qA168 (Lemmers et al, 2007, 2010b).

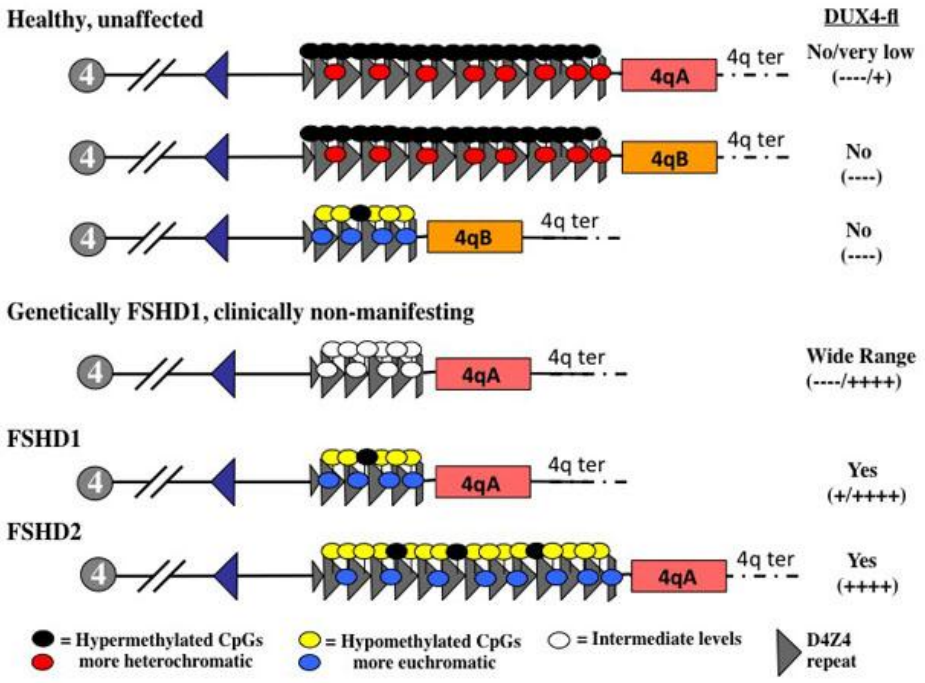


Fig. 8 - Summary of genetic and epigenetic characteristics of 4q35 locus.

However, these haplotypes represent only a permissive condition and by themselves are not sufficient to cause FSHD; indeed, 4qA161 healthy or asymptomatic carriers have been reported (Arashiro et al, 2009).

Further analysis of permissive and non-permissive haplotypes has identified a variant C/T SNP within pLAM sequence, localized distal of D4Z4 array. The SNP can create a consensus polyadenylation signal ATTAAA, found in most FSHD patients. Polyadenylation signal has been observed to be involved in expression of stable DUX4 mRNA. DUX4 open reading frame (ORF) was found to be enclosed within each D4Z4 unit, but it has been hypothesized that its transcription initiates only from the last repeat and proceeds along the distal pLAM region, which provides both an intron and a functional polyadenylation signal, stabilizing the transcript and guaranteeing its translation (Dixit et al, 2007).

Since DUX4 protein was initially detected in FSHD primary muscle cells and only in permissive haplotypes, its expression was considered a consequence of D4Z4 contraction and DUX4 itself became the first gene involved in FSHD etiopathogenesis (Dixit et al, 2007; Bosnakovski et al, 2008; van der Mareel et al, 2011; Tassin et al, 2013). Nevertheless, recent work reported DUX4 also in healthy individuals, taking back expression of this gene in the area of necessary but not sufficient conditions for development of the disease (Jones et al, 2015).

Other FSHD candidate genes localized upstream D4Z4, such as FRG1, FRG2 and SLC24A4 (ANT1), are inappropriately over-expressed in affected muscle (Gabellini et al, 2006), but their role in both initiation and severity of disease is still controversial. Indeed, while initial reports suggested that FRG1 expression levels were elevated in patient muscle cells (Gabellini et al, 2002; Gabellini et al, 2006; van Koningsbruggen et al, 2007), later studies have found no alteration in its expression (Klooster et al, 2009; Arashiro et al, 2009; Masny et al, 2010). Analogously, several analysis of FSHD genetic profile have proposed a causative role for DUX4c, an homologous of DUX4 gene located outside D4Z4 macrosatellite, on chromosome 4 (Bosnakovski et al, 2008a; Ansseau et al, 2009).

However, two FSHD patients were found with a 75 kb deletion that removed both FRG2 and DUX4c on the 4q35 pathogenic allele, suggesting that neither gene is directly responsible of FSHD (Deak et al, 2007).

As mentioned, FSHD has gradually assumed the characteristics of epigenetic-based pathology. Normally, the 4q35 locus is highly methylated, displaying features of unexpressed, strongly heterochromatic region (Jiang et al. 2003;

Yang et al. 2004). Conversely, hypomethylation of both contracted (in FSHD type I) and non-contracted (in FSHD type II) D4Z4 alleles has been reported (van Overveld et al. 2003, de Greef et al. 2007, 2009, 2010). The likelihood is that such epigenetic changes are disease-related and influence the functionality of 4qter region, most probably by the up-regulation of its associated genes (Saccone and Puri, 2010).

1.3.5 DUX4 gene

DUX4 protein was initially described as a double-homeodomain transcription factor (Gabriels et al, 1999) and therefore included within paired class homeodomain protein family, with a function similar to the skeletal muscle stem cell regulators Pax3 and Pax7. In that respect, and in order to explain skeletal myogenic phenotype in FSHD, a model based on DUX4 competition with Pax3/7 for target recognition has been recently proposed (Bosnakovski et al, 2008). However, although DUX4 homeobox sequences are very close to Pax3/7 homeodomain and both contain TAAT core motif, the DUX4 consensus site shows two TAAT motifs in tandem, whereas Pax3/7 site have the same motif in head-to-head orientation (Zhang et al, 2016).

Across D4Z4, each repeated DUX4 gene is composed by one exon terminating with a stop codon; conversely, DUX4 gene in the last repeat of D4Z4 sits next to a sequence polymorphisms harboring two additional untranslated exons, two small introns and a non-canonical polyadenylation signal (pLAM) required for stabilizing DUX4 mRNA.

Intronless DUX4 mapping on first D4Z4 repeats contains two major ORFs: a 135 aa ORF beginning with MER codon that has no homology with proteins deposited in Genbank database; and a 689 aa ORF beginning with MQGR codons, that include two other in-frame ORFs: one beginning with MKG codon and the other corresponding to canonical DUX4 ORF that begins with MAL codon. The last repeat shows sequence variations that extend MER ORF to 256 aa and truncate MQGR after 241 aa (Snider et al, 2009).

Canonical DUX4 mRNA starts upstream MAL ORF in the last D4Z4 unit and undergoes alternative splicing to generate isoforms: a full-length form (DUX4-

fl) containing the two exons and the pLAM region; and a second transcript (DUX4-s) utilizing a cryptic splice donor site in the DUX4 ORF that maintains the N-terminal double-homeobox domains and removes the carboxyterminal end (Snider et al, 2010).

DUX4-fl and DUX4-s are associated to toxic and non-toxic isoforms of DUX4, respectively (Geng et al, 2012; Mistuhashi et al, 2012). An alternative full-length form (DUX4-fl2) characterized by retention of first intron and skipping of stop codon in the first exon has been observed (Dixit et al, 2007; Snider et al, 2009), but its function is still poorly understood.

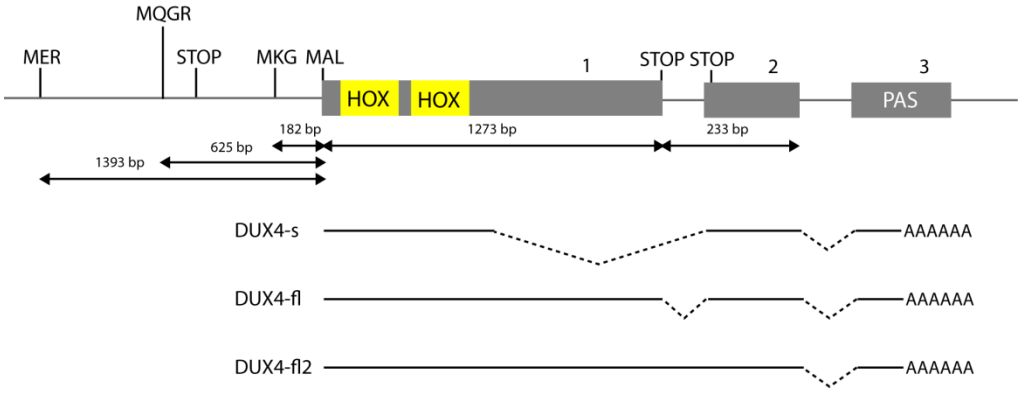


Fig. 9 - DUX4 gene inside D4Z4 last repeat. Schematic representation of DUX4 ORFs and its splicing isoforms. HOX = homeobox. PAS = polyadenylation signal.

Because of the presence of homeodomains and the strong trans-activation property of C-terminus domains (Kawamura et al, 2006), DUX4 was presumed to act as transcription factor. Protein was initially detected into the nucleus by immunofluorescence and three nuclear localization sequences (NLSs) has been reported: NLS1 (RRRR[23]) and NLS2 (RRKR[98]), located at N-terminus portion of homeodomains 1 and 2; and NLS3 (RRAR[148]) at C-terminus of homeobox 2 (Corona et al, 2013). The main target of transcriptional activatory role played by DUX4 is PITX1, a gene specifically up-regulated in FSHD patients (Dixit et al, 2007). However, recently DUX4 has been observed to alter proteostasis by interaction with ubiquitinated protein (Homma et al, 2015) and to translocate into cytoplasm where is involved in binding with a plethora of interactors, such as desmin, α-actinin, tubulin, C1QBP, FUS/TLS, SRSF9, SFPQ and RBM24 (Anseau et al, 2016).

These evidences, in addition to recognition of DUX4-fl forms also in healthy subject (Jones et al, 2015), demonstrate that there are still unclear aspects about the physiopathological role of this protein.

1.3.6 DUX4c

DUX4c (*centromeric* or *contracted*) is a DUX4 homologue, mapping 42 kb proximal to the D4Z4 repeat array, next to FRG2 gene, on chromosome 4. DUX4c is a truncated and inverted D4Z4 unit (Wright et al, 1993) that shares with its homologous a large part of the coding and the proximal sequences (including the two homeoboxes). Indeed, DUX4c ORF extends over 1125 bp as compared to 1273 for DUX4, and both genes share a 1137 bp fragment starting 111 bp upstream their common start codon (MAL), except for three mismatch outside double homeobox domain (Ansseau et al, 2009).

DUX4c is actively transcribed and translated, and was found up-regulated in FSHD myoblasts. Its characterization revealed that mRNA is polyadenylated, although no canonical polyadenylation signal was detected. Moreover, DUX4c has heterogeneous 3'end, whereas 5'RACE data suggests that transcription could be driven by both variant TATAA box and the several GC boxes found in the promoter region (Ansseau et al, 2009).

DUX4c ORF encodes a 47 kDa protein, composed by 347 aa. DUX4c has been shown to regulate Myf5 (Ansseau et al, 2009) and to induce myomiR miR-1, miR-206, miR-133a and miR-133b (Dmitriev et al, 2013) by binding to their promoters. DUX4c is mainly localized into nucleus in proliferating myoblasts, whereas it moves to cytoplasm during myoblast fusion, interacting with several proteins, such as desmin, α-actinin and tubulin (with double homeobox domain), and C1QBP and RBM24 (with C-terminus domains) (Anseau et al, 2016). This data suggested that DUX4c may disturb muscle differentiation, altering cytoskeleton organization and perturbing spliceosome complex.

1.3.7 DUX4-like genes

D4Z4 (4q35.2) has an highly homologous polymorphic repeat array of similar size on subtelomeric region 10q26 (Wijmenga et al,1992; Bakker et al, 1995; Deidda et al, 1995). Indeed, both macrosatellites mapping on chromosomes 4 and 10 are identically tandemly arranged and, in 4qA/10q alleles, followed by a 6.2 kb of 68 bp β-satellite sequences repeated in tandem.

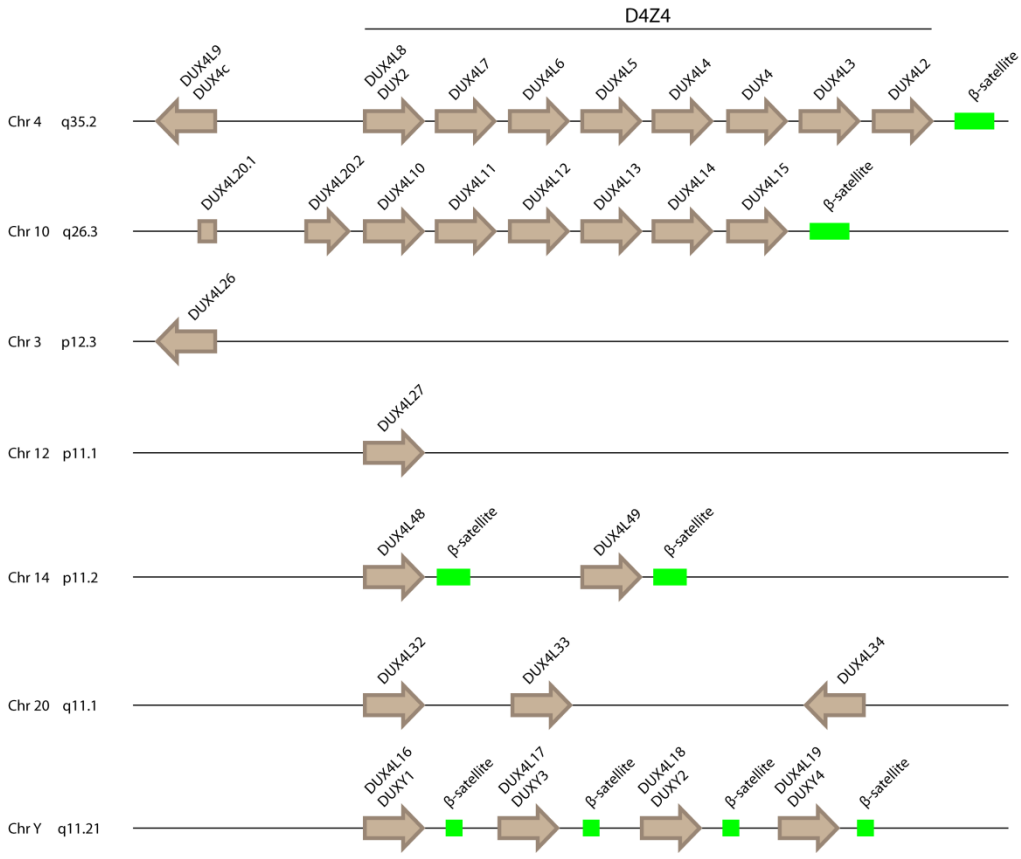


Fig. 10 - Summary of DUX4-like genes and β-satellite sequences in the human genome.

DUX4-like genes on 10q26 conserve their ORF, but the last unit lacks PAS in exon 3 and not produces stable transcripts in skeletal muscle cells (Lemmers et al, 2010). On the other hand, in human testis, a polyadenylated transcript from DUX4-like on 10q array was detected; here 3'RACE experiments confirmed the use of alternative PAS in exon 7, located approximately 6.5 kb distal to pLAM sequence (Snider et al, 2010).

In addition to 10q26 homologous, first southern analysis and experiments of immunofluorescence in situ hybridization (FISH) by using D4Z4-containing

probes suggested the existence of many other sequences related to D4Z4 in the human genome (Wijmenga et al,1992; Hewitt et al, 1994; Winokur et al,1994). Further investigation on somatic cell hybrids initially confirmed the presence on the short arms of acrocentric chromosomes of a 3.3 kb repeat family, having a complex structure, partially similar to D4Z4 (Lyle et al, 1995).

These repeat arrays on acrocentric chromosomes are characterized by a degree of heterogeneity due to the position of 68 bp satellite sequences interspersed between the 3.3 kb repeats (Meneveri et al, 1993) or to distribution of LSau sequences forming part of 3.3 kb repeat itself. However, no active homeobox sequence was found in these D4Z4-related repeat arrays, suggesting that the DUX4 ORF enclosed within 3.3 kb repeats on acrocentrics was disrupted by recombination or transposition events involving other repetitive elements, generating pseudogenes (Lyle et al, 1995).

Tandemly repeated organization of interspersed DUX4-like/satellite sequences was observed also in heterochromatic regions on chromosome 1q12, 9q12, 10cen and 18p11.32, with sequence identities ranging from 80-99% (Winokur et al, 1994, Giussani et al, 2012).

On human chromosome 3p12.3 has been reported an inverted copy (which referred to as DUX4L26) of D4Z4 unit, upstream to FGR2 gene, probably derived from translocation of DUX4c and its proximal regions. Indeed, multiple sequence alignment confirmed sharing identity between DUX4L26 and DUX4c, with putative introns interrupting the coding sequences (Zhang and Holland, 2011).

Several intronless DUX4-like sequences were finally found on 12p11.1 (DUX4L27), 20q11.1 (DUX4L32-34) and Yq11.21 (DUXY1-4), and having assorted degrees of homology with D4Z4 units. In particular, all four Y-chromosomal DUX4-like copies are enframed by β-satellite blocks (in a similar configuration observed on acrocentrics) and lack the potential to encode the second homeodomain (Clapp et al, 2007). Furthermore, DUXY1 ORF shows a 1 bp deletion giving rise to markedly different C-terminus, whereas in the 5' of DUXY4 an Alu insertion was identified (Schmidt et al, 2009).

The presence of an high number of DUX4-like sequences across the human genome suggests the complex evolutionary history of D4Z4 and its coding gene DUX4.

However, unlike FSHD-related DUX4 and DUX4c, transcriptional potential and biological function of other DUX4-like genes remain poorly understood, so much so that they are still considered pseudogenes or junk.

1.3.8 DBE-T

Every D4Z4 unit contains a motif identical to Drosophila PRE sequence (CNGCCATNDNND), overlapping the DBE (D4Z4 binding element) region, recognized by YY1, EZH2 and HMGB2, members of Polycomb protein group (PcG) (Myhaly et al, 1998; Gabellini et al, 2002; Bodega et al, 2009). DBE, as well as the remainder D4Z4, is a CG-rich region, and CpG islands are critical in PcG recruitment (Mendenhall et al, 2010; Neguembor and Gabellini, 2010).
Recently, a lncRNA named DBE-T has been identified, starting from proximal region to D4Z4 macrosatellite, upstream of NDE (nondeleted element). DBE-T was detected as large as 9.8 kb, and it resulted chromatin-associated and over-expressed in FSHD muscle cells versus control (Cabianca et al, 2012).
DBE-T functions *in cis* by directly recruiting the Trithorax group (TrxG) protein Ash1L to 4q35.2 locus. Here Ash1L mediates the de-repression of FSHD locus and overexpression of DUX4, making DBE-T the first activatory lncRNA involved in a human genetic disease (Cabianca et al, 2012).

1.3.9 Other noncoding transcripts from D4Z4 macrosatellite

In addition to DBE-T, other lncRNAs have been recognized inside D4Z4 units, in healthy and FSHD muscle cells: three in sense and three in antisense to DUX4. Furthermore, several mi/siRNA-sized fragments has been observed to originate upstream DUX4 and within both its coding and intronic regions (Snider et al, 2009). Recently, a further study proposed a functional model of DUX4 sense/antisense transcript pairs as precursors of these D4Z4-derived mi/siRNA-sized fragments in a DICER-dependent manner. Researcher hypothesized that these small RNAs can epigenetically suppress DUX4 through gene silencing mediated by AGO2 pathway (Lim et al, 2015).

Interestingly, D4Z4-derived cytoplasmic and/or chromatin/AGO2-associated short RNAs were found also in WI38 primary human fibroblasts and HeLa cells (Benhamed et al, 2012; Ameyar-Zazoua et al, 2012), and tested in turn for their ability to repress DUX4 itself (Lim et al, 2015).

Although the mechanism of action of these transcripts and their involvement in FSHD is still largely unknown, all of these findings highlight the extreme complexity of the macrosatellite D4Z4.

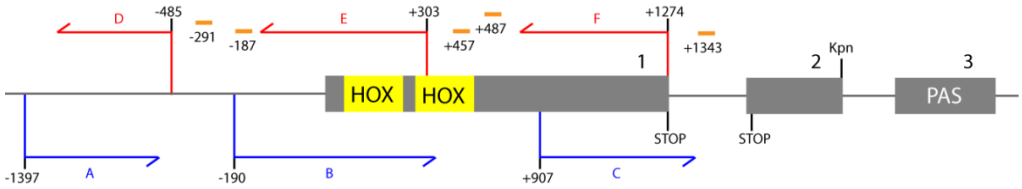


Fig. 11 - Schematic representation of lncRNAs and mi/siRNA-sized fragments crossing DUX4 in healthy and FSHD muscle cells. Numbers represent relative position to MAL start codon of DUX4 ORF. HOX = homeobox. PAS = polyadenylation signal.

Chapter II - Materials and Methods

2.1 Cell lines

Immortalized human muscle cell lines were obtained from Boston Biomedical Research Institute (BBRI, Senator Paul D. Wellstone Muscular Dystrophy Cooperative Research Center for FSHD). Cell lines used in this study have been derived from both biceps and deltoid biopsies of healthy and FSHD subjects. Details about cell strains and clinical characteristics of donors are shown below.

Cohort	Type	Age	Gender	D4Z4 length	Strain	Muscle	MRC
01	CTRL	46	Male	>48 kb	160	Biceps	Full
					175	Deltoid	Full
01	FSHD	42	Male	18 kb	157	Biceps	4+/5
					172	Deltoid	4+/5

Cells were cultured in 0.1% gelatin-coated tissue culture dishes (Corning) with growth LHCN medium (4:1 DMEM : medium 199; 15% FBS; 0.03 μg/ml zinc sulfate; 1.4 μg/ml vit. B12; 0.055 μg/ml dexamethasone; 0.8 mM sodium piruvate; 2.5 ng/ml HGF; 10 ng/ml bFGF; 1% Pen/Strep; 1% Amphotericin B; 0.02 M HEPES), following BBRI's guidelines.

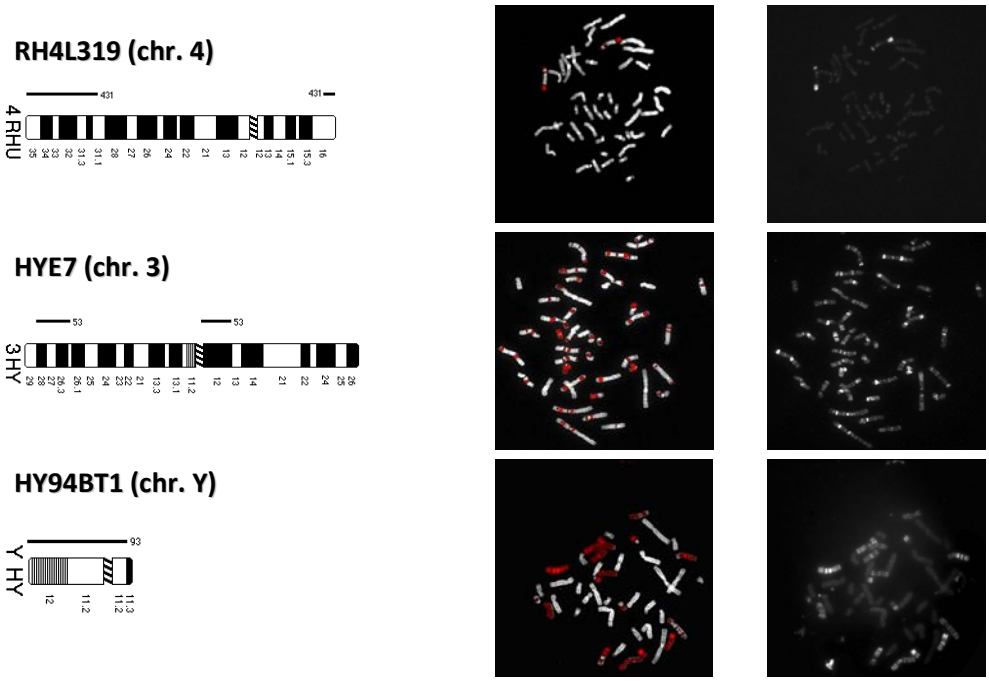
Differentiation was induced at 80-90% of confluence by replacing growth medium with fresh medium containing: 4:1 DMEM : medium 199; 2% horse serum (HS); 1 mM sodium piruvate; 2 mM L-glutamin; 1% Pen/Strep; 1% Amphotericin B; 0.02 M HEPES. We also used 10% KOSR in substitution of 2% HS according to experimental requirements.

We calculate the rate of growth of each cell line by counting the total number of cells in duplicate every day for 7 days, using following Neubauer chamber formula: $[(n° \text{ of cells} \cdot 10^4) / n° \text{ of squares}] \cdot \text{dilution factor}$.

Somatic cell hybrids derived from fusion between human cells and Chinese hamster ovary (CHO) were obtained from Prof. Mariano Rocchi's group at

Department of Anatomical Pathology and Genetics (DAPeG, University of Bari, Italy). Hybrids were cultured in RPMI 1640 supplemented with 10% FBS, 2 mM L-glutamin and 1% Pen/Strep.

Reached 70-80% of confluence, cells were re-plated at lower cell density to prevent cell death. PCP (Partial Chromosome Painting) libraries are free to access at http://biologia.uniba.it/rmc/. Details about human/CHO hybrids used in this study are reported below.



2.2 AZA-TSA treatment

Human/CHO hybrids were seeded at low confluence in growth medium, in order to perform a 72h 5-Aza-2'-deoxycytidine (AZA) and 12h Trichostatin A (TSA) treatment. The day after plating, AZA (Sigma) was added to the medium at final concentration of 1 μM. Every 24h and for other 2 days, the medium was replaced with fresh medium containing 1 μM AZA. For the last 12h of treatment, TSA (Sigma) was added to the medium at final concentration of 1 μM. Cells were collected after 72h AZA and 12h TSA treatment for RNA extraction.

2.3 RNA fractionation

Cells were detached by treating with 1X Tripsin-EDTA (Euroclone), counted and centrifuged at 1000 rpm for 5'. The pellet was lysed with 175 µl/106 cells of cold RLN1 solution (50 mM Tris HCl pH 8.0; 140 mM NaCl; 1.5 mM MgCl2; 0.5% NP-40; 2 mM Vanadyl Ribonucleoside Complex) and incubated 5' in ice. Suspension was centrifuged at 4°C and 2000 rpm for 2', and the supernatant, corresponding to cytoplasmic fraction, was transferred into 1.5 ml fresh tube.

The pellet containing nuclei was treated with 175 µl/106 cells of cold RLN2 solution (50 mM Tris HCl pH 8.0; 500 mM NaCl; 1.5 mM MgCl2; 0.5% NP-40; 2 mM Vanadyl Ribonucleoside Complex) and incubated 5' in ice. Suspension was centrifuged at 4°C and 13000 rpm for 2', and the supernatant, corresponding to nuclear-soluble fraction, was transferred into 1.5 ml fresh tube. Pellet corresponding to chromatin-associated fraction was resuspended in RNase-free water.

2.4 RNA extraction

Total RNA extraction was performed by miRVana miRNA Isolation Kit (Ambion) or by ReliaPrep RNA Miniprep System (Promega). If miRVana Kit was used, DNase I (Thermo Fisher Scientific) treatment was performed separately after purification according .to the instruction of the manufacturer.

All total RNA samples were quantified by using NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific) and the integrity was evaluated on 2100 Bioanalyzer System (Agilent Technologies) by using RNA 6000 Nano Kit (Agilent Technologies). In this study only RNA samples with an RNA Integrity Number (RIN) ≥8 were used.

MicroRNA isolation was achieved by miRVana miRNA Isolation Kit (Ambion) following manufacturer's small RNA species enrichment protocol. The NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific) was used for their quantification.

2.5 3’ Rapid Amplification of cDNA Ends (3’RACE)

3’RACE of polyA+ transcripts was carried out following manufacturer’s instructions (Thermo Fisher Scientific). Retrotranscription was performed on 500 ng of total RNA with SuperScript III First Strand Synthesis System (Thermo Fisher Scientific), using adapter-tailed oligo(dT). First amplification and nested PCR were performed by using Platinum Taq High Fidelity (Thermo Fisher Scientific).

3’ RACE of polyA- was performed by SMART (Switching Mechanism at the 5’ end of RNA Template) protocol, based on using of anchor-tailed random hexamers and riboguanosine oligonucleotides in order to exploit template-switching reverse transcriptase activity. Indeed, this method takes advantage of hybridization between adapter-tailed riboguanosine oligonucleotides and 5’ cytosine tail added during reverse transcription.

In detail, 100 ng of total RNA was added to 1 µl of dNTPs (10 mM each; Thermo Scientific) and 2 µl of pre-mixed anchor-tailed random hexamers (100 µM), adapter-tailed riboguanosine template-switching oligos (100 µM), and trehalose (0.66 M; Sigma-Aldrich) and sorbitol (3.3 M; Sigma-Aldrich), two potent enhancers of cDNA synthesis and priming specificity.

After heating at 65°C for 5’ to melt secondary structures of RNAs and replacing in ice for 1’, we collected the mixture by centrifugation and added the following reagents: 4 µl 5X First Strand Buffer (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl2), 1 µl DTT (0.1 M), 1 µl RNaseOut (40 U/µl) and 1 µl SuperScript III RT (200 U/µl) in a final volume of 20 µl. The reaction mix was pre-incubated at 25°C for 5’ to facilitate random hexamers-template priming, then we proceeded with heating at 50°C for 50’, followed by 15’ at 70°C for RT inactivation. cDNA was finally incubate at 37°C for 20’ with RNase H.

Pre-amplification step was performed in ice, adding 1 µl of cDNA (5 ng) to 3-5 µl 3’-adapter primer (10 µM), 1 µl 5’-adapter primer (10 µM), 2 µl MgSO4 (50 mM), 1 µl dNTPs (10 mM each), 5 µl 10X High Fidelity PCR Buffer (600 mM Tris-SO4 (pH 8.9), 180 mM (NH4)2SO4), 0.5 µl Platinum Taq High Fidelity (5U/µl) and DNase-free distilled water to final volume of 50 µl.

After pre-incubation at 45°C for 1’ and at 94°C for 2’, we set up the following program: 5 x (30’’ 94°C, 30’’ 72°C), 5 x (30’’ 94°C, 30’’ 70°C), 25 x (30’’ 94°C, 1.5’ 68°C), 10’ 68°C. The optimal ratio between 5’- and 3’-adpeter primers was found between 1:3 and 1:5. Respectively, these 5’- and 3’-adapter sequences were used: 5’-GTATCAACGCAGAGTACGATTGATG-3’ and 5’-GACCACGCGTATCGATGTCGAC-3’.

Pre-amplification reaction was useful to enrich low abundant transcripts and reduce background in following nested sequence-specific PCR step. Ever in ice, 1 µl of pre-amplification product was added to 1 µl nested 3’-adapter primer (10 µM), 1 µl tag-specific primer (10 µM), 2 µl MgSO4 (50 mM), 1 µl dNTPs (10 mM each), 5 µl 10X High Fidelity PCR Buffer (600 mM Tris-SO4 (pH 8.9), 180 mM (NH4)2SO4), 3 µl of pre-mixed 1,2-propanediol (0.81 M; Sigma-Aldrich), trehalose (0.66 M; Sigma-Aldrich) and betaine (2.2 M; Sigma-Aldrich), 0.5 µl Platinum Taq High Fidelity (5U/µl) and DNase-free distilled water to final volume of 50 µl. On thermal cycler we set up the following program: 2 min 94°C, 25 x (30’’ 94°C, 30’’ 65°C, 1’ 68°C), 10 min 68°C.

For both 3’RACE standard protocol and alternative method, after every PCR step an aliquot of reaction mix was loaded on 1% TAE (40 mM Tris acetate pH 8.2; 1 mM EDTA) agarose gel, for careful evaluation of amplification patterns.

2.6 Deep sequencing

Sequencing of 3’RACE products (derived from both experimental procedures) was performed on MiSeq sequencing platform (Illumina), in cooperation with ITB-CNR (Segrate, Milan, Italy). Library preparation was carried out by PCR, using barcoded anchor-tailed primers and Platinum Taq High Fidelity protocol (Thermo Fisher Scientific). Libraries was purified by Agencourt AMPure XP (Beckman-Coulter) and evaluated on 2200 TapeStation (Agilent Technologies) by D1000 ScreenTape assay (Agilent Technologies). Library was sequenced using paired-end strategy.

2.7 One-step strand-specific RT-PCR and nested PCR

Strand-specific RT-PCR (ssRT-PCR) was performed combining the cDNA synthesis and amplification reactions in the same tube (one-step). We added 10 µl of total RNA (500 ng) to 2 µl primers mix (anchored strand-specific RT primer (100 µM) and PCR forward and reverse (8 µM each) oligonucleotides), 1 µl dNTPs (10 mM each), 1 µl DTT (0,1 M), 1 µl RNase Out (40U/µl), 5 µl 10X one-step RT-PCR buffer (200 mM Tris HCl, 500 mM KCl, 15 mM MgSO4), 0,5 µl SuperScript III RT (200U/µl), 0,2 µl Platinum Taq High Fidelity (5U/µl) and DNase-free distilled water to final volume of 50 µl.

Reaction mixture was heated at 65°C and temperature was gradually decreased by 1°C every minute. Reached 55°C, temperature was maintained for 45' to synthesize cDNA, then increased to 94°C for 5' to restore polymerase activity, followed by 15 cycles of amplification (30'' 94°C, 30'' 65°C, 1' 68°C) and 5' at 68°C of final elongation step. Terminate the reaction, first amplification product was stored to -20°C or used for nested PCR or qRT-PCR immediately.

Nested PCR was performed added 1 µl of ssRT-PCR product to 1 µl nested forward primer (10 µM), 1 µl nested reverse primer (10 µM), 2 µl MgSO4 (50 mM), 1 µl dNTPs (10 mM each), 5 µl 10X High Fidelity PCR Buffer (600 mM Tris-SO4 (pH 8.9), 180 mM (NH4)2SO4), 0.5 µl Platinum Taq High Fidelity (5U/µl) and DNase-free distilled water to final volume of 50 µl. On thermal cycler we set up the following program: 2 min 94°C, 30 x (30'' 94°C, 30'' 65°C, 1' 68°C), 10 min 68°C. Terminate the reaction, PCR product was then loaded on agarose gel for further evaluations.

The anchored ssRT primers were:

L4 (chr4): 5'-*cgactggagcacgaggacactga*GCCTTTACAAGGGCGGCTGG-3'

L17 (chrY): 5'-*cgactggagcacgaggacactga*TGCAGCCTGCCTGTCTGCGT-3'

L26 (chr3): 5'-*cgactggagcacgaggacactga*GTAGGTCTTACTAAGGGCCT-3'

First amplification primers for all chromosome-specific sequences were:

5'-CGACTGGAGCACGAGGACACTGA-3'

5'-AATCTGGACCCTGGGCTCCGGAATGC-3'

as forward and reverse primers. Nested forward primers for chr4/Y and chr 3 were: 5'-TGGCTGGACCTGCCTGCAGC-3' and 5'-GCTGGCTGGCTGTCCG-3'. Common nested reverse primer was: 5'-CGTTCTCTGGTGGCGATG-3'.

2.8 Cloning and Sanger sequencing

Amplicons of antisense transcripts obtained by nested PCR and loaded on 1.5% TAE (40 mM Tris acetate pH 8.2; 1 mM EDTA) agarose gel was extracted by using Wizard SV Gel and PCR Clean-Up System (Promega). Cloning was performed in DH5α competent cells by TOPO TA Cloning Kit for Sequencing (Thermo Fisher Scientific), following manufacturer's instructions. Positive clones were screened by colony PCR, using T3 and T7 primers, then sequenced at Human Molecular Genetics Consortium (Monza, Italy).

2.9 Quantitative Real-Time PCR

qRT-PCR analysis was performed on 7900 HT Fast Real-Time PCR System (Applied Biosystem) by GoTaq qPCR Master Mix (Promega) to quantify antisense transcripts and the other genes of interest.

In absence of endogenous normalizers for ssRT-PCR, the quantification was achieved by standard curve method. We used serial dilutions of known amount of antisense DNA amplicon (starting from in vitro transcribed RNA) in order to create standard curve and verify efficiency and linearity of reaction. Unknown sample quantity was determined by linear interpolation of Ct on standard curve graph and then represented as fold change over control myoblasts (T0). For gene expression analysis, we used instead comparative $2^{-\Delta\Delta Ct}$ method, with GAPDH as endogenous normalizer.

Thermal cycling conditions were 2' at 94°C, followed by 40 cycles at 94°C for 10'' and 60°C for 30'', for gene amplification, or by 40 cycles at at 94°C for 10' and 65°C for 1', for antisense transcripts amplification. All qRT-PCR primers are listed in following page.

qRT-PCR primers

Target	Sequence (5’-3’)
MYOD-F	CCGCCTGAGCAAAGTAAATGA
MYOD-R	GCAACCGCTGGTTTGGATT
MYF5-F	CGAATGTAACAGTCCTGTCTGG
MYF5-R	AGGTTGCTCTGAGGAGGTGA
MYOG-F	TGCTCAACCCCAACCAGCGG
MYOG-R	TTCACTGGGCACCATGGGCTG
MEF2A-F	GTGTACTCAGCAATGCCGAC
MEF2A-R	AACCCTGAGATAACTGCCCTC
MEF2C-F	TTCCAGTATGCCAGCACCG
MEF2C-R	GGCCCTTCTTTCTCAACGTCTC
DUX4-FL-F	GAGCTCCTGGCGAGCCCGGAGTTTCTG
DUX4-FL-R	CTAAAGCTCCTCCAGCAGAGCCCGGTATTCTTCCTC
TSC1-F	TAGGGCACAATGAAGAGGCA
TSC1-R	CGCTTCTCCCATAGTCGTCT
TSC2-F	GACGCCTTAAAGAGCAGAGC
TSC2-R	GATGGATCTGGTCGAGGAG
GβL-F	ACGGCGTCAACAAGAACATC
GβL-R	GCAGTTAATGGGTGCGTTCA
RAPTOR-F	ACTGATGGAGTCCGAAATGC
RAPTOR-R	TCATCCGATCCTTCATCCTC
RICTOR-F	GGAAGCCTGTTGATGGTGAT
RICTOR-R	GGCAGCCTGTTTTATGGTGT
AKT2-F	CCCTTCTACAACCAGGACCA
AKT2-R	AACCTGTGCTCCATGACCTC
PLD1-F	AATCGTTGGAGGTTGGACTG
PLD1-R	AGACGGTGGATGACACATGA
S6K1-F	GCATGCTCCTACGCTGAACT
S6K1-R	TGTCCTCAGCTTCCCTGTCT
MSTN-F	CGCTACAACGGAAACAATCA
MSTN-R	GAGTCTCGACGGGTCTCAAA
SMAD3-F	GCCTGTGCTGGAACATCATC
SMAD3-R	TTGCCCTCATGTGTGCTCTT
FOXO3-F	AGTGGATGGTGCGCTGTGT
FOXO3-R	CTGTGCAGGGACAGGTTGT
FOLL-F	TGCACTCCTAAAGGCAAGATG
FOLL-R	CTGGGCAATCCGATTACAG
GAPDH-F	TTCACCACCATGGAGAAGGC
GAPDH-R	GGCATGGACTGTGGTCATGA

2.10 MiRNA retrotranscription and quantification

Small RNA fraction isolated from control and FSHD myoblasts and myotubes was used for microRNA retrotranscription and amplification by miRCURY LNA Universal RT microRNA PCR (Exiqon), following manufacturer's instructions. CTX (500 nM) was used as passive reference and qRT-PCR was performed on 7900 HT Fast Real-Time PCR System (Applied Biosystem). Custom primer sets for novel miRNAs was designed by Exiqon with its proprietary tool (https://www.exiqon.com/mirna-qpcr-designer).

Data was analyzed with comparative $2^{-\Delta\Delta Ct}$ method. We used has-miR-103a as endogenous control and nomalizer, and UniSp6 as RNA spike-in for transcription efficiency evaluation. Thermal cycling conditions were 10' at 94°C, followed by 45 cycles at 94°C for 10'' and 60°C for 1'. Melting curve analysis was performed with SDS software v1.4 (Applied Biosystem).

2.11 In vitro transcription

Vector pMK-RQ (GeneArt) containing DUX4-AS2 consensus sequence flanked by T7 and SP6 promoters was treated with restriction enzyme in order to cutting out and use insert (541 bp) as template for in vitro transcription. Purified construct (10 μg) was incubated with 20 U of SfiI enzyme and 1X CutSmart Buffer (NEB) at 50°C for 1h.

Reaction mix was loaded on 1.5% TAE (40 mM Tris acetate pH 8.2; 1 mM EDTA) agarose gel to evaluate cutting efficiency and isolate insert by gel extraction. Insert was purified from agarose slice by using Wizard SV Gel and PCR Clean-Up System (Promega) and assayed by Nanodrop 2000 UV-Vis Spectrophotometer (Thermo Scientific).

About 1.85x1011 copies of insert (0.1 μg) was used to assemble reaction at room temperature with MEGAscript T7 Transcription Kit (Ambion) and MEGAscript SP6 Transcription Kit (Ambion) reagents, respectively for sense and antisense transcript synthesis and following manufacturer's instructions.

Transcription reaction mixture was incubated at 37°C for 4h and finally treated with 2 U of TURBO DNase at 37°C for 15' to totally remove template DNA.

In vitro transcribed RNA was purified by using RNA Clean & Concentrator 100 (Zymo Research), evaluated by run on 1.5% MOPS (20 mM MOPS pH 7.0; 1 mM EDTA; 5 mM NaOAc) agarose-formaldehyde (37% formaldehyde to final concentration of 0.7 M; Sigma) gel and quantified by NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific).

2.12 5'end capping and 3'end biotinylation

After in vitro transcription, 10 μg of purified DUX4-AS2 RNA was heated at 65°C for 5' to remove secondary structure and placed in ice for other 5'. Capping reaction was performed by using Vaccinia Capping System (NEB), following manufacturer's instructions. The reaction time was extended to 1h to improve capping efficiency.

Capped RNA was purified by using RNA Clean & Concentrator 100 (Zymo Research) and resuspended in 6 μl of RNase-free water.

Biotin labeling of DUX4-AS2 RNA was performed by using Pierce RNA 3' End Biotinylation Kit (Thermo Fisher Scientific), following manufacturer's instructions. Reaction was set up with 50 pmol of RNA and incubated overnight at 16°C. Before ligation, RNA was pre-heated at 85°C for 4' in presence of 25% of DMSO to increase efficiency.

Reaction byproducts was removed with Chlorophorm:isoamyl alcohol 24:1 (Sigma), and labeled RNA was precipitated (1h at 20°C) with ice-cold 100% ethanol in presence of 20 μg/μl glycogen and 5M NaCl. Precipitate was centrifuged at 13000 g for 15' at 4°C, supernatant removed and pellet washed with ice-cold 70% ethanol. Finally ethanol was removed and air-dried pellet resuspended in 20 μl RNase-free water.

2.13 RNA-protein pull down assay

Interaction between DUX4-AS2 and DGCR8 complex was validated by RNA-protein interaction assay using Pierce Magnetic RNA-Protein Pull-Down Kit (Thermo Fisher Scientific).

Cell lysate of healthy myoblasts at 90% of confluence in 100 mm treated tissue culture dishes (Corning) was prepared using RIPA buffer [50 mM Tris HCl pH 7.4; 150 mM NaCl, 0.1% SDS, 0.5% Deoxycholate Sodium, 1% NP-40 and protease inhibitor cocktail 1X (Prod. N. P2714-1BTL; Sigma)], extracting both cytoplasmic and nuclear protein by sonication in Branson Ultrasonic Bath (Hach), followed by 15 min of centrifugation a 13000 rpm at 4°C to discard cell debris.

Binding of labeled RNA (100 pmol) to pre-washed Streptavidin Magnetic Beads (15 µl) and loading of protein lysate (2 mg/ml) was performed following manufacturer's instructions. RNA-binding proteins complex was eluted in 50 µl of Elution Buffer provided by Pull-Down Kit for downstream application. Complementary sense RNA of DUX-AS2 was used as negative control.

2.14 Long non-coding RNA, siRNA and miRNA transfection

Myoblasts were plated in 60 mm (surface area 20 cm^2) treated tissue culture dishes (Corning) to be 70-80% confluent at transfection. Capped DUX4-AS2 RNA (5 µg) was transfected by using Lipofectamine MessengerMAX Reagent (Invitrogen) in Opti-MEM medium (Euroclone). Complementary sense RNA, in vitro transcribed from vector pMK-RQ-AST2, was used as negative control.

Single strand miRNA as-miR-276-5p and as-miR-320-3p (60 pmol each), as well as siRNA against DUX4 (s196453, Thermo Scientific) were individually transfected by using Lipofectamine RNAiMAX Reagent (Invitrogen) in Opti-MEM medium (Euroclone). The mirVana miRNA Mimic Negative Control #1 (Ambion) was used as scrambled sequence. Cells were incubated at 37°C and monitored for 2 days, then we proceeded with RNA or protein extraction.

2.15 SDS-PAGE and immunoblotting

Cell lysate was prepared using RIPA buffer [50 mM Tris HCl pH 7.4; 150 mM NaCl, 0.1% SDS, 0.5% Deoxycholate Sodium, 1% NP-40 and protease inhibitor cocktail 1X (Prod. N. P2714-1BTL; Sigma)], extracting cytoplasmic and nuclear protein by sonication in Branson Ultrasonic Bath (Hach), followed by 15 min of centrifugation a 13000 rpm at 4°C to discard cell debris. Sample preparation and quantification, and western blot analyses were performed as previously described (Pisconti et al, 2006). After electrophoresis, polypeptides were electrophoretically transferred to nitrocellulose filters (Thermo Scientific) and antigens revealed by respective primary antibody and the appropriate secondary HRP-linked antibody, through enhanced chemiluminescence (LiteAblot Plus, Euroclone). In immunoblotting analysis, antibodies specific for MyoD (1:1000) (5.8A, monoclonal mouse, Thermo Scientific), MyoG, CAV3, DUX4 (1:1000) (P4H2, monoclonal mouse, Thermo Scientific), MSTN (1:1000) (6H12, monoclonal mouse, Thermo Scientific), TSC1 (1:1000) (PA5-19502, polyclonal rabbit, Thermo Scientific), RAPTOR (1:1000) (24C12, monoclonal rabbit, Cell Signaling), RICTOR (1:1000) (53A2, monoclonal rabbit, Cell Signaling), GβL (1:1000) (86B8, monoclonal rabbit, Cell Signaling) and sarcomeric myosin MHC (1:10) (MF20, from Developmental Studies Hybridoma Bank) were used.

2.16 Immunofluorescence

Cell immunofluorescence was performed as previously described (Brunelli et al, 2004), using antibodies specific for Desmin (1:20) (D8281, polyclonal rabbit, Sigma Aldrich), Ki67 (1:250) (SP6, monoclonal mouse, Abcam), MyoD (1:500) (5.8A, monoclonal mouse, Thermo Scientific) DUX4 (1:100) (P4H2, monoclonal mouse, Thermo Scientific), M-chaderin (1:200) (PA5-47578, polyclonal sheep, Thermo Scientific) and sarcomeric myosin MHC (1:4) (MF20, from Developmental Studies Hybridoma Bank). Appropriate secondary antibodies conjugated with Alexa 488 (green, Cell Signalling) or

Alexa 568 (red; Cell Signalling) were used for fluorescence detection. Nuclei were stained with Hoechst Stain Solution (H6024, SIGMA).
Fluorescent images were taken on confocal laser scanning microscope (Zeiss Lsm 01, Biorad mrc 600, Biorad 1024) by using 12× magnification. Images showing double or triple fluorescence were separately acquired by using appropriate filters, and the different layers were merged and analyzed with ImageJ software.

2.17 Flow cytometry

Cell were detached by treating with 1X Tripsin-EDTA (Euroclone), counted and centrifuged at 1000 rpm for 5 min, then washed with PBS and incubated 15 min in PBS supplemented with 5% of heat-inactivated calf serum to block non-specific sites. Incubation with primary antibody was performed for 15 min at room temperature. When secondary antibody was needed, cells were stained with AlexaFour 488-conjubated goat anti-mouse IgG (1:100) for 30 min at 4°C. Cells were assayed by using PE-conjugated CD56 (HCD56, monoclonal mouse, BioLegend) and anti-fibroblasts antibodies (TE7, monoclonal mouse, Sigma-Aldrich). The acquisition process was stopped when 15,000 events were collected in the population gate.

2.18 Target prediction, gene ontology and pathway analysis

For novel miRNA target prediction we mainly used miRTar tool (Hsu et al, 2011), an integrated system, based on miRanda, PITA and TargetScan algorithms, setting cutoff of minimum free energy (MFE) of the miRNA-target duplex to -14 kcal/mol and threshold of total score to 140. We also used RNAhybrid (Rehmsmeier et al, 2004) for miRNA-3'UTR negative normalized MFE calculation (cutoff of -12 kcal/mol and seed 2–8 allowing G:U wobble pairs) and RNA22 (Miranda et al, 2006) for evaluation of miRNA-5'UTR/CDS interaction, using following parameters: sensitivity of 63%, specificity of 61%, seed size of 7 allowing 1 un-paired bases and no limit of G:U wobble pairs, and

maximum folding energy of -12 kcal/mol. Putative targets were finally tested for miRNA repression strength by using miRmap (Vejnar et al, 2013), then subjected to analysis of Gene Ontology terms and KEGG pathway enrichment by DAVID (Database for Annotation, Visualization and Integrated Discovery) (da Huang et al, 2009).

2.19 Sequence homology-based conserved domain analysis

Search of DUX4-fl conserved domains was performed by using PSI-BLAST tools, to align amino acid sequence of DUX4-fl with non-redundant sequences collected in several datasets: GenBank, RefSeq, PDB, SwissProt, PIR and PRF. Alignment score was determined by BLOSUM62 matrix and setting following gap coasts: 11 for existence and 1 for extension. Statistical significance threshold to include a sequence in PSSM (Position Specific Score Matrix) on next alignment iteration was 0.005. Analysis was aborted after three iteration cycles. Obtained sequences were clustered by using CLANS software, basing on homology level, query covering and phylogenetic distance.

2.20 Secondary structure analysis

Prediction of α-helix and β-sheet motifs was obtained from PSIPRED (Buchan et al, 2013) and Jpred4 (Drozdetskiy et al, 2015) servers. Trans-membrane and coiled-coils topology were sought by TMHMM (Krogh et al, 2001) and COILS (Lupas et al, 1991) algorithms. Low complexity regions were predicted by CAST and intrinsic protein disorder was evaluated by DisEMBL tools, using following algorithm parameters: 8 of Savitzky-Golay smoothing frame, 8 of minimum peak width, 4 of maximum join distance; and predictors stringency values: 1.20 (Coils and Remark465 threshold) and 1.40 (hot loops threshold).

2.21 De novo modeling and structure validation

DUX4-fl was submitted to Rosetta software for full-chain structure prediction. Models and conformational domains were further evaluated by using ProSA (Wiederstein et al, 2007) and QMEAN (Benkert et al, 2009), for z-score calculation, and MolProbity (Chen et al, 2010), for Ramachandran plot and other geometric properties evaluation, structure-validation web services, then visualized by PyMOL Molecular Graphics System, v.1.8.

2.22 Molecular dynamics simulation

We carried out simulations using GROMACS 4.5 (Hess et al, 2008) and we used the trjconv, rms and gyrate tools in the GROMACS package to analyze the simulation data. We performed energy minimization in all-atom OPLS or GROMOS96 43a1force fields, heating from 0 K to 300 K, then running the NPT simulation for 2 ns at 300 K at 1 bar pressure.

2.23 Statistical analysis

For qPCR data of antisense transcript expression a one-tail, unpaired t-test was used. For all other immunofluorescence, qPCR and western blot experiments, p-value was evaluated by two-tail, paired t-test. All error bars on the graphs are referred to standard deviation.

Chapter III - Results

3.1 Characterization of FSHD myogenesis

3.1.1 FSHD myoblasts exhibit early commitment to myogenesis, fusing into disorganized myotubes, when compared to control muscle cells

Previous works reported that affected myoblasts fuse to give rise either thin and branched myotubes or abnormally swelled myotubes with disorganized nuclei distribution, compared to large branched myofibers with well-aligned nuclei deriving from healthy skeletal muscle cells (Barro et al, 2010; Vanderplanck et al, 2011).

Despite of these morphological alterations, the molecular mechanisms driving pathologic differentiation are still poorly understood. Indeed, several studies of gene expression analysis have been focused on identification of dysregulated pathways and toxic effects of 4q35.2 genes, explaining the altered outcome of FSHD myogenesis, but omitting dynamics underlying myoblast commitment, migration and fusion into myotubes.

In this study, we took advantage of immortalized myoblasts derived by biceps and deltoid muscle biopsies recovered from multiple cohorts of FSHD donors and their healthy first-degree relatives (Homma et al, 2012). In particular, we mainly used 157A (affected biceps) and 160U (unaffected biceps) from cohort 01. In necessary, both 172A (affected deltoid) and 175U (unaffected deltoid) cell strains was used as well (see Material and Methods for details).

At the outset, we evaluated cell identity and purity by flow cytometry, in order to exclude the presence of non-muscular contaminants (CD56 negative cells) retained during clonal selection of immortalized strains. As expected, in each sample analyzed, CD56+ cells were more than 99%. (Fig. 1A). Secondly, we analyzed the growth curves to estimate propagation efficiency and velocity of

each cell line in proliferation medium (LHCN), finding that both biceps- and deltoid-derived affected myoblasts are characterized by a slower growth rate than controls (Fig. 1B).

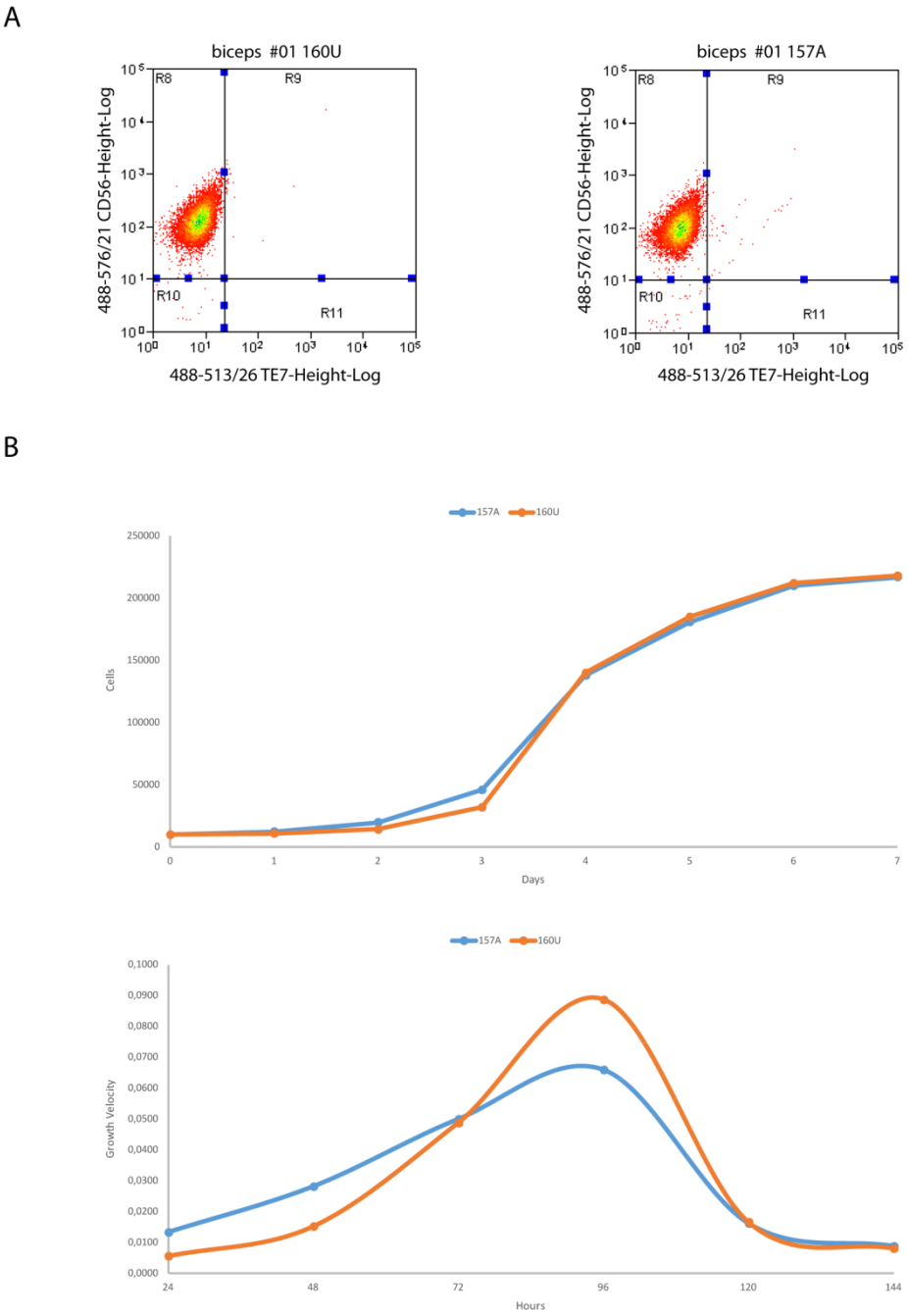


Fig. 1 - A) Analysis by flow cytometry of expression of CD56 (muscle-specific maker) and TE7 (fibroblast-specific marker) in biceps control and FSHD myoblasts. **B)** Seven days growth curves of biceps control and FSHD myoblasts.

We choose to inspect the entire differentiation process, through setting up of time course experiments. Both FSHD and healthy myoblasts were then switched to differentiation medium containing 2% of horse serum (HS) and analyzed every 2 days until the complete maturation of myotubes (6-8 days).

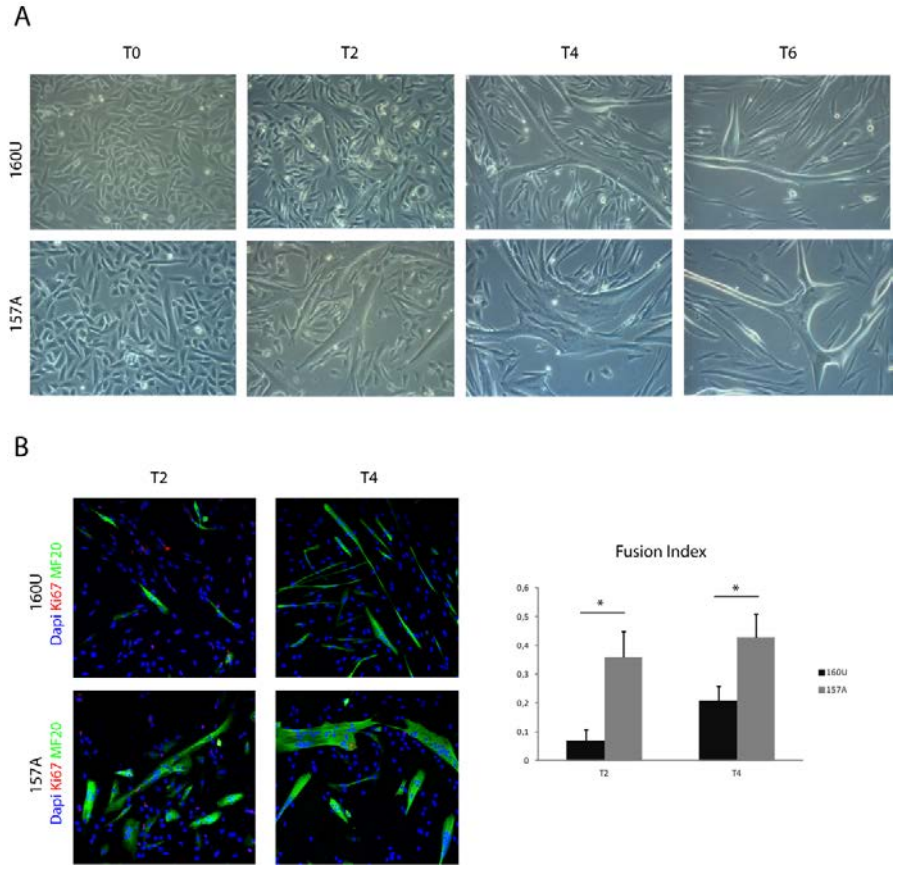


Fig. 2 - **A)** Phase contrast microscope images of control and FSHD muscle cells during different stages of differentiation. **B)** Analysis by immunofluorescence of healthy and FSHD cells after 2-4 days of differentiation for marker expression of cell proliferation (Ki67, in red) and differentiation (MF20, in green). Nuclei were stained with DAPI (blue). Fusion index was compared between control and FSHD. Error bars indicate standard deviation. N = 6; * p-value < 0.05.

Morphological evaluation and fusion rate calculation were performed by microscopy techniques, whereas expression level of different muscle-specific markers was assayed by qRT-PCR, western blot and immunofluorescence too.
Results supported hypothesis of early differentiation in FSHD myoblasts, when compared to healthy controls. Moreover, we observed disorganized migration and uncontrolled fusion in affected cells, which starts after two days in

differentiation medium and proceeds without a correct alignment along the primary myotubes (Fig. 2A).

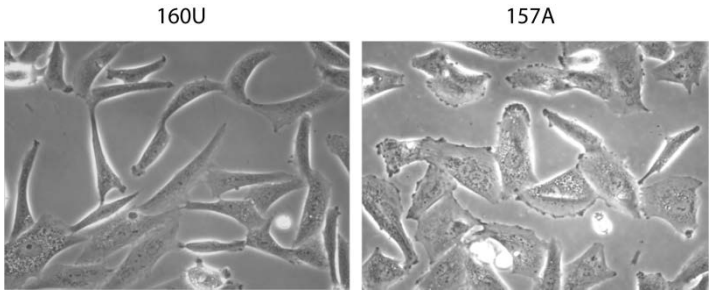


Fig. 3 - Phase contrast microscope images of control and FSHD proliferating myoblasts. Dark flanges are appreciable on plasma membrane of affected cells.

As revealed by immunofluorescence, the fusion index (which is referred to as number of nuclei into myotubes over the total number of nuclei) of FSHD myoblasts is significantly higher than controls (Fig. 2B).

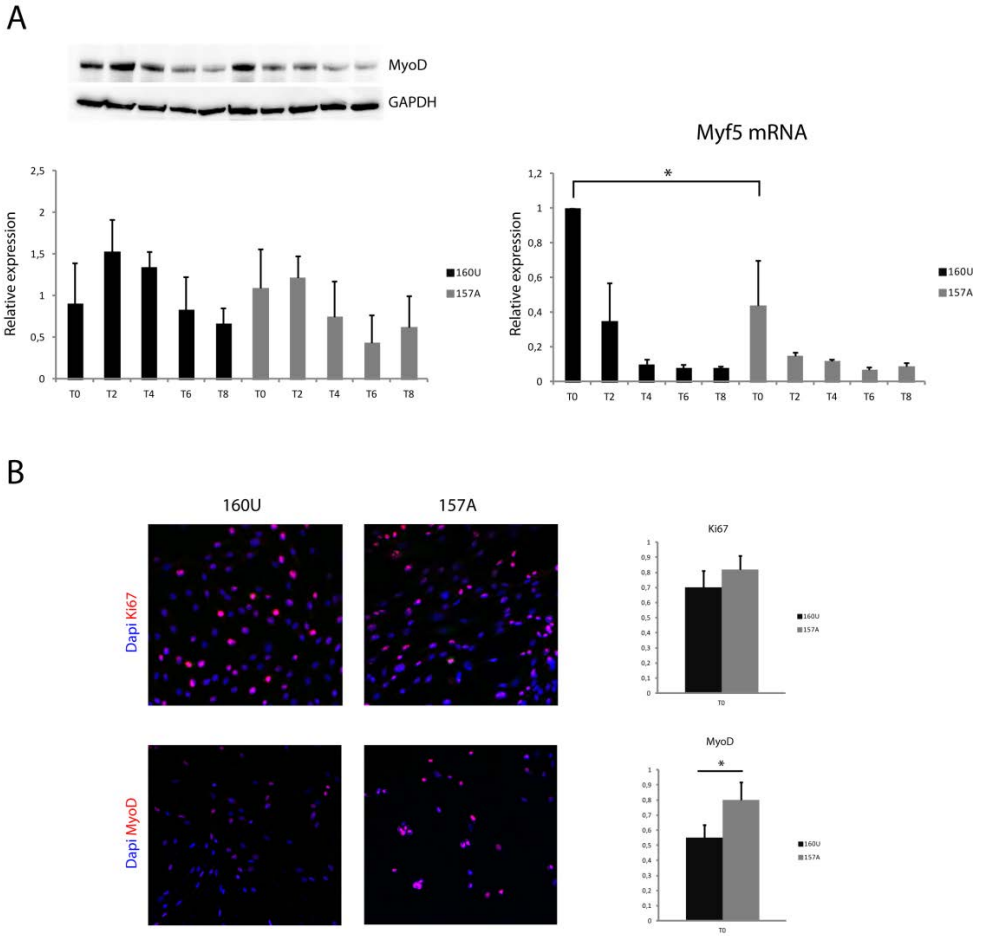


Fig. 4 - **A)** MyoD protein and Myf5 mRNA levels were evaluated by western blot and qRT-PCR respectively, in control and FSHD muscle cells during different stages of differentiation. GAPDH was used as normalizer. **B)** Quantification by immunofluorescence of Ki67+ and MyoD+ control and FSHD proliferating myoblasts. Nuclei were stained with DAPI (blue). Error bars indicate standard deviation. N = 5; * p-value < 0.05.

Overall, some morphological differences have been also recognized during the first stages of differentiation: FSHD myoblasts do not stretch and polarize themselves, showing a disordered topographic patterning of cell-surface fusion interfaces (Fig. 3).

Myogenic determination factors Myf5 is significantly down-regulated in FSHD myoblasts versus controls, suggesting that affected cells are prone to escape cell cycle and start differentiation (Fig. 4A). Particularly, to exclude that affected proliferating cells had already started the differentiation program, we evaluated the ratio of Ki67+ nuclei over total of nuclei, without finding variations (Fig. 4B).

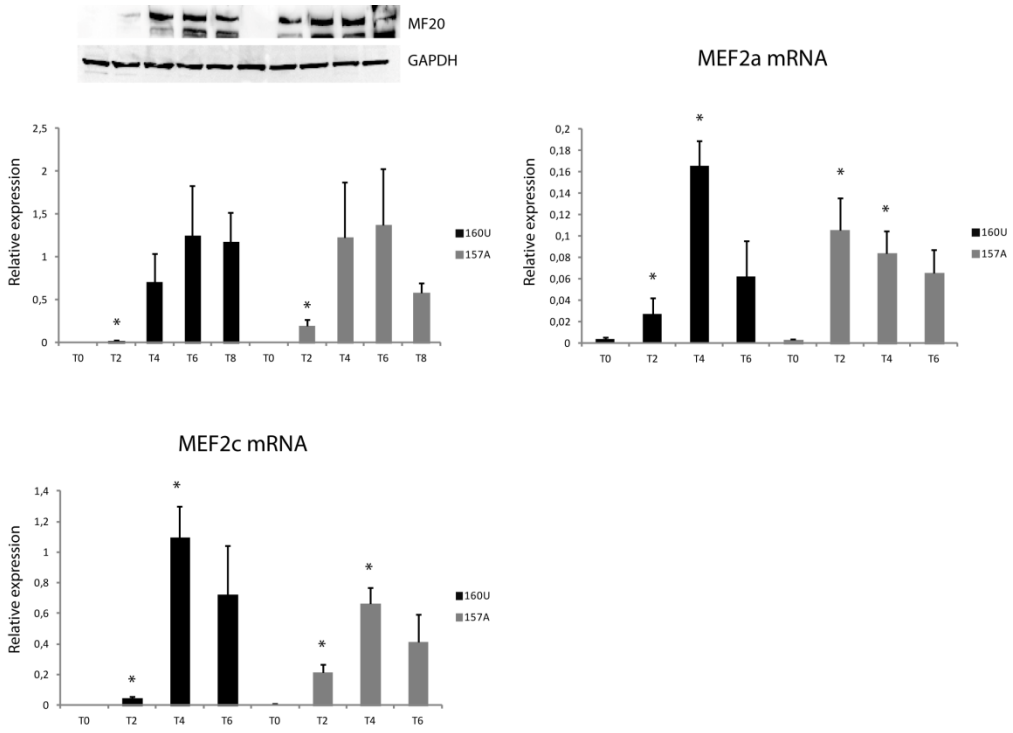


Fig. 5 - MF20 protein and MEF2a/MEF2c mRNA levels were assayed by western blot and qRT-PCR respectively, in biceps control and FSHD muscle cells, during different stages of differentiation. GAPDH was used as endogenous normalizer. Error bars indicate standard deviation. The statistical significance was evaluated between control and FSHD, at same time point. N = 3; * p-value < 0.05.

No significant differences, instead, were observed by western blot in MyoD levels (Fig. 4A), despite of higher number of MyoD+ nuclei was detected in FSHD myoblasts (Fig. 4B). Conversely, other muscle-specific markers, such as Mef2a, Mef2c and MF20, confirmed early differentiation in affected cells. Interestingly, both Mef2a and Mef2c, following a significant increase of their expression levels in FSHD myoblasts at T2, appeared lower than controls at T4 (Fig. 5). Similar results were obtained on deltoid-derived cell lines, although their differentiation rate, in FSHD, appeared slightly slower compared to biceps-derived myoblasts (data not shown).

3.1.2 DUX4 full-length isoforms are expressed in FSHD and control myoblasts and increase during myogenic differentiation

We evaluated our cell model for DUX4 expression, particularly for its toxic full length isoforms. Protein quantification was performed by western blot, during all stages of differentiation. Interestingly, we found DUX4-fl expressed in both control and FSHD cell lines, observing an increase of its level during myoblast fusion and myotubes formation processes.

Surprisingly, no quantitative difference was highlighted for DUX4-fl between control and FSHD, discrediting the exclusively pathological role attributed to this protein (Fig. 6).

By using an antibody (P4H2) against C-terminus of DUX4, we also detected DUX4-fl2, occurring preferentially during late stages of muscle differentiation.

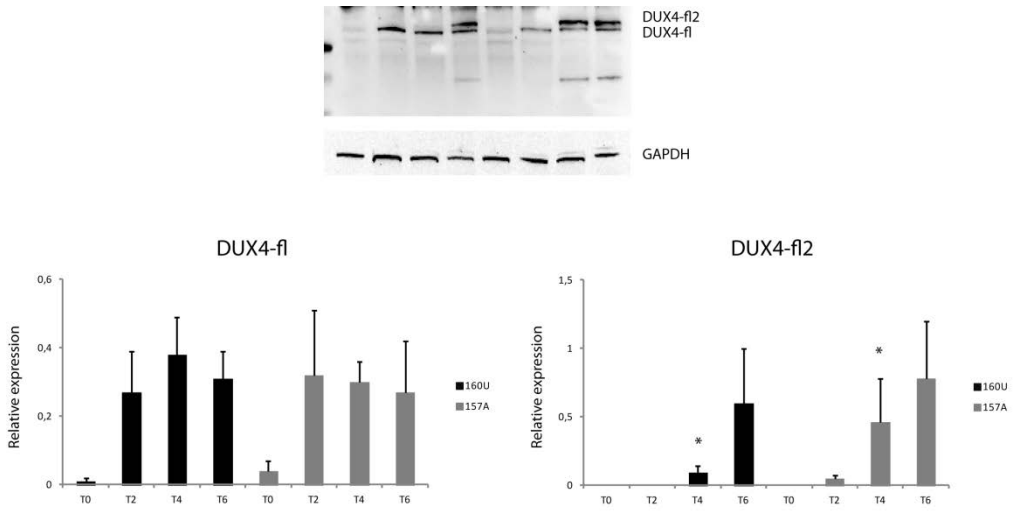


Fig. 6 - Protein levels of DUX4 full length isoforms in control and FSHD muscle cells during differentiation. GAPDH was used as endogenous normalizer. Error bars indicate standard deviation. The statistical significance was evaluated between control and FSHD, at same time point. N = 3; * p-value < 0.05.

Moreover, according to evidences of premature myogenesis in FSHD cell lines, DUX4-fl2 was found from the fourth day, in affected cells, whereas in healthy control appeared after 6 days only (Fig. 6). Results were comparable by using myoblasts deriving from both biceps and deltoid (not shown).

3.1.3 Use of knockout serum replacement (KOSR) as alternative to HS accelerates myogenic differentiation in FSHD and control myoblasts by enhancing DUX4 expression

Previous studies demonstrated that DUX4 expression is negatively regulated by Wnt/β-catenin signaling (Block et al, 2013). The LHCN medium we used to grow myoblasts contains dexamethasone, which is reported to disrupt Wnt/β-catenin signaling (Almeida et al, 2011). Therefore, we expected to find DUX4 expression favored in proliferation medium. On the other hand, dexamethasone has been shown to increase apoptosis and ROS generation (Oshima).

Analogously, Wnt/β-catenin signaling has been observed to be suppressed by serum deprivation in cancer cells (Khan et al, 2007). This finding could explain DUX4 increasing after switch to differentiation medium. However, latest work demonstrated that growing myoblasts without dexamethasone and replacing HS with the artificial serum KOSR (Life Technologies, Frederick, MD, USA) in differentiation medium, DUX4 significantly increased by Wnt suppression and consequently skeletal muscle cells appeared to be protected from oxidative stress (Pandey et al, 2015).

For these reasons, we chose to modify the cell culture conditions according to Pandey's protocol, removing dexamethasone in growth medium and replacing 2% HS with 10% KOSR in differentiation medium. As previously observed (Pandey et al, 2015), the enhancement of DUX4 expression correlates with increased differentiation rate in both control and FSHD muscle cells (Fig. 7A): complete myotube maturation, in KOSR-added medium, occurs after 4-6 days only, instead of 6-8 days in traditional differentiation medium.

However, affected myoblasts remained able to fuse into myotubes faster than healthy cells. Indeed, as confirmed by western blot analysis of several markers of differentiation, such as MyoD, MF20, MyoG and caveolin-3 (CAV3), new culture conditions enhanced FSHD characteristics of altered myoblast fusion and aberrant myotubes formation (Fig. 7B).

The model made time course experiments more reproducible and manageable, compared to conventional culture conditions of both immortalized or primary muscle cells. Consequently, in order to overcome limits of previous culture

methods, mainly concerning negative effects on D4Z4 transcription efficiency, we decided to implement this protocol for most of the following experiments.

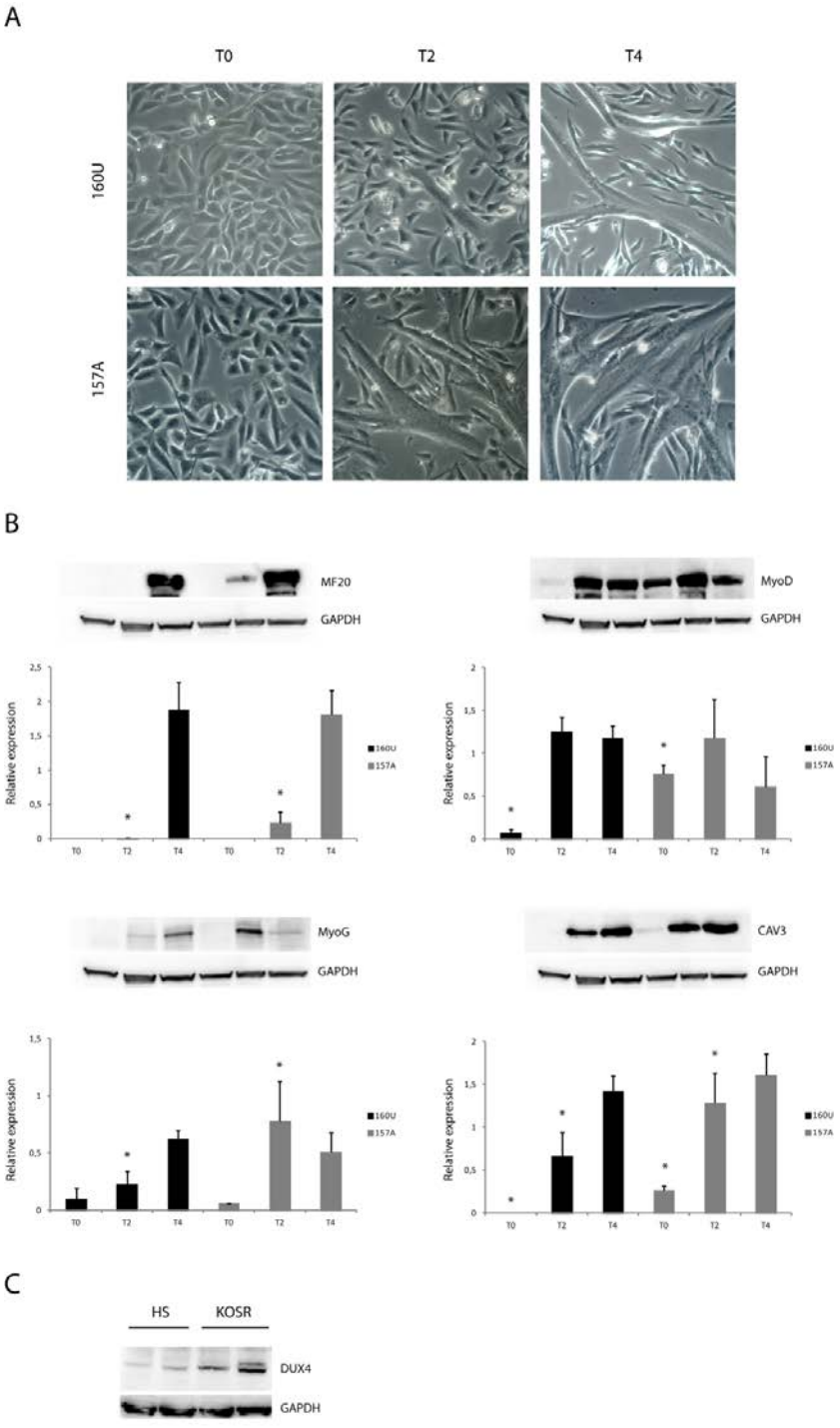


Fig. 7 - A) Phase contrast microscope images of control and FSHD muscle cells during differentiation in KOSR-added medium. **B)** Western blot analysis of muscle-specific markers, during differentiation in KOSR-added medium. GAPDH was used as endogenous normalizer. Error bars indicate standard deviation. The statistical significance was evaluated between control and FSHD, at same time point. N = 3; * p-value < 0.05. **C)** Comparison of DUX4 protein level in myoblasts induced to differentiate in HS- or KOSR-added medium.

3.2 Characterization of D4Z4-derived miRNAs

3.2.1 Analysis of the transcription start site distribution within D4Z4 macrosatellite and DUX4-like genes

Project started by querying FANTOM database (http://fantom.gsc.riken.jp/) for transcription start sites (TSSs), in order to evaluate their distribution patterns in D4Z4 repeat array and other DUX4-like loci in skeletal muscle samples. We used configuration for FANTOM5 human promoterome updated with phase 2 data (Erik et al, 2015) and including tracks for Genecode v19 transcript expression, promoters filtered by Genecode v19 +/- 50bp 5' end, and Entrez gene expression. Data integration, exploration and analysis were performed by using Zenbu, a visualization system improved for RNAseq, CAGE and other next-generation sequencing tag-based data.

All DUX4-like genes annotated in human genome database (hg19, GRCh37) are listed in table 1. We took into account tags included in the intron-less open reading frame of each DUX4 locus plus ~500 bp 5'/3' ends.

Table 1 - DUX4-like genes annotated in the human genome database (hg19)

Gene	Alias	Entrez ID	Chrom.	Genomic coordinates	Strand
DUX4L1	DUX4	22947	4	191005470-191006849	+
DUX4L2	-	728410	4	191012063-191013442	+
DUX4L3	-	653548	4	191008763-191010142	+
DUX4L4	-	441056	4	191002176-191003549	+
DUX4L5	-	653545	4	190998876-191000255	+
DUX4L6	-	653544	4	190995583-190996962	+
DUX4L7	-	653543	4	190992290-190993669	+
DUX4L8	DUX2	26583	4	190988993-190990372	+
DUX4L9	DUX4c	554045	4	190940255-190943820	-
DUX4L10	DUX10	440013	10	135480558-135481931	+
DUX4L11	XX-2136C48.2	399839	10	135483868-135485241	+
DUX4L12	XX-2136C48.3	440014	10	135487177-135488550	+
DUX4L13	XX-2136C48.4	100289581	10	135490476-135491849	+
DUX4L14	XX-2136C48.5	728022	10	135493786-135495159	+
DUX4L15	XX-2136C48.6	440017	10	135497085-135498458	+
DUX4L16	DUXY1	728169	Y	13462499-13463867	+
DUX4L17	DUXY3	643001	Y	13470500-13471872	+
DUX4L18	DUXY2	100132421	Y	13477137-13478507	+
DUX4L19	DUXY4	643034	Y	13487910-13489283	+
DUX4L26	LOC100996350	100996350	3	75718010-75720761	+
DUX4L27	LOC100131369	100131369	12	34361358-34362736	-
DUX4L32	-		20	29318824-29319579	+
DUX4L33	-	107080551	20	29324092-29325049	+
DUX4L34	-	105379479	20	29410284-29411803	-

Datasets we used for TSS pattern analysis have been derived from tag-based sequencing results of CAGE libraries prepared from total RNA of human skeletal muscle samples at different developmental and differentiation stages (Table 2). Platform employed for strand-specific high-precision sequencing of CAGE-defined TSSs was the HeliScope Single Molecule Sequencer.

Table 2 - FANTOM5 human promoterome datasets

DNAse I HSs

Accession	Subject	Cell Type
wgEncodeEH000584_rep1	Healthy	adult primary myoblasts
wgEncodeEH000584_rep2	Healthy	adult primary myoblasts
wgEncodeEH000584_rep3	Healthy	adult primary myoblasts
wgEncodeEH002550_rep1	Healthy	embryonic myoblasts
wgEncodeEH002550_rep2	Healthy	embryonic myoblasts
wgEncodeEH002556_rep1	FSHD	adult primary mioblasts
wgEncodeEH002556_rep2	FSHD	adult primary mioblasts
wgEncodeEH002556_rep3	FSHD	adult primary mioblasts

HeliscopeCAGE

Accession	Cell Type	Sex	Age
CNhs11084	myotubes	N/A	22 week old fetus
CNhs11984	myotubes	Female	22 week old fetus
CNhs12041	myotubes	Female	16 week old fetus
CNhs10870	myoblasts	Female	fetus
CNhs11965	myoblasts	Male	21 years old adult
CNhs11908	myoblasts	Female	59 years old adult

HeliscopeCAGE (time course)

Accession	Cell Type	Sex	Age
CNhs13847	myoblasts day00	Female	3 years old
CNhs14567	myoblasts day00	Female	N/A
CNhs14577	myoblasts day00	Male	N/A
CNhs13848	myoblasts day01	Female	3 years old
CNhs14568	myoblasts day01	Female	N/A
CNhs14578	myoblasts day01	Male	N/A
CNhs13849	myoblasts day02	Female	3 years old
CNhs14570	myoblasts day02	Female	N/A
CNhs14579	myoblasts day02	Male	N/A
CNhs13850	myoblasts day03	Female	3 years old
CNhs14571	myoblasts day03	Female	N/A
CNhs14580	myoblasts day03	Male	N/A
CNhs13851	myoblasts day04	Female	3 years old
CNhs14572	myoblasts day04	Female	N/A
CNhs14581	myoblasts day04	Male	N/A
CNhs13852	myoblasts day06	Female	3 years old
CNhs14573	myoblasts day06	Female	N/A
CNhs14582	myoblasts day06	Male	N/A
CNhs13853	myoblasts day08	Female	3 years old
CNhs14574	myoblasts day08	Female	N/A
CNhs14583	myoblasts day08	Male	N/A
CNhs13854	myoblasts day10	Female	3 years old
CNhs14575	myoblasts day10	Female	N/A
CNhs14584	myoblasts day10	Male	N/A
CNhs14566	myoblasts day12	Female	3 years old
CNhs14576	myoblasts day12	Female	N/A
CNhs14585	myoblasts day12	Male	N/A

In Zenbu genome browser, we collected datasets in several tracks, according to cell type, development stage or experimental features. By signal histogram, each track show genomic position and orientation (green for sense and violet for antisense strand) of the TSSs, and their relative expression level measured as tags per million (tpm).

Altogether, sense and antisense TSS distribution within DUX4 loci localized on different chromosomes appears to be significantly variable. Analogously,

changing of CAGE tag patterns in muscle cells can also occur when compared among different stages of development and differentiation. The TSSs within D4Z4 repeat array, as well as within homologous macrosatellite on 10q26, are equally distributed in sense and antisense orientation along DUX4 locus, without significant differences between adult and fetal muscle cells (Fig. 8A).

Conversely, both DUX4L9 (alias DUX4c) and DUX4L26 (on chromosome 3) show a significant increase of density of the sense/antisense TSS signals at 5' of DUX4 ORF during adult myogenic differentiation when compared to adult myoblasts and fetal muscle cells (Fig. 8B and 9B).

TSSs within DUX4-like loci on chromosome Y (DUX4Y1-4) are particularly enriched during embryonic rather than adult myogenesis (Fig. 9A). Moreover, very few signals of transcriptional activity were found on 12p11.1 (DUX4L27), 20q11.1 (DUX4L32-34) and acrocentrics (data not shown).

Overall, no significant differences at quantitative level were found between sense and antisense tags; indeed, data suggests that transcriptional activity can involve both strands, in particular on chromosomes 3 and Y, potentially giving rise to several transcripts which may be homologous to D4Z4-derived DUX4 mRNA and, analogously, to the three antisense lncRNAs found on 4q35.2.

Of these antisense lncRNAs, in particular, we chose to focus our valuation on transcript overlapping the 5' of DUX4 ORF, that, according to HUGO Gene Nomenclature Committee (HGNC) guidelines, we renamed DUX4-AS2, since it is the second antisense transcript reported along DUX4 gene.

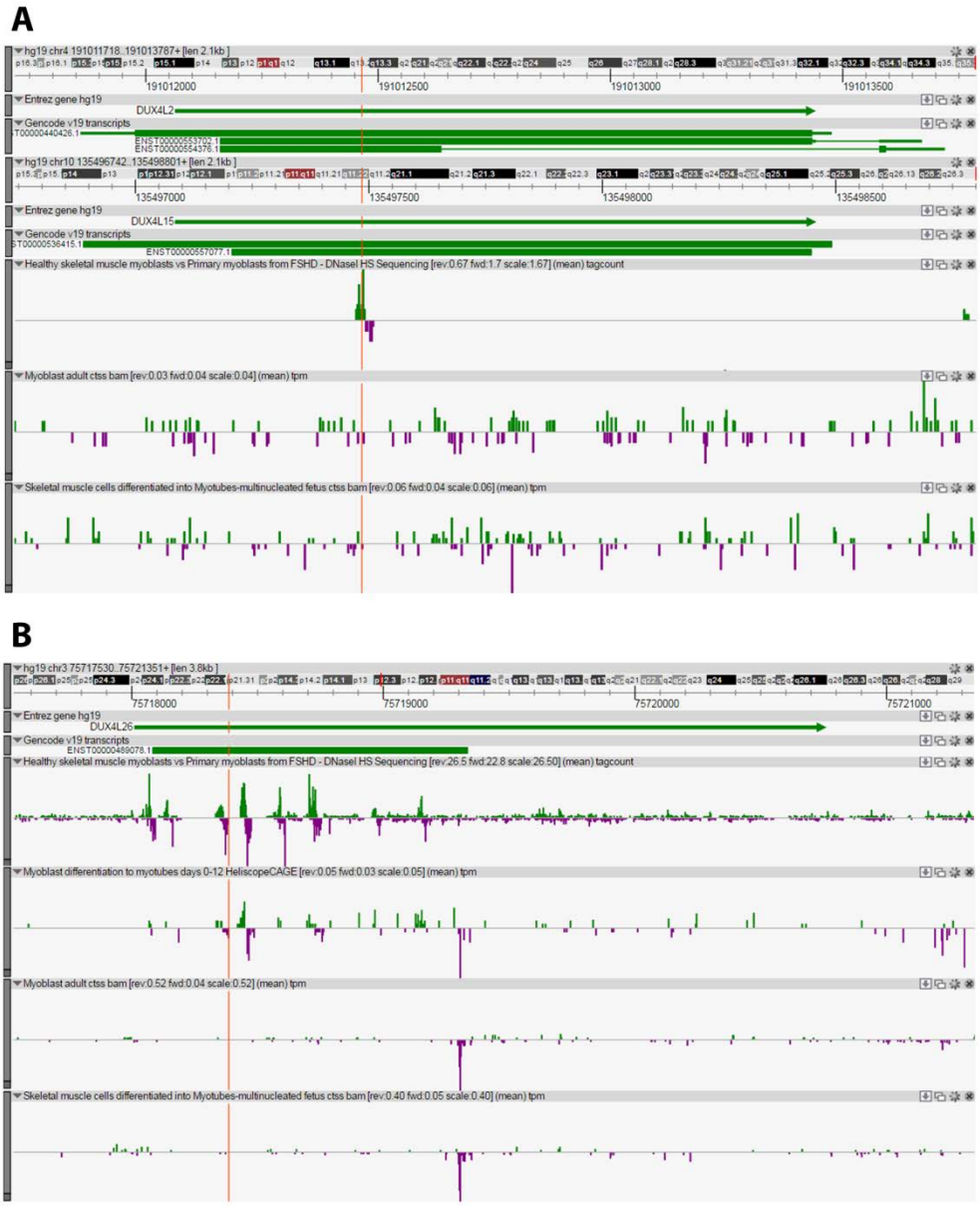


Fig. 8 - A) Graphical view of DUX4L2/DUX4L15 (chromosome 4/10) gene. Tag expression (measured in tpm) is visualized as a signal-height graph along genomic coordinate space, in sense (green) and antisense (violet) orientation. First track reports DNAse I hypersensitive sites in control and FSHD skeletal muscle; second and third tracks show TSSs in adult myoblast and in fetal multinucleated muscle cells, respectively. Vertical orange line highlights the 5' of DUX4-AS2. **B)** Graphical view of DUX4L26 (chromosome 3) gene. First track reports DNAse I hypersensitive sites in control and FSHD skeletal muscle cells; second and third tracks show TSSs in adult myotubes and myoblasts, respectively; fourth track reports those found in fetal multinucleated muscle cells. Vertical orange line highlights the 5' of the putative antisense transcript, homologous to DUX4-AS2.

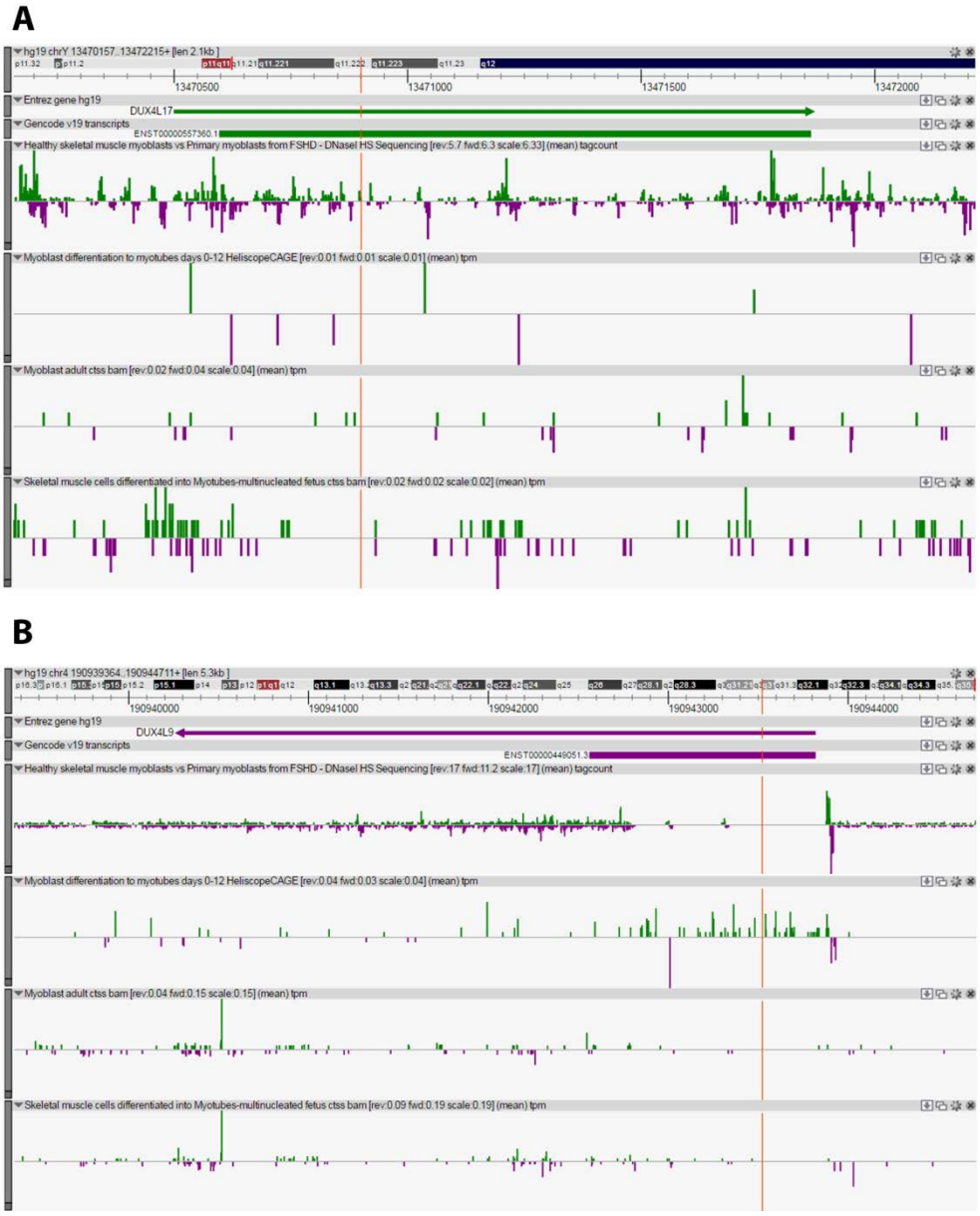


Fig. 9 - Graphical view of **A)** DUX4L9 (or DUX4c - chromosome 4) and **B)** DUX4L17 (chromosome Y) gene. First track reports DNAse I hypersensitive sites in control and FSHD skeletal muscle; second and third tracks show TSSs in adult myotubes and myoblasts, respectively; fourth track reports those found in fetal multinucleated muscle cells. Vertical orange line highlights the 5' of the putative antisense transcripts, homologous to DUX4-AS2.

3.2.2 Evidences of alternative TSSs for DUX4-AS2 (4q35.2) and its putative homologous

Focusing our investigation on DUX4-AS2 transcript, as reported by Snider and colleagues, and analyzing CAGE data on chromosome 4/10 (DUX4L1-8/L10-15), we found three TSSs very close to 5' of DUX4-AS2: one (-3 bp) in fetal and two (-4 and +8) in adult myoblasts (~0.31 tpm), in association with a peak of DNAse I sensitivity reported in both healthy and FSHD samples (Fig 8A).
Moreover, we also tried to map raw data deriving from both skeletal muscle PolyA+ (SRX082584) and PolyA- (SRC084669) RNAseq experiments.
Intriguing, we found the sequence of an antisense transcript identical to DUX4-AS2, but starting 68 bp upstream and perfectly coinciding with another TSS in adult myoblast. In the same datasets (SRX084669), we also come across a sequence of 78 bp in length overlapping one of the proximal miRNA-sized fragments discovered by Snider and colleagues (Fig. 10).

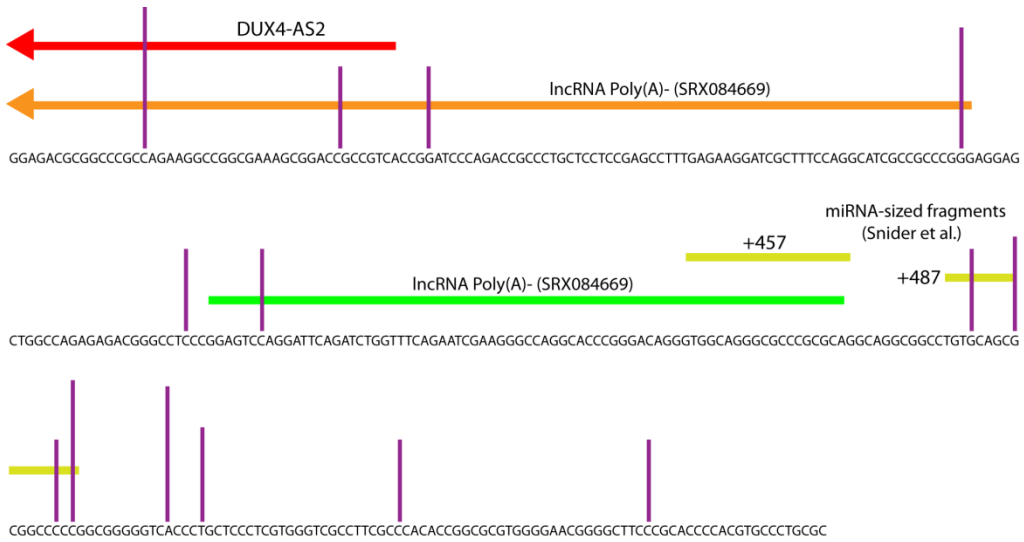


Fig. 10 - Genomic sequence (strand plus) upstream 5' of DUX4-AS2 (red arrow). Mapping of lncRNAs found in dataset SRX084669 are also reported (orange arrow and green line). Yellow lines represents miRNA-sized fragments discovered in this region (numbers refer to nt from MAL codon of DUX4 ORF). Violet vertical lines indicate TSSs observed in adult myoblasts.

These evidences, in addition to the several TSSs located more upstream and downstream (as close as ~31 bp) of the 5' of DUX4-AS2, support hypothesis that this antisense transcript could undergo alternative transcription initiation

and that, furthermore, could be non-polyadenylated, since we have not found matches within the dataset of PolyA+ RNAs.

The putative 5' of homologous transcripts, in antisense orientation to DUX4c and DUX4L26 genes, showed to be closed to several TSSs in adult skeletal muscle cells as well (Fig. 8B and 9B respectively). DNAse hypersensitivity sites corroborated hypothesis of transcriptional activity along region of origin of the putative DUX4-AS2 homologous on DUX4L26, but not on DUX4c locus, where we strangely observed a break in the signal pattern of chromatin accessibility.

No match was found within DUXY repeats, pointing to genomic coordinates corresponding to hypothetical 5' of DUX4-AS2 homologous transcripts. As shown in figure 9A for DUXY3 (DUX4L17), the closest TSS, in adult muscle cells, is over 50 bp downstream. Pattern of chromatin accessibility sites shows a significantly increase approaching 5' of DUXY3 gene, suggestive of putative antisense transcription activity shifted downstream compared to that of other loci. Conversely, on DUX4L27 (chromosome 12), despite of general shortage of transcription signals, the few we found were concentrated around region of homology with the 5' of DUX4-AS2 (data not shown).

Overall, mapping of CAGE tags deriving from repetitive region is affected by low accuracy, hence these evidences remain mere conjectures. We then chose to set up 3'RACE experiments, in order to confirm, or exclude, the existence of this class of homologous antisense transcripts from DUX4-like gene family.

3.2.3 Experiments of 3'RACE reveal that DUX4-AS2, as well as its putative homologous, is non-polyadenylated

The designing of 3'RACE primers was affected by both technical requirements (i.e. accomplishing high stringency conditions in order to reduce bias affecting technique itself) and the need to amplify in the same reaction all hypothetical DUX4-AS2 homologs deriving from different chromosomes. In order to avoid no amplification of some transcripts because of their possibly different transcription initiation sites, we choose to fix the sequence-specific primer annealing position 109 nt downstream the 5' end of DUX4-AS2 characterized by Snider and colleagues, maintaining pair specificity of primer to each putative antisense homologous sequence, as shown in figure 11.

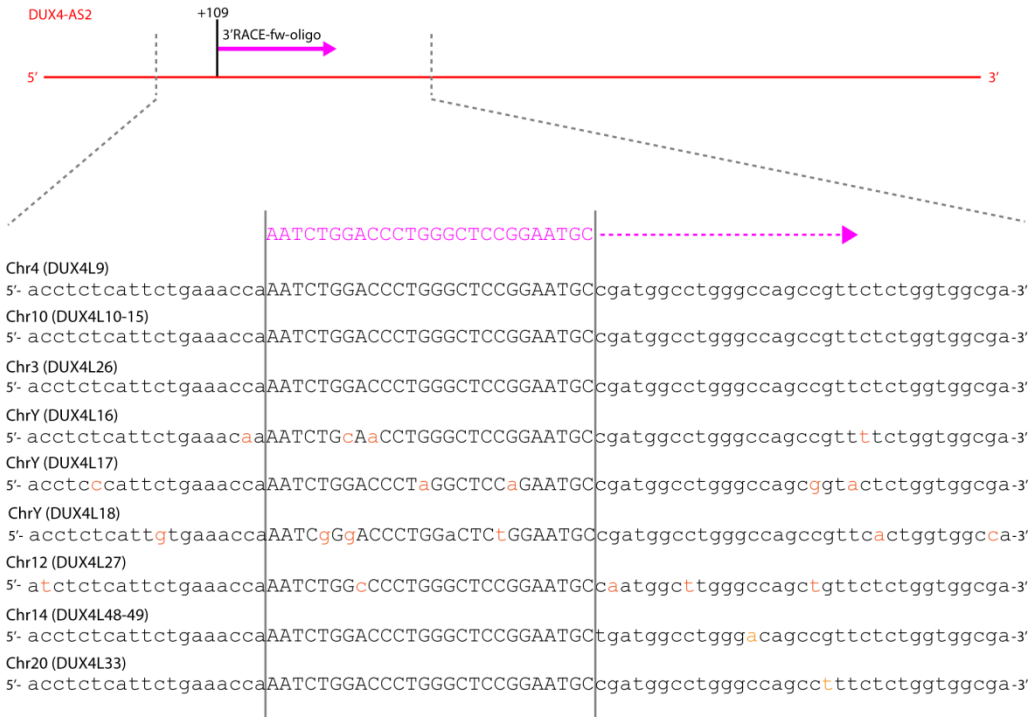


Fig. 11 - 3'RACE forward primer position along DUX4-AS2. Sequence of primer (violet) is aligned to different DUX4-like genes. All mismatch between the sequences are highlighted.

Experiments of 3'RACE was then performed on total RNA extracted from control and FSHD immortalized myoblasts and myotubes (7 days of differentiation). We used both the standard protocol, by using adapter-tailed oligo(dT), and a modified SMART (Switching Mechanism at the 5' end of

RNA Template) approach, by using adapter-linked random hexamers, in order to extend PolyA+ and PolyA+/PolyA- transcripts respectively (see Materials and Methods for details).

The 3'RACE products were evaluated by agarose gel electrophoresis (Fig. 12). For all RNA samples, when SMART approach was used, we observed a less intense smear and a single specific amplicon of about 400-450 bp, whereas the standard protocol led to a variable amplification pattern and most faint bands.

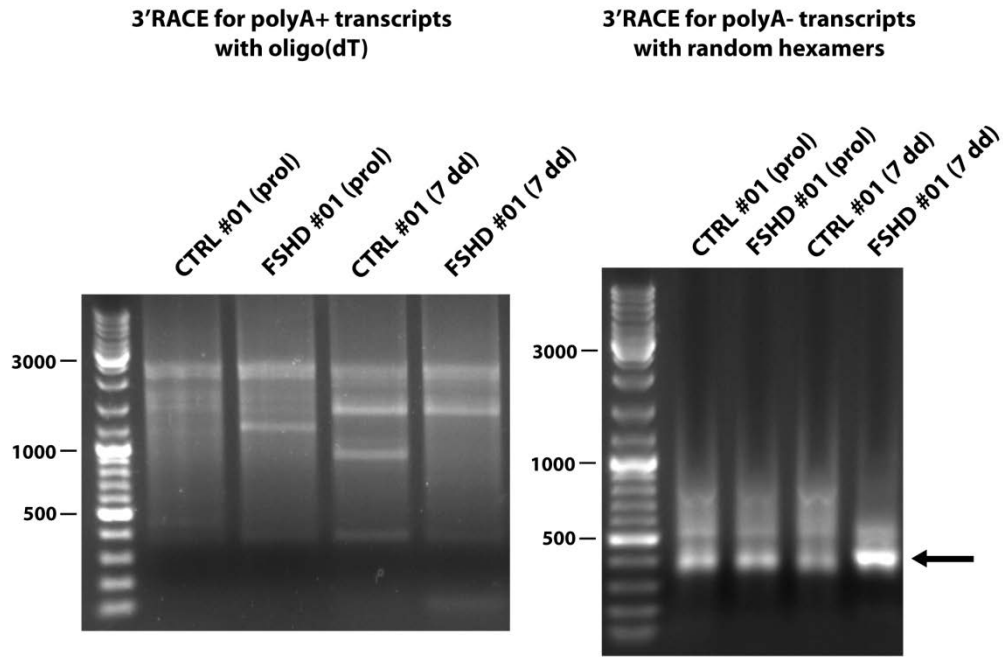


Fig. 12 - Agarose gel electrophoresis of 3'RACE products, using oligo(dT) (left) or random hexamers (right) on total RNA extracted from cohort 01 control and FSHD myoblasts (prol) and myotubes after seven days of differentiation (7 dd). Black arrow indicates specific amplification obtained by polyA- 3'RACE protocol.

By Illumina MiSeq platform, paired-end sequencing of amplification products deriving from both 3' end extension approaches confirmed our hypothesis that DUX4-AS2 was non-polyadenylated. Indeed, preliminary alignment of total reads, performed by MiSeq Reporter analysis software to human genome database (h19) showed specific mapping to DUX4-like loci obtained after using of adapter-linked random hexamers only, as reported in table 3.

Table 3 - Preliminary alignment by MiSeq Reporter analysys software

Poly(A+) 3'RACE[a]		
#	Gene	Reads[b]
1	SMARCA4	505
2	LAMB1	460
3	CD59	293
4	CDK18	191
5	UQCRH	83
6	NHP2L1	75
7	H2AFJ	69
8	LRIG2	27
9	RPS27A	27
10	FAHD2A	25
11	NDUFA13	22
12	PGLS	19
13	STAT2	19
14	CLCC1	18
15	FLYWCH2	17
16	RPLP1	17
17	SERINC5	14
18	PTGDS	11
19	MGRN1	9
20	CLN3	9
21	UQCRHL	8
22	TMEM126A	7
23	TMEM49	7
24	RPS8	6
25	POFUT1	6
..	..	..
..	..	..
79	ABL1	1
80	PLEC	1
81	C8orf33	1
82	YKT6	1
83	PKNOX1	1

Poly(A-) 3'RACE[a]		
#	Gene	Reads[b]
1	DUX4L26	9111
2	DUX4L9	1308
3	DUX4L7	263
4	DUX4L4	252
5	DUX4L17	110
6	DUX4L16	50
7	DUX4L19	19
8	HIST2H2AC	3
9	DUX4L27	5
10	ND4	3
11	ID1	2
12	CCT6A	2
13	AHNAK	2
14	DUX4L18	1
15	HNRNPK	1
16	NEAT1	1
17	CYBRD1	1
18	TRIM25	1
19	SVIL	1

[a] FSHD biceps 7dd (#01)
[b] Unique mapped reads

3.2.4 Analysis of sequencing suggests the presence of at least three different antisense transcripts from DUX4 gene family

As previously explained, the goal of 3'RACE we set up was to amplify several transcripts in the same reaction. For this reason, 3'RACE experiments was designed as multi-template PCR, where a unique specific primer potentially led to enrichment of an heterogeneous class of sequences. Since DUX4-like loci have a high degree of homology, deep sequencing data analysis was performed to ensure assignment of correct chromosomal derivation of most of the sequences, minimizing effects of errors deriving from amplification and sequencing reactions on mapping accuracy.

Initially, we needed to compute the Levenshtein distance between D4Z4 repeat and other DUX4-like loci (reverse strand) reported in human genome database (hg19, GRCh37), from 5' (corresponding to 3'RACE primer position) to 3' end of putative antisense transcripts we are looking for (~445 bp).

Levenshtein distance between two string, also named *edit distance*, is closely related to pairwise string alignment and represent the minimum number of single-character edit (i.e. insertion, deletion or substitution) required to change one string into the other. Mathematically, Levenshtein distance between two sequence *a* and *b* (of length |*a*| and |*b*| respectively) is given by $\mathrm{lev}_{a,b}(|a|,|b|)$, defined as:

$$\mathrm{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \max(i,j) = 0 \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1 \\ \mathrm{lev}_{a,b}(i,j-1)+1 \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise} \end{cases}$$

where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$ and 1 otherwise.

As shown in figure 13, antisense strand of D4Z4 repeats (4q35.2) gradually diverges from its homologous sequences on other DUX4-like loci as they approach to 3' end and moving away from the open reading frame. DUX4-like loci mapping to 10q26 region, which share the same cluster organization with D4Z4 macrosatellite, represent the only exception to this trend.

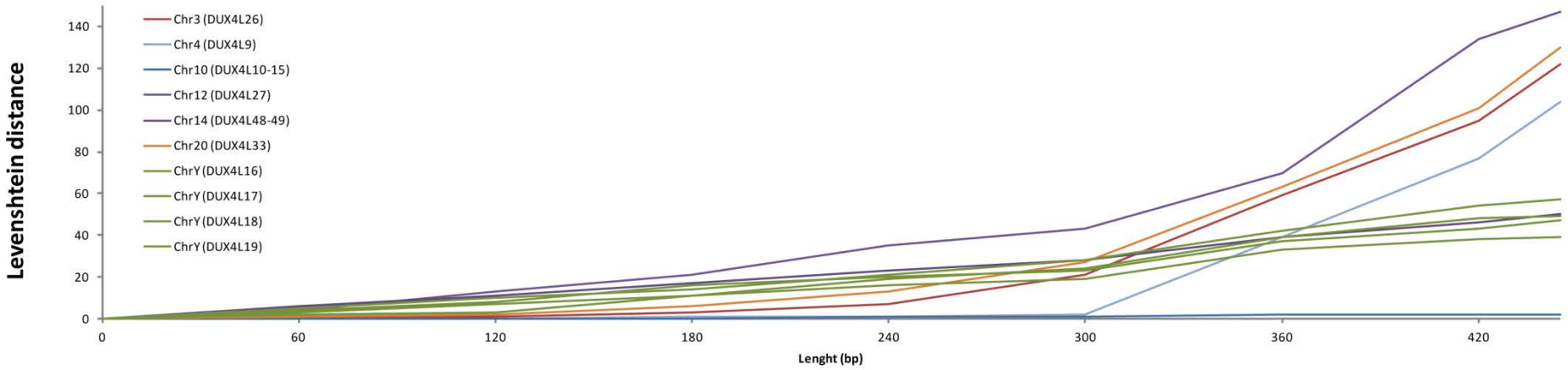


Fig. 13 - Cumulative Levenshtein distance between reverse complement of the genomic sequences of different DUX4-like loci and the D4Z4-deriving sequence ranging from 3'RACE primer annealing position to 3' end of antisense transcript.

Since sequences on 4q35.2 and 10q26 are very close ($lev_{a,b}(i,j) = 2$), we have not been able, during sequencing analysis, to discriminate multi-mapping reads deriving from these two regions, therefore we will refer to them as 4q/10q sequences, without further specification.

Slightly divergent sequences are represented also by those of DUXY1-4 loci, with edit distance of 49, 57, 47 and 39 respectively, and that mapping on short harm of chromosomes 14 ($lev_{a,b}(i,j) = 50$). Unlike other DUX4-like sequences, their distribution of mismatches is particularly uniform and is characterized by several indel polymorphisms. Moreover, pairwise alignment of repeats on chromosome Y was further investigated.

All four DUXY loci, rather than repeated units on 4q35.2 and 10q26, have suffered microduplications and/or microdeletions events which decreased their degree of identity ($\leq 90\%$), as it has been previously reported (Schmidt et al, 2009).

On the other hand, DUX4L26 (chromosome 3), DUX4L27 (chromosome 20) and DUX4L9 (DUX4c), that heavily diverge from D4Z4 repeated sequences ($lev_{a,b}(i,j) > 100$), share a high level of homology ($\geq 95\%$), necessitating a closer study of distribution of pairwise edit distances.

The main problem of univocal alignment of multi-template PCR products derive from both DNA polymerase mutation rate and template-switching occurring during each replication cycle to the extent that they change the Levenshtein distance between two different starting sequences.

Also PCR-based library preparation and sequencing itself introduce variations, therefore all reads take with a given number of artifactual mutations according to a specific probability distribution. A single mutational event affects Levenshtein distance, increasing or decreasing it.

Assume we have two sequences N and M with n and m nucleotides, where $n \neq m$; their distance is equal to $d + \sum g_k$, where d represent mismatches between sequences and $\sum g_k$ the total number of gaps of length k.

If a mutation happens (with mutation rate μ) in any of d nucleotides of sequence N, changing it to the nucleotide which is on the same position of the sequence M, or analogously, if a indel error (with frequency λ) reduces k length of gap between the sequences (without generate a mismatch), then the Levenshtein distance decreases.

The probability to have such distance reduction is given by:

$$P_{dec} = \frac{1}{3}\frac{d}{n}\mu(1-\mu)^{n-1} + \frac{1}{4}\frac{(m-n)}{n}\lambda(1-\lambda)^{n-1} + \frac{(m-n)}{m}\lambda(1-\lambda)^{m-1}$$

However, substitutions can occur in any of other similar nucleotides or one of *d* different nucleotides can mutate to either of two other nucleotides which do not exist in either of sequences; analogously, indel events can extend *k* length of existing gaps, create a new one or insert in shorter sequences one of three other nucleotides which are not present in the longest, then the Levenshtein distance increases with probability defined as:

$$P_{inc} = \mu(1-\mu)^{n-1}\left(\frac{n-d}{n} + \frac{2}{3}\frac{d}{n}\right) +$$

$$+ \lambda(1-\lambda)^{n-1}\left(\frac{3}{4}\frac{(m-n)}{n} + \frac{n-(m-n)}{n} + 1\right) +$$

$$+ \frac{n}{m}\lambda(1-\lambda)^{m-1}$$

From these equations we can infer that probability of decrease is smaller than increase distance, therefore the distribution of probability is not symmetric around the initial Levenshtein distance.

The higher is this difference, the relatively closer are starting sequences, provided that the mutation rate is sufficiently low and assuming it occurs randomly along the sequences.

If the Levenshtein distance between two starting sequences exceeds the upper limit of the finite interval $I = [x, x + dx]$ where the probability density function of the number of mutations assumes non-zero values, such that $P\left(lev_{a,b}(|a|,|b|) \in I\right) = \int_x^{x+dx} p\left(lev_{a,b}(|a|,|b|)\right) dx = 0$, then we could use the Bayesian maximum a posteriori estimate that maximizes, given the data, the probability that changes in Levenshtein distance occur because of errors during PCR and/or sequencing base calling, consequently re-integrating multi-mapping reads into sequencing analysis.

Assume series of random variables $X = [x_1,\ldots x_R]$ of *R* elements, with a $p(X; \theta)$ distribution, and $\theta = [\theta_1,\ldots\ \theta_R]$ as the vector of parameters from which distribution depends. Let $x_1,\ldots x_n$ be *N* observations extracted from such distribution.

Posterior probability is given by:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

where *P(θ)* is the probability *a priori* of θ parameter and *P(X| θ)* is the likelihood function, which is the probability of the observations given the parameters. If we considers $x_1,\ldots x_n$ observations the edit distance of reads from reference sequences and θ the mutation rate that generate this distance, we can maximize estimation of parameter θ, omitting *P(X)* that do not depend on it:

$$\theta_{MAP} = \arg\max_{\theta} P(X|\theta)P(\theta)$$

Since we have not enough information to approximate *P(θ)*, we can assume it as constant, so that maximum a posteriori probability and likelihood coincide, allowing us to resolve estimation of parameter by finding the value that maximize the likelihood.

$$\theta_{ML} = \arg\max_{\theta} P(X|\theta)$$

This expression represents the maximum likelihood, whereas binomial distribution was the probabilistic model we approached to calculate the probability that multi-mapping reads deriving from a specific DUX4 locus rather than another. Maximum likelihood estimate with Bernoulli model (binomial) is given by:

$$L(X|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

$$lnL(X|\theta) = \ln\left[\binom{n}{x}\right] + xln[\theta] + (n-x)\ln[1 - \theta]$$

We use logarithmic formula to build log-likelihood curve along the values of θ between 0 and 1. The point of maximum of log-likelihood curve coincide with the value of θ that maximizes the likelihood of making the observations given the parameters.

Since the range of values of θ around its point of maximum follow a X^2 distribution with a confidence level of $1 - \alpha$, mapping of reads can be resolved by performing log-likelihood ratio-based hypothesis test, where the probability

that a given read derive from two different loci is p = 0,5 represents the null hypothesis.

$$G = 2 \ln \left[\frac{L(Y_{max}|X)}{L(Y_{0,5}|X)} \right]$$

If log-likelihood ratio $G > X^2_{1,0.05}$ then we refused null hypothesis and assume that a given read originated from specific DUX4 locus with a p-value ≥ 0,05 despite of errors introduced by PCR and/or sequencing base calling.

By preliminary alignment report, although < 5% was not specific amplification products, approximately 75% of DUX4 antisense sequences resulted multi-mapping reads. Re-alignment was then performed by using BWA-MEM algorithm and multi-mapping sequences were evaluated by log-likelihood ratio test to exclude suboptimal alignments. By using above-explained statistical method, over 55% were assigned to specific DUX4 locus. Alignment results for each sample were indexed, merged into a unique bam files e visualized by IGV (Integrative Genome Viewer) software.

Through MUSCLE algorithm, univocally mapped overlapping sequence data allowed us to easily generate consensus of three antisense transcripts deriving from D4Z4 (and/or 10q26 homologous cluster), DUX4L17 (DUXY3 on chromosome Y) and DUX4L26 (on chromosome 3), as shown in figure 14.

We derived also a partial contig from multiple alignment of reads mapping to DUX4L9 (DUX4c), but we have not further investigated the existence of antisense transcripts from this locus. Moreover, comparison between data obtained from the different samples did not show differences at qualitative level, therefore both control and FSHD skeletal muscle cells would seem to express same transcripts. Interestingly, we found that DUX4-AS2 overcomes of ~75 nt the antisense transcript reported by Snider and colleagues.

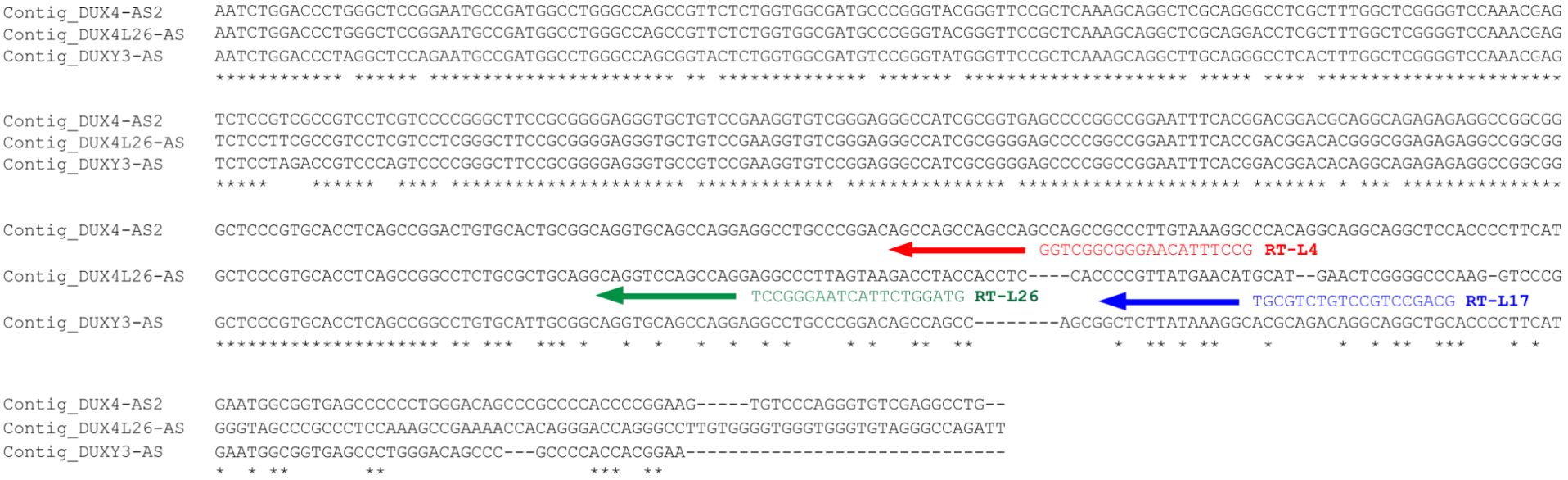


Fig. 14 - Alignment of consensus sequences of the antisense transcripts deriving from chromosomes 4 (DUX4-AS2), 3 (DUX4L26-AS) and Y (DUXY3-AS). Strand-specific RT primers are reported with different colors at their annealing position. Asterisks represent common bases among three contigs.

3.2.5 Strand-specific RT-PCR corroborates existence of three antisense transcripts from chromosome 3, 4 and Y

In order to validate consensus sequences and amplify one by one the antisense transcripts deriving from chromosome 3, 4 and Y, we designed three specific primers (figure XX) and performed three separated one-step strand-specific RT-PCR (ssRT-PCR) experiments, testing oligo specificity on total RNA extracted from Chinese Hamster Ovary (CHO) cell hybrids, each containing fragments of one of the above human chromosomes.

A brief summary of human/CHO hybrids used in this study, including their evaluation by in situ hybridization, is reported in Material and Methods. One-step ssRT-PCR experiments were performed after treatment with AZA/TSA, and followed by nested PCR, as described in detail (see Material and Methods). Total RNA of untreated hybrids was used as negative control.

As displayed on agarose gel (Fig. 15A), DUX4-AS2 amplification product in CHO/Chr4 (RH4L319) cell line was obtained after RT with L4 oligo only; similarly L17 and L26 primers guaranteed specific amplification in CHO/Chr3 (HYE7) and CHO/ChrY (HY94BT1) cell hybrids respectively.

After testing ssRT primers, we performed the same reactions on human FSHD myotube (cohort 01, strain 157A) total RNA, using a neuroblastoma cell line (SH-SY5Y) as negative control (Fig. 15B).

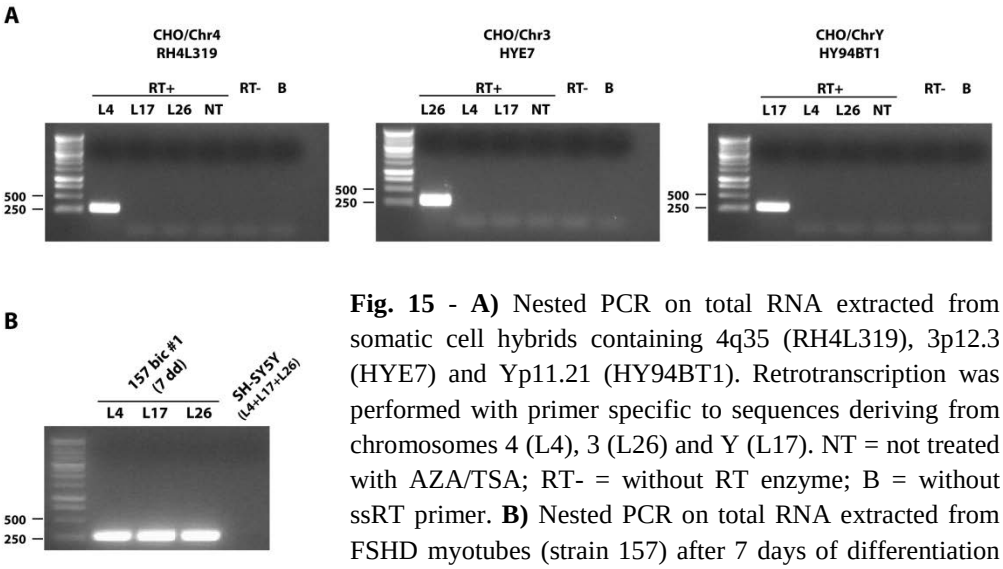


Fig. 15 - A) Nested PCR on total RNA extracted from somatic cell hybrids containing 4q35 (RH4L319), 3p12.3 (HYE7) and Yp11.21 (HY94BT1). Retrotranscription was performed with primer specific to sequences deriving from chromosomes 4 (L4), 3 (L26) and Y (L17). NT = not treated with AZA/TSA; RT- = without RT enzyme; B = without ssRT primer. **B)** Nested PCR on total RNA extracted from FSHD myotubes (strain 157) after 7 days of differentiation (7 dd), retrotranscribed with L4, L17 and L26 ssRT primers. SH-SY5Y was used as negative control.

All nested PCR products from human FSHD myotubes were sub-cloned in TA cloning vector and then sequenced. As reported in following table, sequencing confirms specificity of retrotranscription reaction, validating transcripts found by analysis of 3'RACE products.

Table 4 - BLAT alignment of nested PCR amplicons on human genome (hg19)

SS-RT oligo	Mapping	Identity	Span	Score
RT-L4	4q35.2 (D4Z4)	99,7 %	283	281
	10q26.1	99,0 %	283	277
	4q35.2 (DUX4L9)	99,3 %	263	259
	Yq11.21	93,0 %	283	243
	3p12.3	95,1 %	264	236
	12p11.1	82,2 %	238	159
RT-L17	Yq11.21	99,3 %	281	277
	4q35.2 (D4Z4)	92,4 %	283	233
	10q26.1	92,4 %	283	233
	4q35.2 (DUX4L9)	92,2 %	262	214
	3p12.3	91,2 %	263	212
	12p11.1	87,5 %	218	107
RT-L26	3p12.3	99,3 %	279	275
	4q35.2 (DUX4L9)	93,9 %	278	249
	20q11.1	94,2 %	276	241
	4q35.2 (D4Z4)	94,0 %	263	236
	10q26.1	93,6 %	263	234
	Yq11.21	90,0 %	258	211

According to guidelines defined by HGNC, antisense transcripts deriving from chromosomes 3 and Y, since they are the first to be identified in these loci, unlike transcript from 4q35.2, will be named DUX4L26-AS and DUXY3-AS, respectively.

3.2.6 Antisense transcript expression profiling during muscle differentiation highlights putative role of DUX4-AS2 in FSHD

At first we focused our investigation on expression level of these antisense RNAs in order to identify differences between FSHD and related healthy control. We then compared affected (cohort 01, cell strain 157A) and control (cohort 01, cell strain 160U) bicep-derived immortalized muscle cell lines before to trigger (proliferating myoblasts) and at the end of differentiation process (myotubes differentiated to 7 days). However, no significant difference was found either among affected and healthy samples or between myoblasts and myotubes (not shown).

Basing on the assumption that FSHD phenotype may be characterized by alterations of early stages of skeletal myogenesis, as previously observed, we wondered whether changes in expression of these antisense RNAs may occur during myotube formation.

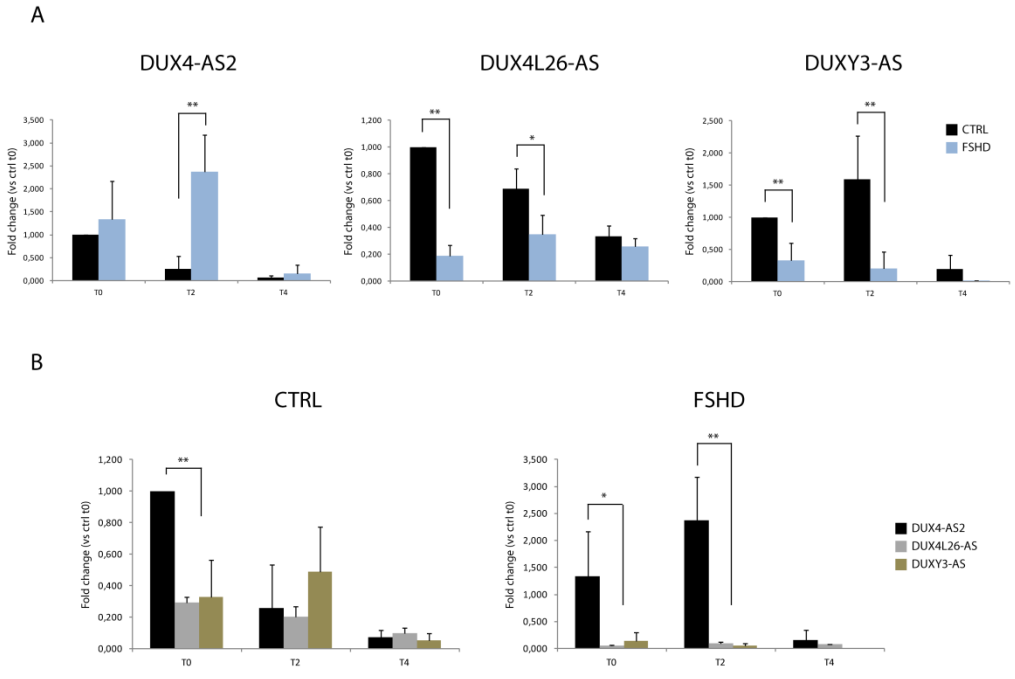


Fig. 16 - Analysis of DUX4-AS2, DUX4L27-AS and DUXY3-AS expression levels by qRT-PCR, during early stages of differentiation, in biceps control and FSHD muscle cells. Expression values are normalized against expression level of control at time 0. Error bars indicate standard deviation. N = 3; * p-value < 0.05; ** p-value < 0.01.

Therefore, we chose to set up a time course, quantifying antisense RNAs at earlier stages of differentiation (from 0 to 4 days in KOSR-added medium) and extending the analysis to immortalized cell lines deriving from deltoids (cohort 01, cell strains 172A and 175U), which, by clinical evaluation, was reported to be less affected than biceps.

The extensive comparison of qRT-PCR results is shown in figures 16 and 17. In confirmation of our hypothesis, in FSHD biceps, DUX4-AS2 increases at early stages (0-2 days) of muscle differentiation, compared to controls (Fig. 16A), showing to be much more expressed than its homologs too (Fig. 16B).

On the other hand, in control samples, during early differentiation steps, we noticed that DUXY3-AS and DUX4L26-AS are generally higher than in FSHD (Fig. 16A). Interestingly, no difference between the three antisense RNAs was found in healthy biceps (except at T0) (Fig. 16B) and in affected deltoids (not shown), whereas in control deltoids this balance is biased towards DUXY3-AS (data not shown).

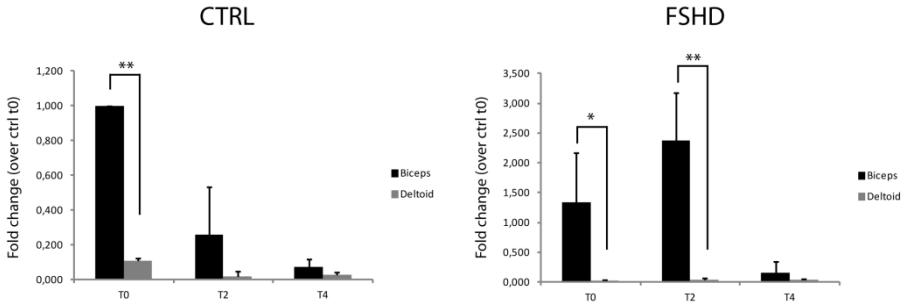


Fig. 17 - Comparison of DUX4-AS2 expression levels by qRT-PCR between biceps and deltoid muscle cells, during early stages of control and FSHD differentiation. Expression values are normalized against expression level of control at time 0. Error bars indicate standard deviation. N = 3; * p-value < 0.05; ** p-value < 0.01.

In further agreement whit our suppositions, preliminary evidence for putative involvement of DUX4-AS2 in early stages of muscle differentiation can be found by comparison of its expression level between biceps and deltoids, as shown in figure 17. Indeed, since biceps is part of most involved musculature in affected patients, in the first days of muscle differentiation of the bicep-derived FSHD myoblasts, level of DUX4-AS2 significantly increases of more than 2 orders of magnitude when compared to deltoid.

3.2.7 DUX4-AS2 is localized into the nucleus, where interacts with DROSHA/DGCR8 complex

In light of results obtained from evaluation of expression profile during early differentiation, functional studies have been focused on DUX4-AS2. At state of art, we have strong evidence of stable transcription and translation from DUX4 gene inside D4Z4 macrosatellite (as well as DUX4c), but no suggestion about activity of widespread DUX4-like genes on other chromosomes. Hypothesizing that DUX4 gene may be affected by its antisense transcript, we assayed effects of DUX4-AS2 overexpression on DUX4 gene regulation.

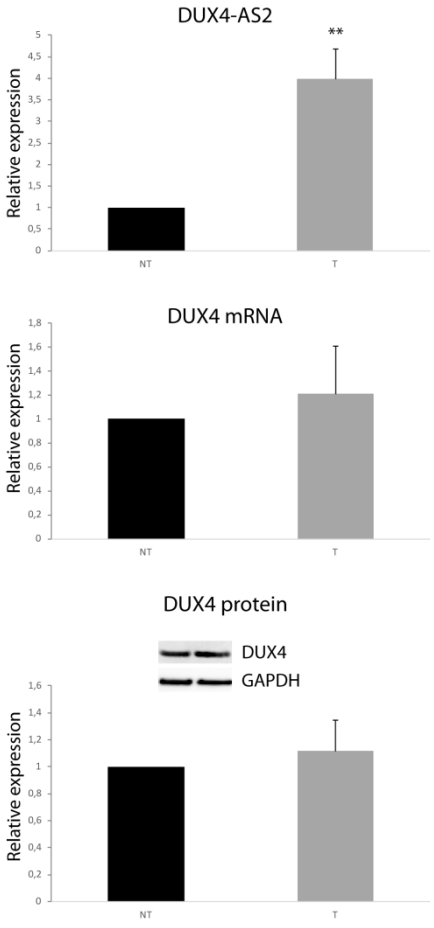


Fig. 18 - Evaluation of DUX4 mRNA and protein levels after transfection (T) of DUX4-AS2 in FSHD myoblasts. The increase of DUX4-AS2 concentration after transfection was assayed by qRT-PCR as well. Values correspond to relative expression against non-transfected cells (NT). GAPDH was used as normalizer. Error bars indicate standard deviation. N = 3; ** p-value < 0.01.

We chosen to transfect DUX4-AS2 RNA as complete as possible, on the basis of the current characterization, then starting in vitro transcription from 5' as reported by Snider and colleagues, to 3' end found by 3'RACE, making a transcript of 554 nt in length. Transfection was performed on FSHD myoblasts and after 48 hours cells were collected to evaluate DUX4 mRNA and protein levels. No differences were found between transfected and non-transfected cells (Fig. 18).

Several mi/siRNA-sized fragments have been reported within D4Z4 (Snider et al, 2009; Benhamed et al, 2012; Ameyar-Zazoua et al, 2012) and DUX4-AS2 overlaps some of these short RNAs. Therefore, as alternative hypothesis, we inferred that this antisense transcript could be involved in the RNAi machinery, in particular as primary miRNA precursor.

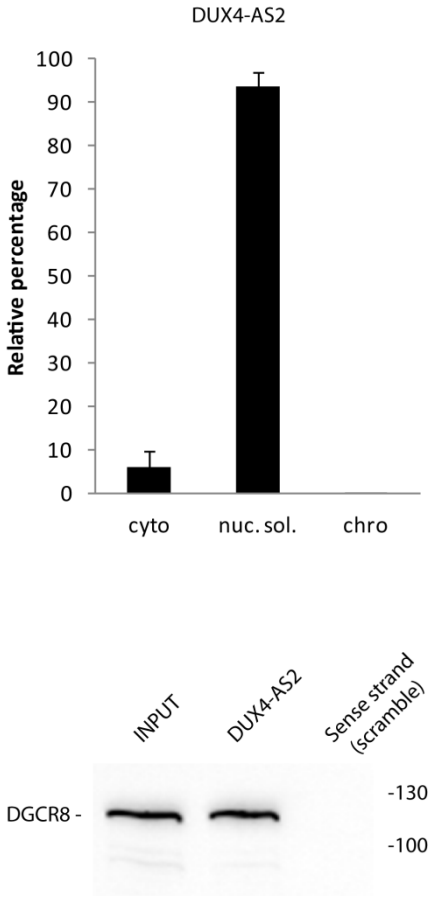


Fig. 19 - **A)** Subcellular localization of DUX4-AS2 in control and FSHD muscle cells. Cyto = cytoplasm; nuc.sol. = nucleoplasm; chro = chromatin-associated. Error bars indicate standard deviation. N = 5. **B)** Results of interaction assay between DGCR8 and biotinylated DUX4-AS2 RNA. Its complementary RNA (sense strand) was used as scramble control.

To corroborate this supposition, we evaluated both subcellular localization of DUX4-AS2 in muscle cells and DGCR8 binding affinity by RNA-protein interaction assays. Interestingly, DUX4-AS2 is recognized by DGCR8 in vitro and appears to be significantly localized into the nucleus, but not associated to chromatin (Fig. 19).

Taking advantage of these evidences, we deepened the role of DUX4-AS2 as miRNA precursor by computational methods, defining its secondary structures and thermodynamic properties, in order to predict all miRNAs that it might generate.

3.2.8 Secondary structure and predictive analysis report several miRNAs originating from DUX4-AS2

The pipeline of method we implemented to predict miRNAs is based on use of previously reported and well-documented PriMir and Miridentalgorithms (He et al, 2008; Liu et al, 2012).

DUX4-AS2 sequence (554 nt) was processed to generate ~2800 overlapping subsequences ranging in length from 60 to 110 nt, thus according to length distribution of miRNA precursors (Krol et al, 2004; Fang et al, 2013). Each subsequence was folded by RNAfold (Vienna Package), obtaining information on intra-molecular interactions and minimum free energy (MFE). In order to predict pre-miRNA stem-loop structures, folded subsequences were screened by score matrix based on eleven features: base pairing of flanking bases, bulge length of pre-miRNA, base pairing of pre-miRNA, length of terminal loop, distance from terminal loop, initial bias of miRNA, base pairing of miRNA, MFE of pre-miRNA, length of pre-miRNA, GC content of pre-miRNA and GC content of miRNA.

All miRNA-encoding candidates with a positive total score (PriMir score ≥ 0) were analyzed by using support vector machine (SVM)-based linear classifier (Miridentl), which integrates sequence-structure motif information to accurately distinguish pre-miRNA from other stem-loop structures. To perform a correct classification, the algorithm was initially instructed with two training sets containing 484 human pre-miRNAs and 484 human non-pre-miRNA stem-loop structures, respectively.

We finally predicted a total of 7 pre-miRNA candidates and, consequently, 16 putative miRNAs originating from DUX4-AS2. Structural configuration and thermodynamic details of pre-miRNAs are reported in figure 20, followed by the list of all putative miRNAs. For one pre-miRNA (as-pre-miR-256/362), in particular, sequence bias of predicted cleavage by both DROSHA and DICER admits generation of two couple of iso-miRs: as-miR-276-5p/as-miR-279-5p and as-miR-316-3p/as-miR-320-3p.

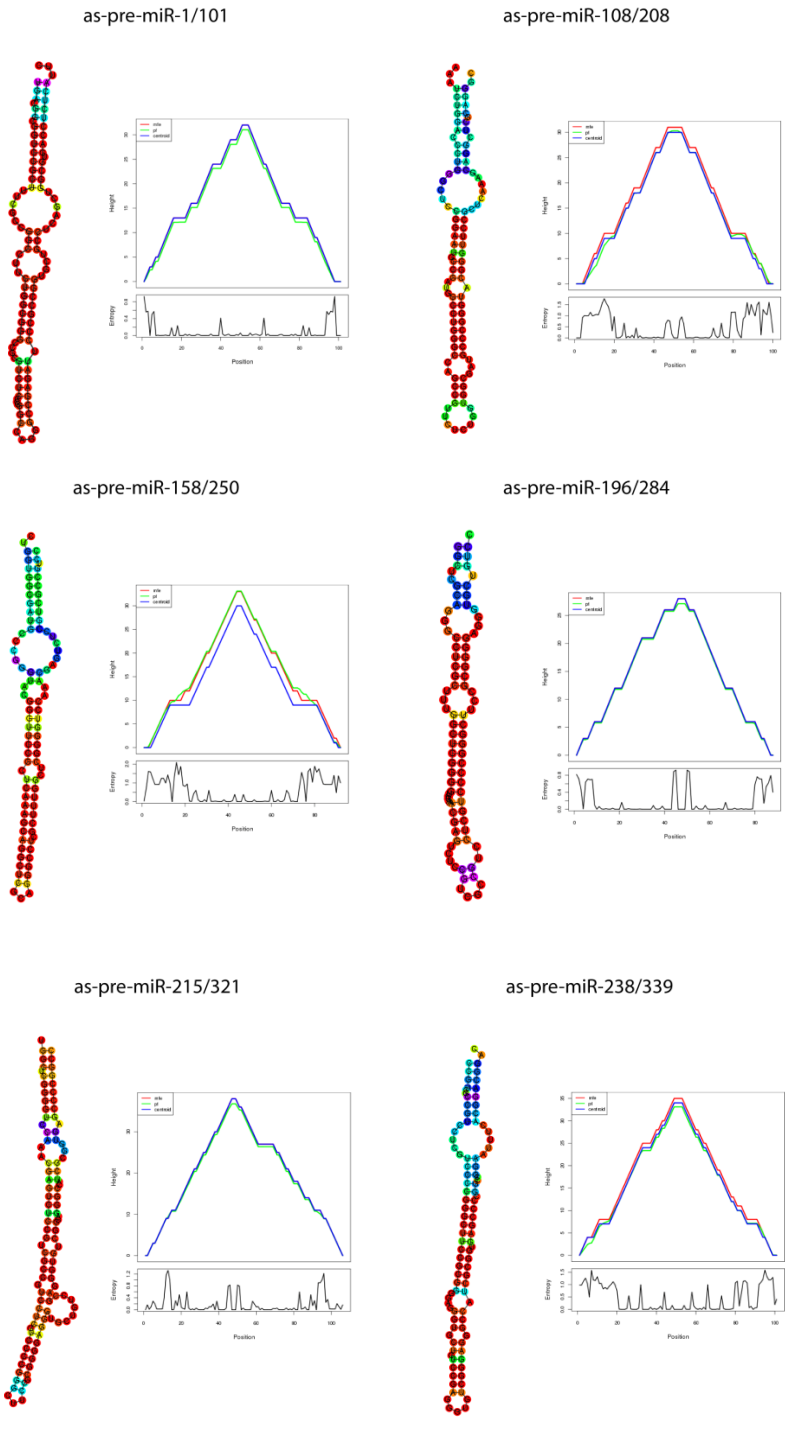


Fig. 20 - Secondary structure of pre-miRNAs, flanked by mountain plot and entropy chart. Bases of structures were colored from blue to red according to increasing level of stability.

Table 5 - Predicted miRNAs from DUX4-AS2

Pre-miRNA	miRNA 5p	miRNA 3p
as-pre-miR-1/101	as-miR-11 CGCUUUCGCCGGCCUUCUGGCG	as-miR-68 CGGUGCUGCCUCAGCUGGCGUG
as-pre-miR-108/208	as-miR-128 GGAAUGCCGAUGGCCUGGGCCA	as-miR-166 GAUGCCCGGGUACGGGUUCCGC
as-pre-miR-158/250	as-miR-178 CGGGUUCCGCUCAAAGCAGGCU	as-miR-208 CUCGCUUUGGCUCGGGGUCCAA
as-pre-miR-196/284	as-miR-216 GGCUCGGGGUCCAAACGAGUCU	as-miR-242 CGCCGUCCUCGUCCCCGGGCUU
as-pre-miR-215/321	as-miR-235 UCUCCGUCGCCGUCCUCGUCCC	as-miR-279 UGUCCGAGGGUGUCGGGAGGGC
as-pre-miR-238/339	as-miR-258 GGGCUUCCGCGGGGAGGGUGCU	as-miR-297 GGGCCAUCGCGGUGAGCCCCGG
as-pre-miR-256/362	as-miR-276 UGCUGUCCGAGGGUGUCGGGAG as-miR-279 UGUCCGAGGGUGUCGGGAGGGC	as-miR-320 CGGAAUUUCACGGACGGACGCG as-miR-316 CGGCCGGAAUUUCACGGACGGA

3.2.9 Six D4Z4-derived miRNAs are expressed in FSHD and/or healthy control muscle cells

Validation of predicted miRNAs was performed by using LNA oligonucleotide made by Exiqon Synthesis Service and its proprietary primer design software. Experiments of qRT-PCR were set up with miRCURY LNA system (Exiqon) on small RNA-enriched fractions, extracted from biceps control and FSHD myoblasts and myotubes (4-6 days of differentiation).

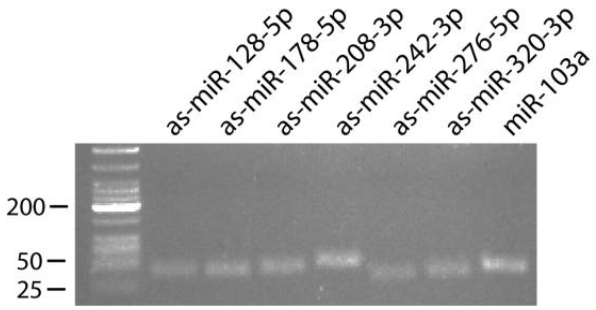


Fig. 21 - Agarose gel electrophoresis of D4Z4-derived miRNAs detected by mercury LNA system in control cells (4 days of differentiation). The housekeeping miR-103a was used as positive control. Amplicon lengths were in accord to those predicted by Exiqon.

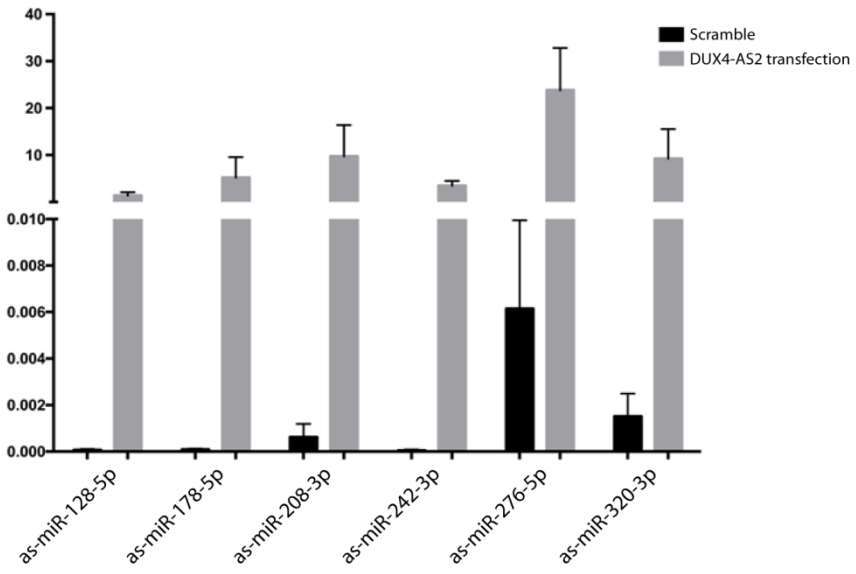


Fig. 22 - Evaluation of miRNA expression levels after transfection of DUX4-AS2 in healthy myoblasts versus control scramble. The miR-103a was used as normalizer. Error bars indicate standard deviation. N = 3.

Amplification products were visualized on agarose gel (Fig. 21) after their evaluation by melting curve analysis. We correctly amplified 6 of 16 predicted miRNAs in at least one of used samples. Moreover, to confirm miRNA derivation from DUX4-AS2, effects of its transfection were assayed on control myoblasts. As we expected, all detected miRNAs significantly increase their amount after DUX4-AS2 transfection, versus scramble control transfected cells (Fig. 22).

3.2.10 Validated miRNAs are differentially expressed during FSHD myogenesis, when compared to control

The expression of 6 validated miRNAs was analyzed as well as their primary precursor, in biceps control and FSHD during different stages of differentiation (from 0 to 4 days in KOSR-added medium).

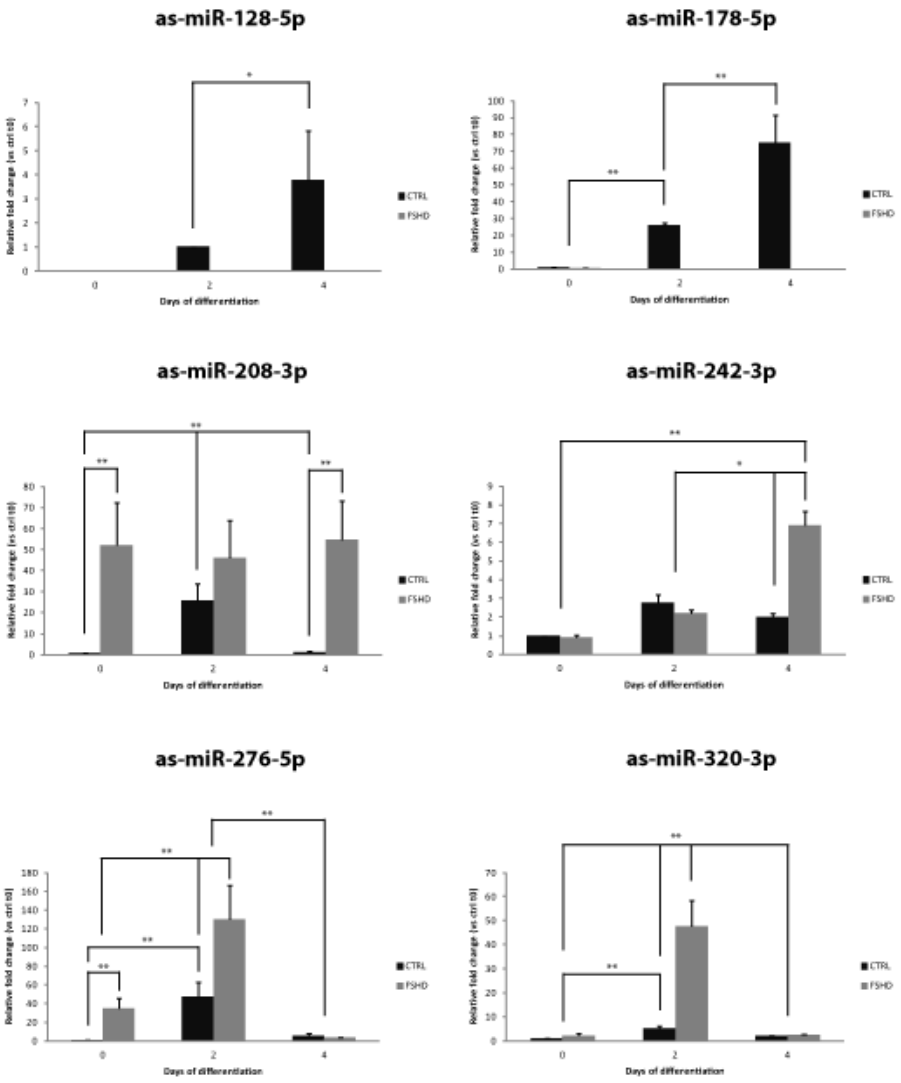


Fig. 23 - Analysis of D4Z4-derived miRNA expression by qRT-PCR, during early stages of differentiation in biceps control and FSHD muscle cells. Values correspond to fold change over expression level of control at time 0. The miR-103a was used as normalizer. Error bars indicate standard deviation. N = 3; * p-value < 0.05; ** p-value < 0.01.

Both as-miR-128-5p and as-miR-178-5p are preferentially expressed in control myoblasts and increase during differentiation. Conversely, as-miR-276-5p and

as-miR-320-3p are significantly expressed in affected cells, particularly after 2 days of differentiation. On the other hand, as-miR-208-3p and as-miR-242-3p show a monotonic trend in FSHD cells, which assist to a slight increase of as-miR-242-3p after 4 days only. In healthy muscle cells, instead, as-miR-242-3p remains constant during entire differentiation, whereas as-miR-208-3p shown a peak of expression at 2 days only (Fig. 23).

Overall, comparison of relative expression levels of all miRNAs reveals that as-miR-276-5p, both in control and FSHD, and as-miR-320-3p, particularly in FSHD, are over 5-8 orders of magnitude more expressed than the others (Fig. 24). For this reason, predictive analysis for gene target was initially conducted in relation to these two DUX4-AS2-derived miRNAs.

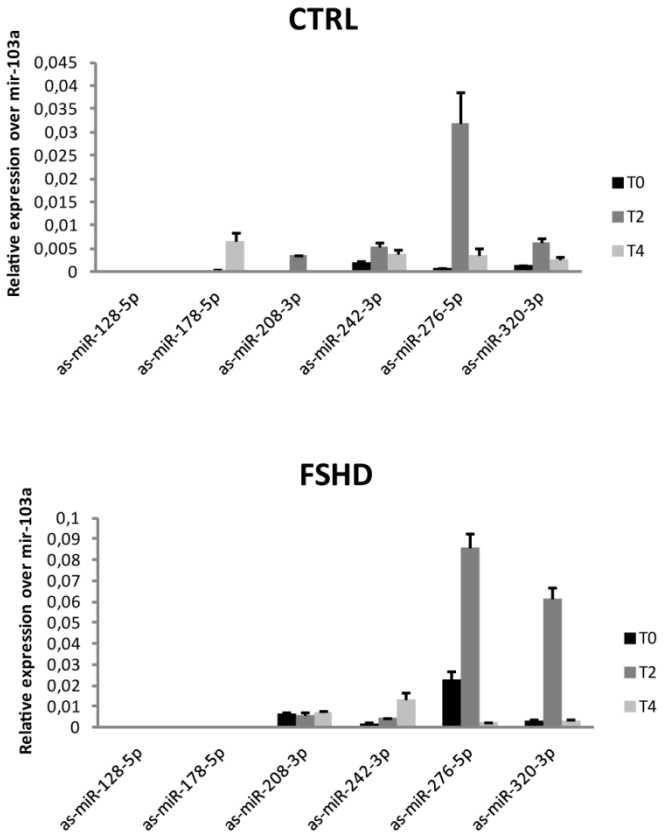


Fig. 24 - Comparison of D4Z4-derived miRNA relative expression levels by qRT-PCR during early stages of biceps control and FSHD differentiation. The miR-103a was used as normalizer. Error bars indicate standard deviation. N = 3.

3.2.11 Target prediction suggests involvement of as-miR-276-5p and as-miR-320-3p in myoblast fusion and early differentiation

To further understand the physiopathological functions concerning muscle differentiation of newly identified D4Z4-derived microRNAs, we oriented predictive analysis to guarantee accurate recognition of target genes involved in signaling pathways playing a pivotal role in orchestrating myoblast fusion and myofiber maturation processes, which we have thoroughly observed to be affected in FSHD.

Several pathways that could influence normal skeletal muscle development and differentiation have been described to be significantly deregulated in FSHD versus normal myoblasts and myotubes: MAPK signaling pathway (Dip et al, 2015), apoptosis (Kowaljow et al, 2007), oxidative phosphorylation (Turki et al, 2012), regulation of actin cytoskeleton (Dmitriev et al, 2011b; Rahimov et al, 2012; Tassin et al, 2012, Dib et al, 2015), calcium signaling pathway (Rahimov et al, 2012; Dib et al, 2015), RNA transport, RNA degradation (Dib et al, 2015) and VEGF signaling pathway (Osborne et al, 2007; Dib et al, 2015). Other dysregulated signaling networks in FSHD are the same involved in hypertrophic cardiomyopathy (Tsuji et al, 2009) and cancer (Dmitriev et al, 2014) too.

Focusing on as-miR-276-5p and as-miR-320-3p, we identified miRNA target sites against 3'UTR, 5'UTR and coding regions (CDS) using a combination of TargetScan, miRanda and PITA algorithms with default parameters. Moreover, thermodynamic stability of miRNA-mRNA interactions and site accessibility against 3'UTR and 5'UTR/CDS were evaluated using RNAHybrid and RNA22 prediction tools respectively (see Materials and Methods). We also cross the data with miRmap software results, for a more accurate assessment of bond energy and miRNA repression strength.

Among annotated mRNA transcripts (derived from GenBank database Release 167 and UniGene database Release 217) we found a total of 1535 and 299 putative targets for as-miR-276-5p and as-miR-320-3p respectively. All predicted target genes were then analyzed through Gene Ontology (GO) and KEGG (Release 53.0) databases.

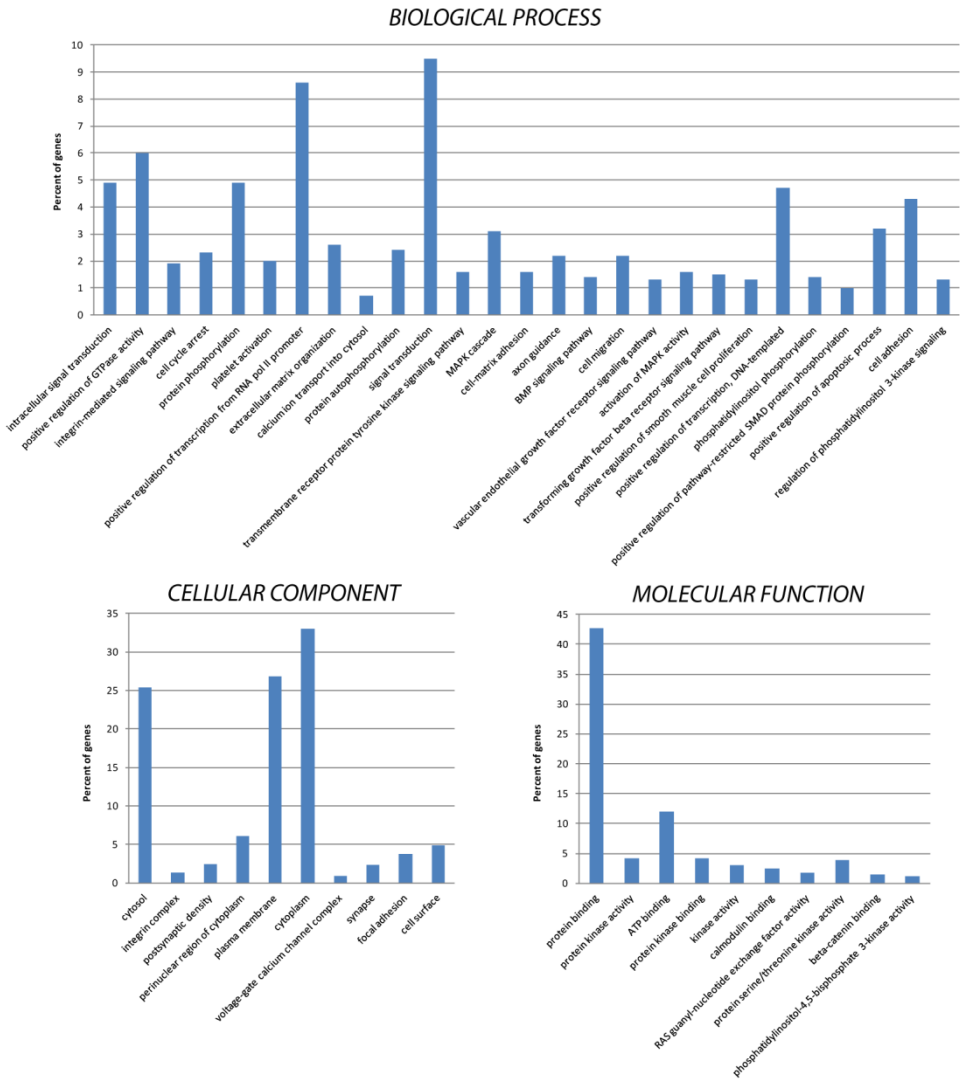


Fig. 25 - Gene ontology results for 1535 target genes of as-miR-276-5p. GO terms in the graphics are ordered from left to right according to their fold enrichment (from highest to lowest).

The GO terms enrichment analysis of the targets of as-miR-276-5p (Fig. 25) showed that among the biological processes (BP) annotated, a relatively high percentage of these were involved in signal transduction (9,5%) and positive regulation of transcription from RNA polymerase II promoter (8,6%) and GTPase activity (6%). Most represented cellular components (CC) were cytoplasm (33%) and plasma membrane (26,8%), whereas protein binding (42,7%) was the mainly enriched molecular function (MF) involving predicted as-miR-276-5p targets.

Analogously, as-miR-320-3p target gene analysis by GO revealed that signal transduction (10,8%) and positive regulation of transcription (9,1-6,1%) were

the BP terms most enriched, as well as protein binding (40,9%) in the MF annotation. On the other hand, as-miR-320-3p targets showed to be mainly localized and equally distributed between cytoplasm (34,1%) and nucleus (33,4%) (Fig. 26).

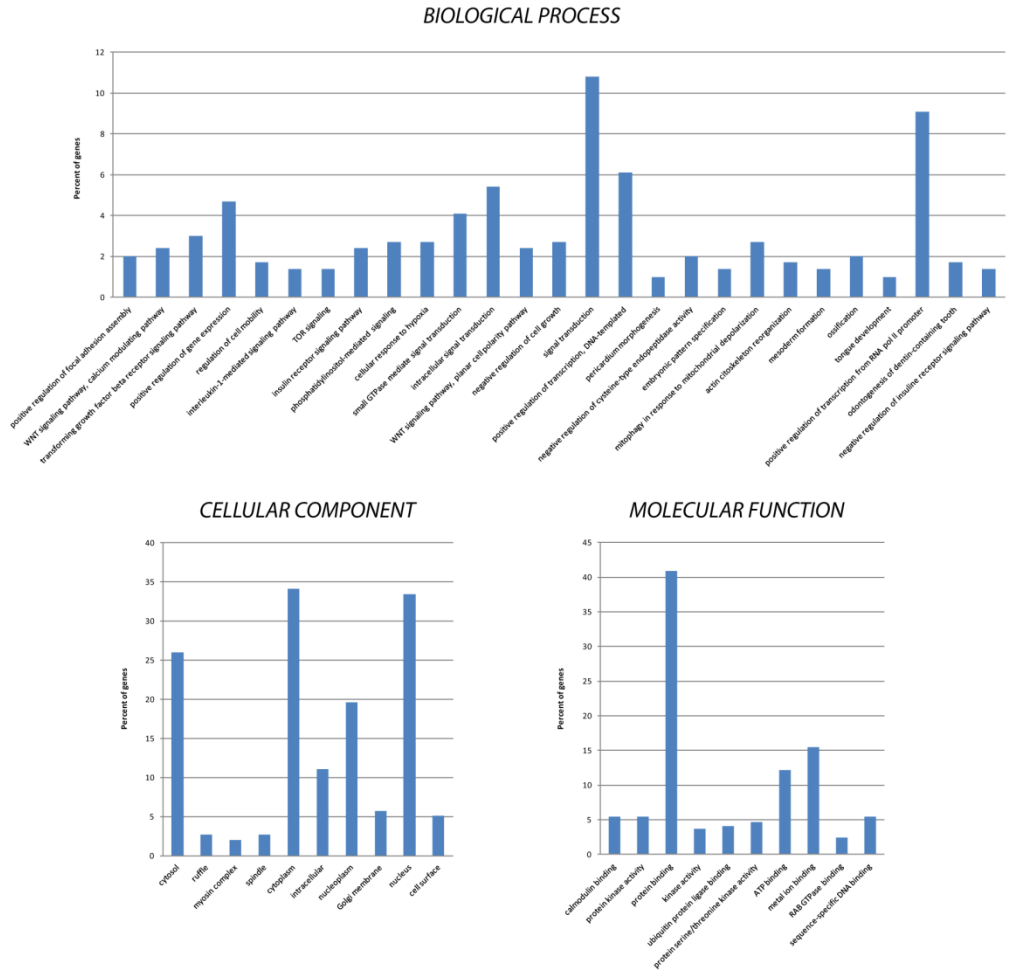


Fig. 26 - Gene ontology results for 299 target genes of as-miR-320-3p. GO terms in the graphics are ordered from left to right according to their fold enrichment (from highest to lowest).

KEGG pathways analysis revealed that approximately 22% of as-miR-276-5p targets fall within the signaling networks driving muscle differentiation and homeostasis, such as MAPK signaling pathway (86 genes), regulation of actin cytoskeleton (76 genes), TGF-beta, calcium and insulin signaling pathways (61, 58 and 57 genes respectively), focal adhesion (56 genes) and WNT, PI3K-AKT and mTOR signaling pathways (45, 37 and 24 genes respectively).

Myogenesis-related target genes of as-miR-320-3p were ~26% instead. Among them we found significantly enriched (FDR < $7{,}3 \cdot 10^{-9}$) the MAPK, WNT and mTOR signaling pathways (19, 16 and 11 genes respectively). Others were

also committed to PI3K-AKT, insulin and FoxO signaling pathways (16, 11 and 10 genes respectively), or to regulation of actin cytoskeleton (13 genes). Notably, more than 10% of total genes targeted by both as-miR-276-5p and as-miR-320-3p were annotated in KEGG pathways as cancer-associated.

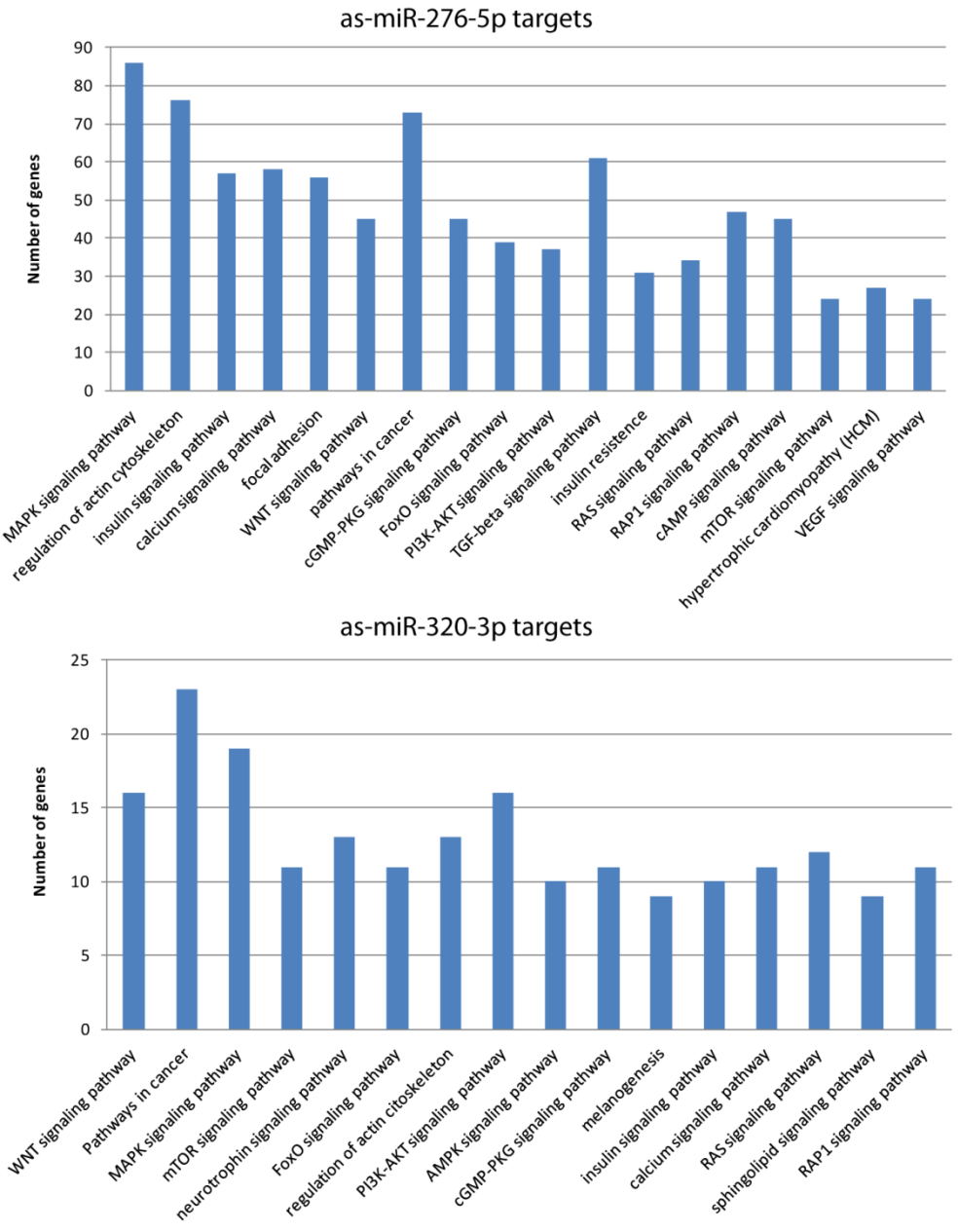


Fig. 27 - KEGG pathway enrichment results of muscle-related target genes of as-miR-276-5p and as-miR-320-3p. Pathways in the graphics are ordered from left to right according to their false discovery rate (FDR) values (from lowest to highest).

3.2.12 Overexpression of as-miR-276-5p and as-miR-320-3p affects members of mTOR signaling pathway

Interestingly, all myogenesis-associated pathways targeted by both as-miR-276-5p and as-miR-320-3p result strongly associated among them, whereas mTOR represents the crossroad between the upstream signaling mediated by MAPK, AMPK, WNT, RAS/ERK and IGF-1/PI3K/AKT, and the downstream effectors involved in cell growth and differentiation, in regulation of actin cytoskeleton, focal adhesion and VEGF signaling networks. Furthermore, mTOR complexes cross-talk with WNT and FoxO pathways in regulation of cell cycle and proliferation, contributing together with atrogin/MAFbx and MuFR-1 (activated by FoxO following insulin resistance), and the calmodulin-dependent kinases signaling cascade in the control of muscle hypertrophy.

We found that within mTOR signaling pathways, several members, such as RAPTOR, GβL, TSC1, TSC2, AKT2 and S6K1, are putative targets of both as-miR-276-5p and as-miR-320-3p; moreover, as-miR-320-3p was predicted to potentially regulate also myostatin and SMAD3.

Before testing effects of overexpression of both miRNAs, we decided to assay their putative targets involved in mTOR activity, in basal conditions, during early stages of differentiation, in control and FSHD muscle cells.

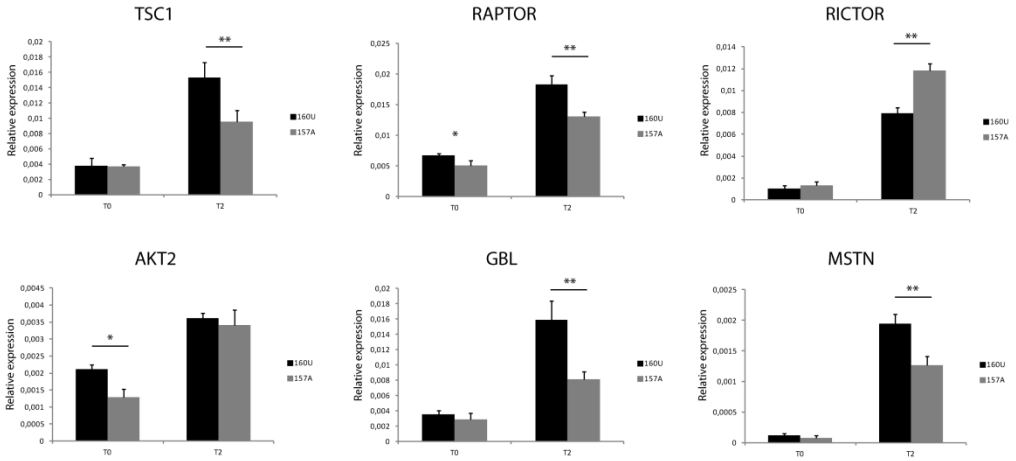


Fig. 28 - Expression of mTOR pathway genes in FSHD proliferating myoblasts and after two days of differentiation, compared to healthy control. GAPDH was used as normalizer. Error bars indicate standard deviation. N = 3; * p-value < 0.05; ** p-value < 0.01.

At two days of differentiation in KOSR-added medium, TSC1, RAPTOR, GβL and myostatin show to be significantly deregulated in FSHD cells compared to controls. Interestingly, deregulation of these genes occurs at the same time that the two D4Z4-derived miRNAs are up-regulated, in FSHD. We also observed AKT2 slightly down-regulated in FSHD myoblasts (T0), but equally expressed after beginning of differentiation. On the contrary, RICTOR, which play a critical role during myoblast fusion by mTORC2 signaling, was found up-regulated in FSHD cells, after two days of differentiation (Fig. 28).

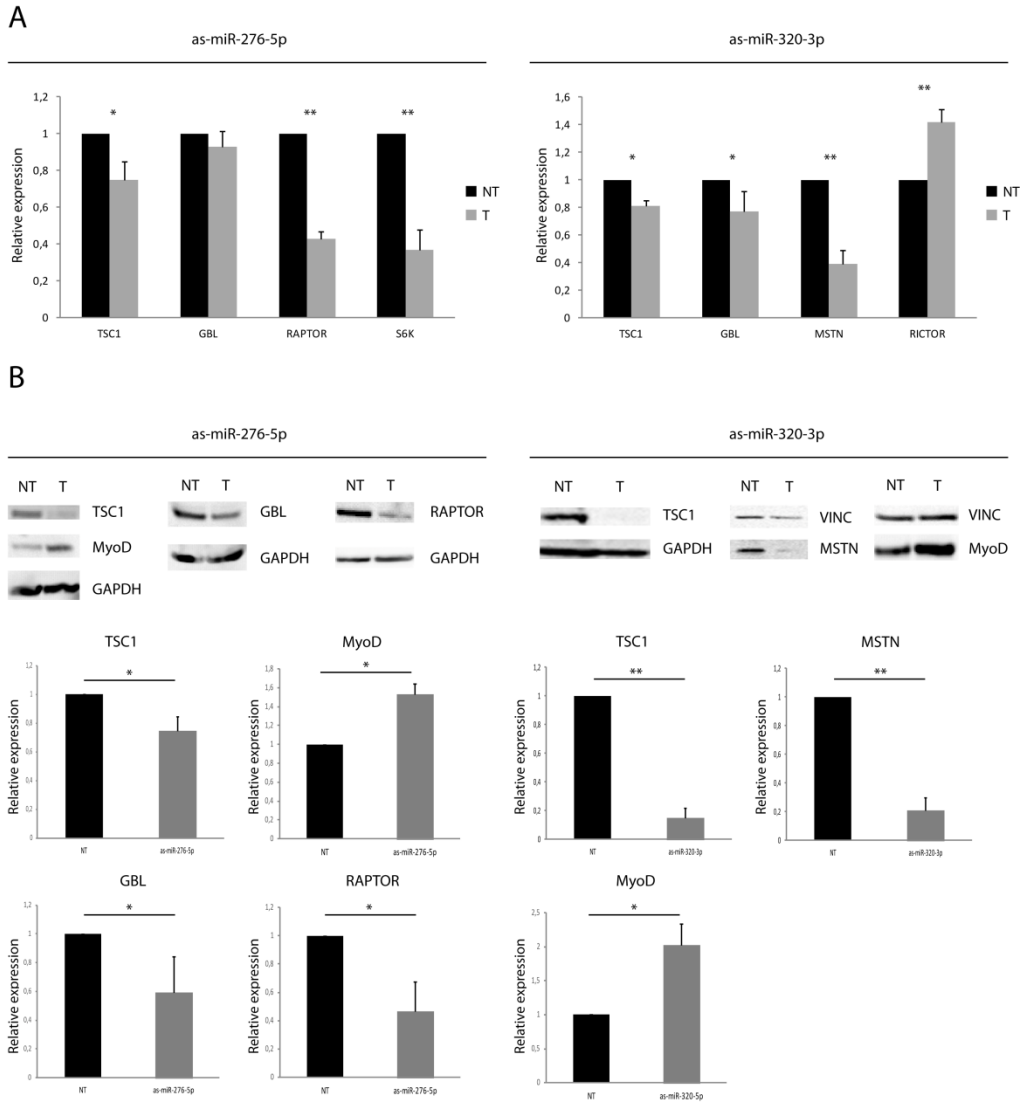


Fig. 29 - **A)** Expression analysis by qRT-PCR and **B)** assessment of protein levels after transfection of as-miR-276-5p and as-miR-320-3p in healthy myoblasts. Values correspond to relative expression against non-transfected cells (NT). GAPDH was used as normalizer. Error bars indicate standard deviation. N = 3; * p-value < 0.05; ** p-value < 0.01.

Transfection of as-miR-276-5p and as-miR-320-3p confirms, albeit indirectly, their involvement in regulation of mTOR signaling pathway.

Indeed, as-miR-276-5p significantly down-regulated mRNA levels of TSC1, RAPTOR, AKT2 and S6K1 (not altered under basal conditions), whereas, as-miR-320-3p overexpression significantly reduced the expression of TSC1, GβL and myostatin, determining concurrently an increase of RICTOR mRNA levels (Fig. 29A). At protein level, instead, deregulation mediated by as-miR-276-5p affects TSC1 and RAPTOR, as expected, and GβL too. Moreover, as-miR-320-3p significantly reduced TSC1 and myostatin, but not GβL protein levels. After transfection of both miRNAs, we observed a significant increase of MyoD, suggesting that negative regulation of such genes can rapidly shift equilibrium toward the myogenic differentiation program (Fig. 29B).

3.3 Structural and functional analysis of DUX4-fl

Physiopathological mechanism of action of the DUX4-fl protein remain still controversial, even though appears increasingly evident its role in myogenic differentiation, strengthening the idea of a direct interplay between D4Z4 and muscle determination proteins. We then worked to deepen the implications of DUX4-fl expression during muscle differentiation and to discover its elective interactors. To this aim, we have started to define structure of DUX4-fl by computational methods in order to predict its functional behavior.

3.3.1 Ab initio modeling of DUX4-fl protein

Infer the tridimensional structure of a protein can aid to predict its functional behavior and interactors, guide experimental efforts to their validation, and assist drug design. Operationally, existing computational methods of prediction can produce a structural model by using either comparative procedures based on sequence homology to proteins with known structure (template-based modeling), or *ab initio* (or *de novo*) modeling, without comparison with existing data.

```
  1       MALPTPSDSTLPAEARGRGRRRRLVWTPSQSEALRACFERNPYPGIATRE
 51       RLAQAIGIPEPRVQIWFQNERSRQLRQHRRESRPWPGRRGPPEGRRKRTA
101       VTGSQTALLLRAFEKDRFPGIAAREELARETGLPESRIQIWFQNRRARHP
151       GQGGRAPAQAGGLCSAAPGGGHPAPSWVAFAHTGAWGTGLPAPHVPCAPG
201       ALPQGAFVSQAARAAPALQPSQAAPAEGISQPAPARGDFAYAAPAPPDGA
251       LSHPQAPRWPPHPGKSREDRDPQRDGLPGPCAVAQPGPAQAGPQGQGVLA
301       PPTSQGSPWWGWGRGPQVAGAAWEPQAGAAPPPQPAPPDASASARQGQMQ
351       GIPAPSQALQEPAPWSALPCGLLLDELLASPEFLQQAQPLLETEAPGELE
401       ASEEAASLEAPLSEEEYRALLEEL
```

Fig. 30 - Sequence of DUX4-fl protein. Underlined residues refer to double homeobox domain.

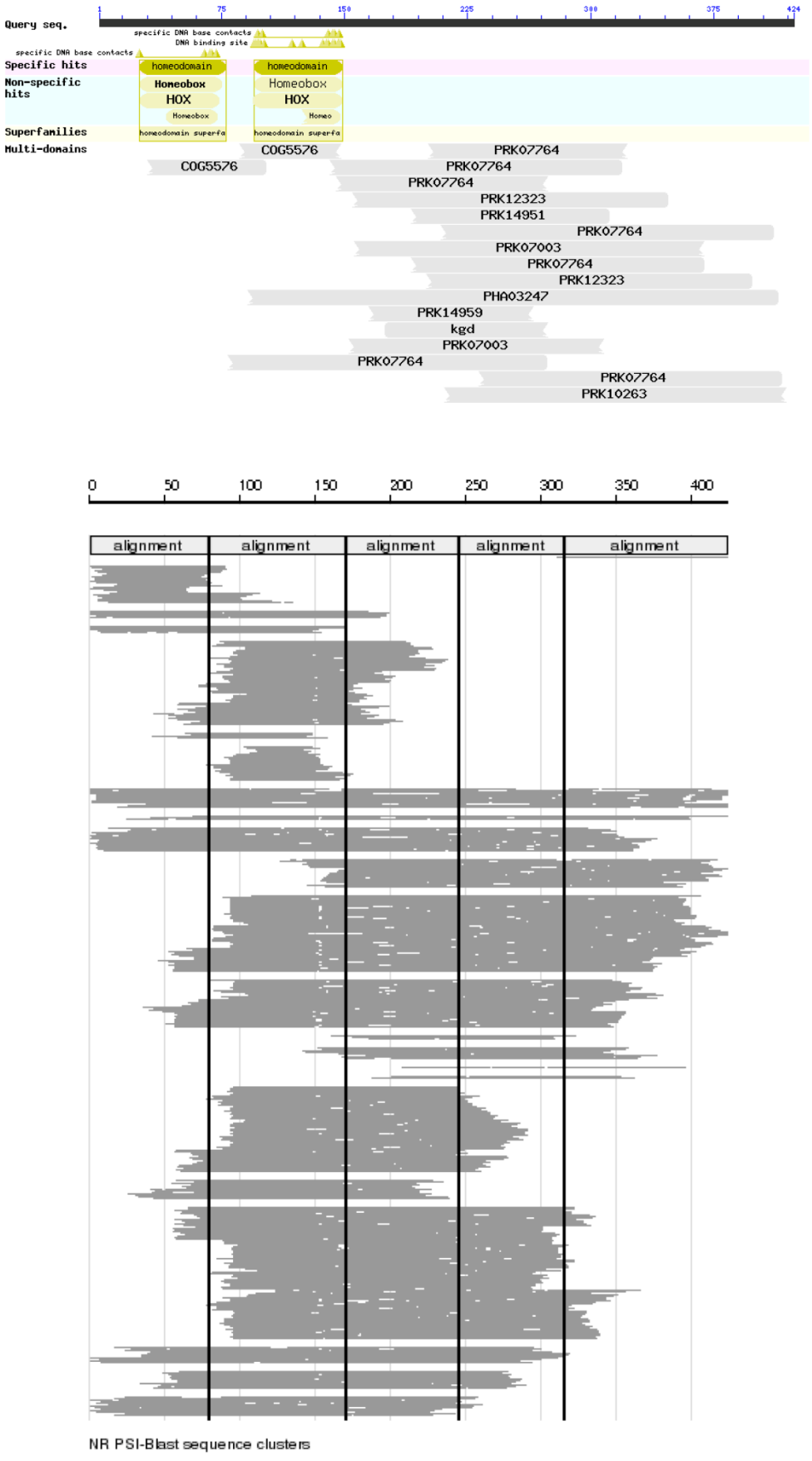


Fig. 31 - **(top)** Sequence homology-based conserved domains and **(bottom)** sequence clusters aligned to DUX4-fl. Vertical black lines separate domains predicted by Ginzu protocol.

Having to define the complete structure of DUX4-fl without information concerning its domain organization, except for N-terminus homeobox (HOX) domains, we preferred to start looking for regions that were homologous to experimentally determined structures, and at first, we used multiple sequence alignment (MSA)-based approach to infer putative domains.

Fig. 32 - Secondary structure of DUX4-fl protein. H = helix; E = beta sheet; C = coils.

DUX4-fl protein sequence of 424 aa in length (Fig. 30), was used as input for MSA analysis. Despite of the search for sequence homology may appear as a waiver of the ab initio modeling, the identification of homology-based structural model, as described below, can be evenly used as a reference, or constrains, for de novo protocol, reducing the processing time. Specifically, the search for homology-based conserved domains within DUX4-fl protein was performed by using PSI-BLAST software.

After three alignment iterative cycles against the NCBI non-redundant protein sequence databases, output sequences with highest score of significance were clustered according to level of homology, query covering and phylogenetic distance, by using CLANS software.

MSA analysis confirms the well-documented two homeodomains at residues 19-78 and 94-153. Furthermore, sequence clustering suggested the putative existence of other three C-terminal domains (Fig. 31), although without a strong correspondence with existing proteins. Homology in DUX4-fl residues from 79 to 419 were indeed found with low level of significance (E-value raging from $1.79 \cdot 10^{7}$ to $7.01 \cdot 10^{3}$) against the DNA polymerase III, subunits gamma and tau (different sub-alignments ranging between residues 79-416), the large tegument protein UL36 (residues 91-414), the alpha-ketoglutarase decarboxylase (residues 175-274) and the provisional DNA translocase Ftsk (residues 211-419).

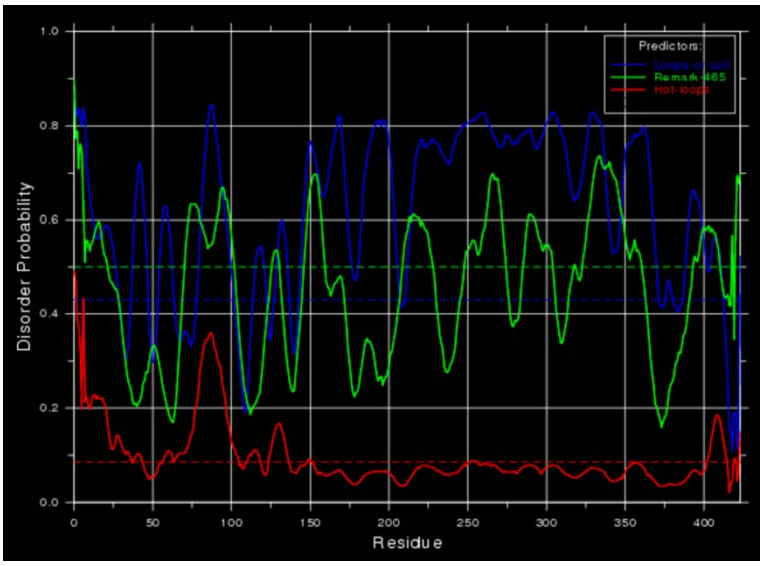


Fig. 33 - Residue probability to be intrinsically disordered, computed by different algorithms.

Secondary structure prediction analysis by PSIPRED and JPred4 confirmed helix-turn-helix motif in homeodomains, connected by short loop or beta-sheet regions (Fig. 32). Three alpha helices were also predicted in leucine-rich C-terminus domain, whereas neither alpha helix trans-membrane nor coiled-coil motifs have been recognized by using TMHMM and COILS algorithms.

Moreover, central domains (3rd and 4th) and the initial portion of last domain (5th) of DUX4-fl showed loop/coil organization and further investigation revealed that the protein might take an intrinsically disordered conformation. Indeed, relying solely on information deriving from secondary structure, both central and C-terminus domains have an high probability (0.6-0.9) to be disordered, although the degree of mobility of *hot loops*, as determined from C-alpha temperature B-factors, does not exceed the threshold of significance in much of the aminoacidic sequence (Fig. 33).

The possible confirmation of DUX4-fl as intrinsically disordered protein could have a major impact on the definition of its biological role and the degree of structural plasticity. Furthermore, low complexity regions (LCRs) found within DUX4-fl sequence can easily suggest which regions are particularly involved in protein-protein interaction. Indeed, LCRs-containing proteins have been shown to have more binding partners than proteins without LCRs (Coletta et al, 2010).

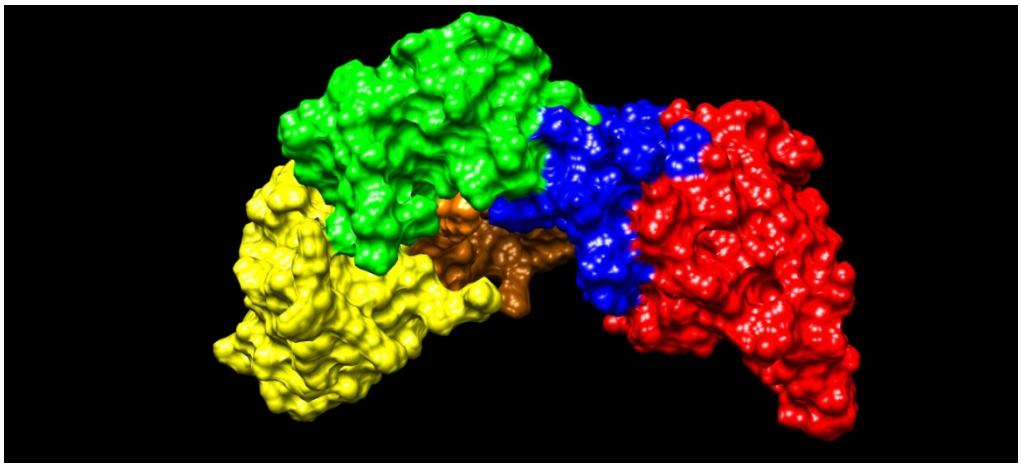


Fig. 34 - Tridimensional representation of best structural model of DUX4-fl. Colors highlight different functional domains: 1st and 2nd HOX domains (yellow and gold); 3rd (green) and 4th (blue) central domains; and 5th (red) C-terminal domain.

Data obtained from MSA analysis for conserved domains and information deriving from secondary structure prediction were used within a hierarchical screening performed by Ginzu protocol, in order to define precise start/end residue position of each putative domain. De novo modeling was then operated by Rosetta software, generating several decoys for each domain (~10.000), filtering out all of them having unlikely strand topologies and then clustering remaining decoys (~2.000) on C-alpha root-mean-square deviation (RMSD) over ungapped position. Finally, we selected five top ranked cluster centers to assemble five alternative models of whole protein.

Choice of final model (Fig. 34) was made by evaluating its prediction quality on ProSA web-server. It displays z-score of longest structural superposition of each model on the plot containing z-scores of the proteins in current PDB database (derived from both NMR and X-ray experiments). The z-score of the best DUX4-fl model (-7.81) represents the likelihood of getting a similar length match between similarly sized proteins by chance (Fig. 35).

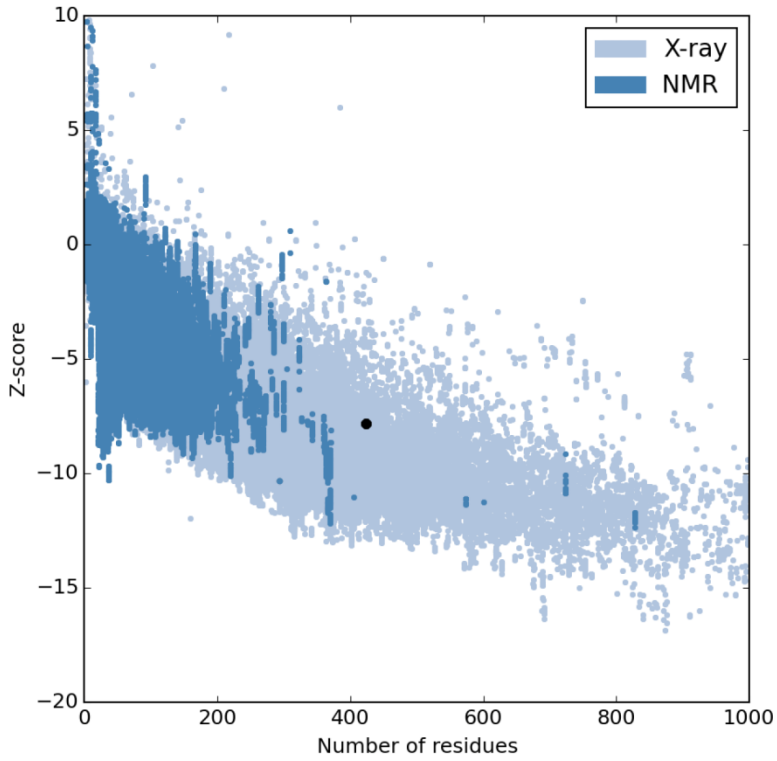


Fig. 35 - Position of DUX4-fl structure (black point) on plot containing the z-score of proteins with known structure deposited in PDB database.

3.3.2 Conformational and functional evaluation of DUX4-fl C-terminus domain reveals homology with ubiquitin-binding CUE domains

The two homeodomains of DUX4 have been extensively studied in the last years, in relation to their DNA-binding sites and specificity (Geng et al, 2011; Young et al, 2013; Zhang et al, 2016), and recently is also emerging a role of HOX domains in both DUX4 sub-cellular localization (Corona et al, 2013) and protein-protein interaction (Ansseau et al, 2016).

DUX4-s, which is considered to be the non-toxic isoform, is composed by homeodomains only, lacking the last three domains we found in full-length forms. Moreover, DUX4-fl differs from DUX4-fl2 by retention of the first intron, which we supposed be able to modify the structure and, consequently, the biological function of 5^{th} domain.

For these reasons, we chose to further investigate structural characteristics of C-terminus of DUX4-fl, in order to infer a functional role that may explain its mechanisms of action during healthy and FSHD skeletal muscle differentiation. Particularly, analysis has been initially addressed to 5^{th} domain (111 residues), deepening its geometric parameters by evaluation of Ramachandran plot (Fig. 36A), angle and bond length of the backbone, C-beta deviations and rotamers.

Quality of 5^{th} domain modeling was assayed by QMEAN, crossing geometric data with structural references and evaluating torsion angle potential in order to recognize if the model is close to its native structure (Fig 36B). Subsequently, through optimization of hydrogen bonds and all-atom energy minimization we were able to improve the model (RMSD: 0.111) (Fig. 37).

Refined model were then aligned to 37.751 tridimensional structures in PDB database (v.Feb17-16 - 90% non-redundant) by using both flexible and rigid alignment models. Superposition of 5^{th} domain after rigid alignment revealed its structural homology with both D-domain (score: 92.92; p-value > 0.001; identity: 5.88%; similarity: 23.53%; RMSD: 4.38) of F-box proteins, involved in protein degradation, and ubiquitin-binding CUE domain (score: 62.73; p-value > 0.05; identity: 7.14%; similarity: 21.43%; RMSD: 3.55) (Fig. 38A).

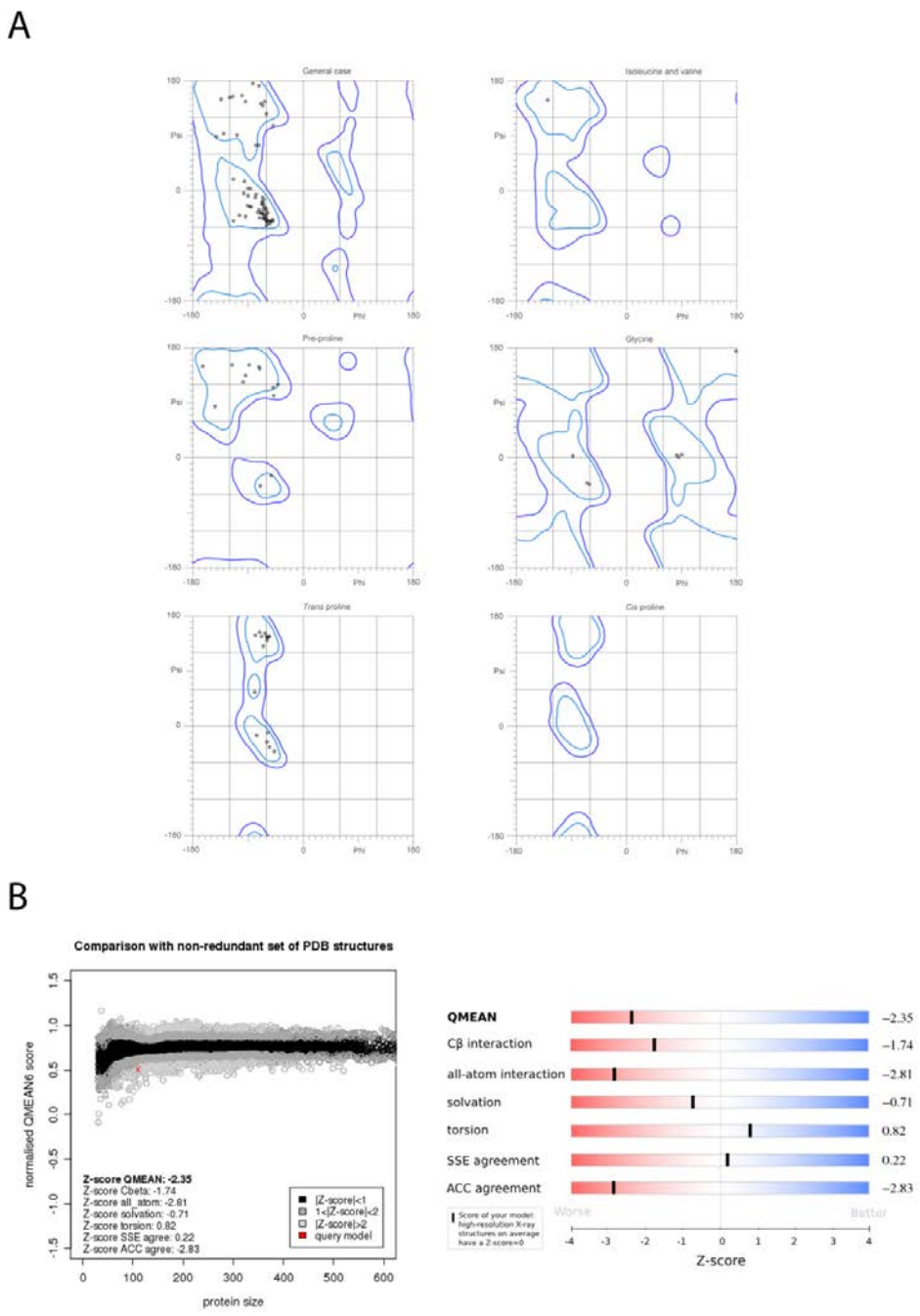


Fig. 36 - A) Ramachandran plot of DUX4-fl C-terminal domain. 96.3% of residues are in favored regions (blue) and 100% of residues are in allowed regions (violet) **B)** Comparison between QMEAN score of DUX4-fl C-terminal domain and PDB structures (left); the multi-parameter scores determining QMEAN value of DUX4-fl C-terminal domain model (right).

Since FSHD-related expression of DUX4-fl has been observed to affect protein ubiquitination, altering proteostasis and inducing aggregation of ubiquitinated proteins (Homma et al, 2015), we focused on homology with CUE domain and tried to verify whether 5th domain of DUX4-fl was able to bind ubiquitin.

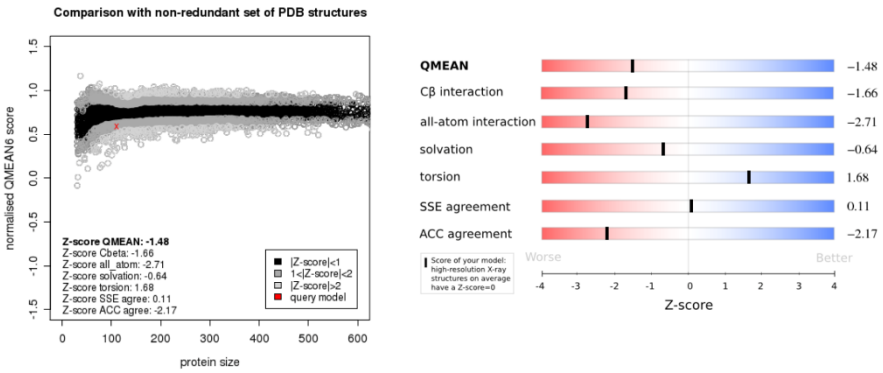
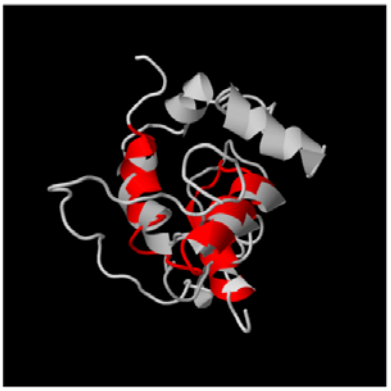


Fig. 37 - QMEAN scores of refined DUX4-fl C-terminal domain and PDB structures (left); the multi-parameter scores determining QMEAN value of refined model (right).

Therefore, we performed multi-alignment with human CUE domain consensus sequences in order to recognize aminoacidic binding motif in the 5th domain of DUX4-fl. Interestingly, multi-alignment shows that DUX4-fl carries with well-conserved di-leucine (LL), but lacks MFP ubiquitin-binding motif of CUE2 proteins (Fig. 38B).

A



B

```
Dux4d5      QALQEPAPWSALPCGLLLDELLASPEFLQQAQPLLETE
Cue1        ETVQNLAPN-LHPEQIRYS-LENTGSVEETVERYLRGD
Cue2.1      SILMDMFPA-ISKSKLQVHLLENNNDLDLTIGLLLKEN
Cue2.2      HQLYDMFPW-LDCSVIKDQFVINEKSVESTISDLLNYE
Cue3        SALMELFPW-FSKYWLSWTLLAYDNNIELVTNKIFEDP
Cue4        EIVMTMAPH-VPQEKVVQD-LRNTGSIEHTMENIFAGK
Cue5        QELKDAFPN-LEEKYIKAVIIASQGVLSPAFNALLFLS
```

Fig. 38 - A) Superposition of rigid structural alignment of DUX4-fl and CUE domain. Region of structural homology is highlighted in red. **B)** Multiple alignment between DUX4-fl C-

terminal domain (DUX4d5) and ubiquitin-binding CUE domains. The conserved residues of interaction with ubiquitin are squared in red.

Instead of MFP motif, DUX4-fl has PAP sequence, quite similar to LAP and MAP motifs found in CUE1 and CUE4, respectively. Alanine, in substitution of Phenylalanine, has been reported to reduce affinity with ubiquitin (Shih et al, 2003); conversely, despite of resulting helix instability, effects on binding affinity of Methionine- or Leucine-to-Proline (LAP –> PAP or MAP –> PAP) substitution in DUX4-fl should be deepened.

After having investigated structural and sequence similarity between 5th domain of DUX4-fl and ubiquitin-binding CUE domains, we proceeded with protein-protein interaction prediction by searching protein-protein interfaces and evaluating docking energy and complex stability by molecular dynamic simulation. For these analyses we used both 4th and 5th domains, assembled together, in order to include in the protein-protein interaction process the effects on 5th domain of its proximal regions. Ubiquitin X-ray structure refined at 1.8 Å resolution was obtained from PDB database (Accession ID: 1UBQ). Prediction confirmed binding to this protein by Leu8/Ile44/Val70 patch with a free energy of bonding interface of -31.57 Kcal/mol (Fig. 39).

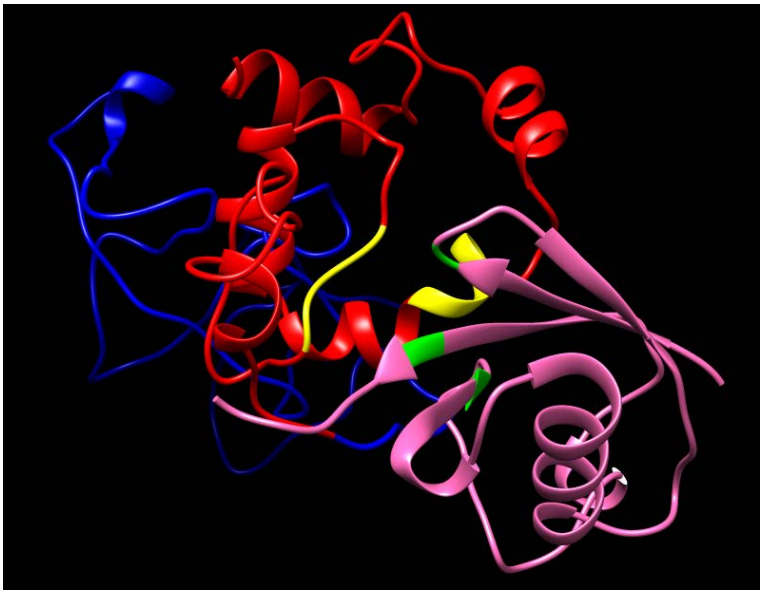


Fig. 39 - Tridimensional representation of interaction between 4th (blue) / 5th (red) C-terminal domain of DUX4-fl and ubiquitin (pink). In yellow are reported both PAP and LL ubiquitin-binding residues. Leu8/Ile44/Val70 hydrophobic patch is highlighted in green.

Molecular simulations of $4^{th}/5^{th}$ domains of DUX4-fl, both before and after interaction with ubiquitin, were performed by using "all-atom OPLS" and "GROMOS96 43a1" force field, respectively, in ion-added solvated system (spc216 model).

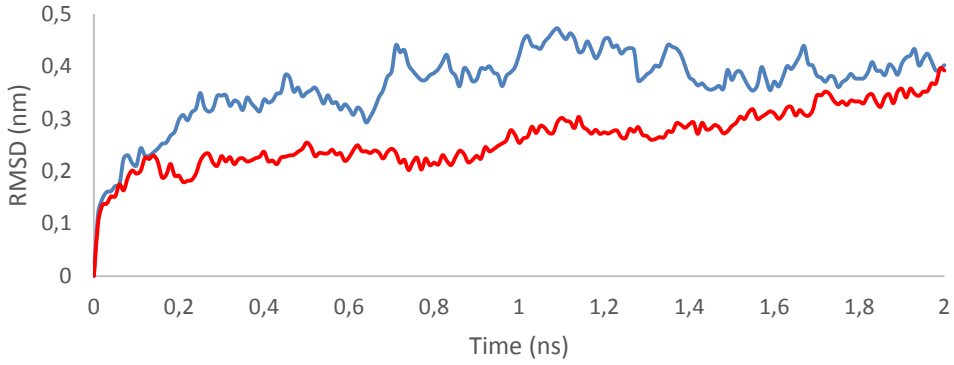


Fig. 40 - RMSD trajectories of the backbone (C-alpha) of $4^{th}/5^{th}$ C-terminal domain unbound (blue) and bound with ubiquitin (red) during simulation of 2 ns.

Energy minimization, in order to relax protein structure, and equilibration of the system at the appropriate temperature (300 K) and pressure (1 bar) values, to ensure its proper density, were set up. Finally, molecular dynamics were started and data collected for 2 ns.

Trajectory analysis corroborated interaction model and bond stability between C-terminus of DUX4-fl and ubiquitin (Fig. 40). Evaluation of fluctuation of the C-alpha residues of DUX4-fl $4^{th}/5^{th}$ domains explained high RMSD (0.4 nm) whose trajectories converge at the end of simulation. Indeed, free N-terminal of 4^{th} domain resulted highly unstable, as expected, whereas the entire backbone suffered fluctuation ranging from 1 to 2.5 Å (Fig. 41).

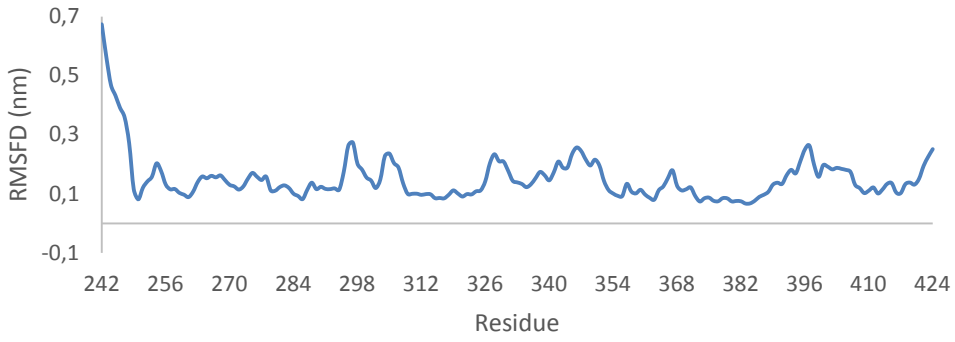


Fig. 41 - Average of backbone (C-alpha) fluctuations of residues of the 4th/ 5th C-terminal domain bound with ubiquitin, during simulation of 2 ns.

3.3.3 DUX4-fl translocates into cytoplasm during early muscle differentiation and its silencing slows myoblast fusion rate

Contrary to what we expected, immunofluorescence on FSHD and control cells revealed increasing cytoplasmic localization of DUX4-fl in fusing myoblasts and myotubes (Fig 42). These results are in agreement with a recent works on interaction between DUX4/DUX4c and several cytoplasmic proteins (Ansseau et al, 2016). Effectively, by comparison of DUX4-fl sequence with database for experimentally validated leucine-rich nuclear export signals (NESs)-containing proteins, we found two NESs within DUX4-fl: one ($LAQAIGI^{58}$) inside first homeodomain; and the other ($LPCGLLL^{374}$) located at the C-terminal domain.

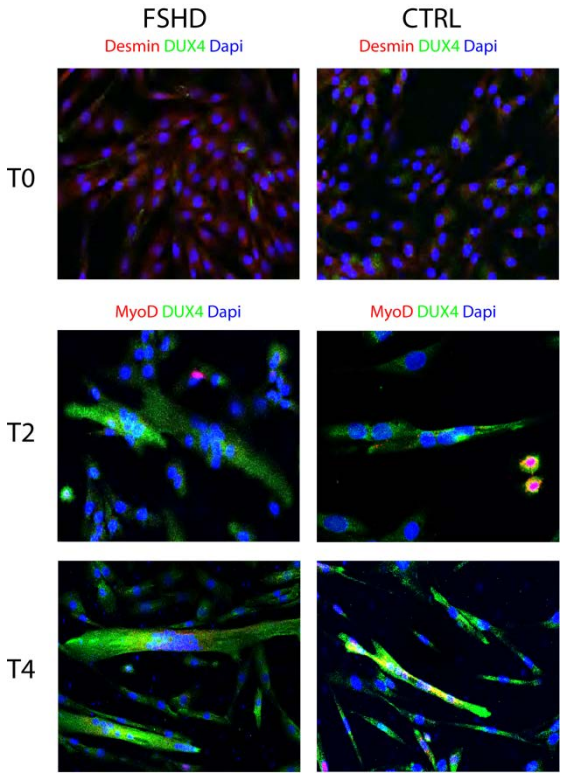


Fig. 42 - Subcellular localization of DUX4-fl (green) by immunofluorescence during FSHD and control differentiation. Desmin (T0) and MyoD (T2-4) are reported in red. Nuclei were stained with DAPI (blue).

In order to understanding the role played by DUX4-fl during differentiation, we performed experiment of gene silencing, by transfection of siRNA against C-terminus of DUX4.

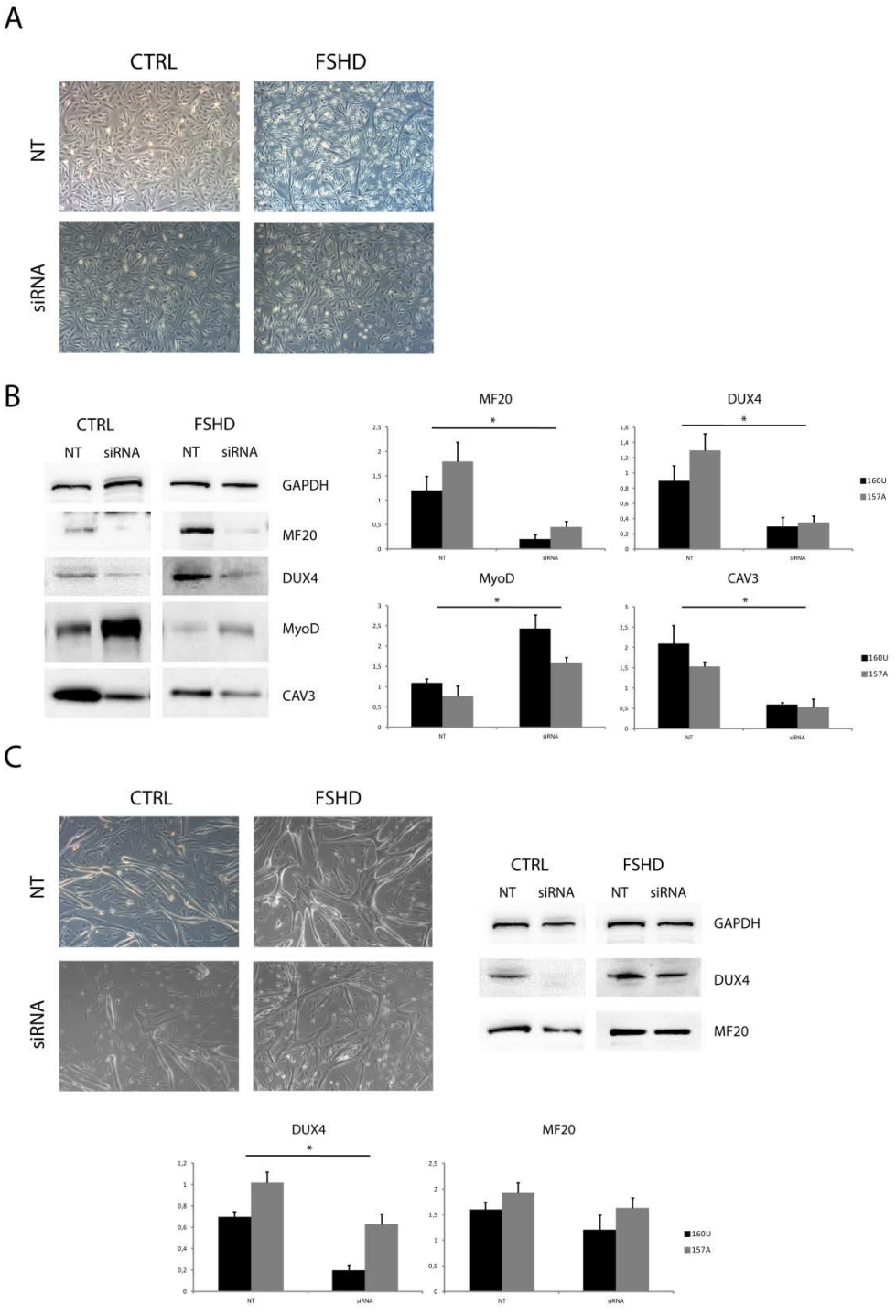


Fig. 43 - A) Phase contrast microscope images of FSHD and control myoblasts at day 2 (T2) of differentiation, after transfection with DUX4 siRNA at T0. B) DUX4-fl and muscle-specific markers evaluation by western blot after 2 days of differentiation of control and FSHD cells transfected at T0. C) Phase contrast microscope images of affected and control myoblasts, followed by both DUX4-fl and MF20 western blot analysis, at day 4 of differentiation, after transfection of DUX4 siRNA at T2. Error bars indicate standard deviation. N = 3; * p-value < 0.05.

Interestingly, we found that transient knockdown of DUX4-fl, when occurring at the beginning of differentiation, slowed myoblast fusion rate in FSHD and control cells (Fig. 43A), as confirmed by evaluation of muscle-specific markers (Fig. 43B). Conversely, the suppression of DUX4-fl after precise time point (T2) did not affect differentiation extent in both cell systems (Fig. 43C).

This experiment confirmed a role of DUX4-fl in early differentiation of healthy myoblasts as well. Besides, following DUX4-fl silencing at T0, we also appreciated reversion of aberrant phenotype occurring in FSHD. (Fig. 44).

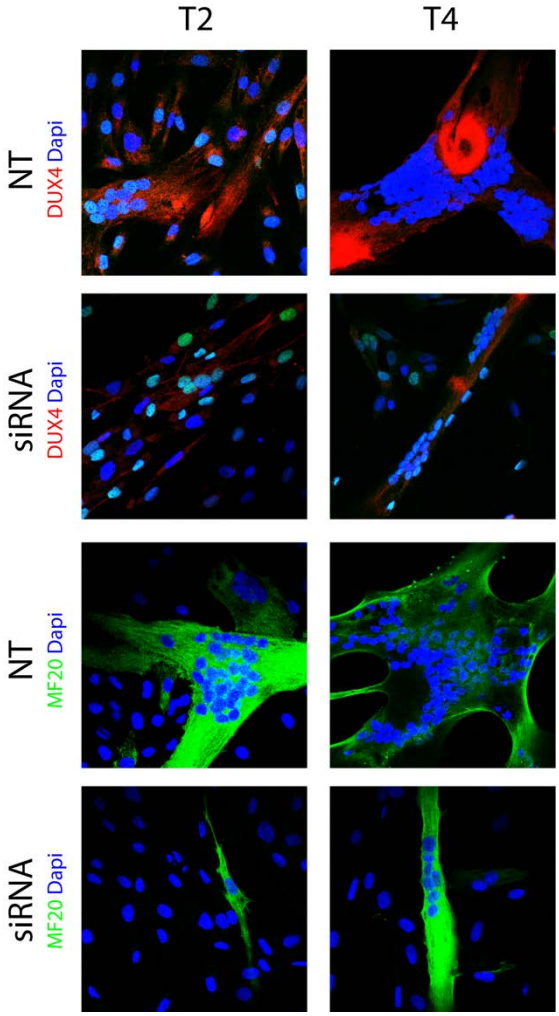


Fig. 44 - Immunofluorescence with antibodies against DUX4-fl (red) and MF20 (green) of FHSD cells, transfected with siRNA against DUX4-fl at T0, and evaluated dafter 2-4 days of differentiation. FSHD non-transfected cells (NT) was used as control. Nuclei were stained with DAPI (blue).

3.3.4 DUX4-fl co-localizes with M-cadherins, resulting in their altered distribution along plasma membrane in FSHD cells

As we revealed, FSHD myotubes display altered nuclei pattern, resulting from disorganized fusion events and inadequate myoblast polarization. We then hypothesized that alterations may affect surface proteins involved in focal adhesion and cell fusion. Among these, M-cadherins play a pivotal role in withdrawal of myoblasts, their fusion into myotubes and in nuclei organization (Zeschnigk et al, 1995).

Assessment of M-cadherins distribution in FSHD and control cell lines showed their delocalization along cell membrane and a significant accumulation in affected myoblasts and myotubes. Furthermore, by co-immunofluorescence, we also observed co-localization of M-cadherins with DUX4-fl, in FSHD muscle cells, compared to healthy control (Fig. 45), suggesting that altered distribution and interaction with DUX4-fl may be directly correlate.

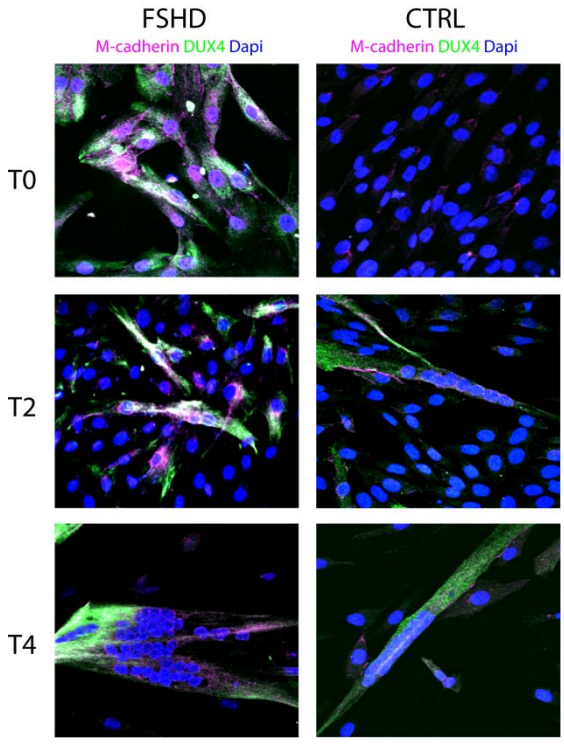


Fig. 45 - Co-immunofluorescence of M-cadherin (violet) and DUX4-fl (green) in FSHD and control cells, during muscle differentiation. Co-localization areas are spotted in white. Nuclei were stained with DAPI (blue).

Chapter IV - Discussion

Our work reports a further deepening on the transcriptional complexity of the D4Z4 macrosatellite, particularly focalized on the production of a long non-coding antisense RNA (DUX4-AS2) acting as a primary miRNA (pri-miRNA).
A previous study carried out by using a strand-specific RT-PCR approach, reported three sense and three antisense transcripts from D4Z4 repeats, in control and FSHD myoblasts and myotubes, as well as several different mi/siRNA-sized fragments from both coding sequence (CDS) and 5'/3'UTR of DUX4 gene, embedded within D4Z4 repeat. Moreover, these noncoding RNAs have been detected also in myoblasts obtained by MyoD-induced fibroblast trans-differentiation, but not in quiescent fibroblasts (Snider et al, 2009), suggesting the existence in D4Z4 macrosatellite of muscle-specific regulatory modules and highlighting a potential role of this region in myogenic determination program.
All three antisense RNAs, as well as their sense counterparts, displayed region of discontinuity with the DUX4 gene, suggesting their different functional activity, and particularly one (which Snider referred to as transcript D), overlapping in *head-to-head* orientation the 5' end of DUX4 gene, has been hypothesized may directly regulate sense gene expression (Lim et al, 2015). Furthermore, all these noncoding RNAs showed polymorphisms, suggesting that different D4Z4 repeats could be actively transcribed (Snider et al, 2009). This evidence introduces an additional element of variability, which results in a greater complexity at functional level.
Our initial approach, based on 3'RACE and deep sequencing, focalized on one of these antisense transcripts (herein named DUX4-AS2 and corresponding to previously reported transcript D) confirmed and extended all these findings. Particularly, we derived that: 1) DUX4-AS2 RNA is effectively transcribed in both FSHD and control muscle cells; 2) it is not polyadenylated, showing a length slightly higher (~75 nt) than that previously reported; and 3) other

D4Z4-like loci, or at least those located on chromosomes 3p12.3 and Yq11.21, transcribe detectable amounts of an homologous antisense transcript.

In addition, alignment of antisense transcripts deriving from different loci evidenced the occurrence of polymorphisms, hence allowing to conclude that different units, with variable efficiency, mainly those located at 4q35.2 and Yq11.21, can be transcribed.

Within D4Z4 locus, asymmetric bidirectional transcription has been recently reported following to identification of a distal upstream TATA-less promoter to DUX4 ORF (Block et al, 2012). TATA-less promoters have been classified as a large class of promoters associated with CpG-island-enriched sequences (Sandelin et al, 2007), driving multiple dispersed transcription start sites (Carninci et al, 2006; Marbach-Bar et al, 2015).

Analysis of CAGE and DNAse I hypersensitivity data in muscle cells, across DUX4 upstream region, revealed an enrichment close to the identified TATA-less promoter. According to its nature, this TATA-less/GC-rich promoter could explain observed clusters of transcription start points along DUX4 gene, particularly supporting our evidences of alternative transcription initiation for DUX4-AS2. Indeed, we found by analysis of SRX084669 dataset an antisense transcript, overlapping DUX4-AS2 itself, but starting 68 bp upstream. Further investigation by 5'RACE experiments will allow to confirm the eventually existence of a bias in TSS of this antisense RNA.

On the other hand, from literature, there are no evidences of transcriptional activity from DUX4-like genes of chromosomes 3 and Y, which are referred to as pseudogenes, and reported to be strongly heterochromatic and unexpressed. For that reason, our validation, by ssRT-PCR, of antisense transcripts deriving from these loci, represents a novelty in itself and opens up the question of possible transcription of gene counterpart and its impact at functional level.

In order to infer their biological role, we evaluated expression levels of DUX4-AS2, DUX4L26-AS and DUXY3-AS during muscle differentiation, in both bicep-deriving control and FSHD cells, confirming their involvement in early stages leading to myoblast fusion. Particularly, at 0-2 days of differentiation, DUX4-AS2 showed an expression level significant higher than DUX4L26-AS and DUXY3-AS, in FSHD cells. Conversely, DUX4L26-AS and DUXY3-AS was over-expressed in control, at 0-2 days, compared to affected cells.

More compelling than above evidences is the comparison between biceps and deltoid muscle, where we found, mainly in FSHD, that DUX4-AS2 is much more expressed in biceps than in less-affected deltoid. Since we have been observed, although not significantly, both an higher disorganization in myotube assembly and a faster fusion rate in biceps than deltoid, we conjectured that DUX4-AS2 may be involved in these processes. This suggestion was further endorsed by differential expression recorded between FSHD and control cells, where disparities in myoblast fusion were much more evident.

Due to sequence organization of DUX4-AS2 within the D4Z4 repeats, showing an overlapping to both 5' end and regulatory region of DUX4 gene, we initially postulated that antisense may affect DUX4 itself, through transcriptional activation, transcriptional repression or via chromatin modification (Ling et al, 2016). We have shown that DUX4 expression increases during differentiation, following the same trend, at least for initial days, of its antisense counterpart. However, over-expression experiments of DUX4-AS2, in FSHD myoblasts, confirmed absence of effects on DUX4 gene expression, as its mRNA and protein levels remained unchanged after antisense transfection.

Hypothesis that DUX4-AS2 may generate mi/siRNA-like fragments has been recently proposed yet (Lim et al, 2015). On the contrary to our results, Lim and colleagues presented evidences of DUX4 negative regulation mediated by antisense-derived siRNAs in a DICER/AGO2 dependent manner.

Nevertheless, all their experiments mainly consisted in transient transfection of several exogenous small RNAs mimicking the putative endogenous siRNAs, but did not show proofs of their effective existence in muscle cells.

Moreover, Lim and colleagues successfully tested repression strength against DUX4 of other D4Z4-derived cytoplasmic and/or chromatin/AGO2-associated short RNAs that are actively produced in WI38 primary human fibroblasts and HeLa cells (Benhamed et al, 2012; Ameyar-Zazoua et al, 2012).

Previous studies have been shown that antisense transcripts and their deriving chromatin-associated AGO2-coupled small RNAs are involved in epigenetic repression of DXZ4 macrosatellite (Chadwick, 2008; Pohlers et al, 2014).

Here we reported, instead, that DUX4-AS2 is localized into the nucleus, not chromatin-associated, and, as supported by results of RNA-protein interaction assay, undergoes to DGCR8/DROSHA processing to generate miRNAs.

On the other hand, unlike canonical miRNA precursors, we demonstrated that DUX4-AS2 is not-polyadenylated. However, a recent paper reported that most lncRNAs functioning as primary miRNA transcripts use a polyadenylation-independent mechanism of Polymerase II transcription termination, mediated by the microprocessor complex (Dhir et al., 2015).

By RNA secondary structure and stem-loop folding analysis, we predicted 16 miRNAs and amplified 6 of them in FSHD and control muscle cells. We also confirmed miRNA source by transfection of DUX4-AS2. Indeed, its over-expression markedly increase levels of all 6 miRNAs.

As we expected, none of these short RNAs corresponded to those previously reported by Snider and colleagues. Conversely, one cytoplasmic and two chromatin-associated small RNAs found in HeLa cells (Ameyar-Zazoua et al, 2012) showed to share sequence with predicted as-miR-11-5p, as-miR-68-3p and as-miR-242-3p.

Among these, only as-miR-242-3p was found in affected and control muscle cells, whereas both as-miR-11-5p and as-miR-68-3p, originating from the same pre-miRNA, were not successfully amplified in muscle system. Interestingly, one of exogenous siRNA, which is able to suppress DUX4 expression (Lim et al, 2015), matches the sequence of as-miR-178-5p we detected in control cells only.

Expression profiling of these miRNAs, moreover, highlighted quantitative and qualitative differences between FSHD and control cells, suggesting variability in processing activity and efficiency. Moreover, pri-miRNA conformational analysis showed that all precursor hairpins generating validated miRNAs can't coexist within the same structure.

Alternative DROSHA processing (Wu et al, 2009) and DROSHA/DICER cleavage shifting have been shown to produce miRNA end polymorphisms (Ruby et al, 2006; Seitz et al, 2008). Several pri-miRNAs have been reported to encode for subsets of clustered miRNAs (Liang et al, 2007), as well as has been recently observed the involvement of spliceosome complex in modulation of alternative biogenesis of several miRNAs, such as the splice-site-overlapping miR-412 (Melamed et al, 2013).

However, our findings represents a novelty, since we introduce a new level of alternative processing of pri-miRNA, depending on stabilization of a specific

subset of suboptimal structures. Indeed, two miRNA miRNAs (as-miR-128-5p and as-miR-178-5p) were detected only in control cells, whereas the others were expressed in both cell types, but showing a significant up-regulation in FSHD cells, during the early stages of differentiation. Particularly, as-miR-276-5p and as-miR-320-3p, after 48 hours of differentiation, exhibited an increase of expression several order of magnitude, compared to other miRNAs. Furthermore, as-miR-11-5p and as-miR-68-3p are encoded in HeLa, but not in muscle cells, and as-miR-242-3p occurs in both cell type, following a miRNA-like, in muscle cell, or a siRNA-like behaviour, in HeLa cell lines.

All together, these data strongly suggest that the unbalance in expressing endogenous miRNAs deriving from the same primary precursor can be tissue-specific and/or influenced by the pathophysiological status of the cell.

As above mentioned, two of the identified D4Z4-derived miRNAs (as-miR-276-5p and as-miR-320-3p) were found significantly up-regulated in FSHD cells during early stages of muscle differentiation. Defects in muscle differentiation of FSHD cells have been already reported (Winokur et al, 2003; Barro et al, 2010; Dib et al, 2015), and also FSHD cells used in this study showed differentiation defects implying a premature formation of disorganized myotubes. In the attempt to define the possible involvement of dysregulated miRNAs into described FSHD differentiation anomalies, we focused our attention to putative gene targets involved in signaling pathways playing a major role in myoblast fusion and myofiber maturation processes.

Our choice of investigating the effect of as-miR-276-5p and as-miR-320-3p on genes of the mTOR pathway was mainly based on several papers highlighting its central role in muscle differentiation and muscular dystrophy (Ge et al, 2009; De Palma et al, 2012, 2014; Wilson et al, 2016).

In particular, knockdown of RAPTOR, a member of mTORC1, was observed to enhance myoblasts differentiation (Jaafar et al, 2011; Ge et al, 2011; Wilson et al, 2016) whereas its over-expression suppressed myotubes formation (Ge et al, 2011). Negative effects of RAPTOR on muscle differentiation was proposed to be mediated by inhibition of RICTOR through S6K1 phosphorylation (Dibble et al, 2009; Julien et al, 2010; Jaafar et al, 2011). On the other hand, mTORC2 activates AKT (or protein kinase B), a critical regulator of skeletal myogenesis and muscle maintenance (Ge and Chen, 2012). Moreover, mTOR

signaling is involved in the finely regulation of both atrophic and hypertrophic stimuli. In this regard, TSC1/TSC2 complex, which is triggered by FoxO3 and myostatin/SMAD3 axis, controls atrophy program and its down-regulation results in enlarged myotubes (Kuleesha et al, 2016).

Effectively, overexpression of as-mirR-276-5p and as-mir-320-3p in muscle cells reduced the expression of molecular markers controlling cell proliferation and atrophy (such as RAPTOR, TSC1 and myostatin), and enhanced RICTOR mRNA and MyoD protein levels, shifting the balance toward differentiation program, hence explaining the observed premature fusion of FSHD myoblasts into swelled myotubes.

Anyway, further investigation will be necessary and experimental efforts will be oriented to validate target genes of as-mirR-276-5p and as-mir-320-3p, as well as of the other predicted miRNAs.

On the other hand, an analogous pro-differentiation effect we also highlighted for the master gene of FSHD: DUX4. Indeed, According to recent findings (Ansseau et al, 2016), we found that DUX4-fl isoform is equally expressed in both affected and healthy myoblasts, increases during muscle differentiation and translocates into the cytoplasm, where it could be involved in focal adhesion and cell fusion signaling, particularly regulating the re-cycling of M-cadherins.

The choice between degradation and recycling can help fine-tune the amount of cadherin present at adherens junctions and the strength of cell-cell adhesion (Nanes and Kowalczyk, 2012). Moreover, the fusion process mediated by the M-cadherin-β-catenin complex is tightly regulated and its persistence might have a dramatically negative effect on terminal myogenic differentiation (Kramerova et al, 2006).

Degradation of M-cadherins is mediated by ubiquitination and proceeds in a proteasome-dependent manner. In our study we found that M-cadherins are unevenly distributed along plasma membrane of FSHD cells, particularly in the first stages of muscle differentiation, where co-localizes with DUX4-fl.

From predictive analysis on tertiary structure of DUX4-fl, we revealed that it may have characteristics of intrinsically disordered proteins (IDPs). Due to their structural plasticity, IDPs can interact with multiple proteins or interface with a single protein through distinct points of interaction (Van Roey et al,

2014; Habchi et al, 2014). These evidences, in particular, could explain results reported by Ansseau et colleagues about validation of a large group of very different DUX4 cytoplasmic interactors.

Furthermore, several protein that regulate ubiquitination possess IDRs and, specifically, de-ubiquitinating enzyme USP36 uses them to scaffold multiple protein interaction sites (Reed et al, 2015). By structural alignment, we found homology between DUX4-fl C-terminal domain and ubiquitin-binding CUE domain, and we have computationally validated interaction with ubiquitin by molecular dynamics simulation. As previously mentioned, the involvement of DUX4-fl in ubiquitin pathway has been recently reported (Homma et al, 2015). For these reasons, and supported by these results, we firstly hypothesized that DUX4-fl may disrupt M-cadherins ubiquitin-mediated degradation signaling, enhancing in affected myoblasts the fusion process.

We demonstrated that DUX4-fl silencing during first stages of differentiation slows fusion rate, driving to normal myotube formation. However, same effects on myogenic differentiation was observed in control cells, suggesting that DUX4-fl is similarly involved in muscle differentiation, but it behaves in a different, more controlled way.

Certainly, these remain preliminary evidences and future aims will be orientate to confirm these observation and corroborate structural and interaction models we inferred. Nevertheless, all reported results represent an additional evidence of the transcriptional complexity of the D4Z4 genomic region and its possible impact on the pathophysiology of skeletal myogenesis.

Bibliography

Abmayr SM, Pavlath GK. (2012) Myoblast fusion: lessons from flies and mice. Development. 139:641-656

Alexander M S, Casar JC, Motohashi N, Myers JA, Eisenberg I, Gonzalez RT, et al. (2011). Regulation of DMD pathology by an ankyrin-encoded miRNA. Skelet. Muscle. 1:27

Almeida M, Han L, Ambrogini E, Weinstein RS, Manolagas SC. (2011) Glucocorticoids and tumor necrosis factor a increase oxidative stress and suppress Wnt protein signaling in osteoblasts. J Biol Chem. 286(52):44326-35

Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau JC, Mathieu J, Hamiche A, Ait-Si-Ali S, Muchardt C, Batsché E, Harel-Bellan A. (2012) Argonaute proteins couple chromatin silencing to alternative splicing. 19(10):998-1004

Analeah BH, Rojas A, Ian S. Harris BLB. (2007) Determinants of Myogenic Specificity within MyoD Are Required for Noncanonical E Box Binding? Mol. Cell Biol. 27(16):5910-5920

Ansseau E, Laoudj-Chenivesse D, Marcowycz A, Tassin A, Vanderplanck C, Sauvage S, Barro M, Mahieu I, Leroy A, Leclercq I, Mainfroid V, Figlewicz D, Mouly V, Butler-Browne G, Belayew A, Coppée F. (2009) DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. PLoS One. 4(10):e7482

Ansseau E, Eidahl JO, Lancelot C, Tassin A1, Matteotti C1, Yip C1, Liu J, Leroy B, Hubeau C, Gerbaux C, Cloet S1, Wauters A, Zorbo S, Meyer P, Pirson I, Laoudj-Chenivesse D, Wattiez R, Harper SQ, Belayew A, Coppée F. (2016) Homologous Transcription Factors DUX4 and DUX4c Associate with Cytoplasmic Proteins during Muscle Differentiation. PLoS One. 11(1):e0146893

Arashiro P, Eisenberg I, Kho AT, Cerqueira AM, Canovas M, Silva HC, Pavanello RC, Verjovski-Almeida S, Kunkel LM, Zatz M. (2009) Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. Proc Natl Acad Sci U S A. 106(15):6220-5

Bae GU, Yang YJ, Jiang G, Hong M, Lee HJ, Tessier-Lavigne M, Kang JS, Krauss RS. (2009) Neogenin regulates skeletal myofiber size and focal adhesion kinase and extracellular signal-regulated kinase activities in vivo and in vitro. Mol Biol Cell. 20:4920-4931

Bakkar N, Wang J, Ladner KJ, Wang H, Dahlman JM, Carathers M, Acharyya S, Rudnicki MA, Hollenbach AD, Guttridge DC. (2008) IKK/NF-kappaB regulates skeletal myogenesis via a signaling switch to inhibit differentiation and promote mitochondrial biogenesis. J Cell Biol. 180(4):787-802

Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, van der Wielen M, Rasmussen K, Frants RR. (1995) The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve Suppl. (2):S39-44

Balakirev ES, Ayala FJ. (2003) Pseudogenes: are they "junk" or functional DNA? Annu Rev Genet. 37:123-51

Ballarino M, Cazzella V, D'Andrea D, Grassi L, Bisceglie L, Cipriano A, Santini , Pinnarò C, Morlando M, Tramontano A, Bozzoni I. (2015) Novel Long Noncoding RNAs (lncRNAs) in Myogenesis: a miR-31 Overlapping lncRNA Transcript Controls Myoblast Differentiation. Molecular and Cellular Biology. 4:728-736

Barro M, Carnac G, Flavier S, Mercier J, Vassetzky Y, Laoudj-Chenivesse D. (2010) Myoblasts from affected and non-affected FSHD muscles exhibit morphological differentiation defects. J Cell Mol Med. 14(1-2):275-89

Beauchamp JR, Heslop L, Yu DS, Tajbakhsh S, Kelly RG, Wernig A, Buckingham ME, Partridge TA, Zammit PS. (2000) Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. J. Cell Biol. 151(6):1221-34

Benhamed M, Herbig U, Ye T, Dejean A, Bischof O. (2012) Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. Nat Cell Biol. 14(3):266-75

Benkert, P, Tosatto, SCE, Schomburg, D. (2008) QMEAN: A comprehensive scoring function for model quality assessment. Proteins: Structure, Function, and Bioinformatics. 71(1):261-277

Bentzinger CF, Wang YX, Rudnicki MA. (2012) Building muscle: molecular regulation of myogenesis. Cold Spring Harb Perspect Biol. 4(2)

Berkes CA, Tapscott SJ. (2005) MyoD and the transcriptional control of myogenesis. Semin Cell Dev Biol. 16:585-595

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F, et al. (2010) A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. Embo J. 29:3082-3093

Bindoff LA, Mjellem N, Sommerfelt K, Krossnes BK, Roberts F, Krohn J, Tranheim RS, Haggerty ID. (2006) Severe fascioscapulohumeral muscular dystrophy presenting with Coats' disease and mental retardation. Neuromuscul Disord. 16(9-10):559-63

Blattler S. M., Cunningham J. T., Verdeguer F., Chim H., Haas W., Liu H., Romanino K., Rüegg M. A., Gygi S. P., Shi Y., Puigserver P. (2012) Yin Yang 1 deficiency in skeletal muscle protects against rapamycin-induced diabetic-like symptoms through activation of insulin/IGF signaling. Cell Metab. 15:505-517

Block GJ, Narayanan D, Amell AM, Petek LM, Davidson KC, Bird TD, Tawil R, Moon RT, Miller DG. (2013) Wnt/ß-catenin signaling suppresses DUX4

expression and prevents apoptosis of FSHD muscle cells. Hum Mol Genet. 22(23):4661-72

Block GJ, Petek LM, Narayanan D, Amell AM, Moore JM, Rabaia NA, Tyler A, van der Maarel SM, Tawil R, Filippova GN, Miller DG. (2012) Asymmetric bidirectional transcription from the FSHD-causing D4Z4 array modulates DUX4 production. PLoS One. 7(4):e35532

Bodega B, Ramirez GD, Grasser F, Cheli S, Brunelli S, Mora M, Meneveri R, Marozzi A, Mueller S, Battaglioli E, Ginelli E. (2009) Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. BMC Biol. 7:41

Bodine SC, Stitt TN, Gonzalez M, Kline WO, Stover GL, Bauerlein R, Zlotchenko E, Scrimgeour A, Lawrence JC, Glass DJ, Yancopoulos GD. (2001) Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. Nat Cell Biol. 3(11):1014-9.

Bosnakovski D, Lamb S, Simsek T, Xu Z, Belayew A, Perlingeiro R, Kyba M. (2008) DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Exp Neurol. 214(1):87-96

Brodersen P, Voinnet O. (2009) Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol. 10(2):141-8

Brouwer OF, Padberg GW, Ruys CJ, Brand R, de Laat JA, Grote JJ. (1991) Hearing loss in facioscapulohumeral muscular dystrophy. Neurology. 41:1878-81
Brunelli S, Tagliafico E, De Angelis FG, Tonlorenzi R, Baesso S, Ferrari S, Niinobe M, Yoshikawa K, Schwartz RJ, Bozzoni I, Ferrari S, Cossu G. (2004) Msx2 and necdin combined activities are required for smooth muscle differentiation in mesoangioblast stem cells. Circ Res. 94(12):1571-8

Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME. (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. Cell. 96(6):857-68.
Bryan B, Cai Y, Wrighton K, Wu G, Feng XH, Liu M (2005) Ubiquitination of RhoA by Smurf1 promotes neurite outgrowth. FEBS Lett. 579:1015-1019

Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 41(W1):W340-W348

Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D. (2012) A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell. 149(4):819-31

Cabianca DS, Casa V, Gabellini D. (2012) A novel molecular mechanism in human genetic disease: a DNA repeat-derived lncRNA. RNA Biol9(10):1211-7

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25(18):1915-27

Cacchiarelli D, Martone J, Girardi E, Cesana M, Incitti T, Morlando M, Nicoletti C, Santini T, Sthandier O, Barberi L, Auricchio A, Musarò A, Bozzoni I. (2010) MicroRNAs involved in molecular circuitries relevant for the Duchenne muscular dystrophy pathogenesis are controlled by the dystrophin/nNOS pathway. Cell Metab. 12(4):341-51.

Cacchiarelli D, Incitti T, Martone J, Cesana M, Cazzella V, Santini T, Sthandier O, Bozzoni I. (2011) miR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. EMBO Rep. 12(2):136-41

Cai X, Cullen BR. (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA. 13(3): 313-316

Caretti G, Schiltz RL, Dilworth FJ, Di Padova M, Zhao P, Ogryzko V, Fuller-Pace FV, Hoffman EP, Tapscott SJ, Sartorelli V. (2006) The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. Dev Cell. 11:547-560

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, et al. (2006) V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. 38:626-635

Castorina A, Szychlinska MA, Marzagalli R, Musumeci G. (2015) Mesenchymal stem cells-based therapy as a potential treatment in neurodegenerative disorders: is the escape from senescence an ansie? Neural Regen Res. 10(6): 850-858

Cech TR, Steitz JA. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 157(1):77-94

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell. 147(2):358-69

Chadwick BP. (2008) DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. Genome Res. 18:1259-1269

Charrasse S, Meriane M, Comunale F, Blangy A, Gauthier-Rouviere C. (2002) N-cadherin-dependent cell–cell contact regulates Rho GTPases and ß-catenin localization in mouse C2C12 myoblasts. J. Cell Biol. 158:953-965

Charrasse S, Comunale F, Grumbach Y, Poulat F, Blangy A, Gauthier-Rouviere C. (2006) RhoA GTPase regulates M-cadherin activity and myoblast fusion. Mol. Biol. Cell. 17:749-759

Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 465(7298):584-9

Chen VB, Arendall WB, Headd JJ, Keedy DA, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 66(Pt 1):12-21

Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP. (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 24(10):992-1009

Clapp J, Mitchell LM, Bolland DJ, Fantes J, Corcoran AE, Scotting PJ, Armour JA, Hewitt JE. (2007) Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am J Hum Genet. 81(2):264-79

Coletta A, Pinney JW, Solís DY, Marsh J, Pettifer SR, Attwood TK. (2010) Low-complexity regions within protein sequences have position-dependent roles. BMC Syst Biol. 4:43.

Core LJ, Waterfall JJ, Lis JT. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 322(5909):1845-8

Cornelison DD, Olwin BB, Rudnicki MA, Wold BJ. (2000). MyoD(-/-) satellite cells in single-fiber culture are differentiation defective and MRF4 deficient. Dev. Biol. 224:122-137

Corona ED, Jacquelin D, Gatica L, Rosa AL. (2013) Multiple protein domains contribute to nuclear import and cell toxicity of DUX4, a candidate pathogenic protein for facioscapulohumeral muscular dystrophy. PLoS One. 8(10):e75614

Cossu G, Tajbakhsh S, Buckingham M. (1996) How is myogenesis initiated in the embryo? Trends Genet. 12:218-223

Crist CG, Montarras D, Buckingham M. (2012) Muscle satellite cells are primed for myogenesis but maintain quiescence with sequestration of Myf5 mRNA targeted by microRNA-31 in mRNP granules. Cell Stem Cell. 11(1):118-26

Cuenda A, Cohen P. (1999) Stress-activated protein kinase-2/p38 and a rapamycin-sensitive pathway a rerequired for C2C12 myogenesis. J. Biol. Chem. 274:4341-4346

da Huang, Sherman BT, Lempicki RA. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4(1):44-57

de Greef JC, Wohlgemuth M, Chan OA, Hansson KB, Smeets D, Frants RR, Weemaes CM, Padberg GW, Van Der Maarel SM. (2007) Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. Neurology. 69(10):1018-26

de Greef JC, Lemmers RJ, van Engelen BG, Sacconi S, Venance SL, Frants RR, Tawil R, van der Maarel SM. (2009) Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum Mutat. 30(10):1449-59

de Greef JC, Lemmers RJ, Camaño P, Day JW, Sacconi S, Dunand M, van Engelen BG, Kiuru-Enari S, Padberg GW, Rosa AL, Desnuelle C, Spuler S, Tarnopolsky M, Venance SL, Frants RR, van der Maarel SM, Tawil R. (2010) Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology. 75(17):1548-54

De Palma C, Morisi F, Cheli S, Pambianco S, Cappello V, Vezzoli M, Rovere-Querini P, Moggio M, Ripolone M, Francolini M, Sandri M, Clementi E. (2012) Autophagy as a new therapeutic target in Duchenne muscular dystrophy. Cell Death Dis. 3:e418

De Palma C, Perrotta C, Pellegrino P, Clementi E, Cervia D. (2014) Skeletal muscle homeostasis in duchenne muscular dystrophy: modulating autophagy as a promising therapeutic strategy. Front Aging Neurosci. 6:188

Deak KL, Lemmers RJ, Stajich JM, Klooster R, Tawil R, Frants RR, Speer MC, van der Maarel SM, Gilbert JR. (2007) Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4. Neurology. 68(8):578-82

Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, Felicetti L. (1995) Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. Eur J Hum Genet. 3(3):155-67

Dey BK, Pfeifer K, Dutta A. (2014) The H19 long noncoding RNA gives rise to microRNAs miR-675-3p and miR-675-5p to promote skeletal muscle differentiation and re generation. Genes Dev. 28(5): 491-501

Dhir A, Dhir S, Proudfoot NJ, Jopling CL. (2015) Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. Nat Struct and Mol Biol. 22:319-327

Di Lazzaro, Pilato F, Dileone M, Profice P, Oliviero A, Mazzone P, Insola A, Ranieri F, Meglio M,Tonali PA, Rothwell JC. (2008) The physiological basis of the effects of intermittent theta burst stimulation of the human motor cortex. J Physiol. 586(Pt 16): 3871-3879

Dib C, Bou Saada Y, Dmitriev P, Richon C, Dessen P, Laoudj-Chenivesse D, Carnac G, Lipinski M, Vassetzky YS. (2016) Correction of the FSHD myoblast differentiation defect by fusion with healthy myoblasts. J Cell Physiol. 231(1):62-71

Dibble CC, Asara JM, Manning BD. (2009) Characterization of Rictor phosphorylation sites reveals direct regulation of mTOR complex 2 by S6K1. Mol Cell Biol. 29(21):5657-70

Dietrich S, Schubert FR, Healy C, Sharpe PT, Lumsden A. (1998). Specification of the hypaxial musculature. Development. 125:2235-2249

Dixit M, Ansseau E, Tassin A, Winokur S, Shi R, Qian H, Sauvage S, Mattéotti C, van Acker AM, Leo O, Figlewicz D, Barro M, Laoudj-Chenivesse D, Belayew A, Coppée F, Chen YW. (2007) DUX4, a candidate gene of facioscapulohumeral

muscular dystrophy, encodes a transcriptional activator of PITX1. Proc Natl Acad Sci U S A. 104(46):18157-62

Dmitriev P, Petrov A, Ansseau E, Stankevicins L, Charron S, Kim E, Bos TJ, Robert T, Turki A, Coppée F, Belayew A, Lazar V, Carnac G, Laoudj D, Lipinski M, Vassetzky YS. (2011) The Krüppel-like factor 15 as a molecular link between myogenic factors and a chromosome 4q transcriptional enhancer implicated in facioscapulohumeral dystrophy. J Biol Chem. 286(52):44620-31

Dmitriev P, Stankevicins L, Ansseau E, Petrov A, Barat A, Dessen P, Robert T, Turki A, Lazar V, Labourer E, Belayew A, Carnac G, Laoudj-Chenivesse D, Lipinski M, Vassetzky YS. (2013) Defective regulation of microRNA target genes in myoblasts from facioscapulohumeral dystrophy patients. J Biol Chem. 288(49):34989-5002

Dmitriev P, Kairov U, Robert T, Barat A, Lazar V, Carnac G, Laoudj-Chenivesse D, Vassetzky YS. (2014) Cancer-related genes in the transcription signature of facioscapulohumeral dystrophy myoblasts and myotubes. J Cell Mol Med. 18(2):208-17

Doherty JT, Lenhart KC, Cameron MV, Mack CP, Conlon FL, Taylor JM. (2011) Skeletal muscle differentiation and fusion are regulated by the BAR-containing Rho-GTPase-activating protein (Rho-GAP). J Biol Chem. 286(29):25903-21

Drozdetskiy A, Cole C, Procter J, Barton GJ. (2015) JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 43(W1):W389-94

Duret L , Chureau C , Samain S , Weissenbach J , Avner P. (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science. 312 (5780): 1653-5

Elisaphenko EA, Kolesnikov NN, Shevchenko AI, Rogozin IB, Nesterova TB, Brockdorff N, Zakian SM. (2008) A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. PLoS One. 3:e2521

Enwere EK, Holbrook J, Lejmi-Mrad R, Vineham J, Timusk K, Sivaraj B, et al. (2012) TWEAK and cIAP1 regulate myoblast fusion through the noncanonical NF-kappaB signaling pathway. Sci Signal. 5(246):ra75

Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA. (2004) B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat. Struct. Mol. Biol. 11:822-829

Fang Z, Rajewsky N. (2011) The impact of miRNA target sites in coding sequences and in 3'UTRs. PLoS One. 6(3):e18067

Fang Z, Du R, Edwards A, Flemington EK, Zhang K. (2013) The Sequence Structures of Human MicroRNA Molecules and Their Implications. PLoS ONE 8(1): e54215

Flanigan KM, Coffeen CM, Sexton L, Stauffer D, Brunner S, Leppert MF. (2001) Genetic characterization of a large, historically significant Utah kindred with facioscapulohumeral dystrophy. Neuromuscul Disord. 11(6-7):525-9

Flynt AS, Chung WJ, Greimann JC, Lima CD, Lai EC (2010) microRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol. Cell. 38:900-907

Fortier M, Comunale F, Kucharczak J, Blangy A, Charrasse S, Gauthier-Rouviere C. (2008) RhoE controls myoblast alignment prior fusion through RhoA and ROCK. Cell Death Differ. 15:1221-1231

Gabellini D, Green MR, Tupler R. (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell. 110(3):339-48

Gabellini D, D'Antona G, Moggio M, Prelle A, Zecca C, Adami R, Angeletti B, Ciscato P, Pellegrino MA, Bottinelli R, Green MR, Tupler R. (2006) Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature. 439(7079):973-7

Gabriëls J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, Padberg GW, Frants RR, Hewitt JE, Collen D, Belayew A. (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene. 236(1):25-32

Gagan J, Dey BK, Layer R, Yan Z, Dutta A. (2012) Notch3 and Mef2c proteins are mutually antagonistic via Mkp1 protein and miR-1/206 microRNAs in differentiating myoblasts. J Biol Chem. 287(48):40360-70

Galbiati F, Volonte D, Engelman JA, Scherer PE, Lisanti MP. (1999) Targeted down-regulation of caveolin-3 is sufficient to inhibit myotube formation in differentiating C2C12 myoblasts. Transient activation of p38 mitogen-activated protein kinase is required for induction of caveolin-3 expression and subsequent myotube formation. J Biol Chem. 274:30315-21

Galetta F, Franzoni F, Sposito R, Plantinga Y, Femia FR, Galluzzi F, Rocchi A, Santoro G, Siciliano G. (2005) Subclinical cardiac involvement in patients with facioscapulohumeral muscular dystrophy. Neuromuscul Disord. 15(6):403-8

Gayraud-Morel B, Chrétien F, Flamant P, Gomès D, Zammit PS, Tajbakhsh S (2007). A role for the myogenic determination gene Myf5 in adult regenerative myogenesis. Dev. Biol. 312:13-28

Ge Y, Wu A, Warnes C, Liu J, Zhang C, et al. (2009) mTOR regulates skeletal muscle regeneration in vivo through kinase-dependent and kinase-independent mechanisms. Am J Physiol Cell Physiol. 297(6):C1434–C1444

Ge Y., Sun Y., Chen J. (2011) IGF-II is regulated by microRNA-125b in skeletal myogenesis. J. Cell Biol. 192:69-81

Ge Y, Yoon MS, Chen J. (2011b) Raptor and Rheb negatively regulate skeletal myogenesis through suppression of insulin receptor substrate 1 (IRS1). J Biol Chem. 286:35675–35682.

Ge Y, Chen J. (2012) Mammalian target of rapamycin (mTOR) signaling network in skeletal myogenesis. J Biol Chem. 287(52):43928-35

Geng LN, Tyler AE, Tapscott SJ (2011) Immunodetection of human double homeobox 4. Hybridoma (Larchmt). 30(2):125-30

Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, Van der Maarel SM, Ruzzo WL, Gentleman RC, Tawil R, Tapscott SJ. (2012) DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev Cell. 22(1):38-51

Giussani M, Cardone MF, Bodega B, Ginelli E, Meneveri R. (2012) Evolutionary history of linked D4Z4 and Beta satellite clusters at the FSHD locus (4q35). Genomics. 100(5):289-96

Goljanek-Whysall K, Sweetman D, Abu-Elmagd M, Chapnik E, Dalmay T, Hornstein E, Münsterberg A. (2011) MicroRNA regulation of the paired-box transcription factor Pax3 confers robustness to developmental timing of myogenesis. Proc Natl Acad Sci U S A. (29):11936-41

Goljanek-Whysall K, Pais H, Rathjen T, Sweetman D, Dalmay T, Münsterberg A. (2012) Regulation of multiple target genes by miR-1 and miR-206 is pivotal for C2C12 myoblast differentiation. J Cell Sci. 125(Pt 15):3590-600

Gomes MD, Lecker SH, Jagoe RT, Navon A, Goldberg AL. (2001) Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy. Proc Natl Acad Sci U S A. 98(25):14440-5

Goodrich JA, Kugel JF. (2009) From bacteria to humans, chromatin to elongation, and activation to repression: The expanding roles of noncoding RNAs in regulating transcription. Crit Rev Biochem Mol Biol. 44(1):3–15

Gredinger E, Gerber AN, Tamir Y, Tapscott, S J and Bengal E. (1998) Mitogen-activated protein kinase pathway is involved in thedifferentiation of muscle cells. J. Biol. Chem. 273:10436-10444

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 458(7235):223-7

Habchi J, Martinho M, Gruet A, Guigliarelli B, Longhi S, Belle V. (2012) Monitoring structural transitions in IDPs by site-directed spin labeling EPR spectroscopy. Methods Mol Biol. 895:361-86

Hammond SM, Caudy AA, Hannon GJ. (2001) Post-transcriptional gene silencing by double-stranded RNA. Nat Rev Genet. 2(2):110-9.

Hasselgren PO. (1999) Glucocorticoids and muscle catabolism. Curr Opin Clin Nutr Metab Care. 2(3):201-5

Hayden MS, Ghosh S. (2004) Signaling to NF-kappaB. Genes Dev. 18(18):2195-224

He S, Su H, Liu C, Skogerbø G, He H, He D, Zhu X, Liu T, Zhao Y, Chen R. (2008) MicroRNA-encoding long non-coding RNAs. BMC Genomics. 9:236

He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. (2008) The antisense transcriptomes of human cells. Science 322:1855-1857.

Heo I, Ha M, Lim J, Yoon MJ, Park JE, Kwon SC, Chang H, Kim VN. (2012) Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell. 151:521-532

Hess B, Kutzner C, van der Spoel D, Lindahl E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4:435-447

Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, van Deutekom JC, Francis F, Sharpe PT, Hofker M, et al. (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet. 3(8):1287-95

Hobson-Webb LD, Caress JB. (2006) Facioscapulohumeral muscular dystrophy can be a cause of isolated childhood cognitive dysfunction. J Child Neurol. 21(3):252-3

Hofmann M, Schuster-Gossler K, Watabe-RudolphM, Aulehla A, Herrmann BG, Gossler A. (2004). WNT signaling, in synergy with T/TBX6, controls Notch signaling by regulating Dll1 expression in the presomitic mesoderm of mouse embryos. Genes Dev. 18:2712-2717

Homma S, Chen JC, Rahimov F, Beermann ML, Hanger K, Bibat GM, Wagner KR, Kunkel LM, Emerson CP Jr, Miller JB. (2012) A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: family, disease and cell function. Eur J Hum Genet. 20(4):404-10

Homma S, Beermann ML, Boyce FM, Miller JB. (2015) Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. Ann Clin Transl Neurol. 2(2):151-66

Horsley V, Pavlath GK. (2002) NFAT: ubiquitous regulator of cell differentiation and adaptation. J Cell Biol. 156(5):771-4

Horsley V, Pavlath GK. (2003) Cell fusion in skeletal muscle--central role of NFATC2 in regulating muscle cell size. Cell Cycle. 2(5):420-3

Hsu JB, Chiu C, Hsu S, Huang W, Chien C, Lee T, HuangEmail H. (2011) miRTar: an integrated system for identifying miRNA-target interactions in human. BMC Bioinformatics. 12:300

Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, et al. (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 47(7):621-9

Jaafar R, Zeiller C, Pirola L, Di Grazia A, Naro F, Vidal H, Lefai E, Némoz G. (2011) Phospholipase D regulates myogenic differentiation through the activation

of both mTORC1 and mTORC2 complexes. J Biol Chem. 2011 Jun 24;286(25):22609-21

Jiang G, Yang F, van Overveld PG, Vedanarayanan V, van der Maarel S, Ehrlich M. (2003) Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Hum Mol Genet. 12(22):2909-21

Jinju Han,1 Jakob S. Pedersen,2,5 S. Chul Kwon,1,5 Cassandra D. Belair,3,5 Young-Kook Kim,1 Kyu-Hyeon Yeom,1 Woo-Young Yang,1 David Haussler,4 Robert Blelloch,3 and V. Narry Kim1. (2009) Posttranscriptional Crossregulation between Drosha and DGCR8. Cell. 136(1):75-84

Jones NC, Fedorov YV, Rosenthal RS, Olwin BB, et al. (2001) ERK1/2 is required for myoblast proliferation but is dispensable for muscle gene expression and cell fusion. J. Cell. Physiol. 186:104-115

Jones TI, King OD, Himeda CL, Homma S, Chen JC, Beermann ML, Yan C, Emerson CP Jr, Miller JB, Wagner KR, Jones PL. (2015) Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. Clin Epigenetics. 7:37

Julien LA, Carriere A, Moreau J, Roux PP. (2010) mTORC1-activated S6K1 phosphorylates Rictor on threonine 1135 and regulates mTORC2 signaling. Mol Cell Biol. 30(4):908-21

Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, Lu L, Liu C, Yi JS, Zhang H, Min W, Bennett AM, Gregory RI, Ding Y, Huang Y. (2013) The imprinted H19 lncRNA antagonizes let-7 microRNAs. Mol Cell. 52(1):101-12

Kanduri C, Thakur N, Pandey RR. (2006) The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. EMBO J. 25:2096-2106

Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD, Whyte WA, Pereira CF, Brookes E, Walker K, Bell GW, Pombo A, Fisher AG, Young RA, Jenner RG. (2010) Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell. 38(5):675-88

Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, Mukai H, Gotoh T, Motoi T, Fukayama M, Aburatani H, Takizawa T, Nakamura T. (2006) Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet. 15(13):2125-37

Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, Smits G, Reik W. (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol. 14(7):659-65

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL. (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 106(28):11667-72

Khan NI, Bradstock KF, Bendall LJ. (2007) Activation of Wnt/beta-catenin pathway mediates growth and survival in B-cell progenitor acute lymphoblastic leukaemia. Br J Haematol. 138(3):338-48.

Kilmer DD, Abresch RT, McCrory MA, Carter GT, Fowler WM Jr, Johnson ER, McDonald CM. (1995) Profiles of neuromuscular diseases. Facioscapulohumeral muscular dystrophy. Am J Phys Med Rehabil. 74(5 Suppl):S131-9

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM., et al. (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465: 182-187

Klinge L, Eagle M, Haggerty ID, Roberts CE, Straub V, Bushby KM. (2006) Severe phenotype in infantile facioscapulohumeral muscular dystrophy. Neuromuscul Disord. 16(9-10):553-8

Klooster R, Straasheijm K, Shah B, Sowden J, Frants R, Thornton C, Tawil R, van der Maarel S. (2009) Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. Eur J Hum Genet. 17(12):1615-24

Kotake Y, Nakagawa T, Kitagawa K, Suzuki S, Liu N, Kitagawa M, Xiong Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. Oncogene. 30:1956-1962

Kowalczyk AP, Nanes BA. (2012) Adherens junction turnover: regulating adhesion through cadherin endocytosis, degradation, and recycling. Subcell Biochem. 2012;60:197-222

Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Mattéotti C, Arias C, Corona ED, Nuñez NG, Leo O, Wattiez R, Figlewicz D, Laoudj-Chenivesse D, Belayew A, Coppée F, Rosa AL. (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 17(8):611-23

Kramerova I, Kudryashova E, Wu B, Spencer MJ. (2006) Regulation of the M-cadherin-beta-catenin complex by calpain 3 during terminal stages of myogenic differentiation. Mol Cell Biol. 26(22):8437-47

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology. 305(3):567-580

Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ. (2004) Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. J Biol Chem. 279(40):42230-9

Kuleesha Y, Puah WC2, Wasser M2. (2016) A model of muscle atrophy based on live microscopy of muscle remodelling in Drosophila metamorphosis. R Soc Open Sci. 3(2):150517

Langley B, Thomas M, Bishop A, Sharma M, Gilmour S, Kambadur R. (2002) Myostatin inhibits myoblast differentiation by down-regulating MyoD expression. J. Biol. Chem. 277:49831-49840

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, et al. (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell. 97:17-27.

Lanz RB., Razani B, Goldberg AD, O'Malley BW. (2002) Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). Proc. Natl. Acad. Sci. USA. 99:16081-16086

Laplante M, Sabatini DM. (2012) mTOR signaling in growth control and disease. Cell. 149(2):274-93

Lechner C, Zahalka MA, Giot JF, et al. (1996) ERK6, a mitogen-activated protein kinase involved in C2C12 myoblast differentiation. Proc Natl Acad Sci USA. 93:4355-4359

Lee JT, Davidow LS, Warshawsky D. (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet. 21(4):400-4

Leidenroth A and Hewitt JE. (2010) A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. BMC Evol Biol. 10:364

Lemmers RJ, de Kievit P, Sandkuijl L, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. (2002) Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet. 32(2):235-6

Lemmers RJ, Wohlgemuth M, van der Gaag KJ, van der Vliet PJ, van Teijlingen CM, de Knijff P, Padberg GW, Frants RR, van der Maarel SM. (2007) Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet. 81(5):884-94

Lemmers RJ, van der Vliet PJ, van der Gaag KJ, Zuniga S, Frants RR, de Knijff P, van der Maarel SM. (2010) Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. Am J Hum Genet. 86(3):364-77

Lemmers RJ, Tawil R, Petek LM, Balog J, Block GJ, Santen GW, Amell AM, van der Vliet PJ, Almomani R, et al. (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 44(12):1370-4

Lemmers RJ, Goeman JJ, van der Vliet PJ, van Nieuwenhuizen MP, Balog J, Vos-Versteeg M, Camano P, Ramos Arroyo MA, Jerico I, Rogers MT, Miller DG, Upadhyaya M, et al. (2014) Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Hum Mol Genet. 24:659-669

Li H, Malhotra S, Kumar AJ. (2008) Nuclear factor-kappa B signaling in skeletal muscle atrophy. J Mol Med (Berl). 86(10):1113-26.

Li Z, Jiang Y, Ulevitch RJ, Han J. The primary structure of p38 gamma: a new member of p38 group of MAP kinases. Biochemical and Biophysical Research Communications. 228(2):334-40

Liang Y, Ridzon D, Wong L, Chen C. (2007) Characterization of microRNA expression profiles in normal human tissues. BMC Genomics. 8:166

Lim JW, Snider L, Yao Z, Tawil R, Van Der Maarel SM, Rigo F, Bennett CF, Filippova GN, Tapscott SJ. (2015) DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD. Hum Mol Genet. 24(17):4817-28

Lin Z, Murtaza I, Wang K, Jiao J, Gao J, Li PF. (2009) miR-23a functions downstream of NFATc3 to regulate cardiac hypertrophy. Proc Natl Acad Sci U S A. 106:12103-12108.

Ling KH, Brautigan PJ, Moore S, Fraser R, Leong MP, Leong JW, Zainal AS, et al. (2016) In depth analysis of the Sox4 gene locus that consists of sense and natural antisense transcripts. Data Brief. 7:282-90

Liu X, He S, Skogerbø G, Gong F, Chen R. (2012) Integrated sequence-structure motifs suffice to identify microRNA precursors. PLoS One. 7(3):e32797

Londhe P, Davie JK (2011) Sequential association of myogenic regulatory factors and E proteins at muscle-specific genes. Skeletal Muscle. 1:14

Louro R, Smirnova AS, Verjovski-Almeida S. (2009). Long intronic noncoding RNA transcription: Expression noise or expression choice? Genomics. 93:291-298.

Lu L, Sun K, Chen X, Zhao Y, Wang L, Zhou L, Sun H, Wang H. (2013) Genome-wide survey by ChIP-seq reveals YY1 regulation of lincRNAs in skeletal myogenesis. EMBO J. 32:2575-2588

Lupas A, Van Dyke M, Stock J. (1991) Predicting Coiled Coils from Protein Sequences. Science. 252:1162-1164

Lyle R, Wright TJ, Clark LN, Hewitt JE. (1995) The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics. 28(3):389-97

Lyle R, Watanabe D, Te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP. (2000) The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. Nat Genet. 25:19-21

Ma XM, Blenis J. (2009) Molecular mechanisms of mTOR-mediated translational control. Nat Rev Mol Cell Biol. 10(5):307-18

Madaro L, Marrocco V, Fiore P, Aulino P, Smeriglio P, Adamo S, Molinaro M, Bouché M, Mark H. Ginsberg, Monitoring Editor University of California, San Diego. (2011) PKCα signaling is required for myoblast fusion by regulating the

expression of caveolin-3 and β1D integrin upstream focal adhesion kinase. Mol Biol Cell. 22(8):1409-1419

Manning BD, Cantley LC. (2003) Rheb fills a GAP between TSC and TOR. Trends Biochem Sci. 28(11):573-6.

Marbach-Bar N, Bahat A, Ashkenazi S, Golan-Mashiach M, Haimov, Wu SY, et al. (2015) DTIE, a novel core promoter element that directs start site selection in TATA-less genes. Nucleic Acids Res. 44(3):1080-94

Marzi MJ, Puggioni EM, Dall'Olio V, Bucci G, Bernard L, Bianchi F, Crescenzi M, Di Fiore PP, Nicassio F. (2012) Differentiation-associated microRNAs antagonize the Rb-E2F pathway to restrict proliferation. J Cell Biol. 199(1):77-95

Masny PS, Chan OY, de Greef JC, Bengtsson U, Ehrlich M, Tawil R, Lock LF, Hewitt JE, Stocksdale J, Martin JH, Van Der Maarel SM, Winokur ST. (2010) Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. Eur J Hum Genet. 18(4):448-56.

Melamed Z, Levy A, Ashwal-Fluss R, Lev-Maor G, Mekahel K, Atias N, Gilad S, Sharan R, Levy C, Kadener S, Ast G. (2013) Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. Mol Cell. 50(6):869-81

Melamed ZE, et al. (2013) Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. Mol Cell. 50:869-881

Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE. (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. 6(12):e1001244

Meneveri R, Agresti A, Marozzi A, Saccone S, Rocchi M, Archidiacono N, Corneo G, Della Valle G, Ginelli E. (1993) Molecular organization and chromosomal location of human GC-rich heterochromatic blocks. Gene. 123(2):227-34

Meriane M, Roux P, Primig M, Fort P, Gauthier-Rouviere C. (2000) Critical activities of Rac1 and Cdc42Hs in skeletal myogenesis: antagonistic effects of JNK and p38 pathways. Mol. Biol. Cell. 11:2513-2528

Mihaly J, Mishra RK, Karch F. (1998) A conserved sequence motif in Polycomb-response elements. Mol Cell. 1(7):1065-6

Minetti C, Sotgia F, Bruno C, Scartezzini P, Broda P, Bado M, Masetti E, et al (1998) Mutations in the caveolin-3 gene cause autosomal dominant limb - girdle muscular dystrophy. Nat Genet. 18:365-368

Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I.(2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell. 126(6):1203-17

Mitch WE, Goldberg AL. (1996) Mechanisms of muscle wasting. The role of the ubiquitin-proteasome pathway. N Engl J Med. 335(25):1897-905.

Mitsuhashi H, Mitsuhashi S, Lynn-Jones T, Kawahara G, Kunkel LM. (2013) Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. Hum Mol Genet. 22(3):568-77

Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. 16(6):720-8

Neguembor MV, Gabellini D. (2010) In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy. Epigenomics. 2(2):271-87

Olson EN, Sternberg E, Hu JS, Spizz G, Wilcox C. (1986). Regulation of myogenic differentiation by type beta transforming growth factor. J. Cell Biol. 103:1799-1805

Osato N, Suzuki Y, Ikeo K, Gojobori T (2007) Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics. 176:1299-1306

Osborne RJ, Welle S, Venance SL, Thornton CA, Tawil R. (2007) Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. Neurology. 68(8):569-77

Padberg GW, Brouwer OF, de Keizer RJ, Dijkman G, Wijmenga C, Grote JJ, Frants RR. (1995) On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. Muscle Nerve Suppl. (2):S73-80

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, et al. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell. 32: 232-246.

Pansters NA, Van Der Velden JL, Kelders MC, Laeremans H, Schols AM, Langen RC. (2011) Segregation of myoblast fusion and muscle-specific gene expression by distinct ligand-dependent inactivation of GSK-3ß. Cell Mol. Life Sci. 68:523-535

Penn BH, Bergstrom DA, Dilworth FJ, Bengal E, Tapscott SJ. (2004) A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. Genes Dev. 18:2348-2353

Pink RC, Wicks K, Caley DP,Punch EK, Jacobs L, Carter DRF. (2011) Pseudogenes: Pseudo-functional or key regulators in health and disease? RNA. 7(5): 792-798

Pisconti A, Brunelli S, Di Padova M, De Palma C, Deponti D, Baesso S, Sartorelli V, Cossu G, Clementi E. (2006) Follistatin induction by nitric oxide through cyclic GMP: a tightly regulated signaling pathway that controls myoblast fusion. J Cell Biol. 172(2):233-44.

Pohlers M, Calabrese JM, Magnuson T. (2014) Small RNA expression from Human Macrosatellite. G3 (Bathesda). 4(10):1981-1989

Polikepahad S, Corry DB (2013) Profiling of T helper cell-derived small RNAs reveals unique antisense transcripts and differential association of miRNAs with argonaute proteins 1 and 2. Nucleic Acids Res. 41(2):1164-1177

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465(7301):1033-8

Quach NL, Biressi S, Reichardt LF, Keller C, Rando TA. (200) Focal adhesion kinase signaling regulates the expression of caveolin 3 and beta1 integrin, genes essential for normal myoblast fusion. Mol. Biol. Cell. 20:3422-3435

Rahimov F, King OD, Leung DG, Bibat GM, Emerson CP Jr, Kunkel LM, Wagner KR. (2012) Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. Proc Nat Acad Sci U S A. 109(40):16234-9

Rantanen J, Hurme T, Lukka R, Heino J, Kalimo H. (1995) Satellite cell proliferation and the expression of myogenin anddesmin in regenerating skeletal muscle : evidence for two differentpopulations of satellite cells. Laboratory Investigation. 72,341-347

Reed BJ, Locke MN, Gardner RG. (2015) A Conserved Deubiquitinating Enzyme Uses Intrinsically Disordered Regions to Scaffold Multiple Protein Interaction Sites. J Biol Chem. 290(33):20601-12

Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. (2004) Fast and effective prediction of microRNA/target duplexes. RNA. 10(10):1507-17

Reynolds TH , Bodine SC, Lawrence JC Jr. (2002) Control of Ser2448 phosphorylation in the mammalian target of rapamycin by insulin and skeletal muscle load. J Biol Chem. 277(20):17657-62

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 129(7):1311-23

Ripoche MA, Kress C, Poirier F, Dandolo L. (1997) Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. Genes & Dev. 11:1596-1604

Rocha ST, Edwards CA, Ito M, Ogata T, Ferguson-Smith AC. (2008) Genomic imprinting at the mammalian Dlk1-Dio3 domain. Trends Genet. 24(6):306-16

Rochat A, Fernandez A, Vandromme M, Moles JP, Bouschet T, Carnac G, Lamb NJ. (2004) Insulin and wnt1 pathways cooperate to induce reserve cell activation in differentiation and myotube hypertrophy. Mol Biol Cell. 15:4544-4555.

Rochlin K, Yu S, Roy S, Baylies MK. (2010) Myoblast fusion: when it takes more to make one. Dev Biol. 341:66-83

Rommel C, Bodine SC, Clarke BA, Rossman R, Nunez L, Stitt TN, Yancopoulos GD, Glass DJ. (2001) Mediation of IGF-1-induced skeletal myotube hypertrophy by PI(3)K/Akt/mTOR and PI(3)K/Akt/GSK3 pathways. Nat Cell Biol. 3(11):1009-13.

Rory Johnson and Roderic Guigó (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. RNA. 20(7):959-976

Rothberg JM1, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature. 475(7356):348-52

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell. 127(6):1193-207

Saccone V, Puri PL. (2010) Epigenetic regulation of skeletal myogenesis. Organogenesis. 6(1):48-53

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat. Rev. Genet. 8:424-436

Schaller MD. (2010) Cellular functions of FAK kinases: insight into molecular mechanisms and novel functions. J Cell Sci. 123(Pt 7):1007-13

Schmidt H, Taubert H, Lange H, Kriese K, Schmitt WD, Hoffmann S, Bartel F, Hauptmann S. (2009) Small polydispersed circular DNA contains strains of mobile genetic elements and occurs more frequently in permanent cell lines of malignant tumors than in normal lymphocytes. Oncol Rep. 22(2):393-400

Schnell JR, Dyson HJ, Wright PE. (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase. Annu Rev Biophys Biomol Struct. 33:119-140.

Schwander M, Leu M, Stumm M, Dorchies OM, Ruegg UT, Schittny J, Muller U. (2003) Beta1 integrins regulate myoblast fusion and sarcomere assembly. Dev Cell. 4:673-685

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008) Divergent transcription from active promoters. Science 322:1849-1851

Seitz H, Ghildiyal M, Zamore PD. (2008) Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. Curr Biol. 22;18(2):147-51

Shih SC, Prag G, Francis SA, Sutanto MA, Hurley JH, Hicke L. (2003) A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. EMBO J. 22(6):1273–1281

Smith CK, Janney MJ, Allen RE. (1994) Temporal expression of myogenic regulatory genes during activation, proliferation, and differentiation of rat skeletal muscle satellite cells. J Cell Physiol. 159:379-385

Snider L, Asawachaicharn A, Tyler AE, Geng LN, Petek LM, Maves L, Miller DG, Lemmers RJ, Winokur ST, Tawil R, Van Der Maarel SM, Filippova GN, Tapscott SJ. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Hum Mol Genet.18(13):2414-30

Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, et al. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. 6(10):e1001181

Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, Nakagawa S. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J. Cell Sci. 120:2498-2506

Sun Y, Ge Y, Drnevich J, Zhao Y, Band M, Chen J. (2012) Mammalian target of rapamycin regulates miRNA–1 and follistatin in skeletal myogenesis. J Cell Biol. 189:1157-1169.

Sunadome K, Yamamoto T, Ebisuya M, Kondoh K, Sehara-Fujisawa A, Nishida E. (2011) ERK5 regulates muscle cell fusion through Klf transcription factors. Dev. Cell. 20:192-205.

Tassin A, Leroy B, Laoudj-Chenivesse D, Wauters A, Vanderplanck C, Le Bihan MC, Coppée F, Wattiez R, Belayew A. (2012) FSHD myotubes with different phenotypes exhibit distinct proteomes. PLoS One. 7(12):e51865

Tassin A, Laoudj-Chenivesse D, Vanderplanck C, Barro M, Charron S, Ansseau E, Chen YW, Mercier J, Coppée F, Belayew A. (2013) DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? J Cell Mol Med. 17(1):76-89

Thomas NS, Wiseman K, Spurlock G, MacDonald M, Ustek D, Upadhyaya M. (2007) A large patient study confirming that facioscapulohumeral muscular dystrophy (FSHD) disease expression is almost exclusively associated with an FSHD locus located on a 4qA-defined 4qter subtelomere. J Med Genet. 44(3):215-8

Tonini MM, Passos-Bueno MR, Cerqueira A, Matioli SR, Pavanello R, Zatz M. (2004) Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromuscul Disord. 14(1):33-8

Tran H, Polakis P. (2012) Reversible modification of adenomatous polyposis coli (APC) with K63-linked polyubiquitin regulates the assembly and activity of the ß-catenin destruction complex. J Biol Chem. 287(34):28552-63

Trendelenburg AU, Meyer A, Jacobi C, Feige JN, Glass DJ. (2012) TAK-1/p38/nNFkappaB signaling inhibits myoblast differentiation by increasing levels of Activin A. Skelet. Muscle 2(1):3

Trendelenburg AU1, Meyer A, Rohner D, Boyle J, Hatakeyama S, Glass DJ (2009) Myostatin reduces Akt/TORC1/p70S6K signaling, inhibiting myoblast differentiation and myotube size. Am J Physiol Cell Physiol. 296(6):C1258-70

Trevisan CP, Pastorello E, Armani M, Angelini C, Nante G, Tomelleri G, Tonin P, Mongini T, Palmucci L, Galluzzi G, Tupler RG, Barchitta A. (2006) Facioscapulohumeral muscular dystrophy and occurrence of heart arrhythmia. Eur Neurol. 56(1):1-5

Tsuji M, Kinoshita M, Imai Y, Kawamoto M, Kohara N. (2009) Facioscapulohumeral muscular dystrophy presenting with hypertrophic cardiomyopathy: a case study. Neuromuscul Disord. 19(2):140-2

Tsuji O, Miura K, Fujiyoshi K, Momoshima S, Nakamura M, Okano H (2011) Cell therapy for spinal cord injury by neural stem/progenitor cells derived from iPS/ES cells. Neurotherapeutics. 8(4):668-76

Turki A, Hayot M, Carnac G, Pillard F, Passerieux E, Bommart S, Raynaud de Mauverger E, Hugon G, Pincemail J, Pietri S, Lambert K, Belayew A, Vassetzky Y, Juntas Morales R, Mercier J, Laoudj-Chenivesse D. (2012) Functional muscle impairment in facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. Free Radic Biol Med. 53(5):1068-79

Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147(7):1537-50

Ustanina S, Carvajal J, Rigby P, Braun T. (2007) The myogenic factor Myf5 supports efficient skeletal muscle regeneration by enabling transient myoblast amplification. Stem Cells. 25:2006-2016

Van Der Maarel SM, Tawil R, Tapscott SJ. (2011) Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol Med. 17(5):252-8

Van Geel M, Dickson MC, Beck AF, Bolland DJ, Frants RR, van der Maarel SM, de Jong PJ, Hewitt JE. (2002) Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics. 79(2):210-7

Van Koningsbruggen S, Straasheijm KR, Sterrenburg E, de Graaf N, Dauwerse HG, Frants RR, van der Maarel SM. (2007) FRG1P-mediated aggregation of proteins involved in pre-mRNA processing. Chromosoma. 116(1):53-64

Van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. (2003) Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet. 35(4):315-7

Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE. (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chem Rev. 114(13):6733-78

Vanderplanck C1, Ansseau E, Charron S, Stricwant N, Tassin A, Laoudj-Chenivesse D, Wilton SD, Coppée F, Belayew A. (2011) The FSHD atrophic myotube phenotype is caused by DUX4 expression. PLoS One. 6(10):e26820

Vejnar CE, Blum M, Zdobnov EM. (2013) miRmap web: Comprehensive microRNA target prediction online. Nucleic Acids Res. 41:W165-8

Von Maltzahn J, Renaud JM, Parise G, Rudnicki MA. (2012) Wnt7a treatment ameliorates muscular dystrophy. Proc. Natl. Acad. Sci. USA. 109:20614-20619

Wada S, Kato Y, Okutsu M, Miyaki S, Suzuki K, Yan Z, Schiaffino S, Asahara H, Ushida T, Akimoto T. (2011) Translational suppression of atrophic regulators by microRNA-23a integrates resistance to skeletal muscle atrophy. J Biol Chem. 286(44):38456-65

Wang T, Diaz-Rosales P, Costa MM, Campbell S, Snow M, Collet B, Martin SA, Secombes CJ. (2011) Functional characterization of a nonmammalian IL-21: rainbow trout Oncorhynchus mykiss IL-21 upregulates the expression of the Th cell signature cytokines IFN-gamma, IL-10, and IL-22. J. Immunol. 186:708-721

Watts R, Johnsen VL, Shearer J, Hittel DS. (2013) Myostatin-induced inhibition of the long noncoding RNA Malat1 is associated with decreased myogenesis. Am J Physiol Cell Physiol. 304(10):C995-1001

Wei L, Zhou W, Croissant JD, Johansen FE, Prywes R, Balasubramanyam A, Schwartz RJ. (1998) RhoA signaling via serum response factor plays an obligatory role in myogenic differentiation. J. Biol. Chem. 273:30287-30294

Wei N, Deng XW. (1998) Characterization and purification of the mammalian COP9 complex, a conserved nuclear regulator initially identified as a repressor of photomorphogenesis in higher plants. Photochem Photobiol. 68(2):237-241.

Werner A, Cockell S, Falconer J, Carlile M, Alnumeir S, Robinson J. (2014) Contribution of natural antisense transcription to an endogenous siRNA signature in human cells. BMC Genomics 15:19

Wiederstein M, Sippl MJ. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35:W407-W410

Wijmenga C, Van Deutekom JC, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR. (1994) Pulsed-field gel electrophoresis of the D4F104S1 locus reveals the size and the parental origin of the facioscapulohumeral muscular dystrophy (FSHD)-associated deletions. Genomics. 19(1):21-6

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science. 309(5740):1570-3.

Wilson RA, Liu J, Xu L, Annis J, Helmig S, Moore G, et al. (2016) Negative regulation of initial steps in skeletal myogenesis by mTOR and other kinases. Sci Rep. 6:20376

Winokur ST, Bengtsson U, Feddersen J, Mathews KD, Weiffenbach B, Bailey H, Markovich RP, Murray JC, Wasmuth JJ, Altherr MR, et al. (1994) The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Res. (3):225-34

Winokur ST, Chen YW, Masny PS, Martin JH, Ehmsen JT, Tapscott SJ, van der Maarel SM, Hayashi Y, Flanigan KM. (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. um Mol Genet. 12(22):2895-907

Wright TJ, Wijmenga C, Clark LN, Frants RR, Williamson R, Hewitt JE. (1993) Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. Hum Mol Genet. 2(10):1673-8

Wu AL, Kim JH, Zhang C, Unterman TG, Chen J. (2008) Forkhead box protein O1 negatively regulates skeletal myocyte differentiation through degradation of mammalian target of rapamycin pathway components. Endocrinology 149:1407-1414

Wu H, Peisley A, Graef IA, Crabtree GR. (2007) NFAT signaling and the invention of vertebrates. Trends Cell Biol. 17(6):251-60

Wu H, Ye C, Ramirez D, Manjunath N. (2009) Alternative processing of primary microRNA transcripts by Drosha generates 5'end variation of mature microRNA. PLoS One. 4(10):e7566

Wu Z, Woodring PJ, Bhakta KS,Tamura K,Wen F, Feramisco JR, Karin M, Wang JY, Puri PL. (2000) p38 and extracellular signal-regulated kinases regulate the myogenic program at multiple steps. Mol. Cell. Biol. 20:3951-3964

Yang F, Shao C, Vedanarayanan V, Ehrlich M. (2004) Cytogenetic and immuno-FISH analysis of the 4q subtelomeric region, which is associated with facioscapulohumeral muscular dystrophy. Chromosoma. 112(7):350-9

Yang S, Maurin T, Lai EC. (2012) Functional parameters of Dicer-independent microRNA biogenesis. RNA. 18(5):945-957

Yap KL1, Li S, Muñoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, Zhou MM. (2010) Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell. 38(5):662-74

Zatz M, Marie SK, Cerqueira A, Vainzof M, Pavanello RC, Passos-Bueno MR. (1998) The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. Am J Med Genet. 77(2):155-61

Zeng X, Tamai K, Doble B, Li S, Huang H, Habas R, Okamura H, Woodgett J, He X. (2005) A dual-kinase mechanism for Wnt co-receptor phosphorylation and activation. Nature. 438:873-877

Zeschnigk M, Kozian D, Kuch C, Schmoll M, Starzinski-Powitz A. (1995) Involvement of M-cadherin in terminal differentiation of skeletal muscle cells. J Cell Sci. 108(Pt 9):2973-81

Zhang J, McCastlain K, Yoshihara H, Xu B, Chang Y, Churchman ML, Wu G, et al. (2016) Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. Nat Genet. 2016 Dec;48(12):1481-1489

Zhang LF, Huynh KD, Lee JT. (2007) Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. Cell. 129:693-706

Zhang Y, Lee JK, Toso EA, Lee JS, Choi SH, Slattery M, Aihara H, Kyba M. (2016) DNA-binding sequence specificity of DUX4. Skelet Muscle. 6:8

Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT. (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell. 40(6):939-53

Zynoviev A, Morozova N, Nonne N, Barillot E, Harel-Bellan A, Gorban AN. (2010) Dynamical modeling of microRNA action on the protein translation process. BMC System Biol. 4:13