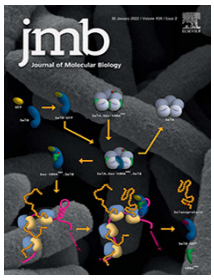

Lnc-SMaRT Translational Regulation of Spire1, A New Player in Muscle Differentiation

Silvia Scalzitti [1†], Davide Mariani [2†], Adriano Setti [1], Alessio Colantoni [3], Michela Lisi [3], Irene Bozzoni [1,2,3] and Julie Martone [4*]

*1 - **Department of Biology and Biotechnology** "**Charles Darwin**" **– Sapienza University of Rome,** Rome, Italy*
*2 - **Center for Human Technologies@Istituto Italiano di Tecnologia,** Genoa, Italy*
*3 - **Center for Life Nano- & Neuro-Science,** Fondazione Istituto Italiano di Tecnologia (IIT), Rome, Italy*
*4 - **Institute of Molecular Biology and Pathology,** National Research Council, Sapienza University of Rome, Rome, Italy*

Correspondence to Julie Martone: Sapienza University of Rome, Department of Biology and Biotechnology, CU026, Piazzale Aldo Moro 5, 00185 Rome (RM), Italy. *julie.martone@cnr.it (J. Martone)* *@Dav_MarPhD (D. Mariani)*
https://doi.org/10.1016/j.jmb.2021.167384
Edited by M. Yaniv

Abstract

The destiny of a messenger RNA is determined from a combination of *in cis* elements, like peculiar secondary structures, and *in trans* modulators, such as RNA binding proteins and non-coding, regulatory RNAs. RNA guanine quadruplexes belong to the first group: these strong secondary structures have been characterized in many mRNAs, and their stabilization or unwinding provides an additional step for the fine tuning of mRNA stability and translation. On the other hand, many cytoplasmic long non-coding RNAs intervene in post-transcriptional regulation, frequently by direct base-pairing with their mRNA targets. We have previously identified the lncRNA SMaRT as a key modulator of the correct timing of murine skeletal muscle differentiation; when expressed, lnc-SMaRT interacts with a G-quadruplex-containing region of Mlx-γ mRNA, therefore inhibiting its translation by counteracting the DHX36 helicase activity. The "smart" mode of action of lnc-SMaRT led us to speculate whether this molecular mechanism could be extended to other targets and conserved in other species. Here, we show that the molecular complex composed by lnc-SMaRT and DHX36 also includes other mRNAs. We prove that lnc-SMaRT is able to repress Spire1 translation through base-pairing with its G-quadruplex-forming sequence, and that Spire1 modulation participates to the regulation of proper skeletal muscle differentiation. Moreover, we demonstrate that the interaction between DHX36 and lnc-SMaRT is indirect and mediated by the mRNAs present in the complex. Finally, we suggest an extendibility of the molecular mechanism of lnc-SMaRT from the mouse model to humans, identifying potential functional analogues.

Introduction

Our perception of the role of RNA in the flow of genetic information has varied over time, especially after the discovery of a high amount of non-coding RNAs transcribed from the genomes of all eukaryotes.[1] Among them, long non-coding RNAs (lncRNAs) represent an eccentric class of transcripts, raising doubts whether these could simply be "noisy" and irrelevant. Several studies answered this question demonstrating that lncRNAs are characterized by very restricted and specific spatio-temporal expression patterns,[2,3] and that a signature of purifying selection supports

their functionality.[4–6] Over the years, diverse roles have been described and defined for these molecular species, which are involved in all the different layers of the regulation of gene expression.[7–9]

As of now, lncRNAs are classified in archetypes according to the molecular functions that they execute and to their nuclear or cytoplasmic subcellular localization.[10,11,8] The regulation of gene expression exerted by lncRNAs is particularly evident in complex tissues such as skeletal muscle: here, a wide population of these transcripts is responsible for the fine-tuning of myoblast proliferation and of their fusion in terminally differentiated myotubes.[12]

In a previous paper, we identified lnc-SMaRT (Skeletal Muscle Regulator of Translation), a murine muscle-specific intergenic long non-coding RNA involved in *in vitro* myoblast differentiation.[13] We demonstrated the role of lnc-SMaRT in the translational repression of the Mlx-γ mRNA isoform (MAX-like protein X). This regulation is mediated by the direct base-pairing between lnc-SMaRT and the Mlx-γ transcript at the level of a sequence able to form a G-quadruplex, a strong non-canonical secondary structure involved in the control of mRNA stability and translation.[14] We demonstrated that lnc-SMaRT antagonizes the activity of DHX36, a DEAH-box RNA helicase, that binds and resolves the G-quadruplex localized on Mlx-γ mRNA, thus allowing its translation. When lnc-SMaRT is expressed, during a specific and restricted time-window, it interacts with the solved G-quadruplex and represses Mlx-γ mRNA translation. Remarkably, the inhibition of Mlx-γ translation, and therefore the reduced abundance of its protein, has repercussions on the subcellular localization of the other two Mlx isoforms (α and β), which in the absence of Mlx-γ are unable to translocate into the nucleus and to activate their pro-differentiating muscle targets.[15,13]

Here we show that the regulatory mechanism involving lnc-SMaRT and DHX36 factors can be extended to at least another mRNA, the actin nucleation factor Spire1. Moreover, we obtain new insights into the lnc-SMaRT mechanism of action, showing that the interaction between the lncRNA and DHX36 is indirect and potentially mediated by the interacting mRNAs. Finally, through DHX36 RNA immunoprecipitation and subsequent sequencing, we identify some lncRNAs that could exert the same function of murine lnc-SMaRT during human myoblast differentiation.

Results

DHX36 helicase and lnc-SMaRT interact with several mRNAs

The previously identified lnc-SMaRT mRNA interactome, obtained by Next Generation Sequencing upon lnc-SMaRT pull down experiment,[13] was confirmed by qPCR. To extend the G-quadruplex-mediated function of the lnc-SMaRT/DHX36 regulatory complex to other mRNA targets, 12 protein-coding genes, characterized by the presence of at least one predicted G-quadruplex with a G-score higher than 35 and with at least three Guanines in tetrads (QGRS Mapper software)[16], were selected for validation assays (see also[13] and Supplementary Table 1). The experiment was performed on extracts derived from C2C12 cells 2 days after the switch to differentiation medium, when the expression of lnc-SMaRT reaches its peak during myogenesis. The interaction was confirmed for 6 out of 12 candidates (Figure 1 (A) and Supplementary Figure 1), using two different sets of biotin-labeled DNA antisense oligonucleotides against lnc-SMaRT (Set-1 and Set-2), while a set of antisense oligonucleotides against LacZ was used as negative control (LacZ). Only the mRNAs recovered with both sets of specific probes were considered as positive (Acad8, Spire1, Ndrg4, Arf5, Rbm45 and Mlx). DHX36 presence in the SMaRT pull-down protein fraction was also checked: as shown in the Western Blot in the lower panel of Figure 1(A), the helicase protein was correctly enriched. Lnc-SMaRT is shown as the positive RNA control of the pull-down experiment.

The presence of Acad8, Spire1, Ndrg4, Arf5, Rbm45 and Mlx (the latter is used as positive control together with Wbp4, a known interactor of DHX36[17]) in RNA samples, obtained upon DHX36 RIP assay, was also assessed (Figure 1(B)). Interestingly, all the 6 selected mRNAs were present among DHX36 RIP interactors, suggesting that they could be part of the same regulative molecular complex, previously identified for Mlx-γ mRNA, composed by lnc-SMaRT, DHX36 helicase and the specific mRNA.

DHX36, lnc-SMaRT, ACAD8, Spire1 molecular complex characterization

Based on the molecular mechanism identified for Mlx mRNA, we focused only on those candidates for which, after performing bioinformatic analysis using the IntaRNA 2.3.0,[18] the predicted base-pairing region with lnc-SMaRT overlapped with a G-quadruplex-forming sequence (Supplementary Figure 2(A) and (B)). All candidates, except for Ndrg4, were characterized by this feature.

The role of lnc-SMaRT in mediating the binding between DHX36 and mRNA candidates was verified by DHX36 RIP experiment performed in the presence (siSCR) or absence of lnc-SMaRT (siSMaRT), in C2C12 cells at day 2 of differentiation (DM2). Accordingly with the previous result obtained for Mlx,[13] we were able to demonstrate that no changes in the enrichment of the selected mRNAs occurred upon lnc-SMaRT depletion, suggesting that DHX36 can bind those mRNAs, likely due to their G-quadruplexes, in a way that is independent from lnc-SMaRT (Figure 2 (A) and Supplementary Figure 2(F)).

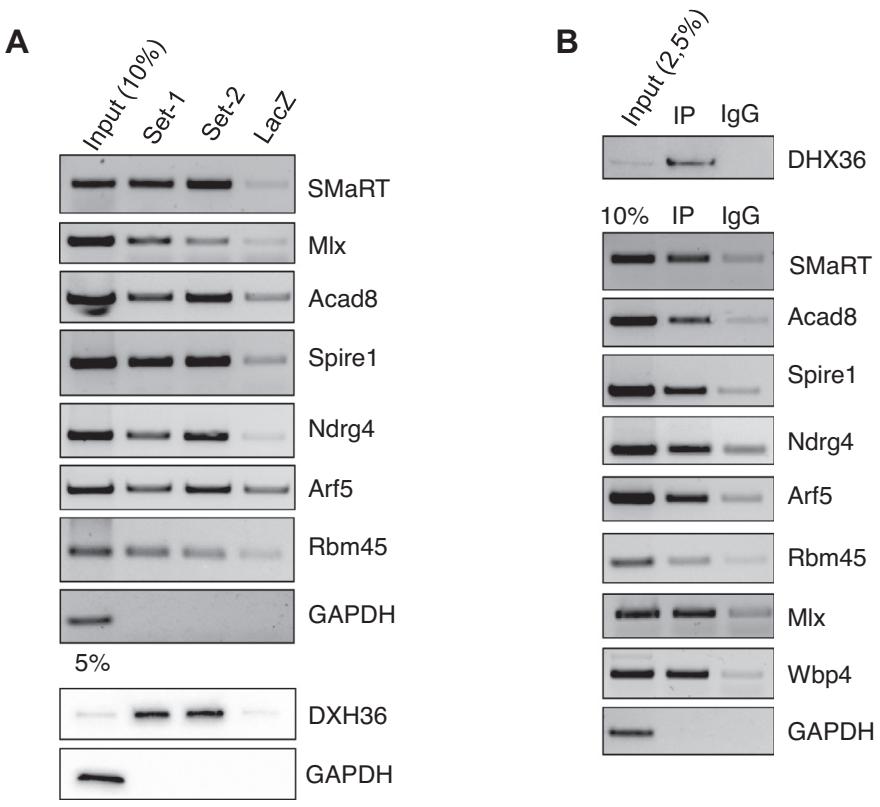

Figure 1. lnc-SMaRT and DHX36 helicase interact with several mRNAs. (A) Upper panel: RT–PCR validation of mRNA enrichment of the indicated candidates upon the lnc-SMaRT pull-down performed with Set-1 and Set-2 probes; a control set of probes against LacZ mRNA (LacZ) was used as negative control. GAPDH was used as negative control. Input sample accounts for 10% of the extract. Lower panel: Western Blot analysis showing the specific enrichment of the DHX36 helicase in lnc-SMaRT pull-down; GAPDH was used as negative control. Input sample accounts for 5% of the extract. (B) Upper panel: Western Blot with DHX36 antibodies on protein extracts from DHX36 RNA immunoprecipitation. Input sample accounts for 2.5% of the extract. Lower panel: RT-PCR analysis of lnc-SMaRT and indicated mRNAs in DHX36 RIP-derived RNA extracts. WBP4 was used as positive control; GAPDH was used as negative control. Input sample accounts for 10% of the extract. Representative results from three independent experiments are shown.

We then selected the candidates showing the G-quadruplex structure embedded within the CDS and containing a putative second region of interaction with lnc-SMaRT in the 3′UTR, as it occurs for Mlx-γ mRNA, for further analysis, supposing that a similarity in the molecular complex structure could be paralleled by the conservation of the regulatory mechanism. Only two mRNAs shared these features with Mlx mRNA: Acad8 and Spire1 (Supplementary Figure 2(C)–(E)).

Conversely to the DHX36 RIP experiment, lnc-SMaRT was not recovered in the immunoprecipitated fraction when we performed a DHX36 CLIP (Cross-linking immunoprecipitation) assay by combining UV cross-linking with immunoprecipitation in C2C12 cells at day 2 of differentiation (DM2) (Figure 2(B) and Supplementary Figure 2(G)). These results suggest that the interaction between DHX36 and lnc-SMaRT is indirect and that the molecular mediators of the complex formation are the mRNA interactors, probably through the second region of interaction. Indeed Mlx, Acad8 and Spire1 were found to be directly bound by the helicase (Figure 2(B)).

DHX36 and lnc-SMaRT control Acad8 and Spire1 translation

In order to evaluate the possibility of a post-transcriptional regulation exerted by lnc-SMaRT, we analyzed the mRNA and protein levels of Spire1 and Acad8 during C2C12 differentiation by qPCR and Western Blot respectively. As shown in Figure 3, the hypothesis of a translational repression mediated by lnc-SMaRT is compatible with the obtained expression profiles. In particular, when the expression of lnc-SMaRT reaches its peak (DM1 and DM2), there is a decrease in SPIRE1 and ACAD8 proteins that is not paralleled by their mRNAs (Figure 3(A) and (B)).

In order to assess the role of DHX36 and lnc-SMaRT in regulating Acad8 and Spire1 translation, we performed an RNAi against DHX36 (siDHX36) and lnc-SMaRT (siSMaRT) on C2C12

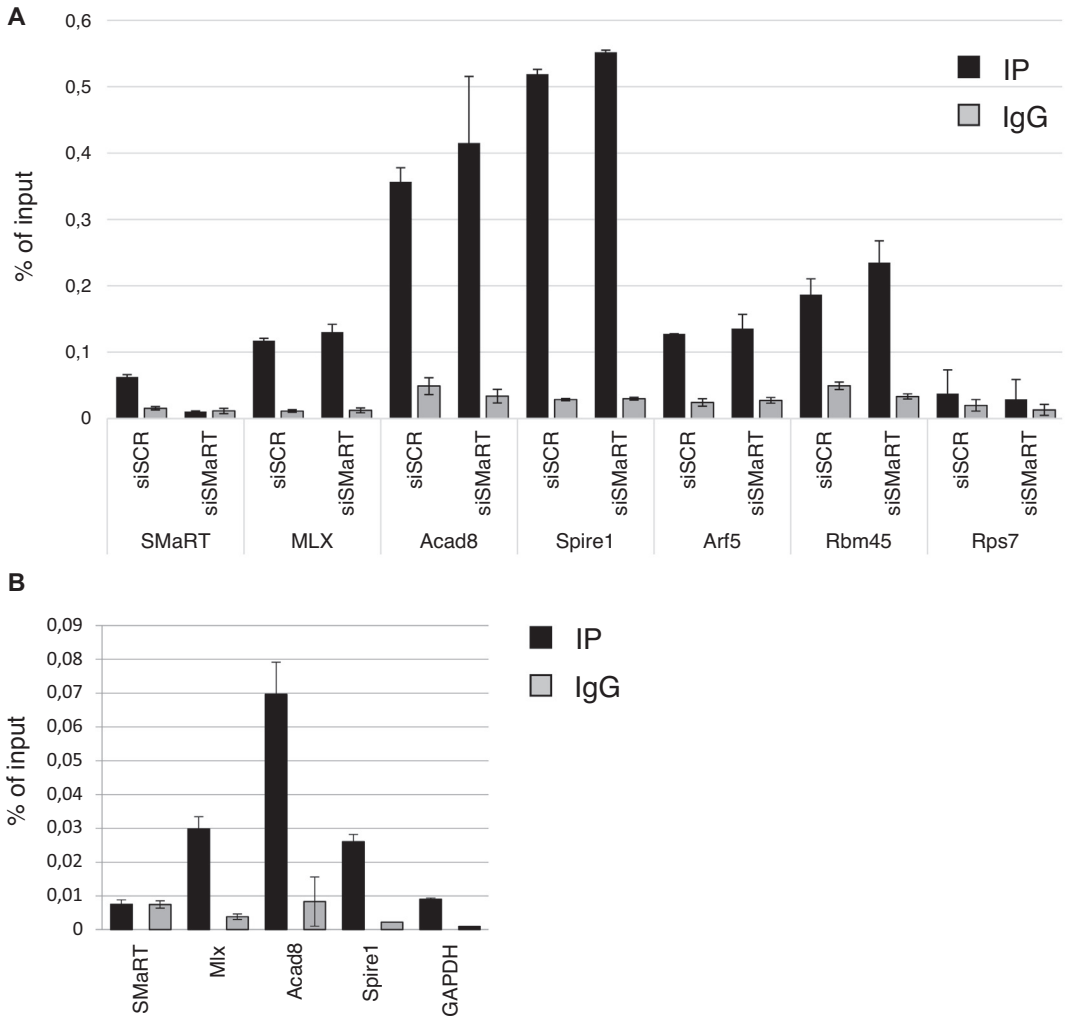

Figure 2. The interaction between DHX36 and lnc-SMaRT is indirect. (A) qPCR analysis of the indicated RNAs enrichment in DHX36 RIP-derived RNA extracts in control condition (siSCR) or upon lnc-SMaRT depletion (siSMaRT). MLX was used as positive control, Rps7 was used as negative control. Data are expressed as percentage of input and presented as the mean ± s.e.m. Representative results from three independent experiments are shown. (B) qPCR analysis of the indicated RNAs enrichment in DHX36 CLIP-derived RNA extracts. MLX was used as positive control, GAPDH was used as negative control. Data are expressed as percentage of input and presented as the mean ± s.e.m of three biological replicates.

cells collected two days after the switch to differentiation medium. The depletion of DHX36 produced a decrease in SPIRE1 and ACAD8 protein levels (Figure 3(C) and Supplementary Figure 3), while lnc-SMaRT knockdown caused an increase of Acad8 and Spire1 translation (Figure 3 (D) and Supplementary Figure 3). In both cases RNA levels were not affected (Supplementary Figure 3), pointing out that lnc-SMaRT and DHX36 act in an antagonistic manner in regulating Acad8 and Spire1 mRNAs translation.

Lnc-SMaRT control on Spire1 translation is mediated by direct base-pairing with the G4-containing region

To verify if the region of interaction that mediates the translational regulation of Acad8 and Spire1 is the one that contains the G-quadruplex structure, as it occurs for Mlx-γ, luciferase reporter assays were performed in proliferating C2C12 cells.

Two different reporter constructs were used for Acad8. The first one contains the Acad8 5′UTR together with the cDNA corresponding to the first three exons and comprising the predicted G4 structure, cloned upstream of the Renilla Luciferase ORF (RLuc), and the Acad8 3′UTR sequence, cloned downstream (Supplementary Figure 4(A)). The second construct was derived by removing the 3′UTR sequence from the first one (Supplementary Figure 4(B)). When those reporters were co-transfected in proliferating C2C12 cells (a condition in which lnc-SMaRT is not expressed) with either an empty vector or with a plasmid expressing lnc-SMaRT, no significant changes in the luciferase activity were observed

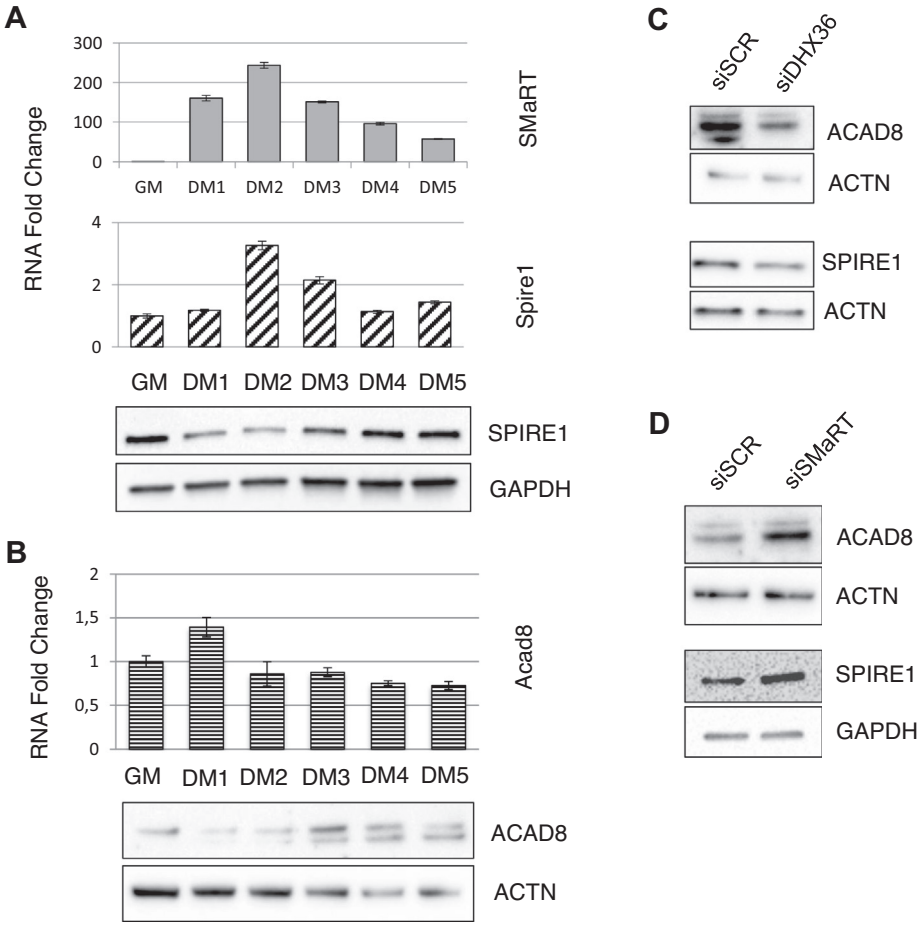

Figure 3. DHX36 and lnc-SMaRT control Acad8 and Spire1 translation. (A) Upper panel: qPCR analysis of lnc-SMaRT RNA expression in C2C12 cells undergoing differentiation at the indicated time points. GM = Growth Medium; DM = days of differentiation after switch to Differentiation Medium. The RNA expression levels were normalized against GAPDH mRNA and presented as relative quantities with respect to GM samples set to a value of 1. Data are presented as the mean ± s.e.m. Middle panel: qPCR analysis of Spire1 mRNA expression in C2C12 cells undergoing differentiation at the indicated time points. The RNA expression levels were normalized against GAPDH mRNA and presented as relative quantities with respect to GM samples set to a value of 1. Data are presented as the mean ± s.e.m. Lower panel: Western Blot with Spire1 antibodies in C2C12 cells undergoing differentiation at the indicated time points. GAPDH was used as loading control. (B) Upper panel: qPCR analysis of Acad8 mRNA expression in C2C12 cells undergoing differentiation at the indicated time points. The RNA expression levels were normalized against GAPDH mRNA and presented as relative quantities with respect to GM samples set to a value of 1. Data are presented as the mean ± s.e.m. Lower panel: Western Blot with Acad8 antibodies in C2C12 cells undergoing differentiation at the indicated time points. Actinin (ACTN) was used as loading control. (C) Western Blot with Acad8 and Spire1 antibodies in C2C12 cells collected two days after switch to differentiation medium in control condition (siSCR) and upon DHX36 depletion (siDHX36). Actinin (ACTN) was used as loading control. (D) Western Blot with Acad8 and Spire1 antibodies in C2C12 cells collected two days after switch to differentiation medium in control condition (siSCR) and upon lnc-SMaRT depletion (siSMaRT). Actinin (ACTN) and GAPDH were used as loading control. Representative results from three independent experiments are shown.

(Supplementary Figure 4(C) and (D)). Those results suggest that the observed effect of lnc-SMaRT on Acad8 translation is not mediated by the identified putative G4 interacting region, meaning that in this case the lncRNA acts in a different way.

Nonetheless, the conservation of the mechanism of action described for Mlx mRNA[13] was tested on Spire1 mRNA. Several luciferase reporter constructs were used. The first one includes both regions of interaction with lnc-SMaRT: the Spire1 5′UTR together with the cDNA corresponding to the first four exons, cloned upstream of the Renilla Luciferase ORF (RLuc), and the Spire1 3′UTR sequence, cloned downstream (Figure 4(A) – left panel). The second construct was derived from the first one by removing the Spire1 3′UTR (Figure 4 (B) – left panel). The third one contains the Spire1 G-quadruplex-forming sequence, that falls within

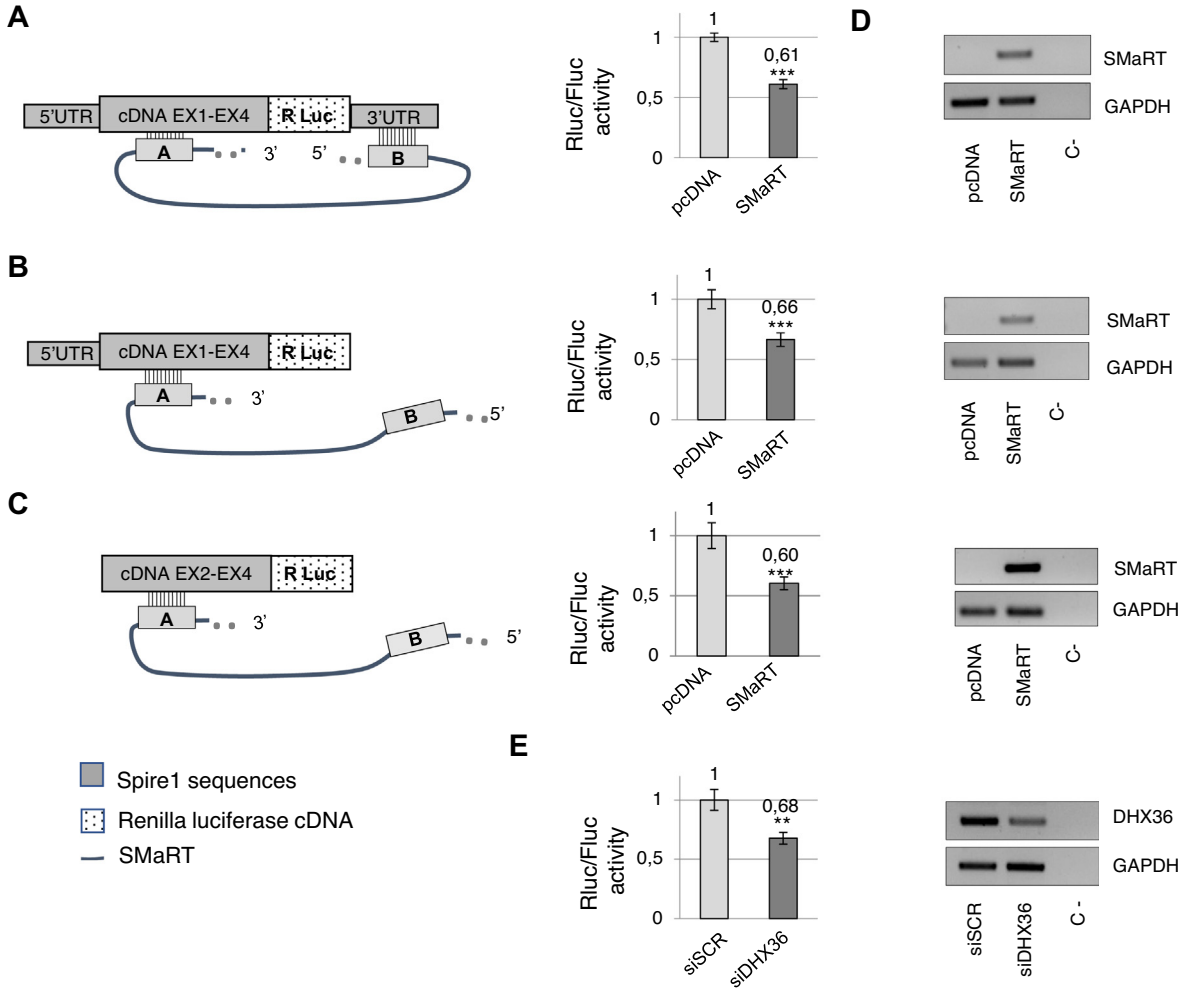

Figure 4. Translation regulation of Spire1 is mediated by its interaction with lnc-SMaRT through the G-quadruplex-forming sequence. (A) Left panel: Schematic representation of the first luciferase construct containing the Spire1 5′UTR together with the cDNA corresponding to the first four exons cloned upstream Renilla Luciferase ORF (RLuc), and the Spire1 3′UTR sequence cloned downstream. Right panel: Luciferase activity recorded upon transfection of the first luciferase construct together with an empty vector (pcDNA) or a vector for the overexpression of lnc-SMaRT (SMaRT). Luciferase activity data are presented as the mean ± s.e.m. of three biological replicates and shown with respect to pcDNA vector set to a value of 1. Statistical analysis was performed with paired two-tailed *t*-test. ***P < 0.001. (B) Left panel: Schematic representation of the second luciferase construct containing the Spire1 5′UTR together with the cDNA corresponding to the first four exons cloned upstream Renilla Luciferase ORF (Rluc). Right panel: Luciferase activity recorded upon transfection of the second luciferase construct together with an empty vector (pcDNA) or a vector for the overexpression of lnc-SMaRT (SMaRT). Luciferase activity data are presented as the mean ± s.e.m. of three biological replicates and shown with respect to pcDNA vector set to a value of 1. Statistical analysis was performed with paired two-tailed *t*-test. ***P < 0.001. (C) Left panel: Schematic representation of the third luciferase construct containing the full-length region from exon 2 to exon 4 (261 bp) of Spire1 cDNA cloned upstream Renilla Luciferase ORF (Rluc). Right panel: Luciferase activity recorded upon transfection of the third luciferase construct together with an empty vector (pcDNA) or a vector for the overexpression of lnc-SMaRT (SMaRT). Luciferase activity data are presented as the mean ± s.e.m. of three biological replicates and shown with respect to pcDNA vector set to a value of 1. Statistical analysis was performed with paired two-tailed *t*-test. ***P < 0.001. (D) RT-PCR for lnc-SMaRT on RNA extract from the luciferase assays described in A, B and C. GAPDH was used as control. Representative results from three independent experiments are shown. (E) Left panel: Luciferase activity recorded upon transfection of the third luciferase construct together with scramble siRNAs (siSCR) or siRNAs against DHX36 (siDHX36). Luciferase activity data are presented as the mean ± s.e.m. of three biological replicates and shown with respect to pcDNA vector set to a value of 1. Statistical analysis was performed with paired two-tailed *t*-test. **P < 0.01. Right panel: RT-PCR for DHX36 mRNA on extract from the luciferase assays. GAPDH was used as control. Representative results from three independent experiments are shown.

the region between exon2 and exon4, together with its flanking regions (Figure 4(C) – left panel). The fourth construct includes only the Spire1 3′UTR sequence cloned downstream of the RLuc ORF (Supplementary Figure 4(E)), while the fifth one was obtained by removing 54 bp, corresponding to the G4-forming sequence (Supplementary Figure 2(A), red box), from the third construct (Supplementary Figure 4(G)).

The different reporter constructs were independently co-transfected in proliferating C2C12 cells with either an empty vector or with a plasmid expressing lnc-SMaRT. A significant reduction of luciferase activity, around 40% (Figure 4(A)–(C)), was observed upon the overexpression of the lncRNA (assessed by RT-PCR and shown in Figure 4(D)) together with Luciferase reporters harboring the G4 sequence (first reporter construct, Figure 4(A); second reporter construct, Figure 4(B); third reporter construct, Figure 4(C)). On the contrary, no variations of luciferase activity were observed when reporter constructs containing only the Spire1 3′UTR or devoid of its G4 sequence were used (3′UTR reporter, Supplementary Figure 4(E) and (F); ΔG4 construct, Supplementary Figure 4 (G) and (H)). Moreover, when a derivative of lncSMaRT, depleted of a region containing the predicted base-pairing sequence with Spire1 (367-542nt), was used for the overexpression together with the third reporter construct, no variations in luciferase activity were detected (Supplementary Figure 4(I) and (L)). Finally, C2C12 cells were first transfected under growth conditions (GM) with the luciferase construct containing the Spire1 G-quadruplex-forming sequence (Figure 4(C)) and were then sent to differentiate for two days (DM2). By comparing luciferase signal levels between GM and DM2 conditions, it was indeed possible to observe a decrease in luciferase signal in the presence of the endogenous SMaRT (DM2) (Supplementary Figure 4(M)).

Taken together, these data suggest that lnc-SMaRT is involved in the translational regulation of Spire1 and that the base-pairing of lnc-SMaRT with the Spire1 G-quadruplex-forming sequence is sufficient and necessary to mediate it.

The role of DHX36 was further inspected by taking advantage of luciferase reporter constructs. Proliferating C2C12 cells were co-transfected with Luciferase reporter constructs containing or lacking the Spire1 G4 sequence (third and fifth reporter constructs, Figure 4(C) and Supplementary Figure 4(G)) together with scramble siRNAs (siSCR) or siRNAs against DHX36 (siDHX36). The reduction of DHX36 only affects the luciferase activity of the G4-containing reporter construct (Figure 4(E) and Supplementary Figure 4(N)), confirming the role of this helicase in the resolution of the Spire1 G-quadruplex-forming sequence.

Role of Spire1 in muscle differentiation

Spire1 is an actin nucleation factor whose role in skeletal muscle has never been characterized before. It belongs to the Spire protein family, that is composed of two members: Spire1 and Spire2. While Spire2 is described as ubiquitously present, previous papers on Spire1 reported a specific expression restricted to nervous system, oocytes and testes.[27] The tissue-specificity of Spire1 and Spire2 was analyzed by qPCR in different mouse samples (Supplementary Figure 5). Spire1 was expressed both in skeletal muscle and heart, even if at lower levels when compared to cerebellum and testis (Supplementary Figure 5(A) – left panel), whereas Spire2 was almost absent in muscle tissues (Supplementary Figure 5(A) – right panel). Their expression was also measured in C2C12 cells in growth condition (GM) and 2 days after the switch to differentiation medium (DM2): while Spire1 was present and its expression increased from GM to DM2, Spire2 was expressed at low levels and showed no variation (Supplementary Figure 5(B)). To assess Spire1 role during *in vitro* skeletal muscle differentiation, C2C12 cells were treated with siRNAs against Spire1 (siSpire1) or a scramble control (siScr). Cells were collected two days after the switch to differentiation medium and the expression of Spire1, Spire2 and several muscle differentiation markers was analyzed by qPCR. An 80% decrease of Spire1 mRNA expression was observed together with a reduction in protein abundance (Figure 5(A) and (B)). Since it was previously observed that Spire1 absence can be compensated by Spire2 upregulation in the mouse oocyte [25], its expression was also checked, however its levels remained unaffected and almost undetectable (Figure 5(C)). Those results suggest that the compensatory mechanism is not maintained in skeletal muscle cells. The expression level of the muscle differentiation marker MyoD was unchanged upon depletion of Spire1, while the one of the downstream factors myogenin (MyoG), Myocyte Enhancer Factor 2C (Mef2C) and Dystrophin (Dmd) decreased, uncovering a delay in differentiation (Figure 5(C)). Interestingly, also lnc-SMaRT expression, that was placed downstream of MyoD and MyoG regulation,[19] was reduced. Those results suggest a role of Spire1 in *in vitro* muscle cell differentiation.

Lnc-SMaRT conservation in human

According to *Basic Local Alignment Search Tool* (BLAST- https://blast.ncbi.nlm.nih.gov/Blast.cgi), the murine lnc-SMaRT sequence has no regions of similarity within the human genome. The same result was obtained when the inquiry was performed using only the lnc-SMaRT Region A sequence (from nucleotide 367 to 542), the one that base-pairs with the G-quadruplex structures identified in the interacting transcripts and that is predicted by bioinformatic analysis to be the most

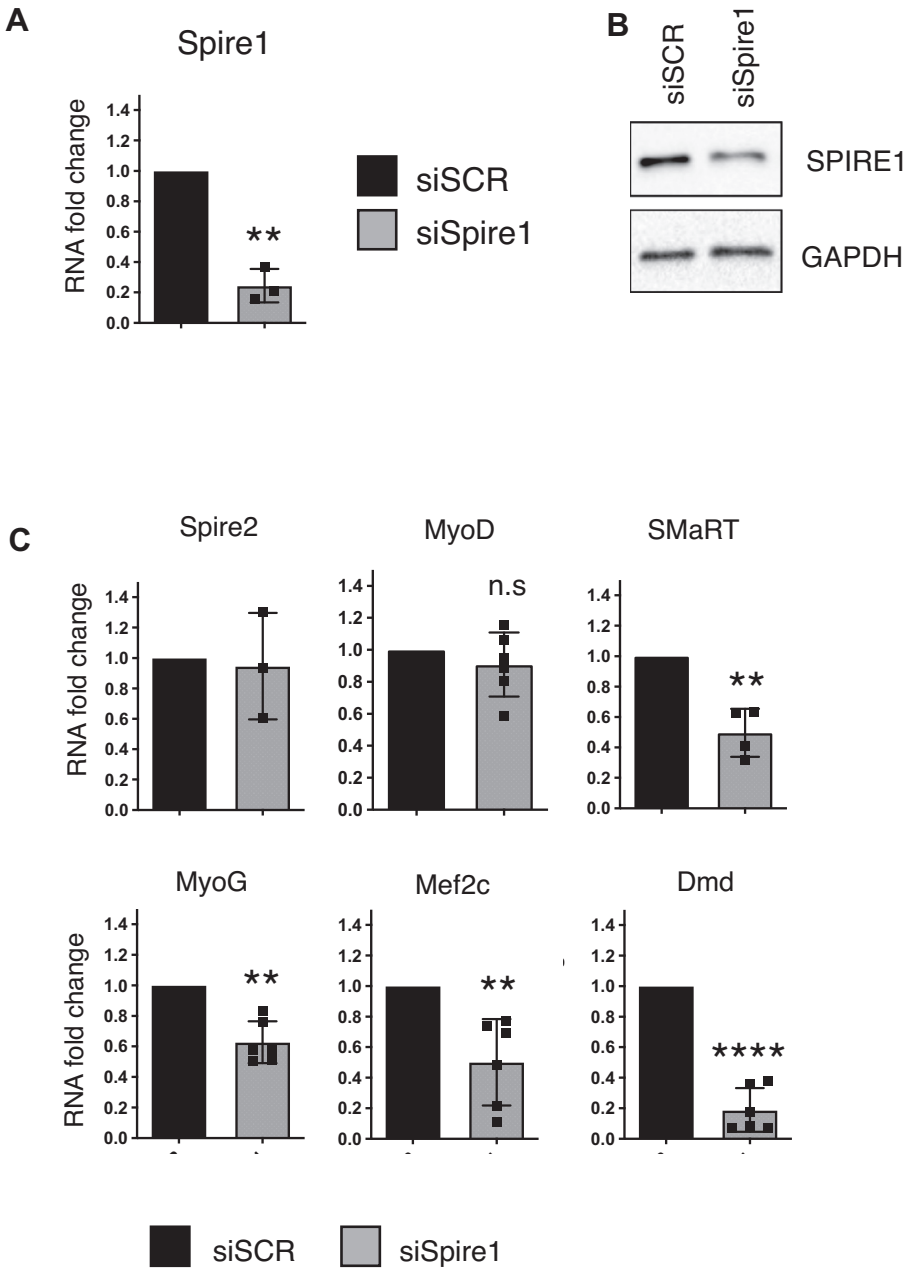

Figure 5. Spire1 is involved in *in vitro* skeletal muscle differentiation. (A) qPCR analysis of Spire1 mRNA expression in C2C12 cells undergoing differentiation (DM2) in control samples (black bars) or samples depleted of Spire1 (gray bars). The RNA expression levels were normalized against GAPDH mRNA and presented as relative quantities with respect to siScr sample set to a value of 1. Data are presented as the mean ± s.e.m of three biological replicates. Statistical analysis was performed with paired two-tailed t-test. **P < 0.01. (B) Western Blot analysis of Spire1 expression in C2C12 cells undergoing differentiation (DM2) in control samples (siSCR) or samples depleted of Spire1 (siSpire1). GAPDH was used as loading control. Representative results from three independent experiments are shown. (C) qPCR analysis of indicated RNAs expression in C2C12 cells undergoing differentiation (DM2) in control samples (black bars) or samples depleted of Spire1 (gray bars). The RNA expression levels were normalized against GAPDH mRNA and presented as relative quantities with respect to siScr sample set to a value of 1. Data are presented as the mean ± s.e.m of at least three biological replicates. Statistical analysis was performed with paired two-tailed *t*-test. **P < 0.01, ****P < 0.0001.

available for mRNA interaction (Supplementary Figure 6(A)). A syntenic conservation analysis was also performed using UCSC Genome Browser.[20] The neighboring genes of murine lncSMaRT are two coding transcripts (Bcor and Atp6ap2), both conserved in human (Supplemen-

tary Figure 6(B)). One gene (LOC101927476), encoding for a lncRNA, placed between human Bcor and Atp6ap2, was identified. However, RT-PCR analyses on human myoblasts showed almost undetectable levels of this RNA molecule, not compatible with Spire1 mRNA regulation (Supplementary Figure 6(C)).

Those results suggest that sequence and synteny of lnc-SMaRT are not conserved in human; however, we cannot rule out the possibility that a human lncRNA could execute functions similar to the ones of lnc-SMaRT during human muscle differentiation. This hypothesis is further supported by the sequence conservation of Mlx and Spire1 mRNAs in human (percentage of sequence identity: 89.3 for Spire1 and 97.9 for Mlx retrived by Biomart[21]).

Moreover, the structure of the human Mlx locus matches the murine one, displaying three main isoforms among which only one (Mlx-γ) contains a predicted G-quadruplex perfectly conserved among the two species (Supplementary Figure 6 (D)). Furthermore, one of the three predicted G-quadruplexes of human Spire1 shows high homology with the murine one and is localized in the CDS of human Spire1 mRNA (Supplementary Figure 6(D)).

The aforementioned conserved features of lnc-SMaRT targets in human allowed us to raise an intriguing hypothesis about mechanism conservation.

In order to find a functional analogue of lnc-SMaRT in human, a DHX36 RIP assay was performed on human myoblasts extract. Indeed, we showed in mouse that, even if DHX36 is not directly bound to lnc-SMaRT, the RIP assay allows the recovery of all directly and indirectly interacting RNAs. Cells were collected in DM1 since the differentiation stage corresponds to that of murine myoblasts (C2C12) two days after the induction of differentiation (DM2), according to cell morphology and differentiation markers (Supplementary Figure 6(E)), which is when lnc-SMaRT reaches its peak expression. By using specific DHX36 antibodies, the helicase was immunoprecipitated and the recovered RNAs were sequenced. The RIP-Seq analysis identified 1765 RNAs interacting with the helicase. Notably, among them we found the human homologs of Mlx and Spire1 mRNAs.

The sequencing analysis revealed that the genes, whose RNAs interact with DHX36, are enriched in G-quadruplex-forming sequences. In order to verify the expected G-quadruplex enrichment, a background containing 8880 genes was created by intersecting the list of genes expressed in human wild-type myoblasts in DM1, representing the INPUT of the experiment, and the list of genes expressed in HeLa cells, used as an experimental system for the RNA G-quadruplex sequencing (rG4-seq) performed by the Balasubramanian lab.[22] They performed the rG4-seq in the presence of K^+ cations alone or in combination with pyridostatin (K^+/PDS), a compound that stabilizes G-quadruplexes. In the generated common background, there were 1520 genes recovered with DHX36 RIP assay and 2039 G-quadruplex-containing genes according to rG4-seq in the presence of K^+ cations. By intersecting these two groups of genes, 399 common genes were found (Fisher Exact Test, p value = 0,0009 – Figure 6(A)). The same intersection was done between the 1520 genes and 4954 G-quadruplex-containing genes recovered with rG4-seq in the presence of K^+/PDS, and we found 925 common genes (Fisher Exact Test, p value = 1,6 * e-05 – Figure 6(A)). These data indicate that the RIP genes list is enriched in G-quadruplex-forming sequences.

In order to find good candidates with similar functions to lnc-SMaRT's, 77 lncRNAs found enriched in the RIP genes list were filtered to retain those ones devoid of G-quadruplex-forming regions, both predicted or experimentally identified (56 lncRNAs). Furthermore, since lnc-SMaRT is expressed only during myoblast differentiation, we selected the 12 lncRNAs described as upregulated in human myoblasts differentiation. 7 lncRNAs, whose predicted interaction with Spire1 and Mlx overlaps with their G-quadruplex-forming sequences, were selected (BX890604, CYTOR, AMZ2P1, SNAI3_AS1, DUBR, MAPKAPK5-AS1 and NNT-AS1; Table 1). After performing Western Blot analysis to assess the specificity of DHX36 immunoprecipitation compared with INPUT, IgG (negative control), IM-IP and IM-IgG (the supernatants recovered after the immunoprecipitation step) (Supplementary Figure 6(F)), the interaction of the seven candidates with DHX36 was validated by RT-PCR on the RNAs recovered upon the RIP assay (Figure 6(B)).

Among them, SNAI3-AS1 was not enriched in the IP fraction compared to IgG, meaning that it does not interact with DHX36; the remaining six lncRNAs, that bind to the helicase, could be potential candidates exerting similar functions to lnc-SMaRT in human muscle cells.

Discussion

In this paper we identified a new mRNA, Spire1, whose translation is regulated by the lnc-SMaRT/DHX36 molecular complex, providing further insights in the formerly proposed mechanisms[13] and proving its extendibility to other targets. In a previous paper, we demonstrated that lnc-SMaRT is able to repress the translation of its molecular target Mlx-γ by base-pairing with a G-quadruplex-forming region harbored within the mRNA. The role of the DHX36 RNA helicase is to solve the G-quadruplex structure on the Mlx-γ mRNA, allowing either its translation or its base-pairing with lnc-

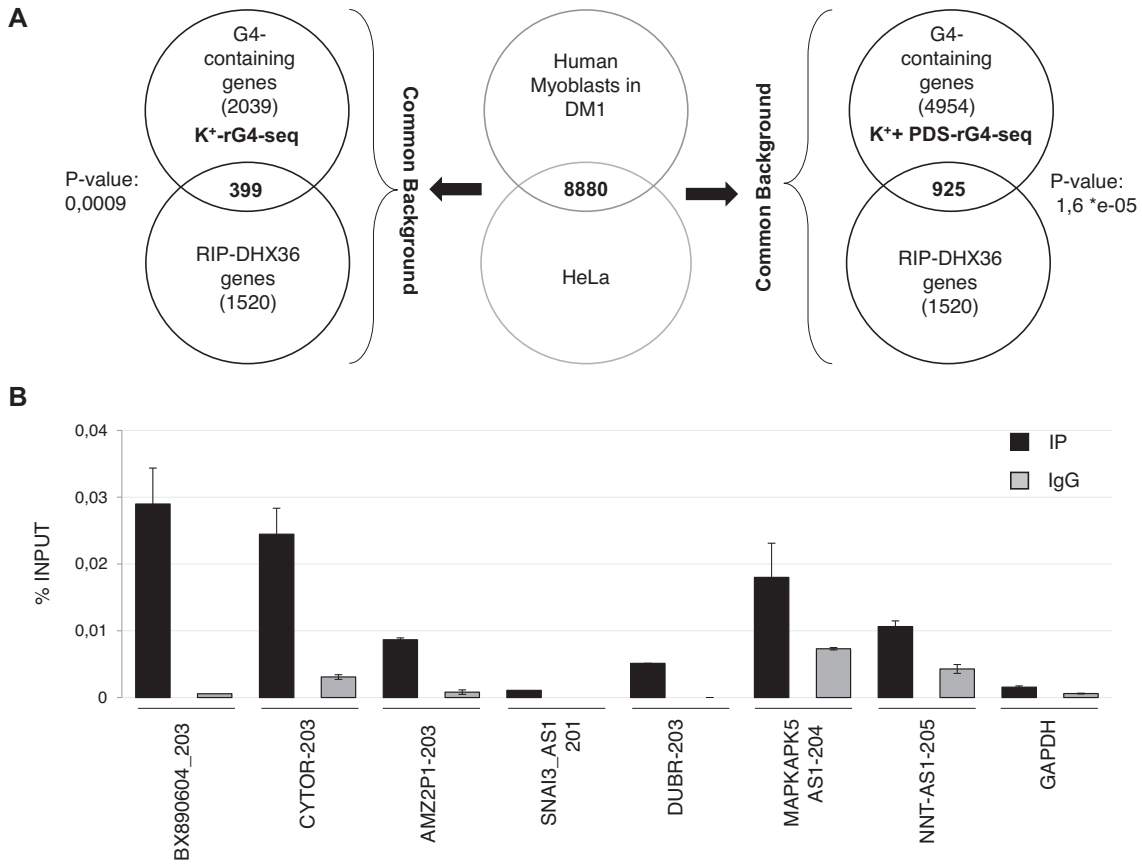

Figure 6. Putative candidates exerting lnc-SMaRT functions in human cells. (A) Schematic representation of RIP G-quadruplex-containing genes enrichment analysis described in the main text. (B) qPCR analysis of the indicated lncRNAs on DHX36 RIP derived RNA performed on human myoblasts one day after switch to differentiation medium. GAPDH was used as negative control. Data are expressed as percentage of input and presented as the mean ± s.e.m. Representative results from three independent experiments are shown.

SMaRT, depending on whether the latter is present or not. Here, we were able to demonstrate that the interaction between the long non-coding RNA SMaRT and the DHX36 RNA helicase is indirect and mediated by the mRNAs present in the complex; moreover, the SMaRT/mRNAs base-pairing overlapping the G4 region is accompanied by a second region of interaction identified in the mRNA 3′UTRs. This interaction could be required to keep lnc-SMaRT in proximity to the target mRNAs, favoring its base-pairing with the G4 region immediately after the resolving activity of DHX36 has occurred, thus inhibiting the subsequent G4 refolding.

Two Spire proteins (Spire1 and Spire2), founding members of Wiskott-Aldrich homology region (WH2)-containing actin nucleators, are encoded in the human genome[23,24] and conserved in mouse.[25] These nucleation factors have a critical role in the actin cytoskeleton assembly and dynamics and are involved in vesicle transport processes, coordinating the association between actin filaments and myosin V motor proteins at vesicles membranes.[26] While Spire2 is characterized by a broad expression pattern, Spire1 was described as prevalently expressed in nervous system, oocytes, and testes.[27] In opposition to this, a wider range of Spire1 expression was described by Gene Expression Database (GXD - Mouse Genome Informatics Web Site: http://www.informatics.jax.org) and we observed significant levels of Spire1 in muscle cells and in muscle tissues, in which Spire2 was almost absent. Previously described Spire1 mutant mice showed no gross abnormalities in motor behavior and memory functions; only a male-specific dysfunction in fear learning was observed.[28] We demonstrated that the depletion of Spire1 produces a delayed *in vitro* muscle differentiation phenotype that was not compensated by Spire2, as observed during mouse oocyte maturation,[25] suggesting a new specific Spire1 role during skeletal muscle differentiation. It is interesting to note that the depletion of Spire1 does not affect the levels of the master regulator MyoD, while decreasing the expression of its downstream targets; this evidence indicates a role in the early phases of myogenesis that immediately follow the induction of the differentiation program. Intriguingly, lnc-SMaRT is expressed exactly in this time window, suggesting

Table 1 The table shows genomic coordinates ("Chr", "Start", "End" and "Strand" columns) and sequences ("G4" column) of human MLX and SPIRE1 predicted G-Quadruplexes. Moreover, the lncRNAs predicted to interact with those putative G4 regions are indicated ("Interacting lncRNA" column) together with their specific isoforms ("Interacting Isoform" column).

Chr	Start	End	Strand	G4	G-score	Gene Name	Interacting lncRNA	Interacting Isoform
chr17	42567196	42567212	+	GGGGAGGGCGGGTCGGG	63	MLX	AMZ2P1	['ENST00000565833']
							BX890604.1	['ENST00000469903', 'ENST00000461011', 'ENST00000490920']
							CYTOR	['ENST00000629257', 'ENST00000331944', 'ENST00000409139']
							DUBR	['ENST00000466734', 'ENST00000473550']
							MAPKAPK5-AS1	['ENST00000456429', 'ENST00000428207', 'ENST00000442119', 'ENST00000443596']
							NNT-AS1	['ENST00000515466']
							SNAI3-AS1	['ENST00000565633', 'ENST00000569786', 'ENST00000568633', 'ENST00000563261', 'ENST00000563475', 'ENST00000567997']
chr18	12506600	12512455	-	GGGCACGATTCTGGGTACAGGTGAT GAGGGATTTGAGGAATGGG	62	SPIRE1	AMZ2P1	['ENST00000565833']
							BX890604.1	['ENST00000469903', 'ENST00000461011', 'ENST00000483854', 'ENST00000475317']
							CYTOR	['ENST00000331944']
							DUBR	['ENST00000473550']
							SNAI3-AS1	['ENST00000565633', 'ENST00000569786', 'ENST00000568633', 'ENST00000563261', 'ENST00000563475']
chr18	12657786	12657813	-	GGGCGGCGAGGGGCCGCGGGAGCCCGGG	63	SPIRE1	BX890604.1	['ENST00000490920']
							MAPKAPK5-AS1	['ENST00000456429', 'ENST00000428207']
							NNT-AS1	['ENST00000515466', 'ENST00000513560']
							SNAI3-AS1	['ENST00000567997']
chr18	12658112	12658130	-	GGGGCGGGGCGGGGCGGGG	84	SPIRE1	/	/

that Spire1 protein levels have to be finely regulated during the first instants of *in vitro* muscle development to ensure proper differentiation of myoblasts. Moreover, we also observed a down-regulation of lnc-SMaRT expression after Spire1 knockdown, probably as a consequence of the observed down-regulation of MyoG, whose binding sites in the lnc-SMaRT promoter have been previously identified [19]. The downregulation of SMaRT in the absence of Spire1 can explain the unexpected concordance of the delayed muscle differentiation phenotype observed after Spire1 or SMaRT depletion, and suggests that Spire1 protein could exert different functions during muscle cell proliferation and differentiation processes. In this work, we also showed that lnc-SMaRT acts as a translational regulator of Acad8 mRNA, which encodes a mitochondrial Isobutyryl-CoA dehydrogenase. ACAD8 activity is crucial for branched-chain amino acid catabolism, a metabolic process for which skeletal muscle is the main contributing tissue.[29] A precise role for ACAD8 in skeletal muscle differentiation has not been elucidated yet, even though mutations in the human gene, causative of Isobutyric Aciduria (Orphanet Code: 79159, Web Site: http://www.orpha.net), are occasionally associated to dilated cardiomyopathy and muscular hypotonia. Although both lnc-SMaRT and DHX36 bind Acad8 mRNA and intervene in its translational regulation, luciferase experiments showed that this modulation is not dependent on the binding of lnc-SMaRT to the G4-forming region or to the 3′UTR. Since luciferase constructs incorporate only some segments of the Acad8 transcript, we cannot rule out the possibility that other sequence motifs or secondary structures, not included in our reporter system, could be important for the lnc-SMaRT-mediated translational regulation.

Long non-coding RNAs are rapidly evolving RNA species characterized by poor sequence conservation, usually limited to short nucleotide stretches.[30] Anyway, despite the absence of sequence homology and synteny, there is the possibility of the independent evolution of different lncRNAs sharing the same function and mechanism of action in separate species.

Even though lnc-SMaRT is not directly conserved in human, its regulated targets Mlx and Spire1 are both highly conserved and are predicted to have G-quadruplex secondary structures. Moreover, these structures arise from high sequence similarity regions with the murine counterparts. Given these considerations, we hypothesized the conservation of lnc-SMaRT mechanism of action in human muscle differentiation.

We identified several lncRNAs up-regulated during human myoblast differentiation that interact with the human homolog of the DHX36 helicase and that are predicted to have thermodynamic propensity to bind both homologs of Mlx and Spire1 in their G-quadruplex predicted sequences. These features suggest a parallelism among these lncRNAs and lnc-SMaRT and give rise to the fascinating hypothesis of the convergent evolution of a similar Mlx and Spire1 regulation mechanism during the skeletal muscle differentiation in different mammalian species.

Materials and Methods

Cell culture and treatments

Mouse myoblasts (C2C12, ATCC) were cultured in growth medium (GM, Dulbecco's modified Eagle's medium with 20% fetal bovine serum, 2 mM L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin) and induced to differentiate in differentiation medium (DM, same as GM medium but with 0.5% fetal bovine serum).

Human myoblasts, obtained from the Telethon Network of Genetic Biobanks, were cultured in Human skeletal muscle growth medium (PromoCell) and differentiated using human skeletal muscle differentiation medium (PromoCell).

Interference with siRNAs was performed as follows: for a 3.5 cm culture dish, 250 μl of Opti-MEM® I Reduced Serum Medium (Gibco) were mixed with 5 μl of Lipofectamine RNAiMAX Reagent (Thermo Scientific); the siRNA was added at a final concentration of 50 nM for a final volume of 2 ml. The transfection mix was incubated at room temperature for 15 min. In order to increase the knockdown efficiency, siRNA transfections were performed in reverse: the transfection mix was spread on the culture dish and 250.000 C2C12 cells were seeded in GM directly on the mix. After 24 h cell confluency was checked, cells were induced to differentiate by replacing the culture medium with DM and then collected after 48 h. Spire1 (GS68166) and DHX36 (SI00979076) FlexiTube siRNA were purchased from Qiagen, while lnc-SMaRT siRNAs were custom synthetized (Qiagen).

RNA extraction and analysis

RNA was extracted using Directzol Miniprep RNA Purification Kit (Zymo Research) with on-column DNase treatment, according to the manufacturer's specifications, and total RNA was retro-transcribed with PrimeScriptTM RT Reagent Kit (Takara). For low RNA input experiments, RNA extraction was performed using QIAzol reagent and miRNEasy spin columns (Qiagen), following manufacturer's instructions, and reverse transcription was performed with the Superscript VILO cDNA Synthesis Kit (Life Technologies).

For samples analysis, qPCRs were performed using PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) and semi-quantitative RT–PCRs were executed using MyTaq™ DNA polymerase (Bioline).

Protein extraction and Western Blot

Protein extract was obtained using standard RIPA buffer, supplied with 1 × Complete Protease Inhibitor Cocktail (Roche). Protein extracts were loaded on 4–15% Mini-PROTEAN TGX Precast acrylamide gels (Bio-Rad) according to the manufacturer's specifications, then proteins were transferred to Immobilon-E PVDF 0.45-μm membrane (Merck-Millipore). Membranes were blocked with 5% non-fat dry milk (Difco) and incubated overnight at 4 °C with primary antibodies. Protein detection was carried out with WesternBright ECL (Advansta) using ChemiDoc™ MP System and images were analyzed using Image Lab™ Software (Bio-Rad). Primary antibodies: anti-ACTININ-4 (G-4, sc-390205, Santa Cruz Biotechnology), anti-GAPDH (6C5, sc-32233, Santa Cruz Biotechnology), and anti-HPRT (FL-218, sc-20975, Santa Cruz Biotechnology), anti-DHX36 (13159-1-AP, Proteintech), anti-Spire1 (H-1, sc-515448, Santa Cruz Biotechnology), anti-Acad8 (A68138, Epigentek). Secondary antibodies: goat anti-rabbit HRP (31460, Invitrogen) and goat anti-mouse HRP (32430, Invitrogen).

DHX36 RNA immunoprecipitation

Two 10 cm plates of day 1-differentiated human myoblasts, or day 2-differentiated C2C12 mouse myoblasts, were scraped in PLB Buffer (KCl 100 mM; MgCl2 5 mM; NP-40 0.5%; DTT 1 mM; Protease and RNase inhibitor), collected, lysed for 15 min in rotation at 4 °C and then centrifuged at 13,000 rpm for 10 min at 4 °C to remove cellular debris. The supernatant was recovered and protein concentration was quantified by Bradford assay; 0,5 mg of extract was used for each sample (IP and IgG). The supernatant was precleared with 40 μl of Protein G Agarose/ Salmon Sperm Beads (Millipore) in a final volume of 1 ml of NT2 buffer (TRIS-HCl pH 7.4 50 mM; NaCl 150 mM; MgCl2 1 mM; NP-40 0.05%; Protease and RNase Inhibitor) in incubation for 2 h in rotation at 4 °C. 10% of the final volume was taken as Input and the remaining precleared lysate was incubated with 5 μg of DHX36 (Proteintech, 13159-1-AP) or IgG (Santa Cruz Biotechnology, sc-2027) antibodies overnight at 4 °C. The following day, 80 μl of Protein G agarose beads were added to each sample and incubated for 2 h at 4 °C. Next, the beads were washed 4 times in NT2 buffer and finally resuspended in 200 μl of NT2 Buffer. 50 μl of beads were pelleted, resuspended in 1x Laemmli Sample Buffer (BioRad) and 50 mM DTT and incubated at 70 °C for 15 min. The eluate was used for Western Blot, in order to check DHX36 immunoprecipitation efficiency. The remaining 150 μl of beads were used to recover RNA by resuspending them in 500 μl of TRI-Reagent (Zymo Research) for the RNA extraction.

DHX36 cross-linking and immunoprecipitation

One 10 cm plate of day 2-differentiated C2C12 cells was rinsed with PBS and UV crosslinked at 254 nm, 4000X100 uJ/cm2. The plate was scraped using NP40-Lysis Buffer (HEPES-KOH 50 mM; KCl 150 mM; EDTA 2 mM; NaF 1 mM; NP-40 0.5%; DTT 0,5 mM; Protease and RNase inhibitor). Cells were collected and lysed for 15 min in rotation at 4 °C, then centrifuged at 13,000 rpm, 10 min, 4 °C to remove cellular debris. The supernatant was recovered, and protein concentration was quantified by Bradford assay: 0.5 mg of extract was used for each sample (IP and IgG). The supernatant was precleared with 40 μl of Dynabeads Protein G (Invitrogen) in a final volume of 1 ml of Wash Buffer (HEPES-KOH 50 mM; KCl 150 mM; NP-40 0.05%; DTT 0.5 mM; Protease and RNase inhibitor) in incubation for 2 h in rotation at 4 °C. 10% of the final volume was taken as Input and the remaining precleared lysate was incubated with 2 μg of DHX36 (Proteintech, 13159-1-AP) or rabbit IgG (Santa Cruz Biotechnology, sc-2027) antibodies overnight at 4 °C. The following day, 80 μl of Dynabeads Protein G were added to each sample (IP and IgG) and incubated for 2 h at 4 °C. After the recovery of antibody-protein complexes, the supernatants were collected to check the effective immunodepletion of the protein from the extract (Immunodepleted-IP, Immunodepleted-IgG). The beads were washed 4 times with High Salt Wash Buffer (HEPES-KOH 50 mM; KCl 500 mM; NP-40 0.05%; DTT 0.5 mM; Protease and RNase inhibitor) and finally resuspended in 200 μl of High Salt Wash Buffer. Protein fraction was recovered from an aliquot of 50 μl of beads form each sample (IP and IgG), which was resuspended in 1x Laemmli Sample Buffer (BioRad) and 50 mM DTT, incubated at 70 °C for 15 min and analyzed through Western Blot. RNA was recovered from the remaining 150 μl of beads: first they were treated with Proteinase K enzyme (Roche), diluted in Proteinase K buffer 2x (Tris-HCl 200 mM, NaCl 300 mM, EDTA 25 mM, SDS 2%), then incubated for 30 min at 50 °C with shaking and lastly resuspended in TRI-Reagent (Zymo Research) for the RNA extraction.

Luciferase reporter constructs

For Spire1 "5′UTR_till_exon4 + RLuc" construct, the full-length region from 5'UTR till exon 4 of Spire1 was amplified from C2C12 cDNA; then, it was cloned upstream the Renilla luciferase coding sequence of Psicheck2 vector (Promega) using the In-Fusion HD Cloning Kit (Clontech), previously depleted of its start codon by inverse PCR. For Spire1 "exon2-exon4 + RLuc" construct,

the full-length region from exon 2 to exon 4 was cloned and fused with the Renilla luciferase coding sequence of Psicheck2 vector, in order to obtain a fusion protein under the Renilla luciferase start codon. For Spire1 "5'UTR_till_exon4 + RLuc + 3'UTR" construct, Spire1 3'UTR sequence was cloned in Spire1 "5'UTR_till_exon4 + RLuc" construct using XhoI and NotI restriction enzymes (Thermo Scientific). Spire1 "delta_G4 + RLuc" construct was generated by inverse PCR with divergent primers. For Spire1 "RLuc + 3'UTR" construct, Spire1 3'UTR sequence was inserted downstream the Renilla luciferase coding sequence of Psicheck2 vector (Promega), using XhoI and NotI restriction enzymes (Thermo Scientific). The mutant lncSMaRT construct "$SMaRT_{\Delta A}$" was generated by inverse PCR with divergent primers.

Similarly, Acad8 "5′UTR_till_exon3 + RLuc" construct was obtained by cloning the full-length region from 5'UTR till exon 3 of Acad8 upstream the Renilla luciferase coding sequence of Psicheck2 vector (Promega), previously depleted of its start codon by inverse PCR. "5'UTR_till_exo n3 + RLuc + 3'UTR" construct was obtained by inserting Acad8 3'UTR in Acad8 "5'UTR_till_exo n3 + RLuc" construct with XhoI and NotI restriction enzymes (Thermo Scientific).

Luciferase assays

C2C12 cells were transiently transfected with the luciferase reporter plasmids in the indicated combinations using Lipofectamine-2000 Reagent (Thermo Scientific). The firefly luciferase (FLuc) gene contained in the reporter plasmids was used to normalize for transfection efficiency. Cells were harvested and lysed 48 h after transfection and RLuc and FLuc activities were measured by Dual Glo Luciferase assay (Promega). Transfection of each construct was performed in triplicate as well as the Luciferase assays. Ratios of RLuc readings to FLuc readings were taken for each experiment, and triplicates were averaged.

DHX36 RNA immunoprecipitation sequencing

Ovation RNA v2 followed by Ovation Ultralow v2 kit was used to prepare cDNA libraries for RIP-Seq experiments. The sequencing reactions, performed on an Illumina NovaSeq 6000 Sequencing system, produced an average of 40.5 million 150 nucleotide long paired end reads per sample.

Adapter sequences and poor quality bases were removed from raw reads using Trimmomatic software[31] with parameters ILLUMINACLIP:path/to/adapter:2:30:10:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:18. Bowtie2 software was used to identify and discard reads mapping to human ribosomal RNAs.[32]

Alignment to human GRCh38 genome and Ensembl 90 transcriptome[33] was performed using STAR aligner version 2.5.2b[34] with parameters --outFilterIntronMotifs RemoveNoncanonical --outSAMstrandField intronMotif --outSAMtype BAM SortedByCoordinate --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04. Alignment BAM files were further processed by filtering out reads mapping to mitochondrial genome, unlocalized and unplaced sequences and alternate loci using SAMtools[35] and removing PCR duplicates using Picard Mark Duplicates (available at http://broadinstitute.github.io/picard). Calculation of transcript-level FPKM and log2 fold change (log2FC) values was performed using Cuffdiff versionv 2.2.1[36] with parameters -c 8 --library-norm-method quartile --dispersion-method blind -u --library-type fr-unstranded -b path/to/genome -M path/to/maskfile, the mask file being a GTF file containing the coordinates of rRNA and tRNA loci; the GTF file used for quantification was a reduced Ensembl 90 annotation, containing only transcripts with GENCODE flag and/or with Transcript Support Level lower than 3, and from which small non-coding RNAs were removed. In parallel, we removed multi-mapped and improperly paired reads from BAM files using BamTools[37] and SAMtools.[35]

We filtered out isoforms belonging to genes with Input or IgG FPKM < 5 and those whose FPKM in the DHX36 RIP sample was lower than 2 and/or lower than the FPKM value of the most abundant isoform from the same gene multiplied by 0.05; furthermore, only isoforms with a Test Status equal to "OK" were retained. To select transcripts bound by DHX36, we set the RIP over Input log2FC cutoff and the RIP over IgG log2FC cutoff to 1.5.

Bioinformatics analyses

Coordinates of G-Quadruplex-forming sequences produced by rG4-Seq experiment were retrieved from GEO dataset GSE77282[22] and transposed to GRCh38 using LiftOver tool.[20] G-Quadruplex-containing genes were obtained intersecting rG4-Seq coordinates with Ensembl gtf (release 90) annotation file using Bedtools intersect with -s option.[38]

Genes and gene isoforms differentially expressed during control human myoblasts differentiation were identified analyzing data available at GEO accession GSE70389[39] with Cuffdiff.[36] Genes with qValue < 0.05, logFC > 0 and with at least 1 FPKM in DM condition were defined as upregulated in differentiation while isoforms with more than 0.5 FPKM in GM or DM samples were selected as expressed.

RNA–RNA interaction and G-Quadruplex predictions were performed using respectively IntaRNA 2.3.0. and QGRS mapping algorithm as described previously in.[13] Intersections between predicted G-Quadruplex-forming regions and predicted interactions were performed using Bedtools intersect.[38] For human analogous candidate identification, the predictions of RNA-RNA interactions and G-Quadruplexes of interesting genes were performed on isoforms expressed in human differentiating myoblasts and the more thermodynamically stable base pairings were selected.

Statistical analyses

Data are shown as mean ± s.e.m. The number of biological replicates is indicated in each Figure legend. Statistical tests used to assess significance of differences between means are indicated in each Figure legend.

P-values below 0.05 were considered significant (*: *P*-value < 0.05; **: *P*-value < 0.01; ***: P-value < 0.001; ****: *P*-value < 0.0001).

Data Availability

The RIP-Seq data from this publication have been deposited to the GEO database (https://www.ncbi.nlm.nih.gov/geo/) with the following identifier GSE181308.

CRediT authorship contribution statement

Silvia Scalzitti: Investigation, Validation, Formal analysis, Visualization, Writing – review & editing. **Davide Mariani:** Investigation, Validation, Formal analysis, Visualization, Writing – review & editing. **Adriano Setti:** Investigation, Formal analysis, Data curation, Visualization, Writing – review & editing. **Alessio Colantoni:** Data curation, Software. **Michela Lisi:** Investigation, Validation. **Irene Bozzoni:** Conceptualization, Supervision, Funding acquisition. **Julie Martone:** Conceptualization, Visualization, Supervision, Project administration, Writing – original draft.

Acknowledgments

The authors acknowledge the Telethon Network Genetic Biobanks (GTB12001F) for control human myoblasts, and Marcella Marchioni for technical support. This work was partially supported by PRIN 2017 (2017P352Z4) to I.B, Sapienza Research Calls (RM118164363B1D21) to J.M. and institutional funding from Fondazione Istituto Italiano di Tecnologia.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167384.

Received 3 August 2021;
Accepted 26 November 2021;
Available online 2 December 2021

Keywords:
long ncRNAs;
DHX36 helicase;
G-quadruplex;
post-transcriptional regulation;
myogenesis

† These authors contributed equally to this work.

References

1. Morris, K., Mattick, J., (2014). The rise of regulatory RNA. *Nature Rev. Genet.* **15**, 423–437.
2. Grote, P., Herrmann, B.G., (2015). Long noncoding RNAs in organogenesis: making the difference. *Trends Genet.* **31**, 329–335.
3. Ransohoff, J., Wei, Y., Khavari, P., (2018). The functions and unique features of long intergenic non-coding RNA. *Nature Rev. Mol. Cell. Biol.* **19**, 143–157.
4. Ponjavic, J., Ponting, C.P., Lunter, G., (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565.
5. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., et al., (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227.
6. Jandura, A., Krause, H.M., (2017). The new RNA world: growing evidence for long noncoding RNA functionality. *Trends Genet.* **33**, 665–676.
7. Fatica, A., Bozzoni, I., (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nature Rev. Genet.* **15**, 7–21.
8. Chen, L., Zhang, S., (2016). Long noncoding RNAs in cell differentiation and pluripotency. *Cell Tissue Res.* **366**, 509–521.
9. Yao, R.W., Wang, Y., Chen, L.L., (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biol.* **21**, 542–551.
10. Wang, K.C., Chang, H.Y., (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell.* **43**, 904–914.
11. Rinn, J.L., Chang, H.Y., (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166.
12. Martone, J., Mariani, D., Desideri, F., Ballarino, M., (2020). Non-coding RNAs shaping muscle. *Front. Cell Dev. Biol.* **7**, 394.
13. Martone, J., Mariani, D., Santini, T., Setti, A., Shamloo, S., Colantoni, A., Capparelli, F., Paiardini, A., et al., (2020). SMaRT lncRNA controls translation of a G-quadruplex-

containing mRNA antagonizing the DHX36 helicase. *EMBO Rep.* 49942.
14. Tran, H., Schilling, M., Wirbelauer, C., Hess, D., Nagamine, Y., (2004). Facilitation of mRNA deadenylation and decay by the exosome-bound, DExH protein RHAU. *Mol. Cell.* **13**, 101–111.
15. Meroni, G., Cairo, S., Merla, G., Messali, S., Brent, R., Ballabio, A., Reymond, A., (2000). Mlx, a new Max-like bHLHZip family member: the center stage of a novel transcription factors regulatory pathway? *Oncogene* **19**, 3266–3277.
16. Kikin, O., D'Antonio, L., Bagga, P.S., (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, 676–682.
17. Lattmann, S., Stadler, M.B., Vaughn, J.P., Akman, S.A., Nagamine, Y., (2011). The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.* **39**, 9390–9404.
18. Mann, M., Wright, P.R., Backofen, R., (2017). IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.* **45**, 435–439.
19. Ballarino, M., Cazzella, V., D'Andrea, D., Grassi, L., Bisceglie, L., Cipriano, A., Santini, T., Pinnarò, C., et al., (2015). Novel long noncoding RNAs (lncRNAs) in myogenesis: a miR-31 overlapping lncRNA transcript controls myoblast differentiation. *Mol. Cell. Biol.* **35**, 728–736.
20. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.
21. Durinck, S., Spellman, P., Birney, E., Huber, W., (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* **4**, 1184–1191.
22. Kwok, C.K., Marsico, G., Sahakyan, A.B., Chambers, V.S., Balasubramanian, S., (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nature Methods* **13**, 841–844.
23. Kerkhoff, E., Simpson, J.C., Leberfinger, C.B., Otto, I.M., Doerks, T., Bork, P., Rapp, U.R., Raabe, T., et al., (2001). The Spir actin organizers are involved in vesicle transport processes. *Curr. Biol.* **11**, 1963–1968.
24. Quinlan, M.E., Heuser, J.E., Kerkhoff, E., Mullins, R.D., (2005). Drosophila Spire is an actin nucleation factor. *Nature* **433**, 382–388.
25. Pfender, S., Kuznetsov, V., Pleiser, S., Kerkhoff, E., Schuh, M., (2011). Spire-type actin nucleators cooperate with Formin-2 to drive asymmetric oocyte division. *Curr. Biol.* **21**, 955–960.
26. Pylypenko, O., Welz, T., Tittel, J., Kollmar, M., Chardon, F., Malherbe, G., Weiss, S., Michel, C.I.L., et al., (2016). Coordinated recruitment of Spir actin nucleators and myosin V motors to Rab11 vesicle membranes. *eLife* **2**, 17523
27. Pleiser, S., Rock, R., Wellmann, J., Gessler, M., Kerkhoff, E., (2010). Expression patterns of the mouse Spir-2 actin nucleator. *Gene Expr. Patterns* **10**, 345–350.
28. Pleiser, S., Banchaabouchi, M.A., Samol-Wolf, A., Farley, D., Welz, T., Wellbourne-Wood, J., Gehring, I., Linkner, J., et al., (2014). Enhanced fear expression in Spir-1 actin organizer mutant mice. *Eur. J. Cell Biol.* **93**, 225–237.
29. Mann, G., Mora, S., Madu, G., Adegoke, O.A.J., (2021). Branched-chain amino acids: catabolism in skeletal muscle and implications for muscle and whole body metabolism. *Front. Physiol.* **12**, 702826
30. Diederichs, S., (2014). The four dimensions of noncoding RNA conservation. *Trends Genet.* **30**, 121–123.
31. Bolger, A.M., Lohse, M., Usadel, B., (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
32. Langmead, B., Salzberg, S.L., (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359.
33. Zerbino, D.R. et al, (2018). Ensembl 2018. *Nucleic Acids Res.* **46**, 754–761.
34. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., et al., (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
35. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079.
36. Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31**, 46–53.
37. Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., Marth, G.T., (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692.
38. Quinlan, A.R., Hall, I.M., (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
39. Martone, J., Briganti, F., Legnini, I., Morlando, M., Picillo, E., Sthandier, O., Politano, L., Bozzoni, I., (2016). The lack of the Celf2a splicing factor converts a Duchenne genotype into a Becker phenotype. *Nature Commun.* **7**, 10488.